I837592

(19)中華民國智慧財產局

(12)發明說明書公告本 **(11)證書號數：TW I837592 B**

(45)公告日： 中華民國 113 (2024) 年 04 月 01 日

(21)申請案號：111105280 (22)申請日： 中華民國 105 (2016) 年 06 月 20 日

(51)Int. Cl.： ***C12N15/11 (2006.01)*** ***C12N15/10 (2006.01)***
C12N15/85 (2006.01) ***C12N15/82 (2006.01)***
C12N9/22 (2006.01)

(30)優先權：2015/06/18 美國 62/181,739
2015/07/16 美國 62/193,507
2015/08/05 美國 62/201,542
2015/08/16 美國 62/205,733
2015/09/24 美國 62/232,067
2015/12/18 美國 14/975,085
2016/01/07 歐洲專利局 16150428.7

(71)申請人：美商博得學院股份有限公司 (美國) THE BROAD INSTITUTE INC. (US)
美國
麻省理工學院 (美國) MASSACHUSETTS INSTITUTE OF TECHNOLOGY (US)
美國
哈佛大學校董委員會 (美國) PRESIDENT AND FELLOWS OF HARVARD COLLEGE (US)
美國

(72)發明人：任 恩方 ZHANG, FENG (US)；任史奇 彼爾德 ZETSCHE, BERND (DE)；古登柏 強納生 GOOTENBERG, JONATHAN (US)；阿布達耶 歐瑪O ABUDAYYEH, OMAR O. (US)；史拉馬可爾 伊恩 SLAYMAKER, IAN (US)

(74)代理人：陳昭誠

(56)參考文獻：

TW 201425580A US 8889356B2

期刊 Gisle Vestergaard, Roger A Garrett & Shiraz A Shah CRISPR adaptive immune systems of Archaea, RNA Biology, 11:2, (2014) 156-167

審查人員：吳思嫻

申請專利範圍項數：40 項 圖式數：116 共 2291 頁

(54)名稱

新型ＣＲＩＳＰＲ酶以及系統

(57)摘要

本發明提供了用於靶向核酸的系統、方法和組成物。具體地說，本發明提供了包含新型 DNA-靶向 CRISPR 效應蛋白和至少一種靶向核酸組分如指導 RNA 的非天然存在的或工程化的 DNA 靶向系統。在此還揭露並且要求保護用於製備和使用此類系統、方法和組成物之方法和此類系統、方法和組成物的用途以及來自此類方法和用途之產物。

The invention provides for systems, methods, and compositions for targeting nucleic acids. In particular, the invention provides non-naturally occurring or engineered DNA-targeting systems comprising a novel DNA-targeting CRISPR effector protein and at least one targeting nucleic acid component like a guide RNA. Methods for making and using and uses of such systems, methods, and compositions and products from such methods and uses are also disclosed and claimed.

公告本

I837592

發明摘要

※ 申請案號：

※ 申請日：　　※IPC 分類：

【發明名稱】(中文/英文)

新型CRISPR酶以及系統

NOVEL CRISPR ENZYMES AND SYSTEMS

【中文】

本發明提供了用於靶向核酸的系統、方法和組成物。具體地說，本發明提供了包含新型DNA-靶向CRISPR效應蛋白和至少一種靶向核酸組分如指導RNA的非天然存在的或工程化的DNA靶向系統。在此還揭露並且要求保護用於製備和使用此類系統、方法和組成物之方法和此類系統、方法和組成物的用途以及來自此類方法和用途之產物。

【英文】

The invention provides for systems, methods, and compositions for targeting nucleic acids. In particular, the invention provides non-naturally occurring or engineered DNA-targeting systems comprising a novel DNA-targeting CRISPR effector protein and at least one targeting nucleic acid component like a guide RNA. Methods for making and using and uses of such systems, methods, and compositions and products from such methods and uses are also disclosed and claimed.

【代表圖】

【本案指定代表圖】：無。

【本代表圖之符號簡單說明】：無。

本案的圖皆為實驗數據，並非本案的代表圖。故本案無指定代表圖。

【本案若有化學式時，請揭示最能顯示發明特徵的化學式】：

無。

發明專利說明書

（本說明書格式、順序，請勿任意更動）

【發明名稱】（中文/英文）

新型CRISPR酶以及系統

NOVEL CRISPR ENZYMES AND SYSTEMS

[相關申請並藉由引用結合]

[0001] 本申請要求於 2015 年 6 月 18 日提交的美國臨時 62/181,739；於 2015 年 7 月 16 日提交的美國臨時 62/193,507；2015 年 8 月 5 日提交的美國臨時 62/201,542；2015 年 8 月 16 日提交的美國臨時 62/205,733；2015 年 9 月 24 日提交的美國臨時 62/232,067;2015 年 12 月 18 日提交的美國申請案序號 14/975,085 以及歐洲申請案號 16150428.7 的權益和優先權。

[0002] 將前述申請及其中或它們的訴訟期間（“應用引用文獻”）所引用的所有文獻或在此引用文獻中引用或參考的所有文獻，以及在此或在藉由引用結合在此的任何文獻中提到的任何產品的任何廠商說明書、描述、產品規格和產品清單均藉由引用結合在此，並且可以應用於本發明的實踐中。更具體地，所有參考的文獻藉由引用結合在此，其程度如同將每個單獨的文獻具體並單獨地指明藉由引用結合在此。

[聯邦資助研究的聲明]

[0003] 本發明係根據由美國國立衛生研究院（National Institutes of Health）授予的批准號MH100706在政府支持下進行。

美國政府享有本發明的某些權利。

[序列表]

[0004] 本申請含有一份已經以ASCII格式電子遞交的序列表並且該序列表藉由引用以其整體結合在此。創建於2015年12月17日的所述ASCII副本名稱為47627.05.2123_SL.txt並且大小為2,467,205位組。

【技術領域】

[0005] 本發明總體涉及用於控制涉及序列靶向諸如基因轉錄物的干擾或核酸編輯的基因表現的系統、方法以及組成物，該等系統、方法以及組成物可以使用與成簇規律間隔短迴文重複序列（CRISPR）及其組分相關的載體系統。

【先前技術】

[0006] 基因體定序技術和分析方法的最新進展明顯加速了對與不同範圍的生物學功能和疾病相關聯的遺傳因子進行編目和映射的能力。精確的基因組靶向技術對於藉由允許個體遺傳元件的選擇性干擾而使得因果性遺傳變異的系統性逆向工程成為可能，以及推進合成生物學、生物技術學和醫療應用係需要的。雖然基因組編輯技術諸如設計師的鋅指、轉錄活化物樣效應物（TALE）或歸巢大範圍核酸酶（homing meganucleases）可以用於產生靶向基因組干擾，但是仍然需要採用新策略和分子機制並且負擔的起的、易於建立的、可擴展的且便於靶向真核基因組內的多個位置的新基因組工程技術。這將為基因組工程和生物技術的新應用提

供主要資源。

[0007] 細菌和古細菌的（archaeal）適應性免疫的CRISPR-Cas系統顯示出蛋白質組成和基因組座位體系結構的極端多樣性。CRISPR-Cas系統座位具有超過50種的基因家族並且不存在嚴格的通用基因，這表明了座位體系結構的快速進化和極端多樣性。到目前為止，採用了多分支方法，針對93種Cas蛋白存在約395種表現譜的全面*cas*基因鑒定。分類包括特徵基因表現譜加上座位體系結構的特徵。提出了新的CRISPR-Cas系統分類，其中該等系統寬泛地分成兩類，具有多亞基效應物複合物的第1類和具有單亞基效應物模組的第2類，藉由Cas9蛋白來舉例說明。與第2類CRISPR-Cas系統相關聯的新型效應蛋白可以被開發為強有力的基因組工程工具並且推定的新型效應蛋白的預測及其工程化和優化係重要的。

[0008] 本申請案中的任何文獻的引用或鑒定並不承認該文獻作為本發明的先前技術而可以獲得。

【發明內容】

[0009] 對於具有一系列廣泛應用的靶向核酸或多核苷酸(例如，DNA或RNA或其任何雜交體或衍生物）的替代性且穩健的系統和技術存在著迫切需要。本發明著手解決這種需要並且提供了相關優點。將本發明的新型DNA或RNA靶向系統添加到基因組和表觀基因組（epigenomic）靶向技術的全能文庫（repertoire）可以藉由直接的檢測、分析和操縱來轉化特定靶位點的研究和干擾或編輯。為了有效而無有害作用地利用本發明的DNA或RNA靶向系

統用於基因組或表觀基因組的靶向，瞭解該等DNA或RNA靶向工具的工程化和優化方面是關鍵的。

[0010] 本發明提供了一種修飾與感興趣的靶座位相關聯的或在該靶座位處的序列之方法，該方法包括將包含推定的V型CRISPR-Cas座位效應蛋白和一種或多種核酸組分的非天然存在或工程化的組成物遞送至所述座位，其中效應蛋白與一種或多種核酸組分形成複合物並且在所述複合物與感興趣的座位結合後，效應蛋白誘導對與感興趣的靶座位相關聯的或在該靶座位處的序列的修飾。在一較佳的實施方式中，修飾係股斷裂的引入。在一較佳的實施方式中，與感興趣的靶座位相關聯的或在該靶座位處的序列包括DNA並且效應蛋白由亞型V-A CRISPR-Cas座位或亞型V-B CRISPR-Cas座位編碼。

[0011] 應瞭解，除非另外表明，否則術語Cas酶、CRISPR酶、CRISPR蛋白、Cas蛋白和CRISPR Cas通常是可以互換使用的並且在所有在此參考方面處以類推方式指進一步描述在本發明中的新型CRISPR效應蛋白，諸如藉由具體參考Cas9。在此所述的CRISPR效應蛋白較佳的是Cpf1效應蛋白。

[0012] 本發明提供了一種修飾與感興趣的靶座位相關聯的或在該靶座位處的序列之方法，該方法包括將包含Cpf1座位效應蛋白和一種或多種核酸組分的非天然存在或工程化的組成物遞送至與座位相關聯或在該座位處的所述序列，其中Cpf1效應蛋白與一種或多種核酸組分形成複合物並且在所述複合物與感興趣的座位結合後，效應蛋白誘導對與感興趣的靶座位相關聯的或在該靶

座位處的序列的修飾。在一較佳的實施方式中，修飾係股斷裂的引入。在一較佳的實施方式中，Cpf1效應蛋白與核酸組分；有利地工程化的或非天然存在的核酸組分形成複合物。對與感興趣的靶座位相關聯的或在該靶座位處的序列的修飾的誘導可以是Cpf1效應蛋白-核酸指導的。在一較佳的實施方式中，核酸組分係CRISPR RNA（crRNA）。在一較佳的實施方式中，核酸組分係成熟crRNA或指導RNA，其中成熟crRNA或指導RNA包含間隔區序列（或指導序列）和同向重複序列或它們的衍生物。在一較佳的實施方式中，間隔區序列或其衍生物包含種子序列，其中種子序列對識別和/或雜交於靶座位處的序列係關鍵的。在一較佳的實施方式中，FnCpf1指導RNA的種子序列大約在間隔區序列（或指導序列）的5'端上的前5個核苷酸（nt）之內。在一較佳的實施方式中，股斷裂係交錯切割的，產生了5'突出端。在一較佳的實施方式中，與感興趣的靶座位相關聯的或在該靶座位處的序列包括直鏈DNA或超螺旋DNA。

[0013] 本發明的方面涉及具有一種或多種非天然存在或工程化或修飾或優化的核酸組分的Cpf1效應蛋白複合物。在一較佳的實施方式中，複合物的核酸組分可以包含連接至同向（direct）重複序列的指導序列，其中同向重複序列包括一個或多個莖環或優化的二級結構。在一較佳的實施方式中，同向重複序列具有16個核苷酸的最小長度並且具有單一莖環。在另外的實施方式中，同向重複序列具有長於16個核苷酸，較佳的是（preferrably）超過17個核苷酸的長度，並且具有超過一個的莖環或優化的二級結構。在一較佳的實施方式中，同向重複序列可以被修飾成包含一種或

多種蛋白質結合的RNA適配體。在一較佳的實施方式中，一個或多個適配體可以被包含作為優化的二級結構的一部分。此類適配體可以能夠結合噬菌體外殼蛋白。噬菌體外殼蛋白可以選自下組，該組包括下項：Qβ、F2、GA、fr、JP501、MS2、M12、R17、BZ13、JP34、JP500、KU1、M11、MX1、TW18、VK、SP、FI、ID2、NL95、TW19、AP205、ϕCb5、ϕCb8r、ϕCb12r、ϕCb23r、7s和PRR1。在一較佳的實施方式中，噬菌體外殼蛋白係MS2。本發明還提供了複合物的核酸組分，該核酸組分的長度為30或更多個、40或更多個，或50或更多個核苷酸。

[0014] 本發明提供了基因組編輯之方法，其中該方法包括兩輪或更多輪的Cpf1效應蛋白靶向和切割。在某些實施方式中，第一輪包括Cpf1效應蛋白切割與遠離種子序列的靶座位相關聯的序列並且第二輪包括Cpf1效應蛋白切割靶座位處的序列。在本發明的較佳的實施方式中，藉由Cpf1效應蛋白進行的第一輪靶向產生了indel並且藉由Cpf1效應蛋白進行的第二輪靶向可以經由同源定向修復（HDR）進行修復。在本發明的一個最較佳的實施方式中，藉由Cpf1效應蛋白進行的一輪或多輪靶向產生了可以藉由修復模板的插入來修復的交錯切割。

[0015] 本發明提供了基因組編輯或修飾與感興趣的靶座位相關聯的或在該靶座位處的序列之方法，其中該方法包括將Cpf1效應蛋白複合物引入到任何所希望的細胞類型，原核細胞或真核細胞中，由此Cpf1效應蛋白複合物有效地用於將DNA插入物整合到原核細胞或真核細胞的基因組中。在較佳的實施方式中，細胞

係真核細胞並且基因組係哺乳動物基因組。在較佳的實施方式中，DNA插入物的整合藉由基於非同源末端連接（NHEJ）的基因插入機制來實現。在較佳的實施方式中，DNA插入物係外源引入的DNA模板或修復模板。在一個較佳的實施方式中，外源引入的DNA模板或修復模板與Cpf1效應蛋白複合物或一種組分或用於表現複合物組分的多核苷酸載體一起遞送。在一個更較佳的實施方式中，真核細胞係不分裂的細胞（例如，其中經由HDR進行基因組編輯係特別具有挑戰性的不分裂細胞）。在人類細胞中的基因組編輯的較佳的方法中，Cpf1效應蛋白可以包括但不限於FnCpf1、AsCpf1和LbCpf1效應蛋白。

[0016] 本發明還提供了一種修飾感興趣的靶座位之方法，該方法包括將包含C2c1座位效應蛋白和一種或多種核酸組分的非天然存在或工程化的組成物遞送至所述座位，其中C2c1效應蛋白與一種或多種核酸組分形成複合物並且在所述複合物與感興趣的座位結合後，效應蛋白誘導對感興趣的靶座位的修飾。在一較佳的實施方式中，修飾係股斷裂的引入。

[0017] 在此類方法中，感興趣的靶座位可以包含在體外的DNA分子中。在一較佳的實施方式中，DNA分子係質粒。

[0018] 在此類方法中，感興趣的靶座位可以包含在細胞內的DNA分子中。細胞可以是原核細胞或真核細胞。細胞可以是哺乳動物細胞。哺乳動物細胞可以是非人類靈長類動物、牛、豬、齧齒動物或小鼠細胞。細胞可以是非哺乳動物真核細胞諸如家禽、魚或蝦的細胞。細胞還可以是植物細胞。植物細胞可以是栽培植

物諸如木薯、玉米、高粱、小麥或稻具有的細胞。植物細胞還可以是藻類、樹或蔬菜具有的細胞。藉由本發明引入到細胞的修飾可以使得細胞和細胞的子代被改變以改進生物產物諸如抗體、澱粉、乙醇或其他所希望的細胞輸出物的產生。藉由本發明引入到細胞的修飾可以使得細胞和細胞的子代包括使所產生的生物產物發生變化的改變。

[0019] 本發明提供了一種修飾感興趣的靶座位之方法，該方法包括將包含VI型CRISPR-Cas座位效應蛋白和一種或多種核酸組分的非天然存在或工程化的組成物遞送至所述座位，其中效應蛋白與一種或多種核酸組分形成複合物並且在所述複合物與感興趣的座位結合後，效應蛋白誘導對感興趣的靶座位的修飾。在一較佳的實施方式中，修飾係股斷裂的引入。

[0020] 在一較佳的實施方式中，感興趣的靶座位包括DNA。

[0021] 在此類方法中，感興趣的靶座位可以包含在細胞內的DNA分子中。細胞可以是原核細胞或真核細胞。細胞可以是哺乳動物細胞。哺乳動物細胞可以是非人類哺乳動物，例如靈長類動物、牛、羊、豬類、犬、齧齒動物、兔科諸如猴、母牛、綿羊、豬、狗、兔、大鼠或小鼠的細胞。細胞可以是非哺乳動物真核細胞諸如家禽鳥類（例如雞）、脊椎動物魚（例如鮭魚）或甲殼類動物（例如牡蠣、蛤（claim）、龍蝦、蝦）的細胞。細胞還可以是植物細胞。植物細胞可以是單子葉植物或雙子葉植物具有的細胞或栽培植物或糧食植物諸如木薯、玉米、高粱、大豆、小麥、燕麥或稻具有的細胞。植物細胞還可以是藻類、樹或生產植物、果實或蔬菜（例

如，樹類諸如柑橘樹，例如桔子樹、葡萄柚樹或檸檬樹；桃樹或油桃樹；蘋果樹或梨樹；堅果樹諸如杏樹或核桃樹或阿月渾子樹；茄屬植物；芸苔屬植物；萵苣屬植物；菠菜屬植物；辣椒屬植物；棉花、煙草、蘆筍、胡蘿蔔、甘藍、青花菜、花椰菜、番茄、茄子、胡椒、萵苣、菠菜、草莓、藍莓、覆盆子、黑莓、葡萄、咖啡、可可等）具有的細胞。

[0022] 在任一所述方法中，感興趣的靶座位可以是感興趣的基因組或表觀基因組的座位。在任一所述方法中，複合物可以使用用於多重用途的多個指導序列進行遞送。在任一所述方法中，可以使用超過一種的蛋白質。

[0023] 在本發明的較佳的實施方式中，在不存在推定的反式啟動crRNA（tracr RNA）序列條件下，發生與感興趣的靶座位相關聯的或在該靶座位處的序列的生物化學的或體外或體內的切割，例如藉由FnCpf1效應蛋白進行切割。在本發明的其他實施方式中，在存在推定的反式啟動crRNA（tracr RNA）序列條件下，可以發生切割，例如藉由其他CRISPR家族效應蛋白進行切割，然而，在評價FnCpf1座位之後，申請人推斷藉由Cpf1效應蛋白複合物進行的靶DNA切割不需要tracrRNA。申請人確定僅包含Cpf1效應蛋白和crRNA（包含同向重複序列和指導序列的指導RNA）的Cpf1效應蛋白複合物足以切割靶DNA。因此，本發明提供了修飾如上文所述的感興趣的靶座位之方法，其中效應蛋白係Cpf1蛋白並且效應蛋白在不存在tracr的條件下與靶序列複合。

[0024] 在任一上述方法中，效應蛋白（例如Cpf1）和核酸組

分可以經由編碼該蛋白質和/或一種或多種核酸組分的一個或多個多核苷酸分子來提供，並且其中一個或多個多核苷酸分子被可操作地構造成用於表現蛋白和/或一種或多種核酸組分。一個或多個多核苷酸分子可以包含被可操作地構造成用於表現蛋白質和/或一種或多種核酸組分的一個或多個調節元件。一個或多個多核苷酸分子可以包含在一個或多個載體中。本發明包括此或此類核苷酸分子例如被可操作地構造成用於表現蛋白質的此類多核苷酸分子，和/或一種或多種核酸組分以及此或此類載體。

[0025] 在任一上述方法中，股斷裂可以是單股斷裂或雙股斷裂。

[0026] 調節元件可以包括誘導型啟動子。多核苷酸和/或載體系統可以包括誘導型系統。

[0027] 在任一上述方法中，一個或多個多核苷酸分子可以包含在遞送系統中，或者一種或多種載體可以包含在遞送系統中。

[0028] 在任一上述方法中，非天然存在或工程化的組成物可以經由脂質體、粒子（例如奈米粒子）、外來體(exosome)、微泡、基因槍或一種或多種載體例如核酸分子或病毒載體遞送。

[0029] 本發明還提供了一種非天然存在或工程化的組成物，該組成物係具有如在此所討論的或任一在此所述方法中所限定的特性的組成物。

[0030] 本發明還提供了一種包含一種或多種載體的載體系統，該等一種或多種載體包含編碼非天然存在或工程化的組成物

（為具有如在此所討論的或任一在此所述方法中所限定的特性的組成物）的組分的一個或多個多核苷酸分子。

[0031] 本發明還提供了一種包含一種或多種載體或一個或多個多核苷酸分子的遞送系統，該等一種或多種載體或一個或多個多核苷酸分子包括編碼非天然存在或工程化的組成物（為具有如在此所討論的或任一在此所述方法中所限定的特性的組成物）的組分的一個或多個多核苷酸分子。

[0032] 本發明還提供了在治療性治療方法中使用的一種非天然存在或工程化的組成物，或編碼所述組成物的組分的一種或多種多核苷酸，或包括編碼所述組成物的組分的一種或多種多核苷酸的載體或遞送系統。治療性治療方法可以包括基因或基因組編輯，或基因治療。

[0033] 本發明還涵蓋用於預測新的第2類CRISPR-Cas系統和鑒定其中組分的計算方法和演算法。

[0034] 本發明還提供了其中效應蛋白的一個或多個胺基酸殘基可以被修飾，例如工程化或非天然存在的效應蛋白或Cpf1的方法和組成物。在一實施方式中，修飾可以包括效應蛋白的一個或多個胺基酸殘基的突變。一個或多個突變可以處於效應蛋白的一個或多個催化活性結構域中。與缺乏所述一個或多個突變的效應蛋白相比，該效應蛋白可以具有降低或廢除的核酸酶活性。效應蛋白不可以引導感興趣的靶座位處的一條或另一條DNA或RNA股的切割。效應蛋白不可以引導感興趣的靶座位處的DNA或RNA股的切割。在一較佳的實施方式中，一個或多個突變可以包括兩

個突變。在一較佳的實施方式中，Cpf1效應蛋白中的一個或多個胺基酸殘基被修飾，例如工程化或非天然存在的效應蛋白或Cpf1。在一較佳的實施方式中，Cpf1效應蛋白係FnCpf1效應蛋白。在一較佳的實施方式中，一個或多個修飾的或突變的胺基酸殘基係參照FnCpf1效應蛋白的胺基酸位置編碼的D917A、E1006A或D1255A。在另外的較佳的實施方式中，一種或多種突變胺基酸殘基係參照AsCpf1中的胺基酸位置的D908A、E993A、D1263A或者是參照LbCpf1中的胺基酸位置的LbD832A、E925A、D947A或D1180A。

[0035] 本發明還提供了處於包含RuvC結構域的效應蛋白的催化活性結構域中的一個或多個突變或兩個或更多個突變。在本發明的一些實施方式中，RuvC結構域可以包括RuvCI、RuvCII或RuvCIII結構域，或與RuvCI、RuvCII或RuvCIII結構域等或與如任一在此所述方法中所述的任何相關結構域同源的催化活性結構域。效應蛋白可以包含一個或多個異源功能結構域。一個或多個異源功能結構域可以包括一個或多個核定位信號（NLS）結構域。一個或多個異源功能結構域可以包括至少兩個或更多個NLS結構域。一個或多個NLS結構域可以定位成處於或靠近或接近效應蛋白（例如Cpf1）的末端，並且如果是兩個或更多個NLS的話，則兩個中的每個可以定位成處於或靠近或接近效應蛋白（例如Cpf1）的末端。一個或多個異源功能結構域可以包括一個或多個轉錄啟動結構域。在一較佳的實施方式中，轉錄啟動結構域可以包括VP64。一個或多個異源功能結構域可以包括一個或多個轉錄阻遏結構域。在一較佳的實施方式中，轉錄阻遏結構域包括KRAB結構域或SID結構域（例如SID4X）。一個或多個異源功能結構域可以包括

一個或多個核酸酶結構域。在一較佳的實施方式中，核酸酶結構域包括Fok1。

[0036] 本發明還提供了具有以下活性中的一種或多種的一個或多個異源功能結構域：修飾烷酶活性、去甲酶活性、轉錄啟動活性、轉錄阻遏活性、轉錄釋放因子活性、組蛋白修飾活性、核酸酶活性、單股RNA切割活性、雙股RNA切割活性、單股DNA切割活性、雙股DNA切割活性以及核酸結合活性。至少一個或多個異源功能結構域可以處於或靠近效應蛋白的胺基末端並且/或者其中至少一個或多個異源功能結構域處於或靠近效應蛋白的羧基末端。一個或多個異源功能結構域可以融合至效應蛋白。一個或多個異源功能結構域可以系接至效應蛋白。一個或多個異源功能結構域可以藉由接頭部分連接至效應蛋白。

[0037] 本發明還提供了效應蛋白（例如Cpf1），包括來自於來自包括下項的屬的生物體的效應蛋白(例如Cpf1)：鏈球菌屬（*Streptococcus*）、彎曲桿菌屬（*Campylobacter*）、硝化裂化菌屬（*Nitratifractor*）、葡萄球菌屬（*Staphylococcus*）、細小棒菌屬（*Parvibaculum*）、羅氏菌屬(*Roseburia*)、奈瑟氏菌屬(*Neisseria*)、葡糖醋桿菌屬（*Gluconacetobacter*）、固氮螺菌屬（*Azospirillum*）、螺旋體屬（*Sphaerochaeta*）、乳酸桿菌屬（*Lactobacillus*）、真細菌屬（*Eubacterium*）、棒狀桿菌屬（*Corynebacter*）、肉桿菌屬（*Carnobacterium*）、紅細菌屬(*Rhodobacter*)、李斯特菌屬(*Listeria*)、帕魯迪菌屬（*Paludibacter*）、梭菌屬（*Clostridium*）、毛螺旋菌科（*Lachnospiraceae*）、梭狀菌屬（*Clostridiaridium*）、纖毛菌屬（*Leptotrichia*）、弗朗西絲菌屬（*Francisella*）、軍團桿菌屬

（*Legionella*）、脂環酸芽孢桿菌屬（*Alicyclobacillus*）、甲烷嗜甲基菌（*Methanomethyophilus*）、卟啉單胞菌屬（*Porphyromonas*）、普雷沃菌屬（*Prevotella*）、擬桿菌門（*Bacteroidetes*）、創傷球菌屬（*Helcococcus*）、鉤端螺旋體屬（*Letospira*）、脫硫弧菌屬（*Desulfovibrio*）、脫硫鹽鹼桿菌屬（*Desulfonatronum*）、豐祐菌科（*Opitutaceae*）、腫塊芽孢桿菌屬（*Tuberibacillus*）、芽孢桿菌屬（*Bacillus*）、短芽孢桿菌屬（*Brevibacilus*）、甲基桿菌屬（*Methylobacterium*）或胺基酸球菌屬（*Acidaminococcus*）。

[0038] 本發明還提供了效應蛋白（例如Cpf1），包括來自於來自下項的生物體的效應蛋白（例如Cpf1）：變異鏈球菌（*S. mutans*）、無乳鏈球菌（*S. agalactiae*）、似馬鏈球菌（*S. equisimilis*）、血鏈球菌（*S. sanguinis*）、肺炎鏈球菌（*S. pneumonia*）；空腸彎曲桿菌（*C. jejuni*）、大腸彎曲桿菌（*C. coli*）；鹽水硝化破壞菌（*N. salsuginis*）、替加硝化破壞菌（*N. tergarcus*）；耳葡萄球菌（*S. auricularis*）、肉葡萄球菌（*S. carnosus*）；腦膜炎奈瑟氏菌（*N. meningitides*）、淋病奈瑟氏菌（*N. gonorrhoeae*）；單核增生李斯特菌（*L. monocytogenes*）、伊氏李斯特菌（*L. ivanovii*）；肉毒梭菌（*C. botulinum*）、艱難梭菌（*C. difficile*）、破傷風梭菌（*C. tetani*）、索氏梭菌（*C. sordellii*）。

[0039] 效應蛋白可以包括嵌合效應蛋白，該嵌合效應蛋白包含來自第一效應蛋白（例如Cpf1）異種同源物的第一片段和來自第二效應蛋白（例如Cpf1）異種同源物的第二片段，並且其中第一效應蛋白異種同源物和第二效應蛋白異種同源物是不同的。第一效應蛋白（例如Cpf1）異種同源物和第二效應蛋白（例如Cpf1）異種同源物中的至少一者可以包括來自於包括下項的生物體的效

應蛋白（例如Cpf1）：鏈球菌屬、彎曲桿菌屬、硝化裂化菌屬、葡萄球菌屬、細小棒菌屬、羅氏菌屬、奈瑟氏菌屬、葡糖醋桿菌屬、固氮螺菌屬、螺旋體屬、乳酸桿菌屬、真細菌屬、棒狀桿菌屬、肉桿菌屬、紅細菌屬、李斯特菌屬、帕魯迪菌屬、梭菌屬、毛螺旋菌科、梭狀菌屬、纖毛菌屬、弗朗西絲菌屬、軍團桿菌屬、脂環酸芽孢桿菌屬、甲烷嗜甲基菌、卟啉單胞菌屬、普雷沃菌屬、擬桿菌門、創傷球菌屬、鉤端螺旋體屬、脫硫弧菌屬、脫硫鹽鹼桿菌屬、豐祐菌科、腫塊芽孢桿菌屬、芽孢桿菌屬、短芽孢桿菌屬、甲基桿菌屬或胺基酸球菌屬；例如包含第一片段和第二片段的嵌合效應蛋白，其中第一片段和第二片段中的每個選自包括下項的生物體的Cpf1：鏈球菌屬、彎曲桿菌屬、硝化裂化菌屬、葡萄球菌屬、細小棒菌屬、羅氏菌屬、奈瑟氏菌屬、葡糖醋桿菌屬、固氮螺菌屬、螺旋體屬、乳酸桿菌屬、真細菌屬、棒狀桿菌屬、肉桿菌屬、紅細菌屬、李斯特菌屬、帕魯迪菌屬、梭菌屬、毛螺旋菌科、梭狀菌屬、纖毛菌屬、弗朗西絲菌屬、軍團桿菌屬、脂環酸芽孢桿菌屬、甲烷嗜甲基菌、卟啉單胞菌屬、普雷沃菌屬、擬桿菌門、創傷球菌屬、鉤端螺旋體屬、脫硫弧菌屬、脫硫鹽鹼桿菌屬、豐祐菌科、腫塊芽孢桿菌屬、芽孢桿菌屬、短芽孢桿菌屬、甲基桿菌屬或胺基酸球菌屬，其中第一片段和第二片段並非來自相同細菌；例如，包含第一片段和第二片段的嵌合效應蛋白，其中第一片段和第二片段中的每個選自下項的Cpf1：變異鏈球菌、無乳鏈球菌、似馬鏈球菌、血鏈球菌、肺炎鏈球菌；空腸彎曲桿菌、大腸彎曲桿菌；鹽水硝化破壞菌、替加硝化破壞菌；耳葡萄球菌、肉葡萄球菌；腦膜炎奈瑟氏菌、淋病奈瑟氏菌；單核增生李斯特菌、伊氏李斯特菌；肉毒梭菌、艱難梭菌、破傷風梭菌、索氏梭菌；土拉熱弗朗西絲菌（*Francisella tularensis*）1、易北普雷沃菌（*Prevotella*

albensis）、毛螺旋菌科細菌（Lachnospiraceae bacterium）*MC2017 1*、解肽丁酸弧菌（*Butyrivibrio proteoclasticus*）、佩萊格裡尼菌科細菌（*Peregrinibacteria bacterium*）*GW2011_GWA2_33_10*、儉菌總門（*Parcubacteria*）細菌*GW2011_GWC2_44_17*、密斯氏菌屬（*Smithella*）某種*SCADC*、胺基酸球菌屬某種*BV3L6*、毛螺旋菌科細菌*MA2020*、候選白蟻甲烷枝原體（*Candidatus Methanoplasma termitum*）、挑剔真細菌（*Eubacterium eligens*）、牛莫拉氏菌（*Moraxella bovoculi*）*237*、稻田鉤端螺旋體（*Leptospira inadai*）、毛螺旋菌科細菌*ND2006*、狗口腔卟啉單胞菌（*Porphyromonas crevioricanis*）*3*、解糖胨普雷沃菌（*Prevotella disiens*）和獼猴卟啉單胞菌（*Porphyromonas macacae*），其中第一片段和第二片段並非來自相同細菌。

[0040] 在本發明的較佳的實施方式中，效應蛋白來源於Cpf1座位（在此此類效應蛋白也稱之為“Cpf1p”），例如Cpf1蛋白（並且此類效應蛋白或Cpf1蛋白或來源於Cpf1座位的蛋白質也稱之為“CRISPR酶”）。Cpf1座位包括但不限於圖64中列出的細菌物種的Cpf1座位。在一更較佳的實施方式中，Cpf1p來源於選自下項的細菌物種：土拉熱弗朗西絲菌*1*、易北普雷沃菌、毛螺旋菌科細菌*MC2017 1*、解肽丁酸弧菌、佩萊格裡尼菌科細菌*GW2011_GWA2_33_10*、儉菌總門細菌*GW2011_GWC2_44_17*、密斯氏菌屬某種*SCADC*、胺基酸球菌屬某種*BV3L6*、毛螺旋菌科細菌*MA2020*、候選白蟻甲烷枝原體、挑剔真細菌、牛莫拉氏菌*237*、稻田鉤端螺旋體、毛螺旋菌科細菌*ND2006*、狗口腔卟啉單胞菌*3*、解糖胨普雷沃菌和獼猴卟啉單胞菌。在某些實施方式中，Cpf1p來源於選自胺基酸球菌屬某種*BV3L6*、毛螺旋菌科細菌*MA2020*的細

菌物種。在某些實施方式中，效應蛋白來源於土拉熱弗朗西絲菌*1*的亞種，包括但不限於土拉熱弗朗西絲菌新殺手(*Novicida*)亞種。

[0041] 在本發明的另外的實施方式中，原型間隔區鄰近模體（PAM）或PAM-樣模體引導效應蛋白複合物與感興趣的靶座位的結合。在本發明的一個較佳的實施方式中，PAM係5' TTN，其中N係A/C/G或T並且效應蛋白係FnCpf1p。在本發明的另一個較佳的實施方式中，PAM係5' TTTV，其中V係A/C或G並且效應蛋白係AsCpf1、LbCpf1或PaCpf1p。在某些實施方式中，PAM係5' TTN，其中N係A/C/G或T，效應蛋白係FnCpf1p，並且PAM位於原型間隔區的5'端的上游。在本發明的某些實施方式中，PAM係5' CTA，其中效應蛋白係FnCpf1p，並且PAM位於原型間隔區或靶座位的5'端的上游。在較佳的實施方式中，本發明提供了用於RNA指導的基因組編輯核酸酶的擴大的靶向範圍，其中Cpf1家族的富含T的PAM允許對富含AT基因組的靶向和編輯。

[0042] 在某些實施方式中，CRISPR酶被工程化並且可以包含降低或消除核酸酶活性的一個或多個突變。FnCpf1p RuvC結構域中的胺基酸位置包括但不限於D917A、E1006A、E1028A、D1227A、D1255A、N1257A、D917A、E1006A、E1028A、D1227A、D1255A和N1257A。申請人還鑒定了與PD-(D/E)XK核酸酶超家族和HincII內切核酸酶樣最類似的推定的第二核酸酶結構域。在此推定的核酸酶結構域中產生的大幅度降低核酸酶活性的點突變包括但不限於N580A、N584A、T587A、W609A、D610A、K613A、E614A、D616A、K624A、D625A、K627A和Y629A。在一較佳的實施方式

中，FnCpf1p RuvC結構域中的突變係D917A或E1006A，其中D917A或E1006A突變使FnCpf1效應蛋白的DNA切割活性完全失活。在另一個實施方式中，FnCpf1p RuvC結構域中的突變係D1255A，其中突變的FnCpf1效應蛋白具有明顯降低的核溶解活性。

[0043] AsCpf1p RuvC結構域中的胺基酸位置包括但不限於908、993和1263。在一較佳的實施方式中，AsCpf1p RuvC結構域中的突變係D908A、E993A和D1263A，其中D908A、E993A和D1263A突變使AsCpf1效應蛋白的DNA切割活性完全失活。LbCpf1p RuvC結構域中的胺基酸位置包括但不限於832、947或1180。在一較佳的實施方式中，LbCpf1p RuvC結構域中的突變係LbD832A、E925A、D947A或D1180A，其中LbD832A、E925A、D947A或D1180A突變使LbCpf1效應蛋白的DNA切割活性完全失活。

[0044] 突變還可以在鄰近殘基處，例如在靠近以上指出的參與核酸酶活性的那些的胺基酸處形成。在一些實施方式中，僅RuvC結構域係失活的，並且在其他實施方式中，另一推定的核酸酶結構域係失活的，其中效應蛋白複合物充當切口酶並且僅切割一條DNA股。在一較佳的實施方式中，其他推定的核酸酶結構域係HincII樣內切核酸酶結構域。在一些實施方式中，使用兩種FnCpf1、AsCpf1或LbCpf1變體（各自不同的切口酶）來增加特異性，使用兩種切口酶變體來切割靶標處的DNA（其中兩種切口酶切割DNA股，同時使脫靶修飾最小化或消除，其中僅一條DNA股被切割並且隨後進行修復）。在較佳的實施方式中，Cpf1效應蛋白

以包含兩個Cpf1效應蛋白分子的同源二聚體形式切割與感興趣的靶座位相關聯的或在該靶座位處的序列。在一較佳的實施方式中，同源二聚體可以包含在其對應RuvC結構域中含有不同的突變的兩個Cpf1效應蛋白分子。

[0045] 本發明涵蓋使用兩種或更多種切口酶的方法，具體地雙或雙重切口酶方法。在一些方面和實施方式中，可以遞送單一類型的FnCpf1、AsCpf1或LbCpf1切口酶，例如如在此所述的修飾的FnCpf1、AsCpf1或LbCpf1或修飾的FnCpf1、AsCpf1或LbCpf1切口酶。這使得靶DNA由兩種FnCpf1切口酶結合。此外，還設想的是可以使用不同的異種同源物，例如DNA的一條股（例如編碼股）上的FnCpf1、AsCpf1或LbCpf1切口酶和非編碼或相反DNA股上的異種同源物。異種同源物可以是但不限於Cas9切口酶諸如SaCas9切口酶或SpCas9切口酶。可能有利的是使用需要不同PAM並且還可以具有不同指導要求的兩種不同的異種同源物，由此允許使用者的更大程度的控制。在某些實施方式中，DNA切割涉及至少四種類型的切口酶，其中每種類型被指導到不同的靶DNA序列，其中每對在一個DNA股中引入第一切口並且第二對在第二條DNA股中引入切口。在此類方法中，至少兩對單股斷裂被引入到靶DNA中，其中在引入第一對單股斷裂和第二對單股斷裂後，第一對單股斷裂與第二對單股斷裂之間的靶序列被切斷。在某些實施方式中，異種同源物中的一者或兩者是可控的，例如是可誘導的。

[0046] 在本發明的某些實施方式中，指導RNA或成熟crRNA包含同向重複序列和指導序列或間隔區序列、基本上由或由同向

重複序列和指導序列或間隔區序列組成。在某些實施方式中，指導RNA或成熟crRNA包含連接至指導序列或間隔區序列的同向重複序列、基本上由或由該同向重複序列組成。在某些實施方式中，指導RNA或成熟crRNA包含19個核苷酸的部分同向重複序列，接著是20-30個核苷酸的指導序列或間隔區序列，有利地約20個核苷酸、23-25個核苷酸或24個核苷酸。在某些實施方式中，效應蛋白係FnCpf1、AsCpf1或LbCpf1效應蛋白並且需要至少16個核苷酸的指導序列以實現可檢測的DNA切割並且需要最小17個核苷酸的指導序列以實現有效的體外DNA切割。在某些實施方式中，同向重複序列位於指導序列或間隔區序列的上游（即5'）。在一較佳的實施方式中，FnCpf1、AsCpf1或LbCpf1指導RNA的種子序列（即為識別和/或雜交於靶座位處的序列所必不可少的序列）大約在指導序列或間隔區序列的5'端上的前5個核苷酸之內。

[0047] 在本發明的較佳的實施方式中，成熟crRNA包括莖環或優化的莖環結構或優化的二級結構。在較佳的實施方式中，成熟crRNA在同向重複序列中包括莖環或優化的莖環結構，其中莖環或優化的莖環結構對切割活性是重要的。在某些實施方式中，成熟crRNA較佳的是包括單一莖環。在某些實施方式中，同向重複序列較佳的是包括單一莖環。在某些實施方式中，效應蛋白複合物的切割活性藉由引入影響莖環RNA雙股體結構的突變來修飾。在較佳的實施方式中，可以引入保持莖環的RNA雙股體的突變，由此效應蛋白複合物的切割活性被保持。在其他較佳的實施方式中，可以引入擾亂莖環的RNA雙股體結構的突變，由此效應蛋白複合物的切割活性被完全廢除。

[0048] 在在此所述方法或組成物中的任一種中，本發明還提供了編碼被密碼子優化為在真核細胞或原核細胞中表現的效應蛋白的核苷酸序列。在本發明的一個實施方式中，密碼子優化的效應蛋白係FnCpf1p、AsCpf1或LbCpf1並且是針對在真核細胞或生物體中的可操作性進行密碼子優化，該真核細胞或生物體為例如如在此任何地方提到的此細胞或生物體，例如但不限於酵母細胞或哺乳動物細胞或生物體，包括小鼠細胞、大鼠細胞和人類細胞或非人類真核生物體，例如植物。

[0049] 在本發明的某些實施方式中，至少一個核定位信號（NLS）附接到編碼Cpf1效應蛋白的核酸序列。在較佳的實施方式中，至少一個或多個C末端或N末端的NLS被附接（並且因此一種或多種核酸分子編碼Cpf1效應蛋白可以包括編碼一個或多個NLS，使得表現的產物具有附接或連接的一個或多個NLS）。在一較佳的實施方式中，C末端的NLS被附接以用於真核細胞較佳的是人類細胞中的最佳表現和核靶向。在一較佳的實施方式中，密碼子優化的效應蛋白係FnCpf1p、AsCpf1或LbCpf1並且指導RNA的間隔區長度為從15至35個核苷酸。在某些實施方式中，指導RNA的間隔區長度為至少16個核苷酸，諸如至少17個核苷酸。在某些實施方式中，間隔區長度為從15至17個核苷酸、從17至20個核苷酸、從20至24個核苷酸，例如20、21、22、23或24個核苷酸、從23至25個核苷酸，例如23、24或25個核苷酸、從24至27個核苷酸、從27-30個核苷酸、從30-35個核苷酸或35個核苷酸或更長。在本發明的某些實施方式中，密碼子優化的效應蛋白係FnCpf1p並且指導RNA的同向重複序列長度為至少16個核苷酸。在某些實施方式中，密碼

子優化的效應蛋白係FnCpf1p並且指導RNA的同向重複序列長度為從16至20個核苷酸，例如16、17、18、19或20個核苷酸。在某些較佳的實施方式中，指導RNA的同向重複長度為19個核苷酸。

[0050] 本發明還涵蓋用於遞送多個核酸組分的方法，其中每個核酸組分對不同的感興趣的靶座位具有特異性，從而修飾多個感興趣的靶座位。複合物的核酸組分可以包含一個或多個蛋白質結合的RNA適配體。一個或多個適配體可以能夠結合噬菌體外殼蛋白。噬菌體外殼蛋白可以選自下組，該組包括下項：Qβ、F2、GA、fr、JP501、MS2、M12、R17、BZ13、JP34、JP500、KU1、M11、MX1、TW18、VK、SP、FI、ID2、NL95、TW19、AP205、ϕCb5、ϕCb8r、ϕCb12r、ϕCb23r、7s和PRR1。在一較佳的實施方式中，噬菌體外殼蛋白係MS2。本發明還提供了複合物的核酸組分，該核酸組分的長度為30或更多個、40或更多個，或50或更多個核苷酸。

[0051] 本發明還涵蓋具有細胞、組分和/或系統中存在的痕量陽離子的本發明的細胞、組分和/或系統。有利地，陽離子係鎂，諸如Mg^{2+}。陽離子可以痕量存在。對於陽離子（有利地是Mg^{2+}），較佳的範圍可以是約1 mM至約15 mM。對於基於人類的細胞、組分和/或系統，較佳的濃度可以是約1 mM，並且對於基於細菌的細胞、組分和/或系統，較佳的濃度可以是約10 mM至約15 mM。參見，例如加西烏納斯（Gasiunas）等人，美國國家科學院院刊（PNAS），2012年9月4日線上公開，www.pnas.org/cgi/doi/10.1073/pnas.1208507109.

[0052] 因此，本發明的目的在於，在本發明內不涵蓋任何先前已知的產品、製備該產品的過程或使用該產品之方法，使得申請人保留和在此揭露放棄任何先前已知的產品、過程或方法的權利。進一步指出的是，在本發明的範圍之內，本發明並非旨在涵蓋任何產品、過程或該產品的製備或使用該產品的方法，其不符合USPTO（35 U.S.C. §112，第一段）或EPO（EPC第83條）的書面說明和實施要求，使得申請人保留和在此揭露放棄任何先前所述的產品、製備該產品的過程或使用該產品的方法之權利。可能有利的是在本發明實踐中遵照EPC第53條 (c) 和EPC規則28 (b) 和(c)。在此沒任何東西被解釋為約定的。

[0053] 指出的是，在本揭露中並且特別是在本申請專利範圍和/或段落中，術語諸如“包括（comprises）”、“包括（comprised）”、“包括（comprising）”等可以具有在美國專利法中屬於它的含義；例如，它們可以意指“包括（includes）”、“包括（included）”、"包括（including）”等；並且該等術語諸如“基本上由.....組成（consisting essentially of）”和“基本上由.....組成（consists essentially of）”具有在美國專利法中歸於它們的含義。

[0054] 該等和其他實施方式揭露於以下詳細說明中或根據其是清楚的並且由其涵蓋。

【圖式簡單說明】

[0055] 本發明的新穎特徵在所附申請專利範圍中具體闡述。藉由參考對說明性實施方式進行闡述的以下詳細說明，將獲得對本發明的特徵和優點的更好理解，在該等實施方式中利用了本發

明的原理，並且在該等附圖中：

[0056] 圖1A-1B描繪了CRISPR-Cas系統的新分類。第1類包括多亞基crRNA-效應物複合物（Cascade）並且第2類包括單亞基crRNA-效應物複合物（Cas9樣）。

[0057] 圖2提供了CRISPR-Cas的分子組構。

[0058] 圖3A-3D提供了I型和III型效應複合物的結構：共同的體系結構/共同的祖先，儘管存在廣泛的序列趨異性。

[0059] 圖4示出了作為以RNA識別模體（RRM）為中心的系統的CRISPR-Cas。

[0060] 圖5A-5D示出了Cas1系統發育，其中自我調整模組和crRNA-效應物模組的重組顯示出CRISPR-Cas進化的主要方面。

[0061] 圖6示出了CRISPR-Cas種群調查，具體地是CRISPR-Cas型/亞型在古生菌和細菌之中的分佈。

[0062] 圖7描繪了用於鑒定Cas候選物的流程。

[0063] 圖8A-8D描繪了第2類系統的完整座位的組構。

[0064] 圖9A-9B描繪了C2c1鄰近群體。

[0065] 圖10A-10C描繪了Cas1樹。

[0066] 圖11A-11B描繪了第2類家族的結構域組構。

[0067] 圖12A-12B描繪了第2類蛋白質（SEQ ID NO 246-428，分別按出現的順序）的TnpB同源區。

[0068] 圖13A-13B描繪了C2c2鄰近群體。

[0069] 圖14A-14E描繪了C2c2家族（SEQ ID NO 429-1032，分別按出現的順序）中的HEPN RxxxxH模體。

[0070] 圖15描繪了C2C1：1.酸土脂環酸芽孢桿菌ATCC 49025（SEQ ID NO 1034-1037，分別按出現的順序）。

[0071] 圖16描繪了C2C1：4.硫岐化脫硫鹽鹼桿菌（*Desulfonatronum thiodismutans*）菌株MLF-1（SEQ ID NO 1038-1041，分別按出現的順序）。

[0072] 圖17描繪了C2C1：5.豐祐菌科細菌TAV5（SEQ ID NO 1042-1045，分別按出現的順序）。

[0073] 圖18描繪了C2C1：7.嗜熱澱粉芽孢桿菌菌株B4166（SEQ ID NO 1046-1049，分別按出現的順序）。

[0074] 圖19描繪了C2C1：9.芽胞桿菌屬某種NSP2.1（SEQ ID NO 1050-1053，分別按出現的順序）。

[0075] 圖20描繪了C2C2：1.毛螺旋菌科細菌MA2020（SEQ ID NO 1054-1057，分別按出現的順序）。

[0076] 圖21描繪了C2C2：2.毛螺旋菌科細菌NK4A179（SEQ ID NO 1058-1064，分別按出現的順序）。

[0077] 圖22描繪了C2C2：3.嗜胺梭菌（[Clostridium] aminophilum）DSM 10710（SEQ ID NO 1065-1068，分別按出現的

順序）。

[0078] 圖23描繪了C2C2：4.毛螺旋菌科細菌NK4A144（SEQ ID NO 1069和1070，分別按出現的順序）。

[0079] 圖24描繪了C2C2：5.雞肉桿菌（Carnobacterium gallinarum）DSM 4847（SEQ ID NO 1071-1074，分別按出現的順序）。

[0080] 圖25描繪了C2C2：6.雞肉桿菌（Carnobacterium gallinarum）DSM 4847（SEQ ID NO 1075-1081，分別按出現的順序）。

[0081] 圖26描繪了C2C2：7.產丙酸帕魯迪菌（*Paludibacter propionicigenes*）WB4（SEQ ID NO: 1082）。

[0082] 圖27描繪了C2C2：8.血清型斯氏李斯特菌1/2b（SEQ ID NO 1083-1086，分別按出現的順序）。

[0083] 圖28描繪了C2C2：9.威氏李斯特菌FSL R9-0317（SEQ ID NO: 1087）。

[0084] 圖29描繪了C2C2：10.李斯特菌屬細菌FSL M6-0635（SEQ ID NO 1088和1091，分別按出現的順序）。

[0085] 圖30描繪了C2C2：11.韋德（wadei）纖毛菌F0279（SEQ ID NO: 1092）。

[0086] 圖31描繪了C2C2：12.韋德纖毛菌F0279（SEQ ID NO 1093-1099，分別按出現的順序）。

[0087] 圖32描繪了C2C2：14.沙氏（shahii）纖毛菌屬DSM 19757（SEQ ID NO 1100-1103，分別按出現的順序）。

[0088] 圖33描繪了C2C2：15.莢膜紅細菌SB 1003（SEQ ID NO 1104和1105，分別按出現的順序）。

[0089] 圖34描繪了C2C2：16.莢膜紅細菌R121（SEQ ID NO 1106和1107，分別按出現的順序）。

[0090] 圖35描繪了C2C2：17.莢膜紅細菌R121（SEQ ID NO 1108和1109，分別按出現的順序）。

[0091] 圖36描繪了DR樹。

[0092] 圖37描繪了C2C2樹。

[0093] 圖38A-38BB示出了Cas-Cpf1異種同源物（SEQ ID NO 1033和1110-1166，分別按出現的順序）的序列比對。

[0094] 圖39A-39B示出了Cpf1座位比對的綜述。

[0095] 圖40A-40X示出了PACYC184 FnCpf1（PY001）載體構建體（SEQ ID NO: 1167和SEQ ID NO 1168-1189，分別按出現的順序）。

[0096] 圖41A-41I示出了人源化PaCpf1的序列，其中核苷酸序列為SEQ ID NO: 1190並且蛋白質序列為SEQ ID NO: 1191。

[0097] 圖42描繪了PAM激發測定。

[0098] 圖43描繪了內源性FnCpf1座位的示意圖。pY0001係具有部分FnCpf1座位的pACY184骨架（來自NEB）。FnCpf1座位被

PCR擴增成三個片段並且使用吉普森（Gibson）元件將該等片段選殖到Xba1和Hind3切割的pACYC184中。PY0001含有從255 bp的乙醯轉移酶3'序列至第四間隔區序列的內源性FnCpf1座位。僅間隔區1-3係潛在地具有活性的，因為間隔區4不再側接同向重複序列。

[0099] 圖44描繪了PAM文庫（library），其按出現的順序分別揭露了SEQ ID NO 1192-1195。兩個PAM文庫（左和右）均處於pUC19中。左PAM文庫的複雜度為48 ~ 65 k並且右PAM文庫的複雜度為47 ~ 16 k。兩個文庫被製備有>500的表現度。

[0100] 圖45A-45E描繪了FnCpf1 PAM篩選計算分析。在對篩選DNA定序之後，提取出對應於左PAM或右PAM的區。對於每個樣品，將定序文庫中存在的PAM數目與文庫中預期的PAM數目（對於左文庫為4^8，對於右文庫為4^7）進行比較。圖44A描繪了左文庫示出PAM缺失。為了量化此缺失，計算了富集比。針對兩種條件（對照pACYC或含有FnCpf1的pACYC），根據$比率 = -\log_2 \frac{樣品 + 0.01}{初始文庫 + 0.01}$針對文庫中的每種PAM計算該比率。繪製的分佈顯示在對照樣品中幾乎沒有富集並且在兩種生物複製本（bioreps）中有富集。圖44B-44D描繪了PAM比率的分佈。圖44E顯示，收集比率為8之上的PAM，並且繪製頻率分佈，揭示出5' YYN PAM。

[0101] 圖46描繪了土拉熱弗朗西絲菌Cpf1座位的RNA定序（RNAseq）分析，該分析顯示CRISPR座位被啟動表現。除Cpf1和Cas基因之外，兩種小的非編碼轉錄物被高度轉錄，這兩種非編碼轉錄物可能是推定的tracrRNA。CRISPR陣列也被表現。兩種推

定的tracrRNA和CRISPR陣列以與Cpf1和Cas基因相同的方向進行轉錄。在此藉由RNA定序實驗鑒定的所有RNA轉錄物映射到座位。在對FnCpf1座位進一步評價之後，申請人推斷藉由Cpf1效應蛋白複合物進行靶DNA切割不需要tracrRNA。申請人確定僅包含Cpf1效應蛋白和crRNA(包含同向重複序列和指導序列的指導RNA)的Cpf1效應蛋白複合物足以切割靶DNA。

[0102] 圖47描繪了放大的Cpf1 CRISPR陣列。可以鑒定出許多不同的短轉錄物。在此繪圖中，將所有鑒定的RNA轉錄物映射到Cpf1座位。

[0103] 圖48描繪了在選擇小於85個核苷酸長的轉錄物之後鑒定出的兩種推定的tracrRNA。

[0104] 圖49描繪了放大的推定的tracrRNA 1（SEQ ID NO: 1196）和CRISPR陣列。

[0105] 圖50描繪了放大的推定的tracrRNA 2，其按出現的順序分別揭露了SEQ ID NO 1197-1203。

[0106] 圖51描繪了推定的crRNA序列（重複序列為藍色，間隔區為黑色）（SEQ ID NO 1205和1206，分別按出現的順序）。

[0107] 圖52示出了用於在體內證實預測的FnCpf1 PAM的測定的示意圖。

[0108] 圖53示出了用編碼具有5' TTN PAM的內源性間隔區1的pUC19轉化的攜帶FnCpf1座位的細胞和對照細胞。

[0109] 圖54示出了指明FnCpf1座位中的推定的tracrRNA序

列位置、crRNA（SEQ ID NO: 1207）以及pUC原型間隔區載體的示意圖。

[0110] 圖55係示出了在細胞裂解物中培養的具有TTa PAM的PCR片段和原型間隔區1序列的凝膠。

[0111] 圖56係示出了在細胞裂解物中培養的具有不同的PAM的pUC-間隔區1的凝膠。

[0112] 圖57係示出了在細胞裂解物中培養之後的BasI消化的凝膠。

[0113] 圖58係示出了三種推定的crRNA序列（SEQ ID NO: 1208）的消化結果的凝膠。

[0114] 圖59係示出了對針對含有靶位點：5'-TTAgagaagtcatttaataaggccactgttaaaa-3'的靶DNA片段（SEQ ID NO: 1209）的不同長度的間隔區的測試的凝膠。結果顯示在體外crRNA 1-7介導了使用FnCpf1對靶DNA的成功切割。crRNA 8-13不有利於靶DNA的切割。分別按出現的順序揭露了SEQ ID NO 1210-1248。

[0115] 圖60係指明了最小FnCpf1座位的示意圖。

[0116] 圖61係指明了最小Cpf1指導序列（SEQ ID NO: 1249）的示意圖。

[0117] 圖62A-62E描繪了PaCpf1 PAM篩選計算分析。在對篩選DNA定序之後，提取出對應於左PAM或右PAM的區。對於每個樣品，將定序文庫中存在的PAM數目與文庫中預期的PAM數目（4^7）進行比較。（圖62A）左文庫顯示出非常輕微的PAM缺失。

為了量化此缺失，計算了富集比。針對兩種條件（對照pACYC或含有PaCpf1的pACYC），根據以下公式針對文庫中的每種PAM計算該比率：

$$\text{比率} = -\log_2 \frac{\text{樣品} + 0.01}{\text{初始文庫} + 0.01}$$

繪製的分佈顯示在對照樣品中幾乎沒有富集並且在兩種生物複製本中有富集。圖62B-62D描繪了PAM比率的分佈。圖62E顯示，收集比率為4.5之上的所有PAM，並且繪製頻率分佈，揭示出5' TTTV PAM，其中V係A或C或G。

[0118] 圖63示出了被描繪為CBh-NLS-huPaCpf1-NLS-3xHA-pA的人類密碼子優化的PaCpf1序列的載體圖譜。

[0119] 圖64A-64B示出了不同細菌中的51 Cpf1座位的系統發育樹。突出顯示框指示基因參考號：1-17。使用預測的成熟crRNA測試加框/編號的異種同源物在體外的切割活性；在其編號周圍具有框的異種同源物在體外測定中顯示出活性。

[0120] 圖65A-65H顯示出具有3849個核苷酸的基因長度的毛螺旋菌科細菌MC2017 1 Cpf1（圖64中的參考號3）的人類密碼子優化序列的細節。圖65A：密碼子適應指數（CAI）。密碼子使用頻率沿基因序列長度的分佈。就高基因表現水平而言，在所希望的表現生物體中1.0 CAI被認為是完美的，並且>0.8的CAI被認為是良好的。圖65B：最佳的密碼子的頻率（FOP）。在計算的密碼子品質組中的密碼子分佈百分比。該值對於在所希望的表現生物體中針對給定胺基酸具有最大使用頻率的密碼子被設定為100。圖

65C：GC含量調整。GC含量的理想百分比範圍為介於30%-70%之間。60 bp視窗中的%GC含量的峰已被移除。圖65D：限制性內切酶和順式作用元件。圖65E：移除重複序列。圖65F-G：優化序列（優化序列長度：3849，GC% 54.70)（SEQ ID NO: 1250）。圖65H：蛋白質序列（SEQ ID NO: 1251）。

[0121] 圖66A-66H顯示出具有3873個核苷酸的基因長度的解朊丁酸弧菌Cpf1（圖64中的參考號4）的人類密碼子優化序列的細節。圖66A：密碼子適應指數（CAI）。密碼子使用頻率沿基因序列長度的分佈。就高基因表現水平而言，在所希望的表現生物體中1.0 CAI被認為是完美的，並且>0.8的CAI被認為是良好的。圖66B：最佳的密碼子的頻率（FOP）。在計算的密碼子品質組中的密碼子分佈百分比。該值對於在所希望的表現生物體中針對給定胺基酸具有最大使用頻率的密碼子被設定為100。圖66C：GC含量調整。GC含量的理想百分比範圍為介於30%-70%之間。60 bp視窗中的%GC含量的峰已被移除。圖66D：限制性內切酶和順式作用元件。圖66E：移除重複序列。圖66F-G：優化序列（優化序列長度：3873，GC% 54.05)（SEQ ID NO: 1252）。圖66H：蛋白質序列（SEQ ID NO: 1253）。

[0122] 圖67A-67H顯示出具有4581個核苷酸的基因長度的佩萊格裡尼菌科細菌GW2011_GWA2_33_10 Cpf1（圖64中的參考號5）的人類密碼子優化序列的細節。圖67A：密碼子適應指數

（CAI）。密碼子使用頻率沿基因序列長度的分佈。就高基因表現水平而言，在所希望的表現生物體中1.0 CAI被認為是完美的，並且>0.8的CAI被認為是良好的。圖67B：最佳的密碼子的頻率（FOP）。在計算的密碼子品質組中的密碼子分佈百分比。該值對於在所希望的表現生物體中針對給定胺基酸具有最大使用頻率的密碼子被設定為100。圖67C：GC含量調整。GC含量的理想百分比範圍為介於30%-70%之間。60 bp視窗中的%GC含量的峰已被移除。圖67D：限制性內切酶和順式作用元件。圖67E：移除重複序列。圖67F-G：優化序列（優化序列長度：4581，GC% 50.81)（SEQ ID NO: 1254）。圖67H：蛋白質序列（SEQ ID NO: 1255）。

[0123] 圖68A-68H顯示出具有4206個核苷酸的基因長度的儉菌總門細菌GW2011_GWC2_44_17 Cpf1（圖64中的參考號6）的人類密碼子優化序列的細節。圖68A：密碼子適應指數（CAI）。密碼子使用頻率沿基因序列長度的分佈。就高基因表現水平而言，在所希望的表現生物體中1.0 CAI被認為是完美的，並且>0.8的CAI被認為是良好的。圖68B：最佳的密碼子的頻率（FOP）。在計算的密碼子品質組中的密碼子分佈百分比。該值對於在所希望的表現生物體中針對給定胺基酸具有最大使用頻率的密碼子被設定為100。圖68C：GC含量調整。GC含量的理想百分比範圍為介於30%-70%之間。60 bp視窗中的%GC含量的峰已被移除。圖68D：限制性內切酶和順式作用元件。圖68E：移除重複序列。圖68F-G：優化序列（優化序列長度：4206，GC% 52.17)（SEQ ID NO: 1256）。圖68H：蛋白質序列（SEQ ID NO: 1257）。

[0124] 圖69A-69H顯示出具有3900個核苷酸的基因長度的密斯氏菌屬某種SCADC Cpf1（圖64中的參考號7）的人類密碼子優化序列的細節。圖69A：密碼子適應指數（CAI）。密碼子使用頻率沿基因序列長度的分佈。就高基因表現水平而言，在所希望的表現生物體中1.0 CAI被認為是完美的，並且>0.8的CAI被認為是良好的。圖69B：最佳的密碼子的頻率（FOP）。在計算的密碼子品質組中的密碼子分佈百分比。該值對於在所希望的表現生物體中針對給定胺基酸具有最大使用頻率的密碼子被設定為100。圖69C：GC含量調整。GC含量的理想百分比範圍為介於30%-70%之間。60 bp視窗中的%GC含量的峰已被移除。圖69D：限制性內切酶和順式作用元件。圖69E：移除重複序列。圖69F-G：優化序列（優化序列長度：3900，GC% 51.56)（SEQ ID NO: 1258）。圖69H：蛋白質序列（SEQ ID NO: 1259）。

[0125] 圖70A-70H顯示出具有4071個核苷酸的基因長度的胺基酸球菌屬某種BV3L6 Cpf1（圖64中的參考號8）的人類密碼子優化序列的細節。圖70A：密碼子適應指數（CAI）。密碼子使用頻率沿基因序列長度的分佈。就高基因表現水平而言，在所希望的表現生物體中1.0 CAI被認為是完美的，並且>0.8的CAI被認為是良好的。圖70B：最佳的密碼子的頻率（FOP）。在計算的密碼子品質組中的密碼子分佈百分比。該值對於在所希望的表現生物體中針對給定胺基酸具有最大使用頻率的密碼子被設定為100。圖70C：GC含量調整。GC含量的理想百分比範圍為介於30%-70%之間。60 bp視窗中的%GC含量的峰已被移除。圖70D：限制性內切酶和順式作用元件。圖70E：移除重複序列。圖70F-G：優化序列（優化序

列長度：4071，GC% 54.89)（SEQ ID NO: 1260）。圖70H：蛋白質序列（SEQ ID NO: 1261）。

[0126] 圖71A-71H顯示出具有3768個核苷酸的基因長度的毛螺旋菌科細菌MA2020 Cpf1（圖64中的參考號9）的人類密碼子優化序列的細節。圖71A：密碼子適應指數（CAI）。密碼子使用頻率沿基因序列長度的分佈。就高基因表現水平而言，在所希望的表現生物體中1.0 CAI被認為是完美的，並且>0.8的CAI被認為是良好的。圖71B：最佳的密碼子的頻率（FOP）。在計算的密碼子品質組中的密碼子分佈百分比。該值對於在所希望的表現生物體中針對給定胺基酸具有最大使用頻率的密碼子被設定為100。圖71C：GC含量調整。GC含量的理想百分比範圍為介於30%-70%之間。60 bp視窗中的%GC含量的峰已被移除。圖71D：限制性內切酶和順式作用元件。圖71E：移除重複序列。圖71F-G：優化序列（優化序列長度：3768，GC% 51.53)（SEQ ID NO: 1262）。圖71H：蛋白質序列（SEQ ID NO: 1263）。

[0127] 圖72A-72H顯示出具有3864個核苷酸的基因長度的候選白蟻甲烷枝原體Cpf1（圖64中的參考號10）的人類密碼子優化序列的細節。圖72A：密碼子適應指數（CAI）。密碼子使用頻率沿基因序列長度的分佈。就高基因表現水平而言，在所希望的表現生物體中1.0 CAI被認為是完美的，並且>0.8的CAI被認為是良好的。圖72B：最佳的密碼子的頻率（FOP）。在計算的密碼子品質組中的密碼子分佈百分比。該值對於在所希望的表現生物體中針對給定胺基酸具有最大使用頻率的密碼子被設定為100。圖72C：

GC含量調整。GC含量的理想百分比範圍為介於30%-70%之間。60 bp視窗中的%GC含量的峰已被移除。圖72D：限制性內切酶和順式作用元件。圖72E：移除重複序列。圖72F-G：優化序列（優化序列長度：3864，GC% 52.67)（SEQ ID NO: 1264）。圖72H：蛋白質序列（SEQ ID NO: 1265）。

[0128] 圖73A-73H顯示出具有3996個核苷酸的基因長度的挑剔真細菌Cpf1（圖64中的參考號11）的人類密碼子優化序列的細節。圖73A：密碼子適應指數（CAI）。密碼子使用頻率沿基因序列長度的分佈。就高基因表現水平而言，在所希望的表現生物體中1.0 CAI被認為是完美的，並且>0.8的CAI被認為是良好的。圖73B：最佳的密碼子的頻率（FOP）。在計算的密碼子品質組中的密碼子分佈百分比。該值對於在所希望的表現生物體中針對給定胺基酸具有最大使用頻率的密碼子被設定為100。圖73C：GC含量調整。GC含量的理想百分比範圍為介於30%-70%之間。60 bp視窗中的%GC含量的峰已被移除。圖73D：限制性內切酶和順式作用元件。圖73E：移除重複序列。圖73F-G：優化序列（優化序列長度：3996，GC% 50.52)（SEQ ID NO: 1266）。圖73H：蛋白質序列（SEQ ID NO: 1267）。

[0129] 圖74A-74H顯示出具有4269個核苷酸的基因長度的牛莫拉氏菌237 Cpf1（圖64中的參考號12）的人類密碼子優化序列的細節。圖74A：密碼子適應指數（CAI）。密碼子使用頻率沿基因序列長度的分佈。就高基因表現水平而言，在所希望的表現生物體中1.0 CAI被認為是完美的，並且>0.8的CAI被認為是良好的。圖

74B：最佳的密碼子的頻率（FOP）。在計算的密碼子品質組中的密碼子分佈百分比。該值對於在所希望的表現生物體中針對給定胺基酸具有最大使用頻率的密碼子被設定為100。圖74C：GC含量調整。GC含量的理想百分比範圍為介於30%-70%之間。60 bp視窗中的%GC含量的峰已被移除。圖74D：限制性內切酶和順式作用元件。圖74E：移除重複序列。圖74F-G：優化序列（優化序列長度：4269，GC% 53.58)（SEQ ID NO: 1268）。圖74H：蛋白質序列（SEQ ID NO: 1269）。

[0130] 圖75A-75H顯示出具有3939個核苷酸的基因長度的稻田鉤端螺旋體Cpf1（圖64中的參考號13）的人類密碼子優化序列的細節。圖75A：密碼子適應指數（CAI）。密碼子使用頻率沿基因序列長度的分佈。就高基因表現水平而言，在所希望的表現生物體中1.0 CAI被認為是完美的，並且>0.8的CAI被認為是良好的。圖75B：最佳的密碼子的頻率（FOP）。在計算的密碼子品質組中的密碼子分佈百分比。該值對於在所希望的表現生物體中針對給定胺基酸具有最大使用頻率的密碼子被設定為100。圖75C：GC含量調整。GC含量的理想百分比範圍為介於30%-70%之間。60 bp視窗中的%GC含量的峰已被移除。圖75D：限制性內切酶和順式作用元件。圖75E：移除重複序列。圖75F-G：優化序列（優化序列長度：3939，GC% 51.30)（SEQ ID NO: 1270）。圖75H：蛋白質序列（SEQ ID NO: 1271）。

[0131] 圖76A-76H顯示出具有3834個核苷酸的基因長度的毛螺旋菌科細菌ND2006 Cpf1（圖64中的參考號14）的人類密碼子

優化序列的細節。圖76A：密碼子適應指數（CAI）。密碼子使用頻率沿基因序列長度的分佈。就高基因表現水平而言，在所希望的表現生物體中1.0 CAI被認為是完美的，並且>0.8的CAI被認為是良好的。圖76B：最佳的密碼子的頻率（FOP）。在計算的密碼子品質組中的密碼子分佈百分比。該值對於在所希望的表現生物體中針對給定胺基酸具有最大使用頻率的密碼子被設定為100。圖76C：GC含量調整。GC含量的理想百分比範圍為介於30%-70%之間。60 bp視窗中的%GC含量的峰已被移除。圖76D：限制性內切酶和順式作用元件。圖76E：移除重複序列。圖76F-G：優化序列（優化序列長度：3834，GC% 51.06)（SEQ ID NO: 1272）。圖76H：蛋白質序列（SEQ ID NO: 1273）。

[0132] 圖77A-77H顯示出具有3930個核苷酸的基因長度的狗口腔卟啉單胞菌3 Cpf1（圖64中的參考號15）的人類密碼子優化序列的細節。圖77A：密碼子適應指數（CAI）。密碼子使用頻率沿基因序列長度的分佈。就高基因表現水平而言，在所希望的表現生物體中1.0 CAI被認為是完美的，並且>0.8的CAI被認為是良好的。圖77B：最佳的密碼子的頻率（FOP）。在計算的密碼子品質組中的密碼子分佈百分比。該值對於在所希望的表現生物體中針對給定胺基酸具有最大使用頻率的密碼子被設定為100。圖77C：GC含量調整。GC含量的理想百分比範圍為介於30%-70%之間。60 bp視窗中的%GC含量的峰已被移除。圖77D：限制性內切酶和順式作用元件。圖77E：移除重複序列。圖77F-G：優化序列（優化序列長度：3930，GC% 54.42)（SEQ ID NO: 1274）。圖77H：蛋白質序列（SEQ ID NO: 1275）。

[0133] 圖78A-78H顯示出具有4119個核苷酸的基因長度的解糖胨普雷沃菌Cpf1（圖64中的參考號16）的人類密碼子優化序列的細節。圖78A：密碼子適應指數（CAI）。密碼子使用頻率沿基因序列長度的分佈。就高基因表現水平而言，在所希望的表現生物體中1.0 CAI被認為是完美的，並且>0.8的CAI被認為是良好的。圖78B：最佳的密碼子的頻率（FOP）。在計算的密碼子品質組中的密碼子分佈百分比。該值對於在所希望的表現生物體中針對給定胺基酸具有最大使用頻率的密碼子被設定為100。圖78C：GC含量調整。GC含量的理想百分比範圍為介於30%-70%之間。60 bp視窗中的%GC含量的峰已被移除。圖78D：限制性內切酶和順式作用元件。圖78E：移除重複序列。圖78F-G：優化序列（優化序列長度：4119，GC% 51.88)（SEQ ID NO: 1276）。圖78H：蛋白質序列（SEQ ID NO: 1277）。

[0134] 圖79A-79H顯示出具有3888個核苷酸的基因長度的獼猴卟啉單胞菌Cpf1（圖64中的參考號17）的人類密碼子優化序列的細節。圖79A：密碼子適應指數（CAI）。密碼子使用頻率沿基因序列長度的分佈。就高基因表現水平而言，在所希望的表現生物體中1.0 CAI被認為是完美的，並且>0.8的CAI被認為是良好的。圖79B：最佳的密碼子的頻率（FOP）。在計算的密碼子品質組中的密碼子分佈百分比。該值對於在所希望的表現生物體中針對給定胺基酸具有最大使用頻率的密碼子被設定為100。圖79C：GC含量調整。GC含量的理想百分比範圍為介於30%-70%之間。60 bp視窗中的%GC含量的峰已被移除。圖79D：限制性內切酶和順式作用元件。圖79E：移除重複序列。圖79F-G：優化序列（優化序列長度：

3888，GC% 53.26)（SEQ ID NO: 1278）。圖79H：蛋白質序列（SEQ ID NO: 1279）。

[0135] 圖80A-80I示出了每個異種同源物（是指圖64中的編碼參考號3-17）的同向重複（DR）序列以及它們的預測折疊結構。分別按出現的順序揭露了SEQ ID NO 1280-1313。

[0136] 圖81示出了人類Emx1座位的PCR擴增子的切割。分別按出現的順序揭露了SEQ ID NO 1314-1318。

[0137] 圖82A-82B示出了5' DR的截短對切割活性的影響。圖82A示出了其中指明了使用5 DR截短物的切割結果的凝膠。圖82B示出了其中crDNA ΔDR5擾亂5'端的莖環的圖。這指明5'端處的莖環係為切割活性所必需的。分別按出現的順序揭露了SEQ ID NO 1319-1324。

[0138] 圖83示出了crRNA-DNA靶錯配對切割效率的影響。分別按出現的順序揭露了SEQ ID NO 1325-1335。

[0139] 圖84示出了使用純化的弗朗西絲菌屬Cpf1和普雷沃菌屬Cpf1的DNA切割。揭露了SEQ ID NO: 1336。

[0140] 圖85A-85B示出了DR二級結構的圖。圖85A示出了FnCpf1 DR二級結構（SEQ ID NO: 1337）（莖環突出顯示）。圖85B示出了PaCpf1 DR二級結構（SEQ ID NO: 1338）（莖環突出顯示，除了環區中的單個鹼基不同之外都相同）。

[0141] 圖86示出了FnCp1座位的RNA定序分析的另外描繪。

[0142] 圖87A-87B示出了成熟crRNA序列的示意圖。圖87A

示出了FnCpf1的成熟crRNA序列。圖87B示出了PaCpf1的成熟crRNA序列。分別按出現的順序揭露了SEQ ID NO 1339-1342。

[0143] 圖88示出了使用人類密碼子優化的新殺手弗朗西絲菌FnCpf1的DNA切割。上條帶對應於未切割的全長片段(606 bp)。預期的 ~345 bp和 ~261 bp大小的切割產物由三角形指示。

[0144] 圖89示出了體外異種同源物測定，展示了Cpf1異種同源物進行的切割。

[0145] 圖90A-90C示出了來自體外切割測定的計算得出的PAM。

[0146] 圖91示出了為交錯方式的Cpf1切割，產生了5'突出端。分別按出現的順序揭露了SEQ ID NO 1343-1345。

[0147] 圖92示出了間隔區長度對切割的影響。分別按出現的順序揭露了SEQ ID NO 1346-1352。

[0148] 圖93示出了HEK293T細胞中FnCpf1介導的indel的SURVEYOR數據。

[0149] 圖94A-94F示出了與野生型FnCpf1座位的轉錄物的加工相比的在缺失FnCpf1座位部分時的轉錄物的加工。圖95B、95D和95F對加工的間隔區進行放大。分別按出現的順序揭露了SEQ ID NO 1353-1401。

[0150] 圖95A-95E顯示土拉熱弗朗西絲菌新殺手亞種*U112* Cpf1 CRISPR座位提供了針對含有側接5'-TTN PAM的原型間隔區的質粒的轉化的免疫。圖95A示出了土拉熱弗朗西絲菌新殺手亞

種*U112*(NC_008601)中發現的兩種CRISPR座位的組構。對FnCas9和FnCpf1的結構域組構進行了比較。圖95B提供了用於發現PAM位置和同一性的質粒缺失測定的示意性說明。用含有側接隨機化5'或3' PAM序列的匹配原型間隔區的質粒的文庫轉化具有異源FnCpf1座位質粒(pFnCpf1)或空載體對照的感受態大腸桿菌並且使用抗生素進行選擇以缺失攜帶成功靶向的PAM的質粒。提取出來自存活群落的質粒並且對該等質粒進行定序以確定缺失的PAM序列。圖95C-95D示出了如藉由質粒缺失測定確定的FnCpf1 PAM的序列圖示。位置處的字母高度由信息量來確定;誤差條顯示出的95%貝葉斯(Bayesian)置信區間。圖95E顯示具有pFnCpf1的大腸桿菌對攜帶5'-TTN PAM的質粒展示出穩健的干擾(n=3,誤差條表示平均值 ± 平均數標準誤差(S.E.M.))。

[0151] 圖96A-96C顯示大腸桿菌中FnCpf1和CRISPR陣列的異源表現足以介導質粒DNA干擾和crRNA成熟。土拉熱弗朗西絲菌新殺手亞種*U112*的小RNA定序(圖96A)揭示出FnCpf1 CRISPR陣列的轉錄和加工。成熟crRNA開始於19個核苷酸的部分同向重複序列,接著是23–25個核苷酸的間隔區序列。用攜帶合成啟動子驅動的FnCpf1和CRISPR陣列的質粒轉化的大腸桿菌的小RNA定序(圖96B)顯示crRNA加工與Cas基因以及FnCpf1座位中其他序列元件無關。圖96C描繪了具有FnCpf1 CRISPR座位的不同截短物的大腸桿菌並且顯示僅FnCpf1和CRISPR陣列係為質粒DNA干擾所需要的(n=3,誤差條顯示平均值 ± 平均數標準誤差)。揭露了SEQ ID NO: 1580。

[0152] 圖97A-97E顯示FnCpf1藉由crRNA來進行靶向以切割體外DNA。圖97A係FnCpf1 crRNA-DNA靶向複合物的示意圖。切割位點由紅色箭頭指示（SEQ ID NO 1402和1403，分別按出現順序揭露）。FnCpf1和crRNA以crRNA和Mg^{2+}依賴性方式單獨介導RNA指導的靶DNA切割（圖97B）。圖97C顯示FnCpf1切割直鏈DNA和超螺旋DNA兩者。圖97D顯示來自FnCpf1消化的靶標的桑格（Sanger）定序痕跡顯示出交錯的突出端（SEQ ID NO 1404和1406，分別按出現的順序揭露）。非模板的另外腺嘌呤（指代為N）的添加係定序中使用的聚合酶的偽影（artifact）。反向引物讀取表示為反向互補序列以有助於視覺化。圖97E顯示切割依賴於5' PAM處的鹼基配對。FnCpf1僅可識別正確地沃森-克裡克配對（Watson-Crick paired）的DNA中的PAM。

[0153] 圖98A-98B顯示FnCpf1的C末端RuvC結構域中的催化殘基係為DNA切割所需的。圖98A示出了FnCpf1結構域結構，其中RuvC催化殘基被突出顯示。基於與嗜熱棲熱菌（*Thermus thermophilus*）RuvC（PDB ID: 4EP5）的序列同源性來鑒定催化殘基。圖98B描繪了天然型TBE PAGE凝膠，其顯示FnCpf1的RuvC催化殘基突變（D917A和E1006A）和SpCas9的RuvC催化殘基突變（D10A）阻止雙股DNA的切割。使TBE-尿素PAGE凝膠變性顯示FnCpf1的RuvC催化殘基突變（D917A和E1006A）阻止DNA切口產生活性，而SpCas9的RuvC催化殘基突變（D10A）使得靶位點產生切口。

[0154] 圖99A-99E示出了體外FnCpf1核酸酶活性的crRNA要

求。圖99A示出了間隔區長度對FnCpf1切割活性的影響。圖99B示出了crRNA-靶DNA錯配對FnCpf1切割活性的影響。圖99C展示了同向重複序列長度對FnCpf1切割活性的影響。圖99D顯示FnCpf1切割活性取決於同向重複序列RNA結構的莖中的二級結構。圖99E顯示FnCpf1切割活性不受環突變影響，而是對同向重複序列的3'-大部分鹼基中的突變具有敏感性。分別按出現的順序揭露了SEQ ID NO 1407-1433。

[0155] 圖100A-100F提供了Cpf1家族蛋白多樣性和功能的分析。圖100A-100B示出了針對功能分析所選擇的16種Cpf1異種同源物的系統發育比較。保守序列以深灰色示出。突出顯示了RuvC結構域、橋螺旋（bridge helix）和鋅指。圖100C示出了來自16種Cpf1家族蛋白的同向重複序列的比對。在crRNA成熟後移除的序列為灰色。非保守序列為紅色。莖雙股體被突出顯示為灰色。圖100D描繪了成熟crRNA中的同向重複序列的RNAfold（勞倫茲（Lorenz）等人，2011）預測。示出了對FnCpf1以及三個保守性較低的異種同源物的預測。圖100E顯示具有類似的同向重複序列的異種同源物crRNA能夠與FnCpf1一起用於介導靶DNA切割。圖100F示出了使用含有隨機化PAM側接的原型間隔區的質粒文庫的體外切割所鑒定的8種Cpf1家族蛋白的PAM序列。分別按出現的順序揭露了SEQ ID NO 1434-1453。

[0156] 圖101A-101E顯示在人類細胞系中Cpf1介導穩健的基因組編輯。圖101A係示出了各個Cpf1家族蛋白在使用CMV驅動的表現載體的HEK 293FT細胞中的表現的示意圖（schemative）。

使用含有融合至crRNA序列的U6啟動子的PCR片段，相應的crRNA被表現。使用Surveyor核酸酶測定或靶向深度定序來分析轉染細胞。圖101B（上面）描繪了DNMT1-靶向crRNA 3的序列，並且定序讀取（下面）示出了代表性的indel。圖101B分別按出現的順序揭露了SEQ ID NO 1454-1465。圖101C提供了體外切割活性和體內切割活性的比較。DNMT1靶區進行PCR擴增並且使用基因組片段來測試Cpf1介導的切割。所有的8種Cpf1家族蛋白質示出了體外DNA切割（上面）。候選物7–AsCpf1和13–Lb3Cpf1促進人類細胞中的穩健的indel形成（底部）。圖101D示出了人類DNMT1座位中的Cpf1和SpCas9靶序列（SEQ ID NO 1466-1473，分別按出現的順序揭露）。圖101E提供了Cpf1和SpCas9基因組編輯效率的比較。靶位點對應於圖101D中所示的序列。

[0157] 圖102A-102D示出了用於鑒定FnCpf1 PAM的體外質粒缺失測定。（還參見圖95）。圖102A：用攜帶隨機化5' PAM序列的質粒的文庫對具有pFnCpf1的大腸桿菌的轉化。質粒的亞群被缺失。繪圖按排序後的順序示出了缺失水平。缺失被測定為相比於pACYC184大腸桿菌對照的標準化豐度倍數比的負$\log_2$。使用閾值為3.5以上的PAM來產生序列圖示。圖102B：用攜帶隨機化3' PAM序列的質粒的文庫對具有pFnCpf1的大腸桿菌的轉化。質粒的亞群被缺失。繪圖按排序後的順序示出了缺失水平。缺失被測定為相比於pACYC184大腸桿菌對照的標準化豐度倍數比的負$\log_2$並且使用閾值為3.5以上的PAM來產生序列圖示。圖102C：攜帶隨機化5' PAM序列的質粒的輸入文庫。繪圖按排序後的順序示出了缺失水平。缺失被測定為相比於pACYC184大腸桿菌對照的標準化豐

度倍數比的負$\log_2$。使用閾值為3.5以上的PAM來產生序列圖示。圖102D：超過5' PAM的2和3位置處的配對鹼基組合的顯著性閾值的獨特PAM的數目。

[0158] 圖103A-103B示出了FnCpf1蛋白純化。(還參見圖97)。圖103A描繪了FnCpf1的考馬斯藍(Coomassie blue)染色的丙烯醯胺凝膠，其示出了分散式純化。從Ni-NTA柱中洗提出恰好高於160 kD的條帶，該帶與MBP-FnCpf1融合物（189.7 kD）的大小一致。在添加TEV蛋白酶之後，出現較低分子量的條帶，該帶與不含FnCpf1的147 kD大小一致。圖103B：fnCpf1的大小排阻凝膠過濾。FnCpf1在大約300 kD大小（62.65 mL）下洗提出，表明Cpf1可能以二聚體存在於溶液中。圖103C示出了用於校準Superdex 200柱的蛋白質標準物。BDex = 藍葡聚糖（空隙體積），Ald = 醛縮酶（158 kD），Ov = 卵白蛋白（44 kD），RibA = 核糖核酸酶（13.7 kD），Apr = 抑肽酶（6.5 kD）。圖103D：Superdex 200柱的校準曲線。K_a被計算為（洗提體積 – 空隙體積）/（幾何柱體積 – 空隙體積）。對標準物進行繪圖並且擬合成對數曲線。

[0159] 圖104A-104E示出了FnCpf1切割型式。(還參見圖97)。來自FnCpf1消化的DNA靶的桑格定序痕跡顯示出交錯的突出端。非模板的另外腺嘌呤(指代為N)的添加係定序中使用的聚合酶的偽影。針對在原型間隔區1（圖104A）、原型間隔區2（圖104B）、原型間隔區3（圖104C）和靶標DNMT1和EMX1（圖104D）情況下的不同TTN PAM示出了桑格痕跡。(–) 股序列係反向互補的以示出上股序列。切割位點由紅色三角形指示。較小的三角形指示推

定的替代性切割位點。圖104E示出了PAM-遠端的crRNA-靶DNA錯配對FnCpf1切割活性的影響。分別按出現的順序揭露了SEQ ID NO 1474-1494。

[0160] 圖105A-105B示出了FnCpf1（SEQ ID NO: 1495）、AsCpf1（SEQ ID NO: 1496）和LbCpf1（SEQ ID NO: 1497）的胺基酸序列比對。（還參見圖100）。保守的殘基用紅色背景突出顯示並且保守的突變用輪廓和紅色字體突出顯示。比對上方（FnCpf1）和下方（LbCpf1）的二級結構預測被突出顯示。α螺旋示出為波紋符號並且β股示出為短劃線。圖95A中鑒定的蛋白質結構域也被突出顯示。

[0161] 圖106A-106D提供了對應於針對哺乳動物實驗所選擇的16種Cpf1家族蛋白的細菌基因組座位圖譜。（還參見圖100）。圖106A-106D分別按出現的順序揭露了SEQ ID NO 1498-1513。

[0162] 圖107A-107E示出了Cpf1家族蛋白的體外特性。圖107A係用於使用Cpf1家族蛋白的體外PAM篩選的示意圖。藉由各種Cpf1家族蛋白以及它們的相應crRNA切割具有隨機化5' PAM序列的質粒的文庫。純化未切割的質粒DNA並且對其進行定序以鑒定缺失的特異性PAM模體。圖107B指示了針對7–AsCpf1超過5' PAM的2和3位置處的配對鹼基組合的顯著性閾值的獨特序列的數目。圖107C指示了針對13–LbCpf1超過5' PAM的2、3和4位置處的三三鹼基組合的顯著性閾值的獨特PAM的數目。圖107D-107E D和E示出了來自7–AsCpf1消化的靶標（圖107D）和13–LbCpf1消化的靶標（圖107E）的桑格（Sanger）定序痕跡並且顯示出交錯的突出

端。非模板的另外腺嘌呤（指代為N）的添加係定序中使用的聚合酶的偽影。切割位點由紅色三角形指示。較小的三角形指示推定的替代性切割位點。圖107D-107E分別按出現的順序揭露了SEQ ID NO 1514-1519。

[0163] 圖108A-108F指示了附加座位處的人類細胞基因組編輯效率。Surveyor凝膠示出了對藉由每種Cpf1家族蛋白在DNMT1靶位點1（圖108A）、2（圖108B）和4（圖108C）處實現的indel效率的量化。圖108A-108C指示了附加座位處的人類細胞基因組編輯效率和DNMT靶位點切割的桑格定序。Surveyor凝膠示出了對藉由每種Cpf1家族蛋白在EMX1靶位點1（圖108D）和2（圖108E）處實現的indel效率的量化。AsCpf1和LbCpf1與DNMT1靶位點2、3和4的indel分佈（圖108F）。青色條表示總的indel覆蓋度；藍色條表示indel的3'端的分佈。對每個靶標，PAM序列用紅色表示並且靶序列用淡藍色表示。

[0164] 圖109A-109C描繪了Cpf1核酸酶初級結構的計算分析，揭示了三個不同的區。第一係C末端RuvC樣結構域，其係僅功能表征的結構域。第二係N末端α-螺旋區並且第三係位於RuvC樣結構域與α-螺旋區之間的混合的α區和β區。

[0165] 圖110A-110E描繪了AsCpf1 Rad50比對（PDB 4W9M）。分別按出現的順序揭露了SEQ ID NO 1520和1521。圖110C描繪了AsCpf1 RuvC比對（PDB 4LD0）。分別按出現的順序揭露了SEQ ID NO 1522和1523。圖110D-110E描繪了AsCpf1和FnCpf1的比對，該比對鑒定出FnCpf1中的Rad50結構域。分別按出現的順序揭露了

SEQ ID NO 1524和1525。

[0166] 圖111描繪了Rad50(4W9M)與DNA複合的結構。DNA相互作用殘基係突出顯示的（為紅色）。

[0167] 圖112描繪了RuvC（4LD0）與霍利迪連接體（holiday junction）複合的結構。DNA相互作用殘基以紅色突出顯示。

[0168] 圖113描繪了AsCpf1與位點特異性重組酶XerD的區比對的blast。XerD活性位點區係LYWTGMR（SEQ ID NO: 1），其中R係催化殘基。分別按出現的順序揭露了SEQ ID NO 1526-1527。

[0169] 圖114描繪了Cpf1異種同源物中保守的區（黃色框）並且雖然R不是保守的，但是高度保守的天冬胺酸（橙色框）恰好係該區的C末端，以及附近的具有絕對保守的精胺酸的保守區（藍色框）。天冬胺酸係LbCpf1中的D732。分別按出現的順序揭露了SEQ ID NO 1204和1528-1579。

[0170] 圖115A示出了實驗，其中在轉染前24 h，每24孔接種150,000個HEK293T細胞。使用Lipofectamin2000用400 ng huAsCpf1質粒和100 ng包含針對GRIN28的一個指導序列和以串聯方式置於U6啟動子後面的針對EMX1的一個指導序列的串聯（tandem）指導質粒轉染細胞。轉染後72 h收穫細胞並且使用SURVEYOR核酸酶測定來測定由串聯指導序列介導的AsCpf1活性。

[0171] 圖115B展示出GRIN28和EMX1基因兩者中的indel資訊。

[0172] 圖116示出了在EDTA濃度增加（和Mg2+濃度降低）的情況下陣列的FnCpf1切割。緩衝液係20 mM TrisHCl pH 7（室溫），50 mM KCl並且包括鼠類RNA酶抑制劑以防止RNA由於從蛋白質純化留下的潛在痕量的非特異性RNA酶而降解。

[0173] 在此的圖僅是出於說明目的並且不一定按比例繪製。

【實施方式】

[0174] 本申請描述了功能上不同於先前所述的CRISPR-Cas9系統的新型RNA指導的核酸內切酶（例如，Cpf1效應蛋白）並且因此在此與該等新型內切核酸酶相關聯的元件術語相應地被修改。在此所述的Cpf1相關CRISPR陣列被加工為成熟crRNA而不需要附加tracrRNA。在此所述的crRNA包含間隔區序列（或指導序列）和同向重複序列並且Cpf1p-crRNA複合物本身足以有效地切割靶DNA。在此所述的種子序列，例如FnCpf1指導RNA的種子序列大約在間隔區序列（或指導序列）的5'端上的前5個核苷酸之內並且種子序列內的突變不利地影響Cpf1效應蛋白複合物的切割活性。

[0175] 總的來說，CRISPR系統特徵在於在靶序列的位點處促進CRISPR複合物形成的元件（也稱之為內源性CRISPR系統情況下的原型間隔區）。在形成CRISPR複合物的情況下，“靶序列”係指指導序列被設計為所靶向的序列，例如與其具有互補性的序列，其中靶序列與指導序列之間的雜交促進CRISPR複合物的形成。指導序列藉由其而與靶序列互補對切割活性係重要的部分在此稱之為種子序列。靶序列可包括任何多核苷酸，諸如DNA或RNA多核

苷酸並且包含在感興趣的靶座位之內。在一些實施方式中，靶序列位於細胞的核或細胞質中。在此所述的本發明涵蓋第2類CRISPR-Cas系統的新型效應蛋白，其中Cas9係示例性效應蛋白並且因此本申請中用於描述新型效應蛋白的術語可以與用於描述CRISPR-Cas9系統的術語相關。

[0176] CRISPR-Cas座位具有超過50種的基因家族並且不存在嚴格的通用基因。因此，單一進化樹係不可行的並且需要多分支方法來鑒定新家族。到目前為止，針對93種Cas蛋白存在395種表現譜的全面*cas*基因鑒定。分類包括特徵基因表現譜加上座位體系結構的特徵。新的CRISPR-Cas系統分類提出在圖1中。第1類包括多亞基crRNA-效應物複合物（Cascade）並且第2類包括單亞基crRNA-效應物複合物（Cas9樣）。圖2提供了CRISPR-Cas的分子組構。圖3提供了I型和III型效應複合物的結構：共同的體系結構/共同的祖先，儘管存在廣泛的序列趨異性。圖4示出了作為以RNA識別模體（RRM）為中心的系統的CRISPR-Cas。圖5示出了Cas1系統發育，其中自我調整模組和crRNA-效應物模組的重組顯示出CRISPR-Cas進化的主要方面。圖6示出了CRISPR-Cas種群調查，具體地在古生菌和細菌之中CRISPR-Cas型/亞型的分佈。

[0177] CRISPR-Cas系統的作用通常被分為三個階段：(1) 自我調整或間隔區整合，(2) CRISPR座位初級轉錄物（前crRNA）的加工和包含間隔區和對應於CRISPR重複序列的5'和3'片段可變區的crRNA的成熟，以及 (3) DNA（或RNA）干擾。大多數的已知CRISPR-Cas系統中存在的兩種蛋白質Cas1和Cas2足以用於將間

隔區插入到CRISPR盒中。這兩種蛋白質形成為此自我調整過程所需要的複合物；Cas1的內切核酸酶活性係為間隔區整合所需要的，而Cas2似乎執行非酶性功能。Cas1-Cas2複合物表示CRISPR-Cas的高度保守的“資訊加工”模組，該模組似乎准自主（quasi-autonomous）於系統的其餘部分。（參見CRISPR-Cas系統的注釋和分類（Annotation and Classification of CRISPR-Cas Systems），馬卡洛夫·KS（Makarova KS）、庫尼恩·EV（Koonin EV），分子生物學方法（Methods Mol Biol.） 2015；1311:47-75）。

[0178] 先前所述的第2類系統，即II型和推定的V型僅由cas操縱子中的三個或四個基因組成，即包括自我調整模組（不參與干擾的cas1-cas2基因對）的cas1和cas2基因，負責干擾但還有助於前crRNA加工和自我調整的單一多結構域效應蛋白，以及常常在至少一些II型系統中可有可無的具有不典型功能的第四基因（並且在一些情況下第四基因係cas4（生物化學或電腦類比證據顯示Cas4係具有三個半胱胺酸C末端簇的PD-(DE)xK超家族核酸酶；具有5'-ssDNA外切核酸酶活性）或編碼失活的ATP酶的csn2）。在大多數情況下，CRISPR陣列和稱為tracrRNA（反式編碼小CRISPR RNA）的不同RNA種類的基因與第2類cas操縱子相鄰。tracrRNA與對應CRISPR陣列內的重複序列部分同源並且是為前crRNA的加工所必需的，該加工由不與CRISPR-Cas座位相關聯的普遍存在的細菌酶RNA酶III催化。

[0179] Cas1係大多數CRISPR-Cas系統中存在的最保守的蛋白質並且相比於其他Cas蛋白進化較慢。因此，Cas1系統發育已用

作CRISPR-Cas系統分類的指南。生物化學或電腦類比證據顯示Cas1係金屬依賴性的去氧核糖核酸酶。大腸桿菌中Cas1的缺失使得對DNA損傷的敏感性增加並且使得染色體分離減弱，如在“CRISPR-Cas系統在細菌抗病毒免疫和DNA修復中的雙重功能（A dual function of the CRISPR-Cassystem in bacterial antivirus immunity and DNA repair）”，巴布·M（Babu M）等人 分子微生物學（Mol Microbiol） 79:484–502 (2011)中所述。生物化學或電腦類比證據顯示Cas 2係對富含U的區具有特異性的RNA酶並且是雙股的DNA酶。

[0180] 本發明的方面涉及與第2類CRISPR-Cas系統相關聯的新型效應蛋白的鑒定和工程化。在一較佳的實施方式中，效應蛋白包含單亞基的效應物模組。在另一個實施方式中，效應蛋白在原核細胞或真核細胞中具有功能以便用於體外、體內或離體應用。本發明的一個方面涵蓋用於預測新的第2類CRISPR-Cas系統和鑒定其中組分的計算方法和演算法。

[0181] 在一個實施方式中，鑒定新型第2類CRISPR-Cas座位的計算方法包括以下步驟：檢測編碼Cas1蛋白的所有疊連群；鑒定20 kB cas1基因內的所有預測蛋白編碼基因；將鑒定的基因與Cas蛋白特異性表現譜和預測的CRISPR陣列進行比較；選擇含有大於500個胺基酸（>500 aa）的蛋白質的未分類候選CRISPR-Cas座位；使用PSI-BLAST和HHPred分析選擇的候選物，從而分離和鑒定新型第2類CRISPR-Cas座位。除以上提到的步驟之外，候選物的另外分析可以藉由搜索基因組學資料庫尋找另外同源物來進行。

[0182] 在一個方面中，檢測編碼Cas1蛋白的所有疊連群藉由GenemarkS進行，GenemarkS為基因預測程序，如“GeneMarkS：用於預測生物基因組中的基因啟動子（starts）的自培訓方法（GeneMarkS: a self-training method for prediction of gene starts in microbial genomes），發現調節區中的序列模體的意義（Implications for finding sequence motifs in regulatory regions）”，約翰·貝瑟麥（John Besemer），亞歷山大·羅明納茲（Alexandre Lomsadze）和馬克·波羅多夫斯基（Mark Borodovsky），核酸研究（Nucleic Acids Research） (2001) 29，第2607-2618頁中所述，該文獻藉由引用結合在此。

[0183] 在一個方面中，鑒定所有預測蛋白編碼基因藉由以下方式進行：將鑒定基因與Cas蛋白特異性表現譜進行比較並且根據NCBI保守結構域資料庫（CDD）注釋它們，該CDD係由用於古結構域和全長蛋白的充分注釋的多重序列比對模型的集合組成的單子注釋資源。該等可用作位置特異性評分矩陣(PSSM)以經由RPS-BLAST快速鑒定蛋白質序列中的保守結構域。CDD內容物包括NCBI管理的（curated）結構域，該結構域使用3D結構資訊來明確地限定結構域邊界並且提供對序列/結構/功能關係的見解，以及結構域模型，該等模型從許多外來源資料庫（Pfam、SMART、COG、PRK、TIGRFAM）導入。在另一個方面中，使用PILER-CR程序預測CRISPR陣列，該程序係用於發現CRISPR重複序列的公共領域軟體，如“PILER-CR：CRISPR重複序列的快速且精確的鑒定”，愛德格，R.C.（Edgar, R.C.） 生物資訊學（Bioinformatics），1月20日；8:18(2007)中所述，該文獻藉由引用結合在此。

[0184] 在另一個方面中，使用PSI-BLAST（位置特異性反覆運算基本局部比對搜索工具）進行逐案分析。PSI-BLAST由使用蛋白質-蛋白質BLAST檢測的在給定得分閾值之上的序列的多重序列比對得出位置特異性評分矩陣（PSSM）或表現譜。此PSSM用於進一步搜索資料庫以尋找新的匹配，並且進行校正以使用該等新檢測到的序列進行後續反覆運算。因此，PSI-BLAST提供了一種檢測蛋白質之間的遠源關係之手段。

[0185] 在另一個方面中，使用HHpred進行逐案分析，該HHpred係一種用於序列資料庫搜索和結構預測之方法，該方法與BLAST或PSI-BLAST一樣易於使用並且同時在發現遠距離同源物方面更具有敏感性。實際上，HHpred敏感性同目前可獲得的用於結構預測的最強有力的服務器具有競爭力。HHpred係第一個基於表現譜隱蔽瑪律科夫（Markov）模型（HMM）的配對比較的伺服器。最常規的序列搜索方法搜索序列資料庫諸如UniProt或NR，而HHpred搜索比對資料庫如Pfam或SMART。這極大簡化了許多序列家族而不是雜亂的單一序列的命中（hit）列表。所有大型可公共獲得的表現譜和比對資料庫可藉由HHpred獲得。HHpred接受單一查詢序列或多重比對作為輸入值。僅在數分鐘內，HHpred以類似於PSI-BLAST格式的易讀格式返回搜索結果。搜索選擇包括局部比對或總體比對和二級結構相似性評分。HHpred可產生配對的查詢模板序列比對、合併的查詢模板多重比對（例如，對於可遞搜索）以及根據HHpred比對藉由MODELLER軟體計算的3D結構模型。

[0186] 其中核酸係DNA或RNA並且在一些方面中還可以是

指DNA-RNA雜交體或其衍生物的術語“核酸靶向系統”總體是指涉及DNA或RNA靶向CRISPR相關(“Cas”)基因的表現或引導該等基因的活性的轉錄物和其他元件，該等轉錄物和其他元件可包括編碼DNA或RNA靶向Cas蛋白的序列和包含CRISPR RNA(crRNA)的DNA或RNA靶向指導RNA以及（在CRISPR-Cas9系統並非所有系統中）反式啟動CRISPR-Cas系統RNA（tracrRNA）序列或來自DNA或RNA靶向CRISPR座位的其他序列或轉錄物。在在此所述的Cpf1 DNA靶向的RNA指導型內切核酸酶系統中，tracrRNA序列不是需要的。總的來說，RNA靶向系統的特徵在於在靶RNA序列的位點處促進RNA靶向複合物形成的元件。在形成DNA或RNA靶向複合物的情況下，“靶序列”係指DNA或RNA靶向指導RNA被設計為與其具有互補性的DNA或RNA序列，其中靶序列與RNA靶向指導RNA之間的雜交促進RNA靶向複合物的形成。在一些實施方式中，靶序列位於細胞的核或細胞質中。

[0187] 在本發明的一個方面中，本申請的新型DNA靶向系統（還稱之為DNA靶向CRISPR-Cas或CRISPR-Cas DNA靶向系統）係基於鑒定的型V（例如亞型V-A和亞型V-B）Cas蛋白的，該等蛋白不需要產生定制蛋白來靶向特異性DNA序列，相反，單一效應蛋白或酶可以由RNA分子程式設計來識別特異性DNA靶，換句話說，可以使用所述RNA分子來將酶募集至特異性DNA靶。本發明的方面具體涉及DNA靶向的RNA指導型Cpf1 CRISPR系統。

[0188] 在本發明的一個方面中，本申請的新型RNA靶向系統（還稱之為RNA或RNA靶向CRISPR-Cas或CRISPR-Cas系統RNA

靶向系統）係基於鑒定的VI型Cas蛋白的，該等蛋白不需要產生定制蛋白來靶向特異性RNA序列，相反，單一酶可由RNA分子程式設計來識別特異性RNA靶，換句話說，可使用所述RNA分子來將酶募集至特異性RNA靶。

[0189] 在此所述的核酸靶向系統、載體系統、載體和組成物可以用於多種核酸靶向應用，改變或修改基因產物諸如蛋白質的合成、核酸切割、核酸編輯、核酸剪接；靶核酸的運輸、靶核酸的追蹤、靶核酸的分離、靶核酸的視覺化等中。

[0190] 如在此所用的，Cas蛋白或CRISPR酶係指新的CRISPR-Cas系統分類中呈現的任一種蛋白質。在一有利的實施方式中，本發明涵蓋V型CRISPR-Cas座位，例如Cpf1編碼座位（指代為亞型V-A）中鑒定的效應蛋白。目前，亞型V-A座位涵蓋cas1、cas2、不同的基因指代的*cpf1*以及CRISPR陣列。Cpf1（CRISPR相關蛋白Cpf1，亞型PREFRAN）係大蛋白質（約1300個胺基酸），其含有與Cas9的相應結構域同源的RuvC樣核酸酶結構域以及Cas9的特徵性富精胺酸簇的對應物。然而，Cpf1缺乏所有Cas9蛋白中存在的HNH核酸酶結構域，並且RuvC樣結構域在Cpf1序列中是連續的，這與其中含有包含HNH結構域的長插入物的Cas9相反。因此，在特定實施方式中，CRISPR-Cas酶僅包含RuvC樣核酸酶結構域。

[0191] Cpf1基因可以見於若干種不同的細菌基因組中，典型地與cas1、cas2和cas4基因以及CRISPR盒（例如，弗朗西絲菌屬cf. 新殺手Fx1（Francisella cf. novicida Fx1）的FNFX1_1431-

FNFX1_1428）在同一座位中。因此，此推定的新型CRISPR-Cas系統的佈置似乎類似於II-B型的佈置。此外，與Cas9類似，Cpf1蛋白含有與轉位子ORF-B同源的便於鑒定的C末端區並且包含活性的RuvC樣核酸酶、富含精胺酸的區和Zn指（不存在於Cas9中）。然而，與Cas9不同，Cpf1還存在於沒有CRISPR-Cas環境的若干種基因組中並且其與ORF-B的相對較高相似性表明其可能是轉位子組分。表明如果此是真正的CRISPR-Cas系統並且Cpf1係Cas9的功能類似物，則其將是新型CRISPR-Cas類型，即V型（參見CRISPR-Cas系統的注釋和分類，馬卡洛夫·KS（Makarova KS）、庫尼恩·EV（Koonin EV），分子生物學方法（Methods Mol Biol.）2015；1311:47-75）。然而，如在此所述，將Cpf1指代為亞型V-A以將其與C2c1p區分，該C2c1p不具有相同的結構域結構並且因此被指代為亞型V-B。

[0192] 在一有利的實施方式中，本發明涵蓋包含在指代為亞型V-A的Cpf1座位中鑒定的效應蛋白的組成物和系統。

[0193] 本發明的方面還涵蓋在此所述的組成物和系統在基因組工程中例如用於在體外、體內或離體中改變或操縱原核細胞或真核細胞中的一種或多種基因的表現或一種或多種基因產物的方法和用途。

[0194] 在本發明的實施方式中，術語成熟crRNA和指導RNA以及單一指導RNA如前面引用的文獻諸如WO 2014/093622（PCT/US2013/074667）中那樣可互換地使用。總的來說，指導序列係與靶多核苷酸序列具有足夠互補性以與靶序列雜交並引導

CRISPR複合物與靶序列的序列特異性結合的任何多核苷酸序列。在一些實施方式中，當使用適合比對演算法進行最佳比對時，指導序列與其相應靶序列之間的互補程度係約或超過約50%、60%、75%、80%、85%、90%、95%、97.5%、99%、或更大。最佳比對可以藉由使用用於比對序列的任何適合的演算法來確定，該等演算法的非限制性實例包括史密斯-沃特曼（Smith-Waterman）演算法、尼德曼-溫施演算法（Needleman-Wunsch algorithm）、基於巴羅斯-惠勒（Burrows-Wheeler）轉換的演算法（例如，巴羅斯-惠勒比對儀）、ClustalW、Clustal X、BLAT、諾沃比對（Novoalign）（諾沃克拉夫特技術公司（Novocraft Technologies）；可在www.novocraft.com處獲得）、ELAND（加利福尼亞州聖迭哥億明達（Illumina, San Diego, CA））、SOAP（可在soap.genomics.org.cn處獲得）、以及Maq（可在maq.sourceforge.net處獲得）。在一些實施方式中，指導序列的長度係約或超過約5、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、35、40、45、50、75、或更多個核苷酸。在一些實施方式中，指導序列的長度係小於約75、50、45、40、35、30、35、25、20、15、12、或更少個核苷酸。較佳的是指導序列係10-30個核苷酸長。指導序列引導CRISPR複合物與靶序列的序列特異性結合的能力可以藉由任何適合的測定來評估。例如，足以形成CRISPR複合物的CRISPR系統的組分（包括有待測試的指導序列）可以諸如藉由用編碼CRISPR序列組分的載體進行轉染來提供給具有相應靶序列的宿主細胞，隨後諸如藉由在此所述的Surveyor測定評估靶序列內的優先切割。類似地，靶多核苷酸序列的切割可以在試管中

藉由以下方式進行評估：藉由提供靶序列、CRISPR複合物的組分（包括有待測試的指導序列）和不同於測試指導序列的對照指導序列並且在測試指導序列反應與對照指導序列反應之間比較靶序列處的結合或切割速率。其他測定係可能的，並且是熟習該項技術者能夠想到的。指導序列可以被選擇為靶向任何靶序列。在一些實施方式中，靶序列係細胞基因組中的序列。示例性靶序列包括靶基因組中獨特的那些。

[0195] 總的來說，在整個說明書中，術語"載體"係指能夠轉運它所連接的另一個核酸的核酸分子。載體包括但不限於，單股、雙股或部分雙股的核酸分子；包含一個或多個游離端、不包含游離端（例如，環狀）的核酸分子；包含DNA、RNA或二者的核酸分子；以及本領域已知的其他種類的多核苷酸。一種類型的載體係"質粒"，該質粒係指一種環狀雙股DNA環，可以諸如藉由標準分子選殖技術向該環中插入另外的DNA區段。另一種類型的載體係病毒載體，其中病毒來源的DNA或RNA序列存在於包裝到病毒（例如，逆轉錄病毒、複製缺陷型逆轉錄病毒、腺病毒、複製缺陷型腺病毒、腺伴隨病毒）中的載體中。病毒載體還包括由病毒攜帶來轉染到宿主細胞中的多核苷酸。某些載體能夠在它們被引入至其中的宿主細胞中自主複製（例如，具有細菌複製起點的細菌載體以及附加型哺乳動物載體）。其他載體（例如，非附加型哺乳動物載體）在引入到宿主細胞後被整合到宿主細胞的基因組中，並且從而隨著宿主基因組一起複製。此外，某些載體能夠引導它們可操作地連接的基因的表現。此類載體在此被稱為"表現載體"。用於真核細胞並且在真核細胞中產生表現的載體可以在此稱之為

“真核表現載體”。在重組DNA技術中採用的常見表現載體常常是質粒形式。

[0196] 重組表現載體可以包含處於適用於在宿主細胞中表現核酸的形式的本發明的核酸，這意味著重組表現載體包含一個或多個調節元件，該等調節元件可以基於用於表現的宿主細胞來選擇，可操作地連接至有待表現的核酸序列。在重組表現載體內，“可操作地連接”旨在意指感興趣的核苷酸序列以允許核苷酸序列表現（例如，在體外轉錄/翻譯系統中或當該載體被引入到宿主細胞時在宿主細胞中）的方式連接至一個或多個調節元件。

[0197] 術語“調節元件”旨在包括啟動子、增強子、內部核糖體進入位點（IRES）以及其他表現控制元件（例如，轉錄終止信號，諸如多聚腺苷酸化信號和聚U序列）。此類調節元件描述於例如高德爾（Goeddel），基因表現技術：酶學方法（GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY）185，學術出版社（Academic Press），加利福尼亞州聖迭哥 (1990)中。調節元件包括引導核苷酸序列在許多類型的宿主細胞中連續表現的那些元件和引導核苷酸序列僅在某些宿主細胞中表現的那些元件（例如，組織特異性調節序列）。組織特異性啟動子可以引導主要在希望的感興趣的組織諸如肌肉、神經元、骨骼、皮膚、血液、特定器官（例如，肝臟、胰臟）、或特定細胞類型（例如，淋巴細胞）中的表現。調節元件還可以時間依賴性方式諸如細胞週期依賴性或發育階段依賴性方式引導表現，這可以是或也可以不是組織特異性或細胞類型特異性的。在一些實施方式中，載體

包含一個或多個pol III啟動子（例如，1、2、3、4、5、或更多個pol III啟動子）、一個或多個pol II啟動子（例如，1、2、3、4、5、或更多個pol II啟動子）、一個或多個pol I啟動子（例如，1、2、3、4、5、或更多個pol I啟動子）、或其組合。pol III啟動子的實例包括但不限於，U6和H1啟動子。pol II啟動子的實例包括但不限於，逆轉錄病毒勞斯氏肉瘤病毒（Rous sarcoma virus）（RSV）LTR啟動子（視情況具有RSV增強子）、巨細胞病毒（CMV）啟動子（視情況具有CMV增強子）[例如，參見博沙特（Boshart）等人，細胞（Cell），41:521-530(1985)]、SV40啟動子、二氫葉酸還原酶啟動子、β-肌動蛋白啟動子、磷酸甘油激酶（PGK）啟動子、以及EF1α啟動子。術語"調節元件"還涵蓋增強子元件，諸如WPRE；CMV增強子；HTLV-I的LTR中的R-U5'區段（分子細胞生物學（Mol.Cell. Biol.），第8(1)卷，第466-472頁，1988）；SV40增強子；以及兔β-球蛋白的外顯子2與3之間的內含子序列（美國國家科學院院刊（Proc.Natl.Acad.Sci.USA.），第78(3)卷，第1527-31頁，1981）。熟習該項技術者將瞭解的是，表現載體的設計可以取決於如有待轉化的宿主細胞的選擇、所希望的表現水平等的此類因素。載體可以引入到宿主細胞中從而產生由在此所述的核酸編碼的轉錄物、蛋白質或肽，包括融合蛋白或肽（例如，成簇規律間隔短迴文重複序列（CRISPR）轉錄物、蛋白質、酶、其突變體形式、其融合蛋白等）。

[0198] 有利的載體包括慢病毒和腺伴隨病毒並且此類載體類型還可以針對靶向的特定細胞類型來選擇。

[0199] 如在此所用，術語型V CRISPR-Cas座位效應蛋白的“crRNA”或“指導RNA”或“單一指導RNA”或“sgRNA”或“一種或多種核酸組分”包括與靶核酸序列具有足夠互補性以與靶核酸序列雜交並引導核酸靶向複合物與靶核酸序列的序列特異性結合的任何多核苷酸序列。在一些實施方式中，當使用適合比對演算法進行最佳比對時，互補程度係約或超過約50%、60%、75%、80%、85%、90%、95%、97.5%、99%、或更大。最佳比對可以藉由使用用於比對序列的任何適合的演算法來確定，該等演算法的非限制性實例包括史密斯-沃特曼演算法、尼德曼-溫施演算法、基於巴羅斯-惠勒轉換的演算法（例如，巴羅斯-惠勒比對儀）、ClustalW、Clustal X、BLAT、諾沃比對（諾沃克拉夫特技術公司；可在www.novocraft.com處獲得）、ELAND（加利福尼亞州聖迭哥億明達）、SOAP（可在soap.genomics.org.cn處獲得）、以及Maq（可在maq.sourceforge.net處獲得）。指導序列（在核酸靶向指導RNA內）引導核酸靶向複合物與靶核酸序列的序列特異性結合的能力可以是藉由任何適合的測定來評估。例如，足以形成核酸靶向複合物的核酸靶向CRISPR系統的組分（包括有待測試的指導序列）可以諸如藉由用編碼核酸靶向複合物組分的載體進行轉染來提供給具有相應靶核酸序列的宿主細胞，隨後諸如藉由在此所述的Surveyor測定評估靶核酸序列內的優先靶向（例如切割）。類似地，靶核酸序列的切割可以在試管中藉由以下方式進行評估：藉由提供靶核酸序列、核酸靶向複合物的組分（包括有待測試的指導序列）和不同於測試指導序列的對照指導序列，並且在測試指導序列反應與對照指導序列反應之間比較靶序列處的結合或切割速率。

其他測定係可能的，並且是熟習該項技術者能夠想到的。指導序列和因此核酸靶向指導RNA可以被選擇成靶向任何靶核苷酸序列。靶序列可以是DNA。靶序列可以是任何RNA序列。在一些實施方式中，靶序列可以是選自下組的RNA分子內的序列，該組由以下各項組成：信使RNA（mRNA）、前mRNA、核糖體RNA（rRNA）、轉移RNA（tRNA）、微小RNA（miRNA）、小干擾RNA（siRNA）、小核RNA（snRNA）、小核仁RNA（snoRNA）、雙股RNA（dsRNA）、非編碼RNA（ncRNA）、長的非編碼RNA（lncRNA）以及細胞質小RNA（scRNA）。在一些較佳的實施方式中，靶序列可以是選自由mRNA、前mRNA和rRNA組成的組的RNA分子內的序列。在一些較佳的實施方式中，靶序列可以是選自由ncRNA和lncRNA組成的組的RNA分子內的序列。在一些更較佳的實施方式中，靶序列可以是mRNA分子或前mRNA分子內的序列。

[0200] 在一些實施方式中，核酸靶向指導RNA被選擇以減小該RNA靶向指導RNA內的二級結構程度。在一些實施方式中，當進行最佳折疊時，核酸靶向指導RNA的約或小於約75%、50%、40%、30%、25%、20%、15%、10%、5%、1%，或更少的核苷酸參與自互補鹼基配對。最佳折疊可以是藉由任何適合的多核苷酸折疊演算法來確定。一些程序係基於計算最小吉布斯自由能。一種這樣的演算法的實例係mFold，如藉由朱克（Zuker）和施蒂格勒（Stiegler）（核酸研究9 (1981), 133-148）。另一個示例性折疊演算法係使用質心結構預測演算法的線上網站伺服器RNAfold，它係維也納大學的理論化學研究所開發的（例如，參見A.R.·格魯伯（A.R. Gruber）等人，2008，細胞 106(1): 23-24；以及PA·凱爾（PA

Carr）和GM·丘奇（GM Church），2009，自然生物技術（Nature Biotechnology）27(12): 1151-62）。

[0201] "tracrRNA"序列或類似術語包括與crRNA序列具有足夠互補性以進行雜交的任何多核苷酸序列。如上文所指出的，在本發明的實施方式中，tracrRNA不是為Cpf1效應蛋白複合物的切割活性所需要的。

[0202] 申請人還進行了激發實驗以驗證V型/VI型蛋白諸如Cpf1/C2c1/C2c2的DNA靶向和切割能力。此實驗與大腸桿菌中的StCas9異源表現的類似工作（薩普拉諾薩克斯，R.（Sapranauskas, R.）等人 核酸研究 39, 9275–9282 (2011)）極為相似。申請人將含有PAM和抗性基因兩者的質粒引入到異源大腸桿菌中,並且然後接種在相應抗生素上。如果存在質粒的DNA切割，則申請人觀察不到有活力的群落。

[0203] 在進一步細節中，如下針對DNA靶進行測定。在此測定中使用兩種大腸桿菌菌株。一種攜帶編碼來自細菌菌株的內源性效應蛋白座位的質粒。另一種菌株攜帶空質粒（例如pACYC184，對照菌株）。將所有可能的7或8 bp PAM序列呈遞在抗生素抗性質粒（具有胺苄青黴素抗性基因的pUC19）上。將PAM定位成靠近原型間隔區1的序列（內源性效應蛋白座位中的第一間隔區的DNA靶）。選殖了兩個PAM文庫。一個具有原型間隔區的8個隨機bp 5'（例如總的65536個不同PAM序列 = 複雜度）。另一個文庫具有原型間隔區的7個隨機bp 3'（例如總複雜度係16384個不同的PAM）。將兩個文庫選殖成具有平均500個質粒/可能的PAM。用5'PAM和

3'PAM文庫在單獨的轉化中轉化測試菌株和對照菌株並且將轉化的細胞分別接種在胺苄青黴素板上。使用質粒的識別和隨後的切割/干擾使得細胞對胺苄青黴素易感並且阻止了生長。轉化後大約12 h，收穫由測試菌株和對照菌株形成的所有群落並且分離出質粒DNA。使用質粒DNA作為用於PCR擴增和隨後的深度定序的模板。未轉化的（untransfomed）文庫中的所有PAM的表現度顯示轉化細胞中的PAM的預期表現度。對照菌株中發現的所有PAM的表現度顯示真實的表現度。測試菌株中的所有PAM的表現度顯示哪個PAM未被酶識別並且與對照菌株的比較允許提取出缺失的PAM的序列。

[0204] 在CRISPR-Cas9系統的一些實施方式中，當進行最佳比對時，tracrRNA序列與crRNA序列之間的互補程度係沿兩者中較短者的長度。如在此所述的，在本發明的實施方式中，tracrRNA係不需要的。在先前所述的CRISPR-Cas系統（例如CRISPR-Cas9系統）的一些實施方式中，嵌合合成的指導RNA（sgRNA）設計可以在crRNA與tracrRNA之間摻入至少12 bp的雙股體結構，然而在在此所述的Cpf1 CRISPR系統中此類嵌合RNA（chi-RNA）係不可能的，因為該系統不利用tracrRNA。

[0205] 為了最小化毒性和脫靶效應，重要的是控制所遞送的核酸靶向指導RNA的濃度。核酸靶向指導RNA的最佳濃度可以藉由以下方式來確定：測試不同濃度的細胞模型或非人類真核動物模型並且使用深度定序分析潛在的脫靶基因組座位處的修飾程度。得到最高的中靶（on-target）修飾水平同時使脫靶修飾水平最小化

的濃度應被選擇用於體內遞送。核酸靶向系統有利地是來源於V型/VI型CRISPR系統。在一些實施方式中，核酸靶向系統的一個或多個元件係來源於包含內源性RNA靶向系統的特定生物體。在本發明的較佳的實施方式中，RNA靶向系統係V型/VI型CRISPR系統。在特定實施方式中，V型/VI型RNA靶向Cas酶係Cpf1/C2c1/C2c2。Cas蛋白的非限制性實例包括Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas6、Cas7、Cas8、Cas9（也稱為Csn1和Csx12）、Cas10、Csy1、Csy2、Csy3、Cse1、Cse2、Csc1、Csc2、Csa5、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx15、Csf1、Csf2、Csf3、Csf4，其同源物，或其修飾版本。在實施方式中，諸如在此提及的Cpf1/C2c1/C2c2的V型/VI型蛋白還涵蓋諸如Cpf1/C2c1/C2c2的V型/VI型蛋白的同源物或異種同源物。術語“異種同源物（orthologue）”（在此也稱之為“異種同源物（ortholog）”）和“同源物（homologue）”（在此也稱之為“同源物（homolog）”）係本領域熟知的。作為進一步指導，如在此所用的蛋白質的“同源物”係屬於同一種類的蛋白質，該蛋白質執行與作為其同源物的蛋白質相同或類似的功能。同源蛋白可以是但不需要是結構相關的，或僅是部分結構相關的。如在此所用的蛋白質的“異種同源物”係屬於不同種類的蛋白質，該蛋白質執行與作為其異種同源物的蛋白質相同或類似的功能。直向同源蛋白可以是但不需要是結構相關的，或僅是部分結構相關的。同源物和異種同源物可以是藉由同源模擬（例如，參見格里爾（Greer），科學（Science） 第228卷 (1985) 1055，和布倫德爾（Blundell）等人

歐洲生物化學雜誌（Eur J Biochem） 第172卷 (1988), 513）或“結構BLAST”（戴伊·F（Dey F），克利夫·張·Q（Cliff Zhang Q），彼德雷·D（Petrey D），霍尼格·B（Honig B），針對“結構BLAST”：使用結構關係推斷功能（Toward a "structural BLAST": using structural relationships to infer function），蛋白質科學（Protein Sci.），2013年4月；22(4):359-66. doi: 10.1002/pro.2225。）來鑒定。還參見CRISPR-Cas座位領域中的什馬科夫（Shmakov）等人（2015）的申請。同源蛋白可以是但不需要是結構相關的，或僅是部分結構相關的。在特定實施方式中，如在此提及的Cpf1的同源物或異種同源物與Cpf1具有至少80%、更較佳的是至少85%、甚至更較佳的是至少90%例如像至少95%的序列同源性或一致性。在另外的實施方式中，如在此提及的Cpf1的同源物或異種同源物與野生型Cpf1具有至少80%、更較佳的是至少85%、甚至更較佳的是至少90%例如像至少95%的序列一致性。在Cpf1具有一個或更多個突變的情況下（突變型），如在此提及的所述Cpf1的同源物或異種同源物與突變型Cpf1具有至少80%、更較佳的是至少85%、甚至更較佳的是至少90%例如像至少95%的序列一致性。

[0206] 在一個實施方式中，V型Cas蛋白可以是包括但不限於下項的屬的生物體的異種同源物：胺基酸球菌屬某種、毛螺旋菌科細菌或牛莫拉氏菌；在特定實施方式中，V型Cas蛋白可以是包括但不限於下項的物種的生物體的異種同源物：胺基酸球菌屬某種*BV3L6*；毛螺旋菌科細菌*ND2006*（LbCpf1）或牛莫拉氏菌*237*。在特定實施方式中，如在此提及的Cpf1的同源物或異種同源物與在此揭露的Cpf1序列中的一種或多種具有至少80%、更較佳的是

至少85%、甚至更較佳的是至少90%例如像至少95%的序列同源性或一致性。在另外的實施方式中，如在此提及的Cpf的同源物或異種同源物與野生型FnCpf1、AsCpf1或LbCpf1具有至少80%、更較佳的是至少85%、甚至更較佳的是至少90%例如像至少95%的序列一致性。

[0207] 在特定實施方式中，本發明的Cpf1蛋白與FnCpf1、AsCpf1或LbCpf1具有至少60%、更具體地至少70諸如至少80%、更較佳的是至少85%、甚至更較佳的是至少90%例如像至少95%的序列同源性或一致性。在另外的實施方式中，如在此提及的Cpf1蛋白與野生型AsCpf1或LbCpf1具有至少60%，諸如至少70%、更具體地至少80%、更較佳的是至少85%、甚至更較佳的是至少90%例如像至少95%的序列一致性。在特定實施方式中，本發明的Cpf1蛋白與FnCpf1具有小於60%的序列一致性。技術人員將理解這包括Cpf1蛋白的截短形式，由此定序一致性係在截短形式的長度上測定。

[0208] 鑒定CRISPR-Cas系統酶的異種同源物的一些方法可以包括鑒定感興趣的基因組中的tracr序列。鑒定tracr序列可以與以下步驟相關：搜索資料庫中的同向重複序列或tracr配對序列以鑒定包含CRISPR酶的CRISPR區。在正義方向和反義方向上搜索側接CRISPR酶的CRISPR區中的同源序列。尋找轉錄終止子和二級結構。鑒定不是同向重複序列或tracr配對序列但與同向重複序列或

tracr配對序列具有超過50%的一致性的任何序列以作為潛在的tracr序列。獲取潛在的tracr序列並且分析與其相關聯的轉錄終止子序列。在此系統中，RNA定序數據揭示藉由計算鑒定的潛在tracrRNA僅有微弱表現，這表明了tracrRNA可能不是為本發明系統的功能所需的可能性。在對FnCpf1座位進一步評價並加上體外切割結果之後，申請人推斷藉由Cpf1效應蛋白複合物進行靶DNA切割不需要tracrRNA。申請人確定僅包含Cpf1效應蛋白和crRNA（包含同向重複序列和指導序列的指導RNA）的Cpf1效應蛋白複合物足以切割靶DNA。

[0209] 應瞭解，在此所述的任一功能可以被工程化到來自其他異種同源物的CRISPR酶中，包括包含來自多個異種同源物的片段的嵌合酶。在此其他地方描述了此類異種同源物的實例。因此，嵌合酶可以包含包括但不限於下項的屬的生物體的CRISPR酶異種同源物的片段：棒狀桿菌屬、薩特氏菌屬（*Sutterella*）、軍團桿菌屬、密螺旋體屬、產線菌屬（*Filifactor*）、真細菌屬、鏈球菌屬、乳酸桿菌屬、支原體屬、擬桿菌屬、黃沃拉菌屬（*Flaviivola*）、黃桿菌屬、螺旋體屬、固氮螺菌屬、葡糖醋桿菌屬、奈瑟氏菌屬、羅氏菌屬、細小棒菌屬、葡萄球菌屬、硝化裂化菌屬、支原體屬以及彎曲桿菌屬。嵌合酶可以包含第一片段和第二片段並且片段可以是屬於在此所提到的屬或在此所提到的物種的生物體的CRISPR酶異種同源物的；有利的是，片段來自不同物種的CRISPR酶異種同源物。

[0210] 在實施方式中，V型/VI型RNA靶向效應蛋白，具體地

是如在此提及的Cpf1/C2c1/C2c2還涵蓋Cpf1/C2c1/C2c2的功能變體或其同源物或其異種同源物。如在此所用的蛋白質的“功能變體”係指此蛋白質的變體，該變體保留該蛋白的至少部分活性。功能變體可以包括突變體（其可以是插入、缺失或置換突變體），包括同質多形體等。還包括在功能變體內的是此蛋白質與別的，通常是不相關的核酸、蛋白質、多肽或肽的融合產物。功能變體可以是天然存在的或可以是人造的。有利的實施方式可以包括工程化的或非天然存在的V型/VI型RNA靶向效應蛋白，例如Cpf1/C2c1/C2c2或其異種同源物或同源物。

[0211] 在一個實施方式中，編碼V型/VI型RNA靶向效應蛋白，具體地是Cpf1/C2c1/C2c2或其異種同源物或同源物的一個或多個核酸分子可以被密碼子優化為在真核細胞中表現。真核生物可以是在此所討論的。一個或多個核酸分子可以是工程化的或非天然存在的。

[0212] 在一個實施方式中，V型/VI型RNA靶向效應蛋白，具體地是Cpf1/C2c1/C2c2或其異種同源物或同源物可以包含一個或多個突變（並且因此編碼所述效應蛋白的一個或多個核酸分子可以具有一個或多個突變）。突變可以是人工引入的突變並且可以包括但不限於催化結構域中的一個或多個突變。參照Cas9酶，催化結構域的實例可以包括但不限於RuvC I、RuvC II、RuvC III以及HNH結構域。

[0213] 在一個實施方式中，V型/VI型蛋白，諸如Cpf1/C2c1/C2c2或其異種同源物或同源物可以用作融合至或可操

作地連接至功能結構域的通用的核酸結合蛋白。示例性功能結構域可以包括但不限於，翻譯起始區、翻譯活化物、翻譯阻遏物、核酸酶（具體地是核糖核酸酶）、剪接體、珠粒、光誘導型/控制型的結構域或化學誘導型/控制型的結構域。

[0214] 在一些實施方式中，未修飾的核酸靶向效應蛋白可以具有切割活性。在一些實施方式中，RNA靶向效應蛋白可以引導靶序列或靠近靶序列的位置處的，諸如靶序列內和/或靶序列補體內或與靶序列相關聯的序列處的一條或兩條核酸（DNA或RNA）股的切割。在一些實施方式中，核酸靶向效應蛋白可以引導從靶序列的第一個或最後一個核苷酸開始的約1、2、3、4、5、6、7、8、9、10、15、20、25、50、100、200、500或更多個鹼基對內的一條或兩條DNA或RNA股的切割。在一些實施方式中，切割可以是交錯的，即產生黏性末端。在一些實施方式中，切割係交錯切割的，產生了5'突出端。在一些實施方式中，切割係交錯切割的，產生了具有1至5個核苷酸，較佳的是4或5個核苷酸的5'突出端。在一些實施方式中，切割位點遠離PAM，例如切割發生在非靶股上的第18核苷酸後面和靶股上的第23核苷酸後面（圖97A）。在一些實施方式中，切割位點出現在非靶股上的第18核苷酸（從PAM開始計數）後面和靶股上的第23核苷酸（從PAM開始計數）後面（圖97A）。在一些實施方式中，載體編碼可以相對於相應野生型酶發生突變的核酸靶向效應蛋白，使得突變型核酸靶向效應蛋白缺乏切割含有靶序列的靶多核苷酸的一條或兩條DNA或RNA股的能力。作為另一個實例，Cas蛋白的兩個或更多個催化結構域（例如Cas9蛋白的RuvC I、RuvC II和RuvC III或HNH結構域）可以突變成產

生實質性缺乏所有DNA切割活性的突變型Cas蛋白。如在此所述，Cpf1效應蛋白的相應催化結構域也可以突變成產生缺乏所有DNA切割活性或具有實質性降低的DNA切割活性的突變型Cpf1效應蛋白。在一些實施方式中，當突變型酶的RNA切割活性不超過該酶的非突變形式的核酸切割活性的約25%、10%、5%、1%、0.1%、0.01%或更少時，核酸靶向效應蛋白可以被認為是實質性缺乏所有RNA切割活性的；一個實例可以是當突變形式的核酸切割活性係零或與非突變形式相比是可忽視的時候。效應蛋白可以參照與具有來自V型/VI型CRISPR系統的多個核酸酶結構域的最大核酸酶享有同源性的酶的一般類別來鑒定。最較佳的是，效應蛋白係V型/VI型蛋白諸如Cpf1/C2c1/C2c2。在另外的實施方式中，效應蛋白係V型蛋白。關於衍生，申請人表示，就與野生型酶具有高度的序列同源性的意思而言，衍生的酶在很大程度上是基於野生型酶的，但是該衍生的酶係已經以本領域已知或在此所述的一些方式發生突變的（修飾的）。

[0215] 同樣，應瞭解，除非另外表明，否則術語Cas和CRISPR酶和CRISPR蛋白和Cas蛋白通常是可以互換使用的並且在所有在此參考方面處以類推方式指進一步描述在本發明中的新型CRISPR效應蛋白，諸如藉由具體參考Cas9。如以上提到的，在此使用的許多殘基編號係指來自V型/VI型CRISPR座位的效應蛋白。然而，應瞭解本發明包括來自其他微生物物種的更多效應蛋白。在某些實施方式中，效應蛋白可以是組成型存在的或誘導型存在的或條件型存在的或被給予或遞送的。效應蛋白優化可以用於增強功能或用於開發新功能，其可產生嵌合效應蛋白。並且如在此

所述的效應蛋白可以被修飾成用作通用的核酸結合蛋白。

[0216] 典型地，在核酸靶向系統的情況下，核酸靶向複合物（包含雜交至靶序列並與一種或多種核酸靶向效應蛋白複合的指導RNA）的形成產生靶序列中或靶序列附近（例如，從靶序列開始的1、2、3、4、5、6、7、8、9、10、20、50、或更多個鹼基對內）的一條或兩條DNA或RNA股的切割。如在此所用的術語"與感興趣的靶座位相關聯的一種或多種序列"係指在靶序列的附近（例如從靶序列開始的1、2、3、4、5、6、7、8、9、10、20、50、或更多個鹼基對內，其中靶序列包含在感興趣的靶座位內）的序列。

[0217] 在本發明中，密碼子優化序列的一個實例係被優化為在真核生物，例如人類（即被優化為在人類中表現），或其他真核生物，如在此所討論的動物或哺乳動物中表現的序列；例如參見作為密碼子優化序列的一個實例的WO 2014/093622（PCT/US2013/074667）中的SaCas9人類密碼子優化序列（根據本領域知識和本揭露，特別是關於效應蛋白（例如Cpf1）的密碼子優化編碼的一個或多個核酸分子係在技術人員的知識範圍內的）。雖然這係較佳的，但是應瞭解，其他實例係可能的並且除人類之外的宿主物種的密碼子優化，或特定器官的密碼子優化係已知的。在一些實施方式中，編碼DNA/RNA靶向Cas蛋白的酶編碼序列被密碼子優化為在特定細胞諸如真核細胞中表現。真核細胞可以是特定生物體具有的或來源於該特定生物體的那些細胞，該特定生物體係諸如植物或哺乳動物，包括但不限於人類或非人類真核生物，或如在此討論的動物或哺乳動物，例如小鼠、大鼠、兔、狗、

家畜或非人類哺乳動物或靈長類動物。在一些實施方式中，對於人類或動物而言很可能使得他們（它們）受苦而沒有任何實質性醫學益處的用於修飾人類的種系遺傳一致性的方法和/或用於修飾動物的遺傳一致性的方法，以及還有作為這樣的方法的結果的動物，可以被排除在外。總的來說，密碼子優化係指在維持天然胺基酸序列的情況下藉由以下方式修飾核酸序列來增強在感興趣的宿主細胞中的表現的方法：藉由用該宿主細胞的基因中更頻繁使用或最頻繁使用的密碼子替換天然序列的至少一個密碼子（例如，約或超過約1、2、3、4、5、10、15、20、25、50或更多個密碼子）。不同的物種對於具有特定胺基酸的某些密碼子表現出特定偏倚性。密碼子偏倚性（生物體之間密碼子使用的差異）常常與信使RNA（mRNA）的翻譯效率相關，而該翻譯效率則被認為依賴於（除其他之外）被翻譯的密碼子的特性和特定轉移RNA（tRNA）分子的可獲得性。細胞中選擇的tRNA的超優勢度通常是肽合成中最頻繁使用的密碼子的反映。因此，基因可以被定制為基於密碼子優化在給定生物體中最佳基因表現。密碼子使用表係易於獲得的，例如在從www.kazusa.orjp/codon/獲得的“密碼子使用資料庫”中並且該等表可以藉由多種方式來調整適用。參見，中村，Y.（Nakamura, Y.）等人，“從國際DNA序列資料庫中製表的密碼子使用：2000年的狀態（Codon usage tabulated from the international DNA sequence databases: status for the year 2000）”核酸研究28:292 (2000)。用於密碼子優化特定序列以在特定宿主細胞中表現的電腦演算法也是可得的，諸如基因製造（Gene Forge）（賓夕法尼亞州雅各斯的Aptagen公司（Aptagen; Jacobus, PA））也是可得

的。在一些實施方式中，編碼DNA/RNA靶向Cas蛋白的序列中的一個或多個密碼子（例如1、2、3、4、5、10、15、20、25、50個、或更多個、或所有密碼子)對應於特定胺基酸的最頻繁使用的密碼子。對於酵母的密碼子使用，參考從http://www.yeastgenome.org/community/codon_usage.shtml獲得的線上酵母基因組資料庫或酵母的密碼子選擇（*Codon selection in yeast*），本特澤恩（Bennetzen）和哈爾（Hall），生物化學雜誌（J Biol Chem.），1982年3月25日；257(6):3026-31。對於植物（包括藻類）的密碼子使用，參考高等植物、綠藻和藍藻細菌的密碼子使用（*Codon usage in higher plants, green algae, and cyanobacteria*），坎貝爾（Campbell）和哥瑞（Gowri），植物生理學（Plant Physiol.），1990年1月；92(1): 1-11；以及植物基因的密碼子使用（*Codon usage in plant genes*），瑪瑞（Murray）等人，核酸研究，1989年1月25日；17(2):477-98；或不同植物和藻類譜系中的葉綠體基因和藍色小體基因的密碼子偏倚性的選擇（*Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages*），莫爾頓·BR（Morton BR），分子進化雜誌（J Mol Evol），1998年4月；46(4):449-59。

[0218] 在一些實施方式中，載體編碼包含一個或多個核酸定位序列（NLS）（諸如約或超過約1、2、3、4、5、6、7、8、9、10個或更多個NLS）的核酸靶向效應蛋白，諸如V型/VI型RNA靶向效應蛋白，具體地是Cpf1/C2c1/C2c2或其異種同源物或同源物。在一些實施方式中，RNA靶向效應蛋白包含處於或靠近胺基末端的約或超過約1、2、3、4、5、6、7、8、9、10個、或更多個NLS，

處於或靠近羧基末端的約或超過約1、2、3、4、5、6、7、8、9、10個、或更多個NLS，或該等的組合（例如在胺基末端處的零個或至少一個或多個NLS以及在羧基末端處的零個或至少一個或多個NLS）。當存在超過一個的NLS時，每一個可以被選擇為不依賴於其他NLS，使得單一NLS可以存在於超過一個的拷貝中和/或與一個或多個其他NLS相組合存在於一個或多個拷貝中。在一些實施方式中，當NLS的最近的胺基酸係在從N末端或C末端沿著該多肽鏈的約1、2、3、4、5、10、15、20、25、30、40、50個、或更多個胺基酸之內時，NLS可以被認為靠近該N末端或C末端。NLS的非限制性實例包括來源於以下項的NLS序列：SV40病毒大T抗原的NLS，其具有胺基酸序列PKKKRKV（SEQ ID NO: 2）；來自核質蛋白的NLS（例如，具有序列KRPAATKKAGQAKKKK（SEQ ID NO: 3）的核質蛋白二分NLS)；c-myc NLS，其具有胺基酸序列PAAKRVKLD（SEQ ID NO: 4）或RQRRNELKRSP（SEQ ID NO: 5）；hRNPA1 M9 NLS，其具有序列NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY（SEQ ID NO: 6）；來自輸入蛋白-α的IBB結構域的序列RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV（SEQ ID NO: 7）；肌瘤T蛋白的序列VSRKRPRP（SEQ ID NO: 8）和PPKKARED（SEQ ID NO: 9）；人類p53的序列PQPKKKPL（SEQ ID NO: 10）；小鼠c-abl IV的序列SALIKKKKKMAP（SEQ ID NO: 11）；流感病毒NS1的序列DRLRR（SEQ ID NO: 12）和PKQKKRK（SEQ ID NO: 13）；肝炎病毒δ抗原的序列RKLKKKIKKL（SEQ ID NO: 14）；小鼠Mx1蛋白的序列REKKKFLKRR（SEQ ID NO: 15）；

人類聚(ADP-核糖)聚合酶的序列KRKGDEVDGVDEVAKKKSKK（SEQ ID NO: 16）；以及類固醇激素受體（人類）糖皮質激素的序列RKCLQAGMNLEARKTKK（SEQ ID NO: 17）。總的來說，一個或多個NLS具有足以驅動DNA/RNA靶向Cas蛋白在真核細胞的核中以可檢測的量積累的強度。總的來說，核定位活性的強度可以來源於核酸靶向效應蛋白中的NLS數目、使用的一個或多個特定NLS、或該等因素的組合。可以藉由任何合適的技術進行核中積累的檢測。例如，可檢測標記可以融合至核酸靶向蛋白，使得細胞內的位置可以被視覺化，諸如與檢測核的位置的手段（例如，對核具有特異性的染料，諸如DAPI)相組合。還可以將細胞核從細胞中分離出來，然後可以藉由任何適合的用於檢測蛋白質的方法分析其內容物，諸如免疫組織化學、西方墨點或酶活性測定。還可以藉由以下方式間接地確定核中的積累：諸如藉由測定核酸靶向複合物形成的作用（例如，測定在靶序列處的DNA或RNA切割或突變、或測定由於DNA或RNA靶向複合物形成和/或DNA或RNA靶向Cas蛋白活性的影響而改變的基因表現活性），與沒有暴露於核酸靶向Cas蛋白或核酸靶向複合物、或暴露於缺乏一個或多個NLS的核酸靶向Cas蛋白的對照進行比較。在在此所述的Cpf1效應蛋白複合物和系統的較佳的實施方式中，密碼子優化的Cpf1效應蛋白包含附接至該蛋白質的C末端的NLS。在某些實施方式中，其他定位標籤可以融合至Cas蛋白，諸如但不限於將Cas定位至細胞中的特定位點，該等特定位點係諸如細胞器，諸如線粒體、質粒、葉綠體、囊泡、高爾基體（核的或細胞的）、細胞膜、核糖體、小核體（nucleoluse）、ER、細胞骨架、液泡、中心體、核小體、顆

粒、中心粒等。

[0219] 在一些實施方式中，驅動核酸靶向系統的一種或多種元件表現的一種或多種載體被引入到宿主細胞中，以使得該核酸靶向系統的該等元件的表現能引導核酸靶向複合物在一個或多個靶位點處形成。例如，核酸靶向效應酶和核酸靶向指導RNA可以各自可操作地連接至單獨載體上的單獨調節元件。核酸靶向系統的一種或多種RNA可以被遞送至轉基因核酸靶向效應蛋白動物或哺乳動物，例如組成型地或誘導型地或條件型地表現核酸靶向效應蛋白的動物或哺乳動物；或以其他方式表現核酸靶向效應蛋白或具有含有核酸靶向效應蛋白的細胞的動物或哺乳動物，諸如藉由在先向該等動物或哺乳動物給予編碼或體內表現核酸靶向效應蛋白的一種或多種載體的方式。可替代地，從相同或不同調節元件表現的該等元件的兩種或更多種可以組合在單一載體中，其中一種或多種另外的載體提供核酸靶向系統在第一載體中不包含的任何組分。組合於單一載體中的核酸靶向系統元件可以佈置為任何適合的取向，諸如一個元件位於相對於第二元件的5'（“上游”）或相對於該第二元件的3'（“下游”）。一個元件的編碼序列可以位於第二元件的編碼序列的相同股或相反股上，並且取向為相同或相反方向。在一些實施方式中，單一啟動子驅動編碼核酸靶向效應蛋白的轉錄物和嵌入一種或多種內含子序列之內（例如，各自在不同內含子中、兩個或更多個在至少一個內含子中，或所有在單一內含子中）的核酸靶向指導RNA的表現。在一些實施方式中，核酸靶向效應蛋白和核酸靶向指導RNA可以可操作地連接至同一個啟動子並且從該同一啟動子表現。用於表現核酸靶向系統的一

個或多個元件的遞送媒介物、載體、粒子、奈米粒子、配製物以及其組分如前述文獻諸如WO 2014/093622（PCT/US2013/074667）中所使用的。在一些實施方式中，載體包含一個或多個插入位點，諸如限制性內切核酸酶識別序列（也稱之為“選殖位點”）。在一些實施方式中，一個或多個插入位點（例如，約或超過約1、2、3、4、5、6、7、8、9、10個、或更多個插入位點）位於一種或多種載體的一個或多個序列元件的上游和/或下游。當使用多個不同的指導序列時，可以使用單一表現構建體來使核酸靶向活性靶向細胞內的多個不同的相應靶序列。例如，單一載體可以包含約或超過約1、2、3、4、5、6、7、8、9、10、15、20個、或更多個指導序列。在一些實施方式中，可以提供約或超過約1、2、3、4、5、6、7、8、9、10個、或更多個含有此指導序列的載體，並且視情況將其遞送至細胞中。在一些實施方式中，載體包含可操作地連接至編碼核酸靶向效應蛋白的酶編碼序列的調節元件。可以單獨地遞送核酸靶向效應蛋白或一個或多個核酸靶向指導RNA；並且有利的是該等中的至少一者經由粒子複合物遞送。核酸靶向效應蛋白mRNA可以在核酸靶向指導RNA在給出時間以待核酸靶向效應蛋白表現之前遞送。核酸靶向效應蛋白mRNA可以在給予核酸靶向指導RNA之前1-12小時（較佳的是約2-6小時）給予。可替代地，核酸靶向效應蛋白mRNA和核酸靶向指導RNA可以一起給予。有利地，指導RNA的第二加強劑量可以在初始給予核酸靶向效應蛋白mRNA + 指導RNA之後1-12小時（較佳的是約2-6小時）給予。為了實現最有效的基因組修飾水平，核酸靶向效應蛋白mRNA和/或指導RNA的附加給予可能是有用的。

[0220] 在一個方面中，本發明提供了用於使用核酸靶向系統的一個或多個元件的方法。本發明的核酸靶向複合物提供了一種用於修飾靶DNA或RNA（單股或雙股、直鏈或超螺旋的)的有效手段。本發明的核酸靶向複合物具有多種多樣的效用，包括修飾（例如，缺失、插入、易位、失活、啟動）許多細胞類型中的靶DNA或RNA。這樣，本發明的核酸靶向複合物在例如基因治療、藥物篩選、疾病診斷以及預後方面具有廣泛的應用。示例性核酸靶向複合物包含與雜交至感興趣的靶座位內的靶序列的指導RNA複合的DNA或RNA靶向效應蛋白。

[0221] 在一個實施方式中，本發明提供了一種切割靶RNA之方法。該方法可以包括使用結合靶RNA的核酸靶向複合物修飾靶RNA並且實施所述靶RNA的切割。在一實施方式中，本發明的核酸靶向複合物在被引入到細胞中時可以產生RNA序列的斷裂（例如單股或雙股斷裂）。例如，該方法可以用於切割細胞中的疾病相關RNA。例如，可以將包含有待整合的側接上游序列和下游序列的序列的外源性RNA模板引入到細胞中。上游序列和下游序列與RNA中整合位點的任一側享有序列相似性。當希望時，供體RNA可以是mRNA。外源性RNA模板包含有待整合的序列（例如，突變型RNA）。供整合的序列可以是對細胞而言內源或外源的序列。有待整合的序列的實例包括編碼蛋白質的RNA或非編碼RNA(例如，微小RNA）。因此，供整合的序列可以可操作地連接至一種或多種適當的控制序列。可替代地，有待整合的序列可以提供調節功能。外源性RNA模板中的上游序列和下游序列被選擇為促進感興趣的RNA序列與供體RNA之間的重組。上游序列係與供整合的靶向位

點的上游的RNA序列享有序列相似性的RNA序列。類似地，下游序列係與整合的靶向位點的下游的RNA序列享有序列相似性的RNA序列。外源性RNA模板中的上游序列和下游序列與靶向的RNA序列可以具有75%、80%、85%、90%、95%或100%序列一致性。較佳的是，外源性RNA模板中的上游序列和下游序列與靶向的RNA序列具有約95%、96%、97%、98%、99%或100%序列一致性。在一些方法中，外源性RNA模板中的上游序列和下游序列與靶向的RNA序列具有約99%或100%序列一致性。上游序列或下游序列可以包含從約20 bp至約2500 bp，例如約50、100、200、300、400、500、600、700、800、900、1000、1100、1200、1300、1400、1500、1600、1700、1800、1900、2000、2100、2200、2300、2400或2500 bp。在一些方法中，示例性上游序列或下游序列具有約200 bp至約2000 bp、約600 bp至約1000 bp，或更具體地約700 bp至約1000 bp。在一些方法中，外源性RNA模板可以進一步包含標記物。此標記物可以使得容易地篩選靶向的整合。適合的標記物的實例包括限制性位點、螢光蛋白或選擇標記物。可以使用重組技術構建本發明的外源性RNA模板（例如參見，薩姆布魯克（Sambrook）等人，2001和奧蘇貝爾（Ausubel）等人，1996）。在用於藉由整合外源性RNA模板來修飾靶RNA的方法中，藉由核酸靶向複合物將斷裂（例如在雙股或單股DNA或RNA中的雙股或單股斷裂）引入到DNA或RNA序列中，經由與外源性RNA模板的同源重組而修復該斷裂，這樣使得將該模板整合到RNA靶中。雙股斷裂的存在促進模板的整合。在其他實施方式中，本發明提供了一種修飾RNA在真核細胞中的表現的方法。該方法包括藉由使用結合DNA或

RNA（例如，mRNA或前mRNA）的核酸靶向複合物增加或減少靶多核苷酸的表現。在一些方法中，可以使靶RNA失活以實施細胞中的表現的修飾。例如，在RNA靶向複合物與細胞中的靶序列結合後，靶RNA失活，這樣使得該序列不被翻譯，該編碼蛋白不被產生，或者該序列不會像野生型序列一樣起作用。例如，可以使蛋白質或微小RNA編碼序列失活，這樣使得該蛋白質或微小RNA或前微小RNA轉錄物不被產生。RNA靶向複合物的靶RNA可以是對真核細胞而言內源或外源的任何RNA。例如，靶RNA可以是駐留在真核細胞核內的RNA。靶RNA可以是編碼基因產物（例如，蛋白質）的序列（mRNA或前mRNA）或非編碼序列（例如，ncRNA、lncRNA、tRNA或rRNA）。靶mRNA的實例包括與傳訊生物化學途徑相關聯的序列，例如傳訊生物化學途徑相關的RNA。靶RNA的實例包括疾病相關的RNA。“疾病相關”RNA係指與非疾病對照的組織或細胞相比，在來源於疾病影響的組織的細胞中產生異常水平或異常形式的翻譯產物的任何RNA。它可以是由以異常高的水平表現的基因轉錄的RNA；它可以是由以異常低的水平表現的基因轉錄的RNA，其中改變的表現與疾病的發生和/或進展相關。疾病相關RNA還是指由下述基因轉錄的RNA：該基因具有一個或多個突變或直接負責或與一個或多個負責疾病的病因學的基因連鎖不平衡的遺傳變異的基因。翻譯的產物可以是已知或未知的，並且可以是處於正常或異常水平。RNA靶向複合物的靶RNA可以是對真核細胞而言內源或外源的任何RNA。例如，靶RNA可以是駐留在真核細胞核內的RNA。靶RNA可以是編碼基因產物（例如，蛋白質）的序列（mRNA或前mRNA）或非編碼序列（例如，ncRNA、

lncRNA、tRNA或rRNA）。

[0222] 在一些實施方式中，該方法可以包括使得核酸靶向複合物結合靶DNA或RNA來實施所述靶DNA或RNA的切割，從而修飾該靶DNA或RNA，其中該核酸靶向複合物包含與雜交至所述靶DNA或RNA內的靶序列的指導RNA複合的核酸靶向效應蛋白。在一個方面中，本發明提供了一種修飾DNA或RNA在真核細胞中的表現之方法。在一些實施方式中，該方法包括使得核酸靶向複合物結合DNA或RNA，以使得所述結合導致所述DNA或RNA的表現增加或減少；其中該核酸靶向複合物包含與指導RNA複合的核酸靶向效應蛋白。類似的考慮因素和條件適用如上文針對修飾靶DNA或RNA的方法。實際上，該等取樣、培養和重新引入選擇跨本發明的多個方面而適用。在一個方面中，本發明提供了修飾真核細胞中的靶DNA或RNA的方法，該等方法可以是在體內、離體或在體外。在一些實施方式中，該方法包括從人類或非人類動物取樣細胞或細胞群體，並且修飾該細胞或該等細胞。培養可以發生在離體的任何階段。該細胞或該等細胞甚至可以被重新引入非人類動物或植物中。對於重新引入的細胞，特別較佳的是該等細胞係幹細胞。

[0223] 實際上，在本發明的任何方面中，核酸靶向複合物可以包含與雜交至靶序列的指導RNA複合的核酸靶向效應蛋白。

[0224] 本發明涉及用於控制涉及DNA或RNA序列靶向的基因表現的系統、方法以及組成物的工程化和優化，該等系統、方法以及組成物與核酸靶向系統及其組分相關。在有利的實施方式

中，效應酶係V型/VI型蛋白諸如Cpf1/C2c1/C2c2。本發明方法的一個優點係CRISPR系統最小化了或避免了脫靶結合及其產生的副作用。這係使用佈置為對靶DNA或RNA具有高度序列特異性的系統來實現的。

[0225] 關於核酸靶向複合物或系統，較佳的是，crRNA序列具有一個或多個莖環或髮夾並且具有30個或更多個核苷酸的長度，40或更多個核苷酸的長度，或50個或更多個核苷酸的長度；crRNA序列的長度介於10個至30個核苷酸之間，核酸靶向效應蛋白係V型/VI型Cas酶。在某些實施方式中，crRNA序列的長度介於42個與44個核苷酸之間，並且核酸靶向Cas蛋白係土拉熱弗朗西絲菌新殺手亞種*U112*的Cpf1。在某些實施方式中，crRNA包含具有19個核苷酸的同向重複序列和具有介於23個與25個之間核苷酸的間隔區序列、基本上由或由該同向重複序列和該間隔區序列組成，並且核酸靶向Cas蛋白係土拉熱弗朗西絲菌新殺手亞種*U112*的Cpf1。

[0226] 使用兩種不同的適配體(各自與不同的核酸靶向指導RNA締合）允許藉由不同的核酸靶向指導RNA來使用活化物-轉接蛋白融合物和阻遏物-轉接蛋白融合物，以啟動一種DNA或RNA的表現，同時阻遏另一種。它們可以與它們的不同指導RNA一起、或大體上一起以多重途徑給予。大量的這樣的修飾核酸靶向指導RNA（例如10或20或30個等）可以同時全部使用，而僅需要遞送一個（或至少最小數目的）效應蛋白分子，因為相對較小數目的效應蛋白分子可以與大量的修飾指導序列一起使用。轉接蛋白可以與一個或多個活化物或一個或多個阻遏物締合（較佳的是連接

或融合）。例如，轉接蛋白可以與第一活化物和第二活化物締合。第一活化物和第二活化物可以是相同的，但是它們較佳的是不同的活化物。可以使用三個或更多個或甚至四個或更多個活化物（或阻遏物），但是包裝尺寸可能限制大於5個不同功能結構域的數目。較佳的是使用接頭，藉由與轉接蛋白的直接融合來使用，其中兩個或更多個結構功能域與轉接蛋白締合。適合接頭可以包括GlySer接頭。

[0227] 還設想的是作為整體的核酸靶向效應蛋白指導RNA複合物可以與兩個或更多個結構功能域締合。例如，可以存在與核酸靶向效應蛋白締合的兩個或更多個功能結構域，或者可以存在與指導RNA締合（經由一種或多種適轉接蛋白）的兩個或更多個功能結構域，或者可以存在與核酸靶向效應蛋白締合的一個或多個功能結構域和與指導RNA締合（經由一種或多種適轉接蛋白）的一個或多個功能結構域。

[0228] 轉接蛋白與活化物或阻遏物之間的融合物可以包含接頭。例如，可以使用GlySer接頭GGGS（SEQ ID NO: 18）。根據需要，它們可以3個（$(GGGGS)_3$（SEQ ID NO: 19））或6個（SEQ ID NO: 20）、9個（SEQ ID NO: 21）或甚至12個（SEQ ID NO: 22）或更多個的重複單元來使用以提供適合的長度。指導RNA與功能結構域（活化物或阻遏物）之間、核酸靶向Cas蛋白（Cas）與功能結構域（活化物或阻遏物）之間可以使用接頭。接頭用於工程化適當的“機械柔性”度。

[0229] 本發明包括包含核酸靶向效應蛋白和指導RNA的核

酸靶向複合物，其中核酸靶向效應蛋白包含至少一個突變，使得核酸靶向效應蛋白具有不超過不具有該至少一個突變的核酸靶向效應蛋白的活性的5%的活性，和視情況至少一個或多個核定位序列；指導RNA包含能夠與細胞中的感興趣的RNA中的靶序列雜交的指導序列；並且其中：核酸靶向效應蛋白與兩個或更多個功能結構域締合；或者指導RNA的至少一個環係藉由插入結合一種或多種轉接蛋白的一種或多種不同的RNA序列來修飾的，並且其中該轉接蛋白與兩個或更多個功能結構域締合；或者核酸靶向Cas蛋白與一個或多個功能結構域締合並且指導RNA的至少一個環係藉由插入結合一種或多種轉接蛋白的一種或多種不同的RNA序列來修飾的，並且其中該轉接蛋白與一個或多個功能結構域締合。

[0230] 在一個方面中，本發明提供了一種產生包含突變型疾病相關基因的模型真核細胞之方法。在一些實施方式中，疾病相關基因係與患病或發展病的風險的增加相關聯的任何基因。在一些實施方式中，該方法包括 (a) 將一種或多種載體引入到真核細胞中，其中一種或多種載體驅動以下項中的一個或多個的表現：Cpf1酶和包含連接至同向重複序列的指導序列的受保護的指導RNA；並且 (b) 使得CRISPR複合物結合靶多核苷酸以實施所述疾病相關基因內的靶多核苷酸的切割，其中CRISPR複合物包含與包含雜交至靶多核苷酸內的靶序列的序列的指導RNA複合的Cpf1酶，從而產生包含突變型疾病相關基因的模型真核細胞。在一些實施方式中，所述切割包括藉由所述Cpf1酶切割靶序列的位置處的一條或兩條股。在一些實施方式中，所述切割使得靶基因的轉錄減少。在一些實施方式中，該方法進一步包括使用外源性模板

多核苷酸藉由基於非同源末端連接（NHEJ）的基因插入機制修復所述切割的靶多核苷酸，其中所述修復產生包括所述靶多核苷酸的一個或多個核苷酸的插入、缺失或取代的突變。在一些實施方式中，所述突變使得來自包含靶序列的基因的蛋白質表現發生一個或多個胺基酸的變化。

[0231] 在一方面中，本發明提供了如在此所討論之方法，其中該宿主係真核細胞。在一方面中，本發明提供了如在此所討論之方法，其中該宿主係哺乳動物細胞。在一方面中，本發明提供了如在此所討論之方法，其中該宿主係非人類真核細胞。在一方面中，本發明提供了如在此所討論之方法，其中該非人類真核細胞係非人類哺乳動物細胞。在一方面中，本發明提供了如在此所討論之方法，其中非人類哺乳動物細胞可以是包括但不限於，靈長類動物、牛、羊、豬類、犬、齧齒動物、兔科諸如猴、母牛、綿羊、豬、狗、兔、大鼠或小鼠的細胞。在一方面中，本發明提供了如在此所討論之方法，該細胞可以是非哺乳動物真核細胞諸如家禽鳥類（例如雞）、脊椎動物魚（例如鮭魚）或甲殼類動物（例如牡蠣、蛤、龍蝦、蝦）的細胞。在一方面中，本發明提供了如在此所討論之方法，該非人類真核細胞係植物細胞。植物細胞可以是單子葉植物或雙子葉植物具有的細胞或栽培植物或糧食植物諸如木薯、玉米、高粱、大豆、小麥、燕麥或稻具有的細胞。植物細胞還可以是藻類、樹或生產植物、果實或蔬菜（例如，樹類諸如柑橘樹，例如桔子樹、葡萄柚樹或檸檬樹；桃樹或油桃樹；蘋果樹或梨樹；堅果樹諸如杏樹或核桃樹或阿月渾子樹；茄屬植物；芸苔屬植物；萵苣屬植物；菠菜屬植物；辣椒屬植物；棉花、煙草、

蘆筍、胡蘿蔔、甘藍、青花菜、花椰菜、番茄、茄子、胡椒、萵苣、菠菜、草莓、藍莓、覆盆子、黑莓、葡萄、咖啡、可可等）具有的細胞。

[0232] 在一個方面中，本發明提供了一種用於開發調控與疾病相關基因相關聯的細胞傳訊事件的生物活性劑之方法。在一些實施方式中，疾病相關基因係與患病或發展病的風險的增加相關聯的任何基因。在一些實施方式中，該方法包括 (a) 使測試化合物與任一種以上所述實施方式的模型細胞接觸；並且 (b) 檢測指示與所述疾病相關基因中的所述突變相關聯的細胞傳訊事件的減少或增加的讀出變化，從而開發調控與所述疾病相關基因相關聯的所述細胞傳訊事件的所述生物活性劑。

[0233] 在一個方面中，本發明提供了一種藉由在一個或多個細胞的基因中引入一個或多個突變來選擇一個或多個細胞之方法，該方法包括：將一種或多種載體引入到一個或多個細胞中，其中一種或多種載體驅動以下項中的一個或多個的表現：Cpf1、連接至同向重複序列的指導序列，以及編輯模板；其中編輯模板包含廢除Cpf1切割的一個或多個突變；使得編輯模板與有待選擇的一個或多個細胞中的靶多核苷酸同源重組；使得Cpf1 CRISPR-Cas複合物結合靶多核苷酸以實施所述基因內的靶多核苷酸的切割，其中Cpf1 CRISPR-Cas複合物包含與 (1) 雜交至靶多核苷酸內的靶序列的指導序列，和 (2) 同向重複序列複合的Cpf1，其中Cpf1 CRISPR-Cas複合物與靶多核苷酸的結合誘導細胞死亡，從而使得其中已引入一個或多個突變的一個或多個細胞被選擇；此Cpf1包

括本發明的拆分的Cpf1。在本發明的另一個較佳的實施方式中，有待選擇的細胞可以是真核細胞。本發明的方面允許在不需要選擇標記物或可能包括反選擇系統的兩步法的情況下選擇特異性細胞。在特定實施方式中，模型真核細胞包含在模型真核生物體之內。

[0234] 在一個方面中，本發明提供了一種包含同向重複序列的下游的指導序列的重組多核苷酸，其中在表現時，指導序列引導Cpf1 CRISPR-Cas複合物與真核細胞中存在的相應靶序列的序列特異性結合。在一些實施方式中，靶序列係真核細胞中存在的病毒序列。在一些實施方式中，靶序列係原癌基因或癌基因。

[0235] 在一個方面中，本發明提供了一種包含以下項的載體系統或真核宿主細胞：(a) 可操作地連接至同向重複序列的第一調節元件和用於將一種或多種指導序列（包括如在此所述的任一種修飾指導序列）插入DR序列的下游的一個或多個插入位點，其中在表現時，指導序列引導Cpf1 CRISPR-Cas複合物與真核細胞中的靶序列的序列特異性結合，其中Cpf1 CRISPR-Cas複合物包含與雜交至靶序列的指導序列（和視情況DR序列）複合的Cpf1（包括如在此所述的任一種修飾酶）；和/或 (b) 可操作地連接至編碼包含核定位序列和/或NES的所述Cpf1酶的酶編碼序列的第二調節元件。在一些實施方式中，宿主細胞包含組分 (a) 和 (b)。在一些實施方式中，組分 (a)、組分 (b) 或組分 (a) 和 (b) 被穩定地整合到宿主真核細胞的基因組中。在一些實施方式中，組分 (a) 進

一步包含可操作地連接至第一調節元件的兩種或更多種指導序列，其中當表現時，兩種或更多種指導序列中的每種引導Cpf1 CRISPR-Cas複合物與真核細胞中的不同靶序列的序列特異性結合。在一些實施方式中，CRISPR酶包含具有足以驅動所述CRISPR酶在真核細胞的核中和/或之外以可檢測的量積累的強度的一個或多個核定位序列和/或核輸出序列或NES。在一些實施方式中，Cpf1酶來源於土拉熱弗朗西絲菌1、土拉熱弗朗西絲菌新殺手亞種、易北普雷沃菌、毛螺旋菌科細菌MC2017 1、解肽丁酸弧菌、佩萊格裡尼菌科細菌GW2011_GWA2_33_10、儉菌總門細菌GW2011_GWC2_44_17、密斯氏菌屬某種SCADC、胺基酸球菌屬某種BV3L6、毛螺旋菌MA2020、候選白蟻甲烷枝原體、挑剔真細菌、牛莫拉氏菌237、稻田鉤端螺旋體、毛螺旋菌科細菌ND2006、狗口腔卟啉單胞菌3、解糖胨普雷沃菌或獼猴卟啉單胞菌的Cpf1，包括如在此所述的任一種修飾酶，並且可以進一步包含Cpf1的改變或突變，並且可以是嵌合Cpf1。在一些實施方式中，CRISPR酶被密碼子優化為在真核細胞中表現。在一些實施方式中，CRISPR酶引導靶序列位置處的一條或兩條股的切割。在一較佳的實施方式中，股斷裂係交錯切割的，產生了5'突出端。在一些實施方式中，Cpf1缺乏DNA股切割活性（例如，與野生型酶或不具有降低核酸酶活性的突變或改變的酶相比的不超過5%的核酸酶活性）。在一些實施方式中，第一調節元件係聚合酶III啟動子。在一些實施方式中，第二調節元件係聚合酶II啟動子。在一些實施方式中，同向重複序列具有16個核苷酸的最小長度並且具有單一莖環。在另外的實施方式中，同向重複序列具有長於16個核苷酸，

較佳的是超過17個核苷酸的長度，並且具有超過一個的莖環或優化的二級結構。在一些實施方式中，指導序列的長度係至少16、17、18、19、20、25個核苷酸，或介於16個-30個、或介於16個-25個、或介於16個-20個核苷酸之間。

[0236] 在一個方面中，本發明提供了一種包含在此所述的一種或多種組分的套組（kit）。在一些實施方式中，套組包括如在此所述的載體系統或宿主細胞和用於使用套組的說明書。

修飾的 Cpf1 酶

[0237] Cpf1核酸酶初級結構的計算分析揭示了三個不同的區（圖1）。第一係C末端RuvC樣結構域，其係僅功能表征的結構域。第二係N末端α-螺旋區並且第三係位於RuvC樣結構域與α-螺旋區之間的混合的α區和β區。

[0238] 預測非結構化區的若干小片段在Cpf1初始結構之內。對於小的蛋白質序列的拆分和插入而言，不同的Cpf1異種同源物內的暴露於溶劑且不保守的非結構化區係較佳的側面（圖2和3）。另外，該等側面可以用於在Cpf1異種同源物之間產生嵌合蛋白。

[0239] 基於以上資訊，可以產生突變體，該等突變體使得酶失活或將雙股核酸酶修飾為具有切口酶活性。在替代實施方式中，此資訊用於開發具有減小的脫靶效應的酶（在此其他地方所述的）。

[0240] 在某些以上所述的Cpf1酶中，酶係藉由一個或多個殘基的突變來修飾的，該等殘基包括但不限於根據FnCpf1蛋白或任何相應的異種同源物的位置D917、E1006、E1028、D1227、D1255A、

N1257。在一個方面中，本發明提供了一種在此所討論的組成物，其中Cpf1酶係失活的酶，該酶包含選自下組的一個或多個突變，該組由下項組成：根據FnCpf1蛋白的D917A、E1006A、E1028A、D1227A、D1255A、N1257A、D917A、E1006A、E1028A、D1227A、D1255A和N1257A，或Cpf1異種同源物中的相應位置。在一個方面中，本發明提供了一種在此所討論的組成物，其中CRISPR酶包含根據FnCpf1蛋白的D917或E1006和D917或D917和D1255，或Cpf1異種同源物中的相應位置。

[0241] 在某些以上所述的Cpf1酶中，酶係藉由一個或多個殘基（在RuvC結構域中）的突變來修飾的，該等殘基包括但不限於參照AsCpf1（胺基酸球菌屬某種BV3L6）的胺基酸位置編碼的位置R909、R912、R930、R947、K949、R951、R955、K965、K968、K1000、K1002、R1003、K1009、K1017、K1022、K1029、K1035、K1054、K1072、K1086、R1094、K1095、K1109、K1118、K1142、K1150、K1158、K1159、R1220、R1226、R1242和/或R1252。

[0242] 在某些以上所述的非天然存在的CRISPR酶中，酶係藉由一個或多個殘基（在RAD50結構域中）的突變來修飾的，該等殘基包括但不限於參照AsCpf1（胺基酸球菌屬某種BV3L6）的胺基酸位置編碼的位置K324、K335、K337、R331、K369、K370、R386、R392、R393、K400、K404、K406、K408、K414、K429、K436、K438、K459、K460、K464、R670、K675、R681、K686、K689、R699、K705、R725、K729、K739、K748和/或K752。

[0243] 在某些Cpf1酶中，酶係藉由一個或多個殘基的突變來

修飾的，該等殘基包括但不限於參照AsCpf1（胺基酸球菌屬某種BV3L6)的胺基酸位置編碼的位置R912、T923、R947、K949、R951、R955、K965、K968、K1000、R1003、K1009、K1017、K1022、K1029、K1072、K1086、R1103、R1226和/或R1252。

[0244] 在某些實施方式中，Cpf1酶係藉由一個或多個殘基的突變來修飾的，該等殘基包括但不限於參照LbCpf1（毛螺旋菌科細菌ND2006)的胺基酸位置編碼的位置R833、R836、K847、K879、K881、R883、R887、K897、K900、K932、R935、K940、K948、K953、K960、K984、K1003、K1017、R1033、R1138、R1165和/或R1252。

[0245] 在某些實施方式中，Cpf1酶係藉由一個或多個殘基的突變來修飾的，該等殘基包括但不限於參照AsCpf1（胺基酸球菌屬某種BV3L6)的胺基酸位置編碼的位置K15、R18、K26、Q34、R43、K48、K51、R56、R84、K85、K87、N93、R103、N104、T118、K123、K134、R176、K177、R192、K200、K226、K273、K275、T291、R301、K307、K369、S404、V409、K414、K436、K438、K468、D482、K516、R518、K524、K530、K532、K548、K559、K570、R574、K592、D596、K603、K607、K613、C647、R681、K686、H720、K739、K748、K757、T766、K780、R790、P791、K796、K809、K815、T816、K860、R862、R863、K868、K897、R909、R912、T923、R947、K949、R951、R955、K965、

K968、K1000、R1003、K1009、K1017、K1022、K1029、A1053、K1072、K1086、F1103、S1209、R1226、R1252、K1273、K1282和/或K1288。

[0246] 在某些實施方式中，酶係藉由一個或多個殘基的突變來修飾的，該等殘基包括但不限於參照FnCpf1（新殺手法蘭西斯菌（*Francisella novicida*）U112）的胺基酸位置編碼的位置K15、R18、K26、R34、R43、K48、K51、K56、K87、K88、D90、K96、K106、K107、K120、Q125、K143、R186、K187、R202、K210、K235、K296、K298、K314、K320、K326、K397、K444、K449、E454、A483、E491、K527、K541、K581、R583、K589、K595、K597、K613、K624、K635、K639、K656、K660、K667、K671、K677、K719、K725、K730、K763、K782、K791、R800、K809、K823、R833、K834、K839、K852、K858、K859、K869、K871、R872、K877、K905、R918、R921、K932、I960、K962、R964、R968、K978、K981、K1013、R1016、K1021、K1029、K1034、K1041、K1065、K1084和/或K1098。

[0247] 在某些實施方式中，酶係藉由一個或多個殘基的突變來修飾的，該等殘基包括但不限於參照LbCpf1（毛螺旋菌科細菌ND2006）的胺基酸位置編碼的位置K15、R18、K26、K34、R43、K48、K51、R56、K83、K84、R86、K92、R102、K103、K116、K121、R158、E159、R174、R182、K206、K251、K253、K269、K271、K278、P342、K380、R385、K390、K415、K421、K457、K471、A506、R508、K514、K520、K522、K538、Y548、K560、

K564、K580、K584、K591、K595、K601、K634、K640、R645、K679、K689、K707、T716、K725、R737、R747、R748、K753、K768、K774、K775、K785、K787、R788、Q793、K821、R833、R836、K847、K879、K881、R883、R887、K897、K900、K932、R935、K940、K948、K953、K960、K984、K1003、K1017、R1033、K1121、R1138、R1165、K1190、K1199和/或K1208。

[0248] 在某些實施方式中，酶係藉由一個或多個殘基的突變來修飾的，該等殘基包括但不限於參照MbCpf1（牛莫拉氏菌237）的胺基酸位置編碼的位置K14、R17、R25、K33、M42、Q47、K50、D55、K85、N86、K88、K94、R104、K105、K118、K123、K131、R174、K175、R190、R198、I221、K267、Q269、K285、K291、K297、K357、K403、K409、K414、K448、K460、K501、K515、K550、R552、K558、K564、K566、K582、K593、K604、K608、K623、K627、K633、K637、E643、K780、Y787、K792、K830、Q846、K858、K867、K876、K890、R900、K901、M906、K921、K927、K928、K937、K939、R940、K945、Q975、R987、R990、K1001、R1034、I1036、R1038、R1042、K1052、K1055、K1087、R1090、K1095、N1103、K1108、K1115、K1139、K1158、R1172、K1188、K1276、R1293、A1319、K1340、K1349和/或K1356。

去活的/失活的Cpf1蛋白

[0249] 在Cpf1蛋白具有核酸酶活性的情況下，Cpf1蛋白可以被修飾成具有減弱的核酸酶活性，例如，與野生型酶相比，具有至少70%、至少80%、至少90%、至少95%、至少97%或100%的核

酸酶失活；或者換句話說，Cpf1酶有利地具有非突變型或野生型Cpf1酶或CRISPR酶的核酸酶活性的約0%，或不超過非突變型或野生型Cpf1酶的核酸酶活性的約3%或約5%或約10%，該等酶例如是屬於非突變型或野生型新殺手法蘭西斯菌*U112*（FnCpf1）、胺基酸球菌屬某種*BV3L6*（AsCpf1）、毛螺旋菌科細菌*ND2006*（LbCpf1）或牛莫拉氏菌*237*（MbCpf1 Cpf1酶或CRISPR酶）的。有可能藉由將突變引入到Cpf1和其異種同源物的核酸酶結構域中實現此舉。

[0250] 更具體地，失活的 Cpf1 酶包括在 AsCpf1 的胺基酸位置 As908、As993、As1263 或 Cpf1 異種同源物中的相應位置突變的酶。另外，失活的 Cpf1 酶包括在 LbCpf1 的胺基酸位置 Lb832、925、947 或 1180 或 Cpf1 異種同源物中的相應位置突變的酶。更具體地，失活的 Cpf1 酶包括包含 AsCpf1 的突變 AsD908A、AsE993A、AsD1263A 或 Cpf1 異種同源物中的相應突變中的一個或多個的酶。另外，失活的 Cpf1 酶包括包含 LbCpf1 的突變 LbD832A、E925A、D947A 或 D1180A 或 Cpf1 異種同源物中的相應突變中的一個或多個的酶。

[0251] 失活的Cpf1 CRISPR酶可以具有締合的（例如經由融合蛋白）一個或多個功能結構域，包括例如來自包括下項，或基本上由或由下項組成的組的一個或多個結構域：修飾烷酶活性、去甲酶活性、轉錄啟動活性、轉錄阻遏活性、轉錄釋放因子活性、組蛋白修飾活性、RNA切割活性、DNA切割活性、核酸結合活性以及分子開關（例如光誘導型的）。較佳的結構域係Fok1、VP64、P65、HSF1、MyoD1。在提供的是Fok1的情況下，有利的是提供多

個Fok1功能結構域以實現功能二聚體並且gRNA被設計為提供適當間隔以用於功能性的使用（Fok1），如蔡（Tsai）等人 自然生物技術，第32卷，第2期，2014年6月）中具體描述的。轉接蛋白可以利用已知的接頭來附接此類功能結構域。在一些情況下，有利的是另外提供至少一個NLS。在一些情況下，將NLS定位在N末端處係有利的。當包含超過一個的功能結構域時，功能結構域可以是相同或不同的。

[0252] 總的來說，一個或多個功能結構域在失活的Cpf1酶上的定位係允許功能結構域的正確空間取向，從而以屬性化的功能效應影響靶的定位。例如，如果功能結構域係轉錄活化物（例如，VP64或p65），則轉錄活化物被定位成允許其影響靶的轉錄的空間取向。同樣地，轉錄阻遏物將有利地定位成影響靶的轉錄，並且核酸酶（例如Fok1）將有利地定位成切割或部分切割靶。此可以包括除CRISPR酶的N末端/C末端之外的位置。

去穩定化的Cpf1

[0253] 在某些實施方式中，如在此所述的根據本發明的效應蛋白（CRISPR酶；Cpf1）與去穩定化的結構域（DD）締合或融合。在一些實施方式中，DD係ER50。在一些實施方式中，此DD的相應的穩定化配位基係4HT。這樣，在一些實施方式中，至少一個DD中的一個係ER50並且因此穩定化配位基係4HT或CMP8。在一些實施方式中，DD係DHFR50。在一些實施方式中，此DD的相應的穩定化配位基係TMP。這樣，在一些實施方式中，至少一個DD中的一個係DHFR50並且因此穩定化配位基係TMP。在一些實施方式中，

DD係ER50。在一些實施方式中，此DD的相應的穩定化配位基係CMP8。因此，在ER50系統中CMP8可以是4HT的替代性穩定化配位基。雖然有可能CMP8和4HT可以/應該使用在競爭事件中，但是一些細胞類型可以對這兩種配位基中的一個或另一個更易感，並且根據本揭露和本領域的知識，技術人員可以使用CMP8和/或4HT。

[0254] 在一些實施方式中，在一個或兩個DD融合至CRISPR酶的C末端的情況下，一個或兩個DD可以融合至CRISPR酶的N末端。在一些實施方式中，至少兩個DD與CRISPR酶締合並且該等DD係相同的DD，即該等DD係同源的。因此，兩個（或兩個或更多個）DD可以是ER50 DD。這在一些實施方式中是較佳的。可替代地，兩個（或兩個或更多個）DD可以是DHFR50 DD。這在一些實施方式中也是較佳的。在一些實施方式中，至少兩個DD與CRISPR酶締合並且該等DD係不同的DD，即該等DD係異源的。因此，一個DD可以是ER50，而一個或多個DD或任何其他DD可以是DHFR50。具有為異源的兩個或更多個DD可能是有利的，因為其將提供更大的降解控制水平。在N末端或C末端處超過一個的DD的銜接融合可以增強降解；並且此銜接融合可以是例如ER50-ER50-C2c2或DHFR-DHFR-Cpf1。設想的是高的降解水平將在不存在任一穩定化配位基的情況下發生，中等的降解水平將在不存在一種穩定化配位基並且存在其他（或另一）穩定化配位基的情況下發生，而低的降解水平將在存在兩種（兩種或更多種）穩定化配位基的情況下發生。控制還可以藉由具有N末端的ER50 DD和C末端的DHFR50 DD來賦予。

[0255] 在一些實施方式中，CRISPR酶與DD的融合物包括在DD與CRISPR酶之間的接頭。在一些實施方式中，接頭為GlySer接頭。在一些實施方式中，DD-CRISPR酶進一步包含至少一個核輸出信號（NES）。在一些實施方式中，DD-CRISPR酶包含兩個或更多個NES。在一些實施方式中，DD-CRISPR酶包含至少一個核定位信號（NLS）。這可以與NES相附加。在一些實施方式中，CRISPR酶包含下項或基本上由或由下項組成：作為CRISPR酶與DD之間的接頭或作為該接頭的一部分的定位（核輸入或輸出）信號。HA或Flag標籤作為接頭也在本發明的範圍之內。申請人使用NLS和/或NES作為接頭並且還使用如最高至(GGGGS)3的GS一樣短的甘胺酸絲胺酸接頭。

[0256] 去穩定化結構域具有向寬範圍的蛋白質賦予不穩定性的一般效用；例如參見宮崎（Miyazaki），美國化學學會雜誌（J Am Chem Soc.），2012年3月7日；134(9): 3942–3945，該文獻藉由引用結合在此。CMP8或4-羥基他莫昔芬（4-hydroxytamoxifen）可以是去穩定化的結構域。更一般地說，哺乳動物DHFR的溫度敏感突變體（DHFRts），N端法則的去穩定化殘基被發現在許可溫度下是穩定的，但是在37°C下是不穩定的。向表現DHFRts的細胞添加甲胺蝶呤即哺乳動物DHFR的高親和力配位基部分地抑制了蛋白質的降解。這重要地證明了小分子配位基可以穩定化細胞中以其他方式被靶向降解的蛋白質。雷帕黴素（rapamycin）衍生物用於穩定化mTOR的FRB結構域（FRB*）的不穩定突變體並且恢復融合激酶GSK-3β.6,7的功能。此系統證明了配位基依賴性的穩定性代表了用於調節複合物生物環境中的特異性蛋白的功能的有吸引力

的策略。用於控制蛋白質活性的系統可以涉及當藉由雷帕黴素誘導的FK506結合蛋白和FKBP12的二聚化發生泛素互補時DD變成功能性的。人類FKBP12或ecDHFR蛋白的突變體可以被工程化為分別在不存在其高親和力配位基Shield-1或甲氧苄啶（TMP）的情況下是代謝不穩定的。該等突變體係可用於本發明的實踐中的一些可能去穩定化的結構域（DD）並且與CRISPR酶形成融合物的DD的不穩定性使得蛋白酶體對整個融合蛋白的CRISPR蛋白進行降解。Shield-1和TMP結合DD並且以劑量依賴性的方式穩定化DD。雌激素受體配位基結合結構域（ERLBD，ERS1的殘基305-549）也可以被工程化為去穩定化的結構域。因為雌激素受體傳訊途徑涉及多種疾病諸如乳腺癌，該途徑已被廣泛研究並且已經開發出雌激素受體的許多激動劑和拮抗劑。因此，相容的ERLBD和藥物對是已知的。存在結合ERLBD的突變體但不結合其野生型形式的配位基。藉由使用該等編碼三個突變（L384M、M421G、G521R）12的突變體結構域中的一個，有可能使用不擾亂內源性雌激素敏感網路的配位基來調節ERLBD源的DD的穩定性。可以引入另外的突變（Y537S）以進一步去穩定化ERLBD並且將其構造為潛在的DD候選物。四突變體係有利的DD改進。突變體ERLBD可以被融合至CRISPR酶並且其穩定性可以使用配位基來調控或擾亂，由此CRISPR酶具有DD。另一種DD可以是基於突變型FKBP蛋白、由Shield1配位基穩定化的12-kDa（107個胺基酸）的標籤；例如參見，自然方法（Nature Methods） 5, (2008)。例如，DD可以是修飾的FK506結合蛋白12（FKBP12），其結合合成的、生物惰性的小分子Shield-1並且由該Shield-1可逆地穩定化；例如參見巴納斯特·LA

（Banaszynski LA）、陳·LC（Chen LC）、梅娜德-史密斯·LA（Maynard-Smith LA）、黃·AG（Ooi AG）、萬德萊斯·TJ（Wandless TJ），一種使用合成的小分子調節活細胞中的蛋白質功能的快速可逆並可調之方法（A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules），細胞，2006;126:995–1004；巴納斯特·LA、塞爾米厄·MA（Sellmyer MA）、康塔格·CH（Contag CH）、萬德萊斯·TJ，索恩·SH（Thorne SH），活小鼠中的蛋白質穩定性和功能的化學控制（Chemical control of protein stability and function in living mice），自然醫學（Nat Med.） 2008;14:1123-1127；梅娜德-史密斯·LA，陳·LC，巴納斯特·LA，黃·AG，萬德萊斯·TJ，一種用於使用生物沈默的小分子工程化條件蛋白穩定性之定向法（A directed approach for engineering conditional protein stability using biologically silent small molecules），生物化學雜誌（The Journal of biological chemistry），2007;282:24866-24872；以及羅德里格斯（Rodriguez），化學生物學（Chem Biol.）2012年3月23日；19(3): 391–398—所有文獻藉由引用結合在此並且在選擇與本發明的實踐中的CRISPR酶締合的DD中可以應用在本發明的實踐中。如可以看出的，本領域知識包括許多DD，並且DD可以有利地藉由接頭與CRISPR酶締合，例如融合，由此DD在存在配位基的情況下可以是穩定化的並且當不存在該配位基時，DD可以是去穩定化的，由此CRISPR酶被完全去穩定化，或者DD在不存在配位基的情況下可以是穩定化的並且當存在配位基時，DD可以變成去穩定化的；DD允許CRISPR酶和因此CRISPR-Cas複合物或系統被調節或控制-可以說開啟或

關閉，從而提供用於例如在體內或體外環境中調節或控制系統的手段。例如，當感興趣的蛋白質與DD標籤作為融合物一起表現時，該蛋白質被去穩定化並且在細胞中例如被蛋白酶體快速降解。因此，不存在穩定化的配位基使得DD締合的Cas被降解。當新的DD被融合至感興趣的蛋白質時，該DD的不穩定性被賦予至感興趣的蛋白質，從而使得整個融合蛋白快速降解。Cas的峰值活性有時對降低脫靶效應係有益的。因此，高活性的突釋（bursts）係較佳的。本發明能夠提供此類峰值。在某種意義上，該系統係誘導型的。在一些其他意義上，在不存在穩定化配位基情況下系統被阻遏並且在存在穩定化配位基的情況下系統被去阻遏。

降低脫靶效應的酶突變

[0257] 在一個方面中，本發明提供了一種非天然發生的或工程化的CRISPR酶，較佳的是第2類CRISPR酶，較佳的是如在此所述的型V或VI CRISPR酶，諸如較佳的是，但不限於如在此其他地方所述的具有使得脫靶效應降低的一個或多個突變的Cpf1，即用於諸如當與指導RNA複合時對靶座位實施修飾但降低或消除朝向脫靶的活性的改進CRISPR酶，以及用於諸如當與指導RNA複合時增加CRISPR酶活性的改進CRISPR酶。應該理解的是，如下文所述的突變型酶可以用於在此其他地方所述的根據本發明的任一方法中。如在此其他地方所述的方法、產物、組成物和用途中的任一種係同樣適用於如下文進一步詳述的突變型CRISPR酶。應該理解的是，在如在此所述的方面和實施方式中，當提及或解讀作為CRISPR酶的Cpf1時，功能CRISPR-Cas系統的重構較佳的是不需要

或不依賴於tracr序列並且/或者同向重複序列係指導（靶或間隔區）序列的5'（上游）。

[0258] 作為進一步指導，提供了以下的特定方面和實施方式。

[0259] 發明者已經出人意料地確定，可以對CRISPR酶進行修飾，該等修飾使得相比於未修飾CRISPR酶脫靶活性降低並且/或者相比於未修飾CRISPR酶靶活性增加。因此，在本發明的某些方面中，在此提供了可以在寬範圍的基因修飾應用中具有效用的改進CRISPR酶。在此還提供了CRISPR複合物、組成物和系統以及方法和用途，所有包含了在此揭露的修飾CRISPR酶。

[0260] 在本揭露中，術語“Cas”可以意指“Cpf1”或CRISPR酶。在本發明的此方面情況下，Cpf1或CRISPR酶係突變的或修飾的，“由此相比於未修飾酶CRISPR複合物中的酶具有降低的修飾一個或多個脫靶座位的能力”（或類似表現）；並且，當閱讀本申請時，術語“Cpf1”或“Cas”或“CRISPR酶”等意指包括根據本發明的突變或修飾的Cpf1或Cas或CRISPR酶，即“由此相比於未修飾酶CRISPR複合物中的酶具有降低的修飾一個或多個脫靶座位的能力”（或類似表現）。

[0261] 在一個方面中，提供了一種如在此所限定的工程化的Cpf1蛋白，諸如Cpf1，其中該蛋白與包含RNA的核酸分子複合以形成CRISPR複合物，其中當處於CRISPR複合物中時，核酸分子靶向一個或多個靶多核苷酸座位，與未修飾Cpf1蛋白相比，該蛋白包含至少一種修飾，並且其中相比於包含未修飾Cpf1蛋白的複合物，包含修飾蛋白的CRISPR複合物具有改變的活性。應該理解的

是，當在此提及CRISPR“蛋白”時，Cpf1蛋白較佳的是修飾的CRISPR酶（例如具有增加或降低（或沒有）酶活性），諸如非限制性地包括Cpf1。術語“CRISPR蛋白”可以與“CRISPR酶”可互換地使用，而不考慮相比於野生型CRISPR蛋白，該CRISPR蛋白是否被改變，諸如增加或降低（或沒有）酶活性。

[0262] 在一個方面中，工程化的CRISPR蛋白的活性改變包括關於包含RNA的核酸分子或靶多核苷酸座位的結合特性改變，關於包含RNA的核酸分子或靶多核苷酸座位的結合動力學改變，或關於包含RNA的核酸分子或靶多核苷酸座位的相比於脫靶多核苷酸座位的結合特異性改變。

[0263] 在一些實施方式中，未修飾Cas具有DNA切割活性，諸如Cpf1。在一些實施方式中，Cas引導靶序列位置處的，諸如靶序列內和/或靶序列補體內的一條或兩條股的切割。在一些實施方式中，Cas引導從靶序列的第一個或最後一個核苷酸開始的約1、2、3、4、5、6、7、8、9、10、15、20、25、50、100、200、500或更多個鹼基對內的一條或兩條股的切割。在一些實施方式中，載體編碼相對於相應野生型酶發生突變的Cas，使得突變型Cas缺乏切割含有靶序列的靶多核苷酸的一條或兩條股的能力。在一些實施方式中，當突變型酶的DNA切割活性不超過該酶的非突變形式的DNA切割活性的約25%、10%、5%、1%、0.1%、0.01%或更少時，Cas被認為是實質性缺乏所有DNA切割活性的；一實例可以是當突變形式的DNA切割活性係零或與非突變形式相比係可忽視的時候。因此，Cas可以包含一個或多個突變並且可以在融合或未融

合至功能結構域的情況下用作通用DNA結合蛋白。該等突變可以是人工引入的突變或者增功能或失功能突變。在本發明的一方面中，Cas酶可以融合至蛋白質，例如TAG，和/或誘導型/控制型結構域諸如化學誘導型/控制型結構域。本發明中的Cas可以是嵌合的Cas蛋白；例如，藉由成為嵌合體而具有增強功能的Cas。嵌合Cas蛋白可以是含有來自超過一種的天然存在的Cas的片段的新Cas。該等可以包括一種Cas9同源物的一個或多個N末端片段與另一種Cas同源物的一個或多個C末端片段的融合物。Cas可以呈mRNA形式被遞送至細胞中。Cas的表現可以在誘導型啟動子的控制下。本發明的目的明確是避免已知突變上的讀取。實際上，短語“由此相比於未修飾酶CRISPR複合物中的酶具有降低的修飾一個或多個脫靶座位的能力並且/或者由此相比於未修飾酶CRISPR複合物中的酶具有增加的修飾一個或多個靶座位的能力”（或類似表現）”不意圖讀取僅產生切口酶或無效Cas的突變或已知的Cas9突變。然而，這不是說本發明的一種或多種修飾或一個或多個突變，“由此相比於未修飾酶CRISPR複合物中的酶具有降低的修飾一個或多個脫靶座位的能力並且/或者由此相比於未修飾酶CRISPR複合物中的酶具有增加的修飾一個或多個靶座位的能力”（或類似表現）”不可以與產生為切口酶或無效酶的突變組合。此無效酶可以是增強的核酸分子結合劑。並且此切口酶可以是增強的切口酶。例如，將溝中和溝附近的一個或多個中性胺基酸和/或Cas中緊密接近核酸（例如，DNA、cDNA、RNA、gRNA）的其他位置中的其他帶電荷的殘基改變為一個或多個帶正電荷的胺基酸可以“由此使得相比於未修飾酶CRISPR複合物中的酶具有降低的

修飾一個或多個脫靶座位的能力並且/或者由此使得相比於未修飾酶CRISPR複合物中的酶具有增加的修飾一個或多個靶座位的能力”，例如產生更多切割。因為此切割可以是增強的中靶切割和脫靶切割兩者（超切割的Cpf1），使用了本領域已知的截短指導序列或截短sgRNA（例如參見付（Fu）等人，“使用截短的指導RNA改進CRISPR-Cas核酸酶特異性（Improving CRISPR-Cas nuclease specificity using truncated guide RNAs）”，自然生物技術，32, 279–284 (2014) doi:10.1038/nbt.2808，2013年12月17日接收，2014年1月06日接受，2014年1月26日線上公開，2014年1月29日線上修正）以使得靶活性增強而沒有較高的脫靶切割或用於產生超切割的切口酶，或用於與使得Cas無效用作超結合劑的突變組合。

[0264] 在某些實施方式中，工程化的Cpf1蛋白的活性改變包括靶向效率增加或脫靶結合減少。在某些實施方式中，工程化的Cpf1蛋白的活性改變包括切割活性的修飾。

[0265] 在某些實施方式中，活性改變包括關於包含RNA的核酸分子或靶多核苷酸座位的結合特性改變，關於包含RNA的核酸分子或靶多核苷酸座位的結合動力學改變，或關於包含RNA的核酸分子或靶多核苷酸座位的相比於脫靶多核苷酸座位的結合特異性改變。

[0266] 在某些實施方式中，活性改變包括靶向效率增加或脫靶結合減少。在某些實施方式中，活性改變包括切割活性的修飾。在某些實施方式中，活性改變包括關於靶多核苷酸座位的切割活性增加。在某些實施方式中，活性改變包括關於靶多核苷酸座位

的切割活性減少。在某些實施方式中，活性改變包括關於脫靶多核苷酸座位的切割活性減少。在某些實施方式中，活性改變包括關於脫靶多核苷酸座位的切割活性增加。

[0267] 因此，在某些實施方式中，相比於脫靶多核苷酸座位，存在對靶多核苷酸座位的特異性的增加。在其他實施方式中，相比於脫靶多核苷酸座位，存在對靶多核苷酸座位的特異性的降低。

[0268] 在本發明的一方面中，工程化的Cpf1蛋白的活性改變包括解旋酶動力學的改變。

[0269] 在本發明的一方面中，工程化的Cpf1蛋白包含改變該蛋白質與包含RNA的核酸分子，或靶多核苷酸座位的股，或脫靶多核苷酸座位的股的締合的修飾。在本發明的一方面中，工程化的Cpf1蛋白包含改變CRISPR複合物的形成的修飾。

[0270] 在某些實施方式中，改變的Cpf1蛋白包含改變核酸分子對多核苷酸座位的靶向的修飾。在某些實施方式中，修飾包括蛋白質中與核酸分子締合的區中的突變。在某些實施方式中，修飾包括蛋白質中與靶多核苷酸座位的股締合的區中的突變。在某些實施方式中，修飾包括蛋白質中與脫靶多核苷酸座位的股締合的區中的突變。在某些實施方式中，修飾或突變包括蛋白質中與包含RNA的核酸分子、或靶多核苷酸座位的股、或脫靶多核苷酸座位的股締合的區中的正電荷的減少。在某些實施方式中，修飾或突變包括蛋白質中與包含RNA的核酸分子、或靶多核苷酸座位的股、或脫靶多核苷酸座位的股締合的區中的負電荷的減少。在某些實施方式中，修飾或突變包括蛋白質中與包含RNA的核酸分

子、或靶多核苷酸座位的股、或脫靶多核苷酸座位的股締合的區中的正電荷的增加。在某些實施方式中，修飾或突變包括蛋白質中與包含RNA的核酸分子、或靶多核苷酸座位的股、或脫靶多核苷酸座位的股締合的區中的負電荷的增加。在某些實施方式中，修飾或突變增加了蛋白質與包含RNA的核酸分子、或靶多核苷酸座位的股、或脫靶多核苷酸座位的股之間的空間位阻。在某些實施方式中，修飾或突變包括Lys、His、Arg、Glu、Asp、Ser、Gly或Thr的取代。在某些實施方式中，修飾或突變包括Gly、Ala、Ile、Glu或Asp的取代。在某些實施方式中，修飾或突變包括結合溝中的胺基酸取代。

[0271] 在一個方面中，本發明提供了：

一種如在此所限定的非天然存在的 CRISPR 酶，諸如 Cpf1，其中：

該酶與指導 RNA 複合以形成 CRISPR 複合物，

當處於 CRISPR 複合物中時，指導 RNA 靶向一個或多個靶多核苷酸座位並且該酶改變該多核苷酸座位，並且

該酶包含至少一個修飾，

由此相比於未修飾酶 CRISPR 複合物中的酶具有降低的修飾一個或多個脫靶座位的能力並且/或者由此相比於未修飾酶 CRISPR 複合物中的酶具有增加的修飾一個或多個靶座位的能力。

[0272] 在任一這樣的非天然存在的CRISPR酶中，修飾可以包括該酶的一個或多個胺基酸殘基的修飾。

[0273] 在任一這樣的非天然存在的CRISPR酶中，修飾可以包括位於包含為未修飾酶中的帶正電荷的殘基的區中的一個或多個胺基酸殘基的修飾。

[0274] 在任一這樣的非天然存在的CRISPR酶中，修飾可以包括未修飾酶中的一個或多個帶正電荷的胺基酸殘基的修飾。

[0275] 在任一這樣的非天然存在的CRISPR酶中，修飾可以包括未修飾酶中的一個或多個不帶正電荷的胺基酸殘基的修飾。

[0276] 修飾可以包括未修飾酶中的一個或多個不帶電荷的胺基酸殘基的修飾。

[0277] 修飾可以包括未修飾酶中的一個或多個帶負電荷的胺基酸殘基的修飾。

[0278] 修飾可以包括未修飾酶中的一個或多個疏水性胺基酸殘基的修飾。

[0279] 修飾可以包括未修飾酶中的一個或多個極性胺基酸殘基的修飾。

[0280] 在某些以上所述的非天然存在的CRISPR酶中，修飾可以包括位於溝中的一個或多個殘基的修飾。

[0281] 在某些以上所述的非天然存在的CRISPR酶中，修飾可以包括位於溝的外面的一個或多個殘基的修飾。

[0282] 在某些以上所述的非天然存在的CRISPR酶中，修飾包括一個或多個殘基的修飾，其中一個或多個殘基包括精胺酸、

組胺酸或賴胺酸。

[0283] 在任一以上所述的非天然存在的CRISPR酶中，該酶可以是藉由所述一個或多個殘基的突變來修飾的。

[0284] 在某些以上所述的非天然存在的CRISPR酶中，該酶係藉由所述一個或多個殘基的突變來修飾的，並且其中突變包括使用丙胺酸取代未修飾酶中的殘基。

[0285] 在某些以上所述的非天然存在的CRISPR酶中，該酶係藉由所述一個或多個殘基的突變來修飾的，並且其中突變包括使用天冬胺酸或穀胺酸取代未修飾酶中的殘基。

[0286] 在某些以上所述的非天然存在的CRISPR酶中，該酶係藉由所述一個或多個殘基的突變來修飾的，並且其中突變包括使用絲胺酸、蘇胺酸、天冬醯胺或穀胺醯胺取代未修飾酶中的殘基。

[0287] 在某些以上所述的非天然存在的CRISPR酶中，該酶係藉由所述一個或多個殘基的突變來修飾的，並且其中突變包括使用丙胺酸、甘胺酸、異亮胺酸、亮胺酸、蛋胺酸、苯丙胺酸、色胺酸、酪胺酸或纈胺酸取代未修飾酶中的殘基。

[0288] 在某些以上所述的非天然存在的CRISPR酶中，該酶係藉由所述一個或多個殘基的突變來修飾的，並且其中突變包括使用極性胺基酸殘基取代未修飾酶中的殘基。

[0289] 在某些以上所述的非天然存在的CRISPR酶中，該酶係藉由所述一個或多個殘基的突變來修飾的，並且其中突變包括

使用為非極性胺基酸殘基的胺基酸殘基取代未修飾酶中的殘基。

[0290] 在某些以上所述的非天然存在的CRISPR酶中，該酶係藉由所述一個或多個殘基的突變來修飾的，並且其中突變包括使用帶負電的胺基酸殘基取代未修飾酶中的殘基。

[0291] 在某些以上所述的非天然存在的CRISPR酶中，該酶係藉由所述一個或多個殘基的突變來修飾的，並且其中突變包括使用為非帶負電荷的胺基酸殘基的胺基酸殘基取代未修飾酶中的殘基。

[0292] 在某些以上所述的非天然存在的CRISPR酶中，該酶係藉由所述一個或多個殘基的突變來修飾的，並且其中突變包括使用不帶電荷的胺基酸殘基取代未修飾酶中的殘基。

[0293] 在某些以上所述的非天然存在的CRISPR酶中，該酶係藉由所述一個或多個殘基的突變來修飾的，並且其中突變包括使用為不帶電荷的胺基酸殘基的胺基酸殘基取代未修飾酶中的殘基。

[0294] 在某些以上所述的非天然存在的CRISPR酶中，該酶係藉由所述一個或多個殘基的突變來修飾的，並且其中突變包括使用疏水性胺基酸殘基取代未修飾酶中的殘基。

[0295] 在某些以上所述的非天然存在的CRISPR酶中，該酶係藉由所述一個或多個殘基的突變來修飾的，並且其中突變包括

使用為非疏水性胺基酸殘基的胺基酸殘基取代未修飾酶中的殘基。

[0296] 在一些實施方式中，CRISPR酶，諸如較佳的是Cpf1酶來源於土拉熱弗朗西絲菌1、土拉熱弗朗西絲菌新殺手亞種、易北普雷沃菌、毛螺旋菌科細菌MC2017 1、解肽丁酸弧菌、佩萊格裡尼菌科細菌GW2011_GWA2_33_10、儉菌總門細菌GW2011_GWC2_44_17、密斯氏菌屬某種SCADC、胺基酸球菌屬某種BV3L6、毛螺旋菌MA2020、候選白蟻甲烷枝原體、挑剔真細菌、牛莫拉氏菌237、稻田鉤端螺旋體、毛螺旋菌科細菌ND2006、狗口腔卟啉單胞菌3、解糖胨普雷沃菌或獼猴卟啉單胞菌的Cpf1（例如，如在此所述修飾的該等生物體中的一種的Cpf1）並且可以進一步包含突變或改變，並且可以是嵌合Cpf1。

[0297] 在某些實施方式中，Cpf1蛋白包含一個或多個核定位信號（NLS）結構域。在某些實施方式中，Cpf1蛋白包含至少兩個或更多個NLS。

[0298] 在某些實施方式中，Cpf1蛋白包括嵌合的CRISPR蛋白，該嵌合的CRISPR蛋白包含來自第一CRISPR異種同源物的第一片段和來自第二CRISPR異種同源物的第二片段，並且第一CRISPR異種同源物和第二CRISPR異種同源物係不同的。

[0299] 在某些實施方式中，酶係藉由在此列出的任一個殘基或對應異種同源物中的相應殘基的突變來修飾的，或包含修飾，例如包括藉由該突變進行的修飾、基本上由或由藉由該突變進行的修飾組成；或者該酶包含根據整個本申請中的揭露內容的任何

一個（單個）、兩個（雙重）、三個（三重）、四個（四重）或更多個位置的，或CRISPR酶異種同源物中的相應殘基或位置的修飾、基本由或由該修飾組成，例如包含在此所列舉的任一種Cpf1殘基的或CRISPR酶異種同源物中的相應殘基或位置的修飾、基本上由或由該修飾組成的酶。在此酶中，每個殘基可以藉由使用丙胺酸殘基的取代來修飾。

[0300] 申請人最近描述了一種用於產生具有增強的特異性的Cas9異種同源物之方法(斯萊馬克爾(Slaymaker)等人 2015“具有提高的特異性的合理工程化Cas9核酸酶（Rationally engineered Cas9 nucleases with improved specificity）”)。此策略可以用於增強Cpf1異種同源物的特異性。用於誘變的初級殘基較佳的是RuvC結構域中的所有帶正電荷的殘基。另外的殘基係在不同異種同源物之間為保守的帶正電荷的殘基。

[0301] 在某些實施方式中，Cpf1的特異性可以藉由使穩定化非靶向DNA股的殘基發生突變來改進。

[0302] 在某些以上所述的非天然存在的Cpf1酶中，酶係藉由一個或多個殘基（在RuvC結構域中）的突變來修飾的，該等殘基包括但不限於參照AsCpf1（胺基酸球菌屬某種BV3L6）的胺基酸位置編碼的位置R909、R912、R930、R947、K949、R951、R955、K965、K968、K1000、K1002、R1003、K1009、K1017、K1022、K1029、K1035、K1054、K1072、K1086、R1094、K1095、K1109、K1118、K1142、K1150、K1158、K1159、R1220、R1226、R1242和/或R1252。

[0303] 在某些以上所述的非天然存在的Cpf1酶中，酶係藉由一個或多個殘基（在RAD50結構域中）的突變來修飾的，該等殘基包括但不限於參照AsCpf1（胺基酸球菌屬某種BV3L6）的胺基酸位置編碼的位置K324、K335、K337、R331、K369、K370、R386、R392、R393、K400、K404、K406、K408、K414、K429、K436、K438、K459、K460、K464、R670、K675、R681、K686、K689、R699、K705、R725、K729、K739、K748和/或K752。

[0304] 在某些以上所述的非天然存在的Cpf1酶中，酶係藉由一個或多個殘基的突變來修飾的，該等殘基包括但不限於參照AsCpf1（胺基酸球菌屬某種BV3L6）的胺基酸位置編碼的位置R912、T923、R947、K949、R951、R955、K965、K968、K1000、R1003、K1009、K1017、K1022、K1029、K1072、K1086、R1103、R1226和/或R1252。

[0305] 在某些實施方式中，酶係藉由一個或多個殘基的突變來修飾的，該等殘基包括但不限於參照LbCpf1（毛螺旋菌科細菌ND2006）的胺基酸位置編碼的位置R833、R836、K847、K879、K881、R883、R887、K897、K900、K932、R935、K940、K948、K953、K960、K984、K1003、K1017、R1033、R1138、R1165和/或R1252。

[0306] 在某些實施方式中，Cpf1酶係藉由一個或多個殘基的突變來修飾的，該等殘基包括但不限於參照AsCpf1（胺基酸球菌屬某種BV3L6）的胺基酸位置編碼的位置K15、R18、K26、Q34、R43、K48、K51、R56、R84、K85、K87、N93、R103、N104、T118、

K123、K134、R176、K177、R192、K200、K226、K273、K275、T291、R301、K307、K369、S404、V409、K414、K436、K438、K468、D482、K516、R518、K524、K530、K532、K548、K559、K570、R574、K592、D596、K603、K607、K613、C647、R681、K686、H720、K739、K748、K757、T766、K780、R790、P791、K796、K809、K815、T816、K860、R862、R863、K868、K897、R909、R912、T923、R947、K949、R951、R955、K965、K968、K1000、R1003、K1009、K1017、K1022、K1029、A1053、K1072、K1086、F1103、S1209、R1226、R1252、K1273、K1282和/或K1288。

[0307] 在某些實施方式中，Cpf1酶係藉由一個或多個殘基的突變來修飾的，該等殘基包括但不限於參照FnCpf1（新殺手法蘭西斯菌U112)的胺基酸位置編碼的位置K15、R18、K26、R34、R43、K48、K51、K56、K87、K88、D90、K96、K106、K107、K120、Q125、K143、R186、K187、R202、K210、K235、K296、K298、K314、K320、K326、K397、K444、K449、E454、A483、E491、K527、K541、K581、R583、K589、K595、K597、K613、K624、K635、K639、K656、K660、K667、K671、K677、K719、K725、K730、K763、K782、K791、R800、K809、K823、R833、K834、K839、K852、K858、K859、K869、K871、R872、K877、K905、R918、R921、K932、I960、K962、R964、R968、K978、K981、K1013、R1016、K1021、K1029、K1034、K1041、K1065、K1084和/或K1098。

[0308] 在某些實施方式中，Cpf1酶係藉由一個或多個殘基的

突變來修飾的，該等殘基包括但不限於參照LbCpf1（毛螺旋菌科細菌ND2006）的胺基酸位置編碼的位置K15、R18、K26、K34、R43、K48、K51、R56、K83、K84、R86、K92、R102、K103、K116、K121、R158、E159、R174、R182、K206、K251、K253、K269、K271、K278、P342、K380、R385、K390、K415、K421、K457、K471、A506、R508、K514、K520、K522、K538、Y548、K560、K564、K580、K584、K591、K595、K601、K634、K640、R645、K679、K689、K707、T716、K725、R737、R747、R748、K753、K768、K774、K775、K785、K787、R788、Q793、K821、R833、R836、K847、K879、K881、R883、R887、K897、K900、K932、R935、K940、K948、K953、K960、K984、K1003、K1017、R1033、K1121、R1138、R1165、K1190、K1199和/或K1208。

[0309] 在某些實施方式中，酶係藉由一個或多個殘基的突變來修飾的，該等殘基包括但不限於參照MbCpf1（牛莫拉氏菌237）的胺基酸位置編碼的位置K14、R17、R25、K33、M42、Q47、K50、D55、K85、N86、K88、K94、R104、K105、K118、K123、K131、R174、K175、R190、R198、I221、K267、Q269、K285、K291、K297、K357、K403、K409、K414、K448、K460、K501、K515、K550、R552、K558、K564、K566、K582、K593、K604、K608、K623、K627、K633、K637、E643、K780、Y787、K792、K830、Q846、K858、K867、K876、K890、R900、K901、M906、K921、K927、K928、K937、K939、R940、K945、Q975、R987、R990、K1001、R1034、I1036、R1038、R1042、K1052、K1055、K1087、R1090、K1095、N1103、K1108、K1115、K1139、K1158、R1172、

K1188、K1276、R1293、A1319、K1340、K1349和/或K1356。

[0310] 在任一非天然存在的CRISPR酶中：

在靶標與一個或多個脫靶座位的相應序列之間可以存在單個錯配；並且/或者

在靶標與一個或多個脫靶座位的相應序列之間可以存在兩個、三個或四個或更多個錯配，並且/或者

其中在 (ii) 中所述兩個、三個或四個或更多個錯配係連續的。

[0311] 在任一非天然存在的CRISPR酶中，相比於未修飾酶CRISPR複合物中的酶可以具有降低的修飾一個或多個脫靶座位的能力並且其中相比於未修飾酶CRISPR複合物中的酶具有增加的修飾所述靶座位的能力。

[0312] 在任一非天然存在的CRISPR酶中，相比於未修飾酶的相對差異，當在CRISPR複合物中時酶在靶標與至少一個脫靶座位之間的修飾能力的相對差異可以是增加的。

[0313] 在任一非天然存在的CRISPR酶中，CRISPR酶可以包含一個或多個另外的突變，其中一個或多個另外的突變係處於一個或多個催化活性結構域中。

[0314] 在此類非天然存在的CRISPR酶中，與缺乏所述一個或多個另外的突變的酶相比，CRISPR酶可以具有降低或廢除的核酸酶活性。

[0315] 在一些此類非天然存在的CRISPR酶中，CRISPR酶不

引導靶序列位置處的一條或另一條DNA股的切割。

[0316] 在CRISPR酶在一個或多個催化活性結構域中包含一個或多個另外的突變的情況下，一個或多個另外的突變可以是處於包含RuvCI、RuvCII或RuvCIII的CRISPR酶的催化活性結構域中。

[0317] 在不受理論束縛的情況下，在本發明的一方面中，描述的方法和突變對產生中靶位點處的切割並避免脫靶位點處的那些構象狀態的位置提供了增強的CRISPR酶結構域（例如Cpf1結構域）的構象重排。CRISPR酶以一系列的協調步驟切割靶DNA。首先，PAM相互作用結構域識別靶DNA的PAM序列5'。PAM結合之後，對靶序列的前10-12個核苷酸（種子序列）進行取樣以用於gRNA:DNA互補，一種依賴於DNA雙股體分離的方法。如果種子序列核苷酸互補gRNA，則DNA的剩餘部分被解旋並且gRNA的全長度與靶DNA股雜交。核苷酸溝可以藉由與DNA磷酸骨架的正電荷的非特異性相互作用來穩定化非靶向的DNA股並且促進解旋。在對抗cDNA：ncDNA再次雜交的競爭中，RNA：cDNA和CRISPR酶:ncDNA相互作用驅動DNA解旋。其他CRISPR酶結構域也可以影響核酸酶結構域的構象，例如連接不同結構域的接頭。因此，所提供的方法和突變涵蓋但不限於RuvCI、RuvCIII、RuvCIII和接頭。藉由靶DNA結合，包括種子序列相互作用以及與靶DNA股和非靶DNA股的相互作用引起的例如Cpf1的構象變化確定了結構域是否被定位成觸發核酸酶活性。因此，在此所提供的方法和突變展示並實現了超過PAM識別和RNA-DNA鹼基配對的修飾。

[0318] 在一方面中，本發明提供了如在此所限定的CRISPR

核酸酶，諸如Cpf1，當涉及中靶相互作用時其具有朝向與切割活性相關聯的構象的改進平衡並且/或者當涉及脫靶相互作用時其具有遠離與切割活性相關聯的構象的改進平衡。在一個方面中，本發明提供了具有改進的校對功能的Cas（例如Cpf1）核酸酶，即採用一在中靶位點處具有核酸酶活性的構象的Cas（例如Cpf1）核酸酶，並且該構象在脫靶位點處具有增加的不利性。斯騰伯格（Sternberg）等人（自然 527(7576):110-3, doi: 10.1038/nature15544，2015年10月28日線上公開，電子版2015年10月28日）使用了螢光共振能量轉移（FRET）實驗來檢測Cas（例如Cpf1）催化結構域在與中靶DNA和脫靶DNA締合時的相對取向，並且這可以外推到本發明的CRISPR酶（例如Cpf1）。

[0319] 本發明進一步提供了用於使用修飾的指導RNA調控核酸酶活性和/或特異性的方法和突變。如所討論的，中靶核酸酶活性可以是增加的或減少的。另外，脫靶核酸酶活性可以是增加的或減少的。此外，對中靶活性的特異性相比於對脫靶活性的特異性可以存在增加或減少。修飾指導RNA包括但不限於，截短的指導RNA、無效指導RNA、化學上修飾的指導RNA、與功能結構域締合的指導RNA、包含功能結構域的修飾指導RNA、包含適配體的修飾指導RNA、包含轉接蛋白的修飾指導RNA以及包含添加或修飾的環的指導RNA。在一些實施方式中，一個或多個功能結構域與無效gRNA（dRNA）締合。在一些實施方式中，與CRISPR酶複合的dRNA引導功能結構域在一個基因座位處的基因調節，同時gRNA引導CRISPR酶在另一個座位處的DNA切割。在一些實施方式中，dRNA被選擇為與脫靶調節相比使對感興趣的基因座位的

調節選擇性最大化。在一些實施方式中，dRNA被選擇為最大化靶基因調節並且最小化靶標切割。

[0320] 出於以下討論的目的，提及功能結構域可以是與CRISPR酶締合的功能結構域或者與轉接蛋白締合的功能結構域。

[0321] 在本發明的實踐中，可以在不與Cas（例如Cpf1）蛋白碰撞的情況下藉由插入可以募集轉接蛋白的一個或多個不同RNA環或一個或多個不同序列來擴展gRNA的環，該等轉接蛋白可以結合一個或多個不同RNA環或者一個或多個不同序列。該等轉接蛋白可以包括但不限於，存在於各種噬菌體外殼蛋白內的正交RNA結合蛋白/適配體組合。此類外殼蛋白列表包括但不限於：Qβ、F2、GA、fr、JP501、M12、R17、BZ13、JP34、JP500、KU1、M11、MX1、TW18、VK、SP、FI、ID2、NL95、TW19、AP205、ϕCb5、ϕCb8r、ϕCb12r、ϕCb23r、7s以及PRR1。該等轉接蛋白或正交RNA結合蛋白可以進一步募集包含一個或多個功能結構域的效應蛋白或融合物。在一些實施方式中，功能結構域可以是選自下組，該組由以下各項組成：易位酶結構域、整合酶結構域、重組酶結構域、解離酶結構域、轉化酶結構域、蛋白酶結構域、DNA甲基轉移酶結構域、DNA羥甲基酶結構域、DNA脫甲基酶結構域、組蛋白乙醯化酶結構域、組蛋白脫乙醯化酶結構域、核酸酶域、阻遏物結構域、活化物結構域、核定位信號結構域、轉錄調節蛋白質（或轉錄複合物募集）結構域、細胞攝取活性相關結構域、核酸結合結構域、抗體呈遞結構域、組蛋白修飾酶、組蛋白修飾酶的募集物；組蛋白修飾酶、組蛋白甲基轉移酶、組蛋白去甲酶、組

蛋白激酶、組蛋白磷酸酶、組蛋白核糖基酶、組蛋白脫核糖基酶、組蛋白泛素酶、組蛋白脫泛素酶、組蛋白生物素酶以及組蛋白尾蛋白酶的抑制劑。在一些較佳的實施方式中，功能結構域係轉錄啟動結構域，諸如但不限於，VP64、p65、MyoD1、HSF1、RTA、SET7/9或組蛋白乙醯轉移酶。在一些實施方式中，功能結構域係轉錄阻遏結構域，較佳的是KRAB。在一些實施方式中，轉錄阻遏結構域係SID或SID的串聯體（例如SID4X）。在一些實施方式中，功能結構域係表觀遺傳修飾結構域，以便提供表觀遺傳修飾酶。在一些實施方式中，功能結構域係啟動結構域，它可以是P65啟動結構域。在一些實施方式中，功能結構域係脫胺酶，諸如胞苷脫胺酶。胞苷脫胺酶（deaminese）可以被引導至靶核酸，在這兒胞苷脫胺酶引導胞苷至尿苷的轉化，產生C至T的取代（在互補股上G變成A）。在此實施方式中，可以在不存在DNA切割的情況下實施核苷酸取代。

[0322] 在一個方面中，本發明還提供了用於調控Cas（例如Cpf1）結合活性和/或結合特異性的方法和突變。在某些實施方式中，使用了缺乏核酸酶活性的Cas（例如Cpf1）蛋白。在某些實施方式中，採用了促進Cas（例如Cpf1）核酸酶的結合但不促進其核酸酶活性的修飾指導RNA。在此類實施方式中，中靶結合可以被增加或減少。另外，在此類實施方式中，脫靶結合可以被增加或減少。另外，對於中靶結合相比於脫靶結合的特異性可以存在增加或減少。

[0323] 在特定實施方式中，脫靶切割的減少係藉由以下方式

確保的：藉由去穩定化股分離，更具體地是藉由在Cpf1酶中引入減少DNA相互作用區中的正電荷的突變（如在此所述的以及斯萊馬克爾等人 2016（科學，1;351(6268):84-8）進一步對Cas9的舉例說明）。在另外的實施方式中，脫靶切割的減少係藉由以下方式確保的：藉由將影響靶股與指導RNA序列之間的相互作用，更具體地是擾亂Cpf1與靶DNA股的磷酸骨架之間的相關作用的突變引入到Cpf1酶中，以此方式使得靶特異性活性保留而脫靶活性降低（如克萊因史迪華（Kleinstiver）等人 2016 自然，28;529(7587):490-5對Cas9描述的）。在特定實施方式中，脫靶活性係藉由修飾Cpf1來降低的，其中相比於野生型Cpf1，與靶股和非靶股的相互作用均被修飾。

[0324] 可以不同組合被採用來增加或減少活性和/或使中靶活性相比於脫靶活性的特異性增加或減少，或增加或減少結合和/或使中靶結合相比於脫靶結合的特異性增加或減少的方法和突變可以用於補償或增強被形成來促進其他效應的突變或修飾。被形成來促進其他效應的此類突變或修飾包括Cas（例如Cpf1）的突變或修飾和/或對指導RNA所進行的突變或修飾。在某些實施方式中，該等方法和突變與化學修飾的指導RNA一起使用。指導RNA化學修飾的實例包括但不限於在一個或多個末端核苷酸處摻入2′-O-甲基（M）、2′-O-甲基3′硫代磷酸酯（MS）或2′-O-甲基3′硫代PACE（MSP）。與未修飾的指導RNA相比，此類化學修飾的指導RNA可以具有增加的穩定性和增加的活性，但是中靶相比於脫靶的特異性係不可測的。（參見，亨戴爾，2015，自然生物技術33(9):985-9, doi: 10.1038/nbt.3290，2015年6月29日線上公開）。化學修飾的指

導RNA進一步包括但不限於具有硫代磷酸酯鍵和鎖定核酸（LNA）核苷酸的RNA，該鎖定核酸（LNA）核苷酸包含在核糖環的2'碳與4'碳之間的亞甲基橋。本發明的方法和突變用於使用化學修飾的指導RNA調控Cas（例如Cpf1）核酸酶活性和/或結合。

[0325] 在一方面中，本發明提供了用於調控如在此所限定的根據本發明的Cas（例如Cpf1）蛋白的結合和/或結合特異性的方法和突變，該等蛋白包含功能結構域諸如核酸酶、轉錄活化物、轉錄阻遏物等。例如，可以使得Cas（例如Cpf1）蛋白無核酸酶或藉由引入突變使得Cas（例如Cpf1）蛋白具有改變或降低的核酸酶活性，該等突變係例如像在此其他地方所述的Cpf1突變，並且包括如在此任何地方所述的例如參照FnCpf1p RuvC結構域的胺基酸位置的D917A、E1006A、E1028A、D1227A、D1255A、N1257A、D917A、E1006A、E1028A、D1227A、D1255A和N1257A；或例如參照推定的第二核酸酶結構域的N580A、N584A、T587A、W609A、D610A、K613A、E614A、D616A、K624A、D625A、K627A和Y629A。核酸酶缺乏的Cas（例如Cpf1）蛋白適用於RNA指導的靶序列依賴性的功能結構域遞送。本發明提供了用於調控Cas（例如Cpf1）蛋白的結合的方法和突變。在一個實施方式中，功能結構域包括VP64，提供了RNA指導的轉錄因子。在另一個實施方式中，功能結構域包括Fok I，提供了RNA指導的核酸酶活性。參考美國專利公開2014/0356959、美國專利公開2014/0342456、美國專利公開2015/0031132，以及瑪裡，P.（Mali, P.）等人，2013，科學339(6121):823-6, doi: 10.1126/science.1232033，2013年1月3日線上公開，並且藉由在此的傳授內容，本發明包括了該等文獻結合在

此的教義應用的方法和材料。在某些實施方式中，中靶結合被增加。在某些實施方式中，脫靶結合被減少。在某些實施方式中，中靶結合被減少。在某些實施方式中，脫靶結合被增加。因此，本發明還提供了功能化的Cas（例如Cpf1）結合蛋白的中靶結合相比於脫靶結合增加或減少的特異性。

[0326] 用作RNA指導的結合蛋白的Cas（例如Cpf1）不限於無核酸酶的Cas（例如Cpf1）。具有核酸酶活性的Cas（例如Cpf1）酶當與某些指導RNA一起使用時也可以用作NA指導的結合蛋白。例如，短指導RNA和包含與靶錯配的核苷酸的指導RNA可以促進RNA引導的Cas（例如Cpf1）與靶序列的結合，幾乎沒有或沒有產生靶切割。（例如參見，達爾曼（Dahlman）等人，2015，自然生物技術33(11):1159-1161, doi: 10.1038/nbt.3390，2015年10月05日線上公開）。在一個方面中，本發明提供了用於調控包含核酸酶活性的Cas（例如Cpf1）蛋白的結合的方法和突變。在某些實施方式中，中靶結合被增加。在某些實施方式中，脫靶結合被減少。在某些實施方式中，中靶結合被減少。在某些實施方式中，脫靶結合被增加。在某些實施方式中，中靶結合相比於脫靶結合的特異性存在增加或減少。在某些實施方式中，指導RNA-Cas（例如Cpf1）酶的核酸酶活性也被調控。

[0327] RNA–DNA異源雙股體形成對整個靶區（不僅僅是最靠近PAM的種子區序列）的切割活性和特異性係重要的。因此，截短的指導RNA顯示出降低的切割活性和特異性。在一個方面中，本發明提供了用於使用改變的指導RNA增加活性和特異性的方法

和突變。

[0328] 本發明還證明可以使得Cas（例如Cpf1）核酸酶特異性的修飾與靶向範圍的修飾一致。Cas（例如Cpf1）突變體可以被設計成具有增加的靶特異性以及PAM識別的調節性修飾，例如藉由選擇改變PAM特異性的突變並且將該等突變與增加（或者如果需要，減少）中靶序列相比於脫靶序列的特異性的核苷酸溝突變相結合來設計。在一個此實施方式中，PI結構域殘基發生突變以調節希望的PAM序列的識別同時一個或多個核苷酸溝胺基酸發生突變以改變靶特異性。在此所述的Cas（例如Cpf1）方法和修飾可以用於抵消由PAM識別的改變產生的特異性損失，增強由PAM識別的改變產生的特異性增益，抵消由PAM識別的改變產生的特異性增益或增強由PAM識別的改變產生的特異性喪失。

[0329] 該等方法和突變可以與具有改變的PAM識別的Cas（例如Cpf1）酶一起使用。包括的PAM的非限制性實例為如在此任何地方所述的。

[0330] 在另外的實施方式中，該等方法和突變用於修飾蛋白質。

[0331] 在任一非天然存在的CRISPR酶中，CRISPR酶可以包含一個或多個異源結構域。

[0332] 一個或多個異源功能結構域可以包括一個或多個核定位信號（NLS）結構域。一個或多個異源功能結構域可以包括至少兩個或更多個NLS。

[0333] 一個或多個異源功能結構域可以包括一個或多個轉錄啟動結構域。轉錄啟動結構域可以包括VP64。

[0334] 一個或多個異源功能結構域可以包括一個或多個轉錄阻遏結構域。轉錄阻遏結構域可以包括KRAB結構域或SID結構域。

[0335] 一個或多個異源功能結構域可以包括一個或多個核酸酶結構域。一個或多個核酸酶結構域可以包括Fok1。

[0336] 一個或多個異源功能結構域可以具有以下活性中的一種或多種：修飾烷酶活性、去甲酶活性、轉錄啟動活性、轉錄阻遏活性、轉錄釋放因子活性、組蛋白修飾活性、核酸酶活性、單股RNA切割活性、雙股RNA切割活性、單股DNA切割活性、雙股DNA切割活性以及核酸結合活性。

[0337] 至少一個或多個異源功能結構域可以處於或靠近酶的胺基末端和/或處於或靠近酶的羧基末端。

[0338] 一個或多個異源功能結構域可以藉由接頭部分融合至CRISPR酶，或系接至CRISPR酶，或連接至CRISPR酶。

[0339] 在任一非天然存在的CRISPR酶中，CRISPR酶可以包括來自於來自包括下項的屬的生物體的CRISPR酶：土拉熱弗朗西絲菌1、土拉熱弗朗西絲菌新殺手亞種、易北普雷沃菌、毛螺旋菌

科細菌*MC2017 1*、解肽丁酸弧菌、佩萊格裡尼菌科細菌*GW2011_GWA2_33_10*、儉菌總門細菌*GW2011_GWC2_44_17*、密斯氏菌屬某種*SCADC*、胺基酸球菌屬某種*BV3L6*、毛螺旋菌*MA2020*、候選白蟻甲烷枝原體、挑剔真細菌、牛莫拉氏菌*237*、稻田鉤端螺旋體、毛螺旋菌科細菌*ND2006*、狗口腔卟啉單胞菌*3*、解糖胨普雷沃菌或獼猴卟啉單胞菌（例如，如在此所述修飾的該等生物體中的一種的Cpf1）並且可以進一步包含突變或改變，並且可以是嵌合Cas（例如Cpf1）。

[0340] 在任一非天然存在的CRISPR酶中，CRISPR酶可以包括嵌合的Cas（例如Cpf1）酶，該嵌合的Cas（例如Cpf1）酶包含來自第一Cas（例如Cpf1）異種同源物的第一片段和來自第二Cas（例如Cpf1）異種同源物的第二片段，並且第一Cas（例如Cpf1）異種同源物和第二Cas（例如Cpf1）異種同源物係不同的。第一Cas（例如Cpf1）異種同源物和第二Cas（例如Cpf1）異種同源物中的至少一者可以包括來自包括下項的生物體的Cas（例如Cpf1）：土拉熱弗朗西絲菌*1*、土拉熱弗朗西絲菌新殺手亞種、易北普雷沃菌、毛螺旋菌科細菌*MC2017 1*、解肽丁酸弧菌、佩萊格裡尼菌科細菌*GW2011_GWA2_33_10*、儉菌總門細菌*GW2011_GWC2_44_17*、密斯氏菌屬某種*SCADC*、胺基酸球菌屬某種*BV3L6*、毛螺旋菌科細菌*MA2020*、候選白蟻甲烷枝原體、挑剔真細菌、牛莫拉氏菌*237*、稻田鉤端螺旋體、毛螺旋菌科細菌*ND2006*、狗口腔卟啉單胞菌*3*、解糖胨普雷沃菌或獼猴卟啉單胞菌。

[0341] 在任一非天然存在的CRISPR酶中，編碼CRISPR酶的

核苷酸序列可以被密碼子優化為在真核生物中表現。

[0342] 在任一非天然存在的CRISPR酶中，細胞可以是真核細胞或原核細胞；其中CRISPR複合物在細胞中是可操作的，並且由此相比於未修飾酶CRISPR複合物的酶具有降低的修飾細胞的一個或多個脫靶座位的能力並且/或者由此相比於未修飾酶CRISPR複合物中的酶具有增加的修飾一個或多個靶座位的能力。

[0343] 因此，在一個方面中，本發明提供了一種包含如在此所限定的工程化的CRISPR蛋白或系統的真核細胞。

[0344] 在某些實施方式中，如在此所述的方法可以包括提供Cas（例如Cpf1）轉基因細胞，其中提供或引入了編碼一個或多個指導RNA的一種或多種核酸，在該細胞中該等核酸可操作地與包含一種或多種感興趣的基因的啟動子的調節元件連接。如在此所用，術語“Cas轉基因細胞”係指細胞諸如真核細胞，其中Cas基因已經是基因組整合的。根據本發明，細胞的性質、類型或來源不是特別限制的。另外，Cas轉基因被引入細胞中的方式可以變化並且可以是如本領域已知的任何方法。在某些實施方式中，Cas轉基因細胞係藉由將Cas轉基因引入分離細胞中來獲得的。在某些其他實施方式中，Cas轉基因細胞係藉由從Cas轉基因生物體中分離細胞來獲得的。作為實例而非限制，如在此提及的Cas轉基因細胞可以是來源於Cas轉基因真核生物，諸如Cas敲入真核生物。參考WO 2014/093622（PCT/US13/74667），該專利藉由引用結合在此。轉讓給桑加莫生物科技公司（Sangamo BioSciences, Inc.）的美國專利公開案號20120017290和20110265198中的針對靶向玫瑰屬（Rosa）

座位的方法可以被修改成利用本發明的CRISPR Cas系統。轉讓給策勒克提斯公司（Cellectis）的美國專利公開案號20130236946中的針對靶向玫瑰屬座位的方法也可以被修改成利用本發明的CRISPR Cas系統。作為另外的實例,參考普萊特（Platt）等人（細胞；159(2):440-455 (2014)），描述了Cas9敲入小鼠，該文獻藉由引用結合在此，並且其可以外推到如在此所限定的本發明CRISPR酶。Cas轉基因可以進一步包含Lox-Stop-polyA-Lox（LSL）盒，從而使得Cas表現藉由Cre重組酶可誘導。可替代地，Cas轉基因細胞可以是藉由將Cas轉基因引入分離細胞中來獲得的。轉基因的遞送系統在本領域是熟知的。作為實例，也如在此其他地方所述的，Cas轉基因可以藉由載體（例如AAV、腺病毒、慢病毒）和/或粒子和/或奈米粒子遞送來遞送。

[0345] 技術人員應理解的是，細胞（諸如在此提及的Cas轉基因細胞）除具有整合的Cas基因或產生自Cas在與能夠將Cas指導至靶座位的RNA複合時的序列特異性作用的突變（例如像一個或多個致癌性突變）之外可以進一步包含基因組的改變，例如但不限於普萊特等人 (2014)；陳等人，(2014)或庫馬爾（Kumar）等人 (2009)。

[0346] 本發明還提供了一種包含如在此所述的諸如在此部分所述的工程化CRISPR蛋白的組成物。

[0347] 本發明還提供了一種包含CRISPR-Cas複合物的非天然存在的工程化組成物，該CRISPR-Cas複合物包含以上所述的任何非天然存在的CRISPR酶。

[0348] 在一個方面中，本發明提供了一種包含一種或多種載體的載體系統，其中一種或多種載體包含：

a) 可操作地連接至編碼如在此所限定的工程化CRISPR蛋白的核苷酸序列的第一調節元件；並且視情況

b) 可操作地連接至編碼包含指導RNA的一個或多個核酸分子的一種或多種核苷酸序列的第二調節元件，該指導RNA包含指導序列、同向重複序列，視情況其中組分 (a)和 (b) 位於相同或不同的載體上。

[0349] 本發明還提供了一種非天然存在的工程化組成物，該組成物包含：

被可操作地構造成用於將 CRISPR-Cas 複合物組分或包含或編碼所述組分的一種或多種多核苷酸序列遞送至細胞中的遞送系統，並且其中所述 CRISPR-Cas 複合物在該細胞中是可操作的，

CRISPR-Cas 複合物組分或編碼細胞中的轉錄和/或翻譯的一種或多種多核苷酸序列，該等 CRISPR-Cas 複合物組分包含：

(I) 如在此所述的非天然存在的 CRISPR 酶（例如，工程化的 Cpf1）；

(II) CRISPR-Cas 指導 RNA，包含：

指導序列，以及

同向重複序列，

其中相比於未修飾酶 CRISPR 複合物中的酶具有降低的修飾

一個或多個脫靶座位的能力並且/或者由此相比於未修飾酶CRISPR複合物中的酶具有增加的修飾一個或多個靶座位的能力。

[0350] 在一個方面中，本發明還提供了一種包含如在此所述的諸如在此部分所述的工程化CRISPR蛋白的系統。

[0351] 在任何此類組成物中，遞送系統可以包括酵母系統、脂質轉染系統、微注射系統、基因槍系統、病毒體、脂質體、免疫脂質體、聚陽離子、脂質：核酸軛合物或人工病毒體，如在此其他地方所限定的。

[0352] 在任何此類組成物中，遞送系統可以包括包含一種或多種載體的載體系統，並且其中組分 (II) 包含可操作地連接至包含指導序列、同向重複序列的多核苷酸序列的第一調節元件，並且視情況，並且其中組分 (I) 包含可操作地連接至編碼CRISPR酶的多核苷酸序列的第二調節元件。

[0353] 在任何此類組成物中，遞送系統可以包括包含一種或多種載體的載體系統，並且其中組分 (II) 包含可操作地連接至指導序列和同向重複序列的第一調節元件，並且其中組分 (I) 包含可操作地連接至編碼CRISPR酶的多核苷酸序列的第二調節元件。

[0354] 在任何此類組成物中，組成物可以包含超過一個的指導RNA，並且每個指導RNA具有不同的靶標，由此存在多重作用。

[0355] 在任何此類組成物中，一種或多種多核苷酸序列可以在一個載體上。

[0356] 本發明還提供了一種包含一種或多種載體的工程化、

非天然存在的成簇規律間隔短迴文重複序列（CRISPR）-CRISPR相關（Cas）（CRISPR-Cas）載體系統，該等載體包含：

a) 可操作地連接至編碼在此的任何一種本發明構建體的非天然存在的 CRISPR 酶的核苷酸序列的第一調節元件；以及

b) 可操作地連接至編碼一個或多個指導 RNA 的一種或多種核苷酸序列的第二調節元件，該指導 RNA 包含指導序列、同向重複序列，

其中：

組分 (a) 和 (b) 位於相同或不同的載體上，

形成了 CRISPR 複合物；

指導 RNA 靶向靶多核苷酸座位並且該酶改變該多核苷酸座位，並且

相比於未修飾酶 CRISPR 複合物中的酶具有降低的修飾一個或多個脫靶座位的能力並且/或者由此相比於未修飾酶 CRISPR 複合物中的酶具有增加的修飾一個或多個靶座位的能力。

[0357] 在此系統中，組分 (I) 可以包含可操作地連接至包含指導序列、同向重複序列的多核苷酸序列的第一調節元件，並且其中組分 (II) 可以包含可操作地連接至編碼CRISPR酶的多核苷酸序列的第二調節元件。在此系統中，其中可適用的指導RNA可以包括嵌合RNA。

[0358] 在此系統中，組分 (I) 可以包含可操作地連接至指

導序列和同向重複序列的第一調節元件，並且其中組分 (II) 可以包含可操作地連接至編碼CRISPR酶的多核苷酸序列的第二調節元件。此系統可以包含超過一個的指導RNA，並且每個指導RNA具有不同的靶標，由此存在多重作用。組分 (a) 和 (b) 可以在相同載體上。

[0359] 在包含載體的任何此類系統中，一種或多種載體可以包括一種或多種病毒載體，諸如一個或多個逆轉錄病毒、慢病毒、腺病毒、腺伴隨病毒或單純皰疹病毒。

[0360] 在包含調節元件的任何此類系統中，至少一種所述調節元件可以包括組織特異性啟動子。組織特異性啟動子可以引導哺乳動物血細胞中、哺乳動物肝細胞中或哺乳動物眼部細胞中的表現。

[0361] 在任一上述組成物或系統中，同向重複序列可以包含一個或多個蛋白質相互作用的RNA適配體。一個或多個適配體可以位於四核苷酸環中。一個或多個適配體可以能夠結合MS2噬菌體外殼蛋白。

[0362] 在任一上述組成物或系統中，細胞可以是真核細胞或原核細胞；其中CRISPR複合物在細胞中是可操作的，並且由此相比於未修飾酶CRISPR複合物的酶具有降低的修飾細胞的一個或多個脫靶座位的能力並且/或者由此相比於未修飾酶CRISPR複合物中的酶具有增加的修飾一個或多個靶座位的能力。

[0363] 本發明還提供了一種任一上述組成物或來自任一上

述系統的CRISPR複合物。

[0364] 本發明還提供了一種在細胞中修飾感興趣的座位之方法，該方法包括使細胞與在此所述的工程化CRISPR酶（例如工程化的Cpf1）、組成物中的任一種或在此所述的系統或載體系統中的任一種接觸，或者其中該細胞包含在該細胞內存在的任一在此所述的CRISPR複合物。在此類方法中，細胞可以是原核細胞或真核細胞，較佳的是真核細胞。在此類方法中，生物體可以包括細胞。在此類方法中，生物體可以不是人類或其他動物。

[0365] 任何此方法可以是離體的或在體外。

[0366] 在某些實施方式中，在細胞中編碼所述指導RNA或Cas蛋白中的至少一種的核苷酸序列可操作地與包含感興趣的基因的啟動子的調節元件連接，由此至少一種CRISPR-Cas系統組分的表現由感興趣的基因的啟動子驅動。“可操作地連接的”旨在意指編碼指導RNA和/或Cas的核苷酸序列以允許核苷酸序列表現的方式被連接至一個或多個調節元件，也如在此其他地方所提及的。術語“調節元件”也被在此其他地方描述。根據本發明，調節元件包括感興趣的基因的啟動子，諸如較佳的是感興趣的內源性基因的啟動子。在某些實施方式中，啟動子處於其內源性基因組位置。在此類實施方式中，編碼CRISPR和/或Cas的核酸在其天然基因組位置處的感興趣的基因的啟動子的轉錄控制下。在某些其他實施方式中，啟動子被提供在（單獨的）核酸分子諸如載體或質粒，或其他染色體外核酸上，即啟動子不被提供在其天然基因組位置處。在某些實施方式中，啟動子被基因組整合在非天然的基因組

位置處。

[0367] 任何此類方法，所述修飾可以包括調控基因表現。所述調節基因表現可以包括啟動基因表現和/或阻遏基因表現。因此，在一個方面中，本發明提供了一種調控基因表現之方法，其中該方法包括將如在此所述的工程化的CRISPR蛋白或系統引入到細胞中。

[0368] 本發明還提供了一種治療有需要的個體的疾病、病症或感染之方法，該方法包括給予有效量的在此所述的工程化CRISPR酶（例如工程化的Cpf1）、組成物、系統或CRISPR複合物中的任一種。疾病、病症或感染可以包括病毒感染。病毒感染可以是HBV。

[0369] 本發明還提供了以上所述的工程化CRISPR酶（例如工程化的Cpf1）、組成物、系統或CRISPR複合物中的任一種用於基因或基因組編輯之用途。

[0370] 本發明還提供了一種改變感興趣的基因組座位在哺乳動物細胞中的表現之方法，該方法包括使細胞與在此所述的工程化CRISPR酶（例如工程化的Cpf1）、組成物、系統或CRISPR複合物接觸並且從而遞送CRISPR-Cas（載體）並且使得CRISPR-Cas複合物形成且結合靶標，並且測定基因組座位的表現是否已經被改變，諸如增加或減少的表現，或基因產物的修飾。

[0371] 本發明還提供了以上所述的用作治療劑的工程化CRISPR酶（例如工程化的Cpf1）、組成物、系統或CRISPR複合物

中的任一種。治療劑可以用於基因或基因組編輯，或基因治療。

[0372] 在某些實施方式中，如在此所述的工程化CRISPR酶（例如工程化的Cpf1）的活性包括基因組DNA切割，視情況使得基因的轉錄減少。

[0373] 在一個方面中，本發明提供了一種根據如在此所述的方法而具有改變的基因組座位表現之分離細胞，其中改變的表現係與未經受改變基因組座位表現的方法的細胞相比較的。在一個相關的方面中，本發明提供了一種由此細胞建立的細胞系。

[0374] 在一個方面中，本發明提供了一種藉由操縱例如HSC（造血幹細胞）的感興趣基因組座位中的靶序列來修飾生物體或非人類生物體的方法，例如其中感興趣的基因組座位與異常的蛋白質表現或與疾病病狀或狀態相關聯的突變相關聯，該方法包括：

遞送至HSC，例如，經由使HSC與含有非天然存在或工程化的組成物的粒子接觸來進行，該組成物包含：

I. CRISPR-Cas系統指導RNA（gRNA）多核苷酸序列，該序列包括：

(a) 能夠雜交至HSC中的靶序列的指導序列，

(b) 同向重複序列，以及

II. CRISPR酶，視情況包含至少一個或多個核定位序列，

其中，指導序列引導CRISPR複合物與靶序列的序列特異性結合，並且

其中CRISPR複合物包含與 (1) 雜交至靶序列的指導序列複合的CRISPR酶；並且

該方法可以視情況還包括遞送HDR模板，例如經由使含有HDR模板的粒子接觸HSC，或使HSC接觸含有HDR模板的另一個粒子來進行，其中HDR模板提供了蛋白質的正常形式或較少異常形式的表現；其中“正常的”係對於野生型來說的，並且“異常的”可以是導致病狀或疾病狀態的蛋白質表現；並且

視情況，該方法可以包括從生物體或非人類生物體分離或獲得HSC，視情況擴增HSC群體，使一個或多個粒子與HSC進行接觸以獲得修飾的HSC群體，視情況擴增修飾HSC的群體，並且視情況向生物體或非人類生物體給予修飾HSC。

[0375] 在一個方面中，本發明提供了一種藉由操縱例如HSC的感興趣基因組座位中的靶序列來修飾生物體或非人類生物體之方法，例如其中感興趣的基因組座位與與異常的蛋白質表現或與疾病病狀或狀態相關聯的突變相關聯，該方法包括：遞送至HSC，例如經由使HSC與含有非天然存在或工程化的組成物的粒子接觸來進行，該組成物包含：I.(a) 能夠雜交至HSC中的靶序列的指導序列，和 (b) 至少一種或多種同向重複序列，以及II.視情況具有一個或多個NLS的CRISPR酶，並且指導序列引導CRISPR複合物與靶序列的序列特異性結合，並且其中CRISPR複合物包含與雜交至靶序列的指導序列複合的CRISPR酶；並且

該方法可以視情況還包括遞送HDR模板，例如經由使含有HDR模板的粒子接觸HSC，或使HSC接觸含有HDR模板的另一個

粒子來進行，其中HDR模板提供了蛋白質的正常形式或較少異常形式的表現；其中“正常的”係對於野生型來說的，並且“異常的”可以是導致病狀或疾病狀態的蛋白質表現；及

視情況，該方法可以包括從生物體或非人類生物體分離或獲得HSC，視情況擴增HSC群體，使一個或多個粒子與HSC進行接觸以獲得修飾的HSC群體，視情況擴增修飾HSC的群體，並且視情況向生物體或非人類生物體給予修飾HSC。

[0376] 遞送可以是對編碼CRISPR複合物中的任何一個或多個或全部的、有利地連接至用於體內表現的一個或多個調節元件的一種或多種多核苷酸的遞送，例如經由含有包含可操作地連接至一個或多個調節元件的一種或多種多核苷酸的載體的一個或多個粒子來進行。編碼CRISPR酶、指導序列、同向重複序列的多核苷酸中的任一種或全部可以是RNA。應瞭解的是，在提及是RNA且被稱為“包含”特徵此同向重複序列的多核苷酸的情況下，RNA序列包含該特徵。在多核苷酸係DNA且被稱為包含特徵此同向重複序列的情況下，DNA序列被或可以被轉錄成包含所討論特徵的RNA。在特徵係蛋白質諸如CRISPR酶的情況下，所提及的DNA或RNA序列被或可以被翻譯(並且在DNA的情況下，首先進行轉錄)。

[0377] 在某些實施方式中，本發明提供了一種藉由操縱HSC的感興趣基因組座位中的靶序列來修飾生物體，例如包括人類的哺乳動物或非人類哺乳動物或生物體之方法，例如其中感興趣的基因組座位與與異常的蛋白質表現或與疾病病狀或狀態相關聯的突變相關聯，該方法包括遞送，例如經由非天然存在或工程化的

組成物與HSC的接觸來進行，其中組成物包含一個或多個粒子，該等粒子包含可操作地編碼組成物以用於對該組成物進行表現的病毒、質粒或一種或多種核酸分子載體（例如RNA），其中該組成物包含：(A) I.可操作地連接至CRISPR-Cas系統RNA多核苷酸序列的第一調節元件，其中多核苷酸序列包含 (a) 能夠與真核細胞中的靶序列雜交的指導序列，(b) 同向重複序列，以及II.可操作地連接至編碼包含至少一個或多個核定位序列（或者如一些實施方式中的視情況至少一個或多個核定位序列可以不涉及NLS）的CRISPR酶的酶編碼序列的第二調節元件，其中 (a)、(b) 和 (c) 被佈置為5'至3'取向，其中組分I和II位於系統的相同或不同載體上，其中當轉錄時，並且指導序列引導CRISPR複合物與靶序列的序列特異性結合，並且其中CRISPR複合物包含與雜交至靶序列的指導序列複合的CRISPR酶，或 (B) 包含含有一種或多種載體的載體系統的非天然存在或工程化的組成物，該等載體包含I.可操作地連接至 (a) 能夠與真核細胞中的靶序列雜交的指導序列和 (b) 至少一種或多種同向重複序列的第一調節元件，II.可操作地連接至編碼CRISPR酶的酶編碼序列的第二調節元件，並且視情況，在可適用的情況下，其中組分I和II位於系統的相同或不同載體上，其中當轉錄時，並且指導序列引導CRISPR複合物與靶序列的序列特異性結合，並且其中CRISPR複合物包含與雜交至靶序列的指導序列複合的CRISPR酶；該方法可以視情況還包括遞送HDR模板，例如經由使含有HDR模板的粒子接觸HSC，或使HSC接觸含有HDR模板的另一個粒子來進行，其中HDR模板提供了蛋白質的正常形式或較少異常形式的表現；其中“正常的”係對於野生型來說的，

並且“異常的”可以是導致病狀或疾病狀態的蛋白質表現；並且視情況，該方法可以包括從生物體或非人類生物體分離或獲得HSC，視情況擴增HSC群體，使一個或多個粒子與HSC進行接觸以獲得修飾的HSC群體，視情況擴增修飾HSC的群體，並且視情況向生物體或非人類生物體給予修飾HSC。在一些實施方式中，組分I、II和III位於相同載體上。在其他實施方式中，組分I和II位於相同載體上，而組分III位於另一載體上。在其他實施方式中，組分I和III位於相同載體上，而組分II位於另一載體上。在其他實施方式中，組分II和III位於相同載體上，而組分I位於另一載體上。在其他實施方式中，組分I、II和III中的每個位於不同載體上。本發明還提供了一種如在此所述的病毒或質粒載體系統。

[0378] 對於操縱靶序列，申請人還意指靶序列的表觀遺傳操縱。這可以是針對靶序列的染色質狀態的，諸如藉由對靶序列的甲基化狀態的修飾（即甲基化或甲基化圖案或CpG島的添加或移除）、組蛋白修飾、增加或降低靶序列的可及性，或藉由促進3D折疊。應瞭解的是，在提及一種藉由操縱感興趣基因組座位中的靶序列來修飾生物體或包括人類的哺乳動物或非人類哺乳動物或生物體的方法的情況下，這可適用於作為整體的生物體（哺乳動物）或僅來自該生物體的單個細胞或細胞群體（如果生物體係多細胞的）。在人類的情況下，例如，申請人尤其設想了單個細胞或細胞群體並且該等可以較佳的是進行離體修飾並且接著重新引入。在此情況下，活組織檢查或其他組織或生物流體樣品可能是必需的。就這一點而言，幹細胞也是特別較佳的。但是，當然，也設想了體內實施方式。並且本發明對於HSC是特別有利的。

[0379] 在一些實施方式中本發明包括一種藉由操縱HSC中感興趣的基因組座位中的DNA雙股體的相反股上的第一靶序列和第二靶序列來修飾生物體或非人類生物體的方法，例如其中感興趣的基因組座位與與異常的蛋白質表現或與疾病病狀或狀態相關聯的突變相關聯，該方法包括遞送，例如藉由使HSC與包含非天然存在或工程化的組成物的一個或多個粒子接觸，該組成物包含：

I. 第一CRISPR-Cas（例如Cpf1）系統RNA多核苷酸序列，其中該第一多核苷酸序列包含：

(a) 能夠與該第一靶序列雜交的第一指導序列，

(b) 第一同向重複序列，以及

II. 第二CRISPR-Cas（例如Cpf1）系統指導RNA多核苷酸序列，其中該第二多核苷酸序列包含：

(a) 能夠與該第二靶序列雜交的第二指導序列，

(b) 第二同向重複序列，以及

III. 編碼包含至少一個或多個核定位序列和包含一個或多個突變的CRISPR酶的多核苷酸序列，其中 (a)、(b) 和 (c) 被佈置為5'至3'取向；或

IV. I.至III.中的一個或多個例如第一同向重複序列和第二同向重複序列、CRISPR酶的一種或多種的表現產物；

其中當轉錄時，第一指導序列和第二指導序列分別引導第一CRISPR複合物和第二CRISPR複合物與第一靶序列和第二靶序列

的序列特異性結合，其中第一CRISPR複合物包含與 (1) 雜交至第一靶序列的第一指導序列複合的CRISPR酶，其中第二CRISPR複合物包含與 (1) 雜交至第二靶序列的第二指導序列複合的CRISPR酶，其中編碼CRISPR酶的多核苷酸序列係DNA或RNA，並且其中第一指導序列引導DNA雙股體中靠近第一靶序列的一條股的切割並且第二指導序列引導靠近第二靶序列的另一條股的切割，從而誘導雙股斷裂，從而修飾生物體或非人類生物體；並且該方法可以視情況還包括遞送HDR模板，例如經由使含有HDR模板的粒子接觸HSC，或使HSC接觸含有HDR模板的另一個粒子來進行，其中HDR模板提供了蛋白質的正常形式或較少異常形式的表現；其中“正常的”係對於野生型來說的，並且“異常的”可以是導致病狀或疾病狀態的蛋白質表現；並且視情況，該方法可以包括從生物體或非人類生物體分離或獲得HSC，視情況擴增HSC群體，使一個或多個粒子與HSC進行接觸以獲得修飾的HSC群體，視情況擴增修飾HSC的群體，並且視情況向生物體或非人類生物體給予修飾HSC。在本發明的一些方法中，編碼CRISPR酶的多核苷酸序列、第一指導序列和第二指導序列、第一同向重複序列和第二同向重複序列中的任一種或全部。在本發明的另外實施方式中，編碼編碼CRISPR酶的序列的多核苷酸序列、第一指導序列和第二指導序列、第一同向重複序列和第二同向重複序列係RNA並且經由脂質體、奈米粒子、外來體、微泡或基因槍來遞送；但是，有利的是遞送經由粒子來遞送。在本發明的某些實施方式中，第一同向重複序列和第二同向重複序列享有100%一致性。在一些實施方式中，多核苷酸可以被包含在包含一種或多種載體的載體系統之內。在

較佳的實施方式中，第一CRISPR酶具有一個或多個突變，使得該酶係互補股切口酶，並且第二CRISPR酶具有一個或多個突變，使得該酶係非互補股切口酶。可替代地，第一酶可以是非互補股切口酶，並且第二酶可以是互補股切口酶。在本發明的較佳的方法中，引導DNA雙股體中的靠近第一靶序列的一條股的切割的第一指導序列和引導靠近第二靶序列的另一條股的切割的第二指導序列使得5'突出端產生。在本發明的實施方式中，5'突出端具有至多200個鹼基對，較佳的是至多100個鹼基對或更較佳的是至多50個鹼基對。在本發明的實施方式中，5'突出端具有至少26個鹼基對，較佳的是至少30個鹼基對或更較佳的是至少34-50個鹼基對。

[0380] 在一些實施方式中本發明包括一種藉由操縱例如HSC中感興趣的基因組座位中的DNA雙股體的相反股上的第一靶序列和第二靶序列來修飾生物體或非人類生物體的方法，例如其中感興趣的基因組座位與與異常的蛋白質表現或與疾病病狀或狀態相關聯的突變相關聯，該方法包括遞送，例如藉由使HSC與包含非天然存在或工程化的組成物的一個或多個粒子接觸，該組成物包含：

I. 可操作地連接至

(a) 能夠與該第一靶序列雜交的第一指導序列，以及

(b) 至少一種或多種指導重複序列的第一調節元件，

II. 可操作地連接至

(a) 能夠與該第二靶序列雜交的第二指導序列，以及

(b) 至少一種或多種指導重複序列的第一調節元件，

III. 可操作地連接至編碼CRISPR酶（例如Cpf1）的酶編碼序列的第三調節元件，以及

V. I.至IV.中的一個或多個例如第一同向重複序列和第二同向重複序列、CRISPR酶的一種或多種的表現產物；

其中組分I、II、III和IV位於系統的相同或不同載體上，當轉錄時，第一指導序列和第二指導序列分別引導第一CRISPR複合物和第二CRISPR複合物與第一靶序列和第二靶序列的序列特異性結合，其中第一CRISPR複合物包含與 (1) 雜交至第一靶序列的第一指導序列複合的CRISPR酶，其中第二CRISPR複合物包含與雜交至第二靶序列的第二指導序列複合的CRISPR酶，其中編碼CRISPR酶的多核苷酸序列係DNA或RNA，並且其中第一指導序列引導DNA雙股體中靠近第一靶序列的一條股的切割並且第二指導序列引導靠近第二靶序列的另一條股的切割，從而誘導雙股斷裂，從而修飾生物體或非人類生物體；並且該方法可以視情況還包括遞送HDR模板，例如經由使含有HDR模板的粒子接觸HSC，或使HSC接觸含有HDR模板的另一個粒子來進行，其中HDR模板提供了蛋白質的正常形式或較少異常形式的表現；其中“正常的”係對於野生型來說的，並且“異常的”可以是導致病狀或疾病狀態的蛋白質表現；並且視情況，該方法可以包括從生物體或非人類生物體分離或獲得HSC，視情況擴增HSC群體，使一個或多個粒子與HSC進行接觸以獲得修飾的HSC群體，視情況擴增修飾HSC的群體，並且視情況向生物體或非人類生物體給予修飾HSC。

[0381] 本發明還提供了一種如在此所述的載體系統。該系統可以包含一個、兩個、三個或四個不同載體。組分I、II、III和IV可以因此位於一個、兩個、三個或四個不同載體上，並且在此設想了可能的組分位置的所有組合，例如：組分I、II、III和IV可以位於同一載體上；組分I、II、III和IV可以各自位於不同載體上；組分I、II、III和IV可以位於總共兩個或三個不同載體上，包括設想的所有位置組合等。在本發明的一些方法中，編碼CRISPR酶的多核苷酸序列、第一指導序列和第二指導序列、第一同向重複序列和第二同向重複序列中的任一種或全部是RNA。在本發明的另外實施方式中，第一同向重複序列和第二同向重複序列享有100%一致性。在較佳的實施方式中，第一CRISPR酶具有一個或多個突變，使得該酶係互補股切口酶，並且第二CRISPR酶具有一個或多個突變，使得該酶係非互補股切口酶。另選地，第一酶可以是非互補股切口酶，並且第二酶可以是互補股切口酶。在本發明的另一個實施方式中，病毒載體中的一個或多個經由脂質體、奈米粒子、外來體、微泡或基因槍來遞送；但是，粒子遞送係有利的。

[0382] 在本發明的較佳的方法中，引導DNA雙股體中的靠近第一靶序列的一條股的切割的第一指導序列和引導靠近第二靶序列的另一條股的切割的第二指導序列使得5'突出端產生。在本發明的實施方式中，5'突出端具有至多200個鹼基對，較佳的是至多100個鹼基對或更較佳的是至多50個鹼基對。在本發明的實施方式中，5'突出端具有至少26個鹼基對，較佳的是至少30個鹼基對或更較佳的是至少34-50個鹼基對。

[0383] 在一些實施方式中本發明包括一種藉由以下方式來修飾例如HSC中的感興趣的基因組座位的方法：例如其中感興趣的基因組座位與與異常的蛋白質表現或與疾病病狀或狀態相關聯的突變相關聯，藉由將該突變引入到HSC中，例如藉由使HSC與包含具有一個或多個突變的Cas蛋白和分別靶向HSC中的DNA分子的第一條股和第二條股的兩個指導RNA的一個或多個粒子接觸，由此指導RNA靶向DNA分子並且Cas蛋白切口DNA分子的第一條股和第二條股中的每條，由此HSC中的靶標被改變；並且，其中Cas蛋白和兩個指導RNA並不是同時天然存在的並且該方法可以視情況還包括遞送HDR模板，例如經由使含有HDR模板的粒子接觸HSC，或使HSC接觸含有HDR模板的另一個粒子來進行，其中HDR模板提供了蛋白質的正常形式或較少異常形式的表現；其中“正常的”係對於野生型來說的，並且“異常的”可以是導致病狀或疾病狀態的蛋白質表現；並且視情況，該方法可以包括從生物體或非人類生物體分離或獲得HSC，視情況擴增HSC群體，使一個或多個粒子與HSC進行接觸以獲得修飾的HSC群體，視情況擴增修飾HSC的群體，並且視情況向生物體或非人類生物體給予修飾HSC。在本發明的較佳的方法中，Cas蛋白切口DNA分子的第一條股和第二條股中的每條使得5'突出端產生。在本發明的實施方式中，5'突出端具有至多200個鹼基對，較佳的是至多100個鹼基對或更較佳的是至多50個鹼基對。在本發明的實施方式中，5'突出端具有至少26個鹼基對，較佳的是至少30個鹼基對或更較佳的是至少34-50個鹼基對。在本發明的一個方面中，Cas蛋白被密碼子優化為在真核細胞，較佳的是哺乳動物細胞或人類細胞中表現。本發明

的方面涉及被減少的基因產物或被進一步引入到編碼基因產物的DNA分子中的模板多核苷酸或藉由允許兩個5'突出端重新退火並連接而被精確切斷的間插序列的表現、或被改變的基因產物的活性或功能、或被增加的基因產物的表現。在本發明的一個實施方式中，基因產物係蛋白質。

[0384] 在一些實施方式中本發明包括一種藉由以下方式來修飾例如HSC中感興趣的基因組座位之方法：例如其中感興趣的基因組座位與與異常的蛋白質表現或與疾病病狀或狀態相關聯的突變相關聯，藉由將該突變引入到HSC中，例如藉由使HSC與包含下項的一個或多個粒子接觸：

a) 可操作地連接至分別靶向HSC的雙股DNA分子的第一條股和第二條股的兩個CRISPR-Cas系統指導RNA中的每個的第一調節元件，以及

b) 可操作地連接至Cas（例如Cpf1）蛋白的第二調節元件，或

c) a) 或b) 的一種或多種表現產物，

其中組分 (a) 和 (b) 位於系統的相同或不同載體上，由此指導RNA靶向HSC的DNA分子並且Cas蛋白切口HSC中的DNA分子的第一條股和第二條股中的每條；並且其中Cas蛋白和兩個指導RNA並不同時天然存在；並且該方法可以視情況還包括遞送HDR模板，例如經由使含有HDR模板的粒子接觸HSC，或使HSC接觸含有HDR模板的另一個粒子來進行，其中HDR模板提供了蛋白質的

正常形式或較少異常形式的表現；其中“正常的”係對於野生型來說的，並且“異常的”可以是導致病狀或疾病狀態的蛋白質表現；並且視情況，該方法可以包括從生物體或非人類生物體分離或獲得HSC，視情況擴增HSC群體，使一個或多個粒子與HSC進行接觸以獲得修飾的HSC群體，視情況擴增修飾HSC的群體，並且視情況向生物體或非人類生物體給予修飾HSC。在本發明的方面中，指導RNA可以包含融合至同向重複序列的指導序列。本發明的方面涉及被減少的基因產物或被進一步引入到編碼基因產物的DNA分子中的模板多核苷酸或藉由允許兩個5’突出端重新退火並連接而被精確切斷的間插序列的表現、或被改變的基因產物的活性或功能、或被增加的基因產物的表現。在本發明的一個實施方式中，基因產物係蛋白質。在本發明的較佳的實施方式中，系統的載體係病毒載體。在另一個實施方式中，系統的載體經由脂質體、奈米粒子、外來體、微泡或基因槍來遞送；並且粒子係較佳的。在一個方面中，本發明提供了一種修飾HSC中的靶多核苷酸的方法。在一些實施方式中，該方法包括使得CRISPR複合物結合靶多核苷酸來實施所述靶多核苷酸的切割，從而修飾該靶多核苷酸，其中該CRISPR複合物包含與雜交至所述靶多核苷酸內的靶序列的指導序列複合的CRISPR酶，其中所述指導序列連接至同向重複序列。在一些實施方式中，所述切割包括藉由所述CRISPR酶切割靶序列的位置處的一條或兩條股。在一些實施方式中，所述切割使得靶基因的轉錄減少。在一些實施方式中，該方法進一步包括使用外源性模板多核苷酸藉由同源重組修復所述切割的靶多核苷酸，其中所述修復產生包括所述靶多核苷酸的一個或多個核苷酸的插入、

缺失或取代的突變。在一些實施方式中，所述突變使得由包含靶序列的基因表現的蛋白質中發生一個或多個胺基酸的變化。在一些實施方式中，該方法進一步包括例如經由一個或多個粒子將一個或多個載體或其一種或多種表現產物遞送至例如所述HSC，其中一個或多個載體驅動以下項中的一個或多個的表現：CRISPR酶、連接至同向重複序列的指導序列。在一些實施方式中，所述載體被遞送至受試者中的例如HSC。在一些實施方式中，所述修飾發生在細胞培養中的所述HSC中。在一些實施方式中，該方法進一步包括在所述修飾之前將所述HSC從受試者分離。在一些實施方式中，該方法進一步包括將所述HSC和/或衍生自其的細胞返回至所述受試者。

[0385] 在一個方面中，本發明提供了一種產生例如包含突變型疾病相關基因的HSC之方法。在一些實施方式中，疾病相關基因係與患病或發展病的風險的增加相關聯的任何基因。在一些實施方式中，該方法包括 (a) 例如經由一個或多個粒子將一種或多種載體或其一種或多種表現產物遞送至HSC中，其中一種或多種載體驅動以下項中的一個或多個的表現：CRISPR酶、連接至同向重複序列的指導序列。並且 (b) 使得CRISPR複合物結合靶多核苷酸以實施所述疾病相關基因內的靶多核苷酸的切割，其中CRISPR複合物包含與雜交至靶多核苷酸內的靶序列的指導序列複合的CRISPR酶，並且視情況，在可適用的情況下，從而產生包含突變型疾病相關基因的HSC。在一些實施方式中，所述切割包括藉由所述CRISPR酶切割靶序列的位置處的一條或兩條股。在一些實施方式中，所述切割使得靶基因的轉錄減少。在一些實施方

式中，該方法進一步包括使用外源性模板多核苷酸藉由同源重組修復所述切割的靶多核苷酸，其中所述修復產生包括所述靶多核苷酸的一個或多個核苷酸的插入、缺失或取代的突變。在一些實施方式中，所述突變使得來自包含靶序列的基因的蛋白質表現發生一個或多個胺基酸的變化。在一些實施方式中，修飾HSC被給予至動物以由此產生動物模型。

[0386] 在一個方面中，本發明提供了修飾例如HSC中的靶多核苷酸的方法。在一些實施方式中，該方法包括使得CRISPR複合物結合靶多核苷酸來實施所述靶多核苷酸的切割，從而修飾該靶多核苷酸，其中該CRISPR複合物包含與雜交至所述靶多核苷酸內的靶序列的指導序列複合的CRISPR酶，其中所述指導序列連接至同向重複序列。在其他實施方式中，本發明提供了一種修飾多核苷酸在來自例如HSC的真核細胞中的表現之方法。該方法包括藉由使用結合HSC中的多核苷酸的CRISPR複合物增加或減少靶多核苷酸的表現；有利的是CRISPR複合物經由一個或多個粒子遞送。

[0387] 在一些方法中，靶多核苷酸可以被失活以實施方式如HSC中表現的修飾。例如，在CRISPR複合物與細胞中的靶序列結合後，靶多核苷酸失活，這樣使得該序列不被轉錄，該編碼蛋白不被產生，或者該序列不會像野生型序列一樣起作用。

[0388] 在一些實施方式中，CRISPR-Cas系統的RNA，例如指導RNA或gRNA可以被修飾；例如以包含適配體或功能結構域。適配體係一種結合特異性靶分子的合成寡核苷酸；例如，已經藉由重複數輪的體外選擇或SELEX（指數富集配位基系統進化法）被

工程化為結合不同分子的核苷酸分子靶向諸如小分子、蛋白質、核酸以及甚至細胞、組織和生物體。適配體係有用的，因為它們提供比得上抗體的分子識別特性。除了其區分識別之外，適配體提供了超過抗體的優勢，包括它們在治療性應用中幾乎不或不引發免疫原性。因此，在本發明的實踐中，酶或RNA中的任一者或兩者可以包含功能結構域。

[0389] 在一些實施方式中，功能結構域係轉錄啟動結構域，較佳的是VP64。在一些實施方式中，功能結構域係轉錄阻遏結構域，較佳的是KRAB。在一些實施方式中，轉錄阻遏結構域係SID或SID的串聯體（例如SID4X）。在一些實施方式中，功能結構域係表觀遺傳修飾結構域，以便提供表觀遺傳修飾酶。在一些實施方式中，功能結構域係啟動結構域，它可以是P65啟動結構域。在一些實施方式中，功能結構域包括核酸酶活性。在一個此實施方式中，功能結構域包括Fok1。

[0390] 本發明還提供了一種包含以上所述的或來自任一以上所述方法的修飾CRISPR酶、組成物、系統或複合物中的任一種的體外或離體細胞。細胞可以是真核細胞或原核細胞。本發明還提供了此類細胞的子代。本發明還提供了一種任何此細胞或任一此子代的產物，其中該產物係如藉由CRISPR複合物的修飾CRISPR酶修飾的所述一個或多個靶座位的產物。產物可以是肽、多肽或蛋白質。一些此類產物可以藉由CRISPR複合物的修飾CRISPR酶來修飾。在一些此類修飾產物中，靶座位的產物物理上不同於未經所述修飾CRISPR酶修飾的所述靶座位的產物。

[0391] 本發明還提供了一種包含編碼以上所述的任一非天然存在的CRISPR酶的多核苷酸序列的多核苷酸分子。

[0392] 任一此多核苷酸可以進一步包含可操作地連接至編碼非天然存在的CRISPR酶的多核苷酸序列的一個或多個調節元件。

[0393] 在包含一個或多個調節元件的任一此多核苷酸中，一個或多個調節元件可以被可操作地構造成用於非天然存在的CRISPR酶在真核細胞中的表現。真核細胞可以是人類細胞。真核細胞可以是齧齒動物細胞，視情況小鼠細胞。真核細胞可以是酵母細胞。真核細胞可以是中國倉鼠卵巢（CHO）細胞。真核細胞可以是昆蟲細胞。

[0394] 在包含一個或多個調節元件的任一此多核苷酸中，一個或多個調節元件可以被可操作地構造成用於非天然存在的CRISPR酶在原核細胞中的表現。

[0395] 在包含一個或多個調節元件的任一此多核苷酸中，一個或多個調節元件可以被可操作地構造成用於非天然存在的CRISPR酶在體外系統中的表現。

[0396] 本發明還提供了一種包含任一上述多核苷酸分子的表現載體。本發明還提供了此或此類核苷酸分子例如被可操作地構造成用於表現蛋白質的此類多核苷酸分子，和/或一種或多種核酸組分，以及此或此類載體。

[0397] 本發明進一步提供了一種使Cas（例如Cpf1）形成突

變或形成突變或修飾的Cas（例如Cpf1）之方法，該突變或修飾的Cas（例如Cpf1）係如在此所述的根據本發明的CRISPR酶的異種同源物，該方法包括確定使得異種同源物可以緊密接近或可以觸及核酸分子，例如DNA、RNA、gRNA等的一個或多個胺基酸，和/或類似於或對應於如在此所述的根據本發明的CRISPR酶中的在此鑒定的一個或多個胺基酸的一個或多個胺基酸，和合成或製備或表現包含下項、由或基本由下項組成的異種同源物：一個或多個修飾和/或一個或多個突變，或如在此所討論的進行突變例如將中性胺基酸修飾例如改變或突變為帶電荷的例如帶正電荷的胺基酸，例如由丙胺酸變為例如賴胺酸。此修飾異種同源物可以用於CRISPR-Cas系統中；並且表現此修飾異種同源物的一個或多個核酸分子可以用於遞送編碼如在此所討論的CRISPR-Cas系統組分的分子的載體或其他遞送系統中。

[0398] 在一個方面中，本發明提供了有效中靶活性並且最小化脫靶活性。在一個方面中，本發明提供了由CRISPR蛋白進行的有效中靶切割並且最小化由CRISPR蛋白進行的脫靶切割。在一個方面中，本發明提供了在無DNA切割情況下CRISPR蛋白在基因座位處的特異性結合。在一個方面中，本發明提供了CRISPR蛋白在基因座位處的有效的指導序列引導的中靶結合並且最小化CRISPR蛋白的脫靶結合。因此，在一個方面中，本發明提供了靶特異性基因調節。在一個方面中，本發明提供了在無DNA切割情況下CRISPR酶在基因座位處的特異性結合。因此，在一個方面中，本發明使用單一CRISPR酶提供一個基因座位處的切割和不同基因座位處的基因調節。在一個方面中，本發明使用一種或多種

CRISPR蛋白和/或酶提供多個靶標的正交啟動和/或抑制和/或切割。

[0399] 在另一個方面中，本發明提供了一種功能性篩選離體或體內細胞庫中的基因組中基因之方法，該方法包括給予或表現包含多個CRISPR-Cas系統指導RNA（gRNA）的文庫並且其中該篩選進一步包括CRISPR酶的使用，其中CRISPR複合物被修飾成包含異源功能結構域。在一個方面中，本發明提供了一種用於篩選基因組之方法，該方法包括向宿主給予文庫或者在宿主體內表現文庫。在一方面中，本發明提供了一種如在此討論之方法，該方法進一步包括向宿主給予活化物或在宿主中表現活化物。在一方面中，本發明提供了如在此所討論之方法，其中該活化物附接至CRISPR蛋白。在一方面中，本發明提供了如在此所討論之方法，其中該活化物附接至CRISPR蛋白的N末端或C末端。在一方面中，本發明提供了如在此所討論之方法，其中該活化物附接至gRNA環。在一方面中，本發明提供了一種如在此討論之方法，該方法進一步包括向宿主給予阻遏劑或在宿主中表現阻遏劑。在一方面中，本發明提供了如在此所討論之方法，其中該篩選包括影響並檢測基因啟動、基因抑制或座位中的切割。

[0400] 在一個方面中，本發明提供了如在此所討論之方法，其中該宿主係真核細胞。在一方面中，本發明提供了如在此所討論之方法，其中該宿主係哺乳動物細胞。在一方面中，本發明提供了如在此所討論之方法，其中該宿主係非人類真核細胞。在一方面中，本發明提供了如在此所討論之方法，其中該非人類真核

細胞係非人類哺乳動物細胞。在一方面中，本發明提供了如在此所討論之方法，其中非人類哺乳動物細胞可以是包括但不限於，靈長類動物、牛、羊、豬類、犬、齧齒動物、兔科諸如猴、母牛、綿羊、豬、狗、兔、大鼠或小鼠的細胞。在一方面中，本發明提供了如在此所討論之方法，該細胞可以是非哺乳動物真核細胞諸如家禽鳥類（例如雞）、脊椎動物魚（例如鮭魚）或甲殼類動物（例如牡蠣、蛤、龍蝦、蝦）的細胞。在一個方面中，本發明提供了如在此所討論之方法，該非人類真核細胞係植物細胞。植物細胞可以是單子葉植物或雙子葉植物具有的細胞或栽培植物或糧食植物諸如木薯、玉米、高粱、大豆、小麥、燕麥或稻具有的細胞。植物細胞還可以是藻類、樹或生產植物、果實或蔬菜（例如，樹類諸如柑橘樹，例如桔子樹、葡萄柚樹或檸檬樹；桃樹或油桃樹；蘋果樹或梨樹；堅果樹諸如杏樹或核桃樹或阿月渾子樹；茄屬植物；芸苔屬植物；萵苣屬植物；菠菜屬植物；辣椒屬植物；棉花、煙草、蘆筍、胡蘿蔔、甘藍、青花菜、花椰菜、番茄、茄子、胡椒、萵苣、菠菜、草莓、藍莓、覆盆子、黑莓、葡萄、咖啡、可可等）具有的細胞。

[0401] 在一方面中，本發明提供了如在此所討論之方法，該方法包括遞送CRISPR-Cas複合物或其一種或多種組分或對其編碼的一個或多個核酸分子，其中所述一個或多個核酸分子可操作地連接至一種或多種調節序列並且在體內表現。在一方面中，本發明提供了如在此所討論之方法，其中該體內表現係經由慢病毒、腺病毒或AAV。在一個方面中，本發明提供了如在此所討論之方法，其中該遞送係經由粒子、奈米粒子、脂質或細胞穿透肽(CPP)。

[0402] 在特定的實施方式中，將CRISPR-Cas複合物靶向葉綠體可能是感興趣的。在許多情況下，此靶向可以是藉由稱為葉綠體轉運肽（CTP）或質體轉運肽的N末端延伸的存在來實現的。如果表現的多肽將要在植物質粒（例如葉綠體）中區室化，則來自細菌來源的染色體轉基因必須具有融合至編碼表現多肽的序列的編碼CTP序列的序列。因此，將外源性多肽定位至葉綠體常常藉由將編碼CTP序列的多核苷酸序列可操作地連接至編碼外源性多肽的多核苷酸的5'區來實現。在易位到葉綠體中的過程中，在處理步驟中去除CTP。然而，處理效率可以受CTP的胺基酸序列和肽的NH2末端處的附近序列影響。已被描述用於靶向至葉綠體的其他選擇係玉米cab-m7信號序列（美國專利7,022,896、WO 97/41228）、豌豆麩胱甘肽還原酶信號序列（WO 97/41228）以及US2009029861中描述的CTP。

[0403] 在一方面中，本發明提供了一對CRISPR-Cas複合物，每個複合物包含含有能夠與細胞中的感興趣的基因組座位中的靶序列雜交的指導序列的指導RNA（gRNA），其中每個sgRNA的至少一個環係藉由插入結合一種或多種轉接蛋白的一種或多種不同RNA序列來修飾的，並且其中該轉接蛋白與一個或多個功能結構域締合，其中每個CRISPR-Cas的每個gRNA包含具有DNA切割活性的功能結構域。在一個方面中，本發明提供了如在此所討論的成對的CRISPR-Cas複合物，其中DNA切割活性係歸因於Fok1核酸酶。

[0404] 在一個方面中，本發明提供了一種用於切割感興趣的

基因組座位中的靶序列之方法，該方法包括向細胞遞送CRISPR-Cas複合物或其一種或多種組分或對其編碼的一個或多個核酸分子，其中所述一個或多個核酸分子可操作地連接至一個或多個調節序列並且在體內表現。在一方面中，本發明提供了如在此所討論之方法，其中該遞送係經由慢病毒、腺病毒或AAV。在一方面中，本發明提供了如在此所討論之方法或如在此所討論的成對的CRISPR-Cas複合物，其中該對中的第一複合物的靶序列係處於雙股DNA的第一條股上並且該對中的第二複合物的靶序列係處於雙股DNA的第二條股上。在一方面中，本發明提供了如在此所討論之方法或如在此所討論的成對的CRISPR-Cas複合物，其中該第一複合物和該第二複合物的靶序列彼此接近，使得DNA以促進同源定向修復的方式切割。在一方面中，一種在此的方法可以進一步包括將模板DNA引入到細胞中。在一個方面中，一種在此的方法或者在此的成對的CRISPR-Cas複合物可以涉及其中每個CRISPR-Cas複合物具有CRISPR酶，該CRISPR酶被突變為使得它具有不超過未突變CRISPR酶的核酸酶活性的約5%的核酸酶活性。

[0405] 在一個方面中，本發明提供了一種如在此所討論的文庫、方法或複合物，其中gRNA被修飾為具有至少一個非編碼功能環，例如其中該至少一個非編碼功能環是具有阻遏作用的；例如，其中該至少一個非編碼功能環包含Alu。

[0406] 在一個方面中，本發明提供了一種用於改變或修飾基因產物的表現之方法。所述方法可以包括將工程化的非天然存在的CRISPR-Cas系統引入到含有並表現編碼基因產物的DNA分子

的細胞中，該CRISPR-Cas系統包含Cas蛋白和靶向DNA分子的指導RNA，由此該指導RNA靶向編碼基因產物的DNA分子並且該Cas蛋白切割編碼該基因產物的DNA分子，由此改變該基因產物的表現；並且其中Cas蛋白和指導RNA並不同時天然存在。本發明進一步包括密碼子優化為在真核細胞中表現的Cas蛋白。在一較佳的實施方式中，真核細胞係哺乳動物細胞，並且在一更較佳的實施方式中，哺乳動物細胞係人類細胞。在本發明的另一個實施方式中，基因產物的表現減少。

[0407] 在一方面中，本發明提供了改變的細胞和那些細胞的子代，以及由該等細胞產生的產物。本發明的CRISPR-Cas（例如Cpf1）蛋白和系統用於產生包含修飾的靶座位的細胞。在一些實施方式中，該方法可以包括使得核酸靶向複合物結合靶DNA或RNA來實施所述靶DNA或RNA的切割，從而修飾該靶DNA或RNA，其中該核酸靶向複合物包含與雜交至所述靶DNA或RNA內的靶序列的指導RNA複合的核酸靶向效應蛋白。在一個方面中，本發明提供了一種修復細胞中的遺傳座位的方法。在另一個方面中，本發明提供了一種修飾DNA或RNA在真核細胞中的表現的方法。在一些實施方式中，該方法包括使得核酸靶向複合物結合DNA或RNA，以使得所述結合導致所述DNA或RNA的表現增加或減少；其中該核酸靶向複合物包含與指導RNA複合的核酸靶向效應蛋白。類似的考慮因素和條件適用如上文針對修飾靶DNA或RNA的方法。實際上，該等取樣、培養和重新引入選擇跨本發明的多個方面而適用。在一個方面中，本發明提供了修飾真核細胞中的靶DNA或RNA的方法，該等方法可以是在體內、離

體或在體外。在一些實施方式中，該方法包括從人類或非人類動物取樣細胞或細胞群體，並且修飾該細胞或該等細胞。培養可以發生在離體的任何階段。此類細胞可以是但不限於植物細胞、動物細胞、任何生物體的特定細胞類型，包括幹細胞、免疫細胞、T細胞、B細胞、樹突細胞、心血管細胞、上皮細胞、幹細胞等。細胞可以根據本發明進行修飾以產生例如受控量的基因產物，該等受控量可以是增加的或減少的，這取決於用途，並且/或者發生突變。在某些實施方式中，細胞的基因組座位被修復。該細胞或該等細胞甚至可以被重新引入非人類動物或植物中。對於重新引入的細胞，可以較佳的是該等細胞係幹細胞。

[0408] 在一方面中，本發明提供了暫態包含CRISPR系統或組分的細胞。例如，CRISPR蛋白或酶和核酸被暫態提供給細胞並且基因組座位被改變，然後CRISPR系統的一種或多種組分的量進行衰減。隨後，獲得了CRISPR介導的遺傳改變的細胞、該等細胞子代以及包含該等細胞的生物體包含減少量的一種或多種CRISPR系統組分，或者不再含有該等一種或多種CRISPR系統組分。一非限制性實例係諸如在此進一步描述的自失活的CRISPR-Cas系統。因此，本發明提供了包含一種或多種CRISPR-Cas系統改變的遺傳座位，但基本上缺乏一種或多種CRISPR系統組分的細胞和生物體，以及該等細胞和生物體的子代。在某些實施方式中，CRISPR系統組分係基本上不存在的。此類細胞、組織和生物體有利地包含所希望的或所選擇的遺傳改變，但是喪失了潛在地可能起非特異性作用、產生安全問題或妨礙監管審批的CRISPR-Cas組分或其剩餘部分。同樣，本發明提供了由細

胞、生物體以及細胞和生物體的子代產生的產物。

誘導型 Cpf1 CRISPR-Cas 系統（"拆分的-Cpf1"）

[0409] 在一方面中，本發明提供了一種非天然存在或工程化的誘導型Cpf1 CRISPR-Cas系統，該系統包含：

附接至誘導型二聚體的第一半部的第一Cpf1融合構建體，以及

附接至誘導型二聚體的第二半部的第二Cpf1融合構建體，

其中第一Cpf1融合構建體被可操作地連接至一個或多個核定位信號，

其中第二Cpf1融合構建體被可操作地連接至一個或多個核輸出信號，

其中與誘導物能量源的接觸使得誘導型二聚體的第一半部和第二半部合在一起，

其中使誘導型二聚體的第一半部和第二半部合在一起允許第一Cpf1融合構建體和第二Cpf1融合構建體組成功能性Cpf1 CRISPR-Cas系統，

其中Cpf1 CRISPR-Cas系統包含含有能夠與細胞中的感興趣的基因組座位中的靶序列雜交的指導序列的指導RNA（gRNA），並且

其中功能性Cpf1 CRISPR-Cas系統結合靶序列並且視情況編輯基因組座位以改變基因表現。

[0410] 在本發明的一個方面中，在誘導型Cpf1 CRISPR-Cas系統中，誘導型二聚體係或包含誘導型異源二聚體或基本上由或由該誘導型異源二聚體組成。在一個方面中，在誘導型Cpf1 CRISPR-Cas系統中，誘導型異源二聚體的第一半部或第一部分或第一片段係或包含FKBP（視情況FKBP12）或由或基本上由該FKBP組成。在本發明的一個方面中，在誘導型Cpf1 CRISPR-Cas系統中，誘導型異源二聚體的第二半部或第二部分或第二片段係或包含FRB或由或基本上由該FRB組成。在本發明的一個方面中，在誘導型Cpf1 CRISPR-Cas系統中，第一Cpf1融合構建體的安排係或包含N' 末端Cpf1部分-FRB-NES或由或基本上由該N'末端Cpf1部分-FRB-NES組成。在本發明的一個方面中，在誘導型Cpf1 CRISPR-Cas系統中，第一Cpf1融合構建體的安排係或包含NES-N'末端Cpf1部分-FRB-NES或由或基本上由該NES-N'末端Cpf1部分-FRB-NES組成。在本發明的一個方面中，在誘導型Cpf1 CRISPR-Cas系統中，第二Cpf1融合構建體的安排係或包含C'末端Cpf1部分-FKBP-NLS或基本上由或由該C'末端Cpf1部分-FKBP-NLS組成。在一個方面中，本發明提供了誘導型Cpf1 CRISPR-Cas系統，第二Cpf1融合構建體的安排係或包含NLS-C'末端Cpf1部分-FKBP-NLS或由或基本上由該NLS-C'末端Cpf1部分-FKBP-NLS組成。在一個方面中，在誘導型Cpf1 CRISPR-Cas系統中，可以存在將Cpf1部分與誘導型二聚體的半部或部分或片段分開的接頭。在一個方面中，在誘導型Cpf1 CRISPR-Cas系統中，誘導物能量源係或包含雷帕黴素或基本上由或由雷帕黴素組成。在一個方面中，在誘導型Cpf1 CRISPR-Cas系統中，誘

導型二聚體係誘導型同源二聚體。在一個方面中，在誘導型Cpf1 CRISPR-Cas系統中，Cpf1係FnCpf1。在一個方面中，在誘導型Cpf1 CRISPR-Cas系統中，一個或多個功能結構域與Cpf1的一部分或兩部分締合，例如功能結構域視情況包括轉錄活化物、轉錄因子(transcriptional)或核酸酶諸如Fok1核酸酶。在一個方面中，在誘導型Cpf1 CRISPR-Cas系統中，功能性Cpf1 CRISPR-Cas系統結合靶序列並且酶係無效Cpf1，該無效Cpf1視情況與不具有至少一個突變的Cpf1相比具有至少97%或100%減弱的核酸酶活性(或不超過3%且有利地0%核酸酶活性)。本發明進一步包括並且本發明的一個方面提供了編碼如在此所討論的誘導型Cpf1 CRISPR-Cas系統的多核苷酸。

[0411] 在一個方面中，本發明提供了一種用於遞送第一Cpf1融合構建體的載體，根據如在此所討論的，該第一Cpf1融合構建體被附接至誘導型二聚體的第一半部或部分或片段並且可操作地連接至一個或多個核定位信號。在一方面中，本發明提供了一種用於遞送第二Cpf1融合構建體的載體，該第二Cpf1融合構建體被附接至誘導型二聚體的第二半部或部分或片段並且可操作地連接至一個或多個核輸出信號。

[0412] 在一方面中，本發明提供了一種用於遞送以下兩者的載體：第一Cpf1融合構建體，如在此所討論的，該第一Cpf1融合構建體被附接至誘導型二聚體的第一半部或部分或片段並且可操作地連接至一個或多個核定位信號；以及第二Cpf1融合構建體，如在此所討論的，該第二Cpf1融合構建體被附接至誘導型

二聚體的第二半部或部分或片段並且可操作地連接至一個或多個核輸出信號。

[0413] 在一方面中，載體可以是單一質粒或表現盒。

[0414] 在一方面中，本發明提供了一種用在此所討論的或表現如在此所討論的誘導型Cpf1 CRISPR-Cas系統的任一載體轉化的真核宿主細胞或細胞系。

[0415] 在一方面中，本發明提供了一種用在此所討論的或表現在此所討論的誘導型Cpf1 CRISPR-Cas系統的任一載體轉化的轉基因生物體，或它們的子代。在一個方面中，本發明提供了一種組成型地表現如在此所討論的誘導型Cpf1 CRISPR-Cas系統的模型生物體。

[0416] 在一方面中，本發明提供了非天然存在或工程化的誘導型Cpf1 CRISPR-Cas系統，該系統包含：

附接至誘導型異源二聚體的第一半部的第一Cpf1融合構建體，以及

附接至誘導型異源二聚體的第二半部的第二Cpf1融合構建體，

其中第一Cpf1融合構建體被可操作地連接至一個或多個核定位信號，

其中第二CPf1融合構建體被可操作地連接至核輸出信號，

其中與誘導物能量源的接觸使得誘導型異源二聚體的第一

半部和第二半部合在一起，

其中使誘導型異源二聚體的第一半部和第二半部合在一起允許第一Cpf1融合構建體和第二Cpf1融合構建體組成功能性Cpf1 CRISPR-Cas系統，

其中Cpf1 CRISPR-Cas系統包含含有能夠與細胞中的感興趣的基因組座位中的靶序列雜交的指導序列的指導RNA(gRNA)，並且

其中功能性Cpf1 CRISPR-Cas系統編輯基因組座位以改變基因表現。

[0417] 在一方面中，本發明提供了一種治療有需要的受試者的方法，該方法包括藉由用如在此所討論的多核苷酸或在此討論的任一載體轉化受試者來誘導基因編輯，並且向受試者給予誘導物能量源。本發明包括此多核苷酸或載體在藥物，例如用於治療受試者的此藥物的製造中，或用於治療受試者的此方法之用途。本發明包括在治療有需要的受試者的方法中使用的如在此所討論的多核苷酸或在此所討論的任一載體，該方法包括誘導基因編輯，其中該方法進一步包括向受試者給予誘導物能量源。在一個方面中，在該方法中，還提供了修復模板，例如藉由包含所述修復模板的載體來遞送。

[0418] 本發明還提供了一種治療有需要的受試者之方法，該方法包括藉由用在此所討論的多核苷酸或在此所討論的任一載體轉化受試者來誘導轉錄啟動或阻遏，其中所述多核苷酸或

載體編碼或包含催化失活的Cpf1和如在此所討論的一個或多個締合的功能結構域；該方法進一步包括向受試者給予誘導物能量源。本發明還提供了在治療有需要的受試者的方法中使用的在此所討論的多核苷酸或在此所討論的任一載體，該方法包括誘導轉錄啟動或阻遏，其中該方法進一步包括向受試者給予誘導物能量源。

[0419] 因此，本發明尤其包括同源二聚體以及異源二聚體、例如藉由突變產生的無效Cpf1或基本上不具有核酸酶活性的Cpf1、其中存在一個或多個NLS和/或一個或多個NES的系統或複合物；連接至拆分Cpf1的一個或多個功能結構域；包括治療方法的方法，以及用途。

[0420] 應瞭解的是，在此提及Cpf1、Cpf1蛋白或Cpf1酶的情況下，這包括本發明的拆分Cpf1。在一個方面中，本發明提供了一種用於改變或修飾基因產物的表現的方法。所述方法可以包括將工程化的非天然存在的Cpf1 CRISPR-Cas系統引入到含有並表現編碼基因產物的DNA分子的細胞中，該Cpf1 CRISPR-Cas系統包含Cpf1蛋白和靶向DNA分子的指導RNA，由此該指導RNA靶向編碼基因產物的DNA分子並且該Cpf1蛋白切割編碼該基因產物的DNA分子，由此改變該基因產物的表現；並且其中Cpf1蛋白和指導RNA並不同時天然存在。本發明包括包含連接至同向重複序列（DR）的指導序列的指導RNA。本發明進一步包括密碼子優化為在真核細胞中表現的Cpf1蛋白。在一較佳的實施方式中，真核細胞係哺乳動物細胞，並且在一更較佳的實施

方式中，哺乳動物細胞係人類細胞。在本發明的另一個實施方式中，基因產物的表現減少。

[0421] 在一個方面中，本發明提供了一種工程化的非天然存在的Cpf1 CRISPR-Cas系統，該Cpf1 CRISPR-Cas系統包含Cpf1蛋白和靶向編碼細胞中基因產物的DNA分子的指導RNA，由此該指導RNA靶向編碼基因產物的DNA分子並且該Cpf1蛋白切割編碼該基因產物的DNA分子，由此改變該基因產物的表現；並且其中Cpf1蛋白和指導RNA並不同時天然存在；此Cpf1包括本發明的拆分的Cpf1。本發明包括包含連接至DR序列的指導序列的指導RNA。本發明進一步包括密碼子優化為在真核細胞中表現的Cpf1蛋白。在一較佳的實施方式中，真核細胞係哺乳動物細胞，並且在一更較佳的實施方式中，哺乳動物細胞係人類細胞。在本發明的另一個實施方式中，基因產物的表現減少。

[0422] 在另一個方面中，本發明提供了一種包含一種或多種載體的工程化的非天然存在的載體系統，該等載體包含可操作地連接至靶向編碼基因產物的DNA分子的Cpf1 CRISPR-Cas系統指導RNA的第一調節元件以及可操作地連接至Cpf1蛋白的第二調節元件；此Cpf1包括本發明的拆分的Cpf1。組分 (a) 和 (b) 可以位於系統的相同或不同的載體上。指導RNA靶向編碼細胞中的基因產物的DNA分子並且Cpf1蛋白切割編碼該基因產物的該DNA分子，由此改變該基因產物的表現；並且其中Cpf1蛋白和指導RNA並不同時天然存在。本發明包括包含連接至DR序列的指導序列的指導RNA。本發明進一步包括密碼子優化為在真

核細胞中表現的Cpf1蛋白。在一較佳的實施方式中，真核細胞係哺乳動物細胞，並且在一更較佳的實施方式中，哺乳動物細胞係人類細胞。在本發明的另一個實施方式中，基因產物的表現減少。

[0423] 在一個方面中，本發明提供了一種包含一種或多種載體的載體系統。在一些實施方式中，該系統包含：(a) 可操作地連接至DR序列的第一調節元件和用於將一種或多種指導序列插入DR序列的下游的一個或多個插入位點，其中在表現時，指導序列引導Cpf1 CRISPR-Cas複合物與真核細胞中的靶序列的序列特異性結合，其中Cpf1 CRISPR-Cas複合物包含與 (1) 雜交至靶序列的指導序列，和 (2) DR序列複合的Cpf1；和 (b) 可操作地連接至編碼包含核定位序列的所述Cpf1酶的酶編碼序列的第二調節元件；其中組分 (a) 和 (b) 位於系統的相同或不同的載體上；此Cpf1包括本發明的拆分的Cpf1。在一些實施方式中，組分 (a) 進一步包含可操作地連接至第一調節元件的兩種或更多種指導序列，其中當表現時，兩種或更多種指導序列中的每種引導Cpf1 CRISPR-Cas複合物與真核細胞中的不同靶序列的序列特異性結合。

[0424] 在一些實施方式中，Cpf1 CRISPR-Cas複合物包含具有足以驅動所述Cpf1 CRISPR-Cas複合物在真核細胞的核中以可檢測的量積累的強度的一個或多個核定位序列。在不希望受到理論約束的情況下，據信核定位序列不是為真核生物中的Cpf1

CRISPR-Cas複合物活性所需要的，但是包括此類序列增強系統的活性，特別是對於靶向核中的核酸分子而言。

[0425] 在一些實施方式中，Cpf1酶係選自下組的細菌物種的Cpf1，該組由下項組成：土拉熱弗朗西絲菌1、土拉熱弗朗西絲菌新殺手亞種、易北普雷沃菌、毛螺旋菌科細菌MC2017 1、解朊丁酸弧菌、佩萊格裡尼菌科細菌GW2011_GWA2_33_10、儉菌總門細菌GW2011_GWC2_44_17、密斯氏菌屬某種SCADC、胺基酸球菌屬某種BV3L6、毛螺旋菌MA2020、候選白蟻甲烷枝原體、挑剔真細菌、牛莫拉氏菌237、稻田鉤端螺旋體、毛螺旋菌科細菌ND2006、狗口腔卟啉單胞菌3、解糖胨普雷沃菌和獼猴卟啉單胞菌，並且可以包括來源於該等生物體的突變型CPf1。酶可以是Cpf1同源物或異種同源物。在一些實施方式中，Cpf1被密碼子優化為在真核細胞中表現。在一些實施方式中，Cpf1引導靶序列位置處的一條或兩條股的切割。在一較佳的實施方式中，股斷裂係交錯切割的，產生了5'突出端。在一些實施方式中，第一調節元件係聚合酶III啟動子。在一些實施方式中，第二調節元件係聚合酶II啟動子。在一些實施方式中，同向重複序列具有16個核苷酸的最小長度並且具有單一莖環。在另外的實施方式中，同向重複序列具有長於16個核苷酸，較佳的是超過17個核苷酸的長度，並且具有超過一個的莖環或優化的二級結構。

[0426] 在一個方面中，本發明提供了一種包含以下項的真核宿主細胞：(a) 可操作地連接至同向重複序列的第一調節元件和用於將一種或多種指導序列插入DR序列的下游的一個或多個插入位

點，其中在表現時，指導序列引導Cpf1 CRISPR-Cas複合物與真核細胞中的靶序列的序列特異性結合，其中Cpf1 CRISPR-Cas複合物包含與 (1) 雜交至靶序列的指導序列，和 (2) DR序列複合的Cpf1；和/或 (b) 可操作地連接至編碼包含核定位序列的所述Cpf1酶的酶編碼序列的第二調節元件。在一些實施方式中，宿主細胞包含組分 (a) 和 (b)；此Cpf1包括本發明的拆分的Cpf1。在一些實施方式中，組分 (a)、組分 (b) 或組分 (a) 和 (b) 被穩定地整合到宿主真核細胞的基因組中。在一些實施方式中，組分 (a) 進一步包含可操作地連接至第一調節元件的兩種或更多種指導序列，其中當表現時，兩種或更多種指導序列中的每種引導Cpf1 CRISPR-Cas複合物與真核細胞中的不同靶序列的序列特異性結合。在一些實施方式中，CPf1被密碼子優化為在真核細胞中表現。在一些實施方式中，Cpf1引導靶序列位置處的一條或兩條股的切割。在一較佳的實施方式中，股斷裂係交錯切割的，產生了5'突出端。在一些實施方式中，Cpf1缺乏DNA股切割活性。在一些實施方式中，第一調節元件係聚合酶III啟動子。在一些實施方式中，同向重複序列具有16個核苷酸的最小長度並且具有單一莖環。在另外的實施方式中，同向重複序列具有長於16個核苷酸，較佳的是超過17個核苷酸的長度，並且具有超過一個的莖環或優化的二級結構。在一方面中，本發明提供了一種非人類真核生物體；較佳的是多細胞真核生物體，該等生物體包含根據任何所述實施方式的真核宿主細胞。在其他方面中，本發明提供了一種真核生物體；較佳的是多細胞真核生物體，該等生物體包含根據任何所述實施方式的真核宿主細胞。在該等方面的一些實施方式中，該生

物體可以是動物；例如，哺乳動物。而且，該生物體可以是節肢動物，諸如昆蟲。生物體還可以是植物。另外，生物體可以是真菌。

[0427] 在一個方面中，本發明提供了一種包含在此所述的一種或多種組分的套組。在一些實施方式中，套組包括載體系統和用於使用套組的說明書。在一些實施方式中，該載體系統包含：(a) 可操作地連接至同向重複序列的第一調節元件和用於將一種或多種指導序列插入DR序列的下游的一個或多個插入位點，其中在表現時，指導序列引導Cpf1 CRISPR-Cas複合物與真核細胞中的靶序列的序列特異性結合，其中Cpf1 CRISPR-Cas複合物包含與(1) 雜交至靶序列的指導序列，和 (2) DR序列複合的Cpf1；和/或(b) 可操作地連接至編碼包含核定位序列的所述Cpf1酶的酶編碼序列的第二調節元件並且有利的是此Cpf1包括本發明的拆分Cpf1。在一些實施方式中，套組包括位於系統的相同或不同的載體上的組分 (a) 和 (b)。在一些實施方式中，組分 (a) 進一步包含可操作地連接至第一調節元件的兩種或更多種指導序列，其中當表現時，兩種或更多種指導序列中的每種引導Cpf1 CRISPR-Cas複合物與真核細胞中的不同靶序列的序列特異性結合。在一些實施方式中，Cpf1包含具有足以驅動所述Cpf1在真核細胞的核中以可檢測的量積累的強度的一個或多個核定位序列。在一些實施方式中，Cpf1酶係選自下組的細菌物種的Cpf1，該組由下項組成：土拉熱弗朗西絲菌1、土拉熱弗朗西絲菌新殺手亞種、易北普雷沃菌、毛螺旋菌科細菌MC2017 1、解肽丁酸弧菌、佩萊格裡尼菌科細菌GW2011_GWA2_33_10、儉菌總門細菌

GW2011_GWC2_44_17、密斯氏菌屬某種SCADC、胺基酸球菌屬某種BV3L6、毛螺旋菌MA2020、候選白蟻甲烷枝原體、挑剔真細菌、牛莫拉氏菌237、稻田鉤端螺旋體、毛螺旋菌科細菌ND2006、狗口腔卟啉單胞菌3、解糖胨普雷沃菌和獼猴卟啉單胞菌，並且可以包括來源於該等生物體的突變型CPf1。酶可以是Cpf1同源物或異種同源物。在一些實施方式中，Cpf1被密碼子優化為在真核細胞中表現。在一些實施方式中，Cpf1引導靶序列位置處的一條或兩條股的切割。在一較佳的實施方式中，股斷裂係交錯切割的，產生了5'突出端。在一些實施方式中，CRISPR酶缺乏DNA股切割活性。在一些實施方式中，同向重複序列具有16個核苷酸的最小長度並且具有單一莖環。在另外的實施方式中，同向重複序列具有長於16個核苷酸，較佳的是超過17個核苷酸的長度，並且具有超過一個的莖環或優化的二級結構。

[0428] 在一個方面中，本發明提供了一種修飾真核細胞中的靶多核苷酸的方法。在一些實施方式中，該方法包括使得Cpf1 CRISPR-Cas複合物結合靶多核苷酸來實施所述靶多核苷酸的切割，從而修飾該靶多核苷酸，其中該Cpf1 CRISPR-Cas複合物包含與雜交至所述靶多核苷酸內的靶序列的指導序列複合的Cpf1，其中所述指導序列連接至同向重複序列。在一些實施方式中，所述切割包括藉由所述Cpf1切割靶序列的位置處的一條或兩條股；此Cpf1包括本發明的拆分的Cpf1。在一些實施方式中，所述切割使得靶基因的轉錄減少。在一些實施方式中，該方法進一步包括使用外源性模板多核苷酸藉由同源重組修復所述切割的靶多核苷酸，其中所述修復產生包括所述靶多核苷酸的一個或多個核苷酸的插

入、缺失或取代的突變。在一些實施方式中，所述突變使得由包含靶序列的基因表現的蛋白質中發生一個或多個胺基酸的變化。在一些實施方式中，該方法進一步包括將一個或多個載體遞送至所述真核細胞，其中一個或多個載體驅動以下項中的一個或多個的表現：Cpf1，和連接至DR序列的指導序列。在一些實施方式中，所述載體被遞送至受試者中的真核細胞。在一些實施方式中，所述修飾發生在細胞培養中的所述真核細胞中。在一些實施方式中，該方法進一步包括在所述修飾之前將所述真核細胞從受試者分離。在一些實施方式中，該方法進一步包括將所述真核細胞和/或衍生自其的細胞返回至所述受試者。

[0429] 在一個方面中，本發明提供了一種修飾多核苷酸在真核細胞中的表現之方法。在一些實施方式中，該方法包括使得Cpf1 CRISPR-Cas複合物結合該多核苷酸，以使得所述結合導致所述多核苷酸的表現增加或減少；其中Cpf1 CRISPR-Cas複合物包含與雜交至所述多核苷酸內的靶序列的指導序列複合的Cpf1，其中所述指導序列連接至同向重複序列；此Cpf1包括本發明的拆分的Cpf1。在一些實施方式中，該方法進一步包括將一個或多個載體遞送至所述真核細胞，其中一個或多個載體驅動以下項中的一個或多個的表現：Cpf1，和連接至DR序列的指導序列。

[0430] 在一個方面中，本發明提供了一種產生包含突變型疾病相關基因的模型真核細胞的方法。在一些實施方式中，疾病相關基因係與患病或發展病的風險的增加相關聯的任何基因。在一些實施方式中，該方法包括 (a) 將一個或多個載體引入到真核細

胞中，其中一個或多個載體驅動以下項中的一個或多個的表現：Cpf1，和連接至同向重複序列的指導序列；並且 (b) 使得Cpf1 CRISPR-Cas複合物結合靶多核苷酸以實施所述疾病相關基因內的靶多核苷酸的切割，其中Cpf1 CRISPR-Cas複合物包含與 (1) 雜交至靶多核苷酸內的靶序列的指導序列，和 (2) DR序列複合的Cpf1，從而產生包含突變型疾病相關基因的模型真核細胞；此Cpf1包括本發明的拆分的Cpf1。在一些實施方式中，所述切割包括藉由所述Cpf1切割靶序列的位置處的一條或兩條股。在一較佳的實施方式中，股斷裂係交錯切割的，產生了5'突出端。在一些實施方式中，所述切割使得靶基因的轉錄減少。在一些實施方式中，該方法進一步包括使用外源性模板多核苷酸藉由同源重組修復所述切割的靶多核苷酸，其中所述修復產生包括所述靶多核苷酸的一個或多個核苷酸的插入、缺失或取代的突變。在一些實施方式中，所述突變使得來自包含靶序列的基因的蛋白質表現發生一個或多個胺基酸的變化。

[0431] 在一個方面中，本發明提供了一種用於開發調控與疾病相關基因相關聯的細胞傳訊事件的生物活性劑之方法。在一些實施方式中，疾病相關基因係與患病或發展病的風險的增加相關聯的任何基因。在一些實施方式中，該方法包括 (a) 使測試化合物與任一種所述實施方式的模型細胞接觸；並且 (b) 檢測指示與所述疾病相關基因中的所述突變相關聯的細胞傳訊事件的減少或增加的讀出變化，從而開發調控與所述疾病相關基因相關聯的所述細胞傳訊事件的所述生物活性劑。

[0432] 在一個方面中，本發明提供了一種包含同向重複序列的下游的指導序列的重組多核苷酸，其中在表現時，指導序列引導Cpf1 CRISPR-Cas複合物與真核細胞中存在的相應靶序列的序列特異性結合。在一些實施方式中，靶序列係真核細胞中存在的病毒序列。在一些實施方式中，靶序列係原癌基因或癌基因。

[0433] 在一個方面中，本發明提供了一種藉由在一個或多個細胞的基因中引入一個或多個突變來選擇一個或多個細胞的方法，該方法包括：將一種或多種載體引入到一個或多個細胞中，其中一種或多種載體驅動以下項中的一個或多個的表現：Cpf1、連接至同向重複序列的指導序列，以及編輯模板；其中編輯模板包含廢除Cpf1切割的一個或多個突變；使得編輯模板與有待選擇的一個或多個細胞中的靶多核苷酸同源重組；使得Cpf1 CRISPR-Cas複合物結合靶多核苷酸以實施所述基因內的靶多核苷酸的切割，其中Cpf1 CRISPR-Cas複合物包含與 (1) 雜交至靶多核苷酸內的靶序列的指導序列，和 (2) 同向重複序列複合的Cpf1，其中Cpf1 CRISPR-Cas複合物與靶多核苷酸的結合誘導細胞死亡，從而使得其中已引入一個或多個突變的一個或多個細胞被選擇；此Cpf1包括本發明的拆分的Cpf1。在本發明的另一個較佳的實施方式中，有待選擇的細胞可以是真核細胞。本發明的方面允許在不需要選擇標記物或可能包括反選擇系統的兩步法的情況下選擇特異性細胞。

[0434] 在此存在短語“此Cpf1包括本發明的拆分Cpf1”或類似文本；並且這係表明在此的實施方式中的Cpf1可以是如在此所

討論的拆分Cpf1。

[0435] 在一方面中，本發明涉及一種非天然存在或工程化的誘導型Cpf1 CRISPR-Cas系統，該系統包含附接至誘導型異源二聚體的第一半部的第一Cpf1融合構建體和附接至誘導型異源二聚體的第二半部的第二Cpf1融合構建體，其中第一CPf1融合構建體被可操作地連接至一個或多個核定位信號，其中第二CPf1融合構建體被可操作地連接至核輸出信號，其中與誘導物能量源的接觸使得誘導型異源二聚體的第一半部和第二半部合在一起，其中使誘導型異源二聚體的第一半部和第二半部合在一起允許第一Cpf1融合構建體和第二Cpf1融合構建體組成功能性Cpf1 CRISPR-Cas系統，其中Cpf1 CRISPR-Cas系統包含含有能夠與細胞中的感興趣的基因組座位中的靶序列雜交的指導序列的指導RNA（gRNA），並且其中功能性Cpf1 CRISPR-Cas系統編輯基因組座位以改變基因表現。在本發明的一個實施方式中，誘導型異源二聚體的第一半部係FKBP12並且誘導型異源二聚體的第二半部係FRB。在本發明的另一個實施方式中，誘導物能量源係雷帕黴素。

[0436] 可以認為誘導物能量源係簡單的誘導物或二聚化劑。術語"誘導物能量源"在此至始至終的使用是一致的。誘導物能量源（或誘導物）用來重構Cpf1。在一些實施方式中，誘導物能量源藉由誘導型二聚體的兩個半部的作用使得Cpf1的兩個部分合在一起。因此在存在誘導物能量源的條件下誘導型二聚體的兩個半部變得更強韌。在不存在誘導物能量源的情況下，二聚體的兩個半部將不形成為二聚體（進行二聚化）。

[0437] 因此，誘導型二聚體的兩個半部與誘導物能量源合作以二聚化二聚體。這進而藉由使得Cpf1的第一部分和第二部分合在一起來重構Cpf1。

[0438] CRISPR酶融合構建體各自包含拆分Cpf1的一部分。該等較佳的是經由接頭諸如在此所述的GlySer接頭融合至二聚體的兩個半部中的一個。二聚體的兩個半部可以是合在一起形成同源二聚體的基本上相同的兩個單體，或者它們可以是合在一起形成異源二聚體的不同單體。這樣，兩個單體可以被認為是全長二聚體的一個半部。

[0439] Cpf1係拆分的，在某種意義上，Cpf1酶的兩個部分基本上包含有功能的Cpf1。Cpf1可以用作基因組編輯酶（當與靶DNA和指導序列形成複合物時），諸如切口酶或核酸酶（切割DNA的兩條股），或者該Cpf1可以是無效Cpf1，該無效Cpf1實質上是典型地由於其催化結構域中的一個或多個突變而具有非常小或沒有催化活性的DNA結合蛋白。

[0440] 拆分Cpf1的兩個部分可以被認為是拆分Cpf1的N'末端部分和C'末端部分。融合典型地是在Cpf1的拆分點處。換句話說，拆分Cpf1的N'末端部分的C'末端融合至一個二聚體半部，而C'末端部分的N'末端融合至另一個二聚體半部。

[0441] Cpf1不是必須被拆分，在某種意思上，斷裂係新創建的。拆分點典型地是經由電腦類比設計的並且選殖到構建體中。合起來，拆分Cpf1的兩個部分N'末端部分和C'末端部分形成全長Cpf1，該全長Cpf1包含較佳的是至少70%或更多的野生型胺基酸

（或編碼它們的核苷酸）、較佳的是至少80%或更多、較佳的是至少90%或更多、較佳的是至少95%或更多，並且最較佳的是至少99%或更多的野生型胺基酸（或編碼它們的核苷酸）。也許可以進行一些修整，並且設想到突變。非功能結構域可以全部被去除。重要的是兩個部分可以被結合在一起並且所希望的Cpf1功能被恢復或重構。

[0442] 二聚體可以是同源二聚體或異源二聚體。

[0443] 一個或多個，較佳的是兩個NLS可以用於可操作地連接至第一Cpf1構建體。一個或多個，較佳的是兩個NES可以用於可操作地連接至第一Cpf1構建體。NLS和/或NES較佳的是側接拆分二聚體（即半二聚體）融合物，即一個NLS可以被定位在第一Cpf1構建體的N'末端處並且一個NLS可以在第一Cpf1構建體的C'末端處。類似地，一個NES可以被定位在第二Cpf1構建體的N'末端處並且一個NES可以在第二Cpf1構建體的C'末端處。在提及N'末端或C'末端的情況下，應瞭解的是該等對應於相應核苷酸序列中的5'端和3'端。

[0444] 較佳的安排係，第一Cpf1構建體被安排為5'-NLS-(N'末端Cpf1部分)-接頭-(二聚體的第一半部)-NLS-3'。較佳的安排係，第二Cpf1構建體被安排為5'-NES-(二聚體的第二半部)-接頭-(C'末端Cpf1部分)-NES-3'。合適的啟動子較佳的是在該等構建體中的每個的上游。兩個構建體可以單獨或一起遞送。

[0445] 在一些實施方式中，可操作地連接至第二CPf1構建體的一個或多個NES中的一個或全部可以對換成NLS。然而，這典

型地可能不是較佳的並且在其他實施方式中，可操作地連接至第二Cpf1構建體的定位信號係一個或多個NES。

[0446] 還應瞭解的是，NES可以被可操作地連接至拆分Cpf1的N'末端片段並且NLS可以被可操作地連接至拆分Cpf1的C'末端片段。然而，可以較佳的安排係其中NLS被可操作地連接至拆分Cpf1的N'末端片段並且NES被可操作地連接至拆分Cpf1的C'末端片段。

[0447] NES用作將第二Cpf1融合構建體定位在核的外面，至少直到提供了誘導物能量源為止（例如，至少直到能量源被提供給誘導物以執行其功能為止）。誘導物的存在刺激了細胞質內兩個Cpf1融合物的二聚化並且使得其熱力學上值得用於將第一Cpf1融合物和第二Cpf1融合物二聚化成定位至核。在不受理論束縛的情況下，申請人相信NES將第二Cpf1融合物隔離至細胞質（即核的外面）。第一Cpf1融合物上的NLS將其定位至核。在兩種情況下，申請人使用NES或NLS來將平衡（核轉運的平衡）移動至所希望的方向。二聚化典型地發生在核的外面（非常少的部分可能發生在核中）並且二聚化複合物上的NLS將核轉運的平衡改變為核定位，所以二聚化的並因此重構的Cpf1進入核。

[0448] 有利地，申請人能夠重構拆分Cpf1的功能。使用暫態轉染來證明該構想並且二聚化發生在存在誘導物能量源的背景下。對於Cpf1的單獨片段，沒有看到活性。然後使用藉由慢病毒遞送的穩定表現來研究這一點並且顯示可以使用拆分Cpf1方法。

[0449] 本發明的拆分Cpf1方法係有益的，因為其使得Cpf1活

性係可誘導的，從而允許時間控制。此外，可以使用不同的定位序列（即，NES和NLS為較佳的）以降低來自自組裝複合物的背景活性。組織特異性啟動子，例如針對第一Cpf1融合構建體和第二Cpf1融合構建體中每個的組織特異性啟動子也可以用於組織特異性靶向，從而提供空間控制。如果需要，兩個不同的組織特異性啟動子可以用於產生更精細程度的控制。對於階段特異性啟動子可以使用相同方法，或者可以存在階段特異性啟動子和組織特異性啟動子的混合物，其中第一Cpf1融合構建體和第二Cpf1融合構建體中的一個在組織特異性啟動子的控制下（即可操作地連接至或包含該組織特異性啟動子），而第一Cpf1融合構建體和第二Cpf1融合構建體中的另一個在階段特異性啟動子的控制下（即可操作地連接至或包含該階段特異性啟動子）。

[0450] 誘導型Cpf1 CRISPR-Cas系統包含如在此所述的例如像可操作地連接至第一Cpf1融合構建體的一個或多個核定位序列（NLS）。理想的是，該等核定位序列具有足以驅動所述第一Cpf1融合構建體在真核細胞的核中以可檢測的量積累的強度。在不希望受到理論約束的情況下，據信核定位序列不是為真核生物中的Cpf1 CRISPR-Cas複合物活性所需要的，但是包括此類序列增強系統的活性，特別是對於靶向核中的核酸分子而言，並且有助於本發明的2-部分系統的操作。

[0451] 同樣地，第二Cpf1融合構建體被可操作地連接至核輸出序列（NES）。實際上，其可以連接至一個或多個核輸出序列。換句話說，與第二Cpf1融合構建體一起使用的輸出序列的數目較

佳的是1或2或3。典型地，2係較佳的，但是1是足夠的並且在一些實施方式中1係較佳的。NLS和NES的適合實例係本領域已知的。例如，較佳的核輸出信號（NES）係人類蛋白酪胺酸激酶2。較佳的信號將是物種特異性的。

[0452] 在使用FRB和FKBP系統的情況下，FKBP較佳的是側接核定位序列（NLS）。在使用FRB和FKBP系統的情況下，較佳的安排係N'末端Cpf1–FRB–NES : C'末端Cpf1-FKBP-NLS。因此，第一Cpf1融合構建體將包含C'末端Cpf1部分並且第二Cpf1融合構建體將包含N'末端Cpf1部分。

[0453] 本發明的另一有益方面係其可以迅速地開啟，即其具有快速應答。據信，在不受理論束縛的情況下，藉由現有（已經存在的）融合構建體（藉由與誘導物能量源接觸）可以比藉由新融合構建體的表現（尤其是翻譯）更迅速地誘導Cpf1活性。這樣，第一Cpf1融合構建體和第二Cpf1融合構建體可以提前即在需要Cpf1活性之前表現在靶細胞中。然後Cpf1活性可以進行時間控制並且然後藉由添加誘導物能量源迅速地進行重構，理想地，該重構比藉由例如載體遞送的Cpf1的表現（包括轉錄的誘導）作用更迅速（以二聚化異源二聚體並且從而提供Cpf1活性）。

[0454] 除非另外表明，否則術語Cpf1或Cpf1酶和CRISPR酶在此可互換地使用。

[0455] 申請人證明出CPf1可以被拆分為兩組分，該兩組分在重新合在一起時重構成功能核酸酶。採用雷帕黴素敏感的二聚化結構域，申請人產生了用於Cpf1介導的基因組編輯和轉錄調控

的時間控制的化學誘導型Cpf1。換言之，申請人證明出可以藉由將Cpf1拆分為兩片段來使得其為化學誘導型的並且證明出雷帕黴素敏感的二聚化結構域可以用於Cpf1的受控重組裝。申請人表明重組裝的Cpf1可以用於介導基因組編輯（藉由核酸酶/切口酶活性）以及轉錄調控（作為DNA結合結構域，所謂的"無效Cpf1"）。

[0456] 這樣，雷帕黴素敏感的二聚化結構域的使用係較佳的。Cpf1的重組裝係較佳的。重組裝可以藉由結合活性的恢復來確定。在Cpf1係切口酶或誘導雙股斷裂的情況下，在此描述了相比於野生型的適合比較百分比。

[0457] 雷帕黴素處理可以持續12天。劑量可以是200 nM。此時間劑量和/或莫耳劑量係用於人類胚腎293FT（HEK293FT）細胞系的適當劑量的一個實例並且此劑量也可以用於其他細胞系。對於體內治療性用途，此數位可以外推為例如mg/kg。然而，還設想，在此也使用用於向受試者給予雷帕黴素的標準劑量。關於"標準劑量"，其意指雷帕黴素的正常治療性用途或初期指示下的劑量（即當給予雷帕黴素以用於防止器官排斥時所用的劑量）。

[0458] 值得注意的是，Cpf1-FRB/FKBP片段的較佳的安排係分開的並且是失活的，直到FRB和FKBP的雷帕黴素誘導的二聚化使得功能性全長Cpf1核酸酶的重組裝產生為止。因此，較佳的是，附接至誘導型異源二聚體的第一半部的第一Cpf1融合構建體與附接至誘導型異源二聚體的第二半部的第二Cpf1融合構建體分開遞送和/或分開定位。

[0459] 為了隔離細胞質中的Cpf1(N)-FRB片段，在該片段不

太可能與核定位的Cpf1(C)-FKBP片段二聚化的情況下，較佳的是在Cpf1(N)-FRB上使用來自人類蛋白酪胺酸激酶2的單個核輸出序列（NES）（Cpf1(N)-FRB-NES）。在雷帕黴素存在下，Cpf1(N)-FRB-NES與Cpf1(C)-FKBP-2xNLS二聚化以重構完全的Cpf1蛋白，這使得核運輸（trafficking）的平衡朝核輸入移動並且允許DNA靶向。

[0460] 高劑量的Cpf1可以加劇表現出與指導股具有很少錯配的脫靶（OT）序列處的indel頻率。如果錯配係非連續的和/或在指導序列的種子區的外面，則此類序列係特別易感的。因此，Cpf1活性的時間控制可以用於減少長期表現實驗中的劑量並且因此與組成型活性Cpf1相比產生降低的脫靶indel。

[0461] 較佳的是病毒遞送。具體地說，設想了慢病毒或AAV遞送載體。申請人產生了類似於慢病毒CRISPR質粒的拆分的-Cpf1慢病毒構建體。拆分片段應該足夠小以適於AAV的~4.7kb大小限制。

[0462] 申請人證明出拆分Cpf1的穩定、低拷貝表現可以用於在靶向的座位處誘導大量indel，而在脫靶位點處沒有明顯的突變。申請人選殖了Cpf1片段（基於拆分5的2部分，在此所述的）。

[0463] 也可以使用包含VP64反式啟動結構域的無效Cpf1，例如該無效Cpf1添加到Cpf1(C)-FKBP-2xNLS中（無效-Cpf1(C)-FKBP-2xNLS-VP64）。該等片段重構成催化失活的Cpf1-VP64融合物（無效-Cpf1-VP64）。轉錄啟動係在雷帕黴素存在下藉由VP64誘導的，以便誘導Cpf1(C)-FKBP融合物與Cpf1(N)-FRB融合物的二聚化。換句話說，申請人測試了拆分的無效-Cpf1-VP64的可誘導

性並且顯示出轉錄啟動係在雷帕黴素存在下藉由拆分的無效-Cpf1-VP64誘導的。這樣，本發明的誘導型Cpf1可以與一個或多個功能結構域，諸如轉錄活化物或阻遏物或核酸酶（諸如Fok1）締合。功能結構域可以與拆分Cpf1的一個部分結合或融合。

[0464] 較佳的安排係第一Cpf1構建體被安排為5'-第一定位信號-(N'末端CPf1部分)-接頭-(二聚體的第一半部)-第一定位信號-3'並且第二Cpf1構建體被安排為5'-第二定位信號-(二聚體的第二半部)-接頭-(C'末端Cpf1部分)-第二定位信號-功能結構域-3'。在此，功能結構域置於第二Cpf1構建體的3'端處。可替代地，功能結構域可以置於第一Cpf1構建體的5'端處。一個或多個功能結構域可以使用在3'端或5'端處或兩個端處。合適的啟動子較佳的是在該等構建體的每個的上游。兩個構建體可以單獨或一起遞送。定位信號可以是NLS或NES，只要它們在每個構建體上不是相互混合的。

[0465] 在一個方面中，本發明提供了一種誘導型Cpf1 CRISPR-Cas系統，其中Cpf1與不具有至少一個突變的Cpf1酶相比具有至少97%或100%減弱的核酸酶活性。

[0466] 因此，還較佳的是，Cpf1係無效Cpf1。理想的是，拆分應該總是使得一個或多個催化結構域不受影響。對於無效-Cpf1，意圖在於發生DNA結合，但不顯示切割或切口酶活性。

[0467] 在一個方面中，本發明提供了一種如在此所討論的誘導型Cpf1 CRISPR-Cas系統，其中一個或多個功能結構域與Cpf1締合。此功能結構域可以與拆分Cpf1的一個部分或兩個部分締合(即結合或融合)。可以存在與拆分Cpf1的兩個部分中的每個締合的功

能結構域。因此該等功能結構域可以典型地被提供作為第一Cpf1融合構建體和/或第二Cpf1融合構建體的一部分，作為該構建體內的融合物。功能結構域典型地經由接頭諸如在此所討論的GlySer接頭來融合。一個或多個功能結構域可以是轉錄啟動結構域或阻遏結構域。儘管它們可以是不同的結構域，但是較佳的是，所有的功能結構域係活化物或阻遏物並且不使用兩者的混合物。

[0468] 轉錄啟動結構域可以包括VP64、p65、MyoD1、HSF1、RTA或SET7/9。

[0469] 在一個方面中，本發明提供了一種如在此所討論的誘導型Cpf1 CRISPR-Cas系統，其中與Cpf1締合的一個或多個功能結構域係轉錄阻遏結構域。

[0470] 在一個方面中，本發明提供了一種如在此所討論的誘導型Cpf1 CRISPR-Cas系統，其中轉錄阻遏結構域係KRAB結構域。

[0471] 在一個方面中，本發明提供了一種如在此所討論的誘導型Cpf1 CRISPR-Cas系統，其中轉錄阻遏結構域係NuE結構域、NcoR結構域、SID結構域或SID4X結構域。

[0472] 在一個方面中，本發明提供了一種如在此所討論的誘導型Cpf1 CRISPR-Cas系統，其中與轉接蛋白締合的一個或多個功能結構域具有一種或多種活性，包括修飾烷酶活性、去甲酶活性、轉錄啟動活性、轉錄阻遏活性、轉錄釋放因子活性、組蛋白修飾活性、RNA切割活性、DNA切割活性、DNA整合活性或核酸結合活性。

[0473] 在一些實施方式中，組蛋白修飾結構域也是較佳的。以下討論了示例性組蛋白修飾結構域。易位酶結構域、HR（同源重組）機構結構域、重組酶結構域、和/或整合酶結構域作為本發明的功能結構域也是較佳的。在一些實施方式中，DNA整合活性包括HR機構結構域、整合酶結構域、重組酶結構域和/或易位酶結構域。

[0474] 在一個方面中，本發明提供了一種如在此所討論的誘導型Cpf1 CRISPR-Cas系統，其中DNA切割活性歸因於核酸酶。

[0475] 在一個方面中，本發明提供了一種如在此所討論的誘導型Cpf1 CRISPR-Cas系統，其中核酸酶包括Fok1核酸酶。

[0476] 對於本發明的拆分Cpf1系統係較佳的此類功能結構域的用途也詳細討論在康爾曼（Konermann）等人（“使用工程化的CRISPR-Cas9複合物的基因組規模的轉錄啟動（Genome-scale transcriptional activation with an engineered CRISPR-Cas9 complex）”自然 2014年12月11日公開）中。

[0477] 本發明的系統可以與任何指導序列一起使用。

[0478] 在某些實施方式中可以使用修飾指導序列。特別較佳的是具體化以上提到的康爾曼 自然 2014年12月11日論文的傳授內容的指導序列。該等指導序列被修飾為使得蛋白結合RNA部分（諸如適配體）被添加。這樣的一個或多個部分可以替換指導序列的一部分。然後相應的RNA結合蛋白結構域可以用於識別RNA並且將功能結構域，諸如在此所述的那些募集至指導序列。此指

導序列主要是與無效-Cpf1一起使用，從而產生轉錄啟動或阻遏或藉由核酸酶諸如Fok1的DNA切割。此類指導序列與無效-Cpf1的組合使用係有力的，並且該組合的使用在Cpf1本身也與其自身功能結構域（如在此所討論的）締合的情況下是特別有力的。當根據本發明誘導無效-Cpf1（具有或不具有其自身締合的功能結構域）進行重構，即該無效-Cpf1係拆分Cpf1時，那麼該工具係特別有用的。

[0479] 也較佳的是用於本發明的指導RNA（gRNA）可以包含能夠與細胞中的感興趣的基因組座位中的靶序列雜交的指導序列，其中gRNA係藉由插入結合一種或多種轉接蛋白的一種或多種不同RNA序列來修飾的，並且其中該轉接蛋白與一個或多個功能結構域締合。Cpf1可以包含至少一個突變，以使得Cpf1酶具有不超過不具有該至少一個突變的Cpf1酶的核酸酶活性的5%的核酸酶活性；和/或至少一個或多個核定位序列。還提供了一種非天然存在或工程化的組成物，該組成物包含：包含能夠與細胞中的感興趣的基因組座位中的靶序列雜交的指導序列的一個或多個指導RNA（gRNA）、包含至少一個或多個核定位序列的Cpf1酶，其中CPf1酶包含至少一個突變，使得Cpf1酶具有不超過不具有該至少一個突變的Cpf1酶的核酸酶活性的5%的核酸酶活性，其中至少一個gRNA係藉由插入結合一種或多種轉接蛋白的一種或多種不同的RNA序列來修飾的，並且其中該轉接蛋白與一個或多個功能結構域締合。

[0480] gRNA較佳的是藉由插入結合一種或多種轉接蛋白的

一種或多種不同的RNA序列來修飾的。插入的結合一種或多種轉接蛋白的一種或多種不同的RNA序列較佳的是一種適配體序列或特異於相同或不同的一種或多種轉接蛋白的兩種或更多種適配體序列。轉接蛋白較佳的是包括：MS2、PP7、Qβ、F2、GA、fr、JP501、M12、R17、BZ13、JP34、JP500、KU1、M11、MX1、TW18、VK、SP、FI、ID2、NL95、TW19、AP205、ϕCb5、ϕCb8r、ϕCb12r、ϕCb23r、7s、PRR1。尤其穩定地表現拆分的無效-Cpf1的細胞系可以是有用的。

[0481] 申請人證明出Cpf1可以被拆分為兩個不同的片段，該等片段在使用化學品誘導重新合在一起時重構成功能性全長Cpf1核酸酶。拆分Cpf1體系結構將有用於多種應用。例如，拆分CPf1可以允許遺傳策略用於藉由將每個片段放在不同的組織特異性啟動子下來將Cpf1活性限制於交叉的細胞群體。另外，不同的化學誘導型二聚化結構域諸如APA和赤黴素也是可以採用的。

[0482] 誘導物能量源較佳的是化學品誘導。

[0483] 拆分部位或位置係Cpf1酶的第一部分與第二部分分開所在的那個點。在一些實施方式中，第一部分包含或編碼胺基酸1至X，而第二部分包含或編碼胺基酸X+1至末端。在此實例中，編碼係連續的，但這並非是總是需要的，因為胺基酸（或編碼它們的核苷酸）可以從拆分末端中的任一個末端開始修整，前提條件係保留了足夠的DNA結合活性和（如果需要）DNA切口酶或切割活性，例如與野生型Cpf1相比保留了至少40%、50%、60%、70%、80%、90%或95%活性。

[0484] 在此提供的示例性編碼可以參考野生型蛋白質，較佳的是野生型FnCpf1。然而，設想的是，可以使用野生型Cpf1諸如FnCpf1蛋白的突變體。編碼還可以不完全遵循FnCpf1編碼，因為例如可以使用一些N'或C'末端截短或缺失，但是這可以使用標準序列比對工具來解決。異種同源物作為序列比對工具也是較佳的。

[0485] 因此，拆分部位可以利用熟習該項技術者例如基於晶體資料和/或計算結構預測來選擇。

[0486] 例如，Cpf1核酸酶初級結構的計算分析揭示了三個不同的區（圖1）。第一係C末端RuvC樣結構域，其係僅功能表征的結構域。第二係N末端α-螺旋區並且第三係位於RuvC樣結構域與α-螺旋區之間的混合的α區和β區。預測非結構化區的若干小片段在Cpf1初始結構之內。不同的Cpf1異種同源物內的暴露於溶劑且不保守的非結構化區可以代表用於拆分的較佳的側面（圖2和圖3）。

[0487] 下表呈現出AsCpf1和LbCpf1內的非限制性潛在的拆分區。此區內的拆分位點可以是合適的。

拆分區	AsCpf1	LbCpf1
1	575-588	566-571
2	631-645	754-757
3	653-664	-
4	818-844	-

[0488] 對於Fn、As和Lb Cpf1突變體，應該很容易理解潛在的拆分位點的相應位置係例如基於序列比對的。對於非Fn、非As和非Lb酶，如果異種同源物與預期Cpf1之間存在相對較高的同源度，則可以使用異種同源物的晶體結構，或可以使用計算預測。

[0489] 理想的是，拆分部位應該位於區或環內。較佳的是，

拆分部位出現在胺基酸序列的中斷不引起結構特徵（例如，α螺旋或β片層）的部分或全部破壞的地方。結構化的區（未在晶體結構中顯現的區，因為該等區係未結構化的從而足以被"凍結"在晶體中）常常是較佳的選擇。申請人可以例如在Cpf1表面上暴露的未結構化區中進行拆分。

[0490] 申請人可以遵循作為較佳的實例及作為指導提供的以下程序。因為未結構化區未在結晶結構中顯現，所以申請人將晶體的周圍胺基酸序列與Cpf1的初級胺基酸序列進行相互參考。每個未結構化區可以由例如約3至10個胺基酸組成，不顯現在晶體中。因此申請人在該等胺基酸之間進行拆分。為了包括更多潛在的拆分側面，申請人包括了位於Cpf1的外面的環中的拆分，使用了與未結構化區相同的標準。

[0491] 在一些實施方式中，拆分部位係在Cpf1的外面環中。在其他較佳的實施方式中，拆分部位係在Cpf1的未結構化區中。未結構化區典型地是高柔性的外面環，該外面環的結構不能容易地由晶體圖案來確定。

[0492] 一旦拆分部位已被鑒定出，可以設計適合的構建體。

[0493] 典型地，NES被定位在拆分胺基酸的第一部分的N'末端（或編碼該部分的核苷酸的5'端）。在那樣的情況下，NLS被定位在拆分胺基酸的第二部分的C'末端（或編碼該部分的核苷酸的3'端）。以這種方式，第一Cpf1融合構建體可以被可操作地連接至一個或多個核輸出信號並且第二Cpf1融合構建體可以被可操作地連接至核定位信號。

[0494] 當然，可以提供相反的安排，其中NLS被定位在拆分胺基酸的第一部分的N'末端（或編碼該部分的核苷酸的5'端）。在那樣的情況下，NES被定位在拆分胺基酸的第二部分的C'末端（或編碼該部分的核苷酸的3'端）。因此，第一Cpf1融合構建體可以被可操作地連接至一個或多個核定位信號並且第二Cpf1融合構建體可以被可操作地連接至核輸出信號。

[0495] 使得兩個部分（拆分的任一側面）具有大致相同的長度的拆分對於包裝目的可能是有利的。例如，認為當轉錄物具有大約相同大小時維持兩個片段之間的化學計量係容易的。

[0496] 在某些實例中，人類密碼子優化的Cpf1諸如FnCpf1的N末端和C末端片段被分別融合至FRB和FKBP二聚化結構域。此安排可以是較佳的。它們可以被轉換（即N'末端融合至FKBP並且C'末端融合至FRB）。

[0497] 在此較佳的是使用接頭諸如$(GGGGS)_3$來將Cpf1片段與二聚化結構域分開。$(GGGGS)_3$係較佳的，因為它係相對長的接頭（15個胺基酸）。甘胺酸殘基係最柔性的並且絲胺酸殘基增加接頭處於蛋白質之外的機會。$(GGGGS)_6$ $(GGGGS)_9$或$(GGGGS)_{12}$可以較佳的是用作替代物。其他較佳的替代物係$(GGGGS)_1$、$(GGGGS)_2$、$(GGGGS)_4$、$(GGGGS)_5$、$(GGGGS)_7$、$(GGGGS)_8$、$(GGGGS)_{10}$、或$(GGGGS)_{11}$。

[0498] 例如，$(GGGGS)_3$可以包含在N'末端Cpf1片段與FRB之間。例如，$(GGGGS)_3$可以包含在FKB與C'末端Cpf1片段之間。

[0499] 替代性接頭係可用的，當高柔性接頭被認為作用最好，以使得Cpf1的2個部分合在一起並因此重構Cpf1活性的機會最大。一個替代方案係核質蛋白的NLS可以用作接頭。

[0500] 接頭也可以用在Cpf1與任何功能結構域之間。同樣，在此可以使用$(GGGGS)_3$接頭（或因此6、9或12個重複版本）或者可以將核質蛋白的NLS用作CPf1與功能結構域之間的接頭。

[0501] 設想了FRB/FKBP系統的替代物。例如ABA和赤黴素系統。

[0502] 因此，FKBP家族的較佳的實例係以下誘導型系統中的任一種。在FK506存在下與鈣調磷酸酶A(CNA)二聚化的FKBP；在FKCsA存在下與CyP-Fas二聚化的FKBP；在雷帕黴素存在下與FRB二聚化的FKBP；在庫馬黴素（Coumermycin）存在下與GryB二聚化的GyrB；在赤黴素（Gibberellin）存在下與GID1二聚化的GAI；或在HaXS存在下與HaloTag二聚化的Snap-tag。

[0503] FKBP家族本身內的替代物也是較佳的。例如，在FK1012存在下進行同源二聚化（即一個FKBP與另一個FKBP二聚化）的FKBP。因此，還提供了一種非天然存在或工程化的誘導型Cpf1 CRISPR-Cas系統，該系統包含：

附接至誘導型同源二聚體（homoodimer）的第一半部的第一Cpf1融合構建體，以及

附接至誘導型同源二聚體的第二半部的第二Cpf1融合構建體，

其中第一Cpf1融合構建體被可操作地連接至一個或多個核定

位信號，

其中第二Cpf1融合構建體被可操作地連接至（視情況一個或多個）核輸出信號，

其中與誘導物能量源的接觸使得誘導型同源二聚體的第一半部和第二半部合在一起，

其中使誘導型同源二聚體的第一半部和第二半部合在一起允許第一CPf1融合構建體和第二CPf1融合構建體組成功能性Cpf1 CRISPR-Cas系統，

其中Cpf1 CRISPR-Cas系統包含含有能夠與細胞中的感興趣的基因組座位中的靶序列雜交的指導序列的指導RNA（gRNA），並且

其中功能性Cpf1 CRISPR-Cas系統結合靶序列並且視情況編輯基因組座位以改變基因表現。

[0504] 在一個實施方式中，同源二聚體較佳的是FKBP並且誘導物能量源較佳的是FK1012。在另一個實施方式中，同源二聚體較佳的是GryB並且誘導物能量源較佳的是庫馬黴素。在另一個實施方式中，同源二聚體較佳的是ABA並且誘導物能量源較佳的是赤黴素。

[0505] 在其他實施方式中，二聚體係異源二聚物。異源二聚體的較佳的實例係以下誘導型系統中的任一種：在FK506存在下與鈣調磷酸酶A（CNA）二聚化的FKBP；在FKCsA存在下與CyP-Fas二聚化的FKBP；在雷帕黴素存在下，在庫馬黴素存在下，與

FRB二聚化的FKBP；在赤黴素存在下與GID1二聚化的GAI；在HaXS存在下與HaloTag二聚化的Snap-tag。

[0506] 申請人使用了FKBP/FRB，因為其係得到充分表徵的，並且兩個結構域係足以小（<100個胺基酸）至有助於包裝的。此外，雷帕黴素已使用較長時間並且副作用係充分瞭解的。大的二聚化結構域（>300 aa）也應該起作用但是可能需要較長的接頭來允許Cpf1重構進行。

[0507] 保羅穆魯甘（Paulmurugan）和甘比爾（Gambhir）（癌症研究（Cancer Res），2005年8月15日 65; 7413）討論了FRB/FKBP/雷帕黴素系統的背景。另一有用論文係克拉布特裡（Crabtree）等人（化學和生物學（Chemistry & Biology） 13, 99-107，2006年1月）。

[0508] 在一個實例中，構建了單一載體，一種表現盒（質粒）。gRNA係在U6啟動子的控制下。使用了兩種不同的Cpf1拆分。拆分Cpf1構建體基於的是側接NLS的第一Cpf1融合構建體，其中FKBP經由GlySer接頭融合至拆分CPf1的C末端部分；以及側接NES的第二CPf1融合構建體，其中FRB經由GlySer接頭與拆分CPf1的N末端部分融合。為了分開第一Cpf1融合構建體和第二Cpf1融合構建體，在轉錄上使用P2A拆分。拆分Cpf1在雷帕黴素存在下顯示出與野生型類似的插入缺失資訊，但是在雷帕黴素不存在下顯示出比野生型顯著更少的indel資訊。

[0509] 因此，提供了單一載體。該載體包含：

附接至誘導型二聚體的第一半部的第一Cpf1融合構建體，以及

附接至誘導型二聚體的第二半部的第二Cpf1融合構建體，

其中第一Cpf1融合構建體被可操作地連接至一個或多個核定位信號，

其中第二CPf1融合構建體被可操作地連接至一個或多個核輸出信號，

其中與誘導物能量源的接觸使得誘導型異源二聚體的第一半部和第二半部合在一起，

其中使誘導型異源二聚體的第一半部和第二半部合在一起允許第一CPf1融合構建體和第二CPf1融合構建體組成功能性Cpf1 CRISPR-Cas系統，

其中Cpf1 CRISPR-Cas系統包含含有能夠與細胞中的感興趣的基因組座位中的靶序列雜交的指導序列的指導RNA（gRNA），並且

其中功能性Cpf1 CRISPR-Cas系統結合靶序列並且視情況編輯基因組座位以改變基因表現。該等元件較佳的是被提供於單一構建體，例如一種表現盒上。

[0510] 第一Cpf1融合構建體較佳的是在每個末端側接至少一個核定位信號。第二CPf1融合構建體較佳的是在每個末端側接至少一個核輸出信號。

[0511] 還提供了一種治療有需要的受試者的方法，該方法包括藉由用編碼系統的多核苷酸或任一本發明載體轉化受試者來誘導基因編輯，並且向受試者給予誘導物能量源。還可以提供適合的修復模板，例如藉由包含所述修復模板的載體來遞送。

[0512] 還提供了一種治療有需要的受試者的方法，該方法包括藉由用編碼本發明的系統的多核苷酸或任一本發明載體轉化受試者來誘導轉錄啟動或阻遏，其中所述多核苷酸或載體編碼或包含催化失活的Cpf1和一個或多個締合的功能結構域；該方法進一步包括向受試者給予誘導物能量源。

[0513] 還提供了包含在所述治療方法中使用的本發明系統的組成物。還提供了本發明系統在用於此類治療方法的藥物的製造中的用途。

[0514] 在此或在在此引用的文獻中描述了由本發明系統可治療的病狀實例。

[0515] 單一載體可以包含轉錄物拆分劑，例如P2A。P2A將轉錄物拆分為兩部分，以分開第一CPf1融合構建體和第二CPf1融合構建體。拆分係歸因於“核糖體跳過（ribosomal skipping）”。實質上，核糖體在翻譯過程中跳過胺基酸，這使蛋白質股斷裂並且產生兩個分開的多肽/蛋白質。單一載體還有用於其中低背景活性不是所關注的但是需要高誘導活性的應用中。

[0516] 一個實例係選殖的胚胎幹細胞系的產生。正常程序係用編碼wt CPf1或Cpf1切口酶的質粒暫態轉染。該等質粒產生Cpf1

分子，該等Cpf1分子持續若干天保留活性並且具有較高的脫靶活性機會。使用單一表現載體用於拆分Cpf1使得將“高Cpf1活性”限制於較短的時間窗（例如，一劑量的誘導物，諸如雷帕黴素）。在不是連續（每日的）誘導物（例如雷帕黴素）治療的情況下，單一表現的拆分Cpf1載體的活性較低並且呈現出減小的產生不想要的脫靶效應的機會。

[0517] 在一些實施方式中誘導的Cpf1活性的峰值係有益的並且使用單一遞送載體可以更容易地發生，但是藉由雙重載體系統（每個載體遞送拆分CPf1的一個半部）也是可能的。峰值可以是高活性並且持續短時間尺度，典型地誘導物的壽命。

[0518] 因此，提供了一種用於產生選殖的胚胎幹細胞系之方法，該方法包括用編碼本發明的系統的多核苷酸或表現本發明的拆分Cpf1的一種本發明載體轉染一個或多個胚胎幹細胞並且給予本發明的誘導物能量源或使一個或多個幹細胞與本發明能量源接觸以誘導Cpf1的重構。可以提供修復模板。

[0519] 對於在此所述的所有方法，應瞭解的是，將需要適合的gRNA或指導序列。

[0520] 在功能結構域等與酶的一個或另一個部分“締合”的情況下，該等典型地是融合物。在此術語“與.....締合”係關於一個分子如何相對於另一個“締合的”，例如Cpf1的部分與功能結構域之間的締合使用的。在此類蛋白質-蛋白質相互作用的情況下，此締合可以按照抗體識別表位的方式進行的識別來觀察。可替代地，一種蛋白質可以與另一種蛋白質經由兩者的融合來締合，例如一

種亞基融合至另一種亞基。典型地藉由將一個蛋白質的胺基酸序列添加到另一個蛋白質的胺基酸序列上，例如經由將編碼每個蛋白質或亞基的核苷酸序列剪接在一起來進行融合。可替代地，這可以實質上視為兩個分子之間的結合或直接連接，諸如融合蛋白。在任何情況下，融合蛋白可以包含兩個感興趣的亞基之間（即酶與功能結構域之間或轉接蛋白與功能結構域之間）的接頭。因此，在一些實施方式中，CPf1的部分藉由結合功能結構域來與該功能結構域締合。在其他實施方式中，CPf1與功能結構域締合，因為兩者視情況經由中間接頭融合在一起。接頭的實例包括在此所討論的GlySer接頭。

[0521] 誘導物的其他實例包括光和激素類。對於光，誘導型二聚體可以是異源二聚體並且包含二聚體的第一光-誘導型半部和二聚體的第二（且互補的）光-誘導型半部。第一光-誘導型二聚體半部和第二光-誘導型二聚體半部的較佳的實例係CIB1和CRY2系統。CIB1結構域係光敏感隱花色素2（CRY2）的異源二聚體結合配偶體。

[0522] 在另一個實例中，藍光-響應的磁體二聚化系統（正磁體和反磁體）可以融合至拆分Cpf1蛋白的兩個部分。響應於光刺激，正磁體和反磁體進行二聚化並且Cpf1重組裝。例如，此類系統結合尼洪佳吉（Nihongaki）等人（自然生物技術 33, 755–790, 2015）中的Cas9進行描述。

[0523] 本發明包括的誘導物能量源可以是熱、超音波、電磁能或化學品。在本發明的一個較佳的實施方式中，誘導物能量源

可以是抗生素、小分子、激素、激素衍生物、類固醇或類固醇衍生物。在一個更較佳的實施方式中，誘導物能量源可以是脫落酸（abscisic acid）（ABA）、多西環素（doxycycline）（DOX）、cumate、雷帕黴素、4-羥基他莫昔芬（4-hydroxytamoxifen）（4OHT）、雌激素或蛻皮激素。本發明提供的是，至少一種開關可以選自下組，該組由下項組成：基於抗生素的誘導型系統、基於電磁能的誘導型系統、基於小分子的誘導型系統、基於核受體的誘導型系統以及基於激素的誘導型系統。在一個更較佳的實施方式中，至少一種開關可以選自下組，該組由下項組成：四環素（Tet）/DOX誘導型系統、光誘導型系統、ABA誘導型系統、cumate阻遏物/操縱子系統、4OHT/雌激素誘導型系統、基於蛻皮激素的誘導型系統以及FKBP12/FRAP（FKBP12-雷帕黴素複合物）誘導型系統。此類誘導物也在此以及在PCT/US2013/051418中被討論，該專利藉由引用結合在此。

[0524] 總的來說，可以由Cpf1，無論是wt、切口酶還是無效-Cpf1（具有或不具有締合的功能結構域），形成的任何用途可以使用本發明的拆分Cpf1方法實現。益處係保持Cpf1活性的可誘導性質。

[0525] 作為另一個實例，可以形成拆分CPf1與螢光蛋白如GFP的融合物。這將允許基因組座位的成像（參見"藉由優化的CRISPR/Cas系統使活人類細胞中的基因組座位動態成像（Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System）"陳·B等人 細胞 2013），但是以

誘導方式進行。這樣，在一些實施方式中，一個或多個Cpf1部分可以與螢光蛋白例如GFP締合（並且具體地是融合）。

[0526] 另外的實驗研究了當中靶切割處於相同水平時，野生型（wt）和拆分Cpf1之間的脫靶切割是否存在差異。為了進行此舉，申請人使用wt和拆分Cpf1質粒的暫態轉染並且在不同時間點進行收穫。在發現一組其中中靶切割在+/-5%之內的樣品之後，申請人尋找脫靶啟動。申請人在沒有指導序列（使用慢病毒）的情況下使細胞系具有wt或拆分Cpf1的穩定表現。在抗生素選擇之後，使用單獨的慢病毒遞送指導序列並且在不同的時間點進行收穫以測定中靶/脫靶切割。

[0527] 申請人將去穩定化序列（PEST，參見"mRNA-和蛋白質-去穩定化元件用於開發高度應答的報導系統之用途（Use of mRNA- and protein-destabilizing elements to develop a highly responsive reporter system）" 文·DC（Voon DC）等人 核酸研究 2005）引入到FRB(N)Cpf1-NES片段中以促進較快的降解並且因此促進拆分無效-Cpf1-VP64複合物穩定性的降低。

[0528] 在本說明書中如其他地方所述的此類去穩定化序列（包括PEST）與拆分Cpf1系統一起使用可能是有利的。

[0529] 產生了穩定地表現拆分的無效-Cpf1-VP64和MS2-p65-HSF1＋ 指導序列的細胞系。PLX抗性篩選可以證明不可逆的、時控轉錄啟動可以用於藥物篩選中。當拆分的無效-Cpf1-VP64係不可逆的時，此方法可能是有利的。

[0530] 在一個方面中，本發明提供了一種非天然存在或工程化的Cpf1 CRISPR-Cas系統，該系統可以包含至少一種開關，其中所述Cpf1 CRISPR-Cas系統的活性係藉由與關於該開關的至少一種誘導物能量源接觸來控制的。在本發明的一個實施方式中，關於至少一種開關的控制或所述Cpf1 CRISPR-Cas系統的活性可以得到啟動、增強、終止或阻遏。與至少一種誘導物能量源的接觸可以產生第一效應和第二效應。第一效應可以是下項中的一種或多種：核輸入、核輸出、次級組分（諸如效應分子）的募集、（蛋白質、DNA或RNA的）構象變化、切割、貨物（cargo）（諸如籠裝分子或輔助因子）的釋放、締合或解離。第二效應可以是下項中的一種或多種：關於至少一種開關的控制或所述Cpf1 CRISPR-Cas系統的活性的啟動、增強、終止或阻遏。在一個實施方式中，第一效應和第二效應可以級聯發生。

[0531] 在本發明的另一個方面中，Cpf1 CRISPR-Cas系統可以進一步包含至少一個或多個核定位信號（NLS）、核輸出信號（NES）、功能結構域、柔性接頭、突變、缺失、改變或截短。NLS、NES或功能結構域中的一個或多個可以是條件型地啟動或失活。在另一個實施方式中，突變可以是下項中的一種或多種：轉錄因子同源區中的突變、DNA結合結構域中的突變（諸如使鹼性螺旋環螺旋的鹼性殘基突變）、內源性NLS中的突變或內源性NES中的突變。本發明包括的誘導物能量源可以是熱、超音波、電磁能或化學品。在本發明的一個較佳的實施方式中，誘導物能量源可以是抗生素、小分子、激素、激素衍生物、類固醇或類固醇衍生物。在一個更較佳的實施方式中，誘導物能量源可以是脫落酸（ABA）、

多西環素（DOX）、cumate、雷帕黴素、4-羥基他莫昔芬（4OHT）、雌激素或蛻皮激素。本發明提供的是，至少一種開關可以選自下組，該組由下項組成：基於抗生素的誘導型系統、基於電磁能的誘導型系統、基於小分子的誘導型系統、基於核受體的誘導型系統以及基於激素的誘導型系統。在一個更較佳的實施方式中，至少一種開關可以選自下組，該組由下項組成：四環素（Tet）/DOX誘導型系統、光誘導型系統、ABA誘導型系統、cumate阻遏物/操縱子系統、4OHT/雌激素誘導型系統、基於蛻皮激素的誘導型系統以及FKBP12/FRAP（FKBP12-雷帕黴素複合物）誘導型系統。

[0532] 如本申請中詳述的控制的方面涉及至少一種或多種開關。如在此所用的術語“開關”係指組分系統或集合，其以協調方式起作用，以影響生物功能的變化，包括生物功能的所有方面，諸如該功能的啟動、阻遏、增強或終止。在一個方面中，術語開關涵蓋基因開關，該基因開關包括基因調節蛋白和該等蛋白質識別的特異性DNA序列的基本組分。在一個方面中，開關涉及在基因調節中使用的誘導和阻遏系統。總的來說，除非存在允許基因表現的一些分子（稱為誘導物），否則誘導型系統可以是關閉的。該分子被稱為“誘導表現”。此發生的方式依賴於控制機制以及細胞類型中的差異。阻遏系統係除了在一些分子（稱為輔阻遏物）的存在下之外，抑制基因表現的系統。該分子被稱為“阻遏表現”。這藉由其發生的方式依賴於控制機制以及細胞類型中的差異。如在此所用的術語“誘導型的”可以涵蓋開關的所有方面，與涉及的分子機制無關。因此，如由本發明包括的開關可以包括但不限於基於抗生素的誘導型系統、基於電磁能的誘導型系統、基於小分

子的誘導型系統、基於核受體的誘導型系統以及基於激素的誘導型系統。在較佳的實施方式中，開關可以是四環素（Tet）/DOX誘導型系統、光誘導型系統、脫落酸（ABA）誘導型系統、cumate阻遏物/操縱子系統、4OHT/雌激素誘導型系統、基於蛻皮激素的誘導型系統或FKBP12/FRAP（FKBP12-雷帕黴素複合物）誘導型系統。

[0533] 本發明的Cpf1 CRISPR-Cas系統可以被設計為以時間和空間精確方式調控或改變個別內源性基因的表現。Cpf1 CRISPR-Cas系統可以被設計為結合感興趣的基因的啟動子序列以改變基因表現。Cpf1可以被拆分為兩個半部，其中一個半部融合至隱花色素異源二聚體（隱花色素-2或CIB1）的一個半部，而剩餘的隱花色素配偶體融合至Cpf1的另一半部。在一些方面中，轉錄效應物結構域還可以包含在Cpf1 CRISPR-Cas系統中。效應物結構域可以是活化物，諸如VP16、VP64或p65，或阻遏物諸如KRAB、EnR或SID。在未受刺激狀態，一個半部的Cpf1-隱花色素2蛋白質定位至感興趣的基因的啟動子，但不結合CIB1-效應蛋白。在用藍光譜光刺激後，隱花色素-2被啟動，經歷構象變化，並且揭示出其結合結構域。CIB1進而結合隱花色素-2，使得Cpf1的第二半部定位至感興趣的基因的啟動子區並且開啟可以產生基因過表現或沈默的基因組編輯。LITE的方面進一步描述於劉，H等人，科學，2008和甘迺迪·M（Kennedy M）等人，自然方法（*Nature Methods*）2010中，該等文獻的內容藉由引用以其整體結合在此。

[0534] 可以進一步調控功能的活化物和阻遏物結構域可以

基於物種、強度、機制、持續時間、大小或任何數目的其他參數進行選擇。較佳的效應物結構域包括但不限於易位酶結構域、整合酶結構域、重組酶結構域、解離酶結構域、轉化酶結構域、蛋白酶結構域、DNA甲基轉移酶結構域、DNA脫甲基酶結構域、組蛋白乙醯化酶結構域、組蛋白脫乙醯化酶結構域、核酸酶結構域、阻遏物結構域、活化物結構域、核定位信號結構域、轉錄-蛋白質募集結構域、細胞攝取活性相關結構域、核酸結合結構域或抗體呈遞結構域。

[0535] 同樣產生化學品誘導型系統存在若干種不同的方式：1.由脫落酸（ABA）誘導的基於ABI-PYL的系統（例如參見，stke.sciencemag.org/cgi/content/abstract/sigtrans;4/164/rs2處的網站），2.由雷帕黴素（或基於雷帕黴素的相關化學品）誘導的基於FKBP-FRB的系統（例如參見，nature.com/nmeth/journal/v2/n6/full/nmeth763.html處的網站），3.由赤黴素（GA）誘導的基於GID1-GAI的系統（例如參見，nature.com/nchembio/journal/v8/n5/full/nchembio.922.html處的網站）。

[0536] 由本發明涵蓋的另一個系統係基於亞細胞定位中的變化的化學品誘導型系統。申請人還瞭解了一種被工程化為靶向感興趣的基因組座位的誘導型Cpf1 CRISPR-Cas系統，其中Cpf1酶被拆分為進一步連接至化學品或能量敏感蛋白質的不同部分的兩個融合構建體。在化學結合或能量轉移至化學品或能量敏感蛋白質之後，此化學品或能量敏感蛋白質使得CPf1酶的任一個半部的

亞細胞定位產生變化（即Cpf1酶的任一個半部從細胞質轉運到細胞的核中）。融合構建體從一個亞細胞區室或細胞器（在其中它的活性由於缺乏用於重構的Cpf1 CRISPR-Cas系統的底物而被隔離）到另一個亞細胞區室或細胞器（在其中存在底物）的這種轉運允許該等組分合在一起並且重構功能活性，並且然後允許該等組分與其所需底物（即哺乳動物核中的基因組DNA）接觸，並且產生靶基因表現的啟動或阻遏。

[0537] 考慮了其他誘導型系統，諸如但不限於藉由下項進行的調節：重金屬[梅奧·KE(Mayo KE)等人，細胞 1982, 29:99-108；瑟爾·PF(Searle PF)等人，分子細胞生物學(Mol Cell Biol) 1985, 5:1480-1489和布林斯特·RL(Brinster RL)等人，自然(倫敦) 1982, 296:39-42]、類固醇激素[海恩斯·NE(Hynes NE)等人，美國國家科學院院刊 1981, 78:2038-2042；克勞克·G(Klock G)等人，自然(倫敦) 1987, 329:734-736和李·F(Lee F)等人，自然(倫敦) 1981, 294:228-232。]、熱休克[努爾·L(Nouer L)：熱休克應答(Heat Shock Response)，波卡拉頓(Boca Raton)，FL: CRC; 1991]並且已經開發了其他試劑[瑪律利克·A(Mullick A)，馬西·B(Massie B)：轉錄、翻譯以及基因表現的控制(Transcription, translation and the control of gene expression)，在由斯皮爾·RE(Speir RE)編輯的細胞技術的百科全書(Encyclopedia of Cell Technology)中，威力公司(Wiley)；2000:1140-1164以及菲斯尼格·M(Fussenegger M)，生物技術進展(Biotechnol Prog) 2001, 17:1-51]。然而，對於該等誘導型哺乳動物啟動子存在局限性，諸如"關閉"狀態的"洩露"和誘導物(熱休克、重金屬、糖皮質激素等)的多效性。昆蟲激素

（蛻皮激素）的使用已在降低哺乳動物細胞中的細胞過程干擾的嘗試中提出[諾·D（No D）等人，美國國家科學院院刊 1996, 93:3346-3351]。另一種優良系統使用雷帕黴素作為誘導物[裡韋拉·VM（Rivera VM）等人，自然醫學 1996，2:1028-1032]，但雷帕黴素作為免疫抑制劑的作用係其體內使用的主要限制，並且因此需要發現用於控制基因表現的生物學惰性化合物[塞斯·E（Saez E）等人，美國國家科學院院刊 2000，97:14512-14517]。

[0538] 在特定實施方式中，在此所述的基因編輯系統處於密碼殺傷開關的控制下，該密碼殺傷開關（passcode kill switch）係當改變細胞條件時有效地殺傷宿主細胞的機制。這係藉由引入雜交體LacI-GalR家族轉錄因子來確保的，該等轉錄因子需要IPTG的存在以進行轉換（尚（Chan）等人 2015 自然 自然化學生物學（*Nature Chemical Biology*）doi:10.1038/nchembio.1979），並且可以用於驅動編碼對細胞存活關鍵的酶的基因。藉由將對不同化學品敏感的不同轉錄因子相結合，可以產生"代碼"。此系統可以用於在空間和時間上控制CRISPR誘導的遺傳修飾的程度，這在包括治療性應用的不同領域中可能是有意義的並且避免GMO從其預定的環境中"逃逸"也可能是有意義的。

自失活系統

[0539] 一旦細胞基因組中的基因的所有拷貝已被編輯，則該細胞中的連續的CRISRP/Cpf1表現不再需要。實際上，持續的表現在非預定基因組位點等處的脫靶效應情況下將是不希望的。因此，時間限制的表現將是有用的。誘導型表現提供了一種途徑，但是

此外，申請人設想了依賴於CRISPR載體本身內的非編碼指導靶序列的用途的自失活CRISPR-Cpf1系統。因此，在表現開始之後，CRISPR系統將使得其自身破壞，但是在完全破壞之前，其將有編輯靶基因的基因組拷貝的時間（在二倍體細胞中的正常點突變的情況下，其需要至多兩次編輯）。簡單地，自失活CRISPR-Cas系統包括附加RNA（即指導RNA），該附加RNA靶向CRISPR酶本身的編碼序列或靶向與存在於以下項中的一種或多種中的獨特序列互補的一種或多種非編碼指導靶序列：

[0540] (a) 在驅動非編碼RNA元件的表現的啟動子之內，

[0541] (b) 在驅動Cpf1基因的表現的啟動子之內，

[0542] (c) 在Cpf1編碼序列中的ATG翻譯起始密碼子的100 bp之內，

[0543] (d) 在病毒遞送載體，例如在AAV基因組中的反向末端重複序列（iTR）之內。

[0544] 此外，RNA可以經由載體，例如單獨的載體或編碼CRISPR複合物的同一載體來遞送。當藉由單獨的載體來提供時，靶向Cpf1表現的CRISPR RNA可以依序或同時給予。當依序給予時，在意圖用於例如基因編輯或基因工程化的CRISPR RNA之後，將遞送靶向Cpf1表現的CRISPR RNA。此時間段可以是數分鐘（例如5分鐘、10分鐘、20分鐘、30分鐘、45分鐘、60分鐘）的時間。此時間段可以是數小時（例如2小時、4小時、6小時、8小時、12小時、24小時）的時間。此時間段可以是數天（例如2天、3天、4天、7

天）的時間。此時間段可以是數週（例如2週、3週、4週）的時間。此時間段可以是數月（例如2個月、4個月、8個月、12個月）的時間。此時間段可以是數年（例如2年、3年、4年）的時間。在此方式中，Cas酶與能夠與第一靶標諸如感興趣的一個基因組座位或多個基因組座位雜交的第一gRNA締合並且負責所希望的CRISPR-Cas系統的一種或多種功能（例如，基因工程化）；並且隨後Cpf1酶可以接著與能夠與包含至少一部分的Cpf1或CRISPR盒的序列雜交的第二gRNA締合。在該gRNA靶向編碼Cpf1蛋白的表現的序列的情況下，該酶受到阻礙並且系統發生自失活。以相同方式，經由如在此解釋的例如脂質體、脂轉染、奈米粒子、微泡施用的靶向Cpf1表現的CRISPR RNA可以依序或同時給予。類似地，自失活可以用於對用來靶向一個或多個靶標的一個或多個指導RNA進行失活。

[0545] 在一些方面中，提供了單一gRNA，該單一gRNA能夠與CRISPR酶起始密碼子下游的序列雜交，由此在一段時間後，存在CRISPR酶表現的喪失。在一些方面中，提供了一個或多個gRNA，該等gRNA能夠與編碼CRISPR-Cas系統的多核苷酸的一個或多個編碼或非編碼區雜交，由此在一段時間後，存在一種或多種、或在一些情況下全部的CRISPR-Cas系統的失活。在系統的一些方面中，並且不受理論限制，細胞可以包含多種CRISPR-Cas複合物，其中第一亞組的CRISPR複合物包含能夠靶向有待編輯的一個基因組座位或多個基因組座位的第一gRNA，並且第二亞組的CRISPR複合物包含能夠靶向編碼CRISPR-Cas系統的多核苷酸的至少一個第二gRNA，其中第一亞組的CRISPR複合物介導靶向的

一個基因組座位或多個基因組座位的編輯並且第二亞組的CRISPR複合物最終使CRISPR-Cas系統失活，從而使細胞中的進一步CRISPR-Cas表現失活。

[0546] 因此，本發明提供了一種包含用於遞送至真核細胞的一種或多種載體，其中一種或多種載體編碼：(i) CRISPR酶，更具體地是Cpf1；(ii) 能夠雜交至細胞中的靶序列的第一指導RNA；以及 (iii) 能夠雜交至編碼CRISPR酶的載體中的一個或多個靶序列的第二指導RNA，當在該細胞中表現時，第一指導RNA引導第一CRISPR複合物與該細胞中的靶序列的序列特異性結合；第二指導RNA引導第二CRISPR複合物與編碼CRISPR酶的載體中的靶序列的序列特異性結合；CRISPR複合物包含結合指導RNA的CRISPR酶，由此指導RNA可以與其靶序列雜交；並且第二CRISPR複合物使CRISPR-Cas系統失活以阻止細胞對CRISPR酶的連續表現。

[0547] 在此其他地方揭露了一種或多種載體、編碼的酶、指導序列等的另外特性。該系統可以編碼 (i) CRISPR酶，更具體地是Cpf1；(ii)包含能夠雜交至細胞中的第一靶序列的序列的第一gRNA，(iii) 能夠雜交至編碼CRISPR酶的載體的第二指導RNA。類似地，酶可以包含一個或多個NLS等。

[0548] 不同編碼序列（CRISPR酶、指導RNA）可以包含在單一載體上或多個載體上。例如，有可能編碼在一個載體上的酶和在另一個載體上的不同RNA序列，或者有可能編碼在一個載體上的酶和一個gRNA以及在另一個載體上的剩餘gRNA或任何其他

前突變。總的來說，使用總共一個或多個不同載體的系統係較佳的。

[0549] 在使用多種載體的情況下，有可能以不相等的數目遞送它們，並且理想的是，其中編碼第一指導RNA的載體相對於編碼第二指導RNA的載體是過量的，從而有助於延遲CRISPR系統的最終失活，直到基因組編輯具有了發生的機會為止。

[0550] 第一指導RNA可以靶向基因組內的感興趣的任何靶序列，如在此其他地方所述的。第二指導RNA靶向編碼CRISPR Cas9酶的載體內的序列，並且從而使來自該載體的酶的表現失活。因此，載體中的靶序列必須能夠使表現失活。適合的靶序列可以是例如在Cpf1編碼序列的翻譯起始密碼子附近或之內，在驅動非編碼RNA元件的表現的啟動子中的非編碼序列中，在驅動Cpf1基因的表現的啟動子之內，在Cpf1編碼序列中的ATG翻譯起始密碼子的100 bp之內，和/或在病毒遞送載體，例如AAV基因組中的反向末端重複序列(iTR)之內。靠近此區域的雙股斷裂可以誘導Cpf1編碼序列的移碼，使得蛋白質表現喪失。用於使指導RNA“自失活”的替代性靶序列將旨在編輯/失活為CRISPR-Cpf1系統的表現或為載體的穩定性所需要的調節區/序列。例如，如果Cpf1編碼序列的啟動子被破壞，那麼轉錄可以被抑制或阻止。類似地，如果載體包含用於複製、維持性或穩定性的序列，那麼有可能靶向該等序列。例如，在AAV載體中，有用的靶序列係在iTR之內。其他有用的供靶向的序列可以是啟動子序列、多聚腺苷酸化（polyadenlyation）位點等。

[0551] 此外,如果指導RNA以陣列格式表現,則同時靶向兩個啟動子的“自失活”指導RNA將使得間插核苷酸從CRISPR-Cas表現構建體內切除,有效地使得其完全失活。類似地,在指導RNA靶向兩個ITR,或同時靶向兩種或更多種其他CRISPR-Cas組分的情況下,發生間插核苷酸的切除。總的來說,如在此解釋的自失活係適用於CRISPR-Cpf1系統的,以便提供CRISPR-Cpf1的調節。例如,如在此解釋的自失活可以適用於如在此解釋的突變,例如擴增病症的CRISPR修復。作為此自失活的結果,CRISPR修復僅僅具有暫態活性。

[0552] 向“自失活”指導RNA的5'端添加非靶向核苷酸(例如1-10個核苷酸,較佳的是1-5個核苷酸)可以用於延遲其加工和/或修飾其效力以作為確保CRISPR-Cpf1停止之前的靶向的基因組座位處的編輯的手段。

[0553] 在自失活AAV-CRISPR-Cpf1系統的一個方面中,可以建立共表現感興趣的一種或多種gRNA靶向基因組序列(例如1-2、1-5、1-10、1-15、1-20、1-30)的質粒,其中靶向LbCpf1序列的“自失活”gRNA處於或靠近工程化的ATG起始位點(例如,在5個核苷酸之內、在15個核苷酸之內、在30個核苷酸之內、在50個核苷酸之內、在100個核苷酸之內)。U6啟動子區中的調節序列也可以用gRNA靶向。U6驅動的gRNA可以被設計為陣列格式,使得多個gRNA序列可以同時被釋放。當首先被遞送至靶組織/細胞(離開的細胞)時,gRNA開始積累,同時Cpf1水平在核中上升。Cpf1與介導CRISPR-Cpf1質粒的基因組編輯和自失活的所有gRNA複合。

[0554] 自失活CRISPR-Cpf1系統的一個方面係由1至4或更多個不同指導序列；例如高達約20或約30個指導序列以單獨或串聯的陣列格式的表現。每個單個自失活指導序列可以靶向不同的靶標。這樣可以從例如一個嵌合pol3轉錄物開始加工。可以使用Pol3啟動子諸如U6或H1啟動子。Pol2啟動子諸如在此所提到的那些。反向末端重複（iTR）序列可以側接Pol3啟動子-一個或多個gRNA-Pol2啟動子-Cpf1。

[0555] 嵌合串聯的陣列轉錄物的一個方面在於一種或多種指導序列編輯一個或多個靶標，而一個或多個自失活指導序列使CRISPR/Cpf1系統失活。因此，例如，用於修復擴增病症的所述CRISPR-Cpf1系統可以直接與在此所述的自失活CRISPR-Cpf1系統相結合。此系統可以例如具有針對供修復的靶區的兩個指導序列以及針對CRISPR-Cpf1的自失活的至少一個第三指導序列。參考申請案序號PCT/US2014/069897，題為"在核苷酸重複病症中使用Crispr-Cas系統的組成物和方法（Compositions And Methods Of Use Of Crispr-Cas Systems In Nucleotide Repeat Disorders）"，2014年12月12日以WO/2015/089351公開。

使用Cpf1的基因編輯或改變靶座位

[0556] 雙股斷裂或一條股的單股斷裂應該有利地足以靠近靶位置，由此使得校正發生。在一實施方式中，距離不超過50、100、200、300、350或400個核苷酸。雖然不希望受理論約束，但是據信斷裂應該足以靠近靶位置，由此使得斷裂在在端切除過程中遭受外切核酸酶介導的移除的區之內。如果靶位置與斷裂之間

的距離太大，則突變可能不被包含在端切除中，並且因此可能不被校正，因為模板核苷酸序列僅可以用於校正端切除區內的序列。

[0557] 在一個實施方式中，其中指導RNA和V型/VI型分子，具體地是Cpf1/C2c1/C2c2或其異種同源物或同源物，較佳的是Cpf1核酸酶誘導雙股斷裂，目的是為了誘導HDR-介導的校正，切割位點在離靶位置的0-200 bp之間（例如0至175、0至150、0至125、0至100、0至75、0至50、0至25、25至200、25至175、25至150、25至125、25至100、25至75、25至50、50至200、50至175、50至150、50至125、50至100、50至75、75至200、75至175、75至150、75至125、75至100 bp）。在一個實施方式中，切割位點係在離靶位置的0-100 bp之間（例如0至75、0至50、0至25、25至100、25至75、25至50、50至100、50至75或75至100 bp）。在另一個實施方式中，與Cpf1或其異種同源物或同源物複合的兩個或更多個指導RNA可以用於誘導多重斷裂，目的是為了誘導HDR介導的校正。

[0558] 同源臂應該至少延伸至其中可以發生端切除的區，例如為了允許切除的單股突出端查找供體模板內的互補區。總體長度可能受參數諸如質粒大小或病毒包裝限制所限制。在一個實施方式中，同源臂可以不延伸到重複元件中。示例性同源臂長度包括至少50、100、250、500、750或1000個核苷酸。

[0559] 如在此所用的，靶位置係指靶核酸或靶基因（例如染色體）上的藉由V型/VI型，具體地是Cpf1/C2c1/C2c2或其異種同源物或同源物，較佳的是Cpf1分子依賴性過程修飾的位點。例如，靶位置可以是進行靶核酸的修飾Cpf1分子切割和模板核酸引導的

修飾例如校正的靶位置。在一實施方式中，靶位置可以是處於靶核酸上的兩個核苷酸例如相鄰核苷酸之間、向其中添加一個或多個核苷酸的的位點。靶位置可以包含藉由模板核酸改變例如校正的一個或多個核苷酸。在一實施方式中，靶位置處於靶序列（例如指導RNA所結合的序列）之內。在一實施方式中，靶位置處於靶序列（例如指導RNA所結合的序列）的上游或下游。

[0560] 如在此所用的術語模板核酸係指可以與V型/VI型分子，具體地是Cpf1/C2c1/C2c2或其異種同源物或同源物，較佳的是Cpf1分子和指導RNA分子結合使用來改變靶位置的結構的核酸序列。在一個實施方式中，靶核酸被修飾為典型地是在或靠近一個或多個切割位點處具有模板核酸的序列的一部分或全部。在一實施方式中，模板核酸係單股的。在一替代性實施方式中，模板核酸係雙股的。在一實施方式中，模板核酸係DNA，例如雙股DNA。在一替代性實施方式中，模板核酸係單股DNA。

[0561] 在一實施方式中，模板核酸藉由參與同源重組改變了靶位置的結構。在一實施方式中，模板核酸改變了靶位置的序列。在一實施方式中，模板核酸使得修飾的或非天然發生的鹼基摻入到靶核酸中。

[0562] 模板序列可以與靶序列一起經歷斷裂介導或催化的重組。在一實施方式中，模板核酸可以包含對應於靶序列上的被Cpf1介導的切割事件切割的位點的序列。在一實施方式中，模板核酸可以包含對應於兩者，即靶序列上的在第一Cpf1介導的事件中被切割的第一位點和靶序列上的在第二Cpf1介導的事件中被切

割的第二位點的序列。

[0563] 在某些實施方式中，模板核酸可以包含使得翻譯序列的編碼序列中發生改變的序列，例如使得蛋白質產物中的一個胺基酸取代另一個胺基酸，例如使得突變型對偶基因轉化為野生型對偶基因，野生型對偶基因轉化為突變型對偶基因，和/或引入終止密碼子、插入胺基酸胺基、缺失胺基酸殘基，或使得無意義突變發生的序列。在某些實施方式中，模板核酸可以包含使得非編碼序列中發生改變，例如外顯子中或5'或3'非翻譯區或非轉錄區中發生改變的序列。此類改變包括控制元件，例如啟動子、增強子中的改變，以及順式作用或反式作用控制元件中的改變。

[0564] 可以使用與靶基因中的靶位置具有同源性的模板核酸以改變靶序列的結構。模板序列可以用於改變不想要的結構，例如不想要的或突變的核苷酸。模板核酸可以包含當整合時產生下項的序列：降低正控制元件的活性；增加正控制元件的活性；降低負控制元件的活性；增加負控制元件的活性；減少基因的表現；增加基因的表現；增加對病症或疾病的抗性；增加對病毒進入的抗性；校正突變或改變不想要的胺基酸殘基，賦予、增加、廢除或減少基因產物的生物特性，例如增加酶的酶活性，或增加基因產物與另一個分子相互作用的能力。

[0565] 模板核酸可以包含使得靶序列的1、2、3、4、5、6、7、8、9、10、11、12或更多個核苷酸的序列中發生變化的序列。在一個實施方式中，模板核酸的長度可以是20+/-10、30+/-10、40+/-10、50+/-10、60+/-10、70+/-10、80+/-10、90+/-10、100+/-10、110+/-

10、120+/-10、130+/-10、140+/-10、150+/-10、160+/-10、170+/-10、180+/-10、190+/-10、200+/-10、210+/-10或220+/-10個核苷酸。在一實施方式中，模板核酸的長度可以是30+/-20、40+/-20、50+/-20、60+/-20、70+/-20、80+/-20、90+/-20、100+/-20、110+/-20、120+/-20、130+/-20、140+/-20、150+/-20、160+/-20、170+/-20、180+/-20、190+/-20、200+/-20、210+/-20或220+/-20個核苷酸。在一實施方式中，模板核酸的長度係10至1,000、20至900、30至800、40至700、50至600、50至500、50至400、50至300、50至200或50至100個核苷酸。

[0566] 模板核酸包含以下組分：[5'同源臂]-[替換序列]-[3'同源臂]。同源臂提供了染色體中的重組，因此用替換序列替換了不希望的元件，例如突變或特徵。在一實施方式中，同源臂側接最遠的切割位點。在一實施方式中，5'同源臂的3'端係靠近替換序列的5'端的位置。在一實施方式中，5'同源臂可以從替換序列的5'端延伸至少10、20、30、40、50、100、200、300、400、500、600、700、800、900、1000、1500或2000個核苷酸5'。在一實施方式中，3'同源臂的5'端係靠近替換序列的3'端的位置。在一實施方式中，3'同源臂可以從替換序列的3'端延伸至少10、20、30、40、50、100、200、300、400、500、600、700、800、900、1000、1500或2000個核苷酸3'。

[0567] 在某些實施方式中，一個或兩個同源臂可以被縮短以避免包括某些序列重複元件。例如，5'同源臂可以被縮短以避免序列重複元件。在其他實施方式中，3'同源臂可以被縮短以避免序列

重複元件。在一些實施方式中，5'同源臂和3'同源臂兩者都可以被縮短以避免包括某些序列重複元件。

[0568] 在某些實施方式中，用於校正突變的模板核酸可以被設計為用作單股的寡核苷酸。當使用單股的寡核苷酸時，5'同源臂和3'同源臂的長度可以在高達約200個鹼基對（bp）的範圍內，例如長度為至少25、50、75、100、125、150、175或200 bp。

Cpf1效應蛋白複合物系統促進的非同源末端連接

[0569] 在某些實施方式中，核酸酶誘導的非同源末端連接（NHEJ）可以用於靶基因特異性敲除。核酸酶誘導的NHEJ還可以用於去除（例如缺失）感興趣的基因中的序列。總體上，NHEJ藉由使兩個末端連接在一起來修復DNA的雙股斷裂；然而，總體上，只要兩個相容末端在恰好它們藉由雙鍵斷裂形成時被完美連接，原始序列就被恢復。在末端重新連接之前，雙鍵斷裂的DNA末端常常是酶加工的受試者，在一條或兩條股處產生核苷酸的添加或去除。這使得NHEJ修復位點處的DNA序列中存在插入和/或缺失（indel）突變。該等突變中的三分之二典型地改變閱讀框並且因此產生非功能蛋白。另外，維持閱讀框但插入或缺失大量的序列的突變可以破壞蛋白質的功能性。這係座位依賴性的，因為關鍵功能結構域中的突變可能比蛋白質的非關鍵區中的突變耐受性低。由NHEJ產生的indel突變在性質上是不可預測的；然而，在給定的斷裂位點處，某些indel序列係有利的並且是以群體來過度表示的，很可能是由於小的微同源區。缺失的長度可以廣泛地變化；最常見是在1-50 bp範圍中，但是它們可以輕易大於50 bp，例如它們可

以輕易達到大於約100-200 bp。插入往往是較短的並且常常包含緊密圍繞斷裂位點的序列的短的重複。然而，有可能獲得大的插入，並且在該等情況下，插入的序列常常被跟蹤至基因組的其他區域或跟蹤至細胞中存在的質粒DNA。

[0570] 因為NHEJ係誘變的方法，所以其還可以用於缺失小序列模體，只要特異性最終序列的產生是不需要的。如果雙股斷裂被靶向靠近短的靶序列，則由NHEJ修復導致的缺失突變常常跨越並且因此去除不想要的核苷酸。對於較大的DNA區段的缺失，引入兩個雙股斷裂（序列的每側上一個雙股斷裂）可以在末端之間產生NHEJ，其中去除了整個間插序列。這兩個方法可以用於缺失特異性DNA序列；然而，NHEJ的易出錯的性質仍可能在修復位點產生indel突變。

[0571] 雙股切割的V型/VI型分子，具體地是Cpf1/C2c1/C2c2或其異種同源物或同源物，較佳的是Cpf1分子和單股或切口酶V型/VI型分子，具體地是Cpf1/C2c1/C2c2或其異種同源物或同源物，較佳的是Cpf1分子兩種均可以用於在此所述的方法和組成物中以產生NHEJ介導的indel。靶向基因，例如編碼區，例如感興趣基因的早期編碼區的NHEJ介導的indel可以用於敲除感興趣的基因（即消除該感興趣基因的表現）。例如，感興趣基因的早期編碼區包含緊跟著轉錄起始位點的序列，在編碼序列的第一外顯子內，或在轉錄起始位點的500 bp內（例如，小於500、450、400、350、300、250、200、150、100或50 bp）。

[0572] 在一個實施方式中，其中指導RNA和V型/VI型分子，

具體地是Cpf1/C2c1/C2c2或其異種同源物或同源物，較佳的是Cpf1核酸酶產生了雙股斷裂，目的是為了誘導NHEJ介導的indel，指導RNA可以被構造成用於將一個雙股斷裂定位成緊密接近靶位置的核苷酸。在一實施方式中，切割位點可以是在離靶位置的0-500 bp之間（例如，離靶位置少於500、400、300、200、100、50、40、30、25、20、15、10、9、8、7、6、5、4、3、2或1 bp）。

[0573] 在一實施方式中，其中與V型/VI型分子，具體地是Cpf1/C2c1/C2c2或其異種同源物或同源物，較佳的是Cpf1切口酶複合的兩個指導RNA誘導了兩個單股斷裂，目的是為了誘導NHEJ介導的indel，兩個指導RNA可以被構造成用於將兩個單股斷裂定位成向靶位置的核苷酸提供NHEJ修復。

Cpf1效應蛋白複合物可以遞送功能效應物

[0574] 與藉由使DNA水平上的基因突變來永久性消除表現的CRISPR-Cas介導的基因敲除不同，CRISPR-Cas敲低（knockdown）允許藉由使用人工轉錄因素來暫時減少基因表現。使Cpf1蛋白諸如FnCpf1蛋白的兩個DNA切割結構域中的關鍵殘基突變（例如，FnCpf1蛋白的D917A和H1006A突變或根據AsCpf1蛋白的D908A、E993A、D1263A或根據LbCpf1蛋白的D832A、E925A、D947A或D1180A）使得催化失活的Cpf1產生。催化失活的Cpf1與指導RNA複合並且定位至由指導RNA的靶向結構域所指定的DNA序列，然而該Cpf1不切割靶DNA。失活的Cpf1蛋白諸如FnCpf1蛋白（例如D917A和H1006A突變）與效應物結構域例如轉錄阻遏結構域的融合能夠將效應物募集至由指導RNA所指定的任何DNA位點。在某

些實施方式中，Cpf1可以被融合至轉錄阻遏結構域並且被募集至基因的啟動子區。特別是對於基因阻遏，在此預期的是，阻斷內源性轉錄因子的結合位點將有助於下調基因表現。在另一個實施方式中，失活的Cpf1可以融合至染色質修飾蛋白。改變染色質狀態可以使得靶基因表現減少。

[0575] 在一實施方式中，指導RNA分子可以被靶向已知的轉錄應答元件（例如，啟動子、增強子等）、已知的上游啟動序列，和/或疑似能夠控制靶DNA的表現的未知或已知功能的序列。

[0576] 在一些方法中，可以使靶多核苷酸失活以實施細胞中的表現的修飾。例如，在CRISPR複合物與細胞中的靶序列結合後，靶多核苷酸失活，這樣使得該序列不被轉錄，該編碼蛋白不被產生，或者該序列不會像野生型序列一樣起作用。例如，可以使蛋白質或微小RNA編碼序列失活，這樣使得該蛋白質不被產生。

[0577] 在某些實施方式中，CRISPR酶包含選自由D917A、E1006A和D1225A組成的組的一個或多個突變，並且/或者一個或多個突變係在CRISPR酶的RuvC結構域中或者是如在此所討論的其他方式的突變。在一些實施方式中，CRISPR酶在催化結構域中具有一個或多個突變，其中當轉錄時，同向重複序列形成單一莖環並且指導序列引導CRISPR複合物與靶序列的序列特異性結合，並且其中酶進一步包含功能結構域。在一些實施方式中，功能結構域係轉錄啟動結構域，較佳的是VP64。在一些實施方式中，功能結構域係轉錄阻遏結構域，較佳的是KRAB。在一些實施方式中，轉錄阻遏結構域係SID或SID的串聯體（例如SID4X）。在一些實施

方式中，功能結構域係表觀遺傳修飾結構域，以便提供表觀遺傳修飾酶。在一些實施方式中，功能結構域係啟動結構域，它可以是P65啟動結構域。

Cpf1效應蛋白複合物或其組分的遞送

[0578] 藉由本揭露和本領域知識，CRISPR-Cas系統，特別是在此所述的新型CRISPR系統或其組分或其核酸分子（包括例如HDR模板）或編碼或提供其組分的核酸分子可以藉由在此一般和詳細描述的遞送系統來遞送。

[0579] 載體遞送，例如質粒、病毒遞送：CRISPR酶，例如Cpf1和/或任一本發明RNA，例如指導RNA可以使用任何適合載體例如質粒或病毒載體諸如腺伴隨病毒（AAV）、慢病毒、腺病毒或其他病毒載體類型、或它們的組合來遞送。Cpf1和一個或多個指導RNA可以包裝到一種或多種載體例如質粒或病毒載體中。在一些實施方式中，載體例如質粒或病毒載體，例如，藉由肌肉注射遞送至感興趣的組織中，而有時遞送係經由靜脈內、經皮、鼻內、經口、黏膜或其他遞送方法進行的。此遞送可以經由單劑量或多劑量來進行。熟習該項技術者應理解的是，在此有待遞送的實際劑量可以在很大程度上取決於多種因素而變化，諸如載體選擇、靶細胞、生物體、或組織、有待治療的受試者的一般狀況、所尋求的轉化/修飾的程度、給藥途徑、給藥方式、所尋求的轉化/修飾的類型等。

[0580] 此劑型可以進一步含有，例如，載體（水、鹽水、乙醇、甘油、乳糖、蔗糖、磷酸鈣、明膠、葡聚糖、瓊脂、果膠、花

生油、芝麻油等等)、稀釋劑、藥學上可接受的載體（例如，磷酸鹽緩衝鹽水)、藥學上可接受的賦形劑、和/或本領域已知的其他化合物。該劑型可以進一步含有一種或多種藥學上可接受的鹽，例如像，無機酸鹽諸如鹽酸鹽、氫溴酸鹽、磷酸鹽、硫酸鹽等；以及有機酸鹽，諸如乙酸鹽、丙酸鹽、丙二酸鹽、苯甲酸鹽等。另外，在此也可以存在輔助物質，諸如潤濕劑或乳化劑、pH緩衝物質、凝膠或膠凝材料、調味劑、著色劑、微球體、聚合物、懸浮劑等。此外，也可以存在一種或多種其他常規藥用成分，諸如防腐劑、保濕劑、懸浮劑、表面活性劑、抗氧化劑、抗結劑、填充劑、螯合劑、包衣劑、化學穩定劑等，尤其是在該劑型係呈可重構形式時。適合的示例性成分包括微晶纖維素、羧甲基纖維素鈉、聚山梨酯80、苯乙醇、三氯三級丁醇、山梨酸鉀、抗壞血酸、二氧化硫、沒食子酸丙酯、對羥基苯甲酸酯、乙基香蘭素、甘油、苯酚、對氯酚、明膠、白蛋白以及它們的組合。藥學上可接受的賦形劑的徹底論述可獲自雷明頓藥物科學（REMINGTON'S PHARMACEUTICAL SCIENCES)（馬克出版公司，紐約 1991)，該文獻藉由引用結合在此。

[0581] 在在此的一個實施方式中，遞送係經由腺病毒進行的，其可以是含有至少1×10^5個腺病毒載體粒子（也稱為粒子單位，pu）的單次加強劑量。在在此的一實施方式中，該劑量較佳的是腺病毒載體的至少約1×10^6個粒子（例如，約1×10^6-1×10^{12}個粒子)，更較佳的是至少約1×10^7個粒子、更較佳的是至少約1×10^8個粒子（例如，約1×10^8-1×10^{11}個粒子或約1×10^8-1×10^{12}個粒子)，並且最較佳的是至少約1×10^0個粒子(例如約1×10^9-1×10^{10}

個粒子或約1×10^{9}-1×10^{12}個粒子），或甚至至少約1×10^{10}個粒子（例如，約1×10^{10}-1×10^{12}個粒子）。可替代地，該劑量包含不超過約1×10^{14}個粒子，較佳的是不超過約1×10^{13}個粒子，甚至更較佳的是不超過約1×10^{12}個粒子，甚至更較佳的是不超過約1×10^{11}個粒子，並且最較佳的是不超過約1×10^{10}個粒子（例如，不超過約1×10^{9}個粒子）。因此，該劑量可以含有單劑量的腺病毒載體，其具有例如約1×10^{6}粒子單位（pu）、約2×10^{6} pu、約4×10^{6} pu、約1×10^{7} pu、約2×10^{7} pu、約4×10^{7} pu、約1×10^{8} pu、約2×10^{8} pu、約4×10^{8} pu、約1×10^{9} pu、約2×10^{9} pu、約4×10^{9} pu、約1×10^{10} pu、約2×10^{10} pu、約4×10^{10} pu、約1×10^{11} pu、約2×10^{11} pu、約4×10^{11} pu、約1×10^{12} pu、約2×10^{12} pu或約4×10^{12} pu的腺病毒。參見，例如，在2013年6月4日授權的授予納貝爾（Nabel）等人的美國專利案號8,454,972 B2中的腺病毒載體；該專利藉由引用結合在此，以及在其第29欄第36-58行的劑量。在在此的一個實施方式中，腺病毒係經由多劑量來遞送的。

[0582] 在在此的一個實施方式中，該遞送係經由AAV進行的。用於針對人類的AAV的體內遞送的治療有效劑量被認為處於含有從約1×10^{10}至約1×10^{10}個功能AAV/ml溶液的從約20至約50 ml的鹽水溶液的範圍內。劑量可以被調整以便相對於任何副作用平衡治療益處。在在此的一個實施方式中，AAV劑量大致處於從約1×10^{5}至1×10^{50}個基因組AAV、從約1×10^{8}至1×10^{20}個基因組AAV、從約1×10^{10}至約1×10^{16}個基因組、或約1×10^{11}至約1×10^{16}個基因組AAV的濃度範圍內。人類劑量可以是約1×10^{13}個基因組AAV。此類濃度可以從約0.001 ml至約100 ml、約0.05至約50 ml、或約10

至約25 ml的載體溶液進行遞送。藉由建立劑量應答曲線的常規試驗，熟習該項技術者可以容易地確立其他有效劑量。參見，例如，2013年3月26日授權的授予哈加（Hajjar）等人的美國專利案號8,404,658 B2，在第27欄，第45-60行。

[0583] 在在此的一實施方式中，該遞送係經由質粒進行的。在此類的質粒組成物中，該劑量應該是足以引發應答的質粒的量。例如，在質粒組成物中的質粒DNA的適當量可以是從約0.1至約2 mg，或從約1 μg至約10 μg/70 kg個體。本發明的質粒大體上包含(i) 啟動子；(ii) 編碼CRISPR酶的序列，該序列可操作地連接至所述啟動子；(iii) 選擇標記物；(iv) 複製起點；以及 (v) 在 (ii) 的下游並可操作地連接至 (ii) 的轉錄終止子。質粒還可以編碼CRISPR複合物的RNA組分，但是該等組分中的一個或多個還可以被編碼在不同的載體上。

[0584] 在此的劑量係基於平均70 kg個體的。給藥頻率在醫學或獸醫學從業者（例如醫師、獸醫師)或本領域熟練的科學家的範圍之內。還應注意的是，實驗中使用的小鼠典型地是約20 g並且來自小鼠實驗的小鼠可以提高至70 kg的個體。

[0585] 用於在此提供的組成物的劑量包括用於重複給予或重複給藥的劑量。在特定實施方式中，在數週、數月或數年的時間內進行重複給予。可以進行合適的測定來獲得最佳劑量方案。重複的給藥可以允許較低劑量的使用，這可以有利地影響脫靶修飾。

[0586] 在一些實施方式中，本發明的RNA分子以脂質體或脂

轉染配製物等遞送並且可以是藉由熟習該項技術者已熟知的方法來製備。此類方法描述於例如美國專利案號5,593,972、5,589,466、以及5,580,859中，該等專利藉由引用結合在此。已開發了特別旨在增強並改進siRNA到哺乳動物細胞的遞送的遞送系統（例如，參見，沈（Shen）等人 FEBS快報（FEBS Let.） 2003, 539: 111-114；夏（Xia）等人，自然生物技術 2002, 20: 1006-1010；賴希（Reich）等人，分子視覺（Mol.Vision.）2003, 9: 210-216；索倫森（Sorensen）等人，分子生物學雜誌（J. Mol. Biol.） 2003, 327: 761-766；路易士（Lewis）等人，自然遺傳學（Nat. Gen.） 2002, 32: 107-108以及西梅奧尼（Simeoni）等人，核酸研究（NAR） 2003, 31, 11: 2717-2724）並且該等遞送系統可以適用於本發明。siRNA最近已成功用於抑制靈長類動物中的基因表現（參見，例如托倫蒂諾（Tolentino）等人，視網膜（Retina） 24(4):660，該文獻也可以適用於本發明）。

[0587] 實際上，RNA遞送係可用的體內遞送方法。有可能使用脂質體或奈米粒子將Cpf1和gRNA（以及例如HR修復模板）遞送至細胞中。因此，本發明的CRISPR酶諸如Cpf1的遞送和/或RNA的遞送可以呈RNA形式並且經由微泡、脂質體或一個粒子或多個粒子來進行。例如，Cpf1 mRNA和gRNA可以包裝到脂質體粒子中以進行體內遞送。脂質體轉染試劑諸如來自生命技術公司（Life Technologies）的lipofectamine和市場上的其他試劑可以有效地將RNA分子遞送至肝臟中。

[0588] RNA遞送手段還較佳的是包括經由粒子的RNA遞送（卓·S.（Cho, S.）、金伯格·M.（Goldberg, M.）、松·S.（Son, S.）、

許·Q.（Xu, Q.）、楊·F.（Yang, F.）、梅·Y.（Mei, Y.）、博加特廖夫·S.（Bogatyrev, S.）、朗格·R.（Langer, R.）和安德森·D.（Anderson, D.），用於將小干擾RNA遞送至內皮細胞的脂質樣奈米粒子（Lipid-like nanoparticles for small interfering RNA delivery to endothelial cells），先進功能材料（Advanced Functional Materials），19: 3112-3118，2010）或外來體（施羅德·A.（Schroeder, A.）、萊文斯·C.（Levins, C.）、科迪斯·C.（Cortez, C.）、朗格·R.和安德森·D.，用於siRNA遞送的基於脂質的奈米治療劑（Lipid-based nanotherapeutics for siRNA delivery），內科醫學雜誌（Journal of Internal Medicine），267:9-21，2010，PMID:20059641）。實際上，已經表明外來體在遞送siRNA中特別有用，其為與CRISPR系統有一些相似之處的系統。例如，艾爾安達盧西·S（El-Andaloussi S）等人（“外來體介導的體外和體內siRNA遞送”（“Exosome-mediated delivery of siRNA in vitro and in vivo.”），自然實驗手冊（Nat Protoc.）2012年12月；7(12):2112-26. doi: 10.1038/nprot.2012.131，電子版2012年11月15日）描述了外來體如何對於跨不同的生物障壁的藥物遞送係有希望的工具並且可以用於體外和體內遞送siRNA。其途徑在於藉由轉染一包含與肽配位基融合的外來體蛋白的表現載體產生靶向的外來體。然後將該等外來體純化並且由轉染的細胞上清液進行表徵，然後將RNA載入到外來體中。根據本發明的遞送或給藥可以使用外來體進行，特別是但不限於腦。維生素E（α-生育酚）可以與CRISPR Cas軛合並且與高密度脂蛋白（HDL）一起遞送至腦，例如以與烏諾（Uno）等人完成的用於將短干擾RNA（siRNA）遞送至腦的類似方式（人類基因治療（HUMAN GENE

THERAPY） 22:711–719（2011年6月））。經由用磷酸鹽緩衝鹽水（PBS）或游離TocsiBACE或Toc-siBACE/HDL充滿的並且與腦灌注套組3（Brain Infusion Kit 3）（Alzet）連接的微滲透壓泵（型號1007D；Alzet，庫栢蒂諾（Cupertino），加利福尼亞州（CA））灌注小鼠。將一腦灌注插管置於在正中線的前囪的後方約0.5 mm，用於灌注到背側第三腦室中。烏諾等人發現，藉由相同的ICV灌注方法，少至3 nmol的Toc-siRNA與HDL可以誘導相當程度的靶減少。在本發明中對於人類可以考慮類似劑量的軛合至α-生育酚並且與HDL共同給予靶向腦的CRISPR Cas，例如，可以考慮靶向腦的約3 nmol至約3 μmol的CRISPR Cas。鄒（Zou）等人（人類基因治療22:465-475（2011年4月））描述了靶向PKCγ的短髮夾RNA的慢病毒介導的遞送方法，其用於在大鼠脊髓中的體內基因沈默。鄒等人藉由鞘內導管給予約10 μl的具有1×10^9個轉導單位（TU）/ml的滴度的重組慢病毒。在本發明中對於人類可以考慮類似劑量的在靶向腦的慢病毒載體中表現的CRISPR Cas，例如，可以考慮靶向腦的在具有1×10^9個轉導單位（TU）/ml的滴度的慢病毒中的約10-50 ml的CRISPR Cas。

[0589] 可以例如藉由電穿孔轉染包含Cpf1和crRNA的預組裝的重組體CRISPR-Cpf1複合物，從而產生高突變率且不存在可檢測的脫靶突變。戶珥，J.K.（Hur, J.K.）等人，藉由電穿孔Cpf1核糖核蛋白在小鼠中進行的靶向誘變（Targeted mutagenesis in mice by electroporation of Cpf1 ribonucleoproteins），自然生物技術，2016年6月6日 doi: 10.1038/nbt.3596. [電子版先於印刷版]

[0590] 就腦的局部遞送而言，這可以藉由不同方式來實現。例如，可以例如藉由注射紋狀體內（intrastriatally）遞送材料。注射可以經由顱骨切開術立體定位地進行。

[0591] 增強NHEJ或HR效率也有助於遞送。較佳的是，藉由共表現末端加工酶諸如Trex2（杜米特拉切（Dumitrache）等人 遺傳學（Genetics），2011年8月；188(4): 787–797）來增強NHEJ效率。較佳的是，藉由暫態地抑制NHEJ機構諸如Ku70和Ku86來增加HR效率。HR效率還可以藉由共表現原核生物或真核生物同源重組酶諸如RecBCD、RecA來增加。

包裝和啟動子

[0592] 將本發明的Cpf1編碼核酸分子例如DNA包裝到載體例如病毒載體中以介導體內修飾的方式包括：

- 為了實現NHEJ介導的基因敲除：
- 單病毒載體：
- 含有兩個或更多個表現盒的載體：
- 啟動子-Cpf1編碼核酸分子-終止子
- 啟動子-gRNA1-終止子
- 啟動子-gRNA2-終止子
- 啟動子-gRNA（N）-終止子（一直到載體的大小限制）
- 雙病毒載體：

- 含有用於驅動Cpf1表現的一表現盒的載體1
- 啟動子-Cpf1編碼核酸分子-終止子
- 含有用於驅動一個或多個指導RNA表現的一個或多個表現盒的載體2
- 啟動子-gRNA1-終止子
- 啟動子-gRNA（N）-終止子（一直到載體的大小限制）
- 用於介導同源定向修復。
- 除了以上所述的單病毒載體和雙病毒載體途徑之外，另外的載體可以用於遞送同源定向修復模板。

[0593] 用於驅動Cpf1編碼核酸分子表現的啟動子可以包括：

—AAV ITR可以充當一啟動子：這對於消除另外的啟動子元件（可能在載體中佔用空間）的需要係有利的。空出來的另外的空間可以用於驅動另外的元件（gRNA等）的表現。另外，ITR活性係相對較弱的，因此可以用於降低由於Cpf1的過表現所致的潛在毒性。

—對於遍存表現，可以使用的啟動子包括：CMV、CAG、CBh、PGK、SV40、鐵蛋白重鏈或輕鏈等。

[0594] 對於腦或其他CNS表現，可以使用啟動子：用於所有神經元的突觸蛋白I（SynapsinI）、用於興奮性神經元的CaMKIIα、用於GABA能神經元的GAD67或GAD65或VGAT等。

[0595] 對於肝臟表現，可以使用白蛋白啟動子。

[0596] 對於肺表現，可以使用SP-B。

[0597] 對於內皮細胞，可以使用ICAM。

[0598] 對於造血細胞，可以使用IFNβ或CD45。

[0599] 對於成骨細胞，可以使用OG-2。

[0600] 用來驅動指導RNA的啟動子可以包括：

—Pol III啟動子諸如U6或H1

—使用Pol II啟動子和內含子盒來表現gRNA

腺伴隨病毒（AAV）

[0601] Cpf1和一個或多個指導RNA可以使用腺伴隨病毒（AAV）、慢病毒、腺病毒或其他質粒或病毒載體類型進行遞送，具體地說，使用來自以下文獻的配方和劑量：例如，美國專利案號8,454,972（針對腺病毒的配方、劑量）、8,404,658（針對AAV的配方、劑量）和5,846,946（針對DNA質粒的配方、劑量）以及來自臨床試驗和關於涉及慢病毒、AAV和腺病毒的臨床試驗的出版物。例如，對於AAV，給藥途徑、配方和劑量可以如美國專利案號8,454,972並且如涉及AAV的臨床試驗。對於腺病毒，給藥途徑、配方和劑量可以如美國專利案號8,404,658並且如涉及腺病毒的臨床試驗。對於質粒遞送，給藥途徑、配方和劑量可以如美國專利案號5,846,946並且如涉及質粒的臨床研究。劑量可以基於或外推為平均70 kg的個體（例如成人男性），並且可以針對患者、受試者、不同重量和物種的哺乳動物進行調整。給藥頻率在醫學或獸醫學從業者（例如醫師、獸醫師）的範圍之內，其取決於常規因素，包

括患者或受試者的年齡、性別、一般健康狀況、其他狀況以及著手解決的特定病狀或症狀。可以將病毒載體注射到感興趣的組織中。對於細胞類型特異性基因組修飾，Cpf1的表現可以由細胞類型特異性啟動子驅動。例如，肝臟特異性表現可以使用白蛋白啟動子，並且神經元特異性表現（例如靶向CNS病症）可以使用突觸蛋白I啟動子。

[0602] 就體內遞送而言，AAV相比於其他病毒載體係有利的，這係由於幾個原因：

低毒性（這可以歸因於純化方法不需要細胞粒子的可以啟動免疫應答的超速離心）以及

引起插入誘變的低概率，原因在於它未整合到宿主基因組中。

[0603] AAV具有4.5或4.75 Kb的包裝限制。這意味著Cpf1以及啟動子和轉錄終止子必須都配合在同一個病毒載體中。大於4.5或4.75 Kb的構建體將導致病毒產生的顯著降低。SpCas9係相當大的，該基因自身超過4.1 Kb，使其難於包裝到AAV中。因此本發明的實施方式包括利用更短的Cpf1同源物。

[0604] 關於AAV，AAV可以是AAV1、AAV2、AAV5或任何其組合。可以相對於有待被靶向的細胞來選擇AAV；例如，可以選擇用於靶向腦或神經元細胞的AAV血清型1、2、5或雜交體衣殼AAV1、AAV2、AAV5或其任何組合；並且可以選擇用於靶向心臟組織的AAV4。AAV8可用於遞送至肝臟。在此的啟動子和載體係單獨較佳的。關於該等細胞的某些AAV血清型的列表（參見，格

林姆·D（Grimm, D.）等人，病毒學雜誌（J. Virol.） 82: 5887-5911 (2008)）如下：

細胞系	AAV-1	AAV-2	AAV-3	AAV-4	AAV-5	AAV-6	AAV-8	AAV-9
Huh-7	13	100	2.5	0.0	0.1	10	0.7	0.0
HEK293	25	100	2.5	0.1	0.1	5	0.7	0.1
HeLa	3	100	2.0	0.1	6.7	1	0.2	0.1
HepG2	3	100	16.7	0.3	1.7	5	0.3	ND
Hep1A	20	100	0.2	1.0	0.1	1	0.2	0.0
911	17	100	11	0.2	0.1	17	0.1	ND
CHO	100	100	14	1.4	333	50	10	1.0
COS	33	100	33	3.3	5.0	14	2.0	0.5
MeWo	10	100	20	0.3	6.7	10	1.0	0.2
NIH3T3	10	100	2.9	2.9	0.3	10	0.3	ND
A549	14	100	20	ND	0.5	10	0.5	0.1
HT1180	20	100	10	0.1	0.3	33	0.5	0.1
單核細胞	1111	100	ND	ND	125	1429	ND	ND
未成熟 DC	2500	100	ND	ND	222	2857	ND	ND
成熟 DC	2222	100	ND	ND	333	3333	ND	ND

慢病毒

[0605] 慢病毒係複雜的反轉錄病毒，其具有在有絲分裂細胞和有絲分裂後細胞兩者中感染並表現其基因的能力。最為人熟知的慢病毒係人類免疫缺陷病毒（HIV），其利用其他病毒的包膜糖蛋白來靶向廣泛範圍的細胞類型。

[0606] 慢病毒可以如下製備。在選殖pCasES10（含有慢病毒轉移質粒骨架）之後，將處於低傳代數（p＝5）的HEK293FT接種在T-75燒瓶中，以在轉染之前的一天在具有10%胎牛血清而沒有抗生素的DMEM中50%匯合。在20小時之後，將培養基更換為OptiMEM（無血清）培養基，並且在4小時後進行轉染。將細胞用10 μg的慢病毒轉移質粒（pCasES10）和下列包裝質粒轉染：5 μg的pMD2.G（VSV-g假型）和7.5 μg的psPAX2（gag/pol/rev/tat）。在具有陽離子脂質遞送劑（50 μL的Lipofectamine 2000和100 μl的

Plus試劑）的4 mL OptiMEM中進行轉染。在6小時之後，將培養基更換為具有10%胎牛血清的無抗生素的DMEM。該等方法在細胞培養過程中使用血清，但是較佳的是無血清的方法。

[0607] 慢病毒可以如下純化。在48小時後收穫病毒上清液。首先清除上清液的碎片並藉由0.45 μm的低蛋白結合（PVDF）過濾器進行過濾。然後將它們在超速離心機中以24,000 rpm旋轉2小時。將病毒沈澱重新懸浮在50 μl的DMEM中，在4°C下過夜。然後將它們等分，並且立即在-80°C下冷凍。

[0608] 在另一個實施方式中，還考慮了基於馬傳染性貧血病毒（EIAV）的最小非靈長類動物慢病毒載體，特別是對於眼部基因治療而言（例如，參見巴拉岡（Balagaan），基因醫學雜誌（J Gene Med） 2006；8: 275–285）。在另一個實施方式中，還考慮了RetinoStat®，經由視網膜下注射遞送用於治療濕型年齡相關性黃斑變性的、表現血管生成抑制性蛋白（內皮抑素和血管抑素）的基於馬傳染性貧血病毒的慢病毒基因治療載體（例如，參見，賓利（Binley）等人，人類基因治療 23:980–991（2012年9月））並且此載體可以被修改用於本發明的CRISPR-Cas系統。

[0609] 在另一個實施方式中，自滅活慢病毒載體可以用於和/或適於本發明的CRISPR-Cas系統，該自滅活慢病毒載體具有靶向由HIV tat/rev共用的共有外顯子的siRNA、核仁定位TAR誘餌、和抗CCR5特異性錘頭狀核酶（例如，參見，迪吉斯托（DiGiusto）等人 (2010) 科學轉化醫學（Sci Transl Med） 2:36ra43）。可以收集最少2.5 × 10^6個CD34+細胞/每千克患者體重並且以2×10^6個細

胞/ml的密度在X-VIVO 15培養基（龍沙公司（Lonza））中預刺激16至20小時，該培養基含有2 μmol/L-穀胺醯胺、幹細胞因子（100 ng/ml）、Flt-3配位基（Flt-3L）（100 ng/ml）和促血小板生成素（10 ng/ml）（CellGenix公司)。可以用慢病毒以感染複數5在75-cm2的包被有纖連蛋白（25 mg/cm2）（重組人纖維連接片斷（RetroNectin），寶生物工程株式會社（Takara Bio Inc.））的組織培養瓶中轉導預刺激的細胞16至24小時。

[0610] 慢病毒載體已揭露於帕金森病的治療中，例如參見美國專利公開案號20120295960以及美國專利案號7303910和7351585。慢病毒載體還已揭露於眼部疾病的治療中，例如參見美國專利公開案號20060281180、20090007284、US20110117189；US20090017543；US20070054961、US20100317109。還已揭露了將慢病毒載體遞送至腦，例如參見美國專利公開案號20110293571；US20110293571、US20040013648、US20070025970、US20090111106和美國專利案號US7259015。

RNA遞送

[0611] RNA遞送：該CRISPR酶，例如Cpf1，和/或任一本發明的RNA，例如指導RNA，也可以RNA的形式遞送。可以使用體外轉錄產生Cpf1 mRNA。例如，可以使用含有下列元件的PCR盒來合成Cpf1 mRNA：來自β球蛋白-polyA尾（一串120個或更多個的腺嘌呤）的T7_啟動子-科紮克（kozak）序列（GCCACC）-Cpf1-3' UTR。該盒可以用於經由T7聚合酶的轉錄。也可以使用體外轉錄從含有T7_啟動子-GG-指導RNA序列的盒來轉錄指導RNA。

[0612] 為了增強表現並且降低可能的毒性，可以例如使用假-U或5-甲基-C將該CRISPR酶編碼序列和/或指導RNA修飾為包含一種或多種修飾核苷酸。

[0613] 目前，mRNA遞送方法係特別有希望用於肝臟遞送。

[0614] 關於RNA遞送的許多臨床工作已集中於RNAi或反義子上，但是該等系統可以適於遞送用於實施本發明的RNA。因此應該相應地理解下文關於RNAi等的參考。

粒子遞送系統和/或配製物：

[0615] 已知若干種類型的粒子遞送系統和/或配製物可用於不同範圍的生物醫學應用中。總的來說，粒子被限定為關於其轉運和特性以整體單位表現的小物體。根據直徑將粒子進一步分類。粗粒子覆蓋介於2,500與10,000奈米之間的範圍。細粒子的大小介於100與2,500奈米之間。超細粒子或奈米粒子的大小大體上介於1與100奈米之間。100-nm限制的基準係在於區分粒子與本體材料的新特性典型地出現在100 nm以下的臨界長度尺度下的事實。

[0616] 如在此所用，粒子遞送系統/配製物被限定為包含根據本發明的粒子的任何生物遞送系統/配製物。根據本發明的粒子係具有小於100微米（μm）的最大尺寸（例如，直徑）的任一實體。在一些實施方式中，本發明粒子具有小於10 μm的最大尺寸。在一些實施方式中，本發明粒子具有小於2000奈米（nm）的最大尺寸。在一些實施方式中，本發明粒子具有小於1000奈米（nm）的最大尺寸。在一些實施方式中，本發明粒子具有小於900 nm、800 nm、

700 nm、600 nm、500 nm、400 nm、300 nm、200 nm或100 nm的最大尺寸。典型地，本發明粒子具有500 nm或更小的最大尺寸（例如，直徑）。在一些實施方式中，本發明粒子具有250 nm或更小的最大尺寸（例如，直徑）。在一些實施方式中，本發明粒子具有200 nm或更小的最大尺寸（例如，直徑）。在一些實施方式中，本發明粒子具有150 nm或更小的最大尺寸（例如，直徑）。在一些實施方式中，本發明粒子具有100 nm或更小的最大尺寸（例如，直徑）。例如具有50 nm或更小的最大尺寸的較小粒子使用在本發明的一些實施方式中。在一些實施方式中，本發明粒子具有介於25 nm與200 nm之間的範圍內的最大尺寸。

[0617] 使用多種不同的技術進行粒子表徵（包括例如表徵形貌、尺寸等）。常見的技術係電子顯微術（TEM、SEM）、原子力顯微術（AFM）、動態光散射（DLS）、X-射線光電子光譜法（XPS）、粉末X-射線衍射（XRD）、傅裡葉變換紅外光譜（FTIR）、基質輔助的雷射解吸/電離-飛行時間質譜（MALDI-TOF）、紫外可見光譜、雙偏振干涉法和核磁共振（NMR）。可以對於天然粒子（即載入前）或在載入貨物（在此貨物係指例如CRISPR-Cas系統的一種或多種組分，例如CRISPR酶或mRNA或指導RNA或它們的任何組合，並且可以包括附加載體和/或賦形劑）之後進行表徵（尺寸測量）以為本發明的任何體外、離體和/或體內應用的遞送提供具有最佳尺寸的粒子。在某些較佳的實施方式中，粒子尺寸（例如直徑）表徵係基於使用動態光散射（DLS）的測量。關於粒子、其製備和使用方法以及其測量參考美國專利案號8,709,843；美國專利案號6,007,845；美國專利案號5,855,913；美國專利案號5,985,309；美

國專利案號5,543,158；以及詹姆斯·E.·達爾曼（James E. Dahlman）和卡曼·巴恩斯（Carmen Barnes）等人 自然奈米科技（Nature Nanotechnology）(2014)的出版物，2014年5月11日線上公開，doi:10.1038/nnano.2014.84。

[0618] 本發明範圍內的粒子遞送系統可以任何形式提供，包括但不限於固體、半固體、乳液或膠體粒子。這樣，在此所述的任一遞送系統，包括但不限於例如基於脂質的系統、脂質體、膠束、微泡、外來體或基因槍可以被提供作為本發明範圍內的粒子遞送系統。

粒子

[0619] 應瞭解的是，在適當情況下，在此關於粒子或奈米粒子的參考可以是可互換的。可以使用粒子或脂質包膜同時遞送CRISPR酶mRNA和指導RNA；例如，本發明的CRISPR酶和RNA，例如作為複合物可以經由如達爾曼等人、WO2015089419 A2和在此引用的文獻中的粒子諸如7C1遞送（例如參見，詹姆斯·E.·達爾曼和卡曼·巴恩斯等人 自然奈米科技 (2014) 2014年5月11日線上公開，doi:10.1038/nnano.2014.84），例如遞送粒子包含脂質或類脂質（lipidoid）和親水性聚合物，例如陽離子脂質和親水性聚合物，例如其中陽離子脂質包括1,2-二油醯基-3-三甲基銨-丙烷（DOTAP）或1,2-二十四醯基-*sn*-甘油基-3-磷酸膽鹼（DMPC）並且/或者其中親水性聚合物包括乙二醇或聚乙二醇（PEG）；並且/或者其中粒子進一步包含膽固醇（例如來自配方1 = DOTAP 100、DMPC 0、PEG 0、膽固醇0；配方編號2 = DOTAP 90、DMPC 0、PEG 10、膽固醇

0；配方編號3 = DOTAP 90、DMPC 0、PEG 5、膽固醇5的粒子），其中粒子使用有效、多步方法形成，其中首先將效應蛋白和RNA例如在室溫下例如以1：1莫耳比於例如無菌無核酸酶的1X PBS中混合在一起；並且單獨地，將適用於配方的DOTAP、DMPC、PEG和膽固醇溶解在醇，例如100%乙醇中；並且將兩種溶液混合在一起以形成含有複合物的粒子）。

[0620] 可以使用粒子或脂質包膜同時遞送核酸靶向效應蛋白（諸如型V蛋白諸如Cpf1）mRNA和指導RNA。適合粒子的實例包括但不限於在US 9,301,923中描述的那些。

[0621] 例如，蘇·X、弗裡克·J（Fricke J），卡瓦納·DG（Kavanagh DG），歐文·DJ（Irvine DJ）（“使用脂質包膜的pH響應性聚合物奈米粒子的體外和體內mRNA遞送（In vitro and in vivo mRNA delivery using lipid-enveloped pH-responsive polymer nanoparticles）” 分子藥劑學（Mol Pharm.）2011年6月6；8(3):774-87. doi: 10.1021/mp100390w，電子版2011年4月1日）描述了生物可降解的核-殼結構化的奈米粒子，其具有由磷脂雙層殼包膜的聚（β-胺基酯）（PBAE）核。該等被開發為用於體內mRNA遞送。該pH響應性PBAE組分被選擇為促進內體破壞，而該脂質表面層被選擇為將聚陽離子核的毒性降低到最低限度。因此，該等對於遞送本發明的RNA是較佳的。

[0622] 在一個實施方式中，考慮了基於自組裝生物黏附聚合物的粒子/奈米粒子，其可以適用於肽的經口遞送、肽的靜脈內遞送以及肽的鼻遞送，均遞送至腦。其他實施方式，還考慮了諸如

疏水性藥物的經口吸收和眼部遞送。分子包膜技術涉及被保護並遞送至疾病位點的工程化聚合物包膜(例如參見,馬薩,M.(Mazza, M.)等人 ACS奈米(ACSNano),2013. 7(2): 1016-1026;秀,A.(Siew, A.)等人 分子藥劑學,2012. 9(1):14-28;拉拉特薩,A.(Lalatsa, A.)等人 控釋雜誌(J Contr Rel),2012. 161(2):523-36;拉拉特薩,A.等人,分子藥劑學,2012. 9(6):1665-80;拉拉特薩,A.等人 分子藥劑學,2012. 9(6):1764-74;加勒特,N.L.(Garrett, N.L.)等人 生物光電雜誌(J Biophotonics),2012. 5(5-6):458-68;加勒特,N.L.等人 拉曼光譜雜誌(J Raman Spect),2012. 43(5):681-688;艾哈邁德,S.(Ahmad, S.)等人 皇家學會介面雜誌(J Royal Soc Interface) 2010. 7:S423-33;烏切克布,I.F.(Uchegbu, I.F.)藥物遞送專家評論(Expert Opin Drug Deliv),2006. 3(5):629-40;曲,X.(Qu, X.)等人 生物大分子,2006.7(12):3452-9和烏切克布I.F.等人 國際藥學雜誌(Int J Pharm),2001. 224:185-199)。考慮了約5 mg/kg的劑量,呈單劑量或多劑量形式,這取決於靶組織。

[0623] 在一個實施方式中,可以使用由丹·安德森實驗室(Dan Anderson's lab)在MIT開發的可以將RNA遞送至癌細胞以便使腫瘤生長停止的粒子/奈米粒子並且/或者使該等粒子/奈米粒子適於本發明的CRISPR Cas系統。具體地說,安德森實驗室開發了用於新生物材料和奈米配製物的合成、純化、表徵和配製的全自動化組合系統。例如參見,阿拉比(Alabi)等人,美國國家科學院院刊,2013年8月6號;110(32):12881-6;張等人,先進材料(Adv Mater.),2013年9月6日;25(33):4641-5;蔣(Jiang)等人,奈米快報(Nano Lett.),2013年3月13日;13(3):1059-64;卡拉吉安尼

斯（Karagiannis）等人，ACS奈米，2012年10月23日；6(10):8484-7；懷特海德（Whitehead）等人，ACS奈米，2012年8月28日；6(8):6922-9以及李等人，自然奈米技術，2012年6月3日；7(6):389-93。

[0624] 美國專利申請20110293703涉及類脂質化合物，該等化合物在多核苷酸的給藥中也是特別有用的，其可以適用於遞送本發明的CRISPR Cas系統。在一個方面中，胺基醇類脂質化合物與有待遞送至細胞或受試者的藥劑結合而形成微粒子、奈米粒子、脂質體、或膠束。有待藉由粒子、脂質體、或膠束遞送的藥劑可以呈氣體、液體、或固體的形式，並且該藥劑可以是一種多核苷酸、蛋白質、肽、或小分子。胺基醇類脂質化合物可以與其他胺基醇類脂質化合物、聚合物（合成的或天然的）、表面活性劑、膽固醇、碳水化合物、蛋白質、脂質等結合而形成該等粒子。然後該等粒子可以視情況與藥物賦形劑結合而形成藥物組成物。

[0625] 美國專利公開案號20110293703也提供了製備胺基醇類脂質化合物的方法。使胺的一種或多種等效物與環氧化物封端化合物的一種或多種等效物在適當條件下反應而形成本發明的胺基醇類脂質化合物。在某些實施方式中，胺的所有胺基基團與環氧化物封端化合物充分反應而形成三級胺。在其他實施方式中，胺的所有胺基基團未與環氧化物封端化合物完全反應形成三級胺，由此生成在胺基醇類脂質化合物中的一級胺或二級胺。該等一級胺或二級胺照原樣留下或者可以與另一種親電子劑諸如一種不同的環氧化物封端化合物反應。正如熟習該項技術者應瞭解的，胺

與未過量的環氧化物封端化合物反應將產生多種不同的具有不同數目的尾部的胺基醇類脂質化合物。某些胺類可以被兩個環氧化物衍生的化合物尾部完全功能化，而其他分子不會被環氧化物衍生的化合物尾部完全功能化。例如，二胺或多胺可以包括離開該分子的不同胺基部分的一個、二個、三個、或四個環氧化物衍生的化合物尾部，從而產生一級胺、二級胺和三級胺。在某些實施方式中，並不是所有胺基基團都被完全功能化。在某些實施方式中，使用了相同類型的環氧化物封端化合物中的兩種。在其他實施方式中，使用了兩種或更多種不同的環氧化物封端化合物。胺基醇類脂質化合物的合成係用或不用溶劑進行的，並且該合成可以在從30°C-100°C的範圍內，較佳的是在大約50°C-90°C的較高溫度下進行。視情況，可以將製備的胺基醇類脂質化合物純化。例如，可以純化胺基醇類脂質化合物的混合物以產生具有特定數目的環氧化物衍生的化合物尾部的胺基醇類脂質化合物。或者，該混合物可以被純化而產生特定的立體異構物或區域異構物。也可以使用鹵代烷（例如，碘甲烷）或其他烷化劑將該等胺基醇類脂質化合物烷化，並且/或者它們可以被醯化。

[0626] 美國專利公開案號20110293703還提供了藉由發明方法製備的胺基醇類脂質化合物的文庫。可以使用涉及液體處理器、機器人、微量滴定板、電腦等的高通量技術製備和/或篩選該等胺基醇類脂質化合物。在某些實施方式中，篩選了該等胺基醇類脂質化合物的將多核苷酸或其他藥劑（例如，蛋白質、肽、小分子）轉染到細胞中的能力。

[0627] 美國專利公開案號20130302401涉及已經使用組合聚合製備的一類聚(β-胺基醇)（PBAA）。該等發明的PBAA可以在生物技術和生物醫學應用中用作塗層（諸如用於醫療裝置或植入物的膜或多層膜的塗層)、添加劑、材料、賦形劑、生物防汙劑（non-biofouling agent）、微圖案化劑、以及細胞封裝劑。當用作表面塗層時，該等PBAA在體外和體內均引發不同水平的炎症，這取決於它們的化學結構。這類材料的巨大化學多樣性允許我們鑒定出在體外抑制巨噬細胞啟動的聚合物塗層。此外，在羧化聚苯乙烯微粒的皮下移植之後，該等塗層減少了炎症細胞的募集，並且減輕了纖維化。該等聚合物可以用於形成用於細胞封裝的聚合電解質複合物膠囊。本發明還可以具有許多其他的生物應用，諸如抗微生物塗層、DNA或siRNA遞送、以及幹細胞組織工程。美國專利公開案號20130302401的傳授內容可以適用於本發明的CRISPR Cas系統。在一些實施方式中，可以使用基於糖的粒子，例如如在此所述的GalNAc並且參考WO2014118272（藉由引用結合在此）以及耐爾，JK（Nair, JK）等人，2014，美國化學學會雜誌（Journal of the American Chemical Society） 136 (49), 16958-16961）以及在此的傳授內容，除非另外表明，否則特別涉及適用於所有粒子的遞送。

[0628] 在另一個實施方式中，考慮了脂質奈米粒子（LNP）。抗轉甲狀腺素蛋白小干擾RNA已被封裝在脂質奈米粒子中並且遞送至人類（例如參見，科爾賀（Coelho）等人，新英格蘭醫學雜

誌（N Engl J Med） 2013；369:819-29），並且此系統可以適於並應用於本發明的CRISPR Cas系統。考慮了靜脈內給予約0.01至約1 mg/kg體重的劑量。考慮了降低輸注相關反應的風險的藥物，諸如考慮到地塞米松（dexamethasone）、對乙醯胺基酚（acetampinophen）、苯海拉明（diphenhydramine）或西替利嗪（cetirizine）、以及雷尼替丁（ ranitidine）。還考慮了約0.3 mg/kg的多劑量，每4週一次，五個劑量。

[0629] LNP已經顯示在將siRNA遞送至肝臟中是高度有效的（例如參見塔韋內羅（Tabernero）等人，癌症發現（Cancer Discovery），2013年4月，第3卷，第4期，第363-470頁），並且因此被考慮用於將編碼CRISPR Cas的RNA遞送至肝臟。可以考慮6 mg/kg的LNP的約四個劑量的用量，每兩週一次。塔韋內羅等人證明，在以0.7 mg/kg給予LNP前2個週期之後，觀察到腫瘤消退，並且在6個週期結束之後，患者已經實現了部分應答，具有淋巴結轉移完全消退以及肝臟腫瘤的顯著萎縮。在此患者中給予40個劑量之後獲得完全應答，在接受經過26個月的劑量之後其保持緩解和完全治療。具有RCC和在用VEGF途徑抑制劑進行的在先治療之後進展的包括腎臟、肺以及淋巴結的肝外位點疾病的兩位患者在所有位點的疾病都保持穩定大約8至12個月，並且一位具有PNET和肝轉移的患者繼續在18個月（36個劑量）的延伸研究中保持疾病穩定。

[0630] 然而，必須將LNP的電荷考慮在內。當陽離子脂質與帶負電的脂質結合時，誘導促進細胞內遞送的非雙層結構。因為

帶電荷的LNP在靜脈內注射之後迅速從循環中清除，所以開發了具有低於7的pKa值的可電離陽離子脂質(例如參見，羅辛(Rosin)等人，分子治療（Molecular Therapy)，第19卷，第12期，第1286-2200頁，2011年12月)。帶負電荷的聚合物諸如RNA可以低pH值（例如，pH 4）載入到LNP中，在此pH時可電離脂質展示出正電荷。然而，在生理學pH值下，LNP表現出與更長的循環時間相容的低表面電荷。已經關注了四種可電離陽離子脂質，即1,2-二亞油醯基-3-二甲基銨-丙烷（DLinDAP)、1,2-二亞油基氧基-3-N,N-二甲基胺基丙烷（DLinDMA)、1,2-二亞油基氧基-酮基-N,N-二甲基-3-胺基丙烷(DLinKDMA)、以及1,2-二亞油基-4-(2-二甲基胺基乙基)-[1,3]-二氧戊環（DLinKC2-DMA)。已經表明，含有該等脂質的LNP siRNA系統在體內肝細胞中表現出顯著不同的基因沈默特性，具有根據採用因子VII基因沈默模型的DLinKC2-DMA>DLinKDMA>DLinDMA>>DLinDAP系列而變化的潛能（例如參見，羅辛等人，分子治療，第19卷，第12期，第1286-2200頁，2011年12月)。可以考慮1 μg/ml LNP或LNP中或與LNP相關聯的CRISPR-Cas RNA的劑量，尤其是對於含有DLinKC2-DMA的配製物而言。

[0631] LNP的製備和CRISPR Cas封裝可以使用和/或改編自羅辛等人，分子治療，第19卷，第12期，第1286-2200頁，2011年12月)。陽離子脂質1,2-二亞油醯基-3-二甲基銨-丙烷(DLinDAP)、1,2-二亞油基氧基-3-N,N-二甲基胺基丙烷（DLinDMA)、1,2-二亞油基氧基酮基-N,N-二甲基-3-胺基丙烷（DLinK-DMA)、1,2-二亞油基-4-(2-二甲基胺基乙基)-[1,3]-二氧戊環(DLinKC2-DMA)、(3-

o-[2"-(甲氧基聚乙二醇2000)琥珀醯]-1,2-二肉豆蔻醯基-sn-乙二醇（PEG-S-DMG）、以及R-3-[(ω-甲氧基-聚(乙二醇)2000)胺甲醯]-1,2-二肉豆蔻醯氧基丙基-3-胺（PEG-C-DOMG）可以由泰米拉製藥公司（Tekmira Pharmaceuticals）（溫哥華（Vancouver），加拿大（Canada））提供或合成。膽固醇可以購自西格瑪公司（Sigma）（聖路易斯（St Louis），密蘇里州（MO））。特異性CRISPR Cas RNA可以封裝在含有DLinDAP、DLinDMA、DLinK-DMA和DLinKC2-DMA的LNP中（陽離子脂質:DSPC:CHOL:PEGS-DMG或PEG-C-DOMG，莫耳比為40：10：40：10）。在需要時，可以摻入0.2% SP-DiOC18（英傑公司（Invitrogen），伯靈頓（Burlington），加拿大）來評估細胞攝取、細胞內遞送和生物分佈。可以藉由以下方式來進行封裝：將由陽離子脂質:DSPC:膽固醇:PEG-c-DOMG（40：10：40：10莫耳比）組成的脂質混合物溶解在乙醇中，直至最終脂質濃度為10 mmol/l。可以將脂質的此乙醇溶液逐滴添加到pH 4.0的50 mmol/l檸檬酸鹽中以形成多層囊泡，從而產生30%（體積/體積）乙醇的終濃度。在使用擠出機（北方脂質公司（Northern Lipids），溫哥華，加拿大）藉由兩個重疊的80 nm Nuclepore聚碳酸酯過濾器擠出多層囊泡之後，可以形成大的單層囊泡。可以藉由如下步驟實現封裝：將溶解在含有30%乙醇(體積/體積）的pH 4.0的50 mmol/l檸檬酸鹽中的2 mg/ml的RNA逐滴添加到擠出的預成形的大單層囊泡中，並且在31°C下培養30分鐘，伴隨持續混合直至最終的RNA/脂質重量比為0.06/1（重量/重量）。藉由使用Spectra/Por 2再生纖維素透析膜在pH 7.4的磷酸鹽緩衝鹽水（PBS）中透析16小時進行乙醇的去除以及配製緩衝液的中和。可以使用NICOMP

370型粒徑分析儀、囊泡/強度模式以及高斯擬合藉由動態光散射測定奈米粒子粒徑分佈（Nicomp粒徑分析儀，聖巴巴拉市（Santa Barbara），加利福尼亞州）。對於所有三個LNP系統的粒徑可以是~70 nm的直徑。可以藉由使用VivaPureD MiniH柱（賽多利斯斯泰迪生物技術公司（Sartorius Stedim Biotech））從分析前後收集的樣品中去除游離RNA來確定RNA封裝效率。可以從洗提的奈米粒子中提取封裝的RNA並且將其在260 nm下量化。藉由使用來自美國瓦克化學公司（Wako Chemicals USA）（里士滿（Richmond），維吉尼亞州（VA））的膽固醇E酶測定法測量囊泡中的膽固醇含量來確定RNA與脂質的比率。與在此的LNP和PEG脂質的論述結合，PEG化的脂質體或LNP同樣適合於CRISPR-Cas系統或其組分的遞送。

[0632] 大LNP的製備可以使用和/或改編自羅辛等人，分子治療，第19卷，第12期，第1286-2200頁，2011年12月。可以在含有50：10：38.5莫耳比的DLinKC2-DMA、DSPC和膽固醇的乙醇中製備脂質預混物溶液（20.4 mg/ml總脂質濃度）。可以0.75：1的莫耳比（乙酸鈉:DLinKC2-DMA）將乙酸鈉添加到脂質預混物中。隨後可以藉由將該混合物與1.85倍體積的檸檬酸鹽緩衝液（10 mmol/l，pH 3.0）在劇烈攪拌下合併來使脂質水合，從而使得自發脂質體在含有35%乙醇的水性緩衝液中形成。可以在37°C下培養該脂質體溶液以允許粒徑的時間依賴性增加。可以藉由動態光散射（奈米粒徑電位分析儀（Zetasizer Nano ZS），瑪律文儀器公司（Malvern Instruments），烏斯特郡（Worcestershire），英國（UK））在培養過程中的不同時間處去除等分試樣來研究脂質體尺寸的變

化。一旦實現所希望的粒徑，可以將水性PEG脂質溶液（儲備溶液＝在35%（體積/體積）乙醇中的10 mg/ml PEG-DMG）添加到該脂質體混合物中，以產生3.5%總脂質的最終PEG莫耳濃度。在添加PEG-脂質之後，該等脂質體應該其大小，有效抑制進一步生長。然後可以大約1：10（重量:重量）的RNA與總脂質比率將RNA添加到空脂質體中，然後在37°C下培養30分鐘以形成載入的LNP。隨後可以將該混合物在PBS中透析過夜，並且用0.45-μm的注射過濾器（syringe filter）進行過濾。

[0633] 球形核酸（SNA™）構建體和其他奈米粒子（尤其是金奈米粒子）也被考慮作為將CRISPR-Cas系統遞送至預期靶標的手段。重要資料表明，基於核酸功能化的金奈米粒子的AuraSense治療性球形核酸（SNA™）構建體係可用的。

[0634] 可以與在此的教授內容結合使用的文獻包括：卡特勒（Cutler）等人，美國化學學會雜誌 2011 133:9254-9257，郝（Hao）等人，Small. 2011 7:3158-3162，張等人，ACS奈米 2011 5:6962-6970，卡特勒等人，美國化學學會雜誌 2012 134:1376-1391，楊（Young）等人，奈米快報 2012 12:3867-71，鄭（Zheng）等人，美國國家科學院院刊 2012 109:11975-80，米爾金（Mirkin），奈米醫學（Nanomedicine） 2012 7:635-638 張等人 美國化學學會雜誌 2012 134:16488-1691，因特勞布（Weintraub），自然 2013 495:S14-S16，崔（Choi）等人，美國國家科學院院刊 2013 110(19):7625-7630，詹森（Jensen）等人，科學轉化醫學 5, 209ra152 (2013)以及米爾金等人，Small, 10:186-192。

[0635] 具有RNA的自組裝奈米粒子可以用PEG化的聚乙烯亞胺（PEI）構建，其中Arg-Gly-Asp（RGD）肽配位基附接在聚乙二醇（PEG）的遠端。此系統已作為例如靶向表現整聯蛋白的腫瘤新血管系統的手段和作為遞送抑制血管內皮生長因子受體2（VEGF R2）表現以及由此實現腫瘤血管新生的siRNA的手段（例如參見，施弗勒斯（Schiffelers）等人，核酸研究，2004，第32卷，第19期）。奈米叢（Nanoplexes）可以藉由以下方式製備：將等體積的陽離子聚合物水性溶液和核酸水性溶液混合，以在2至6範圍內產生可電離氮（聚合物）相比磷酸鹽（核酸）的淨莫耳過量。在陽離子聚合物與核酸之間的靜電相互作用使得聚複合物（polyplexes）形成，該聚複合物具有約100 nm的平均粒徑分佈，因此在此稱之為奈米叢。設想了CRISPR Cas的約100至200 mg的劑量，用於施弗勒斯等人的自組裝奈米粒子中的遞送。

[0636] 巴特利特（Bartlett）等人的奈米叢（美國國家科學院院刊，2007年9月25日，第104卷，第39期）也可以適用於本發明。巴特利特等人的奈米叢藉由以下方式製備：將等體積的陽離子聚合物水性溶液和核酸水性溶液混合，以在2至6範圍內產生可電離氮（聚合物）相比磷酸鹽（核酸）的淨莫耳過量。在陽離子聚合物與核酸之間的靜電相互作用使得聚複合物形成，該聚複合物具有約100 nm的平均粒徑分佈，因此在此稱為奈米叢。巴特利特等人的DOTA-siRNA合成如下：1,4,7,10-四氮雜環十二烷-1,4,7,10-四乙酸單(N-羥基琥珀醯亞胺酯)（DOTA-NHS酯）訂購自Macrocyclics公司（達拉斯（Dallas），德克薩斯州（TX））。將於碳酸鹽緩衝液（pH 9）中的具有100倍莫耳過量的DOTA-NHS-酯的胺修飾的

RNA有義股添加到微量離心管中。藉由在室溫攪拌4 h使該等內容物反應。將該DOTA-RNA有義軛合物用乙醇沈澱，重新懸浮在水中，並且退火到未修飾的反義股上以產生DOTA-siRNA。所有液體用Chelex-100（伯樂公司（Bio-rad），赫庫斯（Hercules），加利福尼亞州）預處理，以便去除痕量金屬污染物。可以藉由使用含有環糊精的聚陽離子形成Tf靶向和非靶向的siRNA奈米粒子。典型地，以3（+/-）的進料比和0.5克/升的siRNA濃度在水中形成奈米粒子。用Tf（金剛烷-PEG-Tf）修飾在靶向奈米粒子表面上的百分之一的金剛烷-PEG分子。將奈米粒子懸浮在用於注射的5%（重量/體積）葡萄糖載體溶液中。

[0637] 大衛斯（Davis）等人（自然，第464卷，2010年4月15日）進行了使用靶向的奈米粒子遞送系統的RNA臨床試驗（臨床試驗登記號NCT00689065）。在21天週期的第1、3、8和10天藉由30 min的靜脈內輸注對患有標準護理治療難治的實體癌的患者給予靶向的奈米粒子劑量。該等奈米粒子由合成遞送系統組成，該系統含有：(1) 線性的、基於環糊精的聚合物（CDP），(2) 展示在奈米粒子外部上的用於接合癌細胞表面上的TF受體（TFR）的人轉鐵蛋白（TF）靶向配位基，(3) 親水性聚合物（用於促進奈米粒子在生物流體中的穩定性的聚乙二醇（PEG）），以及 (4) 被設計為降低RRM2（先前在臨床中使用的序列指代為siR2B+5）表現的siRNA。長久以來已知TFR在惡性細胞中被下調，並且RRM2係一種確立的抗癌靶標。已經顯示該等奈米粒子（臨床版本指代為CALAA-01）在非人類靈長類動物中的多劑量研究中耐受性良好。雖然已經藉由脂質體遞送向患有慢性粒細胞白血病的單一患者給

予了siRNA，但是大衛斯等人的臨床試驗係初期人類試驗，該試驗用靶向遞送系統全身性地遞送siRNA並且治療患有實體癌的患者。為了確定該靶向遞送系統是否能夠將功能性siRNA有效遞送至人類腫瘤，大衛斯等人研究了來自三個不同的劑量組群的三位患者的活組織檢查；患者A、B和C，他們均患有轉移性黑素瘤並且分別接受了18、24和30 mg m^{-2} siRNA的CALAA-01劑量。還可以針對本發明的CRISPR Cas系統考慮類似的劑量。用含有線性的基於環糊精的聚合物（CDP）、展示在奈米粒子外部上的用於接合癌細胞表面上的TF受體（TFR）的人轉鐵蛋白（TF）靶向配位基和/或親水聚合物（例如，用於促進奈米粒子在生物流體中的穩定性的聚乙二醇（PEG））的奈米粒子，可以實現本發明的遞送。

[0638] 就本發明而言，較佳的是使用奈米粒子或脂質包膜遞送CRISPR複合物的一種或多種組分，例如CRISPR酶或mRNA或指導RNA。其他遞送系統或載體可以與本發明奈米粒子方面結合使用。

[0639] 總的來說，“奈米粒子”係指具有小於1000 nm直徑的任何粒子。在某些較佳的實施方式中，本發明奈米粒子具有500 nm或更小的最大尺寸（例如，直徑）。在其他較佳的實施方式中，本發明奈米粒子具有介於25 nm與200 nm之間的範圍內的最大尺寸。在其他較佳的實施方式中，本發明奈米粒子具有100 nm或更小的最大尺寸（例如，直徑）。在其他較佳的實施方式中，本發明奈米粒子具有在35 nm與60 nm之間的範圍內的最大尺寸。

[0640] 本發明中涵蓋的奈米粒子（Nanoarticles）可以提供為

不同的形式，例如為固體奈米粒子（例如金屬（諸如銀、金、鐵、鈦）、非金屬、基於脂質的固體、聚合物）、奈米粒子的懸浮液、或它們的組合。可以製備金屬、絕緣體和半導體奈米粒子，以及雜合結構（例如核-殼奈米粒子）。如果由半導體材料製備的奈米粒子足夠小（典型地亞-10 nm）以至於出現電子能級的量子化，則該等奈米粒子還可以是標記量子點。此類奈米級粒子作為藥物載體或成像劑用於生物醫學應用中並且可以適於本發明中的類似目的。

[0641] 半固體和軟奈米粒子以被製造出並且處於本發明的範圍之內。半固體性質的原型奈米粒子係脂質體。目前，不同類型的脂質體奈米粒子在臨床上用作用於抗癌藥物和疫苗的遞送系統。具有一半親水性和另一半疏水性的奈米粒子被稱為雙面（Janus）粒子並且用於穩定化乳液是特別有效的。它們可以在水/油介面處自組裝並且充當固體表面活性劑。

[0642] 美國專利案號8,709,843（藉由引用結合在此）提供了用於將含有治療劑的粒子靶向遞送至組織、細胞和細胞內區室的藥物遞送系統。本發明提供了包含軛合至表面活性劑、親水性聚合物或脂質的聚合物的靶向粒子。

[0643] 美國專利案號6,007,845（藉由引用結合在此）提供了下述粒子：該等粒子具有藉由將多官能化合物與一種或多種疏水性聚合物和一種或多種親水性聚合物共價地連接而形成的多嵌段共聚合物的核，並且含有生物活性材料。

[0644] 美國專利案號5,855,913（藉由引用結合在此）提供了下述顆粒組成物：該組成物含有具有小於0.4 g/cm3振實密度以及

介於5 μm與30 μm之間的平均直徑的空氣動力學光粒子，在其表面上摻入了表面活性劑，以用於向肺部系統的藥物遞送。

[0645] 美國專利案號5,985,309（藉由引用結合在此）提供了下述粒子：該等粒子摻入了表面活性劑和/或帶正電荷或帶負電荷的治療劑或診斷劑和帶相反電荷的分子的親水性或疏水性複合物，以用於向肺部系統的遞送。

[0646] 美國專利案號5,543,158（藉由引用結合在此）提供了生物可降解可注射粒子，該等粒子具有生物可降解實心核，該核在其表面上含有生物活性材料和聚(烷撐二醇)部分。

[0647] WO2012135025（也以US20120251560公開）（藉由引用結合在此）描述了軛合的聚乙烯亞胺（PEI）聚合物和軛合的氮雜大環（統稱為“一個軛合微質體（lipomer）”或“多個微質體（lipomers）”）。在某些實施方式中，可以設想的是，此類軛合微質體可以用於CRISPR-Cas系統的情況中以在體外、離體和在體內實現基因組干擾，從而修飾基因表現，包括蛋白質表現的調控。

[0648] 在一個實施方式中，奈米粒子可以是環氧化物修飾的脂質-聚合物，有利地是7C1（例如參見，詹姆斯·E.·達爾曼和卡曼·巴恩斯等人，自然奈米科技 (2014) 2014年5月11日線上公開，doi:10.1038/nnano.2014.84）。C71係藉由使C15環氧化物封端脂質與PEI600以14：1莫耳比反應合成的，並且與C14PEG2000一起進行配製以產生奈米粒子（直徑介於35與60 nm之間），該等奈米粒子在PBS溶液中穩定至少40天。

[0649] 環氧化物修飾的脂質-聚合物可以用於將本發明的CRISPR-Cas系統遞送至肺部、心血管或腎細胞，然而，熟習該項技術者可以將該系統適配成遞送其他靶器官。設想了在從約0.05至約0.6 mg/kg範圍內的劑量。還設想了在數天或數週內的劑量，總劑量為約2 mg/kg。

外來體

[0650] 外來體係轉運RNA和蛋白質的內源性奈米囊泡，並且可以將RNA遞送至腦和其他靶器官。為了降低免疫原性，阿爾瓦雷斯-爾維蒂（Alvarez-Erviti）等人（2011，自然生物技術 29:341）使用了用於外來體產生的自我衍生的樹突細胞。藉由將樹突細胞工程化為表現Lamp2b（一種外來體膜蛋白，融合至神經元特異性RVG肽）實現對腦的靶向。藉由電穿孔使純化的外來體載入外源性RNA。靜脈內注射的RVG靶向的外來體將GAPDH siRNA特異性地遞送至腦中的神經元、小膠質細胞、少突神經膠質細胞，導致特異性的基因敲低。預暴露於RVG外來體未減弱敲低，並且在其他組織中未觀察到非特異性攝取。藉由BACE1的強的mRNA（60%）和蛋白質（62%）敲低證明了外來體介導的siRNA遞送的治療潛能，BACE1係一種阿茲海默症中的治療靶標。

[0651] 為了獲得免疫惰性的外來體庫，阿爾瓦雷斯-爾維蒂等人收穫了來自具有同源主要組織相容性複合體（MHC）單倍型的近交C57BL/6小鼠的骨髓。由於未成熟樹突細胞產生大量的缺乏T細胞啟動劑諸如MHC-II和CD86的外來體，阿爾瓦雷斯-爾維蒂（Alvarez-Erviti）等人選擇了具有粒細胞/巨噬細胞集落刺激因子

（GM-CSF）的樹突細胞，持續7天。次日，使用良好建立的超速離心方案從培養上清液中純化外來體。產生的外來體在物理上是均質的，具有直徑為80 nm的粒徑分佈峰，如藉由奈米粒子跟蹤分析（NTA）和電子顯微鏡檢查所測定。阿爾瓦雷斯-爾維蒂等人獲得了6-12 μg的外來體（基於蛋白質濃度測量的）/10^6個細胞。

[0652] 其次，阿爾瓦雷斯-爾維蒂等人研究了使用適於奈米級應用的電穿孔方案給修飾的外來體載入外源性貨物的可能性。由於電穿孔對於奈米級的膜粒子尚未良好表徵，使用非特異性Cy5標記的RNA用於電穿孔方案的經驗優化。在外來體超速離心和溶解之後測定了封裝的RNA量。在400 V和125 μF下的電穿孔產生RNA的最大保留並且用於所有的後續實驗。

[0653] 阿爾瓦雷斯-爾維蒂等人向正常C57BL/6小鼠給予被封裝在150 μg的RVG外來體中的150 μg的每種BACE1 siRNA並且將敲低效率與四個對照進行比較：未處理的小鼠、僅用RVG外來體注射的小鼠、用與一種體內陽離子脂質體試劑複合的BACE1 siRNA注射的小鼠、以及用與RVG-9R複合的BACE1 siRNA注射的小鼠，該RVG肽與靜電結合siRNA的9個D-精胺酸軛合。在給藥之後3天，分析皮層組織樣品，並且在siRNA-RVG-9R處理的和siRNARVG外來體處理的小鼠中均觀察到顯著的蛋白質敲低（45%，$P < 0.05$，相對於62%，$P < 0.01$），這係由於BACE1 mRNA水平的顯著降低（分別為66% [+或-] 15%，$P < 0.001$和61% [+或-] 13%，$P < 0.01$）。此外，申請人證明了在RVG-外來體處理的動物中總[β]-澱粉樣蛋白1-42水平上的顯著降低（55%，$P < 0.05$），該β澱粉樣

蛋白為在阿茲海默症理學中的澱粉樣白斑的主要成分。所觀察到的降低大於在心室內注射BACE1抑制劑之後的正常小鼠中展示的β澱粉樣蛋白1-40降低。阿爾瓦雷斯-爾維蒂等人在BACE1切割產物上進行了5'-cDNA末端快速擴增（RACE），其提供了經由siRNA的RNAi-介導的敲低的證據。

[0654] 最後，阿爾瓦雷斯-爾維蒂等人藉由評估IL-6、IP-10、TNFα和IFN-α血清濃度研究了RNA-RVG外來體是否誘導了體內免疫應答。在外來體處理之後，類似於與強有力地刺激IL-6分泌的siRNA-RVG-9R相反的siRNA轉染試劑處理，登記了在所有細胞介素上的非顯著性變化，證實了該外來體處理的免疫惰性屬性（profile）。假定外來體僅封裝20%的siRNA，用RVG-外來體的遞送比RVG-9R遞送顯得更有效，因為用少五倍的siRNA實現了相當的mRNA敲低和更好的蛋白質敲低，而沒有相應水平的免疫刺激。這個實驗證明了RVG-外來體技術的治療潛力，其潛在地適合於與神經變性疾病相關的基因的長期沈默。阿爾瓦雷斯-爾維蒂等人的外來體遞送系統可以適用於將本發明的CRISPR-Cas系統遞送至治療靶標，尤其是神經變性疾病。對於本發明可以考慮封裝在約100至1000 mg的RVG外來體中的約100至1000 mg的CRISPR Cas的劑量。

[0655] 艾爾·安達盧西（El-Andaloussi）等人（自然實驗手冊（Nature Protocols） 7,2112–2126(2012)）揭露了可以如何利用來源於培養的細胞的外來體用於體外和體內遞送RNA。這個方案首先描述了藉由轉染一種包含與肽配位基融合的外來體蛋白的表現

載體產生靶向的外來體。接著，艾爾·安達盧西等人解釋了如何純化和表徵來自轉染的細胞上清液的外來體。接著，艾爾·安達盧西等人詳述了將RNA載入到外來體中的關鍵步驟。最後，艾爾·安達盧西等人概述了如何使用外來體有效地在體外遞送RNA以及體內遞送至小鼠腦中。還提供了預期結果的實例，其中外來體介導的RNA遞送藉由功能測定和成像來評價。整個方案進行~3週。根據本發明的遞送或給藥可以使用從自我衍生的樹突細胞產生的外來體來進行。根據在此的教授內容，這可以應用在本發明的實踐中。

[0656] 在另一個實施方式中，考慮了瓦爾葛蘭（Wahlgren）等人的血漿外來體（核酸研究，2012年，第40卷，第17期，e130)。外來體係由包括樹突細胞（DC)、B細胞、T細胞、肥大細胞、上皮細胞和腫瘤細胞的許多細胞類型產生的奈米尺寸的囊泡（30-90 nm大小)。該等囊泡藉由晚期內體的向內出芽而形成，並且然後在與質膜融合後釋放到細胞外環境。因為外來體天然地在細胞之間運送RNA，所以這種特性在基因治療中可能有用，並且根據本揭露可以應用在本發明的實踐中。

[0657] 來自血漿的外來體可以藉由以下方式製備：以900 g離心血沈棕黃層20 min以便分離血漿，然後收穫細胞上清液，以300 g離心10 min以便消除細胞，並且以16500 g離心30 min，然後藉由0.22 mm過濾器進行過濾。藉由以120000 g超速離心70 min使外來體沈澱。根據在RNAi人類/小鼠啟動（Starter）套組（凱傑公司（Quiagen)，希爾頓（Hilden)，德國）中的製造商說明進行siRNA到外來體中的化學轉染。siRNA以終濃度2 mmol/ml添加到100ml

PBS中。在添加HiPerFect轉染試劑之後，將該混合物在室溫下培養10 min。為了去除過量的膠束，使用醛/硫酸鹽乳膠珠再分離外來體。可以類似於siRNA進行CRISPR Cas到外來體中的化學轉染。外來體可以與從健康供體的外週血中分離的單核細胞和淋巴細胞共培養。因此，可以考慮的是，可以將含有CRISPR Cas的外來體引入到人類的單核細胞和淋巴細胞中並且以自體方式再引入到人類中。因此，可以使用血漿外來體進行根據本發明的遞送或給藥。

脂質體

[0658] 可以用脂質體進行根據本發明的遞送或給藥。脂質體係球形囊泡結構，其由圍繞內部水性區室的單層或多層脂質雙層以及相對不可滲透的外部親脂性磷脂雙層構成。脂質體作為藥物遞送載體受到了相當的重視，因為它們係生物相容、無毒的，可以遞送親水性和親脂性藥物分子，保護它們的貨物免於被血漿酶降解，並且轉運它們的負荷跨過生物膜和血腦障壁（BBB）（對於評述，例如參見，斯普奇（Spuch）和納瓦羅（Navarro），藥物遞送雜誌（Journal of Drug Delivery），第2011卷，文獻標識碼469679，第12頁，2011. doi:10.1155/2011/469679）。

[0659] 可以由幾種不同類型的脂質製造脂質體；然而，磷脂最常用來產生作為藥物載體的脂質體。雖然當脂質膜與一種水性溶液混合時脂質體形成係自發的，但是也可以藉由使用均質機、超音波破碎器、或擠出設備藉由以振盪的形式施加力使其加速（對於評述，例如參見，斯普奇和納瓦羅，藥物遞送雜誌，第2011卷，文獻標識碼469679，第12頁，2011. doi:10.1155/2011/469679）。

[0660] 可以將幾種其他的添加劑添加到脂質體中，以便修飾其結構和特性。例如，可以將膽固醇或鞘磷脂添加到脂質體混合物中，以便幫助穩定化脂質體結構並且防止脂質體內部貨物的洩漏。此外，脂質體由氫化卵磷脂醯膽鹼或卵磷脂醯膽鹼、膽固醇和磷酸二鯨蠟脂製備，並且脂質體的平均囊泡尺寸被調整到約50 nm和100 nm。(對於評述，例如參見，斯普奇和納瓦羅，藥物遞送雜誌，第2011卷，文獻標識碼469679，第12頁，2011. doi:10.1155/2011/469679)。

[0661] 脂質體配製物可以是主要由天然磷脂和脂質諸如1,2-二硬脂醯-sn-甘油基-3-磷脂醯膽鹼（DSPC）、鞘磷脂、卵磷脂醯膽鹼和單唾液醯神經節苷酯構成。因為這種配製物僅由磷脂組成，所以脂質體配製物已經遇到了許多挑戰，其中之一係在血漿中的不穩定性。已經做出戰勝該等挑戰的若干嘗試，特別是在脂質膜的處理方面。該等嘗試之一集中於膽固醇的處理。將膽固醇添加到常規配製物中減緩了封裝的生物活性化合物到血漿中的迅速釋放，或者添加1,2-二油醯基-sn-甘油基-3-磷酸乙醇胺（DOPE）增加穩定性（對於評述，例如參見，斯普奇和納瓦羅，藥物遞送雜誌，第2011卷，文獻標識碼469679，第12頁，2011. doi:10.1155/2011/469679)。

[0662] 在一個特別有利的實施方式中，特洛伊木馬（Trojan Horse）脂質體（也稱為分子特洛伊木馬）係令人希望的並且方案可見於http://cshprotocols.cshlp.org/content/2010/4/pdb.prot5407.long。該

等粒子允許轉基因在血管內注射之後遞送至整個腦。在不受限制的情況下，據信表面軛合有特異性抗體的中性脂質粒子允許經由胞吞作用跨過血腦障壁。申請人假定利用特洛伊木馬脂質體將核酸酶的CRISPR家族經由血管內注射遞送至腦，這將允許全腦轉基因動物，而不需要胚胎操縱。對於脂質體中的體內給藥，可以考慮約1-5 g的DNA或RNA。

[0663] 在另一個實施方式中，CRISPR Cas系統或其組分可以脂質體諸如穩定的核酸脂質粒子（SNALP）來給予（例如，參見，莫里西等人，自然生物技術，第23卷，第8期，2005年8月）。考慮每日靜脈內注射約1、3或5 mg/kg/天的SNALP中的被靶向的特異性CRISPR Cas。日治療可以經過約三天，並且然後每週治療持續約五週。在另一個實施方式中，還考慮了藉由以約1或2.5 mg/kg的劑量靜脈內注射給予封裝有特異性CRISPR Cas的SNALP（例如參見，齊默爾曼（Zimmerman）等人，自然快報（Nature Letters），第441卷，2006年5月4日）。該SNALP配製物可以含有為2：40：10：48的莫耳百分比的脂質3-N-[(w甲氧基聚(乙二醇)2000)胺甲醯]-1,2-二肉豆蔻氧基-丙胺（PEG-C-DMA）、1,2-二亞油基氧基-N,N-二甲基-3-胺基丙烷（DLinDMA）、1,2-二硬脂醯-sn-甘油基-3-磷酸膽鹼（DSPC）和膽固醇（例如參見，齊默爾曼等人，自然快報，第441卷，2006年5月4日）。

[0664] 在另一個實施方式中，已經證明穩定的核酸-脂質粒子（SNALP）將分子有效地遞送至高度血管化的HepG2-衍生的肝臟腫瘤，但是不遞送至血管化不良的HCT-116衍生的肝臟腫瘤(例

如參見，李，基因治療 (2012) 19, 775–780)。可以藉由以下方式製備SNALP脂質體：使用25：1的脂質/siRNA比率和48/40/10/2的膽固醇/D-Lin-DMA/DSPC/PEG-C-DMA的莫耳比，用二硬脂醯磷脂醯膽鹼(DSPC)、膽固醇和siRNA配製D-Lin-DMA和PEG-C-DMA。生成的SNALP脂質體的大小為約80-100 nm。

[0665] 在又另一個實施方式中，SNALP可以包含合成膽固醇(西格瑪-奧德里奇公司(Sigma-Aldrich)，聖路易斯，密蘇里州，美國)、二棕櫚醯磷脂醯膽鹼（阿凡提極地脂質公司（Avanti Polar Lipids)，阿拉巴斯特（Alabaster)，阿拉巴馬州（AL)，美國)、3-N-[(w-甲氧基聚（乙二醇）2000）胺甲醯]-1,2-二肉豆蔻氧基丙基胺，以及陽離子的1,2-二亞油基氧基-3-N,N二甲基胺基丙烷(例如參見，蓋斯伯特(Geisbert)等人，柳葉刀(Lancet) 2010;375: 1896-905)。例如可以考慮靜脈內推注給予約2 mg/kg總CRISPR Cas/劑的劑量。

[0666] 在又另一個實施方式中，SNALP可以包含合成膽固醇（西格瑪-奧德里奇公司)、1,2-二硬脂醯基-sn-甘油基-3-磷酸膽鹼（DSPC；阿凡提極地脂質公司)、PEG-cDMA，以及1,2-二亞油基氧基-3-(N;N-二甲基)胺基丙烷（DLinDMA)(例如參見，賈奇（Judge)，臨床研究雜誌(J.Clin.Invest.) 119:661-673 (2009))。用於體內研究的配製物可以包含約9：1的最終脂質/RNA質量比。

[0667] 已經由阿爾尼拉姆製藥公司（Alnylam Pharmaceuticals）的巴羅斯（Barros）和格羅布（Gollob）評論了RNAi奈米藥物的安全性曲線（例如參見，先進藥物遞送評論

（Advanced Drug Delivery Reviews） 64 (2012) 1730–1737)。穩定的核酸脂質粒子（SNALP）由四種不同的脂質構成—在低pH下為陽離子的可電離脂質（DLinDMA）、中性輔助脂質、膽固醇以及可擴散的聚乙二醇（PEG）-脂質。該粒子的直徑為大約80 nm並且在生理pH下是電中性的。在配製過程中，該可電離脂質用於在粒子形成過程中使脂質與陰離子RNA縮合。當在漸增的酸性內體條件下帶正電荷時，該可電離的脂質還介導了SNALP與內體膜的融合，從而能夠將RNA釋放到細胞質中。該PEG-脂質在配製過程中穩定化粒子並且減少聚集，並且隨後提供改進藥代動力學特性的中性的親水性外部。

[0668] 到目前為止，已經使用具有RNA的SNALP配製物開始兩個臨床項目。泰米拉製藥公司最近在具有升高的LDL膽固醇的成年志願者中完成了SNALP-ApoB I期單劑量研究。ApoB主要是在肝臟和空腸中表現，並且是為VLDL和LDL的組裝和分泌所必需的。十七位受試者接受了SNALP-ApoB的單劑量（跨7個劑量水平的劑量遞增）。沒有肝臟毒性（預期為基於臨床前研究的潛在劑量限制性毒性）的證據。處於最高劑量的（兩位中的）一位受試者經歷了與免疫系統刺激一致的流感樣症狀，並且做出結束該試驗的決定。

[0669] 阿爾尼拉姆製藥公司已經類似地推出了ALN-TTR01，其採用以上所述的SNALP技術並且靶向突變體和野生型TTR的肝細胞產生，以治療TTR澱粉樣變性（ATTR）。已經描述了三種ATTR綜合症：家族性澱粉樣變性多神經病（FAP）和家族性澱粉樣心肌

病（FAC）—兩者均由TTR中的常染色體顯性突變引起；以及由野生型TTR引起的老年全身性澱粉樣變性（SSA）。最近在具有ATTR的患者中完成了ALN-TTR01的安慰劑對照單劑量遞增I期試驗。向31位患者（23位用研究藥物，8位用安慰劑）在0.01至1.0 mg/kg（基於siRNA）的劑量範圍內以15分鐘靜脈內輸注給予ALN-TTR01。治療耐受性良好，其中在肝功能試驗中沒有顯著增加。在≥0.4 mg/kg時在23位患者的3位中注意到輸注相關反應；所有患者均對減慢輸注速率做出了響應並且所有患者繼續參與研究。在處於1 mg/kg的最高劑量（如根據臨床前和NHP研究預期的）的兩位患者中注意到血清細胞介素IL-6、IP-10和IL-1ra的最小與暫態升高。在1 mg/kg時觀察到ALN-TTR01的預期藥效動力學效應，即，血清TTR的降低。

[0670] 在又另一個實施方式中，可以藉由將陽離子脂質、DSPC、膽固醇以及PEG-脂質例如以40：10：40：10的莫耳比分別溶解在例如乙醇中來製備SNALP（參見，森普爾（Semple）等人，自然生物技術，第28卷，第2期，2010年2月，第172-177頁）。將該脂質混合物添加到水性緩衝液中（50 mM檸檬酸鹽，pH 4），混合至最終的乙醇和脂質濃度分別為30%（體積/體積）和6.1 mg/ml，並且使得其在22°C下平衡2 min，然後擠出。使用Lipex擠出儀（北方脂質公司），在22°C下將水合脂質擠出藉由兩個重疊的80 nm孔徑大小的過濾器（Nuclepore），直到獲得如藉由動態光散射分析測定的70-90 nm直徑的囊泡為止。這大致需要1-3道次。將該siRNA（溶解在50 mM檸檬酸鹽中，pH為4的含有30%乙醇的水性溶液）以約~5 ml/min的速率在混合下添加到預平衡的（35°C）囊泡中。

在達到0.06（重量/重量）的最終靶siRNA/脂質比率之後，將該混合物在35°C下另外培養30 min，以允許囊泡重組和siRNA的封裝。然後去除乙醇並且藉由透析或切向流滲濾用PBS(155 mM NaCl，3 mM Na_2HPO_4，1 mM KH_2PO_4，pH 7.5)替換外部緩衝液。使用受控的逐步稀釋法工藝將siRNA封裝在SNALP中。KC2-SNALP的脂質組分為以57.1 : 7.1 : 34.3 : 1.4的莫耳比使用的DLin-KC2-DMA（陽離子脂質）、二棕櫚醯磷脂醯膽鹼（DPPC；阿凡提極地脂質公司）、合成膽固醇（西格瑪公司)和PEG-C-DMA。在形成載入的粒子後，將SNALP在PBS中透析並且在使用之前藉由0.2 μm的過濾器滅菌過濾。平均粒徑為75-85 nm，並且將90%-95%的siRNA封裝在脂質粒子之內。用於體內測試的在配製物中的最終siRNA/脂質比率係~0.15（重量/重量）。在臨使用之前將含有因子VII siRNA的LNP-siRNA系統在無菌PBS中稀釋到適當濃度，並且藉由側尾靜脈以10 ml/kg的總體積靜脈內給藥。這種方法和該等系統可以外推到本發明的CRISPR Cas系統。

其他脂質

[0671] 其他陽離子脂質，諸如胺基脂質2,2-二亞油基-4-二甲基胺基乙基-[1,3]-二氧戊環（DLin-KC2-DMA）可以類似於SiRNA地用來封裝CRISPR Cas或其組分或對其編碼的一個或多個核酸分子（例如參見，加雅拉曼（Jayaraman），德國應用化學（Angew.Chem.Int.Ed.） 2012, 51, 8529 –8533），並且因此可以應用於本發明的實踐中。可以考慮具有下列脂質組成的預成型囊泡：分別處於莫耳比40/10/40/10的胺基脂質、二硬脂醯磷脂醯膽鹼

（DSPC）、膽固醇和(R)-2,3-雙(十八烷氧基)丙基-1-(甲氧基聚(乙二醇)2000)丙基碳酸酯(PEG-脂質)，以及大約0.05（w/w）的FVII siRNA/總脂質比率。為了確保在70–90 nm範圍內的窄粒徑分佈以及0.11±0.04（n=56）的低多分散性指數，可以在添加指導RNA之前將粒子藉由80 nm的膜擠出達三次。可以使用含有高度有效的胺基脂質16的粒子，其中四種脂質組分16、DSPC、膽固醇和PEG-脂質的莫耳比（50/10/38.5/1.5）可以被進一步優化，以增強體內活性。

[0672] 邁克爾 S D 科爾曼(Michael S D Kormann)等人("在小鼠中遞送化學修飾的mRNA之後治療蛋白的表現（Expression of therapeutic proteins after delivery of chemically modified mRNA in mice: Nature Biotechnology)"：自然生物技術，第29卷，第154-157頁，(2011)）描述了脂質包膜用於遞送RNA的用途。脂質包膜的用途在本發明中也是較佳的。

[0673] 在另一個實施方式中，脂質可以與本發明的CRISPR Cas系統或其一種或多種組分或對其編碼的一個或多個核酸分子一起配製而形成脂質奈米粒子（LNP）。脂質包括但不限於，DLin-KC2-DMA4、C12-200和輔助脂質二硬脂醯磷脂醯膽鹼、膽固醇和PEG-DMG，可以使用自發囊泡形成程序將其與CRISPR Cas而不是siRNA一起配製（例如參見，諾沃布塞瓦（Novobrantseva），分子治療-核酸（Molecular Therapy–Nucleic Acids） (2012) 1, e4; doi:10.1038/mtna.2011.3）。組分莫耳比可以是約50/10/38.5/1.5（DLin-KC2-DMA或C12-200/二硬脂醯磷脂醯膽鹼/膽固醇/PEG-

DMG）。在DLin-KC2-DMA和C12-200脂質奈米粒子（LNP）的情況下，最終脂質:siRNA重量比可以分別是~12：1和9：1。配製物可以具有~80 nm的平均粒子直徑，具有>90%的包封效率。可以考慮3 mg/kg的劑量。

[0674] 泰米拉公司在美國和國外具有一組針對LNP和LNP配製物的不同方面的大約95個同族專利（例如參見，美國專利案號7,982,027；7,799,565；8,058,069；8,283,333；7,901,708；7,745,651；7,803,397；8,101,741；8,188,263；7,915,399；8,236,943和7,838,658以及歐洲專利案號1766035；1519714；1781593和1664316），所有該等專利均可用於和/或適於本發明。

[0675] 該CRISPR Cas系統或其組分或對其編碼的一個或多個核酸分子可以封裝在PLGA微球中進行遞送，諸如進一步描述於美國公開申請20130252281和20130245107以及20130244279（轉讓給Moderna Therapeutics公司）中，該等申請涉及包含修飾的核酸分子的組成物的配製物方面，該等核酸分子可以編碼蛋白質、蛋白質先質、或該蛋白質或該蛋白質先質的部分或完全加工形式。該配製物可以具有50：10：38.5：1.5-3.0（陽離子脂質:融合脂質:膽固醇:PEG脂質）的莫耳比。PEG脂質可以選自但不限於PEG-c-DOMG、PEG-DMG。該融合脂質可以是DSPC。還參見，施魯姆（Schrum）等人，工程化核酸的遞送和配製（Delivery and Formulation of Engineered Nucleic Acids），美國公開申請20120251618。

[0676] Nanomerics公司的技術著手解決針對廣泛治療學的

生物利用度挑戰，包括基於低分子量疏水性藥物、肽以及核酸（質粒、siRNA、miRNA）的治療學。該技術已經證明了明顯優勢的特異性的給藥途徑包括口服途徑、跨血腦障壁的轉運、向實性瘤以及眼部的遞送。例如參見，馬薩等人，2013，ACS奈米，2013年2月26日；7(2):1016-26；烏切克布（Uchegbu）和秀，2013，製藥科學雜誌（J Pharm Sci.） 102(2):305-10和拉拉特薩等人，2012，控釋雜誌，2012年7月20日；161(2):523-36。

[0677] 美國專利公開案號20050019923描述了用於向哺乳動物身體遞送生物活性分子諸如多核苷酸分子、肽和多肽和/或藥劑的陽離子樹狀聚合物。該等樹狀聚合物適合於將生物活性分子的遞送靶向至例如肝臟、脾、肺、腎或心臟（或甚至腦）。樹狀聚合物係由簡單的支化單體單元以逐步方式製備的合成性3維大分子，其性質和功能性可以容易地進行控制和改變。樹狀聚合物經由向多功能核（發散式合成法）或朝向多功能核（收斂式合成法）重複加成結構單元來合成，並且結構單元的3維殼的每次加成使得更高級別的樹狀聚合物形成。聚丙烯亞胺樹狀聚合物從二胺基丁烷核開始，藉由對一級胺的丙烯腈的雙邁克爾加成反應向其上添加兩倍數目的胺基基團，然後進行腈的氫化。這導致胺基基團的加倍。聚丙烯亞胺樹狀聚合物含有100%的可質子化氮以及高達64個末端胺基基團（5級，DAB 64）。可質子化基團通常是能夠在中性pH下接受質子的胺基。樹狀聚合物作為基因遞送劑的用途在很大程度上集中於聚醯胺-胺和含磷化合物的用途，其中胺/醯胺的混合物或N--P(O_2)S分別作為軛合單元，沒有報導關於更低級別的聚丙烯亞胺樹狀聚合物用於基因遞送的用途的工作。還研究了作為pH

敏感的控制釋放系統的聚丙烯亞胺樹狀聚合物，其用於藥物遞送以及當被外週胺基酸基團化學修飾時用於它們的客體分子的封裝。還研究了聚丙烯亞胺樹狀聚合物的細胞毒性和其與DNA的相互作用以及DAB 64的轉染效力。

[0678] 美國專利公開案號20050019923係基於與早期報導相反的觀察：陽離子樹狀聚合物諸如聚丙烯亞胺樹狀聚合物展示出適當的特性，諸如特異性靶向和低毒性，其用於靶向遞送生物活性分子諸如遺傳物質。此外，陽離子樹狀聚合物的衍生物也展示出適用於生物活性分子的靶向遞送的特性。還參見，生物活性聚合物（Bioactive Polymers）、美國公開申請20080267903，其揭露“不同的聚合物，包括陽離子聚胺聚合物和樹枝狀聚合物顯示出具有抗增殖活性，並且因此可用於治療特徵為不希望的細胞增殖的病症，諸如新生物和腫瘤、炎性病症（包括自身免疫性病症）、牛皮癬和動脈粥樣硬化。該等聚合物可以作為活性劑單獨使用，或者作為其他治療劑（諸如藥物分子或用於基因治療的核酸）的遞送載體。在此類情況下，聚合物的自身固有的抗腫瘤活性可以補足有待遞送的藥劑的活性。該等專利出版物的揭露內容可以與在此的教授內容結合使用，以用於遞送一種或多種CRISPR Cas系統或其一種或多種組分或對其編碼的一個或多個核酸分子。

超電荷蛋白

[0679] 超電荷蛋白係一類具有非常高的正或負的理論淨電荷的工程化或天然存在的蛋白質並且可以用於遞送一種或多種CRISPR Cas系統或其一種或多種組分或對其編碼的一個或多個核

酸分子。超負電荷蛋白和超正電荷蛋白兩者都表現出顯著的抵抗熱誘導或化學誘導的聚集的能力。超正電荷蛋白還能夠穿透哺乳動物細胞。使貨物諸如質粒DNA、RNA或其他蛋白質與該等蛋白質締合可以使得該等大分子到體外和體內的哺乳動物細胞中的功能遞送成為可能。劉大衛實驗室（David Liu's lab）在2007年報導了超電荷蛋白的創建和表徵（勞倫斯（Lawrence）等人，2007，美國化學學會雜誌 129, 10110–10112）。

[0680] RNA和質粒DNA到哺乳動物細胞中的非病毒遞送對於研究和治療應用都是有價值的（阿肯克（Akinc）等人，2010，自然生物技術 26, 561–569）。純化的+36 GFP蛋白（或其他超正電荷蛋白）與RNA在適當的無血清培養基中混合並且使得其在添加到細胞中之前複合。在這個階段血清的包含抑制超電荷蛋白-RNA複合物的形成並且降低治療效果。已經發現以下方案對於多種細胞系係有效的（麥克諾頓（McNaughton）等人，2009，美國國家科學院院刊 106, 6111–6116）（然而，應當進行改變蛋白質和RNA劑量的預試驗來優化用於特異性細胞系的程序）：

[0681] (1) 在治療前一天，以1×10^5個細胞/孔接種於48孔板中。

[0682] (2) 在治療當天，將純化的+36 GFP蛋白在無血清的培養基中稀釋至終濃度200 nM。添加RNA到50 nM的終濃度。渦旋混合並且在室溫下培養10 min。

[0683] (3) 在培養過程中，從細胞抽出培養基並且再次用PBS洗滌。

[0684] (4) 在培養+36 GFP和RNA之後，向細胞添加蛋白質-RNA複合物。

[0685] (5) 將細胞與複合物在37°C下培養4 h。

[0686] (6) 在培養之後，抽出培養基並且用20 U/mL的肝素PBS洗滌三次。用含血清的培養基另外培養細胞48 h或更長，這取決於用於活性的測定。

[0687] (7) 藉由免疫印跡、qPCR、表型分析或其他適當的方法分析細胞。

[0688] 劉大衛實驗室已經進一步發現+36 GFP在一系列細胞中是有效的質粒遞送試劑。由於質粒DNA係比siRNA大的貨物，有效複合質粒需要成比例地更大的+36 GFP蛋白。為了有效質粒遞送，申請人已經開發了帶有C末端HA2肽標籤的+36 GFP變體，這種肽係已知的來源於流感病毒血凝素蛋白的內體破壞肽。以下方案在多種細胞中是有效的，但是如上所述，建議針對特異性細胞系和遞送應用優化質粒DNA和超電荷蛋白的劑量。

[0689] (1) 在治療前一天，以1×10^5/孔接種於48孔板中。(2) 在治療當天，

[0690] 將純化的þ36 GFP蛋白在無血清的培養基中稀釋至終濃度2 mM。添加1 mg質粒DNA。渦旋混合並且在室溫下培養10 min。

[0691] (3) 在培養過程中，從細胞抽出培養基並且再次用PBS洗滌。

[0692] (4) 在培養þ36 GFP和質粒DNA之後，向細胞輕輕添

加蛋白質-DNA複合物。

[0693] (5) 將細胞與複合物在37°C下培養4 h。

[0694] (6) 在培養之後，抽出培養基並且用PBS洗滌。在含血清培養基中培養細胞，並且另外培養24-48 h。

[0695] (7) 在適當時分析質粒遞送（例如，藉由質粒驅動的基因表現）。

[0696] 還參見，例如，麥克諾頓等人，美國國家科學院院刊106, 6111-6116 (2009)；克羅尼肯（Cronican）等人，ACS化學生物學（ACS Chemical Biology） 5, 747-752 (2010)；克羅尼肯等人，化學與生物學（Chemistry & Biology） 18, 833-838 (2011)；湯普森（Thompson）等人，酶學方法（Methods in Enzymology） 503, 293-319 (2012)；湯普森，D.B.等人，化學與生物學 19 (7), 831-843 (2012)。超電荷蛋白的該等方法可以用於和/或適於本發明的CRISPR Cas系統的遞送。劉博士的該等系統和在此的文獻結合在此的教授內容可以用於遞送一種或多種CRISPR Cas系統或其一種或多種組分或對其編碼的一個或多個核酸分子。

細胞穿透肽（CPP）

[0697] 在又另一個實施方式中，考慮了細胞穿透肽（CPP）用於CRISPR Cas系統的遞送。CPP係促進不同分子貨物（從奈米級粒子至小化學分子和大的DNA片段）的細胞攝取的短肽。如在此所用的術語"貨物"包括但不限於下組，該組由下項組成：治療劑、診斷性探針、肽、核酸、反義寡核苷酸、質粒、蛋白質、粒子

（包括奈米粒子）、脂質體、發色團、小分子以及放射性物質。在本發明的方面中，貨物還可以包括CRISPR Cas系統的任何組分或整個功能性CRISPR Cas系統。本發明的方面進一步提供了用於將所希望的貨物遞送至受試者中的方法，該等方法包括：(a) 製備包含本發明的細胞穿透肽和所希望的貨物的複合物，並且 (b) 向受試者口服地、關節內地、腹膜內地、鞘內地、動脈內地（intrarterially）、鼻內地、實質內地（intraparenchymally）、皮下地、肌內地、靜脈內地、真皮地、直腸內地或局部地給予複合物。貨物藉由經由共價鍵的化學連接或藉由非共價的相互作用與肽締合。

[0698] CPP的功能是將貨物遞送至細胞中，這係通常藉由胞吞作用發生的過程，其中貨物被遞送至活哺乳動物細胞的內體。細胞穿透肽具有不同的尺寸、胺基酸序列並且帶電荷，但是所有CPP具有獨特的特性，該特性係易位質膜的能力，並且促進不同分子貨物到細胞質或細胞器的遞送。CPP易位可以被分類成三種主要的進入機制：直接穿透膜中、胞吞作用介導的進入，以及藉由暫態結構的形成的易位。CPP在醫學中發現了許多應用，在治療不同疾病包括癌症中作為藥物遞送劑，和病毒抑制劑以及用於細胞標記的造影劑。後者的實例包括充當用於GFP、MRI造影劑或量子點的載體。CPP作為用於研究和醫學的體外及體內遞送載體具有極大潛力。CPP典型地具有下述的胺基酸組成，該胺基酸組成含有高相對豐度的帶正電荷的胺基酸諸如賴胺酸或精胺酸或具有含有極性/帶電荷胺基酸和非極性、疏水性胺基酸的交替圖案的序列。這兩種結構類型被分別稱之為聚陽離子或兩親性分子。CPP的第

三種類別係疏水性肽，其僅含有具有低淨電荷極性殘基或具有對細胞攝取係關鍵的疏水性胺基酸基團。所發現的原始CPP中之一係來自人類免疫缺陷病毒1（HIV-1）的反啟動轉錄活化物（Tat），發現該Tat被培養物中的多種細胞類型從周圍培養基中有效攝取。從此以後，多種已知的CPP得到了相當地擴展並且產生了具有更有效的效應蛋白轉導特性的小分子合成類似物。CPP包括但不限於穿透素、Tat(48-60)、轉運素和(R-AhX-R4)(Ahx=胺基己醯基)。

[0699] 美國專利8,372,951提供了來源於嗜酸性粒細胞陽離子蛋白（ECP）的CPP，該CPP表現出非常高的細胞穿透效率和低毒性。還提供了將帶有其貨物的CPP遞送至脊椎動物受試者中的方面。CPP的另外方面和其遞送描述於美國專利8,575,305；8,614,194和8,044,019中。CPP可以用於遞送CRISPR-Cas系統或其組分。可以用於遞送CRISPR-Cas系統或其組分的CPP也被提供於蘇雷什·羅摩克裡希納(Suresh Ramakrishna)、阿布誇庫戴德(Abu-Bonsrah Kwaku Dad)、賈格迪什·拜洛兒（Jagadish Beloor）等人寫的手稿“藉由細胞穿透肽介導的Cas9蛋白和指導RNA的遞送進行的基因破壞（Gene disruption by cell-penetrating peptide-mediated delivery of Cas9protein and guide RNA）”，基因組研究（Genome Res.），2014年4月2日，[電子版先於印刷版]，該文獻藉由引用以其整體結合在此，其中展示出的是用CPP軛合的重組Cas9蛋白和CPP複合的指導RNA的處理導致人類細胞系中的內源性基因破壞。在論文中，Cas9蛋白經由硫醚鍵軛合至CPP，而指導RNA與CPP複合，形成了稠合的帶正電荷的粒子。已表明，用修飾的Cas9和指導RNA對人類細胞(包括胚胎幹細胞、真皮成纖維細胞、HEK293T

細胞、HeLa細胞和胚胎癌細胞）的同時處理和依序處理產生有效的基因破壞，其中相對於質粒轉染脫靶突變減少。

可植入裝置

[0700] 在另一個實施方式中，還考慮了可植入裝置用於遞送CRISPR Cas系統或其一種或多種組分或對其編碼的一個或多個核酸分子。例如，美國專利公開20110195123揭露了可植入的醫療器械，其局部地且在一長時間段內洗提藥物，包括了若干種類型的此設備、實施的治療方式和植入方法。該裝置包含聚合物基材，諸如，例如用作裝置主體的基質，以及藥物，並且在一些情況下包含另外的支架材料，諸如金屬或另外的聚合物，以及增強能見度和成像的材料。可植入的遞送裝置在提供局部且一長時間段內的釋放方面可能是有利的，其中藥物直接釋放到患病區域諸如腫瘤、炎症、退化的細胞外基質（ECM），或用於針對症狀的目的，或者釋放到損傷的平滑肌細胞，或者用於預防。一種藥物係如以上揭露的RNA，並且這個系統可以用於和/或適於本發明的CRISPR Cas系統。在一些實施方式中，植入方式係針對包括近距離放射療法和針吸活組織檢查的其他治療的當今開發和使用的現有植入程序。在這樣的情況下，在本發明中描述的新植入物的尺寸類似於原始植入物。典型地，在同一的治療程序中，植入了幾個裝置。

[0701] 美國專利公開20110195123提供了一種藥物遞送可植入或可插入系統，包括適用於空腔諸如腹腔和/或其中藥物遞送系統未被錨定或附接的任何其他類型的給藥的系統，該等系統包括

生物穩定的和/或可降解的和/或生物可吸收的聚合物基材，該基材可以例如視情況是基質。應當指出的是術語“插入”也包括植入。該藥物遞送系統較佳的是如美國專利公開20110195123中描述的“裝填器（Loder）”那樣實施。

[0702] 聚合物或多種聚合物係生物相容的，其結合一種藥劑和/或多種藥劑，使得藥劑以控制的速率釋放，其中該聚合物基材諸如基質的總體積，例如在一些實施方式中是視情況並且較佳的是不大於容許達到該藥劑的治療水平的最大體積。作為一非限制性實例，這樣的體積較佳的是在0.1 m^3至1000 mm^3的範圍內，正如該藥劑負荷的體積所要求的。該裝填器視情況是較大的，例如當結合有其尺寸由功能性決定的裝置例如而不限於，膝關節、宮內節育環或子宮頸環等時。

[0703] 在一些實施方式中，該藥物遞送系統（用於遞送該組成物）被設計為較佳的是採用可降解聚合物，其中主要釋放機制係本體溶蝕（bulk erosion）；或者在一些實施方式中，使用了不可降解的、或緩慢降解的聚合物，其中主要釋放機制係擴散而不是本體溶蝕，使得外部部分用作膜，並且其內部部分用作藥物貯庫，該藥物貯庫在延長的時間段內（例如從約一週至約幾個月）實際上不受環境的影響。還可以視情況使用具有不同釋放機制的不同聚合物的組合。在總藥物釋放期的重要時段期間，在表面處的濃度梯度較佳的是維持為有效恒定，並且因此擴散速率係有效恒定的（稱為“零模式”擴散）。關於術語“恒定”，它意指較佳的是維持在治療有效性的低閾值以上的擴散速率，但是可以仍然視情況具

有初期突釋的特徵和/或可以發生波動，例如增加和降低到一定程度。擴散速率較佳的是被如此維持一長時間段，並且可以考慮使它相對於一定的水平是恒定的，以便優化治療有效期，例如有效沈默期。

[0704] 藥物遞送系統視情況並且較佳的是被設計為保護基於核苷酸的治療劑免於降解，而無論是化學性質還是由於受試者體內的酶和其他因素的攻擊。

[0705] 美國專利公開20110195123的藥物遞送系統視情況與感測和/或啟動設備相關聯，該等設備藉由啟動和/或加速/減速的無創和/或微創方法在該裝置的植入之時和/或之後被操作，該等方法例如視情況包括但不限於熱力加熱和冷卻、雷射光束和超音波，包括聚焦超音波和/或RF（射頻）方法或裝置。

[0706] 根據美國專利公開20110195123的一些實施方式，用於局部遞送的位點可以視情況包括特徵為高度異常的細胞增殖和受抑制的細胞凋亡的靶位點，包括腫瘤、活動性和/或慢性炎症和感染，包括自身免疫性疾病狀態、退化組織（包括肌肉和神經組織）、慢性疼痛、退行性位點，以及用於增強組織再生的骨折位置以及其他傷口位置，以及損傷的心肌、平滑肌和橫紋肌。

[0707] 用於植入該組成物的位點、或靶位點，較佳的是其特徵為用於靶向局部遞送的足夠小的半徑、面積和/或體積。例如，該靶位點視情況具有在從約0.1 mm至約5 cm範圍內的直徑。

[0708] 該靶位點的位置較佳的是針對最大治療效力而選擇。

例如，該藥物遞送系統的組成物（視情況與如上所述的用於植入的裝置一起）視情況並且較佳的是被植入在腫瘤環境或與腫瘤環境相關聯的血供之內或附近。

[0709] 例如該組成物（視情況與該裝置一起）視情況植入在胰臟、前列腺、乳房、肝臟之內或附近，經由接管（nipple）進行，植入在血管系統之內，等等。

[0710] 靶位置視情況選自下組，該組包括下項、基本上由、或由下項組成（僅僅作為非限制性實例，因為視情況身體內的任何位點可以適合於植入裝填器）：1.在退行性位點處的腦，像在帕金森病或阿茲海默症中在基底神經節、白質和灰質處；2.如在肌萎縮側索硬化（ALS)的情況下的脊柱；3.預防HPV感染的子宮頸；4.活動性或慢性炎性關節；5.在牛皮癬情況下的真皮；6.用於止痛作用的交感神經位點和感覺神經位點；7.骨內植入；8.急性和慢性感染位點；9.陰道內；10.耳內--聽覺系統、內耳的迷路、前庭系統；11.氣管內；12.心內；冠狀動脈、心外膜；13.膀胱；14.膽道系統；15.實質組織，包括但不限於腎、肝臟、脾；16.淋巴結；17.唾液腺；18.牙齦；19.關節內（進入關節）；20.眼內；21.腦組織；22.腦室；23.空腔，包括腹腔（例如但不限於，卵巢癌）；24.食管內以及25.直腸內。

[0711] 視情況，該系統（例如含有該組成物的裝置）的插入與向在靶位點處和該位點附近的ECM注射材料相關聯，從而影響該靶位點和此位點附近的ECM中的局部pH和/或溫度和/或影響該藥物擴散和/或藥物動力學的其他生物因素。

[0712] 視情況，根據一些實施方式，所述藥劑的釋放可以與感測和/或啟動設備相關聯，該等設備藉由啟動和/或加速/減速的無創和/或微創方法和/或別的方法在插入之前和/或之時和/或之後被操作，所述方法包括雷射光束、放射、熱力加熱和冷卻、和超音波，包括聚焦超音波和/或RF（射頻）方法或裝置、以及化學啟動劑。

[0713] 根據美國專利公開20110195123的其他實施方式，藥物較佳的是包括RNA，例如，對於局限性癌症情況，在乳房、胰臟、腦、腎、膀胱、肺以及前列腺中，如下文所述。儘管使用RNAi進行舉例說明，但是許多藥物係適用於封裝在裝填器中的，並且可以與本發明結合使用，只要此類藥物可以用裝填器基材例如像基質封裝，並且此系統可以用於和/或適於遞送本發明的CRISPR Cas系統。

[0714] 作為特殊應用的另一個實例，神經肌肉退行性疾病由於異常基因表現而發生。RNA的局部遞送可以具有干擾此異常基因表現的治療特性。包括小藥物和大分子的抗凋亡、抗炎症和抗退行性藥物的局部遞送也可以視情況是治療性的。在這樣的情況下，該裝填器用於以恒定速率和/或藉由單獨植入的專用裝置延長釋放。這都可以用於和/或適於本發明的CRISPR Cas系統。

[0715] 作為特殊應用的又另一個實例，用基因修飾劑治療精神和認知障礙。基因敲低係一治療選擇。向中樞神經系統位點局部遞送藥劑的裝填器係對於精神障礙和認知障礙的治療選擇，該等精神障礙和認知障礙包括但不限於，精神病、雙極疾病、神經

性病症和行為疾病（behavioral maladies）。該等裝填器也可以在特定腦位點進行植入時局部遞送包括小藥物和大分子的藥物。這都可以用於和/或適於本發明的CRISPR Cas系統。

[0716] 作為特殊應用的另一個實例，在局部位點的先天性和/或適應性免疫介質的沈默能夠預防器官移植排斥。用植入到移植器官和/或植入位點中的裝填器局部遞送RNA和免疫調節試劑使得藉由排斥性免疫細胞（諸如針對移植器官而被啟動的CD8）產生局部免疫抑制。這都可以用於和/或適於本發明的CRISPR Cas系統。

[0717] 作為特殊應用的另一個實例，包括VEGF和血管生成素及其他的血管生長因子對於新血管形成係必需的。該等因子、肽、肽類比物的局部遞送或抑制它們的阻遏物係重要的治療模式；使阻遏物沈默以及用裝填器局部遞送刺激血管發生的該等因子、肽、大分子和小藥物對於周圍血管疾病、全身性血管疾病和心血管疾病係具有治療性的。

[0718] 插入的方法，諸如植入，可以視情況已用於其他類型的組織植入和/或用於插入和/或用於組織取樣，視情況在此類方法中沒有修改，或者可替代地視情況僅僅具有非重點修改。此類方法視情況包括但不限於，近距離放射療法、活組織檢查、用和/或不用超音波的內窺鏡檢查諸如ERCP、進入腦組織的立體定位法、腹腔鏡檢查，包括用腹腔鏡進入關節、腹器官、膀胱壁和體腔的植入。

[0719] 在此所討論的可植入裝置技術可以與在此的教授內容一起使用並且因此，藉由本揭露和本領域知識，CRISPR-Cas系

統或其組分或其核酸分子或編碼組分或提供組分的核酸分子可以經由可植入裝置遞送。

患者特異性篩選方法

[0720] 靶向DNA，例如三核苷酸重複序列的核酸靶向系統可以用於篩選存在此類重複序列的患者或患者樣品。重複序列可以是核酸靶向系統的RNA的靶標，並且如果藉由核酸靶向系統與其存在結合，則可以檢測出該結合，從而表明此類序列存在。因此，核酸靶向系統可以用於篩選存在此類重複序列的患者或患者樣品。然後可以向患者給予一種或多種適合的化合物以解決此病狀；或可以給予核酸靶向系統以結合此病狀並且產生插入、缺失或突變並且緩解此病狀。

[0721] 本發明使用核酸結合靶DNA序列。

CRISPR效應蛋白mRNA和指導RNA

[0722] 也可以單獨遞送CRISPR酶mRNA和指導RNA。CRISPR酶mRNA可以在指導RNA在給出時間以待CRISPR酶表現之前遞送。CRISPR酶mRNA可以在給予指導RNA之前1-12小時（較佳的是約2-6小時）給予。

[0723] 可替代地，CRISPR酶mRNA和指導RNA可以一起給予。有利地，指導RNA的第二加強劑量可以在初始給予CRISPR酶mRNA + 指導RNA之後1-12小時（較佳的是約2-6小時）給予。

[0724] 本發明的CRISPR效應蛋白，即Cpf1效應蛋白在此有時稱之為CRISPR酶。應瞭解的是，效應蛋白係基於或來源於酶，

所以術語“效應蛋白”當然包括一些實施方式中的“酶”。然而，還應瞭解的是，根據在一些實施方式中的需要，效應蛋白可以具有DNA或RNA結合，但是不一定具有切割或切口活性，包括無效Cas效應蛋白功能。

[0725] 為了實現最有效的基因組修飾水平，CRISPR酶mRNA和/或指導RNA的附加給予可能是有用的。在一些實施方式中，當特別是在治療方法中遺傳疾病被靶向時，表型改變較佳的是基因組修飾的結果並且較佳的是其中提供了修復模板以校正或改變表型。

[0726] 在一些實施方式中，可以被靶向的疾病包括與引起疾病的剪接缺陷相關的那些。

[0727] 在一些實施方式中，細胞靶標包括造血幹細胞/祖細胞（CD34+）；人類T細胞；以及眼（視網膜細胞）-例如光受體先質細胞。

[0728] 在一些實施方式中，基因靶包括：人類β球蛋白-HBB（用於治療鐮狀細胞貧血，包括藉由刺激基因轉變（使用緊密相關的HBD基因作為內源性模板）進行）；CD3（T細胞）；以及CEP920-視網膜（眼）。

[0729] 在一些實施方式中，疾病靶標也可以包括：癌症；鐮狀細胞貧血（基於點突變）；HIV；β-地中海貧血；以及眼睛或眼部疾病-例如引起萊伯氏先天性黑矇（Leber Congenital Amaurosis）（LCA）的剪接缺陷。

[0730] 在一些實施方式中，遞送方法包括：酶-指導序列複合物（核糖核蛋白）的陽離子脂質介導的“直接遞送”以及質粒DNA的電穿孔。

[0731] 本發明方法可以進一步包括模板的遞送，諸如修復模板，該等修復模板可以是dsODN或ssODN，參見下文。模板遞送可以是經由與任一或所有CRISPR酶或指導序列的遞送同時發生或分開並且經由相同遞送機制或不同遞送機制。在一些實施方式中，較佳的是，模板與指導序列一起遞送，並且也較佳的是與CRISPR酶一起遞送。一實例可以是AAV載體。

[0732] 本發明方法可以進一步包括：(a) 將包含互補於由所述雙股斷裂創建的突出端的突出端的雙股寡去氧核苷酸（dsODN）遞送至細胞，其中所述dsODN被整合到感興趣的座位中；或–(b)將單股寡去氧核苷酸（ssODN）遞送至細胞，其中所述ssODN充當用於所述雙股斷裂的同源定向修復的模板。本發明方法可以用於預防或治療個體的疾病，視情況其中所述疾病由所述感興趣的座位中的缺陷導致。本發明方法可以在個體中體內進行或者在取自個體的細胞上離體進行，視情況其中所述細胞被返回至個體。

[0733] 為了最小化毒性和脫靶效應，重要的是控制所遞送的CRISPR酶mRNA和指導RNA的濃度。CRISPR酶mRNA和指導RNA的最佳濃度可以藉由以下方式來確定：測試不同濃度的細胞模型或動物模型並且使用深度定序分析潛在的脫靶基因組座位處的修飾程度。例如，對於人類基因組的EMX1基因中的指導序列靶向5'-GAGTCCGAGCAGAAGAAGAA-3'（SEQ ID NO: 23），深度定序可

以用於評估以下兩個脫靶座位處的修飾水平：1：5'-GAGTCCTAGCAGGAGAAGAA-3'（SEQ ID NO: 24）和2：5'-GAGTCTAAGCAGAAGAAGAA-3'（SEQ ID NO: 25）。得到最高的中靶修飾水平同時使脫靶修飾水平最小化的濃度應被選擇用於體內遞送。

誘導型系統

[0734] 在一些實施方式中，CRISPR酶可以形成誘導型系統的組分。該系統的誘導性質允許使用能量形式時間空間控制基因編輯或基因表現。能量形式可以包括但不限於，電磁輻射、聲能、化學能以及熱能。誘導型系統的實例包括四環素誘導型啟動子（Tet-開或Tet-關）、小分子雙雜交轉錄啟動系統（FKBP、ABA等）、或光誘導型系統（光敏色素、LOV結構域或隱花色素）。在一個實施方式中，CRISPR酶可以是光誘導型轉錄效應物（LITE）的一部分，從而以序列特異性方式引導轉錄活性的變化。光誘導型系統的組分可以包括CRISPR酶、光反應性細胞色素異源二聚體（例如，來自阿拉伯芥）、以及轉錄啟動/阻遏結構域。誘導型DNA結合蛋白及其使用方法的其他實例提供於US 61/736,465和US 61/721,283以及WO 2014/018423 A2中，該等專利藉由引用以其整體結合在此。

自失活系統

[0735] 一旦細胞基因組中的基因的所有拷貝已被編輯,則該細胞中的連續的CRISRP/Cpf1p表現不再需要。實際上，持續的表現在非預定基因組位點等處的脫靶效應情況下將是不希望的。因此，時間限制的表現將是有用的。誘導型表現提供了一途徑，但

是此外，申請人已經工程化出依賴於CRISPR載體本身內的非編碼指導靶序列的用途的自失活CRISPR系統。因此，在表現開始之後，CRISPR-Cas系統將使得其自身破壞，但是在完全破壞之前，其將有編輯靶基因的基因組拷貝的時間（在二倍體細胞中的正常點突變的情況下，其需要至多兩次編輯）。簡單地，自失活CRISPR-Cas系統包括靶向CRISPR酶本身的編碼序列或靶向與存在於以下項中的一種或多種中的獨特序列互補的一種或多種非編碼指導靶序列的附加RNA（即指導RNA）：

(a) 在驅動非編碼RNA元件的表現的啟動子之內，

(b) 在驅動Cpf1效應蛋白基因的表現的啟動子之內，

(c) 在Cpf1效應蛋白編碼序列中的ATG翻譯起始密碼子的100 bp之內，

(d) 在病毒遞送載體，例如在AAV基因組中反向末端重複序列（iTR）之內。

[0736] 此外，RNA可以經由載體，例如單獨的載體或編碼CRISPR複合物的同一載體來遞送。當藉由單獨的載體來遞送時，靶向Cas表現的CRISPR RNA可以依序或同時給予。當依序給予時，在意圖用於例如基因編輯或基因工程化的CRISPR RNA之後，將遞送靶向Cas表現的CRISPR RNA。此時間段可以是數分鐘（例如5分鐘、10分鐘、20分鐘、30分鐘、45分鐘、60分鐘）的時間。此時間段可以是數小時（例如2小時、4小時、6小時、8小時、12小時、24小時）的時間。此時間段可以是數天（例如2天、3天、4天、7天）

的時間。此時間段可以是數週（例如2週、3週、4週）的時間。此時間段可以是數月（例如2個月、4個月、8個月、12個月）的時間。此時間段可以是數年（例如2年、3年、4年）的時間。在此方式中，Cas酶與能夠與第一靶標諸如感興趣的一個基因組座位或多個基因組座位雜交的第一gRNA締合並且負責所希望的CRISPR-Cas系統的一種或多種功能（例如，基因工程化）；並且隨後Cas酶可以接著與能夠與包含至少一部分的Cas或CRISPR盒的序列雜交的第二gRNA締合。在指導RNA靶向編碼Cas蛋白的表現的序列的情況下，該酶受到阻礙並且系統發生自失活。以相同方式，經由如在此解釋的例如脂質體、脂轉染、粒子、微泡施用的靶向Cas表現的CRISPR RNA可以依序或同時給予。類似地，自失活可以用於對用來靶向一個或多個靶標的一個或多個指導RNA進行失活。

[0737] 在一些方面中，提供了單一gRNA，該單一gRNA能夠與CRISPR酶起始密碼子下游的序列雜交，由此在一段時間後，存在CRISPR酶表現的喪失。在一些方面中，提供了一個或多個gRNA，該等gRNA能夠與編碼CRISPR-Cas系統的多核苷酸的一個或多個編碼或非編碼區雜交，由此在一段時間後，存在一種或多種、或在一些情況下全部的CRISPR-Cas系統的失活。在系統的一些方面中，並且不受理論限制，細胞可以包含多種CRISPR-Cas複合物，其中第一亞組的CRISPR複合物包含能夠靶向有待編輯的一個基因組座位或多個基因組座位的第一指導RNA，並且第二亞組的CRISPR複合物包含能夠靶向編碼CRISPR-Cas系統的多核苷酸的至少一個第二指導RNA，其中第一亞組的CRISPR複合物介導靶向的一個基因組座位或多個基因組座位的編輯並且第二亞組的

CRISPR複合物最終使CRISPR-Cas系統失活，從而使細胞中的進一步CRISPR-Cas表現失活。

[0738] 因此，本發明提供了一種包含用於遞送至真核細胞的一種或多種載體，其中一種或多種載體編碼：(i) CRISPR酶；(ii) 能夠雜交至細胞中的靶序列的第一指導RNA；(iii) 能夠雜交至編碼CRISPR酶的載體中的一個或多個靶序列的第二指導RNA，當在該細胞中表現時，第一指導RNA引導第一CRISPR複合物與該細胞中的靶序列的序列特異性結合；第二指導RNA引導第二CRISPR複合物與編碼CRISPR酶的載體中的靶序列的序列特異性結合；CRISPR複合物包含結合指導RNA的CRISPR酶，由此使得指導RNA可以與其靶序列雜交；並且第二CRISPR複合物使CRISPR-Cas系統失活以阻止細胞對CRISPR酶的連續表現。

[0739] 不同編碼序列（CRISPR酶和指導RNA）可以包含在單一載體上或多個載體上。例如，有可能編碼在一個載體上的酶和在另一個載體上的不同RNA序列，或者有可能編碼在一個載體上的酶和一個指導RNA以及在另一個載體上的剩餘指導RNA或任何其他前突變。總的來說，使用總共一個或多個不同載體的系統係較佳的。

[0740] 在使用多種載體的情況下，有可能以不相等的數目遞送它們，並且理想的是，其中編碼第一指導RNA的載體相對於編碼第二指導RNA的載體係過量的，從而有助於延遲CRISPR系統的最終失活，直到基因組編輯具有了發生的機會為止。

[0741] 第一指導RNA可以靶向基因組內的感興趣的任何靶

序列，如在此其他地方所述的。第二指導RNA靶向編碼CRISPR Cpf1酶的載體內的序列，並且從而使來自該載體的酶的表現失活。因此，載體中的靶序列必須能夠使表現失活。適合的靶序列可以是例如在Cpf1p編碼序列的翻譯起始密碼子附近或之內，在驅動非編碼RNA元件的表現的啟動子中的非編碼序列中，在驅動Cpf1p基因的表現的啟動子之內，在Cas編碼序列中的ATG翻譯起始密碼子的100 bp之內，和/或在病毒遞送載體，例如AAV基因組中的反向末端重複序列（iTR）之內。靠近此區域的雙股斷裂可以誘導Cas編碼序列的移碼，使得蛋白質表現喪失。用於使指導RNA“自失活”的替代性靶序列將旨在編輯/失活為CRISPR-Cpf1系統的表現或為載體的穩定性所需要的調節區/序列。例如，如果Cas編碼序列的啟動子被破壞，那麼轉錄可以被抑制或阻止。類似地，如果載體包含用於複製、維持性或穩定性的序列，那麼有可能靶向該等序列。例如，在AAV載體中，有用的靶序列係在iTR之內。其他有用的供靶向的序列可以是啟動子序列、多聚腺苷酸化（polyadenlyation）位點等。

[0742] 此外，如果指導RNA以陣列格式表現，則同時靶向兩個啟動子的“自失活”指導RNA將使得間插核苷酸從CRISPR-Cas表現構建體內切除，有效地使得其完全失活。類似地，在指導RNA靶向兩個ITR，或同時靶向兩種或更多種其他CRISPR-Cas組分的情況下，發生間插核苷酸的切除。總的來說，如在此解釋的自失活係適用於CRISPR-Cas系統的，以便提供CRISPR-Cas的調節。例如，如在此解釋的自失活可以適用於如在此解釋的突變，例如擴增病症的CRISPR修復。作為此自失活的結果，CRISPR修復僅僅具

有暫態活性。

[0743] 向“自失活”指導RNA的5'端添加非靶向核苷酸（例如1-10個核苷酸，較佳的是1-5個核苷酸）可以用於延遲其加工和/或修飾其效力以作為確保CRISPR-Cas停止之前的靶向的基因組座位處的編輯的手段。

[0744] 在自失活AAV-CRISPR-Cas系統的一個方面中，可以建立共表現感興趣的一種或多種指導RNA靶向基因組序列（例如1-2、1-5、1-10、1-15、1-20、1-30）的質粒，其中靶向SpCas9序列的“自失活”指導RNA處於或靠近工程化的ATG起始位點(例如，在5個核苷酸之內、在15個核苷酸之內、在30個核苷酸之內、在50個核苷酸之內、在100個核苷酸之內)。U6啟動子區中的調節序列也可以用指導RNA靶向。U6驅動的指導RNA可以被設計為陣列格式，使得多個指導RNA序列可以同時被釋放。當首先被遞送至靶組織/細胞（離開的細胞）時，指導RNA開始積累，同時Cas水平在核中上升。Cas與介導CRISPR-Cas質粒的基因組編輯和自失活的所有指導RNA複合。

[0745] 自失活CRISPR-Cas系統的一個方面係由1至4或更多個不同指導序列；例如高達約20或約30個指導序列以單獨或串聯的陣列格式的表現。每個單個自失活指導序列可以靶向不同的靶標。這樣可以從例如一嵌合pol3轉錄物開始加工。可以使用Pol3啟動子諸如U6或H1啟動子。Pol2啟動子諸如在此所提到的那些。反向末端重複(iTR)序列可以側接Pol3啟動子-一個或多個指導RNA-Pol2啟動子-Cas。

[0746] 串聯的陣列轉錄物的一個方面在於一種或多種指導序列編輯一個或多個靶標，而一個或多個自失活指導序列使CRISPR-Cas系統失活。因此，例如，用於修復擴增病症的所述CRISPR-Cas系統可以直接與在此所述的自失活CRISPR-Cas系統相結合。此系統可以例如具有針對供修復的靶區的兩個指導序列以及針對CRISPR-Cas的自失活的至少一個第三指導序列。參考申請案序號PCT/US2014/069897，題為“在核苷酸重複病症中使用Crispr-Cas系統的組成物和方法”，2014年12月12日以WO/2015/089351公開。

[0747] 指導RNA可以是控制指導序列。例如，其可以被工程化為靶向編碼CRISPR酶本身的核酸序列，如US2015232881A1中所描述，該專利的揭露內容藉由引用結合在此。在一些實施方式中，系統或組成物可以僅提供有被工程化為靶向編碼CRISPR酶的核酸序列的指導RNA。此外，該系統或組成物可以提供有被工程化為靶向編碼CRISPR酶的核酸序列以及編碼CRISPR和視情況第二指導RNA和另外視情況修復模板的核酸序列的指導RNA。第二指導RNA可以是CRISPR系統或組成物（如在此所限定的此治療性、診斷性、敲除性等）的主要靶標。以這種方式，該系統或組成物是自失活的。這係關於US2015232881A1（也如在此其他地方引用的WO2015070083 (A1)所公開的）中的Cas9進行舉例說明的，並且可以外推到Cpf1。

在多重（串聯）靶向方法中使用的根據本發明的酶

[0748] 發明者已證實，在此所限定的CRISPR酶可以採用沒

有喪失活性的超過一個的RNA指導序列。這使得能夠使用如在此所限定的CRISPR酶、系統或複合物用於靶向多個DNA靶標、基因或基因座位，其中單一酶、系統或複合物如在此所限定的。指導RNA可以是串聯地安排的，視情況藉由核苷酸序列諸如如在此所限定的同向重複序列分開。不同指導RNA的位置係使得串聯不影響活性。應注意的是，術語“CRISPR-Cas系統”、“CRISP-Cas複合物”、“CRISPR複合物”和“CRISPR系統”係可互換使用的。另外，術語“CRISPR酶”、“Cas酶”或“CRISPR-Cas酶”可以是可互換使用的。在較佳的實施方式中，所述CRISPR酶、CRISP-Cas酶或Cas酶係Cpf1，或者是在此其他地方所述的Cpf1的修飾或突變型變體中的任一種。

[0749] 在一個方面中，本發明提供了一非天然存在或工程化的CRISPR酶，較佳的是第2類CRISPR酶，較佳的是如在此所述的型V或VI CRISPR酶，諸如但不限於如在此其他地方所述的Cpf1，以用於串聯或多重靶向。應該理解的是，如在此其他地方所述的根據本發明的CRISPR（或CRISPR-Cas或Cas)酶、複合物或系統中的任一種可以用於此方法中。如在此其他地方所述的方法、產物、組成物和用途中的任一種係同樣適用於如下文進一步詳述的多重或串聯靶向方法。作為進一步指導，提供了以下的特定方面和實施方式。

[0750] 在一個方面中，本發明提供了如在此所限定的Cpf1酶、複合物或系統用於靶向多個基因座位的用途。在一個實施方式中，這可以藉由使用多個（串聯或多重）指導RNA（gRNA）序列來建

立。

[0751] 在一個方面中，本發明提供了用於使用如在所限定的Cpf1酶、複合物或系統中的一個或多個元件用於串聯或多重靶向的方法，其中所述CRISP系統包含多個指導RNA序列。較佳的是，所述gRNA序列藉由核苷酸序列，諸如如在此其他地方所限定的同向重複序列分開。

[0752] 如在此所限定的Cpf1酶、系統或複合物提供了一種用於修飾多個靶多核苷酸的有效手段。如在此所限定的Cpf1酶、系統或複合物具有多種多樣的效用，包括修飾（例如，缺失、插入、易位、失活、啟動）許多細胞類型中的一個或多個靶多核苷酸。這樣，如在此所限定的本發明Cpf1酶、系統或複合物在例如基因治療、藥物篩選、疾病診斷以及預後方面具有廣泛的應用，包括靶向單一CRISPR系統內的多個基因座位。

[0753] 在一個方面中，本發明提供了如在此所限定的Cpf1酶、系統或複合物，即具有Cpf1蛋白和多個指導RNA的Cpf1 CRISPR-Cas複合物，該Cpf1蛋白具有與其締合的至少一個去穩定化結構域，該等指導RNA靶向多個核酸分子諸如DNA分子，由此每個所述多個指導RNA特異性地靶向其相應的核酸分子，例如DNA分子。每個核酸分子靶向例如可以編碼基因產物或包括基因座位的DNA分子。因此使用多個指導RNA能夠靶向多個基因座位或多基因。在一些實施方式中，Cpf1酶可以切割編碼基因產物的DNA分子。在一些實施方式中，基因產物的表現被改變。Cpf1蛋白和指導RNA並不同時天然存在。本發明包括包含串聯地安排的指導序列的指

導RNA。本發明進一步包括密碼子優化為在真核細胞中表現的Cpf1蛋白的編碼序列。在一較佳的實施方式中，真核細胞係哺乳動物細胞、植物細胞或酵母細胞，並且在一個更較佳的實施方式中，哺乳動物細胞係人類細胞。基因產物的表現可以被減少。Cpf1酶可以形成CRISPR系統或複合物的一部分，該CRISPR系統或複合物進一步包含串聯安排的指導RNA（gRNA），包括2、3、4、5、6、7、8、9、10、15、25、25、30個或超過30個的一組指導序列，每個指導序列能夠與細胞中感興趣的基因組座位中的靶序列特異性地雜交。在一些實施方式中，功能性Cpf1 CRISPR系統或複合物結合多個靶序列。在一些實施方式中，功能性CRISPR系統或複合物可以編輯多種靶序列，例如靶序列可以包含基因組座位，並且在一些實施方式中，可以存在基因表現的改變。在一些實施方式中，功能性CRISPR系統或複合物可以進一步包含功能結構域。在一些實施方式中，本發明提供了一種用於改變或修飾多種基因產物的表現的方法。該方法可以包括引入到含有所述靶核酸，例如DNA分子，或含有和表現靶核酸，例如DNA分子的細胞中；例如，靶核酸可以編碼基因產物或提供基因產物（例如，調節序列）的表現。

[0754] 在較佳的實施方式中，用於多重靶向的CRISPR酶係Cpf1，或者CRISPR系統或複合物包含Cpf1。在一些實施方式中，用於多重靶向的CRISPR酶係AsCpf1，或者用於多重靶向的CRISPR系統或複合物包含AsCpf1。在一些實施方式中，CRISPR酶係LbCpf1，或者CRISPR系統或複合物包含LbCpf1。在一些實施方式中，用於多重靶向的Cpf1酶切割DNA的兩條股以產生雙股斷裂

（DSB）。在一些實施方式中，用於多重靶向的CRISPR酶係切口酶。在一些實施方式中，用於多重靶向的Cpf1酶係雙重切口酶。在一些實施方式中，用於多重靶向的Cpf1酶係Cpf1酶，諸如如在此其他地方所限定的DD Cpf1酶。

[0755] 在一些一般實施方式中，用於多重靶向的Cpf1酶與一個或多個功能結構域締合。在一些更具體的實施方式中，用於多重靶向的CRISPR酶係如在此其他地方所限定的無效Cpf1。

[0756] 在一個方面中，本發明提供了一種用於遞送如在此所限定的多靶向中使用的Cpf1酶、系統或複合物，或如在此所限定的多核苷酸的手段。此類遞送手段的非限制性實例係例如一個或多個粒子，該一個或多個粒子遞送複合物的一種或多種組分、包含在此所討論的一種或多種多核苷酸（例如編碼CRISPR酶，提供編碼CRISPR複合物的核苷酸）的一種或多種載體。在一些實施方式中，載體可以是質粒或病毒載體諸如AAV或慢病毒。使用質粒暫態轉染到例如HEK細胞中可能是有利的，特別是考慮到AAV的大小限制以及考慮到雖然Cpf1適配於AAV，但是可能達到另外的指導RNA的上限。

[0757] 還提供了一種組成型地表現如在此使用的Cpf1酶、複合物或系統的模型，該模型用於多重靶向中。生物體可以是轉基因的並且可以用本發明的載體轉染或者可以是如此轉染的生物體的後代。在另一個方面中，本發明提供了包含如在此所限定的CRISPR酶、系統和複合物，或如在此所述的多核苷酸或載體的組成物。還提供了包含較佳的是呈串聯安排格式的多個指導RNA的

Cpf1 CRISPR系統或複合物。所述不同的指導RNA可以藉由核苷酸序列諸如同向重複序列分開。

[0758] 還提供了一種治療受試者，例如有需要的受試者之方法，該方法包括藉由用編碼Cpf1 CRISPR系統或複合物的多核苷酸或在此所述的多核苷酸或載體中的任一種轉化受試者來誘導基因編輯，並且向受試者給予它們。還可以提供適合的修復模板，例如藉由包含所述修復模板的載體來遞送。還提供了一種治療受試者，例如有需要的受試者的之方法，該方法包括藉由用在此所述的多核苷酸或載體轉化受試者來誘導多個靶基因座位的轉錄啟動或阻遏，其中所述多核苷酸或載體編碼或包含Cpf1酶、包含較佳的是呈串聯安排的多個指導RNA的複合物或系統。在離體發生，例如在細胞培養物中發生任何處理的情況下，那麼應瞭解的是，術語“受試者”可以藉由短語“細胞或細胞培養物”來替換。

[0759] 提供了用於如在此任何地方所限定的治療方法中的組成物，該等組成物包含Cpf1酶、包含較佳的是呈串聯安排的多個指導RNA的複合物或系統，或編碼或包含所述Cpf1酶、包含較佳的是呈串聯安排的多個指導RNA的複合物或系統的多核苷酸或載體。可以提供包含此類組成物的成套套組。還提供了所述組成物在用於此類治療方法的藥物的製造中的用途。還藉由本發明提供了Cpf1 CRISPR系統在篩選，例如增功能篩選中的用途。被人工驅使過表現基因的細胞能夠藉由負反饋回路來下調隨時間變化的基因（重新建立平衡）。在開始篩選的時候，未被調節的基因可能再一次減少。使用誘導型Cpf1活化物允許就在篩選之前誘導轉錄

並且因此使假陰性命中的可能性最小化。因此，藉由本發明在篩選，例如增功能篩選中的使用，假陰性結果的可能性可以被最小化。

[0760] 在一個方面中，本發明提供了一種工程化的非天然存在的CRISPR系統，該CRISPR系統包含Cpf1蛋白和各自特異性地靶向編碼細胞中基因產物的DNA分子的多個指導RNA，由此多個指導RNA各自靶向它們的編碼基因產物的特異性DNA分子並且該Cpf1蛋白切割編碼該基因產物的靶DNA分子，由此改變該基因產物的表現；並且其中CRISPR蛋白和指導RNA並不同時天然存在。本發明包括包含較佳的是藉由核苷酸序列諸如同向重複序列分開的多個指導序列的多個指導RNA。在本發明的一實施方式中，CRISPR蛋白係V型或VI型CRISPR-Cas蛋白並且在一更較佳的實施方式中，CRISPR蛋白係Cpf1蛋白。本發明進一步包括密碼子優化為在真核細胞中表現的Cpf1蛋白。在一較佳的實施方式中，真核細胞係哺乳動物細胞，並且在一更較佳的實施方式中，哺乳動物細胞係人類細胞。在本發明的另一個實施方式中，基因產物的表現減少。

[0761] 在另一個方面中，本發明提供了一種包含一種或多種載體的工程化的非天然存在的載體系統，該等載體包含可操作地連接至各自特異性地靶向編碼基因產物的DNA分子的多個Cpf1 CRISPR系統指導RNA的第一調節元件以及編碼CRISPR蛋白的可操作地連接的第二調節元件。兩個調節元件可以位於系統的相同載體上或不同載體上。多個指導RNA靶向編碼細胞中多種基因產

物的多個DNA分子並且CRISPR蛋白可以切割編碼基因產物的多個DNA分子（該CRISPR蛋白可以切割一條或兩條股或者基本上不具有核酸酶活性），由此改變多種基因產物；並且其中CRISPR蛋白和多個指導RNA並不同時天然存在。在一較佳的實施方式中，CRISPR蛋白係Cpf1蛋白，其視情況被密碼子優化為在真核細胞中表現。在一較佳的實施方式中，真核細胞係哺乳動物細胞、植物細胞或酵母細胞，並且在一個更較佳的實施方式中，哺乳動物細胞係人類細胞。在本發明的另一個實施方式中，多種基因產物中的每種的表現被改變，較佳的是被減少。

[0762] 在一個方面中，本發明提供了一種包含一種或多種載體的載體系統。在一些實施方式中，該系統包含：(a) 可操作地連接至同向重複序列的第一調節元件和用於將一種或多種指導序列插入同向重複序列的上游或下游（無論哪一種都適用）的一個或多個插入位點，其中在表現時，一種或多種指導序列引導CRISPR複合物與真核細胞中的一個或多個靶序列的序列特異性結合，其中CRISPR複合物包含與雜交至一個或多個靶序列的一種或多種指導序列複合的Cpf1酶；和 (b) 可操作地連接至編碼較佳的是包含至少一個核定位序列和/或至少一個NES的所述Cpf1酶的酶編碼序列的第二調節元件；其中組分 (a) 和 (b) 位於系統的相同或不同的載體上。在一些實施方式中，組分 (a) 進一步包含可操作地連接至第一調節元件的兩種或更多種指導序列，其中當表現時，兩種或更多種指導序列中的每種引導Cpf1 CRISPR複合物與真核細胞中的不同靶序列的序列特異性結合。在一些實施方式中，CRISPR複合物包含具有足以驅動所述Cpf1 CRISPR複合物在真核

細胞的核中和/或之外以可檢測的量積累的強度的一個或多個核定位序列和/或一個或多個NES。在一些實施方式中，第一調節元件係聚合酶III啟動子。在一些實施方式中，第二調節元件係聚合酶II啟動子。在一些實施方式中，每個指導序列的長度係至少16、17、18、19、20、25個核苷酸，或介於16個-30個、或介於16個-25個、或介於16個-20個核苷酸之間。

[0763] 重組表現載體可以包含編碼用於如在此所限定的多靶向中的Cpf1酶、系統或複合物的多核苷酸，該等多核苷酸處於適用於在宿主細胞中表現核酸的形式，這意味著重組表現載體包含一個或多個調節元件，該等調節元件可以基於用於表現的宿主細胞來選擇，可操作地連接至有待表現的核酸序列。在重組表現載體內，“可操作地連接”旨在意指感興趣的核苷酸序列以允許核苷酸序列表現（例如，在體外轉錄/翻譯系統中或當該載體被引入到宿主細胞時在宿主細胞中）的方式連接至一個或多個調節元件。

[0764] 在一些實施方式中，宿主細胞係用包含編碼用於如在此所限定的多靶向中的Cpf1酶、系統或複合物的多核苷酸的一種或多種載體暫態轉染或非暫態轉染的。在一些實施方式中，細胞當天然存在在受試者中時被轉染。在一些實施方式中，轉染的細胞係從受試者中獲得的。在一些實施方式中，細胞係來源於從受試者中獲得的細胞，諸如細胞系。用於組織培養的多種多樣的細胞系係本領域已知的並且在此在其他地方舉例說明。細胞系可從

熟習該項技術者已知的多種來源獲得（例如，參見美國典型培養物保藏中心（ATCC）（維吉尼亞州馬納薩斯（Manassus, Va.））)。在一些實施方式中，將用包含編碼用於如在此所限定的多靶向中的Cpf1酶、系統或複合物的多核苷酸的一種或多種載體轉染的細胞用於建立一種包含一種或多種載體衍生序列的新細胞系。在一些實施方式中，將用如在此所述的用於多靶向的Cpf1 CRISPR系統或複合物的組分暫態轉染（諸如藉由暫態轉染一種或多種載體或用RNA轉染）並且藉由Cpf1 CRISPR系統或複合物的活性修飾的細胞用於建立一種包括含有修飾但缺乏任何其他外源性序列的細胞的細胞系。在一些實施方式中，將用包含編碼用於如在此所限定的多靶向中的Cpf1酶、系統或複合物的多核苷酸的一種或多種載體暫態轉染或非暫態轉染的細胞，或來源於此類細胞的細胞系用於評估一種或多種測試化合物。

[0765] 術語“調節元件”如在此其他地方所限定。

[0766] 有利的載體包括慢病毒和腺伴隨病毒並且所述載體類型還可以針對靶向的特定細胞類型來選擇。

[0767] 在一個方面中，本發明提供了一種包含以下項的真核宿主細胞：(a) 可操作地連接至同向重複序列的第一調節元件和用於將一個或多個指導RNA序列插入同向重複序列的上游或下游（無論哪一種都適用）的一個或多個插入位點，其中在表現時，

一種或多種指導序列引導Cpf1 CRISPR複合物與真核細胞中的對應的一個或多個靶序列的序列特異性結合，其中Cpf1 CRISPR複合物包含與雜交至對應的一個或多個靶序列的一種或多種指導序列複合的Cpf1酶；和/或 (b) 可操作地連接至編碼包含較佳的是至少一個核定位序列和/或NES的所述Cpf1酶的酶編碼序列的第二調節元件。在一些實施方式中，宿主細胞包含組分 (a) 和 (b)。在一些實施方式中，組分 (a)、組分 (b) 或組分 (a) 和 (b) 被穩定地整合到宿主真核細胞的基因組中。在一些實施方式中，組分 (a) 進一步包含可操作地連接至第一調節元件，並且視情況藉由同向重複序列分開的兩種或更多種指導序列，其中當表現時，兩種或更多種指導序列中的每種引導Cpf1 CRISPR複合物與真核細胞中的不同靶序列的序列特異性結合。在一些實施方式中，Cpf1酶包含具有足以驅動所述CRISPR酶在真核細胞的核中和/或之外以可檢測的量積累的強度的一個或多個核定位序列和/或核輸出序列或NES。

[0768] 在一些實施方式中，Cpf1酶係V型或VI型CRISPR系統酶。在一些實施方式中，Cpf1酶係Cpf1酶。在一些實施方式中，Cpf1酶來源於土拉熱弗朗西絲菌1、土拉熱弗朗西絲菌新殺手亞種、易北普雷沃菌、毛螺旋菌科細菌MC2017 1、解朊丁酸弧菌、佩萊格裡尼菌科細菌GW2011_GWA2_33_10、儉菌總門細菌GW2011_GWC2_44_17、密斯氏菌屬某種SCADC、胺基酸球菌屬某種BV3L6、毛螺旋菌MA2020、候選白蟻甲烷枝原體、挑剔真細菌、牛莫拉氏菌237、稻田鉤端螺旋體、毛螺旋菌科細菌ND2006、狗口腔卟啉單胞菌3、解糖胨普雷沃菌或獼猴卟啉單胞菌的Cpf1，

並且可以進一步包含如在此其他地方所限定的Cpf1的改變或突變，並且可以是嵌合Cpf1。在一些實施方式中，Cpf1酶被密碼子優化為在真核細胞中表現。在一些實施方式中，CRISPR酶引導靶序列位置處的一條或兩條股的切割。在一些實施方式中，第一調節元件係聚合酶III啟動子。在一些實施方式中，第二調節元件係聚合酶II啟動子。在一些實施方式中，一種或多種指導序列的長度（各自）係至少16、17、18、19、20、25個核苷酸，或介於16個-30個、或介於16個-25個、或介於16個-20個核苷酸之間。當使用多個指導RNA時，它們較佳的是藉由同向重複序列分開。在一個方面中，本發明提供了一非人類真核生物體；較佳的是多細胞真核生物體，該等生物體包含根據任何所述實施方式的真核宿主細胞。在其他方面中，本發明提供了一真核生物體；較佳的是多細胞真核生物體，該等生物體包含根據任何所述實施方式的真核宿主細胞。在該等方面的一些實施方式中，該生物體可以是動物；例如，哺乳動物。而且，該生物體可以是節肢動物，諸如昆蟲。生物體還可以是植物。此外，生物體可以是真菌。

[0769] 在一個方面中，本發明提供了一種包含在此所述的一種或多種組分的套組。在一些實施方式中，套組包括載體系統和用於使用套組的說明書。在一些實施方式中，該載體系統包含：(a) 可操作地連接至同向重複序列的第一調節元件和用於將一種或多種指導序列插入同向重複序列的上游或下游（無論哪一種都適用）的一個或多個插入位點，其中在表現時，指導序列引導Cpf1 CRISPR複合物與真核細胞中的靶序列的序列特異性結合，其中Cpf1 CRISPR複合物包含與雜交至靶序列的指導序列複合的Cpf1

酶;和/或 (b) 可操作地連接至編碼包含核定位序列的所述Cpf1酶的酶編碼序列的第二調節元件。在一些實施方式中，套組包括位於系統的相同或不同的載體上的組分 (a) 和 (b)。在一些實施方式中，組分 (a) 進一步包含可操作地連接至第一調節元件的兩種或更多種指導序列，其中當表現時，兩種或更多種指導序列中的每種引導CRISPR複合物與真核細胞中的不同靶序列的序列特異性結合。在一些實施方式中，Cpf1酶包含具有足以驅動所述CRISPR酶在真核細胞的核中以可檢測的量積累的強度的一個或多個核定位序列。在一些實施方式中，CRISPR酶係V型或VI型CRISPR系統酶。在一些實施方式中，CRISPR酶係Cpf1酶。在一些實施方式中，Cpf1酶來源於土拉熱弗朗西絲菌1、土拉熱弗朗西絲菌新殺手亞種、易北普雷沃菌、毛螺旋菌科細菌MC2017 1、解肽丁酸弧菌、佩萊格裡尼菌科細菌GW2011_GWA2_33_10、儉菌總門細菌GW2011_GWC2_44_17、密斯氏菌屬某種SCADC、胺基酸球菌屬某種BV3L6、毛螺旋菌MA2020、候選白蟻甲烷枝原體、挑剔真細菌、牛莫拉氏菌237、稻田鉤端螺旋體、毛螺旋菌科細菌ND2006、狗口腔卟啉單胞菌3、解糖胨普雷沃菌或獼猴卟啉單胞菌的Cpf1（例如，被修飾成具有或締合於至少一個DD），並且可以進一步包含Cpf1的改變或突變，並且可以是嵌合Cpf1。在一些實施方式中，DD-CRISPR酶被密碼子優化為在真核細胞中表現。在一些實施方式中，DD-CRISPR酶引導靶序列位置處的一條或兩條股的切割。在一些實施方式中，DD-CRISPR酶缺乏或基本上缺乏DNA股切割活性（例如，與野生型酶或不具有降低核酸酶活性的突變或改變的酶相比的不超過5%的核酸酶活

性)。在一些實施方式中，第一調節元件係聚合酶III啟動子。在一些實施方式中，第二調節元件係聚合酶II啟動子。在一些實施方式中，指導序列的長度係至少16、17、18、19、20、25個核苷酸，或介於16個-30個、或介於16個-25個、或介於16個-20個核苷酸之間。

[0770] 在一個方面中，本發明提供了一種修飾宿主細胞諸如真核細胞中的多種靶多核苷酸之方法。在一些實施方式中，該方法包括使得Cpf1CRISPR複合物結合多種靶多核苷酸，例如來實施所述多種靶多核苷酸的切割，從而修飾多種靶多核苷酸，其中該Cpf1CRISPR複合物包含與多個指導序列複合的Cpf1酶，每個指導序列雜交至所述靶多核苷酸內的特異性靶序列，其中所述多個指導序列連接至同向重複序列。在一些實施方式中，所述切割包括藉由所述Cpf1酶切割每個靶序列的位置處的一條或兩條股。在一些實施方式中，所述切割使得多個靶基因的轉錄減少。在一些實施方式中，該方法進一步包括使用外源性模板多核苷酸藉由同源重組修復一種或多種所述所述切割的靶多核苷酸，其中所述修復產生包括一種或多種所述靶多核苷酸的一個或多個核苷酸的插入、缺失或取代的突變。在一些實施方式中，所述突變使得由包含一個或多個靶序列的基因表現的蛋白質中發生一個或多個胺基酸的變化。在一些實施方式中，該方法進一步包括將一個或多個載體遞送至所述真核細胞，其中一個或多個載體驅動以下項中的一個或多個的表現：Cpf1酶和連接至同向重複序列的多個指導RNA序列。在一些實施方式中，所述載體被遞送至受試者中的真核細胞。在一些實施方式中，所述修飾發生在細胞培養中的所述真核細胞

中。在一些實施方式中，該方法進一步包括在所述修飾之前將所述真核細胞從受試者分離。在一些實施方式中，該方法進一步包括將所述真核細胞和/或衍生自其的細胞返回至所述受試者。

[0771] 在一個方面中，本發明提供了一種修飾多種多核苷酸在真核細胞中的表現的方法。在一些實施方式中，該方法包括使得Cpf1 CRISPR複合物結合多種多核苷酸，以使得所述結合導致所述多核苷酸的表現增加或減少；其中Cpf1 CRISPR複合物包含與多個指導序列複合的Cpf1酶，每個指導序列雜交至其自身的在所述多核苷酸內的靶序列，其中所述指導序列連接至同向重複序列。在一些實施方式中，該方法進一步包括將一個或多個載體遞送至所述真核細胞，其中一個或多個載體驅動以下項中的一個或多個的表現：Cpf1酶和連接至同向重複序列的多個指導序列。

[0772] 在一個方面中，本發明提供了一種包含同向重複序列的上游或下游（無論哪一種都適用）的多個指導RNA序列的重組多核苷酸，其中在表現時，每個指導序列引導Cpf1CRISPR複合物與其在真核細胞中存在的相應靶序列的序列特異性結合。在一些實施方式中，靶序列係真核細胞中存在的病毒序列。在一些實施方式中，靶序列係原癌基因或癌基因。

[0773] 本發明的方面涵蓋可以包含指導RNA（gRNA）和如在此所限定的Cpf1酶的非天然存在或工程化的組成物，該指導RNA包含能夠與細胞中感興趣的基因組座位中的靶序列雜交的指導序列，該Cpf1酶可以包含至少一個或多個核定位序列。

[0774] 本發明的一個方面涵蓋藉由將在此所述的任一組成

物引入到細胞中來修飾感興趣的基因組座位以改變該細胞中的基因表現的方法。

[0775] 本發明的一個方面在於以上元件包含在單一組成物中或包含在單獨組成物中。該等組成物可以有利地應用於宿主以引起基因組水平上的功能效應。

[0776] 如在此所用，術語“指導RNA”或“gRNA”具有如在此其他地方使用的含義（leaning）並且包括與靶核酸序列具有足夠互補性以與靶核酸序列雜交並引導核酸靶向複合物與靶核酸序列的序列特異性結合的任何多核苷酸序列。每個gRNA可以被設計成包含特異於相同或不同轉接蛋白的多個結合識別位點（例如適配體）。每個gRNA可以被設計成結合位於轉錄起始位點（即TSS）的上游的啟動子區-1000 - +1核酸，較佳的是-200核酸。此定位改善了影響基因啟動（例如，轉錄活化物）或基因抑制（例如，轉錄阻遏物）的功能結構域。修飾gRNA可以是組成物中包含的靶向一個或多個靶座位的一個或多個修飾gRNA（例如至少1個gRNA、至少2個gRNA、至少5個gRNA、至少10個gRNA、至少20個gRNA、至少30個gRNA、至少50個gRNA）。所述多個gRNA序列可以是串聯安排的並且較佳的是藉由同向重複序列分開。

[0777] 因此，如在此所限定的gRNA、CRISPR酶可以各自單獨地包含在組成物中並且單獨或共同地給予至宿主。可替代地，該等組分可以提供於單一組成物中以用於向宿主給予。對宿主的給藥可以經由技術人員已知的或在此描述的用於遞送至宿主的病毒載體（例如，慢病毒載體、腺病毒載體、AAV載體）進行。如在

此解釋的，使用不同的選擇標記物（例如，對於慢病毒gRNA選擇）和gRNA濃度（例如，取決於是否使用多個gRNA）可能對引起改善的效應是有利的。基於此構想，若干種變體適合用於引起基因組座位事件，包括DNA切割、基因啟動或基因失活。使用提供的組成物，熟習該項技術者可以使用相同或不同功能結構域有利地且特異性地靶向單一或多個座位以引起一個或多個基因組座位事件。組成物可以應用在用於篩選細胞中文庫和體內功能模型（例如，lincRNA的基因啟動和功能的鑒定；增功能模型；失功能模型；使用本發明組成物來建立細胞系和轉基因動物以用於優化和篩選目的）的多種多樣的方法中。

[0778] 本發明包括本發明的組成物用於建立和利用條件型或誘導型CRISPR轉基因細胞/動物的用途；例如參見，普萊特（Platt）等人，細胞 (2014), 159(2): 440-455，或在此引用的PCT專利出版物，諸如WO 2014/093622（PCT/US2013/074667）。例如，細胞或動物諸如非人類動物，例如脊椎動物或哺乳動物，諸如齧齒動物，例如小鼠、大鼠或其他實驗室或野生動物，例如貓、狗、羊等，可以進行“敲入”，由此類似於普萊特等人該動物條件型地或可誘導型地表現Cpf1。因此靶細胞或動物包含條件型或可誘導型（例如呈Cre依賴性構建體的形式）的CRISRP酶（例如Cpf1），在引入到靶細胞中的載體表現時，載體表現Cre，從而誘導或產生了靶細胞中的CRISRP酶（例如Cpf1）表現的條件。藉由應用如在此所限定的教授內容和組成物以及創建CRISPR複合物的已知方法，誘導型基因組事件還可以是本發明的一個方面。此類誘導型事件的實例已在此其他地方有所描述。

[0779] 在一些實施方式中，當特別是在治療方法中遺傳疾病被靶向時，表型改變較佳的是基因組修飾的結果並且較佳的是其中提供了修復模板以校正或改變表型。

[0780] 在一些實施方式中，可以被靶向的疾病包括與引起疾病的剪接缺陷相關的那些。

[0781] 在一些實施方式中，細胞靶標包括造血幹細胞/祖細胞（CD34+）；人類T細胞；以及眼（視網膜細胞）-例如光受體先質細胞。

[0782] 在一些實施方式中，基因靶包括：人類β球蛋白-HBB（用於治療鐮狀細胞貧血，包括藉由刺激基因轉變（使用緊密相關的HBD基因作為內源性模板）進行）；CD3（T細胞）；以及CEP920-視網膜（眼）。

[0783] 在一些實施方式中，疾病靶標也可以包括：癌症；鐮狀細胞貧血（基於點突變）；HBV、HIV；β-地中海貧血；以及眼睛或眼部疾病-例如引起萊伯氏先天性黑矇（LCA）的剪接缺陷。

在一些實施方式中，遞送方法包括：酶-指導序列複合物（核糖核蛋白）的陽離子脂質介導的“直接遞送”以及質粒DNA的電穿孔。

[0784] 在此所述的方法、產物和用途可以用於非治療性目的。此外，任一在此所述方法可以應用於體外離體中。

[0785] 在一個方面中，提供了非天然存在或工程化的組成物，該組成物包含：

I.兩種或更多種CRISPR-Cas系統多核苷酸序列，該等多核苷酸序列包含：

(a) 能夠雜交至多核苷酸座位中的第一靶序列的第一指導序列，

(a) 能夠雜交至多核苷酸座位中的第二靶序列的第二指導序列，

(c) 同向重複序列，

以及

II.Cpf1酶或編碼它的第二多核苷酸序列，

其中當轉錄時，第一指導序列和第二指導序列分別引導第一Cpf1 CRISPR複合物和第二Cpf1 CRISPR複合物與第一靶序列和第二靶序列的序列特異性結合，

其中第一 CRISPR 複合物包含與可雜交至第一靶序列的第一指導序列複合的 Cpf1 酶，

其中第二 CRISPR 複合物包含與可雜交至第二靶序列的第二指導序列複合的 Cpf1 酶，並且

其中第一指導序列引導DNA雙股體中的靠近第一靶序列的一條股的切割並且第二指導序列引導靠近第二靶序列的另一條股的切割，誘導雙股斷裂，從而修飾生物體或非人類或非動物生物體。類似地，可以設想包含超過兩個的指導RNA的組成物，例如每個RNA 特異於一靶標，並且串聯安排在如在此所述的組成物或

CRISPR 系統或複合物中。

[0786] 在另一個實施方式中，Cpf1作為蛋白質遞送至細胞中。在另一個且特別較佳的實施方式中，Cpf1作為蛋白質或作為編碼它的核苷酸序列遞送至細胞中。作為蛋白質向細胞的遞送可以包括核糖核蛋白（RNP）複合物的遞送，其中蛋白質與多個指導序列複合。

[0787] 在一個方面中，提供了藉由本發明的組成物、系統或修飾酶修飾或包含該等組成物、系統或修飾酶的宿主細胞和細胞系，包括幹細胞及其子代。

[0788] 在一個方面中，提供了細胞治療的方法，其中例如取樣或培養了單個細胞或細胞群體，其中一個或多個細胞係或已經係如在此所述進行離體修飾的，並且然後被重新引入（取樣的細胞）或引入（培養的細胞）到生物體中。就這一點而言，幹細胞，無論胚胎幹細胞或誘導多能幹細胞或全能幹細胞也是特別較佳的。但是，當然，也設想了體內實施方式。

[0789] 本發明方法可以進一步包括模板的遞送，諸如修復模板，該等修復模板可以是dsODN或ssODN，參見下文。模板遞送可以是經由與任一或所有CRISPR酶或指導RNA的遞送同時發生或分開並且經由相同遞送機制或不同遞送機制。在一些實施方式中，較佳的是，模板與指導RNA一起遞送，並且也較佳的是與CRISPR酶一起遞送。一個實例可以是AAV載體，其中CRISPR酶係AsCpf1或LbCpf1。

[0790] 本發明方法可以進一步包括：(a) 將包含互補於由所述雙股斷裂創建的突出端的突出端的雙股寡去氧核苷酸（dsODN）遞送至細胞，其中所述dsODN被整合到感興趣的座位中；或–(b)將單股寡去氧核苷酸（ssODN）遞送至細胞，其中所述ssODN充當用於所述雙股斷裂的同源定向修復的模板。本發明方法可以用於預防或治療個體的疾病，視情況其中所述疾病由所述感興趣的座位中的缺陷導致。本發明方法可以在個體中體內進行或者在取自個體的細胞上離體進行，視情況其中所述細胞被返回至個體。

[0791] 本發明還包括根據使用用於如在此所限定的串聯或多靶向中的CRISPR酶或Cas酶或Cpf1酶或CRISPR-CRISPR酶或CRISPR-Cas系統或CRISPR-Cpf1系統獲得產物。

套組

[0792] 在一個方面中，本發明提供了含有在以上方法和組成物中所揭露的任何一個或多個元件的套組。在一些實施方式中，套組包括如在此教授的載體系統和用於使用套組的說明書。元件可以單獨地或組合地提供，並且可以被提供於任何適合的容器中，諸如小瓶、瓶子或管。套組可以包括gRNA和如在此所述的非結合保護子股。套組可以包括gRNA以及與指導序列至少部分地結合的保護子股（即pgRNA）。因此，套組可以包括呈如在此所述的部分雙股核苷酸序列的pgRNA。在一些實施方式中，套組包括一種或多種語言，例如超過一種語言的說明書。說明書可以是針對在此所述的應用和方法的。

[0793] 在一些實施方式中，套組包括在利用在此所述的一個

或多個元件的方法中使用的一種或多種試劑。試劑可以提供於任何適合容器中。例如，套組可以提供一種或多種反應或存儲緩衝液。可以按在具體測定中可用的形式或按在使用之前需要添加一種或多種其他組分的形式（例如按濃縮或凍乾形式）提供試劑。緩衝液可以是任何緩衝液，包括但不限於碳酸鈉緩衝液、碳酸氫鈉緩衝液、硼酸鹽緩衝液、Tris緩衝液、MOPS緩衝液、HEPES緩衝液及其組合。在一些實施方式中，緩衝液係鹼性的。在一些實施方式中，緩衝液具有從約7至約10的pH。在一些實施方式中，套組包括一種或多種寡核苷酸，該一種或多種寡核苷酸對應於用於插入到載體中的指導序列，以便可操作地連接該指導序列和調節元件。在一些實施方式中，套組包括同源重組模板多核苷酸。在一些實施方式中，套組包括在此所述的一種或多種載體和/或一種或多種多核苷酸。套組可以有利地允許提供本發明的系統的所有元件。

[0794] 在一個方面中，本發明提供了用於使用CRISPR系統的一個或多個元件的方法。本發明的CRISPR複合物提供了一種用於修飾靶多核苷酸的有效手段。本發明的CRISPR複合物具有多種多樣的效用，包括修飾（例如，缺失、插入、易位、失活、啟動）許多細胞類型中的靶多核苷酸。這樣，本發明的CRISPR複合物在例如基因治療、藥物篩選、疾病診斷以及預後方面具有廣泛的應用。示例性CRISPR複合物包含與雜交至靶多核苷酸內的靶序列的指導序列複合的CRISPR效應蛋白。在某些實施方式中，同向重複序列連接至指導序列。

[0795] 在一個實施方式中，本發明提供了一種切割靶多核苷酸的方法。該方法包括使用結合靶多核苷酸的CRISPR複合物修飾靶多核苷酸並且實施所述靶多核苷酸的切割。典型地，本發明的CRISPR複合物在被引入到細胞中時產生基因組序列的斷裂（例如單股或雙股斷裂）。例如，該方法可以用於切割細胞中的疾病相關基因。

[0796] 藉由CRISPR複合物產生的斷裂可以藉由修復過程來修復，諸如易出錯的非同源末端連接（NHEJ）途徑或高保真性同源定向修復（HDR）。在該等修復過程期間，可以將一個外源性多核苷酸模板引入到基因組序列中。在一些方法中，該HDR過程被用於修飾基因組序列。例如，可以將包含有待整合的側接有一個上游序列和一個下游序列的序列的外源性多核苷酸模板引入到細胞中。上游序列和下游序列與染色體中整合位點的任一側享有序列相似性。

[0797] 在希望的情況下，供體多核苷酸可以是DNA，例如DNA質粒、細菌人工染色體（BAC）、酵母人工染色體（YAC）、病毒載體、一段線性DNA、PCR片段、裸核酸或與遞送媒介物（諸如脂質體或泊洛沙姆）複合的核酸。

[0798] 外源性多核苷酸模板包含有待整合的序列（例如，突變型基因）。供整合的序列可以是對細胞而言內源或外源的序列。有待整合的序列的實例包括編碼蛋白質的多核苷酸或非編碼RNA（例如，微小RNA）。因此，供整合的序列可以可操作地連接至一種或多種適當的控制序列。可替代地，有待整合的序列可以提供

調節功能。

[0799] 外源性多核苷酸模板中的上游序列和下游序列被選擇為促進感興趣的染色體序列與供體多核苷酸之間的重組。上游序列係與供整合的靶向位點的上游的基因組序列享有序列相似性的核酸序列。類似地，下游序列係與整合的靶向位點的下游的染色體序列享有序列相似性的核酸序列。外源性多核苷酸模板中的上游序列和下游序列與靶向的基因組序列可以具有75%、80%、85%、90%、95%或100%序列一致性。較佳的是，外源性多核苷酸模板中的上游序列和下游序列與靶向的基因組序列具有約95%、96%、97%、98%、99%或100%序列一致性。在一些方法中，外源性多核苷酸模板中的上游序列和下游序列與靶向的基因組序列具有約99%或100%序列一致性。

[0800] 上游序列或下游序列可以包含從約20 bp至約2500 bp，例如約50、100、200、300、400、500、600、700、800、900、1000、1100、1200、1300、1400、1500、1600、1700、1800、1900、2000、2100、2200、2300、2400或2500 bp。在一些方法中，示例性上游序列或下游序列具有約200 bp至約2000 bp、約600 bp至約1000 bp，或更具體地約700 bp至約1000 bp。

[0801] 在一些方法中，外源性多核苷酸模板可以進一步包含標記物。此標記物可以使得容易地篩選靶向的整合。適合的標記物的實例包括限制位點、螢光蛋白或選擇標記物。可以使用重組技術構建本發明的外源性多核苷酸模板（例如參見，薩姆布魯克（Sambrook）等人，2001和奧蘇貝爾（Ausubel）等人，1996）。

[0802] 在一用於藉由整合外源性多核苷酸模板來修飾靶多核苷酸的示例性方法中，藉由CRISPR複合物將雙股斷裂引入到基因組序列中，經由同源重組外源性多核苷酸模板而修復該斷裂，這樣使得將該模板整合到基因組中。雙股斷裂的存在促進模板的整合。

[0803] 在其他實施方式中，本發明提供了一種修飾多核苷酸在真核細胞中的表現之方法。該方法包括藉由使用結合多核苷酸的CRISPR複合物增加或減少靶多核苷酸的表現。

[0804] 在一些方法中，可以使靶多核苷酸失活以實施細胞中的表現的修飾。例如，在CRISPR複合物與細胞中的靶序列結合後，靶多核苷酸失活，這樣使得該序列不被轉錄，該編碼蛋白不被產生，或者該序列不會像野生型序列一樣起作用。例如，可以使蛋白質或微小RNA編碼序列失活，這樣使得該蛋白質不被產生。

[0805] 在一些方法中，控制序列可以失活，使得其不再作為控制序列起作用。如在此使用，“控制序列”係指影響核酸序列的轉錄、翻譯或可及性的任何核酸序列。控制序列的實例包括啟動子、轉錄終止子和增強子，它們係控制序列。失活的靶序列可以包括缺失突變（即，缺失一個或多個核苷酸）、插入突變（即，插入一個或多個核苷酸）或無義突變（即，用另一個核苷酸取代一個單核苷酸，這樣使得引入終止密碼子）。在一些方法中，靶序列的失活導致該靶序列的“敲除”。

CRISPR Cas系統的示例性使用方法

[0806] 本發明提供了一種非天然存在的或工程化的組成物、或編碼所述組成物的組分的一種或多種多核苷酸、或含有編碼所述組成物的組分一種或多種多核苷酸的載體或遞送系統，其用於體內、離體或體外修飾靶細胞並且該修飾可以改變細胞使得一旦修飾，CRISPR修飾細胞的子代或細胞系保留改變的表型的方式實施。該等修飾的細胞和子代可以是多細胞生物體的一部分，諸如在將CRISPR系統應用於所希望的細胞類型的情況下的植物或動物。CRISPR發明可以是一種治療性治療方法。治療性治療方法可以包括基因或基因組編輯，或基因治療。

失活的CRISPR Cpf1酶用於檢測方法諸如FISH之用途

[0807] 在一個方面中，本發明提供了一種包含在此所述的催化失活Cas蛋白，較佳的是失活Cpf1（dCpf1）的工程化、非天然發生的CRISPR-Cas系統，以及此系統在檢測方法諸如螢光原位雜交（FISH）的檢測方法中之用途。缺乏產生DNA雙股斷裂的能力的dCpf1可以與標記物諸如螢光蛋白，諸如增強型綠色螢光蛋白（eEGFP）融合，並且與小指導RNA共表現以靶向體內的臂間、中心和端粒的（teleomeric）重複序列。dCpf1系統可以用於視覺化人類基因組中的重複序列和單個基因兩者。標記的dCpf1 CRISPR-cas系統的此類新應用在成像細胞和研究機能核體系結構中，特別是在小核體積或複合物3-D結構的情況下可能是重要的。（陳·B、吉伯·LA（Gilbert LA）、克伊米尼·BA（Cimini BA）、斯尼茨鮑爾·J（Schnitzbauer J）、張·W、李·GW、派克·J（Park J）、布萊克本·EH（Blackburn EH）、魏斯曼·JS（Weissman JS）、齊·LS（Qi LS）、

黃·B（Huang B），2013，藉由優化的CRISPR/Cas系統使活人類細胞中的基因組座位動態成像，細胞 155(7):1479-91. doi: 10.1016/j.cell.2013.12.001.）

用CRISPR Cas系統或複合物（例如，Cpf1-RNA複合物）修飾靶標

[0808] 在一個方面中，本發明提供了修飾真核細胞中的靶多核苷酸的方法，該等方法可以是在體內、離體或在體外。在一些實施方式中，該方法包括從人類或非人類動物取樣細胞或細胞群體，並且修飾該細胞或該等細胞。培養可以發生在離體的任何階段。該細胞或該等細胞甚至可以被重新引入非人類動物或植物中。對於重新引入的細胞，特別較佳的是該等細胞係幹細胞。

[0809] 在一些實施方式中，該方法包括使得CRISPR複合物結合靶多核苷酸來實施所述靶多核苷酸的切割，從而修飾該靶多核苷酸，其中該CRISPR複合物包含與雜交至或可雜交至所述靶多核苷酸內的靶序列的指導序列複合的CRISPR酶。

[0810] 在一個方面中，本發明提供了一種修飾多核苷酸在真核細胞中的表現之方法。在一些實施方式中，該方法包括使得CRISPR複合物結合該多核苷酸，以使得所述結合導致所述多核苷酸的表現增加或減少；其中CRISPR複合物包含與雜交至或可雜交至所述多核苷酸內的靶序列的指導序列複合的CRISPR酶。類似的考慮因素和條件適用如上文針對修飾靶多核苷酸之方法。實際上，該等取樣、培養和重新引入選擇跨本發明的多個方面而適用。

[0811] 實際上，在本發明的任何方面中，CRISPR複合物可以包含與雜交至或可雜交至靶序列的指導序列複合的CRISPR酶。類似的考慮因素和條件適用如上文針對修飾靶多核苷酸的方法。

[0812] 因此，在在此所述的任一非天然存在的CRISPR酶中包含至少一種修飾並且由此該酶具有某些改善的能力。具體地說，任一酶能夠與指導RNA形成CRISPR複合物。當形成此複合物時，指導RNA能夠結合靶多核苷酸序列並且酶能夠修飾靶座位。此外，與未修飾酶相比，CRISPR複合物中的酶具有降低的修飾一個或多個脫靶座位的能力。

[0813] 此外，在此所述的修飾CRISPR酶涵蓋下述的酶：由此在CRISPR複合物中該酶與未修飾酶相比具有增加的修飾一個或多個靶座位的能力。此功能可以單獨提供或者與以上所述的降低的修飾一個或多個脫靶座位的能力的功能組合提供。任何此類酶可以提供有如在此所述的對CRISPR酶的任一另外修飾，諸如與藉由一個或多個締合的異源功能結構域提供的任何活性、任何降低核酸酶活性的另外突變等組合。

[0814] 在本發明的有利實施方式中，修飾CRISPR酶相比於未修飾酶被提供有降低的修飾一個或多個脫靶座位的能力並且相比於未修飾酶被提供有增加的修飾一個或多個靶座位的能力。在與對酶的另外修飾組合的情況下，可以實現顯著增強的特異性。例如，提供了此類有利實施方式與一個或多個另外的突變的組合，其中一個或多個另外的突變係處於一個或多個催化活性結構域之中。此類另外的催化突變可以賦予如在此其他地方詳細所述的切

口酶功能性。在此類酶中，可以實現增強的特異性，這歸因於關於酶活性的改善特異性。

[0815] 可以對坐落於位於RuvC-III結構域與HNH結構域之間的帶正電荷的區/溝中的胺基酸殘基進行如以上所述的降低脫靶效應和/或增強中靶效應的突變。應瞭解的是，任一以上所述的功能效應可以藉由上述溝內的胺基酸的修飾來實現，但是還藉由相鄰於此溝或在此溝之外的胺基酸的修飾來實現。

[0816] 可以被工程化到如在此所述的修飾CRISPR酶的另外功能包括以下項。1.破壞DNA:蛋白質相互作用而不影響蛋白質三級或二級結構的修飾CRISPR酶。此CRISPR酶包含接觸RNA:DNA雙股體的任一部分的殘基。2.響應於DNA結合（中靶或脫靶）削弱內部蛋白質相互作用使Cpf1保持為核酸酶切割所必需的構象的修飾CRISPR酶。例如，微弱抑制，但是仍允許HNH結構域（定位在易裂的磷酸鹽處）的核酸酶構象的修飾。3.響應於DNA結合（中靶或脫靶）增強內部蛋白質相互作用使Cpf1保持為抑制核酸酶活性的構象的修飾CRISPR酶。例如：將HNH結構域穩定為呈遠離易裂的磷酸鹽的構象的修飾。可以與如在此其他地方詳細所述的對CRISPR酶的任何其他修飾組合來提供任何此另外的功能增強。

[0817] 任一在此所述的改善功能性可以對任何CRISPR酶，諸如Cpf1酶進行。然而，應瞭解，在此所述的任一功能性可以被工程化到來自其他異種同源物的Cpf1酶中，包括包含來自多個異種同源物的片段的嵌合酶。

核酸、胺基酸和蛋白質、調節序列、載體等

[0818] 本發明使用核酸結合靶DNA序列。這係有利的，因為核酸的製備比蛋白質更容易且更便宜，並且特異性根據其中尋求同源性的一段序列長度而改變。多靶指的複雜3-D定位例如是不需要的。術語“多核苷酸”、“核苷酸”、“核苷酸序列”、“核酸”以及“寡核苷酸”係可互換使用的。它們係指任何長度的核苷酸（去氧核糖核苷酸或核糖核苷酸）的聚合形式或其類似物。多核苷酸可以具有任何三維結構並且可以執行任何已知或未知的功能。以下各項係多核苷酸的非限制性實例：基因或基因片段的編碼區或非編碼區、由連鎖分析定義的多個座位（一個座位）、外顯子、內含子、信使RNA(mRNA)、轉移RNA、核糖體RNA、短干擾RNA(siRNA)、短髮夾RNA（shRNA）、微RNA（miRNA）、核糖核酸酶、cDNA、重組多核苷酸、分枝多核苷酸、質粒、載體、分離的任何序列的DNA、分離的任何序列的RNA、核酸探針以及引物。該術語還涵蓋具有合成骨架的核酸樣結構，參見例如埃克斯坦（Eckstein），1991；巴塞折（Baserga）等人，1992；米利根（Milligan），1993；WO 97/03211；WO 96/39154；馬塔（Mata），1997；施特勞斯-紹庫普（Strauss-Soukup），1997；以及紮姆斯塔格（Samstag），1996。多核苷酸可以包含一個或多個修飾核苷酸，諸如甲基化核苷酸和核苷酸類似物。如果存在的話，對核苷酸結構的修飾可以在聚合物組裝之前或之後賦予。核苷酸的序列可以被非核苷酸組分中斷。多核苷酸可以在聚合之後諸如藉由與標記組分軛合來進一步修飾。如在此所用，術語“野生型”係熟習該項技術者理解的技術術語並且意指在自然界中出現的典型式的生物體、菌株、基因或特徵，與突變體或變體形式區分。“野生型”可以是底線。如在此所用，術

語“變體”應理解為意指具有源自自然界中存在的模式的性質展示。術語“非天然存在的”或“工程化的”係可互換使用的並且係指涉及人工處理。該等術語當提及核酸分子或多肽時意指核酸分子或多肽至少基本上與至少一種其他組分分離，該至少一種其他組分在自然界中與該核酸分子或多肽天然締合並且如自然界中發現的。“互補”係指核酸藉由傳統的沃森-克裡克鹼基配對或其他非傳統類型來與另一個核酸序列形成氫鍵的能力。互補百分比表示核酸分子中可與第二核酸序列形成氫鍵（例如，沃森-克裡克鹼基配對）的殘基百分比（例如，10分之5、6、7、8、9、10係50%、60%、70%、80%、90%以及100%互補）。“完美互補”意指核酸序列的所有連續殘基都將與第二核酸序列中同樣數目的連續殘基形成氫鍵。如在此所用的“基本上互補”係指在8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、30、35、40、45、50、或更多個核苷酸上的至少60%、65%、70%、75%、80%、85%、90%、95%、97%、98%、99%、或100%的互補程度，或者是指在嚴格條件下雜交的兩個核酸。如在此所用，用於雜交的“嚴格條件”係指與靶序列具有互補的核酸與靶序列顯著雜交並且基本上不與非靶序列雜交的條件。嚴格條件通常是依賴序列的並且根據多種因素而改變。總的來說，該序列越長，該序列與其靶序列特異性雜交的溫度越高。嚴格條件的非限制性實例詳細描述於媞撒（Tijssen）(1993)，生物化學和分子生物學實驗室技術—與核酸探針雜交第I部分，第二章“雜交理論和核酸探針測定策略的綜述”（Laboratory Techniques In Biochemistry And Molecular Biology-Hybridization With Nucleic Acid Probes Part I, Second Chapter “Overview of

principles of hybridization and the strategy of nucleic acid probe assay")，紐約愛思維爾公司（Elsevier, N.Y.）。在參考多核苷酸序列時，那麼還設想互補或部分互補的序列。該等序列較佳的是能夠在高嚴格條件下與參考序列雜交。總體上，為了使雜交率最大化，選擇相對低的嚴格雜交條件：低於熔點（T_m）約20°C至25°C。該T_m係50%特定靶序列在限定的離子強度和pH的溶液中與完美互補探針雜交的溫度。總體上，為了要求至少約85%核苷酸互補的雜交序列，選擇高嚴格洗滌條件為低於該T_m約5°C至15°C。為了要求至少約70%核苷酸互補的雜交序列，選擇中等嚴格洗滌條件為低於該T_m約15°C至30°C。高容許（極低嚴格）洗滌條件可以是低至在該T_m之下50°C，從而允許在雜交序列之間高錯配水平。熟習該項技術者將認識到，雜交和洗滌階段的其他物理和化學參數也可被改變為影響來自靶標與探針序列之間特定同源性水平的可檢測雜交信號的結果。較佳的是高嚴格條件包括在50%甲醯胺、5×SSC以及1% SDS中在42°C下培養或者在5×SSC和1% SDS中在65°C下培養，在0.2×SSC和0.1% SDS中在65°C下洗滌。"雜交"係指其中一個或多個多核苷酸反應形成經由核苷酸殘基的鹼基之間氫鍵鍵合而穩定的複合物的反應。氫鍵可以藉由沃森-克裡克鹼基配對、Hoogstein鍵合或以任何其他序列特異性方式形成該複合物可包含形成雙股體結構的兩條股、形成多股複合物的三條或更多條股、單一自雜交股或該等的任何組合。雜交反應可以構成在更廣泛的方法中的步驟，諸如PCR起始、或酶切割多核苷酸。能夠與給定序列雜交的序列被稱為給定序列的"補體"。如在此所用，術語"基因組座位（genomic locus）"或"座位（locus）"（複數係座位（loci））

係染色體上的基因或DNA序列的特定位點。“基因”係指編碼在生物體中具有功能作用的多肽或RNA股的DNA或RNA片段並且因此係活生物體中的遺傳分子單元。出於本發明的目的，可以認為基因包含調節基因產物產生的區域，無論此類調節序列是否與編碼序列和/或轉錄序列相鄰。因此，基因包括但不一定限於，啟動子序列、終止子、翻譯調節序列諸如核糖體結合位點和內部核糖體進入位點、增強子、沈默子、絕緣子、邊界元件、複製起點、基質附著位點以及座位控制區。如在此所用，“基因組座位的表現”或“基因表現”係來自基因的資訊用於合成功能基因產物所藉由的過程。基因表現產物常常是蛋白質，但在非蛋白編碼基因諸如rRNA基因或tRNA基因中，該產物係功能RNA。所有已知生命（真核生物（包括多細胞生物體）、原核生物（細菌和古生菌）以及病毒）使用基因表現過程產生功能產物以生存。如在此所用，基因或核酸的“表現”不僅涵蓋細胞基因表現，而且涵蓋選殖系統和任何其他背景中的核酸轉錄和翻譯。如在此所用，“表現”還是指多核苷酸從DNA模板轉錄（諸如轉錄成mRNA或其他RNA轉錄物）所藉由的過程和/或轉錄的mRNA隨後翻譯成肽、多肽或蛋白質所藉由的過程。轉錄物和編碼的多肽可以統稱為“基因產物”。如果多核苷酸來源於基因組DNA，則在真核細胞中表現可以包括mRNA的剪接。在此可互換使用的術語“多肽”、“肽”和“蛋白質”係指具有任何長度的胺基酸聚合物。該聚合物可以是線性或支化的，它可以包含修飾胺基酸，並且它可以被非胺基酸中斷。該等術語還涵蓋已修飾的胺基酸聚合物；例如，二硫鍵形成、糖基化、脂化、乙醯化、磷酸化活任何其他操縱，諸如與標記組分軛合。如在此所用，

術語“胺基酸”包括天然和/或非天然或合成的胺基酸，包括甘胺酸和D或L光學異構物，以及胺基酸類似物和肽模擬物。如在此所用，術語“結構域”或“蛋白質結構域”係指可以存在並且獨立於其餘蛋白質鏈起作用的一部分蛋白質序列。如本發明的多個方面中所述的，序列一致性與序列同源性相關。同源性比較可以藉由眼睛來進行，或者更通常地借助於容易獲得的序列比較程式來進行。該等商業上可用的電腦程式可以計算兩個或更多更序列之間的同源性百分比(%)並且還可以計算兩個或更多更胺基酸或核酸序列所享有的序列一致性。

[0819] 在本發明的多個方面中，術語“指導RNA”係指包含推定或鑒定的crRNA序列或指導序列的多核苷酸序列。

[0820] 如在此所用，術語“野生型”係熟習該項技術者理解的技術術語並且意指在自然界中出現的典型式的生物體、菌株、基因或特徵，與突變體或變體形式區分。“野生型”可以是底線。

[0821] 如在此所用，術語“變體”應理解為意指具有源自自然界中存在的模式的性質展示。

[0822] 術語“非天然存在的”或“工程化的”係可互換使用的並且係指涉及人工處理。該等術語當提及核酸分子或多肽時意指核酸分子或多肽至少基本上與至少一種其他組分分離，該至少一種其他組分在自然界中與該核酸分子或多肽天然締合並且如自然界中發現的。在所有方面和實施方式中，無論它們是否包括該等術語，都應當理解，較佳的是，它們可以是視情況並且因此較佳的是包括或並不較佳的是不包括。此外，術語“非天然存在的”和

“工程化的”可以是可互換使用的並且因此也可以單獨或組合使用，並且在兩者一起提及時，它們可以彼此替換。具體地說，“工程化的”較佳的是代替“非天然存在的”或“非天然存在的和/或工程化的”。

[0823] 序列同源性可以藉由任何本領域已知的多種電腦程式例如BLAST或FASTA來生成。用於進行此種比對的適合電腦程式係GCG威斯康辛·貝斯菲特（Wisconsin Bestfit）套裝軟體（美國威斯康辛大學（University of Wisconsin, U.S.A）；德弗羅（Devereux）等人，1984，核酸研究（Nucleic Acids Res.） 12:387）。可以進行序列比較的其他軟體的實例包括但不限於，BLAST套裝軟體（參見奧蘇貝爾等人，1999同上—第18章）、FASTA（安特斯庫爾等人，1990，分子生物學（J. Mol.Biol.）,403-410）和GENEWORKS比較工具套件。BLAST和FASTA兩種均可用於離線和線上搜索（參見奧蘇貝爾等人，1999同上，第7-58頁至第7-60頁）。然而，較佳的是使用GCG貝斯菲特程式。序列同源性百分比（%）可以對連續序列計算，即將一個序列與另一個序列比對並且將在一個序列中的每個胺基酸或核苷酸與另一個序列中的相應胺基酸或核苷酸直接比較，每次一個殘基。這被稱為“無空位”比對。典型地，此類無空位比對僅對相對短的多個殘基進行。儘管這係一種非常簡單且連貫的方法，但是它未能考慮到例如，在序列的另外相同鹼基對中，一個插入或缺失可以引起後面的胺基酸殘基無法比對，因此可能導致在進行總體比對時同源性%大大減小。因此，大部分序列比較方法被設計為產生最佳比對，從而考慮到可能的插入和缺失而不會過度不利於總體同源性或一致性得分。這藉由在序列比對中插

入“空位”以試圖最大化局部同源性或一致性來實現。然而，該等更複雜的方法將“空位罰分”分配給出現在比對中的每個空位，以使得對於相同數目的相同胺基酸，與盡可能少的空位（影響兩個比較的序列之間的較高相關性）的序列比對可以實現比具有許多空位的序列更高的得分。“親和力空位成本（Affinity gap cost）”典型地用於對空位的存在承擔相對高的成本並且對空位中的每個後續殘基施加較小的罰分。這係最常使用的空位評分系統。高空位罰分當然可以產生與較少空位的最佳比對。大部分比對程式允許修改空位罰分。然而，較佳的是當使用此類軟體進行序列比較時使用預設值。例如，當使用GCG威斯康辛·貝斯菲特套裝軟體時，胺基酸的默認空位罰分對於空位係-12並且對於每個延伸係-4。因此最大同源性%的計算首先需要產生最佳比對，考慮到空位罰分。用於進行此種比對的適合的電腦軟體係GCG威斯康辛·貝斯菲特套裝軟體（德弗羅等人，1984核酸研究 12 p387）。可以進行序列比較的其他軟體的實例包括但不限於，BLAST套裝軟體（參見奧蘇貝爾等人，1999，分子生物學短方案，第4版—第18章）、FASTA（安特斯庫爾等人，1990分子雜誌生物學403-410）和GENEWORKS比較工具套件。BLAST和FASTA兩種均可用於離線和線上搜索（參見奧蘇貝爾等人，1999，分子生物學短方案，第7-58頁至第7-60頁）。然而，對於一些應用，較佳的是使用GCG貝斯菲特程式。一種稱為BLAST 2序列的新工具也可用於比較蛋白質和核苷酸序列（參見*FEMS*微生物學概述（*FEMS Microbiol Lett.*）1999 174(2): 247-50；*FEMS微生物學概述*1999 177(1): 187-8以及國家健康研究所網站的國家生物技術資訊中心網站）。儘管可以根

據一致性測量最終同源性%，但是比對方法本身典型地並不是基於不全則無的成對比較。相反，通常使用標準的相似性評分矩陣，它基於化學相似性或進化距離將得分分配給每個成對比較。通常使用的這種矩陣的實例係BLOSUM62矩陣—這係BLAST程式套件的預設矩陣。GCG威斯康辛程式通常使用公用的預設值或自訂符號比較表，如果提供的話（詳細內容，參見使用者手冊）。對於一些應用，較佳的是對於GCG套裝軟體使用公用預設值，並且在其他軟體的情況下，使用預設矩陣，諸如BLOSUM62。可替代地，可以使用DNASIS™（日立軟體公司）中的多重比對特徵，基於與CLUSTAL類似的演算法計算同源性百分比（希金斯·DG（Higgins DG）和夏普·PM（Sharp PM）(1988)，基因73(1), 237-244)。一旦軟體生成最佳比對，可以計算同源性%，較佳的是序列一致性%。軟體典型地進行作為序列比較的一部分的此計算並且生成多個結果。該等序列還可以具有胺基酸殘基的缺失、插入或取代，這產生沈默變化並且形成功能上等效的物質。可以基於胺基酸特徵（諸如殘基的極性、電荷、可溶性、疏水性、親水性、和/或兩親性）的類似性來進行有目的的胺基酸取代並且因此將胺基酸以官能團分組在一起是有用的。胺基酸可以基於其單獨的側鏈的特徵來分組在一起。然而，包括突變資料也會更有用。出於結構原因，因此衍生的幾組胺基酸可能是保守的。這幾組可以卞氏圖表形式描述（利文斯敦·C.D.（Livingstone C.D.）和巴頓·G.J.（Barton G.J.）(1993)“蛋白質序列比對：用於分層分析殘基保守性的策略（Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation）”*生物科學中的電腦*應用

（Comput. Appl Biosci.）9: 745-756）（泰勒·W.R.（Taylor W.R.）(1986)“胺基酸保守性分類（The classification of amino acid conservation）”理論生物學雜誌（*J. Theor.Biol.*）119;205-218）。可以例如根據下表進行保守性取代，該表描述了普遍接受的胺基酸卞氏圖表分組。

組		子組	
疏水性	F W Y H K M I L V A G C	芳香族	F W Y H
		脂肪族	I L V
極性	W Y H K R E D C S T N Q	帶電荷	H K R E D
		帶正電荷	H K R
		帶負電荷	E D
小	V C A G S P T N D	微小	A G S

[0824] 在此可互換使用的術語“受試者”、“個體”和“患者”係指脊椎動物，較佳的是哺乳動物，更較佳的是人類。哺乳動物包括但不限於，鼠類、猴類、人類、家畜、競技動物、以及寵物。還涵蓋體內獲得或體外培養的生物實體的組織、細胞以及其子代。

[0825] 術語“治療藥”、“能治療的藥劑”或“治療劑”係可互換使用的並且係指當給予受試者時賦予一些有利作用的分子或化合物。有利作用包括實現診斷確定；緩解疾病、症狀、病症或病理病狀；減少或預防疾病、症狀、病症或病狀的發作；以及大體上消除疾病、症狀、病症或病理病狀。

[0826] 如在此所用，“治療”或“進行治療”或“減輕”或“緩解”係在此可互換使用的。該等術語係指一種用於獲得有利或希望的結果的方法，該等結果包括但不限於治療益處和/或預防益處。治療益處意指在治療中的一種或多種疾病、病狀、症狀中的任何治療上的相關改進或對該等疾病的作用。對於預防益處，組成物可

以給予至處於發展具體的疾病、病狀或症狀的風險中的受試者，或給予報告疾病的一種或多種生理學症狀的受試者，儘管這種疾病、病狀或症狀可能還未得到證實。

[0827] 術語“有效量”或“治療有效量”係指藥劑足以實現有利或希望的結果的量。治療有效量可以根據以下各項中的一種或多種來改變：受試者和正在治療的疾病病狀、受試者的體重和年齡、疾病病狀的嚴重性、給藥方式等，該等可以容易藉由熟習該項技術者確定。該術語還適合於將藉由在此所述的任何一種成像方法提供檢測圖像的劑量。該特定劑量可以根據以下各項中的一種或多種來改變：所選擇的具體藥劑、隨後的給藥方案（無論它是否與其他化合物組合）、給藥時間、成像的組織、以及其中攜帶它的物理遞送系統。

[0828] 本發明的若干方面涉及包含一種或多種載體的載體系統或這樣的載體。載體可以被設計為在原核細胞或真核細胞中表現CRISPR轉錄物（例如，核酸轉錄物、蛋白質或酶）。例如，CRISPR轉錄物可以在細菌細胞諸如大腸桿菌、昆蟲細胞（使用桿狀病毒表現載體）、酵母細胞、或哺乳動物細胞中表現。適合的宿主細胞在高德爾，基因表現技術：酶學方法185，學術出版社，加利福尼亞州聖迭哥(1990)中進行進一步討論。可替代地，重組表現載體可以例如使用T7啟動子調節序列和T7聚合酶來進行體外轉錄和翻譯。

[0829] 本發明的實施方式包括可以含有可發生的同源取代（取代和替換二者在此用於意指存在的胺基酸殘基或核苷酸與替

代性殘基或核苷酸的互換）的序列（多核苷酸或多肽二者），該同源取代即在胺基酸的情況下的同比取代，諸如鹼對鹼、酸對酸、極性對極性等。也可以發生非同源性取代，即從一類殘基到另一類殘基或者可替代地涉及包含非天然胺基酸諸如鳥胺酸（在下文中稱為Z）、二胺基丁酸鳥胺酸（在下文中稱為B）、正亮胺酸鳥胺酸（在下文中稱為O）、吡啶基丙胺酸、噻吩丙胺酸、萘基丙胺酸以及苯基甘胺酸。變體胺基酸序列可以包含適合的間隔基團，該等間隔基團可以插入在該序列的任何兩個胺基酸殘基之間，包括烷基諸如甲基、乙基或丙基以及胺基酸間隔物諸如甘胺酸或β-丙胺酸殘基。涉及在類肽形式中存在一個或多個胺基酸殘基的另一種變型形式可以被熟習該項技術者很好地理解。為免生疑，"類肽形式"用於指示變體胺基酸殘基，其中α-碳取代基係在殘基的氮原子上而不是α-碳上。用於製備類肽形式的肽的方法係本領域已知的，例如西蒙·RJ（Simon RJ）等人，美國國家科學院院刊（*PNAS*）(1992) 89(20), 9367-9371以及奧爾韋爾·DC（Horwell DC），生物技術趨勢（*Trends Biotechnol.*）(1995) 13(4), 132-134。

[0830] 同源建模：在其他Cpf1異種同源物中的相應殘基可以藉由以下方法來鑒定：張等人，2012（自然（Nature）；490(7421): 556-60）和陳等人，2015（科學公共圖書館計算生物學（PLoS Comput Biol）；11(5): e1004248）—預測由結構域模體介面介導的相互作用的計算蛋白質-蛋白質相互作用（PPI）方法。PrePPI（預測的PPI）係一種基於結構的PPI預測方法，該方法使用貝葉斯統計框架將結構證據與非結構證據組合。該方法涉及查詢蛋白質的鹼基對並且使用結構比對鑒定與其實驗上確定的結構或同源模型

相對應的結構示意圖。結構比對進一步用於藉由考慮整體和局部幾何關係來鑒定近處和遠處的結構鄰近物。無論何時結構示意圖的兩個鄰近物形成蛋白質資料庫中報導的複合物，這定義了用於建模兩種查詢蛋白質之間的相互作用的模板。複合物模型係藉由在模板的相應結構鄰近物上疊加代表性結構來創建的。此方法進一步描述於戴伊等人，2013(蛋白質科學(Prot Sci)；22: 359-66)。

[0831] 出於本發明的目的，擴增意指採用能夠以適當保真度複製靶序列的引物和聚合酶的任何方法。擴增可以是藉由天然或重組DNA聚合酶諸如TaqGold™、T7 DNA聚合酶、大腸桿菌DNA聚合酶的克列諾片段、以及逆轉錄酶。一種較佳的擴增方法係PCR。

[0832] 在某些方面中，本發明涉及載體。如在此所用，“載體”係一種允許或說明實體從一個環境轉移到另一個環境中的工具。它係複製子，諸如質粒、噬菌體或黏粒，可以向該複製子中插入另一個DNA區段以便使得該插入的區段複製。通常，載體在與適當控制元件締合時能夠複製。總的來說，術語“載體”係指能夠轉運它所連接的另一個核酸的核酸分子。載體包括但不限於，單股、雙股或部分雙股的核酸分子；包含一個或多個游離端、不包含游離端（例如，環狀）的核酸分子；包含DNA、RNA或二者的核酸分子；以及本領域已知的其他種類多核苷酸。一種類型的載體係“質粒”，該質粒係指環狀雙股DNA環，可以諸如藉由標準分子選殖技術向該環中插入另外的DNA區段。另一種類型的載體係病毒載體，其中病毒來源的DNA或RNA序列存在於包裝到病毒（例如，逆轉錄病毒、複製缺陷型逆轉錄病毒、腺病毒、複製缺

陷型腺病毒、以及腺相關病毒（AAV））中的載體中。病毒載體還包括由轉染到宿主細胞中的病毒攜帶的多核苷酸。某些載體能夠在引入它們的宿主細胞中自主複製（例如，具有細菌複製起點的細菌載體和附加型哺乳動物載體）。其他載體（例如，非附加型哺乳動物載體）在引入到宿主細胞後被整合到宿主細胞的基因組中，並且因此隨著宿主基因組一起複製。此外，某些載體能夠引導它們可操作地連接的基因的表現。此類載體在此被稱為"表現載體"。在重組DNA技術中實用的常見表現載體常常是質粒形式。

[0833] 重組表現載體可以包含處於適用於核酸在宿主細胞中表現的形式的本發明的核酸，這意味著重組表現載體包含一個或多個調節元件，該等調節元件可以基於用於表現的宿主細胞來選擇，可操作地連接至有待表現的核酸序列。在重組表現載體內，"可操作地連接"旨在意指感興趣的核苷酸序列以允許核苷酸序列表現（例如，在體外轉錄/翻譯系統中或當該載體被引入到宿主細胞時在宿主細胞中）的方式連接至一個或多個調節元件。關於重組和選殖方法，參考2004年9月2日作為US 2004-0171156 A1公開的美國專利申請10/815,730，該等專利的內容藉由引用以其整體結合在此。

[0834] 本發明的多個方面涉及指導RNA和（任選修飾或突變的）CRISPR酶（例如，Cpf1）的雙順反子載體。指導RNA和（任選修飾或突變的）CRISPR酶的雙順反子表現載體係較佳的。總的來說並且具體的說，在此實施方式中（任選修飾或突變的）CRISPR酶係較佳的是藉由CBh啟動子驅動。該RNA可以是較佳的是藉由

Pol III啟動子諸如U6啟動子驅動。理想的是，將兩者組合。

[0835] 在一些實施方式中，提供了指導RNA中的環。這可以是髮夾環或四員環。該環較佳的是GAAA，但不限於此序列或者確實係僅4 bp的長度。實際上，用於髮夾結構中的較佳的成環序列的長度係四個核苷酸，並且最較佳的是具有序列GAAA。然而，也可以使用較長或較短的環序列，如可以是替代性序列。該等序列較佳的是包含核苷酸三聯體（例如，AAA）和另一個核苷酸（例如C或G）。成環序列的實例包括CAAA和AAAG。在實踐在此揭露的任何方法時，可以經由本領域已知的一種或多種方法來將適合的載體引入到細胞或胚胎中，該等方法包括但不限於，微注射、電穿孔、聲孔效應、基因槍、磷酸鈣介導的轉染、陽離子轉染、脂質體轉染、樹枝狀轉染、熱激轉染、核轉染、磁轉染、脂轉染、刺穿轉染、光學轉染、專有劑增強的核酸攝取、以及經由脂質體、免疫脂質體、病毒體或人工病毒體遞送。在一些方法中，載體藉由微注射引入到胚胎中。這種或該等載體可以微注射到胚胎的核或胞質中。在一些方法中，這種或該等載體可以藉由核轉染引入到細胞中。

[0836] 術語“調節元件”旨在包括啟動子、增強子、內部核糖體進入位點（IRES）以及其他表現控制元件（例如，轉錄終止信號，諸如多聚腺苷酸化信號和聚U序列）。此類調節元件描述於例如高德爾，基因表現技術：酶學方法185，學術出版社，加利福尼亞州聖迭哥(1990)中。調節元件包括引導核苷酸序列在許多類型的宿主細胞中連續表現的那些元件和引導核苷酸序列僅在某些宿

主細胞中表現的那些元件（例如，組織特異性調節序列）。組織特異性啟動子可以引導主要在希望的感興趣的組織諸如肌肉、神經元、骨骼、皮膚、血液、特定器官（例如，肝臟、胰臟）、或特定細胞類型（例如，淋巴細胞）中的表現。調節元件還可以時間依賴性方式諸如細胞週期依賴性或發育階段依賴性方式引導表現，這可以是或也可以不是組織特異性或細胞類型特異性的。在一些實施方式中，載體包含一個或多個pol III啟動子（例如，1、2、3、4、5、或更多個pol III啟動子）、一個或多個pol II啟動子（例如，1、2、3、4、5、或更多個pol II啟動子）、一個或多個pol I啟動子（例如，1、2、3、4、5、或更多個pol I啟動子）、或其組合。pol III啟動子的實例包括但不限於，U6和H1啟動子。pol II啟動子的實例包括但不限於，逆轉錄病毒勞斯氏肉瘤病毒（RSV）LTR啟動子（視情況具有RSV增強子）、巨細胞病毒（CMV）啟動子（視情況具有CMV增強子）[例如，參見博沙特等人，細胞，41:521-530(1985)]、SV40啟動子、二氫葉酸還原酶啟動子、β-肌動蛋白啟動子、磷酸甘油激酶（PGK）啟動子、以及EF1α啟動子。術語“調節元件”還涵蓋增強子元件，諸如WPRE；CMV增強子；HTLV-I的LTR中的R-U5’區段（分子細胞生物學，第8(1)卷，第466-472頁，1988）；SV40增強子；以及兔β-球蛋白的外顯子2與3之間的內含子序列（美國國家科學院院刊，第78(3)卷，第1527-31頁，1981）。熟習該項技術者將瞭解的是，表現載體的設計可以取決於諸如有待轉化的宿主細胞的選擇、所希望的表現水平等因素。載體可以引入到宿主細胞中從而產生由在此所述的核酸編碼的轉錄物、蛋白質或肽，包括融合蛋白或肽（例如，成簇規律間隔短迴文重複

序列（CRISPR）轉錄物、蛋白質、酶、其突變體形式、其融合蛋白等）。關於調節序列，參考美國專利申請10/491,026，該專利申請的內容藉由引用以其整體結合在此。關於啟動子，參考PCT公開WO 2011/028929和美國申請12/511,940，該等專利申請的內容藉由引用以其整體結合在此。

[0837] 載體可以被設計為在原核細胞或真核細胞中表現CRISPR轉錄物（例如，核酸轉錄物、蛋白質或酶）。例如，CRISPR轉錄物可以在細菌細胞諸如大腸桿菌、昆蟲細胞（使用桿狀病毒表現載體）、酵母細胞、或哺乳動物細胞中表現。適合的宿主細胞在高德爾，基因表現技術：酶學方法185，學術出版社，加利福尼亞州聖迭哥(1990)中進行進一步討論。可替代地，重組表現載體可以例如使用T7啟動子調節序列和T7聚合酶來進行體外轉錄和翻譯。

[0838] 載體可以在原核生物或原核細胞中引入並增殖。在一些實施方式中，使用原核生物擴增有待引入到真核細胞的載體拷貝或者作為產生有待引入真核細胞的載體的中間載體（例如，擴增作為病毒載體包裝系統的一部分的質粒）。在一些實施方式，使用原核生物擴增載體拷貝並表現一種或多種核酸，以便提供用於遞送至宿主細胞或宿主生物體的一種或多種蛋白質來源。原核生物中的蛋白質表現最常在具有載體的大腸桿菌中進行，該等載體含有引導融合蛋白或非融合蛋白表現的組成型啟動子或誘導型啟動子。融合載體將許多胺基酸添加到其中編碼的蛋白質，諸如添加到重組蛋白的胺基末端。該等融合載體可以用於一種或多種目

的，諸如：(i) 增加重組蛋白的表現；(ii) 增加重組蛋白的溶解度；以及 (iii) 藉由充當親和純化中的配位基來幫助純化重組蛋白。通常，在融合表現載體中，蛋白水解切割位點被引入在融合部分與重組蛋白的接點處，以使得重組蛋白能夠與融合部分分離，從而隨後純化該融合蛋白。此類酶及其同源識別序列包括凝血因子Xa、凝血酶以及腸激酶。示例性融合表現載體包括pGEX（法瑪西亞生物技術公司（Pharmacia Biotech Inc）；史密斯和詹森（Johnson），1988.基因67: 31-40）、pMAL（普利茅斯貝弗莉的新英格蘭生物實驗室（New England Biolabs, Beverly, Mass.））以及pRIT5（新澤西州皮斯卡塔韋的法瑪西亞公司（Pharmacia, Piscataway, N.J.）），它們將麩胱甘肽 S-轉移酶（GST）、麥芽糖E結合蛋白或蛋白A融合至靶重組蛋白。適合的誘導型非融合大腸桿菌表現載體的實例包括pTrc（阿蘭恩（Amrann）等人，(1988)基因69:301-315）和pET 11d（司圖登爾（Studier）等人，基因表現技術：酶學方法185，學術出版社，加利福尼亞州聖迭哥(1990) 60-89）。在一些實施方式中，載體係酵母表現載體。用於在酵母釀酒酵母中表現的載體的實例包括pYepSec1（班得瑞（Baldari）等人，1987.歐洲分子生物學學會雜誌（EMBO J.）6: 229-234）、pMFa（奎讓（Kuijan）和赫斯奎茲（Herskowitz），1982.細胞 30: 933-943）、pJRY88（舒爾茨（Schultz）等人，1987.基因 54: 113-123）、pYES2（加利福尼亞州聖達戈的英傑公司（Invitrogen Corporation, San Diego, Calif.））以及picZ（加利福尼亞州聖達戈的英傑公司）。在一些實施方式中，載體在使用桿狀病毒表現載體的昆蟲細胞中驅動蛋白質表現。可用於在培養的昆蟲細胞（例如，SF9細胞）表現蛋白質的桿狀病毒

載體包括pAc系列（史密斯等人，1983分子細胞生物學 3: 2156-2165）和pVL系列（拉克樓（Lucklow）和薩默斯（Summers），1989.病毒學（Virology）170: 31-39）。

[0839] 在一些實施方式中，載體能夠使用哺乳動物表現載體驅動一種或多種序列在哺乳動物細胞中表現。哺乳動物表現載體的實例包括pCDM8（錫德（Seed），1987.自然329: 840）和pMT2PC（考夫曼（Kaufman）等人，1987.歐洲分子生物學學會雜誌6: 187-195）。當用於哺乳動物細胞時，表現載體的控制功能典型地是由一個或多個調節元件提供的。例如，常用的啟動子係來源於多瘤、腺病毒2、巨細胞病毒、猿猴病毒40、以及在此揭露和本領域已知的其他來源。對於原核細胞和真核細胞二者的其他適合表現系統，參見例如薩姆布魯克等人，分子選殖：實驗室手冊（MOLECULAR CLONING: A LABORATORY MANUAL.）第2版，冷泉港實驗室（Cold Spring Harbor Laboratory），冷泉港實驗室出版社（Cold Spring Harbor Laboratory Press），紐約冷泉港（Cold Spring Harbor, N.Y.），1989中的第16章和第17章。

[0840] 在一些實施方式中，重組哺乳動物表現載體能夠引導核酸優先在特定細胞類型中表現（例如，組織特異性調節元件用於表現核酸）。組織特異性調節元件係本領域已知的。適合的組織特異性啟動子的非限制性實例包括白蛋白啟動子（肝特異性；平克特（Pinkert）等人，1987.*基因與發育（Genes Dev.）*1: 268-277）、淋巴特異性啟動子（卡拉梅（Calame）和伊頓（Eaton），1988.*免疫學進展（Adv.Immunol.）*43: 235-275）（具體地說T細胞受體（維

諾托（Winoto）和巴爾的摩（Baltimore），1989.歐洲分子生物學學會雜誌8: 729-733）和免疫球蛋白類（班恩吉（Baneiji）等人，1983.細胞33: 729-740；奎因（Queen）和巴爾的摩，1983.細胞 33: 741-748）的啟動子）、神經元特異性啟動子（例如，神經絲啟動子；伯恩（Byrne）和瑞德爾（Ruddle），1989.*美國國家科學院院刊*）86: 5473-5477）、胰臟特異性啟動子（埃德隆德（Edlund）等人，1985.科學（*Science*）230: 912-916）、以及乳腺特異性啟動子（乳清啟動子；美國專利案號4,873,316和歐洲申請公開號264,166）。還涵蓋發育調節啟動子，例如鼠科hox啟動子（克塞爾（Kessel ）和（格魯斯），1990.科學249: 374-379）和α-胎蛋白啟動子（康珀斯（Campes）和蒂爾曼（Tilghman），1989.*基因與發育*3: 537-546）。關於原核載體和真核載體，參考美國專利6,750,059，該專利的內容藉由引用以其整體結合在此。本發明的其他實施方式可以涉及病毒載體的使用，關於此使用參考美國專利申請13/092,085，該專利申請的內容藉由引用以其整體結合在此。組織特異性調節元件係本領域已知的並且就這一點而言，參考美國專利7,776,321，該專利的內容藉由引用以其整體結合在此。在一些實施方式中，調節元件可操作地連接至CRISPR系統的一個或多個元件，以便驅動該CRISPR系統的一個或多個元件表現。總的來說，CRISPR（成簇規律間隔短迴文重複序列）也稱為SPIDR（間隔區間隔同向重複序列），它構成通常對特定細菌種類特異的DNA座位家族。CRISPR座位包括大腸桿菌中識別的不同類別的間隔短序列重複序列（SSR）（石野（Ishino）等人，細菌學雜誌（J. Bacteriol.），169:5429-5433 [1987]；以及中田（Nakata）等人，細菌學雜誌，171:3553-

3556 [1989])、以及相關基因。類似的間隔SSR已在地中海富鹽菌、釀膿鏈球菌、魚腥藻屬、以及結核分枝桿菌中鑒定（參見，葛籣恩（Groenen）等人，分子微生物學（Mol.Microbiol.），10:1057-1065 [1993]；霍（Hoe）等人，新發傳染病（Emerg.Infect.Dis.），5:254-263 [1999]；馬塞波爾（Masepohl）等人，生物化學與生物物理學雜誌（Biochim.Biophys.Acta 1307:26-30 [1996]；以及莫西卡（Mojica）等人，分子微生物學），17:85-93 [1995]）。CRISPR座位與其他SSR的典型不同之處在於重複序列結構，該結構稱為短規律間隔重複序列（SRSR）（詹森（Janssen）等人，組學：整合生物學雜誌（OMICS J. Integ.Biol.），6:23-33 [2002]；以及莫西卡等人，分子微生物學），36:244-246 [2000]）。總的來說，該等重複序列係成簇出現的短元件，它們由具有基本上恒定的長度的獨特間插序列規律地間隔開（莫西卡等人，[2000]，同上）。儘管該等重複序列在菌株之間係高度保守的，但是間隔重複序列的數目和間隔區的序列典型地因菌株不同而不同（凡埃姆登（van Embden）等人，細菌學雜誌，182:2393-2401 [2000]）。CRISPR座位已在超過40種原核生物中鑒定（例如，參見，詹森等人，分子微生物學，43:1565-1575 [2002]；以及莫西卡等人，[2005]），包括但不限於，氣火菌屬、火棒菌屬、硫化葉菌屬、古生球菌屬、鹽盒菌屬、甲烷桿菌屬、甲烷球菌屬、甲烷八疊球菌屬、甲烷八疊球菌屬、火球菌屬、嗜酸菌屬、熱原體屬、棒狀桿菌屬、分枝桿菌屬、鏈黴菌屬、產水菌屬、卟啉單胞菌屬、綠硫菌屬、棲熱菌屬、芽孢桿菌屬、李斯特菌屬、葡萄球菌屬、梭菌屬、高溫厭氧桿菌屬、支原體屬、梭菌屬、固氮弓菌屬（Azarcus）、色桿菌屬、奈瑟氏菌屬、亞硝化單胞

菌屬、脫硫弧菌屬、地桿菌屬、黏球菌屬、彎曲桿菌屬、沃廉菌屬、不動桿菌屬、歐文氏菌屬、埃希菌屬、軍團桿菌屬、甲基球菌屬、巴氏桿菌屬、發光桿菌屬、沙門氏菌屬、黃單胞菌屬、耶爾森菌屬、密螺旋體屬、以及熱袍菌屬。

[0841] 總的來說，如本申請所用的“核酸靶向系統”總體上是指涉及表現核酸靶向CRISPR相關（“Cas”）基因或引導該等基因活性的轉錄物和其他元件（在此也稱為效應蛋白），該等基因包括編碼核酸靶向Cas（效應）蛋白和指導RNA的序列或來自核酸靶向CRISPR座位的其他序列和轉錄物。在一些實施方式中，核酸靶向系統的一個或多個元件係來源於V型/VI型核酸靶向CRISPR系統。在一些實施方式中，核酸靶向系統的一個或多個元件係來源於包含內源性核酸靶向CRISPR系統的特定生物體。總的來說，核酸靶向系統的特徵係在靶序列的位點處促進核酸靶向複合物形成的元件。在形成核酸靶向複合物的情況下，“靶序列”係指指導序列被設計為與其具有互補性的序列，其中靶序列與指導RNA之間的雜交促進了DNA或RNA靶向複合物的形成。並不一定需要完全互補，只要存在引起雜交並且促進核酸靶向複合物形成的足夠互補性。靶序列可以包括RNA多核苷酸。在一些實施方式中，靶序列位於細胞的核或胞質中。在一些實施方式中，該靶序列可以是在真核細胞的細胞器中，例如線粒體或葉綠體。可以用於重組到包含靶序列的靶向座位中的序列或模板被稱為“編輯模板”或“編輯RNA”或“編輯序列”。在本發明的多個方面中，外源性模板RNA可以被稱為編輯模板。在本發明的一個方面中，重組係同源重組。

[0842] 典型地，在內源性核酸靶向系統的情況下，核酸靶向複合物（包含雜交至靶序列並與一種或多種核酸靶向效應蛋白複合的指導RNA）的形成產生靶序列中或靶序列附近（例如，從靶序列開始的1、2、3、4、5、6、7、8、9、10、20、50、或更多個鹼基對內）的一條或兩條RNA股的切割。在一些實施方式中，驅動核酸靶向系統的一個或多個元件表現的一種或多種載體被引入到宿主細胞中，以使得該核酸靶向系統的該等元件的表現能引導核酸靶向複合物在一個或多個靶位點處形成。例如，核酸靶向效應蛋白和指導RNA可以各自可操作地連接至單獨載體上的單獨調節元件。可替代地，由相同或不同調節元件表現的該等元件的兩種或更多種可以組合在單一載體中，其中一種或多種另外的載體提供核酸靶向系統在第一載體中不包含的任何組分。在單一載體中組合的核酸靶向系統元件可以佈置為任何適合的取向，諸如一個元件位於相對於第二元件的5'（“上游”）或相對於該第二元件的3'（“下游”）。一個元件的編碼序列可以位於第二元件的編碼序列的相同股或相反股上，並且取向為相同或相反方向。在一些實施方式中，單一啟動子驅動編碼核酸靶向效應蛋白的轉錄物和嵌入一種或多種內含子序列之內（例如，各自在不同內含子中、兩種或更多種在至少一個內含子中，或所有在單一內含子中）的指導RNA的表現。在一些實施方式中，核酸靶向效應蛋白和指導RNA可操作地連接至同一啟動子並且從該同一啟動子表現。

[0843] 總的來說，指導序列係與靶多核苷酸序列具有足夠互補性以與靶序列雜交並引導核酸靶向複合物與靶序列的序列特異性結合的任何多核苷酸序列。在一些實施方式中，當使用適合比

對演算法進行最佳比對時，指導序列與其相應靶序列之間的互補程度係約或超過約50%、60%、75%、80%、85%、90%、95%、97.5%、99%、或更大。最佳比對可以藉由使用用於比對序列的任何適合的演算法來確定，該等演算法的非限制性實例包括史密斯-沃特曼演算法、尼德曼-溫施演算法、基於巴羅斯-惠勒轉換的演算法(例如，巴羅斯-惠勒比對儀)、ClustalW、Clustal X、BLAT、諾沃比對（諾沃克拉夫特技術公司，ELAND（加利福尼亞州聖迭哥億明達公司))、SOAP（可在soap.genomics.org.cn處獲得）、以及Maq（可在maq.sourceforge.net處獲得）。在一些實施方式中，指導序列的長度係約或超過約5、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、35、40、45、50、75、或更多個核苷酸。在一些實施方式中，指導序列的長度係小於約75、50、45、40、35、30、25、20、15、12、或更少個核苷酸。指導序列引導核酸靶向複合物與靶序列的序列特異性結合的能力可以是藉由任何適合的測定來評估。例如，足以形成核酸靶向複合物的核酸靶向系統的組分（包括有待測試的指導序列）可以諸如藉由用編碼核酸靶向CRISPR序列的該等組分的載體進行轉染來提供給具有相應靶序列的宿主細胞，隨後諸如藉由在此所述的Surveyor測定評估靶序列內或附近的優先切割。類似地，靶多核苷酸序列（或其附近的序列）的切割可以在試管中藉由以下方式進行評估：提供靶序列、核酸靶向複合物的組分（包括有待測試的指導序列）和不同於測試指導序列的對照指導序列並且在測試指導序列反應與對照指導序列反應之間比較靶序列處或附近的結合或切割速率。其他測定係可能的，並且將是熟習該項技術者能夠

想到的。

[0844] 指導序列可以被選擇為靶向任何靶序列。在一些實施方式中，靶序列係基因轉錄物或mRNA中的序列。

[0845] 在一些實施方式中，靶序列係細胞基因組中的序列。

[0846] 在一些實施方式中，指導序列被選擇為減小該指導序列內的二級結構程度。二級結構可以是藉由任何適合的多核苷酸折疊演算法來確定。一些程式係基於計算最小吉布斯自由能。一種這樣演算法的實例係mFold，如藉由朱克和施蒂格勒（核酸研究9 (1981)，133-148）。另一個示例性折疊演算法係維也納大學的理論化學研究所使用質心結構預測演算法開發的線上網站伺服器RNAfold（例如，參見A.R.·格魯伯（A.R. Gruber）等人，2008，細胞106(1): 23-24；以及PA·凱爾（PA Carr）和GM·丘奇（GM Church），2009，自然生物技術（*Nature Biotechnology*）27(12): 1151-62）。其他演算法可以見於美國申請案序號TBA（代理檔號44790.11.2022；廣泛參考BI-2013/004A）；藉由引用結合在此。

[0847] 在一些實施方式中，還提供了重組模板。重組模板可以是如在此所述的另一種載體的組分，它包含在單獨的載體中或者作為單獨的多核苷酸提供。在一些實施方式中，重組模板被設計為充當同源重組中的模板，諸如在由作為核酸靶向複合物的一部分的核酸靶向效應蛋白切割或分解的靶序列內或附近。模板多核苷酸可以具有任何適合的長度，諸如長度係約或超過約10、15、20、25、50、75、100、150、200、500、1000或更多個核苷酸。在一些實施方式中，模板多核苷酸係與包含該靶序列的多核苷酸部

分互補的。當最佳比對時，模板多核苷酸可以與靶序列的一個或多個核苷酸重疊（例如約或超過約1、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100或更多個核苷酸）。在一些實施方式中，當模板序列和包含靶序列的多核苷酸最佳比對時，模板多核苷酸最近的核苷酸係在來自靶序列的約1、5、10、15、20、25、50、75、100、200、300、400、500、1000、5000、10000或更多個核苷酸中。

[0848] 在一些實施方式中，核酸靶向效應蛋白係包含一個或多個異源蛋白結構域（例如，約或超過約1、2、3、4、5、6、7、8、9、10或更多個結構域，還有核酸靶向效應蛋白）的融合蛋白的一部分。在一些實施方式中，CRISPR效應蛋白係包含一個或多個異源蛋白結構域（例如，約或超過約1、2、3、4、5、6、7、8、9、10或更多個結構域，還有CRISPR酶）的融合蛋白的一部分。CRISPR酶融合蛋白可以包含任何另外的蛋白序列和視情況在任何兩個結構域之間的接頭序列。可以融合至CRISPR酶的蛋白結構域的實例包括但不限於，表位標籤、報導基因序列、以及具有以下活性中的一種或多種活性的蛋白結構域：修飾烷酶活性、去甲酶活性、轉錄啟動活性、轉錄阻遏活性、轉錄釋放因子活性、組蛋白修飾活性、RNA切割活性以及核酸結合活性。表位標籤的非限制性實例包括組胺酸（His）標籤、V5標籤、FLAG標籤、流感血球凝集素（HA）標籤、Myc標籤、VSV-G標籤、以及硫氧還蛋白（Trx）標籤。報導基因的實例包括但不限於，麩胱甘肽-S-轉移酶（GST）、辣根過氧化物酶（HRP）、氯黴素乙醯轉移酶（CAT）、β-半乳糖苷酶、β-葡萄糖醛酸酶、螢光素酶、綠色螢光蛋白（GFP）、

HcRed、DsRed、青色螢光蛋白（CFP）、黃色螢光蛋白（YFP）、以及自身螢光蛋白（包括藍色螢光蛋白（BFP））。CRISPR酶可以融合至編碼結合DNA分子或結合其他細胞分子的蛋白質或蛋白質片段的基因序列，該蛋白質包括但不限於，麥芽糖結合蛋白（MBP）、S-標籤、Lex A DNA結合結構域（DBD）融合物、GAL4 DNA結合結構域融合物、以及單純皰疹病毒（HSV）BP16蛋白融合物。可以形成包含CRISPR酶的融合蛋白的一部分的附加結構域描述於US20110059502，該專利藉由引用結合在此。在一些實施方式中，標記的CRISPR酶用於識別靶序列的位置。

[0849] 在一些實施方式中，CRISPR酶可以形成誘導型系統的一種組分。該系統的誘導性質允許使用能量形式時間空間控制基因編輯或基因表現。能量形式可以包括但不限於，電磁輻射、聲能、化學能以及熱能。誘導型系統的實例包括四環素誘導型啟動子（Tet-開或Tet-關）、小分子雙雜交轉錄啟動系統（FKBP、ABA等）、或光誘導型系統（光敏色素、LOV結構域或隱花色素）。在一個實施方式中，CRISPR酶可以是以序列特異性方式引導轉錄活性改變的光誘導型轉錄效應物（LITE）的一部分。光誘導型系統的組分可以包括CRISPR酶、光反應性細胞色素異源二聚體（例如，來自阿拉伯芥）、以及轉錄啟動/阻遏結構域。誘導型DNA結合蛋白及其使用方法的其他實例提供於US 61/736465和US 61/721,283以及WO 2014/018423以及US8889418、US8895308、US20140186919、US20140242700、US20140273234、US20140335620、WO2014093635中，該等專利藉由引用以其整體結合在此。

[0850]　遞送

[0851]　在一些方面中，本發明提供了包括以下各項的方法：向宿主細胞遞送一種或多種多核苷酸，諸如在此所述的一種或多種載體、其一種或多種轉錄物和/或由其轉錄的一種或多種蛋白質。在一些方面中，本發明進一步提供了藉由此類方法產生的細胞以及包含此類細胞或由此類細胞產生的生物體（諸如動物、植物或真菌）。在一些實施方式中，將核酸靶向效應蛋白與指導RNA組合（以及視情況與其複合）遞送到細胞中。常規的基於病毒和非病毒的基因轉移方法可以用於在哺乳動物細胞或靶組織中引入核酸。此類方法可以用於向培養基或宿主生物體中的細胞給予編碼核酸靶向系統的組分的核酸。非病毒載體遞送系統包括DNA質粒、RNA（例如在此所述的載體的轉錄物）、裸核酸、以及與遞送媒介物諸如脂質體複合的核酸。病毒載體遞送系統包括DNA和RNA病毒，該等病毒在遞送至細胞後具有附加型基因組或整合型基因組。對於基因治療程序的綜述，參見安德森（Anderson），科學 256:808-813 (1992)；納貝爾（Nabel）和費爾格納（Felgner），TIBTECH 11:211-217 (1993)；三穀（Mitani）和卡斯基（Caskey），TIBTECH 11:162-166 (1993)；狄龍（Dillon），TIBTECH 11:167-175 (1993)；米勒（Miller），自然 357:455-460 (1992)；範布倫特（Van Brunt），生物技術（Biotechnology） 6(10):1149-1154 (1988)；比涅（Vigne），恢復神經學和神經科學（Restorative Neurology and Neuroscience） 8:35-36 (1995)；克雷默（Kremer）和佩里科代（Perricaudet），英國醫學公報（British Medical Bulletin） 51(1):31-44 (1995)；哈塔塔（Haddada）等人，微生物學和免疫學的前沿課題（Current Topics

in Microbiology and Immunology)，竇爾弗勒（Doerfler）和博姆（Böhm)(編輯)(1995)；以及餘(Yu)等人，基因治療(Gene Therapy) 1:13-26 (1994)。

[0852] 非病毒遞送核酸的方法包括包括脂質轉染、核轉染、微注射、基因槍、病毒體、脂質體、免疫脂質體、聚陽離子或脂質:核酸軛合物、裸DNA、人工病毒體遞送、以及藥劑增強DNA攝取。脂質轉染描述於例如美國專利案號5,049,386、4,946,787；以及4,897,355）並且脂質轉染試劑係商業上銷售的（例如，Transfectam™和Lipofectin™)。適用於多核苷酸的有效受體識別脂質轉染的陽離子脂質和中性脂質包括以下各項中的那些：費爾格納WO 91/17424；WO 91/16024。遞送可以是遞送至細胞（例如，體外或離體給予）或靶組織（例如，體內給予）。

[0853] 脂質:核酸複合物（包括靶向脂質體，諸如免疫脂質複合物）的製備係熟習該項技術者已熟知的（例如，參見克裡斯特爾（Crystal)，科學270:404-410 (1995)；布萊澤（Blaese）等人，癌症基因治療(Cancer Gene Ther.)2:291-297 (1995)；貝爾(Behr)等人，生物共軛化學（Bioconjugate Chem.）5:382-389 (1994)；雷米（Remy）等人，生物共軛化學5:647-654 (1994)；高（Gao）等人，基因治療2:710-722 (1995)；艾哈邁德（Ahmad）等人,癌症研究52:4817-4820 (1992)；美國專利案號4,186,183、4,217,344、4,235,871、4,261,975、4,485,054、4,501,728、4,774,085、4,837,028、以及4,946,787)。

[0854] 使用用於遞送核酸的基於RNA或DNA病毒的系統利

用了用於將病毒靶向身體內的特定細胞並且將病毒有效負載運輸到核內的高度進化的方法。病毒載體可以直接給予至患者(體內)或者它們可以用於體外處理細胞，並且修飾細胞可以視情況給予至患者(離體)。常規基於病毒的系統可以包括用於基因轉移的逆轉錄病毒、慢病毒、腺病毒、腺相關病毒以及單純皰疹病毒載體。藉由逆轉錄病毒、慢病毒、以及腺相關病毒基因轉移方法整合在宿主基因組中是可能的，這常常導致插入的轉基因長期表現。另外，高轉導效率已在許多不同細胞類型和靶組織中觀察到。

[0855] 逆轉錄病毒的趨向性可以藉由結合外源包膜蛋白來改變，從而擴增靶細胞的潛在靶群體。慢病毒載體係能夠轉導或感染非分裂細胞並典型地產生高病毒滴度的逆轉錄病毒載體。逆轉錄病毒基因轉移系統的選擇因此將取決於靶組織。逆轉錄病毒載體包含具有多至6-10 kb外源序列的包裝容量的順式作用長末端重複序列。最小量順式作用LTR係足以複製並包裝該等載體的，然後使用該等載體將治療基因整合到靶細胞中，以提高永久性轉基因表現。廣泛使用的逆轉錄病毒載體包括基於以下各項的那些：鼠白血病病毒(MuLV)、長臂猿白血病病毒(GaLV)、猴免疫缺陷病毒(SIV)、人免疫缺陷病毒(HIV)、以及其組合(例如，參見，布奇謝爾(Buchscher)等人，病毒學雜誌(J. Virol.) 66:2731-2739 (1992)；約翰(Johann)等人，病毒學雜誌66:1635-1640 (1992)；紹姆內爾費爾特(Sommnerfelt)等人，病毒學176:58-59 (1990)；威爾遜(Wilson)等人，病毒學雜誌63:2374-2378 (1989)；米勒等人，病毒學雜誌65:2220-2224 (1991)；PCT/US94/05700)。在其中短暫表現係較佳的應用中，可以使用基於腺病毒的系統。基於腺

病毒的載體能夠在許多細胞類型中具有極高轉導效率並且並不要求細胞分裂。使用此類載體，已獲得高滴度和表現水平。此載體可以在相對簡單的系統中大量產生。腺相關病毒（"AAV"）載體也可以用於使用靶核酸轉導細胞，例如用於體外產生核酸和肽並且用於體內和離體基因治療程序（例如，參見，韋斯特（West）等人，病毒學（Virology）160:38-47 (1987)；美國專利案號4,797,368；WO 93/24641；科廷（Kotin）等人，人類基因治療（Human Gene Therapy）5:793-801 (1994)；繆斯茲卡（Muzyczka），臨床研究雜誌（J. Clin.Invest.）94:1351 (1994)。重組AAV載體的結構描述於許多出版物中，包括美國專利案號5,173,414；特拉辛（Tratschin）等人，分子細胞生物學5:3251-3260 (1985)；特拉辛等人，分子細胞生物學4:2072-2081 (1984)；賀莫納特和繆斯茲卡（Hermonat & Muzyczka），美國國家科學院院刊81:6466-6470 (1984)；以及薩莫爾斯基（Samulski）等人，病毒學雜誌63:03822-3828 (1989)。

對於DNA/RNA或DNA/DNA或RNA/RNA或蛋白質/RNA的選項

[0856] 在一些實施方式中，CRISPR系統的組分可以不同形式遞送，諸如DNA/RNA或RNA/RNA或蛋白質/RNA的組合。例如，Cpf1可以作為DNA編碼多核苷酸或RNA-編碼多核苷酸或作為蛋白質遞送。指導序列可以作為DNA編碼多核苷酸或RNA遞送。預想所有可能的組合，包括混合遞送形式。

[0857] 在一些實施方式中，所有此類組合（DNA/RNA或DNA/DNA或RNA/RNA或蛋白質/RNA）。

[0858] 在一些實施方式中，當Cpf1以蛋白質形式遞送時，可以將它與一種或多種指導序列預先組裝。

奈米線團

[0859] 另外，CRISPR可以使用奈米線團（nanoclew）遞送，例如如孫·W（Sun W）等人，用抗癌藥物遞送的繭樣可自降解DNA線團（*Cocoon-like self-degradable DNA nanoclew for anticancer drug delivery*），美國化學學會雜誌（J Am Chem Soc.），2014年10月22日；136(42):14722-5. doi: 10.1021/ja5088024. 電子版2014年10月13日；或者在孫·W等人，用於有效遞送基因組編輯的*CRISPR-Cas9*的自組裝DNA奈米線團（*Self-Assembled DNA Nanoclews for the Efficient Delivery of CRISPR-Cas9 for Genome Editing.*），應用化學國際英語版（Angew Chem Int Ed Engl.），2015年10月5日；54(41):12029-33. doi: 10.1002/anie.201506030. 電子版2015年8月27日。

[0860] 除非另外指明，本發明的實施採用處於本領域技能範圍內的免疫學、生物化學、化學、分子生物學、微生物學、細胞生物學、基因組學以及重組DNA的常規技術。參見薩姆布魯克、弗裡奇（Fritsch）和馬尼亞蒂斯（Maniatis），分子選殖：實驗室手冊，第2版(1989)；分子生物學通用方法（F. M.·奧蘇貝爾等人編著(1987)）；系列叢書 酶學方法（學術出版社公司）：PCR 2：實踐方法（M.J.·麥克弗森（M.J.MacPherson）、B.D.·黑姆斯（B.D.Hames）以及G.R.泰勒編著(1995)），哈洛（Harlow）和拉內（Lane）編著(1988)抗體、實驗室手冊和動物細胞培養（ANTIBODIES, A

LABORATORY MANUAL, and ANIMAL CELL CULTURE）（R.I.費施奈伊（Freshney）編著(1987)）。

遺傳和表觀遺傳條件的模型

[0861] 本發明的一種方法可以用於創建可用於建模和/或研究感興趣的遺傳或表觀遺傳條件的植物、動物或細胞，諸如藉由感興趣的突變模型或疾病模型。如在此所用，“疾病”係指受試者的疾病、病狀或適應症。例如，本發明的一種方法可以用於創建包含與疾病相關聯的一種或多種核酸序列中的修飾的動物或細胞、或者其中與疾病相關聯的一種或多種核酸序列的表現發生改變的植物、動物或細胞。這種核酸序列可以編碼疾病相關蛋白序列或者可以是疾病相關控制序列。因此，應理解在本發明的多個實施方式中，植物、受試者、患者、生物體或細胞可以是非人類受試者、患者、生物體或細胞。因此，本發明提供了藉由本發明產生的植物、動物或細胞、或其子代。該子代可以是產生的植物或動物的選殖，或者可以由藉由與相同種類的其他個體雜交以使另外希望的性狀滲入其後代來進行的有性繁殖產生。在多細胞生物體（具體的是動物或植物）的情況下，該細胞可以是體內或離體的。在其中細胞處於培養中的情況下，如果滿足適當的培養條件並且較佳的是如果細胞適合地適用於此目的（例如幹細胞），則可以建立細胞系。還設想藉由本發明產生的細菌細胞系。因此，還設想細胞系。

[0862] 在一些方法中，該疾病模型可以用於使用該疾病研究中常用的措施研究突變對動物或細胞和疾病的發展和/或進展的

影響。可替代地，這種疾病模型適用於研究藥物活性化合物對該疾病的影響。

[0863] 在一些方法中，該疾病模型可以用於評定可能的基因治療策略的效力。即，疾病相關基因或多核苷酸可以被修飾為使得疾病發展和/或進展受到抑制或減少。具體地說，該方法包括修飾疾病相關基因或多核苷酸，以使得能產生改變的蛋白質並因此使得動物或細胞具有改變的應答。因此，在一些方法中，基因修飾動物可以與易於發展疾病的動物比較，以使得可以評定基因治療事件的作用。

[0864] 在另一個實施方式中，本發明提供了一種開發調控與疾病相關基因相關聯的細胞傳訊事件的生物活性劑之方法。該方法包括使測試化合物與包含驅動一種或多種CRISPR酶表現的一種或多種載體和連接至指導序列的同向重複序列的細胞接觸；並且檢測指示與例如細胞所含有的疾病相關基因中的突變相關聯的細胞傳訊事件的減少或增加的讀出變化。

[0865] 細胞模型或動物模型可以與本發明用於篩查細胞功能變化的方法組合來構造。這種模型可以用於研究藉由本發明的CRISPR複合物修飾的基因組序列對感興趣的細胞功能的影響。例如，細胞功能模型可以用研究修飾的基因組序列對細胞內傳訊或細胞外傳訊的影響。可替代地，細胞功能模型可以用研究修飾的基因組序列對感官知覺的影響。在一些此類模型中，修飾模型中與傳訊生物化學途徑相關聯的一個或多個基因組序列。

[0866] 已特別研究了幾種疾病模型。該等包括從新自閉症危

險基因 CHD8、KATNAL2 和 SCN2A；以及綜合症自閉症（安格曼綜合症）基因 UBE3A。該等基因和所得自閉症模型當然是較佳的，但足以顯示本發明對基因和相應模型的廣泛適用性。當與傳訊生物化學途徑相關聯的一個或多個基因組序列與候選藥劑接觸時，該等基因組序列的改變的表現可以是藉由評定測試模型細胞與對照細胞之間相應基因的 mRNA 水平差異來確定。可替代地，與傳訊生物活性途徑相關聯的序列的差異表現係藉由檢測編碼的多肽或基因產物的水平的差異來確定的。

[0867] 為了評定mRNA轉錄物或相應多核苷酸水平的試劑誘導的改變，樣品中含有的核酸首先根據本領域的標準方法來提取。例如，mRNA可以根據薩姆布魯克等人(1989)所列出的程序使用不同的分解酶或化學品溶液來分離，或者藉由核酸結合樹脂遵循製造商提供的附帶說明來提取。然後藉由擴增程序或常規雜交測定（例如，RNA印跡分析）根據本領域廣泛已知的方法或者基於在此舉例說明的方法來檢測提取的核酸樣品中含有的mRNA。

[0868] 出於本發明的目的，擴增意指採用能夠以適當保真度複製靶序列的引物和聚合酶的任何方法。擴增可以是藉由天然或重組DNA聚合酶諸如TaqGold™、T7 DNA聚合酶、大腸桿菌DNA聚合酶的克列諾片段、以及逆轉錄酶。一種較佳的擴增方法係PCR。具體地說，分離的RNA可以經受逆轉錄測定，該測定與定量聚合酶鏈反應（RT-PCR）結合以便量化與傳訊生物化學途徑相關聯的序列的表現水平。

[0869] 基因表現水平的檢測可以是在擴增測定中即時進行

的。在一個方面中，擴增的產物可以用螢光DNA結合劑直接視覺化的，該等結合劑包括但不限於DNA嵌入劑和DNA溝結合劑。因為結合到雙股DNA分子中的嵌入劑的量典型地與擴增的DNA產物的量成比例，所以可以藉由使用本領域的常規光學系統量化嵌入染料的螢光來確定擴增產物的量。適用於此應用的DNA結合染料包括SYBR綠、SYBR藍、DAPI、碘化丙啶、Hoeste, SYBR金、溴化乙錠（ethidium bromide）、吖啶、普羅黃素、吖啶橙、吖啶黃、氟香豆素、玫瑰樹鹼（ellipticine）、道諾黴素（daunomycin）、氯喹、遠端黴素（distamycin）D、色黴素（chromomycin）、胡米溴銨（homidium）、光輝黴素（mithramycin）、聚吡啶釕（ruthenium polypyridyl）、安麯黴素（anthramycin）、以及類似物。

[0870] 在另一個方面中，其他螢光標記諸如序列特異性探針可以用於擴增反應中，以幫助檢測和量化擴增產物。基於探針的定量擴增依賴於希望的擴增產物的序列特異性檢測。它利用螢光的靶特異性探針（例如，TaqMan®探針），從而獲得增加的特異性和靈敏度。本領域中已建立了用於進行基於探針的定量擴增的方法並且在美國專利案號5,210,015中教授了該等方法。

[0871] 在另一個方面中，進行使用雜交探針的常規雜交測定，該等雜交探針與和傳訊生物化學途徑相關聯的序列享有序列同源性。典型地，在雜交反應中允許探針與和來源於測試受試者的生物樣品內的傳訊生物化學途徑相關聯的序列形成穩定的複合物。熟習該項技術者將瞭解的是，在使用反義股作為探針核酸的情況下，樣品中提供的靶多核苷酸被選擇為與反義核酸的序列互補。

相反地，在核苷酸探針係有義核酸的情況下，靶多核苷酸被選擇為與有義核酸的序列互補。

[0872] 雜交可以是在不同嚴格性的條件下進行。用於實踐本發明的適合雜交條件係使得探針與和傳訊生物化學途徑相關聯的序列之間的識別相互作用係足夠特異的且足夠穩定的。增加雜交反應嚴格性的條件係本領域廣泛已知和公開的。例如，參見，(薩姆布魯克等人，(1989)；非輻射原位雜交應用手冊（Nonradioactive In Situ Hybridization Application Manual），德國寶靈曼公司（Boehringer Mannheim），第二版）雜交測定可以使用任何固相支撐體上固定的探針來形成，該固相支撐體包括但不限於硝化纖維、玻璃、矽、以及各種各樣的基因陣列。較佳的雜交測定市在高密度基因晶片上進行的，如美國專利案號5,445,934所述的。

[0873] 對於在雜交測定過程中形成的探針靶標複合物的常規檢測，核苷酸探針被軛合至可檢測標記。適用於本發明的可檢測標記包括藉由光化學手段、生物化學手段、光譜手段、免疫化學手段、電學手段、光學手段或化學手段可檢測的任何組成物。多種多樣的適當可檢測標記係本領域已知的，它們包括螢光標記或化學發光標記、放射性同位素標記、酶標記或其他配位基。在較佳的實施方式中，將可能希望採用螢光標記或酶標記，諸如地地麩新配質、β半乳糖苷酶、脲酶、鹼性磷酸酶或過氧化物酶、抗生物素蛋白/生物素複合物。

[0874] 用於檢測或量化雜交強度的檢測方法將典型地取決於以上選擇的標記。例如，放射性標記可以是使用攝影膠片或感

光成像儀檢測的。螢光標記物可以是使用檢測發射光的光探測器檢測和量化的。酶標記典型地是藉由提供具有底物的酶並測量由酶對該底物的作用產生的反應產物來檢測的；並且最後比色標記係藉由簡單視覺化染色標記來檢測的。

[0875] 與傳訊生物化學途徑相關聯的序列表現的藥劑誘導的變化也可以藉由檢查相應基因產物來確定。測定蛋白質水平典型地涉及 a) 將生物樣品中含有的蛋白質與特異性結合和信號生傳導生物化學途徑相關聯的蛋白質的藥劑接觸；並且 (b) 鑒定所形成的任何藥劑:蛋白質複合物。在此實施方式的一個方面中，特異性結合與傳訊生物化學途徑相關聯的蛋白質的藥劑係抗體，較佳的是單株抗體。

[0876] 該反應係藉由以下各項來進行的:在將允許藥劑與和傳訊生物化學途徑相關聯的蛋白質之間形成複合物的條件下，將該藥劑與和來源於測試樣品的傳訊生物化學途徑相關的蛋白質樣品接觸。複合物的形成可以是根據本領域的程序直接或間接檢測的。在直接檢測方法中，該等藥劑供應有可檢測標記並且未反應的藥劑可以從該複合物中去除；剩餘標記的量因此指示所形成的複合物的量。對於這種方法，較佳的是選擇甚至在嚴格洗滌條件過程中仍然附接至該等藥劑的標記。較佳的是，該標記並不干擾結合反應。在替代方案中，間接檢測程序可以使用含有以化學方式或酶方式引入的標記的藥劑。希望的標記通常並不干擾所得藥劑:多肽複合物的結合或穩定性。然而，該標記典型地被設計為易於接近抗體，以有效結合並因此生成可檢測信號。

[0877] 適用於檢測蛋白質水平的多種多樣的標記係本領域已知的。非限制性實例包括放射性同位素、酶類、膠態金屬、螢光化合物、生物發光化合物、以及化學發光化合物。

[0878] 在結合反應過程中形成的藥劑:多肽複合物的量可以是藉由標準定量測定來量化的。如上所述，藥劑:多肽複合物的形成可以是藉由保留在結合位點處的標記的量來直接測量的。在一個替代方案中，測試與傳訊生物化學途徑相關聯的蛋白質與標記的類似物競爭特異性藥劑上的結合位點的能力。在此競爭性測定中，捕獲的標記的量與和存在於測試樣品中的傳訊生物化學途徑相關聯的蛋白質序列的量成反比。

[0879] 用於基於以上列出的一般原則進行蛋白質分析的多種技術係本領域可用的。該等技術包括但不限於，放射免疫測定、ELISA（酶聯免疫放射測定法）、"夾層"免疫測定、免疫放射測定法、原位免疫測定（例如使用膠態金、酶或放射性同位素標記）、蛋白印跡分析、免疫沈澱測定、免疫螢光測定、以及SDS-PAGE。

[0880] 特異性識別或結合與傳訊生物化學途徑相關聯的蛋白質的抗體係較佳的是用於進行上述蛋白質分析。當希望時，可以使用識別特定類型的翻譯後修飾（例如傳訊生物化學途徑誘導的修飾）的抗體。翻譯後修飾包括但不限於，糖基化、脂化、乙醯化、以及磷酸化。該等抗體可以從商業供應商購買。例如，特異性識別酪胺酸磷酸化蛋白質的抗磷酸酪胺酸抗體可從許多供應商購買，該等供應商包括英傑公司（Invitrogen）和珀金埃爾默公司（Perkin Elmer）。抗磷酸酪胺酸抗體特別適用於檢測響應於ER脅

迫而在其酪胺酸殘基上有差異地磷醯化的蛋白質。此類蛋白質包括但不限於，真核生物翻譯起始因子2α（eIF-2α）。可替代地，該等抗體可以是使用多株抗體或單株抗體技術藉由用展示希望的翻譯後修飾的靶蛋白免疫宿主動物或抗體產物細胞來生成。

[0881] 在實踐主題方法中，可能希望的是辨別與傳訊生物化學途徑相關聯的蛋白質在不同身體組織、不同細胞類型和/或不同亞細胞結構中的表現模式。該等研究可以藉由使用能夠結合蛋白質標記物的組織特異性、細胞特異性或亞細胞結構特異性抗體來進行，該等蛋白質標記物優先在某些組織、細胞類型或亞細胞結構中表現。

[0882] 與傳訊生物化學途徑相關聯的基因的改變的表現也可以是藉由檢查基因產物活性相對於對照細胞的變化來確定。用於與傳訊生物化學途徑相關聯的蛋白質的藥劑誘導性活性變化的測定將依賴於研究中的生物活性和/或信號轉導途徑。例如，在該蛋白質係激酶的情況下，該蛋白質使一種或多種下游底物磷酸化的能力的變化可以是藉由本領域已知的多種測定來確定的。代表性測定包括但不限於，使用識別磷酸化蛋白的抗體諸如抗磷酸酪胺酸抗體的免疫印跡和免疫沈澱反應。此外，激酶活性可以是藉由高流通量化學發光測定諸如AlphaScreen™（可商購自珀金埃爾默公司）和eTag™測定（錢-慧（Chan-Hui）等人，(2003)臨床免疫學（Clinical Immunology）111: 162-174)來檢測的。

[0883] 在與傳訊生物化學途徑相關聯的蛋白質係導致細胞內pH條件波動的信號級聯放大的一部分的情況下，pH敏感分子諸

如螢光pH染料可以用作報導分子。在另一個實例中，在與傳訊生物化學途徑相關聯的蛋白質係離子通道的情況下，可以監測膜電位和/或細胞內離子濃度的波動。許多商用套組和高流通量裝置係特別適用於快速且強勁地篩選離子通道的調節劑。代表性儀器包括FLIPRTM(分子機器公司(Molecular Devices, Inc.))和VIPR(極光生物科學公司(Aurora Biosciences))。該等儀器能夠同時檢測微板的1000個樣品孔中的反應並且在一秒或甚至一微秒內提供即時測量值和功能資料。

[0884] 在實踐在此揭露的任何方法時，可以經由本領域已知的一種或多種方法來將適合的載體引入到細胞或胚胎中，該等方法包括但不限於，微注射、電穿孔、聲孔效應、基因槍、磷酸鈣介導的轉染、陽離子轉染、脂質體轉染、樹枝狀轉染、熱激轉染、核轉染、磁轉染、脂轉染、刺穿轉染、光學轉染、專有劑增強的核酸攝取、以及經由脂質體、免疫脂質體、病毒體或人工病毒體遞送。在一些方法中，載體藉由微注射引入到胚胎中。這種或該等載體可以微注射到胚胎的核或胞質中。在一些方法中，這種或該等載體可以藉由核轉染引入到細胞中。

[0885] CRISPR複合物的靶多核苷酸可以是對於真核細胞而言內源或外源的任何多核苷酸。例如，靶多核苷酸可以是駐留在真核細胞核內的多核苷酸。靶多核苷酸可以是編碼基因產物(例如，蛋白質)的序列或非編碼序列(例如，調節多核苷酸或無用DNA)。

[0886] 靶多核苷酸的實例包括與傳訊生物化學途徑相關聯

的序列，例如傳訊生物化學途徑相關基因或多核苷酸。靶多核苷酸的實例包括疾病相關基因或多核苷酸。“疾病相關”基因或多核苷酸係指與非疾病對照的組織或細胞相比，在來源於疾病影響的組織的細胞中產生異常水平或異常形式的轉錄產物或翻譯產物的任何基因或多核苷酸。它可以是以異常高的水平表現的基因；它可以是以異常低的水平表現的基因，其中改變的表現與疾病的發生和/或進展相關。疾病相關基因還是指具有一個或多個突變或直接負責或與一個或多個負責疾病的病因學的基因連鎖不平衡的遺傳變異的基因。轉錄或翻譯的產物可以是已知或未知的，並且可以是處於正常或異常水平。

[0887] CRISPR複合物的靶多核苷酸可以是對於真核細胞而言內源或外源的任何多核苷酸。例如，靶多核苷酸可以是駐留在真核細胞核內的多核苷酸。靶多核苷酸可以是編碼基因產物（例如，蛋白質）的序列或非編碼序列（例如，調節多核苷酸或無用DNA）。在不希望受到理論約束的情況下，認為靶序列應該與PAM（原型間隔區相鄰模體）相關聯；即，由CRISPR複合物識別的短序列。對於PAM的精確序列和長度要求根據所使用的CRISPR酶而不同，但PAM典型地是與原型間隔區（即，靶序列）相鄰的2-5個鹼基對序列。PAM序列的實例給出在以下實例部分中，並且技術人員將能夠鑒定與給定的CRISPR酶一起使用的其他PAM序列。此外，PAM相互作用（PI）結構域的工程化可以允許程式設計PAM特異性，提高靶位點識別保真度，並且增加Cas（例如Cas9）基因組工程化平臺的多功能性。Cas蛋白諸如Cas9蛋白可以被工程化以改變其PAM特異性，例如如克萊因史迪華 BP(Kleinstiver BP)等人，

具有改變的*PAM*特異性的工程化*CRISPR-Cas9*（*Engineered CRISPR-Cas9 nucleases with altered PAM specificities*），自然，2015年7月23日；523(7561):481-5. doi: 10.1038/nature14592。

[0888] CRISPR複合物的靶多核苷酸可以包括多個疾病相關基因和多核苷酸以及傳訊生物化學途徑相關基因和多核苷酸，如以下各項中列出的：美國臨時專利申請61/736,527和61/748,427，這兩份專利申請分別具有廣泛的參考文獻BI-2011/008/WSGR檔號44063-701.101和BI-2011/008/WSGR文件號44063-701.102，這兩份專利申請均題為用於序列操縱的系統方法和組成物（SYSTEMS METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION），它們分別在2012年12月12日和2013年1月2日提交，以及PCT申請PCT/US2013/074667，該申請題為用於序列操縱和治療應用的系統、方法和組成物的遞送、工程化和優化（DELIVERY, ENGINEERING AND OPTIMIZATION OF SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION AND THERAPEUTIC APPLICATIONS）且在2013年12月12日提交，所有該等專利申請的內容均藉由引用以其整體結合在此。

[0889] 靶多核苷酸的實例包括與傳訊生物化學途徑相關聯的序列，例如傳訊生物化學途徑相關基因或多核苷酸。靶多核苷酸的實例包括疾病相關基因或多核苷酸。“疾病相關”基因或多核苷酸係指與非疾病對照的組織或細胞相比，在來源於疾病影響的組織的細胞中產生異常水平或異常形式的轉錄產物或翻譯產物的

任何基因或多核苷酸。它可以是以異常高的水平表現的基因；它可以是以異常低的水平表現的基因，其中改變的表現與疾病的發生和/或進展相關。疾病相關基因還是指具有一個或多個突變或直接負責或與一個或多個負責疾病的病因學的基因連鎖不平衡的遺傳變異的基因。轉錄或翻譯的產物可以是已知或未知的，並且可以是處於正常或異常水平。

全基因組敲除篩選

[0890] 在此所述的CRISPR蛋白質和系統可以用於進行有效且性價比高的功能基因組篩選。此類篩選可以利用基於CRISPR效應蛋白的全基因組文庫。此類篩選和文庫可以提供對基因功能的確定，涉及細胞途徑基因，並且基因表現中的任何改變可以如何形成特定生物過程。本發明的優點係CRISPR系統避免了脫靶結合及其產生的副作用。這係使用安排為對靶DNA具有高度序列特異性的系統來實現的。在本發明的較佳的實施方式中，CRISPR效應蛋白複合物係Cpf1效應蛋白複合物。

[0891] 在本發明的實施方式中，全基因組文庫可以包含多個在此所述的Cpf1指導RNA，該等RNA包含能夠靶向真核細胞群體中的多個基因組座位中的多個靶序列的指導序列。細胞群體可以是胚胎幹細胞（ES）群體。基因組座位中的靶序列可以是非編碼序列。非編碼序列可以是內含子、調節序列、剪接位點、3' UTR、5' UTR、或多聚腺苷酸化信號。一種或多種基因產物的基因功能可以是藉由所述靶向來改變。該靶向可以導致基因功能敲除。基因產物的靶向可以包含超過一個的指導RNA。基因產物可以被每

個基因2、3、4、5、6、7、8、9或10個指導RNA，較佳的是3至4個指導RNA靶向。脫靶修飾可以是藉由採用由Cpf1效應蛋白複合物生成的交錯雙股斷裂或者藉由利用類似於CRISPR-Cas9系統中使用的方法來最小化（例如，參見，RNA-指導Cas9核酸酶的DNA靶向特異性（DNA targeting specificity of RNA-guided Cas9 nucleases）徐，P.（Hsu, P.,）、斯科特，D.（Scott, D.）、溫斯坦，J.（Weinstein, J.）、拉恩，FA.（Ran, FA.）、康爾曼，S.（Konermann, S.）、瓦拉，V.（Agarwala, V.）、李，Y.（Li, Y.）、法恩，E.（Fine, E.）、吳，X.（Wu, X.）、謝萊姆，O.（Shalem, O.）、科瑞迪克，TJ.（Cradick, TJ.）、瑪律拉菲尼，LA.（Marraffini, LA.）、包，G.（Bao, G.）、以及張，F.（Zhang, F.），自然生物技術（Nat Biotechnol）doi:10.1038/nbt.2647 (2013)），文獻藉由引用結合在此。該靶向可以是針對約100個或更多個序列。該靶向可以是針對約1000個或更多個序列。該靶向可以是針對約20,000個或更多個序列。該靶向可以是針對整個基因組。該靶向可以是針對集中於相關或希望的途徑中的一組靶序列。該途徑可以是免疫途徑。該途徑可以是細胞分裂途徑。

[0892] 本發明的一個方面包括全基因組文庫，該全基因組文庫可以包含多個Cpf1指導RNA，該等指導RNA可以包含能夠靶向多個基因組座位中的多個靶序列的指導序列，其中所述靶向導致基因功能敲除/敲低。此文庫可以潛在地包含靶向生物體基因組中的各個和每個基因的指導RNA。

[0893] 在本發明的一些實施方式中，生物體或受試者係真核

生物（包括哺乳動物，包括人類）或非人類真核生物或非人類動物或非人類哺乳動物。在一些實施方式中，生物體或受試者係非人類動物，並且可以是節肢動物，例如昆蟲，或者可以是線蟲。在本發明的一些方法中，生物體或受試者係植物。在本發明的一些方法中，生物體或受試者係哺乳動物或非人類哺乳動物。非人類哺乳動物可以是例如齧齒動物（較佳的是小鼠或大鼠）、有蹄動物或靈長類動物。在本發明的一些方法中，生物體或受試者係藻類，包括微藻類，或者是真菌類。

[0894] 基因功能的敲除/敲低可以包括：在細胞群體的每個細胞中引入包含工程化非天然存在的Cpf1效應蛋白系統的一種或多種載體的一個載體系統，該效應蛋白系統包含I. Cpf1效應蛋白和II.一個或多個指導RNA，其中組分I和組分II可以是處於該系統的相同或不同載體上；將組分I和II整合到每個細胞中，其中該指導序列靶向每個細胞中的獨特基因，其中該Cpf1效應蛋白可操作地連接至調節元件，其中在轉錄時，包含指導序列的該指導RNA引導Cpf1效應蛋白系統與對應於該獨特基因的基因組座位的靶序列的序列特異性結合；藉由Cpf1效應蛋白誘導基因組座位切割；並且確認細胞群體的每個細胞中的多個獨特基因中的不同敲除/敲低突變，從而生成基因敲除/敲低細胞文庫。本發明包括細胞群體係真核細胞群體，並且在較佳的實施方式中，細胞群體係胚胎幹（ES）細胞群體。

[0895] 該一種或多種載體可以是質粒載體。該載體可以是包含Cpf1效應蛋白、gRNA和視情況進入靶細胞中的選擇標記物的單

一載體。在不受理論約束的情況下，藉由單一載體同時遞送Cpf1效應蛋白和gRNA的能力使得能夠應用於任何感興趣的細胞類型，而不需要首先生成表現Cpf1效應蛋白的細胞系。調節元件可以是誘導型啟動子。誘導型啟動子可以是多西環素誘導型啟動子。在本發明的一些方法中，指導序列的表現係處於T7啟動子的控制下並且係由T7聚合酶的表現來驅動。不同敲除/敲低突變的確認可以是藉由全外顯子組定序進行的。敲除/敲低突變可以在100個或更多個獨特基因中實現。敲除/敲低突變可以在1000個或更多個獨特基因中實現。敲除/敲低突變可以在20,000個或更多個獨特基因中實現。敲除/敲低突變可以在整個基因組中實現。基因功能的敲除/敲低可以在多個獨特基因中實現，該等獨特基因在特定生理途徑或條件下起作用。該途徑或條件可以是免疫途徑或條件。該途徑或條件可以是細胞分裂途徑或條件。

[0896] 本發明還提供了包含在此提及的全基因組文庫的套組。該套組可以包含單個容器，該容器包含含有本發明的文庫的載體或質粒。該套組還可以包含含有獨特Cpf1效應蛋白系統指導RNA的選擇的面板，該指導RNA包含來自本發明的文庫的指導序列，其中該選擇指示特定生理條件。本發明包括的是，靶向係針對約100個或更多個序列、約1000或更多個序列或者約20,000個或更多個序列或整個基因組。另外，一組靶序列可以集中於相關或希望的途徑中，諸如免疫途徑或細胞分裂。

[0897] 在本發明的另一個方面中，Cpf1效應蛋白可以包含一個或多個突變並且可以在融合或未融合至功能結構域的情況下用

作通用DNA結合蛋白。該等突變可以是人工引入的突變或者增功能突變或失功能突變。該等突變已如在此所述地進行表徵。在本發明的一個方面中，功能結構域可以是轉錄啟動結構域，該結構域可以是VP64。在本發明的其他方面中，功能結構域可以是轉錄阻遏蛋白結構域，該結構域可以是KRAB或SID4X。本發明的其他方面涉及融合至結構域的突變的Cpf1效應蛋白，該等結構域包括但不限於，轉錄活化物、阻遏物、重組酶、易位酶、組蛋白改型物、去甲酶、DNA甲基轉移酶、隱花色素、光誘導型/控制型結構域或者化學誘導型/控制型結構域。本發明的一些方法可以包括誘導靶基因的表現。在一個實施方式中，藉由靶向真核細胞群體中的多個基因組座位中的多個靶序列誘導表現係藉由使用功能結構域進行的。

[0898] 用於CRISPR-Cas9系統中的方法適用於利用Cpf1效應蛋白複合物實踐本發明並且參考以下各項：

[0899] 人類細胞中的基因組規模的CRISPR-Cas9敲除篩選（Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells）。沙萊姆，O.（Shalem, O.）、珊亞納，NE.（Sanjana, NE.）、哈特諾斯，E.（Hartenian, E.）、石，X.（Shi, X.）、斯科特，DA.（Scott, DA.）、邁克爾森，T.（Mikkelson, T.）、赫克爾，D.（Heckl, D.）、埃伯特，BL.（Ebert, BL.）、羅特，DE.（Root, DE.）、多恩奇，JG.（Doench, JG.）、張，F.（Zhang, F.），科學，12月12日(2013)。[電子版先於印刷版]；以最終編輯形式出版為：科學，2014年1月3日；343(6166): 84–87。

[0900] 沙萊姆等人涉及一種探察全基因組規模的基因功能的新方式。他們的研究顯示具有64,751個獨特指導序列的基因組規模的CRISPR-Cas9敲除（GeCKO）文庫靶向的18,080個基因的遞送能夠在人類細胞中進行陰性和陽性選擇篩選。首先，作者們證實使用GeCKO文庫鑒定了癌症和多能幹細胞中的細胞活力所必須的基因。接著，在黑色素瘤模型中，作者們篩選其喪失涉及對維羅非尼（抑制蛋白激酶BRAF的治療劑）的抗性的基因。他們的研究顯示最高評級的候選物包括先前驗證的基因NF1和MED12以及新型命中基因NF2、CUL3、TADA2B、以及TADA1。作者們在靶向相同基因的獨立的指導RNA與高命中確認率之間觀察到高水平的一致性，並且因此證實允許用Cas9進行基因組規模篩選。

[0901] 還參考美國專利公開案號US20140357530；以及PCT專利公開WO2014093701，該等專利藉由引用結合在此。還參考2015年10月22日題為"研究者鑒定了CRISPR-Cas基因組編輯工具的可潛在替代方案：新Cas酶闡明CRISPR-Cas系統的進化（Researchers identify potential alternative to CRISPR-Cas genome editing tools: New Cas enzymes shed light on evolution of CRISPR-Cas systems）"的NIH通訊稿，該通訊稿藉由引用結合在此。

功能變化和篩選

[0902] 在另一個方面中，本發明提供了一種功能性評定和篩選基因的方法。使用本發明的CRISPR系統精確遞送功能結構域以藉由精確改變特定感興趣的座位上的甲基化位點來啟動或阻遏基因或者改變外遺傳狀態，這可以與一個或多個指導RNA一起應用

於單個細胞或細胞群體或者與文庫一起應用於離體或體內細胞庫中的基因組，包括給予或表現包含多個指導RNA（gRNA）的文庫並且其中該篩選進一步包括使用Cpf1效應蛋白，其中包含該Cpf1效應蛋白的CRISPR複合物被修飾為包含異源功能結構域。在一個方面中，本發明提供了一種用於篩選基因組之方法，該方法包括向宿主給予文庫或者在宿主體內表現文庫。在一方面中，本發明提供了如在此討論之方法，該方法進一步包括向宿主給予活化物或在宿主中表現活化物。在一方面中，本發明提供了如在此所討論之方法，其中該活化物附接至Cpf1效應蛋白。在一方面中，本發明提供了如在此所討論之方法，其中該活化物附接至Cpf1效應蛋白的N末端或C末端。在一方面中，本發明提供了如在此所討論之方法，其中該活化物附接至gRNA環。在一個方面中，本發明提供了如在此討論之方法，該方法進一步包括向宿主給予阻遏劑或在宿主中表現阻遏劑。在一方面中，本發明提供了如在此所討論之方法，其中該篩選包括影響並檢測基因啟動、基因抑制或座位中的切割。

[0903] 在一方面中，本發明提供了有效中靶活性並且最小化脫靶活性。在一方面中，本發明提供了由Cpf1效應蛋白進行的有效中靶切割並且最小化由Cpf1效應蛋白進行的脫靶切割。在一方面中，本發明提供了在無DNA切割情況下Cpf1效應蛋白在基因座位處的特異性結合。因此，在一方面中，本發明提供了靶特異性基因調節。在一方面中，本發明提供了在無DNA切割情況下Cpf1效應蛋白在基因座位處的特異性結合。因此，在一方面中，本發明使用單一Cpf1效應蛋白提供在一個基因座位處的切割和在一個

不同的基因座位處的基因調節。在一方面中，本發明使用一種或多種Cpf1效應蛋白和/或酶提供多個靶標的正交啟動和/或抑制和/或切割。

[0904] 在一方面中，本發明提供了如在此所討論之方法，其中該宿主係真核細胞。在一個方面中，本發明提供了如在此所討論之方法，其中該宿主係哺乳動物細胞。在一個方面中，本發明提供了如在此所討論之方法，其中該宿主係非人類真核生物。在一個方面中，本發明提供了如在此所討論之方法，其中該非人類真核生物係非人類哺乳動物。在一方面中，本發明提供了如在此所討論之方法，其中該非人類哺乳動物係小鼠。在一方面中，本發明提供了如在此所討論之方法，該方法包括遞送Cpf1效應蛋白複合物或其一種或多種組分或對其編碼的一個或多個核酸分子，其中所述一個或多個核酸分子可操作地連接至一個或多個調節序列並且在體內表現。在一方面中，本發明提供了如在此所討論之方法，其中該體內表現係經由慢病毒、腺病毒或AAV。在一方面中，本發明提供了如在此所討論之方法，其中該遞送係經由粒子、奈米粒子、脂質或細胞穿透肽（CPP）。

[0905] 在一方面中，本發明提供了一對包含Cpf1效應蛋白的CRISPR複合物，每個複合物包含含有能夠與細胞中感興趣的基因組座位中的靶序列雜交的指導序列的指導RNA（gRNA），其中每個gRNA的至少一個環係藉由插入結合一種或多種轉接蛋白的一種或多種不同RNA序列來修飾的，並且其中該轉接蛋白與一個或多個功能結構域締合，其中每個Cpf1效應蛋白複合物的每個gRNA

包含具有DNA切割活性的功能結構域。在一方面中，本發明提供了如在此所討論的成對的Cpf1效應蛋白複合物，其中DNA切割活性係歸因於Fok1核酸酶。

[0906] 在一方面中，本發明提供了一種用於切割感興趣的基因組座位中的靶序列之方法，該方法包括向細胞遞送Cpf1效應蛋白複合物或其一種或多種組分或對其編碼的一個或多個核酸分子，其中所述一個或多個核酸分子可操作地連接至一個或多個調節序列並且在體內表現。在一方面中，本發明提供了如在此所討論之方法，其中該遞送係經由慢病毒、腺病毒或AAV。在一方面中，本發明提供了如在此所討論之方法或如在此所討論的成對的Cpf1效應蛋白複合物，其中該對中的第一複合物的靶序列係處於雙股DNA的第一條股上並且該對中的第二複合物的靶序列係處於雙股DNA的第二條股上。在一方面中，本發明提供了如在此所討論之方法或如在此所討論的成對的Cpf1效應蛋白複合物，其中該第一複合物和該第二複合物的靶序列彼此接近，使得DNA以促進同源定向修復的方式切割。在一方面中，在此的方法可以進一步包括將模板DNA引入到細胞中。在一個方面中，一種在此的方法或者在此的成對的Cpf1效應蛋白複合物可以涉及其中每個Cpf1效應蛋白複合物具有Cpf1效應酶，該Cpf1效應酶被突變為使得它具有不超過未突變Cpf1效應酶的核酸酶活性的約5%的核酸酶活性。

[0907] 在一方面中，本發明提供了如在此所討論的文庫、方法或複合物，其中gRNA被修飾為具有至少一個非編碼功能環，例如其中該至少一個非編碼功能環係有阻遏作用的；例如，其中該

至少一個非編碼功能環包含Alu。

[0908] 在一個方面中，本發明提供了用於改變或修飾基因產物的表現之方法。所述方法可以包括將工程化的非天然存在的CRISPR系統引入到含有並表現編碼基因產物的DNA分子的細胞中，該CRISPR系統包含Cpf1效應蛋白和靶向DNA分子的指導RNA，由此該指導RNA靶向編碼基因產物的DNA分子並且該Cpf1效應蛋白切割編碼該基因產物的DNA分子，由此改變該基因產物的表現；並且其中Cpf1效應蛋白和指導RNA並不一起天然存在。本發明包括含有連接至同向重複序列的指導序列的指導RNA。本發明進一步包括密碼子優化為在真核細胞中表現的Cpf1效應蛋白。在一較佳的實施方式中，真核細胞係哺乳動物細胞，並且在一更較佳的實施方式中，哺乳動物細胞係人類細胞。在本發明的另一個實施方式中，基因產物的表現減少。

[0909] 在一些實施方式中，一個或多個功能結構域與Cpf1效應蛋白締合。在一些實施方式中，一個或多個功能結構域與轉接蛋白，例如如與康爾曼等人（自然 517，583–588，2015年1月29日）的修飾指導序列一起使用。在一些實施方式中，一個或多個功能結構域與無效gRNA（dRNA）締合。在一些實施方式中，與活性Cpf1效應蛋白的dRNA複合物藉由基因座位上的功能結構域來引導基因調節，而gRNA藉由另一個座位處的活性Cpf1效應蛋白來引導DNA切割，例如由達爾曼（Dahlman）等人“使用催化活性的Cas9核酸酶的正交基因控制（Orthogonal gene control with a catalytically active Cas9 nuclease）”（出版中）在CRISPR-Cas9系

統中類似描述的。在一些實施方式中，dRNA被選擇為與脫靶調節相比使對感興趣的基因座位的調節選擇性最大化。在一些實施方式中，dRNA被選擇為最大化靶基因調節並且最小化靶標切割。

[0910] 出於以下討論的目的，提及功能結構域可以是與Cpf1效應蛋白締合的功能結構域或者與轉接蛋白締合的功能結構域。

[0911] 在本發明的實踐中，可以在不與Cpf1蛋白碰撞的情況下藉由插入可以募集轉接蛋白的一個或多個不同RNA環或一個或多個不同序列擴展gRNA的環，該等轉接蛋白可以結合一個或多個不同RNA環或者一個或多個不同序列。該等轉接蛋白可以包括但不限於，存在於各種噬菌體外殼蛋白內的正交RNA結合蛋白/適配體組合。此類外殼蛋白列表包括但不限於：Qβ、F2、GA、fr、JP501、M12、R17、BZ13、JP34、JP500、KU1、M11、MX1、TW18、VK、SP、FI、ID2、NL95、TW19、AP205、ϕCb5、ϕCb8r、ϕCb12r、ϕCb23r、7s以及PRR1。該等轉接蛋白或正交RNA結合蛋白可以進一步募集包含一個或多個功能結構域的效應蛋白或融合物。在一些實施方式中，功能結構域可以是選自下組，該組由以下各項組成：易位酶結構域、整合酶結構域、重組酶結構域、解離酶結構域、轉化酶結構域、蛋白酶結構域、DNA甲基轉移酶結構域、DNA羥甲基酶結構域、DNA脫甲基酶結構域、組蛋白乙醯化酶結構域、組蛋白脫乙醯化酶結構域、核酸酶域、阻遏物結構域、活化物結構域、核定位信號結構域、轉錄調節蛋白質（或轉錄複合物募集）結構域、細胞攝取活動相關結構域、核酸結合結構域、抗體呈遞結構域、組蛋白修飾酶、組蛋白修飾酶的募集物；組蛋白修飾酶、組

蛋白甲基轉移酶、組蛋白去甲酶、組蛋白激酶、組蛋白磷酸酶、組蛋白核糖基酶、組蛋白脫核糖基酶、組蛋白泛素酶、組蛋白脫泛素酶、組蛋白生物素酶以及組蛋白尾蛋白酶的抑制劑。在一些較佳的實施方式中，功能結構域係轉錄啟動結構域，諸如但不限於，VP64、p65、MyoD1、HSF1、RTA、SET7/9或組蛋白乙醯轉移酶。在一些實施方式中，功能結構域係轉錄阻遏結構域，較佳的是KRAB。在一些實施方式中，轉錄阻遏結構域係SID或SID的串聯體（例如SID4X）。在一些實施方式中，功能結構域係表觀遺傳修飾結構域，以便提供表觀遺傳修飾酶。在一些實施方式中，功能結構域係啟動結構域，它可以是P65啟動結構域。

[0912] 在一些實施方式中，一個或多個功能結構域係NLS（核定位序列）或NES（核輸出信號）。在一些實施方式中，一個或多個功能結構域係轉錄啟動結構域，包括VP64、p65、MyoD1、HSF1、RTA、SET7/9以及組蛋白乙醯轉移酶。在此提及的其他啟動（或活化物）結構域（關於與CRISPR酶締合的那些結構域）包括任何已知的轉錄啟動結構域並且確切的是VP64、p65、MyoD1、HSF1、RTA、SET7/9或組蛋白乙醯轉移酶。

[0913] 在一些實施方式中，一個或多個功能結構域係轉錄阻遏蛋白結構域。在一些實施方式中，轉錄阻遏蛋白結構域係KRAB結構域。在一些實施方式中，轉錄阻遏蛋白結構域係NuE結構域、NcoR結構域、SID結構域或SID4X結構域。

[0914] 在一些實施方式中，一個或多個功能結構域具有一種或多種活性，包括修飾烷酶活性、去甲酶活性、轉錄啟動活性、

轉錄阻遏活性、轉錄釋放因子活性、組蛋白修飾活性、RNA切割活性、DNA切割活性、DNA整合活性或核酸結合活性。

[0915] 在一些實施方式中，組蛋白修飾結構域也是較佳的。以下討論了示例性組蛋白修飾結構域。易位酶結構域、HR（同源重組）機構結構域、重組酶結構域、和/或整合酶結構域作為本發明功能結構域也是較佳的。在一些實施方式中，DNA整合活性包括HR機構結構域、整合酶結構域、重組酶結構域和/或易位酶結構域。在一些實施方式中，組蛋白乙醯轉移酶係較佳的。

[0916] 在一些實施方式中，DNA切割活性係歸因於核酸酶。在一些實施方式中，核酸酶包括Fok1核酸酶。參見，“用於高特異性基因組編輯的二聚CRISPR RNA指導FokI核酸酶（Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing）”，盛達爾·Q.·蔡（Shengdar Q. Tsai）、尼古拉斯·維肯恩（Nicolas Wyvekens）、采德·凱特爾（Cyd Khayter）、詹尼弗·A.·福登布（Jennifer A. Foden）、維沙爾·撒帕爾（Vishal Thapar）、迪派克·雷恩（Deepak Reyon）、馬修·J.·古德溫（Mathew J. Goodwin）、馬丁·J.·阿裡耶（Martin J. Aryee）、J.·基斯·薑俊（J. Keith Joung），自然生物技術32(6): 569-77 (2014)，它涉及在人類細胞中識別擴展序列並以高效率編輯內源性基因的二聚RNA指導FokI核酸酶。

[0917] 在一些實施方式中，一個或多個功能結構域附接至Cpf1效應蛋白，以使得在結合sgRNA和靶標時，功能結構域係呈允許功能結構域以其屬性功能起作用的空間取向。

[0918] 在一些實施方式中，一個或多個功能結構域附接至轉

接蛋白，以使得在Cpf1效應蛋白結合gRNA和靶標時，功能結構域係處於允許功能結構域以其屬性功能起作用的空間定向中。

[0919] 在一方面中，本發明提供了如在此所討論的組成物，其中一個或多個功能結構域經由如在此所討論的接頭、視情況GlySer接頭附接至Cpf1效應蛋白或轉接蛋白。

[0920] 內源性轉錄阻遏常常是藉由染色質修飾酶諸如組蛋白甲基轉移酶（HMT）和脫乙醯化酶（HDAC）介導的。阻遏組蛋白效應物結構域係已知的並且下文提供了一示例性列表。在該示例性表格中，優先使用有助於有效病毒包裝（例如經由AAV）的小型蛋白質和功能截短物。然而，總的來說，該等結構域可以包括HDAC、組蛋白甲基轉移酶（HMT）和組蛋白乙醯轉移酶（HAT）抑制劑、以及HDAC和HMT募集蛋白。在一些實施方式中，功能結構域可以是或者包括HDAC效應物結構域、HDAC募集物效應物結構域、組蛋白甲基轉移酶（HMT）效應物結構域、組蛋白甲基轉移酶（HMT）募集物效應物結構域、或組蛋白乙醯轉移酶抑制劑效應物結構域。

HDAC 效應物結構域

亞型/複合物	**名稱**	底物（如果已知的話）	修飾（如果已知的話）	生物體	全尺寸（aa）	選定的截短物（aa）	最終尺寸（aa）	催化結構域
HDAC I	**HDAC8**	-	-	*非洲爪蟾*	325	1-325	325	1-272: HDAC
HDAC I	**RPD3**	-	-	釀酒酵母	433	19-340	322（維尼爾（Vannier））	19-331: HDAC

HDAC IV	**Meso Lo4**	-	-	*百脈根根瘤菌*	300	1-300(格雷戈雷蒂(Gregoretti))	300	-
HDAC IV	**HDAC11**	-	-	*智人*	347	1-347(高(Gao))	347	14-326: HDAC
HD2	**HDT1**	-	-	*擬南芥*	245	1-211(吳(Wu))	211	-
SIRT I	**SIRT3**	H3K9Ac H4K16Ac H3K56Ac	-	*智人*	399	143-399(謝爾(Scher))	257	126-382: SIRT
SIRT I	**HST2**	-	-	*白色念珠菌*	331	1-331(尼斯茲(Hnisz))	331	-
SIRT I	**CobB**	-	-	*大腸桿菌(K12)*	242	1-242(蘭德裡(Landry))	242	-
SIRT I	**HST2**	-	-	*釀酒酵母*	357	8-298(威爾遜(Wilson))	291	-
SIRT III	**SIRT5**	H4K8Ac H4K16Ac	-	*智人*	310	37-310(格茨(Gertz))	274	41-309: SIRT
SIRT III	**Sir2A**	-	-	*惡性瘧原蟲*	273	1-273(朱(Zhu))	273	19-273: SIRT
SIRT IV	**SIRT6**	H3K9Ac H3K56Ac	-	*智人*	355	1-289(特內恩(Tennen))	289	35-274: SIRT

[0921] 因此，本發明的阻遏蛋白結構域可以是選自組蛋白甲基轉移酶（HMT）、組蛋白脫乙醯酶（HDAC）、組蛋白乙醯轉移酶（HAT）抑制劑、以及HDAC和HMT募集蛋白。

[0922] HDAC結構域可以是以上表中的那些結構域中的任一種，即：HDAC8、RPD3、MesoLo4、HDAC11、HDT1、SIRT3、HST2、CobB、HST2、SIRT5、Sir2A、或SIRT6。

[0923] 在一些實施方式中，功能結構域可以是HDAC募集物效應物結構域。較佳的實例包括以下表中的那些，即MeCP2、MBD2b、Sin3a、NcoR、SALL1、RCOR1。NcoR係本發明實例中舉例說明的，並且儘管它係較佳的，但設想的是該類別中的其他結構域也將是有用的。

HDAC募集物效應物結構域的表

亞型/複合物	**名稱**	底物（如果已知的話）	修飾（如果已知的話）	生物體	全尺寸（aa）	選定的截短物（aa）	最終尺寸（aa）	催化結構域
Sin3a	**MeCP2**	-	-	*褐家鼠*	492	207-492（南（Nan））	286	-
Sin3a	**MBD2b**	-	-	*智人*	262	45-262（伯克（Boeke））	218	-
Sin3a	**Sin3a**	-	-	*智人*	1273	524-851（拉赫蒂（Lahertty））	328	627-829: HDAC1相互作用
NcoR	**NcoR**	-	-	*智人*	2440	420-488（張（Zhang））	69	-
NuRD	**SALL1**	-	-	*小家鼠*	1322	1-93（勞伯思（Lauberth））	93	-
CoREST	**RCOR1**	-	-	*智人*	482	81-300（古，歐	220	-

						陽（Gu, Ouyang））		

[0924] 在一些實施方式中，功能結構域可以是甲基轉移酶（HMT）效應物結構域。較佳的實例包括下表中的那些，即NUE、vSET、EHMT2/G9A、SUV39H1、dim-5、KYP、SUVR4、SET4、SET1、SETD8、以及TgSET8。NUE係本發明實例中舉例說明的，並且儘管它係較佳的，但設想的是該類別中的其他結構域也將是有用的。

組蛋白甲基轉移酶（HMT）效應物結構域的表

亞型/複合物	名稱	底物（如果已知的話）	修飾（如果已知的話）	生物體	全尺寸（aa）	選定的截短物（aa）	最終尺寸（aa）	催化結構域
SET	**NUE**	H2B、H3、H4	-	*沙眼衣原體*	219	1-219（彭尼尼（Pennini））	219	-
SET	**vSET**	-	H3K27me3	*綠草履蟲小球藻病毒*	119	1-119（穆傑塔巴（Mujtaba））	119	4-112: SET2
SUV39家族	**EHMT2/G9A**	H1.4K2、H3K9、H3K27	H3K9me1/2、H1K25me1	*小家鼠*	1263	969-1263（泰奇巴納（Tachibana））	295	1025-1233: preSET、SET、postSET
SUV39	**SUV39H1**	-	H3K9me2/3	*智人*	412	79-412（斯諾登（Snowden））	334	172-412: preSET、SET、postSET
Suvar3-9	**dim-5**	-	H3K9me3	*粗糙脈孢菌*	331	1-331（拉思	331	77-331: preSET、

						厄特（Rathert））		SET、postSET
Suvar 3-9（SUVH子族）	**KYP**	-	H3K9me1/2	*擬南芥*	624	335-601	267（傑克遜（Jackson））	-
Suvar 3-9（SUVR子族）	**SUVR4**	H3K9me1	H3K9me2/3	*擬南芥*	492	180-492	313（托斯滕森（Thorstensen））	192-462: preSET、SET、postSET
Suvar 4-20	**SET4**	-	H4K20me3	*秀麗隱桿線蟲*	288	1-288（維耶勒（Vielle））	288	-
SET8	**SET1**	-	H4K20me1	*秀麗隱桿線蟲*	242	1-242（維耶勒（Vielle））	242	-
SET8	**SETD8**	-	H4K20me1	*智人*	393	185-393	209（科圖雷（Couture））	256-382: SET
SET8	**TgSET8**	-	H4K20me1/2/3	*剛地弓形蟲*	1893	1590-1893（索泰（Saute1））	304	1749-1884: SET

[0925] 在一些實施方式中，功能結構域可以是組蛋白甲基轉

移酶（HMT）募集物效應物結構域。較佳的實例包括以下表中的那些，即Hp1a、PHF19、以及NIPP1。

[0926] 組蛋白甲基轉移酶(HMT)募集物效應物結構域的表

亞型/複合物	**名稱**	底物（如果已知的話）	修飾（如果已知的話）	生物體	全尺寸（aa）	選定的截短物（aa）	最終尺寸（aa）	催化結構域
-	**Hp1a**	-	H3K9me3	*小家鼠*	191	73-191	119(哈撒韋（Hathaway))	121-179:染色陰影區（chromoshadow）
-	**PHF19**	-	H3K27me3	*智人*	580	(1-250)+GGSG接頭+(500-580)	335(巴羅瑞（Ballaré))	163-250:PHD2
-	**NIPP1**	-	H3K27me3	*智人*	351	1-329（金（Jin))	329	310-329:EED

[0927] 在一些實施方式中，功能結構域可以是組蛋白乙醯轉移酶抑制劑效應物結構域。較佳的實例包括以下表中列出的SET/TAF-1β。

[0928] 組蛋白乙醯轉移酶抑制劑效應物結構域的表

亞型/複合物	**名稱**	底物（如果已知的話）	修飾（如果已知的話）	生物體	全尺寸（aa）	選定的截短物（aa）	最終尺寸（aa）	催化結構域
-	**SET/TAF-1β**	-	-	*小家鼠*	289	1-289（切爾沃尼（Cervoni))	289	-

[0929] 還較佳的是靶向除啟動子或啟動子近側元件之外的內源性（調節）控制元件（諸如增強子和沈默子）。因此，除靶向啟動子之外，本發明還可以用於靶向內源性控制元件（包括增強子和沈默子）。該等控制元件可以位於轉錄起始位點（TSS）上游和下游，從距離TSS的200 bp開始到100 kb遠。已知控制元件的靶向可以用於啟動或阻遏感興趣的基因。在一些情況下，單一控制元件可以影響多個靶基因的轉錄。單一控制元件的靶向可以因此用於同時控制多基因的轉錄。

[0930] 在另一個方面中，推定的控制元件的靶向（例如，藉由針對推定的控制元件區域以及該元件周圍200 bp至100 kB）可以用作驗證此類元件（藉由測量感興趣的基因的轉錄）或者檢測新型控制元件（例如，藉由針對感興趣的基因的TSS的上游和下游的100 kb）的手段。此外，推定的控制元件的靶向可以適用於理解疾病遺傳原因的情況。與疾病表型相關聯的許多突變和常見SNP變體位於編碼區之外。使用在此所述的啟動或阻遏系統靶向此類區域，可以接著讀出 a) 一組推定的靶標（例如，最緊密接近控制元件來定位的一組基因）的轉錄或者 b) 藉由RNAseq或微陣列進行整體轉錄組讀出。這允許鑒定疾病表型中涉及的最可能的候選基因。此類候選基因可以適用作新型藥物靶標。

[0931] 在此提及了組蛋白乙醯轉移酶（HAT）抑制劑。然而，在一些實施方式中，替代物係針對包含乙醯轉移酶、較佳的是組蛋白乙醯轉移酶的一個或多個功能結構域。該等適用於表觀基因組學領域，例如適用於探察表觀基因組的方法。探察表觀基因組

的方法可以包括例如靶向表觀基因組序列。靶向表觀基因組序列可以包括將指導序列引導至表觀基因組靶序列。在一些實施方式中，表觀基因組靶序列可以包括啟動子、沈默子或增強子序列。

[0932] 使用連接至在此所述的Cpf1效應蛋白、較佳的是無效Cpf1效應蛋白、更較佳的是無效-FnCpf1效應蛋白的功能結構域靶向表觀基因組序列，可以用於啟動或阻遏啟動子、沈默子或增強子。

[0933] 乙醯轉移酶的實例係已知的，但在一些實施方式中可以包括組蛋白乙醯轉移酶。在一些實施方式中，組蛋白乙醯轉移酶可以包含人類乙醯轉移酶p300的催化核心（傑爾巴遲（Gerbasch）和雷迪（Reddy），自然生物技術，2015年4月6日）。

[0934] 在一些較佳的實施方式中，功能結構域連接至無效-Cpf1效應蛋白，以靶向並啟動表觀基因組序列，諸如啟動子或增強子。還可以提供引導至此類啟動子或增強子的一種或多種指導序列，以引導CRISPR酶與此類啟動子或增強子結合。

[0935] 術語“與...締合”在此用於係指功能結構域與Cpf1效應蛋白或轉接蛋白締合。它係關於一個分子如何與另一個分子“締合”，例如轉接蛋白與功能結構域之間或者Cpf1效應蛋白與功能結構域之間。在此類蛋白質-蛋白質相互作用的情況下，此締合可以按照抗體識別表位的方式進行的識別來觀察。可替代地，一種蛋白質可以與另一種蛋白質經由兩者的融合來締合，例如一種亞基融合至另一種亞基。典型地藉由將一種蛋白質的胺基酸序列添加到另一種蛋白質的胺基酸序列上，例如經由將編碼每種蛋白質或

亞基的核苷酸序列剪接在一起來發生融合。可替代地，這可以實質上視為兩個分子之間的結合或直接連接，諸如融合蛋白。在任何情況下，融合蛋白可以包含兩個感興趣的亞基之間（即酶與功能結構域之間或轉接蛋白與功能結構域之間）的接頭。因此，在一些實施方式中，Cpf1效應蛋白或轉接蛋白藉由結合功能結構域來與該功能結構域締合。在其他實施方式中，Cpf1效應蛋白或轉接蛋白與功能結構域締合，因為兩者視情況經由中間接頭融合在一起。

[0936] 功能結構域或融合蛋白的附接可以是經由接頭，例如柔性甘胺酸-絲胺酸(GlyGlyGlySer)或$(GGGS)_3$或者硬性α-螺旋形接頭諸如(Ala(GluAlaAlaAlaLys)Ala)。在此較佳的是使用接頭諸如(GGGGS)3分開蛋白質或肽結構域。$(GGGGS)_3$係較佳的，因為它係相對長的接頭（15個胺基酸）。甘胺酸殘基係最柔性的並且絲胺酸殘基增加接頭處於蛋白質之外的機會。$(GGGGS)_6$、$(GGGGS)_9$或$(GGGGS)_{12}$可以較佳的是用作替代物。其他較佳的替代物係$(GGGGS)_1$、$(GGGGS)_2$、$(GGGGS)_4$、$(GGGGS)_5$、$(GGGGS)_7$、$(GGGGS)_8$、$(GGGGS)_{10}$、或$(GGGGS)_{11}$。替代性接頭係可用的，但高柔性接頭被認為是作用最好，以使得Cpf1的2個部分合在一起並因此重構Cpf1活性的機會最大。一替代方案係核質蛋白的NLS可以用作接頭。例如，接頭也可以用在Cpf1與任何功能結構域之間。同樣，在此可以使用$(GGGGS)_3$接頭（或因此6、9或12個重複版本）或者可以將核質蛋白的NLS用作Cpf1與功能結構域之間的接頭。

飽和誘變

[0937] 在此所述的一種或多種Cpf1效應蛋白系統可以用於進行基因組座位連同細胞表型的飽和誘變或深度掃描誘變—例如以用於測定基因表現、藥物抗性和疾病逆轉所需要的功能元件的關鍵性最小特徵和不連續易損性。飽和誘變或深度掃描誘導意指在基因組座位中每個或基本上每個DNA鹼基被切割。Cpf1效應蛋白指導RNA的文庫可以被引入到細胞群體中。該文庫可以被引入為使得每個細胞接收單一指導RNA（gRNA）。在其中該文庫藉由轉導如在此所述的病毒載體來引入的情況下，使用低感染複數（MOI）。該文庫可以包括靶向在基因組座位中的（原型間隔區相鄰模體）（PAM）序列上游的每個序列的gRNA。該文庫對於基因組座位中的每1000個鹼基對可以包括PAM序列上游的至少100個非重疊基因組座位序列。該文庫可以包括靶向至少一個不同PAM序列上游的序列的gRNA。Cpf1效應蛋白系統可以包含超過一個Cpf1蛋白。可以使用在此所述的任何Cpf1效應蛋白，包括識別不同PAM序列的異種同源物或工程化Cpf1效應蛋白。gRNA的脫靶位點的頻率可以是小於500。可以生成脫靶得分以選擇具有最低脫靶位點的gRNA。在單個實驗中，確定與gRNA靶位點處的切割相關聯的任何表型可以是藉由使用靶向相同位點的gRNA來證實。靶位點的確認也可以是藉由使用如在此所述的修飾的Cpf1效應蛋白和靶向感興趣的基因組位點的兩個gRNA來進行的。在不希望受到理論約束的情況下，如果在確認實驗中觀察到表型的變化，則靶位點係準確命中的。

[0938] 基因組座位可以包含至少一個連續的基因組區域。該至少一個連續基因組區域可以包含多至整個基因組。該至少一個

連續基因組區域可以包含該基因組的功能元件。該功能元件可以是處於非編碼區、編碼基因、內含子區域、啟動子或增強子。該至少一個連續的基因組區域可以包含至少1 kb、較佳的是至少50 kb的基因組DNA。該至少一個連續基因組區域可以包含轉錄因子結合位點。該至少一個連續基因組區域可以包含DNA酶I超敏區域。該至少一個連續基因組區域可以包含轉錄增強子或阻遏子元件。該至少一個連續基因組區域可以包含富含表觀遺傳特徵的位點。該至少一個連續基因組DNA區域可以包含表觀遺傳絕緣子。該至少一個連續基因組區域可以包含物理上相互作用的兩個或更多個基因組區域。相互作用的基因組區域可以是藉由“4C技術”來確定。4C技術允許對於與選擇的DNA片段物理相互作用的DNA區段以無偏性方式篩選整個基因組，如趙（Zhao）等人（(2006)自然遺傳學（Nat Genet）38，1341-7）和美國專利8,642,295中所述的，這兩份文獻藉由引用以其整體結合在此。表觀遺傳特徵可以是組蛋白乙醯化、組蛋白甲基化、組蛋白泛素化、組蛋白磷酸化、DNA甲基化或其缺失。

[0939] 用於飽和誘變或深度掃描誘變的一種或多種Cpf1效應蛋白系統可以用於細胞群體中。一種或多種Cpf1效應蛋白系統可以用於真核細胞中，該等真核細胞包括但不限於哺乳動物細胞和植物細胞。細胞群體可以是真核細胞。真核細胞群體可以是胚胎幹細胞（ES）、神經元細胞、上皮細胞、免疫細胞、內分泌細胞、肌肉細胞、紅細胞、淋巴細胞、植物細胞或酵母細胞。

[0940] 在一個方面中，本發明提供了篩選與表型變化相關聯

的功能元件之方法。文庫可以被引入到適於含有Cpf1效應蛋白的細胞群體。該等細胞基於表型可以被分成至少兩組。該表型可以是基因表現、細胞生長或細胞活力。確定存在於每組中的指導RNA的相對表現度，由此藉由每組中的指導RNA的表現度來確定與表型變化相關聯的基因組位點。表型變化可以是感興趣的基因表現的變化。感興趣的基因可以是上調、下調或敲除的。該等細胞可以被分為高表現組和低表現組。細胞群體可以包括用於確定表型的報導基因構建體。該報導基因構建體可以包含可檢測標記物。細胞可以藉由使用可檢測標記物來分類。

[0941] 在另一個方面中，本發明提供了篩選與對化學化合物的抗性相關聯的基因組位點之方法。化學化合物可以是藥物或殺蟲劑。文庫可以被引入到適於含有Cpf1效應蛋白的細胞群體中，其中該群體中的每個細胞含有不超過一個的指導RNA；用化學化合物處理細胞群體；並且與早時間點相比，在較晚時間點用化學化合物處理後確定指導RNA的表現度，由此藉由富集指導RNA來確定與對化學化合物的抗性相關聯的基因組位點。gRNA的表現度可以是藉由深度定序方法確定的。

[0942] 用於CRISPR-Cas9系統中的方法適用於利用Cpf1效應蛋白複合物實踐本發明並且參考以下文章：題為藉由Cas9-誘導的原位飽和誘變進行BCL11A增強子分割（BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis）康維爾M.C.（Canver, M.C.）、史密斯E.C.（Smith,E.C.）、謝爾F.（Sher, F.）、派因洛L.（Pinello, L.）、珊亞納N.E.（Sanjana, N.E.）、沙萊姆O.

（Shalem, O.）、陳D.D.（Chen, D.D.）、舒普P.G.（Schupp, P.G.）、維佳莫爾D.S.（Vinjamur, D.S.）、加西亞S.P.（Garcia, S.P.）、呂克S.（Luc, S.）、栗田R.（Kurita, R.）、納卡穆拉Y.（Nakamura, Y.）、藤原Y.（Fujiwara, Y.）、馬艾達T.（Maeda, T.）、元G.（Yuan, G.）、張F.（Zhang, F.）、奧爾金S.H.（Orkin, S.H.）、以及鮑爾D.E.（Bauer, D.E.）DOI:10.1038/nature15521，2015年9月16日網上公開，該文章藉由引用結合在此並且在下文中簡要討論：

[0943] 康維爾等人涉及進行人和小鼠BCL11A紅系增強子的原位飽和誘變的新型合併的CRISPR-Cas9指導RNA文庫，該等紅系增強子先前被鑒定為係與胎血紅蛋白（HbF）水平相關聯的增強子並且該增強子的小鼠異種同源物係紅系BCL11A表現所必需的。此方法揭示了該等增強子關鍵性最小特徵和離散缺點。藉由編輯原代人祖細胞和小鼠誘變，作者們確認BCL11A紅系增強子作為用於HbF再誘導的靶標。作者們製成報告治療性基因組編輯的詳細增強子圖。

使用Cpf1系統修飾細胞或生物體之方法

[0944] 在一些實施方式中，本發明包括修飾細胞或生物體之方法。細胞可以是原核細胞或真核細胞。細胞可以是哺乳動物細胞。哺乳動物細胞可以是非人類靈長類動物、牛、豬、齧齒動物或小鼠細胞。細胞可以是非哺乳動物真核細胞諸如家禽、魚或蝦的細胞。細胞還可以是植物細胞。植物細胞可以是栽培植物諸如木薯、玉米、高粱、小麥或稻具有的細胞。植物細胞還可以是藻類、樹或蔬菜具有的細胞。藉由本發明引入到細胞的修飾可以使

得細胞和細胞的子代被改變以改進生物產物諸如抗體、澱粉、乙醇或其他所希望的細胞輸出物的產生。藉由本發明引入到細胞的修飾可以使得細胞和細胞的子代包括使所產生的生物產物發生變化的改變。

[0945] 該系統可以包含一種或多種不同載體。在本發明的一方面中，Cas蛋白被密碼子優化為在所希望的細胞類型，優先地是真核細胞，較佳的是哺乳動物細胞或人類細胞中表現。

[0946] 包裝細胞典型地用於形成能夠感染宿主細胞的病毒粒子。此類細胞包括包裝腺病毒的293細胞和包裝逆轉錄病毒的ψ2細胞或PA317細胞。基因治療中使用的病毒載體通常是藉由產生將核酸載體包裝到病毒粒子中的細胞系來生成的。該等載體典型地含有包裝並隨後整合到宿主中所需要的最小病毒序列、被有待表現的一個或多個多核苷酸的表現盒替換的其他病毒序列。失去的病毒功能典型地是由包裝的細胞系反向供應的。例如，基因治療中使用的AAV載體典型地僅具有來自AAV基因組的ITR序列，該等序列係包裝並整合到宿主基因組中所需要的。病毒DNA被包裝在一個細胞系中，該細胞系含有編碼其他AAV基因即rep和cap但缺乏ITR序列的輔助質粒。該細胞系還可以用作為輔助物的腺病毒感染。輔助病毒促進AAV載體複製和來自輔助質粒的AAV基因表現。輔助質粒由於缺乏ITR序列而未大量包裝。腺病毒的污染可以是藉由例如進行腺病毒比AAV更敏感的熱處理來減少的。

遞送

[0947] 本發明涉及經由至少一種奈米粒子複合物遞送的

CRISPR複合物的至少一種組分，例如RNA。在一些方面中，本發明提供了包括以下各項的方法：向宿主細胞遞送一種或多種多核苷酸，諸如在此所述的一種或多種載體、其一種或多種轉錄物和/或由其轉錄的一種或多種蛋白質。在一些方面中，本發明進一步提供了藉由此類方法產生的細胞以及包含此類細胞或由此類細胞產生的動物。在一些實施方式中，將CRISPR酶與指導序列組合（以及視情況與其複合）遞送到細胞中。常規的基於病毒和非病毒的基因轉移方法可以用於在哺乳動物細胞或靶組織中引入核酸。此類方法可以用於向培養基或宿主生物體中的細胞給予編碼CRISPR系統的組分的核酸。非病毒載體遞送系統包括DNA質粒、RNA（例如在此所述的載體的轉錄物）、裸核酸、以及與遞送媒介物脂質體複合的核酸。病毒載體遞送系統包括DNA和RNA病毒，該等病毒在遞送至細胞後具有附加型基因組或整合型基因組。對於基因治療程序的綜述，參見安德森，科學256:808-813 (1992)；納貝爾和費爾格納，TIBTECH 11:211-217 (1993)；三穀和卡斯基，TIBTECH 11:162-166 (1993)；狄龍，TIBTECH 11:167-175 (1993)；米勒，自然357:455-460 (1992)；範布倫特，生物技術 6(10):1149-1154 (1988)；比涅，恢復神經學和神經科學 8:35-36 (1995)；克雷默和佩里科代，英國醫學公報 51(1):31-44 (1995)；哈嗒嗒等人，微生物學和免疫學的前沿課題，竇爾弗勒和博姆（編輯）(1995)；以及余等人，基因治療 1:13-26 (1994)。

[0948] 非病毒遞送核酸的方法包括包括脂質轉染、微注射、基因槍、病毒體、脂質體、免疫脂質體、聚陽離子或脂質:核酸軛合物、裸DNA、人工病毒體遞送、以及藥劑增強DNA攝取。脂質

轉染描述於例如美國專利案號5,049,386、4,946,787；以及4,897,355）並且脂質轉染試劑係商業上銷售的（例如，Transfectam™和Lipofectin™）。適用於多核苷酸的有效受體識別脂質轉染的陽離子脂質和中性脂質包括以下各項中的那些：費爾格納WO 91/17424；WO 91/16024。遞送可以是遞送至細胞（例如，體外或離體給予）或靶組織（例如，體內給予）。

[0949] 脂質:核酸複合物（包括靶向脂質體，諸如免疫脂質複合物）的製備係熟習該項技術者已熟知的（例如，參見克裡斯特爾（Crystal），科學270:404-410 (1995)；布萊澤等人，癌症基因治療2:291-297 (1995)；貝爾等人，生物共軛化學5:382-389 (1994)；雷米等人，生物共軛化學5:647-654 (1994)；高等人，基因治療2:710-722 (1995)；艾哈邁德等人,癌症研究52:4817-4820 (1992)；美國專利案號4,186,183、4,217,344、4,235,871、4,261,975、4,485,054、4,501,728、4,774,085、4,837,028、以及4,946,787）。

[0950] 使用用於遞送核酸的基於RNA或DNA病毒的系統利用了用於將病毒靶向身體內的特定細胞並且將病毒有效負載運輸到核內的高度進化的方法。病毒載體可以直接給予至患者（體內）或者它們可以用於體外處理細胞，並且修飾細胞可以視情況給予至患者（離體）。常規基於病毒的系統可以包括用於基因轉移的逆轉錄病毒、慢病毒、腺病毒、腺相關病毒以及單純皰疹病毒載體。藉由逆轉錄病毒、慢病毒、以及腺相關病毒基因轉移方法整合在宿主基因組中是可能的，這常常導致插入的轉基因長期表現。另外，高轉導效率已在許多不同細胞類型和靶組織中觀察到。

[0951] 逆轉錄病毒的趨向性可以藉由結合外源包膜蛋白來改變，從而擴增靶細胞的潛在靶群體。慢病毒載體係能夠轉導或感染非分裂細胞並典型地產生高病毒滴度的逆轉錄病毒載體。逆轉錄病毒基因轉移系統的選擇因此將取決於靶組織。逆轉錄病毒載體包含具有多至6-10 kb外源序列的包裝容量的順式作用長末端重複序列。最小量順式作用LTR係足以複製並包裝該等載體的，然後使用該等載體將治療基因整合到靶細胞中，以提高永久性轉基因表現。廣泛使用的逆轉錄病毒載體包括基於以下各項的那些：鼠白血病病毒（MuLV）、長臂猿白血病病毒（GaLV）、猴免疫缺陷病毒（SIV）、人免疫缺陷病毒（HIV）、以及其組合（例如，參見，布奇謝爾等人，病毒學雜誌66:2731-2739 (1992)；約翰等人，病毒學雜誌66:1635-1640 (1992)；紹姆內爾費爾特等人，病毒學176:58-59 (1990)；威爾遜等人，病毒學雜誌63:2374-2378 (1989)；米勒等人，病毒學雜誌65:2220-2224 (1991)；PCT/US94/05700）。

[0952] 在另一個實施方式中，預期可卡耳水泡病毒包膜假型化逆轉錄病毒載體粒子（例如，參見轉讓給福瑞德哈金森腫瘤研究中心（Fred Hutchinson Cancer Research Center）的美國專利公開案號20120164118）。可卡耳病毒係在水泡病毒屬中，並且係哺乳動物的水泡性口炎的致病物。可卡耳病毒最初從千里達拉島（Trinidad）的蠕蟲中分離（喬恩克爾（Jonkers）等人，美國獸醫研究雜誌（Am. J. Vet.Res.）25:236-242 (1964)），並且在千里達拉島、巴西和阿根廷已從昆蟲、牛和馬中鑒定到感染。感染哺乳動物的許多水泡病毒已從天然感染的節肢動物中分離，這表明它們係載體傳播（vector-borne）的。水泡病毒抗體在生活於農村地區

的人中是常見，而該等病毒係地方性的並且係實驗室採集的；人類的感染通常導致流感樣症狀。可卡耳病毒包膜糖蛋白與VSV-G Indiana享有71.5%胺基酸水平的一致性，並且水泡病毒包膜基因的系統發育比較顯示可卡耳病毒在血清學上不同於水泡病毒內的VSV-G Indiana病毒株，但與其緊密相關。喬恩克爾等人，美國獸醫研究雜誌25:236-242 (1964)和特拉瓦索斯達羅薩（Travassos da Rosa）等人，美國熱帶醫學和衛生雜誌（Am. J. Tropical Med.& Hygiene）33:999-1006 (1984)。可卡耳水泡病毒包膜蛋白假型化逆轉錄病毒載體粒子可以包括慢病毒、α逆轉錄病毒、β逆轉錄病毒、γ逆轉錄病毒、δ逆轉錄病毒以及ε逆轉錄病毒載體粒子，該等載體可以可以包含逆轉錄病毒Gag、Pol、和/或一種或多種輔助蛋白以及可卡耳水泡病毒包膜蛋白。在該等實施方式的某些方面中，Gag、Pol和輔助蛋白係慢病毒和/或γ逆轉錄病毒的。本發明提供了含有以下各項或基本上由以下各項組成的AAV：編碼CRISPR系統的外源性核酸分子，例如包含第一盒或基本上由該第一盒組成的多個盒，該第一盒包含啟動子、編碼CRISPR-相關（Cas）蛋白（推定的核酸酶或解旋酶蛋白）例如Cpf1的核酸分子和終止子或基本上尤其組成，以及兩個或更多個、有利地多至包裝尺寸限度的載體，例如總計（包括該第一盒）五個包含啟動子、編碼指導RNA（gRNA）的核酸分子和終止子或基本上由其組成的盒（例如，每個盒示意性表示為啟動子-gRNA1-終止子、啟動子-gRNA2-終止子...啟動子-gRNA(N)-終止子（其中N係可以插入的數值，該數值係載體包裝尺寸限度的上限），或者兩個或更多個單獨的rAAV，各自含有CRISPR系統的一個或超過一個盒，例如第一個rAAV含有啟動子、

編碼Cas例如Cas（Cpf1）的核酸分子和終止子或基本上由其組成的第一盒，並且第二rAAV含有多個、四個包含啟動子、編碼指導RNA(gRNA)的核酸分子和終止子或基本上由其組成的盒(例如，每個盒示意性表示為啟動子-gRNA1-終止子、啟動子-gRNA2-終止子...啟動子-gRNA(N)-終止子（其中N係可以插入的數值，該數值係在該載體的包裝尺寸限度的上限內）。由於rAAV係DNA病毒，所以在此關於AAV或rAAV的討論中的核酸分子有利地是DNA。在一些實施方式中，該啟動子有利地是人類突觸蛋白I啟動子(hSyn)。用於將核酸遞送至細胞的其他方法係熟習該項技術者已知的。例如，參見US20030087817，該專利藉由引用結合在此。

[0953] 在一些實施方式中，宿主細胞係用在此所述的一種或多種載體暫態轉染或非暫態轉染的。在一些實施方式中，細胞當天然存在於受試者中時被轉染。在一些實施方式中，轉染的細胞係從受試者中獲得的。在一些實施方式中，細胞係來源於從受試者中獲得的細胞，諸如細胞系。用於組織培養的多種多樣的細胞系係本領域已知的。細胞系包括但不限於，C8161、CCRF-CEM、MOLT、mIMCD-3、NHDF、HeLa-S3、Huh1、Huh4、Huh7、HUVEC、HASMC、HEKn、HEKa、MiaPaCell、Panc1、PC-3、TF1、CTLL-2、C1R、Rat6、CV1、RPTE、A10、T24、J82、A375、ARH-77、Calu1、SW480、SW620、SKOV3、SK-UT、CaCo2、P388D1、SEM-K2、WEHI-231、HB56、TIB55、Jurkat、J45.01、LRMB、Bcl-1、BC-3、IC21、DLD2、Raw264.7、NRK、NRK-52E、MRC5、MEF、Hep G2、HeLa B、HeLa T4、COS、COS-1、COS-6、COS-M6A、BS-C-1猴腎上皮細胞、BALB/ 3T3小鼠胚成纖維細胞、3T3 Swiss、

3T3-L1、132-d5人胎兒成纖維細胞；10.1小鼠成纖維細胞、293-T、3T3、721、9L、A2780、A2780ADR、A2780cis、A172、A20、A253、A431、A-549、ALC、B16、B35、BCP-1 cells、BEAS-2B、bEnd.3、BHK-21、BR 293、BxPC3、C3H-10T1/2、C6/36、Cal-27、CHO、CHO-7、CHO-IR、CHO-K1、CHO-K2、CHO-T、CHO Dhfr -/-、COR-L23、COR-L23/CPR、COR-L23/5010、COR-L23/R23、COS-7、COV-434、CML T1、CMT、CT26、D17、DH82、DU145、DuCaP、EL4、EM2、EM3、EMT6/AR1、EMT6/AR10.0、FM3、H1299、H69、HB54、HB55、HCA2、HEK-293、HeLa、Hepa1c1c7、HL-60、HMEC、HT-29、Jurkat、JY細胞、K562細胞、Ku812、KCL22、KG1、KYO1、LNCap、Ma-Mel 1-48、MC-38、MCF-7、MCF-10A、MDA-MB-231、MDA-MB-468、MDA-MB-435、MDCK II、MDCK II、MOR/0.2R、MONO-MAC 6、MTD-1A、MyEnd、NCI-H69/CPR、NCI-H69/LX10、NCI-H69/LX20、NCI-H69/LX4、NIH-3T3、NALM-1、NW-145、OPCN / OPCT細胞系、Peer、PNT-1A / PNT 2、RenCa、RIN-5F、RMA/RMAS、Saos-2細胞、Sf-9、SkBr3、T2、T-47D、T84、THP1細胞系、U373、U87、U937、VCaP、Vero細胞、WM39、WT-49、X63、YAC-1、YAR、以及其轉基因品種。細胞系可從熟習該項技術者已知的多種來源獲得（例如，參見美國典型培養物保藏中心（ATCC）（維吉尼亞州馬納薩斯））。在一些實施方式中，用在此所述的一種或多種載體轉染的細胞用於建立包含一種或多種載體來源的序列的新細胞系。在一些實施方式中，將用如在此所述的CRISPR系統的組分暫態轉染（諸如藉由暫態轉染一種或多種載體或用RNA轉染）並且藉由CRISPR複合物的活性修飾的細胞用於建立包括含有修

飾但缺乏任何其他外源性序列的細胞的細胞系。在一些實施方式中，將用在此所述的一種或多種載體暫態轉染或非暫態轉染的細胞或來源於此類細胞的細胞系用於評定一種或多種測試化合物。

[0954] 在一些實施方式中，在此所述的一種或多種載體用於產生非人類轉基因動物或轉基因植物。在一些實施方式中，轉基因動物係哺乳動物，諸如小鼠、大鼠或兔。用於產生轉基因動物和植物的方法係本領域已知的並且通常以細胞轉染方法諸如在此所述的方法開始。在另一個實施方式中，可以預期具有針陣列的流體遞送裝置（例如，參見轉讓給福瑞德哈金森腫瘤研究中心的美國專利公開案號20110230839）用於將CRISPR Cas遞送至實體組織。美國專利公開案號20110230839的用於將流體遞送至實體組織的裝置可以包括按陣列安排的多個針；多個內存，每個內存與該多個針中的一個對應針流體連通；以及多個制動器，該多個制動器可操作地連接至該多個內存中的對應內存並且被配置為控制該內存內的流體壓力。在某些實施方式中，該多個制動器中的每個制動器可以包括多個柱塞之一，該多個柱塞中的每個柱塞的第一端被接收在該多個內存中的對應內存中，並且在某些另外的實施方式中，該多個柱塞中的該等柱塞在第二端可操作地連接在一起，以便可同時下壓。某些另外的實施方式可以包括柱塞驅動器，該柱塞驅動器被配置為以選擇性可變速率壓下所有該多個柱塞。在其他實施方式中，該多個制動器中的每個制動器可以包括具有第一端和第二端的多個流體傳輸線中的一個流體傳輸線，該多個流體傳輸線中的每個流體傳輸線的第一端連接至該多個內存中的一個對應內存。在其他實施方式中，該裝置可以包括一個流體壓力

來源，並且該多個制動器中的每個制動器包括該流體壓力來源與該多個內存中的一個對應內存之間的流體連接。在另外的實施方式中，流體壓力來源可以包括以下各項中的至少一種：壓縮器、真空貯氣筒、蠕動泵、主缸、微流體泵、以及閥。在另一個實施方式中，該多個針中的每個針可以包括沿著其長度分佈的多個埠。

[0955] 在一個方面中，本發明提供了修飾真核細胞中的靶多核苷酸的方法。在一些實施方式中，該方法包括使得核酸靶向複合物結合靶多核苷酸來實施所述靶多核苷酸的切割，從而修飾該靶多核苷酸，其中該核酸靶向複合物包含與雜交至所述靶多核苷酸內的靶序列的指導RNA複合的核酸靶向效應蛋白。

[0956] 在另一個方面中，本發明提供了修飾多核苷酸在真核細胞中的表現之方法。在一些實施方式中，該方法包括允許核酸靶向複合物結合該多核苷酸，以使得所述結合導致所述多核苷酸的表現增加或減少；其中該核酸靶向複合物包含與雜交至所述多核苷酸內的靶序列的指導RNA複合的核酸靶向效應蛋白。

[0957] CRISPR複合物組分可以是藉由與轉運部分軛合或締合來遞送（例如，改編自美國專利案號8,106,022；8,313,772）。核酸遞送策略可以例如用於改進指導RNA或信使RNA或編碼CRISPR複合物組分的編碼DNA的遞送。例如，RNA可以結合修飾的RNA核苷酸來提高穩定性、減小免疫刺激並且/或者改進特異性（參見德勒埃維，葛蘭·F.（Deleavey, Glen F.）等人，2012，化學與生物學（Chemistry & Biology），第19卷，第8期，937-954；紮利皮斯科（Zalipsky），1995，先進藥物遞送評論（Advanced Drug

Delivery Reviews）16: 157-182；卡利色提（Caliceti）和威羅尼（Veronese），2003，先進藥物輸送評論55: 1261-1277）。已描述可以用於修飾核酸諸如gRNA以進行更有效的遞送的不同構建體，諸如可以適於修飾gRNA以便具有更大疏水性和非離子性從而提高細胞進入的可逆性電荷中和磷酸三酯骨架修飾（米德·BR（Meade BR）等人，2014，自然生物技術 32，1256–1261）。在其他替代實施方式中，選定的RNA模體可以適用於介導細胞轉染（麥哲倫·M.（Magalhães M.）等人，分子治療（Molecular Therapy）(2012)；20 3, 616–624）。類似地，適配體可以適於例如藉由將適配體附加至gRNA來遞送CRISPR複合物組分（塔恩（Tan）等人，2011，生物技術趨勢（Trends in Biotechnology），2011年12月，第29卷，第12期）。

[0958] 在一些實施方式中，三觸角N-乙醯基半乳糖胺（GalNAc）與寡核苷酸組分的軛合可以用於改進遞送，例如選擇細胞類型例如肝細胞的遞送（參見WO2014118272，該專利藉由引用結合在此；耐爾，JK（Nair, JK）等人，2014，美國化學學會雜誌（Journal of the American Chemical Society）136 (49), 16958-16961）。這可以被認為是基於糖的粒子並且在此提供了關於其他粒子遞送系統和/或配製物的其他詳情。GalNAc因此可以被認為是在此所述的其他粒子的意義上的粒子，以使得一般用途和其他考慮因素（例如所述粒子的遞送）也應用於GalNAc粒子。溶液相軛合策略可以例如用於將作為PFP (五氟苯酚)酯啟動的三觸角GalNAc簇（分子量~2000）附接到5′-己基胺基修飾的寡核苷酸上（5′-HA ASO，分子量~8000 Da；奧斯塔蓋得（Østergaard）等人，

生物共軛化學，2015, 26 (8),第1451-1455頁）。類似地，已描述用於體內核酸遞送的聚(丙烯酸酯)聚合物（參見WO2013158141，該專利藉由引用結合在此）。在其他替代實施方式中，為了改進遞送，可以使用預先混合的CRISPR奈米粒子（或蛋白質複合物）與天然存在的血清蛋白（阿克尼克·A（Akinc A）等人，2010，分子治療（Molecular Therapy），第18卷第7期，1357–1364）。

[0959] 篩選技術可用於鑒定遞送增強子，例如藉由篩選化學文庫（吉爾埃倫·J.（Gilleron J.）等人，2015，核酸研究43 (16): 7984-8001）。還已描述了用於測定遞送媒介物諸如奈米粒子的效率的方法，該等方法可以用於鑒定對於CRISPR組分的有效遞送媒介物（參見薩哈義·G.（Sahay G.）等人，2013，自然生物技術31, 653–658）。

[0960] 在一些實施方式中，蛋白質CRISPR組分的遞送可以是藉由將功能肽（諸如改變蛋白質疏水性的肽）添加到該蛋白質中，例如以便改進體內功能來實現。CRISPR組分蛋白可以類似地被修飾為促進隨後的化學反應。例如，胺基酸可以被添加到具有經受點擊化學的基團的蛋白質中（尼基克·I.（Nikić I.）等人，2015，自然實驗手冊（Nature Protocols）10,780–791）。在這種類型的實施方式中，點擊化學基團那麼可以用於添加多種多樣的替代性結構，諸如用於穩定的聚(乙二醇)、細胞穿透肽、RNA適配體、脂質、或碳水化合物諸如GalNAc。在其他替代方案中，CRISPR組分蛋白可以被修飾為適應進入細胞蛋白（參見斯文森（Svensen）等人，2012，藥理學趨勢（Trends in Pharmacological Sciences），第33卷，

第4期），例如藉由將細胞穿透肽添加到該蛋白質（參見考夫曼，W.·伯克利（Kauffman, W. Berkeley）等人，2015，生物化學趨勢（Trends in Biochemical Sciences），第40卷，第12期，749-764；科倫（Koren）和托爾基林（Torchilin），2012，分子醫學趨勢（Trends in Molecular Medicine），第18卷，第7期）。在另一個替代實施方式中，患者或受試者可以用有助於CRISPR組分隨後遞送的化合物或配製物預處理。

Cpf1效應蛋白複合物可以用於植物中

[0961] 一種或多種Cpf1效應蛋白系統（例如，單一或多重）可以與農作物基因組的研究進展結合來使用。在此所述的系統可以用於進行有效且性價比高的植物基因或基因組探察或編輯或操縱，例如，快速研究並且/或者選擇並且/或者探察並且/或者比較並且/或者操縱並且/或者轉化植物基因或基因組；例如，以為一種或多種植物創建、鑒定、發展、優化或賦予一種或多種性狀或一種或多種特徵或者以轉化植物基因組。因此可以存在植物、具有新性狀或特徵組合的新植物或具有增強的性狀的新植物的改進的產生方法。一種或多種Cpf1效應蛋白系統可以用於植物中，以定點整合（SDI）或基因編輯（GE）或任何近反向育種（Near Reverse Breeding，NRB）或反向育種（RB）技術。利用在此所述的Cpf1效應蛋白系統的方面可能類似於CRISPR-Cas（例如，CRISPR-Cas9）系統在植物中的使用，並且參考亞利桑那大學（University of Arizona）網站"CRISPR-PLANT"（http://www.genome.arizona.edu/crispr/）（得到賓州州立大學

（Penn State）和AGI的支持）。本發明的實施方式可以用於在植物中或在先前已使用RNAi或類似基因組編輯技術的情況中進行基因組編輯。例如，參見涅克拉索夫（Nekrasov），“植物基因組編輯一點通：在模型和農作物植物中使用CRISPR-Cas系統進行靶向誘變（Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR-Cas system）”，植物方法（Plant Methods）2013, 9:39 (doi:10.1186/1746-4811-9-39)；布魯克斯（Brooks），“在第一代番茄中使用CRISPR-Cas9系統進行有效基因編輯（Efficient gene editing in tomato in the first generation using the CRISPR-Cas9 system）”，植物生理學（ Plant Physiology），2014年9月，第114.247577頁；單（Shan），“使用CRISPR-Cas系統進行農作物植物的靶向基因組修飾（Targeted genome modification of crop plants using a CRISPR-Cas system）”，自然生物技術31, 686-688 (2013)；馮（Feng），“在植物中使用CRISPR/Cas系統進行有效基因組編輯（Efficient genome editing in plants using a CRISPR/Cas system）”，細胞研究（Cell Research）(2013) 23:1229–1232. doi:10.1038/cr.2013.114；2013年8月20日線上公開；謝（Xie），“在植物中使用CRISPR-Cas系統進行RNA指導的基因組編輯（RNA-guided genome editing in plants using a CRISPR-Cas system）”，分子植物（Mol Plant.），2013年11月；6(6):1975-83. doi: 10.1093/mp/sst119. 電子版2013年8月17日；許（Xu），“在稻中使用根癌土壤桿菌介導的CRISPR-Cas系統進行基因靶向（Gene targeting using the Agrobacterium tumefaciens-mediated CRISPR-Cas system in rice）”，稻（Rice）2014, 7:5 (2014)，週（Zhou）等

人，“在異型雜交木多年生楊樹中利用SNP進行二對偶基因CRISPR突變揭示了4-香豆酸：CoA連接酶特異性和豐餘性（Exploiting SNPs for biallelic CRISPR mutations in the outcrossing woody perennial Populus reveals 4-coumarate: CoA ligase specificity and Redundancy）”，新植物學家（New Phytologist）(2015)（論壇）1-4（僅可在www.newphytologist.com處線上獲得）；卡利安多（Caliando）等人，“使用在宿主基因組中穩定攜帶的CRISPR裝置進行靶向DNA降解（Targeted DNA degradation using a CRISPR device stably carried in the host genome）”，自然通訊（NATURE COMMUNICATIONS）6:6989, DOI: 10.1038/ncomms7989, www.nature.com/naturecommunications DOI: 10.1038/ncomms7989；美國專利案號6,603,061—土壤桿菌屬介導的植物轉化方法（Agrobacterium-Mediated Plant Transformation Method）；美國專利案號7,868,149—植物基因組序列及其用途（Plant Genome Sequences and Uses Thereof）以及US 2009/0100536—具有增強的農藝性狀的轉基因植物（Transgenic Plants with Enhanced Agronomic Traits），每份文獻的所有內容和揭露均藉由引用以其整體結合在此。在本發明的實踐中，莫雷爾（Morrell）等人“農作物基因組：發展與應用（Crop genomics: advances and applications）”，遺傳學自然評論（Nat Rev Genet.），2011年12月29日；13(2):85-96的內容和揭露；每份文獻藉由引用結合在此，包括關於在此的實施方式可以如何用於植物一樣。因此，除非另外表明，否則在此提及動物細胞也可以將必要的修正應用於植物細胞；並且，在此具有減小的脫靶效應的酶和採用此類酶的系統可以用於植物應用，

包括在此提及的那些。

Cpf1-CRISPR 系統對植物和酵母的應用

定義：

[0962] 總的來說，術語“植物”涉及植物界中藉由細胞分裂特徵性生長、含有葉綠體並具有包含纖維素的細胞壁的任何不同光合作用生物體、真核生物體、單細胞生物體或多細胞生物體。術語植物涵蓋單子葉植物和雙子葉植物。確切的說，該等植物旨在包括但不限於被子植物和裸子植物，諸如刺槐、苜蓿、莧、蘋果、杏、洋薊、灰樹、蘆筍、鱷梨、香蕉、大麥、豆、甜菜、樺樹、山毛櫸、黑莓、藍莓、花椰菜、芽球甘藍、捲心菜、油菜（canola）、哈密瓜、胡蘿蔔、木薯、菜花、雪松、穀類、芹菜、栗樹、櫻桃、大白菜、柑橘、克萊門小柑橘、三葉草、咖啡、穀物、棉花、豇豆、黃瓜、柏樹、茄子、榆樹、菊苣、桉樹、茴香、無花果、樅樹、天竺葵、葡萄、葡萄柚、落花生（groundnut）、地櫻桃、樹膠鐵杉、山核桃、羽衣甘藍、奇異果、甘藍、落葉松、生菜、韭蔥、檸檬、酸橙、洋槐、松樹、掌葉鐵線蕨、玉米（maize）、芒果、楓樹、甜瓜、小米、蘑菇、芥菜、堅果、橡樹、燕麥、油棕樹、秋葵、洋蔥、桔子、觀賞植物或花或樹、木瓜、棕櫚樹、歐芹、防風草、豌豆、桃樹、花生（peanut）、梨樹、泥煤苔（peat）、胡椒、柿子樹、木豆、松樹、鳳梨、車前草、李子、石榴、馬鈴薯、南瓜、菊苣、蘿蔔、油菜籽、覆盆子、稻穀、黑麥、高粱、紅花、黃華柳、大豆、菠菜、雲杉、南瓜、草莓、甜菜、甘蔗、向日葵、甘薯、甜玉米、橘子、茶、煙草、番茄、樹木、黑小麥、草坪草、蕪菁、葡萄樹、

胡桃、西洋菜、西瓜、小麥、山藥、紫杉、以及綠皮西葫蘆。術語植物還涵蓋藻類，該等藻類主要係統一標準主要為缺乏根、葉和表徵高等植物的其他器官的光合自養生物。

[0963] 用於使用在此所述的Cpf1系統進行基因組編輯的方法可以用於對基本上任何植物賦予所希望的性狀。各種各樣的植物和植物細胞系統可以使用本揭露的核酸構建體和以上提及的各種轉化方法來工程化為如在此所述的希望的生理學和農藝學特徵。在較佳的實施方式中，用於工程化的靶植物和植物細胞包括但不限於，那些單子葉植物和雙子葉植物，諸如農作物，包括穀類作物（例如，小麥、玉米、稻米、小米、大麥）、果實農作物（例如，番茄、蘋果、梨、草莓、桔子）、草料作物（例如，苜蓿）、根用蔬菜作物（例如，胡蘿蔔、馬鈴薯、甜萊、山藥）、葉用蔬菜作物（例如，生菜、菠菜）；有花植物（例如，矮牽牛、玫瑰、菊花）、針葉樹和松樹（例如，冷杉、雲杉）；植物治理法中使用的植物（例如，重金屬積累的植物）；油料作物（例如，向日葵、油菜籽）和用於實驗目的的植物（例如，擬南芥屬）。因此，該等方法和CRISPR-Cas系統可以用於廣泛範圍的植物，例如像屬於以下目的雙子葉植物：木蘭目（Magniolales）、八角目（Illiciales）、樟目（Laurales）、胡椒目（Piperales）、馬兜鈴目（Aristolochiales）、睡蓮目（Nymphaeales）、毛茛目（Ranunculales）、罌粟目（Papaverales）、瓶子草科（Sarraceniaceae）、昆欄樹目（Trochodendrales）、金縷梅目（Hamamelidales）、杜仲目（Eucommiales）、塞子木目（Leitneriales）、楊梅目（Myricales）、殼鬥目（Fagales）、木麻黃目（Casuarinales）、石竹目（Caryophyllales）、肉穗果目（Batales）、

蓼目（Polygonales）、藍雪目（Plumbaginales）、五桠果目（Dilleniales）、山茶目（Theales）、錦葵目（Malvales）、蕁麻目（Urticales）、玉蕊目（Lecythidales）、堇菜目（Violales）、楊柳目（Salicales）、白花菜目（Capparales）、杜鵑花目（Ericales）、岩梅目（Diapensiales）、柿樹目（Ebenales）、報春花目（Primulales）、薔薇目（Rosales）、豆目（Fabales）、川草目（Podostemales）、小二仙草目（Haloragales）、桃金娘目（Myrtales）、山茱萸目（Cornales）、山龍眼目（Proteales）、檀香目（San tales）、大花草目（Rafflesiales）、衛矛目（Celastrales）、大戟目（Euphorbiales）、鼠李目（Rhamnales）、無患子目（Sapindales）、胡桃目（Juglandales）、牻牛兒苗目（Geraniales）、遠志目（Polygalales）、傘形目（Umbellales）、龍膽目（Gentianales）、花蔥目（Polemoniales）、唇形目（Lamiales）、車前草目（Plantaginales）、玄參目（Scrophulariales）、桔梗目（Campanulales）、茜草目（Rubiales）、川續斷目（Dipsacales）以及菊目（Asterales）；該等方法和CRISPR-Cas可以用於單子葉植物，諸如屬於以下目的單子葉植物：澤瀉目（Alismatales）、水鱉目（Hydrocharitales）、茨藻目（Najadales）、黴草目（Triuridales）、鴨跖草目（Commelinales）、穀精草目（Eriocaulales）、帚燈草目（Restionales）、禾本目（Poales）、燈芯草目（Juncales）、莎草科（Cyperales）、香蒲目（Typhales）、鳳梨目（Bromeliales）、薑目（Zingiberales）、檳榔目（Arecales）、環花目（Cyclanthales）、露兜樹目（Pandanales）、天南星目（Arales）、百合目（Lilliales）以及蘭目（Orchid ales），或者用於屬於裸子植物的植物，例如屬於松杉目（Pinales）、銀杏目（Ginkgoales）、蘇鐵目（Cycadales）、

南洋杉目(Araucariales)、柏目(Cupressales)以及麻黃目(Gnetales)。

[0964] 在此所述的Cpf1 CRISPR系統和使用方法可以用於廣泛範圍的植物種類，包括在下面的雙子葉植物、單子葉植物或裸子植物屬的非限制性列表中：顛茄屬（*Atropa*）、油丹屬（*Alseodaphne*）、腰果屬（*Anacardium*）、落花生屬（*Arachis*）、瓊楠屬（*Beilschmiedia*）、芸苔屬（*Brassica*）、紅花屬（*Carthamus*）、木防己屬（*Cocculus*）、巴豆屬（*Croton*）、甜瓜屬（*Cucumis*）、柑橘屬（*Citrus*）、西瓜屬（*Citrullus*）、辣椒屬（*Capsicum*）、長春花屬（*Catharanthus*）、椰子屬（*Cocos*）、咖啡屬（*Coffea*）、南瓜屬（*Cucurbita*）、胡蘿蔔屬（*Daucus*）、杜氏木屬（*Duguetia*）、花菱草屬（*Eschscholzia*）、榕屬（*Ficus*）、草莓屬（*Fragaria*）、海罌粟屬（*Glaucium*）、大豆屬（*Glycine*）、棉屬（*Gossypium*）、向日葵屬（*Helianthus*）、橡膠樹屬（*Hevea*）、天仙子屬（*Hyoscyamus*）、萵苣屬（*Lactuca*）、卷枝藤屬（*Landolphia*）、亞麻屬（*Linum*）、木薑子屬（*Litsea*）、番茄屬（*Lycopersicon*）、羽扇豆屬（*Lupinus*）、木薯屬（*Manihot*）、馬郁蘭屬（*Majorana*）、蘋果屬（*Malus*）、苜蓿屬（*Medicago*）、煙草屬（*Nicotiana*）、木犀欖屬（*Olea*）、銀膠菊屬（*Parthenium*）、罌粟屬（*Papaver*）、鱷梨屬（*Persea*）、菜豆屬（*Phaseolus*）、黃連木屬(*Pistacia*)、豌豆屬(*Pisum*)、梨屬(*Pyrus*)、李屬（*Prunus*）、蘿蔔屬（*Raphanus*）、蓖麻屬（*Ricinus*）、千里光屬（*Senecio*）、防己屬（*Sinomenium*）、千金藤屬（*Stephania*）、歐白芥屬（*Sinapis*）、茄屬（*Solanum*）、可可屬（*Theobroma*）、三葉

草屬（*Trifolium*）、胡蘆巴屬（*Trigonella*）、蠶豆屬（*Vicia*）、蔓長春花屬（*Vinca*）、葡萄屬（*Vilis*）以及豇豆屬（*Vigna*）；以及蔥屬（*Allium*）、須芒草屬（*Andropogon*）、畫眉草屬（*Aragrostis*）、天門冬屬（*Asparagus*）、燕麥屬（*Avena*）、狗牙根屬（*Cynodon*）、油棕屬（*Elaeis*）、羊茅屬（*Festuca*）、羊茅黑麥草屬（*Festulolium*）、萱草屬（*Heterocallis*）、大麥屬（*Hordeum*）、浮萍屬（*Lemna*）、毒麥屬（*Lolium*）、芭蕉屬（*Musa*）、稻屬（*Oryza*）、黍屬（*Panicum*）、狼尾草屬（*Pannesetum*）、梯牧草屬（*Phleum*）、早熟禾屬（*Poa*）、黑麥屬（*Secale*）、高粱屬（*Sorghum*）、小麥屬（*Triticum*）、玉蜀黍屬（*Zea*）、冷杉屬（*Abies*）、杉木屬（*Cunninghamia*）、麻黃屬（*Ephedra*）、雲杉屬（*Picea*）、松屬（*Pinus*）、以及黃杉屬（*Pseudotsuga*）。

[0965] Cpf1 CRISPR系統和使用方法也可以用於廣泛範圍的“藻類”或“藻類細胞”；包括例如選自若干真核生物門的藻類，包括紅藻門（紅藻）、綠藻門（綠藻）、褐藻門（褐藻）、矽藻門（矽藻）、真眼點藻綱以及溝鞭藻類，以及原核生物門藍藻細菌（藍綠藻類）。術語“藻類”包括例如選自以下各項的藻類：雙眉藻屬（Amphora）、魚腥藻屬（Anabaena）、纖維藻屬（Anikstrodesmis）、叢粒藻屬（Botryococcus）、角毛藻屬（Chaetoceros）、衣藻屬（Chlamydomonas）、綠藻屬（Chlorella）、綠球藻屬（Chlorococcum）、小環藻屬（Cyclotella）、筒柱藻屬（Cylindrotheca）、杜氏藻屬（Dunaliella）、球石藻屬（Emiliana）、眼蟲屬（Euglena）、紅球藻屬（Hematococcus）、等鞭金藻屬

（Isochrysis）、單鞭金藻屬（Monochrysis）、單針藻屬（Monoraphidium）、微綠球藻屬（Nannochloris）、擬微綠球藻屬（Nannnochloropsis）、舟形藻屬（Navicula）、腎鞭藻屬（Nephrochloris）、腎片藻屬（Nephroselmis）、菱形藻屬（Nitzschia）、節球藻屬（Nodularia）、念珠藻屬（Nostoc）、髓球藻屬（Oochromonas）、卵囊藻（Oocystis）、顫藻（Oscillartoria）、巴夫藻屬（Pavlova）、褐指藻屬（Phaeodactylum）、扁藻屬（Playtmonas）、顆石藻屬（Pleurochrysis）、紫菜屬（Porhyra）、假魚腥藻（Pseudoanabaena）、塔胞藻屬（Pyramimonas）、裂絲藻屬（Stichococcus）、聚球藻屬（Synechococcus）、集胞藻屬（Synechocystis）、四片藻屬（Tetraselmis）、海鏈藻屬（Thalassiosira）、以及束毛藻屬（Trichodesmium）。

[0966] 植物的一部分即“植物組織”可以根據本發明的方法進行處理以產生改進的植物。植物組織還涵蓋植物細胞。如在此所用的術語“植物細胞”係指活體植物的個體單元，在完整全株或在體外組織培養基中、在培養基或瓊脂上、在生長培養基或緩衝液的懸浮液中或作為高等組織單元一部分生長的分離形式，例如像植物組織、植物器官或全株。

[0967] “原生質體”係指植物細胞使用例如機械或酶促方式完全去除或部分去除保護性細胞壁從而形成的活體植物的完整生物化學活性單元，該活性單元在適當生長條件下可以重新形成細胞壁、增殖並再生成全株。

[0968] 術語“轉化”廣泛地是指植物宿主藉由借助於土壤桿

菌或各種化學或物理方法之一來引入DNA從而進行遺傳修飾的過程。如在此所用，術語“植物宿主”係指植物，包括植物的任何細胞、組織、器官或子代。許多適合的植物組織或植物細胞可以被轉化，並且包括但不限於，原生質體、體細胞胚胎、花粉、葉、幼苗、莖、愈傷組織、匍伏莖、試管塊莖、以及胚芽。植物組織還是指這種植物的任何殖株、種子、子代、繁殖體（無論是有性繁殖產生或無性繁殖產生）、以及任何該等的後代諸如切塊或種子。

[0969] 如在此所用術語“轉化的”係指已引入外源DNA分子諸如構建體的細胞、組織、器官或生物體。引入的DNA分子可以整合到受體細胞、組織、器官或生物體的基因組DNA中，以使得引入的DNA分子被傳輸到隨後的子代中。在該等實施方式中，“轉化的”或“轉基因的”細胞或植物也可以包括細胞或植物的子代以及藉由育種程式採用這種轉化的植物作為雜交的母體並表現出因引入的DNA分子的存在而產生的改變的表型的子代。較佳的是，轉基因植物係能育的並且能夠將引入的DNA藉由有性繁殖傳輸到子代中。

[0970] 術語“子代”諸如轉基因植物的子代係由植物或轉基因植物生出的、由其產生的或從其來源的子代。引入的DNA分子還可以暫態轉染到受體細胞中，以使得引入的DNA分子未被隨後的子代繼承並因此不認為是“轉基因的”。因此，如在此所用，“非轉基因”植物或植物細胞係不含有穩定整合到其基因組中的外源DNA的植物。

[0971] 如在此所用，術語“植物啟動子”係能夠開啟植物細胞

內的轉錄（無論其起源是否是植物細胞）的啟動子。示例性的適合植物啟動子包括但不限於，從植物、植物病毒和包含在植物細胞內表現的基因的細菌諸如土壤桿菌屬或根瘤菌屬中獲得的啟動子。

[0972] 如在此所用，“真菌細胞”係指真菌界內的任何類型的真核細胞。真菌界內的種系包括子囊菌門、擔子菌門、芽枝黴門（Blastocladiomycota）、壺菌門、球囊菌門（Glomeromycota）、微孢子蟲目、以及新美鞭菌門（Neocallimastigomycota）。真菌細胞可以包括酵母、黴菌、以及絲狀真菌。在一些實施方式中，該真菌細胞係酵母細胞。

[0973] 如在此所用，術語“酵母細胞”係指子囊菌門和擔子菌門內的任何真菌細胞。酵母細胞可以包括芽殖酵母細胞、裂殖酵母細胞、以及黴菌細胞。在不限於該等生物體的情況下，實驗室和工業環境中使用的許多類型的酵母係子囊菌門的一部分。在一些實施方式中，酵母細胞係釀酒酵母、馬克思克魯維酵母（Kluyveromyces marxianus）、或東方伊薩酵母（Issatchenkia orientalis）細胞。其他酵母細胞可以包括但不限於假絲酵母屬某些種（例如，白色念珠菌）、亞羅酵母屬某些種（例如，亞羅解脂酵母）、畢赤酵母屬某些種（例如，巴斯德畢赤酵母）、克魯維酵母屬某些種（例如，產乳糖酶酵母和馬克思克魯維酵母）、鏈孢黴屬某些種（例如，粗糙脈孢菌）、鐮刀菌某些種（例如，尖孢鐮刀菌）、以及伊薩酵母屬某些種（例如，東方伊薩酵母，又稱為庫德里阿茲威畢赤酵母（Pichia kudriavzevii）和酸性嗜熱假絲酵母（Candida

acidothermophilum））。在一些實施方式中，該真菌細胞係絲狀真菌細胞。如在此所用，術語“絲狀真菌細胞”係指以細絲（即菌絲或菌絲體）生長的任何類型的真菌細胞。絲狀真菌細胞的實例可以包括但不限於，麴黴屬某些種（例如，黑麴黴）、木黴屬某些種（例如，裡氏木黴）、根黴屬某些種（例如，稻根黴菌）、以及被孢黴屬某些種（例如，深黃被孢黴）。

[0974] 在一些實施方式中，該真菌細胞係工業菌株。如在此所用，“工業菌株”係指工業過程中使用的或由工業過程分離的任何真菌細胞菌株，該工業過程例如以商業或工業規模生產產品。工業菌株可以是典型地用於工業過程的真菌種類，或者它可以是指也可用於非工業目的（例如，實驗室研究）的真菌種類分離株。工業過程的實例可以包括發酵（例如，在食品或飲料產品生產中）、蒸餾、生物燃料生產、化合物生產、以及多肽生產。工業菌株的實例可以包括但不限於，JAY270和ATCC4124。

[0975] 在一些實施方式中，該真菌細胞係多倍體細胞。如在此所用，“多倍體”細胞可以是指其基因組以超過一個拷貝存在的任何細胞。多倍體細胞可以是指以多倍體狀態天然發現的細胞類型，或者它可以是指已誘導為以多倍體狀態存在（例如，藉由特異性調節、改變、滅活、啟動、或者減數分裂、胞質分裂或DNA複製的修飾）的細胞。多倍體細胞可以是指其整個基因組為多倍體的細胞，或者它可以是指在特定感興趣的基因組座位中為多倍體的細胞。在不希望受到理論約束的情況下，認為與在單倍體細胞中相比指導RNA的豐度在多倍體細胞的基因組工程中可能是更

常見的速率限制性組分，並且因此使用在此所述的Cpf1 CRISPRS系統的方法可以利用使用某種真菌細胞類型的優點。

[0976] 在一些實施方式中，該真菌細胞係二倍體細胞。如在此所用，“二倍體”細胞可以是指其基因組以兩個拷貝存在的任何細胞。二倍體細胞可以是指以二倍體狀態天然發現的細胞類型，或者它可以是指已誘導為以二倍體狀態存在（例如，藉由特異性調節、改變、滅活、啟動、或者減數分裂、胞質分裂或DNA複製的修飾）的細胞。例如，釀酒酵母菌株S228C可以維持在單倍體狀態或二倍體狀態中。二倍體細胞可以是指其整個基因組為二倍體的細胞，或者它可以是指在特定感興趣的基因組座位中為二倍體的細胞。在一些實施方式中，該真菌細胞係單倍體細胞。如在此所用，“單倍體”細胞可以是指其基因組以一個拷貝存在的任何細胞。單倍體細胞可以是指以單倍體狀態天然發現的細胞類型，或者它可以是指已誘導為以單倍體狀態存在（例如，藉由特異性調節、改變、滅活、啟動、或者減數分裂、胞質分裂或DNA複製的修飾）的細胞。例如，釀酒酵母菌株S228C可以維持在單倍體狀態或二倍體狀態中。單倍體細胞可以是指其整個基因組為單倍體的細胞，或者它可以是指在特定感興趣的基因組座位中為單倍體的細胞。

[0977] 如在此所用，“酵母表現載體”係指含有編碼RNA和/或多肽的一個或多個序列的核酸並且可以進一步含有控制一個或多個核酸表現的任何希望的元件、以及使得能夠在酵母細胞內複製並維持表現載體的任何元件。許多適合的酵母表現載體及其特

徵係本領域已知的；例如，不同載體和技術示出在酵母方案（Yeast Protocols），第2版，肖，W.（Xiao, W.）編輯（胡馬納出版社（Humana Press），紐約，2007）和巴克霍爾茲，R.G.（Buckholz, R.G.）和格利森，M.A.（Gleeson, M.A.）(1991)生物技術(NY) 9(11): 1067-72。酵母載體可以包含但不限於，著絲粒（CEN）序列、自主性複製序列（ARS）、可操作地連接至感興趣的序列或基因的啟動子諸如RNA聚合酶III啟動子、終止子諸如RNA聚合物III終止子、複製起點、以及標記物基因（例如，營養缺陷型、抗生素型或其他選擇標記物）。用於酵母的表現載體的實例可以包括質粒、酵母人工染色體、2μ質粒、酵母整合型質粒、酵母複製型質粒、穿梭載體、以及附加型質粒。

Cpf1 CRISP 系統組分穩定整合在植物和植物細胞的基因組中

[0978] 在特定實施方式中，設想的是引入編碼Cpf1 CRISPR系統組分的多核苷酸，以穩定整合到植物細胞基因組中。在該等實施方式中，轉化載體或表現系統的設計可以根據指導RNA和/或Cpf1基因表現的時間、位置和條件來調整。

[0979] 在特定實施方式中，設想的是將Cpf1 CRISPR系統的組分穩定引入到植物細胞的基因組DNA中。另外地或可替代地，設想的是引入Cpf1 CRISPR系統的組分，以穩定整合到植物細胞器的DNA中，該細胞器諸如但不限於質粒、線粒體或葉綠體。

[0980] 用於穩定整合到植物細胞基因組中的表現系統可以包含以下元件中的一個或多個：可以用於在植物細胞中表現RNA和/或Cpf1酶的啟動子元件；增強表現的5'非翻譯區；在某些細胞

諸如單子葉植物細胞內進一步增強表現的內含子元件；提供了用於插入指導RNA和/或Cpf1基因序列以及其他希望的元件的便利限制性位點的多株位點；以及提供對表現轉錄物的有效終止的3'非翻譯區。

[0981] 表現系統的元件可以是處於一個或多個表現構建體上，該一個或多個表現構建體係環狀的，諸如質粒或轉化載體，或者是非環狀的，諸如線性雙股DNA。

[0982] 在一個特定實施方式中，Cfp1 CRISPR表現系統包含至少：

(a) 編碼與植物中的靶序列雜交的指導RNA（gRNA）的核苷酸序列，並且其中該指導RNA包含指導序列和同向重複序列，以及

(b) 編碼Cpf1蛋白的核苷酸序列，

其中組分 (a) 或 (b) 位於相同或不同構建體上，並且由此不同核苷酸序列可以處於植物細胞內可操作的相同或不同調節元件的控制下。

[0983] 含有Cpf1 CRISPR系統的組分的一個或多個DNA構建體和（在適用情況下）模板序列可以藉由多種常規技術引入到植物、植物部分或植物細胞的基因組中。該過程大體上包括以下步驟：選擇一適合的宿主細胞或宿主組織、將該一個或多個構建體引入到該宿主細胞或宿主組織中、以及由其再生植物細胞或植物。

[0984] 在特定實施方式中，DNA構建體可以使用諸如但不限於電穿孔、微注射、植物細胞原生質體的氣溶膠波束注射來引入到植物細胞中，或者DNA構建體可以使用基因槍方法諸如DNA粒子轟擊來直接引入到植物組織（還參見，付（Fu）等人，轉基因研究（Transgenic Res.）2000年2月；9(1):11-9）。粒子轟擊的基礎係使包覆有感興趣的一種或多種基因的粒子朝向細胞加速，從而導致粒子穿透原生質並且典型地穩定整合到基因組中。（例如，參見克萊因（Klein）等人，自然(1987)；克萊因等人，生物/技術（Bio/Technology）(1992)；卡薩斯（Casas）等人，美國國家科學院院刊(1993).）。

[0985] 在特定實施方式中，含有Cpf1 CRISPR系統組分的DNA構建體可以是藉由土壤桿菌介導的轉化引入到該植物中。該DNA構建體可以與適合T-DNA側翼區組合並且被引入到常規根瘤土壤桿菌宿主載體中。外源DNA可以是藉由感染植物或藉由將植物原生質體與含有一個或多個Ti（根瘤誘導）質粒的土壤桿菌一起培育來結合到植物基因組中。（例如，參見弗拉麗（Fraley）等人，(1985)；羅傑斯（Rogers）等人，(1987)；以及美國專利案號5,563,055）。

植物啟動子

[0986] 為了確保植物細胞內的適當表現，在此描述的Cpf1 CRISPR系統的組分典型地是置於植物啟動子即植物細胞內可操作的啟動子的控制下。設想使用不同類型的啟動子。

[0987] 植物組成型啟動子係能夠表現它在所有或幾乎所有

植物組織中在植物的所有或幾乎所有發育階段過程中控制的開放讀碼框（ORF）的啟動子（稱為“組成型表現”）。組成型啟動子的一個非限制性實例係花椰菜花葉病毒35S啟動子。“調節型啟動子”係指不以組成性方式而是以時間和/或空間調節方式引導基因表現的啟動子並且包括組織特異型啟動子、組織優選型啟動子和誘導型啟動子。不同啟動子可以在不同組織或細胞類型中、或在不同發育階段、或響應於不同環境條件來引導基因表現。在特定實施方式中，一種或多種Cpf1 CRISPR組分在組成型啟動子諸如花椰菜花葉病毒35S啟動子的控制下表現，組織優選型啟動子可以用於靶向在特定植物組織內的某些細胞類型，例如葉或根的維管細胞或種子的特定細胞內的增強的表現。用於Cpf1 CRISPR系統的特定啟動子的實例可見於川又（Kawamata）等人，(1997)植物細胞生理學（Plant Cell Physiol）38:792-803；山本（Yamamoto）等人，(1997)植物雜誌（Plant J）12:255-65；海厄（Hire）等人，(1992)植物分子生物學（Plant Mol Biol）20:207-18；庫斯特（Kuster）等人，(1995)植物分子生物學29:759-72；以及卡帕那（Capana）等人，(1994)植物分子生物學25:681 -91。

[0988] 允許時間空間控制基因編輯或基因表現的誘導型啟動子的實例可以使用能量形式。能量形式可以包括但不限於，聲能、電磁輻射、化學能和/或熱能。誘導型系統的實例包括四環素誘導型啟動子（Tet-On或Tet-Off）、小分子雙雜交轉錄啟動系統（FKBP、ABA等）、或光誘導型系統（光敏色素、LOV結構域或隱花色素），諸如以序列特異性方式引導轉錄活性改變的光誘導型轉錄效應物（LITE）。光誘導型系統的組分可以包括Cpf1 CRISPR酶、

光反應性細胞色素異源二聚體（例如，來自阿拉伯芥）、以及轉錄啟動/阻遏結構域。誘導型DNA結合蛋白及其使用方法的其他實例提供於US 61/736465和US 61/721,283中，該等專利藉由引用以其整體結合在此。

[0989] 在特定實施方式中，暫態表現或誘導型表現可以是藉由使用例如化學品調節啟動子來實現的，即由此外源性化學品的應用誘導基因表現。基因表現的調節可以是藉由化學物阻抑型啟動子來獲得，其中化學物的應用阻遏基因表現。化學品誘導型啟動子包括但不限於，由苯磺醯胺除草劑安全劑啟動的玉米ln2-2啟動子（德·威力德爾（De Veylder）等人，(1997)植物細胞生理學38:568-77）、由用作苗前除草劑的疏水性親電子化合物啟動的玉米GST啟動子（GST-ll-27，WO93/01294）、以及由水楊酸啟動的煙草PR-1啟動子（奧諾（Ono）等人，(2004)生物科學、生物技術和生物化學（Biosci Biotechnol Biochem） 68:803-7）。在此還可以使用藉由抗生素調節的啟動子，諸如四環素誘導型啟動子和四環素阻抑型啟動子（加茨（Gatz）等人，(1991)分子遺傳學和普通遺傳學（Mol Gen Genet） 227:229-37；美國專利案號5,814,618和5,789,156）。

特定植物細胞器中的易位和/表現

[0990] 表現系統可以包含在特定植物細胞器中易位和/或表現的元件。

葉綠體靶向

[0991] 在特定實施方式中，設想的是Cpf1 CRISPR系統用於特別修飾葉綠體基因或者確保葉綠體中表現。出於此目的，使用葉綠體轉化方法或者將Cpf1 CRISPR組分區室化至葉綠體的方法。例如，在質粒基因組中遺傳修飾的引入可以減少生物安全性問題，諸如藉由花粉的基因流。

[0992] 葉綠體轉化方法係本領域已知的並且包括粒子轟擊、PEG處理和微注射。另外，涉及轉化盒從核基因組易位到質粒的方法可以如WO2010061186所述地使用。

[0993] 可替代地，設想的是，將一種或多種Cpf1 CRISPR組分靶向植物葉綠體。這係藉由在表現構建體中結合編碼葉綠體轉運肽（CTP）或質體轉運肽的序列來實現的，該序列可操作地連接至編碼Cpf1蛋白的序列的5'區。在易位到葉綠體中的過程中，在處理步驟中去除CTP。表現蛋白的葉綠體靶向係技術人員已熟知的（例如，參見蛋白質轉運到葉綠體中（Protein Transport into Chloroplasts），2010，植物生物學年評（Annual Review of Plant Biology），第61卷:157-180）。在此類實施方式中，還希望將指導RNA靶向植物葉綠體。例如在US 20040142476中描述了可以用於借助於葉綠體定位序列來將指導RNA易位到葉綠體中的方法和構建體，該專利藉由引用結合在此。構建體的此類變型可以結合到本發明的表現系統中，以有效易位Cpf1-指導RNA。

在藻類細胞中引入編碼 CRISPR-Cpf1 系統的多核苷酸。

[0994] 轉基因藻類（或其他植物諸如芸苔）可以特別適用於生產植物油或生物燃料諸如醇（具體地是甲醇和乙醇）或其他產

品。該等藻類可以被工程化以表現或過量表現用於油或生物燃料工業中的高水平油或醇。

[0995] US 8945839描述了一種用於使用Cas9工程化微藻(萊茵衣藻細胞)種類)的方法。使用類似工具，在此所述的Cpf1 CRISPR系統的方法可以應用於衣藻屬種類和其他藻類上。在特定實施方式中，Cpf1和指導RNA引入使用載體表現的藻類中，該載體在組成型啟動子的控制下表現Cpf1，諸如Hsp70A-Rbc S2或βBeta2-微管蛋白。指導RNA視情況使用含有T7啟動子的載體遞送。可替代地，Cas9 mRNA和體外轉錄的指導RNA可以是遞送至藻類細胞中。電穿孔方法對於技術人員而言是可用的，諸如來自基因領域衣藻屬工程化套組(GeneArt Chlamydomonas Engineering kit)的標準推薦方法。

[0996] 在特定實施方式中，在此使用的內切核酸酶係拆分的Cpf1酶。拆分的Cpf1酶優先用於藻類中以進行靶向基因組修飾，如在WO 2015086795中對於Cas9已描述的。使用Cpf1拆分系統係特別適用於基因組靶向的誘導型方法，並且避免了藻類細胞中Cpf1過量表現的潛在毒性作用。在特定實施方式中，所述Cpf1拆分結構域(RuvC和HNH結構域)可以同時或依次引入到細胞中，以使得所述一個或多個拆分的Cpf1結構域具有藻類細胞中的靶核酸序列。拆分的Cpf1與野生型Cpf1相比減小的尺寸允許使用將CRISPR系統遞送至細胞的其他方法，諸如使用如在此所述的細胞穿透肽。用於生成遺傳修飾性藻類的此方法係特別感興趣的。

在酵母細胞中引入編碼 Cpf1 組分的多核苷酸。

[0997] 在特定實施方式中，本發明涉及使用Cpf1 CRISPR系統進行酵母細胞的基因組編輯。用於轉化可用於引入編碼Cpf1 CRISPR系統組分的多核苷酸的酵母細胞的方法係技術人員已熟知的並藉由河合(Kawai)等人，2010，生物工程缺陷(Bioeng Bugs.)，2010年11月-12月；1(6): 395–403)。非限制性實例包括藉由乙酸鋰處理(可以進一步包括攜帶者DNA和PEG處理)、轟擊或藉由電穿孔來轉化酵母細胞。

在植物和植物細胞中暫態表現 Cpf1 CRISP 系統組分

[0998] 在特定實施方式中，設想的是，在植物細胞中暫態表現指導RNA和/或Cpf1基因。在該等實施方式中，Cpf1 CRISPR系統可以確保僅當指導RNA和Cpf1蛋白二者均存在於細胞中時修飾靶基因，以使得基因組修飾可以得到進一步控制。當Cpf1酶的表現係暫態的時，由此類植物細胞再生的植物典型地不含有外源DNA。在特定實施方式中，Cpf1酶係由植物細胞穩定表現的並且指導序列係暫態表現的。

[0999] 在特定實施方式中，Cpf1 CRISPR系統組分可以是使用植物病毒載體引入在植物細胞中(蘇爾他弗(Scholthof)等人，1996, 植物病理學年度評審(Annu Rev Phytopathol.)1996 ;34:299-323)。在另外的特定實施方式中，所述病毒載體係來自DNA病毒的載體。例如，雙粒病毒組(例如，捲心菜曲葉病毒、豆黃矮病毒、小麥矮化病毒、番茄曲葉病毒、玉米條紋病毒、煙草曲葉病毒或番茄金色花葉病毒)或矮縮病毒組(例如蠶豆壞死黃脈病毒)。在另外的特定實施方式中，所述病毒載體係來自RNA病毒的載體。

例如，煙草脆裂病毒組（例如，煙草擾亂病毒、煙草花葉病毒）、馬鈴薯X病毒組（例如，馬鈴薯X病毒）、或大麥病毒組（例如，大麥條紋花葉病毒）。植物病毒複製基因組係非整合型載體。

[1000] 在特定實施方式中，用於暫態表現Cpf1 CRISPR構建體的載體例如是pEAQ載體，該載體被專門定制用於在原生質體中進行土壤桿菌介導的暫態表現（塞恩思伯裡·F.（Sainsbury F.）等人，植物生物技術雜誌(Plant Biotechnol J.），2009年9月；7(7):682-93）。使用修飾的捲心菜曲葉病毒（CaLCuV）載體證明基因組位置的精確靶向，以在表現CRISPR酶的穩定轉基因植物中表現gRNA（科技報告（Scientific Reports）5，文章編號：14926 (2015), doi:10.1038/srep14926）。

[1001] 在特定實施方式中，編碼指導RNA和/或Cpf1基因的雙股DNA片段可以被暫態引入到植物細胞中。在此類實施方式中，以足夠的量提供引入的雙股DNA片段以修飾細胞，但在預期時間段過去之後或者在一次或多次細胞分裂之後不再持續。用於在植物中直接DNA轉移的方法係技術人員已知的（例如，參見大衛（Davey）等人，植物分子生物學，1989年9月；13(3):273-85。）

[1002] 在其他實施方式中，編碼Cpf1蛋白的RNA多核苷酸被引入到植物細胞中，然後藉由生成足夠量的蛋白質的宿主細胞翻譯並加工以修飾該細胞（在至少一個指導RNA存在下），該引入在預期時間段過去之後或者在一次或多次細胞分裂之後不再持續。用於將mRNA引入到植物原生質體以進行暫態表現的方法係技術人員已知的（例如，參見加利耶（Gallie），植物細胞報告（Plant

Cell Reports）(1993), 13;119-122）。

[1003] 還設想了以上描述的不同方法的組合。

將 Cpf1 CRISPR 組分遞送至植物細胞

[1004] 在特定實施方式中，感興趣的是將Cpf1 CRISPR系統的一種或多種組分直接遞送至植物細胞。這尤其對於生成非轉基因植物係感興趣的（參見下文）。在特定實施方式中，在植物或植物細胞外製備一種或多種Cpf1組分並且將該一種或多種Cpf1組分遞送至細胞。例如，在特定實施方式中，體外製備Cpf1蛋白，之後引入到植物細胞中。Cpf1蛋白可以是藉由熟習該項技術者已知的不同方法來製備並且包括重組產生。在表現之後，Cpf1蛋白被分離，在需要時被折疊，被純化並視情況處理以去除任何純化標籤諸如His-標籤。一旦獲得粗的、部分純化的、或更完全純化的Cpf1蛋白，就可以將該蛋白引入到植物細胞中。

[1005] 在特定實施方式中，該Cpf1蛋白與靶向感興趣的基因的指導RNA混合，以形成預組裝的核糖核蛋白。

[1006] 單獨組分或預組裝核糖核蛋白可以經由電穿孔、藉由用Cpf1相關基因產品包覆的粒子轟擊、藉由化學轉染或藉由轉運穿過細胞膜的一些其他方式來引入到植物細胞中。例如，已證明用預組裝CRISPR核糖核蛋白轉染植物原生質體確保了植物基因組的靶向修飾（如藉由吳（Woo）等人，自然生物技術，2015；DOI: 10.1038/nbt.3389所述的）。

[1007] 在特定實施方式中，Cpf1 CRISPR系統組分係使用奈

米粒子引入到植物細胞中。該等組分，無論是蛋白質或核酸或其組合都可以上載到奈米粒子上或包裝在奈米粒子中並且適用於該等植物（例如像WO 2008042156和US 20130185823所述的）。具體地說，本發明的實施方式包括上載有以下各項或包裝有以下各項的奈米粒子：編碼Cpf1蛋白的一個或多個DNA分子、編碼指導RNA的DNA分子和/或如WO2015089419所述的分離的指導RNA。

[1008] 將Cpf1 CRISPR系統的一種或多種組分引入到植物細胞中的其他方式係藉由使用細胞穿透肽（CPP）。因此，在特定實施方式中，本發明包括含有連接至Cpf1蛋白的細胞穿透肽的組成物。在本發明的特定實施方式中，Cpf1蛋白和/或指導RNA連接一個或多個CPP，以在植物原生質體內有效轉運它們；**還參見** 羅摩克裡希納（Ramakrishna）（20140基因組研究（Genome Res.），2014年6月；24(6):1020-7，人類細胞中的Cas9（Cas9 in human cells））。在其他實施方式中，該Cpf1基因和/或指導RNA係藉由連接至一個或多個CPP以進行植物原生質體遞送的一個或多個環狀或非環狀DNA分子來編碼。該等植物原生質體然後再生成植物細胞並進一步再生成植物。CPP通常被描述為來源於蛋白質或來源於能夠以受體獨立性方式轉運生物分子穿過細胞膜的嵌合序列的小於35個胺基酸的短肽。CPP可以是陽離子肽、具有疏水性序列的肽、兩親性肽、具有脯胺酸富集序列和抗微生物序列的肽、以及嵌合肽或二分肽（普吉（Pooga）和朗格爾（Langel）2005）。CPP能夠穿透生物膜並且同樣引發不同生物分子移動穿過細胞膜進入細胞質並改進其細胞內路線，並且因此有助於生物分子與靶標相互作用。CPP的實例包括其他各項：Tat（它係一種1型HIV進行病

毒複製所需要的核轉錄啟動蛋白）、穿膜肽、卡波西成纖維細胞增長因子（FGF）信號肽序列、整聯蛋白β3信號肽序列；聚精胺酸肽Args序列、富含鳥嘌呤分子轉運體、甜箭頭肽等...

使用Cpf1 CRISPR系統製備遺傳修飾的非轉基因植物

[1009] 在特定實施方式中，在此所述的方法用於修飾內源性基因或修飾其表現而不會永久性引入到任何外源性基因的植物的基因組中，包括編碼CRISPR組分的外源性基因，以便避免在植物基因組中存在外源DNA。這可能是感興趣的，因為非轉基因植物的調節要求較不嚴格。

[1010] 在特定實施方式中，這係藉由Cpf1 CRISPR組分的暫態表現來保證的。在特定實施方式中，一種或多種Cpf1 CRISPR組分係在一種或多種病毒載體上表現的，該一種或多種表現載體產生足夠的Cpf1蛋白和指導RNA，以一致地穩定地確保根據在此所述的方法修飾感興趣的基因。

[1011] 在特定實施方式中，在植物原生質體中確保Cpf1 CRISPR構建體的暫態表現並且因此該構建體並未整合到基因組中。有限表現窗可以足以允許Cpf1 CRISPR系統確保如在此所述的靶基因的修飾。

[1012] 在特定實施方式中，Cpf1 CRISPR系統的不同組分借助於上文所述的粒子遞送分子諸如奈米粒子或CPP分子單獨或混合地引入在植物細胞、原生質體或植物組織中。

[1013] Cpf1 CRISPR組分的表現可以藉由Cpf1核酸酶的直

接活性和視情況引入模板DNA或者藉由修飾使用如在此所述的Cpf1 CRISPR系統靶向的基因來誘導基因組的靶向修飾。上文所述的不同策略允許Cpf1介導的靶向基因組編輯而不需要將Cpf1 CRISPR組分引入到植物基因組中。暫態引入到植物細胞中的組分典型地在雜交時去除。

檢測植物基因組選擇標記物的修飾

[1014] 在特定實施方式中，當方法涉及植物基因組內源性靶基因的修飾時，任何適合的方法可以用於在植物、植物部分或植物細胞用Cpf1 CRISPR系統感染或轉染之後確定基因靶向或靶向誘變是否發生在靶位點。在該方法涉及靶基因引入的情況下，轉化的植物細胞、愈傷組織、組織或植物可以是藉由選擇或篩選存在轉基因或由轉基因編碼的性狀的工程化植物材料來鑒定和分離。物理方法和生物化學方法可以用於鑒定含有插入的基因構建體或內源性DNA修飾的植物或植物細胞轉化株。該等方法包括但不限於：1) 用於檢測並確定重組DNA插入物或修飾的內源性基因的結構的dna印跡分析或PCR擴增；2) 用於檢測並檢查基因構建體的RNA轉錄物的rna印跡、S1 RNA酶保護、引物延伸或逆轉錄酶PCR擴增；3) 用於檢測酶或核糖核酸酶的酶法測定，其中此類基因產物係由基因構建體編碼的或者表現受到遺傳修飾的影響；4) 蛋白質凝膠電泳、西方墨點技術、免疫沈澱反應或酶聯免疫分析，其中基因構建體或內源性基因產物係蛋白質。附加技術諸如原位雜交、酶染色以及免疫染色也可以用於檢測重組構建體的存在或表現或者檢測特定植物器官和組織中的內源性基因修飾。用於完成

所有該等測定的方法係熟習該項技術者熟知的。

[1015] 另外地（或者可替代地），編碼Cpf1 CRISPR組分的表現系統典型地被設計為包含一種或多種選擇標記物或可檢測標記物，該等標記物提供一種分離或有效選擇含有Cpf1 CRISPR系統並且/或者已在早期階段且以大規模地被該系統修飾的細胞的方式。

在土壤桿菌介導的轉化的情況下，標記物盒可以與側接T-DNA邊界相鄰或處於該等邊界之間並且包含在二元載體之內。在另一個實施方式中，標記物盒可以處於T-DNA之外。選擇標記物盒也可以處於與表現盒相同的T-DNA邊界內或與該等邊界相鄰或者可以處於二元載體上的第二T-DNA內的其他位置處（例如，2 T-DNA系統）。

[1016] 對於粒子轟擊或使用原生質體轉化，表現系統可以包含一種或多種分離的線性片段或者可以是含有細菌複製元件、細菌選擇標記物或其他可檢測元件的較大構建體的一部分。包含編碼指導序列和/或Cpf1的多核苷酸的這個或該等表現盒可以是物理連接至標記物盒或者可以是與編碼標記物盒的第二核酸分子混合。標記物盒包含表現允許有效選擇轉化細胞的可檢測標記物或選擇標記物的必需元件。

[1017] 基於選擇標記物的細胞選擇程序將取決於標記物基因的性質。在特定實施方式中，使用選擇標記物，即允許基於標記物的表現直接選擇細胞的標記物。選擇標記物可以賦予陽性選擇或陰性選擇並且對於外部底物的存在係有條件的或沒有條件的（米琪（Miki）等人2004, 107(3): 193–232）。常見地，將抗生素或

除草劑抗性基因用作標記物，由此藉由在含有抑制量的抗生素或除草劑（標記物基因對其賦予抗性）的培養基上生長工程化植物材料來進行選擇。此類細菌的實例係對抗生素諸如潮黴素（hpt）和潮黴素（nptII）賦予抗性的基因和對除草劑諸如草銨膦（bar）和氯磺隆（als）賦予抗性的基因，

[1018] 轉化植物和植物細胞也可以藉由篩選可見標記物，典型地為能夠處理有色底物（例如，β-葡糖醛酸糖苷酶、螢蟲素酶、B或C1基因）的活性來鑒定的。此類選擇和篩選方法係熟習該項技術者所熟知的。

植物培養和再生

[1019] 在特定實施方式中，具有修飾基因組並且藉由在此所述的任何方法產生或獲得的植物細胞可以被培養至再生成具有轉化或修飾表型並因此具有所希望的表型的全株。常規再生技術係熟習該項技術者熟知的。此類再生技術的特定實例依賴於組織培養生長培養基中某些植物激素的操縱，並且典型地依賴於已與所希望的核苷酸序列一起引入的殺生物劑和/或除草劑標記物。在另外的特定實施方式中，植物再生係從培養的原生質體、植物愈傷組織、外植體、器官、花粉、胚胎或其部分獲得的（例如，參見埃文斯（Evans）等人(1983)，植物細胞培養手冊（Handbook of Plant Cell Culture），克萊（Klee）等人(1987)植物生物學年評（Ann.Rev. of Plant Phys.））。

[1020] 在特定實施方式中，如在此所述的轉化或改進的植物可以自體受精以提供本發明的純合改進植物（對於DNA修飾係純

合的）種子，或者可以與非轉基因植物或不同的改進植物雜交以提供純合植物的種子。當重組DNA引入到植物細胞中時，這種雜交所得植物係對於重組DNA分子為雜合的植物。藉由與改進植物雜交並包含遺傳修飾（可以是重組DNA）的此類純合植物和雜合植物在此被稱為“子代”。子代植物係從原始轉基因植物傳代並且含有藉由在此提供的方法引入的基因組修飾或重組DNA分子的植物。可替代地，遺傳修飾植物可以是藉由以上所述方法之一使用Cfp1酶來獲得的，因此無外源DNA結合到該基因組中。藉由進一步育種獲得的此類植物的子代也可以含有遺傳修飾。育種係藉由常用於不同農作物的任何育種方法來進行(例如，阿拉爾(Allard)，植物育種原則（Principles of Plant Breeding），紐約約翰威立國際出版公司（John Wiley & Sons, NY, U. of CA），美國加利福尼亞州大衛斯（Davis, CA）50-98 (1960)。

生成具有增強的農藝性狀的植物

[1021] 在此提供的基於Cpf1的CRISPR系統可以用於引入靶向雙股或單股斷裂並且/或者引入基因活化物和或阻遏物並且（不限於）可以用於基因靶向、基因置換、靶向誘變、靶向缺失或插入、靶向倒位和／或靶向易位。藉由在單個細胞中共表現涉及實現多個修飾的多個靶向RNA，可以確保多重基因組修飾。此技術可以用於高度精確工程化植物以使其具有改進的特徵，包括增強的營養品質、增加的對疾病的抗性和對生物和非生物脅迫的抗性、以及增加的有商業價值的植物產品或異源化合物的產生。

[1022] 在特定實施方式中，如在此所述的Cpf1 CRISPR系統

可以用於在內源性DNA序列中引入靶向雙股斷裂（DSB）。該DSB啟動細胞DNA修復途徑，該修復途徑可以用於實現所希望的斷裂位點附近的DNA序列修飾。當內源性基因失活可以賦予或促成所希望的性狀時，這係感興趣的。在特定實施方式中，在DSB位點處促成使用模板序列的同源重組，以便引入感興趣的基因。

[1023] 在特定實施方式中，Cpf1 CRISPR系統可以用作融合至或可操作地連接至功能結構域以啟動和/或阻遏內源性植物基因的通用核酸結合蛋白。示例性功能結構域可以包括但不限於，翻譯起始區、翻譯活化物、翻譯阻遏物、核酸酶（具體地是核糖核酸酶）、剪接體、珠粒、光誘導型/控制型結構域或化學誘導型/控制型結構域。典型地，在該等實施方式中，該Cpf1蛋白包含至少一個突變，以使得它具有不超過不具有該至少一個突變的Cpf1蛋白的活性的5%的活性；指導RNA包含能夠與靶序列雜交的指導序列。

[1024] 在此所述的方法通常導致生成"改進植物"，在這點上它們與野生型植物相比具有一種或多種希望的性狀。在特定實施方式中，獲得的該等植物、植物細胞或植物部分係包含整合到所有或部分植物細胞的基因組中的內源性DNA序列的轉基因植物。在特定實施方式中，獲得非轉基因遺傳修飾植物、植物部分或細胞，在這點上沒有內源性DNA序列結合到植物的任何植物細胞的基因組中。在此類實施方式中，改進植物係非轉基因的。當僅確保內源性基因的修飾並且在植物基因組中未引入或維持外源性基因時，所得遺傳修飾農作物不含有外源基因並且因此可以基本上

認為是非轉基因的。在下文中更詳細地描述了用於植物基因組編輯的Cpf1 CRISPR系統的不同應用：

a) 引入一種或多種外源基因以賦予一感興趣的農藝性狀

[1025] 本發明提供了基因組編輯或修飾與感興趣的靶座位相關聯的或在該靶座位處的序列之方法，其中該方法包括將Cpf1效應蛋白複合物引入到植物細胞中，由此Cpf1效應蛋白複合物有效地用於將DNA插入物（例如編碼感興趣的外源基因的插入物）整合到植物細胞的基因組中。在較佳的實施方式中，DNA插入物的整合係藉由用外源引入的DNA模板或修復模板進行HR來促成的。典型地，外源引入的DNA模板或修復模板與Cpf1效應蛋白複合物或一組分或用於表現複合物組分的多核苷酸載體一起來遞送。

在此提供的 Cpf1 CRISPR 系統允許靶向基因遞送。已經越來越清楚的是，表現感興趣的基因的效率在很大程度上是由整合到基因組中的位置來確定的。本發明方法允許將外源基因靶向整合到基因組中希望的位置處。該位置可以是基於先前生成事件的資訊來選擇的或者可以藉由在此任何位置揭露的方法來選擇的。

[1026] 在特定實施方式中，在此提供的方法包括 (a) 將包含指導RNA（包含同向重複序列和指導序列）的Cpf1 CRISPR複合物引入到細胞中，其中該指導序列與植物細胞內源性靶序列雜交；(b) 將在指導序列與靶序列雜交時與指導RNA複合並且誘導處於或靠近指導序列所靶向的序列的雙股斷裂的Cpf1效應分子引入到該植物細胞中；並且 (c) 將編碼HDR修復模板的核苷酸序列引入到細胞中，該修復模板編碼感興趣的基因並且作為HDR的結果被

引入到DS斷裂位置中。在特定實施方式中，引入的步驟可以包括將編碼Cpf1效應蛋白、指導RNA和修復模板的一個或多個多核苷酸遞送到植物細胞中。在特定實施方式中，該等多核苷酸係藉由DNA病毒（例如，雙粒病毒組）或RNA病毒（例如，煙草脆裂病毒組）來遞送到細胞中的。在特定實施方式中，引入步驟包括將含有編碼Cpf1效應蛋白、指導RNA和修復模板的一個或多個多核苷酸序列的T-DNA遞送到植物細胞中，其中該遞送係經由土壤桿菌。編碼Cpf1效應蛋白的核酸序列可以是可操作地連接至啟動子，諸如組成型啟動子（例如，花椰菜花葉病毒35S啟動子）或細胞特異型或誘導型啟動子。在特定實施方式中，多核苷酸係藉由微粒轟擊來引入的。在特定實施方式中，該方法進一步包括在引入步驟後篩選植物細胞，以確定是否引入修復模板，即模板基因。在特定實施方式中，該等方法包括由植物細胞再生植物的步驟。在另外的實施方式中，該等方法包括雜交育種該植物以獲得遺傳上希望的植物譜系。編碼感興趣的性狀的外源基因的實例列出在下文中。

b) 編輯內源性基因以賦予感興趣的農藝性狀

[1027] 本發明提供了基因組編輯或修飾與感興趣的靶座位相關聯的或在該靶座位處的序列之方法，其中該方法包括將Cpf1效應蛋白複合物引入到植物細胞中，由此Cpf1效應蛋白複合物修飾植物內源性基因的表現。這可以不同方式來實現。在特定實施方式中，消除內源性基因的表現係希望的並且使用Cpf1 CRISPR複合物靶向並裂解內源性基因，以便修飾基因表現。在該等實施方

式中，在此提供的方法包括 (a) 將包含指導RNA（包含同向重複序列和指導序列）的Cpf1 CRISPR複合物引入到植物細胞中，其中該指導序列與植物細胞基因組的感興趣的基因內的靶序列雜交；並且 (b) 將Cpf1效應分子引入到該細胞中，當結合指導RNA時，該效應蛋白包含與靶序列雜交、確保處於或靠近指導序列所靶向的序列的雙股斷裂的指導序列；在特定實施方式中，引入的步驟可以包括將編碼Cpf1效應蛋白和指導RNA的一個或多個多核苷酸遞送到植物細胞中。

[1028] 在特定實施方式中，該等多核苷酸係藉由DNA病毒（例如，雙粒病毒組）或RNA病毒（例如，煙草脆裂病毒組）來遞送到細胞中的。在特定實施方式中，引入步驟包括將含有編碼Cpf1效應蛋白和指導RNA的一個或多個多核苷酸序列的T-DNA遞送到植物細胞中，其中該遞送係經由土壤桿菌。編碼Cpf1 CRISPR系統組分的多核苷酸序列可以是可操作地連接至啟動子，諸如組成型啟動子（例如，花椰菜花葉病毒35S啟動子）或細胞特異型或誘導型啟動子。在特定實施方式中，多核苷酸係藉由微粒轟擊來引入的。在特定實施方式中，該方法進一步包括在引入步驟後篩選植物細胞，以確定是否修飾感興趣的基因的表現。在特定實施方式中，該等方法包括由植物細胞再生植物的步驟。在另外的實施方式中，該等方法包括雜交育種該植物以獲得遺傳上希望的植物譜系。

[1029] 在以上所述方法的特定實施方式中，抗病性農作物係藉由靶向突變疾病易感性基因或編碼植物防衛基因的負調節物的

基因（例如，Mlo基因）來獲得的。在一特定實施方式中，抗除草劑農作物係藉由靶向取代植物基因諸如編碼乙醯乳酸合酶（ALS）和原卟啉原氧化酶（PPO）的基因的特定核苷酸來生成的。在特定實施方式中，藉由靶向突變編碼非生物脅迫耐受性的負調節物的基因而產生的乾旱耐鹽農作物、藉由靶向突變Waxy基因而產生的低直鏈澱粉穀物、藉由靶向突變糊粉層中的主要脂肪酶基因而產生的具有降低的酸敗性的稻穀或其他穀物等。在特定實施方式中，編碼感興趣的性狀的內源性基因的更廣泛列表列出在下文中。

c) 藉由 Cpf1 CRISPR 系統調節內源性基因以賦予感興趣的農藝性狀

[1030] 在此還提供了用於使用在此提供的Cpf1蛋白調節（即，啟動或阻遏）內源性基因表現的方法。此類方法利用藉由Cpf1複合物靶向植物基因組的一個或多個不同RNA序列。更具體地說，一個或多個不同RNA序列結合兩個或更多個轉接蛋白（例如，適配體），由此每個轉接蛋白與一個或多個功能結構域締合並且其中與該轉接蛋白締合的一個或多個結構域中的至少一個功能結構域具有一種或多種活性，包括修飾烷酶活性、去甲酶活性、轉錄啟動活性、轉錄阻遏活性、轉錄釋放因子活性、組蛋白修飾活性、DNA整合活性、RNA切割活性、DNA切割活性或核酸結合活性。該等功能結構域用於調控內源性植物基因的表現以便獲得所希望的性狀。典型地，在該等實施方式中，該Cpf1蛋白具有一個或多個突變，以使得它具有不超過不具有該至少一個突變的Cpf1效應蛋白的核酸酶活性的5%的核酸酶活性。

[1031] 在特定實施方式中，在此提供的方法包括以下步驟：(a) 將包含指導RNA（包含同向重複序列和指導序列）的Cpf1 CRISPR複合物引入到細胞中，其中該指導序列與植物細胞內源性靶序列雜交；(b) 將在指導序列與靶序列雜交時與指導RNA複合的Cpf1效應分子引入到植物細胞中；並且其中指導RNA被修飾為包含結合功能結構域的不同RNA序列（適配體）和/或Cpf1效應蛋白被修飾為使得它連接至功能結構域。在特定實施方式中，引入的步驟可以包括將編碼（修飾的）Cpf1效應蛋白和（修飾的）指導RNA的一個或多個多核苷酸遞送到植物細胞中。在此任何位置處描述了用於該等方法中的Cpf1 CRISPR系統組分的詳情。

[1032] 在特定實施方式中，該等多核苷酸係藉由DNA病毒（例如，雙粒病毒組）或RNA病毒（例如，煙草脆裂病毒組）來遞送到細胞中的。在特定實施方式中，引入步驟包括將含有編碼Cpf1效應蛋白和指導RNA的一個或多個多核苷酸序列的T-DNA遞送到植物細胞中，其中該遞送係經由土壤桿菌。編碼Cpf1 CRISPR系統的一種或多種組分的核酸序列可以是可操作地連接至啟動子，諸如組成型啟動子（例如，花椰菜花葉病毒35S啟動子）或細胞特異型或誘導型啟動子。在特定實施方式中，多核苷酸係藉由微粒轟擊來引入的。在特定實施方式中，該方法進一步包括在引入步驟後篩選植物細胞，以確定是否修飾感興趣的基因的表現。在特定實施方式中，該等方法包括由植物細胞再生植物的步驟。在另外的實施方式中，該等方法包括雜交育種該植物以獲得遺傳上希望的植物譜系。編碼感興趣的性狀的內源性基因的更廣泛列表列出在下文中。

使用Cpf1修飾多倍體植物

[1033] 許多植物係多倍體，這意味著它們攜帶其基因組的複製拷貝，有時多至六個，像在小麥中。根據本發明利用Cpf1 CRISPR效應蛋白的方法可以"多重"影響基因的所有拷貝或者一次靶向許多基因。例如，在特定實施方式中，本發明的方法用於同時確保不同基因中負責抑制針對疾病的防衛的失功能突變。在特定實施方式中，本發明的方法用於同時抑制小麥植物細胞內TaMLO-A1、TaMLO-B1和TaMLO-D1核酸序列的表現並且由該細胞再生小麥植物，以便確保該小麥植物抵抗白粉病（還參見WO2015109752）。

賦予農藝性狀

[1034] 如上文所述的，在特定實施方式中，本發明涵蓋如在此所述的Cpf1 CRISPR用於插入感興趣的DNA（包括一個或多個植物可表現基因）的用途。在另外的特定實施方式中，本發明涵蓋使用如在此所述的Cpf1系統用於部分或完全缺失一個或多個植物表現基因的方法和工具。在另外的特定實施方式中，本發明涵蓋使用如在此所述的Cpf1系統確保一種或多種植物表現基因藉由突變、驅動、插入一個或多個核苷酸來修飾的方法和工具。在另外的特定實施方式中，本發明涵蓋如在此所述的Cpf1 CRISPR系統確保藉由特定修飾引導一個或多個植物表現基因的表現的一種或多種調節元件來修飾所述基因的表現。

[1035] 在特定實施方式中，本發明涵蓋涉及引入內源性基因和/或靶向內源性基因以及其調節元件的方法，諸如以下列出的基因：

[1036] 1.賦予對害蟲或疾病的抗性的基因：

• 植物疾病抗性基因。植物可以用選殖的抗性基因轉化以工程化對特定病原體菌株具有抗性的植物。參見，例如，鐘斯（Jones）等人，科學 266:789 (1994)（對黃枝孢黴的抗性的番茄 Cf-9 基因的選殖（cloning of the tomato Cf- 9 gene for resistance to Cladosporium fulvum））；馬丁（Martin）等人，科學 262:1432 (1993)（對丁香假單胞菌番茄致病變種的抗性的番茄 Pto 基因編碼蛋白激酶（tomato Pto gene for resistance to Pseudomonas syringae pv. tomato encodes a protein kinase））；麥德瑞諾斯（Mindrinos）等人，細胞 78:1089 (1994)（擬南芥可以是對丁香假單胞菌的抗性的 RSP2 基因（Arabidopsmay be RSP2 gene for resistance to Pseudomonas syringae））。

• 賦予對害蟲諸如大豆囊胞線蟲的抗性的基因。例如，參見 PCT 公開號 96/30517；PCT 申請 WO 93/19181。

• 蘇雲金芽孢桿菌蛋白，例如，參見，熱塞（Geiser）等人，基因 48:109 (1986)。

• 凝集素，例如，參見，凡·達默（Van Damme）等人，植物分子生物學 24:25 (1994。

• 維生素結合蛋白，諸如抗生物素蛋白，參見 PCT 申請 US93/06487，該申請教授了抗生物素蛋白和抗生物素蛋白同源物作為針對害蟲的殺幼蟲劑的用途。

• 酶抑制劑諸如蛋白酶或朊酶抑制劑或澱粉酶抑制劑。參

見，例如，亞伯（Abe）等人，生物化學雜誌 262:16793 (1987)；赫伯(Huub)等人，植物分子生物學 21:985 (1993)，角穀(Sumitani)等人，生物科學、生物技術和生物化學 57:1243 (1993)以及美國專利案號 5,494,813。

- 昆蟲特有的激素或資訊素，諸如蛻皮甾類或保幼激素、或其變體、基於它的模擬物、或其拮抗劑或激動劑。參見,例如，哈莫克（Hammock）等人，自然 344:458 (1990)。

- 昆蟲特有的肽或神經肽，該等肽在表現時破壞受影響害蟲的生理學。例如，雷根（Regan），生物化學雜誌 269:9 (1994)以及普拉特（Pratt）等人，生物化學和生物物理學研究通訊（Biochem.Biophys.Res.Comm.）163:1243 (1989)。還參見美國專利案號 5,266,317。

- 在自然界中由蛇、黃蜂或任何其他生物體產生的昆蟲特有的毒液。例如，參見龐（Pang）等人，基因 116: 165 (1992)。

- 引起單萜、倍半萜、類固醇、異羥肟酸、苯丙素衍生物或具有殺昆蟲活性的另一種非蛋白質分子超積累的酶。

- 涉及生物活性分子修飾（包括翻譯後修飾）的酶；例如，糖解酶、蛋白水解酶、脂解酶、核酸酶、環化酶、轉胺酶、酯酶、水解酶、磷酸酶、激酶、磷酸化酶、聚合酶、彈性蛋白酶、殼多糖酶以及葡聚糖酶，無論是天然還是合成的。參見 PCT 申請 WO93/02197，克雷默（Kramer）等人，昆蟲生物化學與分子生物學（Insect Biochem.Molec.Biol.23:691 (1993)以及卡瓦萊克

（Kawalleck）等人，植物分子生物學（Plant Molec.Biol.）21 :673 (1993)。

• 刺激信號轉導的分子。例如，參見博泰拉（Botella）等人，植物分子生物學 24:757 (1994)以及格裡斯（Griess）等人，植物生理學（Plant Physiol.）104:1467 (1994)。

• 病毒侵入蛋白或源於此的複合物毒素。參見比奇（Beachy）等人，植物病理學年度回顧（Ann. rev. Phytopathol.）28:451 (1990)。

• 在自然界中由病原體或寄生蟲產生的發育阻滯蛋白（Developmental-arrestive protein）。參見蘭布（Lamb）等人，生物/技術 10:1436 (1992)以及杜巴特（Toubart）等人，植物雜誌 2:367 (1992)。

• 在自然界中由植物產生的發育阻滯蛋白。例如，洛格曼（Logemann）等人，生物/技術 10: 305 (1992)。

• 在植物中，病原體常常是宿主特異性的。例如，一些鐮刀菌種類將引起番茄枯萎病但僅攻擊番茄，而其他鐮刀菌種類僅攻擊小麥。植物具有現存和誘導的防衛以抵抗大部分病原體。跨植物各代的突變和重組事件導致引起易感性的遺傳變異性，特別是當病原體以比植物更大頻率繁殖時。在植物中可以存在非宿主抗性，例如宿主和病原體係不相容的或者可以存在針對所有病原體種族的部分抗性，該等抗性典型地是藉由許多基因來控制的，並且/或者也存在對一些病原體種族而不是其他種族的完全抗性。此抗性典型地是藉由幾種基因控制的。使用多種方法和 CRISP-cpf1

系統組分，現在存在在此預先誘導特異性突變的新工具。因此，可以分析抗性基因來源基因組，並且在具有所希望的特徵或形狀的植物中，使用誘導抗性基因增加的方法和 Cpf1 CRISPR 系統組分。本發明系統可以比先前的誘變劑更精確地完成此分析並且因此加速並改進植物育種程式。

[1037] 2.涉及植物疾病的基因，諸如WO 2013046247列出的基因：

- 稻穀病害：稻梨孢、宮部旋孢腔菌、立枯絲核菌、藤倉赤黴；小麥病害：白粉病菌、禾穀鐮刀菌、燕麥鐮刀菌（F. avenaceum）、黃色鐮刀菌、雪黴葉枯菌、條形柄鏽菌、禾柄鏽菌、隱匿柄鏽菌、粉紅雪腐病菌（Micronectriella nivale）、核瑚菌屬某種（Typhula sp.）、小麥黑粉菌、小麥網腥黑穗病菌（Tilletia caries）、小麥基腐病菌、禾生球腔菌、小麥殼多孢、偃麥草核腔菌；大麥病害：白粉病菌（Erysiphe graminis）、禾穀鐮刀菌、燕麥鐮刀菌、黃色鐮刀菌、雪黴葉枯菌、條形柄鏽菌、禾柄鏽菌、大麥柄鏽菌、裸黑粉菌、大麥雲紋斑病菌、圓核腔菌、禾旋孢腔菌、麥類核腔菌、立枯絲核菌：玉米病害：玉米黑粉菌、異旋孢腔菌、高粱膠尾孢、多堆柄鏽菌、玉米灰斑病菌、立枯絲核菌；

- 柑橘病害：柑橘間座殼菌（Diaporthe citri）、柑橘痂囊腔菌（Elsinoe fawcetti）、指狀青黴菌、桔青黴菌（P. italicum）、寄生疫黴、柑橘褐腐疫黴；蘋果病害：蘋果鏈核盤菌（Monilinia mali）、蘋果樹腐爛病菌（Valsa ceratosperma）、蘋果白粉病菌、互隔交鏈孢菌蘋果致病型、蘋果黑星病菌、尖孢炭疽（Colletotrichum

acutatum)、惡疫黴；

- 梨病害：梨黑星病菌（Venturia nashicola）、梨黑星菌（V. pirina）、互隔交鏈孢黴日本梨致病型、梨膠鏽菌（Gymnosporangium haraeanum）、惡疫黴；

- 桃病害：褐腐病菌、嗜果枝孢菌（Cladosporium carpophilum）、擬莖點黴屬某種（Phomopsis sp.）；

- 葡萄病害：痂囊腔菌（Elsinoe ampelina）、檬果炭疽病菌、葡萄白粉菌（Uninula necator）、葡萄銹病菌（Phakopsora ampelopsidis）、葡萄球座菌、葡萄霜黴菌；

- 柿子病害：柿盤長孢（Gloesporium kaki）、柿角斑病菌（Cercospora kaki）、柿葉球腔菌（Mycosphaerela nawae）；

- 瓠果病害：瓜類炭疽菌、黃瓜白粉病菌、甜瓜球腔菌（Mycosphaerella melonis）、尖孢鐮刀菌、黃瓜霜黴病菌、疫黴屬某種、腐黴屬某種；

- 番茄病害：茄鏈格孢菌、番茄葉黴病菌（Cladosporium fulvum）、致病疫黴菌；

- 茄子病害：茄褐紋病菌（Phomopsis vexans）、二孢白粉菌；十字花科蔬菜病害：蘿蔔鏈格孢菌（Alternaria japonica）、白菜白斑病菌（Cercosporella brassicae）、根腫病菌（Plasmodiophora brassicae）、寄生霜黴菌；

- 大蔥病害：蔥柄鏽菌（Puccinia allii）、大蔥霜黴（Peronospora destructor）；

- 大豆病害：大豆紫斑病菌、大豆痂囊腔菌（Elsinoe glycines）、菜豆間座殼大豆變種、大豆殼針孢、大豆尾孢、豆薯層鏽菌、大豆疫黴病菌、立枯絲核菌、棒抱葉斑病菌（Corynespora casiicola）、核盤菌；

- 菜豆病害：菜豆炭疽病菌；

- 花生病害：花生黑斑病菌（Cercospora personata）、花生褐斑病菌、齊整小核菌；

- 豌豆病害豌豆：豌豆白粉菌；

- 馬鈴薯病害：茄鏈格孢菌、致病疫黴菌、馬鈴薯疫黴緋腐病菌、馬鈴薯粉狀瘡痂病菌（Spongospora subterranean, f. sp. Subterranean）；

- 草莓病害：薄草單絲殼菌（Sphaerotheca humuli）、檬果炭疽病菌；

- 茶病害；茶網餅病菌（Exobasidium reticulatum）、茶白星病菌（Elsinoe leucospila）、擬盤多毛孢屬某種、茶炭疽菌（Colletotrichum theae-sinensis）；

- 煙草病害：煙草赤星病菌（Alternaria longipes）、二孢白粉菌、煙草炭疽病菌（Colletotrichum tabacum）、煙草霜黴菌、煙草疫黴菌（Phytophthora nicotianae）；

- 油菜籽病害：核盤菌、立枯絲核菌；

- 棉花病害：立枯絲核菌；

- 甜菜病害：甜菜尾孢菌（Cercospora beticola）、水稻紋枯病菌、螺殼狀絲囊黴（Aphanomyces cochlioides）；

- 玫瑰病害：薔薇雙殼菌（Diplocarpon rosae）、薔薇單絲殼茵（Sphaerotheca pannosa）、薔薇霜黴（Peronospora sparsa）；

- 菊花和菊科病害：萵苣盤枝黴、野菊殼針抱（Septoria chrysanthemi-indici）、堀氏菊柄鏽菌（Puccinia horiana）；

- 各種植物的病害：瓜果腐黴病菌、德巴厘氏腐黴（Pythium debarianum）、禾草腐黴、畸雌腐黴、終極腐黴、灰葡萄孢菌、核盤菌；

- 蘿蔔病害：甘藍鏈格孢；

- 結縷草病害：同果核盤菌、立枯絲核菌；

- 香蕉病害：香蕉黑條葉斑病菌、香蕉黃條葉斑病菌；

- 向日葵病害：向日葵霜黴病菌；

- 在不同植物生長早期階段由以下各項引起的種子疾病或疾病：麴黴屬某些種、青黴屬某些種、鐮刀菌某些種、赤黴菌某些種、木黴屬某些種、根串珠黴屬某些種、根黴屬某些種、毛黴菌某些種、伏革菌屬某些種、莖點黴屬某些種、絲核菌某些種、色二孢屬某些種等；

- 各種植物由桿菌屬某些種、油壺菌屬某些種等介導的病毒病。

[1038] 3.賦予對除草劑的抗性的基因的實例：

- 對抑制生長點或分生組織的除草劑的抗性，該除草劑諸如咪唑啉酮或硫醯脲，例如分別在李等人，歐洲分子生物學學會雜誌 7:1241 (1988)以及米琪等人，理論與應用遺傳學（Theor.Appl.Genet.）80:449 (1990)。

- 草甘膦耐受性（分別由例如突變體 5-烯醇丙酮莽草酸-3-磷酸合酶（EPSP）基因、aroA 基因和草甘膦乙醯轉移酶（GAT）基因賦予的抗性）、或者對於其他膦羧基化合物諸如草銨膦的抗性（由來自鏈黴菌種類的草丁膦乙醯基轉移酶（PAT）基因賦予，該鏈黴菌種類包括吸水鏈黴菌和綠色產色鏈黴菌（Streptomyces viridichromogenes））、以及對吡啶氧基或苯氧基丙酸和環異己酮的抗性（由 ACC 酶抑制劑編碼基因賦予）。參見，例如美國專利案號 4,940,835 和美國專利 6,248,876、美國專利案號 4,769,061、歐洲專利案號 0 333 033 以及美國專利案號 4,975,374。還參見歐洲專利案號 0242246，德格瑞夫（DeGreef）等人，生物/技術 7:61 (1989)，馬歇爾（Marshall）等人，理論與應用遺傳學 83:435 (1992)，卡斯爾（Castle）等人的 WO 2005012515，以及 WO 2005107437。

- 對抑制光合成的除草劑的抗性，該除草劑諸如三咔（psbA 和 gs+基因）或苯基氰（腈水解酶基因），以及麩胱甘肽 s-轉移酶，在瑞茲伯勒（Przibila）等人，植物細胞（Plant Cell）3:169 (1991)，美國專利案號 4,810,648，以及海耶斯等人，生物化學雜誌 285: 173 (1992)。

- 編碼使除草劑去毒的酶或對抑制具有抗性的突變體穀胺醯胺合酶的基因，例如在美國專利申請案序號 11/760,602 中。或

者去毒酶係編碼草丁膦乙醯轉移酶的酶（諸如來自鏈黴菌種類的 bar 或 pat 蛋白）。草丁膦乙醯轉移酶係例如描述於美國專利案號 5,561,236；5,648,477；5,646,024；5,273,894；5,637,489；5,276,268；5,739,082；5,908,810 以及 7,112,665。

• 羥基苯丙酮酸雙氧化酶（HPPD）抑制劑，即天然存在的 HPPD 抗病性酶，或者編碼突變或嵌合 HPPD 酶的基因，如 WO 96/38567、WO 99/24585、以及 WO 99/24586、WO 2009/144079、WO 2002/046387、或美國專利案號 6,768,044 中所述的。

[1039] 4.涉及非生物脅迫耐受性的基因的實例：

• 能夠在植物細胞或植物中減少聚(ADP 核糖)聚合酶（PARP）基因的表現/或活性的轉基因，如 WO 00/04173 或 WO/2006/045633 所述的。

• 能夠減少該等植物或植物細胞的 PARG 編碼基因的表現和/或活性的轉基因，例如在 WO 2004/090140 中。

• 編碼煙醯胺腺嘌呤二核苷酸補救合成途徑的植物功能酶的轉基因，該等酶包括煙醯胺酶、煙醯酸磷酸核糖基轉移酶、煙酸單核苷酸腺嘌呤轉移酶、煙醯胺腺嘌呤二核苷酸合成酶或煙醯胺磷酸核糖基轉移酶，例如在 EP 04077624.7、WO 2006/133827、PCT/EP07/002,433、EP 1999263、或 WO 2007/107326 中所述的。

• 涉及碳水化合物生物合成的酶包括例如 EP 0571427、WO 95/04826、EP 0719338、WO 96/15248、WO 96/19581、WO 96/27674、WO 97/11188、WO 97/26362、WO 97/32985、WO 97/42328、WO

97/44472、WO 97/45545、WO 98/27212、WO 98/40503、WO99/58688、WO 99/58690、WO 99/58654、WO 00/08184、WO 00/08185、WO 00/08175、WO 00/28052、WO 00/77229、WO 01/12782、WO 01/12826、WO 02/101059、WO 03/071860、WO 2004/056999、WO 2005/030942、WO 2005/030941、WO 2005/095632、WO 2005/095617、WO 2005/095619、WO 2005/095618、WO 2005/123927、WO 2006/018319、WO 2006/103107、WO 2006/108702、WO 2007/009823、WO 00/22140、WO 2006/063862、WO 2006/072603、WO 02/034923、EP 06090134.5、EP 06090228.5、EP 06090227.7、EP 07090007.1、EP 07090009.7、WO 01/14569、WO 02/79410、WO 03/33540、WO 2004/078983、WO 01/19975、WO 95/26407、WO 96/34968、WO 98/20145、WO 99/12950、WO 99/66050、WO 99/53072、美國專利案號 6,734,341、WO 00/11192、WO 98/22604、WO 98/32326、WO 01/98509、WO 01/98509、WO 2005/002359、美國專利案號 5,824,790、美國專利案號 6,013,861、WO 94/04693、WO 94/09144、WO 94/11520、WO 95/35026 或 WO 97/20936 中所述的酶，或者如 EP 0663956、WO 96/01904、WO 96/21023、WO 98/39460、以及 WO 99/24593 所揭露的涉及多聚果糖（尤其是菊粉和果聚糖類型）的產生的酶，如 WO 95/31553、US 2002031826、美國專利案號 6,284,479、美國專利案號 5,712,107、WO 97/47806, WO 97/47807, WO 97/47808 以及 WO 00/14249 揭露的涉及 α-1,4-葡聚糖的產生的酶，如 WO 00/73422 所揭露的涉及 α-1,6 分支 α-1,4-葡聚糖的產生的酶，如在例如 WO 00/47727、WO 00/73422、EP 06077301.7、美國專利案號 5,908,975 以及 EP 0728213 所揭露的涉及交替糖的

產生的酶，如例如在 WO 2006/032538、WO 2007/039314、WO 2007/039315、WO 2007/039316、JP 2006304779 以及 WO 2005/012529 所揭露的涉及透明質酸的產生的酶。

- 改進抗旱性的基因。例如，WO 2013122472 揭露了功能性泛素蛋白連接酶蛋白（UPL）、更確切地說是 UPL3 的缺乏或水平降低導致所述植物對水的需求減少或者對乾旱的抗性提高。具有增加的耐旱性的轉基因植物的其他實例揭露於例如 US 2009/0144850、US 2007/0266453、以及 WO 2002/083911。US2009/0144850 描述了由於 DR02 核酸表現的改變而顯示耐旱性表型的植物。US 2007/0266453 描述了由於 DR03 核酸表現的改變而顯示耐旱性表型的植物並且 WO 2002/08391 1 描述了由於保衛細胞中表現的 ABC 轉運體活性降低而具有增加的對乾旱脅迫的耐受性的植物。另一個實例係春日（Kasuga）和合著者(1999)的著作，他們描述了在正常生長條件下編碼 DREB1 A 的 cDNA 在轉基因植物中的過表現啟動了許多脅迫耐受性基因的表現並且導致對乾旱、鹽負荷以及寒冷的耐受性提高。然而，在正常生長條件下 DREB1A 的表現也導致嚴重生長遲緩（春日(1999)自然生物技術 17(3) 287-291）。

[1040] 在另外的特定實施方式中，農作物植物可以是藉由影響特定植物性狀來改進的。例如，藉由開發耐殺蟲劑植物、提高植物的抗病性、提高昆蟲和線蟲抗性、提高植物針對寄生雜草的抗性、提高植物耐旱性、提高植物營養價值、提高植物脅迫耐受性、避免自花授粉、植物飼料可消化性生物質、穀物產量等等。

在下文中提供了若干特定的非限制性實例。

[1041] 除單一基因的靶向突變之外，Cpf1CRISPR 複合物可以被設計允許在植物中靶向突變多基因、缺失染色體片段、位點特異性整合轉基因、體內定點誘變、以及精確基因置換或對偶基因交換。因此，在此所述的方法在基因發現和驗證、突變和順基因育種、以及雜交育種中具有廣泛的應用。該等應用有助於產生新一代的具有各種改進的農藝形狀的遺傳修飾農作物，該等農藝形狀諸如除草劑耐受性、抗病性、非生物脅迫耐受性、高產率、以及優等品質。

使用Cpf1基因創建雄性不育植物

[1042] 雜交植物與自交植物相比典型地具有有利的農藝形狀。然而，對於自花授粉植物，雜種傳代可能是有挑戰的。在不同植物類型中，基因已被修飾，這對於植物能育性，更具體地說雄性能育性係重要的。例如，在玉米中，至少兩種基因已被修飾，這在能育性方面係重要的（關於新植物育種分子技術的阿米塔布莫漢蒂國際會議（Amitabh Mohanty International Conference on New Plant Breeding Molecular Technologies），技術發展與管理，2014年10月9-10日，印度齋蒲爾（Jaipur, India）；斯維塔佘（Svitashev）等人，植物生理學，2015年10月；169(2):931-45；久卡諾維奇（Djukanovic）等人，植物雜誌，2013年12月；76(5):888-99）。在此提供的方法可以用於靶向雄性能育性所需要的基因，以便生成雄性不育植物，該等植物可以易於雜交以生成雜種。在特定實施方式中，在此提供的Cpf1 CRISPR系統用於靶向誘變細胞色

素P450-樣基因（MS26）或大範圍核酸酶基因（MS45），從而向玉米植物賦予雄性不育性。像這樣遺傳改變的玉米植物可以用於雜交育種程式中。

增加植物的能育性階段

[1043] 在特定實施方式中，在此提供的方法用於延長植物諸如稻米植物的能育性階段。例如，可以靶向稻米能育性階段基因諸如Ehd3，以便生成基因中的突變並且可以選擇小植物以延長再生植物能育性階段（如CN 104004782中所述的）

使用Cpf1生成感興趣的農作物的遺傳變異

[1044] 作物植物中野生種質和遺傳變異的可用性係農作物改進程序的關鍵，但是來自作物植物的種質的可用多樣性係有限的。本發明設想了用於生成感興趣的種質的遺傳變異的多樣性的方法。在Cpf1 CRISPR系統的此應用中，提供了靶向植物基因組中的不同位置的指導RNA文庫並且該文庫與Cpf1效應蛋白一起引入到植物細胞中。以這種方式，可以生成基因組規模的點突變和基因敲除的集合。在特定實施方式中，該等方法包括由如此獲得的細胞生成植物部分或植物並且篩選感興趣的性狀的細胞。靶基因可以包含編碼區和非編碼區二者。在特定實施方式中，該性狀係脅迫耐受性的並且該方法係用於生成脅迫耐受性農作物種類。

使用Cpf1影響果實催熟

[1045] 催熟（Ripening）係果實和蔬菜成熟過程中的正常階段。僅在催熟開始後的幾天，催熟致使果實或蔬菜不可食用。此

過程對農民和消費者造成大量損失。在特定實施方式中，本發明的方法可以用於減少乙烯產生。這係藉由確保以下各項中的一種或多種來保證的：a.抑制ACC合酶基因表現。ACC（1-胺基環丙烷-1-羧酸）合酶係負責將S-腺苷甲硫胺酸（SAM）轉化成ACC的酶；這係乙烯生物合成中的第二個步驟至最後一個步驟。當合酶基因的反義（“鏡像”）或截短的拷貝插入到植物基因組中時會阻礙酶表現；b.插入ACC脫胺酶基因。從一種常見的非致病性土壤細菌綠針假單胞菌（Pseudomonas chlororaphis）獲得編碼該酶的基因。它將ACC轉化為一種不同的化合物，從而減少可用於產生乙烯的ACC量；c.插入SAM水解酶基因。此方法類似於ACC脫胺酶，其中當乙烯先質代謝物的量減少時乙烯產生受到阻礙；在此情況下，SAM被轉化為高絲胺酸。從大腸桿菌T3噬菌體獲得編碼該酶的基因，以及d.抑制ACC氧化酶基因表現。ACC氧化酶係催化ACC氧化成乙烯的酶，這係乙烯生物合成途徑中的最後一個步驟。使用在此所述的方法，下調ACC氧化酶基因，導致乙烯產生受到抑制，從而延遲果實催熟。在特定實施方式中，對於以上所述修飾另外地或可替代地，在此所述方法用於修飾乙烯受體，以便干擾由果實獲得的乙烯信號。在特定實施方式中，修飾，更具體地說抑制編碼乙烯結合蛋白的ETR1基因的表現。在特定實施方式中，對於在此所述修飾附加地或可替代地，在此所述的方法用於修飾編碼多聚半乳糖醛酸酶（PG）的基因的表現，該PG係負責分解果膠（維持植物細胞壁完整性的物質）的酶。果膠分解發生在催熟過程開始時，從而導致水果軟化。因此，在特定實施方式中，在此所述的方法用於在PG基因中引入突變或者用於抑制PG基因的啟動，以

便減少所產生的PG酶的量，從而延遲果膠降解。

[1046] 因此在特定實施方式中，該等方法包括使用Cpf1 CRISPR系統確保如上所述的植物細胞基因組的一種或多種修飾並且由該細胞再生植物。在特定實施方式中，該植物係番茄植物。

增加植物的保存期限

[1047] 在特定實施方式中，本發明的方法用於修飾涉及產生影響植物或植物部分的保存期限的化合物的基因。更具體地說，該修飾係在防止馬鈴薯塊莖中的還原糖積累的基因中。在高溫處理時，該等還原糖與游離胺基酸反應，從而產生棕色苦味產物和高水平的丙烯醯胺，該丙烯醯胺係潛在致癌物。在特定實施方式中，在此提供的方法用於減少或抑制液泡轉化酶基因（VInv）的表現，該液泡轉化酶基因編碼將蔗糖分解為葡萄糖和果糖的蛋白質（克拉森（Clasen）等人，DOI: 10.1111/pbi.12370）。

使用Cpf1 CRISPR系統確保增值的性狀

[1048] 在特定實施方式中，Cpf1 CRISPR系統用於產生營養提高的農作物。在特定實施方式中，在此提供的方法適於生成“功能性食品”，即可以提供超過它所含有的傳統營養物的健康益處的修飾的食品或食品成分，並且/或者適於生成“營養食品”，即可以被視為食品或食品的一部分並且提供健康益處（包括預防和治療疾病）的物質。在特定實施方式中，營養食品適用於預防和/或治療癌症、糖尿病、心血管疾病以及高血壓中的一種或多種。

[1049] 營養提高的農作物實例包括（紐厄爾麥格盧林

（Newell-McGloughlin），植物生理學，2008年7月，第147卷，第939–953頁）；

- 修飾蛋白品質、含量和/或胺基酸組成，諸如對於以下各項所描述的：百喜草（盧西亞尼（Luciani）等人，2005，佛羅里達遺傳會議海報（Florida Genetics Conference Poster））、油菜（勒斯勒爾（Roesler）等人，1997, 植物生理學 113 75–81）、玉米（克倫威爾（Cromwell）等人，1967, 1969 農林科學（J Anim Sci） 26 1325–1331，歐·昆（O'Quin）等人，2000 農林科學 78 2144–2149，陽（Yang）等人 2002，轉基因研究 11 11–20，楊（Young）等人 2004，植物雜誌 38 910–922）、馬鈴薯（餘·J（Yu J）和奧（Ao），1997 植物學報（Acta Bot Sin） 39 329–334；查克拉博蒂（Chakraborty）等人 2000，美國國家科學院院刊 97 3724–3729；李等人(2001)中國科學通報（Chin Sci Bull） 46 482–484）、稻米（卡茲伯（Katsube）等人，1999, 植物生理學 120 1063–1074）、大豆（丁金斯（Dinkins）等人，2001，拉普（Rapp），2002，植物體外細胞與發育生物學（In Vitro Cell Dev Biol Plant） 37 742–747）、甘薯（厄尼安（Egnin）和普拉卡什（Prakash）1997，體外細胞與發育生物學（In Vitro Cell Dev Biol） 33 52A）。

- 必需胺基酸含量，諸如對於以下各項所描述的：油菜（法爾科（Falco）等人，1995,生物/技術 13 577–582）、羽扇豆（懷特（White）等人，2001，食品與農業科學雜誌（J Sci Food Agric） 81 147–154）、玉米（萊（Lai）和梅辛（Messing），2002，2008 年農業與生物技術戰略公司轉基因作物資料庫（Agbios 2008 GM

crop database)(2008 年 3 月 11 日))、馬鈴薯(澤(Zeh)等人,2001,植物生理學 127 792–802)、高粱(趙等人,2003,克拉維爾科學出版社(Kluwer Academic Publishers),荷蘭多德雷赫特(Dordrecht, The Netherlands),第 413–416 頁)、大豆(法爾科等人,1995 生物/技術 13 577–582;加利爾(Galili)等人,2002 植物科學的關鍵評論(Crit Rev Plant Sci) 21 167–204)。

- 油類和脂肪酸,例如對於油菜(德赫士(Dehesh)等人,(1996)植物雜誌 9 167–172 [PubMed];德韋基奧(Del Vecchio)(1996)關於脂肪、油類和相關材料的國際新聞通告(INFORM International News on Fats, Oils and Related Materials) 7 230–243;勒斯勒爾等人,(1997)植物生理學 113 75–81 [PMC 免費文獻] [PubMed];弗羅曼(Froman)和於爾森(Ursin)(2002, 2003),美國化學學會論文摘要(Abstracts of Papers of the American Chemical Society) 223 U35;詹姆斯(James)等人,(2003)美國臨床營養學雜誌(Am J Clin Nutr) 77 1140–1145 [PubMed];農業與生物技術戰略公司(2008,同上);棉花(coton)(查普曼(Chapman)等人,(2001).美國石油化學家協會雜誌(J Am Oil Chem Soc) 78 941–947;劉等人,(2002)美國營養學院雜誌(J Am Coll Nutr) 21 205S–211S [PubMed];歐尼爾(O'Neill)(2007)澳大利亞生命科學家(Australian Life Scientist). http://www.biotechnews.com.au/index.php/id;866694817;fp;4;fpid;2 (2008 年 6 月 17 日)、亞麻籽(阿巴迪(Abbadi)等人,2004,植物細胞 16: 2734–2748)、玉米(楊等人,2004,植物雜誌 38 910–922)、油棕(嘉拉尼(Jalani)等人,1997,美國石油化學家協會雜

誌 74 1451–1455；帕爾維姿（Parveez），2003，農業生物科技網（AgBiotechNet） 113 1–8）、稻米（阿奈（Anai）等人，2003，植物細胞報告（Plant Cell Rep） 21 988–992）、大豆（雷迪（Reddy）和湯瑪斯（Thomas），1996, 自然生物技術 14 639–642；金尼（Kinney）和克沃爾頓（Kwolton），1998，黑人學術和專業（Blackie Academic and Professional），倫敦，第 193–213 頁）、向日葵（阿爾卡迪亞（Arcadia），生物科學（Biosciences） 2008）

- 碳水化合物，諸如對於以下各項所描述的果聚糖：菊苣（斯米肯恩（Smeekens）(1997)植物科學趨勢（Trends Plant Sci） 2 286–287，施普倫格（Sprenger）等人，(1997)FEBS 快報 400 355–358，思維尼爾（Sévenier）等人，(1998)自然生物技術 16 843–846）、玉米（菜密（Caimi）等人，(1996)植物生理學 110 355–363）、馬鈴薯（黑爾韋格（Hellwege）等人，1997 植物雜誌 12 1057–1065）、甜菜（斯米肯恩等人，1997，同上）；菊粉，諸如對於馬鈴薯（赫勒韋格（Hellewege）等人，2000, 美國國家科學院院刊 97 8699–8704）所述的；澱粉，諸如對於稻米所述的（施瓦爾（Schwall）等人，(2000) 自然生物技術 18 551–554，蔣（Chiang）等人，(2005) 分子育種（Mol Breed） 15 125–143），

- 維生素類和類葫蘿蔔素，諸如對於以下各項所述的：油菜（新穀（Shintani）和黛拉彭娜（DellaPenna）(1998)科學 282 2098–2100）、玉米（羅徹福德（Rocheford）等人，(2002) .美國營養學院雜誌 21 191S–198S，卡洪（Cahoon）等人，(2003)自然生物技術 21 1082–1087，陳等人，(2003)美國國家科學院院刊 100 3525–

3530)、芥菜籽(休梅克(Shewmaker)等人,(1999) 植物雜誌 20 401–412)、馬鈴薯(杜克勒(Ducreux)等人,2005,實驗植物學雜誌(J Exp Bot) 56 81–89)、稻米(葉(Ye)等人,(2000) 科學 287 303–305)、草莓(阿吉厄斯(Agius)等人,(2003),自然生物技術 21 177–181)、番茄(羅薩蒂(Rosati)等人,(2000) 植物雜誌 24 413–419,弗雷澤(Fraser)等人,(2001) 食品與農業科學雜誌 81 822–827,梅塔(Mehta)等人,(2002) 自然生物技術 20 613–618,迪則·德·拉·加爾薩(Díaz de la Garza)等人,(2004) 美國國家科學院院刊 101 13720–13725,恩非斯(Enfissi)等人,(2005) 植物生物技術雜誌 3 17–27,黛拉彭娜 (2007)美國國家科學院院刊 104 3675–3676。

- 功能性次級代謝產物,諸如對於以下各項描述的:蘋果(芪類,斯贊卡斯基(Szankowski)等人,(2003) 植物細胞報告 22: 141–149)、苜蓿(白藜蘆醇,希普斯金德(Hipskind)和派瓦(Paiva)(2000)分子植物微生物的相互作用(Mol Plant Microbe Interact) 13 551–562)、獼猴桃(白藜蘆醇,小林(Kobayashi)等人,(2000)植物細胞報告 19 904–910)、玉米和大豆(黃酮類,餘等人,(2000)植物生理學 124 781–794)、馬鈴薯(花青素和生物鹼糖苷,魯卡斯瑟維克茨(Lukaszewicz)等人,(2004)農業與食品化學雜誌(J Agric Food Chem) 52 1526–1533)、稻米(黃酮類和白藜蘆醇,斯塔克-勞任森(Stark-Lorenzen)等人,(1997) 植物細胞報告 16 668–673,信(Shin)等人,(2006) 植物生物技術雜誌 4 303–315)、番茄(+白藜蘆醇、綠原酸、黃酮類、芪;羅薩蒂等人,(2000)同上,繆爾(Muir)等人,(2001)自然 19 470–474,

尼吉威戈（Niggeweg）等人，(2004)自然生物技術 22 746–754，焦維納佐(Giovinazzo)等人，(2005)植物生物技術雜誌 3 57–69)、小麥（咖啡酸和咖啡酸、白藜蘆醇；美國合眾國際新聞社（United Press International） (2002)）；以及

- 礦物質可用性，諸如對於以下各項所述的：苜蓿(植酸酶，奧斯丁-菲力浦斯（Austin-Phillips）等人，(1999) http://www.molecularfarming.com/nonmedical.html)、生菜(Lettuse)（鐵，戈托等人，(2000)理論與應用遺傳學（Theor Appl Genet）100 658–664)、稻米（鐵，盧卡（Lucca）等人，(2002)美國營養學院雜誌 21 184S–190S)、玉米、大豆和小麥（植酸酶，德拉卡卡基（Drakakaki）等人，(2005)植物分子生物學 59 869–880，丹茂（Denbow）等人，(1998)養禽科學（Poult Sci） 77 878–881，布林克-佩德森(Brinch-Pedersen)等人，(2000)分子育種 6 195–206)。

[1050] 在特定實施方式中，增值的性狀與存在於植物中的化合物的設想的健康益處相關。例如，在特定實施方式中，藉由應用本發明的方法來確保以下化合物中的一種或多種化合物的合成的修改或者誘導/增加它們的合成，以獲得增值的農作物：

- 類葫蘿蔔素，諸如存在於胡蘿蔔中的 α-胡蘿蔔素，該 α-胡蘿蔔素中和可引起對細胞的損害的自由基；或者存在於各種果實和蔬菜中的 β-胡蘿蔔素，該 β-胡蘿蔔素中和自由基

- 存在於綠色蔬菜中的葉黃素，該葉黃素有助於維持健康視力

- 存在於番茄和番茄產品中的番茄紅素，該番茄紅素認為降低前列腺癌風險

- 存在於柑橘和玉米中的玉米黃素，該玉米黃素有助於維持健康視力

- 膳食纖維，諸如存在於麥麩中的不溶性纖維，該不溶性纖維可以降低乳腺癌和/或結腸癌風險；以及存在於燕麥中的β葡聚糖；存在於車前子（Psylium）和全穀粒中的可溶性纖維，該可溶性纖維可以降低心血管疾病（CVD）風險

- 脂肪酸，諸如ω-3脂肪酸，該等ω-3脂肪酸可以降低CVD風險並提高心理功能和視功能；共軛亞油酸，該共軛亞油酸可以改進身體組成，可以減小某些癌症風險；以及GLA，該GLA可以降低癌症和CVD的炎症風險，可以改進身體組成

- 黃酮類，諸如存在於小麥中的羥基苯乙烯，該等羥基苯乙烯具有抗氧化劑樣活性，可以降低退行性疾病風險；存在於果實和蔬菜中的黃酮醇、兒茶酚類和鞣酸類，該等黃酮醇、兒茶酚類和鞣酸類中和自由基並且可以降低癌症風險

- 葡萄糖異硫氰酸酯、吲哚、異硫氰酸酯，諸如存在於十字花科蔬菜（花椰菜、羽衣甘藍）、辣根屬中的蘿蔔硫素，該蘿蔔硫素中和自由基，可以降低癌症風險

- 酚類，諸如存在於葡萄中的芪類，該等芪類可以降低退行性疾病、心臟病和癌症的風險，可以延年益壽功效；以及存在於蔬菜和柑橘中的咖啡酸和阿魏酸，它們具有抗氧化劑樣活性，可

以降低退行性疾病、心臟病和眼病的風險；以及存在於可可中的表兒茶素，該表兒茶素具有抗氧化劑樣活性，可以降低退行性疾病和心臟病的風險

- 存在於玉米、大豆、小麥以及木制油中的植物甾烷醇/固醇類，它們可以藉由降低血膽固醇水平來降低冠心病的風險

- 存在於洋姜、胡蔥、洋蔥粉中的果聚糖、菊糖、低聚果糖，它們可以提高胃腸道健康

- 存在於大豆中的皂苷類，它們可以降低 LDL 膽固醇

- 存在於大豆中的大豆蛋白質，它可以降低心臟病風險

- 植物雌激素，諸如存在於大豆中的異黃酮，該等異黃酮可以減少絕經期症狀（諸如熱潮紅），可以減少骨質疏鬆症和 CVD；以及存在於亞麻、黑麥和蔬菜中的木脂素類，該等木脂素類可以防止心臟病和一些癌症，可以降低 LDL 膽固醇、總膽固醇

- 硫化物和硫醇類，諸如存在於洋蔥、大蒜、橄欖、韭蔥以及青蔥（scallon）中的二烯丙基硫；以及存在於十字花科蔬菜中的烯丙基甲基三硫、二硫醇硫酮，它們可以降低 LDL 膽固醇，說明維持健康免疫系統

- 鞣酸，諸如存在於蔓越橘、可可中的原花色素，它可以提高泌尿道健康，可以降低 CVD 和高血壓風險

- 等等。

[1051] 此外，本發明的方法還設想了修改蛋白質/澱粉功能

性、保質期、味道/美學、纖維品質、以及減少過敏原、抗營養素以及毒素的形狀。

[1052] 因此，本發明涵蓋了用於產生具有營養增加價值的植物的方法，所述方法包括使用如在此所述的Cpf1 CRISPR系統將編碼涉及產生增加的營養價值的組分的酶的基因引入到植物細胞中並且由所述植物細胞再生植物，所述植物的特徵在於增加的營養價值的所述組分的表現增加。在特定實施方式中，Cpf1 CRISPR系統用於例如藉由修飾控制此化合物代謝的一種或多種轉錄因子來間接修改該等化合物的內源性合成。上文描述了用於將感興趣的基因引入到植物細胞並且/或者使用Cpf1 CRISPR系統修飾內源性基因的方法。

[1053] 在已修飾為賦予增值性狀的植物中的一些特定修飾實例係：例如藉由用硬脂醯-ACP去飽和酶的反義基因轉化植物以增加植物硬脂酸含量的具有修飾的脂肪酸代謝的植物。參見，庫內爾特頓（Knultzon）等人，美國國家科學院院刊89:2624 (1992)。另一個實例涉及例如藉由選殖並且然後再引入與可以負責特徵為低水平植酸的玉米突變體的單一對偶基因相關聯的DNA來減小植酸酯含量。參見瑞博（Raboy）等人，Maydica 35:383 (1990)。

[1054] 類似地，在強啟動子控制下調節玉米糊粉層中黃酮類的產生的玉米（玉蜀黍）Tfs C1和R的表現導致擬南芥屬（Arabidopsis）（阿拉伯芥）中的花色素苷高積累速率，推測係藉由啟動整個途徑（布魯斯（Bruce）等人，2000，植物細胞12:65–80）。黛拉彭娜（韋爾施（Welsch）等人，2007植物生物學年評57:

711–738）發現Tf RAP2.2及其相互作用配位基SINAT2增加擬南芥葉中的胡蘿蔔素形成作用。在轉基因擬南芥中表現Tf Dof1誘導了編碼用於產生碳架、標記性增加胺基酸含量以及減少Glc水平的酶的基因的上調（柳澤（Yanagisawa），2004植物細胞生理學 45: 386–391），並且DOF Tf AtDof1.1（OBP2）上調了擬南芥的葡萄糖異硫氰酸酯生物合成途徑中的所有步驟（希瑞克茲（Skirycz）等人，2006植物雜誌47: 10–24）。

減少植物中的過敏原

[1055] 在特定實施方式中，在此提供的方法可以用於生成具有減少的水平的過敏原的植物，從而使得它們對於消費者而言更安全。在特定實施方式中，該等方法包括修飾負責產生植物過敏原的一種或多種基因的表現。例如，在特定實施方式中，該等方法包括下調植物細胞諸如黑麥草植物細胞中的Lol p5基因的表現並且由該細胞再生植物以便減少所述植物的花粉的過敏原性（巴拉（Bhalla）等人，1999，美國國家科學院院刊，第96卷: 11676–11680）。

[1056] 花生過敏和對豆類過敏總體上是真實而嚴重的健康問題。本發明的Cpf1效應蛋白系統可以用於鑒定並且然後編碼或沈默編碼此類豆類的過敏原性蛋白的基因。在不限於此類基因和蛋白質的情況下，尼柯拉烏（Nicolaou）等人鑒定了花生、大豆、扁豆、羽扇豆、青豆、以及綠豆中的過敏原性蛋白。參見，尼柯拉烏等人，過敏症及臨床免疫學當代觀點（Current Opinion in Allergy and Clinical Immunology）2011;11(3):222)。

用於感興趣的內源性基因的篩選方法

[1057] 在此提供的方法進一步允許鑒定編碼涉及產生增加的營養價值的組分的酶的有價值基因或者通常是影響跨種類、門和植物界的感興趣的農藝性狀的基因。藉由使用如在此所用的Cpf1 CRISPR系統選擇性靶向例如編碼植物代謝途徑的酶的基因，可以鑒定負責植物某些營養方面的基因。類似地，藉由選擇性靶向可以影響所希望的農藝性狀的基因，可以鑒定相關基因。因此，本發明涵蓋了用於編碼涉及產生具有特定營養價值和/或農藝性狀的化合物的酶的基因的篩選方法。

在植物和酵母中進一步應用Cpf1-CRISPR系統

在生物燃料生產中使用Cpf1 CRISPR系統

[1058] 如在此所述的“生物燃料”係由植物和植物來源的資源製成的代用燃料。可再生生物燃料可以從有機物質中提取，該有機物質的能量已藉由碳固定方法來獲得或者藉由使用或轉化生物質來製成。此生物質可以直接用於生物燃料或者可以藉由熱轉化、化學轉化和生物化學轉化來轉化成含有能量的物質。此生物質轉化可以形成固體、液體或氣體形式的燃料。存在兩種類型的生物燃料：生物乙醇和生物柴油。生物乙醇主要係藉由纖維素（澱粉）的糖發酵過程來產生的，該纖維素大部分來源於玉米和甘蔗。在另一個方面中，生物柴油主要係由油料作物諸如油菜籽、棕櫚和大豆產生的。生物燃料主要用於運輸。

增強用於生物燃料生產的植物特性

[1059] 在特定實施方式中，使用在此所述的Cpf1 CRISPR系統的方法用於改變細胞壁的特性，以便促進關鍵性水解劑進入，從而更有效地釋放用於發酵的糖。在特定實施方式中，修改纖維素和/或木質素的生物合成。纖維素係細胞壁的主要組分。纖維素和木質素的生物合成係共調節的。藉由減少植物中的木質素比例，可以增加纖維素的比例。在特定實施方式中，在此所述的方法用於下調植物中的木質素生物合成，以便增加可發酵的碳水化合物。更具體地說，在此所述的方法用於下調選自下組的至少一種第一木質素生物合成基因，該組由以下各項組成：4-香豆酸酯3-羥化酶（C3H）、苯丙胺酸胺裂解酶（PAL）、肉桂酸酯4-羥化酶（C4H）、羥基肉桂醯轉移酶（HCT）、咖啡酸O-甲基轉移酶（COMT）、咖啡醯氧基CoA 3-O-甲基轉移酶（CCoAOMT）、阿魏酸酯5-羥化酶（F5H）、肉桂醇脫氫酶（CAD）、肉桂醯CoA-還原酶（CCR）、4-香豆酸酯-CoA連接酶（4CL）、單木質醇-木質素-特異性糖基轉移酶、以及醛脫氫酶（ALDH），如WO 2008064289 A2所揭露的。

[1060] 在特定實施方式中，在此所述的方法用於產生在發酵過程中生成低水平乙酸的植物生物質（還參見WO 2010096488）。更具體地說，在此揭露的方法用於生成與CaslL同源的突變，以減少多糖乙醯化。

修飾用於生物燃料生產的酵母

[1061] 在特定實施方式中，在此提供的Cpf1酶用於藉由重組微生物進行生物乙醇生產。例如，Cpf1可以用於工程化微生物，諸如酵母，以由可發酵糖類生成生物燃料或生物聚合物並且視情

況能夠降解來源於作為可發酵糖來源的農業廢棄物的植物來源的木質纖維素。更具體地說，本發明提供了多種方法，憑藉該等方法Cpf1 CRISPR複合物用於將生物燃料生產所需要的外源基因引入到微生物中並且/或者修飾可能干擾生物燃料合成的內源性基因。更具體地說，該等方法包括將編碼涉及丙酮酸酯轉化為乙醇或另一種感興趣的產物的酶的一種或多種核苷酸序列引入到微生物諸如酵母中。在特定實施方式中，該等方法確保引入允許微生物降解纖維素的一種或多種酶諸如纖維素酶。在另外的實施方式中，Cpf1 CRISPR複合物用於修飾與生物燃料產生途徑競爭的內源性代謝途徑。

[1062] 因此，在更特定的實施方式中，在此所述的方法用於如下地修飾微生物：

[1063] 以引入至少一種異源核酸或增加編碼植物細胞壁降解酶的至少一種內源性核酸的表現，以使得所述微生物能夠表現所述核酸並且能夠產生並分泌所述植物細胞壁降解酶；

[1064] 以引入至少一種異源核酸或增加編碼將丙酮酸酯轉化為乙醛的酶的至少一種內源性核酸視情況連同編碼將乙醛轉化為乙醇的酶的至少一種異源核酸的表現，以使得所述宿主細胞能夠表現所述核酸；並且/或者

[1065] 以修飾編碼所述宿主細胞的代謝途徑中的酶的至少一種核酸，其中所述途徑產生除來自丙酮酸酯的乙醛或來自乙醛的乙醇之外的代謝物，並且其中所述修飾導致所述代謝物的產生減少，或者以引入編碼所述酶的抑制劑的至少一種核酸。

修飾用於生產植物油或生物燃料的藻類和植物

[1066] 轉基因藻類或其他植物諸如芸苔可以例如特別適用於生產植物油或生物燃料諸如醇（具體地是甲醇和乙醇）。該等藻類可以被工程化以表現或過量表現用於油或生物燃料工業中的高水平油或醇。

[1067] 根據本發明的特定實施方式，Cpf1 CRISPR系統用於生成適用於生物燃料生產的富含脂質的矽藻類。

[1068] 在特定實施方式中，設想的是特異性修飾涉及改變由藻類細胞產生的脂質的量和/或脂質的品質的基因。編碼涉及脂肪酸合成途徑的酶的基因的實例可以編碼具有例如以下活性的蛋白質：乙醯CoA羧化酶、脂肪酸合酶、3-酮乙基_醯基-載體蛋白合酶III、甘油-3-磷酸脫氫酶（G3PDH）、烯醯-醯基載體蛋白還原酶（烯醯-ACP-還原酶）、甘油-3-磷酸醯基轉移酶、溶血磷脂醯基轉移酶或二醯甘油醯基轉移酶、磷脂:二醯甘油二醯甘油、磷脂酸磷酸酶、脂肪酸硫酯酶諸如軟脂醯蛋白硫酯酶、或者蘋果酸酶活性。在另外的實施方式中，設想的是生成具有增加的脂質積累的矽藻類。這可以是藉由靶向減少脂質異化的基因來實現的。對於用於本發明的方法中特別感興趣的是涉及啟動三醯甘油和游離脂肪酸的基因，以及直接涉及脂肪酸的β氧化的基因，諸如醯基-CoA合成酶、3-酮乙基-CoA硫解酶、醯基-CoA氧化酶活性以及磷葡萄糖變位酶。在此所述的Cpf1 CRISPR系統和方法可以用於特異性啟動矽藻類中的此類基因，以便增加其脂質含量。

[1069] 諸如微藻的生物體廣泛用於合成生物學。斯迪維塞克

（Stovicek）等人（代謝工程通信（Metab.Eng. Comm.），2015;2:13描述了工業用酵母諸如釀酒酵母的基因組編輯，以有效產生用於工業生產的有力菌株。斯迪維塞克使用了對酵母密碼子優化的CRISPR-Cas9系統來同時破壞內源性基因的兩個對偶基因並且敲除異源基因。Cas9和gRNA由基因組或附加型2μ基載體位置表現。作者們還證實基因破壞效率可以藉由優化Cas9和gRNA表現水平來提高。拉維瓦（Hlavová）等人（生物技術進展（Biotechnol.Adv.）2015）討論了使用諸如CRISPR的技術靶向核基因和葉綠體基因進行插入誘變和篩選來開發微藻種類或菌株。斯迪維塞克和拉維瓦的方法可以適用於本發明的Cpf1效應蛋白系統。

[1070] US 8945839描述了用於使用Cas9工程化微藻（萊茵衣藻細胞）種類）的方法。使用類似工具，在此所述的Cpf1 CRISPR系統的方法可以應用於衣藻屬種類和其他藻類上。在特定實施方式中，Cpf1和指導RNA引入使用載體表現的藻類中，該載體在組成型啟動子的控制下表現Cpf1，諸如Hsp70A-Rbc S2或βBeta2-微管蛋白。指導RNA將使用含有T7啟動子的載體遞送。可替代地，Cpf1 mRNA和體外轉錄的指導RNA可以是遞送至藻類細胞中。電穿孔方法遵循來自基因領域衣藻屬工程化套組的標準推薦方法。

使用Cpf1生成能夠進行脂肪酸生產的微生物

[1071] 在特定實施方式中，本發明的方法用於生成能夠產生脂肪酸酯諸如脂肪酸甲酯（“FAME”）和脂肪酸乙酯（“FAEE”）的遺傳工程化微生物。

[1072] 典型地，宿主細胞可以被工程化以藉由表現或過表現

編碼硫酯酶的基因、編碼醯基-CoA合酶的基因以及編碼酯合酶的基因來由存在於培養基中的碳源諸如醇產生脂肪酸酯。因此，在此提供的方法用於修飾微生物，以便過表現或引入硫酯酶基因、編碼脂醯CoA合酶的基因、以及編碼酯合酶的基因。在特定實施方式中，硫酯酶基因係選自tesA、'tesA、tesB,fatB、fatB2,fatB3,fatA1、或fatA。在特定實施方式中，編碼醯基-CoA合酶的基因係選自fadDJadK、BH3103、pfl-4354、EAV15023、fadD1、fadD2、RPC_4074、fadDD35、fadDD22、faa39、或編碼具有相同特性的酶的鑒定的基因。在特定實施方式中，編碼酯合酶的基因係編碼來自以下各項的合酶/醯基-CoA :二醯基甘油醯基轉移酶或其變體的基因：霍霍巴、不動桿菌屬某種ADP、泊庫島食烷菌、銅綠假單胞菌、亞德海床桿菌（Fundibacter jadensis）、阿拉伯芥、或真養產鹼桿菌。

另外地或可替代地，在此提供的方法用於減少以下各項中的至少一種基因在所述微生物中的表現：編碼醯基-CoA脫氫酶的基因、編碼外膜蛋白受體的基因、以及編碼脂肪酸生物合成轉錄調節因子的基因。在特定實施方式中，諸如藉由引入突變來滅活該等基因中的一種或多種。

在特定實施方式中，編碼醯基-CoA脫氫酶的基因係fadE。在特定實施方式中，編碼脂肪酸生物合成的轉錄調節因子的基因編碼DNA轉錄阻遏因子，例如fabR。

[1073] 另外地或可替代地，所述微生物被修飾為減少以下各項中的至少一種基因的表現：編碼丙酮酸甲酸裂解酶的基因、編碼乳酸脫氫酶的基因或二者。在特定實施方式中，編碼丙酮酸甲酸裂解酶的基因係pflB。在特定實施方式中，編碼乳酸脫氫酶的基

因係IdhA。在特定實施方式中，諸如藉由在其中引入突變來滅活該等基因中的一種或多種。

[1074] 在特定實施方式中，微生物係選自埃希菌屬、芽孢桿菌屬、乳酸桿菌屬、紅球菌屬、聚球藍細菌屬、集胞藻屬（*Synechoystis*）、假單胞菌屬、麴黴屬、木黴屬、鏈孢黴屬、鐮刀菌屬、腐質黴屬（*Humicola*）、根毛黴屬、克魯維酵母屬、畢赤酵母屬、毛黴菌屬、蝕絲黴屬、青黴菌屬、平革菌屬、側耳屬（*Pleurotus*）、栓菌屬（*Trametes*）、金孢子菌屬、酵母菌屬、寡養單胞菌（*Stenotrophamonas*）、裂殖酵母屬、亞羅酵母屬或鏈黴菌屬。

使用Cpf1生成能夠進行有機酸生產的微生物

[1075] 在此提供的方法進一步用於工程化能夠更具體地說由戊糖或己糖生產有機酸的微生物。在特定實施方式中，該等方法包括將外源性LDH基因引入到微生物中。在特定實施方式中，所述微生物中的有機酸生產另外地或可替代地藉由滅活編碼涉及內源性代謝途徑的蛋白質的內源性基因來增加，該代謝途徑產生除感興趣的有機酸之外的代謝物，並且/或者其中該內源性代謝途徑消耗有機酸。在特定實施方式中，該修飾確保減少除感興趣的有機酸之外的代謝物的產生。根據特定實施方式，該等方法用於引入其中消耗有機酸的內源性途徑或編碼涉及產生除感興趣的有機酸之外的代謝物的內源性途徑的產物的基因的至少一種工程化基因缺失和/或滅活。在特定實施方式中，該至少一種工程化基因缺失或滅活係處於編碼選自下組的酶的一種或多種基因中：丙酮

酸脫羧酶（pdc）、延胡索酸還原酶、醇脫氫酶（adh）、乙醛脫氫酶、磷酸烯醇丙酮酸羧化酶（ppc）、D-乳酸脫氫酶（d-ldh）、L-乳酸脫氫酶（l-ldh）、乳酸2-單加氧酶。

在其他實施方式中，該至少一種工程化基因缺失和/或滅活係處於編碼丙酮酸脫羧酶（pdc）的內源性基因中。

[1076] 在另外的實施方式中，微生物被工程化以產生乳酸，並且該至少一種工程化基因缺失和/或滅活係處於編碼乳酸脫氫酶的內源性基因中。另外地或可替代地，微生物包含至少一種工程化基因缺失或者編碼細胞色素依賴性乳酸脫氫酶諸如細胞色素B2依賴性L-乳酸脫氫酶的內源性基因的滅活。

使用Cpf1生成改進的利用木糖或纖維二糖的酵母菌株

[1077] 在特定實施方式中，Cpf1 CRISPR系統可以用於選擇改進的利用木糖或纖維二糖的酵母菌株。易錯PCR可以用於擴增涉及木糖利用或纖維二糖利用途徑的一種（或多種）基因。涉及木糖利用途徑和纖維二糖利用途徑的基因的實例可以包括但不限於，以下所述的那些：巴，S.J.（Ha, S.J.）等人，(2011)美國國家科學院院刊 108(2):504-9和加拉茲卡，J.M.（Galazka, J.M.）等人，(2010)科學 330(6000):84-6。各自在這種選擇的基因中包含隨機突變的雙股DNA分子的所得文庫可以與Cpf1 CRISPR系統的組分共轉化到酵母菌株（例如S288C）中並且可以選擇具有增加的木糖或纖維二糖利用能力的菌株，如WO2015138855所述的。

使用Cpf1生成用於類異戊二烯生物合成的改進的酵母菌株

[1078] 達塔司傑克奇納思（Tadas Jakočiūnas）等人描述了多種CRISPR/Cas9系統在麵包酵母釀酒酵母的一個轉化步驟中用於基因組工程化多至5個不同基因組座位的成功應用（代謝工程 第28卷，2015年3月，第213–222頁），從而得到具有高甲羥戊酸酯（它係工業上重要的異戊二烯生物合成途徑的關鍵性中間體）產量的菌株。在特定實施方式中，Cpf1 CRISPR系統可以應用於如在此所述的用於鑒定在異戊二烯合成中使用的另外高產的酵母菌株的多種基因組工程化方法中。

使用Cpf1生成產乳酸酵母菌株

[1079] 在另一個實施方式中，涵蓋多種 Cpf1 CRISPR 系統的成功應用。與弗羅茨瓦夫斯迪維塞克（Vratislav Stovicek）等人（代謝工程通訊（Metabolic Engineering Communications），第 2 卷，2015 年 12 月，第 13-22 頁）類似地，改進的產乳酸菌株可以單一轉化事件來設計並獲得。在一個特定實施方式中，Cpf1 CRISPR 系統用於同時插入異源乳酸脫氫酶基因並破壞兩個內源性基因 PDC1 和 PDC5 基因。

在植物中進一步應用Cpf1 CRISPR系統

[1080] 在特定實施方式中，CRISPR系統以及較佳的是在此所述的Cpf1 CRISPR系統可以用於視覺化遺傳元件動力學。例如，CRISPR成像可以視覺化重複或非重複基因組序列，報導端粒長度變化和端粒移動，並且監控整個細胞週期中的基因座位動力學（陳等人，細胞，2013）。該等方法也可以適用於植物。

[1081] CRISPR系統以及較佳的是在此所述的Cpf1 CRISPR系統的其他應用係體外和體內靶向基因破壞陽性選擇篩選（馬利娜（Malina）等人，基因與發育（Genes and Development），2013）。該等方法也可以適用於植物。

[1082] 在特定實施方式中，失活Cpf1內切核酸酶與組蛋白修飾酶的融合可以在複雜的表觀基因組中引入自訂變化（魯斯克（Rusk）等人，自然方法（Nature Methods），2014）。該等方法也可以適用於植物。

[1083] 在特定實施方式中，CRISPR系統以及較佳的是在此所述的Cpf1 CRISPR系統可以用於純化一個特定部分的染色質並且鑒定相關蛋白，從而闡明它們在轉錄中的調節作用（瓦爾錐普（Waldrip）等人，表觀遺傳學（Epigenetics），2014）。該等方法也可以適用於植物。

[1084] 在特定實施方式中，本發明可以用在植物系統中的病毒清除療法，因為它能夠切割病毒DNA和RNA。以前的人類系統研究已證明利用CRISPR靶向丙型肝炎的單股RNA病毒（A.·普裡斯（A. Price）等人，美國國家科學院院刊，2015）以及乙型肝炎的雙股DNA病毒（V.·拉曼那（V. Ramanan）等人，科技報告（Sci.Rep），2015）係成功的。該等方法還可以適於在植物中使用Cpf1 CRISPR系統。

[1085] 在特定實施方式中，本發明可以用於改變基因組複雜度。在另一個特定實施方式中，CRISPR系統以及較佳的是在此所述的Cpf1 CRISPR系統可以用於破壞或改變染色體數目並且生成

僅含有來自一個母體的染色體的單倍體植物。此類植物可以被誘導以經受染色體複製並且被轉化成僅含有純合對偶基因的二倍體植物（卡裡米-阿石提亞尼（Karimi-Ashtiyani）等人，美國國家科學院院刊，2015；安東（Anton）等人，2014）。該等方法也可以適用於植物。

[1086] 在特定實施方式中，在此所述的Cpf1 CRISPR系統可以用於自切割。在該等實施方式中，Cpf1酶和gRNA的啟動子可以是組成型啟動子並且第二gRNA被引入在相同轉化盒中，但受到誘導型啟動子的控制。此第二gRNA可以被設計以誘導Cpf1基因中的位點特異性切割，以便創建非功能性Cpf1。在另一個特定實施方式中，第二gRNA在轉化盒的兩端誘導切割，從而使得該盒從宿主基因組中去除。此系統提供受控的細胞暴露於Cas酶的持續時間並且進一步最小化脫靶編輯。另外，CRISPR/Cas盒兩端的切割可以用於生成具有雙對偶基因突變的無轉基因T0植物（如對於Cas9所描述的，例如，莫耳（Moore）等人，核酸研究，2014；舍費爾（Schaeffer）等人，植物科學，2015）。莫耳等人的方法可以適用於在此所述的Cpf1 CRISPR系統。菅野（Sugano）等人（植物生理學2014年3月；55(3):475-81. doi: 10.1093/pcp/pcu014. 電子版2014年1月18日）報導了CRISPR-Cas9對於地錢屬地錢（Marchantia polymorpha L.）中的靶向誘導的應用，該地錢已做完用於研究陸生植物進化的模型種類。地錢的U6啟動子被鑒定並選殖，以表現gRNA。gRNA的靶序列被設計為破壞編碼地錢中的生長素響應因子1（ARF1）的基因。使用土壤桿菌介導的轉化，菅野等人分類了配子體世代的地錢的穩定突變體。基於CRISPR-Cas9的體內定點

誘變係使用表現Cas9的花椰菜花葉病毒35S或地錢EF1α啟動子來實現的。顯示生長素耐受表型的分離突變體個體係非嵌合的。此外，穩定的突變體係藉由T1植物的無性繁殖來產生的。arf1複對偶基因容易使用基於CRIPSR-Cas9的靶向誘變來建立。菅野等人的方法可以適用於本發明的Cpf1效應蛋白系統。

[1087] 卡巴迪（Kabadi）等人（核酸研究，2014年10月29日；42(19):e147.doi: 10.1093/nar/gku749. 電子版2014年8月13日）開發了由藉由常規金門（Golden Gate）選殖方法結合到載體中的獨立RNA聚合酶III啟動子表現Cas9變體、報導基因和多至四個sgRNA的單一慢病毒系統。每個sgRNA被有效地表現並且可以介導無限增殖細胞和原代人類細胞中的多重基因編輯和持續的轉錄啟動。卡巴迪等人的方法可以適用於本發明的Cpf1效應蛋白系統。

[1088] 林（Ling）等人（BMC植物生物學（BMC Plant Biology）2014，14:327）開發了基於pGreen或pCAMBIA骨架以及gRNA的CRISPR-Cas9二元載體集合。此工具包不需要除BsaI之外的限制酶來在僅僅一個選殖步驟中以高效率生成具有玉米密碼子優化的Cas9和一種或更多種gRNA的最終構建體。此工具包係使用玉米原生質體、轉基因玉米品系和轉基因擬南芥品系來驗證的並且顯示表現出高效率和高特異性。更重要地是，使用此工具包，檢測T1代轉基因幼苗中的三種擬南芥基因的靶向突變。此外，多個基因突變可以由下一代繼承。（指導RNA）模組載體集合，作為用於植物多重基因組編輯的工具包。林等人的工具包可以適用於本發明的Cpf1效應蛋白系統。

[1089] 用於經由CRISPR-Cpf1進行靶向植物基因組編輯的方案基於系列文獻分子生物學方法（ Methods in Molecular Biology，第239-255頁，2015年2月10日的）第1284卷中對於CRISPR-Cas9系統所揭露的那些方法也是可用的。描述了使用阿拉伯芥和本塞姆氏煙草原生質體來設計、構建並評價植物密碼子優化的Cas9（pcoCas9）介導的基因組編輯的雙gRNA的詳細程序。還討論了在全株植物中應用CRISPR-Cas9系統生成靶向基因組修飾的策略。在此章中描述的方案可以適用於本發明的Cpf1效應蛋白系統。

[1090] 彼得森（Petersen）（“朝向精確乙二醇工程化植物（Towards precisely glycol engineered plants）”，植物生物技術丹麥年會（Plant Biotech Denmark Annual meeting）2015，丹麥哥本哈根（Copenhagen, Denmark））開發了使用CRISPR/Cas9工程化擬南芥中的基因組變化以便例如乙二醇工程化擬南芥以用於生產具有所希望的翻譯後修飾的蛋白質和產物的方法。赫布林斯察普（Hebelstrup）等人（植物科學前沿（Front Plant Sci.），2015年4月23日；6:247）列出了在植物中提供澱粉生物工程的農作物，該等農作物表現澱粉修飾酶並且直接產生通常藉由工業化學處理和/或物理處理澱粉來製成的產物。彼得森和赫布林斯察普的方法可以適用於本發明的Cpf1效應蛋白系統。

[1091] 馬（Ma）等人（分子植物（Mol Plant.），2015年8月3日；8(8):1274-84. doi: 10.1016/j.molp.2015.04.007）報導了利用植物密碼子優化的Cas9基因以在單子葉植物和雙子葉植物中方便且

高效地進行多重基因組編輯的穩健CRISPR-Cas9載體系統。馬等人設計了快速生成多個sgRNA表現盒的基於PCR的程序，該等表現盒可以在一輪選殖中藉由金門連接或吉布森（Gibson）組裝來組裝到二元CRISPR-Cas9載體中。使用此系統，馬等人編輯了在具有平均85.4%突變率的稻米中的46個靶位點，該等突變大部分係處於雙對偶基因狀態和純合狀態。馬等人提供了藉由同時靶向多個（多至八個）基因家族成員、生物合成途徑中的多基因或者單一基因中的多個位點來進行T0稻米和T1擬南芥植物中的失功能基因突變的實例。馬等人的方法可以適用於本發明的Cpf1效應蛋白系統。

[1092] 勞德邇（Lowder）等人（植物生理學，2015年8月21日.pii: pp.00636.2015）還開發了能夠在植物中對表現的基因、沈默的基因或非編碼基因進行多重基因組編輯和轉錄調節的CRISPR-Cas9工具包。此工具包為研究者提供了使用金門選殖方法和通路（Gateway）選殖方法快速且有效地組裝單子葉植物和雙子葉植物的功能性CRISPR-Cas9 T-DNA構建體的方案和試劑。它具有一套完整的能力，包括植物內源性基因的多重基因編輯和轉錄啟動或阻遏。基於T-DNA的轉化技術係現代植物生物技術、遺傳學、分子生物學和生理學的基礎。像這樣，申請人開發了用於將Cas9（WT、切口酶或dCas9）和一種或多種gRNA組裝到感興趣的T-DNA目標載體中的方法。組裝方法係基於金門組裝和多位點通路（MultiSite Gateway）重組。對於組裝需要三態模組。第一模組係含有無啟動子Cas9或其側接attL1和attR5位點的衍生物基因的Cas9入門載體。第二模組係含有側接attL5和attL2位點的入門

gRNA表現盒的gRNA入門載體。第三模組包括為Cas9表現提供啟動子選擇的含attR1-attR2目標T-DNA載體。勞德邇等人的工具包可以適用於本發明的Cpf1效應蛋白系統。

[1093] 在一有利實施方式中，該植物可以是樹。本發明還利用在此揭露的CRISPR Cas系統用於草本系統（例如，參見，貝爾哈吉（Belhaj）等人，植物方法（Plant Methods） 9: 39和哈里森（Harrison）等人，基因與發育 28: 1859–1872）。在一特別有利的實施方式中，本發明的CRISPR Cas系統可以靶向樹的單核苷酸多態性（SNP）（例如，參見，週等人，新植物學家（New Phytologist），第208卷，第2期，第298–301頁，2015年10月）。在週等人的研究中，作者們在木質多年生楊樹中使用4-香豆酸酯:CoA連接酶（4CL）基因家族作為個案研究來採用CRISPR Cas系統並且對於所靶向的兩個4CL基因實現100%突變效率，其中每個轉化株檢查攜帶的雙對偶基因修飾。在週等人的研究中，CRISPR-Cas9系統對於單核苷酸多態性（SNP）高度敏感，因為對第三4CL基因的切割因靶序列中的SNP而消除。該等方法可以適用於本發明的Cpf1效應蛋白系統。

[1094] 週等人的方法（新植物學家，第208卷，第2期，第298–301頁，2015年10月）可以如下地適用於本發明。對於CRISPR-Cas9編輯，靶向分別與木質素和黃酮類生物合成相關聯的兩種4CL基因4CL1和4CL2。通常用於轉化的歐洲山楊×銀白楊（Populus tremula × alba）殖株717-1B4係與基因體定序的毛果楊趨異的。因此，由參考基因組設計的4CL1和4CL2 gRNA用內部717 RNA-序列

資料探察，以確保不存在可能限制Cas效率的SNP。還包括對於4CL1的基因組複製物4CL5設計的第三gRNA。相應717序列在PAM附近/內部的每個對偶基因中具有一個SNP，對偶基因二者預期消除了由4CL5-gRNA進行的靶向。所有三個gRNA靶向位點都係位於第一外顯子內。對於717轉化，gRNA由苜蓿屬U6.6啟動子表現，連同在二元載體內的CaMV 35S啟動子控制下表現人類密碼子優化的Cas。使用僅Cas的載體的轉化可以充當對照。隨機選擇的4CL1和4CL2品系經受擴增子定序。然後處理資料並且確認所有情況中的雙對偶基因突變。該等方法可以適用於本發明的Cpf1效應蛋白系統。

[1095] 在植物中，病原體常常是宿主特異性的。例如，尖孢鐮刀菌某種番茄專化型（*Fusarium oxysporum* f. sp. *lycopersici*）引起番茄枯萎病但僅攻擊番茄，而尖孢鐮刀菌某種柄鏽菌小麥專化型（*F. oxysporum f. dianthii Puccinia graminis* f. sp. *tritici*）僅攻擊小麥。植物具有現存和誘導的防衛以抵抗大部分病原體。跨植物各代的突變和重組事件導致引起易感性的遺傳變異性，特別是當病原體以比植物更大頻率繁殖時。在植物中，可以存在非宿主抗性，例如，宿主和病原體係不相容的。還可以存在典型地受到許多基因控制的水平抗性，例如針對所有病原體種族的部分抗性，以及典型地受到幾種基因控制的垂直抗性，例如對一些病原體種族而不是其他種族的競爭性抗性。在基因對基因的水平中，植物和病原體一起進化，並且在一者中的遺傳變化與另一者中的變化平衡。因此，使用自然變異，培育者組合大部分對於產量、品質、均勻性、抵抗力、抗性的可用基因。抗性基因的來源包括天然或

外來品種、祖傳品種、野生植物近緣種、以及誘發突變，例如用誘變劑處理植物材料。使用本發明，為植物育種者提供誘導突變的新工具。因此，熟習該項技術者可以分析抗性基因來源的基因組，並且在具有所希望的特徵或性狀的品種中採用本發明誘導產生抗性基因，這具有比先前的誘變劑更大的精確度，並且因此加速並改進植物育種程式。

改進的植物和酵母細胞

[1096] 本發明還提供了藉由在此所述的方法可獲得並藉由該等方法獲得的植物和酵母細胞。藉由在此所述的方法獲得的改進的植物可以適用於藉由表現確保例如對植物害蟲、除草劑、乾旱、低溫或高溫、過量水等耐受的基因來進行食品或飼料生產。

[1097] 藉由在此所述的方法獲得的改進的植物，具體地是農作物和藻類可以適用於藉由表現例如比野生型中通常所見更高的蛋白質、碳水化合物、營養素或維生素水平來進行食品或飼料生產。就這一點而言，改進的植物，具體地是豆類和塊莖類係較佳的。

[1098] 改進的藻類或其他植物諸如芸苔可以例如特別適用於生產植物油或生物燃料諸如醇（具體地是甲醇和乙醇）。該等藻類可以被工程化以表現或過量表現用於油或生物燃料工業中的高水平油或醇。

[1099] 本發明還提供了改進的植物部分。植物部分包括但不限於，葉、莖、根、塊莖、種子、胚乳、胚珠、以及花粉。如在此

所設想的植物部分可以是有活力的、無活力的、可再生的、和/或不可再生的。

[1100] 在此還涵蓋的是提供根據本發明的方法生成的植物細胞和植物。在本發明的範圍內還包括藉由傳統育種方法產生的含有遺傳修飾的植物的配子、種子、胚胎（合子胚或體細胞胚）、子代或雜種。此類植物可以含有插入在靶序列處或代替靶序列的異源或外源DNA序列。可替代地，此類植物可以僅含有在一個或多個中的變化（突變、缺失、插入、取代）。這樣，此類植物與祖代植物的不同之處僅在於特定修飾的存在。

[1101] 因此，本發明提供了藉由本發明產生的植物、動物或細胞、或其子代。該子代可以是產生的植物或動物的選殖，或者可以由藉由與相同種類的其他個體雜交以使另外希望的性狀滲入其後代來進行的有性繁殖產生。在多細胞生物體（具體的是動物或植物）的情況下，該細胞可以是體內或離體的。

Cpf1效應蛋白複合物可以用於非人類生物體/動物

[1102] 在一個方面中，本發明提供了非人類真核生物體；較佳的是多細胞真核生物體，該等生物體包含根據任何所述實施方式的真核宿主細胞。在其他方面中，本發明提供了真核生物體；較佳的是多細胞真核生物體，該等生物體包含根據任何所述實施方式的真核宿主細胞。在該等方面的一些實施方式中，該生物體可以是動物；例如，哺乳動物。而且，該生物體可以是節肢動物，諸如昆蟲。生物體還可以是植物。另外，生物體可以是真菌。

[1103] 本發明還可以擴展到其他農業應用，例如像農場和生產動物。例如，豬具有許多特徵，該等特徵使得它們作為生物醫學模型係有吸引力的，尤其是在再生醫學中。具體地說，具有重症聯合免疫缺陷（SCID）的豬可以提供用於再生醫學、異種移植（也在此的其他位置討論）以及腫瘤發展的有用模型並且將有助於開發用人類SCID患者的治療。李等人（美國國家科學院院刊，2014年5月20日；111(20):7260-5）利用報導基因指導的轉錄活化物樣效應核酸酶（TALEN）系統，以高效率生成體細胞中的重組啟動基因（RAG）2的靶向修飾，包括影響兩種對偶基因的一些修飾。Cpf1效應蛋白可以適用於類似的系統。

[1104] 李等人的方法（美國國家科學院院刊，2014年5月20日；111(20):7260-5）可以與如下類似地適用於本發明。突變的豬係藉由靶向修飾胎兒成纖維細胞中的RAG2，隨後進行SCNT和胚胎轉移來產生的。編碼CRISPR Cas和報導基因的構建體被電穿孔到胎兒來源的成纖維細胞中。在48 h後，表現綠色螢光蛋白的轉染細胞以估計每孔一個單一細胞的稀釋分到96孔板的單個孔中。RAG2的靶向修飾係藉由擴增側接任何CRISPR Cas切割位點的基因組DNA片段隨後對PCR產物進行定序來篩選的。在篩選並確保不存在位點外突變之後，將攜帶RAG2的靶向修飾的細胞用於SCNT。去除極體連同卵母細胞的一部分相鄰細胞質（推測含有中期II板），並且使供體細胞置於卵黃周隙中。然後電穿孔重構的胚胎，以將供體細胞與卵母細胞融合，並且然後化學啟動。將啟動的胚胎在具有0.5 μM Scriptaid（S7817；西格馬阿德里奇公司）豬受精卵培養基（Porcine Zygote Medium） 3（PZM3）中培養14–16

h。然後洗滌胚胎以去除Scriptaid並且在PZM3中培養，直到它們轉移到代孕豬的輸卵管為止。

[1105] 本發明還可應用於修飾其他動物諸如牛的SNP。塔恩（Tan）等人（美國國家科學院院刊，2013年10月8日；110(41): 16526–16531）使用質粒、rAAV和寡核苷酸模板擴增家畜基因編輯工具包，以包括轉錄活化物樣（TAL）效應核酸酶（TALEN）和成簇規律間隔短迴文重複序列（CRISPR）/Cas9-刺激性同源定向修復（HDR）。根據他們的方法將基因特異性gRNA序列選殖到丘奇實驗室（Church lab）gRNA載體（Addgene ID: 41824）中（馬里·P（Mali P）等人，(2013) 經由Cas9進行RNA指導的人類基因組工程化（RNA-Guided Human Genome Engineering via Cas9）。科學339(6121):823-826）。Cas9核酸酶係藉由共轉染hCas9質粒（Addgene ID: 41815）或由RCIScript-hCas9合成的mRNA來提供的。此RCIScript-hCas9係藉由將來自hCas9質粒（涵蓋hCas9 cDNA）的XbaI-AgeI片段亞選殖到RCIScript質粒中來構建。

[1106] 霍（Heo）等人（幹細胞與發育（Stem Cells Dev.），2015年2月1日；24(3):393-402. doi: 10.1089/scd.2014.0278. 電子版2014年11月3日）報導了在牛基因組中使用牛多能細胞和成簇規律間隔短迴文重複序列（CRISPR）/Cas9核酸酶的高效基因靶向。首先，霍等人藉由異位表現山中因子（yamanaka factor）並且進行GSK3β和MEK抑制劑（2i）處理來由牛體成纖維細胞生成誘導的多能幹細胞（iPSC）。霍等人觀察到，該等牛iPSC在畸胎瘤的基因表現和發育潛力方面高度類似於天然多能幹細胞。此外，對於牛

NANOG座位特異的CRISPR-Cas9核酸酶在牛iPSC和胚胎的牛基因組中顯示高度有效的編輯。

[1107] Igenity®提供了對諸如牛的動物執行並傳播經濟上重要的經濟性狀的性狀的譜圖分析，該等性狀諸如胴體組成、胴體質量、母體和繁殖性狀以及平均日增重。綜合性Igenity®譜圖的分析以DNA標記物（最常是單核苷酸多態性或SNP）的發現開始。在Igenity®譜圖之後的所有標記物係藉由科研機構的獨立科學家發現的，該等研究機構包括大學、研究團體以及政府機構諸如USDA。然後在驗證群體中分析Igenity®的標記物。Igenity®使用代表各種生產環境和生物類型的多種資源種群，通常與來自牛肉產業的種畜、母犢牛、飼育場和/或包裝部門的行業夥伴一起工作，以收集不能普遍獲得的表型。牛基因組資料庫係廣泛可用的，例如，參見NAGRP牛基因組協調程序（http://www.animalgenome.org/cattle/maps/db.html）。因此，本發明可以適用於靶向牛SNP。熟習該項技術者可以利用用於靶向SNP的以上方案並且將它們應用於牛SNP，例如，如塔恩等人或霍等人所述的。

[1108] 清澗·鄒（Qingjian Zou）等人（分子細胞生物學雜誌（Journal of Molecular Cell Biology），在2015年10月12日線上先行公開）證明藉由靶向狗肌生成抑制蛋白（MSTN）基因的第一外顯子（骨骼肌質量的負調節物）增加狗的肌肉質量。首先，藉由將sgRNA靶向的MST與Cas9載體共轉染到犬胚胎成纖維細胞（CEF）來驗證sgRNA的效率。之後，藉由微注射具有正常形態學的胚胎

與Cas9 mRNA和MSTN sgRNA的混合物並且將受精卵自身移植到同一母狗的輸卵管來生成MSTN KO狗。與其野生型同窩出生母狗相比，敲除小狗在大腿上顯示明顯的肌肉表型。這也可以使用在此提供的Cpf1 CRISPR系統來進行。

家畜—豬

[1109] 在一些實施方式中，家畜中的病毒靶標可以包括豬CD163，例如在豬巨噬細胞上。CD163與PRRSv（豬繁殖與呼吸綜合症病毒，它係一種動脈炎病毒）的感染（認為是藉由病毒細胞侵入）相關聯。PRRSv的感染，特別是對豬肺泡巨噬細胞（可見於肺中）的感染導致先前不能治癒的豬綜合症（“神秘病”或“藍耳病”），從而使得家豬遭受（包括）生殖障礙、體重減輕和高死亡率。常常可見機會性感染諸如流行性肺炎、腦膜炎和耳腫脹，這係因為藉由巨噬細胞活性喪失會引起免疫缺陷。由於抗生素使用的增加和經濟損失（估計每年660百萬美元），這也具有重大的經濟和環境影響。

[1110] 如密蘇裡大學（University of Missouri）的克莉絲汀·W·惠特沃思和蘭德爾·普萊瑟（Randall Prather）博士等人與Genus公司合作(自然生物技術 3434 ,2015年12月07日線上公開）報導的 ,使用CRISPR-Cas9靶向CD163並且編輯的豬的後代當暴露於PRRSv時是有抗性的。在CD163的外顯子7中均具有突變的一個雄性起始者和一種雌性起始者二者繁殖產生後代。雄性起始者在對偶基因的外顯子7中具有11-bp的缺失，這導致移碼突變以及結構域5中的胺基酸45的錯義翻譯和胺基酸64處的後一個提前終止

密碼子。另一個對偶基因具有外顯子7中的2-bp添加和前述內含子中的377-bp缺失，這被預測為引起結構域5的前49個胺基酸的表現，隨後是在胺基酸85處的提前終止密碼子。母豬在一個對偶基因中具有7 bp添加，該添加在翻譯時預測表現結構域5的前48個胺基酸，隨後是在胺基酸70處的提前終止密碼子。母豬的另一個對偶基因係不可擴增的。預測選定的後代係無效突變動物（CD163–/–），即CD163敲除。

[1111] 因此，在一些實施方式中，豬肺泡巨噬細胞可以被CRISPR蛋白靶向。在一些實施方式中，豬CD163可以被CRISPR蛋白靶向。在一些實施方式中，豬CD163可以藉由誘導DSB或藉由插入或缺失來敲除，例如靶向外顯子7的缺失或修飾，包括以上所述的那些缺失或修飾中的一種或多種，或者在該基因的其他區域中，例如外顯子5的缺失或修飾。

[1112] 還設想了編輯的豬及其子代，例如CD163敲除豬。這可以是出於家畜、育種或建模目的（即，豬模型）。還提供了包含基因敲除的精液。

[1113] CD163係清道夫受體富含半胱胺酸（SRCR）超家族的成員。基於體外研究，蛋白質的SRCR結構域5係負責啟封和釋放病毒基因組的結構域。這樣，也可以靶向SRCR超家族的其他成員，以便評定對其他病毒的抗性。PRRSV也是哺乳動物動脈炎病毒組的成員，該病毒組還包括鼠科乳酸脫氫酶病毒、猴出血熱病毒以及馬動脈炎病毒。該等動脈炎病毒享有重要的發病機理特徵，包括巨噬細胞向性和引起嚴重疾病和持續感染二者的能力。因此，

動脈炎病毒以及具體地乳酸脫氫酶病毒、猴出血熱病毒和馬動脈炎病毒可以例如藉由豬CD163或其在其他種類中的同源物，並且還提供鼠科、猴和馬的模型以及敲除。

[1114] 實際上，此方法可以擴展到引起其他家畜疾病且可以傳播到人類的病毒或細菌，諸如豬流感病毒（SIV）菌株，包括丙型流感和稱為H1N1、H1N2、H2N1、H3N1、H3N2以及H2N3的甲型流感亞型，以及以上提及的肺炎、腦膜炎和水腫。

使用RNA-指導的Cpf1效應蛋白複合物進行治療性靶向

[1115] 如將清楚的，設想的是本發明系統可以用於靶向任何感興趣的多核苷酸序列。本發明提供了非天然存在的或工程化的組成物、或編碼所述組成物的組分的一種或多種多核苷酸、或含有編碼所述組成物的組分一種或多種多核苷酸的載體或遞送系統，其用於體內、離體或體外修飾靶細胞並且該修飾可以改變細胞使得一旦修飾，CRISPR修飾細胞的子代或細胞系保留改變的表型的方式實施。該等修飾的細胞和子代可以是多細胞生物體的一部分，諸如在將CRISPR系統應用於所希望的細胞類型的情況下的植物或動物。CRISPR發明可以是治療性治療方法。治療性治療方法可以包括基因或基因組編輯，或基因治療。

治療病原體，如細菌、真菌和寄生蟲病原體

[1116] 本發明還可以適用於治療細菌、真菌和寄生蟲病原體。大部分研究工作集中於開發新的抗生素，然而一旦開發，就會經受相同的抗藥性問題。本發明提供了克服該等困難的新型基於

CRISPR的替代方案。另外，與現存的抗生素不同，基於CRISPR的治療可以是製成病原體特異性的，從而誘導靶病原體的細菌細胞死亡同時避免有益細菌死亡。

[1117] 蔣（Jiang）等人（“使用CRISPR-Cas對細菌基因組進行RNA指導編輯（RNA-guided editing of bacterial genomes using CRISPR-Cas systems）”，自然生物技術，第31卷，第233-9頁，2013年3月）使用CRISPR-Cas9系統來突變或殺死肺炎鏈球菌或大腸桿菌。將精確突變引入到基因組中的工作依賴於靶基因組位點處的雙-RNA：Cas9-引導的切割以殺死未突變細胞，並且不再需要選擇標記物或反選擇系統。CRISPR系統已用於逆轉抗生素抗性並且消除各菌株之間的抗性轉移。比克考爾德（Bickard）等人證實重編程式以靶向致病基因的Cas9殺死了致命的金黃色葡萄球菌而不是無毒的金黃色葡萄球菌。將核酸酶重編程式以靶向抗生素抗性基因，破壞了具有抗生素抗性基因的葡萄球菌質粒並且針對質粒攜帶的抗性基因的傳播進行免疫。（參見，比卡爾德（Bikard）等人，“探索產生序列特異性抗微生物劑的CRISPR-Cas核酸酶（Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials）”，自然生物技術，第32卷，1146–1150, doi:10.1038/nbt.3043，2014年10月05日線上公開）。比卡爾德證實CRISPR-Cas9抗微生物劑在體內起到殺死老鼠皮膚定位模型中的金黃色葡萄球菌的作用。類似地，優素福（Yosef）等人使用CRISPR系統來靶向編碼賦予β內醯胺類抗生素抗性的酶的基因（參見優素福等人，“程式設計以敏化並殺死抗生素抗性細菌的溫和和烈性噬菌體（Temperate and lytic bacteriophages programmed to sensitize

and kill antibiotic-resistant bacteria）”，美國國家科學院院刊，第112卷，第7267–7272頁，doi: 10.1073/pnas.1500107112，2015年5月18日線上公開）。

[1118] CRISPR系統可以用於編輯對其他遺傳方法具有抗性的寄生蟲基因組。例如，已顯示CRISPR-Cas9系統能將雙股斷裂引入到*約氏瘧原蟲*基因組（參見，張等人，“使用CRISPR/Cas9系統有效編輯瘧原蟲基因組（Efficient Editing of Malaria Parasite Genome Using the CRISPR/Cas9 System）”，mBio.，第5卷，e01414-14，2014年7月-8月）。古爾巴爾（Ghorbal）等人（“使用CRISPR-Cas9系統在人類瘧原蟲鐮狀瘧原蟲中進行基因組編輯（Genome editing in the human malaria parasite Plasmodium falciparumusing the CRISPR-Cas9 system）”，自然生物技術，第32卷，第819-821頁，doi: 10.1038/nbt.2925，2014年6月1日線上公開）修飾了兩種基因*orc1*和*kelch13*的序列，這兩種基因分別在基因沈默和形成對青蒿素的抗性中具有推定的作用。儘管對於該修飾沒有直接選擇，但是在適當位點改變的寄生蟲以極高效率恢復，這表明使用此系統可以生成中性突變或甚至有害突變。CRISPR-Cas9還用於修飾其他致病性寄生蟲的基因組，包括*剛地弓形蟲（參見沈等人，“使用CRISPR/CAS9在剛地弓形蟲的不同菌株中進行有效基因破壞（Efficient gene disruption in diverse strains of Toxoplasma gondii using CRISPR/CAS9）”，mBio，第5卷:e01114-14, 2014；以及西迪克（Sidik）等人*，“使用CRISPR/Cas9進行剛地弓形蟲的有效基因組工程化（Efficient Genome Engineering of *Toxoplasma gondii* Using CRISPR/Cas9）”，公共科學圖書館綜合，第9卷，e100450,

doi: 10.1371/journal.pone.0100450，2014年6月27日線上公開）。

[1119] 維亞斯（Vyas）等人（“白色念珠菌CRISPR系統允許對必需基因和基因家族進行遺傳工程化（A *Candida albicans* CRISPR system permits genetic engineering of essential genes and gene families）”，科學進展（Science Advances），第1卷，e1500248, DOI: 10.1126/sciadv.1500248，2015年4月3日）採用CRISPR系統來克服白色念珠菌中長期存在的遺傳工程化障礙並且在幾種不同基因的兩個拷貝的單一實驗中進行有效突變。在其他幾種機制促成抗藥性的有機體中，維亞斯產生不再顯示母體臨床分離物Can90顯示的對氟康唑或放線菌酮的抗性的純合雙突變體。維亞斯還藉由創建條件對偶基因來獲得*白色念珠菌*的必需基因中的純合的失功能突變。對於核糖體RNA加工所需要的*DCR1*無效對偶基因在低溫下是致命的，但是在高溫下是可存活的。維亞斯使用引入無效突變的修復模板和不能在16°C下生長的分離的*dcr1/dcr1*突變體。

[1120] 本發明的CRISPR系統藉由破壞染色體座位來用於鐮狀瘧原蟲 中。古爾巴爾等人（“使用CRISPR-Cas9系統在人類瘧原蟲鐮狀瘧原蟲中進行基因組編輯”，自然生物技術，32, 819-821 (2014), DOI: 10.1038/nbt.2925，2014年6月1日）採用CRISPR系統來在瘧疾基因組中引入特異性基因敲除和單一核苷酸取代。為了使CRISPR-Cas9系統適於 鐮狀瘧原蟲，古爾巴爾等人生成在也攜帶藥物選擇標記物*ydhodh*的pUF1-Cas9附加體中的瘧原蟲調節元件控制下並且用於轉錄sgRNA的表現載體，該附加體給予對鐮狀瘧原蟲 二氫乳清酸脫氫酶（PfDHODH）抑制劑DSM1的抗性，使

用 鐮狀瘧原蟲 U6小核（sn）RNA調節元件，將用於同源重組修復的指導RNA和供體DNA模板置於相同質粒pL7上。還參見，張·C.等人（“使用CRISPR/Cas9系統有效編輯瘧原蟲基因組（Efficient editing of malaria parasite genome using the CRISPR/Cas9 system）”，MBio，2014年7月1日；5(4):E01414-14, doi: 10.1128/MbIO.01414-14）和瓦格納等人（“鐮狀瘧原蟲中的有效CRISPR-Cas9介導的基因組編輯（Efficient CRISPR-Cas9-mediated genome editing in *Plasmodium falciparum*）”，自然方法 11, 915-918 (2014), DOI: 10.1038/nmeth.3063）。

治療病原體，如病毒病原體，諸如HIV

[1121] Cas-介導的基因組編輯可以用於在軀體組織中引入保護性突變，以對抗非遺傳性疾病或複雜疾病。例如，淋巴細胞中NHEJ-介導的CCR5受體滅活（隆巴爾多（Lombardo）等人，自然生物技術，2007年11月；25(11):1298-306）可以是用於避免HIV感染的可行策略，而PCSK9（科恩（Cohen）等人，自然遺傳學，2005年2月；37(2):161-5）或血管生成素（木蘇努魯（Musunuru）等人，新英格蘭醫學雜誌，2010年12月2日；363(23):2220-7）的缺失可以提供針對具有他汀類抗性的血膽固醇過多或高血脂症的治療作用。儘管該等靶標也可以使用siRNA介導的蛋白質敲低來解決，但是NHEJ介導的基因失活優點係實現永久性治療益處而不需要持續治療的能夠。正如所有基因治療一樣，確定每個提出的治療性用途具有有利的益處-危險比率當然是重要的。

[1122] 將編碼Cas9和指導RNA的質粒DNA連同修復模板一

起流體動力學遞送到酪胺酸血症成年小鼠模型的肝臟中，已顯示此遞送能夠校正突變體Fah基因並且在約250分之一個細胞中挽救野生型Fah蛋白質的表現（自然生物技術，2014年6月；32(6):551-3）。另外，臨床試驗成功地使用ZF核酸酶，藉由離體敲除CCR5受體來對抗HIV感染。在所有患者中，HIV DNA水平降低，並且在四分之一的患者中，HIV RNA變得不可檢測（特巴斯（Tebas）等人，新英格蘭醫學雜誌，2014年3月6日；370(10):901-10）。這兩種結果都表明了可程式設計核酸酶作為一種新治療平臺的希望。

[1123] 在另一個實施方式中，自滅活慢病毒載體可以用於並且/或者適於本發明的CRISPR-Cas系統，該自滅活慢病毒載體具有靶向由HIV tat/rev共用的共有外顯子的siRNA、核仁定位TAR誘餌、和抗CCR5特異性錘頭狀核酶（例如，參見，迪吉斯托等人 (2010) 科學轉化醫學 2:36ra43）。可以收集最少2.5 × 10^6個CD34+細胞/每千克患者體重並且以2×10^6個細胞/ml的密度在X-VIVO 15培養基（龍沙公司）中預刺激16至20小時，該培養基含有2 μmol/L-穀胺醯胺、幹細胞因子（100 ng/ml）、Flt-3配位基（Flt-3L）（100 ng/ml）和促血小板生成素（10 ng/ml）（CellGenix公司）。可以用慢病毒以感染複數5在75-cm^2的包被有纖連蛋白（25 mg/cm^2）（重組人纖維連接片斷（RetroNectin），寶生物工程株式會社）的組織培養瓶中轉導預刺激的細胞16至24小時。

[1124] 藉由本領域的知識和本揭露中的教義，技術人員可以校正HSC至免疫缺陷條件，諸如HIV/AIDS，包括使HSC與靶向並敲除CCR5的CRISPR-Cas9系統接觸。靶向並敲除含有CCR5-和-

Cpf1蛋白的粒子的指導RNA（以及有利地雙指導方法，例如一對不同的指導RNA；例如，靶向原代人類CD4+ T細胞和CD34+造血幹細胞以及祖細胞（HSPC）中的兩種臨床相關基因B2M和CCR5的指導RNA）與HSC接觸。可以給予如此接觸的細胞；並且視情況處理/擴增；參考卡蒂亞（Cartier）。還參見，凱門（Kiem），“用於HIV疾病的基於造血幹細胞的基因治療（Hematopoietic stem cell-based gene therapy for HIV disease）”，細胞幹細胞（Cell Stem Cell.），2012年2月3日；10(2): 137–147；該文獻連同其引用的文獻一起藉由引用結合在此；曼達爾（Mandal）等人，“使用CRISPR/Cas9有效消除人類造血幹細胞和效應細胞中的基因（Efficient Ablation of Genes in Human Hematopoietic Stem and Effector Cells using CRISPR/Cas9），”細胞幹細胞，第15卷，第5期，第643–652頁，2014年11月6日；該文獻連同其引用的文獻一起藉由引用結合在此。還參考“藉由編輯HIV-1整合的前病毒DNA來抑制HIV-1表現的CRISPR/Cas9系統（CRISPR/Cas9 system to suppress HIV-1 expression by editing HIV-1 integrated proviral DNA）”科技報告| 3 : 2510 | DOI: 10.1038/srep02510，該文獻連同其所引用的文獻一起藉由引用結合在此，作為用於使用CRISPR-Cpf1系統對抗HIV/AIDS的另一種方式。

[1125] 用於HIV治療的基因組編輯的基本原理起源於以下觀察：對於CCR5（病毒的細胞共受體）中的失功能突變純合的個體對感染具有高抗性並且以其他方式獲得健康，這表明藉由基因組編輯模擬此突變可能是安全且有效的治療策略[劉，R.等人，細胞 86, 367-377 (1996)]。當HIV感染的患者被給予來自對失功能

CCR5突變純合的供體的同種異體骨髓移植，從而導致不可檢測的HIV水平和正常CD4 T細胞計數的恢復時，這個想法在臨床上得到證實[胡特爾，G.（Hutter, G.）等人，新英格蘭醫學雜誌（The New England journal of medicine） 360, 692-698 (2009)]。儘管骨髓移植對於大部分HIV患者而言由於成本和潛在移植物對抗宿主疾病而是不現實的治療策略，但將患者自身T細胞轉化為CCR5的HIV治療則係希望的。

[1126] 使用ZFN和NHEJ敲除人源化小鼠HIV模型中的CCR5的早期研究顯示CCR5編輯的CD4 T細胞的移植提高了病毒載量和CD4 T-細胞計數[佩雷斯，E.E.（Perez, E.E.）等人，自然生物技術 26, 808-816 (2008)]。重要的是，該等模型還顯示HIV感染導致CCR5裸細胞的選擇，這表明編輯賦予了適合的優點並且潛在地允許少量編輯的細胞形成治療效果。

[1127] 作為這個研究和其他希望的臨床前研究的結果，現在已在人類中測試了敲除人類T細胞中的CCR5的基因組編輯治療[霍爾特，N.（Holt, N.）等人，自然生物技術 28, 839-847 (2010)；李，L.等人，分子治療：美國基因治療協會雜誌（Molecular therapy : the journal of the American Society of Gene Therapy） 21, 1259-1269 (2013)]。在最近的I期臨床試驗中，來自患有HIV的患者的CD4+ T細胞被去除，用設計敲除CCR5基因的ZFN編輯，並且自身移植回到患者中[特巴斯，P.等人，新英格蘭醫學雜誌 370, 901-910 (2014)]。

[1128] 在另一個研究（曼達爾等人，細胞幹細胞，第15卷，

第5期，第643–652頁，2014年11月6日）中，CRISPR-Cas9已靶向人類CD4+ T細胞和CD34+造血幹細胞和祖細胞（HSPC）中的兩種臨床相關基因B2M和CCR5。使用單一RNA指導序列引起HSPC中而不是T細胞中的高效誘變。雙重指導方法提高了兩種細胞類型中的基因缺失效力。經受使用CRISPR-Cas9的基因組編輯的HSPC保留多譜系潛能。預測的中靶和脫靶突變係經由HSPC中的靶序列定序來檢查的並且低水平的脫靶誘變僅在一個位點處觀察到。該等結果表明CRISPR-Cas9可以有效消除HSPC中具有最小脫靶誘變的基因，該等HSPC對於具有基於造血細胞的治療具有廣泛的適用性。

[1129] 王等人（公共科學圖書館綜合，2014年12月26日；9(12):e115987. doi: 10.1371/journal.pone.0115987）使用表現CRISPR相關蛋白9（Cas9）和CCR5指導RNA的慢病毒載體經由Cas9和單一指導RNA（指導RNA）來沈默CCR5。王等人證實表現Cas9和CCR5指導RNA的慢病毒載體到HIV-1易感性人類CD4+細胞中的單輪轉導產生高頻率的CCR5基因破壞。CCR5基因破壞的細胞不僅抵抗R5-向性HIV-1，包括傳輸/起始者(T/F) HIV-1分離株，而且在R5-向性HIV-1感染過程中比CCR5基因未破壞細胞具有選擇優勢。與在甚至轉導後84天仍穩定轉導的細胞中的該等CCR5指導RNA高度同源的潛在脫靶位點處的基因組突變藉由T7內切核酸酶I測定未檢測到。

[1130] 法恩（Fine）等人（科技報告，2015年7月1日；5:10777. doi: 10.1038/srep10777）鑒定了表現化膿鏈球菌Cas9（SpCas9）蛋

白片段的雙盒系統，該等蛋白片段在細胞中拼接在一起形成能夠進行位點特異性DNA切割的功能蛋白。使用特定CRISPR指導股，法恩等人證明此系統作為單一Cas9並且作為一對Cas9切割酶來切割人類HEK-293T細胞內的HBB和CCR5基因的效力。反式拼接的SpCas9（tsSpCas9）與野生型SpCas9（wtSpCas9）相比在標準轉染劑量下展示~35%的核酸酶活性，但是在較低給藥劑量下具有基本上降低的活性。tsSpCas9相對於wtSpCas9大大減小的開放閱讀框長度潛在地允許更複雜且更長的遺傳元件包裝到AAV載體中，包括組織特異性啟動子、多重指導RNA表現、以及與SpCas9的效應物結構域融合物。

李等人（普通病毒學雜誌（J Gen Virol.），2015年8月；96(8):2381-93. doi: 10.1099/vir.0.000139. 電子版2015年4月8日）證明CRISPR-Cas9可以有效介導細胞系中的CCR5座位的編輯，從而導致細胞表面上的CCR5表現的敲除。下一代定序揭示在預測的CCR5切割位點周圍引入了不同的突變。對於所分析的三種最有效的指導RNA中的每一種，在15個最高得分的潛在位點處未檢測到脫靶效應。藉由構建攜帶CRISPR-Cas9組分的嵌合Ad5F35腺病毒，李等人有效轉導原代CD4+ T-淋巴細胞並且破壞CCR5表現，並且正性轉導細胞被賦予了HIV-1抗性。

[1131] 熟習該項技術者可以利用例如霍爾特，N.等人，自然生物技術 28, 839-847 (2010)；李，L.等人，分子治療：美國基因治療協會雜誌 21, 1259-1269 (2013)；曼達爾等人，細胞幹細胞，第15卷，第5期，第643–652頁，2014年11月6日；王等人（公共科

學圖書館綜合，2014年12月26日；9(12):e115987. doi: 10.1371/journal.pone.0115987）；法恩等人（科技報告，2015年7月1日；5:10777. doi: 10.1038/srep10777）以及李等人（普通病毒學雜誌，2015年8月；96(8):2381-93. doi: 10.1099/vir.0.000139. 電子版2015年4月8日）的以上研究用於使用本發明的CRISPR Cas系統靶向CCR5。

治療病原體，如病毒病原體，諸如HBV

[1132] 本發明還可以適用於治療乙型肝炎病毒（HBV）。然而，CRISPR Cas系統必須適於藉由例如優化劑量和序列來避免RNAi的缺點，諸如過度緊張的（oversatring）內源性小RNA途徑的風險（例如，參見，格林姆（Grimm）等人，自然，第441卷，2006年5月26日）例如，考慮諸如每位人類約1-10 × 10^{14}個粒子的低劑量。在另一個實施方式中，針對HBV的CRISPR Cas系統可以脂質體諸如穩定的核酸脂質粒子（SNALP）來給予（例如，參見，莫里西等人，自然生物技術，第23卷，第8期，2005年8月）。考慮了每日靜脈內注射約1、3或5 mg/kg/天的靶向SNALP中的HBV RNA的CRISPR Cas。日治療可以經過約三天，並且然後每週治療持續約五週。在另一個實施方式中，陳等人（基因治療(2007) 14, 11–19）的系統可以用於並且/或者適於本發明的CRISPR Cas系統。陳等人使用雙股腺相關病毒8-假病毒載體（dsAAV2/8）遞送shRNA。單次給予攜帶HBV-特異性shRNAA的dsAAV2/8載體（每隻小鼠1 × 10^{12}個載體基因組）有效抑制HBV轉基因小鼠肝臟中的HBV蛋白、mRNA和複製DNA的穩定水平，從而使得循環中的HBV載量降低

多至2–3 log_{10}。在載體給予後顯著HBV抑制持續至少120天。shRNA的治療效果係靶序列依賴性的並且並不涉及干擾素啟動。對於本發明，針對HBV的CRISPR Cas系統可以選殖到AAV載體諸如dsAAV2/8載體中並且例如以每位患者約1×10^{15}個載體染色體至約1×10^{16}個載體染色體的劑量給予。在另一個實施方式中，伍德爾（Wooddell）等人（分子治療，第21卷，第5期，973–985，2013年5月）的方法可以用於並且/或者適於本發明的CRISPR Cas系統。伍德爾等人還證實肝細胞靶向的N-乙醯半乳糖胺軛合的蜂毒肽樣肽（NAG-MLP）與肝臟向性膽固醇軛合的siRNA（chol-siRNA）靶向性凝血因子VII（F7）的簡單共注射引起小鼠和非人類靈長類動物中的有效F7敲低而不會有臨床化學變化或細胞介素的誘導。使用HBV感染的暫態和轉基因小鼠模型，伍德爾等人證實NAG-MLP與靶向保守性HBV序列的強效chol-siRNA的單次共注射引起病毒RNA、蛋白質和病毒DNA的複對數抑制，該效果持續時間長。對於本發明，可以設想例如約6 mg/kg NAG-MLP與6 mg/kg HBV特異性CRISPR Cas的靜脈內共注射。在替代方案中，可以在第一天遞送約3 mg/kg NAG-MLP和3 mg/kg HBV特異性CRISPR Cas，隨後在兩週後給予約2-3 mg/kg NAG-MLP和2-3 mg/kg HBV特異性CRISPR Cas。

[1133] 林等人（分子治療—核酸（Mol Ther Nucleic Acids.）2014年8月19日；3:e186. doi: 10.1038/mtna.2014.38）設計了針對基因型A HBV的八種gRNA。在使用HBV-特異性gRNA時，CRISPR-Cas9系統在用HBV表現載體轉染的Huh-7細胞中顯著減少HBV核心和表面蛋白。在八種篩選的gRNA中，鑒定了兩種有效的gRNA。

靶向保守性HBV序列的gRNA作用於不同基因型。使用流體動力學-HBV持久性小鼠模型，林等人進一步證明此系統可以切割含有肝內HBV基因組的質粒並且促進其在體內清除，從而使得血清表面抗原水平降低。該等資料表明CRISPR-Cas9系統可以破壞體外和體內二者的HBV表現模板，這表明其根除持久性HBV感染的可能性。

[1134] 董（Dong）等人（抗病毒研究（Antiviral Res.），2015年6月；118:110-7. doi: 10.1016/j.antiviral.2015.03.015. 電子版2015年4月3日）使用CRISPR-Cas9系統靶向HBV基因組並且有效抑制HBV感染。董等人合成了靶向HBV保守區的四種單一指導RNA（指導RNA）。具有Cas9的該等指導RNA的表現減少了在Huh7細胞以及HBV-複製細胞HepG2.2.15中的病毒產生。董等人進一步證明CRISPR-Cas9直接切割和切割介導的誘變發生在轉染細胞的HBV cccDNA中。在攜帶HBV cccDNA的小鼠模型中，經由快速尾靜脈內注射指導RNA-Cas9質粒引起低水平的cccDNA和HBV蛋白。

[1135] 林等人（普通病毒學雜誌，2015年8月；96(8):2252-61. doi: 10.1099/vir.0.000159. 電子版2015年4月22日）設計了靶向不同HBV基因型的保守區的八種指導RNA（gRNA），該等指導RNA可以在體外和體內顯著抑制HBV複製，以研究使用CRISPR-Cas9系統破壞HBV DNA模板的可能性。HBV特異性gRNA/Cpf1系統可以抑制細胞內不同基因型的HBV的複製，並且病毒DNA藉由單一gRNA/Cpf1系統顯著減少並且藉由不同gRNA/Cpf1系統的組合來清除。

[1136] 王等人（世界胃腸病學雜誌（World J Gastroenterol.），2015年8月28日；21(32):9554-65. doi: 10.3748/wjg.v21.i32.9554）設計了針對基因型A-D HBV的15種gRNA。選擇涵蓋HBV調節區域的兩種gRNA（雙-gRNA）的十一種組合。每種gRNA和11種雙gRNA對抑制HBV（基因型A-D）複製的效率係藉由測量培養基上清液中的HBV表面抗原（HBsAg）或e抗原（HBeAg）來檢查的。HBV-表現載體的破壞係在用雙-gRNA和HBV-表現載體共轉染的HuH7細胞中使用聚合酶鏈反應（PCR）和定序方法檢查的，並且cccDNA的破壞係在HepAD38細胞中使用KCl沈澱法、質粒安全性ATP依賴性DNA酶（PSAD）消化法、滾環擴增法以及定量PCR組合法來檢查的。該等gRNA的細胞毒性係藉由線粒體四唑測定來評估的。所有gRNA可以顯著減少培養上清液中的HBsAg或HBeAg產生，該產生依賴於gRNA所針對的區域。所有雙gRNA均可以有效抑制基因型A-D HBV的HBsAg和/或HBeAg產生，並且雙gRNA抑制HBsAg和/或HBeAg產生的效力與單獨使用的單一gRNA相比顯著增加。另外，藉由PCR直接定序，我們確認該等雙gRNA可以藉由去除兩種使用的gRNA的切割位點之間的片段來特異性破壞HBV表現模板。最重要的是，gRNA-5和gRNA-12組合不僅可以有效抑制HBsAg和/或HBeAg產生，而且破壞HepAD38細胞中的cccDNA儲層。

[1137] 卡利莫瓦（Karimova）等人（科技報告，2015年9月3日；5:13734. doi: 10.1038/srep13734）鑒定了在HBV基因組的S和X區域內由Cas9切割酶進行的特異性和有效性切割所靶向的交叉基因保守性HBV序列。此方法不僅破壞了報導細胞系中的附加型

cccDNA和染色體整合HBV靶位點，而且破壞長期感染和重新感染的肝癌細胞系中的HBV複製。

[1138] 熟習該項技術者可以利用例如林等人（分子治療—核酸，2014年8月19日；3:e186. doi: 10.1038/mtna.2014.38）、董等人（抗病毒研究，2015年6月；118:110-7. doi: 10.1016/j.antiviral.2015.03.015. 電子版2015年4月3日）、劉等人（普通病毒學雜誌，2015年8月；96(8):2252-61. doi: 10.1099/vir.0.000159. 電子版2015年4月22日）、王等人（世界胃腸病學雜誌，2015年8月28日；21(32):9554-65. doi: 10.3748/wjg.v21.i32.9554）、以及卡利莫瓦等人（科技報告，2015年9月3日；5:13734. doi: 10.1038/srep13734）的以上研究用於使用本發明的CRISPR Cas系統靶向HBV。

[1139] 慢性乙型肝炎病毒（HBV）感染係流行的、致命的並且極少治癒的，這係因為病毒附加型DNA（cccDNA）在感染的細胞中持久性存在。拉曼那等人（拉曼那·V、舍羅曼·A（Shlomai A）、考克斯·DB（Cox DB）、施瓦茲·RE（Schwartz RE）、米凱利迪斯·E（Michailidis E）、巴塔·A（Bhatta A）、斯科特·DA、張·F、賴斯·CM（Rice CM）、巴蒂亞·SN（Bhatia SN），科技報告，2015年6月2日；5:10833. doi: 10.1038/srep10833，2015年6月2日線上公開）顯示CRISPR/Cas9系統可以特異性靶向並切割HBV基因組中的保守區，從而能夠強烈抑制病毒基因表現和複製。當Cas9和適當選擇的指導RNA持續表現時，他們證明Cas9切割cccDNA並且cccDNA和病毒基因表現和複製的其他參數顯著降低。因此，他們證實直接靶

向病毒附加型DNA係一種控制病毒並且可能治癒患者的新型治療方法。這也描述於在布羅德研究所（Broad Institute）等人名下的WO2015089465 A1中，該專利的內容藉由引用結合在此。

[1140] 在一些實施方式中，這樣靶向HBV中的病毒附加型DNA係較佳的。

[1141] 本發明還可以適用於治療病原體，例如細菌、真菌和寄生蟲病原體。大部分研究工作集中於開發新的抗生素，然而一旦開發，就會經受相同的抗藥性問題。本發明提供了克服該等困難的新型基於CRISPR的替代方案。另外，與現存的抗生素不同，基於CRISPR的治療可以是製成病原體特異性的，從而誘導靶病原體的細菌細胞死亡同時避免有益細菌死亡。

[1142] 本發明還可以適用於治療丙型肝炎病毒（HCV）。羅埃爾維奇（Roelvinki）等人（分子治療，第20卷，第9期，1737-1749，2012年9月）的方法可以適用於本發明的CRISPR Cas系統。例如,AAV載體諸如AAV8可以是考慮的載體並且例如可以考慮每千克體重約1.25×10^{11}至1.25×10^{13}個載體基因組(vg/kg)。本發明還可以適用於治療病原體，諸如細菌、真菌和寄生蟲病原體。大部分研究工作集中於開發新的抗生素，然而一旦開發，就會經受相同的抗藥性問題。本發明提供了克服該等困難的新型基於CRISPR的替代方案。另外，與現存的抗生素不同，基於CRISPR的治療可以是製成病原體特異性的，從而誘導靶病原體的細菌細胞死亡同時避免有益細菌死亡。

[1143] 蔣等人（“使用CRISPR-Cas對細菌基因組進行RNA指

導編輯”，自然生物技術，第31卷，第233-9頁，2013年3月）使用CRISPR-Cas9系統來突變或殺死肺炎鏈球菌或大腸桿菌。將精確突變引入到基因組中的工作依賴於靶基因組位點處的雙-RNA：Cas9-引導的切割以殺死未突變細胞，並且不再需要選擇標記物或反選擇系統。CRISPR系統已用於逆轉抗生素抗性並且消除各菌株之間的抗性轉移。比克考爾德（Bickard）等人證實重編程式以靶向致病基因的Cas9殺死了致命的金黃色葡萄球菌而不是無毒的金黃色葡萄球菌。將核酸酶重編程式以靶向抗生素抗性基因，破壞了具有抗生素抗性基因的葡萄球菌質粒並且針對質粒攜帶的抗性基因的傳播進行免疫。（參見，比卡爾德等人，“探索產生序列特異性抗微生物劑的CRISPR-Cas核酸酶”，自然生物技術，第32卷，1146–1150, doi:10.1038/nbt.3043，2014年10月05日線上公開）。比卡爾德證實CRISPR-Cas9抗微生物劑在體內起到殺死老鼠皮膚定位模型中的金黃色葡萄球菌的作用。類似地，優素福等人使用一種CRISPR系統來靶向編碼賦予β內醯胺類抗生素抗性的酶的基因（參見優素福等人，“程式設計以敏化並殺死抗生素抗性細菌的溫和和烈性噬菌體”，美國國家科學院院刊，第112卷，第7267–7272頁，doi: 10.1073/pnas.1500107112，2015年5月18日線上公開）。

[1144] CRISPR系統可以用於編輯對其他遺傳方法具有抗性的寄生蟲基因組。例如，已顯示CRISPR-Cas9系統能將雙股斷裂引入到*約氏瘧原蟲*基因組（參見，張等人，“使用CRISPR/Cas9系統有效編輯瘧原蟲基因組”，mBio.，第5卷，e01414-14，2014年7月-8月）。古爾巴爾等人（“使用CRISPR-Cas9系統在人類瘧原蟲鐮狀瘧原蟲中進行基因組編輯”，自然生物技術，第32卷，第819-821頁，

doi: 10.1038/nbt.2925，2014年6月1日線上公開）修飾了兩種基因*orc1*和*kelch13*的序列，這兩種基因分別在基因沈默和形成對青蒿素的抗性中具有推定的作用。儘管對於該修飾沒有直接選擇，但是在適當位點改變的寄生蟲以極高效率恢復，這表明使用此系統可以生成中性突變或甚至有害突變。CRISPR-Cas9還用於修飾其他致病性寄生蟲的基因組，包括*剛地弓形蟲（參見沈等人，"使用CRISPR/CAS9在剛地弓形蟲的不同菌株中進行有效基因破壞"，mBio，第5卷:e01114-14, 2014；以及西迪克等人，*"使用CRISPR/Cas9進行剛地弓形蟲的有效基因組工程化"，公共科學圖書館綜合，第9卷，e100450, doi: 10.1371/journal.pone.0100450，2014年6月27日線上公開）。

[1145] 維亞斯等人（"白色念珠菌CRISPR系統允許對必需基因和基因家族進行遺傳工程化"，科學進展，第1卷，e1500248, DOI: 10.1126/sciadv.1500248，2015年4月3日）採用CRISPR系統來克服白色念珠菌中長期存在的遺傳工程化障礙並且在幾種不同基因的兩個拷貝的單一實驗中進行有效突變。在其他幾種機制促成抗藥性的有機體中，維亞斯產生不再顯示母體臨床分離物Can90顯示的對氟康唑或放線菌酮的抗性的純合雙突變體。維亞斯還藉由創建條件對偶基因來獲得*白色念珠菌*的必需基因中的純合的失功能突變。對於核糖體RNA加工所需要的*DCR1*無效對偶基因在低溫下是致命的，但是在高溫下是可存活的。維亞斯使用引入無效突變的修復模板和不能在16°C下生長的分離的*dcr1/dcr1*突變體。

用遺傳方面或表觀遺傳方面治療疾病

[1146] 本發明的 CRISPR-Cas 系統可以用於校正先前使用 TALEN 和 ZFN 嘗試時有限成功的遺傳突變，並且已被鑒定為 Cas9 系統的潛在靶標，包括如在描述使用 Cas9 系統靶向座位以使用基因治療來治療性解決疾病的方法的愛迪塔斯醫藥公司（Editas Medicine）的公開申請中，包括格盧克曼（Gluckmann）等人的 WO 2015/048577 CRISPR 相關方法和組成物（CRISPR-RELATED METHODS AND COMPOSITIONS）；格盧克曼等人的 WO 2015/070083 具有控制的 gRNA 的 CRISPR 相關方法和組成物（CRISPR-RELATED METHODS AND COMPOSITIONS WITH GOVERNING gRNAS）；在一些實施方式中，提供了原發性開角型青光眼（POAG）的治療、預防或診斷。該靶標較佳的是 MYOC 基因。這描述於 WO2015153780 中，該專利的揭露內容藉由引用結合在此。

[1147] 參考馬埃德爾（Maeder）等人的WO2015/134812 用於治療Usher症候群和色素性視網膜炎的CRISPR/CAS相關方法和組成物（CRISPR/CAS-RELATED METHODS AND COMPOSITIONS FOR TREATING USHER SYNDROME AND RETINITIS PIGMENTOSA）。藉由在此的教義，本發明包括了該等文獻結合在此的教義應用的方法和材料。在眼睛和聽覺基因治療的一方面中，用於治療Usher症候群和色素性視網膜炎的方法和組成物可以適於本發明的CRISPR-Cas系統（例如，參見WO 2015/134812）。在一實施方式中，WO 2015/134812涉及藉由基因編輯治療IIA型Usher症候群（USH2A、USH11A）和色素性視網膜炎39（RP39）或延遲其發作或進展，該基因編輯例如使用CRISPR-

Cas9介導的方法校正USH2A基因的位置2299處的鳥嘌呤缺失（例如，置換USH2A基因的位置2299處的缺失的鳥嘌呤殘基）。使用Cpf1可以實現類似的效果。在一相關方面中，藉由使用一種或多種核酸酶、一種或多種切割酶或其組合來靶向突變，例如以誘導使用供體模板的HDR，以校正點突變（例如，單一核苷酸，例如鳥嘌呤缺失）。突變體USH2A基因的改變或校正可以是藉由任何機制介導的。可能與突變體HSH2A基因的改變（例如，校正）相關聯的示例性機制包括但不限於，非同源性末端接合、微同源性介導的末端接合（MMEJ）、同源定向修復（例如，內源性供體模板介導的）、SDSA（合成依賴性鏈退火）、單股退火或單股侵入。在一實施方式中，用於治療Usher症候群和色素性視網膜炎的方法可以包括獲取由受試者攜帶的突變的知識，例如藉由定序USH2A基因的適當部分。

[1148] 還參考WO 2015/138510並且藉由在此的教義，本發明（使用CRISPR-Cas9系統）包括提供對萊伯氏先天性黑內障10（LCA 10）的治療或對其發作或進展的延遲。LCA 10係由CEP290基因的突變引起的，該突變例如a c.2991+1655，係在內含子26中引起隱蔽剪接位點的CEP290基因中的腺嘌呤到鳥嘌呤的突變。這係CEP290的內含子26的核苷酸1655處的突變，例如A到G的突變。CEP290也稱為：CT87；MKS4；POC3；rd16；BBS14；JBTS5；LCAJO；NPHP6；SLSN6；以及3H11Ag（例如，參見，WO 2015/138510）。在基因治療的一個方面中，本發明涉及在CEP290基因的至少一個對偶基因的LCA靶位置位點附近引入一個或多個斷裂（例如，c.2991 + 1655；A至G）。改變LCA10靶位置係指 (1) 緊密接近或

包括LCA10靶位置處斷裂誘導性引入indel（在此也稱為NHEJ-介導的indel引入（例如c.2991+1655A到G），或者 (2) 基因組序列的斷裂引入的缺失（在此也稱為NHEJ-介導的缺失），包括LCA10靶位置處的突變（例如，c.2991+1655A到G）。兩種方法都產生由LCA10靶位置處的突變引起的隱蔽剪接位點的喪失或破壞。因此，特別設想Cpf1治療LCA的用途。

[1149] 研究者考慮基因治療是否可以用於治療廣泛範圍的疾病。預想本發明基於Cpf1效應蛋白的CRISPR系統用於此類治療用途，包括但不限於進一步列舉的靶區域並且使用如下遞送方法。可以使用本發明系統有效治療的病狀或疾病的一些實例被包括於基因的實例和在此所包含的參考文獻中並且還提供目前與那些病狀相關聯的基因。所舉例說明的基因和病狀並不是詳盡的。

治療循環系統疾病

[1150] 本發明還考慮向血液或造血幹細胞遞送CRISPR-Cas系統，具體地是在此所述的新型CRISPR效應蛋白系統。先前已描述了瓦爾葛蘭（Wahlgren）等人（核酸研究，2012，第40卷，第17期 e130）的血漿外來體並且該等血漿外來體可以用於將CRISPR Cas系統遞送至血液。本發明的核酸靶向系統還考慮治療血紅蛋白病，諸如地中海貧血和鐮狀細胞疾病。例如，參見，關於可以被本發明的CRISPR Cas系統靶向的潛在靶標的國際專利公開案號WO 2013/126794。

[1151] 德拉科布盧（Drakopoulou）“評論文章，基於造血幹細胞的基因治療用於β地中海貧血的持續挑戰（Review Article,

The Ongoing Challenge of Hematopoietic Stem Cell-Based Gene Therapy for β-Thalassemia），”國際幹細胞雜誌（Stem Cells International），第2011卷，文章ID 987980，10頁，doi:10.4061/2011/987980，該文獻連同其引用的文獻如同完全列出一樣藉由引用結合在此，該文獻討論了使用遞送β-球蛋白或γ-球蛋白的基因的慢病毒修飾HSC。與使用慢病毒相比，藉由本領域的知識和本揭露的教義，技術人員可以使用靶向並校正突變的CRISPR-Cas系統（例如，具有遞送β-球蛋白或γ-球蛋白，有利地是非鐮狀β-球蛋白或γ-球蛋白的編碼序列的適合HDR模板）校正關於β地中海貧血的HSC；具體地說，指導RNA可以靶向引起β地中海貧血的突變，並且HDR可以提供對於β-球蛋白或γ-球蛋白的適當表現的編碼。使靶向含有突變-和-Cas蛋白的粒子的指導RNA與攜帶突變的HSC接觸。該粒子還可以含有校正對於β-球蛋白或γ-球蛋白的適當表現的突變的適合HDR；或者HSC可以與含有或遞送HDR模板的第二粒子或載體接觸。可以給予如此接觸的細胞；並且視情況處理/擴增；參考卡蒂亞。就這一點而言，參考：卡瓦紮娜（Cavazzana），“主要經由移植用慢病毒β^{A-T87Q}-球蛋白載體離體轉導的自體造血幹細胞來進行β地中海貧血的基因治療的結果（Outcomes of Gene Therapy for β-Thalassemia Major via ransplantation of AutologousHematopoietic Stem Cells Transduced Ex Vivo with a Lentiviral β^{A-T87Q}-Globin Vector）。” tif2014.org/abstractFiles/Jean%20Antoine%20Ribeil_Abstract.pdf；卡瓦紮娜-卡爾沃（Calvo），“在基因治療人類β-地中海貧血後的輸血自主性和HMGA2啟動（Transfusion independence and HMGA2

activation after gene therapy of human β-thalassaemia）”，自然 467, 318–322（2010年9月16日）doi:10.1038/nature09328；尼安慧思（Nienhuis），“地中海貧血的基因治療的發展（Development of Gene Therapy for Thalassemia）”，冷泉港醫學觀點（Cold Spring Harbor Perpsectives in Medicine），doi: 10.1101/cshperspect.a011833 (2012)，LentiGlobin BB305，它係含有工程化β-球蛋白基因的慢病毒載體（βA-T87Q）；以及謝（Xie）等人，“在患者特異性iPSC中使用CRISPR/Cas9和分段控制來無縫基因校正β地中海貧血性突變（Seamless gene correction of β-thalassaemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyback）”，基因組研究 gr.173427.114 (2014) http://www.genome.org/cgi/doi/10.1101/gr.173427.114（冷泉港實驗室出版社）；這係卡瓦紮娜涉及人類β-地中海貧血的工作主題和謝的工作主題，所有該等文獻連同其中引用或者與其相關聯的所有文獻一起藉由引用結合在此。在本發明中，HDR模板可以提供表現工程化β球蛋白基因（例如βA-T87Q）或者如謝所述的β-球蛋白的HSC。

[1152] 許（Xu）等人（科技報告，2015年7月9日；5:12065. doi: 10.1038/srep12065）已設計直接靶向球蛋白基因中的內含子2突變位點IVS2-654的TALEN和CRISPR-Cas9。許等人使用TALEN和CRISPR-Cas9觀察到在IVS2-654座位處的不同雙股斷裂（DSB）頻率，並且當與piggyBac轉位子供體組合時TALEN介導與CRISPR-Cas9相比更高的同源基因靶向效率。另外，與TALEN相比，對於CRISPR-Cas9觀察到更明顯的脫靶事件。最終，使用OP9共培

養系統選擇用於成紅細胞分化的TALEN校正的iPSC選殖並且檢測到比未校正細胞相對更高的HBB轉錄。

[1153] 宋（Song）等人（幹細胞與發育，2015年5月1日；24(9):1053-65. doi: 10.1089/scd.2014.0347. 電子版2015年2月5日）使用CRISPR/Cas9校正β-Thal iPSC；基因校正的細胞展現出正常的核型和完整的多能性，因為人類胚胎幹細胞（hESC）未顯示脫靶效應。然後，宋等人評價了基因校正的β-Thal iPSC的分化效率。宋等人發現在造血分化過程中，基因校正的β-Thal iPSC顯示增加的胚狀體比率和不同的造血祖細胞百分比。更重要地是，基因校正的β-Thal iPSC品系恢復了HBB表現並且與未校正組相比減少了活性氧產生。宋等人的研究表明β-Thal iPSC的造血分化效率一旦藉由CRISPR-Cas9系統校正就會極大地提高。類似的方法可以利用在此所述的CRISPR-Cas系統，例如包含Cpf1效應蛋白的系統來進行。

[1154] 鐮狀細胞性貧血係常染色體隱性遺傳性疾病，其中紅血細胞變成鐮刀狀。它係由位於染色體11的短臂上的β球蛋白基因中的單鹼基取代所引起的。因此，產生纈胺酸而不是產生引起鐮狀血紅蛋白（HbS）產生的穀胺酸。這導致形成扭曲性狀的紅細胞。由於此異常性狀，可以阻斷小血管，從而引起對骨骼、脾臟和皮膚組織的嚴重損害。這可以導致疼痛發作、頻繁感染、手足綜合症或甚至多器官衰竭。扭曲紅細胞也更易於發生紅細胞溶解，從而導致嚴重貧血。如在β地中海貧血的情況下，鐮狀細胞貧血可以是藉由用CRISPR-Cas系統修飾HSC來校正的。該系統允許藉由切

割其DNA並且然後讓其自身修復來特異性編輯細胞的基因組。Cas蛋白藉由RNA指導序列插入並引導至突變點並且然後它在該點切割DNA。同時，插入健康版本的序列。此序列藉由細胞自身修復系統使用來固定誘導的切割。以這種方式，CRISPR-Cas允許校正在先前獲得的幹細胞中的突變。藉由本領域的知識和本揭露的教義，技術人員可以使用靶向並校正突變的CRISPR-Cas系統（例如，具有遞送β-球蛋白，有利地是非鐮狀β-球蛋白的編碼序列的適合HDR模板）校正關於鐮狀細胞性貧血的HSC；具體地說，指導RNA可以靶向引起鐮狀細胞性貧血的突變，並且HDR可以提供對於β-球蛋白的適當表現的編碼。使靶向含有突變-和-Cas蛋白的粒子的指導RNA與攜帶突變的HSC接觸。該粒子還可以含有校正對於β-球蛋白的適當表現的突變的適合HDR；或者HSC可以與含有或遞送HDR模板的第二粒子或載體接觸。可以給予如此接觸的細胞；並且視情況處理/擴增；參考卡蒂亞。HDR模板可以提供表現工程化β球蛋白基因（例如βA-T87Q）或者如謝所述的β-球蛋白的HSC。

[1155] 威廉姆斯（Williams）“擴展用於造血幹細胞基因治療的適應症（Broadening the Indications for Hematopoietic Stem Cell Genetic Therapies）”，細胞幹細胞 13:263-264 (2013)，該文獻連同其所引用的文獻如同完全列出一樣藉由引用結合在此，該文獻報導了到來自患有溶酶體貯積症、異染性腦白質營養不良疾病（MLD）、由芳基硫酸酯酶A缺乏症（ARSA）引起的遺傳疾病的患者的HSC/P細胞中的慢病毒介導的基因轉移，從而導致神經脫髓鞘；以及到患有威斯科特-奧爾德里奇綜合症（Wiskott-Aldrich syndrome，WAS）的患者（患有WAS蛋白缺乏症的患者，該WAS

蛋白係調節血細胞譜系中的細胞骨架功能的小GTP酶CDC42的效應蛋白，並且因此該等患者罹患免疫缺陷伴隨復發性感染、自身免疫性症狀、以及血小板減少伴隨異常小且功能失調的血小板，從而導致過量出血和白血病與淋巴瘤的風險增加）的HSC中的慢病毒介導的基因轉移。與使用慢病毒相比，藉由本領域的知識和本揭露的教義，技術人員可以使用靶向並校正突變（芳基硫酸酯酶A缺乏症（ARSA））的CRISPR-Cas系統（例如，具有遞送ARSA的編碼序列的適合HDR模板）校正關於MLD（芳基硫酸酯酶A缺乏症（ARSA））的HSC；具體地說，指導RNA可以靶向引起MLD（缺陷性ARSA）的突變，並且HDR可以提供對於ARSA的適當表現的編碼。使靶向含有突變-和-Cas蛋白的粒子的指導RNA與攜帶突變的HSC接觸。該粒子還可以含有校正對於ARSA的適當表現的突變的適合HDR；或者HSC可以與含有或遞送HDR模板的第二粒子或載體接觸。可以給予如此接觸的細胞；並且視情況處理/擴增；參考卡蒂亞。與使用慢病毒相比，藉由本領域的知識和本揭露的教義，技術人員可以使用靶向並校正突變（WAS蛋白缺乏症）的CRISPR-Cas系統（例如，具有遞送WAS蛋白的編碼序列的適合HDR模板）校正關於WAS的HSC；具體地說，指導RNA可以靶向引起WAS（WAS蛋白缺乏症）的突變，並且HDR可以提供對於WAS蛋白的適當表現的編碼。使靶向含有突變-和-Cpf1蛋白的粒子的指導RNA與攜帶突變的HSC接觸。該粒子還可以含有校正對於WAS蛋白的適當表現的突變的適合HDR；或者HSC可以與含有或遞送HDR模板的第二粒子或載體接觸。可以給予如此接觸的細胞；並且視情況處理/擴增；參考卡蒂亞。

[1156] 瓦特（Watts），“造血幹細胞擴增和基因治療（Hematopoietic Stem Cell Expansion and Gene Therapy）”，細胞療法（Cytotherapy）13(10):1164–1171. doi:10.3109/14653249.2011.620748 (2011)，該文獻連同其所引用的文獻如同完全列出一樣藉由引用結合在此，該文獻討論了作為許多病症的有高度吸引力的治療選項的造血幹細胞（HSC）基因治療，例如病毒介導的HSC基因治療，該等病症包括血液學病狀、免疫缺陷（包括HIV/AIDS）、以及其他遺傳病症，如溶酶體儲存病，包括SCID-X1、ADA-SCID、β-地中海貧血、X-連鎖的CGD、威斯科特-奧爾德里奇綜合症、範科尼貧血、腎上腺腦白質營養不良（ALD）、以及異染性腦白質營養不良（MLD）。

[1157] 轉讓給策勒克提斯公司（Cellectis）的美國專利公開案號20110225664、20110091441、20100229252、20090271881以及20090222937涉及CREI變體，其中兩個I-CreI單體中的至少一個具有至少兩個取代，一個取代係在分別位於I-CreI的從位置26至40和從位置44至77的LAGLIDADG（SEQ ID NO: 26）核心結構域的兩個功能子結構域的每一個中，所述變體能夠切割來自人類白細胞介素-2受體γ鏈（IL2RG）基因（也稱為常見細胞介素受體γ鏈基因或γC基因）的DNA靶序列。在美國專利公開案號20110225664、20110091441、20100229252、20090271881以及20090222937中鑒定的靶序列可以用於本發明的核酸靶向系統。

[1158] 幾種聯合免疫缺陷（SCID）由T淋巴細胞成熟的缺陷引起，通常與B淋巴細胞的功能缺陷相關聯（卡瓦紮娜-卡爾沃等

人，醫學年鑒（Annu.Rev. Med.），2005, 56, 585-602；費舍爾等人，免疫學綜述（Immunol.Rev.），2005, 203, 98-109）。據估計總發病率係75 000出生兒中有一例。患有未治療的SCID的患者經受多重機會的微生物感染，並且通常活不過一年。SCID可以是藉由進行來自家族供體的造血幹細胞移植來治療。與供體的組織相容性可以廣泛地變化。在腺甙脫胺酶（ADA）缺乏症（係SCID的形式之一）的情況下，患者可以藉由注射重組腺甙脫胺酶來治療。

[1159] 由於ADA基因已在SCID患者中顯示係突變的（吉布列托（Giblett）等人，柳葉刀（Lancet），1972, 2, 1067-1069），所以已鑒定涉及SCID的幾種其他基因（卡瓦紮娜-卡爾沃等人，醫學年鑒，2005, 56, 585-602；費舍爾等人，免疫學綜述，2005, 203, 98-109）。對於SCID存在四種主要原因：(i) 最頻繁形式的SCID係SCID-X1（X-連鎖SCID或X-SCID），它由IL2RG基因的突變引起，從而導致成熟型T淋巴細胞和NK細胞的缺乏。IL2RG編碼γC蛋白（野口（Noguchi）等人，細胞，1993, 73, 147-157），該蛋白係至少五種介白素受體複合物的常見組分。該等受體藉由JAK3激酶（馬基（Macchi）等人，自然，1995, 377, 65-68）啟動幾種靶標，滅活形成與γC滅活相同的綜合症；(ii) ADA基因中的突變導致嘌呤代謝缺陷，這對於淋巴細胞先質係致命的，進而導致B細胞、T細胞和NK細胞的准缺乏；(iii) V(D)J重組係免疫球蛋白和T淋巴細胞受體（TCR）成熟中的必需步驟。涉及此過程的三種基因重組啟動基因1和2（RAG1和RAG2）以及Artemis中的突變導致成熟T淋巴細胞和B淋巴細胞的缺乏；並且 (iv) 還已報導了參與T細胞特異性傳訊的其他基因諸如CD45中的突變，儘管它們代表少數情

況（卡瓦紮娜-卡爾沃等人，醫學年鑒，2005, 56, 585-602；費舍爾等人，免疫學綜述，2005, 203, 98-109）。因為當鑒定它們的遺傳鹼基時，不同SCID形式出於兩種主要的原因已成為基因治療方法的範例（費舍爾等人，免疫學綜述，2005, 203, 98-109）。首先，如在所有血液疾病中，可以設想離體治療。造血幹細胞（HSC）可以從骨髓中恢復，並且保持對於少量細胞分裂的多能性特徵。因此，它們可以在體外處理，並且然後重新注射到患者中，在患者中它們重新填充骨髓。其次，由於在SCID患者中淋巴細胞的成熟受到損害，所以校正的細胞具有選擇性優點。因此，少量校正的細胞可以恢復功能免疫系統。此假設藉由以下各項得到幾次證實：(i) 免疫功能的部分恢復與SCID患者中的突變逆轉相關聯（赫希霍恩（Hirschhorn）等人，自然遺傳學，1996, 13, 290-295；斯蒂芬（Stephan）等人，新英格蘭醫學雜誌，1996, 335, 1563-1567；布索（Bousso）等人，美國國家科學院院刊2000, 97, 274-278；瓦達（Wada）等人，美國國家科學院院刊2001, 98, 8697-8702；西小森（Nishikomori）等人，血液，2004, 103, 4565-4572）；(ii) 造血細胞體外SCID-X1缺陷的校正（坎多迪（Candotti）等人，血液，1996, 87, 3097-3102；卡瓦紮娜-卡爾沃等人，血液，1996，血液，88, 3901-3909；泰勒等人，血液，1996, 87, 3103-3107；哈斯因-貝伊（Hacein-Bey）等人，血液，1998, 92, 4090-4097）；(iii) 動物模型體內SCID-X1（蘇達斯（Soudais）等人，血液，2000, 95, 3071-3077；蔡等人，血液，2002, 100, 72-79）、JAK-3（邦廷（Bunting）等人，自然醫學，1998, 4, 58-64；邦廷等人，人類基因治療（Hum.Gene Ther.），2000, 11, 2353-2364）以及RAG2（耶茨（Yates）等人，血液，2002,

100, 3942-3949）缺乏的校正；以及 (iv) 基因治療臨床試驗的結果（卡瓦紮娜-卡爾沃等人，科學，2000, 288, 669-672；愛烏蒂（Aiuti）等人，自然醫學，2002；8, 423-425；加斯帕（Gaspar）等人，柳葉刀，2004, 364, 2181-2187）。

[1160] 轉讓給兒童醫學中心社團和哈佛大學校長和校友會的美國專利公開案號20110182867涉及經由BCL11A表現或活性抑制劑諸如RNAi和抗體調控造血祖細胞內的胎兒血紅蛋白表現（HbF）的方法和用途。美國專利公開案號20110182867中所揭露的靶標諸如BCL11A可以被本發明的CRISPR Cas系統靶向以用於調控胎兒血紅蛋白表現。對於另外的BCL11A靶標，還參見鮑爾等人（科學，2013年10月11日：第342卷，第6155期，第253-257頁）和許等人（科學，2011年11月18日：第334卷，第6058期，第993-996頁）。

[1161] 藉由本領域的知識和本揭露的教義，技術人員可以校正關於遺傳性血液學病症的HSC，例如β地中海貧血、血友病、或遺傳性溶酶體貯積病。

HSC—遞送並編輯造血幹細胞；以及特定疾病。

[1162] 術語“造血幹細胞”或“HSC”意指廣泛包括認為是HSC的那些細胞，例如產生所有其他血細胞並且來源於中胚層；位於大部分骨骼核心所含有的紅骨髓中的血細胞。本發明的HSC包括具有藉由以下各項鑒定的造血幹細胞表型的細胞：小尺寸、譜系（lin）標記物的缺乏、以及屬於分化系列簇的標記物，如：CD34、CD38、CD90、CD133、CD105、CD45、以及還有c-kit（幹

細胞因子的受體)。造血幹細胞對於用於檢測譜系定型的標記物呈陰性，並且因此稱為Lin-；並且在它們藉由FACS純化的過程中，對於人類有許多多至14種不同的成熟血液譜系標記物，例如對於髓細胞的CD13和CD33、對於紅系細胞的CD71、對於B細胞的CD19、對於巨核細胞的CD61等人；以及，對於B細胞的B220(鼠科CD45)、對於單核細胞的Mac-1（CD11b/CD18）、對於粒細胞的Gr-1、對於紅系細胞的Ter119、對於T細胞的Il7Ra、CD3、CD4、CD5、CD8等等。小鼠HSC標記物：CD34lo/-、SCA-1+、Thy1.1+/lo、CD38+、C-kit+、lin-，並且人類HSC標記物：CD34+、CD59+、Thy1/CD90+、CD38lo/-、C-kit/CD117+、以及lin-。HSC係藉由標記物鑒定的。因此，於在此討論的實施方式中，HSC可以是CD34+細胞。HSC也可以是造血幹細胞CD34-/CD38-。在細胞表面缺乏c-kit且在本領域視為HSC的幹細胞係處於本發明的範圍內，與本領域同樣視為HSC的CD133+細胞一樣。

[1163] CRISPR-Cas（例如Cpf1）系統可以被工程化以靶向HSC內的一個遺傳座位或多個遺傳座位。可以製備有利地密碼子優化用於真核細胞以及具體地哺乳動物細胞，例如人類細胞，例如HSC的Cas（例如Cpf1）蛋白和靶向HSC中的一個座位或多個座位，例如基因EMX1的sgRNA。該等可以經由粒子來遞送。該等粒子可以是藉由混合的Cas（例如Cpf1）蛋白和gRNA來形成。gRNA和Cas（例如Cpf1）蛋白混合物可以例如與包含以下各項或基本上由以下各項組成或者由以下各項組成的混合物混合：表面活性劑、磷脂、生物可降解聚合物、脂蛋白以及醇，由此可以形成含有gRNA和Cas（例如Cpf1）蛋白的粒子。本發明包括如此製備的粒子和來

自這種方法的粒子以及其用途。

[1164] 更普遍地說，粒子可以使用有效方法來形成。首先，可以按適合的莫耳比，例如3：1至1：3或者2：1至1：2或1：1莫耳比，在適合的溫度下，例如15-30°C，例如20-25°C，例如室溫，持續適合的時間，例如15-45分鐘，例如30分鐘，有利地在無菌無核酸酶緩衝液，例如1X PBS中，混合Cas（例如Cpf1）蛋白和靶向基因EMX1或控制基因LacZ的gRNA。單獨地，可以將粒子組分，諸如或者包括：表面活性劑，例如陽離子脂質，例如1,2-二油醯-3-三甲基銨-丙烷（DOTAP）；磷脂，例如，二豆蔻醯磷脂醯膽鹼（DMPC）；生物可降解聚合物，諸如乙二醇聚合物或PEG，以及脂蛋白，諸如低密度脂蛋白，例如膽固醇，溶解於醇中，有利地是C1-6烷基醇，諸如甲醇、乙醇、異丙醇，例如100%乙醇。兩種溶液可以混合在一起以形成含有Cas（例如Cpf1）-gRNA複合物的粒子。在某些實施方式中，該粒子可以含有HDR模板。該粒子可以是與含有gRNA+Cas（例如Cpf1）蛋白的粒子聯合給予的粒子，或者即除了使HSC與含有gRNA+Cas（例如Cpf1）蛋白的粒子接觸之外，還使HSC與含有HDR模板的粒子接觸；或者使HSC與含有所有gRNA、Cas（例如Cpf1）和HDR模板的粒子接觸。該HDR模板可以藉由單獨的載體給予，由此在第一種情況下粒子穿透HSC細胞並且單獨的載體也穿透該細胞，其中該HSC基因組係藉由gRNA+Cas（例如Cpf1）修飾的並且該HDR模板也是存在，由此藉由該HDR修飾基因組座位；例如，這可以導致校正突變。

[1165] 在該等粒子形成之後，可以用每孔15 ug Cas（例如

Cpf1）蛋白轉染96孔板中的HSC。在轉染後三天，可以收穫HSC，並且可以定量在EMX1座位處的插入和缺失（indel）的數目。

[1166] 這說明了HSC可以如何使用靶向HSC中感興趣的一個基因組座位或多個基因組座位的CRISPR-Cas（例如Cpf1）來修飾。有待修飾的HSC可以是在體內，即在生物體中，例如在人類或非人類真核細胞中，例如動物，諸如魚，例如斑馬魚；哺乳動物，例如靈長類動物，例如猿、黑猩猩、獼猴；齧齒動物，例如小鼠、兔、大鼠、犬或狗；家畜（母牛/牛、羊/綿羊、山羊或豬）；家禽或家禽類，例如雞。有待修飾的HSC可以是在體外，例如這種生物體體外。並且，修飾的HSC可以離體使用，即這種生物體的一種或多種HSC可以從該生物體獲得或從該生物體分離，視情況這種或該等HSC可以被擴增，這種或該等HSC係藉由包含靶向HSC中的一個遺傳座位或多個遺傳座位的CRISPR-Cas（例如Cpf1）的一種組成物，例如藉由使這種或該等HSC與該組成物接觸來修飾的，例如其中該組成物包含含有CRISPR酶和靶向HSC中的一個遺傳座位或多個遺傳座位的一種或多種gRNA的粒子，諸如藉由將gRNA和Cas（例如Cpf1）蛋白混合物與包含以下各項或基本上由以下各項組成或由以下各項組成的一種混合物混合來獲得或可獲得的粒子：磷脂、生物可降解聚合物、脂蛋白以及醇（其中一種或多種gRNA靶向HSC中的一個遺傳座位或多個遺傳座位），從而視情況擴增所產生的修飾的HSC並且向該生物體給予產生的修飾的HSC。在一些情況下，分離或獲得的HSC可以是來自第一生物體，諸如來自與第二生物體相同種類的生物體，並且第二生物體可以是向其給予所產生的修飾的HSC的生物體，例如第一生物體係該

第二生物體的供體(諸如第二生物體的親屬,如父母或兄弟姐妹)。修飾的HSC可以具有解決或緩解或減輕個體或受試者或患者的疾病或病狀狀態的症狀的遺傳修飾。例如在第一生物體供體到第二生物體的情況下,修飾的HSC可以具有使得HSC含有與第二生物體更類似的一種或多種蛋白質例如表面標記物或蛋白質的遺傳修飾。修飾的HSC可以具有刺激個體或受試者或患者的疾病或病狀狀態的遺傳修飾並且再次給予非人類生物體,以便製備一種動物模型。根據本揭露和本領域的知識,HSC的擴增係在技術人員的知識範圍內,例如,參見,李,"藉由克服CUL4-介導的HOXB4降解來改進成人造血幹細胞的離體擴增(Improved ex vivo expansion of adult hematopoietic stem cells by overcoming CUL4-mediated degradation of HOXB4)"。血液,2013年5月16日;121(20):4082-9. doi: 10.1182/blood-2012-09-455204. 電子版2013年3月21日。

[1167] 如對於提高活性所指示的,gRNA可以是與Cas(例如Cpf1)蛋白預先複合,之後將整個複合物配製在粒子中。可以按已知有助於將核酸遞送至細胞的不同組分的不同莫耳比來製備配製物(例如1,2-二油醯基-3-三甲基銨-丙烷(DOTAP)、1,2-雙十四醯基-sn-丙三醇-3-磷酸膽鹼(DMPC)、聚乙二醇(PEG)以及膽固醇)。例如DOTAP : DMPC : PEG :膽固醇莫耳比可以是DOTAP 100、DMPC 0、PEG 0、膽固醇0;或者DOTAP 90、DMPC 0、PEG 10、膽固醇0;或者DOTAP 90、DMPC 0、PEG 5、膽固醇5;DOTAP 100、DMPC 0、PEG 0、膽固醇0。本發明因此包括將gRNA、Cas(例如Cpf1)蛋白和形成粒子的組分混合;以及來自此混合的粒子。

[1168] 在一較佳的實施方式中，含有Cas（例如Cpf1)-gRNA複合物的粒子可以是藉由較佳的是以酶:指導RNA的1:1莫耳比將Cas（例如Cpf1）蛋白和一種或多種gRNA混合在一起來形成。單獨地，將已知有助於遞送核酸的不同組分（例如，DOTAP、DMPC、PEG以及膽固醇）較佳的是溶解於乙醇中。將兩種溶液混合在一起以形成含有Cas（例如Cpf1）-gRNA複合物的粒子。在形成粒子後，Cas（例如Cpf1）-gRNA複合物可以轉染到細胞（例如HSC）中。可以採用條形編碼（Bar coding）。可以條形編碼該等粒子、Cas-9和/或gRNA。

[1169] 在一實施方式中，本發明包括製備含有gRNA-和-Cas（例如Cpf1）蛋白的粒子之方法，該方法包括將gRNA和Cas（例如Cpf1）蛋白混合物與包含以下各項或基本上由以下各項組成或者由以下各項組成的混合物混合：表面活性劑、磷脂、生物可降解聚合物、脂蛋白以及醇。一實施方式包括來自該方法的含有gRNA-和-Cas（例如Cpf1）蛋白的粒子。在一實施方式中，本發明包括該粒子在藉由操縱感興趣的基因組座位中的靶序列來修飾感興趣的基因組座位或者生物體或非人類生物體的方法中的用途，該方法包括使含有感興趣的基因組座位的細胞與其中該gRNA靶向該感興趣的基因組座位的該粒子接觸；或者包括藉由操縱感興趣的基因組座位中的靶序列來修飾感興趣的基因組座位或者生物體或非人類生物體之方法，該方法包括使含有感興趣的基因組座位的細胞與其中該gRNA靶向該感興趣的基因組座位的該粒子接觸。在該等實施方式中，感興趣的基因組座位有利地是HSC中的基因組座位。

[1170] 治療應用的考慮因素：基因組編輯治療的考慮因素係序列特異性核酸酶諸如Cpf1核酸酶變體的選擇。每種核酸酶變體可以具有其自身獨特的一組優勢和缺點，許多該等優勢和缺點在治療情況下必須平衡以使治療益處最大化。迄今為止，使用核酸酶的兩種治療編輯方法已表現出很大的希望：基因破壞和基因校正。基因破壞涉及在遺傳元件內創建靶向的indel的NHEJ刺激，常常引起對患者有益的失功能突變。相反，基因校正使用HDR直接逆轉引起疾病的突變，從而恢復功能同時保留校正的元件的生理調節。HDR也可以用於將治療的轉基因插入到基因組內限定的"安全港"座位中，以恢復失去的基因功能。對於有效的特異性編輯治療，必須在靶細胞群體中實現足夠高水平的修飾以逆轉疾病症狀。此治療修飾"閾值"係藉由在治療之後編輯的細胞的適合度和逆轉症狀所需要的基因產物的量來決定的。關於適合度，相對於其未編輯對應物，編輯對靶細胞產生三種潛在結果：適合度增加、中度或者減小。在適合度增加的情況下，例如在SCID-X1的治療中，修飾的造血祖細胞相對於其未編輯的對應物來選擇性擴增。SCID-X1係由IL2RG基因中的突變引起的疾病，該基因的功能對於造血淋巴細胞譜系的適當發育係需要的[倫納德，W.J.（Leonard, W.J.）等人，免疫學綜述（Immunological reviews） 138, 61-86 (1994)；考杉斯基，K.（Kaushansky, K.）和威廉姆斯，W.J.威廉姆斯血液學（Williams hematology），（麥格勞-希爾醫學出版社（McGraw-Hill Medical），紐約，2010）]在患者接受SCID-X1的病毒基因治療的臨床試驗中並且在SCID-X1突變自發校正的少量實例中，校正的造血祖細胞能夠克服此發育阻斷並且相對於其疾病對應物而擴

增以介導治療[布索，P.等人，美國國家科學院院刊 97, 274-278 (2000)；哈斯因-貝伊-阿比納，S.（Hacein-Bey-Abina, S.）等人，新英格蘭醫學雜誌 346, 1185-1193 (2002)；加斯帕，H.B.等人，柳葉刀 364, 2181-2187 (2004)]。在此情況下，當編輯的細胞具有一選擇性優點時，甚至低數目的編輯細胞也可以藉由擴增來增殖，從而為患者提供治療益處。相比之下，對於其他造血疾病如慢性肉芽腫病症（CGD）的編輯可以不誘導對於編輯的造血祖細胞的適合度的變化，從而增加治療修飾閾值。CGD係藉由編碼吞噬細胞氧化酶蛋白的基因中的突變引起的，該等氧化酶蛋白通常被中性粒細胞用來產生殺死病原體的活性氧[慕克吉，S.（Mukherjee, S.）和思拉舍，A.J.（Thrasher, A.J.）基因 525, 174-181 (2013)]。由於該等基因的功能障礙並不影響造血祖細胞適合度或發育，而僅影響成熟造血細胞類型抵禦感染的能力，所以在此疾病中編輯細胞可能並不優先擴增。實際上，在基因治療試驗中未觀察到CGD中基因校正細胞的選擇性優點，從而引起長期細胞移植的困難[梅爾奇，H.L.（Malech, H.L.）等人，美國國家科學院院刊 94，12133-12138 (1997)；康，H.J.（Kang, H.J.）等人，分子治療：美國基因治療協會雜誌 19, 2092-2101 (2011)]。這樣，相對於其中編輯對於靶細胞形成增加的適合度的疾病，治療其中編輯形成中度適合度優點的疾病諸如CGD可能需要顯著更高水平的編輯。如果編輯加強適合度優點，如恢復癌細胞中的腫瘤抑制基因的功能的情況，修飾細胞可以戰勝其疾病對應物，從而使得治療益處相對於編輯率較低。此後一類疾病特別難以用基因組編輯療法來治療。

[1171] 除細胞適合度之外，治療疾病所需要的基因產物的量

還影響必須實現逆轉症狀的治療性基因組編輯的最小水平。B型血友病係其中基因產物水平少量變化可以導致臨床結果顯著變化的疾病。此疾病係藉由編輯因子IX的基因的突變引起的，該因子IX係通常由肝臟分泌到血液中的蛋白質，其中它充當凝血級聯組分。B型血友病的臨床嚴重性與因子IX活性的量相關。嚴重疾病與小於正常活性的1%相關聯，而較輕微形式的疾病與大於因子IX活性的1%相關聯[考杉斯基，K.和威廉姆斯，W.J.威廉姆斯血液學，(麥格勞-希爾醫學出版社，紐約，2010)；洛夫維斯特，T.（Lofqvist, T.）等人，內科醫學雜誌（Journal of internal medicine） 241, 395-400 (1997)]。這表明可以使甚至小百分比的肝細胞恢復因子IX表現的編輯治療可以對臨床結果具有極大的影響。在出生後立即使用ZFN校正B型血友病小鼠模型的研究表明3%-7%校正足以逆轉疾病症狀，從而為此假設提供臨床前證據[李，H.等人，自然 475, 217-221 (2011)]。

[1172] 其中基因產物水平的少量變化可以影響臨床結果的病症和其中對於編輯細胞存在適合度優點的疾病係基因組編輯治療的理想靶標，因為治療修飾閾值係低到足以允許給予當前技術高成功機會。現在已在臨床前水平和I期臨床試驗中成功使用編輯治療靶向該等疾病。延長對於編輯細胞具有中度適合度優點的疾病的該等希望的結果，或者在治療需要更大量的基因產物的情況下，需要改進DSB修復途徑操縱和核酸酶遞送。以下表示出基因組編輯對於治療模型的應用的一些實例，並且以下表的參考文獻和那些參考文獻所引用的文獻如同完全列出一樣藉由引用結合在此。

疾病類型	採用的核酸酶平臺	治療策略	參考文獻
B型血友病	ZFN	HDR介導的校正基因序列的插入	李，H.等人，自然 475, 217-221 (2011)
SCID	ZFN	HDR介導的校正基因序列的插入	吉諾維斯，P.等人，自然 510, 235-240 (2014)
遺傳性高酪胺酸血症	CRISPR	HDR介導的肝臟突變校正	殷，H.（Yin, H.）等人，自然生物技術 32, 551-553 (2014)

[1173] 有利地經由在此所述的遞送系統例如粒子遞送系統，使用CRISPR-Cas（例如Cpf1）系統藉由HDR介導的突變校正或者HDR介導的校正基因序列插入進行靶向，解決以上表中的每種病狀，係根據本揭露和本領域知識的技術人員的知識範圍內。因此，一實施方式包括使攜帶B型血友病、SCID（例如，SCID-X1、ADA-SCID）或遺傳性高酪胺酸血症突變的HSC與含有gRNA-和-Cas（例如Cpf1）蛋白的粒子接觸，從而靶向關於B型血友病、SCID（例如，SCID-X1、ADA-SCID）或遺傳性高酪胺酸血症的感興趣的基因組座位（例如，如李、吉諾維斯或殷所述）。該粒子還可以含有校正突變的適合HDR模板；或者HSC可以與含有或遞送HDR模板的第二粒子或載體接觸。就這一點而言，必須提及的是B型血友病係一種藉由編碼因子IX（決定性凝血級聯組分）的基因的失功能突變引起的X連鎖隱性病症。在嚴重受影響個體中使因子IX活性恢復至超過其水平的1%可以使該疾病轉變成嚴重程度較輕的形式，因為重組因子IX輸注到從年輕時期開始預防的此類患者中以達到此類水平，極大地改善了臨床併發症。藉由本領域的知識和本揭露的教義，技術人員可以使用靶向並校正突變（由編碼因子IX的基因中的失功能突變引起的X連鎖隱性病症）的CRISPR-Cas（例如Cpf1）系統（例如，具有遞送因子IX的編碼序列的適合HDR模板）

校正關於B型血友病的HSC；具體地說，gRNA可以靶向引起B型血友病的突變，並且HDR可以提供對於因子IX的適當表現的編碼。使靶向含有突變-和-Cas（例如Cpf1）蛋白的粒子的gRNA與攜帶突變的HSC接觸。該粒子還可以含有校正對於因子IX的適當表現的突變的適合HDR；或者HSC可以與含有或遞送HDR模板的第二粒子或載體接觸。可以給予如此接觸的細胞；並且視情況處理/擴增；參考卡蒂亞，在此討論的。

[1174] 在卡蒂亞，“小型討論會：X連鎖腎上腺腦白質營養不良、X連鎖腎上腺腦白質營養不良的造血幹細胞移植和造血幹細胞基因治療（MINI-SYMPOSIUM: X-Linked Adrenoleukodystrophypa, Hematopoietic Stem Cell Transplantation and Hematopoietic Stem Cell Gene Therapy in X-Linked Adrenoleukodystrophy）”，腦病理學（Brain Pathology） 20 (2010) 857–862中，該文獻連同其所引用的文獻如同完全列出一樣藉由引用結合在此，認識到同種異體造血幹細胞移植（HSCT）用於將正常溶酶體酶遞送到患有賀勒氏疾病的患者大腦中，並且討論了治療ALD的HSC基因治療。在兩位患者中，在粒細胞集落刺激因子（G-CSF）轉移後收集外週CD34+細胞並且用骨髓增生性肉瘤病毒增強子、缺失的負控制區域、dl587rev引物結合位點取代的(MND)-ALD慢病毒載體轉導。在低濃度細胞介素存在下在16 h過程中用MND-ALD載體轉導來自患者的CD34+細胞。在轉導後冷凍轉導的CD34+細胞，以對5%細胞進行各種安全性測試，該等安全性測試具體地說包括三種複製能力的慢病毒（RCL）測定。CD34+細胞的轉導效率範圍係35%至50%，其中慢病毒整合拷貝的平均值

係0.65與0.70之間。在融化轉導的CD34+細胞之後，用超過4.106轉導的CD34+細胞再輸注患者，然後用白消安和環磷醯胺進行完全骨髓消除。消除患者的HSC以利於移植基因校正的HSC。兩位患者在第13天與第15天之間出現造血恢復。第一位患者在12個月出現幾乎完全的免疫恢復，而第二位患者在9個月出現此恢復。與使用慢病毒相比，藉由本領域的知識和本揭露的教義，技術人員可以使用靶向並校正突變的CRISPR-Cas（Cpf1）系統（例如，具有適合HDR模板）校正關於ALD的HSC；確切地說，gRNA可以靶向ABCD1中的突變，該ABCD1係位於X染色體上的編碼過氧化物酶體膜轉運體蛋白ALD的基因，並且HDR可以提供對於該蛋白質的適當表現的編碼。使靶向含有突變-和-Cas（Cpf1）蛋白的粒子的gRNA與攜帶突變的HSC例如CD34+細胞接觸，如卡蒂亞所述。該粒子還可以含有校正對於過氧化物酶體膜轉運體蛋白的表現的突變的適合HDR；或者HSC可以與含有或遞送HDR模板的第二粒子或載體接觸。如此接觸的細胞視情況可以如卡蒂亞所述地處理。如此接觸的細胞可以如卡蒂亞所述地給予。

[1175] 參考WO 2015/148860，藉由在此的教義，本發明包括該等文獻結合在此的教義應用的方法和材料。在血液相關疾病基因治療的一方面中，用於治療β地中海貧血的方法和組成物可以適於本發明的CRISPR-Cas系統（例如，參見WO 2015/148860）。在一實施方式中，WO 2015/148860涉及例如藉由改變B-細胞CLL/淋巴瘤11A（BCL11A）的基因來治療或預防β地中海貧血或其症狀。該BCL11A基因也稱為B-細胞CLL/淋巴瘤11A、BCL11A-L、BCL11A-S、BCL11AXL、CTIP 1、HBFQTL5以及ZNF。BCL11A編碼涉及

調節球蛋白基因表現的鋅指蛋白。藉由改變BCL11A基因（例如，BCL11A基因的一個或兩個對偶基因），可以增加β球蛋白的水平。γ球蛋白可以替換血紅蛋白複合物中的β球蛋白並且有效攜帶氧氣到組織中，從而改善β地中海貧血病表型。

[1176] 參考WO 2015/148863，並且藉由在此的教義，本發明包括該等文獻中適於本發明的CRISPR-Cas系統的方法和材料。在治療和預防遺傳性血液疾病鐮狀細胞疾病的一個方面中，WO 2015/148863包括改變BCL11A基因。藉由改變BCL11A基因（例如，BCL11A基因的一個或兩個對偶基因），可以增加β球蛋白的水平。γ球蛋白可以替換血紅蛋白複合物中的β球蛋白並且有效攜帶氧氣到組織中，從而改善鐮狀細胞疾病表型。

[1177] 在本發明的一方面中，藉由調整本發明的CRISPR-Cas系統來包括涉及編輯靶核酸序列或者調解靶核酸序列的表現的方法和組成物以及其結合癌症免疫治療的應用。參考WO 2015/161276中的基因治療的應用，該專利涉及可以用於藉由改變一種或多種T細胞表現的基因來影響T細胞增殖、存活和/或功能的方法和組成物，該等T細胞表現的基因例如是FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、TRAC和/或TRBC基因中的一種或多種。在一個相關方面中，T-細胞增殖係藉由改變一種或多種T細胞表現的基因，例如CBLB和/或PTPN6基因、FAS和/或BID基因、CTLA4和/或PDCDI、和/或TRAC和/或TRBC基因來影響。

[1178] 在患者惡性腫瘤中嵌合抗原受體（CAR）19 T細胞展示出抗白血病作用。然而，白血病患者常常並不具有足夠收集的

T細胞，這意味著治療必須涉及來自供體的修飾的T細胞。因此，存在建立供體T細胞銀行的興趣。凱西姆（Qasim）等人（“Talen工程化通用CAR19 T細胞在B-ALL中的第一臨床應用（First Clinical Application of Talen Engineered Universal CAR19 T Cells in B-ALL）”，第57屆ASH年會和博覽會（ASH 57th Annual Meeting and Exposition），2015年12月5-8日，摘要2046（https://ash.confex.com/ash/2015/webprogram/Paper81653.html，2015年11月線上公開）討論了修飾CAR19 T細胞以藉由破壞T細胞受體表現和CD52靶向來消除移植物抗宿主病風險。此外，靶向CD52細胞以使得它們對阿侖單抗不敏感，並因此允許阿侖單抗預防宿主介導的人類白細胞抗原（HLA）錯配的CAR19 T-細胞排斥。研究者使用第三代自滅活載體編碼連接至RQR8的4g7 CAR19（CD19 scFv-4-1BB-CD3ζ），然後用兩對TALEN mRNA電穿孔細胞，以對T細胞受體（TCR）α恒定鏈座位和CD52基因座位進行多重靶向。使用CliniMacs α/β TCR缺失來消耗離體擴增之後仍表現TCR的細胞，從而產生具有<1% TCR表現的T細胞產物（UCART19），85%的細胞表現CAR19並且64%變成CD52陰性。給予修飾的CAR19 T細胞，以治療患者的復發性急性淋巴母細胞白血病。在此提供的教義提供了用於提供修飾的造血幹細胞及其子代的有效方法，該等細胞包括但不限於，血液骨髓系和淋巴系的細胞，包括T細胞、B細胞、單核細胞、巨噬細胞、中性粒細胞、嗜鹼性粒細胞、嗜酸性粒細胞、紅細胞、樹突細胞、以及巨核細胞或血小板、以及中性殺傷性細胞以及其先質或祖細胞。此類細胞可以是藉由敲除、敲入或以其他方式調控靶標，例如以去除或調控如上

所述的CD52，以及其他靶標（包括但不限於CXCR4和PD-1）來修飾。因此，本發明的組成物、細胞和方法可以用於調控免疫反應並且用於結合向患者給予T細胞或其他細胞的修改來治療(不限於)惡性腫瘤、病毒感染以及免疫症狀。

[1179] 參考WO 2015/148670並且藉由在此的教義，本發明包括此文獻結合在此的教義應用的方法和材料。在基因治療的一個方面中，包括用於編輯與人類免疫缺陷病毒（HIV）和獲得性免疫缺陷綜合症（AIDS）相關或有關的靶序列的方法和組成物。，在此描述的發明包括藉由在C-C趨化因子受體類型5（CCR5）的基因中引入一個或多個突變來預防和治療HIV感染和AIDS。該CCR5基因也稱為CKR5、CCR-5、CD195、CKR-5、CCCKR5、CMKBR5、IDDM22、以及CC-CKR-5。在另一個方面中，在此描述的發明包括提供對於HIV感染的預防或減少和/或對HIV進入宿主細胞的能力的預防或減小，例如在已感染的受試者中。HIV的示例性宿主細胞包括但不限於，CD4細胞、T細胞、腸道相關淋巴組織（GALT）、巨噬細胞、樹突細胞、骨髓先質細胞、以及小膠質細胞。病毒進入宿主細胞需要病毒糖蛋白gp41和gp120與CD4受體和共受體例如CCR5相互作用。如果在宿主細胞表面不存在共受體例如CCR5，則病毒不能結合並進入宿主。疾病的進展因此受到阻礙。藉由敲除或敲下宿主細胞內的CCR5，例如藉由引入保護性突變（諸如CCR5δ32突變），阻止了HIV病毒進入宿主細胞。

[1180] X連鎖慢性肉芽腫病（CGD）係由於吞噬細胞NADPH氧化酶活性缺乏或減小而產生的宿主防禦性遺傳病症。使用靶向

或校正突變(吞噬細胞NADPH氧化酶活性缺乏或減小)的CRISPR-Cas(Cpf1)系統(例如，具有遞送吞噬細胞NADPH氧化酶的編碼序列的HDR模板)；具體地說，gRNA可以靶向引起CGD(吞噬細胞NADPH氧化酶缺乏症)的突變，並且HDR可以提供對於吞噬細胞NADPH氧化酶的適當表現的編碼。使靶向含有突變-和-Cas(Cpf1)蛋白的粒子的gRNA與攜帶突變的HSC接觸。該粒子還可以含有校正對於吞噬細胞NADPH氧化酶的適當表現的突變的適合HDR；或者HSC可以與含有或遞送HDR模板的第二粒子或載體接觸。可以給予如此接觸的細胞；並且視情況處理/擴增；參考卡蒂亞。

[1181] 範科尼貧血：在至少15種基因(FANCA、FANCB、FANCC、FANCD1/BRCA2、FANCD2、FANCE、FANCF、FANCG、FANCI、FANCJ/BACH1/BRIP1、FANCL/PHF9/POG、FANCM、FANCN/PALB2、FANCO/Rad51C、以及FANCP/SLX4/BTBD12)上的突變可以引起範科尼貧血。由該等基因產生的蛋白質涉及稱為FA途徑的細胞過程。當形成新DNA拷貝(稱為DNA複製)的過程由於DNA損害而阻斷時，該FA途徑接通(啟動)。該FA途徑將某些蛋白質發送到損害區域，從而觸發DNA修復，以便可以繼續DNA複製。FA途徑具體負責稱為鏈間交聯(ICL)的某種類型的DNA損害。當DNA相反股上的兩個DNA構建塊(核苷酸)異常附接或連接在一起時，會停止DNA複製過程，從而出現ICL。ICL可以是藉由體內產生的有毒物質累積或者藉由用某些癌症治療藥物進行治療引起的。與范科尼貧血相關聯的八種蛋白質組合在一起形成稱為FA核心複合物的複合物。該FA核心複合物啟動稱為FANCD2和

FANCI的兩種蛋白質。這兩種蛋白質的啟動將DNA修復蛋白帶到ICL區域，以便去除交聯並且可以繼續DNA修復。FA核心複合物。更具體地說，FA核心複合物係由FANCA、FANCB、FANCC、FANCE、FANCF、FANCG、FANCL、以及FANCM組成的核多蛋白複合物，充當E3泛素連接酶並且介導ID複合物的啟動，該ID複合物係由FANCD2和FANCI組成的異源二聚體。一旦單泛素化，它就與FA途徑下游的經典腫瘤抑制物相互作用，該等抑制物包括FANCD1/BRCA2、FANCN/PALB2、FANCJ/BRIP1、以及FANCO/Rad51C，並且由此藉由同源重組（HR）幫助DNA修復。80%至90%的FA情況係由於三種基因FANCA、FANCC、以及FANCG之一的突變。該等基因提供用於產生FA核心複合物組分的指示。與FA核心複合物相關聯的此類基因中的突變將使得複合物無功能並且破壞整個FA途徑。因此，DNA損害未得到有效修復並且ICL隨著時間累積。蓋澤爾哈特（Geiselhart），“評論文章，藉由範科尼貧血途徑破壞的傳訊導致造血幹細胞生物學功能障礙：基本機制和潛在治療策略（Review Article, Disrupted Signaling through the Fanconi Anemia Pathway Leads to Dysfunctional Hematopoietic Stem Cell Biology: Underlying Mechanisms and Potential Therapeutic Strategies）”，貧血（Anemia），第2012卷（2012），文章ID 265790, http://dx.doi.org/10.1155/2012/265790討論了FA和涉及編碼FANCC基因的引起體內HSC校正的慢病毒股骨內排斥的動物實驗。使用靶向並校正與FA相關的一個或多個突變的CRISPR-Cas（Cpf1）系統，例如具有分別靶向引起FA的突變FANCA、FANCC、或FANCG中的一個或多個並且提供FANCA、FANCC、或

FANCG中的一個或多個的校正表現的一種或多種gRNA和一種或HDR模板的CRISPR-Cas（Cpf1）系統，；例如，gRNA可以靶向關於FANCC的突變，並且HDR可以提供對於FANCC的適當表現的編碼。使靶向含有一個或多個突變（例如，涉及FA的一個或多個突變，諸如關於FANCA、FANCC或FANCG的任何一種或多種的一個或多個突變）-和-Cas（Cpf1）蛋白的粒子的gRNA與攜帶這種或該等突變的HSC接觸。該粒子還可以含有校正對於涉及FA的一種或多種蛋白質的適當表現的突變，諸如FANCA、FANCC或FANCG中的一個或多個突變的適合HDR模板；或者HSC可以與含有或遞送HDR模板的第二粒子或載體接觸。可以給予如此接觸的細胞；並且視情況處理/擴增；參考卡蒂亞。

[1182] 在此討論的粒子（例如，關於含有一種或多種gRNA和Cas（Cpf1），視情況一種或多種HDR模板，或者一種或多種HDR模板的粒子；例如關於B型血友病、SCID、SCID-X1、ADA-SCID、遺傳性高酪胺酸血症、β-地中海貧血、X連鎖CGD、威斯科特-奧爾德里奇綜合症（Wiskott-Aldrich syndrome）、範科尼貧血（Fanconi anemia）、腎上腺腦白質營養不良（ALD）、異染性腦白質營養不良（MLD）、HIV/AIDS、免疫缺陷疾病、血液疾病、或者遺傳性溶酶體貯積病）有利地藉由將一種或多種gRNA和Cas（Cpf1）蛋白的混合物（視情況含有一種或多種HDR模板或者當關於一種或多種模板的單獨粒子係希望的時此類混合物僅含有一種或多種HDR模板）與包含表面活性劑、磷脂、生物可降解聚合物、脂蛋白以及醇或基本上由其組成或由其組成的混合物混合來獲得或可獲得（其中一種或多種gRNA靶向HSC中的一個遺傳座位或多個遺傳

座位）。

[1183] 實際上，本發明尤其適用於使用基因組編輯治療造血性遺傳病症，以及適用於治療免疫缺陷病症，諸如遺傳性免疫缺陷病症，尤其藉由使用在此討論的粒子技術。遺傳性免疫缺陷病係其中本發明的基因組編輯干預可以成功的疾病。原因包括：其中免疫細胞係子集的造血細胞係治療可進入的。它們可以從身體內去除並且自體或同種異體移植。另外，某些遺傳免疫缺陷病例如重症綜合性免疫缺陷（SCID）形成免疫細胞的增殖性缺點。藉由少見的自發性"逆轉"突變校正引起SCID的遺傳性病變表明校正甚至一個淋巴細胞祖細胞可以足以恢復患者的免疫功能 .../../../Users/t_kowalski/AppData/Local/Microsoft/Windows/Temporary Internet Files/Content.Outlook/GA8VY8LK/Treating SCID for Ellen.docx - _ENREF_1。參見布索，P.等人，來源於體內單個人類T細胞先質的T細胞全部組成部分的多樣性、功能性和穩定性（Diversity, functionality, and stability of the T cell repertoire derived in vivo from a single human T cell precursor）。美國國家科學院院刊 97, 274-278 (2000)。編輯細胞的選擇性優點使得甚至低水平的編輯引起治療效果。本發明的此效果可以見於SCID、威斯科特-奧爾德里奇綜合症、以及在此提及的其他病狀中，包括其他遺傳性造血病症，諸如α地中海貧血和β地中海貧血，其中血紅細胞缺乏不利地影響紅系祖細胞的適合度。

[1184] NHEJ和HDR DSB修復的活性隨著細胞類型和細胞狀態而顯著改變。NHEJ不是藉由細胞週期高度調節的並且它在各細

胞類型中均是有效的，從而在可進入靶細胞群體中允許高水平的基因破壞。相比之下，HDR主要在S/G2期過程中起作用，並且因此限於活躍分裂的細胞，從而、限制了需要有絲分裂細胞的精確基因組修飾的治療[奇奇阿，A.(Ciccia, A.)和埃利奇，S.J.(Elledge, S.J.）分子細胞（Molecular cell） 40, 179-204 (2010)；查普曼，J.R.等人，分子細胞 47, 497-510 (2012)]。

[1185] 經由HDR校正的效率可以是藉由靶向座位的表觀遺傳狀態或序列或所使用的特定修復模板構造（單股對比雙股、長同源臂對比短同源臂）來控制的[哈斯因-貝伊-阿比納，S.等人，新英格蘭醫學雜誌 346, 1185-1193 (2002)；加斯帕，H.B.等人，柳葉刀 364, 2181-2187 (2004)；博伊默，K.J.（Beumer, K.J.）等人，G3 (2013)]。靶細胞中NHEJ和HDR機構的相對活性也可以影響基因校正效率，因為該等途徑可以競爭恢復DSB[博伊默，K.J.等人，美國國家科學院院刊 105, 19821-19826 (2008)]。HDR還增強NHEJ策略中不可見的遞送挑戰，因為它需要同時遞送核酸酶和修復模板。在實踐中，該等限制到目前為止導致治療上相關的細胞類型中的HDR水平較低。臨床翻譯因此主要集中於治療疾病的NHEJ策略，儘管對於B型血友病和遺傳性高酪胺酸血症的小鼠模型目前已描述了概念驗證的臨床前HDR治療[李，H.等人，自然 475, 217-221 (2011)；殷，H.等人，自然生物技術 32, 551-553 (2014)]。

[1186] 任何給定的基因組編輯應用可以包括蛋白、小RNA分子、和/或修復模板的組合，使得該等多個部分的遞送基本上比小分子治療劑更具有挑戰性。已開發用於遞送基因組編輯工具的兩

種主要策略：離體和體內。在離體治療中，從身體內取出患病細胞，進行編輯並且然後將它移植回到患者中。離體編輯具有允許良好限定靶細胞群體並且指定遞送到細胞的治療分子的特定劑量的優點。當脫靶修飾係所關心的問題時，後一種考慮因素可能是特別重要的，因為滴定核酸酶的量可以減少突變（徐等人，2013）。離體方法的另一個優點典型地是由於蛋白質和核酸到用於研究和基因治療應用的培養基的細胞中的有效遞送系統的開發而可以實現的高編輯率。

[1187] 離體方法可能存在限制應用於少數疾病的缺點。例如，靶細胞必須能夠存活於身體之外的操縱。對於許多組織，如大腦，在身體之外培養細胞係主要的挑戰，因為細胞難以存活或者失去其體內功能所需要的特徵。因此，鑒於本揭露和本領域的知識，關於具有適於離體培養和操縱的成體幹細胞群體的組織（諸如造血系統）的離體治療，可以藉由CRISPR-Cas（Cpf1）系統。[邦恩，H.F.（Bunn, H.F.）和阿斯特爾，J.（Aster, J.），血液病的病理生理學（Pathophysiology of blood disorders），（麥格勞-希爾出版社，紐約，2011)]

[1188] 體內基因組編輯設計將編輯系統直接遞送到其天然組織的細胞類型中。體內編輯允許治療其中受影響細胞群不適於離體操縱的疾病。另外，將核酸酶遞送至原位細胞允許治療多個組織和細胞類型。該等特徵可能允許體內治療應用於比離體治療更廣泛的疾病範圍。

[1189] 到此為止，體內編輯已大部分藉由使用具有限定的組

織特異性趨向性的病毒載體來實現。此類載體目前在貨物負載力和趨向性方面係有限的，這使得這種模式的治療局限於其中使用臨床上有用的載體的轉導係有效的器官系統，諸如肝臟、肌肉和眼睛[科特曼，M.A.（Kotterman, M.A.）和謝弗，D.V.（Schaffer, D.V.）遺傳學自然評論（Nature reviews.Genetics） 15, 445-451 (2014)；阮，T.H.（Nguyen, T.H.）和費裡，N.（Ferry, N.），基因治療11增刊1，S76-84 (2004)；博伊，S.E.（Boye, S.E.）等人，分子治療：美國基因治療協會雜誌 21, 509-519 (2013)]。

[1190] 體內遞送的潛在障壁係可以響應於治療所需要的大量病毒而形成的免疫反應，但是此現象並不是基因編輯所獨有的並且在其他基於病毒的基因治療中也觀察到[貝西，N.（Bessis, N.）等人，基因治療11增刊1，S10-17 (2004)]。還可能的是來自編輯的核酸酶本身的肽在MHC I類分子上呈遞以刺激免疫反應，儘管幾乎沒有證據支援此情況以臨床前水平發生。關於此治療模式的另一個主要難題係控制體內基因組編輯核酸酶的分佈以及隨後的劑量，從而導致可能難以預測的脫靶突變特徵。然而，鑒於本揭露和本領域的知識，包括使用用於治療癌症的基於病毒-和-粒子的治療，例如藉由粒子或病毒遞送的HSC體內修飾係技術人員的知識範圍內。

[1191] 離體編輯治療：關於造血細胞純化、培養和移植的長期存在的臨床專業知識使得影響血液系統的疾病諸如SCID、範科尼貧血、威斯科特-奧爾德里奇綜合症以及鐮狀細胞貧血成為離體編輯治療的集中點。集中於造血細胞的另一個原因係，由於先前

設計血液病症的基因治療的努力，已經存在相對高效率的遞送系統。在該等優點的情況下，這種治療模式可以適用於其中編輯細胞具有適合度優點的疾病，以使得少量移植的編輯細胞可以擴展並且治療疾病。一種這樣的疾病係HIV，其中感染導致對CD4+ T細胞的適合度缺點。

[1192] 離體編輯治療最近擴展包括基因校正策略。在來自吉諾維斯和同事的最新論文中克服了對於HDR的離體障壁，他們實現在從罹患SCID-X1的患者獲得的造血幹細胞（HSC）中突變的IL2RG基因的基因校正[吉諾維斯，P.等人，自然 510, 235-240 (2014)]。吉諾維斯等人使用多模態策略完成HSC中的基因校正。首先，使用含有編碼IL2RG的治療性cDNA的HDR模板的整合有效性慢病毒轉導HSC。在轉導之後，用編碼靶向IL2RG中的突變熱點的ZFN的mRNA電穿孔細胞，以刺激基於HDR的基因校正。為了增加HDR比率，使用小分子優化培養條件以鼓勵HSC分裂。在優化的培養條件下，在培養中以治療相關速率獲得核酸酶和HDR模板、來自SCID-X1患者的基因校正的HSC。來自未受影響的個體的經受相同基因校正程序的HSC可以在小鼠中維持長期造血，這係HSC功能的優質標準。HSC能夠產生所有造血細胞類型並且可以自體移植，從而使得它們成為所有造血性遺傳病症的極其有價值的細胞群體[魏斯曼，I.L.（Weissman, I.L.）和靜留，J.A.（Shizuru, J.A.）血液 112, 3543-3553 (2008)]。理論上，基因校正的HSC可以用於治療廣泛範圍的遺傳性血液病症，這使得此研究成為治療性基因組編輯的令人興奮的重大發現。

[1193] 體內編輯治療：體內編輯可以有利地根據本揭露和本領域的治療來使用。對於其中遞送有效的器官系統，已存在許多令人興奮的臨床前治療成功案例。成功的體內編輯治療的第一個實例在B型血友病的小鼠模型中得到證實[李，H.等人，自然 475, 217-221 (2011)]。如前所述，B型血友病係藉由編碼因子IX（決定性凝血級聯組分）的基因的失功能突變引起的X連鎖隱性病症。在嚴重受影響個體中使因子IX活性恢復至超過其水平的1%可以使該疾病轉變成嚴重程度較輕的形式，因為重組因子IX輸注到從年輕時期開始預防的此類患者中以達到此類水平，極大地改善了臨床併發症[洛夫維斯特，T.等人，內科醫學雜誌 241, 395-400 (1997)]。因此，改變患者臨床結果僅需要低水平的HDR基因校正。此外，因子IX係由肝臟合成和分泌的，該肝臟係可以藉由編碼編輯系統的病毒載體有效轉導的器官。

[1194] 使用編碼ZFN和校正的HDR模板的親肝腺相關病毒（AAV）血清型，獲得小鼠肝臟內突變的人源化因子IX基因的多至7%基因校正[李，H.等人，自然 475, 217-221 (2011)]。這使得凝血因子形成動力學（凝血級聯功能的測量值）得到改進，這首先表明體內編輯治療不僅是可行的而且是有效的。如在此所討論的，技術人員根據在此的教義和本領域的知識（例如，李）使用含有HDR模板的粒子和靶向X連鎖隱性病症的突變以逆轉失功能突變的CRISPR-Cas（Cpf1）系統來解決B型血友病。

[1195] 建立此研究，其他組最近使用利用CRISPR-Cas的肝臟體內基因組編輯，以成功治療遺傳性高酪胺酸血症的小鼠模型

並且形成提供針對心血管疾病的保護的突變。這兩種不同的應用表明用於涉及肝功能缺陷的病症的這種方法的多功能性[殷，H.等人，自然生物技術 32, 551-553 (2014)；丁，Q.（Ding, Q.）等人，循環研究（Circulation research） 115, 488-492 (2014)]。體內編輯對於其他器官系統的應用係必需的，以證明此策略係廣泛可用的。目前，已進行優化病毒和非病毒載體的努力，以擴展可以用此治療模式治療的病症範圍[科特曼，M.A.和謝弗，D.V.遺傳學自然評論 15, 445-451 (2014)；殷，H.等人，遺傳學自然評論 15, 541-555 (2014)]。如在此所討論的，技術人員根據在此的教義和本領域的知識（例如，殷）使用含有HDR模板的粒子和靶向突變的CRISPR-Cas（Cpf1）系統來解決遺傳性高酪胺酸血症。

[1196] 靶向缺失，治療應用：基因的靶向缺失可以是較佳的。因此，較佳的是涉及免疫缺陷病症、血液學病狀或遺傳性溶酶體貯積病，例如B型血友病、SCID、SCID-X1、ADA-SCID、遺傳性高酪胺酸血症、β-地中海貧血、X連鎖 CGD、威斯科特-奧爾德里奇綜合症、範科尼貧血、腎上腺腦白質營養不良（ALD）、異染性腦白質營養不良（MLD）、HIV/AIDS、其他代謝性病症的基因，編碼涉及疾病的錯誤折疊蛋白質的基因，導致涉及疾病的功能缺失的基因；總體上，在HSC中可以使用具有認為有利的粒子系統的任何在此討論的遞送系統靶向的突變。

[1197] 在本發明中，CRISPR酶的免疫原性具體地可以根據騰格裡（Tangri）等人相對於紅細胞生成素首先陳敘並隨後發展的方法來減小。因此，定向進化或合理設計可以用於減少大部分種

類（人類或其他種類）中的CRISPR酶（例如Cpf1）的免疫原性。

[1198] 基因組編輯：本發明的CRISPR/Cas（Cpf1）系統可以用於校正先前使用TALEN和ZFN以及慢病毒（包括如在此所討論的）嘗試時有限成功的基因突變。還參見WO2013163628。

治療大腦、中樞神經和免疫系統的疾病

[1199] 本發明還考慮將CRISPR-Cas系統遞送到大腦或神經元。例如，RNA干擾（RNAi）藉由減少杭丁頓氏症的致病基因*HTT*的表現來提供針對這種病症的治療潛力（例如，參見，麥克布賴德（McBride）等人，分子治療，第19卷，第12期，2011年12月，第2152-2162頁），因此申請者假設它可以用於並且/或者適於CRISPR-Cas系統。該CRISPR-Cas系統可以使用減去反義序列脫靶可能性的演算法來生成。CRISPR-Cas序列可以靶向小鼠、恒河猴或人類亨廷頓蛋白的外顯子52中的序列並且在病毒載體諸如AAV中表現。動物（包括人類）可以使用每個腦半球約三次顯微注射（總計六次注射）：前連台的頭側前1 mm（12 μl）並且其餘兩次注射（分別是12 μl和10 μl）與前一次注射的尾側間隔3 mm和6 mm，其中AAV係1e12 vg/ml，速率係約1 μl/min，並且將針再放置於原處5分鐘以允許注入，以從針尖擴散。

[1200] 迪非莉婭（DiFiglia）等人（美國國家科學院院刊，2007年10月23日，第104期，第43卷，17204–17209）觀察到單次給予到成人紋狀體的siRNA靶向Htt中可以沈默突變體Htt，減輕神經元病理，並且延遲在快速起效的病毒轉基因小鼠HD模型中觀察到的異常行為表型。迪非莉婭用2 μl的10 μM Cy3-標記的cc-siRNA-

Htt或未軛合siRNA-Htt注射到小鼠紋狀體內中。在本發明中對於人類可以考慮類似劑量的靶向Htt的CRISPR Cas，例如可以紋狀體內注射約5-10 ml 10 μM靶向Htt的CRISPR Cas。

[1201] 在另一個實例中，布德羅（Boudreau）等人（分子治療，第17卷第6期，2009年6月）將5 μl表現htt-特異性RNAi病毒（在4 × 10^{12}個病毒基因組/ml下）的重組AAV血清型2/1載體注射到紋狀體中。在本發明中對於人類可以考慮類似劑量的靶向Htt的CRISPR Cas，例如可以紋狀體內注射約10-20 ml（4 × 10^{12}個病毒基因組/ml）靶向Htt的CRISPR Cas。

[1202] 在另一個實例中，可以連續給予靶向HTT的CRISPR Cas（例如，參見，餘等人，細胞 150, 895–908，2012年8月31日）。余等人利用遞送0.25 ml/h的滲透泵（型號2004）遞送300 mg/天的ss-siRNA或磷酸鹽緩衝鹽水（PBS）（西格瑪奧德里奇公司），持續28天，並且使用設計為遞送0.5 μl/h的泵（型號2002）遞送75 mg/天的陽性對照MOE ASO，持續14天。用以無菌PBS稀釋的ss-siRNA或MOE填充泵（杜雷克特公司（Durect Corporation）），並且然後在37°C培養24或48小時（型號2004），之後移植。用2.5%異氟烷麻醉小鼠並且然後在顱底做出中線切口。使用立體定位引導件，將插管移植到右側腦室並且用樂泰膠固定。將附接至Alzet微型滲透泵的導管附接到該插管，並且將泵置於中肩胛區域皮下。切口用5.0尼龍縫線閉合。在本發明中對於人類可以考慮類似劑量的靶向Htt的CRISPR Cas，例如可以給予約500至1000 g/天的靶向Htt的CRISPR Cas。

[1203] 在連續輸注的另一個實例中，斯泰爾斯（Stiles）等人（實驗神經病學（Experimental Neurology） 233 (2012) 463–471）將具有鈦針尖的腦實質內導管移植到右核殼中。將導管連接到皮下植入腹部的SynchroMed® II型泵（美敦力公司（Medtronic Neurological），明尼蘇達州明尼阿波利斯市（Minneapolis, MN））。在7天輸注6 μL/天的磷酸鹽緩衝鹽水之後，用測試品再次填充泵並且程式設計以用於連續遞送7天。以約0.1至0.5 μL/min的不同輸注速率輸注約2.3至11.52 mg/d的siRNA。在本發明中對於人類可以考慮類似劑量的靶向Htt的CRISPR Cas，例如可以給予約20至200 mg/天的靶向Htt的CRISPR Cas。在另一個實例中，轉讓給桑加莫公司的美國專利公開案號20130253040的方法也可以從TALES修改為本發明的核酸靶向系統，以用於治療杭丁頓氏症。

[1204] 在另一個實例中，轉讓給桑加莫公司的美國專利公開案號20130253040（WO2013130824）的方法也可以從TALES修改為本發明的CRISPR Cas系統，以用於治療杭丁頓氏症。

[1205] 在布羅德研究所等人名下的WO2015089354 A1（藉由引用結合在此）描述針對杭丁頓氏症（HP）的靶標。關於杭丁頓氏症的CRISPR複合物的可能的靶基因：PRKCE；IGF1；EP300；RCOR1；PRKCZ；HDAC4；以及TGM2。因此，在本發明的一些實施方式中，PRKCE；IGF1；EP300；RCOR1；PRKCZ；HDAC4；以及TGM2中的一種或多種可以被選擇為用於杭丁頓氏症的靶標。

[1206] 其他三核苷酸重複序列病症。該等病症可以包括以下各項中的任一種：I類包括杭丁頓氏症（HD）和脊髓小腦性共濟失

調；II類擴增係在表型上多樣的，其中異種擴增通常數量較小但也可見於基因外顯子。並且III類包括脆性X綜合症、強直性肌營養不良、兩種脊髓小腦性共濟失調、青少年肌陣攣癲癇、以及弗裡德賴希氏共濟失調。

[1207] 本發明的另一個方面涉及利用CRISPR-Cas系統校正已鑒定為與拉福拉病相關聯的EMP2A和EMP2B基因中的缺陷。拉福拉病係常染色體隱性病狀，它的特徵在於在青年期可以作為癲癇發作開始的進行性肌陣攣性癲癇。該疾病的幾種情況可以是由已鑒定的基因中突變引起的。該疾病引起驚厥、肌肉痙攣、行走困難、癡呆、以及最終死亡。目前沒有已證明針對疾病進展有效的治療。與癲癇相關聯的其他遺傳性異常也可以是藉由CRISPR-Cas系統靶向的並且潛在遺傳學在癲癇遺傳學和遺傳性癲癇（Genetics of Epilepsy and Genetic Epilepsies中進一步描述，該文獻由朱利亞諾·阿文濟尼（Giuliano Avanzini）、傑佛瑞L.諾貝爾斯（Jeffrey L. Noebels）編輯，馬里亞尼兒科神經學基礎（Mariani Foundation Paediatric Neurology）:20；2009）。

[1208] 轉讓給桑加莫生物科技公司的美國專利公開案號20110158957中涉及滅活T細胞受體（TCR）基因的方法也可以被修改成本發明的CRISPR Cas系統。在另一個實例中，轉讓給桑加莫生物科技公司的美國專利公開案號20100311124和轉讓給策勒克提斯公司的美國專利公開案號20110225664中均涉及滅活穀胺醯胺合成酶基因表現基因的方法也可以被修改成本發明的CRISPR Cas系統。

[1209] 用於大腦的遞送選項包括將 DNA 或 RNA 形式的 CRISPR 酶和指導 RNA 封裝成脂質體並且與特洛伊木馬分子軛合以進行跨血腦障壁（BBB）遞送。已顯示特洛伊木馬分子有效於將 B-gal 表現載體遞送到非人類靈長動物大腦中。相同方法可以用於遞送含有 CRISPR 酶和指導 RNA 的載體。例如，夏 CF（Xia CF）和博阿多 RJ（Boado RJ），巴德里奇 WM（Pardridge WM）（“經由人類胰島素受體使用抗生物素蛋白-生物素技術對 siRNA 進行抗體介導的靶向（Antibody-mediated targeting of siRNA via the human insulin receptor using avidin-biotin technology）”。分子藥劑學（Mol Pharm.），2009 年 5 月-6 月；6(3):747-51. doi: 10.1021/mp800194）描述了將短干擾 RNA（siRNA）遞送到培養基中的細胞中的方式，並且在體內藉由組合使用受體特異性單株抗體（mAb）和抗生物素蛋白-生物素技術係可能的。作者們還報導了因為靶向 mAb 與 siRNA 之間的鍵藉由抗生物素蛋白-生物素技術係穩定的，所以在靜脈內給予靶向的 siRNA 之後在體內觀察到遠距離位點諸如大腦處的 RNAi 效果。

[1210] 張等人（分子治療，2003 年 1 月；7(1):11-8.）描述了編碼報導物諸如螢光素酶的表現質粒封裝在包含 85 nm 聚乙二醇化免疫脂質體的“人工病毒”內部的方式，該免疫脂質體與人胰島素受體（HIR）的單株抗體（MAb）一起靶向獼猴大腦中。在靜脈內注射之後，HIRMAb 使得攜帶外源性基因的脂質體經受跨血腦障壁的轉胞吞作用和跨神經元質膜的胞吞作用。與大鼠相比，在獼猴中大腦螢光素酶基因表現水平高 50 倍。藉由組織化學和共焦顯微術二者證實了靈長類動物大腦中 β-半乳糖苷酶基因的廣泛

神經元表現。作者們指示此方法在 24 小時內形成可行的可逆成人轉基因。因此，使用免疫脂質體係較佳的。該等可以結合抗體一起用於靶向特異性組織或細胞表面蛋白。

阿茲海默症

[1211] 美國專利公開案號 20110023153 描述了使用鋅指核酸酶遺傳性修飾與阿茲海默症相關聯的細胞、動物和蛋白質。一旦修飾，細胞和動物就可以使用已知方法進一步測試，以使用 AD 研究中常用的測量（諸如但不限於，學習和記憶、焦慮、抑鬱、成癮、以及感覺運動功能）以及測量行為、功能、病理學、代謝和生物化學功能的測定來研究靶向突變對於 AD 的發展和/或進展的作用。

[1212] 本揭露包括編輯編碼與 AD 相關聯的蛋白質的任何染色體序列。AD 相關蛋白典型地是基於 AD 相關蛋白與 AD 病症的實驗相關性來選擇。例如，AD 相關蛋白的產生率或循環濃度在患有 AD 病症的群體中相對於不存在 AD 病症的群體有所升高或降低。蛋白質水平的差異可以適於蛋白質組學技術來評估，該等技術包括但不限於，西方墨點法、免疫組織化學染色、酶聯免疫吸附測定（ELISA）、以及質譜法。可替代地，AD 相關蛋白可以是藉由使用基因組技術獲得編碼該等蛋白質的基因的基因表現譜來鑒定，該等基因組技術包括但不限於 DNA 微陣列分析、基因表現序列分析（SAGE）、以及定量即時聚合酶鏈反應（Q-PCR）。

[1213] 例如，阿茲海默症相關蛋白的實例可以包括由 VLDLR 基因編碼的極低密度脂蛋白受體蛋白（VLDLR）、由 UBA1

基因編碼的泛素樣修飾劑啟動酶 1（UBA1）、或由 UBA3 基因編碼的 NEDD8 啟動酶 E1 催化亞基蛋白（UBE1C）。

[1214] 作為非限制性實例，與 AD 相關聯的蛋白質包括但不限於如下所列出的蛋白質：染色體序列編碼蛋白質，ALAS2 δ-胺乙醯丙酸合酶 2（ALAS2）；ABCA1 ATP-結合盒式轉運體（ABCA1）；ACE 血管緊張素轉化酶 I（ACE）；APOE 載脂蛋白 E 先質（APOE）；APP 澱粉樣先質蛋白（APP）；AQP1 水通道蛋白 1（AQP1）；BIN1 Myc 盒依賴性相互作用蛋白 1 或橋連整合蛋白 1（BIN1）；BDNF 腦衍生神經元營養因子（BDNF）；BTNL8 嗜乳脂蛋白樣蛋白 8（BTNL8）；C1ORF49 染色體 1 開放閱讀框 49；CDH4 鈣黏蛋白-4；CHRNB2 神經元乙醯膽鹼受體亞基 β-2；CKLFSF2 CKLF 樣 MARVEL 跨膜結構域蛋白 2（CKLFSF2）；CLEC4E C-型凝集素結構域家族 4 成員 e（CLEC4E）；CLU 聚集素蛋白（也稱為載脂蛋白 J）；CR1 紅細胞補體受體 1（CR1，也稱為 CD35；C3b/C4b 受體和免疫黏附受體）；CR1L 紅細胞補體受體 1（CR1L）；CSF3R 粒細胞集落刺激因子 3 受體（CSF3R）；CST3 血清胱抑素 C 或血清胱抑素 3；CYP2C 細胞色素 P450 2C；DAPK1 死亡相關蛋白激酶 1（DAPK1）；ESR1 雌激素受體 1；FCAR IgA 受體的 Fc 片段（FCAR，也稱為 CD89）；FCGR3B IgG Fc 片段低親和力受體 IIIb（FCGR3B 或 CD16b）；FFA2 游離脂肪酸受體 2（FFA2）；FGA 纖維蛋白原（因子 I）；GAB2 GRB2-相關性結合蛋白 2（GAB2）；GAB2 GRB2-相關性結合蛋白 2（GAB2）；GALP 甘丙肽樣肽；GAPDHS 精子發生的甘油醛-3-磷酸脫氫酶（GAPDHS）；GMPB GMBP；HP 結合珠蛋白（HP）；HTR7 5-羥色胺（血清素）受體 7

（腺苷酸環化酶連接）；IDE 胰島素降解酶；IF127 IF127；IFI6 干擾素 α 誘導型蛋白 6（IFI6）；IFIT2 具有三角形四肽（tetratricopeptide）重複單元的干擾素誘導型蛋白 2（IFIT2）；IL1RN 介白素-1 受體拮抗劑（IL-1RA）；IL8RA 介白素 8 受體 α（IL8RA 或 CD181）；IL8RB 介白素 8 受體 β（IL8RB）；JAG1 Jagged 1(JAG1)；KCNJ15 整流鉀通道子家族 J 成員 15(KCNJ15)；LRP6 低密度脂蛋白受體相關蛋白 6(LRP6)；MAPT 微管相關蛋白 τ（MAPT）；MARK4 MAP/微管親和力調節激酶 4（MARK4）；MPHOSPH1 M 期磷蛋白 1；MTHFR 5,10-亞甲基四氫葉酸還原酶；MX2 干擾素誘導型 GTP-結合蛋白 Mx2；NBN Nibrin（也稱為 NBN）；NCSTN 呆蛋白；NIACR2 煙酸受體 2（NIACR2，也稱為 GPR109B）；NMNAT3 煙醯胺核苷酸腺苷轉移酶 3；NTM Neurotrimin（或 HNT）；ORM1 血清類黏蛋白（Orosmucoid）1（ORM1）或 α-1-酸糖蛋白 1；P2RY13 P2Y 嘌呤受體 13(P2RY13)；PBEF1 煙醯胺磷酸核糖轉移酶（NAmPRT 酶或 Nampt）也稱為前-B-細胞集落增強因子 1（PBEF1）或內脂素；PCK1 磷酸烯醇丙酮酸羧化激酶；PICALM 磷脂醯肌醇結合網格蛋白裝配蛋白（PICALM）；PLAU 尿激酶纖溶酶原啟動物（PLAU）；PLXNC1 叢狀蛋白 C1（PLXNC1）；PRNP 朊病毒蛋白；PSEN1 早老素蛋白 1（PSEN1）；PSEN2 早老素蛋白 2（PSEN2）；PTPRA 蛋白酪胺酸激酶磷酸酶受體 A 型蛋白（PTPRA）；RALGPS2 具有 PH 結構域和 SH3 結合模體的 Ral GEF 2（RALGPS2）；RGSL2 G 蛋白傳訊樣調節因子 2(RGSL2)；SELENBP1 硒結合蛋白 1(SELNBP1)；SLC25A37 線粒體轉鐵蛋白（Mitoferrin）-1；SORL1 選蛋白

（sortilin）相關受體 L（DLR 類）含 A 重複單元蛋白（SORL1）；TF 運鐵蛋白；TFAM 線粒體轉錄因子 A；TNF 腫瘤壞死因子；TNFRSF10C 腫瘤壞死因子受體超家族成員 10C（TNFRSF10C）；TNFSF10 腫瘤壞死因子受體超家族（TRAIL）成員 10a（TNFSF10）；UBA1 泛素樣修飾物啟動酶 1（UBA1）；UBA3 NEDD8-啟動酶 E1 催化亞基蛋白（UBE1C）；UBB 泛素 B 蛋白（UBB）；UBQLN1 泛醌蛋白（Ubiquilin）-1；UCHL1 泛素羧基末端酯酶 L1 蛋白（UCHL1）；UCHL3 泛素羧基末端水解酶同工酶 L3 蛋白（UCHL3）；VLDLR 極低密度脂蛋白受體蛋白（VLDLR）

[1215] 在示例性實施方式中，與 AD 相關聯且染色體序列被編輯的蛋白質可以是由 VLDLR 基因編碼的極低密度脂蛋白受體蛋白（VLDLR）、由 UBA1 基因編碼的泛素樣修飾物啟動酶 1（UBA1）、由 UBA3 基因編碼的 NEDD8-啟動酶 E1 催化亞基蛋白（UBE1C）、由 AQP1 基因編碼的水通道蛋白 1（AQP1）、由 UCHL1 基因編碼的泛素羧基末端酯酶 L1 蛋白（UCHL1）、由 UCHL3 基因編碼的泛素羧基末端水解酶同工酶 L3 蛋白（UCHL3）、由 UBB 基因編碼的泛素 B 蛋白（UBB）、由 MAPT 基因編碼的微管相關蛋白 τ（MAPT）、由 PTPRA 基因編碼的蛋白酪胺酸磷酸酶受體 A 型蛋白（PTPRA）、由 PICALM 基因編碼的磷脂醯肌醇結合網格蛋白裝配蛋白（PICALM）、由 CLU 基因編碼的聚集素蛋白（也稱為載脂蛋白 J）、由 PSEN1 基因編碼的早老素蛋白 1、由 PSEN2 基因編碼的早老素蛋白 2、由 SORL1 基因編碼的選蛋白相關受體 L（DLR 類）含 A 重複單元蛋白（SORL1）、由 APP 基因編碼的澱粉樣前蛋白（APP）、由 APOE 基因編碼的載脂蛋白 E 先質（APOE）、

或者由 BDNF 基因編碼的腦衍生神經元營養因子（BDNF）。在一示例性實施方式中，遺傳修飾的動物係大鼠，並且編碼與 AD 相關聯的蛋白質的編輯的染色體序列如下：APP 澱粉樣前蛋白（APP）NM_019288；AQP1 水通道蛋白 1（AQP1） NM_012778；BDNF 腦衍生神經元營養因子 NM_012513；CLU 聚集素蛋白（也稱為載脂蛋白 J） NM_053021；MAPT 微管相關蛋白 τ（MAPT）NM_017212；PICALM 磷脂醯肌醇結合網格蛋白裝配蛋白（PICALM） NM_053554 ；PSEN1 早老素蛋白 1（PSEN1）NM_019163；PSEN2 早老素蛋白 2（PSEN2） NM_031087；PTPRA 蛋白酪胺酸激酶磷酸酶受體 A 型蛋白（PTPRA） NM_012763；SORL1 選蛋白相關受體 L（DLR 類）含 A 重複單元蛋白（SORL1）NM_053519；XM_001065506；XM_217115；UBA1 泛素樣修飾物啟動酶 1（UBA1） NM_001014080；UBA3 NEDD8-啟動酶 E1 催化亞基蛋白（UBE1C） NM_057205；UBB 泛素 B 蛋白（UBB）NM_138895；UCHL1 泛素羧基末端酯酶 L1 蛋白（UCHL1）NM_017237；UCHL3 泛素羧基末端水解酶同工酶 L3 蛋白（UCHL3） NM_001110165；VLDLR 極低密度脂蛋白受體蛋白（VLDLR） NM_013155

[1216] 動物或細胞可以包含編碼與 AD 相關聯的蛋白質的 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15 或更多個破壞的染色體序列和編碼與 AD 相關聯的蛋白質的 0、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15 或更多個染色體整合序列。

[1217] 編輯或整合的染色體序列可以被修飾為編碼與 AD 相關聯的改變的蛋白質。在 AD 相關染色體序列中的許多突變已與 AD 相關聯。例如，APP 中的 V7171（即，在位置 717 處的纈胺酸被改變成異亮胺酸）錯義突變引起家族性 AD。在早老素蛋白-1 中的多重突變諸如 H163R（即在位置 163 處的組胺酸被改變成精胺酸）、A246E（即在位置 246 處的丙胺酸被改變成穀胺酸）、L286V（即在位置 286 處的亮胺酸被改變成纈胺酸）以及 C410Y（即在位置 410 處的半胱胺酸被改變成酪胺酸）引起家族性 3 型阿茲海默症。在早老素蛋白-2 中的突變諸如 N141I（即在位置 141 處的天冬醯胺被改變成異亮胺酸）、M239V（即在位置 239 處的甲硫胺酸被改變成纈胺酸）、以及 D439A（即在位置 439 處的天冬胺酸被改變成丙胺酸）引起家族性 4 型阿茲海默症。AD 相關基因的遺傳性變型與疾病的其他相關性係本領域已知的。參見，例如華陵（Waring）等人（2008），神經學檔案（Arch.Neurol.）65:329-334，該文獻的揭露內容藉由引用以其整體結合在此。

分泌酶病症

[1218] 美國專利公開案號20110023146描述了使用鋅指核酸酶遺傳性修飾與分泌酶相關病症相關聯的細胞、動物和蛋白質。分泌酶係將前蛋白加工成其生物活性形式所必需的。分泌酶途徑的不同組分的缺陷導致許多病症，具體地是具有標誌性澱粉狀蛋白生成或澱粉樣蛋白斑的那些病症，諸如阿茲海默症（AD）。

[1219] 分泌酶病症和與該等病症相關聯的蛋白質係造成許多病症的易感性、病症的存在、病症的嚴重性或其任何組合的一

組不同的蛋白質。本揭露包括編輯編碼與分泌酶病症相關聯的蛋白質的任何染色體序列。與分泌酶病症相關聯的蛋白質典型地是基於分泌酶相關蛋白與分泌酶病症的發展的實驗相關性來選擇的。例如，與分泌酶病症相關聯的蛋白質的產生率或循環濃度在患有分泌酶病症的群體中相對於未患有分泌酶病症的群體有所升高或降低。蛋白質水平的差異可以適於蛋白質組學技術來評估，該等技術包括但不限於，西方墨點法、免疫組織化學染色、酶聯免疫吸附測定（ELISA）、以及質譜法。可替代地，與分泌酶病症相關聯的蛋白質可以是藉由使用基因組技術獲得編碼該等蛋白質的基因的基因表現譜來鑒定，該等基因組技術包括但不限於DNA微陣列分析、基因表現序列分析（SAGE）、以及定量即時聚合酶鏈反應（Q-PCR）。

[1220] 作為非限制性實例，與分泌酶病症相關聯的蛋白質包括PSENEN（早老素增強子2同源物（秀麗隱桿線蟲））、CTSB（組織蛋白酶B）、PSEN1（早老素1）、APP（澱粉樣β（A4）先質蛋白）、APH1B（前咽缺陷性1同源物B（秀麗隱桿線蟲））、PSEN2（早老素2（阿茲海默症4））、BACE1（β-位點APP-切割酶1）、ITM2B（整合膜蛋白2B）、CTSD（組織蛋白酶D）、NOTCH1（Notch同源物1，易位相關（果蠅））、TNF（腫瘤壞死因子（TNF超家族，成員2））、INS（胰島素）、DYT10（肌張力障礙10）、ADAM17（ADAM金屬肽酶結構域17）、APOE（載脂蛋白E）、ACE（血管緊張素轉化酶1（肽基二肽酶A）1）、STN（他汀類）、TP53（腫瘤蛋白p53）、IL6（介白素6（干擾素，β2））、NGFR（神經生長因子受體（TNFR超家族，成員16））、IL1B（介白素1，β）、ACHE（乙醯膽鹼酯酶（Yt

血型））、CTNNB1（鏈蛋白（鈣黏蛋白相關蛋白），β1，88kDa）、IGF1（胰島素樣生長因子1（生長調節素C））、IFNG（干擾素，γ）、NRG1（神經調節蛋白1）、CASP3（半胱天冬酶3，細胞凋亡相關半胱胺酸肽酶）、MAPK1（絲裂原啟動蛋白激酶1）、CDH1（鈣黏蛋白1，1型，E-鈣黏蛋白（上皮））、APBB1（澱粉樣β（A4）先質蛋白結合，家族B，成員1（Fe65））、HMGCR（3-羥基-3-甲基戊二醯-輔酶A還原酶）、CREB1（cAMP應答元件結合蛋白1）、PTGS2（前列腺素-內過氧化物合酶2（前列腺素G/H合酶和環氧合酶））、HES1（長毛和裂口增強蛋白1，（果蠅））、CAT（過氧化氫酶）、TGFB1（轉化生長因子，β1）、ENO2（烯醇酶2（γ，神經元））、ERBB4（v-erb-a成紅細胞白血病病毒癌基因同源物4（鳥類））、TRAPPC10（運輸蛋白顆粒複合物10）、MAOB（單胺氧化酶B）、NGF（神經生長因子（β多肽））、MMP12（基質金屬蛋白酶12（巨噬細胞彈力蛋白酶））、JAG1（jagged 1（阿拉吉歐綜合症））、CD40LG（CD40配位基）、PPARG（過氧化物酶體增殖物啟動受體γ）、FGF2（成纖維細胞生長因子2（基底））、IL3（介白素3（集落刺激因子，多個））、LRP1（低密度脂蛋白受體相關蛋白1）、NOTCH4（Notch同源物4（果蠅））、MAPK8（絲裂原-啟動蛋白激酶8）、PREP（脯肽醯內肽酶）、NOTCH3（Notch同源物3（果蠅））、PRNP（朊病毒蛋白）、CTSG（組織蛋白酶G）、EGF（表皮生長因子（β-尿抑胃素））、REN（腎素）、CD44（CD44分子（印度血型））、SELP（選擇素P（顆粒膜蛋白140 kDa，抗原CD62））、GHR（生長激素受體）、ADCYAP1（腺苷酸環化酶啟動多肽1（垂體））、INSR（胰島素受體）、GFAP（膠質原酸性纖維蛋白）、MMP3（基質金屬蛋白酶3（溶基質素1，

前明膠酶(progelatinase))、MAPK10(絲裂原-啟動蛋白激酶10)、SP1(Sp1轉錄因子)、MYC(v-myc髓細胞組織增生病毒癌基因同源物(鳥類))、CTSE(組織蛋白酶E)、PPARA(過氧化物酶體增殖物-啟動受體α)、JUN(jun癌基因)、TIMP1(TIMP金屬肽酶抑制劑1)、IL5(介白素5(集落刺激因子,嗜酸性粒細胞))、IL1A(介白素1,α)、MMP9(基質金屬肽酶9(明膠酶B,92 kDa明膠酶,92 kDa IV型膠原酶))、HTR4(5-羥色胺(血清素)受體4)、HSPG2(硫酸類肝素蛋白多糖2)、KRAS(v-Ki-ras2柯爾斯頓大鼠肉瘤病毒癌基因同源物)、CYCS(細胞色素c,軀體)、SMG1(SMG1同源物,磷脂醯肌醇3-激酶相關激酶(秀麗隱桿線蟲))、IL1R1(介白素1受體,I型)、PROK1(前動力蛋白1)、MAPK3(絲裂原-啟動蛋白激酶3)、NTRK1(神經營養性酪胺酸激酶,受體,1型)、IL13(介白素13)、MME(膜金屬內肽酶)、TKT(轉酮醇酶)、CXCR2(趨化因子(C-X-C模體)受體2)、IGF1R(胰島素樣生長因子1受體)、RARA(視黃酸受體,α)、CREBBP(CREB結合蛋白)、PTGS1(前列腺素-內過氧化物合酶1(前列腺素G/H合酶和環氧合酶))、GALT(半乳糖-1磷酸尿甙基轉化酶)、CHRM1(膽鹼能受體,毒蕈鹼型1)、ATXN1(共濟失調蛋白(ataxin)1)、PAWR(PRKC,細胞凋亡,WT1,調節劑)、NOTCH2(Notch同源物2(果蠅))、M6PR(甘露糖-6-磷酸受體(陽離子依賴型))、CYP46A1(細胞色素P450,家族46,子家族A,多肽1)、CSNK1 D(酪蛋白激酶1,δ)、MAPK14(絲裂原-啟動蛋白激酶14)、PRG2(蛋白多糖2,骨髓(自然殺傷細胞啟動劑,嗜酸性粒細胞顆粒樣主要鹼性蛋白))、PRKCA(蛋白激酶C,α)、L1 CAM(L1細胞黏附分子)、CD40(CD40

分子，TNF受體超家族成員5）、NR1I2（核受體子家族1，I組，成員2）、JAG2（jagged 2）、CTNND1（連環素（鈣黏蛋白相關蛋白），δ 1）、CDH2（鈣黏蛋白2，1型，N-鈣黏蛋白（神經元））、CMA1（糜酶1，肥大細胞）、SORT1（選蛋白1）、DLK1（δ-樣1同源物（果蠅））、THEM4（硫酯酶超家族成員4）、JUP（接合斑珠蛋白）、CD46（CD46分子，補體調節蛋白）、CCL11（趨化因子（C-C模體）配位基11）、CAV3（小窩蛋白3）、RNASE3（核糖核酸酶，RNA酶A家族，3（嗜酸性粒細胞陽離子蛋白））、HSPA8（熱休克70kDa蛋白8）、CASP9（半胱天冬酶9，細胞凋亡相關半胱胺酸肽酶）、CYP3A4（細胞色素P450，家族3，子家族A，多肽4）、CCR3（趨化因子（C-C模體）受體3）、TFAP2A（轉錄因子AP-2α（啟動增強子結合蛋白2α））、SCP2（固醇載體蛋白2）、CDK4（週期蛋白依賴性激酶4）、HIF1A（低氧誘導因子1，α亞基（鹼性螺旋-環-螺旋轉錄因子））、TCF7L2（轉錄因子7-樣2（T-細胞特異性，HMG-盒））、IL1R2（介白素1受體，II型）、B3GALTL（β 1,3-半乳糖基轉移酶樣）、MDM2（Mdm2 p53結合蛋白同源物（小鼠））、RELA（v-rel網狀內皮組織增殖病毒癌基因同源物A（鳥類））、CASP7（半胱天冬酶7，細胞凋亡相關半胱胺酸肽酶）、IDE（胰島素降解酶）、FABP4（脂肪酸結合4，脂肪細胞）、CASK（鈣/鈣調蛋白依賴性絲胺酸蛋白激酶（MAGUK家族））、ADCYAP1R1（腺苷酸環化酶啟動多肽1（垂體）受體I型）、ATF4（啟動轉錄因子4（tax-應答增強子元件B67））、PDGFA（血小板源生長因子α多肽）、C21或f33（染色體21開放閱讀框33）、SCG5（分泌粒蛋白V（7B2蛋白））、RNF123（環指蛋白123）、NFKB1（B細胞中輕κ多肽基因增強子的核因子1）、

ERBB2（v-erb-b2成紅細胞白血病病毒癌基因同源物2，神經/成膠質細胞瘤源癌基因同源物（鳥類））、CAV1（小窩蛋白1，小凹蛋白，22 kDa）、MMP7（基質金屬肽酶7（基質溶解因子，子宮））、TGFA（轉化生長因子，α）、RXRA（類視黃醇X受體，α）、STX1A（突觸融合蛋白1A（大腦））、PSMC4（蛋白酶體（先質，巨蛋白因子）26S亞基，ATP酶，4）、P2RY2（嘌呤受體P2Y，G-蛋白偶聯，2）、TNFRSF21（腫瘤壞死因子受體超家族，成員21）、DLG1（盤狀，大同源物1（果蠅））、NUMBL（numb同源物（果蠅）樣）、SPN（載唾液酸蛋白）、PLSCR1（磷脂促翻轉酶1）、UBQLN2（泛醌蛋白2）、UBQLN1（泛醌蛋白1）、PCSK7（前蛋白轉化酶枯草桿菌蛋白酶/kexin 7型）、SPON1（脊椎蛋白1，細胞外基質蛋白）、SILV（銀同源物（小鼠））、QPCT（穀胺醯胺醯肽環轉移酶）、HESS（長毛和裂口增強蛋白5（果蠅））、GCC1（含GRIP和捲曲螺旋結構域1）、以及其任何組合。

[1221] 遺傳修飾的動物或細胞可以包含編碼與分泌酶病症相關聯的蛋白質的 1、2、3、4、5、6、7、8、9,10 或更多個破壞的染色體序列和編碼與分泌酶病症相關聯的破壞蛋白質的 0、1、2、3、4、5、6、7、8、9、10 或更多個染色體整合序列。

ALS

[1222] 美國專利公開案號 20110023144 描述了使用鋅指核酸酶遺傳性修飾與肌萎縮性側索硬化（ALS）疾病相關聯的細胞、動物和蛋白質。ALS 的特徵在於涉及隨意運動的大腦皮層、腦幹和脊髓中某些神經細胞的逐漸穩定的變性。

[1223] 運動神經元紊亂和與該等病症相關聯的蛋白質係造成對發展運動神經元紊亂的易感性、運動神經元紊亂的存在、運動神經元紊亂的嚴重性或其任何組合的一組不同的蛋白質。本揭露包括編輯編碼與一特定運動神經元紊亂 ALS 病相關聯的蛋白質的任何染色體序列。與 ALS 相關聯的蛋白質典型地是基於 ALS 相關蛋白與 ALS 的實驗相關性來選擇的。例如，與 ALS 相關聯的蛋白質的產生率或循環濃度在患有 ALS 的群體中相對於未患有 ALS 的群體有所升高或降低。蛋白質水平的差異可以適於蛋白質組學技術來評估，該等技術包括但不限於，西方墨點法、免疫組織化學染色、酶聯免疫吸附測定（ELISA）、以及質譜法。可替代地，與 ALS 相關聯的蛋白質可以是藉由使用基因組技術獲得編碼該等蛋白質的基因的基因表現譜來鑒定，該等基因組技術包括但不限於 DNA 微陣列分析、基因表現序列分析（SAGE）、以及定量即時聚合酶鏈反應（Q-PCR）。

[1224] 作為非限制性實例，與 ALS 相關聯的蛋白質包括但不限於以下蛋白質：SOD1 超氧化物歧化酶 1（可溶性）、ALS3 肌萎縮側索硬化 3；SETX senataxin ALS5 肌萎縮側索硬化 5；FUS 肉瘤融合蛋白 ALS7 肌萎縮側索硬化 7；ALS2 肌萎縮側索硬化 2 DPP6 二肽基-肽酶 6；NEFH 神經絲，重多肽 PTGS1 前列腺素內過氧化物合酶 1；SLC1A2 溶質運載蛋白家族 1（神經膠質高親和力穀胺酸轉運體）成員 2 TNFRSF10B 腫瘤壞死因子受體超家族成員 10b；PRPH 外週蛋白 HSP90AA1 熱休克蛋白 90 kDa α（細胞溶質），類別 A 成員 1；GRIA2 穀胺酸受體，離子移變，AMPA 2 IFNG 干擾素，γ；S100B S100 鈣結合蛋白 B FGF2 成纖

維細胞生長因子 2；AOX1 醛氧化酶 1 CS 檸檬酸合酶；TARDBP TAR DNA 結合蛋白 TXN 硫氧還蛋白；RAPH1 Ras 締合（RaIGDS/AF-6）和普列克底物蛋白同源結構域 1 MAP3K5 絲裂原啟動蛋白激酶 5；NBEAL1 蛋白激酶錨定蛋白（neurobeachin）樣 1 GPX1 麩胱甘肽過氧化物酶 1;ICA1L 胰島細胞自身抗原 1.69 kDa-樣 RAC1 ras 相關 C3 肉毒菌毒素底物 1；MAPT 微管相關蛋白 tau ITPR2 肌醇 1,4,5-三磷酸鹽受體，2 型；ALS2CR4 肌萎縮側索硬化 2（青少年）染色體區，候選物 4GLS 穀胺醯胺酶；ALS2CR8 肌萎縮側索硬化 2(青少年)染色體區，候選物 8 CNTFR 睫狀神經營養因子受體；ALS2CR11 肌萎縮側索硬化 2（青少年）染色體區，候選物 11 FOLH1 葉酸水解酶 1；FAM117B 具有序列相似性 117 的家族，成員 B P4HB 脯胺醯 4-羥化酶 β 多肽；CNTF 睫狀神經營養因子 SQSTM1 死骨片（sequestosome） 1；STRADB STE20-相關激酶接合蛋白 β NAIP NLR 家族，細胞凋亡抑制蛋白；YWHAQ 酪胺酸 3-單氧酶/色胺酸（tryptoph）5-單氧酶啟動蛋白，Θ 多肽 SLC33A1 溶質運載蛋白家族 33（乙醯-CoA 轉運體），成員 1；TRAK2 運輸蛋白驅動蛋白結合 2 圖 4 圖 4 同源物，SAC1 脂質磷酸酶結構域；NIF3L1 NIF3 NGG1 相互作用因子 3-樣 1 INA 互聯蛋白神經元中間絲狀體蛋白，α；PARD3B par-3 分區缺陷性 3 同源物 B COX8A 細胞色素 c 氧化酶亞基 VIIIA；CDK15 週期蛋白依賴性激酶 15 HECW1 含有 HECT、C2 和 WW 結構域 E3 泛素蛋白連接酶 1；NOS1 一氧化氮合酶 1 MET met 原癌基因；SOD2 超氧化物歧化酶 2，線粒體 HSPB1 熱休克 27 kDa 蛋白 1；NEFL 神經絲，輕多肽 CTSB 組織蛋白酶 B；ANG 血管生成素，核糖

核酸酶，RNA 酶家族，5 HSPA8 熱休克 70 kDa 蛋白 8；VAPB VAMP（囊泡相關膜蛋白）相關蛋白 B 和 C ESR1 雌激素受體 1；SNCA 突觸核蛋白，α HGF 肝細胞生長因子；CAT 過氧化氫酶 ACTB 肌動蛋白，β；NEFM 神經絲，中等多肽 TH 酪胺酸羥化酶；BCL2 B-細胞 CLL/淋巴瘤 2 FAS Fas（TNF 受體超家族，成員 6）；CASP3 半胱天冬酶 3，細胞凋亡相關半胱胺酸肽酶 CLU 叢生蛋白；SMN1 調聚的運動神經元 1 的存活 G6PD 葡萄糖-6-磷酸脫氫酶；BAX BCL2-締合 X 蛋白 HSF1 熱休克轉錄因子 1；RNF19A 環指蛋白 19A JUN jun 癌基因；ALS2CR12 肌萎縮側索硬化 2（青少年）染色體區，候選物 12 HSPA5 熱休克 70 kDa 蛋白 5；MAPK14 絲裂原啟動蛋白激酶 14 IL10 介白素 10；APEX1 APEX 核酸酶（多功能 DNA 修復酶）1 TXNRD1 硫氧還蛋白還原酶 1；NOS2 一氧化氮合酶 2，誘導型 TIMP1 TIMP 金屬肽酶抑制劑 1；CASP9 半胱天冬酶 9，細胞凋亡相關半胱胺酸肽酶 XIAP 細胞凋亡 X-連鎖抑制劑；GLG1 golgi 糖蛋白 1 EPO 紅細胞生成素；VEGFA 血管內皮生長因子 A ELN 彈性蛋白；GDNF 膠質細胞源性神經營養因子 NFE2L2 核因子（紅細胞源性 2）樣 2；SLC6A3 溶質運載蛋白家族 6（神經遞質蛋白 4 轉運體，多巴胺），成員 3 HSPA4 熱休克 70 kDa 蛋白 4；APOE 載脂蛋白 E PSMB8 蛋白酶體（先質、巨蛋白因子（macropain））亞基，β 型，8；DCTN1 動力蛋白啟動蛋白 1 TIMP3 TIMP 金屬肽酶抑制劑 3；KIFAP3 驅動蛋白締合蛋白 3 SLC1A1 溶質運載蛋白家族 1（神經元/上皮高親和力穀胺酸轉運體，系統 Xag），成員 1；SMN2 運動神經元 2 的存活，著絲粒 CCNC 週期蛋白 C；MPP4 膜蛋白，棕櫚醯化 4

STUB1 STIP1 同源物和含 U-框蛋白 1；ALS2 澱粉樣 β（A4）先質蛋白 PRDX6 過氧化物氧化還原酶 6;SYP 突觸素 CABIN1 鈣調磷酸酶結合蛋白 1；CASP1 半胱天冬酶 1，細胞凋亡相關半胱胺酸肽酶 GART 磷酸核糖基甘胺醯脫甲醯基轉移酶、磷酸核糖基甘胺醯脫合成酶、磷酸核糖基甘胺醯唑基合成酶；CDK5 週期蛋白依賴性激酶 5 ATXN3 共濟失調蛋白 3；RTN4 網狀內皮素 4 C1QB 補體組分 1，q 亞組分，B 鏈；VEGFC 神經生長因子受體 HTT 亨廷頓蛋白；PARK7 帕金森病 7 XDH 黃嘌呤脫氫酶；GFAP 膠質原酸性纖維蛋白 MAP2 微管相關蛋白 2;CYCS 細胞色素 c，軀體 FCGR3B IgG 的 Fc 片段，低親和力 IIIb；CCS 超氧化物歧化酶的銅分子伴侶 UBL5 泛素樣 5；MMP9 基質金屬肽酶 9 SLC18A3 溶質運載蛋白家族 18（囊泡乙醯膽鹼）成員 3；TRPM7 暫態型受體潛在陽離子通道，超家族 M，成員 7 HSPB2 熱休克 27 kDa 蛋白 2；AKT1 v-akt 小鼠胸腺瘤病毒癌基因同源物 1 DERL1 Der1-樣結構域家族，成員 1；CCL2 趨化因子（C—C 模體）配位基 2 NGRN 神經元蛋白（neugrin），神經突增生相關；GSR 麩胱甘肽還原酶 TPPP3 促微管蛋白聚合的蛋白質家族成員 3；APAF1 細胞凋亡肽酶活化物 1 BTBD10 BTB（POZ）結構域活化物 10；GLUD1 穀胺酸脫氫酶 1 CXCR4 趨化因子（C--X—C 模體）受體 4；SLC1A3 溶質運載蛋白家族 1（神經膠質高親和力穀胺酸轉運體）成員 3 FLT1 fms-相關酪胺酸激酶 1；PON1 對氧磷酶 1 AR 雄激素受體；LIF 白血病抑制因子 ERBB3 v-erb-b2 成紅細胞病毒癌基因同源物 3；LGALS1 凝集素，半乳糖苷-結合，溶質，1CD44 CD44 分子；TP53 腫瘤蛋白 p53 TLR3 toll 樣受體 3；GRIA1 穀

胺酸受體，離子移變，AMPA 1GAPDH 甘油醛-3-磷酸脫氫酶；GRIK1 穀胺酸受體離子移變，紅藻胺酸 1 DES 肌間線蛋白；CHAT 膽鹼乙醯轉移酶 FLT4 fms 相關酪胺酸激酶 4；CHMP2B 染色質修飾蛋白 BAG1 BCL2 相關永生基因；MT3 金屬硫蛋白 3 CHRNA4 膽鹼能受體，煙酸，α4；GSS 麩胱甘肽合成酶 BAK1 BCL2-拮抗劑/殺傷細胞 1；KDR 激酶插入結構域受體（III 型受體酪胺酸激酶） GSTP1 麩胱甘肽 s-轉移酶 pi 1；OGG1 8-氧橋鳥嘌呤 DNA 糖苷酶 IL6 介白素 6（干擾素，β2）。

[1225] 動物或細胞可以包含編碼與 ALS 相關聯的蛋白質的 1、2、3、4、5、6、7、8、9,10 或更多個破壞的染色體序列和編碼與 ALS 相關聯的破壞蛋白質的 0、1、2、3、4、5、6、7、8、9、10 或更多個染色體整合序列。與 ALS 相關聯的較佳的蛋白質包括 SOD1（超氧化物歧化酶 1）、ALS2（肌萎縮性側索硬化 2）、FUS（肉瘤融合蛋白）、TARDBP（TAR DNA 結合蛋白）、VAGFA（血管內皮生長因子 A）、VAGFB（血管內皮生長因子 B）、以及 VAGFC（血管內皮生長因子 C）、以及其任何組合。

自閉症

[1226] 美國專利公開案號20110023145描述了使用鋅指核酸酶遺傳性修飾與自閉症譜系障礙（ASD）相關聯的細胞、動物和蛋白質。自閉症譜系障礙（ASD）係一組特徵在於社交交互和溝通的定性損傷、以及行為、興趣和活動的限制性重複和刻板模式的病症。三種病症自閉症、亞斯伯格症候群（AS）和未另行規定的廣泛性發育障礙（PDD-NOS）係具有不同嚴重程度、相關智力功能

和醫學病狀的連續性病症。ASD係主要的遺傳確定的病症，其中遺傳力係約90%。

[1227] 美國專利公開案號20110023145包括編輯編碼與ASD相關聯的蛋白質的任何染色體序列，它們可以適用於本發明的CRISPR Cas系統。與ASD相關聯的蛋白質典型地是基於和ASD相關聯的蛋白質與ASD的發病率或指征的實驗相關性來選擇。例如，與ASD相關聯的蛋白質的產生率或循環濃度在患有ASD的群體中相對於不存在ASD的群體有所升高或降低。蛋白質水平的差異可以適於蛋白質組學技術來評估，該等技術包括但不限於，西方墨點法、免疫組織化學染色、酶聯免疫吸附測定（ELISA）、以及質譜法。可替代地，與ASD相關聯的蛋白質可以是藉由使用基因組技術獲得編碼該等蛋白質的基因的基因表現譜來鑒定，該等基因組技術包括但不限於DNA微陣列分析、基因表現序列分析（SAGE）、以及定量即時聚合酶鏈反應（Q-PCR）。

[1228] 可能與和ASD相關聯的蛋白質相關的疾病狀態或病症的非限制性實例包括自閉症、亞斯伯格症候群（AS）、未另行規定的廣泛性發育障礙（PDD-NOS）、雷特氏綜合症（Rett's syndrome）、結節性硬化、苯酮尿症（phenylketonuria）、史-倫-奧三氏綜合症（Smith-Lemli-Opitz syndrome）、以及脆性X綜合症。作為非限制性實例，與ASD相關聯的蛋白質包括但不限於以下蛋白質：ATP10C 胺磷脂轉運的ATP酶（ATP10C） METMET受體酪胺酸激酶；BZRAP1 MGLUR5（GRM5）代謝型穀胺酸受體5（MGLUR5）；CDH10 鈣黏蛋白-10 MGLUR6（GRM6）代謝型穀

胺酸受體6（MGLUR6）；CDH9 鈣黏蛋白-9 NLGN1 神經連接蛋白-1；CNTN4 接觸蛋白-4 NLGN2 神經連接蛋白-2；CNTNAP2接觸蛋白相關蛋白樣2（CNTNAP2） SEMA5A 神經連接蛋白-3；DHCR7 7-脫氫膽甾醇還原酶（DHCR7） NLGN4X 神經連接蛋白-4 X-連接；DOC2A 含雙C2-樣結構域蛋白α NLGN4Y 神經連接蛋白-4 Y-連接；DPP6 二肽基胺肽酶樣蛋白6 NLGN5 神經連接蛋白-5；EN2 鋸齒蛋白2（EN2） NRCAM 神經元細胞黏附分子（NRCAM）；MDGA2 脆性X精神發育遲緩1（MDGA2） NRXN1 軸突蛋白-1；FMR2（AFF2） AF4/FMR2家族成員2 OR4M2 嗅覺受體4M2；FOXP2 叉頭框蛋白P2（FOXP2） OR4N4 嗅覺受體4N4；FXR1 脆性X精神發育遲緩，常染色體同源物1（FXR1） OXTR 催產素受體（OXTR）；FXR2 脆性X精神發育遲緩，常染色體同源物2（FXR2） PAH 苯基丙胺酸羥化酶（PAH）；GABRA1 γ胺基丁酸受體亞基α-1（GABRA1） PTEN 磷酸酶和張力蛋白同源物（PTEN）；GABRA5 GABAA（γ胺基丁酸）受體α5亞基（GABRA5） PTPRZ1 受體型酪胺酸蛋白磷酸酶ζ（PTPRZ1）；GABRB1 γ胺基丁酸受體亞基β-1（GABRB1） RELN 顫蛋白（Reelin）；GABRB3 GABAA （γ胺基丁酸）受體β3亞基（GABRB3） RPL10 60S核糖體蛋白L10；GABRG1 γ胺基丁酸受體亞基γ-1（GABRG1） SEMA5A 臂板蛋白-5A（SEMA5A）；HIRIP3 HIRA-相互作用蛋白3 SEZ6L2 癲癇相關6同源物（小鼠）樣2；HOXA1 同源框蛋白Hox-A1（HOXA1） SHANK3 SH3和多個錨蛋白重複結構域3（SHANK3）；IL6 介白素-6 SHBZRAP1 SH3和多個錨蛋白重複結構域3（SHBZRAP1）；LAMB1 層黏連蛋白亞基β-1（LAMB1） SLC6A4

血清素轉運體（SERT）；MAPK3絲裂原啟動蛋白激酶3 TAS2R1 味覺受體2型成員1 TAS2R1 ；MAZ Myc-相關鋅指蛋白 TSC1 結節性硬化蛋白1； MDGA2 含MAM結構域的糖基磷脂醯肌醇錨2（MDGA2）TSC2結節性硬化蛋白2；MECP2 甲基CpG結合蛋白2（MECP2） UBE3A 泛素蛋白連接酶E3A（UBE3A）；MECP2 甲基CpG結合蛋白2（MECP2） WNT2 無翼型（Wingless-type）蛋白MMTV整合位點家族，成員2（WNT2）

[1229] 與 ASD 相關聯且其染色體序列被編輯的蛋白質的性質可以並且將發生改變。在較佳的實施方式中，與 ASD 相關聯且其染色體序列被編輯的蛋白質可以是由 BZRAP1 基因編碼的苯二氮䓬類受體（外週）相關蛋白 1（BZRAP1）、由 AFF2 基因（也稱為 MFR2）編碼的 AF4/FMR2 家族成員 2 蛋白（AFF2）、由 FXR1 基因編碼的脆性 X 精神發育遲緩常染色體同源蛋白 1（FXR1）、由 FXR2 基因編碼的脆性 X 精神發育遲緩常染色體同源蛋白 2（FXR2）、由 MDGA2 基因編碼的含 MAM 結構域的糖基磷脂醯肌醇錨蛋白 2（MDGA2）、由 MECP2 基因編碼的甲基 CpG 結合蛋白 2（MECP2）、由 MGLUR5-1 基因（也稱為 GRM5）編碼的代謝型穀胺酸受體 5（MGLUR5）、由 NRXN1 基因編碼的軸突蛋白 1、或者由 SEMA5A 基因編碼的臂板蛋白-5A（SEMA5A）。在一示例性實施方式中，遺傳修飾的動物係大鼠，並且編碼與 ASD 相關聯的蛋白質的編輯的染色體序列係如以下所列出的：BZRAP1 苯二氮䓬類受體（外週）相關蛋白 1（BZRAP1）XM_002727789、XM_213427、XM_002724533、XM_001081125； AFF2（FMR2）AF4/FMR2 家族成員（AFF2）2 XM_219832、XM_001054673；FXR1

脆性 X 精神發育遲緩常染色體同源蛋白 1（FXR1）NM_001012179 ；FXR2 脆性 X 精神發育遲緩常染色體同源蛋白 2（FXR2）NM_001100647；MDGA2 含 MAM 結構域的糖基磷脂醯肌醇錨蛋白 2（MDGA2）NM_199269；MECP2 甲基 CpG 結合蛋白 2（MECP2）NM_022673；MGLUR5 代謝型穀胺酸受體 5（MGLUR5）NM_017012 (GRM5)；NRXN1 軸突蛋白 1 NM_021767；SEMA5A 臂板蛋白-5A（SEMA5A）NM_001107659。

三核苷酸重複序列擴增病症

[1230] 美國專利公開案號20110016540描述了使用鋅指核酸酶遺傳性修飾與三核苷酸重複序列擴增病症相關聯的細胞、動物和蛋白質。三核苷酸重複序列擴增病症係涉及發育神經生物學的複雜的進展性病症並且常常影響認知功能以及感知運動功能。

[1231] 三核苷酸重複序列擴增蛋白係與發展三核苷酸重複序列擴增病症的易感性、三核苷酸重複序列擴增病症的存在、三核苷酸重複序列擴增病症的嚴重性或其任何組合相關聯的一組不同的蛋白質。三核苷酸重複序列擴增病症被分成藉由重複序列類型確定的兩類。最常見的重複序列係三聯體CAG，該三聯體當存在於基因的編碼區時編碼胺基酸穀胺醯胺（Q）。因此，該等病症被稱為聚穀胺醯胺（polyQ）病症並且包括以下疾病：杭丁頓氏症（HD）；脊延髓肌萎縮症（SBMA）；脊髓小腦性共濟失調（1、2、3、6、7、以及17型SCA）；以及齒狀核紅核蒼白球雷維體萎縮（Dentatorubro-Pallidoluysian Atrophy，DRPLA）。其餘三核苷酸重複序列擴增病症並不涉及CAG三聯體或者該CAG三聯體並不在

該基因的編碼區內，並且因此被稱為非聚穀胺醯胺病症。非聚穀胺醯胺病症包括脆性X綜合症（FRAXA）；脆性XE精神發育遲緩（FRAXE）；弗裡德賴希氏共濟失調（FRDA）；強直性肌營養不良（DM）；以及脊髓小腦性共濟失調（8型和12型SCA）。

[1232] 與三核苷酸重複序列擴增病症相關聯的蛋白質典型地是基於和三核苷酸重複序列擴增病症相關聯的蛋白質與三核苷酸重複序列擴增病症的實驗相關性來選擇的。例如，與三核苷酸重複序列擴增病症相關聯的蛋白質的產生率或循環濃度在患有三核苷酸重複序列擴增病症的群體中相對於不存在三核苷酸重複序列擴增病症的群體有所升高或降低。蛋白質水平的差異可以適於蛋白質組學技術來評估，該等技術包括但不限於，西方墨點法、免疫組織化學染色、酶聯免疫吸附測定（ELISA）、以及質譜法。可替代地，與三核苷酸重複序列擴增病症相關聯的蛋白質可以是藉由使用基因組技術獲得編碼該等蛋白質的基因的基因表現譜來鑒定，該等基因組技術包括但不限於DNA微陣列分析、基因表現序列分析（SAGE）、以及定量即時聚合酶鏈反應（Q-PCR）。

[1233] 與三核苷酸重複序列擴增病症相關聯的蛋白質的非限制性實例包括AR（雄激素受體）、FMR1（脆性X精神發育遲緩1）、HTT（亨廷頓蛋白）、DMPK（肌營養不良性肌強直症蛋白激酶）、FXN（費氏共濟失調蛋白（frataxin））、ATXN2（共濟失調蛋白2）、ATN1（萎縮蛋白（atrophin）1）、FEN1（翼結構特異性內切核酸酶1）、TNRC6A（含三核苷酸重複序列6A）、PABPN1（聚(A)結合蛋白，核1）、JPH3（親聯蛋白3）、MED15（仲介體複合物

亞基15）、ATXN1（共濟失調蛋白1）、ATXN3（共濟失調蛋白3）、TBP（TATA框結合蛋白）、CACNA1A（鈣通道，電壓依賴型，P/Q型，α1A亞基）、ATXN80S（ATXN8相反股（非蛋白質編碼股））、PPP2R2B（蛋白磷酸酶2，調節亞基B，β）、ATXN7（共濟失調蛋白7）、TNRC6B（含三核苷酸重複序列6B）、TNRC6C（含三核苷酸重複序列6C）、CELF3（CUGBP，Elav-樣家族成員3）、MAB21L1（mab-21-樣1（秀麗隱桿線蟲））、MSH2（mutS同源物2，結腸癌，非息肉病性1型（大腸桿菌））、TMEM185A（跨膜蛋白185A）、SIX5（SIX同源框5）、CNPY3（冠層3同源物（斑馬魚））、FRAXE（脆性位點，葉酸類型，少見，fra(X)(q28) E）、GNB2（鳥嘌呤核苷酸結合蛋白（G蛋白），β多肽2）、RPL14（核糖體蛋白L14）、ATXN8（共濟失調蛋白8）、INSR（胰島素受體）、TTR（甲狀腺素運載蛋白）、EP400（E1A結合蛋白p400）、GIGYF2（GRB10相互作用GYF蛋白2）、OGG1（8-氧橋鳥嘌呤DNA糖苷酶）、STC1（斯鈣素1）、CNDP1（肌肽二肽酶1（金屬肽酶M20家族））、C10orf2（染色體10開放閱讀框2）、MAML3策劃蛋白（mastermind）樣3（果蠅）、DKC1（先天性角化不良1，角化不良蛋白）、PAXIP1（PAX相互作用（與轉錄啟動結構域相互作用）蛋白1）、CASK（鈣/鈣調蛋白依賴性絲胺酸蛋白激酶（MAGUK家族））、MAPT（微管相關蛋白tau）、SP1（Sp1轉錄因子）、POLG（聚合酶（DNA定向），γ）、AFF2（AF4/FMR2家族，成員2）、THBS1（血小板反應蛋白1）、TP53（腫瘤蛋白p53）、ESR1（雌激素受體1）、CGGBP1（CGG三聯體重複序列結合蛋白1）、ABT1（基本轉錄啟動物1）、KLK3（激肽釋放酶相關肽酶3）、PRNP（朊病毒蛋白）、JUN（jun癌基因）、KCNN3（鉀中間/小電

導鈣啟動通道，子家族N，成員3）、BAX（BCL2相關X蛋白）、FRAXA（脆性位點，葉酸類型，少見，fra(X)(q27.3) A（巨睾丸症，精神發育遲緩））、KBTBD10（含kelch重複序列和BTB（POZ）結構域10）、MBNL1（盲肌樣（果蠅））、RAD51（RAD51同源物（RecA同源物，大腸桿菌）（釀酒酵母））、NCOA3（核受體共啟動物3）、ERDA1（擴增的重複序列結構域，CAG/CTG 1）、TSC1（結節性硬化1）、COMP（軟骨寡聚基質蛋白）、GCLC（穀胺酸-半胱胺酸連接酶，催化亞基）、RRAD（與糖尿病相關聯的Ras相關）、MSH3（mutS同源物3（大腸桿菌））、DRD2（多巴胺受體D2）、CD44（CD44分子（印度血型））、CTCF（CCCTC-結合因子（鋅指蛋白））、CCND1（週期蛋白D1）、CLSPN（卡環（claspin）同源物（非洲爪蟾））、MEF2A（肌細胞增強因子2A）、PTPRU（蛋白質酪胺酸磷酸酶，受體類型，U）、GAPDH（甘油醛-3-磷酸脫氫酶）、TRIM22（含三重模體22）、WT1（韋爾姆斯氏瘤1）、AHR（芳香烴受體）、GPX1（麩胱甘肽過氧化物酶1）、TPMT（硫嘌呤S-甲基轉移酶）、NDP（諾裡病（假神經膠質瘤））、ARX（無芒相關同源框）、MUS81（MUS81內切核酸酶同源物（釀酒酵母））、TYR（酪胺酸酶（眼皮膚白化病IA））、EGR1（早期生長反應因子 1）、UNG（尿嘧啶-DNA糖苷酶）、NUMBL（numb同源物（果蠅）-樣）、FABP2（脂肪酸結合蛋白2，腸）、EN2（鋸齒形同源框2）、CRYGC（晶狀體蛋白，γC）、SRP14（信號識別顆粒14 kDa（同源Alu RNA結合蛋白））、CRYGB（晶狀體蛋白，γB）、PDCD1（程式性細胞死亡1）、HOXA1（同源框A1）、ATXN2L（共濟失調蛋白2-樣）、PMS2（增加的PMS2減數分裂後分離2（釀酒酵母））、GLA（半乳糖苷酶，α）、CBL（Cas-Br-

M（鼠）同向性逆轉錄病毒轉化序列）、FTH1（鐵蛋白，重多肽1）、IL12RB2（介白素12受體，β2）、OTX2（正小齒同源框2）、HOXA5（同源框A5）、POLG2（聚合酶（DNA定向），γ2，輔助亞基）、DLX2（末梢更少（distal-less）同源框2）、SIRPA（信號調節蛋白α）、OTX1（正小齒同源框1）、AHRR（芳香烴受體阻遏物）、MANF（中腦星形膠質細胞源性神經元營養因子）、TMEM158（跨膜蛋白158（基因/偽基因））、以及ENSG00000078687。

[1234] 與三核苷酸重複序列擴增病症相關聯的較佳的蛋白質包括HTT（亨廷頓蛋白）、AR（雄激素受體）、FXN（費氏共濟失調蛋白）、Atxn3（共濟失調蛋白）、Atxn1（共濟失調蛋白）、Atxn2（共濟失調蛋白）、Atxn7（共濟失調蛋白）、Atxn10（共濟失調蛋白）、DMPK（肌營養不良性肌強直症蛋白激酶）、Atn1（萎縮蛋白1）、CBP（creb結合蛋白）、VLDLR（極低密度脂蛋白受體）、以及其任何組合。

治療聽力疾病

[1235] 本發明還考慮將CRISPR-Cas系統遞送到一隻耳朵或兩隻耳朵。

[1236] 研究者調查基因治療是否可以用於幫助進行目前的耳聾治療，即，電子耳蝸。耳聾往往係由不能將信號傳播到聽覺神經元的毛細胞喪失或損傷來引起的。在此類情況下，電子耳蝸可以用於響應於聲音並且將電信號傳輸到神經元。但是該等神經元常常變性並且從耳蝸縮回，因為受損的毛細胞釋放較少的生長因子。

[1237] 美國專利申請20120328580描述了使用注射器諸如單劑量注射器將藥物組成物注射到耳朵（例如，耳朵給藥），諸如注射到耳蝸腔（luminae）（例如，膜蝸管、Sc前庭、以及Sc定音鼓）。例如，可以藉由鼓室內注射（例如，注射到中耳）和/或注射到外耳、中耳和/或內耳來給予一種或多種在此所述的化合物。此類方法在本領域中常規用於例如將類固醇和抗生素給予到人耳朵中。注射可以是例如藉由耳朵的圓窗或藉由耳蝸膠囊。其他內耳給藥方法係本領域已知的（例如，參見，索爾特（Salt）和普龍特科（Plontke），今日藥物發現（Drug Discovery Today），10:1299-1306, 2005）。

[1238] 在另一種給藥模式中，藥物組成物可以經由導管或泵原位給藥。導管或泵可以例如將藥物組成物引導到耳蝸腔或耳朵圓窗和/或結腸（colon）腔。適用於將一種或多種在此所述的化合物給予到耳朵例如人類耳朵的示例性藥物遞送裝置和方法係由麥肯納（McKenna）等人（美國公佈號2006/0030837）和雅各森（Jacobsen）等人（美國專利案號7,206,639）描述的。在一些實施方式中，導管或泵在手術程序過程中可以定位在例如患者的耳朵（例如，外耳、中耳和/或內耳）。在一些實施方式中，導管或泵在不需要進行手術程序的情況下可以定位在例如患者的耳朵（例如，外耳、中耳和/或內耳）。

[1239] 可替代地，一種或多種在此所述的化合物可以與戴在外耳的機械裝置諸如電子耳蝸或助聽器結合給予。適用於本發明的示例性電子耳蝸係由埃奇（Edge）等人（美國公開號

2007/0093878）描述的。

[1240] 在一些實施方式中，以上所述給藥模式可以任何順序組合並且可以是同時的或交替的。

[1241] 可替代地或另外地，本發明可以根據食品與任何藥品管理局批准的方法來給予，例如，如CDER資料標準手冊（CDER Data Standards Manual）版本號004（在fda.give/cder/dsm/DRG/drg00301.htm處可獲得）所述的。

[1242] 總的來說，美國專利申請20120328580所述的細胞治療方法可以用於促進細胞體外完全分化或部分分化成或分化為內耳成熟細胞類型（例如，毛細胞）。由此類方法獲得的細胞然後可以移植或植入到需要此治療的患者中。以下描述了實踐該等方法所需要的細胞培養方法，包括用於鑒定和選擇適合細胞類型的方法、用於促進所選擇細胞完全分化或部分分化的方法、以及用於植入完全或部分分化的細胞的方法。

[1243] 適用於本發明的細胞包括但不限於，當例如在體外與一種或多種在此所述的化合物接觸時能夠完全分化或部分分化成內耳成熟細胞例如毛細胞（例如，內和/或外毛細胞）的細胞。能夠分化成毛細胞的示例性細胞包括但不限於，幹細胞（例如，內耳幹細胞、成體幹細胞、骨髓源性幹細胞、胚胎幹細胞、間充質幹細胞、皮膚乾細胞、iPS細胞、以及脂肪源性幹細胞）、祖細胞（例如，內耳祖細胞）、支援細胞（例如，戴特斯細胞（Deiters' cell）、柱細胞、內指細胞、覆蓋細胞（tectal cell）以及漢森細胞（Hensen's cell））、和/或生殖細胞。李等人（美國公開號2005/0287127）和李

等人（美國專利序號11/953,797）描述了使用幹細胞替換內耳感覺細胞。在埃奇等人PCT/US2007/084654中描述使用骨髓源性幹細胞替換內耳感覺細胞。在例如高橋等人，細胞，第131卷，第5期，第861-872頁(2007)；高橋和山中，細胞 126, 663-76 (2006)；沖田（Okita）等人，自然 448:260-262（2007）；餘，J.等人，科學318(5858):1917-1920 (2007)；納卡加瓦（Nakagawa）等人，自然生物技術26:101-106 (2008)；以及卓瑞斯（Zaehres）和肖勒（Scholer），細胞 131(5):834-835 (2007)。此類適合的細胞可以是藉由分析（例如，定性或定量）一種或多種組織特異性基因的存在來鑒定的。例如，基因表現可以是藉由檢測一種或多種組織特異性基因的蛋白質產物來檢測的。蛋白質檢測技術涉及使用針對適當抗原的抗體染色蛋白質（例如，使用細胞提取物或全細胞）。在此情況下，適當抗原係組織特異性基因表現的蛋白質產物。儘管，理論上，第一抗體（即，結合抗原的抗體）可以被標記，但是更常見的是（並且提高視覺化）使用針對該第一抗體的第二抗體（例如，抗IgG）。此第二抗體與螢光染料或用於比色反應的適當的酶或金珠（用於電子顯微術）或者與生物素-抗生物素蛋白系統軛合，以使得第一抗體的位置以及因此抗原的位置可以被識別。

[1244] 本發明的CRISPR Cas分子可以藉由將藥物組成物直接應用於外耳來遞送到耳朵，其中組成物藉由美國公開的申請20110142917來修改。在一些實施方式中，將藥物組成物應用於耳道。遞送到耳朵還可以被稱為耳朵遞送或耳部遞送。

[1245] 在一些實施方式中，本發明的RNA分子以脂質體或脂

轉染配製物等遞送並且可以是藉由熟習該項技術者已熟知的方法來製備。此類方法描述於例如美國專利案號5,593,972、5,589,466、以及5,580,859中，該等專利藉由引用結合在此。

[1246] 已開發了特別旨在增強並改進siRNA到哺乳動物細胞的遞送的遞送系統（例如，參見，沈等人，FEBS快報 2003, 539:111-114；夏（Xia）等人，自然生物技術2002, 20:1006-1010；賴希（Reich）等人，分子視覺（Mol.Vision.）2003, 9: 210-216；索倫森（Sorensen）等人，分子生物學雜誌2003, 327: 761-766；路易士（Lewis）等人，自然遺傳學2002, 32: 107-108以及西梅奧尼（Simeoni）等人，核酸研究（NAR） 2003, 31, 11: 2717-2724）並且該等遞送系統可以適用於本發明。siRNA最近已成功用於抑制靈長類動物中的基因表現（參見，例如托倫蒂諾（Tolentino）等人，視網膜（Retina） 24(4):660，該文獻也可以適用於本發明）。

[1247] 齊（Qi）等人揭露了用於藉由可以適用於本發明的核酸靶向系統的新型蛋白遞送技術來經由完整圓窗有效siRNA轉染到內耳中（參見，例如，齊等人，基因治療 (2013), 1-9）。具體地說，可以藉由完整圓窗滲透將Cy3-標記的siRNA轉染到內耳（包括內外毛細胞）、壺腹脊、橢圓囊斑以及球囊斑的細胞中的TAT雙股RNA結合結構域（TAT-DRBD）成功用於體內遞送雙股siRNA，以用於治療各種內耳疾病並且保護聽力功能。可以考慮約40 μl 10 mM RNA作為給予至耳朵的劑量。

[1248] 根據雷亞利（Rejali）等人（聽覺研究（Hear Res.），2007年6月；228(1-2):180-7），電子耳蝸功能可以是藉由良好地保

留螺旋神經節神經元來提高的，該等神經元係電子耳蝸電刺激的靶標，並且腦源性神經營養因子（BDNF）先前已顯示增強實驗性耳聾耳朵中存活的螺旋神經節。雷亞利等人測試了改進的電子耳蝸電極設計，該設計包括藉由具有BDNF基因插入物的病毒載體轉導的成纖維細胞塗層。為了完成這種類型的離體基因轉移，雷亞利等人使用具有BDNF基因盒插入物的腺病毒轉導豚鼠成纖維細胞，並且確定該等細胞分泌BDNF並且然後使BDNF分泌細胞經由瓊脂糖凝膠附接到電子耳蝸電極並且將電極植入在鼓階中。雷亞利等人確定BDNF表現電極當與對照電極相比時在植入48小時後能夠保留顯著更多的耳蝸基底轉彎處的螺旋神經節神經元，並且證實了將電子耳蝸與離體基因轉移組合用於增強螺旋神經節神經元存活的可行性。這種系統可以適用於本發明的核酸靶向系統，以用於遞送到耳朵。

[1249] 穆克埃爾西（Mukherjea）等人（抗氧劑與氧化還原信號（Antioxidants & Redox Signaling），第13卷，第5期，2010）用文獻證明使用短干擾（si）RNA敲低NOX3消除了順鉑耳毒性，如藉由防止OHC損傷並且減小聽腦幹反應（ABR）中的閾值移位來證實的。向大鼠給予不同劑量的siNOX3（0.3、0.6、以及0.9 μg）並且藉由即時RT-PCR評價NOX3表現。所使用的最低劑量的NOX3 siRNA（0.3 μg）當與經鼓膜給予亂序siRNA或未處理的耳蝸相比時並未顯示NOX3 mRNA的任何抑制。然而，與對照的亂序siRNA相比，給予較高劑量的NOX3 siRNA（0.6和0.9 μg）減少了NOX3表現。這種系統可以適用於本發明的CRISPR Cas系統，以用於經鼓膜給予約2 mg至約4 mg劑量的CRISPR Cas，以給予至人類。

[1250] 榮格（Jung）等人（分子治療，第21卷，第4期，834–841，2013年4月）證實橢圓囊中的Hes5水平在應用siRNA之後有所減小並且該等橢圓囊中的毛細胞數目顯著高於隨後的對照治療。資料表明siRNA技術可以適用於誘導內耳的修復和再生並且Notch傳訊途徑係對於特異性基因表現抑制可能有用的靶標。榮格等人將2 μl體積的8 μg Hes5 siRNA（藉由將無菌生理鹽水添加到凍幹siRNA中來製備）注射到耳朵的前庭上皮中。這種系統可以適用於本發明的CRISPR Cas系統，以用於將約1 mg至約30 mg劑量的CRISPR Cas給予到耳朵的前庭上皮中，以給予至人類。

不分裂細胞中的基因靶向（神經元和肌肉）

[1251] 不分裂（尤其是不分裂的完全分化）細胞類型存在例如基因靶向或基因組工程化的問題，因為同源重組（HR）在G1細胞週期階段通常受到抑制。然而，當研究細胞控制正常DNA修復系統的機制時，迪羅謝（Durocher）發現了使得HR在不分裂細胞中保持"關閉"的先前未知的開關並且設計了將此開關切換回來的策略。奧爾特威恩（Orthwein）等人（在加拿大渥太華的西奈山醫院的丹尼爾·迪羅謝實驗室（Daniel Durocher's lab at the Mount Sinai Hospital in Ottawa, Canada）最近的報導（自然 16142，2015年12月9日線上公開）顯示HR的抑制可以取消並且基因靶向在腎臟（293T）細胞和骨肉瘤（U2OS）細胞中成功地推斷。已知腫瘤抑制基因BRCA1、PALB2和BRAC2促進HR進行DNA DSB修復。發現BRCA1與PALB2-BRAC2的複合物形成係藉由PALB2上的泛素位點來控制的，以使得E3泛素連接酶對該位點起作用。此E3泛素

連接酶係由與滯蛋白-3（CUL3）-RBX1複合的KEAP1（PALB2相互作用蛋白）組成。PALB2泛素化抑制了它與BRCA1的相互作用並且被去泛素化酶USP11抵消，該去泛素化酶自身處於細胞週期控制下。BRCA1-PALB2相互作用的恢復與DNA末端切除的啟動組合足以誘導G1中的同源重組，如藉由許多方法測量的，該等方法包括針對USP11或KEAP1（由pX459載體表現）的基於CRISPR-Cas9的基因靶向測定。然而，當在切除-感受態G1細胞中BRCA1–PALB2相互作用使用KEAP1缺失或PALB2-KR突變體的表現來恢復時，檢測到基因靶向事件的強勁增加。

[1252] 因此，在一些實施方式中，在細胞，尤其是不分裂的完全分化細胞類型中的HR重啟動係較佳的。在一些實施方式中，BRCA1–PALB2相互作用的促進在一些實施方式中是較佳的。在一些實施方式中，靶細胞係不分裂細胞。在一些實施方式中，靶細胞係神經元或肌細胞。在某些實施方式中，靶細胞係在體內靶向的。在一些實施方式中，細胞係在G1中並且HR被抑制。在一些實施方式中，使用KEAP1缺失，例如抑制KEAP1活性的表現係較佳的。KEAP1缺失可以是藉由siRNA來實現的，例如，如奧爾特威恩等人所述的。可替代地，PALB2-KR突變體（在BRCA1-相互作用結構域中不存在所有八個Lys殘基）的表現係較佳的，它與KEAP1缺失組合或者單獨。PALB2-KR與BRCA1相互作用，不論係在哪個細胞週期位置。因此，在一些實施方式中，BRCA1-PALB2相互作用的促進或恢復（尤其在G1細胞中）係較佳的，尤其在靶細胞係不分裂的情況下或者在去除和還原（離體基因靶向）係有問題的（例如神經元或肌肉細胞）的情況下。KEAP1 siRNA係從賽默飛

世爾公司（ThermoFischer）獲得的。在一些實施方式中，BRCA1-PALB2複合物可以被遞送到G1細胞。在一些實施方式中，可以例如藉由增加去泛素化酶USP11的表現來促進PALB2去泛素化，因此設想的是可以提供構建體來促進或上調去泛素化酶USP11的表現或活性。

治療眼睛疾病

[1253] 本發明還考慮將CRISPR-Cas系統遞送到一隻眼睛或兩隻眼睛。

[1254] 在本發明的特定實施方式中，CRISPR-Cas系統可以用於校正由幾種遺傳性突變引起的眼部缺陷，該等遺傳性突變在遺傳性眼睛疾病（Genetic Diseases of the Eye），第二版，埃利阿斯I.特拉布勒西（Elias I. Traboulsi），哈佛大學出版社，2012中進一步描述。

[1255] 對於給予至眼睛，慢病毒載體，具體地是馬傳染性貧血病毒（EIAV）係特別較佳的。

[1256] 在另一個實施方式中，還考慮基於馬傳染性貧血病毒（EIAV）的最小非靈長類動物慢病毒載體，特別是對於眼部基因治療（例如，參見巴拉岡，基因醫學雜誌 2006；8: 275-285，2005年11月21日威力出版公司（Wiley InterScience）線上公開（www.interscience.wiley.com）。DOI: 10.1002/jgm.845）。考慮具有驅動靶基因表現的巨細胞病毒（CMV）啟動子的載體。考慮所有前房內、視網膜下、眼內以及玻璃體內注射（例如，參見，巴拉

岡，基因醫學雜誌 2006；8: 275-285，2005年11月21日威力出版公司線上公開（www.interscience.wiley.com）。DOI: 10.1002/jgm.845）。可以借助於手術顯微鏡進行眼內注射。對於視網膜下注射和玻璃體內注射，眼睛可以藉由輕微數位壓力和使用接觸透鏡系統查看的眼底來脫垂，該接觸透鏡系統由在用玻璃顯微鏡載片蓋玻片覆蓋的角膜上的一滴偶合介質溶液組成。對於視網膜下注射，安裝在5-μl漢密爾頓氏注射器上的10-mm 34號針尖端可以在直接視覺化下藉由上部赤道鞏膜朝向後極切向推進，直到針孔在視網膜下間隙中可見為止。然後，可以注射2 μl載體懸浮液，以產生上部大泡狀視網膜脫離，因此證實視網膜下載體給予。此方法創建自封合鞏膜切開術，允許載體懸浮液保留在視網膜下間隙中，直到它被RPE吸收為止，通常在該程序的48 h內。此程序可以在下眼半球中重複，以產生下部視網膜脫離。此技術使得約70%視網膜神經感覺層和RPE暴露於載體懸浮液中。對於玻璃體內注射，針尖可以推進藉由鞏膜到角鞏膜緣後部1 mm，並且將2 μl載體懸浮液注射到玻璃體腔中。對於前房內注射，針尖可以藉由角鞏膜緣穿刺，朝向角膜中央推進，並且可以注射2 μl載體懸浮液。對於前房內注射，針尖可以藉由角鞏膜緣穿刺，朝向角膜中央推進，並且可以注射2 μl載體懸浮液。該等載體可以$1.0\text{-}1.4 \times 10^{10}$或者$1.0\text{-}1.4 \times 10^{9}$轉導單位（TU）/ml注射。

[1257] 在另一個實施方式中，還考慮了RetinoStat®，一種經由視網膜下注射遞送用於治療濕型年齡相關性黃斑變性的、表現血管生成抑制性蛋白（內皮抑素和血管抑素）的基於馬傳染性貧血病毒的慢病毒基因治療載體（例如，參見，賓利（Binley）等人，

人類基因治療 23:980–991(2012年9月))。這種載體可以被修飾用於本發明的CRISPR-Cas系統。每隻眼睛可以用總體積100 μl的1.1 × 10^5轉導單位/眼睛（TU/眼睛）的劑量的RetinoStat®治療。

[1258] 在另一個實施方式中，可以考慮將E1-缺失、部分E3-缺失、E4-缺失腺病毒載體遞送至眼睛。向患有晚期新生血管性年齡相關性黃斑變性（AMD）的二十八位患者給予表現人類色素上皮細胞源性因子（AdPEDF.11）的E1-缺失、部分E3-缺失、E4-缺失腺病毒載體的單次靜脈內注射（參見，例如，坎波基亞羅（Campochiaro）等人，人類基因治療 17:167-176(2006年2月))。研究10^6至$10^{9.5}$粒子單位（PU）範圍內的劑量，並且不存在與AdPEDF.11相關的嚴重不良事件並且不存在劑量限制性毒性（參見，例如，坎波基亞羅等人，人類基因治療 17:167-176(2006年2月))。腺病毒載體介導的眼部基因轉移似乎係一種用於治療眼部病症的可行方法並且可以適用於CRISPR Cas系統。

[1259] 在另一個實施方式中，RXi製藥公司（RXi Pharmaceuticals）的sd-rxRNA®系統可以用於和/或適於將CRISPR Cas遞送至眼睛。在此系統中，單次玻璃體內給予3 μg sd-rxRNA，導致PPIB mRNA水平的序列特異性減小，持續14天。sd-rxRNA®系統可以適用於本發明的核酸靶向系統，考慮CRISPR給予人類的約3至20 mg的劑量。

[1260] 米林頓-瓦爾德（Millington-Ward）等人（分子治療，第19卷，第4期，642–649，2011年4月）描述了將基於RNA干擾（RNAi）的視紫紅質抑制物和由於變性位置處的核苷酸變化而抵

抗抑制的密碼子修飾的視紫紅質替換基因遞送到RNAi靶位點的腺相關病毒（AAV）載體。藉由米林頓-瓦爾德等人將6.0×10^8 vp或1.8×10^{10} vp AAV注射液視網膜下注射到眼睛。米林頓-瓦爾德等人的AAV載體可以適用於本發明的CRISPR Cas系統，考慮給予至人類的約2×10^{11}至約6×10^{13} vp的劑量。

[1261] 達爾卡拉（Dalkara）等人（科學轉化醫學（Sci Transl Med） 5, 189ra76 (2013)）也涉及改變無害注射到眼睛玻璃體液中之後將野生型版本的缺陷基因遞送到整個視網膜的AAV載體的體內定向進化。達爾卡拉描述了7聚體肽展示文庫和藉由DNA改組來自AAV1、2、4、5、6、8、以及9的cap基因來構建的AAV文庫。包裝在CAG或Rho啟動子下表現GFP的rcAAV文庫和rAAV載體，並且藉由定量PCR獲得抗去氧核糖核酸酶的基因組滴度。合併該等文庫，並且進行兩輪進化，每次進化由初始文庫多樣性和隨後的三次體內選擇步驟組成。在每個此步驟中，向P30 rho-GFP小鼠玻璃體內注射2 ml碘克沙醇純化的磷酸鹽緩衝鹽水（PBS）透析文庫，其中基因組滴度係約1×10^{12} vg/ml。達爾卡拉等人的AAV系統可以適用於本發明的核酸靶向系統，考慮給予人類的約1×10^{15}至約1×10^{16} vg/ml的劑量。

[1262] 在一特定實施方式中，可以靶向視紫紅質基因以用於治療色素性視網膜炎（RP），其中轉讓給桑加莫生物科技公司的美國專利案號20120204282的系統可以根據本發明的CRISPR Cas系統進行修改。

[1263] 在另一個實施方式中，轉讓給策勒克提斯公司的美國

專利公開案號20130183282的方法係涉及切割來自人類視紫紅質基因的靶序列的方法，該方法也可以被修改成本發明的核酸靶向系統。

[1264] 轉讓給中央研究院（Academia Sinica）的美國專利公開案號20130202678涉及用於治療視網膜病變和威脅視力的眼科病症的方法，該等方法涉及將Puf-A基因（該基因在視網膜神經節和眼組織的色素細胞中表現並且展示獨特的抗細胞凋亡活性）遞送到眼睛的視網膜下間隙或玻璃體內間隙。具體地說，希望的靶標係zgc:193933、prdm1a、spata2、tex10、rbb4、ddx3、zp2.2、Blimp-1以及HtrA2，所有該等靶標均可以被本發明的核酸靶向系統靶向。

[1265] 吳（細胞幹細胞，13:659–62, 2013）設計了一種將Cas9引導到在小鼠中引起白內障的單一鹼基對突變的指導RNA，其中它誘導DNA切割。然後使用另一個野生型對偶基因或者給予受精卵的寡核苷酸，修復機制校正破壞的對偶基因的序列並且校正突變體小鼠中引起白內障的遺傳性缺陷。

[1266] 美國專利公開案號20120159653描述了使用鋅指核酸酶遺傳性修飾與黃斑變性（MD）相關聯的細胞、動物和蛋白質。黃斑變性（MD）係老年人視力缺損的主要原因，但是也是兒童疾病諸如眼底黃色斑點症、索斯比氏眼底病（Sorsby fundus）以及致命性兒童神經變性疾病的標誌性症狀，其中發作年齡最早到嬰兒期。黃斑變性由於視網膜損害而導致視野（黃斑）中心的視力喪失。目前存在的動物模型並未概括疾病在人類中所觀察到的主要標誌。包含編碼與MD相關聯的蛋白質的突變體基因的可用動物模

型也產生高度可變的表型，這使得對人類疾病的翻譯和治療發展成為問題。

[1267] 美國專利公開案號20120159653的一方面涉及編輯編碼與MD相關聯的蛋白質的任何染色體序列，它們可以適用於本發明的核酸靶向系統。與MD相關聯的蛋白質典型地是基於和MD相關聯的蛋白質與MD病症的實驗相關性來選擇。例如，與MD相關聯的蛋白質的產生率或循環濃度在患有MD病症的群體中相對於不存在MD病症的群體有所升高或降低。蛋白質水平的差異可以適於蛋白質組學技術來評估，該等技術包括但不限於，西方墨點法、免疫組織化學染色、酶聯免疫吸附測定（ELISA）、以及質譜法。可替代地，與MD相關聯的蛋白質可以是藉由使用基因組技術獲得編碼該等蛋白質的基因的基因表現譜來鑒定，該等基因組技術包括但不限於DNA微陣列分析、基因表現序列分析（SAGE）、以及定量即時聚合酶鏈反應（Q-PCR）。

[1268] 作為非限制性實例，與MD相關聯的蛋白質包括但不限於以下蛋白質：（ABCA4）ATP-結合盒，子家族A（ABC1），成員4；ACHM1 全色盲（桿狀細胞單色型色覺（rod monochromacy））1；ApoE 載脂蛋白E（ApoE）；C1QTNF5（CTRP5）C1q和腫瘤壞死因子相關蛋白5（C1QTNF5）；C2補體組分2（C2）；C3補體組分（C3）；CCL2趨化因子（C-C模體）配位基2（CCL2）；CCR2趨化因子（C-C模體）受體2（CCR2）；CD36分化抗原簇36；CFB補體因子B；CFH補體因子；CFH H；CFHR1補體因子H相關1；CFHR3補體因子H相關3；CNGB3環狀核苷酸閘控通道β3；CP血漿銅藍蛋

白（CP）；CRP C反應蛋白（CRP）；CST3 抑半胱胺酸蛋白酶蛋白C或抑半胱胺酸蛋白酶蛋白3（CST3）；CTSD組織蛋白酶D（CTSD）；CX3CR1趨化因子（C-X3-C模體）受體1；ELOVL4伸長的極長鏈脂肪酸4；ERCC6切補修復交叉互補齧齒動物修復缺陷，互補群6；FBLN5腓骨蛋白-5；FBLN5腓骨蛋白5；FBLN6腓骨蛋白6；FSCN2成束蛋白（FSCN2）；HMCN1半椎蛋白（Hemicentrin）1；HMCN1半椎蛋白1；HTRA1 HtrA絲胺酸肽酶1（HTRA1）；HTRA1 HtrA絲胺酸肽酶1；IL-6介白素6；IL-8介白素8；LOC387715假定蛋白；PLEKHA1含普列克底物蛋白同源結構域家族A成員1（PLEKHA1）；PROM1 Prominin 1（PROM1或CD133）；PRPH2外週蛋白-2；RPGR色素性視網膜炎GTP酶調節劑；SERPING1絲胺酸蛋白酶抑制劑肽酶抑制劑，進化支（clade）G，成員1（C1抑制劑）；TCOF1糖蜜TIMP3金屬蛋白酶抑制劑3（TIMP3）；TLR3 Toll樣受體3。

[1269] 與MD相關聯且其染色體序列被編輯的蛋白質的性質可以並且將發生改變。在較佳的實施方式中，與MD相關聯且其染色體序列被編輯的蛋白質可以是由ABCR基因編碼的ATP結合盒子家族A（ABC1）成員4蛋白（ABCA4）、由APOE基因編碼的載脂蛋白E蛋白（APOE）、由CCL2基因編碼的趨化因子（C-C模體）配位基2蛋白（CCL2）、由CCR2基因編碼的趨化因子（C-C模體）受體2蛋白（CCR2）、由CP基因編碼的血漿銅藍蛋白（CP）、由CTSD基因編碼的組織蛋白酶D蛋白（CTSD）、或者由TIMP3基因編碼的金屬蛋白酶抑制劑3蛋白質（TIMP3）。在一示例性實施方式中，遺傳修飾的動物係大鼠，並且編碼與MD相關聯的蛋白質的編輯的染色體序列可以是：（ABCA4）ATP結合盒子家族A（ABC1）成員4

NM_000350；APOE載脂蛋白E（APOE） NM_138828；CCL2趨化因子（C-C模體）配位基2（CCL2） NM_031530；CCR2趨化因子（C-C模體）受體2（CCR2） NM_021866 ；CP血漿銅藍蛋白（CP） NM_012532；CTSD組織蛋白酶D（CTSD） NM_134334；TIMP3金屬蛋白酶抑制劑3（TIMP3） NM_012886 。動物或細胞可以包含編碼與MD相關聯的蛋白質的1、2、3、4、5、6、7或更多個破壞的染色體序列和編碼與MD相關聯的破壞蛋白質的0、1、2、3、4、5、6、7或更多個染色體整合序列。

[1270] 編輯或整合的染色體序列可以被修飾為編碼與MD相關聯的改變的蛋白質。在MD相關染色體序列中的許多突變已與MD相關聯。與MD相關聯的染色體序列中的突變的非限制性實例包括可以引起MD的那些突變，包括在ABCR蛋白質，E471K（即在位置471處的穀胺酸被改變成賴胺酸）、R1129L（即在位置1129處的精胺酸被改變成亮胺酸）、T1428M（即在位置1428處的蘇胺酸被改變成甲硫胺酸）、R1517S（即在位置1517處的精胺酸被改變成絲胺酸）、I1562T（即在位置1562處的異亮胺酸被改變成蘇胺酸）、以及G1578R（即在位置1578處的甘胺酸被改變成精胺酸）；在CCR2蛋白質中，V64I（即在位置192處的纈胺酸被改變成異亮胺酸）；在CP蛋白質中，G969B（即在位置969處的甘胺酸被改變成天冬醯胺或天冬胺酸）；在TIMP3蛋白質，S156C（即在位置156處的絲胺酸被改變成半胱胺酸）、G166C（即在位置166處的甘胺酸被改變成半胱胺酸）、G167C（即在位置167處的甘胺酸被改變成半胱胺酸）、Y168C（即在位置168處的酪胺酸被改變成半胱胺酸）、S170C（即在位置170處的絲胺酸被改變成半胱胺酸）、Y172C（即

在位置172處的酪胺酸被改變成半胱胺酸）、以及S181C（即在位置181處的絲胺酸被改變成半胱胺酸）。MD相關基因的遺傳性變型與疾病的其他相關性係本領域已知的。

[1271] CRISPR系統適用於校正由常染色體顯性基因引起的疾病。例如，使用CRISPR/Cas9去除引起眼睛受體喪失的常染色體顯性基因。巴科迪，B.（Bakondi, B.）等人，體內CRISPR/Cas9基因編輯常染色體顯性色素性視網膜炎的S334ter-3大鼠模型中的視網膜營養性萎縮症（*In Vivo CRISPR/Cas9 Gene Editing Corrects Retinal Dystrophy in the S334ter-3 Rat Model of Autosomal Dominant Retinitis Pigmentosa*）。分子治療，2015；DOI: 10.1038/mt.2015.220。

治療循環系統疾病和肌肉疾病

[1272] 本發明還考慮將在此所述的CRISPR-Cas系統例如Cpf1效應蛋白系統遞送到心臟。對於心臟，心肌向性腺相關病毒（AAVM）係較佳的，具體地是在心臟中顯示優先基因轉移的AAVM41（參見，例如，林-陽尕（Lin-Yanga）等人，美國國家科學院院刊，2009年3月10日，第106卷，第10期）。給藥可以是全身性的或局部的。對於全身性給藥，考慮約1-10 × 10^{14}個載體基因組的劑量。還參見，例如，艾拉裡奧（Eulalio）等人，(2012)自然 492: 376和蘇摩素塔拉姆（Somasuntharam）等人(2013)生物材料（Biomaterials） 34: 7790。

[1273] 例如美國專利公開案號20110023139描述了使用鋅指核酸酶遺傳性修飾與心血管疾病相關聯的細胞、動物和蛋白質。心血管疾病通常包括高血壓、心臟病發作、心力衰竭、以及卒中

和TIA。涉及心血管疾病的任何染色體序列或者由涉及心血管疾病的任何染色體序列編碼的蛋白質可以用於本揭露描述的方法中。心血管相關蛋白典型地是基於心血管相關蛋白與心血管疾病發展的實驗相關性來選擇。例如，心血管相關蛋白的產生率或循環濃度在患有心血管病症的群體中相對於不存在心血管病症的群體有所升高或降低。蛋白質水平的差異可以適於蛋白質組學技術來評估，該等技術包括但不限於，西方墨點法、免疫組織化學染色、酶聯免疫吸附測定（ELISA）、以及質譜法。可替代地，心血管相關蛋白可以是藉由使用基因組技術獲得編碼該等蛋白質的基因的基因表現譜來鑒定，該等基因組技術包括但不限於DNA微陣列分析、基因表現序列分析（SAGE）、以及定量即時聚合酶鏈反應（Q-PCR）。

[1274] 作為非限制性實例，染色體序列可以包括但不限於，IL1B（介白素1，β）、XDH（黃嘌呤脫氫酶）、TP53（腫瘤蛋白p53）、PTGIS（前列腺素12（前列環素）合酶）、MB（肌紅蛋白）、IL4（介白素4）、ANGPT1（血管生成素1）、ABCG8（ATP-結合盒，子家族G（WHITE），成員8）、CTSK（組織蛋白酶K）、PTGIR（前列腺素12（前列環素）受體（IP））、KCNJ11（鉀整流通道，子家族J，成員11）、INS（胰島素）、CRP（C-反應蛋白，穿透素相關）、PDGFRB（血小板源生長因子受體，β多肽）、CCNA2（週期蛋白A2）、PDGFB（血小板源生長因子β多肽（猴肉瘤病毒（v-sis）癌基因同源物））、KCNJ5（鉀整流通道，子家族J，成員5）、KCNN3（鉀中間體/小電導鈣啟動通道，子家族N，成員3）、CAPN10（鈣蛋白酶10）、PTGES（前列腺素E合酶）、ADRA2B（腎上腺素性，α-2B-，受體）、ABCG5

（ATP-結合盒，子家族G（WHITE）、成員5）、PRDX2（過氧化物還原酶2）、CAPN5（鈣蛋白酶5）、PARP14（聚(ADP-核糖)聚合酶家族，成員14）、MEX3C（mex-3同源物C（秀麗隱桿線蟲））、ACE血管緊張素I轉化酶（肽基-二肽酶A）1）、TNF（腫瘤壞死因子（TNF超家族，成員2））、IL6（介白素6（干擾素，β2））、STN（他汀類）、絲胺酸蛋白酶抑制劑E1（絲胺酸蛋白酶抑制劑肽酶抑制劑，進化支E（微管連接蛋白，纖溶酶原啟動物抑制劑1型）、成員1）、ALB（白蛋白）、ADIPOQ（脂聯素，含有C1Q和膠原蛋白結構域）、APOB（載脂蛋白B（包含Ag(x)抗原））、APOE（載脂蛋白E）、LEP（瘦蛋白）、MTHFR（5，10-亞甲基四氫葉酸還原酶（NADPH））、APOA1（載脂蛋白A-I）、EDN1（內皮素1）、NPPB（利鈉肽先質B）、NOS3（一氧化氮合酶3（內皮細胞））、PPARG（過氧化物酶體增殖物啟動受體γ）、PLAT（纖溶酶原啟動物，組織）、PTGS2（前列腺素-內過氧化物合酶2（前列腺素G/H合酶和環氧合酶））、CETP（膽固醇酯轉移蛋白，血漿）、AGTR1（血管緊張素II受體，1型）、HMGCR（3-羥基-3-甲基戊二醯-輔酶A還原酶）、IGF1（胰島素樣生長因子1（生長調節素C））、SELE（選擇素E）、REN（腎素）、PPARA（過氧化物酶體增殖物啟動受體α）、PON1（對氧磷酶1）、KNG1（激肽原1）、CCL2（趨化因子（C-C模體）配位基2）、LPL（脂蛋白連接酶）、VWF（血管假性血友病因子）、F2（凝血因子II（凝血酶））、ICAM1（細胞間黏附分子1）、TGFB1（轉化生長因子，β1）、NPPA（利鈉肽先質A）、IL10（介白素10）、EPO（紅細胞生成素）、SOD1（超氧化物歧化酶1，可溶）、VCAM1（血管細胞黏附分子1）、IFNG（干擾素，γ）、LPA（脂蛋白，Lp(a)）、MPO（髓過氧化物酶）、

ESR1（雌激素受體1）、MAPK1（絲裂原啟動蛋白激酶1）、HP（結合珠蛋白）、F3（凝血因子III（促凝血酶原激酶，組織因子））、CST3（半胱胺酸蛋白酶抑制劑C）、COG2（低聚高爾基體複合物組分2）、MMP9（基質金屬肽酶9（明膠酶B，92kDa明膠酶，92kDaIV型膠原蛋白酶））、絲胺酸蛋白酶抑制劑C1（絲胺酸蛋白酶抑制劑肽酶抑制劑，進化支C（抗凝血酶）、成員1）、F8（凝血因子VIII，促凝血組分）、HMOX1（血紅素氧化酶（脫環）1）、APOC3（載脂蛋白C-III）、IL8（介白素8）、PROK1（脫環1）、CBS（胱硫醚-β-合酶）、NOS2（一氧化氮合酶2，誘導型）、TLR4（toll-樣受體4）、SELP（選擇素P（顆粒膜蛋白140kDa，抗原CD62））、ABCA1（ATP-結合盒，子家族A（ABC1）、成員1）、AGT（血管緊張素原（絲胺酸蛋白酶抑制劑肽酶抑制劑，進化支A，成員8））、LDLR（低密度脂蛋白受體）、GPT（穀胺酸-丙酮酸轉胺酶（丙胺酸轉胺酶））、VEGFA（血管內皮生長因子A）、NR3C2（核受體子家族3，C組，成員2）、IL18（介白素18（干擾素-γ-誘導因子））、NOS1（一氧化氮合酶1（神經元））、NR3C1（核受體子家族3，C組，成員1（糖皮質激素受體））、FGB（纖維蛋白原β鏈）、HGF（肝細胞生長因子（肝細胞生成素A；散射因子））、IL1A（介白素1，α）、RETN（抵抗素）、AKT1（v-akt鼠胸腺瘤病毒癌基因同源物1）、LIPC（脂肪酶，肝臟）、HSPD1（熱休克60 kDa蛋白1（伴侶蛋白））、MAPK14（絲裂原啟動蛋白激酶14）、SPP1（分泌型磷蛋白1）、ITGB3（整聯蛋白，β3（血小板糖蛋白111a，抗原CD61））、CAT（過氧化氫酶）、UTS2（尿緊張素2）、THBD（血栓調節蛋白）、F10（凝血因子X）、CP（血漿銅藍蛋白（鐵氧化酶））、TNFRSF11B（腫瘤壞死因子受體超家族，

成員11b)、EDNRA(內皮素受體A型)、EGFR(表皮生長因子受體(成紅細胞白血病病毒(v-erb-b)癌基因同源物,鳥類))、MMP2(基質金屬肽酶2(明膠酶A,72 kDa明膠酶,72 kDaIV型膠原酶))、PLG(纖溶酶原)、NPY(神經肽Y)、RHOD(ras同源物基因家族,成員D)、MAPK8(絲裂原啟動蛋白激酶8)、MYC(v-myc髓細胞組織增生病毒癌基因同源物(鳥類))、FN1(纖連蛋白1)、CMA1(糜酶1,肥大細胞)、PLAU(纖溶酶原啟動物,尿激酶)、GNB3(鳥嘌呤核苷酸結合蛋白(G蛋白),β多肽3)、ADRB2(腎上腺素性,β-2-,受體,表面)、APOA5(載脂蛋白A-V)、SOD2(超氧化物歧化酶2,線粒體)、F5(凝血因子V(促凝血球蛋白原,易變因子))、VDR(維生素D(1,25-二羥基維生素D3)受體)、ALOX5(花生四烯酸5-脂加氧酶)、HLA-DRB1(主要組織相容性複合物,II類,DRβ1)、PARP1(聚(ADP-核糖)聚合酶1)、CD40LG(CD40配位基)、PON2(對氧磷酶2)、AGER(高級糖基化末端產物特異性受體)、IRS1(胰島素受體底物1)、PTGS1(前列腺素-內過氧化物合酶1(前列腺素G/H合酶和環氧合酶))、ECE1(內皮素轉化酶1)、F7(凝血因子VII(血清凝血素轉化加速物))、URN(介白素1受體拮抗劑)、EPHX2(環氧化物酶2,細胞質)、IGFBP1(胰島素-樣生長因子結合蛋白1)、MAPK10(絲裂原啟動蛋白激酶10)、FAS(Fas(TNF受體超家族,成員6))、ABCB1(ATP-結合盒,子家族B(MDR/TAP),成員1)、JUN(jun癌基因)、IGFBP3(胰島素-樣生長因子結合蛋白3)、CD14(CD14分子)、PDE5A(磷酸二酯酶5A,cGMP-特異性)、AGTR2(血管緊張素II受體,2型)、CD40(CD40分子,TNF受體超家族成員5)、LCAT(卵磷脂-膽固醇醯基轉移酶)、

CCR5（趨化因子（C-C模體）受體5）、MMP1（基質金屬肽酶1（間質膠原酶））、TIMP1（TIMP金屬肽酶抑制劑1）、ADM（腎上腺髓質素）、DYT10（肌張力障礙10）、STAT3（信號轉導物和轉錄啟動物3（急性期反應因子））、MMP3（基質金屬肽酶3（基質溶素1，前明膠酶））、ELN（彈性蛋白）、USF1（上游轉錄因子1）、CFH（補體因子H）、HSPA4（熱休克70 kDa蛋白4）、MMP12（基質金屬肽酶12（巨噬細胞彈性蛋白酶））、MME（膜金屬內肽酶）、F2R（凝血因子II（凝血酶）受體）、SELL（選擇素L）、CTSB（組織蛋白酶B）、ANXA5（膜聯蛋白A5）、ADRB1（腎上腺素性，β-1-，受體）、CYBA（細胞色素b-245，α多肽）、FGA（纖維蛋白原α鏈）、GGT1（γ-穀胺醯轉移酶1）、LIPG（脂肪酶，內皮）、HIF1A（低氧誘導因子1，α亞基（鹼性螺旋環螺旋轉錄因子））、CXCR4（趨化因子（C-X-C模體）受體4）、PROC（蛋白C（凝血因子滅活物Va和VIIIa））、SCARB1（清道夫受體B類，成員1）、CD79A（CD79a分子，免疫球蛋白-相關α）、PLTP（磷脂轉移蛋白）、ADD1（內收蛋白1(α)）、FGG（纖維蛋白原γ鏈）、SAA1（血清澱粉樣蛋白A1）、KCNH2（鉀電壓閘控通道，子家族H（eag-相關），成員2）、DPP4（二肽基-肽酶4）、G6PD（葡萄糖-6-磷酸脫氫酶）、NPR1（利鈉肽受體A/鳥苷酸環化酶A（心房利鈉肽受體A））、VTN（玻連蛋白）、KIAA0101（KIAA0101）、FOS（FBJ鼠骨肉瘤病毒癌基因同源物）、TLR2（toll-樣受體2）、PPIG（肽基脯胺醯異構酶G（親環蛋白G））、IL1R1（介白素1受體，I型）、AR（雄激素受體）、CYP1A1（細胞色素P450,家族1，子家族A，多肽1）、絲胺酸蛋白酶抑制劑A1（絲胺酸蛋白酶抑制劑肽酶抑制劑，進化支A（α-1抗蛋白酶，抗胰蛋

白酶），成員1）、MTR（5-甲基四氫葉酸-高半胱胺酸甲基轉移酶）、RBP4（視黃醇結合蛋白4，血漿）、APOA4（載脂蛋白A-IV）、CDKN2A（週期蛋白依賴性激酶抑制劑2A（黑色素瘤，p16，抑制CDK4））、FGF2（成纖維細胞生長因子2（鹼性））、EDNRB（內皮素受體B型）、ITGA2（整聯蛋白，α2（CD49B，VLA-2受體的α2亞基））、CABIN1（鈣調磷酸酶結合蛋白1）、SHBG（性激素-結合球蛋白）、HMGB1（高遷移率族蛋白1）、HSP90B2P（熱休克蛋白90 kDaβ（Grp94），成員2（假基因））、CYP3A4（細胞色素P450,家族3，子家族A，多肽4）、GJA1（間隙連接蛋白，α1，43 kDa）、CAV1（小窩蛋白1，小凹蛋白，22 kDa）、ESR2（雌激素受體2（ERβ））、LTA（淋巴毒素α（TNF超家族，成員1））、GDF15（生長分化因子15）、BDNF（腦源性神經營養因子）、CYP2D6（細胞色素P450,家族2，子家族D，多肽6）、NGF（神經生長因子（β多肽））、SP1（Sp1轉錄因子）、TGIF1（TGFB-誘導因子同源框1）、SRC（v-src肉瘤（施密特-魯平（Schmidt-Ruppin）A-2）病毒癌基因同源物（鳥類））、EGF（表皮生長因子（β-尿抑胃素））、PIK3CG（磷酸肌醇-3-激酶，催化型，γ多肽）、HLA-A（主要組織相容性複合物，I類，A）、KCNQ1（鉀電壓閘控通道，KQT-樣子家族，成員1）、CNR1（大麻素受體1（大腦））、FBN1（原纖蛋白1）、CHKA（膽鹼激酶α）、BEST1（斑萎蛋白1）、APP（澱粉樣蛋白β(A4)先質蛋白）、CTNNB1（鏈蛋白（鈣黏蛋白-相關蛋白）、β1，88 kDa）、IL2（介白素2）、CD36（CD36分子（血小板反應蛋白受體））、PRKAB1（蛋白激酶，AMP-啟動型，β1非催化型亞基）、TPO（甲狀腺過氧化物酶）、ALDH7A1（醛脫氫酶7家族，成員A1）、CX3CR1（趨化因子（C-X3-C模體）受體

1)、TH(酪胺酸羥化酶)、F9(凝血因子IX)、GH1(生長激素1)、TF(轉鐵蛋白)、HFE(血色素沈著)、IL17A(介白素17A)、PTEN(磷酸酶和張力蛋白同源物)、GSTM1(麩胱甘肽s-轉移酶mu1)、DMD(肌萎縮蛋白)、GATA4(GATA結合蛋白4)、F13A1(凝血因子XIII，A1多肽)、TTR(甲狀腺素運載蛋白)、FABP4(脂肪酸結合蛋白4，脂肪細胞)、PON3(對氧磷酶3)、APOC1(載脂蛋白C-I)、INSR(胰島素受體)、TNFRSF1B(腫瘤壞死因子受體超家族，成員1B)、HTR2A(5-羥色胺(血清素)受體2A)、CSF3(集落刺激因子3(粒細胞))、CYP2C9(細胞色素P450,家族2，子家族C，多肽9)、TXN(硫氧還蛋白)、CYP11B2(細胞色素P450,家族11，子家族B，多肽2)、PTH(甲狀旁腺素)、CSF2(集落刺激因子2(粒細胞-巨噬細胞))、 KDR(激酶插入結構域受體(III型受體酪胺酸激酶))、PLA2G2A(磷脂酶A2，IIA族(血小板，滑液))、B2M(β-2-微球蛋白)、THBS1(血小板反應蛋白1)、GCG(胰高血糖素)、RHOA(ras同源物基因家族，成員A)、ALDH2(醛脫氫酶2家族(線粒體))、TCF7L2(轉錄因子7-樣2(T-細胞特異性，HMG-框))、BDKRB2(緩激肽受體B2)、NFE2L2(核因子(紅細胞源性2)-樣2)、NOTCH1(Notch同源物1，易位-相關(果蠅))、UGT1A1(UDP葡萄糖醛酸轉移酶1家族，多肽A1)、IFNA1(干擾素，α1)、PPARD(過氧物酶體增生啟動受體δ)、SIRT1(去乙醯化酶(沈默交配型資訊調節2同源物) 1(釀酒酵母))、GNRH1(促性腺素釋放激素1(黃體化釋放激素))、PAPPA(妊娠相關血漿蛋白A，冠毛素1)、ARR3(抑制蛋白3，視黃醛(X-抑制蛋白))、NPPC(利鈉肽先質C)、AHSP(α血紅蛋白穩定蛋白)、PTK2(PTK2蛋白酪

胺酸激酶2）、IL13（介白素13）、MTOR（雷帕黴素機制靶標（絲胺酸/蘇胺酸激酶））、ITGB2（整聯蛋白，β2（補體組分3受體3和4亞基））、GSTT1（麩胱甘肽s-轉移酶θ1）、IL6ST（介白素6信號轉導蛋白（gp130，制癌蛋白M受體））、CPB2（羧基肽酶B2（血漿））、CYP1A2（細胞色素P450,家族1，子家族A，多肽2）、HNF4A（肝蛋白核因子4，α）、SLC6A4（溶質運載蛋白家族6（神經遞質轉運體，血清素），成員4）、PLA2G6（磷脂酶A2，VI組（細胞溶質，鈣不依賴性））、TNFSF11（腫瘤壞死因子（配位基）超家族，成員11）、SLC8A1（溶質運載蛋白家族8（納/鈣交換蛋白），成員1）、F2RL1（凝血因子II（凝血酶）受體-樣1）、AKR1A1（醛酮還原酶家族1，成員A1（醛還原酶））、ALDH9A1（醛脫氫酶9家族，成員A1）、BGLAP（骨髓γ-羧基穀胺酸（gla）蛋白）、MTTP（線粒體甘油三酯轉移蛋白）、MTRR（5-甲基四氫葉酸-高半胱胺酸甲基轉移酶還原酶）、SULT1A3（磺基轉移酶家族，細胞溶質，1A，苯酚優選型，成員3）、RAGE（腎腫瘤抗原）、C4B（補體組分4B（Chido血型）、P2RY12（嘌呤受體P2Y，G-蛋白偶聯，12）、RNLS（腎胺酶，FAD-依賴型胺氧化酶）、CREB1（cAMP反應元件結合蛋白1）、POMC（阿黑皮素原）、RAC1（ras-相關C3肉毒毒素底物1（rho家族，小GTP結合蛋白Rac1））、LMNA（核纖層蛋白NC）、CD59（CD59分子，補體調節蛋白）、SCN5A（鈉通道，電壓閘控，V型，α亞基）、CYP1B1（細胞色素P450,家族1，子家族B，多肽1）、MIF（巨噬細胞遷移抑制劑y因子（糖基化-抑制因子））、MMP13（基質金屬肽酶13（膠原酶3））、TIMP2（TIMP金屬肽酶抑制劑2）、CYP19A1（細胞色素P450,家族19，子家族A，多肽1）、CYP21A2（細胞色素P450,

家族21，子家族A，多肽2）、PTPN22（蛋白酪胺酸磷酸酶，非-受體型22（淋巴樣））、MYH14（肌球蛋白，重鏈14，非肌肉）、MBL2（甘露糖-結合凝集素（蛋白C）2，可溶（調理素缺陷））、SELPLG（選擇素P配位基）、AOC3（胺氧化酶，含銅3（血管黏附蛋白1））、CTSL1（組織蛋白酶L1）、PCNA（增生細胞核抗原）、IGF2（胰島素樣生長因子2（生長調節素A））、ITGB1（整聯蛋白，β1（纖連蛋白受體，β多肽，抗原CD29包括MDF2、MSK12））、CAST（鈣蛋白酶抑素）、CXCL12（趨化因子（C-X-C模體）配位基12（基質細胞衍生因子1））、IGHE（免疫球蛋白重結構域ε）、KCNE1（鉀電壓閘控通道，Isk-相關家族，成員1）、TFRC（轉鐵蛋白受體（p90，CD71））、COL1A1（膠原，I型，α1）、COL1A2（膠原，I型，α2）、IL2RB（介白素2受體，β）、PLA2G10（磷脂酶A2，X組）、ANGPT2（血管生成素2）、PROCR（蛋白C受體，內皮（EPCR））、NOX4（NADPH氧化酶4）、HAMP（鐵調素抗微生物肽）、PTPN11（蛋白酪胺酸磷酸酶，非受體型11）、SLC2A1（溶質運載蛋白家族2（促葡萄糖轉運體），成員1）、IL2RA（介白素2受體，α）、CCL5（趨化因子（C-C模體）配位基5）、IRF1（干擾素調節因子1）、CFLAR（CASP8和FADD樣細胞凋亡調節劑）、CALCA（降鈣素-相關多肽α）、EIF4E（真核細胞翻譯起始因子4E）、GSTP1（麩胱甘肽s-轉移酶pi1）、JAK2（Janus激酶2）、CYP3A5（細胞色素P450,家族3，子家族A，多肽5）、HSPG2（硫酸類肝素蛋白多糖2）、CCL3（趨化因子（C-C模體）配位基3）、MYD88（骨髓分化主反應基因（88））、VIP（血管活性腸肽）、SOAT1（甾醇O-醯基轉移酶1）、ADRBK1（腎上腺素性，β，受體激酶1）、NR4A2（核受體子家族4，A組，成員2）、

MMP8（基質金屬肽酶8（中性粒細胞膠原酶））、NPR2（利鈉肽受體B/鳥苷酸環化酶B（心房利鈉肽受體B））、GCH1（GTP環化水解酶1）、EPRS（穀醯基-脯胺醯-tRNA合成酶）、PPARGC1A（過氧物酶體增生啟動受體γ，共啟動物1α）、F12（凝血因子XII（接觸因子））、PECAM1（血小板/內皮細胞黏附分子）、CCL4（趨化因子（C-C模體）配位基4）、絲胺酸蛋白酶抑制劑A3（絲胺酸蛋白酶抑制劑肽酶抑制劑，進化支A（α-1抗蛋白酶，抗胰蛋白酶），成員3）、CASR（鈣敏感受體）、GJA5（間隙連接蛋，α5，40 kDa）、FABP2（脂肪酸結合蛋白2，腸）、TTF2（轉錄終止因子，RNA聚合酶II）、PROS1（蛋白S(α)）、CTF1（心肌營養蛋白1）、SGCB（肌聚糖，β（43 kDa肌萎縮蛋白-相關糖蛋白））、YME1L1（YME1樣1（釀酒酵母））、CAMP（抗菌肽抗微生物肽）、ZC3H12A（含鋅指CCCH-型12A）、AKR1B1（醛酮還原酶家族1，成員B1（醛糖還原酶））、DES（肌間線蛋白）、MMP7（基質金屬肽酶7（基質溶解因子，子宮））、AHR（芳香烴受體）、CSF1（集落刺激因子1（巨噬細胞））、HDAC9（組蛋白脫乙醯酶9）、CTGF（結締組織生長因子）、KCNMA1（大電導鈣啟動鉀通道，子家族M，α成員1）、UGT1A（UDP葡萄糖醛酸轉移酶1家族，多肽A複合物座位）、PRKCA（蛋白激酶C，α）、COMT（兒茶酚-.β.-甲基轉移酶）、S100B（S100鈣結合蛋白B）、EGR1（早期生長反應因子1）、PRL（催乳素）、IL15（介白素15）、DRD4（多巴胺受體D4）、CAMK2G（鈣/鈣調蛋白-依賴性蛋白激酶IIγ）、SLC22A2（溶質運載蛋白家族22（有機陽離子轉運體），成員2）、CCL11（趨化因子（C-C模體）配位基11）、PGF（B321胎盤生長因子）、THPO（促血小板生成素）、GP6（糖蛋白VI（血

小板））、TACR1（速激肽受體1）、NTS（神經降壓素）、HNF1A（HNF1同源框A）、SST（生長激素抑制素）、KCND1（鉀電壓閘控通道，Shal-相關子家族，成員1）、LOC646627（磷脂酶抑制劑）、TBXAS1（凝血脂素A合酶1（血小板））、CYP2J2（細胞色素P450,家族2，子家族J，多肽2）、TBXA2R（凝血脂素A2受體）、ADH1C（醇脫氫酶1C（I類）、γ多肽）、ALOX12（花生四烯酸12-脂加氧酶）、AHSG（α-2-HS-糖蛋白）、BHMT（甜菜鹼-高半胱胺酸甲基轉移酶）、GJA4（間隙連接蛋，α4，37 kDa）、SLC25A4（溶質運載蛋白家族25（線粒體運載蛋白；腺嘌呤核苷酸轉運體），成員4）、ACLY（ATP檸檬酸裂合酶）、ALOX5AP（花生四烯酸5-脂加氧酶-啟動蛋白）、NUMA1（核有絲分裂器蛋白1）、CYP27B1（細胞色素P450,家族27，子家族B，多肽1）、CYSLTR2（半胱胺醯白三烯受體2）、SOD3（超氧化物歧化酶3，細胞外）、LTC4S（白三烯C4合酶）、UCN（尿皮素）、GHRL（饑餓素/肥胖抑制素前多肽原）、APOC2（載脂蛋白C-II）、CLEC4A（C-型凝集素結構域家族4，成員A）、KBTBD10（含有kelch重複和BTB（POZ）結構域10）、TNC（腱生蛋白C）、TYMS（胸苷酸合酶）、SHC1（SHC（含有Src同源物y2結構域）轉化蛋白1）、LRP1（低密度脂蛋白受體-相關蛋白1）、SOCS3（細胞介素傳訊抑制物3）、ADH1B（醇脫氫酶1B（I類）、β多肽）、KLK3（激肽釋放酶-相關肽酶3）、HSD11B1（羥基類固醇(11-β)脫氫酶1）、VKORC1（維生素K環氧化物還原酶複合物，亞基1）、絲胺酸蛋白酶抑制劑B2（絲胺酸蛋白酶抑制劑肽酶抑制劑，進化支B（卵白蛋白），成員2）、TNS1（張力蛋白1）、RNF19A（環指蛋白19A）、EPOR（紅細胞生成素受體）、ITGAM（整聯蛋白，αM（補體組分3受體

3亞基））、PITX2（成對樣同源結構域2）、MAPK7（絲裂原啟動蛋白激酶7）、FCGR3A（IgG的Fc片段，低親和力111a，受體（CD16a））、LEPR（瘦蛋白受體）、ENG（內皮因子）、GPX1（麩胱甘肽過氧化物酶1）、GOT2（穀胺酸草醯轉胺酶2，線粒體（天冬胺酸胺基轉移酶2））、HRH1（組胺受體H1）、NR112（核受體子家族1，I組，成員2）、CRH（促腎上腺激素釋放激素）、HTR1A（5-羥色胺（血清素）受體1A）、VDAC1（電壓-依賴性陰離子通道1）、HPSE（乙醯肝素酶）、SFTPD（表面活性劑蛋白D）、TAP2（轉運體2，ATP-結合盒，子家族B（MDR/TAP））、RNF123（環指蛋白123）、PTK2B（PTK2B蛋白酪胺酸激酶2β）、NTRK2（神經營養性酪胺酸激酶，受體，2型）、IL6R（介白素6受體）、ACHE（乙醯膽鹼酯酶（Yt血型））、GLP1R（胰高血糖素樣肽1受體）、GHR（生長激素受體）、GSR（麩胱甘肽還原酶）、NQO1（NAD(P)H脫氫酶，醌1）、NR5A1（核受體子家族5，A組，成員1）、GJB2（間隙連接蛋，β2，26 kDa）、SLC9A1（溶質運載蛋白家族9（鈉/氫交換蛋白），成員1）、MAOA（單胺氧化酶A）、PCSK9（前蛋白轉化酶枯草桿菌蛋白酶/kexin9型）、FCGR2A（IgG的Fc片段，低親和力IIa，受體（CD32））、絲胺酸蛋白酶抑制劑F1（絲胺酸蛋白酶抑制劑肽酶抑制劑，進化支F（α-2抗纖維蛋白溶素，色素上皮源性因子），成員1）、EDN3（內皮素3）、DHFR（二氫葉酸還原酶）、GAS6（生長停滯-特異性6）、SMPD1（鞘磷脂磷酸二酯酶1，酸性溶酶體）、UCP2（解偶聯蛋白2（線粒體，質子運載蛋白））、TFAP2A（轉錄因子AP-2α（啟動增強子結合蛋白2α））、C4BPA（補體組分4結合蛋白，α）、絲胺酸蛋白酶抑制劑F2（絲胺酸蛋白酶抑制劑肽酶抑制劑，進化支F（α-2

抗纖維蛋白溶素，色素上皮源性因子），成員2）、TYMP（胸苷磷酸化酶）、ALPP（鹼性磷酸酶，胎盤（Regan同工酶））、CXCR2（趨化因子（C-X-C模體）受體2）、SLC39A3（溶質運載蛋白家族39（鋅轉運體），成員3）、ABCG2（ATP-結合盒，子家族G（WHITE），成員2）、ADA（腺苷脫胺酶）、JAK3（Janus激酶3）、HSPA1A（熱休克70 kDa蛋白1A）、FASN（脂肪酸合酶）、FGF1（成纖維細胞生長因子1（酸性））、F11（凝血因子XI）、ATP7A（ATP酶，Cu++轉運，α多肽）、CR1（補體組分（3b/4b）受體1（Knops血型））、GFAP（膠質纖維酸性蛋白）、ROCK1（Rho-相關，含捲曲螺旋蛋白激酶1）、MECP2（甲基CpG結合蛋白2（蕾特氏綜合症））、MYLK（肌球蛋白輕鏈激酶）、BCHE（丁醯膽鹼酯酶）、LIPE（脂肪酶，激素敏感性）、PRDX5（過氧化物還原酶5）、ADORA1（腺苷A1受體）、WRN（維爾納綜合症，RecQ螺旋酶樣）、CXCR3（趨化因子（C-X-C模體）受體3）、CD81（CD81分子）、SMAD7（SMAD家族成員7）、LAMC2（核纖層蛋白in，γ2）、MAP3K5（絲裂原啟動蛋白激酶激酶激酶5）、CHGA（嗜鉻粒蛋白A（甲狀旁腺分泌蛋白1））、IAPP（胰島澱粉樣多肽）、RHO（視紫紅質）、ENPP1（核苷酸內焦磷酸酶/磷酸二酯酶1）、PTHLH（甲狀旁腺素樣激素）、NRG1（神經調節蛋白1）、VEGFC（血管內皮生長因子C）、ENPEP（穀醯基胺基肽酶（胺基肽酶A））、CEBPB（CCAAT/增強子結合蛋白（C/EBP）、β）、NAGLU（N-乙醯葡糖胺糖苷酶，α-）、F2RL3（凝血因子II（凝血酶）受體樣3）、CX3CL1（趨化因子（C-X3-C模體）配位基1）、BDKRB1（緩激肽受體B1）、ADAMTS13（具有血小板反應蛋白1型模體的ADAM金屬肽酶，13）、ELANE（彈性蛋白酶，

中性粒細胞表現）、ENPP2（核苷酸內焦磷酸酶/磷酸二酯酶2）、CISH（含有細胞介素誘導型SH2的蛋白）、GAST（胃泌素）、MYOC（肌纖蛋白，小梁網誘導型糖皮質激素反應）、ATP1A2（ATP酶，Na+/K+轉運，α2多肽）、NF1（神經纖維瘤蛋白1）、GJB1（間隙連接蛋，β1，32 kDa）、MEF2A（肌細胞增強因子2A）、VCL（黏著斑蛋白）、BMPR2（骨形態蛋白受體，II型（絲胺酸/蘇胺酸激酶））、TUBB（微管蛋白，β）、CDC42（細胞分裂週期42（GTP結合蛋白，25 kDa））、KRT18（角蛋白18）、HSF1（熱休克轉錄因子1）、MYB（v-myb成髓細胞血症病毒癌基因同源物（鳥類））、PRKAA2（蛋白激酶，AMP-啟動，α2催化型亞基）、ROCK2（Rho-相關，含捲曲螺旋蛋白激酶2）、TFPI（組織因子途徑抑制劑（脂蛋白-相關凝血抑制劑））、PRKG1（蛋白激酶，cGMP-依賴性，I型）、BMP2（骨形態蛋白2）、CTNND1（鏈蛋白（鈣黏蛋白-相關蛋白）、δ1）、CTH（胱胺醚酶（胱硫醚γ-裂合酶））、CTSS（組織蛋白酶S）、VAV2（vav2鳥嘌呤核苷酸交換因子）、NPY2R（神經肽Y受體Y2）、IGFBP2（胰島素樣生長因子結合蛋白2，36 kDa）、CD28（CD28分子）、GSTA1（麩胱甘肽s-轉移酶α1）、PPIA（肽基脯胺醯異構酶A（親環蛋白A））、APOH（載脂蛋白H（β-2-糖蛋白I））、S100A8（S100鈣結合蛋白A8）、IL11（介白素11）、ALOX15（花生四烯酸15-脂加氧酶）、FBLN1（腓骨蛋白1）、NR1H3（核受體子家族1，H組，成員3）、SCD（硬脂醯-CoA去飽和酶（δ-9-去飽和酶））、GIP（胃抑制劑y多肽）、CHGB（嗜鉻粒蛋白B（分泌粒蛋白1））、PRKCB（蛋白激酶C，β）、SRD5A1（類固醇-5-α-還原酶，α多肽1（3-氧代-5α-類固醇δ4-脫氫酶α1））、HSD11B2（羥基類固醇(11-β)脫氫酶2）、CALCRL

（降鈣素受體樣）、GALNT2（UDP-N-乙醯基-α-D-半乳糖胺:多肽N-乙醯基胺基半乳糖轉移酶2（GalNAc-T2））、ANGPTL4（血管生成素樣4）、KCNN4（鉀中間體/小電導鈣-啟動通道，子家族N，成員4）、PIK3C2A（磷酸肌醇-3-激酶，2類，α多肽）、HBEGF（肝素結合EGF樣生長因子）、CYP7A1（細胞色素P450,家族7，子家族A，多肽1）、HLA-DRB5（主要組織相容性複合物，II類，DRβ5）、BNIP3（BCL2/腺病毒E1B19 kDa相互作用蛋白3）、GCKR（葡糖激酶（己糖激酶4）調節劑）、S100A12（S100鈣結合蛋白A12）、PADI4（肽基精胺酸脫亞胺酶，IV型）、HSPA14（熱休克70 kDa蛋白14）、CXCR1（趨化因子（C-X-C模體）受體1）、H19（H19，母系印記表現的轉錄物（非蛋白編碼））、KRTAP19-3（角蛋白相關蛋白19-3）、IDDM2（胰島素-依賴性糖尿病2）、RAC2（ras-相關C3肉毒毒素底物2（rho家族，小GTP結合蛋白Rac2））、RYR1（利阿諾定（ryanodine）受體1（骨骼））、CLOCK（clock同源物（小鼠））、NGFR（神經生長因子受體（TNFR超家族，成員16））、DBH（多巴胺β-羥化酶（多巴胺β-單加氧酶））、CHRNA4（膽鹼能受體，煙鹼型，α4）、CACNA1C（鈣通道，電壓-依賴性，L型，α1C亞基）、PRKAG2（蛋白激酶，AMP-啟動，γ2非-催化型亞基）、CHAT（膽鹼乙醯轉移酶）、PTGDS（前列腺素D2合酶21 kDa（大腦））、NR1H2（核受體子家族1，H組，成員2）、TEK（TEK酪胺酸激酶，內皮）、VEGFB（血管內皮生長因子B）、MEF2C（肌細胞增強因子2C）、MAPKAPK2（絲裂原啟動蛋白激酶-啟動蛋白激酶2）、TNFRSF11A（腫瘤壞死因子受體超家族，成員11a，NFKB啟動物）、HSPA9（熱休克70 kDa蛋白9（致死蛋白））、CYSLTR1（半胱胺醯白三烯受體

1)、MAT1A(甲硫胺酸腺苷轉移酶I，α)、OPRL1(鴉片受體樣1)、IMPA1(肌醇(myo)-1(或4)-單磷酸酶1)、CLCN2(氯通道2)、DLD(二氫硫辛醯胺脫氫酶)、PSMA6(蛋白酶體(先質，巨蛋白因子)亞基，α型，6)、PSMB8(蛋白酶體(先質，巨蛋白因子)亞基，β型，8(大多功能肽酶7))、CHI3L1(殼多糖酶3樣1(軟骨糖蛋白-39))、ALDH1B1(醛脫氫酶1家族，成員B1)、PARP2(聚(ADP-核糖)聚合酶2)、STAR(生成類固醇的急性調節蛋白)、LBP(脂多糖結合蛋白)、ABCC6(ATP-結合盒，子家族C(CFTR/MRP)，成員6)、RGS2(G-蛋白傳訊調節劑2，24 kDa)、EFNB2(肝配蛋白-B2)、GJB6(間隙連接蛋，β6，30 kDa)、APOA2(載脂蛋白A-II)、AMPD1(腺苷單磷酸單磷酸1)、DYSF(dysferlin，肢帶肌肉萎縮症2B(常染色體隱性))、FDFT1(法呢基二磷酸法呢基轉移酶1)、EDN2(內皮素2)、CCR6(趨化因子(C-C模體)受體6)、GJB3(間隙連接蛋，β3，31 kDa)、IL1RL1(介白素1受體樣1)、ENTPD1(核苷酸內三磷酸酯二磷酸水解酶1)、BBS4(巴比二氏綜合症(Bardet-Biedlsyndrome)4)、CELSR2(鈣黏蛋白，EGFLAG七次跨膜G-型受體2(flamingo同源物，果蠅))、F11R(F11受體)、RAPGEF3(Rap鳥嘌呤核苷酸交換因子(GEF)3)、HYAL1(透明質酸胺基葡糖苷酶1)、ZNF259(鋅指蛋白259)、ATOX1(ATX1抗氧化劑蛋白1同源物(酵母))、ATF6(啟動轉錄因子6)、KHK(已酮糖激酶(果糖激酶))、SAT1(亞精胺/精胺N1-乙醯轉移酶1)、GGH(γ-穀醯基水解酶(軛合酶，葉醯聚γ穀醯基水解酶))、TIMP4(TIMP金屬肽酶抑制劑4)、SLC4A4(溶質運載蛋白家族4，碳酸氫鈉共轉運體，成員4)、PDE2A(磷酸二酯酶2A，cGMP-刺激)、

PDE3B（磷酸二酯酶3B，cGMP-抑制）、FADS1（脂肪酸去飽和酶1）、FADS2（脂肪酸去飽和酶2）、TMSB4X（胸腺素β4，X-連鎖）、TXNIP（硫氧還蛋白相互作用蛋白）、LIMS1（LIM和衰老細胞抗原樣結構域1）、RHOB（ras同源物基因家族，成員B）、LY96（淋巴細胞抗原96）、FOXO1（叉頭框O1）、PNPLA2（含馬鈴薯糖蛋白樣磷脂酶結構域2）、TRH（促甲狀腺激素釋放激素）、GJC1（間隙連接蛋，γ1，45 kDa）、SLC17A5（溶質運載蛋白家族17（陰離子/糖轉運體），成員5）、FTO（脂肪量和肥胖相關）、GJD2（間隙連接蛋，δ2，36 kDa）、PSRC1（脯胺酸/絲胺酸富集捲曲螺旋1）、CASP12（盒12（基因/假基因））、GPBAR1（G蛋白-偶聯膽汁酸受體1）、PXK（含PX結構域絲胺酸/蘇胺酸激酶）、IL33（介白素33）、TRIB1（tribbles同源物1（果蠅））、PBX4（前-B-細胞白血病同源框4）、NUPR1（核蛋白，轉錄調節劑，1）、15-Sep（15 kDa硒蛋白）、CILP2（軟骨中間層蛋白2）、TERC（端粒酶RNA組分）、GGT2（γ-穀胺醯轉移酶2）、MT-CO1（線粒體編碼的細胞色素c氧化酶I）、以及UOX（尿酸氧化酶，假基因）。任何該等序列可以是CRISPR-Cas系統的靶標,例如以處理突變。

[1275] 在另一個實施方式中，染色體序列可以是進一步選自Pon1（對氧磷酶1）、LDLR（LDL受體）、ApoE（載脂蛋白E）、Apo B-100（載脂蛋白B-100）、ApoA（載脂蛋白(a)）、ApoA1（載脂蛋白A1）、CBS（胱硫醚B-合酶）、糖蛋白IIb/IIb、MTHRF（5,10-亞甲四氫葉酸還原酶（NADPH）、以及其組合。在一次反覆運算中，染色體序列和由涉及心血管疾病的染色體序列編碼的蛋白可以是選自Cacna1C、Sod1、Pten、Ppar(alpha)、Apo E、Leptin、以及其

作為CRISPR-Cas系統的一個或多個靶標的組合。

治療肝臟和腎臟的疾病

[1276] 本發明還考慮將在此所述的CRISPR-Cas系統例如Cpf1效應蛋白系統遞送到肝臟和/或腎臟。誘導治療性核酸的細胞攝取的遞送策略包括物理力或載體系統，諸如基於病毒、脂質或複合物的遞送系統或奈米載體。從最初具有不太可能的臨床相關性的應用開始，當核酸使用全身性流體動力學高壓注射來發送到腎細胞時，廣泛範圍的基因治療病毒和非病毒載體已經用於靶向不同動物腎臟疾病模型的體內轉錄後事件（喬鮑·由裡夫斯（Csaba Révész）和皮特· 哈馬爾(Péter Hamar) (2011).靶向腎臟中的RNA的遞送方法（Delivery Methods to Target RNAs in the Kidney），基因治療應用(Gene Therapy Applications)，春生·康(Chunsheng Kang)教授（編輯），ISBN: 978-953-307-541-9, InTech, 可獲自：http://www.intechopen.com/books/gene-therapy-applications/delivery-methods-to-target-rnas-inthe-kidney)。到腎臟中的遞送方法可以包括袁（Yuan）等人（美國腎臟生理學雜誌(Am J Physiol Renal Physiol） 295: F605–F617, 2008）中所述的那些方法，他們研究了靶向花生四烯酸代謝的12/15-脂加氧酶(12/15-LO)途徑的小干擾RNA(siRNA)的體內遞送是否可以改善鏈脲黴素注射的1型糖尿病小鼠模型中的腎損傷和糖尿病腎病（DN)。為了實現更大的體內進入和腎臟中的siRNA表現，袁等人使用與膽固醇軛合的雙股12/15-LO siRNA寡核苷酸。將約400 μg siRNA皮下注射到小鼠中。袁等人的方法可以適用於本發明的CRISPR Cas系統，

考慮向人皮下注射1-2 g與膽固醇軛合的CRISPR Cas，以用於遞送到腎臟。

[1277] 莫里托裡斯(Molitoris)等人(美國腎臟病學會雜誌(J Am Soc Nephrol) 20: 1754–1764, 2009)利用近端小管細胞(PTC)作為腎臟內的寡核苷酸重吸收位點，以測試siRNA靶向細胞凋亡途徑中的關鍵蛋白p53的效率，從而防止腎臟損傷。在缺血性損傷後4 h靜脈內注射對於p53的合成的裸siRNA最大限度地保護了PTC和腎功能。莫里托裡斯等人的資料表明將siRNA快速遞送到近端小管細胞採用靜脈內注射。對於劑量反應性分析，用0.33；1、3或5 mg/kg劑量的siP53注射大鼠，在相同的四個時間點給予，分別產生累積劑量1.32；4、12以及20 mg/kg。與PBS處理的缺血對照大鼠相比較，所有測試的siRNA 劑量在第一天產生了SCr 降低作用，其中更高的劑量在經過大約五天中更有效12 mg/kg和20 mg/kg的累積劑量提供了最好的保護作用。莫里托裡斯等人的方法可以適用於本發明的CRISPR Cas系統，對於人類考慮用於遞送到腎臟的12和20 mg/kg累積。

[1278] 湯普森（Thompson）等人（核酸治療（Nucleic Acid Therapeutics），第22卷，第4期，2012）報導了在齧齒動物和非人類靈長類動物中靜脈內注射之後合成的小干擾RNA I5NP的毒物學特徵和藥代動力學特徵。I5NP被設計經由RNA干擾（RNAi）途徑作用，以暫時抑制促細胞凋亡蛋白p53的表現，並且被開發來防止細胞經受在主要心臟手術過程中可能出現的急性缺血/再灌注損傷諸如急性腎損傷以及在腎臟移植之後可能出現的移植物功能

延遲。在齧齒類中的800 mg/kg I5NP以及在非人類靈長動物中的1,000mg/kg I5NP的劑量對於引起不良作用係需要的，在猴中被分離為引導對血液的作用，包括補體的亞臨床啟動和凝血時間的輕度增加。在大鼠中，使用I5NP的大鼠類似物未觀察到另外的不良作用，這表明該等作用可能表示合成型RNA雙股體的分類作用，而不是與I5NP的預期藥理活性相關的毒性。總之，該等資料支援用於在急性缺血/再灌注損傷之後保留腎功能的I5NP的靜脈內給藥的臨床測試。在猴子中無觀察到的不良反應的水平（NOAEL）係500 mg/kg。在猴子中在以多至25 mg/kg的劑量水平靜脈內給藥之後未觀察到對心血管、呼吸和神經系統參數的作用。因此，對於向人類的腎臟靜脈內給予CRISPR Cas可以考慮類似劑量。

[1279] 清水（Shimizu）等人（美國腎臟病學會雜誌 21: 622–633, 2010）開發了經由基於聚(乙二醇)-聚(L-賴胺酸)的媒介物將siRNA靶向遞送到腎小球的系統。該siRNA/奈米載體複合物的直徑係約10至20 nm，該直徑係將允許它移動跨過穿孔內皮細胞而接近腎小球膜的大小。在腹膜內注射螢光標記的siRNA/ 奈米載體複合物之後，清水等人在血液循環中檢測到siRNA，持續一段延長的時間。在腎小球腎炎的小鼠模型中，絲裂原啟動蛋白激酶1（MAPK1）siRNA/ 奈米載體複合物的重複腹膜內給藥抑制了腎小球MAPK1 mRNA和蛋白質表現。為了研究siRNA累積，向BALBc小鼠給予與PIC奈米載體複合的Cy5標記的siRNA（0.5 ml，5 nmol的siRNA含量）、裸露的Cy5標記的siRNA（0.5 ml，5 nmol）、或封裝在HVJ-E 中的Cy5標記的siRNA（0.5 ml，5 nmol的siRNA含量）。清水等人的方法可以適用於本發明的CRISPR Cas系統，對於人類

考慮在約1-2升內約10-20 μmol與奈米載體複合的CRISPR Cas，以腹膜內給藥並且遞送到腎臟。

[1280] 到腎臟的遞送方法概括如下：

遞送方法	載體	靶RNA	疾病	模型	功能測定	作者
流體動力學/脂質	TransIT體內基因遞送系統，DOTAP	p85α	急性腎損傷	缺血再灌注	攝取，生物分佈	拉森（Larson）等人，手術（Surgery），（2007年8月），第142卷，第2期，第（262-269）頁
流體動力學/脂質	Lipofectamine 2000	Fas	急性腎損傷	缺血再灌注	血尿素氮、Fas免疫組織化學、細胞凋亡、組織評分	哈馬爾（Hamar）等人，美國國家科學院院刊，（2004年10月），第101卷，第41期，第（14883-14888）頁
流體動力學	n.a.	細胞凋亡級聯元件	急性腎損傷	缺血再灌注	n.a.	鄭（Zheng）等人，美國病理學雜誌，（2008年10月），第173卷，第4期，第（973-980）頁
流體動力學	n.a.	核因子κ-b（NFkB）	急性腎損傷	缺血再灌注	n.a.	馮等人，移植（Transplantation），（2009年5月），第87卷，第9

遞送方法	載體	靶RNA	疾病	模型	功能測定	作者
						期，第（1283-1289）頁
流體動力學/病毒	Lipofectamine 2000	拮抗凋亡轉錄因子（AATF）	急性腎損傷	缺血再灌注	細胞凋亡、氧化脅迫、半胱天冬酶啟動、膜脂過氧化作用	謝和郭（Guo），美國腎臟病學會雜誌（Am Soc Nephrol），（2006年12月），第17卷，第12期，第（3336–3346）頁
流體動力學	pBAsi mU6 Neo/ TransIT-EE流體動力學遞送系統	Gremlin	糖尿病腎病	鏈佐黴素誘導型糖尿病	蛋白尿、血清肌酐、腎小球和腎小管直徑、膠原IV型/BMP7表現	Q.·張等人，公共科學圖書館綜合，（2010年7月），第5卷，第7期，e11709，第（1-13）頁
病毒/脂質	pSUPER載體/脂質體	TGF-β II型受體	間質性腎纖維變性	單側輸尿管梗阻	α-SMA表現、膠原含量，	櫛引（Kushibikia）等人，控制釋放雜誌（J Controlled Release），（2005年7月），第105卷，第3期，第（318-331）頁
病毒	腺相關病毒-2	鹽皮質類固醇受體	高血壓引起的腎損傷	冷誘導型高血壓	血壓、血清白蛋白、血清尿素氮、血清肌酐、腎重、尿鈉	王（Wang）等人，基因治療，（2006年7月），第13卷，第14期，第

遞送方法	載體	靶RNA	疾病	模型	功能測定	作者
						（1097-1103）頁
流體動力學/病毒	pU6載體	螢光素酶	n.a.	n.a.	攝取	小林等人，藥理與實驗治療學雜（Journal of Pharmacology and Experimental Therapeutics），（2004年2月），第308期，第2卷，第（688-693）頁
脂質	脂蛋白、白蛋白	apoB1、apoM	n.a.	n.a.	攝取，與脂蛋白和白蛋白的結合親和力	沃爾夫魯姆（Wolfrum）等人，自然生物技術，（2007年9月），第25卷，第10期，第（1149-1157）頁
脂質	Lipofectamine2000	p53	急性腎損傷	缺血和順鉑誘導型急性損傷	組織學評分、細胞凋亡	莫里托裡斯等人，美國腎臟病學會雜誌，（2009年8月），第20卷，第8期，第（1754-1764）頁
脂質	DOTAP/DOPE、DOTAP/DOPE/DOPE-PEG2000	COX-2	乳腺腺癌	攜帶MDA-MB-231乳腺癌異種移植物的小鼠	細胞活力，攝取	米克哈洛瓦（Mikhaylova）等人，癌症基因治療（Cancer Gene Therapy），

遞送方法	載體	靶RNA	疾病	模型	功能測定	作者
						（2011年3月），第16卷，第3期，第（217-226）頁
脂質	膽固醇	12/15-脂加氧酶	糖尿病腎病	鏈佐黴素誘導型糖尿病	蛋白尿、尿肌酐、組織學、I型和IV型膠原、TGF-β、纖連蛋白、纖溶酶原啟動物抑制劑1	袁等人，美國腎臟生理學雜誌，（2008年6月），第295卷，第（F605-F617）頁
脂質	Lipofectamine 2000	線粒體膜44（TIM44）	糖尿病腎病	鏈佐黴素誘導型糖尿病	細胞增殖和細胞凋亡、組織學、ROS、Mn-SOD和麩胱甘肽過氧化物酶的線粒體輸入、細胞膜極化	Y.·張等人，美國腎臟病學會雜誌，（2006年4月），第17卷，第4期，第（1090-1101）頁
流體動力學/脂質	脂蛋白體	RLIP76	腎癌	攜帶Caki-2腎癌異種移植物的小鼠	攝取	辛格哈爾（Singhal）等人，癌症研究，（2009年5月），第69卷，第10期，第（4244-4251）頁
聚合物	PEG化PEI	螢光素酶pGL3	n.a.	n.a.	攝取、生物分佈、紅細胞聚集	馬利克（Malek）等人，毒物學與應用藥理學，（2009年4月），第236

遞送方法	載體	靶RNA	疾病	模型	功能測定	作者
						卷，第1期，第（97-108）頁
聚合物	PEG化聚-L-賴胺酸	MAPK1	狼瘡性腎小球腎炎	腎小球腎炎	蛋白尿、腎小球硬化症、TGF-β、纖連蛋白、纖溶酶原啟動物抑制劑1	清水等人，美國腎臟病學會雜誌，（2010年4月），第21卷，第4期，第（622-633）頁
聚合物/奈米粒子	透明質酸/量子點/PEI	VEGF	腎癌/黑色素瘤	攜帶B16F1黑色素腫瘤的小鼠	生物分佈、細胞毒性、腫瘤體積、內吞作用	蔣等人，分子藥劑學（Molecular Pharmaceutics），（2009年5月-6月），第6卷，第3期，第（727-737）頁
聚合物/奈米粒子	PEG化聚已酸內酯奈米纖維	GAPDH	n.a.	n.a.	細胞活力，攝取	曹（Cao）等人，控制釋放雜誌，（2010年6月），第144卷，第2期，第（203-212）頁
適配體	Spiegelmer mNOX-E36	CC趨化因子配位基2	腎小球硬化	單側腎切除後小鼠	尿白蛋白、尿肌酐、組織病理學、腎小球濾過率、巨噬細胞計數、血清Ccl2、Mac- 2+、Ki-67+	尼尼楚克（Ninichuk）等人，美國病理學雜誌，（2008年3月），第172卷，第3期，第（628-637）頁

遞送方法	載體	靶RNA	疾病	模型	功能測定	作者
適配體	Aptamer NOX-F37	血管加壓素（AVP）	充血性心力衰竭	n.a.	與D-AVP的結合親和力、AVP傳訊的抑制、尿滲透壓以及鈉濃度	普爾施克（Purschke）等人，美國國家科學院院刊，（2006年3月），第103卷，第13期，第（5173-5178）頁

靶向肝臟或肝臟細胞

[1281] 提供了靶向的肝臟細胞。此細胞可以是在體外或在體內。肝細胞係較佳的。CRISPR蛋白質諸如在此的Cpf1的遞送可以是經由病毒載體，尤其是AAV（並且具體地是AAV2/6）載體。該等載體可以是藉由靜脈內注射來給予的。

[1282] 肝臟的較佳的靶標（無論是在體外或在體內）係白蛋白基因。這係所謂的"安全港"，因為白蛋白以極高水平表現並且因此耐受在成功基因編輯之後的白蛋白產生的一些減少。它也是較佳的，因為由白蛋白啟動子/增強子看出的高水平表現允許實現有用水平的校正或轉基因產生（由插入的供體模板產生），即使僅一小部分肝細胞被編輯。

[1283] 韋克斯勒（Wechsler）等人證實白蛋白的內含子1（報導在美國血液學會第57屆年會和博覽會（57th Annual Meeting and Exposition of the American Society of Hematology）—摘要在https://ash.confex.com/ash/2015/webprogram/Paper86495.html處可線上獲得並且在2015年12月6日給出）係適合的靶位點。他們的研

究使用Zn指切割DNA的此靶位點，並且可以生成適合的指導序列以指導CRISPR蛋白在相同位點處的切割。

使用高度表現的基因（具有高活性增強子/啟動子的基因）諸如白蛋白內的靶標也可以允許使用無啟動子供體模板，如藉由韋克斯勒報導的，並且這在肝臟靶向之外也是廣泛適用的。高度表現基因的其他實例係已知的。

其他肝臟疾病

[1284] 在特定實施方式中，本發明的CRISPR蛋白用於治療肝臟病症，諸如甲狀腺素運載蛋白澱粉樣變性（ATTR）、α-1抗胰蛋白酶缺乏症以及其他基於肝臟的先天性代謝障礙。FAP係由編碼甲狀腺素運載蛋白（TTR）的基因中的突變引起的。雖然它係常染色體顯性疾病，但是並非所有的載體都發展該疾病。在TTR基因中存在已知與該疾病相關聯的超過100個突變。常見突變的實例包括V30M。藉由使用iRNA的研究已證明基於基因沈默的TTR治療原則（上田（Ueda）等人，2014，翻譯性神經變性（Transl Neurogener.）3:19）。威爾森氏病（WD）係由編碼肝細胞中專一可見的ATP7B的基因中的突變引起的。存在與WD相關聯的500個突變，其中它在特定地區諸如東亞的患病率增加。其他實例係A1ATD（由SERPINA1基因中的突變引起的常染色體隱性疾病）和PKU（由苯丙胺酸羥化酶（*PAH*）基因中的突變引起的常染色體隱性疾病）。

肝臟相關血液病症，尤其是血友病並且具體地是B型血友病

[1285] 在小鼠（體外和體內）和非人類靈長類動物（體內）

中已實現成功的肝細胞基因編輯，這顯示藉由肝細胞中的基因編輯/基因組工程化來治療血液病症係可行的。具體地說，在非人類靈長類動物中已顯示肝細胞的人類F9（hF9）基因的表現，這指示人類B型血友病的治療。

[1286] 韋克斯勒等人在美國血液學會第57屆年會和博覽會（摘要在2015年12月6日給出並且在https://ash.confex.com/ash/2015/webprogram/Paper86495.html處可線上獲得）報導了他們藉由體內基因編輯成功表現了來自非人類靈長類動物的肝細胞的F9（hF9）。這係使用 1) 靶向白蛋白座位的內含子1的兩種鋅指核酸酶（ZFN）以及 2) 人類F9供體模板來實現的。在靜脈內注射的單獨親肝性腺相關病毒血清型2/6（AAV2/6）上編碼ZFN和供體模板，從而使得hF9基因的校正拷貝靶向插入到一定比例的肝臟肝細胞的白蛋白座位中。

[1287] 白蛋白座位被選擇為"安全港"，因為此最豐富的血漿蛋白質的產生超過10 g/天，並且良好耐受那些水平的中等降低。藉由高活性白蛋白增強子/啟動子驅動，基因組編輯的肝細胞產生治療量的正常hFIX（hF9），而不是白蛋白。示出hF9轉基因在白蛋白座位處的靶向整合和此基因到白蛋白轉錄物中的剪接。

[1288] 小鼠研究：經由尾靜脈注射向C57BL/6小鼠給予媒介物（n=20）或編碼小鼠代理試劑的1.0×10^{13}個載體基因組（vg）/kg的AAV2/6載體（n=25）。在治療的小鼠中的血漿hFIX的ELISA分析顯示持續6個月研究持續時間的50-1053 ng/mL峰值水平。來自小鼠血漿的FIX活性水平證實與表現水平相稱的生物活性。

[1289] 非人類靈長類動物（NHP）研究：以$1.2x10^{13}$ vg/kg單次靜脈內共輸注編碼NHP靶向白蛋白特異性ZFN的AAV2/6載體和人類F9供體（n=5/組），導致在此大動物模型中>50 ng/mL（>正常水平的1%）。使用較高AAV2/6劑量（多至$1.5x10^{14}$ vg/kg）在幾隻動物中產生多至1000 ng/ml（或正常水平的20%）的血漿hFIX水平並且在單一動物中產生多至2000 ng/ml（或正常水平的50%）的血漿水平，持續研究的持續時間（3個月）。

[1290] 治療在小鼠和NHP中良好耐受，其中沒有顯著毒物學發現與治療劑量的兩種種類之一的AAV2/6 ZFN +供體治療相關。桑加莫公司（美國加利福尼亞州）已經適用於FDA許可並且已經得到批准，進行世界首次體內基因組編輯應用的人類臨床試驗。這在脂蛋白脂肪酶缺乏症的Glybera基因療法治療的EMEA批准之後進行。

[1291] 因此，在一些實施方式中，較佳的是使用任何或所有以下各項：

AAV（具體地是 AAV2/6）載體，較佳的是藉由靜脈內注射來給予；

作為用於基因編輯/插入轉基因/模板的靶標的白蛋白—尤其是在白蛋白的內含子 1 處；

人類 F9 供體模板；和/或

無啟動子供體模板。

B型血友病

[1292] 因此，在一些實施方式中，較佳的是本發明用於治療B型血友病。這樣較佳的是提供一種模板並且此模板係人類F9基因。將瞭解的是hF9模板包含wt版本或“校正”版本的hF9，以使得該治療係有效的。

[1293] 在一替代實施方式中，可以遞送F9的B型血友病版本，以便創建模型生物體、細胞或細胞系（例如鼠或非人類靈長類動物模型生物體、細胞或細胞系），該模型生物體、細胞或細胞系具有或攜帶B型血友病表型，即沒有產生wt F9的能力。

A型血友病

[1294] 在一些實施方式中，F9（因子IX）基因可以被以上所述F8（因子VIII）基因替換，從而使得能夠治療A型血友病（藉由提供校正的F8基因）並且/或者創建A型血友病模型生物體、細胞或細胞系（藉由提供未校正的A型血友病版本的F8基因）。

C型血友病

[1295] 在一些實施方式中，F9（因子IX）基因可以被以上所述F11（因子XI）基因替換，從而使得能夠治療C型血友病（藉由提供校正的F11基因）並且/或者創建C型血友病模型生物體、細胞或細胞系（藉由提供未校正的C型血友病版本的F11基因）。

治療上皮細胞和肺部疾病

[1296] 本發明還考慮將在此所述的CRISPR-Cas系統例如Cpf1效應蛋白系統遞送到一側或兩側肺部。

[1297] 儘管基於AAV-2的載體最初提出用於CFTR遞送到CF

氣道，但是其他血清型諸如AAV-1、AAV-5、AAV-6、以及AAV-9在各種各樣的肺上皮細胞模型中展現出提高的基因轉移效率（參見，例如，分子治療，第17卷，第12期，2067-2077，2009年12月）。證實在體外轉導的人類氣道上皮細胞中AAV-1比AAV-2和AAV-5更有效~100倍，5儘管體內AAV-1轉導的鼠氣管內氣道上皮細胞具有等於AAV-5的效率。其他研究已顯示，在針對體外人類氣道上皮（HAE）的基因遞送上，AAV-5比AAV-2更有效50倍，並且在體內小鼠肺氣道上皮中顯著更有效。還已顯示在體外人類氣道上皮細胞中和在體內鼠類氣道中，AAV-6比AAV-2更有效。8更為近期的分離物AAV-9顯示在體內鼠類鼻和肺泡上皮中展示了比AAV-5更大的基因轉移效率，其中持續超過9個月檢測出基因表現，這表明AAV可以使得能夠在體內進行長期基因表現，這對於CFTR基因遞送載體而言是一種理想特性。此外，證明了AAV-9可以被再次給予至鼠類的肺部，而不喪失CFTR表現並且具有最低限度的免疫結果。可以在CF和非CF HAE培養物的頂面上用100 μl 的AAV載體接種，持續數小時（參見，例如，李等人，分子治療，第17卷，第12期，2067-2077，2009年12月）。MOI可以從1×10^3到4×10^5個載體基因組/細胞而變化，這取決於病毒濃度和該等實驗的目的。以上引用的載體被考慮用於本發明的遞送和/或給藥。

[1298] 薩莫拉（Zamora）等人（美國呼吸道與危重護理學雜誌（Am J Respir Crit Care Med），第183卷，第531-538頁，2011）報導了針對人類感染性疾病治療的RNA干擾治療法的應用實例以及抗病毒藥物在呼吸道合胞病毒（RSV）感染的肺移植受體中的隨機試驗。薩莫拉等人進行

了一項在具有RSV呼吸道感染的LTX受體中的隨機化、雙盲、安慰劑對照的試驗。允許患者接受針對RSV的護理標準。每天給予霧化的ALN-RSV01（0.6 mg/kg）或安慰劑，持續3天。此研究表明可以向患有RSV感染的LTX受體安全地給予靶向RSV的RNAi治療劑。ALN-RSV01的三個每日劑量並不導致任何呼吸道症狀的加重或肺功能的損害，並且未展示任何出全身性致炎作用，諸如細胞介素或CRP的誘導。在吸入之後，藥代動力學僅顯示低的、短暫的全身性暴露，與臨床前動物資料一致，表明靜脈內或藉由吸入給予的ALN-RSV01藉由外切核酸酶介導的消化和腎臟排泄而從循環中快速清除。薩莫拉等人的方法可以適用於本發明的CRISPR Cas系統，並且對於本發明可以考慮霧化的CRISPR Cas，例如使用0.6 mg/kg的劑量。

[1299] 對肺病進行治療的受試者可以例如每側肺部接受藥物有效量的支氣管遞送的霧化的AAV載體系統，同時自發地呼吸。這樣，總的來說，對於AAV遞送，霧化的遞送係較佳的。腺病毒或AAV粒子可以用於遞送。可以將適合的基因構建體選殖到遞送載體中，該等基因構建體各自可操作地連接到一種或多種調節序列。在此情況下，提供以下構建體作為實例：對於Cas（Cpf1）的Cbh或EF1a啟動子、對於指導RNA的U6或H1啟動子。較佳的安排係使用CFTRδ508靶向指導、δF508突變的修復模板以及密碼子優化的Cpf1酶，其中具有視情況一種或多種核定位信號或序列（一個或多個NLS），例如兩個（2）NLS。設想沒有NLS的構建體。

治療肌肉系統疾病

[1300] 本發明還考慮將在此所述的CRISPR-Cas系統例如Cpf1效應蛋白系統遞送到一個或多個肌肉。

[1301] 博爾特蘭薩（Bortolanza）等人（分子治療，第19卷，第11期，2055-2064，2011年11月）證實，在FRG1小鼠中在面肩肱型肌營養不良（FSHD）發作之後，RNA干擾表現盒的全身性遞送導致劑量依賴性長期FRG1敲低，而沒有毒性跡象。博爾特蘭薩等人發現，單次靜脈內注射5×10^{12}vg的rAAV6-sh1FRG1挽救了FRG1小鼠的肌肉組織病理學和肌肉功能。詳細地說，使用25號泰爾茂（Terumo）注射器將200 μl含有2×10^{12}或5×10^{12} vg載體的生理溶液注射到尾靜脈中。博爾特蘭薩等人的方法可以適用於表現CRISPR Cas的AAV並且可以約2×10^{15}或2×10^{16} vg載體的劑量注射到人類中。

[1302] 杜蒙特科（Dumonceaux）等人（分子治療，第18卷，第5期，881-887，2010年5月）使用針對肌肉生長抑制素受體AcvRIIb mRNA（sh-AcvRIIb）的RNA干擾技術抑制肌肉生長抑制素途徑。由載體化U7外顯子跳躍技術（U7-DYS）介導准肌營養不良蛋白（quasi-dystrophin）的恢復。將攜帶單獨的sh-AcvrIIb構建體、單獨的U7-DYS構建體、或這兩種構建體的組合的腺相關載體注射到營養不良mdx小鼠的脛骨前肌（TA）肌肉中。以10^{11}個AAV病毒基因組進行注射。杜蒙特科等人的方法可以適用於表現CRISPR Cas的AAV並且以約10^{14}或10^{15} vg載體的劑量注射到人類中。

[1303] 木內（Kinouchi）等人（基因治療 (2008)15，1126-

1130）報導了藉由未經化學修飾的siRNA 與缺端膠原（ATCOL）形成奈米粒子來體內siRNA遞送到正常或患病小鼠骨骼肌的有效性。ATCOL介導的靶向肌肉生長抑制素（骨骼肌生長的負調節劑）的siRNA在小鼠骨骼肌中的局部應用或者靜脈內應用在應用之後幾週內引起肌肉質量的顯著增加。該等結果顯示siRNA的ATCOL介導的應用係用於包括肌肉萎縮在內的疾病的未來治療用途的強大工具。根據製造商的說明，將MstsiRNA（終濃度，10 mM）與ATCOL（對於局部給藥的終濃度，0.5%）（AteloGene，高研株式會社（Kohken），日本東京（Tokyo, Japan））混合。在藉由耐波他（Nembutal）（25 mg/kg，腹膜內注射）麻醉小鼠（20週大的雄性C57BL/6）之後，將Mst-siRNA/ATCOL複合物注射到咬肌和股二頭肌中。木內等人的方法可以適用於CRISPR Cas並且可以注射到人類中，例如以40 μM溶液的約500至1000 ml的劑量注射到肌肉中。哈格斯特龍（Hagstrom）等人（分子治療，第10卷，第2期，2004年8月）描述了使得能夠將核酸有效且可重複地遞送到遍及哺乳動物四肢肌肉的肌細胞（肌纖維）的血管內、非病毒方法。該程式涉及將裸質粒DNA或siRNA注射到暫時由止血帶或血壓袖帶分離的肢體的遠端靜脈中。藉由以足夠的體積將其迅速注射來促進向肌纖維的核酸遞送，使得該核酸溶液能夠溢出到肌肉組織中。在小動物和大動物中都以最低毒性實現在骨骼肌中的高水平轉基因表現。還獲得了向四肢肌肉遞送siRNA的證據。為了將質粒DNA靜脈內注射到恒河猴中，將三通旋塞連接到各自載入有單個注射器的兩個注射器泵（型號PHD 2000；哈佛儀器公司（Harvard Instruments））上。在罌粟鹼注射五分鐘之後，以1.7或2.0 ml/s的速

率注射pDNA（15.5到25.7 mg，在40-100 ml鹽水中）。對於表現本發明的CRISPR Cas的質粒DNA，這可以按比例增加，其中對於人類注射在800到2000 ml鹽水中的約300到500 mg。對於將腺病毒載體注射到大鼠中，注射在3 ml正常生理鹽水溶液（NSS）中的2×10^9個感染粒子。對於表現本發明的CRISPR Cas的質粒DNA，這可以按比例增加，其中對於人類注射在10升NSS中的約1×10^{13}個感染粒子。對於siRNA，以12.5 μg的siRNA注射到大鼠的大隱靜脈中，並且以750 μg的siRNA注射到靈長類動物的大隱靜脈中。對於本發明的CRISPR Cas，這可以按比例增加，例如，其中將約15至約50 mg注射到大隱靜脈中。

[1304] 例如，還參見WO2013163628 A2，突變基因的遺傳校正（Genetic Correction of Mutated Genes），杜克大學（Duke University）的公開申請，描述了例如校正框移突變的努力，該框移突變引起提前終止密碼子和可以經由核酸酶介導的非同源末端接合進行校正的截短基因產物，該基因產物諸如引起杜氏肌營養不良（“DMD”）的那些基因產物，該杜氏肌營養不良係一種隱性遺傳的、致命的、X連鎖疾病，其導致由肌營養不良蛋白基因突變所致的肌肉變性。引起DMD的大多數肌營養不良蛋白突變係破壞閱讀框並且引起肌營養不良蛋白基因的提前翻譯終止的外顯子缺失。肌營養不良蛋白係細胞質蛋白，它提供負責調節肌細胞完整性和功能的細胞膜肌營養不良蛋白聚糖複合物的結構穩定性。如在此可互換地使用的肌營養不良蛋白基因或“DMD基因”係在座位Xp21處的2.2兆鹼基。初級轉錄測量了約2,400 kb，其中成熟mRNA係約14 kb。79個外顯子編碼超過3500個胺基酸的蛋白質。在DMD

患者中，外顯子51常常接近破壞框的缺失並且已在臨床試驗中被靶向用於基於寡核苷酸的外顯子跳躍。對於外顯子51跳躍化合物依替利森（eteplirsen）的臨床試驗，最近報導了跨48週的顯著功能益處，與基線相比具平均47%的肌營養不良蛋白陽性纖維。外顯子51中的突變理想地適合於藉由基於NHEJ的基因組編輯進行永久性校正。

[1305] 轉讓給策勒克提斯公司的美國專利公開案號20130145487的方法涉及切割來自人類肌營養不良蛋白基因（DND）的靶序列的大範圍核酸酶變體，該方法也可以被修改成本發明的核酸靶向系統。

治療皮膚疾病

[1306] 本發明還考慮將在此所述的CRISPR-Cas系統例如Cpf1效應蛋白系統遞送到皮膚。

[1307] 希克森（Hickerson）等人（分子治療-核酸 (2013)2，e129）涉及用於向人類和鼠類皮膚遞送自我遞送（sd）-siRNA的機動化的微針陣列皮膚遞送裝置。將基於siRNA的皮膚治療劑轉化到臨床的主要挑戰係有效遞送系統的開發。在多種皮膚遞送技術中已經投入了大量的努力，但是成功有限。在其中用siRNA治療皮膚的臨床研究中，與皮下針注射相關聯的劇烈疼痛排除了試驗中額外患者的納入，這凸顯了對於改進的、更為“患者友好的”（ 即，很少或沒有疼痛） 遞送方法的需要。微針代表將包括siRNA在內的大帶電貨物遞送穿過一級障壁角質層的有效途徑，並且通常被認為比常規皮下針疼痛更少。機動化的“衝壓型”微針裝置，包括由

希克森等人使用的機動化微針陣列（MMNA）裝置，已經顯示在無毛小鼠研究中是安全的並且引起很少的疼痛或不引起疼痛，其證據為：(i) 在美容業中廣泛使用以及 (ii) 其中幾乎所有志願者都發現使用該裝置比流感疫苗針劑（flushot）疼痛少得多的有限測試，這表明使用此裝置的siRNA遞送將產生比使用皮下針注射的先前臨床試驗中所體驗的少得多的疼痛。該MMNA裝置（作為Triple-M或Tri-M由韓國首爾（Seoul, South Korea）的Bomtech電子有限公司銷售）適於將siRNA遞送到小鼠和人類皮膚。將sd-siRNA溶液（高達300 μl的0.1 mg/mlRNA）引入到設定為0.1 mm深度的一次性Tri-M針盒（Bomtech公司）的腔室中。為了處理人類皮膚，在處理之前將未鑒定的皮膚（在外科手術之後立即獲得）手動拉伸並且釘在軟木平臺上。使用具有28號0.5英吋針頭的胰島素注射器進行所有皮內注射。該MMNA裝置和希克森等人的方法可以用於並且/或者適於例如以高達300 μl的0.1 mg/mlCRISPR Cas的劑量將本發明的CRISPR Cas遞送到皮膚。

[1308] 裡奇曼（Leachman）等人（分子治療，第18卷，第2期，442-446，2010年2月）涉及利用基於第一短干擾RNA（siRNA）的皮膚治療劑用於治療罕見皮膚病症先天性厚甲（PC）的Ib期臨床試驗，先天性厚甲係常染色體顯性綜合症，包括致殘性的掌蹠角化病。此siRNA（稱為TD101）特異性地並強有力地靶向角蛋白6a（K6a）N171K突變體mRNA，而不影響野生型K6a mRNA。

[1309] 鄭等人（美國國家科學院院刊，2012年7月24日，第109卷，第30期，11975–11980）證實，球形核酸奈米粒子軛合物

（SNA-NC）係由高度定向的、共價固定的siRNA的緻密殼圍繞的金核，它在應用之後數小時內自由地穿透幾乎100%體外角化細胞、小鼠皮膚、以及人類表皮。鄭等人證明，在人類皮膚中單次應用25 nM的表皮生長因子受體（EGFR）SNA-NC持續60小時顯示出有效的基因敲低。對於向皮膚給予的在SNA-NC中固定的CRISPR Cas，可以考慮類似劑量。

癌症

[1310] 在一些實施方式中，提供了癌症的治療、預防或診斷。靶標較佳的是以下各項中的一種或多種：FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、TRAC或TRBC基因。該癌症可以是以下各種中的一種或多種：淋巴瘤、慢性淋巴細胞白血病（CLL）、B細胞急性淋巴細胞白血病（B-ALL）、急性成淋巴細胞性白血病、急性髓性白血病、非霍奇金淋巴瘤（NHL）、彌漫性大細胞淋巴瘤（DLCL）、多發性骨髓瘤、腎細胞癌（RCC）、成神經細胞瘤、結腸直腸癌、乳腺癌、卵巢癌、黑色素瘤、肉瘤、前列腺癌、肺癌、食管癌、肝細胞癌、胰腺癌、星形細胞瘤、間皮瘤、頭頸癌、以及成神經管細胞瘤。這可以用工程化嵌合抗原受體（CAR）T細胞執行。這描述於WO2015161276中，該專利的揭露內容藉由引用結合在此並且如下文所述。

[1311] 在一些實施方式中，適用於治療或預防癌症的靶基因可以包括WO2015048577中所述的那些靶基因，該專利的揭露內容藉由引用結合在此。

Usher症候群或色素性視網膜炎-39

[1312] 在一些實施方式中，提供了Usher症候群或色素性視網膜炎-39的治療、預防或診斷。該靶標較佳的是USH2A基因。在一些實施方式中，提供了在位置2299處的G缺失（2299delG）的校正。這描述於WO2015134812A1中，該專利的揭露內容藉由引用結合在此。

囊性纖維化（CF）

[1313] 在一些實施方式中，提供了囊性纖維化的治療、預防或診斷。該靶標較佳的是SCNN1A或CFTR基因。這描述於WO2015157070中，該專利的揭露內容藉由引用結合在此。

[1314] 施萬克(Schwank)等人(細胞幹細胞，13:653–58, 2013)使用CRISPR-Cas9校正與人類幹細胞的囊性纖維化相關聯的缺陷。該組的靶標係離子通道囊性纖維化跨膜導體受體(CFTR)的基因。CFTR中的缺失引起囊性纖維化患者中的蛋白質錯誤折疊。使用由來自患有囊性纖維化的兩位兒童的細胞樣品開發的培養的腸道幹細胞，施萬克等人能夠使用CRISPR連同含有有待插入的修復性序列的供體質粒校正該缺陷。然後研究者將該等細胞生長成腸道“細胞器”或微型腸，並且證實它們能夠正常起作用。在此情況下，約一半選殖的細胞器經受適當的遺傳校正。

HIV和AIDS

[1315] 在一些實施方式中，提供了HIV和AIDS的治療、預防或診斷。該靶標較佳的是HIV中的CCR5基因。這描述於WO2015148670A1中，該專利的揭露內容藉由引用結合在此。

β地中海貧血

[1316] 在一些實施方式中，提供了β地中海貧血的治療、預防或診斷。該靶標較佳的是BCL11A基因。這描述於WO2015148860中，該專利的揭露內容藉由引用結合在此。

鐮狀細胞疾病（SCD）

[1317] 在一些實施方式中，提供了鐮狀細胞疾病（SCD）的治療、預防或診斷。該靶標較佳的是HBB或BCL11A基因。這描述於WO2015148863中，該專利的揭露內容藉由引用結合在此。

單純性皰疹病毒1和2

[1318] 在一些實施方式中，提供了HSV-1（單純性皰疹病毒1）的治療、預防或診斷。該靶標較佳的是在HSV-1中的UL19、UL30、UL48或UL50基因。這描述於WO2015153789中，該專利的揭露內容藉由引用結合在此。

[1319] 在其他實施方式中，提供了HSV-2（單純性皰疹病毒2）的治療、預防或診斷。該靶標較佳的是在HSV-2中的UL19、UL30、UL48或UL50基因。這描述於WO2015153791中，該專利的揭露內容藉由引用結合在此。

[1320] 在一些實施方式中，提供了原發性開角型青光眼（POAG）的治療、預防或診斷。該靶標較佳的是MYOC基因。這描述於WO2015153780中，該專利的揭露內容藉由引用結合在此。

過繼細胞治療

[1321] 本發明還考慮使用在此所述的CRISPR-Cas系統例如Cpf1效應蛋白系統修飾用於過繼治療的細胞。本發明的多個方面因此涉及過繼性轉移對於選定抗原諸如腫瘤相關抗原特異的免疫系統細胞諸如T細胞（參見，毛斯（Maus）等人，2014，用於癌症或病毒的過繼性免疫治療（Adoptive Immunotherapy for Cancer or Viruses），免疫學年度綜述（Annual Review of Immunology），第32卷：189-225；羅森伯格（Rosenberg）和羅斯替弗（Restifo），2015，作為用於人類癌症的個體化免疫治療的過繼性細胞轉移（Adoptive cell transfer as personalized immunotherapy for human cancer），科學，第348卷，第6230期，第62-68頁；以及羅斯替弗等人，2015，用於癌症的過繼性免疫治療：利用T細胞反應（Adoptive immunotherapy for cancer: harnessing the T cell response.）自然綜述免疫學12(4): 269-281；以及詹森（Jenson）和裡德爾（Riddell），2014，使用嵌合抗原受體修飾性T細胞設計並實施過繼治療（Design and implementation of adoptive therapy with chimeric antigen receptor-modified T cells.）免疫學綜述 257(1): 127–144）。不同策略可以例如用於藉由改變T細胞受體（TCR）的特異性，例如藉由引起具有選定的肽特異性的新TCR α和β鏈來遺傳性修飾T細胞（參見，美國專利案號8,697,854；PCT專利公開：WO2003020763、WO2004033685、WO2004044004、WO2005114215、WO2006000830、WO2008038002、WO2008039818、WO2004074322、WO2005113595、WO2006125962、WO2013166321、WO2013039889、WO2014018863、WO2014083173；美國專利案號8,088,379）。

[1322] 作為TCR修飾的替代方案或者除TCR修飾之外，可以

使用嵌合抗原受體（CAR），以便生成對於選定靶標諸如惡性腫瘤細胞特異的免疫反應細胞諸如T細胞，其中已經描述了各種各樣的受體嵌合構建體（參見美國專利案號5,843,728；5,851,828；5,912,170；6,004,811；6,284,240；6,392,013；6,410,014；6,753,162；8,211,422；以及PCT公開WO9215322）。替代性CAR構建體可以被表徵為屬於連續世代。第一代CAR典型地由對於抗原特異的抗體單股可變片段組成，例如包括連接至特異性抗體的VH的VL，藉由柔性接頭連接，例如藉由CD8α鉸鏈區和CD8α跨膜結構域，連接至CD3ζ或FcRγ的跨膜和細胞內傳訊結構域（scFv-CD3ζ或scFv-FcRγ；參見美國專利案號7,741,465；美國專利案號5,912,172；美國專利案號5,906,936）。第二代CAR結合一種或多種共刺激分子的細胞內結構域，諸如內結構域內的CD28、OX40（CD134）或4-1BB（CD137）（例如scFv-CD28/OX40/4-1BB-CD3ζ；參見美國專利案號8,911,993；8,916,381；8,975,071；9,101,584；9,102,760；9,102,761）。第三代CAR包括共刺激內結構域的組合，諸如CD3ζ-鏈、CD97、GDI la-CD18、CD2、ICOS、CD27、CD154、CDS、OX40、4-1BB、或CD28傳訊結構域（例如scFv-CD28-4-1BB-CD3ζ或scFv-CD28-OX40-CD3ζ；參見美國專利案號8,906,682；美國專利案號8,399,645；美國專利案號5,686,281；PCT公開號WO2014134165；PCT公開號WO2012079000）。可替代地，共刺激可以藉由在抗原特異性T細胞中表現CAR來調控，該T細胞被選擇為在其天然αβTCR接合之後啟動並擴增，例如在伴隨共刺激的情況下藉由專職抗原呈遞細胞上的抗原。此外，另外的工程化受體可以被提供在免疫反應細胞中，例如以提高T細胞攻擊的靶向並且/或者最小化副作用。

[1323] 替代技術可以用於轉化免疫反應靶細胞，諸如原生質體融合、脂轉染、轉染或電穿孔。可以使用各種各樣的載體，諸如逆轉錄病毒載體、慢病毒載體、腺病毒載體、腺相關病毒載體、質粒或轉位子，諸如睡美人轉位子（參見美國專利案號6,489,458；7,148,203；7,160,682；7,985,739；8,227,432），該等載體可以用於例如使用藉由CD3ζ和CD28或CD137傳導信號的第2代抗原特異性CAR來引入CAR。病毒載體可以例如包括基於HIV、SV40、EBV、HSV或BPV的載體。

[1324] 靶向轉化的細胞可以例如包括T細胞、自然殺傷細胞（NK）、細胞毒性T淋巴細胞（CTL）、調節T細胞、人類胚胎幹細胞、腫瘤浸潤淋巴細胞（TIL）或淋巴樣細胞可以由其分化的多能幹細胞。表現希望的CAR的T細胞可以例如藉由與γ-輻射啟動和增殖細胞（AaPC）共培養來選擇，該等啟動和增殖細胞共表現癌症抗原和共刺激分子。過程化CAR T細胞可以例如藉由在可溶性因子諸如IL-2和IL-21存在下在AaPC上共培養來擴增。此擴增可以例如被執行來提供記憶CAR+ T細胞（該等細胞可以例如藉由非酶數位陣列和/或多板型（multi-panel）流式細胞術來測定）。以這種方式，可以提供具有針對攜帶抗原的腫瘤的特異性細胞毒性活性的CAR T細胞（視情況與希望的趨化因子諸如干擾素-γ的產生相結合）。這種類型的CAR T細胞例如可以用於動物模型中，例如以威懾腫瘤異種移植物。

[1325] 方法諸如上述方法可以適於提供例如藉由給予有效量的免疫反應細胞來治療患有諸如瘤形成的疾病的患者並且/或

者增加該患者的存活的方法，該免疫反應細胞包含結合選定抗原的抗原識別受體，其中該結合啟動免疫反應細胞，從而治療或預防該疾病（諸如瘤形成、病原體感染、自身免疫病症或同種異體移植反應）。在具有或不具有淋巴細胞耗竭過程的情況下，例如在使用環磷醯胺的情況下，CAR T細胞治療的給藥可以例如涉及給予10^6至10^9個細胞/kg。

[1326] 在一個實施方式中，可以向經受免疫抑制治療的患者給予該治療。該等細胞或細胞群體可以被製成由於編碼至少一種免疫抑制劑的受體的基因滅活而抵抗此免疫抑制劑。在不受理論約束的情況下，在患者內免疫抑制治療應說明選擇並擴增根據本發明的免疫反應細胞或T細胞。

[1327] 可以任何常規方式執行根據本發明的細胞或細胞群體的給藥，該等方式包括藉由霧化吸入、注射、攝取、輸血、植入或移植。可以向患者皮下、真皮內、瘤內、節點內、髓內、肌內、藉由靜脈內注射或淋巴管內注射、或者腹膜內給予該等細胞或細胞群體。在一個實施方式中，本發明的細胞組成物較佳的是藉由靜脈內注射來給予。

[1328] 該等細胞或細胞群體的給予可以由10^4-10^9個細胞/kg體重、較佳的是10^5至10^6個細胞/kg體重（包括該等範圍內細胞數目的所有整數值）的給予組成。在具有或不具有淋巴細胞耗竭過程的情況下，例如在使用環磷醯胺的情況下，CAR T細胞治療的給藥可以例如涉及給予10^6至10^9個細胞/kg。該等細胞或細胞群體可以一個或多個劑量給予。在另一個實施方式中，細胞的有效劑

量係作為單一劑量給予。在另一個實施方式中，細胞的有效劑量係超過一個劑量持續一段時間來給予。給藥的時間係在管理醫生的判斷內並且取決於患者的臨床病狀。該等細胞或細胞群體可以是從任何來源諸如血庫或供體獲得。雖然個體需要不同，但是特定疾病或病狀的給定細胞類型的有效量的最佳範圍的確定係在熟習該項技術者知識內的。有效量意指提供治療或預防益處的量。給予的劑量將取決於接受者的年齡、健康和體重、同期治療（如果有的話）的類型、治療的頻率以及希望的作用性質。

[1329] 在另一個實施方式中，胃腸外給予有效量的細胞或包含那些細胞的組成物。給藥可以是靜脈內給藥。給藥可以是藉由在腫瘤內注射來直接進行的。

[1330] 為了防止可能的不良反應，可以用轉基因安全開關裝備工程化免疫反應細胞，該轉基因安全開關係致使該等細胞易於暴露於特定信號的轉基因形式。例如，單純皰疹病毒胸苷激酶（TK）可以這種方式使用，例如藉由在幹細胞移植之後引入到用作供體淋巴細胞輸注的同種異體T淋巴細胞中（格雷科（Greco）等人，使用TK自殺基因提高細胞治療的安全性（Improving the safety of cell therapy with the TK-suicide gene.）藥理學前沿（Front.Pharmacol.） 2015；6:95）。在此類細胞中，核苷前藥諸如更昔洛韋或阿昔洛韋的給予引起細胞死亡。替代性安全開關構建體包括誘導型半胱天冬酶9，例如藉由給予使兩個非功能icasp9分子連接在一起形成活性酶的小分子二聚體來觸發。已描述了實施細胞增殖控制的各種各樣的替代方法（參見美國專利公開案號

20130071414；PCT專利公開WO2011146862；PCT專利公開WO2014011987;PCT專利公開WO2013040371;週等人,血液 2014, 123/25:3895-3905；迪·史塔西（Di Stasi）等人，新英格蘭醫學雜誌 2011；365:1673-1683;薩德薩德·M（Sadelain M）等人，新英格蘭醫學雜誌 2011；365:1735-173;拉莫斯（Ramos）等人, 幹細胞 28(6):1107-15 (2010))。

[1331] 在過繼治療的另一個改進方案中，如在此所述的CRISPR-Cas系統進行的基因組編輯可以用於使免疫反應細胞適於替代性實施方案，例如提供編輯的CAR T細胞（參見，白羅（Poirot）等人，2015，用於"現成的"過繼性T細胞免疫反應的多重基因組編輯T細胞製造平臺（Multiplex genome edited T-cell manufacturing platform for "off-the-shelf" adoptive T-cell immunotherapies)，癌症研究 75 (18): 3853)。例如，免疫反應細胞可以被編輯為缺失一些或所有類別的HLA II型和/或I型分子的表現，或者敲除可以抑制所希望的免疫反應的選定基因諸如PD1基因。

[1332] 細胞可以使用如在此所述的任何CRISPR系統及其使用方法來編輯。CRISPR系統可以藉由在此所述的任何方法來遞送到免疫細胞。在較佳的實施方式中，細胞被離體編輯並且轉移到有需要的受試者中。可以編輯免疫反應細胞、CAR T細胞或用於過繼性細胞轉移的任何細胞。編輯可以被進行來消除潛在的同種異體反應性T-細胞受體（TCR）、破壞化學治療劑的靶標、阻斷免疫校驗點、啟動T細胞並且/或者增加功能耗盡或功能障礙的CD8+

T-細胞的分化和/或增殖（參見PCT專利公開：WO2013176915、WO2014059173、WO2014172606、WO2014184744、以及WO2014191128）。編輯可以導致基因滅活。

[1333] 藉由滅活基因，期望感興趣的基因不以功能蛋白形式表現。在一個特定實施方式中，CRISPR系統特異性催化在一個靶向基因中的切割，從而滅活所述靶基因。引起的核酸股斷裂通常藉由同源重組或非同源末端接合（NHEJ）的不同機制來修復。然而，NHEJ係不完美的修復過程，它常常導致切割位點處的DNA序列改變。藉由非同源末端接合（NHEJ）的修復常常形成小插入或缺失（Indel）並且可以用於創建特定基因敲除。其中發生切割誘導型誘變事件的細胞可以藉由本領域已熟知的方法來鑒定和/或選擇。

[1334] T細胞受體（TCR）係響應於抗原呈遞而參與啟動T細胞的細胞表面受體。TCR通常是由組裝形成異源二聚體的兩條股α和β形成，並且與CD3-轉導亞基締合形成存在於細胞表面的T細胞受體複合物。TCR的每條α和β鏈由免疫球蛋白樣N-末端可變區（V）和恒定（C）區、疏水性跨膜結構域、以及短胞質區。如對於免疫球蛋白分子，α和β鏈的可變區係藉由V(D)J重組從而在T細胞群體內形成多種抗原特異性來生成。然而，與識別完整抗原的免疫球蛋白相比，T細胞藉由與MHC分子締合的加工肽片段來啟動，從而將額外維度引入到由T細胞進行的抗原識別中，這被稱為MHC限制。藉由T細胞受體識別供體與受體之間的MHC差異導致T細胞增殖和移植物抗宿主疾病（GVHD）的潛在發展。TCRα或TCRβ的滅

活可以導致TCR從T細胞表面消除，從而阻止了同種抗原的識別並因此產生GVHD。然而，TCR破壞通常導致CD3傳訊組分的消除並且改變其他T細胞擴增的方式。

[1335] 同種異體細胞被宿主免疫細胞快速排斥。已證明存在於非輻射血液產品中的同種異體淋巴細胞將持續不超過5至6天（博尼，穆拉斯基（Boni, Muranski）等人，2008 血液1;112(12):4746-54）。因此，為防止同種異體細胞的排斥，宿主免疫系統通常不得不受到一定程度的抑制。然而，在過繼性細胞轉移的情況下，使用免疫抑制藥物也對引入的治療性T細胞具有有害作用。因此，為了在該等病狀中有效使用過繼性免疫治療方法，引入的細胞將需要抵抗免疫抑制治療。因此，在一個特定實施方式中，本發明進一步包括修飾T細胞較佳的是藉由滅活編碼免疫抑制劑的靶標的至少一種基因以使其抵抗免疫抑制劑的步驟。免疫抑制劑係藉由幾種作用機制之一來抑制免疫功能的藥劑。免疫抑制劑可以是但不限於，鈣調磷酸酶抑制劑、雷帕黴素的靶標、介白素-2受體α-鏈阻斷劑、肌苷單磷酸脫氫酶的抑制劑、二氫葉酸還原酶的抑制劑、皮質類固醇或免疫抑制抗代謝物。本發明允許藉由滅活免疫抑制劑在T細胞中的靶標來對用於免疫治療的T細胞賦予免疫抑制抗性。作為非限制性實例，免疫抑制劑的靶標可以是免疫抑制劑的受體，諸如：CD52、糖皮質激素受體（GR）、FKBP家族基因成員以及親環蛋白家族基因成員。

[1336] 免疫檢驗點係減慢或停止免疫反應並且防止受到免疫細胞的未控制活性的過度組織損害的抑制途徑。在某些實施方

式中，靶向的免疫檢驗點係程式性死亡-1（PD-1或CD279）基因（*PDCD1*）。在其他實施方式中，靶向的免疫檢驗點係細胞毒性T-淋巴細胞相關抗原（CTLA-4）。在另外的實施方式中，靶向的免疫檢驗點係CD28和CTLA4 Ig超家族的另一個成員，諸如BTLA、LAG3、ICOS、PDL1或KIR。在另外的實施方式中，靶向的免疫檢驗點係TNFR超家族的成員，諸如CD40、OX40、CD137、GITR、CD27或TIM-3。

[1337] 另外的免疫檢驗點包括含有Src同源2結構域的蛋白酪胺酸磷酸酶1（SHP-1）（沃森·HA等人，SHP-1：用於癌症免疫治療的下一個檢驗點靶標（SHP-1: the next checkpoint target for cancer immunotherapy）？生物化學學會彙報（Biochem Soc Trans.），2016年4月15日；44(2):356-62）。SHP-1係一種廣泛表現的抑制性蛋白酪胺酸磷酸酶（PTP）。在T細胞中，它係抗原依賴性啟動和增殖的負調節劑。它係一種細胞溶質蛋白，並且因此不適於抗體介導的治療，但是它在啟動和增殖中的作用使得它成為過繼轉移策略中用於遺傳操縱的有吸引力靶標，諸如嵌合抗原受體（CAR）T細胞。免疫檢驗點還可以包括具有Ig和ITIM結構域的T細胞免疫受體（TIGIT/Vstm3/WUCAM/VSIG9）和VISTA（勒·梅西埃·I（Le Mercier I）等人，(2015) 除CTLA-4和PD-1之外的第Z代負檢驗點調節劑（Beyond CTLA-4 and PD-1, the generation Z of negative checkpoint regulators）。免疫學前沿6:418）。

[1338] WO2014172606涉及使用MT1和/或MT1抑制劑增加耗盡的CD8+ T-細胞的增殖和/或活性並且減少CD8+ T-細胞耗盡

（例如，減少功能耗盡或不反應的CD8+免疫細胞）。在某些實施方式中，在過繼轉移的T細胞中藉由基因編輯靶向金屬硫蛋白。

[1339] 在某些實施方式中，基因編輯的靶標可以是涉及免疫檢驗點蛋白的表現的至少一個靶向座位。此類靶標可以包括但不限於，CTLA4、PPP2CA、PPP2CB、PTPN6、PTPN22、PDCD1、ICOS（CD278）、PDL1、KIR、LAG3、HAVCR2、BTLA、CD160、TIGIT、CD96、CRTAM、LAIR1、SIGLEC7、SIGLEC9、CD244（2B4）、TNFRSF10B、TNFRSF10A、CASP8、CASP10、CASP3、CASP6、CASP7、FADD、FAS、TGFBRII、TGFRBRI、SMAD2、SMAD3、SMAD4、SMAD10、SKI、SKIL、TGIF1、IL10RA、IL10RB、HMOX2、IL6R、IL6ST、EIF2AK4、CSK、PAG1、SIT1、FOXP3、PRDM1、BATF、VISTA、GUCY1A2、GUCY1A3、GUCY1B2、GUCY1B3、MT1、MT2、CD40、OX40、CD137、GITR、CD27、SHP-1或TIM-3。在較佳的實施方式中，靶向涉及PD-1或CTLA-4基因的表現的基因座位。在其他較佳的實施方式中，靶向基因的組合，諸如但不限於PD-1和TIGIT。

[1340] 在其他實施方式中，編輯至少兩種基因。基因對可以包括但不限於，PD1和TCRα、PD1和TCRβ、CTLA-4和TCRα、CTLA-4和TCRβ、LAG3和TCRα、LAG3和TCRβ、Tim3和TCRα、Tim3和TCRβ、BTLA和TCRα、BTLA和TCRβ、BY55和TCRα、BY55和TCRβ、TIGIT和TCRα、TIGIT和TCRβ、B7H5和TCRα、B7H5和TCRβ、LAIR1和TCRα、LAIR1和TCRβ、SIGLEC10和TCRα、SIGLEC10和TCRβ、2B4和TCRα、2B4和TCRβ。

[1341] 無論是在T細胞的遺傳修飾之前還是之後，T細胞都可以通常使用例如以下各項所述的方法來啟動並擴增：美國專利6,352,694；6,534,055；6,905,680；5,858,358；6,887,466；6,905,681；7,144,575；7,232,566；7,175,843；5,883,223；6,905,874；6,797,514；6,867,041；以及7,572,631。該細胞可以在體外或在體內擴增。

[1342] 除非另外指明，本發明的實施採用處於本領域技能範圍內的免疫學、生物化學、化學、分子生物學、微生物學、細胞生物學、基因組學以及重組DNA的常規技術。參見，分子選殖：實驗手冊，第2版 (1989)（薩姆布魯克、弗裡奇和馬尼亞蒂斯）；分子選殖：實驗手冊，第4版 (2012)（格林和薩姆布魯克）；分子生物學通用方法 (1987)（F. M.奧蘇貝爾等人編著）；系列叢書 酶學方法（學術出版社公司）；PCR 2：實踐方法 (1995)（M.J.·麥克弗森、B.D.·黑姆斯以及G.R.·泰勒編著）；抗體、實驗室手冊 (1988)（哈洛和拉內編著）；抗體、實驗室手冊，第2版 (2013)（E.A.·格林菲爾德（E.A. Greenfield）編著）；以及動物細胞培養 (1987)（R.I.費施奈伊編著）。

[1343] 除非另外指示，否則本發明的實踐採用用於生成遺傳修飾小鼠的常見技術。參見瑪律滕·H.·霍夫卡爾（Marten H. Hofker）和讓·凡·德瑞森（Jan van Deursen），轉基因小鼠方法和方案（TRANSGENIC MOUSE METHODS AND PROTOCOLS），第2版 (2011)。

基因驅動

[1344] 本發明還考慮使用在此所述的CRISPR-Cas系統例如

Cpf1效應蛋白系統來提供RNA指導的基因驅動，例如在與PCT專利公開WO 2015/105928中所述的基因驅動類似的系統中。這種類型的系統可以例如提供用於藉由將編碼RNA指導的DNA核酸酶和一種或多種指導RNA的核酸序列引入到生殖細胞中來改變真核生殖細胞的方法。指導RNA可以被設計為與生殖細胞的基因組DNA上的一個或多個靶位置互補。編碼RNA指導的DNA核酸酶的核酸序列和編碼指導RNA的核酸序列可以被提供在構建體上的側翼序列之間，其中啟動子被安排為使得生殖細胞可以表現RNA指導的DNA核酸酶和指導RNA，連同也位於側翼序列之間的任何希望的貨物編碼序列。側翼序列將典型地包括與選定的靶染色體上的相應序列相同的序列，以使得側翼序列與由該構建體編碼的元件一起作用，以促進外源核酸構建體序列藉由諸如同源重組的機制插入到基因組DNA中，以致使生殖細胞對於外源核酸序列係純合的。以這種方式，基因驅動系統能夠使希望的貨物基因滲入到整個育種群體中（甘茨（Gantz）等人，2015，用於群體修飾瘧疾載體蚊子斯氏按蚊的高效Cas9-介導的基因驅動（Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi），美國國家科學院院刊 2015，2015年11月23日先於印刷版公開，doi:10.1073/pnas.1521077112；厄斯維特（Esvelt）等人，2014，關於用於改變野生群體的RNA指導的基因驅動（Concerning RNA-guided gene drives for the alteration of wild populations） eLife 2014;3:e03401）。在選擇的實施方式中，可以選擇在基因組中具有幾個潛在脫靶位點的靶序列。使用多個指導RNA靶向靶座位內的多個位點可以增加切割頻率並且阻礙驅

動抵抗對偶基因的進化。截短的指導RNA可以減少脫靶切割。可以使用成對的切口酶代替核酸酶來進一步增加特異性。基因驅動構建體可以包括編碼轉錄調節劑的貨物序列，例如以啟動同源重組基因並且/或者阻遏非同源末端接合。靶位點可以在必需基因內選擇，以使得非同源末端接合事件可以引起致死性而不是形成驅動抵抗的對偶基因。基因驅動構建體可以被工程化為在一系列溫度下在一系列宿主中起作用（給（Cho）等人，2013，使用小分子快速且可調地控制秀麗隱桿線蟲中的蛋白質穩定性（Rapid and Tunable Control of Protein Stability in Caenorhabditis elegans Using a Small Molecule），公共科學圖書館綜合 8(8): e72393. doi:10.1371/journal.pone.0072393）。

異種移植

[1345] 本發明還考慮使用在此所述的CRISPR-Cas系統例如Cpf1效應蛋白系統來提供RNA指導的DNA核酸酶，該核酸酶適於用於提供用於移植的修飾組織。例如，RNA指導的DNA核酸酶可以用於例如藉由破壞編碼由人類免疫系統識別的表位的基因（即異種抗原基因）的表現來敲除、敲低或破壞動物諸如轉基因豬（諸如人類血紅素加氧酶-1轉基因豬系列）中的選定基因。用於破壞的候選豬基因可以例如包括α(1,3)-半乳糖基轉移酶和單磷酸胞苷-N-乙醯神經胺酸羥化酶基因（參見PCT專利公開WO 2014/066505）。此外，可以破壞編碼內源性逆轉錄病毒的基因，例如編碼所有豬內源性逆轉錄病毒的基因（參見楊等人，2015，豬內源性逆轉錄病毒（PERV）的全基因組滅活（Genome-wide inactivation of porcine

endogenous retroviruses (PERVs))，科學，2015年11月27日：第350卷，第6264期，第1101-1104頁）。此外，RNA-指導的DNA核酸酶可以用於靶向一個位點，以用於整合異種移植供體動物中的額外基因諸如人類CD55基因，以提高防止超急排斥。

一般基因治療考慮因素

[1346] 疾病相關基因和多核苷酸的實例和疾病特定資訊係從麥考斯克-南森遺傳醫學研究所（McKusick-Nathans Institute of Genetic Medicine）、約翰霍普金斯大學（馬里蘭州巴爾的摩）（Johns Hopkins University (Baltimore, Md.)）和國家生物技術資訊中心（National Center for Biotechnology Information）、國家醫學圖書館（馬里蘭州貝塞斯達）（National Library of Medicine (Bethesda, Md.)）可獲得的，在世界互聯網上可獲得的。

[1347] 該等基因和途徑中的突變可能導致產生影響功能的不適當蛋白質或不適當量的蛋白質。來自2012年12月12日提交的美國臨時申請61/736,527的基因、疾病和蛋白質的另外實例藉由引用結合在此。此類基因、蛋白質和途徑可以是本發明的CRISPR複合物的靶多核苷酸。疾病相關基因和多核苷酸的實例列出在表A和表B中。傳訊生物化學途徑相關基因和多核苷酸的實例列出在表C中。

[1348] 表A

疾病/病症	基因
瘤形成	PTEN；ATM；ATR；EGFR；ERBB2；ERBB3；ERBB4；
	Notch1；Notch2；Notch3；Notch4；AKT；AKT2；AKT3；HIF；
	HIF1a；HIF3a；Met；HRG；Bcl2；PPARα；PPAR

	γ；WT1（維爾姆斯瘤）；FGF受體家族
	成員（5個成員：1、2、3、4、5）；CDKN2a；APC；RB
	（成視網膜細胞瘤）；MEN1；VHL；BRCA1；BRCA2；AR
	（雄激素受體）TSG101；IGF；IGF受體；Igf1（4
	變體）；Igf2（3變體）；Igf 1受體；Igf 2受體；
	Bax；Bcl2；半胱天冬酶家族（9個成員：
	1、2、3、4、6、7、8、9、12）；Kras；Apc
年齡相關性黃斑	Abcr；Ccl2；Cc2；cp（血漿銅藍蛋白）；Timp3；組織蛋白酶D；
變性	Vldlr；Ccr2
精神分裂症	神經調節蛋白1（Nrg1）；Erb4（神經調節蛋白受體）；
	Complexin1（Cplx1）；Tph1色胺酸羥化酶；Tph2
	色胺酸羥化酶2；軸突蛋白1；GSK3；GSK3a；
	GSK3b
病症	5-HTT(Slc6a4)；COMT；DRD(Drd1a)；SLC6A3；DAOA；
	DTNBP1；Dao（Dao1）
三核苷酸重複序列	HTT(杭丁頓氏症(Huntington's Dx))；SBMA/SMAX1/AR（甘迺迪
病症	病）；FXN/X25（弗裡德賴希氏共濟失調）；ATX3（馬查多-約瑟夫病（Machado-
	Joseph's Dx））；ATXN1和ATXN2（脊髓小腦
	共濟失調）；DMPK（肌強直性營養不良）；萎縮蛋白-1和Atn1
	（DRPLA病）；CBP（Creb-BP -整體不穩定性）；VLDLR
	（阿茲海默症）；Atxn7；Atxn10
脆性X綜合症	FMR2；FXR1；FXR2；mGLUR5
分泌酶相關	APH-1（α和β）；早老素（Psen1）；呆蛋白
病症	（Ncstn）；PEN-2
其他	Nos1；Parp1；Nat1；Nat2
朊病毒相關病症	Prp
ALS	SOD1；ALS2；STEX；FUS；TARDBP；VEGF（VEGF-a；
	VEGF-b；VEGF-c）
藥物成癮	Prkce（醇）；Drd2；Drd4；ABAT（醇）；GRIA2；
	Grm5；Grin1；Htr1b；Grin2a；Drd3；Pdyn；Gria1（醇）；
自閉症	Mecp2；BZRAP1；MDGA2；Sema5A；軸突蛋白1；脆性X
	（FMR2（AFF2）；FXR1；FXR2；Mglur5）
阿茲海默症	E1；CHIP；UCH；UBB；Tau；LRP；PICALM；叢生蛋白；PS1；
	SORL1；CR1；Vldlr；Uba1；Uba3；CHIP28（Aqp1、
	水通道蛋白1）；Uchl1；Uchl3；APP
炎症	IL-10；IL-1(IL-1a；IL-1b)；IL-13；IL-17(IL-17a(CTLA8)；IL-
	17b；IL-17c；IL-17d；IL-17f)；II-23；Cx3cr1；ptpn22；TNFa；

	用於IBD的NOD2/CARD15 ;IL-6 ;IL-12(IL-12a ;IL-12b)；
	CTLA4； Cx3cl1
帕金森病	x-突觸核蛋白； DJ-1； LRRK2； Parkin； PINK1

[1349] 表B：

血液和凝血疾病和病症	貧血(CDAN1、CDA1、RPS19、DBA、PKLR、PK1、NT5C3、UMPH1、PSN1、RHAG、RH50A、NRAMP2、SPTB、ALAS2、ANH1、ASB、ABCB7、ABC7、ASAT)；裸淋巴細胞綜合症(TAPBP、TPSN、TAP2、ABCB3、PSF2、RING11、MHC2TA、C2TA、RFX5、RFXAP、RFX5))、出血病症(TBXA2R、P2RX1、P2X1)；因子H和因子H-樣1(HF1、CFH、HUS)；因子V和因子VIII(MCFD2)；因子VII缺乏症(F7)；因子X缺乏症(F10)；因子XI缺乏症(F11)；因子XII缺乏症(F12)；因子XIIIA缺乏症(F13A1、F13A)；因子XIIIB缺乏症(F13B)；範科尼貧血(FANCA、FACA、FA1、FA、FAA、FAAP95、FAAP90、FLJ34064、FANCB、FANCC、FACC、BRCA2、FANCD1、FANCD2、FANCD、FACD、FAD、FANCE、FACE、FANCF、XRCC9、FANCG、BRIP1、BACH1、FANCJ、PHF9、FANCL、FANCM、KIAA1596)；噬血細胞性淋巴組織細胞增生病症(PRF1、HPLH2、UNC13D、MUNC13-4、HPLH3、HLH3、FHL3)；A型血友病(F8、F8C、HEMA)；B型血友病(F9、HEMB)、出血性病症(PI、ATT、F5)；白細胞缺乏症和病症(ITGB2、CD18、LCAMB、LAD、EIF2B1、EIF2BA、EIF2B2、EIF2B3、EIF2B5、LVWM、CACH、CLE、EIF2B4)；鐮狀細胞貧血(HBB)；地中海貧血(HBA2、HBB、HBD、LCRB、HBA1)。
細胞調節異常和腫瘤疾病和病症	B細胞非霍奇金淋巴瘤(BCL7A、BCL7)；白血病(TAL1、TCL5、SCL、TAL2、FLT3、NBS1、NBS、ZNFN1A1、IK1、LYF1、HOXD4、HOX4B、BCR、CML、PHL、ALL、ARNT、KRAS2、RASK2、GMPS、AF10、ARHGEF12、LARG、KIAA0382、CALM、CLTH、CEBPA、CEBP、CHIC2、BTL、FLT3、KIT、PBT、LPP、NPM1、NUP214、D9S46E、CAN、CAIN、RUNX1、CBFA2、AML1、WHSC1L1、NSD3、FLT3、AF1Q、NPM1、NUMA1、ZNF145、PLZF、PML、MYL、STAT5B、AF10、CALM、CLTH、ARL11、ARLTS1、P2RX7、P2X7、BCR、CML、PHL、ALL、GRAF、NF1、VRNF、WSS、NFNS、PTPN11、PTP2C、SHP2、NS1、BCL2、CCND1、PRAD1、BCL1、TCRA、GATA1、GF1、ERYF1、NFE1、ABL1、NQO1、DIA4、NMOR1、NUP214、D9S46E、CAN、CAIN)。
炎症和免疫相關疾病和病症	AIDS(KIR3DL1、NKAT3、NKB1、AMB11、KIR3DS1、IFNG、CXCL12、SDF1)；自身免疫性淋巴組織增生性綜合症(TNFRSF6、APT1、FAS、CD95、ALPS1A)；聯合免疫缺陷(IL2RG、SCIDX1、SCIDX、IMD4)；HIV-1(CCL5、SCYA5、D17S136E、TCP228)、HIV易感性或感染(IL10、CSIF、

	CMKBR2、CCR2、CMKBR5、CCCKR5（CCR5））；免疫缺陷（CD3E、CD3G、AICDA、AID、HIGM2、TNFRSF5、CD40、UNG、DGU、HIGM4、TNFSF5、CD40LG、HIGM1、IGM、FOXP3、IPEX、AIID、XPID、PIDX、TNFRSF14B、TACI）；炎症（IL-10、IL-1（IL-1a、IL-1b）、IL-13、IL-17（IL-17a（CTLA8）、IL-17b、IL-17c、IL-17d、IL-17f）、II-23、Cx3cr1、ptpn22、TNFa、用於IBD的NOD2/CARD15、IL-6、IL-12（IL-12a、IL-12b）、CTLA4、Cx3cl1）；嚴重聯合免疫缺陷（SCID）（JAK3、JAKL、DCLRE1C、ARTEMIS、SCIDA、RAG1、RAG2、ADA、PTPRC、CD45、LCA、IL7R、CD3D、T3D、IL2RG、SCIDX1、SCIDX、IMD4）。
代謝性、肝臟、腎臟以及蛋白質疾病和病症	澱粉樣變性神經病（TTR、PALB）；澱粉樣變性（APOA1、APP、AAA、CVAP、AD1、GSN、FGA、LYZ、TTR、PALB）；肝硬化（KRT18、KRT8、CIRH1A、NAIC、TEX292、KIAA1988）；囊性纖維化（CFTR、ABCC7、CF、MRP7）；肝糖貯積症（SLC2A2、GLUT2、G6PC、G6PT、G6PT1、GAA、LAMP2、LAMPB、 AGL、GDE、GBE1、GYS2、PYGL、PFKM）；肝腺瘤、142330 （TCF1、HNF1A、MODY3）、肝衰竭、早髮型和神經系統病症（SCOD1、SCO1）、肝脂酶缺乏症 （LIPC）、肝母細胞癌、癌症和癌（CTNNB1、PDGFRL、PDGRL、PRLTS、AXIN1、AXIN、CTNNB1、TP53、P53、LFS1、IGF2R、MPRI、MET、CASP8、MCH5;髓質囊性腎病（UMOD、HNFJ、FJHN、MCKD2、ADMCKD2）；苯丙酮尿症（PAH、PKU1、QDPR、DHPR、PTS）；多囊腎和肝臟疾病（FCYT、PKHD1、ARPKD、PKD1、PKD2、PKD4、PKDTS、PRKCSH、G19P1、PCLD、SEC63）。
肌肉/骨骼疾病和病症	貝克肌營養不良 （DMD、BMD、MYF6）、迪謝內肌營養不良（DMD、BMD）；埃-德二氏肌營養不良（Emery-Dreifuss muscular dystrophy） （LMNA、LMN1、EMD2、FPLD、CMD1A、HGPS、LGMD1B、LMNA、LMN1、EMD2、FPLD、CMD1A）；面肩胛臂肌營養不良（FSHMD1A、FSHD1A）；肌營養不良（FKRP、MDC1C、LGMD2I、LAMA2、LAMM、LARGE、KIAA0609、MDC1D、FCMD、TTID、MYOT、CAPN3、CANP3、DYSF、LGMD2B、SGCG、LGMD2C、DMDA1、SCG3、SGCA、ADL、DAG2、LGMD2D、DMDA2、SGCB、LGMD2E、SGCD、SGD、LGMD2F、CMD1L、TCAP、LGMD2G、CMD1N、TRIM32、HT2A、LGMD2H、FKRP、MDC1C、LGMD2I、TTN、CMD1G、TMD、LGMD2J、POMT1、CAV3、LGMD1C、SEPN1、SELN、RSMD1、PLEC1、PLTN、EBS1）；骨硬化病（LRP5、BMND1、LRP7、LR3、OPPG、VBCH2、CLCN7、CLC7、OPTA2、OSTM1、GL、TCIRG1、TIRC7、OC116、OPTB1）；肌肉萎縮（VAPB、VAPC、ALS8、SMN1、SMA1、SMA2、SMA3、SMA4、BSCL2、SPG17、GARS、SMAD1、CMT2D、HEXB、IGHMBP2、SMUBP2、

	CATF1、SMARD1）。
神經學和神經元疾病和病症	ALS（SOD1、ALS2、STEX、FUS、TARDBP、VEGF（VEGF-a、VEGF-b、VEGF-c）；阿茲海默症（APP、AAA、CVAP、AD1、APOE、AD2、PSEN2、AD4、STM2、APBB2、FE65L1、NOS3、PLAU、URK、ACE、DCP1、ACE1、MPO、PACIP1、PAXIP1L、PTIP、A2M、BLMH、BMH、PSEN1、AD3）；自閉症（Mecp2、BZRAP1、MDGA2、Sema5A、軸突蛋白1、GLO1、MECP2、RTT、PPMX、MRX16、MRX79、NLGN3、NLGN4、KIAA1260、AUTSX2）；脆性X綜合症（FMR2、FXR1、FXR2、mGLUR5）；杭丁頓氏症和病樣障礙（HD、IT15、PRNP、PRIP、JPH3、JP3、HDL2、TBP、SCA17）；帕金森病（NR4A2、NURR1、NOT、TINUR、SNCAIP、TBP、SCA17、SNCA、NACP、PARK1、PARK4、DJ1、PARK7、LRRK2、PARK8、PINK1、PARK6、UCHL1、PARK5、SNCA、NACP、PARK1、PARK4、PRKN、PARK2、PDJ、DBH、NDUFV2）；雷特氏綜合症（MECP2、RTT、PPMX、MRX16、MRX79、CDKL5、STK9、MECP2、RTT、PPMX、MRX16、MRX79、x-Synuclein、DJ-1）；精神分裂症（神經調節蛋白1（Nrg1）、Erb4（神經調節蛋白的受體）、Complexin1（Cplx1）、Tph1色胺酸羥化酶、Tph2、色胺酸羥化酶2、軸突蛋白1、GSK3、GSK3a、GSK3b、5-HTT（Slc6a4）、COMT、DRD（Drd1a）、SLC6A3、DAOA、DTNBP1、Dao（Dao1））；分泌酶相關病症（APH-1（α和β）、早老蛋白（Psen1）、呆蛋白（Ncstn）、PEN-2、Nos1、Parp1、Nat1、Nat2）；三核苷酸重複序列病症（HTT（杭丁頓氏症）、SBMA/SMAX1/AR（甘迺迪病）、FXN/X25（弗裡德賴希氏共濟失調）、ATX3（馬查多-約瑟夫病）、ATXN1和ATXN2（脊髓小腦共濟失調）、DMPK（肌強直性營養不良）、萎縮蛋白-1和Atn1（DRPLA Dx）、CBP（Creb-BP-整體不穩定性）、VLDLR（阿茲海默症）、Atxn7、Atxn10）。
眼部疾病和病症	年齡相關性黃斑變性（Abcr、Ccl2、Cc2、cp（血漿銅藍蛋白）、Timp3、組織蛋白酶D、Vldlr、Ccr2）；白內障（CRYAA、CRYA1、CRYBB2、CRYB2、PITX3、BFSP2、CP49、CP47、CRYAA、CRYA1、PAX6、AN2、MGDA、CRYBA1、CRYB1、CRYGC、CRYG3、CCL、LIM2、MP19、CRYGD、CRYG4、BFSP2、CP49、CP47、HSF4、CTM、HSF4、CTM、MIP、AQP0、CRYAB、CRYA2、CTPP2、CRYBB1、CRYGD、CRYG4、CRYBB2、CRYB2、CRYGC、CRYG3、CCL、CRYAA、CRYA1、GJA8、CX50、CAE1、GJA3、CX46、CZP3、CAE3、CCM1、CAM、KRIT1）；角膜渾濁和營養不良（APOA1、TGFBI、CSD2、CDGG1、CSD、BIGH3、CDG2、TACSTD2、TROP2、M1S1、VSX1、RINX、PPCD、PPD、KTCN、COL8A2、FECD、PPCD2、PIP5K3、CFD）；先天性扁平角膜（KERA、CNA2）；青光眼（MYOC、TIGR、GLC1A、JOAG、GPOA、OPTN、

	GLC1E、FIP2、HYPL、NRP、CYP1B1、GLC3A、OPA1、NTG、NPG、CYP1B1、GLC3A）；萊伯氏先天性黑矇（CRB1、RP12、CRX、CORD2、CRD、RPGRIP1、LCA6、CORD9、RPE65、RP20、AIPL1、LCA4、GUCY2D、GUC2D、LCA1、CORD6、RDH12、LCA3）；黃斑營養不良（ELOVL4、ADMD、STGD2、STGD3、RDS、RP7、PRPH2、PRPH、AVMD、AOFMD、VMD2）。

[1350]

[1351] 表C：

細胞功能	基因
PI3K/AKT傳訊	PRKCE ;ITGAM ;ITGA5 ;IRAK1 ;PRKAA2 ;EIF2AK2 ;
	PTEN ;EIF4E ;PRKCZ ;GRK6 ;MAPK1 ;TSC1 ;PLK1 ;
	AKT2 ; IKBKB ; PIK3CA ; CDK8 ; CDKN1B ; NFKB2 ; BCL2 ;
	PIK3CB ;PPP2R1A ;MAPK8 ;BCL2L1 ;MAPK3 ;TSC2 ;
	ITGA1 ; KRAS ; EIF4EBP1 ; RELA ; PRKCD ; NOS3 ;
	PRKAA1 ; MAPK9 ; CDK2 ; PPP2CA ; PIM1 ; ITGB7 ;
	YWHAZ ;ILK ;TP53 ;RAF1 ;IKBKG ;RELB ;DYRK1A ;
	CDKN1A ; ITGB1 ; MAP2K2 ; JAK1 ; AKT1 ; JAK2 ; PIK3R1 ;
	CHUK ; PDPK1 ; PPP2R5C ; CTNNB1 ; MAP2K1 ; NFKB1 ;
	PAK3 ;ITGB3 ;CCND1 ;GSK3A ;FRAP1 ;SFN ;ITGA2 ;
	TTK ; CSNK1A1 ; BRAF ; GSK3B ; AKT3 ; FOXO1 ; SGK ;
	HSP90AA1 ; RPS6KB1
ERK/MAPK傳訊	PRKCE ; ITGAM ; ITGA5 ; HSPB1 ; IRAK1 ; PRKAA2 ;
	EIF2AK2 ;RAC1 ;RAP1A ;TLN1 ;EIF4E ;ELK1 ;GRK6 ;
	MAPK1 ;RAC2 ;PLK1 ;AKT2 ;PIK3CA ;CDK8 ;CREB1 ;
	PRKCI ; PTK2 ; FOS ; RPS6KA4 ; PIK3CB ; PPP2R1A ;
	PIK3C3 ; MAPK8 ; MAPK3 ; ITGA1 ; ETS1 ; KRAS ; MYCN ;
	EIF4EBP1 ;PPARG ;PRKCD ;PRKAA1 ;MAPK9 ;SRC ;
	CDK2 ; PPP2CA ; PIM1 ; PIK3C2A ; ITGB7 ; YWHAZ ;
	PPP1CC ;KSR1 ;PXN ;RAF1 ;FYN ;DYRK1A ;ITGB1 ;
	MAP2K2 ;PAK4 ;PIK3R1 ;STAT3 ;PPP2R5C ;MAP2K1 ;
	PAK3 ;ITGB3 ;ESR1 ;ITGA2 ;MYC ;TTK ;CSNK1A1 ;
	CRKL ; BRAF ; ATF4 ; PRKCA ; SRF ; STAT1 ; SGK
糖皮質激素受體	RAC1 ;TAF4B ;EP300 ;SMAD2 ;TRAF6 ;PCAF ;ELK1 ;
傳訊	MAPK1 ; SMAD3 ; AKT2 ; IKBKB ; NCOR2 ; UBE2I ;
	PIK3CA ; CREB1 ; FOS ; HSPA5 ; NFKB2 ; BCL2 ;
	MAP3K14 ; STAT5B ; PIK3CB ; PIK3C3 ; MAPK8 ;

	BCL2L1 ;
	MAPK3 ;TSC22D3 ;MAPK10 ;NRIP1 ;KRAS ;MAPK13 ;
	RELA ; STAT5A ; MAPK9 ; NOS2A ; PBX1 ; NR3C1 ;
	PIK3C2A ; CDKN1C ; TRAF2 ; SERPINE1 ; NCOA3 ;
	MAPK14 ; TNF ; RAF1 ; IKBKG ; MAP3K7 ; CREBBP ;
	CDKN1A ; MAP2K2 ; JAK1 ; IL8 ; NCOA2 ; AKT1 ; JAK2 ;
	PIK3R1 ;CHUK ;STAT3 ;MAP2K1 ;NFKB1 ;TGFBR1 ;
	ESR1 ; SMAD4 ; CEBPB ; JUN ; AR ; AKT3 ; CCL2 ; MMP1 ;
	STAT1 ; IL6 ; HSP90AA1
軸突指導傳訊	PRKCE ;ITGAM ;ROCK1 ;ITGA5 ;CXCR4 ;ADAM12 ;
	IGF1 ;RAC1 ;RAP1A ;EIF4E ;PRKCZ ;NRP1 ;NTRK2 ;
	ARHGEF7 ; SMO ; ROCK2 ; MAPK1 ; PGF ; RAC2 ;
	PTPN11 ; GNAS ; AKT2 ; PIK3CA ; ERBB2 ; PRKCI ; PTK2 ;
	CFL1 ; GNAQ ; PIK3CB ; CXCL12 ; PIK3C3 ; WNT11 ;
	PRKD1 ; GNB2L1 ; ABL1 ; MAPK3 ; ITGA1 ; KRAS ; RHOA ;
	PRKCD ;PIK3C2A ;ITGB7 ;GLI2 ;PXN ;VASP ;RAF1 ;
	FYN ; ITGB1 ; MAP2K2 ; PAK4 ; ADAM17 ; AKT1 ; PIK3R1 ;
	GLI1 ; WNT5A ; ADAM10 ; MAP2K1 ; PAK3 ; ITGB3 ;
	CDC42 ; VEGFA ; ITGA2 ; EPHA8 ; CRKL ; RND1 ; GSK3B ;
	AKT3 ; PRKCA
肝配蛋白受體傳訊	PRKCE ; ITGAM ; ROCK1 ; ITGA5 ; CXCR4 ; IRAK1 ;
	PRKAA2 ;EIF2AK2 ;RAC1 ;RAP1A ;GRK6 ;ROCK2 ;
	MAPK1 ;PGF ;RAC2 ;PTPN11 ;GNAS ;PLK1 ;AKT2 ;
	DOK1 ; CDK8 ; CREB1 ; PTK2 ; CFL1 ; GNAQ ; MAP3K14 ;
	CXCL12 ;MAPK8 ;GNB2L1 ;ABL1 ;MAPK3 ;ITGA1 ;
	KRAS ; RHOA ; PRKCD ; PRKAA1 ; MAPK9 ; SRC ; CDK2 ;
	PIM1 ;ITGB7 ;PXN ;RAF1 ;FYN ;DYRK1A ;ITGB1 ;
	MAP2K2 ; PAK4 ; AKT1 ; JAK2 ; STAT3 ; ADAM10 ;
	MAP2K1 ; PAK3 ; ITGB3 ; CDC42 ; VEGFA ; ITGA2 ;
	EPHA8 ; TTK ; CSNK1A1 ; CRKL ; BRAF ; PTPN13 ; ATF4 ;
	AKT3 ; SGK
肌動蛋白細胞骨架	ACTN4 ; PRKCE ; ITGAM ; ROCK1 ; ITGA5 ; IRAK1 ;
傳訊	PRKAA2 ; EIF2AK2 ; RAC1 ; INS ; ARHGEF7 ; GRK6 ;
	ROCK2 ; MAPK1 ; RAC2 ; PLK1 ; AKT2 ; PIK3CA ; CDK8 ;
	PTK2 ; CFL1 ; PIK3CB ; MYH9 ; DIAPH1 ; PIK3C3 ;

	MAPK8；
	F2R； MAPK3； SLC9A1； ITGA1； KRAS； RHOA； PRKCD；
	PRKAA1；MAPK9；CDK2；PIM1；PIK3C2A；ITGB7；
	PPP1CC；PXN；VIL2；RAF1；GSN；DYRK1A；ITGB1；
	MAP2K2；PAK4；PIP5K1A；PIK3R1；MAP2K1；PAK3；
	ITGB3；CDC42；APC；ITGA2；TTK；CSNK1A1；CRKL；
	BRAF； VAV3； SGK
杭丁頓氏症	PRKCE； IGF1； EP300； RCOR1； PRKCZ； HDAC4； TGM2；
傳訊	MAPK1； CAPNS1； AKT2； EGFR； NCOR2； SP1； CAPN2；
	PIK3CA； HDAC5； CREB1； PRKCI； HSPA5； REST；
	GNAQ；PIK3CB；PIK3C3；MAPK8；IGF1R；PRKD1；
	GNB2L1；BCL2L1；CAPN1；MAPK3；CASP8；HDAC2；
	HDAC7A； PRKCD； HDAC11； MAPK9； HDAC9； PIK3C2A；
	HDAC3； TP53； CASP9； CREBBP； AKT1； PIK3R1；
	PDPK1；CASP1；APAF1；FRAP1；CASP2；JUN；BAX；
	ATF4；AKT3；PRKCA；CLTC；SGK；HDAC6；CASP3
細胞凋亡傳訊	PRKCE；ROCK1；BID；IRAK1；PRKAA2；EIF2AK2； BAK1；
	BIRC4； GRK6； MAPK1； CAPNS1； PLK1； AKT2； IKBKB；
	CAPN2； CDK8； FAS； NFKB2； BCL2； MAP3K14； MAPK8；
	BCL2L1； CAPN1； MAPK3； CASP8； KRAS； RELA；
	PRKCD； PRKAA1； MAPK9； CDK2； PIM1； TP53； TNF；
	RAF1；IKBKG；RELB；CASP9；DYRK1A；MAP2K2；
	CHUK； APAF1； MAP2K1； NFKB1； PAK3； LMNA； CASP2；
	BIRC2；TTK；CSNK1A1；BRAF；BAX；PRKCA；SGK；
	CASP3； BIRC3； PARP1
B細胞受體傳訊	RAC1；PTEN；LYN；ELK1；MAPK1；RAC2；PTPN11；
	AKT2； IKBKB； PIK3CA； CREB1； SYK； NFKB2； CAMK2A；
	MAP3K14；PIK3CB；PIK3C3；MAPK8；BCL2L1；ABL1；
	MAPK3； ETS1； KRAS； MAPK13； RELA； PTPN6； MAPK9；
	EGR1； PIK3C2A； BTK； MAPK14； RAF1； IKBKG； RELB；
	MAP3K7；MAP2K2；AKT1；PIK3R1；CHUK；MAP2K1；
	NFKB1； CDC42； GSK3A； FRAP1； BCL6； BCL10； JUN；

	GSK3B；ATF4；AKT3；VAV3；RPS6KB1
白細胞外滲	ACTN4；CD44；PRKCE；ITGAM；ROCK1；CXCR4；CYBA；
傳訊	RAC1；RAP1A；PRKCZ；ROCK2；RAC2；PTPN11；
	MMP14；PIK3CA；PRKCI；PTK2；PIK3CB；CXCL12；
	PIK3C3；MAPK8；PRKD1；ABL1；MAPK10；CYBB；
	MAPK13；RHOA；PRKCD；MAPK9；SRC；PIK3C2A；BTK；
	MAPK14；NOX1；PXN；VIL2；VASP；ITGB1；MAP2K2；
	CTNND1；PIK3R1；CTNNB1；CLDN1；CDC42；F11R；ITK；
	CRKL；VAV3；CTTN；PRKCA；MMP1；MMP9
整聯蛋白傳訊	ACTN4；ITGAM；ROCK1；ITGA5；RAC1；PTEN；RAP1A；
	TLN1；ARHGEF7；MAPK1；RAC2；CAPNS1；AKT2；
	CAPN2；PIK3CA；PTK2；PIK3CB；PIK3C3；MAPK8；
	CAV1；CAPN1；ABL1；MAPK3；ITGA1；KRAS；RHOA；
	SRC；PIK3C2A；ITGB7；PPP1CC；ILK；PXN；VASP；
	RAF1；FYN；ITGB1；MAP2K2；PAK4；AKT1；PIK3R1；
	TNK2；MAP2K1；PAK3；ITGB3；CDC42；RND3；ITGA2；
	CRKL；BRAF；GSK3B；AKT3
急相反應	IRAK1；SOD2；MYD88；TRAF6；ELK1；MAPK1；PTPN11；
傳訊	AKT2；IKBKB；PIK3CA；FOS；NFKB2；MAP3K14；
	PIK3CB；MAPK8；RIPK1；MAPK3；IL6ST；KRAS；
	MAPK13；IL6R；RELA；SOCS1；MAPK9；FTL；NR3C1；
	TRAF2；SERPINE1；MAPK14；TNF；RAF1；PDK1；
	IKBKG；RELB；MAP3K7；MAP2K2；AKT1；JAK2；PIK3R1；
	CHUK；STAT3；MAP2K1；NFKB1；FRAP1；CEBPB；JUN；
	AKT3；IL1R1；IL6
PTEN傳訊	ITGAM；ITGA5；RAC1；PTEN；PRKCZ；BCL2L11；
	MAPK1；RAC2；AKT2；EGFR；IKBKB；CBL；PIK3CA；
	CDKN1B；PTK2；NFKB2；BCL2；PIK3CB；BCL2L1；
	MAPK3；ITGA1；KRAS；ITGB7；ILK；PDGFRB；INSR；
	RAF1；IKBKG；CASP9；CDKN1A；ITGB1；MAP2K2；
	AKT1；PIK3R1；CHUK；PDGFRA；PDPK1；MAP2K1；
	NFKB1；ITGB3；CDC42；CCND1；GSK3A；ITGA2；
	GSK3B；AKT3；FOXO1；CASP3；RPS6KB1
p53傳訊	PTEN；EP300；BBC3；PCAF；FASN；BRCA1；GADD45A；
	BIRC5；AKT2；PIK3CA；CHEK1；TP53INP1；BCL2；
	PIK3CB；PIK3C3；MAPK8；THBS1；ATR；BCL2L1；

	E2F1 ;
	PMAIP1 ; CHEK2 ; TNFRSF10B ; TP73 ; RB1 ; HDAC9 ;
	CDK2 ; PIK3C2A ; MAPK14 ; TP53 ; LRDD ; CDKN1A ;
	HIPK2 ; AKT1 ; PIK3R1 ; RRM2B ; APAF1 ; CTNNB1 ;
	SIRT1 ; CCND1 ; PRKDC ; ATM ; SFN ; CDKN2A ; JUN ;
	SNAI2 ; GSK3B ; BAX ; AKT3
芳香烴受體	HSPB1 ; EP300 ; FASN ; TGM2 ; RXRA ; MAPK1 ; NQO1 ;
傳訊	NCOR2 ; SP1 ; ARNT ; CDKN1B ; FOS ; CHEK1 ;
	SMARCA4 ; NFKB2 ; MAPK8 ; ALDH1A1 ; ATR ; E2F1 ;
	MAPK3 ; NRIP1 ; CHEK2 ; RELA ; TP73 ; GSTP1 ; RB1 ;
	SRC ; CDK2 ; AHR ; NFE2L2 ; NCOA3 ; TP53 ; TNF ;
	CDKN1A ; NCOA2 ; APAF1 ; NFKB1 ; CCND1 ; ATM ; ESR1 ;
	CDKN2A ; MYC ; JUN ; ESR2 ; BAX ; IL6 ; CYP1B1 ;
	HSP90AA1
異生物質代謝	PRKCE ; EP300 ; PRKCZ ; RXRA ; MAPK1 ; NQO1 ;
傳訊	NCOR2 ; PIK3CA ; ARNT ; PRKCI ; NFKB2 ; CAMK2A ;
	PIK3CB ; PPP2R1A ; PIK3C3 ; MAPK8 ; PRKD1 ;
	ALDH1A1 ; MAPK3 ; NRIP1 ; KRAS ; MAPK13 ; PRKCD ;
	GSTP1 ; MAPK9 ; NOS2A ; ABCB1 ; AHR ; PPP2CA ; FTL ;
	NFE2L2 ; PIK3C2A ; PPARGC1A ; MAPK14 ; TNF ; RAF1 ;
	CREBBP ; MAP2K2 ; PIK3R1 ; PPP2R5C ; MAP2K1 ;
	NFKB1 ; KEAP1 ; PRKCA ; EIF2AK3 ; IL6 ; CYP1B1 ;
	HSP90AA1
SAPK/JNK傳訊	PRKCE ; IRAK1 ; PRKAA2 ; EIF2AK2 ; RAC1 ; ELK1 ;
	GRK6 ; MAPK1 ; GADD45A ; RAC2 ; PLK1 ; AKT2 ; PIK3CA ;
	FADD ; CDK8 ; PIK3CB ; PIK3C3 ; MAPK8 ; RIPK1 ;
	GNB2L1 ; IRS1 ; MAPK3 ; MAPK10 ; DAXX ; KRAS ;
	PRKCD ; PRKAA1 ; MAPK9 ; CDK2 ; PIM1 ; PIK3C2A ;
	TRAF2 ; TP53 ; LCK ; MAP3K7 ; DYRK1A ; MAP2K2 ;
	PIK3R1 ; MAP2K1 ; PAK3 ; CDC42 ; JUN ; TTK ; CSNK1A1 ;
	CRKL ; BRAF ; SGK
PPAr/RXR傳訊	PRKAA2 ; EP300 ; INS ; SMAD2 ; TRAF6 ; PPARA ; FASN ;
	RXRA ; MAPK1 ; SMAD3 ; GNAS ; IKBKB ; NCOR2 ;
	ABCA1 ; GNAQ ; NFKB2 ; MAP3K14 ; STAT5B ; MAPK8 ;
	IRS1 ; MAPK3 ; KRAS ; RELA ; PRKAA1 ; PPARGC1A ;
	NCOA3 ; MAPK14 ; INSR ; RAF1 ; IKBKG ; RELB ; MAP3K7 ;
	CREBBP ; MAP2K2 ; JAK2 ; CHUK ; MAP2K1 ; NFKB1 ;
	TGFBR1 ; SMAD4 ; JUN ; IL1R1 ; PRKCA ; IL6 ;

	HSP90AA1；
	ADIPOQ
NF-KB傳訊	IRAK1； EIF2AK2； EP300； INS； MYD88； PRKCZ； TRAF6；
	TBK1；AKT2；EGFR；IKBKB；PIK3CA；BTRC；NFKB2；
	MAP3K14；PIK3CB；PIK3C3；MAPK8；RIPK1；HDAC2；
	KRAS； RELA； PIK3C2A； TRAF2； TLR4； PDGFRB； TNF；
	INSR； LCK； IKBKG； RELB； MAP3K7； CREBBP； AKT1；
	PIK3R1； CHUK； PDGFRA； NFKB1； TLR2； BCL10；
	GSK3B； AKT3； TNFAIP3； IL1R1
神經調節蛋白傳訊	ERBB4； PRKCE； ITGAM； ITGA5； PTEN； PRKCZ； ELK1；
	MAPK1； PTPN11； AKT2； EGFR； ERBB2； PRKCI；
	CDKN1B；STAT5B；PRKD1；MAPK3；ITGA1；KRAS；
	PRKCD； STAT5A； SRC； ITGB7； RAF1； ITGB1； MAP2K2；
	ADAM17；AKT1；PIK3R1；PDPK1；MAP2K1；ITGB3；
	EREG；FRAP1；PSEN1；ITGA2；MYC；NRG1；CRKL；
	AKT3； PRKCA； HSP90AA1； RPS6KB1
Wnt和β鏈蛋白	CD44；EP300；LRP6；DVL3；CSNK1E；GJA1；SMO；
傳訊	AKT2； PIN1； CDH1； BTRC； GNAQ； MARK2； PPP2R1A；
	WNT11；SRC；DKK1；PPP2CA；SOX6；SFRP2；ILK；
	LEF1； SOX9； TP53； MAP3K7； CREBBP； TCF7L2； AKT1；
	PPP2R5C； WNT5A； LRP5； CTNNB1； TGFBR1； CCND1；
	GSK3A； DVL1； APC； CDKN2A； MYC； CSNK1A1； GSK3B；
	AKT3； SOX2
胰島素受體傳訊	PTEN；INS；EIF4E；PTPN1；PRKCZ；MAPK1；TSC1；
	PTPN11； AKT2； CBL； PIK3CA； PRKCI； PIK3CB； PIK3C3；
	MAPK8； IRS1； MAPK3； TSC2； KRAS； EIF4EBP1；
	SLC2A4； PIK3C2A； PPP1CC； INSR； RAF1； FYN；
	MAP2K2； JAK1； AKT1； JAK2； PIK3R1； PDPK1； MAP2K1；
	GSK3A； FRAP1； CRKL； GSK3B； AKT3； FOXO1； SGK；
	RPS6KB1
IL-6傳訊	HSPB1；TRAF6；MAPKAPK2；ELK1；MAPK1；PTPN11；
	IKBKB；FOS；NFKB2；MAP3K14；MAPK8；MAPK3；
	MAPK10； IL6ST； KRAS； MAPK13； IL6R； RELA；

	SOCS1；
	MAPK9； ABCB1； TRAF2； MAPK14； TNF； RAF1； IKBKG；
	RELB；MAP3K7；MAP2K2；IL8；JAK2；CHUK；STAT3；
	MAP2K1； NFKB1； CEBPB； JUN； IL1R1； SRF； IL6
肝膽汁淤積	PRKCE； IRAK1； INS； MYD88； PRKCZ； TRAF6； PPARA；
	RXRA；IKBKB；PRKCI；NFKB2；MAP3K14；MAPK8；
	PRKD1；MAPK10；RELA；PRKCD；MAPK9；ABCB1；
	TRAF2；TLR4；TNF；INSR；IKBKG；RELB；MAP3K7； IL8；
	CHUK；NR1H2；TJP2；NFKB1；ESR1；SREBF1；FGFR4；
	JUN； IL1R1； PRKCA； IL6
IGF-1傳訊	IGF1； PRKCZ； ELK1； MAPK1； PTPN11； NEDD4； AKT2；
	PIK3CA； PRKCI； PTK2； FOS； PIK3CB； PIK3C3； MAPK8；
	IGF1R； IRS1； MAPK3； IGFBP7； KRAS； PIK3C2A；
	YWHAZ； PXN； RAF1； CASP9； MAP2K2； AKT1； PIK3R1；
	PDPK1；MAP2K1；IGFBP2；SFN；JUN；CYR61；AKT3；
	FOXO1； SRF； CTGF； RPS6KB1
NRF2-介導的氧化	PRKCE；EP300；SOD2；PRKCZ；MAPK1；SQSTM1；
脅迫應答	NQO1； PIK3CA； PRKCI； FOS； PIK3CB； PIK3C3； MAPK8；
	PRKD1； MAPK3； KRAS； PRKCD； GSTP1； MAPK9； FTL；
	NFE2L2； PIK3C2A； MAPK14； RAF1； MAP3K7； CREBBP；
	MAP2K2； AKT1； PIK3R1； MAP2K1； PPIB； JUN； KEAP1；
	GSK3B； ATF4； PRKCA； EIF2AK3； HSP90AA1
肝纖維化/肝	EDN1； IGF1； KDR； FLT1； SMAD2； FGFR1； MET； PGF；
星狀細胞啟動	SMAD3；EGFR；FAS；CSF1；NFKB2；BCL2；MYH9；
	IGF1R；IL6R；RELA；TLR4；PDGFRB；TNF；RELB； IL8；
	PDGFRA；NFKB1；TGFBR1；SMAD4；VEGFA；BAX；
	IL1R1； CCL2； HGF； MMP1； STAT1； IL6； CTGF； MMP9
PPAR傳訊	EP300；INS；TRAF6；PPARA；RXRA；MAPK1；IKBKB；
	NCOR2；FOS；NFKB2；MAP3K14；STAT5B；MAPK3；
	NRIP1； KRAS； PPARG； RELA； STAT5A； TRAF2；
	PPARGC1A；PDGFRB；TNF；INSR；RAF1；IKBKG；
	RELB； MAP3K7； CREBBP； MAP2K2； CHUK；

	PDGFRA；
	MAP2K1；NFKB1；JUN；IL1R1；HSP90AA1
Fc ε RI傳訊	PRKCE；RAC1；PRKCZ；LYN；MAPK1；RAC2；PTPN11；
	AKT2；PIK3CA；SYK；PRKCI；PIK3CB；PIK3C3；MAPK8；
	PRKD1；MAPK3；MAPK10；KRAS；MAPK13；PRKCD；
	MAPK9；PIK3C2A；BTK；MAPK14；TNF；RAF1；FYN；
	MAP2K2；AKT1；PIK3R1；PDPK1；MAP2K1；AKT3；
	VAV3；PRKCA
G蛋白偶聯	PRKCE；RAP1A；RGS16；MAPK1；GNAS；AKT2；IKBKB；
受體傳訊	PIK3CA；CREB1；GNAQ；NFKB2；CAMK2A；PIK3CB；
	PIK3C3；MAPK3；KRAS；RELA；SRC；PIK3C2A；RAF1；
	IKBKG；RELB；FYN；MAP2K2；AKT1；PIK3R1；CHUK；
	PDPK1；STAT3；MAP2K1；NFKB1；BRAF；ATF4；AKT3；
	PRKCA
磷酸肌醇	PRKCE；IRAK1；PRKAA2；EIF2AK2；PTEN；GRK6；
代謝	MAPK1；PLK1；AKT2；PIK3CA；CDK8；PIK3CB；PIK3C3；
	MAPK8；MAPK3；PRKCD；PRKAA1；MAPK9；CDK2；
	PIM1；PIK3C2A；DYRK1A；MAP2K2；PIP5K1A；PIK3R1；
	MAP2K1；PAK3；ATM；TTK；CSNK1A1；BRAF；SGK
PDGF傳訊	EIF2AK2；ELK1；ABL2；MAPK1；PIK3CA；FOS；PIK3CB；
	PIK3C3；MAPK8；CAV1；ABL1；MAPK3；KRAS；SRC；
	PIK3C2A；PDGFRB；RAF1；MAP2K2；JAK1；JAK2；
	PIK3R1；PDGFRA；STAT3；SPHK1；MAP2K1；MYC；
	JUN；CRKL；PRKCA；SRF；STAT1；SPHK2
VEGF傳訊	ACTN4；ROCK1；KDR；FLT1；ROCK2；MAPK1；PGF；
	AKT2；PIK3CA；ARNT；PTK2；BCL2；PIK3CB；PIK3C3；
	BCL2L1；MAPK3；KRAS；HIF1A；NOS3；PIK3C2A；PXN；
	RAF1；MAP2K2；ELAVL1；AKT1；PIK3R1；MAP2K1；SFN；
	VEGFA；AKT3；FOXO1；PRKCA
自然殺傷細胞傳訊	PRKCE；RAC1；PRKCZ；MAPK1；RAC2；PTPN11；
	KIR2DL3；AKT2；PIK3CA；SYK；PRKCI；PIK3CB；

	PIK3C3；PRKD1；MAPK3；KRAS；PRKCD；PTPN6；
	PIK3C2A；LCK；RAF1；FYN；MAP2K2；PAK4；AKT1；
	PIK3R1；MAP2K1；PAK3；AKT3；VAV3；PRKCA
細胞週期：G1/S	HDAC4；SMAD3；SUV39H1；HDAC5；CDKN1B；BTRC；
檢驗點調節	ATR；ABL1；E2F1；HDAC2；HDAC7A；RB1；HDAC11；
	HDAC9；CDK2；E2F2；HDAC3；TP53；CDKN1A；CCND1；
	E2F4；ATM；RBL2；SMAD4；CDKN2A；MYC；NRG1；
	GSK3B；RBL1；HDAC6
T細胞受體傳訊	RAC1；ELK1；MAPK1；IKBKB；CBL；PIK3CA；FOS；
	NFKB2；PIK3CB；PIK3C3；MAPK8；MAPK3；KRAS；
	RELA；PIK3C2A；BTK；LCK；RAF1；IKBKG；RELB；FYN；
	MAP2K2；PIK3R1；CHUK；MAP2K1；NFKB1；ITK；BCL10；
	JUN；VAV3
死亡受體傳訊	CRADD；HSPB1；BID；BIRC4；TBK1；IKBKB；FADD；
	FAS；NFKB2；BCL2；MAP3K14；MAPK8；RIPK1；CASP8；
	DAXX；TNFRSF10B；RELA；TRAF2；TNF；IKBKG；RELB；
	CASP9；CHUK；APAF1；NFKB1；CASP2；BIRC2；CASP3；
	BIRC3
FGF傳訊	RAC1；FGFR1；MET；MAPKAPK2；MAPK1；PTPN11；
	AKT2；PIK3CA；CREB1；PIK3CB；PIK3C3；MAPK8；
	MAPK3；MAPK13；PTPN6；PIK3C2A；MAPK14；RAF1；
	AKT1；PIK3R1；STAT3；MAP2K1；FGFR4；CRKL；ATF4；
	AKT3；PRKCA；HGF
GM-CSF傳訊	LYN；ELK1；MAPK1；PTPN11；AKT2；PIK3CA；CAMK2A；
	STAT5B；PIK3CB；PIK3C3；GNB2L1；BCL2L1；MAPK3；
	ETS1；KRAS；RUNX1；PIM1；PIK3C2A；RAF1；MAP2K2；
	AKT1；JAK2；PIK3R1；STAT3；MAP2K1；CCND1；AKT3；
	STAT1
肌萎縮側索	BID；IGF1；RAC1；BIRC4；PGF；CAPNS1；CAPN2；
硬化傳訊	PIK3CA；BCL2；PIK3CB；PIK3C3；BCL2L1；CAPN1；
	PIK3C2A；TP53；CASP9；PIK3R1；RAB5A；CASP1；
	APAF1；VEGFA；BIRC2；BAX；AKT3；CASP3；BIRC3
JAK/Stat傳訊	PTPN1；MAPK1；PTPN11；AKT2；PIK3CA；STAT5B；

	PIK3CB ;PIK3C3 ;MAPK3 ;KRAS ;SOCS1 ;STAT5A ;
	PTPN6 ;PIK3C2A ;RAF1 ;CDKN1A ;MAP2K2 ;JAK1 ;
	AKT1 ; JAK2 ; PIK3R1 ; STAT3 ; MAP2K1 ; FRAP1 ; AKT3 ;
	STAT1
煙酸鹽和煙醯胺	PRKCE ;IRAK1 ;PRKAA2 ;EIF2AK2 ;GRK6 ;MAPK1 ;
代謝	PLK1 ; AKT2 ; CDK8 ; MAPK8 ; MAPK3 ; PRKCD ; PRKAA1 ;
	PBEF1 ; MAPK9 ; CDK2 ; PIM1 ; DYRK1A ; MAP2K2 ;
	MAP2K1 ; PAK3 ; NT5E ; TTK ; CSNK1A1 ; BRAF ; SGK
趨化因子傳訊	CXCR4 ;ROCK2 ;MAPK1 ;PTK2 ;FOS ;CFL1 ;GNAQ ;
	CAMK2A ; CXCL12 ; MAPK8 ; MAPK3 ; KRAS ; MAPK13 ;
	RHOA ;CCR3 ;SRC ;PPP1CC ;MAPK14 ;NOX1 ;RAF1 ;
	MAP2K2 ; MAP2K1 ; JUN ; CCL2 ; PRKCA
IL-2傳訊	ELK1 ;MAPK1 ;PTPN11 ;AKT2 ;PIK3CA ;SYK ;FOS ;
	STAT5B ;PIK3CB ;PIK3C3 ;MAPK8 ;MAPK3 ;KRAS ;
	SOCS1 ; STAT5A ; PIK3C2A ; LCK ; RAF1 ; MAP2K2 ;
	JAK1 ; AKT1 ; PIK3R1 ; MAP2K1 ; JUN ; AKT3
突觸性長期	PRKCE ;IGF1 ;PRKCZ ;PRDX6 ;LYN ;MAPK1 ;GNAS ;
抑鬱	PRKCI ; GNAQ ; PPP2R1A ; IGF1R ; PRKD1 ; MAPK3 ;
	KRAS ; GRN ; PRKCD ; NOS3 ; NOS2A ; PPP2CA ;
	YWHAZ ;RAF1 ;MAP2K2 ;PPP2R5C ;MAP2K1 ;PRKCA
雌激素受體	TAF4B ; EP300 ; CARM1 ; PCAF ; MAPK1 ; NCOR2 ;
傳訊	SMARCA4 ; MAPK3 ; NRIP1 ; KRAS ; SRC ; NR3C1 ;
	HDAC3 ;PPARGC1A ;RBM9 ;NCOA3 ;RAF1 ;CREBBP ;
	MAP2K2 ; NCOA2 ; MAP2K1 ; PRKDC ; ESR1 ; ESR2
蛋白質泛素化	TRAF6 ;SMURF1 ;BIRC4 ;BRCA1 ;UCHL1 ;NEDD4 ;
途徑	CBL ;UBE2I ;BTRC ;HSPA5 ;USP7 ;USP10 ;FBXW7 ;
	USP9X ;STUB1 ;USP22 ;B2M ;BIRC2 ;PARK2 ;USP8 ;
	USP1 ; VHL ; HSP90AA1 ; BIRC3
IL-10傳訊	TRAF6 ; CCR1 ; ELK1 ; IKBKB ; SP1 ; FOS ; NFKB2 ;
	MAP3K14 ;MAPK8 ;MAPK13 ;RELA ;MAPK14 ;TNF ;
	IKBKG ; RELB ; MAP3K7 ; JAK1 ; CHUK ; STAT3 ; NFKB1 ;
	JUN ; IL1R1 ; IL6
VDR/RXR啟動	PRKCE ; EP300 ; PRKCZ ; RXRA ; GADD45A ; HES1 ;
	NCOR2 ; SP1 ; PRKCI ; CDKN1B ; PRKD1 ; PRKCD ;
	RUNX2 ; KLF4 ; YY1 ; NCOA3 ; CDKN1A ; NCOA2 ; SPP1 ;
	LRP5 ; CEBPB ; FOXO1 ; PRKCA
TGF-β傳訊	EP300 ;SMAD2 ;SMURF1 ;MAPK1 ;SMAD3 ;SMAD1 ;
	FOS ; MAPK8 ; MAPK3 ; KRAS ; MAPK9 ; RUNX2 ;
	SERPINE1 ; RAF1 ; MAP3K7 ; CREBBP ; MAP2K2 ;

	MAP2K1；TGFBR1；SMAD4；JUN；SMAD5
Toll樣受體傳訊	IRAK1；EIF2AK2；MYD88；TRAF6；PPARA；ELK1；
	IKBKB；FOS；NFKB2；MAP3K14；MAPK8；MAPK13；
	RELA；TLR4；MAPK14；IKBKG；RELB；MAP3K7；CHUK；
	NFKB1；TLR2；JUN
p38 MAPK傳訊	HSPB1；IRAK1；TRAF6；MAPKAPK2；ELK1；FADD；FAS；
	CREB1；DDIT3；RPS6KA4；DAXX；MAPK13；TRAF2；
	MAPK14；TNF；MAP3K7；TGFBR1；MYC；ATF4；IL1R1；
	SRF；STAT1
神經營養蛋白/TRK傳訊	NTRK2；MAPK1；PTPN11；PIK3CA；CREB1；FOS；
	PIK3CB；PIK3C3；MAPK8；MAPK3；KRAS；PIK3C2A；
	RAF1；MAP2K2；AKT1；PIK3R1；PDPK1；MAP2K1；
	CDC42；JUN；ATF4
FXR/RXR啟動	INS；PPARA；FASN；RXRA；AKT2；SDC1；MAPK8；
	APOB；MAPK10；PPARG；MTTP；MAPK9；PPARGC1A；
	TNF；CREBBP；AKT1；SREBF1；FGFR4；AKT3；FOXO1
突觸性長期	PRKCE；RAP1A；EP300；PRKCZ；MAPK1；CREB1；
增強	PRKCI；GNAQ；CAMK2A；PRKD1；MAPK3；KRAS；
	PRKCD；PPP1CC；RAF1；CREBBP；MAP2K2；MAP2K1；
	ATF4；PRKCA
鈣傳訊	RAP1A；EP300；HDAC4；MAPK1；HDAC5；CREB1；
	CAMK2A；MYH9；MAPK3；HDAC2；HDAC7A；HDAC11；
	HDAC9；HDAC3；CREBBP；CALR；CAMKK2；ATF4；
	HDAC6
EGF傳訊	ELK1；MAPK1；EGFR；PIK3CA；FOS；PIK3CB；PIK3C3；
	MAPK8；MAPK3；PIK3C2A；RAF1；JAK1；PIK3R1；
	STAT3；MAP2K1；JUN；PRKCA；SRF；STAT1
心血管系統中的低氧	EDN1；PTEN；EP300；NQO1；UBE2I；CREB1；ARNT；
傳訊	HIF1A；SLC2A4；NOS3；TP53；LDHA；AKT1；ATM；
	VEGFA；JUN；ATF4；VHL；HSP90AA1
RXR功能的LPS/IL-1介導	IRAK1；MYD88；TRAF6；PPARA；RXRA；ABCA1；
的抑制	MAPK8；ALDH1A1；GSTP1；MAPK9；ABCB1；TRAF2；
	TLR4；TNF；MAP3K7；NR1H2；SREBF1；JUN；IL1R1
LXR/RXR啟動	FASN；RXRA；NCOR2；ABCA1；NFKB2；IRF3；RELA；
	NOS2A；TLR4；TNF；RELB；LDLR；NR1H2；NFKB1；
	SREBF1；IL1R1；CCL2；IL6；MMP9

類澱粉蛋白加工	PRKCE ;CSNK1E ;MAPK1 ;CAPNS1 ;AKT2 ;CAPN2 ;
	CAPN1 ;MAPK3 ;MAPK13 ;MAPT ;MAPK14 ;AKT1 ;
	PSEN1 ; CSNK1A1 ; GSK3B ; AKT3 ; APP
IL-4傳訊	AKT2 ; PIK3CA ; PIK3CB ; PIK3C3 ; IRS1 ; KRAS ; SOCS1 ;
	PTPN6 ; NR3C1 ; PIK3C2A ; JAK1 ; AKT1 ; JAK2 ; PIK3R1 ;
	FRAP1 ; AKT3 ; RPS6KB1
細胞週期：G2/M DNA	EP300 ; PCAF ; BRCA1 ; GADD45A ; PLK1 ; BTRC ;
損害檢驗點	CHEK1 ; ATR ; CHEK2 ; YWHAZ ; TP53 ; CDKN1A ;
調節	PRKDC ; ATM ; SFN ; CDKN2A
心血管系統中的一氧化氮	KDR ;FLT1 ;PGF ;AKT2 ;PIK3CA ;PIK3CB ;PIK3C3 ;
傳訊	CAV1 ; PRKCD ; NOS3 ; PIK3C2A ; AKT1 ; PIK3R1 ;
	VEGFA ; AKT3 ; HSP90AA1
嘌呤代謝	NME2 ;SMARCA4 ;MYH9 ;RRM2 ;ADAR ;EIF2AK4 ;
	PKM2 ; ENTPD1 ; RAD51 ; RRM2B ; TJP2 ; RAD51C ;
	NT5E ; POLD1 ; NME1
cAMP-介導的傳訊	RAP1A ;MAPK1 ;GNAS ;CREB1 ;CAMK2A ;MAPK3 ;
	SRC ; RAF1 ; MAP2K2 ; STAT3 ; MAP2K1 ; BRAF ; ATF4
線粒體功能異常	SOD2 ;MAPK8 ;CASP8 ;MAPK10 ;MAPK9 ;CASP9 ;
	PARK7 ; PSEN1 ; PARK2 ; APP ; CASP3
Notch傳訊	HES1 ;JAG1 ;NUMB ;NOTCH4 ;ADAM17 ;NOTCH2 ;
	PSEN1 ; NOTCH3 ; NOTCH1 ; DLL4
內質網	HSPA5 ;MAPK8 ;XBP1 ;TRAF2 ;ATF6 ;CASP9 ;ATF4 ;
脅迫途徑	EIF2AK3 ; CASP3
嘧啶代謝	NME2 ;AICDA ;RRM2 ;EIF2AK4 ;ENTPD1 ;RRM2B ;
	NT5E ; POLD1 ; NME1
帕金森氏傳訊	UCHL1 ;MAPK8 ;MAPK13 ;MAPK14 ;CASP9 ;PARK7 ;
	PARK2 ; CASP3
心臟和β腎上腺素能傳訊	GNAS ;GNAQ ;PPP2R1A ;GNB2L1 ;PPP2CA ;PPP1CC ; PPP2R5C
酵解/葡糖異生	HK2 ; GCK ; GPI ; ALDH1A1 ; PKM2 ; LDHA ; HK1
干擾素傳訊	IRF1 ; SOCS1 ; JAK1 ; JAK2 ; IFITM1 ; STAT1 ; IFIT3
音蝟因子傳訊	ARRB2 ; SMO ; GLI2 ; DYRK1A ; GLI1 ; GSK3B ; DYRK1B
甘油磷脂	PLD1 ; GRN ; GPAM ; YWHAZ ; SPHK1 ; SPHK2
代謝	
磷脂降解	PRDX6 ; PLD1 ; GRN ; YWHAZ ; SPHK1 ; SPHK2
色胺酸代謝	SIAH2 ;PRMT5 ;NEDD4 ;ALDH1A1 ;CYP1B1 ;SIAH1
賴胺酸降解	SUV39H1 ; EHMT2 ; NSD1 ; SETD7 ; PPP2R5C
核苷酸切除修復	ERCC5 ; ERCC4 ; XPA ; XPC ; ERCC1
途徑	

澱粉和蔗糖	UCHL1；HK2；GCK；GPI；HK1
代謝	
胺基糖代謝	NQO1；HK2；GCK；HK1
花生四烯酸	PRDX6；GRN；YWHAZ；CYP1B1
代謝	
晝夜間律傳訊	CSNK1E；CREB1；ATF4；NR1D1
凝血系統	BDKRB1；F2R；SERPINE1；F3
多巴胺受體	PPP2R1A；PPP2CA；PPP1CC；PPP2R5C
傳訊	
麩胱甘肽代謝	IDH2；GSTP1；ANPEP；IDH1
甘油酯類代謝	ALDH1A1；GPAM；SPHK1；SPHK2
亞油酸代謝	PRDX6；GRN；YWHAZ；CYP1B1
甲硫胺酸代謝	DNMT1；DNMT3B；AHCY；DNMT3A
丙酮酸代謝	GLO1；ALDH1A1；PKM2；LDHA
精胺酸和脯胺酸	ALDH1A1；NOS3；NOS2A
代謝	
類二十烷酸傳訊	PRDX6；GRN；YWHAZ
果糖和甘露糖	HK2；GCK；HK1
代謝	
半乳糖代謝	HK2；GCK；HK1
芪、香豆素和	PRDX6；PRDX1；TYR
木質素生物合成	
抗原呈遞	CALR；B2M
途徑	
類固醇的生物合成	NQO1；DHCR7
丁酸酯代謝	ALDH1A1；NLGN1
檸檬酸循環	IDH2；IDH1
脂肪酸代謝	ALDH1A1；CYP1B1
甘油磷脂	PRDX6；CHKA
代謝	
組胺酸代謝	PRMT5；ALDH1A1
肌醇代謝	ERO1L；APEX1
藉由細胞色素p450進行的外源物	GSTP1；CYP1B1
代謝	
甲烷代謝	PRDX6；PRDX1
苯丙胺酸代謝	PRDX6；PRDX1
丙酸酯代謝	ALDH1A1；LDHA
硒胺基酸	PRMT5；AHCY
代謝	
鞘脂代謝	SPHK1；SPHK2
胺基磷糖	PRMT5
代謝	
雄激素和雌激素	PRMT5
代謝	

抗壞血酸酯和醛糖酸酯（Aldarate）	ALDH1A1
代謝	
膽汁酸生物合成	ALDH1A1
半胱胺酸代謝	LDHA
脂肪酸生物合成	FASN
穀胺酸受體	GNB2L1
傳訊	
NRF2-介導的氧化	PRDX1
脅迫應答	
戊糖磷酸酯	GPI
途徑	
戊糖和葡糖醛酸酯	UCHL1
互變	
視黃醇代謝	ALDH1A1
核黃素代謝	TYR
酪胺酸代謝	PRMT5、TYR
泛醌生物合成	PRMT5
纈胺酸、亮胺酸和	ALDH1A1
異亮胺酸降解	
甘胺酸、絲胺酸和	CHKA
蘇胺酸代謝	
賴胺酸降解	ALDH1A1
疼痛/味道	TRPM5；TRPA1
疼痛	TRPM7；TRPC5；TRPC6；TRPC1；Cnr1；cnr2；Grk2；
	Trpa1；Pomc；Cgrp；Crf；Pka；Era；Nr2b；TRPM5；Prkaca；
	Prkacb；Prkar1a；Prkar2a
線粒體功能	AIF；CytC；SMAC（Diablo）；Aifm-1；Aifm-2
發育神經學	BMP-4；脊索發生素（Chrd）；頭髮生素（Nog）；WNT（Wnt2；
	Wnt2b；Wnt3a；Wnt4；Wnt5a；Wnt6；Wnt7b；Wnt8b；
	Wnt9a；Wnt9b；Wnt10a；Wnt10b；Wnt16）；β-鏈蛋白
	Dkk-1；捲曲相關蛋白；Otx-2；Gbx2；FGF-8；
	顫蛋白；Dab1；unc-86（Pou4f1或Brn3a）；Numb；Reln

[1352] 本發明的實施方式還涉及牽涉敲除基因、擴增基因並修復與DNA重複序列不穩定性和神經病症相關聯的特定突變的方法和組成物（羅伯特·D.·威爾斯、鐵雄藎沢（Tetsuo Ashizawa），基因不穩定性和神經疾病（Genetic Instabilities and Neurological Diseases），第二版，學術出版社，2011年10月13日－ 醫學

（Medical））。已發現串聯重複序列的特定方面係超過二十種人類疾病的原因（重複序列不穩定性的新觀點：RNA•DNA雜交體的作用（New insights into repeat instability: role of RNA•DNA hybrids）。麥基弗·EI（McIvor EI）、波拉克·U（Polak U）、納皮爾拉拉·M（Napierala M.），RNA生物學（RNA Biol.），2010年9月-10月；7(5):551-8）。本發明效應蛋白系統可以用於校正基因組不穩定性的該等缺陷。

[1353] 本發明的幾個其他方面涉及校正與廣泛範圍的遺傳疾病相關聯的缺陷，該等疾病在國立衛生研究院（National Institutes of Health）網站上的主題小節遺傳病症下進一步描述（網址係在health.nih.gov/topic/GeneticDisorders）。遺傳性腦病可以包括但不限於，腎上腺腦白質營養不良、胼胝體發育不全、艾卡爾迪綜合症、阿爾珀斯病、阿耳茨海默病、巴斯綜合症、貝敦氏病、CADASIL、小腦變性、法布裡病、格-施-沙病（Gerstmann-Straussler-Scheinker Disease）、杭丁頓氏症、以及其他三聯體重複序列病症、利氏病、萊施-奈恩綜合症、門克斯病、線粒體肌病以及NINDS空洞腦。該等疾病在國立衛生研究院網站上的小節遺傳性腦部病症下進一步描述。

Cas9發展和使用

[1354] 本發明可以基於以下文獻中列出的CRISPR-Cas9發展和使用的方面來說明和擴展，並且特別是涉及細胞和生物體中的CRISPR蛋白複合物的遞送和RNA指導內切核酸酶的使用：

使用 CRISPR/Cas 系統的多重基因組工程化（Multiplex

genome engineering using CRISPR/Cas systems）。叢，L.（Cong, L.）、拉恩，F.A.（Ran, F.A.）、科克斯，D.（Cox, D.）、林，S.、巴雷德，R.（Barretto, R.）、哈比蔔，N.（Habib, N.）、徐，P.D.、吳，X.、蔣，W.、馬拉非尼，L.A.（Marraffini, L.A.）、以及張，F.，科學，2 月 15 日；339(6121):819-23 (2013)；

使用 CRISPR-Cas 對細菌基因組進行 RNA 指導編輯。蔣·W.、畢卡德·D（Bikard D.）、科克斯·D.、張·F、馬拉非尼·L.A. 自然生物技術，3 月；31(3):233-9 (2013)；

藉由 CRISPR/Cas 介導的基因組工程化來一步生成攜帶多基因中的突變的小鼠（One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering）。王·H.、楊·H.、希瓦里拉·CS.（Shivalila CS.）、道拉提·MM.（Dawlaty MM.）、成·AW.、張·F.、耶尼施·R（Jaenisch R.），細胞，5 月 9 日；153(4):910-8 (2013)；

哺乳動物內源性轉錄和外遺傳狀態的光控制（Optical control of mammalian endogenous transcription and epigenetic states）。康爾曼·S、布裡格姆·MD（Brigham MD）、特雷維諾·AE（Trevino AE）、徐·PD、海頓裡希·M（Heidenreich M）、叢·L、普萊特·RJ（Platt RJ）、斯科特·DA、丘奇·GM、張·F，自然 8 月 22 日；500(7463):472-6. doi: 10.1038/Nature12466. 電子版 2013 年 8 月 23 日(2013)；

用於增強基因組編輯特異性的 RNA 指導的 CRISPR Cas9 的雙重切口（Double Nicking by RNA-Guided CRISPR Cas9 for

Enhanced Genome Editing Specificity)。拉恩，FA.、徐，PD.、林，CY.、根特柏格，JS.（Gootenberg, JS.）、康爾曼，S.、特雷維諾，AE.、斯科特，DA.、井上，A.（Inoue, A.）、的場，S.（Matoba, S.）、張，Y.、以及張，F.，細胞，8 月 28 日. pii: S0092-8674(13)01015-5 (2013-A)；

靶向 RNA-指導的 Cas9 核酸酶的特異性的 DNA（DNA targeting specificity of RNA-guided Cas9 nucleases） 。徐，P.、斯科特，D.、溫斯坦，J.、拉恩，FA.、康爾曼，S.、瓦拉，V.、李，Y.、法恩，E.、吳，X.、謝萊姆，O.、科瑞迪克，TJ.、瑪律拉菲尼，LA.、包，G.、以及張，F.，自然生物技術 doi:10.1038/nbt.2647 (2013)；

使用 CRISPR-Cas9 系統的基因組工程化（Genome engineering using the CRISPR-Cas9 system)。拉恩，FA.、徐，PD.、賴特，J.、瓦拉，V.、斯科特，DA.、張，F.，自然實驗手冊，11 月；8(11):2281-308 (2013-B)；

人類細胞中的基因組規模的 CRISPR-Cas9 敲除篩選。沙萊姆，O.、珊亞納，NE.、哈特諾斯，E.、石，X.、斯科特，DA.、邁克爾森，T.、赫克爾，D.、埃伯特，BL.、羅特，DE.、多恩奇，JG.、張，F.，科學，12 月 12 日(2013)。[電子版先於印刷版]；

與指導 RNA 和靶向 RNA 複合的 cas9 的晶體結構（Crystal structure of cas9 in complex with guide RNA and target DNA)。西松，H.（ishimasu, H.）、拉恩，FA.、徐，PD.、康爾曼，S.、舍哈塔，SI.（Shehata, SI.）、多哈曼，N（Dohmae, N.）、石穀，R.（shitani, R.）、張，F.、Nureki，O.，細胞，2 月 27 日，156(5):935-49 (2014)；

哺乳動物細胞中 CRISPR 內切核酸酶 Cas9 的全基因組結合吳·X.、斯科特·DA.、克裡茨·AJ.（Kriz AJ.）、邱·AC.、徐·PD.、達頓·DB（Dadon DB.）、成·AW.、特雷維諾·AE.、康爾曼·S.、陳·S.、耶尼施·R.、張·F.、夏普·PA.自然生物技術 4 月 20 日 doi: 10.1038/nbt.2889 (2014)；

用於基因組編輯和癌症建模的 CRISPR-Cas9 敲入小鼠（CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling）。普萊特·RJ、陳·S、週·Y、嚴·MJ（Yim MJ）、斯維奇·L（Swiech L）、肯普頓·HR（Kempton HR）、達爾曼·JE（Dahlman JE）、帕納斯·O（Parnas O）、艾森哈爾·TM（Eisenhaure TM）、約瓦諾維奇·M（Jovanovic M）、格雷厄姆·DB（Graham DB）、卷卷瓦拉·S（Jhunjhunwala S）、海頓裡希·M、賽維爾·RJ（Xavier RJ）、朗格·R、安德森·DG、哈科恩·N（Hacohen N）、雷格夫·A（Regev A）、馮·G、夏普·PA、張·F，細胞 159(2): 440-455 DOI: 10.1016/j.cell.2014.09.014(2014)；

用基因組工程化的 CRISPR-Cas9 的發展和應用（Development and Applications of CRISPR-Cas9 for Genome Engineering），徐·PD、朗格·ES、張·F.，細胞 6 月 5 日；157(6):1262-78 (2014)。

使用 CRISPR/Cas9 系統的人類細胞的遺傳篩選（Genetic screens in human cells using the CRISPR/Cas9 system），王·T、魏·JJ、薩巴蒂尼·DM（Sabatini DM）、朗格·ES.，科學，1 月 3 日；343(6166): 80–84. doi:10.1126/science.1246981 (2014)；

用於 CRISPR-Cas9-介導的基因滅活的高活性 sgRNA 的合理設計（Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation），多恩奇·JG、哈特諾斯·E、格雷厄姆·DB、托特瓦·Z（Tothova Z）、赫格德·M（Hegde M）、史密斯·I、蘇倫德·M（Sullender M）、埃伯特·BL、賽維爾·RJ、羅特·DE.，（2014年9月3日線上公開）自然生物技術 12 月；32(12):1262-7 (2014)；

使用 CRISPR-Cas9 體內探察哺乳動物大腦的基因功能（*In vivo* interrogation of gene function in the mammalian brain using CRISPR-Cas9），斯維奇·L、海頓裡希·M、海頓裡希·A（Banerjee A）、哈比卜·N、李·Y、特龍貝塔·J（Trombetta J）、蘇爾·M（Sur M）、張·F.，（2014 年 10 月 19 日線上公開）自然生物技術 1 月；33(1):102-6 (2015)；

藉由工程化 CRISPR-Cas9 複合物進行基因組規模的轉錄啟動（Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex），康爾曼·S、布裡格姆·MD、特雷維諾·AE、薑俊·J、阿布德耶爾·OO（Abudayyeh OO）、巴爾塞納·C（Barcena C）、徐·PD、哈比蔔·N、根特柏格·JS、西松·H、Nureki O、張·F.，自然 1 月 29 日；517(7536):583-8 (2015)。

用於誘導型基因組編輯和轉錄調節的拆分 Cas9 體系結構（A split-Cas9 architecture for inducible genome editing and transcription modulation），蔡徹·B（Zetsche B）、沃爾茲·SE（Volz SE）、張·F.，（2015 年 2 月 02 日線上公開）自然生物技術 2 月；33(2):139-42 (2015)；

腫瘤生長和轉移的小鼠模型中的全基因組 CRISPR 篩選（Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis），陳·S、珊亞納·NE、鄭·K、沙萊姆·O、李·K、石·X、斯科特·DA、宋·J、潘·JQ、韋斯萊德爾·R（Weissleder R）、李·H、張·F、夏普·PA，細胞 160, 1246–1260，2015 年 3 月 12 日（小鼠中的多重篩選），以及

使用金黃色葡萄球菌 Cas9 的體內基因組編輯（*In vivo* genome editing using *Staphylococcus aureus* Cas9），拉恩·FA.、宋·L、嚴·WX、斯科特·DA, 根特柏格·JS、克裡茨·AJ、蔡徹·B、沙萊姆·O、吳·X、馬卡洛夫·KS、庫尼恩·EV、夏普·PA、張·F.，（2015 年 4 月 01 日線上公開），自然，4 月 9 日；520(7546):186-91 (2015)。

沙萊姆等人，“使用 CRISPR-Cas9 的高通量功能基因組學（High-throughput functional genomics using CRISPR-Cas9）”，遺傳學自然評論（Nature Reviews Genetics） 16, 299-311（2015 年 5 月）。

許等人，“改進的 CRISPR sgRNA 設計的序列決定簇（Sequence determinants of improved CRISPR sgRNA design）”，基因組研究 25, 1147-1157（2015 年 8 月）。

帕納斯等人，“在原代免疫細胞解剖調節網路的全基因組 CRISPR 篩選（A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks）”，細胞 162, 675-686（2015 年 7 月 30 日）。

拉曼那等人，“病毒 DNA 的 CRISPR/Cas9 分割有效抑制了乙型肝炎病毒（CRISPR/Cas9 cleavage of viral DNA efficiently suppresses hepatitis B virus）”，科技報告 5:10833. doi: 10.1038/srep10833（2015 年 6 月 2 日）

西松等人，“金黃色葡萄球菌 Cas9 的晶體結構（Crystal Structure of Staphylococcus aureus Cas9）”，細胞 162, 1113-1126（2015 年 8 月 27 日）

藉由 Cas9-誘導的原位飽和誘變進行 BCL11A 增強子分割，康維爾等人，自然 527(7577):192-7（2015 年 11 月 12 日）doi: 10.1038/nature15521.電子版 2015 年 9 月 16 日。

Cpf1 係第 2 類 CRISPR-Cas 系統的單一 RNA 指導的內切核酸酶（*Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System*），蔡徹等人，細胞 163, 759-71（2015 年 9 月 25 日）。

不同的第 2 類 CRISPR-Cas 系統的發現和功能表征（*Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems*），什馬科夫等人，分子細胞，60(3), 385–397 doi: 10.1016/j.molcel.2015.10.008 電子版 2015 年 10 月 22 日。

具有提高的特異性的合理工程化的 Cas9 核酸酶（*Rationally engineered Cas9 nucleases with improved specificity*），斯萊馬克爾等人，科學，2016 年 1 月 1 日 351(6268): 84-88 doi: 10.1126/science.aad5227. 電子版 2015 年 12 月 1 日。[電子版先於

印刷版]。

該等文獻各自藉由引用結合在此，它們可以被考慮用於實踐本發明，並且如以下簡要討論的：

叢等人基於嗜熱鏈球菌Cas9並且還基於釀膿鏈球菌Cas9來工程化用於真核細胞的II型CRISPR-Cas系統，並且證明Cas9核酸酶可以藉由短RNA來指導以誘導人類和小鼠細胞中的精確DNA切割。他們的研究顯示Cas9在轉化成切口酶後可以用於促進具有最小誘變活性的真核細胞中的同源定向修復。另外，他們的研究證實多個指導序列可以被編碼成單一CRISPR陣列，以使得能夠在哺乳動物基因組內的多個內源性基因組座位位點處同時進行編輯，這證明RNA指導的核酸酶技術的容易可程式設計性和廣泛適用性。使用RNA程式設計細胞內的序列特異性DNA切割的這種能力限定了一類新的基因組工程化工具。該等研究進一步顯示其他CRISPR作為可能可移植到哺乳動物細胞中並且也可以介導哺乳動物基因組切割。重要的是，可以設想CRISPR-Cas系統的幾個方面可以被進一步改進以增加其效率和多功能性。

蔣等人使用與雙RNA複合的成簇的、規律間隔的、短迴文重複序列（CRISPR）相關Cas9內切核酸酶，在肺炎鏈球菌和大腸桿菌的基因組中引入精確突變。該方法依賴於靶基因組位點處的雙-RNA：Cas9-引導的切割以殺死未突變細胞，並且不再需要選擇標記物或反選擇系統。該研究報導了藉由改變短CRISPR RNA（crRNA）的序列以形成編輯模板上攜帶的單個核苷酸或多個核苷酸來重新程式設計雙RNA：Cas9特異性。該研究顯示同時使用

兩個crRNA能夠進行多重誘變。另外，當該方法與重組組合使用時，在肺炎鏈球菌中幾乎100%使用所述方法恢復的細胞含有希望的突變，並且在大腸桿菌中65%恢復的細胞含有該突變。

王等人（2013）使用用於一步生成攜帶多基因中的突變的小鼠的CRISPR-Cas系統，該等突變通常是在多個步驟中藉由在胚胎幹細胞中連續重組和/或具有單個突變的小鼠的耗時互交來生成的。CRISPR-Cas系統將極大地加速功能冗余基因和上位基因相互作用的體內研究。

康爾曼等人（2013）解決了本領域中對於通用和穩健技術的需要，該等技術使得能夠光學和化學調節基於DNA結合結構域的CRISPR Cas9酶並且還調節轉錄啟動物樣效應物。

拉恩等人（2013-A）描述了將Cas9切口酶突變體與成對的指導RNA組合以引入靶向雙股斷裂之方法。這解決了來自微生物CRISPR-Cas系統的Cas9核酸酶藉由指導序列被靶向特異性基因組座位的問題，該等指導序列可以耐受與DNA靶標的某些錯配並且因此促成不希望的脫靶誘變。因為基因組中的個別切口以高保真度修復，因此經由適當偏移的指導RNA進行同時切口對於雙股斷裂係需要的並且擴大了用於靶向切割的特異性識別的鹼基數目。作者們證實使用成對切口可以在細胞系中減少50倍至1,500倍的脫靶活性並且促進小鼠受精卵中的基因敲除，而不用犧牲中靶切割效率。此通用策略使得能夠進行需要高特異性的各種各樣的基因組編輯應用。

徐等人（2013）表徵了人類細胞中的SpCas9靶向特異性，

以告知靶位點的選擇並避免脫靶效應。該研究評價了>700種指導RNA變體和293T和293FT細胞中>100個預測的基因組脫靶座位處的SpCas9誘導型indel突變水平。作者們指示SpCas9以序列依賴性方式容忍指導RNA與靶DNA之間的不同位置處的錯配，對錯配的數目、位置和分佈敏感。作者們進一步證實SpCas9-介導的切割未受到DNA甲基化的影響並且SpCas9和gRNA的劑量可以被滴定來最小化脫靶修飾。另外，為了促進哺乳動物基因組工程化應用，作者們的報導提供了一種指導靶序列的選擇和驗證以及脫靶分析的基於網路的軟體工具。

拉恩等人（2013-B）描述了一組用於在哺乳動物細胞中經由非同源末端接合（NHEJ）或同源定向修復（HDR）進行Cas9介導的基因組編輯以及生成用於下游功能研究的修飾的細胞系的工具。為了最小化脫靶切割，作者們進一步描述一種使用具有成對指導RNA的Cas9切口酶的雙切口策略。該等作者們提供的方案在實驗上推導了用於選擇靶位點、評價切割效率並分析脫靶活性的準則。該等研究顯示以靶向設計開始，基因修飾可以在僅僅1-2週內實現，並且修飾的選殖細胞系可以在2-3週內得到。

沙萊姆等人描述了探察全基因組規模的基因功能的新方式。他們的研究顯示具有64,751個獨特指導序列的基因組規模的CRISPR-Cas9敲除（GeCKO）文庫靶向的18,080個基因的遞送能夠在人類細胞中進行陰性和陽性選擇篩選。首先，作者們證實使用GeCKO文庫鑒定了癌症和多能幹細胞中的細胞活力所必須的基因。接著，在黑色素瘤模型中，作者們篩選其喪失涉及對維羅非

尼（抑制蛋白激酶BRAF的治療劑）的抗性的基因。他們的研究顯示最高評級的候選物包括先前驗證的基因NF1和MED12以及新型命中基因NF2、CUL3、TADA2B、以及TADA1。作者們在靶向相同基因的獨立的指導RNA與高命中確認率之間觀察到高水平的一致性，並且因此證實允許用Cas9進行基因組規模篩選。

西松等人報導了與sgRNA複合的釀膿鏈球菌Cas9的晶體結構以及其在2.5 A°解析度下的靶DNA。該結構揭示了由靶向識別和核酸酶裂片（lobe）組成的二裂片體系結構，在其介面處帶正電荷的溝中提供sgRNA：DNA異源雙股核酸分子。識別裂片對於結合sgRNA和DNA係必需的，而核酸酶裂片含有HNH和RuvC核酸酶結構域，該等結構域被適當地定位來分別切割靶DNA的互補股和非互補股。核酸酶裂片還含有負責與原型間隔區相鄰模體（PAM）相互作用的羧基末端結構域。此高分辨結構和伴隨的功能分析已揭示了藉由Cas9進行的RNA指導的DNA靶向的分子機制，從而為新的通用型基因組編輯技術的合理設計做準備。

吳等人由裝載有小鼠胚胎幹細胞（mESC）中的單一指導RNA（sgRNA）的釀膿鏈球菌繪製無催化活性Cas9（dCas9）的全基因組結合位點。作者們證實四種測試的sgRNA各自使得dCas9靶向幾十個與幾千個之間的基因組位點，頻繁地表徵為sgRNA和NGG原型間隔區相鄰模體（PAM）中的5-核苷酸種子區。染色質難接近性減少了dCas9與具有匹配的種子序列的其他位點的結合；因此70%脫靶位點與基因相關聯。作者們證實在用催化活性的Cas9轉染的mESC中295 dCas9結合位點的靶向定序鑒定了在背景水平

上突變的僅一個位點。作者們提出用於Cas9結合和切割的兩階段模型，其中種子匹配觸發結合但擴大的與靶DNA的配對對於切割是需要的。

普萊特等人建立了Cre-依賴性Cas9敲入小鼠。作者們證實了使用神經元、免疫細胞和內皮細胞中的指導RNA的腺相關病毒（AAV）、慢病毒或粒子介導的遞送進行的體內以及離體基因組編輯。

徐等人（2014）係大體上討論了CRISPR-Cas9從酸乳到基因組編輯的歷史（包括細胞遺傳篩選）的評論性文章。

王等人（2014）涉及適用於使用基因組規模的慢病毒單一指導RNA（sgRNA）文庫的陽性選擇和陰性選擇的合併的失功能遺傳篩選方法。

多恩奇等人創建了sgRNA庫，鋪在一組六種內源性小鼠基因和三種內源性人類基因的所有可能的靶位點上，並且藉由抗體染色和流式細胞術定量評定它們產生靶基因的無效對偶基因的能力。作者們證實PAM的優化提高活性並且也提高一組用於設計sgRNA的線上工具。

斯維奇等人證實AAV介導的SpCas9基因組編輯可以能夠進行大腦中的基因功能的反向遺傳學研究。

康爾曼等人（2015）討論了在使用或不使用接頭的情況下將多個效應物結構域例如轉錄啟動物、功能和表觀基因組調節物附接在指導序列諸如莖環或四核苷酸環上的適當位置處的能力。

蔡徹等人證實Cas9酶可以拆分成兩個並且因此可以控制用於啟動的Cas9的組裝。

陳等人涉及藉由證實小鼠全基因組體內CRISPR-Cas9篩選揭示了基因調節肺部轉移來進行多重篩選。

拉恩等人（2015）涉及SaCas9以及其編輯基因組的能力並且證實不能從生物化學測定外推。

沙萊姆等人（2015）描述其中無催化活性Cas9（dCas9）融合物用於在合成上阻遏（CRISPRi）或啟動（CRISPRa）表現，從而顯示使用基因組規模的篩選的Cas9的進展的方式，該等篩選包括排列和合併的篩選、滅活基因組座位的敲除方法以及調節轉錄活性的策略。

許等人（2015）評定了有助於基於CRISPR的篩選的單一指導RNA（sgRNA）效率的DNA序列特徵。作者們探尋了CRISPR/Cas9敲除的效率和切割位點處的核苷酸偏好。作者們還發現CRISPRi/a的序列偏好基本上不同於CRISPR/Cas9敲除的序列偏好。

帕納斯等人（2015）將全基因組合並的CRISPR-Cas9文庫引入到樹突細胞（DC）中，以鑒定控制細菌性脂多糖（LPS）對腫瘤壞死因子（Tnf）的誘導。鑒定Tlr4傳訊的已知調節劑和先前未知的候選物並且將其分成對於對LPS的正則反應具有不同作用的三種功能模組。

拉曼那等人（2015）證實了感染細胞中的病毒附加體DNA

（cccDNA）的切割。HBV基因組在感染的肝細胞的核中作為稱為共價閉環DNA（cccDNA）的3.2 kb雙股附加型DNA種類存在，該共價閉環DNA係在HBV生命週期中其複製不受當前治療的抑制的關鍵性組分。作者們證實特異性靶向HBV的高度保守區的sgRNA強烈抑制了病毒複製和缺失的cccDNA。

西松等人（2015）報導了與單一指導RNA（sgRNA）及其雙股DNA靶標複合的SaCas9的晶體結構，該單一指導RNA含有5'-TTGAAT-3' PAM和5'-TTGGGT-3' PAM。SaCas9與SpCas9的結構比較突出顯示了結構保守性和趨異性，這解釋了它們不同的PAM特異性和直源sgRNA識別。

康維爾等人（2015）證實基於CRISPR-Cas9的非編碼基因組元件的功能研究。作者們開發了進行人類和小鼠BCL11A增強子的原位飽和誘變的合併的CRISPR-Cas9指導RNA文庫，這揭示了該等增強子的關鍵性特徵。

蔡徹等人（2015）報導了來自新殺手法蘭西斯菌U112的、具有不同於Cas9的特徵的第2類CRISPR核酸酶Cpf1的表徵。Cpf1係一種缺乏tracrRNA的單一RNA指導的內切核酸酶，它利用T富集的原型間隔區相鄰模體，並且經由交錯雙股斷裂來切割DNA。

什馬科夫等人（2015）報導了三種不同的第2類CRISPR-Cas系統。兩種系統性CRISPR酶（C2c1和C2c3）含有與Cpf1相關性較遠的RuvC樣內切核酸酶結構域。不同於Cpf1，C2c1取決於用於DNA切割的crRNA和tracrRNA。第三種酶（C2c2）含有兩個預測的HEPN RNA酶結構域並且係tracrRNA獨立的。

斯萊馬克爾等人（2016）報導了使用結構指導的蛋白質工程化來提高釀膿鏈球菌Cas9（SpCas9）的特異性。作者們開發了維持強勁的中靶切割同時具有減小的脫靶效應的“特異性增強”的SpCas9（eSpCas9）變體。

[1355] 同樣地，“用於高特異性基因組編輯的二聚CRISPR RNA指導FokI核酸酶”，盛達爾·Q.·蔡、尼古拉斯·維肯恩、采德·凱特爾、詹尼弗·A.·福登布、維沙爾·撒帕爾、迪派克·雷恩、馬修·J.·古德溫、馬丁·J.·阿裡耶、J.·基斯·姜俊，自然生物技術 32(6): 569-77 (2014)，涉及在人類細胞中識別擴展序列並以高效率編輯內源性基因的二聚RNA指導FokI核酸酶。

[1356] 美國專利案號8,697,359、8,771,945、8,795,965、8,865,406、8,871,445、8,889,356、8,889,418、8,895,308、8,906,616、8,932,814、8,945,839、8,993,233以及8,999,641；美國專利公開2014-0310830（美國申請案序號14/105,031）、US 2014-0287938 A1（美國申請案序號14/213,991）、US 2014-0273234 A1（美國申請案序號14/293,674）、US2014-0273232 A1（美國申請案序號14/290,575）、US 2014-0273231（美國申請案序號14/259,420）、US 2014-0256046 A1（美國申請案序號14/226,274）、US 2014-0248702 A1（美國申請案序號14/258,458）、US 2014-0242700 A1（美國申請案序號14/222,930）、US 2014-0242699 A1（美國申請案序號14/183,512）、US 2014-0242664 A1（美國申請案序號14/104,990）、US 2014-0234972 A1（美國申請案序號14/183,471）、US 2014-0227787 A1（美國申請案序號14/256,912）、US 2014-0189896 A1（美國申請

案序號14/105,035）、US 2014-0186958（美國申請案序號14/105,017）、US 2014-0186919 A1（美國申請案序號14/104,977）、US 2014-0186843 A1（美國申請案序號14/104,900）、US 2014-0179770 A1（美國申請案序號14/104,837）以及US 2014-0179006 A1（美國申請案序號14/183,486）、US 2014-0170753（美國申請案序號14/183,429）;US 2015-0184139（美國申請案序號14/324,960）；14/054,414，歐洲專利申請EP 2 771 468（EP13818570.7）EP 2 764 103（EP13824232.6）、以及EP 2 784 162（EP14170383.5）；以及PCT專利公開WO 2014/093661（PCT/US2013/074743）、WO 2014/093694（PCT/US2013/074790）、WO 2014/093595（PCT/US2013/074611）、WO 2014/093718（PCT/US2013/074825）、WO 2014/093709（PCT/US2013/074812）、WO 2014/093622（PCT/US2013/074667）、WO 2014/093635（PCT/US2013/074691）、WO 2014/093655（PCT/US2013/074736）、WO 2014/093712（PCT/US2013/074819）、WO 2014/093701（PCT/US2013/074800）、WO 2014/018423（PCT/US2013/051418）、WO 2014/204723（PCT/US2014/041790）、WO 2014/204724（PCT/US2014/041800）、WO 2014/204725（PCT/US2014/041803）、WO 2014/204726（PCT/US2014/041804）、WO 2014/204727（PCT/US2014/041806）、WO 2014/204728（PCT/US2014/041808）、WO 2014/204729（PCT/US2014/041809）、WO 2015/089351（PCT/US2014/069897）、WO 2015/089354（PCT/US2014/069902）、WO 2015/089364（PCT/US2014/069925）、WO 2015/089427（PCT/US2014/070068）、WO 2015/089462（PCT/US2014/070127）、WO 2015/089419

（PCT/US2014/070057）、WO 2015/089465（PCT/US2014/070135）、WO 2015/089486（PCT/US2014/070175）、PCT/US2015/051691、PCT/US2015/051830。還參考美國臨時專利申請61/758,468；61/802,174；61/806,375；61/814,263；61/819,803和61/828,130，它們分別在2013年1月30日；2013年3月15日；2013年3月28日；2013年4月20日；2013年5月6日以及2013年5月28日提交。還參考在2013年6月17日提交的美國臨時專利申請61/836,123。還另外參考美國臨時專利申請61/835,931、61/835,936、61/835,973、61/836,080、61/836,101、以及61/836,127，該等專利各自在2013年6月17日提交。還參考2013年8月5日提交的美國臨時專利申請61/862,468和61/862,355；在2013年8月28日提交的61/871,301；在2013年9月25日提交的61/960,777以及2013年10月28日提交的61/961,980。進一步參考：在2014年10月28日提交的PCT/US2014/62558，以及美國臨時專利申請案序號：各自在2013年12月12日提交的61/915,148、61/915,150、61/915,153、61/915,203、61/915,251、61/915,301、61/915,267、61/915,260、and 61/915,397；在2013年1月29日和2013年2月25日提交的61/757,972和61/768,959；二者均在2014年6月11日提交的62/010,888和62/010,879；各自在2014年6月10日提交的62/010,329、62/010,439和62/010,441；各自在2014年2月12日提交的61/939,228和61/939,242；在2014年4月15日提交的61/980,012；在2014年8月17日提交的62/038,358；各自在2014年9月25日提交的62/055,484、62/055,460和62/055,487；以及在2014年10月27日提交的62/069,243。參考指定尤其是美國的2014年6月10日提交的申請案號PCT/US14/41806的PCT申請。參考2014年1月22日提交的美國

臨時專利申請61/930,214。參考指定尤其是美國的2014年6月10日提交的申請案號PCT/US14/41806的PCT申請。

[1357] 還參考2015年6月17日的美國申請62/180,709，保護性指導RNA（PGRNA）（PROTECTED GUIDE RNAS (PGRNAS)）；2014年12月12日提交的美國申請62/091,455，保護性指導RNA（PGRNA）；2014年12月24日提交的美國申請62/096,708，保護性指導RNA（PGRNA）；2014年12月12日的美國申請62/091,462、2014年12月23日的62/096,324、2015年6月17日的62/180,681、以及2015年10月5日的62/237,496，用於CRISPR轉錄因子的無效指導序列（DEAD GUIDES FOR CRISPR TRANSCRIPTION FACTORS）；2014年12月12日的美國申請62/091,456和2015年6月17日的62/180,692，用於CRISPR-CAS系統的護送指導序列和功能化指導序列（ESCORTED AND FUNCTIONALIZED GUIDES FOR CRISPR-CAS SYSTEMS）；2014年12月12日的美國申請62/091,461，用於關於造血幹細胞（HSC）的基因組編輯的CRISPR-CAS系統和組成物的遞送、使用和治療性應用（DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR GENOME EDITING AS TO HEMATOPOETIC STEM CELLS (HSCs)）；2014年12月19日的美國申請62/094,903，藉由基因組規格的插入物捕獲定序來無偏地鑒定雙股斷裂和基因組重新排列（UNBIASED IDENTIFICATION OF DOUBLE-STRAND BREAKS AND GENOMIC REARRANGEMENT BY GENOME-WISE INSERT CAPTURE SEQUENCING）；2014年12月24日的美國申請62/096,761，用於序列操縱的系統、方法和最

佳酶和指導支架的工程化（ENGINEERING OF SYSTEMS, METHODS AND OPTIMIZED ENZYME AND GUIDE SCAFFOLDS FOR SEQUENCE MANIPULATION）；2014年12月30日的美國申請62/098,059、2015年6月18日的62/181,641、以及2015年6月18日的62/181,667，RNA靶向系統（RNA-TARGETING SYSTEM）；2014年12月24日的美國申請62/096,656和2015年6月17日的62/181,151，具有不穩定結構域或與該結構域締合的CRISPR（CRISPR HAVING OR ASSOCIATED WITH DESTABILIZATION DOMAINS）；2014年12月24日的美國申請62/096,697，具有AAV或與該AAV締合的CRISPR（CRISPR HAVING OR ASSOCIATED WITH AAV）；2014年12月30日的美國申請62/098,158，工程化CRISPR複合物插入的靶向系統（ENGINEERED CRISPR COMPLEX INSERTIONAL TARGETING SYSTEMS）；2015年4月22日的美國申請62/151,052，用於細胞外外來體報導的細胞靶向（CELLULAR TARGETING FOR EXTRACELLULAR EXOSOMAL REPORTING）；2014年9月24日的美國申請62/054,490，用於使用粒子遞送組成物靶向病症和疾病的CRISPR-CAS系統和組成物的遞送、使用和治療性應用（DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING PARTICLE DELIVERY COMPONENTS）；2014年2月12日的美國申請61/939,154，用於使用最佳功能的CRISPR-CAS系統進行序列操縱的系統、方法和組成物（SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH

OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS）；2014年9月25日的美國申請62/055,484，用於使用最佳功能的CRISPR-CAS系統進行序列操縱的系統、方法和組成物；2014年12月4日的美國申請62/087,537，用於使用最佳功能的CRISPR-CAS系統進行序列操縱的系統、方法和組成物；2014年9月24日的美國申請62/054,651，用於體內調節多種癌症突變的競爭的CRISPR-CAS系統和組成物的遞送、使用和治療性應用（DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO）；2014年10月23日的美國申請62/067,886，用於體內調節多種癌症突變的競爭的CRISPR-CAS系統和組成物的遞送、使用和治療性應用；2014年9月24日的美國申請62/054,675和2015年6月17日的62/181,002，CRISPR-CAS系統和組成物在神經元細胞/組織中的遞送、使用和治療性應用（DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN NEURONAL CELLS/TISSUES）；2014年9月24日的美國申請62/054,528，CRISPR-CAS系統和組成物在免疫疾病或病症中的遞送、使用和治療性應用（ELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN IMMUNE DISEASES OR DISORDERS）；2014年9月25日的美國申請62/055,454，用於使用細胞穿透肽（CPP）靶向病症和疾病的CRISPR-CAS系統和組成物的遞送、使用和治療性應用（DELIVERY, USE AND THERAPEUTIC APPLICATIONS

OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING CELL PENETRATION PEPTIDES (CPP)）；2014年9月25日的美國申請62/055,460，多功能-CRISPR複合物和/或最佳酶連接的功能性-CRISPR複合物（MULTIFUNCTIONAL-CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES）；2014年12月4日的美國申請62/087,475和2015年6月18日的62/181,690，使用最佳CRISPR-CAS系統的功能性篩選（FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS）；2014年9月25日的美國申請62/055,487，使用最佳CRISPR-CAS系統的功能性篩選；2014年12月4日的美國申請62/087,546和2015年6月18日的62/181,687，多功能-CRISPR複合物和/或最佳酶連接的功能性-CRISPR複合物；以及2014年12月30日的美國申請62/098,285，腫瘤生長和轉移的CRISPR介導的體內調節和遺傳篩選（CRISPR MEDIATED IN VIVO MODELING AND GENETIC SCREENING OF TUMOR GROWTH AND METASTASIS）。

[1358] 還參考2015年6月18日的美國申請62/181,659和2015年8月19日的62/207,318，用於序列操縱的CAS9異種同源物和變體的系統、方法、酶以及指導支架的工程化和優化（ENGINEERING AND OPTIMIZATION OF SYSTEMS, METHODS, ENZYME AND GUIDE SCAFFOLDS OF CAS9 ORTHOLOGS AND VARIANTS FOR SEQUENCE MANIPULATION）。參考2015年6月18日的美國申請62/181,663和2015年10月22日的62/245,264，新型CRISPR酶以

及系統（NOVEL CRISPR ENZYMES AND SYSTEMS）；2015年6月18日的美國申請62/181,675、2015年10月22日的62/285,349、2016年2月17日的62/296,522、以及2016年4月8日的62/320,231，新型CRISPR酶以及系統；2015年9月24日的美國申請62/232,067、2015年12月18日的美國申請14/975,085、歐洲申請案號16150428.7、2015年8月16日的美國申請62/205,733、2015年8月5日的美國申請62/201,542、2015年7月16日的美國申請62/193,507、以及2015年6月18日的美國申請62/181,739，該等專利各自題為新型CRISPR酶以及系統；以及2015年10月22日的美國申請62/245,270，新型CRISPR酶以及系統。還參考2014年2月12日的美國申請61/939,256和2014年12月12日的WO 2015/089473（PCT/US2014/070152），該等專利各自題為具有用於序列操縱的新體系結構的系統、方法和最佳指導組成物的工程化（ENGINEERING OF SYSTEMS, METHODS AND OPTIMIZED GUIDE COMPOSITIONS WITH NEW ARCHITECTURES FOR SEQUENCE MANIPULATION）。還參考2015年8月15日的PCT/US2015/045504、2015年6月17日的美國申請62/180,699以及2014年8月17日的美國申請62/038,358，該等專利各自題為使用CAS9切口酶的基因組編輯（GENOME EDITING USING CAS9 NICKASES）。

[1359] 每一份該等專利、專利公開和申請、以及其中或它們的訴訟期間（“應用引用文獻”）所引用的所有文獻或在此應用引用文獻中引用或參考的所有文獻，連同在此或在藉由引用結合在此的任何文獻中提到的任何產品的任何說明書、描述、產品規格和產品清單均藉由引用結合在此，並且可以應用於本發明的實踐

中。所有文獻（例如，該等專利、專利公開和申請以及應用引用文獻）藉由引用結合在此，其程度如同將每個單獨的文獻具體並單獨地指明藉由引用結合在此。

[1360] 本發明的有效性已經得到證實。可以例如藉由電穿孔轉染包含Cpf1和crRNA的預組裝的重組體CRISPR-Cpf1複合物，從而產生高突變率且不存在可檢測的脫靶突變。戶珥，J.K.等人，藉由電穿孔Cpf1核糖核蛋白在小鼠中進行的靶向誘變，自然生物技術2016年6月6日 doi: 10.1038/nbt.3596。[電子版先於印刷版]。全基因組分析顯示Cpf1係高度特異的。藉由一個測量，對於人類HEK293T細胞中的SpCas9確定的體外切割位點顯著少於對於SpCas9的體外切割位點。基姆，D等人，全基因組分析揭示了人類細胞中的Cpf1內切核酸酶的特異性（Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells），自然生物技術，2016年6月6日 doi: 10.1038/nbt.3609。[電子版先於印刷版]。在採用由含有發明的tRNA的陣列加工的gRNA的果蠅中證實了一種採用Cpf1的有效複用系統。波特，F.（Port, F.）等人，在具有tRNA-側接的Cas9和Cpf1 gRNA的動物中的CRISPR工具箱的擴增（Expansion of the CRISPR toolbox in an animal with tRNA-flanked Cas9 and Cpf1 gRNAs）。doi: http://dx.doi.org/10.1101/046417。

[1361] 本發明將在以下實例中進一步說明，該等實例僅出於說明目的給出並且不旨在以任何方式限制本發明。

實例

實例1：適應性免疫系統的起源和進化

[1362] CRISPR-Cas系統在古細菌和細菌基因組中的分類和注解。CRISPR-Cas座位具有超過50種的基因家族並且不存在嚴格的通用基因，這表明了座位體系結構的快速進化、極端多樣性。因此，單一進化樹係不可行的並且需要多分支方法。到目前為止，針對93種Cas蛋白存在395種表現譜的全面*cas*基因鑒定。分類包括特徵基因表現譜加上座位體系結構的特徵。

[1363] 新的CRISPR-Cas系統分類提出在圖1中。第1類包括多亞基crRNA-效應物複合物（Cascade）並且第2類包括單亞基crRNA-效應物複合物（Cas9樣）。圖2提供了CRISPR-Cas的分子組構。圖3提供了I型和III型效應物複合物的結構：共同的體系結構/共同的祖先，儘管存在廣泛的序列趨異性。圖4示出了作為以RNA識別模體（RRM）為中心的系統的CRISPR-Cas。圖5示出了Cas1系統發育，其中自我調整模組和crRNA-效應物模組的重組顯示出CRISPR-Cas進化的主要方面。圖F示出了CRISPR-Cas種群調查，具體地是CRISPR-Cas型/亞型在古生菌和細菌之中的分佈。

[1364] Cas1並不總是連接至CRISPR-Cas系統，因此可能存在“單獨”Cas1的兩條分支，這說明功能和起源與可能的新型可動遺傳因子可能存在差異（參見馬卡洛夫、格魯帕維克（Krupovic）、庫尼恩，遺傳學前沿（Frontiers Genet） 2014）。三個casposon家族的基因組組構可以提供一些線索。除Cas1和PolB之外，casposon結合不同基因，包括各種核酸酶（格魯帕維克等人，BMC生物學（BMC Biology）2014）。一個家族具有蛋白質引發的聚合酶，另一個家族具有RNA引發的聚合酶。除不同廣古菌和奇古菌之外，

casposon可見於表明橫向流動的幾種細菌中。Casposon Cas1（易位酶/整合酶）顯示Cas1種系發生中的basal進化枝。

[1365] 細菌和古細菌利用CRISPR在原核生物和真核生物中經由基因組操縱來進行適應性免疫。Cas 1提供了用於基因組操縱的現成工具。casposon和CRISPR中存在類似整合機制，特別是藉由複製/黏貼不能切割複製部分來進行複製依賴性獲取（格魯帕維克等人，BMC生物學 2014）。Cas1係真實的整合酶（努涅斯·JK（Nuñez JK）、李·AS（Lee AS）、昂熱爾曼·A（Engelman A）、竇得那·JA（Doudna JA），在CRISPR-Cas適應性免疫過程中的整合酶介導的間隔區獲取（Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity），自然，2015年2月18日）。在casposon與CRISPRT的末端反向重複序列之間存在相似性（格魯帕維克等人，BMC生物學 2014）。CRISPR-Cas可以來源於casposon和先天性免疫座位（庫尼恩、格魯帕維克，遺傳學自然評論，2015）。原核生物和動物中的適應性免疫系統的進化可以是已沿著與先天性免疫座位處的轉位子整合平行的過程進行（庫尼恩、格魯帕維克，遺傳學自然評論，2015）。RAG1易位酶（脊椎動物中的V(D)J重組的關鍵酶）可以來源於Transib轉位子（卡皮托諾夫·VV（Kapitonov VV）、朱卡·J（Jurka J），RAG1核心和V(D)J重組信號序列時來源於Transib轉位子（RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons），公共科學圖書館生物學（PLoS Biol.）2005年6月；3(6):e181），然而，沒有Transib編碼RAG2。RAG1和RAG2編碼的轉位子描述於卡皮托諾夫、庫尼恩，生物學指導（Biol Direct） 2015並且*Transib*易位酶

系統發育被給出在卡皮托諾夫、庫尼恩，生物學指導 2015中。纖毛蟲中的防禦性DNA消除從PiggyMAc轉位子和先天性免疫系統RNAi演變而來（斯瓦特·EC（Swart EC）、諾瓦茨基·M（Nowacki M），藉由sRNA-靶向的DNA缺失來抵禦並編輯基因組的真核生物方式（The eukaryotic way to defend and edit genomes by sRNA-targeted DNA deletion），紐約科學院年鑒2015）。

[1366] 分類的相對穩定性說明最流行的CRISPR-Cas系統變體係已經知道的。然而，少數目前不可分類的變體的存在說明另外類型和亞型仍有待表徵（馬卡洛夫等人2015。CRISPR-Cas系統和cas基因的進化性分類（Evolutionary classification of CRISPR-Cas systems and cas genes））。

[1367] 轉位子對於適應性免疫和涉及DNA操縱的其他系統的進化起到關鍵性作用。第1類CRISPR-Cas來源於轉位子但僅用於自我調整模組。第2類CRISPR-Cas具有自我調整功能和效應物功能二者，其中模組可以從不同轉位子進化。

實例2：新預測的第2類CRISPR-Cas系統以及其來自可易位件的獨立起源的證據

[1368] 細菌和古細菌的適應性免疫的CRISPR-Cas系統顯示出蛋白質組成和基因組座位體系結構的極端多樣性。該等系統寬泛地分成兩類，具有多亞單元效應物複合物的第1類和具有單亞單元效應物模組的第2類，藉由Cas9蛋白來舉例說明。申請人開發了用於預測推定的新第2類CRISPR-Cas系統的簡單計算流程。使用此流程分析完整細菌基因組的資料庫使得能夠鑒定兩種新的變體，

該等變體各自在不同細菌中表現出來並且含有*cas1*和*cas2*基因連同編碼預測作為效應物模組起作用的大蛋白質的第三基因。在該等座位的第一個座位中，推定的效應蛋白（C2c1p）含有RuvC-樣核酸酶結構域並且類似於先前描述的Cpf1蛋白，即V型CRISPR-Cas系統的預測效應物；因此，新推定系統被分類為V-B亞型。蛋白質序列的深入比較表明含RuvC的效應蛋白Cas9、Cpf1和C2C1p獨立地從不同組的轉位子編碼的TnpB蛋白質進化。第二組新推定的CRISPR-Cas座位涵蓋了含有具有預測的RNA酶活性的兩個高度趨異性HEPN結構域的大蛋白質。考慮到預測的效應蛋白的新穎性，該等座位被分類為可能靶向mRNA的新VI型CRISPR-Cas。總之，此分析的結果顯示第2類CRISPR-Cas系統在多種獨立的情景上藉由將不同Cas1-Cas2-編碼的自我調整模組與來源於不同可動遺傳因子的效應蛋白組合來進化。此進化路徑最可能產生仍有待發現的第2類系統的多種變體。

[1369] CRISPR-Cas適應性免疫系統存在於~45%細菌和~90%古細菌基因組中並且顯示Cas蛋白質組成和序列以及基因組座位體系結構的極端多樣性。基於其crRNA-效應物複合物的結構組織化，該等系統被分類兩類，即具有多亞基效應物複合物的第1類和具有單一亞基效應物複合物的第2類（馬卡洛夫，2015）。第1類系統係更常見的並且不同於第2類系統。第1類目前藉由由許多古細菌和細菌基因組編碼的12種不同亞型表示，而第2類系統包括II型系統和推定的V型系統的三種亞型，該等亞型總體可見於約10%的定序細菌基因組（具有涵蓋推定型系統的單一古細菌基因組）。第2類系統典型地僅含有*cas*操縱子中的三個或四個基因，即

涉及自我調整而不涉及干擾的*cas1-cas2*基因對、負責干擾而且有助於前crRNA加工和自我調整的單一多結構域效應蛋白、以及常常是具有在至少一些II型系統中可分配的未表徵功能的第四種基因。在大多數情況下，CRISPR陣列和稱為tracrRNA（反式編碼的小CRISPR RNA）的不同RNA種類的基因與第2類cas操縱子相鄰（吉林斯基（Chylinski），2014）。tracrRNA與對應CRISPR陣列內的重複序列部分同源並且係為前crRNA的加工所必需的，該加工不與CRISPR-Cas座位相關聯的普遍存在的細菌酶RNA酶III催化（德爾特切瓦（Deltcheva），2011）（吉林斯基，2014；吉林斯基，2013)。

[1370] II型多結構域效應蛋白Cas9已在功能和結構方面以精巧詳情表徵。在不同細菌中，Cas9蛋白涵蓋約950個與1,400個之間的胺基酸並且含有兩個核酸酶結構域，即RuvC-樣（RNA酶H折疊）和HNH（McrA-樣）核酸酶（馬卡洛夫，2011）。Cas9的晶體結構揭示了具有不同靶向識別和核酸酶裂片的蛋白質的二裂片組構，其中後者具有RuvC和HNH結構域（西松，2014）（金科（Jinek），2014）。每一個Cas9核酸酶結構域對於一條靶DNA股的切割係需要的（金科，2012；撒普拉那斯卡（Sapranauskas），2011）。最近，Cas9已顯示有助於CRISPR反應的所有三個階段，它不僅靶向DNA切割（干擾）而且靶向自我調整和前-crRNA加工（金科，2012）。更確切地說，Cas9的核酸酶裂片中的不同結構域已顯示在自我調整階段過程中識別並結合病毒DNA中的原型間隔區相關模體（PAM）（西松，2014）（金科，2014）（黑勒爾（Heler），2015；魏，2015）。在CRISPR反應的此階段，Cas9與涉及所有CRISPR-Cas

系統中的間隔區獲取的兩種蛋白質Cas1和Cas2形成複合物（黑勒爾，2015；魏，2015）。

[1371] 與tracrRNA組合的Cas9蛋白最近已成為用於新生成的基因編輯和工程化方法的關鍵性工具（加西烏納斯，2013；馬里，2013；桑普森，2014；叢，2015）。Cas9在基因組編輯中的此效用取決於以下事實，在II型CRISPR-Cas系統中，與其他類型的CRISPR-Cas系統不同，對於靶DNA識別和切割所需要的所有活性在單一、儘管較大的多結構域蛋白質中組裝。II型系統的此特徵極大地促進用於基因組操縱的有效工具的設計。重要的是，並非所有的Cas9變體都係相等的。迄今為止大部分工作已使用來自釀膿鏈球菌的Cas9來完成，但是其他Cas9種類可以提供實質性優點。作為適當的例子，使用來自金黃色葡萄球菌的、比化膿鏈球菌蛋白質短約300個胺基酸的Cas9的最近實驗已允許將Cas9包裝到腺相關病毒載體中，從而引起用於體內基因組編輯的CRISPR-Cas效用的主要增強（拉恩，2015）。

[1372] II型CRISPR-Cas系統目前被分成3種亞型（II-A、II-B和II-C）（馬卡洛夫，2011）（方法拉（Fonfara），2014；吉林斯基，2013；吉林斯基，2014）。除所有II型座位所共用的cas1、cas2和cas9基因之外，II-A亞型的特徵係編碼滅活ATP酶的額外基因*csn2*（納姆（Nam），2011；古（Koo），2012；李，2012），該ATP酶在間隔區獲取中起到很少的表徵作用（巴蘭岡（Barrangou），2007；阿爾斯蘭（Arslan），2013）（黑勒爾，2015）。II-B亞型系統缺乏*csn2*但是反而含有另外具有I型系統典型特徵並且編碼recB家族5'-3'外

切核酸酶的*cas4*基因，該外切核酸酶有助於藉由生成重組基因座（recombinogeneci ）DNA末端進行間隔區獲取（張，2012）（萊馬克（Lemak），2013；萊馬克，2014）。II-B亞型的*cas1*和*cas2*基因與提示此II型亞型的重組起源的I型CRISPR-Cas系統的對應蛋白質最密切相關（吉林斯基，2014）。

[1373] II-C亞型CRISPR-Cas系統係僅由*cas1*、*cas2*和*cas9*基因組成的最少種類（吉林斯基，2013；庫尼恩，2013；吉林斯基，2014）。然而，值得注意地是，已顯示在空腸彎麴菌中，藉由II-C系統進行間隔區獲取需要由噬菌體編碼的Cas4的參與（胡頓（Hooton），2014）。II-C亞型的另一種不同特徵係藉由轉錄形成一些crRNA，該轉錄涉及由內部替代性啟動子進行的轉錄，與所有其他實驗表徵的CRISPR-Cas系統中觀察到的加工相反（張，2013）。

[1374] 最近，V型CRISPR-Cas系統的存在已藉由細菌基因組的比較性分析來預測到。該等推定的新型CRISPR-Cas系統在幾種細菌基因組中表現出來，特別是來自弗朗西絲菌屬和古細菌*alvus*甲烷嗜甲基菌（*Methanomethylophilus alvus*）的那些細菌基因組（韋斯特高（Vestergaard），2014）。所有推定的V型座位涵蓋cas1、cas2、表示為*cpf1*的不同基因以及CRISPR陣列（舒德爾（Schunder），2013）（馬卡洛夫，2015）。Cpf1係一種大蛋白質（約1300個胺基酸），它含有與Cas9的相應結構域同源的RuvC樣核酸酶結構域以及Cas9的特徵性富精胺酸簇的對應物。然而，Cpf1缺乏所有Cas9蛋白中存在的HNH核酸酶結構域，並且RuvC樣結構域在Cpf1序列中是連續的，這與其中含有包含HNH結構域的長插入物的Cas9相反（吉

林斯基，2014；馬卡洛夫，2015）。在Cas9和Cpf1的結構域體系結構中的該等主要差異表明含有Cpf1的系統應被分成一個新型。推定的V型系統的組成提示Cpf1係單一亞基效應物複合物並且因此該等系統被分配給第2類CRISPR-Cas。一些推定的V型座位編碼Cas4並且因此類似於II-B亞型座位，而其他座位缺乏Cas4並且因此類似於II-C亞型。

[1375] 已顯示Cas9和Cpf1蛋白質的最近同源物係在IS605家族轉位子中編碼的並且含有RuvC-樣核酸酶結構域以及在Cpf1中具有對應物的Zn-指的TnpB蛋白質。此外，已鑒定含有插入到RuvC-樣結構域的HNH結構域並且顯示與Cas9的高序列相似性的TnpB同源物。TnpB在轉位子中的作用仍不確定，因為已顯示此蛋白質並不是易位所需要的。

[1376] 考慮到Cas9和Cpf1與轉位子編碼的蛋白質的同源性，申請人假設第2類CRISPR-Cas系統已在多種情景中因為轉位子與cas1-cas2座位之間的重組而進化。因此，申請人設計了一種鑒定可以是第2類新型變體的候選物的基因組座位的簡單計算策略。在此申請人描述了此方法用於鑒定此類候選物的兩個組別的第一種應用，這兩個組別之一似乎係V型的不同亞型而第二個組別似乎具有VI型的品質。第2類CRISPR-Cas系統的新變體作為用於基因組編輯和表現調節的潛在工具受到明顯的關注。

[1377] 用於檢測候選的新型第2類CRISPR-Cas座位的資料庫搜索策略。申請人實施了鑒定候選的新型第2類CRISPR-Cas系統的直接計算方法（圖7.流程）。因為絕大多數CRISPR-Cas座位涵

蓋*cas1*基因（馬卡洛夫，2011；馬卡洛夫，2015）並且Cas1序列係所有Cas蛋白質中最高度保守的序列（健內，2012），申請人推斷cas1係使用利用Cas1譜圖進行的翻譯PSI-BLAST搜索鑒定候選的新座位的最可能的錨點。在檢測編碼Cas1的所有疊連群之後，使用GenemarkS預測*cas1*基因上游和下游的20 KB區域內的蛋白質編碼基因。使用NCBI CDD和Cas蛋白特異性譜圖標注該等預測的基因，並且使用PILER-CR程序預測CRISPR陣列。此程式提供了檢測的CRISPR-Cas座位到已知亞基的分配。選擇含有大（>500 aa）蛋白質的未分類的候選CRISPR-Cas座位作為新型第2類系統的候選物，假定此類蛋白質特徵性存在於在II型和V型中（分別是Cas9和Cpf1）。根據具體情況使用PSI-BLAST和HHpred分析用此標準檢測的所有34個候選座位。在候選座位中編碼的蛋白質序列進一步用作搜索另外的同源物的巨集基因組學資料庫的查詢，並且如上所指示地分析該等搜索中檢測的長疊連群。此分析流程產生牢固連接至CRISPR-Cas系統的兩組座位。

[1378] 推定的V-B型系統。第一組候選座位暫時指示命名為C2c1（第2類候選物1），它在來自四種主要門，包括桿菌門、疣微菌門、α變形菌門以及δ變形菌門的細菌基因組中表現出來（圖8“第2類系統的完整座位的組構”）。所有C2c1座位編碼了Cas1-Cas4融合物、Cas2以及申請人表示為C2c1p的大蛋白質，並且典型地與CRISPR陣列相鄰（圖9，C2c1鄰近群體）。在Cas1的系統發育樹中，對應Cas1蛋白與I-U型系統（圖10，Cas1樹）成簇，該系統係其中可見Cas1-Cas4融合物的唯一系統。C2c1p蛋白由大約1200個胺基酸組成，並且HHpred搜索檢測了此蛋白質的C末端部分與IS605家

族的轉位子中編碼的TnpB蛋白子集之間的顯著相似性。相反，在和其他組的TnpB蛋白質類似的C2c1p與Cas9或Cpf1之間未檢測到顯著相似性(吉林斯基，2014)(馬卡洛夫，2015；馬卡洛夫，2015)。因此，C2c1p的結構域體系結構類似於Cpf1的結構域體系結構並且不同於Cas9的結構域體系結構，儘管所有三種Cas蛋白似乎都係從TnpB家族進化（圖11“第2類家族的結構域組構”）。C2c1p的N末端區域未顯示與其他蛋白質的顯著相似性。二級結構預測指示此區域主要採用α螺旋構象。與TnpB具有相似性的兩個區段涵蓋具有D..E..D特徵的RuvC樣核酸酶的三個催化模體（圖12“第2類蛋白質中的TnpB同源區”）；與Cas9蛋白中的橋螺旋（也稱為富精胺酸簇）相對應的區域涉及crRNA結合；並且似乎係TnpB的Zn指的對應物的小區域（然而，在C2C1p中替換Zn結合半胱胺酸殘基，這指示此蛋白質並不結合鋅）。C2c1p與Cpf1的結構域體系結構的相似性表明C2c1座位最好分類成V-B亞型，在此情況下編碼Cpf1的座位變成V-A亞型。

[1379] 儘管與此系統相關聯的*cas1*基因具有相似性，但是在對應陣列中的CRISPR重複序列係高度異質的，雖然所有該等重複序列都係36-37 bp長並且可以被分類成未結構化的（折疊能量ΔG係-0.5-4.5 kcal/mol，而高迴文CRISPR具有低於-7的ΔG）。根據CRISPRmap（蘭格（Lange），2013）分類方案，幾種V-B亞型重複序列與II型重複序列共用一些序列相似性或結構相似性。

[1380] 考慮到推定的V-B亞型CRISPR-Cas系統在機制上類似於II型系統的可能性，申請人試圖鑒定對應基因組座位中的

tracrRNA。

[1381] 來自V-B型CRISPR陣列的間隔區與非冗餘核苷酸序列資料庫的比較鑒定了與不同細菌基因組的幾種匹配。考慮到對於具有推定的V-B型CRISPR-Cas系統的細菌而言沒有噬菌體係已知的，該等匹配的相關性難以評定。

[1382] 推定的VI型系統。第二組候選CRISPR-Cas座位表示為C2c2，係在來自5種主要細菌門α變形菌門、桿菌門、梭菌綱、梭桿菌門以及擬桿菌門的基因組中鑒定的（圖8"第2類系統的完整座位的組構"）。與c2c1類似的是，C2c2座位涵蓋*cas1*和*cas2*基因連同大蛋白質（C2c2p）和CRISPR陣列；然而，與C2c1不同的是，C2c2p常常緊挨著CRISPR陣列而不是cas1-cas2來編碼（圖13，C2c2鄰近群體）。在Cas1系統發育樹中，來自C2c2座位的Cas1蛋白分佈在兩種進化枝中。第一進化枝包括來自梭菌綱的Cas1並且連同小III-A型分支位於II型子樹內（圖10，Cas1樹）。第二進化枝由來自纖毛菌屬的C2c2座位的Cas1蛋白組成並且安置在主要含有來自III-A型CRISPR-Cas系統的Cas1蛋白的混合型分支內。使用HHpred和PSI-BLAST的資料庫搜索未在C2c2p與其他蛋白質之間檢測到序列相似性。然而，檢查C2c2p蛋白序列的多重比對能夠鑒定具有HEPN結構域特徵的兩種嚴格保守的RxxxxH模體（阿南他拉曼（Anantharaman），2013）。二級結構預測指示該等模體位於與HEPN結構域結構相容的結構環境內，如與C2c2p的對應部分的總體二級結構預測一樣。HEPN結構係已顯示或預測具有RNA酶活性並且通常與各種防禦系統相關聯的小（~150 aa）α螺旋結構域（阿

南他拉曼，2013）（圖14，C2c2家族中的HEPN RxxxxH模體）。除催化性RxxxxH模體之外，HEPN結構域序列顯示少許保守性。因此，可能出現C2c2p含有兩個活性HEPN結構域。HEPN結構域對於CRISPR-Cas系統並不是新的，因為它常常與存在於許多III型CRISPR-Cas系統中的Csm6和Csx1蛋白質的CARF（CRISPR相關羅斯曼折疊）結構域締合（馬卡洛夫，2014）。該等蛋白質並不屬於自我調整模組或效應物複合物，而是似乎係存在於大部分CRISPR-Cas系統中並且涉及程式性細胞死亡以及在CRISPR反應過程中的調節功能的相關免疫模組的組分（庫尼恩，2013；馬卡洛夫，2012；馬卡洛夫，2013）。然而，C2c2p與Csm6和Csx1的不同之處在於除Cas1和Cas2之外，此大許多的蛋白質係C2c2座位中編碼的唯一蛋白質。因此，似乎C2c2p係該等推定的新型CRISPR-Cas系統的效應物並且HEPN結構域係其催化部分。在該等預測的HEPN結構域之外，C2c1p序列未顯示與其他蛋白質的可檢測相似性並且預測採用混合型α/β二級結構。

[1383] C2c2座位中的CRISPR陣列係高度異質的，其中長度係35至39 bp，並且係未結構化的（折疊能量係-0.9至4.7 kcal/mol）。根據CRISPRmap（蘭格，2013），該等CRISPR並不屬於任何建立的結構類別並且被分配至6個超類中的3個超類。僅來自斯氏利斯特菌（Listeria seeligeri）的CRISPR被分配至通常與II-C型系統相關聯的序列家族24。

[1384] C2c2座位的間隔區分析鑒定了與來自威氏李斯特菌（Listeria weihenstephanensis）的基因組序列和對於噬菌體基因組

的兩個未完成命中相同的30個核苷酸區域。

[1385] 考慮到C2c2獨特預測的效應物複合物，該等系統似乎具有作為推定的VI型CRISPR-Cas的品質。此外，考慮到所有實驗表徵和酶促活性的HEPN結構域係RNA酶，VI型系統可能在mRNA水平下起作用。

[1386] 申請人申請預測新第2類CRISPR-cas系統的簡單的直接計算策略。先前描述的第2類系統，即II型和推定的V型，由包括自我調整模組和含有效應物模組的單一大蛋白質的cas1和cas2基因（並且在一些情況下也有cas4）組成。因此，申請人推測含有cas1和大蛋白質的任何基因組座位可以是應進行詳細研究的新型第2類系統的潛在候選物。使用用於蛋白質序列比較的靈敏方法的此類分析能夠鑒定兩種強候選物，其中一種係先前描述的推定的V型的亞型，而另一種在新預測的效應蛋白存在強度方面具有作為一種新的推定的VI型的品質。許多該等新系統出現在未涵蓋其他CRISPR-Cas座位的細菌基因組中，這說明V型和VI型系統可以自發起作用。

[1387] 與先前的分析結果組合，（吉林斯基，2014；馬卡洛夫，2011），推定的V-B型的鑒定揭示了第2類CRISPR-Cas系統進化的主導主題。此類別的所有目前已知的系統的效應蛋白似乎從編碼含有RuvC-樣結構域的TnpB蛋白質的可易位件庫進化。TnpB的RuvC-樣結構域和第2類效應蛋白的同源結構域的序列對於可靠的系統發育分析而言太趨異。儘管如此，但是對於II型系統的效應蛋白Cas9，特定祖蛋白（ancestor）似乎容易鑒定，即在藍細菌中特

別豐富的TnpB樣蛋白家族，該蛋白家族與Cas9顯示相對高的序列相似性並且與它共用整個結構域體系結構，即RuvC樣和HNH核酸酶結構域和富精胺酸橋螺旋（吉林斯基，2014）（圖11，“第2類家族的結構域組構”；圖12，“第2類蛋白質中的TnpB同源區”）。與Cas9不同，它不可能將Cpf1和C2c1追蹤至特定TnpB家族；儘管位於RuvC-樣核酸酶的催化殘基中心的所有模體均具有保守性，但是該等蛋白質僅與TnpB屬性譜圖顯示有限的相似性。然而，考慮到C2c1p與Cpf1未顯示可檢測的序列相似性，在RuvC-模體與明確不相關的N-末端區域之間含有不同插入，似乎最可能的是Cpf1和C2c1獨立地來源於TnpB編碼元件庫內的不同家族。

[1388] 有意義的是TnpB蛋白似乎被“預先設計”用於第2類CRISPR-Cas效應物複合物，以使得它們明顯在多種不同的情景中募集。可想像地，TnpB蛋白的此類效用必須使用其預測的能力來切割單股DNA同時經由在Cas9中已顯示結合crRNA的富R橋螺旋結合RNA分子（金科，2014；西松，2014）。對於TnpB的功能瞭解較少。此蛋白質並不是易位所需要的，並且在一種情況中，已顯示下調易位（帕斯捷爾納克（Pasternak），2013）但是它們的作用機制仍未知。TnpB的實驗研究可能解釋了第2類CRISPR-Cas系統的機制方面。應注意Cpf1和C2c1的機制可以彼此類似但結合基本上不同於Cas9，因為前兩種蛋白質缺乏在Cas9中負責切割靶DNA股之一的HNH結構域（加西烏納斯，2012）（金科，2012）（陳，2014）。因此，Cpf1和C2c1的利用可以帶來另外的基因組編輯可能性。

[1389] 在進化方面中，引人注意的是第2類CRISPR-Cas似乎完全來源於不同的可易位件，給出了關於來自不同轉位子家族的*cas1*基因的可能起源的最新證據（庫尼恩，2015；格魯帕維克，2014）。此外，來自不同TnpB家族的效應蛋白的可能獨立起源連同對應cas1蛋白的不同系統發育親和性強烈提示第2類系統已在多種情景中藉由不同自我調整模組與產生效應蛋白的轉位子來源核酸酶組合來進化。此進化模式似乎係作為CRISPR-Cas進化特徵的最終的模組性表現形式（馬卡洛夫，2015），這提示自我調整模組和效應物模組的另外組合可能存在於自然界中。

[1390] 推定的VI型CRISPR-Cas系統涵蓋了含有可能具有RNA酶活性的兩個預測的HEPN結構域的預測的新型效應蛋白。HEPN結構不是其他CRISPR-Cas系統中的效應物複合物的部分，但是涉及許多防禦功能，包括預測的在不同CRISPR-Cas系統中的輔助作用（阿南他拉曼，2013）（馬卡洛夫，2015）。HEPN結構域作為預測的效應物模組的催化部分存在提示VI型系統靶向並切割mRNA。先前，已報導了對於某些III型CRISPR-Cas系統的mRNA靶向（黑爾（Hale），2014；黑爾，2009）（彭（Peng），2015）。儘管HEPN結構域迄今為止未在真實易位件中檢測到，但是它們的特徵係高橫向流動並且與可動遺傳因子諸如毒素抗毒素單元成整體（阿南他拉曼，2013）。因此，推定的VI型系統似乎適合第2類CRISPR-Cas來自可動部件的模組進化的一般範例，並且預期藉由分析基因組和巨集基因組學資料來發現另外的變體和新類型。

[1391] 模組進化係CRISPR-Cas系統的關鍵性特徵。此進化

模式似乎係第2類系統中最明顯的，它藉由來自不同的其他CRISPR-Cas系統的自我調整模組與似乎在多種獨立的情景下從可動遺傳因子募集的效應蛋白組合來進化。考慮到可動遺傳因子在細菌中的極端多樣性，似乎可能的是第2類CRISPR-Cas系統的效應物模組也是高度不同的。在此申請人採用描述CRISPR-Cas系統的兩種新變體的簡單計算方法，但是更多變體可能存在於已定序的細菌基因組中。儘管預期大部分（如果不是所有）該等新CRISPR-Cas系統係少見的，但是它們可以採用新策略和分子機制並且可以為基因組工程化和生物技術中的新應用提供主要資源。

[1392] 使用TBLASTN程式，使用Cas1譜圖作為查詢來搜索NCBI WGS資料庫。其中已鑒定Cas1命中的疊連群或全基因組部分的序列從同一資料庫中檢索。使用GENMARK切割並翻譯Cas1基因周圍的區域。針對來自CDD資料庫（瑪琪樂-鮑爾（Marchler-Bauer），2009）的特徵集合和在FTP處可獲得的特定Cas特徵搜索每種基因的預測的蛋白質，其中命中優先於Cas蛋白質。將先前開發的鑒定CRISPR座位的完整性的程序應用於每個座位。

[1393] 使用CRISPRmap（蘭格，2013）進行重複序列分類。

[1394] 反覆運算性譜圖用PSI-BLAST（安特斯庫爾，1997）搜索並且使用基於組成的統計學和低複雜度過濾關閉（filtering turned off）搜索兩個NCBI非冗餘（NR）資料庫的相似性較遠的序列。使用TBLAST程式針對WGS搜索每種鑒定的非冗餘蛋白質。使用HHpred，使用默認參數鑒定遠端序列相似性（塞汀（Soding），2005）。使用MUSCLE（愛德格，2004）構建多重序列比對。使用

Jpred 4預測蛋白質二級結構(德羅茲德斯基(Drozdetskiy),2015)。

[1395] 選擇的基因候選物

[1396] 基因ID：A；基因類型：C2C1；生物體：5.豐祐菌科細菌TAV5；間隔區長度—模式（範圍）：34（33至37）；DR1：GCCGCAGCGAAUGCCGUUUCACGAAUCGUCAGGCGG(SEQ ID NO: 27)；DR2：無；tracrRNA1：GCUGGAGACGUUUUUUGAAACGGCGAGUGCUGCGGAUAGCGAGUUUCUCUUGGGGAGGCGCUCGCGGCCACUUUU（SEQ ID NO: 28）；tracrRNA2：無；蛋白質序列：MSLNRIYQGRVAAVETGTALAKGNVEWMPAAGGDEVLWQHHELFQAAINYYLVALLALADKNNPVLGPLISQMDNPQSPYHVWGSFRRQGRQRTGLSQAVAPYITPGNNAPTLDEVFRSILAGNPTDRATLDAALMQLLKACDGAGAIQQEGRSYWPKFCDPDSTANFAGDPAMLRREQHRLLLPQVLHDPAITHDSPALGSFDTYSIATPDTRTPQLTGPKARARLEQAITLWRVRLPESAADFDRLASSLKKIPDDDSRLNLQGYVGSSAKGEVQARLFALLLFRHLERSSFTLGLLRSATPPPKNAETPPPAGVPLPAASAADPVRIARGKRSFVFRAFTSLPCWHGGDNIHPTWKSFDIAAFKYALTVINQIEEKTKERQKECAELETDFDYMHGRLAKIPVKYTTGEAEPPPILANDLRIPLLRELLQNIKVDTALTDGEAVSYGLQRRTIRGFRELRRIWRGHAPAGTVFSSELKEKLAGELRQFQTDNSTTIGSVQLFNELIQNPKYWPIWQAPDVETARQWADAGFADDPLAALVQEAELQEDIDALKAPVKLTPADPEYSRRQYDFNAVSKFGAGSRSANRHEPGQTERGHNTFTTEIAARNAADGNRWRA

THVRIHYSAPRLLRDGLRRPDTDGNEALEAVPWLQPMMEALAP
LPTLPQDLTGMPVFLMPDVTLSGERRILLNLPVTLEPAALVEQLG
NAGRWQNQFFGSREDPFALRWPADGAVKTAKGKTHIPWHQDRD
HFTVLGVDLGTRDAGALALLNVTAQKPAKPVHRIIGEADGRTW
YASLADARMIRLPGEDARLFVRGKLVQEPYGERGRNASLLEWE
DARNIILRLGQNPDELLGADPRRHSYPEINDKLLVALRRAQARLA
RLQNRSWRLRDLAESDKALDEIHAERAGEKPSPLPPLARDDAIK
STDEALLSQRDIIRRSFVQIANLILPLRGRRWEWRPHVEVPDCHI
LAQSDPGTDDTKRLVAGQRGISHERIEQIEELRRRCQSLNRALRH
KPGERPVLGRPAKGEEIADPCPALLEKINRLRDQRVDQTAHAILA
AALGVRLRAPSKDRAERRHRDIHGEYERFRAPADFVVIENLSRY
LSSQDRARSENTRLMQWCHRQIVQKLRQLCETYGIPVLAVPAAY
SSRFSSRDGSAGFRAVHLTPDHRHRMPWSRILARLKAHEEDGKR
LEKTVLDEARAVRGLFDRLDRFNAGHVPGKPWRTLLAPLPGGP
VFVPLGDATPMQADLNAAINIALRGIAAPDRHDIHHRLRAENKK
RILSLRLGTQREKARWPGGAPAVTLSTPNNGASPEDSDALPERVS
NLFVDIAGVANFERVTIEGVSQKFATGRGLWASVKQRAWNRVAR
LNETVTDNNRNEEEDDIPM（SEQ ID NO: 29）

[1397] 基因ID：B；基因類型：C2C1；生物體：7.嗜熱澱粉芽孢桿菌菌株B4166；間隔區長度—模式（範圍）：37（35-38）；DR1：GUCCAAGAAAAAAGAAAUGAUACGAGGCAUUAGCAC（SEQ ID NO: 30）；DR2：無；tracrRNA1：CUGGACGAUGUCUCUUUUAUUUCUUUUUUCUUGGAUCUGAG
UACGAGCACCCACAUUGGACAUUUCGCAUGGUGGGUGCUCG

UACUAUAGGUAAAACAAACCUUUUU（SEQ ID NO: 31）；tracrRNA2：無；蛋白質序列：

MATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEA
IYEHHEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDK
DVVFNILRELYEELVPSSVEKKGEANQLSNKFLYPLVDPNSQSGK
GTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILG
KLAEYGLIPLFIPFTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFI
QALERFLSWESWNLKVKEEYEKVEKEHKTLEERIKEDIQAFKSL
EQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDE
NEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNH
PEYPYLYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEERSG
SNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGKV
DIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGAR
VQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIH
RDDFPKFVNFKPKELTEWIKDSKGKKLKSGIESLEIGLRVMSIDL
GQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKL
PGETLVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITE
REKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLK
QLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKF
LLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMAN
TIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYNPYEERSR
FENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKT
GSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYP
DKGGEKFISLSKDRKLVTTHADINAAQNLQKRFWTRTHGFYKV

YCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWGN
AGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDP
SGNVFPSDKWMAAGVFFGKLERILISKLTNQYSISTIEDDSSKQS
M（SEQ ID NO: 32）

[1398] 基因ID：C；基因類型：C2C1；生物體：9.芽孢桿菌某種NSP2.1；間隔區長度—模式（範圍）：36（35-42）；DR1：GUUCGAAAGCUUAGUGGAAAGCUUCGUGGUUAGCAC（SEQ ID NO: 33）；DR2：無；tracrRNA1：CACGGAUAAUCACGACUUUCCACUAAGCUUUCGAAUUUUAU
GAUGCGAGCAUCCUCUCAGGUCAAAAAA（SEQ ID NO: 34）；tracrRNA2：無；蛋白質序列：
MAIRSIKLKLKTHTGPEAQNLRKGIWRTHRLLNEGVAYYMKML
LLFRQESTGERPKEELQEELICHIREQQQRNQADKNTQALPLDK
ALEALRQLYELLVPSSVGQSGDAQIISRKFLSPLVDPNSEGGKGT
SKAGAKPTWQKKKEANDPTWEQDYEKWKKRREEDPTASVITTL
EEYGIRPIFPLYTNTVTDIAWLPLQSNQFVRTWDRDMLQQAIERL
LSWESWNKRVQEEYAKLKEKMAQLNEQLEGGQEWISLLEQYEE
NRERELRENMTAANDKYRITKRQMKGWNELYELWSTFPASASH
EQYKEALKRVQQRLRGRFGDAHFFQYLMEEKNRLIWKGNPQRI
HYFVARNELTKRLEEAKQSATMTLPNARKHPLWVRFDARGGNL
QDYYLTAEADKPRSRRFVTFSQLIWPSESGWMEKKDVEVELALS
RQFYQQVKLLKNDKGKQKIEFKDKGSGSTFNGHLGGAKLQLER
GDLEKEEKNFEDGEIGSVYLNVVIDFEPLQEVKNGRVQAPYGQV
LQLIRRPNEFPKVTTYKSEQLVEWIKASPQHSAGVESLASGFRV

MSIDLGLRAAAATSIFSVEESSDKNAADFSYWIEGTPLVAVHQRS
YMLRLPGEQVEKQVMEKRDERFQLHQRVKFQIRVLAQIMRMAN
KQYGDRWDELDSLKQAVEQKKSPLDQTDRTFWEGIVCDLTKVL
PRNEADWEQAVVQIHRKAEEYVGKAVQAWRKRFAADERKGIA
GLSMWNIEELEGLRKLLISWSRRTRNPQEVNRFERGHTSHQRLL
THIQNVKEDRLKQLSHAIVMTALGYVYDERKQEWCAEYPACQV
ILFENLSQYRSNLDRSTKENSTLMKWAHRSIPKYVHMQAEPYGI
QIGDVRAEYSSRFYAKTGTPGIRCKKVRGQDLQGRRFENLQKRL
VNEQFLTEEQVKQLRPGDIVPDDSGELFMTLTDGSGSKEVVFLQ
ADINAAHNLQKRFWQRYNELFKVSCRVIVRDEEEYLVPKTKSVQ
AKLGKGLFVKKSDTAWKDVYVWDSQAKLKGKTTFTEESESPEQ
LEDFQEIIEEAEEAKGTYRTLFRDPSGVFFPESVWYPQKDFWGE
VKRKLYGKLRERFLTKAR（SEQ ID NO: 35）

[1399] 基因ID：D；基因類型：C2C2；生物體：4.毛螺菌科細菌（Lachnospiraceae bacterium）NK4A144 G619；間隔區長度—模式（範圍）：35；DR1：GUUUUGAGAAUAGCCCGACAUAGAGGGCAAUAGAC（SEQ ID NO: 36）；DR2：GUUAUGAAAACAGCCCGACAUAGAGGGCAAUAGACA（SEQ ID NO: 37）；tracrRNA1：無；tracrRNA2：無；蛋白質序列：

MKISKVDHTRMAVAKGNQHRRDEISGILYKDPTKTGSIDFDERF
KKLNCSAKILYHVFNGIAEGSNKYKNIVDKVNNNLDRVLFTGKS
YDRKSIIDIDTVLRNVEKINAFDRISTEEREQIIDDLLEIQLRKGLR
KGKAGLREVLLIGAGVIVRTDKKQEIADFLEILDEDFNKTNQAK

NIKLSIENQGLVVSPVSRGEERIFDVSGAQKGKSSKKAQEKEALS
AFLLDYADLDKNVRFEYLRKIRRLINLYFYVKNDDVMSLTEIPAE
VNLEKDFDIWRDHEQRKEENGDFVGCPDILLADRDVKKSNSKQ
VKIAERQLRESIREKNIKRYRFSIKTIEKDDGTYFFANKQISVFWI
HRIENAVERILGSINDKKLYRLRLGYLGEKVWKDILNFLSIKYIA
VGKAVFNFAMDDLQEKDRDIEPGKISENAVNGLTSFDYEQIKAD
EMLQREVAVNVAFAANNLARVTVDIPQNGEKEDILLWNKSDIKK
YKKNSKKGILKSILQFFGGASTWNMKMFEIAYHDQPGDYEENY
LYDIIQIIYSLRNKSFHFKTYDHGDKNWNRELIGKMIEHDAERVI
SVEREKFHSNNLPMFYKDADLKKILDLLYSDYAGRASQVPAFNT
VLVRKNFPEFLRKDMGYKVHFNNPEVENQWHSAVYYLYKEIYY
NLFLRDKEVKNLFYTSLKNIRSEVSDKKQKLASDDFASRCEEIED
RSLPEICQIIMTEYNAQNFGNRKVKSQRVIEKNKDIFRHYKMLLI
KTLAGAFSLYLKQERFAFIGKATPIPYETTDVKNFLPEWKSGMYA
SFVEEIKNNLDLQEWYIVGRFLNGRMLNQLAGSLRSYIQYAEDI
ERRAAENRNKLFSKPDEKIEACKKAVRVLDLCIKISTRISAEFTD
YFDSEDDYADYLEKYLKYQDDAIKELSGSSYAALDHFCNKDDL
KFDIYVNAGQKPILQRNIVMAKLFGPDNILSEVMEKVTESAIREY
YDYLKKVSGYRVRGKCSTEKEQEDLLKFQRLKNAVEFRDVTEY
AEVINELLGQLISWSYLRERDLLYFQLGFHYMCLKNKSFKPAEY
VDIRRNNGTIIHNAILYQIVSMYINGLDFYSCDKEGKTLKPIETGK
GVGSKIGQFIKYSQYLYNDPSYKLEIYNAGLEVFENIDEHDNITD
LRKYVDHFKYYAYGNKMSLLDLYSEFFDRFFTYDMKYQKNVV
NVLENILLRHFVIFYPKFGSGKKDVGIRDCKKERAQIEISEQSLTS

EDFMFKLDDKAGEEAKKFPARDERYLQTIAKLLYYPNEIEDMNR
FMKKGETINKKVQFNRKKKITRKQKNNSSNEVLSSTMGYLFKNI
KL（SEQ ID NO: 38）

[1400] 基因ID：E；基因類型：C2C2；生物體：8.斯氏利斯特菌血清變型1/2b str.SLCC3954；間隔區長度—模式(範圍)：30；DR1：GUUUUAGUCCUCUUUCAUAUAGAGGUAGUCUCUUAC（SEQ ID NO: 39）；DR2：無；tracrRNA1：AUGAAAAGAGGACUAAAACUGAAAGAGGACUAAAACACCA
GAUGUGGAUAACUAUAUUAGUGGCUAUUAAAAAUUCGUCG
AUAUUAGAGAGGAAACUUU(SEQ ID NO: 40)；tracrRNA2：無；蛋白質序列：
MWISIKTLIHHLGVLFFCDYMYNRREKKIIEVKTMRITKVEVDR
KKVLISRDKNGGKLVYENEMQDNTEQIMHHKKSSFYKSVVNKT
ICRPEQKQMKKLVHGLLQENSQEKIKVSDVTKLNISNFLNHRFK
KSLYYFPENSPDKSEEYRIEINLSQLLEDSLKKQQGTFICWESFSK
DMELYINWAENYISSKTKLIKKSIRNNRIQSTESRSGQLMDRYMK
DILNKNKPFDIQSVSEKYQLEKLTSALKATFKEAKKNDKEINYKL
KSTLQNHERQIIEELKENSELNQFNIEIRKHLETYFPIKKTNRKVG
DIRNLEIGEIQKIVNHRLKNKIVQRILQEGKLASYEIESTVNSNSL
QKIKIEEAFALKFINACLFASNNLRNMVYPVCKKDILMIGEFKNS
FKEIKHKKFIRQWSQFFSQEITVDDIELASWGLRGAIAPIRNEIIH
LKKHSWKKFFNNPTFKVKKSKIINGKTKDVTSEFLYKETLFKDY
FYSELDSVPELIINKMESSKILDYYSSDQLNQVFTIPNFELSLLTS
AVPFAPSFKRVYLKGFDYQNQDEAQPDYNLKLNIYNEKAFNSEA

FQAQYSLFKMVYYQVFLPQFTTNNDLFKSSVDFILTLNKERKGY
AKAFQDIRKMNKDEKPSEYMSYIQSQLMLYQKKQEEKEKINHF
EKFINQVFIKGFNSFIEKNRLTYICHPTKNTVPENDNIEIPFHTDM
DDSNIAFWLMCKLLDAKQLSELRNEMIKFSCSLQSTEEISTFTKA
REVIGLALLNGEKGCNDWKELFDDKEAWKKNMSLYVSEELLQS
LPYTQEDGQTPVINRSIDLVKKYGTETILEKLFSSSDDYKVSAKD
IAKLHEYDVTEKIAQQESLHKQWIEKPGLARDSAWTKKYQNVIN
DISNYQWAKTKVELTQVRHLHQLTIDLLSRLAGYMSIADRDFQF
SSNYILERENSEYRVTSWILLSENKNKNKYNDYELYNLKNASIK
VSSKNDPQLKVDLKQLRLTLEYLELFDNRLKEKRNNISHFNYLN
GQLGNSILELFDDARDVLSYDRKLKNAVSKSLKEILSSHGMEVT
FKPLYQTNHHLKIDKLQPKKIHHLGEKSTVSSNQVSNEYCQLVR
TLLTMK（SEQ ID NO: 41）

[1401] 基因ID：F；基因類型：C2C2；生物體：12.wadei纖毛菌屬F0279；間隔區長度—模式（範圍）：31；DR1：GUUUUAGUCCCCUUCGUUUUUGGGGUAGUCUAAAUC（SEQ ID NO: 42）；DR2：無；tracrRNA1：GAUUUAGAGCACCCCAAAAGUAAUGAAAAUUUGCAAUUAA AUAAGGAAUAUUAAAAAAAUGUGAUUUUAAAAAAAUUGAA GAAAUUAAAUGAAAAAAUUGUCCAAGUAAAAAAAA（SEQ ID NO: 43）；tracrRNA2：AUUUAGAUUACCCCUUUAAUUUAUUUUACCAUAUUUUUCUC AUAAUGCAAACUAAUAUUCCAAAAUUUUU（SEQ ID NO: 44）；蛋白質序列：

MGNLFGHKRWYEVRDKKDFKIKRKVKVKRNYDGNKYILNINE
NNNKEKIDNNKFIRKYINYKKNDNILKEFTRKFHAGNILFKLKG
KEGIIRIENNDDFLETEEVVLYIEAYGKSEKLKALGITKKKIIDEAI
RQGITKDDKKIEIKRQENEEEIEIDIRDEYTNKTLNDCSIILRIIEN
DELETKKSIYEIFKNINMSLYKIIEKIIENETEKVFENRYYEEHLRE
KLLKDDKIDVILTNFMEIREKIKSNLEILGFVKFYLNVGGDKKKS
KNKKMLVEKILNINVDLTVEDIADFVIKELEFWNITKRIEKVKKV
NNEFLEKRRNRTYIKSYVLLDKHEKFKIERENKKDKIVKFFVENI
KNNSIKEKIEKILAEFKIDELIKKLEKELKKGNCDTEIFGIFKKHY
KVNFDSKKFSKKSDEEKELYKIIYRYLKGRIEKILVNEQKVRLKK
MEKIEIEKILNESILSEKILKRVKQYTLEHIMYLGKLRHNDIDMTT
VNTDDFSRLHAKEELDLELITFFASTNMELNKIFSRENINNDENI
DFFGGDREKNYVLDKKILNSKIKIIRDLDFIDNKNNITNNFIRKFT
KIGTNERNRILHAISKERDLQGTQDDYNKVINIIQNLKISDEEVSK
ALNLDVVFKDKKNIITKINDIKISEENNNDIKYLPSFSKVLPEILN
LYRNNPKNEPFDTIETEKIVLNALIYVNKELYKKLILEDDLEENES
KNIFLQELKKTLGNIDEIDENIIENYYKNAQISASKGNNKAIKKY
QKKVIECYIGYLRKNYEELFDFSDFKMNIQEIKKQIKDINDNKTY
ERITVKTSDKTIVINDDFEYIISIFALLNSNAVINKIRNRFFATSVW
LNTSEYQNIIDILDEIMQLNTLRNECITENWNLNLEEFIQKMKEIE
KDFDDFKIQTKKEIFNNYYEDIKNNILTEFKDDINGCDVLEKKLE
KIVIFDDETKFEIDKKSNILQDEQRKLSNINKKDLKKKVDQYIKD
KDQEIKSKILCRIIFNSDFLKKYKKEIDNLIEDMESENENKFQEIY
YPKERKNELYIYKKNLFLNIGNPNFDKIYGLISNDIKMADAKFLF

NIDGKNIRKNKISEIDAILKNLNDKLNGYSKEYKEKYIKKLKEND
DFFAKNIQNKNYKSFEKDYNRVSEYKKIRDLVEFNYLNKIESYLI
DINWKLAIQMARFERDMHYIVNGLRELGIIKLSGYNTGISRAYPK
RNGSDGFYTTTAYYKFFDEESYKKFEKICYGFGIDLSENSEINKP
ENESIRNYISHFYIVRNPFADYSIAEQIDRVSNLLSYSTRYNNSTY
ASVFEVFKKDVNLDYDELKKKFKLIGNNDILERLMKPKKVSVLE
LESYNSDYIKNLIIELLTKIENTNDTL（SEQ ID NO: 45）

[1402] 基因ID：G；基因類型：C2C2；生物體：14.shahii纖毛菌屬DSM 19757 B031；間隔區長度—模式（範圍）：30（30-32）；DR1：GUUUUAGUCCCCUUCGAUAUUGGGGUGGUCUAUAUC（SEQ ID NO: 46）；DR2：無；tracrRNA1：AUUGAUGUGGUAUACUAAAAAUGGAAAAUUGUAUUUUUGA
UUAGAAAGAUGUAAAAUUGAUUUAAUUUAAAAAUAUUUUA
UUAGAUUAAAGUAGA（SEQ ID NO: 47）；tracrRNA2：無；蛋白質序列：
MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAK
DYKKAKQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSD
DDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQ
ESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFK
GFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKA
PEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFN
NYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKT
LKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQI
AAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQ

QVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAK
YLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQN
KDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLH
KLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRN
YITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKY
YLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPK
VFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDC
RKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGY
KLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLY
WKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIAN
KNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
NDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTF
NIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGY
LSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLE
KMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQT
GIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICY
NLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNH
NWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFA
KLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKN
MPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYF
EFVQNRNN（SEQ ID NO: 48）

[1403] 基因ID：H；基因類型：Cpf1；生物體：土拉熱弗朗西絲菌新殺手亞種U112間隔區長度—模式（範圍）：31；DR1：GUCUAAGAACUUUAAAUAAUUUCUACUGUUGUAGAU（SEQ

ID NO: 49) ; DR2 : 無 ; tracrRNA1 : AUCUACAAAAUUAUAAACUAAAUAAAGAUUCUUAUAAUAA CUUUAUAUAUAAUCGAAAUGUAGAGAAUUUU (SEQ ID NO: 50) ; tracrRNA2 : 無 ; 蛋 白 質 序 列 : MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAK DYKKAKQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSD DDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQ ESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFK GFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKA PEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFN NYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKT LKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQI AAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQ QVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAK YLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQN KDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLH KLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRN YITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKY YLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPK VFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDC RKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGY KLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLY WKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIAN KNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF

NDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTF
NIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGY
LSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLE
KMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQT
GIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICY
NLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNH
NWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFA
KLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKN
MPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYF
EFVQNRNN（SEQ ID NO: 51）

[1404] 用於合成的基因

[1405] 對於基因A至H，優化以進行人類表現並且將以下DNA序列附加到每個基因末端。注意此DNA序列含有終止密碼子（已加底線），因此不用將任何終止密碼子添加到密碼子優化的基因序列：

AAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAA
GAAAAAGgatccTACCCATACGATGTTCCAGATTACGCTTATCC
CTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACT
ATGCCTAA（SEQ ID NO: 52）

[1406] 對於優化，避免以下限制性位點：BamHI、EcoRI、HindIII、BsmBI、BsaI、BbsI、AgeI、XhoI、NdeI、NotI、KpnI、BsrGI、SpeI、XbaI、NheI

[1407] 將該等基因選殖到簡單的哺乳動物表現載體中：

[1408] >A

MSLNRIYQGRVAAVETGTALAKGNVEWMPAAGGDEVLWQ
HHELFQAAINYYLVALLALADKNNPVLGPLISQMDNPQSPYHV
WGSFRRQGRQRTGLSQAVAPYITPGNNAPTLDEVFRSILAGNPTD
RATLDAALMQLLKACDGAGAIQQEGRSYWPKFCDPDSTANFAG
DPAMLRREQHRLLLPQVLHDPAITHDSPALGSFDTYSIATPDTRT
PQLTGPKARARLEQAITLWRVRLPESAADFDRLASSLKKIPDDDS
RLNLQGYVGSSAKGEVQARLFALLLFRHLERSSFTLGLLRSATPP
PKNAETPPPAGVPLPAASAADPVRIARGKRSFVFRAFTSLPCWH
GGDNIHPTWKSFDIAAFKYALTVINQIEEKTKERQKECAELETDF
DYMHGRLAKIPVKYTTGEAEPPPILANDLRIPLLRELLQNIKVDT
ALTDGEAVSYGLQRRTIRGFRELRRIWRGHAPAGTVFSSELKEKL
AGELRQFQTDNSTTIGSVQLFNELIQNPKYWPIWQAPDVETARQ
WADAGFADDPLAALVQEAELQEDIDALKAPVKLTPADPEYSRRQ
YDFNAVSKFGAGSRSANRHEPGQTERGHNTFTTEIAARNAADG
NRWRATHVRIHYSAPRLLRDGLRRPDTDGNEALEAVPWLQPMM
EALAPLPTLPQDLTGMPVFLMPDVTLSGERRILLNLPVTLEPAAL
VEQLGNAGRWQNQFFGSREDPFALRWPADGAVKTAKGKTHIPW
HQDRDHFTVLGVDLGTRDAGALALLNVTAQKPAKPVHRIIGEA
DGRTWYASLADARMIRLPGEDARLFVRGKLVQEPYGERGRNAS
LLEWEDARNIILRLGQNPDELLGADPRRHSYPEINDKLLVALRRA
QARLARLQNRSWRLRDLAESDKALDEIHAERAGEKPSPLPPLAR
DDAIKSTDEALLSQRDIIRRSFVQIANLILPLRGRRWEWRPHVEV
PDCHILAQSDPGTDDTKRLVAGQRGISHERIEQIEELRRRCQSLN

RALRHKPGERPVLGRPAKGEEIADPCPALLEKINRLRDQRVDQT
AHAILAAALGVRLRAPSKDRAERRHRDIHGEYERFRAPADFVVI
ENLSRYLSSQDRARSENTRLMQWCHRQIVQKLRQLCETYGIPVL
AVPAAYSSRFSSRDGSAGFRAVHLTPDHRHRMPWSRILARLKAH
EEDGKRLEKTVLDEARAVRGLFDRLDRFNAGHVPGKPWRTLLA
PLPGGPVFVPLGDATPMQADLNAAINIALRGIAAPDRHDIHHRLR
AENKKRILSLRLGTQREKARWPGGAPAVTLSTPNNGASPEDSDA
LPERVSNLFVDIAGVANFERVTIEGVSQKFATGRGLWASVKQRA
WNRVARLNETVTDNNRNEEEDDIPM (SEQ ID NO: 53)

[1409] >B

[1410]

MATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIR
QEAIYEHHEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHE
VDKDVVFNILRELYEELVPSSVEKKGEANQLSNKFLYPLVDPNSQ
SGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAK
ILGKLAEYGLIPLFIPFTDSNEPIVKEIKWMEKSRNQSVRRLDKD
MFIQALERFLSWESWNLKVKEEYEKVEKEHKTLEERIKEDIQAF
KSLEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLK
MDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIW
RNHPEYPYLYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFE
ERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEE
KGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTL
GGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSK

SLKIHRDDFPKFVNFKPKELTEWIKDSKGKKLKSGIESLEIGLRV
MSIDLGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRAS
FNIKLPGETLVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQF
EDITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKDW
VAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEI
DRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDR
LKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYN
PYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFS
SRFHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAVL
KEGDLYPDKGGEKFISLSKDRKLVTTHADINAAQNLQKRFWTRT
HGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDG
VYEWGNAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEK
LMLYRDPSGNVFPSDKWMAAGVFFGKLERILISKLTNQYSISTIE
DDSSKQSM (SEQ ID NO: 54)

[1411] >C

[1412]

MAIRSIKLKLKTHTGPEAQNLRKGIWRTHRLLNEGVAYYMK
MLLLFRQESTGERPKEELQEELICHIREQQQRNQADKNTQALPL
DKALEALRQLYELLVPSSVGQSGDAQIISRKFLSPLVDPNSEGGK
GTSKAGAKPTWQKKKEANDPTWEQDYEKWKKRREEDPTASVI
TTLEEYGIRPIFPLYTNTVTDIAWLPLQSNQFVRTWDRDMLQQAI
ERLLSWESWNKRVQEEYAKLKEKMAQLNEQLEGGQEWISLLEQ
YEENRERELRENMTAANDKYRITKRQMKGWNELYELWSTFPAS

ASHEQYKEALKRVQQRLRGRFGDAHFFQYLMEEKNRLIWKGNP
QRIHYFVARNELTKRLEEAKQSATMTLPNARKHPLWVRFDARG
GNLQDYYLTAEADKPRSRRFVTFSQLIWPSESGWMEKKDVEVE
LALSRQFYQQVKLLKNDKGKQKIEFKDKGSGSTFNGHLGGAKL
QLERGDLEKEEKNFEDGEIGSVYLNVVIDFEPLQEVKNGRVQAP
YGQVLQLIRRPNEFPKVTTYKSEQLVEWIKASPQHSAGVESLAS
GFRVMSIDLGLRAAAATSIFSVEESSDKNAADFSYWIEGTPLVAV
HQRSYMLRLPGEQVEKQVMEKRDERFQLHQRVKFQIRVLAQIM
RMANKQYGDRWDELDSLKQAVEQKKSPLDQTDRTFWEGIVCD
LTKVLPRNEADWEQAVVQIHRKAEEYVGKAVQAWRKRFAADER
KGIAGLSMWNIEELEGLRKLLISWSRRTRNPQEVNRFERGHTSH
QRLLTHIQNVKEDRLKQLSHAIVMTALGYVYDERKQEWCAEYP
ACQVILFENLSQYRSNLDRSTKENSTLMKWAHRSIPKYVHMQA
EPYGIQIGDVRAEYSSRFYAKTGTPGIRCKKVRGQDLQGRRFEN
LQKRLVNEQFLTEEQVKQLRPGDIVPDDSGELFMTLTDGSGSKE
VVFLQADINAAHNLQKRFWQRYNELFKVSCRVIVRDEEEYLVPK
TKSVQAKLGKGLFVKKSDTAWKDVYVWDSQAKLKGKTTFTEE
SESPEQLEDFQEIIEEAEEAKGTYRTLFRDPSGVFFPESVWYPQK
DFWGEVKRKLYGKLRERFLTKAR (SEQ ID NO: 55)

[1413] >D

[1414]

MKISKVDHTRMAVAKGNQHRRDEISGILYKDPTKTGSIDFD
ERFKKLNCSAKILYHVFNGIAEGSNKYKNIVDKVNNNLDRVLFT

GKSYDRKSIIDIDTVLRNVEKINAFDRISTEEREQIIDDLLEIQLRK
GLRKGKAGLREVLLIGAGVIVRTDKKQEIADFLEILDEDFNKTN
QAKNIKLSIENQGLVVSPVSRGEERIFDVSGAQKGKSSKKAQEK
EALSAFLLDYADLDKNVRFEYLRKIRRLINLYFYVKNDDVMSLT
EIPAEVNLEKDFDIWRDHEQRKEENGDFVGCPDILLADRDVKKS
NSKQVKIAERQLRESIREKNIKRYRFSIKTIEKDDGTYFFANKQIS
VFWIHRIENAVERILGSINDKKLYRLRLGYLGEKVWKDILNFLSI
KYIAVGKAVFNFAMDDLQEKDRDIEPGKISENAVNGLTSFDYEQI
KADEMLQREVAVNVAFAANNLARVTVDIPQNGEKEDILLWNKS
DIKKYKKNSKKGILKSILQFFGGASTWNMKMFEIAYHDQPGDYE
ENYLYDIIQIIYSLRNKSFHFKTYDHGDKNWNRELIGKMIEHDAE
RVISVEREKFHSNNLPMFYKDADLKKILDLLYSDYAGRASQVPA
FNTVLVRKNFPEFLRKDMGYKVHFNNPEVENQWHSAVYYLYKE
IYYNLFLRDKEVKNLFYTSLKNIRSEVSDKKQKLASDDFASRCE
EIEDRSLPEICQIIMTEYNAQNFGNRKVKSQRVIEKNKDIFRHYK
MLLIKTLAGAFSLYLKQERFAFIGKATPIPYETTDVKNFLPEWKS
GMYASFVEEIKNNLDLQEWYIVGRFLNGRMLNQLAGSLRSYIQ
YAEDIERRAAENRNKLFSKPDEKIEACKKAVRVLDLCIKISTRISA
EFTDYFDSEDDYADYLEKYLKYQDDAIKELSGSSYAALDHFCNK
DDLKFDIYVNAGQKPILQRNIVMAKLFGPDNILSEVMEKVTESAI
REYYDYLKKVSGYRVRGKCSTEKEQEDLLKFQRLKNAVEFRDV
TEYAEVINELLGQLISWSYLRERDLLYFQLGFHYMCLKNKSFKP
AEYVDIRRNNGTIIHNAILYQIVSMYINGLDFYSCDKEGKTLKPIE
TGKGVGSKIGQFIKYSQYLYNDPSYKLEIYNAGLEVFENIDEHDN

ITDLRKYVDHFKYYAYGNKMSLLDLYSEFFDRFFTYDMKYQKN
VVNVLENILLRHFVIFYPKFGSGKKDVGIRDCKKERAQIEISEQS
LTSEDFMFKLDDKAGEEAKKFPARDERYLQTIAKLLYYPNEIED
MNRFMKKGETINKKVQFNRKKKITRKQKNNSSNEVLSSTMGYL
FKNIKL (SEQ ID NO: 56)

[1415] >E

[1416]

MWISIKTLIHHLGVLFFCDYMYNRREKKIIEVKTMRITKVE
VDRKKVLISRDKNGGKLVYENEMQDNTEQIMHHKKSSFYKSV
VNKTICRPEQKQMKKLVHGLLQENSQEKIKVSDVTKLNISNFL
NHRFKKSLYYFPENSPDKSEEYRIEINLSQLLEDSLKKQQGTFIC
WESFSKDMELYINWAENYISSKTKLIKKSIRNNRIQSTESRSGQL
MDRYMKDILNKNKPFDIQSVSEKYQLEKLTSALKATFKEAKKN
DKEINYKLKSTLQNHERQIIEELKENSELNQFNIEIRKHLETYFP
IKKTNRKVGDIRNLEIGEIQKIVNHRLKNKIVQRILQEGKLASY
EIESTVNSNSLQKIKIEEAFALKFINACLFASNNLRNMVYPVCK
KDILMIGEFKNSFKEIKHKKFIRQWSQFFSQEITVDDIELASWGL
RGAIAPIRNEIIHLKKHSWKKFFNNPTFKVKKSKIINGKTKDVT
SEFLYKETLFKDYFYSELDSVPELIINKMESSKILDYYSSDQLNQ
VFTIPNFELSLLTSAVPFAPSFKRVYLKGFDYQNQDEAQPDYNL
KLNIYNEKAFNSEAFQAQYSLFKMVYYQVFLPQFTTNNDLFKS
SVDFILTLNKERKGYAKAFQDIRKMNKDEKPSEYMSYIQSQLM
LYQKKQEEKEKINHFEKFINQVFIKGFNSFIEKNRLTYICHPTKN

TVPENDNIEIPFHTDMDDSNIAFWLMCKLLDAKQLSELRNEMI
KFSCSLQSTEEISTFTKAREVIGLALLNGEKGCNDWKELFDDKE
AWKKNMSLYVSEELLQSLPYTQEDGQTPVINRSIDLVKKYGTE
TILEKLFSSSDDYKVSAKDIAKLHEYDVTEKIAQQESLHKQWIE
KPGLARDSAWTKKYQNVINDISNYQWAKTKVELTQVRHLHQL
TIDLLSRLAGYMSIADRDFQFSSNYILERENSEYRVTSWILLSEN
KNKNKYNDYELYNLKNASIKVSSKNDPQLKVDLKQLRLTLEY
LELFDNRLKEKRNNISHFNYLNGQLGNSILELFDDARDVLSYD
RKLKNAVSKSLKEILSSHGMEVTFKPLYQTNHHLKIDKLQPKKI
HHLGEKSTVSSNQVSNEYCQLVRTLLTMK (SEQ ID NO: 57)

[1417] >F

[1418]

MKVTKVDGISHKKYIEEGKLVKSTSEENRTSERLSELLSIR
LDIYIKNPDNASEEENRIRRENLKKFFSNKVLHLKDSVLYLKNR
KEKNAVQDKNYSEEDISEYDLKNKNSFSVLKKILLNEDVNSEE
LEIFRKDVEAKLNKINSLKYSFEENKANYQKINENNVEKVGGK
SKRNIIYDYYRESAKRNDYINNVQEAFDKLYKKEDIEKLFFLIE
NSKKHEKYKIREYYHKIIGRKNDKENFAKIIYEEIQNVNNIKELI
EKIPDMSELKKSQVFYKYYLDKEELNDKNIKYAFCHFVEIEMS
QLLKNYVYKRLSNISNDKIKRIFEYQNLKKLIENKLLNKLDTYV
RNCGKYNYYLQVGEIATSDFIARNRQNEAFLRNIIGVSSVAYFS
LRNILETENENDITGRMRGKTVKNNKGEEKYVSGEVDKIYNEN
KQNEVKENLKMFYSYDFNMDNKNEIEDFFANIDEAISSIRHGIV

HFNLELEGKDIFAFKNIAPSEISKKMFQNEINEKKLKLKIFKQLN
SANVFNYYEKDVIIKYLKNTKFNFVNKNIPFVPSFTKLYNKIED
LRNTLKFFWSVPKDKEEKDAQIYLLKNIYYGEFLNKFVKNSKV
FFKITNEVIKINKQRNQKTGHYKYQKFENIEKTVPVEYLAIIQSR
EMINNQDKEEKNTYIDFIQQIFLKGFIDYLNKNNLKYIESNNNN
DNNDIFSKIKIKKDNKEKYDKILKNYEKHNRNKEIPHEINEFVR
EIKLGKILKYTENLNMFYLILKLLNHKELTNLKGSLEKYQSANK
EETFSDELELINLLNLDNNRVTEDFELEANEIGKFLDFNENKIK
DRKELKKFDTNKIYFDGENIIKHRAFYNIKKYGMLNLLEKIAD
KAKYKISLKELKEYSNKKNEIEKNYTMQQNLHRKYARPKKDE
KFNDEDYKEYEKAIGNIQKYTHLKNKVEFNELNLLQGLLLKIL
HRLVGYTSIWERDLRFRLKGEFPENHYIEEIFNFDNSKNVKYKS
GQIVEKYINFYKELYKDNVEKRSIYSDKKVKKLKQEKKDLYIR
NYIAHFNYIPHAEISLLEVLENLRKLLSYDRKLKNAIMKSIVDIL
KEYGFVATFKIGADKKIEIQTLESEKIVHLKNLKKKKLMTDRNS
EELCELVKVMFEYKALE (SEQ ID NO: 58)

[1419] >G

[1420]

MGNLFGHKRWYEVRDKKDFKIKRKVKVKRNYDGNKYIL
NINENNNKEKIDNNKFIRKYINYKKNDNILKEFTRKFHAGNILF
KLKGKEGIIRIENNDDFLETEEVVLYIEAYGKSEKLKALGITKK
KIIDEAIRQGITKDDKKIEIKRQENEEEIEIDIRDEYTNKTLNDCS
IILRIIENDELETKKSIYEIFKNINMSLYKIIEKIIENETEKVFENRY

YEEHLREKLLKDDKIDVILTNFMEIREKIKSNLEILGFVKFYLNV
GGDKKKSKNKKMLVEKILNINVDLTVEDIADFVIKELEFWNIT
KRIEKVKKVNNEFLEKRRNRTYIKSYVLLDKHEKFKIERENKK
DKIVKFFVENIKNNSIKEKIEKILAEFKIDELIKKLEKELKKGNC
DTEIFGIFKKHYKVNFDSKKFSKKSDEEKELYKIIYRYLKGRIEK
ILVNEQKVRLKKMEKIEIEKILNESILSEKILKRVKQYTLEHIMY
LGKLRHNDIDMTTVNTDDFSRLHAKEELDLELITFFASTNMEL
NKIFSRENINNDENIDFFGGDREKNYVLDKKILNSKIKIIRDLDF
IDNKNNITNNFIRKFTKIGTNERNRILHAISKERDLQGTQDDYN
KVINIIQNLKISDEEVSKALNLDVVFKDKKNIITKINDIKISEENN
NDIKYLPSFSKVLPEILNLYRNNPKNEPFDTIETEKIVLNALIYV
NKELYKKLILEDDLEENESKNIFLQELKKTLGNIDEIDENIIENY
YKNAQISASKGNNKAIKKYQKKVIECYIGYLRKNYEELFDFSD
FKMNIQEIKKQIKDINDNKTYERITVKTSDKTIVINDDFEYIISIF
ALLNSNAVINKIRNRFFATSVWLNTSEYQNIIDILDEIMQLNTLR
NECITENWNLNLEEFIQKMKEIEKDFDDFKIQTKKEIFNNYYED
IKNNILTEFKDDINGCDVLEKKLEKIVIFDDETKFEIDKKSNILQ
DEQRKLSNINKKDLKKKVDQYIKDKDQEIKSKILCRIIFNSDFL
KKYKKEIDNLIEDMESENENKFQEIYYPKERKNELYIYKKNLFL
NIGNPNFDKIYGLISNDIKMADAKFLFNIDGKNIRKNKISEIDAI
LKNLNDKLNGYSKEYKEKYIKKLKENDDFFAKNIQNKNYKSFE
KDYNRVSEYKKIRDLVEFNYLNKIESYLIDINWKLAIQMARFER
DMHYIVNGLRELGIIKLSGYNTGISRAYPKRNGSDGFYTTTAYY
KFFDEESYKKFEKICYGFGIDLSENSEINKPENESIRNYISHFYIV

RNPFADYSIAEQIDRVSNLLSYSTRYNNSTYASVFEVFKKDVNL
DYDELKKKFKLIGNNDILERLMKPKKVSVLELESYNSDYIKNLI
IELLTKIENTNDTL (SEQ ID NO: 59)

[1421] >H

[1422]

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDE
KRAKDYKKAKQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFK
LKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLID
AKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKG
WTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKY
ESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDE
VFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLY
SQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTT
MQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKN
DKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQE
LIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFA
AIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIK
DLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFE
LANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKE
PDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKI
VYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNG
SPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQ
RYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIY

NKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELF
YRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTE
DKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGE
RHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKD
RDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFE
DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKT
GGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFV
NQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGD
KAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLL
KDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKT
GTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLK
GLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN（SEQ ID
NO: 60）

[1423] 對於A-座位至G-座位，將該等基因選殖並插入到低拷貝質粒中。使用不含有Amp抗性的載體。

[1424] >A-座位

[1425]

TATCCGGTCGAATCGAGAATGACGACCGCTACGTCTTGG
ACTACGAAGCCGTGGCCCTTGCCGATGCTCTCGGTGTGGATG
TTGCCGACCTGTTCCGCAAGATCGATTGCCCCAAGAACCTGC
TGCGCAGGCGGGCAGGGTAGGGGAGCGGTTTCCGGCGGAGA
TTTTCGGAGGCGCCGGTAACGTTATGTCGGGGAATTTGCTATA
CATCGACGATAATTAGTTTTGTTGATTCAGGATCGAAATGCGC

TCAAACAAAGAACGTTCCGCGTTTCCCTCATGCGCTACTACG

CCCACACCGCCATCTTTCGGCACGCAAACAAAGCAGATGGG

TTGCCTGTCAATGGGTGATCATTGCCTGAAGTTACCATCCATC

AATAATATAAATCATCCTTACTCCGAATGTCCCTCAATCGCATC

TATCAAGGCCGCGTGGCGGCCGTCGAAACAGGAACGGCCTT

AGCGAAAGGTAATGTCGAATGGATGCCTGCCGCAGGAGGCG

ACGAAGTTCTCTGGCAGCACCACGAACTTTTCCAAGCTGCCA

TCAACTACTATCTCGTCGCCCTGCTCGCACTCGCCGACAAAA

ACAATCCCGTACTTGGCCCGCTGATCAGCCAGATGGATAATC

CCCAAAGCCCTTACCATGTCTGGGGAAGTTTCCGCCGCCAAG

GACGTCAGCGCACAGGTCTCAGTCAAGCCGTTGCACCTTATA

TCACGCCGGGCAATAACGCTCCCACCCTTGACGAAGTTTTCC

GCTCCATTCTTGCGGGCAACCCAACCGACCGCGCAACTTTGG

ACGCTGCACTCATGCAATTGCTCAAGGCTTGTGACGGCGCGG

GCGCTATCCAGCAGGAAGGTCGTTCCTACTGGCCCAAATTCT

GCGATCCTGACTCCACTGCCAACTTCGCGGGAGATCCGGCCA

TGCTCCGGCGTGAACAACACCGCCTCCTCCTTCCGCAAGTTC

TCCACGATCCGGCGATTACTCACGACAGTCCTGCCCTTGGCT

CGTTCGACACTTATTCGATTGCTACCCCCGACACCAGAACTC

CTCAACTCACCGGCCCCAAGGCACGCGCCCGTCTTGAGCAG

GCGATCACCCTCTGGCGCGTCCGTCTTCCCGAATCGGCTGCT

GACTTCGATCGCCTTGCCAGTTCCCTCAAAAAAATTCCGGAC

GACGATTCTCGCCTTAACCTTCAGGGCTACGTCGGCAGCAGT

GCGAAAGGCGAAGTTCAGGCCCGTCTTTTCGCCCTTCTGCTA

TTCCGTCACCTGGAGCGTTCCTCCTTTACGCTTGGCCTTCTCC

GTTCCGCCACCCCGCCGCCCAAGAACGCTGAAACACCTCCT

CCCGCCGGCGTTCCTTTACCTGCGGCGTCCGCAGCCGATCCG

GTGCGGATAGCCCGTGGCAAACGCAGTTTTGTTTTTCGCGCA

TTCACCAGTCTCCCCTGCTGGCATGGCGGTGATAACATCCATC

CCACCTGGAAGTCATTCGACATCGCAGCGTTCAAATATGCCC

TCACGGTCATCAACCAGATCGAGGAAAAGACGAAAGAACGC

CAAAAAGAATGTGCGGAACTTGAAACTGATTTCGACTACATG

CACGGACGGCTCGCCAAGATTCCGGTAAAATACACGACCGG

CGAAGCCGAACCGCCCCCCATTCTCGCAAACGATCTCCGCAT

CCCCCTCCTCCGCGAACTTCTCCAGAATATCAAGGTCGACAC

CGCACTCACCGATGGCGAAGCCGTCTCCTATGGTCTCCAACG

CCGCACCATTCGCGGTTTCCGCGAGCTGCGCCGCATCTGGCG

CGGCCATGCCCCCGCTGGCACGGTCTTTTCCAGCGAGTTGAA

AGAAAAACTAGCCGGCGAACTCCGCCAGTTCCAGACCGACA

ACTCCACCACCATCGGCAGCGTCCAACTCTTCAACGAACTCA

TCCAAAACCCGAAATACTGGCCCATCTGGCAGGCTCCTGACG

TCGAAACCGCCCGCCAATGGGCCGATGCCGGTTTTGCCGACG

ATCCGCTCGCCGCCCTTGTGCAAGAAGCCGAACTCCAGGAA

GACATCGACGCCCTCAAGGCTCCAGTCAAACTCACTCCGGCC

GATCCTGAGTATTCAAGAAGGCAATACGATTTCAATGCCGTC

AGCAAATTCGGGGCCGGCTCCCGCTCCGCCAATCGCCACGA

ACCCGGGCAGACGGAGCGCGGCCACAACACCTTTACCACCG

AAATCGCCGCCCGTAACGCGGCGGACGGGAACCGCTGGCGG

GCAACCCACGTCCGCATCCATTACTCCGCTCCCCGCCTTCTTC

GTGACGGACTCCGCCGACCTGACACCGACGGCAACGAAGCC

CTGGAAGCCGTCCCTTGGCTCCAGCCCATGATGGAAGCCCTC

GCCCCTCTCCCGACGCTTCCGCAAGACCTCACAGGCATGCCG

GTCTTCCTCATGCCCGACGTCACCCTTTCCGGTGAGCGTCGC

ATCCTCCTCAATCTTCCTGTCACCCTCGAACCAGCCGCTCTTG

TCGAACAACTGGGCAACGCCGGTCGCTGGCAAAACCAGTTC

TTCGGCTCCCGCGAAGATCCATTCGCTCTCCGATGGCCCGCC

GACGGTGCTGTAAAAACCGCCAAGGGGAAAACCCACATACC

TTGGCACCAGGACCGCGATCACTTCACCGTACTCGGCGTGGA

TCTCGGCACGCGCGATGCCGGGGCGCTCGCTCTTCTCAACGT

CACTGCGCAAAAACCGGCCAAGCCGGTCCACCGCATCATTG

GTGAGGCCGACGGACGCACCTGGTATGCCAGCCTTGCCGAC

GCTCGCATGATCCGCCTGCCCGGGGAGGATGCCCGGCTCTTT

GTCCGGGGAAAACTCGTTCAGGAACCCTATGGTGAACGCGG

GCGAAACGCGTCTCTTCTCGAATGGGAAGACGCCCGCAATAT

CATCCTTCGCCTTGGCCAAAATCCCGACGAACTCCTCGGCGC

CGATCCCCGGCGCCATTCGTATCCGGAAATAAACGATAAACT

TCTCGTCGCCCTTCGCCGCGCTCAGGCCCGTCTTGCCCGTCT

CCAGAACCGGAGCTGGCGGTTGCGCGACCTTGCAGAATCGG

ACAAGGCCCTTGATGAAATCCATGCCGAGCGTGCCGGGGAG

AAGCCTTCTCCGCTTCCGCCCTTGGCTCGCGACGATGCCATC

AAAAGCACCGACGAAGCCCTCCTTTCCCAGCGTGACATCATC

CGGCGATCCTTCGTTCAGATCGCCAACTTGATCCTTCCCCTTC

GCGGACGCCGATGGGAATGGCGGCCCCATGTCGAGGTCCCG

GATTGCCACATCCTTGCGCAGAGCGATCCCGGTACGGATGAC

ACCAAGCGTCTTGTCGCCGGACAACGCGGCATCTCTCACGA

GCGTATCGAGCAAATCGAAGAACTCCGTCGTCGCTGCCAATC

CCTCAACCGTGCCCTGCGTCACAAACCCGGAGAGCGTCCCG

TGCTCGGACGCCCCGCCAAGGGCGAGGAAATCGCCGATCCC

TGTCCCGCGCTCCTCGAAAAGATCAACCGTCTCCGGGACCAG

CGCGTTGACCAAACCGCGCATGCCATCCTCGCCGCCGCTCTC

GGTGTTCGACTCCGCGCCCCCTCAAAAGACCGCGCCGAACG

CCGCCATCGCGACATCCATGGCGAATACGAACGCTTTCGTGC

GCCCGCTGATTTTGTCGTCATCGAAAACCTCTCCCGTTATCTC

AGCTCGCAGGATCGTGCTCGTAGTGAAAACACCCGTCTCATG

CAGTGGTGCCATCGCCAGATCGTGCAAAAACTCCGTCAGCTC

TGCGAGACCTACGGCATCCCCGTCCTCGCCGTCCCGGCGGCC

TACTCATCGCGTTTTTCTTCCCGGGACGGCTCGGCCGGATTCC

GGGCCGTCCATCTGACACCGGACCACCGTCACCGGATGCCAT

GGAGCCGCATCCTCGCCCGCCTCAAGGCCCACGAGGAAGAC

GGAAAAAGACTCGAAAAGACGGTGCTCGACGAGGCTCGCGC

CGTCCGGGGACTCTTTGACCGGCTCGACCGGTTCAACGCCG

GGCATGTCCCGGGAAAACCTTGGCGCACGCTCCTCGCGCCG

CTCCCCGGCGGCCCTGTGTTTGTCCCCCTCGGGGACGCCACA

CCCATGCAGGCCGATCTGAACGCCGCCATCAACATCGCCCTC

CGGGGCATCGCGGCTCCCGACCGCCACGACATCCATCACCGG

CTCCGTGCCGAAAACAAAAAACGCATCCTGAGCTTGCGTCTC

GGCACTCAGCGCGAGAAAGCCCGCTGGCCTGGAGGAGCTCC

GGCGGTGACACTCTCCACTCCGAACAACGGCGCCTCTCCCG

AAGATTCCGATGCGTTGCCCGAACGGGTATCCAACCTGTTTG

TGGACATCGCCGGTGTCGCCAACTTCGAGCGAGTCACGATCG

AAGGAGTCTCGCAAAAATTCGCCACCGGGCGTGGCCTTTGG

GCCTCCGTCAAGCAACGTGCATGGAACCGCGTTGCCAGACT

CAACGAGACAGTAACAGATAACAACAGGAACGAAGAGGAG

GACGACATTCCGATGTAACCATTGCTTCATTACATCTGAGTCT

CCCCTCAATCCCTCTGCCCCATGCGTGATATAACCTCCACCTC

ATGTCCCGGATCGGCGCCGGCAACCTGTAGTTCCCTTCCATC

CTCCAACACTCCCGCAGATCGCGATCCGCTGCCGCCGATGCC

GGTGCGCCGCCTTCACAACTATCTCTACTGTCCGCGGCTTTTT

TATCTCCAGTGGGTCGAGAATCTCTTTGAGGAAAATGCCGAC

ACCATTGCCGGCAGCGCCGTGCATCGTCACGCCGACAAACCT

ACGCGTTACGATGATGAAAAAGCCGAGGCACTTCGCACTGGT

CTCCCTGAAGGCGCGCACATACGCAGCCTTCGCCTGGAAAAC

GCCCAACTCGGTCTCGTTGGCGTGGTGGATATCGTGGAGGGA

GGCCCCGACGGACTCGAACTCGTCGACTACAAAAAAGGTTC

CGCCTTCCGCCTCGACGACGGCACGCTCGCTCCCAAGGAAA

ACGACACCGTGCAACTTGCCGCCTACGCTCTTCTCCTGGCTG

CCGATGGTGCGCGCGTTGCGCCCATGGCGACGGTCTATTACG

CTGCCGATCGCCGGCGTGTCACCTTCCCGCTCGATGACGCCC

TCTACGCCCGCACCCGTTCCGCCCTCGAAGAGGCCCGCGCCG

TTGCAACCTCGGGGCGCATACCTCCGCCGCTCGTCTCTGACG

TCCGCTGCCTCCATTGTTCCTCCTATGCGCTTTGCCTTCCCCG

CGAGTCCGCCTGGTGGTGCCGCCATCGCAGCACGCCGCGGG

GAGCCGGCCACACCCCCATGTTGCCGGGCTTTGAGGATGACG

CCGCCGCCATTCACCAAATCTCCGAACCTGACACCGAGCCAC

CACCCGATCTTGCCAGCCAGCCTCCCCGTCCCCCGCGGCTCG

ATGGAGAATTGTTGGTTGTCCAGACTCCGGGAGCGATGATCG

GACAAAGCGGCGGTGAGTTTACCGTGTCCGTCAAGGGTGAG

GTTTTGCGCAAGCTTCCGGTTCATCAACTCCGGGCCATTTAC

GTTTACGGAGCCGTGCAACTCACGGCGCATGCTGTGCAGACC

GCCCTTGAGGAGGATATCGACGTCTCCTATTTTGCGCCCAGC

GGCCGCTTTCTTGGCCTCCTCCGCGGCCTGCCCGCATCCGGC

GTGGATGCGCGTCTCGGGCAATACACCCTGTTTCGCGAACCC

TTTGGCCGTCTCCGTCTCGCCTGCGAGGCGATTCGGGCCAAG

ATCCATAACCAGCGCGTCCTCCTCATGCGTAACGGCGAGCCC

GGGGAGGGCGTCTTGCGCGAACTCGCCCGTCTGCGCGACGC

CACCAGTGAGGCGACTTCGCTCGACGAACTCCTCGGCATCG

AGGGCATCGCCGCGCATTTCTATTTCCAGTATTTTCCCACCAT

GCTGAAAGAACGGGCGGCCTGGGCCTTTGATTTTTCCGGACG

CAATCGCCGCCCGCCGCGCGACCCGGTCAACGCCCTGCTTTC

GTTCGGTTACAGCGTGTTGTCCAAGGAACTTGCCGGCGTCTG

CCACGCTGTTGGCCTAGACCCGTTTTTCGGCTTCATGCACCA

GCCGCGTTACGGGCGCCCCGCACTCGCTCTCGATCTGATGGA

GGAGTTTCGCCCTCTCATCGCCGACAGTGTTGCCCTGAATCT

CATCAACCGTGGCGAACTCGACGAAGGGGACTTTATCCGGTC

GGCCAATGGCACCGCGCTCAATGATCGGGGCCGCCGGCGTTT

TTGGGAGGCATGGTTCCGGCGTCTCGACAGCGAAGTCAGCC

ATCCTGAATTTGGTTACAAGATGAGCTATCGACGGATGCTTGA

AGTGCAGGCGCGCCAGCTATGGCGCTATGTGCGCGGTGACGC

CTTCCGCTACCACGGATTCACCACCCGTTGATTCCGATGTCA

GATCCCCGCCGCCGTTATCTTGTGTGTTACGACATCGCCAATC

CGAAGCGATTGCGCCAAGTGGCCAAGCTGCTGGAGAGCTAT

GGCACGCGTCTGCAATACTCGGTTTTCGAATGTCCTTTGGAC

GATCTTCGTCTTGAACAGGCGAAGGCTGATTTGCGCGACACG

ATTAATGCCGACCAAGACCAGGTGTTATTTGTTTCGCTTGGCC

CCGAAGCCAACGATGCCACGTTGATCATCGCCACGCTTGGGC

TCCCTTATACCGTGCGCTCGCGAGTGACGATTATCTGACCCAT

AACCCACGTGTTGAAGAGGCTGAAAACAGACGGACCTCTAT

GAAGAACAATTGACGTTTTGGCCGAACTCAGCAGACCTTTAT

GCGGCTAAGGCCAATGATCATCCATCCTACCGCCATTGGGCT

GGAGACGTTTTTTGAAACGGCGAGTGCTGCGGATAGCGAGTT

TCTCTTGGGGAGGCGCTCGCGGCCACTTTTACAGAGGAGATG

TTCGGGCGAACTGGCCGACCTAACAAGGCGTACCCGGCTCA

AAATCGAGGCACGCTCGCACGGGATGATGTAATTCGTTGTTT

TTCAGCATACCGTGCGAGCACGGGCCGCAGCGAATGCCGTTT

CACGAATCGTCAGGCGGCGGGGAGAAGTCATTTAATAAGGCC

ACTGTTAAAAGCCGCAGCGAATGCCGTTTCACGAATCGTCAG

GCGGGCAGTGGATGTTTTTCCATGAGGCGAAGAATTTCATCG

CCGCAGTGAATGCCGTTTCACCATTGATGAAGAATGCGAGGT

GAAAACAGAGAAATTGGGTCAACTCTATCACTCTTATTCAGC
CATCGTTTCAAGAAAGGATACCTCGTATTGGATACAACACAG
CTCGTTCGTTCTCTCTACCTCCCTCGACAATCTCAAGGA（SEQ
ID NO: 61）

[1426] >B-座位

[1427]

TAATAAAATTGAAATATCACTATGGATTATTGTAATATTAC
CATAAAGATAGGTGACGTTTTTTTGAAAATTGTAAACCTAATT
TGAAGAAAACCAATTAAAAATCGCTTCGGCTTTTTTTTAAGT
GCCAGGTAGCATTGATGCTAACCCATGTGTAATAAAGGTTTGT
TTTCCTTCGGGGCACGAACACATTATAAGGGAAACCTAAAGA
TTCCCTTTCTTGTTTAATATTATAACCAGTGAAAATAAGAATA
ATGCACCTAAAACTAATATACAGAAAATAAGAATTAAAAGTA
CTAATATATACATCATATGTTATCCTCCAATGCTTTATTTTTTAA
TAATTGATGTTAGTATTAGTTTTATTTTAATTTCTAAACATAAG
AATTTGAAAAGGATGTGTTTATTATGGCGACACGCAGTTTTAT
TTTAAAAATTGAACCAAATGAAGAAGTTAAAAAGGGATTATG
GAAGACGCATGAGGTATTGAATCATGGAATTGCCTACTACAT
GAATATTCTGAAACTAATTAGACAGGAAGCTATTTATGAACAT
CATGAACAAGATCCTAAAAATCCGAAAAAAGTTTCAAAAGC
AGAAATACAAGCCGAGTTATGGGATTTTGTTTTAAAAATGCA
AAAATGTAATAGTTTTACACATGAAGTTGACAAAGATGTTGT
TTTTAACATCCTGCGTGAACTATATGAAGAGTTGGTCCCTAGT

TCAGTCGAGAAAAAGGGTGAAGCCAATCAATTATCGAATAAG

TTTCTGTACCCGCTAGTTGATCCGAACAGTCAAAGTGGGAAA

GGGACGGCATCATCCGGACGTAAACCTCGGTGGTATAATTTA

AAAATAGCAGGCGACCCATCGTGGGAGGAAGAAAAGAAAAA

ATGGGAAGAGGATAAAAAGAAAGATCCCCTTGCTAAAATCTT

AGGTAAGTTAGCAGAATATGGGCTTATTCCGCTATTTATTCCA

TTTACTGACAGCAACGAACCAATTGTAAAAGAAATTAAATGG

ATGGAAAAAAGTCGTAATCAAAGTGTCCGGCGACTTGATAAG

GATATGTTTATCCAAGCATTAGAGCGTTTTCTTTCATGGGAAA

GCTGGAACCTTAAAGTAAAGGAAGAGTATGAAAAAGTTGAA

AAGGAACACAAAACACTAGAGGAAAGGATAAAAGAGGACAT

TCAAGCATTTAAATCCCTTGAACAATATGAAAAAGAACGGCA

GGAGCAACTTCTTAGAGATACATTGAATACAAATGAATACCG

ATTAAGCAAAAGAGGATTACGTGGTTGGCGTGAAATTATCCA

AAAATGGCTAAAGATGGATGAAAATGAACCATCAGAAAAATA

TTTAGAAGTATTTAAAGATTATCAACGGAAACATCCACGAGA

AGCCGGGGACTATTCTGTCTATGAATTTTTAAGCAAGAAAGA

AAATCATTTTATTTGGCGAAATCATCCTGAATATCCTTATTTGT

ATGCTACATTTTGTGAAATTGACAAAAAAAAAGAAAGACGCTA

AGCAACAGGCAACTTTTACTTTGGCTGACCCGATTAACCATC

CGTTATGGGTACGATTTGAAGAAAGAAGCGGTTCGAACTTAA

ACAAATATCGAATTTTAACAGAGCAATTACACACTGAAAAGT

TAAAAAAGAAATTAACAGTTCAACTTGATCGTTTAATTTATCC

AACTGAATCCGGCGGTTGGGAGGAAAAAGGTAAAGTAGATA

TCGTTTTGTTGCCGTCAAGACAATTTTATAATCAAATCTTCCT

TGATATAGAAGAAAAGGGGAAACATGCTTTTACTTATAAGGA

TGAAAGTATTAAATTCCCCCTTAAAGGTACACTTGGTGGTGC

AAGAGTGCAGTTTGACCGTGACCATTTGCGGAGATATCCGCA

TAAAGTAGAATCAGGAAATGTTGGACGGATTTATTTTAACATG

ACAGTAAATATTGAACCAACTGAGAGCCCTGTTAGTAAGTCT

TTGAAAATACATAGGGACGATTTCCCCAAGTTCGTTAATTTTA

AACCGAAAGAGCTCACCGAATGGATAAAAGATAGTAAAGGG

AAAAAATTAAAAAGTGGTATAGAATCCCTTGAAATTGGTCTA

CGGGTGATGAGTATCGACTTAGGTCAACGTCAAGCGGCTGCT

GCATCGATTTTTGAAGTAGTTGATCAGAAACCGGATATTGAA

GGGAAGTTATTTTTTCCAATCAAAGGAACTGAGCTTTATGCT

GTTCACCGGGCAAGTTTTAACATTAAATTACCGGGTGAAACA

TTAGTAAAATCACGGGAAGTATTGCGGAAAGCTCGGGAGGA

CAACTTAAAATTAATGAATCAAAAGTTAAACTTTCTAAGAAA

TGTTCTACATTTCCAACAGTTTGAAGATATCACAGAAAGAGA

GAAGCGTGTAACTAAATGGATTTCTAGACAAGAAAATAGTGA

TGTTCCTCTTGTATATCAAGATGAGCTAATTCAAATTCGTGAA

TTAATGTATAAACCCTATAAAGATTGGGTTGCCTTTTTAAAAC

AACTCCATAAACGGCTAGAAGTCGAGATTGGCAAAGAGGTTA

AGCATTGGCGAAAATCATTAAGTGACGGGAGAAAAGGTCTTT

ACGGAATCTCCCTAAAAAATATTGATGAAATTGATCGAACAA

GGAAATTCCTTTTAAGATGGAGCTTACGTCCAACAGAACCTG

GGGAAGTAAGACGCTTGGAACCAGGACAGCGTTTTGCGATT

GATCAATTAAACCACCTAAATGCATTAAAAGAAGATCGATTA

AAAAAGATGGCAAATACGATTATCATGCATGCCTTAGGTTACT

GTTATGATGTAAGAAAGAAAAAGTGGCAGGCAAAAAATCCA

GCATGTCAAATTATTTTATTTGAAGATTTATCTAACTACAATCC

TTACGAGGAAAGGTCCCGTTTTGAAAACTCAAAACTGATGA

AGTGGTCACGGAGAGAAATTCCACGACAAGTCGCCTTACAA

GGTGAAATTTACGGATTACAAGTTGGGGAAGTAGGTGCCCAA

TTCAGTTCAAGATTCCATGCGAAAACCGGGTCGCCGGGAATT

CGTTGCAGTGTTGTAACGAAAGAAAAATTGCAGGATAATCGC

TTTTTTAAAAATTTACAAAGAGAAGGACGACTTACTCTTGAT

AAAATCGCAGTTTTAAAAGAAGGAGACTTATATCCAGATAAA

GGTGGAGAAAAGTTTATTTCTTTATCAAAGGATCGAAAGTTG

GTAACTACGCATGCTGATATTAACGCGGCCCAAAATTTACAG

AAGCGTTTTTGGACAAGAACACATGGATTTTATAAAGTTTAC

TGCAAAGCCTATCAGGTTGATGGACAAACTGTTTATATTCCGG

AGAGCAAGGACCAAAAACAAAAAATAATTGAAGAATTTGGG

GAAGGCTATTTTATTTTAAAAGATGGTGTATATGAATGGGGTA

ATGCGGGGAAACTAAAAATTAAAAAAGGTTCCTCTAAACAAT

CATCGAGTGAATTAGTAGATTCGGACATACTGAAAGATTCATT

TGATTTAGCAAGTGAACTTAAGGGAGAGAAACTCATGTTATA

TCGAGATCCGAGTGGAAACGTATTTCCTTCCGACAAGTGGAT

GGCAGCAGGAGTATTTTTTGGCAAATTAGAAAGAATATTGAT

TTCTAAGTTAACAAATCAATACTCAATATCAACAATAGAAGAT

GATTCTTCAAAACAATCAATGTAAAAGTTTGCCCGTATAAGA

ACTTAATTAATTAGGATGGTAGGATGTTACTAAATATGTCTGTA

GGCATCATTCCTACTATCCGTTTTGTCCGAATATCAGAGCATT

AGGTGAGGAATGGTAAGAAAGGAAAATTTATATGAACCAACC

GATTCCTATTCGAATGTTAAATGAAATACAATATTGTGAGCGA

CTTTTTTACTTTATGCATGTCCAAAAGCTATTTGATGAGAATG

CAGATACAGTTGAAGGAAGTGCACAGCATGAGCGGGCAGAA

AGAAGCAAAAGACCAAGTAAAATGGGACCAAAGGAATTATG

GGGTGAGGCGCCAAGAAGTCTTAAGCTTGGTGATGAGCTGTT

AAATATTACCGGTGTTCTTGATGCCATAAGTCATGAAGAGAA

CAGTTGGATCCCGGTTGAATCAAAACACAGTTCCGCACCGGA

TGGATTGAACCCTTTTAAAGTAGATGGCTTTCTACTTGACGG

GTCTGCATGGCCAAACGATCAAATTCAACTTTGTGCACAAGG

CTTGCTCTTGAATGCCAATGGATACCCGTGTGATTATGGGTAT

TTATTTTATCGTGGTAATAAGAAAAAGGTGAAAATTTATTTTA

CTGAAGATTTAATCGCTGCCACAAAGTACTATATTAAAAAAG

CACACGAGATACTAGTATTATCTGGTGATGAATCAGCTATTCC

TAAGCCTTTAATTGATTCTAATAAGTGTTTTCGCTGTTCTTTAA

ACTATATCTGTCTTCCGGATGAAACGAACTATCTATTAGGGGC

AAGTTCAACAATTCGTAAAATTGTGCCTTCAAGGACAGATGG

TGGCGTTTTATATGTATCAGAGTCTGGTACAAAATTAGGAAAA

TCGGGTGAGGAGTTAATCATTCAGTATAAAGATGGCCAAAAG

CAGGGTGTTCCTATAAAAGATATTATTCAAGTTTCGTTAATTG

GAAATGTTCAATGCTCAACGCAATTACTTCATTTTTTAATGCA

ATCAAATATTCCTGTAAGTTATTTATCATCCCACGGTCGTTTGA

TTGGTGTCAGTTCATCTTTAGTTACAAAAAATGTTTTAACAAG

GCAGCAACAGTTCATTAAATTTACAAATCCTGAGTTTGGACT

AAATCTAGCAAAACAAATTGTTTATGCCAAGATTCGAAATCA

ACGAACTTTACTTAGAAGAAATGGGGGGGAGTGAGGTAAAGG

AGATTTTAACAGATTTAAAATCTTTAAGTGACAGTGCACTGA

ACGCAATATCAATAGAACAATTACGGGGTATTGAAGGGATTT

CTGCAAAACATTATTTCGCAGGATTTCCGTTTATGTTGAAAAA

TGAATTACGTGAATTGAATTTAATGAAAGGGCGTAATAGGAG

ACCGCCAAAAGATCCTGTAAATGTACTTCTTTCTCTTGGTTAT

ACTTTATTGACACGTGATATTCATGCTGCGTGTGGTTCAGTCG

GATTGGATCCGATGTTTGGTTGTTACCATCGTCCAGAAGCAG

GTCGACCGGCTCTAGTATTAGATGTTATGGAAACATTTCGACC

ACTTATTGTAGACAGTATTGTCATCCGAGCTTTGAATACGGGT

GAAATCTCATTAAAAGATTTTTATATAGGAAAAGATAGTTGTC

AATTATTAAAACATGGCCGCGATTCCTTTTTTGCCATTTATGA

AAGAAGAATGCATGAAACTATTACCGATCCAATTTTCGGCTAT

AAGATTAGCTATCGCCGTATGCTCGATTTGCACATTCGAATGC

TTGCAAGGTTTATTGAAGGGGAACTGCCGGAATATAAACCAT

TAATGACCCGGTGAGTTTGTTTATTAGGTTAAAAGAAGGTGA

AGACATGCAGCAATACGTCCTTGTTTCTTATGATATTTCGGAC

CAAAAAAGATGGAGAAAAGTATTTAAACTGATGAAAGGATAC

GGAGAACATGTTCAATATTCCGTATTCATATGCCAGTTAACTG

AATTACAGAAGGCAAAATTACAAGCCTCTTTAGAAGACATTA

TCCATCATAAGAATGACCAAGTAATGTTTGTTCACATCGGGCC

AGTGAAAGATGGTCAACTATCTAAAAAAATCTCAACAATTGG
GAAAGAATTTGTTCCATTGGATTTAAAGCGGCTTATATTTTGA
AAAGATATAGCAAAGAAATCTTATGAAAAAAATACAAAAATA
TATTGTTAAAAAAATAGGGAATATTATATAATGGACTTACGAGG
TTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGC
AGGGGTGTGAGAAACTCCTATTGCTGGACGATGTCTCTTTTAT
TTCTTTTTTCTTGGATCTGAGTACGAGCACCCACATTGGACAT
TTCGCATGGTGGGTGCTCGTACTATAGGTAAAACAAACCTTT
TTAAGAAGAATACAAAAATAACCACAATATTTTTTAAAAGGA
ATTTTGATGGATTTACATAACCTCTCGCAACATGCTTCTAAAA
CCCAAGCCCACCATAGCCCAAAACCCCCTGCGGTCCAAGAA
AAAAGAAATGATACGAGGCATTAGCACCGGGGAGAAGTCAT
TTAATAAGGCCACTGTTAAAAGTCCAAGAAAAAAGAAATGAT
ACGAGGCATTAGCACAACAATATAAACGACTACTTTACCGTG
TTCAAGAAAAAAGAAATGATATGAGGCATTAGCACGATGGGA
TGGGAGAGAGAGGACAGTTCTACTCTTGCTGTATCCAGCTTC
TTTTACTTTATCCGGTATCATTTCTTCACTTCTTTCTGCACATA
AAAAAGCACCTAACTATTTGGATAAGTTAAGTGCTTTTATTTC
CGTTTGAAGTTGTCTATTGCTTTTTTCTTCATATCTTCAAATTT
TTTCTGTTTCTCAGAGTCAACTTTACCAACTGTAATCCCTTTT
CTTTTTGGCATTGGGGTATCTTTCCACCTTAGTGTGTTCATAA
GGCTTATATTTATCACTCATTGTATTCCTCCAACACAATTATAA
TTTTTCCGTCATCCTCAATCCAACCGTCAACTGTGACAAAAG
ACGAATCTCTCTTAT (SEQ ID NO: 62)

[1428] >C-座位

[1429]

GTTTCATTTGGAAAGGGAGAGCATTGGCTTTTCTCTTTGT
AAATAAAGTGCAAGCTTTGTAATAAGCTTCTAGTGGAGAAGT
GATTGTTTGAATCACCCAATGCACACGCACTAAAGTTAGACG
AACCTATAATTCGTATTAGTAAGTATAGTACATGAAGAAAAAT
GCAACAAGCATTTACTCTCTTTTAAATAAAGAATTGATAGCTG
TTAATATTGATAGTATATTATACCTTATAGATGTTCGATTTTTTT
TGAAATTCAAAAATCATACTTAGTAAAGAAAGGAAATAACGT
CATGGACAAGCGAAAGCGTAGAAGTTACGAGTTTAGGTGGG
AAGCGGGAGGCACCAGTCATGGCAATCCGTAGCATAAAACTA
AAACTAAAAACCCACACAGGCCCGGAAGCGCAAAACCTCCG
AAAAGGAATATGGCGGACGCATCGGTTGTTAAATGAAGGCGT
CGCCTATTACATGAAAATGCTCCTGCTCTTTCGTCAGGAAAG
CACTGGTGAACGGCCAAAAGAAGAACTACAGGAAGAACTGA
TTTGTCACATACGCGAACAGCAACAACGAAATCAGGCAGATA
AAAATACGCAAGCGCTTCCGCTAGATAAGGCACTGGAAGCTT
TGCGCCAACTATATGAACTGCTTGTCCCCTCCTCGGTCGGAC
AAAGTGGCGACGCCCAGATCATCAGCCGAAAGTTTCTCAGC
CCGCTCGTCGATCCGAACAGCGAAGGCGGCAAAGGTACTTC
GAAGGCAGGGGCAAAACCCACTTGGCAGAAGAAAAAAGAA
GCGAACGACCCAACCTGGGAACAGGATTACGAAAAATGGAA
AAAAAGACGCGAGGAAGACCCAACCGCTTCTGTGATTACTA
CTTTGGAGGAATACGGCATTAGACCGATCTTTCCCCTGTACAC

GAACACCGTAACAGATATCGCGTGGTTGCCACTTCAATCCAA

TCAGTTTGTGCGAACCTGGGACAGAGACATGCTTCAACAAG

CGATTGAAAGACTGCTCAGTTGGGAGAGCTGGAACAAACGT

GTCCAGGAAGAGTATGCCAAGCTGAAAGAAAAAATGGCTCA

ACTGAACGAGCAACTCGAAGGCGGTCAGGAATGGATCAGCT

TGCTAGAGCAGTACGAAGAAAACCGAGAGCGAGAGCTTAGG

GAAAACATGACCGCTGCCAATGACAAGTATCGGATTACCAAG

CGGCAAATGAAAGGCTGGAACGAGCTGTACGAGCTATGGTC

AACCTTTCCCGCCAGTGCCAGTCACGAGCAATACAAAGAGG

CGCTCAAGCGTGTGCAGCAGCGACTGAGAGGGCGGTTTGGG

GATGCTCATTTCTTCCAGTATCTGATGGAAGAGAAGAACCGC

CTGATCTGGAAGGGGAATCCGCAGCGTATCCATTATTTTGTCG

CGCGCAACGAACTGACGAAACGGCTGGAGGAAGCCAAGCA

AAGCGCCACGATGACGTTGCCCAATGCCAGGAAGCATCCATT

GTGGGTGCGCTTCGATGCACGGGGAGGAAATTTGCAAGACT

ACTACTTGACGGCTGAAGCGGACAAACCGAGAAGCAGACGT

TTTGTAACGTTTAGTCAGTTGATATGGCCAAGCGAATCGGGAT

GGATGGAAAAGAAAGACGTCGAGGTCGAGCTAGCTTTGTCC

AGGCAGTTTTACCAGCAGGTGAAGTTGCTGAAAAATGACAA

AGGCAAGCAGAAAATCGAGTTCAAGGATAAAGGTTCGGGCT

CGACGTTTAACGGACACTTGGGGGGAGCAAAGCTACAACTG

GAGCGGGGCGATTTGGAGAAGGAAGAAAAAAACTTCGAGG

ACGGGGAAATCGGCAGCGTTTACCTTAACGTTGTCATTGATT

TCGAACCTTTGCAAGAAGTGAAAAATGGCCGCGTGCAGGCG

CCGTATGGACAAGTACTGCAACTCATTCGTCGCCCCAACGAG

TTTCCCAAGGTCACTACCTATAAGTCGGAGCAACTTGTTGAA

TGGATAAAAGCTTCGCCACAACACTCGGCTGGGGTGGAGTC

GCTGGCATCCGGTTTTCGTGTAATGAGCATAGACCTTGGGCT

GCGCGCGGCTGCAGCGACTTCTATTTTTTCTGTAGAAGAGAG

TAGCGATAAAAATGCGGCTGATTTTTCCTACTGGATTGAAGG

AACGCCGCTGGTCGCTGTCCATCAGCGGAGCTATATGCTCAG

GTTGCCTGGTGAACAGGTAGAAAAACAGGTGATGGAAAAAC

GGGACGAGCGGTTCCAGCTACACCAACGTGTGAAGTTTCAA

ATCAGAGTGCTCGCCCAAATCATGCGTATGGCAAATAAGCAG

TATGGAGATCGCTGGGATGAACTCGACAGCCTGAAACAAGC

GGTTGAGCAGAAAAAGTCGCCGCTCGATCAAACAGACCGGA

CATTTTGGGAGGGGATTGTCTGCGACTTAACAAAGGTTTTGC

CTCGAAACGAAGCGGACTGGGAACAAGCGGTAGTGCAAATA

CACCGAAAAGCAGAGGAATACGTCGGAAAAGCCGTTCAGGC

ATGGCGCAAGCGCTTTGCTGCTGACGAGCGAAAAGGCATCG

CAGGTCTGAGCATGTGGAACATAGAAGAATTGGAGGGCTTGC

GCAAGCTGTTGATTTCCTGGAGCCGCAGGACGAGGAATCCG

CAGGAGGTTAATCGCTTTGAGCGAGGCCATACCAGCCACCAG

CGTCTGTTGACCCATATCCAAAACGTCAAAGAGGATCGCCTG

AAGCAGTTAAGTCACGCCATTGTCATGACTGCCTTGGGGTAT

GTTTACGACGAGCGGAAACAAGAGTGGTGCGCCGAATACCC

GGCTTGCCAGGTCATTCTGTTTGAAAATCTGAGCCAGTACCG

TTCTAACCTGGATCGCTCGACCAAAGAAAACTCCACCTTGAT

GAAGTGGGCGCATCGCAGCATTCCGAAATACGTCCACATGCA

GGCGGAGCCATACGGGATTCAGATTGGCGATGTCCGGGCGGA

ATATTCCTCTCGTTTTTACGCCAAGACAGGAACGCCAGGCAT

TCGTTGTAAAAAGGTGAGAGGCCAAGACCTGCAGGGCAGAC

GGTTTGAGAACTTGCAGAAGAGGTTAGTCAACGAGCAATTTT

TGACGGAAGAACAAGTGAAACAGCTAAGGCCCGGCGACATT

GTCCCGGATGATAGCGGAGAACTGTTCATGACCTTGACAGAC

GGAAGCGGAAGCAAGGAGGTCGTGTTTCTCCAGGCCGATATT

AACGCGGCGCACAATCTGCAAAAACGTTTTTGGCAGCGATAC

AATGAACTGTTCAAGGTTAGCTGCCGCGTCATCGTCCGAGAC

GAGGAAGAGTATCTCGTTCCCAAGACAAAATCGGTGCAGGC

AAAGCTGGGCAAAGGGCTTTTTGTGAAAAAATCGGATACAG

CCTGGAAAGATGTATATGTGTGGGACAGCCAGGCAAAGCTTA

AAGGTAAAACAACCTTTACAGAAGAGTCTGAGTCGCCCGAA

CAACTGGAAGACTTTCAGGAGATCATCGAGGAAGCAGAAGA

GGCGAAAGGAACATACCGTACACTGTTCCGCGATCCTAGCGG

AGTCTTTTTTCCCGAATCCGTATGGTATCCCCAAAAAGATTTT

TGGGGCGAGGTGAAAAGGAAGCTGTACGGAAAATTGCGGGA

ACGGTTTTTGACAAAGGCTCGGTAAGGGTGTGCAAGGAGAG

TGAATGGCTTGTCCTGGATACCTGTCCGCATGCTAAATGAAAT

TCAGTATTGTGAGCGACTGTACCATATTATGCATGTGCAGGGG

CTGTTTGAGGAAAGCGCAGACACGGTCGAAGGAGCAGCACA

ACACAAGCGTGCAGAGACACATCTGCGCAAAAGCAAGGCAG

CGCCGGAAGAGATGTGGGGGGACGCTCCGTTTAGCTTGCAG

CTCGGCGACCCTGTGCTTGGCATTACGGGAAAGCTGGATGCC

GTCTGTCTGGAAGAAGGTAAGCAGTGGATTCCGGTAGAAGG

AAAGCATTCGGCGTCGCCAGAAGGCGGGCAGATGTTCACTG

TAGGCGTGTATTCGCTGGACGGTTCTGCCTGGCCCAACGACC

AAATCCAATTGTGTGCGCAAGGCTTGCTGCTTCGCGCGAATG

GATATGAATCCGATTATGGCTACTTATACTACCGTGGCAATAA

AAAGAAGGTTCGCATTCCTTTTTCGCAGGAACTCATAGCGGC

TACTCACGCCTGCATTCAAAAAGCTCATCAGCTTCGGGAAGC

CGAAATTCCCCCTCCGTTGCAGGAGTCGAAAAAGTGCTTTCG

ATGCTCGTTAAATTACGTATGCATGCCTGACGAGACGAATTAC

ATGTTGGGGTTGAGCGCAAACATCAGAAAGATTGTGCCCAGT

CGTCCAGATGGCGGGGTACTGTATGTTACAGAGCAGGGGGCA

AAACTGGGCAGAAGCGGAGAAAGCTTGACCATCACCTGCCG

GGGCGAAAAGATAGACGAAATCCCGATCAAAGACTTGATTCA

CGTGAGCTTGATGGGGCATGTGCAATGCTCTACGCAGCTTCT

GCACACCTTGATGAACTGTGGCGTCCACGTCAGCTACTTGAC

TACGCATGGCACATTGACAGGAATAATGACTCCCCCTTTATCG

AAAAACATTCGAACAAGAGCCAAGCAGTTTATCAAATTTCAG

CACGCGGAGATCGCCCTTGGAATCGCGAGAAGGGTCGTGTAT

GCGAAAATTTCCAATCAGCGCACGATGCTGCGCCGCAATGGC

TCACCAGATAAAGCAGTTTTAAAAGAGTTAAAAGAGCTTAGA

GATCGCGCGTGGGAGGCGCCATCACTGGAAATAGTGAGAGG

TATCGAGGGACGTGCAGCACAGTTGTACATGCAGTTTTTCCC

TACCATGTTAAAGCACCCAGTAGTAGACGGTATGGCGATCAT

GAACGGTCGCAACCGTCGCCCGCCCAAAGATCCGGTCAATG

CGCTGCTCTCCCTCGGCTATACGCTTCTTTCACGGGATGTTTA

CTCCGCATGTGCCAATGTCGGACTCGATCCACTGTTCGGCTTT

TTCCATACGATGGAGCCGGGCAGACCAGCTTTGGCACTCGAT

CTGATGGAACCGTTCCGCGCCTTGATTGCCGATAGCGTAGCG

ATACGTACCTTGAATACGGAGGAACTCACCCTCGGGGACTTT

TATTGGGGAAAAGACAGTTGTTATTTGAAAAAGGCAGGAAG

ACAAACGTATTTCGCTGCCTATGAAAGACGGATGAACGAGAC

GCTGACGCATCCGCAATTTGGGTATAAGCTCAGCTATCGCCGT

ATGCTGGAGCTGGAAGCAAGGTTTTTGGCCCGGTATCTGGAT

GGAGAGCTGGTGGAATATACGCCGCTCATGACAAGGTAGGAA

ATGACCATGCGACAATTTGTTCTGGTAAGCTATGATATTGCCG

ATCAAAAACGTTGGAGAAAAGTATTCAAGCTGATGAAGGGG

CAAGGCGAGCACGTCCAGTACTCGGTGTTTCTGTGCCAACTC

ACCGAGATTCAGCAAGCCAAGCTAAAGGTAAGCCTGGCGGA

GCTGGTTCACCATGGAGAAGACCAGGTCATGTTTGTAAAAAT

CGGCCCAGTGACGAGAGATCAACTGGACAAGCGGATATCTAC

TGTTGGCAGGGAGTTTCTGCCTCGCGATTTGACCAAATTTATC

TATTAAGGAATGAAGAAAGCTAGTTGTAACAAAAGTGGAAA

AAGAGTAAAATAAAGGTGTCAGTCGCACGCTATAGGCCATAA

GTCGACTTACATATCCGTGCGTGTGCATTATGGGCCCATCCAC

AGGTCTATTCCCACGGATAATCACGACTTTCCACTAAGCTTTC

GAATTTTATGATGCGAGCATCCTCTCAGGTCAAAAAAGCCGG

GGGATGCTCGAACTCTTTGTGGGCGTAGGCTTTCCAGAGTTT

TTTAGGGGAAGAGGCAGCCGATGGATAAGAGGAATGGCGAT
TGAATTTTGGCTTGCTCGAAAAACGGGTCTGTAAGGCTTGCG
GCTGTAGGGGTTGAGTGGGAAGGAGTTCGAAAGCTTAGTGG
AAAGCTTCGTGGTTAGCACCGGGGAGAAGTCATTTAATAAGG
CCACTGTTAAAAGTTCGAAAGCTTAGTGGAAAGCTTCGTGGT
TAGCACGCTAAAGTCCGTCTAAACTACTGAGATCTTAAATCG
GCGCTCAAATAAAAAACCTCGCTAATGCGAGGTTTCAGC
（SEQ ID NO: 63）

[1430] >D-座位

[1431]

GAAGTTATGTTGATAAAATGGTTTATGAAAACGTGAGTCT
GTGGTAGTATTATAAACAATGATGGAATAAAGTGTTTTTTGCG
CCGCACGGCATGAATTCAGGGGTTAGCTTGGTTTTGTGTATA
AATAAATGTTCTACATATTTATTTTGTTTTTTGCGCCGCAAAAT
GCAACTGAAAGCCGCATCTAGAGCACCCTGTAGAAGACAGG
GTTTTGAGAATAGCCCGACATAGAGGGCAATAGACACGGGGA
GAAGTCATTTAATAAGGCCACTGTTAAAAGTTTTGAGAATAG
CCCGACATAGAGGGCAATAGACTTTTGCTTCGTCACGGATGG
ACTTCACAATGGCAACAACGTTTTGAGAATAGCCCGACATAG
TTATAGAGATGTATAAATATAACCGATAAACATTGACTAATTTG
TTGAAGTCAGTGTTTATCGGTTTTTTGTGTAAATATAGGAGTT
GTTAGAATGATACTTTTTGCCTAATTTTGGAACTTTATGAGGA
TATAAGATAGACTTGATAAAAAGGTAAAAGAAAGGTTAAAGA

GCATGGCAGGAATAGTGACCTGTGATGAAGATGATGGTAGAA

TTAAAAGTGTTCTTAAAGAAAAACAATATTGGATAAGGAAAA

TAATTCAATAGATAAAAAATTTAGGGGGAAAAATGAAAATAT

CAAAAGTCGATCATACCAGAATGGCGGTTGCTAAAGGTAATC

AACACAGGAGAGATGAGATTAGTGGGATTCTCTATAAGGATC

CGACAAAGACAGGAAGTATAGATTTTGATGAACGATTCAAAA

AACTGAATTGTTCGGCGAAGATACTTTATCATGTATTCAATGG

AATTGCTGAGGGAAGCAATAAATACAAAAATATTGTTGATAA

AGTAAATAACAATTTAGATAGGGTCTTATTTACAGGTAAGAGC

TATGATCGAAAATCTATCATAGACATAGATACTGTTCTTAGAA

ATGTTGAGAAAATTAATGCATTTGATCGAATTTCAACAGAGG

AAAGAGAACAAATAATTGACGATTTGTTAGAAATACAATTGA

GGAAGGGGTTAAGGAAAGGAAAAGCTGGATTAAGAGAGGTA

TTACTAATTGGTGCTGGTGTAATAGTTAGAACCGATAAGAAG

CAGGAAATAGCTGATTTTCTGGAGATTTTAGATGAAGATTTCA

ATAAGACGAATCAGGCTAAGAACATAAAATTGTCTATTGAGA

ATCAGGGGTTGGTGGTCTCGCCTGTATCAAGGGGAGAGGAA

CGGATTTTTGATGTCAGTGGCGCACAAAAGGGAAAAAGCAG

CAAAAAAGCGCAGGAGAAAGAGGCACTATCTGCATTTCTGTT

AGATTATGCTGATCTTGATAAGAATGTCAGGTTTGAGTATTTA

CGTAAAATTAGAAGACTGATAAATCTATATTTCTATGTCAAAA

ATGATGATGTTATGTCTTTAACTGAAATTCCGGCAGAAGTGAA

TCTGGAAAAAGATTTTGATATCTGGAGAGATCACGAACAAAG

AAAGGAAGAGAATGGAGATTTTGTTGGATGTCCGGACATACT

TTTGGCAGATCGTGATGTGAAGAAAAGTAACAGTAAGCAGG

TAAAAATTGCAGAGAGGCAATTAAGGGAGTCAATACGTGAA

AAAAATATAAAACGATATAGATTTAGCATAAAAACGATTGAA

AAGGATGATGGAACATACTTTTTTGCAAATAAGCAGATAAGT

GTATTTTGGATTCATCGCATTGAAAATGCTGTAGAACGTATAT

TAGGATCTATTAATGATAAAAAACTGTATAGATTACGTTTAGG

ATATCTAGGAGAAAAAGTATGGAAGGACATACTCAATTTTCT

CAGCATAAAATACATTGCAGTAGGCAAGGCAGTATTCAATTTT

GCAATGGATGATCTGCAGGAGAAGGATAGAGATATAGAACCC

GGCAAGATATCAGAAAATGCAGTAAATGGATTGACTTCGTTT

GATTATGAGCAAATAAAGGCAGATGAGATGCTGCAGAGAGA

AGTTGCTGTTAATGTAGCATTCGCAGCAAATAATCTTGCTAGA

GTAACTGTAGATATTCCGCAAAATGGAGAAAAAGAGGATATC

CTTCTTTGGAATAAAAGTGACATAAAAAAAATACAAAAAGAAT

TCAAAGAAAGGTATTCTGAAATCTATACTTCAGTTTTTTGGTG

GTGCTTCAACTTGGAATATGAAAATGTTTGAGATTGCATATCA

TGATCAGCCAGGTGATTACGAAGAAAACTACCTATATGACAT

TATTCAGATCATTTACTCGCTCAGAAATAAGAGCTTTCATTTC

AAGACATATGATCATGGGGATAAGAATTGGAATAGAGAACTG

ATAGGAAAGATGATTGAGCATGATGCTGAAAGAGTCATTTCT

GTTGAGAGGGAAAAGTTTCATTCCAATAACCTGCCGATGTTT

TATAAAGACGCTGATCTAAAGAAAATATTGGATCTCTTGTATA

GCGATTATGCAGGACGTGCATCTCAGGTTCCGGCATTTAACA

CTGTCTTGGTTCGAAAGAACTTTCCGGAATTTCTTAGGAAAG

ATATGGGCTACAAGGTTCATTTTAACAATCCTGAAGTAGAGA

ATCAGTGGCACAGTGCGGTGTATTACCTATATAAAGAGATTTA

TTACAATCTATTTTTGAGAGATAAAGAGGTAAAGAATCTTTTT

TATACTTCATTAAAAAATATAAGAAGTGAAGTTTCGGACAAA

AAACAAAAGTTAGCTTCAGATGATTTTGCATCCAGGTGTGAA

GAAATAGAGGATAGAAGTCTTCCGGAAATTTGTCAGATAATA

ATGACAGAATACAATGCGCAGAACTTTGGTAATAGAAAAGTT

AAATCTCAGCGTGTTATTGAAAAAAATAAGGATATTTTCAGA

CATTATAAAATGCTTTTGATAAAGACTTTAGCAGGTGCTTTTT

CTCTTTATTTGAAGCAGGAAAGATTTGCATTTATTGGTAAGGC

AACACCTATACCATACGAAACAACCGATGTTAAGAATTTTTTG

CCTGAATGGAAATCCGGAATGTATGCATCGTTTGTAGAGGAG

ATAAAGAATAATCTTGATCTTCAAGAATGGTATATCGTCGGAC

GATTCCTTAATGGGAGGATGCTCAATCAATTGGCAGGAAGCC

TGCGGTCATACATACAGTATGCGGAAGATATAGAACGTCGTG

CTGCAGAAAATAGGAATAAGCTTTTCTCCAAGCCTGATGAAA

AGATTGAAGCATGTAAAAAAGCGGTCAGAGTGCTTGATTTGT

GTATAAAAATTTCAACTAGAATATCTGCGGAATTTACTGACTA

TTTTGATAGTGAAGATGATTATGCAGATTATCTTGAAAAATAT

CTCAAGTATCAGGATGATGCCATTAAGGAATTGTCAGGATCTT

CGTATGCTGCGTTGGATCATTTTTGCAACAAGGATGATCTGAA

ATTTGATATCTATGTAAATGCCGGACAGAAGCCTATCTTACAG

AGAAATATCGTGATGGCAAAGCTTTTTGGACCAGATAACATT

TTGTCTGAAGTTATGGAAAAGGTAACAGAAAGTGCCATACGA

GAATACTATGACTATCTGAAGAAAGTTTCAGGATATCGGGTAA

GGGGAAAATGTAGTACAGAGAAAGAACAGGAAGATCTGCTA

AAGTTCCAAAGATTGAAAAACGCAGTAGAATTCCGGGATGTT

ACTGAATATGCTGAGGTTATTAATGAGCTTTTAGGACAGTTGA

TAAGTTGGTCATATCTTAGGGAGAGGGATCTATTATATTTCCA

GCTGGGATTCCATTACATGTGTCTGAAAAACAAATCTTTCAA

ACCGGCAGAATATGTGGATATTCGTAGAAATAATGGTACGATT

ATACATAATGCGATACTTTACCAGATTGTTTCGATGTATATTAA

TGGACTGGATTTCTATAGTTGTGATAAAGAAGGGAAAACGCT

CAAACCAATTGAAACAGGAAAGGGCGTAGGAAGTAAGATAG

GACAATTTATAAAGTATTCCCAGTATTTATACAATGATCCGTCA

TATAAGCTTGAGATCTATAATGCAGGATTAGAAGTTTTTGAAA

ACATTGATGAACATGATAATATTACAGATCTTAGAAAGTATGT

GGATCATTTTAAGTATTATGCATATGGTAATAAAATGAGCCTG

CTTGATCTGTATAGTGAATTCTTCGATCGTTTCTTTACATATGA

TATGAAGTATCAGAAGAATGTAGTGAATGTGTTGGAGAATAT

CCTTTTAAGGCATTTTGTAATTTTCTATCCGAAGTTTGGATCA

GGAAAAAAAGATGTTGGAATTAGGGATTGTAAAAAAGAAAG

AGCTCAGATTGAAATAAGTGAGCAGAGCCTCACATCGGAAG

ACTTCATGTTTAAGCTTGACGACAAAGCAGGAGAAGAAGCA

AAGAAGTTTCCGGCAAGGGATGAACGTTATCTCCAGACAATA

GCCAAGTTGCTCTATTATCCTAACGAAATTGAGGATATGAACA

GATTCATGAAGAAAGGAGAAACGATAAATAAAAAAGTTCAG

TTTAATAGAAAAAAGAAGATAACCAGGAAACAAAAGAATAA

TTCATCAAACGAGGTATTGTCTTCAACTATGGGTTATTTATTTA

AGAACATTAAATTGTAAAAAAGATTCGTTGTAGATAATTGATA

GGTAAAAGCTGACCGGAGCCTTTGGCTCCGGACAGTTGTATA

TAAGAGGATATTAATGACTGAAAATGATTTTTGTTGGAAGTCA

GTTTTTTCTGTGGAAAGCGAAATCGAATATGATGAGTATGCAT

ATGGCAGAAGAGCTGTAGAAGGCGAGAATACATATGATTACA

TTACTAAGGAAGAAAGACCGGAACTTAATGACGAATATGTAG

CGAGACGTTGCATTTTCGGTAAAAAAGCAGGAAAAATATCCA

GGTCGGATTTTAGTAGGATAAGATCTGCGTTGGATCATGCGAT

GATAAATAATACACATACAGCATTTGCCAGATTTATCACTGAA

AATCTGACGAGACTCAATCACAAAGAACATTTTCTGAATGTG

ACACGTGCATATTCTAAACCTGATTCTGAAAAATTGATACAAC

CGAGATACTGGCAGTCGCCTGTAGTTCCAAAGGATAAACAAA

TATATTATAGCAAGAATGCGATTAAAAAATGGTGTGGTTACGA

AGATGATATTCCGCCTCGTTCTGTGATAGTTCAGATGTGTCTA

TTGTGGGGGACTGATCATGAAGAGGCAGATCATATCCTTCGC

AGTTCAGGATACGCGGCGCTTAGTCCTGTTGTACTTCGAGAT

CTTATCTATATGTATTATCTGGATCATCAGGATTTGCAAAAAAA

TGAGTTGATATGGGAAGTAAAAAAGCAGTTGGATCACTTCGA

TTTGACAAATAGAAATTATGATACAAATCCTTTTGATGTAGGG

GGCAGCGTAAATGATCATATCTGTGAACTGAGCGAGCATATA

GCGAAGGCTCATTATATTTATGAGAGGGCTAAGGAAGGACCA

TTGCAAAATGTAATTCGGGATATTTTGGGAGATACACCTGCCC

TTTATTCTGAAATGGCATTTCCTCAGCTAGCATCTATAAACAG

GTGTGCTTGCAATTCGCTTTCTTCATATCAAAAAAATATTTTT

GATACTGACATAGCTATATATGCAGATGAAAAGGACACAAGA

GGTAAATCAGACCGTATCCTTGTTGAGGGCGCATCTTCGAAA

TGGTATGAATTGAAGAAACGCGATGCTAATAATGTCAAAATTT

CTGAAAAGCTGAGTATACTCAATACTATTCTTAAATTTAATAG

TGTTTTTTGGGAAGAATGTTACCTTGATGGAAATATAAAACA

ATCGAGCGGAAAGCGATCTGAGGCAGGAAAAATTCTTTATGG

TCGCGACAACGGAAAAGAAAATGTCGGAGTTTCAAAATTGG

AATTGGTGCGGTATATGATAGCTGCAGGTCAGGAACAAAATC

TGGGAAATTACCTGGTGAGTTCAGGATTTTGGAGAAAAAATC

ATATGCTGTCATTTATACAAGGCAATGATATAGCGCTTGATGA

GATGGATGAATTGGATCTCTTAGACTATATTCTGATATATGCAT

GGGGATTTAGGGAAAATATCATTAAAAAGAACAGTAATGTGA

ATTCTTTGGATGAAAAGACTAGAAAAGTGCAGTTTCCGTTTA

TAAAGTTACTCATGGCAATTGCAAGAGATATCCAGATACTTAT

ATGTTCAGCACATGAAAAAACAGTCGATGAGTCATCTCGAAA

TGCAGCAAAGAAGATAGATATATTGGGAAATTATATTCCTTTT

CAGATTCATCTTCAGAGAACTAAAAAAGATGGTGGAAGAGT

GGTAATGGATACATTGTGTGCTGATTGGATTGCGGATTATGAA

TGGTACATTGATCTTGAGAAAGGAACACTTGGATGAGCAGTG

ATGAAAGGATATTTAAAAAATTTTTGGAAAAAGGATCGATTT

CTGAGCAGAAAAAGATGCTTTTAGAAGAAAAGAAATGTTCG

GATAAACTAACTGCACTGCTTGGGAATTACTGCATACCGATAG

ACAATATTTCAGAGTCAGACGGAAAAATATATGCGGTCTATAA

GCTTCCAAAAAATGTTAAACCTTTGTCCGAAATCATTAATGAT

GTATCCTTTTCTGATTGTACGATGAGAGTACGTTTGCTTCTCA

TAAAGAGAATTCTGGAACTCGTGTGTGCTTTTCACGAAAAAA

AATGGTATTGTCTCAGTATTTCACCGGGAATGCTCATGGTTGA

AGATTTTGATATACCGATGGGAAATGTCGGAAAAGTATTGATA

TATGATTTCAGAAATCCTGTTCCGTTCGAGTCAGTAAATGAAA

GACATAATTTTAACGTTTCAAATAAATACACTTCACCGGAGCT

GCTCATCCATTCAAGATATGACGAGTCGAAATCTGTGAGTGA

AAAATCAGATTTGTATTCTGTTGCAAAAATTGCGGAAACAAT

AATAGGAGATTTTAACAGTATTATTGCAAATGGAAATTTGATA

CTACTTGCAATGCTTAGAGTTTTTATCAGTACAGGGAAAAGT

CCGGAACCTGAGTATCGGTTTGAATCGTCGGAAAATATGCTT

TCAGTATTTGAAAATTTGATCAAAGAAAATTGTTTTTTTGAAA

AAAACGATTATACATCTATGTTTCATCAGGCGTATGACAATTT

TTTTGAATGGCAGGAATGTTTGATATCACCGGATCACTTGGAT

AAAAATATGTTCGAGGCAGCTTTATCAAATCTTGAGGATCAG

CTGCTTAGGGTTGATATTGATAAGTATAGAGCAGAGTACTTCT

ATAAGCTTCTCCGAGAGTTGTCTAATAAATATAAAAATACAAT

TACTGATGAACAAAAGGTAAGGTTGGCAATACTTGGAATCAG

AGCGAAAAATAATCTGGGAAAAAGTTTTGATGCATTGGAAAT

ATATGAGTCAGTACGTGATTTAGAAACTATGTTGGAGGAGAT

GGCAGAGCTTAGTCCTGTCATTGCTTCGACATATATGGATTGC

TACCGATATGCAGATGCGCAGAAAGTGGCGGAAGAAAACATT

ATCAGGCTTCATAATAGTAATATTCGTATGGAGAAAAAAAGA

ATACTGCTTGGAAGGTCATATAGTTCAAAAGGGTGCAGCATG

GGGTTTCAGCATATTCTTGGTGCGGATGAGTCATTTGAACAG

GCTTTATATTTCTTTAACGAAAAGGACAATTTTTGGAAAGAA

ATATTTGAGAGCAGAAATTTAGAGGACAGCGATAGACTTATA

AAGTCTTTACGAAGCAATACGCATATTACGCTGTTTCATTACA

TGCAATATGCATGTGAAACAAGGAGAAAGGAATTATATGGAG

CACTTTCAGACAAATATTTTATAGGTAAAGAATGGACAGAAA

GACTCAAAGCATATATAAGCAACAAGGATATATGGAAAAACT

ATTATGAGATATATATTCTGCTAAAGGGTATTTATTGCTTCTAT

CCAGAAGTCATGTGTTCGTCTGCGTTTTATGATGAAATCCAA

AAAATGTACGATCTTGAATTTGAAAAGGAAAAAATGTTTTAC

CCATTGAGTCTGATAGAACTGTATCTTGCTCTGATAGAGATAA

AAGTTAATGGGAGTCTGACGGAGAATGCCGAGAAGTTGTTTA

AACAGGCATTGACACATGACAATGAAGTCAAAAAAGGAAAT

ATGAATATTCAGACCGCCATTTGGTATCGAATATATGCACTGT

ATAACGATGTAAAAGATGAAACTGATAAGAATAAAAGGCTTT

TAAAACGGCTTATGATTCTTTGCCGACGATTTGGTTGGGCGG

ATATGTATAGTGCTTTGGAGAAGGATGGGAAGTTAATTGATTT

TTTGAGATTTGAGGTATGTTAAATGATAACACTTGCATTAGAT

GAAAATGGCAAATTTGAAGATGCTTTTTCTAAAAAAAAATGAA

AAACCGATAATGATTGCGGGGATAATCTATGATGACAAGGGG

AAAGAGTATGATGCTGAGAATGAACGCTACAGGATATCCAGT

TATCTGCGAGCAGTATGTGACAGTTTGGGTGCGAAATACCCT

CAGGATCTACATTCAAATAGTAATGGAAATAAGGCGACTGTT

GGGAAAGTAAAATGTAAAATTGGTGAAACACTAAAGGAATTC

TTGAGAGAAGGAACCTATGAAAAAAAGGAATTGCCGACAAA

GAACGGTTATTTAAATAAGAGATCTGGAAAATATGTAATGTTT

GCAGAACTCAGGAGTAGTCAGGGAGTTAAAAAGCGTGTTAG

TGGTTGGAATGACAATGATCTGACTCAGGATGAAAAGGTCAG

CAATCTGTACCTTCATATGGCAGAAAATGCCGTTGTCAGAATG

CTCTTCCATAATCCTATATATGAAGATGTAACAGATGTAAATCT

CTATTTTCCCACGCGAAAAGTTGTTCTGAAAGATAGAGATAG

AGAATACGATAAACAAGATTTCAAAATATATGGTGATAAGGA

CAAGTGCGAAGCAGAAAGCGGGAGATTGGTGCATTATGATAT

CGTGTCATCGGATTTTTACCGTACGATAATGGAGAACGAATGT

ACAAGAATTAATAAAAAGCAATTAAATGTTCATTATATGAACA

CAAGCCCAATTTCGTACTGGGAGAAAAATGAAAAATATAATA

CATTTTTATATTTGGCTGACATAGTTTGTTCTATGCTGGATTAT

TACAAAAAGGGTTCGAGTCCGGCAGAGTGGATGGATTCTTTT

GCCGAATGGGGAAACAAATATTTTGGTGATGATCAGATAATCT

TATTTGGGTATGATGATATAGATGACAAATACATGGAGGCTGT

AGATGCAGTAGGACAGGGAGAGTATTTTCATGCGCTGGATAT

TATATATGATGCGGAATGTAGTGGAAGTGAATTTGAGAAGCA

CTACAAAGATTATTGGTTTCCAAAGCTTATAAAAAAGATACG

AATAACAGCAACTGTGGATAATTTATGCAGATCGATCTCAGAT

CTGGAGAGTTTTACATATCGAAGTAATCTTGATCAGCAGAAA

CTTTTGTGGATTTTTGAGGAAATCAAAGCTATCGTCGATAAGG

GAGATTTTGGAAAGAAATATCATACAGATCAGGTTATGTTTGA

TATGTGTAATGCCGGTATTGCTGTGTACAATCATATCGGAGAT

TTTGGGACTGCAAAGGAATACTATGATGAGTGCATGAAACAC

ACTGGGGATGTGGATCTGGTAAAGATACTTCGTGCATCAAAT

AAAATGGTGGTCTTTCTTGACGATGCTTTTAGGTATGGTGACG

CGACAGAACGTGCCAGGAAGAATGTTGAATACCAAAAAGCT

TTGCACGATATAAAGAGTGAGATTTGTCCGGAAAAGAAAGAT

GAAGACTTGAACTATGCCATATCGCTCAGTCAATTTGGACAG

GCGCTTGCGTGTGAAAAAAATTCTGATGCAGAGAGTGTTTTC

CTAGAGTCGTTGCGGCATATGAGGAAAGGGACTGCCAATTAT

CAGATTACTCTTTCATATTTACTCCATTTTTATCTGGATATGGG

AATGACAGATTCTTATCGAGAAAAAACAAAGGACTATTTTGG

AAGTGAAAAACCAAAGGAACAGCTGAAAGAATTGCTGAAGT

TATCGGGAAAGGATGATAGTATAGTTACTTTCAAATTTGCAAT

GTATGTCTATTTACGTGCACTTTGGGTATTACAGGAACCGCTT

ACTGATTTTATCAGAACAAGATTAGAGGACATACGTGAGACT

CTTGTAAAGAAGAAAATGAGTGAACATATGGTTGGACATCCG

TGGGAGTTGATTTATAAATATCTGGCATTTCTTTTTTATCGTGA

TGGAAATTGTGAAGCTGCTGAAAAATATATTCATAAAAGTGA

AGAGTGCTTGGAAACACAAGGACTGACTATAGATGCGATTAT

TCATAATGGTAAGTATGAATATGCAGAATTGTCAGGTGACGA

GGAGATGATGGCAAGAGAGAAAGCGTACTTTGATGAAAAAG

GGATAGATAGAAAAAATGTTTGTACTTTTATGTATCATTGATG

TTTAATAAGATTTGACCGAGGAGTGACAGGTAATCGCCGGTA

TATCTGGTATTACCTGTCATTTTTTGATGAAATAAGCTACTTTT

TGCCTAAAAAACGAAACTGTTGGTGTTTTATGATGATTGTGTC

AACAAAAGAGAGCAAAAGAAGAGGAGAAAAGTAATGTCAA

TGATTTCATGTCCGAATTGTGGTGGAGAGATATCTGAAAGGT

CAAAGAAATGTGTTCATTGTGGATATGTGTTAGTCGAAGAAG

CTAAAGTAGTGTGCACAGAATGTGGAACTGAGGTAGAGAGT

GGCGCTGCTGTATGTCCGAAGTGCGGCTGTCCTGTAAATGAT

AGTGAGACGCCTCAGAAAGTTGAAGTGACTAGGGTAAATGT

ATCTTCCGTAATCAGCAAAAAAGTCGTTGTAAGCATACTGAT

CGCAGTGATTACAATTGCAGGTTTTTTCTATGGAGTGAAGTAT

TCGCAGGAAAAGAAAGCAATTGAAGAGTCAGTAAAGCAGAA

GGAAGACTATCAAAGTACGCTAGAGCTTGCTTCGCTAATGAT

GCTTCAAGGAGCTTCGGATGCAGAAACTTGTGGGAATTTGGT

TAGGAAAGTGTGGAGCAACTGCATTTATAAGGAGAGGGATGA

AGAAACCGACAAGTATACGTGTGATAGCAGGGGTGCAGGAT

GGTTTTATGATGATTTTAATGATGCATTAATGGCTCTTTACAGT

GACAGCAGTTTTGGCAAGAAGATAAATGAAATCAAAAACGG

TCAGGAAACCGTTGCGGCGATGATGAAAGATCTGAAAAATCC

GCCGGATGAGATGGCAGATGCCTATGAGGATATTCAAAATTTT

TATGTGTCCTATCTAACGCTGACAGAAATGGTTGTGAATCCAA

CTGGAAGTTTGAGTTCTTTTTCATCTGATTTTTCCGATGCGGA

TACGGAGGTGTCCAATGCCTATAGCCGGATGAAGTTGTATTTA

GATTAAACTATTGAGGAAAAAATGGAGGTGCTTTAATGCGGG

GGAGAAACTGTGGAGGGTCATCAGGCGACGGACTGCTGGTA

CTTCTCGTACTGCTTGTCCTTTTTTATAAAATCATGCCATTCAT

AGGTTTATGGATTTTAATTTTTGGTGATGCTGAACGTAAAGAT
CTGGGTATGGGTATGATTATTGTCGGGATAGTTCTATATGTATT
ATTAGAGGTTTTTTAATGTGAGTTTCTGTGGTAAACTATAAAA
GTACAAGCTTTTGCGCCGCACCGCATAAATAGCGGATTTATG
ACCATTATTTGGTGAAAAAAATGGTGTACACCTGTGTTTTTTT
GTTTTGCGCCGCAAAATGCGCCACGGAACCGCATGCAGAGC
ACCCTGCAAGAGACAGGGTTATGAAAACAGCCCGACATAGA
GGGCAATAGACACGGGGAGAAGTCATTTAATAAGGCCACTGT
TAAAAGTTATGAAAACAGCCCGACATAGAGGGCAATAGACAT
AAAGACCAAAAACAGGTCATCTGCATACTGTGTTATGAAAAC
AGCCCGATATAGAGGGTGTGAGAGATATAGTTCTCGTCACAG
TGCAGAAAATGACCTATTATGTGCCGAAAAACAAAATGAAAA
AAGAATGGAAAGGCGTATTTAATGAAATGCTGATCTGTTGATT
TGAATTAACAAAAAAAGGTCGCCCCACGGATGACAAAAACA
TCCGGGGGCGACCCTTTT (SEQ ID NO: 64)

[1432] >E-座位

[1433]

TACTGTGTGCATAAGTCTTCCTTAGATCCATAGGTACAGC
AGTTTTATTTATTAGCCTTAGAAAATGGAAAATAGAGCTTATA
AATGATATGATATTTATGAATAAAATGATTGCATTCTCGTGCAA
ACTTTAAATATATTGATTATATCCTTTACATTGGTTGTTTTAATT
ACTATTATTAAGTAGGAATACGATATACCTCTAAATGAAAGAG
GACTAAAACCCGCCAAAAGTATCAGAAAATGTTATTGCAGTA

AGAGACTACCTCTATATGAAAGAGGACTAAAACTTTTAACAG

TGGCCTTATTAAATGACTTCTGTAAGAGACTACCTCTATATGA

AAGAGGACTAAAACGTCTAATGTGGATAAGTATAAAAACGCT

TATCCATCATTTAGGTGTTTTATTTTTTTGTGATTATATGTACA

ATAGAAGAGAGAAAAAAATCATTGAGGTGAAAACTATGAGA

ATTACTAAAGTAGAGGTTGATAGAAAAAAAGTACTAATTTCT

AGGGATAAAAACGGGGGCAAGTTAGTTTATGAAAATGAAATG

CAAGATAATACAGAACAAATCATGCATCACAAAAAAAGTTCT

TTTTACAAAAGTGTGGTAAACAAAACTATTTGTCGTCCTGAA

CAAAAACAAATGAAAAAATTAGTTCATGGATTATTACAAGAA

AATAGTCAAGAAAAAATAAAAGTTTCAGATGTCACTAAACTT

AATATCTCAAATTTCTTAAATCATCGTTTCAAAAAAAGTTTAT

ATTATTTTCCTGAAAATAGTCCTGACAAAAGCGAAGAATACA

GAATAGAAATAAATCTCTCCCAATTGTTAGAAGATAGCTTAAA

AAAACAGCAAGGGACATTTATATGTTGGGAATCTTTTAGCAA

AGACATGGAATTATACATTAATTGGGCGGAAAATTATATTTCA

TCAAAAACGAAGCTAATAAAAAAAATCCATTCGAAACAATAGA

ATTCAATCTACTGAATCAAGAAGTGGACAACTAATGGATAGA

TATATGAAAGACATTTTAAATAAAAACAAACCTTTCGATATCC

AATCAGTTAGCGAAAAGTACCAACTTGAAAAATTGACTAGTG

CTTTAAAAGCTACTTTTAAAGAAGCGAAGAAAAACGACAAA

GAGATTAACTATAAGCTTAAGTCCACTCTCCAAAACCATGAA

AGACAAATAATAGAAGAATTGAAGGAAAATTCCGAACTGAA

CCAATTTAATATAGAAATAAGAAAACATCTTGAAACTTATTTT

CCTATTAAGAAAACAAACAGAAAAGTTGGAGATATAAGGAAT

TTAGAAATAGGAGAAATCCAAAAAATAGTAAATCATCGGTTG

AAAAATAAAATAGTTCAACGCATTCTCCAAGAAGGGAAATTA

GCTTCTTATGAGATTGAATCAACAGTTAACTCTAATTCCTTAC

AAAAAATTAAAATTGAAGAAGCATTTGCCTTAAAGTTTATCA

ATGCTTGTTTATTTGCTTCTAACAATTTAAGGAATATGGTATAT

CCTGTTTGCAAAAAGGATATATTAATGATAGGTGAATTTAAAA

ATAGTTTTAAAGAAATAAAACACAAAAAATTCATTCGTCAAT

GGTCGCAATTCTTCTCTCAAGAAATAACTGTTGATGACATTG

AATTAGCTTCATGGGGGCTGAGAGGAGCCATTGCACCAATAA

GAAATGAAATAATTCATTTAAAGAAGCATAGCTGGAAAAAAT

TTTTTAATAACCCTACTTTCAAAGTGAAAAAAAGTAAAATAAT

AAATGGGAAAACGAAAGATGTTACATCTGAATTCCTTTATAA

AGAAACTTTATTTAAGGATTATTTCTATAGTGAGTTAGATTCT

GTTCCAGAATTGATTATTAATAAAATGGAAAGTAGCAAAATTT

TAGATTATTATTCCAGTGACCAGCTTAACCAAGTTTTTACAAT

TCCGAATTTCGAATTATCTTTACTGACTTCGGCCGTTCCCTTT

GCACCTAGCTTTAAACGAGTTTATTTGAAAGGCTTTGATTATC

AGAATCAAGATGAAGCACAACCGGATTATAATCTTAAATTAA

ATATCTATAACGAAAAAGCCTTTAATTCGGAGGCATTTCAGGC

GCAATATTCATTATTTAAAATGGTTTATTATCAAGTCTTTTTAC

CGCAATTCACTACAAATAACGATTTATTTAAGTCAAGTGTGGA

TTTTATTTTAACATTAAACAAAGAACGGAAAGGTTACGCCAA

AGCATTTCAAGATATTCGAAAGATGAATAAAGATGAAAAGCC

CTCAGAATATATGAGTTACATTCAGAGTCAATTAATGCTCTAT

CAAAAAAAGCAAGAAGAAAAAGAGAAAATTAATCATTTTGA

AAAATTTATAAATCAAGTGTTTATTAAAGGTTTCAATTCTTTT

ATAGAAAAGAATAGATTAACCTATATTTGCCATCCAACCAAAA

ACACAGTGCCAGAAAATGATAATATAGAAATACCTTTCCACA

CGGATATGGATGATTCCAATATTGCATTTTGGCTTATGTGTAAA

TTATTAGATGCTAAACAACTTAGCGAATTACGTAATGAAATGA

TAAAATTCAGTTGTTCCTTACAATCAACTGAAGAAATAAGCA

CATTTACCAAGGCGCGAGAAGTGATTGGTTTAGCTCTTTTAA

ATGGCGAAAAAGGATGTAATGATTGGAAAGAACTTTTTGATG

ATAAAGAAGCTTGGAAAAAGAACATGTCCTTATATGTTTCCG

AGGAATTGCTTCAATCATTGCCGTACACACAAGAAGATGGTC

AAACACCTGTAATTAATCGAAGTATCGATTTAGTAAAAAAATA

CGGTACAGAAACAATACTAGAGAAATTATTTTCCTCCTCAGAT

GATTATAAAGTTTCAGCTAAAGATATCGCAAAATTACATGAAT

ATGATGTAACGGAGAAAATAGCACAGCAAGAGAGTCTACATA

AGCAATGGATAGAAAAGCCCGGTTTAGCCCGTGACTCAGCAT

GGACAAAAAAAATACCAAAATGTGATTAATGATATTAGTAATTA

CCAATGGGCTAAGACAAAGGTCGAATTAACACAAGTAAGGC

ATCTTCATCAATTAACTATTGATTTGCTTTCAAGGTTAGCAGG

ATATATGTCTATCGCTGACCGTGATTTCCAGTTTTCTAGTAATT

ATATTTTAGAAAGAGAGAACTCTGAGTATAGAGTTACAAGTT

GGATATTATTAAGTGAAAATAAAAATAAAAATAAATATAACGA

CTACGAATTGTATAATCTAAAAAATGCCTCTATAAAAGTATCA

TCAAAAAATGATCCCCAGTTAAAAGTTGATCTTAAGCAATTA

CGATTAACCTTAGAGTACTTAGAACTTTTTGATAACCGATTGA

AAGAAAAACGAAATAACATTTCACATTTTAATTACCTTAACG

GACAGTTAGGGAACTCTATTTTAGAATTATTTGACGATGCTCG

AGATGTACTTTCCTATGATCGTAAACTAAAGAATGCGGTGTCT

AAATCTTTGAAAGAAATTTTAAGCTCTCATGGAATGGAAGTG

ACATTTAAACCACTATATCAAACCAATCATCATTTAAAAATTG

ATAAACTCCAACCTAAAAAAAATACACCACTTAGGTGAAAAAA

GTACTGTTTCTTCAAATCAAGTTTCTAATGAATACTGTCAACT

AGTAAGAACGCTATTAACGATGAAGTAATTCTTTTAAAGCAC

ATTAATTACCTCTAAATGAAAAGAGGACTAAAACTGAAAGAG

GACTAAAACACCAGATGTGGATAACTATATTAGTGGCTATTAA

AAATTCGTCGATATTAGAGAGGAAACTTTAGATGAAGATGAA

ATGGAAATTAAAAGAAAATGACGTTCGCAAAGGGGTGGTGG

TCATTGAGTAAAATTGACATCGGAGAAGTAACCCACTTTTTA

CAAGGTCTAAAGAAAAGTAACGAAAACGCCCGAAAAATGAT

AGAAGACATTCAATCGGCTGTCAAAGCCTACGCTGATGATAC

AACTTTAAAAGGAAAAGCAGTGGATTCTTCACAAAGATACTT

TGATGAAACGTATACTGTTATTTGTAAAAGTATCATAGAAGCA

TTAGATGAAAGCGAAGAGAGATTACAACAATATATTCATGATT

TTGGAGATCAAGTGGATTCTTCACCTAACGCACGAATTGATG

CGGAATTACTACAAGAAGCAATGAGTAGGTTAGCTGACATAA

AGCGGAAGCAAGAAGCACTTATGCAATCCTTATCTTCTTCTA

CAGCAACGCTTTACGAAGGCAAGCAACAAGCGTTACACACT

CAATTCACGGATGCGCTGGAGCAAGAAAAAATATTGGAACGC

TATATTACTTTTGAACAAACTCACGGGAATTTTTTTGACTCAT

TTGGAGAACTTGTCTATCGAACGGGACAAGCAGTGCGTGAAT

TAGCTAATAACGTCACATTCGAGAGCCAAACAGGAAGCTATC

ATTTTGATAAAATAGATGCTTCTAGATTCCAAACTTTGCAAGA

AATGTTGCCAAAGGCAAAGAAAAAAGCATTTAATTTTAATGA

CTACCAAATAACATGGAATGGCACCACGCACCTTTTATGGAA

AAATGGTAAAGTGGATGCAGAAGCAACCAAAGCTTATAACG

AGGCGAAACTGAATGGAAAGCTACCAAAGGAAGGTAATGTA

GCAACACAAGATGCAGAACTATTAAAAGGCATTTTGGCTTCA

CTGAAAAACAAGAAAGATCCTATCACTGGAGCAGATATAAGC

AGTGTGCATGTATTATCTATCCTTAGCGGGCTCGCATTCTCCT

ATACAGCTGGGAATTATAAGGGAAGAAAACTTACTGTTCCAA

AAAGTTTCTTAGACAAATTAAAGAAAAACCGAAAATCTAAAG

TACCTAAACTATCTAGTTTATCAGAAAAACAACAACTAAAAC

TCGCAAATAAATACAAGAAAAAATCACCTATTCCAATTCCAG

ATGATGCTAAAATCAAAGCTCAGACGAAAAAGGCTGGTTATG

AACAAATATCTTATAAATGGAAAGAGAATGGGATAACCTTTG

AAGTTAGATGGCATACTAGGACACCAGGTGCACCAAAGGAA

CAAGGAAATACGTTTGTTATAGAAAGAAAAATTCAGGGTACA

GCAGAAGGGAAAACAAAAGTTCAACAAATATTGGTTGGAGA

TAATAAGTGGGTGAGTAAAAGTGAGTGGCAAAAGGCTATAAC

TGATAAGAAAAATGGTGTAAGTACCTCGGAGCAAAATAAAAT

GTTGTCTGATGGACATTGGAAAGAATAGAAAGGAGCAAAAT

GATGGAAGATTATTATAAAGGTTTTGAGGGATATCCAGAGATA
GATTTTTATACGTATATAGATGATATGAAATTGGGTATAGCAAT
GTGGGAAGGATACTTTGACAACATTATGAAAGAAATTAATCC
AAGTAACGGAAGATGGACTTCATTAGCGTATTATTATCATTTA
GATGAGGGGTGGTATGATGAAAGTCCTTGGGAAATACCAAGT
AATACAGAAGCATTAGAATTATTGGAAACAATCCATATATCTA
ATCTAGATACTATCACACAAGAGATATTACTTAAATTAATAAAT
TTATTAAAGAAGAATATAAATAGACAAGTTTATATTGAATACT
CATAAAAAAAGATGATTATGATATATTATAGAACAAACGAACAA
GCCCCAAATACGAGGTTTGTTCGTTTGTTTTCAATATAATTATT
TGCCACCAAGTGAGATATTACGGTTTTAAATAGCTTATTTGAC
GATACCAAACCCTGATAAGAGAAAGAAGAAAGAGAAAGCTG
GTGTAGTTGTTTTAAGTGAACTAGATAAAAAATTAATAGCAA
AACTTGAAAAAGATGGTGTGAAAATATCAAAAGAAGATGTTA
TAGGAATAAAATAATTGCCAGATGATGAGAAATCGTTTGGCT
GGAAAAAGGAAATCCATCCGCTGGATTTGAGCATATTCTTATT
GAACATGGTGAACAATTTGCTAAATAGGGAATTTCAAAAGCT
GAGTTACCTGATTTTTTGATGACTGCTTTAGAAAAGGAAA
（SEQ ID NO: 65）

[1434] >F-座位

[1435]

ATTCTTTAAAAATATCTAATAATTTATTTACTATATACTCTA
ATACATCTTTTAACCTATCTAAAACATCATCACCTACAACATC

CCAAAAATCATCTAAAAAGTTAAAAAAATCCATCTTTATCAA

CTCCTATATCTATTTTTTATTGTGTAATTCCTGAGTTACAAAAC

CATTATAACACGTATTACACACGTAGTCAATACTTCAAAAAAA

TTTTTTGTATATTTTTTTGAATAAGTAAATAAAAAGAGCTGTG

TAGCTCTTTATTAAAATCAATATTTTTATTTTGTTAACAAACTT

AGACAACATTAAATTTAGAAACCTATATATATTTCAGTACTTTT

CATTTTTAGGTAGTCTAAATCAGAAATGGTTTTGTCTAAATGA

TGTATGTAAGTTTTAGTCCCCTTCGTTTTTAGGGTAGTCTAAA

TCAGAAGTCATTTAATAAGGCCACTGTTAAAAGTTTTAGTCC

CCTTCGTTTTTAGGGTAGTCTAAATCCCATCCAAATTATGGGA

TAATATGTTACTTTTTATTTTAATATTTGATTATTTATTGTTTTTT

TACTGATTTAGATTACCCCTTTAATTTATTTTACCATATTTTTCT

CATAATGCAAACTAATATTCCAAAATTTTTGTTTCTTTTCTTAT

GATCTTTTCTCCGATAGTTATTTCTCCAGATAAGATTTTCATTT

TTTTGAATTGATCTTCTGTTAGAATTAATGTTCTTACTGATGA

ATTTTCTGGAACTATCATTGACAACTGATTTTCATAGGAAATT

ATTTTTTCTTTTGTGCTAGAACTTACAATGTATACTGATTTTTG

TACCTGATAATATCCTTTTCTTATAATTTCTTTTCTAAATTTTGC

ATATTCTTTTTTTTCTTTTCCTGTTTGCATTGGAAAATCATACA

TTAGAATCCCTACATAATTAGTACTCATAATCCTCTATCCTTAA

CTCAGGAATTTCTACTTCTGACATTTCTCCTGTAAAATAATTT

CTAATATTATCTAAAAAATAATCAATCACTTGAGCCAATTCATA

TTTTTTATTTTTCCAATAAACTTTTTGTGTTAATACCAATAACA

ATTTTTGTCTTAATGATTTATTCAAACTTACTTCTTCCTGTTGA

TTAAAATATACGATATAATCTACCATTGGACGAAATATTTCAAT

AATATCATCTGCAAAATTATAATTATTAAATTGTGAACTGTGAT

GTATTCCCAAACTTGGATGAAATCCTTTAGCCACAATTTTTGA

AGAGATTAAGCTTCTCAAAACCATATACCCATAATTTAATGCC

GAATTTGTCCCGTCTTCACCAAATCTCTTAAATTTTTTCCCAA

AAAGTTCACCAAAATACATTCTTGCAGCAATTGCTTCCTGAT

GTTCCGCTTCTTTTCCTTTTAATCTAATATTATTTTCATATGCTT

CCAACTTATATGATACTTCCTGAGATTTTTTCAAAAACTGCAA

TAAATTTCTTTGATTTTCTATTTTTCTCATTACAATTTTTCTCC

AGATTTCTTCTTTTTTATCGTCAATCCAGCTCACTTGCTCATTA

ATTCTTGTTGTTACTTGAAAATGATTATACAGTCCTAATGAAT

GTAAAACTGGCTGATGTTTTTCATTACAAATTATCAGTGGAAT

ATTATGTTCTGATAATCTTAACTGTAATATTCCGCTAATTTTAC

ATCTGCAATTTTCAACTACAATTGCCATGATATCATTTAAAGA

TACTTTATCAGCCTTATTTTCATCATCTTCATTTATCATCACAA

GCTGGTTATTTAAAACTGATAATTCATTGACTCTTGTTACATG

GATAATATTAGACATTTTTATTACTCCTTTACTCTAAAGCTTTA

TATTCAAACATAACTTTCACAAGTTCACACAATTCTTCTGAAT

TTCTATCAGTCATTAATTTTTTCTTTTTTAAATTTTTCAAATGT

ACAATTTTTTCCGATTCTAAAGTCTGAATTTCTATTTTCTTATC

TGCTCCTATTTTAAATGTTGCTACAAAACCATATTCCTTTAATA

TATCCACTATTGATTTCATAATTGCATTTTTAAGTTTTCTATCAT

AAGAAAGTAATTTTCTTAAATTTTCCAGCACTTCTAAAAGTG

AAATTTCAGCATGCGGAATATAGTTAAAATGTGCAATATAGTT

TCGTATATACAAATCTTTTTTCTCTTGTTTTAATTTTTTTACTTT

TTTATCAGAATAGATGCTTCTTTTTTCTACATTATCTTTGTATA

ATTCTTTATAAAAATTTATATATTTTTCAACAATTTGCCCACTT

TTATATTTTACATTTTTACTGTTATCAAAATTAAATATTTCTTCA

ATATAATGATTTTCAGGAAATTCACCTTTCAATCTAAATCTTAA

GTCCCTTTCCCAGATCGAAGTATATCCCACAAGTCTGTGGAG

TATTTTTAATAACAAGCCTTGCAACAAGTTTAATTCATTAAAT

TCCACTTTATTTTTCAAATGAGTATATTTTTGTATATTTCCAATT

GCTTTTTCATATTCTTTATAATCTTCATCATTAAATTTTTCATCT

TTTTTAGGTCTTGCATATTTTCTATGTAAATTTTGCTGCATTGT

ATAATTTTTTTCTATTTCATTTTTTTTTATTGCTGTATTCTTTCAA

TTCTTTTAAACTTATTTTATACTTCGCTTTATCAGCTATTTTTTC

AAGTAAATTTAACATCCCATATTTTTTTATATTATAAAAAGCTC

TATGCTTTATAATATTTTCTCCATCAAAATATATTTTATTTGTGT

CAAATTTCTTCAATTCTTTCCTATCTTTTATTTTATTTTCATTAA

AATCTAAAAATTTTCCAATTTCATTCGCTTCTAATTCAAAATC

TTCTGTTACTCTATTATTATCTAAATTTAAAAGATTTATAAGTT

CAAGTTCATCTGAAAAAGTTTCTTCTTTATTTGCACTCTGATA

TTTTTCAAGACTTCCCTTCAAATTAGTCAATTCTTTATGATTA

AGCAATTTTAAAATTAAATAAAACATATTCAAATTTTCAGTGT

ATTTTAATATCTTTCCTAATTTTATCTCTCTTACAAATTCATTTA

TTTCATGTGGAATTTCTTTATTCCTATTATGTTTTTCATAATTTT

TTAAAATTTTATCATATTTTTCTTTATTATCTTTTTTTATTTTTAT

TTTAGAAAATATATCATTATTATCATTGTTATTATTACTTTCTAT

ATATTTTAAATTATTTTTATTCAAATAATCTATAAAACCTTTTAA

AAATATTTGTTGTATAAAATCAATGTATGTATTTTTTTCTTCTT

TATCTTGATTATTAATCATCTCCCTACTTTGTATAATAGCAAGA

TATTCTACTGGTACAGTTTTTTCTATATTTTCAAATTTTTGATAT

TTATAATGTCCTGTTTTTTTGATTTCTTTGTTTATTTATTTTTATT

ACTTCATTAGTTATTTTAAAAAAAACTTTACTATTTTTAACAA

ATTTATTAAGAAATTCACCATAATAAATATTTTTCAAAAGATAT

ATTTGAGCATCTTTTTCTTCTTTATCCTTAGGAACACTCCAAA

AAAATTTTAAAGTATTTCTTAAATCTTCTATTTTATTATATAATT

TCGTAAAAGAAGGAACAAAAGGAATATTCTTATTTACAAAAT

TAAATTTTGTATTTTTTAAATATTTAATTATCACATCCTTTTCAT

AATAATTAAATACATTTGCACTATTTAACTGCTTAAATATCTTC

AATTTCAATTTTTTCTCATTTATTTCATTTTGAAACATTTTTTT

TGAAATTTCAGAAGGAGCTATATTTTTAAATGCAAATATATCT

TTCCCTTCTAATTCCAAATTAAAATGCACAATCCCATGTCTAA

TACTGCTAATAGCTTCATCAATATTTGCAAAAAAATCTTCTAT

CTCATTTTTATTATCCATATTAAAATCATAACTATAGAACATTTT

TAAATTTTCTTTTACTTCATTTTGCTTGTTTTCATTATATATTTT

ATCAACTTCTCCAGAAACATATTTTTCTTCGCCCTTATTATTTT

TTACAGTTTTTCCTCTCATTCTACCTGTAATATCATTCTCATTT

TCAGTTTCAAGAATATTTCTCAATGAAAAATATGCAACCGAA

GAAACTCCAATTATATTTCGTAAAAATGCTTCATTTTGTCTATT

CCTAGCAATAAAATCACTTGTTGCAATCTCTCCAACTTGTAAA

TAATAATTGTATTTCCCACAATTTCTTACATAAGTATCCAATTT

ATTTAGTAATTTGTTTTCAATTAATTTTTTTAAATTTTGATATTC

AAATATTCTCTTAATTTTATCGTTACTTATGTTACTCAGTCTTT

TATACACATAATTTTTCAAAAGCTGACTCATTTCAATTTCCAC

AAAATGACAAAAAAGCATATTTTATATTTTTATCATTAAGTTCTT

CTTTATCCAAATAATATTTATAAAACACTTGTGATTTTTTTAAT

TCACTCATATCCGGAATTTTTTCAATTAATTCTTTTATATTATTT

ACATTTTGTATTTCTTCGTAAATAATTTTAGCAAAATTTTCTTT

ATCATTTTTTCTTCCAATTATTTTGTGATAGTATTCTCTTATTTT

ATATTTTTCATGTTTTTTTGAATTTTCTATTAAAAAAAATAACT

TCTCAATATCTTCTTTTTTATACAATTTATCAAATGCTTCCTGT

ACATTATTTATATAATCATTACGCTTTGCTGATTCTCTATAATAA

TCATAAATAATATTTCTTTTGCTCTTCCCTCCAACTTTTTCAAC

ATTATTTTCATTAATTTTCTGATAATTAGCCTTATTTTCTTCAAA

TGAATATTTTAAAGAATTTATCTTATTCAATTTTGCCTCAACAT

CTTTTCTAAATATTTCTAATTCTTCAGAGTTCACATCTTCATTT

AACAATATTTTCTTTAAAACTGAAAAACTATTTTTATTTTTTAA

ATCATATTCTGAAATATCTTCTTCAGAATAATTTTTATCCTGTA

CTGCATTTTTCTCTTTCCTATTCTTTAAATACAGAACACTATCT

TTTAGATGCAATACTTTATTTGAAAAAAACTTTTTTAAATTTT

CTCTTCTTATTCTATTTTCTTCTTCACTTGCATTATCAGGATTT

TTTATATATATATCCAGTCTTATACTTAAAAGCTCTGACAATCT

CTCACTAGTCCTATTTTCTTCGCTCGTACTTTTTACTAATTTTC

CCTCTTCAATATATTTTTTATGCGAAATTCCATCAACTTTTGTA

ACTTTCATATATAAAAACCTCCTAATATCTATATTTTTTACTCA

ATACCTAATTCTTTTTTCAATGCTTTTTGTAAAATTTGTGAAA
AATTCAGATTTTTTTCCTGTGCCAATATATCTAACCAAACAGG
AATTGTTAAAGTTTTCTTTTTAAGTGCATTTGTAACTTTTGCC
ACTTCATACACTGGATCAACAGATAAAATATACAAATACTGAT
TTTCTTTCAGTTTCACATCCTCCACTTTTGAAGGCTCAGGAA
ATTTTTTTCTTACATCCAAAAAATCAGCCAAATGCAGACCCA
ATGTCTCTCTCAAATTGGAAACAGCCTCCTCCATGCTATCTCC
AAATGTAGCATAATAATTTATCTCTCCATCTTCAAACTTATCAA
AATCAACAATACAACCATAATAAGTCCCATCTTCCTTAGTTAC
CACTGCTGGATAAAATACATCCATTTTAATTATCTCCAATCTAT
ACCACGTGTTAAATACGTGTTTAAAAATATTTATAAAATTTTTT
AGCATCTCTGCTAAAATAAAACAATTATTTCAAATTTTTCTATT
CCTTAATCACTCATTGTTAGTGATTCTTTTTTTACTTGGACAAT
TTTTCATTTAATTTCTTCAATTTTTTTAAAATCACATTTTTTTA
ATATTCCTTATTTAATTGCAAATTTTCATTACTTTTGGGGTGCT
CTAAATCCCATCCAAATTATGGGATAATAATTTTTAGTGAAAG
CAAGAAGGGACTAGAATTTAATCCCAACTTGTTTTTCAATACT
TCTTAATGTTCCTACAGGTATATCTTTTGAATATGGTACTGTGA
CCACACCTTCCACACCTGGGATCATCCATTGATAATGACTACC
TCTTATACGCACAACTTTTCCGCCTAATTTTCTAAATCTTTTTT
CGAT（SEQ ID NO: 66）

[1436] >G-座位

[1437]

CTTTCTATCTTTTTCAAATAAAATTAGGCTCTAGTTAGCC
TAATCGCATAATTATTTATTATAGTATAATTCTTATTTTTTTTCA
ACCTAAAAATTTAAAACATCTCCAAAAATTTTCGTTTCAGAA
CAACCAAGCAACCATATTCAAAAAACAATAAAAAATGAGCA
AGAATTGAAATTTTATTCTCACTCAGAAGTTATTTTTATTAAA
TATCACTTTTCGATATTGGGGTGGTCTATATCAATTTAAAAGA
CAGAATAGATAATTCTTTAGAGTTTTAGTCCCCTTCGATATTG
GGGTGGTCTATATCAGAAGTCATTTAATAAGGCCACTGTTAA
AAGTTTTAGTCCCCTTCGATATTGGGGTGGTCTATATCCCATC
CTAATTTCTTGCTGATGAGATATTTATTTCTAATTTTTCTATTT
TGTCTTTATTTTCAATACTTTCAATCCTATTTTTCTCTTTATTA
ATAATATAGAACCACCCTATACTATTATACCATATTTTTTGATT
TTTCAAAATTCCAATATTTTGTTTTGTGAAATTTTTTCTCCCA
TTGTCACTTCTCCTGCAAGTACCTTCATTTTTTGAAACTGATC
TTCTGTCAGGATAATGGAACGGATTGATGAATTTTCTGGAGC
GAGCATTGATAACTGTTTTTCTGCCAGTTCGATTTTTTCTTTT
GTTTTCGACCTCATTATATATACCGATTTTTGAAGCTGATAAT
ATCCCTTTTCTATCAATTTTTTCCTAAAAGTCCTATATTCAAAT
CTCTCAACATCTGTCTGCATAGGAAAATCATACATAAGCAGA
CCAAAATACTCAATACTCATAGTCCATCACGCTCAATGTCGG
AATTATCACTTCTTCATCTTTTACAAAATAATTTCGTATACTAT
CCAAATAATAGTCTACCGCTTGGAAAAAATCATATTTCTTATT

GTTAAATAATACCTTCTGCTGTGCTACAAGAAGTATTTTTTGC

CTTATTTCCTTACTTAATTTCACTTCATTCAAAATATCCTTGTA

CATATAAACAAGATAATCCACCATAGGACGAAAAACCTCTAT

TATATCATCAGAAAAATTATAGGCATTAAACTGTGACTTATGA

TGTAATCCTAAACTTGGATGAAATCCTTTTGCTACAATCTTTG

ATGATATTATAGCTCTTAAAATCATATATCCATAATTAAGTGCA

GAATTCACTCCATCTTCATCAAATCTTTTAAAACTATTACTAT

ACAATTCCTGAAAATATATCCTTGAAGCTATTGCTTCCTGATG

TTCTGCACTCGCATCATCTTTTTTCAAGTTTTCCTTATATGTTT

TCAGTCTTTCAATGGAAATATCACTTTTTTCAAGATACTCTAA

CAATGCTCTTTGATTTTCAATCTTATTCTCCACTATCCTGCTC

CACAATTTTTCCTTTTTCTCTTTTTCCCACTCAATCTGCTCAT

TTATTCGTAAAGTCACTTGAAAATGATTAAATAATCCCAGCG

AATGAATTTCAGGCTGATGTTTCTCGTTGCAAATAATAATCG

GAATGTTATTTTCCACCAGCCTCAACTGCAAAATCGCACTAA

TCTTACAATAGCAGTTTTCAATAACTATCGCAGATATATCATT

CAAAGAAATCTTATTTTTCTCATCATTATTGTCTTCATCAACC

ATTATAAGCTGATTATTCGATATTGACAAATCATCAGCCCTTG

TTATGTGAATTATATTGGGCATTTTAATCATACTCCTTATAAAT

TTCATTCTTATAACGTATCATTCGTATTTTCTATTTTTGTTAAA

AGTTCTATTATCAAGTTTTTAATATAATCAGAATTATAACTTTC

TAATTCTAAAACAGAAACTTTTTTAGGTTTCATTAATCTTTCA

AGTATATCATTATTACCGATAAGTTTAAATTTTTTCTTTAATTC

ATCATAATCTAAATTCACATCTTTTTTAAATACTTCAAATACA

CTTGCATAAGTTGAATTATTATAACGTGTACTATATGATAATAA

ATTAGAAACTCTATCAATTTGTTCTGCAATACTGTAATCAGCA

AACGGATTTCTTACAATATAGAAATGTGAAATATAGTTTCTAA

TACTTTCATTTTCCGGCTTATTAATTTCAGAATTTTCAGACAA

ATCAATTCCAAATCCATAACATATTTTCTCAAATTTTTTATAA

GATTCTTCATCAAAAAATTTATAGTATGCTGTTGTTGTATAAA

AGCCATCAGATCCATTACGCTTAGGATAAGCTCTACTTATTCC

AGTATTGTAGCCACTTAACTTAATAATTCCTAATTCTCTTAGC

CCATTTACAATATAGTGCATATCTCTTTCAAATCTAGCCATTT

GAATAGCAAGTTTCCAATTTATATCTATCAAATAACTTTCTAT

TTTATTCAAATAATTAAATTCTACCAAATCTCTAATTTTTTTGT

ATTCAGAAACTCTATTATAATCTTTTTCAAATGATTTATAGTTT

TTATTTTGTATATTTTTTGCAAAAAAGTCATCATTTTCTTTCA

ATTTTTTTATATACTTCTCTTTGTATTCTTTAGAATATCCATTTA

GTTTATCATTTAGATTTTTCAATATTGCATCAATTTCAGATATT

TTATTTTTTCTAATATTTTTACCATCAATATTAAATAAAAATTT

TGCATCAGCCATTTTAATATCATTTGAAATTAATCCATAAATTT

TATCAAAATTTGGATTTCCAATATTTAAAAATAAATTCTTTTT

ATAAATATATAATTCATTCTTACGTTCTTTAGGATAATATATTT

CTTGAAATTTATTTTCATTCTCTGATTCCATATCTTCTATTAAA

TTATCTATTTCTTTTTTGTATTTTTTTAAAAAATCAGAATTAAA

TATTATTCTACACAATATTTTACTCTTTATTTCCTGATCTTTATC

TTTTATATACTGATCAACCTTTTTTTTCAAATCCTTTTTATTTA

TGTTTGATAACTTTCTTTGTTCATCTTGTAATATATTCGATTTT

TTATCTATCTCAAATTTAGTTTCATCATCAAAAATTACAATTTT

TTCTAATTTTTTCTCTAAAACATCACAACCATTAATATCATCT

TTAAATTCAGTTAATATATTATTTTTTATATCCTCATAATAATTA

TTAAAAATTTCTTTTTTAGTTTGTATTTTAAAATCATCAAAGT

CTTTTTCTATCTCTTTCATTTTTTGAATAAATTCTTCTAAATTA

AGATTCCAATTTTCAGTTATACATTCATTTCTCAAAGTATTTA

ATTGCATTATTTCATCTAAAATATCTATAATATTTTGATATTCTG

AAGTATTTAACCAAACTGATGTTGCAAAAAATCTATTTCTAAT

TTTATTTATAACCGCATTACTATTTAACAGTGCAAATATTGAA

ATTATATATTCAAAATCATCATTTATTACTATAGTTTTATCACTA

GTCTTTACAGTTATTCTTTCGTAAGTTTTATTATCATTAATGTC

TTTTATTTGTTTCTTAATTTCTTGAATATTCATTTTAAAATCTG

AAAAATCAAAAAGTTCCTCATAATTTTTTCTCAAATATCCAAT

ATAACATTCTATTACTTTTTTCTGATATTTTTTAATAGCTTTATT

ATTACCTTTTGAAGCAGAAATCTGAGCATTTTTATAATAATTT

TCTATAATATTTTCATCTATTTCATCAATGTTTCCTAAAGTTTT

CTTTAATTCTTGTAAAAATATATTCTTACTTTCATTTTCTTCTA

AATCATCTTCTAAAATTAATTTCTTATACAATTCTTTATTCACA

TATATTAAAGCATTTAATACTATTTTTTCTGTTTCTATAGTATC

AAATGGTTCATTCTTAGGATTATTCCTATATAAATTTAATATTT

CAGGAAGTACTTTAGAAAAGGATGGTAAATATTTAATATCATT

ATTATTTTCTTCTGAAATTTTAATATCATTTATTTTAGTAATTAT

ATTTTTTTTATCTTTAAATACTACATCTAAATTTAATGCTTTTG

ACACTTCTTCATCTGATATTTTTAAATTTTGAATTATATTTATG

ACTTTATTATAGTCATCTTGCGTTCCTTGTAAATCTCTTTCCTT

GCTAATCGCATGTAATATCCTGTTTCTTTCATTTGTTCCTATCT

TTGTAAATTTCCTAATAAAATTATTTGTAATGTTATTTTTATTA

TCTATAAAATCTAAGTCTCTTATTATTTTTATTTTTGAATTTAA

AATTTTTTTATCAAGTACGTAATTTTTTTCTCGATCTCCTCCA

AAGAAATCTATATTTTCATCATTATTTATATTTTCTCTAGAAAA

AATCTTATTTAATTCCATATTGGTAGAAGCAAAAAAAAGTAATC

AATTCTAAATCCAATTCCTCTTTAGCGTGAAGTCTAGAAAAA

TCATCAGTATTTACTGTTGTCATATCTATATCATTATGTCTTAA

TTTCCCTAAATACATAATATGCTCTAACGTATATTGCTTAACTC

TTTTTAAAATTTTTTCAGATAATATACTTTCATTTAAAATTTTT

TCTATTTCTATTTTTTCCATTTTCTTTAATCTGACTTTTTGTTC

ATTTACCAATATTTTTTCAATTCTTCCTTTCAAATATCGATATA

TGATTTTATATAGTTCTTTTTCTTCATCAGATTTCTTTGAAAAT

TTTTTCGAATCAAAATTAACTTTATAATGTTTTTTAAATATTCC

AAAAATTTCTGTATCACAATTTCCTTTTTTTAGTTCTTTTTCT

AATTTTTTTATTAATTCATCTATTTTAAATTCTGCTAAAATTTT

TTCTATTTTTTCTTTTATACTATTATTTTTTATATTTTCTACAAA

AAATTTTACAATTTTATCTTTTTTATTTTCTCTTTCTATTTTAA

ATTTTTCGTGCTTATCTAATAGTACATAAGATTTTATATATGTT

CTATTTCTTCTCTTTTCAAGAAATTCATTATTAACTTTTTTTAC

TTTTTCAATTCTTTTAGTAATATTCCAAAATTCTAACTCTTTTA

TAACAAAATCAGCTATATCTTCTACTGTTAAATCTACATTTAT

ATTTAAAATTTTTTCAACAAGCATTTTTTTATTTTTAGATTTCT

TTTTATCACCACCAACATTAAGATAAAATTTTACAAAACCCA

GAATTTCTAAATTACTTTTTATTTTTTCTCTTATTTCCATAAAA

TTAGTCAAAATAACATCTATTTTATCATCTTTCAATAATTTTTC

TCTTAAATGTTCTTCATAATATCGATTTTCAAATACTTTTTCTG

TTTCATTTTCAATTATTTTTTCTATAATCTTATATAAACTCATG

TTAATATTTTTAAAAATTTCGTAAATTGATTTTTTTGTTTCTAA

TTCATCATTTTCTATTATTCTTAATATTATTGAACAATCATTTA

GTGTTTTATTAGTATACTCATCTCTGATATCTATCTCTATTTCT

TCTTCATTCTCTTGTCTCTTTATTTCTATTTTTTTATCATCTTTA

GTTATTCCTTGCCTAATTGCTTCATCTATTATTTTCTTTTTTGT

AATCCCCAATGCTTTCAATTTCTCAGATTTTCCATATGCTTCT

ATATATAATACAACTTCTTCTGTTTCCAAAAAATCATCATTATT

TTCTATTCTTATGATTCCTTCTTTACCTTTCAACTTAAATAGAA

TATTTCCTGCATGAAATTTTCTTGTAAATTCTTTAAGAATATTA

TCATTTTTTTTGTAATTAATATATTTTCTAATAAATTTATTATTA

TCAATTTTTTCTTTATTATTATTTTCATTAATATTTAAAATGTAT

TTGTTTCCATCATAGTTCCTTTTAACTTTTACTTTCCGTTTTAT

TTTAAAATCTTTTTTATCACGAACTTCATACCATCTCTTATGT

CCAAATAAATTTCCCATTCCAATCTCCTCGTTTCTACTTTAAT

CTAATAAAATATTTTTAAATTAAATCAATTTTACATCTTTCTAAT

CAAAAATACAATTTTCCATTTTTAGTATACCACATCAATATTAA

ATCTCAAAAAAATAAGGAGCCGTCAAACATAGCTCCCTACTTC

TATTTACTCATAATCCCCATCTATCCTTACTTTTCGTAAAATCAA

TCCTTCTTTCGCCTTTAGATCCAACTTAATTTTCCCATTTGAA

CCTGTTCTAAATGTTCTGCCTTCTGTTACCAAATCAATAAATCT
TTCATCCTGATAATTTGTTTCAAATTCCACATTTTCCCAGCTGTT
AAACGAATTATTTATTACAACAATAATTAAATGATCCTCGATTAC
TCTTTCATACACAATTATTT（SEQ ID NO: 67）

實例3：Cpf1和相關組分的進一步評價

[1438] 申請人使用Cas-Cpf1異種同源物進行序列比對並且比較結構域結構和組構（圖38A-圖38N）。Cpf1座位比對的綜述示出在圖39中。

[1439] 以下列出了不同異種同源物中的Cpf1座位序列：

[1440] >KKP36646_（修飾的）假定蛋白UR27_C0015G0004 [儉菌總門細菌GW2011_GWA2_33_10]

[1441]

MSNFFKNFTNLYELSKTLRFELKPVGDTLTNMKDHLEYDEK
LQTFLKDQNIDDAYQALKPQFDEIHEEFITDSLESKKAKEIDFSEY
LDLFQEKKELNDSEKKLRNKIGETFNKAGEKWKKEKYPQYEWK
KGSKIANGADILSCQDMLQFIKYKNPEDEKIKNYIDDTLKGFFTY
FGGFNQNRANYYETKKEASTAVATRIVHENLPKFCDNVIQFKHII
KRKKDGTVEKTERKTEYLNAYQYLKNNNKITQIKDAETEKMIES
TPIAEKIFDVYYFSSCLSQKQIEEYNRIIGHYNLLINLYNQAKRSE
GKHLSANEKKYKDLPKFKTLYKQIGCGKKKDLFYTIKCDTEEEA
NKSRNEGKESHSVEEIINKAQEAINKYFKSNNDCENINTVPDFIN
YILTKENYEGVYWSKAAMNTISDKYFANYHDLQDRLKEAKVFQ

KADKKSEDDIKIPEAIELSGLFGVLDSLADWQTTLFKSSILSNED
KLKIITDSQTPSEALLKMIFNDIEKNMESFLKETNDIITLKKYKGN
KEGTEKIKQWFDYTLAINRMLKYFLVKENKIKGNSLDTNISEAL
KTLIYSDDAEWFKWYDALRNYLTQKPQDEAKENKLKLNFDNPS
LAGGWDVNKECSNFCVILKDKNEKKYLAIMKKGENTLFQKEW
TEGRGKNLTKKSNPLFEINNCEILSKMEYDFWADVSKMIPKCST
QLKAVVNHFKQSDNEFIFPIGYKVTSGEKFREECKISKQDFELNN
KVFNKNELSVTAMRYDLSSTQEKQYIKAFQKEYWELLFKQEKR
DTKLTNNEIFNEWINFCNKKYSELLSWERKYKDALTNWINFCKY
FLSKYPKTTLFNYSFKESENYNSLDEFYRDVDICSYKLNINTTIN
KSILDRLVEEGKLYLFEIKNQDSNDGKSIGHKNNLHTIYWNAIFE
NFDNRPKLNGEAEIFYRKAISKDKLGIVKGKKTKNGTEIIKNYRF
SKEKFILHVPITLNFCSNNEYVNDIVNTKFYNFSNLHFLGIDRGE
KHLAYYSLVNKNGEIVDQGTLNLPFTDKDGNQRSIKKEKYFYNK
QEDKWEAKEVDCWNYNDLLDAMASNRDMARKNWQRIGTIKE
AKNGYVSLVIRKIADLAVNNERPAFIVLEDLNTGFKRSRQKIDKS
VYQKFELALAKKLNFLVDKNAKRDEIGSPTKALQLTPPVNNYGD
IENKKQAGIMLYTRANYTSQTDPATGWRKTIYLKAGPEETTYKK
DGKIKNKSVKDQIIETFTDIGFDGKDYYFEYDKGEFVDEKTGEIK
PKKWRLYSGENGKSLDRFRGEREKDKYEWKIDKIDIVKILDDLF
VNFDKNISLLKQLKEGVELTRNNEHGTGESLRFAINLIQQIRNTG
NNERDNDFILSPVRDENGKHFDSREYWDKETKGEKISMPSSGDA
NGAFNIARKGIIMNAHILANSDSKDLSLFVSDEEWDLHLNNKTE
WKKQLNIFSSRKAMAKRKK (SEQ ID NO: 68)

[1442] >KKR91555_（修飾的）假定蛋白UU43_C0004G0003 [儉菌總門（佛爾扣菌（Falkowbacteria））細菌GW2011_GWA2_41_14]

[1443]

MLFFMSTDITNKPREKGVFDNFTNLYEFSKTLTFGLIPLKWD
DNKKMIVEDEDFSVLRKYGVIEEDKRIAESIKIAKFYLNILHRELI
GKVLGSLKFEKKNLENYDRLLGEIEKNNKNENISEDKKKEIRKN
FKKELSIAQDILLKKVGEVFESNGSGILSSKNCLDELTKRFTRQEV
DKLRRENKDIGVEYPDVAYREKDGKEETKSFFAMDVGYLDDFH
KNRKQLYSVKGKKNSLGRRILDNFEIFCKNKKLYEKYKNLDIDFS
EIERNFNLTLEKVFDFDNYNERLTQEGLDEYAKILGGESNKQERT
ANIHGLNQIINLYIQKKQSEQKAEQKETGKKKIKFNKKDYPTFTC
LQKQILSQVFRKEIIIESDRDLIRELKFFVEESKEKVDKARGIIEFL
LNHEENDIDLAMVYLPKSKINSFVYKVFKEPQDFLSVFQDGASN
LDFVSFDKIKTHLENNKLTYKIFFKTLIKENHDFESFLILLQQEIDL
LIDGGETVTLGGKKESITSLDEKKNRLKEKLGWFEGKVRENEKM
KDEEEGEFCSTVLAYSQAVLNITKRAEIFWLNEKQDAKVGEDNK
DMIFYKKFDEFADDGFAPFFYFDKFGNYLKRRSRNTTKEIKLHFG
NDDLLEGWDMNKEPEYWSFILRDRNQYYLGIGKKDGEIFHKKL
GNSVEAVKEAYELENEADFYEKIDYKQLNIDRFEGIAFPKKTKTE
EAFRQVCKKRADEFLGGDTYEFKILLAIKKEYDDFKARRQKEKD
WDSKFSKEKMSKLIEYYITCLGKRDDWKRFNLNFRQPKEYEDRS
DFVRHIQRQAYWIDPRKVSKDYVDKKVAEGEMFLFKVHNKDFY
DFERKSEDKKNHTANLFTQYLLELFSCENIKNIKSKDLIESIFELD

GKAEIRFRPKTDDVKLKIYQKKGKDVTYADKRDGNKEKEVIQH
RRFAKDALTLHLKIRLNFGKHVNLFDFNKLVNTELFAKVPVKIL
GMDRGENNLIYYCFLDEHGEIENGKCGSLNRVGEQIITLEDDKK
VKEPVDYFQLLVDREGQRDWEQKNWQKMTRIKDLKKAYLGNV
VSWISKEMLSGIKEGVVTIGVLEDLNSNFKRTRFFRERQVYQGF
EKALVNKLGYLVDKKYDNYRNVYQFAPIVDSVEEMEKNKQIGT
LVYVPASYTSKICPHPKCGWRERLYMKNSASKEKIVGLLKSDGI
KISYDQKNDRFYFEYQWEQEHKSDGKKKKYSGVDKVFSNVSR
MRWDVEQKKSIDFVDGTDGSITNKLKSLLKGKGIELDNINQQIV
NQQKELGVEFFQSIIFYFNLIMQIRNYDKEKSGSEADYIQCPSCLF
DSRKPEMNGKLSAITNGDANGAYNIARKGFMQLCRIRENPQEP
MKLITNREWDEAVREWDIYSAAQKIPVLSEEN(SEQ ID NO: 69)

[1444] >KDN25524_（修飾的）假定蛋白MBO_03467 [牛莫拉氏菌237]

[1445]

MLFQDFTHLYPLSKTVRFELKPIDRTLEHIHAKNFLSQDETM
ADMHQKVKVILDDYHRDFIADMMGEVKLTKLAEFYDVYLKFR
KNPKDDELQKQLKDLQAVLRKEIVKPIGNGGKYKAGYDRLFGA
KLFKDGKELGDLAKFVIAQEGESSPKLAHLAHFEKFSTYFTGFH
DNRKNMYSDEDKHTAIAYRLIHENLPRFIDNLQILTTIKQKHSAL
YDQIINELTASGLDVSLASHLDGYHKLLTQEGITAYNTLLGGISGE
AGSPKIQGINELINSHHNQHCHKSERIAKLRPLHKQILSDGMSVS
FLPSKFADDSEMCQAVNEFYRHYADVFAKVQSLFDGFDDHQKD

GIYVEHKNLNELSKQAFGDFALLGRVLDGYYVDVVNPEFNERFA
KAKTDNAKAKLTKEKDKFIKGVHSLASLEQAIEHYTARHDDESV
QAGKLGQYFKHGLAGVDNPIQKIHNNHSTIKGFLERERPAGERA
LPKIKSGKNPEMTQLRQLKELLDNALNVAHFAKLLTTKTTLDNQ
DGNFYGEFGVLYDELAKIPTLYNKVRDYLSQKPFSTEKYKLNFG
NPTLLNGWDLNKEKDNFGVILQKDGCYYLALLDKAHKKVFDN
APNTGKSIYQKMIYKYLEVRKQFPKVFFSKEAIAINYHPSKELVE
IKDKGRQRSDDERLKLYRFILECLKIHPKYDKKFEGAIGDIQLFK
KDKKGREVPISEKDLFDKINGIFSSKPKLEMEDFFIGEFKRYNPSQ
DLVDQYNIYKKIDSNDNRKKENFYNNHPKFKKDLVRYYYESMC
KHEEWEESFEFSKKLQDIGCYVDVNELFTEIETRRLNYKISFCNI
NADYIDELVEQGQLYLFQIYNKDFSPKAHGKPNLHTLYFKALFS
EDNLADPIYKLNGEAQIFYRKASLDMNETTIHRAGEVLENKNPD
NPKKRQFVYDIIKDKRYTQDKFMLHVPITMNFGVQGMTIKEFNK
KVNQSIQQYDEVNVIGIDRGERHLLYLTVINSKGEILEQCSLNDIT
TASANGTQMTTPYHKILDKREIERLNARVGWGEIETIKELKSGYL
SHVVHQISQLMLKYNAIVVLEDLNFGFKRGRFKVEKQIYQNFEN
ALIKKLNHLVLKDKADDEIGSYKNALQLTNNFTDLKSIGKQTGF
LFYVPAWNTSKIDPETGFVDLLKPRYENIAQSQAFFGKFDKICYN
ADKDYFEFHIDYAKFTDKAKNSRQIWTICSHGDKRYVYDKTAN
QNKGAAKGINVNDELKSLFARHHINEKQPNLVMDICQNNDKEF
HKSLMYLLKTLLALRYSNASSDEDFILSPVANDEGVFFNSALAD
DTQPQNADANGAYHIALKGLWLLNELKNSDDLNKVKLAIDNQT
WLNFAQNR (SEQ ID NO: 70)

[1446] >KKT48220_（修飾的）假定蛋白UW39_C0001G0044 [儉菌總門細菌GW2011_GWC2_44_17]

[1447]

MENIFDQFIGKYSLSKTLRFELKPVGKTEDFLKINKVFEKDQ
TIDDSYNQAKFYFDSLHQKFIDAALASDKTSELSFQNFADVLEKQ
NKIILDKKREMGALRKRDKNAVGIDRLQKEINDAEDIIQKEKEKI
YKDVRTLFDNEAESWKTYYQEREVDGKKITFSKADLKQKGADF
LTAAGILKVLKYEFPEEKEKEFQAKNQPSLFVEEKENPGQKRYIF
DSFDKFAGYLTKFQQTKKNLYAADGTSTAVATRIADNFIIFHQNT
KVFRDKYKNNHTDLGFDEENIFEIERYKNCLLQREIEHIKNENSY
NKIIGRINKKIKEYRDQKAKDTKLTKSDFPFFKNLDKQILGEVEK
EKQLIEKTREKTEEDVLIERFKEFIENNEERFTAAKKLMNAFCNG
EFESEYEGIYLKNKAINTISRRWFVSDRDFELKLPQQKSKNKSEK
NEPKVKKFISIAEIKNAVEELDGDIFKAVFYDKKIIAQGGSKLEQF
LVIWKYEFEYLFRDIERENGEKLLGYDSCLKIAKQLGIFPQEKEA
REKATAVIKNYADAGLGIFQMMKYFSLDDKDRKNTPGQLSTNFY
AEYDGYYKDFEFIKYYNEFRNFITKKPFDEDKIKLNFENGALLKG
WDENKEYDFMGVILKKEGRLYLGIMHKNHRKLFQSMGNAKGD
NANRYQKMIYKQIADASKDVPRLLLTSKKAMEKFKPSQEILRIK
KEKTFKRESKNFSLRDLHALIEYYRNCIPQYSNWSFYDFQFQDTG
KYQNIKEFTDDVQKYGYKISFRDIDDEYINQALNEGKMYLFEVV
NKDIYNTKNGSKNLHTLYFEHILSAENLNDPVFKLSGMAEIFQRQ
PSVNEREKITTQKNQCILDKGDRAYKYRRYTEKKIMFHMSLVLN
TGKGEIKQVQFNKIINQRISSSDNEMRVNVIGIDRGEKNLLYYSV

VKQNGEIIEQASLNEINGVNYRDKLIEREKERLKNRQSWKPVVK
IKDLKKGYISHVIHKICQLIEKYSAIVVLEDLNMRFKQIRGGIERS
VYQQFEKALIDKLGYLVFKDNRDLRAPGGVLNGYQLSAPFVSFE
KMRKQTGILFYTQAEYTSKTDPITGFRKNVYISNSASLDKIKEAV
KKFDAIGWDGKEQSYFFKYNPYNLADEKYKNSTVSKEWAIFAS
APRIRRQKGEDGYWKYDRVKVNEEFEKLLKVWNFVNPKATDIK
QEIIKKEKAGDLQGEKELDGRLRNFWHSFIYLFNLVLELRNSFSL
QIKIKAGEVIAVDEGVDFIASPVKPFFTTPNPYIPSNLCWLAVENA
DANGAYNIARKGVMILKKIREHAKKDPEFKKLPNLFISNAEWDE
AARDWGKYAGTTALNLDH（SEQ ID NO: 71）

[1448] >WP_031492824_（修飾的）假定蛋白[溶糊精琥珀酸弧菌]

[1449]

MSSLTKFTNKYSKQLTIKNELIPVGKTLENIKENGLIDGDEQL
NENYQKAKIIVDDFLRDFINKALNNTQIGNWRELADALNKEDED
NIEKLQDKIRGIIVSKFETFDLFSSYSIKKDEKIIDDDNDVEEEELD
LGKKTSSFKYIFKKNLFKLVLPSYLKTTNQDKLKIISSFDNFSTYF
RGFFENRKNIFTKKPISTSIAYRIVHDNFPKFLDNIRCFNVWQTEC
PQLIVKADNYLKSKNVIAKDKSLANYFTVGAYDYFLSQNGIDFY
NNIIGGLPAFAGHEKIQGLNEFINQECQKDSELKSKLKNRHAFKM
AVLFKQILSDREKSFVIDEFESDAQVIDAVKNFYAEQCKDNNVIF
NLLNLIKNIAFLSDDELDGIFIEGKYLSSVSQKLYSDWSKLRNDIE
DSANSKQGNKELAKKIKTNKGDVEKAISKYEFSLSELNSIVHDNT

KFSDLLSCTLHKVASEKLVKVNEGDWPKHLKNNEEKQKIKEPLD
ALLEIYNTLLIFNCKSFNKNGNFYVDYDRCINELSSVVYLYNKTR
NYCTKKPYNTDKFKLNFNSPQLGEGFSKSKENDCLTLLFKKDDN
YYVGIIRKGAKINFDDTQAIADNTDNCIFKMNYFLLKDAKKFIPK
CSIQLKEVKAHFKKSEDDYILSDKEKFASPLVIKKSTFLLATAHVK
GKKGNIKKFQKEYSKENPTEYRNSLNEWIAFCKEFLKTYKAATIF
DITTLKKAEEYADIVEFYKDVDNLCYKLEFCPIKTSFIENLIDNGD
LYLFRINNKDFSSKSTGTKNLHTLYLQAIFDERNLNNPTIMLNGG
AELFYRKESIEQKNRITHKAGSILVNKVCKDGTSLDDKIRNEIYQ
YENKFIDTLSDEAKKVLPNVIKKEATHDITKDKRFTSDKFFFHCP
LTINYKEGDTKQFNNEVLSFLRGNPDINIIGIDRGERNLIYVTVIN
QKGEILDSVSFNTVTNKSSKIEQTVDYEEKLAVREKERIEAKRSW
DSISKIATLKEGYLSAIVHEICLLMIKHNAIVVLENLNAGFKRIRG
GLSEKSVYQKFEKMLINKLNYFVSKKESDWNKPSGLLNGLQLSD
QFESFEKLGIQSGFIFYVPAAYTSKIDPTTGFANVLNLSKVRNVDA
IKSFFSNFNEISYSKKEALFKFSFDLDSLSKKGFSSFVKFSKSKWN
VYTFGERIIKPKNKQGYREDKRINLTFEMKKLLNEYKVSFDLEN
NLIPNLTSANLKDTFWKELFFIFKTTLQLRNSVTNGKEDVLISPVK
NAKGEFFVSGTHNKTLPQDCDANGAYHIALKGLMILERNNLVRE
EKDTKKIMAISNVDWFEYVQKRRGVL（SEQ ID NO: 72）

[1450] >KKT50231_（修飾的）假定蛋白UW40_C0007G0006 [儉菌總門細菌GW2011_GWF2_44_17]

[1451]

MKPVGKTEDFLKINKVFEKDQTIDDSYNQAKFYFDSLHQKF
IDAALASDKTSELSFQNFADVLEKQNKIILDKKREMGALRKRDK
NAVGIDRLQKEINDAEDIIQKEKEKIYKDVRTLFDNEAESWKTY
YQEREVDGKKITFSKADLKQKGADFLTAAGILKVLKYEFPEEKE
KEFQAKNQPSLFVEEKENPGQKRYIFDSFDKFAGYLTKFQQTKK
NLYAADGTSTAVATRIADNFIIFHQNTKVFRDKYKNNHTDLGFDE
ENIFEIERYKNCLLQREIEHIKNENSYNKIIGRINKKIKEYRDQKA
KDTKLTKSDFPFFKNLDKQILGEVEKEKQLIEKTREKTEEDVLIE
RFKEFIENNEERFTAAKKLMNAFCNGEFESEYEGIYLKNKAINTI
SRRWFVSDRDFELKLPQQKSKNKSEKNEPKVKKFISIAEIKNAVE
ELDGDIFKAVFYDKKIIAQGGSKLEQFLVIWKYEFEYLFRDIERE
NGEKLLGYDSCLKIAKQLGIFPQEKEAREKATAVIKNYADAGLGI
FQMMKYFSLDDKDRKNTPGQLSTNFYAEYDGYYKDFEFIKYYN
EFRNFITKKPFDEDKIKLNFENGALLKGWDENKEYDFMGVILKK
EGRLYLGIMHKNHRKLFQSMGNAKGDNANRYQKMIYKQIADA
SKDVPRLLLTSKKAMEKFKPSQEILRIKKEKTFKRESKNFSLRDL
HALIEYYRNCIPQYSNWSFYDFQFQDTGKYQNIKEFTDDVQKYG
YKISFRDIDDEYINQALNEGKMYLFEVVNKDIYNTKNGSKNLHT
LYFEHILSAENLNDPVFKLSGMAEIFQRQPSVNEREKITTQKNQC
ILDKGDRAYKYRRYTEKKIMFHMSLVLNTGKGEIKQVQFNKIIN
QRISSSDNEMRVNVIGIDRGEKNLLYYSVVKQNGEIIEQASLNEI
NGVNYRDKLIEREKERLKNRQSWKPVVKIKDLKKGYISHVIHKI
CQLIEKYSAIVVLEDLNMRFKQIRGGIERSVYQQFEKALIDKLGY
LVFKDNRDLRAPGGVLNGYQLSAPFVSFEKMRKQTGILFYTQAE

YTSKTDPITGFRKNVYISNSASLDKIKEAVKKFDAIGWDGKEQS
YFFKYNPYNLADEKYKNSTVSKEWAIFASAPRIRRQKGEDGYW
KYDRVKVNEEFEKLLKVWNFVNPKATDIKQEIIKKEKAGDLQGE
KELDGRLRNFWHSFIYLFNLVLELRNSFSLQIKIKAGEVIAVDEG
VDFIASPVKPFFTTPNPYIPSNLCWLAVENADANGAYNIARKGV
MILKKIREHAKKDPEFKKLPNLFISNAEWDEAARDWGKYAGTT
ALNLDH（SEQ ID NO: 73）

[1452] >WP_004356401_（修飾的）假定蛋白[解糖腺普雷沃菌]

[1453]

MENYQEFTNLFQLNKTLRFELKPIGKTCELLEEGKIFASGSF
LEKDKVRADNVSYVKKEIDKKHKIFIEETLSSFSISNDLLKQYFD
CYNELKAFKKDCKSDEEEVKKTALRNKCTSIQRAMREAISQAFL
KSPQKKLLAIKNLIENVFKADENVQHFSEFTSYFSGFETNRENFY
SDEEKSTSIAYRLVHDNLPIFIKNIYIFEKLKEQFDAKTLSEIFENY
KLYVAGSSLDEVFSLEYFNNTLTQKGIDNYNAVIGKIVKEDKQEI
QGLNEHINLYNQKHKDRRLPFFISLKKQILSDREALSWLPDMFK
NDSEVIKALKGFYIEDGFENNVLTPLATLLSSLDKYNLNGIFIRN
NEALSSLSQNVYRNFSIDEAIDANAELQTFNNYELIANALRAKIK
KETKQGRKSFEKYEEYIDKKVKAIDSLSIQEINELVENYVSEFNS
NSGNMPRKVEDYFSLMRKGDFGSNDLIENIKTKLSAAEKLLGTK
YQETAKDIFKKDENSKLIKELLDATKQFQHFIKPLLGTGEEADRD
LVFYGDFLPLYEKFEELTLLYNKVRNRLTQKPYSKDKIRLCFNKP

KLMTGWVDSKTEKSDNGTQYGGYLFRKKNEIGEYDYFLGISSK
AQLFRKNEAVIGDYERLDYYQPKANTIYGSAYEGENSYKEDKK
RLNKVIIAYIEQIKQTNIKKSIIESISKYPNISDDDKVTPSSLLEKIK
KVSIDSYNGILSFKSFQSVNKEVIDNLLKTISPLKNKAEFLDLINK
DYQIFTEVQAVIDEICKQKTFIYFPISNVELEKEMGDKDKPLCLF
QISNKDLSFAKTFSANLRKKRGAENLHTMLFKALMEGNQDNLD
LGSGAIFYRAKSLDGNKPTHPANEAIKCRNVANKDKVSLFTYDI
YKNRRYMENKFLFHLSIVQNYKAANDSAQLNSSATEYIRKADD
LHIIGIDRGERNLLYYSVIDMKGNIVEQDSLNIIRNNDLETDYHD
LLDKREKERKANRQNWEAVEGIKDLKKGYLSQAVHQIAQLMLK
YNAIIALEDLGQMFVTRGQKIEKAVYQQFEKSLVDKLSYLVDKK
RPYNELGGILKAYQLASSITKNNSDKQNGFLFYVPAWNTSKIDPV
TGFTDLLRPKAMTIKEAQDFFGAFDNISYNDKGYFEFETNYDKF
KIRMKSAQTRWTICTFGNRIKRKKDKNYWNYEEVELTEEFKKLF
KDSNIDYENCNLKEEIQNKDNRKFFDDLIKLLQLTLQMRNSDDK
GNDYIISPVANAEGQFFDSRNGDKKLPLDADANGAYNIARKGLW
NIRQIKQTKNDKKLNLSISSTEWLDFVREKPYLK（SEQ ID NO: 74）

[1454] >CCB70584_（修飾的）未知功能蛋白質[嗜鰓黃桿菌（Flavobacterium branchiophilum） FL-15]

[1455]

MTNKFTNQYSLSKTLRFELIPQGKTLEFIQEKGLLSQDKQRA
ESYQEMKKTIDKFHKYFIDLALSNAKLTHLETYLELYNKSAETK
KEQKFKDDLKKVQDNLRKEIVKSFSDGDAKSIFAILDKKELITVE

LEKWFENNEQKDIYFDEKFKTFTTYFTGFHQNRKNMYSVEPNST
AIAYRLIHENLPKFLENAKAFEKIKQVESLQVNFRELMGEFGDEG
LIFVNELEEMFQINYYNDVLSQNGITIYNSIISGFTKNDIKYKGLN
EYINNYNQTKDKKDRLPKLKQLYKQILSDRISLSFLPDAFTDGKQ
VLKAIFDFYKINLLSYTIEGQEESQNLLLLIRQTIENLSSFDTQKIY
LKNDTHLTTISQQVFGDFSVFSTALNYWYETKVNPKFETEYSKA
NEKKREILDKAKAVFTKQDYFSIAFLQEVLSEYILTLDHTSDIVK
KHSSNCIADYFKNHFVAKKENETDKTFDFIANITAKYQCIQGILE
NADQYEDELKQDQKLIDNLKFFLDAILELLHFIKPLHLKSESITE
KDTAFYDVFENYYEALSLLTPLYNMVRNYVTQKPYSTEKIKLNF
ENAQLLNGWDANKEGDYLTTILKKDGNYFLAIMDKKHNKAFQ
KFPEGKENYEKMVYKLLPGVNKMLPKVFFSNKNIAYFNPSKELL
ENYKKETHKKGDTFNLEHCHTLIDFFKDSLNKHEDWKYFDFQF
SETKSYQDLSGFYREVEHQGYKINFKNIDSEYIDGLVNEGKLFLF
QIYSKDFSPFSKGKPNMHTLYWKALFEEQNLQNVIYKLNGQAEI
FFRKASIKPKNIILHKKKIKIAKKHFIDKKTKTSEIVPVQTIKNLN
MYYQGKISEKELTQDDLRYIDNFSIFNEKNKTIDIIKDKRFTVDK
FQFHVPITMNFKATGGSYINQTVLEYLQNNPEVKIIGLDRGERHL
VYLTLIDQQGNILKQESLNTITDSKISTPYHKLLDNKENERDLAR
KNWGTVENIKELKEGYISQVVHKIATLMLEENAIVVMEDLNFGF
KRGRFKVEKQIYQKLEKMLIDKLNYLVLKDKQPQELGGLYNAL
QLTNKFESFQKMGKQSGFLFYVPAWNTSKIDPTTGFVNYFYTKY
ENVDKAKAFFEKFEAIRFNAEKKYFEFEVKKYSDFNPKAEGTQQ
AWTICTYGERIETKRQKDQNNKFVSTPINLTEKIEDFLGKNQIVY

GDGNCIKSQIASKDDKAFFETLLYWFKMTLQMRNSETRTDIDYL
ISPVMNDNGTFYNSRDYEKLENPTLPKDADANGAYHIAKKGLM
LLNKIDQADLTKKVDLSISNRDWLQFVQKNK（SEQ ID NO: 75）

[1456] >WP_005398606_（修飾的）假定蛋白[孔茲氏創傷球菌（Helcococcus kunzii）]

[1457]

MFEKLSNIVSISKTIRFKLIPVGKTLENIEKLGKLEKDFERSD
FYPILKNISDDYYRQYIKEKLSDLNLDWQKLYDAHELLDSSKKE
SQKNLEMIQAQYRKVLFNILSGELDKSGEKNSKDLIKNNKALYG
KLFKKQFILEVLPDFVNNNDSYSEEDLEGLNLYSKFTTRLKNFW
ETRKNVFTDKDIVTAIPFRAVNENFGFYYDNIKIFNKNIEYLENKI
PNLENELKEADILDDNRSVKDYFTPNGFNYVITQDGIDVYQAIR
GGFTKENGEKVQGINEILNLTQQQLRRKPETKNVKLGVLTKLRK
QILEYSESTSFLIDQIEDDNDLVDRINKFNVSFFESTEVSPSLFEQI
ERLYNALKSIKKEEVYIDARNTQKFSQMLFGQWDVIRRGYTVKI
TEGSKEEKKKYKEYLELDETSKAKRYLNIREIEELVNLVEGFEEV
DVFSVLLEKFKMNNIERSEFEAPIYGSPIKLEAIKEYLEKHLEEYH
KWKLLLIGNDDLDTDETFYPLLNEVISDYYIIPLYNLTRNYLTRK
HSDKDKIKVNFDFPTLADGWSESKISDNRSIILRKGGYYYLGILI
DNKLLINKKNKSKKIYEILIYNQIPEFSKSIPNYPFTKKVKEHFKN
NVSDFQLIDGYVSPLIITKEIYDIKKEKKYKKDFYKDNNTNKNYL
YTIYKWIEFCKQFLYKYKGPNKESYKEMYDFSTLKDTSLYVNLN
DFYADVNSCAYRVLFNKIDENTIDNAVEDGKLLLFQIYNKDFSPE

SKGKKNLHTLYWLSMFSEENLRTRKLKLNGQAEIFYRKKLEKKP
IIHKEGSILLNKIDKEGNTIPENIYHECYRYLNKKIGREDLSDEAI
ALFNKDVLKYKEARFDIIKDRRYSESQFFFHVPITFNWDIKTNKN
VNQIVQGMIKDGEIKHIIGIDRGERHLLYYSVIDLEGNIVEQGSLN
TLEQNRFDNSTVKVDYQNKLRTREEDRDRARKNWTNINKIKEL
KDGYLSHVVHKLSRLIIKYEAIVIMENLNQGFKRGRFKVERQVY
QKFELALMNKLSALSFKEKYDERKNLEPSGILNPIQACYPVDAY
QELQGQNGIVFYLPAAYTSVIDPVTGFTNLFRLKSINSSKYEEFIK
KFKNIYFDNEEEDFKFIFNYKDFAKANLVILNNIKSKDWKISTRG
ERISYNSKKKEYFYVQPTEFLINKLKELNIDYENIDIIPLIDNLEEK
AKRKILKALFDTFKYSVQLRNYDFENDYIISPTADDNGNYYNSN
EIDIDKTNLPNNGDANGAFNIARKGLLLKDRIVNSNESKVDLKIK
NEDWINFIIS（SEQ ID NO: 76）

[1458] >WP_021736722_（修飾的）CRISPR相關蛋白Cpf1，PREFRAN亞型[胺基酸球菌屬某種BV3L6]

[1459]

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKA
RNDHYKELKPIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKE
KTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKG
LFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYEN
RKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHF
ENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISRE
AGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRN

TLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSID
LTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAK
EKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALD
QPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEF
SARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTL
ASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPT
EKTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPIL
LSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREA
LCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPL
LYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHT
LYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEK
MLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPN
VITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVN
AYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDY
QKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMI
HYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVL
KDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSK
IDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHF
KMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIV
PVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLEND
DSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDS
RFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGIS
NQDWLAYIQELRN （SEQ ID NO: 77）

[1460] >WP_004339290_（修飾的）假定蛋白[土拉熱弗朗西

絲菌]

[1461]

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEK
RAKDYKKAKQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLK
KSDDDNLQKDFKSAKDTIKKQISKYINDSEKFKNLFNQNLIDAK
KGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT
TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESL
KDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEI
ANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQI
NDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLT
DLSQQVFDDYSVIGTAVLEYITQQVAPKNLDNPSKKEQDLIAKK
TEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILSNFAAIPMIFD
EIAQNKDNLAQISIKYQNQGKKDLLQASAEEDVKAIKDLLDQTN
NLLHRLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYN
KIRNYITQKPYSDEKFKLNFENSTLASGWDKNKESANTAILFIKD
DKYYLGIMDKKHNKIFSDKAIEENKGEGYKKIVYKQIADASKDI
QNLMIIDGKTVCKKGRKDRNGVNRQLLSLKRKHLPENIYRIKET
KSYLKNEARFSRKDLYDFIDYYKDRLDYYDFEFELKPSNEYSDF
NDFTNHIGSQGYKLTFENISQDYINSLVNEGKLYLFQIYSKDFSAY
SKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPK
KITHPAKETIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPI
TINFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLV
DGKGNIIKQDNFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWK

KINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRF
KVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAP
FETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKS
QEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGS
RLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKA
AICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNG
NFFDSRQAPKNMPQDADANGAYHIGLKGLMLLDRIKNNQEGKK
LNLVIKNEEYFEFVQNRNN（SEQ ID NO: 78）

[1462] >WP_022501477_（修飾的）假定蛋白[真細菌某種 CAG:76]

[1463]

MNKAADNYTGGNYDEFIALSKVQKTLRNELKPTPFTAEHIK
QRGIISEDEYRAQQSLELKKIADEYYRNYITHKLNDINNLDFYNL
FDAIEEKYKKNDKDNRDKLDLVEKSKRGEIAKMLSADDNFKSM
FEAKLITKLLPDYVERNYTGEDKEKALETLALFKGFTTYFKGYF
KTRKNMFSGEGGASSICHRIVNVNASIFYDNLKTFMRIQEKAGD
EIALIEEELTEKLDGWRLEHIFSRDYYNEVLAQKGIDYYNQICGD
INKHMNLYCQQNKFKANIFKMMKIQKQIMGISEKAFEIPPMYQN
DEEVYASFNEFISRLEEVKLTDRLINILQNINIYNTAKIYINARYYT
NVSSYVYGGWGVIDSAIERYLYNTIAGKGQSKVKKIENAKKDN
KFMSVKELDSIVAEYEPDYFNAPYIDDDDNAVKAFGGQGVLGYF
NKMSELLADVSLYTIDYNSDDSLIENKESALRIKKQLDDIMSLYH
WLQTFIIDEVVEKDNAFYAELEDICCELENVVTLYDRIRNYVTKK

PYSTQKFKLNFASPTLAAGWSRSKEFDNNAIILLRNNKYYIAIFN
VNNKPDKQIIKGSEEQRLSTDYKKMVYNLLPGPNKMLPKVFIKS
DTGKRDYNPSSYILEGYEKNRHIKSSGNFDINYCHDLIDYYKACI
NKHPEWKNYGFKFKETNQYNDIGQFYKDVEKQGYSISWAYISEE
DINKLDEEGKIYLFEIYNKDLSAHSTGRDNLHTMYLKNIFSEDNL
KNICIELNGEAELFYRKSSMKSNITHKKDTILVNKTYINETGVRV
SLSDEDYMKVYNYYNNNYVIDTENDKNLIDIIEKIGHRKSKIDIV
KDKRYTEDKYFLYLPITINYGIEDENVNSKIIEYIAKQDNMNVIGI
DRGERNLIYISVIDNKGNIIEQKSFNLVNNYDYKNKLKNMEKTR
DNARKNWQEIGKIKDVKSGYLSGVISKIARMVIDYNAIIVMEDL
NKGFKRGRFKVERQVYQKFENMLISKLNYLVFKERKADENGGI
LRGYQLTYIPKSIKNVGKQCGCIFYVPAAYTSKIDPATGFINIFDF
KKYSGSGINAKVKDKKEFLMSMNSIRYINECSEEYEKIGHRELFA
FSFDYNNFKTYNVSSPVNEWTAYTYGERIKKLYKDGRWLRSEVL
NLTENLIKLMEQYNIEYKDGHDIREDISHMDETRNADFICSLFEE
LKYTVQLRNSKSEAEDENYDRLVSPILNSSNGFYDSSDYMENEN
NTTHTMPKDADANGAYCIALKGLYEINKIKQNWSDDKKFKENE
LYINVTEWLDYIQNRRFE（SEQ ID NO: 79）

[1464] >WP_014550095_（修飾的）假定蛋白[土拉熱弗朗西絲菌]

[1465]

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEK
RAKDYKKAKQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLK

KSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAK
KGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT
TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESL
KDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEI
ANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQI
NDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLT
DLSQQVFDDYSVIGTAVLEYITQQVAPKNLDNPSKKEQDLIAKK
TEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFD
EIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTN
NLLHRLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYN
KIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKD
DKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANK
MLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGNPQKGYEKFEF
NIEDCRKFIDFYKESISKHPEWKDFGFRFSDTQRYNSIDEFYREV
ENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPN
LHTLYWKALFDERNLQDVVYKLNGEAELFYRKKSIPKKITHPAK
EAIANKNKDNPKKESFFEYDLIKDKRFTEDKFFFHCPITINFKSSG
ANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIK
QDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEM
KEGYLSQVVHEIAKLVIEHNAIVVFEDLNFGFKRGRFKVEKQVY
QKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMG
KQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFD
KICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNS

DKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDK
KFFAKLTSILNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQA
PKNMPQDADANGAYHIGLKGLMLLDRIKNNQEGKKLNLVIKNE
EYFEFVQNRNN（SEQ ID NO: 80）

[1466] >WP_003034647_（修飾的）假定蛋白[土拉熱弗朗西絲菌]

[1467]

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEK
RAKDYKKAKQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLK
KSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAK
KGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT
TYFKGFHENRKNVYSSDDIPTSIIYRIVDDNLPKFLENKAKYESL
KDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEI
ANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQI
NDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLT
DLSQQVFDDYSVIGTAVLEYITQQVAPKNLDNPSKKEQDLIAKK
TEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFD
EIAQNKDNLAQISLKYQNQGKKDLLQASAEEDVKAIKDLLDQT
NNLLHRLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLY
NKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIK
DDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGAN
KMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGNPQKGYEKFE

FNIEDCRKFIDFYKESISKHPEWKDFGFRFSDTQRYNSIDEFYREV
ENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPN
LHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAK
EAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSS
GANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNII
KQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKE
MKEGYLSQVVHEIAKLVIEHNAIVVFEDLNFGFKRGRFKVEKQV
YQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKM
GKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKF
DKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRN
SDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESD
KKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQ
APKNMPQDADANGAYHIGLKGLMLLDRIKNNQEGKKLNLVIKN
EEYFEFVQNRNN（SEQ ID NO: 81）

[1468] >FnCpf1土拉熱弗朗西絲菌新殺手亞種U112，全基因組

[1469]

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEK
RAKDYKKAKQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLK
KSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAK
KGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT
TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESL
KDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEI

ANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQI
NDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLT
DLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKT
EKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDE
IAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTN
NLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLY
NKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIK
DDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGAN
KMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFE
FNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYRE
VENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRP
NLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPA
KEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKS
SGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGN
IIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIK
EMKEGYLSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKVEKQ
VYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKK
MGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSK
FDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFR
NSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGES
DKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSR
QAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIK
NEEYFEFVQNRNN (SEQ ID NO: 82)

[1470] >KKQ38174_（修飾的）假定蛋白US54_C0016G0015[Microgenomates(羅斯曼菌(Roizmanbacteria))細菌GW2011_GWA2_37_7]

[1471]

MKSFDSFTNLYSLSKTLKFEMRPVGNTQKMLDNAGVFEKD
KLIQKKYGKTKPYFDRLHREFIEEALTGVELIGLDENFRTLVDWQ
KDKKNNVAMKAYENSLQRLRTEIGKIFNLKAEDWVKNKYPILG
LKNKNTDILFEEAVFGILKARYGEEKDTFIEVEEIDKTGKSKINQI
SIFDSWKGFTGYFKKFFETRKNFYKNDGTSTAIATRIIDQNLKRFI
DNLSIVESVRQKVDLAETEKSFSISLSQFFSIDFYNKCLLQDGIDY
YNKIIGGETLKNGEKLIGLNELINQYRQNNKDQKIPFFKLLDKQI
LSEKILFLDEIKNDTELIEALSQFAKTAEEKTKIVKKLFADFVENN
SKYDLAQIYISQEAFNTISNKWTSETETFAKYLFEAMKSGKLAK
YEKKDNSYKFPDFIALSQMKSALLSISLEGHFWKEKYYKISKFQ
EKTNWEQFLAIFLYEFNSLFSDKINTKDGETKQVGYYLFAKDLH
NLILSEQIDIPKDSKVTIKDFADSVLTIYQMAKYFAVEKKRAWLA
EYELDSFYTQPDTGYLQFYDNAYEDIVQVYNKLRNYLTKKPYSE
EKWKLNFENSTLANGWDKNKESDNSAVILQKGGKYYLGLITKG
HNKIFDDRFQEKFIVGIEGGKYEKIVYKFFPDQAKMFPKVCFSA
KGLEFFRPSEEILRIYNNAEFKKGETYSIDSMQKLIDFYKDCLTK
YEGWACYTFRHLKPTEEYQNNIGEFFRDVAEDGYRIDFQGISDQ
YIHEKNEKGELHLFEIHNKDWNLDKARDGKSKTTQKNLHTLYF
ESLFSNDNVVQNFPIKLNGQAEIFYRPKTEKDKLESKKDKKGNK
VIDHKRYSENKIFFHVPLTLNRTKNDSYRFNAQINNFLANNKDIN

IIGVDRGEKHLVYYSVITQASDILESGSLNELNGVNYAEKLGKKA
ENREQARRDWQDVQGIKDLKKGYISQVVRKLADLAIKHNAIIIL
EDLNMRFKQVRGGIEKSIYQQLEKALIDKLSFLVDKGEKNPEQA
GHLLKAYQLSAPFETFQKMGKQTGIIFYTQASYTSKSDPVTGWR
PHLYLKYFSAKKAKDDIAKFTKIEFVNDRFELTYDIKDFQQAKE
YPNKTVWKVCSNVERFRWDKNLNQNKGGYTHYTNITENIQELF
TKYGIDITKDLLTQISTIDEKQNTSFFRDFIFYFNLICQIRNTDDSEI
AKKNGKDDFILSPVEPFFDSRKDNGNKLPENGDDNGAYNIARKG
IVILNKISQYSEKNENCEKMKWGDLYVSNIDWDNFVTQANARH
（SEQ ID NO: 83）

[1472] >WP_022097749_（修飾的）假定蛋白[挑剔真細菌CAG:72]

[1473]

MNGNRSIVYREFVGVTPVAKTLRNELRPVGHTQEHIIQNGLI
QEDELRQEKSTELKNIMDDYYREYIDKSLSGLTDLDFTLLFELM
NSVQSSLSKDNKKALEKEHNKMREQICTHLQSDSDYKNMFNAK
LFKEILPDFIKNYNQYDVKDKAGKLETLALFNGFSTYFTDFFEKR
KNVFTKEAVSTSIAYRIVHENSLIFLANMTSYKKISEKALDEIEVI
EKNNQDKMGDWELNQIFNPDFYNMVLIQSGIDFYNEICGVVNA
HMNLYCQQTKNNYNLFKMRKLHKQILAYTSTSFEVPKMFEDDM
SVYNAVNAFIDETEKGNIIGKLKDIVNKYDELDEKRIYISKDFYE
TLSCFMSGNWNLITGCVENFYDENIHAKGKSKEEKVKKAVKED
KYKSINDVNDLVEKYIDEKERNEFKNSNAKQYIREISNIITDTETA

HLEYDEHISLIESEEKADEIKKRLDMYMNMYHWVKAFIVDEVL
DRDEMFYSDIDDIYNILENIVPLYNRVRNYVTQKPYTSKKIKLNF
QSPTLANGWSQSKEFDNNAIILIRDNKYYLAIFNAKNKPDKKIIQ
GNSDKKNDNDYKKMVYNLLPGANKMLPKVFLSKKGIETFKPSD
YIISGYNAHKHIKTSENFDISFCRDLIDYFKNSIEKHAEWRKYEF
KFSATDSYNDISEFYREVEMQGYRIDWTYISEADINKLDEEGKIY
LFQIYNKDFAENSTGKENLHTMYFKNIFSEENLKNIVIKLNGQAE
LFYRKASVKNPVKHKKDSVLVNKTYKNQLDNGDVVRIPIPDDIY
NEIYKMYNGYIKESDLSEAAKEYLDKVEVRTAQKDIVKDYRYT
VDKYFIHTPITINYKVTARNNVNDMAVKYIAQNDDIHVIGIDRGE
RNLIYISVIDSHGNIVKQKSYNILNNYDYKKKLVEKEKTREYARK
NWKSIGNIKELKEGYISGVVHEIAMLMVEYNAIIAMEDLNYGFK
RGRFKVERQVYQKFESMLINKLNYFASKGKSVDEPGGLLKGYQ
LTYVPDNIKNLGKQCGVIFYVPAAFTSKIDPSTGFISAFNFKSIST
NASRKQFFMQFDEIRYCAEKDMFSFGFDYNNFDTYNITMGKTQ
WTVYTNGERLQSEFNNARRTGKTKSINLTETIKLLLEDNEINYAD
GHDVRIDMEKMYEDKNSEFFAQLLSLYKLTVQMRNSYTEAEEQ
EKGISYDKIISPVINDEGEFFDSDNYKESDDKECKMPKDADANG
AYCIALKGLYEVLKIKSEWTEDGFDRNCLKLPHAEWLDFIQNKR
YE（SEQ ID NO: 84）

[1474] >WP_012739647_(修飾的)假定蛋白 [[真細菌]挑剔]

[1475]

MNGNRSIVYREFVGVIPVAKTLRNELRPVGHTQEHIIQNGLI

QEDELRQEKSTELKNIMDDYYREYIDKSLSGVTDLDFTLLFELM
NLVQSSPSKDNKKALEKEQSKMREQICTHLQSDSNYKNIFNAKL
LKEILPDFIKNYNQYDVKDKAGKLETLALFNGFSTYFTDFFEKR
KNVFTKEAVSTSIAYRIVHENSLIFLANMTSYKKISEKALDEIEVI
EKNNQDKMGDWELNQIFNPDFYNMVLIQSGIDFYNEICGVVNA
HMNLYCQQTKNNYNLFKMRKLHKQILAYTSTSFEVPKMFEDDM
SVYNAVNAFIDETEKGNIIGKLKDIVNKYDELDEKRIYISKDFYE
TLSCFMSGNWNLITGCVENFYDENIHAKGKSKEEKVKKAVKED
KYKSINDVNDLVEKYIDEKERNEFKNSNAKQYIREISNIITDTETA
HLEYDDHISLIESEEKADEMKKRLDMYMNMYHWAKAFIVDEVL
DRDEMFYSDIDDIYNILENIVPLYNRVRNYVTQKPYNSKKIKLNF
QSPTLANGWSQSKEFDNNAIILIRDNKYYLAIFNAKNKPDKKIIQ
GNSDKKNDNDYKKMVYNLLPGANKMLPKVFLSKKGIETFKPSD
YIISGYNAHKHIKTSENFDISFCRDLIDYFKNSIEKHAEWRKYEF
KFSATDSYSDISEFYREVEMQGYRIDWTYISEADINKLDEEGKIY
LFQIYNKDFAENSTGKENLHTMYFKNIFSEENLKDIIIKLNGQAE
LFYRRASVKNPVKHKKDSVLVNKTYKNQLDNGDVVRIPIPDDIY
NEIYKMYNGYIKESDLSEAAKEYLDKVEVRTAQKDIVKDYRYT
VDKYFIHTPITINYKVTARNNVNDMVVKYIAQNDDIHVIGIDRGE
RNLIYISVIDSHGNIVKQKSYNILNNYDYKKKLVEKEKTREYARK
NWKSIGNIKELKEGYISGVVHEIAMLIVEYNAIIAMEDLNYGFKR
GRFKVERQVYQKFESMLINKLNYFASKEKSVDEPGGLLKGYQLT
YVPDNIKNLGKQCGVIFYVPAAFTSKIDPSTGFISAFNFKSISTNA
SRKQFFMQFDEIRYCAEKDMFSFGFDYNNFDTYNITMGKTQWT

VYTNGERLQSEFNNARRTGKTKSINLTETIKLLLEDNEINYADGH
DIRIDMEKMDEDKKSEFFAQLLSLYKLTVQMRNSYTEAEEQENG
ISYDKIISPVINDEGEFFDSDNYKESDDKECKMPKDADANGAYCI
ALKGLYEVLKIKSEWTEDGFDRNCLKLPHAEWLDFIQNKRYE
（SEQ ID NO: 85）

[1476] >WP_045971446_（修飾的）假定蛋白[黃桿菌屬某種316]

[1477]

MKNFSNLYQVSKTVRFELKPIGNTLENIKNKSLLKNDSIRAE
SYQKMKKTIDEFHKYFIDLALNNKKLSYLNEYIALYTQSAEAKK
EDKFKADFKKVQDNLRKEIVSSFTEGEAKAIFSVLDKKELITIEL
EKWKNENNLAVYLDESFKSFTTYFTGFHQNRKNMYSAEANSTA
IAYRLIHENLPKFIENSKAFEKSSQIAELQPKIEKLYKEFEAYLNV
NSISELFEIDYFNEVLTQKGITVYNNIIGGRTATEGKQKIQGLNEII
NLYNQTKPKNERLPKLKQLYKQILSDRISLSFLPDAFTEGKQVLK
AVFEFYKINLLSYKQDGVEESQNLLELIQQVVKNLGNQDVNKIY
LKNDTSLTTIAQQLFGDFSVFSAALQYRYETVVNPKYTAEYQKA
NEAKQEKLDKEKIKFVKQDYFSIAFLQEVVADYVKTLDENLDW
KQKYTPSCIADYFTTHFIAKKENEADKTFNFIANIKAKYQCIQGI
LEQADDYEDELKQDQKLIDNIKFFLDAILEVVHFIKPLHLKSESIT
EKDNAFYDVFENYYEALNVVTPLYNMVRNYVTQKPYSTEKIKL
NFENAQLLNGWDANKEKDYLTTILKRDGNYFLAIMDKKHNKTF
QQFTEDDENYEKIVYKLLPGVNKMLPKVFFSNKNIAFFNPSKEIL

DNYKNNTHKKGATFNLKDCHALIDFFKDSLNKHEDWKYFDFQF
SETKTYQDLSGFYKEVEHQGYKINFKKVSVSQIDTLIEEGKMYL
FQIYNKDFSPYAKGKPNMHTLYWKALFETQNLENVIYKLNGQA
EIFFRKASIKKKNIITHKAHQPIAAKNPLTPTAKNTFAYDLIKDKR
YTVDKFQFHVPITMNFKATGNSYINQDVLAYLKDNPEVNIIGLD
RGERHLVYLTLIDQKGTILLQESLNVIQDEKTHTPYHTLLDNKEI
ARDKARKNWGSIESIKELKEGYISQVVHKITKMMIEHNAIVVME
DLNFGFKRGRFKVEKQIYQKLEKMLIDKLNYLVLKDKQPHELG
GLYNALQLTNKFESFQKMGKQSGFLFYVPAWNTSKIDPTTGFVN
YFYTKYENVEKAKTFFSKFDSILYNKTKGYFEFVVKNYSDFNPK
AADTRQEWTICTHGERIETKRQKEQNNNFVSTTIQLTEQFVNFFE
KVGLDLSKELKTQLIAQNEKSFFEELFHLLKLTLQMRNSESHTEI
DYLISPVANEKGIFYDSRKATASLPIDADANGAYHIAKKGLWIME
QINKTNSEDDLKKVKLAISNREWLQYVQQVQKK（SEQ ID NO:
86）

[1478] >WP_044110123_（修飾的）假定蛋白[短普雷沃菌（Prevotella brevis）]

[1479]

MKQFTNLYQLSKTLRFELKPIGKTLEHINANGFIDNDAHRA
ESYKKVKKLIDDYHKDYIENVLNNFKLNGEYLQAYFDLYSQDT
KDKQFKDIQDKLRKSIASALKGDDRYKTIDKKELIRQDMKTFLK
KDTDKALLDEFYEFTTYFTGYHENRKNMYSDEAKSTAIAYRLIH
DNLPKFIDNIAVFKKIANTSVADNFSTIYKNFEEYLNVNSIDEIFSL

DYYNIVLTQTQIEVYNSIIGGRTLEDDTKIQGINEFVNLYNQQLA
NKKDRLPKLKPLFKQILSDRVQLSWLQEEFNTGADVLNAVKEYC
TSYFDNVEESVKVLLTGISDYDLSKIYITNDLALTDVSQRMFGE
WSIIPNAIEQRLRSDNPKKTNEKEEKYSDRISKLKKLPKSYSLGYI
NECISELNGIDIADYYATLGAINTESKQEPSIPTSIQVHYNALKPIL
DTDYPREKNLSQDKLTVMQLKDLLDDFKALQHFIKPLLGNGDE
AEKDEKFYGELMQLWEVIDSITPLYNKVRNYCTRKPFSTEKIKV
NFENAQLLDGWDENKESTNASIILRKNGMYYLGIMKKEYRNILT
KPMPSDGDCYDKVVYKFFKDITTMVPKCTTQMKSVKEHFSNSN
DDYTLFEKDKFIAPVVITKEIFDLNNVLYNGVKKFQIGYLNNTG
DSFGYNHAVEIWKSFCLKFLKAYKSTSIYDFSSIEKNIGCYNDLN
SFYGAVNLLLYNLTYRKVSVDYIHQLVDEDKMYLFMIYNKDFST
YSKGTPNMHTLYWKMLFDESNLNDVVYKLNGQAEVFYRKKSIT
YQHPTHPANKPIDNKNVNNPKKQSNFEYDLIKDKRYTVDKFMF
HVPITLNFKGMGNGDINMQVREYIKTTDDLHFIGIDRGERHLLYI
CVINGKGEIVEQYSLNEIVNNYKGTEYKTDYHTLLSERDKKRKE
ERSSWQTIEGIKELKSGYLSQVIHKITQLMIKYNAIVLLEDLNMG
FKRGRQKVESSVYQQFEKALIDKLNYLVDKNKDANEIGGLLHA
YQLTNDPKLPNKNSKQSGFLFYVPAWNTSKIDPVTGFVNLLDTR
YENVAKAQAFFKKFDSIRYNKEYDRFEFKFDYSNFTAKAEDTRT
QWTLCTYGTRIETFRNAEKNSNWDSREIDLTTEWKTLFTQHNIP
LNANLKEAILLQANKNFYTDILHLMKLTLQMRNSVTGTDIDYM
VSPVANECGEFFDSRKVKEGLPVNADANGAYNIARKGLWLAQQ
IKNANDLSDVKLAITNKEWLQFAQKKQYLKD (SEQ ID NO: 87)

[1480] >WP_036388671_（修飾的）假定蛋白[山羊莫拉氏菌（Moraxella caprae）]

[1481]

MLFQDFTHLYPLSKTMRFELKPIGKTLEHIHAKNFLSQDET
MADMYQKVKAILDDYHRDFIADMMGEVKLTKLAEFYDVYLKF
RKNPKDDGLQKQLKDLQAVLRKEIVKPIGNGGKYKAGYDRLFG
AKLFKDGKELGDLAKFVIAQEGESSPKLAHLAHFEKFSTYFTGF
HDNRKNMYSDEDKHTAITYRLIHENLPRFIDNLQILATIKQKHSA
LYDQIINELTASGLDVSLASHLDGYHKLLTQEGITAYNTLLGGISG
EAGSRKIQGINELINSHHNQHCHKSERIAKLRPLHKQILSDGMGV
SFLPSKFADDSEMCQAVNEFYRHYADVFAKVQSLFDGFDDHQK
DGIYVEHKNLNELSKQAFGDFALLGRVLDGYYVDVVNPEFNER
FAKAKTDNAKAKLTKEKDKFIKGVHSLASLEQAIEHYTARHDDE
SVQAGKLGQYFKHGLAGVDNPIQKIHNNHSTIKGFLERERPAGE
RALPKIKSGKNPEMTQLRQLKELLDNALNVAHFAKLLTTKTTLD
NQDGNFYGEFGALYDELAKIPTLYNKVRDYLSQKPFSTEKYKLN
FGNPTLLNGWDLNKEKDNFGIILQKDGCYYLALLDKAHKKVFD
NAPNTGKNVYQKMIYKLLPGPNKMLPKVFFAKSNLDYYNPSAE
LLDKYAQGTHKKGNNFNLKDCHALIDFFKAGINKHPEWQHFGF
KFSPTSSYQDLSDFYREVEPQGYQVKFVDINADYINELVEQGQL
YLFQIYNKDFSPKAHGKPNLHTLYFKALFSKDNLANPIYKLNGE
AQIFYRKASLDMNETTIHRAGEVLENKNPDNPKKRQFVYDIIKD
KRYTQDKFMLHVPITMNFGVQGMTIKEFNKKVNQSIQQYDEVN
VIGIDRGERHLLYLTVINSKGEILEQRSLNDITTASANGTQMTTPY

HKILDKREIERLNARVGWGEIETIKELKSGYLSHVVHQISQLMLK
YNAIVVLEDLNFGFKRGRFKVEKQIYQNFENALIKKLNHLVLKD
EADDEIGSYKNALQLTNNFTDLKSIGKQTGFLFYVPAWNTSKIDP
ETGFVDLLKPRYENIAQSQAFFGKFDKICYNADKDYFEFHIDYA
KFTDKAKNSRQIWKICSHGDKRYVYDKTANQNKGATKGINVND
ELKSLFARHHINDKQPNLVMDICQNNDKEFHKSLIYLLKTLLAL
RYSNASSDEDFILSPVANDEGMFFNSALADDTQPQNADANGAYH
IALKGLWVLEQIKNSDDLNKVKLAIDNQTWLNFAQNR（SEQ ID
NO: 88）

[1482] >WP_020988726_（修飾的）CRISPR相關蛋白Cpf1，PREFRAN亞型[稻田鉤端螺旋體]

[1483]

MEDYSGFVNIYSIQKTLRFELKPVGKTLEHIEKKGFLKKDKI
RAEDYKAVKKIIDKYHRAYIEEVFDSVLHQKKKKDKTRFSTQFI
KEIKEFSELYYKTEKNIPDKERLEALSEKLRKMLVGAFKGEFSEE
VAEKYKNLFSKELIRNEIEKFCETDEERKQVSNFKSFTTYFTGFH
SNRQNIYSDEKKSTAIGYRIIHQNLPKFLDNLKIIESIQRRFKDFP
WSDLKKNLKKIDKNIKLTEYFSIDGFVNVLNQKGIDAYNTILGG
KSEESGEKIQGLNEYINLYRQKNNIDRKNLPNVKILFKQILGDRE
TKSFIPEAFPDDQSVLNSITEFAKYLKLDKKKKSIIAELKKFLSSF
NRYELDGIYLANDNSLASISTFLFDDWSFIKKSVSFKYDESVGDP
KKKIKSPLKYEKEKEKWLKQKYYTISFLNDAIESYSKSQDEKRV
KIRLEAYFAEFKSKDDAKKQFDLLERIEEAYAIVEPLLGAEYPRD

RNLKADKKEVGKIKDFLDSIKSLQFFLKPLLSAEIFDEKDLGFYN
QLEGYYEEIDSIGHLYNKVRNYLTGKIYSKEKFKLNFENSTLLKG
WDENREVANLCVIFREDQKYYLGVMDKENNTILSDIPKVKPNEL
FYEKMVYKLIPTPHMQLPRIIFSSDNLSIYNPSKSILKIREAKSFKE
GKNFKLKDCHKFIDFYKESISKNEDWSRFDFKFSKTSSYENISEF
YREVERQGYNLDFKKVSKFYIDSLVEDGKLYLFQIYNKDFSIFSK
GKPNLHTIYFRSLFSKENLKDVCLKLNGEAEMFFRKKSINYDEK
KKREGHHPELFEKLKYPILKDKRYSEDKFQFHLPISLNFKSKERL
NFNLKVNEFLKRNKDINIIGIDRGERNLLYLVMINQKGEILKQTL
LDSMQSGKGRPEINYKEKLQEKEIERDKARKSWGTVENIKELKE
GYLSIVIHQISKLMVENNAIVVLEDLNIGFKRGRQKVERQVYQK
FEKMLIDKLNFLVFKENKPTEPGGVLKAYQLTDEFQSFEKLSKQT
GFLFYVPSWNTSKIDPRTGFIDFLHPAYENIEKAKQWINKFDSIRF
NSKMDWFEFTADTRKFSENLMLGKNRVWVICTTNVERYFTSKT
ANSSIQYNSIQITEKLKELFVDIPFSNGQDLKPEILRKNDAVFFKS
LLFYIKTTLSLRQNNGKKGEEEKDFILSPVVDSKGRFFNSLEASD
DEPKDADANGAYHIALKGLMNLLVLNETKEENLSRPKWKIKNK
DWLEFVWERNR（SEQ ID NO: 89）

[1484] >WP_023936172_（修飾的）外切核酸酶SbcC [狗口腔卟啉單胞菌]

[1485]

MPWIDLKDFTNLYPVSKTLRFELKPVGKTLENIEKAGILKED
EHRAESYRRVKKIIDTYHKVFIDSSLENMAKMGIENEIKAMLQSF

CELYKKDHRTEGEDKALDKIRAVLRGLIVGAFTGVCGRRENTVQ
NEKYESLFKEKLIKEILPDFVLSTEAESLPFSVEEATRSLKEFDSFT
SYFAGFYENRKNIYSTKPQSTAIAYRLIHENLPKFIDNILVFQKIKE
PIAKELEHIRADFSAGGYIKKDERLEDIFSLNYYIHVLSQAGIEKY
NALIGKIVTEGDGEMKGLNEHINLYNQQRGREDRLPLFRPLYKQI
LSDREQLSYLPESFEKDEELLRALKEFYDHIAEDILGRTQQLMTS
ISEYDLSRIYVRNDSQLTDISKKMLGDWNAIYMARERAYDHEQA
PKRITAKYERDRIKALKGEESISLANLNSCIAFLDNVRDCRVDTY
LSTLGQKEGPHGLSNLVENVFASYHEAEQLLSFPYPEENNLIQDK
DNVVLIKNLLDNISDLQRFLKPLWGMGDEPDKDERFYGEYNYIR
GALDQVIPLYNKVRNYLTRKPYSTRKVKLNFGNSQLLSGWDRN
KEKDNSCVILRKGQNFYLAIMNNRHKRSFENKVLPEYKEGEPYF
EKMDYKFLPDPNKMLPKVFLSKKGIEIYEPSPKLLEQYGHGTHK
KGDTFSMDDLHELIDFFKHSIEAHEDWKQFGFKFSDTATYENVS
SFYREVEDQGYKLSFRKVSESYVYSLIDQGKLYLFQIYNKDFSPC
SKGTPNLHTLYWRMLFDERNLADVIYKLDGKAEIFFREKSLKND
HPTHPAGKPIKKKSRQKKGEESLFEYDLVKDRRYTMDKFQFHVP
ITMNFKCSAGSKVNDMVNAHIREAKDMHVIGIDRGERNLLYICV
IDSRGTILDQISLNTINDIDYHDLLESRDKDRQQERRNWQTIEGIK
ELKQGYLSQAVHRIAELMVAYKAVVALEDLNMGFKRGRQKVES
SVYQQFEKQLIDKLNYLVDKKKRPEDIGGLLRAYQFTAPFKSFK
EMGKQNGFLFYIPAWNTSNIDPTTGFVNLFHAQYENVDKAKSFF
QKFDSISYNPKKDWFEFAFDYKNFTKKAEGSRSMWILCTHGSRI
KNFRNSQKNGQWDSEEFALTEAFKSLFVRYEIDYTADLKTAIVD

EKQKDFFVDLLKLFKLTVQMRNSWKEKDLDYLISPVAGADGRF
FDTREGNKSLPKDADANGAYNIALKGLWALRQIRQTSEGGKLKL
AISNKEWLQFVQERSYEKD（SEQ ID NO: 90）

[1486] >WP_009217842_（修飾的）假定蛋白[口腔類擬桿菌（Bacteroidetes oral taxon）274]

[1487]

MRKFNEFVGLYPISKTLRFELKPIGKTLEHIQRNKLLEHDAV
RADDYVKVKKIIDKYHKCLIDEALSGFTFDTEADGRSNNSLSEY
YLYYNLKKRNEQEQKTFKTIQNNLRKQIVNKLTQSEKYKRIDKK
ELITTDLPDFLTNESEKELVEKFKNFTTYFTEFHKNRKNMYSKEE
KSTAIAFRLINENLPKFVDNIAAFEKVVSSPLAEKINALYEDFKEY
LNVEEISRVFRLDYYDELLTQKQIDLYNAIVGGRTEEDNKIQIKG
LNQYINEYNQQQTDRSNRLPKLKPLYKQILSDRESVSWLPPKFD
SDKNLLIKIKECYDALSEKEKVFDKLESILKSLSTYDLSKIYISND
SQLSYISQKMFGRWDIISKAIREDCAKRNPQKSRESLEKFAERID
KKLKTIDSISIGDVDECLAQLGETYVKRVEDYFVAMGESEIDDEQ
TDTTSFKKNIEGAYESVKELLNNADNITDNNLMQDKGNVEKIKT
LLDAIKDLQRFIKPLLGKGDEADKDGVFYGEFTSLWTKLDQVTP
LYNMVRNYLTSKPYSTKKIKLNFENSTLMDGWDLNKEPDNTTVI
FCKDGLYYLGIMGKKYNRVFVDREDLPHDGECYDKMEYKLLPG
ANKMLPKVFFSETGIQRFLPSEELLGKYERGTHKKGAGFDLGDC
RALIDFFKKSIERHDDWKKFDFKFSDTSTYQDISEFYREVEQQGY
KMSFRKVSVDYIKSLVEEGKLYLFQIYNKDFSAHSKGTPNMHTL

YWKMLFDEENLKDVVYKLNGEAEVFFRKSSITVQSPTHPANSPI
KNKNKDNQKKESKFEYDLIKDRRYTVDKFLFHVPITMNFKSVG
GSNINQLVKRHIRSATDLHIIGIDRGERHLLYLTVIDSRGNIKEQFS
LNEIVNEYNGNTYRTDYHELLDTREGERTEARRNWQTIQNIREL
KEGYLSQVIHKISELAIKYNAVIVLEDLNFGFMRSRQKVEKQVY
QKFEKMLIDKLNYLVDKKKPVAETGGLLRAYQLTGEFESFKTLG
KQSGILFYVPAWNTSKIDPVTGFVNLFDTHYENIEKAKVFFDKFK
SIRYNSDKDWFEFVVDDYTRFSPKAEGTRRDWTICTQGKRIQIC
RNHQRNNEWEGQEIDLTKAFKEHFEAYGVDISKDLREQINTQNK
KEFFEELLRLLRLTLQMRNSMPSSDIDYLISPVANDTGCFFDSRK
QAELKENAVLPMNADANGAYNIARKGLLAIRKMKQEENDSAKI
SLAISNKEWLKFAQTKPYLED（SEQ ID NO: 91）

[1488] >WP_036890108_（修飾的）假定蛋白[狗口腔卟啉單胞菌]

[1489]

MDSLKDFTNLYPVSKTLRFELKPVGKTLENIEKAGILKEDEH
RAESYRRVKKIIDTYHKVFIDSSLENMAKMGIENEIKAMLQSFCE
LYKKDHRTEGEDKALDKIRAVLRGLIVGAFTGVCGRRENTVQNE
KYESLFKEKLIKEILPDFVLSTEAESLPFSVEEATRSLKEFDSFTSY
FAGFYENRKNIYSTKPQSTAIAYRLIHENLPKFIDNILVFQKIKEPI
AKELEHIRADFSAGGYIKKDERLEDIFSLNYYIHVLSQAGIEKYN
ALIGKIVTEGDGEMKGLNEHINLYNQQRGREDRLPLFRPLYKQIL
SDREQLSYLPESFEKDEELLRALKEFYDHIAEDILGRTQQLMTSIS

EYDLSRIYVRNDSQLTDISKKMLGDWNAIYMARERAYDHEQAP
KRITAKYERDRIKALKGEESISLANLNSCIAFLDNVRDCRVDTYL
STLGQKEGPHGLSNLVENVFASYHEAEQLLSFPYPEENNLIQDK
DNVVLIKNLLDNISDLQRFLKPLWGMGDEPDKDERFYGEYNYIR
GALDQVIPLYNKVRNYLTRKPYSTRKVKLNFGNSQLLSGWDRN
KEKDNSCVILRKGQNFYLAIMNNRHKRSFENKMLPEYKEGEPY
FEKMDYKFLPDPNKMLPKVFLSKKGIEIYKPSPKLLEQYGHGTH
KKGDTFSMDDLHELIDFFKHSIEAHEDWKQFGFKFSDTATYENV
SSFYREVEDQGYKLSFRKVSESYVYSLIDQGKLYLFQIYNKDFSP
CSKGTPNLHTLYWRMLFDERNLADVIYKLDGKAEIFFREKSLKN
DHPTHPAGKPIKKKSRQKKGEESLFEYDLVKDRRYTMDKFQFHV
PITMNFKCSAGSKVNDMVNAHIREAKDMHVIGIDRGERNLLYIC
VIDSRGTILDQISLNTINDIDYHDLLESRDKDRQQEHRNWQTIEGI
KELKQGYLSQAVHRIAELMVAYKAVVALEDLNMGFKRGRQKVE
SSVYQQFEKQLIDKLNYLVDKKKRPEDIGGLLRAYQFTAPFKSFK
EMGKQNGFLFYIPAWNTSNIDPTTGFVNLFHVQYENVDKAKSFF
QKFDSISYNPKKDWFEFAFDYKNFTKKAEGSRSMWILCTHGSRI
KNFRNSQKNGQWDSEEFALTEAFKSLFVRYEIDYTADLKTAIVD
EKQKDFFVDLLKLFKLTVQMRNSWKEKDLDYLISPVAGADGRF
FDTREGNKSLPKDADANGAYNIALKGLWALRQIRQTSEGGKLKL
AISNKEWLQFVQERSYEKD（SEQ ID NO: 92）

[1490] >WP_036887416_（修飾的）假定蛋白[狗口腔卟啉單胞菌]

[1491]

MDSLKDFTNLYPVSKTLRFELKPVGKTLENIEKAGILKEDEH
RAESYRRVKKIIDTYHKVFIDSSLENMAKMGIENEIKAMLQSFCE
LYKKDHRTEGEDKALDKIRAVLRGLIVGAFTGVCGRRENTVQNE
KYESLFKEKLIKEILPDFVLSTEAESLPFSVEEATRSLKEFDSFTSY
FAGFYENRKNIYSTKPQSTAIAYRLIHENLPKFIDNILVFQKIKEPI
AKELEHIRADFSAGGYIKKDERLEDIFSLNYYIHVLSQAGIEKYN
ALIGKIVTEGDGEMKGLNEHINLYNQQRGREDRLPLFRPLYKQIL
SDREQLSYLPESFEKDEELLRALKEFYDHIAEDILGRTQQLMTSIS
EYDLSRIYVRNDSQLTDISKKMLGDWNAIYMARERAYDHEQAP
KRITAKYERDRIKALKGEESISLANLNSCIAFLDNVRDCRVDTYL
STLGQKEGPHGLSNLVENVFASYHEAEQLLSFPYPEENNLIQDK
DNVVLIKNLLDNISDLQRFLKPLWGMGDEPDKDERFYGEYNYIR
GALDQVIPLYNKVRNYLTRKPYSTRKVKLNFGNSQLLSGWDRN
KEKDNSCVILRKGQNFYLAIMNNRHKRSFENKVLPEYKEGEPYF
EKMDYKFLPDPNKMLPKVFLSKKGIEIYKPSPKLLEQYGHGTHK
KGDTFSMDDLHELIDFFKHSIEAHEDWKQFGFKFSDTATYENVS
SFYREVEDQGYKLSFRKVSESYVYSLIDQGKLYLFQIYNKDFSPC
SKGTPNLHTLYWRMLFDERNLADVIYKLDGKAEIFFREKSLKND
HPTHPAGKPIKKKSRQKKGEESLFEYDLVKDRHYTMDKFQFHVP
ITMNFKCSAGSKVNDMVNAHIREAKDMHVIGIDRGERNLLYICV
IDSRGTILDQISLNTINDIDYHDLLESRDKDRQQERRNWQTIEGIK
ELKQGYLSQAVHRIAELMVAYKAVVALEDLNMGFKRGRQKVES
SVYQQFEKQLIDKLNYLVDKKKRPEDIGGLLRAYQFTAPFKSFK

EMGKQNGFLFYIPAWNTSNIDPTTGFVNLFHAQYENVDKAKSFF
QKFDSISYNPKKDWFEFAFDYKNFTKKAEGSRSMWILCTHGSRI
KNFRNSQKNGQWDSEEFALTEAFKSLFVRYEIDYTADLKTAIVD
EKQKDFFVDLLKLFKLTVQMRNSWKEKDLDYLISPVAGADGRF
FDTREGNKSLPKDADANGAYNIALKGLWALRQIRQTSEGGKLKL
AISNKEWLQFVQERSYEKD （SEQ ID NO: 93）

[1492] >WP_023941260_（修飾的）外切核酸酶SbcC [犬口腔卟啉單胞菌]

[1493]

MDSLKDFTNLYPVSKTLRFELKPVGKTLENIEKAGILKEDEH
RAESYRRVKKIIDTYHKVFIDSSLENMAKMGIENEIKAMLQSFCE
LYKKDHRTEGEDKALDKIRAVLRGLIVGAFTGVCGRRENTVQNE
KYESLFKEKLIKEILPDFVLSTEAESLPFSVEEATRSLKEFDSFTSY
FAGFYENRKNIYSTKPQSTAIAYRLIHENLPKFIDNILVFQKIKEPI
AKELEHIRADFSAGGYIKKDERLEDIFSLNYYIHVLSQAGIEKYN
ALIGKIVTEGDGEMKGLNEHINLYNQQRGREDRLPLFRPLYKQIL
SDREQLSYLPESFEKDEELLRALKEFYDHIAEDILGRTQQLMTSIS
EYDLSRIYVRNDSQLTDISKKMLGDWNAIYMARERAYDHEQAP
KRITAKYERDRIKALKGEESISLANLNSCIAFLDNVRDCRVDTYL
STLGQKEGPHGLSNLVENVFASYHEAEQLLSFPYPEENNLIQDK
DNVVLIKNLLDNISDLQRFLKPLWGMGDEPDKDERFYGEYNYIR
GALDQVIPLYNKVRNYLTRKPYSTRKVKLNFGNSQLLSGWDRN
KEKDNSCVILRKGQNFYLAIMNNRHKRSFENKVLPEYKEGEPYF

EKMDYKFLPDPNKMLPKVFLSKKGIEIYKPSPKLLEQYGHGTHK
KGDTFSMDDLHELIDFFKHSIEAHEDWKQFGFKFSDTATYENVS
SFYREVEDQGYKLSFRKVSESYVYSLIDQGKLYLFQIYNKDFSPC
SKGTPNLHTLYWRMLFDERNLADVIYKLDGKAEIFFREKSLKND
HPTHPAGKPIKKKSRQKKGEESLFEYDLVKDRRYTMDKFQFHVP
ITMNFKCSAGSKVNDMVNAHIREAKDMHVIGIDRGERNLLYICV
IDSRGTILDQISLNTINDIDYHDLLESRDKDRQQERRNWQTIEGIK
ELKQGYLSQAVHRIAELMVAYKAVVALEDLNMGFKRGRQKVES
SVYQQFEKQLIDKLNYLVDKKKRPEDIGGLLRAYQFTAPFKSFK
EMGKQNGFLFYIPAWNTSNIDPTTGFVNLFHAQYENVDKAKSFF
QKFDSISYNPKKDWFEFAFDYKNFTKKAEGSRSMWILCTHGSRI
KNFRNSQKNGQWDSEEFALTEAFKSLFVRYEIDYTADLKTAIVD
EKQKDFFVDLLKLFKLTVQMRNSWKEKDLDYLISPVAGADGRF
FDTREGNKSLPKDADANGAYNIALKGLWALRQIRQTSEGGKLKL
AISNKEWLQFVQERSYEKD（SEQ ID NO: 94）

[1494] >WP_037975888_（修飾的）假定蛋白[窮氏互養菌（Synergistes jonesii）]

[1495]

MANSLKDFTNIYQLSKTLRFELKPIGKTEEHINRKLIIMHDE
KRGEDYKSVTKLIDDYHRKFIHETLDPAHFDWNPLAEALIQSGS
KNNKALPAEQKEMREKIISMFTSQAVYKKLFKKELFSELLPEMIK
SELVSDLEKQAQLDAVKSFDKFSTYFTGFHENRKNIYSKKDTSTS
IAFRIVHQNFPKFLANVRAYTLIKERAPEVIDKAQKELSGILGGK

TLDDIFSIESFNNVLTQDKIDYYNQIIGGVSGKAGDKKLRGVNEF
SNLYRQQHPEVASLRIKMVPLYKQILSDRTTLSFVPEALKDDEQA
INAVDGLRSELERNDIFNRIKRLFGKNNLYSLDKIWIKNSSISAFS
NELFKNWSFIEDALKEFKENEFNGARSAGKKAEKWLKSKYFSFA
DIDAAVKSYSEQVSADISSAPSASYFAKFTNLIETAAENGRKFSYF
AAESKAFRGDDGKTEIIKAYLDSLNDILHCLKPFETEDISDIDTEF
YSAFAEIYDSVKDVIPVYNAVRNYTTQKPFSTEKFKLNFENPALA
KGWDKNKEQNNTAIILMKDGKYYLGVIDKNNKLRADDLADDG
SAYGYMKMNYKFIPTPHMELPKVFLPKRAPKRYNPSREILLIKEN
KTFIKDKNFNRTDCHKLIDFFKDSINKHKDWRTFGFDFSDTDSYE
DISDFYMEVQDQGYKLTFTRLSAEKIDKWVEEGRLFLFQIYNKD
FADGAQGSPNLHTLYWKAIFSEENLKDVVLKLNGEAELFFRRKS
IDKPAVHAKGSMKVNRRDIDGNPIDEGTYVEICGYANGKRDMAS
LNAGARGLIESGLVRITEVKHELVKDKRYTIDKYFFHVPFTINFK
AQGQGNINSDVNLFLRNNKDVNIIGIDRGERNLVYVSLIDRDGHI
KLQKDFNIIGGMDYHAKLNQKEKERDTARKSWKTIGTIKELKEG
YLSQVVHEIVRLAVDNNAVIVMEDLNIGFKRGRFKVEKQVYQKF
EKMLIDKLNYLVFKDAGYDAPCGILKGLQLTEKFESFTKLGKQC
GIIFYIPAGYTSKIDPTTGFVNLFNINDVSSKEKQKDFIGKLDSIRF
DAKRDMFTFEFDYDKFRTYQTSYRKKWAVWTNGKRIVREKDK
DGKFRMNDRLLTEDMKNILNKYALAYKAGEDILPDVISRDKSLA
SEIFYVFKNTLQMRNSKRDTGEDFIISPVLNAKGRFFDSRKTDAA
LPIDADANGAYHIALKGSLVLDAIDEKLKEDGRIDYKDMAVSNP
KWFEFMQTRKFDF (SEQ ID NO: 95)

[1496] >EFI70750_（修飾的）保守性假定蛋白[布氏普雷沃菌（Prevotella bryantii）B14]

[1497]

MQINNLKIIYMKFTDFTGLYSLSKTLRFELKPIGKTLENIKKA
GLLEQDQHRADSYKKVKKIIDEYHKAFIEKSLSNFELKYQSEDK
LDSLEEYLMYYSMKRIEKTEKDKFAKIQDNLRKQIADHLKGDES
YKTIFSKDLIRKNLPDFVKSDEERTLIKEFKDFTTYFKGFYENRE
NMYSAEDKSTAISHRIIHENLPKFVDNINAFSKIILIPELREKLNQI
YQDFEEYLNVESIDEIFHLDYFSMVMTQKQIEVYNAIIGGKSTND
KKIQGLNEYINLYNQKHKDCKLPKLKLLFKQILSDRIAISWLPDN
FKDDQEALDSIDTCYKNLLNDGNVLGEGNLKLLLENIDTYNLK
GIFIRNDLQLTDISQKMYASWNVIQDAVILDLKKQVSRKKKESAE
DYNDRLKKLYTSQESFSIQYLNDCLRAYGKTENIQDYFAKLGAV
NNEHEQTINLFAQVRNAYTSVQAILTTPYPENANLAQDKETVALI
KNLLDSLKRLQRFIKPLLGKGDESDKDERFYGDFTPLWETLNQI
TPLYNMVRNYMTRKPYSQEKIKLNFENSTLLGGWDLNKEHDNT
AIILRKNGLYYLAIMKKSANKIFDKDKLDNSGDCYEKMVYKLLP
GANKMLPKVFFSKSRIDEFKPSENIIENYKKGTHKKGANFNLAD
CHNLIDFFKSSISKHEDWSKFNFHFSDTSSYEDLSDFYREVEQQG
YSISFCDVSVEYINKMVEKGDLYLFQIYNKDFSEFSKGTPNMHTL
YWNSLFSKENLNNIIYKLNGQAEIFFRKKSLNYKRPTHPAHQAIK
NKNKCNEKKESIFDYDLVKDKRYTVDKFQFHVPITMNFKSTGNT
NINQQVIDYLRTEDDTHIIGIDRGERHLLYLVVIDSHGKIVEQFTL
NEIVNEYGGNIYRTNYHDLLDTREQNREKARESWQTIENIKELK

EGYISQVIHKITDLMQKYHAVVVLEDLNMGFMRGRQKVEKQVY
QKFEEMLINKLNYLVNKKADQNSAGGLLHAYQLTSKFESFQKL
GKQSGFLFYIPAWNTSKIDPVTGFVNLFDTRYESIDKAKAFFGKF
DSIRYNADKDWFEFAFDYNNFTTKAEGTRTNWTICTYGSRIRTF
RNQAKNSQWDNEEIDLTKAYKAFFAKHGINIYDNIKEAIAMETE
KSFFEDLLHLLKLTLQMRNSITGTTTDYLISPVHDSKGNFYDSRI
CDNSLPANADANGAYNIARKGLMLIQQIKDSTSSNRFKFSPITNK
DWLIFAQEKPYLND （SEQ ID NO: 96）

[1498] >WP_024988992_（修飾的）假定蛋白[易北普雷沃菌]

[1499]

MNIKNFTGLYPLSKTLRFELKPIGKTKENIEKNGILTKDEQR
AKDYLIVKGFIDEYHKQFIKDRLWDFKLPLESEGEKNSLEEYQE
LYELTKRNDAQEADFTEIKDNLRSSITEQLTKSGSAYDRIFKKEFI
REDLVNFLEDEKDKNIVKQFEDFTTYFTGFYENRKNMYSSEEKS
TAIAYRLIHQNLPKFMDNMRSFAKIANSSVSEHFSDIYESWKEYL
NVNSIEEIFQLDYFSETLTQPHIEVYNYIIGKKVLEDGTEIKGINE
YVNLYNQQQKDKSKRLPFLVPLYKQILSDREKLSWIAEEFDSDK
KMLSAITESYNHLHNVLMGNENESLRNLLLNIKDYNLEKINITN
DLSLTEISQNLFGRYDVFTNGIKNKLRVLTPRKKKETDENFEDRI
NKIFKTQKSFSIAFLNKLPQPEMEDGKPRNIEDYFITQGAINTKSI
QKEDIFAQIENAYEDAQVFLQIKDTDNKLSQNKTAVEKIKTLLDA
LKELQHFIKPLLGSGEENEKDELFYGSFLAIWDELDTITPLYNKV
RNWLTRKPYSTEKIKLNFDNAQLLGGWDVNKEHDCAGILLRKN

DSYYLGIINKKTNHIFDTDITPSDGECYDKIDYKLLPGANKMLPK
VFFSKSRIKEFEPSEAIINCYKKGTHKKGKNFNLTDCHRLINFFKT
SIEKHEDWSKFGFKFSDTETYEDISGFYREVEQQGYRLTSHPVSA
SYIHSLVKEGKLYLFQIWNKDFSQFSKGTPNLHTLYWKMLFDKR
NLSDVVYKLNGQAEVFYRKSSIEHQNRIIHPAQHPITNKNELNKK
HTSTFKYDIIKDRRYTVDKFQFHVPITINFKATGQNNINPIVQEVI
RQNGITHIIGIDRGERHLLYLSLIDLKGNIIKQMTLNEIINEYKGV
TYKTNYHNLLEKREKERTEARHSWSSIESIKELKDGYMSQVIHKI
TDMMVKYNAIVVLEDLNGGFMRGRQKVEKQVYQKFEKKLIDK
LNYLVDKKLDANEVGGVLNAYQLTNKFESFKKIGKQSGFLFYIP
AWNTSKIDPITGFVNLFNTRYESIKETKVFWSKFDIIRYNKEKNW
FEFVFDYNTFTTKAEGTRTKWTLCTHGTRIQTFRNPEKNAQWD
NKEINLTESFKALFEKYKIDITSNLKESIMQETEKKFFQELHNLLH
LTLQMRNSVTGTDIDYLISPVADEDGNFYDSRINGKNFPENADA
NGAYNIARKGLMLIRQIKQADPQKKFKFETITNKDWLKFAQDKP
YLKD（SEQ ID NO: 97）

[1500] >WP_039658684_（修飾的）假定蛋白[密斯氏菌屬某種SC_K08D17]

[1501]

MQTLFENFTNQYPVSKTLRFELIPQGKTKDFIEQKGLLKKD
EDRAEKYKKVKNIIDEYHKDFIEKSLNGLKLDGLEKYKTLYLKQ
EKDDKDKKAFDKEKENLRKQIANAFRNNEKFKTLFAKELIKNDL
MSFACEEDKKNVKEFEAFTTYFTGFHQNRANMYVADEKRTAIAS

RLIHENLPKFIDNIKIFEKMKKEAPELLSPFNQTLKDMKDVIKGT
TLEEIFSLDYFNKTLTQSGIDIYNSVIGGRTPEEGKTKIKGLNEYI
NTDFNQKQTDKKKRQPKFKQLYKQILSDRQSLSFIAEAFKNDTEI
LEAIEKFYVNELLHFSNEGKSTNVLDAIKNAVSNLESFNLTKMYF
RSGASLTDVSRKVFGEWSIINRALDNYYATTYPIKPREKSEKYEE
RKEKWLKQDFNVSLIQTAIDEYDNETVKGKNSGKVIADYFAKFC
DDKETDLIQKVNEGYIAVKDLLNTPCPENEKLGSNKDQVKQIKA
FMDSIMDIMHFVRPLSLKDTDKEKDETFYSLFTPLYDHLTQTIAL
YNKVRNYLTQKPYSTEKIKLNFENSTLLGGWDLNKETDNTAIIL
RKDNLYYLGIMDKRHNRIFRNVPKADKKDFCYEKMVYKLLPGA
NKMLPKVFFSQSRIQEFTPSAKLLENYANETHKKGDNFNLNHCH
KLIDFFKDSINKHEDWKNFDFRFSATSTYADLSGFYHEVEHQGY
KISFQSVADSFIDDLVNEGKLYLFQIYNKDFSPFSKGKPNLHTLY
WKMLFDENNLKDVVYKLNGEAEVFYRKKSIAEKNTTIHKANES
IINKNPDNPKATSTFNYDIVKDKRYTIDKFQFHIPITMNFKAEGIF
NMNQRVNQFLKANPDINIIGIDRGERHLLYYALINQKGKILKQDT
LNVIANEKQKVDYHNLLDKKEGDRATARQEWGVIETIKELKEG
YLSQVIHKLTDLMIENNAIIVMEDLNFGFKRGRQKVEKQVYQKF
EKMLIDKLNYLVDKNKKANELGGLLNAFQLANKFESFQKMGK
QNGFIFYVPAWNTSKTDPATGFIDFLKPRYENLNQAKDFFEKFDS
IRLNSKADYFEFAFDFKNFTEKADGGRTKWTVCTTNEDRYAWN
RALNNNRGSQEKYDITAELKSLFDGKVDYKSGKDLKQQIASQES
ADFFKALMKNLSITLSLRHNNGEKGDNEQDYILSPVADSKGRFF
DSRKADDDMPKNADANGAYHIALKGLWCLEQISKTDDLKKVKL

AISNKEWLEFVQTLKG （SEQ ID NO: 98）

[1502] >WP_037385181_（修飾的）假定蛋白[密斯氏菌屬某種SCADC]

[1503]

MQTLFENFTNQYPVSKTLRFELIPQGKTKDFIEQKGLLKKD
EDRAEKYKKVKNIIDEYHKDFIEKSLNGLKLDGLEEYKTLYLKQ
EKDDKDKKAFDKEKENLRKQIANAFRNNEKFKTLFAKELIKNDL
MSFACEEDKKNVKEFEAFTTYFTGFHQNRANMYVADEKRTAIAS
RLIHENLPKFIDNIKIFEKMKKEAPELLSPFNQTLKDMKDVIKGT
TLEEIFSLDYFNKTLTQSGIDIYNSVIGGRTPEEGKTKIKGLNEYI
NTDFNQKQTDKKKRQPKFKQLYKQILSDRQSLSFIAEAFKNDTEI
LEAIEKFYVNELLHFSNEGKSTNVLDAIKNAVSNLESFNLTKIYF
RSGTSLTDVSRKVFGEWSIINRALDNYYATTYPIKPREKSEKYEE
RKEKWLKQDFNVSLIQTAIDEYDNETVKGKNSGKVIVDYFAKFC
DDKETDLIQKVNEGYIAVKDLLNTPYPENEKLGSNKDQVKQIKA
FMDSIMDIMHFVRPLSLKDTDKEKDETFYSLFTPLYDHLTQTIAL
YNKVRNYLTQKPYSTEKIKLNFENSTLLGGWDLNKETDNTAIIL
RKENLYYLGIMDKRHNRIFRNVPKADKKDSCYEKMVYKLLPGA
NKMLPKVFFSQSRIQEFTPSAKLLENYENETHKKGDNFNLNHCH
QLIDFFKDSINKHEDWKNFDFRFSATSTYADLSGFYHEVEHQGY
KISFQSIADSFIDDLVNEGKLYLFQIYNKDFSPFSKGKPNLHTLYW
KMLFDENNLKDVVYKLNGEAEVFYRKKSIAEKNTTIHKANESII
NKNPDNPKATSTFNYDIVKDKRYTIDKFQFHVPITMNFKAEGIFN

MNQRVNQFLKANPDINIIGIDRGERHLLYYTLINQKGKILKQDTL
NVIANEKQKVDYHNLLDKKEGDRATARQEWGVIETIKELKEGY
LSQVIHKLTDLMIENNAIIVMEDLNFGFKRGRQKVEKQVYQKFE
KMLIDKLNYLVDKNKKANELGGLLNAFQLANKFESFQKMGKQ
NGFIFYVPAWNTSKTDPATGFIDFLKPRYENLKQAKDFFEKFDSI
RLNSKADYFEFAFDFKNFTGKADGGRTKWTVCTTNEDRYAWNR
ALNNNRGSQEKYDITAELKSLFDGKVDYKSGKDLKQQIASQELA
DFFRTLMKYLSVTLSLRHNNGEKGETEQDYILSPVADSMGKFFD
SRKAGDDMPKNADANGAYHIALKGLWCLEQISKTDDLKKVKLA
ISNKEWLEFMQTLKG （SEQ ID NO: 99）

[1504] >WP_039871282_(修飾的)假定蛋白[布氏普雷沃菌]

[1505]

MKFTDFTGLYSLSKTLRFELKPIGKTLENIKKAGLLEQDQHR
ADSYKKVKKIIDEYHKAFIEKSLSNFELKYQSEDKLDSLEEYLM
YYSMKRIEKTEKDKFAKIQDNLRKQIADHLKGDESYKTIFSKDLI
RKNLPDFVKSDEERTLIKEFKDFTTYFKGFYENRENMYSAEDKS
TAISHRIIHENLPKFVDNINAFSKIILIPELREKLNQIYQDFEEYLN
VESIDEIFHLDYFSMVMTQKQIEVYNAIIGGKSTNDKKIQGLNEY
INLYNQKHKDCKLPKLKLLFKQILSDRIAISWLPDNFKDDQEALD
SIDTCYKNLLNDGNVLGEGNLKLLLENIDTYNLKGIFIRNDLQLT
DISQKMYASWNVIQDAVILDLKKQVSRKKKESAEDYNDRLKKL
YTSQESFSIQYLNDCLRAYGKTENIQDYFAKLGAVNNEHEQTINL
FAQVRNAYTSVQAILTTPYPENANLAQDKETVALIKNLLDSLKRL

QRFIKPLLGKGDESDKDERFYGDFTPLWETLNQITPLYNMVRNY
MTRKPYSQEKIKLNFENSTLLGGWDLNKEHDNTAIILRKNGLYY
LAIMKKSANKIFDKDKLDNSGDCYEKMVYKLLPGANKMLPKVF
FSKSRIDEFKPSENIIENYKKGTHKKGANFNLADCHNLIDFFKSSI
SKHEDWSKFNFHFSDTSSYEDLSDFYREVEQQGYSISFCDVSVE
YINKMVEKGDLYLFQIYNKDFSEFSKGTPNMHTLYWNSLFSKEN
LNNIIYKLNGQAEIFFRKKSLNYKRPTHPAHQAIKNKNKCNEKK
ESIFDYDLVKDKRYTVDKFQFHVPITMNFKSTGNTNINQQVIDYL
RTEDDTHIIGIDRGERHLLYLVVIDSHGKIVEQFTLNEIVNEYGGN
IYRTNYHDLLDTREQNREKARESWQTIENIKELKEGYISQVIHKI
TDLMQKYHAVVVLEDLNMGFMRGRQKVEKQVYQKFEEMLINK
LNYLVNKKADQNSAGGLLHAYQLTSKFESFQKLGKQSGFLFYIP
AWNTSKIDPVTGFVNLFDTRYESIDKAKAFFGKFDSIRYNADKD
WFEFAFDYNNFTTKAEGTRTNWTICTYGSRIRTFRNQAKNSQWD
NEEIDLTKAYKAFFAKHGINIYDNIKEAIAMETEKSFFEDLLHLLK
LTLQMRNSITGTTTDYLISPVHDSKGNFYDSRICDNSLPANADAN
GAYNIARKGLMLIQQIKDSTSSNRFKFSPITNKDWLIFAQEKPYL
ND （SEQ ID NO: 100）

[1506] >EKE28449_（修飾的）假定蛋白ACD_3C00058G0015 [未培養細菌（gcode 4）]

[1507]

MFKGDAFTGLYEVQKTLRFELVPIGLTQSYLENDWVIQKDK
EVEENYGKIKAYFDLIHKEFVRQSLENAWLCQLDDFYEKYIELH

NSLETRKDKNLAKQFEKVMKSLKKEFVSFFDAKWNEWKQKFSF
LKKWWIDVLNEKEVLDLMAEFYPDEKELFDKFDKFFTYFSNFK
ESRKNFYADDGRAWAIATRAIDENLITFIKNIEDFKKLNSSFREFV
NDNFSEEDKQIFEIDFYNNCLLQPWIDKYNKIVWWYSLENWEK
VQWLNEKINNFKQNQNKSNSKDLKFPRMKLLYKQILGDKEKKV
YIDEIRDDKNLIDLIDNSKRRNQIKIDNANDIINDFINNNAKFELD
KIYLTRQSINTISSKYFSSWDYIRWYFWTGELQEFVSFYDLKETF
WKIEYETLENIFKDCYVKGINTESQNNIVFETQGIYENFLNIFKFE
FNQNISQISLLEWELDKIQNEDIKKNEKQVEVIKNYFDSVMSVY
KMTKYFSLEKWKKRVELDTDNNFYNDFNEYLEGFEIWKDYNLV
RNYITKKQVNTDKIKLNFDNSQFLTWWDKDKENERLGIILRREW
KYYLWILKKWNTLNFGDYLQKEWEIFYEKMNYKQLNNVYRQL
PRLLFPLTKKLNELKWDELKKYLSKYIQNFWYNEEIAQIKIEFDI
FQESKEKWEKFDIDKLRKLIEYYKKWVLALYSDLYDLEFIKYKN
YDDLSIFYSDVEKKMYNLNFTKIDKSLIDGKVKSWELYLFQIYN
KDFSESKKEWSTENIHTKYFKLLFNEKNLQNLVVKLSWWADIFF
RDKTENLKFKKDKNGQEILDHRRFSQDKIMFHISITLNANCWDK
YWFNQYVNEYMNKERDIKIIWIDRWEKHLAYYCVIDKSWKIFN
NEIWTLNELNWVNYLEKLEKIESSRKDSRISWWEIENIKELKNG
YISQVINKLTELIVKYNAIIVFEDLNIWFKRWRQKIEKQIYQKLEL
ALAKKLNYLTQKDKKDDEILWNLKALQLVPKVNDYQDIWNYK
QSWIMFYVRANYTSVTCPNCWLRKNLYISNSATKENQKKSLNSI
AIKYNDWKFSFSYEIDDKSWKQKQSLNKKKFIVYSDIERFVYSP
LEKLTKVIDVNKKLLELFRDFNLSLDINKQIQEKDLDSVFFKSLT

HLFNLILQLRNSDSKDNKDYISCPSCYYHSNNWLQWFEFNWDA
NWAYNIARKGIILLDRIRKNQEKPDLYVSDIDWDNFVQSNQFPNT
IIPIQNIEKQVPLNIKI（SEQ ID NO: 101）

[1508] >WP_018359861_（修飾的）假定蛋白[獼猴卟啉單胞菌]

[1509]

MKTQHFFEDFTSLYSLSKTIRFELKPIGKTLENIKKNGLIRRD
EQRLDDYEKLKKVIDEYHEDFIANILSSFSFSEEILQSYIQNLSESE
ARAKIEKTMRDTLAKAFSEDERYKSIFKKELVKKDIPVWCPAYK
SLCKKFDNFTTSLVPFHENRKNLYTSNEITASIPYRIVHVNLPKFI
QNIEALCELQKKMGADLYLEMMENLRNVWPSFVKTPDDLCNLK
TYNHLMVQSSISEYNRFVGGYSTEDGTKHQGINEWINIYRQRNK
EMRLPGLVFLHKQILAKVDSSSFISDTLENDDQVFCVLRQFRKLF
WNTVSSKEDDAASLKDLFCGLSGYDPEAIYVSDAHLATISKNIF
DRWNYISDAIRRKTEVLMPRKKESVERYAEKISKQIKKRQSYSL
AELDDLLAHYSEESLPAGFSLLSYFTSLGGQKYLVSDGEVILYEE
GSNIWDEVLIAFRDLQVILDKDFTEKKLGKDEEAVSVIKKALDS
ALRLRKFFDLLSGTGAEIRRDSSFYALYTDRMDKLKGLLKMYDK
VRNYLTKKPYSIEKFKLHFDNPSLLSGWDKNKELNNLSVIFRQN
GYYYLGIMTPKGKNLFKTLPKLGAEEMFYEKMEYKQIAEPMLM
LPKVFFPKKTKPAFAPDQSVVDIYNKKTFKTGQKGFNKKDLYRL
IDFYKEALTVHEWKLFNFSFSPTEQYRNIGEFFDEVREQAYKVS
MVNVPASYIDEAVENGKLYLFQIYNKDFSPYSKGIPNLHTLYWK

ALFSEQNQSRVYKLCGGGELFYRKASLHMQDTTVHPKGISIHKK
NLNKKGETSLFNYDLVKDKRFTEDKFFFHVPISINYKNKKITNVN
QMVRDYIAQNDDLQIIGIDRGERNLLYISRIDTRGNLLEQFSLNVI
ESDKGDLRTDYQKILGDREQERLRRRQEWKSIESIKDLKDGYMS
QVVHKICNMVVEHKAIVVLENLNLSFMKGRKKVEKSVYEKFER
MLVDKLNYLVVDKKNLSNEPGGLYAAYQLTNPLFSFEELHRYPQ
SGILFFVDPWNTSLTDPSTGFVNLLGRINYTNVGDARKFFDRFNA
IRYDGKGNILFDLDLSRFDVRVETQRKLWTLTTFGSRIAKSKKSG
KWMVERIENLSLCFLELFEQFNIGYRVEKDLKKAILSQDRKEFY
VRLIYLFNLMMQIRNSDGEEDYILSPALNEKNLQFDSRLIEAKDL
PVDADANGAYNVARKGLMVVQRIKRGDHESIHRIGRAQWLRYV
QEGIVE（SEQ ID NO: 102）

[1510] >WP_013282991_(修飾的)假定蛋白[解朊丁酸弧菌]

[1511]

MLLYENYTKRNQITKSLRLELRPQGKTLRNIKELNLLEQDK
AIYALLERLKPVIDEGIKDIARDTLKNCELSFEKLYEHFLSGDKK
AYAKESERLKKEIVKTLIKNLPEGIGKISEINSAKYLNGVLYDFID
KTHKDSEEKQNILSDILETKGYLALFSKFLTSRITTLEQSMPKRVI
ENFEIYAANIPKMQDALERGAVSFAIEYESICSVDYYNQILSQEDI
DSYNRLISGIMDEDGAKEKGINQTISEKNIKIKSEHLEEKPFRILK
QLHKQILEEREKAFTIDHIDSDEEVVQVTKEAFEQTKEQWENIK
KINGFYAKDPGDITLFIVVGPNQTHVLSQLIYGEHDRIRLLLEEYE
KNTLEVLPRRTKSEKARYDKFVNAVPKKVAKESHTFDGLQKMT

GDDRLFILYRDELARNYMRIKEAYGTFERDILKSRRGIKGNRDV
QESLVSFYDELTKFRSALRIINSGNDEKADPIFYNTFDGIFEKANR
TYKAENLCRNYVTKSPADDARIMASCLGTPARLRTHWWNGEEN
FAINDVAMIRRGDEYYYFVLTPDVKPVDLKTKDETDAQIFVQRK
GAKSFLGLPKALFKCILEPYFESPEHKNDKNCVIEEYVSKPLTID
RRAYDIFKNGTFKKTNIGIDGLTEEKFKDDCRYLIDVYKEFIAVY
TRYSCFNMSGLKRADEYNDIGEFFSDVDTRLCTMEWIPVSFERI
NDMVDKKEGLLFLVRSMFLYNRPRKPYERTFIQLFSDSNMEHTS
MLLNSRAMIQYRAASLPRRVTHKKGSILVALRDSNGEHIPMHIRE
AIYKMKNNFDISSEDFIMAKAYLAEHDVAIKKANEDIIRNRRYTE
DKFFLSLSYTKNADISARTLDYINDKVEEDTQDSRMAVIVTRNL
KDLTYVAVVDEKNNVLEEKSLNEIDGVNYRELLKERTKIKYHDK
TRLWQYDVSSKGLKEAYVELAVTQISKLATKYNAVVVVESMSST
FKDKFSFLDEQIFKAFEARLCARMSDLSFNTIKEGEAGSISNPIQV
SNNNGNSYQDGVIYFLNNAYTRTLCPDTGFVDVFDKTRLITMQS
KRQFFAKMKDIRIDDGEMLFTFNLEEYPTKRLLDRKEWTVKIAG
DGSYFDKDKGEYVYVNDIVREQIIPALLEDKAVFDGNMAEKFLD
KTAISGKSVELIYKWFANALYGIITKKDGEKIYRSPITGTEIDVSK
NTTYNFGKKFMFKQEYRGDGDFLDAFLNYMQAQDIAV（SEQ ID
NO: 103）

[1512] >AIZ56868_（修飾的）假定蛋白Mpt1_c09950 [候選白蟻甲烷枝原體]

[1513]

MNNYDEFTKLYPIQKTIRFELKPQGRTMEHLETFNFFEEDRD
RAEKYKILKEAIDEYHKKFIDEHLTNMSLDWNSLKQISEKYYKS
REEKDKKVFLSEQKRMRQEIVSEFKKDDRFKDLFSKKLFSELLK
EEIYKKGNHQEIDALKSFDKFSGYFIGLHENRKNMYSDGDEITAI
SNRIVNENFPKFLDNLQKYQEARKKYPEWIIKAESALVAHNIKM
DEVFSLEYFNKVLNQEGIQRYNLALGGYVTKSGEKMMGLNDAL
NLAHQSEKSSKGRIHMTPLFKQILSEKESFSYIPDVFTEDSQLLPS
IGGFFAQIENDKDGNIFDRALELISSYAEYDTERIYIRQADINRVS
NVIFGEWGTLGGLMREYKADSINDINLERTCKKVDKWLDSKEF
ALSDVLEAIKRTGNNDAFNEYISKMRTAREKIDAARKEMKFISE
KISGDEESIHIIKTLLDSVQQFLHFFNLFKARQDIPLDGAFYAEFD
EVHSKLFAIVPLYNKVRNYLTKNNLNTKKIKLNFKNPTLANGWD
QNKVYDYASLIFLRDGNYYLGIINPKRKKNIKFEQGSGNGPFYR
KMVYKQIPGPNKNLPRVFLTSTKGKKEYKPSKEIIEGYEADKHIR
GDKFDLDFCHKLIDFFKESIEKHKDWSKFNFYFSPTESYGDISEF
YLDVEKQGYRMHFENISAETIDEYVEKGDLFLFQIYNKDFVKAA
TGKKDMHTIYWNAAFSPENLQDVVVKLNGEAELFYRDKSDIKE
IVHREGEILVNRTYNGRTPVPDKIHKKLTDYHNGRTKDLGEAKE
YLDKVRYFKAHYDITKDRRYLNDKIYFHVPLTLNFKANGKKNL
NKMVIEKFLSDEKAHIIGIDRGERNLLYYSIIDRSGKIIDQQSLNVI
DGFDYREKLNQREIEMKDARQSWNAIGKIKDLKEGYLSKAVHEI
TKMAIQYNAIVVMEELNYGFKRGRFKVEKQIYQKFENMLIDKM
NYLVFKDAPDESPGGVLNAYQLTNPLESFAKLGKQTGILFYVPA
AYTSKIDPTTGFVNLFNTSSKTNAQERKEFLQKFESISYSAKDGG

IFAFAFDYRKFGTSKTDHKNVWTAYTNGERMRYIKEKKRNELFD
PSKEIKEALTSSGIKYDGGQNILPDILRSNNNGLIYTMYSSFIAAI
QMRVYDGKEDYIISPIKNSKGEFFRTDPKRRELPIDADANGAYNI
ALRGELTMRAIAEKFDPDSEKMAKLELKHKDWFEFMQTRGD
（SEQ ID NO: 104）

[1514] >WP_027407524_（修飾的）假定蛋白[厭氧弧菌屬（Anaerovibrio）某種RM50]

[1515]

MVAFIDEFVGQYPVSKTLRFEARPVPETKKWLESDQCSVLF
NDQKRNEYYGVLKELLDDYYRAYIEDALTSFTLDKALLENAYD
LYCNRDTNAFSSCCEKLRKDLVKAFGNLKDYLLGSDQLKDLVK
LKAKVDAPAGKGKKKIEVDSRLINWLNNNAKYSAEDREKYIKA
IESFEGFVTYLTNYKQARENMFSSEDKSTAIAFRVIDQNMVTYFG
NIRIYEKIKAKYPELYSALKGFEKFFSPTAYSEILSQSKIDEYNYQ
CIGRPIDDADFKGVNSLINEYRQKNGIKARELPVMSMLYKQILSD
RDNSFMSEVINRNEEAIECAKNGYKVSYALFNELLQLYKKIFTED
NYGNIYVKTQPLTELSQALFGDWSILRNALDNGKYDKDIINLAE
LEKYFSEYCKVLDADDAAKIQDKFNLKDYFIQKNALDATLPDLD
KITQYKPHLDAMLQAIRKYKLFSMYNGRKKMDVPENGIDFSNE
FNAIYDKLSEFSILYDRIRNFATKKPYSDEKMKLSFNMPTMLAG
WDYNNETANGCFLFIKDGKYFLGVADSKSKNIFDFKKNPHLLDK
YSSKDIYYKVKYKQVSGSAKMLPKVVFAGSNEKIFGHLISKRILE
IREKKLYTAAAGDRKAVAEWIDFMKSAIAIHPEWNEYFKFKFKN

TAEYDNANKFYEDIDKQTYSLEKVEIPTEYIDEMVSQHKLYLFQ
LYTKDFSDKKKKKGTDNLHTMYWHGVFSDENLKAVTEGTQPII
KLNGEAEMFMRNPSIEFQVTHEHNKPIANKNPLNTKKESVFNYD
LIKDKRYTERKFYFHCPITLNFRADKPIKYNEKINRFVENNPDVC
IIGIDRGERHLLYYTVINQTGDILEQGSLNKISGSYTNDKGEKVN
KETDYHDLLDRKEKGKHVAQQAWETIENIKELKAGYLSQVVYK
LTQLMLQYNAVIVLENLNVGFKRGRTKVEKQVYQKFEKAMIDK
LNYLVFKDRGYEMNGSYAKGLQLTDKFESFDKIGKQTGCIYYVI
PSYTSHIDPKTGFVNLLNAKLRYENITKAQDTIRKFDSISYNAKA
DYFEFAFDYRSFGVDMARNEWVVCTCGDLRWEYSAKTRETKA
YSVTDRLKELFKAHGIDYVGGENLVSHITEVADKHFLSTLLFYLR
LVLKMRYTVSGTENENDFILSPVEYAPGKFFDSREATSTEPMNAD
ANGAYHIALKGLMTIRGIEDGKLHNYGKGGENAAWFKFMQNQE
YKNNG（SEQ ID NO: 105）

[1516] >WP_044910712_（修飾的）假定蛋白[毛螺菌科細菌MC2017]

[1517]

MDYGNGQFERRAPLTKTITLRLKPIGETRETIREQKLLEQDA
AFRKLVETVTPIVDDCIRKIADNALCHFGTEYDFSCLGNAISKND
SKAIKKETEKVEKLLAKVLTENLPDGLRKVNDINSAAFIQDTLTS
FVQDDADKRVLIQELKGKTVLMQRFLTTRITALTVWLPDRVFEN
FNIFIENAEKMRILLDSPLNEKIMKFDPDAEQYASLEFYGQCLSQ
KDIDSYNLIISGIYADDEVKNPGINEIVKEYNQQIRGDKDESPLPK

LKKLHKQILMPVEKAFFVRVLSNDSDARSILEKILKDTEMLPSKII
EAMKEADAGDIAVYGSRLHELSHVIYGDHGKLSQIIYDKESKRIS
ELMETLSPKERKESKKRLEGLEEHIRKSTYTFDELNRYAEKNVM
AAYIAAVEESCAEIMRKEKDLRTLLSKEDVKIRGNRHNTLIVKN
YFNAWTVFRNLIRILRRKSEAEIDSDFYDVLDDSVEVLSLTYKGE
NLCRSYITKKIGSDLKPEIATYGSALRPNSRWWSPGEKFNVKFHT
IVRRDGRLYYFILPKGAKPVELEDMDGDIECLQMRKIPNPTIFLP
KLVFKDPEAFFRDNPEADEFVFLSGMKAPVTITRETYEAYRYKL
YTVGKLRDGEVSEEEYKRALLQVLTAYKEFLENRMIYADLNFGF
KDLEEYKDSSEFIKQVETHNTFMCWAKVSSSQLDDLVKSGNGLL
FEIWSERLESYYKYGNEKVLRGYEGVLLSILKDENLVSMRTLLN
SRPMLVYRPKESSKPMVVHRDGSRVVDRFDKDGKYIPPEVHDE
LYRFFNNLLIKEKLGEKARKILDNKKVKVKVLESERVKWSKFYD
EQFAVTFSVKKNADCLDTTKDLNAEVMEQYSESNRLILIRNTTDI
LYYLVLDKNGKVLKQRSLNIINDGARDVDWKERFRQVTKDRNE
GYNEWDYSRTSNDLKEVYLNYALKEIAEAVIEYNAILIIEKMSNA
FKDKYSFLDDVTFKGFETKLLAKLSDLHFRGIKDGEPCSFTNPLQ
LCQNDSNKILQDGVIFMVPNSMTRSLDPDTGFIFAINDHNIRTKK
AKLNFLSKFDQLKVSSEGCLIMKYSGDSLPTHNTDNRVWNCCC
NHPITNYDRETKKVEFIEEPVEELSRVLEENGIETDTELNKLNER
ENVPGKVVDAIYSLVLNYLRGTVSGVAGQRAVYYSPVTGKKYDI
SFIQAMNLNRKCDYYRIGSKERGEWTDFVAQLIN（SEQ ID NO:
106）

[1518] >WP_027216152_（修飾的）假定蛋白[溶纖維丁酸弧

菌]

[1519]

MYYESLTKLYPIKKTIRNELVPIGKTLENIKKNNILEADEDRK
IAYIRVKAIMDDYHKRLINEALSGFALIDLDKAANLYLSRSKSAD
DIESFSRFQDKLRKAIAKRLREHENFGKIGNKDIIPLLQKLSENED
DYNALESFKNFYTYFESYNDVRLNLYSDKEKSSTVAYRLINENLP
RFLDNIRAYDAVQKAGITSEELSSEAQDGLFLVNTFNNVLIQDGI
NTYNEDIGKLNVAINLYNQKNASVQGFRKVPKMKVLYKQILSDR
EESFIDEFESDTELLDSLESHYANLAKYFGSNKVQLLFTALRESK
GVNVYVKNDIAKTSFSNVVFGSWSRIDELINGEYDDNNNRKKD
EKYYDKRQKELKKNKSYTIEKIITLSTEDVDVIGKYIEKLESDID
DIRFKGKNFYEAVLCGHDRSKKLSKNKGAVEAIKGYLDSVKDFE
RDLKLINGSGQELEKNLVVYGEQEAVLSELSGIDSLYNMTRNYL
TKKPFSTEKIKLNFNKPTFLDGWDYGNEEAYLGFFMIKEGNYFL
AVMDANWNKEFRNIPSVDKSDCYKKVIYKQISSPEKSIQNLMVI
DGKTVKKNGRKEKEGIHSGENLILEELKNTYLPKKINDIRKRRS
YLNGDTFSKKDLTEFIGYYKQRVIEYYNGYSFYFKSDDDYASFK
EFQEDVGRQAYQISYVDVPVSFVDDLINSGKLYLFRVYNKDFSE
YSKGRLNLHTLYFKMLFDERNLKNVVYKLNGQAEVFYRPSSIK
KEELIVHRAGEEIKNKNPKRAAQKPTRRLDYDIVKDRRYSQDKF
MLHTSIIMNFGAEENVSFNDIVNGVLRNEDKVNVIGIDRGERNL
LYVVVIDPEGKILEQRSLNCITDSNLDIETDYHRLLDEKESDRKIA
RRDWTTIENIKELKAGYLSQVVHIVAELVLKYNAIICLEDLNFGF
KRGRQKVEKQVYQKFEKMLIDKLNYLVMDKSREQLSPEKISGA

LNALQLTPDFKSFKVLGKQTGIIYYVPAYLTSKIDPMTGFANLFY
VKYENVDKAKEFFSKFDSIKYNKDGKNWNTKGYFEFAFDYKKF
TDRAYGRVSEWTVCTVGERIIKFKNKEKNNSYDDKVIDLTNSLK
ELFDSYKVTYESEVDLKDAILAIDDPAFYRDLTRRLQQTLQMRN
SSCDGSRDYIISPVKNSKGEFFCSDNNDDTTPNDADANGAFNIAR
KGLWVLNEIRNSEEGSKINLAMSNAQWLEYAQDNTI（SEQ ID
NO: 107）

[1520] >WP_016301126_（修飾的）假定蛋白[毛螺菌科細菌
COE1]

[1521]

MHENNGKIADNFIGIYPVSKTLRFELKPVGKTQEYIEKHGIL
DEDLKRAGDYKSVKKIIDAYHKYFIDEALNGIQLDGLKNYYELY
EKKRDNNEEKEFQKIQMSLRKQIVKRFSEHPQYKYLFKKELIKN
VLPEFTKDNAEEQTLVKSFQEFTTYFEGFHQNRKNMYSDEEKST
AIAYRVVHQNLPKYIDNMRIFSMILNTDIRSDLTELFNNLKTKMD
ITIVEEYFAIDGFNKVVNQKGIDVYNTILGAFSTDDNTKIKGLNE
YINLYNQKNKAKLPKLKPLFKQILSDRDKISFIPEQFDSDTEVLEA
VDMFYNRLLQFVIENEGQITISKLLTNFSAYDLNKIYVKNDTTIS
AISNDLFDDWSYISKAVRENYDSENVDKNKRAAAYEEKKEKAL
SKIKMYSIEELNFFVKKYSCNECHIEGYFERRILEILDKMRYAYES
CKILHDKGLINNISLCQDRQAISELKDFLDSIKEVQWLLKPLMIG
QEQADKEEAFYTELLRIWEELEPITLLYNKVRNYVTKKPYTLEK
VKLNFYKSTLLDGWDKNKEKDNLGIILLKDGQYYLGIMNRRNN

KIADDAPLAKTDNVYRKMEYKLLTKVSANLPRIFLKDKYNPSEE
MLEKYEKGTHLKGENFCIDDCRELIDFFKKGIKQYEDWGQFDFK
FSDTESYDDISAFYKEVEHQGYKITFRDIDETYIDSLVNEGKLYLF
QIYNKDFSPYSKGTKNLHTLYWEMLFSQQNLQNIVYKLNGNAEI
FYRKASINQKDVVVHKADLPIKNKDPQNSKKESMFDYDIIKDKR
FTCDKYQFHVPITMNFKALGENHFNRKVNRLIHDAENMHIIGID
RGERNLIYLCMIDMKGNIVKQISLNEIISYDKNKLEHKRNYHQLL
KTREDENKSARQSWQTIHTIKELKEGYLSQVIHVITDLMVEYNA
IVVLEDLNFGFKQGRQKFERQVYQKFEKMLIDKLNYLVDKSKG
MDEDGGLLHAYQLTDEFKSFKQLGKQSGFLYYIPAWNTSKLDPT
TGFVNLFYTKYESVEKSKEFINNFTSILYNQEREYFEFLFDYSAFT
SKAEGSRLKWTVCSKGERVETYRNPKKNNEWDTQKIDLTFELK
KLFNDYSISLLDGDLREQMGKIDKADFYKKFMKLFALIVQMRNS
DEREDKLISPVLNKYGAFFETGKNERMPLDADANGAYNIARKG
LWIIEKIKNTDVEQLDKVKLTISNKEWLQYAQEHIL（SEQ ID NO:
108）

[1522] >WP_035635841_（修飾的）假定蛋白[毛螺菌科細菌
ND2006]

[1523]

MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDE
KRAEDYKGVKKLLDRYYLSFINDVLHSIKLKNLNNYISLFRKKT
RTEKENKELENLEINLRKEIAKAFKGNEGYKSLFKKDIIETILPEF
LDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRCI

NENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFE
GEFFNFVLTQEGIDVYNAIIGGFVTESGEKIKGLNEYINLYNQKT
KQKLPKFKPLYKQVLSDRESLSFYGEGYTSDEEVLEVFRNTLNK
NSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNV
IRDKWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQ
EYADADLSVVEKLKEIIIQKVDEIYKVYGSSEKLFDADFVLEKSL
KKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFYGDF
VLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGG
WDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNG
NYEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNG
TFKKGDMFNLNDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKY
KDIAGFYREVEEQGYKVSFESASKKEVDKLVEEGKLYMFQIYNK
DFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASL
KKEELVVHPANSPIANKNPDNPKKTTTLSYDVYKDKRFSEDQYE
LHIPIAINKCPKNIFKINTEVRVLLKHDDNPYVIGIDRGERNLLYI
VVVDGKGNIVEQYSLNEIINNFNGIRIKTDYHSLLDKKEKERFEA
RQNWTSIENIKELKAGYISQVVHKICELVEKYDAVIALEDLNSGF
KNSRVKVEKQVYQKFEKMLIDKLNYMVDKKSNPCATGGALKG
YQITNKFESFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLKTKYT
SIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWK
LYSYGNRIRIFRNPKKNNVFDWEEVCLTSAYKELFNKYGINYQQ
GDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDFLIS
PVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIG
QFKKAEDEKLDKVKIAISNKEWLEYAQTSVKH (SEQ ID NO: 109)

[1524] >WP_015504779_（修飾的）外切核酸酶SbcC [候選*alvus*甲烷嗜甲基菌（Candidatus Methanomethylophilus alvus）]

[1525]

MDAKEFTGQYPLSKTLRFELRPIGRTWDNLEASGYLAEDRH
RAECYPRAKELLDDNHRAFLNRVLPQIDMDWHPIAEAFCKVHK
NPGNKELAQDYNLQLSKRRKEISAYLQDADGYKGLFAKPALDE
AMKIAKENGNESDIEVLEAFNGFSVYFTGYHESRENIYSDEDMV
SVAYRITEDNFPRFVSNALIFDKLNESHPDIISEVSGNLGVDDIGK
YFDVSNYNNFLSQAGIDDYNHIIGGHTTEDGLIQAFNVVLNLRH
QKDPGFEKIQFKQLYKQILSVRTSKSYIPKQFDNSKEMVDCICDY
VSKIEKSETVERALKLVRNISSFDLRGIFVNKKNLRILSNKLIGD
WDAIETALMHSSSSENDKKSVYDSAEAFTLDDIFSSVKKFSDAS
AEDIGNRAEDICRVISETAPFINDLRAVDLDSLNDDGYEAAVSKIR
ESLEPYMDLFHELEIFSVGDEFPKCAAFYSELEEVSEQLIEIIPLFN
KARSFCTRKRYSTDKIKVNLKFPTLADGWDLNKERDNKAAILR
KDGKYYLAILDMKKDLSSIRTSDEDESSFEKMEYKLLPSPVKML
PKIFVKSKAAKEKYGLTDRMLECYDKGMHKSGSAFDLGFCHELI
DYYKRCIAEYPGWDVFDFKFRETSDYGSMKEFNEDVAGAGYY
MSLRKIPCSEVYRLLDEKSIYLFQIYNKDYSENAHGNKNMHTM
YWEGLFSPQNLESPVFKLSGGAELFFRKSSIPNDAKTVHPKGSVL
VPRNDVNGRRIPDSIYRELTRYFNRGDCRISDEAKSYLDKVKTK
KADHDIVKDRRFTVDKMMFHVPIAMNFKAISKPNLNKKVIDGII
DDQDLKIIGIDRGERNLIYVTMVDRKGNILYQDSLNILNGYDYR
KALDVREYDNKEARRNWTKVEGIRKMKEGYLSLAVSKLADMII

ENNAIIVMEDLNHGFKAGRSKIEKQVYQKFESMLINKLGYMVL
KDKSIDQSGGALHGYQLANHVTTLASVGKQCGVIFYIPAAFTSKI
DPTTGFADLFALSNVKNVASMREFFSKMKSVIYDKAEGKFAFTF
DYLDYNVKSECGRTLWTVYTVGERFTYSRVNREYVRKVPTDIIY
DALQKAGISVEGDLRDRIAESDGDTLKSIFYAFKYALDMRVENR
EEDYIQSPVKNASGEFFCSKNAGKSLPQDSDANGAYNIALKGIL
QLRMLSEQYDPNAESIRLPLITNKAWLTFMQSGMKTWKN（SEQ
ID NO: 110）

[1526] >WP_044910713_（修飾的）假定蛋白[毛螺菌科細菌
MC2017]

[1527]

MGLYDGFVNRYSVSKTLRFELIPQGRTREYIETNGILSDDEE
RAKDYKTIKRLIDEYHKDYISRCLKNVNISCLEEYYHLYNSSNRD
KRHEELDALSDQMRGEIASFLTGNDEYKEQKSRDIIINERIINFAS
TDEELAAVKRFRKFTSYFTGFFTNRENMYSAEKKSTAIAHRIIDV
NLPKYVDNIKAFNTAIEAGVFDIAEFESNFKAITDEHEVSDLLDIT
KYSRFIRNEDIIIYNTLLGGISMKDEKIQGLNELINLHNQKHPGKK
VPLLKVLYKQILGDSQTHSFVDDQFEDDQQVINAVKAVTDTFSE
TLLGSLKIIINNIGHYDLDRIYIKAGQDITTLSKRALNDWHIITECL
ESEYDDKFPKNKKSDTYEEMRNRYVKSFKSFSIGRLNSLVTTYT
EQACFLENYLGSFGGDTDKNCLTDFTNSLMEVEHLLNSEYPVTN
RLITDYESVRILKRLLDSEMEVIHFLKPLLGNGNESDKDLVFYGE
FEAEYEKLLPVIKVYNRVRNYLTRKPFSTEKIKLNFNSPTLLCGW

SQSKEKEYMGVILRKDGQYYLGIMTPSNKKIFSEAPKPDEDCYE
KMVLRYIPHPYQMLPKVFFSKSNIAFFNPSDEILRIKKQESFKKG
KSFNRDDCHKFIDFYKDSINRHEEWRKFNFKFSDTDSYEDISRFY
KEVENQAFSMSFTKIPTVYIDSLVDEGKLYLFKLHNKDFSEHSKG
KPNLHTVYWNALFSEYNLQNTVYQLNGSAEIFFRKASIPENERVI
HKKNVPITRKVAELNGKKEVSVFPYDIIKNRRYTVDKFQFHVPL
KMNFKADEKKRINDDVIEAIRSNKGIHVIGIDRGERNLLYLSLIN
EEGRIIEQRSLNIIDSGEGHTQNYRDLLDSREKDREKARENWQEI
QEIKDLKTGYLSQAIHTITKWMKEYNAIIVLEDLNDRFTNGRKK
VEKQVYQKFEKMLIDKLNYYVDKDEEFDRMGGTHRALQLTEKF
ESFQKLGRQTGFIFYVPAWNTSKLDPTTGFVDLLYPKYKSVDAT
KDFIKKFDFIRFNSEKNYFEFGLHYSNFTERAIGCRDEWILCSYG
NRIVNFRNAAKNNSWDYKEIDITKQLLDLFEKNGIDVKQENLID
SICEMKDKPFFKSLIANIKLILQIRNSASGTDIDYMISPAMNDRGE
FFDTRKGLQQLPLDADANGAYNIAKKGLWIVDQIRNTTGNNVK
MAMSNREWMHFAQESRLA（SEQ ID NO: 111）

[1528] >KKQ36153_（修飾的）假定蛋白US52_C0007G0008 [候選的分裂WS6細菌GW2011_GWA2_37_6]

[1529]

MKNVFGGFTNLYSLTKTLRFELKPTSKTQKLMKRNNVIQTD
EEIDKLYHDEMKPILDEIHRRFINDALAQKIFISASLDNFLKVVKN
YKVESAKKNIKQNQVKLLQKEITIKTLGLRREVVSGFITVSKKW
KDKYVGLGIKLKGDGYKVLTEQAVLDILKIEFPNKAKYIDKFRG

FWTYFSGFNENRKNYYSEEDKATSIANRIVNENLSRYIDNIIAFEE
ILQKIPNLKKFKQDLDITSYNYYLNQAGIDKYNKIIGGYIVDKDK
KIQGINEKVNLYTQQTKKKLPKLKFLFKQIGSERKGFGIFEIKEG
KEWEQLGDLFKLQRTKINSNGREKGLFDSLRTMYREFFDEIKRD
SNSQARYSLDKIYFNKASVNTISNSWFTNWNKFAELLNIKEDKK
NGEKKIPEQISIEDIKDSLSIIPKENLEELFKLTNREKHDRTRFFGS
NAWVTFLNIWQNEIEESFNKLEEKEKDFKKNAAIKFQKNNLVQK
NYIKEVCDRMLAIERMAKYHLPKDSNLSREEDFYWIIDNLSEQR
EIYKYYNAFRNYISKKPYNKSKMKLNFENGNLLGGWSDGQERN
KAGVILRNGNKYYLGVLINRGIFRTDKINNEIYRTGSSKWERLIL
SNLKFQTLAGKGFLGKHGVSYGNMNPEKSVPSLQKFIRENYLK
KYPQLTEVSNTKFLSKKDFDAAIKEALKECFTMNFINIAENKLLE
AEDKGDLYLFEITNKDFSGKKSGKDNIHTIYWKYLFSESNCKSPI
IGLNGGAEIFFREGQKDKLHTKLDKKGKKVFDAKRYSEDKLFFH
VSITINYGKPKNIKFRDIINQLITSMNVNIIGIDRGEKHLLYYSVID
SNGIILKQGSLNKIRVGDKEVDFNKKLTERANEMKKARQSWEQI
GNIKNFKEGYLSQAIHEIYQLMIKYNAIIVLEDLNTEFKAKRLSK
VEKSVYKKFELKLARKLNHLILKDRNTNEIGGVLKAYQLTPTIG
GGDVSKFEKAKQWGMMFYVRANYTSTTDPVTGWRKHLYISNF
SNNSVIKSFFDPTNRDTGIEIFYSGKYRSWGFRYVQKETGKKWE
LFATKELERFKYNQTTKLCEKINLYDKFEELFKGIDKSADIYSQL
CNVLDFRWKSLVYLWNLLNQIRNVDKNAEGNKNDFIQSPVYPF
FDSRKTDGKTEPINGDANGALNIARKGLMLVERIKNNPEKYEQL
IRDTEWDAWIQNFNKVN (SEQ ID NO: 112)

[1530] >WP_044919442_(修飾的)假定蛋白[毛螺菌科細菌MA2020]

[1531]

MYYESLTKQYPVSKTIRNELIPIGKTLDNIRQNNILESDVKR
KQNYEHVKGILDEYHKQLINEALDNCTLPSLKIAAEIYLKNQKE
VSDREDFNKTQDLLRKEVVEKLKAHENFTKIGKKDILDLLEKLP
SISEDDYNALESFRNFYTYFTSYNKVRENLYSDKEKSSTVAYRLI
NENFPKFLDNVKSYRFVKTAGILADGLGEEEQDSLFIVETFNKTL
TQDGIDTYNSQVGKINSSINLYNQKNQKANGFRKIPKMKMLYKQ
ILSDREESFIDEFQSDEVLIDNVESYGSVLIESLKSSKVSAFFDAL
RESKGKNVYVKNDLAKTAMSNIVFENWRTFDDLLNQEYDLANE
NKKKDDKYFEKRQKELKKNKSYSLEHLCNLSEDSCNLIENYIHQ
ISDDIENIIINNETFLRIVINEHDRSRKLAKNRKAVKAIKDFLDSIK
VLERELKLINSSGQELEKDLIVYSAHEELLVELKQVDSLYNMTR
NYLTKKPFSTEKVKLNFNRSTLLNGWDRNKETDNLGVLLLKDG
KYYLGIMNTSANKAFVNPPVAKTEKVFKKVDYKLLPVPNQMLP
KVFFAKSNIDFYNPSSEIYSNYKKGTHKKGNMFSLEDCHNLIDFF
KESISKHEDWSKFGFKFSDTASYNDISEFYREVEKQGYKLTYTDI
DETYINDLIERNELYLFQIYNKDFSMYSKGKLNLHTLYFMMLFD
QRNIDDVVYKLNGEAEVFYRPASISEDELIIHKAGEEIKNKNPNR
ARTKETSTFSYDIVKDKRYSKDKFTLHIPITMNFGVDEVKRFND
AVNSAIRIDENVNVIGIDRGERNLLYVVVIDSKGNILEQISLNSIIN
KEYDIETDYHALLDEREGGRDKARKDWNTVENIRDLKAGYLSQ
VVNVVAKLVLKYNAIICLEDLNFGFKRGRQKVEKQVYQKFEKM

LIDKLNYLVIDKSREQTSPKELGGALNALQLTSKFKSFKELGKQS
GVIYYVPAYLTSKIDPTTGFANLFYMKCENVEKSKRFFDGFDFIR
FNALENVFEFGFDYRSFTQRACGINSKWTVCTNGERIIKYRNPD
KNNMFDEKVVVVTDEMKNLFEQYKIPYEDGRNVKDMIISNEEA
EFYRRLYRLLQQTLQMRNSTSDGTRDYIISPVKNKREAYFNSELS
DGSVPKDADANGAYNIARKGLWVLEQIRQKSEGEKINLAMTNA
EWLEY AQTHLL（SEQ ID NO: 113）

[1532] >WP_035798880_（修飾的）假定蛋白[丁酸弧菌屬某種NC3005]

[1533]

MYYQNLTKKYPVSKTIRNELIPIGKTLENIRKNNILESDVKR
KQDYEHVKGIMDEYHKQLINEALDNYMLPSLNQAAEIYLKKHV
DVEDREEFKKTQDLLRREVTGRLKEHENYTKIGKKDILDLLEKL
PSISEEDYNALESFRNFYTYFTSYNKVRENLYSDEEKSSTVAYRLI
NENLPKFLDNIKSYAFVKAAGVLADCIEEEEQDALFMVETFNMT
LTQEGIDMYNYQIGKVNSAINLYNQKNHKVEEFKKIPKMKVLYK
QILSDREEVFIGEFKDDETLLSSIGAYGNVLMTYLKSEKINIFFDA
LRESEGKNVYYVKNDLSKTTMSNIVFGSWSAFDELLNQEYDLAN
ENKKKDDKYFEKRQKELKKNKSYTLEQMSNLSKEDISPIENYIE
RISEDIEKICIYNGEFEKIVVNEHDSSRKLSKNIKAVKVIKDYLDSI
KELEHDIKLINGSGQELEKNLVVYVGQEEALEQLRPVDSLYNLT
RNYLTKKPFSTEKVKLNFNKSTLLNGWDKNKETDNLGILFFKDG
KYYLGIMNTTANKAFVNPPAAKTENVFKKVDYKLLPGSNKMLP

KVFFAKSNIGYYNPSTELYSNYKKGTHKKGPSFSIDDCHNLIDFF
KESIKKHEDWSKFGFEFSDTADYRDISEFYREVEKQGYKLTFTDI
DESYINDLIEKNELYLFQIYNKDFSEYSKGKLNLHTLYFMMLFDQ
RNLDNVVYKLNGEAEVFYRPASIAENELVIHKAGEGIKNKNPNR
AKVKETSTFSYDIVKDKRYSKYKFTLHIPITMNFGVDEVRRFND
VINNALRTDDNVNVIGIDRGERNLLYVVVINSEGKILEQISLNSII
NKEYDIETNYHALLDEREDDRNKARKDWNTIENIKELKTGYLS
QVVNVVAKLVLKYNAIICLEDLNFGFKRGRQKVEKQVYQKFEK
MLIEKLNYLVIDKSREQVSPEKMGGALNALQLTSKFKSFAELGK
QSGIIYYVPAYLTSKIDPTTGFVNLFYIKYENIEKAKQFFDGFDFI
RFNKKDDMFEFSFDYKSFTQKACGIRSKWIVYTNGERIIKYPNPE
KNNLFDEKVINVTDEIKGLFKQYRIPYENGEDIKEIIISKAEADFY
KRLFRLLHQTLQMRNSTSDGTRDYIISPVKNDRGEFFCSEFSEGT
MPKDADANGAYNIARKGLWVLEQIRQKDEGEKVNLSMTNAEW
LKYAQLHLL（SEQ ID NO: 114）

[1534] >WP_027109509_（修飾的）假定蛋白[毛螺菌科細菌NC2008]

[1535]

MENYYDSLTRQYPVTKTIRQELKPVGKTLENIKNAEIIEADK
QKKEAYVKVKELMDEFHKSIIEKSLVGIKLDGLSEFEKLYKIKTK
TDEDKNRISELFYYMRKQIADALKNSRDYGYVDNKDLIEKILPE
RVKDENSLNALSCFKGFTTYFTDYYKNRKNIYSDEEKHSTVGYR
CINENLLIFMSNIEVYQIYKKANIKNDNYDEETLDKTFMIESFNE

CLTQSGVEAYNSVVASIKTATNLYIQKNNKEENFVRVPKMKVLF
KQILSDRTSLFDGLIIESDDELLDKLCSFSAEVDKFLPINIDRYIKT
LMDSNNGTGIYVKNDSSLTTLSNYLTDSWSSIRNAFNENYDAKY
TGKVNDKYEEKREKAYKSNDSFELNYIQNLLGINVIDKYIERINF
DIKEICEAYKEMTKNCFEDHDKTKKLQKNIKAVASIKSYLDSLK
NIERDIKLLNGTGLESRNEFFYGEQSTVLEEITKVDELYNITRNYL
TKKPFSTEKMKLNFNNPQLLGGWDVNKERDCYGVILIKDNNYY
LGIMDKSANKSFLNIKESKNENAYKKVNCKLLPGPNKMFPKVFF
AKSNIDYYDPTHEIKKLYDKGTFKKGNSFNLEDCHKLIDFYKESI
KKNDDWKNFNFNFSDTKDYEDISGFFREVEAQNYKITYTNVSC
DFIESLVDEGKLYLFQIYNKDFSEYATGNLNLHTLYLKMLFDERN
LKDLCIKMNGEAEVFYRPASILDEDKVVHKANQKITNKNTNSKK
KESIFSYDIVKDKRYTVDKFFIHLPITLNYKEQNVSRFNDYIREIL
KKSKNIRVIGIDRGERNLLYVVVCDSDGSILYQRSINEIVSGSHKT
DYHKLLDNKEKERLSSRRDWKTIENIKDLKAGYMSQVVNEIYN
LILKYNAIVVLEDLNIGFKNGRKKVEKQVYQNFEKALIDKLNYL
CIDKTREQLSPSSPGGVLNAYQLTAKFESFEKIGKQTGCIFYVPAY
LTSQIDPTTGFVNLFYQKDTSKQGLQLFFRKFKKINFDKVASNFE
FVFDYNDFTNKAEGTKTNWTISTQGTRIAKYRSDDANGKWISRT
VHPTDIIKEALNREKINYNDGHDLIDEIVSIEKSAVLKEIYYGFKL
TLQLRNSTLANEEEQEDYIISPVKNSSGNYFDSRITSKELPCDAD
ANGAYNIARKGLWALEQIRNSENVSKVKLAISNKEWFEYTQNNI
PSL（SEQ ID NO: 1581）

[1536] >WP_029202018_（修飾的）假定蛋白[口小桿菌

屬某種（Oribacterium sp.）NK2B42]

[1537]

MYYDGLTKQYALSKTIRNELVPIGKTLDNIKKNRILEADIKR
KSDYEHVKKLMDMYHKKIINEALDNFKLSVLEDAADIYFNKQN
DERDIDAFLKIQDKLRKEIVEQLKGHTDYSKVGNKDFLGLLKAA
STEEDRILIESFDNFYTYFTSYNKVRSNLYSAEDKSSTVAYRLINE
NLPKFFDNIKAYRTVRNAGVISGDMSIVEQDELFEVDTFNHTLTQ
YGIDTYNHMIGQLNSAINLYNQKMHGAGSFKKLPKMKELYKQL
LTEREEEFIEEYTDDEVLITSVHNYVSYLIDYLNSDKVESFFDTLR
KSDGKEVFIKNDVSKTTMSNILFDNWSTIDDLINHEYDSAPENV
KKTKDDKYFEKRQKDLKKNKSYSLSKIAALCRDTTILEKYIRRL
VDDIEKIYTSNNVFSDIVLSKHDRSKKLSKNTNAVQAIKNMLDSI
KDFEHDVMLINGSGQEIKKNLNVYSEQEALAGILRQVDHIYNLT
RNYLTKKPFSTEKIKLNFNRPTFLDGWDKNKEEANLGILLIKDNR
YYLGIMNTSSNKAFVNPPKAISNDIYKKVDYKLLPGPNKMLPKV
FFATKNIAYYAPSEELLSKYRKGTHKKGDSFSIDDCRNLIDFFKSS
INKNTDWSTFGFNFSDTNSYNDISDFYREVEKQGYKLSFTDIDA
CYIKDLVDNNELYLFQIYNKDFSPYSKGKLNLHTLYFKMLFDQR
NLDNVVYKLNGEAEVFYRPASIESDEQIIHKSGQNIKNKNQKRS
NCKKTSTFDYDIVKDRRYCKDKFMLHLPITVNFGTNESGKFNEL
VNNAIRADKDVNVIGIDRGERNLLYVVVVDPCGKIIEQISLNTIV
DKEYDIETDYHQLLDEKEGSRDKARKDWNTIENIKELKEGYLSQ
VVNIIAKLVLKYDAIICLEDLNFGFKRGRQKVEKQVYQKFEKML
IDKMNYLVLDKSRKQESPQKPGGALNALQLTSAFKSFKELGKQT

GIIYYVPAYLTSKIDPTTGFANLFYIKYESVDKARDFFSKFDFIRY
NQMDNYFEFGFDYKSFTERASGCKSKWIACTNGERIVKYRNSD
KNNSFDDKTVILTDEYRSLFDKYLQNYIDEDDLKDQILQIDSADF
YKNLIKLFQLTLQMRNSSSDGKRDYIISPVKNYREEFFCSEFSDD
TFPRDADANGAYNIARKGLWVIKQIRETKSGTKINLAMSNSEWL
EYAQCNLL（SEQ ID NO: 115）

[1538] >WP_028248456_（修飾的）假定蛋白[瘤胃假丁酸弧菌（Pseudobutyrivibrio ruminis）]

[1539]

MYYQNLTKMYPISKTLRNELIPVGKTLENIRKNGILEADIQR
KADYEHVKKLMDNYHKQLINEALQGVHLSDLSDAYDLY

[1540]

FNLSKEKNSVDAFSKCQDKLRKEIVSLLKNHENFPKIGNKEI
IKLLQSLYDNDTDYKALDSFSNFYTYFSSYNEVRKNLYSDEEKSS
TVAYRLINENLPKFLDNIKAYAIAKKAGVRAEGLSEEDQDCLFIIE
TFERTLTQDGIDNYNAAIGKLNTAINLFNQQNKKQEGFRKVPQM
KCLYKQILSDREEAFIDEFSDDEDLITNIESFAENMNVFLNSEIITD
FKIALVESDGSLVYIKNDVSKTSFSNIVFGSWNAIDEKLSDEYDL
ANSKKKKDEKYYEKRQKELKKNKSYDLETIIGLFDDNSDVIGKY
IEKLESDITAIAEAKNDFDEIVLRKHDKNKSLRKNTNAVEAIKSY
LDTVKDFERDIKLINGSGQEVEKNLVVYAEQENILAEIKNVDSLY
NMSRNYLTQKPFSTEKFKLNFNRATLLNGWDKNKETDNLGILFE
KDGMYYLGIMNTKANKIFVNIPKATSNDVYHKVNYKLLPGPNK

MLPKVFFAQSNLDYYKPSEELLAKYKAGTHKKGDNFSLEDCHA
LIDFFKASIEKHPDWSSFGFEFSETCTYEDLSGFYREVEKQGYKI
TYTDVDADYITSLVERDELYLFQIYNKDFSPYSKGNLNLHTIYLQ
MLFDQRNLNNVVYKLNGEAEVFYRPASINDEEVIIHKAGEEIKN
KNSKRAVDKPTSKFGYDIIKDRRYSKDKFMLHIPVTMNFGVDET
RRFNDVVNDALRNDEKVRVIGIDRGERNLLYVVVVDTDGTILEQ
ISLNSIINNEYSIETDYHKLLDEKEGDRDRARKNWTTIENIKELKE
GYLSQVVNVIAKLVLKYNAIICLEDLNFGFKRGRQKVEKQVYQ
KFEKMLIDKLNYLVIDKSRKQDKPEEFGGALNALQLTSKFTSFK
DMGKQTGIIYYVPAYLTSKIDPTTGFANLFYVKYENVEKAKEFFS
RFDSISYNNESGYFEFAFDYKKFTDRACGARSQWTVCTYGERII
KFRNTEKNNSFDDKTIVLSEEFKELFSIYGISYEDGAELKNKIMS
VDEADFFRSLTRLFQQTMQMRNSSNDVTRDYIISPIMNDRGEFF
NSEACDASKPKDADANGAFNIARKGLWVLEQIRNTPSGDKLNL
AMSNAEWLEYAQRNQI （SEQ ID NO: 116）

[1541] >WP_028830240_（修飾的）假定蛋白[Proteocatella sphenisci]

[1542]

MENFKNLYPINKTLRFELRPYGKTLENFKKSGLLEKDAFKA
NSRRSMQAIIDEKFKETIEERLKYTEFSECDLGNMTSKDKKITDK
AATNLKKQVILSFDDEIFNNYLKPDKNIDALFKNDPSNPVISTFK
GFTTYFVNFFEIRKHIFKGESSGSMAYRIIDENLTTYLNNIEKIKK
LPEELKSQLEGIDQIDKLNNYNEFITQSGITHYNEIIGGISKSENV

KIQGINEGINLYCQKNKVKLPRLTPLYKMILSDRVSNSFVLDTIEN
DTELIEMISDLINKTEISQDVIMSDIQNIFIKYKQLGNLPGISYSSI
VNAICSDYDNNFGDGKRKKSYENDRKKHLETNVYSINYISELLT
DTDVSSNIKMRYKELEQNYQVCKENFNATNWMNIKNIKQSEKT
NLIKDLLDILKSIQRFYDLFDIVDEDKNPSAEFYTWLSKNAEKLD
FEFNSVYNKSRNYLTRKQYSDKKIKLNFDSPTLAKGWDANKEID
NSTIIMRKFNNDRGDYDYFLGIWNKSTPANEKIIPLEDNGLFEKM
QYKLYPDPSKMLPKQFLSKIWKAKHPTTPEFDKKYKEGRHKKG
PDFEKEFLHELIDCFKHGLVNHDEKYQDVFGFNLRNTEDYNSYT
EFLEDVERCNYNLSFNKIADTSNLINDGKLYVFQIWSKDFSIDSK
GTKNLNTIYFESLFSEENMIEKMFKLSGEAEIFYRPASLNYCEDII
KKGHHHAELKDKFDYPIIKDKRYSQDKFFFHVPMVINYKSEKLN
SKSLNNRTNENLGQFTHIIGIDRGERHLIYLTVVDVSTGEIVEQK
HLDEIINTDTKGVEHKTHYLNKLEEKSKTRDNERKSWEAIETIKE
LKEGYISHVINEIQKLQEKYNALIVMENLNYGFKNSRIKVEKQV
YQKFETALIKKFNYIIDKKDPETYIHGYQLTNPITTLDKIGNQSGI
VLYIPAWNTSKIDPVTGFVNLLYADDLKYKNQEQAKSFIQKIDNI
YFENGEFKFDIDFSKWNNRYSISKTKWTLTSYGTRIQTFRNPQKN
NKWDSAEYDLTEEFKLILNIDGTLKSQDVETYKKFMSLFKLMLQ
LRNSVTGTDIDYMISPVTDKTGTHFDSRENIKNLPADADANGAY
NIARKGIMAIENIMNGISDPLKISNEDYLKYIQNQQE(SEQ ID NO:
117)

[1543] 申請人生成如圖40A-圖40L(例如，PACYC184 fnCpf1（PY001））和圖41A-圖41E(例如，PaCpf1)所示的載體構建體。

[1544] 用於檢測FnCpf1的推定的PAM序列的PAM激發測定（圖42）：申請人從新殺手法蘭西斯菌（Fn）中分離Cpf1座位（圖43）並且將其轉化到大腸桿菌中。在大腸桿菌中表現來自pACYC184的座位，類似於撒普拉那斯卡等人所述的實驗。

具有pACYC-FnCpf1座位的大腸桿菌= Cpf1+

具有空pACYC184的大腸桿菌=對照

[1545] 申請人用PAM文庫質粒轉化Cpf1+和對照大腸桿菌。獲得兩個PAM文庫（圖44）。PAM文庫係含有31 bp原型間隔區序列的pUC19質粒，該原型間隔區序列匹配FnCpf1座位中的間隔區1。PAM左文庫具有在原型間隔區的5'端處的8 nt簡並性PAM。PAM右文庫具有在原型間隔區的3'端處的7 nt簡並性PAM。申請人接種Cpf1+和對照大腸桿菌並且在~12 h之後收穫所有菌落。每個菌落代表不會引起Cpf1進行切割/干擾的PAM-pUC19轉化事件。該等PAM-pUC19質粒不會攜帶可識別PAM。申請人藉由所有菌落的定序確定與對照相比哪些PAM-pUC19質粒不再存在，並且鑒定該等質粒含有可識別PAM。

[1546] pY0001的選殖:pY0001係具有部分FnCpf1座位的pACYC184骨架（來自NEB）。pY0001含有來自第4間隔區序列的255 bp乙醯轉移酶3'序列的內源性FnCpf1座位。僅間隔區1-3係潛在地具有活性的，因為間隔區4不再側接同向重複序列。

[1547] 申請人PCR擴增3個片段的FnCpf1座位並且使用吉布森元件將該等座位選殖到Xba1和Hind3切割的pACYC184中。

[1548] Cpf1 PAM篩選計算分析

在對篩選DNA定序之後，申請人提取出對應於左PAM或右PAM的區域。對於每個樣品，將定序文庫中存在的PAM數目與文庫中預期的PAM數目（對於左文庫為4^8，對於右文庫為4^7）進行比較。

左文庫示出PAM缺失。為了量化此缺失，申請人計算了富集比。針對兩種條件（對照pACYC或含有PaCpf1的pACYC），申請人將文庫中的每種PAM比率計算為：

$$\text{比率} = -\log_2 \frac{\text{樣品} + 0.01}{\text{初始文庫} + 0.01}$$

[1549] 申請人確定繪製的分佈顯示在對照樣品中幾乎沒有富集並且在兩種生物複製本中有富集。申請人收集比率超過8的所有PAM，並且繪製頻率分佈圖，揭示了5' YYN PAM（圖45A-圖45E）。申請人確認PAM係TTN，其中N係A/C/G或T。

[1550] 申請人對tolerances弗朗西絲菌Cpf1座位進行RNA定序並且RNAseq分析顯示CRISPR座位表現活躍（圖46）。FnCpf1座位的RNAseq分析的另一個描繪示出在圖86中。除Cpf1和Cas基因之外，兩種小的非編碼轉錄物被高度轉錄，申請人推測的這兩種非編碼轉錄物係推定的tracrRNA。CRISPR陣列也被表現。兩種推定的tracrRNA和CRISPR陣列以與Cpf1和Cas基因相同的方向進行轉錄。在此藉由RNA定序實驗鑒定的所有RNA轉錄物映射到座位。放大Cpf1 CRISPR陣列，申請人鑒定了許多不同的短轉錄物。在此圖中，將所有鑒定的RNA轉錄物映射到Cpf1座位（圖47）。在選擇

小於85個核苷酸長度的轉錄物之後，申請人鑒定兩種推定的tracrRNA（圖48）。圖49示出了推定的tracrRNA 1和CRISPR陣列的放大的透視圖。圖50示出了推定的tracrRNA 2的放大的透視圖。圖51中指示了推定的crRNA序列。

[1551] 申請人使用U6 PCR產物測試哺乳動物細胞中的功能：間隔區（DR-間隔區-DR）（在某些方面間隔區可以被稱為crRNA或指導RNA或者如本申請所述的類似術語）並且追蹤其他鑒定的Cpf1座位。

實例4：對於FnCpf1的其他確認實驗

[1552] 申請人藉由使用圖52所列出的測定證實預測的FnCpf1 PAM係在體內的TTN。申請人用編碼具有5' TTN PAM的內源性間隔區1的pUC19轉化攜帶FnCpf1座位的細胞和對照細胞（圖53）。簡言之，在體內PAM確認測定中，用10 ng攜帶原型間隔區1的質粒轉化50 μl具有FnCpf1座位（測試菌株）或具有空pACYC184（對照菌株）的感受態大腸桿菌。前述原型間隔區序列係預測的PAM序列（TTC、TTG、TTA以及TTT）。在轉化之後，將細胞稀釋1：2000並且接種在含有氨苄西林和氯黴素的LB瓊脂板上。僅具有完整原型間隔區質粒的細胞可以形成菌落。在接種~14 h後對具有菌落的板成像，並且使用ImageJ軟體計數菌落。

[1553] 申請人進行細胞裂解物切割測定，以進一步驗證FnCpf1切割。用於細胞裂解物切割測定的方案如下：

[1554] 體外切割反應。切割緩衝液：100 mM HEPES pH 7.5、

500 mM KCl、25 mM MgCl2、5 mM DTT、25%甘油。可以製備沒有DTT的儲備液。

[1555] 製備細胞裂解物

[1556] 裂解緩衝液：20 mM Hepes pH 7.5、100 mM氯化鉀[KCl]、5 mM氯化鎂[$MgCl_2$]、1 mM二硫蘇糖醇[DTT]、5%甘油、0.1% Triton X-100，補充有10x羅氏(Roche)蛋白酶抑制劑混合物。可以維持裂解緩衝液w/o羅氏蛋白酶抑制劑和DTT的濃儲備液。保持在-20°C下。

[1557] 用推薦量的具有Lipofectamine 2000的DNA轉染HEK細胞。

- 500 ng/24孔

- 2000 ng/6孔

[1558] 在轉染後24-72小時收穫具有裂解緩衝液的細胞

-吸掉培養基

-用DPBS輕輕沖洗

-吸掉DPBS

-使用50 ul裂解緩衝液/24孔或250 ul/6孔

-置於冰上5 min

-轉移到埃彭道夫管中

-冰凍15分鐘

-在高功率下超音波處理，50%工作循環，持續5-10 min

-以最大速率旋轉減慢冷卻，持續20 min

-將上清液轉移到新管中

-在PCR條管中等分試樣，每個條管10 ul並且在-80°C下冷凍

[1559] 指導RNA的體外轉錄

[1560] 套組方案：可以在網址www.neb.com/products/e2030-hiscribe-t7-in-vitro-transcription-kit處獲得資訊

取用100 uM儲備寡核苷酸

在10 ul反應物中退火：

1 ul T7“正向”股=“XRP2649”

1 ul T7“反向”寡核苷酸

1 ul TaqB緩衝液

7 ul水

[1561] 在未進行37°C培養步驟的情況下運行PNK PCR程式（基本上加熱至95°C，持續5 min並且緩慢冷卻至4°C但不像surveyor退火一樣慢）。Nanodrop退火的寡核苷酸：用水標準化至500 ng/ul（通常對於120nt寡核苷酸係1000-2000 ng/ul）

[1562] 對於T7轉錄，遵循套組說明書（但大小減短4x）

[1563] 10 ul反應物

1 ul 10x緩衝液

1 ul T7轉錄酶

0.5 ul rNTP

0.5 ul HMW mix

1 ul DNA模板（退火）

6 ul水

[1564] 在42°C(較佳的是循環變溫器)中轉錄至少2-3小時，運行過夜。產率應係約1000-2000 ng/ul的RNA。形成白色殘餘物係正常的。

[1565] DNA的製備

對於pUC19，用HindIII線性化並且進行柱純化

→將需要300-400 ng質粒/反應，因此切割需要的量

對於gDNA，用PCR擴增wt細胞DNA

→進行多個PCR反應，合併並進行柱純化

→將產物濃縮至約100-200 ng/ul

保持在-20°C

[1566] 20 ul反應物

10 ul裂解物（這係預先等分的試樣）

2 ul裂解緩衝液（NEB緩衝液3）

1 ul RNA（直接來自以上產物；不需要純化）

1 ul DNA（直接來自以上產物）

6 ul水

在37°C下培養1-2小時（30 min係足夠的）

[1567] 柱純化該反應物

[1568] 在2% E-gel上進行

[1569] 細胞裂解物切割測定使用如圖54所示的位置1、2、3、4以及5處的tracrRNA。細胞裂解物切割測定（1）（圖55）係指示了在細胞裂解物中培養的具有TTa PAM的PCR片段和原型間隔區1序列的凝膠。細胞裂解物切割測定（2）（圖56）係示出了在細胞裂解物中培養的具有不同的PAM的pUC-間隔區1的凝膠。細胞裂解物切割測定（3）（圖57）係示出了在細胞裂解物中培養之後的BasI消化的凝膠。細胞裂解物切割測定（4）（圖58）係示出了三種推定的crRNA序列的消化結果的凝膠。

[1570] 申請人還確定了間隔區長度對於切割效率的影響。申請人測試了對針對含有靶位點：5'-TTAgagaagtcatttaataaggccactgttaaaa-3'的靶DNA片段（SEQ ID NO: 119）的不同長度的間隔區的凝膠。對於此實驗，將含有間隔區（5'-TTcgagaagucauuuaauaaggccacuguuaaaa-3'（SEQ ID NO: 120））的pUC19質粒處理成以下狀態：

2 ul 含有Cpf1的細胞裂解物

2 ul 具有間隔區的pUC19 DNA（300 ng）

1 ul crRNA（500 ng）

2 ul NEBuffer 3

2 ul 40 mM DTT

0.3 ul BsaI

10.7 ul ddH2O

[1571] 在37°C下培養30分鐘，接著用RNA酶處理5分鐘。然後使用Qiagen PCR純化套組純化反應物並且在2% Invitrogen E-gel EX上分析。圖59係顯示在體外crRNA 1-7介導了使用FnCpf1對靶DNA的成功切割而crRNA 8-13並不利於靶DNA的切割的凝膠。

[1572] 申請人得到了最少的Fn Cpf1座位（圖60）並且還闡明了最少的Cpf1指導（圖61）。申請人還切割了人類Emx1座位的PCR擴增子（圖81）。將EMX擴增子處理成以下狀態：

2 ul 含有Cpf1的細胞裂解物

3 ul 具有間隔區的pUC19 DNA（300 ng）

1 ul crRNA（500 ng）

2 ul NEBuffer 3

2 ul 40 mM DTT

0.3 ul BsaI

9.7 ul ddH_2O

[1573] 在37°C下培養30分鐘，接著用RNA酶處理5分鐘。然

後使用Qiagen PCR純化套組純化反應物並且在2% Invitrogen E-gel EX上分析。

[1574] 申請人進一步研究了5' DR的截短對切割活性的影響（圖82A-圖82B）。對於此實驗，將含有間隔區（5'-TTcgagaagucauuuaauaaggccacuguuaaaa-3'（SEQ ID NO: 121））的pUC19質粒處理成以下狀態：

2 ul 含有Cpf1的細胞裂解物

2 ul 具有間隔區的pUC19 DNA（300 ng）

1 ul crRNA（500 ng）

2 ul NEBuffer 3

2 ul 40 mM DTT

0.3 ul BsaI

10.7 ul ddH2O

[1575] 在37°C下培養30分鐘，接著用RNA酶處理5分鐘。然後使用Qiagen PCR純化套組純化反應物並且在2% Invitrogen E-gel EX上分析。申請人確定crDNA δDR5破壞了5'端的莖環並且這顯示5'端的莖環係對於切割活性所必需的（圖82B）。

[1576] 申請人研究了crRNA-DNA靶錯配對切割效率的影響（圖83）。對於此實驗，將含有間隔區（5'-TTcgagaagucauuuaauaaggccacuguuaaaa-3'（SEQ ID NO: 122））的pUC19質粒處理成以下狀態：

2 ul 含有Cpf1的細胞裂解物

2 ul 具有間隔區的pUC19 DNA（300 ng）

1 ul crRNA（500 ng）

2 ul NEBuffer 3

2 ul 40 mM DTT

0.3 ul BsaI

10.7 ul ddH2O

[1577] 在37°C下培養30分鐘，接著用RNA酶處理5分鐘。然後使用Qiagen PCR純化套組純化反應物並且在2% Invitrogen E-gel EX上分析。圖83所示的凝膠中的各泳道由含有Cpf1的細胞裂解物、具有TTc原型間隔區的pUC19、以及相應crRNA組成，如1-11所指示的。

[1578] 申請人研究了FnCpf1p RuvC結構域並且已鑒定可以將FnCpf1效應蛋白轉化成切口酶的胺基酸突變，由此該效應蛋白具有基本上減小的核酸酶活性並且僅一條DNA股被切口和/或切割。FnCpf1p RuvC結構域中的胺基酸位置包括但不限於D917A、E1006A、E1028A、D1227A、D1255A、N1257A、D917A、E1006A、E1028A、D1227A、D1255A和N1257A。AsCpf1中的胺基酸**位置**對應於AsD908A、AsE993A、AsD1263A。LbCpf1中的胺基酸**位置**對應於LbD832A。

[1579] 申請人還鑒定了與PD-(D/E)XK核酸酶超家族和

HincII內切核酸酶樣最類似的推定的第二核酸酶結構域。在此推定的核酸酶結構域中產生的大幅度降低核酸酶活性的點突變包括但不限於N580A、N584A、T587A、W609A、D610A、K613A、E614A、D616A、K624A、D625A、K627A和Y629A。

[1580] 申請人使用FnCpf1p進行質粒切割實驗並且所述質粒的定序將提供關於切割位點係黏性還是鈍性的資訊。申請人將由適合複合物中的FnCpf1p的晶體結構闡明關於此蛋白質的不同結構域的其他詳情。對於FnCpf1座位件在人類細胞中的活性的優化，申請人將嘗試不同的crRNA體系結構並且嘗試比在此所述更多的靶標。

[1581] 申請人使用純化的弗朗西絲菌和普雷沃菌Cpf1切割DNA（圖84）。對於此實驗，將含有間隔區（5'-TTcgagaagucauuuaauaaggccacuguuaaaa-3'（SEQ ID NO: 123））的pUC19質粒處理成以下狀態：

2 ul 純化的蛋白質溶液

2 ul具有間隔區的pUC19 DNA（300 ng）

1 ulcrRNA（500 ng）

2 ulNEBuffer 3

2 ul40 mM DTT

0.3 ulBsaI

10.7 ulddH2O

[1582] 在37°C下培養30分鐘，接著用RNA酶處理5分鐘。然後使用Qiagen PCR純化套組純化反應物並且在2% Invitrogen E-gel EX上分析。在圖84中示出的凝膠的分析指示PaCpf1可以與FnCpf1 crRNA一起作用，儘管該活性不像FnCpf1一樣高。申請人推斷這係有意義的，考慮到PaCpf1和FnCpf1的莖環序列係幾乎相同的（僅1個鹼基不同）（參見圖85A-圖85B）。這在圖87A-圖87B中所示的FnCpf1和PaCpf1的成熟crRNA序列中進一步突出。在本發明的較佳的實施方式中，生物化學或體外切割可以不需要用於Cpf1p CRISPR系統的有效功能的tracr序列。包含莖環或進一步優化的莖環結構對於切割活性而言是重要的。

[1583] 藉由人類密碼子優化的新殺手弗朗西絲菌FnCpf1p進行DNA切割。

[1584] 申請人還顯示FnCpf1p切割在人類細胞中的DNA。將400 ng人類密碼子優化的FnCpf1p和100 ng U6::crRNA轉染到24孔板中的每個孔中的HEK293T細胞（~240,000個細胞）。採用包含基於5'-ctgatggtccatgtctgttactcg-3'（SEQ ID NO: 124）（即，前20、21、22、23或所有24 nt）的長度為20-24 nt的間隔區序列的五種crRNA。crRNA進一步包含PaCpf1在間隔區的5'處的20 nt 5'重複序列。申請人早期確定來自PaCpf1的重複序列可以被FnCpf1識別。

[1585] 在~60 h之後收穫DNA並且藉由SURVEYOR核酸酶測定進行分析。用於DNMT1的SURVEYOR引物係5'-ctgggactcaggcgggtcac-3'（SEQ ID NO: 125）（正向）和5'-cctcacacaacagcttcatgtcagc-3'（SEQ ID NO: 126）（反向）。對於

所有五種crRNA（間隔區長度20-24 nt）觀察到與預期的~345 bp和~261 bp的切割產物符合的切割的DNA片段。（圖88）。

實例5：對於PaCpf1的其他確認實驗

[1586] 對於易北普雷沃菌Cpf1（PaCpf1）進行PAM計算篩選，這與如實例3詳述的對於FnCpf1進行的篩選類似。在對篩選DNA定序之後，提取出對應於左PAM或右PAM的區。對於每個樣品，將定序文庫中存在的PAM數目與文庫中預期的PAM數目（4^7）進行比較。左文庫顯示出非常輕微的PAM缺失。為了量化此缺失，計算了富集比。針對兩種條件（對照pACYC或含有PaCpf1的pACYC），根據以下公式針對文庫中的每種PAM計算該比率：

[1587] $$\text{比率} = -\log_2 \frac{\text{樣品} + 0.01}{\text{初始文庫} + 0.01}$$

[1588] 繪製的分佈顯示在對照樣品中幾乎沒有富集並且在兩種生物複製本中有富集。收集比率超過4.5的所有PAM，並且繪製頻率分佈圖，揭示了5' TTTV PAM，其中V係A或C或G（圖62A-圖62E）。

[1589] 申請人將由適合複合物中的PaCpf1p的晶體結構闡明關於此蛋白質的不同結構域的其他詳情。對於PaCpf1座位件在人類細胞中的活性的優化，申請人將以不同的crRNA（指導RNA）體系結構和不同優化的PaCpf1效應蛋白為工作物件。申請人已如下地人類密碼子優化PaCpf1序列：

[1590] NLS（底線）

[1591] GS接頭（**粗體**）

[1592] 3× HA標籤（*斜體*）

[1593]

ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGA
GTCCCAGCAGCC**ggtagt**AACATCAAAAACTTTACCGGGCTCTA
CCCCCTCAGCAAAACTTTGCGCTTTGAACTCAAGCCTATTGGC
AAAACCAAGGAAAACATCGAGAAAAATGGCATCCTGACCAAG
GACGAGCAACGGGCTAAAGACTACCTCATAGTCAAAGGCTTT
ATTGACGAGTATCACAAGCAGTTCATCAAAGACAGGCTTTGGG
ACTTTAAATTGCCTCTCGAAAGTGAGGGGGAGAAGAACAGTC
TCGAAGAATACCAGGAACTGTACGAGCTCACTAAGCGCAACG
ATGCCCAGGAGGCCGACTTCACCGAGATTAAAGATAACCTTCG
CAGCTCTATTACCGAACAGCTCACGAAGTCTGGATCTGCGTAC
GATCGGATTTTTAAAAAAAGAGTTCATTAGAGAAGACCTGGTCA
ACTTCCTCGAAGATGAAAAAGATAAAAATATCGTGAAACAGTT
CGAGGACTTTACTACATATTTTACGGGTTTTTATGAAAATAGGA
AGAACATGTACTCTAGCGAAGAGAAGTCCACGGCCATCGCATA
CCGGCTTATCCATCAGAATCTGCCAAAATTCATGGACAACATG
AGAAGTTTTGCCAAAATTGCAAATTCCAGTGTTTCCGAGCACT
TTAGCGACATCTATGAAAGCTGGAAGGAATATCTGAATGTAAA
TAGCATCGAGGAAATCTTCCAGCTCGACTATTTTAGCGAAACC
TTGACTCAGCCACATATTGAGGTGTATAACTATATTATCGGGAA
GAAAGTCCTGGAAGACGGAACCGAGATAAAGGGCATCAACGA
GTATGTGAACCTCTACAATCAGCAGCAGAAAGATAAGAGTAA
ACGACTGCCTTTCCTGGTGCCACTGTATAAGCAAATTTTGTCT

GATAGGGAAAAACTCTCCTGGATTGCTGAAGAGTTCGACAGC

GACAAGAAGATGCTGAGCGCTATCACCGAGTCTTACAACCAC

CTGCACAACGTGTTGATGGGTAACGAGAACGAAAGCCTGCGA

AATCTGCTGCTGAATATTAAGGACTATAACCTGGAGAAAATTA

ATATCACAAACGACTTGTCTCTCACCGAAATCTCCCAGAATCT

TTTTGGCCGATATGATGTATTCACAAATGGGATCAAAAACAAG

CTGAGAGTGTTGACTCCAAGGAAGAAAAAGGAGACGGACGA

AAATTTTGAGGACCGCATTAACAAAATTTTTAAGACCCAGAAG

TCCTTCAGCATCGCTTTTCTGAACAAGCTGCCTCAGCCCGAAA

TGGAGGATGGGAAGCCCCGGAACATTGAGGACTATTTCATTAC

ACAGGGGGCGATTAACACCAAATCTATACAGAAAGAAGATATC

TTCGCCCAAATTGAGAATGCATACGAGGATGCACAGGTGTTCC

TGCAAATTAAGGACACCGACAACAAACTTAGCCAGAACAAGA

CGGCGGTGGAAAAGATCAAAACTTTGCTGGACGCCTTGAAGG

AACTCCAGCACTTCATCAAACCGCTGCTGGGCTCTGGGGAGG

AGAACGAGAAAGACGAACTGTTCTACGGTTCCTTCCTGGCCA

TCTGGGACGAACTGGACACCATTACACCACTTTATAACAAAGT

GAGAAATTGGCTGACCCGAAAACCATATTCAACAGAAAAAAT

CAAATTGAATTTCGACAACGCTCAGCTGCTGGGAGGGTGGGA

TGTCAATAAAGAACACGACTGTGCAGGTATCTTGTTGCGGAAA

AACGATAGCTACTATCTCGGAATTATCAATAAGAAAACCAACC

ACATCTTTGATACGGATATTACGCCATCAGATGGCGAGTGCTAT

GACAAAATCGACTACAAGCTCCTTCCCGGGGCGAACAAAATG

CTTCCAAAGGTGTTTTTTAGTAAGTCCCGAATCAAAGAGTTCG

AGCCATCAGAGGCCATAATCAATTGCTATAAGAAGGGGACACA

CAAAAAAGGAAAAAACTTTAACCTGACGGACTGTCACCGCCT

GATCAACTTTTTTAAGACCTCAATCGAGAAACACGAGGATTGG

TCAAAATTCGGATTCAAGTTCTCCGATACCGAAACGTATGAGG

ATATTAGCGGTTTTTATAGAGAGGTCGAGCAGCAGGGATACAG

GCTGACGAGCCATCCAGTCAGTGCCAGCTATATACATAGTCTG

GTCAAGGAAGGAAAACTGTACCTCTTCCAAATCTGGAACAAG

GACTTTTCTCAATTCTCCAAGGGGACCCCTAACTTGCACACTC

TCTATTGGAAGATGCTGTTTGACAAACGGAATCTTAGCGATGT

GGTTTATAAGCTGAATGGCCAGGCTGAAGTGTTCTATAGAAAG

AGCTCCATTGAACACCAGAACCGAATTATCCACCCCGCTCAGC

ATCCCATCACAAATAAGAATGAGCTTAACAAAAAGCACACTAG

CACCTTCAAATACGATATCATCAAAGATCGCAGATACACGGTG

GATAAATTCCAGTTCCATGTGCCCATTACTATAAATTTTAAGGC

GACCGGGCAGAACAACATCAACCCAATCGTCCAAGAGGTGAT

TCGCCAAAACGGTATCACCCACATCATAGGCATCGATCGAGGT

GAACGCCATCTTCTGTACCTCTCTCTCATCGATTTGAAAGGCA

ACATCATCAAGCAGATGACTCTCAACGAAATTATTAATGAGTAT

AAGGGTGTGACCTATAAGACCAACTACCATAACCTCCTGGAGA

AGAGGGAGAAGGAGCGGACCGAGGCCAGACACTCCTGGAGT

AGTATTGAAAGCATAAAAGAACTGAAGGATGGATACATGTCAC

AGGTGATTCACAAAATTACGGACATGATGGTTAAGTACAATGC

GATTGTGGTCCTGGAGGACCTCAACGGGGGGTTTATGCGAGG

CCGCCAGAAGGTCGAGAAGCAGGTGTACCAGAAATTTGAAAA

AAAGTTGATCGACAAGCTGAACTATCTCGTTGACAAGAAACT
CGACGCTAACGAGGTCGGCGGAGTACTGAATGCTTATCAGCTG
ACCAACAAGTTCGAGTCTTTCAAGAAGATTGGGAAACAAAGC
GGATTTTTGTTCTACATCCCCGCCTGGAACACAAGCAAAATCG
ATCCTATAACAGGGTTCGTTAATCTGTTCAACACCAGGTACGA
GTCTATCAAGGAGACAAAAGTTTTTTGGTCTAAGTTTGATATTA
TCCGATACAATAAAGAGAAGAATTGGTTCGAGTTCGTCTTCGA
TTACAATACCTTTACGACTAAAGCGGAGGGAACACGCACTAAG
TGGACTCTGTGCACCCACGGCACTCGCATCCAGACATTCCGGA
ACCCAGAAAAGAATGCCCAGTGGGACAATAAAGAGATCAATT
TGACTGAGTCCTTCAAAGCTCTGTTTGAAAAGTACAAGATCGA
TATCACCAGTAATCTCAAGGAATCCATCATGCAGGAAACCGAG
AAGAAGTTCTTCCAGGAACTGCATAATCTGCTCCACCTGACCC
TGCAGATGAGGAATAGCGTTACTGGAACCGACATAGACTATTT
GATCAGCCCCGTTGCCGATGAGGATGGAAATTTCTATGATAGT
CGCATAAATGGCAAAAATTTTCCGGAGAATGCCGATGCCAATG
GCGCGTACAACATCGCACGAAAGGGTCTGATGCTTATTCGGCA
GATCAAGCAAGCAGATCCACAGAAGAAATTCAAGTTTGAGAC
AATCACCAATAAAGACTGGCTGAAATTCGCCCAAGACAAGCC
CTATCTTAAAGAT**ggcagcggg**AAAAGGCCGGCGGCCACGAAAA
AGGCCGGCCAGGCAAAAAAGAAAAAG**ggatcc***TACCCATACGAT*
GTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACC
CATACGATGTCCCCGACTATGCCTAA（SEQ ID NO: 127）

[1594] 用於人類密碼子優化的PaCpf1序列的載體圖譜提供

在圖63中。

實例6：Cpf1異種同源物

[1595] 申請人分析Cpf1異種同源物的擴增庫（圖64）。對於幾種Cpf1座位件獲得人類密碼子優化序列（圖65-圖79）。申請人也得到了每種異種同源物的同向重複（DR）序列及其預測的折疊結構（圖80A-圖80I）。

[1596] 申請人基於效應蛋白的大小來進一步研究Cpf1異種同源物，即較小效應蛋白允許更容易包裝到載體中並且包裝在PAM組成上。所有方面允許在原核細胞和真核細胞中進一步優化，較佳的是對於哺乳動物細胞（即人類細胞）中的有效活性。

[1597] 申請人證實以下座位的效應蛋白異種同源物在體外切割測定中顯示活性：佩萊格裡尼菌科細菌GW2011_GWA2_33_10 Cpf1、胺基酸球菌屬某種BV3L6 Cpf1、土拉熱弗朗西絲菌1 Cpf1、牛莫拉氏菌237 Cpf1、毛螺菌科細菌ND2006 Cpf1、毛螺菌科細菌MA2020 Cpf1、獼猴卟啉單胞菌Cpf1、狗口腔卟啉單胞菌3 Cpf1、易北普雷沃菌Cpf1（圖64）。

[1598] 在藉由異種同源物進行的體外切割測定中，收穫了表現Cpf1異種同源物的HEK293細胞並且用靶向選殖到pUC19質粒的人工間隔區的預測的成熟crRNA培養裂解物。間隔區在8個簡並性鹼基之前，以允許經由定序確定PAM。較低的帶表示由Cpf1酶進行的切割（圖89）。

[1599] 申請人由體外切割測定確定計算得出的PAM（圖90）。

切斷來自圖89的未切割DNA（較高的帶）並且擴增以進行下一代定序。計算每種8聚體的豐度並且使用與輸入文庫相比的對數比率，以量化富集。彙編具有大於4的對數比率的單個8聚體並且將其用於使用Weblogo測定共有PAM。

[1600] 申請人進一步確定Cpf1p效應蛋白以交錯切割的方式切割，產生了5'突出端。收穫純化的FnCpf1蛋白並且用crRNA培養並且將相應靶標選殖到pUC19中。凝膠提取切割的產物並且提交以進行桑格定序。不對稱的讀取顯示存在交錯切割（圖91）。在本發明的一較佳的實施方式中，申請人證實與模板（例如，外源性模板）的體內交錯連接。

[1601] 申請人還確定了間隔區長度對於效應蛋白切割能力的影響（圖92）。收穫純化的FnCpf1蛋白並且用crRNA培養並且將相應靶標選殖到pUC19中。大於17 nt的間隔區長度切割至完成，而17 nt間隔區顯示減少的活性並且小於17 nt的間隔區係無活性的。

[1602] 申請人證實FnCpf1了HEK293T細胞中的indel形成。

[1603] 用350 ng huFnCpf1質粒和150 ng U6::crRNA轉染~280,000 HEK細胞/24孔。在轉染後三天收穫細胞並且藉由SURVEYOR核酸酶測定進行分析。未切割的PCR片段大小係606 bp。對於crRNA DNMT1-1，預期的片段大小係~418 bp和~188 bp並且對於crRNA DNMT1-3係~362 bp和~244 bp（圖93）。

[1604] DNMT1-1間隔區序列：cctcactcctgctcggtgaattt（SEQ ID NO: 128）

[1605] DNMT1-3間隔區序列：ctgatggtccatgtctgttactc（SEQ ID NO: 129）

[1606] 申請人藉由確定當座位的某些序列缺失時轉錄物是否被加工來鑒定Cpf1系統實現切割所需要的元件(圖94A-圖94F)。缺失的序列可以包括但不限於，Cas1基因、Cas2基因和tracr。因此，在本發明的一較佳的實施方式中，申請人證實該tracr不是功能性Cpf1系統或複合物實現切割所需要的元件。

實例7：異源

質粒的程序性生成

[1607] 為了生成用於異源表現的FnCpf1座位，使用Herculase II聚合酶（安捷倫科技公司）PCR擴增來自新殺手弗朗西絲菌的基因組DNA並且使用吉布森選殖(新英格蘭生物實驗室）將其選殖到pACYC-184中。將具有質粒的細胞用Z-感受態套組（Zymo公司）製成感受態的。

細菌RNA定序

[1608] 藉由首先將新殺手弗朗西絲菌（由大衛·魏斯（David Weiss）惠贈）或大腸桿菌重新懸浮在TRIzol中並且然後在BeadBeater（BioSpec Products公司）中用氧化鋯/二氧化矽珠粒（BioSpec Products公司）均質化該細菌持續3個一分鐘循環，將RNA從固定相細菌中分離。藉由Direct-Zol RNA小量製備方案（Zymo公司）從均質化樣品中純化總的RNA，用TURBO DNA酶（生命技術公司）對其進行DNA酶處理，並且用T4多核苷酸激酶

（新英格蘭生物實驗室）進行3'脫磷酸化。用細菌Ribo-Zero rRNA去除套組（億明達公司）去除rRNA。使用億明達的NEBNext® Small RNA Library Prep Set（新英格蘭生物實驗室）由rRNA-缺失的RNA製備RNA文庫並且使用Pippin Prep（聖徒科學公司（Sage Science））進行大小選擇。

[1609] 對於FnCpf1座位的異源大腸桿菌表現，使用先前描述的CRISPR RNA定序方法的衍生物由rRNA-缺失RNA製備RNA定序的文庫（海德里希（Heidrich）等人，2015。簡言之，轉錄物以大腸桿菌Poly(A)聚合酶（新英格蘭生物實驗室）進行poly-A加尾，使用T4 RNA連接酶1（ssRNA連接酶）高濃縮物（High Concentration）（新英格蘭生物實驗室）將其與5' RNA銜接子連接，並且使用AffinityScript多溫度逆轉錄酶（安捷倫技術公司）進行逆轉錄。使用條形編碼的引物使用Herculase II聚合酶（安捷倫技術公司）PCR擴增cDNA。*RNA-定序分析*

[1610] 在MiSeq（億明達公司）上對製備的cDNA文庫進行定序。基於相關條形編碼鑒定每種樣品的讀取，並且使用BWA將其與適當RefSeq參考基因組進行比對（李和德賓（Durbin），2009）。使用成對末端比對，使用皮卡（Picard）工具（http://broadinstitute.github.io/picard）提取整個轉錄物序列，並且使用Geneious 8.1.5.分析哲學序列。

體內FnCpf1 PAM篩選

[1611] 使用合成的寡核苷酸（IDT）構建隨機化PAM質粒文庫，該寡核苷酸由間隔區1靶標上游或下游的7個隨機化核苷酸組

成（輔助表S8）。藉由退火到短引物並且使用大的克列諾片段（新英格蘭生物實驗室）用於第二條股合成來製成雙股隨機化ssDNA寡核苷酸。使用吉布森選殖將dsDNA產物組裝到線性pUC19中（新英格蘭生物實驗室）。用選殖的產物轉化感受態的Stbl3大腸桿菌（英傑公司），並且收集併合並超過10^7個細胞。使用Maxi-prep套組（凱傑公司）收穫質粒DNA。我們將360 ng合併的文庫轉化到攜帶FnCpf1座位或pACYC184對照的大腸桿菌中。在轉化之後，將細胞接種在氨苄西林上。在生長16小時後，收穫>$4*10^6$個細胞並且使用Maxi-prep套組（凱傑公司）提取質粒DNA。擴增靶PAM區域並且以單端150個循環使用MiSeq（億明達公司）進行定序。

計算的PAM發現流程

[1612] 將PAM區域提取、計數並且標準化成每種樣品的總讀取。對於給定的PAM，將富集測量為與pACYC184對照相比的對數比率，其中具有0.01假計數調整。收集超過3.5富集閾值的PAM並且將其用於生成序列圖示（克魯克斯（Crooks）等人，2004）。

PAM確認

[1613] 將與PAM、非-PAM二者相對應的序列選殖到消化的pUC19並且將其與T4連接酶（Enzymatics公司）連接。用20 ng PAM質粒轉化具有FnCpf1座位質粒或pACYC184對照質粒的感受態大腸桿菌，並且將其接種在補充有氨苄西林和氯黴素的LB瓊脂板上。在18小時計數菌落。

crRNA和gRNA的合成

[1614] 使用HiScribe™ T7高產率RNA合成套組（NEB）合成體外使用的所有crRNA和gRNA。由IDT合成與靶RNA序列的反向互補序列相對應的ssDNA寡核苷酸並且退火到短T7引物序列上。進行4小時的T7轉錄並且然後使用MEGAclear™轉錄純化套組（Ambion公司）純化RNA。

Cpf1蛋白質的純化

[1615] 將FnCpf1蛋白質選殖到細菌表現載體（6-His-MBP-TEV-Cpf1，它係一種由道格·丹尼爾斯（Doug Daniels）惠贈給申請人的基於pET的載體）（"6-His"被揭露為SEQ ID NO: 130）中。用10 mL含有Cpf1表現構建體的過夜培養的Rosetta(DE3) pLyseS（EMD密理博公司）接種兩升具有100 μg/mL氨苄西林的Terrific Broth生長培養基。生長培養基加上接種體在37°C下生長，直到細胞密度達到0.2 OD600為止，然後將溫度降低到21°C。繼續生長直到OD600達到0.6為止，之後添加終濃度500 μM IPTG以誘導MBP-Cpf1表現。將培養物誘導14-18小時，之後收穫細胞並在-80°C下冷凍，直到進行純化。

[1616] 將細胞糊狀物重新懸浮在200 mL補充有蛋白酶抑制劑(羅氏cOmplete，無EDTA)和溶菌酶的裂解緩衝液(50 mM Hepes pH 7、2M NaCl、5 mM $MgCl_2$、20 mM咪唑)。一旦均質化，就藉由超音波處理（必能信（Branson）超音波儀450）來裂解細胞，然後在10,000g下離心1小時來清出裂解物。藉由0.22微米過濾器（密理博公司，Stericup）過濾裂解物，並且將其應用於鎳柱（HisTrap FF，5 mL），用梯度的咪唑洗滌並且然後洗提。將含有預期大小的

蛋白質的級分合併，添加TEV蛋白酶（西格瑪公司），並且將樣品在TEV緩衝液（500 mM NaCl、50 mM Hepes pH 7、5 mM MgCl、2 mM DTT）中透析過夜。在透析之後，藉由SDS-PAGE證實TEV切割，並且將樣品濃縮至500 μL，之後經由FPLC（AKTA Pure）裝載到凝膠過濾柱（HiLoad 16/600 Superdex 200）上。藉由SDS-PAGE分析來自凝膠過濾的級分；將含有Cpf1的級分合併並濃縮至200 μL，並且將其直接用於生物化學測定或在-80°C下冷凍儲存。在以2M NaCl、Hepes pH 7.0平衡的相同柱上運行凝膠過濾標準，以計算FnCpf1的近似大小。

Cpf1蛋白裂解物的生成

[1617] 用N-末端核定位標籤合成密碼子優化用於人類表現的Cpf1蛋白並且將其藉由Genscript選殖到pcDNA3.1表現質粒中。使用Lipofectamine 2000試劑（生命技術公司）將2000 ng Cpf1表現質粒以90%融合度轉染到6孔板的HEK293FT細胞中。48小時後，藉由用DPBS（生命技術公司）洗滌一次並且在裂解緩衝液[20 mM Hepes pH 7.5、100 mM KCl、5mM $MgCl_2$、1 mM DTT、5%甘油、0.1% Triton X-100、1X cOmplete蛋白酶抑制劑混合片劑（羅氏公司）]中擦洗來收穫細胞。將裂解物在Biorupter超音波破碎器（Diagenode公司）中超音波處理10分鐘並且然後離心。將上清液冷凍以隨後用於體外切割測定。

體外切割測定

[1618] 使用純化的蛋白質或具有蛋白質的哺乳動物裂解物在37°C下在切割緩衝液（NEBuffer 3，5 mM DTT）中進行體外切

割，持續20分鐘。切割反應使用500 ng合成的crRNA或sgRNA以及200 ng靶DNA。靶DNA涉及選殖到pUC19的原型間隔區或來自從HEK293細胞分離的基因組DNA的基因區域的PCR擴增子。使用PCR純化柱（凱傑公司）純化反應物並且在2%瓊脂E-gel（生命技術公司）上運行。對於分析核酸酶突變體進行的切割的天然和變性凝膠，在TBE 6%聚丙烯醯胺或TBE-Urea 6%聚丙烯醯胺凝膠（生命技術公司）上運行純化反應。

體外Cpf1-家族蛋白PAM篩選

[1619] 在2%瓊脂E-gel（生命技術公司）上運行使用Cpf1-家族蛋白的體外切割反應。QIAquick凝膠提取套組（凱傑公司）來凝膠提取與未切割靶標相對應的帶，並且以單端150個循環使用MiSeq（億明達公司）對靶PAM區域進行擴增和定序。將定序結果輸入到PAM發現流程中。

Cpf1切割在293FT細胞中的活性

[1620] 用N-末端核定位標籤合成密碼子優化用於人類表現的Cpf1蛋白並且將其藉由Genscript選殖到pcDNA3.1 CMV表現質粒中。使用Herculase II（安捷倫技術公司）生成包含驅動crRNA序列表現的U6啟動子的PCR擴增子。使用Lipofectamine 2000試劑（生命技術公司）將400ng Cpf1表現質粒和100 ng crRNA PCR產物以75%-90%融合度轉染到24孔板的HEK293FT細胞中。使用QuickExtract™ DNA提取溶液（Epicentre公司）收穫基因組DNA。

用於基因組修飾的SURVEYOR核酸酶測定

[1621] 使用Lipofectamine 2000試劑（生命技術公司），用400 ng Cpf1表現質粒和100 ng U6::crRNA PCR片段轉染293FT細胞。轉染後72 h在37°C下培養細胞，之後進行基因組DNA提取。根據製造商方案使用QuickExtract DNA提取溶液（Epicentre公司）提取基因組DNA。對側接每個基因的CRISPR靶位點的基因組區域進行PCR擴增，並且根據製造商方案使用QiaQuick自旋柱（凱傑公司）純化產物。將總計200-500 ng純化的PCR產物與1 μl 10× 耐熱性DNA聚合酶PCR緩衝液（Enzymatics公司）和超純水混合至10 μl的最終體積，並且經受重退火過程，以使得能夠形成異源雙股體：95°C持續10 min，以−2°C/s從95°C降溫至85°C，以−0.25°C/s從85°C降溫至25°C，並且25°C維持1 min。在重退火之後，根據製造商推薦的方案，用SURVEYOR核酸酶和SURVEYOR增強子S（綜合DNA技術公司（Integrated DNA Technologies））處理產物，並且在4%-20% Novex TBE聚丙烯醯胺凝膠（生命技術公司）上進行分析。將凝膠用SYBR Gold DNA著色劑（生命技術公司）染色10分鐘並且用Gel Doc凝膠成像系統（伯樂生命醫學產品公司（Bio-rad））進行成像。基於相對帶強度進行量化。藉由等式100 × (1 − (1 − (b + c)/(a + b + c))1/2)確定Indel百分比，其中a係未消化的PCR產物的積分強度，並且b和c係每種切割產物的積分強度。

表徵293FT細胞中的Cpf1 indel模式的深度定序

[1622] 如所述地轉染並收穫HEK293FT細胞，以用於評定Cpf1切割的活性。使用兩輪PCR區域來擴增側接DNMT1靶標的基因組區域，以將Illumina P5銜接子以及獨特樣品特異性條形編碼

添加到靶擴增子中。在2% E-gel（英傑公司）運行PCR產物並且按照製造商推薦的方案使用QiaQuick自旋柱（凱傑公司）進行凝膠提取。將樣品合併並且藉由Qubit 2.0螢光計（生命技術公司）進行量化。在MiSeq（億明達公司）上對製備的cDNA文庫進行定序。使用Geneious 6.0.3讀取測繪儀的Python實現方式繪製Indel。

Cpf1座位的計算分析

[1623] 使用PSI-BLAST程式（安特斯庫爾，1997），使用幾種已知的Cpf1序列作為Cpf1的查詢來鑒定NCBI NR資料庫中的Cpf1同源物，其中具有0.01的截止E-值和低複雜度過濾以及基於組成的統計學關閉。使用具有0.01的截止E值和低複雜度過濾關閉參數的TBLASTN程序，使用Cpf1譜圖（瑪拉柯瓦（Marakova）等人，2015）作為查詢，搜索NCBI WGS資料庫。將所有搜索的結果組合。使用具有預設參數的HHpred程式，使用代表性Cpf1查詢的子集，鑒定遠距離序列相似性（塞汀等人，2006）。使用MUSCLE（愛德格，2004）構建多序列比對，其中基於使用PSI-BLAST和HHpred程序獲得的成對比對進行手動校正。使用具有WAG進化模式和離散性γ模式（具有20比率類別）的FastTree程式進行系統發育性分析（普裡斯等人，2010）。使用Jpred 4預測蛋白質二級結構（德羅茲德斯基，2015）。

[1624] 使用PILER-CR（愛德格，2007）和CRISPRfinder（格麗莎（Grissa）等人，2007）鑒定CRISPR重複序列。使用具有預設參數（除字體大小設為20並且E截止值為0.0001之外）的MEGABLAST（莫耳古利斯（Morgulis）等人，2008）搜索NCBI核

苷酸NR資料庫中的間隔區序列。

表1.內源性新殺手弗朗西絲菌間隔區序列	
間隔區編號	序列
1	GAGAAGTCATTTAATAAGGCCACTGTTAAAA（SEQ ID NO: 131）
2	GCTACTATTCCTGTGCCTTCAGATAATTCA（SEQ ID NO: 132）
3	GTCTAGAGCCTTTTGTATTAGTAGCCG（SEQ ID NO: 133）

表2. 用於生成PAM文庫的ssDNA寡核苷酸和引物	
寡核苷酸/引物名稱	序列
PAM文庫5’（+）	GGCCAGTGAATTCGAGCTCGGTACCCGGGNNNNNNNNGAGAAGTCATTTAATAAGGCCACTGTTAAAAAGCTTGGCGTAATCATGGTCATAGCTGTTT（SEQ ID NO: 134）
PAM 文庫 3’（+）	GGCCAGTGAATTCGAGCTCGGTACCCGGGGAGAAGTCATTTAATAAGGCCACTGTTAAAANNNNNNNNNAGCTTGGCGTAATCATGGTCATAGCTGTTT（SEQ ID NO: 135）
PAM 文庫（–）	GCTGACATGAAGCTGTTGTGTGAGG（SEQ ID NO: 136）

表3，用於pUC19定序和SURVEYOR測定的引物	
引物名稱	序列
NGS pUC For	GGCCAGTGAATTCGAGCTCGG（SEQ ID NO: 137）
NGS pUC Rev	CAATTTCACACAGGAAACAGCTATGACC（SEQ ID NO: 138）
桑格pUC For	CGGGGCTGGCTTAACTATGCG（SEQ ID NO: 139）
桑格pUC Rev	GCCCAATACGCAAACCGCCT（SEQ ID NO: 140）
EMX1 For	CCATCCCCTTCTGTGAATGT（SEQ ID NO: 141）
EMX1 Rev	TCTCCGTGTCTCCAATCTCC（SEQ ID NO: 142）
DNMT1 For	CTGGGACTCAGGCGGGTCAC（SEQ ID NO: 143）
DNMT1 Rev	GCTGACATGAAGCTGTTGTGTGAGG（SEQ ID NO: 144）

表4.用於體外切割測定的截短指導序列	
截短指導序列編號	序列
1	GAGAAGTCATTTAATAAGGCCACT（SEQ ID NO: 145）
2	GAGAAGTCATTTAATAAGGCCA（SEQ ID NO: 146）
3	GAGAAGTCATTTAATAAGGC（SEQ ID NO: 147）
4	GAGAAGTCATTTAATAAG（SEQ ID NO: 148）
5	GAGAAGTCATTTAATAA（SEQ ID NO: 149）
6	GAGAAGTCATTTAATA（SEQ ID NO: 150）

表5.用於體外切割測定的錯配指導序列	
錯配指導序列編號	序列
1	GATAAGTCATTTAATAAGGCCACT（SEQ ID NO: 151）
2	GAGAAGGCATTTAATAAGGCCACT（SEQ ID NO: 152）
3	GAGAAGTCATGTAATAAGGCCACT（SEQ ID NO: 153）
4	GAGAAGTCATTTAAGAAGGCCACT（SEQ ID NO: 154）
5	GAGAAGTCATTTAATAAGTCCACT（SEQ ID NO: 155）
6	GAGAAGTCATTTAATAAGGCCAAT（SEQ ID NO: 156）

表6.用於體外切割測定的截短同向重複指導序列	
同向重複序列長度	序列
+18	ATTTCTACTGTTGTAGATGAGAAGTCATTTAATAAGGCCACT（SEQ ID NO: 157）
+17	TTTCTACTGTTGTAGATGAGAAGTCATTTAATAAGGCCACT（SEQ ID NO: 158）
+16	TTCTACTGTTGTAGATGAGAAGTCATTTAATAAGGCCACT（SEQ ID NO: 159）
+15	TCTACTGTTGTAGATGAGAAGTCATTTAATAAGGCCACT（SEQ ID NO: 160）
+11	CTGTTGTAGATGAGAAGTCATTTAATAAGGCCACT

	（SEQ ID NO: 161）
+7	TGTAGATGAGAAGTCATTTAATAAGGCCACT（SEQ ID NO: 162）

表7.用於體外切割測定的同向重複莖突變	
同向重複莖突變體編號	序列
1	AATTTCTGCTGTTGCAGAT（SEQ ID NO: 163）
2	AATTTCCACTGTTGTGGAT（SEQ ID NO: 164）
3	AATTCCTACTGTTGTAGGT(SEQ ID NO: 165）
4	AATTTATACTGTTGTAGAT(SEQ ID NO: 166）
5	AATTTCGACTGTTGTAGAT AATTTCGACTGTTGTAGAT（SEQ ID NO: 167）
6	AATTTCTAGTGTTGTAGAT（SEQ ID NO: 168）

表8.用於體外切割測定的同向重複環突變	
同向重複環突變體編號	序列
1	AATTTCTACTATTGTAGAT（SEQ ID NO: 169）
2	AATTTCTACTGCTGTAGAT（SEQ ID NO: 170）
3	AATTTCTACTTTGTAGAT（SEQ ID NO: 171）
4	AATTTCTACTTGTAGAT（SEQ ID NO: 172）
5	AATTTCTACTTTTGTAGAA（SEQ ID NO: 173）
6	AATTTCTACTTTTGTAGAC（SEQ ID NO: 174）

表9.用於哺乳動物細胞的異種同源物特異性DNMT1靶向指導序列			
核酸酶	名稱	5’同向重複	序列
AsCpf1	DNMT1靶標1	5’同向重複	序列
AsCpf1	DNMT1靶標2	TAATTTCTACTGTTGTAGAT（SEQ ID NO: 175）	CCTCACTCCTGCTCGGTGAATTT（SEQ ID NO:

			176）
AsCpf1	DNMT1靶標3	TAATTTCTACTGTTGTAGAT（SEQ ID NO: 177）	AGGAGTGTTCAGTCTCCGTGAAC（SEQ ID NO: 178）
AsCpf1	DNMT1靶標4	TAATTTCTACTGTTGTAGAT（SEQ ID NO: 179）	CTGATGGTCCATGTCTGTTACTC（SEQ ID NO: 180）
Lb3Cpf1	DNMT1靶標1	TAATTTCTACTGTTGTAGAT（SEQ ID NO: 181）	TTTCCCTTCAGCTAAAATAAAGG（SEQ ID NO: 182）
Lb3Cpf1	DNMT1靶標2	TAATTTCTACTAAGTGTAGAT（SEQ ID NO: 183）	CCTCACTCCTGCTCGGTGAATTT（SEQ ID NO: 184）
Lb3Cpf1	DNMT1靶標3	TAATTTCTACTAAGTGTAGAT（SEQ ID NO: 185）	AGGAGTGTTCAGTCTCCGTGAAC（SEQ ID NO: 186）
Lb3Cpf1	DNMT1靶標4	TAATTTCTACTAAGTGTAGAT（SEQ ID NO: 187）	CTGATGGTCCATGTCTGTTACTC（SEQ ID NO: 188）
SpCas9	DNMT1靶標1	TAATTTCTACTAAGTGTAGAT（SEQ ID NO: 189）	TTTCCCTTCAGCTAAAATAAAGG（SEQ ID NO: 190）
SpCas9	DNMT1靶標2	na	TCACTCCTGCTCGGTGAATT（SEQ ID NO: 191）
SpCas9	DNMT1靶標3	na	AACCCTCTGGGGACCGTTTG（SEQ ID NO: 192）
SpCas9	DNMT1靶標4	na	AGTACGTTAATGTTTCCTGA（SEQ ID NO: 193）

表10.對於靶向原型間隔區1和DNMT1靶標3的crRNA的異種同源物特異性

同向重複序列	
同向重複序列起點	序列
FnCpf1	TAATTTCTACTGTTGTAGAT（SEQ ID NO: 195）
Lb1Cpf1	AGAAATGCATGGTTCTCATGC（SEQ ID NO: 196）

BpCpf1	AAAATTACCTAGTAATTAGGT(SEQ ID NO: 197)
PeCpf1	GGATTTCTACTTTTGTAGAT (SEQ ID NO: 198)
PbCpf1	AAATTTCTACTTTTGTAGAT (SEQ ID NO: 199)
SsCpf1	CGCGCCCACGCGGGGCGCGAC (SEQ ID NO: 200)
AsCpf1	TAATTTCTACTCTTGTAGAT (SEQ ID NO: 201)
Lb2Cpf1	GAATTTCTACTATTGTAGAT (SEQ ID NO: 202)
CMtCpf1	GAATCTCTACTCTTTGTAGAT(SEQ ID NO: 203)
EeCpf1	TAATTTCTACTTTGTAGAT (SEQ ID NO: 204)
MbCpf1	AAATTTCTACTGTTTGTAGAT(SEQ ID NO: 205)
LiCpf1	GAATTTCTACTTTTGTAGAT (SEQ ID NO: 206)
Lb3Cpf1	TAATTTCTACTAAGTGTAGAT(SEQ ID NO: 207)
PcCpf1	TAATTTCTACTATTGTAGAT (SEQ ID NO: 208)
PdCpf1	TAATTTCTACTTCGGTAGAT (SEQ ID NO: 209)
PmCpf1	TAATTTCTACTATTGTAGAT (SEQ ID NO: 210)

實例8：土拉熱弗朗西絲菌新殺手亞種U112 Cpf1（FnCpf1）的選殖

[1625] 申請人將土拉熱弗朗西絲菌新殺手亞種*U112*（圖95A）Cpf1（FnCpf1）座位選殖到低拷貝質粒（pFnCpf1）中，以允許在大腸桿菌中異源性重構。典型地，在目前表徵的CRISPR-Cas系統中，對於DNA干擾存在兩種要求：(i) 靶序列必須與存在於對應CRISPR陣列中的間隔區匹配，並且 (ii) 與該間隔區（在下文中是原型間隔區）互補的靶序列必須側接適當的原型間隔區相鄰模體

（PAM）。考慮到FnCpf1 CRISPR座位的完全未表徵的功能，設計質粒缺失測定以確定Cpf1的活性並且鑒定PAM序列及其相對於原型間隔區（5’或3’）的位置（圖95B）。用隨機化的5’或3’7 bp序列構建攜帶匹配FnCpf1 CRISPR陣列中的第一間隔區的原型間隔區的兩個質粒文庫。將每個質粒文庫轉化到異源表現FnCpf1座位的大腸桿菌或者攜帶空載體的對照大腸桿菌菌株中。使用此測定，藉由鑒定在異源表現FnCpf1座位的細胞中優先缺失的核苷酸模體來確定PAM序列和位置。發現FnCpf1的PAM位於原型間隔區的替換股5’端上游並且具有序列5’-TTN（圖95C-圖95D和圖102）。在I型CRISPR系統中也觀察到PAM的5’位置，但是在II型系統中未觀察到，其中Cas9採用位於原型間隔區3’端的PAM序列（莫西卡等人，2009；加爾諾（Garneau）等人，2010）。除PAM鑒定之外，缺失測定的結果明確指示異源表現的Cpf1座位能夠被質粒DNA有效干擾。

[1626] 為了進一步表徵PAM，藉由用攜帶側接5’-TTN PAM的原型間隔區1的質粒轉化表現cpf1座位的細胞來分析質粒干擾活性。有效靶向所有5’-TTN PAM（圖1E）。此外，還有效靶向5’-CTA而不是5’-TCA（圖95E），這表明中間T對於PAM識別比第一個T更關鍵，並且與PAM發現測定（圖102D）中缺失的序列模體一致，該PAM可能比5’-TTN更鬆弛。

實例9：Cpf1 CRISPR陣列獨立於tracrRNA來加工

[1627] 使用小RNAseq確定由基於*cpf1*的CRISPR座位產生的crRNA的完整性質。藉由對從土拉熱弗朗西絲菌新殺手亞種U112

培養物中提取的小RNA進行定序，發現CRISPR陣列被加工成長度為42-44 nt的短成熟crRNA。每個成熟crRNA開始於同向重複序列的19 nt，接著係間隔區序列的23–25 nt（圖96A）。此crRNA安排與II型CRISPR-Cas系統中的安排相反，在該系統中成熟crRNA開始於間隔區序列的20-24 nt，接著係同向重複序列的~22 nt（德爾特切瓦等人，2011；吉林斯基等人，2013）。出乎意料地是，除crRNA之外，我們沒有在可以對應於tracrRNA的弗朗西絲菌cpf1座位附近觀察到任何強勁表現的小轉錄物，該等轉錄物與基於Cas9的系統相關聯。

[1628] 為了證實crRNA成熟和DNA干擾不需要另外的RNA，使用合成的啟動子構建表現質粒以驅動弗朗西絲菌*cpf1*（FnCpf1）和CRISPR陣列（pFnCpf1_min）的表現。表現此質粒的大腸桿菌的小RNAseq仍顯示將CRISPR陣列強勁加工成成熟crRNA（圖96B），這指示FnCpf1及其CRISPR陣列足以實現crRNA加工。另外，表現pFnCpf1_min以及pFnCpf1_ΔCas的大腸桿菌係所有cas基因都已去除但保留驅動FnCpf1和CRISPR陣列表現的天然啟動子的質粒，它也展示了強勁DNA干擾，這表明FnCpf1和crRNA足以用於介導DNA靶向（圖96C）。相反，Cas9需要crRNA和tracrRNA二者來介導靶向的DNA干擾（德爾特切瓦等人，2011；張等人，2013）。

實例10：Cpf1係單一的crRNA指導性內切核酸酶。

[1629] 考慮到Cas9藉由crRNA與tracrRNA之間的雙股結構，FnCpf1可以與單獨的crRNA介導DNA干擾的發現結果係非常令人驚奇的（金科等人，2012；西松等人，2014），以及tracrRNA的3'

二級結構（徐等人，2013；西松等人，2014）來識別crRNA。為了確保crRNA確實足以與FnCpf1形成活性複合物並且介導RNA-指導的DNA切割，針對靶DNA的體外切割測試僅供應有crRNA的FnCpf1。針對其切割在細菌DNA干擾實驗（圖97A）中使用的含有相同原型間隔區1的質粒的能力測定純化的FnCpf1（圖103）。具有體外轉錄的成熟crRNA靶向原型間隔區1的FnCpf1能夠以Mg^{2+}-和crRNA-依賴性方式有效切割靶質粒（圖97B）。此外，FnCpf1能夠切割超螺旋和線性靶DNA（圖97C）。該等結果明確證實FnCpf1和crRNA足以用於RNA指導的DNA切割。

[1630] 還使用切割的DNA末端的桑格定序來繪製FnCpf1的切割位點。FnCpf1介導的切割形成5-nt 5'突出端（圖97A、圖97D和圖104），該突出端不同於由Cas9生成的鈍切割產物（加爾諾等人，2010；金科等人，2012；加西烏納斯等人，2012）。FnCpf1的交錯切割位點遠離PAM：在非靶向（+）股上的第18個鹼基之後和在靶向（–）股上的第23個鹼基之後發生切割（圖97A、圖97D和圖104）。使用具有不同PAM序列的雙股寡核苷酸底物，我們還發現當5'-TTN PAM處於雙股體形式（圖97E）時FnCpf1切割靶DNA，這與Cas9的PAM相反（斯騰伯格等人，2014）。

實例11：Cpf1的RuvC樣結構域介導了RNA指導的DNA切割

[1631] Cpf1的RuvC樣結構域保留內切核酸酶的此家族的所有催化殘基（圖98A和圖105）並且因此預測係活性核酸酶。生成三種突變體FnCpf1（D917A）、FnCpf1（E1006A）和FnCpf1（D1225A）（圖98A），以測試保守性催化殘基是否是FnCpf1的核酸酶活性所

必需的。D917A和E1006A突變完全滅活FnCpf1的DNA切割活性，並且D1255A顯著降低了溶核活性（圖98B）。該等結果與釀膿鏈球菌Cas9（SpCas9）的誘變結果相反，其中RuvC（D10A）和HNH（N863A）核酸酶結構域的突變將SpCas9轉化成DNA切口酶（即這兩種核酸酶結構域各自的滅活消除了一條DNA股的切割）（金科等人，2012；加西烏納斯等人，2012）（圖98B）。該等發現結果表明FnCpf1的RuvC樣結構域切割可能處於二聚體構型中的靶DNA的兩條股（圖103B）。

實例12：Cpf1 crRNA的序列和結構

[1632] 與具有和Cas9相互作用的精細RNA二級結構特徵的Cas9的指導RNA相比（西松等人，2014），FnCpf1的指導RNA係顯著更簡單的並且僅包含同向重複序列中的單一莖環（圖97A）。

[1633] 探尋了用於與FnCpf1一起介導DNA切割的crRNA的序列和結構要求。檢查指導序列的長度。觀察16 nt指導序列以實現可檢測的DNA切割並且18 nt的指導序列實現有效的體外DNA切割（圖99A）。該等長度類似於對於SpCas9證實的長度，其中16至17 nt間隔區序列足以用於DNA切割（森席克（Cencic）等人，2014；付等人，2014）。FnCpf1指導RNA的種子區在間隔區序列的5'端上的前6或7 nt內觀察到（圖99B）。

[1634] 研究同向重複序列突變對於RNA指導的DNA切割活性的影響。成熟crRNA的同向重複序列部分係19 nt長（圖96A）。同向重複序列的截短顯示16 nt係足夠的，但最佳地超過17 nt的同向重複序列有效用於切割。保存RNA雙股體的莖環中的突變並不

影響切割活性，而破壞莖環雙股體結構的突變消除了切割（圖99D）。最終，環區域中的鹼基取代並不影響核酸酶活性，而緊接著在間隔區序列的5'的U取代基本上降低了活性（圖5E）。總之，該等結果表明FnCpf1藉由莖環的序列特異性和結構特徵的組合來識別crRNA。

實例13：來自不同細菌的Cpf1家族蛋白共用了一般crRNA結構和PAM

[1635] 為了研究Cpf1作為基因組編輯工具的用途，探索了在公共序列資料庫中可獲得的Cpf1家族蛋白的多樣性。在NCBI除的WGS資料庫的BLAST搜索揭示了46種非冗餘Cpf1家族蛋白（圖64）。基於系統發育重構（圖64）選擇16種作為Cpf1多樣性的代表（圖100A-圖100B和圖106）。該等Cpf1家族蛋白跨越~1200個與~1500個胺基酸之間的長度範圍。

[1636] 該等Cpf1家族蛋白各自的同向重複序列顯示在該同向重複序列的3'處的19個核苷酸中的強保守性，這係重複序列包括在加工的crRNA中的部分（圖100C）。同向重複序列的5'序列係更加多樣的。選擇用於分析的16種Cpf1家族蛋白中的三種（2-毛螺菌科細菌*MC2017*，Lb3Cpf1；3-解朊丁酸弧菌，BpCpf1；以及6-密斯氏菌屬某種*SC_K08D17*，SsCpf1）與和FnCpf1同向重複序列顯著趨異的同向重複序列相關聯（圖100C）。值得注意地是，該等同向重複序列保留與FnCpf1同向重複序列相同或幾乎相同的莖環結構（圖100D）。

[1637] 測試直源同向重複序列支援體外FnCpf1核酸酶活性

的能力。含有保守性莖序列的同向重複序列能夠與FnCpf1可互換地起作用。來自候選物3（BpCpf1）的同向重複序列支持低水平的FnCpf1核酸酶活性（圖100E），這可能是由於3'-大部分U的保守性。

[1638] 使用體外PAM鑒定測定（圖107A）確定每種Cpf1家族蛋白的PAM序列。鑒定7種新Cpf1家族蛋白的PAM序列（圖100E和圖107B-圖107C），並且篩選證實FnCpf1的PAM為5'-TTN。Cpf1-家族蛋白的PAM序列主要係T富集的，主要造構成每種PAM的T數目方面改變（圖100F和圖107B-圖107C）。

實例14：Cpf1可以用於促進人類細胞中的基因組編輯

[1639] 對Cpf1家族蛋白進行密碼子優化並且使其連接用於最佳表現的C-末端核定位信號（NLS）與人類細胞中的核靶向（圖101A）。為了測試每種Cpf1家族蛋白的活性，在DNMT1基因內選擇指導RNA靶位點（圖101B）。每種Cpf1家族蛋白連同其被設計成靶向*DNMT1*的對應crRNA能夠體外切割*DNMT1*基因組區域的PCR擴增子（圖101C）。當在人類胚腎293FT（HEK 293FT）細胞中測試中，2種Cpf1家族蛋白（7－AsCpf1和13－LbCpf1）在所採用的條件下展現出可檢測水平的核酸酶誘導的indel（圖101C和圖101D）。

[1640] 測試具有另外基因組靶標的每種Cpf1-家族蛋白。AsCpf1和LbCpf1一致地介導HEK293FT細胞中強勁基因編輯（圖101E和圖108）。當與Cas9相比時，AsCpf1和LbCpf1介導了可比較水平的indel形成（圖101E）。另外，我們使用體外切割，隨後進行切割DNA端的桑格定序，並且發現7 - AsCpf1和13 - LbCpf1也生成

了交錯切割位點（圖101D和圖107E）。

[1641] 以下是FnCpf1構建體和異種同源物的核苷酸和胺基酸序列。

[1642] **FnCpf1座位序列**

[1643] **pFnCpf1**

[1644] *內源性新殺手弗朗西絲菌乙醯轉移酶的5'端（FnCpf1座位上游）*

[1645] FnCpf1

[1646] Cas4

[1647] Cas1

[1648] Cas2

[1649] **同向重複序列**

[1650] 間隔區

[1651]

CATCAAGGAATTGGTTCTAAGCTTATAGAAGCAATGATTAAGG
AAGCCAAAAAAAATAATATTGATGCAATATTTGTCTTAGGTCATCCA
AGTTATTATCCAAAATTTGGTTTTAAACCAGCCACAGAATATCAGAT
AAAATGTGAATATGATGTCCCAGCGGATGTTTTTATGGTACTAGATT
TGTCAGCTAAACTAGCTAGTTTAAAAGGACAAACTGTCTACTATGC
*CGATGAGTTTGGCAAAATTTTTTAG*ATCTACAAAATTATAAACTA
AATAAAGATTCTTATAATAACTTTATATATAATCGAAATGTAGAG

AATTTTATAAGGAGTCTTTATCATGTCAATTTATCAAGAATTTGT

TAATAAATATAGTTTAAGTAAAACTCTAAGATTTGAGTTAATCC

CACAGGGTAAAACACTTGAAAACATAAAAGCAAGAGGTTTGA

TTTTAGATGATGAGAAAAGAGCTAAAGACTACAAAAAGGCTA

AACAAATAATTGATAAATATCATCAGTTTTTTATAGAGGAGATA

TTAAGTTCGGTTTGTATTAGCGAAGATTTATTACAAAACTATTC

TGATGTTTATTTTAAACTTAAAAAGAGTGATGATGATAATCTAC

AAAAAGATTTTAAAAGTGCAAAAGATACGATAAAGAAACAAA

TATCTGAATATATAAAGGACTCAGAGAAATTTAAGAATTTGTTT

AATCAAAACCTTATCGATGCTAAAAAAGGGCAAGAGTCAGATT

TAATTCTATGGCTAAAGCAATCTAAGGATAATGGTATAGAACTA

TTTAAAGCCAATAGTGATATCACAGATATAGATGAGGCGTTAGA

AATAATCAAATCTTTTAAAGGTTGGACAACTTATTTTAAGGGTT

TTCATGAAAATAGAAAAAATGTTTATAGTAGCAATGATATTCCT

ACATCTATTATTTATAGGATAGTAGATGATAATTTGCCTAAATTT

CTAGAAAATAAAGCTAAGTATGAGAGTTTAAAAGACAAAGCT

CCAGAAGCTATAAACTATGAACAAATTAAAAAAGATTTGGCAG

AAGAGCTAACCTTTGATATTGACTACAAAACATCTGAAGTTAA

TCAAAGAGTTTTTTCACTTGATGAAGTTTTTGAGATAGCAAAC

TTTAATAATTATCTAAATCAAAGTGGTATTACTAAATTTAATACT

ATTATTGGTGGTAAATTTGTAAATGGTGAAAATACAAAGAGAA

AAGGTATAAATGAATATATAAATCTATACTCACAGCAAATAAAT

GATAAAACACTCAAAAAATATAAAATGAGTGTTTTATTTAAGC

AAATTTTAAGTGATACAGAATCTAAATCTTTTGTAATTGATAAG

TTAGAAGATGATAGTGATGTAGTTACAACGATGCAAAGTTTTT

ATGAGCAAATAGCAGCTTTTAAAACAGTAGAAGAAAAATCTAT

TAAAGAAACACTATCTTTATTATTTGATGATTTAAAAGCTCAAA

AACTTGATTTGAGTAAAATTTATTTTAAAAATGATAAATCTCTT

ACTGATCTATCACAACAAGTTTTTGATGATTATAGTGTTATTGG

TACAGCGGTACTAGAATATATAACTCAACAAATAGCACCTAAA

AATCTTGATAACCCTAGTAAGAAAGAGCAAGAATTAATAGCCA

AAAAAACTGAAAAAGCAAAATACTTATCTCTAGAAACTATAAA

GCTTGCCTTAGAAGAATTTAATAAGCATAGAGATATAGATAAAC

AGTGTAGGTTTGAAGAAATACTTGCAAACTTTGCGGCTATTCC

GATGATATTTGATGAAATAGCTCAAAACAAAGACAATTTGGCA

CAGATATCTATCAAATATCAAAATCAAGGTAAAAAAAGACCTAC

TTCAAGCTAGTGCGGAAGATGATGTTAAAGCTATCAAGGATCT

TTTAGATCAAACTAATAATCTCTTACATAAACTAAAAATATTTC

ATATTAGTCAGTCAGAAGATAAGGCAAATATTTTAGACAAGGA

TGAGCATTTTTATCTAGTATTTGAGGAGTGCTACTTTGAGCTAG

CGAATATAGTGCCTCTTTATAACAAAATTAGAAACTATATAACT

CAAAAGCCATATAGTGATGAGAAATTTAAGCTCAATTTTGAGA

ACTCGACTTTGGCTAATGGTTGGGATAAAAATAAAGAGCCTGA

CAATACGGCAATTTTATTTATCAAAGATGATAAATATTATCTGGG

TGTGATGAATAAGAAAAATAACAAAATATTTGATGATAAAGCT

ATCAAAGAAAATAAAGGCGAGGGTTATAAAAAAATTGTTTATA

AACTTTTACCTGGCGCAAATAAAATGTTACCTAAGGTTTTCTTT

TCTGCTAAATCTATAAAATTTTATAATCCTAGTGAAGATATACTT

AGAATAAGAAATCATTCCACACATACAAAAAATGGTAGTCCTC

AAAAAGGATATGAAAAATTTGAGTTTAATATTGAAGATTGCCG

AAAATTTATAGATTTTTATAAACAGTCTATAAGTAAGCATCCGG

AGTGGAAAGATTTTGGATTTAGATTTTCTGATACTCAAAGATAT

AATTCTATAGATGAATTTTATAGAGAAGTTGAAAATCAAGGCTA

CAAACTAACTTTTGAAAATATATCAGAGAGCTATATTGATAGCG

TAGTTAATCAGGGTAAATTGTACCTATTCCAAATCTATAATAAA

GATTTTTCAGCTTATAGCAAAGGGCGACCAAATCTACATACTTT

ATATTGGAAAGCGCTGTTTGATGAGAGAAATCTTCAAGATGTG

GTTTATAAGCTAAATGGTGAGGCAGAGCTTTTTTATCGTAAAC

AATCAATACCTAAAAAAATCACTCACCCAGCTAAAGAGGCAAT

AGCTAATAAAAACAAAGATAATCCTAAAAAAGAGAGTGTTTTT

GAATATGATTTAATCAAAGATAAACGCTTTACTGAAGATAAGTT

TTTCTTTCACTGTCCTATTACAATCAATTTTAAATCTAGTGGAG

CTAATAAGTTTAATGATGAAATCAATTTATTGCTAAAAGAAAAA

GCAAATGATGTTCATATATTAAGTATAGATAGAGGTGAAAGACA

TTTAGCTTACTATACTTTGGTAGATGGTAAAGGCAATATCATCA

AACAAGATACTTTCAACATCATTGGTAATGATAGAATGAAAAC

AAACTACCATGATAAGCTTGCTGCAATAGAGAAAGATAGGGAT

TCAGCTAGGAAAGACTGGAAAAAGATAAATAACATCAAAGAG

ATGAAAGAGGGCTATCTATCTCAGGTAGTTCATGAAATAGCTA

AGCTAGTTATAGAGTATAATGCTATTGTGGTTTTTGAGGATTTA

AATTTTGGATTTAAAAGAGGGCGTTTCAAGGTAGAGAAGCAG

GTCTATCAAAAGTTAGAAAAAATGCTAATTGAGAAACTAAACT

ATCTAGTTTTCAAAGATAATGAGTTTGATAAAACTGGGGGAGT

GCTTAGAGCTTATCAGCTAACAGCACCTTTTGAGACTTTTAAA

AAGATGGGTAAACAAACAGGTATTATCTACTATGTACCAGCTG

GTTTTACTTCAAAAATTTGTCCTGTAACTGGTTTTGTAAATCAG

TTATATCCTAAGTATGAAAGTGTCAGCAAATCTCAAGAGTTCTT

TAGTAAGTTTGACAAGATTTGTTATAACCTTGATAAGGGCTATT

TTGAGTTTAGTTTTGATTATAAAAACTTTGGTGACAAGGCTGC

CAAAGGCAAGTGGACTATAGCTAGCTTTGGGAGTAGATTGATT

AACTTTAGAAATTCAGATAAAAATCATAATTGGGATACTCGAG

AAGTTTATCCAACTAAAGAGTTGGAGAAATTGCTAAAAGATTA

TTCTATCGAATATGGGCATGGCGAATGTATCAAAGCAGCTATTT

GCGGTGAGAGCGACAAAAAGTTTTTTGCTAAGCTAACTAGTG

TCCTAAATACTATCTTACAAATGCGTAACTCAAAAACAGGTAC

TGAGTTAGATTATCTAATTTCACCAGTAGCAGATGTAAATGGCA

ATTTCTTTGATTCGCGACAGGCGCCAAAAAATATGCCTCAAGA

TGCTGATGCCAATGGTGCTTATCATATTGGGCTAAAAGGTCTGA

TGCTACTAGGTAGGATCAAAAATAATCAAGAGGGCAAAAAAC

TCAATTTGGTTATCAAAAATGAAGAGTATTTTGAGTTCGTGCA

GAATAGGAATAACTAATTCATTCAAGAATATATTACCCTGTCAG

TTTAGCGACTATTACCTCTTTAATAATTTGCAGGGGAATTATTTT

AGTAATAGTAATATACACAAGAGTTATTGATTATATGGAAAATT

ATATTTAGATAACATGGTTAAATGATTTTATATTCTGTCCTTACT

CGATATATTTGCATAATATCTATAGTAATGCCTCAGATACTACAT

ACTATTCATCTAGCCAAACAAAAGGGCGCGATGCTCATAAAAG

TATCGATAAAGGAATCTATAGTACCAAAAAAGATGACCTGATC

GGTATCGATGTTATTAACCATAAATATGGTTTGGTTGGTAAAAT

TGATGTTTTTCATAAAGATAAGGGCTTACTTGTGGAGAGAAAA

AGGCAAATCAAGACTATCTATGATGGCTATAAATATCAGCTTTA

TGCGCAATATTTTTGTCTCCAAGAGATGGGCTATGATGTCAAA

GCCATTAAATTTTATTCGATGGTTGATAATAAATCATACCCAATA

GCTATACCAACTTCAGCTGAGTTAGAAAAGTTTGAAAAACATA

TTCAAACAATCAAGCAATATAATCCAATGGATAACTCATTTAGG

CAAAATATTGAAAAGTGTAAATTTTGTATATATGCAAACTTATG

TGATAAAACGGACTTGTAGATTATGTTTAGTAAAAATGATATTG

AATCAAAGAATATAGTTTTTGTTAATATTTTTGATGGAGTGAAA

CTTAGTCTATCATTGGGGAATATAGTTATAAAAGATAAAGAAAC

TGATGAGGTGAAAACTAAGCTTTCTGTTCATAAAGTTCTTGCA

TTGTTTATCGTAGGTAATATGACGATGACCTCGCAACTTTTAGA

GACCTGTAAGAAAAATGCTATACAGCTAGTTTTTATGAAAAAT

AGCTTTAGACCATATCTATGTTTTGGTGATATTGCTGAGGCTAA

TTTTTTAGCTAGATATAAGCAATATAGTGTAGTTGAGCAAGATA

TAAGTTTAGCAAGGATTTTTATAACATCAAAGATACGCAATCAA

CATAACTTAGTCAAAAGCCTAAGAGATAAAACTCCAGAGCAG

CAAGAGATAGTCAAAAAGAATAAACAGCTAATAGCAGAGTTA

GAAAATACAACAAGCCTAGCGGAGCTAATGGGTATAGAGGGC

AATGTTGCCAAAAATTTCTTCAAAGGATTCTATGGACATTTAG

ATAGTTGGCAAGGGCGCAAACCTAGAATAAAACAGGATCCATA

TAATGTTGTTTTAGACTTGGGCTATAGTATGTTGTTTAATTTTGT

AGAGTGTTTTTTGCGACTTTTTGGCTTTGATTTATACAAGGGCT
TTTGTCATCAGACTTGGTATAAGCGTAAATCCCTAGTTTGTGAC
TTTGTTGAGCCATTTAGATGTATAGTGGATAACCAAGTTAGAAA
ATCATGGAATCTCGGGCAATTTTCTGTAGAGGATTTTGGTTGCA
AAAATGAGCAGTTTTATATAAAAAAAGATAAAACAAAAGACTA
CTCAAAAATACTTTTTGCCGAGATTATCAGCTACAAGCTAGAG
ATATTTGAATATGTAAGAGAATTTTATCGTGCCTTTATGCGAGG
CAAAGAAATTGCAGAGTATCCAATATTTTGTTATGAAACTAGG
AGGGTGTATGTTGATAGTCAGTTATGATTTTAGTAATAATAAAG
TACGTGCAAAGTTTGCCAAATTTCTAGAAAGTTATGGTGTACG
TTTACAATATTCGGTATTTGAGCTCAAATATAGCAAGAGAATGT
TAGACTTGATTTTAGCTGAGATAGAAAATAACTATGTACCACTA
TTTACAAATGCTGATAGTGTTTTAATCTTTAATGCTCCAGATAA
AGATGTGATAAAATATGGTTATGCGATTCATAGAGAACAAGAG
GTTGTTTTTATAGACTAAAAATTGCAAACCTTAGTCTTTATGTT
AAAATAACTACTAAGTTCTTAGAGATATTTAAAAATATGACTGT
TGTTATATATCAAAATGCTAAAAAAATCATAGATTTTAGGTCTT
TTTTTGCTGATTTAGGCAAAAACGG**GTCTAAGAACTTTAAAT**
AATTTCTACTGTTGTAGATGAGAAGTCATTTAATAAGGCCAC
TGTTAAAA**GTCTAAGAACTTTAAATAATTTCTACTGTTGTAG**
ATGCTACTATTCCTGTGCCTTCAGATAATTCA**GTCTAAGAACT**
TTAAATAATTTCTACTGTTGTAGATGTCTAGAGCCTTTTGTAT
TAGTAGCCG**GTCTAAGAACTTTAAATAATTTCTACTGTTGTA**
GATTAGCGATTTATGAAGGTCATTTTTTTGTCT (SEQ ID NO: 211)

[1652] **pFnCpf1_min**

[1653] Lac啟動子

[1654] *夏因-達爾加諾序列*

[1655] FnCpf1

[1656] J23119啟動子

[1657] **同向重複序列**

[1658] 間隔區

[1659]

TTTACACTTTATGCTTCCGGCTCGTATGTT*AGGAGG*TCTTT
ATCATGTCAATTTATCAAGAATTTGTTAATAAATATAGTTTAAGT
AAAACTCTAAGATTTGAGTTAATCCCACAGGGTAAAACACTTG
AAAACATAAAAGCAAGAGGTTTGATTTTAGATGATGAGAAAA
GAGCTAAAGACTACAAAAAGGCTAAACAAATAATTGATAAATA
TCATCAGTTTTTTATAGAGGAGATATTAAGTTCGGTTTGTATTA
GCGAAGATTTATTACAAAACTATTCTGATGTTTATTTTAAACTT
AAAAAGAGTGATGATGATAATCTACAAAAAGATTTTAAAAGTG
CAAAAGATACGATAAAGAAACAAATATCTGAATATATAAAGGA
CTCAGAGAAATTTAAGAATTTGTTTAATCAAAACCTTATCGATG
CTAAAAAAGGGCAAGAGTCAGATTTAATTCTATGGCTAAAGCA
ATCTAAGGATAATGGTATAGAACTATTTAAAGCCAATAGTGATA
TCACAGATATAGATGAGGCGTTAGAAATAATCAAATCTTTTAAA
GGTTGGACAACTTATTTTAAGGGTTTTCATGAAAATAGAAAAA

ATGTTTATAGTAGCAATGATATTCCTACATCTATTATTTATAGGAT

AGTAGATGATAATTTGCCTAAATTTCTAGAAAATAAAGCTAAGT

ATGAGAGTTTAAAAGACAAAGCTCCAGAAGCTATAAACTATGA

ACAAATTAAAAAAGATTTGGCAGAAGAGCTAACCTTTGATATT

GACTACAAAACATCTGAAGTTAATCAAAGAGTTTTTTCACTTG

ATGAAGTTTTTGAGATAGCAAACTTTAATAATTATCTAAATCAA

AGTGGTATTACTAAATTTAATACTATTATTGGTGGTAAATTTGTA

AATGGTGAAAATACAAAGAGAAAAGGTATAAATGAATATATAA

ATCTATACTCACAGCAAATAAATGATAAAACACTCAAAAAATAT

AAAATGAGTGTTTTATTTAAGCAAATTTTAAGTGATACAGAATC

TAAATCTTTTGTAATTGATAAGTTAGAAGATGATAGTGATGTAG

TTACAACGATGCAAAGTTTTTATGAGCAAATAGCAGCTTTTAA

AACAGTAGAAGAAAAATCTATTAAAGAAACACTATCTTTATTAT

TTGATGATTTAAAAGCTCAAAAACTTGATTTGAGTAAAATTTAT

TTTAAAAATGATAAATCTCTTACTGATCTATCACAACAAGTTTT

TGATGATTATAGTGTTATTGGTACAGCGGTACTAGAATATATAA

CTCAACAAATAGCACCTAAAAATCTTGATAACCCTAGTAAGAA

AGAGCAAGAATTAATAGCCAAAAAAACTGAAAAAGCAAAATA

CTTATCTCTAGAAACTATAAAGCTTGCCTTAGAAGAATTTAATA

AGCATAGAGATATAGATAAACAGTGTAGGTTTGAAGAAATACT

TGCAAACTTTGCGGCTATTCCGATGATATTTGATGAAATAGCTC

AAAACAAAGACAATTTGGCACAGATATCTATCAAATATCAAAA

TCAAGGTAAAAAAAGACCTACTTCAAGCTAGTGCGGAAGATGA

TGTTAAAGCTATCAAGGATCTTTTAGATCAAACTAATAATCTCT

TACATAAACTAAAAATATTTCATATTAGTCAGTCAGAAGATAAG

GCAAATATTTTAGACAAGGATGAGCATTTTTATCTAGTATTTGA

GGAGTGCTACTTTGAGCTAGCGAATATAGTGCCTCTTTATAACA

AAATTAGAAACTATATAACTCAAAAGCCATATAGTGATGAGAA

ATTTAAGCTCAATTTTGAGAACTCGACTTTGGCTAATGGTTGG

GATAAAAATAAAGAGCCTGACAATACGGCAATTTTATTTATCAA

AGATGATAAATATTATCTGGGTGTGATGAATAAGAAAAATAACA

AAATATTTGATGATAAAGCTATCAAAGAAAATAAAGGCGAGGG

TTATAAAAAAAATTGTTTATAAACTTTTACCTGGCGCAAATAAAA

TGTTACCTAAGGTTTTCTTTTCTGCTAAATCTATAAAATTTTATA

ATCCTAGTGAAGATATACTTAGAATAAGAAATCATTCCACACAT

ACAAAAAATGGTAGTCCTCAAAAAGGATATGAAAAATTTGAGT

TTAATATTGAAGATTGCCGAAAATTTATAGATTTTTATAAACAG

TCTATAAGTAAGCATCCGGAGTGGAAAGATTTTGGATTTAGATT

TTCTGATACTCAAAGATATAATTCTATAGATGAATTTTATAGAGA

AGTTGAAAATCAAGGCTACAAACTAACTTTTGAAAATATATCA

GAGAGCTATATTGATAGCGTAGTTAATCAGGGTAAATTGTACCT

ATTCCAAATCTATAATAAAGATTTTTCAGCTTATAGCAAAGGGC

GACCAAATCTACATACTTTATATTGGAAAGCGCTGTTTGATGAG

AGAAATCTTCAAGATGTGGTTTATAAGCTAAATGGTGAGGCAG

AGCTTTTTTATCGTAAACAATCAATACCTAAAAAAATCACTCAC

CCAGCTAAAGAGGCAATAGCTAATAAAAACAAAGATAATCCTA

AAAAAGAGAGTGTTTTTGAATATGATTTAATCAAAGATAAACG

CTTTACTGAAGATAAGTTTTTCTTTCACTGTCCTATTACAATCA

ATTTTAAATCTAGTGGAGCTAATAAGTTTAATGATGAAATCAAT

TTATTGCTAAAAGAAAAAGCAAATGATGTTCATATATTAAGTAT

AGATAGAGGTGAAAGACATTTAGCTTACTATACTTTGGTAGAT

GGTAAAGGCAATATCATCAAACAAGATACTTTCAACATCATTG

GTAATGATAGAATGAAAACAAACTACCATGATAAGCTTGCTGC

AATAGAGAAAGATAGGGATTCAGCTAGGAAAGACTGGAAAAA

GATAAATAACATCAAAGAGATGAAAGAGGGCTATCTATCTCAG

GTAGTTCATGAAATAGCTAAGCTAGTTATAGAGTATAATGCTAT

TGTGGTTTTTGAGGATTTAAATTTTGGATTTAAAAGAGGGCGT

TTCAAGGTAGAGAAGCAGGTCTATCAAAAGTTAGAAAAAATG

CTAATTGAGAAACTAAACTATCTAGTTTTCAAAGATAATGAGTT

TGATAAAACTGGGGGAGTGCTTAGAGCTTATCAGCTAACAGCA

CCTTTTGAGACTTTTAAAAAGATGGGTAAACAAACAGGTATTA

TCTACTATGTACCAGCTGGTTTTACTTCAAAAATTTGTCCTGTA

ACTGGTTTTGTAAATCAGTTATATCCTAAGTATGAAAGTGTCAG

CAAATCTCAAGAGTTCTTTAGTAAGTTTGACAAGATTTGTTATA

ACCTTGATAAGGGCTATTTTGAGTTTAGTTTTGATTATAAAAAC

TTTGGTGACAAGGCTGCCAAAGGCAAGTGGACTATAGCTAGC

TTTGGGAGTAGATTGATTAACTTTAGAAATTCAGATAAAAATC

ATAATTGGGATACTCGAGAAGTTTATCCAACTAAAGAGTTGGA

GAAATTGCTAAAAGATTATTCTATCGAATATGGGCATGGCGAAT

GTATCAAAGCAGCTATTTGCGGTGAGAGCGACAAAAAGTTTTT

TGCTAAGCTAACTAGTGTCCTAAATACTATCTTACAAATGCGTA

ACTCAAAAACAGGTACTGAGTTAGATTATCTAATTTCACCAGT

AGCAGATGTAAATGGCAATTTCTTTGATTCGCGACAGGCGCCA
AAAAATATGCCTCAAGATGCTGATGCCAATGGTGCTTATCATAT
TGGGCTAAAAGGTCTGATGCTACTAGGTAGGATCAAAAATAAT
CAAGAGGGCAAAAAACTCAATTTGGTTATCAAAAATGAAGAG
TATTTTGAGTTCGTGCAGAATAGGAATAACTAATTGACAGCTA
GCTCAGTCCTAGGTATAATGCTAGCGCTGATTTAGGCAAAAAC
GG**GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT**GAGA
AGTCATTTAATAAGGCCACTGTTAAAA**GTCTAAGAACTTTAA**
ATAATTTCTACTGTTGTAGATGCTACTATTCCTGTGCCTTCAG
ATAATTCA**GTCTAAGAACTTTAAATAATTTCTACTGTTGTAG**
A（SEQ ID NO: 212）

[1660] **pFnCpf1_ΔCas**

[1661] *內源性新殺手弗朗西絲菌乙醯轉移酶的5'端（FnCpf1座位上游）*

[1662] FnCpf1

[1663] **同向重複序列**

[1664] 間隔區

[1665]

*CTGTCTACTATGCCGATGAGTTTGGCAAAATTTTTTAG*ATCTA
CAAAATTATAAACTAAATAAAGATTCTTATAATAACTTTATATAT
AATCGAAATGTAGAGAATTTTATAAGGAGTCTTTATCATGTCAA
TTTATCAAGAATTTGTTAATAAATATAGTTTAAGTAAAACTCTA

AGATTTGAGTTAATCCCACAGGGTAAAACACTTGAAAACATAA

AAGCAAGAGGTTTGATTTTAGATGATGAGAAAAGAGCTAAAG

ACTACAAAAAGGCTAAACAAATAATTGATAAATATCATCAGTTT

TTTATAGAGGAGATATTAAGTTCGGTTTGTATTAGCGAAGATTT

ATTACAAAACTATTCTGATGTTTATTTTAAACTTAAAAAGAGTG

ATGATGATAATCTACAAAAAGATTTTAAAAGTGCAAAAGATAC

GATAAAGAAACAAATATCTGAATATATAAAGGACTCAGAGAAA

TTTAAGAATTTGTTTAATCAAAACCTTATCGATGCTAAAAAAG

GGCAAGAGTCAGATTTAATTCTATGGCTAAAGCAATCTAAGGA

TAATGGTATAGAACTATTTAAAGCCAATAGTGATATCACAGATA

TAGATGAGGCGTTAGAAATAATCAAATCTTTTAAAGGTTGGAC

AACTTATTTTAAGGGTTTTCATGAAAATAGAAAAAATGTTTATA

GTAGCAATGATATTCCTACATCTATTATTTATAGGATAGTAGATG

ATAATTTGCCTAAATTTCTAGAAAATAAAGCTAAGTATGAGAGT

TTAAAAGACAAAGCTCCAGAAGCTATAAACTATGAACAAATTA

AAAAAGATTTGGCAGAAGAGCTAACCTTTGATATTGACTACAA

AACATCTGAAGTTAATCAAAGAGTTTTTTCACTTGATGAAGTT

TTTGAGATAGCAAACTTTAATAATTATCTAAATCAAAGTGGTAT

TACTAAATTTAATACTATTATTGGTGGTAAATTTGTAAATGGTGA

AAATACAAAGAGAAAAGGTATAAATGAATATATAAATCTATACT

CACAGCAAATAAATGATAAAACACTCAAAAAATATAAAATGAG

TGTTTTATTTAAGCAAATTTTAAGTGATACAGAATCTAAATCTT

TTGTAATTGATAAGTTAGAAGATGATAGTGATGTAGTTACAACG

ATGCAAAGTTTTTATGAGCAAATAGCAGCTTTTAAAACAGTAG

AAGAAAAATCTATTAAAGAAACACTATCTTTATTATTTGATGAT

TTAAAAGCTCAAAAACTTGATTTGAGTAAAATTTATTTTAAAA

ATGATAAATCTCTTACTGATCTATCACAACAAGTTTTTGATGAT

TATAGTGTTATTGGTACAGCGGTACTAGAATATATAACTCAACA

AATAGCACCTAAAAATCTTGATAACCCTAGTAAGAAAGAGCAA

GAATTAATAGCCAAAAAAACTGAAAAAGCAAAATACTTATCTC

TAGAAACTATAAAGCTTGCCTTAGAAGAATTTAATAAGCATAG

AGATATAGATAAACAGTGTAGGTTTGAAGAAATACTTGCAAAC

TTTGCGGCTATTCCGATGATATTTGATGAAATAGCTCAAAACAA

AGACAATTTGGCACAGATATCTATCAAATATCAAAATCAAGGT

AAAAAAGACCTACTTCAAGCTAGTGCGGAAGATGATGTTAAA

GCTATCAAGGATCTTTTAGATCAAACTAATAATCTCTTACATAA

ACTAAAAATATTTCATATTAGTCAGTCAGAAGATAAGGCAAATA

TTTTAGACAAGGATGAGCATTTTTATCTAGTATTTGAGGAGTGC

TACTTTGAGCTAGCGAATATAGTGCCTCTTTATAACAAAATTAG

AAACTATATAACTCAAAAGCCATATAGTGATGAGAAATTTAAG

CTCAATTTTGAGAACTCGACTTTGGCTAATGGTTGGGATAAAA

ATAAAGAGCCTGACAATACGGCAATTTTATTTATCAAAGATGAT

AAATATTATCTGGGTGTGATGAATAAGAAAAATAACAAAATATT

TGATGATAAAGCTATCAAAGAAAATAAAGGCGAGGGTTATAAA

AAAATTGTTTATAAACTTTTACCTGGCGCAAATAAAATGTTACC

TAAGGTTTTCTTTTCTGCTAAATCTATAAAATTTTATAATCCTAG

TGAAGATATACTTAGAATAAGAAATCATTCCACACATACAAAA

AATGGTAGTCCTCAAAAAGGATATGAAAAATTTGAGTTTAATA

TTGAAGATTGCCGAAAATTTATAGATTTTTATAAACAGTCTATA

AGTAAGCATCCGGAGTGGAAAGATTTTGGATTTAGATTTTCTG

ATACTCAAAGATATAATTCTATAGATGAATTTTATAGAGAAGTT

GAAAATCAAGGCTACAAACTAACTTTTGAAAATATATCAGAGA

GCTATATTGATAGCGTAGTTAATCAGGGTAAATTGTACCTATTC

CAAATCTATAATAAAGATTTTTCAGCTTATAGCAAAGGGCGACC

AAATCTACATACTTTATATTGGAAAGCGCTGTTTGATGAGAGA

AATCTTCAAGATGTGGTTTATAAGCTAAATGGTGAGGCAGAGC

TTTTTTATCGTAAACAATCAATACCTAAAAAAATCACTCACCCA

GCTAAAGAGGCAATAGCTAATAAAAACAAAGATAATCCTAAAA

AAGAGAGTGTTTTTGAATATGATTTAATCAAAGATAAACGCTTT

ACTGAAGATAAGTTTTTCTTTCACTGTCCTATTACAATCAATTT

TAAATCTAGTGGAGCTAATAAGTTTAATGATGAAATCAATTTAT

TGCTAAAAGAAAAAGCAAATGATGTTCATATATTAAGTATAGAT

AGAGGTGAAAGACATTTAGCTTACTATACTTTGGTAGATGGTA

AAGGCAATATCATCAAACAAGATACTTTCAACATCATTGGTAAT

GATAGAATGAAAACAAACTACCATGATAAGCTTGCTGCAATAG

AGAAAGATAGGGATTCAGCTAGGAAAGACTGGAAAAAGATAA

ATAACATCAAAGAGATGAAAGAGGGCTATCTATCTCAGGTAGT

TCATGAAATAGCTAAGCTAGTTATAGAGTATAATGCTATTGTGG

TTTTTGAGGATTTAAATTTTGGATTTAAAAGAGGGCGTTTCAA

GGTAGAGAAGCAGGTCTATCAAAAGTTAGAAAAAATGCTAATT

GAGAAACTAAACTATCTAGTTTTCAAAGATAATGAGTTTGATA

AAACTGGGGGAGTGCTTAGAGCTTATCAGCTAACAGCACCTTT

TGAGACTTTTAAAAAGATGGGTAAACAAACAGGTATTATCTAC

TATGTACCAGCTGGTTTTACTTCAAAAATTTGTCCTGTAACTGG

TTTTGTAAATCAGTTATATCCTAAGTATGAAAGTGTCAGCAAAT

CTCAAGAGTTCTTTAGTAAGTTTGACAAGATTTGTTATAACCTT

GATAAGGGCTATTTTGAGTTTAGTTTTGATTATAAAAACTTTGG

TGACAAGGCTGCCAAAGGCAAGTGGACTATAGCTAGCTTTGG

GAGTAGATTGATTAACTTTAGAAATTCAGATAAAAATCATAATT

GGGATACTCGAGAAGTTTATCCAACTAAAGAGTTGGAGAAATT

GCTAAAAGATTATTCTATCGAATATGGGCATGGCGAATGTATCA

AAGCAGCTATTTGCGGTGAGAGCGACAAAAAGTTTTTTGCTAA

GCTAACTAGTGTCCTAAATACTATCTTACAAATGCGTAACTCAA

AAACAGGTACTGAGTTAGATTATCTAATTTCACCAGTAGCAGA

TGTAAATGGCAATTTCTTTGATTCGCGACAGGCGCCAAAAAAT

ATGCCTCAAGATGCTGATGCCAATGGTGCTTATCATATTGGGCT

AAAAGGTCTGATGCTACTAGGTAGGATCAAAAATAATCAAGAG

GGCAAAAAACTCAATTTGGTTATCAAAAATGAAGAGTATTTTG

AGTTCGTGCAGAATAGGAATAACTAATTCATTCAAGAATATATT

ACCCTGTCAGTTTAGCGACTATTACCTCTTTAATAATTTGCAGG

GGAATTATTTTAGTAATAGTAATATACACAAGAGTTATTGATTAT

ATGGAAAATTATATTTAGATAACATGGTTAAATGATTTTATATTC

TGTCCTTACTCGATATATTTTTTATAGACTAAAAATTGCAAACC

TTAGTCTTTATGTTAAAATAACTACTAAGTTCTTAGAGATATTTA

AAAATATGACTGTTGTTATATATCAAAATGCTAAAAAAATCATA

GATTTTAGGTCTTTTTTTGCTGATTTAGGCAAAAACGG**GTCTA**

AGAACTTTAAATAATTTCTACTGTTGTAGATGAGAAGTCATT
TAATAAGGCCACTGTTAAAA**GTCTAAGAACTTTAAATAATTT**
CTACTGTTGTAGATGCTACTATTCCTGTGCCTTCAGATAATTC
A**GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT**GTCTA
GAGCCTTTTGTATTAGTAGCCG**GTCTAAGAACTTTAAATAATT**
TCTACTGTTGTAGATTAGCGATTTATGAAGGTCATTTTTTTGT
CT（SEQ ID NO: 213）

[1666] **人類密碼子優化的Cpf1異種同源物的核苷酸序列**

[1667] *核定位信號（NLS）*

[1668] 甘胺酸-絲胺酸接頭

[1669] **3× HA標籤**

[1670] **1-土拉熱弗朗西絲菌新殺手亞種U112（FnCpf1）**

[1671]

ATGAGCATCTACCAGGAGTTCGTCAACAAGTATTCACTGA
GTAAGACACTGCGGTTCGAGCTGATCCCACAGGGCAAGACAC
TGGAGAACATCAAGGCCCGAGGCCTGATTCTGGACGATGAGA
AGCGGGCAAAAGACTATAAGAAAGCCAAGCAGATCATTGATA
AATACCACCAGTTCTTTATCGAGGAAATTCTGAGCTCCGTGTG
CATCAGTGAGGATCTGCTGCAGAATTACTCAGACGTGTACTTC
AAGCTGAAGAAGAGCGACGATGACAACCTGCAGAAGGACTTC
AAGTCCGCCAAGGACACCATCAAGAAACAGATTAGCGAGTAC
ATCAAGGACTCCGAAAAGTTTAAAAATCTGTTCAACCAGAATC

TGATCGATGCTAAGAAAGGCCAGGAGTCCGACCTGATCCTGTG

GCTGAAACAGTCTAAGGACAATGGGATTGAACTGTTCAAGGC

TAACTCCGATATCACTGATATTGACGAGGCACTGGAAATCATC

AAGAGCTTCAAGGGATGGACCACATACTTTAAAGGCTTCCACG

AGAACCGCAAGAACGTGTACTCCAGCAACGACATTCCTACCT

CCATCATCTACCGAATCGTCGATGACAATCTGCCAAAGTTCCT

GGAGAACAAGGCCAAATATGAATCTCTGAAGGACAAAGCTCC

CGAGGCAATTAATTACGAACAGATCAAGAAAGATCTGGCTGA

GGAACTGACATTCGATATCGACTATAAGACTAGCGAGGTGAAC

CAGAGGGTCTTTTCCCTGGACGAGGTGTTTGAAATCGCCAATT

TCAACAATTACCTGAACCAGTCCGGCATTACTAAATTCAATAC

CATCATTGGCGGGAAGTTTGTGAACGGGGAGAATACCAAGCG

CAAGGGAATTAACGAATACATCAATCTGTATAGCCAGCAGATC

AACGACAAAACTCTGAAGAAATACAAGATGTCTGTGCTGTTC

AAACAGATCCTGAGTGATACCGAGTCCAAGTCTTTTGTCATTG

ATAAACTGGAAGATGACTCAGACGTGGTCACTACCATGCAGA

GCTTTTATGAGCAGATCGCCGCTTTCAAGACAGTGGAGGAAA

AATCTATTAAGGAAACTCTGAGTCTGCTGTTCGATGACCTGAA

AGCCCAGAAGCTGGACCTGAGTAAGATCTACTTCAAAAACGA

TAAGAGTCTGACAGACCTGTCACAGCAGGTGTTTGATGACTAT

TCCGTGATTGGGACCGCCGTCCTGGAGTACATTACACAGCAGA

TCGCTCCAAAGAACCTGGATAATCCCTCTAAGAAAGAGCAGG

AACTGATCGCTAAGAAAACCGAGAAGGCAAAATATCTGAGTC

TGGAAACAATTAAGCTGGCACTGGAGGAGTTCAACAAGCACA

GGGATATTGACAAACAGTGCCGCTTTGAGGAAATCCTGGCCA

ACTTCGCAGCCATCCCCATGATTTTTGATGAGATCGCCCAGAA

CAAAGACAATCTGGCTCAGATCAGTATTAAGTACCAGAACCAG

GGCAAGAAAGACCTGCTGCAGGCTTCAGCAGAAGATGACGTG

AAAGCCATCAAGGATCTGCTGGACCAGACCAACAATCTGCTG

CACAAGCTGAAAATCTTCCATATTAGTCAGTCAGAGGATAAGG

CTAATATCCTGGATAAAGACGAACACTTCTACCTGGTGTTCGA

GGAATGTTACTTCGAGCTGGCAAACATTGTCCCCCTGTATAAC

AAGATTAGGAACTACATCACACAGAAGCCTTACTCTGACGAG

AAGTTTAAACTGAACTTCGAAAATAGTACCCTGGCCAACGGGT

GGGATAAGAACAAGGAGCCTGACAACACAGCTATCCTGTTCAT

CAAGGATGACAAGTACTATCTGGGAGTGATGAATAAGAAAAA

CAATAAGATCTTCGATGACAAAGCCATTAAGGAGAACAAAGG

GGAAGGATACAAGAAAATCGTGTATAAGCTGCTGCCCGGCGC

AAATAAGATGCTGCCTAAGGTGTTCTTCAGCGCCAAGAGTATC

AAATTCTACAACCCATCCGAGGACATCCTGCGGATTAGAAATC

ACTCAACACATACTAAGAACGGGAGCCCCCAGAAGGGATATG

AGAAATTTGAGTTCAACATCGAGGATTGCAGGAAGTTTATTGA

CTTCTACAAGCAGAGCATCTCCAAACACCCTGAATGGAAGGAT

TTTGGCTTCCGGTTTTCCGACACACAGAGATATAACTCTATCGA

CGAGTTCTACCGCGAGGTGGAAAATCAGGGGTATAAGCTGAC

TTTTGAGAACATTTCTGAAAGTTACATCGACAGCGTGGTCAAT

CAGGGAAAGCTGTACCTGTTCCAGATCTATAACAAAGATTTTT

CAGCATACAGCAAGGGCAGACCAAACCTGCATACACTGTACT

GGAAGGCCCTGTTCGATGAGAGGAATCTGCAGGACGTGGTCT

ATAAACTGAACGGAGAGGCCGAACTGTTTTACCGGAAGCAGT

CTATTCCTAAGAAAATCACTCACCCAGCTAAGGAGGCCATCGC

TAACAAGAACAAGGACAATCCTAAGAAAGAGAGCGTGTTCGA

ATACGATCTGATTAAGGACAAGCGGTTCACCGAAGATAAGTTC

TTTTTCCATTGTCCAATCACCATTAACTTCAAGTCAAGCGGCG

CTAACAAGTTCAACGACGAGATCAATCTGCTGCTGAAGGAAA

AAGCAAACGATGTGCACATCCTGAGCATTGACCGAGGAGAGC

GGCATCTGGCCTACTATACCCTGGTGGATGGCAAAGGGAATAT

CATTAAGCAGGATACATTCAACATCATTGGCAATGACCGGATG

AAAACCAACTACCACGATAAACTGGCTGCAATCGAGAAGGAT

AGAGACTCAGCTAGGAAGGACTGGAAGAAAATCAACAACATT

AAGGAGATGAAGGAAGGCTATCTGAGCCAGGTGGTCCATGAG

ATTGCAAAGCTGGTCATCGAATACAATGCCATTGTGGTGTTCG

AGGATCTGAACTTCGGCTTTAAGAGGGGGCGCTTTAAGGTGG

AAAAACAGGTCTATCAGAAGCTGGAGAAAATGCTGATCGAAA

AGCTGAATTACCTGGTGTTTAAAGATAACGAGTTCGACAAGAC

CGGAGGCGTCCTGAGAGCCTACCAGCTGACAGCTCCCTTTGA

AACTTTCAAGAAAATGGGAAAACAGACAGGCATCATCTACTAT

GTGCCAGCCGGATTCACTTCCAAGATCTGCCCCGTGACCGGCT

TTGTCAACCAGCTGTACCCTAAATATGAGTCAGTGAGCAAGTC

CCAGGAATTTTTCAGCAAGTTCGATAAGATCTGTTATAATCTGG

ACAAGGGGTACTTCGAGTTTTCCTTCGATTACAAGAACTTCGG

CGACAAGGCCGCTAAGGGGAAATGGACCATTGCCTCCTTCGG

ATCTCGCCTGATCAACTTTCGAAATTCCGATAAAAACCACAAT
TGGGACACTAGGGAGGTGTACCCAACCAAGGAGCTGGAAAAG
CTGCTGAAAGACTACTCTATCGAGTATGGACATGGCGAATGCA
TCAAGGCAGCCATCTGTGGCGAGAGTGATAAGAAATTTTTCGC
CAAGCTGACCTCAGTGCTGAATACAATCCTGCAGATGCGGAAC
TCAAAGACCGGGACAGAACTGGACTATCTGATTAGCCCCGTG
GCTGATGTCAACGGAAACTTCTTCGACAGCAGACAGGCACCC
AAAAATATGCCTCAGGATGCAGACGCCAACGGGGCCTACCAC
ATCGGGCTGAAGGGACTGATGCTGCTGGGCCGGATCAAGAAC
AATCAGGAGGGGAAGAAGCTGAACCTGGTCATTAAGAACGAG
GAATACTTCGAGTTTGTCCAGAATAGAAATAAC*AAAAGGCCGG*
*CGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAA*GGGATCC
TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGT
GCCTGATTATGCATACCCATATGATGTCCCCGACTATGCC
（SEQ ID NO: 214）

[1672]　3-毛螺菌科細菌MC2017（Lb3Cpf1）

[1673]

ATGGATTACGGCAACGGCCAGTTTGAGCGGAGAGCCCCC
CTGACCAAGACAATCACCCTGCGCCTGAAGCCTATCGGCGAG
ACACGGGAGACAATCCGCGAGCAGAAGCTGCTGGAGCAGGA
CGCCGCCTTCAGAAAGCTGGTGGAGACAGTGACCCCTATCGT
GGACGATTGTATCAGGAAGATCGCCGATAACGCCCTGTGCCAC
TTTGGCACCGAGTATGACTTCAGCTGTCTGGGCAACGCCATCT

CTAAGAATGACAGCAAGGCCATCAAGAAGGAGACAGAGAAG

GTGGAGAAGCTGCTGGCCAAGGTGCTGACCGAGAATCTGCCA

GATGGCCTGCGCAAGGTGAACGACATCAATTCCGCCGCCTTTA

TCCAGGATACACTGACCTCTTTCGTGCAGGACGATGCCGACAA

GCGGGTGCTGATCCAGGAGCTGAAGGGCAAGACCGTGCTGAT

GCAGCGGTTCCTGACCACACGGATCACAGCCCTGACCGTGTG

GCTGCCCGACAGAGTGTTCGAGAACTTTAATATCTTCATCGAG

AACGCCGAGAAGATGAGAATCCTGCTGGACTCCCCTCTGAAT

GAGAAGATCATGAAGTTTGACCCAGATGCCGAGCAGTACGCC

TCTCTGGAGTTCTATGGCCAGTGCCTGTCTCAGAAGGACATCG

ATAGCTACAACCTGATCATCTCCGGCATCTATGCCGACGATGAG

GTGAAGAACCCTGGCATCAATGAGATCGTGAAGGAGTACAAT

CAGCAGATCCGGGGCGACAAGGATGAGTCCCCACTGCCCAAG

CTGAAGAAGCTGCACAAGCAGATCCTGATGCCAGTGGAGAAG

GCCTTCTTTGTGCGCGTGCTGTCTAACGACAGCGATGCCCGGA

GCATCCTGGAGAAGATCCTGAAGGACACAGAGATGCTGCCCT

CCAAGATCATCGAGGCCATGAAGGAGGCAGATGCAGGCGACA

TCGCCGTGTACGGCAGCCGGCTGCACGAGCTGAGCCACGTGA

TCTACGGCGATCACGGCAAGCTGTCCCAGATCATCTATGACAA

GGAGTCCAAGAGGATCTCTGAGCTGATGGAGACACTGTCTCC

AAAGGAGCGCAAGGAGAGCAAGAAGCGGCTGGAGGGCCTGG

AGGAGCACATCAGAAAGTCTACATACACCTTCGACGAGCTGA

ACAGGTATGCCGAGAAGAATGTGATGGCAGCATACATCGCAGC

AGTGGAGGAGTCTTGTGCCGAGATCATGAGAAAGGAGAAGGA

TCTGAGGACCCTGCTGAGCAAGGAGGACGTGAAGATCCGGGG

CAACAGACACAATACACTGATCGTGAAGAACTACTTTAATGCC

TGGACCGTGTTCCGGAACCTGATCAGAATCCTGAGGCGCAAG

TCCGAGGCCGAGATCGACTCTGACTTCTACGATGTGCTGGACG

ATTCCGTGGAGGTGCTGTCTCTGACATACAAGGGCGAGAATCT

GTGCCGCAGCTATATCACCAAGAAGATCGGCTCCGACCTGAAG

CCCGAGATCGCCACATACGGCAGCGCCCTGAGGCCTAACAGC

CGCTGGTGGTCCCCAGGAGAGAAGTTTAATGTGAAGTTCCAC

ACCATCGTGCGGAGAGATGGCCGGCTGTACTATTTCATCCTGC

CCAAGGGCGCCAAGCCTGTGGAGCTGGAGGACATGGATGGCG

ACATCGAGTGTCTGCAGATGAGAAAGATCCCTAACCCAACAAT

CTTTCTGCCCAAGCTGGTGTTCAAGGACCCTGAGGCCTTCTTT

AGGGATAATCCAGAGGCCGACGAGTTCGTGTTTCTGAGCGGC

ATGAAGGCCCCCGTGACAATCACCAGAGAGACATACGAGGCC

TACAGGTATAAGCTGTATACCGTGGGCAAGCTGCGCGATGGCG

AGGTGTCCGAAGAGGAGTACAAGCGGGCCCTGCTGCAGGTGC

TGACCGCCTACAAGGAGTTTCTGGAGAACAGAATGATCTATGC

CGACCTGAATTTCGGCTTTAAGGATCTGGAGGAGTATAAGGAC

AGCTCCGAGTTTATCAAGCAGGTGGAGACACACAACACCTTC

ATGTGCTGGGCCAAGGTGTCTAGCTCCCAGCTGGACGATCTGG

TGAAGTCTGGCAACGGCCTGCTGTTCGAGATCTGGAGCGAGC

GCCTGGAGTCCTACTATAAGTACGGCAATGAGAAGGTGCTGCG

GGGCTATGAGGGCGTGCTGCTGAGCATCCTGAAGGATGAGAA

CCTGGTGTCCATGCGGACCCTGCTGAACAGCCGGCCCATGCTG

GTGTACCGGCCAAAGGAGTCTAGCAAGCCTATGGTGGTGCAC

CGGGATGGCAGCAGAGTGGTGGACAGGTTTGATAAGGACGGC

AAGTACATCCCCCCTGAGGTGCACGACGAGCTGTATCGCTTCT

TTAACAATCTGCTGATCAAGGAGAAGCTGGGCGAGAAGGCCC

GGAAGATCCTGGACAACAAGAAGGTGAAGGTGAAGGTGCTG

GAGAGCGAGAGAGTGAAGTGGTCCAAGTTCTACGATGAGCAG

TTTGCCGTGACCTTCAGCGTGAAGAAGAACGCCGATTGTCTG

GACACCACAAAGGACCTGAATGCCGAAGTGATGGAGCAGTAT

AGCGAGTCCAACAGACTGATCCTGATCAGGAATACCACAGATA

TCCTGTACTATCTGGTGCTGGACAAGAATGGCAAGGTGCTGAA

GCAGAGATCCCTGAACATCATCAATGACGGCGCCAGGGATGTG

GACTGGAAGGAGAGGTTCCGCCAGGTGACAAAGGATAGAAA

CGAGGGCTACAATGAGTGGGATTATTCCAGGACCTCTAACGAC

CTGAAGGAGGTGTACCTGAATTATGCCCTGAAGGAGATCGCCG

AGGCCGTGATCGAGTACAACGCCATCCTGATCATCGAGAAGAT

GTCTAATGCCTTTAAGGACAAGTATAGCTTCCTGGACGACGTG

ACCTTCAAGGGCTTCGAGACAAAGCTGCTGGCCAAGCTGAGC

GATCTGCACTTTAGGGGCATCAAGGACGGCGAGCCATGTTCCT

TCACAAACCCCCTGCAGCTGTGCCAGAACGATTCTAATAAGAT

CCTGCAGGACGGCGTGATCTTTATGGTGCCAAATTCTATGACA

CGGAGCCTGGACCCCGACACCGGCTTCATCTTTGCCATCAACG

ACCACAATATCAGGACCAAGAAGGCCAAGCTGAACTTTCTGA

GCAAGTTCGATCAGCTGAAGGTGTCCTCTGAGGGCTGCCTGAT

CATGAAGTACAGCGGCGATTCCCTGCCTACACACAACACCGAC

AATCGCGTGTGGAACTGCTGTTGCAATCACCCAATCACAAACT
ATGACCGGGAGACAAAGAAGGTGGAGTTCATCGAGGAGCCCG
TGGAGGAGCTGTCCCGCGTGCTGGAGGAGAATGGCATCGAGA
CAGACACCGAGCTGAACAAGCTGAATGAGCGGGAGAACGTG
CCTGGCAAGGTGGTGGATGCCATCTACTCTCTGGTGCTGAATT
ATCTGCGCGGCACAGTGAGCGGAGTGGCAGGACAGAGGGCC
GTGTACTATAGCCCTGTGACCGGCAAGAAGTACGATATCTCCT
TTATCCAGGCCATGAACCTGAATAGGAAGTGTGACTACTATAG
GATCGGCTCCAAGGAGAGGGGAGAGTGGACCGATTTCGTGGC
CCAGCTGATCAAC*AAAAGGCCGGCGGCCACGAAAAAGGCCGGC*
*CAGGCAAAAAAGAAAAA*GGGATCC**TACCCATACGATGTTCCAG**
ATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCAT
ATGATGTCCCCGACTATGCC（SEQ ID NO: 215）

[1674]　4-解朊丁酸弧菌（BpCpf1）

[1675]

ATGAGCATCTACCAGGAGTTCGTCAACAAGTATTCACTGA
GTAAGACACTGCGGTTCGAGCTGATCCCACAGGGCAAGACAC
TGGAGAACATCAAGGCCCGAGGCCTGATTCTGGACGATGAGA
AGCGGGCAAAAGACTATAAGAAAGCCAAGCAGATCATTGATA
AATACCACCAGTTCTTTATCGAGGAAATTCTGAGCTCCGTGTG
CATCAGTGAGGATCTGCTGCAGAATTACTCAGACGTGTACTTC
AAGCTGAAGAAGAGCGACGATGACAACCTGCAGAAGGACTTC
AAGTCCGCCAAGGACACCATCAAGAAACAGATTAGCGAGTAC

ATCAAGGACTCCGAAAAGTTTAAAAATCTGTTCAACCAGAATC

TGATCGATGCTAAGAAAGGCCAGGAGTCCGACCTGATCCTGTG

GCTGAAACAGTCTAAGGACAATGGGATTGAACTGTTCAAGGC

TAACTCCGATATCACTGATATTGACGAGGCACTGGAAATCATC

AAGAGCTTCAAGGGATGGACCACATACTTTAAAGGCTTCCACG

AGAACCGCAAGAACGTGTACTCCAGCAACGACATTCCTACCT

CCATCATCTACCGAATCGTCGATGACAATCTGCCAAAGTTCCT

GGAGAACAAGGCCAAATATGAATCTCTGAAGGACAAAGCTCC

CGAGGCAATTAATTACGAACAGATCAAGAAAGATCTGGCTGA

GGAACTGACATTCGATATCGACTATAAGACTAGCGAGGTGAAC

CAGAGGGTCTTTTCCCTGGACGAGGTGTTTGAAATCGCCAATT

TCAACAATTACCTGAACCAGTCCGGCATTACTAAATTCAATAC

CATCATTGGCGGGAAGTTTGTGAACGGGGAGAATACCAAGCG

CAAGGGAATTAACGAATACATCAATCTGTATAGCCAGCAGATC

AACGACAAAACTCTGAAGAAATACAAGATGTCTGTGCTGTTC

AAACAGATCCTGAGTGATACCGAGTCCAAGTCTTTTGTCATTG

ATAAACTGGAAGATGACTCAGACGTGGTCACTACCATGCAGA

GCTTTTATGAGCAGATCGCCGCTTTCAAGACAGTGGAGGAAA

AATCTATTAAGGAAACTCTGAGTCTGCTGTTCGATGACCTGAA

AGCCCAGAAGCTGGACCTGAGTAAGATCTACTTCAAAAACGA

TAAGAGTCTGACAGACCTGTCACAGCAGGTGTTTGATGACTAT

TCCGTGATTGGGACCGCCGTCCTGGAGTACATTACACAGCAGA

TCGCTCCAAAGAACCTGGATAATCCCTCTAAGAAAGAGCAGG

AACTGATCGCTAAGAAAACCGAGAAGGCAAAATATCTGAGTC

TGGAAACAATTAAGCTGGCACTGGAGGAGTTCAACAAGCACA

GGGATATTGACAAACAGTGCCGCTTTGAGGAAATCCTGGCCA

ACTTCGCAGCCATCCCCATGATTTTTGATGAGATCGCCCAGAA

CAAAGACAATCTGGCTCAGATCAGTATTAAGTACCAGAACCAG

GGCAAGAAAGACCTGCTGCAGGCTTCAGCAGAAGATGACGTG

AAAGCCATCAAGGATCTGCTGGACCAGACCAACAATCTGCTG

CACAAGCTGAAAATCTTCCATATTAGTCAGTCAGAGGATAAGG

CTAATATCCTGGATAAAGACGAACACTTCTACCTGGTGTTCGA

GGAATGTTACTTCGAGCTGGCAAACATTGTCCCCCTGTATAAC

AAGATTAGGAACTACATCACACAGAAGCCTTACTCTGACGAG

AAGTTTAAACTGAACTTCGAAAATAGTACCCTGGCCAACGGGT

GGGATAAGAACAAGGAGCCTGACAACACAGCTATCCTGTTCAT

CAAGGATGACAAGTACTATCTGGGAGTGATGAATAAGAAAAA

CAATAAGATCTTCGATGACAAAGCCATTAAGGAGAACAAAGG

GGAAGGATACAAGAAAATCGTGTATAAGCTGCTGCCCGGCGC

AAATAAGATGCTGCCTAAGGTGTTCTTCAGCGCCAAGAGTATC

AAATTCTACAACCCATCCGAGGACATCCTGCGGATTAGAAATC

ACTCAACACATACTAAGAACGGGAGCCCCCAGAAGGGATATG

AGAAATTTGAGTTCAACATCGAGGATTGCAGGAAGTTTATTGA

CTTCTACAAGCAGAGCATCTCCAAACACCCTGAATGGAAGGAT

TTTGGCTTCCGGTTTTCCGACACACAGAGATATAACTCTATCGA

CGAGTTCTACCGCGAGGTGGAAAATCAGGGGTATAAGCTGAC

TTTTGAGAACATTTCTGAAAGTTACATCGACAGCGTGGTCAAT

CAGGGAAAGCTGTACCTGTTCCAGATCTATAACAAAGATTTTT

CAGCATACAGCAAGGGCAGACCAAACCTGCATACACTGTACT

GGAAGGCCCTGTTCGATGAGAGGAATCTGCAGGACGTGGTCT

ATAAACTGAACGGAGAGGCCGAACTGTTTTACCGGAAGCAGT

CTATTCCTAAGAAAATCACTCACCCAGCTAAGGAGGCCATCGC

TAACAAGAACAAGGACAATCCTAAGAAAGAGAGCGTGTTCGA

ATACGATCTGATTAAGGACAAGCGGTTCACCGAAGATAAGTTC

TTTTTCCATTGTCCAATCACCATTAACTTCAAGTCAAGCGGCG

CTAACAAGTTCAACGACGAGATCAATCTGCTGCTGAAGGAAA

AAGCAAACGATGTGCACATCCTGAGCATTGACCGAGGAGAGC

GGCATCTGGCCTACTATACCCTGGTGGATGGCAAAGGGAATAT

CATTAAGCAGGATACATTCAACATCATTGGCAATGACCGGATG

AAAACCAACTACCACGATAAACTGGCTGCAATCGAGAAGGAT

AGAGACTCAGCTAGGAAGGACTGGAAGAAAATCAACAACATT

AAGGAGATGAAGGAAGGCTATCTGAGCCAGGTGGTCCATGAG

ATTGCAAAGCTGGTCATCGAATACAATGCCATTGTGGTGTTCG

AGGATCTGAACTTCGGCTTTAAGAGGGGGCGCTTTAAGGTGG

AAAAACAGGTCTATCAGAAGCTGGAGAAAATGCTGATCGAAA

AGCTGAATTACCTGGTGTTTAAAGATAACGAGTTCGACAAGAC

CGGAGGCGTCCTGAGAGCCTACCAGCTGACAGCTCCCTTTGA

AACTTTCAAGAAAATGGGAAAACAGACAGGCATCATCTACTAT

GTGCCAGCCGGATTCACTTCCAAGATCTGCCCCGTGACCGGCT

TTGTCAACCAGCTGTACCCTAAATATGAGTCAGTGAGCAAGTC

CCAGGAATTTTTCAGCAAGTTCGATAAGATCTGTTATAATCTGG

ACAAGGGGTACTTCGAGTTTTCCTTCGATTACAAGAACTTCGG

CGACAAGGCCGCTAAGGGGAAATGGACCATTGCCTCCTTCGG
ATCTCGCCTGATCAACTTTCGAAATTCCGATAAAAACCACAAT
TGGGACACTAGGGAGGTGTACCCAACCAAGGAGCTGGAAAAG
CTGCTGAAAGACTACTCTATCGAGTATGGACATGGCGAATGCA
TCAAGGCAGCCATCTGTGGCGAGAGTGATAAGAAATTTTTCGC
CAAGCTGACCTCAGTGCTGAATACAATCCTGCAGATGCGGAAC
TCAAAGACCGGGACAGAACTGGACTATCTGATTAGCCCCGTG
GCTGATGTCAACGGAAACTTCTTCGACAGCAGACAGGCACCC
AAAAATATGCCTCAGGATGCAGACGCCAACGGGGCCTACCAC
ATCGGGCTGAAGGGACTGATGCTGCTGGGCCGGATCAAGAAC
AATCAGGAGGGGAAGAAGCTGAACCTGGTCATTAAGAACGAG
GAATACTTCGAGTTTGTCCAGAATAGAAATAAC*AAAAGGCCGG*
*CGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAG*GGATCC
TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGT
GCCTGATTATGCATACCCATATGATGTCCCCGACTATGCC
（SEQ ID NO: 216）

[1676] 5-佩萊格裡尼菌科細菌GW2011_GWA_33_10（PeCpf1）

[1677]

ATGTCCAACTTCTTTAAGAATTTCACCAACCTGTATGAGCT
GTCCAAGACACTGAGGTTTGAGCTGAAGCCCGTGGGCGACAC

CCTGACAAACATGAAGGACCACCTGGAGTACGATGAGAAGCT
GCAGACCTTCCTGAAGGATCAGAATATCGACGATGCCTATCAG
GCCCTGAAGCCTCAGTTCGACGAGATCCACGAGGAGTTTATCA
CAGATTCTCTGGAGAGCAAGAAGGCCAAGGAGATCGACTTCT
CCGAGTACCTGGATCTGTTTCAGGAGAAGAAGGAGCTGAACG
ACTCTGAGAAGAAGCTGCGCAACAAGATCGGCGAGACATTCA
ACAAGGCCGGCGAGAAGTGGAAGAAGGAGAAGTACCCTCAG
TATGAGTGGAAGAAGGGCTCCAAGATCGCCAATGGCGCCGAC
ATCCTGTCTTGCCAGGATATGCTGCAGTTTATCAAGTATAAGAA
CCCAGAGGATGAGAAGATCAAGAATTACATCGACGATACACTG
AAGGGCTTCTTTACCTATTTCGGCGGCTTTAATCAGAACAGGG
CCAACTACTATGAGACAAAGAAGGAGGCCTCCACCGCAGTGG
CAACAAGGATCGTGCACGAGAACCTGCCAAAGTTCTGTGACA
ATGTGATCCAGTTTAAGCACATCATCAAGCGGAAGAAGGATGG
CACCGTGGAGAAAACCGAGAGAAAGACCGAGTACCTGAACG
CCTACCAGTATCTGAAGAACAATAACAAGATCACACAGATCAA
GGACGCCGAGACAGAGAAGATGATCGAGTCTACACCCATCGC
CGAGAAGATCTTCGACGTGTACTACTTCAGCAGCTGCCTGAGC
CAGAAGCAGATCGAGGAGTACAACCGGATCATCGGCCACTATA
ATCTGCTGATCAACCTGTATAACCAGGCCAAGAGATCTGAGGG
CAAGCACCTGAGCGCCAACGAGAAGAAGTATAAGGACCTGCC
TAAGTTCAAGACCCTGTATAAGCAGATCGGCTGCGGCAAGAA
GAAGGACCTGTTTTACACAATCAAGTGTGATACCGAGGAGGA
GGCCAATAAGTCCCGGAACGAGGGCAAGGAGTCCCACTCTGT
GGAGGAGATCATCAACAAGGCCCAGGAGGCCATCAATAAGTA
CTTCAAGTCTAATAACGACTGTGAGAATATCAACACCGTGCCC

GACTTCATCAACTATATCCTGACAAAGGAGAATTACGAGGGCG

TGTATTGGAGCAAGGCCGCCATGAACACCATCTCCGACAAGTA

CTTCGCCAATTATCACGACCTGCAGGATAGACTGAAGGAGGCC

AAGGTGTTTCAGAAGGCCGATAAGAAGTCCGAGGACGATATC

AAGATCCCAGAGGCCATCGAGCTGTCTGGCCTGTTCGGCGTGC

TGGACAGCCTGGCCGATTGGCAGACCACACTGTTTAAGTCTAG

CATCCTGAGCAACGAGGACAAGCTGAAGATCATCACAGATTC

CCAGACCCCCTCTGAGGCCCTGCTGAAGATGATCTTCAATGAC

ATCGAGAAGAACATGGAGTCCTTTCTGAAGGAGACAAACGAT

ATCATCACCCTGAAGAAGTATAAGGGCAATAAGGAGGGCACC

GAGAAGATCAAGCAGTGGTTCGACTATACACTGGCCATCAACC

GGATGCTGAAGTACTTTCTGGTGAAGGAGAATAAGATCAAGG

GCAACTCCCTGGATACCAATATCTCTGAGGCCCTGAAAACCCT

GATCTACAGCGACGATGCCGAGTGGTTCAAGTGGTACGACGC

CCTGAGAAACTATCTGACCCAGAAGCCTCAGGATGAGGCCAA

GGAGAATAAGCTGAAGCTGAATTTCGACAACCCATCTCTGGCC

GGCGGCTGGGATGTGAACAAGGAGTGCAGCAATTTTTGCGTG

ATCCTGAAGGACAAGAACGAGAAGAAGTACCTGGCCATCATG

AAGAAGGGCGAGAATACCCTGTTCCAGAAGGAGTGGACAGA

GGGCCGGGGCAAGAACCTGACAAAGAAGTCTAATCCACTGTT

CGAGATCAATAACTGCGAGATCCTGAGCAAGATGGAGTATGAC

TTTTGGGCCGACGTGAGCAAGATGATCCCCAAGTGTAGCACCC

AGCTGAAGGCCGTGGTGAACCACTTCAAGCAGTCCGACAATG

AGTTCATCTTTCCTATCGGCTACAAGGTGACAAGCGGCGAGAA

GTTTAGGGAGGAGTGCAAGATCTCCAAGCAGGACTTCGAGCT

GAATAACAAGGTGTTTAATAAGAACGAGCTGAGCGTGACCGC

CATGCGCTACGATCTGTCCTCTACACAGGAGAAGCAGTATATC

AAGGCCTTCCAGAAGGAGTACTGGGAGCTGCTGTTTAAGCAG

GAGAAGCGGGACACCAAGCTGACAAATAACGAGATCTTCAAC

GAGTGGATCAATTTTTGCAACAAGAAGTATAGCGAGCTGCTGT

CCTGGGAGAGAAAGTACAAGGATGCCCTGACCAATTGGATCA

ACTTCTGTAAGTACTTTCTGAGCAAGTATCCCAAGACCACACT

GTTCAACTACTCTTTTAAGGAGAGCGAGAATTATAACTCCCTG

GACGAGTTCTACCGGGACGTGGATATCTGTTCTTACAAGCTGA

ATATCAACACCACAATCAATAAGAGCATCCTGGATAGACTGGT

GGAGGAGGGCAAGCTGTACCTGTTTGAGATCAAGAATCAGGA

CAGCAACGATGGCAAGTCCATCGGCCACAAGAATAACCTGCA

CACCATCTACTGGAACGCCATCTTCGAGAATTTTGACAACAGG

CCTAAGCTGAATGGCGAGGCCGAGATCTTCTATCGCAAGGCCA

TCTCCAAGGATAAGCTGGGCATCGTGAAGGGCAAGAAAACCA

AGAACGGCACCGAGATCATCAAGAATTACAGATTCAGCAAGG

AGAAGTTTATCCTGCACGTGCCAATCACCCTGAACTTCTGCTC

CAATAACGAGTATGTGAATGACATCGTGAACACAAAGTTCTAC

AATTTTTCCAACCTGCACTTTCTGGGCATCGATAGGGGCGAGA

AGCACCTGGCCTACTATTCTCTGGTGAATAAGAACGGCGAGAT

CGTGGACCAGGGCACACTGAACCTGCCTTTCACCGACAAGGA

TGGCAATCAGCGCAGCATCAAGAAGGAGAAGTACTTTTATAAC

AAGCAGGAGGACAAGTGGGAGGCCAAGGAGGTGGATTGTTG

GAATTATAACGACCTGCTGGATGCCATGGCCTCTAACCGGGAC

ATGGCCAGAAAGAATTGGCAGAGGATCGGCACCATCAAGGAG

GCCAAGAACGGCTACGTGAGCCTGGTCATCAGGAAGATCGCC

GATCTGGCCGTGAATAACGAGCGCCCCGCCTTCATCGTGCTGG

AGGACCTGAATACAGGCTTTAAGCGGTCCAGACAGAAGATCG

ATAAGAGCGTGTACCAGAAGTTCGAGCTGGCCCTGGCCAAGA

AGCTGAACTTTCTGGTGGACAAGAATGCCAAGCGCGATGAGA

TCGGCTCCCCTACAAAGGCCCTGCAGCTGACCCCCCCTGTGA

ATAACTACGGCGACATTGAGAACAAGAAGCAGGCCGGCATCA

TGCTGTATACCCGGGCCAATTATACCTCTCAGACAGATCCAGC

CACAGGCTGGAGAAAGACCATCTATCTGAAGGCCGGCCCCGA

GGAGACAACATACAAGAAGGACGGCAAGATCAAGAACAAGA

GCGTGAAGGACCAGATCATCGAGACATTCACCGATATCGGCTT

TGACGGCAAGGATTACTATTTCGAGTACGACAAGGGCGAGTTT

GTGGATGAGAAAACCGGCGAGATCAAGCCCAAGAAGTGGCG

GCTGTACTCCGGCGAGAATGGCAAGTCCCTGGACAGGTTCCG

CGGAGAGAGGGAGAAGGATAAGTATGAGTGGAAGATCGACAA

GATCGATATCGTGAAGATCCTGGACGATCTGTTCGTGAATTT

TGACAAGAACATCAGCCTGCTGAAGCAGCTGAAGGAGGGC

GTGGAGCTGACCCGGAATAACGAGCACGGCACAGGCGAGTC

CCTGAGATTCGCCATCAACCTGATCCAGCAGATCCGGAATAC

CGGCAATAACGAGAGAGACAACGATTTCATCCTGTCCCCAG

TGAGGGACGAGAATGGCAAGCACTTTGACTCTCGCGAGTAC

TGGGATAAGGAGACAAAGGGCGAGAAGATCAGCATGCCCAG

CTCCGGCGATGCCAATGGCGCCTTCAACATCGCCCGGAAGG

GCATCATCATGAACGCCCACATCCTGGCCAATAGCGACTCCA

AGGATCTGTCCCTGTTCGTGTCTGACGAGGAGTGGGATCTGC

ACCTGAATAACAAGACCGAGTGGAAGAAGCAGCTGAACATC

TTTTCTAGCAGGAAGGCCATGGCCAAGCGCAAGAAG*AAAAG*

GCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAA

*G*GGATCC**TACCCATACGATGTTCCAGATTACGCTTATCCCTA CGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTA TGCCTAAGAATTC（SEQ ID NO: 217）**

[1678] 6-儉菌總門細菌GWC2011_GWC2_44_17（PbCpf1）

[1679]

ATGGAGAACATCTTCGACCAGTTTATCGGCAAGTACAGCC TGTCCAAGACCCTGAGATTCGAGCTGAAGCCCGTGGGCAAGA CAGAGGACTTCCTGAAGATCAACAAGGTGTTTGAGAAGGATC AGACCATCGACGATAGCTACAATCAGGCCAAGTTCTATTTTGA TTCCCTGCACCAGAAGTTTATCGACGCCGCCCTGGCCTCCGAT AAGACATCCGAGCTGTCTTTCCAGAACTTTGCCGACGTGCTGG AGAAGCAGAATAAGATCATCCTGGATAAGAAGAGAGAGATGG GCGCCCTGAGGAAGCGCGACAAGAACGCCGTGGGCATCGATA GGCTGCAGAAGGAGATCAATGACGCCGAGGATATCATCCAGA AGGAGAAGGAGAAGATCTACAAGGACGTGCGCACCCTGTTCG ATAACGAGGCCGAGTCTTGGAAAACCTACTATCAGGAGCGGG AGGTGGACGGCAAGAAGATCACCTTCAGCAAGGCCGACCTGA AGCAGAAGGGCGCCGATTTTCTGACAGCCGCCGGCATCCTGA AGGTGCTGAAGTATGAGTTCCCCGAGGAGAAGGAGAAGGAGT TTCAGGCCAAGAACCAGCCCTCCCTGTTCGTGGAGGAGAAGG AGAATCCTGGCCAGAAGAGGTACATCTTCGACTCTTTTGATAA GTTCGCCGGCTATCTGACCAAGTTTCAGCAGACAAAGAAGAA

TCTGTACGCAGCAGACGGCACCAGCACAGCAGTGGCCACCCG

CATCGCCGATAACTTTATCATCTTCCACCAGAATACCAAGGTGT

TCCGGGACAAGTACAAGAACAATCACACAGACCTGGGCTTCG

ATGAGGAGAACATCTTTGAGATCGAGAGGTATAAGAATTGCCT

GCTGCAGCGCGAGATCGAGCACATCAAGAATGAGAATAGCTA

CAACAAGATCATCGGCCGGATCAATAAGAAGATCAAGGAGTAT

CGGGACCAGAAGGCCAAGGATACCAAGCTGACAAAGTCCGAC

TTCCCTTTCTTTAAGAACCTGGATAAGCAGATCCTGGGCGAGG

TGGAGAAGGAGAAGCAGCTGATCGAGAAAACCCGGGAGAAA

ACCGAGGAGGACGTGCTGATCGAGCGGTTCAAGGAGTTCATC

GAGAACAATGAGGAGAGGTTCACCGCCGCCAAGAAGCTGATG

AATGCCTTCTGTAACGGCGAGTTTGAGTCCGAGTACGAGGGC

ATCTATCTGAAGAATAAGGCCATCAACACAATCTCCCGGAGAT

GGTTCGTGTCTGACAGAGATTTTGAGCTGAAGCTGCCTCAGCA

GAAGTCCAAGAACAAGTCTGAGAAGAATGAGCCAAAGGTGA

AGAAGTTCATCTCCATCGCCGAGATCAAGAACGCCGTGGAGG

AGCTGGACGGCGATATCTTTAAGGCCGTGTTCTACGACAAGAA

GATCATCGCCCAGGGCGGCTCTAAGCTGGAGCAGTTCCTGGTC

ATCTGGAAGTACGAGTTTGAGTATCTGTTCCGGGACATCGAGA

GAGAGAACGGCGAGAAGCTGCTGGGCTATGATAGCTGCCTGA

AGATCGCCAAGCAGCTGGGCATCTTCCCACAGGAGAAGGAGG

CCCGCGAGAAGGCAACCGCCGTGATCAAGAATTACGCCGACG

CCGGCCTGGGCATCTTCCAGATGATGAAGTATTTTTCTCTGGA

CGATAAGGATCGGAAGAACACCCCCGGCCAGCTGAGCACAAA

TTTCTACGCCGAGTATGACGGCTACTACAAGGATTTCGAGTTT

ATCAAGTACTACAACGAGTTTAGGAACTTCATCACCAAGAAGC

CTTTCGACGAGGATAAGATCAAGCTGAACTTTGAGAATGGCGC

CCTGCTGAAGGGCTGGGACGAGAACAAGGAGTACGATTTCAT

GGGCGTGATCCTGAAGAAGGAGGGCCGCCTGTATCTGGGCAT

CATGCACAAGAACCACCGGAAGCTGTTTCAGTCCATGGGCAA

TGCCAAGGGCGACAACGCCAATAGATACCAGAAGATGATCTAT

AAGCAGATCGCCGACGCCTCTAAGGATGTGCCCAGGCTGCTG

CTGACCAGCAAGAAGGCCATGGAGAAGTTCAAGCCTTCCCAG

GAGATCCTGAGAATCAAGAAGGAGAAAACCTTCAAGCGGGAG

AGCAAGAACTTTTCCCTGAGAGATCTGCACGCCCTGATCGAGT

ACTATAGGAACTGCATCCCTCAGTACAGCAATTGGTCCTTTTAT

GACTTCCAGTTTCAGGATACCGGCAAGTACCAGAATATCAAGG

AGTTCACAGACGATGTGCAGAAGTACGGCTATAAGATCTCCTT

TCGCGACATCGACGATGAGTATATCAATCAGGCCCTGAACGAG

GGCAAGATGTACCTGTTCGAGGTGGTGAACAAGGATATCTATA

ACACCAAGAATGGCTCCAAGAATCTGCACACACTGTACTTTGA

GCACATCCTGTCTGCCGAGAACCTGAATGACCCAGTGTTCAAG

CTGTCTGGCATGGCCGAGATCTTTCAGCGGCAGCCCAGCGTGA

ACGAAAGAGAGAAGATCACCACACAGAAGAATCAGTGTATCC

TGGACAAGGGCGATAGAGCCTACAAGTATAGGCGCTACACCG

AGAAGAAGATCATGTTCCACATGAGCCTGGTGCTGAACACAG

GCAAGGGCGAGATCAAGCAGGTGCAGTTTAATAAGATCATCA

ACCAGAGGATCAGCTCCTCTGACAACGAGATGAGGGTGAATG

TGATCGGCATCGATCGCGGCGAGAAGAACCTGCTGTACTATAG

CGTGGTGAAGCAGAATGGCGAGATCATCGAGCAGGCCTCCCT

GAACGAGATCAATGGCGTGAACTACCGGGACAAGCTGATCGA

GAGGGAGAAGGAGCGCCTGAAGAACCGGCAGAGCTGGAAGC

CTGTGGTGAAGATCAAGGATCTGAAGAAGGGCTACATCTCCC

ACGTGATCCACAAGATCTGCCAGCTGATCGAGAAGTATTCTGC

CATCGTGGTGCTGGAGGACCTGAATATGAGATTCAAGCAGATC

AGGGGAGGAATCGAGCGGAGCGTGTACCAGCAGTTCGAGAA

GGCCCTGATCGATAAGCTGGGCTATCTGGTGTTTAAGGACAAC

AGGGATCTGAGGGCACCAGGAGGCGTGCTGAATGGCTACCAG

CTGTCTGCCCCCTTTGTGAGCTTCGAGAAGATGCGCAAGCAG

ACCGGCATCCTGTTCTACACACAGGCCGAGTATACCAGCAAGA

CAGACCCAATCACCGGCTTTCGGAAGAACGTGTATATCTCTAA

TAGCGCCTCCCTGGATAAGATCAAGGAGGCCGTGAAGAAGTT

CGACGCCATCGGCTGGGATGGCAAGGAGCAGTCTTACTTCTTT

AAGTACAACCCTTACAACCTGGCCGACGAGAAGTATAAGAAC

TCTACCGTGAGCAAGGAGTGGGCCATCTTTGCCAGCGCCCCA

AGAATCCGGAGACAGAAGGGCGAGGACGGCTACTGGAAGTAT

GATAGGGTGAAAGTGAATGAGGAGTTCGAGAAGCTGCTGAAG

GTCTGGAATTTTGTGAACCCAAAGGCCACAGATATCAAGCAG

GAGATCATCAAGAAGGAGAAGGCAGGCGACCTGCAGGGAGA

GAAGGAGCTGGATGGCCGGCTGAGAAACTTTTGGCACTCTTT

CATCTACCTGTTTAACCTGGTGCTGGAGCTGCGCAATTCTTTC

AGCCTGCAGATCAAGATCAAGGCAGGAGAAGTGATCGCAGTG

GACGAGGGCGTGGACTTCATCGCCAGCCCAGTGAAGCCCTTC
TTTACCACACCCAACCCTTACATCCCCTCCAACCTGTGCTGGC
TGGCCGTGGAGAATGCAGACGCAAACGGAGCCTATAATATCGC
CAGGAAGGGCGTGATGATCCTGAAGAAGATCCGCGAGCACGC
CAAGAAGGACCCCGAGTTCAAGAAGCTGCCAAACCTGTTTAT
CAGCAATGCAGAGTGGGACGAGGCAGCCCGGGATTGGGGCAA
GTACGCAGGCACCACAGCCCTGAACCTGGACCAC*AAAAGGCC*
*GGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAG*GGA
TCC**TACCCATACGATGTTCCAGATTACGCTTATCCCTACGAC**
GTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGC
CTAAGAATTC（SEQ ID NO: 218）

[1680]　7-密斯氏菌屬某種SC_K08D17（SsCpf1）

[1681]

ATGCAGACCCTGTTTGAGAACTTCACAAATCAGTACCCAG
TGTCCAAGACCCTGCGCTTTGAGCTGATCCCCCAGGGCAAGA
CAAAGGACTTCATCGAGCAGAAGGGCCTGCTGAAGAAGGATG
AGGACCGGGCCGAGAAGTATAAGAAGGTGAAGAACATCATCG
ATGAGTACCACAAGGACTTCATCGAGAAGTCTCTGAATGGCCT
GAAGCTGGACGGCCTGGAGAAGTACAAGACCCTGTATCTGAA
GCAGGAGAAGGACGATAAGGATAAGAAGGCCTTTGACAAGGA
GAAGGAGAACCTGCGCAAGCAGATCGCCAATGCCTTCCGGAA
CAATGAGAAGTTTAAGACACTGTTCGCCAAGGAGCTGATCAA
GAACGATCTGATGTCTTTCGCCTGCGAGGAGGACAAGAAGAA

TGTGAAGGAGTTTGAGGCCTTCACCACATACTTCACCGGCTTC

CACCAGAACCGCGCCAATATGTACGTGGCCGATGAGAAGAGA

ACAGCCATCGCCAGCAGGCTGATCCACGAGAACCTGCCAAAG

TTTATCGACAATATCAAGATCTTCGAGAAGATGAAGAAGGAGG

CCCCCGAGCTGCTGTCTCCTTTCAACCAGACCCTGAAGGATAT

GAAGGACGTGATCAAGGGCACCACACTGGAGGAGATCTTTAG

CCTGGATTATTTCAACAAGACCCTGACACAGAGCGGCATCGAC

ATCTACAATTCCGTGATCGGCGGCAGAACCCCTGAGGAGGGC

AAGACAAAGATCAAGGGCCTGAACGAGTACATCAATACCGAC

TTCAACCAGAAGCAGACAGACAAGAAGAAGCGGCAGCCAAA

GTTCAAGCAGCTGTATAAGCAGATCCTGAGCGATAGGCAGAGC

CTGTCCTTTATCGCCGAGGCCTTCAAGAACGACACCGAGATCC

TGGAGGCCATCGAGAAGTTTTACGTGAATGAGCTGCTGCACTT

CAGCAATGAGGGCAAGTCCACAAACGTGCTGGACGCCATCAA

GAATGCCGTGTCTAACCTGGAGAGCTTTAACCTGACCAAGATG

TATTTCCGCTCCGGCGCCTCTCTGACAGACGTGAGCCGGAAGG

TGTTTGGCGAGTGGAGCATCATCAATAGAGCCCTGGACAACTA

CTATGCCACCACATATCCAATCAAGCCCAGAGAGAAGTCTGAG

AAGTACGAGGAGAGGAAGGAGAAGTGGCTGAAGCAGGACTT

CAACGTGAGCCTGATCCAGACCGCCATCGATGAGTACGACAA

CGAGACAGTGAAGGGCAAGAACAGCGGCAAAGTGATCGCCG

ATTATTTTGCCAAGTTCTGCGACGATAAGGAGACAGACCTGAT

CCAGAAGGTGAACGAGGGCTACATCGCCGTGAAGGATCTGCT

GAATACACCCTGTCCTGAGAACGAGAAGCTGGGCAGCAATAA

GGACCAGGTGAAGCAGATCAAGGCCTTTATGGATTCTATCATG

GACATCATGCACTTCGTGCGCCCCCTGAGCCTGAAGGATACCG

ACAAGGAGAAGGATGAGACATTCTACTCCCTGTTCACACCTCT

GTACGACCACCTGACCCAGACAATCGCCCTGTATAACAAGGTG

CGGAACTATCTGACCCAGAAGCCTTACAGCACAGAGAAGATC

AAGCTGAACTTCGAGAACAGCACCCTGCTGGGCGGCTGGGAT

CTGAATAAGGAGACAGACAACACAGCCATCATCCTGAGGAAG

GATAACCTGTACTATCTGGGCATCATGGACAAGAGGCACAATC

GCATCTTTCGGAACGTGCCCAAGGCCGATAAGAAGGACTTCT

GCTACGAGAAGATGGTGTATAAGCTGCTGCCTGGCGCCAACA

AGATGCTGCCAAAGGTGTTCTTTTCTCAGAGCAGAATCCAGGA

GTTTACCCCTTCCGCCAAGCTGCTGGAGAACTACGCCAATGAG

ACACACAAGAAGGGCGATAATTTCAACCTGAATCACTGTCAC

AAGCTGATCGATTTCTTTAAGGACTCTATCAACAAGCACGAGG

ATTGGAAGAATTTCGACTTTAGGTTCAGCGCCACCTCCACCTA

CGCCGACCTGAGCGGCTTTTACCACGAGGTGGAGCACCAGGG

CTACAAGATCTCTTTTCAGAGCGTGGCCGATTCCTTCATCGAC

GATCTGGTGAACGAGGGCAAGCTGTACCTGTTCCAGATCTATA

ATAAGGACTTTTCCCCATTCTCTAAGGGCAAGCCCAACCTGCA

CACCCTGTACTGGAAGATGCTGTTTGATGAGAACAATCTGAAG

GACGTGGTGTATAAGCTGAATGGCGAGGCCGAGGTGTTCTACC

GCAAGAAGAGCATTGCCGAGAAGAACACCACAATCCACAAGG

CCAATGAGTCCATCATCAACAAGAATCCTGATAACCCAAAGGC

CACCAGCACCTTCAACTATGATATCGTGAAGGACAAGAGATAC

ACCATCGACAAGTTTCAGTTCCACATCCCAATCACAATGAACT

TTAAGGCCGAGGGCATCTTCAACATGAATCAGAGGGTGAATCA

GTTCCTGAAGGCCAATCCCGATATCAACATCATCGGCATCGAC

AGAGGCGAGAGGCACCTGCTGTACTATGCCCTGATCAACCAG

AAGGGCAAGATCCTGAAGCAGGATACCCTGAATGTGATCGCC

AACGAGAAGCAGAAGGTGGACTACCACAATCTGCTGGATAAG

AAGGAGGGCGACCGCGCAACCGCAAGGCAGGAGTGGGGCGT

GATCGAGACAATCAAGGAGCTGAAGGAGGGCTATCTGTCCCA

GGTCATCCACAAGCTGACCGATCTGATGATCGAGAACAATGCC

ATCATCGTGATGGAGGACCTGAACTTTGGCTTCAAGCGGGGCA

GACAGAAGGTGGAGAAGCAGGTGTATCAGAAGTTTGAGAAG

ATGCTGATCGATAAGCTGAATTACCTGGTGGACAAGAATAAGA

AGGCAAACGAGCTGGGAGGCCTGCTGAACGCATTCCAGCTGG

CCAATAAGTTTGAGTCCTTCCAGAAGATGGGCAAGCAGAACG

GCTTTATCTTCTACGTGCCCGCCTGGAATACCTCTAAGACAGAT

CCTGCCACCGGCTTTATCGACTTCCTGAAGCCCCGCTATGAGA

ACCTGAATCAGGCCAAGGATTTCTTTGAGAAGTTTGACTCTAT

CCGGCTGAACAGCAAGGCCGATTACTTTGAGTTCGCCTTTGAC

TTCAAGAATTTCACCGAGAAGGCCGATGGCGGCAGAACCAAG

TGGACAGTGTGCACCACAAACGAGGACAGATATGCCTGGAAT

AGGGCCCTGAACAATAACAGGGGCAGCCAGGAGAAGTACGAC

ATCACAGCCGAGCTGAAGTCCCTGTTCGATGGCAAGGTGGAC

TATAAGTCTGGCAAGGATCTGAAGCAGCAGATCGCCAGCCAG

GAGTCCGCCGACTTCTTTAAGGCCCTGATGAAGAACCTGTCCA

TCACCCTGTCTCTGAGACACAATAACGGCGAGAAGGGCGATA
ATGAGCAGGACTACATCCTGTCCCCTGTGGCCGATTCTAAGGG
CCGCTTCTTTGACTCCCGGAAGGCCGACGATGACATGCCAAA
GAATGCCGACGCCAACGGCGCCTATCACATCGCCCTGAAGGG
CCTGTGGTGTCTGGAGCAGATCAGCAAGACCGATGACCTGAA
GAAGGTGAAGCTGGCCATCTCCAACAAGGAGTGGCTGGAGTT
CGTGCAGACACTGAAGGGC*AAAAGGCCGGCGGCCACGAAAAAG*
*GCCGGCCAGGCAAAAAAGAAAAAG*GGATCC**TACCCATACGATG**
TTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCA
TACCCATATGATGTCCCCGACTATGCCTAAGAATTC（SEQ ID
NO: 219）

[1682]　8-胺基酸球菌屬某種BV3L6（AsCpf1）

[1683]

ATGACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGA
GCAAGACACTGCGGTTTGAGCTGATCCCACAGGGCAAGACCC
TGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAGGACAAGG
CCCGCAATGATCACTACAAGGAGCTGAAGCCCATCATCGATCG
GATCTACAAGACCTATGCCGACCAGTGCCTGCAGCTGGTGCAG
CTGGATTGGGAGAACCTGAGCGCCGCCATCGACTCCTATAGAA
AGGAGAAAACCGAGGAGACAAGGAACGCCCTGATCGAGGAG
CAGGCCACATATCGCAATGCCATCCACGACTACTTCATCGGCC
GGACAGACAACCTGACCGATGCCATCAATAAGAGACACGCCG
AGATCTACAAGGGCCTGTTCAAGGCCGAGCTGTTTAATGGCAA

GGTGCTGAAGCAGCTGGGCACCGTGACCACAACCGAGCACGA

GAACGCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTC

TCCGGCTTTTATGAGAACAGGAAGAACGTGTTCAGCGCCGAG

GATATCAGCACAGCCATCCCACACCGCATCGTGCAGGACAACT

TCCCCAAGTTTAAGGAGAATTGTCACATCTTCACACGCCTGAT

CACCGCCGTGCCCAGCCTGCGGGAGCACTTTGAGAACGTGAA

GAAGGCCATCGGCATCTTCGTGAGCACCTCCATCGAGGAGGT

GTTTTCCTTCCCTTTTTATAACCAGCTGCTGACACAGACCCAG

ATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAGG

CAGGCACCGAGAAGATCAAGGGCCTGAACGAGGTGCTGAATC

TGGCCATCCAGAAGAATGATGAGACAGCCCACATCATCGCCTC

CCTGCCACACAGATTCATCCCCCTGTTTAAGCAGATCCTGTCC

GATAGGAACACCCTGTCTTTCATCCTGGAGGAGTTTAAGAGCG

ACGAGGAAGTGATCCAGTCCTTCTGCAAGTACAAGACACTGC

TGAGAAACGAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTA

ACGAGCTGAACAGCATCGACCTGACACACATCTTCATCAGCCA

CAAGAAGCTGGAGACAATCAGCAGCGCCCTGTGCGACCACTG

GGATACACTGAGGAATGCCCTGTATGAGCGGAGAATCTCCGAG

CTGACAGGCAAGATCACCAAGTCTGCCAAGGAGAAGGTGCAG

CGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCATCT

CTGCCGCAGGCAAGGAGCTGAGCGAGGCCTTCAAGCAGAAA

ACCAGCGAGATCCTGTCCCACGCACACGCCGCCCTGGATCAG

CCACTGCCTACAACCCTGAAGAAGCAGGAGGAGAAGGAGATC

CTGAAGTCTCAGCTGGACAGCCTGCTGGGCCTGTACCACCTG

CTGGACTGGTTTGCCGTGGATGAGTCCAACGAGGTGGACCCC

GAGTTCTCTGCCCGGCTGACCGGCATCAAGCTGGAGATGGAG

CCTTCTCTGAGCTTCTACAACAAGGCCAGAAATTATGCCACCA

AGAAGCCCTACTCCGTGGAGAAGTTCAAGCTGAACTTTCAGA

TGCCTACACTGGCCTCTGGCTGGGACGTGAATAAGGAGAAGA

ACAATGGCGCCATCCTGTTTGTGAAGAACGGCCTGTACTATCT

GGGCATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAG

CTTCGAGCCCACAGAGAAAACCAGCGAGGGCTTTGATAAGAT

GTACTATGACTACTTCCCTGATGCCGCCAAGATGATCCCAAAG

TGCAGCACCCAGCTGAAGGCCGTGACAGCCCACTTTCAGACC

CACACAACCCCCATCCTGCTGTCCAACAATTTCATCGAGCCTC

TGGAGATCACAAAGGAGATCTACGACCTGAACAATCCTGAGA

AGGAGCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAACC

GGCGACCAGAAGGGCTACAGAGAGGCCCTGTGCAAGTGGATC

GACTTCACAAGGGATTTTCTGTCCAAGTATACCAAGACAACCT

CTATCGATCTGTCTAGCCTGCGGCCATCCTCTCAGTATAAGGAC

CTGGGCGAGTACTATGCCGAGCTGAATCCCCTGCTGTACCACA

TCAGCTTCCAGAGAATCGCCGAGAAGGAGATCATGGATGCCG

TGGAGACAGGCAAGCTGTACCTGTTCCAGATCTATAACAAGGA

CTTTGCCAAGGGCCACCACGGCAAGCCTAATCTGCACACACT

GTATTGGACCGGCCTGTTTTCTCCAGAGAACCTGGCCAAGACA

AGCATCAAGCTGAATGGCCAGGCCGAGCTGTTCTACCGCCCTA

AGTCCAGGATGAAGAGGATGGCACACCGGCTGGGAGAGAAG

ATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCC

GACACCCTGTACCAGGAGCTGTACGACTATGTGAATCACAGAC

TGTCCCACGACCTGTCTGATGAGGCCAGGGCCCTGCTGCCCA

ACGTGATCACCAAGGAGGTGTCTCACGAGATCATCAAGGATA

GGCGCTTTACCAGCGACAAGTTCTTTTTCCACGTGCCTATCAC

ACTGAACTATCAGGCCGCCAATTCCCCATCTAAGTTCAACCAG

AGGGTGAATGCCTACCTGAAGGAGCACCCCGAGACACCTATC

ATCGGCATCGATCGGGGCGAGAGAAACCTGATCTATATCACAG

TGATCGACTCCACCGGCAAGATCCTGGAGCAGCGGAGCCTGA

ACACCATCCAGCAGTTTGATTACCAGAAGAAGCTGGACAACA

GGGAGAAGGAGAGGGTGGCAGCAAGGCAGGCCTGGTCTGTG

GTGGGCACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAG

GTCATCCACGAGATCGTGGACCTGATGATCCACTACCAGGCCG

TGGTGGTGCTGGAGAACCTGAATTTCGGCTTTAAGAGCAAGA

GGACCGGCATCGCCGAGAAGGCCGTGTACCAGCAGTTCGAGA

AGATGCTGATCGATAAGCTGAATTGCCTGGTGCTGAAGGACTA

TCCAGCAGAGAAAGTGGGAGGCGTGCTGAACCCATACCAGCT

GACAGACCAGTTCACCTCCTTTGCCAAGATGGGCACCCAGTCT

GGCTTCCTGTTTTACGTGCCTGCCCCATATACATCTAAGATCGA

TCCCCTGACCGGCTTCGTGGACCCCTTCGTGTGGAAAACCATC

AAGAATCACGAGAGCCGCAAGCACTTCCTGGAGGGCTTCGAC

TTTCTGCACTACGACGTGAAAACCGGCGACTTCATCCTGCACT

TTAAGATGAACAGAAATCTGTCCTTCCAGAGGGGCCTGCCCG

GCTTTATGCCTGCATGGGATATCGTGTTCGAGAAGAACGAGAC

ACAGTTTGACGCCAAGGGCACCCCTTTCATCGCCGGCAAGAG

AATCGTGCCAGTGATCGAGAATCACAGATTCACCGGCAGATAC
CGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTGCTGGAGG
AGAAGGGCATCGTGTTCAGGGATGGCTCCAACATCCTGCCAA
AGCTGCTGGAGAATGACGATTCTCACGCCATCGACACCATGGT
GGCCCTGATCCGCAGCGTGCTGCAGATGCGGAACTCCAATGCC
GCCACAGGCGAGGACTATATCAACAGCCCCGTGCGCGATCTGA
ATGGCGTGTGCTTCGACTCCCGGTTTCAGAACCCAGAGTGGCC
CATGGACGCCGATGCCAATGGCGCCTACCACATCGCCCTGAAG
GGCCAGCTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCTG
AAGCTGCAGAACGGCATCTCCAATCAGGACTGGCTGGCCTAC
ATCCAGGAGCTGCGCAAC*AAAAGGCCGGCGGCCACGAAAAAGG*
*CCGGCCAGGCAAAAAAGAAAAAG*GGATCC**TACCCATACGATGT**
TCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCAT
ACCCATATGATGTCCCCGACTATGCCTAAGAATTC（SEQ ID
NO: 220）

[1684]　9-毛螺菌科細菌MA2020（Lb2Cpf1）

[1685]

ATGTACTATGAGTCCCTGACCAAGCAGTACCCCGTGTCTA
AGACAATCCGGAATGAGCTGATCCCTATCGGCAAGACACTGG
ATAACATCCGCCAGAACAATATCCTGGAGAGCGACGTGAAGC
GGAAGCAGAACTACGAGCACGTGAAGGGCATCCTGGATGAGT
ATCACAAGCAGCTGATCAACGAGGCCCTGGACAATTGCACCC
TGCCATCCCTGAAGATCGCCGCCGAGATCTACCTGAAGAATCA

GAAGGAGGTGTCTGACAGAGAGGATTTCAACAAGACACAGG

ACCTGCTGAGGAAGGAGGTGGTGGAGAAGCTGAAGGCCCAC

GAGAACTTTACCAAGATCGGCAAGAAGGACATCCTGGATCTG

CTGGAGAAGCTGCCTTCCATCTCTGAGGACGATTACAATGCCC

TGGAGAGCTTCCGCAACTTTTACACCTATTTCACATCCTACAA

CAAGGTGCGGGAGAATCTGTATTCTGATAAGGAGAAGAGCTC

CACAGTGGCCTACAGACTGATCAACGAGAATTTCCCAAAGTTT

CTGGACAATGTGAAGAGCTATAGGTTTGTGAAAACCGCAGGC

ATCCTGGCAGATGGCCTGGGAGAGGAGGAGCAGGACTCCCTG

TTCATCGTGGAGACATTCAACAAGACCCTGACACAGGACGGC

ATCGATACCTACAATTCTCAAGTGGGCAAGATCAACTCTAGCA

TCAATCTGTATAACCAGAAGAATCAGAAGGCCAATGGCTTCAG

AAAGATCCCCAAGATGAAGATGCTGTATAAGCAGATCCTGTCC

GATAGGGAGGAGTCTTTCATCGACGAGTTTCAGAGCGATGAG

GTGCTGATCGACAACGTGGAGTCTTATGGCAGCGTGCTGATCG

AGTCTCTGAAGTCCTCTAAGGTGAGCGCCTTCTTTGATGCCCT

GAGAGAGTCTAAGGGCAAGAACGTGTACGTGAAGAATGACCT

GGCCAAGACAGCCATGAGCAACATCGTGTTCGAGAATTGGAG

GACCTTTGACGATCTGCTGAACCAGGAGTACGACCTGGCCAA

CGAGAACAAGAAGAAGGACGATAAGTATTTCGAGAAGCGCCA

GAAGGAGCTGAAGAAGAATAAGAGCTACTCCCTGGAGCACCT

GTGCAACCTGTCCGAGGATTCTTGTAACCTGATCGAGAATTAT

ATCCACCAGATCTCCGACGATATCGAGAATATCATCATCAACAA

TGAGACATTCCTGCGCATCGTGATCAATGAGCACGACAGGTCC

CGCAAGCTGGCCAAGAACCGGAAGGCCGTGAAGGCCATCAA

GGACTTTCTGGATTCTATCAAGGTGCTGGAGCGGGAGCTGAA

GCTGATCAACAGCTCCGGCCAGGAGCTGGAGAAGGATCTGAT

CGTGTACTCTGCCCACGAGGAGCTGCTGGTGGAGCTGAAGCA

GGTGGACAGCCTGTATAACATGACCAGAAATTATCTGACAAAG

AAGCCTTTCTCTACCGAGAAGGTGAAGCTGAACTTTAATCGCA

GCACACTGCTGAACGGCTGGGATCGGAATAAGGAGACAGACA

ACCTGGGCGTGCTGCTGCTGAAGGACGGCAAGTACTATCTGG

GCATCATGAACACAAGCGCCAATAAGGCCTTCGTGAATCCCCC

TGTGGCCAAGACCGAGAAGGTGTTTAAGAAGGTGGATTACAA

GCTGCTGCCAGTGCCCAACCAGATGCTGCCAAAGGTGTTCTTT

GCCAAGAGCAATATCGACTTCTATAACCCCTCTAGCGAGATCT

ACTCCAATTATAAGAAGGGCACCCACAAGAAGGGCAATATGTT

TTCCCTGGAGGATTGTCACAACCTGATCGACTTCTTTAAGGAG

TCTATCAGCAAGCACGAGGACTGGAGCAAGTTCGGCTTTAAG

TTCAGCGATACAGCCTCCTACAACGACATCTCCGAGTTCTATC

GCGAGGTGGAGAAGCAGGGCTACAAGCTGACCTATACAGACA

TCGATGAGACATACATCAATGATCTGATCGAGCGGAACGAGCT

GTACCTGTTCCAGATCTATAATAAGGACTTTAGCATGTACTCCA

AGGGCAAGCTGAACCTGCACACACTGTATTTCATGATGCTGTT

TGATCAGCGCAATATCGACGACGTGGTGTATAAGCTGAACGGA

GAGGCAGAGGTGTTCTATAGGCCAGCCTCCATCTCTGAGGACG

AGCTGATCATCCACAAGGCCGGCGAGGAGATCAAGAACAAGA

ATCCTAACCGGGCCAGAACCAAGGAGACAAGCACCTTCAGCT

ACGACATCGTGAAGGATAAGCGGTATAGCAAGGATAAGTTTAC

CCTGCACATCCCCATCACAATGAACTTCGGCGTGGATGAGGTG

AAGCGGTTCAACGACGCCGTGAACAGCGCCATCCGGATCGAT

GAGAATGTGAACGTGATCGGCATCGACCGGGGCGAGAGAAAT

CTGCTGTACGTGGTGGTCATCGACTCTAAGGGCAACATCCTGG

AGCAGATCTCCCTGAACTCTATCATCAATAAGGAGTACGACAT

CGAGACAGATTATCACGCACTGCTGGATGAGAGGGAGGGCGG

CAGAGATAAGGCCCGGAAGGACTGGAACACCGTGGAGAATAT

CAGGGACCTGAAGGCCGGCTACCTGAGCCAGGTGGTGAACGT

GGTGGCCAAGCTGGTGCTGAAGTATAATGCCATCATCTGCCTG

GAGGACCTGAACTTTGGCTTCAAGAGGGGCCGCCAGAAGGTG

GAGAAGCAGGTGTACCAGAAGTTCGAGAAGATGCTGATCGAT

AAGCTGAATTACCTGGTCATCGACAAGAGCCGCGAGCAGACA

TCCCCTAAGGAGCTGGGAGGCGCCCTGAACGCACTGCAGCTG

ACCTCTAAGTTCAAGAGCTTTAAGGAGCTGGGCAAGCAGTCC

GGCGTGATCTACTATGTGCCTGCCTACCTGACCTCTAAGATCG

ATCCAACCACAGGCTTCGCCAATCTGTTTTATATGAAGTGTGA

GAACGTGGAGAAGTCCAAGAGATTCTTTGACGGCTTTGATTTC

ATCAGGTTCAACGCCCTGGAGAACGTGTTCGAGTTCGGCTTTG

ACTACCGGAGCTTCACCCAGAGGGCCTGCGGCATCAATTCCA

AGTGGACCGTGTGCACCAACGGCGAGCGCATCATCAAGTATC

GGAATCCAGATAAGAACAATATGTTCGACGAGAAGGTGGTGG

TGGTGACCGATGAGATGAAGAACCTGTTTGAGCAGTACAAGA

TCCCCTATGAGGATGGCAGAAATGTGAAGGACATGATCATCAG

CAACGAGGAGGCCGAGTTCTACCGGAGACTGTATAGGCTGCT
GCAGCAGACCCTGCAGATGAGAAACAGCACCTCCGACGGCAC
AAGGGATTACATCATCTCCCCTGTGAAGAATAAGAGAGAGGCC
TACTTCAACAGCGAGCTGTCCGACGGCTCTGTGCCAAAGGAC
GCCGATGCCAACGGCGCCTACAATATCGCCAGAAAGGGCCTGT
GGGTGCTGGAGCAGATCAGGCAGAAGAGCGAGGGCGAGAAG
ATCAATCTGGCCATGACCAACGCCGAGTGGCTGGAGTATGCCC
AGACACACCTGCTG*AAAAGGCCGGCGGCCACGAAAAAGGCCGG*
*CCAGGCAAAAAAGAAAAAG*GGATCC**TACCCATACGATGTTCCA**
GATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCC
ATATGATGTCCCCGACTATGCCTAAGAATTC （SEQ ID NO:
221）

[1686] 10-候選白蟻甲烷枝原體（CMtCpf1）

[1687]

ATGAACAATTACGACGAGTTCACCAAGCTGTATCCTATCC
AGAAAACCATCCGGTTTGAGCTGAAGCCACAGGGCAGAACCA
TGGAGCACCTGGAGACATTCAACTTCTTTGAGGAGGACCGGG
ATAGAGCCGAGAAGTATAAGATCCTGAAGGAGGCCATCGACG
AGTACCACAAGAAGTTTATCGATGAGCACCTGACCAATATGTC
CCTGGATTGGAACTCTCTGAAGCAGATCAGCGAGAAGTACTAT
AAGAGCAGGGAGGAGAAGGACAAGAAGGTGTTCCTGTCCGA
GCAGAAGAGGATGCGCCAGGAGATCGTGTCTGAGTTTAAGAA
GGACGATCGCTTCAAGGACCTGTTTTCCAAGAAGCTGTTCTCT

GAGCTGCTGAAGGAGGAGATCTACAAGAAGGGCAACCACCA

GGAGATCGACGCCCTGAAGAGCTTCGATAAGTTTTCCGGCTAT

TTCATCGGCCTGCACGAGAATAGGAAGAACATGTACTCCGACG

GCGATGAGATCACCGCCATCTCCAATCGCATCGTGAATGAGAA

CTTCCCCAAGTTTCTGGATAACCTGCAGAAGTACCAGGAGGCC

AGGAAGAAGTATCCTGAGTGGATCATCAAGGCCGAGAGCGCC

CTGGTGGCCCACAATATCAAGATGGACGAGGTGTTCTCCCTGG

AGTACTTTAATAAGGTGCTGAACCAGGAGGGCATCCAGCGGTA

CAACCTGGCCCTGGGCGGCTATGTGACCAAGAGCGGCGAGAA

GATGATGGGCCTGAATGATGCCCTGAACCTGGCCCACCAGTCC

GAGAAGAGCTCCAAGGGCAGAATCCACATGACCCCCCTGTTC

AAGCAGATCCTGTCCGAGAAGGAGTCCTTCTCTTACATCCCCG

ACGTGTTTACAGAGGATTCTCAGCTGCTGCCTAGCATCGGCGG

CTTCTTTGCCCAGATCGAGAATGACAAGGATGGCAACATCTTC

GACCGGGCCCTGGAGCTGATCTCTAGCTACGCCGAGTATGATA

CCGAGCGGATCTATATCAGACAGGCCGACATCAATAGAGTGTC

CAACGTGATCTTTGGAGAGTGGGGCACCCTGGGAGGCCTGAT

GAGGGAGTACAAGGCCGACTCTATCAATGATATCAACCTGGAG

CGCACATGCAAGAAGGTGGACAAGTGGCTGGATTCTAAGGAG

TTTGCCCTGAGCGATGTGCTGGAGGCCATCAAGAGGACCGGC

AACAATGACGCCTTCAACGAGTATATCTCCAAGATGCGGACAG

CCAGAGAGAAGATCGATGCCGCCCGCAAGGAGATGAAGTTCA

TCAGCGAGAAGATCTCCGGCGATGAGGAGTCTATCCACATCAT

CAAGACCCTGCTGGACAGCGTGCAGCAGTTCCTGCACTTCTTT

AATCTGTTTAAGGCAAGGCAGGACATCCCACTGGATGGAGCCT

TCTACGCCGAGTTTGACGAGGTGCACAGCAAGCTGTTTGCCAT

CGTGCCCCTGTATAACAAGGTGCGGAACTATCTGACCAAGAAC

AATCTGAACACAAAGAAGATCAAGCTGAATTTCAAGAACCCT

ACACTGGCCAATGGCTGGGACCAGAACAAGGTGTACGATTAT

GCCTCCCTGATCTTTCTGCGGGACGGCAATTACTATCTGGGCAT

CATCAATCCTAAGAGAAAGAAGAACATCAAGTTCGAGCAGGG

CTCTGGCAACGGCCCCTTCTACCGGAAGATGGTGTATAAGCAG

ATCCCCGGCCCTAATAAGAACCTGCCAAGAGTGTTCCTGACCT

CCACAAAGGGCAAGAAGGAGTATAAGCCCTCTAAGGAGATCA

TCGAGGGCTACGAGGCCGACAAGCACATCAGGGGCGATAAGT

TCGACCTGGATTTTTGTCACAAGCTGATCGATTTCTTTAAGGA

GTCCATCGAGAAGCACAAGGACTGGTCTAAGTTCAACTTCTAC

TTCAGCCCAACCGAGAGCTATGGCGACATCTCTGAGTTCTACC

TGGATGTGGAGAAGCAGGGCTATCGCATGCACTTTGAGAATAT

CAGCGCCGAGACAATCGACGAGTATGTGGAGAAGGGCGATCT

GTTTCTGTTCCAGATCTACAACAAGGATTTTGTGAAGGCCGCC

ACCGGCAAGAAGGACATGCACACAATCTACTGGAATGCCGCC

TTCAGCCCCGAGAACCTGCAGGACGTGGTGGTGAAGCTGAAC

GGCGAGGCCGAGCTGTTTTATAGGGACAAGTCCGATATCAAGG

AGATCGTGCACCGCGAGGGCGAGATCCTGGTGAATAGGACCT

ACAACGGCCGCACACCAGTGCCCGACAAGATCCACAAGAAGC

TGACCGATTATCACAATGGCCGGACAAAGGACCTGGGCGAGG

CCAAGGAGTACCTGGATAAGGTGAGATACTTCAAGGCCCACTA

TGACATCACCAAGGATCGGAGATACCTGAACGACAAGATCTAT

TTCCACGTGCCTCTGACCCTGAACTTCAAGGCCAACGGCAAG

AAGAATCTGAACAAGATGGTCATCGAGAAGTTCCTGTCCGATG

AGAAGGCCCACATCATCGGCATCGACAGGGGCGAGCGCAATC

TGCTGTACTATTCCATCATCGACAGGTCTGGCAAGATCATCGAT

CAGCAGAGCCTGAATGTGATCGACGGCTTTGATTATCGGGAGA

AGCTGAACCAGAGAGAGATCGAGATGAAGGATGCCCGCCAGT

CTTGGAACGCCATCGGCAAGATCAAGGACCTGAAGGAGGGCT

ACCTGAGCAAGGCCGTGCACGAGATCACCAAGATGGCCATCC

AGTATAATGCCATCGTGGTCATGGAGGAGCTGAACTACGGCTT

CAAGCGGGGCCGGTTCAAGGTGGAGAAGCAGATCTATCAGAA

GTTCGAGAATATGCTGATCGATAAGATGAACTACCTGGTGTTTA

AGGACGCACCTGATGAGTCCCCAGGAGGCGTGCTGAATGCCT

ACCAGCTGACAAACCCACTGGAGTCTTTCGCCAAGCTGGGCA

AGCAGACCGGCATCCTGTTTTACGTGCCAGCCGCCTATACATC

CAAGATCGACCCCACCACAGGCTTCGTGAATCTGTTTAACACC

TCCTCTAAGACAAACGCCCAGGAGCGGAAGGAGTTCCTGCAG

AAGTTTGAGAGCATCTCCTATTCTGCCAAGGATGGCGGCATCT

TTGCCTTCGCCTTTGACTACAGAAAGTTCGGCACCAGCAAGA

CAGATCACAAGAACGTGTGGACCGCCTATACAAACGGCGAGA

GGATGCGCTACATCAAGGAGAAGAAGCGGAATGAGCTGTTTG

ACCCTTCTAAGGAGATCAAGGAGGCCCTGACCAGCTCCGGCA

TCAAGTACGATGGCGGCCAGAACATCCTGCCAGACATCCTGA

GGAGCAACAATAACGGCCTGATCTACACAATGTATTCTAGCTT

CATCGCCGCCATCCAGATGCGCGTGTACGACGGCAAGGAGGA
TTATATCATCAGCCCCATCAAGAACTCCAAGGGCGAGTTCTTT
AGGACCGACCCCAAGAGGCGCGAGCTGCCTATCGACGCCGAT
GCCAATGGCGCCTACAACATCGCCCTGAGGGGAGAGCTGACA
ATGAGGGCAATCGCAGAGAAGTTCGACCCTGATAGCGAGAAG
ATGGCCAAGCTGGAGCTGAAGCACAAGGATTGGTTCGAGTTT
ATGCAGACCAGAGGCGAC*AAAAGGCCGGCGGCCACGAAAAAGG*
*CCGGCCAGGCAAAAAAGAAAAAG*GGATCC**TACCCATACGATGT**
TCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCAT
ACCCATATGATGTCCCCGACTATGCCTAAGAATTC（SEQ ID
NO: 222）

[1688]　11-挑剔真細菌（EeCpf1）

[1689]

ATGAACGGCAATAGGTCCATCGTGTACCGCGAGTTCGTGG
GCGTGATCCCCGTGGCCAAGACCCTGAGGAATGAGCTGCGCC
CTGTGGGCCACACACAGGAGCACATCATCCAGAACGGCCTGA
TCCAGGAGGACGAGCTGCGGCAGGAGAAGAGCACCGAGCTG
AAGAACATCATGGACGATTACTATAGAGAGTACATCGATAAGT
CTCTGAGCGGCGTGACCGACCTGGACTTCACCCTGCTGTTCGA
GCTGATGAACCTGGTGCAGAGCTCCCCCTCCAAGGACAATAA
GAAGGCCCTGGAGAAGGAGCAGTCTAAGATGAGGGAGCAGAT
CTGCACCCACCTGCAGTCCGACTCTAACTACAAGAATATCTTT
AACGCCAAGCTGCTGAAGGAGATCCTGCCTGATTTCATCAAGA

ACTACAATCAGTATGACGTGAAGGATAAGGCCGGCAAGCTGG

AGACACTGGCCCTGTTTAATGGCTTCAGCACATACTTTACCGA

CTTCTTTGAGAAGAGGAAGAACGTGTTCACCAAGGAGGCCGT

GAGCACATCCATCGCCTACCGCATCGTGCACGAGAACTCCCTG

ATCTTCCTGGCCAATATGACCTCTTATAAGAAGATCAGCGAGA

AGGCCCTGGATGAGATCGAAGTGATCGAGAAGAACAATCAGG

ACAAGATGGGCGATTGGGAGCTGAATCAGATCTTTAACCCTGA

CTTCTACAATATGGTGCTGATCCAGTCCGGCATCGACTTCTACA

ACGAGATCTGCGGCGTGGTGAATGCCCACATGAACCTGTACTG

TCAGCAGACCAAGAACAATTATAACCTGTTCAAGATGCGGAA

GCTGCACAAGCAGATCCTGGCCTACACCAGCACCAGCTTCGA

GGTGCCCAAGATGTTCGAGGACGATATGAGCGTGTATAACGCC

GTGAACGCCTTCATCGACGAGACAGAGAAGGGCAACATCATC

GGCAAGCTGAAGGATATCGTGAATAAGTACGACGAGCTGGAT

GAGAAGAGAATCTATATCAGCAAGGACTTTTACGAGACACTGA

GCTGCTTCATGTCCGGCAACTGGAATCTGATCACAGGCTGCGT

GGAGAACTTCTACGATGAGAACATCCACGCCAAGGGCAAGTC

CAAGGAGGAGAAGGTGAAGAAGGCCGTGAAGGAGGACAAGT

ACAAGTCTATCAATGACGTGAACGATCTGGTGGAGAAGTATAT

CGATGAGAAGGAGAGGAATGAGTTCAAGAACAGCAATGCCAA

GCAGTACATCCGCGAGATCTCCAACATCATCACCGACACAGAG

ACAGCCCACCTGGAGTATGACGATCACATCTCTCTGATCGAGA

GCGAGGAGAAGGCCGACGAGATGAAGAAGCGGCTGGATATGT

ATATGAACATGTACCACTGGGCCAAGGCCTTTATCGTGGACGA

GGTGCTGGACAGAGATGAGATGTTCTACAGCGATATCGACGAT

ATCTATAATATCCTGGAGAACATCGTGCCACTGTATAATCGGGT

GAGAAACTACGTGACCCAGAAGCCCTACAACTCTAAGAAGAT

CAAGCTGAATTTCCAGAGCCCTACACTGGCCAATGGCTGGTCC

CAGTCTAAGGAGTTCGACAACAATGCCATCATCCTGATCAGAG

ATAACAAGTACTATCTGGCCATCTTCAATGCCAAGAACAAGCC

AGACAAGAAGATCATCCAGGGCAACTCCGATAAGAAGAACGA

CAACGATTACAAGAAGATGGTGTATAACCTGCTGCCAGGCGCC

AACAAGATGCTGCCCAAGGTGTTTCTGTCTAAGAAGGGCATC

GAGACATTCAAGCCCTCCGACTATATCATCTCTGGCTACAACG

CCCACAAGCACATCAAGACAAGCGAGAATTTTGATATCTCCTT

CTGTCGGGACCTGATCGATTACTTCAAGAACAGCATCGAGAAG

CACGCCGAGTGGAGAAAGTATGAGTTCAAGTTTTCCGCCACC

GACAGCTACTCCGATATCTCTGAGTTCTATCGGGAGGTGGAGA

TGCAGGGCTACAGAATCGACTGGACATATATCAGCGAGGCCGA

CATCAACAAGCTGGATGAGGAGGGCAAGATCTATCTGTTTCAG

ATCTACAATAAGGATTTCGCCGAGAACAGCACCGGCAAGGAG

AATCTGCACACAATGTACTTTAAGAACATCTTCTCCGAGGAGA

ATCTGAAGGACATCATCATCAAGCTGAACGGCCAGGCCGAGC

TGTTTTATCGGAGAGCCTCTGTGAAGAATCCCGTGAAGCACAA

GAAGGATAGCGTGCTGGTGAACAAGACCTACAAGAATCAGCT

GGACAACGGCGACGTGGTGAGAATCCCCATCCCTGACGATATC

TATAACGAGATCTACAAGATGTATAATGGCTACATCAAGGAGTC

CGACCTGTCTGAGGCCGCCAAGGAGTACCTGGATAAGGTGGA

GGTGAGGACCGCCCAGAAGGACATCGTGAAGGATTACCGCTA

TACAGTGGACAAGTACTTCATCCACACACCTATCACCATCAAC

TATAAGGTGACCGCCCGCAACAATGTGAATGATATGGTGGTGA

AGTACATCGCCCAGAACGACGATATCCACGTGATCGGCATCGA

CCGGGGCGAGAGAAACCTGATCTACATCTCCGTGATCGATTCT

CACGGCAACATCGTGAAGCAGAAATCCTACAACATCCTGAAC

AACTACGACTACAAGAAGAAGCTGGTGGAGAAGGAGAAAAC

CCGGGAGTACGCCAGAAAGAACTGGAAGAGCATCGGCAATAT

CAAGGAGCTGAAGGAGGGCTATATCTCCGGCGTGGTGCACGA

GATCGCCATGCTGATCGTGGAGTACAACGCCATCATCGCCATG

GAGGACCTGAATTATGGCTTTAAGAGGGGCCGCTTCAAGGTG

GAGCGGCAGGTGTACCAGAAGTTTGAGAGCATGCTGATCAAT

AAGCTGAACTATTTCGCCAGCAAGGAGAAGTCCGTGGACGAG

CCAGGAGGCCTGCTGAAGGGCTATCAGCTGACCTACGTGCCC

GATAATATCAAGAACCTGGGCAAGCAGTGCGGCGTGATCTTTT

ACGTGCCTGCCGCCTTCACCAGCAAGATCGACCCATCCACAG

GCTTTATCTCTGCCTTCAACTTTAAGTCTATCAGCACAAATGCC

TCTCGGAAGCAGTTCTTTATGCAGTTTGACGAGATCAGATACT

GTGCCGAGAAGGATATGTTCAGCTTTGGCTTCGACTACAACAA

CTTCGATACCTACAACATCACAATGGGCAAGACACAGTGGACC

GTGTATACAAACGGCGAGAGACTGCAGTCTGAGTTCAACAAT

GCCAGGCGCACCGGCAAGACAAAGAGCATCAATCTGACAGAG

ACAATCAAGCTGCTGCTGGAGGACAATGAGATCAACTACGCC

GACGGCCACGATATCAGGATCGATATGGAGAAGATGGACGAG

GATAAGAAGAGCGAGTTCTTTGCCCAGCTGCTGAGCCTGTATA
AGCTGACCGTGCAGATGCGCAATTCCTATACAGAGGCCGAGG
AGCAGGAGAACGGCATCTCTTACGACAAGATCATCAGCCCTGT
GATCAATGATGAGGGCGAGTTCTTTGACTCCGATAACTATAAG
GAGTCTGACGATAAGGAGTGCAAGATGCCAAAGGACGCCGAT
GCCAACGGCGCCTACTGTATCGCCCTGAAGGGCCTGTATGAGG
TGCTGAAGATCAAGAGCGAGTGGACCGAGGACGGCTTTGATA
GGAATTGCCTGAAGCTGCCACACGCAGAGTGGCTGGACTTCA
TCCAGAACAAGCGGTACGAG*AAAAGGCCGGCGGCCACGAAAAA*
*GGCCGGCCAGGCAAAAAAGAAAAAG*GGATCC**TACCCATACGAT**
GTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGC
ATACCCATATGATGTCCCCGACTATGCCTAAGAATTC （SEQ
ID NO: 223）

[1690] 12-牛莫拉氏菌237（MbCpf1）

[1691]

ATGCTGTTCCAGGACTTTACCCACCTGTATCCACTGTCCAA
GACAGTGAGATTTGAGCTGAAGCCCATCGATAGGACCCTGGA
GCACATCCACGCCAAGAACTTCCTGTCTCAGGACGAGACAAT
GGCCGATATGCACCAGAAGGTGAAAGTGATCCTGGACGATTAC
CACCGCGACTTCATCGCCGATATGATGGGCGAGGTGAAGCTGA
CCAAGCTGGCCGAGTTCTATGACGTGTACCTGAAGTTTCGGAA
GAACCCAAAGGACGATGAGCTGCAGAAGCAGCTGAAGGATCT
GCAGGCCGTGCTGAGAAAGGAGATCGTGAAGCCCATCGGCAA

TGGCGGCAAGTATAAGGCCGGCTACGACAGGCTGTTCGGCGC

CAAGCTGTTTAAGGACGGCAAGGAGCTGGGCGATCTGGCCAA

GTTCGTGATCGCACAGGAGGGAGAGAGCTCCCCAAAGCTGGC

CCACCTGGCCCACTTCGAGAAGTTTTCCACCTATTTCACAGGC

TTTCACGATAACCGGAAGAATATGTATTCTGACGAGGATAAGC

ACACCGCCATCGCCTACCGCCTGATCCACGAGAACCTGCCCCG

GTTTATCGACAATCTGCAGATCCTGACCACAATCAAGCAGAAG

CACTCTGCCCTGTACGATCAGATCATCAACGAGCTGACCGCCA

GCGGCCTGGACGTGTCTCTGGCCAGCCACCTGGATGGCTATCA

CAAGCTGCTGACACAGGAGGGCATCACCGCCTACAATACACT

GCTGGGAGGAATCTCCGGAGAGGCAGGCTCTCCTAAGATCCA

GGGCATCAACGAGCTGATCAATTCTCACCACAACCAGCACTGC

CACAAGAGCGAGAGAATCGCCAAGCTGAGGCCACTGCACAA

GCAGATCCTGTCCGACGGCATGAGCGTGTCCTTCCTGCCCTCT

AAGTTTGCCGACGATAGCGAGATGTGCCAGGCCGTGAACGAG

TTCTATCGCCACTACGCCGACGTGTTCGCCAAGGTGCAGAGCC

TGTTCGACGGCTTTGACGATCACCAGAAGGATGGCATCTACGT

GGAGCACAAGAACCTGAATGAGCTGTCCAAGCAGGCCTTCGG

CGACTTTGCACTGCTGGGACGCGTGCTGGACGGATACTATGTG

GATGTGGTGAATCCAGAGTTCAACGAGCGGTTTGCCAAGGCC

AAGACCGACAATGCCAAGGCCAAGCTGACAAAGGAGAAGGA

TAAGTTCATCAAGGGCGTGCACTCCCTGGCCTCTCTGGAGCAG

GCCATCGAGCACTATACCGCAAGGCACGACGATGAGAGCGTG

CAGGCAGGCAAGCTGGGACAGTACTTCAAGCACGGCCTGGCC

GGAGTGGACAACCCCATCCAGAAGATCCACAACAATCACAGC

ACCATCAAGGGCTTTCTGGAGAGGGAGCGCCCTGCAGGAGAG

AGAGCCCTGCCAAAGATCAAGTCCGGCAAGAATCCTGAGATG

ACACAGCTGAGGCAGCTGAAGGAGCTGCTGGATAACGCCCTG

AATGTGGCCCACTTCGCCAAGCTGCTGACCACAAAGACCACA

CTGGACAATCAGGATGGCAACTTCTATGGCGAGTTTGGCGTGC

TGTACGACGAGCTGGCCAAGATCCCCACCCTGTATAACAAGGT

GAGAGATTACCTGAGCCAGAAGCCTTTCTCCACCGAGAAGTA

CAAGCTGAACTTTGGCAATCCAACACTGCTGAATGGCTGGGA

CCTGAACAAGGAGAAGGATAATTTCGGCGTGATCCTGCAGAA

GGACGGCTGCTACTATCTGGCCCTGCTGGACAAGGCCCACAA

GAAGGTGTTTGATAACGCCCCTAATACAGGCAAGAGCATCTAT

CAGAAGATGATCTATAAGTACCTGGAGGTGAGGAAGCAGTTCC

CCAAGGTGTTCTTTTCCAAGGAGGCCATCGCCATCAACTACCA

CCCTTCTAAGGAGCTGGTGGAGATCAAGGACAAGGGCCGGCA

GAGATCCGACGATGAGCGCCTGAAGCTGTATCGGTTTATCCTG

GAGTGTCTGAAGATCCACCCTAAGTACGATAAGAAGTTCGAG

GGCGCCATCGGCGACATCCAGCTGTTTAAGAAGGATAAGAAG

GGCAGAGAGGTGCCAATCAGCGAGAAGGACCTGTTCGATAAG

ATCAACGGCATCTTTTCTAGCAAGCCTAAGCTGGAGATGGAGG

ACTTCTTTATCGGCGAGTTCAAGAGGTATAACCCAAGCCAGGA

CCTGGTGGATCAGTATAATATCTACAAGAAGATCGACTCCAAC

GATAATCGCAAGAAGGAGAATTTCTACAACAATCACCCCAAGT

TTAAGAAGGATCTGGTGCGGTACTATTACGAGTCTATGTGCAA

GCACGAGGAGTGGGAGGAGAGCTTCGAGTTTTCCAAGAAGCT

GCAGGACATCGGCTGTTACGTGGATGTGAACGAGCTGTTTACC

GAGATCGAGACACGGAGACTGAATTATAAGATCTCCTTCTGCA

ACATCAATGCCGACTACATCGATGAGCTGGTGGAGCAGGGCC

AGCTGTATCTGTTCCAGATCTACAACAAGGACTTTTCCCCAAA

GGCCCACGGCAAGCCCAATCTGCACACCCTGTACTTCAAGGC

CCTGTTTTCTGAGGACAACCTGGCCGATCCTATCTATAAGCTG

AATGGCGAGGCCCAGATCTTCTACAGAAAGGCCTCCCTGGAC

ATGAACGAGACAACAATCCACAGGGCCGGCGAGGTGCTGGAG

AACAAGAATCCCGATAATCCTAAGAAGAGACAGTTCGTGTACG

ACATCATCAAGGATAAGAGGTACACACAGGACAAGTTCATGCT

GCACGTGCCAATCACCATGAACTTTGGCGTGCAGGGCATGAC

AATCAAGGAGTTCAATAAGAAGGTGAACCAGTCTATCCAGCA

GTATGACGAGGTGAACGTGATCGGCATCGATCGGGGCGAGAG

ACACCTGCTGTACCTGACCGTGATCAATAGCAAGGGCGAGATC

CTGGAGCAGTGTTCCCTGAACGACATCACCACAGCCTCTGCC

AATGGCACACAGATGACCACACCTTACCACAAGATCCTGGATA

AGAGGGAGATCGAGCGCCTGAACGCCCGGGTGGGATGGGGCG

AGATCGAGACAATCAAGGAGCTGAAGTCTGGCTATCTGAGCC

ACGTGGTGCACCAGATCAGCCAGCTGATGCTGAAGTACAACG

CCATCGTGGTGCTGGAGGACCTGAATTTCGGCTTTAAGAGGGG

CCGCTTTAAGGTGGAGAAGCAGATCTATCAGAACTTCGAGAAT

GCCCTGATCAAGAAGCTGAACCACCTGGTGCTGAAGGACAAG

GCCGACGATGAGATCGGCTCTTACAAGAATGCCCTGCAGCTGA

CCAACAATTTCACAGATCTGAAGAGCATCGGCAAGCAGACCG
GCTTCCTGTTTTATGTGCCCGCCTGGAACACCTCTAAGATCGA
CCCTGAGACAGGCTTTGTGGATCTGCTGAAGCCAAGATACGA
GAACATCGCCCAGAGCCAGGCCTTCTTTGGCAAGTTCGACAA
GATCTGCTATAATGCCGACAAGGATTACTTCGAGTTTCACATCG
ACTACGCCAAGTTTACCGATAAGGCCAAGAATAGCCGCCAGAT
CTGGACAATCTGTTCCCACGGCGACAAGCGGTACGTGTACGAT
AAGACAGCCAACCAGAATAAGGGCGCCGCCAAGGGCATCAAC
GTGAATGATGAGCTGAAGTCCCTGTTCGCCCGCCACCACATCA
ACGAGAAGCAGCCCAACCTGGTCATGGACATCTGCCAGAACA
ATGATAAGGAGTTTCACAAGTCTCTGATGTACCTGCTGAAAAC
CCTGCTGGCCCTGCGGTACAGCAACGCCTCCTCTGACGAGGAT
TTCATCCTGTCCCCCGTGGCAAACGACGAGGGCGTGTTCTTTA
ATAGCGCCCTGGCCGACGATACACAGCCTCAGAATGCCGATGC
CAACGGCGCCTACCACATCGCCCTGAAGGGCCTGTGGCTGCT
GAATGAGCTGAAGAACTCCGACGATCTGAACAAGGTGAAGCT
GGCCATCGACAATCAGACCTGGCTGAATTTCGCCCAGAACAG
G*AAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAG*
*AAAAAG*GGATCC**TACCCATACGATGTTCCAGATTACGCTTAT**
CCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCC
CGACTATGCCTAAGAATTC （SEQ ID NO: 224）

[1692] 13-稻田鉤端螺旋體（LiCpf1）

[1693]

ATGGAGGACTATTCCGGCTTTGTGAACATCTACTCTATCCA
GAAAACCCTGAGGTTCGAGCTGAAGCCAGTGGGCAAGACACT
GGAGCACATCGAGAAGAAGGGCTTCCTGAAGAAGGACAAGAT
CCGGGCCGAGGATTACAAGGCCGTGAAGAAGATCATCGATAA
GTACCACAGAGCCTATATCGAGGAGGTGTTTGATTCCGTGCTG
CACCAGAAGAAGAAGAAGGACAAGACCCGCTTTTCTACACAG
TTCATCAAGGAGATCAAGGAGTTCAGCGAGCTGTACTATAAGA
CCGAGAAGAACATCCCCGACAAGGAGAGGCTGGAGGCCCTG
AGCGAGAAGCTGCGCAAGATGCTGGTGGGCGCCTTTAAGGGC
GAGTTCTCCGAGGAGGTGGCCGAGAAGTATAAGAACCTGTTT
TCTAAGGAGCTGATCAGGAATGAGATCGAGAAGTTCTGCGAG
ACAGACGAGGAGCGCAAGCAGGTGTCTAACTTCAAGAGCTTC
ACCACATACTTTACCGGCTTCCACTCCAACAGGCAGAATATCT
ATTCCGACGAGAAGAAGTCTACAGCCATCGGCTACCGCATCAT
CCACCAGAACCTGCCTAAGTTCCTGGATAATCTGAAGATCATC
GAGTCCATCCAGCGGCGGTTCAAGGACTTCCCATGGTCTGATC
TGAAGAAGAACCTGAAGAAGATCGATAAGAATATCAAGCTGA
CCGAGTACTTCAGCATCGACGGCTTCGTGAACGTGCTGAATCA
GAAGGGCATCGATGCCTACAACACAATCCTGGGCGGCAAGTC
CGAGGAGTCTGGCGAGAAGATCCAGGGCCTGAACGAGTACAT
CAATCTGTATCGGCAGAAGAACAATATCGACAGAAAGAACCT
GCCCAATGTGAAGATCCTGTTTAAGCAGATCCTGGGCGATAGG
GAGACAAAGAGCTTTATCCCTGAGGCCTTCCCAGACGATCAGT
CCGTGCTGAACTCTATCACAGAGTTCGCCAAGTACCTGAAGCT

GGATAAGAAGAAGAAGAGCATCATCGCCGAGCTGAAGAAGTT

TCTGAGCTCCTTCAATCGCTACGAGCTGGACGGCATCTATCTG

GCCAACGATAATAGCCTGGCCTCTATCAGCACCTTCCTGTTTG

ACGATTGGTCCTTTATCAAGAAGTCCGTGTCTTTCAAGTATGA

CGAGTCCGTGGGCGACCCCAAGAAGAAGATCAAGTCTCCCCT

GAAGTACGAGAAGGAGAAGGAGAAGTGGCTGAAGCAGAAGT

ACTATACAATCTCTTTCCTGAACGATGCCATCGAGAGCTATTCC

AAGTCTCAGGACGAGAAGAGGGTGAAGATCCGCCTGGAGGCC

TACTTTGCCGAGTTCAAGAGCAAGGACGATGCCAAGAAGCAG

TTCGACCTGCTGGAGAGGATCGAGGAGGCCTATGCCATCGTGG

AGCCTCTGCTGGGAGCAGAGTACCCAAGGGACCGCAACCTGA

AGGCCGATAAGAAGGAAGTGGGCAAGATCAAGGACTTCCTGG

ATAGCATCAAGTCCCTGCAGTTCTTTCTGAAGCCTCTGCTGTC

CGCCGAGATCTTTGACGAGAAGGATCTGGGCTTCTACAATCAG

CTGGAGGGCTACTATGAGGAGATCGATTCTATCGGCCACCTGT

ATAACAAGGTGCGGAATTATCTGACCGGCAAGATCTACAGCAA

GGAGAAGTTTAAGCTGAACTTCGAGAACAGCACCCTGCTGAA

GGGCTGGGACGAGAACCGGGAGGTGGCCAATCTGTGCGTGAT

CTTCAGAGAGGACCAGAAGTACTATCTGGGCGTGATGGATAAG

GAGAACAATACCATCCTGTCCGACATCCCCAAGGTGAAGCCTA

ACGAGCTGTTTTACGAGAAGATGGTGTATAAGCTGATCCCCAC

ACCTCACATGCAGCTGCCCCGGATCATCTTCTCTAGCGACAAC

CTGTCTATCTATAATCCTAGCAAGTCCATCCTGAAGATCAGAGA

GGCCAAGAGCTTTAAGGAGGGCAAGAACTTCAAGCTGAAGG

ACTGTCACAAGTTTATCGATTTCTACAAGGAGTCTATCAGCAA

GAATGAGGACTGGAGCAGATTCGACTTCAAGTTCAGCAAGAC

CAGCAGCTACGAGAACATCAGCGAGTTTTACCGGGAGGTGGA

GAGACAGGGCTATAACCTGGACTTCAAGAAGGTGTCTAAGTT

CTACATCGACAGCCTGGTGGAGGATGGCAAGCTGTACCTGTTC

CAGATCTATAACAAGGACTTTTCTATCTTCAGCAAGGGCAAGC

CCAATCTGCACACCATCTATTTTCGGTCCCTGTTCTCTAAGGAG

AACCTGAAGGACGTGTGCCTGAAGCTGAATGGCGAGGCCGAG

ATGTTCTTTCGGAAGAAGTCCATCAACTACGATGAGAAGAAG

AAGCGGGAGGGCCACCACCCCGAGCTGTTTGAGAAGCTGAAG

TATCCTATCCTGAAGGACAAGAGATACAGCGAGGATAAGTTTC

AGTTCCACCTGCCCATCAGCCTGAACTTCAAGTCCAAGGAGC

GGCTGAACTTTAATCTGAAAGTGAATGAGTTCCTGAAGAGAA

ACAAGGACATCAATATCATCGGCATCGATCGGGGCGAGAGAA

ACCTGCTGTACCTGGTCATGATCAATCAGAAGGGCGAGATCCT

GAAGCAGACCCTGCTGGACAGCATGCAGTCCGGCAAGGGCCG

GCCTGAGATCAACTACAAGGAGAAGCTGCAGGAGAAGGAGAT

CGAGAGGGATAAGGCCCGCAAGAGCTGGGGCACAGTGGAGA

ATATCAAGGAGCTGAAGGAGGGCTATCTGTCTATCGTGATCCA

CCAGATCAGCAAGCTGATGGTGGAGAACAATGCCATCGTGGT

GCTGGAGGACCTGAACATCGGCTTTAAGCGGGGCAGACAGAA

GGTGGAGCGGCAGGTGTACCAGAAGTTCGAGAAGATGCTGAT

CGATAAGCTGAACTTTCTGGTGTTCAAGGAGAATAAGCCAACC

GAGCCAGGAGGCGTGCTGAAGGCCTATCAGCTGACAGACGAG

TTTCAGTCTTTCGAGAAGCTGAGCAAGCAGACCGGCTTTCTGT
TCTACGTGCCAAGCTGGAACACCTCCAAGATCGACCCCAGAA
CAGGCTTTATCGATTTCCTGCACCCTGCCTACGAGAATATCGAG
AAGGCCAAGCAGTGGATCAACAAGTTTGATTCCATCAGGTTCA
ATTCTAAGATGGACTGGTTTGAGTTCACCGCCGATACACGCAA
GTTTTCCGAGAACCTGATGCTGGGCAAGAATCGGGTGTGGGT
CATCTGCACCACAAATGTGGAGCGGTACTTCACCAGCAAGAC
CGCCAACAGCTCCATCCAGTACAATAGCATCCAGATCACCGAG
AAGCTGAAGGAGCTGTTTGTGGACATCCCTTTCAGCAACGGC
CAGGATCTGAAGCCAGAGATCCTGAGGAAGAATGACGCCGTG
TTCTTTAAGAGCCTGCTGTTTTACATCAAGACCACACTGTCCC
TGCGCCAGAACAATGGCAAGAAGGGCGAGGAGGAGAAGGAC
TTCATCCTGAGCCCAGTGGTGGATTCCAAGGGCCGGTTCTTTA
ACTCTCTGGAGGCCAGCGACGATGAGCCCAAGGACGCCGATG
CCAATGGCGCCTACCACATCGCCCTGAAGGGCCTGATGAACCT
GCTGGTGCTGAATGAGACAAAGGAGGAGAACCTGAGCAGAC
CAAAGTGGAAGATCAAGAATAAGGACTGGCTGGAGTTCGTGT
GGGAGAGGAACCGC*AAAAGGCCGGCGGCCACGAAAAAGGCCG*
*GCCAGGCAAAAAAGAAAAAG*GGATCC**TACCCATACGATGTTCC**
AGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACC
CATATGATGTCCCCGACTATGCCTAAGAATTC（SEQ ID NO: 225）

[1694] 14-毛螺菌科細菌ND2006（LbCpf1）

[1695]

ATGAGCAAGCTGGAGAAGTTTACAAACTGCTACTCCCTGT
CTAAGACCCTGAGGTTCAAGGCCATCCCTGTGGGCAAGACCC
AGGAGAACATCGACAATAAGCGGCTGCTGGTGGAGGACGAGA
AGAGAGCCGAGGATTATAAGGGCGTGAAGAAGCTGCTGGATC
GCTACTATCTGTCTTTTATCAACGACGTGCTGCACAGCATCAA
GCTGAAGAATCTGAACAATTACATCAGCCTGTTCCGGAAGAA
AACCAGAACCGAGAAGGAGAATAAGGAGCTGGAGAACCTGG
AGATCAATCTGCGGAAGGAGATCGCCAAGGCCTTCAAGGGCA
ACGAGGGCTACAAGTCCCTGTTTAAGAAGGATATCATCGAGAC
AATCCTGCCAGAGTTCCTGGACGATAAGGACGAGATCGCCCTG
GTGAACAGCTTCAATGGCTTTACCACAGCCTTCACCGGCTTCT
TTGATAACAGAGAGAATATGTTTTCCGAGGAGGCCAAGAGCA
CATCCATCGCCTTCAGGTGTATCAACGAGAATCTGACCCGCTA
CATCTCTAATATGGACATCTTCGAGAAGGTGGACGCCATCTTT
GATAAGCACGAGGTGCAGGAGATCAAGGAGAAGATCCTGAAC
AGCGACTATGATGTGGAGGATTTCTTTGAGGGCGAGTTCTTTA
ACTTTGTGCTGACACAGGAGGGCATCGACGTGTATAACGCCAT
CATCGGCGGCTTCGTGACCGAGAGCGGCGAGAAGATCAAGGG
CCTGAACGAGTACATCAACCTGTATAATCAGAAAACCAAGCAG
AAGCTGCCTAAGTTTAAGCCACTGTATAAGCAGGTGCTGAGCG
ATCGGGAGTCTCTGAGCTTCTACGGCGAGGGCTATACATCCGA
TGAGGAGGTGCTGGAGGTGTTTAGAAACACCCTGAACAAGAA
CAGCGAGATCTTCAGCTCCATCAAGAAGCTGGAGAAGCTGTT

CAAGAATTTTGACGAGTACTCTAGCGCCGGCATCTTTGTGAAG

AACGGCCCCGCCATCAGCACAATCTCCAAGGATATCTTCGGCG

AGTGGAACGTGATCCGGGACAAGTGGAATGCCGAGTATGACG

ATATCCACCTGAAGAAGAAGGCCGTGGTGACCGAGAAGTACG

AGGACGATCGGAGAAAGTCCTTCAAGAAGATCGGCTCCTTTT

CTCTGGAGCAGCTGCAGGAGTACGCCGACGCCGATCTGTCTG

TGGTGGAGAAGCTGAAGGAGATCATCATCCAGAAGGTGGATG

AGATCTACAAGGTGTATGGCTCCTCTGAGAAGCTGTTCGACGC

CGATTTTGTGCTGGAGAAGAGCCTGAAGAAGAACGACGCCGT

GGTGGCCATCATGAAGGACCTGCTGGATTCTGTGAAGAGCTTC

GAGAATTACATCAAGGCCTTCTTTGGCGAGGGCAAGGAGACA

AACAGGGACGAGTCCTTCTATGGCGATTTTGTGCTGGCCTACG

ACATCCTGCTGAAGGTGGACCACATCTACGATGCCATCCGCAA

TTATGTGACCCAGAAGCCCTACTCTAAGGATAAGTTCAAGCTG

TATTTTCAGAACCCTCAGTTCATGGGCGGCTGGGACAAGGATA

AGGAGACAGACTATCGGGCCACCATCCTGAGATACGGCTCCA

AGTACTATCTGGCCATCATGGATAAGAAGTACGCCAAGTGCCT

GCAGAAGATCGACAAGGACGATGTGAACGGCAATTACGAGAA

GATCAACTATAAGCTGCTGCCCGGCCCTAATAAGATGCTGCCA

AAGGTGTTCTTTTCTAAGAAGTGGATGGCCTACTATAACCCCA

GCGAGGACATCCAGAAGATCTACAAGAATGGCACATTCAAGA

AGGGCGATATGTTTAACCTGAATGACTGTCACAAGCTGATCGA

CTTCTTTAAGGATAGCATCTCCCGGTATCCAAAGTGGTCCAATG

CCTACGATTTCAACTTTTCTGAGACAGAGAAGTATAAGGACAT

CGCCGGCTTTTACAGAGAGGTGGAGGAGCAGGGCTATAAGGT

GAGCTTCGAGTCTGCCAGCAAGAAGGAGGTGGATAAGCTGGT

GGAGGAGGGCAAGCTGTATATGTTCCAGATCTATAACAAGGAC

TTTTCCGATAAGTCTCACGGCACACCCAATCTGCACACCATGT

ACTTCAAGCTGCTGTTTGACGAGAACAATCACGGACAGATCA

GGCTGAGCGGAGGAGCAGAGCTGTTCATGAGGCGCGCCTCCC

TGAAGAAGGAGGAGCTGGTGGTGCACCCAGCCAACTCCCCTA

TCGCCAACAAGAATCCAGATAATCCCAAGAAAACCACAACCC

TGTCCTACGACGTGTATAAGGATAAGAGGTTTTCTGAGGACCA

GTACGAGCTGCACATCCCAATCGCCATCAATAAGTGCCCCAAG

AACATCTTCAAGATCAATACAGAGGTGCGCGTGCTGCTGAAGC

ACGACGATAACCCCTATGTGATCGGCATCGATAGGGGCGAGCG

CAATCTGCTGTATATCGTGGTGGTGGACGGCAAGGGCAACATC

GTGGAGCAGTATTCCCTGAACGAGATCATCAACAACTTCAACG

GCATCAGGATCAAGACAGATTACCACTCTCTGCTGGACAAGA

AGGAGAAGGAGAGGTTCGAGGCCCGCCAGAACTGGACCTCC

ATCGAGAATATCAAGGAGCTGAAGGCCGGCTATATCTCTCAGG

TGGTGCACAAGATCTGCGAGCTGGTGGAGAAGTACGATGCCG

TGATCGCCCTGGAGGACCTGAACTCTGGCTTTAAGAATAGCCG

CGTGAAGGTGGAGAAGCAGGTGTATCAGAAGTTCGAGAAGAT

GCTGATCGATAAGCTGAACTACATGGTGGACAAGAAGTCTAAT

CCTTGTGCAACAGGCGGCGCCCTGAAGGGCTATCAGATCACC

AATAAGTTCGAGAGCTTTAAGTCCATGTCTACCCAGAACGGCT

TCATCTTTTACATCCCTGCCTGGCTGACATCCAAGATCGATCCA

TCTACCGGCTTTGTGAACCTGCTGAAAACCAAGTATACCAGCA
TCGCCGATTCCAAGAAGTTCATCAGCTCCTTTGACAGGATCAT
GTACGTGCCCGAGGAGGATCTGTTCGAGTTTGCCCTGGACTAT
AAGAACTTCTCTCGCACAGACGCCGATTACATCAAGAAGTGG
AAGCTGTACTCCTACGGCAACCGGATCAGAATCTTCCGGAATC
CTAAGAAGAACAACGTGTTCGACTGGGAGGAGGTGTGCCTGA
CCAGCGCCTATAAGGAGCTGTTCAACAAGTACGGCATCAATTA
TCAGCAGGGCGATATCAGAGCCCTGCTGTGCGAGCAGTCCGA
CAAGGCCTTCTACTCTAGCTTTATGGCCCTGATGAGCCTGATGC
TGCAGATGCGGAACAGCATCACAGGCCGCACCGACGTGGATT
TTCTGATCAGCCCTGTGAAGAACTCCGACGGCATCTTCTACGA
TAGCCGGAACTATGAGGCCCAGGAGAATGCCATCCTGCCAAA
GAACGCCGACGCCAATGGCGCCTATAACATCGCCAGAAAGGT
GCTGTGGGCCATCGGCCAGTTCAAGAAGGCCGAGGACGAGAA
GCTGGATAAGGTGAAGATCGCCATCTCTAACAAGGAGTGGCTG
GAGTACGCCCAGACCAGCGTGAAGCAC*AAAAGGCCGGCGGCC*
*ACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAG*GGATCC**TACCC**
ATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTG
ATTATGCATACCCATATGATGTCCCCGACTATGCCTAAGAAT
TC （SEQ ID NO: 226）

[1696] 15-狗口腔卟啉單胞菌（PcCpf1）

[1697]

ATGGACAGCCTGAAGGATTTCACCAACCTGTACCCCGTGT

CCAAGACACTGCGGTTTGAGCTGAAGCCTGTGGGCAAGACCC

TGGAGAATATCGAGAAGGCCGGCATCCTGAAGGAGGATGAGC

ACAGAGCCGAGAGCTACCGGAGAGTGAAGAAGATCATCGATA

CATATCACAAGGTGTTCATCGACAGCTCCCTGGAGAACATGGC

CAAGATGGGCATCGAGAATGAGATCAAGGCCATGCTGCAGTC

CTTTTGCGAGCTGTATAAGAAGGACCACAGGACCGAGGGAGA

GGACAAGGCCCTGGATAAGATCAGGGCCGTGCTGAGGGGCCT

GATCGTGGGAGCCTTCACCGGCGTGTGCGGCCGGCGGGAGAA

CACAGTGCAGAATGAGAAGTATGAGAGCCTGTTTAAGGAGAA

GCTGATCAAGGAGATCCTGCCAGATTTCGTGCTGTCTACAGAG

GCCGAGTCCCTGCCCTTTTCTGTGGAGGAGGCCACCAGAAGC

CTGAAGGAGTTCGACTCCTTTACATCTTACTTCGCCGGCTTTTA

TGAGAACCGGAAGAATATCTACTCTACCAAGCCCCAGAGCAC

AGCCATCGCCTATAGACTGATCCACGAGAACCTGCCTAAGTTC

ATCGATAATATCCTGGTGTTTCAGAAGATCAAGGAGCCAATCG

CCAAGGAGCTGGAGCACATCAGGGCAGACTTCAGCGCCGGCG

GCTACATCAAGAAGGATGAGCGCCTGGAGGACATCTTTTCCCT

GAACTACTATATCCACGTGCTGTCTCAGGCCGGCATCGAGAAG

TACAATGCCCTGATCGGCAAGATCGTGACCGAGGGCGATGGC

GAGATGAAGGGCCTGAACGAGCACATCAACCTGTATAATCAGC

AGAGGGGCCGCGAGGACCGGCTGCCACTGTTCAGACCCCTGT

ATAAGCAGATCCTGTCTGATAGGGAGCAGCTGTCCTATCTGCC

AGAGTCTTTCGAGAAGGACGAGGAGCTGCTGAGGGCCCTGAA

GGAGTTTTACGATCACATCGCAGAGGACATCCTGGGAAGGAC

CCAGCAGCTGATGACAAGCATCTCCGAGTACGATCTGTCCCGG

ATCTATGTGAGAAACGATAGCCAGCTGACCGACATCTCCAAGA

AGATGCTGGGCGATTGGAATGCCATCTACATGGCCCGGGAGAG

AGCCTATGACCACGAGCAGGCCCCCAAGCGCATCACAGCCAA

GTACGAGAGGGACCGCATCAAGGCCCTGAAGGGCGAGGAGTC

TATCAGCCTGGCCAACCTGAACAGCTGCATCGCCTTCCTGGAC

AACGTGAGGGATTGTCGCGTGGACACCTATCTGTCTACACTGG

GACAGAAGGAGGGACCTCACGGCCTGAGCAACCTGGTGGAG

AACGTGTTCGCCTCCTACCACGAGGCCGAGCAGCTGCTGTCTT

TTCCCTATCCTGAGGAGAACAATCTGATCCAGGACAAGGATAA

CGTGGTGCTGATCAAGAACCTGCTGGATAATATCAGCGACCTG

CAGAGGTTCCTGAAGCCACTGTGGGGCATGGGCGATGAGCCC

GACAAGGATGAGAGGTTTTACGGCGAGTACAATTATATCAGGG

GCGCCCTGGACCAGGTCATCCCTCTGTATAACAAGGTGCGGAA

TTATCTGACCCGCAAGCCATACTCCACACGCAAGGTGAAGCTG

AACTTCGGCAATAGCCAGCTGCTGTCCGGCTGGGATAGGAAC

AAGGAGAAGGACAATTCTTGCGTGATCCTGCGCAAGGGCCAG

AACTTCTACCTGGCCATCATGAACAATCGGCACAAGCGGAGCT

TCGAGAATAAGATGCTGCCCGAGTATAAGGAGGGCGAGCCTTA

CTTCGAGAAGATGGATTATAAGTTTCTGCCAGACCCCAACAAG

ATGCTGCCCAAGGTGTTCCTGTCTAAGAAGGGCATCGAGATCT

ACAAGCCTAGCCCAAAGCTGCTGGAGCAGTATGGCCACGGCA

CCCACAAGAAGGGCGATACCTTCAGCATGGACGATCTGCACG

AGCTGATCGACTTCTTTAAGCACTCCATCGAGGCCCACGAGGA

TTGGAAGCAGTTCGGCTTTAAGTTCAGCGACACCGCCACATAC

GAGAACGTGAGCAGCTTCTACCGGGAGGTGGAGGACCAGGG

CTACAAGCTGTCTTTTAGAAAGGTGTCCGAGTCTTACGTGTAT

AGCCTGATCGATCAGGGCAAGCTGTACCTGTTCCAGATCTATA

ACAAGGACTTTAGCCCTTGTTCCAAGGGCACCCCAAATCTGCA

CACACTGTACTGGCGGATGCTGTTCGATGAGAGAAACCTGGC

CGACGTGATCTATAAGCTGGATGGCAAGGCCGAGATCTTCTTT

CGGGAGAAGTCCCTGAAGAATGACCACCCAACCCACCCTGCA

GGCAAGCCCATCAAGAAGAAGAGCCGGCAGAAGAAGGGCGA

GGAGAGCCTGTTCGAGTACGATCTGGTGAAGGACCGGAGATA

TACCATGGATAAGTTTCAGTTCCACGTGCCAATCACAATGAAC

TTTAAGTGCTCTGCCGGCAGCAAGGTGAACGACATGGTGAAT

GCCCACATCAGGGAGGCCAAGGACATGCACGTGATCGGCATC

GATAGGGGCGAGCGCAATCTGCTGTATATCTGCGTGATCGACA

GCCGCGGCACCATCCTGGATCAGATCTCCCTGAACACAATCAA

TGACATCGATTATCACGATCTGCTGGAGTCCAGGGACAAGGAT

CGCCAGCAGGAGCACAGGAACTGGCAGACCATCGAGGGCATC

AAGGAGCTGAAGCAGGGCTACCTGTCTCAGGCCGTGCACCGC

ATCGCCGAGCTGATGGTGGCCTATAAGGCCGTGGTGGCCCTGG

AGGACCTGAACATGGGCTTCAAGCGGGGCAGACAGAAGGTG

GAGAGCAGCGTGTACCAGCAGTTTGAGAAGCAGCTGATCGAC

AAGCTGAATTATCTGGTGGATAAGAAGAAGCGGCCCGAGGAC

ATCGGAGGCCTGCTGAGAGCCTACCAGTTCACCGCCCCTTTCA

AGAGCTTTAAGGAGATGGGCAAGCAGAACGGCTTTCTGTTCT

ATATCCCTGCCTGGAACACATCCAATATCGACCCAACCACAGG

CTTCGTGAACCTGTTTCACGTGCAGTACGAGAATGTGGATAAG

GCCAAGAGCTTCTTTCAGAAGTTCGACAGCATCTCCTACAACC

CTAAGAAGGATTGGTTTGAGTTCGCCTTTGACTATAAGAACTT

CACCAAGAAGGCCGAGGGCTCTAGGAGCATGTGGATTCTGTG

CACCCACGGCTCCCGGATCAAGAACTTCAGAAATTCTCAGAA

GAATGGCCAGTGGGATAGCGAGGAGTTTGCCCTGACCGAGGC

CTTCAAGTCCCTGTTTGTGCGGTACGAGATCGATTATACCGCC

GACCTGAAAACCGCCATCGTGGACGAGAAGCAGAAGGATTTC

TTTGTGGACCTGCTGAAGCTGTTCAAGCTGACCGTGCAGATGA

GAAACTCCTGGAAGGAGAAGGACCTGGATTACCTGATCTCTC

CAGTGGCCGGCGCCGATGGCAGGTTCTTTGACACACGCGAGG

GCAATAAGAGCCTGCCCAAGGACGCAGATGCAAACGGAGCCT

ATAATATCGCCCTGAAGGGCCTGTGGGCACTGAGGCAGATCAG

ACAGACCTCCGAGGGCGGCAAGCTGAAGCTGGCCATCTCTAA

CAAGGAGTGGCTGCAGTTTGTGCAGGAGAGATCCTACGAGAA

GGACAAAAGGCCGGCGGCCACG*AAAAAGGCCGGCCAGGCAAA*

*AAAGAAAAAG*GGATCC**TACCCATACGATGTTCCAGATTACGC**

TTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGT

CCCCGACTATGCCTAAGAATTC （SEQ ID NO: 227）

[1698] 16-解糖胨普雷沃菌（PdCpf1）

[1699]

ATGGAGAACTATCAGGAGTTCACCAACCTGTTTCAGCTGA

ATAAGACACTGAGATTCGAGCTGAAGCCCATCGGCAAGACCT

GCGAGCTGCTGGAGGAGGGCAAGATCTTCGCCAGCGGCTCCT

TTCTGGAGAAGGACAAGGTGAGGGCCGATAACGTGAGCTACG

TGAAGAAGGAGATCGACAAGAAGCACAAGATCTTTATCGAGG

AGACACTGAGCTCCTTCTCTATCAGCAACGATCTGCTGAAGCA

GTACTTTGACTGCTATAATGAGCTGAAGGCCTTCAAGAAGGAC

TGTAAGAGCGATGAGGAGGAGGTGAAGAAAACCGCCCTGCGC

AACAAGTGTACCTCCATCCAGAGGGCCATGCGCGAGGCCATCT

CTCAGGCCTTTCTGAAGAGCCCCCAGAAGAAGCTGCTGGCCA

TCAAGAACCTGATCGAGAACGTGTTCAAGGCCGACGAGAATG

TGCAGCACTTCTCCGAGTTTACCAGCTATTTCTCCGGCTTTGA

GACAAACAGAGAGAATTTCTACTCTGACGAGGAGAAGTCCAC

ATCTATCGCCTATAGGCTGGTGCACGATAACCTGCCTATCTTCA

TCAAGAACATCTACATCTTCGAGAAGCTGAAGGAGCAGTTCG

ACGCCAAGACCCTGAGCGAGATCTTCGAGAACTACAAGCTGT

ATGTGGCCGGCTCTAGCCTGGATGAGGTGTTCTCCCTGGAGTA

CTTTAACAATACCCTGACACAGAAGGGCATCGACAACTATAAT

GCCGTGATCGGCAAGATCGTGAAGGAGGATAAGCAGGAGATC

CAGGGCCTGAACGAGCACATCAACCTGTATAATCAGAAGCAC

AAGGACCGGAGACTGCCCTTCTTTATCTCCCTGAAGAAGCAG

ATCCTGTCCGATCGGGAGGCCCTGTCTTGGCTGCCTGACATGT

TCAAGAATGATTCTGAAGTGATCAAGGCCCTGAAGGGCTTCTA

CATCGAGGACGGCTTTGAGAACAATGTGCTGACACCTCTGGC

CACCCTGCTGTCCTCTCTGGATAAGTACAACCTGAATGGCATC

TTTATCCGCAACAATGAGGCCCTGAGCTCCCTGTCCCAGAACG

TGTATCGGAATTTTTCTATCGACGAGGCCATCGATGCCAACGC

CGAGCTGCAGACCTTCAACAATTACGAGCTGATCGCCAATGCC

CTGCGCGCCAAGATCAAGAAGGAGACAAAGCAGGGCCGGAA

GTCTTTCGAGAAGTACGAGGAGTATATCGATAAGAAGGTGAAG

GCCATCGACAGCCTGTCCATCCAGGAGATCAACGAGCTGGTG

GAGAATTACGTGAGCGAGTTTAACTCTAATAGCGGCAACATGC

CAAGAAAGGTGGAGGACTACTTCAGCCTGATGAGGAAGGGCG

ACTTCGGCTCCAACGATCTGATCGAAAATATCAAGACCAAGCT

GAGCGCCGCAGAGAAGCTGCTGGGCACAAAGTACCAGGAGA

CAGCCAAGGACATCTTCAAGAAGGATGAGAACTCCAAGCTGA

TCAAGGAGCTGCTGGACGCCACCAAGCAGTTCCAGCACTTTA

TCAAGCCACTGCTGGGCACAGGCGAGGAGGCAGATCGGGACC

TGGTGTTCTACGGCGATTTTCTGCCCCTGTATGAGAAGTTTGA

GGAGCTGACCCTGCTGTATAACAAGGTGCGGAATAGACTGAC

ACAGAAGCCCTATTCCAAGGACAAGATCCGCCTGTGCTTCAAC

AAGCCTAAGCTGATGACAGGCTGGGTGGATTCCAAGACCGAG

AAGTCTGACAACGGCACACAGTACGGCGGCTATCTGTTTCGG

AAGAAGAATGAGATCGGCGAGTACGATTATTTTCTGGGCATCT

CTAGCAAGGCCCAGCTGTTCAGAAAGAACGAGGCCGTGATCG

GCGACTACGAGAGGCTGGATTACTATCAGCCAAAGGCCAATAC

CATCTACGGCTCTGCCTATGAGGGCGAGAACAGCTACAAGGA

GGACAAGAAGCGGCTGAACAAAGTGATCATCGCCTATATCGA

GCAGATCAAGCAGACAAACATCAAGAAGTCTATCATCGAGTC

CATCTCTAAGTATCCTAATATCAGCGACGATGACAAGGTGACC

CCATCCTCTCTGCTGGAGAAGATCAAGAAGGTGTCTATCGACA

GCTACAACGGCATCCTGTCCTTCAAGTCTTTTCAGAGCGTGAA

CAAGGAAGTGATCGATAACCTGCTGAAAACCATCAGCCCCCT

GAAGAACAAGGCCGAGTTTCTGGACCTGATCAATAAGGATTAT

CAGATCTTCACCGAGGTGCAGGCCGTGATCGACGAGATCTGC

AAGCAGAAAACCTTCATCTACTTTCCAATCTCCAACGTGGAGC

TGGAGAAGGAGATGGGCGATAAGGACAAGCCCCTGTGCCTGT

TCCAGATCAGCAATAAGGATCTGTCCTTCGCCAAGACCTTTAG

CGCCAACCTGCGGAAGAAGAGAGGCGCCGAGAATCTGCACAC

AATGCTGTTTAAGGCCCTGATGGAGGGCAACCAGGATAATCTG

GACCTGGGCTCTGGCGCCATCTTCTACAGAGCCAAGAGCCTG

GACGGCAACAAGCCCACACACCCTGCCAATGAGGCCATCAAG

TGTAGGAACGTGGCCAATAAGGATAAGGTGTCCCTGTTCACCT

ACGACATCTATAAGAACAGGCGCTACATGGAGAATAAGTTCCT

GTTTCACCTGAGCATCGTGCAGAACTATAAGGCCGCCAATGAC

TCCGCCCAGCTGAACAGCTCCGCCACCGAGTATATCAGAAAG

GCCGATGACCTGCACATCATCGGCATCGATAGGGGCGAGCGCA

ATCTGCTGTACTATTCCGTGATCGATATGAAGGGCAACATCGTG

GAGCAGGACTCTCTGAATATCATCAGGAACAATGACCTGGAG

ACAGATTACCACGACCTGCTGGATAAGAGGGAGAAGGAGCGC

AAGGCCAACCGGCAGAATTGGGAGGCCGTGGAGGGCATCAAG

GACCTGAAGAAGGGCTACCTGAGCCAGGCCGTGCACCAGATC

GCCCAGCTGATGCTGAAGTATAACGCCATCATCGCCCTGGAGG

ATCTGGGCCAGATGTTTGTGACCCGCGGCCAGAAGATCGAGA

AGGCCGTGTACCAGCAGTTCGAGAAGAGCCTGGTGGATAAGC

TGTCCTACCTGGTGGACAAGAAGCGGCCTTATAATGAGCTGGG

CGGCATCCTGAAGGCCTACCAGCTGGCCTCTAGCATCACCAAG
AACAATTCTGACAAGCAGAACGGCTTCCTGTTTTATGTGCCAG
CCTGGAATACAAGCAAGATCGATCCCGTGACCGGCTTTACAGA
CCTGCTGCGGCCCAAGGCCATGACCATCAAGGAGGCCCAGGA
CTTCTTTGGCGCCTTCGATAACATCTCTTACAATGACAAGGGC
TATTTCGAGTTTGAGACAAACTACGACAAGTTTAAGATCAGAA
TGAAGAGCGCCCAGACCAGGTGGACAATCTGCACCTTCGGCA
ATCGGATCAAGAGAAAGAAGGATAAGAACTACTGGAATTATG
AGGAGGTGGAGCTGACCGAGGAGTTCAAGAAGCTGTTTAAGG
ACAGCAACATCGATTACGAGAACTGTAATCTGAAGGAGGAGA
TCCAGAACAAGGACAATCGCAAGTTCTTTGATGACCTGATCAA
GCTGCTGCAGCTGACACTGCAGATGCGGAACTCCGATGACAA
GGGCAATGATTATATCATCTCTCCTGTGGCCAACGCCGAGGGC
CAGTTCTTTGACTCCCGCAATGGCGATAAGAAGCTGCCACTGG
ATGCAGACGCAAACGGAGCCTACAATATCGCCCGCAAGGGCC
TGTGGAACATCCGGCAGATCAAGCAGACCAAGAACGACAAGA
AGCTGAATCTGAGCATCTCCTCTACAGAGTGGCTGGATTTCGT
GCGGGAGAAGCCTTACCTGAAG*AAAAGGCCGGCGGCCACGAA*
*AAAGGCCGGCCAGGCAAAAAAGAAAAA*GGGATCC**TACCCATAC**
GATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTA
TGCATACCCATATGATGTCCCCGACTATGCCTAAGAATTC
（SEQ ID NO: 228）

[1700]　17-獼猴卟啉單胞菌（PmCpf1）

[1701]

ATGAAAACCCAGCACTTCTTTGAGGACTTCACAAGCCTGT
ACTCTCTGAGCAAGACCATCCGGTTTGAGCTGAAGCCAATCG
GCAAGACCCTGGAGAACATCAAGAAGAATGGCCTGATCCGGA
GAGATGAGCAGAGACTGGACGATTACGAGAAGCTGAAGAAA
GTGATCGACGAGTATCACGAGGATTTCATCGCCAACATCCTGA
GCTCCTTTTCCTTCTCTGAGGAGATCCTGCAGTCCTACATCCA
GAATCTGAGCGAGTCCGAGGCCAGGGCCAAGATCGAGAAAAC
CATGCGCGACACACTGGCCAAGGCCTTCTCTGAGGATGAGAG
GTACAAGAGCATCTTTAAGAAGGAGCTGGTGAAGAAGGACAT
CCCCGTGTGGTGCCCTGCCTATAAGAGCCTGTGCAAGAAGTTC
GATAACTTTACCACATCTCTGGTGCCCTTCCACGAGAACAGGA
AGAACCTGTATACCAGCAATGAGATCACAGCCTCTATCCCTTAT
CGCATCGTGCACGTGAACCTGCCAAAGTTTATCCAGAATATCG
AGGCCCTGTGCGAGCTGCAGAAGAAGATGGGCGCCGACCTGT
ACCTGGAGATGATGGAGAACCTGCGCAACGTGTGGCCCAGCT
TCGTGAAAACCCCAGACGACCTGTGCAACCTGAAAACCTATA
ATCACCTGATGGTGCAGTCTAGCATCAGCGAGTACAACAGGTT
TGTGGGCGGCTATTCCACCGAGGACGGCACAAAGCACCAGGG
CATCAACGAGTGGATCAATATCTACAGACAGAGGAATAAGGAG
ATGCGCCTGCCTGGCCTGGTGTTCCTGCACAAGCAGATCCTGG
CCAAGGTGGACTCCTCTAGCTTCATCAGCGATACACTGGAGAA
CGACGATCAGGTGTTTTGCGTGCTGAGACAGTTCAGGAAGCT
GTTTTGGAATACCGTGTCCTCTAAGGAGGACGATGCCGCCTCC
CTGAAGGACCTGTTCTGTGGCCTGTCTGGCTATGACCCTGAGG
CCATCTACGTGAGCGATGCCCACCTGGCCACAATCTCCAAGAA

CATCTTTGACAGATGGAATTACATCTCCGATGCCATCAGGCGC

AAGACCGAGGTGCTGATGCCACGGAAGAAGGAGAGCGTGGA

GAGATATGCCGAGAAGATCTCCAAGCAGATCAAGAAGAGACA

GTCTTACAGCCTGGCCGAGCTGGACGATCTGCTGGCCCACTAT

AGCGAGGAGTCCCTGCCCGCAGGCTTCTCTCTGCTGAGCTACT

TTACATCTCTGGGCGGCCAGAAGTATCTGGTGAGCGACGGCGA

AGTGATCCTGTACGAGGAGGGCAGCAACATCTGGGACGAGGT

GCTGATCGCCTTCAGGGATCTGCAGGTCATCCTGGACAAGGAC

TTCACCGAGAAGAAGCTGGGCAAGGATGAGGAGGCCGTGTCT

GTGATCAAGAAGGCCCTGGACAGCGCCCTGCGCCTGCGGAAG

TTCTTTGATCTGCTGTCCGGCACAGGCGCAGAGATCAGGAGA

GACAGCTCCTTCTATGCCCTGTATACCGACCGGATGGATAAGC

TGAAGGGCCTGCTGAAGATGTATGATAAGGTGAGAAACTACCT

GACCAAGAAGCCTTATTCCATCGAGAAGTTCAAGCTGCACTTT

GACAACCCATCCCTGCTGTCTGGCTGGGATAAGAATAAGGAGC

TGAACAATCTGTCTGTGATCTTCCGGCAGAACGGCTACTATTA

CCTGGGCATCATGACACCCAAGGGCAAGAATCTGTTCAAGAC

CCTGCCTAAGCTGGGCGCCGAGGAGATGTTTTATGAGAAGATG

GAGTACAAGCAGATCGCCGAGCCTATGCTGATGCTGCCAAAG

GTGTTCTTTCCCAAGAAAACCAAGCCAGCCTTCGCCCCAGAC

CAGAGCGTGGTGGATATCTACAACAAGAAAACCTTCAAGACA

GGCCAGAAGGGCTTTAATAAGAAGGACCTGTACCGGCTGATC

GACTTCTACAAGGAGGCCCTGACAGTGCACGAGTGGAAGCTG

TTTAACTTCTCCTTTTCTCCAACCGAGCAGTATCGGAATATCGG

CGAGTTCTTTGACGAGGTGAGAGAGCAGGCCTACAAGGTGTC

CATGGTGAACGTGCCCGCCTCTTATATCGACGAGGCCGTGGAG

AACGGCAAGCTGTATCTGTTCCAGATCTACAATAAGGACTTCA

GCCCCTACTCCAAGGGCATCCCTAACCTGCACACACTGTATTG

GAAGGCCCTGTTCAGCGAGCAGAATCAGAGCCGGGTGTATAA

GCTGTGCGGAGGAGGAGAGCTGTTTTATAGAAAGGCCAGCCT

GCACATGCAGGACACCACAGTGCACCCCAAGGGCATCTCTATC

CACAAGAAGAACCTGAATAAGAAGGGCGAGACAAGCCTGTTC

AACTACGACCTGGTGAAGGATAAGAGGTTTACCGAGGACAAG

TTCTTTTTCCACGTGCCTATCTCTATCAACTACAAGAATAAGAA

GATCACCAACGTGAATCAGATGGTGCGCGATTATATCGCCCAG

AACGACGATCTGCAGATCATCGGCATCGACCGCGGCGAGCGG

AATCTGCTGTATATCAGCCGGATCGATACAAGGGGCAACCTGC

TGGAGCAGTTCAGCCTGAATGTGATCGAGTCCGACAAGGGCG

ATCTGAGAACCGACTATCAGAAGATCCTGGGCGATCGCGAGC

AGGAGCGGCTGAGGCGCCGGCAGGAGTGGAAGTCTATCGAGA

GCATCAAGGACCTGAAGGATGGCTACATGAGCCAGGTGGTGC

ACAAGATCTGTAACATGGTGGTGGAGCACAAGGCCATCGTGG

TGCTGGAGAACCTGAATCTGAGCTTCATGAAGGGCAGGAAGA

AGGTGGAGAAGTCCGTGTACGAGAAGTTTGAGCGCATGCTGG

TGGACAAGCTGAACTATCTGGTGGTGGATAAGAAGAACCTGT

CCAATGAGCCAGGAGGCCTGTATGCAGCATACCAGCTGACCA

ATCCACTGTTCTCTTTTGAGGAGCTGCACAGATACCCCCAGAG

CGGCATCCTGTTTTTCGTGGACCCATGGAACACCTCTCTGACA

GATCCCAGCACAGGCTTCGTGAATCTGCTGGGCAGAATCAACT

ACACCAATGTGGGCGACGCCCGCAAGTTTTTCGATCGGTTTAA

CGCCATCAGATATGACGGCAAGGGCAATATCCTGTTCGACCTG
GATCTGTCCAGATTTGATGTGAGGGTGGAGACACAGAGGAAG
CTGTGGACACTGACCACATTCGGCTCTCGCATCGCCAAATCCA
AGAAGTCTGGCAAGTGGATGGTGGAGCGGATCGAGAACCTGA
GCCTGTGCTTTCTGGAGCTGTTCGAGCAGTTTAATATCGGCTA
CAGAGTGGAGAAGGACCTGAAGAAGGCCATCCTGAGCCAGG
ATAGGAAGGAGTTCTATGTGCGCCTGATCTACCTGTTTAACCT
GATGATGCAGATCCGGAACAGCGACGGCGAGGAGGATTATATC
CTGTCTCCCGCCCTGAACGAGAAGAATCTGCAGTTCGACAGC
AGGCTGATCGAGGCCAAGGATCTGCCTGTGGACGCAGATGCA
AACGGAGCATACAATGTGGCCCGCAAGGGCCTGATGGTGGTG
CAGAGAATCAAGAGGGGCGACCACGAGTCCATCCACAGGATC
GGAAGGGCACAGTGGCTGAGATATGTGCAGGAGGGCATCGTG
GAG*AAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAA*
*AGAAAAAG*GGATCC**TACCCATACGATGTTCCAGATTACGCTT**
ATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTC
CCCGACTATGCCTAAGAATTC （SEQ ID NO: 229）

[1702] **人類密碼子優化的Cpf1異種同源物的胺基酸序列**

[1703] *核定位信號（NLS）*

[1704] 甘胺酸-絲胺酸接頭

[1705] **3× HA標籤**

[1706] **1-土拉熱弗朗西絲菌新殺手亞種U112（FnCpf1）**

[1707]

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEK
RAKDYKKAKQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLK
KSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAK
KGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT
TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESL
KDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEI
ANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQI
NDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLT
DLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKT
EKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDE
IAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTN
NLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLY
NKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIK
DDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGAN
KMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFE
FNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYRE
VENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRP
NLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPA
KEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKS
SGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGN
IIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIK
EMKEGYLSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKVEKQ
VYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKK
MGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSK

FDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFR
NSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGES
DKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSR
QAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIK
NEEYFEFVQNRNN*KRPAATKKAGQAKKKK*GS**YPYDVPDYAYPYD**
VPDYAYPYDVPDYA（SEQ ID NO: 230）

[1708]　3-毛螺菌科細菌MC2017（Lb3Cpf1）

[1709]

MDYGNGQFERRAPLTKTITLRLKPIGETRETIREQKLLEQDA
AFRKLVETVTPIVDDCIRKIADNALCHFGTEYDFSCLGNAISKND
SKAIKKETEKVEKLLAKVLTENLPDGLRKVNDINSAAFIQDTLTS
FVQDDADKRVLIQELKGKTVLMQRFLTTRITALTVWLPDRVFEN
FNIFIENAEKMRILLDSPLNEKIMKFDPDAEQYASLEFYGQCLSQ
KDIDSYNLIISGIYADDEVKNPGINEIVKEYNQQIRGDKDESPLPK
LKKLHKQILMPVEKAFFVRVLSNDSDARSILEKILKDTEMLPSKII
EAMKEADAGDIAVYGSRLHELSHVIYGDHGKLSQIIYDKESKRIS
ELMETLSPKERKESKKRLEGLEEHIRKSTYTFDELNRYAEKNVM
AAYIAAVEESCAEIMRKEKDLRTLLSKEDVKIRGNRHNTLIVKN
YFNAWTVFRNLIRILRRKSEAEIDSDFYDVLDDSVEVLSLTYKGE
NLCRSYITKKIGSDLKPEIATYGSALRPNSRWWSPGEKFNVKFHT
IVRRDGRLYYFILPKGAKPVELEDMDGDIECLQMRKIPNPTIFLP
KLVFKDPEAFFRDNPEADEFVFLSGMKAPVTITRETYEAYRYKL
YTVGKLRDGEVSEEEYKRALLQVLTAYKEFLENRMIYADLNFGF
KDLEEYKDSSEFIKQVETHNTFMCWAKVSSSQLDDLVKSGNGLL

FEIWSERLESYYKYGNEKVLRGYEGVLLSILKDENLVSMRTLLN
SRPMLVYRPKESSKPMVVHRDGSRVVDRFDKDGKYIPPEVHDE
LYRFFNNLLIKEKLGEKARKILDNKKVKVKVLESERVKWSKFYD
EQFAVTFSVKKNADCLDTTKDLNAEVMEQYSESNRLILIRNTTDI
LYYLVLDKNGKVLKQRSLNIINDGARDVDWKERFRQVTKDRNE
GYNEWDYSRTSNDLKEVYLNYALKEIAEAVIEYNAILIIEKMSNA
FKDKYSFLDDVTFKGFETKLLAKLSDLHFRGIKDGEPCSFTNPLQ
LCQNDSNKILQDGVIFMVPNSMTRSLDPDTGFIFAINDHNIRTKK
AKLNFLSKFDQLKVSSEGCLIMKYSGDSLPTHNTDNRVWNCCC
NHPITNYDRETKKVEFIEEPVEELSRVLEENGIETDTELNKLNER
ENVPGKVVDAIYSLVLNYLRGTVSGVAGQRAVYYSPVTGKKYDI
SFIQAMNLNRKCDYYRIGSKERGEWTDFVAQLIN*KRPAATKKAG*
*QAKKKK*GS**YPYDVPDYAYPYDVPDYAYPYDVPDYA**(SEQ ID NO:
231)

[1710] 4-解朊丁酸弧菌（BpCpf1）

[1711]

MLLYENYTKRNQITKSLRLELRPQGKTLRNIKELNLLEQDK
AIYALLERLKPVIDEGIKDIARDTLKNCELSFEKLYEHFLSGDKK
AYAKESERLKKEIVKTLIKNLPEGIGKISEINSAKYLNGVLYDFID
KTHKDSEEKQNILSDILETKGYLALFSKFLTSRITTLEQSMPKRVI
ENFEIYAANIPKMQDALERGAVSFAIEYESICSVDYYNQILSQEDI
DSYNRLISGIMDEDGAKEKGINQTISEKNIKIKSEHLEEKPFRILK
QLHKQILEEREKAFTIDHIDSDEEVVQVTKEAFEQTKEQWENIK
KINGFYAKDPGDITLFIVVGPNQTHVLSQLIYGEHDRIRLLLEEYE

KNTLEVLPRRTKSEKARYDKFVNAVPKKVAKESHTFDGLQKMT
GDDRLFILYRDELARNYMRIKEAYGTFERDILKSRRGIKGNRDV
QESLVSFYDELTKFRSALRIINSGNDEKADPIFYNTFDGIFEKANR
TYKAENLCRNYVTKSPADDARIMASCLGTPARLRTHWWNGEEN
FAINDVAMIRRGDEYYYFVLTPDVKPVDLKTKDETDAQIFVQRK
GAKSFLGLPKALFKCILEPYFESPEHKNDKNCVIEEYVSKPLTID
RRAYDIFKNGTFKKTNIGIDGLTEEKFKDDCRYLIDVYKEFIAVY
TRYSCFNMSGLKRADEYNDIGEFFSDVDTRLCTMEWIPVSFERI
NDMVDKKEGLLFLVRSMFLYNRPRKPYERTFIQLFSDSNMEHTS
MLLNSRAMIQYRAASLPRRVTHKKGSILVALRDSNGEHIPMHIRE
AIYKMKNNFDISSEDFIMAKAYLAEHDVAIKKANEDIIRNRRYTE
DKFFLSLSYTKNADISARTLDYINDKVEEDTQDSRMAVIVTRNL
KDLTYVAVVDEKNNVLEEKSLNEIDGVNYRELLKERTKIKYHDK
TRLWQYDVSSKGLKEAYVELAVTQISKLATKYNAVVVVESMSST
FKDKFSFLDEQIFKAFEARLCARMSDLSFNTIKEGEAGSISNPIQV
SNNNGNSYQDGVIYFLNNAYTRTLCPDTGFVDVFDKTRLITMQS
KRQFFAKMKDIRIDDGEMLFTFNLEEYPTKRLLDRKEWTVKIAG
DGSYFDKDKGEYVYVNDIVREQIIPALLEDKAVFDGNMAEKFLD
KTAISGKSVELIYKWFANALYGIITKKDGEKIYRSPITGTEIDVSK
NTTYNFGKKFMFKQEYRGDGDFLDAFLNYMQAQDIAV*KRPAAT*
*KKAGQAKKKK*GS**YPYDVPDYAYPYDVPDYAYPYDVPDYA**

（SEQ ID NO: 232）

[1712] 5-佩萊格裡尼菌科細菌GW2011_GWA_33_10（PeCpf1）

[1713]

MSNFFKNFTNLYELSKTLRFELKPVGDTLTNMKDHLEYDEK
LQTFLKDQNIDDAYQALKPQFDEIHEEFITDSLESKKAKEIDFSE
YLDLFQEKKELNDSEKKLRNKIGETFNKAGEKWKKEKYPQYEW
KKGSKIANGADILSCQDMLQFIKYKNPEDEKIKNYIDDTLKGFFT
YFGGFNQNRANYYETKKEASTAVATRIVHENLPKFCDNVIQFKHI
IKRKKDGTVEKTERKTEYLNAYQYLKNNNKITQIKDAETEKMIE
STPIAEKIFDVYYFSSCLSQKQIEEYNRIIGHYNLLINLYNQAKRS
EGKHLSANEKKYKDLPKFKTLYKQIGCGKKKDLFYTIKCDTEEE
ANKSRNEGKESHSVEEIINKAQEAINKYFKSNNDCENINTVPDFI
NYILTKENYEGVYWSKAAMNTISDKYFANYHDLQDRLKEAKVF
QKADKKSEDDIKIPEAIELSGLFGVLDSLADWQTTLFKSSILSNE
DKLKIITDSQTPSEALLKMIFNDIEKNMESFLKETNDIITLKKYKG
NKEGTEKIKQWFDYTLAINRMLKYFLVKENKIKGNSLDTNISEA
LKTLIYSDDAEWFKWYDALRNYLTQKPQDEAKENKLKLNFDNP
SLAGGWDVNKECSNFCVILKDKNEKKYLAIMKKGENTLFQKEW
TEGRGKNLTKKSNPLFEINNCEILSKMEYDFWADVSKMIPKCSTQ
LKAVVNHFKQSDNEFIFPIGYKVTSGEKFREECKISKQDFELNNKV
FNKNELSVTAMRYDLSSTQEKQYIKAFQKEYWELLFKQEKRDTK
LTNNEIFNEWINFCNKKYSELLSWERKYKDALTNWINFCKYFLSK
YPKTTLFNYSFKESENYNSLDEFYRDVDICSYKLNINTTINKSILDR
LVEEGKLYLFEIKNQDSNDGKSIGHKNNLHTIYWNAIFENFDNRP
KLNGEAEIFYRKAISKDKLGIVKGKKTKNGTEIIKNYRFSKEKFIL

HVPITLNFCSNNEYVNDIVNTKFYNFSNLHFLGIDRGEKHLAYYS
LVNKNGEIVDQGTLNLPFTDKDGNQRSIKKEKYFYNKQEDKWE
AKEVDCWNYNDLLDAMASNRDMARKNWQRIGTIKEAKNGYVS
LVIRKIADLAVNNERPAFIVLEDLNTGFKRSRQKIDKSVYQKFEL
ALAKKLNFLVDKNAKRDEIGSPTKALQLTPPVNNYGDIENKKQA
GIMLYTRANYTSQTDPATGWRKTIYLKAGPEETTYKKDGKIKNK
SVKDQIIETFTDIGFDGKDYYFEYDKGEFVDEKTGEIKPKKWRLY
SGENGKSLDRFRGEREKDKYEWKIDKIDIVKILDDLFVNFDKNIS
LLKQLKEGVELTRNNEHGTGESLRFAINLIQQIRNTGNNERDNDF
ILSPVRDENGKHFDSREYWDKETKGEKISMPSSGDANGAFNIAR
KGIIMNAHILANSDSKDLSLFVSDEEWDLHLNNKTEWKKQLNIF
SSRKAMA*KRKKKRPAATKKAGQAKKKK*GS**YPYDVPDYAYPYDVP**
DYAYPYDVPDYA（SEQ ID NO: 233）

[1714] 6-儉菌總門細菌GWC2011_GWC2_44_17（PbCpf1）

[1715]

MENIFDQFIGKYSLSKTLRFELKPVGKTEDFLKINKVFEKDQ
TIDDSYNQAKFYFDSLHQKFIDAALASDKTSELSFQNFADVLEKQ
NKIILDKKREMGALRKRDKNAVGIDRLQKEINDAEDIIQKEKEKI
YKDVRTLFDNEAESWKTYYQEREVDGKKITFSKADLKQKGADF
LTAAGILKVLKYEFPEEKEKEFQAKNQPSLFVEEKENPGQKRYIF
DSFDKFAGYLTKFQQTKKNLYAADGTSTAVATRIADNFIIFHQNT
KVFRDKYKNNHTDLGFDEENIFEIERYKNCLLQREIEHIKNENSY

NKIIGRINKKIKEYRDQKAKDTKLTKSDFPFFKNLDKQILGEVEK
EKQLIEKTREKTEEDVLIERFKEFIENNEERFTAAKKLMNAFCNG
EFESEYEGIYLKNKAINTISRRWFVSDRDFELKLPQQKSKNKSEK
NEPKVKKFISIAEIKNAVEELDGDIFKAVFYDKKIIAQGGSKLEQF
LVIWKYEFEYLFRDIERENGEKLLGYDSCLKIAKQLGIFPQEKEA
REKATAVIKNYADAGLGIFQMMKYFSLDDKDRKNTPGQLSTNFY
AEYDGYYKDFEFIKYYNEFRNFITKKPFDEDKIKLNFENGALLK
GWDENKEYDFMGVILKKEGRLYLGIMHKNHRKLFQSMGNAKG
DNANRYQKMIYKQIADASKDVPRLLLTSKKAMEKFKPSQEILRI
KKEKTFKRESKNFSLRDLHALIEYYRNCIPQYSNWSFYDFQFQD
TGKYQNIKEFTDDVQKYGYKISFRDIDDEYINQALNEGKMYLFE
VVNKDIYNTKNGSKNLHTLYFEHILSAENLNDPVFKLSGMAEIF
QRQPSVNEREKITTQKNQCILDKGDRAYKYRRYTEKKIMFHMSL
VLNTGKGEIKQVQFNKIINQRISSSDNEMRVNVIGIDRGEKNLLY
YSVVKQNGEIIEQASLNEINGVNYRDKLIEREKERLKNRQSWKP
VVKIKDLKKGYISHVIHKICQLIEKYSAIVVLEDLNMRFKQIRGG
IERSVYQQFEKALIDKLGYLVFKDNRDLRAPGGVLNGYQLSAPF
VSFEKMRKQTGILFYTQAEYTSKTDPITGFRKNVYISNSASLDKI
KEAVKKFDAIGWDGKEQSYFFKYNPYNLADEKYKNSTVSKEWA
IFASAPRIRRQKGEDGYWKYDRVKVNEEFEKLLKVWNFVNPKA
TDIKQEIIKKEKAGDLQGEKELDGRLRNFWHSFIYLFNLVLELRN
SFSLQIKIKAGEVIAVDEGVDFIASPVKPFFTTPNPYIPSNLCWLA
VENADANGAYNIARKGVMILKKIREHAKKDPEFKKLPNLFISNA
EWDEAARDWGKYAGTTALNLDH*KRPAATKKAGQAKKKK*GS**YPY**

DVPDYAYPYDVPDYAYPYDVPDYA （SEQ ID NO: 234）

[1716] 7-密斯氏菌屬某種SC_K08D17（SsCpf1）

[1717]

MQTLFENFTNQYPVSKTLRFELIPQGKTKDFIEQKGLLKKD
EDRAEKYKKVKNIIDEYHKDFIEKSLNGLKLDGLEKYKTLYLKQ
EKDDKDKKAFDKEKENLRKQIANAFRNNEKFKTLFAKELIKNDL
MSFACEEDKKNVKEFEAFTTYFTGFHQNRANMYVADEKRTAIAS
RLIHENLPKFIDNIKIFEKMKKEAPELLSPFNQTLKDMKDVIKGT
TLEEIFSLDYFNKTLTQSGIDIYNSVIGGRTPEEGKTKIKGLNEYI
NTDFNQKQTDKKKRQPKFKQLYKQILSDRQSLSFIAEAFKNDTEI
LEAIEKFYVNELLHFSNEGKSTNVLDAIKNAVSNLESFNLTKMYF
RSGASLTDVSRKVFGEWSIINRALDNYYATTYPIKPREKSEKYEE
RKEKWLKQDFNVSLIQTAIDEYDNETVKGKNSGKVIADYFAKFC
DDKETDLIQKVNEGYIAVKDLLNTPCPENEKLGSNKDQVKQIKA
FMDSIMDIMHFVRPLSLKDTDKEKDETFYSLFTPLYDHLTQTIAL
YNKVRNYLTQKPYSTEKIKLNFENSTLLGGWDLNKETDNTAIIL
RKDNLYYLGIMDKRHNRIFRNVPKADKKDFCYEKMVYKLLPGA
NKMLPKVFFSQSRIQEFTPSAKLLENYANETHKKGDNFNLNHCH
KLIDFFKDSINKHEDWKNFDFRFSATSTYADLSGFYHEVEHQGY
KISFQSVADSFIDDLVNEGKLYLFQIYNKDFSPFSKGKPNLHTLY
WKMLFDENNLKDVVYKLNGEAEVFYRKKSIAEKNTTIHKANES
IINKNPDNPKATSTFNYDIVKDKRYTIDKFQFHIPITMNFKAEGIF
NMNQRVNQFLKANPDINIIGIDRGERHLLYYALINQKGKILKQDT

LNVIANEKQKVDYHNLLDKKEGDRATARQEWGVIETIKELKEG
YLSQVIHKLTDLMIENNAIIVMEDLNFGFKRGRQKVEKQVYQKF
EKMLIDKLNYLVDKNKKANELGGLLNAFQLANKFESFQKMGK
QNGFIFYVPAWNTSKTDPATGFIDFLKPRYENLNQAKDFFEKFDS
IRLNSKADYFEFAFDFKNFTEKADGGRTKWTVCTTNEDRYAWN
RALNNNRGSQEKYDITAELKSLFDGKVDYKSGKDLKQQIASQES
ADFFKALMKNLSITLSLRHNNGEKGDNEQDYILSPVADSKGRFF
DSRKADDDMPKNADANGAYHIALKGLWCLEQISKTDDLKKVKL
AISNKEWLEFVQTLKG*KRPAATKKAGQAKKKK*GS**YPYDVPDYAY**
PYDVPDYAYPYDVPDYA（SEQ ID NO: 235）

[1718]　8-胺基酸球菌屬某種BV3L6（AsCpf1）

[1719]

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKA
RNDHYKELKPIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKE
KTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKG
LFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYEN
RKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHF
ENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISRE
AGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRN
TLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSID
LTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAK
EKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALD
QPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEF

SARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTL
ASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPT
EKTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPIL
LSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREA
LCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPL
LYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHT
LYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEK
MLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPN
VITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVN
AYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDY
QKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMI
HYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVL
KDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSK
IDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHF
KMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIV
PVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLEND
DSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDS
RFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGIS
NQDWLAYIQELRN*KRPAATKKAGQAKKKK*GS**YPYDVPDYAYPY**
DVPDYAYPYDVPDYA（SEQ ID NO: 236）

[1720]　9-毛螺菌科細菌MA2020（Lb2Cpf1）

[1721]

MYYESLTKQYPVSKTIRNELIPIGKTLDNIRQNNILESDVKR

KQNYEHVKGILDEYHKQLINEALDNCTLPSLKIAAEIYLKNQKE
VSDREDFNKTQDLLRKEVVEKLKAHENFTKIGKKDILDLLEKLP
SISEDDYNALESFRNFYTYFTSYNKVRENLYSDKEKSSTVAYRLI
NENFPKFLDNVKSYRFVKTAGILADGLGEEEQDSLFIVETFNKTL
TQDGIDTYNSQVGKINSSINLYNQKNQKANGFRKIPKMKMLYKQ
ILSDREESFIDEFQSDEVLIDNVESYGSVLIESLKSSKVSAFFDAL
RESKGKNVYVKNDLAKTAMSNIVFENWRTFDDLLNQEYDLANE
NKKKDDKYFEKRQKELKKNKSYSLEHLCNLSEDSCNLIENYIHQ
ISDDIENIIINNETFLRIVINEHDRSRKLAKNRKAVKAIKDFLDSIK
VLERELKLINSSGQELEKDLIVYSAHEELLVELKQVDSLYNMTR
NYLTKKPFSTEKVKLNFNRSTLLNGWDRNKETDNLGVLLLKDG
KYYLGIMNTSANKAFVNPPVAKTEKVFKKVDYKLLPVPNQMLP
KVFFAKSNIDFYNPSSEIYSNYKKGTHKKGNMFSLEDCHNLIDFF
KESISKHEDWSKFGFKFSDTASYNDISEFYREVEKQGYKLTYTDI
DETYINDLIERNELYLFQIYNKDFSMYSKGKLNLHTLYFMMLFD
QRNIDDVVYKLNGEAEVFYRPASISEDELIIHKAGEEIKNKNPNR
ARTKETSTFSYDIVKDKRYSKDKFTLHIPITMNFGVDEVKRFND
AVNSAIRIDENVNVIGIDRGERNLLYVVVIDSKGNILEQISLNSIIN
KEYDIETDYHALLDEREGGRDKARKDWNTVENIRDLKAGYLSQ
VVNVVAKLVLKYNAIICLEDLNFGFKRGRQKVEKQVYQKFEKM
LIDKLNYLVIDKSREQTSPKELGGALNALQLTSKFKSFKELGKQS
GVIYYVPAYLTSKIDPTTGFANLFYMKCENVEKSKRFFDGFDFIR
FNALENVFEFGFDYRSFTQRACGINSKWTVCTNGERIIKYRNPD
KNNMFDEKVVVVTDEMKNLFEQYKIPYEDGRNVKDMIISNEEA

EFYRRLYRLLQQTLQMRNSTSDGTRDYIISPVKNKREAYFNSELS
DGSVPKDADANGAYNIARKGLWVLEQIRQKSEGEKINLAMTNA
EWLEYAQTHLL*KRPAATKKAGQAKKKK*GS**YPYDVPDYAYPYDVP**
DYAYPYDVPDYA（SEQ ID NO: 237）

[1722]　10-候選白蟻甲烷枝原體（CMtCpf1）

[1723]

MNNYDEFTKLYPIQKTIRFELKPQGRTMEHLETFNFFEEDRD
RAEKYKILKEAIDEYHKKFIDEHLTNMSLDWNSLKQISEKYYKS
REEKDKKVFLSEQKRMRQEIVSEFKKDDRFKDLFSKKLFSELLK
EEIYKKGNHQEIDALKSFDKFSGYFIGLHENRKNMYSDGDEITAI
SNRIVNENFPKFLDNLQKYQEARKKYPEWIIKAESALVAHNIKM
DEVFSLEYFNKVLNQEGIQRYNLALGGYVTKSGEKMMGLNDAL
NLAHQSEKSSKGRIHMTPLFKQILSEKESFSYIPDVFTEDSQLLPS
IGGFFAQIENDKDGNIFDRALELISSYAEYDTERIYIRQADINRVS
NVIFGEWGTLGGLMREYKADSINDINLERTCKKVDKWLDSKEF
ALSDVLEAIKRTGNNDAFNEYISKMRTAREKIDAARKEMKFISE
KISGDEESIHIIKTLLDSVQQFLHFFNLFKARQDIPLDGAFYAEFD
EVHSKLFAIVPLYNKVRNYLTKNNLNTKKIKLNFKNPTLANGWD
QNKVYDYASLIFLRDGNYYLGIINPKRKKNIKFEQGSGNGPFYR
KMVYKQIPGPNKNLPRVFLTSTKGKKEYKPSKEIIEGYEADKHIR
GDKFDLDFCHKLIDFFKESIEKHKDWSKFNFYFSPTESYGDISEF
YLDVEKQGYRMHFENISAETIDEYVEKGDLFLFQIYNKDFVKAA
TGKKDMHTIYWNAAFSPENLQDVVVKLNGEAELFYRDKSDIKE

IVHREGEILVNRTYNGRTPVPDKIHKKLTDYHNGRTKDLGEAKE
YLDKVRYFKAHYDITKDRRYLNDKIYFHVPLTLNFKANGKKNL
NKMVIEKFLSDEKAHIIGIDRGERNLLYYSIIDRSGKIIDQQSLNVI
DGFDYREKLNQREIEMKDARQSWNAIGKIKDLKEGYLSKAVHEI
TKMAIQYNAIVVMEELNYGFKRGRFKVEKQIYQKFENMLIDKM
NYLVFKDAPDESPGGVLNAYQLTNPLESFAKLGKQTGILFYVPA
AYTSKIDPTTGFVNLFNTSSKTNAQERKEFLQKFESISYSAKDGG
IFAFAFDYRKFGTSKTDHKNVWTAYTNGERMRYIKEKKRNELFD
PSKEIKEALTSSGIKYDGGQNILPDILRSNNNGLIYTMYSSFIAAI
QMRVYDGKEDYIISPIKNSKGEFFRTDPKRRELPIDADANGAYNI
ALRGELTMRAIAEKFDPDSEKMAKLELKHKDWFEFMQTRGD*KR*
*PAATKKAGQAKKKK*GS**YPYDVPDYAYPYDVPDYAYPYDVPDYA**
（SEQ ID NO: 238）

[1724]　11-挑剔真細菌（EeCpf1）

[1725]

MNGNRSIVYREFVGVIPVAKTLRNELRPVGHTQEHIIQNGLI
QEDELRQEKSTELKNIMDDYYREYIDKSLSGVTDLDFTLLFELM
NLVQSSPSKDNKKALEKEQSKMREQICTHLQSDSNYKNIFNAKL
LKEILPDFIKNYNQYDVKDKAGKLETLALFNGFSTYFTDFFEKR
KNVFTKEAVSTSIAYRIVHENSLIFLANMTSYKKISEKALDEIEVI
EKNNQDKMGDWELNQIFNPDFYNMVLIQSGIDFYNEICGVVNA
HMNLYCQQTKNNYNLFKMRKLHKQILAYTSTSFEVPKMFEDDM
SVYNAVNAFIDETEKGNIIGKLKDIVNKYDELDEKRIYISKDFYE

TLSCFMSGNWNLITGCVENFYDENIHAKGKSKEEKVKKAVKED
KYKSINDVNDLVEKYIDEKERNEFKNSNAKQYIREISNIITDTETA
HLEYDDHISLIESEEKADEMKKRLDMYMNMYHWAKAFIVDEVL
DRDEMFYSDIDDIYNILENIVPLYNRVRNYVTQKPYNSKKIKLNF
QSPTLANGWSQSKEFDNNAIILIRDNKYYLAIFNAKNKPDKKIIQ
GNSDKKNDNDYKKMVYNLLPGANKMLPKVFLSKKGIETFKPSD
YIISGYNAHKHIKTSENFDISFCRDLIDYFKNSIEKHAEWRKYEF
KFSATDSYSDISEFYREVEMQGYRIDWTYISEADINKLDEEGKIY
LFQIYNKDFAENSTGKENLHTMYFKNIFSEENLKDIIIKLNGQAE
LFYRRASVKNPVKHKKDSVLVNKTYKNQLDNGDVVRIPIPDDIY
NEIYKMYNGYIKESDLSEAAKEYLDKVEVRTAQKDIVKDYRYT
VDKYFIHTPITINYKVTARNNVNDMVVKYIAQNDDIHVIGIDRGE
RNLIYISVIDSHGNIVKQKSYNILNNYDYKKKLVEKEKTREYARK
NWKSIGNIKELKEGYISGVVHEIAMLIVEYNAIIAMEDLNYGFKR
GRFKVERQVYQKFESMLINKLNYFASKEKSVDEPGGLLKGYQLT
YVPDNIKNLGKQCGVIFYVPAAFTSKIDPSTGFISAFNFKSISTNA
SRKQFFMQFDEIRYCAEKDMFSFGFDYNNFDTYNITMGKTQWT
VYTNGERLQSEFNNARRTGKTKSINLTETIKLLLEDNEINYADGH
DIRIDMEKMDEDKKSEFFAQLLSLYKLTVQMRNSYTEAEEQENG
ISYDKIISPVINDEGEFFDSDNYKESDDKECKMPKDADANGAYCI
ALKGLYEVLKIKSEWTEDGFDRNCLKLPHAEWLDFIQNKRYE*KR*
*PAATKKAGQAKKKK*G**SYPYDVPDYAYPYDVPDYAYPYDVPDYA**

（SEQ ID NO: 239）

[1726] 12-牛莫拉氏菌237（MbCpf1）

[1727]

MLFQDFTHLYPLSKTVRFELKPIDRTLEHIHAKNFLSQDETM
ADMHQKVKVILDDYHRDFIADMMGEVKLTKLAEFYDVYLKFR
KNPKDDELQKQLKDLQAVLRKEIVKPIGNGGKYKAGYDRLFGA
KLFKDGKELGDLAKFVIAQEGESSPKLAHLAHFEKFSTYFTGFH
DNRKNMYSDEDKHTAIAYRLIHENLPRFIDNLQILTTIKQKHSAL
YDQIINELTASGLDVSLASHLDGYHKLLTQEGITAYNTLLGGISGE
AGSPKIQGINELINSHHNQHCHKSERIAKLRPLHKQILSDGMSVS
FLPSKFADDSEMCQAVNEFYRHYADVFAKVQSLFDGFDDHQKD
GIYVEHKNLNELSKQAFGDFALLGRVLDGYYVDVVNPEFNERFA
KAKTDNAKAKLTKEKDKFIKGVHSLASLEQAIEHYTARHDDESV
QAGKLGQYFKHGLAGVDNPIQKIHNNHSTIKGFLERERPAGERA
LPKIKSGKNPEMTQLRQLKELLDNALNVAHFAKLLTTKTTLDNQ
DGNFYGEFGVLYDELAKIPTLYNKVRDYLSQKPFSTEKYKLNFG
NPTLLNGWDLNKEKDNFGVILQKDGCYYLALLDKAHKKVFDN
APNTGKSIYQKMIYKYLEVRKQFPKVFFSKEAIAINYHPSKELVE
IKDKGRQRSDDERLKLYRFILECLKIHPKYDKKFEGAIGDIQLFK
KDKKGREVPISEKDLFDKINGIFSSKPKLEMEDFFIGEFKRYNPSQ
DLVDQYNIYKKIDSNDNRKKENFYNNHPKFKKDLVRYYYESMC
KHEEWEESFEFSKKLQDIGCYVDVNELFTEIETRRLNYKISFCNI
NADYIDELVEQGQLYLFQIYNKDFSPKAHGKPNLHTLYFKALFS
EDNLADPIYKLNGEAQIFYRKASLDMNETTIHRAGEVLENKNPD
NPKKRQFVYDIIKDKRYTQDKFMLHVPITMNFGVQGMTIKEFNK
KVNQSIQQYDEVNVIGIDRGERHLLYLTVINSKGEILEQCSLNDIT

TASANGTQMTTPYHKILDKREIERLNARVGWGEIETIKELKSGYL
SHVVHQISQLMLKYNAIVVLEDLNFGFKRGRFKVEKQIYQNFEN
ALIKKLNHLVLKDKADDEIGSYKNALQLTNNFTDLKSIGKQTGF
LFYVPAWNTSKIDPETGFVDLLKPRYENIAQSQAFFGKFDKICYN
ADKDYFEFHIDYAKFTDKAKNSRQIWTICSHGDKRYVYDKTAN
QNKGAAKGINVNDELKSLFARHHINEKQPNLVMDICQNNDKEF
HKSLMYLLKTLLALRYSNASSDEDFILSPVANDEGVFFNSALAD
DTQPQNADANGAYHIALKGLWLLNELKNSDDLNKVKLAIDNQT
WLNFAQNR*KRPAATKKAGQAKKKK*GS**YPYDVPDYAYPYDVPDY**
AYPYDVPDYA（**SEQ ID NO: 240**）

[1728]　13-稻田鉤端螺旋體（LiCpf1）

[1729]

MEDYSGFVNIYSIQKTLRFELKPVGKTLEHIEKKGFLKKDKI
RAEDYKAVKKIIDKYHRAYIEEVFDSVLHQKKKKDKTRFSTQFI
KEIKEFSELYYKTEKNIPDKERLEALSEKLRKMLVGAFKGEFSEE
VAEKYKNLFSKELIRNEIEKFCETDEERKQVSNFKSFTTYFTGFH
SNRQNIYSDEKKSTAIGYRIIHQNLPKFLDNLKIIESIQRRFKDFP
WSDLKKNLKKIDKNIKLTEYFSIDGFVNVLNQKGIDAYNTILGG
KSEESGEKIQGLNEYINLYRQKNNIDRKNLPNVKILFKQILGDRE
TKSFIPEAFPDDQSVLNSITEFAKYLKLDKKKKSIIAELKKFLSSF
NRYELDGIYLANDNSLASISTFLFDDWSFIKKSVSFKYDESVGDP
KKKIKSPLKYEKEKEKWLKQKYYTISFLNDAIESYSKSQDEKRV
KIRLEAYFAEFKSKDDAKKQFDLLERIEEAYAIVEPLLGAEYPRD

RNLKADKKEVGKIKDFLDSIKSLQFFLKPLLSAEIFDEKDLGFYN
QLEGYYEEIDSIGHLYNKVRNYLTGKIYSKEKFKLNFENSTLLKG
WDENREVANLCVIFREDQKYYLGVMDKENNTILSDIPKVKPNEL
FYEKMVYKLIPTPHMQLPRIIFSSDNLSIYNPSKSILKIREAKSFKE
GKNFKLKDCHKFIDFYKESISKNEDWSRFDFKFSKTSSYENISEF
YREVERQGYNLDFKKVSKFYIDSLVEDGKLYLFQIYNKDFSIFSK
GKPNLHTIYFRSLFSKENLKDVCLKLNGEAEMFFRKKSINYDEK
KKREGHHPELFEKLKYPILKDKRYSEDKFQFHLPISLNFKSKERL
NFNLKVNEFLKRNKDINIIGIDRGERNLLYLVMINQKGEILKQTL
LDSMQSGKGRPEINYKEKLQEKEIERDKARKSWGTVENIKELKE
GYLSIVIHQISKLMVENNAIVVLEDLNIGFKRGRQKVERQVYQK
FEKMLIDKLNFLVFKENKPTEPGGVLKAYQLTDEFQSFEKLSKQT
GFLFYVPSWNTSKIDPRTGFIDFLHPAYENIEKAKQWINKFDSIRF
NSKMDWFEFTADTRKFSENLMLGKNRVWVICTTNVERYFTSKT
ANSSIQYNSIQITEKLKELFVDIPFSNGQDLKPEILRKNDAVFFKS
LLFYIKTTLSLRQNNGKKGEEEKDFILSPVVDSKGRFFNSLEASD
DEPKDADANGAYHIALKGLMNLLVLNETKEENLSRPKWKIKNK
DWLEFVWERNR*KRPAATKKAGQAKKKK*G**SYPYDVPDYAYPYDV**
PDYAYPYDVPDYA （SEQ ID NO: 241）

[1730] 14-毛螺菌科細菌ND2006（LbCpf1）

[1731]

MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDE
KRAEDYKGVKKLLDRYYLSFINDVLHSIKLKNLNNYISLFRKKT

RTEKENKELENLEINLRKEIAKAFKGNEGYKSLFKKDIIETILPEF
LDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRCI
NENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFE
GEFFNFVLTQEGIDVYNAIIGGFVTESGEKIKGLNEYINLYNQKT
KQKLPKFKPLYKQVLSDRESLSFYGEGYTSDEEVLEVFRNTLNK
NSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNV
IRDKWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQ
EYADADLSVVEKLKEIIIQKVDEIYKVYGSSEKLFDADFVLEKSL
KKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFYGDF
VLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGG
WDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNG
NYEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNG
TFKKGDMFNLNDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKY
KDIAGFYREVEEQGYKVSFESASKKEVDKLVEEGKLYMFQIYNK
DFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASL
KKEELVVHPANSPIANKNPDNPKKTTTLSYDVYKDKRFSEDQYE
LHIPIAINKCPKNIFKINTEVRVLLKHDDNPYVIGIDRGERNLLYI
VVVDGKGNIVEQYSLNEIINNFNGIRIKTDYHSLLDKKEKERFEA
RQNWTSIENIKELKAGYISQVVHKICELVEKYDAVIALEDLNSGF
KNSRVKVEKQVYQKFEKMLIDKLNYMVDKKSNPCATGGALKG
YQITNKFESFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLKTKYT
SIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWK
LYSYGNRIRIFRNPKKNNVFDWEEVCLTSAYKELFNKYGINYQQ
GDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDFLIS

PVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIG
QFKKAEDEKLDKVKIAISNKEWLEYAQTSVKH*KRPAATKKAGQA*
*KKKK*GS**YPYDVPDYAYPYDVPDYAYPYDVPDYA（SEQ ID NO: 242）**

[1732] 15-狗口腔卟啉單胞菌（PcCpf1）

[1733]

MDSLKDFTNLYPVSKTLRFELKPVGKTLENIEKAGILKEDEH
RAESYRRVKKIIDTYHKVFIDSSLENMAKMGIENEIKAMLQSFCE
LYKKDHRTEGEDKALDKIRAVLRGLIVGAFTGVCGRRENTVQNE
KYESLFKEKLIKEILPDFVLSTEAESLPFSVEEATRSLKEFDSFTSY
FAGFYENRKNIYSTKPQSTAIAYRLIHENLPKFIDNILVFQKIKEPI
AKELEHIRADFSAGGYIKKDERLEDIFSLNYYIHVLSQAGIEKYN
ALIGKIVTEGDGEMKGLNEHINLYNQQRGREDRLPLFRPLYKQIL
SDREQLSYLPESFEKDEELLRALKEFYDHIAEDILGRTQQLMTSIS
EYDLSRIYVRNDSQLTDISKKMLGDWNAIYMARERAYDHEQAP
KRITAKYERDRIKALKGEESISLANLNSCIAFLDNVRDCRVDTYL
STLGQKEGPHGLSNLVENVFASYHEAEQLLSFPYPEENNLIQDK
DNVVLIKNLLDNISDLQRFLKPLWGMGDEPDKDERFYGEYNYIR
GALDQVIPLYNKVRNYLTRKPYSTRKVKLNFGNSQLLSGWDRN
KEKDNSCVILRKGQNFYLAIMNNRHKRSFENKMLPEYKEGEPY
FEKMDYKFLPDPNKMLPKVFLSKKGIEIYKPSPKLLEQYGHGTH
KKGDTFSMDDLHELIDFFKHSIEAHEDWKQFGFKFSDTATYENV
SSFYREVEDQGYKLSFRKVSESYVYSLIDQGKLYLFQIYNKDFSP

CSKGTPNLHTLYWRMLFDERNLADVIYKLDGKAEIFFREKSLKN
DHPTHPAGKPIKKKSRQKKGEESLFEYDLVKDRRYTMDKFQFHV
PITMNFKCSAGSKVNDMVNAHIREAKDMHVIGIDRGERNLLYIC
VIDSRGTILDQISLNTINDIDYHDLLESRDKDRQQEHRNWQTIEGI
KELKQGYLSQAVHRIAELMVAYKAVVALEDLNMGFKRGRQKVE
SSVYQQFEKQLIDKLNYLVDKKKRPEDIGGLLRAYQFTAPFKSFK
EMGKQNGFLFYIPAWNTSNIDPTTGFVNLFHVQYENVDKAKSFF
QKFDSISYNPKKDWFEFAFDYKNFTKKAEGSRSMWILCTHGSRI
KNFRNSQKNGQWDSEEFALTEAFKSLFVRYEIDYTADLKTAIVD
EKQKDFFVDLLKLFKLTVQMRNSWKEKDLDYLISPVAGADGRF
FDTREGNKSLPKDADANGAYNIALKGLWALRQIRQTSEGGKLKL
AISNKEWLQFVQERSYEKD*KRPAATKKAGQAKKKK*GS**YPYDVPD**
YAYPYDVPDYAYPYDVPDYA （SEQ ID NO: 243）

[1734] 16-解糖胨普雷沃菌（PdCpf1）

[1735]

MENYQEFTNLFQLNKTLRFELKPIGKTCELLEEGKIFASGSF
LEKDKVRADNVSYVKKEIDKKHKIFIEETLSSFSISNDLLKQYFD
CYNELKAFKKDCKSDEEEVKKTALRNKCTSIQRAMREAISQAFL
KSPQKKLLAIKNLIENVFKADENVQHFSEFTSYFSGFETNRENFY
SDEEKSTSIAYRLVHDNLPIFIKNIYIFEKLKEQFDAKTLSEIFENY
KLYVAGSSLDEVFSLEYFNNTLTQKGIDNYNAVIGKIVKEDKQEI
QGLNEHINLYNQKHKDRRLPFFISLKKQILSDREALSWLPDMFK
NDSEVIKALKGFYIEDGFENNVLTPLATLLSSLDKYNLNGIFIRN

NEALSSLSQNVYRNFSIDEAIDANAELQTFNNYELIANALRAKIK
KETKQGRKSFEKYEEYIDKKVKAIDSLSIQEINELVENYVSEFNS
NSGNMPRKVEDYFSLMRKGDFGSNDLIENIKTKLSAAEKLLGTK
YQETAKDIFKKDENSKLIKELLDATKQFQHFIKPLLGTGEEADRD
LVFYGDFLPLYEKFEELTLLYNKVRNRLTQKPYSKDKIRLCFNKP
KLMTGWVDSKTEKSDNGTQYGGYLFRKKNEIGEYDYFLGISSK
AQLFRKNEAVIGDYERLDYYQPKANTIYGSAYEGENSYKEDKK
RLNKVIIAYIEQIKQTNIKKSIIESISKYPNISDDDKVTPSSLLEKIK
KVSIDSYNGILSFKSFQSVNKEVIDNLLKTISPLKNKAEFLDLINK
DYQIFTEVQAVIDEICKQKTFIYFPISNVELEKEMGDKDKPLCLF
QISNKDLSFAKTFSANLRKKRGAENLHTMLFKALMEGNQDNLD
LGSGAIFYRAKSLDGNKPTHPANEAIKCRNVANKDKVSLFTYDI
YKNRRYMENKFLFHLSIVQNYKAANDSAQLNSSATEYIRKADD
LHIIGIDRGERNLLYYSVIDMKGNIVEQDSLNIIRNNDLETDYHD
LLDKREKERKANRQNWEAVEGIKDLKKGYLSQAVHQIAQLMLK
YNAIIALEDLGQMFVTRGQKIEKAVYQQFEKSLVDKLSYLVDKK
RPYNELGGILKAYQLASSITKNNSDKQNGFLFYVPAWNTSKIDPV
TGFTDLLRPKAMTIKEAQDFFGAFDNISYNDKGYFEFETNYDKF
KIRMKSAQTRWTICTFGNRIKRKKDKNYWNYEEVELTEEFKKLF
KDSNIDYENCNLKEEIQNKDNRKFFDDLIKLLQLTLQMRNSDDK
GNDYIISPVANAEGQFFDSRNGDKKLPLDADANGAYNIARKGLW
NIRQIKQTKNDKKLNLSISSTEWLDFVREKPYLK*KRPAATKKAGQ*
*AKKKK*GS**YPYDVPDYAYPYDVPDYAYPYDVPDYA（SEQ ID NO:**
244）

[1736] 17-獼猴卟啉單胞菌（PmCpf1）

[1737]

MKTQHFFEDFTSLYSLSKTIRFELKPIGKTLENIKKNGLIRRD
EQRLDDYEKLKKVIDEYHEDFIANILSSFSFSEEILQSYIQNLSESE
ARAKIEKTMRDTLAKAFSEDERYKSIFKKELVKKDIPVWCPAYK
SLCKKFDNFTTSLVPFHENRKNLYTSNEITASIPYRIVHVNLPKFI
QNIEALCELQKKMGADLYLEMMENLRNVWPSFVKTPDDLCNLK
TYNHLMVQSSISEYNRFVGGYSTEDGTKHQGINEWINIYRQRNK
EMRLPGLVFLHKQILAKVDSSSFISDTLENDDQVFCVLRQFRKLF
WNTVSSKEDDAASLKDLFCGLSGYDPEAIYVSDAHLATISKNIF
DRWNYISDAIRRKTEVLMPRKKESVERYAEKISKQIKKRQSYSL
AELDDLLAHYSEESLPAGFSLLSYFTSLGGQKYLVSDGEVILYEE
GSNIWDEVLIAFRDLQVILDKDFTEKKLGKDEEAVSVIKKALDS
ALRLRKFFDLLSGTGAEIRRDSSFYALYTDRMDKLKGLLKMYDK
VRNYLTKKPYSIEKFKLHFDNPSLLSGWDKNKELNNLSVIFRQN
GYYYLGIMTPKGKNLFKTLPKLGAEEMFYEKMEYKQIAEPMLM
LPKVFFPKKTKPAFAPDQSVVDIYNKKTFKTGQKGFNKKDLYRL
IDFYKEALTVHEWKLFNFSFSPTEQYRNIGEFFDEVREQAYKVS
MVNVPASYIDEAVENGKLYLFQIYNKDFSPYSKGIPNLHTLYWK
ALFSEQNQSRVYKLCGGGELFYRKASLHMQDTTVHPKGISIHKK
NLNKKGETSLFNYDLVKDKRFTEDKFFFHVPISINYKNKKITNVN
QMVRDYIAQNDDLQIIGIDRGERNLLYISRIDTRGNLLEQFSLNVI
ESDKGDLRTDYQKILGDREQERLRRRQEWKSIESIKDLKDGYMS
QVVHKICNMVVEHKAIVVLENLNLSFMKGRKKVEKSVYEKFER

MLVDKLNYLVVDKKNLSNEPGGLYAAYQLTNPLFSFEELHRYPQ SGILFFVDPWNTSLTDPSTGFVNLLGRINYTNVGDARKFFDRFNA IRYDGKGNILFDLDLSRFDVRVETQRKLWTLTTFGSRIAKSKKSG KWMVERIENLSLCFLELFEQFNIGYRVEKDLKKAILSQDRKEFY VRLIYLFNLMMQIRNSDGEEDYILSPALNEKNLQFDSRLIEAKDL PVDADANGAYNVARKGLMVVQRIKRGDHESIHRIGRAQWLRYV QEGIVE*KRPAATKKAGQAKKKK*GS**YPYDVPDYAYPYDVPDYAYP YDVPDYA （SEQ ID NO: 245）**

實例 15：Cpf1 結構的計算分析

[1738] Cpf1核酸酶初級結構的計算分析揭示了三個不同的區（圖109）。第一係C末端RuvC樣結構域，其係僅功能表征的結構域。第二係N末端α-螺旋區並且第三係位於RuvC樣結構域與α-螺旋區之間的混合的α區和β區。

[1739] 預測非結構化區的若干小片段在Cpf1初始結構之內。對於小蛋白質序列的拆分和插入而言，不同的Cpf1異種同源物內的暴露於溶劑且不保守的非結構化區係較佳的側面。另外，該等側面可以用於在Cpf1異種同源物之間產生嵌合蛋白。

實例 16：生成具有增強的特異性的 Cpf1 突變體

[1740] 最近描述了用於生成具有增強的特異性的Cas9異種同源物的方法（斯萊馬克爾等人，2015）。此策略可以用於增強Cpf1異種同源物的特異性。

[1741] 用於誘變的主要殘基係RuvC結構域內所有帶正電荷

殘基，因為這係在不存在晶體時僅已知的結構並且我們知道RuvC中的特異性突變體在Cas9中起作用（參見以下表：RuvC內的保守性賴胺酸和精胺酸殘基）。

[1742] 在不希望受到理論約束的情況下，Cpf1的此區域的帶正電荷殘基可以用於藉由與DNA非靶股的帶負電荷磷酸二酯骨架相互作用來穩定酶與DNA之間的相互作用。藉由取代Cpf1的帶正電荷殘基，可以破壞與非靶股的相互作用。此相互作用的足夠破壞可以維持針對靶位點的適當活性，但是減小了針對非靶位點的酶活性（考慮到與靶序列相比的一個或多個錯配，這通常將預期與指導序列具有較弱的相互作用）。

[1743] 其他結構域展示出類似的特徵。感興趣的區域係REC1結構域，包括但不限於與SpCas9的N497、R661、Q695、以及Q926類似的一個或多個胺基酸殘基的突變，並且包括但不限於在那些位置處對丙胺酸的突變。在此類殘基處的突變還破壞了酶-DNA磷酸酯骨架相互作用。此外，可以採用位於相同或不同結構域中的突變的組合。

表：RuvC 內的保守性賴胺酸和精胺酸殘基。

AsCpf1	**LbCpf1**
R912	R833
T923	R836
R947	K847
K949	K879
R951	K881
R955	R883
K965	R887
K968	K897
K1000	K900
R1003	K932
K1009	R935

K1017	K940
K1022	K948
K1029	K953
K1072	K960
K1086	K984
F1103	K1003
R1226	K1017
R1252	R1033
	R1138
	R1165

另外的候選物係以下表中提供的在不同異種同源物之間保守的帶正電荷殘基。

表：保守性賴胺酸和精胺酸殘基。

殘基	**AsCpf1**	**FnCpf1**	**LbCpf1**	**MbCpf1**
Lys	K15	K15	K15	K14
Arg	R18	R18	R18	R17
Lys/Arg	K26	K26	K26	R25
Lys/Arg	Q34	R34	K34	K33
Arg	R43	R43	R43	M42
Lys	K48	K48	K48	Q47
Lys	K51	K51	K51	K50
Lys/Arg	R56	K56	R56	D55
Lys/Arg	R84	K87	K83	K85
Lys/Arg	K85	K88	K84	N86
Lys/Arg	K87	D90	R86	K88
Arg	N93	K96	K92	K94
Lys/Arg	R103	K106	R102	R104
Lys	N104	K107	K103	K105
Lys	T118	K120	K116	K118
Lys/Arg	K123	Q125	K121	K123
Lys	K134	K143	-	K131
Arg	R176	R186	R158	R174
Lys	K177	K187	E159	K175
Arg	R192	R202	R174	R190
Lys/Arg	K200	K210	R182	R198
Lys	K226	K235	K206	I221
Lys	K273	K296	K251	K267
Lys	K275	K298	K253	Q269
Lys	T291	K314	K269	K285
Lys/Arg	R301	K320	K271	K291
Lys	K307	K326	K278	K297
Lys	K369	K397	P342	K357
Lys	S404	K444	K380	K403

Lys/Arg	V409	K449	R385	K409
Lys	K414	E454	K390	K414
Lys	K436	A483	K415	K448
Lys	K438	E491	K421	K460
Lys	K468	K527	K457	K501
Lys	D482	K541	K471	K515
Lys	K516	K581	A506	K550
Arg	R518	R583	R508	R552
Lys	K524	K589	K514	K558
Lys	K530	K595	K520	K564
Lys	K532	K597	K522	K566
Lys	K548	K613	K538	K582
Lys	K559	K624	Y548	K593
Lys	K570	K635	K560	K604
Lys/Arg	R574	K639	K564	K608
Lys	K592	K656	K580	K623
Lys	D596	K660	K584	K627
Lys	K603	K667	K591	K633
Lys	K607	K671	K595	K637
Lys	K613	K677	K601	E643
Lys	C647	K719	K634	K780
Lys/Arg	R681	K725	K640	Y787
Lys/Arg	K686	K730	R645	K792
Lys	H720	K763	K679	K830
Lys	K739	K782	K689	Q846
Lys	K748	K791	K707	K858
Lys/Arg	K757	R800	T716	K867
Lys/Arg	T766	K809	K725	K876
Lys/Arg	K780	K823	R737	K890
Arg	R790	R833	R747	R900
Lys/Arg	P791	K834	R748	K901
Lys	K796	K839	K753	M906
Lys	K809	K852	K768	K921
Lys	K815	K858	K774	K927
Lys	T816	K859	K775	K928
Lys	K860	K869	K785	K937
Lys/Arg	R862	K871	K787	K939
Arg	R863	R872	R788	R940
Lys	K868	K877	Q793	K945
Lys	K897	K905	K821	Q975
Arg	R909	R918	R833	R987
Arg	R912	R921	R836	R990
Lys	T923	K932	K847	K1001
Lys/Arg	R947	I960	K879	R1034
Lys	K949	K962	K881	I1036
Arg	R951	R964	R883	R1038

Arg	R955	R968	R887	R1042
Lys	K965	K978	K897	K1052
Lys	K968	K981	K900	K1055
Lys	K1000	K1013	K932	K1087
Arg	R1003	R1016	R935	R1090
Lys	K1009	K1021	K940	K1095
Lys	K1017	K1029	K948	N1103
Lys	K1022	K1034	K953	K1108
Lys	K1029	K1041	K960	K1115
Lys	A1053	K1065	K984	K1139
Lys	K1072	K1084	K1003	K1158
Lys/Arg	K1086	K1098	K1017	R1172
Lys/Arg	F1103	K1114	R1033	K1188
Lys	S1209	K1201	K1121	K1276
Arg	R1226	R1218	R1138	R1293
Arg	R1252	R1244	R1165	A1319
Lys	K1273	K1265	K1190	K1340
Lys	K1282	K1274	K1199	K1349
Lys	K1288	K1281	K1208	K1356

[1744] 以上表提供了在來自新殺手弗朗西絲菌*U112*（FnCpf1）、-胺基酸球菌屬某種*BV3L6*（AsCpf1）、毛螺菌科細菌*ND2006*（LbCpf1）以及牛莫拉氏菌*237*（MbCpf1）的Cpf1核酸酶比對中保守性賴胺酸和精胺酸殘基的位置。該等可以用於生成具有增強的特異性的Cpf1突變體。

實例 17：提供 Cpf1 結合的特異性

[1745] 在用於提高Cas9特異性的類似策略中，Cpf1的特異性可以藉由使穩定化非靶向DNA股的殘基發生突變來改進。這可以在無晶體結構的情況下藉由使用線性結構比對以預測 1) 哪個Cpf1結構域結合DNA的哪條股，以及 2) 該等結構域內的哪些殘基接觸DNA來完成。

[1746] 然而，此方法可能由於Cpf1與已知蛋白質的不良保守性而受到限制。因此，希望的是以探針探測所有可能的DNA相互

作用胺基酸（賴胺酸、組胺酸以及精胺酸）的功能。

[1747] RuvC結構域內帶正電荷殘基在整個Cpf1中比在Rad50結構域中更保守，這指示RuvC殘基係具有較小進化彈性的。這表明在此結構域中需要核酸結合的嚴格控制（相對於Rad50結構域）。因此，可能此結構域由於需要RNA：DNA雙股體穩定而切割靶向的DNA股（Cas9中的前例）。此外，更多精胺酸存在於RuvC結構域中（5%的RuvC殘基904至1307對比提議的Rad50結構域中的3.8%），這表明RuvC靶向一條DNA股。精胺酸涉及結合核酸大溝和小溝（羅氏自然（Rohs Nature） 2009：http://rohslab.cmb.usc.edu/Papers/Rohs_etal_Nature.pdf）。大溝/小溝可以僅存在於雙股體（諸如DNA：RNA靶向雙股體）中，這進一步表明RuvC可以涉及切割。

[1748] 圖110、圖111和圖112提供了與Cpf1中發現的結構域類似的兩種結構域的晶體結構（RuvC霍利迪連結體解離酶和Rad50DNA修復蛋白）。基於該等結構，可以推斷Cpf1中哪些相關結構域看起來像並且推斷哪些區域和殘基可以接觸DNA。在每種結構中，突出了接觸DNA的殘基。在圖113的比對中，注解與該等DNA結合區域相對應的AsCpf1區域。以下表中的殘基清單係兩種結合結構域中可見的殘基。

表–可能的 DNA 相互作用殘基的列表

RuvC 結構域可能的DNA相互作用殘基：	Rad50 結構域可能的 DNA 相互作用殘基：
AsCpf1	AsCpf1
R909	K324
R912	K335

R930	K337
R947	R331
K949	K369
R951	K370
R955	R386
K965	R392
K968	R393
K1000	K400
K1002	K404
R1003	K406
K1009	K408
K1017	K414
K1022	K429
K1029	K436
K1035	K438
K1054	K459
K1072	K460
K1086	K464
R1094	R670
K1095	K675
K1109	R681
K1118	K686
K1142	K689
K1150	R699
K1158	K705
K1159	R725
R1220	K729
R1226	K739
R1242	K748
R1252	K752
	R670

[1749] 由關於AsCpf1的該等特定觀察，我們可以藉由序列比對鑒定來自其他種類的Cpf1中的類似殘基。比對圖114中給出的AsCpf1和FnCpf1的實例，鑒定了Rad50結合結構域和其中的精胺酸和賴胺酸。

實例 18：使用串聯指導序列使用 Cpf1 進行多重作用

[1750] 考慮使用Cpf1酶進行多重作用是否可能。出於此目的，開發指導RNA，由此在相同啟動子下串聯定位不同指導序列，並且確定該等指導序列指導基因組編輯到其對應靶標的能力。

[1751] 在轉染前24 h，每24孔接種150,000個HEK293T細胞。使用Lipofectamin2000，用400 ng huAsCpf1質粒和100 ng包含針對GRIN28的一個指導序列和以串聯方式置於U6啟動子（圖115A）後面的針對EMX1的指導序列的串聯指導質粒轉染細胞。轉染後72 h收穫細胞並且使用SURVEYOR核酸酶測定來測定由串聯指導序列介導的AsCpf1活性。

[1752] 該等結果在圖115B中證實，該圖證明了GRIN28和EMX1基因中的INDEL形成。

[1753] 因此確定AsCpf1和類似的LbCpf1可以採用由相同U6啟動子表現的兩種指導序列而無活性喪失。串聯中的位置對indel形成沒有影響。這表明Cpf1可以用於使用兩種或更多種指導序列進行多重作用。

[1754] 本發明進一步藉由如下編號段落進行描述。

1. 一種工程化的、非天然存在的成簇規律間隔短迴文重複序列（CRISPR）-CRISPR 相關（Cas）（CRISPR-Cas）系統，包含

a) 一種或多種包含指導 RNA 的 V 型 CRISPR-Cas 多核苷酸序列，該指導 RNA 包含連接至同向重複序列的指導序列，其中該指導序列能夠與靶序列雜交，或者一種或多種編碼該一種或多種 V 型 CRISPR-Cas 多核苷酸序列的核苷酸序列，以及

b) Cpf1 效應蛋白或者編碼該 Cpf1 效應蛋白的一種或多種核苷酸序列；

其中該一種或多種指導序列與所述靶序列雜交，所述靶序列

係原型間隔區相鄰模體（PAM）的 3’，並且所述指導 RNA 與該 Cpf1 效應蛋白形成複合物。

2. 一種包含一種或多種載體的工程化的、非天然存在的成簇規律間隔短迴文重複序列（CRISPR）-CRISPR 相關（Cas）（CRISPR-Cas）載體系統，包含

c) 可操作地連接至一種或多種核苷酸序列的第一調節元件，該一種或多種核苷酸序列編碼包含指導 RNA 的一種或多種 V 型 CRISPR-Cas 多核苷酸序列，該指導 RNA 包含連接至同向重複序列的指導序列，其中該指導序列能夠與靶序列雜交，

d) 可操作地連接至編碼 Cpf1 效應蛋白的核苷酸序列的第二調節元件；

其中組分 (a) 和 (b) 位於該系統的相同或不同的載體上，

其中當轉錄時，該一種或多種指導序列與所述靶序列雜交，所述靶序列係原型間隔區相鄰模體（PAM）的 3’，並且所述指導 RNA 與該 Cpf1 效應蛋白形成複合物。

3. 如段落 1 或 2 所述之系統，其中該靶序列係在細胞內。

4. 如段落 3 所述之系統，其中該細胞包括真核細胞。

5. 如段落 1-4 中任一項所述之系統，其中當轉錄時，該一種或多種指導序列與該靶序列雜交並且該指導 RNA 與該 Cpf1 效應蛋白形成一種複合物，該複合物引起該靶序列遠端切割。

6. 如段落 5 所述之系統，其中所述切割生成具有 4 或 5-nt 5’突出端的交錯雙股斷裂。

7. 如段落 1-6 中任一項所述之系統，其中該 PAM 包含 5' T-富集模體。

8. 如段落 1-7 中任一項所述之系統，其中該效應蛋白係來源於圖 64 所列出的細菌種類的 Cpf1 效應蛋白。

9. 如段落 8 所述之系統，其中該 Cpf1 效應蛋白係來源於選自下組的細菌種類，該組由以下各項組成：土拉熱弗朗西絲菌 1、土拉熱弗朗西絲菌新殺手亞種、易北普雷沃菌、毛螺旋菌科細菌 MC2017 1、解朊丁酸弧菌、佩萊格裡尼菌科細菌 GW2011_GWA2_33_10、儉菌總門細菌 GW2011_GWC2_44_17、密斯氏菌屬某種 SCADC、胺基酸球菌屬某種 BV3L6、毛螺旋菌科細菌 MA2020、候選白蟻甲烷枝原體、挑剔真細菌、牛莫拉氏菌 237、稻田鉤端螺旋體、毛螺旋菌科細菌 ND2006、狗口腔卟啉單胞菌 3、解糖胨普雷沃菌和獼猴卟啉單胞菌。

10. 如段落 9 所述之系統，其中該 PAM 序列係 TTN，其中 N 係 A/C/G 或 T 並且該效應蛋白係 FnCpf1 或者其中該 PAM 序列係 TTTV，其中 V 係 A/C 或 G 並且該效應蛋白係 PaCpf1p、LbCpf1 或 AsCpf1。

11. 如段落 1-10 中任一項所述之系統，其中該 Cpf1 效應蛋白包含一個或多個核定位信號。

12. 如段落 1-11 中任一項所述之系統，其中編碼該 Cpf1 效應蛋白的該等核酸序列被密碼子優化用於在真核細胞中表現。

13. 如段落 1-12 中任一項所述之系統，其中組分 (a) 和 (b) 或該等核苷酸序列係在載體上。

14. 一種修飾感興趣的靶座位之方法，該方法包括將如段落 1-13 中任一項所述之系統遞送到所述座位或含有該座位的細胞中。

15. 一種修飾感興趣的靶座位之方法，該方法包括將包含 Cpf1 效應蛋白和一種或多種核酸組分的非天然存在或工程化的組成物遞送至所述座位，其中該 Cpf1 效應蛋白與該一種或多種核酸組分形成複合物並且在所述複合物與係原型間隔區相鄰模體（PAM）的 3'的感興趣的靶座位結合後，該效應蛋白誘導對該感興趣的靶座位的修飾，其中該複合物包含 Mg^{2+}。

16. 如段落 14 或 15 所述之方法，其中該感興趣的靶座位係在細胞內。

17. 如段落 16 所述之方法，其中該細胞係真核細胞。

18. 如段落 16 所述之方法，其中該細胞係動物細胞或人類細胞。

19. 如段落 16 所述之方法，其中該細胞係植物細胞。

20. 如段落 14 或 15 所述之方法，其中該感興趣的靶座位被包含體外 DNA 分子中。

21. 如段落 15-20 中任一項所述之方法，其中包含 Cpf1 效應蛋白和一種或多種核酸組分的所述非天然存在或工程化的組成物作為一個或多個多核苷酸分子遞送到該細胞中。

22. 如段落 14-21 中任一項所述之方法，其中該感興趣的靶座位包含 DNA。

23. 如段落 22 所述之方法，其中該 DNA 係鬆弛的或超螺旋的。

24. 如段落 14-23 中任一項所述之方法，其中該組成物包含單一核酸組分。

25. 如段落 24 所述之方法，其中該單一核酸組分包含連接至同向重複序列的指導序列。

26. 如段落 14-25 中任一項所述之方法，其中該感興趣的靶座位的該修飾係股斷裂。

27. 如段落 26 所述之方法，其中該股斷裂包括具有 4 或 5-nt 5'突出端的交錯 DNA 雙股斷裂。

28. 如段落 26 或 27 所述之方法，其中該感興趣的靶座位係藉由將 DNA 插入物整合到該交錯 DNA 雙股斷裂中來修飾的。

29. 如段落 14-28 中任一項所述之方法，其中該 Cpf1 效應蛋白包含一個或多個核定位信號（一個或多個 NLS）。

30. 如段落 21-29 中任一項所述之方法，其中該一個或多個多核苷酸分子被包含在一種或多種載體中。

31. 如段落 21-30 中任一項所述之方法，其中該一個或多個多核苷酸分子包含可操作地配置為表現該 Cpf1 效應蛋白和/或該一種或多種核酸組分的一種或多種調節元件，視情況其中該一種或

多種調節元件包括誘導型啟動子。

32. 如段落 21 至 31 中任一項所述之方法，其中該一個或多個多核苷酸分子或該一種或多種載體被包含在一遞送系統中。

33. 如段落 14-30 中任一項所述之方法，其中系統或該一個或多個多核苷酸分子係經由粒子、囊泡或一種或多種病毒載體來遞送的。

34. 如段落 33 所述之方法，其中該等粒子包含脂質、糖、金屬或蛋白質。

35. 如段落 33 所述之方法，其中該等囊泡包含外來體或脂質體。

36. 如段落 33 所述之方法，其中該一種或多種病毒質粒包括一種或多種腺病毒、一種或多種慢病毒或一種或多種腺相關病毒。

37. 如段落 14-36 中任一項所述之方法，該方法係一種藉由操縱感興趣的基因組座位處的一個或多個靶序列來修飾細胞、細胞系或生物體的方法。

38. 一種來自如段落 37 所述的方法之細胞或其子代，其中該細胞包含在未經受該方法的細胞中不存在的修飾。

39. 如段落 38 所述之細胞或其子代，其中未經受該方法的該細胞包含異常並且來自該方法的該細胞的該異常已得到解決或校正。

40. 一種來自如段落 38 所述之細胞或其子代的細胞產物，其

中該產物係以相對於來自未經受該方法的細胞的細胞產物的性質或量來修飾的。

41. 如段落 40 所述之細胞產物，其中未經受該方法的該細胞包含異常並且該細胞產物反映了該異常已藉由該方法解決或校正。

42. 一種包含如段落 1-13 中任一項所述之系統之體外、離體或體內宿主細胞或細胞系或其子代。

43. 如段落 42 所述之宿主細胞或細胞系或其子代，其中該細胞係真核細胞。

44. 如段落 43 所述之宿主細胞或細胞系或其子代，其中該細胞係動物細胞。

45. 如段落 33 所述之宿主細胞或細胞系或其子代，其中該細胞係人類細胞。

46. 如段落 31 所述之宿主細胞或細胞系或其子代，包含幹細胞或幹細胞系。

47. 如段落 30 所述之宿主細胞或細胞系或其子代，其中該細胞係植物細胞。

48. 一種產生具有由感興趣的基因編碼的修飾的感興趣的性狀的植物之方法，所述方法包括將植物細胞與如段落 1-13 中任一項所述之系統接觸或者使該植物細胞經受如段落 14-17 或 19 至 37 所述之方法，由此修飾或引入所述感興趣的基因並且由所述植物細胞再生植物。

49. 一種鑒定植物中感興趣的性狀之方法，所述感興趣的性狀由感興趣的基因編碼，所述方法包括將植物細胞與如段落 1-13 中任一項所述之系統接觸或者使該植物細胞經受如段落 14-17 或 19 至 37 所述之方法，由此鑒定所述感興趣的基因。

50. 如段落 49 所述之方法，進一步包括將該鑒定的感興趣的基因引入到植物細胞或植物細胞系或植物種質中並且由其生成植物，由此該植物含有該感興趣的基因。

51. 如段落 50 所述之方法，其中該植物展現出該感興趣的性狀。

52. 一種包含如段落 1-13 中任一項所述之系統的粒子。

53. 如段落 52 所述的粒子，其中該粒子含有與該指導 RNA 複合的該 Cpf1 效應蛋白。

54. 如任何以上段落所述之系統或方法，其中該複合物、指導 RNA 或蛋白質被軛合至至少一種糖部分，視情況是 N-乙醯半乳糖胺（GalNAc），具體地是三觸角 GalNAc。

55. 如任何以上段落所述的系統或方法，其中 Mg^{2+}的濃度係約 1 mM 至約 15 mM。

56. 一種分離的蛋白質，該分離的蛋白質與AsCpf1或LbCpf1具有至少60%序列一致性並且能夠藉由與包含同向重複序列和指導序列的指導RNA複合來結合靶DNA，而不需要存在tracrRNA。

57. 一種編碼如段落56所述之蛋白質的分離的核酸。

58. 如段落17所述之方法，該方法係一種治療由所述細胞中的遺傳缺陷引起的疾病之方法。

59. 如段落58所述之方法，其中所述方法係在體內或離體細胞上進行的。

60. 一種非天然存在或工程化的組成物，該組成物包含Cpf1效應蛋白以及含有同向重複序列和能夠與在感興趣的座位處的靶DNA雜交的指導序列的一種或多種RNA，其中該Cpf1效應蛋白與該一種或多種指導RNA形成複合物並且在所述複合物與係原型間隔區相鄰模體（PAM）的3'的感興趣的靶座位結合後，該效應蛋白誘導該感興趣的靶座位的修飾。

61. 一種非天然存在或工程化的組成物，該組成物包含編碼Cpf1效應蛋白的多核苷酸序列以及含有同向重複序列和能夠與在感興趣的座位處的靶DNA雜交的指導序列的一種或多種RNA，其中該Cpf1效應蛋白在表現時與該一種或多種指導RNA形成複合物並且在所述複合物與係原型間隔區相鄰模體（PAM）的3'的感興趣的靶座位結合後，該效應蛋白誘導該感興趣的靶座位的修飾。

62. 如段落60或61所述的組成物，該組成物係醫藥組成物。

63. 如段落60或61所述的組成物，該組成物用作藥物。

64. 如段落60或61所述的組成物，該組成物用於治療由該感興趣的靶座位處的遺傳缺陷引起的疾病或病症。

65. 如段落58所述的方法或如陳述64所述地使用的組成物，其中該細胞係HSC細胞。

66. 如段落58所述的方法或如陳述64所述地使用的組成物，其中該疾病或病症係血細胞病症。

* * *

[1755] 儘管在此已經顯示並說明了本發明的較佳的實施方式，但是熟習該項技術者將清楚的是僅作為舉例而提供了此類實施方式。熟習該項技術者現在將會想到眾多變體、變化、以及替代，而不背離本發明。應該理解的是，在此說明的本發明的實施方式的不同替代方案可以用於實施本發明。預期的是以下申請專利範圍限定了本發明的範圍以及由此覆蓋在該等申請專利範圍和它們的等效物的範圍內的方法和結構。

【符號說明】

無。

序列表

<110> 博得學院股份有限公司 (THE BROAD INSTITUTE INC.)
麻省理工學院 (MASSACHUSETTS INSTITUTE OF TECHNOLOGY)
哈佛大學校董委員會 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE)

<120> 新型CRISPR酶以及系統

<140> 14/975,085
<141> 2015-12-18

<150> 62/232,067
<151> 2015-09-24

<150> 62/205,733
<151> 2015-08-16

<150> 62/201,542
<151> 2015-08-05

<150> 62/193,507
<151> 2015-07-16

<150> 62/181,739
<151> 2015-06-18

<160> 1595

<170> PatentIn版本3.5

<210> 1
<211> 7
<212> PRT
<213> 未知

<220>
<223> 未知的描述:
XerD序列的活性位點區域

<400> 1
Leu Tyr Trp Thr Gly Met Arg
1 5

<210> 2
<211> 7
<212> PRT
<213> 猿猴病毒40

<400> 2
Pro Lys Lys Lys Arg Lys Val
1 5

<210> 3
<211> 16
<212> PRT
<213> 未知

<220>
<223> 未知的描述:
核質蛋白二分NLS序列

<400> 3
Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys
1 5 10 15

<210> 4
<211> 9
<212> PRT
<213> 未知

<220>
<223> 未知的描述:
C-myc NLS序列

<400> 4
Pro Ala Ala Lys Arg Val Lys Leu Asp
1 5

<210> 5
<211> 11
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
C-myc NLS序列

<400> 5
Arg Gln Arg Arg Asn Glu Leu Lys Arg Ser Pro
1 5 10

<210> 6
<211> 38
<212> PRT
<213> 智人（Homo sapiens）

<400> 6
Asn Gln Ser Ser Asn Phe Gly Pro Met Lys Gly Gly Asn Phe Gly Gly
1 5 10 15

Arg Ser Ser Gly Pro Tyr Gly Gly Gly Gly Gln Tyr Phe Ala Lys Pro
20 25 30

Arg Asn Gln Gly Gly Tyr
35

<210> 7
<211> 42
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
來自輸入素α序列的IBB結構域

<400> 7
Arg Met Arg Ile Glx Phe Lys Asn Lys Gly Lys Asp Thr Ala Glu Leu
1 5 10 15

Arg Arg Arg Arg Val Glu Val Ser Val Glu Leu Arg Lys Ala Lys Lys
20 25 30

Asp Glu Gln Ile Leu Lys Arg Arg Asn Val
35 40

<210> 8
<211> 8
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
肌瘤T蛋白序列

<400> 8
Val Ser Arg Lys Arg Pro Arg Pro
1 5

<210> 9
<211> 8
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
肌瘤T蛋白序列

<400> 9
Pro Pro Lys Lys Ala Arg Glu Asp
1 5

<210> 10
<211> 8
<212> PRT
<213> 智人

<400> 10
Pro Gln Pro Lys Lys Lys Pro Leu
1 5

<210> 11
<211> 12
<212> PRT
<213> 小家鼠（Mus musculus）

<400> 11
Ser Ala Leu Ile Lys Lys Lys Lys Lys Met Ala Pro
1 5 10

<210> 12
<211> 5
<212> PRT
<213> 流感病毒

<400> 12
Asp Arg Leu Arg Arg
1 5

<210> 13
<211> 7
<212> PRT
<213> 流感病毒

<400> 13
Pro Lys Gln Lys Lys Arg Lys
1 5

<210> 14
<211> 10
<212> PRT
<213> 肝炎δ病毒

<400> 14
Arg Lys Leu Lys Lys Lys Ile Lys Lys Leu
1 5 10

<210> 15
<211> 10
<212> PRT
<213> 小家鼠

<400> 15
Arg Glu Lys Lys Lys Phe Leu Lys Arg Arg
1 5 10

<210> 16
<211> 20
<212> PRT
<213> 智人

<400> 16
Lys Arg Lys Gly Asp Glu Val Asp Gly Val Asp Glu Val Ala Lys Lys
1 5 10 15

Lys Ser Lys Lys
20

<210> 17
<211> 17
<212> PRT
<213> 智人

<400> 17
Arg Lys Cys Leu Gln Ala Gly Met Asn Leu Glu Ala Arg Lys Thr Lys
1 5 10 15

Lys

<210> 18
<211> 4
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
肽

<400> 18
Gly Gly Gly Ser
1

<210> 19
<211> 15
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
肽

<400> 19
Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
1 5 10 15

<210> 20
<211> 30
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 20
Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
1 5 10 15

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
20 25 30

<210> 21
<211> 45
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 21
Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
1 5 10 15

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
20 25 30

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
35 40 45

<210> 22
<211> 60
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 22
Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
1 5 10 15

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
20 25 30

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly
35 40 45

Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
50 55 60

<210> 23
<211> 20
<212> DNA
<213> 智人

<400> 23
gagtccgagc agaagaagaa 20

<210> 24
<211> 20
<212> DNA
<213> 智人

<400> 24
gagtcctagc aggagaagaa 20

<210> 25
<211> 20
<212> DNA
<213> 智人

<400> 25
gagtctaagc agaagaagaa 20

<210> 26
<211> 9
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
“LAGLIDADG”家族模體肽

<400> 26
Leu Ala Gly Leu Ile Asp Ala Asp Gly
1 5

<210> 27
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 27
gccgcagcga augccguuuc acgaaucguc aggcgg 36

<210> 28
<211> 75
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 28
gcuggagacg uuuuuugaaa cggcgagugc ugcggauagc gaguuucucu uggggaggcg 60

cucgcggcca cuuuu 75

<210> 29

<211> 1388
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
豐祐菌科（Opitutaceae）細菌序列

<400> 29

```
Met Ser Leu Asn Arg Ile Tyr Gln Gly Arg Val Ala Ala Val Glu Thr
1               5                  10                  15

Gly Thr Ala Leu Ala Lys Gly Asn Val Glu Trp Met Pro Ala Ala Gly
            20                  25                  30

Gly Asp Glu Val Leu Trp Gln His His Glu Leu Phe Gln Ala Ala Ile
        35                  40                  45

Asn Tyr Tyr Leu Val Ala Leu Leu Ala Leu Ala Asp Lys Asn Asn Pro
    50                  55                  60

Val Leu Gly Pro Leu Ile Ser Gln Met Asp Asn Pro Gln Ser Pro Tyr
65                  70                  75                  80

His Val Trp Gly Ser Phe Arg Arg Gln Gly Arg Gln Arg Thr Gly Leu
                85                  90                  95

Ser Gln Ala Val Ala Pro Tyr Ile Thr Pro Gly Asn Asn Ala Pro Thr
            100                 105                 110

Leu Asp Glu Val Phe Arg Ser Ile Leu Ala Gly Asn Pro Thr Asp Arg
        115                 120                 125

Ala Thr Leu Asp Ala Ala Leu Met Gln Leu Leu Lys Ala Cys Asp Gly
    130                 135                 140

Ala Gly Ala Ile Gln Gln Glu Gly Arg Ser Tyr Trp Pro Lys Phe Cys
145                 150                 155                 160

Asp Pro Asp Ser Thr Ala Asn Phe Ala Gly Asp Pro Ala Met Leu Arg
                165                 170                 175

Arg Glu Gln His Arg Leu Leu Leu Pro Gln Val Leu His Asp Pro Ala
            180                 185                 190

Ile Thr His Asp Ser Pro Ala Leu Gly Ser Phe Asp Thr Tyr Ser Ile
        195                 200                 205

Ala Thr Pro Asp Thr Arg Thr Pro Gln Leu Thr Gly Pro Lys Ala Arg
    210                 215                 220

Ala Arg Leu Glu Gln Ala Ile Thr Leu Trp Arg Val Arg Leu Pro Glu
225                 230                 235                 240

Ser Ala Ala Asp Phe Asp Arg Leu Ala Ser Ser Leu Lys Lys Ile Pro
                245                 250                 255

Asp Asp Asp Ser Arg Leu Asn Leu Gln Gly Tyr Val Gly Ser Ser Ala
            260                 265                 270

Lys Gly Glu Val Gln Ala Arg Leu Phe Ala Leu Leu Leu Phe Arg His
        275                 280                 285

Leu Glu Arg Ser Ser Phe Thr Leu Gly Leu Leu Arg Ser Ala Thr Pro
    290                 295                 300
```

```
Pro Pro Lys Asn Ala Glu Thr Pro Pro Pro Ala Gly Val Pro Leu Pro
305                 310                 315                 320

Ala Ala Ser Ala Ala Asp Pro Val Arg Ile Ala Arg Gly Lys Arg Ser
                325                 330                 335

Phe Val Phe Arg Ala Phe Thr Ser Leu Pro Cys Trp His Gly Gly Asp
            340                 345                 350

Asn Ile His Pro Thr Trp Lys Ser Phe Asp Ile Ala Ala Phe Lys Tyr
        355                 360                 365

Ala Leu Thr Val Ile Asn Gln Ile Glu Glu Lys Thr Lys Glu Arg Gln
    370                 375                 380

Lys Glu Cys Ala Glu Leu Glu Thr Asp Phe Asp Tyr Met His Gly Arg
385                 390                 395                 400

Leu Ala Lys Ile Pro Val Lys Tyr Thr Thr Gly Glu Ala Glu Pro Pro
                405                 410                 415

Pro Ile Leu Ala Asn Asp Leu Arg Ile Pro Leu Leu Arg Glu Leu Leu
            420                 425                 430

Gln Asn Ile Lys Val Asp Thr Ala Leu Thr Asp Gly Glu Ala Val Ser
        435                 440                 445

Tyr Gly Leu Gln Arg Arg Thr Ile Arg Gly Phe Arg Glu Leu Arg Arg
    450                 455                 460

Ile Trp Arg Gly His Ala Pro Ala Gly Thr Val Phe Ser Ser Glu Leu
465                 470                 475                 480

Lys Glu Lys Leu Ala Gly Glu Leu Arg Gln Phe Gln Thr Asp Asn Ser
                485                 490                 495

Thr Thr Ile Gly Ser Val Gln Leu Phe Asn Glu Leu Ile Gln Asn Pro
            500                 505                 510

Lys Tyr Trp Pro Ile Trp Gln Ala Pro Asp Val Glu Thr Ala Arg Gln
        515                 520                 525

Trp Ala Asp Ala Gly Phe Ala Asp Asp Pro Leu Ala Ala Leu Val Gln
    530                 535                 540

Glu Ala Glu Leu Gln Glu Asp Ile Asp Ala Leu Lys Ala Pro Val Lys
545                 550                 555                 560

Leu Thr Pro Ala Asp Pro Glu Tyr Ser Arg Arg Gln Tyr Asp Phe Asn
                565                 570                 575

Ala Val Ser Lys Phe Gly Ala Gly Ser Arg Ser Ala Asn Arg His Glu
            580                 585                 590

Pro Gly Gln Thr Glu Arg Gly His Asn Thr Phe Thr Thr Glu Ile Ala
        595                 600                 605

Ala Arg Asn Ala Ala Asp Gly Asn Arg Trp Arg Ala Thr His Val Arg
    610                 615                 620

Ile His Tyr Ser Ala Pro Arg Leu Leu Arg Asp Gly Leu Arg Arg Pro
625                 630                 635                 640
```

Asp Thr Asp Gly Asn Glu Ala Leu Glu Ala Val Pro Trp Leu Gln Pro
645 650 655

Met Met Glu Ala Leu Ala Pro Leu Pro Thr Leu Pro Gln Asp Leu Thr
660 665 670

Gly Met Pro Val Phe Leu Met Pro Asp Val Thr Leu Ser Gly Glu Arg
675 680 685

Arg Ile Leu Leu Asn Leu Pro Val Thr Leu Glu Pro Ala Ala Leu Val
690 695 700

Glu Gln Leu Gly Asn Ala Gly Arg Trp Gln Asn Gln Phe Phe Gly Ser
705 710 715 720

Arg Glu Asp Pro Phe Ala Leu Arg Trp Pro Ala Asp Gly Ala Val Lys
725 730 735

Thr Ala Lys Gly Lys Thr His Ile Pro Trp His Gln Asp Arg Asp His
740 745 750

Phe Thr Val Leu Gly Val Asp Leu Gly Thr Arg Asp Ala Gly Ala Leu
755 760 765

Ala Leu Leu Asn Val Thr Ala Gln Lys Pro Ala Lys Pro Val His Arg
770 775 780

Ile Ile Gly Glu Ala Asp Gly Arg Thr Trp Tyr Ala Ser Leu Ala Asp
785 790 795 800

Ala Arg Met Ile Arg Leu Pro Gly Glu Asp Ala Arg Leu Phe Val Arg
805 810 815

Gly Lys Leu Val Gln Glu Pro Tyr Gly Glu Arg Gly Arg Asn Ala Ser
820 825 830

Leu Leu Glu Trp Glu Asp Ala Arg Asn Ile Ile Leu Arg Leu Gly Gln
835 840 845

Asn Pro Asp Glu Leu Leu Gly Ala Asp Pro Arg Arg His Ser Tyr Pro
850 855 860

Glu Ile Asn Asp Lys Leu Leu Val Ala Leu Arg Arg Ala Gln Ala Arg
865 870 875 880

Leu Ala Arg Leu Gln Asn Arg Ser Trp Arg Leu Arg Asp Leu Ala Glu
885 890 895

Ser Asp Lys Ala Leu Asp Glu Ile His Ala Glu Arg Ala Gly Glu Lys
900 905 910

Pro Ser Pro Leu Pro Pro Leu Ala Arg Asp Asp Ala Ile Lys Ser Thr
915 920 925

Asp Glu Ala Leu Leu Ser Gln Arg Asp Ile Ile Arg Arg Ser Phe Val
930 935 940

Gln Ile Ala Asn Leu Ile Leu Pro Leu Arg Gly Arg Arg Trp Glu Trp
945 950 955 960

Arg Pro His Val Glu Val Pro Asp Cys His Ile Leu Ala Gln Ser Asp
965 970 975

```
Pro Gly Thr Asp Asp Thr Lys Arg Leu Val Ala Gly Gln Arg Gly Ile
           980                 985                 990

Ser His Glu Arg Ile Glu Gln Ile Glu Glu Leu Arg Arg Arg Cys Gln
       995                1000                1005

Ser Leu  Asn Arg Ala Leu Arg  His Lys Pro Gly Glu  Arg Pro Val
    1010                1015                1020

Leu Gly  Arg Pro Ala Lys Gly  Glu Glu Ile Ala Asp  Pro Cys Pro
    1025                1030                1035

Ala Leu  Leu Glu Lys Ile Asn  Arg Leu Arg Asp Gln  Arg Val Asp
    1040                1045                1050

Gln Thr  Ala His Ala Ile Leu  Ala Ala Ala Leu Gly  Val Arg Leu
    1055                1060                1065

Arg Ala  Pro Ser Lys Asp Arg  Ala Glu Arg Arg His  Arg Asp Ile
    1070                1075                1080

His Gly  Glu Tyr Glu Arg Phe  Arg Ala Pro Ala Asp  Phe Val Val
    1085                1090                1095

Ile Glu  Asn Leu Ser Arg Tyr  Leu Ser Ser Gln Asp  Arg Ala Arg
    1100                1105                1110

Ser Glu  Asn Thr Arg Leu Met  Gln Trp Cys His Arg  Gln Ile Val
    1115                1120                1125

Gln Lys  Leu Arg Gln Leu Cys  Glu Thr Tyr Gly Ile  Pro Val Leu
    1130                1135                1140

Ala Val  Pro Ala Ala Tyr Ser  Ser Arg Phe Ser Ser  Arg Asp Gly
    1145                1150                1155

Ser Ala  Gly Phe Arg Ala Val  His Leu Thr Pro Asp  His Arg His
    1160                1165                1170

Arg Met  Pro Trp Ser Arg Ile  Leu Ala Arg Leu Lys  Ala His Glu
    1175                1180                1185

Glu Asp  Gly Lys Arg Leu Glu  Lys Thr Val Leu Asp  Glu Ala Arg
    1190                1195                1200

Ala Val  Arg Gly Leu Phe Asp  Arg Leu Asp Arg Phe  Asn Ala Gly
    1205                1210                1215

His Val  Pro Gly Lys Pro Trp  Arg Thr Leu Leu Ala  Pro Leu Pro
    1220                1225                1230

Gly Gly  Pro Val Phe Val Pro  Leu Gly Asp Ala Thr  Pro Met Gln
    1235                1240                1245

Ala Asp  Leu Asn Ala Ala Ile  Asn Ile Ala Leu Arg  Gly Ile Ala
    1250                1255                1260

Ala Pro  Asp Arg His Asp Ile  His His Arg Leu Arg  Ala Glu Asn
    1265                1270                1275

Lys Lys  Arg Ile Leu Ser Leu  Arg Leu Gly Thr Gln  Arg Glu Lys
    1280                1285                1290
```

Ala Arg Trp Pro Gly Gly Ala Pro Ala Val Thr Leu Ser Thr Pro
1295 1300 1305

Asn Asn Gly Ala Ser Pro Glu Asp Ser Asp Ala Leu Pro Glu Arg
1310 1315 1320

Val Ser Asn Leu Phe Val Asp Ile Ala Gly Val Ala Asn Phe Glu
1325 1330 1335

Arg Val Thr Ile Glu Gly Val Ser Gln Lys Phe Ala Thr Gly Arg
1340 1345 1350

Gly Leu Trp Ala Ser Val Lys Gln Arg Ala Trp Asn Arg Val Ala
1355 1360 1365

Arg Leu Asn Glu Thr Val Thr Asp Asn Asn Arg Asn Glu Glu Glu
1370 1375 1380

Asp Asp Ile Pro Met
1385

<210> 30
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 30
guccaagaaa aaagaaauga uacgaggcau uagcac 36

<210> 31
<211> 107
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 31
cuggacgaug ucucuuuuau uucuuuuuuc uuggaucuga guacgagcac ccacauugga 60

cauuucgcau ggugggugcu cguacuauag guaaaacaaa ccuuuuu 107

<210> 32
<211> 1108
<212> PRT
<213> 嗜熱澱粉芽孢桿菌（Bacillus thermoamylovorans）

<400> 32
Met Ala Thr Arg Ser Phe Ile Leu Lys Ile Glu Pro Asn Glu Glu Val
1 5 10 15

Lys Lys Gly Leu Trp Lys Thr His Glu Val Leu Asn His Gly Ile Ala
20 25 30

Tyr Tyr Met Asn Ile Leu Lys Leu Ile Arg Gln Glu Ala Ile Tyr Glu
35 40 45

His His Glu Gln Asp Pro Lys Asn Pro Lys Lys Val Ser Lys Ala Glu
50 55 60

Ile Gln Ala Glu Leu Trp Asp Phe Val Leu Lys Met Gln Lys Cys Asn
65 70 75 80

Ser Phe Thr His Glu Val Asp Lys Asp Val Val Phe Asn Ile Leu Arg

85 90 95

Glu Leu Tyr Glu Glu Leu Val Pro Ser Ser Val Glu Lys Lys Gly Glu
100 105 110

Ala Asn Gln Leu Ser Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn
115 120 125

Ser Gln Ser Gly Lys Gly Thr Ala Ser Ser Gly Arg Lys Pro Arg Trp
130 135 140

Tyr Asn Leu Lys Ile Ala Gly Asp Pro Ser Trp Glu Glu Glu Lys Lys
145 150 155 160

Lys Trp Glu Glu Asp Lys Lys Lys Asp Pro Leu Ala Lys Ile Leu Gly
165 170 175

Lys Leu Ala Glu Tyr Gly Leu Ile Pro Leu Phe Ile Pro Phe Thr Asp
180 185 190

Ser Asn Glu Pro Ile Val Lys Glu Ile Lys Trp Met Glu Lys Ser Arg
195 200 205

Asn Gln Ser Val Arg Arg Leu Asp Lys Asp Met Phe Ile Gln Ala Leu
210 215 220

Glu Arg Phe Leu Ser Trp Glu Ser Trp Asn Leu Lys Val Lys Glu Glu
225 230 235 240

Tyr Glu Lys Val Glu Lys Glu His Lys Thr Leu Glu Glu Arg Ile Lys
245 250 255

Glu Asp Ile Gln Ala Phe Lys Ser Leu Glu Gln Tyr Glu Lys Glu Arg
260 265 270

Gln Glu Gln Leu Leu Arg Asp Thr Leu Asn Thr Asn Glu Tyr Arg Leu
275 280 285

Ser Lys Arg Gly Leu Arg Gly Trp Arg Glu Ile Ile Gln Lys Trp Leu
290 295 300

Lys Met Asp Glu Asn Glu Pro Ser Glu Lys Tyr Leu Glu Val Phe Lys
305 310 315 320

Asp Tyr Gln Arg Lys His Pro Arg Glu Ala Gly Asp Tyr Ser Val Tyr
325 330 335

Glu Phe Leu Ser Lys Lys Glu Asn His Phe Ile Trp Arg Asn His Pro
340 345 350

Glu Tyr Pro Tyr Leu Tyr Ala Thr Phe Cys Glu Ile Asp Lys Lys Lys
355 360 365

Lys Asp Ala Lys Gln Gln Ala Thr Phe Thr Leu Ala Asp Pro Ile Asn
370 375 380

His Pro Leu Trp Val Arg Phe Glu Glu Arg Ser Gly Ser Asn Leu Asn
385 390 395 400

Lys Tyr Arg Ile Leu Thr Glu Gln Leu His Thr Glu Lys Leu Lys Lys
405 410 415

Lys Leu Thr Val Gln Leu Asp Arg Leu Ile Tyr Pro Thr Glu Ser Gly

420 425 430

Gly Trp Glu Glu Lys Gly Lys Val Asp Ile Val Leu Leu Pro Ser Arg
435 440 445

Gln Phe Tyr Asn Gln Ile Phe Leu Asp Ile Glu Glu Lys Gly Lys His
450 455 460

Ala Phe Thr Tyr Lys Asp Glu Ser Ile Lys Phe Pro Leu Lys Gly Thr
465 470 475 480

Leu Gly Gly Ala Arg Val Gln Phe Asp Arg Asp His Leu Arg Arg Tyr
485 490 495

Pro His Lys Val Glu Ser Gly Asn Val Gly Arg Ile Tyr Phe Asn Met
500 505 510

Thr Val Asn Ile Glu Pro Thr Glu Ser Pro Val Ser Lys Ser Leu Lys
515 520 525

Ile His Arg Asp Asp Phe Pro Lys Phe Val Asn Phe Lys Pro Lys Glu
530 535 540

Leu Thr Glu Trp Ile Lys Asp Ser Lys Gly Lys Lys Leu Lys Ser Gly
545 550 555 560

Ile Glu Ser Leu Glu Ile Gly Leu Arg Val Met Ser Ile Asp Leu Gly
565 570 575

Gln Arg Gln Ala Ala Ala Ala Ser Ile Phe Glu Val Val Asp Gln Lys
580 585 590

Pro Asp Ile Glu Gly Lys Leu Phe Phe Pro Ile Lys Gly Thr Glu Leu
595 600 605

Tyr Ala Val His Arg Ala Ser Phe Asn Ile Lys Leu Pro Gly Glu Thr
610 615 620

Leu Val Lys Ser Arg Glu Val Leu Arg Lys Ala Arg Glu Asp Asn Leu
625 630 635 640

Lys Leu Met Asn Gln Lys Leu Asn Phe Leu Arg Asn Val Leu His Phe
645 650 655

Gln Gln Phe Glu Asp Ile Thr Glu Arg Glu Lys Arg Val Thr Lys Trp
660 665 670

Ile Ser Arg Gln Glu Asn Ser Asp Val Pro Leu Val Tyr Gln Asp Glu
675 680 685

Leu Ile Gln Ile Arg Glu Leu Met Tyr Lys Pro Tyr Lys Asp Trp Val
690 695 700

Ala Phe Leu Lys Gln Leu His Lys Arg Leu Glu Val Glu Ile Gly Lys
705 710 715 720

Glu Val Lys His Trp Arg Lys Ser Leu Ser Asp Gly Arg Lys Gly Leu
725 730 735

Tyr Gly Ile Ser Leu Lys Asn Ile Asp Glu Ile Asp Arg Thr Arg Lys
740 745 750

Phe Leu Leu Arg Trp Ser Leu Arg Pro Thr Glu Pro Gly Glu Val Arg

```
        755                 760                 765

Arg Leu Glu Pro Gly Gln Arg Phe Ala Ile Asp Gln Leu Asn His Leu
    770                 775                 780

Asn Ala Leu Lys Glu Asp Arg Leu Lys Lys Met Ala Asn Thr Ile Ile
785                 790                 795                 800

Met His Ala Leu Gly Tyr Cys Tyr Asp Val Arg Lys Lys Lys Trp Gln
                805                 810                 815

Ala Lys Asn Pro Ala Cys Gln Ile Ile Leu Phe Glu Asp Leu Ser Asn
            820                 825                 830

Tyr Asn Pro Tyr Glu Glu Arg Ser Arg Phe Glu Asn Ser Lys Leu Met
        835                 840                 845

Lys Trp Ser Arg Arg Glu Ile Pro Arg Gln Val Ala Leu Gln Gly Glu
    850                 855                 860

Ile Tyr Gly Leu Gln Val Gly Glu Val Gly Ala Gln Phe Ser Ser Arg
865                 870                 875                 880

Phe His Ala Lys Thr Gly Ser Pro Gly Ile Arg Cys Ser Val Val Thr
                885                 890                 895

Lys Glu Lys Leu Gln Asp Asn Arg Phe Phe Lys Asn Leu Gln Arg Glu
            900                 905                 910

Gly Arg Leu Thr Leu Asp Lys Ile Ala Val Leu Lys Glu Gly Asp Leu
        915                 920                 925

Tyr Pro Asp Lys Gly Gly Glu Lys Phe Ile Ser Leu Ser Lys Asp Arg
    930                 935                 940

Lys Leu Val Thr Thr His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln
945                 950                 955                 960

Lys Arg Phe Trp Thr Arg Thr His Gly Phe Tyr Lys Val Tyr Cys Lys
                965                 970                 975

Ala Tyr Gln Val Asp Gly Gln Thr Val Tyr Ile Pro Glu Ser Lys Asp
            980                 985                 990

Gln Lys Gln Lys Ile Ile Glu Glu Phe Gly Glu Gly Tyr Phe Ile Leu
        995                 1000                1005

Lys Asp  Gly Val Tyr Glu Trp Gly Asn Ala Gly  Lys Leu Lys Ile
    1010                1015                1020

Lys Lys  Gly Ser Ser Lys Gln Ser Ser Ser Glu  Leu Val Asp Ser
    1025                1030                1035

Asp Ile  Leu Lys Asp Ser Phe Asp Leu Ala Ser  Glu Leu Lys Gly
    1040                1045                1050

Glu Lys  Leu Met Leu Tyr Arg Asp Pro Ser Gly  Asn Val Phe Pro
    1055                1060                1065

Ser Asp  Lys Trp Met Ala Ala Gly Val Phe Phe  Gly Lys Leu Glu
    1070                1075                1080

Arg Ile  Leu Ile Ser Lys Leu Thr Asn Gln Tyr  Ser Ile Ser Thr
```

1085 1090 1095

Ile Glu Asp Asp Ser Ser Lys Gln Ser Met
1100 1105

<210> 33
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 33
guucgaaagc uuaguggaaa gcuucguggu uagcac 36

<210> 34
<211> 69
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 34
cacggauaau cacgacuuuc cacuaagcuu ucgaauuuua ugaugcgagc auccucucag 60

gucaaaaaa 69

<210> 35
<211> 1108
<212> PRT
<213> 芽孢桿菌屬某種（Bacillus sp.）

<400> 35
Met Ala Ile Arg Ser Ile Lys Leu Lys Leu Lys Thr His Thr Gly Pro
1 5 10 15

Glu Ala Gln Asn Leu Arg Lys Gly Ile Trp Arg Thr His Arg Leu Leu
20 25 30

Asn Glu Gly Val Ala Tyr Tyr Met Lys Met Leu Leu Leu Phe Arg Gln
35 40 45

Glu Ser Thr Gly Glu Arg Pro Lys Glu Glu Leu Gln Glu Glu Leu Ile
50 55 60

Cys His Ile Arg Glu Gln Gln Gln Arg Asn Gln Ala Asp Lys Asn Thr
65 70 75 80

Gln Ala Leu Pro Leu Asp Lys Ala Leu Glu Ala Leu Arg Gln Leu Tyr
85 90 95

Glu Leu Leu Val Pro Ser Ser Val Gly Gln Ser Gly Asp Ala Gln Ile
100 105 110

Ile Ser Arg Lys Phe Leu Ser Pro Leu Val Asp Pro Asn Ser Glu Gly
115 120 125

Gly Lys Gly Thr Ser Lys Ala Gly Ala Lys Pro Thr Trp Gln Lys Lys
130 135 140

Lys Glu Ala Asn Asp Pro Thr Trp Glu Gln Asp Tyr Glu Lys Trp Lys
145 150 155 160

Lys Arg Arg Glu Glu Asp Pro Thr Ala Ser Val Ile Thr Thr Leu Glu
165 170 175

```
Glu Tyr Gly Ile Arg Pro Ile Phe Pro Leu Tyr Thr Asn Thr Val Thr
            180                 185                 190

Asp Ile Ala Trp Leu Pro Leu Gln Ser Asn Gln Phe Val Arg Thr Trp
        195                 200                 205

Asp Arg Asp Met Leu Gln Gln Ala Ile Glu Arg Leu Leu Ser Trp Glu
    210                 215                 220

Ser Trp Asn Lys Arg Val Gln Glu Glu Tyr Ala Lys Leu Lys Glu Lys
225                 230                 235                 240

Met Ala Gln Leu Asn Glu Gln Leu Glu Gly Gly Gln Glu Trp Ile Ser
                245                 250                 255

Leu Leu Glu Gln Tyr Glu Glu Asn Arg Glu Arg Glu Leu Arg Glu Asn
            260                 265                 270

Met Thr Ala Ala Asn Asp Lys Tyr Arg Ile Thr Lys Arg Gln Met Lys
        275                 280                 285

Gly Trp Asn Glu Leu Tyr Glu Leu Trp Ser Thr Phe Pro Ala Ser Ala
    290                 295                 300

Ser His Glu Gln Tyr Lys Glu Ala Leu Lys Arg Val Gln Gln Arg Leu
305                 310                 315                 320

Arg Gly Arg Phe Gly Asp Ala His Phe Phe Gln Tyr Leu Met Glu Glu
                325                 330                 335

Lys Asn Arg Leu Ile Trp Lys Gly Asn Pro Gln Arg Ile His Tyr Phe
            340                 345                 350

Val Ala Arg Asn Glu Leu Thr Lys Arg Leu Glu Glu Ala Lys Gln Ser
        355                 360                 365

Ala Thr Met Thr Leu Pro Asn Ala Arg Lys His Pro Leu Trp Val Arg
    370                 375                 380

Phe Asp Ala Arg Gly Gly Asn Leu Gln Asp Tyr Tyr Leu Thr Ala Glu
385                 390                 395                 400

Ala Asp Lys Pro Arg Ser Arg Arg Phe Val Thr Phe Ser Gln Leu Ile
                405                 410                 415

Trp Pro Ser Glu Ser Gly Trp Met Glu Lys Lys Asp Val Glu Val Glu
            420                 425                 430

Leu Ala Leu Ser Arg Gln Phe Tyr Gln Gln Val Lys Leu Leu Lys Asn
        435                 440                 445

Asp Lys Gly Lys Gln Lys Ile Glu Phe Lys Asp Lys Gly Ser Gly Ser
    450                 455                 460

Thr Phe Asn Gly His Leu Gly Gly Ala Lys Leu Gln Leu Glu Arg Gly
465                 470                 475                 480

Asp Leu Glu Lys Glu Glu Lys Asn Phe Glu Asp Gly Glu Ile Gly Ser
                485                 490                 495

Val Tyr Leu Asn Val Val Ile Asp Phe Glu Pro Leu Gln Glu Val Lys
            500                 505                 510
```

Asn Gly Arg Val Gln Ala Pro Tyr Gly Gln Val Leu Gln Leu Ile Arg
515 520 525

Arg Pro Asn Glu Phe Pro Lys Val Thr Thr Tyr Lys Ser Glu Gln Leu
530 535 540

Val Glu Trp Ile Lys Ala Ser Pro Gln His Ser Ala Gly Val Glu Ser
545 550 555 560

Leu Ala Ser Gly Phe Arg Val Met Ser Ile Asp Leu Gly Leu Arg Ala
565 570 575

Ala Ala Ala Thr Ser Ile Phe Ser Val Glu Glu Ser Ser Asp Lys Asn
580 585 590

Ala Ala Asp Phe Ser Tyr Trp Ile Glu Gly Thr Pro Leu Val Ala Val
595 600 605

His Gln Arg Ser Tyr Met Leu Arg Leu Pro Gly Glu Gln Val Glu Lys
610 615 620

Gln Val Met Glu Lys Arg Asp Glu Arg Phe Gln Leu His Gln Arg Val
625 630 635 640

Lys Phe Gln Ile Arg Val Leu Ala Gln Ile Met Arg Met Ala Asn Lys
645 650 655

Gln Tyr Gly Asp Arg Trp Asp Glu Leu Asp Ser Leu Lys Gln Ala Val
660 665 670

Glu Gln Lys Lys Ser Pro Leu Asp Gln Thr Asp Arg Thr Phe Trp Glu
675 680 685

Gly Ile Val Cys Asp Leu Thr Lys Val Leu Pro Arg Asn Glu Ala Asp
690 695 700

Trp Glu Gln Ala Val Val Gln Ile His Arg Lys Ala Glu Glu Tyr Val
705 710 715 720

Gly Lys Ala Val Gln Ala Trp Arg Lys Arg Phe Ala Ala Asp Glu Arg
725 730 735

Lys Gly Ile Ala Gly Leu Ser Met Trp Asn Ile Glu Glu Leu Glu Gly
740 745 750

Leu Arg Lys Leu Leu Ile Ser Trp Ser Arg Arg Thr Arg Asn Pro Gln
755 760 765

Glu Val Asn Arg Phe Glu Arg Gly His Thr Ser His Gln Arg Leu Leu
770 775 780

Thr His Ile Gln Asn Val Lys Glu Asp Arg Leu Lys Gln Leu Ser His
785 790 795 800

Ala Ile Val Met Thr Ala Leu Gly Tyr Val Tyr Asp Glu Arg Lys Gln
805 810 815

Glu Trp Cys Ala Glu Tyr Pro Ala Cys Gln Val Ile Leu Phe Glu Asn
820 825 830

Leu Ser Gln Tyr Arg Ser Asn Leu Asp Arg Ser Thr Lys Glu Asn Ser
835 840 845

```
Thr Leu Met Lys Trp Ala His Arg Ser Ile Pro Lys Tyr Val His Met
    850                 855                 860

Gln Ala Glu Pro Tyr Gly Ile Gln Ile Gly Asp Val Arg Ala Glu Tyr
865                 870                 875                 880

Ser Ser Arg Phe Tyr Ala Lys Thr Gly Thr Pro Gly Ile Arg Cys Lys
                885                 890                 895

Lys Val Arg Gly Gln Asp Leu Gln Gly Arg Arg Phe Glu Asn Leu Gln
            900                 905                 910

Lys Arg Leu Val Asn Glu Gln Phe Leu Thr Glu Glu Gln Val Lys Gln
        915                 920                 925

Leu Arg Pro Gly Asp Ile Val Pro Asp Asp Ser Gly Glu Leu Phe Met
    930                 935                 940

Thr Leu Thr Asp Gly Ser Gly Ser Lys Glu Val Val Phe Leu Gln Ala
945                 950                 955                 960

Asp Ile Asn Ala Ala His Asn Leu Gln Lys Arg Phe Trp Gln Arg Tyr
                965                 970                 975

Asn Glu Leu Phe Lys Val Ser Cys Arg Val Ile Val Arg Asp Glu Glu
            980                 985                 990

Glu Tyr Leu Val Pro Lys Thr Lys Ser Val Gln Ala Lys Leu Gly Lys
        995                 1000                1005

Gly Leu  Phe Val Lys Lys Ser Asp Thr Ala Trp  Lys Asp Val Tyr
    1010                1015                1020

Val Trp  Asp Ser Gln Ala Lys Leu Lys Gly Lys  Thr Thr Phe Thr
    1025                1030                1035

Glu Glu  Ser Glu Ser Pro Glu Gln Leu Glu Asp  Phe Gln Glu Ile
    1040                1045                1050

Ile Glu  Glu Ala Glu Glu Ala Lys Gly Thr Tyr  Arg Thr Leu Phe
    1055                1060                1065

Arg Asp  Pro Ser Gly Val Phe Phe Pro Glu Ser  Val Trp Tyr Pro
    1070                1075                1080

Gln Lys  Asp Phe Trp Gly Glu Val Lys Arg Lys  Leu Tyr Gly Lys
    1085                1090                1095

Leu Arg  Glu Arg Phe Leu Thr Lys Ala Arg
    1100                1105
```

<210> 36
<211> 35
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 36
guuuugagaa uagcccgaca uagagggcaa uagac 35

<210> 37
<211> 36

<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 37
guuaugaaaa cagcccgaca uagagggcaa uagaca 36

<210> 38
<211> 1334
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科（Lachnospiraceae）細菌序列

<400> 38
Met Lys Ile Ser Lys Val Asp His Thr Arg Met Ala Val Ala Lys Gly
1 5 10 15

Asn Gln His Arg Arg Asp Glu Ile Ser Gly Ile Leu Tyr Lys Asp Pro
20 25 30

Thr Lys Thr Gly Ser Ile Asp Phe Asp Glu Arg Phe Lys Lys Leu Asn
35 40 45

Cys Ser Ala Lys Ile Leu Tyr His Val Phe Asn Gly Ile Ala Glu Gly
50 55 60

Ser Asn Lys Tyr Lys Asn Ile Val Asp Lys Val Asn Asn Asn Leu Asp
65 70 75 80

Arg Val Leu Phe Thr Gly Lys Ser Tyr Asp Arg Lys Ser Ile Ile Asp
85 90 95

Ile Asp Thr Val Leu Arg Asn Val Glu Lys Ile Asn Ala Phe Asp Arg
100 105 110

Ile Ser Thr Glu Glu Arg Glu Gln Ile Ile Asp Asp Leu Leu Glu Ile
115 120 125

Gln Leu Arg Lys Gly Leu Arg Lys Gly Lys Ala Gly Leu Arg Glu Val
130 135 140

Leu Leu Ile Gly Ala Gly Val Ile Val Arg Thr Asp Lys Lys Gln Glu
145 150 155 160

Ile Ala Asp Phe Leu Glu Ile Leu Asp Glu Asp Phe Asn Lys Thr Asn
165 170 175

Gln Ala Lys Asn Ile Lys Leu Ser Ile Glu Asn Gln Gly Leu Val Val
180 185 190

Ser Pro Val Ser Arg Gly Glu Glu Arg Ile Phe Asp Val Ser Gly Ala
195 200 205

Gln Lys Gly Lys Ser Ser Lys Lys Ala Gln Glu Lys Glu Ala Leu Ser
210 215 220

Ala Phe Leu Leu Asp Tyr Ala Asp Leu Asp Lys Asn Val Arg Phe Glu
225 230 235 240

Tyr Leu Arg Lys Ile Arg Arg Leu Ile Asn Leu Tyr Phe Tyr Val Lys
245 250 255

```
Asn Asp Asp Val Met Ser Leu Thr Glu Ile Pro Ala Glu Val Asn Leu
            260                 265                 270

Glu Lys Asp Phe Asp Ile Trp Arg Asp His Glu Gln Arg Lys Glu Glu
        275                 280                 285

Asn Gly Asp Phe Val Gly Cys Pro Asp Ile Leu Leu Ala Asp Arg Asp
    290                 295                 300

Val Lys Lys Ser Asn Ser Lys Gln Val Lys Ile Ala Glu Arg Gln Leu
305                 310                 315                 320

Arg Glu Ser Ile Arg Glu Lys Asn Ile Lys Arg Tyr Arg Phe Ser Ile
                325                 330                 335

Lys Thr Ile Glu Lys Asp Asp Gly Thr Tyr Phe Phe Ala Asn Lys Gln
            340                 345                 350

Ile Ser Val Phe Trp Ile His Arg Ile Glu Asn Ala Val Glu Arg Ile
        355                 360                 365

Leu Gly Ser Ile Asn Asp Lys Lys Leu Tyr Arg Leu Arg Leu Gly Tyr
    370                 375                 380

Leu Gly Glu Lys Val Trp Lys Asp Ile Leu Asn Phe Leu Ser Ile Lys
385                 390                 395                 400

Tyr Ile Ala Val Gly Lys Ala Val Phe Asn Phe Ala Met Asp Asp Leu
                405                 410                 415

Gln Glu Lys Asp Arg Asp Ile Glu Pro Gly Lys Ile Ser Glu Asn Ala
            420                 425                 430

Val Asn Gly Leu Thr Ser Phe Asp Tyr Glu Gln Ile Lys Ala Asp Glu
        435                 440                 445

Met Leu Gln Arg Glu Val Ala Val Asn Val Ala Phe Ala Ala Asn Asn
    450                 455                 460

Leu Ala Arg Val Thr Val Asp Ile Pro Gln Asn Gly Glu Lys Glu Asp
465                 470                 475                 480

Ile Leu Leu Trp Asn Lys Ser Asp Ile Lys Lys Tyr Lys Lys Asn Ser
                485                 490                 495

Lys Lys Gly Ile Leu Lys Ser Ile Leu Gln Phe Phe Gly Gly Ala Ser
            500                 505                 510

Thr Trp Asn Met Lys Met Phe Glu Ile Ala Tyr His Asp Gln Pro Gly
        515                 520                 525

Asp Tyr Glu Glu Asn Tyr Leu Tyr Asp Ile Ile Gln Ile Ile Tyr Ser
    530                 535                 540

Leu Arg Asn Lys Ser Phe His Phe Lys Thr Tyr Asp His Gly Asp Lys
545                 550                 555                 560

Asn Trp Asn Arg Glu Leu Ile Gly Lys Met Ile Glu His Asp Ala Glu
                565                 570                 575

Arg Val Ile Ser Val Glu Arg Glu Lys Phe His Ser Asn Asn Leu Pro
            580                 585                 590
```

```
Met Phe Tyr Lys Asp Ala Asp Leu Lys Lys Ile Leu Asp Leu Leu Tyr
        595                 600                 605

Ser Asp Tyr Ala Gly Arg Ala Ser Gln Val Pro Ala Phe Asn Thr Val
    610                 615                 620

Leu Val Arg Lys Asn Phe Pro Glu Phe Leu Arg Lys Asp Met Gly Tyr
625                 630                 635                 640

Lys Val His Phe Asn Asn Pro Glu Val Glu Asn Gln Trp His Ser Ala
                645                 650                 655

Val Tyr Tyr Leu Tyr Lys Glu Ile Tyr Tyr Asn Leu Phe Leu Arg Asp
            660                 665                 670

Lys Glu Val Lys Asn Leu Phe Tyr Thr Ser Leu Lys Asn Ile Arg Ser
        675                 680                 685

Glu Val Ser Asp Lys Lys Gln Lys Leu Ala Ser Asp Asp Phe Ala Ser
    690                 695                 700

Arg Cys Glu Glu Ile Glu Asp Arg Ser Leu Pro Glu Ile Cys Gln Ile
705                 710                 715                 720

Ile Met Thr Glu Tyr Asn Ala Gln Asn Phe Gly Asn Arg Lys Val Lys
                725                 730                 735

Ser Gln Arg Val Ile Glu Lys Asn Lys Asp Ile Phe Arg His Tyr Lys
            740                 745                 750

Met Leu Leu Ile Lys Thr Leu Ala Gly Ala Phe Ser Leu Tyr Leu Lys
        755                 760                 765

Gln Glu Arg Phe Ala Phe Ile Gly Lys Ala Thr Pro Ile Pro Tyr Glu
    770                 775                 780

Thr Thr Asp Val Lys Asn Phe Leu Pro Glu Trp Lys Ser Gly Met Tyr
785                 790                 795                 800

Ala Ser Phe Val Glu Glu Ile Lys Asn Asn Leu Asp Leu Gln Glu Trp
                805                 810                 815

Tyr Ile Val Gly Arg Phe Leu Asn Gly Arg Met Leu Asn Gln Leu Ala
            820                 825                 830

Gly Ser Leu Arg Ser Tyr Ile Gln Tyr Ala Glu Asp Ile Glu Arg Arg
        835                 840                 845

Ala Ala Glu Asn Arg Asn Lys Leu Phe Ser Lys Pro Asp Glu Lys Ile
    850                 855                 860

Glu Ala Cys Lys Lys Ala Val Arg Val Leu Asp Leu Cys Ile Lys Ile
865                 870                 875                 880

Ser Thr Arg Ile Ser Ala Glu Phe Thr Asp Tyr Phe Asp Ser Glu Asp
                885                 890                 895

Asp Tyr Ala Asp Tyr Leu Glu Lys Tyr Leu Lys Tyr Gln Asp Asp Ala
            900                 905                 910

Ile Lys Glu Leu Ser Gly Ser Ser Tyr Ala Ala Leu Asp His Phe Cys
        915                 920                 925
```

```
Asn Lys Asp Asp Leu Lys Phe Asp Ile Tyr Val Asn Ala Gly Gln Lys
    930                 935                 940

Pro Ile Leu Gln Arg Asn Ile Val Met Ala Lys Leu Phe Gly Pro Asp
945                 950                 955                 960

Asn Ile Leu Ser Glu Val Met Glu Lys Val Thr Glu Ser Ala Ile Arg
                965                 970                 975

Glu Tyr Tyr Asp Tyr Leu Lys Lys Val Ser Gly Tyr Arg Val Arg Gly
            980                 985                 990

Lys Cys Ser Thr Glu Lys Glu Gln Glu Asp Leu Leu Lys Phe Gln Arg
        995                 1000                1005

Leu Lys  Asn Ala Val Glu Phe  Arg Asp Val Thr Glu  Tyr Ala Glu
    1010                1015                1020

Val Ile  Asn Glu Leu Leu Gly  Gln Leu Ile Ser Trp  Ser Tyr Leu
    1025                1030                1035

Arg Glu  Arg Asp Leu Leu Tyr  Phe Gln Leu Gly Phe  His Tyr Met
    1040                1045                1050

Cys Leu  Lys Asn Lys Ser Phe  Lys Pro Ala Glu Tyr  Val Asp Ile
    1055                1060                1065

Arg Arg  Asn Asn Gly Thr Ile  Ile His Asn Ala Ile  Leu Tyr Gln
    1070                1075                1080

Ile Val  Ser Met Tyr Ile Asn  Gly Leu Asp Phe Tyr  Ser Cys Asp
    1085                1090                1095

Lys Glu  Gly Lys Thr Leu Lys  Pro Ile Glu Thr Gly  Lys Gly Val
    1100                1105                1110

Gly Ser  Lys Ile Gly Gln Phe  Ile Lys Tyr Ser Gln  Tyr Leu Tyr
    1115                1120                1125

Asn Asp  Pro Ser Tyr Lys Leu  Glu Ile Tyr Asn Ala  Gly Leu Glu
    1130                1135                1140

Val Phe  Glu Asn Ile Asp Glu  His Asp Asn Ile Thr  Asp Leu Arg
    1145                1150                1155

Lys Tyr  Val Asp His Phe Lys  Tyr Tyr Ala Tyr Gly  Asn Lys Met
    1160                1165                1170

Ser Leu  Leu Asp Leu Tyr Ser  Glu Phe Phe Asp Arg  Phe Phe Thr
    1175                1180                1185

Tyr Asp  Met Lys Tyr Gln Lys  Asn Val Val Asn Val  Leu Glu Asn
    1190                1195                1200

Ile Leu  Leu Arg His Phe Val  Ile Phe Tyr Pro Lys  Phe Gly Ser
    1205                1210                1215

Gly Lys  Lys Asp Val Gly Ile  Arg Asp Cys Lys Lys  Glu Arg Ala
    1220                1225                1230

Gln Ile  Glu Ile Ser Glu Gln  Ser Leu Thr Ser Glu  Asp Phe Met
    1235                1240                1245
```

Phe Lys Leu Asp Asp Lys Ala Gly Glu Glu Ala Lys Lys Phe Pro
1250 1255 1260

Ala Arg Asp Glu Arg Tyr Leu Gln Thr Ile Ala Lys Leu Leu Tyr
1265 1270 1275

Tyr Pro Asn Glu Ile Glu Asp Met Asn Arg Phe Met Lys Lys Gly
1280 1285 1290

Glu Thr Ile Asn Lys Lys Val Gln Phe Asn Arg Lys Lys Lys Ile
1295 1300 1305

Thr Arg Lys Gln Lys Asn Asn Ser Ser Asn Glu Val Leu Ser Ser
1310 1315 1320

Thr Met Gly Tyr Leu Phe Lys Asn Ile Lys Leu
1325 1330

<210> 39
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 39
guuuuagucc ucuuucauau agagguaguc ucuuac 36

<210> 40
<211> 99
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 40
augaaaagag gacuaaaacu gaaagaggac uaaaacacca gauguggaua acuauauuag 60

uggcuauuaa aaauucgucg auauuagaga ggaaacuuu 99

<210> 41
<211> 1120
<212> PRT
<213> 斯氏李斯特菌（Listeria seeligeri）

<400> 41
Met Trp Ile Ser Ile Lys Thr Leu Ile His His Leu Gly Val Leu Phe
1 5 10 15

Phe Cys Asp Tyr Met Tyr Asn Arg Arg Glu Lys Lys Ile Ile Glu Val
20 25 30

Lys Thr Met Arg Ile Thr Lys Val Glu Val Asp Arg Lys Lys Val Leu
35 40 45

Ile Ser Arg Asp Lys Asn Gly Gly Lys Leu Val Tyr Glu Asn Glu Met
50 55 60

Gln Asp Asn Thr Glu Gln Ile Met His His Lys Lys Ser Ser Phe Tyr
65 70 75 80

Lys Ser Val Val Asn Lys Thr Ile Cys Arg Pro Glu Gln Lys Gln Met
85 90 95

Lys Lys Leu Val His Gly Leu Leu Gln Glu Asn Ser Gln Glu Lys Ile

100 105 110

Lys Val Ser Asp Val Thr Lys Leu Asn Ile Ser Asn Phe Leu Asn His
115 120 125

Arg Phe Lys Lys Ser Leu Tyr Tyr Phe Pro Glu Asn Ser Pro Asp Lys
130 135 140

Ser Glu Glu Tyr Arg Ile Glu Ile Asn Leu Ser Gln Leu Leu Glu Asp
145 150 155 160

Ser Leu Lys Lys Gln Gln Gly Thr Phe Ile Cys Trp Glu Ser Phe Ser
165 170 175

Lys Asp Met Glu Leu Tyr Ile Asn Trp Ala Glu Asn Tyr Ile Ser Ser
180 185 190

Lys Thr Lys Leu Ile Lys Lys Ser Ile Arg Asn Asn Arg Ile Gln Ser
195 200 205

Thr Glu Ser Arg Ser Gly Gln Leu Met Asp Arg Tyr Met Lys Asp Ile
210 215 220

Leu Asn Lys Asn Lys Pro Phe Asp Ile Gln Ser Val Ser Glu Lys Tyr
225 230 235 240

Gln Leu Glu Lys Leu Thr Ser Ala Leu Lys Ala Thr Phe Lys Glu Ala
245 250 255

Lys Lys Asn Asp Lys Glu Ile Asn Tyr Lys Leu Lys Ser Thr Leu Gln
260 265 270

Asn His Glu Arg Gln Ile Ile Glu Glu Leu Lys Glu Asn Ser Glu Leu
275 280 285

Asn Gln Phe Asn Ile Glu Ile Arg Lys His Leu Glu Thr Tyr Phe Pro
290 295 300

Ile Lys Lys Thr Asn Arg Lys Val Gly Asp Ile Arg Asn Leu Glu Ile
305 310 315 320

Gly Glu Ile Gln Lys Ile Val Asn His Arg Leu Lys Asn Lys Ile Val
325 330 335

Gln Arg Ile Leu Gln Glu Gly Lys Leu Ala Ser Tyr Glu Ile Glu Ser
340 345 350

Thr Val Asn Ser Asn Ser Leu Gln Lys Ile Lys Ile Glu Glu Ala Phe
355 360 365

Ala Leu Lys Phe Ile Asn Ala Cys Leu Phe Ala Ser Asn Asn Leu Arg
370 375 380

Asn Met Val Tyr Pro Val Cys Lys Lys Asp Ile Leu Met Ile Gly Glu
385 390 395 400

Phe Lys Asn Ser Phe Lys Glu Ile Lys His Lys Lys Phe Ile Arg Gln
405 410 415

Trp Ser Gln Phe Phe Ser Gln Glu Ile Thr Val Asp Asp Ile Glu Leu
420 425 430

Ala Ser Trp Gly Leu Arg Gly Ala Ile Ala Pro Ile Arg Asn Glu Ile

435 440 445

Ile His Leu Lys Lys His Ser Trp Lys Lys Phe Phe Asn Asn Pro Thr
450 455 460

Phe Lys Val Lys Lys Ser Lys Ile Ile Asn Gly Lys Thr Lys Asp Val
465 470 475 480

Thr Ser Glu Phe Leu Tyr Lys Glu Thr Leu Phe Lys Asp Tyr Phe Tyr
485 490 495

Ser Glu Leu Asp Ser Val Pro Glu Leu Ile Ile Asn Lys Met Glu Ser
500 505 510

Ser Lys Ile Leu Asp Tyr Tyr Ser Ser Asp Gln Leu Asn Gln Val Phe
515 520 525

Thr Ile Pro Asn Phe Glu Leu Ser Leu Leu Thr Ser Ala Val Pro Phe
530 535 540

Ala Pro Ser Phe Lys Arg Val Tyr Leu Lys Gly Phe Asp Tyr Gln Asn
545 550 555 560

Gln Asp Glu Ala Gln Pro Asp Tyr Asn Leu Lys Leu Asn Ile Tyr Asn
565 570 575

Glu Lys Ala Phe Asn Ser Glu Ala Phe Gln Ala Gln Tyr Ser Leu Phe
580 585 590

Lys Met Val Tyr Tyr Gln Val Phe Leu Pro Gln Phe Thr Thr Asn Asn
595 600 605

Asp Leu Phe Lys Ser Ser Val Asp Phe Ile Leu Thr Leu Asn Lys Glu
610 615 620

Arg Lys Gly Tyr Ala Lys Ala Phe Gln Asp Ile Arg Lys Met Asn Lys
625 630 635 640

Asp Glu Lys Pro Ser Glu Tyr Met Ser Tyr Ile Gln Ser Gln Leu Met
645 650 655

Leu Tyr Gln Lys Lys Gln Glu Glu Lys Glu Lys Ile Asn His Phe Glu
660 665 670

Lys Phe Ile Asn Gln Val Phe Ile Lys Gly Phe Asn Ser Phe Ile Glu
675 680 685

Lys Asn Arg Leu Thr Tyr Ile Cys His Pro Thr Lys Asn Thr Val Pro
690 695 700

Glu Asn Asp Asn Ile Glu Ile Pro Phe His Thr Asp Met Asp Asp Ser
705 710 715 720

Asn Ile Ala Phe Trp Leu Met Cys Lys Leu Leu Asp Ala Lys Gln Leu
725 730 735

Ser Glu Leu Arg Asn Glu Met Ile Lys Phe Ser Cys Ser Leu Gln Ser
740 745 750

Thr Glu Glu Ile Ser Thr Phe Thr Lys Ala Arg Glu Val Ile Gly Leu
755 760 765

Ala Leu Leu Asn Gly Glu Lys Gly Cys Asn Asp Trp Lys Glu Leu Phe

770 775 780

Asp Asp Lys Glu Ala Trp Lys Lys Asn Met Ser Leu Tyr Val Ser Glu
785 790 795 800

Glu Leu Leu Gln Ser Leu Pro Tyr Thr Gln Glu Asp Gly Gln Thr Pro
805 810 815

Val Ile Asn Arg Ser Ile Asp Leu Val Lys Lys Tyr Gly Thr Glu Thr
820 825 830

Ile Leu Glu Lys Leu Phe Ser Ser Ser Asp Asp Tyr Lys Val Ser Ala
835 840 845

Lys Asp Ile Ala Lys Leu His Glu Tyr Asp Val Thr Glu Lys Ile Ala
850 855 860

Gln Gln Glu Ser Leu His Lys Gln Trp Ile Glu Lys Pro Gly Leu Ala
865 870 875 880

Arg Asp Ser Ala Trp Thr Lys Lys Tyr Gln Asn Val Ile Asn Asp Ile
885 890 895

Ser Asn Tyr Gln Trp Ala Lys Thr Lys Val Glu Leu Thr Gln Val Arg
900 905 910

His Leu His Gln Leu Thr Ile Asp Leu Leu Ser Arg Leu Ala Gly Tyr
915 920 925

Met Ser Ile Ala Asp Arg Asp Phe Gln Phe Ser Ser Asn Tyr Ile Leu
930 935 940

Glu Arg Glu Asn Ser Glu Tyr Arg Val Thr Ser Trp Ile Leu Leu Ser
945 950 955 960

Glu Asn Lys Asn Lys Asn Lys Tyr Asn Asp Tyr Glu Leu Tyr Asn Leu
965 970 975

Lys Asn Ala Ser Ile Lys Val Ser Ser Lys Asn Asp Pro Gln Leu Lys
980 985 990

Val Asp Leu Lys Gln Leu Arg Leu Thr Leu Glu Tyr Leu Glu Leu Phe
995 1000 1005

Asp Asn Arg Leu Lys Glu Lys Arg Asn Asn Ile Ser His Phe Asn
1010 1015 1020

Tyr Leu Asn Gly Gln Leu Gly Asn Ser Ile Leu Glu Leu Phe Asp
1025 1030 1035

Asp Ala Arg Asp Val Leu Ser Tyr Asp Arg Lys Leu Lys Asn Ala
1040 1045 1050

Val Ser Lys Ser Leu Lys Glu Ile Leu Ser Ser His Gly Met Glu
1055 1060 1065

Val Thr Phe Lys Pro Leu Tyr Gln Thr Asn His His Leu Lys Ile
1070 1075 1080

Asp Lys Leu Gln Pro Lys Lys Ile His His Leu Gly Glu Lys Ser
1085 1090 1095

Thr Val Ser Ser Asn Gln Val Ser Asn Glu Tyr Cys Gln Leu Val

1100 1105 1110

Arg Thr Leu Leu Thr Met Lys
1115 1120

<210> 42
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 42
guuuuagucc ccuucguuuu ugggguaguc uaaauc 36

<210> 43
<211> 113
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 43
gauuuagagc accccaaaag uaaugaaaau uugcaauuaa auaaggaaua uuaaaaaaau 60

gugauuuuaa aaaaauugaa gaaauuaaau gaaaaauugu ccaaguaaaa aaa 113

<210> 44
<211> 70
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 44
auuuagauua ccccuuuaau uuauuuuacc auauuuuucu cauaaugcaa acuaauauuc 60

caaaauuuuu 70

<210> 45
<211> 1389
<212> PRT
<213> 韋德纖毛菌（Leptotrichia wadei）

<400> 45
Met Gly Asn Leu Phe Gly His Lys Arg Trp Tyr Glu Val Arg Asp Lys
1 5 10 15

Lys Asp Phe Lys Ile Lys Arg Lys Val Lys Val Lys Arg Asn Tyr Asp
20 25 30

Gly Asn Lys Tyr Ile Leu Asn Ile Asn Glu Asn Asn Asn Lys Glu Lys
35 40 45

Ile Asp Asn Asn Lys Phe Ile Arg Lys Tyr Ile Asn Tyr Lys Lys Asn
50 55 60

Asp Asn Ile Leu Lys Glu Phe Thr Arg Lys Phe His Ala Gly Asn Ile
65 70 75 80

Leu Phe Lys Leu Lys Gly Lys Glu Gly Ile Ile Arg Ile Glu Asn Asn
85 90 95

Asp Asp Phe Leu Glu Thr Glu Glu Val Val Leu Tyr Ile Glu Ala Tyr
100 105 110

```
Gly Lys Ser Glu Lys Leu Lys Ala Leu Gly Ile Thr Lys Lys Lys Ile
        115                 120                 125

Ile Asp Glu Ala Ile Arg Gln Gly Ile Thr Lys Asp Asp Lys Lys Ile
    130                 135                 140

Glu Ile Lys Arg Gln Glu Asn Glu Glu Glu Ile Glu Ile Asp Ile Arg
145                 150                 155                 160

Asp Glu Tyr Thr Asn Lys Thr Leu Asn Asp Cys Ser Ile Ile Leu Arg
                165                 170                 175

Ile Ile Glu Asn Asp Glu Leu Glu Thr Lys Lys Ser Ile Tyr Glu Ile
            180                 185                 190

Phe Lys Asn Ile Asn Met Ser Leu Tyr Lys Ile Ile Glu Lys Ile Ile
        195                 200                 205

Glu Asn Glu Thr Glu Lys Val Phe Glu Asn Arg Tyr Tyr Glu Glu His
    210                 215                 220

Leu Arg Glu Lys Leu Leu Lys Asp Asp Lys Ile Asp Val Ile Leu Thr
225                 230                 235                 240

Asn Phe Met Glu Ile Arg Glu Lys Ile Lys Ser Asn Leu Glu Ile Leu
                245                 250                 255

Gly Phe Val Lys Phe Tyr Leu Asn Val Gly Gly Asp Lys Lys Lys Ser
            260                 265                 270

Lys Asn Lys Lys Met Leu Val Glu Lys Ile Leu Asn Ile Asn Val Asp
        275                 280                 285

Leu Thr Val Glu Asp Ile Ala Asp Phe Val Ile Lys Glu Leu Glu Phe
    290                 295                 300

Trp Asn Ile Thr Lys Arg Ile Glu Lys Val Lys Lys Val Asn Asn Glu
305                 310                 315                 320

Phe Leu Glu Lys Arg Arg Asn Arg Thr Tyr Ile Lys Ser Tyr Val Leu
                325                 330                 335

Leu Asp Lys His Glu Lys Phe Lys Ile Glu Arg Glu Asn Lys Lys Asp
            340                 345                 350

Lys Ile Val Lys Phe Phe Val Glu Asn Ile Lys Asn Asn Ser Ile Lys
        355                 360                 365

Glu Lys Ile Glu Lys Ile Leu Ala Glu Phe Lys Ile Asp Glu Leu Ile
    370                 375                 380

Lys Lys Leu Glu Lys Glu Leu Lys Lys Gly Asn Cys Asp Thr Glu Ile
385                 390                 395                 400

Phe Gly Ile Phe Lys Lys His Tyr Lys Val Asn Phe Asp Ser Lys Lys
                405                 410                 415

Phe Ser Lys Lys Ser Asp Glu Glu Lys Glu Leu Tyr Lys Ile Ile Tyr
            420                 425                 430

Arg Tyr Leu Lys Gly Arg Ile Glu Lys Ile Leu Val Asn Glu Gln Lys
        435                 440                 445
```

```
Val Arg Leu Lys Lys Met Glu Lys Ile Glu Ile Glu Lys Ile Leu Asn
    450                 455                 460

Glu Ser Ile Leu Ser Glu Lys Ile Leu Lys Arg Val Lys Gln Tyr Thr
465                 470                 475                 480

Leu Glu His Ile Met Tyr Leu Gly Lys Leu Arg His Asn Asp Ile Asp
                485                 490                 495

Met Thr Thr Val Asn Thr Asp Asp Phe Ser Arg Leu His Ala Lys Glu
            500                 505                 510

Glu Leu Asp Leu Glu Leu Ile Thr Phe Phe Ala Ser Thr Asn Met Glu
        515                 520                 525

Leu Asn Lys Ile Phe Ser Arg Glu Asn Ile Asn Asn Asp Glu Asn Ile
    530                 535                 540

Asp Phe Phe Gly Gly Asp Arg Glu Lys Asn Tyr Val Leu Asp Lys Lys
545                 550                 555                 560

Ile Leu Asn Ser Lys Ile Lys Ile Ile Arg Asp Leu Asp Phe Ile Asp
                565                 570                 575

Asn Lys Asn Asn Ile Thr Asn Asn Phe Ile Arg Lys Phe Thr Lys Ile
            580                 585                 590

Gly Thr Asn Glu Arg Asn Arg Ile Leu His Ala Ile Ser Lys Glu Arg
        595                 600                 605

Asp Leu Gln Gly Thr Gln Asp Asp Tyr Asn Lys Val Ile Asn Ile Ile
    610                 615                 620

Gln Asn Leu Lys Ile Ser Asp Glu Glu Val Ser Lys Ala Leu Asn Leu
625                 630                 635                 640

Asp Val Val Phe Lys Asp Lys Lys Asn Ile Ile Thr Lys Ile Asn Asp
                645                 650                 655

Ile Lys Ile Ser Glu Glu Asn Asn Asn Asp Ile Lys Tyr Leu Pro Ser
            660                 665                 670

Phe Ser Lys Val Leu Pro Glu Ile Leu Asn Leu Tyr Arg Asn Asn Pro
        675                 680                 685

Lys Asn Glu Pro Phe Asp Thr Ile Glu Thr Glu Lys Ile Val Leu Asn
    690                 695                 700

Ala Leu Ile Tyr Val Asn Lys Glu Leu Tyr Lys Lys Leu Ile Leu Glu
705                 710                 715                 720

Asp Asp Leu Glu Glu Asn Glu Ser Lys Asn Ile Phe Leu Gln Glu Leu
                725                 730                 735

Lys Lys Thr Leu Gly Asn Ile Asp Glu Ile Asp Glu Asn Ile Ile Glu
            740                 745                 750

Asn Tyr Tyr Lys Asn Ala Gln Ile Ser Ala Ser Lys Gly Asn Asn Lys
        755                 760                 765

Ala Ile Lys Lys Tyr Gln Lys Lys Val Ile Glu Cys Tyr Ile Gly Tyr
    770                 775                 780
```

```
Leu Arg Lys Asn Tyr Glu Glu Leu Phe Asp Phe Ser Asp Phe Lys Met
785                 790                 795                 800

Asn Ile Gln Glu Ile Lys Lys Gln Ile Lys Asp Ile Asn Asp Asn Lys
                805                 810                 815

Thr Tyr Glu Arg Ile Thr Val Lys Thr Ser Asp Lys Thr Ile Val Ile
            820                 825                 830

Asn Asp Asp Phe Glu Tyr Ile Ile Ser Ile Phe Ala Leu Leu Asn Ser
        835                 840                 845

Asn Ala Val Ile Asn Lys Ile Arg Asn Arg Phe Phe Ala Thr Ser Val
    850                 855                 860

Trp Leu Asn Thr Ser Glu Tyr Gln Asn Ile Ile Asp Ile Leu Asp Glu
865                 870                 875                 880

Ile Met Gln Leu Asn Thr Leu Arg Asn Glu Cys Ile Thr Glu Asn Trp
                885                 890                 895

Asn Leu Asn Leu Glu Glu Phe Ile Gln Lys Met Lys Glu Ile Glu Lys
            900                 905                 910

Asp Phe Asp Asp Phe Lys Ile Gln Thr Lys Lys Glu Ile Phe Asn Asn
        915                 920                 925

Tyr Tyr Glu Asp Ile Lys Asn Asn Ile Leu Thr Glu Phe Lys Asp Asp
    930                 935                 940

Ile Asn Gly Cys Asp Val Leu Glu Lys Lys Leu Glu Lys Ile Val Ile
945                 950                 955                 960

Phe Asp Asp Glu Thr Lys Phe Glu Ile Asp Lys Lys Ser Asn Ile Leu
                965                 970                 975

Gln Asp Glu Gln Arg Lys Leu Ser Asn Ile Asn Lys Lys Asp Leu Lys
            980                 985                 990

Lys Lys Val Asp Gln Tyr Ile Lys Asp Lys Asp Gln Glu Ile Lys Ser
        995                 1000                1005

Lys Ile  Leu Cys Arg Ile Ile Phe Asn Ser Asp  Phe Leu Lys Lys
    1010                1015                1020

Tyr Lys  Lys Glu Ile Asp Asn Leu Ile Glu Asp  Met Glu Ser Glu
    1025                1030                1035

Asn Glu  Asn Lys Phe Gln Glu Ile Tyr Tyr Pro  Lys Glu Arg Lys
    1040                1045                1050

Asn Glu  Leu Tyr Ile Tyr Lys Lys Asn Leu Phe  Leu Asn Ile Gly
    1055                1060                1065

Asn Pro  Asn Phe Asp Lys Ile Tyr Gly Leu Ile  Ser Asn Asp Ile
    1070                1075                1080

Lys Met  Ala Asp Ala Lys Phe Leu Phe Asn Ile  Asp Gly Lys Asn
    1085                1090                1095

Ile Arg  Lys Asn Lys Ile Ser Glu Ile Asp Ala  Ile Leu Lys Asn
    1100                1105                1110
```

Leu Asn Asp Lys Leu Asn Gly Tyr Ser Lys Glu Tyr Lys Glu Lys
1115 1120 1125

Tyr Ile Lys Lys Leu Lys Glu Asn Asp Asp Phe Phe Ala Lys Asn
1130 1135 1140

Ile Gln Asn Lys Asn Tyr Lys Ser Phe Glu Lys Asp Tyr Asn Arg
1145 1150 1155

Val Ser Glu Tyr Lys Lys Ile Arg Asp Leu Val Glu Phe Asn Tyr
1160 1165 1170

Leu Asn Lys Ile Glu Ser Tyr Leu Ile Asp Ile Asn Trp Lys Leu
1175 1180 1185

Ala Ile Gln Met Ala Arg Phe Glu Arg Asp Met His Tyr Ile Val
1190 1195 1200

Asn Gly Leu Arg Glu Leu Gly Ile Ile Lys Leu Ser Gly Tyr Asn
1205 1210 1215

Thr Gly Ile Ser Arg Ala Tyr Pro Lys Arg Asn Gly Ser Asp Gly
1220 1225 1230

Phe Tyr Thr Thr Thr Ala Tyr Tyr Lys Phe Phe Asp Glu Glu Ser
1235 1240 1245

Tyr Lys Lys Phe Glu Lys Ile Cys Tyr Gly Phe Gly Ile Asp Leu
1250 1255 1260

Ser Glu Asn Ser Glu Ile Asn Lys Pro Glu Asn Glu Ser Ile Arg
1265 1270 1275

Asn Tyr Ile Ser His Phe Tyr Ile Val Arg Asn Pro Phe Ala Asp
1280 1285 1290

Tyr Ser Ile Ala Glu Gln Ile Asp Arg Val Ser Asn Leu Leu Ser
1295 1300 1305

Tyr Ser Thr Arg Tyr Asn Asn Ser Thr Tyr Ala Ser Val Phe Glu
1310 1315 1320

Val Phe Lys Lys Asp Val Asn Leu Asp Tyr Asp Glu Leu Lys Lys
1325 1330 1335

Lys Phe Lys Leu Ile Gly Asn Asn Asp Ile Leu Glu Arg Leu Met
1340 1345 1350

Lys Pro Lys Lys Val Ser Val Leu Glu Leu Glu Ser Tyr Asn Ser
1355 1360 1365

Asp Tyr Ile Lys Asn Leu Ile Ile Glu Leu Leu Thr Lys Ile Glu
1370 1375 1380

Asn Thr Asn Asp Thr Leu
1385

<210> 46
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 46
guuuuagucc ccuucgauau ugggggugguc uauauc 36

<210> 47
<211> 95
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 47
auugaugugg uauacuaaaa auggaaaauu guauuuuuga uuagaaagau guaaaauuga 60

uuuaauuuaa aaauauuuua uuagauuaaa guaga 95

<210> 48
<211> 1300
<212> PRT
<213> 沙氏纖毛菌（Leptotrichia shahii）

<400> 48
Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
20 25 30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
35 40 45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
50 55 60

Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65 70 75 80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
85 90 95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
100 105 110

Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
115 120 125

Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
130 135 140

Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145 150 155 160

Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
165 170 175

Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190

Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205

Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
210 215 220

Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240

```
Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
                245                 250                 255

Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
            260                 265                 270

Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
        275                 280                 285

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
    290                 295                 300

Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305                 310                 315                 320

Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
                325                 330                 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
            340                 345                 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
        355                 360                 365

Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
    370                 375                 380

Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385                 390                 395                 400

Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
                405                 410                 415

Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
            420                 425                 430

Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
        435                 440                 445

Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
    450                 455                 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465                 470                 475                 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
                485                 490                 495

Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
            500                 505                 510

Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
        515                 520                 525

Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
    530                 535                 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545                 550                 555                 560

Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
                565                 570                 575
```

```
Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
            580                 585                 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
        595                 600                 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
    610                 615                 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625                 630                 635                 640

Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
                645                 650                 655

Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
            660                 665                 670

Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
        675                 680                 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln
    690                 695                 700

Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
705                 710                 715                 720

Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp
                725                 730                 735

Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
            740                 745                 750

Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
        755                 760                 765

Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
    770                 775                 780

Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
785                 790                 795                 800

Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
                805                 810                 815

Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
            820                 825                 830

Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
        835                 840                 845

Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu
    850                 855                 860

Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
865                 870                 875                 880

His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
                885                 890                 895

Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
            900                 905                 910
```

```
Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
        915                 920                 925

Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
    930                 935                 940

Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945                 950                 955                 960

Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
                965                 970                 975

Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
            980                 985                 990

Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu
        995                 1000                1005

Asn Phe  Gly Phe Lys Arg Gly  Arg Phe Lys Val Glu  Lys Gln Val
    1010                1015                1020

Tyr Gln  Lys Leu Glu Lys Met  Leu Ile Glu Lys Leu  Asn Tyr Leu
    1025                1030                1035

Val Phe  Lys Asp Asn Glu Phe  Asp Lys Thr Gly Gly  Val Leu Arg
    1040                1045                1050

Ala Tyr  Gln Leu Thr Ala Pro  Phe Glu Thr Phe Lys  Lys Met Gly
    1055                1060                1065

Lys Gln  Thr Gly Ile Ile Tyr  Tyr Val Pro Ala Gly  Phe Thr Ser
    1070                1075                1080

Lys Ile  Cys Pro Val Thr Gly  Phe Val Asn Gln Leu  Tyr Pro Lys
    1085                1090                1095

Tyr Glu  Ser Val Ser Lys Ser  Gln Glu Phe Phe Ser  Lys Phe Asp
    1100                1105                1110

Lys Ile  Cys Tyr Asn Leu Asp  Lys Gly Tyr Phe Glu  Phe Ser Phe
    1115                1120                1125

Asp Tyr  Lys Asn Phe Gly Asp  Lys Ala Ala Lys Gly  Lys Trp Thr
    1130                1135                1140

Ile Ala  Ser Phe Gly Ser Arg  Leu Ile Asn Phe Arg  Asn Ser Asp
    1145                1150                1155

Lys Asn  His Asn Trp Asp Thr  Arg Glu Val Tyr Pro  Thr Lys Glu
    1160                1165                1170

Leu Glu  Lys Leu Leu Lys Asp  Tyr Ser Ile Glu Tyr  Gly His Gly
    1175                1180                1185

Glu Cys  Ile Lys Ala Ala Ile  Cys Gly Glu Ser Asp  Lys Lys Phe
    1190                1195                1200

Phe Ala  Lys Leu Thr Ser Val  Leu Asn Thr Ile Leu  Gln Met Arg
    1205                1210                1215

Asn Ser  Lys Thr Gly Thr Glu  Leu Asp Tyr Leu Ile  Ser Pro Val
    1220                1225                1230
```

Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1235 1240 1245

Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly
1250 1255 1260

Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu
1265 1270 1275

Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1280 1285 1290

Phe Val Gln Asn Arg Asn Asn
1295 1300

<210> 49
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 49
gucuaagaac uuuaaauaau uucuacuguu guagau 36

<210> 50
<211> 71
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 50
aucuacaaaa uuauaaacua aauaaagauu cuuauaauaa cuuuauauau aaucgaaaug 60

uagagaauuu u 71

<210> 51
<211> 1300
<212> PRT
<213> 土拉熱弗朗西絲菌（Francisella ularensis）

<400> 51
Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
20 25 30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
35 40 45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
50 55 60

Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65 70 75 80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
85 90 95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
100 105 110

```
Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
        115                 120                 125

Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
    130                 135                 140

Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145                 150                 155                 160

Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
                165                 170                 175

Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
            180                 185                 190

Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
        195                 200                 205

Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
    210                 215                 220

Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225                 230                 235                 240

Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
                245                 250                 255

Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
            260                 265                 270

Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
        275                 280                 285

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
    290                 295                 300

Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305                 310                 315                 320

Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
                325                 330                 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
            340                 345                 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
        355                 360                 365

Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
    370                 375                 380

Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385                 390                 395                 400

Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
                405                 410                 415

Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
            420                 425                 430

Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
        435                 440                 445
```

```
Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
    450                 455                 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465                 470                 475                 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
                485                 490                 495

Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
            500                 505                 510

Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
        515                 520                 525

Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
    530                 535                 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545                 550                 555                 560

Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
                565                 570                 575

Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
            580                 585                 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
        595                 600                 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
    610                 615                 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625                 630                 635                 640

Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
                645                 650                 655

Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
            660                 665                 670

Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
        675                 680                 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln
    690                 695                 700

Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
705                 710                 715                 720

Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp
                725                 730                 735

Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
            740                 745                 750

Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
        755                 760                 765

Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
    770                 775                 780
```

```
Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
785                 790                 795                 800

Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
                805                 810                 815

Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
            820                 825                 830

Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
        835                 840                 845

Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu
    850                 855                 860

Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
865                 870                 875                 880

His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
                885                 890                 895

Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
            900                 905                 910

Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
        915                 920                 925

Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
    930                 935                 940

Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945                 950                 955                 960

Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
                965                 970                 975

Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
            980                 985                 990

Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu
        995                 1000                1005

Asn Phe  Gly Phe Lys Arg Gly Arg Phe Lys Val  Glu Lys Gln Val
    1010                1015                1020

Tyr Gln  Lys Leu Glu Lys Met Leu Ile Glu Lys  Leu Asn Tyr Leu
    1025                1030                1035

Val Phe  Lys Asp Asn Glu Phe Asp Lys Thr Gly  Gly Val Leu Arg
    1040                1045                1050

Ala Tyr  Gln Leu Thr Ala Pro Phe Glu Thr Phe  Lys Lys Met Gly
    1055                1060                1065

Lys Gln  Thr Gly Ile Ile Tyr Tyr Val Pro Ala  Gly Phe Thr Ser
    1070                1075                1080

Lys Ile  Cys Pro Val Thr Gly Phe Val Asn Gln  Leu Tyr Pro Lys
    1085                1090                1095

Tyr Glu  Ser Val Ser Lys Ser Gln Glu Phe Phe  Ser Lys Phe Asp
    1100                1105                1110
```

Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe
1115 1120 1125

Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr
1130 1135 1140

Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp
1145 1150 1155

Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu
1160 1165 1170

Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly
1175 1180 1185

Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe
1190 1195 1200

Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg
1205 1210 1215

Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val
1220 1225 1230

Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1235 1240 1245

Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly
1250 1255 1260

Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu
1265 1270 1275

Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1280 1285 1290

Phe Val Gln Asn Arg Asn Asn
1295 1300

<210> 52
<211> 138
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 52
aaaaggccgg cggccacgaa aaaggccggc caggcaaaaa agaaaaaggg atcctaccca 60

tacgatgttc cagattacgc ttatccctac gacgtgcctg attatgcata cccatatgat 120

gtccccgact atgcctaa 138

<210> 53
<211> 1388
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 53
Met Ser Leu Asn Arg Ile Tyr Gln Gly Arg Val Ala Ala Val Glu Thr
1 5 10 15

Gly Thr Ala Leu Ala Lys Gly Asn Val Glu Trp Met Pro Ala Ala Gly

```
            20                  25                  30

Gly Asp Glu Val Leu Trp Gln His His Glu Leu Phe Gln Ala Ala Ile
        35                  40                  45

Asn Tyr Tyr Leu Val Ala Leu Leu Ala Leu Ala Asp Lys Asn Asn Pro
    50                  55                  60

Val Leu Gly Pro Leu Ile Ser Gln Met Asp Asn Pro Gln Ser Pro Tyr
65                  70                  75                  80

His Val Trp Gly Ser Phe Arg Arg Gln Gly Arg Gln Arg Thr Gly Leu
                85                  90                  95

Ser Gln Ala Val Ala Pro Tyr Ile Thr Pro Gly Asn Asn Ala Pro Thr
            100                 105                 110

Leu Asp Glu Val Phe Arg Ser Ile Leu Ala Gly Asn Pro Thr Asp Arg
        115                 120                 125

Ala Thr Leu Asp Ala Ala Leu Met Gln Leu Leu Lys Ala Cys Asp Gly
    130                 135                 140

Ala Gly Ala Ile Gln Gln Glu Gly Arg Ser Tyr Trp Pro Lys Phe Cys
145                 150                 155                 160

Asp Pro Asp Ser Thr Ala Asn Phe Ala Gly Asp Pro Ala Met Leu Arg
                165                 170                 175

Arg Glu Gln His Arg Leu Leu Leu Pro Gln Val Leu His Asp Pro Ala
            180                 185                 190

Ile Thr His Asp Ser Pro Ala Leu Gly Ser Phe Asp Thr Tyr Ser Ile
        195                 200                 205

Ala Thr Pro Asp Thr Arg Thr Pro Gln Leu Thr Gly Pro Lys Ala Arg
    210                 215                 220

Ala Arg Leu Glu Gln Ala Ile Thr Leu Trp Arg Val Arg Leu Pro Glu
225                 230                 235                 240

Ser Ala Ala Asp Phe Asp Arg Leu Ala Ser Ser Leu Lys Lys Ile Pro
                245                 250                 255

Asp Asp Asp Ser Arg Leu Asn Leu Gln Gly Tyr Val Gly Ser Ser Ala
            260                 265                 270

Lys Gly Glu Val Gln Ala Arg Leu Phe Ala Leu Leu Leu Phe Arg His
        275                 280                 285

Leu Glu Arg Ser Ser Phe Thr Leu Gly Leu Leu Arg Ser Ala Thr Pro
    290                 295                 300

Pro Pro Lys Asn Ala Glu Thr Pro Pro Pro Ala Gly Val Pro Leu Pro
305                 310                 315                 320

Ala Ala Ser Ala Ala Asp Pro Val Arg Ile Ala Arg Gly Lys Arg Ser
                325                 330                 335

Phe Val Phe Arg Ala Phe Thr Ser Leu Pro Cys Trp His Gly Gly Asp
            340                 345                 350

Asn Ile His Pro Thr Trp Lys Ser Phe Asp Ile Ala Ala Phe Lys Tyr
```

355 360 365

Ala Leu Thr Val Ile Asn Gln Ile Glu Glu Lys Thr Lys Glu Arg Gln
370 375 380

Lys Glu Cys Ala Glu Leu Glu Thr Asp Phe Asp Tyr Met His Gly Arg
385 390 395 400

Leu Ala Lys Ile Pro Val Lys Tyr Thr Thr Gly Glu Ala Glu Pro Pro
405 410 415

Pro Ile Leu Ala Asn Asp Leu Arg Ile Pro Leu Leu Arg Glu Leu Leu
420 425 430

Gln Asn Ile Lys Val Asp Thr Ala Leu Thr Asp Gly Glu Ala Val Ser
435 440 445

Tyr Gly Leu Gln Arg Arg Thr Ile Arg Gly Phe Arg Glu Leu Arg Arg
450 455 460

Ile Trp Arg Gly His Ala Pro Ala Gly Thr Val Phe Ser Ser Glu Leu
465 470 475 480

Lys Glu Lys Leu Ala Gly Glu Leu Arg Gln Phe Gln Thr Asp Asn Ser
485 490 495

Thr Thr Ile Gly Ser Val Gln Leu Phe Asn Glu Leu Ile Gln Asn Pro
500 505 510

Lys Tyr Trp Pro Ile Trp Gln Ala Pro Asp Val Glu Thr Ala Arg Gln
515 520 525

Trp Ala Asp Ala Gly Phe Ala Asp Asp Pro Leu Ala Ala Leu Val Gln
530 535 540

Glu Ala Glu Leu Gln Glu Asp Ile Asp Ala Leu Lys Ala Pro Val Lys
545 550 555 560

Leu Thr Pro Ala Asp Pro Glu Tyr Ser Arg Arg Gln Tyr Asp Phe Asn
565 570 575

Ala Val Ser Lys Phe Gly Ala Gly Ser Arg Ser Ala Asn Arg His Glu
580 585 590

Pro Gly Gln Thr Glu Arg Gly His Asn Thr Phe Thr Thr Glu Ile Ala
595 600 605

Ala Arg Asn Ala Ala Asp Gly Asn Arg Trp Arg Ala Thr His Val Arg
610 615 620

Ile His Tyr Ser Ala Pro Arg Leu Leu Arg Asp Gly Leu Arg Arg Pro
625 630 635 640

Asp Thr Asp Gly Asn Glu Ala Leu Glu Ala Val Pro Trp Leu Gln Pro
645 650 655

Met Met Glu Ala Leu Ala Pro Leu Pro Thr Leu Pro Gln Asp Leu Thr
660 665 670

Gly Met Pro Val Phe Leu Met Pro Asp Val Thr Leu Ser Gly Glu Arg
675 680 685

Arg Ile Leu Leu Asn Leu Pro Val Thr Leu Glu Pro Ala Ala Leu Val

690 695 700

Glu Gln Leu Gly Asn Ala Gly Arg Trp Gln Asn Gln Phe Phe Gly Ser
705 710 715 720

Arg Glu Asp Pro Phe Ala Leu Arg Trp Pro Ala Asp Gly Ala Val Lys
725 730 735

Thr Ala Lys Gly Lys Thr His Ile Pro Trp His Gln Asp Arg Asp His
740 745 750

Phe Thr Val Leu Gly Val Asp Leu Gly Thr Arg Asp Ala Gly Ala Leu
755 760 765

Ala Leu Leu Asn Val Thr Ala Gln Lys Pro Ala Lys Pro Val His Arg
770 775 780

Ile Ile Gly Glu Ala Asp Gly Arg Thr Trp Tyr Ala Ser Leu Ala Asp
785 790 795 800

Ala Arg Met Ile Arg Leu Pro Gly Glu Asp Ala Arg Leu Phe Val Arg
805 810 815

Gly Lys Leu Val Gln Glu Pro Tyr Gly Glu Arg Gly Arg Asn Ala Ser
820 825 830

Leu Leu Glu Trp Glu Asp Ala Arg Asn Ile Ile Leu Arg Leu Gly Gln
835 840 845

Asn Pro Asp Glu Leu Leu Gly Ala Asp Pro Arg Arg His Ser Tyr Pro
850 855 860

Glu Ile Asn Asp Lys Leu Leu Val Ala Leu Arg Arg Ala Gln Ala Arg
865 870 875 880

Leu Ala Arg Leu Gln Asn Arg Ser Trp Arg Leu Arg Asp Leu Ala Glu
885 890 895

Ser Asp Lys Ala Leu Asp Glu Ile His Ala Glu Arg Ala Gly Glu Lys
900 905 910

Pro Ser Pro Leu Pro Pro Leu Ala Arg Asp Asp Ala Ile Lys Ser Thr
915 920 925

Asp Glu Ala Leu Leu Ser Gln Arg Asp Ile Ile Arg Arg Ser Phe Val
930 935 940

Gln Ile Ala Asn Leu Ile Leu Pro Leu Arg Gly Arg Arg Trp Glu Trp
945 950 955 960

Arg Pro His Val Glu Val Pro Asp Cys His Ile Leu Ala Gln Ser Asp
965 970 975

Pro Gly Thr Asp Asp Thr Lys Arg Leu Val Ala Gly Gln Arg Gly Ile
980 985 990

Ser His Glu Arg Ile Glu Gln Ile Glu Glu Leu Arg Arg Arg Cys Gln
995 1000 1005

Ser Leu Asn Arg Ala Leu Arg His Lys Pro Gly Glu Arg Pro Val
1010 1015 1020

Leu Gly Arg Pro Ala Lys Gly Glu Glu Ile Ala Asp Pro Cys Pro

1025 1030 1035

Ala Leu Leu Glu Lys Ile Asn Arg Leu Arg Asp Gln Arg Val Asp
1040 1045 1050

Gln Thr Ala His Ala Ile Leu Ala Ala Ala Leu Gly Val Arg Leu
1055 1060 1065

Arg Ala Pro Ser Lys Asp Arg Ala Glu Arg Arg His Arg Asp Ile
1070 1075 1080

His Gly Glu Tyr Glu Arg Phe Arg Ala Pro Ala Asp Phe Val Val
1085 1090 1095

Ile Glu Asn Leu Ser Arg Tyr Leu Ser Ser Gln Asp Arg Ala Arg
1100 1105 1110

Ser Glu Asn Thr Arg Leu Met Gln Trp Cys His Arg Gln Ile Val
1115 1120 1125

Gln Lys Leu Arg Gln Leu Cys Glu Thr Tyr Gly Ile Pro Val Leu
1130 1135 1140

Ala Val Pro Ala Ala Tyr Ser Ser Arg Phe Ser Ser Arg Asp Gly
1145 1150 1155

Ser Ala Gly Phe Arg Ala Val His Leu Thr Pro Asp His Arg His
1160 1165 1170

Arg Met Pro Trp Ser Arg Ile Leu Ala Arg Leu Lys Ala His Glu
1175 1180 1185

Glu Asp Gly Lys Arg Leu Glu Lys Thr Val Leu Asp Glu Ala Arg
1190 1195 1200

Ala Val Arg Gly Leu Phe Asp Arg Leu Asp Arg Phe Asn Ala Gly
1205 1210 1215

His Val Pro Gly Lys Pro Trp Arg Thr Leu Leu Ala Pro Leu Pro
1220 1225 1230

Gly Gly Pro Val Phe Val Pro Leu Gly Asp Ala Thr Pro Met Gln
1235 1240 1245

Ala Asp Leu Asn Ala Ala Ile Asn Ile Ala Leu Arg Gly Ile Ala
1250 1255 1260

Ala Pro Asp Arg His Asp Ile His His Arg Leu Arg Ala Glu Asn
1265 1270 1275

Lys Lys Arg Ile Leu Ser Leu Arg Leu Gly Thr Gln Arg Glu Lys
1280 1285 1290

Ala Arg Trp Pro Gly Gly Ala Pro Ala Val Thr Leu Ser Thr Pro
1295 1300 1305

Asn Asn Gly Ala Ser Pro Glu Asp Ser Asp Ala Leu Pro Glu Arg
1310 1315 1320

Val Ser Asn Leu Phe Val Asp Ile Ala Gly Val Ala Asn Phe Glu
1325 1330 1335

Arg Val Thr Ile Glu Gly Val Ser Gln Lys Phe Ala Thr Gly Arg

1340 1345 1350

Gly Leu Trp Ala Ser Val Lys Gln Arg Ala Trp Asn Arg Val Ala
1355 1360 1365

Arg Leu Asn Glu Thr Val Thr Asp Asn Asn Arg Asn Glu Glu Glu
1370 1375 1380

Asp Asp Ile Pro Met
1385

<210> 54
<211> 1108
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 54
Met Ala Thr Arg Ser Phe Ile Leu Lys Ile Glu Pro Asn Glu Glu Val
1 5 10 15

Lys Lys Gly Leu Trp Lys Thr His Glu Val Leu Asn His Gly Ile Ala
20 25 30

Tyr Tyr Met Asn Ile Leu Lys Leu Ile Arg Gln Glu Ala Ile Tyr Glu
35 40 45

His His Glu Gln Asp Pro Lys Asn Pro Lys Lys Val Ser Lys Ala Glu
50 55 60

Ile Gln Ala Glu Leu Trp Asp Phe Val Leu Lys Met Gln Lys Cys Asn
65 70 75 80

Ser Phe Thr His Glu Val Asp Lys Asp Val Val Phe Asn Ile Leu Arg
85 90 95

Glu Leu Tyr Glu Glu Leu Val Pro Ser Ser Val Glu Lys Lys Gly Glu
100 105 110

Ala Asn Gln Leu Ser Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn
115 120 125

Ser Gln Ser Gly Lys Gly Thr Ala Ser Ser Gly Arg Lys Pro Arg Trp
130 135 140

Tyr Asn Leu Lys Ile Ala Gly Asp Pro Ser Trp Glu Glu Glu Lys Lys
145 150 155 160

Lys Trp Glu Glu Asp Lys Lys Lys Asp Pro Leu Ala Lys Ile Leu Gly
165 170 175

Lys Leu Ala Glu Tyr Gly Leu Ile Pro Leu Phe Ile Pro Phe Thr Asp
180 185 190

Ser Asn Glu Pro Ile Val Lys Glu Ile Lys Trp Met Glu Lys Ser Arg
195 200 205

Asn Gln Ser Val Arg Arg Leu Asp Lys Asp Met Phe Ile Gln Ala Leu
210 215 220

Glu Arg Phe Leu Ser Trp Glu Ser Trp Asn Leu Lys Val Lys Glu Glu
225 230 235 240

```
Tyr Glu Lys Val Glu Lys Glu His Lys Thr Leu Glu Glu Arg Ile Lys
                245                 250                 255

Glu Asp Ile Gln Ala Phe Lys Ser Leu Glu Gln Tyr Glu Lys Glu Arg
            260                 265                 270

Gln Glu Gln Leu Leu Arg Asp Thr Leu Asn Thr Asn Glu Tyr Arg Leu
        275                 280                 285

Ser Lys Arg Gly Leu Arg Gly Trp Arg Glu Ile Ile Gln Lys Trp Leu
    290                 295                 300

Lys Met Asp Glu Asn Glu Pro Ser Glu Lys Tyr Leu Glu Val Phe Lys
305                 310                 315                 320

Asp Tyr Gln Arg Lys His Pro Arg Glu Ala Gly Asp Tyr Ser Val Tyr
                325                 330                 335

Glu Phe Leu Ser Lys Lys Glu Asn His Phe Ile Trp Arg Asn His Pro
            340                 345                 350

Glu Tyr Pro Tyr Leu Tyr Ala Thr Phe Cys Glu Ile Asp Lys Lys Lys
        355                 360                 365

Lys Asp Ala Lys Gln Gln Ala Thr Phe Thr Leu Ala Asp Pro Ile Asn
    370                 375                 380

His Pro Leu Trp Val Arg Phe Glu Glu Arg Ser Gly Ser Asn Leu Asn
385                 390                 395                 400

Lys Tyr Arg Ile Leu Thr Glu Gln Leu His Thr Glu Lys Leu Lys Lys
                405                 410                 415

Lys Leu Thr Val Gln Leu Asp Arg Leu Ile Tyr Pro Thr Glu Ser Gly
            420                 425                 430

Gly Trp Glu Glu Lys Gly Lys Val Asp Ile Val Leu Leu Pro Ser Arg
        435                 440                 445

Gln Phe Tyr Asn Gln Ile Phe Leu Asp Ile Glu Glu Lys Gly Lys His
    450                 455                 460

Ala Phe Thr Tyr Lys Asp Glu Ser Ile Lys Phe Pro Leu Lys Gly Thr
465                 470                 475                 480

Leu Gly Gly Ala Arg Val Gln Phe Asp Arg Asp His Leu Arg Arg Tyr
                485                 490                 495

Pro His Lys Val Glu Ser Gly Asn Val Gly Arg Ile Tyr Phe Asn Met
            500                 505                 510

Thr Val Asn Ile Glu Pro Thr Glu Ser Pro Val Ser Lys Ser Leu Lys
        515                 520                 525

Ile His Arg Asp Asp Phe Pro Lys Phe Val Asn Phe Lys Pro Lys Glu
    530                 535                 540

Leu Thr Glu Trp Ile Lys Asp Ser Lys Gly Lys Lys Leu Lys Ser Gly
545                 550                 555                 560

Ile Glu Ser Leu Glu Ile Gly Leu Arg Val Met Ser Ile Asp Leu Gly
                565                 570                 575
```

```
Gln Arg Gln Ala Ala Ala Ala Ser Ile Phe Glu Val Val Asp Gln Lys
            580                 585                 590

Pro Asp Ile Glu Gly Lys Leu Phe Phe Pro Ile Lys Gly Thr Glu Leu
        595                 600                 605

Tyr Ala Val His Arg Ala Ser Phe Asn Ile Lys Leu Pro Gly Glu Thr
    610                 615                 620

Leu Val Lys Ser Arg Glu Val Leu Arg Lys Ala Arg Glu Asp Asn Leu
625                 630                 635                 640

Lys Leu Met Asn Gln Lys Leu Asn Phe Leu Arg Asn Val Leu His Phe
                645                 650                 655

Gln Gln Phe Glu Asp Ile Thr Glu Arg Glu Lys Arg Val Thr Lys Trp
            660                 665                 670

Ile Ser Arg Gln Glu Asn Ser Asp Val Pro Leu Val Tyr Gln Asp Glu
        675                 680                 685

Leu Ile Gln Ile Arg Glu Leu Met Tyr Lys Pro Tyr Lys Asp Trp Val
    690                 695                 700

Ala Phe Leu Lys Gln Leu His Lys Arg Leu Glu Val Glu Ile Gly Lys
705                 710                 715                 720

Glu Val Lys His Trp Arg Lys Ser Leu Ser Asp Gly Arg Lys Gly Leu
                725                 730                 735

Tyr Gly Ile Ser Leu Lys Asn Ile Asp Glu Ile Asp Arg Thr Arg Lys
            740                 745                 750

Phe Leu Leu Arg Trp Ser Leu Arg Pro Thr Glu Pro Gly Glu Val Arg
        755                 760                 765

Arg Leu Glu Pro Gly Gln Arg Phe Ala Ile Asp Gln Leu Asn His Leu
    770                 775                 780

Asn Ala Leu Lys Glu Asp Arg Leu Lys Lys Met Ala Asn Thr Ile Ile
785                 790                 795                 800

Met His Ala Leu Gly Tyr Cys Tyr Asp Val Arg Lys Lys Lys Trp Gln
                805                 810                 815

Ala Lys Asn Pro Ala Cys Gln Ile Ile Leu Phe Glu Asp Leu Ser Asn
            820                 825                 830

Tyr Asn Pro Tyr Glu Glu Arg Ser Arg Phe Glu Asn Ser Lys Leu Met
        835                 840                 845

Lys Trp Ser Arg Arg Glu Ile Pro Arg Gln Val Ala Leu Gln Gly Glu
    850                 855                 860

Ile Tyr Gly Leu Gln Val Gly Glu Val Gly Ala Gln Phe Ser Ser Arg
865                 870                 875                 880

Phe His Ala Lys Thr Gly Ser Pro Gly Ile Arg Cys Ser Val Val Thr
                885                 890                 895

Lys Glu Lys Leu Gln Asp Asn Arg Phe Phe Lys Asn Leu Gln Arg Glu
            900                 905                 910
```

Gly Arg Leu Thr Leu Asp Lys Ile Ala Val Leu Lys Glu Gly Asp Leu
915 920 925

Tyr Pro Asp Lys Gly Gly Glu Lys Phe Ile Ser Leu Ser Lys Asp Arg
930 935 940

Lys Leu Val Thr Thr His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln
945 950 955 960

Lys Arg Phe Trp Thr Arg Thr His Gly Phe Tyr Lys Val Tyr Cys Lys
965 970 975

Ala Tyr Gln Val Asp Gly Gln Thr Val Tyr Ile Pro Glu Ser Lys Asp
980 985 990

Gln Lys Gln Lys Ile Ile Glu Glu Phe Gly Glu Gly Tyr Phe Ile Leu
995 1000 1005

Lys Asp Gly Val Tyr Glu Trp Gly Asn Ala Gly Lys Leu Lys Ile
1010 1015 1020

Lys Lys Gly Ser Ser Lys Gln Ser Ser Ser Glu Leu Val Asp Ser
1025 1030 1035

Asp Ile Leu Lys Asp Ser Phe Asp Leu Ala Ser Glu Leu Lys Gly
1040 1045 1050

Glu Lys Leu Met Leu Tyr Arg Asp Pro Ser Gly Asn Val Phe Pro
1055 1060 1065

Ser Asp Lys Trp Met Ala Ala Gly Val Phe Phe Gly Lys Leu Glu
1070 1075 1080

Arg Ile Leu Ile Ser Lys Leu Thr Asn Gln Tyr Ser Ile Ser Thr
1085 1090 1095

Ile Glu Asp Asp Ser Ser Lys Gln Ser Met
1100 1105

<210> 55
<211> 1108
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 55
Met Ala Ile Arg Ser Ile Lys Leu Lys Leu Lys Thr His Thr Gly Pro
1 5 10 15

Glu Ala Gln Asn Leu Arg Lys Gly Ile Trp Arg Thr His Arg Leu Leu
20 25 30

Asn Glu Gly Val Ala Tyr Tyr Met Lys Met Leu Leu Leu Phe Arg Gln
35 40 45

Glu Ser Thr Gly Glu Arg Pro Lys Glu Glu Leu Gln Glu Glu Leu Ile
50 55 60

Cys His Ile Arg Glu Gln Gln Gln Arg Asn Gln Ala Asp Lys Asn Thr
65 70 75 80

Gln Ala Leu Pro Leu Asp Lys Ala Leu Glu Ala Leu Arg Gln Leu Tyr

85 90 95

Glu Leu Leu Val Pro Ser Ser Val Gly Gln Ser Gly Asp Ala Gln Ile
100 105 110

Ile Ser Arg Lys Phe Leu Ser Pro Leu Val Asp Pro Asn Ser Glu Gly
115 120 125

Gly Lys Gly Thr Ser Lys Ala Gly Ala Lys Pro Thr Trp Gln Lys Lys
130 135 140

Lys Glu Ala Asn Asp Pro Thr Trp Glu Gln Asp Tyr Glu Lys Trp Lys
145 150 155 160

Lys Arg Arg Glu Glu Asp Pro Thr Ala Ser Val Ile Thr Thr Leu Glu
165 170 175

Glu Tyr Gly Ile Arg Pro Ile Phe Pro Leu Tyr Thr Asn Thr Val Thr
180 185 190

Asp Ile Ala Trp Leu Pro Leu Gln Ser Asn Gln Phe Val Arg Thr Trp
195 200 205

Asp Arg Asp Met Leu Gln Gln Ala Ile Glu Arg Leu Leu Ser Trp Glu
210 215 220

Ser Trp Asn Lys Arg Val Gln Glu Glu Tyr Ala Lys Leu Lys Glu Lys
225 230 235 240

Met Ala Gln Leu Asn Glu Gln Leu Glu Gly Gly Gln Glu Trp Ile Ser
245 250 255

Leu Leu Glu Gln Tyr Glu Glu Asn Arg Glu Arg Glu Leu Arg Glu Asn
260 265 270

Met Thr Ala Ala Asn Asp Lys Tyr Arg Ile Thr Lys Arg Gln Met Lys
275 280 285

Gly Trp Asn Glu Leu Tyr Glu Leu Trp Ser Thr Phe Pro Ala Ser Ala
290 295 300

Ser His Glu Gln Tyr Lys Glu Ala Leu Lys Arg Val Gln Gln Arg Leu
305 310 315 320

Arg Gly Arg Phe Gly Asp Ala His Phe Phe Gln Tyr Leu Met Glu Glu
325 330 335

Lys Asn Arg Leu Ile Trp Lys Gly Asn Pro Gln Arg Ile His Tyr Phe
340 345 350

Val Ala Arg Asn Glu Leu Thr Lys Arg Leu Glu Glu Ala Lys Gln Ser
355 360 365

Ala Thr Met Thr Leu Pro Asn Ala Arg Lys His Pro Leu Trp Val Arg
370 375 380

Phe Asp Ala Arg Gly Gly Asn Leu Gln Asp Tyr Tyr Leu Thr Ala Glu
385 390 395 400

Ala Asp Lys Pro Arg Ser Arg Arg Phe Val Thr Phe Ser Gln Leu Ile
405 410 415

Trp Pro Ser Glu Ser Gly Trp Met Glu Lys Lys Asp Val Glu Val Glu

420 425 430

Leu Ala Leu Ser Arg Gln Phe Tyr Gln Gln Val Lys Leu Leu Lys Asn
435 440 445

Asp Lys Gly Lys Gln Lys Ile Glu Phe Lys Asp Lys Gly Ser Gly Ser
450 455 460

Thr Phe Asn Gly His Leu Gly Gly Ala Lys Leu Gln Leu Glu Arg Gly
465 470 475 480

Asp Leu Glu Lys Glu Glu Lys Asn Phe Glu Asp Gly Glu Ile Gly Ser
485 490 495

Val Tyr Leu Asn Val Val Ile Asp Phe Glu Pro Leu Gln Glu Val Lys
500 505 510

Asn Gly Arg Val Gln Ala Pro Tyr Gly Gln Val Leu Gln Leu Ile Arg
515 520 525

Arg Pro Asn Glu Phe Pro Lys Val Thr Thr Tyr Lys Ser Glu Gln Leu
530 535 540

Val Glu Trp Ile Lys Ala Ser Pro Gln His Ser Ala Gly Val Glu Ser
545 550 555 560

Leu Ala Ser Gly Phe Arg Val Met Ser Ile Asp Leu Gly Leu Arg Ala
565 570 575

Ala Ala Ala Thr Ser Ile Phe Ser Val Glu Glu Ser Ser Asp Lys Asn
580 585 590

Ala Ala Asp Phe Ser Tyr Trp Ile Glu Gly Thr Pro Leu Val Ala Val
595 600 605

His Gln Arg Ser Tyr Met Leu Arg Leu Pro Gly Glu Gln Val Glu Lys
610 615 620

Gln Val Met Glu Lys Arg Asp Glu Arg Phe Gln Leu His Gln Arg Val
625 630 635 640

Lys Phe Gln Ile Arg Val Leu Ala Gln Ile Met Arg Met Ala Asn Lys
645 650 655

Gln Tyr Gly Asp Arg Trp Asp Glu Leu Asp Ser Leu Lys Gln Ala Val
660 665 670

Glu Gln Lys Lys Ser Pro Leu Asp Gln Thr Asp Arg Thr Phe Trp Glu
675 680 685

Gly Ile Val Cys Asp Leu Thr Lys Val Leu Pro Arg Asn Glu Ala Asp
690 695 700

Trp Glu Gln Ala Val Val Gln Ile His Arg Lys Ala Glu Glu Tyr Val
705 710 715 720

Gly Lys Ala Val Gln Ala Trp Arg Lys Arg Phe Ala Ala Asp Glu Arg
725 730 735

Lys Gly Ile Ala Gly Leu Ser Met Trp Asn Ile Glu Glu Leu Glu Gly
740 745 750

Leu Arg Lys Leu Leu Ile Ser Trp Ser Arg Arg Thr Arg Asn Pro Gln

```
        755                 760                 765

Glu Val Asn Arg Phe Glu Arg Gly His Thr Ser His Gln Arg Leu Leu
    770                 775                 780

Thr His Ile Gln Asn Val Lys Glu Asp Arg Leu Lys Gln Leu Ser His
785                 790                 795                 800

Ala Ile Val Met Thr Ala Leu Gly Tyr Val Tyr Asp Glu Arg Lys Gln
                805                 810                 815

Glu Trp Cys Ala Glu Tyr Pro Ala Cys Gln Val Ile Leu Phe Glu Asn
            820                 825                 830

Leu Ser Gln Tyr Arg Ser Asn Leu Asp Arg Ser Thr Lys Glu Asn Ser
        835                 840                 845

Thr Leu Met Lys Trp Ala His Arg Ser Ile Pro Lys Tyr Val His Met
    850                 855                 860

Gln Ala Glu Pro Tyr Gly Ile Gln Ile Gly Asp Val Arg Ala Glu Tyr
865                 870                 875                 880

Ser Ser Arg Phe Tyr Ala Lys Thr Gly Thr Pro Gly Ile Arg Cys Lys
                885                 890                 895

Lys Val Arg Gly Gln Asp Leu Gln Gly Arg Arg Phe Glu Asn Leu Gln
            900                 905                 910

Lys Arg Leu Val Asn Glu Gln Phe Leu Thr Glu Glu Gln Val Lys Gln
        915                 920                 925

Leu Arg Pro Gly Asp Ile Val Pro Asp Asp Ser Gly Glu Leu Phe Met
    930                 935                 940

Thr Leu Thr Asp Gly Ser Gly Ser Lys Glu Val Val Phe Leu Gln Ala
945                 950                 955                 960

Asp Ile Asn Ala Ala His Asn Leu Gln Lys Arg Phe Trp Gln Arg Tyr
                965                 970                 975

Asn Glu Leu Phe Lys Val Ser Cys Arg Val Ile Val Arg Asp Glu Glu
            980                 985                 990

Glu Tyr Leu Val Pro Lys Thr Lys Ser Val Gln Ala Lys Leu Gly Lys
        995                 1000                1005

Gly Leu  Phe Val Lys Lys Ser Asp Thr Ala Trp  Lys Asp Val Tyr
    1010                1015                1020

Val Trp  Asp Ser Gln Ala Lys Leu Lys Gly Lys  Thr Thr Phe Thr
    1025                1030                1035

Glu Glu  Ser Glu Ser Pro Glu Gln Leu Glu Asp  Phe Gln Glu Ile
    1040                1045                1050

Ile Glu  Glu Ala Glu Glu Ala Lys Gly Thr Tyr  Arg Thr Leu Phe
    1055                1060                1065

Arg Asp  Pro Ser Gly Val Phe Phe Pro Glu Ser  Val Trp Tyr Pro
    1070                1075                1080

Gln Lys  Asp Phe Trp Gly Glu Val Lys Arg Lys  Leu Tyr Gly Lys
```

1085 1090 1095

Leu Arg Glu Arg Phe Leu Thr Lys Ala Arg
1100 1105

<210> 56
<211> 1334
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 56
Met Lys Ile Ser Lys Val Asp His Thr Arg Met Ala Val Ala Lys Gly
1 5 10 15

Asn Gln His Arg Arg Asp Glu Ile Ser Gly Ile Leu Tyr Lys Asp Pro
20 25 30

Thr Lys Thr Gly Ser Ile Asp Phe Asp Glu Arg Phe Lys Lys Leu Asn
35 40 45

Cys Ser Ala Lys Ile Leu Tyr His Val Phe Asn Gly Ile Ala Glu Gly
50 55 60

Ser Asn Lys Tyr Lys Asn Ile Val Asp Lys Val Asn Asn Asn Leu Asp
65 70 75 80

Arg Val Leu Phe Thr Gly Lys Ser Tyr Asp Arg Lys Ser Ile Ile Asp
85 90 95

Ile Asp Thr Val Leu Arg Asn Val Glu Lys Ile Asn Ala Phe Asp Arg
100 105 110

Ile Ser Thr Glu Glu Arg Glu Gln Ile Ile Asp Asp Leu Leu Glu Ile
115 120 125

Gln Leu Arg Lys Gly Leu Arg Lys Gly Lys Ala Gly Leu Arg Glu Val
130 135 140

Leu Leu Ile Gly Ala Gly Val Ile Val Arg Thr Asp Lys Lys Gln Glu
145 150 155 160

Ile Ala Asp Phe Leu Glu Ile Leu Asp Glu Asp Phe Asn Lys Thr Asn
165 170 175

Gln Ala Lys Asn Ile Lys Leu Ser Ile Glu Asn Gln Gly Leu Val Val
180 185 190

Ser Pro Val Ser Arg Gly Glu Glu Arg Ile Phe Asp Val Ser Gly Ala
195 200 205

Gln Lys Gly Lys Ser Ser Lys Lys Ala Gln Glu Lys Glu Ala Leu Ser
210 215 220

Ala Phe Leu Leu Asp Tyr Ala Asp Leu Asp Lys Asn Val Arg Phe Glu
225 230 235 240

Tyr Leu Arg Lys Ile Arg Arg Leu Ile Asn Leu Tyr Phe Tyr Val Lys
245 250 255

Asn Asp Asp Val Met Ser Leu Thr Glu Ile Pro Ala Glu Val Asn Leu
260 265 270

```
Glu Lys Asp Phe Asp Ile Trp Arg Asp His Glu Gln Arg Lys Glu Glu
        275                 280                 285

Asn Gly Asp Phe Val Gly Cys Pro Asp Ile Leu Leu Ala Asp Arg Asp
    290                 295                 300

Val Lys Lys Ser Asn Ser Lys Gln Val Lys Ile Ala Glu Arg Gln Leu
305                 310                 315                 320

Arg Glu Ser Ile Arg Glu Lys Asn Ile Lys Arg Tyr Arg Phe Ser Ile
                325                 330                 335

Lys Thr Ile Glu Lys Asp Asp Gly Thr Tyr Phe Phe Ala Asn Lys Gln
            340                 345                 350

Ile Ser Val Phe Trp Ile His Arg Ile Glu Asn Ala Val Glu Arg Ile
        355                 360                 365

Leu Gly Ser Ile Asn Asp Lys Lys Leu Tyr Arg Leu Arg Leu Gly Tyr
    370                 375                 380

Leu Gly Glu Lys Val Trp Lys Asp Ile Leu Asn Phe Leu Ser Ile Lys
385                 390                 395                 400

Tyr Ile Ala Val Gly Lys Ala Val Phe Asn Phe Ala Met Asp Asp Leu
                405                 410                 415

Gln Glu Lys Asp Arg Asp Ile Glu Pro Gly Lys Ile Ser Glu Asn Ala
            420                 425                 430

Val Asn Gly Leu Thr Ser Phe Asp Tyr Glu Gln Ile Lys Ala Asp Glu
        435                 440                 445

Met Leu Gln Arg Glu Val Ala Val Asn Val Ala Phe Ala Ala Asn Asn
    450                 455                 460

Leu Ala Arg Val Thr Val Asp Ile Pro Gln Asn Gly Glu Lys Glu Asp
465                 470                 475                 480

Ile Leu Leu Trp Asn Lys Ser Asp Ile Lys Lys Tyr Lys Lys Asn Ser
                485                 490                 495

Lys Lys Gly Ile Leu Lys Ser Ile Leu Gln Phe Phe Gly Gly Ala Ser
            500                 505                 510

Thr Trp Asn Met Lys Met Phe Glu Ile Ala Tyr His Asp Gln Pro Gly
        515                 520                 525

Asp Tyr Glu Glu Asn Tyr Leu Tyr Asp Ile Ile Gln Ile Ile Tyr Ser
    530                 535                 540

Leu Arg Asn Lys Ser Phe His Phe Lys Thr Tyr Asp His Gly Asp Lys
545                 550                 555                 560

Asn Trp Asn Arg Glu Leu Ile Gly Lys Met Ile Glu His Asp Ala Glu
                565                 570                 575

Arg Val Ile Ser Val Glu Arg Glu Lys Phe His Ser Asn Asn Leu Pro
            580                 585                 590

Met Phe Tyr Lys Asp Ala Asp Leu Lys Lys Ile Leu Asp Leu Leu Tyr
        595                 600                 605
```

```
Ser Asp Tyr Ala Gly Arg Ala Ser Gln Val Pro Ala Phe Asn Thr Val
    610                 615                 620

Leu Val Arg Lys Asn Phe Pro Glu Phe Leu Arg Lys Asp Met Gly Tyr
625                 630                 635                 640

Lys Val His Phe Asn Asn Pro Glu Val Glu Asn Gln Trp His Ser Ala
                645                 650                 655

Val Tyr Tyr Leu Tyr Lys Glu Ile Tyr Tyr Asn Leu Phe Leu Arg Asp
            660                 665                 670

Lys Glu Val Lys Asn Leu Phe Tyr Thr Ser Leu Lys Asn Ile Arg Ser
        675                 680                 685

Glu Val Ser Asp Lys Lys Gln Lys Leu Ala Ser Asp Asp Phe Ala Ser
    690                 695                 700

Arg Cys Glu Glu Ile Glu Asp Arg Ser Leu Pro Glu Ile Cys Gln Ile
705                 710                 715                 720

Ile Met Thr Glu Tyr Asn Ala Gln Asn Phe Gly Asn Arg Lys Val Lys
                725                 730                 735

Ser Gln Arg Val Ile Glu Lys Asn Lys Asp Ile Phe Arg His Tyr Lys
            740                 745                 750

Met Leu Leu Ile Lys Thr Leu Ala Gly Ala Phe Ser Leu Tyr Leu Lys
        755                 760                 765

Gln Glu Arg Phe Ala Phe Ile Gly Lys Ala Thr Pro Ile Pro Tyr Glu
    770                 775                 780

Thr Thr Asp Val Lys Asn Phe Leu Pro Glu Trp Lys Ser Gly Met Tyr
785                 790                 795                 800

Ala Ser Phe Val Glu Glu Ile Lys Asn Asn Leu Asp Leu Gln Glu Trp
                805                 810                 815

Tyr Ile Val Gly Arg Phe Leu Asn Gly Arg Met Leu Asn Gln Leu Ala
            820                 825                 830

Gly Ser Leu Arg Ser Tyr Ile Gln Tyr Ala Glu Asp Ile Glu Arg Arg
        835                 840                 845

Ala Ala Glu Asn Arg Asn Lys Leu Phe Ser Lys Pro Asp Glu Lys Ile
    850                 855                 860

Glu Ala Cys Lys Lys Ala Val Arg Val Leu Asp Leu Cys Ile Lys Ile
865                 870                 875                 880

Ser Thr Arg Ile Ser Ala Glu Phe Thr Asp Tyr Phe Asp Ser Glu Asp
                885                 890                 895

Asp Tyr Ala Asp Tyr Leu Glu Lys Tyr Leu Lys Tyr Gln Asp Asp Ala
            900                 905                 910

Ile Lys Glu Leu Ser Gly Ser Ser Tyr Ala Ala Leu Asp His Phe Cys
        915                 920                 925

Asn Lys Asp Asp Leu Lys Phe Asp Ile Tyr Val Asn Ala Gly Gln Lys
    930                 935                 940
```

```
Pro Ile Leu Gln Arg Asn Ile Val Met Ala Lys Leu Phe Gly Pro Asp
945                 950                 955                 960

Asn Ile Leu Ser Glu Val Met Glu Lys Val Thr Glu Ser Ala Ile Arg
                965                 970                 975

Glu Tyr Tyr Asp Tyr Leu Lys Lys Val Ser Gly Tyr Arg Val Arg Gly
            980                 985                 990

Lys Cys Ser Thr Glu Lys Glu Gln Glu Asp Leu Leu Lys Phe Gln Arg
        995                1000                1005

Leu Lys  Asn Ala Val Glu Phe  Arg Asp Val Thr Glu  Tyr Ala Glu
    1010                1015                1020

Val Ile  Asn Glu Leu Leu Gly  Gln Leu Ile Ser Trp  Ser Tyr Leu
    1025                1030                1035

Arg Glu  Arg Asp Leu Leu Tyr  Phe Gln Leu Gly Phe  His Tyr Met
    1040                1045                1050

Cys Leu  Lys Asn Lys Ser Phe  Lys Pro Ala Glu Tyr  Val Asp Ile
    1055                1060                1065

Arg Arg  Asn Asn Gly Thr Ile  Ile His Asn Ala Ile  Leu Tyr Gln
    1070                1075                1080

Ile Val  Ser Met Tyr Ile Asn  Gly Leu Asp Phe Tyr  Ser Cys Asp
    1085                1090                1095

Lys Glu  Gly Lys Thr Leu Lys  Pro Ile Glu Thr Gly  Lys Gly Val
    1100                1105                1110

Gly Ser  Lys Ile Gly Gln Phe  Ile Lys Tyr Ser Gln  Tyr Leu Tyr
    1115                1120                1125

Asn Asp  Pro Ser Tyr Lys Leu  Glu Ile Tyr Asn Ala  Gly Leu Glu
    1130                1135                1140

Val Phe  Glu Asn Ile Asp Glu  His Asp Asn Ile Thr  Asp Leu Arg
    1145                1150                1155

Lys Tyr  Val Asp His Phe Lys  Tyr Tyr Ala Tyr Gly  Asn Lys Met
    1160                1165                1170

Ser Leu  Leu Asp Leu Tyr Ser  Glu Phe Phe Asp Arg  Phe Phe Thr
    1175                1180                1185

Tyr Asp  Met Lys Tyr Gln Lys  Asn Val Val Asn Val  Leu Glu Asn
    1190                1195                1200

Ile Leu  Leu Arg His Phe Val  Ile Phe Tyr Pro Lys  Phe Gly Ser
    1205                1210                1215

Gly Lys  Lys Asp Val Gly Ile  Arg Asp Cys Lys Lys  Glu Arg Ala
    1220                1225                1230

Gln Ile  Glu Ile Ser Glu Gln  Ser Leu Thr Ser Glu  Asp Phe Met
    1235                1240                1245

Phe Lys  Leu Asp Asp Lys Ala  Gly Glu Glu Ala Lys  Lys Phe Pro
    1250                1255                1260
```

Ala Arg Asp Glu Arg Tyr Leu Gln Thr Ile Ala Lys Leu Leu Tyr
1265 1270 1275

Tyr Pro Asn Glu Ile Glu Asp Met Asn Arg Phe Met Lys Lys Gly
1280 1285 1290

Glu Thr Ile Asn Lys Lys Val Gln Phe Asn Arg Lys Lys Lys Ile
1295 1300 1305

Thr Arg Lys Gln Lys Asn Asn Ser Ser Asn Glu Val Leu Ser Ser
1310 1315 1320

Thr Met Gly Tyr Leu Phe Lys Asn Ile Lys Leu
1325 1330

<210> 57
<211> 1120
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 57
Met Trp Ile Ser Ile Lys Thr Leu Ile His His Leu Gly Val Leu Phe
1 5 10 15

Phe Cys Asp Tyr Met Tyr Asn Arg Arg Glu Lys Lys Ile Ile Glu Val
20 25 30

Lys Thr Met Arg Ile Thr Lys Val Glu Val Asp Arg Lys Lys Val Leu
35 40 45

Ile Ser Arg Asp Lys Asn Gly Gly Lys Leu Val Tyr Glu Asn Glu Met
50 55 60

Gln Asp Asn Thr Glu Gln Ile Met His His Lys Lys Ser Ser Phe Tyr
65 70 75 80

Lys Ser Val Val Asn Lys Thr Ile Cys Arg Pro Glu Gln Lys Gln Met
85 90 95

Lys Lys Leu Val His Gly Leu Leu Gln Glu Asn Ser Gln Glu Lys Ile
100 105 110

Lys Val Ser Asp Val Thr Lys Leu Asn Ile Ser Asn Phe Leu Asn His
115 120 125

Arg Phe Lys Lys Ser Leu Tyr Tyr Phe Pro Glu Asn Ser Pro Asp Lys
130 135 140

Ser Glu Glu Tyr Arg Ile Glu Ile Asn Leu Ser Gln Leu Leu Glu Asp
145 150 155 160

Ser Leu Lys Lys Gln Gln Gly Thr Phe Ile Cys Trp Glu Ser Phe Ser
165 170 175

Lys Asp Met Glu Leu Tyr Ile Asn Trp Ala Glu Asn Tyr Ile Ser Ser
180 185 190

Lys Thr Lys Leu Ile Lys Lys Ser Ile Arg Asn Asn Arg Ile Gln Ser
195 200 205

Thr Glu Ser Arg Ser Gly Gln Leu Met Asp Arg Tyr Met Lys Asp Ile

210 215 220

Leu Asn Lys Asn Lys Pro Phe Asp Ile Gln Ser Val Ser Glu Lys Tyr
225 230 235 240

Gln Leu Glu Lys Leu Thr Ser Ala Leu Lys Ala Thr Phe Lys Glu Ala
245 250 255

Lys Lys Asn Asp Lys Glu Ile Asn Tyr Lys Leu Lys Ser Thr Leu Gln
260 265 270

Asn His Glu Arg Gln Ile Ile Glu Glu Leu Lys Glu Asn Ser Glu Leu
275 280 285

Asn Gln Phe Asn Ile Glu Ile Arg Lys His Leu Glu Thr Tyr Phe Pro
290 295 300

Ile Lys Lys Thr Asn Arg Lys Val Gly Asp Ile Arg Asn Leu Glu Ile
305 310 315 320

Gly Glu Ile Gln Lys Ile Val Asn His Arg Leu Lys Asn Lys Ile Val
325 330 335

Gln Arg Ile Leu Gln Glu Gly Lys Leu Ala Ser Tyr Glu Ile Glu Ser
340 345 350

Thr Val Asn Ser Asn Ser Leu Gln Lys Ile Lys Ile Glu Glu Ala Phe
355 360 365

Ala Leu Lys Phe Ile Asn Ala Cys Leu Phe Ala Ser Asn Asn Leu Arg
370 375 380

Asn Met Val Tyr Pro Val Cys Lys Lys Asp Ile Leu Met Ile Gly Glu
385 390 395 400

Phe Lys Asn Ser Phe Lys Glu Ile Lys His Lys Lys Phe Ile Arg Gln
405 410 415

Trp Ser Gln Phe Phe Ser Gln Glu Ile Thr Val Asp Asp Ile Glu Leu
420 425 430

Ala Ser Trp Gly Leu Arg Gly Ala Ile Ala Pro Ile Arg Asn Glu Ile
435 440 445

Ile His Leu Lys Lys His Ser Trp Lys Lys Phe Phe Asn Asn Pro Thr
450 455 460

Phe Lys Val Lys Lys Ser Lys Ile Ile Asn Gly Lys Thr Lys Asp Val
465 470 475 480

Thr Ser Glu Phe Leu Tyr Lys Glu Thr Leu Phe Lys Asp Tyr Phe Tyr
485 490 495

Ser Glu Leu Asp Ser Val Pro Glu Leu Ile Ile Asn Lys Met Glu Ser
500 505 510

Ser Lys Ile Leu Asp Tyr Tyr Ser Ser Asp Gln Leu Asn Gln Val Phe
515 520 525

Thr Ile Pro Asn Phe Glu Leu Ser Leu Leu Thr Ser Ala Val Pro Phe
530 535 540

Ala Pro Ser Phe Lys Arg Val Tyr Leu Lys Gly Phe Asp Tyr Gln Asn

545 550 555 560

Gln Asp Glu Ala Gln Pro Asp Tyr Asn Leu Lys Leu Asn Ile Tyr Asn
565 570 575

Glu Lys Ala Phe Asn Ser Glu Ala Phe Gln Ala Gln Tyr Ser Leu Phe
580 585 590

Lys Met Val Tyr Tyr Gln Val Phe Leu Pro Gln Phe Thr Thr Asn Asn
595 600 605

Asp Leu Phe Lys Ser Ser Val Asp Phe Ile Leu Thr Leu Asn Lys Glu
610 615 620

Arg Lys Gly Tyr Ala Lys Ala Phe Gln Asp Ile Arg Lys Met Asn Lys
625 630 635 640

Asp Glu Lys Pro Ser Glu Tyr Met Ser Tyr Ile Gln Ser Gln Leu Met
645 650 655

Leu Tyr Gln Lys Lys Gln Glu Glu Lys Glu Lys Ile Asn His Phe Glu
660 665 670

Lys Phe Ile Asn Gln Val Phe Ile Lys Gly Phe Asn Ser Phe Ile Glu
675 680 685

Lys Asn Arg Leu Thr Tyr Ile Cys His Pro Thr Lys Asn Thr Val Pro
690 695 700

Glu Asn Asp Asn Ile Glu Ile Pro Phe His Thr Asp Met Asp Asp Ser
705 710 715 720

Asn Ile Ala Phe Trp Leu Met Cys Lys Leu Leu Asp Ala Lys Gln Leu
725 730 735

Ser Glu Leu Arg Asn Glu Met Ile Lys Phe Ser Cys Ser Leu Gln Ser
740 745 750

Thr Glu Glu Ile Ser Thr Phe Thr Lys Ala Arg Glu Val Ile Gly Leu
755 760 765

Ala Leu Leu Asn Gly Glu Lys Gly Cys Asn Asp Trp Lys Glu Leu Phe
770 775 780

Asp Asp Lys Glu Ala Trp Lys Lys Asn Met Ser Leu Tyr Val Ser Glu
785 790 795 800

Glu Leu Leu Gln Ser Leu Pro Tyr Thr Gln Glu Asp Gly Gln Thr Pro
805 810 815

Val Ile Asn Arg Ser Ile Asp Leu Val Lys Lys Tyr Gly Thr Glu Thr
820 825 830

Ile Leu Glu Lys Leu Phe Ser Ser Ser Asp Asp Tyr Lys Val Ser Ala
835 840 845

Lys Asp Ile Ala Lys Leu His Glu Tyr Asp Val Thr Glu Lys Ile Ala
850 855 860

Gln Gln Glu Ser Leu His Lys Gln Trp Ile Glu Lys Pro Gly Leu Ala
865 870 875 880

Arg Asp Ser Ala Trp Thr Lys Lys Tyr Gln Asn Val Ile Asn Asp Ile

885 890 895

Ser Asn Tyr Gln Trp Ala Lys Thr Lys Val Glu Leu Thr Gln Val Arg
900 905 910

His Leu His Gln Leu Thr Ile Asp Leu Leu Ser Arg Leu Ala Gly Tyr
915 920 925

Met Ser Ile Ala Asp Arg Asp Phe Gln Phe Ser Ser Asn Tyr Ile Leu
930 935 940

Glu Arg Glu Asn Ser Glu Tyr Arg Val Thr Ser Trp Ile Leu Leu Ser
945 950 955 960

Glu Asn Lys Asn Lys Asn Lys Tyr Asn Asp Tyr Glu Leu Tyr Asn Leu
965 970 975

Lys Asn Ala Ser Ile Lys Val Ser Ser Lys Asn Asp Pro Gln Leu Lys
980 985 990

Val Asp Leu Lys Gln Leu Arg Leu Thr Leu Glu Tyr Leu Glu Leu Phe
995 1000 1005

Asp Asn Arg Leu Lys Glu Lys Arg Asn Asn Ile Ser His Phe Asn
1010 1015 1020

Tyr Leu Asn Gly Gln Leu Gly Asn Ser Ile Leu Glu Leu Phe Asp
1025 1030 1035

Asp Ala Arg Asp Val Leu Ser Tyr Asp Arg Lys Leu Lys Asn Ala
1040 1045 1050

Val Ser Lys Ser Leu Lys Glu Ile Leu Ser Ser His Gly Met Glu
1055 1060 1065

Val Thr Phe Lys Pro Leu Tyr Gln Thr Asn His His Leu Lys Ile
1070 1075 1080

Asp Lys Leu Gln Pro Lys Lys Ile His His Leu Gly Glu Lys Ser
1085 1090 1095

Thr Val Ser Ser Asn Gln Val Ser Asn Glu Tyr Cys Gln Leu Val
1100 1105 1110

Arg Thr Leu Leu Thr Met Lys
1115 1120

<210> 58
<211> 1152
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 58
Met Lys Val Thr Lys Val Asp Gly Ile Ser His Lys Lys Tyr Ile Glu
1 5 10 15

Glu Gly Lys Leu Val Lys Ser Thr Ser Glu Glu Asn Arg Thr Ser Glu
20 25 30

Arg Leu Ser Glu Leu Leu Ser Ile Arg Leu Asp Ile Tyr Ile Lys Asn
35 40 45

```
Pro Asp Asn Ala Ser Glu Glu Glu Asn Arg Ile Arg Arg Glu Asn Leu
    50                  55                  60

Lys Lys Phe Phe Ser Asn Lys Val Leu His Leu Lys Asp Ser Val Leu
65                  70                  75                  80

Tyr Leu Lys Asn Arg Lys Glu Lys Asn Ala Val Gln Asp Lys Asn Tyr
                85                  90                  95

Ser Glu Glu Asp Ile Ser Glu Tyr Asp Leu Lys Asn Lys Asn Ser Phe
            100                 105                 110

Ser Val Leu Lys Lys Ile Leu Leu Asn Glu Asp Val Asn Ser Glu Glu
        115                 120                 125

Leu Glu Ile Phe Arg Lys Asp Val Glu Ala Lys Leu Asn Lys Ile Asn
    130                 135                 140

Ser Leu Lys Tyr Ser Phe Glu Glu Asn Lys Ala Asn Tyr Gln Lys Ile
145                 150                 155                 160

Asn Glu Asn Asn Val Glu Lys Val Gly Gly Lys Ser Lys Arg Asn Ile
                165                 170                 175

Ile Tyr Asp Tyr Tyr Arg Glu Ser Ala Lys Arg Asn Asp Tyr Ile Asn
            180                 185                 190

Asn Val Gln Glu Ala Phe Asp Lys Leu Tyr Lys Lys Glu Asp Ile Glu
        195                 200                 205

Lys Leu Phe Phe Leu Ile Glu Asn Ser Lys Lys His Glu Lys Tyr Lys
    210                 215                 220

Ile Arg Glu Tyr Tyr His Lys Ile Ile Gly Arg Lys Asn Asp Lys Glu
225                 230                 235                 240

Asn Phe Ala Lys Ile Ile Tyr Glu Glu Ile Gln Asn Val Asn Asn Ile
                245                 250                 255

Lys Glu Leu Ile Glu Lys Ile Pro Asp Met Ser Glu Leu Lys Lys Ser
            260                 265                 270

Gln Val Phe Tyr Lys Tyr Tyr Leu Asp Lys Glu Glu Leu Asn Asp Lys
        275                 280                 285

Asn Ile Lys Tyr Ala Phe Cys His Phe Val Glu Ile Glu Met Ser Gln
    290                 295                 300

Leu Leu Lys Asn Tyr Val Tyr Lys Arg Leu Ser Asn Ile Ser Asn Asp
305                 310                 315                 320

Lys Ile Lys Arg Ile Phe Glu Tyr Gln Asn Leu Lys Lys Leu Ile Glu
                325                 330                 335

Asn Lys Leu Leu Asn Lys Leu Asp Thr Tyr Val Arg Asn Cys Gly Lys
            340                 345                 350

Tyr Asn Tyr Tyr Leu Gln Val Gly Glu Ile Ala Thr Ser Asp Phe Ile
        355                 360                 365

Ala Arg Asn Arg Gln Asn Glu Ala Phe Leu Arg Asn Ile Ile Gly Val
    370                 375                 380
```

```
Ser Ser Val Ala Tyr Phe Ser Leu Arg Asn Ile Leu Glu Thr Glu Asn
385                 390                 395                 400

Glu Asn Asp Ile Thr Gly Arg Met Arg Gly Lys Thr Val Lys Asn Asn
                405                 410                 415

Lys Gly Glu Glu Lys Tyr Val Ser Gly Glu Val Asp Lys Ile Tyr Asn
            420                 425                 430

Glu Asn Lys Gln Asn Glu Val Lys Glu Asn Leu Lys Met Phe Tyr Ser
        435                 440                 445

Tyr Asp Phe Asn Met Asp Asn Lys Asn Glu Ile Glu Asp Phe Phe Ala
    450                 455                 460

Asn Ile Asp Glu Ala Ile Ser Ser Ile Arg His Gly Ile Val His Phe
465                 470                 475                 480

Asn Leu Glu Leu Glu Gly Lys Asp Ile Phe Ala Phe Lys Asn Ile Ala
                485                 490                 495

Pro Ser Glu Ile Ser Lys Lys Met Phe Gln Asn Glu Ile Asn Glu Lys
            500                 505                 510

Lys Leu Lys Leu Lys Ile Phe Lys Gln Leu Asn Ser Ala Asn Val Phe
        515                 520                 525

Asn Tyr Tyr Glu Lys Asp Val Ile Ile Lys Tyr Leu Lys Asn Thr Lys
    530                 535                 540

Phe Asn Phe Val Asn Lys Asn Ile Pro Phe Val Pro Ser Phe Thr Lys
545                 550                 555                 560

Leu Tyr Asn Lys Ile Glu Asp Leu Arg Asn Thr Leu Lys Phe Phe Trp
                565                 570                 575

Ser Val Pro Lys Asp Lys Glu Glu Lys Asp Ala Gln Ile Tyr Leu Leu
            580                 585                 590

Lys Asn Ile Tyr Tyr Gly Glu Phe Leu Asn Lys Phe Val Lys Asn Ser
        595                 600                 605

Lys Val Phe Phe Lys Ile Thr Asn Glu Val Ile Lys Ile Asn Lys Gln
    610                 615                 620

Arg Asn Gln Lys Thr Gly His Tyr Lys Tyr Gln Lys Phe Glu Asn Ile
625                 630                 635                 640

Glu Lys Thr Val Pro Val Glu Tyr Leu Ala Ile Ile Gln Ser Arg Glu
                645                 650                 655

Met Ile Asn Asn Gln Asp Lys Glu Glu Lys Asn Thr Tyr Ile Asp Phe
            660                 665                 670

Ile Gln Gln Ile Phe Leu Lys Gly Phe Ile Asp Tyr Leu Asn Lys Asn
        675                 680                 685

Asn Leu Lys Tyr Ile Glu Ser Asn Asn Asn Asn Asp Asn Asn Asp Ile
    690                 695                 700

Phe Ser Lys Ile Lys Ile Lys Lys Asp Asn Lys Glu Lys Tyr Asp Lys
705                 710                 715                 720
```

```
Ile Leu Lys Asn Tyr Glu Lys His Asn Arg Asn Lys Glu Ile Pro His
                725                 730                 735

Glu Ile Asn Glu Phe Val Arg Glu Ile Lys Leu Gly Lys Ile Leu Lys
            740                 745                 750

Tyr Thr Glu Asn Leu Asn Met Phe Tyr Leu Ile Leu Lys Leu Leu Asn
        755                 760                 765

His Lys Glu Leu Thr Asn Leu Lys Gly Ser Leu Glu Lys Tyr Gln Ser
    770                 775                 780

Ala Asn Lys Glu Glu Thr Phe Ser Asp Glu Leu Glu Leu Ile Asn Leu
785                 790                 795                 800

Leu Asn Leu Asp Asn Asn Arg Val Thr Glu Asp Phe Glu Leu Glu Ala
                805                 810                 815

Asn Glu Ile Gly Lys Phe Leu Asp Phe Asn Glu Asn Lys Ile Lys Asp
            820                 825                 830

Arg Lys Glu Leu Lys Lys Phe Asp Thr Asn Lys Ile Tyr Phe Asp Gly
        835                 840                 845

Glu Asn Ile Ile Lys His Arg Ala Phe Tyr Asn Ile Lys Lys Tyr Gly
    850                 855                 860

Met Leu Asn Leu Leu Glu Lys Ile Ala Asp Lys Ala Lys Tyr Lys Ile
865                 870                 875                 880

Ser Leu Lys Glu Leu Lys Glu Tyr Ser Asn Lys Lys Asn Glu Ile Glu
                885                 890                 895

Lys Asn Tyr Thr Met Gln Gln Asn Leu His Arg Lys Tyr Ala Arg Pro
            900                 905                 910

Lys Lys Asp Glu Lys Phe Asn Asp Glu Asp Tyr Lys Glu Tyr Glu Lys
        915                 920                 925

Ala Ile Gly Asn Ile Gln Lys Tyr Thr His Leu Lys Asn Lys Val Glu
    930                 935                 940

Phe Asn Glu Leu Asn Leu Leu Gln Gly Leu Leu Leu Lys Ile Leu His
945                 950                 955                 960

Arg Leu Val Gly Tyr Thr Ser Ile Trp Glu Arg Asp Leu Arg Phe Arg
                965                 970                 975

Leu Lys Gly Glu Phe Pro Glu Asn His Tyr Ile Glu Glu Ile Phe Asn
            980                 985                 990

Phe Asp Asn Ser Lys Asn Val Lys Tyr Lys Ser Gly Gln Ile Val Glu
        995                 1000                1005

Lys Tyr Ile Asn Phe Tyr Lys Glu Leu Tyr Lys Asp Asn Val Glu
    1010                1015                1020

Lys Arg Ser Ile Tyr Ser Asp Lys Lys Val Lys Lys Leu Lys Gln
    1025                1030                1035

Glu Lys Lys Asp Leu Tyr Ile Arg Asn Tyr Ile Ala His Phe Asn
    1040                1045                1050
```

Tyr Ile Pro His Ala Glu Ile Ser Leu Leu Glu Val Leu Glu Asn
1055 1060 1065

Leu Arg Lys Leu Leu Ser Tyr Asp Arg Lys Leu Lys Asn Ala Ile
1070 1075 1080

Met Lys Ser Ile Val Asp Ile Leu Lys Glu Tyr Gly Phe Val Ala
1085 1090 1095

Thr Phe Lys Ile Gly Ala Asp Lys Lys Ile Glu Ile Gln Thr Leu
1100 1105 1110

Glu Ser Glu Lys Ile Val His Leu Lys Asn Leu Lys Lys Lys Lys
1115 1120 1125

Leu Met Thr Asp Arg Asn Ser Glu Glu Leu Cys Glu Leu Val Lys
1130 1135 1140

Val Met Phe Glu Tyr Lys Ala Leu Glu
1145 1150

<210> 59
<211> 1389
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 59
Met Gly Asn Leu Phe Gly His Lys Arg Trp Tyr Glu Val Arg Asp Lys
1 5 10 15

Lys Asp Phe Lys Ile Lys Arg Lys Val Lys Val Lys Arg Asn Tyr Asp
20 25 30

Gly Asn Lys Tyr Ile Leu Asn Ile Asn Glu Asn Asn Asn Lys Glu Lys
35 40 45

Ile Asp Asn Asn Lys Phe Ile Arg Lys Tyr Ile Asn Tyr Lys Lys Asn
50 55 60

Asp Asn Ile Leu Lys Glu Phe Thr Arg Lys Phe His Ala Gly Asn Ile
65 70 75 80

Leu Phe Lys Leu Lys Gly Lys Glu Gly Ile Ile Arg Ile Glu Asn Asn
85 90 95

Asp Asp Phe Leu Glu Thr Glu Glu Val Val Leu Tyr Ile Glu Ala Tyr
100 105 110

Gly Lys Ser Glu Lys Leu Lys Ala Leu Gly Ile Thr Lys Lys Lys Ile
115 120 125

Ile Asp Glu Ala Ile Arg Gln Gly Ile Thr Lys Asp Asp Lys Lys Ile
130 135 140

Glu Ile Lys Arg Gln Glu Asn Glu Glu Glu Ile Glu Ile Asp Ile Arg
145 150 155 160

Asp Glu Tyr Thr Asn Lys Thr Leu Asn Asp Cys Ser Ile Ile Leu Arg
165 170 175

Ile Ile Glu Asn Asp Glu Leu Glu Thr Lys Lys Ser Ile Tyr Glu Ile

180 185 190

Phe Lys Asn Ile Asn Met Ser Leu Tyr Lys Ile Ile Glu Lys Ile Ile
195 200 205

Glu Asn Glu Thr Glu Lys Val Phe Glu Asn Arg Tyr Tyr Glu Glu His
210 215 220

Leu Arg Glu Lys Leu Leu Lys Asp Asp Lys Ile Asp Val Ile Leu Thr
225 230 235 240

Asn Phe Met Glu Ile Arg Glu Lys Ile Lys Ser Asn Leu Glu Ile Leu
245 250 255

Gly Phe Val Lys Phe Tyr Leu Asn Val Gly Gly Asp Lys Lys Lys Ser
260 265 270

Lys Asn Lys Lys Met Leu Val Glu Lys Ile Leu Asn Ile Asn Val Asp
275 280 285

Leu Thr Val Glu Asp Ile Ala Asp Phe Val Ile Lys Glu Leu Glu Phe
290 295 300

Trp Asn Ile Thr Lys Arg Ile Glu Lys Val Lys Lys Val Asn Asn Glu
305 310 315 320

Phe Leu Glu Lys Arg Arg Asn Arg Thr Tyr Ile Lys Ser Tyr Val Leu
325 330 335

Leu Asp Lys His Glu Lys Phe Lys Ile Glu Arg Glu Asn Lys Lys Asp
340 345 350

Lys Ile Val Lys Phe Phe Val Glu Asn Ile Lys Asn Asn Ser Ile Lys
355 360 365

Glu Lys Ile Glu Lys Ile Leu Ala Glu Phe Lys Ile Asp Glu Leu Ile
370 375 380

Lys Lys Leu Glu Lys Glu Leu Lys Lys Gly Asn Cys Asp Thr Glu Ile
385 390 395 400

Phe Gly Ile Phe Lys Lys His Tyr Lys Val Asn Phe Asp Ser Lys Lys
405 410 415

Phe Ser Lys Lys Ser Asp Glu Glu Lys Glu Leu Tyr Lys Ile Ile Tyr
420 425 430

Arg Tyr Leu Lys Gly Arg Ile Glu Lys Ile Leu Val Asn Glu Gln Lys
435 440 445

Val Arg Leu Lys Lys Met Glu Lys Ile Glu Ile Glu Lys Ile Leu Asn
450 455 460

Glu Ser Ile Leu Ser Glu Lys Ile Leu Lys Arg Val Lys Gln Tyr Thr
465 470 475 480

Leu Glu His Ile Met Tyr Leu Gly Lys Leu Arg His Asn Asp Ile Asp
485 490 495

Met Thr Thr Val Asn Thr Asp Asp Phe Ser Arg Leu His Ala Lys Glu
500 505 510

Glu Leu Asp Leu Glu Leu Ile Thr Phe Phe Ala Ser Thr Asn Met Glu

515 520 525

Leu Asn Lys Ile Phe Ser Arg Glu Asn Ile Asn Asn Asp Glu Asn Ile
530 535 540

Asp Phe Phe Gly Gly Asp Arg Glu Lys Asn Tyr Val Leu Asp Lys Lys
545 550 555 560

Ile Leu Asn Ser Lys Ile Lys Ile Ile Arg Asp Leu Asp Phe Ile Asp
565 570 575

Asn Lys Asn Asn Ile Thr Asn Asn Phe Ile Arg Lys Phe Thr Lys Ile
580 585 590

Gly Thr Asn Glu Arg Asn Arg Ile Leu His Ala Ile Ser Lys Glu Arg
595 600 605

Asp Leu Gln Gly Thr Gln Asp Asp Tyr Asn Lys Val Ile Asn Ile Ile
610 615 620

Gln Asn Leu Lys Ile Ser Asp Glu Glu Val Ser Lys Ala Leu Asn Leu
625 630 635 640

Asp Val Val Phe Lys Asp Lys Lys Asn Ile Ile Thr Lys Ile Asn Asp
645 650 655

Ile Lys Ile Ser Glu Glu Asn Asn Asn Asp Ile Lys Tyr Leu Pro Ser
660 665 670

Phe Ser Lys Val Leu Pro Glu Ile Leu Asn Leu Tyr Arg Asn Asn Pro
675 680 685

Lys Asn Glu Pro Phe Asp Thr Ile Glu Thr Glu Lys Ile Val Leu Asn
690 695 700

Ala Leu Ile Tyr Val Asn Lys Glu Leu Tyr Lys Lys Leu Ile Leu Glu
705 710 715 720

Asp Asp Leu Glu Glu Asn Glu Ser Lys Asn Ile Phe Leu Gln Glu Leu
725 730 735

Lys Lys Thr Leu Gly Asn Ile Asp Glu Ile Asp Glu Asn Ile Ile Glu
740 745 750

Asn Tyr Tyr Lys Asn Ala Gln Ile Ser Ala Ser Lys Gly Asn Asn Lys
755 760 765

Ala Ile Lys Lys Tyr Gln Lys Lys Val Ile Glu Cys Tyr Ile Gly Tyr
770 775 780

Leu Arg Lys Asn Tyr Glu Glu Leu Phe Asp Phe Ser Asp Phe Lys Met
785 790 795 800

Asn Ile Gln Glu Ile Lys Lys Gln Ile Lys Asp Ile Asn Asp Asn Lys
805 810 815

Thr Tyr Glu Arg Ile Thr Val Lys Thr Ser Asp Lys Thr Ile Val Ile
820 825 830

Asn Asp Asp Phe Glu Tyr Ile Ile Ser Ile Phe Ala Leu Leu Asn Ser
835 840 845

Asn Ala Val Ile Asn Lys Ile Arg Asn Arg Phe Phe Ala Thr Ser Val

```
    850                 855                 860

Trp Leu Asn Thr Ser Glu Tyr Gln Asn Ile Ile Asp Ile Leu Asp Glu
865                 870                 875                 880

Ile Met Gln Leu Asn Thr Leu Arg Asn Glu Cys Ile Thr Glu Asn Trp
                885                 890                 895

Asn Leu Asn Leu Glu Glu Phe Ile Gln Lys Met Lys Glu Ile Glu Lys
            900                 905                 910

Asp Phe Asp Asp Phe Lys Ile Gln Thr Lys Lys Glu Ile Phe Asn Asn
        915                 920                 925

Tyr Tyr Glu Asp Ile Lys Asn Asn Ile Leu Thr Glu Phe Lys Asp Asp
    930                 935                 940

Ile Asn Gly Cys Asp Val Leu Glu Lys Lys Leu Glu Lys Ile Val Ile
945                 950                 955                 960

Phe Asp Asp Glu Thr Lys Phe Glu Ile Asp Lys Lys Ser Asn Ile Leu
                965                 970                 975

Gln Asp Glu Gln Arg Lys Leu Ser Asn Ile Asn Lys Lys Asp Leu Lys
            980                 985                 990

Lys Lys Val Asp Gln Tyr Ile Lys Asp Lys Asp Gln Glu Ile Lys Ser
        995                 1000                1005

Lys Ile  Leu Cys Arg Ile Ile  Phe Asn Ser Asp Phe  Leu Lys Lys
    1010                1015                1020

Tyr Lys  Lys Glu Ile Asp Asn  Leu Ile Glu Asp Met  Glu Ser Glu
    1025                1030                1035

Asn Glu  Asn Lys Phe Gln Glu  Ile Tyr Tyr Pro Lys  Glu Arg Lys
    1040                1045                1050

Asn Glu  Leu Tyr Ile Tyr Lys  Lys Asn Leu Phe Leu  Asn Ile Gly
    1055                1060                1065

Asn Pro  Asn Phe Asp Lys Ile  Tyr Gly Leu Ile Ser  Asn Asp Ile
    1070                1075                1080

Lys Met  Ala Asp Ala Lys Phe  Leu Phe Asn Ile Asp  Gly Lys Asn
    1085                1090                1095

Ile Arg  Lys Asn Lys Ile Ser  Glu Ile Asp Ala Ile  Leu Lys Asn
    1100                1105                1110

Leu Asn  Asp Lys Leu Asn Gly  Tyr Ser Lys Glu Tyr  Lys Glu Lys
    1115                1120                1125

Tyr Ile  Lys Lys Leu Lys Glu  Asn Asp Asp Phe Phe  Ala Lys Asn
    1130                1135                1140

Ile Gln  Asn Lys Asn Tyr Lys  Ser Phe Glu Lys Asp  Tyr Asn Arg
    1145                1150                1155

Val Ser  Glu Tyr Lys Lys Ile  Arg Asp Leu Val Glu  Phe Asn Tyr
    1160                1165                1170

Leu Asn  Lys Ile Glu Ser Tyr  Leu Ile Asp Ile Asn  Trp Lys Leu
```

1175 1180 1185

Ala Ile Gln Met Ala Arg Phe Glu Arg Asp Met His Tyr Ile Val
1190 1195 1200

Asn Gly Leu Arg Glu Leu Gly Ile Ile Lys Leu Ser Gly Tyr Asn
1205 1210 1215

Thr Gly Ile Ser Arg Ala Tyr Pro Lys Arg Asn Gly Ser Asp Gly
1220 1225 1230

Phe Tyr Thr Thr Thr Ala Tyr Tyr Lys Phe Phe Asp Glu Glu Ser
1235 1240 1245

Tyr Lys Lys Phe Glu Lys Ile Cys Tyr Gly Phe Gly Ile Asp Leu
1250 1255 1260

Ser Glu Asn Ser Glu Ile Asn Lys Pro Glu Asn Glu Ser Ile Arg
1265 1270 1275

Asn Tyr Ile Ser His Phe Tyr Ile Val Arg Asn Pro Phe Ala Asp
1280 1285 1290

Tyr Ser Ile Ala Glu Gln Ile Asp Arg Val Ser Asn Leu Leu Ser
1295 1300 1305

Tyr Ser Thr Arg Tyr Asn Asn Ser Thr Tyr Ala Ser Val Phe Glu
1310 1315 1320

Val Phe Lys Lys Asp Val Asn Leu Asp Tyr Asp Glu Leu Lys Lys
1325 1330 1335

Lys Phe Lys Leu Ile Gly Asn Asn Asp Ile Leu Glu Arg Leu Met
1340 1345 1350

Lys Pro Lys Lys Val Ser Val Leu Glu Leu Glu Ser Tyr Asn Ser
1355 1360 1365

Asp Tyr Ile Lys Asn Leu Ile Ile Glu Leu Leu Thr Lys Ile Glu
1370 1375 1380

Asn Thr Asn Asp Thr Leu
1385

<210> 60
<211> 1300
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 60
Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
20 25 30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
35 40 45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
50 55 60

```
Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65                  70                  75                  80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
                85                  90                  95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
            100                 105                 110

Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
        115                 120                 125

Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
    130                 135                 140

Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145                 150                 155                 160

Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
                165                 170                 175

Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
            180                 185                 190

Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
        195                 200                 205

Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
    210                 215                 220

Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225                 230                 235                 240

Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
                245                 250                 255

Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
            260                 265                 270

Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
        275                 280                 285

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
    290                 295                 300

Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305                 310                 315                 320

Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
                325                 330                 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
            340                 345                 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
        355                 360                 365

Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
    370                 375                 380

Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385                 390                 395                 400
```

```
Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
                405                 410                 415

Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
            420                 425                 430

Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
        435                 440                 445

Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
    450                 455                 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465                 470                 475                 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
                485                 490                 495

Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
            500                 505                 510

Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
        515                 520                 525

Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
    530                 535                 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545                 550                 555                 560

Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
                565                 570                 575

Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
            580                 585                 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
        595                 600                 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
    610                 615                 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625                 630                 635                 640

Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
                645                 650                 655

Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
            660                 665                 670

Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
        675                 680                 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln
    690                 695                 700

Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
705                 710                 715                 720

Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp
                725                 730                 735
```

Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
740 745 750

Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
755 760 765

Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
770 775 780

Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
785 790 795 800

Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
805 810 815

Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
820 825 830

Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
835 840 845

Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu
850 855 860

Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
865 870 875 880

His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
885 890 895

Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
900 905 910

Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
915 920 925

Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
930 935 940

Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945 950 955 960

Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
965 970 975

Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
980 985 990

Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu
995 1000 1005

Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val
1010 1015 1020

Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu
1025 1030 1035

Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg
1040 1045 1050

Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly
1055 1060 1065

Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser
1070 1075 1080

Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys
1085 1090 1095

Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp
1100 1105 1110

Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe
1115 1120 1125

Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr
1130 1135 1140

Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp
1145 1150 1155

Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu
1160 1165 1170

Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly
1175 1180 1185

Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe
1190 1195 1200

Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg
1205 1210 1215

Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val
1220 1225 1230

Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1235 1240 1245

Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly
1250 1255 1260

Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu
1265 1270 1275

Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1280 1285 1290

Phe Val Gln Asn Arg Asn Asn
1295 1300

<210> 61
<211> 7403
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 61
tatccggtcg aatcgagaat gacgaccgct acgtcttgga ctacgaagcc gtggcccttg 60

ccgatgctct cggtgtggat gttgccgacc tgttccgcaa gatcgattgc cccaagaacc 120

tgctgcgcag gcgggcaggg taggggagcg gtttccggcg gagattttcg gaggcgccgg 180

taacgttatg tcggggaatt tgctatacat cgacgataat tagttttgtt gattcaggat 240

cgaaatgcgc tcaaacaaag aacgttccgc gtttccctca tgcgctacta cgcccacacc 300

gccatctttc ggcacgcaaa caaagcagat gggttgcctg tcaatgggtg atcattgcct 360

gaagttacca tccatcaata atataaatca tccttactcc gaatgtccct caatcgcatc 420

tatcaaggcc gcgtggcggc cgtcgaaaca ggaacggcct tagcgaaagg taatgtcgaa 480

tggatgcctg ccgcaggagg cgacgaagtt ctctggcagc accacgaact tttccaagct 540

gccatcaact actatctcgt cgccctgctc gcactcgccg acaaaaacaa tcccgtactt 600

ggcccgctga tcagccagat ggataatccc caaagcccтt accatgtctg gggaagtttc 660

cgccgccaag gacgtcagcg cacaggtctc agtcaagccg ttgcacctta tatcacgccg 720

ggcaataacg ctcccaccct tgacgaagtt ttccgctcca ttcttgcggg caacccaacc 780

gaccgcgcaa ctttggacgc tgcactcatg caattgctca aggcttgtga cggcgcgggc 840

gctatccagc aggaaggtcg ttcctactgg cccaaattct gcgatcctga ctccactgcc 900

aacttcgcgg gagatccggc catgctccgg cgtgaacaac accgcctcct ccttccgcaa 960

gttctccacg atccggcgat tactcacgac agtcctgccc ttggctcgtt cgacacttat 1020

tcgattgcta cccccgacac cagaactcct caactcaccg gccccaaggc acgcgcccgt 1080

cttgagcagg cgatcaccct ctggcgcgtc cgtcttcccg aatcggctgc tgacttcgat 1140

cgccttgcca gttccctcaa aaaaattccg gacgacgatt ctcgccttaa ccttcagggc 1200

tacgtcggca gcagtgcgaa aggcgaagtt caggcccgtc ttttcgccct tctgctattc 1260

cgtcacctgg agcgttcctc ctttacgctt ggccttctcc gttccgccac cccgccgccc 1320

aagaacgctg aaacacctcc tcccgccggc gttcctttac ctgcggcgtc cgcagccgat 1380

ccggtgcgga tagcccgtgg caaacgcagt tttgtttttc gcgcattcac cagtctcccc 1440

tgctggcatg gcggtgataa catccatccc acctggaagt cattcgacat cgcagcgttc 1500

aaatatgccc tcacggtcat caaccagatc gaggaaaaga cgaaagaacg ccaaaaagaa 1560

tgtgcggaac ttgaaactga tttcgactac atgcacggac ggctcgccaa gattccggta 1620

aaatacacga ccggcgaagc cgaaccgccc cccattctcg caaacgatct ccgcatcccc 1680

ctcctccgcg aacttctcca gaatatcaag gtcgacaccg cactcaccga tggcgaagcc 1740

gtctcctatg gtctccaacg ccgcaccatt cgcggtttcc gcgagctgcg ccgcatctgg 1800

cgcggccatg cccccgctgg cacggtcttt tccagcgagt tgaaagaaaa actagccggc 1860

gaactccgcc agttccagac cgacaactcc accaccatcg gcagcgtcca actcttcaac 1920

gaactcatcc aaaacccgaa atactggccc atctggcagg ctcctgacgt cgaaaccgcc 1980

cgccaatggg ccgatgccgg ttttgccgac gatccgctcg ccgcccttgt gcaagaagcc 2040

gaactccagg aagacatcga cgccctcaag gctccagtca aactcactcc ggccgatcct 2100

gagtattcaa gaaggcaata cgatttcaat gccgtcagca aattcggggc cggctcccgc 2160

tccgccaatc gccacgaacc cgggcagacg gagcgcggcc acaacacctt taccaccgaa 2220

atcgccgccc gtaacgcggc ggacgggaac cgctggcggg caacccacgt ccgcatccat 2280

tactccgctc cccgccttct tcgtgacgga ctccgccgac ctgacaccga cggcaacgaa 2340

gccctggaag ccgtcccttg gctccagccc atgatggaag ccctcgcccc tctcccgacg 2400

cttccgcaag acctcacagg catgccggtc ttcctcatgc ccgacgtcac cctttccggt 2460

gagcgtcgca tcctcctcaa tcttcctgtc accctcgaac cagccgctct tgtcgaacaa 2520

ctgggcaacg ccggtcgctg gcaaaaccag ttcttcggct cccgcgaaga tccattcgct 2580

ctccgatggc ccgccgacgg tgctgtaaaa accgccaagg ggaaaaccca cataccttgg 2640

caccaggacc gcgatcactt caccgtactc ggcgtggatc tcggcacgcg cgatgccggg 2700

gcgctcgctc ttctcaacgt cactgcgcaa aaaccggcca agccggtcca ccgcatcatt 2760

ggtgaggccg acggacgcac ctggtatgcc agccttgccg acgctcgcat gatccgcctg 2820

cccggggagg atgcccggct ctttgtccgg ggaaaactcg ttcaggaacc ctatggtgaa 2880
cgcgggcgaa acgcgtctct tctcgaatgg gaagacgccc gcaatatcat ccttcgcctt 2940
ggccaaaatc ccgacgaact cctcggcgcc gatccccggc gccattcgta tccggaaata 3000
aacgataaac ttctcgtcgc ccttcgccgc gctcaggccc gtcttgcccg tctccagaac 3060
cggagctggc ggttgcgcga ccttgcagaa tcggacaagg cccttgatga aatccatgcc 3120
gagcgtgccg gggagaagcc ttctccgctt ccgcccttgg ctcgcgacga tgccatcaaa 3180
agcaccgacg aagccctcct ttcccagcgt gacatcatcc ggcgatcctt cgttcagatc 3240
gccaacttga tccttcccct tcgcggacgc cgatgggaat ggcggcccca tgtcgaggtc 3300
ccggattgcc acatccttgc gcagagcgat cccggtacgg atgacaccaa gcgtcttgtc 3360
gccggacaac gcggcatctc tcacgagcgt atcgagcaaa tcgaagaact ccgtcgtcgc 3420
tgccaatccc tcaaccgtgc cctgcgtcac aaacccggag agcgtcccgt gctcggacgc 3480
cccgccaagg gcgaggaaat cgccgatccc tgtcccgcgc tcctcgaaaa gatcaaccgt 3540
ctccgggacc agcgcgttga ccaaaccgcg catgccatcc tcgccgccgc tctcggtgtt 3600
cgactccgcg ccccctcaaa agaccgcgcc gaacgccgcc atcgcgacat ccatggcgaa 3660
tacgaacgct ttcgtgcgcc cgctgatttt gtcgtcatcg aaaacctctc ccgttatctc 3720
agctcgcagg atcgtgctcg tagtgaaaac acccgtctca tgcagtggtg ccatcgccag 3780
atcgtgcaaa aactccgtca gctctgcgag acctacggca tccccgtcct cgccgtcccg 3840
gcggcctact catcgcgttt ttcttcccgg gacggctcgg ccggattccg ggccgtccat 3900
ctgacaccgg accaccgtca ccggatgcca tggagccgca tcctcgcccg cctcaaggcc 3960
cacgaggaag acggaaaaag actcgaaaag acggtgctcg acgaggctcg cgccgtccgg 4020
ggactctttg accggctcga ccggttcaac gccgggcatg tcccgggaaa accttggcgc 4080
acgctcctcg cgccgctccc cggcggccct gtgtttgtcc ccctcgggga cgccacaccc 4140
atgcaggccg atctgaacgc cgccatcaac atcgccctcc ggggcatcgc ggctcccgac 4200
cgccacgaca tccatcaccg gctccgtgcc gaaaacaaaa aacgcatcct gagcttgcgt 4260
ctcggcactc agcgcgagaa agcccgctgg cctggaggag ctccggcggt gacactctcc 4320
actccgaaca acggcgcctc tcccgaagat tccgatgcgt tgcccgaacg ggtatccaac 4380
ctgtttgtgg acatcgccgg tgtcgccaac ttcgagcgag tcacgatcga aggagtctcg 4440
caaaaattcg ccaccgggcg tggcctttgg gcctccgtca agcaacgtgc atggaaccgc 4500
gttgccagac tcaacgagac agtaacagat aacaacagga acgaagagga ggacgacatt 4560
ccgatgtaac cattgcttca ttacatctga gtctcccctc aatccctctg ccccatgcgt 4620
gatataacct ccacctcatg tcccggatcg gcgccggcaa cctgtagttc ccttccatcc 4680
tccaacactc ccgcagatcg cgatccgctg ccgccgatgc cggtgcgccg ccttcacaac 4740
tatctctact gtccgcggct tttttatctc cagtgggtcg agaatctctt tgaggaaaat 4800
gccgacacca ttgccggcag cgccgtgcat cgtcacgccg acaaacctac gcgttacgat 4860
gatgaaaaag ccgaggcact tcgcactggt ctccctgaag gcgcgcacat acgcagcctt 4920
cgcctggaaa acgcccaact cggtctcgtt ggcgtggtgg atatcgtgga gggaggcccc 4980
gacggactcg aactcgtcga ctacaaaaaa ggttccgcct tccgcctcga cgacggcacg 5040
ctcgctccca aggaaaacga caccgtgcaa cttgccgcct acgctcttct cctggctgcc 5100
gatggtgcgc gcgttgcgcc catggcgacg gtctattacg ctgccgatcg ccggcgtgtc 5160
accttcccgc tcgatgacgc cctctacgcc cgcacccgtt ccgccctcga agaggcccgc 5220
gccgttgcaa cctcggggcg catacctccg ccgctcgtct ctgacgtccg ctgcctccat 5280
tgttcctcct atgcgctttg ccttccccgc gagtccgcct ggtggtgccg ccatcgcagc 5340

acgccgcggg gagccggcca cacccccatg ttgccgggct ttgaggatga cgccgccgcc 5400

attcaccaaa tctccgaacc tgacaccgag ccaccacccg atcttgccag ccagcctccc 5460

cgtcccccgc ggctcgatgg agaattgttg gttgtccaga ctccgggagc gatgatcgga 5520

caaagcggcg gtgagtttac cgtgtccgtc aagggtgagg ttttgcgcaa gcttccggtt 5580

catcaactcc gggccattta cgtttacgga gccgtgcaac tcacggcgca tgctgtgcag 5640

accgcccttg aggaggatat cgacgtctcc tattttgcgc ccagcggccg ctttcttggc 5700

ctcctccgcg gcctgcccgc atccggcgtg gatgcgcgtc tcgggcaata caccctgttt 5760

cgcgaaccct ttggccgtct ccgtctcgcc tgcgaggcga ttcgggccaa gatccataac 5820

cagcgcgtcc tcctcatgcg taacggcgag cccggggagg gcgtcttgcg cgaactcgcc 5880

cgtctgcgcg acgccaccag tgaggcgact tcgctcgacg aactcctcgg catcgagggc 5940

atcgccgcgc atttctattt ccagtatttt cccaccatgc tgaaagaacg ggcggcctgg 6000

gcctttgatt tttccggacg caatcgccgc ccgccgcgcg acccggtcaa cgccctgctt 6060

tcgttcggtt acagcgtgtt gtccaaggaa cttgccggcg tctgccacgc tgttggccta 6120

gacccgtttt tcggcttcat gcaccagccg cgttacgggc gccccgcact cgctctcgat 6180

ctgatggagg agtttcgccc tctcatcgcc gacagtgttg ccctgaatct catcaaccgt 6240

ggcgaactcg acgaagggga ctttatccgg tcggccaatg gcaccgcgct caatgatcgg 6300

ggccgccggc gtttttggga ggcatggttc cggcgtctcg acagcgaagt cagccatcct 6360

gaatttggtt acaagatgag ctatcgacgg atgcttgaag tgcaggcgcg ccagctatgg 6420

cgctatgtgc gcggtgacgc cttccgctac cacggattca ccacccgttg attccgatgt 6480

cagatccccg ccgccgttat cttgtgtgtt acgacatcgc caatccgaag cgattgcgcc 6540

aagtggccaa gctgctggag agctatggca cgcgtctgca atactcggtt ttcgaatgtc 6600

ctttggacga tcttcgtctt gaacaggcga aggctgattt gcgcgacacg attaatgccg 6660

accaagacca ggtgttattt gtttcgcttg gccccgaagc caacgatgcc acgttgatca 6720

tcgccacgct tgggctccct tataccgtgc gctcgcgagt gacgattatc tgacccataa 6780

cccacgtgtt gaagaggctg aaaacagacg gacctctatg aagaacaatt gacgttttgg 6840

ccgaactcag cagaccttta tgcggctaag gccaatgatc atccatccta ccgccattgg 6900

gctggagacg ttttttgaaa cggcgagtgc tgcggatagc gagtttctct tggggaggcg 6960

ctcgcggcca cttttacaga ggagatgttc gggcgaactg gccgacctaa caaggcgtac 7020

ccggctcaaa atcgaggcac gctcgcacgg gatgatgtaa ttcgttgttt ttcagcatac 7080

cgtgcgagca cgggccgcag cgaatgccgt ttcacgaatc gtcaggcggc ggggagaagt 7140

catttaataa ggccactgtt aaaagccgca gcgaatgccg tttcacgaat cgtcaggcgg 7200

gcagtggatg tttttccatg aggcgaagaa tttcatcgcc gcagtgaatg ccgtttcacc 7260

attgatgaag aatgcgaggt gaaaacagag aaattgggtc aactctatca ctcttattca 7320

gccatcgttt caagaaagga tacctcgtat tggatacaac acagctcgtt cgttctctct 7380

acctccctcg acaatctcaa gga 7403

<210> 62
<211> 6789
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 62
taataaaatt gaaatatcac tatggattat tgtaatatta ccataaagat aggtgacgtt 60

ttttgaaaa ttgtaaacct aatttgaaga aaaccaatta aaaatcgctt cggctttttt 120

ttaagtgcca ggtagcattg atgctaaccc atgtgtaata aaggtttgtt ttccttcggg 180

gcacgaacac attataaggg aaacctaaag attccctttc ttgtttaata ttataaccag 240

tgaaaataag aataatgcac ctaaaactaa tatacagaaa ataagaatta aaagtactaa 300

tatatacatc atatgttatc ctccaatgct ttatttttta ataattgatg ttagtattag 360

ttttatttta atttctaaac ataagaattt gaaaaggatg tgtttattat ggcgacacgc 420

agttttattt taaaaattga accaaatgaa gaagttaaaa agggattatg gaagacgcat 480

gaggtattga atcatggaat tgcctactac atgaatattc tgaaactaat tagacaggaa 540

gctatttatg aacatcatga acaagatcct aaaaatccga aaaaagtttc aaaagcagaa 600

atacaagccg agttatggga ttttgtttta aaaatgcaaa aatgtaatag ttttacacat 660

gaagttgaca aagatgttgt ttttaacatc ctgcgtgaac tatatgaaga gttggtccct 720

agttcagtcg agaaaaaggg tgaagccaat caattatcga ataagtttct gtacccgcta 780

gttgatccga acagtcaaag tgggaaaggg acggcatcat ccggacgtaa acctcggtgg 840

tataatttaa aaatagcagg cgacccatcg tgggaggaag aaaagaaaaa atgggaagag 900

gataaaaaga aagatcccct tgctaaaatc ttaggtaagt tagcagaata tgggcttatt 960

ccgctattta ttccatttac tgacagcaac gaaccaattg taaaagaaat taaatggatg 1020

gaaaaaagtc gtaatcaaag tgtccggcga cttgataagg atatgtttat ccaagcatta 1080

gagcgttttc tttcatggga aagctggaac cttaaagtaa aggaagagta tgaaaaagtt 1140

gaaaaggaac acaaaacact agaggaaagg ataaaagagg acattcaagc atttaaatcc 1200

cttgaacaat atgaaaaaga acggcaggag caacttctta gagatacatt gaatacaaat 1260

gaataccgat taagcaaaag aggattacgt ggttggcgtg aaattatcca aaaatggcta 1320

aagatggatg aaaatgaacc atcagaaaaa tatttagaag tatttaaaga ttatcaacgg 1380

aaacatccac gagaagccgg ggactattct gtctatgaat ttttaagcaa gaaagaaaat 1440

cattttattt ggcgaaatca tcctgaatat ccttatttgt atgctacatt ttgtgaaatt 1500

gacaaaaaaa agaaagacgc taagcaacag gcaactttta ctttggctga cccgattaac 1560

catccgttat gggtacgatt tgaagaaaga agcggttcga acttaaacaa atatcgaatt 1620

ttaacagagc aattacacac tgaaaagtta aaaaagaaat taacagttca acttgatcgt 1680

ttaatttatc caactgaatc cggcggttgg gaggaaaaag gtaaagtaga tatcgttttg 1740

ttgccgtcaa gacaatttta taatcaaatc ttccttgata tagaagaaaa ggggaaacat 1800

gcttttactt ataaggatga aagtattaaa ttccccctta aaggtacact tggtggtgca 1860

agagtgcagt ttgaccgtga ccatttgcgg agatatccgc ataaagtaga atcaggaaat 1920

gttggacgga tttattttaa catgacagta aatattgaac caactgagag ccctgttagt 1980

aagtctttga aaatacatag ggacgatttc cccaagttcg ttaattttaa accgaaagag 2040

ctcaccgaat ggataaaaga tagtaaaggg aaaaaattaa aaagtggtat agaatccctt 2100

gaaattggtc tacgggtgat gagtatcgac ttaggtcaac gtcaagcggc tgctgcatcg 2160

atttttgaag tagttgatca gaaaccggat attgaaggga agttattttt tccaatcaaa 2220

ggaactgagc tttatgctgt tcaccgggca agttttaaca ttaaattacc gggtgaaaca 2280

ttagtaaaat cacgggaagt attgcggaaa gctcgggagg acaacttaaa attaatgaat 2340

caaaagttaa actttctaag aaatgttcta catttccaac agtttgaaga tatcacagaa 2400

agagagaagc gtgtaactaa atggatttct agacaagaaa atagtgatgt tcctcttgta 2460

tatcaagatg agctaattca aattcgtgaa ttaatgtata aaccctataa agattgggtt 2520

gccttttaa aacaactcca taaacggcta gaagtcgaga ttggcaaaga ggttaagcat 2580

tggcgaaaat cattaagtga cgggagaaaa ggtctttacg gaatctccct aaaaaatatt 2640

gatgaaattg atcgaacaag gaaattcctt ttaagatgga gcttacgtcc aacagaacct 2700

ggggaagtaa gacgcttgga accaggacag cgttttgcga ttgatcaatt aaaccaccta 2760

aatgcattaa aagaagatcg attaaaaaag atggcaaata cgattatcat gcatgcctta 2820

ggttactgtt atgatgtaag aaagaaaaag tggcaggcaa aaaatccagc atgtcaaatt 2880

attttatttg aagatttatc taactacaat ccttacgagg aaaggtcccg ttttgaaaac 2940

tcaaaactga tgaagtggtc acggagagaa attccacgac aagtcgcctt acaaggtgaa 3000

atttacggat tacaagttgg ggaagtaggt gcccaattca gttcaagatt ccatgcgaaa 3060

accgggtcgc cgggaattcg ttgcagtgtt gtaacgaaag aaaaattgca ggataatcgc 3120

ttttttaaaa atttacaaag agaaggacga cttactcttg ataaaatcgc agttttaaaa 3180

gaaggagact tatatccaga taaaggtgga gaaaagttta tttctttatc aaaggatcga 3240

aagttggtaa ctacgcatgc tgatattaac gcggcccaaa atttacagaa gcgtttttgg 3300

acaagaacac atggatttta taaagtttac tgcaaagcct atcaggttga tggacaaact 3360

gtttatattc cggagagcaa ggaccaaaaa caaaaaataa ttgaagaatt tggggaaggc 3420

tattttattt taaaagatgg tgtatatgaa tggggtaatg cggggaaact aaaaattaaa 3480

aaaggttcct ctaaacaatc atcgagtgaa ttagtagatt cggacatact gaaagattca 3540

tttgatttag caagtgaact taagggagag aaactcatgt tatatcgaga tccgagtgga 3600

aacgtatttc cttccgacaa gtggatggca gcaggagtat tttttggcaa attagaaaga 3660

atattgattt ctaagttaac aaatcaatac tcaatatcaa caatagaaga tgattcttca 3720

aaacaatcaa tgtaaaagtt tgcccgtata agaacttaat taattaggat ggtaggatgt 3780

tactaaatat gtctgtaggc atcattccta ctatccgttt tgtccgaata tcagagcatt 3840

aggtgaggaa tggtaagaaa ggaaaattta tatgaaccaa ccgattccta ttcgaatgtt 3900

aaatgaaata caatattgtg agcgactttt ttactttatg catgtccaaa agctatttga 3960

tgagaatgca gatacagttg aaggaagtgc acagcatgag cgggcagaaa gaagcaaaag 4020

accaagtaaa atgggaccaa aggaattatg gggtgaggcg ccaagaagtc ttaagcttgg 4080

tgatgagctg ttaaatatta ccggtgttct tgatgccata agtcatgaag agaacagttg 4140

gatcccggtt gaatcaaaac acagttccgc accggatgga ttgaaccctt ttaaagtaga 4200

tggctttcta cttgacgggt ctgcatggcc aaacgatcaa attcaacttt gtgcacaagg 4260

cttgctcttg aatgccaatg gatacccgtg tgattatggg tatttatttt atcgtggtaa 4320

taagaaaaag gtgaaaattt attttactga agatttaatc gctgccacaa agtactatat 4380

taaaaaagca cacgagatac tagtattatc tggtgatgaa tcagctattc ctaagccttt 4440

aattgattct aataagtgtt ttcgctgttc tttaaactat atctgtcttc cggatgaaac 4500

gaactatcta ttaggggcaa gttcaacaat tcgtaaaatt gtgccttcaa ggacagatgg 4560

tggcgtttta tatgtatcag agtctggtac aaaattagga aaatcgggtg aggagttaat 4620

cattcagtat aaagatggcc aaaagcaggg tgttcctata aaagatatta ttcaagtttc 4680

gttaattgga aatgttcaat gctcaacgca attacttcat tttttaatgc aatcaaatat 4740

tcctgtaagt tatttatcat cccacggtcg tttgattggt gtcagttcat ctttagttac 4800

aaaaaatgtt ttaacaaggc agcaacagtt cattaaattt acaaatcctg agtttggact 4860

aaatctagca aaacaaattg tttatgccaa gattcgaaat caacgaactt tacttagaag 4920

aaatgggggg agtgaggtaa aggagatttt aacagattta aaatctttaa gtgacagtgc 4980

actgaacgca atatcaatag aacaattacg gggtattgaa gggatttctg caaaacatta 5040

tttcgcagga tttccgttta tgttgaaaaa tgaattacgt gaattgaatt taatgaaagg 5100

gcgtaatagg agaccgccaa aagatcctgt aaatgtactt ctttctcttg gttatacttt 5160

attgacacgt gatattcatg ctgcgtgtgg ttcagtcgga ttggatccga tgtttggttg 5220

ttaccatcgt ccagaagcag gtcgaccggc tctagtatta gatgttatgg aaacatttcg 5280

accacttatt gtagacagta ttgtcatccg agctttgaat acgggtgaaa tctcattaaa 5340

agatttttat ataggaaaag atagttgtca attattaaaa catggccgcg attccttttt 5400

tgccatttat gaaagaagaa tgcatgaaac tattaccgat ccaattttcg gctataagat 5460

tagctatcgc cgtatgctcg atttgcacat tcgaatgctt gcaaggttta ttgaagggga 5520

actgccggaa tataaaccat taatgacccg gtgagtttgt ttattaggtt aaaagaaggt 5580

gaagacatgc agcaatacgt ccttgtttct tatgatattt cggaccaaaa aagatggaga 5640

aaagtattta aactgatgaa aggatacgga gaacatgttc aatattccgt attcatatgc 5700

cagttaactg aattacagaa ggcaaaatta caagcctctt tagaagacat tatccatcat 5760

aagaatgacc aagtaatgtt tgttcacatc gggccagtga aagatggtca actatctaaa 5820

aaaatctcaa caattgggaa agaatttgtt ccattggatt taaagcggct tatattttga 5880

aaagatatag caaagaaatc ttatgaaaaa aatacaaaaa tatattgtta aaaaataggg 5940

aatattatat aatggactta cgaggttctg tcttttggtc aggacaaccg tctagctata 6000

agtgctgcag gggtgtgaga aactcctatt gctggacgat gtctctttta tttctttttt 6060

cttggatctg agtacgagca cccacattgg acatttcgca tggtgggtgc tcgtactata 6120

ggtaaaacaa acctttttaa gaagaataca aaaataacca caatattttt taaaaggaat 6180

tttgatggat ttacataacc tctcgcaaca tgcttctaaa acccaagccc accatagccc 6240

aaaaccccct gcggtccaag aaaaaagaaa tgatacgagg cattagcacc ggggagaagt 6300

catttaataa ggccactgtt aaaagtccaa gaaaaaagaa atgatacgag gcattagcac 6360

aacaatataa acgactactt taccgtgttc aagaaaaaag aaatgatatg aggcattagc 6420

acgatgggat gggagagaga ggacagttct actcttgctg tatccagctt cttttacttt 6480

atccggtatc atttcttcac ttctttctgc acataaaaaa gcacctaact atttggataa 6540

gttaagtgct tttatttccg tttgaagttg tctattgctt ttttcttcat atcttcaaat 6600

tttttctgtt tctcagagtc aactttacca actgtaatcc cttttctttt tggcattggg 6660

gtatctttcc accttagtgt gttcataagg cttatattta tcactcattg tattcctcca 6720

acacaattat aatttttccg tcatcctcaa tccaaccgtc aactgtgaca aaagacgaat 6780

ctctcttat 6789

<210> 63
<211> 6214
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 63
gtttcatttg gaaagggaga gcattggctt ttctctttgt aaataaagtg caagctttgt 60

aataagcttc tagtggagaa gtgattgttt gaatcaccca atgcacacgc actaaagtta 120

gacgaaccta taattcgtat tagtaagtat agtacatgaa gaaaaatgca acaagcattt 180

actctctttt aaataaagaa ttgatagctg ttaatattga tagtatatta taccttatag 240

atgttcgatt ttttttgaaa ttcaaaaatc atacttagta aagaaaggaa ataacgtcat 300

ggacaagcga aagcgtagaa gttacgagtt taggtgggaa gcgggaggca ccagtcatgg 360

caatccgtag cataaaacta aaactaaaaa cccacacagg cccggaagcg caaaacctcc 420

gaaaaggaat atggcggacg catcggttgt taaatgaagg cgtcgcctat tacatgaaaa 480

tgctcctgct ctttcgtcag gaaagcactg gtgaacggcc aaaagaagaa ctacaggaag 540

aactgatttg tcacatacgc gaacagcaac aacgaaatca ggcagataaa aatacgcaag 600

cgcttccgct agataaggca ctggaagctt tgcgccaact atatgaactg cttgtcccct 660

cctcggtcgg acaaagtggc gacgcccaga tcatcagccg aaagtttctc agcccgctcg 720

tcgatccgaa cagcgaaggc ggcaaaggta cttcgaaggc aggggcaaaa cccacttggc 780

agaagaaaaa agaagcgaac gacccaacct gggaacagga ttacgaaaaa tggaaaaaaa 840

gacgcgagga agacccaacc gcttctgtga ttactacttt ggaggaatac ggcattagac 900

cgatctttcc cctgtacacg aacaccgtaa cagatatcgc gtggttgcca cttcaatcca 960

atcagtttgt gcgaacctgg gacagagaca tgcttcaaca agcgattgaa agactgctca 1020

gttgggagag ctggaacaaa cgtgtccagg aagagtatgc caagctgaaa gaaaaaatgg 1080

ctcaactgaa cgagcaactc gaaggcggtc aggaatggat cagcttgcta gagcagtacg 1140

aagaaaaccg agagcgagag cttagggaaa acatgaccgc tgccaatgac aagtatcgga 1200

ttaccaagcg gcaaatgaaa ggctggaacg agctgtacga gctatggtca acctttcccg 1260

ccagtgccag tcacgagcaa tacaaagagg cgctcaagcg tgtgcagcag cgactgagag 1320

ggcggtttgg ggatgctcat ttcttccagt atctgatgga agagaagaac cgcctgatct 1380

ggaaggggaa tccgcagcgt atccattatt ttgtcgcgcg caacgaactg acgaaacggc 1440

tggaggaagc caagcaaagc gccacgatga cgttgcccaa tgccaggaag catccattgt 1500

gggtgcgctt cgatgcacgg ggaggaaatt tgcaagacta ctacttgacg gctgaagcgg 1560

acaaaccgag aagcagacgt tttgtaacgt ttagtcagtt gatatggcca agcgaatcgg 1620

gatggatgga aaagaaagac gtcgaggtcg agctagcttt gtccaggcag ttttaccagc 1680

aggtgaagtt gctgaaaaat gacaaaggca agcagaaaat cgagttcaag gataaaggtt 1740

cgggctcgac gtttaacgga cacttggggg gagcaaagct acaactggag cggggcgatt 1800

tggagaagga agaaaaaaac ttcgaggacg gggaaatcgg cagcgtttac cttaacgttg 1860

tcattgattt cgaacctttg caagaagtga aaaatggccg cgtgcaggcg ccgtatggac 1920

aagtactgca actcattcgt cgccccaacg agtttcccaa ggtcactacc tataagtcgg 1980

agcaacttgt tgaatggata aaagcttcgc cacaacactc ggctggggtg gagtcgctgg 2040

catccggttt tcgtgtaatg agcatagacc ttgggctgcg cgcggctgca gcgacttcta 2100

ttttttctgt agaagagagt agcgataaaa atgcggctga tttttcctac tggattgaag 2160

gaacgccgct ggtcgctgtc catcagcgga gctatatgct caggttgcct ggtgaacagg 2220

tagaaaaaca ggtgatggaa aaacgggacg agcggttcca gctacaccaa cgtgtgaagt 2280

ttcaaatcag agtgctcgcc caaatcatgc gtatggcaaa taagcagtat ggagatcgct 2340

gggatgaact cgacagcctg aaacaagcgg ttgagcagaa aaagtcgccg ctcgatcaaa 2400

cagaccggac attttgggag gggattgtct gcgacttaac aaaggttttg cctcgaaacg 2460

aagcggactg ggaacaagcg gtagtgcaaa tacaccgaaa agcagaggaa tacgtcggaa 2520

aagccgttca ggcatggcgc aagcgctttg ctgctgacga gcgaaaaggc atcgcaggtc 2580

tgagcatgtg gaacatagaa gaattggagg gcttgcgcaa gctgttgatt tcctggagcc 2640

gcaggacgag gaatccgcag gaggttaatc gctttgagcg aggccatacc agccaccagc 2700

gtctgttgac ccatatccaa aacgtcaaag aggatcgcct gaagcagtta agtcacgcca 2760

ttgtcatgac tgccttgggg tatgtttacg acgagcggaa acaagagtgg tgcgccgaat 2820

acccggcttg ccaggtcatt ctgtttgaaa atctgagcca gtaccgttct aacctggatc 2880

gctcgaccaa agaaaactcc accttgatga agtgggcgca tcgcagcatt ccgaaatacg 2940

tccacatgca ggcggagcca tacgggattc agattggcga tgtccgggcg gaatattcct 3000

ctcgttttta cgccaagaca ggaacgccag gcattcgttg taaaaaggtg agaggccaag 3060

acctgcaggg cagacggttt gagaacttgc agaagaggtt agtcaacgag caatttttga 3120

cggaagaaca agtgaaacag ctaaggcccg gcgacattgt cccggatgat agcggagaac 3180

tgttcatgac cttgacagac ggaagcggaa gcaaggaggt cgtgtttctc caggccgata 3240

ttaacgcggc gcacaatctg caaaaacgtt tttggcagcg atacaatgaa ctgttcaagg 3300

ttagctgccg cgtcatcgtc cgagacgagg aagagtatct cgttcccaag acaaaatcgg 3360

tgcaggcaaa gctgggcaaa gggctttttg tgaaaaaatc ggatacagcc tggaaagatg 3420

tatatgtgtg ggacagccag gcaaagctta aaggtaaaac aacctttaca gaagagtctg 3480

agtcgcccga acaactggaa gactttcagg agatcatcga ggaagcagaa gaggcgaaag 3540

gaacataccg tacactgttc cgcgatccta gcggagtctt ttttcccgaa tccgtatggt 3600

atccccaaaa agatttttgg ggcgaggtga aaaggaagct gtacggaaaa ttgcgggaac 3660

ggtttttgac aaaggctcgg taagggtgtg caaggagagt gaatggcttg tcctggatac 3720

ctgtccgcat gctaaatgaa attcagtatt gtgagcgact gtaccatatt atgcatgtgc 3780

aggggctgtt tgaggaaagc gcagacacgg tcgaaggagc agcacaacac aagcgtgcag 3840

agacacatct gcgcaaaagc aaggcagcgc cggaagagat gtgggggggac gctccgttta 3900

gcttgcagct cggcgaccct gtgcttggca ttacgggaaa gctggatgcc gtctgtctgg 3960

aagaaggtaa gcagtggatt ccggtagaag gaaagcattc ggcgtcgcca gaaggcgggc 4020

agatgttcac tgtaggcgtg tattcgctgg acggttctgc ctggcccaac gaccaaatcc 4080

aattgtgtgc gcaaggcttg ctgcttcgcg cgaatggata tgaatccgat tatggctact 4140

tatactaccg tggcaataaa aagaaggttc gcattccttt ttcgcaggaa ctcatagcgg 4200

ctactcacgc ctgcattcaa aaagctcatc agcttcggga agccgaaatt ccccctccgt 4260

tgcaggagtc gaaaaagtgc tttcgatgct cgttaaatta cgtatgcatg cctgacgaga 4320

cgaattacat gttggggttg agcgcaaaca tcagaaagat tgtgcccagt cgtccagatg 4380

gcggggtact gtatgttaca gagcaggggg caaaactggg cagaagcgga gaaagcttga 4440

ccatcacctg ccggggcgaa aagatagacg aaatcccgat caaagacttg attcacgtga 4500

gcttgatggg gcatgtgcaa tgctctacgc agcttctgca caccttgatg aactgtggcg 4560

tccacgtcag ctacttgact acgcatggca cattgacagg aataatgact ccccctttat 4620

cgaaaaacat tcgaacaaga gccaagcagt ttatcaaatt tcagcacgcg gagatcgccc 4680

ttggaatcgc gagaagggtc gtgtatgcga aaatttccaa tcagcgcacg atgctgcgcc 4740

gcaatggctc accagataaa gcagttttaa aagagttaaa agagcttaga gatcgcgcgt 4800

gggaggcgcc atcactggaa atagtgagag gtatcgaggg acgtgcagca cagttgtaca 4860

tgcagttttt ccctaccatg ttaaagcacc cagtagtaga cggtatggcg atcatgaacg 4920

gtcgcaaccg tcgcccgccc aaagatccgg tcaatgcgct gctctccctc ggctatacgc 4980

ttctttcacg ggatgtttac tccgcatgtg ccaatgtcgg actcgatcca ctgttcggct 5040

ttttccatac gatggagccg ggcagaccag ctttggcact cgatctgatg gaaccgttcc 5100

gcgccttgat tgccgatagc gtagcgatac gtaccttgaa tacggaggaa ctcaccctcg 5160

gggacttta ttggggaaaa gacagttgtt atttgaaaaa ggcaggaaga caaacgtatt 5220

tcgctgccta tgaaagacgg atgaacgaga cgctgacgca tccgcaattt gggtataagc 5280

tcagctatcg ccgtatgctg gagctggaag caaggttttt ggcccggtat ctggatggag 5340

agctggtgga atatacgccg ctcatgacaa ggtaggaaat gaccatgcga caatttgttc 5400

tggtaagcta tgatattgcc gatcaaaaac gttggagaaa agtattcaag ctgatgaagg 5460

ggcaaggcga gcacgtccag tactcggtgt ttctgtgcca actcaccgag attcagcaag 5520

ccaagctaaa ggtaagcctg gcggagctgg ttcaccatgg agaagaccag gtcatgtttg 5580

taaaaatcgg cccagtgacg agagatcaac tggacaagcg gatatctact gttggcaggg 5640

agtttctgcc tcgcgatttg accaaattta tctattaagg aatgaagaaa gctagttgta 5700

acaaaagtgg aaaaagagta aaataaaggt gtcagtcgca cgctataggc cataagtcga 5760

cttacatatc cgtgcgtgtg cattatgggc ccatccacag gtctattccc acggataatc 5820

acgactttcc actaagcttt cgaattttat gatgcgagca tcctctcagg tcaaaaaagc 5880

cgggggatgc tcgaactctt tgtgggcgta ggctttccag agtttttag gggaagaggc 5940

agccgatgga taagaggaat ggcgattgaa ttttggcttg ctcgaaaaac gggtctgtaa 6000

ggcttgcggc tgtaggggtt gagtgggaag gagttcgaaa gcttagtgga aagcttcgtg 6060

gttagcaccg gggagaagtc atttaataag gccactgtta aaagttcgaa agcttagtgg 6120

aaagcttcgt ggttagcacg ctaaagtccg tctaaactac tgagatctta aatcggcgct 6180

caaataaaaa acctcgctaa tgcgaggttt cagc 6214

<210> 64
<211> 12338
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 64
gaagttatgt tgataaaatg gtttatgaaa acgtgagtct gtggtagtat tataaacaat 60

gatggaataa agtgtttttt gcgccgcacg gcatgaattc aggggttagc ttggttttgt 120

gtataaataa atgttctaca tatttatttt gtttttgcg ccgcaaaatg caactgaaag 180

ccgcatctag agcaccctgt agaagacagg gttttgagaa tagcccgaca tagagggcaa 240

tagacacggg gagaagtcat ttaataaggc cactgttaaa agttttgaga atagcccgac 300

atagagggca atagactttt gcttcgtcac ggatggactt cacaatggca acaacgtttt 360

gagaatagcc cgacatagtt atagagatgt ataaatataa ccgataaaca ttgactaatt 420

tgttgaagtc agtgtttatc ggtttttgt gtaaatatag gagttgttag aatgatactt 480

tttgcctaat tttggaactt tatgaggata taagatagac ttgataaaaa ggtaaaagaa 540

aggttaaaga gcatggcagg aatagtgacc tgtgatgaag atgatggtag aattaaaagt 600

gttcttaaag aaaaacaata ttggataagg aaaataattc aatagataaa aaatttaggg 660

ggaaaaatga aaatatcaaa agtcgatcat accagaatgg cggttgctaa aggtaatcaa 720

cacaggagag atgagattag tgggattctc tataaggatc cgacaaagac aggaagtata 780

gattttgatg aacgattcaa aaaactgaat tgttcggcga agatacttta tcatgtattc 840

aatggaattg ctgagggaag caataaatac aaaaatattg ttgataaagt aaataacaat 900

ttagataggg tcttatttac aggtaagagc tatgatcgaa aatctatcat agacatagat 960

actgttctta gaaatgttga gaaaattaat gcatttgatc gaatttcaac agaggaaaga 1020

gaacaaataa ttgacgattt gttagaaata caattgagga aggggttaag gaaaggaaaa 1080

gctggattaa gagaggtatt actaattggt gctggtgtaa tagttagaac cgataagaag 1140

caggaaatag ctgattttct ggagatttta gatgaagatt tcaataagac gaatcaggct 1200

aagaacataa aattgtctat tgagaatcag gggttggtgg tctcgcctgt atcaagggga 1260

gaggaacgga tttttgatgt cagtggcgca caaaagggaa aaagcagcaa aaaagcgcag 1320

gagaaagagg cactatctgc atttctgtta gattatgctg atcttgataa gaatgtcagg 1380

tttgagtatt tacgtaaaat tagaagactg ataaatctat atttctatgt caaaaatgat 1440

gatgttatgt ctttaactga aattccggca gaagtgaatc tggaaaaaga ttttgatatc 1500

tggagagatc acgaacaaag aaaggaagag aatggagatt ttgttggatg tccggacata 1560

cttttggcag atcgtgatgt gaagaaaagt aacagtaagc aggtaaaaat tgcagagagg 1620

caattaaggg agtcaatacg tgaaaaaaat ataaaacgat atagatttag cataaaaacg 1680

attgaaaagg atgatggaac atacttttttt gcaaataagc agataagtgt attttggatt 1740

catcgcattg aaaatgctgt agaacgtata ttaggatcta ttaatgataa aaaactgtat 1800

agattacgtt taggatatct aggagaaaaa gtatggaagg acatactcaa ttttctcagc 1860

ataaaataca ttgcagtagg caaggcagta ttcaattttg caatggatga tctgcaggag 1920

aaggatagag atatagaacc cggcaagata tcagaaaatg cagtaaatgg attgacttcg 1980

tttgattatg agcaaataaa ggcagatgag atgctgcaga gagaagttgc tgttaatgta 2040

gcattcgcag caaataatct tgctagagta actgtagata ttccgcaaaa tggagaaaaa 2100

gaggatatcc ttctttggaa taaaagtgac ataaaaaaat acaaaaagaa ttcaaagaaa 2160

ggtattctga aatctatact tcagtttttt ggtggtgctt caacttggaa tatgaaaatg 2220

tttgagattg catatcatga tcagccaggt gattacgaag aaaactacct atatgacatt 2280

attcagatca tttactcgct cagaaataag agctttcatt tcaagacata tgatcatggg 2340

gataagaatt ggaatagaga actgatagga aagatgattg agcatgatgc tgaaagagtc 2400

atttctgttg agagggaaaa gtttcattcc aataacctgc cgatgtttta taaagacgct 2460

gatctaaaga aaatattgga tctcttgtat agcgattatg caggacgtgc atctcaggtt 2520

ccggcattta acactgtctt ggttcgaaag aactttccgg aatttcttag gaaagatatg 2580

ggctacaagg ttcattttaa caatcctgaa gtagagaatc agtggcacag tgcggtgtat 2640

tacctatata aagagattta ttacaatcta tttttgagag ataaagaggt aaagaatctt 2700

ttttatactt cattaaaaaa tataagaagt gaagtttcgg acaaaaaaca aaagttagct 2760

tcagatgatt ttgcatccag gtgtgaagaa atagaggata gaagtcttcc ggaaatttgt 2820

cagataataa tgacagaata caatgcgcag aactttggta atagaaaagt taaatctcag 2880

cgtgttattg aaaaaaataa ggatattttc agacattata aaatgctttt gataaagact 2940

ttagcaggtg ctttttctct ttatttgaag caggaaagat ttgcatttat tggtaaggca 3000

acacctatac catacgaaac aaccgatgtt aagaattttt tgcctgaatg gaaatccgga 3060

atgtatgcat cgtttgtaga ggagataaag aataatcttg atcttcaaga atggtatatc 3120

gtcggacgat tccttaatgg gaggatgctc aatcaattgg caggaagcct gcggtcatac 3180

atacagtatg cggaagatat agaacgtcgt gctgcagaaa ataggaataa gcttttctcc 3240

aagcctgatg aaaagattga agcatgtaaa aaagcggtca gagtgcttga tttgtgtata 3300

aaaatttcaa ctagaatatc tgcggaattt actgactatt ttgatagtga agatgattat 3360

gcagattatc ttgaaaaata tctcaagtat caggatgatg ccattaagga attgtcagga 3420

tcttcgtatg ctgcgttgga tcatttttgc aacaaggatg atctgaaatt tgatatctat 3480

gtaaatgccg gacagaagcc tatcttacag agaaatatcg tgatggcaaa gctttttgga 3540

ccagataaca ttttgtctga agttatggaa aaggtaacag aaagtgccat acgagaatac 3600

tatgactatc tgaagaaagt ttcaggatat cgggtaaggg gaaaatgtag tacagagaaa 3660

gaacaggaag atctgctaaa gttccaaaga ttgaaaaacg cagtagaatt ccgggatgtt 3720

actgaatatg ctgaggttat taatgagctt ttaggacagt tgataagttg gtcatatctt 3780

agggagaggg atctattata tttccagctg ggattccatt acatgtgtct gaaaaacaaa 3840

tctttcaaac cggcagaata tgtggatatt cgtagaaata atggtacgat tatacataat 3900

gcgatacttt accagattgt ttcgatgtat attaatggac tggatttcta tagttgtgat 3960

aaagaaggga aaacgctcaa accaattgaa acaggaaagg gcgtaggaag taagatagga 4020

caatttataa agtattccca gtatttatac aatgatccgt catataagct tgagatctat 4080

aatgcaggat tagaagtttt tgaaaacatt gatgaacatg ataatattac agatcttaga 4140

aagtatgtgg atcattttaa gtattatgca tatggtaata aaatgagcct gcttgatctg 4200

tatagtgaat tcttcgatcg tttctttaca tatgatatga agtatcagaa gaatgtagtg 4260

aatgtgttgg agaatatcct tttaaggcat tttgtaattt tctatccgaa gtttggatca 4320

ggaaaaaaag atgttggaat tagggattgt aaaaaagaaa gagctcagat tgaaataagt 4380

gagcagagcc tcacatcgga agacttcatg tttaagcttg acgacaaagc aggagaagaa 4440

gcaaagaagt ttccggcaag ggatgaacgt tatctccaga caatagccaa gttgctctat 4500

tatcctaacg aaattgagga tatgaacaga ttcatgaaga aaggagaaac gataaataaa 4560

aaagttcagt ttaatagaaa aaagaagata accaggaaac aaaagaataa ttcatcaaac 4620

gaggtattgt cttcaactat gggttattta tttaagaaca ttaaattgta aaaaagattc 4680

gttgtagata attgataggt aaaagctgac cggagccttt ggctccggac agttgtatat 4740

aagaggatat taatgactga aaatgatttt tgttggaagt cagtttttc tgtggaaagc 4800

gaaatcgaat atgatgagta tgcatatggc agaagagctg tagaaggcga gaatacatat 4860

gattacatta ctaaggaaga aagaccggaa cttaatgacg aatatgtagc gagacgttgc 4920

attttcggta aaaaagcagg aaaaatatcc aggtcggatt ttagtaggat aagatctgcg 4980

ttggatcatg cgatgataaa taatacacat acagcatttg ccagatttat cactgaaaat 5040

ctgacgagac tcaatcacaa agaacatttt ctgaatgtga cacgtgcata ttctaaacct 5100

gattctgaaa aattgataca accgagatac tggcagtcgc ctgtagttcc aaaggataaa 5160

caaatatatt atagcaagaa tgcgattaaa aaatggtgtg gttacgaaga tgatattccg 5220

cctcgttctg tgatagttca gatgtgtcta ttgtggggga ctgatcatga agaggcagat 5280

catatccttc gcagttcagg atacgcggcg cttagtcctg ttgtacttcg agatcttatc 5340

tatatgtatt atctggatca tcaggatttg caaaaaaatg agttgatatg ggaagtaaaa 5400

aagcagttgg atcacttcga tttgacaaat agaaattatg atacaaatcc ttttgatgta 5460

gggggcagcg taaatgatca tatctgtgaa ctgagcgagc atatagcgaa ggctcattat 5520

atttatgaga gggctaagga aggaccattg caaaatgtaa ttcgggatat tttgggagat 5580

acacctgccc tttattctga aatggcattt cctcagctag catctataaa caggtgtgct 5640

tgcaattcgc tttcttcata tcaaaaaaat atttttgata ctgacatagc tatatatgca 5700

gatgaaaagg acacaagagg taaatcagac cgtatccttg ttgagggcgc atcttcgaaa 5760

tggtatgaat tgaagaaacg cgatgctaat aatgtcaaaa tttctgaaaa gctgagtata 5820

ctcaatacta ttcttaaatt taatagtgtt ttttgggaag aatgttacct tgatggaaat 5880

ataaaacaat cgagcggaaa gcgatctgag gcaggaaaaa ttctttatgg tcgcgacaac 5940

ggaaaagaaa atgtcggagt ttcaaaattg gaattggtgc ggtatatgat agctgcaggt 6000

caggaacaaa atctgggaaa ttacctggtg agttcaggat tttggagaaa aaatcatatg 6060

ctgtcattta tacaaggcaa tgatatagcg cttgatgaga tggatgaatt ggatctctta 6120

gactatattc tgatatatgc atggggattt agggaaaata tcattaaaaa gaacagtaat 6180

gtgaattctt tggatgaaaa gactagaaaa gtgcagtttc cgtttataaa gttactcatg 6240

gcaattgcaa gagatatcca gatacttata tgttcagcac atgaaaaaac agtcgatgag 6300

tcatctcgaa atgcagcaaa gaagatagat atattgggaa attatattcc ttttcagatt 6360

catcttcaga gaactaaaaa agatggtgga agagtggtaa tggatacatt gtgtgctgat 6420

tggattgcgg attatgaatg gtacattgat cttgagaaag gaacacttgg atgagcagtg 6480

atgaaaggat atttaaaaaa tttttggaaa aaggatcgat ttctgagcag aaaaagatgc 6540

ttttagaaga aaagaaatgt tcggataaac taactgcact gcttgggaat tactgcatac 6600

cgatagacaa tatttcagag tcagacggaa aaatatatgc ggtctataag cttccaaaaa 6660

atgttaaacc tttgtccgaa atcattaatg atgtatcctt ttctgattgt acgatgagag 6720

tacgtttgct tctcataaag agaattctgg aactcgtgtg tgcttttcac gaaaaaaaat 6780

ggtattgtct cagtatttca ccgggaatgc tcatggttga agattttgat ataccgatgg 6840

gaaatgtcgg aaaagtattg atatatgatt tcagaaatcc tgttccgttc gagtcagtaa 6900

atgaaagaca taattttaac gtttcaaata aatacacttc accggagctg ctcatccatt 6960

caagatatga cgagtcgaaa tctgtgagtg aaaaatcaga tttgtattct gttgcaaaaa 7020

ttgcggaaac aataatagga gattttaaca gtattattgc aaatggaaat ttgatactac 7080

ttgcaatgct tagagttttt atcagtacag ggaaaagtcc ggaacctgag tatcggtttg 7140

aatcgtcgga aaatatgctt tcagtatttg aaaatttgat caaagaaaat tgtttttttg 7200

aaaaaaacga ttatacatct atgtttcatc aggcgtatga caatttttt gaatggcagg 7260

aatgtttgat atcaccggat cacttggata aaaatatgtt cgaggcagct ttatcaaatc 7320

ttgaggatca gctgcttagg gttgatattg ataagtatag agcagagtac ttctataagc 7380

ttctccgaga gttgtctaat aaatataaaa atacaattac tgatgaacaa aaggtaaggt 7440

tggcaatact tggaatcaga gcgaaaaata atctgggaaa aagttttgat gcattggaaa 7500

tatatgagtc agtacgtgat ttagaaacta tgttggagga gatggcagag cttagtcctg 7560

tcattgcttc gacatatatg gattgctacc gatatgcaga tgcgcagaaa gtggcggaag 7620

aaaacattat caggcttcat aatagtaata ttcgtatgga gaaaaaaaga atactgcttg 7680

gaaggtcata tagttcaaaa gggtgcagca tggggtttca gcatattctt ggtgcggatg 7740

agtcatttga acaggcttta tatttcttta acgaaaagga caatttttgg aaagaaatat 7800

ttgagagcag aaatttagag gacagcgata gacttataaa gtctttacga agcaatacgc 7860

atattacgct gtttcattac atgcaatatg catgtgaaac aaggagaaag gaattatatg 7920

gagcactttc agacaaatat tttataggta aagaatggac agaaagactc aaagcatata 7980

taagcaacaa ggatatatgg aaaaactatt atgagatata tattctgcta aagggtattt 8040

attgcttcta tccagaagtc atgtgttcgt ctgcgtttta tgatgaaatc caaaaaatgt 8100

acgatcttga atttgaaaag gaaaaaatgt tttacccatt gagtctgata gaactgtatc 8160

ttgctctgat agagataaaa gttaatggga gtctgacgga gaatgccgag aagttgttta 8220

aacaggcatt gacacatgac aatgaagtca aaaaaggaaa tatgaatatt cagaccgcca 8280

tttggtatcg aatatatgca ctgtataacg atgtaaaaga tgaaactgat aagaataaaa 8340

ggcttttaaa acggcttatg attctttgcc gacgatttgg ttgggcggat atgtatagtg 8400

ctttggagaa ggatgggaag ttaattgatt ttttgagatt tgaggtatgt taaatgataa 8460

cacttgcatt agatgaaaat ggcaaatttg aagatgcttt ttctaaaaaa aatgaaaaac 8520

cgataatgat tgcggggata atctatgatg acaaggggaa agagtatgat gctgagaatg 8580

aacgctacag gatatccagt tatctgcgag cagtatgtga cagtttgggt gcgaaatacc 8640

ctcaggatct acattcaaat agtaatggaa ataaggcgac tgttgggaaa gtaaaatgta 8700

aaattggtga aacactaaag gaattcttga gagaaggaac ctatgaaaaa aaggaattgc 8760

cgacaaagaa cggttattta aataagagat ctggaaaata tgtaatgttt gcagaactca 8820

ggagtagtca gggagttaaa aagcgtgtta gtggttggaa tgacaatgat ctgactcagg 8880

atgaaaaggt cagcaatctg taccttcata tggcagaaaa tgccgttgtc agaatgctct 8940

tccataatcc tatatatgaa gatgtaacag atgtaaatct ctattttccc acgcgaaaag 9000

ttgttctgaa agatagagat agagaatacg ataaacaaga tttcaaaata tatggtgata 9060

aggacaagtg cgaagcagaa agcgggagat tggtgcatta tgatatcgtg tcatcggatt 9120

tttaccgtac gataatggag aacgaatgta caagaattaa taaaaagcaa ttaaatgttc 9180

attatatgaa cacaagccca atttcgtact gggagaaaaa tgaaaaatat aatacatttt 9240

tatatttggc tgacatagtt tgttctatgc tggattatta caaaaagggt tcgagtccgg 9300

cagagtggat ggattctttt gccgaatggg gaaacaaata ttttggtgat gatcagataa 9360

tcttatttgg gtatgatgat atagatgaca aatacatgga ggctgtagat gcagtaggac 9420

agggagagta ttttcatgcg ctggatatta tatatgatgc ggaatgtagt ggaagtgaat 9480

ttgagaagca ctacaaagat tattggtttc caaagcttat aaaaaagata cgaataacag 9540

caactgtgga taatttatgc agatcgatct cagatctgga gagttttaca tatcgaagta 9600

atcttgatca gcagaaactt ttgtggattt ttgaggaaat caaagctatc gtcgataagg 9660

gagattttgg aaagaaatat catacagatc aggttatgtt tgatatgtgt aatgccggta 9720

ttgctgtgta caatcatatc ggagattttg ggactgcaaa ggaatactat gatgagtgca 9780

tgaaacacac tggggatgtg gatctggtaa agatacttcg tgcatcaaat aaaatggtgg 9840

tctttcttga cgatgctttt aggtatggtg acgcgacaga acgtgccagg aagaatgttg 9900

aataccaaaa agctttgcac gatataaaga gtgagatttg tccggaaaag aaagatgaag 9960

acttgaacta tgccatatcg ctcagtcaat ttggacaggc gcttgcgtgt gaaaaaaatt 10020

ctgatgcaga gagtgttttc ctagagtcgt tgcggcatat gaggaaaggg actgccaatt 10080

atcagattac tctttcatat ttactccatt tttatctgga tatgggaatg acagattctt 10140

atcgagaaaa aacaaaggac tattttggaa gtgaaaaacc aaaggaacag ctgaaagaat 10200

tgctgaagtt atcgggaaag gatgatagta tagttacttt caaatttgca atgtatgtct 10260

atttacgtgc actttgggta ttacaggaac cgcttactga ttttatcaga acaagattag 10320

aggacatacg tgagactctt gtaaagaaga aaatgagtga acatatggtt ggacatccgt 10380

gggagttgat ttataaatat ctggcatttc ttttttatcg tgatggaaat tgtgaagctg 10440

ctgaaaaata tattcataaa agtgaagagt gcttggaaac acaaggactg actatagatg 10500

cgattattca taatggtaag tatgaatatg cagaattgtc aggtgacgag gagatgatgg 10560

caagagagaa agcgtacttt gatgaaaaag ggatagatag aaaaaatgtt tgtactttta 10620

tgtatcattg atgtttaata agatttgacc gaggagtgac aggtaatcgc cggtatatct 10680

ggtattacct gtcatttttt gatgaaataa gctacttttt gcctaaaaaa cgaaactgtt 10740

ggtgttttat gatgattgtg tcaacaaaag agagcaaaag aagaggagaa aagtaatgtc 10800

aatgatttca tgtccgaatt gtggtggaga gatatctgaa aggtcaaaga aatgtgttca 10860

ttgtggatat gtgttagtcg aagaagctaa agtagtgtgc acagaatgtg gaactgaggt 10920

agagagtggc gctgctgtat gtccgaagtg cggctgtcct gtaaatgata gtgagacgcc 10980

tcagaaagtt gaagtgacta gggtaaatgt atcttccgta atcagcaaaa aagtcgttgt 11040

aagcatactg atcgcagtga ttacaattgc aggtttttc tatggagtga agtattcgca 11100

ggaaaagaaa gcaattgaag agtcagtaaa gcagaaggaa gactatcaaa gtacgctaga 11160

gcttgcttcg ctaatgatgc ttcaaggagc ttcggatgca gaaacttgtg ggaatttggt 11220

taggaaagtg tggagcaact gcatttataa ggagagggat gaagaaaccg acaagtatac 11280

gtgtgatagc aggggtgcag gatggtttta tgatgatttt aatgatgcat taatggctct 11340

ttacagtgac agcagttttg gcaagaagat aaatgaaatc aaaaacggtc aggaaaccgt 11400

tgcggcgatg atgaaagatc tgaaaaatcc gccggatgag atggcagatg cctatgagga 11460

tattcaaaat ttttatgtgt cctatctaac gctgacagaa atggttgtga atccaactgg 11520

aagtttgagt tctttttcat ctgatttttc cgatgcggat acggaggtgt ccaatgccta 11580

tagccggatg aagttgtatt tagattaaac tattgaggaa aaaatggagg tgctttaatg 11640

cgggggagaa actgtggagg gtcatcaggc gacggactgc tggtacttct cgtactgctt 11700

gtccttttttt ataaaatcat gccattcata ggtttatgga ttttaatttt tggtgatgct 11760

gaacgtaaag atctgggtat gggtatgatt attgtcggga tagttctata tgtattatta 11820

gaggtttttt aatgtgagtt tctgtggtaa actataaaag tacaagcttt tgcgccgcac 11880

cgcataaata gcggatttat gaccattatt tggtgaaaaa aatggtgtac acctgtgttt 11940

ttttgttttg cgccgcaaaa tgcgccacgg aaccgcatgc agagcaccct gcaagagaca 12000

gggttatgaa aacagcccga catagagggc aatagacacg gggagaagtc atttaataag 12060

gccactgtta aaagttatga aaacagcccg acatagaggg caatagacat aaagaccaaa 12120

aacaggtcat ctgcatactg tgttatgaaa acagcccgat atagagggtg tgagagatat 12180

agttctcgtc acagtgcaga aaatgaccta ttatgtgccg aaaaacaaaa tgaaaaaaga 12240

atggaaaggc gtatttaatg aaatgctgat ctgttgattt gaattaacaa aaaaaggtcg 12300

ccccacggat gacaaaaaca tccgggggcg acccttttt 12338

<210> 65
<211> 6098
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 65
tactgtgtgc ataagtcttc cttagatcca taggtacagc agttttattt attagcctta 60

gaaaatggaa aatagagctt ataaatgata tgatatttat gaataaaatg attgcattct 120

cgtgcaaact ttaaatatat tgattatatc ctttacattg gttgttttaa ttactattat 180

taagtaggaa tacgatatac ctctaaatga aagaggacta aaacccgcca aaagtatcag 240

aaaatgttat tgcagtaaga gactacctct atatgaaaga ggactaaaac ttttaacagt 300

ggccttatta aatgacttct gtaagagact acctctatat gaaagaggac taaaacgtct 360

aatgtggata agtataaaaa cgcttatcca tcatttaggt gttttatttt tttgtgatta 420

tatgtacaat agaagagaga aaaaaatcat tgaggtgaaa actatgagaa ttactaaagt 480

agaggttgat agaaaaaaag tactaatttc tagggataaa aacgggggca agttagttta 540

tgaaaatgaa atgcaagata atacagaaca aatcatgcat cacaaaaaaa gttctttta 600

caaaagtgtg gtaaacaaaa ctatttgtcg tcctgaacaa aaacaaatga aaaaattagt 660

tcatggatta ttacaagaaa atagtcaaga aaaaataaaa gtttcagatg tcactaaact 720

taatatctca aatttcttaa atcatcgttt caaaaaaagt ttatattatt ttcctgaaaa 780

tagtcctgac aaaagcgaag aatacagaat agaaataaat ctctcccaat tgttagaaga 840

tagcttaaaa aaacagcaag ggacatttat atgttgggaa tcttttagca aagacatgga 900

attatacatt aattgggcgg aaaattatat ttcatcaaaa acgaagctaa taaaaaaatc 960

cattcgaaac aatagaattc aatctactga atcaagaagt ggacaactaa tggatagata 1020

tatgaaagac attttaaata aaaacaaacc tttcgatatc caatcagtta gcgaaaagta 1080

ccaacttgaa aaattgacta gtgctttaaa agctactttt aaagaagcga agaaaaacga 1140

caaagagatt aactataagc ttaagtccac tctccaaaac catgaaagac aaataataga 1200

agaattgaag gaaaattccg aactgaacca atttaatata gaaataagaa aacatcttga 1260

aacttatttt cctattaaga aaacaaacag aaaagttgga gatataagga atttagaaat 1320

aggagaaatc caaaaaatag taaatcatcg gttgaaaaat aaaatagttc aacgcattct 1380

ccaagaaggg aaattagctt cttatgagat tgaatcaaca gttaactcta attccttaca 1440

aaaaattaaa attgaagaag catttgcctt aaagtttatc aatgcttgtt tatttgcttc 1500

taacaattta aggaatatgg tatatcctgt ttgcaaaaag gatatattaa tgataggtga 1560

atttaaaaat agttttaaag aaataaaaca caaaaaattc attcgtcaat ggtcgcaatt 1620

cttctctcaa gaaataactg ttgatgacat tgaattagct tcatgggggc tgagaggagc 1680

cattgcacca ataagaaatg aaataattca tttaaagaag catagctgga aaaaattttt 1740

taataaccct actttcaaag tgaaaaaaag taaaataata aatgggaaaa cgaaagatgt 1800

tacatctgaa ttccttttata aagaaacttt atttaaggat tatttctata gtgagttaga 1860

ttctgttcca gaattgatta ttaataaaat ggaaagtagc aaaattttag attattattc 1920

cagtgaccag cttaaccaag tttttacaat tccgaatttc gaattatctt tactgacttc 1980

ggccgttccc tttgcaccta gctttaaacg agtttatttg aaaggctttg attatcagaa 2040

tcaagatgaa gcacaaccgg attataatct taaattaaat atctataacg aaaaagcctt 2100

taattcggag gcatttcagg cgcaatattc attatttaaa atggtttatt atcaagtctt 2160

tttaccgcaa ttcactacaa ataacgattt atttaagtca agtgtggatt ttattttaac 2220

attaaacaaa gaacggaaag gttacgccaa agcatttcaa gatattcgaa agatgaataa 2280

agatgaaaag ccctcagaat atatgagtta cattcagagt caattaatgc tctatcaaaa 2340

aaagcaagaa gaaaaagaga aaattaatca ttttgaaaaa tttataaatc aagtgtttat 2400

taaaggtttc aattctttta tagaaaagaa tagattaacc tatatttgcc atccaaccaa 2460

aaacacagtg ccagaaaatg ataatataga aatacctttc cacacggata tggatgattc 2520

caatattgca ttttggctta tgtgtaaatt attagatgct aaacaactta gcgaattacg 2580

taatgaaatg ataaaattca gttgttcctt acaatcaact gaagaaataa gcacatttac 2640

caaggcgcga gaagtgattg gtttagctct tttaaatggc gaaaaaggat gtaatgattg 2700

gaaagaactt tttgatgata aagaagcttg gaaaaagaac atgtccttat atgtttccga 2760

ggaattgctt caatcattgc cgtacacaca agaagatggt caaacacctg taattaatcg 2820

aagtatcgat ttagtaaaaa aatacggtac agaaacaata ctagagaaat tattttcctc 2880

ctcagatgat tataaagttt cagctaaaga tatcgcaaaa ttacatgaat atgatgtaac 2940

ggagaaaata gcacagcaag agagtctaca taagcaatgg atagaaaagc ccggtttagc 3000

ccgtgactca gcatggacaa aaaaatacca aaatgtgatt aatgatatta gtaattacca 3060

atgggctaag acaaaggtcg aattaacaca agtaaggcat cttcatcaat taactattga 3120

tttgctttca aggttagcag gatatatgtc tatcgctgac cgtgatttcc agttttctag 3180

taattatatt ttagaaagag agaactctga gtatagagtt acaagttgga tattattaag 3240

tgaaaataaa aataaaaata aatataacga ctacgaattg tataatctaa aaaatgcctc 3300

tataaaagta tcatcaaaaa atgatcccca gttaaaagtt gatcttaagc aattacgatt 3360

aaccttagag tacttagaac tttttgataa ccgattgaaa gaaaaacgaa ataacatttc 3420

acattttaat taccttaacg gacagttagg gaactctatt ttagaattat ttgacgatgc 3480

tcgagatgta ctttcctatg atcgtaaact aaagaatgcg gtgtctaaat ctttgaaaga 3540

aattttaagc tctcatggaa tggaagtgac atttaaacca ctatatcaaa ccaatcatca 3600

tttaaaaatt gataaactcc aacctaaaaa aatacaccac ttaggtgaaa aaagtactgt 3660

ttcttcaaat caagtttcta atgaatactg tcaactagta agaacgctat taacgatgaa 3720

gtaattcttt taaagcacat taattacctc taaatgaaaa gaggactaaa actgaaagag 3780

gactaaaaca ccagatgtgg ataactatat tagtggctat taaaaattcg tcgatattag 3840

agaggaaact ttagatgaag atgaaatgga aattaaaaga aaatgacgtt cgcaaagggg 3900

tggtggtcat tgagtaaaat tgacatcgga gaagtaaccc acttttaca aggtctaaag 3960

aaaagtaacg aaaacgcccg aaaaatgata gaagacattc aatcggctgt caaagcctac 4020

gctgatgata caactttaaa aggaaaagca gtggattctt cacaaagata ctttgatgaa 4080

acgtatactg ttatttgtaa aagtatcata gaagcattag atgaaagcga agagagatta 4140

caacaatata ttcatgattt tggagatcaa gtggattctt cacctaacgc acgaattgat 4200

gcggaattac tacaagaagc aatgagtagg ttagctgaca taaagcggaa gcaagaagca 4260

cttatgcaat ccttatcttc ttctacagca acgctttacg aaggcaagca acaagcgtta 4320

cacactcaat tcacggatgc gctggagcaa gaaaaaatat tggaacgcta tattactttt 4380

gaacaaactc acgggaattt ttttgactca tttggagaac ttgtctatcg aacgggacaa 4440

gcagtgcgtg aattagctaa taacgtcaca ttcgagagcc aaacaggaag ctatcatttt 4500

gataaaatag atgcttctag attccaaact ttgcaagaaa tgttgccaaa ggcaaagaaa 4560

aaagcattta attttaatga ctaccaaata acatggaatg gcaccacgca ccttttatgg 4620

aaaaatggta aagtggatgc agaagcaacc aaagcttata acgaggcgaa actgaatgga 4680

aagctaccaa aggaaggtaa tgtagcaaca caagatgcag aactattaaa aggcattttg 4740

gcttcactga aaaacaagaa agatcctatc actggagcag atataagcag tgtgcatgta 4800

ttatctatcc ttagcgggct cgcattctcc tatacagctg ggaattataa gggaagaaaa 4860

cttactgttc caaaaagttt cttagacaaa ttaaagaaaa accgaaaatc taaagtacct 4920

aaactatcta gtttatcaga aaaacaacaa ctaaaactcg caaataaata caagaaaaaa 4980

tcacctattc caattccaga tgatgctaaa atcaaagctc agacgaaaaa ggctggttat 5040

gaacaaatat cttataaatg gaaagagaat gggataacct ttgaagttag atggcatact 5100

aggacaccag gtgcaccaaa ggaacaagga aatacgtttg ttatagaaag aaaaattcag 5160

ggtacagcag aagggaaaac aaaagttcaa caaatattgg ttggagataa taagtgggtg 5220

agtaaaagtg agtggcaaaa ggctataact gataagaaaa atggtgtaag tacctcggag 5280

caaaataaaa tgttgtctga tggacattgg aaagaataga aaggagcaaa atgatggaag 5340

attattataa aggttttgag ggatatccag agatagattt ttatacgtat atagatgata 5400

tgaaattggg tatagcaatg tgggaaggat actttgacaa cattatgaaa gaaattaatc 5460

caagtaacgg aagatggact tcattagcgt attattatca tttagatgag gggtggtatg 5520

atgaaagtcc ttgggaaata ccaagtaata cagaagcatt agaattattg gaaacaatcc 5580

atatatctaa tctagatact atcacacaag agatattact taaattaata aatttattaa 5640

agaagaatat aaatagacaa gtttatattg aatactcata aaaaagatga ttatgatata 5700

ttatagaaca aacgaacaag ccccaaatac gaggtttgtt cgtttgtttt caatataatt 5760

atttgccacc aagtgagata ttacggtttt aaatagctta tttgacgata ccaaaccctg 5820

ataagagaaa gaagaaagag aaagctggtg tagttgtttt aagtgaacta gataaaaaat 5880

taatagcaaa acttgaaaaa gatggtgtga aaatatcaaa agaagatgtt ataggaataa 5940

aataattgcc agatgatgag aaatcgtttg gctggaaaaa ggaaatccat ccgctggatt 6000

tgagcatatt cttattgaac atggtgaaca atttgctaaa tagggaattt caaaagctga 6060

gttacctgat tttttgatga ctgctttaga aaaggaaa 6098

<210> 66
<211> 6222
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 66

attctttaaa aatatctaat aatttattta ctatatactc taatacatct tttaacctat 60

ctaaaacatc atcacctaca acatcccaaa aatcatctaa aaagttaaaa aaatccatct 120

ttatcaactc ctatatctat tttttattgt gtaattcctg agttacaaaa ccattataac 180

acgtattaca cacgtagtca atacttcaaa aaaatttttt gtatattttt ttgaataagt 240

aaataaaaag agctgtgtag ctctttatta aaatcaatat ttttatttg ttaacaaact 300

tagacaacat taaatttaga aacctatata tatttcagta cttttcattt ttaggtagtc 360

taaatcagaa atggttttgt ctaaatgatg tatgtaagtt ttagtcccct tcgtttttag 420

ggtagtctaa atcagaagtc atttaataag gccactgtta aaagttttag tcccccttcgt 480

ttttagggta gtctaaatcc catccaaatt atgggataat atgttacttt ttatttttaat 540

atttgattat ttattgtttt tttactgatt tagattaccc ctttaattta ttttaccata 600

tttttctcat aatgcaaact aatattccaa aatttttgtt tcttttctta tgatcttttc 660

tccgatagtt atttctccag ataagatttt catttttttg aattgatctt ctgttagaat 720

taatgttctt actgatgaat tttctggaac tatcattgac aactgatttt cataggaaat 780

tatttttct tttgtgctag aacttacaat gtatactgat ttttgtacct gataatatcc 840

ttttcttata atttcttttc taaattttgc atattctttt ttttcttttc ctgtttgcat 900

tggaaaatca tacattagaa tccctacata attagtactc ataatcctct atccttaact 960

caggaatttc tacttctgac atttctcctg taaaataatt tctaatatta tctaaaaaat 1020

aatcaatcac ttgagccaat tcatattttt tatttttcca ataaactttt tgtgttaata 1080

ccaataacaa tttttgtctt aatgatttat tcaaacttac ttcttcctgt tgattaaaat 1140

atacgatata atctaccatt ggacgaaata tttcaataat atcatctgca aaattataat 1200

tattaaattg tgaactgtga tgtattccca aacttggatg aaatccttta gccacaattt 1260

ttgaagagat taagcttctc aaaaccatat acccataatt taatgccgaa tttgtcccgt 1320

cttcaccaaa tctcttaaat tttttcccaa aaagttcacc aaaatacatt cttgcagcaa 1380

ttgcttcctg atgttccgct tcttttcctt ttaatctaat attattttca tatgcttcca 1440

acttatatga tacttcctga gatttttca aaaactgcaa taaatttctt tgattttcta 1500

tttttctcat tacaattttt ctccagattt cttctttttt atcgtcaatc cagctcactt 1560

gctcattaat tcttgttgtt acttgaaaat gattatacag tcctaatgaa tgtaaaactg 1620

gctgatgttt ttcattacaa attatcagtg gaatattatg ttctgataat cttaactgta 1680

atattccgct aattttacat ctgcaatttt caactacaat tgccatgata tcatttaaag 1740

atactttatc agccttattt tcatcatctt catttatcat cacaagctgg ttatttaaaa 1800

ctgataattc attgactctt gttacatgga taatattaga catttttatt actcctttac 1860

tctaaagctt tatattcaaa cataactttc acaagttcac acaattcttc tgaatttcta 1920

tcagtcatta atttttttctt ttttaaattt ttcaaatgta caatttttttc cgattctaaa 1980

gtctgaattt ctattttctt atctgctcct attttaaatg ttgctacaaa accatattcc 2040

tttaatatat ccactattga tttcataatt gcatttttaa gttttctatc ataagaaagt 2100

aattttctta aattttccag cacttctaaa agtgaaattt cagcatgcgg aatatagtta 2160

aaatgtgcaa tatagtttcg tatatacaaa tctttttttct cttgttttaa ttttttttact 2220

tttttatcag aatagatgct tctttttttct acattatctt tgtataattc tttataaaaa 2280

tttatatatt tttcaacaat ttgcccactt ttatatttta catttttact gttatcaaaa 2340

ttaaatattt cttcaatata atgattttca ggaaattcac ctttcaatct aaatcttaag 2400

tccctttccc agatcgaagt atatcccaca agtctgtgga gtatttttaa taacaagcct 2460

tgcaacaagt ttaattcatt aaattccact ttatttttca aatgagtata tttttgtata 2520

tttccaattg ctttttcata ttctttataa tcttcatcat taaatttttc atcttttta 2580

ggtcttgcat attttctatg taaattttgc tgcattgtat aatttttttc tatttcattt 2640

tttttattgc tgtattcttt caattctttt aaacttattt tatacttcgc tttatcagct 2700

atttttcaa gtaaatttaa catcccatat ttttttatat tataaaaagc tctatgcttt 2760

ataatatttt ctccatcaaa atatatttta tttgtgtcaa atttcttcaa ttctttccta 2820

tcttttattt tattttcatt aaaatctaaa aattttccaa tttcattcgc ttctaattca 2880

aaatcttctg ttactctatt attatctaaa tttaaaagat ttataagttc aagttcatct 2940

gaaaaagttt cttctttatt tgcactctga tatttttcaa gacttccctt caaattagtc 3000

aattctttat gattaagcaa ttttaaaatt aaataaaaca tattcaaatt ttcagtgtat 3060

tttaatatct ttcctaattt tatctctctt acaaattcat ttatttcatg tggaatttct 3120

ttattcctat tatgtttttc ataatttttt aaaattttat catatttttc tttattatct 3180

ttttttattt ttattttaga aaatatatca ttattatcat tgttattatt actttctata 3240

tattttaaat tatttttatt caaataatct ataaaacctt ttaaaaatat ttgttgtata 3300

aaatcaatgt atgtattttt ttcttcttta tcttgattat taatcatctc cctactttgt 3360

ataatagcaa gatattctac tggtacagtt ttttctatat tttcaaattt ttgatattta 3420

taatgtcctg ttttttgatt tctttgttta tttatttta ttacttcatt agttatttta 3480

aaaaaaactt tactatttt aacaaattta ttaagaaatt caccataata aatatttttc 3540

aaaagatata tttgagcatc tttttcttct ttatccttag gaacactcca aaaaaatttt 3600

aaagtatttc ttaaatcttc tattttatta tataatttcg taaaagaagg aacaaaagga 3660

atattcttat ttacaaaatt aaattttgta ttttttaaat atttaattat cacatccttt 3720

tcataataat taaatacatt tgcactattt aactgcttaa atatcttcaa tttcaatttt 3780

ttctcattta tttcattttg aaacattttt tttgaaattt cagaaggagc tatattttta 3840

aatgcaaata tatctttccc ttctaattcc aaattaaaat gcacaatccc atgtctaata 3900

ctgctaatag cttcatcaat atttgcaaaa aaatcttcta tctcattttt attatccata 3960

ttaaaatcat aactatagaa cattttaaa ttttctttta cttcattttg cttgttttca 4020

ttatatattt tatcaacttc tccagaaaca tatttttctt cgcccttatt atttttaca 4080

gtttttcctc tcattctacc tgtaatatca ttctcatttt cagtttcaag aatatttctc 4140

aatgaaaaat atgcaaccga agaaactcca attatatttc gtaaaaatgc ttcattttgt 4200

ctattcctag caataaaatc acttgttgca atctctccaa cttgtaaata ataattgtat 4260

ttcccacaat ttcttacata agtatccaat ttatttagta atttgttttc aattaatttt 4320

tttaaatttt gatattcaaa tattctctta attttatcgt tacttatgtt actcagtctt 4380

ttatacacat aatttttcaa aagctgactc atttcaattt ccacaaaatg acaaaaagca 4440

tattttatat ttttatcatt aagttcttct ttatccaaat aatatttata aaacacttgt 4500

gatttttta attcactcat atccggaatt ttttcaatta attcttttat attatttaca 4560

ttttgtattt cttcgtaaat aattttagca aaatttttctt tatcattttt tcttccaatt 4620

attttgtgat agtattctct tattttatat ttttcatgtt tttttgaatt ttctattaaa 4680

aaaaataact tctcaatatc ttctttttta tacaatttat caaatgcttc ctgtacatta 4740

tttatataat cattacgctt tgctgattct ctataataat cataaataat atttcttttg 4800

ctcttccctc caacttttttc aacattattt tcattaattt tctgataatt agccttattt 4860

tcttcaaatg aatattttaa agaatttatc ttattcaatt ttgcctcaac atcttttcta 4920

aatatttcta attcttcaga gttcacatct tcatttaaca atattttctt taaaactgaa 4980

aaactatttt tatttttaa atcatattct gaaatatctt cttcagaata atttttatcc 5040

tgtactgcat ttttctcttt cctattcttt aaatacagaa cactatcttt tagatgcaat 5100

actttatttg aaaaaaactt ttttaaattt tctcttctta ttctattttc ttcttcactt 5160

gcattatcag gatttttttat atatatatcc agtcttatac ttaaaagctc tgacaatctc 5220

tcactagtcc tattttcttc gctcgtactt tttactaatt ttccctcttc aatatatttt 5280

ttatgcgaaa ttccatcaac ttttgtaact ttcatatata aaaacctcct aatatctata 5340

ttttttactc aatacctaat tctttttttca atgctttttg taaaatttgt gaaaaattca 5400

gatttttttc ctgtgccaat atatctaacc aaacaggaat tgttaaagtt ttctttttaa 5460

gtgcatttgt aacttttgcc acttcataca ctggatcaac agataaaata tacaaatact 5520

gattttcttt cagtttcaca tcctccactt ttgaaggctc aggaaatttt tttcttacat 5580

ccaaaaaatc agccaaatgc agacccaatg tctctctcaa attggaaaca gcctcctcca 5640

tgctatctcc aaatgtagca taataattta tctctccatc ttcaaactta tcaaaatcaa 5700

caatacaacc ataataagtc ccatcttcct tagttaccac tgctggataa aatacatcca 5760

ttttaattat ctccaatcta taccacgtgt taaatacgtg tttaaaaata tttataaaat 5820

tttttagcat ctctgctaaa ataaaacaat tatttcaaat ttttctattc cttaatcact 5880

cattgttagt gattcttttt ttacttggac aatttttcat ttaatttctt caatttttttt 5940

aaaatcacat ttttttaata ttccttattt aattgcaaat tttcattact tttggggtgc 6000

tctaaatccc atccaaatta tgggataata atttttagtg aaagcaagaa gggactagaa 6060

tttaatccca acttgttttt caatacttct taatgttcct acaggtatat cttttgaata 6120

tggtactgtg accacacctt ccacacctgg gatcatccat tgataatgac tacctcttat 6180

acgcacaact tttccgccta attttctaaa tctttttttcg at 6222

<210> 67
<211> 6337
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 67
ctttctatct ttttcaaata aaattaggct ctagttagcc taatcgcata attatttatt 60

atagtataat tcttatttttt tttcaaccta aaaatttaaa acatctccaa aaattttcgt 120

ttcagaacaa ccaagcaacc atattcaaaa aacaataaaa aatgagcaag aattgaaatt 180

ttattctcac tcagaagtta tttttattaa atatcacttt tcgatattgg ggtggtctat 240

atcaatttaa aagacagaat agataattct ttagagtttt agtccccttc gatattgggg 300

tggtctatat cagaagtcat ttaataaggc cactgttaaa agttttagtc cccttcgata 360

ttggggtggt ctatatccca tcctaatttc ttgctgatga gatatttatt tctaattttt 420

ctattttgtc tttattttca atactttcaa tcctatttttt ctctttatta ataatataga 480

accaccctat actattatac catatttttt gatttttcaa aattccaata ttttgttttg 540

tgaaattttt tctcccattg tcacttctcc tgcaagtacc ttcatttttt gaaactgatc 600

ttctgtcagg ataatggaac ggattgatga attttctgga gcgagcattg ataactgttt 660

ttctgccagt tcgatttttt cttttgtttt cgacctcatt atatataccg atttttgaag 720

ctgataatat ccctttttcta tcaatttttt cctaaaagtc ctatattcaa atctctcaac 780

atctgtctgc ataggaaaat catacataag cagaccaaaa tactcaatac tcatagtcca 840

tcacgctcaa tgtcggaatt atcacttctt catcttttac aaaataattt cgtatactat 900

ccaaataata gtctaccgct tggaaaaaat catatttctt attgttaaat aataccttct 960

gctgtgctac aagaagtatt ttttgcctta tttccttact taatttcact tcattcaaaa 1020

tatccttgta catataaaca agataatcca ccataggacg aaaaacctct attatatcat 1080

cagaaaaatt ataggcatta aactgtgact tatgatgtaa tcctaaactt ggatgaaatc 1140

cttttgctac aatctttgat gatattatag ctcttaaaat catatatcca taattaagtg 1200

cagaattcac tccatcttca tcaaatcttt taaaactatt actatacaat tcctgaaaat 1260

atatccttga agctattgct tcctgatgtt ctgcactcgc atcatctttt ttcaagtttt 1320

ccttatatgt tttcagtctt tcaatggaaa tatcactttt ttcaagatac tctaacaatg 1380

ctctttgatt ttcaatctta ttctccacta tcctgctcca caatttttcc tttttctctt 1440

tttcccactc aatctgctca tttattcgta aagtcacttg aaaatgatta aataatccca 1500

gcgaatgaat ttcaggctga tgtttctcgt tgcaaataat aatcggaatg ttattttcca 1560

ccagcctcaa ctgcaaaatc gcactaatct tacaatagca gttttcaata actatcgcag 1620

atatatcatt caaagaaatc ttatttttct catcattatt gtcttcatca accattataa 1680

gctgattatt cgatattgac aaatcatcag cccttgttat gtgaattata ttgggcattt 1740

taatcatact ccttataaat ttcattctta taacgtatca ttcgtatttt ctatttttgt 1800

taaaagttct attatcaagt ttttaatata atcagaatta taactttcta attctaaaac 1860

agaaactttt ttaggtttca ttaatctttc aagtatatca ttattaccga taagtttaaa 1920

ttttttcttt aattcatcat aatctaaatt cacatctttt ttaaatactt caaatacact 1980

tgcataagtt gaattattat aacgtgtact atatgataat aaattagaaa ctctatcaat 2040

ttgttctgca atactgtaat cagcaaacgg atttcttaca atatagaaat gtgaaatata 2100

gtttctaata ctttcatttt ccggcttatt aatttcagaa ttttcagaca aatcaattcc 2160

aaatccataa catattttct caaatttttt ataagattct tcatcaaaaa atttatagta 2220

tgctgttgtt gtataaaagc catcagatcc attacgctta ggataagctc tacttattcc 2280

agtattgtag ccacttaact taataattcc taattctctt agcccattta caatatagtg 2340

catatctctt tcaaatctag ccatttgaat agcaagtttc caatttatat ctatcaaata 2400

actttctatt ttattcaaat aattaaattc taccaaatct ctaatttttt tgtattcaga 2460

aactctatta taatcttttt caaatgattt atagttttta ttttgtatat tttttgcaaa 2520

aaagtcatca ttttctttca atttttttat atacttctct ttgtattctt tagaatatcc 2580

atttagttta tcatttagat ttttcaatat tgcatcaatt tcagatattt tatttttttct 2640

aatattttta ccatcaatat taaataaaaa ttttgcatca gccattttaa tatcatttga 2700

aattaatcca taaattttat caaaatttgg atttccaata tttaaaaata aattcttttt 2760

ataaatatat aattcattct tacgttcttt aggataatat atttcttgaa atttattttc 2820

attctctgat tccatatctt ctattaaatt atctatttct tttttgtatt tttttaaaaa 2880

atcagaatta aatattattc tacacaatat tttactcttt atttcctgat ctttatcttt 2940

tatatactga tcaacctttt ttttcaaatc ctttttattt atgtttgata actttctttg 3000

ttcatcttgt aatatattcg atttttttatc tatctcaaat ttagtttcat catcaaaaat 3060

tacaattttt tctaattttt tctctaaaac atcacaacca ttaatatcat ctttaaattc 3120

agttaatata ttatttttta tatcctcata ataattatta aaaatttctt ttttagtttg 3180

tattttaaaa tcatcaaagt ctttttctat ctctttcatt ttttgaataa attcttctaa 3240

attaagattc caatttttcag ttatacattc atttctcaaa gtatttaatt gcattatttc 3300

atctaaaata tctataatat tttgatattc tgaagtattt aaccaaaactg atgttgcaaa 3360

aaatctattt ctaattttat ttataaccgc attactattt aacagtgcaa atattgaaat 3420

tatatattca aaatcatcat ttattactat agttttatca ctagtcttta cagttattct 3480

ttcgtaagtt ttattatcat taatgtcttt tatttgtttc ttaatttctt gaatattcat 3540

tttaaaatct gaaaaatcaa aaagttcctc ataatttttt ctcaaatatc caatataaca 3600

ttctattact ttttttctgat atttttttaat agctttatta ttacctttttg aagcagaaat 3660

ctgagcattt ttataataat tttctataat attttcatct atttcatcaa tgtttcctaa 3720

agttttcttt aattcttgta aaaatatatt cttactttca ttttcttcta aatcatcttc 3780

taaaattaat ttcttataca attctttatt cacatatatt aaagcattta atactatttt 3840

ttctgtttct atagtatcaa atggttcatt cttaggatta ttcctatata aatttaatat 3900

ttcaggaagt actttagaaa aggatggtaa atatttaata tcattattat tttcttctga 3960

aattttaata tcatttattt tagtaattat attttttttа tctttaaata ctacatctaa 4020

atttaatgct tttgacactt cttcatctga tatttttaaa ttttgaatta tatttatgac 4080

tttattatag tcatcttgcg ttccttgtaa atctctttcc ttgctaatcg catgtaatat 4140

cctgtttctt tcatttgttc ctatctttgt aaatttccta ataaaattat ttgtaatgtt 4200

atttttatta tctataaaat ctaagtctct tattattttt atttttgaat ttaaaatttt 4260

tttatcaagt acgtaatttt tttctcgatc tcctccaaag aaatctatat tttcatcatt 4320

atttatattt tctctagaaa aaatcttatt taattccata ttggtagaag caaaaaaagt 4380

aatcaattct aaatccaatt cctctttagc gtgaagtcta gaaaaatcat cagtatttac 4440

tgttgtcata tctatatcat tatgtcttaa tttccctaaa tacataatat gctctaacgt 4500

atattgctta actcttttta aaatttttтc agataatata ctttcattta aaatttttтc 4560

tatttctatt ttttccattt tctttaatct gacttttтgt tcatttacca atatttttтc 4620

aattcttcct ttcaaatatc gatatatgat tttatatagt tctttttctt catcagattt 4680

ctttgaaaat tttttcgaat caaaattaac tttataatgt tttttaaata ttccaaaaat 4740

ttctgtatca caatttcctt tttttagttc tttttctaat ttttttatta attcatctat 4800

tttaaattct gctaaaattt tttctatttt ttcttttata ctattatttt ttatattttc 4860

tacaaaaaat tttacaattt tatctttttt attttctctt tctattttaa atttttcgtg 4920

cttatctaat agtacataag attttatata tgttctattt cttctctttt caagaaattc 4980

attattaact ttttttactt tttcaattct tttagtaata ttccaaaatt ctaactcttt 5040

tataacaaaa tcagctatat cttctactgt taaatctaca tttatattta aaatttttтc 5100

aacaagcatt tttttatttt tagatttctt tttatcacca ccaacattaa gataaaattt 5160

tacaaaaccc agaatttcta aattactttt tatttttтct cttatttcca taaaattagt 5220

caaaataaca tctattttat catctttcaa taatttttct cttaaatgtt cttcataata 5280

tcgattttca aatacttttt ctgtttcatt ttcaattatt ttttctataa tcttatataa 5340

actcatgtta atatttttaa aaatttcgta aattgatttt tttgtttcta attcatcatt 5400

ttctattatt cttaatatta ttgaacaatc atttagtgtt ttattagtat actcatctct 5460

gatatctatc tctatttctt cttcattctc ttgtctcttt atttctattt ttttatcatc 5520

tttagttatt ccttgcctaa ttgcttcatc tattattttc ttttttgtaa tccccaatgc 5580

tttcaatttc tcagattttc catatgcttc tatatataat acaacttctt ctgtttccaa 5640

aaaatcatca ttatttttcta ttcttatgat tccttcttta cctttcaact taaatagaat 5700

atttcctgca tgaaattttc ttgtaaattc tttaagaata ttatcatttt ttttgtaatt 5760

aatatatttt ctaataaatt tattattatc aatttttttct ttattattat tttcattaat 5820

atttaaaatg tatttgtttc catcatagtt ccttttaact tttactttcc gttttatttt 5880

aaaatctttt ttatcacgaa cttcatacca tctcttatgt ccaaataaat ttcccattcc 5940

aatctcctcg tttctacttt aatctaataa aatattttta aattaaatca attttacatc 6000

tttctaatca aaaatacaat tttccatttt tagtatacca catcaatatt aaatctcaaa 6060

aaaataagga gccgtcaaac atagctccct acttctattt actcataatc cccatctatc 6120

cttacttttc gtaaaatcaa tccttctttc gcctttagat ccaacttaat tttcccattt 6180

gaacctgttc taaatgttct gccttctgtt accaaatcaa taaatctttc atcctgataa 6240

tttgtttcaa attccacatt ttcccagctg ttaaacgaat tatttattac aacaataatt 6300

aaatgatcct cgattactct ttcatacaca attattt 6337

<210> 68
<211> 1477
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選佩萊格裡尼（Peregrinibacteria）細菌序列

<400> 68

Met Ser Asn Phe Phe Lys Asn Phe Thr Asn Leu Tyr Glu Leu Ser Lys
1 5 10 15

Thr Leu Arg Phe Glu Leu Lys Pro Val Gly Asp Thr Leu Thr Asn Met
20 25 30

Lys Asp His Leu Glu Tyr Asp Glu Lys Leu Gln Thr Phe Leu Lys Asp
35 40 45

Gln Asn Ile Asp Asp Ala Tyr Gln Ala Leu Lys Pro Gln Phe Asp Glu
50 55 60

Ile His Glu Glu Phe Ile Thr Asp Ser Leu Glu Ser Lys Lys Ala Lys
65 70 75 80

Glu Ile Asp Phe Ser Glu Tyr Leu Asp Leu Phe Gln Glu Lys Lys Glu
85 90 95

Leu Asn Asp Ser Glu Lys Lys Leu Arg Asn Lys Ile Gly Glu Thr Phe
100 105 110

Asn Lys Ala Gly Glu Lys Trp Lys Lys Glu Lys Tyr Pro Gln Tyr Glu
115 120 125

Trp Lys Lys Gly Ser Lys Ile Ala Asn Gly Ala Asp Ile Leu Ser Cys
130 135 140

Gln Asp Met Leu Gln Phe Ile Lys Tyr Lys Asn Pro Glu Asp Glu Lys
145 150 155 160

Ile Lys Asn Tyr Ile Asp Asp Thr Leu Lys Gly Phe Phe Thr Tyr Phe
165 170 175

Gly Gly Phe Asn Gln Asn Arg Ala Asn Tyr Tyr Glu Thr Lys Lys Glu
180 185 190

Ala Ser Thr Ala Val Ala Thr Arg Ile Val His Glu Asn Leu Pro Lys
195 200 205

Phe Cys Asp Asn Val Ile Gln Phe Lys His Ile Ile Lys Arg Lys Lys
210 215 220

Asp Gly Thr Val Glu Lys Thr Glu Arg Lys Thr Glu Tyr Leu Asn Ala
225 230 235 240

```
Tyr Gln Tyr Leu Lys Asn Asn Asn Lys Ile Thr Gln Ile Lys Asp Ala
                245                 250                 255

Glu Thr Glu Lys Met Ile Glu Ser Thr Pro Ile Ala Glu Lys Ile Phe
            260                 265                 270

Asp Val Tyr Tyr Phe Ser Ser Cys Leu Ser Gln Lys Gln Ile Glu Glu
        275                 280                 285

Tyr Asn Arg Ile Ile Gly His Tyr Asn Leu Leu Ile Asn Leu Tyr Asn
    290                 295                 300

Gln Ala Lys Arg Ser Glu Gly Lys His Leu Ser Ala Asn Glu Lys Lys
305                 310                 315                 320

Tyr Lys Asp Leu Pro Lys Phe Lys Thr Leu Tyr Lys Gln Ile Gly Cys
                325                 330                 335

Gly Lys Lys Lys Asp Leu Phe Tyr Thr Ile Lys Cys Asp Thr Glu Glu
            340                 345                 350

Glu Ala Asn Lys Ser Arg Asn Glu Gly Lys Glu Ser His Ser Val Glu
        355                 360                 365

Glu Ile Ile Asn Lys Ala Gln Glu Ala Ile Asn Lys Tyr Phe Lys Ser
    370                 375                 380

Asn Asn Asp Cys Glu Asn Ile Asn Thr Val Pro Asp Phe Ile Asn Tyr
385                 390                 395                 400

Ile Leu Thr Lys Glu Asn Tyr Glu Gly Val Tyr Trp Ser Lys Ala Ala
                405                 410                 415

Met Asn Thr Ile Ser Asp Lys Tyr Phe Ala Asn Tyr His Asp Leu Gln
            420                 425                 430

Asp Arg Leu Lys Glu Ala Lys Val Phe Gln Lys Ala Asp Lys Lys Ser
        435                 440                 445

Glu Asp Asp Ile Lys Ile Pro Glu Ala Ile Glu Leu Ser Gly Leu Phe
    450                 455                 460

Gly Val Leu Asp Ser Leu Ala Asp Trp Gln Thr Thr Leu Phe Lys Ser
465                 470                 475                 480

Ser Ile Leu Ser Asn Glu Asp Lys Leu Lys Ile Ile Thr Asp Ser Gln
                485                 490                 495

Thr Pro Ser Glu Ala Leu Leu Lys Met Ile Phe Asn Asp Ile Glu Lys
            500                 505                 510

Asn Met Glu Ser Phe Leu Lys Glu Thr Asn Asp Ile Ile Thr Leu Lys
        515                 520                 525

Lys Tyr Lys Gly Asn Lys Glu Gly Thr Glu Lys Ile Lys Gln Trp Phe
    530                 535                 540

Asp Tyr Thr Leu Ala Ile Asn Arg Met Leu Lys Tyr Phe Leu Val Lys
545                 550                 555                 560

Glu Asn Lys Ile Lys Gly Asn Ser Leu Asp Thr Asn Ile Ser Glu Ala
                565                 570                 575
```

Leu Lys Thr Leu Ile Tyr Ser Asp Asp Ala Glu Trp Phe Lys Trp Tyr
580 585 590

Asp Ala Leu Arg Asn Tyr Leu Thr Gln Lys Pro Gln Asp Glu Ala Lys
595 600 605

Glu Asn Lys Leu Lys Leu Asn Phe Asp Asn Pro Ser Leu Ala Gly Gly
610 615 620

Trp Asp Val Asn Lys Glu Cys Ser Asn Phe Cys Val Ile Leu Lys Asp
625 630 635 640

Lys Asn Glu Lys Lys Tyr Leu Ala Ile Met Lys Lys Gly Glu Asn Thr
645 650 655

Leu Phe Gln Lys Glu Trp Thr Glu Gly Arg Gly Lys Asn Leu Thr Lys
660 665 670

Lys Ser Asn Pro Leu Phe Glu Ile Asn Asn Cys Glu Ile Leu Ser Lys
675 680 685

Met Glu Tyr Asp Phe Trp Ala Asp Val Ser Lys Met Ile Pro Lys Cys
690 695 700

Ser Thr Gln Leu Lys Ala Val Val Asn His Phe Lys Gln Ser Asp Asn
705 710 715 720

Glu Phe Ile Phe Pro Ile Gly Tyr Lys Val Thr Ser Gly Glu Lys Phe
725 730 735

Arg Glu Glu Cys Lys Ile Ser Lys Gln Asp Phe Glu Leu Asn Asn Lys
740 745 750

Val Phe Asn Lys Asn Glu Leu Ser Val Thr Ala Met Arg Tyr Asp Leu
755 760 765

Ser Ser Thr Gln Glu Lys Gln Tyr Ile Lys Ala Phe Gln Lys Glu Tyr
770 775 780

Trp Glu Leu Leu Phe Lys Gln Glu Lys Arg Asp Thr Lys Leu Thr Asn
785 790 795 800

Asn Glu Ile Phe Asn Glu Trp Ile Asn Phe Cys Asn Lys Lys Tyr Ser
805 810 815

Glu Leu Leu Ser Trp Glu Arg Lys Tyr Lys Asp Ala Leu Thr Asn Trp
820 825 830

Ile Asn Phe Cys Lys Tyr Phe Leu Ser Lys Tyr Pro Lys Thr Thr Leu
835 840 845

Phe Asn Tyr Ser Phe Lys Glu Ser Glu Asn Tyr Asn Ser Leu Asp Glu
850 855 860

Phe Tyr Arg Asp Val Asp Ile Cys Ser Tyr Lys Leu Asn Ile Asn Thr
865 870 875 880

Thr Ile Asn Lys Ser Ile Leu Asp Arg Leu Val Glu Glu Gly Lys Leu
885 890 895

Tyr Leu Phe Glu Ile Lys Asn Gln Asp Ser Asn Asp Gly Lys Ser Ile
900 905 910

```
Gly His Lys Asn Asn Leu His Thr Ile Tyr Trp Asn Ala Ile Phe Glu
       915                 920                 925

Asn Phe Asp Asn Arg Pro Lys Leu Asn Gly Glu Ala Glu Ile Phe Tyr
    930                 935                 940

Arg Lys Ala Ile Ser Lys Asp Lys Leu Gly Ile Val Lys Gly Lys Lys
945                 950                 955                 960

Thr Lys Asn Gly Thr Glu Ile Ile Lys Asn Tyr Arg Phe Ser Lys Glu
                965                 970                 975

Lys Phe Ile Leu His Val Pro Ile Thr Leu Asn Phe Cys Ser Asn Asn
           980                 985                 990

Glu Tyr Val Asn Asp Ile Val Asn Thr Lys Phe Tyr Asn Phe Ser Asn
       995                 1000                1005

Leu His  Phe Leu Gly Ile Asp  Arg Gly Glu Lys His  Leu Ala Tyr
    1010                1015                1020

Tyr Ser  Leu Val Asn Lys Asn  Gly Glu Ile Val Asp  Gln Gly Thr
    1025                1030                1035

Leu Asn  Leu Pro Phe Thr Asp  Lys Asp Gly Asn Gln  Arg Ser Ile
    1040                1045                1050

Lys Lys  Glu Lys Tyr Phe Tyr  Asn Lys Gln Glu Asp  Lys Trp Glu
    1055                1060                1065

Ala Lys  Glu Val Asp Cys Trp  Asn Tyr Asn Asp Leu  Leu Asp Ala
    1070                1075                1080

Met Ala  Ser Asn Arg Asp Met  Ala Arg Lys Asn Trp  Gln Arg Ile
    1085                1090                1095

Gly Thr  Ile Lys Glu Ala Lys  Asn Gly Tyr Val Ser  Leu Val Ile
    1100                1105                1110

Arg Lys  Ile Ala Asp Leu Ala  Val Asn Asn Glu Arg  Pro Ala Phe
    1115                1120                1125

Ile Val  Leu Glu Asp Leu Asn  Thr Gly Phe Lys Arg  Ser Arg Gln
    1130                1135                1140

Lys Ile  Asp Lys Ser Val Tyr  Gln Lys Phe Glu Leu  Ala Leu Ala
    1145                1150                1155

Lys Lys  Leu Asn Phe Leu Val  Asp Lys Asn Ala Lys  Arg Asp Glu
    1160                1165                1170

Ile Gly  Ser Pro Thr Lys Ala  Leu Gln Leu Thr Pro  Pro Val Asn
    1175                1180                1185

Asn Tyr  Gly Asp Ile Glu Asn  Lys Lys Gln Ala Gly  Ile Met Leu
    1190                1195                1200

Tyr Thr  Arg Ala Asn Tyr Thr  Ser Gln Thr Asp Pro  Ala Thr Gly
    1205                1210                1215

Trp Arg  Lys Thr Ile Tyr Leu  Lys Ala Gly Pro Glu  Glu Thr Thr
    1220                1225                1230
```

Tyr Lys Lys Asp Gly Lys Ile Lys Asn Lys Ser Val Lys Asp Gln
1235 1240 1245

Ile Ile Glu Thr Phe Thr Asp Ile Gly Phe Asp Gly Lys Asp Tyr
1250 1255 1260

Tyr Phe Glu Tyr Asp Lys Gly Glu Phe Val Asp Glu Lys Thr Gly
1265 1270 1275

Glu Ile Lys Pro Lys Lys Trp Arg Leu Tyr Ser Gly Glu Asn Gly
1280 1285 1290

Lys Ser Leu Asp Arg Phe Arg Gly Glu Arg Glu Lys Asp Lys Tyr
1295 1300 1305

Glu Trp Lys Ile Asp Lys Ile Asp Ile Val Lys Ile Leu Asp Asp
1310 1315 1320

Leu Phe Val Asn Phe Asp Lys Asn Ile Ser Leu Leu Lys Gln Leu
1325 1330 1335

Lys Glu Gly Val Glu Leu Thr Arg Asn Asn Glu His Gly Thr Gly
1340 1345 1350

Glu Ser Leu Arg Phe Ala Ile Asn Leu Ile Gln Gln Ile Arg Asn
1355 1360 1365

Thr Gly Asn Asn Glu Arg Asp Asn Asp Phe Ile Leu Ser Pro Val
1370 1375 1380

Arg Asp Glu Asn Gly Lys His Phe Asp Ser Arg Glu Tyr Trp Asp
1385 1390 1395

Lys Glu Thr Lys Gly Glu Lys Ile Ser Met Pro Ser Ser Gly Asp
1400 1405 1410

Ala Asn Gly Ala Phe Asn Ile Ala Arg Lys Gly Ile Ile Met Asn
1415 1420 1425

Ala His Ile Leu Ala Asn Ser Asp Ser Lys Asp Leu Ser Leu Phe
1430 1435 1440

Val Ser Asp Glu Glu Trp Asp Leu His Leu Asn Asn Lys Thr Glu
1445 1450 1455

Trp Lys Lys Gln Leu Asn Ile Phe Ser Ser Arg Lys Ala Met Ala
1460 1465 1470

Lys Arg Lys Lys
1475

<210> 69
<211> 1403
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選佛爾扣菌（Falkowbacteria）細菌序列

<400> 69
Met Leu Phe Phe Met Ser Thr Asp Ile Thr Asn Lys Pro Arg Glu Lys
1 5 10 15

Gly Val Phe Asp Asn Phe Thr Asn Leu Tyr Glu Phe Ser Lys Thr Leu
20 25 30

```
Thr Phe Gly Leu Ile Pro Leu Lys Trp Asp Asp Asn Lys Lys Met Ile
        35                  40                  45

Val Glu Asp Glu Asp Phe Ser Val Leu Arg Lys Tyr Gly Val Ile Glu
    50                  55                  60

Glu Asp Lys Arg Ile Ala Glu Ser Ile Lys Ile Ala Lys Phe Tyr Leu
65                  70                  75                  80

Asn Ile Leu His Arg Glu Leu Ile Gly Lys Val Leu Gly Ser Leu Lys
                85                  90                  95

Phe Glu Lys Lys Asn Leu Glu Asn Tyr Asp Arg Leu Leu Gly Glu Ile
            100                 105                 110

Glu Lys Asn Asn Lys Asn Glu Asn Ile Ser Glu Asp Lys Lys Lys Glu
        115                 120                 125

Ile Arg Lys Asn Phe Lys Lys Glu Leu Ser Ile Ala Gln Asp Ile Leu
    130                 135                 140

Leu Lys Lys Val Gly Glu Val Phe Glu Ser Asn Gly Ser Gly Ile Leu
145                 150                 155                 160

Ser Ser Lys Asn Cys Leu Asp Glu Leu Thr Lys Arg Phe Thr Arg Gln
                165                 170                 175

Glu Val Asp Lys Leu Arg Arg Glu Asn Lys Asp Ile Gly Val Glu Tyr
            180                 185                 190

Pro Asp Val Ala Tyr Arg Glu Lys Asp Gly Lys Glu Glu Thr Lys Ser
        195                 200                 205

Phe Phe Ala Met Asp Val Gly Tyr Leu Asp Asp Phe His Lys Asn Arg
    210                 215                 220

Lys Gln Leu Tyr Ser Val Lys Gly Lys Lys Asn Ser Leu Gly Arg Arg
225                 230                 235                 240

Ile Leu Asp Asn Phe Glu Ile Phe Cys Lys Asn Lys Lys Leu Tyr Glu
                245                 250                 255

Lys Tyr Lys Asn Leu Asp Ile Asp Phe Ser Glu Ile Glu Arg Asn Phe
            260                 265                 270

Asn Leu Thr Leu Glu Lys Val Phe Asp Phe Asp Asn Tyr Asn Glu Arg
        275                 280                 285

Leu Thr Gln Glu Gly Leu Asp Glu Tyr Ala Lys Ile Leu Gly Gly Glu
    290                 295                 300

Ser Asn Lys Gln Glu Arg Thr Ala Asn Ile His Gly Leu Asn Gln Ile
305                 310                 315                 320

Ile Asn Leu Tyr Ile Gln Lys Lys Gln Ser Glu Gln Lys Ala Glu Gln
                325                 330                 335

Lys Glu Thr Gly Lys Lys Lys Ile Lys Phe Asn Lys Lys Asp Tyr Pro
            340                 345                 350

Thr Phe Thr Cys Leu Gln Lys Gln Ile Leu Ser Gln Val Phe Arg Lys
        355                 360                 365
```

Glu Ile Ile Ile Glu Ser Asp Arg Asp Leu Ile Arg Glu Leu Lys Phe
370 375 380

Phe Val Glu Glu Ser Lys Glu Lys Val Asp Lys Ala Arg Gly Ile Ile
385 390 395 400

Glu Phe Leu Leu Asn His Glu Glu Asn Asp Ile Asp Leu Ala Met Val
405 410 415

Tyr Leu Pro Lys Ser Lys Ile Asn Ser Phe Val Tyr Lys Val Phe Lys
420 425 430

Glu Pro Gln Asp Phe Leu Ser Val Phe Gln Asp Gly Ala Ser Asn Leu
435 440 445

Asp Phe Val Ser Phe Asp Lys Ile Lys Thr His Leu Glu Asn Asn Lys
450 455 460

Leu Thr Tyr Lys Ile Phe Phe Lys Thr Leu Ile Lys Glu Asn His Asp
465 470 475 480

Phe Glu Ser Phe Leu Ile Leu Leu Gln Gln Glu Ile Asp Leu Leu Ile
485 490 495

Asp Gly Gly Glu Thr Val Thr Leu Gly Gly Lys Lys Glu Ser Ile Thr
500 505 510

Ser Leu Asp Glu Lys Lys Asn Arg Leu Lys Glu Lys Leu Gly Trp Phe
515 520 525

Glu Gly Lys Val Arg Glu Asn Glu Lys Met Lys Asp Glu Glu Glu Gly
530 535 540

Glu Phe Cys Ser Thr Val Leu Ala Tyr Ser Gln Ala Val Leu Asn Ile
545 550 555 560

Thr Lys Arg Ala Glu Ile Phe Trp Leu Asn Glu Lys Gln Asp Ala Lys
565 570 575

Val Gly Glu Asp Asn Lys Asp Met Ile Phe Tyr Lys Lys Phe Asp Glu
580 585 590

Phe Ala Asp Asp Gly Phe Ala Pro Phe Phe Tyr Phe Asp Lys Phe Gly
595 600 605

Asn Tyr Leu Lys Arg Arg Ser Arg Asn Thr Thr Lys Glu Ile Lys Leu
610 615 620

His Phe Gly Asn Asp Asp Leu Leu Glu Gly Trp Asp Met Asn Lys Glu
625 630 635 640

Pro Glu Tyr Trp Ser Phe Ile Leu Arg Asp Arg Asn Gln Tyr Tyr Leu
645 650 655

Gly Ile Gly Lys Lys Asp Gly Glu Ile Phe His Lys Lys Leu Gly Asn
660 665 670

Ser Val Glu Ala Val Lys Glu Ala Tyr Glu Leu Glu Asn Glu Ala Asp
675 680 685

Phe Tyr Glu Lys Ile Asp Tyr Lys Gln Leu Asn Ile Asp Arg Phe Glu
690 695 700

Gly Ile Ala Phe Pro Lys Lys Thr Lys Thr Glu Glu Ala Phe Arg Gln
705 710 715 720

Val Cys Lys Lys Arg Ala Asp Glu Phe Leu Gly Gly Asp Thr Tyr Glu
725 730 735

Phe Lys Ile Leu Leu Ala Ile Lys Lys Glu Tyr Asp Asp Phe Lys Ala
740 745 750

Arg Arg Gln Lys Glu Lys Asp Trp Asp Ser Lys Phe Ser Lys Glu Lys
755 760 765

Met Ser Lys Leu Ile Glu Tyr Tyr Ile Thr Cys Leu Gly Lys Arg Asp
770 775 780

Asp Trp Lys Arg Phe Asn Leu Asn Phe Arg Gln Pro Lys Glu Tyr Glu
785 790 795 800

Asp Arg Ser Asp Phe Val Arg His Ile Gln Arg Gln Ala Tyr Trp Ile
805 810 815

Asp Pro Arg Lys Val Ser Lys Asp Tyr Val Asp Lys Lys Val Ala Glu
820 825 830

Gly Glu Met Phe Leu Phe Lys Val His Asn Lys Asp Phe Tyr Asp Phe
835 840 845

Glu Arg Lys Ser Glu Asp Lys Lys Asn His Thr Ala Asn Leu Phe Thr
850 855 860

Gln Tyr Leu Leu Glu Leu Phe Ser Cys Glu Asn Ile Lys Asn Ile Lys
865 870 875 880

Ser Lys Asp Leu Ile Glu Ser Ile Phe Glu Leu Asp Gly Lys Ala Glu
885 890 895

Ile Arg Phe Arg Pro Lys Thr Asp Asp Val Lys Leu Lys Ile Tyr Gln
900 905 910

Lys Lys Gly Lys Asp Val Thr Tyr Ala Asp Lys Arg Asp Gly Asn Lys
915 920 925

Glu Lys Glu Val Ile Gln His Arg Arg Phe Ala Lys Asp Ala Leu Thr
930 935 940

Leu His Leu Lys Ile Arg Leu Asn Phe Gly Lys His Val Asn Leu Phe
945 950 955 960

Asp Phe Asn Lys Leu Val Asn Thr Glu Leu Phe Ala Lys Val Pro Val
965 970 975

Lys Ile Leu Gly Met Asp Arg Gly Glu Asn Asn Leu Ile Tyr Tyr Cys
980 985 990

Phe Leu Asp Glu His Gly Glu Ile Glu Asn Gly Lys Cys Gly Ser Leu
995 1000 1005

Asn Arg Val Gly Glu Gln Ile Ile Thr Leu Glu Asp Asp Lys Lys
1010 1015 1020

Val Lys Glu Pro Val Asp Tyr Phe Gln Leu Leu Val Asp Arg Glu
1025 1030 1035

```
Gly Gln  Arg Asp Trp Glu Gln  Lys Asn Trp Gln Lys  Met Thr Arg
    1040                1045                1050

Ile Lys  Asp Leu Lys Lys Ala  Tyr Leu Gly Asn Val  Val Ser Trp
    1055                1060                1065

Ile Ser  Lys Glu Met Leu Ser  Gly Ile Lys Glu Gly  Val Val Thr
    1070                1075                1080

Ile Gly  Val Leu Glu Asp Leu  Asn Ser Asn Phe Lys  Arg Thr Arg
    1085                1090                1095

Phe Phe  Arg Glu Arg Gln Val  Tyr Gln Gly Phe Glu  Lys Ala Leu
    1100                1105                1110

Val Asn  Lys Leu Gly Tyr Leu  Val Asp Lys Lys Tyr  Asp Asn Tyr
    1115                1120                1125

Arg Asn  Val Tyr Gln Phe Ala  Pro Ile Val Asp Ser  Val Glu Glu
    1130                1135                1140

Met Glu  Lys Asn Lys Gln Ile  Gly Thr Leu Val Tyr  Val Pro Ala
    1145                1150                1155

Ser Tyr  Thr Ser Lys Ile Cys  Pro His Pro Lys Cys  Gly Trp Arg
    1160                1165                1170

Glu Arg  Leu Tyr Met Lys Asn  Ser Ala Ser Lys Glu  Lys Ile Val
    1175                1180                1185

Gly Leu  Leu Lys Ser Asp Gly  Ile Lys Ile Ser Tyr  Asp Gln Lys
    1190                1195                1200

Asn Asp  Arg Phe Tyr Phe Glu  Tyr Gln Trp Glu Gln  Glu His Lys
    1205                1210                1215

Ser Asp  Gly Lys Lys Lys Lys  Tyr Ser Gly Val Asp  Lys Val Phe
    1220                1225                1230

Ser Asn  Val Ser Arg Met Arg  Trp Asp Val Glu Gln  Lys Lys Ser
    1235                1240                1245

Ile Asp  Phe Val Asp Gly Thr  Asp Gly Ser Ile Thr  Asn Lys Leu
    1250                1255                1260

Lys Ser  Leu Leu Lys Gly Lys  Gly Ile Glu Leu Asp  Asn Ile Asn
    1265                1270                1275

Gln Gln  Ile Val Asn Gln Gln  Lys Glu Leu Gly Val  Glu Phe Phe
    1280                1285                1290

Gln Ser  Ile Ile Phe Tyr Phe  Asn Leu Ile Met Gln  Ile Arg Asn
    1295                1300                1305

Tyr Asp  Lys Glu Lys Ser Gly  Ser Glu Ala Asp Tyr  Ile Gln Cys
    1310                1315                1320

Pro Ser  Cys Leu Phe Asp Ser  Arg Lys Pro Glu Met  Asn Gly Lys
    1325                1330                1335

Leu Ser  Ala Ile Thr Asn Gly  Asp Ala Asn Gly Ala  Tyr Asn Ile
    1340                1345                1350
```

```
Ala Arg Lys Gly Phe Met Gln Leu Cys Arg Ile Arg Glu Asn Pro
    1355                1360                1365

Gln Glu Pro Met Lys Leu Ile Thr Asn Arg Glu Trp Asp Glu Ala
    1370                1375                1380

Val Arg Glu Trp Asp Ile Tyr Ser Ala Ala Gln Lys Ile Pro Val
    1385                1390                1395

Leu Ser Glu Glu Asn
    1400
```

<210> 70
<211> 1373
<212> PRT
<213> 牛莫拉氏菌（Moraxella bovoculi）

<400> 70

```
Met Leu Phe Gln Asp Phe Thr His Leu Tyr Pro Leu Ser Lys Thr Val
1               5                   10                  15

Arg Phe Glu Leu Lys Pro Ile Asp Arg Thr Leu Glu His Ile His Ala
            20                  25                  30

Lys Asn Phe Leu Ser Gln Asp Glu Thr Met Ala Asp Met His Gln Lys
        35                  40                  45

Val Lys Val Ile Leu Asp Asp Tyr His Arg Asp Phe Ile Ala Asp Met
    50                  55                  60

Met Gly Glu Val Lys Leu Thr Lys Leu Ala Glu Phe Tyr Asp Val Tyr
65                  70                  75                  80

Leu Lys Phe Arg Lys Asn Pro Lys Asp Asp Glu Leu Gln Lys Gln Leu
                85                  90                  95

Lys Asp Leu Gln Ala Val Leu Arg Lys Glu Ile Val Lys Pro Ile Gly
            100                 105                 110

Asn Gly Gly Lys Tyr Lys Ala Gly Tyr Asp Arg Leu Phe Gly Ala Lys
        115                 120                 125

Leu Phe Lys Asp Gly Lys Glu Leu Gly Asp Leu Ala Lys Phe Val Ile
    130                 135                 140

Ala Gln Glu Gly Glu Ser Ser Pro Lys Leu Ala His Leu Ala His Phe
145                 150                 155                 160

Glu Lys Phe Ser Thr Tyr Phe Thr Gly Phe His Asp Asn Arg Lys Asn
                165                 170                 175

Met Tyr Ser Asp Glu Asp Lys His Thr Ala Ile Ala Tyr Arg Leu Ile
            180                 185                 190

His Glu Asn Leu Pro Arg Phe Ile Asp Asn Leu Gln Ile Leu Thr Thr
        195                 200                 205

Ile Lys Gln Lys His Ser Ala Leu Tyr Asp Gln Ile Ile Asn Glu Leu
    210                 215                 220

Thr Ala Ser Gly Leu Asp Val Ser Leu Ala Ser His Leu Asp Gly Tyr
225                 230                 235                 240
```

```
His Lys Leu Leu Thr Gln Glu Gly Ile Thr Ala Tyr Asn Thr Leu Leu
                245                 250                 255

Gly Gly Ile Ser Gly Glu Ala Gly Ser Pro Lys Ile Gln Gly Ile Asn
            260                 265                 270

Glu Leu Ile Asn Ser His His Asn Gln His Cys His Lys Ser Glu Arg
        275                 280                 285

Ile Ala Lys Leu Arg Pro Leu His Lys Gln Ile Leu Ser Asp Gly Met
    290                 295                 300

Ser Val Ser Phe Leu Pro Ser Lys Phe Ala Asp Asp Ser Glu Met Cys
305                 310                 315                 320

Gln Ala Val Asn Glu Phe Tyr Arg His Tyr Ala Asp Val Phe Ala Lys
                325                 330                 335

Val Gln Ser Leu Phe Asp Gly Phe Asp Asp His Gln Lys Asp Gly Ile
            340                 345                 350

Tyr Val Glu His Lys Asn Leu Asn Glu Leu Ser Lys Gln Ala Phe Gly
        355                 360                 365

Asp Phe Ala Leu Leu Gly Arg Val Leu Asp Gly Tyr Tyr Val Asp Val
    370                 375                 380

Val Asn Pro Glu Phe Asn Glu Arg Phe Ala Lys Ala Lys Thr Asp Asn
385                 390                 395                 400

Ala Lys Ala Lys Leu Thr Lys Glu Lys Asp Lys Phe Ile Lys Gly Val
                405                 410                 415

His Ser Leu Ala Ser Leu Glu Gln Ala Ile Glu His Tyr Thr Ala Arg
            420                 425                 430

His Asp Asp Glu Ser Val Gln Ala Gly Lys Leu Gly Gln Tyr Phe Lys
        435                 440                 445

His Gly Leu Ala Gly Val Asp Asn Pro Ile Gln Lys Ile His Asn Asn
    450                 455                 460

His Ser Thr Ile Lys Gly Phe Leu Glu Arg Glu Arg Pro Ala Gly Glu
465                 470                 475                 480

Arg Ala Leu Pro Lys Ile Lys Ser Gly Lys Asn Pro Glu Met Thr Gln
                485                 490                 495

Leu Arg Gln Leu Lys Glu Leu Leu Asp Asn Ala Leu Asn Val Ala His
            500                 505                 510

Phe Ala Lys Leu Leu Thr Thr Lys Thr Thr Leu Asp Asn Gln Asp Gly
        515                 520                 525

Asn Phe Tyr Gly Glu Phe Gly Val Leu Tyr Asp Glu Leu Ala Lys Ile
    530                 535                 540

Pro Thr Leu Tyr Asn Lys Val Arg Asp Tyr Leu Ser Gln Lys Pro Phe
545                 550                 555                 560

Ser Thr Glu Lys Tyr Lys Leu Asn Phe Gly Asn Pro Thr Leu Leu Asn
                565                 570                 575
```

```
Gly Trp Asp Leu Asn Lys Glu Lys Asp Asn Phe Gly Val Ile Leu Gln
            580                 585                 590

Lys Asp Gly Cys Tyr Tyr Leu Ala Leu Leu Asp Lys Ala His Lys Lys
        595                 600                 605

Val Phe Asp Asn Ala Pro Asn Thr Gly Lys Ser Ile Tyr Gln Lys Met
    610                 615                 620

Ile Tyr Lys Tyr Leu Glu Val Arg Lys Gln Phe Pro Lys Val Phe Phe
625                 630                 635                 640

Ser Lys Glu Ala Ile Ala Ile Asn Tyr His Pro Ser Lys Glu Leu Val
                645                 650                 655

Glu Ile Lys Asp Lys Gly Arg Gln Arg Ser Asp Asp Glu Arg Leu Lys
            660                 665                 670

Leu Tyr Arg Phe Ile Leu Glu Cys Leu Lys Ile His Pro Lys Tyr Asp
        675                 680                 685

Lys Lys Phe Glu Gly Ala Ile Gly Asp Ile Gln Leu Phe Lys Lys Asp
    690                 695                 700

Lys Lys Gly Arg Glu Val Pro Ile Ser Glu Lys Asp Leu Phe Asp Lys
705                 710                 715                 720

Ile Asn Gly Ile Phe Ser Ser Lys Pro Lys Leu Glu Met Glu Asp Phe
                725                 730                 735

Phe Ile Gly Glu Phe Lys Arg Tyr Asn Pro Ser Gln Asp Leu Val Asp
            740                 745                 750

Gln Tyr Asn Ile Tyr Lys Lys Ile Asp Ser Asn Asp Asn Arg Lys Lys
        755                 760                 765

Glu Asn Phe Tyr Asn Asn His Pro Lys Phe Lys Lys Asp Leu Val Arg
    770                 775                 780

Tyr Tyr Tyr Glu Ser Met Cys Lys His Glu Glu Trp Glu Glu Ser Phe
785                 790                 795                 800

Glu Phe Ser Lys Lys Leu Gln Asp Ile Gly Cys Tyr Val Asp Val Asn
                805                 810                 815

Glu Leu Phe Thr Glu Ile Glu Thr Arg Arg Leu Asn Tyr Lys Ile Ser
            820                 825                 830

Phe Cys Asn Ile Asn Ala Asp Tyr Ile Asp Glu Leu Val Glu Gln Gly
        835                 840                 845

Gln Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Pro Lys Ala
    850                 855                 860

His Gly Lys Pro Asn Leu His Thr Leu Tyr Phe Lys Ala Leu Phe Ser
865                 870                 875                 880

Glu Asp Asn Leu Ala Asp Pro Ile Tyr Lys Leu Asn Gly Glu Ala Gln
                885                 890                 895

Ile Phe Tyr Arg Lys Ala Ser Leu Asp Met Asn Glu Thr Thr Ile His
            900                 905                 910
```

```
Arg Ala Gly Glu Val Leu Glu Asn Lys Asn Pro Asp Asn Pro Lys Lys
        915                 920                 925

Arg Gln Phe Val Tyr Asp Ile Ile Lys Asp Lys Arg Tyr Thr Gln Asp
    930                 935                 940

Lys Phe Met Leu His Val Pro Ile Thr Met Asn Phe Gly Val Gln Gly
945                 950                 955                 960

Met Thr Ile Lys Glu Phe Asn Lys Lys Val Asn Gln Ser Ile Gln Gln
                965                 970                 975

Tyr Asp Glu Val Asn Val Ile Gly Ile Asp Arg Gly Glu Arg His Leu
            980                 985                 990

Leu Tyr Leu Thr Val Ile Asn Ser Lys Gly Glu Ile Leu Glu Gln Cys
        995                 1000                1005

Ser Leu  Asn Asp Ile Thr Thr  Ala Ser Ala Asn Gly  Thr Gln Met
    1010                1015                1020

Thr Thr  Pro Tyr His Lys Ile  Leu Asp Lys Arg Glu  Ile Glu Arg
    1025                1030                1035

Leu Asn  Ala Arg Val Gly Trp  Gly Glu Ile Glu Thr  Ile Lys Glu
    1040                1045                1050

Leu Lys  Ser Gly Tyr Leu Ser  His Val Val His Gln  Ile Ser Gln
    1055                1060                1065

Leu Met  Leu Lys Tyr Asn Ala  Ile Val Val Leu Glu  Asp Leu Asn
    1070                1075                1080

Phe Gly  Phe Lys Arg Gly Arg  Phe Lys Val Glu Lys  Gln Ile Tyr
    1085                1090                1095

Gln Asn  Phe Glu Asn Ala Leu  Ile Lys Lys Leu Asn  His Leu Val
    1100                1105                1110

Leu Lys  Asp Lys Ala Asp Asp  Glu Ile Gly Ser Tyr  Lys Asn Ala
    1115                1120                1125

Leu Gln  Leu Thr Asn Asn Phe  Thr Asp Leu Lys Ser  Ile Gly Lys
    1130                1135                1140

Gln Thr  Gly Phe Leu Phe Tyr  Val Pro Ala Trp Asn  Thr Ser Lys
    1145                1150                1155

Ile Asp  Pro Glu Thr Gly Phe  Val Asp Leu Leu Lys  Pro Arg Tyr
    1160                1165                1170

Glu Asn  Ile Ala Gln Ser Gln  Ala Phe Phe Gly Lys  Phe Asp Lys
    1175                1180                1185

Ile Cys  Tyr Asn Ala Asp Lys  Asp Tyr Phe Glu Phe  His Ile Asp
    1190                1195                1200

Tyr Ala  Lys Phe Thr Asp Lys  Ala Lys Asn Ser Arg  Gln Ile Trp
    1205                1210                1215

Thr Ile  Cys Ser His Gly Asp  Lys Arg Tyr Val Tyr  Asp Lys Thr
    1220                1225                1230
```

Ala Asn Gln Asn Lys Gly Ala Ala Lys Gly Ile Asn Val Asn Asp
1235 1240 1245

Glu Leu Lys Ser Leu Phe Ala Arg His His Ile Asn Glu Lys Gln
1250 1255 1260

Pro Asn Leu Val Met Asp Ile Cys Gln Asn Asn Asp Lys Glu Phe
1265 1270 1275

His Lys Ser Leu Met Tyr Leu Leu Lys Thr Leu Leu Ala Leu Arg
1280 1285 1290

Tyr Ser Asn Ala Ser Ser Asp Glu Asp Phe Ile Leu Ser Pro Val
1295 1300 1305

Ala Asn Asp Glu Gly Val Phe Phe Asn Ser Ala Leu Ala Asp Asp
1310 1315 1320

Thr Gln Pro Gln Asn Ala Asp Ala Asn Gly Ala Tyr His Ile Ala
1325 1330 1335

Leu Lys Gly Leu Trp Leu Leu Asn Glu Leu Lys Asn Ser Asp Asp
1340 1345 1350

Leu Asn Lys Val Lys Leu Ala Ile Asp Asn Gln Thr Trp Leu Asn
1355 1360 1365

Phe Ala Gln Asn Arg
1370

<210> 71
<211> 1352
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
Parcubacteria群細菌序列

<400> 71
Met Glu Asn Ile Phe Asp Gln Phe Ile Gly Lys Tyr Ser Leu Ser Lys
1 5 10 15

Thr Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr Glu Asp Phe Leu
20 25 30

Lys Ile Asn Lys Val Phe Glu Lys Asp Gln Thr Ile Asp Asp Ser Tyr
35 40 45

Asn Gln Ala Lys Phe Tyr Phe Asp Ser Leu His Gln Lys Phe Ile Asp
50 55 60

Ala Ala Leu Ala Ser Asp Lys Thr Ser Glu Leu Ser Phe Gln Asn Phe
65 70 75 80

Ala Asp Val Leu Glu Lys Gln Asn Lys Ile Ile Leu Asp Lys Lys Arg
85 90 95

Glu Met Gly Ala Leu Arg Lys Arg Asp Lys Asn Ala Val Gly Ile Asp
100 105 110

Arg Leu Gln Lys Glu Ile Asn Asp Ala Glu Asp Ile Ile Gln Lys Glu
115 120 125

Lys Glu Lys Ile Tyr Lys Asp Val Arg Thr Leu Phe Asp Asn Glu Ala
130 135 140

```
Glu Ser Trp Lys Thr Tyr Tyr Gln Glu Arg Glu Val Asp Gly Lys Lys
145                 150                 155                 160

Ile Thr Phe Ser Lys Ala Asp Leu Lys Gln Lys Gly Ala Asp Phe Leu
                165                 170                 175

Thr Ala Ala Gly Ile Leu Lys Val Leu Lys Tyr Glu Phe Pro Glu Glu
            180                 185                 190

Lys Glu Lys Glu Phe Gln Ala Lys Asn Gln Pro Ser Leu Phe Val Glu
        195                 200                 205

Glu Lys Glu Asn Pro Gly Gln Lys Arg Tyr Ile Phe Asp Ser Phe Asp
    210                 215                 220

Lys Phe Ala Gly Tyr Leu Thr Lys Phe Gln Gln Thr Lys Lys Asn Leu
225                 230                 235                 240

Tyr Ala Ala Asp Gly Thr Ser Thr Ala Val Ala Thr Arg Ile Ala Asp
                245                 250                 255

Asn Phe Ile Ile Phe His Gln Asn Thr Lys Val Phe Arg Asp Lys Tyr
            260                 265                 270

Lys Asn Asn His Thr Asp Leu Gly Phe Asp Glu Glu Asn Ile Phe Glu
        275                 280                 285

Ile Glu Arg Tyr Lys Asn Cys Leu Leu Gln Arg Glu Ile Glu His Ile
    290                 295                 300

Lys Asn Glu Asn Ser Tyr Asn Lys Ile Ile Gly Arg Ile Asn Lys Lys
305                 310                 315                 320

Ile Lys Glu Tyr Arg Asp Gln Lys Ala Lys Asp Thr Lys Leu Thr Lys
                325                 330                 335

Ser Asp Phe Pro Phe Phe Lys Asn Leu Asp Lys Gln Ile Leu Gly Glu
            340                 345                 350

Val Glu Lys Glu Lys Gln Leu Ile Glu Lys Thr Arg Glu Lys Thr Glu
        355                 360                 365

Glu Asp Val Leu Ile Glu Arg Phe Lys Glu Phe Ile Glu Asn Asn Glu
    370                 375                 380

Glu Arg Phe Thr Ala Ala Lys Lys Leu Met Asn Ala Phe Cys Asn Gly
385                 390                 395                 400

Glu Phe Glu Ser Glu Tyr Glu Gly Ile Tyr Leu Lys Asn Lys Ala Ile
                405                 410                 415

Asn Thr Ile Ser Arg Arg Trp Phe Val Ser Asp Arg Asp Phe Glu Leu
            420                 425                 430

Lys Leu Pro Gln Gln Lys Ser Lys Asn Lys Ser Glu Lys Asn Glu Pro
        435                 440                 445

Lys Val Lys Lys Phe Ile Ser Ile Ala Glu Ile Lys Asn Ala Val Glu
    450                 455                 460

Glu Leu Asp Gly Asp Ile Phe Lys Ala Val Phe Tyr Asp Lys Lys Ile
465                 470                 475                 480
```

```
Ile Ala Gln Gly Gly Ser Lys Leu Glu Gln Phe Leu Val Ile Trp Lys
                485                 490                 495

Tyr Glu Phe Glu Tyr Leu Phe Arg Asp Ile Glu Arg Glu Asn Gly Glu
            500                 505                 510

Lys Leu Leu Gly Tyr Asp Ser Cys Leu Lys Ile Ala Lys Gln Leu Gly
        515                 520                 525

Ile Phe Pro Gln Glu Lys Glu Ala Arg Glu Lys Ala Thr Ala Val Ile
    530                 535                 540

Lys Asn Tyr Ala Asp Ala Gly Leu Gly Ile Phe Gln Met Met Lys Tyr
545                 550                 555                 560

Phe Ser Leu Asp Asp Lys Asp Arg Lys Asn Thr Pro Gly Gln Leu Ser
                565                 570                 575

Thr Asn Phe Tyr Ala Glu Tyr Asp Gly Tyr Tyr Lys Asp Phe Glu Phe
            580                 585                 590

Ile Lys Tyr Tyr Asn Glu Phe Arg Asn Phe Ile Thr Lys Lys Pro Phe
        595                 600                 605

Asp Glu Asp Lys Ile Lys Leu Asn Phe Glu Asn Gly Ala Leu Leu Lys
    610                 615                 620

Gly Trp Asp Glu Asn Lys Glu Tyr Asp Phe Met Gly Val Ile Leu Lys
625                 630                 635                 640

Lys Glu Gly Arg Leu Tyr Leu Gly Ile Met His Lys Asn His Arg Lys
                645                 650                 655

Leu Phe Gln Ser Met Gly Asn Ala Lys Gly Asp Asn Ala Asn Arg Tyr
            660                 665                 670

Gln Lys Met Ile Tyr Lys Gln Ile Ala Asp Ala Ser Lys Asp Val Pro
        675                 680                 685

Arg Leu Leu Leu Thr Ser Lys Lys Ala Met Glu Lys Phe Lys Pro Ser
    690                 695                 700

Gln Glu Ile Leu Arg Ile Lys Lys Glu Lys Thr Phe Lys Arg Glu Ser
705                 710                 715                 720

Lys Asn Phe Ser Leu Arg Asp Leu His Ala Leu Ile Glu Tyr Tyr Arg
                725                 730                 735

Asn Cys Ile Pro Gln Tyr Ser Asn Trp Ser Phe Tyr Asp Phe Gln Phe
            740                 745                 750

Gln Asp Thr Gly Lys Tyr Gln Asn Ile Lys Glu Phe Thr Asp Asp Val
        755                 760                 765

Gln Lys Tyr Gly Tyr Lys Ile Ser Phe Arg Asp Ile Asp Asp Glu Tyr
    770                 775                 780

Ile Asn Gln Ala Leu Asn Glu Gly Lys Met Tyr Leu Phe Glu Val Val
785                 790                 795                 800

Asn Lys Asp Ile Tyr Asn Thr Lys Asn Gly Ser Lys Asn Leu His Thr
                805                 810                 815
```

```
Leu Tyr Phe Glu His Ile Leu Ser Ala Glu Asn Leu Asn Asp Pro Val
            820                 825                 830

Phe Lys Leu Ser Gly Met Ala Glu Ile Phe Gln Arg Gln Pro Ser Val
        835                 840                 845

Asn Glu Arg Glu Lys Ile Thr Thr Gln Lys Asn Gln Cys Ile Leu Asp
    850                 855                 860

Lys Gly Asp Arg Ala Tyr Lys Tyr Arg Arg Tyr Thr Glu Lys Lys Ile
865                 870                 875                 880

Met Phe His Met Ser Leu Val Leu Asn Thr Gly Lys Gly Glu Ile Lys
                885                 890                 895

Gln Val Gln Phe Asn Lys Ile Ile Asn Gln Arg Ile Ser Ser Ser Asp
            900                 905                 910

Asn Glu Met Arg Val Asn Val Ile Gly Ile Asp Arg Gly Glu Lys Asn
        915                 920                 925

Leu Leu Tyr Tyr Ser Val Val Lys Gln Asn Gly Glu Ile Ile Glu Gln
    930                 935                 940

Ala Ser Leu Asn Glu Ile Asn Gly Val Asn Tyr Arg Asp Lys Leu Ile
945                 950                 955                 960

Glu Arg Glu Lys Glu Arg Leu Lys Asn Arg Gln Ser Trp Lys Pro Val
                965                 970                 975

Val Lys Ile Lys Asp Leu Lys Lys Gly Tyr Ile Ser His Val Ile His
            980                 985                 990

Lys Ile Cys Gln Leu Ile Glu Lys Tyr Ser Ala Ile Val Val Leu Glu
        995                 1000                1005

Asp Leu  Asn Met Arg Phe Lys Gln Ile Arg Gly  Gly Ile Glu Arg
    1010                1015                1020

Ser Val  Tyr Gln Gln Phe Glu Lys Ala Leu Ile  Asp Lys Leu Gly
    1025                1030                1035

Tyr Leu  Val Phe Lys Asp Asn Arg Asp Leu Arg  Ala Pro Gly Gly
    1040                1045                1050

Val Leu  Asn Gly Tyr Gln Leu Ser Ala Pro Phe  Val Ser Phe Glu
    1055                1060                1065

Lys Met  Arg Lys Gln Thr Gly Ile Leu Phe Tyr  Thr Gln Ala Glu
    1070                1075                1080

Tyr Thr  Ser Lys Thr Asp Pro Ile Thr Gly Phe  Arg Lys Asn Val
    1085                1090                1095

Tyr Ile  Ser Asn Ser Ala Ser Leu Asp Lys Ile  Lys Glu Ala Val
    1100                1105                1110

Lys Lys  Phe Asp Ala Ile Gly Trp Asp Gly Lys  Glu Gln Ser Tyr
    1115                1120                1125

Phe Phe  Lys Tyr Asn Pro Tyr Asn Leu Ala Asp  Glu Lys Tyr Lys
    1130                1135                1140
```

Asn Ser Thr Val Ser Lys Glu Trp Ala Ile Phe Ala Ser Ala Pro
1145 1150 1155

Arg Ile Arg Arg Gln Lys Gly Glu Asp Gly Tyr Trp Lys Tyr Asp
1160 1165 1170

Arg Val Lys Val Asn Glu Glu Phe Glu Lys Leu Leu Lys Val Trp
1175 1180 1185

Asn Phe Val Asn Pro Lys Ala Thr Asp Ile Lys Gln Glu Ile Ile
1190 1195 1200

Lys Lys Glu Lys Ala Gly Asp Leu Gln Gly Glu Lys Glu Leu Asp
1205 1210 1215

Gly Arg Leu Arg Asn Phe Trp His Ser Phe Ile Tyr Leu Phe Asn
1220 1225 1230

Leu Val Leu Glu Leu Arg Asn Ser Phe Ser Leu Gln Ile Lys Ile
1235 1240 1245

Lys Ala Gly Glu Val Ile Ala Val Asp Glu Gly Val Asp Phe Ile
1250 1255 1260

Ala Ser Pro Val Lys Pro Phe Phe Thr Thr Pro Asn Pro Tyr Ile
1265 1270 1275

Pro Ser Asn Leu Cys Trp Leu Ala Val Glu Asn Ala Asp Ala Asn
1280 1285 1290

Gly Ala Tyr Asn Ile Ala Arg Lys Gly Val Met Ile Leu Lys Lys
1295 1300 1305

Ile Arg Glu His Ala Lys Lys Asp Pro Glu Phe Lys Lys Leu Pro
1310 1315 1320

Asn Leu Phe Ile Ser Asn Ala Glu Trp Asp Glu Ala Ala Arg Asp
1325 1330 1335

Trp Gly Lys Tyr Ala Gly Thr Thr Ala Leu Asn Leu Asp His
1340 1345 1350

<210> 72
<211> 1334
<212> PRT
<213> 溶糊精琥珀酸弧菌（Succinivibrio dextrinosolvens）

<400> 72
Met Ser Ser Leu Thr Lys Phe Thr Asn Lys Tyr Ser Lys Gln Leu Thr
1 5 10 15

Ile Lys Asn Glu Leu Ile Pro Val Gly Lys Thr Leu Glu Asn Ile Lys
20 25 30

Glu Asn Gly Leu Ile Asp Gly Asp Glu Gln Leu Asn Glu Asn Tyr Gln
35 40 45

Lys Ala Lys Ile Ile Val Asp Asp Phe Leu Arg Asp Phe Ile Asn Lys
50 55 60

Ala Leu Asn Asn Thr Gln Ile Gly Asn Trp Arg Glu Leu Ala Asp Ala
65 70 75 80

```
Leu Asn Lys Glu Asp Glu Asp Asn Ile Glu Lys Leu Gln Asp Lys Ile
                85                  90                  95

Arg Gly Ile Ile Val Ser Lys Phe Glu Thr Phe Asp Leu Phe Ser Ser
            100                 105                 110

Tyr Ser Ile Lys Lys Asp Glu Lys Ile Ile Asp Asp Asp Asn Asp Val
        115                 120                 125

Glu Glu Glu Glu Leu Asp Leu Gly Lys Lys Thr Ser Ser Phe Lys Tyr
    130                 135                 140

Ile Phe Lys Lys Asn Leu Phe Lys Leu Val Leu Pro Ser Tyr Leu Lys
145                 150                 155                 160

Thr Thr Asn Gln Asp Lys Leu Lys Ile Ile Ser Ser Phe Asp Asn Phe
                165                 170                 175

Ser Thr Tyr Phe Arg Gly Phe Phe Glu Asn Arg Lys Asn Ile Phe Thr
            180                 185                 190

Lys Lys Pro Ile Ser Thr Ser Ile Ala Tyr Arg Ile Val His Asp Asn
        195                 200                 205

Phe Pro Lys Phe Leu Asp Asn Ile Arg Cys Phe Asn Val Trp Gln Thr
    210                 215                 220

Glu Cys Pro Gln Leu Ile Val Lys Ala Asp Asn Tyr Leu Lys Ser Lys
225                 230                 235                 240

Asn Val Ile Ala Lys Asp Lys Ser Leu Ala Asn Tyr Phe Thr Val Gly
                245                 250                 255

Ala Tyr Asp Tyr Phe Leu Ser Gln Asn Gly Ile Asp Phe Tyr Asn Asn
            260                 265                 270

Ile Ile Gly Gly Leu Pro Ala Phe Ala Gly His Glu Lys Ile Gln Gly
        275                 280                 285

Leu Asn Glu Phe Ile Asn Gln Glu Cys Gln Lys Asp Ser Glu Leu Lys
    290                 295                 300

Ser Lys Leu Lys Asn Arg His Ala Phe Lys Met Ala Val Leu Phe Lys
305                 310                 315                 320

Gln Ile Leu Ser Asp Arg Glu Lys Ser Phe Val Ile Asp Glu Phe Glu
                325                 330                 335

Ser Asp Ala Gln Val Ile Asp Ala Val Lys Asn Phe Tyr Ala Glu Gln
            340                 345                 350

Cys Lys Asp Asn Asn Val Ile Phe Asn Leu Leu Asn Leu Ile Lys Asn
        355                 360                 365

Ile Ala Phe Leu Ser Asp Asp Glu Leu Asp Gly Ile Phe Ile Glu Gly
    370                 375                 380

Lys Tyr Leu Ser Ser Val Ser Gln Lys Leu Tyr Ser Asp Trp Ser Lys
385                 390                 395                 400

Leu Arg Asn Asp Ile Glu Asp Ser Ala Asn Ser Lys Gln Gly Asn Lys
                405                 410                 415
```

```
Glu Leu Ala Lys Lys Ile Lys Thr Asn Lys Gly Asp Val Glu Lys Ala
            420                 425                 430

Ile Ser Lys Tyr Glu Phe Ser Leu Ser Glu Leu Asn Ser Ile Val His
        435                 440                 445

Asp Asn Thr Lys Phe Ser Asp Leu Leu Ser Cys Thr Leu His Lys Val
    450                 455                 460

Ala Ser Glu Lys Leu Val Lys Val Asn Glu Gly Asp Trp Pro Lys His
465                 470                 475                 480

Leu Lys Asn Asn Glu Glu Lys Gln Lys Ile Lys Glu Pro Leu Asp Ala
                485                 490                 495

Leu Leu Glu Ile Tyr Asn Thr Leu Leu Ile Phe Asn Cys Lys Ser Phe
            500                 505                 510

Asn Lys Asn Gly Asn Phe Tyr Val Asp Tyr Asp Arg Cys Ile Asn Glu
        515                 520                 525

Leu Ser Ser Val Val Tyr Leu Tyr Asn Lys Thr Arg Asn Tyr Cys Thr
    530                 535                 540

Lys Lys Pro Tyr Asn Thr Asp Lys Phe Lys Leu Asn Phe Asn Ser Pro
545                 550                 555                 560

Gln Leu Gly Glu Gly Phe Ser Lys Ser Lys Glu Asn Asp Cys Leu Thr
                565                 570                 575

Leu Leu Phe Lys Lys Asp Asp Asn Tyr Tyr Val Gly Ile Ile Arg Lys
            580                 585                 590

Gly Ala Lys Ile Asn Phe Asp Asp Thr Gln Ala Ile Ala Asp Asn Thr
        595                 600                 605

Asp Asn Cys Ile Phe Lys Met Asn Tyr Phe Leu Leu Lys Asp Ala Lys
    610                 615                 620

Lys Phe Ile Pro Lys Cys Ser Ile Gln Leu Lys Glu Val Lys Ala His
625                 630                 635                 640

Phe Lys Lys Ser Glu Asp Asp Tyr Ile Leu Ser Asp Lys Glu Lys Phe
                645                 650                 655

Ala Ser Pro Leu Val Ile Lys Lys Ser Thr Phe Leu Leu Ala Thr Ala
            660                 665                 670

His Val Lys Gly Lys Lys Gly Asn Ile Lys Lys Phe Gln Lys Glu Tyr
        675                 680                 685

Ser Lys Glu Asn Pro Thr Glu Tyr Arg Asn Ser Leu Asn Glu Trp Ile
    690                 695                 700

Ala Phe Cys Lys Glu Phe Leu Lys Thr Tyr Lys Ala Ala Thr Ile Phe
705                 710                 715                 720

Asp Ile Thr Thr Leu Lys Lys Ala Glu Glu Tyr Ala Asp Ile Val Glu
                725                 730                 735

Phe Tyr Lys Asp Val Asp Asn Leu Cys Tyr Lys Leu Glu Phe Cys Pro
            740                 745                 750
```

```
Ile Lys Thr Ser Phe Ile Glu Asn Leu Ile Asp Asn Gly Asp Leu Tyr
        755                 760                 765

Leu Phe Arg Ile Asn Asn Lys Asp Phe Ser Ser Lys Ser Thr Gly Thr
    770                 775                 780

Lys Asn Leu His Thr Leu Tyr Leu Gln Ala Ile Phe Asp Glu Arg Asn
785                 790                 795                 800

Leu Asn Asn Pro Thr Ile Met Leu Asn Gly Gly Ala Glu Leu Phe Tyr
                805                 810                 815

Arg Lys Glu Ser Ile Glu Gln Lys Asn Arg Ile Thr His Lys Ala Gly
            820                 825                 830

Ser Ile Leu Val Asn Lys Val Cys Lys Asp Gly Thr Ser Leu Asp Asp
        835                 840                 845

Lys Ile Arg Asn Glu Ile Tyr Gln Tyr Glu Asn Lys Phe Ile Asp Thr
    850                 855                 860

Leu Ser Asp Glu Ala Lys Lys Val Leu Pro Asn Val Ile Lys Lys Glu
865                 870                 875                 880

Ala Thr His Asp Ile Thr Lys Asp Lys Arg Phe Thr Ser Asp Lys Phe
                885                 890                 895

Phe Phe His Cys Pro Leu Thr Ile Asn Tyr Lys Glu Gly Asp Thr Lys
            900                 905                 910

Gln Phe Asn Asn Glu Val Leu Ser Phe Leu Arg Gly Asn Pro Asp Ile
        915                 920                 925

Asn Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu Ile Tyr Val Thr
    930                 935                 940

Val Ile Asn Gln Lys Gly Glu Ile Leu Asp Ser Val Ser Phe Asn Thr
945                 950                 955                 960

Val Thr Asn Lys Ser Ser Lys Ile Glu Gln Thr Val Asp Tyr Glu Glu
                965                 970                 975

Lys Leu Ala Val Arg Glu Lys Glu Arg Ile Glu Ala Lys Arg Ser Trp
            980                 985                 990

Asp Ser Ile Ser Lys Ile Ala Thr Leu Lys Glu Gly Tyr Leu Ser Ala
        995                 1000                1005

Ile Val  His Glu Ile Cys Leu Leu Met Ile Lys  His Asn Ala Ile
    1010                1015                1020

Val Val  Leu Glu Asn Leu Asn Ala Gly Phe Lys  Arg Ile Arg Gly
    1025                1030                1035

Gly Leu  Ser Glu Lys Ser Val Tyr Gln Lys Phe  Glu Lys Met Leu
    1040                1045                1050

Ile Asn  Lys Leu Asn Tyr Phe Val Ser Lys Lys  Glu Ser Asp Trp
    1055                1060                1065

Asn Lys  Pro Ser Gly Leu Leu Asn Gly Leu Gln  Leu Ser Asp Gln
    1070                1075                1080
```

Phe Glu Ser Phe Glu Lys Leu Gly Ile Gln Ser Gly Phe Ile Phe
1085 1090 1095

Tyr Val Pro Ala Ala Tyr Thr Ser Lys Ile Asp Pro Thr Thr Gly
1100 1105 1110

Phe Ala Asn Val Leu Asn Leu Ser Lys Val Arg Asn Val Asp Ala
1115 1120 1125

Ile Lys Ser Phe Phe Ser Asn Phe Asn Glu Ile Ser Tyr Ser Lys
1130 1135 1140

Lys Glu Ala Leu Phe Lys Phe Ser Phe Asp Leu Asp Ser Leu Ser
1145 1150 1155

Lys Lys Gly Phe Ser Ser Phe Val Lys Phe Ser Lys Ser Lys Trp
1160 1165 1170

Asn Val Tyr Thr Phe Gly Glu Arg Ile Ile Lys Pro Lys Asn Lys
1175 1180 1185

Gln Gly Tyr Arg Glu Asp Lys Arg Ile Asn Leu Thr Phe Glu Met
1190 1195 1200

Lys Lys Leu Leu Asn Glu Tyr Lys Val Ser Phe Asp Leu Glu Asn
1205 1210 1215

Asn Leu Ile Pro Asn Leu Thr Ser Ala Asn Leu Lys Asp Thr Phe
1220 1225 1230

Trp Lys Glu Leu Phe Phe Ile Phe Lys Thr Thr Leu Gln Leu Arg
1235 1240 1245

Asn Ser Val Thr Asn Gly Lys Glu Asp Val Leu Ile Ser Pro Val
1250 1255 1260

Lys Asn Ala Lys Gly Glu Phe Phe Val Ser Gly Thr His Asn Lys
1265 1270 1275

Thr Leu Pro Gln Asp Cys Asp Ala Asn Gly Ala Tyr His Ile Ala
1280 1285 1290

Leu Lys Gly Leu Met Ile Leu Glu Arg Asn Asn Leu Val Arg Glu
1295 1300 1305

Glu Lys Asp Thr Lys Lys Ile Met Ala Ile Ser Asn Val Asp Trp
1310 1315 1320

Phe Glu Tyr Val Gln Lys Arg Arg Gly Val Leu
1325 1330

<210> 73
<211> 1331
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
Parcubacteria群細菌序列

<400> 73
Met Lys Pro Val Gly Lys Thr Glu Asp Phe Leu Lys Ile Asn Lys Val
1 5 10 15

Phe Glu Lys Asp Gln Thr Ile Asp Asp Ser Tyr Asn Gln Ala Lys Phe
20 25 30

```
Tyr Phe Asp Ser Leu His Gln Lys Phe Ile Asp Ala Ala Leu Ala Ser
        35                  40                  45

Asp Lys Thr Ser Glu Leu Ser Phe Gln Asn Phe Ala Asp Val Leu Glu
    50                  55                  60

Lys Gln Asn Lys Ile Ile Leu Asp Lys Lys Arg Glu Met Gly Ala Leu
65                  70                  75                  80

Arg Lys Arg Asp Lys Asn Ala Val Gly Ile Asp Arg Leu Gln Lys Glu
                85                  90                  95

Ile Asn Asp Ala Glu Asp Ile Ile Gln Lys Glu Lys Glu Lys Ile Tyr
            100                 105                 110

Lys Asp Val Arg Thr Leu Phe Asp Asn Glu Ala Glu Ser Trp Lys Thr
        115                 120                 125

Tyr Tyr Gln Glu Arg Glu Val Asp Gly Lys Lys Ile Thr Phe Ser Lys
    130                 135                 140

Ala Asp Leu Lys Gln Lys Gly Ala Asp Phe Leu Thr Ala Ala Gly Ile
145                 150                 155                 160

Leu Lys Val Leu Lys Tyr Glu Phe Pro Glu Glu Lys Glu Lys Glu Phe
                165                 170                 175

Gln Ala Lys Asn Gln Pro Ser Leu Phe Val Glu Glu Lys Glu Asn Pro
            180                 185                 190

Gly Gln Lys Arg Tyr Ile Phe Asp Ser Phe Asp Lys Phe Ala Gly Tyr
        195                 200                 205

Leu Thr Lys Phe Gln Gln Thr Lys Lys Asn Leu Tyr Ala Ala Asp Gly
    210                 215                 220

Thr Ser Thr Ala Val Ala Thr Arg Ile Ala Asp Asn Phe Ile Ile Phe
225                 230                 235                 240

His Gln Asn Thr Lys Val Phe Arg Asp Lys Tyr Lys Asn Asn His Thr
                245                 250                 255

Asp Leu Gly Phe Asp Glu Glu Asn Ile Phe Glu Ile Glu Arg Tyr Lys
            260                 265                 270

Asn Cys Leu Leu Gln Arg Glu Ile Glu His Ile Lys Asn Glu Asn Ser
        275                 280                 285

Tyr Asn Lys Ile Ile Gly Arg Ile Asn Lys Lys Ile Lys Glu Tyr Arg
    290                 295                 300

Asp Gln Lys Ala Lys Asp Thr Lys Leu Thr Lys Ser Asp Phe Pro Phe
305                 310                 315                 320

Phe Lys Asn Leu Asp Lys Gln Ile Leu Gly Glu Val Glu Lys Glu Lys
                325                 330                 335

Gln Leu Ile Glu Lys Thr Arg Glu Lys Thr Glu Glu Asp Val Leu Ile
            340                 345                 350

Glu Arg Phe Lys Glu Phe Ile Glu Asn Asn Glu Glu Arg Phe Thr Ala
        355                 360                 365
```

```
Ala Lys Lys Leu Met Asn Ala Phe Cys Asn Gly Glu Phe Glu Ser Glu
    370                 375                 380

Tyr Glu Gly Ile Tyr Leu Lys Asn Lys Ala Ile Asn Thr Ile Ser Arg
385                 390                 395                 400

Arg Trp Phe Val Ser Asp Arg Asp Phe Glu Leu Lys Leu Pro Gln Gln
                405                 410                 415

Lys Ser Lys Asn Lys Ser Glu Lys Asn Glu Pro Lys Val Lys Lys Phe
            420                 425                 430

Ile Ser Ile Ala Glu Ile Lys Asn Ala Val Glu Glu Leu Asp Gly Asp
        435                 440                 445

Ile Phe Lys Ala Val Phe Tyr Asp Lys Lys Ile Ile Ala Gln Gly Gly
    450                 455                 460

Ser Lys Leu Glu Gln Phe Leu Val Ile Trp Lys Tyr Glu Phe Glu Tyr
465                 470                 475                 480

Leu Phe Arg Asp Ile Glu Arg Glu Asn Gly Glu Lys Leu Leu Gly Tyr
                485                 490                 495

Asp Ser Cys Leu Lys Ile Ala Lys Gln Leu Gly Ile Phe Pro Gln Glu
            500                 505                 510

Lys Glu Ala Arg Glu Lys Ala Thr Ala Val Ile Lys Asn Tyr Ala Asp
        515                 520                 525

Ala Gly Leu Gly Ile Phe Gln Met Met Lys Tyr Phe Ser Leu Asp Asp
    530                 535                 540

Lys Asp Arg Lys Asn Thr Pro Gly Gln Leu Ser Thr Asn Phe Tyr Ala
545                 550                 555                 560

Glu Tyr Asp Gly Tyr Tyr Lys Asp Phe Glu Phe Ile Lys Tyr Tyr Asn
                565                 570                 575

Glu Phe Arg Asn Phe Ile Thr Lys Lys Pro Phe Asp Glu Asp Lys Ile
            580                 585                 590

Lys Leu Asn Phe Glu Asn Gly Ala Leu Leu Lys Gly Trp Asp Glu Asn
        595                 600                 605

Lys Glu Tyr Asp Phe Met Gly Val Ile Leu Lys Lys Glu Gly Arg Leu
    610                 615                 620

Tyr Leu Gly Ile Met His Lys Asn His Arg Lys Leu Phe Gln Ser Met
625                 630                 635                 640

Gly Asn Ala Lys Gly Asp Asn Ala Asn Arg Tyr Gln Lys Met Ile Tyr
                645                 650                 655

Lys Gln Ile Ala Asp Ala Ser Lys Asp Val Pro Arg Leu Leu Leu Thr
            660                 665                 670

Ser Lys Lys Ala Met Glu Lys Phe Lys Pro Ser Gln Glu Ile Leu Arg
        675                 680                 685

Ile Lys Lys Glu Lys Thr Phe Lys Arg Glu Ser Lys Asn Phe Ser Leu
    690                 695                 700
```

```
Arg Asp Leu His Ala Leu Ile Glu Tyr Tyr Arg Asn Cys Ile Pro Gln
705                 710                 715                 720

Tyr Ser Asn Trp Ser Phe Tyr Asp Phe Gln Phe Gln Asp Thr Gly Lys
                725                 730                 735

Tyr Gln Asn Ile Lys Glu Phe Thr Asp Asp Val Gln Lys Tyr Gly Tyr
            740                 745                 750

Lys Ile Ser Phe Arg Asp Ile Asp Asp Glu Tyr Ile Asn Gln Ala Leu
        755                 760                 765

Asn Glu Gly Lys Met Tyr Leu Phe Glu Val Val Asn Lys Asp Ile Tyr
    770                 775                 780

Asn Thr Lys Asn Gly Ser Lys Asn Leu His Thr Leu Tyr Phe Glu His
785                 790                 795                 800

Ile Leu Ser Ala Glu Asn Leu Asn Asp Pro Val Phe Lys Leu Ser Gly
                805                 810                 815

Met Ala Glu Ile Phe Gln Arg Gln Pro Ser Val Asn Glu Arg Glu Lys
            820                 825                 830

Ile Thr Thr Gln Lys Asn Gln Cys Ile Leu Asp Lys Gly Asp Arg Ala
        835                 840                 845

Tyr Lys Tyr Arg Arg Tyr Thr Glu Lys Lys Ile Met Phe His Met Ser
    850                 855                 860

Leu Val Leu Asn Thr Gly Lys Gly Glu Ile Lys Gln Val Gln Phe Asn
865                 870                 875                 880

Lys Ile Ile Asn Gln Arg Ile Ser Ser Ser Asp Asn Glu Met Arg Val
                885                 890                 895

Asn Val Ile Gly Ile Asp Arg Gly Glu Lys Asn Leu Leu Tyr Tyr Ser
            900                 905                 910

Val Val Lys Gln Asn Gly Glu Ile Ile Glu Gln Ala Ser Leu Asn Glu
        915                 920                 925

Ile Asn Gly Val Asn Tyr Arg Asp Lys Leu Ile Glu Arg Glu Lys Glu
    930                 935                 940

Arg Leu Lys Asn Arg Gln Ser Trp Lys Pro Val Val Lys Ile Lys Asp
945                 950                 955                 960

Leu Lys Lys Gly Tyr Ile Ser His Val Ile His Lys Ile Cys Gln Leu
                965                 970                 975

Ile Glu Lys Tyr Ser Ala Ile Val Val Leu Glu Asp Leu Asn Met Arg
            980                 985                 990

Phe Lys Gln Ile Arg Gly Gly Ile Glu Arg Ser Val Tyr Gln Gln Phe
        995                 1000                1005

Glu Lys Ala Leu Ile Asp Lys Leu Gly Tyr Leu Val Phe Lys Asp
    1010                1015                1020

Asn Arg Asp Leu Arg Ala Pro Gly Gly Val Leu Asn Gly Tyr Gln
    1025                1030                1035
```

```
Leu Ser  Ala Pro Phe Val Ser  Phe Glu Lys Met Arg  Lys Gln Thr
    1040                1045                1050

Gly Ile  Leu Phe Tyr Thr Gln  Ala Glu Tyr Thr Ser  Lys Thr Asp
    1055                1060                1065

Pro Ile  Thr Gly Phe Arg Lys  Asn Val Tyr Ile Ser  Asn Ser Ala
    1070                1075                1080

Ser Leu  Asp Lys Ile Lys Glu  Ala Val Lys Lys Phe  Asp Ala Ile
    1085                1090                1095

Gly Trp  Asp Gly Lys Glu Gln  Ser Tyr Phe Phe Lys  Tyr Asn Pro
    1100                1105                1110

Tyr Asn  Leu Ala Asp Glu Lys  Tyr Lys Asn Ser Thr  Val Ser Lys
    1115                1120                1125

Glu Trp  Ala Ile Phe Ala Ser  Ala Pro Arg Ile Arg  Arg Gln Lys
    1130                1135                1140

Gly Glu  Asp Gly Tyr Trp Lys  Tyr Asp Arg Val Lys  Val Asn Glu
    1145                1150                1155

Glu Phe  Glu Lys Leu Leu Lys  Val Trp Asn Phe Val  Asn Pro Lys
    1160                1165                1170

Ala Thr  Asp Ile Lys Gln Glu  Ile Ile Lys Lys Glu  Lys Ala Gly
    1175                1180                1185

Asp Leu  Gln Gly Glu Lys Glu  Leu Asp Gly Arg Leu  Arg Asn Phe
    1190                1195                1200

Trp His  Ser Phe Ile Tyr Leu  Phe Asn Leu Val Leu  Glu Leu Arg
    1205                1210                1215

Asn Ser  Phe Ser Leu Gln Ile  Lys Ile Lys Ala Gly  Glu Val Ile
    1220                1225                1230

Ala Val  Asp Glu Gly Val Asp  Phe Ile Ala Ser Pro  Val Lys Pro
    1235                1240                1245

Phe Phe  Thr Thr Pro Asn Pro  Tyr Ile Pro Ser Asn  Leu Cys Trp
    1250                1255                1260

Leu Ala  Val Glu Asn Ala Asp  Ala Asn Gly Ala Tyr  Asn Ile Ala
    1265                1270                1275

Arg Lys  Gly Val Met Ile Leu  Lys Lys Ile Arg Glu  His Ala Lys
    1280                1285                1290

Lys Asp  Pro Glu Phe Lys Lys  Leu Pro Asn Leu Phe  Ile Ser Asn
    1295                1300                1305

Ala Glu  Trp Asp Glu Ala Ala  Arg Asp Trp Gly Lys  Tyr Ala Gly
    1310                1315                1320

Thr Thr  Ala Leu Asn Leu Asp  His
    1325                1330
```

<210> 74
<211> 1323

<212> PRT
<213> 解糖腺普雷沃菌（Prevotella disiens）

<400> 74

```
Met Glu Asn Tyr Gln Glu Phe Thr Asn Leu Phe Gln Leu Asn Lys Thr
1               5                  10                  15

Leu Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Cys Glu Leu Leu Glu
            20                  25                  30

Glu Gly Lys Ile Phe Ala Ser Gly Ser Phe Leu Glu Lys Asp Lys Val
        35                  40                  45

Arg Ala Asp Asn Val Ser Tyr Val Lys Lys Glu Ile Asp Lys Lys His
    50                  55                  60

Lys Ile Phe Ile Glu Glu Thr Leu Ser Ser Phe Ser Ile Ser Asn Asp
65                  70                  75                  80

Leu Leu Lys Gln Tyr Phe Asp Cys Tyr Asn Glu Leu Lys Ala Phe Lys
                85                  90                  95

Lys Asp Cys Lys Ser Asp Glu Glu Glu Val Lys Lys Thr Ala Leu Arg
            100                 105                 110

Asn Lys Cys Thr Ser Ile Gln Arg Ala Met Arg Glu Ala Ile Ser Gln
        115                 120                 125

Ala Phe Leu Lys Ser Pro Gln Lys Lys Leu Leu Ala Ile Lys Asn Leu
    130                 135                 140

Ile Glu Asn Val Phe Lys Ala Asp Glu Asn Val Gln His Phe Ser Glu
145                 150                 155                 160

Phe Thr Ser Tyr Phe Ser Gly Phe Glu Thr Asn Arg Glu Asn Phe Tyr
                165                 170                 175

Ser Asp Glu Glu Lys Ser Thr Ser Ile Ala Tyr Arg Leu Val His Asp
            180                 185                 190

Asn Leu Pro Ile Phe Ile Lys Asn Ile Tyr Ile Phe Glu Lys Leu Lys
        195                 200                 205

Glu Gln Phe Asp Ala Lys Thr Leu Ser Glu Ile Phe Glu Asn Tyr Lys
    210                 215                 220

Leu Tyr Val Ala Gly Ser Ser Leu Asp Glu Val Phe Ser Leu Glu Tyr
225                 230                 235                 240

Phe Asn Asn Thr Leu Thr Gln Lys Gly Ile Asp Asn Tyr Asn Ala Val
                245                 250                 255

Ile Gly Lys Ile Val Lys Glu Asp Lys Gln Glu Ile Gln Gly Leu Asn
            260                 265                 270

Glu His Ile Asn Leu Tyr Asn Gln Lys His Lys Asp Arg Arg Leu Pro
        275                 280                 285

Phe Phe Ile Ser Leu Lys Lys Gln Ile Leu Ser Asp Arg Glu Ala Leu
    290                 295                 300

Ser Trp Leu Pro Asp Met Phe Lys Asn Asp Ser Glu Val Ile Lys Ala
305                 310                 315                 320
```

Leu Lys Gly Phe Tyr Ile Glu Asp Gly Phe Glu Asn Asn Val Leu Thr
325 330 335

Pro Leu Ala Thr Leu Leu Ser Ser Leu Asp Lys Tyr Asn Leu Asn Gly
340 345 350

Ile Phe Ile Arg Asn Asn Glu Ala Leu Ser Ser Leu Ser Gln Asn Val
355 360 365

Tyr Arg Asn Phe Ser Ile Asp Glu Ala Ile Asp Ala Asn Ala Glu Leu
370 375 380

Gln Thr Phe Asn Asn Tyr Glu Leu Ile Ala Asn Ala Leu Arg Ala Lys
385 390 395 400

Ile Lys Lys Glu Thr Lys Gln Gly Arg Lys Ser Phe Glu Lys Tyr Glu
405 410 415

Glu Tyr Ile Asp Lys Lys Val Lys Ala Ile Asp Ser Leu Ser Ile Gln
420 425 430

Glu Ile Asn Glu Leu Val Glu Asn Tyr Val Ser Glu Phe Asn Ser Asn
435 440 445

Ser Gly Asn Met Pro Arg Lys Val Glu Asp Tyr Phe Ser Leu Met Arg
450 455 460

Lys Gly Asp Phe Gly Ser Asn Asp Leu Ile Glu Asn Ile Lys Thr Lys
465 470 475 480

Leu Ser Ala Ala Glu Lys Leu Leu Gly Thr Lys Tyr Gln Glu Thr Ala
485 490 495

Lys Asp Ile Phe Lys Lys Asp Glu Asn Ser Lys Leu Ile Lys Glu Leu
500 505 510

Leu Asp Ala Thr Lys Gln Phe Gln His Phe Ile Lys Pro Leu Leu Gly
515 520 525

Thr Gly Glu Glu Ala Asp Arg Asp Leu Val Phe Tyr Gly Asp Phe Leu
530 535 540

Pro Leu Tyr Glu Lys Phe Glu Glu Leu Thr Leu Leu Tyr Asn Lys Val
545 550 555 560

Arg Asn Arg Leu Thr Gln Lys Pro Tyr Ser Lys Asp Lys Ile Arg Leu
565 570 575

Cys Phe Asn Lys Pro Lys Leu Met Thr Gly Trp Val Asp Ser Lys Thr
580 585 590

Glu Lys Ser Asp Asn Gly Thr Gln Tyr Gly Gly Tyr Leu Phe Arg Lys
595 600 605

Lys Asn Glu Ile Gly Glu Tyr Asp Tyr Phe Leu Gly Ile Ser Ser Lys
610 615 620

Ala Gln Leu Phe Arg Lys Asn Glu Ala Val Ile Gly Asp Tyr Glu Arg
625 630 635 640

Leu Asp Tyr Tyr Gln Pro Lys Ala Asn Thr Ile Tyr Gly Ser Ala Tyr
645 650 655

```
Glu Gly Glu Asn Ser Tyr Lys Glu Asp Lys Lys Arg Leu Asn Lys Val
            660                 665                 670

Ile Ile Ala Tyr Ile Glu Gln Ile Lys Gln Thr Asn Ile Lys Lys Ser
        675                 680                 685

Ile Ile Glu Ser Ile Ser Lys Tyr Pro Asn Ile Ser Asp Asp Asp Lys
    690                 695                 700

Val Thr Pro Ser Ser Leu Leu Glu Lys Ile Lys Lys Val Ser Ile Asp
705                 710                 715                 720

Ser Tyr Asn Gly Ile Leu Ser Phe Lys Ser Phe Gln Ser Val Asn Lys
                725                 730                 735

Glu Val Ile Asp Asn Leu Leu Lys Thr Ile Ser Pro Leu Lys Asn Lys
            740                 745                 750

Ala Glu Phe Leu Asp Leu Ile Asn Lys Asp Tyr Gln Ile Phe Thr Glu
        755                 760                 765

Val Gln Ala Val Ile Asp Glu Ile Cys Lys Gln Lys Thr Phe Ile Tyr
    770                 775                 780

Phe Pro Ile Ser Asn Val Glu Leu Glu Lys Glu Met Gly Asp Lys Asp
785                 790                 795                 800

Lys Pro Leu Cys Leu Phe Gln Ile Ser Asn Lys Asp Leu Ser Phe Ala
                805                 810                 815

Lys Thr Phe Ser Ala Asn Leu Arg Lys Lys Arg Gly Ala Glu Asn Leu
            820                 825                 830

His Thr Met Leu Phe Lys Ala Leu Met Glu Gly Asn Gln Asp Asn Leu
        835                 840                 845

Asp Leu Gly Ser Gly Ala Ile Phe Tyr Arg Ala Lys Ser Leu Asp Gly
    850                 855                 860

Asn Lys Pro Thr His Pro Ala Asn Glu Ala Ile Lys Cys Arg Asn Val
865                 870                 875                 880

Ala Asn Lys Asp Lys Val Ser Leu Phe Thr Tyr Asp Ile Tyr Lys Asn
                885                 890                 895

Arg Arg Tyr Met Glu Asn Lys Phe Leu Phe His Leu Ser Ile Val Gln
            900                 905                 910

Asn Tyr Lys Ala Ala Asn Asp Ser Ala Gln Leu Asn Ser Ser Ala Thr
        915                 920                 925

Glu Tyr Ile Arg Lys Ala Asp Asp Leu His Ile Ile Gly Ile Asp Arg
    930                 935                 940

Gly Glu Arg Asn Leu Leu Tyr Tyr Ser Val Ile Asp Met Lys Gly Asn
945                 950                 955                 960

Ile Val Glu Gln Asp Ser Leu Asn Ile Ile Arg Asn Asn Asp Leu Glu
                965                 970                 975

Thr Asp Tyr His Asp Leu Leu Asp Lys Arg Glu Lys Glu Arg Lys Ala
            980                 985                 990
```

```
Asn Arg Gln Asn Trp Glu Ala Val Glu Gly Ile Lys Asp Leu Lys Lys
        995                1000                1005

Gly Tyr  Leu Ser Gln Ala Val  His Gln Ile Ala Gln  Leu Met Leu
    1010                1015                1020

Lys Tyr  Asn Ala Ile Ile Ala  Leu Glu Asp Leu Gly  Gln Met Phe
    1025                1030                1035

Val Thr  Arg Gly Gln Lys Ile  Glu Lys Ala Val Tyr  Gln Gln Phe
    1040                1045                1050

Glu Lys  Ser Leu Val Asp Lys  Leu Ser Tyr Leu Val  Asp Lys Lys
    1055                1060                1065

Arg Pro  Tyr Asn Glu Leu Gly  Gly Ile Leu Lys Ala  Tyr Gln Leu
    1070                1075                1080

Ala Ser  Ser Ile Thr Lys Asn  Asn Ser Asp Lys Gln  Asn Gly Phe
    1085                1090                1095

Leu Phe  Tyr Val Pro Ala Trp  Asn Thr Ser Lys Ile  Asp Pro Val
    1100                1105                1110

Thr Gly  Phe Thr Asp Leu Leu  Arg Pro Lys Ala Met  Thr Ile Lys
    1115                1120                1125

Glu Ala  Gln Asp Phe Phe Gly  Ala Phe Asp Asn Ile  Ser Tyr Asn
    1130                1135                1140

Asp Lys  Gly Tyr Phe Glu Phe  Glu Thr Asn Tyr Asp  Lys Phe Lys
    1145                1150                1155

Ile Arg  Met Lys Ser Ala Gln  Thr Arg Trp Thr Ile  Cys Thr Phe
    1160                1165                1170

Gly Asn  Arg Ile Lys Arg Lys  Lys Asp Lys Asn Tyr  Trp Asn Tyr
    1175                1180                1185

Glu Glu  Val Glu Leu Thr Glu  Glu Phe Lys Lys Leu  Phe Lys Asp
    1190                1195                1200

Ser Asn  Ile Asp Tyr Glu Asn  Cys Asn Leu Lys Glu  Glu Ile Gln
    1205                1210                1215

Asn Lys  Asp Asn Arg Lys Phe  Phe Asp Asp Leu Ile  Lys Leu Leu
    1220                1225                1230

Gln Leu  Thr Leu Gln Met Arg  Asn Ser Asp Asp Lys  Gly Asn Asp
    1235                1240                1245

Tyr Ile  Ile Ser Pro Val Ala  Asn Ala Glu Gly Gln  Phe Phe Asp
    1250                1255                1260

Ser Arg  Asn Gly Asp Lys Lys  Leu Pro Leu Asp Ala  Asp Ala Asn
    1265                1270                1275

Gly Ala  Tyr Asn Ile Ala Arg  Lys Gly Leu Trp Asn  Ile Arg Gln
    1280                1285                1290

Ile Lys  Gln Thr Lys Asn Asp  Lys Lys Leu Asn Leu  Ser Ile Ser
    1295                1300                1305
```

Ser Thr Glu Trp Leu Asp Phe Val Arg Glu Lys Pro Tyr Leu Lys
1310 1315 1320

<210> 75
<211> 1318
<212> PRT
<213> 嗜鰓黃桿菌（Flavobacterium branchiophilum）

<400> 75
Met Thr Asn Lys Phe Thr Asn Gln Tyr Ser Leu Ser Lys Thr Leu Arg
1 5 10 15

Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Phe Ile Gln Glu Lys
20 25 30

Gly Leu Leu Ser Gln Asp Lys Gln Arg Ala Glu Ser Tyr Gln Glu Met
35 40 45

Lys Lys Thr Ile Asp Lys Phe His Lys Tyr Phe Ile Asp Leu Ala Leu
50 55 60

Ser Asn Ala Lys Leu Thr His Leu Glu Thr Tyr Leu Glu Leu Tyr Asn
65 70 75 80

Lys Ser Ala Glu Thr Lys Lys Glu Gln Lys Phe Lys Asp Asp Leu Lys
85 90 95

Lys Val Gln Asp Asn Leu Arg Lys Glu Ile Val Lys Ser Phe Ser Asp
100 105 110

Gly Asp Ala Lys Ser Ile Phe Ala Ile Leu Asp Lys Lys Glu Leu Ile
115 120 125

Thr Val Glu Leu Glu Lys Trp Phe Glu Asn Asn Glu Gln Lys Asp Ile
130 135 140

Tyr Phe Asp Glu Lys Phe Lys Thr Phe Thr Thr Tyr Phe Thr Gly Phe
145 150 155 160

His Gln Asn Arg Lys Asn Met Tyr Ser Val Glu Pro Asn Ser Thr Ala
165 170 175

Ile Ala Tyr Arg Leu Ile His Glu Asn Leu Pro Lys Phe Leu Glu Asn
180 185 190

Ala Lys Ala Phe Glu Lys Ile Lys Gln Val Glu Ser Leu Gln Val Asn
195 200 205

Phe Arg Glu Leu Met Gly Glu Phe Gly Asp Glu Gly Leu Ile Phe Val
210 215 220

Asn Glu Leu Glu Glu Met Phe Gln Ile Asn Tyr Tyr Asn Asp Val Leu
225 230 235 240

Ser Gln Asn Gly Ile Thr Ile Tyr Asn Ser Ile Ile Ser Gly Phe Thr
245 250 255

Lys Asn Asp Ile Lys Tyr Lys Gly Leu Asn Glu Tyr Ile Asn Asn Tyr
260 265 270

Asn Gln Thr Lys Asp Lys Lys Asp Arg Leu Pro Lys Leu Lys Gln Leu
275 280 285

Tyr Lys Gln Ile Leu Ser Asp Arg Ile Ser Leu Ser Phe Leu Pro Asp
290 295 300

```
Ala Phe Thr Asp Gly Lys Gln Val Leu Lys Ala Ile Phe Asp Phe Tyr
305                 310                 315                 320

Lys Ile Asn Leu Leu Ser Tyr Thr Ile Glu Gly Gln Glu Glu Ser Gln
                325                 330                 335

Asn Leu Leu Leu Leu Ile Arg Gln Thr Ile Glu Asn Leu Ser Ser Phe
            340                 345                 350

Asp Thr Gln Lys Ile Tyr Leu Lys Asn Asp Thr His Leu Thr Thr Ile
        355                 360                 365

Ser Gln Gln Val Phe Gly Asp Phe Ser Val Phe Ser Thr Ala Leu Asn
    370                 375                 380

Tyr Trp Tyr Glu Thr Lys Val Asn Pro Lys Phe Glu Thr Glu Tyr Ser
385                 390                 395                 400

Lys Ala Asn Glu Lys Lys Arg Glu Ile Leu Asp Lys Ala Lys Ala Val
                405                 410                 415

Phe Thr Lys Gln Asp Tyr Phe Ser Ile Ala Phe Leu Gln Glu Val Leu
            420                 425                 430

Ser Glu Tyr Ile Leu Thr Leu Asp His Thr Ser Asp Ile Val Lys Lys
        435                 440                 445

His Ser Ser Asn Cys Ile Ala Asp Tyr Phe Lys Asn His Phe Val Ala
    450                 455                 460

Lys Lys Glu Asn Glu Thr Asp Lys Thr Phe Asp Phe Ile Ala Asn Ile
465                 470                 475                 480

Thr Ala Lys Tyr Gln Cys Ile Gln Gly Ile Leu Glu Asn Ala Asp Gln
                485                 490                 495

Tyr Glu Asp Glu Leu Lys Gln Asp Gln Lys Leu Ile Asp Asn Leu Lys
            500                 505                 510

Phe Phe Leu Asp Ala Ile Leu Glu Leu Leu His Phe Ile Lys Pro Leu
        515                 520                 525

His Leu Lys Ser Glu Ser Ile Thr Glu Lys Asp Thr Ala Phe Tyr Asp
    530                 535                 540

Val Phe Glu Asn Tyr Tyr Glu Ala Leu Ser Leu Leu Thr Pro Leu Tyr
545                 550                 555                 560

Asn Met Val Arg Asn Tyr Val Thr Gln Lys Pro Tyr Ser Thr Glu Lys
                565                 570                 575

Ile Lys Leu Asn Phe Glu Asn Ala Gln Leu Leu Asn Gly Trp Asp Ala
            580                 585                 590

Asn Lys Glu Gly Asp Tyr Leu Thr Thr Ile Leu Lys Lys Asp Gly Asn
        595                 600                 605

Tyr Phe Leu Ala Ile Met Asp Lys Lys His Asn Lys Ala Phe Gln Lys
    610                 615                 620

Phe Pro Glu Gly Lys Glu Asn Tyr Glu Lys Met Val Tyr Lys Leu Leu
625                 630                 635                 640
```

Pro Gly Val Asn Lys Met Leu Pro Lys Val Phe Phe Ser Asn Lys Asn
645 650 655

Ile Ala Tyr Phe Asn Pro Ser Lys Glu Leu Leu Glu Asn Tyr Lys Lys
660 665 670

Glu Thr His Lys Lys Gly Asp Thr Phe Asn Leu Glu His Cys His Thr
675 680 685

Leu Ile Asp Phe Phe Lys Asp Ser Leu Asn Lys His Glu Asp Trp Lys
690 695 700

Tyr Phe Asp Phe Gln Phe Ser Glu Thr Lys Ser Tyr Gln Asp Leu Ser
705 710 715 720

Gly Phe Tyr Arg Glu Val Glu His Gln Gly Tyr Lys Ile Asn Phe Lys
725 730 735

Asn Ile Asp Ser Glu Tyr Ile Asp Gly Leu Val Asn Glu Gly Lys Leu
740 745 750

Phe Leu Phe Gln Ile Tyr Ser Lys Asp Phe Ser Pro Phe Ser Lys Gly
755 760 765

Lys Pro Asn Met His Thr Leu Tyr Trp Lys Ala Leu Phe Glu Glu Gln
770 775 780

Asn Leu Gln Asn Val Ile Tyr Lys Leu Asn Gly Gln Ala Glu Ile Phe
785 790 795 800

Phe Arg Lys Ala Ser Ile Lys Pro Lys Asn Ile Ile Leu His Lys Lys
805 810 815

Lys Ile Lys Ile Ala Lys Lys His Phe Ile Asp Lys Lys Thr Lys Thr
820 825 830

Ser Glu Ile Val Pro Val Gln Thr Ile Lys Asn Leu Asn Met Tyr Tyr
835 840 845

Gln Gly Lys Ile Ser Glu Lys Glu Leu Thr Gln Asp Asp Leu Arg Tyr
850 855 860

Ile Asp Asn Phe Ser Ile Phe Asn Glu Lys Asn Lys Thr Ile Asp Ile
865 870 875 880

Ile Lys Asp Lys Arg Phe Thr Val Asp Lys Phe Gln Phe His Val Pro
885 890 895

Ile Thr Met Asn Phe Lys Ala Thr Gly Gly Ser Tyr Ile Asn Gln Thr
900 905 910

Val Leu Glu Tyr Leu Gln Asn Asn Pro Glu Val Lys Ile Ile Gly Leu
915 920 925

Asp Arg Gly Glu Arg His Leu Val Tyr Leu Thr Leu Ile Asp Gln Gln
930 935 940

Gly Asn Ile Leu Lys Gln Glu Ser Leu Asn Thr Ile Thr Asp Ser Lys
945 950 955 960

Ile Ser Thr Pro Tyr His Lys Leu Leu Asp Asn Lys Glu Asn Glu Arg
965 970 975

```
Asp Leu Ala Arg Lys Asn Trp Gly Thr Val Glu Asn Ile Lys Glu Leu
            980                 985                 990

Lys Glu Gly Tyr Ile Ser Gln Val Val His Lys Ile Ala Thr Leu Met
        995                1000                1005

Leu Glu  Glu Asn Ala Ile Val  Val Met Glu Asp Leu  Asn Phe Gly
    1010                1015                1020

Phe Lys  Arg Gly Arg Phe Lys  Val Glu Lys Gln Ile  Tyr Gln Lys
    1025                1030                1035

Leu Glu  Lys Met Leu Ile Asp  Lys Leu Asn Tyr Leu  Val Leu Lys
    1040                1045                1050

Asp Lys  Gln Pro Gln Glu Leu  Gly Gly Leu Tyr Asn  Ala Leu Gln
    1055                1060                1065

Leu Thr  Asn Lys Phe Glu Ser  Phe Gln Lys Met Gly  Lys Gln Ser
    1070                1075                1080

Gly Phe  Leu Phe Tyr Val Pro  Ala Trp Asn Thr Ser  Lys Ile Asp
    1085                1090                1095

Pro Thr  Thr Gly Phe Val Asn  Tyr Phe Tyr Thr Lys  Tyr Glu Asn
    1100                1105                1110

Val Asp  Lys Ala Lys Ala Phe  Phe Glu Lys Phe Glu  Ala Ile Arg
    1115                1120                1125

Phe Asn  Ala Glu Lys Lys Tyr  Phe Glu Phe Glu Val  Lys Lys Tyr
    1130                1135                1140

Ser Asp  Phe Asn Pro Lys Ala  Glu Gly Thr Gln Gln  Ala Trp Thr
    1145                1150                1155

Ile Cys  Thr Tyr Gly Glu Arg  Ile Glu Thr Lys Arg  Gln Lys Asp
    1160                1165                1170

Gln Asn  Asn Lys Phe Val Ser  Thr Pro Ile Asn Leu  Thr Glu Lys
    1175                1180                1185

Ile Glu  Asp Phe Leu Gly Lys  Asn Gln Ile Val Tyr  Gly Asp Gly
    1190                1195                1200

Asn Cys  Ile Lys Ser Gln Ile  Ala Ser Lys Asp Asp  Lys Ala Phe
    1205                1210                1215

Phe Glu  Thr Leu Leu Tyr Trp  Phe Lys Met Thr Leu  Gln Met Arg
    1220                1225                1230

Asn Ser  Glu Thr Arg Thr Asp  Ile Asp Tyr Leu Ile  Ser Pro Val
    1235                1240                1245

Met Asn  Asp Asn Gly Thr Phe  Tyr Asn Ser Arg Asp  Tyr Glu Lys
    1250                1255                1260

Leu Glu  Asn Pro Thr Leu Pro  Lys Asp Ala Asp Ala  Asn Gly Ala
    1265                1270                1275

Tyr His  Ile Ala Lys Lys Gly  Leu Met Leu Leu Asn  Lys Ile Asp
    1280                1285                1290
```

Gln Ala Asp Leu Thr Lys Lys Val Asp Leu Ser Ile Ser Asn Arg
1295 1300 1305

Asp Trp Leu Gln Phe Val Gln Lys Asn Lys
1310 1315

<210> 76
<211> 1310
<212> PRT
<213> 孔茲氏創傷球菌（Helcococcus kunzii）

<400> 76
Met Phe Glu Lys Leu Ser Asn Ile Val Ser Ile Ser Lys Thr Ile Arg
1 5 10 15

Phe Lys Leu Ile Pro Val Gly Lys Thr Leu Glu Asn Ile Glu Lys Leu
20 25 30

Gly Lys Leu Glu Lys Asp Phe Glu Arg Ser Asp Phe Tyr Pro Ile Leu
35 40 45

Lys Asn Ile Ser Asp Asp Tyr Tyr Arg Gln Tyr Ile Lys Glu Lys Leu
50 55 60

Ser Asp Leu Asn Leu Asp Trp Gln Lys Leu Tyr Asp Ala His Glu Leu
65 70 75 80

Leu Asp Ser Ser Lys Lys Glu Ser Gln Lys Asn Leu Glu Met Ile Gln
85 90 95

Ala Gln Tyr Arg Lys Val Leu Phe Asn Ile Leu Ser Gly Glu Leu Asp
100 105 110

Lys Ser Gly Glu Lys Asn Ser Lys Asp Leu Ile Lys Asn Asn Lys Ala
115 120 125

Leu Tyr Gly Lys Leu Phe Lys Lys Gln Phe Ile Leu Glu Val Leu Pro
130 135 140

Asp Phe Val Asn Asn Asn Asp Ser Tyr Ser Glu Glu Asp Leu Glu Gly
145 150 155 160

Leu Asn Leu Tyr Ser Lys Phe Thr Thr Arg Leu Lys Asn Phe Trp Glu
165 170 175

Thr Arg Lys Asn Val Phe Thr Asp Lys Asp Ile Val Thr Ala Ile Pro
180 185 190

Phe Arg Ala Val Asn Glu Asn Phe Gly Phe Tyr Tyr Asp Asn Ile Lys
195 200 205

Ile Phe Asn Lys Asn Ile Glu Tyr Leu Glu Asn Lys Ile Pro Asn Leu
210 215 220

Glu Asn Glu Leu Lys Glu Ala Asp Ile Leu Asp Asp Asn Arg Ser Val
225 230 235 240

Lys Asp Tyr Phe Thr Pro Asn Gly Phe Asn Tyr Val Ile Thr Gln Asp
245 250 255

Gly Ile Asp Val Tyr Gln Ala Ile Arg Gly Gly Phe Thr Lys Glu Asn
260 265 270

```
Gly Glu Lys Val Gln Gly Ile Asn Glu Ile Leu Asn Leu Thr Gln Gln
        275                 280                 285

Gln Leu Arg Arg Lys Pro Glu Thr Lys Asn Val Lys Leu Gly Val Leu
    290                 295                 300

Thr Lys Leu Arg Lys Gln Ile Leu Glu Tyr Ser Glu Ser Thr Ser Phe
305                 310                 315                 320

Leu Ile Asp Gln Ile Glu Asp Asp Asn Asp Leu Val Asp Arg Ile Asn
                325                 330                 335

Lys Phe Asn Val Ser Phe Phe Glu Ser Thr Glu Val Ser Pro Ser Leu
            340                 345                 350

Phe Glu Gln Ile Glu Arg Leu Tyr Asn Ala Leu Lys Ser Ile Lys Lys
        355                 360                 365

Glu Glu Val Tyr Ile Asp Ala Arg Asn Thr Gln Lys Phe Ser Gln Met
    370                 375                 380

Leu Phe Gly Gln Trp Asp Val Ile Arg Arg Gly Tyr Thr Val Lys Ile
385                 390                 395                 400

Thr Glu Gly Ser Lys Glu Glu Lys Lys Lys Tyr Lys Glu Tyr Leu Glu
                405                 410                 415

Leu Asp Glu Thr Ser Lys Ala Lys Arg Tyr Leu Asn Ile Arg Glu Ile
            420                 425                 430

Glu Glu Leu Val Asn Leu Val Glu Gly Phe Glu Glu Val Asp Val Phe
        435                 440                 445

Ser Val Leu Leu Glu Lys Phe Lys Met Asn Asn Ile Glu Arg Ser Glu
    450                 455                 460

Phe Glu Ala Pro Ile Tyr Gly Ser Pro Ile Lys Leu Glu Ala Ile Lys
465                 470                 475                 480

Glu Tyr Leu Glu Lys His Leu Glu Glu Tyr His Lys Trp Lys Leu Leu
                485                 490                 495

Leu Ile Gly Asn Asp Asp Leu Asp Thr Asp Glu Thr Phe Tyr Pro Leu
            500                 505                 510

Leu Asn Glu Val Ile Ser Asp Tyr Tyr Ile Ile Pro Leu Tyr Asn Leu
        515                 520                 525

Thr Arg Asn Tyr Leu Thr Arg Lys His Ser Asp Lys Asp Lys Ile Lys
    530                 535                 540

Val Asn Phe Asp Phe Pro Thr Leu Ala Asp Gly Trp Ser Glu Ser Lys
545                 550                 555                 560

Ile Ser Asp Asn Arg Ser Ile Ile Leu Arg Lys Gly Gly Tyr Tyr Tyr
                565                 570                 575

Leu Gly Ile Leu Ile Asp Asn Lys Leu Leu Ile Asn Lys Lys Asn Lys
            580                 585                 590

Ser Lys Lys Ile Tyr Glu Ile Leu Ile Tyr Asn Gln Ile Pro Glu Phe
        595                 600                 605
```

Ser Lys Ser Ile Pro Asn Tyr Pro Phe Thr Lys Lys Val Lys Glu His
610 615 620

Phe Lys Asn Asn Val Ser Asp Phe Gln Leu Ile Asp Gly Tyr Val Ser
625 630 635 640

Pro Leu Ile Ile Thr Lys Glu Ile Tyr Asp Ile Lys Lys Glu Lys Lys
645 650 655

Tyr Lys Lys Asp Phe Tyr Lys Asp Asn Asn Thr Asn Lys Asn Tyr Leu
660 665 670

Tyr Thr Ile Tyr Lys Trp Ile Glu Phe Cys Lys Gln Phe Leu Tyr Lys
675 680 685

Tyr Lys Gly Pro Asn Lys Glu Ser Tyr Lys Glu Met Tyr Asp Phe Ser
690 695 700

Thr Leu Lys Asp Thr Ser Leu Tyr Val Asn Leu Asn Asp Phe Tyr Ala
705 710 715 720

Asp Val Asn Ser Cys Ala Tyr Arg Val Leu Phe Asn Lys Ile Asp Glu
725 730 735

Asn Thr Ile Asp Asn Ala Val Glu Asp Gly Lys Leu Leu Leu Phe Gln
740 745 750

Ile Tyr Asn Lys Asp Phe Ser Pro Glu Ser Lys Gly Lys Lys Asn Leu
755 760 765

His Thr Leu Tyr Trp Leu Ser Met Phe Ser Glu Glu Asn Leu Arg Thr
770 775 780

Arg Lys Leu Lys Leu Asn Gly Gln Ala Glu Ile Phe Tyr Arg Lys Lys
785 790 795 800

Leu Glu Lys Lys Pro Ile Ile His Lys Glu Gly Ser Ile Leu Leu Asn
805 810 815

Lys Ile Asp Lys Glu Gly Asn Thr Ile Pro Glu Asn Ile Tyr His Glu
820 825 830

Cys Tyr Arg Tyr Leu Asn Lys Lys Ile Gly Arg Glu Asp Leu Ser Asp
835 840 845

Glu Ala Ile Ala Leu Phe Asn Lys Asp Val Leu Lys Tyr Lys Glu Ala
850 855 860

Arg Phe Asp Ile Ile Lys Asp Arg Arg Tyr Ser Glu Ser Gln Phe Phe
865 870 875 880

Phe His Val Pro Ile Thr Phe Asn Trp Asp Ile Lys Thr Asn Lys Asn
885 890 895

Val Asn Gln Ile Val Gln Gly Met Ile Lys Asp Gly Glu Ile Lys His
900 905 910

Ile Ile Gly Ile Asp Arg Gly Glu Arg His Leu Leu Tyr Tyr Ser Val
915 920 925

Ile Asp Leu Glu Gly Asn Ile Val Glu Gln Gly Ser Leu Asn Thr Leu
930 935 940

```
Glu Gln Asn Arg Phe Asp Asn Ser Thr Val Lys Val Asp Tyr Gln Asn
945                 950                 955                 960

Lys Leu Arg Thr Arg Glu Glu Asp Arg Asp Arg Ala Arg Lys Asn Trp
                965                 970                 975

Thr Asn Ile Asn Lys Ile Lys Glu Leu Lys Asp Gly Tyr Leu Ser His
            980                 985                 990

Val Val His Lys Leu Ser Arg Leu Ile Ile Lys Tyr Glu Ala Ile Val
        995                 1000                1005

Ile Met  Glu Asn Leu Asn Gln  Gly Phe Lys Arg Gly  Arg Phe Lys
    1010                1015                1020

Val Glu  Arg Gln Val Tyr Gln  Lys Phe Glu Leu Ala  Leu Met Asn
    1025                1030                1035

Lys Leu  Ser Ala Leu Ser Phe  Lys Glu Lys Tyr Asp  Glu Arg Lys
    1040                1045                1050

Asn Leu  Glu Pro Ser Gly Ile  Leu Asn Pro Ile Gln  Ala Cys Tyr
    1055                1060                1065

Pro Val  Asp Ala Tyr Gln Glu  Leu Gln Gly Gln Asn  Gly Ile Val
    1070                1075                1080

Phe Tyr  Leu Pro Ala Ala Tyr  Thr Ser Val Ile Asp  Pro Val Thr
    1085                1090                1095

Gly Phe  Thr Asn Leu Phe Arg  Leu Lys Ser Ile Asn  Ser Ser Lys
    1100                1105                1110

Tyr Glu  Glu Phe Ile Lys Lys  Phe Lys Asn Ile Tyr  Phe Asp Asn
    1115                1120                1125

Glu Glu  Glu Asp Phe Lys Phe  Ile Phe Asn Tyr Lys  Asp Phe Ala
    1130                1135                1140

Lys Ala  Asn Leu Val Ile Leu  Asn Asn Ile Lys Ser  Lys Asp Trp
    1145                1150                1155

Lys Ile  Ser Thr Arg Gly Glu  Arg Ile Ser Tyr Asn  Ser Lys Lys
    1160                1165                1170

Lys Glu  Tyr Phe Tyr Val Gln  Pro Thr Glu Phe Leu  Ile Asn Lys
    1175                1180                1185

Leu Lys  Glu Leu Asn Ile Asp  Tyr Glu Asn Ile Asp  Ile Ile Pro
    1190                1195                1200

Leu Ile  Asp Asn Leu Glu Glu  Lys Ala Lys Arg Lys  Ile Leu Lys
    1205                1210                1215

Ala Leu  Phe Asp Thr Phe Lys  Tyr Ser Val Gln Leu  Arg Asn Tyr
    1220                1225                1230

Asp Phe  Glu Asn Asp Tyr Ile  Ile Ser Pro Thr Ala  Asp Asp Asn
    1235                1240                1245

Gly Asn  Tyr Tyr Asn Ser Asn  Glu Ile Asp Ile Asp  Lys Thr Asn
    1250                1255                1260
```

Leu Pro Asn Asn Gly Asp Ala Asn Gly Ala Phe Asn Ile Ala Arg
1265 1270 1275

Lys Gly Leu Leu Leu Lys Asp Arg Ile Val Asn Ser Asn Glu Ser
1280 1285 1290

Lys Val Asp Leu Lys Ile Lys Asn Glu Asp Trp Ile Asn Phe Ile
1295 1300 1305

Ile Ser
1310

<210> 77
<211> 1307
<212> PRT
<213> 胺基酸球菌屬某種（Acidaminococcus sp.）

<400> 77
Met Thr Gln Phe Glu Gly Phe Thr Asn Leu Tyr Gln Val Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Lys His Ile Gln
20 25 30

Glu Gln Gly Phe Ile Glu Glu Asp Lys Ala Arg Asn Asp His Tyr Lys
35 40 45

Glu Leu Lys Pro Ile Ile Asp Arg Ile Tyr Lys Thr Tyr Ala Asp Gln
50 55 60

Cys Leu Gln Leu Val Gln Leu Asp Trp Glu Asn Leu Ser Ala Ala Ile
65 70 75 80

Asp Ser Tyr Arg Lys Glu Lys Thr Glu Glu Thr Arg Asn Ala Leu Ile
85 90 95

Glu Glu Gln Ala Thr Tyr Arg Asn Ala Ile His Asp Tyr Phe Ile Gly
100 105 110

Arg Thr Asp Asn Leu Thr Asp Ala Ile Asn Lys Arg His Ala Glu Ile
115 120 125

Tyr Lys Gly Leu Phe Lys Ala Glu Leu Phe Asn Gly Lys Val Leu Lys
130 135 140

Gln Leu Gly Thr Val Thr Thr Thr Glu His Glu Asn Ala Leu Leu Arg
145 150 155 160

Ser Phe Asp Lys Phe Thr Thr Tyr Phe Ser Gly Phe Tyr Glu Asn Arg
165 170 175

Lys Asn Val Phe Ser Ala Glu Asp Ile Ser Thr Ala Ile Pro His Arg
180 185 190

Ile Val Gln Asp Asn Phe Pro Lys Phe Lys Glu Asn Cys His Ile Phe
195 200 205

Thr Arg Leu Ile Thr Ala Val Pro Ser Leu Arg Glu His Phe Glu Asn
210 215 220

Val Lys Lys Ala Ile Gly Ile Phe Val Ser Thr Ser Ile Glu Glu Val
225 230 235 240

Phe Ser Phe Pro Phe Tyr Asn Gln Leu Leu Thr Gln Thr Gln Ile Asp
245 250 255

```
Leu Tyr Asn Gln Leu Leu Gly Gly Ile Ser Arg Glu Ala Gly Thr Glu
            260                 265                 270

Lys Ile Lys Gly Leu Asn Glu Val Leu Asn Leu Ala Ile Gln Lys Asn
        275                 280                 285

Asp Glu Thr Ala His Ile Ile Ala Ser Leu Pro His Arg Phe Ile Pro
    290                 295                 300

Leu Phe Lys Gln Ile Leu Ser Asp Arg Asn Thr Leu Ser Phe Ile Leu
305                 310                 315                 320

Glu Glu Phe Lys Ser Asp Glu Glu Val Ile Gln Ser Phe Cys Lys Tyr
                325                 330                 335

Lys Thr Leu Leu Arg Asn Glu Asn Val Leu Glu Thr Ala Glu Ala Leu
            340                 345                 350

Phe Asn Glu Leu Asn Ser Ile Asp Leu Thr His Ile Phe Ile Ser His
        355                 360                 365

Lys Lys Leu Glu Thr Ile Ser Ser Ala Leu Cys Asp His Trp Asp Thr
    370                 375                 380

Leu Arg Asn Ala Leu Tyr Glu Arg Arg Ile Ser Glu Leu Thr Gly Lys
385                 390                 395                 400

Ile Thr Lys Ser Ala Lys Glu Lys Val Gln Arg Ser Leu Lys His Glu
                405                 410                 415

Asp Ile Asn Leu Gln Glu Ile Ile Ser Ala Ala Gly Lys Glu Leu Ser
            420                 425                 430

Glu Ala Phe Lys Gln Lys Thr Ser Glu Ile Leu Ser His Ala His Ala
        435                 440                 445

Ala Leu Asp Gln Pro Leu Pro Thr Thr Leu Lys Lys Gln Glu Glu Lys
    450                 455                 460

Glu Ile Leu Lys Ser Gln Leu Asp Ser Leu Leu Gly Leu Tyr His Leu
465                 470                 475                 480

Leu Asp Trp Phe Ala Val Asp Glu Ser Asn Glu Val Asp Pro Glu Phe
                485                 490                 495

Ser Ala Arg Leu Thr Gly Ile Lys Leu Glu Met Glu Pro Ser Leu Ser
            500                 505                 510

Phe Tyr Asn Lys Ala Arg Asn Tyr Ala Thr Lys Lys Pro Tyr Ser Val
        515                 520                 525

Glu Lys Phe Lys Leu Asn Phe Gln Met Pro Thr Leu Ala Ser Gly Trp
    530                 535                 540

Asp Val Asn Lys Glu Lys Asn Asn Gly Ala Ile Leu Phe Val Lys Asn
545                 550                 555                 560

Gly Leu Tyr Tyr Leu Gly Ile Met Pro Lys Gln Lys Gly Arg Tyr Lys
                565                 570                 575

Ala Leu Ser Phe Glu Pro Thr Glu Lys Thr Ser Glu Gly Phe Asp Lys
            580                 585                 590
```

Met Tyr Tyr Asp Tyr Phe Pro Asp Ala Ala Lys Met Ile Pro Lys Cys
595 600 605

Ser Thr Gln Leu Lys Ala Val Thr Ala His Phe Gln Thr His Thr Thr
610 615 620

Pro Ile Leu Leu Ser Asn Asn Phe Ile Glu Pro Leu Glu Ile Thr Lys
625 630 635 640

Glu Ile Tyr Asp Leu Asn Asn Pro Glu Lys Glu Pro Lys Lys Phe Gln
645 650 655

Thr Ala Tyr Ala Lys Lys Thr Gly Asp Gln Lys Gly Tyr Arg Glu Ala
660 665 670

Leu Cys Lys Trp Ile Asp Phe Thr Arg Asp Phe Leu Ser Lys Tyr Thr
675 680 685

Lys Thr Thr Ser Ile Asp Leu Ser Ser Leu Arg Pro Ser Ser Gln Tyr
690 695 700

Lys Asp Leu Gly Glu Tyr Tyr Ala Glu Leu Asn Pro Leu Leu Tyr His
705 710 715 720

Ile Ser Phe Gln Arg Ile Ala Glu Lys Glu Ile Met Asp Ala Val Glu
725 730 735

Thr Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ala Lys
740 745 750

Gly His His Gly Lys Pro Asn Leu His Thr Leu Tyr Trp Thr Gly Leu
755 760 765

Phe Ser Pro Glu Asn Leu Ala Lys Thr Ser Ile Lys Leu Asn Gly Gln
770 775 780

Ala Glu Leu Phe Tyr Arg Pro Lys Ser Arg Met Lys Arg Met Ala His
785 790 795 800

Arg Leu Gly Glu Lys Met Leu Asn Lys Lys Leu Lys Asp Gln Lys Thr
805 810 815

Pro Ile Pro Asp Thr Leu Tyr Gln Glu Leu Tyr Asp Tyr Val Asn His
820 825 830

Arg Leu Ser His Asp Leu Ser Asp Glu Ala Arg Ala Leu Leu Pro Asn
835 840 845

Val Ile Thr Lys Glu Val Ser His Glu Ile Ile Lys Asp Arg Arg Phe
850 855 860

Thr Ser Asp Lys Phe Phe Phe His Val Pro Ile Thr Leu Asn Tyr Gln
865 870 875 880

Ala Ala Asn Ser Pro Ser Lys Phe Asn Gln Arg Val Asn Ala Tyr Leu
885 890 895

Lys Glu His Pro Glu Thr Pro Ile Ile Gly Ile Asp Arg Gly Glu Arg
900 905 910

Asn Leu Ile Tyr Ile Thr Val Ile Asp Ser Thr Gly Lys Ile Leu Glu
915 920 925

```
Gln Arg Ser Leu Asn Thr Ile Gln Gln Phe Asp Tyr Gln Lys Lys Leu
    930                 935                 940

Asp Asn Arg Glu Lys Glu Arg Val Ala Ala Arg Gln Ala Trp Ser Val
945                 950                 955                 960

Val Gly Thr Ile Lys Asp Leu Lys Gln Gly Tyr Leu Ser Gln Val Ile
                965                 970                 975

His Glu Ile Val Asp Leu Met Ile His Tyr Gln Ala Val Val Val Leu
            980                 985                 990

Glu Asn Leu Asn Phe Gly Phe Lys Ser Lys Arg Thr Gly Ile Ala Glu
        995                 1000                1005

Lys Ala  Val Tyr Gln Gln Phe  Glu Lys Met Leu Ile  Asp Lys Leu
    1010                1015                1020

Asn Cys  Leu Val Leu Lys Asp  Tyr Pro Ala Glu Lys  Val Gly Gly
    1025                1030                1035

Val Leu  Asn Pro Tyr Gln Leu  Thr Asp Gln Phe Thr  Ser Phe Ala
    1040                1045                1050

Lys Met  Gly Thr Gln Ser Gly  Phe Leu Phe Tyr Val  Pro Ala Pro
    1055                1060                1065

Tyr Thr  Ser Lys Ile Asp Pro  Leu Thr Gly Phe Val  Asp Pro Phe
    1070                1075                1080

Val Trp  Lys Thr Ile Lys Asn  His Glu Ser Arg Lys  His Phe Leu
    1085                1090                1095

Glu Gly  Phe Asp Phe Leu His  Tyr Asp Val Lys Thr  Gly Asp Phe
    1100                1105                1110

Ile Leu  His Phe Lys Met Asn  Arg Asn Leu Ser Phe  Gln Arg Gly
    1115                1120                1125

Leu Pro  Gly Phe Met Pro Ala  Trp Asp Ile Val Phe  Glu Lys Asn
    1130                1135                1140

Glu Thr  Gln Phe Asp Ala Lys  Gly Thr Pro Phe Ile  Ala Gly Lys
    1145                1150                1155

Arg Ile  Val Pro Val Ile Glu  Asn His Arg Phe Thr  Gly Arg Tyr
    1160                1165                1170

Arg Asp  Leu Tyr Pro Ala Asn  Glu Leu Ile Ala Leu  Leu Glu Glu
    1175                1180                1185

Lys Gly  Ile Val Phe Arg Asp  Gly Ser Asn Ile Leu  Pro Lys Leu
    1190                1195                1200

Leu Glu  Asn Asp Asp Ser His  Ala Ile Asp Thr Met  Val Ala Leu
    1205                1210                1215

Ile Arg  Ser Val Leu Gln Met  Arg Asn Ser Asn Ala  Ala Thr Gly
    1220                1225                1230

Glu Asp  Tyr Ile Asn Ser Pro  Val Arg Asp Leu Asn  Gly Val Cys
    1235                1240                1245
```

Phe Asp Ser Arg Phe Gln Asn Pro Glu Trp Pro Met Asp Ala Asp
1250 1255 1260

Ala Asn Gly Ala Tyr His Ile Ala Leu Lys Gly Gln Leu Leu Leu
1265 1270 1275

Asn His Leu Lys Glu Ser Lys Asp Leu Lys Leu Gln Asn Gly Ile
1280 1285 1290

Ser Asn Gln Asp Trp Leu Ala Tyr Ile Gln Glu Leu Arg Asn
1295 1300 1305

<210> 78
<211> 1307
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 78
Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
20 25 30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
35 40 45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
50 55 60

Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65 70 75 80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
85 90 95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Lys Tyr
100 105 110

Ile Asn Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
115 120 125

Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
130 135 140

Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145 150 155 160

Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
165 170 175

Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190

Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205

Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
210 215 220

Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240

```
Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
                245                 250                 255

Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
            260                 265                 270

Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
        275                 280                 285

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
    290                 295                 300

Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305                 310                 315                 320

Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
                325                 330                 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
            340                 345                 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
        355                 360                 365

Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
    370                 375                 380

Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385                 390                 395                 400

Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
                405                 410                 415

Val Leu Glu Tyr Ile Thr Gln Gln Val Ala Pro Lys Asn Leu Asp Asn
            420                 425                 430

Pro Ser Lys Lys Glu Gln Asp Leu Ile Ala Lys Lys Thr Glu Lys Ala
        435                 440                 445

Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
    450                 455                 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ser
465                 470                 475                 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
                485                 490                 495

Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
            500                 505                 510

Asp Leu Leu Gln Ala Ser Ala Glu Glu Asp Val Lys Ala Ile Lys Asp
        515                 520                 525

Leu Leu Asp Gln Thr Asn Asn Leu Leu His Arg Leu Lys Ile Phe His
    530                 535                 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545                 550                 555                 560

Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
                565                 570                 575
```

```
Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
            580                 585                 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Ser Gly
        595                 600                 605

Trp Asp Lys Asn Lys Glu Ser Ala Asn Thr Ala Ile Leu Phe Ile Lys
    610                 615                 620

Asp Asp Lys Tyr Tyr Leu Gly Ile Met Asp Lys Lys His Asn Lys Ile
625                 630                 635                 640

Phe Ser Asp Lys Ala Ile Glu Glu Asn Lys Gly Glu Gly Tyr Lys Lys
                645                 650                 655

Ile Val Tyr Lys Gln Ile Ala Asp Ala Ser Lys Asp Ile Gln Asn Leu
            660                 665                 670

Met Ile Ile Asp Gly Lys Thr Val Cys Lys Lys Gly Arg Lys Asp Arg
        675                 680                 685

Asn Gly Val Asn Arg Gln Leu Leu Ser Leu Lys Arg Lys His Leu Pro
    690                 695                 700

Glu Asn Ile Tyr Arg Ile Lys Glu Thr Lys Ser Tyr Leu Lys Asn Glu
705                 710                 715                 720

Ala Arg Phe Ser Arg Lys Asp Leu Tyr Asp Phe Ile Asp Tyr Tyr Lys
                725                 730                 735

Asp Arg Leu Asp Tyr Tyr Asp Phe Glu Phe Glu Leu Lys Pro Ser Asn
            740                 745                 750

Glu Tyr Ser Asp Phe Asn Asp Phe Thr Asn His Ile Gly Ser Gln Gly
        755                 760                 765

Tyr Lys Leu Thr Phe Glu Asn Ile Ser Gln Asp Tyr Ile Asn Ser Leu
    770                 775                 780

Val Asn Glu Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Ser Lys Asp Phe
785                 790                 795                 800

Ser Ala Tyr Ser Lys Gly Arg Pro Asn Leu His Thr Leu Tyr Trp Lys
                805                 810                 815

Ala Leu Phe Asp Glu Arg Asn Leu Gln Asp Val Val Tyr Lys Leu Asn
            820                 825                 830

Gly Glu Ala Glu Leu Phe Tyr Arg Lys Gln Ser Ile Pro Lys Lys Ile
        835                 840                 845

Thr His Pro Ala Lys Glu Thr Ile Ala Asn Lys Asn Lys Asp Asn Pro
    850                 855                 860

Lys Lys Glu Ser Val Phe Glu Tyr Asp Leu Ile Lys Asp Lys Arg Phe
865                 870                 875                 880

Thr Glu Asp Lys Phe Phe Phe His Cys Pro Ile Thr Ile Asn Phe Lys
                885                 890                 895

Ser Ser Gly Ala Asn Lys Phe Asn Asp Glu Ile Asn Leu Leu Leu Lys
            900                 905                 910
```

```
Glu Lys Ala Asn Asp Val His Ile Leu Ser Ile Asp Arg Gly Glu Arg
        915                 920                 925

His Leu Ala Tyr Tyr Thr Leu Val Asp Gly Lys Gly Asn Ile Ile Lys
    930                 935                 940

Gln Asp Asn Phe Asn Ile Ile Gly Asn Asp Arg Met Lys Thr Asn Tyr
945                 950                 955                 960

His Asp Lys Leu Ala Ala Ile Glu Lys Asp Arg Asp Ser Ala Arg Lys
                965                 970                 975

Asp Trp Lys Lys Ile Asn Asn Ile Lys Glu Met Lys Glu Gly Tyr Leu
            980                 985                 990

Ser Gln Val Val His Glu Ile Ala Lys Leu Val Ile Glu Tyr Asn Ala
        995                 1000                1005

Ile Val  Val Phe Glu Asp Leu  Asn Phe Gly Phe Lys  Arg Gly Arg
    1010                1015                1020

Phe Lys  Val Glu Lys Gln Val  Tyr Gln Lys Leu Glu  Lys Met Leu
    1025                1030                1035

Ile Glu  Lys Leu Asn Tyr Leu  Val Phe Lys Asp Asn  Glu Phe Asp
    1040                1045                1050

Lys Thr  Gly Gly Val Leu Arg  Ala Tyr Gln Leu Thr  Ala Pro Phe
    1055                1060                1065

Glu Thr  Phe Lys Lys Met Gly  Lys Gln Thr Gly Ile  Ile Tyr Tyr
    1070                1075                1080

Val Pro  Ala Gly Phe Thr Ser  Lys Ile Cys Pro Val  Thr Gly Phe
    1085                1090                1095

Val Asn  Gln Leu Tyr Pro Lys  Tyr Glu Ser Val Ser  Lys Ser Gln
    1100                1105                1110

Glu Phe  Phe Ser Lys Phe Asp  Lys Ile Cys Tyr Asn  Leu Asp Lys
    1115                1120                1125

Gly Tyr  Phe Glu Phe Ser Phe  Asp Tyr Lys Asn Phe  Gly Asp Lys
    1130                1135                1140

Ala Ala  Lys Gly Lys Trp Thr  Ile Ala Ser Phe Gly  Ser Arg Leu
    1145                1150                1155

Ile Asn  Phe Arg Asn Ser Asp  Lys Asn His Asn Trp  Asp Thr Arg
    1160                1165                1170

Glu Val  Tyr Pro Thr Lys Glu  Leu Glu Lys Leu Leu  Lys Asp Tyr
    1175                1180                1185

Ser Ile  Glu Tyr Gly His Gly  Glu Cys Ile Lys Ala  Ala Ile Cys
    1190                1195                1200

Gly Glu  Ser Asp Lys Lys Phe  Phe Ala Lys Leu Thr  Ser Val Leu
    1205                1210                1215

Asn Thr  Ile Leu Gln Met Arg  Asn Ser Lys Thr Gly  Thr Glu Leu
    1220                1225                1230
```

Asp Tyr Leu Ile Ser Pro Val Ala Asp Val Asn Gly Asn Phe Phe
1235 1240 1245

Asp Ser Arg Gln Ala Pro Lys Asn Met Pro Gln Asp Ala Asp Ala
1250 1255 1260

Asn Gly Ala Tyr His Ile Gly Leu Lys Gly Leu Met Leu Leu Asp
1265 1270 1275

Arg Ile Lys Asn Asn Gln Glu Gly Lys Lys Leu Asn Leu Val Ile
1280 1285 1290

Lys Asn Glu Glu Tyr Phe Glu Phe Val Gln Asn Arg Asn Asn
1295 1300 1305

<210> 79
<211> 1305
<212> PRT
<213> 真細菌某種（Eubacterium sp.）

<400> 79
Met Asn Lys Ala Ala Asp Asn Tyr Thr Gly Gly Asn Tyr Asp Glu Phe
1 5 10 15

Ile Ala Leu Ser Lys Val Gln Lys Thr Leu Arg Asn Glu Leu Lys Pro
20 25 30

Thr Pro Phe Thr Ala Glu His Ile Lys Gln Arg Gly Ile Ile Ser Glu
35 40 45

Asp Glu Tyr Arg Ala Gln Gln Ser Leu Glu Leu Lys Lys Ile Ala Asp
50 55 60

Glu Tyr Tyr Arg Asn Tyr Ile Thr His Lys Leu Asn Asp Ile Asn Asn
65 70 75 80

Leu Asp Phe Tyr Asn Leu Phe Asp Ala Ile Glu Glu Lys Tyr Lys Lys
85 90 95

Asn Asp Lys Asp Asn Arg Asp Lys Leu Asp Leu Val Glu Lys Ser Lys
100 105 110

Arg Gly Glu Ile Ala Lys Met Leu Ser Ala Asp Asp Asn Phe Lys Ser
115 120 125

Met Phe Glu Ala Lys Leu Ile Thr Lys Leu Leu Pro Asp Tyr Val Glu
130 135 140

Arg Asn Tyr Thr Gly Glu Asp Lys Glu Lys Ala Leu Glu Thr Leu Ala
145 150 155 160

Leu Phe Lys Gly Phe Thr Thr Tyr Phe Lys Gly Tyr Phe Lys Thr Arg
165 170 175

Lys Asn Met Phe Ser Gly Glu Gly Gly Ala Ser Ser Ile Cys His Arg
180 185 190

Ile Val Asn Val Asn Ala Ser Ile Phe Tyr Asp Asn Leu Lys Thr Phe
195 200 205

Met Arg Ile Gln Glu Lys Ala Gly Asp Glu Ile Ala Leu Ile Glu Glu
210 215 220

Glu Leu Thr Glu Lys Leu Asp Gly Trp Arg Leu Glu His Ile Phe Ser
225 230 235 240

```
Arg Asp Tyr Tyr Asn Glu Val Leu Ala Gln Lys Gly Ile Asp Tyr Tyr
                245                 250                 255

Asn Gln Ile Cys Gly Asp Ile Asn Lys His Met Asn Leu Tyr Cys Gln
            260                 265                 270

Gln Asn Lys Phe Lys Ala Asn Ile Phe Lys Met Met Lys Ile Gln Lys
        275                 280                 285

Gln Ile Met Gly Ile Ser Glu Lys Ala Phe Glu Ile Pro Pro Met Tyr
    290                 295                 300

Gln Asn Asp Glu Glu Val Tyr Ala Ser Phe Asn Glu Phe Ile Ser Arg
305                 310                 315                 320

Leu Glu Glu Val Lys Leu Thr Asp Arg Leu Ile Asn Ile Leu Gln Asn
                325                 330                 335

Ile Asn Ile Tyr Asn Thr Ala Lys Ile Tyr Ile Asn Ala Arg Tyr Tyr
            340                 345                 350

Thr Asn Val Ser Ser Tyr Val Tyr Gly Gly Trp Gly Val Ile Asp Ser
        355                 360                 365

Ala Ile Glu Arg Tyr Leu Tyr Asn Thr Ile Ala Gly Lys Gly Gln Ser
    370                 375                 380

Lys Val Lys Lys Ile Glu Asn Ala Lys Lys Asp Asn Lys Phe Met Ser
385                 390                 395                 400

Val Lys Glu Leu Asp Ser Ile Val Ala Glu Tyr Glu Pro Asp Tyr Phe
                405                 410                 415

Asn Ala Pro Tyr Ile Asp Asp Asp Asp Asn Ala Val Lys Ala Phe Gly
            420                 425                 430

Gly Gln Gly Val Leu Gly Tyr Phe Asn Lys Met Ser Glu Leu Leu Ala
        435                 440                 445

Asp Val Ser Leu Tyr Thr Ile Asp Tyr Asn Ser Asp Asp Ser Leu Ile
    450                 455                 460

Glu Asn Lys Glu Ser Ala Leu Arg Ile Lys Lys Gln Leu Asp Asp Ile
465                 470                 475                 480

Met Ser Leu Tyr His Trp Leu Gln Thr Phe Ile Ile Asp Glu Val Val
                485                 490                 495

Glu Lys Asp Asn Ala Phe Tyr Ala Glu Leu Glu Asp Ile Cys Cys Glu
            500                 505                 510

Leu Glu Asn Val Val Thr Leu Tyr Asp Arg Ile Arg Asn Tyr Val Thr
        515                 520                 525

Lys Lys Pro Tyr Ser Thr Gln Lys Phe Lys Leu Asn Phe Ala Ser Pro
    530                 535                 540

Thr Leu Ala Ala Gly Trp Ser Arg Ser Lys Glu Phe Asp Asn Asn Ala
545                 550                 555                 560

Ile Ile Leu Leu Arg Asn Asn Lys Tyr Tyr Ile Ala Ile Phe Asn Val
                565                 570                 575
```

```
Asn Asn Lys Pro Asp Lys Gln Ile Ile Lys Gly Ser Glu Glu Gln Arg
            580                 585                 590

Leu Ser Thr Asp Tyr Lys Lys Met Val Tyr Asn Leu Leu Pro Gly Pro
        595                 600                 605

Asn Lys Met Leu Pro Lys Val Phe Ile Lys Ser Asp Thr Gly Lys Arg
    610                 615                 620

Asp Tyr Asn Pro Ser Ser Tyr Ile Leu Glu Gly Tyr Glu Lys Asn Arg
625                 630                 635                 640

His Ile Lys Ser Ser Gly Asn Phe Asp Ile Asn Tyr Cys His Asp Leu
                645                 650                 655

Ile Asp Tyr Tyr Lys Ala Cys Ile Asn Lys His Pro Glu Trp Lys Asn
            660                 665                 670

Tyr Gly Phe Lys Phe Lys Glu Thr Asn Gln Tyr Asn Asp Ile Gly Gln
        675                 680                 685

Phe Tyr Lys Asp Val Glu Lys Gln Gly Tyr Ser Ile Ser Trp Ala Tyr
    690                 695                 700

Ile Ser Glu Glu Asp Ile Asn Lys Leu Asp Glu Glu Gly Lys Ile Tyr
705                 710                 715                 720

Leu Phe Glu Ile Tyr Asn Lys Asp Leu Ser Ala His Ser Thr Gly Arg
                725                 730                 735

Asp Asn Leu His Thr Met Tyr Leu Lys Asn Ile Phe Ser Glu Asp Asn
            740                 745                 750

Leu Lys Asn Ile Cys Ile Glu Leu Asn Gly Glu Ala Glu Leu Phe Tyr
        755                 760                 765

Arg Lys Ser Ser Met Lys Ser Asn Ile Thr His Lys Lys Asp Thr Ile
    770                 775                 780

Leu Val Asn Lys Thr Tyr Ile Asn Glu Thr Gly Val Arg Val Ser Leu
785                 790                 795                 800

Ser Asp Glu Asp Tyr Met Lys Val Tyr Asn Tyr Tyr Asn Asn Asn Tyr
                805                 810                 815

Val Ile Asp Thr Glu Asn Asp Lys Asn Leu Ile Asp Ile Ile Glu Lys
            820                 825                 830

Ile Gly His Arg Lys Ser Lys Ile Asp Ile Val Lys Asp Lys Arg Tyr
        835                 840                 845

Thr Glu Asp Lys Tyr Phe Leu Tyr Leu Pro Ile Thr Ile Asn Tyr Gly
    850                 855                 860

Ile Glu Asp Glu Asn Val Asn Ser Lys Ile Ile Glu Tyr Ile Ala Lys
865                 870                 875                 880

Gln Asp Asn Met Asn Val Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu
                885                 890                 895

Ile Tyr Ile Ser Val Ile Asp Asn Lys Gly Asn Ile Ile Glu Gln Lys
            900                 905                 910
```

```
Ser Phe Asn Leu Val Asn Asn Tyr Asp Tyr Lys Asn Lys Leu Lys Asn
        915                 920                 925

Met Glu Lys Thr Arg Asp Asn Ala Arg Lys Asn Trp Gln Glu Ile Gly
    930                 935                 940

Lys Ile Lys Asp Val Lys Ser Gly Tyr Leu Ser Gly Val Ile Ser Lys
945                 950                 955                 960

Ile Ala Arg Met Val Ile Asp Tyr Asn Ala Ile Ile Val Met Glu Asp
                965                 970                 975

Leu Asn Lys Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Arg Gln Val
            980                 985                 990

Tyr Gln Lys Phe Glu Asn Met Leu Ile Ser Lys Leu Asn Tyr Leu Val
        995                 1000                1005

Phe Lys  Glu Arg Lys Ala Asp  Glu Asn Gly Gly Ile  Leu Arg Gly
    1010                1015                1020

Tyr Gln  Leu Thr Tyr Ile Pro  Lys Ser Ile Lys Asn  Val Gly Lys
    1025                1030                1035

Gln Cys  Gly Cys Ile Phe Tyr  Val Pro Ala Ala Tyr  Thr Ser Lys
    1040                1045                1050

Ile Asp  Pro Ala Thr Gly Phe  Ile Asn Ile Phe Asp  Phe Lys Lys
    1055                1060                1065

Tyr Ser  Gly Ser Gly Ile Asn  Ala Lys Val Lys Asp  Lys Lys Glu
    1070                1075                1080

Phe Leu  Met Ser Met Asn Ser  Ile Arg Tyr Ile Asn  Glu Cys Ser
    1085                1090                1095

Glu Glu  Tyr Glu Lys Ile Gly  His Arg Glu Leu Phe  Ala Phe Ser
    1100                1105                1110

Phe Asp  Tyr Asn Asn Phe Lys  Thr Tyr Asn Val Ser  Ser Pro Val
    1115                1120                1125

Asn Glu  Trp Thr Ala Tyr Thr  Tyr Gly Glu Arg Ile  Lys Lys Leu
    1130                1135                1140

Tyr Lys  Asp Gly Arg Trp Leu  Arg Ser Glu Val Leu  Asn Leu Thr
    1145                1150                1155

Glu Asn  Leu Ile Lys Leu Met  Glu Gln Tyr Asn Ile  Glu Tyr Lys
    1160                1165                1170

Asp Gly  His Asp Ile Arg Glu  Asp Ile Ser His Met  Asp Glu Thr
    1175                1180                1185

Arg Asn  Ala Asp Phe Ile Cys  Ser Leu Phe Glu Glu  Leu Lys Tyr
    1190                1195                1200

Thr Val  Gln Leu Arg Asn Ser  Lys Ser Glu Ala Glu  Asp Glu Asn
    1205                1210                1215

Tyr Asp  Arg Leu Val Ser Pro  Ile Leu Asn Ser Ser  Asn Gly Phe
    1220                1225                1230
```

```
Tyr Asp Ser Ser Asp Tyr  Met Glu Asn Glu Asn  Asn Thr Thr His
    1235                 1240                 1245

Thr Met Pro Lys Asp Ala  Asp Ala Asn Gly Ala  Tyr Cys Ile Ala
    1250                 1255                 1260

Leu Lys Gly Leu Tyr Glu  Ile Asn Lys Ile Lys  Gln Asn Trp Ser
    1265                 1270                 1275

Asp Asp Lys Lys Phe Lys  Glu Asn Glu Leu Tyr  Ile Asn Val Thr
    1280                 1285                 1290

Glu Trp Leu Asp Tyr Ile  Gln Asn Arg Arg Phe  Glu
    1295                 1300                 1305
```

<210> 80
<211> 1300
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 80

```
Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1               5                   10                  15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
            20                  25                  30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
        35                  40                  45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
    50                  55                  60

Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65                  70                  75                  80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
                85                  90                  95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
            100                 105                 110

Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
        115                 120                 125

Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
    130                 135                 140

Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145                 150                 155                 160

Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
                165                 170                 175

Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
            180                 185                 190

Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
        195                 200                 205

Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
    210                 215                 220
```

```
Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225                 230                 235                 240

Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
                245                 250                 255

Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
            260                 265                 270

Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
        275                 280                 285

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
    290                 295                 300

Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305                 310                 315                 320

Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
                325                 330                 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
            340                 345                 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
        355                 360                 365

Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
    370                 375                 380

Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385                 390                 395                 400

Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
                405                 410                 415

Val Leu Glu Tyr Ile Thr Gln Gln Val Ala Pro Lys Asn Leu Asp Asn
            420                 425                 430

Pro Ser Lys Lys Glu Gln Asp Leu Ile Ala Lys Lys Thr Glu Lys Ala
        435                 440                 445

Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
    450                 455                 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465                 470                 475                 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
                485                 490                 495

Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
            500                 505                 510

Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
        515                 520                 525

Leu Leu Asp Gln Thr Asn Asn Leu Leu His Arg Leu Lys Ile Phe His
    530                 535                 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545                 550                 555                 560
```

```
Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
                565                 570                 575

Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
            580                 585                 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
        595                 600                 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
    610                 615                 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625                 630                 635                 640

Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
                645                 650                 655

Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
            660                 665                 670

Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
        675                 680                 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Asn Pro Gln
    690                 695                 700

Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
705                 710                 715                 720

Ile Asp Phe Tyr Lys Glu Ser Ile Ser Lys His Pro Glu Trp Lys Asp
                725                 730                 735

Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
            740                 745                 750

Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
        755                 760                 765

Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
    770                 775                 780

Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
785                 790                 795                 800

Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
                805                 810                 815

Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
            820                 825                 830

Arg Lys Lys Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
        835                 840                 845

Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Phe Phe Glu
    850                 855                 860

Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
865                 870                 875                 880

His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
                885                 890                 895
```

```
Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
           900                 905                 910

Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
       915                 920                 925

Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
   930                 935                 940

Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945                 950                 955                 960

Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
               965                 970                 975

Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
           980                 985                 990

Ala Lys Leu Val Ile Glu His Asn Ala Ile Val Val Phe Glu Asp Leu
       995                 1000                1005

Asn Phe  Gly Phe Lys Arg Gly  Arg Phe Lys Val Glu  Lys Gln Val
   1010                1015                1020

Tyr Gln  Lys Leu Glu Lys Met  Leu Ile Glu Lys Leu  Asn Tyr Leu
   1025                1030                1035

Val Phe  Lys Asp Asn Glu Phe  Asp Lys Thr Gly Gly  Val Leu Arg
   1040                1045                1050

Ala Tyr  Gln Leu Thr Ala Pro  Phe Glu Thr Phe Lys  Lys Met Gly
   1055                1060                1065

Lys Gln  Thr Gly Ile Ile Tyr  Tyr Val Pro Ala Gly  Phe Thr Ser
   1070                1075                1080

Lys Ile  Cys Pro Val Thr Gly  Phe Val Asn Gln Leu  Tyr Pro Lys
   1085                1090                1095

Tyr Glu  Ser Val Ser Lys Ser  Gln Glu Phe Phe Ser  Lys Phe Asp
   1100                1105                1110

Lys Ile  Cys Tyr Asn Leu Asp  Lys Gly Tyr Phe Glu  Phe Ser Phe
   1115                1120                1125

Asp Tyr  Lys Asn Phe Gly Asp  Lys Ala Ala Lys Gly  Lys Trp Thr
   1130                1135                1140

Ile Ala  Ser Phe Gly Ser Arg  Leu Ile Asn Phe Arg  Asn Ser Asp
   1145                1150                1155

Lys Asn  His Asn Trp Asp Thr  Arg Glu Val Tyr Pro  Thr Lys Glu
   1160                1165                1170

Leu Glu  Lys Leu Leu Lys Asp  Tyr Ser Ile Glu Tyr  Gly His Gly
   1175                1180                1185

Glu Cys  Ile Lys Ala Ala Ile  Cys Gly Glu Ser Asp  Lys Lys Phe
   1190                1195                1200

Phe Ala  Lys Leu Thr Ser Ile  Leu Asn Thr Ile Leu  Gln Met Arg
   1205                1210                1215
```

```
Asn Ser Lys Thr Gly Thr  Glu Leu Asp Tyr Leu  Ile Ser Pro Val
    1220                 1225                 1230

Ala Asp Val Asn Gly Asn  Phe Phe Asp Ser Arg  Gln Ala Pro Lys
    1235                 1240                 1245

Asn Met Pro Gln Asp Ala  Asp Ala Asn Gly Ala  Tyr His Ile Gly
    1250                 1255                 1260

Leu Lys Gly Leu Met Leu  Leu Asp Arg Ile Lys  Asn Asn Gln Glu
    1265                 1270                 1275

Gly Lys Lys Leu Asn Leu  Val Ile Lys Asn Glu  Glu Tyr Phe Glu
    1280                 1285                 1290

Phe Val Gln Asn Arg Asn  Asn
    1295                 1300
```

<210> 81
<211> 1300
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 81

```
Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1               5                  10                  15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
            20                  25                  30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
        35                  40                  45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
    50                  55                  60

Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65                  70                  75                  80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
                85                  90                  95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
            100                 105                 110

Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
        115                 120                 125

Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
    130                 135                 140

Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145                 150                 155                 160

Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
                165                 170                 175

Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
            180                 185                 190

Asp Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
        195                 200                 205

Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
    210                 215                 220
```

```
Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225                 230                 235                 240

Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
                245                 250                 255

Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
            260                 265                 270

Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
        275                 280                 285

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
    290                 295                 300

Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305                 310                 315                 320

Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
                325                 330                 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
            340                 345                 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
        355                 360                 365

Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
    370                 375                 380

Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385                 390                 395                 400

Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
                405                 410                 415

Val Leu Glu Tyr Ile Thr Gln Gln Val Ala Pro Lys Asn Leu Asp Asn
            420                 425                 430

Pro Ser Lys Lys Glu Gln Asp Leu Ile Ala Lys Lys Thr Glu Lys Ala
        435                 440                 445

Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
    450                 455                 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465                 470                 475                 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
                485                 490                 495

Asp Asn Leu Ala Gln Ile Ser Leu Lys Tyr Gln Asn Gln Gly Lys Lys
            500                 505                 510

Asp Leu Leu Gln Ala Ser Ala Glu Glu Asp Val Lys Ala Ile Lys Asp
        515                 520                 525

Leu Leu Asp Gln Thr Asn Asn Leu Leu His Arg Leu Lys Ile Phe His
    530                 535                 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545                 550                 555                 560
```

Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
565 570 575

Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
580 585 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
595 600 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
610 615 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625 630 635 640

Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
645 650 655

Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
660 665 670

Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
675 680 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Asn Pro Gln
690 695 700

Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
705 710 715 720

Ile Asp Phe Tyr Lys Glu Ser Ile Ser Lys His Pro Glu Trp Lys Asp
725 730 735

Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
740 745 750

Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
755 760 765

Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
770 775 780

Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
785 790 795 800

Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
805 810 815

Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
820 825 830

Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
835 840 845

Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu
850 855 860

Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
865 870 875 880

His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
885 890 895

```
Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
           900                 905                 910

Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
       915                 920                 925

Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
   930                 935                 940

Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945                 950                 955                 960

Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
               965                 970                 975

Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
           980                 985                 990

Ala Lys Leu Val Ile Glu His Asn Ala Ile Val Val Phe Glu Asp Leu
       995                 1000                1005

Asn Phe  Gly Phe Lys Arg Gly  Arg Phe Lys Val Glu  Lys Gln Val
   1010                1015                1020

Tyr Gln  Lys Leu Glu Lys Met  Leu Ile Glu Lys Leu  Asn Tyr Leu
   1025                1030                1035

Val Phe  Lys Asp Asn Glu Phe  Asp Lys Thr Gly Gly  Val Leu Arg
   1040                1045                1050

Ala Tyr  Gln Leu Thr Ala Pro  Phe Glu Thr Phe Lys  Lys Met Gly
   1055                1060                1065

Lys Gln  Thr Gly Ile Ile Tyr  Tyr Val Pro Ala Gly  Phe Thr Ser
   1070                1075                1080

Lys Ile  Cys Pro Val Thr Gly  Phe Val Asn Gln Leu  Tyr Pro Lys
   1085                1090                1095

Tyr Glu  Ser Val Ser Lys Ser  Gln Glu Phe Phe Ser  Lys Phe Asp
   1100                1105                1110

Lys Ile  Cys Tyr Asn Leu Asp  Lys Gly Tyr Phe Glu  Phe Ser Phe
   1115                1120                1125

Asp Tyr  Lys Asn Phe Gly Asp  Lys Ala Ala Lys Gly  Lys Trp Thr
   1130                1135                1140

Ile Ala  Ser Phe Gly Ser Arg  Leu Ile Asn Phe Arg  Asn Ser Asp
   1145                1150                1155

Lys Asn  His Asn Trp Asp Thr  Arg Glu Val Tyr Pro  Thr Lys Glu
   1160                1165                1170

Leu Glu  Lys Leu Leu Lys Asp  Tyr Ser Ile Glu Tyr  Gly His Gly
   1175                1180                1185

Glu Cys  Ile Lys Ala Ala Ile  Cys Gly Glu Ser Asp  Lys Lys Phe
   1190                1195                1200

Phe Ala  Lys Leu Thr Ser Val  Leu Asn Thr Ile Leu  Gln Met Arg
   1205                1210                1215
```

Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val
1220 1225 1230

Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1235 1240 1245

Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly
1250 1255 1260

Leu Lys Gly Leu Met Leu Leu Asp Arg Ile Lys Asn Asn Gln Glu
1265 1270 1275

Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1280 1285 1290

Phe Val Gln Asn Arg Asn Asn
1295 1300

<210> 82
<211> 1300
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 82
Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
20 25 30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
35 40 45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
50 55 60

Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65 70 75 80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
85 90 95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
100 105 110

Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
115 120 125

Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
130 135 140

Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145 150 155 160

Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
165 170 175

Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190

Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205

```
Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
    210                 215                 220

Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225                 230                 235                 240

Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
                245                 250                 255

Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
            260                 265                 270

Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
        275                 280                 285

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
    290                 295                 300

Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305                 310                 315                 320

Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
                325                 330                 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
            340                 345                 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
        355                 360                 365

Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
    370                 375                 380

Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385                 390                 395                 400

Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
                405                 410                 415

Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
            420                 425                 430

Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
        435                 440                 445

Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
    450                 455                 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465                 470                 475                 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
                485                 490                 495

Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
            500                 505                 510

Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
        515                 520                 525

Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
    530                 535                 540
```

```
Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545                 550                 555                 560

Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
                565                 570                 575

Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
            580                 585                 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
        595                 600                 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
    610                 615                 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625                 630                 635                 640

Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
                645                 650                 655

Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
            660                 665                 670

Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
        675                 680                 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln
    690                 695                 700

Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
705                 710                 715                 720

Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp
                725                 730                 735

Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
            740                 745                 750

Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
        755                 760                 765

Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
    770                 775                 780

Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
785                 790                 795                 800

Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
                805                 810                 815

Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
            820                 825                 830

Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
        835                 840                 845

Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu
    850                 855                 860

Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
865                 870                 875                 880
```

```
His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
                885                 890                 895

Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
            900                 905                 910

Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
        915                 920                 925

Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
    930                 935                 940

Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945                 950                 955                 960

Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
                965                 970                 975

Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
            980                 985                 990

Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu
        995                 1000                1005

Asn Phe  Gly Phe Lys Arg Gly  Arg Phe Lys Val Glu  Lys Gln Val
    1010                1015                1020

Tyr Gln  Lys Leu Glu Lys Met  Leu Ile Glu Lys Leu  Asn Tyr Leu
    1025                1030                1035

Val Phe  Lys Asp Asn Glu Phe  Asp Lys Thr Gly Gly  Val Leu Arg
    1040                1045                1050

Ala Tyr  Gln Leu Thr Ala Pro  Phe Glu Thr Phe Lys  Lys Met Gly
    1055                1060                1065

Lys Gln  Thr Gly Ile Ile Tyr  Tyr Val Pro Ala Gly  Phe Thr Ser
    1070                1075                1080

Lys Ile  Cys Pro Val Thr Gly  Phe Val Asn Gln Leu  Tyr Pro Lys
    1085                1090                1095

Tyr Glu  Ser Val Ser Lys Ser  Gln Glu Phe Phe Ser  Lys Phe Asp
    1100                1105                1110

Lys Ile  Cys Tyr Asn Leu Asp  Lys Gly Tyr Phe Glu  Phe Ser Phe
    1115                1120                1125

Asp Tyr  Lys Asn Phe Gly Asp  Lys Ala Ala Lys Gly  Lys Trp Thr
    1130                1135                1140

Ile Ala  Ser Phe Gly Ser Arg  Leu Ile Asn Phe Arg  Asn Ser Asp
    1145                1150                1155

Lys Asn  His Asn Trp Asp Thr  Arg Glu Val Tyr Pro  Thr Lys Glu
    1160                1165                1170

Leu Glu  Lys Leu Leu Lys Asp  Tyr Ser Ile Glu Tyr  Gly His Gly
    1175                1180                1185

Glu Cys  Ile Lys Ala Ala Ile  Cys Gly Glu Ser Asp  Lys Lys Phe
    1190                1195                1200
```

Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg
1205 1210 1215

Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val
1220 1225 1230

Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1235 1240 1245

Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly
1250 1255 1260

Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu
1265 1270 1275

Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1280 1285 1290

Phe Val Gln Asn Arg Asn Asn
1295 1300

<210> 83
<211> 1285
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候选罗斯曼菌（Roizmanbacteria）細菌序列

<400> 83
Met Lys Ser Phe Asp Ser Phe Thr Asn Leu Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Lys Phe Glu Met Arg Pro Val Gly Asn Thr Gln Lys Met Leu Asp
20 25 30

Asn Ala Gly Val Phe Glu Lys Asp Lys Leu Ile Gln Lys Lys Tyr Gly
35 40 45

Lys Thr Lys Pro Tyr Phe Asp Arg Leu His Arg Glu Phe Ile Glu Glu
50 55 60

Ala Leu Thr Gly Val Glu Leu Ile Gly Leu Asp Glu Asn Phe Arg Thr
65 70 75 80

Leu Val Asp Trp Gln Lys Asp Lys Lys Asn Asn Val Ala Met Lys Ala
85 90 95

Tyr Glu Asn Ser Leu Gln Arg Leu Arg Thr Glu Ile Gly Lys Ile Phe
100 105 110

Asn Leu Lys Ala Glu Asp Trp Val Lys Asn Lys Tyr Pro Ile Leu Gly
115 120 125

Leu Lys Asn Lys Asn Thr Asp Ile Leu Phe Glu Glu Ala Val Phe Gly
130 135 140

Ile Leu Lys Ala Arg Tyr Gly Glu Glu Lys Asp Thr Phe Ile Glu Val
145 150 155 160

Glu Glu Ile Asp Lys Thr Gly Lys Ser Lys Ile Asn Gln Ile Ser Ile
165 170 175

Phe Asp Ser Trp Lys Gly Phe Thr Gly Tyr Phe Lys Lys Phe Phe Glu
180 185 190

```
Thr Arg Lys Asn Phe Tyr Lys Asn Asp Gly Thr Ser Thr Ala Ile Ala
        195                 200                 205

Thr Arg Ile Ile Asp Gln Asn Leu Lys Arg Phe Ile Asp Asn Leu Ser
    210                 215                 220

Ile Val Glu Ser Val Arg Gln Lys Val Asp Leu Ala Glu Thr Glu Lys
225                 230                 235                 240

Ser Phe Ser Ile Ser Leu Ser Gln Phe Phe Ser Ile Asp Phe Tyr Asn
                245                 250                 255

Lys Cys Leu Leu Gln Asp Gly Ile Asp Tyr Tyr Asn Lys Ile Ile Gly
            260                 265                 270

Gly Glu Thr Leu Lys Asn Gly Glu Lys Leu Ile Gly Leu Asn Glu Leu
        275                 280                 285

Ile Asn Gln Tyr Arg Gln Asn Asn Lys Asp Gln Lys Ile Pro Phe Phe
    290                 295                 300

Lys Leu Leu Asp Lys Gln Ile Leu Ser Glu Lys Ile Leu Phe Leu Asp
305                 310                 315                 320

Glu Ile Lys Asn Asp Thr Glu Leu Ile Glu Ala Leu Ser Gln Phe Ala
                325                 330                 335

Lys Thr Ala Glu Glu Lys Thr Lys Ile Val Lys Lys Leu Phe Ala Asp
            340                 345                 350

Phe Val Glu Asn Asn Ser Lys Tyr Asp Leu Ala Gln Ile Tyr Ile Ser
        355                 360                 365

Gln Glu Ala Phe Asn Thr Ile Ser Asn Lys Trp Thr Ser Glu Thr Glu
    370                 375                 380

Thr Phe Ala Lys Tyr Leu Phe Glu Ala Met Lys Ser Gly Lys Leu Ala
385                 390                 395                 400

Lys Tyr Glu Lys Lys Asp Asn Ser Tyr Lys Phe Pro Asp Phe Ile Ala
                405                 410                 415

Leu Ser Gln Met Lys Ser Ala Leu Leu Ser Ile Ser Leu Glu Gly His
            420                 425                 430

Phe Trp Lys Glu Lys Tyr Tyr Lys Ile Ser Lys Phe Gln Glu Lys Thr
        435                 440                 445

Asn Trp Glu Gln Phe Leu Ala Ile Phe Leu Tyr Glu Phe Asn Ser Leu
    450                 455                 460

Phe Ser Asp Lys Ile Asn Thr Lys Asp Gly Glu Thr Lys Gln Val Gly
465                 470                 475                 480

Tyr Tyr Leu Phe Ala Lys Asp Leu His Asn Leu Ile Leu Ser Glu Gln
                485                 490                 495

Ile Asp Ile Pro Lys Asp Ser Lys Val Thr Ile Lys Asp Phe Ala Asp
            500                 505                 510

Ser Val Leu Thr Ile Tyr Gln Met Ala Lys Tyr Phe Ala Val Glu Lys
        515                 520                 525
```

Lys Arg Ala Trp Leu Ala Glu Tyr Glu Leu Asp Ser Phe Tyr Thr Gln
530 535 540

Pro Asp Thr Gly Tyr Leu Gln Phe Tyr Asp Asn Ala Tyr Glu Asp Ile
545 550 555 560

Val Gln Val Tyr Asn Lys Leu Arg Asn Tyr Leu Thr Lys Lys Pro Tyr
565 570 575

Ser Glu Glu Lys Trp Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn
580 585 590

Gly Trp Asp Lys Asn Lys Glu Ser Asp Asn Ser Ala Val Ile Leu Gln
595 600 605

Lys Gly Gly Lys Tyr Tyr Leu Gly Leu Ile Thr Lys Gly His Asn Lys
610 615 620

Ile Phe Asp Asp Arg Phe Gln Glu Lys Phe Ile Val Gly Ile Glu Gly
625 630 635 640

Gly Lys Tyr Glu Lys Ile Val Tyr Lys Phe Phe Pro Asp Gln Ala Lys
645 650 655

Met Phe Pro Lys Val Cys Phe Ser Ala Lys Gly Leu Glu Phe Phe Arg
660 665 670

Pro Ser Glu Glu Ile Leu Arg Ile Tyr Asn Asn Ala Glu Phe Lys Lys
675 680 685

Gly Glu Thr Tyr Ser Ile Asp Ser Met Gln Lys Leu Ile Asp Phe Tyr
690 695 700

Lys Asp Cys Leu Thr Lys Tyr Glu Gly Trp Ala Cys Tyr Thr Phe Arg
705 710 715 720

His Leu Lys Pro Thr Glu Glu Tyr Gln Asn Asn Ile Gly Glu Phe Phe
725 730 735

Arg Asp Val Ala Glu Asp Gly Tyr Arg Ile Asp Phe Gln Gly Ile Ser
740 745 750

Asp Gln Tyr Ile His Glu Lys Asn Glu Lys Gly Glu Leu His Leu Phe
755 760 765

Glu Ile His Asn Lys Asp Trp Asn Leu Asp Lys Ala Arg Asp Gly Lys
770 775 780

Ser Lys Thr Thr Gln Lys Asn Leu His Thr Leu Tyr Phe Glu Ser Leu
785 790 795 800

Phe Ser Asn Asp Asn Val Val Gln Asn Phe Pro Ile Lys Leu Asn Gly
805 810 815

Gln Ala Glu Ile Phe Tyr Arg Pro Lys Thr Glu Lys Asp Lys Leu Glu
820 825 830

Ser Lys Lys Asp Lys Lys Gly Asn Lys Val Ile Asp His Lys Arg Tyr
835 840 845

Ser Glu Asn Lys Ile Phe Phe His Val Pro Leu Thr Leu Asn Arg Thr
850 855 860

```
Lys Asn Asp Ser Tyr Arg Phe Asn Ala Gln Ile Asn Asn Phe Leu Ala
865                 870                 875                 880

Asn Asn Lys Asp Ile Asn Ile Ile Gly Val Asp Arg Gly Glu Lys His
                885                 890                 895

Leu Val Tyr Tyr Ser Val Ile Thr Gln Ala Ser Asp Ile Leu Glu Ser
            900                 905                 910

Gly Ser Leu Asn Glu Leu Asn Gly Val Asn Tyr Ala Glu Lys Leu Gly
        915                 920                 925

Lys Lys Ala Glu Asn Arg Glu Gln Ala Arg Arg Asp Trp Gln Asp Val
    930                 935                 940

Gln Gly Ile Lys Asp Leu Lys Lys Gly Tyr Ile Ser Gln Val Val Arg
945                 950                 955                 960

Lys Leu Ala Asp Leu Ala Ile Lys His Asn Ala Ile Ile Ile Leu Glu
                965                 970                 975

Asp Leu Asn Met Arg Phe Lys Gln Val Arg Gly Gly Ile Glu Lys Ser
            980                 985                 990

Ile Tyr Gln Gln Leu Glu Lys Ala Leu Ile Asp Lys Leu Ser Phe Leu
        995                 1000                1005

Val Asp  Lys Gly Glu Lys Asn  Pro Glu Gln Ala Gly  His Leu Leu
    1010                1015                1020

Lys Ala  Tyr Gln Leu Ser Ala  Pro Phe Glu Thr Phe  Gln Lys Met
    1025                1030                1035

Gly Lys  Gln Thr Gly Ile Ile  Phe Tyr Thr Gln Ala  Ser Tyr Thr
    1040                1045                1050

Ser Lys  Ser Asp Pro Val Thr  Gly Trp Arg Pro His  Leu Tyr Leu
    1055                1060                1065

Lys Tyr  Phe Ser Ala Lys Lys  Ala Lys Asp Asp Ile  Ala Lys Phe
    1070                1075                1080

Thr Lys  Ile Glu Phe Val Asn  Asp Arg Phe Glu Leu  Thr Tyr Asp
    1085                1090                1095

Ile Lys  Asp Phe Gln Gln Ala  Lys Glu Tyr Pro Asn  Lys Thr Val
    1100                1105                1110

Trp Lys  Val Cys Ser Asn Val  Glu Arg Phe Arg Trp  Asp Lys Asn
    1115                1120                1125

Leu Asn  Gln Asn Lys Gly Gly  Tyr Thr His Tyr Thr  Asn Ile Thr
    1130                1135                1140

Glu Asn  Ile Gln Glu Leu Phe  Thr Lys Tyr Gly Ile  Asp Ile Thr
    1145                1150                1155

Lys Asp  Leu Leu Thr Gln Ile  Ser Thr Ile Asp Glu  Lys Gln Asn
    1160                1165                1170

Thr Ser  Phe Phe Arg Asp Phe  Ile Phe Tyr Phe Asn  Leu Ile Cys
    1175                1180                1185
```

Gln Ile Arg Asn Thr Asp Asp Ser Glu Ile Ala Lys Lys Asn Gly
1190 1195 1200

Lys Asp Asp Phe Ile Leu Ser Pro Val Glu Pro Phe Phe Asp Ser
1205 1210 1215

Arg Lys Asp Asn Gly Asn Lys Leu Pro Glu Asn Gly Asp Asp Asn
1220 1225 1230

Gly Ala Tyr Asn Ile Ala Arg Lys Gly Ile Val Ile Leu Asn Lys
1235 1240 1245

Ile Ser Gln Tyr Ser Glu Lys Asn Glu Asn Cys Glu Lys Met Lys
1250 1255 1260

Trp Gly Asp Leu Tyr Val Ser Asn Ile Asp Trp Asp Asn Phe Val
1265 1270 1275

Thr Gln Ala Asn Ala Arg His
1280 1285

<210> 84
<211> 1282
<212> PRT
<213> 挑剔真細菌（Eubacterium eligens）

<400> 84
Met Asn Gly Asn Arg Ser Ile Val Tyr Arg Glu Phe Val Gly Val Thr
1 5 10 15

Pro Val Ala Lys Thr Leu Arg Asn Glu Leu Arg Pro Val Gly His Thr
20 25 30

Gln Glu His Ile Ile Gln Asn Gly Leu Ile Gln Glu Asp Glu Leu Arg
35 40 45

Gln Glu Lys Ser Thr Glu Leu Lys Asn Ile Met Asp Asp Tyr Tyr Arg
50 55 60

Glu Tyr Ile Asp Lys Ser Leu Ser Gly Leu Thr Asp Leu Asp Phe Thr
65 70 75 80

Leu Leu Phe Glu Leu Met Asn Ser Val Gln Ser Ser Leu Ser Lys Asp
85 90 95

Asn Lys Lys Ala Leu Glu Lys Glu His Asn Lys Met Arg Glu Gln Ile
100 105 110

Cys Thr His Leu Gln Ser Asp Ser Asp Tyr Lys Asn Met Phe Asn Ala
115 120 125

Lys Leu Phe Lys Glu Ile Leu Pro Asp Phe Ile Lys Asn Tyr Asn Gln
130 135 140

Tyr Asp Val Lys Asp Lys Ala Gly Lys Leu Glu Thr Leu Ala Leu Phe
145 150 155 160

Asn Gly Phe Ser Thr Tyr Phe Thr Asp Phe Phe Glu Lys Arg Lys Asn
165 170 175

Val Phe Thr Lys Glu Ala Val Ser Thr Ser Ile Ala Tyr Arg Ile Val
180 185 190

```
His Glu Asn Ser Leu Ile Phe Leu Ala Asn Met Thr Ser Tyr Lys Lys
        195                 200                 205

Ile Ser Glu Lys Ala Leu Asp Glu Ile Glu Val Ile Glu Lys Asn Asn
    210                 215                 220

Gln Asp Lys Met Gly Asp Trp Glu Leu Asn Gln Ile Phe Asn Pro Asp
225                 230                 235                 240

Phe Tyr Asn Met Val Leu Ile Gln Ser Gly Ile Asp Phe Tyr Asn Glu
                245                 250                 255

Ile Cys Gly Val Val Asn Ala His Met Asn Leu Tyr Cys Gln Gln Thr
            260                 265                 270

Lys Asn Asn Tyr Asn Leu Phe Lys Met Arg Lys Leu His Lys Gln Ile
        275                 280                 285

Leu Ala Tyr Thr Ser Thr Ser Phe Glu Val Pro Lys Met Phe Glu Asp
    290                 295                 300

Asp Met Ser Val Tyr Asn Ala Val Asn Ala Phe Ile Asp Glu Thr Glu
305                 310                 315                 320

Lys Gly Asn Ile Ile Gly Lys Leu Lys Asp Ile Val Asn Lys Tyr Asp
                325                 330                 335

Glu Leu Asp Glu Lys Arg Ile Tyr Ile Ser Lys Asp Phe Tyr Glu Thr
            340                 345                 350

Leu Ser Cys Phe Met Ser Gly Asn Trp Asn Leu Ile Thr Gly Cys Val
        355                 360                 365

Glu Asn Phe Tyr Asp Glu Asn Ile His Ala Lys Gly Lys Ser Lys Glu
    370                 375                 380

Glu Lys Val Lys Lys Ala Val Lys Glu Asp Lys Tyr Lys Ser Ile Asn
385                 390                 395                 400

Asp Val Asn Asp Leu Val Glu Lys Tyr Ile Asp Glu Lys Glu Arg Asn
                405                 410                 415

Glu Phe Lys Asn Ser Asn Ala Lys Gln Tyr Ile Arg Glu Ile Ser Asn
            420                 425                 430

Ile Ile Thr Asp Thr Glu Thr Ala His Leu Glu Tyr Asp Glu His Ile
        435                 440                 445

Ser Leu Ile Glu Ser Glu Glu Lys Ala Asp Glu Ile Lys Lys Arg Leu
    450                 455                 460

Asp Met Tyr Met Asn Met Tyr His Trp Val Lys Ala Phe Ile Val Asp
465                 470                 475                 480

Glu Val Leu Asp Arg Asp Glu Met Phe Tyr Ser Asp Ile Asp Asp Ile
                485                 490                 495

Tyr Asn Ile Leu Glu Asn Ile Val Pro Leu Tyr Asn Arg Val Arg Asn
            500                 505                 510

Tyr Val Thr Gln Lys Pro Tyr Thr Ser Lys Lys Ile Lys Leu Asn Phe
        515                 520                 525
```

```
Gln Ser Pro Thr Leu Ala Asn Gly Trp Ser Gln Ser Lys Glu Phe Asp
    530                 535                 540

Asn Asn Ala Ile Ile Leu Ile Arg Asp Asn Lys Tyr Tyr Leu Ala Ile
545                 550                 555                 560

Phe Asn Ala Lys Asn Lys Pro Asp Lys Lys Ile Ile Gln Gly Asn Ser
                565                 570                 575

Asp Lys Lys Asn Asp Asn Asp Tyr Lys Lys Met Val Tyr Asn Leu Leu
            580                 585                 590

Pro Gly Ala Asn Lys Met Leu Pro Lys Val Phe Leu Ser Lys Lys Gly
        595                 600                 605

Ile Glu Thr Phe Lys Pro Ser Asp Tyr Ile Ile Ser Gly Tyr Asn Ala
    610                 615                 620

His Lys His Ile Lys Thr Ser Glu Asn Phe Asp Ile Ser Phe Cys Arg
625                 630                 635                 640

Asp Leu Ile Asp Tyr Phe Lys Asn Ser Ile Glu Lys His Ala Glu Trp
                645                 650                 655

Arg Lys Tyr Glu Phe Lys Phe Ser Ala Thr Asp Ser Tyr Asn Asp Ile
            660                 665                 670

Ser Glu Phe Tyr Arg Glu Val Glu Met Gln Gly Tyr Arg Ile Asp Trp
        675                 680                 685

Thr Tyr Ile Ser Glu Ala Asp Ile Asn Lys Leu Asp Glu Glu Gly Lys
    690                 695                 700

Ile Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ala Glu Asn Ser Thr
705                 710                 715                 720

Gly Lys Glu Asn Leu His Thr Met Tyr Phe Lys Asn Ile Phe Ser Glu
                725                 730                 735

Glu Asn Leu Lys Asn Ile Val Ile Lys Leu Asn Gly Gln Ala Glu Leu
            740                 745                 750

Phe Tyr Arg Lys Ala Ser Val Lys Asn Pro Val Lys His Lys Lys Asp
        755                 760                 765

Ser Val Leu Val Asn Lys Thr Tyr Lys Asn Gln Leu Asp Asn Gly Asp
    770                 775                 780

Val Val Arg Ile Pro Ile Pro Asp Asp Ile Tyr Asn Glu Ile Tyr Lys
785                 790                 795                 800

Met Tyr Asn Gly Tyr Ile Lys Glu Ser Asp Leu Ser Glu Ala Ala Lys
                805                 810                 815

Glu Tyr Leu Asp Lys Val Glu Val Arg Thr Ala Gln Lys Asp Ile Val
            820                 825                 830

Lys Asp Tyr Arg Tyr Thr Val Asp Lys Tyr Phe Ile His Thr Pro Ile
        835                 840                 845

Thr Ile Asn Tyr Lys Val Thr Ala Arg Asn Asn Val Asn Asp Met Ala
    850                 855                 860
```

```
Val Lys Tyr Ile Ala Gln Asn Asp Asp Ile His Val Ile Gly Ile Asp
865                 870                 875                 880

Arg Gly Glu Arg Asn Leu Ile Tyr Ile Ser Val Ile Asp Ser His Gly
                885                 890                 895

Asn Ile Val Lys Gln Lys Ser Tyr Asn Ile Leu Asn Asn Tyr Asp Tyr
            900                 905                 910

Lys Lys Lys Leu Val Glu Lys Glu Lys Thr Arg Glu Tyr Ala Arg Lys
        915                 920                 925

Asn Trp Lys Ser Ile Gly Asn Ile Lys Glu Leu Lys Glu Gly Tyr Ile
    930                 935                 940

Ser Gly Val Val His Glu Ile Ala Met Leu Met Val Glu Tyr Asn Ala
945                 950                 955                 960

Ile Ile Ala Met Glu Asp Leu Asn Tyr Gly Phe Lys Arg Gly Arg Phe
                965                 970                 975

Lys Val Glu Arg Gln Val Tyr Gln Lys Phe Glu Ser Met Leu Ile Asn
            980                 985                 990

Lys Leu Asn Tyr Phe Ala Ser Lys Gly Lys Ser Val Asp Glu Pro Gly
        995                 1000                1005

Gly Leu  Leu Lys Gly Tyr Gln  Leu Thr Tyr Val Pro  Asp Asn Ile
    1010                1015                1020

Lys Asn  Leu Gly Lys Gln Cys  Gly Val Ile Phe Tyr  Val Pro Ala
    1025                1030                1035

Ala Phe  Thr Ser Lys Ile Asp  Pro Ser Thr Gly Phe  Ile Ser Ala
    1040                1045                1050

Phe Asn  Phe Lys Ser Ile Ser  Thr Asn Ala Ser Arg  Lys Gln Phe
    1055                1060                1065

Phe Met  Gln Phe Asp Glu Ile  Arg Tyr Cys Ala Glu  Lys Asp Met
    1070                1075                1080

Phe Ser  Phe Gly Phe Asp Tyr  Asn Asn Phe Asp Thr  Tyr Asn Ile
    1085                1090                1095

Thr Met  Gly Lys Thr Gln Trp  Thr Val Tyr Thr Asn  Gly Glu Arg
    1100                1105                1110

Leu Gln  Ser Glu Phe Asn Asn  Ala Arg Arg Thr Gly  Lys Thr Lys
    1115                1120                1125

Ser Ile  Asn Leu Thr Glu Thr  Ile Lys Leu Leu Leu  Glu Asp Asn
    1130                1135                1140

Glu Ile  Asn Tyr Ala Asp Gly  His Asp Val Arg Ile  Asp Met Glu
    1145                1150                1155

Lys Met  Tyr Glu Asp Lys Asn  Ser Glu Phe Phe Ala  Gln Leu Leu
    1160                1165                1170

Ser Leu  Tyr Lys Leu Thr Val  Gln Met Arg Asn Ser  Tyr Thr Glu
    1175                1180                1185
```

Ala Glu Glu Gln Glu Lys Gly Ile Ser Tyr Asp Lys Ile Ile Ser
1190 1195 1200

Pro Val Ile Asn Asp Glu Gly Glu Phe Phe Asp Ser Asp Asn Tyr
1205 1210 1215

Lys Glu Ser Asp Asp Lys Glu Cys Lys Met Pro Lys Asp Ala Asp
1220 1225 1230

Ala Asn Gly Ala Tyr Cys Ile Ala Leu Lys Gly Leu Tyr Glu Val
1235 1240 1245

Leu Lys Ile Lys Ser Glu Trp Thr Glu Asp Gly Phe Asp Arg Asn
1250 1255 1260

Cys Leu Lys Leu Pro His Ala Glu Trp Leu Asp Phe Ile Gln Asn
1265 1270 1275

Lys Arg Tyr Glu
1280

<210> 85
<211> 1282
<212> PRT
<213> 挑剔真細菌

<400> 85
Met Asn Gly Asn Arg Ser Ile Val Tyr Arg Glu Phe Val Gly Val Ile
1 5 10 15

Pro Val Ala Lys Thr Leu Arg Asn Glu Leu Arg Pro Val Gly His Thr
20 25 30

Gln Glu His Ile Ile Gln Asn Gly Leu Ile Gln Glu Asp Glu Leu Arg
35 40 45

Gln Glu Lys Ser Thr Glu Leu Lys Asn Ile Met Asp Asp Tyr Tyr Arg
50 55 60

Glu Tyr Ile Asp Lys Ser Leu Ser Gly Val Thr Asp Leu Asp Phe Thr
65 70 75 80

Leu Leu Phe Glu Leu Met Asn Leu Val Gln Ser Ser Pro Ser Lys Asp
85 90 95

Asn Lys Lys Ala Leu Glu Lys Glu Gln Ser Lys Met Arg Glu Gln Ile
100 105 110

Cys Thr His Leu Gln Ser Asp Ser Asn Tyr Lys Asn Ile Phe Asn Ala
115 120 125

Lys Leu Leu Lys Glu Ile Leu Pro Asp Phe Ile Lys Asn Tyr Asn Gln
130 135 140

Tyr Asp Val Lys Asp Lys Ala Gly Lys Leu Glu Thr Leu Ala Leu Phe
145 150 155 160

Asn Gly Phe Ser Thr Tyr Phe Thr Asp Phe Phe Glu Lys Arg Lys Asn
165 170 175

Val Phe Thr Lys Glu Ala Val Ser Thr Ser Ile Ala Tyr Arg Ile Val
180 185 190

His Glu Asn Ser Leu Ile Phe Leu Ala Asn Met Thr Ser Tyr Lys Lys
195 200 205

```
Ile Ser Glu Lys Ala Leu Asp Glu Ile Glu Val Ile Glu Lys Asn Asn
    210                 215                 220

Gln Asp Lys Met Gly Asp Trp Glu Leu Asn Gln Ile Phe Asn Pro Asp
225                 230                 235                 240

Phe Tyr Asn Met Val Leu Ile Gln Ser Gly Ile Asp Phe Tyr Asn Glu
                245                 250                 255

Ile Cys Gly Val Val Asn Ala His Met Asn Leu Tyr Cys Gln Gln Thr
            260                 265                 270

Lys Asn Asn Tyr Asn Leu Phe Lys Met Arg Lys Leu His Lys Gln Ile
        275                 280                 285

Leu Ala Tyr Thr Ser Thr Ser Phe Glu Val Pro Lys Met Phe Glu Asp
    290                 295                 300

Asp Met Ser Val Tyr Asn Ala Val Asn Ala Phe Ile Asp Glu Thr Glu
305                 310                 315                 320

Lys Gly Asn Ile Ile Gly Lys Leu Lys Asp Ile Val Asn Lys Tyr Asp
                325                 330                 335

Glu Leu Asp Glu Lys Arg Ile Tyr Ile Ser Lys Asp Phe Tyr Glu Thr
            340                 345                 350

Leu Ser Cys Phe Met Ser Gly Asn Trp Asn Leu Ile Thr Gly Cys Val
        355                 360                 365

Glu Asn Phe Tyr Asp Glu Asn Ile His Ala Lys Gly Lys Ser Lys Glu
    370                 375                 380

Glu Lys Val Lys Lys Ala Val Lys Glu Asp Lys Tyr Lys Ser Ile Asn
385                 390                 395                 400

Asp Val Asn Asp Leu Val Glu Lys Tyr Ile Asp Glu Lys Glu Arg Asn
                405                 410                 415

Glu Phe Lys Asn Ser Asn Ala Lys Gln Tyr Ile Arg Glu Ile Ser Asn
            420                 425                 430

Ile Ile Thr Asp Thr Glu Thr Ala His Leu Glu Tyr Asp Asp His Ile
        435                 440                 445

Ser Leu Ile Glu Ser Glu Glu Lys Ala Asp Glu Met Lys Lys Arg Leu
    450                 455                 460

Asp Met Tyr Met Asn Met Tyr His Trp Ala Lys Ala Phe Ile Val Asp
465                 470                 475                 480

Glu Val Leu Asp Arg Asp Glu Met Phe Tyr Ser Asp Ile Asp Asp Ile
                485                 490                 495

Tyr Asn Ile Leu Glu Asn Ile Val Pro Leu Tyr Asn Arg Val Arg Asn
            500                 505                 510

Tyr Val Thr Gln Lys Pro Tyr Asn Ser Lys Lys Ile Lys Leu Asn Phe
        515                 520                 525

Gln Ser Pro Thr Leu Ala Asn Gly Trp Ser Gln Ser Lys Glu Phe Asp
    530                 535                 540
```

Asn Asn Ala Ile Ile Leu Ile Arg Asp Asn Lys Tyr Tyr Leu Ala Ile
545 550 555 560

Phe Asn Ala Lys Asn Lys Pro Asp Lys Lys Ile Ile Gln Gly Asn Ser
565 570 575

Asp Lys Lys Asn Asp Asn Asp Tyr Lys Lys Met Val Tyr Asn Leu Leu
580 585 590

Pro Gly Ala Asn Lys Met Leu Pro Lys Val Phe Leu Ser Lys Lys Gly
595 600 605

Ile Glu Thr Phe Lys Pro Ser Asp Tyr Ile Ile Ser Gly Tyr Asn Ala
610 615 620

His Lys His Ile Lys Thr Ser Glu Asn Phe Asp Ile Ser Phe Cys Arg
625 630 635 640

Asp Leu Ile Asp Tyr Phe Lys Asn Ser Ile Glu Lys His Ala Glu Trp
645 650 655

Arg Lys Tyr Glu Phe Lys Phe Ser Ala Thr Asp Ser Tyr Ser Asp Ile
660 665 670

Ser Glu Phe Tyr Arg Glu Val Glu Met Gln Gly Tyr Arg Ile Asp Trp
675 680 685

Thr Tyr Ile Ser Glu Ala Asp Ile Asn Lys Leu Asp Glu Glu Gly Lys
690 695 700

Ile Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ala Glu Asn Ser Thr
705 710 715 720

Gly Lys Glu Asn Leu His Thr Met Tyr Phe Lys Asn Ile Phe Ser Glu
725 730 735

Glu Asn Leu Lys Asp Ile Ile Ile Lys Leu Asn Gly Gln Ala Glu Leu
740 745 750

Phe Tyr Arg Arg Ala Ser Val Lys Asn Pro Val Lys His Lys Lys Asp
755 760 765

Ser Val Leu Val Asn Lys Thr Tyr Lys Asn Gln Leu Asp Asn Gly Asp
770 775 780

Val Val Arg Ile Pro Ile Pro Asp Asp Ile Tyr Asn Glu Ile Tyr Lys
785 790 795 800

Met Tyr Asn Gly Tyr Ile Lys Glu Ser Asp Leu Ser Glu Ala Ala Lys
805 810 815

Glu Tyr Leu Asp Lys Val Glu Val Arg Thr Ala Gln Lys Asp Ile Val
820 825 830

Lys Asp Tyr Arg Tyr Thr Val Asp Lys Tyr Phe Ile His Thr Pro Ile
835 840 845

Thr Ile Asn Tyr Lys Val Thr Ala Arg Asn Asn Val Asn Asp Met Val
850 855 860

Val Lys Tyr Ile Ala Gln Asn Asp Asp Ile His Val Ile Gly Ile Asp
865 870 875 880

Arg Gly Glu Arg Asn Leu Ile Tyr Ile Ser Val Ile Asp Ser His Gly
885 890 895

Asn Ile Val Lys Gln Lys Ser Tyr Asn Ile Leu Asn Asn Tyr Asp Tyr
900 905 910

Lys Lys Lys Leu Val Glu Lys Glu Lys Thr Arg Glu Tyr Ala Arg Lys
915 920 925

Asn Trp Lys Ser Ile Gly Asn Ile Lys Glu Leu Lys Glu Gly Tyr Ile
930 935 940

Ser Gly Val Val His Glu Ile Ala Met Leu Ile Val Glu Tyr Asn Ala
945 950 955 960

Ile Ile Ala Met Glu Asp Leu Asn Tyr Gly Phe Lys Arg Gly Arg Phe
965 970 975

Lys Val Glu Arg Gln Val Tyr Gln Lys Phe Glu Ser Met Leu Ile Asn
980 985 990

Lys Leu Asn Tyr Phe Ala Ser Lys Glu Lys Ser Val Asp Glu Pro Gly
995 1000 1005

Gly Leu Leu Lys Gly Tyr Gln Leu Thr Tyr Val Pro Asp Asn Ile
1010 1015 1020

Lys Asn Leu Gly Lys Gln Cys Gly Val Ile Phe Tyr Val Pro Ala
1025 1030 1035

Ala Phe Thr Ser Lys Ile Asp Pro Ser Thr Gly Phe Ile Ser Ala
1040 1045 1050

Phe Asn Phe Lys Ser Ile Ser Thr Asn Ala Ser Arg Lys Gln Phe
1055 1060 1065

Phe Met Gln Phe Asp Glu Ile Arg Tyr Cys Ala Glu Lys Asp Met
1070 1075 1080

Phe Ser Phe Gly Phe Asp Tyr Asn Asn Phe Asp Thr Tyr Asn Ile
1085 1090 1095

Thr Met Gly Lys Thr Gln Trp Thr Val Tyr Thr Asn Gly Glu Arg
1100 1105 1110

Leu Gln Ser Glu Phe Asn Asn Ala Arg Arg Thr Gly Lys Thr Lys
1115 1120 1125

Ser Ile Asn Leu Thr Glu Thr Ile Lys Leu Leu Leu Glu Asp Asn
1130 1135 1140

Glu Ile Asn Tyr Ala Asp Gly His Asp Ile Arg Ile Asp Met Glu
1145 1150 1155

Lys Met Asp Glu Asp Lys Lys Ser Glu Phe Phe Ala Gln Leu Leu
1160 1165 1170

Ser Leu Tyr Lys Leu Thr Val Gln Met Arg Asn Ser Tyr Thr Glu
1175 1180 1185

Ala Glu Glu Gln Glu Asn Gly Ile Ser Tyr Asp Lys Ile Ile Ser
1190 1195 1200

Pro Val Ile Asn Asp Glu Gly Glu Phe Phe Asp Ser Asp Asn Tyr
1205 1210 1215

Lys Glu Ser Asp Asp Lys Glu Cys Lys Met Pro Lys Asp Ala Asp
1220 1225 1230

Ala Asn Gly Ala Tyr Cys Ile Ala Leu Lys Gly Leu Tyr Glu Val
1235 1240 1245

Leu Lys Ile Lys Ser Glu Trp Thr Glu Asp Gly Phe Asp Arg Asn
1250 1255 1260

Cys Leu Lys Leu Pro His Ala Glu Trp Leu Asp Phe Ile Gln Asn
1265 1270 1275

Lys Arg Tyr Glu
1280

<210> 86
<211> 1273
<212> PRT
<213> 黃桿菌屬某種（Flavobacterium sp.）

<400> 86
Met Lys Asn Phe Ser Asn Leu Tyr Gln Val Ser Lys Thr Val Arg Phe
1 5 10 15

Glu Leu Lys Pro Ile Gly Asn Thr Leu Glu Asn Ile Lys Asn Lys Ser
20 25 30

Leu Leu Lys Asn Asp Ser Ile Arg Ala Glu Ser Tyr Gln Lys Met Lys
35 40 45

Lys Thr Ile Asp Glu Phe His Lys Tyr Phe Ile Asp Leu Ala Leu Asn
50 55 60

Asn Lys Lys Leu Ser Tyr Leu Asn Glu Tyr Ile Ala Leu Tyr Thr Gln
65 70 75 80

Ser Ala Glu Ala Lys Lys Glu Asp Lys Phe Lys Ala Asp Phe Lys Lys
85 90 95

Val Gln Asp Asn Leu Arg Lys Glu Ile Val Ser Ser Phe Thr Glu Gly
100 105 110

Glu Ala Lys Ala Ile Phe Ser Val Leu Asp Lys Lys Glu Leu Ile Thr
115 120 125

Ile Glu Leu Glu Lys Trp Lys Asn Glu Asn Asn Leu Ala Val Tyr Leu
130 135 140

Asp Glu Ser Phe Lys Ser Phe Thr Thr Tyr Phe Thr Gly Phe His Gln
145 150 155 160

Asn Arg Lys Asn Met Tyr Ser Ala Glu Ala Asn Ser Thr Ala Ile Ala
165 170 175

Tyr Arg Leu Ile His Glu Asn Leu Pro Lys Phe Ile Glu Asn Ser Lys
180 185 190

Ala Phe Glu Lys Ser Ser Gln Ile Ala Glu Leu Gln Pro Lys Ile Glu
195 200 205

```
Lys Leu Tyr Lys Glu Phe Glu Ala Tyr Leu Asn Val Asn Ser Ile Ser
    210                 215                 220

Glu Leu Phe Glu Ile Asp Tyr Phe Asn Glu Val Leu Thr Gln Lys Gly
225                 230                 235                 240

Ile Thr Val Tyr Asn Asn Ile Ile Gly Gly Arg Thr Ala Thr Glu Gly
                245                 250                 255

Lys Gln Lys Ile Gln Gly Leu Asn Glu Ile Ile Asn Leu Tyr Asn Gln
            260                 265                 270

Thr Lys Pro Lys Asn Glu Arg Leu Pro Lys Leu Lys Gln Leu Tyr Lys
        275                 280                 285

Gln Ile Leu Ser Asp Arg Ile Ser Leu Ser Phe Leu Pro Asp Ala Phe
    290                 295                 300

Thr Glu Gly Lys Gln Val Leu Lys Ala Val Phe Glu Phe Tyr Lys Ile
305                 310                 315                 320

Asn Leu Leu Ser Tyr Lys Gln Asp Gly Val Glu Glu Ser Gln Asn Leu
                325                 330                 335

Leu Glu Leu Ile Gln Gln Val Val Lys Asn Leu Gly Asn Gln Asp Val
            340                 345                 350

Asn Lys Ile Tyr Leu Lys Asn Asp Thr Ser Leu Thr Thr Ile Ala Gln
        355                 360                 365

Gln Leu Phe Gly Asp Phe Ser Val Phe Ser Ala Ala Leu Gln Tyr Arg
    370                 375                 380

Tyr Glu Thr Val Val Asn Pro Lys Tyr Thr Ala Glu Tyr Gln Lys Ala
385                 390                 395                 400

Asn Glu Ala Lys Gln Glu Lys Leu Asp Lys Glu Lys Ile Lys Phe Val
                405                 410                 415

Lys Gln Asp Tyr Phe Ser Ile Ala Phe Leu Gln Glu Val Val Ala Asp
            420                 425                 430

Tyr Val Lys Thr Leu Asp Glu Asn Leu Asp Trp Lys Gln Lys Tyr Thr
        435                 440                 445

Pro Ser Cys Ile Ala Asp Tyr Phe Thr Thr His Phe Ile Ala Lys Lys
    450                 455                 460

Glu Asn Glu Ala Asp Lys Thr Phe Asn Phe Ile Ala Asn Ile Lys Ala
465                 470                 475                 480

Lys Tyr Gln Cys Ile Gln Gly Ile Leu Glu Gln Ala Asp Asp Tyr Glu
                485                 490                 495

Asp Glu Leu Lys Gln Asp Gln Lys Leu Ile Asp Asn Ile Lys Phe Phe
            500                 505                 510

Leu Asp Ala Ile Leu Glu Val Val His Phe Ile Lys Pro Leu His Leu
        515                 520                 525

Lys Ser Glu Ser Ile Thr Glu Lys Asp Asn Ala Phe Tyr Asp Val Phe
    530                 535                 540
```

```
Glu Asn Tyr Tyr Glu Ala Leu Asn Val Val Thr Pro Leu Tyr Asn Met
545                 550                 555                 560

Val Arg Asn Tyr Val Thr Gln Lys Pro Tyr Ser Thr Glu Lys Ile Lys
                565                 570                 575

Leu Asn Phe Glu Asn Ala Gln Leu Leu Asn Gly Trp Asp Ala Asn Lys
            580                 585                 590

Glu Lys Asp Tyr Leu Thr Thr Ile Leu Lys Arg Asp Gly Asn Tyr Phe
        595                 600                 605

Leu Ala Ile Met Asp Lys Lys His Asn Lys Thr Phe Gln Gln Phe Thr
    610                 615                 620

Glu Asp Asp Glu Asn Tyr Glu Lys Ile Val Tyr Lys Leu Leu Pro Gly
625                 630                 635                 640

Val Asn Lys Met Leu Pro Lys Val Phe Phe Ser Asn Lys Asn Ile Ala
                645                 650                 655

Phe Phe Asn Pro Ser Lys Glu Ile Leu Asp Asn Tyr Lys Asn Asn Thr
            660                 665                 670

His Lys Lys Gly Ala Thr Phe Asn Leu Lys Asp Cys His Ala Leu Ile
        675                 680                 685

Asp Phe Phe Lys Asp Ser Leu Asn Lys His Glu Asp Trp Lys Tyr Phe
    690                 695                 700

Asp Phe Gln Phe Ser Glu Thr Lys Thr Tyr Gln Asp Leu Ser Gly Phe
705                 710                 715                 720

Tyr Lys Glu Val Glu His Gln Gly Tyr Lys Ile Asn Phe Lys Lys Val
                725                 730                 735

Ser Val Ser Gln Ile Asp Thr Leu Ile Glu Glu Gly Lys Met Tyr Leu
            740                 745                 750

Phe Gln Ile Tyr Asn Lys Asp Phe Ser Pro Tyr Ala Lys Gly Lys Pro
        755                 760                 765

Asn Met His Thr Leu Tyr Trp Lys Ala Leu Phe Glu Thr Gln Asn Leu
    770                 775                 780

Glu Asn Val Ile Tyr Lys Leu Asn Gly Gln Ala Glu Ile Phe Phe Arg
785                 790                 795                 800

Lys Ala Ser Ile Lys Lys Lys Asn Ile Ile Thr His Lys Ala His Gln
                805                 810                 815

Pro Ile Ala Ala Lys Asn Pro Leu Thr Pro Thr Ala Lys Asn Thr Phe
            820                 825                 830

Ala Tyr Asp Leu Ile Lys Asp Lys Arg Tyr Thr Val Asp Lys Phe Gln
        835                 840                 845

Phe His Val Pro Ile Thr Met Asn Phe Lys Ala Thr Gly Asn Ser Tyr
    850                 855                 860

Ile Asn Gln Asp Val Leu Ala Tyr Leu Lys Asp Asn Pro Glu Val Asn
865                 870                 875                 880
```

```
Ile Ile Gly Leu Asp Arg Gly Glu Arg His Leu Val Tyr Leu Thr Leu
                885                 890                 895

Ile Asp Gln Lys Gly Thr Ile Leu Leu Gln Glu Ser Leu Asn Val Ile
            900                 905                 910

Gln Asp Glu Lys Thr His Thr Pro Tyr His Thr Leu Leu Asp Asn Lys
        915                 920                 925

Glu Ile Ala Arg Asp Lys Ala Arg Lys Asn Trp Gly Ser Ile Glu Ser
    930                 935                 940

Ile Lys Glu Leu Lys Glu Gly Tyr Ile Ser Gln Val Val His Lys Ile
945                 950                 955                 960

Thr Lys Met Met Ile Glu His Asn Ala Ile Val Val Met Glu Asp Leu
                965                 970                 975

Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Ile Tyr
            980                 985                 990

Gln Lys Leu Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr Leu Val Leu
        995                 1000                1005

Lys Asp  Lys Gln Pro His Glu  Leu Gly Gly Leu Tyr  Asn Ala Leu
    1010                1015                1020

Gln Leu  Thr Asn Lys Phe Glu  Ser Phe Gln Lys Met  Gly Lys Gln
    1025                1030                1035

Ser Gly  Phe Leu Phe Tyr Val  Pro Ala Trp Asn Thr  Ser Lys Ile
    1040                1045                1050

Asp Pro  Thr Thr Gly Phe Val  Asn Tyr Phe Tyr Thr  Lys Tyr Glu
    1055                1060                1065

Asn Val  Glu Lys Ala Lys Thr  Phe Phe Ser Lys Phe  Asp Ser Ile
    1070                1075                1080

Leu Tyr  Asn Lys Thr Lys Gly  Tyr Phe Glu Phe Val  Val Lys Asn
    1085                1090                1095

Tyr Ser  Asp Phe Asn Pro Lys  Ala Ala Asp Thr Arg  Gln Glu Trp
    1100                1105                1110

Thr Ile  Cys Thr His Gly Glu  Arg Ile Glu Thr Lys  Arg Gln Lys
    1115                1120                1125

Glu Gln  Asn Asn Asn Phe Val  Ser Thr Thr Ile Gln  Leu Thr Glu
    1130                1135                1140

Gln Phe  Val Asn Phe Phe Glu  Lys Val Gly Leu Asp  Leu Ser Lys
    1145                1150                1155

Glu Leu  Lys Thr Gln Leu Ile  Ala Gln Asn Glu Lys  Ser Phe Phe
    1160                1165                1170

Glu Glu  Leu Phe His Leu Leu  Lys Leu Thr Leu Gln  Met Arg Asn
    1175                1180                1185

Ser Glu  Ser His Thr Glu Ile  Asp Tyr Leu Ile Ser  Pro Val Ala
    1190                1195                1200
```

Asn Glu Lys Gly Ile Phe Tyr Asp Ser Arg Lys Ala Thr Ala Ser
1205 1210 1215

Leu Pro Ile Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Ala Lys
1220 1225 1230

Lys Gly Leu Trp Ile Met Glu Gln Ile Asn Lys Thr Asn Ser Glu
1235 1240 1245

Asp Asp Leu Lys Lys Val Lys Leu Ala Ile Ser Asn Arg Glu Trp
1250 1255 1260

Leu Gln Tyr Val Gln Gln Val Gln Lys Lys
1265 1270

<210> 87
<211> 1264
<212> PRT
<213> 短普雷沃菌（Prevotella brevis）

<400> 87
Met Lys Gln Phe Thr Asn Leu Tyr Gln Leu Ser Lys Thr Leu Arg Phe
1 5 10 15

Glu Leu Lys Pro Ile Gly Lys Thr Leu Glu His Ile Asn Ala Asn Gly
20 25 30

Phe Ile Asp Asn Asp Ala His Arg Ala Glu Ser Tyr Lys Lys Val Lys
35 40 45

Lys Leu Ile Asp Asp Tyr His Lys Asp Tyr Ile Glu Asn Val Leu Asn
50 55 60

Asn Phe Lys Leu Asn Gly Glu Tyr Leu Gln Ala Tyr Phe Asp Leu Tyr
65 70 75 80

Ser Gln Asp Thr Lys Asp Lys Gln Phe Lys Asp Ile Gln Asp Lys Leu
85 90 95

Arg Lys Ser Ile Ala Ser Ala Leu Lys Gly Asp Asp Arg Tyr Lys Thr
100 105 110

Ile Asp Lys Lys Glu Leu Ile Arg Gln Asp Met Lys Thr Phe Leu Lys
115 120 125

Lys Asp Thr Asp Lys Ala Leu Leu Asp Glu Phe Tyr Glu Phe Thr Thr
130 135 140

Tyr Phe Thr Gly Tyr His Glu Asn Arg Lys Asn Met Tyr Ser Asp Glu
145 150 155 160

Ala Lys Ser Thr Ala Ile Ala Tyr Arg Leu Ile His Asp Asn Leu Pro
165 170 175

Lys Phe Ile Asp Asn Ile Ala Val Phe Lys Lys Ile Ala Asn Thr Ser
180 185 190

Val Ala Asp Asn Phe Ser Thr Ile Tyr Lys Asn Phe Glu Glu Tyr Leu
195 200 205

Asn Val Asn Ser Ile Asp Glu Ile Phe Ser Leu Asp Tyr Tyr Asn Ile
210 215 220

Val Leu Thr Gln Thr Gln Ile Glu Val Tyr Asn Ser Ile Ile Gly Gly
225 230 235 240

```
Arg Thr Leu Glu Asp Asp Thr Lys Ile Gln Gly Ile Asn Glu Phe Val
                245                 250                 255

Asn Leu Tyr Asn Gln Gln Leu Ala Asn Lys Lys Asp Arg Leu Pro Lys
            260                 265                 270

Leu Lys Pro Leu Phe Lys Gln Ile Leu Ser Asp Arg Val Gln Leu Ser
        275                 280                 285

Trp Leu Gln Glu Glu Phe Asn Thr Gly Ala Asp Val Leu Asn Ala Val
    290                 295                 300

Lys Glu Tyr Cys Thr Ser Tyr Phe Asp Asn Val Glu Glu Ser Val Lys
305                 310                 315                 320

Val Leu Leu Thr Gly Ile Ser Asp Tyr Asp Leu Ser Lys Ile Tyr Ile
                325                 330                 335

Thr Asn Asp Leu Ala Leu Thr Asp Val Ser Gln Arg Met Phe Gly Glu
            340                 345                 350

Trp Ser Ile Ile Pro Asn Ala Ile Glu Gln Arg Leu Arg Ser Asp Asn
        355                 360                 365

Pro Lys Lys Thr Asn Glu Lys Glu Glu Lys Tyr Ser Asp Arg Ile Ser
    370                 375                 380

Lys Leu Lys Lys Leu Pro Lys Ser Tyr Ser Leu Gly Tyr Ile Asn Glu
385                 390                 395                 400

Cys Ile Ser Glu Leu Asn Gly Ile Asp Ile Ala Asp Tyr Tyr Ala Thr
                405                 410                 415

Leu Gly Ala Ile Asn Thr Glu Ser Lys Gln Glu Pro Ser Ile Pro Thr
            420                 425                 430

Ser Ile Gln Val His Tyr Asn Ala Leu Lys Pro Ile Leu Asp Thr Asp
        435                 440                 445

Tyr Pro Arg Glu Lys Asn Leu Ser Gln Asp Lys Leu Thr Val Met Gln
    450                 455                 460

Leu Lys Asp Leu Leu Asp Asp Phe Lys Ala Leu Gln His Phe Ile Lys
465                 470                 475                 480

Pro Leu Leu Gly Asn Gly Asp Glu Ala Glu Lys Asp Glu Lys Phe Tyr
                485                 490                 495

Gly Glu Leu Met Gln Leu Trp Glu Val Ile Asp Ser Ile Thr Pro Leu
            500                 505                 510

Tyr Asn Lys Val Arg Asn Tyr Cys Thr Arg Lys Pro Phe Ser Thr Glu
        515                 520                 525

Lys Ile Lys Val Asn Phe Glu Asn Ala Gln Leu Leu Asp Gly Trp Asp
    530                 535                 540

Glu Asn Lys Glu Ser Thr Asn Ala Ser Ile Ile Leu Arg Lys Asn Gly
545                 550                 555                 560

Met Tyr Tyr Leu Gly Ile Met Lys Lys Glu Tyr Arg Asn Ile Leu Thr
                565                 570                 575
```

Lys Pro Met Pro Ser Asp Gly Asp Cys Tyr Asp Lys Val Val Tyr Lys
580 585 590

Phe Phe Lys Asp Ile Thr Thr Met Val Pro Lys Cys Thr Thr Gln Met
595 600 605

Lys Ser Val Lys Glu His Phe Ser Asn Ser Asn Asp Asp Tyr Thr Leu
610 615 620

Phe Glu Lys Asp Lys Phe Ile Ala Pro Val Val Ile Thr Lys Glu Ile
625 630 635 640

Phe Asp Leu Asn Asn Val Leu Tyr Asn Gly Val Lys Lys Phe Gln Ile
645 650 655

Gly Tyr Leu Asn Asn Thr Gly Asp Ser Phe Gly Tyr Asn His Ala Val
660 665 670

Glu Ile Trp Lys Ser Phe Cys Leu Lys Phe Leu Lys Ala Tyr Lys Ser
675 680 685

Thr Ser Ile Tyr Asp Phe Ser Ser Ile Glu Lys Asn Ile Gly Cys Tyr
690 695 700

Asn Asp Leu Asn Ser Phe Tyr Gly Ala Val Asn Leu Leu Leu Tyr Asn
705 710 715 720

Leu Thr Tyr Arg Lys Val Ser Val Asp Tyr Ile His Gln Leu Val Asp
725 730 735

Glu Asp Lys Met Tyr Leu Phe Met Ile Tyr Asn Lys Asp Phe Ser Thr
740 745 750

Tyr Ser Lys Gly Thr Pro Asn Met His Thr Leu Tyr Trp Lys Met Leu
755 760 765

Phe Asp Glu Ser Asn Leu Asn Asp Val Val Tyr Lys Leu Asn Gly Gln
770 775 780

Ala Glu Val Phe Tyr Arg Lys Lys Ser Ile Thr Tyr Gln His Pro Thr
785 790 795 800

His Pro Ala Asn Lys Pro Ile Asp Asn Lys Asn Val Asn Asn Pro Lys
805 810 815

Lys Gln Ser Asn Phe Glu Tyr Asp Leu Ile Lys Asp Lys Arg Tyr Thr
820 825 830

Val Asp Lys Phe Met Phe His Val Pro Ile Thr Leu Asn Phe Lys Gly
835 840 845

Met Gly Asn Gly Asp Ile Asn Met Gln Val Arg Glu Tyr Ile Lys Thr
850 855 860

Thr Asp Asp Leu His Phe Ile Gly Ile Asp Arg Gly Glu Arg His Leu
865 870 875 880

Leu Tyr Ile Cys Val Ile Asn Gly Lys Gly Glu Ile Val Glu Gln Tyr
885 890 895

Ser Leu Asn Glu Ile Val Asn Asn Tyr Lys Gly Thr Glu Tyr Lys Thr
900 905 910

```
Asp Tyr His Thr Leu Leu Ser Glu Arg Asp Lys Lys Arg Lys Glu Glu
        915                 920                 925

Arg Ser Ser Trp Gln Thr Ile Glu Gly Ile Lys Glu Leu Lys Ser Gly
    930                 935                 940

Tyr Leu Ser Gln Val Ile His Lys Ile Thr Gln Leu Met Ile Lys Tyr
945                 950                 955                 960

Asn Ala Ile Val Leu Leu Glu Asp Leu Asn Met Gly Phe Lys Arg Gly
                965                 970                 975

Arg Gln Lys Val Glu Ser Ser Val Tyr Gln Gln Phe Glu Lys Ala Leu
            980                 985                 990

Ile Asp Lys Leu Asn Tyr Leu Val Asp Lys Asn Lys Asp Ala Asn Glu
        995                 1000                1005

Ile Gly  Gly Leu Leu His Ala  Tyr Gln Leu Thr Asn  Asp Pro Lys
    1010                1015                1020

Leu Pro  Asn Lys Asn Ser Lys  Gln Ser Gly Phe Leu  Phe Tyr Val
    1025                1030                1035

Pro Ala  Trp Asn Thr Ser Lys  Ile Asp Pro Val Thr  Gly Phe Val
    1040                1045                1050

Asn Leu  Leu Asp Thr Arg Tyr  Glu Asn Val Ala Lys  Ala Gln Ala
    1055                1060                1065

Phe Phe  Lys Lys Phe Asp Ser  Ile Arg Tyr Asn Lys  Glu Tyr Asp
    1070                1075                1080

Arg Phe  Glu Phe Lys Phe Asp  Tyr Ser Asn Phe Thr  Ala Lys Ala
    1085                1090                1095

Glu Asp  Thr Arg Thr Gln Trp  Thr Leu Cys Thr Tyr  Gly Thr Arg
    1100                1105                1110

Ile Glu  Thr Phe Arg Asn Ala  Glu Lys Asn Ser Asn  Trp Asp Ser
    1115                1120                1125

Arg Glu  Ile Asp Leu Thr Thr  Glu Trp Lys Thr Leu  Phe Thr Gln
    1130                1135                1140

His Asn  Ile Pro Leu Asn Ala  Asn Leu Lys Glu Ala  Ile Leu Leu
    1145                1150                1155

Gln Ala  Asn Lys Asn Phe Tyr  Thr Asp Ile Leu His  Leu Met Lys
    1160                1165                1170

Leu Thr  Leu Gln Met Arg Asn  Ser Val Thr Gly Thr  Asp Ile Asp
    1175                1180                1185

Tyr Met  Val Ser Pro Val Ala  Asn Glu Cys Gly Glu  Phe Phe Asp
    1190                1195                1200

Ser Arg  Lys Val Lys Glu Gly  Leu Pro Val Asn Ala  Asp Ala Asn
    1205                1210                1215

Gly Ala  Tyr Asn Ile Ala Arg  Lys Gly Leu Trp Leu  Ala Gln Gln
    1220                1225                1230
```

Ile Lys Asn Ala Asn Asp Leu Ser Asp Val Lys Leu Ala Ile Thr
1235 1240 1245

Asn Lys Glu Trp Leu Gln Phe Ala Gln Lys Lys Gln Tyr Leu Lys
1250 1255 1260

Asp

<210> 88
<211> 1264
<212> PRT
<213> 山羊莫拉氏菌（Moraxella caprae）

<400> 88
Met Leu Phe Gln Asp Phe Thr His Leu Tyr Pro Leu Ser Lys Thr Met
1 5 10 15

Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Leu Glu His Ile His Ala
20 25 30

Lys Asn Phe Leu Ser Gln Asp Glu Thr Met Ala Asp Met Tyr Gln Lys
35 40 45

Val Lys Ala Ile Leu Asp Asp Tyr His Arg Asp Phe Ile Ala Asp Met
50 55 60

Met Gly Glu Val Lys Leu Thr Lys Leu Ala Glu Phe Tyr Asp Val Tyr
65 70 75 80

Leu Lys Phe Arg Lys Asn Pro Lys Asp Asp Gly Leu Gln Lys Gln Leu
85 90 95

Lys Asp Leu Gln Ala Val Leu Arg Lys Glu Ile Val Lys Pro Ile Gly
100 105 110

Asn Gly Gly Lys Tyr Lys Ala Gly Tyr Asp Arg Leu Phe Gly Ala Lys
115 120 125

Leu Phe Lys Asp Gly Lys Glu Leu Gly Asp Leu Ala Lys Phe Val Ile
130 135 140

Ala Gln Glu Gly Glu Ser Ser Pro Lys Leu Ala His Leu Ala His Phe
145 150 155 160

Glu Lys Phe Ser Thr Tyr Phe Thr Gly Phe His Asp Asn Arg Lys Asn
165 170 175

Met Tyr Ser Asp Glu Asp Lys His Thr Ala Ile Thr Tyr Arg Leu Ile
180 185 190

His Glu Asn Leu Pro Arg Phe Ile Asp Asn Leu Gln Ile Leu Ala Thr
195 200 205

Ile Lys Gln Lys His Ser Ala Leu Tyr Asp Gln Ile Ile Asn Glu Leu
210 215 220

Thr Ala Ser Gly Leu Asp Val Ser Leu Ala Ser His Leu Asp Gly Tyr
225 230 235 240

His Lys Leu Leu Thr Gln Glu Gly Ile Thr Ala Tyr Asn Thr Leu Leu
245 250 255

```
Gly Gly Ile Ser Gly Glu Ala Gly Ser Arg Lys Ile Gln Gly Ile Asn
            260                 265                 270

Glu Leu Ile Asn Ser His His Asn Gln His Cys His Lys Ser Glu Arg
        275                 280                 285

Ile Ala Lys Leu Arg Pro Leu His Lys Gln Ile Leu Ser Asp Gly Met
    290                 295                 300

Gly Val Ser Phe Leu Pro Ser Lys Phe Ala Asp Asp Ser Glu Met Cys
305                 310                 315                 320

Gln Ala Val Asn Glu Phe Tyr Arg His Tyr Ala Asp Val Phe Ala Lys
                325                 330                 335

Val Gln Ser Leu Phe Asp Gly Phe Asp Asp His Gln Lys Asp Gly Ile
            340                 345                 350

Tyr Val Glu His Lys Asn Leu Asn Glu Leu Ser Lys Gln Ala Phe Gly
        355                 360                 365

Asp Phe Ala Leu Leu Gly Arg Val Leu Asp Gly Tyr Tyr Val Asp Val
    370                 375                 380

Val Asn Pro Glu Phe Asn Glu Arg Phe Ala Lys Ala Lys Thr Asp Asn
385                 390                 395                 400

Ala Lys Ala Lys Leu Thr Lys Glu Lys Asp Lys Phe Ile Lys Gly Val
                405                 410                 415

His Ser Leu Ala Ser Leu Glu Gln Ala Ile Glu His Tyr Thr Ala Arg
            420                 425                 430

His Asp Asp Glu Ser Val Gln Ala Gly Lys Leu Gly Gln Tyr Phe Lys
        435                 440                 445

His Gly Leu Ala Gly Val Asp Asn Pro Ile Gln Lys Ile His Asn Asn
    450                 455                 460

His Ser Thr Ile Lys Gly Phe Leu Glu Arg Glu Arg Pro Ala Gly Glu
465                 470                 475                 480

Arg Ala Leu Pro Lys Ile Lys Ser Gly Lys Asn Pro Glu Met Thr Gln
                485                 490                 495

Leu Arg Gln Leu Lys Glu Leu Leu Asp Asn Ala Leu Asn Val Ala His
            500                 505                 510

Phe Ala Lys Leu Leu Thr Thr Lys Thr Thr Leu Asp Asn Gln Asp Gly
        515                 520                 525

Asn Phe Tyr Gly Glu Phe Gly Ala Leu Tyr Asp Glu Leu Ala Lys Ile
    530                 535                 540

Pro Thr Leu Tyr Asn Lys Val Arg Asp Tyr Leu Ser Gln Lys Pro Phe
545                 550                 555                 560

Ser Thr Glu Lys Tyr Lys Leu Asn Phe Gly Asn Pro Thr Leu Leu Asn
                565                 570                 575

Gly Trp Asp Leu Asn Lys Glu Lys Asp Asn Phe Gly Ile Ile Leu Gln
            580                 585                 590
```

```
Lys Asp Gly Cys Tyr Tyr Leu Ala Leu Leu Asp Lys Ala His Lys Lys
        595                 600                 605

Val Phe Asp Asn Ala Pro Asn Thr Gly Lys Asn Val Tyr Gln Lys Met
    610                 615                 620

Ile Tyr Lys Leu Leu Pro Gly Pro Asn Lys Met Leu Pro Lys Val Phe
625                 630                 635                 640

Phe Ala Lys Ser Asn Leu Asp Tyr Tyr Asn Pro Ser Ala Glu Leu Leu
                645                 650                 655

Asp Lys Tyr Ala Gln Gly Thr His Lys Lys Gly Asn Asn Phe Asn Leu
            660                 665                 670

Lys Asp Cys His Ala Leu Ile Asp Phe Phe Lys Ala Gly Ile Asn Lys
        675                 680                 685

His Pro Glu Trp Gln His Phe Gly Phe Lys Phe Ser Pro Thr Ser Ser
    690                 695                 700

Tyr Gln Asp Leu Ser Asp Phe Tyr Arg Glu Val Glu Pro Gln Gly Tyr
705                 710                 715                 720

Gln Val Lys Phe Val Asp Ile Asn Ala Asp Tyr Ile Asn Glu Leu Val
                725                 730                 735

Glu Gln Gly Gln Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser
            740                 745                 750

Pro Lys Ala His Gly Lys Pro Asn Leu His Thr Leu Tyr Phe Lys Ala
        755                 760                 765

Leu Phe Ser Lys Asp Asn Leu Ala Asn Pro Ile Tyr Lys Leu Asn Gly
    770                 775                 780

Glu Ala Gln Ile Phe Tyr Arg Lys Ala Ser Leu Asp Met Asn Glu Thr
785                 790                 795                 800

Thr Ile His Arg Ala Gly Glu Val Leu Glu Asn Lys Asn Pro Asp Asn
                805                 810                 815

Pro Lys Lys Arg Gln Phe Val Tyr Asp Ile Ile Lys Asp Lys Arg Tyr
            820                 825                 830

Thr Gln Asp Lys Phe Met Leu His Val Pro Ile Thr Met Asn Phe Gly
        835                 840                 845

Val Gln Gly Met Thr Ile Lys Glu Phe Asn Lys Lys Val Asn Gln Ser
    850                 855                 860

Ile Gln Gln Tyr Asp Glu Val Asn Val Ile Gly Ile Asp Arg Gly Glu
865                 870                 875                 880

Arg His Leu Leu Tyr Leu Thr Val Ile Asn Ser Lys Gly Glu Ile Leu
                885                 890                 895

Glu Gln Arg Ser Leu Asn Asp Ile Thr Thr Ala Ser Ala Asn Gly Thr
            900                 905                 910

Gln Met Thr Thr Pro Tyr His Lys Ile Leu Asp Lys Arg Glu Ile Glu
        915                 920                 925
```

```
Arg Leu Asn Ala Arg Val Gly Trp Gly Glu Ile Glu Thr Ile Lys Glu
    930                 935                 940

Leu Lys Ser Gly Tyr Leu Ser His Val Val His Gln Ile Ser Gln Leu
945                 950                 955                 960

Met Leu Lys Tyr Asn Ala Ile Val Val Leu Glu Asp Leu Asn Phe Gly
                965                 970                 975

Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Ile Tyr Gln Asn Phe
            980                 985                 990

Glu Asn Ala Leu Ile Lys Lys Leu Asn His Leu Val Leu Lys Asp Glu
        995                 1000                1005

Ala Asp  Asp Glu Ile Gly Ser  Tyr Lys Asn Ala Leu  Gln Leu Thr
    1010                1015                1020

Asn Asn  Phe Thr Asp Leu Lys  Ser Ile Gly Lys Gln  Thr Gly Phe
    1025                1030                1035

Leu Phe  Tyr Val Pro Ala Trp  Asn Thr Ser Lys Ile  Asp Pro Glu
    1040                1045                1050

Thr Gly  Phe Val Asp Leu Leu  Lys Pro Arg Tyr Glu  Asn Ile Ala
    1055                1060                1065

Gln Ser  Gln Ala Phe Phe Gly  Lys Phe Asp Lys Ile  Cys Tyr Asn
    1070                1075                1080

Ala Asp  Lys Asp Tyr Phe Glu  Phe His Ile Asp Tyr  Ala Lys Phe
    1085                1090                1095

Thr Asp  Lys Ala Lys Asn Ser  Arg Gln Ile Trp Lys  Ile Cys Ser
    1100                1105                1110

His Gly  Asp Lys Arg Tyr Val  Tyr Asp Lys Thr Ala  Asn Gln Asn
    1115                1120                1125

Lys Gly  Ala Thr Lys Gly Ile  Asn Val Asn Asp Glu  Leu Lys Ser
    1130                1135                1140

Leu Phe  Ala Arg His His Ile  Asn Asp Lys Gln Pro  Asn Leu Val
    1145                1150                1155

Met Asp  Ile Cys Gln Asn Asn  Asp Lys Glu Phe His  Lys Ser Leu
    1160                1165                1170

Ile Tyr  Leu Leu Lys Thr Leu  Leu Ala Leu Arg Tyr  Ser Asn Ala
    1175                1180                1185

Ser Ser  Asp Glu Asp Phe Ile  Leu Ser Pro Val Ala  Asn Asp Glu
    1190                1195                1200

Gly Met  Phe Phe Asn Ser Ala  Leu Ala Asp Asp Thr  Gln Pro Gln
    1205                1210                1215

Asn Ala  Asp Ala Asn Gly Ala  Tyr His Ile Ala Leu  Lys Gly Leu
    1220                1225                1230

Trp Val  Leu Glu Gln Ile Lys  Asn Ser Asp Asp Leu  Asn Lys Val
    1235                1240                1245
```

Lys Leu Ala Ile Asp Asn Gln Thr Trp Leu Asn Phe Ala Gln Asn
1250 1255 1260

Arg

<210> 89
<211> 1263
<212> PRT
<213> 稻田鉤端螺旋體（Leptospira inadai）

<400> 89
Met Glu Asp Tyr Ser Gly Phe Val Asn Ile Tyr Ser Ile Gln Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr Leu Glu His Ile Glu
20 25 30

Lys Lys Gly Phe Leu Lys Lys Asp Lys Ile Arg Ala Glu Asp Tyr Lys
35 40 45

Ala Val Lys Lys Ile Ile Asp Lys Tyr His Arg Ala Tyr Ile Glu Glu
50 55 60

Val Phe Asp Ser Val Leu His Gln Lys Lys Lys Lys Asp Lys Thr Arg
65 70 75 80

Phe Ser Thr Gln Phe Ile Lys Glu Ile Lys Glu Phe Ser Glu Leu Tyr
85 90 95

Tyr Lys Thr Glu Lys Asn Ile Pro Asp Lys Glu Arg Leu Glu Ala Leu
100 105 110

Ser Glu Lys Leu Arg Lys Met Leu Val Gly Ala Phe Lys Gly Glu Phe
115 120 125

Ser Glu Glu Val Ala Glu Lys Tyr Lys Asn Leu Phe Ser Lys Glu Leu
130 135 140

Ile Arg Asn Glu Ile Glu Lys Phe Cys Glu Thr Asp Glu Glu Arg Lys
145 150 155 160

Gln Val Ser Asn Phe Lys Ser Phe Thr Thr Tyr Phe Thr Gly Phe His
165 170 175

Ser Asn Arg Gln Asn Ile Tyr Ser Asp Glu Lys Lys Ser Thr Ala Ile
180 185 190

Gly Tyr Arg Ile Ile His Gln Asn Leu Pro Lys Phe Leu Asp Asn Leu
195 200 205

Lys Ile Ile Glu Ser Ile Gln Arg Arg Phe Lys Asp Phe Pro Trp Ser
210 215 220

Asp Leu Lys Lys Asn Leu Lys Lys Ile Asp Lys Asn Ile Lys Leu Thr
225 230 235 240

Glu Tyr Phe Ser Ile Asp Gly Phe Val Asn Val Leu Asn Gln Lys Gly
245 250 255

Ile Asp Ala Tyr Asn Thr Ile Leu Gly Gly Lys Ser Glu Glu Ser Gly
260 265 270

Glu Lys Ile Gln Gly Leu Asn Glu Tyr Ile Asn Leu Tyr Arg Gln Lys
275 280 285

```
Asn Asn Ile Asp Arg Lys Asn Leu Pro Asn Val Lys Ile Leu Phe Lys
    290                 295                 300

Gln Ile Leu Gly Asp Arg Glu Thr Lys Ser Phe Ile Pro Glu Ala Phe
305                 310                 315                 320

Pro Asp Asp Gln Ser Val Leu Asn Ser Ile Thr Glu Phe Ala Lys Tyr
                325                 330                 335

Leu Lys Leu Asp Lys Lys Lys Lys Ser Ile Ile Ala Glu Leu Lys Lys
            340                 345                 350

Phe Leu Ser Ser Phe Asn Arg Tyr Glu Leu Asp Gly Ile Tyr Leu Ala
        355                 360                 365

Asn Asp Asn Ser Leu Ala Ser Ile Ser Thr Phe Leu Phe Asp Asp Trp
    370                 375                 380

Ser Phe Ile Lys Lys Ser Val Ser Phe Lys Tyr Asp Glu Ser Val Gly
385                 390                 395                 400

Asp Pro Lys Lys Lys Ile Lys Ser Pro Leu Lys Tyr Glu Lys Glu Lys
                405                 410                 415

Glu Lys Trp Leu Lys Gln Lys Tyr Tyr Thr Ile Ser Phe Leu Asn Asp
            420                 425                 430

Ala Ile Glu Ser Tyr Ser Lys Ser Gln Asp Glu Lys Arg Val Lys Ile
        435                 440                 445

Arg Leu Glu Ala Tyr Phe Ala Glu Phe Lys Ser Lys Asp Asp Ala Lys
    450                 455                 460

Lys Gln Phe Asp Leu Leu Glu Arg Ile Glu Glu Ala Tyr Ala Ile Val
465                 470                 475                 480

Glu Pro Leu Leu Gly Ala Glu Tyr Pro Arg Asp Arg Asn Leu Lys Ala
                485                 490                 495

Asp Lys Lys Glu Val Gly Lys Ile Lys Asp Phe Leu Asp Ser Ile Lys
            500                 505                 510

Ser Leu Gln Phe Phe Leu Lys Pro Leu Leu Ser Ala Glu Ile Phe Asp
        515                 520                 525

Glu Lys Asp Leu Gly Phe Tyr Asn Gln Leu Glu Gly Tyr Tyr Glu Glu
    530                 535                 540

Ile Asp Ser Ile Gly His Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr
545                 550                 555                 560

Gly Lys Ile Tyr Ser Lys Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser
                565                 570                 575

Thr Leu Leu Lys Gly Trp Asp Glu Asn Arg Glu Val Ala Asn Leu Cys
            580                 585                 590

Val Ile Phe Arg Glu Asp Gln Lys Tyr Tyr Leu Gly Val Met Asp Lys
        595                 600                 605

Glu Asn Asn Thr Ile Leu Ser Asp Ile Pro Lys Val Lys Pro Asn Glu
    610                 615                 620
```

```
Leu Phe Tyr Glu Lys Met Val Tyr Lys Leu Ile Pro Thr Pro His Met
625                 630                 635                 640

Gln Leu Pro Arg Ile Ile Phe Ser Ser Asp Asn Leu Ser Ile Tyr Asn
                645                 650                 655

Pro Ser Lys Ser Ile Leu Lys Ile Arg Glu Ala Lys Ser Phe Lys Glu
            660                 665                 670

Gly Lys Asn Phe Lys Leu Lys Asp Cys His Lys Phe Ile Asp Phe Tyr
        675                 680                 685

Lys Glu Ser Ile Ser Lys Asn Glu Asp Trp Ser Arg Phe Asp Phe Lys
    690                 695                 700

Phe Ser Lys Thr Ser Ser Tyr Glu Asn Ile Ser Glu Phe Tyr Arg Glu
705                 710                 715                 720

Val Glu Arg Gln Gly Tyr Asn Leu Asp Phe Lys Lys Val Ser Lys Phe
                725                 730                 735

Tyr Ile Asp Ser Leu Val Glu Asp Gly Lys Leu Tyr Leu Phe Gln Ile
            740                 745                 750

Tyr Asn Lys Asp Phe Ser Ile Phe Ser Lys Gly Lys Pro Asn Leu His
        755                 760                 765

Thr Ile Tyr Phe Arg Ser Leu Phe Ser Lys Glu Asn Leu Lys Asp Val
    770                 775                 780

Cys Leu Lys Leu Asn Gly Glu Ala Glu Met Phe Phe Arg Lys Lys Ser
785                 790                 795                 800

Ile Asn Tyr Asp Glu Lys Lys Lys Arg Glu Gly His His Pro Glu Leu
                805                 810                 815

Phe Glu Lys Leu Lys Tyr Pro Ile Leu Lys Asp Lys Arg Tyr Ser Glu
            820                 825                 830

Asp Lys Phe Gln Phe His Leu Pro Ile Ser Leu Asn Phe Lys Ser Lys
        835                 840                 845

Glu Arg Leu Asn Phe Asn Leu Lys Val Asn Glu Phe Leu Lys Arg Asn
    850                 855                 860

Lys Asp Ile Asn Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu Leu
865                 870                 875                 880

Tyr Leu Val Met Ile Asn Gln Lys Gly Glu Ile Leu Lys Gln Thr Leu
                885                 890                 895

Leu Asp Ser Met Gln Ser Gly Lys Gly Arg Pro Glu Ile Asn Tyr Lys
            900                 905                 910

Glu Lys Leu Gln Glu Lys Glu Ile Glu Arg Asp Lys Ala Arg Lys Ser
        915                 920                 925

Trp Gly Thr Val Glu Asn Ile Lys Glu Leu Lys Glu Gly Tyr Leu Ser
    930                 935                 940

Ile Val Ile His Gln Ile Ser Lys Leu Met Val Glu Asn Asn Ala Ile
945                 950                 955                 960
```

```
Val Val Leu Glu Asp Leu Asn Ile Gly Phe Lys Arg Gly Arg Gln Lys
                965                 970                 975

Val Glu Arg Gln Val Tyr Gln Lys Phe Glu Lys Met Leu Ile Asp Lys
            980                 985                 990

Leu Asn Phe Leu Val Phe Lys Glu Asn Lys Pro Thr Glu Pro Gly Gly
        995                1000                1005

Val Leu  Lys Ala Tyr Gln Leu  Thr Asp Glu Phe Gln  Ser Phe Glu
    1010                1015                1020

Lys Leu  Ser Lys Gln Thr Gly  Phe Leu Phe Tyr Val  Pro Ser Trp
    1025                1030                1035

Asn Thr  Ser Lys Ile Asp Pro  Arg Thr Gly Phe Ile  Asp Phe Leu
    1040                1045                1050

His Pro  Ala Tyr Glu Asn Ile  Glu Lys Ala Lys Gln  Trp Ile Asn
    1055                1060                1065

Lys Phe  Asp Ser Ile Arg Phe  Asn Ser Lys Met Asp  Trp Phe Glu
    1070                1075                1080

Phe Thr  Ala Asp Thr Arg Lys  Phe Ser Glu Asn Leu  Met Leu Gly
    1085                1090                1095

Lys Asn  Arg Val Trp Val Ile  Cys Thr Thr Asn Val  Glu Arg Tyr
    1100                1105                1110

Phe Thr  Ser Lys Thr Ala Asn  Ser Ser Ile Gln Tyr  Asn Ser Ile
    1115                1120                1125

Gln Ile  Thr Glu Lys Leu Lys  Glu Leu Phe Val Asp  Ile Pro Phe
    1130                1135                1140

Ser Asn  Gly Gln Asp Leu Lys  Pro Glu Ile Leu Arg  Lys Asn Asp
    1145                1150                1155

Ala Val  Phe Phe Lys Ser Leu  Leu Phe Tyr Ile Lys  Thr Thr Leu
    1160                1165                1170

Ser Leu  Arg Gln Asn Asn Gly  Lys Lys Gly Glu Glu  Glu Lys Asp
    1175                1180                1185

Phe Ile  Leu Ser Pro Val Val  Asp Ser Lys Gly Arg  Phe Phe Asn
    1190                1195                1200

Ser Leu  Glu Ala Ser Asp Asp  Glu Pro Lys Asp Ala  Asp Ala Asn
    1205                1210                1215

Gly Ala  Tyr His Ile Ala Leu  Lys Gly Leu Met Asn  Leu Leu Val
    1220                1225                1230

Leu Asn  Glu Thr Lys Glu Glu  Asn Leu Ser Arg Pro  Lys Trp Lys
    1235                1240                1245

Ile Lys  Asn Lys Asp Trp Leu  Glu Phe Val Trp Glu  Arg Asn Arg
    1250                1255                1260
```

<210> 90
<211> 1262

<212> PRT
<213> 狗口腔卟啉單胞菌（Porphyromonas crevioricanis）

<400> 90

```
Met Pro Trp Ile Asp Leu Lys Asp Phe Thr Asn Leu Tyr Pro Val Ser
1               5                  10                  15

Lys Thr Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr Leu Glu Asn
            20                  25                  30

Ile Glu Lys Ala Gly Ile Leu Lys Glu Asp Glu His Arg Ala Glu Ser
        35                  40                  45

Tyr Arg Arg Val Lys Lys Ile Ile Asp Thr Tyr His Lys Val Phe Ile
    50                  55                  60

Asp Ser Ser Leu Glu Asn Met Ala Lys Met Gly Ile Glu Asn Glu Ile
65                  70                  75                  80

Lys Ala Met Leu Gln Ser Phe Cys Glu Leu Tyr Lys Lys Asp His Arg
                85                  90                  95

Thr Glu Gly Glu Asp Lys Ala Leu Asp Lys Ile Arg Ala Val Leu Arg
            100                 105                 110

Gly Leu Ile Val Gly Ala Phe Thr Gly Val Cys Gly Arg Arg Glu Asn
        115                 120                 125

Thr Val Gln Asn Glu Lys Tyr Glu Ser Leu Phe Lys Glu Lys Leu Ile
    130                 135                 140

Lys Glu Ile Leu Pro Asp Phe Val Leu Ser Thr Glu Ala Glu Ser Leu
145                 150                 155                 160

Pro Phe Ser Val Glu Glu Ala Thr Arg Ser Leu Lys Glu Phe Asp Ser
                165                 170                 175

Phe Thr Ser Tyr Phe Ala Gly Phe Tyr Glu Asn Arg Lys Asn Ile Tyr
            180                 185                 190

Ser Thr Lys Pro Gln Ser Thr Ala Ile Ala Tyr Arg Leu Ile His Glu
        195                 200                 205

Asn Leu Pro Lys Phe Ile Asp Asn Ile Leu Val Phe Gln Lys Ile Lys
    210                 215                 220

Glu Pro Ile Ala Lys Glu Leu Glu His Ile Arg Ala Asp Phe Ser Ala
225                 230                 235                 240

Gly Gly Tyr Ile Lys Lys Asp Glu Arg Leu Glu Asp Ile Phe Ser Leu
                245                 250                 255

Asn Tyr Tyr Ile His Val Leu Ser Gln Ala Gly Ile Glu Lys Tyr Asn
            260                 265                 270

Ala Leu Ile Gly Lys Ile Val Thr Glu Gly Asp Gly Glu Met Lys Gly
        275                 280                 285

Leu Asn Glu His Ile Asn Leu Tyr Asn Gln Gln Arg Gly Arg Glu Asp
    290                 295                 300

Arg Leu Pro Leu Phe Arg Pro Leu Tyr Lys Gln Ile Leu Ser Asp Arg
305                 310                 315                 320
```

```
Glu Gln Leu Ser Tyr Leu Pro Glu Ser Phe Glu Lys Asp Glu Glu Leu
                325                 330                 335

Leu Arg Ala Leu Lys Glu Phe Tyr Asp His Ile Ala Glu Asp Ile Leu
            340                 345                 350

Gly Arg Thr Gln Gln Leu Met Thr Ser Ile Ser Glu Tyr Asp Leu Ser
        355                 360                 365

Arg Ile Tyr Val Arg Asn Asp Ser Gln Leu Thr Asp Ile Ser Lys Lys
    370                 375                 380

Met Leu Gly Asp Trp Asn Ala Ile Tyr Met Ala Arg Glu Arg Ala Tyr
385                 390                 395                 400

Asp His Glu Gln Ala Pro Lys Arg Ile Thr Ala Lys Tyr Glu Arg Asp
                405                 410                 415

Arg Ile Lys Ala Leu Lys Gly Glu Glu Ser Ile Ser Leu Ala Asn Leu
            420                 425                 430

Asn Ser Cys Ile Ala Phe Leu Asp Asn Val Arg Asp Cys Arg Val Asp
        435                 440                 445

Thr Tyr Leu Ser Thr Leu Gly Gln Lys Glu Gly Pro His Gly Leu Ser
    450                 455                 460

Asn Leu Val Glu Asn Val Phe Ala Ser Tyr His Glu Ala Glu Gln Leu
465                 470                 475                 480

Leu Ser Phe Pro Tyr Pro Glu Glu Asn Asn Leu Ile Gln Asp Lys Asp
                485                 490                 495

Asn Val Val Leu Ile Lys Asn Leu Leu Asp Asn Ile Ser Asp Leu Gln
            500                 505                 510

Arg Phe Leu Lys Pro Leu Trp Gly Met Gly Asp Glu Pro Asp Lys Asp
        515                 520                 525

Glu Arg Phe Tyr Gly Glu Tyr Asn Tyr Ile Arg Gly Ala Leu Asp Gln
    530                 535                 540

Val Ile Pro Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Arg Lys Pro
545                 550                 555                 560

Tyr Ser Thr Arg Lys Val Lys Leu Asn Phe Gly Asn Ser Gln Leu Leu
                565                 570                 575

Ser Gly Trp Asp Arg Asn Lys Glu Lys Asp Asn Ser Cys Val Ile Leu
            580                 585                 590

Arg Lys Gly Gln Asn Phe Tyr Leu Ala Ile Met Asn Asn Arg His Lys
        595                 600                 605

Arg Ser Phe Glu Asn Lys Val Leu Pro Glu Tyr Lys Glu Gly Glu Pro
    610                 615                 620

Tyr Phe Glu Lys Met Asp Tyr Lys Phe Leu Pro Asp Pro Asn Lys Met
625                 630                 635                 640

Leu Pro Lys Val Phe Leu Ser Lys Lys Gly Ile Glu Ile Tyr Glu Pro
                645                 650                 655
```

Ser Pro Lys Leu Leu Glu Gln Tyr Gly His Gly Thr His Lys Lys Gly
660 665 670

Asp Thr Phe Ser Met Asp Asp Leu His Glu Leu Ile Asp Phe Phe Lys
675 680 685

His Ser Ile Glu Ala His Glu Asp Trp Lys Gln Phe Gly Phe Lys Phe
690 695 700

Ser Asp Thr Ala Thr Tyr Glu Asn Val Ser Ser Phe Tyr Arg Glu Val
705 710 715 720

Glu Asp Gln Gly Tyr Lys Leu Ser Phe Arg Lys Val Ser Glu Ser Tyr
725 730 735

Val Tyr Ser Leu Ile Asp Gln Gly Lys Leu Tyr Leu Phe Gln Ile Tyr
740 745 750

Asn Lys Asp Phe Ser Pro Cys Ser Lys Gly Thr Pro Asn Leu His Thr
755 760 765

Leu Tyr Trp Arg Met Leu Phe Asp Glu Arg Asn Leu Ala Asp Val Ile
770 775 780

Tyr Lys Leu Asp Gly Lys Ala Glu Ile Phe Phe Arg Glu Lys Ser Leu
785 790 795 800

Lys Asn Asp His Pro Thr His Pro Ala Gly Lys Pro Ile Lys Lys Lys
805 810 815

Ser Arg Gln Lys Lys Gly Glu Glu Ser Leu Phe Glu Tyr Asp Leu Val
820 825 830

Lys Asp Arg Arg Tyr Thr Met Asp Lys Phe Gln Phe His Val Pro Ile
835 840 845

Thr Met Asn Phe Lys Cys Ser Ala Gly Ser Lys Val Asn Asp Met Val
850 855 860

Asn Ala His Ile Arg Glu Ala Lys Asp Met His Val Ile Gly Ile Asp
865 870 875 880

Arg Gly Glu Arg Asn Leu Leu Tyr Ile Cys Val Ile Asp Ser Arg Gly
885 890 895

Thr Ile Leu Asp Gln Ile Ser Leu Asn Thr Ile Asn Asp Ile Asp Tyr
900 905 910

His Asp Leu Leu Glu Ser Arg Asp Lys Asp Arg Gln Gln Glu Arg Arg
915 920 925

Asn Trp Gln Thr Ile Glu Gly Ile Lys Glu Leu Lys Gln Gly Tyr Leu
930 935 940

Ser Gln Ala Val His Arg Ile Ala Glu Leu Met Val Ala Tyr Lys Ala
945 950 955 960

Val Val Ala Leu Glu Asp Leu Asn Met Gly Phe Lys Arg Gly Arg Gln
965 970 975

Lys Val Glu Ser Ser Val Tyr Gln Gln Phe Glu Lys Gln Leu Ile Asp
980 985 990

Lys Leu Asn Tyr Leu Val Asp Lys Lys Lys Arg Pro Glu Asp Ile Gly
995 1000 1005

Gly Leu Leu Arg Ala Tyr Gln Phe Thr Ala Pro Phe Lys Ser Phe
1010 1015 1020

Lys Glu Met Gly Lys Gln Asn Gly Phe Leu Phe Tyr Ile Pro Ala
1025 1030 1035

Trp Asn Thr Ser Asn Ile Asp Pro Thr Thr Gly Phe Val Asn Leu
1040 1045 1050

Phe His Ala Gln Tyr Glu Asn Val Asp Lys Ala Lys Ser Phe Phe
1055 1060 1065

Gln Lys Phe Asp Ser Ile Ser Tyr Asn Pro Lys Lys Asp Trp Phe
1070 1075 1080

Glu Phe Ala Phe Asp Tyr Lys Asn Phe Thr Lys Lys Ala Glu Gly
1085 1090 1095

Ser Arg Ser Met Trp Ile Leu Cys Thr His Gly Ser Arg Ile Lys
1100 1105 1110

Asn Phe Arg Asn Ser Gln Lys Asn Gly Gln Trp Asp Ser Glu Glu
1115 1120 1125

Phe Ala Leu Thr Glu Ala Phe Lys Ser Leu Phe Val Arg Tyr Glu
1130 1135 1140

Ile Asp Tyr Thr Ala Asp Leu Lys Thr Ala Ile Val Asp Glu Lys
1145 1150 1155

Gln Lys Asp Phe Phe Val Asp Leu Leu Lys Leu Phe Lys Leu Thr
1160 1165 1170

Val Gln Met Arg Asn Ser Trp Lys Glu Lys Asp Leu Asp Tyr Leu
1175 1180 1185

Ile Ser Pro Val Ala Gly Ala Asp Gly Arg Phe Phe Asp Thr Arg
1190 1195 1200

Glu Gly Asn Lys Ser Leu Pro Lys Asp Ala Asp Ala Asn Gly Ala
1205 1210 1215

Tyr Asn Ile Ala Leu Lys Gly Leu Trp Ala Leu Arg Gln Ile Arg
1220 1225 1230

Gln Thr Ser Glu Gly Gly Lys Leu Lys Leu Ala Ile Ser Asn Lys
1235 1240 1245

Glu Trp Leu Gln Phe Val Gln Glu Arg Ser Tyr Glu Lys Asp
1250 1255 1260

<210> 91
<211> 1262
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
口腔類擬桿菌（Bacteroidetes oral taxon）274序列

<400> 91
Met Arg Lys Phe Asn Glu Phe Val Gly Leu Tyr Pro Ile Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Leu Glu His Ile Gln
20 25 30

Arg Asn Lys Leu Leu Glu His Asp Ala Val Arg Ala Asp Asp Tyr Val
35 40 45

Lys Val Lys Lys Ile Ile Asp Lys Tyr His Lys Cys Leu Ile Asp Glu
50 55 60

Ala Leu Ser Gly Phe Thr Phe Asp Thr Glu Ala Asp Gly Arg Ser Asn
65 70 75 80

Asn Ser Leu Ser Glu Tyr Tyr Leu Tyr Tyr Asn Leu Lys Lys Arg Asn
85 90 95

Glu Gln Glu Gln Lys Thr Phe Lys Thr Ile Gln Asn Asn Leu Arg Lys
100 105 110

Gln Ile Val Asn Lys Leu Thr Gln Ser Glu Lys Tyr Lys Arg Ile Asp
115 120 125

Lys Lys Glu Leu Ile Thr Thr Asp Leu Pro Asp Phe Leu Thr Asn Glu
130 135 140

Ser Glu Lys Glu Leu Val Glu Lys Phe Lys Asn Phe Thr Thr Tyr Phe
145 150 155 160

Thr Glu Phe His Lys Asn Arg Lys Asn Met Tyr Ser Lys Glu Glu Lys
165 170 175

Ser Thr Ala Ile Ala Phe Arg Leu Ile Asn Glu Asn Leu Pro Lys Phe
180 185 190

Val Asp Asn Ile Ala Ala Phe Glu Lys Val Val Ser Ser Pro Leu Ala
195 200 205

Glu Lys Ile Asn Ala Leu Tyr Glu Asp Phe Lys Glu Tyr Leu Asn Val
210 215 220

Glu Glu Ile Ser Arg Val Phe Arg Leu Asp Tyr Tyr Asp Glu Leu Leu
225 230 235 240

Thr Gln Lys Gln Ile Asp Leu Tyr Asn Ala Ile Val Gly Gly Arg Thr
245 250 255

Glu Glu Asp Asn Lys Ile Gln Ile Lys Gly Leu Asn Gln Tyr Ile Asn
260 265 270

Glu Tyr Asn Gln Gln Gln Thr Asp Arg Ser Asn Arg Leu Pro Lys Leu
275 280 285

Lys Pro Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Ser Val Ser Trp
290 295 300

Leu Pro Pro Lys Phe Asp Ser Asp Lys Asn Leu Leu Ile Lys Ile Lys
305 310 315 320

Glu Cys Tyr Asp Ala Leu Ser Glu Lys Glu Lys Val Phe Asp Lys Leu
325 330 335

Glu Ser Ile Leu Lys Ser Leu Ser Thr Tyr Asp Leu Ser Lys Ile Tyr
340 345 350

Ile Ser Asn Asp Ser Gln Leu Ser Tyr Ile Ser Gln Lys Met Phe Gly
355 360 365

Arg Trp Asp Ile Ile Ser Lys Ala Ile Arg Glu Asp Cys Ala Lys Arg
370 375 380

Asn Pro Gln Lys Ser Arg Glu Ser Leu Glu Lys Phe Ala Glu Arg Ile
385 390 395 400

Asp Lys Lys Leu Lys Thr Ile Asp Ser Ile Ser Ile Gly Asp Val Asp
405 410 415

Glu Cys Leu Ala Gln Leu Gly Glu Thr Tyr Val Lys Arg Val Glu Asp
420 425 430

Tyr Phe Val Ala Met Gly Glu Ser Glu Ile Asp Asp Glu Gln Thr Asp
435 440 445

Thr Thr Ser Phe Lys Lys Asn Ile Glu Gly Ala Tyr Glu Ser Val Lys
450 455 460

Glu Leu Leu Asn Asn Ala Asp Asn Ile Thr Asp Asn Asn Leu Met Gln
465 470 475 480

Asp Lys Gly Asn Val Glu Lys Ile Lys Thr Leu Leu Asp Ala Ile Lys
485 490 495

Asp Leu Gln Arg Phe Ile Lys Pro Leu Leu Gly Lys Gly Asp Glu Ala
500 505 510

Asp Lys Asp Gly Val Phe Tyr Gly Glu Phe Thr Ser Leu Trp Thr Lys
515 520 525

Leu Asp Gln Val Thr Pro Leu Tyr Asn Met Val Arg Asn Tyr Leu Thr
530 535 540

Ser Lys Pro Tyr Ser Thr Lys Lys Ile Lys Leu Asn Phe Glu Asn Ser
545 550 555 560

Thr Leu Met Asp Gly Trp Asp Leu Asn Lys Glu Pro Asp Asn Thr Thr
565 570 575

Val Ile Phe Cys Lys Asp Gly Leu Tyr Tyr Leu Gly Ile Met Gly Lys
580 585 590

Lys Tyr Asn Arg Val Phe Val Asp Arg Glu Asp Leu Pro His Asp Gly
595 600 605

Glu Cys Tyr Asp Lys Met Glu Tyr Lys Leu Leu Pro Gly Ala Asn Lys
610 615 620

Met Leu Pro Lys Val Phe Phe Ser Glu Thr Gly Ile Gln Arg Phe Leu
625 630 635 640

Pro Ser Glu Glu Leu Leu Gly Lys Tyr Glu Arg Gly Thr His Lys Lys
645 650 655

Gly Ala Gly Phe Asp Leu Gly Asp Cys Arg Ala Leu Ile Asp Phe Phe
660 665 670

Lys Lys Ser Ile Glu Arg His Asp Asp Trp Lys Lys Phe Asp Phe Lys
675 680 685

```
Phe Ser Asp Thr Ser Thr Tyr Gln Asp Ile Ser Glu Phe Tyr Arg Glu
    690                 695                 700

Val Glu Gln Gln Gly Tyr Lys Met Ser Phe Arg Lys Val Ser Val Asp
705                 710                 715                 720

Tyr Ile Lys Ser Leu Val Glu Glu Gly Lys Leu Tyr Leu Phe Gln Ile
                725                 730                 735

Tyr Asn Lys Asp Phe Ser Ala His Ser Lys Gly Thr Pro Asn Met His
            740                 745                 750

Thr Leu Tyr Trp Lys Met Leu Phe Asp Glu Glu Asn Leu Lys Asp Val
        755                 760                 765

Val Tyr Lys Leu Asn Gly Glu Ala Glu Val Phe Phe Arg Lys Ser Ser
    770                 775                 780

Ile Thr Val Gln Ser Pro Thr His Pro Ala Asn Ser Pro Ile Lys Asn
785                 790                 795                 800

Lys Asn Lys Asp Asn Gln Lys Lys Glu Ser Lys Phe Glu Tyr Asp Leu
                805                 810                 815

Ile Lys Asp Arg Arg Tyr Thr Val Asp Lys Phe Leu Phe His Val Pro
            820                 825                 830

Ile Thr Met Asn Phe Lys Ser Val Gly Gly Ser Asn Ile Asn Gln Leu
        835                 840                 845

Val Lys Arg His Ile Arg Ser Ala Thr Asp Leu His Ile Ile Gly Ile
    850                 855                 860

Asp Arg Gly Glu Arg His Leu Leu Tyr Leu Thr Val Ile Asp Ser Arg
865                 870                 875                 880

Gly Asn Ile Lys Glu Gln Phe Ser Leu Asn Glu Ile Val Asn Glu Tyr
                885                 890                 895

Asn Gly Asn Thr Tyr Arg Thr Asp Tyr His Glu Leu Leu Asp Thr Arg
            900                 905                 910

Glu Gly Glu Arg Thr Glu Ala Arg Arg Asn Trp Gln Thr Ile Gln Asn
        915                 920                 925

Ile Arg Glu Leu Lys Glu Gly Tyr Leu Ser Gln Val Ile His Lys Ile
    930                 935                 940

Ser Glu Leu Ala Ile Lys Tyr Asn Ala Val Ile Val Leu Glu Asp Leu
945                 950                 955                 960

Asn Phe Gly Phe Met Arg Ser Arg Gln Lys Val Glu Lys Gln Val Tyr
                965                 970                 975

Gln Lys Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr Leu Val Asp
            980                 985                 990

Lys Lys Lys Pro Val Ala Glu Thr Gly Gly Leu Leu Arg Ala Tyr Gln
        995                 1000                1005

Leu Thr Gly Glu Phe Glu Ser Phe Lys Thr Leu Gly Lys Gln Ser
    1010                1015                1020
```

```
Gly Ile  Leu Phe Tyr Val Pro  Ala Trp Asn Thr Ser  Lys Ile Asp
    1025                1030                1035

Pro Val  Thr Gly Phe Val Asn  Leu Phe Asp Thr His  Tyr Glu Asn
    1040                1045                1050

Ile Glu  Lys Ala Lys Val Phe  Phe Asp Lys Phe Lys  Ser Ile Arg
    1055                1060                1065

Tyr Asn  Ser Asp Lys Asp Trp  Phe Glu Phe Val Val  Asp Asp Tyr
    1070                1075                1080

Thr Arg  Phe Ser Pro Lys Ala  Glu Gly Thr Arg Arg  Asp Trp Thr
    1085                1090                1095

Ile Cys  Thr Gln Gly Lys Arg  Ile Gln Ile Cys Arg  Asn His Gln
    1100                1105                1110

Arg Asn  Asn Glu Trp Glu Gly  Gln Glu Ile Asp Leu  Thr Lys Ala
    1115                1120                1125

Phe Lys  Glu His Phe Glu Ala  Tyr Gly Val Asp Ile  Ser Lys Asp
    1130                1135                1140

Leu Arg  Glu Gln Ile Asn Thr  Gln Asn Lys Lys Glu  Phe Phe Glu
    1145                1150                1155

Glu Leu  Leu Arg Leu Leu Arg  Leu Thr Leu Gln Met  Arg Asn Ser
    1160                1165                1170

Met Pro  Ser Ser Asp Ile Asp  Tyr Leu Ile Ser Pro  Val Ala Asn
    1175                1180                1185

Asp Thr  Gly Cys Phe Phe Asp  Ser Arg Lys Gln Ala  Glu Leu Lys
    1190                1195                1200

Glu Asn  Ala Val Leu Pro Met  Asn Ala Asp Ala Asn  Gly Ala Tyr
    1205                1210                1215

Asn Ile  Ala Arg Lys Gly Leu  Leu Ala Ile Arg Lys  Met Lys Gln
    1220                1225                1230

Glu Glu  Asn Asp Ser Ala Lys  Ile Ser Leu Ala Ile  Ser Asn Lys
    1235                1240                1245

Glu Trp  Leu Lys Phe Ala Gln  Thr Lys Pro Tyr Leu  Glu Asp
    1250                1255                1260
```

<210> 92
<211> 1260
<212> PRT
<213> 狗口腔卟啉單胞菌

<400> 92

```
Met Asp Ser Leu Lys Asp Phe Thr Asn Leu Tyr Pro Val Ser Lys Thr
1               5                  10                  15

Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr Leu Glu Asn Ile Glu
            20                  25                  30

Lys Ala Gly Ile Leu Lys Glu Asp Glu His Arg Ala Glu Ser Tyr Arg
        35                  40                  45
```

```
Arg Val Lys Lys Ile Ile Asp Thr Tyr His Lys Val Phe Ile Asp Ser
    50                  55                  60

Ser Leu Glu Asn Met Ala Lys Met Gly Ile Glu Asn Glu Ile Lys Ala
65                  70                  75                  80

Met Leu Gln Ser Phe Cys Glu Leu Tyr Lys Lys Asp His Arg Thr Glu
                85                  90                  95

Gly Glu Asp Lys Ala Leu Asp Lys Ile Arg Ala Val Leu Arg Gly Leu
            100                 105                 110

Ile Val Gly Ala Phe Thr Gly Val Cys Gly Arg Arg Glu Asn Thr Val
        115                 120                 125

Gln Asn Glu Lys Tyr Glu Ser Leu Phe Lys Glu Lys Leu Ile Lys Glu
    130                 135                 140

Ile Leu Pro Asp Phe Val Leu Ser Thr Glu Ala Glu Ser Leu Pro Phe
145                 150                 155                 160

Ser Val Glu Glu Ala Thr Arg Ser Leu Lys Glu Phe Asp Ser Phe Thr
                165                 170                 175

Ser Tyr Phe Ala Gly Phe Tyr Glu Asn Arg Lys Asn Ile Tyr Ser Thr
            180                 185                 190

Lys Pro Gln Ser Thr Ala Ile Ala Tyr Arg Leu Ile His Glu Asn Leu
        195                 200                 205

Pro Lys Phe Ile Asp Asn Ile Leu Val Phe Gln Lys Ile Lys Glu Pro
    210                 215                 220

Ile Ala Lys Glu Leu Glu His Ile Arg Ala Asp Phe Ser Ala Gly Gly
225                 230                 235                 240

Tyr Ile Lys Lys Asp Glu Arg Leu Glu Asp Ile Phe Ser Leu Asn Tyr
                245                 250                 255

Tyr Ile His Val Leu Ser Gln Ala Gly Ile Glu Lys Tyr Asn Ala Leu
            260                 265                 270

Ile Gly Lys Ile Val Thr Glu Gly Asp Gly Glu Met Lys Gly Leu Asn
        275                 280                 285

Glu His Ile Asn Leu Tyr Asn Gln Gln Arg Gly Arg Glu Asp Arg Leu
    290                 295                 300

Pro Leu Phe Arg Pro Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Gln
305                 310                 315                 320

Leu Ser Tyr Leu Pro Glu Ser Phe Glu Lys Asp Glu Glu Leu Leu Arg
                325                 330                 335

Ala Leu Lys Glu Phe Tyr Asp His Ile Ala Glu Asp Ile Leu Gly Arg
            340                 345                 350

Thr Gln Gln Leu Met Thr Ser Ile Ser Glu Tyr Asp Leu Ser Arg Ile
        355                 360                 365

Tyr Val Arg Asn Asp Ser Gln Leu Thr Asp Ile Ser Lys Lys Met Leu
    370                 375                 380
```

```
Gly Asp Trp Asn Ala Ile Tyr Met Ala Arg Glu Arg Ala Tyr Asp His
385                 390                 395                 400

Glu Gln Ala Pro Lys Arg Ile Thr Ala Lys Tyr Glu Arg Asp Arg Ile
                405                 410                 415

Lys Ala Leu Lys Gly Glu Glu Ser Ile Ser Leu Ala Asn Leu Asn Ser
            420                 425                 430

Cys Ile Ala Phe Leu Asp Asn Val Arg Asp Cys Arg Val Asp Thr Tyr
        435                 440                 445

Leu Ser Thr Leu Gly Gln Lys Glu Gly Pro His Gly Leu Ser Asn Leu
    450                 455                 460

Val Glu Asn Val Phe Ala Ser Tyr His Glu Ala Glu Gln Leu Leu Ser
465                 470                 475                 480

Phe Pro Tyr Pro Glu Glu Asn Asn Leu Ile Gln Asp Lys Asp Asn Val
                485                 490                 495

Val Leu Ile Lys Asn Leu Leu Asp Asn Ile Ser Asp Leu Gln Arg Phe
            500                 505                 510

Leu Lys Pro Leu Trp Gly Met Gly Asp Glu Pro Asp Lys Asp Glu Arg
        515                 520                 525

Phe Tyr Gly Glu Tyr Asn Tyr Ile Arg Gly Ala Leu Asp Gln Val Ile
    530                 535                 540

Pro Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Arg Lys Pro Tyr Ser
545                 550                 555                 560

Thr Arg Lys Val Lys Leu Asn Phe Gly Asn Ser Gln Leu Leu Ser Gly
                565                 570                 575

Trp Asp Arg Asn Lys Glu Lys Asp Asn Ser Cys Val Ile Leu Arg Lys
            580                 585                 590

Gly Gln Asn Phe Tyr Leu Ala Ile Met Asn Asn Arg His Lys Arg Ser
        595                 600                 605

Phe Glu Asn Lys Met Leu Pro Glu Tyr Lys Glu Gly Glu Pro Tyr Phe
    610                 615                 620

Glu Lys Met Asp Tyr Lys Phe Leu Pro Asp Pro Asn Lys Met Leu Pro
625                 630                 635                 640

Lys Val Phe Leu Ser Lys Lys Gly Ile Glu Ile Tyr Lys Pro Ser Pro
                645                 650                 655

Lys Leu Leu Glu Gln Tyr Gly His Gly Thr His Lys Lys Gly Asp Thr
            660                 665                 670

Phe Ser Met Asp Asp Leu His Glu Leu Ile Asp Phe Phe Lys His Ser
        675                 680                 685

Ile Glu Ala His Glu Asp Trp Lys Gln Phe Gly Phe Lys Phe Ser Asp
    690                 695                 700

Thr Ala Thr Tyr Glu Asn Val Ser Ser Phe Tyr Arg Glu Val Glu Asp
705                 710                 715                 720
```

```
Gln Gly Tyr Lys Leu Ser Phe Arg Lys Val Ser Glu Ser Tyr Val Tyr
                725                 730                 735

Ser Leu Ile Asp Gln Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys
            740                 745                 750

Asp Phe Ser Pro Cys Ser Lys Gly Thr Pro Asn Leu His Thr Leu Tyr
        755                 760                 765

Trp Arg Met Leu Phe Asp Glu Arg Asn Leu Ala Asp Val Ile Tyr Lys
    770                 775                 780

Leu Asp Gly Lys Ala Glu Ile Phe Phe Arg Glu Lys Ser Leu Lys Asn
785                 790                 795                 800

Asp His Pro Thr His Pro Ala Gly Lys Pro Ile Lys Lys Lys Ser Arg
                805                 810                 815

Gln Lys Lys Gly Glu Glu Ser Leu Phe Glu Tyr Asp Leu Val Lys Asp
            820                 825                 830

Arg Arg Tyr Thr Met Asp Lys Phe Gln Phe His Val Pro Ile Thr Met
        835                 840                 845

Asn Phe Lys Cys Ser Ala Gly Ser Lys Val Asn Asp Met Val Asn Ala
    850                 855                 860

His Ile Arg Glu Ala Lys Asp Met His Val Ile Gly Ile Asp Arg Gly
865                 870                 875                 880

Glu Arg Asn Leu Leu Tyr Ile Cys Val Ile Asp Ser Arg Gly Thr Ile
                885                 890                 895

Leu Asp Gln Ile Ser Leu Asn Thr Ile Asn Asp Ile Asp Tyr His Asp
            900                 905                 910

Leu Leu Glu Ser Arg Asp Lys Asp Arg Gln Gln Glu His Arg Asn Trp
        915                 920                 925

Gln Thr Ile Glu Gly Ile Lys Glu Leu Lys Gln Gly Tyr Leu Ser Gln
    930                 935                 940

Ala Val His Arg Ile Ala Glu Leu Met Val Ala Tyr Lys Ala Val Val
945                 950                 955                 960

Ala Leu Glu Asp Leu Asn Met Gly Phe Lys Arg Gly Arg Gln Lys Val
                965                 970                 975

Glu Ser Ser Val Tyr Gln Gln Phe Glu Lys Gln Leu Ile Asp Lys Leu
            980                 985                 990

Asn Tyr Leu Val Asp Lys Lys Lys Arg Pro Glu Asp Ile Gly Gly Leu
        995                1000                1005

Leu Arg Ala Tyr Gln Phe Thr Ala Pro Phe Lys Ser Phe Lys Glu
    1010                1015                1020

Met Gly Lys Gln Asn Gly Phe Leu Phe Tyr Ile Pro Ala Trp Asn
    1025                1030                1035

Thr Ser Asn Ile Asp Pro Thr Thr Gly Phe Val Asn Leu Phe His
    1040                1045                1050
```

Val Gln Tyr Glu Asn Val Asp Lys Ala Lys Ser Phe Phe Gln Lys
1055 1060 1065

Phe Asp Ser Ile Ser Tyr Asn Pro Lys Lys Asp Trp Phe Glu Phe
1070 1075 1080

Ala Phe Asp Tyr Lys Asn Phe Thr Lys Lys Ala Glu Gly Ser Arg
1085 1090 1095

Ser Met Trp Ile Leu Cys Thr His Gly Ser Arg Ile Lys Asn Phe
1100 1105 1110

Arg Asn Ser Gln Lys Asn Gly Gln Trp Asp Ser Glu Glu Phe Ala
1115 1120 1125

Leu Thr Glu Ala Phe Lys Ser Leu Phe Val Arg Tyr Glu Ile Asp
1130 1135 1140

Tyr Thr Ala Asp Leu Lys Thr Ala Ile Val Asp Glu Lys Gln Lys
1145 1150 1155

Asp Phe Phe Val Asp Leu Leu Lys Leu Phe Lys Leu Thr Val Gln
1160 1165 1170

Met Arg Asn Ser Trp Lys Glu Lys Asp Leu Asp Tyr Leu Ile Ser
1175 1180 1185

Pro Val Ala Gly Ala Asp Gly Arg Phe Phe Asp Thr Arg Glu Gly
1190 1195 1200

Asn Lys Ser Leu Pro Lys Asp Ala Asp Ala Asn Gly Ala Tyr Asn
1205 1210 1215

Ile Ala Leu Lys Gly Leu Trp Ala Leu Arg Gln Ile Arg Gln Thr
1220 1225 1230

Ser Glu Gly Gly Lys Leu Lys Leu Ala Ile Ser Asn Lys Glu Trp
1235 1240 1245

Leu Gln Phe Val Gln Glu Arg Ser Tyr Glu Lys Asp
1250 1255 1260

<210> 93
<211> 1260
<212> PRT
<213> 狗口腔卟啉單胞菌

<400> 93
Met Asp Ser Leu Lys Asp Phe Thr Asn Leu Tyr Pro Val Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr Leu Glu Asn Ile Glu
20 25 30

Lys Ala Gly Ile Leu Lys Glu Asp Glu His Arg Ala Glu Ser Tyr Arg
35 40 45

Arg Val Lys Lys Ile Ile Asp Thr Tyr His Lys Val Phe Ile Asp Ser
50 55 60

Ser Leu Glu Asn Met Ala Lys Met Gly Ile Glu Asn Glu Ile Lys Ala
65 70 75 80

Met Leu Gln Ser Phe Cys Glu Leu Tyr Lys Lys Asp His Arg Thr Glu
85 90 95

```
Gly Glu Asp Lys Ala Leu Asp Lys Ile Arg Ala Val Leu Arg Gly Leu
            100                 105                 110

Ile Val Gly Ala Phe Thr Gly Val Cys Gly Arg Arg Glu Asn Thr Val
        115                 120                 125

Gln Asn Glu Lys Tyr Glu Ser Leu Phe Lys Glu Lys Leu Ile Lys Glu
    130                 135                 140

Ile Leu Pro Asp Phe Val Leu Ser Thr Glu Ala Glu Ser Leu Pro Phe
145                 150                 155                 160

Ser Val Glu Glu Ala Thr Arg Ser Leu Lys Glu Phe Asp Ser Phe Thr
                165                 170                 175

Ser Tyr Phe Ala Gly Phe Tyr Glu Asn Arg Lys Asn Ile Tyr Ser Thr
            180                 185                 190

Lys Pro Gln Ser Thr Ala Ile Ala Tyr Arg Leu Ile His Glu Asn Leu
        195                 200                 205

Pro Lys Phe Ile Asp Asn Ile Leu Val Phe Gln Lys Ile Lys Glu Pro
    210                 215                 220

Ile Ala Lys Glu Leu Glu His Ile Arg Ala Asp Phe Ser Ala Gly Gly
225                 230                 235                 240

Tyr Ile Lys Lys Asp Glu Arg Leu Glu Asp Ile Phe Ser Leu Asn Tyr
                245                 250                 255

Tyr Ile His Val Leu Ser Gln Ala Gly Ile Glu Lys Tyr Asn Ala Leu
            260                 265                 270

Ile Gly Lys Ile Val Thr Glu Gly Asp Gly Glu Met Lys Gly Leu Asn
        275                 280                 285

Glu His Ile Asn Leu Tyr Asn Gln Gln Arg Gly Arg Glu Asp Arg Leu
    290                 295                 300

Pro Leu Phe Arg Pro Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Gln
305                 310                 315                 320

Leu Ser Tyr Leu Pro Glu Ser Phe Glu Lys Asp Glu Glu Leu Leu Arg
                325                 330                 335

Ala Leu Lys Glu Phe Tyr Asp His Ile Ala Glu Asp Ile Leu Gly Arg
            340                 345                 350

Thr Gln Gln Leu Met Thr Ser Ile Ser Glu Tyr Asp Leu Ser Arg Ile
        355                 360                 365

Tyr Val Arg Asn Asp Ser Gln Leu Thr Asp Ile Ser Lys Lys Met Leu
    370                 375                 380

Gly Asp Trp Asn Ala Ile Tyr Met Ala Arg Glu Arg Ala Tyr Asp His
385                 390                 395                 400

Glu Gln Ala Pro Lys Arg Ile Thr Ala Lys Tyr Glu Arg Asp Arg Ile
                405                 410                 415

Lys Ala Leu Lys Gly Glu Glu Ser Ile Ser Leu Ala Asn Leu Asn Ser
            420                 425                 430
```

```
Cys Ile Ala Phe Leu Asp Asn Val Arg Asp Cys Arg Val Asp Thr Tyr
        435                 440                 445

Leu Ser Thr Leu Gly Gln Lys Glu Gly Pro His Gly Leu Ser Asn Leu
    450                 455                 460

Val Glu Asn Val Phe Ala Ser Tyr His Glu Ala Glu Gln Leu Leu Ser
465                 470                 475                 480

Phe Pro Tyr Pro Glu Glu Asn Asn Leu Ile Gln Asp Lys Asp Asn Val
                485                 490                 495

Val Leu Ile Lys Asn Leu Leu Asp Asn Ile Ser Asp Leu Gln Arg Phe
            500                 505                 510

Leu Lys Pro Leu Trp Gly Met Gly Asp Glu Pro Asp Lys Asp Glu Arg
        515                 520                 525

Phe Tyr Gly Glu Tyr Asn Tyr Ile Arg Gly Ala Leu Asp Gln Val Ile
    530                 535                 540

Pro Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Arg Lys Pro Tyr Ser
545                 550                 555                 560

Thr Arg Lys Val Lys Leu Asn Phe Gly Asn Ser Gln Leu Leu Ser Gly
                565                 570                 575

Trp Asp Arg Asn Lys Glu Lys Asp Asn Ser Cys Val Ile Leu Arg Lys
            580                 585                 590

Gly Gln Asn Phe Tyr Leu Ala Ile Met Asn Asn Arg His Lys Arg Ser
        595                 600                 605

Phe Glu Asn Lys Val Leu Pro Glu Tyr Lys Glu Gly Glu Pro Tyr Phe
    610                 615                 620

Glu Lys Met Asp Tyr Lys Phe Leu Pro Asp Pro Asn Lys Met Leu Pro
625                 630                 635                 640

Lys Val Phe Leu Ser Lys Lys Gly Ile Glu Ile Tyr Lys Pro Ser Pro
                645                 650                 655

Lys Leu Leu Glu Gln Tyr Gly His Gly Thr His Lys Lys Gly Asp Thr
            660                 665                 670

Phe Ser Met Asp Asp Leu His Glu Leu Ile Asp Phe Phe Lys His Ser
        675                 680                 685

Ile Glu Ala His Glu Asp Trp Lys Gln Phe Gly Phe Lys Phe Ser Asp
    690                 695                 700

Thr Ala Thr Tyr Glu Asn Val Ser Ser Phe Tyr Arg Glu Val Glu Asp
705                 710                 715                 720

Gln Gly Tyr Lys Leu Ser Phe Arg Lys Val Ser Glu Ser Tyr Val Tyr
                725                 730                 735

Ser Leu Ile Asp Gln Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys
            740                 745                 750

Asp Phe Ser Pro Cys Ser Lys Gly Thr Pro Asn Leu His Thr Leu Tyr
        755                 760                 765
```

```
Trp Arg Met Leu Phe Asp Glu Arg Asn Leu Ala Asp Val Ile Tyr Lys
    770                 775                 780

Leu Asp Gly Lys Ala Glu Ile Phe Phe Arg Glu Lys Ser Leu Lys Asn
785                 790                 795                 800

Asp His Pro Thr His Pro Ala Gly Lys Pro Ile Lys Lys Lys Ser Arg
                805                 810                 815

Gln Lys Lys Gly Glu Glu Ser Leu Phe Glu Tyr Asp Leu Val Lys Asp
            820                 825                 830

Arg His Tyr Thr Met Asp Lys Phe Gln Phe His Val Pro Ile Thr Met
        835                 840                 845

Asn Phe Lys Cys Ser Ala Gly Ser Lys Val Asn Asp Met Val Asn Ala
    850                 855                 860

His Ile Arg Glu Ala Lys Asp Met His Val Ile Gly Ile Asp Arg Gly
865                 870                 875                 880

Glu Arg Asn Leu Leu Tyr Ile Cys Val Ile Asp Ser Arg Gly Thr Ile
                885                 890                 895

Leu Asp Gln Ile Ser Leu Asn Thr Ile Asn Asp Ile Asp Tyr His Asp
            900                 905                 910

Leu Leu Glu Ser Arg Asp Lys Asp Arg Gln Gln Glu Arg Arg Asn Trp
        915                 920                 925

Gln Thr Ile Glu Gly Ile Lys Glu Leu Lys Gln Gly Tyr Leu Ser Gln
    930                 935                 940

Ala Val His Arg Ile Ala Glu Leu Met Val Ala Tyr Lys Ala Val Val
945                 950                 955                 960

Ala Leu Glu Asp Leu Asn Met Gly Phe Lys Arg Gly Arg Gln Lys Val
                965                 970                 975

Glu Ser Ser Val Tyr Gln Gln Phe Glu Lys Gln Leu Ile Asp Lys Leu
            980                 985                 990

Asn Tyr Leu Val Asp Lys Lys Lys Arg Pro Glu Asp Ile Gly Gly Leu
        995                 1000                1005

Leu Arg  Ala Tyr Gln Phe Thr Ala Pro Phe Lys  Ser Phe Lys Glu
    1010                1015                1020

Met Gly  Lys Gln Asn Gly Phe Leu Phe Tyr Ile  Pro Ala Trp Asn
    1025                1030                1035

Thr Ser  Asn Ile Asp Pro Thr Thr Gly Phe Val  Asn Leu Phe His
    1040                1045                1050

Ala Gln  Tyr Glu Asn Val Asp Lys Ala Lys Ser  Phe Phe Gln Lys
    1055                1060                1065

Phe Asp  Ser Ile Ser Tyr Asn Pro Lys Lys Asp  Trp Phe Glu Phe
    1070                1075                1080

Ala Phe  Asp Tyr Lys Asn Phe Thr Lys Lys Ala  Glu Gly Ser Arg
    1085                1090                1095
```

Ser Met Trp Ile Leu Cys Thr His Gly Ser Arg Ile Lys Asn Phe
1100 1105 1110

Arg Asn Ser Gln Lys Asn Gly Gln Trp Asp Ser Glu Glu Phe Ala
1115 1120 1125

Leu Thr Glu Ala Phe Lys Ser Leu Phe Val Arg Tyr Glu Ile Asp
1130 1135 1140

Tyr Thr Ala Asp Leu Lys Thr Ala Ile Val Asp Glu Lys Gln Lys
1145 1150 1155

Asp Phe Phe Val Asp Leu Leu Lys Leu Phe Lys Leu Thr Val Gln
1160 1165 1170

Met Arg Asn Ser Trp Lys Glu Lys Asp Leu Asp Tyr Leu Ile Ser
1175 1180 1185

Pro Val Ala Gly Ala Asp Gly Arg Phe Phe Asp Thr Arg Glu Gly
1190 1195 1200

Asn Lys Ser Leu Pro Lys Asp Ala Asp Ala Asn Gly Ala Tyr Asn
1205 1210 1215

Ile Ala Leu Lys Gly Leu Trp Ala Leu Arg Gln Ile Arg Gln Thr
1220 1225 1230

Ser Glu Gly Gly Lys Leu Lys Leu Ala Ile Ser Asn Lys Glu Trp
1235 1240 1245

Leu Gln Phe Val Gln Glu Arg Ser Tyr Glu Lys Asp
1250 1255 1260

<210> 94
<211> 1260
<212> PRT
<213> 犬口腔卟啉單胞菌（Porphyromonas cansulci）

<400> 94
Met Asp Ser Leu Lys Asp Phe Thr Asn Leu Tyr Pro Val Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr Leu Glu Asn Ile Glu
20 25 30

Lys Ala Gly Ile Leu Lys Glu Asp Glu His Arg Ala Glu Ser Tyr Arg
35 40 45

Arg Val Lys Lys Ile Ile Asp Thr Tyr His Lys Val Phe Ile Asp Ser
50 55 60

Ser Leu Glu Asn Met Ala Lys Met Gly Ile Glu Asn Glu Ile Lys Ala
65 70 75 80

Met Leu Gln Ser Phe Cys Glu Leu Tyr Lys Lys Asp His Arg Thr Glu
85 90 95

Gly Glu Asp Lys Ala Leu Asp Lys Ile Arg Ala Val Leu Arg Gly Leu
100 105 110

Ile Val Gly Ala Phe Thr Gly Val Cys Gly Arg Arg Glu Asn Thr Val
115 120 125

```
Gln Asn Glu Lys Tyr Glu Ser Leu Phe Lys Glu Lys Leu Ile Lys Glu
    130                 135                 140

Ile Leu Pro Asp Phe Val Leu Ser Thr Glu Ala Glu Ser Leu Pro Phe
145                 150                 155                 160

Ser Val Glu Glu Ala Thr Arg Ser Leu Lys Glu Phe Asp Ser Phe Thr
                165                 170                 175

Ser Tyr Phe Ala Gly Phe Tyr Glu Asn Arg Lys Asn Ile Tyr Ser Thr
            180                 185                 190

Lys Pro Gln Ser Thr Ala Ile Ala Tyr Arg Leu Ile His Glu Asn Leu
        195                 200                 205

Pro Lys Phe Ile Asp Asn Ile Leu Val Phe Gln Lys Ile Lys Glu Pro
    210                 215                 220

Ile Ala Lys Glu Leu Glu His Ile Arg Ala Asp Phe Ser Ala Gly Gly
225                 230                 235                 240

Tyr Ile Lys Lys Asp Glu Arg Leu Glu Asp Ile Phe Ser Leu Asn Tyr
                245                 250                 255

Tyr Ile His Val Leu Ser Gln Ala Gly Ile Glu Lys Tyr Asn Ala Leu
            260                 265                 270

Ile Gly Lys Ile Val Thr Glu Gly Asp Gly Glu Met Lys Gly Leu Asn
        275                 280                 285

Glu His Ile Asn Leu Tyr Asn Gln Gln Arg Gly Arg Glu Asp Arg Leu
    290                 295                 300

Pro Leu Phe Arg Pro Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Gln
305                 310                 315                 320

Leu Ser Tyr Leu Pro Glu Ser Phe Glu Lys Asp Glu Glu Leu Leu Arg
                325                 330                 335

Ala Leu Lys Glu Phe Tyr Asp His Ile Ala Glu Asp Ile Leu Gly Arg
            340                 345                 350

Thr Gln Gln Leu Met Thr Ser Ile Ser Glu Tyr Asp Leu Ser Arg Ile
        355                 360                 365

Tyr Val Arg Asn Asp Ser Gln Leu Thr Asp Ile Ser Lys Lys Met Leu
    370                 375                 380

Gly Asp Trp Asn Ala Ile Tyr Met Ala Arg Glu Arg Ala Tyr Asp His
385                 390                 395                 400

Glu Gln Ala Pro Lys Arg Ile Thr Ala Lys Tyr Glu Arg Asp Arg Ile
                405                 410                 415

Lys Ala Leu Lys Gly Glu Glu Ser Ile Ser Leu Ala Asn Leu Asn Ser
            420                 425                 430

Cys Ile Ala Phe Leu Asp Asn Val Arg Asp Cys Arg Val Asp Thr Tyr
        435                 440                 445

Leu Ser Thr Leu Gly Gln Lys Glu Gly Pro His Gly Leu Ser Asn Leu
    450                 455                 460
```

```
Val Glu Asn Val Phe Ala Ser Tyr His Glu Ala Glu Gln Leu Leu Ser
465                 470                 475                 480

Phe Pro Tyr Pro Glu Glu Asn Asn Leu Ile Gln Asp Lys Asp Asn Val
                485                 490                 495

Val Leu Ile Lys Asn Leu Leu Asp Asn Ile Ser Asp Leu Gln Arg Phe
            500                 505                 510

Leu Lys Pro Leu Trp Gly Met Gly Asp Glu Pro Asp Lys Asp Glu Arg
        515                 520                 525

Phe Tyr Gly Glu Tyr Asn Tyr Ile Arg Gly Ala Leu Asp Gln Val Ile
    530                 535                 540

Pro Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Arg Lys Pro Tyr Ser
545                 550                 555                 560

Thr Arg Lys Val Lys Leu Asn Phe Gly Asn Ser Gln Leu Leu Ser Gly
                565                 570                 575

Trp Asp Arg Asn Lys Glu Lys Asp Asn Ser Cys Val Ile Leu Arg Lys
            580                 585                 590

Gly Gln Asn Phe Tyr Leu Ala Ile Met Asn Asn Arg His Lys Arg Ser
        595                 600                 605

Phe Glu Asn Lys Val Leu Pro Glu Tyr Lys Glu Gly Glu Pro Tyr Phe
    610                 615                 620

Glu Lys Met Asp Tyr Lys Phe Leu Pro Asp Pro Asn Lys Met Leu Pro
625                 630                 635                 640

Lys Val Phe Leu Ser Lys Lys Gly Ile Glu Ile Tyr Lys Pro Ser Pro
                645                 650                 655

Lys Leu Leu Glu Gln Tyr Gly His Gly Thr His Lys Lys Gly Asp Thr
            660                 665                 670

Phe Ser Met Asp Asp Leu His Glu Leu Ile Asp Phe Phe Lys His Ser
        675                 680                 685

Ile Glu Ala His Glu Asp Trp Lys Gln Phe Gly Phe Lys Phe Ser Asp
    690                 695                 700

Thr Ala Thr Tyr Glu Asn Val Ser Ser Phe Tyr Arg Glu Val Glu Asp
705                 710                 715                 720

Gln Gly Tyr Lys Leu Ser Phe Arg Lys Val Ser Glu Ser Tyr Val Tyr
                725                 730                 735

Ser Leu Ile Asp Gln Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys
            740                 745                 750

Asp Phe Ser Pro Cys Ser Lys Gly Thr Pro Asn Leu His Thr Leu Tyr
        755                 760                 765

Trp Arg Met Leu Phe Asp Glu Arg Asn Leu Ala Asp Val Ile Tyr Lys
    770                 775                 780

Leu Asp Gly Lys Ala Glu Ile Phe Phe Arg Glu Lys Ser Leu Lys Asn
785                 790                 795                 800
```

```
Asp His Pro Thr His Pro Ala Gly Lys Pro Ile Lys Lys Lys Ser Arg
                805                 810                 815

Gln Lys Lys Gly Glu Glu Ser Leu Phe Glu Tyr Asp Leu Val Lys Asp
            820                 825                 830

Arg Arg Tyr Thr Met Asp Lys Phe Gln Phe His Val Pro Ile Thr Met
        835                 840                 845

Asn Phe Lys Cys Ser Ala Gly Ser Lys Val Asn Asp Met Val Asn Ala
    850                 855                 860

His Ile Arg Glu Ala Lys Asp Met His Val Ile Gly Ile Asp Arg Gly
865                 870                 875                 880

Glu Arg Asn Leu Leu Tyr Ile Cys Val Ile Asp Ser Arg Gly Thr Ile
                885                 890                 895

Leu Asp Gln Ile Ser Leu Asn Thr Ile Asn Asp Ile Asp Tyr His Asp
            900                 905                 910

Leu Leu Glu Ser Arg Asp Lys Asp Arg Gln Gln Glu Arg Arg Asn Trp
        915                 920                 925

Gln Thr Ile Glu Gly Ile Lys Glu Leu Lys Gln Gly Tyr Leu Ser Gln
    930                 935                 940

Ala Val His Arg Ile Ala Glu Leu Met Val Ala Tyr Lys Ala Val Val
945                 950                 955                 960

Ala Leu Glu Asp Leu Asn Met Gly Phe Lys Arg Gly Arg Gln Lys Val
                965                 970                 975

Glu Ser Ser Val Tyr Gln Gln Phe Glu Lys Gln Leu Ile Asp Lys Leu
            980                 985                 990

Asn Tyr Leu Val Asp Lys Lys Lys Arg Pro Glu Asp Ile Gly Gly Leu
        995                1000                1005

Leu Arg  Ala Tyr Gln Phe Thr Ala Pro Phe Lys  Ser Phe Lys Glu
    1010                1015                1020

Met Gly  Lys Gln Asn Gly Phe Leu Phe Tyr Ile  Pro Ala Trp Asn
    1025                1030                1035

Thr Ser  Asn Ile Asp Pro Thr Thr Gly Phe Val  Asn Leu Phe His
    1040                1045                1050

Ala Gln  Tyr Glu Asn Val Asp Lys Ala Lys Ser  Phe Phe Gln Lys
    1055                1060                1065

Phe Asp  Ser Ile Ser Tyr Asn Pro Lys Lys Asp  Trp Phe Glu Phe
    1070                1075                1080

Ala Phe  Asp Tyr Lys Asn Phe Thr Lys Lys Ala  Glu Gly Ser Arg
    1085                1090                1095

Ser Met  Trp Ile Leu Cys Thr His Gly Ser Arg  Ile Lys Asn Phe
    1100                1105                1110

Arg Asn  Ser Gln Lys Asn Gly Gln Trp Asp Ser  Glu Glu Phe Ala
    1115                1120                1125
```

Leu Thr Glu Ala Phe Lys Ser Leu Phe Val Arg Tyr Glu Ile Asp
1130 1135 1140

Tyr Thr Ala Asp Leu Lys Thr Ala Ile Val Asp Glu Lys Gln Lys
1145 1150 1155

Asp Phe Phe Val Asp Leu Leu Lys Leu Phe Lys Leu Thr Val Gln
1160 1165 1170

Met Arg Asn Ser Trp Lys Glu Lys Asp Leu Asp Tyr Leu Ile Ser
1175 1180 1185

Pro Val Ala Gly Ala Asp Gly Arg Phe Phe Asp Thr Arg Glu Gly
1190 1195 1200

Asn Lys Ser Leu Pro Lys Asp Ala Asp Ala Asn Gly Ala Tyr Asn
1205 1210 1215

Ile Ala Leu Lys Gly Leu Trp Ala Leu Arg Gln Ile Arg Gln Thr
1220 1225 1230

Ser Glu Gly Gly Lys Leu Lys Leu Ala Ile Ser Asn Lys Glu Trp
1235 1240 1245

Leu Gln Phe Val Gln Glu Arg Ser Tyr Glu Lys Asp
1250 1255 1260

<210> 95
<211> 1259
<212> PRT
<213> 窮氏互養菌（Synergistes jonesii）

<400> 95
Met Ala Asn Ser Leu Lys Asp Phe Thr Asn Ile Tyr Gln Leu Ser Lys
1 5 10 15

Thr Leu Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Glu Glu His Ile
20 25 30

Asn Arg Lys Leu Ile Ile Met His Asp Glu Lys Arg Gly Glu Asp Tyr
35 40 45

Lys Ser Val Thr Lys Leu Ile Asp Asp Tyr His Arg Lys Phe Ile His
50 55 60

Glu Thr Leu Asp Pro Ala His Phe Asp Trp Asn Pro Leu Ala Glu Ala
65 70 75 80

Leu Ile Gln Ser Gly Ser Lys Asn Asn Lys Ala Leu Pro Ala Glu Gln
85 90 95

Lys Glu Met Arg Glu Lys Ile Ile Ser Met Phe Thr Ser Gln Ala Val
100 105 110

Tyr Lys Lys Leu Phe Lys Lys Glu Leu Phe Ser Glu Leu Leu Pro Glu
115 120 125

Met Ile Lys Ser Glu Leu Val Ser Asp Leu Glu Lys Gln Ala Gln Leu
130 135 140

Asp Ala Val Lys Ser Phe Asp Lys Phe Ser Thr Tyr Phe Thr Gly Phe
145 150 155 160

His Glu Asn Arg Lys Asn Ile Tyr Ser Lys Lys Asp Thr Ser Thr Ser
165 170 175

```
Ile Ala Phe Arg Ile Val His Gln Asn Phe Pro Lys Phe Leu Ala Asn
            180                 185                 190

Val Arg Ala Tyr Thr Leu Ile Lys Glu Arg Ala Pro Glu Val Ile Asp
        195                 200                 205

Lys Ala Gln Lys Glu Leu Ser Gly Ile Leu Gly Gly Lys Thr Leu Asp
    210                 215                 220

Asp Ile Phe Ser Ile Glu Ser Phe Asn Asn Val Leu Thr Gln Asp Lys
225                 230                 235                 240

Ile Asp Tyr Tyr Asn Gln Ile Ile Gly Gly Val Ser Gly Lys Ala Gly
                245                 250                 255

Asp Lys Lys Leu Arg Gly Val Asn Glu Phe Ser Asn Leu Tyr Arg Gln
            260                 265                 270

Gln His Pro Glu Val Ala Ser Leu Arg Ile Lys Met Val Pro Leu Tyr
        275                 280                 285

Lys Gln Ile Leu Ser Asp Arg Thr Thr Leu Ser Phe Val Pro Glu Ala
    290                 295                 300

Leu Lys Asp Asp Glu Gln Ala Ile Asn Ala Val Asp Gly Leu Arg Ser
305                 310                 315                 320

Glu Leu Glu Arg Asn Asp Ile Phe Asn Arg Ile Lys Arg Leu Phe Gly
                325                 330                 335

Lys Asn Asn Leu Tyr Ser Leu Asp Lys Ile Trp Ile Lys Asn Ser Ser
            340                 345                 350

Ile Ser Ala Phe Ser Asn Glu Leu Phe Lys Asn Trp Ser Phe Ile Glu
        355                 360                 365

Asp Ala Leu Lys Glu Phe Lys Glu Asn Glu Phe Asn Gly Ala Arg Ser
    370                 375                 380

Ala Gly Lys Lys Ala Glu Lys Trp Leu Lys Ser Lys Tyr Phe Ser Phe
385                 390                 395                 400

Ala Asp Ile Asp Ala Ala Val Lys Ser Tyr Ser Glu Gln Val Ser Ala
                405                 410                 415

Asp Ile Ser Ser Ala Pro Ser Ala Ser Tyr Phe Ala Lys Phe Thr Asn
            420                 425                 430

Leu Ile Glu Thr Ala Ala Glu Asn Gly Arg Lys Phe Ser Tyr Phe Ala
        435                 440                 445

Ala Glu Ser Lys Ala Phe Arg Gly Asp Asp Gly Lys Thr Glu Ile Ile
    450                 455                 460

Lys Ala Tyr Leu Asp Ser Leu Asn Asp Ile Leu His Cys Leu Lys Pro
465                 470                 475                 480

Phe Glu Thr Glu Asp Ile Ser Asp Ile Asp Thr Glu Phe Tyr Ser Ala
                485                 490                 495

Phe Ala Glu Ile Tyr Asp Ser Val Lys Asp Val Ile Pro Val Tyr Asn
            500                 505                 510
```

```
Ala Val Arg Asn Tyr Thr Thr Gln Lys Pro Phe Ser Thr Glu Lys Phe
        515                 520                 525

Lys Leu Asn Phe Glu Asn Pro Ala Leu Ala Lys Gly Trp Asp Lys Asn
    530                 535                 540

Lys Glu Gln Asn Asn Thr Ala Ile Ile Leu Met Lys Asp Gly Lys Tyr
545                 550                 555                 560

Tyr Leu Gly Val Ile Asp Lys Asn Asn Lys Leu Arg Ala Asp Asp Leu
                565                 570                 575

Ala Asp Asp Gly Ser Ala Tyr Gly Tyr Met Lys Met Asn Tyr Lys Phe
            580                 585                 590

Ile Pro Thr Pro His Met Glu Leu Pro Lys Val Phe Leu Pro Lys Arg
        595                 600                 605

Ala Pro Lys Arg Tyr Asn Pro Ser Arg Glu Ile Leu Leu Ile Lys Glu
    610                 615                 620

Asn Lys Thr Phe Ile Lys Asp Lys Asn Phe Asn Arg Thr Asp Cys His
625                 630                 635                 640

Lys Leu Ile Asp Phe Phe Lys Asp Ser Ile Asn Lys His Lys Asp Trp
                645                 650                 655

Arg Thr Phe Gly Phe Asp Phe Ser Asp Thr Asp Ser Tyr Glu Asp Ile
            660                 665                 670

Ser Asp Phe Tyr Met Glu Val Gln Asp Gln Gly Tyr Lys Leu Thr Phe
        675                 680                 685

Thr Arg Leu Ser Ala Glu Lys Ile Asp Lys Trp Val Glu Glu Gly Arg
    690                 695                 700

Leu Phe Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ala Asp Gly Ala Gln
705                 710                 715                 720

Gly Ser Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Ile Phe Ser Glu
                725                 730                 735

Glu Asn Leu Lys Asp Val Val Leu Lys Leu Asn Gly Glu Ala Glu Leu
            740                 745                 750

Phe Phe Arg Arg Lys Ser Ile Asp Lys Pro Ala Val His Ala Lys Gly
        755                 760                 765

Ser Met Lys Val Asn Arg Arg Asp Ile Asp Gly Asn Pro Ile Asp Glu
    770                 775                 780

Gly Thr Tyr Val Glu Ile Cys Gly Tyr Ala Asn Gly Lys Arg Asp Met
785                 790                 795                 800

Ala Ser Leu Asn Ala Gly Ala Arg Gly Leu Ile Glu Ser Gly Leu Val
                805                 810                 815

Arg Ile Thr Glu Val Lys His Glu Leu Val Lys Asp Lys Arg Tyr Thr
            820                 825                 830

Ile Asp Lys Tyr Phe Phe His Val Pro Phe Thr Ile Asn Phe Lys Ala
        835                 840                 845
```

```
Gln Gly Gln Gly Asn Ile Asn Ser Asp Val Asn Leu Phe Leu Arg Asn
    850                 855                 860

Asn Lys Asp Val Asn Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu
865                 870                 875                 880

Val Tyr Val Ser Leu Ile Asp Arg Asp Gly His Ile Lys Leu Gln Lys
                885                 890                 895

Asp Phe Asn Ile Ile Gly Gly Met Asp Tyr His Ala Lys Leu Asn Gln
            900                 905                 910

Lys Glu Lys Glu Arg Asp Thr Ala Arg Lys Ser Trp Lys Thr Ile Gly
        915                 920                 925

Thr Ile Lys Glu Leu Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu
    930                 935                 940

Ile Val Arg Leu Ala Val Asp Asn Asn Ala Val Ile Val Met Glu Asp
945                 950                 955                 960

Leu Asn Ile Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val
                965                 970                 975

Tyr Gln Lys Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr Leu Val
            980                 985                 990

Phe Lys Asp Ala Gly Tyr Asp Ala Pro Cys Gly Ile Leu Lys Gly Leu
        995                 1000                1005

Gln Leu  Thr Glu Lys Phe Glu  Ser Phe Thr Lys Leu  Gly Lys Gln
    1010                1015                1020

Cys Gly  Ile Ile Phe Tyr Ile  Pro Ala Gly Tyr Thr  Ser Lys Ile
    1025                1030                1035

Asp Pro  Thr Thr Gly Phe Val  Asn Leu Phe Asn Ile  Asn Asp Val
    1040                1045                1050

Ser Ser  Lys Glu Lys Gln Lys  Asp Phe Ile Gly Lys  Leu Asp Ser
    1055                1060                1065

Ile Arg  Phe Asp Ala Lys Arg  Asp Met Phe Thr Phe  Glu Phe Asp
    1070                1075                1080

Tyr Asp  Lys Phe Arg Thr Tyr  Gln Thr Ser Tyr Arg  Lys Lys Trp
    1085                1090                1095

Ala Val  Trp Thr Asn Gly Lys  Arg Ile Val Arg Glu  Lys Asp Lys
    1100                1105                1110

Asp Gly  Lys Phe Arg Met Asn  Asp Arg Leu Leu Thr  Glu Asp Met
    1115                1120                1125

Lys Asn  Ile Leu Asn Lys Tyr  Ala Leu Ala Tyr Lys  Ala Gly Glu
    1130                1135                1140

Asp Ile  Leu Pro Asp Val Ile  Ser Arg Asp Lys Ser  Leu Ala Ser
    1145                1150                1155

Glu Ile  Phe Tyr Val Phe Lys  Asn Thr Leu Gln Met  Arg Asn Ser
    1160                1165                1170
```

Lys Arg Asp Thr Gly Glu Asp Phe Ile Ile Ser Pro Val Leu Asn
1175 1180 1185

Ala Lys Gly Arg Phe Phe Asp Ser Arg Lys Thr Asp Ala Ala Leu
1190 1195 1200

Pro Ile Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Ala Leu Lys
1205 1210 1215

Gly Ser Leu Val Leu Asp Ala Ile Asp Glu Lys Leu Lys Glu Asp
1220 1225 1230

Gly Arg Ile Asp Tyr Lys Asp Met Ala Val Ser Asn Pro Lys Trp
1235 1240 1245

Phe Glu Phe Met Gln Thr Arg Lys Phe Asp Phe
1250 1255

<210> 96
<211> 1257
<212> PRT
<213> 布氏普雷沃菌（Prevotella bryantii）

<400> 96
Met Gln Ile Asn Asn Leu Lys Ile Ile Tyr Met Lys Phe Thr Asp Phe
1 5 10 15

Thr Gly Leu Tyr Ser Leu Ser Lys Thr Leu Arg Phe Glu Leu Lys Pro
20 25 30

Ile Gly Lys Thr Leu Glu Asn Ile Lys Lys Ala Gly Leu Leu Glu Gln
35 40 45

Asp Gln His Arg Ala Asp Ser Tyr Lys Lys Val Lys Lys Ile Ile Asp
50 55 60

Glu Tyr His Lys Ala Phe Ile Glu Lys Ser Leu Ser Asn Phe Glu Leu
65 70 75 80

Lys Tyr Gln Ser Glu Asp Lys Leu Asp Ser Leu Glu Glu Tyr Leu Met
85 90 95

Tyr Tyr Ser Met Lys Arg Ile Glu Lys Thr Glu Lys Asp Lys Phe Ala
100 105 110

Lys Ile Gln Asp Asn Leu Arg Lys Gln Ile Ala Asp His Leu Lys Gly
115 120 125

Asp Glu Ser Tyr Lys Thr Ile Phe Ser Lys Asp Leu Ile Arg Lys Asn
130 135 140

Leu Pro Asp Phe Val Lys Ser Asp Glu Glu Arg Thr Leu Ile Lys Glu
145 150 155 160

Phe Lys Asp Phe Thr Thr Tyr Phe Lys Gly Phe Tyr Glu Asn Arg Glu
165 170 175

Asn Met Tyr Ser Ala Glu Asp Lys Ser Thr Ala Ile Ser His Arg Ile
180 185 190

Ile His Glu Asn Leu Pro Lys Phe Val Asp Asn Ile Asn Ala Phe Ser
195 200 205

```
Lys Ile Ile Leu Ile Pro Glu Leu Arg Glu Lys Leu Asn Gln Ile Tyr
    210                 215                 220

Gln Asp Phe Glu Glu Tyr Leu Asn Val Glu Ser Ile Asp Glu Ile Phe
225                 230                 235                 240

His Leu Asp Tyr Phe Ser Met Val Met Thr Gln Lys Gln Ile Glu Val
                245                 250                 255

Tyr Asn Ala Ile Ile Gly Gly Lys Ser Thr Asn Asp Lys Lys Ile Gln
            260                 265                 270

Gly Leu Asn Glu Tyr Ile Asn Leu Tyr Asn Gln Lys His Lys Asp Cys
        275                 280                 285

Lys Leu Pro Lys Leu Lys Leu Leu Phe Lys Gln Ile Leu Ser Asp Arg
    290                 295                 300

Ile Ala Ile Ser Trp Leu Pro Asp Asn Phe Lys Asp Asp Gln Glu Ala
305                 310                 315                 320

Leu Asp Ser Ile Asp Thr Cys Tyr Lys Asn Leu Leu Asn Asp Gly Asn
                325                 330                 335

Val Leu Gly Glu Gly Asn Leu Lys Leu Leu Leu Glu Asn Ile Asp Thr
            340                 345                 350

Tyr Asn Leu Lys Gly Ile Phe Ile Arg Asn Asp Leu Gln Leu Thr Asp
        355                 360                 365

Ile Ser Gln Lys Met Tyr Ala Ser Trp Asn Val Ile Gln Asp Ala Val
    370                 375                 380

Ile Leu Asp Leu Lys Lys Gln Val Ser Arg Lys Lys Lys Glu Ser Ala
385                 390                 395                 400

Glu Asp Tyr Asn Asp Arg Leu Lys Lys Leu Tyr Thr Ser Gln Glu Ser
                405                 410                 415

Phe Ser Ile Gln Tyr Leu Asn Asp Cys Leu Arg Ala Tyr Gly Lys Thr
            420                 425                 430

Glu Asn Ile Gln Asp Tyr Phe Ala Lys Leu Gly Ala Val Asn Asn Glu
        435                 440                 445

His Glu Gln Thr Ile Asn Leu Phe Ala Gln Val Arg Asn Ala Tyr Thr
    450                 455                 460

Ser Val Gln Ala Ile Leu Thr Thr Pro Tyr Pro Glu Asn Ala Asn Leu
465                 470                 475                 480

Ala Gln Asp Lys Glu Thr Val Ala Leu Ile Lys Asn Leu Leu Asp Ser
                485                 490                 495

Leu Lys Arg Leu Gln Arg Phe Ile Lys Pro Leu Leu Gly Lys Gly Asp
            500                 505                 510

Glu Ser Asp Lys Asp Glu Arg Phe Tyr Gly Asp Phe Thr Pro Leu Trp
        515                 520                 525

Glu Thr Leu Asn Gln Ile Thr Pro Leu Tyr Asn Met Val Arg Asn Tyr
    530                 535                 540
```

```
Met Thr Arg Lys Pro Tyr Ser Gln Glu Lys Ile Lys Leu Asn Phe Glu
545                 550                 555                 560

Asn Ser Thr Leu Leu Gly Gly Trp Asp Leu Asn Lys Glu His Asp Asn
                565                 570                 575

Thr Ala Ile Ile Leu Arg Lys Asn Gly Leu Tyr Tyr Leu Ala Ile Met
            580                 585                 590

Lys Lys Ser Ala Asn Lys Ile Phe Asp Lys Asp Lys Leu Asp Asn Ser
        595                 600                 605

Gly Asp Cys Tyr Glu Lys Met Val Tyr Lys Leu Leu Pro Gly Ala Asn
    610                 615                 620

Lys Met Leu Pro Lys Val Phe Phe Ser Lys Ser Arg Ile Asp Glu Phe
625                 630                 635                 640

Lys Pro Ser Glu Asn Ile Ile Glu Asn Tyr Lys Lys Gly Thr His Lys
                645                 650                 655

Lys Gly Ala Asn Phe Asn Leu Ala Asp Cys His Asn Leu Ile Asp Phe
            660                 665                 670

Phe Lys Ser Ser Ile Ser Lys His Glu Asp Trp Ser Lys Phe Asn Phe
        675                 680                 685

His Phe Ser Asp Thr Ser Ser Tyr Glu Asp Leu Ser Asp Phe Tyr Arg
    690                 695                 700

Glu Val Glu Gln Gln Gly Tyr Ser Ile Ser Phe Cys Asp Val Ser Val
705                 710                 715                 720

Glu Tyr Ile Asn Lys Met Val Glu Lys Gly Asp Leu Tyr Leu Phe Gln
                725                 730                 735

Ile Tyr Asn Lys Asp Phe Ser Glu Phe Ser Lys Gly Thr Pro Asn Met
            740                 745                 750

His Thr Leu Tyr Trp Asn Ser Leu Phe Ser Lys Glu Asn Leu Asn Asn
        755                 760                 765

Ile Ile Tyr Lys Leu Asn Gly Gln Ala Glu Ile Phe Phe Arg Lys Lys
    770                 775                 780

Ser Leu Asn Tyr Lys Arg Pro Thr His Pro Ala His Gln Ala Ile Lys
785                 790                 795                 800

Asn Lys Asn Lys Cys Asn Glu Lys Lys Glu Ser Ile Phe Asp Tyr Asp
                805                 810                 815

Leu Val Lys Asp Lys Arg Tyr Thr Val Asp Lys Phe Gln Phe His Val
            820                 825                 830

Pro Ile Thr Met Asn Phe Lys Ser Thr Gly Asn Thr Asn Ile Asn Gln
        835                 840                 845

Gln Val Ile Asp Tyr Leu Arg Thr Glu Asp Asp Thr His Ile Ile Gly
    850                 855                 860

Ile Asp Arg Gly Glu Arg His Leu Leu Tyr Leu Val Val Ile Asp Ser
865                 870                 875                 880
```

```
His Gly Lys Ile Val Glu Gln Phe Thr Leu Asn Glu Ile Val Asn Glu
                885                 890                 895

Tyr Gly Gly Asn Ile Tyr Arg Thr Asn Tyr His Asp Leu Leu Asp Thr
            900                 905                 910

Arg Glu Gln Asn Arg Glu Lys Ala Arg Glu Ser Trp Gln Thr Ile Glu
        915                 920                 925

Asn Ile Lys Glu Leu Lys Glu Gly Tyr Ile Ser Gln Val Ile His Lys
    930                 935                 940

Ile Thr Asp Leu Met Gln Lys Tyr His Ala Val Val Val Leu Glu Asp
945                 950                 955                 960

Leu Asn Met Gly Phe Met Arg Gly Arg Gln Lys Val Glu Lys Gln Val
                965                 970                 975

Tyr Gln Lys Phe Glu Glu Met Leu Ile Asn Lys Leu Asn Tyr Leu Val
            980                 985                 990

Asn Lys Lys Ala Asp Gln Asn Ser Ala Gly Gly Leu Leu His Ala Tyr
        995                 1000                1005

Gln Leu  Thr Ser Lys Phe Glu  Ser Phe Gln Lys Leu  Gly Lys Gln
    1010                1015                1020

Ser Gly  Phe Leu Phe Tyr Ile  Pro Ala Trp Asn Thr  Ser Lys Ile
    1025                1030                1035

Asp Pro  Val Thr Gly Phe Val  Asn Leu Phe Asp Thr  Arg Tyr Glu
    1040                1045                1050

Ser Ile  Asp Lys Ala Lys Ala  Phe Phe Gly Lys Phe  Asp Ser Ile
    1055                1060                1065

Arg Tyr  Asn Ala Asp Lys Asp  Trp Phe Glu Phe Ala  Phe Asp Tyr
    1070                1075                1080

Asn Asn  Phe Thr Thr Lys Ala  Glu Gly Thr Arg Thr  Asn Trp Thr
    1085                1090                1095

Ile Cys  Thr Tyr Gly Ser Arg  Ile Arg Thr Phe Arg  Asn Gln Ala
    1100                1105                1110

Lys Asn  Ser Gln Trp Asp Asn  Glu Glu Ile Asp Leu  Thr Lys Ala
    1115                1120                1125

Tyr Lys  Ala Phe Phe Ala Lys  His Gly Ile Asn Ile  Tyr Asp Asn
    1130                1135                1140

Ile Lys  Glu Ala Ile Ala Met  Glu Thr Glu Lys Ser  Phe Phe Glu
    1145                1150                1155

Asp Leu  Leu His Leu Leu Lys  Leu Thr Leu Gln Met  Arg Asn Ser
    1160                1165                1170

Ile Thr  Gly Thr Thr Thr Asp  Tyr Leu Ile Ser Pro  Val His Asp
    1175                1180                1185

Ser Lys  Gly Asn Phe Tyr Asp  Ser Arg Ile Cys Asp  Asn Ser Leu
    1190                1195                1200
```

Pro Ala Asn Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala Arg Lys
1205 1210 1215

Gly Leu Met Leu Ile Gln Gln Ile Lys Asp Ser Thr Ser Ser Asn
1220 1225 1230

Arg Phe Lys Phe Ser Pro Ile Thr Asn Lys Asp Trp Leu Ile Phe
1235 1240 1245

Ala Gln Glu Lys Pro Tyr Leu Asn Asp
1250 1255

<210> 97
<211> 1253
<212> PRT
<213> 易北普雷沃菌（Prevotella albensis）

<400> 97
Met Asn Ile Lys Asn Phe Thr Gly Leu Tyr Pro Leu Ser Lys Thr Leu
1 5 10 15

Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Lys Glu Asn Ile Glu Lys
20 25 30

Asn Gly Ile Leu Thr Lys Asp Glu Gln Arg Ala Lys Asp Tyr Leu Ile
35 40 45

Val Lys Gly Phe Ile Asp Glu Tyr His Lys Gln Phe Ile Lys Asp Arg
50 55 60

Leu Trp Asp Phe Lys Leu Pro Leu Glu Ser Glu Gly Glu Lys Asn Ser
65 70 75 80

Leu Glu Glu Tyr Gln Glu Leu Tyr Glu Leu Thr Lys Arg Asn Asp Ala
85 90 95

Gln Glu Ala Asp Phe Thr Glu Ile Lys Asp Asn Leu Arg Ser Ser Ile
100 105 110

Thr Glu Gln Leu Thr Lys Ser Gly Ser Ala Tyr Asp Arg Ile Phe Lys
115 120 125

Lys Glu Phe Ile Arg Glu Asp Leu Val Asn Phe Leu Glu Asp Glu Lys
130 135 140

Asp Lys Asn Ile Val Lys Gln Phe Glu Asp Phe Thr Thr Tyr Phe Thr
145 150 155 160

Gly Phe Tyr Glu Asn Arg Lys Asn Met Tyr Ser Ser Glu Glu Lys Ser
165 170 175

Thr Ala Ile Ala Tyr Arg Leu Ile His Gln Asn Leu Pro Lys Phe Met
180 185 190

Asp Asn Met Arg Ser Phe Ala Lys Ile Ala Asn Ser Ser Val Ser Glu
195 200 205

His Phe Ser Asp Ile Tyr Glu Ser Trp Lys Glu Tyr Leu Asn Val Asn
210 215 220

Ser Ile Glu Glu Ile Phe Gln Leu Asp Tyr Phe Ser Glu Thr Leu Thr
225 230 235 240

Gln Pro His Ile Glu Val Tyr Asn Tyr Ile Ile Gly Lys Lys Val Leu
245 250 255

```
Glu Asp Gly Thr Glu Ile Lys Gly Ile Asn Glu Tyr Val Asn Leu Tyr
            260                 265                 270

Asn Gln Gln Gln Lys Asp Lys Ser Lys Arg Leu Pro Phe Leu Val Pro
        275                 280                 285

Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Lys Leu Ser Trp Ile Ala
    290                 295                 300

Glu Glu Phe Asp Ser Asp Lys Lys Met Leu Ser Ala Ile Thr Glu Ser
305                 310                 315                 320

Tyr Asn His Leu His Asn Val Leu Met Gly Asn Glu Asn Glu Ser Leu
                325                 330                 335

Arg Asn Leu Leu Leu Asn Ile Lys Asp Tyr Asn Leu Glu Lys Ile Asn
            340                 345                 350

Ile Thr Asn Asp Leu Ser Leu Thr Glu Ile Ser Gln Asn Leu Phe Gly
        355                 360                 365

Arg Tyr Asp Val Phe Thr Asn Gly Ile Lys Asn Lys Leu Arg Val Leu
    370                 375                 380

Thr Pro Arg Lys Lys Lys Glu Thr Asp Glu Asn Phe Glu Asp Arg Ile
385                 390                 395                 400

Asn Lys Ile Phe Lys Thr Gln Lys Ser Phe Ser Ile Ala Phe Leu Asn
                405                 410                 415

Lys Leu Pro Gln Pro Glu Met Glu Asp Gly Lys Pro Arg Asn Ile Glu
            420                 425                 430

Asp Tyr Phe Ile Thr Gln Gly Ala Ile Asn Thr Lys Ser Ile Gln Lys
        435                 440                 445

Glu Asp Ile Phe Ala Gln Ile Glu Asn Ala Tyr Glu Asp Ala Gln Val
    450                 455                 460

Phe Leu Gln Ile Lys Asp Thr Asp Asn Lys Leu Ser Gln Asn Lys Thr
465                 470                 475                 480

Ala Val Glu Lys Ile Lys Thr Leu Leu Asp Ala Leu Lys Glu Leu Gln
                485                 490                 495

His Phe Ile Lys Pro Leu Leu Gly Ser Gly Glu Glu Asn Glu Lys Asp
            500                 505                 510

Glu Leu Phe Tyr Gly Ser Phe Leu Ala Ile Trp Asp Glu Leu Asp Thr
        515                 520                 525

Ile Thr Pro Leu Tyr Asn Lys Val Arg Asn Trp Leu Thr Arg Lys Pro
    530                 535                 540

Tyr Ser Thr Glu Lys Ile Lys Leu Asn Phe Asp Asn Ala Gln Leu Leu
545                 550                 555                 560

Gly Gly Trp Asp Val Asn Lys Glu His Asp Cys Ala Gly Ile Leu Leu
                565                 570                 575

Arg Lys Asn Asp Ser Tyr Tyr Leu Gly Ile Ile Asn Lys Lys Thr Asn
            580                 585                 590
```

```
His Ile Phe Asp Thr Asp Ile Thr Pro Ser Asp Gly Glu Cys Tyr Asp
        595                 600                 605

Lys Ile Asp Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys
    610                 615                 620

Val Phe Phe Ser Lys Ser Arg Ile Lys Glu Phe Glu Pro Ser Glu Ala
625                 630                 635                 640

Ile Ile Asn Cys Tyr Lys Lys Gly Thr His Lys Lys Gly Lys Asn Phe
                645                 650                 655

Asn Leu Thr Asp Cys His Arg Leu Ile Asn Phe Phe Lys Thr Ser Ile
            660                 665                 670

Glu Lys His Glu Asp Trp Ser Lys Phe Gly Phe Lys Phe Ser Asp Thr
        675                 680                 685

Glu Thr Tyr Glu Asp Ile Ser Gly Phe Tyr Arg Glu Val Glu Gln Gln
    690                 695                 700

Gly Tyr Arg Leu Thr Ser His Pro Val Ser Ala Ser Tyr Ile His Ser
705                 710                 715                 720

Leu Val Lys Glu Gly Lys Leu Tyr Leu Phe Gln Ile Trp Asn Lys Asp
                725                 730                 735

Phe Ser Gln Phe Ser Lys Gly Thr Pro Asn Leu His Thr Leu Tyr Trp
            740                 745                 750

Lys Met Leu Phe Asp Lys Arg Asn Leu Ser Asp Val Val Tyr Lys Leu
        755                 760                 765

Asn Gly Gln Ala Glu Val Phe Tyr Arg Lys Ser Ser Ile Glu His Gln
    770                 775                 780

Asn Arg Ile Ile His Pro Ala Gln His Pro Ile Thr Asn Lys Asn Glu
785                 790                 795                 800

Leu Asn Lys Lys His Thr Ser Thr Phe Lys Tyr Asp Ile Ile Lys Asp
                805                 810                 815

Arg Arg Tyr Thr Val Asp Lys Phe Gln Phe His Val Pro Ile Thr Ile
            820                 825                 830

Asn Phe Lys Ala Thr Gly Gln Asn Asn Ile Asn Pro Ile Val Gln Glu
        835                 840                 845

Val Ile Arg Gln Asn Gly Ile Thr His Ile Ile Gly Ile Asp Arg Gly
    850                 855                 860

Glu Arg His Leu Leu Tyr Leu Ser Leu Ile Asp Leu Lys Gly Asn Ile
865                 870                 875                 880

Ile Lys Gln Met Thr Leu Asn Glu Ile Ile Asn Glu Tyr Lys Gly Val
                885                 890                 895

Thr Tyr Lys Thr Asn Tyr His Asn Leu Leu Glu Lys Arg Glu Lys Glu
            900                 905                 910

Arg Thr Glu Ala Arg His Ser Trp Ser Ser Ile Glu Ser Ile Lys Glu
        915                 920                 925
```

```
Leu Lys Asp Gly Tyr Met Ser Gln Val Ile His Lys Ile Thr Asp Met
    930                 935                 940

Met Val Lys Tyr Asn Ala Ile Val Val Leu Glu Asp Leu Asn Gly Gly
945                 950                 955                 960

Phe Met Arg Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe
                965                 970                 975

Glu Lys Lys Leu Ile Asp Lys Leu Asn Tyr Leu Val Asp Lys Lys Leu
            980                 985                 990

Asp Ala Asn Glu Val Gly Gly Val Leu Asn Ala Tyr Gln Leu Thr Asn
        995                 1000                1005

Lys Phe  Glu Ser Phe Lys Lys  Ile Gly Lys Gln Ser  Gly Phe Leu
    1010                1015                1020

Phe Tyr  Ile Pro Ala Trp Asn  Thr Ser Lys Ile Asp  Pro Ile Thr
    1025                1030                1035

Gly Phe  Val Asn Leu Phe Asn  Thr Arg Tyr Glu Ser  Ile Lys Glu
    1040                1045                1050

Thr Lys  Val Phe Trp Ser Lys  Phe Asp Ile Ile Arg  Tyr Asn Lys
    1055                1060                1065

Glu Lys  Asn Trp Phe Glu Phe  Val Phe Asp Tyr Asn  Thr Phe Thr
    1070                1075                1080

Thr Lys  Ala Glu Gly Thr Arg  Thr Lys Trp Thr Leu  Cys Thr His
    1085                1090                1095

Gly Thr  Arg Ile Gln Thr Phe  Arg Asn Pro Glu Lys  Asn Ala Gln
    1100                1105                1110

Trp Asp  Asn Lys Glu Ile Asn  Leu Thr Glu Ser Phe  Lys Ala Leu
    1115                1120                1125

Phe Glu  Lys Tyr Lys Ile Asp  Ile Thr Ser Asn Leu  Lys Glu Ser
    1130                1135                1140

Ile Met  Gln Glu Thr Glu Lys  Lys Phe Phe Gln Glu  Leu His Asn
    1145                1150                1155

Leu Leu  His Leu Thr Leu Gln  Met Arg Asn Ser Val  Thr Gly Thr
    1160                1165                1170

Asp Ile  Asp Tyr Leu Ile Ser  Pro Val Ala Asp Glu  Asp Gly Asn
    1175                1180                1185

Phe Tyr  Asp Ser Arg Ile Asn  Gly Lys Asn Phe Pro  Glu Asn Ala
    1190                1195                1200

Asp Ala  Asn Gly Ala Tyr Asn  Ile Ala Arg Lys Gly  Leu Met Leu
    1205                1210                1215

Ile Arg  Gln Ile Lys Gln Ala  Asp Pro Gln Lys Lys  Phe Lys Phe
    1220                1225                1230

Glu Thr  Ile Thr Asn Lys Asp  Trp Leu Lys Phe Ala  Gln Asp Lys
    1235                1240                1245
```

Pro Tyr Leu Lys Asp
1250

<210> 98
<211> 1250
<212> PRT
<213> 密斯氏菌屬某種（Smithella sp.）

<400> 98
Met Gln Thr Leu Phe Glu Asn Phe Thr Asn Gln Tyr Pro Val Ser Lys
1 5 10 15

Thr Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Lys Asp Phe Ile
20 25 30

Glu Gln Lys Gly Leu Leu Lys Lys Asp Glu Asp Arg Ala Glu Lys Tyr
35 40 45

Lys Lys Val Lys Asn Ile Ile Asp Glu Tyr His Lys Asp Phe Ile Glu
50 55 60

Lys Ser Leu Asn Gly Leu Lys Leu Asp Gly Leu Glu Lys Tyr Lys Thr
65 70 75 80

Leu Tyr Leu Lys Gln Glu Lys Asp Asp Lys Asp Lys Lys Ala Phe Asp
85 90 95

Lys Glu Lys Glu Asn Leu Arg Lys Gln Ile Ala Asn Ala Phe Arg Asn
100 105 110

Asn Glu Lys Phe Lys Thr Leu Phe Ala Lys Glu Leu Ile Lys Asn Asp
115 120 125

Leu Met Ser Phe Ala Cys Glu Glu Asp Lys Lys Asn Val Lys Glu Phe
130 135 140

Glu Ala Phe Thr Thr Tyr Phe Thr Gly Phe His Gln Asn Arg Ala Asn
145 150 155 160

Met Tyr Val Ala Asp Glu Lys Arg Thr Ala Ile Ala Ser Arg Leu Ile
165 170 175

His Glu Asn Leu Pro Lys Phe Ile Asp Asn Ile Lys Ile Phe Glu Lys
180 185 190

Met Lys Lys Glu Ala Pro Glu Leu Leu Ser Pro Phe Asn Gln Thr Leu
195 200 205

Lys Asp Met Lys Asp Val Ile Lys Gly Thr Thr Leu Glu Glu Ile Phe
210 215 220

Ser Leu Asp Tyr Phe Asn Lys Thr Leu Thr Gln Ser Gly Ile Asp Ile
225 230 235 240

Tyr Asn Ser Val Ile Gly Gly Arg Thr Pro Glu Glu Gly Lys Thr Lys
245 250 255

Ile Lys Gly Leu Asn Glu Tyr Ile Asn Thr Asp Phe Asn Gln Lys Gln
260 265 270

Thr Asp Lys Lys Lys Arg Gln Pro Lys Phe Lys Gln Leu Tyr Lys Gln
275 280 285

```
Ile Leu Ser Asp Arg Gln Ser Leu Ser Phe Ile Ala Glu Ala Phe Lys
    290                 295                 300

Asn Asp Thr Glu Ile Leu Glu Ala Ile Glu Lys Phe Tyr Val Asn Glu
305                 310                 315                 320

Leu Leu His Phe Ser Asn Glu Gly Lys Ser Thr Asn Val Leu Asp Ala
                325                 330                 335

Ile Lys Asn Ala Val Ser Asn Leu Glu Ser Phe Asn Leu Thr Lys Met
            340                 345                 350

Tyr Phe Arg Ser Gly Ala Ser Leu Thr Asp Val Ser Arg Lys Val Phe
        355                 360                 365

Gly Glu Trp Ser Ile Ile Asn Arg Ala Leu Asp Asn Tyr Tyr Ala Thr
    370                 375                 380

Thr Tyr Pro Ile Lys Pro Arg Glu Lys Ser Glu Lys Tyr Glu Glu Arg
385                 390                 395                 400

Lys Glu Lys Trp Leu Lys Gln Asp Phe Asn Val Ser Leu Ile Gln Thr
                405                 410                 415

Ala Ile Asp Glu Tyr Asp Asn Glu Thr Val Lys Gly Lys Asn Ser Gly
            420                 425                 430

Lys Val Ile Ala Asp Tyr Phe Ala Lys Phe Cys Asp Asp Lys Glu Thr
        435                 440                 445

Asp Leu Ile Gln Lys Val Asn Glu Gly Tyr Ile Ala Val Lys Asp Leu
    450                 455                 460

Leu Asn Thr Pro Cys Pro Glu Asn Glu Lys Leu Gly Ser Asn Lys Asp
465                 470                 475                 480

Gln Val Lys Gln Ile Lys Ala Phe Met Asp Ser Ile Met Asp Ile Met
                485                 490                 495

His Phe Val Arg Pro Leu Ser Leu Lys Asp Thr Asp Lys Glu Lys Asp
            500                 505                 510

Glu Thr Phe Tyr Ser Leu Phe Thr Pro Leu Tyr Asp His Leu Thr Gln
        515                 520                 525

Thr Ile Ala Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Gln Lys Pro
    530                 535                 540

Tyr Ser Thr Glu Lys Ile Lys Leu Asn Phe Glu Asn Ser Thr Leu Leu
545                 550                 555                 560

Gly Gly Trp Asp Leu Asn Lys Glu Thr Asp Asn Thr Ala Ile Ile Leu
                565                 570                 575

Arg Lys Asp Asn Leu Tyr Tyr Leu Gly Ile Met Asp Lys Arg His Asn
            580                 585                 590

Arg Ile Phe Arg Asn Val Pro Lys Ala Asp Lys Lys Asp Phe Cys Tyr
        595                 600                 605

Glu Lys Met Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro
    610                 615                 620
```

```
Lys Val Phe Phe Ser Gln Ser Arg Ile Gln Glu Phe Thr Pro Ser Ala
625                 630                 635                 640

Lys Leu Leu Glu Asn Tyr Ala Asn Glu Thr His Lys Lys Gly Asp Asn
                645                 650                 655

Phe Asn Leu Asn His Cys His Lys Leu Ile Asp Phe Phe Lys Asp Ser
            660                 665                 670

Ile Asn Lys His Glu Asp Trp Lys Asn Phe Asp Phe Arg Phe Ser Ala
        675                 680                 685

Thr Ser Thr Tyr Ala Asp Leu Ser Gly Phe Tyr His Glu Val Glu His
    690                 695                 700

Gln Gly Tyr Lys Ile Ser Phe Gln Ser Val Ala Asp Ser Phe Ile Asp
705                 710                 715                 720

Asp Leu Val Asn Glu Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys
                725                 730                 735

Asp Phe Ser Pro Phe Ser Lys Gly Lys Pro Asn Leu His Thr Leu Tyr
            740                 745                 750

Trp Lys Met Leu Phe Asp Glu Asn Asn Leu Lys Asp Val Val Tyr Lys
        755                 760                 765

Leu Asn Gly Glu Ala Glu Val Phe Tyr Arg Lys Lys Ser Ile Ala Glu
    770                 775                 780

Lys Asn Thr Thr Ile His Lys Ala Asn Glu Ser Ile Ile Asn Lys Asn
785                 790                 795                 800

Pro Asp Asn Pro Lys Ala Thr Ser Thr Phe Asn Tyr Asp Ile Val Lys
                805                 810                 815

Asp Lys Arg Tyr Thr Ile Asp Lys Phe Gln Phe His Ile Pro Ile Thr
            820                 825                 830

Met Asn Phe Lys Ala Glu Gly Ile Phe Asn Met Asn Gln Arg Val Asn
        835                 840                 845

Gln Phe Leu Lys Ala Asn Pro Asp Ile Asn Ile Ile Gly Ile Asp Arg
    850                 855                 860

Gly Glu Arg His Leu Leu Tyr Tyr Ala Leu Ile Asn Gln Lys Gly Lys
865                 870                 875                 880

Ile Leu Lys Gln Asp Thr Leu Asn Val Ile Ala Asn Glu Lys Gln Lys
                885                 890                 895

Val Asp Tyr His Asn Leu Leu Asp Lys Lys Glu Gly Asp Arg Ala Thr
            900                 905                 910

Ala Arg Gln Glu Trp Gly Val Ile Glu Thr Ile Lys Glu Leu Lys Glu
        915                 920                 925

Gly Tyr Leu Ser Gln Val Ile His Lys Leu Thr Asp Leu Met Ile Glu
    930                 935                 940

Asn Asn Ala Ile Ile Val Met Glu Asp Leu Asn Phe Gly Phe Lys Arg
945                 950                 955                 960
```

```
Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu Lys Met
                965                 970                 975

Leu Ile Asp Lys Leu Asn Tyr Leu Val Asp Lys Asn Lys Lys Ala Asn
            980                 985                 990

Glu Leu Gly Gly Leu Leu Asn Ala Phe Gln Leu Ala Asn Lys Phe Glu
        995                1000                1005

Ser Phe  Gln Lys Met Gly Lys  Gln Asn Gly Phe Ile  Phe Tyr Val
    1010                1015                1020

Pro Ala  Trp Asn Thr Ser Lys  Thr Asp Pro Ala Thr  Gly Phe Ile
    1025                1030                1035

Asp Phe  Leu Lys Pro Arg Tyr  Glu Asn Leu Asn Gln  Ala Lys Asp
    1040                1045                1050

Phe Phe  Glu Lys Phe Asp Ser  Ile Arg Leu Asn Ser  Lys Ala Asp
    1055                1060                1065

Tyr Phe  Glu Phe Ala Phe Asp  Phe Lys Asn Phe Thr  Glu Lys Ala
    1070                1075                1080

Asp Gly  Gly Arg Thr Lys Trp  Thr Val Cys Thr Thr  Asn Glu Asp
    1085                1090                1095

Arg Tyr  Ala Trp Asn Arg Ala  Leu Asn Asn Asn Arg  Gly Ser Gln
    1100                1105                1110

Glu Lys  Tyr Asp Ile Thr Ala  Glu Leu Lys Ser Leu  Phe Asp Gly
    1115                1120                1125

Lys Val  Asp Tyr Lys Ser Gly  Lys Asp Leu Lys Gln  Gln Ile Ala
    1130                1135                1140

Ser Gln  Glu Ser Ala Asp Phe  Phe Lys Ala Leu Met  Lys Asn Leu
    1145                1150                1155

Ser Ile  Thr Leu Ser Leu Arg  His Asn Asn Gly Glu  Lys Gly Asp
    1160                1165                1170

Asn Glu  Gln Asp Tyr Ile Leu  Ser Pro Val Ala Asp  Ser Lys Gly
    1175                1180                1185

Arg Phe  Phe Asp Ser Arg Lys  Ala Asp Asp Asp Met  Pro Lys Asn
    1190                1195                1200

Ala Asp  Ala Asn Gly Ala Tyr  His Ile Ala Leu Lys  Gly Leu Trp
    1205                1210                1215

Cys Leu  Glu Gln Ile Ser Lys  Thr Asp Asp Leu Lys  Lys Val Lys
    1220                1225                1230

Leu Ala  Ile Ser Asn Lys Glu  Trp Leu Glu Phe Val  Gln Thr Leu
    1235                1240                1245

Lys Gly
    1250
```

<210> 99
<211> 1250
<212> PRT
<213> 密斯氏菌屬某種

<400> 99

```
Met Gln Thr Leu Phe Glu Asn Phe Thr Asn Gln Tyr Pro Val Ser Lys
1               5                  10                  15

Thr Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Lys Asp Phe Ile
            20                  25                  30

Glu Gln Lys Gly Leu Leu Lys Lys Asp Glu Asp Arg Ala Glu Lys Tyr
        35                  40                  45

Lys Lys Val Lys Asn Ile Ile Asp Glu Tyr His Lys Asp Phe Ile Glu
    50                  55                  60

Lys Ser Leu Asn Gly Leu Lys Leu Asp Gly Leu Glu Glu Tyr Lys Thr
65                  70                  75                  80

Leu Tyr Leu Lys Gln Glu Lys Asp Asp Lys Asp Lys Lys Ala Phe Asp
                85                  90                  95

Lys Glu Lys Glu Asn Leu Arg Lys Gln Ile Ala Asn Ala Phe Arg Asn
            100                 105                 110

Asn Glu Lys Phe Lys Thr Leu Phe Ala Lys Glu Leu Ile Lys Asn Asp
        115                 120                 125

Leu Met Ser Phe Ala Cys Glu Glu Asp Lys Lys Asn Val Lys Glu Phe
    130                 135                 140

Glu Ala Phe Thr Thr Tyr Phe Thr Gly Phe His Gln Asn Arg Ala Asn
145                 150                 155                 160

Met Tyr Val Ala Asp Glu Lys Arg Thr Ala Ile Ala Ser Arg Leu Ile
                165                 170                 175

His Glu Asn Leu Pro Lys Phe Ile Asp Asn Ile Lys Ile Phe Glu Lys
            180                 185                 190

Met Lys Lys Glu Ala Pro Glu Leu Leu Ser Pro Phe Asn Gln Thr Leu
        195                 200                 205

Lys Asp Met Lys Asp Val Ile Lys Gly Thr Thr Leu Glu Glu Ile Phe
    210                 215                 220

Ser Leu Asp Tyr Phe Asn Lys Thr Leu Thr Gln Ser Gly Ile Asp Ile
225                 230                 235                 240

Tyr Asn Ser Val Ile Gly Gly Arg Thr Pro Glu Glu Gly Lys Thr Lys
                245                 250                 255

Ile Lys Gly Leu Asn Glu Tyr Ile Asn Thr Asp Phe Asn Gln Lys Gln
            260                 265                 270

Thr Asp Lys Lys Lys Arg Gln Pro Lys Phe Lys Gln Leu Tyr Lys Gln
        275                 280                 285

Ile Leu Ser Asp Arg Gln Ser Leu Ser Phe Ile Ala Glu Ala Phe Lys
    290                 295                 300

Asn Asp Thr Glu Ile Leu Glu Ala Ile Glu Lys Phe Tyr Val Asn Glu
305                 310                 315                 320

Leu Leu His Phe Ser Asn Glu Gly Lys Ser Thr Asn Val Leu Asp Ala
                325                 330                 335
```

Ile Lys Asn Ala Val Ser Asn Leu Glu Ser Phe Asn Leu Thr Lys Ile
340 345 350

Tyr Phe Arg Ser Gly Thr Ser Leu Thr Asp Val Ser Arg Lys Val Phe
355 360 365

Gly Glu Trp Ser Ile Ile Asn Arg Ala Leu Asp Asn Tyr Tyr Ala Thr
370 375 380

Thr Tyr Pro Ile Lys Pro Arg Glu Lys Ser Glu Lys Tyr Glu Glu Arg
385 390 395 400

Lys Glu Lys Trp Leu Lys Gln Asp Phe Asn Val Ser Leu Ile Gln Thr
405 410 415

Ala Ile Asp Glu Tyr Asp Asn Glu Thr Val Lys Gly Lys Asn Ser Gly
420 425 430

Lys Val Ile Val Asp Tyr Phe Ala Lys Phe Cys Asp Asp Lys Glu Thr
435 440 445

Asp Leu Ile Gln Lys Val Asn Glu Gly Tyr Ile Ala Val Lys Asp Leu
450 455 460

Leu Asn Thr Pro Tyr Pro Glu Asn Glu Lys Leu Gly Ser Asn Lys Asp
465 470 475 480

Gln Val Lys Gln Ile Lys Ala Phe Met Asp Ser Ile Met Asp Ile Met
485 490 495

His Phe Val Arg Pro Leu Ser Leu Lys Asp Thr Asp Lys Glu Lys Asp
500 505 510

Glu Thr Phe Tyr Ser Leu Phe Thr Pro Leu Tyr Asp His Leu Thr Gln
515 520 525

Thr Ile Ala Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Gln Lys Pro
530 535 540

Tyr Ser Thr Glu Lys Ile Lys Leu Asn Phe Glu Asn Ser Thr Leu Leu
545 550 555 560

Gly Gly Trp Asp Leu Asn Lys Glu Thr Asp Asn Thr Ala Ile Ile Leu
565 570 575

Arg Lys Glu Asn Leu Tyr Tyr Leu Gly Ile Met Asp Lys Arg His Asn
580 585 590

Arg Ile Phe Arg Asn Val Pro Lys Ala Asp Lys Lys Asp Ser Cys Tyr
595 600 605

Glu Lys Met Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro
610 615 620

Lys Val Phe Phe Ser Gln Ser Arg Ile Gln Glu Phe Thr Pro Ser Ala
625 630 635 640

Lys Leu Leu Glu Asn Tyr Glu Asn Glu Thr His Lys Lys Gly Asp Asn
645 650 655

Phe Asn Leu Asn His Cys His Gln Leu Ile Asp Phe Phe Lys Asp Ser
660 665 670

```
Ile Asn Lys His Glu Asp Trp Lys Asn Phe Asp Phe Arg Phe Ser Ala
        675                 680                 685

Thr Ser Thr Tyr Ala Asp Leu Ser Gly Phe Tyr His Glu Val Glu His
    690                 695                 700

Gln Gly Tyr Lys Ile Ser Phe Gln Ser Ile Ala Asp Ser Phe Ile Asp
705                 710                 715                 720

Asp Leu Val Asn Glu Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys
                725                 730                 735

Asp Phe Ser Pro Phe Ser Lys Gly Lys Pro Asn Leu His Thr Leu Tyr
            740                 745                 750

Trp Lys Met Leu Phe Asp Glu Asn Asn Leu Lys Asp Val Val Tyr Lys
        755                 760                 765

Leu Asn Gly Glu Ala Glu Val Phe Tyr Arg Lys Lys Ser Ile Ala Glu
    770                 775                 780

Lys Asn Thr Thr Ile His Lys Ala Asn Glu Ser Ile Ile Asn Lys Asn
785                 790                 795                 800

Pro Asp Asn Pro Lys Ala Thr Ser Thr Phe Asn Tyr Asp Ile Val Lys
                805                 810                 815

Asp Lys Arg Tyr Thr Ile Asp Lys Phe Gln Phe His Val Pro Ile Thr
            820                 825                 830

Met Asn Phe Lys Ala Glu Gly Ile Phe Asn Met Asn Gln Arg Val Asn
        835                 840                 845

Gln Phe Leu Lys Ala Asn Pro Asp Ile Asn Ile Ile Gly Ile Asp Arg
    850                 855                 860

Gly Glu Arg His Leu Leu Tyr Tyr Thr Leu Ile Asn Gln Lys Gly Lys
865                 870                 875                 880

Ile Leu Lys Gln Asp Thr Leu Asn Val Ile Ala Asn Glu Lys Gln Lys
                885                 890                 895

Val Asp Tyr His Asn Leu Leu Asp Lys Lys Glu Gly Asp Arg Ala Thr
            900                 905                 910

Ala Arg Gln Glu Trp Gly Val Ile Glu Thr Ile Lys Glu Leu Lys Glu
        915                 920                 925

Gly Tyr Leu Ser Gln Val Ile His Lys Leu Thr Asp Leu Met Ile Glu
    930                 935                 940

Asn Asn Ala Ile Ile Val Met Glu Asp Leu Asn Phe Gly Phe Lys Arg
945                 950                 955                 960

Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu Lys Met
                965                 970                 975

Leu Ile Asp Lys Leu Asn Tyr Leu Val Asp Lys Asn Lys Lys Ala Asn
            980                 985                 990

Glu Leu Gly Gly Leu Leu Asn Ala Phe Gln Leu Ala Asn Lys Phe Glu
        995                 1000                1005
```

Ser Phe Gln Lys Met Gly Lys Gln Asn Gly Phe Ile Phe Tyr Val
1010 1015 1020

Pro Ala Trp Asn Thr Ser Lys Thr Asp Pro Ala Thr Gly Phe Ile
1025 1030 1035

Asp Phe Leu Lys Pro Arg Tyr Glu Asn Leu Lys Gln Ala Lys Asp
1040 1045 1050

Phe Phe Glu Lys Phe Asp Ser Ile Arg Leu Asn Ser Lys Ala Asp
1055 1060 1065

Tyr Phe Glu Phe Ala Phe Asp Phe Lys Asn Phe Thr Gly Lys Ala
1070 1075 1080

Asp Gly Gly Arg Thr Lys Trp Thr Val Cys Thr Thr Asn Glu Asp
1085 1090 1095

Arg Tyr Ala Trp Asn Arg Ala Leu Asn Asn Asn Arg Gly Ser Gln
1100 1105 1110

Glu Lys Tyr Asp Ile Thr Ala Glu Leu Lys Ser Leu Phe Asp Gly
1115 1120 1125

Lys Val Asp Tyr Lys Ser Gly Lys Asp Leu Lys Gln Gln Ile Ala
1130 1135 1140

Ser Gln Glu Leu Ala Asp Phe Phe Arg Thr Leu Met Lys Tyr Leu
1145 1150 1155

Ser Val Thr Leu Ser Leu Arg His Asn Asn Gly Glu Lys Gly Glu
1160 1165 1170

Thr Glu Gln Asp Tyr Ile Leu Ser Pro Val Ala Asp Ser Met Gly
1175 1180 1185

Lys Phe Phe Asp Ser Arg Lys Ala Gly Asp Asp Met Pro Lys Asn
1190 1195 1200

Ala Asp Ala Asn Gly Ala Tyr His Ile Ala Leu Lys Gly Leu Trp
1205 1210 1215

Cys Leu Glu Gln Ile Ser Lys Thr Asp Asp Leu Lys Lys Val Lys
1220 1225 1230

Leu Ala Ile Ser Asn Lys Glu Trp Leu Glu Phe Met Gln Thr Leu
1235 1240 1245

Lys Gly
1250

<210> 100
<211> 1247
<212> PRT
<213> 布氏普雷沃菌

<400> 100
Met Lys Phe Thr Asp Phe Thr Gly Leu Tyr Ser Leu Ser Lys Thr Leu
1 5 10 15

Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Leu Glu Asn Ile Lys Lys
20 25 30

```
Ala Gly Leu Leu Glu Gln Asp Gln His Arg Ala Asp Ser Tyr Lys Lys
        35                  40                  45

Val Lys Lys Ile Ile Asp Glu Tyr His Lys Ala Phe Ile Glu Lys Ser
    50                  55                  60

Leu Ser Asn Phe Glu Leu Lys Tyr Gln Ser Glu Asp Lys Leu Asp Ser
65                  70                  75                  80

Leu Glu Glu Tyr Leu Met Tyr Tyr Ser Met Lys Arg Ile Glu Lys Thr
                85                  90                  95

Glu Lys Asp Lys Phe Ala Lys Ile Gln Asp Asn Leu Arg Lys Gln Ile
            100                 105                 110

Ala Asp His Leu Lys Gly Asp Glu Ser Tyr Lys Thr Ile Phe Ser Lys
        115                 120                 125

Asp Leu Ile Arg Lys Asn Leu Pro Asp Phe Val Lys Ser Asp Glu Glu
    130                 135                 140

Arg Thr Leu Ile Lys Glu Phe Lys Asp Phe Thr Thr Tyr Phe Lys Gly
145                 150                 155                 160

Phe Tyr Glu Asn Arg Glu Asn Met Tyr Ser Ala Glu Asp Lys Ser Thr
                165                 170                 175

Ala Ile Ser His Arg Ile Ile His Glu Asn Leu Pro Lys Phe Val Asp
            180                 185                 190

Asn Ile Asn Ala Phe Ser Lys Ile Ile Leu Ile Pro Glu Leu Arg Glu
        195                 200                 205

Lys Leu Asn Gln Ile Tyr Gln Asp Phe Glu Glu Tyr Leu Asn Val Glu
    210                 215                 220

Ser Ile Asp Glu Ile Phe His Leu Asp Tyr Phe Ser Met Val Met Thr
225                 230                 235                 240

Gln Lys Gln Ile Glu Val Tyr Asn Ala Ile Ile Gly Gly Lys Ser Thr
                245                 250                 255

Asn Asp Lys Lys Ile Gln Gly Leu Asn Glu Tyr Ile Asn Leu Tyr Asn
            260                 265                 270

Gln Lys His Lys Asp Cys Lys Leu Pro Lys Leu Lys Leu Leu Phe Lys
        275                 280                 285

Gln Ile Leu Ser Asp Arg Ile Ala Ile Ser Trp Leu Pro Asp Asn Phe
    290                 295                 300

Lys Asp Asp Gln Glu Ala Leu Asp Ser Ile Asp Thr Cys Tyr Lys Asn
305                 310                 315                 320

Leu Leu Asn Asp Gly Asn Val Leu Gly Glu Gly Asn Leu Lys Leu Leu
                325                 330                 335

Leu Glu Asn Ile Asp Thr Tyr Asn Leu Lys Gly Ile Phe Ile Arg Asn
            340                 345                 350

Asp Leu Gln Leu Thr Asp Ile Ser Gln Lys Met Tyr Ala Ser Trp Asn
        355                 360                 365
```

```
Val Ile Gln Asp Ala Val Ile Leu Asp Leu Lys Lys Gln Val Ser Arg
    370                 375                 380

Lys Lys Lys Glu Ser Ala Glu Asp Tyr Asn Asp Arg Leu Lys Lys Leu
385                 390                 395                 400

Tyr Thr Ser Gln Glu Ser Phe Ser Ile Gln Tyr Leu Asn Asp Cys Leu
                405                 410                 415

Arg Ala Tyr Gly Lys Thr Glu Asn Ile Gln Asp Tyr Phe Ala Lys Leu
            420                 425                 430

Gly Ala Val Asn Asn Glu His Glu Gln Thr Ile Asn Leu Phe Ala Gln
        435                 440                 445

Val Arg Asn Ala Tyr Thr Ser Val Gln Ala Ile Leu Thr Thr Pro Tyr
    450                 455                 460

Pro Glu Asn Ala Asn Leu Ala Gln Asp Lys Glu Thr Val Ala Leu Ile
465                 470                 475                 480

Lys Asn Leu Leu Asp Ser Leu Lys Arg Leu Gln Arg Phe Ile Lys Pro
                485                 490                 495

Leu Leu Gly Lys Gly Asp Glu Ser Asp Lys Asp Glu Arg Phe Tyr Gly
            500                 505                 510

Asp Phe Thr Pro Leu Trp Glu Thr Leu Asn Gln Ile Thr Pro Leu Tyr
        515                 520                 525

Asn Met Val Arg Asn Tyr Met Thr Arg Lys Pro Tyr Ser Gln Glu Lys
    530                 535                 540

Ile Lys Leu Asn Phe Glu Asn Ser Thr Leu Leu Gly Gly Trp Asp Leu
545                 550                 555                 560

Asn Lys Glu His Asp Asn Thr Ala Ile Ile Leu Arg Lys Asn Gly Leu
                565                 570                 575

Tyr Tyr Leu Ala Ile Met Lys Lys Ser Ala Asn Lys Ile Phe Asp Lys
            580                 585                 590

Asp Lys Leu Asp Asn Ser Gly Asp Cys Tyr Glu Lys Met Val Tyr Lys
        595                 600                 605

Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val Phe Phe Ser Lys
    610                 615                 620

Ser Arg Ile Asp Glu Phe Lys Pro Ser Glu Asn Ile Ile Glu Asn Tyr
625                 630                 635                 640

Lys Lys Gly Thr His Lys Lys Gly Ala Asn Phe Asn Leu Ala Asp Cys
                645                 650                 655

His Asn Leu Ile Asp Phe Phe Lys Ser Ser Ile Ser Lys His Glu Asp
            660                 665                 670

Trp Ser Lys Phe Asn Phe His Phe Ser Asp Thr Ser Ser Tyr Glu Asp
        675                 680                 685

Leu Ser Asp Phe Tyr Arg Glu Val Glu Gln Gln Gly Tyr Ser Ile Ser
    690                 695                 700
```

```
Phe Cys Asp Val Ser Val Glu Tyr Ile Asn Lys Met Val Glu Lys Gly
705                 710                 715                 720

Asp Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Glu Phe Ser
                725                 730                 735

Lys Gly Thr Pro Asn Met His Thr Leu Tyr Trp Asn Ser Leu Phe Ser
            740                 745                 750

Lys Glu Asn Leu Asn Asn Ile Ile Tyr Lys Leu Asn Gly Gln Ala Glu
        755                 760                 765

Ile Phe Phe Arg Lys Lys Ser Leu Asn Tyr Lys Arg Pro Thr His Pro
    770                 775                 780

Ala His Gln Ala Ile Lys Asn Lys Asn Lys Cys Asn Glu Lys Lys Glu
785                 790                 795                 800

Ser Ile Phe Asp Tyr Asp Leu Val Lys Asp Lys Arg Tyr Thr Val Asp
                805                 810                 815

Lys Phe Gln Phe His Val Pro Ile Thr Met Asn Phe Lys Ser Thr Gly
            820                 825                 830

Asn Thr Asn Ile Asn Gln Gln Val Ile Asp Tyr Leu Arg Thr Glu Asp
        835                 840                 845

Asp Thr His Ile Ile Gly Ile Asp Arg Gly Glu Arg His Leu Leu Tyr
    850                 855                 860

Leu Val Val Ile Asp Ser His Gly Lys Ile Val Glu Gln Phe Thr Leu
865                 870                 875                 880

Asn Glu Ile Val Asn Glu Tyr Gly Gly Asn Ile Tyr Arg Thr Asn Tyr
                885                 890                 895

His Asp Leu Leu Asp Thr Arg Glu Gln Asn Arg Glu Lys Ala Arg Glu
            900                 905                 910

Ser Trp Gln Thr Ile Glu Asn Ile Lys Glu Leu Lys Glu Gly Tyr Ile
        915                 920                 925

Ser Gln Val Ile His Lys Ile Thr Asp Leu Met Gln Lys Tyr His Ala
    930                 935                 940

Val Val Val Leu Glu Asp Leu Asn Met Gly Phe Met Arg Gly Arg Gln
945                 950                 955                 960

Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu Glu Met Leu Ile Asn
                965                 970                 975

Lys Leu Asn Tyr Leu Val Asn Lys Lys Ala Asp Gln Asn Ser Ala Gly
            980                 985                 990

Gly Leu Leu His Ala Tyr Gln Leu Thr Ser Lys Phe Glu Ser Phe Gln
        995                 1000                1005

Lys Leu Gly Lys Gln Ser Gly Phe Leu Phe Tyr Ile Pro Ala Trp
    1010                1015                1020

Asn Thr Ser Lys Ile Asp Pro Val Thr Gly Phe Val Asn Leu Phe
    1025                1030                1035
```

Asp Thr Arg Tyr Glu Ser Ile Asp Lys Ala Lys Ala Phe Phe Gly
1040 1045 1050

Lys Phe Asp Ser Ile Arg Tyr Asn Ala Asp Lys Asp Trp Phe Glu
1055 1060 1065

Phe Ala Phe Asp Tyr Asn Asn Phe Thr Thr Lys Ala Glu Gly Thr
1070 1075 1080

Arg Thr Asn Trp Thr Ile Cys Thr Tyr Gly Ser Arg Ile Arg Thr
1085 1090 1095

Phe Arg Asn Gln Ala Lys Asn Ser Gln Trp Asp Asn Glu Glu Ile
1100 1105 1110

Asp Leu Thr Lys Ala Tyr Lys Ala Phe Phe Ala Lys His Gly Ile
1115 1120 1125

Asn Ile Tyr Asp Asn Ile Lys Glu Ala Ile Ala Met Glu Thr Glu
1130 1135 1140

Lys Ser Phe Phe Glu Asp Leu Leu His Leu Leu Lys Leu Thr Leu
1145 1150 1155

Gln Met Arg Asn Ser Ile Thr Gly Thr Thr Thr Asp Tyr Leu Ile
1160 1165 1170

Ser Pro Val His Asp Ser Lys Gly Asn Phe Tyr Asp Ser Arg Ile
1175 1180 1185

Cys Asp Asn Ser Leu Pro Ala Asn Ala Asp Ala Asn Gly Ala Tyr
1190 1195 1200

Asn Ile Ala Arg Lys Gly Leu Met Leu Ile Gln Gln Ile Lys Asp
1205 1210 1215

Ser Thr Ser Ser Asn Arg Phe Lys Phe Ser Pro Ile Thr Asn Lys
1220 1225 1230

Asp Trp Leu Ile Phe Ala Gln Glu Lys Pro Tyr Leu Asn Asp
1235 1240 1245

<210> 101
<211> 1247
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
非培養的細菌序列

<400> 101
Met Phe Lys Gly Asp Ala Phe Thr Gly Leu Tyr Glu Val Gln Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Val Pro Ile Gly Leu Thr Gln Ser Tyr Leu Glu
20 25 30

Asn Asp Trp Val Ile Gln Lys Asp Lys Glu Val Glu Glu Asn Tyr Gly
35 40 45

Lys Ile Lys Ala Tyr Phe Asp Leu Ile His Lys Glu Phe Val Arg Gln
50 55 60

Ser Leu Glu Asn Ala Trp Leu Cys Gln Leu Asp Asp Phe Tyr Glu Lys
65 70 75 80

```
Tyr Ile Glu Leu His Asn Ser Leu Glu Thr Arg Lys Asp Lys Asn Leu
                85                  90                  95

Ala Lys Gln Phe Glu Lys Val Met Lys Ser Leu Lys Lys Glu Phe Val
            100                 105                 110

Ser Phe Phe Asp Ala Lys Trp Asn Glu Trp Lys Gln Lys Phe Ser Phe
        115                 120                 125

Leu Lys Lys Trp Trp Ile Asp Val Leu Asn Glu Lys Glu Val Leu Asp
    130                 135                 140

Leu Met Ala Glu Phe Tyr Pro Asp Glu Lys Glu Leu Phe Asp Lys Phe
145                 150                 155                 160

Asp Lys Phe Phe Thr Tyr Phe Ser Asn Phe Lys Glu Ser Arg Lys Asn
                165                 170                 175

Phe Tyr Ala Asp Asp Gly Arg Ala Trp Ala Ile Ala Thr Arg Ala Ile
            180                 185                 190

Asp Glu Asn Leu Ile Thr Phe Ile Lys Asn Ile Glu Asp Phe Lys Lys
        195                 200                 205

Leu Asn Ser Ser Phe Arg Glu Phe Val Asn Asp Asn Phe Ser Glu Glu
    210                 215                 220

Asp Lys Gln Ile Phe Glu Ile Asp Phe Tyr Asn Asn Cys Leu Leu Gln
225                 230                 235                 240

Pro Trp Ile Asp Lys Tyr Asn Lys Ile Val Trp Trp Tyr Ser Leu Glu
                245                 250                 255

Asn Trp Glu Lys Val Gln Trp Leu Asn Glu Lys Ile Asn Asn Phe Lys
            260                 265                 270

Gln Asn Gln Asn Lys Ser Asn Ser Lys Asp Leu Lys Phe Pro Arg Met
        275                 280                 285

Lys Leu Leu Tyr Lys Gln Ile Leu Gly Asp Lys Glu Lys Lys Val Tyr
    290                 295                 300

Ile Asp Glu Ile Arg Asp Asp Lys Asn Leu Ile Asp Leu Ile Asp Asn
305                 310                 315                 320

Ser Lys Arg Arg Asn Gln Ile Lys Ile Asp Asn Ala Asn Asp Ile Ile
                325                 330                 335

Asn Asp Phe Ile Asn Asn Asn Ala Lys Phe Glu Leu Asp Lys Ile Tyr
            340                 345                 350

Leu Thr Arg Gln Ser Ile Asn Thr Ile Ser Ser Lys Tyr Phe Ser Ser
        355                 360                 365

Trp Asp Tyr Ile Arg Trp Tyr Phe Trp Thr Gly Glu Leu Gln Glu Phe
    370                 375                 380

Val Ser Phe Tyr Asp Leu Lys Glu Thr Phe Trp Lys Ile Glu Tyr Glu
385                 390                 395                 400

Thr Leu Glu Asn Ile Phe Lys Asp Cys Tyr Val Lys Gly Ile Asn Thr
                405                 410                 415
```

```
Glu Ser Gln Asn Asn Ile Val Phe Glu Thr Gln Gly Ile Tyr Glu Asn
            420                 425                 430

Phe Leu Asn Ile Phe Lys Phe Glu Phe Asn Gln Asn Ile Ser Gln Ile
        435                 440                 445

Ser Leu Leu Glu Trp Glu Leu Asp Lys Ile Gln Asn Glu Asp Ile Lys
    450                 455                 460

Lys Asn Glu Lys Gln Val Glu Val Ile Lys Asn Tyr Phe Asp Ser Val
465                 470                 475                 480

Met Ser Val Tyr Lys Met Thr Lys Tyr Phe Ser Leu Glu Lys Trp Lys
                485                 490                 495

Lys Arg Val Glu Leu Asp Thr Asp Asn Asn Phe Tyr Asn Asp Phe Asn
            500                 505                 510

Glu Tyr Leu Glu Gly Phe Glu Ile Trp Lys Asp Tyr Asn Leu Val Arg
        515                 520                 525

Asn Tyr Ile Thr Lys Lys Gln Val Asn Thr Asp Lys Ile Lys Leu Asn
    530                 535                 540

Phe Asp Asn Ser Gln Phe Leu Thr Trp Trp Asp Lys Asp Lys Glu Asn
545                 550                 555                 560

Glu Arg Leu Gly Ile Ile Leu Arg Arg Glu Trp Lys Tyr Tyr Leu Trp
                565                 570                 575

Ile Leu Lys Lys Trp Asn Thr Leu Asn Phe Gly Asp Tyr Leu Gln Lys
            580                 585                 590

Glu Trp Glu Ile Phe Tyr Glu Lys Met Asn Tyr Lys Gln Leu Asn Asn
        595                 600                 605

Val Tyr Arg Gln Leu Pro Arg Leu Leu Phe Pro Leu Thr Lys Lys Leu
    610                 615                 620

Asn Glu Leu Lys Trp Asp Glu Leu Lys Lys Tyr Leu Ser Lys Tyr Ile
625                 630                 635                 640

Gln Asn Phe Trp Tyr Asn Glu Glu Ile Ala Gln Ile Lys Ile Glu Phe
                645                 650                 655

Asp Ile Phe Gln Glu Ser Lys Glu Lys Trp Glu Lys Phe Asp Ile Asp
            660                 665                 670

Lys Leu Arg Lys Leu Ile Glu Tyr Tyr Lys Lys Trp Val Leu Ala Leu
        675                 680                 685

Tyr Ser Asp Leu Tyr Asp Leu Glu Phe Ile Lys Tyr Lys Asn Tyr Asp
    690                 695                 700

Asp Leu Ser Ile Phe Tyr Ser Asp Val Glu Lys Lys Met Tyr Asn Leu
705                 710                 715                 720

Asn Phe Thr Lys Ile Asp Lys Ser Leu Ile Asp Gly Lys Val Lys Ser
                725                 730                 735

Trp Glu Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Glu Ser
            740                 745                 750
```

Lys Lys Glu Trp Ser Thr Glu Asn Ile His Thr Lys Tyr Phe Lys Leu
755 760 765

Leu Phe Asn Glu Lys Asn Leu Gln Asn Leu Val Val Lys Leu Ser Trp
770 775 780

Trp Ala Asp Ile Phe Phe Arg Asp Lys Thr Glu Asn Leu Lys Phe Lys
785 790 795 800

Lys Asp Lys Asn Gly Gln Glu Ile Leu Asp His Arg Arg Phe Ser Gln
805 810 815

Asp Lys Ile Met Phe His Ile Ser Ile Thr Leu Asn Ala Asn Cys Trp
820 825 830

Asp Lys Tyr Trp Phe Asn Gln Tyr Val Asn Glu Tyr Met Asn Lys Glu
835 840 845

Arg Asp Ile Lys Ile Ile Trp Ile Asp Arg Trp Glu Lys His Leu Ala
850 855 860

Tyr Tyr Cys Val Ile Asp Lys Ser Trp Lys Ile Phe Asn Asn Glu Ile
865 870 875 880

Trp Thr Leu Asn Glu Leu Asn Trp Val Asn Tyr Leu Glu Lys Leu Glu
885 890 895

Lys Ile Glu Ser Ser Arg Lys Asp Ser Arg Ile Ser Trp Trp Glu Ile
900 905 910

Glu Asn Ile Lys Glu Leu Lys Asn Gly Tyr Ile Ser Gln Val Ile Asn
915 920 925

Lys Leu Thr Glu Leu Ile Val Lys Tyr Asn Ala Ile Ile Val Phe Glu
930 935 940

Asp Leu Asn Ile Trp Phe Lys Arg Trp Arg Gln Lys Ile Glu Lys Gln
945 950 955 960

Ile Tyr Gln Lys Leu Glu Leu Ala Leu Ala Lys Lys Leu Asn Tyr Leu
965 970 975

Thr Gln Lys Asp Lys Lys Asp Asp Glu Ile Leu Trp Asn Leu Lys Ala
980 985 990

Leu Gln Leu Val Pro Lys Val Asn Asp Tyr Gln Asp Ile Trp Asn Tyr
995 1000 1005

Lys Gln Ser Trp Ile Met Phe Tyr Val Arg Ala Asn Tyr Thr Ser
1010 1015 1020

Val Thr Cys Pro Asn Cys Trp Leu Arg Lys Asn Leu Tyr Ile Ser
1025 1030 1035

Asn Ser Ala Thr Lys Glu Asn Gln Lys Lys Ser Leu Asn Ser Ile
1040 1045 1050

Ala Ile Lys Tyr Asn Asp Trp Lys Phe Ser Phe Ser Tyr Glu Ile
1055 1060 1065

Asp Asp Lys Ser Trp Lys Gln Lys Gln Ser Leu Asn Lys Lys Lys
1070 1075 1080

Phe Ile Val Tyr Ser Asp Ile Glu Arg Phe Val Tyr Ser Pro Leu
1085 1090 1095

Glu Lys Leu Thr Lys Val Ile Asp Val Asn Lys Lys Leu Leu Glu
1100 1105 1110

Leu Phe Arg Asp Phe Asn Leu Ser Leu Asp Ile Asn Lys Gln Ile
1115 1120 1125

Gln Glu Lys Asp Leu Asp Ser Val Phe Phe Lys Ser Leu Thr His
1130 1135 1140

Leu Phe Asn Leu Ile Leu Gln Leu Arg Asn Ser Asp Ser Lys Asp
1145 1150 1155

Asn Lys Asp Tyr Ile Ser Cys Pro Ser Cys Tyr Tyr His Ser Asn
1160 1165 1170

Asn Trp Leu Gln Trp Phe Glu Phe Asn Trp Asp Ala Asn Trp Ala
1175 1180 1185

Tyr Asn Ile Ala Arg Lys Gly Ile Ile Leu Leu Asp Arg Ile Arg
1190 1195 1200

Lys Asn Gln Glu Lys Pro Asp Leu Tyr Val Ser Asp Ile Asp Trp
1205 1210 1215

Asp Asn Phe Val Gln Ser Asn Gln Phe Pro Asn Thr Ile Ile Pro
1220 1225 1230

Ile Gln Asn Ile Glu Lys Gln Val Pro Leu Asn Ile Lys Ile
1235 1240 1245

<210> 102
<211> 1246
<212> PRT
<213> 獼猴卟啉單胞菌（Porphyromonas macacae）

<400> 102
Met Lys Thr Gln His Phe Phe Glu Asp Phe Thr Ser Leu Tyr Ser Leu
1 5 10 15

Ser Lys Thr Ile Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Leu Glu
20 25 30

Asn Ile Lys Lys Asn Gly Leu Ile Arg Arg Asp Glu Gln Arg Leu Asp
35 40 45

Asp Tyr Glu Lys Leu Lys Lys Val Ile Asp Glu Tyr His Glu Asp Phe
50 55 60

Ile Ala Asn Ile Leu Ser Ser Phe Ser Phe Ser Glu Glu Ile Leu Gln
65 70 75 80

Ser Tyr Ile Gln Asn Leu Ser Glu Ser Glu Ala Arg Ala Lys Ile Glu
85 90 95

Lys Thr Met Arg Asp Thr Leu Ala Lys Ala Phe Ser Glu Asp Glu Arg
100 105 110

Tyr Lys Ser Ile Phe Lys Lys Glu Leu Val Lys Lys Asp Ile Pro Val
115 120 125

```
Trp Cys Pro Ala Tyr Lys Ser Leu Cys Lys Lys Phe Asp Asn Phe Thr
    130                 135                 140

Thr Ser Leu Val Pro Phe His Glu Asn Arg Lys Asn Leu Tyr Thr Ser
145                 150                 155                 160

Asn Glu Ile Thr Ala Ser Ile Pro Tyr Arg Ile Val His Val Asn Leu
                165                 170                 175

Pro Lys Phe Ile Gln Asn Ile Glu Ala Leu Cys Glu Leu Gln Lys Lys
            180                 185                 190

Met Gly Ala Asp Leu Tyr Leu Glu Met Met Glu Asn Leu Arg Asn Val
        195                 200                 205

Trp Pro Ser Phe Val Lys Thr Pro Asp Asp Leu Cys Asn Leu Lys Thr
    210                 215                 220

Tyr Asn His Leu Met Val Gln Ser Ser Ile Ser Glu Tyr Asn Arg Phe
225                 230                 235                 240

Val Gly Gly Tyr Ser Thr Glu Asp Gly Thr Lys His Gln Gly Ile Asn
                245                 250                 255

Glu Trp Ile Asn Ile Tyr Arg Gln Arg Asn Lys Glu Met Arg Leu Pro
            260                 265                 270

Gly Leu Val Phe Leu His Lys Gln Ile Leu Ala Lys Val Asp Ser Ser
        275                 280                 285

Ser Phe Ile Ser Asp Thr Leu Glu Asn Asp Asp Gln Val Phe Cys Val
    290                 295                 300

Leu Arg Gln Phe Arg Lys Leu Phe Trp Asn Thr Val Ser Ser Lys Glu
305                 310                 315                 320

Asp Asp Ala Ala Ser Leu Lys Asp Leu Phe Cys Gly Leu Ser Gly Tyr
                325                 330                 335

Asp Pro Glu Ala Ile Tyr Val Ser Asp Ala His Leu Ala Thr Ile Ser
            340                 345                 350

Lys Asn Ile Phe Asp Arg Trp Asn Tyr Ile Ser Asp Ala Ile Arg Arg
        355                 360                 365

Lys Thr Glu Val Leu Met Pro Arg Lys Lys Glu Ser Val Glu Arg Tyr
    370                 375                 380

Ala Glu Lys Ile Ser Lys Gln Ile Lys Lys Arg Gln Ser Tyr Ser Leu
385                 390                 395                 400

Ala Glu Leu Asp Asp Leu Leu Ala His Tyr Ser Glu Glu Ser Leu Pro
                405                 410                 415

Ala Gly Phe Ser Leu Leu Ser Tyr Phe Thr Ser Leu Gly Gly Gln Lys
            420                 425                 430

Tyr Leu Val Ser Asp Gly Glu Val Ile Leu Tyr Glu Glu Gly Ser Asn
        435                 440                 445

Ile Trp Asp Glu Val Leu Ile Ala Phe Arg Asp Leu Gln Val Ile Leu
    450                 455                 460
```

```
Asp Lys Asp Phe Thr Glu Lys Lys Leu Gly Lys Asp Glu Glu Ala Val
465                 470                 475                 480

Ser Val Ile Lys Lys Ala Leu Asp Ser Ala Leu Arg Leu Arg Lys Phe
                485                 490                 495

Phe Asp Leu Leu Ser Gly Thr Gly Ala Glu Ile Arg Arg Asp Ser Ser
            500                 505                 510

Phe Tyr Ala Leu Tyr Thr Asp Arg Met Asp Lys Leu Lys Gly Leu Leu
        515                 520                 525

Lys Met Tyr Asp Lys Val Arg Asn Tyr Leu Thr Lys Lys Pro Tyr Ser
    530                 535                 540

Ile Glu Lys Phe Lys Leu His Phe Asp Asn Pro Ser Leu Leu Ser Gly
545                 550                 555                 560

Trp Asp Lys Asn Lys Glu Leu Asn Asn Leu Ser Val Ile Phe Arg Gln
                565                 570                 575

Asn Gly Tyr Tyr Tyr Leu Gly Ile Met Thr Pro Lys Gly Lys Asn Leu
            580                 585                 590

Phe Lys Thr Leu Pro Lys Leu Gly Ala Glu Glu Met Phe Tyr Glu Lys
        595                 600                 605

Met Glu Tyr Lys Gln Ile Ala Glu Pro Met Leu Met Leu Pro Lys Val
    610                 615                 620

Phe Phe Pro Lys Lys Thr Lys Pro Ala Phe Ala Pro Asp Gln Ser Val
625                 630                 635                 640

Val Asp Ile Tyr Asn Lys Lys Thr Phe Lys Thr Gly Gln Lys Gly Phe
                645                 650                 655

Asn Lys Lys Asp Leu Tyr Arg Leu Ile Asp Phe Tyr Lys Glu Ala Leu
            660                 665                 670

Thr Val His Glu Trp Lys Leu Phe Asn Phe Ser Phe Ser Pro Thr Glu
        675                 680                 685

Gln Tyr Arg Asn Ile Gly Glu Phe Phe Asp Glu Val Arg Glu Gln Ala
    690                 695                 700

Tyr Lys Val Ser Met Val Asn Val Pro Ala Ser Tyr Ile Asp Glu Ala
705                 710                 715                 720

Val Glu Asn Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe
                725                 730                 735

Ser Pro Tyr Ser Lys Gly Ile Pro Asn Leu His Thr Leu Tyr Trp Lys
            740                 745                 750

Ala Leu Phe Ser Glu Gln Asn Gln Ser Arg Val Tyr Lys Leu Cys Gly
        755                 760                 765

Gly Gly Glu Leu Phe Tyr Arg Lys Ala Ser Leu His Met Gln Asp Thr
    770                 775                 780

Thr Val His Pro Lys Gly Ile Ser Ile His Lys Lys Asn Leu Asn Lys
785                 790                 795                 800
```

```
Lys Gly Glu Thr Ser Leu Phe Asn Tyr Asp Leu Val Lys Asp Lys Arg
                805                 810                 815

Phe Thr Glu Asp Lys Phe Phe Phe His Val Pro Ile Ser Ile Asn Tyr
            820                 825                 830

Lys Asn Lys Lys Ile Thr Asn Val Asn Gln Met Val Arg Asp Tyr Ile
        835                 840                 845

Ala Gln Asn Asp Asp Leu Gln Ile Ile Gly Ile Asp Arg Gly Glu Arg
    850                 855                 860

Asn Leu Leu Tyr Ile Ser Arg Ile Asp Thr Arg Gly Asn Leu Leu Glu
865                 870                 875                 880

Gln Phe Ser Leu Asn Val Ile Glu Ser Asp Lys Gly Asp Leu Arg Thr
                885                 890                 895

Asp Tyr Gln Lys Ile Leu Gly Asp Arg Glu Gln Glu Arg Leu Arg Arg
            900                 905                 910

Arg Gln Glu Trp Lys Ser Ile Glu Ser Ile Lys Asp Leu Lys Asp Gly
        915                 920                 925

Tyr Met Ser Gln Val Val His Lys Ile Cys Asn Met Val Val Glu His
    930                 935                 940

Lys Ala Ile Val Val Leu Glu Asn Leu Asn Leu Ser Phe Met Lys Gly
945                 950                 955                 960

Arg Lys Lys Val Glu Lys Ser Val Tyr Glu Lys Phe Glu Arg Met Leu
                965                 970                 975

Val Asp Lys Leu Asn Tyr Leu Val Val Asp Lys Lys Asn Leu Ser Asn
            980                 985                 990

Glu Pro Gly Gly Leu Tyr Ala Ala Tyr Gln Leu Thr Asn Pro Leu Phe
        995                 1000                1005

Ser Phe  Glu Glu Leu His Arg Tyr Pro Gln Ser  Gly Ile Leu Phe
    1010                1015                1020

Phe Val  Asp Pro Trp Asn Thr Ser Leu Thr Asp  Pro Ser Thr Gly
    1025                1030                1035

Phe Val  Asn Leu Leu Gly Arg Ile Asn Tyr Thr  Asn Val Gly Asp
    1040                1045                1050

Ala Arg  Lys Phe Phe Asp Arg Phe Asn Ala Ile  Arg Tyr Asp Gly
    1055                1060                1065

Lys Gly  Asn Ile Leu Phe Asp Leu Asp Leu Ser  Arg Phe Asp Val
    1070                1075                1080

Arg Val  Glu Thr Gln Arg Lys Leu Trp Thr Leu  Thr Thr Phe Gly
    1085                1090                1095

Ser Arg  Ile Ala Lys Ser Lys Lys Ser Gly Lys  Trp Met Val Glu
    1100                1105                1110

Arg Ile  Glu Asn Leu Ser Leu Cys Phe Leu Glu  Leu Phe Glu Gln
    1115                1120                1125
```

Phe Asn Ile Gly Tyr Arg Val Glu Lys Asp Leu Lys Lys Ala Ile
1130 1135 1140

Leu Ser Gln Asp Arg Lys Glu Phe Tyr Val Arg Leu Ile Tyr Leu
1145 1150 1155

Phe Asn Leu Met Met Gln Ile Arg Asn Ser Asp Gly Glu Glu Asp
1160 1165 1170

Tyr Ile Leu Ser Pro Ala Leu Asn Glu Lys Asn Leu Gln Phe Asp
1175 1180 1185

Ser Arg Leu Ile Glu Ala Lys Asp Leu Pro Val Asp Ala Asp Ala
1190 1195 1200

Asn Gly Ala Tyr Asn Val Ala Arg Lys Gly Leu Met Val Val Gln
1205 1210 1215

Arg Ile Lys Arg Gly Asp His Glu Ser Ile His Arg Ile Gly Arg
1220 1225 1230

Ala Gln Trp Leu Arg Tyr Val Gln Glu Gly Ile Val Glu
1235 1240 1245

<210> 103
<211> 1241
<212> PRT
<213> 解朊丁酸弧菌（Butyrivibrio proteoclasticus）

<400> 103
Met Leu Leu Tyr Glu Asn Tyr Thr Lys Arg Asn Gln Ile Thr Lys Ser
1 5 10 15

Leu Arg Leu Glu Leu Arg Pro Gln Gly Lys Thr Leu Arg Asn Ile Lys
20 25 30

Glu Leu Asn Leu Leu Glu Gln Asp Lys Ala Ile Tyr Ala Leu Leu Glu
35 40 45

Arg Leu Lys Pro Val Ile Asp Glu Gly Ile Lys Asp Ile Ala Arg Asp
50 55 60

Thr Leu Lys Asn Cys Glu Leu Ser Phe Glu Lys Leu Tyr Glu His Phe
65 70 75 80

Leu Ser Gly Asp Lys Lys Ala Tyr Ala Lys Glu Ser Glu Arg Leu Lys
85 90 95

Lys Glu Ile Val Lys Thr Leu Ile Lys Asn Leu Pro Glu Gly Ile Gly
100 105 110

Lys Ile Ser Glu Ile Asn Ser Ala Lys Tyr Leu Asn Gly Val Leu Tyr
115 120 125

Asp Phe Ile Asp Lys Thr His Lys Asp Ser Glu Glu Lys Gln Asn Ile
130 135 140

Leu Ser Asp Ile Leu Glu Thr Lys Gly Tyr Leu Ala Leu Phe Ser Lys
145 150 155 160

Phe Leu Thr Ser Arg Ile Thr Thr Leu Glu Gln Ser Met Pro Lys Arg
165 170 175

Val Ile Glu Asn Phe Glu Ile Tyr Ala Ala Asn Ile Pro Lys Met Gln
180 185 190

```
Asp Ala Leu Glu Arg Gly Ala Val Ser Phe Ala Ile Glu Tyr Glu Ser
        195                 200                 205

Ile Cys Ser Val Asp Tyr Tyr Asn Gln Ile Leu Ser Gln Glu Asp Ile
    210                 215                 220

Asp Ser Tyr Asn Arg Leu Ile Ser Gly Ile Met Asp Glu Asp Gly Ala
225                 230                 235                 240

Lys Glu Lys Gly Ile Asn Gln Thr Ile Ser Glu Lys Asn Ile Lys Ile
                245                 250                 255

Lys Ser Glu His Leu Glu Glu Lys Pro Phe Arg Ile Leu Lys Gln Leu
            260                 265                 270

His Lys Gln Ile Leu Glu Glu Arg Glu Lys Ala Phe Thr Ile Asp His
        275                 280                 285

Ile Asp Ser Asp Glu Glu Val Val Gln Val Thr Lys Glu Ala Phe Glu
    290                 295                 300

Gln Thr Lys Glu Gln Trp Glu Asn Ile Lys Lys Ile Asn Gly Phe Tyr
305                 310                 315                 320

Ala Lys Asp Pro Gly Asp Ile Thr Leu Phe Ile Val Val Gly Pro Asn
                325                 330                 335

Gln Thr His Val Leu Ser Gln Leu Ile Tyr Gly Glu His Asp Arg Ile
            340                 345                 350

Arg Leu Leu Leu Glu Glu Tyr Glu Lys Asn Thr Leu Glu Val Leu Pro
        355                 360                 365

Arg Arg Thr Lys Ser Glu Lys Ala Arg Tyr Asp Lys Phe Val Asn Ala
    370                 375                 380

Val Pro Lys Lys Val Ala Lys Glu Ser His Thr Phe Asp Gly Leu Gln
385                 390                 395                 400

Lys Met Thr Gly Asp Asp Arg Leu Phe Ile Leu Tyr Arg Asp Glu Leu
                405                 410                 415

Ala Arg Asn Tyr Met Arg Ile Lys Glu Ala Tyr Gly Thr Phe Glu Arg
            420                 425                 430

Asp Ile Leu Lys Ser Arg Arg Gly Ile Lys Gly Asn Arg Asp Val Gln
        435                 440                 445

Glu Ser Leu Val Ser Phe Tyr Asp Glu Leu Thr Lys Phe Arg Ser Ala
    450                 455                 460

Leu Arg Ile Ile Asn Ser Gly Asn Asp Glu Lys Ala Asp Pro Ile Phe
465                 470                 475                 480

Tyr Asn Thr Phe Asp Gly Ile Phe Glu Lys Ala Asn Arg Thr Tyr Lys
                485                 490                 495

Ala Glu Asn Leu Cys Arg Asn Tyr Val Thr Lys Ser Pro Ala Asp Asp
            500                 505                 510

Ala Arg Ile Met Ala Ser Cys Leu Gly Thr Pro Ala Arg Leu Arg Thr
        515                 520                 525
```

```
His Trp Trp Asn Gly Glu Glu Asn Phe Ala Ile Asn Asp Val Ala Met
    530                 535                 540

Ile Arg Arg Gly Asp Glu Tyr Tyr Tyr Phe Val Leu Thr Pro Asp Val
545                 550                 555                 560

Lys Pro Val Asp Leu Lys Thr Lys Asp Glu Thr Asp Ala Gln Ile Phe
                565                 570                 575

Val Gln Arg Lys Gly Ala Lys Ser Phe Leu Gly Leu Pro Lys Ala Leu
            580                 585                 590

Phe Lys Cys Ile Leu Glu Pro Tyr Phe Glu Ser Pro Glu His Lys Asn
        595                 600                 605

Asp Lys Asn Cys Val Ile Glu Glu Tyr Val Ser Lys Pro Leu Thr Ile
    610                 615                 620

Asp Arg Arg Ala Tyr Asp Ile Phe Lys Asn Gly Thr Phe Lys Lys Thr
625                 630                 635                 640

Asn Ile Gly Ile Asp Gly Leu Thr Glu Glu Lys Phe Lys Asp Asp Cys
                645                 650                 655

Arg Tyr Leu Ile Asp Val Tyr Lys Glu Phe Ile Ala Val Tyr Thr Arg
            660                 665                 670

Tyr Ser Cys Phe Asn Met Ser Gly Leu Lys Arg Ala Asp Glu Tyr Asn
        675                 680                 685

Asp Ile Gly Glu Phe Phe Ser Asp Val Asp Thr Arg Leu Cys Thr Met
    690                 695                 700

Glu Trp Ile Pro Val Ser Phe Glu Arg Ile Asn Asp Met Val Asp Lys
705                 710                 715                 720

Lys Glu Gly Leu Leu Phe Leu Val Arg Ser Met Phe Leu Tyr Asn Arg
                725                 730                 735

Pro Arg Lys Pro Tyr Glu Arg Thr Phe Ile Gln Leu Phe Ser Asp Ser
            740                 745                 750

Asn Met Glu His Thr Ser Met Leu Leu Asn Ser Arg Ala Met Ile Gln
        755                 760                 765

Tyr Arg Ala Ala Ser Leu Pro Arg Arg Val Thr His Lys Lys Gly Ser
    770                 775                 780

Ile Leu Val Ala Leu Arg Asp Ser Asn Gly Glu His Ile Pro Met His
785                 790                 795                 800

Ile Arg Glu Ala Ile Tyr Lys Met Lys Asn Asn Phe Asp Ile Ser Ser
                805                 810                 815

Glu Asp Phe Ile Met Ala Lys Ala Tyr Leu Ala Glu His Asp Val Ala
            820                 825                 830

Ile Lys Lys Ala Asn Glu Asp Ile Ile Arg Asn Arg Arg Tyr Thr Glu
        835                 840                 845

Asp Lys Phe Phe Leu Ser Leu Ser Tyr Thr Lys Asn Ala Asp Ile Ser
    850                 855                 860
```

```
Ala Arg Thr Leu Asp Tyr Ile Asn Asp Lys Val Glu Glu Asp Thr Gln
865                 870                 875                 880

Asp Ser Arg Met Ala Val Ile Val Thr Arg Asn Leu Lys Asp Leu Thr
                885                 890                 895

Tyr Val Ala Val Val Asp Glu Lys Asn Asn Val Leu Glu Glu Lys Ser
            900                 905                 910

Leu Asn Glu Ile Asp Gly Val Asn Tyr Arg Glu Leu Leu Lys Glu Arg
        915                 920                 925

Thr Lys Ile Lys Tyr His Asp Lys Thr Arg Leu Trp Gln Tyr Asp Val
    930                 935                 940

Ser Ser Lys Gly Leu Lys Glu Ala Tyr Val Glu Leu Ala Val Thr Gln
945                 950                 955                 960

Ile Ser Lys Leu Ala Thr Lys Tyr Asn Ala Val Val Val Val Glu Ser
                965                 970                 975

Met Ser Ser Thr Phe Lys Asp Lys Phe Ser Phe Leu Asp Glu Gln Ile
            980                 985                 990

Phe Lys Ala Phe Glu Ala Arg Leu Cys Ala Arg Met Ser Asp Leu Ser
        995                 1000                1005

Phe Asn  Thr Ile Lys Glu Gly  Glu Ala Gly Ser Ile  Ser Asn Pro
    1010                1015                1020

Ile Gln  Val Ser Asn Asn Asn  Gly Asn Ser Tyr Gln  Asp Gly Val
    1025                1030                1035

Ile Tyr  Phe Leu Asn Asn Ala  Tyr Thr Arg Thr Leu  Cys Pro Asp
    1040                1045                1050

Thr Gly  Phe Val Asp Val Phe  Asp Lys Thr Arg Leu  Ile Thr Met
    1055                1060                1065

Gln Ser  Lys Arg Gln Phe Phe  Ala Lys Met Lys Asp  Ile Arg Ile
    1070                1075                1080

Asp Asp  Gly Glu Met Leu Phe  Thr Phe Asn Leu Glu  Glu Tyr Pro
    1085                1090                1095

Thr Lys  Arg Leu Leu Asp Arg  Lys Glu Trp Thr Val  Lys Ile Ala
    1100                1105                1110

Gly Asp  Gly Ser Tyr Phe Asp  Lys Asp Lys Gly Glu  Tyr Val Tyr
    1115                1120                1125

Val Asn  Asp Ile Val Arg Glu  Gln Ile Ile Pro Ala  Leu Leu Glu
    1130                1135                1140

Asp Lys  Ala Val Phe Asp Gly  Asn Met Ala Glu Lys  Phe Leu Asp
    1145                1150                1155

Lys Thr  Ala Ile Ser Gly Lys  Ser Val Glu Leu Ile  Tyr Lys Trp
    1160                1165                1170

Phe Ala  Asn Ala Leu Tyr Gly  Ile Ile Thr Lys Lys  Asp Gly Glu
    1175                1180                1185
```

Lys Ile Tyr Arg Ser Pro Ile Thr Gly Thr Glu Ile Asp Val Ser
1190 1195 1200

Lys Asn Thr Thr Tyr Asn Phe Gly Lys Lys Phe Met Phe Lys Gln
1205 1210 1215

Glu Tyr Arg Gly Asp Gly Asp Phe Leu Asp Ala Phe Leu Asn Tyr
1220 1225 1230

Met Gln Ala Gln Asp Ile Ala Val
1235 1240

<210> 104
<211> 1238
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選白蟻甲烷枝原體（Candidatus Methanoplasma termitum）序列

<400> 104
Met Asn Asn Tyr Asp Glu Phe Thr Lys Leu Tyr Pro Ile Gln Lys Thr
1 5 10 15

Ile Arg Phe Glu Leu Lys Pro Gln Gly Arg Thr Met Glu His Leu Glu
20 25 30

Thr Phe Asn Phe Phe Glu Glu Asp Arg Asp Arg Ala Glu Lys Tyr Lys
35 40 45

Ile Leu Lys Glu Ala Ile Asp Glu Tyr His Lys Lys Phe Ile Asp Glu
50 55 60

His Leu Thr Asn Met Ser Leu Asp Trp Asn Ser Leu Lys Gln Ile Ser
65 70 75 80

Glu Lys Tyr Tyr Lys Ser Arg Glu Glu Lys Asp Lys Lys Val Phe Leu
85 90 95

Ser Glu Gln Lys Arg Met Arg Gln Glu Ile Val Ser Glu Phe Lys Lys
100 105 110

Asp Asp Arg Phe Lys Asp Leu Phe Ser Lys Lys Leu Phe Ser Glu Leu
115 120 125

Leu Lys Glu Glu Ile Tyr Lys Lys Gly Asn His Gln Glu Ile Asp Ala
130 135 140

Leu Lys Ser Phe Asp Lys Phe Ser Gly Tyr Phe Ile Gly Leu His Glu
145 150 155 160

Asn Arg Lys Asn Met Tyr Ser Asp Gly Asp Glu Ile Thr Ala Ile Ser
165 170 175

Asn Arg Ile Val Asn Glu Asn Phe Pro Lys Phe Leu Asp Asn Leu Gln
180 185 190

Lys Tyr Gln Glu Ala Arg Lys Lys Tyr Pro Glu Trp Ile Ile Lys Ala
195 200 205

Glu Ser Ala Leu Val Ala His Asn Ile Lys Met Asp Glu Val Phe Ser
210 215 220

```
Leu Glu Tyr Phe Asn Lys Val Leu Asn Gln Glu Gly Ile Gln Arg Tyr
225                 230                 235                 240

Asn Leu Ala Leu Gly Gly Tyr Val Thr Lys Ser Gly Glu Lys Met Met
                245                 250                 255

Gly Leu Asn Asp Ala Leu Asn Leu Ala His Gln Ser Glu Lys Ser Ser
            260                 265                 270

Lys Gly Arg Ile His Met Thr Pro Leu Phe Lys Gln Ile Leu Ser Glu
        275                 280                 285

Lys Glu Ser Phe Ser Tyr Ile Pro Asp Val Phe Thr Glu Asp Ser Gln
    290                 295                 300

Leu Leu Pro Ser Ile Gly Gly Phe Phe Ala Gln Ile Glu Asn Asp Lys
305                 310                 315                 320

Asp Gly Asn Ile Phe Asp Arg Ala Leu Glu Leu Ile Ser Ser Tyr Ala
                325                 330                 335

Glu Tyr Asp Thr Glu Arg Ile Tyr Ile Arg Gln Ala Asp Ile Asn Arg
            340                 345                 350

Val Ser Asn Val Ile Phe Gly Glu Trp Gly Thr Leu Gly Gly Leu Met
        355                 360                 365

Arg Glu Tyr Lys Ala Asp Ser Ile Asn Asp Ile Asn Leu Glu Arg Thr
    370                 375                 380

Cys Lys Lys Val Asp Lys Trp Leu Asp Ser Lys Glu Phe Ala Leu Ser
385                 390                 395                 400

Asp Val Leu Glu Ala Ile Lys Arg Thr Gly Asn Asn Asp Ala Phe Asn
                405                 410                 415

Glu Tyr Ile Ser Lys Met Arg Thr Ala Arg Glu Lys Ile Asp Ala Ala
            420                 425                 430

Arg Lys Glu Met Lys Phe Ile Ser Glu Lys Ile Ser Gly Asp Glu Glu
        435                 440                 445

Ser Ile His Ile Ile Lys Thr Leu Leu Asp Ser Val Gln Gln Phe Leu
    450                 455                 460

His Phe Phe Asn Leu Phe Lys Ala Arg Gln Asp Ile Pro Leu Asp Gly
465                 470                 475                 480

Ala Phe Tyr Ala Glu Phe Asp Glu Val His Ser Lys Leu Phe Ala Ile
                485                 490                 495

Val Pro Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Lys Asn Asn Leu
            500                 505                 510

Asn Thr Lys Lys Ile Lys Leu Asn Phe Lys Asn Pro Thr Leu Ala Asn
        515                 520                 525

Gly Trp Asp Gln Asn Lys Val Tyr Asp Tyr Ala Ser Leu Ile Phe Leu
    530                 535                 540

Arg Asp Gly Asn Tyr Tyr Leu Gly Ile Ile Asn Pro Lys Arg Lys Lys
545                 550                 555                 560
```

```
Asn Ile Lys Phe Glu Gln Gly Ser Gly Asn Gly Pro Phe Tyr Arg Lys
                565                 570                 575

Met Val Tyr Lys Gln Ile Pro Gly Pro Asn Lys Asn Leu Pro Arg Val
            580                 585                 590

Phe Leu Thr Ser Thr Lys Gly Lys Lys Glu Tyr Lys Pro Ser Lys Glu
        595                 600                 605

Ile Ile Glu Gly Tyr Glu Ala Asp Lys His Ile Arg Gly Asp Lys Phe
    610                 615                 620

Asp Leu Asp Phe Cys His Lys Leu Ile Asp Phe Phe Lys Glu Ser Ile
625                 630                 635                 640

Glu Lys His Lys Asp Trp Ser Lys Phe Asn Phe Tyr Phe Ser Pro Thr
                645                 650                 655

Glu Ser Tyr Gly Asp Ile Ser Glu Phe Tyr Leu Asp Val Glu Lys Gln
            660                 665                 670

Gly Tyr Arg Met His Phe Glu Asn Ile Ser Ala Glu Thr Ile Asp Glu
        675                 680                 685

Tyr Val Glu Lys Gly Asp Leu Phe Leu Phe Gln Ile Tyr Asn Lys Asp
    690                 695                 700

Phe Val Lys Ala Ala Thr Gly Lys Lys Asp Met His Thr Ile Tyr Trp
705                 710                 715                 720

Asn Ala Ala Phe Ser Pro Glu Asn Leu Gln Asp Val Val Val Lys Leu
                725                 730                 735

Asn Gly Glu Ala Glu Leu Phe Tyr Arg Asp Lys Ser Asp Ile Lys Glu
            740                 745                 750

Ile Val His Arg Glu Gly Glu Ile Leu Val Asn Arg Thr Tyr Asn Gly
        755                 760                 765

Arg Thr Pro Val Pro Asp Lys Ile His Lys Lys Leu Thr Asp Tyr His
    770                 775                 780

Asn Gly Arg Thr Lys Asp Leu Gly Glu Ala Lys Glu Tyr Leu Asp Lys
785                 790                 795                 800

Val Arg Tyr Phe Lys Ala His Tyr Asp Ile Thr Lys Asp Arg Arg Tyr
                805                 810                 815

Leu Asn Asp Lys Ile Tyr Phe His Val Pro Leu Thr Leu Asn Phe Lys
            820                 825                 830

Ala Asn Gly Lys Lys Asn Leu Asn Lys Met Val Ile Glu Lys Phe Leu
        835                 840                 845

Ser Asp Glu Lys Ala His Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn
    850                 855                 860

Leu Leu Tyr Tyr Ser Ile Ile Asp Arg Ser Gly Lys Ile Ile Asp Gln
865                 870                 875                 880

Gln Ser Leu Asn Val Ile Asp Gly Phe Asp Tyr Arg Glu Lys Leu Asn
                885                 890                 895
```

```
Gln Arg Glu Ile Glu Met Lys Asp Ala Arg Gln Ser Trp Asn Ala Ile
            900                 905                 910

Gly Lys Ile Lys Asp Leu Lys Glu Gly Tyr Leu Ser Lys Ala Val His
        915                 920                 925

Glu Ile Thr Lys Met Ala Ile Gln Tyr Asn Ala Ile Val Val Met Glu
    930                 935                 940

Glu Leu Asn Tyr Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln
945                 950                 955                 960

Ile Tyr Gln Lys Phe Glu Asn Met Leu Ile Asp Lys Met Asn Tyr Leu
                965                 970                 975

Val Phe Lys Asp Ala Pro Asp Glu Ser Pro Gly Gly Val Leu Asn Ala
            980                 985                 990

Tyr Gln Leu Thr Asn Pro Leu Glu Ser Phe Ala Lys Leu Gly Lys Gln
        995                 1000                1005

Thr Gly  Ile Leu Phe Tyr Val  Pro Ala Ala Tyr Thr  Ser Lys Ile
    1010                1015                1020

Asp Pro  Thr Thr Gly Phe Val  Asn Leu Phe Asn Thr  Ser Ser Lys
    1025                1030                1035

Thr Asn  Ala Gln Glu Arg Lys  Glu Phe Leu Gln Lys  Phe Glu Ser
    1040                1045                1050

Ile Ser  Tyr Ser Ala Lys Asp  Gly Gly Ile Phe Ala  Phe Ala Phe
    1055                1060                1065

Asp Tyr  Arg Lys Phe Gly Thr  Ser Lys Thr Asp His  Lys Asn Val
    1070                1075                1080

Trp Thr  Ala Tyr Thr Asn Gly  Glu Arg Met Arg Tyr  Ile Lys Glu
    1085                1090                1095

Lys Lys  Arg Asn Glu Leu Phe  Asp Pro Ser Lys Glu  Ile Lys Glu
    1100                1105                1110

Ala Leu  Thr Ser Ser Gly Ile  Lys Tyr Asp Gly Gly  Gln Asn Ile
    1115                1120                1125

Leu Pro  Asp Ile Leu Arg Ser  Asn Asn Asn Gly Leu  Ile Tyr Thr
    1130                1135                1140

Met Tyr  Ser Ser Phe Ile Ala  Ala Ile Gln Met Arg  Val Tyr Asp
    1145                1150                1155

Gly Lys  Glu Asp Tyr Ile Ile  Ser Pro Ile Lys Asn  Ser Lys Gly
    1160                1165                1170

Glu Phe  Phe Arg Thr Asp Pro  Lys Arg Arg Glu Leu  Pro Ile Asp
    1175                1180                1185

Ala Asp  Ala Asn Gly Ala Tyr  Asn Ile Ala Leu Arg  Gly Glu Leu
    1190                1195                1200

Thr Met  Arg Ala Ile Ala Glu  Lys Phe Asp Pro Asp  Ser Glu Lys
    1205                1210                1215
```

Met Ala Lys Leu Glu Leu Lys His Lys Asp Trp Phe Glu Phe Met
1220 1225 1230

Gln Thr Arg Gly Asp
1235

<210> 105
<211> 1235
<212> PRT
<213> 厭氧弧菌屬某種（Anaerovibrio sp.）

<400> 105
Met Val Ala Phe Ile Asp Glu Phe Val Gly Gln Tyr Pro Val Ser Lys
1 5 10 15

Thr Leu Arg Phe Glu Ala Arg Pro Val Pro Glu Thr Lys Lys Trp Leu
20 25 30

Glu Ser Asp Gln Cys Ser Val Leu Phe Asn Asp Gln Lys Arg Asn Glu
35 40 45

Tyr Tyr Gly Val Leu Lys Glu Leu Leu Asp Asp Tyr Tyr Arg Ala Tyr
50 55 60

Ile Glu Asp Ala Leu Thr Ser Phe Thr Leu Asp Lys Ala Leu Leu Glu
65 70 75 80

Asn Ala Tyr Asp Leu Tyr Cys Asn Arg Asp Thr Asn Ala Phe Ser Ser
85 90 95

Cys Cys Glu Lys Leu Arg Lys Asp Leu Val Lys Ala Phe Gly Asn Leu
100 105 110

Lys Asp Tyr Leu Leu Gly Ser Asp Gln Leu Lys Asp Leu Val Lys Leu
115 120 125

Lys Ala Lys Val Asp Ala Pro Ala Gly Lys Gly Lys Lys Lys Ile Glu
130 135 140

Val Asp Ser Arg Leu Ile Asn Trp Leu Asn Asn Asn Ala Lys Tyr Ser
145 150 155 160

Ala Glu Asp Arg Glu Lys Tyr Ile Lys Ala Ile Glu Ser Phe Glu Gly
165 170 175

Phe Val Thr Tyr Leu Thr Asn Tyr Lys Gln Ala Arg Glu Asn Met Phe
180 185 190

Ser Ser Glu Asp Lys Ser Thr Ala Ile Ala Phe Arg Val Ile Asp Gln
195 200 205

Asn Met Val Thr Tyr Phe Gly Asn Ile Arg Ile Tyr Glu Lys Ile Lys
210 215 220

Ala Lys Tyr Pro Glu Leu Tyr Ser Ala Leu Lys Gly Phe Glu Lys Phe
225 230 235 240

Phe Ser Pro Thr Ala Tyr Ser Glu Ile Leu Ser Gln Ser Lys Ile Asp
245 250 255

Glu Tyr Asn Tyr Gln Cys Ile Gly Arg Pro Ile Asp Asp Ala Asp Phe
260 265 270

Lys Gly Val Asn Ser Leu Ile Asn Glu Tyr Arg Gln Lys Asn Gly Ile
275 280 285

```
Lys Ala Arg Glu Leu Pro Val Met Ser Met Leu Tyr Lys Gln Ile Leu
    290                 295                 300

Ser Asp Arg Asp Asn Ser Phe Met Ser Glu Val Ile Asn Arg Asn Glu
305                 310                 315                 320

Glu Ala Ile Glu Cys Ala Lys Asn Gly Tyr Lys Val Ser Tyr Ala Leu
                325                 330                 335

Phe Asn Glu Leu Leu Gln Leu Tyr Lys Lys Ile Phe Thr Glu Asp Asn
            340                 345                 350

Tyr Gly Asn Ile Tyr Val Lys Thr Gln Pro Leu Thr Glu Leu Ser Gln
        355                 360                 365

Ala Leu Phe Gly Asp Trp Ser Ile Leu Arg Asn Ala Leu Asp Asn Gly
    370                 375                 380

Lys Tyr Asp Lys Asp Ile Ile Asn Leu Ala Glu Leu Glu Lys Tyr Phe
385                 390                 395                 400

Ser Glu Tyr Cys Lys Val Leu Asp Ala Asp Asp Ala Ala Lys Ile Gln
                405                 410                 415

Asp Lys Phe Asn Leu Lys Asp Tyr Phe Ile Gln Lys Asn Ala Leu Asp
            420                 425                 430

Ala Thr Leu Pro Asp Leu Asp Lys Ile Thr Gln Tyr Lys Pro His Leu
        435                 440                 445

Asp Ala Met Leu Gln Ala Ile Arg Lys Tyr Lys Leu Phe Ser Met Tyr
    450                 455                 460

Asn Gly Arg Lys Lys Met Asp Val Pro Glu Asn Gly Ile Asp Phe Ser
465                 470                 475                 480

Asn Glu Phe Asn Ala Ile Tyr Asp Lys Leu Ser Glu Phe Ser Ile Leu
                485                 490                 495

Tyr Asp Arg Ile Arg Asn Phe Ala Thr Lys Lys Pro Tyr Ser Asp Glu
            500                 505                 510

Lys Met Lys Leu Ser Phe Asn Met Pro Thr Met Leu Ala Gly Trp Asp
        515                 520                 525

Tyr Asn Asn Glu Thr Ala Asn Gly Cys Phe Leu Phe Ile Lys Asp Gly
    530                 535                 540

Lys Tyr Phe Leu Gly Val Ala Asp Ser Lys Ser Lys Asn Ile Phe Asp
545                 550                 555                 560

Phe Lys Lys Asn Pro His Leu Leu Asp Lys Tyr Ser Ser Lys Asp Ile
                565                 570                 575

Tyr Tyr Lys Val Lys Tyr Lys Gln Val Ser Gly Ser Ala Lys Met Leu
            580                 585                 590

Pro Lys Val Val Phe Ala Gly Ser Asn Glu Lys Ile Phe Gly His Leu
        595                 600                 605

Ile Ser Lys Arg Ile Leu Glu Ile Arg Glu Lys Lys Leu Tyr Thr Ala
    610                 615                 620
```

```
Ala Ala Gly Asp Arg Lys Ala Val Ala Glu Trp Ile Asp Phe Met Lys
625                 630                 635                 640

Ser Ala Ile Ala Ile His Pro Glu Trp Asn Glu Tyr Phe Lys Phe Lys
                645                 650                 655

Phe Lys Asn Thr Ala Glu Tyr Asp Asn Ala Asn Lys Phe Tyr Glu Asp
            660                 665                 670

Ile Asp Lys Gln Thr Tyr Ser Leu Glu Lys Val Glu Ile Pro Thr Glu
        675                 680                 685

Tyr Ile Asp Glu Met Val Ser Gln His Lys Leu Tyr Leu Phe Gln Leu
    690                 695                 700

Tyr Thr Lys Asp Phe Ser Asp Lys Lys Lys Lys Lys Gly Thr Asp Asn
705                 710                 715                 720

Leu His Thr Met Tyr Trp His Gly Val Phe Ser Asp Glu Asn Leu Lys
                725                 730                 735

Ala Val Thr Glu Gly Thr Gln Pro Ile Ile Lys Leu Asn Gly Glu Ala
            740                 745                 750

Glu Met Phe Met Arg Asn Pro Ser Ile Glu Phe Gln Val Thr His Glu
        755                 760                 765

His Asn Lys Pro Ile Ala Asn Lys Asn Pro Leu Asn Thr Lys Lys Glu
    770                 775                 780

Ser Val Phe Asn Tyr Asp Leu Ile Lys Asp Lys Arg Tyr Thr Glu Arg
785                 790                 795                 800

Lys Phe Tyr Phe His Cys Pro Ile Thr Leu Asn Phe Arg Ala Asp Lys
                805                 810                 815

Pro Ile Lys Tyr Asn Glu Lys Ile Asn Arg Phe Val Glu Asn Asn Pro
            820                 825                 830

Asp Val Cys Ile Ile Gly Ile Asp Arg Gly Glu Arg His Leu Leu Tyr
        835                 840                 845

Tyr Thr Val Ile Asn Gln Thr Gly Asp Ile Leu Glu Gln Gly Ser Leu
    850                 855                 860

Asn Lys Ile Ser Gly Ser Tyr Thr Asn Asp Lys Gly Glu Lys Val Asn
865                 870                 875                 880

Lys Glu Thr Asp Tyr His Asp Leu Leu Asp Arg Lys Glu Lys Gly Lys
                885                 890                 895

His Val Ala Gln Gln Ala Trp Glu Thr Ile Glu Asn Ile Lys Glu Leu
            900                 905                 910

Lys Ala Gly Tyr Leu Ser Gln Val Val Tyr Lys Leu Thr Gln Leu Met
        915                 920                 925

Leu Gln Tyr Asn Ala Val Ile Val Leu Glu Asn Leu Asn Val Gly Phe
    930                 935                 940

Lys Arg Gly Arg Thr Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu
945                 950                 955                 960
```

```
Lys Ala Met Ile Asp Lys Leu Asn Tyr Leu Val Phe Lys Asp Arg Gly
                965                 970                 975

Tyr Glu Met Asn Gly Ser Tyr Ala Lys Gly Leu Gln Leu Thr Asp Lys
            980                 985                 990

Phe Glu Ser Phe Asp Lys Ile Gly Lys Gln Thr Gly Cys Ile Tyr Tyr
        995                 1000                1005

Val Ile  Pro Ser Tyr Thr Ser  His Ile Asp Pro Lys  Thr Gly Phe
    1010                1015                1020

Val Asn  Leu Leu Asn Ala Lys  Leu Arg Tyr Glu Asn  Ile Thr Lys
    1025                1030                1035

Ala Gln  Asp Thr Ile Arg Lys  Phe Asp Ser Ile Ser  Tyr Asn Ala
    1040                1045                1050

Lys Ala  Asp Tyr Phe Glu Phe  Ala Phe Asp Tyr Arg  Ser Phe Gly
    1055                1060                1065

Val Asp  Met Ala Arg Asn Glu  Trp Val Val Cys Thr  Cys Gly Asp
    1070                1075                1080

Leu Arg  Trp Glu Tyr Ser Ala  Lys Thr Arg Glu Thr  Lys Ala Tyr
    1085                1090                1095

Ser Val  Thr Asp Arg Leu Lys  Glu Leu Phe Lys Ala  His Gly Ile
    1100                1105                1110

Asp Tyr  Val Gly Gly Glu Asn  Leu Val Ser His Ile  Thr Glu Val
    1115                1120                1125

Ala Asp  Lys His Phe Leu Ser  Thr Leu Leu Phe Tyr  Leu Arg Leu
    1130                1135                1140

Val Leu  Lys Met Arg Tyr Thr  Val Ser Gly Thr Glu  Asn Glu Asn
    1145                1150                1155

Asp Phe  Ile Leu Ser Pro Val  Glu Tyr Ala Pro Gly  Lys Phe Phe
    1160                1165                1170

Asp Ser  Arg Glu Ala Thr Ser  Thr Glu Pro Met Asn  Ala Asp Ala
    1175                1180                1185

Asn Gly  Ala Tyr His Ile Ala  Leu Lys Gly Leu Met  Thr Ile Arg
    1190                1195                1200

Gly Ile  Glu Asp Gly Lys Leu  His Asn Tyr Gly Lys  Gly Gly Glu
    1205                1210                1215

Asn Ala  Ala Trp Phe Lys Phe  Met Gln Asn Gln Glu  Tyr Lys Asn
    1220                1225                1230

Asn Gly
    1235
```

<210> 106
<211> 1233
<212> PRT
<213> 未知

<220>

<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 106

```
Met Asp Tyr Gly Asn Gly Gln Phe Glu Arg Arg Ala Pro Leu Thr Lys
1               5                  10                  15

Thr Ile Thr Leu Arg Leu Lys Pro Ile Gly Glu Thr Arg Glu Thr Ile
            20                  25                  30

Arg Glu Gln Lys Leu Leu Glu Gln Asp Ala Ala Phe Arg Lys Leu Val
        35                  40                  45

Glu Thr Val Thr Pro Ile Val Asp Asp Cys Ile Arg Lys Ile Ala Asp
    50                  55                  60

Asn Ala Leu Cys His Phe Gly Thr Glu Tyr Asp Phe Ser Cys Leu Gly
65                  70                  75                  80

Asn Ala Ile Ser Lys Asn Asp Ser Lys Ala Ile Lys Lys Glu Thr Glu
                85                  90                  95

Lys Val Glu Lys Leu Leu Ala Lys Val Leu Thr Glu Asn Leu Pro Asp
            100                 105                 110

Gly Leu Arg Lys Val Asn Asp Ile Asn Ser Ala Ala Phe Ile Gln Asp
        115                 120                 125

Thr Leu Thr Ser Phe Val Gln Asp Asp Ala Asp Lys Arg Val Leu Ile
    130                 135                 140

Gln Glu Leu Lys Gly Lys Thr Val Leu Met Gln Arg Phe Leu Thr Thr
145                 150                 155                 160

Arg Ile Thr Ala Leu Thr Val Trp Leu Pro Asp Arg Val Phe Glu Asn
                165                 170                 175

Phe Asn Ile Phe Ile Glu Asn Ala Glu Lys Met Arg Ile Leu Leu Asp
            180                 185                 190

Ser Pro Leu Asn Glu Lys Ile Met Lys Phe Asp Pro Asp Ala Glu Gln
        195                 200                 205

Tyr Ala Ser Leu Glu Phe Tyr Gly Gln Cys Leu Ser Gln Lys Asp Ile
    210                 215                 220

Asp Ser Tyr Asn Leu Ile Ile Ser Gly Ile Tyr Ala Asp Asp Glu Val
225                 230                 235                 240

Lys Asn Pro Gly Ile Asn Glu Ile Val Lys Glu Tyr Asn Gln Gln Ile
                245                 250                 255

Arg Gly Asp Lys Asp Glu Ser Pro Leu Pro Lys Leu Lys Lys Leu His
            260                 265                 270

Lys Gln Ile Leu Met Pro Val Glu Lys Ala Phe Phe Val Arg Val Leu
        275                 280                 285

Ser Asn Asp Ser Asp Ala Arg Ser Ile Leu Glu Lys Ile Leu Lys Asp
    290                 295                 300

Thr Glu Met Leu Pro Ser Lys Ile Ile Glu Ala Met Lys Glu Ala Asp
305                 310                 315                 320
```

```
Ala Gly Asp Ile Ala Val Tyr Gly Ser Arg Leu His Glu Leu Ser His
                325                 330                 335

Val Ile Tyr Gly Asp His Gly Lys Leu Ser Gln Ile Ile Tyr Asp Lys
            340                 345                 350

Glu Ser Lys Arg Ile Ser Glu Leu Met Glu Thr Leu Ser Pro Lys Glu
        355                 360                 365

Arg Lys Glu Ser Lys Lys Arg Leu Glu Gly Leu Glu Glu His Ile Arg
    370                 375                 380

Lys Ser Thr Tyr Thr Phe Asp Glu Leu Asn Arg Tyr Ala Glu Lys Asn
385                 390                 395                 400

Val Met Ala Ala Tyr Ile Ala Ala Val Glu Glu Ser Cys Ala Glu Ile
                405                 410                 415

Met Arg Lys Glu Lys Asp Leu Arg Thr Leu Leu Ser Lys Glu Asp Val
            420                 425                 430

Lys Ile Arg Gly Asn Arg His Asn Thr Leu Ile Val Lys Asn Tyr Phe
        435                 440                 445

Asn Ala Trp Thr Val Phe Arg Asn Leu Ile Arg Ile Leu Arg Arg Lys
    450                 455                 460

Ser Glu Ala Glu Ile Asp Ser Asp Phe Tyr Asp Val Leu Asp Asp Ser
465                 470                 475                 480

Val Glu Val Leu Ser Leu Thr Tyr Lys Gly Glu Asn Leu Cys Arg Ser
                485                 490                 495

Tyr Ile Thr Lys Lys Ile Gly Ser Asp Leu Lys Pro Glu Ile Ala Thr
            500                 505                 510

Tyr Gly Ser Ala Leu Arg Pro Asn Ser Arg Trp Trp Ser Pro Gly Glu
        515                 520                 525

Lys Phe Asn Val Lys Phe His Thr Ile Val Arg Arg Asp Gly Arg Leu
    530                 535                 540

Tyr Tyr Phe Ile Leu Pro Lys Gly Ala Lys Pro Val Glu Leu Glu Asp
545                 550                 555                 560

Met Asp Gly Asp Ile Glu Cys Leu Gln Met Arg Lys Ile Pro Asn Pro
                565                 570                 575

Thr Ile Phe Leu Pro Lys Leu Val Phe Lys Asp Pro Glu Ala Phe Phe
            580                 585                 590

Arg Asp Asn Pro Glu Ala Asp Glu Phe Val Phe Leu Ser Gly Met Lys
        595                 600                 605

Ala Pro Val Thr Ile Thr Arg Glu Thr Tyr Glu Ala Tyr Arg Tyr Lys
    610                 615                 620

Leu Tyr Thr Val Gly Lys Leu Arg Asp Gly Glu Val Ser Glu Glu Glu
625                 630                 635                 640

Tyr Lys Arg Ala Leu Leu Gln Val Leu Thr Ala Tyr Lys Glu Phe Leu
                645                 650                 655
```

Glu Asn Arg Met Ile Tyr Ala Asp Leu Asn Phe Gly Phe Lys Asp Leu
660 665 670

Glu Glu Tyr Lys Asp Ser Ser Glu Phe Ile Lys Gln Val Glu Thr His
675 680 685

Asn Thr Phe Met Cys Trp Ala Lys Val Ser Ser Ser Gln Leu Asp Asp
690 695 700

Leu Val Lys Ser Gly Asn Gly Leu Leu Phe Glu Ile Trp Ser Glu Arg
705 710 715 720

Leu Glu Ser Tyr Tyr Lys Tyr Gly Asn Glu Lys Val Leu Arg Gly Tyr
725 730 735

Glu Gly Val Leu Leu Ser Ile Leu Lys Asp Glu Asn Leu Val Ser Met
740 745 750

Arg Thr Leu Leu Asn Ser Arg Pro Met Leu Val Tyr Arg Pro Lys Glu
755 760 765

Ser Ser Lys Pro Met Val Val His Arg Asp Gly Ser Arg Val Val Asp
770 775 780

Arg Phe Asp Lys Asp Gly Lys Tyr Ile Pro Pro Glu Val His Asp Glu
785 790 795 800

Leu Tyr Arg Phe Phe Asn Asn Leu Leu Ile Lys Glu Lys Leu Gly Glu
805 810 815

Lys Ala Arg Lys Ile Leu Asp Asn Lys Lys Val Lys Val Lys Val Leu
820 825 830

Glu Ser Glu Arg Val Lys Trp Ser Lys Phe Tyr Asp Glu Gln Phe Ala
835 840 845

Val Thr Phe Ser Val Lys Lys Asn Ala Asp Cys Leu Asp Thr Thr Lys
850 855 860

Asp Leu Asn Ala Glu Val Met Glu Gln Tyr Ser Glu Ser Asn Arg Leu
865 870 875 880

Ile Leu Ile Arg Asn Thr Thr Asp Ile Leu Tyr Tyr Leu Val Leu Asp
885 890 895

Lys Asn Gly Lys Val Leu Lys Gln Arg Ser Leu Asn Ile Ile Asn Asp
900 905 910

Gly Ala Arg Asp Val Asp Trp Lys Glu Arg Phe Arg Gln Val Thr Lys
915 920 925

Asp Arg Asn Glu Gly Tyr Asn Glu Trp Asp Tyr Ser Arg Thr Ser Asn
930 935 940

Asp Leu Lys Glu Val Tyr Leu Asn Tyr Ala Leu Lys Glu Ile Ala Glu
945 950 955 960

Ala Val Ile Glu Tyr Asn Ala Ile Leu Ile Ile Glu Lys Met Ser Asn
965 970 975

Ala Phe Lys Asp Lys Tyr Ser Phe Leu Asp Asp Val Thr Phe Lys Gly
980 985 990

Phe Glu Thr Lys Leu Leu Ala Lys Leu Ser Asp Leu His Phe Arg Gly
995 1000 1005

Ile Lys Asp Gly Glu Pro Cys Ser Phe Thr Asn Pro Leu Gln Leu
1010 1015 1020

Cys Gln Asn Asp Ser Asn Lys Ile Leu Gln Asp Gly Val Ile Phe
1025 1030 1035

Met Val Pro Asn Ser Met Thr Arg Ser Leu Asp Pro Asp Thr Gly
1040 1045 1050

Phe Ile Phe Ala Ile Asn Asp His Asn Ile Arg Thr Lys Lys Ala
1055 1060 1065

Lys Leu Asn Phe Leu Ser Lys Phe Asp Gln Leu Lys Val Ser Ser
1070 1075 1080

Glu Gly Cys Leu Ile Met Lys Tyr Ser Gly Asp Ser Leu Pro Thr
1085 1090 1095

His Asn Thr Asp Asn Arg Val Trp Asn Cys Cys Cys Asn His Pro
1100 1105 1110

Ile Thr Asn Tyr Asp Arg Glu Thr Lys Lys Val Glu Phe Ile Glu
1115 1120 1125

Glu Pro Val Glu Glu Leu Ser Arg Val Leu Glu Glu Asn Gly Ile
1130 1135 1140

Glu Thr Asp Thr Glu Leu Asn Lys Leu Asn Glu Arg Glu Asn Val
1145 1150 1155

Pro Gly Lys Val Val Asp Ala Ile Tyr Ser Leu Val Leu Asn Tyr
1160 1165 1170

Leu Arg Gly Thr Val Ser Gly Val Ala Gly Gln Arg Ala Val Tyr
1175 1180 1185

Tyr Ser Pro Val Thr Gly Lys Lys Tyr Asp Ile Ser Phe Ile Gln
1190 1195 1200

Ala Met Asn Leu Asn Arg Lys Cys Asp Tyr Tyr Arg Ile Gly Ser
1205 1210 1215

Lys Glu Arg Gly Glu Trp Thr Asp Phe Val Ala Gln Leu Ile Asn
1220 1225 1230

<210> 107
<211> 1231
<212> PRT
<213> 溶纤维丁酸弧菌（Butyrivibrio fibrisolvens）

<400> 107
Met Tyr Tyr Glu Ser Leu Thr Lys Leu Tyr Pro Ile Lys Lys Thr Ile
1 5 10 15

Arg Asn Glu Leu Val Pro Ile Gly Lys Thr Leu Glu Asn Ile Lys Lys
20 25 30

Asn Asn Ile Leu Glu Ala Asp Glu Asp Arg Lys Ile Ala Tyr Ile Arg
35 40 45

Val Lys Ala Ile Met Asp Asp Tyr His Lys Arg Leu Ile Asn Glu Ala
50 55 60

Leu Ser Gly Phe Ala Leu Ile Asp Leu Asp Lys Ala Ala Asn Leu Tyr
65 70 75 80

Leu Ser Arg Ser Lys Ser Ala Asp Asp Ile Glu Ser Phe Ser Arg Phe
85 90 95

Gln Asp Lys Leu Arg Lys Ala Ile Ala Lys Arg Leu Arg Glu His Glu
100 105 110

Asn Phe Gly Lys Ile Gly Asn Lys Asp Ile Ile Pro Leu Leu Gln Lys
115 120 125

Leu Ser Glu Asn Glu Asp Asp Tyr Asn Ala Leu Glu Ser Phe Lys Asn
130 135 140

Phe Tyr Thr Tyr Phe Glu Ser Tyr Asn Asp Val Arg Leu Asn Leu Tyr
145 150 155 160

Ser Asp Lys Glu Lys Ser Ser Thr Val Ala Tyr Arg Leu Ile Asn Glu
165 170 175

Asn Leu Pro Arg Phe Leu Asp Asn Ile Arg Ala Tyr Asp Ala Val Gln
180 185 190

Lys Ala Gly Ile Thr Ser Glu Glu Leu Ser Ser Glu Ala Gln Asp Gly
195 200 205

Leu Phe Leu Val Asn Thr Phe Asn Asn Val Leu Ile Gln Asp Gly Ile
210 215 220

Asn Thr Tyr Asn Glu Asp Ile Gly Lys Leu Asn Val Ala Ile Asn Leu
225 230 235 240

Tyr Asn Gln Lys Asn Ala Ser Val Gln Gly Phe Arg Lys Val Pro Lys
245 250 255

Met Lys Val Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Glu Ser Phe
260 265 270

Ile Asp Glu Phe Glu Ser Asp Thr Glu Leu Leu Asp Ser Leu Glu Ser
275 280 285

His Tyr Ala Asn Leu Ala Lys Tyr Phe Gly Ser Asn Lys Val Gln Leu
290 295 300

Leu Phe Thr Ala Leu Arg Glu Ser Lys Gly Val Asn Val Tyr Val Lys
305 310 315 320

Asn Asp Ile Ala Lys Thr Ser Phe Ser Asn Val Val Phe Gly Ser Trp
325 330 335

Ser Arg Ile Asp Glu Leu Ile Asn Gly Glu Tyr Asp Asp Asn Asn Asn
340 345 350

Arg Lys Lys Asp Glu Lys Tyr Tyr Asp Lys Arg Gln Lys Glu Leu Lys
355 360 365

Lys Asn Lys Ser Tyr Thr Ile Glu Lys Ile Ile Thr Leu Ser Thr Glu
370 375 380

Asp Val Asp Val Ile Gly Lys Tyr Ile Glu Lys Leu Glu Ser Asp Ile
385 390 395 400

```
Asp Asp Ile Arg Phe Lys Gly Lys Asn Phe Tyr Glu Ala Val Leu Cys
                405                 410                 415

Gly His Asp Arg Ser Lys Lys Leu Ser Lys Asn Lys Gly Ala Val Glu
            420                 425                 430

Ala Ile Lys Gly Tyr Leu Asp Ser Val Lys Asp Phe Glu Arg Asp Leu
        435                 440                 445

Lys Leu Ile Asn Gly Ser Gly Gln Glu Leu Glu Lys Asn Leu Val Val
    450                 455                 460

Tyr Gly Glu Gln Glu Ala Val Leu Ser Glu Leu Ser Gly Ile Asp Ser
465                 470                 475                 480

Leu Tyr Asn Met Thr Arg Asn Tyr Leu Thr Lys Lys Pro Phe Ser Thr
                485                 490                 495

Glu Lys Ile Lys Leu Asn Phe Asn Lys Pro Thr Phe Leu Asp Gly Trp
            500                 505                 510

Asp Tyr Gly Asn Glu Glu Ala Tyr Leu Gly Phe Phe Met Ile Lys Glu
        515                 520                 525

Gly Asn Tyr Phe Leu Ala Val Met Asp Ala Asn Trp Asn Lys Glu Phe
    530                 535                 540

Arg Asn Ile Pro Ser Val Asp Lys Ser Asp Cys Tyr Lys Lys Val Ile
545                 550                 555                 560

Tyr Lys Gln Ile Ser Ser Pro Glu Lys Ser Ile Gln Asn Leu Met Val
                565                 570                 575

Ile Asp Gly Lys Thr Val Lys Lys Asn Gly Arg Lys Glu Lys Glu Gly
            580                 585                 590

Ile His Ser Gly Glu Asn Leu Ile Leu Glu Glu Leu Lys Asn Thr Tyr
        595                 600                 605

Leu Pro Lys Lys Ile Asn Asp Ile Arg Lys Arg Arg Ser Tyr Leu Asn
    610                 615                 620

Gly Asp Thr Phe Ser Lys Lys Asp Leu Thr Glu Phe Ile Gly Tyr Tyr
625                 630                 635                 640

Lys Gln Arg Val Ile Glu Tyr Tyr Asn Gly Tyr Ser Phe Tyr Phe Lys
                645                 650                 655

Ser Asp Asp Asp Tyr Ala Ser Phe Lys Glu Phe Gln Glu Asp Val Gly
            660                 665                 670

Arg Gln Ala Tyr Gln Ile Ser Tyr Val Asp Val Pro Val Ser Phe Val
        675                 680                 685

Asp Asp Leu Ile Asn Ser Gly Lys Leu Tyr Leu Phe Arg Val Tyr Asn
    690                 695                 700

Lys Asp Phe Ser Glu Tyr Ser Lys Gly Arg Leu Asn Leu His Thr Leu
705                 710                 715                 720

Tyr Phe Lys Met Leu Phe Asp Glu Arg Asn Leu Lys Asn Val Val Tyr
                725                 730                 735
```

Lys Leu Asn Gly Gln Ala Glu Val Phe Tyr Arg Pro Ser Ser Ile Lys
740 745 750

Lys Glu Glu Leu Ile Val His Arg Ala Gly Glu Glu Ile Lys Asn Lys
755 760 765

Asn Pro Lys Arg Ala Ala Gln Lys Pro Thr Arg Arg Leu Asp Tyr Asp
770 775 780

Ile Val Lys Asp Arg Arg Tyr Ser Gln Asp Lys Phe Met Leu His Thr
785 790 795 800

Ser Ile Ile Met Asn Phe Gly Ala Glu Glu Asn Val Ser Phe Asn Asp
805 810 815

Ile Val Asn Gly Val Leu Arg Asn Glu Asp Lys Val Asn Val Ile Gly
820 825 830

Ile Asp Arg Gly Glu Arg Asn Leu Leu Tyr Val Val Val Ile Asp Pro
835 840 845

Glu Gly Lys Ile Leu Glu Gln Arg Ser Leu Asn Cys Ile Thr Asp Ser
850 855 860

Asn Leu Asp Ile Glu Thr Asp Tyr His Arg Leu Leu Asp Glu Lys Glu
865 870 875 880

Ser Asp Arg Lys Ile Ala Arg Arg Asp Trp Thr Thr Ile Glu Asn Ile
885 890 895

Lys Glu Leu Lys Ala Gly Tyr Leu Ser Gln Val Val His Ile Val Ala
900 905 910

Glu Leu Val Leu Lys Tyr Asn Ala Ile Ile Cys Leu Glu Asp Leu Asn
915 920 925

Phe Gly Phe Lys Arg Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln
930 935 940

Lys Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr Leu Val Met Asp
945 950 955 960

Lys Ser Arg Glu Gln Leu Ser Pro Glu Lys Ile Ser Gly Ala Leu Asn
965 970 975

Ala Leu Gln Leu Thr Pro Asp Phe Lys Ser Phe Lys Val Leu Gly Lys
980 985 990

Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Tyr Leu Thr Ser Lys Ile
995 1000 1005

Asp Pro Met Thr Gly Phe Ala Asn Leu Phe Tyr Val Lys Tyr Glu
1010 1015 1020

Asn Val Asp Lys Ala Lys Glu Phe Phe Ser Lys Phe Asp Ser Ile
1025 1030 1035

Lys Tyr Asn Lys Asp Gly Lys Asn Trp Asn Thr Lys Gly Tyr Phe
1040 1045 1050

Glu Phe Ala Phe Asp Tyr Lys Lys Phe Thr Asp Arg Ala Tyr Gly
1055 1060 1065

```
Arg Val  Ser Glu Trp Thr Val  Cys Thr Val Gly Glu  Arg Ile Ile
    1070                 1075                 1080

Lys Phe  Lys Asn Lys Glu Lys  Asn Asn Ser Tyr Asp  Asp Lys Val
    1085                 1090                 1095

Ile Asp  Leu Thr Asn Ser Leu  Lys Glu Leu Phe Asp  Ser Tyr Lys
    1100                 1105                 1110

Val Thr  Tyr Glu Ser Glu Val  Asp Leu Lys Asp Ala  Ile Leu Ala
    1115                 1120                 1125

Ile Asp  Asp Pro Ala Phe Tyr  Arg Asp Leu Thr Arg  Arg Leu Gln
    1130                 1135                 1140

Gln Thr  Leu Gln Met Arg Asn  Ser Ser Cys Asp Gly  Ser Arg Asp
    1145                 1150                 1155

Tyr Ile  Ile Ser Pro Val Lys  Asn Ser Lys Gly Glu  Phe Phe Cys
    1160                 1165                 1170

Ser Asp  Asn Asn Asp Asp Thr  Thr Pro Asn Asp Ala  Asp Ala Asn
    1175                 1180                 1185

Gly Ala  Phe Asn Ile Ala Arg  Lys Gly Leu Trp Val  Leu Asn Glu
    1190                 1195                 1200

Ile Arg  Asn Ser Glu Glu Gly  Ser Lys Ile Asn Leu  Ala Met Ser
    1205                 1210                 1215

Asn Ala  Gln Trp Leu Glu Tyr  Ala Gln Asp Asn Thr  Ile
    1220                 1225                 1230
```

<210> 108
<211> 1230
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 108

```
Met His Glu Asn Asn Gly Lys Ile Ala Asp Asn Phe Ile Gly Ile Tyr
1               5                  10                  15

Pro Val Ser Lys Thr Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr
            20                  25                  30

Gln Glu Tyr Ile Glu Lys His Gly Ile Leu Asp Glu Asp Leu Lys Arg
        35                  40                  45

Ala Gly Asp Tyr Lys Ser Val Lys Lys Ile Ile Asp Ala Tyr His Lys
    50                  55                  60

Tyr Phe Ile Asp Glu Ala Leu Asn Gly Ile Gln Leu Asp Gly Leu Lys
65                  70                  75                  80

Asn Tyr Tyr Glu Leu Tyr Glu Lys Lys Arg Asp Asn Asn Glu Glu Lys
                85                  90                  95

Glu Phe Gln Lys Ile Gln Met Ser Leu Arg Lys Gln Ile Val Lys Arg
            100                 105                 110
```

```
Phe Ser Glu His Pro Gln Tyr Lys Tyr Leu Phe Lys Lys Glu Leu Ile
        115                 120                 125

Lys Asn Val Leu Pro Glu Phe Thr Lys Asp Asn Ala Glu Glu Gln Thr
    130                 135                 140

Leu Val Lys Ser Phe Gln Glu Phe Thr Thr Tyr Phe Glu Gly Phe His
145                 150                 155                 160

Gln Asn Arg Lys Asn Met Tyr Ser Asp Glu Glu Lys Ser Thr Ala Ile
                165                 170                 175

Ala Tyr Arg Val Val His Gln Asn Leu Pro Lys Tyr Ile Asp Asn Met
            180                 185                 190

Arg Ile Phe Ser Met Ile Leu Asn Thr Asp Ile Arg Ser Asp Leu Thr
        195                 200                 205

Glu Leu Phe Asn Asn Leu Lys Thr Lys Met Asp Ile Thr Ile Val Glu
    210                 215                 220

Glu Tyr Phe Ala Ile Asp Gly Phe Asn Lys Val Val Asn Gln Lys Gly
225                 230                 235                 240

Ile Asp Val Tyr Asn Thr Ile Leu Gly Ala Phe Ser Thr Asp Asp Asn
                245                 250                 255

Thr Lys Ile Lys Gly Leu Asn Glu Tyr Ile Asn Leu Tyr Asn Gln Lys
            260                 265                 270

Asn Lys Ala Lys Leu Pro Lys Leu Lys Pro Leu Phe Lys Gln Ile Leu
        275                 280                 285

Ser Asp Arg Asp Lys Ile Ser Phe Ile Pro Glu Gln Phe Asp Ser Asp
    290                 295                 300

Thr Glu Val Leu Glu Ala Val Asp Met Phe Tyr Asn Arg Leu Leu Gln
305                 310                 315                 320

Phe Val Ile Glu Asn Glu Gly Gln Ile Thr Ile Ser Lys Leu Leu Thr
                325                 330                 335

Asn Phe Ser Ala Tyr Asp Leu Asn Lys Ile Tyr Val Lys Asn Asp Thr
            340                 345                 350

Thr Ile Ser Ala Ile Ser Asn Asp Leu Phe Asp Asp Trp Ser Tyr Ile
        355                 360                 365

Ser Lys Ala Val Arg Glu Asn Tyr Asp Ser Glu Asn Val Asp Lys Asn
    370                 375                 380

Lys Arg Ala Ala Ala Tyr Glu Glu Lys Lys Glu Lys Ala Leu Ser Lys
385                 390                 395                 400

Ile Lys Met Tyr Ser Ile Glu Glu Leu Asn Phe Phe Val Lys Lys Tyr
                405                 410                 415

Ser Cys Asn Glu Cys His Ile Glu Gly Tyr Phe Glu Arg Arg Ile Leu
            420                 425                 430

Glu Ile Leu Asp Lys Met Arg Tyr Ala Tyr Glu Ser Cys Lys Ile Leu
        435                 440                 445
```

```
His Asp Lys Gly Leu Ile Asn Asn Ile Ser Leu Cys Gln Asp Arg Gln
    450                 455                 460

Ala Ile Ser Glu Leu Lys Asp Phe Leu Asp Ser Ile Lys Glu Val Gln
465                 470                 475                 480

Trp Leu Leu Lys Pro Leu Met Ile Gly Gln Glu Gln Ala Asp Lys Glu
                485                 490                 495

Glu Ala Phe Tyr Thr Glu Leu Leu Arg Ile Trp Glu Glu Leu Glu Pro
            500                 505                 510

Ile Thr Leu Leu Tyr Asn Lys Val Arg Asn Tyr Val Thr Lys Lys Pro
        515                 520                 525

Tyr Thr Leu Glu Lys Val Lys Leu Asn Phe Tyr Lys Ser Thr Leu Leu
    530                 535                 540

Asp Gly Trp Asp Lys Asn Lys Glu Lys Asp Asn Leu Gly Ile Ile Leu
545                 550                 555                 560

Leu Lys Asp Gly Gln Tyr Tyr Leu Gly Ile Met Asn Arg Arg Asn Asn
                565                 570                 575

Lys Ile Ala Asp Asp Ala Pro Leu Ala Lys Thr Asp Asn Val Tyr Arg
            580                 585                 590

Lys Met Glu Tyr Lys Leu Leu Thr Lys Val Ser Ala Asn Leu Pro Arg
        595                 600                 605

Ile Phe Leu Lys Asp Lys Tyr Asn Pro Ser Glu Glu Met Leu Glu Lys
    610                 615                 620

Tyr Glu Lys Gly Thr His Leu Lys Gly Glu Asn Phe Cys Ile Asp Asp
625                 630                 635                 640

Cys Arg Glu Leu Ile Asp Phe Phe Lys Lys Gly Ile Lys Gln Tyr Glu
                645                 650                 655

Asp Trp Gly Gln Phe Asp Phe Lys Phe Ser Asp Thr Glu Ser Tyr Asp
            660                 665                 670

Asp Ile Ser Ala Phe Tyr Lys Glu Val Glu His Gln Gly Tyr Lys Ile
        675                 680                 685

Thr Phe Arg Asp Ile Asp Glu Thr Tyr Ile Asp Ser Leu Val Asn Glu
    690                 695                 700

Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Pro Tyr
705                 710                 715                 720

Ser Lys Gly Thr Lys Asn Leu His Thr Leu Tyr Trp Glu Met Leu Phe
                725                 730                 735

Ser Gln Gln Asn Leu Gln Asn Ile Val Tyr Lys Leu Asn Gly Asn Ala
            740                 745                 750

Glu Ile Phe Tyr Arg Lys Ala Ser Ile Asn Gln Lys Asp Val Val Val
        755                 760                 765

His Lys Ala Asp Leu Pro Ile Lys Asn Lys Asp Pro Gln Asn Ser Lys
    770                 775                 780
```

```
Lys Glu Ser Met Phe Asp Tyr Asp Ile Ile Lys Asp Lys Arg Phe Thr
785                 790                 795                 800

Cys Asp Lys Tyr Gln Phe His Val Pro Ile Thr Met Asn Phe Lys Ala
                805                 810                 815

Leu Gly Glu Asn His Phe Asn Arg Lys Val Asn Arg Leu Ile His Asp
            820                 825                 830

Ala Glu Asn Met His Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu
        835                 840                 845

Ile Tyr Leu Cys Met Ile Asp Met Lys Gly Asn Ile Val Lys Gln Ile
    850                 855                 860

Ser Leu Asn Glu Ile Ile Ser Tyr Asp Lys Asn Lys Leu Glu His Lys
865                 870                 875                 880

Arg Asn Tyr His Gln Leu Leu Lys Thr Arg Glu Asp Glu Asn Lys Ser
                885                 890                 895

Ala Arg Gln Ser Trp Gln Thr Ile His Thr Ile Lys Glu Leu Lys Glu
            900                 905                 910

Gly Tyr Leu Ser Gln Val Ile His Val Ile Thr Asp Leu Met Val Glu
        915                 920                 925

Tyr Asn Ala Ile Val Val Leu Glu Asp Leu Asn Phe Gly Phe Lys Gln
    930                 935                 940

Gly Arg Gln Lys Phe Glu Arg Gln Val Tyr Gln Lys Phe Glu Lys Met
945                 950                 955                 960

Leu Ile Asp Lys Leu Asn Tyr Leu Val Asp Lys Ser Lys Gly Met Asp
                965                 970                 975

Glu Asp Gly Gly Leu Leu His Ala Tyr Gln Leu Thr Asp Glu Phe Lys
            980                 985                 990

Ser Phe Lys Gln Leu Gly Lys Gln Ser Gly Phe Leu Tyr Tyr Ile Pro
        995                 1000                1005

Ala Trp  Asn Thr Ser Lys Leu Asp Pro Thr Thr  Gly Phe Val Asn
    1010                1015                1020

Leu Phe  Tyr Thr Lys Tyr Glu Ser Val Glu Lys  Ser Lys Glu Phe
    1025                1030                1035

Ile Asn  Asn Phe Thr Ser Ile Leu Tyr Asn Gln  Glu Arg Glu Tyr
    1040                1045                1050

Phe Glu  Phe Leu Phe Asp Tyr Ser Ala Phe Thr  Ser Lys Ala Glu
    1055                1060                1065

Gly Ser  Arg Leu Lys Trp Thr Val Cys Ser Lys  Gly Glu Arg Val
    1070                1075                1080

Glu Thr  Tyr Arg Asn Pro Lys Lys Asn Asn Glu  Trp Asp Thr Gln
    1085                1090                1095

Lys Ile  Asp Leu Thr Phe Glu Leu Lys Lys Leu  Phe Asn Asp Tyr
    1100                1105                1110
```

Ser Ile Ser Leu Leu Asp Gly Asp Leu Arg Glu Gln Met Gly Lys
1115 1120 1125

Ile Asp Lys Ala Asp Phe Tyr Lys Lys Phe Met Lys Leu Phe Ala
1130 1135 1140

Leu Ile Val Gln Met Arg Asn Ser Asp Glu Arg Glu Asp Lys Leu
1145 1150 1155

Ile Ser Pro Val Leu Asn Lys Tyr Gly Ala Phe Phe Glu Thr Gly
1160 1165 1170

Lys Asn Glu Arg Met Pro Leu Asp Ala Asp Ala Asn Gly Ala Tyr
1175 1180 1185

Asn Ile Ala Arg Lys Gly Leu Trp Ile Ile Glu Lys Ile Lys Asn
1190 1195 1200

Thr Asp Val Glu Gln Leu Asp Lys Val Lys Leu Thr Ile Ser Asn
1205 1210 1215

Lys Glu Trp Leu Gln Tyr Ala Gln Glu His Ile Leu
1220 1225 1230

<210> 109
<211> 1228
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 109
Met Ser Lys Leu Glu Lys Phe Thr Asn Cys Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Arg Phe Lys Ala Ile Pro Val Gly Lys Thr Gln Glu Asn Ile Asp
20 25 30

Asn Lys Arg Leu Leu Val Glu Asp Glu Lys Arg Ala Glu Asp Tyr Lys
35 40 45

Gly Val Lys Lys Leu Leu Asp Arg Tyr Tyr Leu Ser Phe Ile Asn Asp
50 55 60

Val Leu His Ser Ile Lys Leu Lys Asn Leu Asn Asn Tyr Ile Ser Leu
65 70 75 80

Phe Arg Lys Lys Thr Arg Thr Glu Lys Glu Asn Lys Glu Leu Glu Asn
85 90 95

Leu Glu Ile Asn Leu Arg Lys Glu Ile Ala Lys Ala Phe Lys Gly Asn
100 105 110

Glu Gly Tyr Lys Ser Leu Phe Lys Lys Asp Ile Ile Glu Thr Ile Leu
115 120 125

Pro Glu Phe Leu Asp Asp Lys Asp Glu Ile Ala Leu Val Asn Ser Phe
130 135 140

Asn Gly Phe Thr Thr Ala Phe Thr Gly Phe Phe Asp Asn Arg Glu Asn
145 150 155 160

Met Phe Ser Glu Glu Ala Lys Ser Thr Ser Ile Ala Phe Arg Cys Ile
165 170 175

Asn Glu Asn Leu Thr Arg Tyr Ile Ser Asn Met Asp Ile Phe Glu Lys
180 185 190

Val Asp Ala Ile Phe Asp Lys His Glu Val Gln Glu Ile Lys Glu Lys
195 200 205

Ile Leu Asn Ser Asp Tyr Asp Val Glu Asp Phe Phe Glu Gly Glu Phe
210 215 220

Phe Asn Phe Val Leu Thr Gln Glu Gly Ile Asp Val Tyr Asn Ala Ile
225 230 235 240

Ile Gly Gly Phe Val Thr Glu Ser Gly Glu Lys Ile Lys Gly Leu Asn
245 250 255

Glu Tyr Ile Asn Leu Tyr Asn Gln Lys Thr Lys Gln Lys Leu Pro Lys
260 265 270

Phe Lys Pro Leu Tyr Lys Gln Val Leu Ser Asp Arg Glu Ser Leu Ser
275 280 285

Phe Tyr Gly Glu Gly Tyr Thr Ser Asp Glu Glu Val Leu Glu Val Phe
290 295 300

Arg Asn Thr Leu Asn Lys Asn Ser Glu Ile Phe Ser Ser Ile Lys Lys
305 310 315 320

Leu Glu Lys Leu Phe Lys Asn Phe Asp Glu Tyr Ser Ser Ala Gly Ile
325 330 335

Phe Val Lys Asn Gly Pro Ala Ile Ser Thr Ile Ser Lys Asp Ile Phe
340 345 350

Gly Glu Trp Asn Val Ile Arg Asp Lys Trp Asn Ala Glu Tyr Asp Asp
355 360 365

Ile His Leu Lys Lys Lys Ala Val Val Thr Glu Lys Tyr Glu Asp Asp
370 375 380

Arg Arg Lys Ser Phe Lys Lys Ile Gly Ser Phe Ser Leu Glu Gln Leu
385 390 395 400

Gln Glu Tyr Ala Asp Ala Asp Leu Ser Val Val Glu Lys Leu Lys Glu
405 410 415

Ile Ile Ile Gln Lys Val Asp Glu Ile Tyr Lys Val Tyr Gly Ser Ser
420 425 430

Glu Lys Leu Phe Asp Ala Asp Phe Val Leu Glu Lys Ser Leu Lys Lys
435 440 445

Asn Asp Ala Val Val Ala Ile Met Lys Asp Leu Leu Asp Ser Val Lys
450 455 460

Ser Phe Glu Asn Tyr Ile Lys Ala Phe Phe Gly Glu Gly Lys Glu Thr
465 470 475 480

Asn Arg Asp Glu Ser Phe Tyr Gly Asp Phe Val Leu Ala Tyr Asp Ile
485 490 495

Leu Leu Lys Val Asp His Ile Tyr Asp Ala Ile Arg Asn Tyr Val Thr
500 505 510

```
Gln Lys Pro Tyr Ser Lys Asp Lys Phe Lys Leu Tyr Phe Gln Asn Pro
        515                 520                 525

Gln Phe Met Gly Gly Trp Asp Lys Asp Lys Glu Thr Asp Tyr Arg Ala
    530                 535                 540

Thr Ile Leu Arg Tyr Gly Ser Lys Tyr Tyr Leu Ala Ile Met Asp Lys
545                 550                 555                 560

Lys Tyr Ala Lys Cys Leu Gln Lys Ile Asp Lys Asp Asp Val Asn Gly
                565                 570                 575

Asn Tyr Glu Lys Ile Asn Tyr Lys Leu Leu Pro Gly Pro Asn Lys Met
            580                 585                 590

Leu Pro Lys Val Phe Phe Ser Lys Lys Trp Met Ala Tyr Tyr Asn Pro
        595                 600                 605

Ser Glu Asp Ile Gln Lys Ile Tyr Lys Asn Gly Thr Phe Lys Lys Gly
    610                 615                 620

Asp Met Phe Asn Leu Asn Asp Cys His Lys Leu Ile Asp Phe Phe Lys
625                 630                 635                 640

Asp Ser Ile Ser Arg Tyr Pro Lys Trp Ser Asn Ala Tyr Asp Phe Asn
                645                 650                 655

Phe Ser Glu Thr Glu Lys Tyr Lys Asp Ile Ala Gly Phe Tyr Arg Glu
            660                 665                 670

Val Glu Glu Gln Gly Tyr Lys Val Ser Phe Glu Ser Ala Ser Lys Lys
        675                 680                 685

Glu Val Asp Lys Leu Val Glu Glu Gly Lys Leu Tyr Met Phe Gln Ile
    690                 695                 700

Tyr Asn Lys Asp Phe Ser Asp Lys Ser His Gly Thr Pro Asn Leu His
705                 710                 715                 720

Thr Met Tyr Phe Lys Leu Leu Phe Asp Glu Asn Asn His Gly Gln Ile
                725                 730                 735

Arg Leu Ser Gly Gly Ala Glu Leu Phe Met Arg Arg Ala Ser Leu Lys
            740                 745                 750

Lys Glu Glu Leu Val Val His Pro Ala Asn Ser Pro Ile Ala Asn Lys
        755                 760                 765

Asn Pro Asp Asn Pro Lys Lys Thr Thr Thr Leu Ser Tyr Asp Val Tyr
    770                 775                 780

Lys Asp Lys Arg Phe Ser Glu Asp Gln Tyr Glu Leu His Ile Pro Ile
785                 790                 795                 800

Ala Ile Asn Lys Cys Pro Lys Asn Ile Phe Lys Ile Asn Thr Glu Val
                805                 810                 815

Arg Val Leu Leu Lys His Asp Asp Asn Pro Tyr Val Ile Gly Ile Asp
            820                 825                 830

Arg Gly Glu Arg Asn Leu Leu Tyr Ile Val Val Val Asp Gly Lys Gly
        835                 840                 845
```

Asn Ile Val Glu Gln Tyr Ser Leu Asn Glu Ile Ile Asn Asn Phe Asn
850 855 860

Gly Ile Arg Ile Lys Thr Asp Tyr His Ser Leu Leu Asp Lys Lys Glu
865 870 875 880

Lys Glu Arg Phe Glu Ala Arg Gln Asn Trp Thr Ser Ile Glu Asn Ile
885 890 895

Lys Glu Leu Lys Ala Gly Tyr Ile Ser Gln Val Val His Lys Ile Cys
900 905 910

Glu Leu Val Glu Lys Tyr Asp Ala Val Ile Ala Leu Glu Asp Leu Asn
915 920 925

Ser Gly Phe Lys Asn Ser Arg Val Lys Val Glu Lys Gln Val Tyr Gln
930 935 940

Lys Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr Met Val Asp Lys
945 950 955 960

Lys Ser Asn Pro Cys Ala Thr Gly Gly Ala Leu Lys Gly Tyr Gln Ile
965 970 975

Thr Asn Lys Phe Glu Ser Phe Lys Ser Met Ser Thr Gln Asn Gly Phe
980 985 990

Ile Phe Tyr Ile Pro Ala Trp Leu Thr Ser Lys Ile Asp Pro Ser Thr
995 1000 1005

Gly Phe Val Asn Leu Leu Lys Thr Lys Tyr Thr Ser Ile Ala Asp
1010 1015 1020

Ser Lys Lys Phe Ile Ser Ser Phe Asp Arg Ile Met Tyr Val Pro
1025 1030 1035

Glu Glu Asp Leu Phe Glu Phe Ala Leu Asp Tyr Lys Asn Phe Ser
1040 1045 1050

Arg Thr Asp Ala Asp Tyr Ile Lys Lys Trp Lys Leu Tyr Ser Tyr
1055 1060 1065

Gly Asn Arg Ile Arg Ile Phe Arg Asn Pro Lys Lys Asn Asn Val
1070 1075 1080

Phe Asp Trp Glu Glu Val Cys Leu Thr Ser Ala Tyr Lys Glu Leu
1085 1090 1095

Phe Asn Lys Tyr Gly Ile Asn Tyr Gln Gln Gly Asp Ile Arg Ala
1100 1105 1110

Leu Leu Cys Glu Gln Ser Asp Lys Ala Phe Tyr Ser Ser Phe Met
1115 1120 1125

Ala Leu Met Ser Leu Met Leu Gln Met Arg Asn Ser Ile Thr Gly
1130 1135 1140

Arg Thr Asp Val Asp Phe Leu Ile Ser Pro Val Lys Asn Ser Asp
1145 1150 1155

Gly Ile Phe Tyr Asp Ser Arg Asn Tyr Glu Ala Gln Glu Asn Ala
1160 1165 1170

Ile Leu Pro Lys Asn Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala
1175 1180 1185

Arg Lys Val Leu Trp Ala Ile Gly Gln Phe Lys Lys Ala Glu Asp
1190 1195 1200

Glu Lys Leu Asp Lys Val Lys Ile Ala Ile Ser Asn Lys Glu Trp
1205 1210 1215

Leu Glu Tyr Ala Gln Thr Ser Val Lys His
1220 1225

<210> 110
<211> 1227
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候选alvus甲烷嗜甲基菌（Candidatus Methanomethylophilus alvus）序列

<400> 110
Met Asp Ala Lys Glu Phe Thr Gly Gln Tyr Pro Leu Ser Lys Thr Leu
1 5 10 15

Arg Phe Glu Leu Arg Pro Ile Gly Arg Thr Trp Asp Asn Leu Glu Ala
20 25 30

Ser Gly Tyr Leu Ala Glu Asp Arg His Arg Ala Glu Cys Tyr Pro Arg
35 40 45

Ala Lys Glu Leu Leu Asp Asp Asn His Arg Ala Phe Leu Asn Arg Val
50 55 60

Leu Pro Gln Ile Asp Met Asp Trp His Pro Ile Ala Glu Ala Phe Cys
65 70 75 80

Lys Val His Lys Asn Pro Gly Asn Lys Glu Leu Ala Gln Asp Tyr Asn
85 90 95

Leu Gln Leu Ser Lys Arg Arg Lys Glu Ile Ser Ala Tyr Leu Gln Asp
100 105 110

Ala Asp Gly Tyr Lys Gly Leu Phe Ala Lys Pro Ala Leu Asp Glu Ala
115 120 125

Met Lys Ile Ala Lys Glu Asn Gly Asn Glu Ser Asp Ile Glu Val Leu
130 135 140

Glu Ala Phe Asn Gly Phe Ser Val Tyr Phe Thr Gly Tyr His Glu Ser
145 150 155 160

Arg Glu Asn Ile Tyr Ser Asp Glu Asp Met Val Ser Val Ala Tyr Arg
165 170 175

Ile Thr Glu Asp Asn Phe Pro Arg Phe Val Ser Asn Ala Leu Ile Phe
180 185 190

Asp Lys Leu Asn Glu Ser His Pro Asp Ile Ile Ser Glu Val Ser Gly
195 200 205

Asn Leu Gly Val Asp Asp Ile Gly Lys Tyr Phe Asp Val Ser Asn Tyr
210 215 220

```
Asn Asn Phe Leu Ser Gln Ala Gly Ile Asp Asp Tyr Asn His Ile Ile
225                 230                 235                 240

Gly Gly His Thr Thr Glu Asp Gly Leu Ile Gln Ala Phe Asn Val Val
                245                 250                 255

Leu Asn Leu Arg His Gln Lys Asp Pro Gly Phe Glu Lys Ile Gln Phe
            260                 265                 270

Lys Gln Leu Tyr Lys Gln Ile Leu Ser Val Arg Thr Ser Lys Ser Tyr
        275                 280                 285

Ile Pro Lys Gln Phe Asp Asn Ser Lys Glu Met Val Asp Cys Ile Cys
    290                 295                 300

Asp Tyr Val Ser Lys Ile Glu Lys Ser Glu Thr Val Glu Arg Ala Leu
305                 310                 315                 320

Lys Leu Val Arg Asn Ile Ser Ser Phe Asp Leu Arg Gly Ile Phe Val
                325                 330                 335

Asn Lys Lys Asn Leu Arg Ile Leu Ser Asn Lys Leu Ile Gly Asp Trp
            340                 345                 350

Asp Ala Ile Glu Thr Ala Leu Met His Ser Ser Ser Ser Glu Asn Asp
        355                 360                 365

Lys Lys Ser Val Tyr Asp Ser Ala Glu Ala Phe Thr Leu Asp Asp Ile
    370                 375                 380

Phe Ser Ser Val Lys Lys Phe Ser Asp Ala Ser Ala Glu Asp Ile Gly
385                 390                 395                 400

Asn Arg Ala Glu Asp Ile Cys Arg Val Ile Ser Glu Thr Ala Pro Phe
                405                 410                 415

Ile Asn Asp Leu Arg Ala Val Asp Leu Asp Ser Leu Asn Asp Asp Gly
            420                 425                 430

Tyr Glu Ala Ala Val Ser Lys Ile Arg Glu Ser Leu Glu Pro Tyr Met
        435                 440                 445

Asp Leu Phe His Glu Leu Glu Ile Phe Ser Val Gly Asp Glu Phe Pro
    450                 455                 460

Lys Cys Ala Ala Phe Tyr Ser Glu Leu Glu Glu Val Ser Glu Gln Leu
465                 470                 475                 480

Ile Glu Ile Ile Pro Leu Phe Asn Lys Ala Arg Ser Phe Cys Thr Arg
                485                 490                 495

Lys Arg Tyr Ser Thr Asp Lys Ile Lys Val Asn Leu Lys Phe Pro Thr
            500                 505                 510

Leu Ala Asp Gly Trp Asp Leu Asn Lys Glu Arg Asp Asn Lys Ala Ala
        515                 520                 525

Ile Leu Arg Lys Asp Gly Lys Tyr Tyr Leu Ala Ile Leu Asp Met Lys
    530                 535                 540

Lys Asp Leu Ser Ser Ile Arg Thr Ser Asp Glu Asp Glu Ser Ser Phe
545                 550                 555                 560
```

Glu Lys Met Glu Tyr Lys Leu Leu Pro Ser Pro Val Lys Met Leu Pro
565 570 575

Lys Ile Phe Val Lys Ser Lys Ala Ala Lys Glu Lys Tyr Gly Leu Thr
580 585 590

Asp Arg Met Leu Glu Cys Tyr Asp Lys Gly Met His Lys Ser Gly Ser
595 600 605

Ala Phe Asp Leu Gly Phe Cys His Glu Leu Ile Asp Tyr Tyr Lys Arg
610 615 620

Cys Ile Ala Glu Tyr Pro Gly Trp Asp Val Phe Asp Phe Lys Phe Arg
625 630 635 640

Glu Thr Ser Asp Tyr Gly Ser Met Lys Glu Phe Asn Glu Asp Val Ala
645 650 655

Gly Ala Gly Tyr Tyr Met Ser Leu Arg Lys Ile Pro Cys Ser Glu Val
660 665 670

Tyr Arg Leu Leu Asp Glu Lys Ser Ile Tyr Leu Phe Gln Ile Tyr Asn
675 680 685

Lys Asp Tyr Ser Glu Asn Ala His Gly Asn Lys Asn Met His Thr Met
690 695 700

Tyr Trp Glu Gly Leu Phe Ser Pro Gln Asn Leu Glu Ser Pro Val Phe
705 710 715 720

Lys Leu Ser Gly Gly Ala Glu Leu Phe Phe Arg Lys Ser Ser Ile Pro
725 730 735

Asn Asp Ala Lys Thr Val His Pro Lys Gly Ser Val Leu Val Pro Arg
740 745 750

Asn Asp Val Asn Gly Arg Arg Ile Pro Asp Ser Ile Tyr Arg Glu Leu
755 760 765

Thr Arg Tyr Phe Asn Arg Gly Asp Cys Arg Ile Ser Asp Glu Ala Lys
770 775 780

Ser Tyr Leu Asp Lys Val Lys Thr Lys Lys Ala Asp His Asp Ile Val
785 790 795 800

Lys Asp Arg Arg Phe Thr Val Asp Lys Met Met Phe His Val Pro Ile
805 810 815

Ala Met Asn Phe Lys Ala Ile Ser Lys Pro Asn Leu Asn Lys Lys Val
820 825 830

Ile Asp Gly Ile Ile Asp Asp Gln Asp Leu Lys Ile Ile Gly Ile Asp
835 840 845

Arg Gly Glu Arg Asn Leu Ile Tyr Val Thr Met Val Asp Arg Lys Gly
850 855 860

Asn Ile Leu Tyr Gln Asp Ser Leu Asn Ile Leu Asn Gly Tyr Asp Tyr
865 870 875 880

Arg Lys Ala Leu Asp Val Arg Glu Tyr Asp Asn Lys Glu Ala Arg Arg
885 890 895

```
Asn Trp Thr Lys Val Glu Gly Ile Arg Lys Met Lys Glu Gly Tyr Leu
            900                 905                 910

Ser Leu Ala Val Ser Lys Leu Ala Asp Met Ile Ile Glu Asn Asn Ala
        915                 920                 925

Ile Ile Val Met Glu Asp Leu Asn His Gly Phe Lys Ala Gly Arg Ser
    930                 935                 940

Lys Ile Glu Lys Gln Val Tyr Gln Lys Phe Glu Ser Met Leu Ile Asn
945                 950                 955                 960

Lys Leu Gly Tyr Met Val Leu Lys Asp Lys Ser Ile Asp Gln Ser Gly
                965                 970                 975

Gly Ala Leu His Gly Tyr Gln Leu Ala Asn His Val Thr Thr Leu Ala
            980                 985                 990

Ser Val Gly Lys Gln Cys Gly Val Ile Phe Tyr Ile Pro Ala Ala Phe
        995                 1000                1005

Thr Ser  Lys Ile Asp Pro Thr  Thr Gly Phe Ala Asp  Leu Phe Ala
    1010                1015                1020

Leu Ser  Asn Val Lys Asn Val  Ala Ser Met Arg Glu  Phe Phe Ser
    1025                1030                1035

Lys Met  Lys Ser Val Ile Tyr  Asp Lys Ala Glu Gly  Lys Phe Ala
    1040                1045                1050

Phe Thr  Phe Asp Tyr Leu Asp  Tyr Asn Val Lys Ser  Glu Cys Gly
    1055                1060                1065

Arg Thr  Leu Trp Thr Val Tyr  Thr Val Gly Glu Arg  Phe Thr Tyr
    1070                1075                1080

Ser Arg  Val Asn Arg Glu Tyr  Val Arg Lys Val Pro  Thr Asp Ile
    1085                1090                1095

Ile Tyr  Asp Ala Leu Gln Lys  Ala Gly Ile Ser Val  Glu Gly Asp
    1100                1105                1110

Leu Arg  Asp Arg Ile Ala Glu  Ser Asp Gly Asp Thr  Leu Lys Ser
    1115                1120                1125

Ile Phe  Tyr Ala Phe Lys Tyr  Ala Leu Asp Met Arg  Val Glu Asn
    1130                1135                1140

Arg Glu  Glu Asp Tyr Ile Gln  Ser Pro Val Lys Asn  Ala Ser Gly
    1145                1150                1155

Glu Phe  Phe Cys Ser Lys Asn  Ala Gly Lys Ser Leu  Pro Gln Asp
    1160                1165                1170

Ser Asp  Ala Asn Gly Ala Tyr  Asn Ile Ala Leu Lys  Gly Ile Leu
    1175                1180                1185

Gln Leu  Arg Met Leu Ser Glu  Gln Tyr Asp Pro Asn  Ala Glu Ser
    1190                1195                1200

Ile Arg  Leu Pro Leu Ile Thr  Asn Lys Ala Trp Leu  Thr Phe Met
    1205                1210                1215
```

Gln Ser Gly Met Lys Thr Trp Lys Asn
1220 1225

<210> 111
<211> 1224
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 111
Met Gly Leu Tyr Asp Gly Phe Val Asn Arg Tyr Ser Val Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Arg Thr Arg Glu Tyr Ile Glu
20 25 30

Thr Asn Gly Ile Leu Ser Asp Asp Glu Glu Arg Ala Lys Asp Tyr Lys
35 40 45

Thr Ile Lys Arg Leu Ile Asp Glu Tyr His Lys Asp Tyr Ile Ser Arg
50 55 60

Cys Leu Lys Asn Val Asn Ile Ser Cys Leu Glu Glu Tyr Tyr His Leu
65 70 75 80

Tyr Asn Ser Ser Asn Arg Asp Lys Arg His Glu Glu Leu Asp Ala Leu
85 90 95

Ser Asp Gln Met Arg Gly Glu Ile Ala Ser Phe Leu Thr Gly Asn Asp
100 105 110

Glu Tyr Lys Glu Gln Lys Ser Arg Asp Ile Ile Ile Asn Glu Arg Ile
115 120 125

Ile Asn Phe Ala Ser Thr Asp Glu Glu Leu Ala Ala Val Lys Arg Phe
130 135 140

Arg Lys Phe Thr Ser Tyr Phe Thr Gly Phe Phe Thr Asn Arg Glu Asn
145 150 155 160

Met Tyr Ser Ala Glu Lys Lys Ser Thr Ala Ile Ala His Arg Ile Ile
165 170 175

Asp Val Asn Leu Pro Lys Tyr Val Asp Asn Ile Lys Ala Phe Asn Thr
180 185 190

Ala Ile Glu Ala Gly Val Phe Asp Ile Ala Glu Phe Glu Ser Asn Phe
195 200 205

Lys Ala Ile Thr Asp Glu His Glu Val Ser Asp Leu Leu Asp Ile Thr
210 215 220

Lys Tyr Ser Arg Phe Ile Arg Asn Glu Asp Ile Ile Ile Tyr Asn Thr
225 230 235 240

Leu Leu Gly Gly Ile Ser Met Lys Asp Glu Lys Ile Gln Gly Leu Asn
245 250 255

Glu Leu Ile Asn Leu His Asn Gln Lys His Pro Gly Lys Lys Val Pro
260 265 270

Leu Leu Lys Val Leu Tyr Lys Gln Ile Leu Gly Asp Ser Gln Thr His
275 280 285

```
Ser Phe Val Asp Asp Gln Phe Glu Asp Asp Gln Gln Val Ile Asn Ala
    290                 295                 300

Val Lys Ala Val Thr Asp Thr Phe Ser Glu Thr Leu Leu Gly Ser Leu
305                 310                 315                 320

Lys Ile Ile Ile Asn Asn Ile Gly His Tyr Asp Leu Asp Arg Ile Tyr
                325                 330                 335

Ile Lys Ala Gly Gln Asp Ile Thr Thr Leu Ser Lys Arg Ala Leu Asn
            340                 345                 350

Asp Trp His Ile Ile Thr Glu Cys Leu Glu Ser Glu Tyr Asp Asp Lys
        355                 360                 365

Phe Pro Lys Asn Lys Lys Ser Asp Thr Tyr Glu Glu Met Arg Asn Arg
    370                 375                 380

Tyr Val Lys Ser Phe Lys Ser Phe Ser Ile Gly Arg Leu Asn Ser Leu
385                 390                 395                 400

Val Thr Thr Tyr Thr Glu Gln Ala Cys Phe Leu Glu Asn Tyr Leu Gly
                405                 410                 415

Ser Phe Gly Gly Asp Thr Asp Lys Asn Cys Leu Thr Asp Phe Thr Asn
            420                 425                 430

Ser Leu Met Glu Val Glu His Leu Leu Asn Ser Glu Tyr Pro Val Thr
        435                 440                 445

Asn Arg Leu Ile Thr Asp Tyr Glu Ser Val Arg Ile Leu Lys Arg Leu
    450                 455                 460

Leu Asp Ser Glu Met Glu Val Ile His Phe Leu Lys Pro Leu Leu Gly
465                 470                 475                 480

Asn Gly Asn Glu Ser Asp Lys Asp Leu Val Phe Tyr Gly Glu Phe Glu
                485                 490                 495

Ala Glu Tyr Glu Lys Leu Leu Pro Val Ile Lys Val Tyr Asn Arg Val
            500                 505                 510

Arg Asn Tyr Leu Thr Arg Lys Pro Phe Ser Thr Glu Lys Ile Lys Leu
        515                 520                 525

Asn Phe Asn Ser Pro Thr Leu Leu Cys Gly Trp Ser Gln Ser Lys Glu
    530                 535                 540

Lys Glu Tyr Met Gly Val Ile Leu Arg Lys Asp Gly Gln Tyr Tyr Leu
545                 550                 555                 560

Gly Ile Met Thr Pro Ser Asn Lys Lys Ile Phe Ser Glu Ala Pro Lys
                565                 570                 575

Pro Asp Glu Asp Cys Tyr Glu Lys Met Val Leu Arg Tyr Ile Pro His
            580                 585                 590

Pro Tyr Gln Met Leu Pro Lys Val Phe Phe Ser Lys Ser Asn Ile Ala
        595                 600                 605

Phe Phe Asn Pro Ser Asp Glu Ile Leu Arg Ile Lys Lys Gln Glu Ser
    610                 615                 620
```

```
Phe Lys Lys Gly Lys Ser Phe Asn Arg Asp Asp Cys His Lys Phe Ile
625                 630                 635                 640

Asp Phe Tyr Lys Asp Ser Ile Asn Arg His Glu Glu Trp Arg Lys Phe
                645                 650                 655

Asn Phe Lys Phe Ser Asp Thr Asp Ser Tyr Glu Asp Ile Ser Arg Phe
            660                 665                 670

Tyr Lys Glu Val Glu Asn Gln Ala Phe Ser Met Ser Phe Thr Lys Ile
        675                 680                 685

Pro Thr Val Tyr Ile Asp Ser Leu Val Asp Glu Gly Lys Leu Tyr Leu
    690                 695                 700

Phe Lys Leu His Asn Lys Asp Phe Ser Glu His Ser Lys Gly Lys Pro
705                 710                 715                 720

Asn Leu His Thr Val Tyr Trp Asn Ala Leu Phe Ser Glu Tyr Asn Leu
                725                 730                 735

Gln Asn Thr Val Tyr Gln Leu Asn Gly Ser Ala Glu Ile Phe Phe Arg
            740                 745                 750

Lys Ala Ser Ile Pro Glu Asn Glu Arg Val Ile His Lys Lys Asn Val
        755                 760                 765

Pro Ile Thr Arg Lys Val Ala Glu Leu Asn Gly Lys Lys Glu Val Ser
    770                 775                 780

Val Phe Pro Tyr Asp Ile Ile Lys Asn Arg Arg Tyr Thr Val Asp Lys
785                 790                 795                 800

Phe Gln Phe His Val Pro Leu Lys Met Asn Phe Lys Ala Asp Glu Lys
                805                 810                 815

Lys Arg Ile Asn Asp Asp Val Ile Glu Ala Ile Arg Ser Asn Lys Gly
            820                 825                 830

Ile His Val Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu Leu Tyr Leu
        835                 840                 845

Ser Leu Ile Asn Glu Glu Gly Arg Ile Ile Glu Gln Arg Ser Leu Asn
    850                 855                 860

Ile Ile Asp Ser Gly Glu Gly His Thr Gln Asn Tyr Arg Asp Leu Leu
865                 870                 875                 880

Asp Ser Arg Glu Lys Asp Arg Glu Lys Ala Arg Glu Asn Trp Gln Glu
                885                 890                 895

Ile Gln Glu Ile Lys Asp Leu Lys Thr Gly Tyr Leu Ser Gln Ala Ile
            900                 905                 910

His Thr Ile Thr Lys Trp Met Lys Glu Tyr Asn Ala Ile Ile Val Leu
        915                 920                 925

Glu Asp Leu Asn Asp Arg Phe Thr Asn Gly Arg Lys Lys Val Glu Lys
    930                 935                 940

Gln Val Tyr Gln Lys Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr
945                 950                 955                 960
```

```
Tyr Val Asp Lys Asp Glu Glu Phe Asp Arg Met Gly Gly Thr His Arg
                965                 970                 975

Ala Leu Gln Leu Thr Glu Lys Phe Glu Ser Phe Gln Lys Leu Gly Arg
            980                 985                 990

Gln Thr Gly Phe Ile Phe Tyr Val Pro Ala Trp Asn Thr Ser Lys Leu
        995                 1000                1005

Asp Pro  Thr Thr Gly Phe Val  Asp Leu Leu Tyr Pro  Lys Tyr Lys
    1010                1015                1020

Ser Val  Asp Ala Thr Lys Asp  Phe Ile Lys Lys Phe  Asp Phe Ile
    1025                1030                1035

Arg Phe  Asn Ser Glu Lys Asn  Tyr Phe Glu Phe Gly  Leu His Tyr
    1040                1045                1050

Ser Asn  Phe Thr Glu Arg Ala  Ile Gly Cys Arg Asp  Glu Trp Ile
    1055                1060                1065

Leu Cys  Ser Tyr Gly Asn Arg  Ile Val Asn Phe Arg  Asn Ala Ala
    1070                1075                1080

Lys Asn  Asn Ser Trp Asp Tyr  Lys Glu Ile Asp Ile  Thr Lys Gln
    1085                1090                1095

Leu Leu  Asp Leu Phe Glu Lys  Asn Gly Ile Asp Val  Lys Gln Glu
    1100                1105                1110

Asn Leu  Ile Asp Ser Ile Cys  Glu Met Lys Asp Lys  Pro Phe Phe
    1115                1120                1125

Lys Ser  Leu Ile Ala Asn Ile  Lys Leu Ile Leu Gln  Ile Arg Asn
    1130                1135                1140

Ser Ala  Ser Gly Thr Asp Ile  Asp Tyr Met Ile Ser  Pro Ala Met
    1145                1150                1155

Asn Asp  Arg Gly Glu Phe Phe  Asp Thr Arg Lys Gly  Leu Gln Gln
    1160                1165                1170

Leu Pro  Leu Asp Ala Asp Ala  Asn Gly Ala Tyr Asn  Ile Ala Lys
    1175                1180                1185

Lys Gly  Leu Trp Ile Val Asp  Gln Ile Arg Asn Thr  Thr Gly Asn
    1190                1195                1200

Asn Val  Lys Met Ala Met Ser  Asn Arg Glu Trp Met  His Phe Ala
    1205                1210                1215

Gln Glu  Ser Arg Leu Ala
    1220
```

<210> 112
<211> 1214
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選分裂WS6細菌序列

<400> 112

```
Met Lys Asn Val Phe Gly Gly Phe Thr Asn Leu Tyr Ser Leu Thr Lys
1               5                  10                  15

Thr Leu Arg Phe Glu Leu Lys Pro Thr Ser Lys Thr Gln Lys Leu Met
            20                  25                  30

Lys Arg Asn Asn Val Ile Gln Thr Asp Glu Glu Ile Asp Lys Leu Tyr
        35                  40                  45

His Asp Glu Met Lys Pro Ile Leu Asp Glu Ile His Arg Arg Phe Ile
    50                  55                  60

Asn Asp Ala Leu Ala Gln Lys Ile Phe Ile Ser Ala Ser Leu Asp Asn
65                  70                  75                  80

Phe Leu Lys Val Val Lys Asn Tyr Lys Val Glu Ser Ala Lys Lys Asn
                85                  90                  95

Ile Lys Gln Asn Gln Val Lys Leu Leu Gln Lys Glu Ile Thr Ile Lys
            100                 105                 110

Thr Leu Gly Leu Arg Arg Glu Val Val Ser Gly Phe Ile Thr Val Ser
        115                 120                 125

Lys Lys Trp Lys Asp Lys Tyr Val Gly Leu Gly Ile Lys Leu Lys Gly
    130                 135                 140

Asp Gly Tyr Lys Val Leu Thr Glu Gln Ala Val Leu Asp Ile Leu Lys
145                 150                 155                 160

Ile Glu Phe Pro Asn Lys Ala Lys Tyr Ile Asp Lys Phe Arg Gly Phe
                165                 170                 175

Trp Thr Tyr Phe Ser Gly Phe Asn Glu Asn Arg Lys Asn Tyr Tyr Ser
            180                 185                 190

Glu Glu Asp Lys Ala Thr Ser Ile Ala Asn Arg Ile Val Asn Glu Asn
        195                 200                 205

Leu Ser Arg Tyr Ile Asp Asn Ile Ile Ala Phe Glu Glu Ile Leu Gln
    210                 215                 220

Lys Ile Pro Asn Leu Lys Lys Phe Lys Gln Asp Leu Asp Ile Thr Ser
225                 230                 235                 240

Tyr Asn Tyr Tyr Leu Asn Gln Ala Gly Ile Asp Lys Tyr Asn Lys Ile
                245                 250                 255

Ile Gly Gly Tyr Ile Val Asp Lys Asp Lys Lys Ile Gln Gly Ile Asn
            260                 265                 270

Glu Lys Val Asn Leu Tyr Thr Gln Gln Thr Lys Lys Lys Leu Pro Lys
        275                 280                 285

Leu Lys Phe Leu Phe Lys Gln Ile Gly Ser Glu Arg Lys Gly Phe Gly
    290                 295                 300

Ile Phe Glu Ile Lys Glu Gly Lys Glu Trp Glu Gln Leu Gly Asp Leu
305                 310                 315                 320

Phe Lys Leu Gln Arg Thr Lys Ile Asn Ser Asn Gly Arg Glu Lys Gly
                325                 330                 335
```

Leu Phe Asp Ser Leu Arg Thr Met Tyr Arg Glu Phe Phe Asp Glu Ile
340 345 350

Lys Arg Asp Ser Asn Ser Gln Ala Arg Tyr Ser Leu Asp Lys Ile Tyr
355 360 365

Phe Asn Lys Ala Ser Val Asn Thr Ile Ser Asn Ser Trp Phe Thr Asn
370 375 380

Trp Asn Lys Phe Ala Glu Leu Leu Asn Ile Lys Glu Asp Lys Lys Asn
385 390 395 400

Gly Glu Lys Lys Ile Pro Glu Gln Ile Ser Ile Glu Asp Ile Lys Asp
405 410 415

Ser Leu Ser Ile Ile Pro Lys Glu Asn Leu Glu Glu Leu Phe Lys Leu
420 425 430

Thr Asn Arg Glu Lys His Asp Arg Thr Arg Phe Phe Gly Ser Asn Ala
435 440 445

Trp Val Thr Phe Leu Asn Ile Trp Gln Asn Glu Ile Glu Glu Ser Phe
450 455 460

Asn Lys Leu Glu Glu Lys Glu Lys Asp Phe Lys Lys Asn Ala Ala Ile
465 470 475 480

Lys Phe Gln Lys Asn Asn Leu Val Gln Lys Asn Tyr Ile Lys Glu Val
485 490 495

Cys Asp Arg Met Leu Ala Ile Glu Arg Met Ala Lys Tyr His Leu Pro
500 505 510

Lys Asp Ser Asn Leu Ser Arg Glu Glu Asp Phe Tyr Trp Ile Ile Asp
515 520 525

Asn Leu Ser Glu Gln Arg Glu Ile Tyr Lys Tyr Tyr Asn Ala Phe Arg
530 535 540

Asn Tyr Ile Ser Lys Lys Pro Tyr Asn Lys Ser Lys Met Lys Leu Asn
545 550 555 560

Phe Glu Asn Gly Asn Leu Leu Gly Gly Trp Ser Asp Gly Gln Glu Arg
565 570 575

Asn Lys Ala Gly Val Ile Leu Arg Asn Gly Asn Lys Tyr Tyr Leu Gly
580 585 590

Val Leu Ile Asn Arg Gly Ile Phe Arg Thr Asp Lys Ile Asn Asn Glu
595 600 605

Ile Tyr Arg Thr Gly Ser Ser Lys Trp Glu Arg Leu Ile Leu Ser Asn
610 615 620

Leu Lys Phe Gln Thr Leu Ala Gly Lys Gly Phe Leu Gly Lys His Gly
625 630 635 640

Val Ser Tyr Gly Asn Met Asn Pro Glu Lys Ser Val Pro Ser Leu Gln
645 650 655

Lys Phe Ile Arg Glu Asn Tyr Leu Lys Lys Tyr Pro Gln Leu Thr Glu
660 665 670

```
Val Ser Asn Thr Lys Phe Leu Ser Lys Lys Asp Phe Asp Ala Ala Ile
        675                 680                 685

Lys Glu Ala Leu Lys Glu Cys Phe Thr Met Asn Phe Ile Asn Ile Ala
    690                 695                 700

Glu Asn Lys Leu Leu Glu Ala Glu Asp Lys Gly Asp Leu Tyr Leu Phe
705                 710                 715                 720

Glu Ile Thr Asn Lys Asp Phe Ser Gly Lys Lys Ser Gly Lys Asp Asn
                725                 730                 735

Ile His Thr Ile Tyr Trp Lys Tyr Leu Phe Ser Glu Ser Asn Cys Lys
            740                 745                 750

Ser Pro Ile Ile Gly Leu Asn Gly Gly Ala Glu Ile Phe Phe Arg Glu
        755                 760                 765

Gly Gln Lys Asp Lys Leu His Thr Lys Leu Asp Lys Lys Gly Lys Lys
    770                 775                 780

Val Phe Asp Ala Lys Arg Tyr Ser Glu Asp Lys Leu Phe Phe His Val
785                 790                 795                 800

Ser Ile Thr Ile Asn Tyr Gly Lys Pro Lys Asn Ile Lys Phe Arg Asp
                805                 810                 815

Ile Ile Asn Gln Leu Ile Thr Ser Met Asn Val Asn Ile Ile Gly Ile
            820                 825                 830

Asp Arg Gly Glu Lys His Leu Leu Tyr Tyr Ser Val Ile Asp Ser Asn
        835                 840                 845

Gly Ile Ile Leu Lys Gln Gly Ser Leu Asn Lys Ile Arg Val Gly Asp
    850                 855                 860

Lys Glu Val Asp Phe Asn Lys Lys Leu Thr Glu Arg Ala Asn Glu Met
865                 870                 875                 880

Lys Lys Ala Arg Gln Ser Trp Glu Gln Ile Gly Asn Ile Lys Asn Phe
                885                 890                 895

Lys Glu Gly Tyr Leu Ser Gln Ala Ile His Glu Ile Tyr Gln Leu Met
            900                 905                 910

Ile Lys Tyr Asn Ala Ile Ile Val Leu Glu Asp Leu Asn Thr Glu Phe
        915                 920                 925

Lys Ala Lys Arg Leu Ser Lys Val Glu Lys Ser Val Tyr Lys Lys Phe
    930                 935                 940

Glu Leu Lys Leu Ala Arg Lys Leu Asn His Leu Ile Leu Lys Asp Arg
945                 950                 955                 960

Asn Thr Asn Glu Ile Gly Gly Val Leu Lys Ala Tyr Gln Leu Thr Pro
                965                 970                 975

Thr Ile Gly Gly Gly Asp Val Ser Lys Phe Glu Lys Ala Lys Gln Trp
            980                 985                 990

Gly Met Met Phe Tyr Val Arg Ala Asn Tyr Thr Ser Thr Thr Asp Pro
        995                 1000                1005
```

Val Thr Gly Trp Arg Lys His Leu Tyr Ile Ser Asn Phe Ser Asn
1010 1015 1020

Asn Ser Val Ile Lys Ser Phe Phe Asp Pro Thr Asn Arg Asp Thr
1025 1030 1035

Gly Ile Glu Ile Phe Tyr Ser Gly Lys Tyr Arg Ser Trp Gly Phe
1040 1045 1050

Arg Tyr Val Gln Lys Glu Thr Gly Lys Lys Trp Glu Leu Phe Ala
1055 1060 1065

Thr Lys Glu Leu Glu Arg Phe Lys Tyr Asn Gln Thr Thr Lys Leu
1070 1075 1080

Cys Glu Lys Ile Asn Leu Tyr Asp Lys Phe Glu Glu Leu Phe Lys
1085 1090 1095

Gly Ile Asp Lys Ser Ala Asp Ile Tyr Ser Gln Leu Cys Asn Val
1100 1105 1110

Leu Asp Phe Arg Trp Lys Ser Leu Val Tyr Leu Trp Asn Leu Leu
1115 1120 1125

Asn Gln Ile Arg Asn Val Asp Lys Asn Ala Glu Gly Asn Lys Asn
1130 1135 1140

Asp Phe Ile Gln Ser Pro Val Tyr Pro Phe Phe Asp Ser Arg Lys
1145 1150 1155

Thr Asp Gly Lys Thr Glu Pro Ile Asn Gly Asp Ala Asn Gly Ala
1160 1165 1170

Leu Asn Ile Ala Arg Lys Gly Leu Met Leu Val Glu Arg Ile Lys
1175 1180 1185

Asn Asn Pro Glu Lys Tyr Glu Gln Leu Ile Arg Asp Thr Glu Trp
1190 1195 1200

Asp Ala Trp Ile Gln Asn Phe Asn Lys Val Asn
1205 1210

<210> 113
<211> 1200
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 113
Met Tyr Tyr Glu Ser Leu Thr Lys Gln Tyr Pro Val Ser Lys Thr Ile
1 5 10 15

Arg Asn Glu Leu Ile Pro Ile Gly Lys Thr Leu Asp Asn Ile Arg Gln
20 25 30

Asn Asn Ile Leu Glu Ser Asp Val Lys Arg Lys Gln Asn Tyr Glu His
35 40 45

Val Lys Gly Ile Leu Asp Glu Tyr His Lys Gln Leu Ile Asn Glu Ala
50 55 60

Leu Asp Asn Cys Thr Leu Pro Ser Leu Lys Ile Ala Ala Glu Ile Tyr
65 70 75 80

```
Leu Lys Asn Gln Lys Glu Val Ser Asp Arg Glu Asp Phe Asn Lys Thr
                85                  90                  95

Gln Asp Leu Leu Arg Lys Glu Val Val Glu Lys Leu Lys Ala His Glu
            100                 105                 110

Asn Phe Thr Lys Ile Gly Lys Lys Asp Ile Leu Asp Leu Leu Glu Lys
        115                 120                 125

Leu Pro Ser Ile Ser Glu Asp Asp Tyr Asn Ala Leu Glu Ser Phe Arg
    130                 135                 140

Asn Phe Tyr Thr Tyr Phe Thr Ser Tyr Asn Lys Val Arg Glu Asn Leu
145                 150                 155                 160

Tyr Ser Asp Lys Glu Lys Ser Ser Thr Val Ala Tyr Arg Leu Ile Asn
                165                 170                 175

Glu Asn Phe Pro Lys Phe Leu Asp Asn Val Lys Ser Tyr Arg Phe Val
            180                 185                 190

Lys Thr Ala Gly Ile Leu Ala Asp Gly Leu Gly Glu Glu Glu Gln Asp
        195                 200                 205

Ser Leu Phe Ile Val Glu Thr Phe Asn Lys Thr Leu Thr Gln Asp Gly
    210                 215                 220

Ile Asp Thr Tyr Asn Ser Gln Val Gly Lys Ile Asn Ser Ser Ile Asn
225                 230                 235                 240

Leu Tyr Asn Gln Lys Asn Gln Lys Ala Asn Gly Phe Arg Lys Ile Pro
                245                 250                 255

Lys Met Lys Met Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Glu Ser
            260                 265                 270

Phe Ile Asp Glu Phe Gln Ser Asp Glu Val Leu Ile Asp Asn Val Glu
        275                 280                 285

Ser Tyr Gly Ser Val Leu Ile Glu Ser Leu Lys Ser Ser Lys Val Ser
    290                 295                 300

Ala Phe Phe Asp Ala Leu Arg Glu Ser Lys Gly Lys Asn Val Tyr Val
305                 310                 315                 320

Lys Asn Asp Leu Ala Lys Thr Ala Met Ser Asn Ile Val Phe Glu Asn
                325                 330                 335

Trp Arg Thr Phe Asp Asp Leu Leu Asn Gln Glu Tyr Asp Leu Ala Asn
            340                 345                 350

Glu Asn Lys Lys Lys Asp Asp Lys Tyr Phe Glu Lys Arg Gln Lys Glu
        355                 360                 365

Leu Lys Lys Asn Lys Ser Tyr Ser Leu Glu His Leu Cys Asn Leu Ser
    370                 375                 380

Glu Asp Ser Cys Asn Leu Ile Glu Asn Tyr Ile His Gln Ile Ser Asp
385                 390                 395                 400

Asp Ile Glu Asn Ile Ile Ile Asn Asn Glu Thr Phe Leu Arg Ile Val
                405                 410                 415
```

```
Ile Asn Glu His Asp Arg Ser Arg Lys Leu Ala Lys Asn Arg Lys Ala
            420                 425                 430

Val Lys Ala Ile Lys Asp Phe Leu Asp Ser Ile Lys Val Leu Glu Arg
        435                 440                 445

Glu Leu Lys Leu Ile Asn Ser Ser Gly Gln Glu Leu Glu Lys Asp Leu
    450                 455                 460

Ile Val Tyr Ser Ala His Glu Glu Leu Leu Val Glu Leu Lys Gln Val
465                 470                 475                 480

Asp Ser Leu Tyr Asn Met Thr Arg Asn Tyr Leu Thr Lys Lys Pro Phe
                485                 490                 495

Ser Thr Glu Lys Val Lys Leu Asn Phe Asn Arg Ser Thr Leu Leu Asn
            500                 505                 510

Gly Trp Asp Arg Asn Lys Glu Thr Asp Asn Leu Gly Val Leu Leu Leu
        515                 520                 525

Lys Asp Gly Lys Tyr Tyr Leu Gly Ile Met Asn Thr Ser Ala Asn Lys
    530                 535                 540

Ala Phe Val Asn Pro Pro Val Ala Lys Thr Glu Lys Val Phe Lys Lys
545                 550                 555                 560

Val Asp Tyr Lys Leu Leu Pro Val Pro Asn Gln Met Leu Pro Lys Val
                565                 570                 575

Phe Phe Ala Lys Ser Asn Ile Asp Phe Tyr Asn Pro Ser Ser Glu Ile
            580                 585                 590

Tyr Ser Asn Tyr Lys Lys Gly Thr His Lys Lys Gly Asn Met Phe Ser
        595                 600                 605

Leu Glu Asp Cys His Asn Leu Ile Asp Phe Phe Lys Glu Ser Ile Ser
    610                 615                 620

Lys His Glu Asp Trp Ser Lys Phe Gly Phe Lys Phe Ser Asp Thr Ala
625                 630                 635                 640

Ser Tyr Asn Asp Ile Ser Glu Phe Tyr Arg Glu Val Glu Lys Gln Gly
                645                 650                 655

Tyr Lys Leu Thr Tyr Thr Asp Ile Asp Glu Thr Tyr Ile Asn Asp Leu
            660                 665                 670

Ile Glu Arg Asn Glu Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe
        675                 680                 685

Ser Met Tyr Ser Lys Gly Lys Leu Asn Leu His Thr Leu Tyr Phe Met
    690                 695                 700

Met Leu Phe Asp Gln Arg Asn Ile Asp Asp Val Val Tyr Lys Leu Asn
705                 710                 715                 720

Gly Glu Ala Glu Val Phe Tyr Arg Pro Ala Ser Ile Ser Glu Asp Glu
                725                 730                 735

Leu Ile Ile His Lys Ala Gly Glu Glu Ile Lys Asn Lys Asn Pro Asn
            740                 745                 750
```

```
Arg Ala Arg Thr Lys Glu Thr Ser Thr Phe Ser Tyr Asp Ile Val Lys
        755                 760                 765

Asp Lys Arg Tyr Ser Lys Asp Lys Phe Thr Leu His Ile Pro Ile Thr
    770                 775                 780

Met Asn Phe Gly Val Asp Glu Val Lys Arg Phe Asn Asp Ala Val Asn
785                 790                 795                 800

Ser Ala Ile Arg Ile Asp Glu Asn Val Asn Val Ile Gly Ile Asp Arg
                805                 810                 815

Gly Glu Arg Asn Leu Leu Tyr Val Val Val Ile Asp Ser Lys Gly Asn
            820                 825                 830

Ile Leu Glu Gln Ile Ser Leu Asn Ser Ile Ile Asn Lys Glu Tyr Asp
        835                 840                 845

Ile Glu Thr Asp Tyr His Ala Leu Leu Asp Glu Arg Glu Gly Gly Arg
    850                 855                 860

Asp Lys Ala Arg Lys Asp Trp Asn Thr Val Glu Asn Ile Arg Asp Leu
865                 870                 875                 880

Lys Ala Gly Tyr Leu Ser Gln Val Val Asn Val Val Ala Lys Leu Val
                885                 890                 895

Leu Lys Tyr Asn Ala Ile Ile Cys Leu Glu Asp Leu Asn Phe Gly Phe
            900                 905                 910

Lys Arg Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu
        915                 920                 925

Lys Met Leu Ile Asp Lys Leu Asn Tyr Leu Val Ile Asp Lys Ser Arg
    930                 935                 940

Glu Gln Thr Ser Pro Lys Glu Leu Gly Gly Ala Leu Asn Ala Leu Gln
945                 950                 955                 960

Leu Thr Ser Lys Phe Lys Ser Phe Lys Glu Leu Gly Lys Gln Ser Gly
                965                 970                 975

Val Ile Tyr Tyr Val Pro Ala Tyr Leu Thr Ser Lys Ile Asp Pro Thr
            980                 985                 990

Thr Gly Phe Ala Asn Leu Phe Tyr Met Lys Cys Glu Asn Val Glu Lys
        995                 1000                1005

Ser Lys  Arg Phe Phe Asp Gly Phe Asp Phe Ile  Arg Phe Asn Ala
    1010                1015                1020

Leu Glu  Asn Val Phe Glu Phe Gly Phe Asp Tyr  Arg Ser Phe Thr
    1025                1030                1035

Gln Arg  Ala Cys Gly Ile Asn Ser Lys Trp Thr  Val Cys Thr Asn
    1040                1045                1050

Gly Glu  Arg Ile Ile Lys Tyr Arg Asn Pro Asp  Lys Asn Asn Met
    1055                1060                1065

Phe Asp  Glu Lys Val Val Val Val Thr Asp Glu  Met Lys Asn Leu
    1070                1075                1080
```

```
Phe Glu Gln Tyr Lys Ile  Pro Tyr Glu Asp Gly  Arg Asn Val Lys
    1085                 1090                 1095

Asp Met Ile Ile Ser Asn  Glu Glu Ala Glu Phe  Tyr Arg Arg Leu
    1100                 1105                 1110

Tyr Arg Leu Leu Gln Gln  Thr Leu Gln Met Arg  Asn Ser Thr Ser
    1115                 1120                 1125

Asp Gly Thr Arg Asp Tyr  Ile Ile Ser Pro Val  Lys Asn Lys Arg
    1130                 1135                 1140

Glu Ala Tyr Phe Asn Ser  Glu Leu Ser Asp Gly  Ser Val Pro Lys
    1145                 1150                 1155

Asp Ala Asp Ala Asn Gly  Ala Tyr Asn Ile Ala  Arg Lys Gly Leu
    1160                 1165                 1170

Trp Val Leu Glu Gln Ile  Arg Gln Lys Ser Glu  Gly Glu Lys Ile
    1175                 1180                 1185

Asn Leu Ala Met Thr Asn  Ala Glu Trp Leu Glu  Tyr
    1190                 1195                 1200
```

<210> 114
<211> 1206
<212> PRT
<213> 丁酸弧菌屬某種（Butyrivibrio sp.）

<400> 114

```
Met Tyr Tyr Gln Asn Leu Thr Lys Lys Tyr Pro Val Ser Lys Thr Ile
1               5                  10                  15

Arg Asn Glu Leu Ile Pro Ile Gly Lys Thr Leu Glu Asn Ile Arg Lys
            20                  25                  30

Asn Asn Ile Leu Glu Ser Asp Val Lys Arg Lys Gln Asp Tyr Glu His
        35                  40                  45

Val Lys Gly Ile Met Asp Glu Tyr His Lys Gln Leu Ile Asn Glu Ala
    50                  55                  60

Leu Asp Asn Tyr Met Leu Pro Ser Leu Asn Gln Ala Ala Glu Ile Tyr
65                  70                  75                  80

Leu Lys Lys His Val Asp Val Glu Asp Arg Glu Glu Phe Lys Lys Thr
                85                  90                  95

Gln Asp Leu Leu Arg Arg Glu Val Thr Gly Arg Leu Lys Glu His Glu
            100                 105                 110

Asn Tyr Thr Lys Ile Gly Lys Lys Asp Ile Leu Asp Leu Leu Glu Lys
        115                 120                 125

Leu Pro Ser Ile Ser Glu Glu Asp Tyr Asn Ala Leu Glu Ser Phe Arg
    130                 135                 140

Asn Phe Tyr Thr Tyr Phe Thr Ser Tyr Asn Lys Val Arg Glu Asn Leu
145                 150                 155                 160

Tyr Ser Asp Glu Glu Lys Ser Ser Thr Val Ala Tyr Arg Leu Ile Asn
                165                 170                 175
```

```
Glu Asn Leu Pro Lys Phe Leu Asp Asn Ile Lys Ser Tyr Ala Phe Val
            180                 185                 190

Lys Ala Ala Gly Val Leu Ala Asp Cys Ile Glu Glu Glu Glu Gln Asp
        195                 200                 205

Ala Leu Phe Met Val Glu Thr Phe Asn Met Thr Leu Thr Gln Glu Gly
    210                 215                 220

Ile Asp Met Tyr Asn Tyr Gln Ile Gly Lys Val Asn Ser Ala Ile Asn
225                 230                 235                 240

Leu Tyr Asn Gln Lys Asn His Lys Val Glu Glu Phe Lys Lys Ile Pro
                245                 250                 255

Lys Met Lys Val Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Glu Val
            260                 265                 270

Phe Ile Gly Glu Phe Lys Asp Asp Glu Thr Leu Leu Ser Ser Ile Gly
        275                 280                 285

Ala Tyr Gly Asn Val Leu Met Thr Tyr Leu Lys Ser Glu Lys Ile Asn
    290                 295                 300

Ile Phe Phe Asp Ala Leu Arg Glu Ser Glu Gly Lys Asn Val Tyr Val
305                 310                 315                 320

Lys Asn Asp Leu Ser Lys Thr Thr Met Ser Asn Ile Val Phe Gly Ser
                325                 330                 335

Trp Ser Ala Phe Asp Glu Leu Leu Asn Gln Glu Tyr Asp Leu Ala Asn
            340                 345                 350

Glu Asn Lys Lys Lys Asp Asp Lys Tyr Phe Glu Lys Arg Gln Lys Glu
        355                 360                 365

Leu Lys Lys Asn Lys Ser Tyr Thr Leu Glu Gln Met Ser Asn Leu Ser
    370                 375                 380

Lys Glu Asp Ile Ser Pro Ile Glu Asn Tyr Ile Glu Arg Ile Ser Glu
385                 390                 395                 400

Asp Ile Glu Lys Ile Cys Ile Tyr Asn Gly Glu Phe Glu Lys Ile Val
                405                 410                 415

Val Asn Glu His Asp Ser Ser Arg Lys Leu Ser Lys Asn Ile Lys Ala
            420                 425                 430

Val Lys Val Ile Lys Asp Tyr Leu Asp Ser Ile Lys Glu Leu Glu His
        435                 440                 445

Asp Ile Lys Leu Ile Asn Gly Ser Gly Gln Glu Leu Glu Lys Asn Leu
    450                 455                 460

Val Val Tyr Val Gly Gln Glu Glu Ala Leu Glu Gln Leu Arg Pro Val
465                 470                 475                 480

Asp Ser Leu Tyr Asn Leu Thr Arg Asn Tyr Leu Thr Lys Lys Pro Phe
                485                 490                 495

Ser Thr Glu Lys Val Lys Leu Asn Phe Asn Lys Ser Thr Leu Leu Asn
            500                 505                 510
```

Gly Trp Asp Lys Asn Lys Glu Thr Asp Asn Leu Gly Ile Leu Phe Phe
515 520 525

Lys Asp Gly Lys Tyr Tyr Leu Gly Ile Met Asn Thr Thr Ala Asn Lys
530 535 540

Ala Phe Val Asn Pro Pro Ala Ala Lys Thr Glu Asn Val Phe Lys Lys
545 550 555 560

Val Asp Tyr Lys Leu Leu Pro Gly Ser Asn Lys Met Leu Pro Lys Val
565 570 575

Phe Phe Ala Lys Ser Asn Ile Gly Tyr Tyr Asn Pro Ser Thr Glu Leu
580 585 590

Tyr Ser Asn Tyr Lys Lys Gly Thr His Lys Lys Gly Pro Ser Phe Ser
595 600 605

Ile Asp Asp Cys His Asn Leu Ile Asp Phe Phe Lys Glu Ser Ile Lys
610 615 620

Lys His Glu Asp Trp Ser Lys Phe Gly Phe Glu Phe Ser Asp Thr Ala
625 630 635 640

Asp Tyr Arg Asp Ile Ser Glu Phe Tyr Arg Glu Val Glu Lys Gln Gly
645 650 655

Tyr Lys Leu Thr Phe Thr Asp Ile Asp Glu Ser Tyr Ile Asn Asp Leu
660 665 670

Ile Glu Lys Asn Glu Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe
675 680 685

Ser Glu Tyr Ser Lys Gly Lys Leu Asn Leu His Thr Leu Tyr Phe Met
690 695 700

Met Leu Phe Asp Gln Arg Asn Leu Asp Asn Val Val Tyr Lys Leu Asn
705 710 715 720

Gly Glu Ala Glu Val Phe Tyr Arg Pro Ala Ser Ile Ala Glu Asn Glu
725 730 735

Leu Val Ile His Lys Ala Gly Glu Gly Ile Lys Asn Lys Asn Pro Asn
740 745 750

Arg Ala Lys Val Lys Glu Thr Ser Thr Phe Ser Tyr Asp Ile Val Lys
755 760 765

Asp Lys Arg Tyr Ser Lys Tyr Lys Phe Thr Leu His Ile Pro Ile Thr
770 775 780

Met Asn Phe Gly Val Asp Glu Val Arg Arg Phe Asn Asp Val Ile Asn
785 790 795 800

Asn Ala Leu Arg Thr Asp Asp Asn Val Asn Val Ile Gly Ile Asp Arg
805 810 815

Gly Glu Arg Asn Leu Leu Tyr Val Val Val Ile Asn Ser Glu Gly Lys
820 825 830

Ile Leu Glu Gln Ile Ser Leu Asn Ser Ile Ile Asn Lys Glu Tyr Asp
835 840 845

```
Ile Glu Thr Asn Tyr His Ala Leu Leu Asp Glu Arg Glu Asp Asp Arg
    850                 855                 860

Asn Lys Ala Arg Lys Asp Trp Asn Thr Ile Glu Asn Ile Lys Glu Leu
865                 870                 875                 880

Lys Thr Gly Tyr Leu Ser Gln Val Val Asn Val Val Ala Lys Leu Val
                885                 890                 895

Leu Lys Tyr Asn Ala Ile Ile Cys Leu Glu Asp Leu Asn Phe Gly Phe
            900                 905                 910

Lys Arg Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu
        915                 920                 925

Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu Val Ile Asp Lys Ser Arg
    930                 935                 940

Glu Gln Val Ser Pro Glu Lys Met Gly Gly Ala Leu Asn Ala Leu Gln
945                 950                 955                 960

Leu Thr Ser Lys Phe Lys Ser Phe Ala Glu Leu Gly Lys Gln Ser Gly
                965                 970                 975

Ile Ile Tyr Tyr Val Pro Ala Tyr Leu Thr Ser Lys Ile Asp Pro Thr
            980                 985                 990

Thr Gly Phe Val Asn Leu Phe Tyr Ile Lys Tyr Glu Asn Ile Glu Lys
        995                 1000                1005

Ala Lys  Gln Phe Phe Asp Gly  Phe Asp Phe Ile Arg  Phe Asn Lys
    1010                1015                1020

Lys Asp  Asp Met Phe Glu Phe  Ser Phe Asp Tyr Lys  Ser Phe Thr
    1025                1030                1035

Gln Lys  Ala Cys Gly Ile Arg  Ser Lys Trp Ile Val  Tyr Thr Asn
    1040                1045                1050

Gly Glu  Arg Ile Ile Lys Tyr  Pro Asn Pro Glu Lys  Asn Asn Leu
    1055                1060                1065

Phe Asp  Glu Lys Val Ile Asn  Val Thr Asp Glu Ile  Lys Gly Leu
    1070                1075                1080

Phe Lys  Gln Tyr Arg Ile Pro  Tyr Glu Asn Gly Glu  Asp Ile Lys
    1085                1090                1095

Glu Ile  Ile Ile Ser Lys Ala  Glu Ala Asp Phe Tyr  Lys Arg Leu
    1100                1105                1110

Phe Arg  Leu Leu His Gln Thr  Leu Gln Met Arg Asn  Ser Thr Ser
    1115                1120                1125

Asp Gly  Thr Arg Asp Tyr Ile  Ile Ser Pro Val Lys  Asn Asp Arg
    1130                1135                1140

Gly Glu  Phe Phe Cys Ser Glu  Phe Ser Glu Gly Thr  Met Pro Lys
    1145                1150                1155

Asp Ala  Asp Ala Asn Gly Ala  Tyr Asn Ile Ala Arg  Lys Gly Leu
    1160                1165                1170
```

```
Trp Val Leu Glu Gln Ile Arg Gln Lys Asp Glu Gly Glu Lys Val
    1175                1180                1185

Asn Leu Ser Met Thr Asn Ala Glu Trp Leu Lys Tyr Ala Gln Leu
    1190                1195                1200

His Leu Leu
    1205
```

<210> 115
<211> 1206
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 115

```
Met Glu Asn Tyr Tyr Asp Ser Leu Thr Arg Gln Tyr Pro Val Thr Lys
1               5                   10                  15

Thr Ile Arg Gln Glu Leu Lys Pro Val Gly Lys Thr Leu Glu Asn Ile
            20                  25                  30

Lys Asn Ala Glu Ile Ile Glu Ala Asp Lys Gln Lys Lys Glu Ala Tyr
        35                  40                  45

Val Lys Val Lys Glu Leu Met Asp Glu Phe His Lys Ser Ile Ile Glu
    50                  55                  60

Lys Ser Leu Val Gly Ile Lys Leu Asp Gly Leu Ser Glu Phe Glu Lys
65                  70                  75                  80

Leu Tyr Lys Ile Lys Thr Lys Thr Asp Glu Asp Lys Asn Arg Ile Ser
                85                  90                  95

Glu Leu Phe Tyr Tyr Met Arg Lys Gln Ile Ala Asp Ala Leu Lys Asn
            100                 105                 110

Ser Arg Asp Tyr Gly Tyr Val Asp Asn Lys Asp Leu Ile Glu Lys Ile
        115                 120                 125

Leu Pro Glu Arg Val Lys Asp Glu Asn Ser Leu Asn Ala Leu Ser Cys
    130                 135                 140

Phe Lys Gly Phe Thr Thr Tyr Phe Thr Asp Tyr Tyr Lys Asn Arg Lys
145                 150                 155                 160

Asn Ile Tyr Ser Asp Glu Glu Lys His Ser Thr Val Gly Tyr Arg Cys
                165                 170                 175

Ile Asn Glu Asn Leu Leu Ile Phe Met Ser Asn Ile Glu Val Tyr Gln
            180                 185                 190

Ile Tyr Lys Lys Ala Asn Ile Lys Asn Asp Asn Tyr Asp Glu Glu Thr
        195                 200                 205

Leu Asp Lys Thr Phe Met Ile Glu Ser Phe Asn Glu Cys Leu Thr Gln
    210                 215                 220

Ser Gly Val Glu Ala Tyr Asn Ser Val Val Ala Ser Ile Lys Thr Ala
225                 230                 235                 240

Thr Asn Leu Tyr Ile Gln Lys Asn Asn Lys Glu Glu Asn Phe Val Arg
                245                 250                 255
```

```
Val Pro Lys Met Lys Val Leu Phe Lys Gln Ile Leu Ser Asp Arg Thr
            260                 265                 270

Ser Leu Phe Asp Gly Leu Ile Ile Glu Ser Asp Asp Glu Leu Leu Asp
        275                 280                 285

Lys Leu Cys Ser Phe Ser Ala Glu Val Asp Lys Phe Leu Pro Ile Asn
    290                 295                 300

Ile Asp Arg Tyr Ile Lys Thr Leu Met Asp Ser Asn Asn Gly Thr Gly
305                 310                 315                 320

Ile Tyr Val Lys Asn Asp Ser Ser Leu Thr Thr Leu Ser Asn Tyr Leu
                325                 330                 335

Thr Asp Ser Trp Ser Ser Ile Arg Asn Ala Phe Asn Glu Asn Tyr Asp
            340                 345                 350

Ala Lys Tyr Thr Gly Lys Val Asn Asp Lys Tyr Glu Glu Lys Arg Glu
        355                 360                 365

Lys Ala Tyr Lys Ser Asn Asp Ser Phe Glu Leu Asn Tyr Ile Gln Asn
    370                 375                 380

Leu Leu Gly Ile Asn Val Ile Asp Lys Tyr Ile Glu Arg Ile Asn Phe
385                 390                 395                 400

Asp Ile Lys Glu Ile Cys Glu Ala Tyr Lys Glu Met Thr Lys Asn Cys
                405                 410                 415

Phe Glu Asp His Asp Lys Thr Lys Lys Leu Gln Lys Asn Ile Lys Ala
            420                 425                 430

Val Ala Ser Ile Lys Ser Tyr Leu Asp Ser Leu Lys Asn Ile Glu Arg
        435                 440                 445

Asp Ile Lys Leu Leu Asn Gly Thr Gly Leu Glu Ser Arg Asn Glu Phe
    450                 455                 460

Phe Tyr Gly Glu Gln Ser Thr Val Leu Glu Glu Ile Thr Lys Val Asp
465                 470                 475                 480

Glu Leu Tyr Asn Ile Thr Arg Asn Tyr Leu Thr Lys Lys Pro Phe Ser
                485                 490                 495

Thr Glu Lys Met Lys Leu Asn Phe Asn Asn Pro Gln Leu Leu Gly Gly
            500                 505                 510

Trp Asp Val Asn Lys Glu Arg Asp Cys Tyr Gly Val Ile Leu Ile Lys
        515                 520                 525

Asp Asn Asn Tyr Tyr Leu Gly Ile Met Asp Lys Ser Ala Asn Lys Ser
    530                 535                 540

Phe Leu Asn Ile Lys Glu Ser Lys Asn Glu Asn Ala Tyr Lys Lys Val
545                 550                 555                 560

Asn Cys Lys Leu Leu Pro Gly Pro Asn Lys Met Phe Pro Lys Val Phe
                565                 570                 575

Phe Ala Lys Ser Asn Ile Asp Tyr Tyr Asp Pro Thr His Glu Ile Lys
            580                 585                 590
```

```
Lys Leu Tyr Asp Lys Gly Thr Phe Lys Lys Gly Asn Ser Phe Asn Leu
        595                 600                 605

Glu Asp Cys His Lys Leu Ile Asp Phe Tyr Lys Glu Ser Ile Lys Lys
    610                 615                 620

Asn Asp Asp Trp Lys Asn Phe Asn Phe Asn Phe Ser Asp Thr Lys Asp
625                 630                 635                 640

Tyr Glu Asp Ile Ser Gly Phe Phe Arg Glu Val Glu Ala Gln Asn Tyr
                645                 650                 655

Lys Ile Thr Tyr Thr Asn Val Ser Cys Asp Phe Ile Glu Ser Leu Val
            660                 665                 670

Asp Glu Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser
        675                 680                 685

Glu Tyr Ala Thr Gly Asn Leu Asn Leu His Thr Leu Tyr Leu Lys Met
    690                 695                 700

Leu Phe Asp Glu Arg Asn Leu Lys Asp Leu Cys Ile Lys Met Asn Gly
705                 710                 715                 720

Glu Ala Glu Val Phe Tyr Arg Pro Ala Ser Ile Leu Asp Glu Asp Lys
                725                 730                 735

Val Val His Lys Ala Asn Gln Lys Ile Thr Asn Lys Asn Thr Asn Ser
            740                 745                 750

Lys Lys Lys Glu Ser Ile Phe Ser Tyr Asp Ile Val Lys Asp Lys Arg
        755                 760                 765

Tyr Thr Val Asp Lys Phe Phe Ile His Leu Pro Ile Thr Leu Asn Tyr
    770                 775                 780

Lys Glu Gln Asn Val Ser Arg Phe Asn Asp Tyr Ile Arg Glu Ile Leu
785                 790                 795                 800

Lys Lys Ser Lys Asn Ile Arg Val Ile Gly Ile Asp Arg Gly Glu Arg
                805                 810                 815

Asn Leu Leu Tyr Val Val Val Cys Asp Ser Asp Gly Ser Ile Leu Tyr
            820                 825                 830

Gln Arg Ser Ile Asn Glu Ile Val Ser Gly Ser His Lys Thr Asp Tyr
        835                 840                 845

His Lys Leu Leu Asp Asn Lys Glu Lys Glu Arg Leu Ser Ser Arg Arg
    850                 855                 860

Asp Trp Lys Thr Ile Glu Asn Ile Lys Asp Leu Lys Ala Gly Tyr Met
865                 870                 875                 880

Ser Gln Val Val Asn Glu Ile Tyr Asn Leu Ile Leu Lys Tyr Asn Ala
                885                 890                 895

Ile Val Val Leu Glu Asp Leu Asn Ile Gly Phe Lys Asn Gly Arg Lys
            900                 905                 910

Lys Val Glu Lys Gln Val Tyr Gln Asn Phe Glu Lys Ala Leu Ile Asp
        915                 920                 925
```

```
Lys Leu Asn Tyr Leu Cys Ile Asp Lys Thr Arg Glu Gln Leu Ser Pro
    930                 935                 940

Ser Ser Pro Gly Gly Val Leu Asn Ala Tyr Gln Leu Thr Ala Lys Phe
945                 950                 955                 960

Glu Ser Phe Glu Lys Ile Gly Lys Gln Thr Gly Cys Ile Phe Tyr Val
                965                 970                 975

Pro Ala Tyr Leu Thr Ser Gln Ile Asp Pro Thr Thr Gly Phe Val Asn
            980                 985                 990

Leu Phe Tyr Gln Lys Asp Thr Ser Lys Gln Gly Leu Gln Leu Phe Phe
        995                 1000                1005

Arg Lys  Phe Lys Lys Ile Asn  Phe Asp Lys Val Ala  Ser Asn Phe
    1010                1015                1020

Glu Phe  Val Phe Asp Tyr Asn  Asp Phe Thr Asn Lys  Ala Glu Gly
    1025                1030                1035

Thr Lys  Thr Asn Trp Thr Ile  Ser Thr Gln Gly Thr  Arg Ile Ala
    1040                1045                1050

Lys Tyr  Arg Ser Asp Asp Ala  Asn Gly Lys Trp Ile  Ser Arg Thr
    1055                1060                1065

Val His  Pro Thr Asp Ile Ile  Lys Glu Ala Leu Asn  Arg Glu Lys
    1070                1075                1080

Ile Asn  Tyr Asn Asp Gly His  Asp Leu Ile Asp Glu  Ile Val Ser
    1085                1090                1095

Ile Glu  Lys Ser Ala Val Leu  Lys Glu Ile Tyr Tyr  Gly Phe Lys
    1100                1105                1110

Leu Thr  Leu Gln Leu Arg Asn  Ser Thr Leu Ala Asn  Glu Glu Glu
    1115                1120                1125

Gln Glu  Asp Tyr Ile Ile Ser  Pro Val Lys Asn Ser  Ser Gly Asn
    1130                1135                1140

Tyr Phe  Asp Ser Arg Ile Thr  Ser Lys Glu Leu Pro  Cys Asp Ala
    1145                1150                1155

Asp Ala  Asn Gly Ala Tyr Asn  Ile Ala Arg Lys Gly  Leu Trp Ala
    1160                1165                1170

Leu Glu  Gln Ile Arg Asn Ser  Glu Asn Val Ser Lys  Val Lys Leu
    1175                1180                1185

Ala Ile  Ser Asn Lys Glu Trp  Phe Glu Tyr Thr Gln  Asn Asn Ile
    1190                1195                1200

Pro Ser  Leu
    1205
```

<210> 116
<211> 1205
<212> PRT
<213> 口小桿菌屬某種（Oribacterium sp.）

<400> 116

```
Met Tyr Tyr Asp Gly Leu Thr Lys Gln Tyr Ala Leu Ser Lys Thr Ile
1               5                  10                  15

Arg Asn Glu Leu Val Pro Ile Gly Lys Thr Leu Asp Asn Ile Lys Lys
            20                  25                  30

Asn Arg Ile Leu Glu Ala Asp Ile Lys Arg Lys Ser Asp Tyr Glu His
        35                  40                  45

Val Lys Lys Leu Met Asp Met Tyr His Lys Lys Ile Ile Asn Glu Ala
    50                  55                  60

Leu Asp Asn Phe Lys Leu Ser Val Leu Glu Asp Ala Ala Asp Ile Tyr
65                  70                  75                  80

Phe Asn Lys Gln Asn Asp Glu Arg Asp Ile Asp Ala Phe Leu Lys Ile
                85                  90                  95

Gln Asp Lys Leu Arg Lys Glu Ile Val Glu Gln Leu Lys Gly His Thr
            100                 105                 110

Asp Tyr Ser Lys Val Gly Asn Lys Asp Phe Leu Gly Leu Leu Lys Ala
        115                 120                 125

Ala Ser Thr Glu Glu Asp Arg Ile Leu Ile Glu Ser Phe Asp Asn Phe
    130                 135                 140

Tyr Thr Tyr Phe Thr Ser Tyr Asn Lys Val Arg Ser Asn Leu Tyr Ser
145                 150                 155                 160

Ala Glu Asp Lys Ser Ser Thr Val Ala Tyr Arg Leu Ile Asn Glu Asn
                165                 170                 175

Leu Pro Lys Phe Phe Asp Asn Ile Lys Ala Tyr Arg Thr Val Arg Asn
            180                 185                 190

Ala Gly Val Ile Ser Gly Asp Met Ser Ile Val Glu Gln Asp Glu Leu
        195                 200                 205

Phe Glu Val Asp Thr Phe Asn His Thr Leu Thr Gln Tyr Gly Ile Asp
    210                 215                 220

Thr Tyr Asn His Met Ile Gly Gln Leu Asn Ser Ala Ile Asn Leu Tyr
225                 230                 235                 240

Asn Gln Lys Met His Gly Ala Gly Ser Phe Lys Lys Leu Pro Lys Met
                245                 250                 255

Lys Glu Leu Tyr Lys Gln Leu Leu Thr Glu Arg Glu Glu Glu Phe Ile
            260                 265                 270

Glu Glu Tyr Thr Asp Asp Glu Val Leu Ile Thr Ser Val His Asn Tyr
        275                 280                 285

Val Ser Tyr Leu Ile Asp Tyr Leu Asn Ser Asp Lys Val Glu Ser Phe
    290                 295                 300

Phe Asp Thr Leu Arg Lys Ser Asp Gly Lys Glu Val Phe Ile Lys Asn
305                 310                 315                 320

Asp Val Ser Lys Thr Thr Met Ser Asn Ile Leu Phe Asp Asn Trp Ser
                325                 330                 335
```

Thr Ile Asp Asp Leu Ile Asn His Glu Tyr Asp Ser Ala Pro Glu Asn
340 345 350

Val Lys Lys Thr Lys Asp Asp Lys Tyr Phe Glu Lys Arg Gln Lys Asp
355 360 365

Leu Lys Lys Asn Lys Ser Tyr Ser Leu Ser Lys Ile Ala Ala Leu Cys
370 375 380

Arg Asp Thr Thr Ile Leu Glu Lys Tyr Ile Arg Arg Leu Val Asp Asp
385 390 395 400

Ile Glu Lys Ile Tyr Thr Ser Asn Asn Val Phe Ser Asp Ile Val Leu
405 410 415

Ser Lys His Asp Arg Ser Lys Lys Leu Ser Lys Asn Thr Asn Ala Val
420 425 430

Gln Ala Ile Lys Asn Met Leu Asp Ser Ile Lys Asp Phe Glu His Asp
435 440 445

Val Met Leu Ile Asn Gly Ser Gly Gln Glu Ile Lys Lys Asn Leu Asn
450 455 460

Val Tyr Ser Glu Gln Glu Ala Leu Ala Gly Ile Leu Arg Gln Val Asp
465 470 475 480

His Ile Tyr Asn Leu Thr Arg Asn Tyr Leu Thr Lys Lys Pro Phe Ser
485 490 495

Thr Glu Lys Ile Lys Leu Asn Phe Asn Arg Pro Thr Phe Leu Asp Gly
500 505 510

Trp Asp Lys Asn Lys Glu Glu Ala Asn Leu Gly Ile Leu Leu Ile Lys
515 520 525

Asp Asn Arg Tyr Tyr Leu Gly Ile Met Asn Thr Ser Ser Asn Lys Ala
530 535 540

Phe Val Asn Pro Pro Lys Ala Ile Ser Asn Asp Ile Tyr Lys Lys Val
545 550 555 560

Asp Tyr Lys Leu Leu Pro Gly Pro Asn Lys Met Leu Pro Lys Val Phe
565 570 575

Phe Ala Thr Lys Asn Ile Ala Tyr Tyr Ala Pro Ser Glu Glu Leu Leu
580 585 590

Ser Lys Tyr Arg Lys Gly Thr His Lys Lys Gly Asp Ser Phe Ser Ile
595 600 605

Asp Asp Cys Arg Asn Leu Ile Asp Phe Phe Lys Ser Ser Ile Asn Lys
610 615 620

Asn Thr Asp Trp Ser Thr Phe Gly Phe Asn Phe Ser Asp Thr Asn Ser
625 630 635 640

Tyr Asn Asp Ile Ser Asp Phe Tyr Arg Glu Val Glu Lys Gln Gly Tyr
645 650 655

Lys Leu Ser Phe Thr Asp Ile Asp Ala Cys Tyr Ile Lys Asp Leu Val
660 665 670

```
Asp Asn Asn Glu Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser
        675                 680                 685

Pro Tyr Ser Lys Gly Lys Leu Asn Leu His Thr Leu Tyr Phe Lys Met
    690                 695                 700

Leu Phe Asp Gln Arg Asn Leu Asp Asn Val Val Tyr Lys Leu Asn Gly
705                 710                 715                 720

Glu Ala Glu Val Phe Tyr Arg Pro Ala Ser Ile Glu Ser Asp Glu Gln
                725                 730                 735

Ile Ile His Lys Ser Gly Gln Asn Ile Lys Asn Lys Asn Gln Lys Arg
            740                 745                 750

Ser Asn Cys Lys Lys Thr Ser Thr Phe Asp Tyr Asp Ile Val Lys Asp
        755                 760                 765

Arg Arg Tyr Cys Lys Asp Lys Phe Met Leu His Leu Pro Ile Thr Val
    770                 775                 780

Asn Phe Gly Thr Asn Glu Ser Gly Lys Phe Asn Glu Leu Val Asn Asn
785                 790                 795                 800

Ala Ile Arg Ala Asp Lys Asp Val Asn Val Ile Gly Ile Asp Arg Gly
                805                 810                 815

Glu Arg Asn Leu Leu Tyr Val Val Val Val Asp Pro Cys Gly Lys Ile
            820                 825                 830

Ile Glu Gln Ile Ser Leu Asn Thr Ile Val Asp Lys Glu Tyr Asp Ile
        835                 840                 845

Glu Thr Asp Tyr His Gln Leu Leu Asp Glu Lys Glu Gly Ser Arg Asp
    850                 855                 860

Lys Ala Arg Lys Asp Trp Asn Thr Ile Glu Asn Ile Lys Glu Leu Lys
865                 870                 875                 880

Glu Gly Tyr Leu Ser Gln Val Val Asn Ile Ile Ala Lys Leu Val Leu
                885                 890                 895

Lys Tyr Asp Ala Ile Ile Cys Leu Glu Asp Leu Asn Phe Gly Phe Lys
            900                 905                 910

Arg Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu Lys
        915                 920                 925

Met Leu Ile Asp Lys Met Asn Tyr Leu Val Leu Asp Lys Ser Arg Lys
    930                 935                 940

Gln Glu Ser Pro Gln Lys Pro Gly Gly Ala Leu Asn Ala Leu Gln Leu
945                 950                 955                 960

Thr Ser Ala Phe Lys Ser Phe Lys Glu Leu Gly Lys Gln Thr Gly Ile
                965                 970                 975

Ile Tyr Tyr Val Pro Ala Tyr Leu Thr Ser Lys Ile Asp Pro Thr Thr
            980                 985                 990

Gly Phe Ala Asn Leu Phe Tyr Ile Lys Tyr Glu Ser Val Asp Lys Ala
        995                 1000                1005
```

Arg Asp Phe Phe Ser Lys Phe Asp Phe Ile Arg Tyr Asn Gln Met
1010 1015 1020

Asp Asn Tyr Phe Glu Phe Gly Phe Asp Tyr Lys Ser Phe Thr Glu
1025 1030 1035

Arg Ala Ser Gly Cys Lys Ser Lys Trp Ile Ala Cys Thr Asn Gly
1040 1045 1050

Glu Arg Ile Val Lys Tyr Arg Asn Ser Asp Lys Asn Asn Ser Phe
1055 1060 1065

Asp Asp Lys Thr Val Ile Leu Thr Asp Glu Tyr Arg Ser Leu Phe
1070 1075 1080

Asp Lys Tyr Leu Gln Asn Tyr Ile Asp Glu Asp Asp Leu Lys Asp
1085 1090 1095

Gln Ile Leu Gln Ile Asp Ser Ala Asp Phe Tyr Lys Asn Leu Ile
1100 1105 1110

Lys Leu Phe Gln Leu Thr Leu Gln Met Arg Asn Ser Ser Ser Asp
1115 1120 1125

Gly Lys Arg Asp Tyr Ile Ile Ser Pro Val Lys Asn Tyr Arg Glu
1130 1135 1140

Glu Phe Phe Cys Ser Glu Phe Ser Asp Asp Thr Phe Pro Arg Asp
1145 1150 1155

Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala Arg Lys Gly Leu Trp
1160 1165 1170

Val Ile Lys Gln Ile Arg Glu Thr Lys Ser Gly Thr Lys Ile Asn
1175 1180 1185

Leu Ala Met Ser Asn Ser Glu Trp Leu Glu Tyr Ala Gln Cys Asn
1190 1195 1200

Leu Leu
1205

<210> 117
<211> 1125
<212> PRT
<213> 瘤胃假丁酸弧菌（Pseudobutyrivibrio ruminis）

<400> 117
Phe Asn Leu Ser Lys Glu Lys Asn Ser Val Asp Ala Phe Ser Lys Cys
1 5 10 15

Gln Asp Lys Leu Arg Lys Glu Ile Val Ser Leu Leu Lys Asn His Glu
20 25 30

Asn Phe Pro Lys Ile Gly Asn Lys Glu Ile Ile Lys Leu Leu Gln Ser
35 40 45

Leu Tyr Asp Asn Asp Thr Asp Tyr Lys Ala Leu Asp Ser Phe Ser Asn
50 55 60

Phe Tyr Thr Tyr Phe Ser Ser Tyr Asn Glu Val Arg Lys Asn Leu Tyr
65 70 75 80

Ser Asp Glu Glu Lys Ser Ser Thr Val Ala Tyr Arg Leu Ile Asn Glu
85 90 95

```
Asn Leu Pro Lys Phe Leu Asp Asn Ile Lys Ala Tyr Ala Ile Ala Lys
            100                 105                 110

Lys Ala Gly Val Arg Ala Glu Gly Leu Ser Glu Glu Asp Gln Asp Cys
        115                 120                 125

Leu Phe Ile Ile Glu Thr Phe Glu Arg Thr Leu Thr Gln Asp Gly Ile
    130                 135                 140

Asp Asn Tyr Asn Ala Ala Ile Gly Lys Leu Asn Thr Ala Ile Asn Leu
145                 150                 155                 160

Phe Asn Gln Gln Asn Lys Lys Gln Glu Gly Phe Arg Lys Val Pro Gln
                165                 170                 175

Met Lys Cys Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Glu Ala Phe
            180                 185                 190

Ile Asp Glu Phe Ser Asp Asp Glu Asp Leu Ile Thr Asn Ile Glu Ser
        195                 200                 205

Phe Ala Glu Asn Met Asn Val Phe Leu Asn Ser Glu Ile Ile Thr Asp
    210                 215                 220

Phe Lys Ile Ala Leu Val Glu Ser Asp Gly Ser Leu Val Tyr Ile Lys
225                 230                 235                 240

Asn Asp Val Ser Lys Thr Ser Phe Ser Asn Ile Val Phe Gly Ser Trp
                245                 250                 255

Asn Ala Ile Asp Glu Lys Leu Ser Asp Glu Tyr Asp Leu Ala Asn Ser
            260                 265                 270

Lys Lys Lys Lys Asp Glu Lys Tyr Tyr Glu Lys Arg Gln Lys Glu Leu
        275                 280                 285

Lys Lys Asn Lys Ser Tyr Asp Leu Glu Thr Ile Ile Gly Leu Phe Asp
    290                 295                 300

Asp Asn Ser Asp Val Ile Gly Lys Tyr Ile Glu Lys Leu Glu Ser Asp
305                 310                 315                 320

Ile Thr Ala Ile Ala Glu Ala Lys Asn Asp Phe Asp Glu Ile Val Leu
                325                 330                 335

Arg Lys His Asp Lys Asn Lys Ser Leu Arg Lys Asn Thr Asn Ala Val
            340                 345                 350

Glu Ala Ile Lys Ser Tyr Leu Asp Thr Val Lys Asp Phe Glu Arg Asp
        355                 360                 365

Ile Lys Leu Ile Asn Gly Ser Gly Gln Glu Val Glu Lys Asn Leu Val
    370                 375                 380

Val Tyr Ala Glu Gln Glu Asn Ile Leu Ala Glu Ile Lys Asn Val Asp
385                 390                 395                 400

Ser Leu Tyr Asn Met Ser Arg Asn Tyr Leu Thr Gln Lys Pro Phe Ser
                405                 410                 415

Thr Glu Lys Phe Lys Leu Asn Phe Asn Arg Ala Thr Leu Leu Asn Gly
            420                 425                 430
```

Trp Asp Lys Asn Lys Glu Thr Asp Asn Leu Gly Ile Leu Phe Glu Lys
435 440 445

Asp Gly Met Tyr Tyr Leu Gly Ile Met Asn Thr Lys Ala Asn Lys Ile
450 455 460

Phe Val Asn Ile Pro Lys Ala Thr Ser Asn Asp Val Tyr His Lys Val
465 470 475 480

Asn Tyr Lys Leu Leu Pro Gly Pro Asn Lys Met Leu Pro Lys Val Phe
485 490 495

Phe Ala Gln Ser Asn Leu Asp Tyr Tyr Lys Pro Ser Glu Glu Leu Leu
500 505 510

Ala Lys Tyr Lys Ala Gly Thr His Lys Lys Gly Asp Asn Phe Ser Leu
515 520 525

Glu Asp Cys His Ala Leu Ile Asp Phe Phe Lys Ala Ser Ile Glu Lys
530 535 540

His Pro Asp Trp Ser Ser Phe Gly Phe Glu Phe Ser Glu Thr Cys Thr
545 550 555 560

Tyr Glu Asp Leu Ser Gly Phe Tyr Arg Glu Val Glu Lys Gln Gly Tyr
565 570 575

Lys Ile Thr Tyr Thr Asp Val Asp Ala Asp Tyr Ile Thr Ser Leu Val
580 585 590

Glu Arg Asp Glu Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser
595 600 605

Pro Tyr Ser Lys Gly Asn Leu Asn Leu His Thr Ile Tyr Leu Gln Met
610 615 620

Leu Phe Asp Gln Arg Asn Leu Asn Asn Val Val Tyr Lys Leu Asn Gly
625 630 635 640

Glu Ala Glu Val Phe Tyr Arg Pro Ala Ser Ile Asn Asp Glu Glu Val
645 650 655

Ile Ile His Lys Ala Gly Glu Glu Ile Lys Asn Lys Asn Ser Lys Arg
660 665 670

Ala Val Asp Lys Pro Thr Ser Lys Phe Gly Tyr Asp Ile Ile Lys Asp
675 680 685

Arg Arg Tyr Ser Lys Asp Lys Phe Met Leu His Ile Pro Val Thr Met
690 695 700

Asn Phe Gly Val Asp Glu Thr Arg Arg Phe Asn Asp Val Val Asn Asp
705 710 715 720

Ala Leu Arg Asn Asp Glu Lys Val Arg Val Ile Gly Ile Asp Arg Gly
725 730 735

Glu Arg Asn Leu Leu Tyr Val Val Val Val Asp Thr Asp Gly Thr Ile
740 745 750

Leu Glu Gln Ile Ser Leu Asn Ser Ile Ile Asn Asn Glu Tyr Ser Ile
755 760 765

```
Glu Thr Asp Tyr His Lys Leu Leu Asp Glu Lys Glu Gly Asp Arg Asp
    770                 775                 780

Arg Ala Arg Lys Asn Trp Thr Thr Ile Glu Asn Ile Lys Glu Leu Lys
785                 790                 795                 800

Glu Gly Tyr Leu Ser Gln Val Val Asn Val Ile Ala Lys Leu Val Leu
                805                 810                 815

Lys Tyr Asn Ala Ile Ile Cys Leu Glu Asp Leu Asn Phe Gly Phe Lys
            820                 825                 830

Arg Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu Lys
        835                 840                 845

Met Leu Ile Asp Lys Leu Asn Tyr Leu Val Ile Asp Lys Ser Arg Lys
    850                 855                 860

Gln Asp Lys Pro Glu Glu Phe Gly Gly Ala Leu Asn Ala Leu Gln Leu
865                 870                 875                 880

Thr Ser Lys Phe Thr Ser Phe Lys Asp Met Gly Lys Gln Thr Gly Ile
                885                 890                 895

Ile Tyr Tyr Val Pro Ala Tyr Leu Thr Ser Lys Ile Asp Pro Thr Thr
            900                 905                 910

Gly Phe Ala Asn Leu Phe Tyr Val Lys Tyr Glu Asn Val Glu Lys Ala
        915                 920                 925

Lys Glu Phe Phe Ser Arg Phe Asp Ser Ile Ser Tyr Asn Asn Glu Ser
    930                 935                 940

Gly Tyr Phe Glu Phe Ala Phe Asp Tyr Lys Lys Phe Thr Asp Arg Ala
945                 950                 955                 960

Cys Gly Ala Arg Ser Gln Trp Thr Val Cys Thr Tyr Gly Glu Arg Ile
                965                 970                 975

Ile Lys Phe Arg Asn Thr Glu Lys Asn Asn Ser Phe Asp Asp Lys Thr
            980                 985                 990

Ile Val Leu Ser Glu Glu Phe Lys Glu Leu Phe Ser Ile Tyr Gly Ile
        995                 1000                1005

Ser Tyr  Glu Asp Gly Ala Glu Leu Lys Asn Lys  Ile Met Ser Val
    1010                1015                1020

Asp Glu  Ala Asp Phe Phe Arg Ser Leu Thr Arg  Leu Phe Gln Gln
    1025                1030                1035

Thr Met  Gln Met Arg Asn Ser Ser Asn Asp Val  Thr Arg Asp Tyr
    1040                1045                1050

Ile Ile  Ser Pro Ile Met Asn Asp Arg Gly Glu  Phe Phe Asn Ser
    1055                1060                1065

Glu Ala  Cys Asp Ala Ser Lys Pro Lys Asp Ala  Asp Ala Asn Gly
    1070                1075                1080

Ala Phe  Asn Ile Ala Arg Lys Gly Leu Trp Val  Leu Glu Gln Ile
    1085                1090                1095
```

```
Arg Asn Thr Pro Ser Gly Asp Lys Leu Asn Leu Ala Met Ser Asn
    1100                1105                1110

Ala Glu Trp Leu Glu Tyr Ala Gln Arg Asn Gln Ile
    1115                1120                1125
```

<210> 118
<211> 1154
<212> PRT
<213> Proteocatella sphenisci

<400> 118

```
Met Glu Asn Phe Lys Asn Leu Tyr Pro Ile Asn Lys Thr Leu Arg Phe
1               5                   10                  15

Glu Leu Arg Pro Tyr Gly Lys Thr Leu Glu Asn Phe Lys Lys Ser Gly
            20                  25                  30

Leu Leu Glu Lys Asp Ala Phe Lys Ala Asn Ser Arg Arg Ser Met Gln
        35                  40                  45

Ala Ile Ile Asp Glu Lys Phe Lys Glu Thr Ile Glu Glu Arg Leu Lys
    50                  55                  60

Tyr Thr Glu Phe Ser Glu Cys Asp Leu Gly Asn Met Thr Ser Lys Asp
65                  70                  75                  80

Lys Lys Ile Thr Asp Lys Ala Ala Thr Asn Leu Lys Lys Gln Val Ile
                85                  90                  95

Leu Ser Phe Asp Asp Glu Ile Phe Asn Asn Tyr Leu Lys Pro Asp Lys
            100                 105                 110

Asn Ile Asp Ala Leu Phe Lys Asn Asp Pro Ser Asn Pro Val Ile Ser
        115                 120                 125

Thr Phe Lys Gly Phe Thr Thr Tyr Phe Val Asn Phe Phe Glu Ile Arg
    130                 135                 140

Lys His Ile Phe Lys Gly Glu Ser Ser Gly Ser Met Ala Tyr Arg Ile
145                 150                 155                 160

Ile Asp Glu Asn Leu Thr Thr Tyr Leu Asn Asn Ile Glu Lys Ile Lys
                165                 170                 175

Lys Leu Pro Glu Glu Leu Lys Ser Gln Leu Glu Gly Ile Asp Gln Ile
            180                 185                 190

Asp Lys Leu Asn Asn Tyr Asn Glu Phe Ile Thr Gln Ser Gly Ile Thr
        195                 200                 205

His Tyr Asn Glu Ile Ile Gly Gly Ile Ser Lys Ser Glu Asn Val Lys
    210                 215                 220

Ile Gln Gly Ile Asn Glu Gly Ile Asn Leu Tyr Cys Gln Lys Asn Lys
225                 230                 235                 240

Val Lys Leu Pro Arg Leu Thr Pro Leu Tyr Lys Met Ile Leu Ser Asp
                245                 250                 255

Arg Val Ser Asn Ser Phe Val Leu Asp Thr Ile Glu Asn Asp Thr Glu
            260                 265                 270
```

```
Leu Ile Glu Met Ile Ser Asp Leu Ile Asn Lys Thr Glu Ile Ser Gln
        275                 280                 285

Asp Val Ile Met Ser Asp Ile Gln Asn Ile Phe Ile Lys Tyr Lys Gln
    290                 295                 300

Leu Gly Asn Leu Pro Gly Ile Ser Tyr Ser Ser Ile Val Asn Ala Ile
305                 310                 315                 320

Cys Ser Asp Tyr Asp Asn Asn Phe Gly Asp Gly Lys Arg Lys Lys Ser
                325                 330                 335

Tyr Glu Asn Asp Arg Lys Lys His Leu Glu Thr Asn Val Tyr Ser Ile
            340                 345                 350

Asn Tyr Ile Ser Glu Leu Leu Thr Asp Thr Asp Val Ser Ser Asn Ile
        355                 360                 365

Lys Met Arg Tyr Lys Glu Leu Glu Gln Asn Tyr Gln Val Cys Lys Glu
    370                 375                 380

Asn Phe Asn Ala Thr Asn Trp Met Asn Ile Lys Asn Ile Lys Gln Ser
385                 390                 395                 400

Glu Lys Thr Asn Leu Ile Lys Asp Leu Leu Asp Ile Leu Lys Ser Ile
                405                 410                 415

Gln Arg Phe Tyr Asp Leu Phe Asp Ile Val Asp Glu Asp Lys Asn Pro
            420                 425                 430

Ser Ala Glu Phe Tyr Thr Trp Leu Ser Lys Asn Ala Glu Lys Leu Asp
        435                 440                 445

Phe Glu Phe Asn Ser Val Tyr Asn Lys Ser Arg Asn Tyr Leu Thr Arg
    450                 455                 460

Lys Gln Tyr Ser Asp Lys Lys Ile Lys Leu Asn Phe Asp Ser Pro Thr
465                 470                 475                 480

Leu Ala Lys Gly Trp Asp Ala Asn Lys Glu Ile Asp Asn Ser Thr Ile
                485                 490                 495

Ile Met Arg Lys Phe Asn Asn Asp Arg Gly Asp Tyr Asp Tyr Phe Leu
            500                 505                 510

Gly Ile Trp Asn Lys Ser Thr Pro Ala Asn Glu Lys Ile Ile Pro Leu
        515                 520                 525

Glu Asp Asn Gly Leu Phe Glu Lys Met Gln Tyr Lys Leu Tyr Pro Asp
    530                 535                 540

Pro Ser Lys Met Leu Pro Lys Gln Phe Leu Ser Lys Ile Trp Lys Ala
545                 550                 555                 560

Lys His Pro Thr Thr Pro Glu Phe Asp Lys Lys Tyr Lys Glu Gly Arg
                565                 570                 575

His Lys Lys Gly Pro Asp Phe Glu Lys Glu Phe Leu His Glu Leu Ile
            580                 585                 590

Asp Cys Phe Lys His Gly Leu Val Asn His Asp Glu Lys Tyr Gln Asp
        595                 600                 605
```

Val Phe Gly Phe Asn Leu Arg Asn Thr Glu Asp Tyr Asn Ser Tyr Thr
610 615 620

Glu Phe Leu Glu Asp Val Glu Arg Cys Asn Tyr Asn Leu Ser Phe Asn
625 630 635 640

Lys Ile Ala Asp Thr Ser Asn Leu Ile Asn Asp Gly Lys Leu Tyr Val
645 650 655

Phe Gln Ile Trp Ser Lys Asp Phe Ser Ile Asp Ser Lys Gly Thr Lys
660 665 670

Asn Leu Asn Thr Ile Tyr Phe Glu Ser Leu Phe Ser Glu Glu Asn Met
675 680 685

Ile Glu Lys Met Phe Lys Leu Ser Gly Glu Ala Glu Ile Phe Tyr Arg
690 695 700

Pro Ala Ser Leu Asn Tyr Cys Glu Asp Ile Ile Lys Lys Gly His His
705 710 715 720

His Ala Glu Leu Lys Asp Lys Phe Asp Tyr Pro Ile Ile Lys Asp Lys
725 730 735

Arg Tyr Ser Gln Asp Lys Phe Phe Phe His Val Pro Met Val Ile Asn
740 745 750

Tyr Lys Ser Glu Lys Leu Asn Ser Lys Ser Leu Asn Asn Arg Thr Asn
755 760 765

Glu Asn Leu Gly Gln Phe Thr His Ile Ile Gly Ile Asp Arg Gly Glu
770 775 780

Arg His Leu Ile Tyr Leu Thr Val Val Asp Val Ser Thr Gly Glu Ile
785 790 795 800

Val Glu Gln Lys His Leu Asp Glu Ile Ile Asn Thr Asp Thr Lys Gly
805 810 815

Val Glu His Lys Thr His Tyr Leu Asn Lys Leu Glu Glu Lys Ser Lys
820 825 830

Thr Arg Asp Asn Glu Arg Lys Ser Trp Glu Ala Ile Glu Thr Ile Lys
835 840 845

Glu Leu Lys Glu Gly Tyr Ile Ser His Val Ile Asn Glu Ile Gln Lys
850 855 860

Leu Gln Glu Lys Tyr Asn Ala Leu Ile Val Met Glu Asn Leu Asn Tyr
865 870 875 880

Gly Phe Lys Asn Ser Arg Ile Lys Val Glu Lys Gln Val Tyr Gln Lys
885 890 895

Phe Glu Thr Ala Leu Ile Lys Lys Phe Asn Tyr Ile Ile Asp Lys Lys
900 905 910

Asp Pro Glu Thr Tyr Ile His Gly Tyr Gln Leu Thr Asn Pro Ile Thr
915 920 925

Thr Leu Asp Lys Ile Gly Asn Gln Ser Gly Ile Val Leu Tyr Ile Pro
930 935 940

Ala Trp Asn Thr Ser Lys Ile Asp Pro Val Thr Gly Phe Val Asn Leu
945 950 955 960

Leu Tyr Ala Asp Asp Leu Lys Tyr Lys Asn Gln Glu Gln Ala Lys Ser
965 970 975

Phe Ile Gln Lys Ile Asp Asn Ile Tyr Phe Glu Asn Gly Glu Phe Lys
980 985 990

Phe Asp Ile Asp Phe Ser Lys Trp Asn Asn Arg Tyr Ser Ile Ser Lys
995 1000 1005

Thr Lys Trp Thr Leu Thr Ser Tyr Gly Thr Arg Ile Gln Thr Phe
1010 1015 1020

Arg Asn Pro Gln Lys Asn Asn Lys Trp Asp Ser Ala Glu Tyr Asp
1025 1030 1035

Leu Thr Glu Glu Phe Lys Leu Ile Leu Asn Ile Asp Gly Thr Leu
1040 1045 1050

Lys Ser Gln Asp Val Glu Thr Tyr Lys Lys Phe Met Ser Leu Phe
1055 1060 1065

Lys Leu Met Leu Gln Leu Arg Asn Ser Val Thr Gly Thr Asp Ile
1070 1075 1080

Asp Tyr Met Ile Ser Pro Val Thr Asp Lys Thr Gly Thr His Phe
1085 1090 1095

Asp Ser Arg Glu Asn Ile Lys Asn Leu Pro Ala Asp Ala Asp Ala
1100 1105 1110

Asn Gly Ala Tyr Asn Ile Ala Arg Lys Gly Ile Met Ala Ile Glu
1115 1120 1125

Asn Ile Met Asn Gly Ile Ser Asp Pro Leu Lys Ile Ser Asn Glu
1130 1135 1140

Asp Tyr Leu Lys Tyr Ile Gln Asn Gln Gln Glu
1145 1150

<210> 119
<211> 34
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 119
ttagagaagt catttaataa ggccactgtt aaaa 34

<210> 120
<211> 34
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>
<223> 组合的DNA/RNA分子的描述：合成的
寡核苷酸

<400> 120
ttcgagaagu cauuuaauaa ggccacuguu aaaa 34

<210> 121
<211> 34
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>
<223> 组合的DNA/RNA分子的描述：合成的
寡核苷酸

<400> 121
ttcgagaagu cauuuaauaa ggccacuguu aaaa 34

<210> 122
<211> 34
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>
<223> 组合的DNA/RNA分子的描述：合成的
寡核苷酸

<400> 122
ttcgagaagu cauuuaauaa ggccacuguu aaaa 34

<210> 123
<211> 34
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>
<223> 组合的DNA/RNA分子的描述：合成的
寡核苷酸

<400> 123
ttcgagaagu cauuuaauaa ggccacuguu aaaa 34

<210> 124
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 124
ctgatggtcc atgtctgtta ctcg 24

<210> 125
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
引物

<400> 125
ctgggactca ggcgggtcac 20

<210> 126
<211> 25
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
引物

<400> 126
cctcacacaa cagcttcatg tcagc 25

<210> 127
<211> 3960
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 127
atggccccaa agaagaagcg gaaggtcggt atccacggag tcccagcagc cggtagtaac 60

atcaaaaact ttaccgggct ctacccccctc agcaaaactt tgcgctttga actcaagcct 120

attggcaaaa ccaaggaaaa catcgagaaa aatggcatcc tgaccaagga cgagcaacgg 180

gctaaagact acctcatagt caaaggcttt attgacgagt atcacaagca gttcatcaaa 240

gacaggcttt gggactttaa attgcctctc gaaagtgagg gggagaagaa cagtctcgaa 300

gaataccagg aactgtacga gctcactaag cgcaacgatg cccaggaggc cgacttcacc 360

gagattaaag ataaccttcg cagctctatt accgaacagc tcacgaagtc tggatctgcg 420

tacgatcgga tttttaaaaa agagttcatt agagaagacc tggtcaactt cctcgaagat 480

gaaaaagata aaaatatcgt gaaacagttc gaggacttta ctacatattt tacgggtttt 540

tatgaaaata ggaagaacat gtactctagc gaagagaagt ccacggccat cgcataccgg 600

cttatccatc agaatctgcc aaaattcatg gacaacatga gaagttttgc caaaattgca 660

aattccagtg tttccgagca ctttagcgac atctatgaaa gctggaagga atatctgaat 720

gtaaatagca tcgaggaaat cttccagctc gactatttta gcgaaacctt gactcagcca 780

catattgagg tgtataacta tattatcggg aagaaagtcc tggaagacgg aaccgagata 840

aagggcatca acgagtatgt gaacctctac aatcagcagc agaaagataa gagtaaacga 900

ctgcctttcc tggtgccact gtataagcaa attttgtctg atagggaaaa actctcctgg 960

attgctgaag agttcgacag cgacaagaag atgctgagcg ctatcaccga gtcttacaac 1020

cacctgcaca acgtgttgat gggtaacgag aacgaaagcc tgcgaaatct gctgctgaat 1080

attaaggact ataacctgga gaaaattaat atcacaaacg acttgtctct caccgaaatc 1140

tcccagaatc tttttggccg atatgatgta ttcacaaatg ggatcaaaaa caagctgaga 1200

gtgttgactc caaggaagaa aaaggagacg gacgaaaatt ttgaggaccg cattaacaaa 1260

atttttaaga cccagaagtc cttcagcatc gcttttctga acaagctgcc tcagcccgaa 1320

atggaggatg ggaagccccg gaacattgag gactatttca ttacacaggg ggcgattaac 1380

accaaatcta tacagaaaga agatatcttc gcccaaattg agaatgcata cgaggatgca 1440

caggtgttcc tgcaaattaa ggacaccgac aacaaactta gccagaacaa gacggcggtg 1500

gaaaagatca aaactttgct ggacgccttg aaggaactcc agcacttcat caaaccgctg 1560

ctgggctctg gggaggagaa cgagaaagac gaactgttct acggttcctt cctggccatc 1620

tgggacgaac tggacaccat tacaccactt tataacaaag tgagaaattg gctgacccga 1680

aaaccatatt caacagaaaa aatcaaattg aatttcgaca acgctcagct gctgggaggg 1740

tgggatgtca ataaagaaca cgactgtgca ggtatcttgt tgcggaaaaa cgatagctac 1800

tatctcggaa ttatcaataa gaaaaccaac cacatctttg atacggatat tacgccatca 1860

gatggcgagt gctatgacaa aatcgactac aagctccttc ccggggcgaa caaaatgctt 1920

ccaaaggtgt tttttagtaa gtcccgaatc aaagagttcg agccatcaga ggccataatc 1980

aattgctata agaaggggac acacaaaaaa ggaaaaaact ttaacctgac ggactgtcac 2040

cgcctgatca actttttttaa gacctcaatc gagaaacacg aggattggtc aaaattcgga 2100

ttcaagttct ccgataccga aacgtatgag gatattagcg gtttttatag agaggtcgag 2160

cagcagggat acaggctgac gagccatcca gtcagtgcca gctatataca tagtctggtc 2220

aaggaaggaa aactgtacct cttccaaatc tggaacaagg acttttctca attctccaag 2280

gggaccccta acttgcacac tctctattgg aagatgctgt ttgacaaacg gaatcttagc 2340

gatgtggttt ataagctgaa tggccaggct gaagtgttct atagaaagag ctccattgaa 2400

caccagaacc gaattatcca ccccgctcag catcccatca caaataagaa tgagcttaac 2460

aaaaagcaca ctagcacctt caaatacgat atcatcaaag atcgcagata cacggtggat 2520

aaattccagt tccatgtgcc cattactata aattttaagg cgaccgggca gaacaacatc 2580

aacccaatcg tccaagaggt gattcgccaa aacggtatca cccacatcat aggcatcgat 2640

cgaggtgaac gccatcttct gtacctctct ctcatcgatt tgaaaggcaa catcatcaag 2700

cagatgactc tcaacgaaat tattaatgag tataagggtg tgacctataa gaccaactac 2760

cataacctcc tggagaagag ggagaaggag cggaccgagg ccagacactc ctggagtagt 2820

attgaaagca taaaagaact gaaggatgga tacatgtcac aggtgattca caaaattacg 2880

gacatgatgg ttaagtacaa tgcgattgtg gtcctggagg acctcaacgg ggggtttatg 2940

cgaggccgcc agaaggtcga gaagcaggtg taccagaaat ttgaaaaaaa gttgatcgac 3000

aagctgaact atctcgttga caagaaactc gacgctaacg aggtcggcgg agtactgaat 3060

gcttatcagc tgaccaacaa gttcgagtct ttcaagaaga ttgggaaaca aagcggattt 3120

ttgttctaca tccccgcctg gaacacaagc aaaatcgatc ctataacagg gttcgttaat 3180

ctgttcaaca ccaggtacga gtctatcaag gagacaaaag ttttttggtc taagtttgat 3240

attatccgat acaataaaga gaagaattgg ttcgagttcg tcttcgatta caataccttt 3300

acgactaaag cggagggaac acgcactaag tggactctgt gcacccacgg cactcgcatc 3360

cagacattcc ggaacccaga aaagaatgcc cagtgggaca ataaagagat caatttgact 3420

gagtccttca aagctctgtt tgaaaagtac aagatcgata tcaccagtaa tctcaaggaa 3480

tccatcatgc aggaaaccga gaagaagttc ttccaggaac tgcataatct gctccacctg 3540

accctgcaga tgaggaatag cgttactgga accgacatag actatttgat cagccccgtt 3600

gccgatgagg atggaaattt ctatgatagt cgcataaatg gcaaaaattt tccggagaat 3660

gccgatgcca atggcgcgta caacatcgca cgaaagggtc tgatgcttat tcggcagatc 3720

aagcaagcag atccacagaa gaaattcaag tttgagacaa tcaccaataa agactggctg 3780

aaattcgccc aagacaagcc ctatcttaaa gatggcagcg ggaaaaggcc ggcggccacg 3840

aaaaaggccg gccaggcaaa aaagaaaaag ggatcctacc catacgatgt tccagattac 3900

gcttatccct acgacgtgcc tgattatgca tacccatacg atgtccccga ctatgcctaa 3960

<210> 128
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 128
cctcactcct gctcggtgaa ttt 23

<210> 129
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 129
ctgatggtcc atgtctgtta ctc 23

<210> 130
<211> 6
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
6xHis标签

<400> 130
His His His His His His
1 5

<210> 131
<211> 31
<212> DNA
<213> 新殺手弗朗西絲菌（Francisella novicida）

<400> 131
gagaagtcat ttaataaggc cactgttaaa a 31

<210> 132
<211> 30
<212> DNA
<213> 新殺手弗朗西絲菌

<400> 132
gctactattc ctgtgccttc agataattca 30

<210> 133
<211> 27
<212> DNA
<213> 新殺手弗朗西絲菌

<400> 133
gtctagagcc ttttgtatta gtagccg 27

<210> 134
<211> 98
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>
<221> 修飾的鹼基
<222> (30)..(37)
<223> a、c、t、g、未知或其他

<400> 134
ggccagtgaa ttcgagctcg gtacccgggn nnnnnnngag aagtcattta ataaggccac 60

tgttaaaaag cttggcgtaa tcatggtcat agctgttt 98

<210> 135
<211> 98
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>
<221> 修飾的鹼基
<222> (61)..(68)

<223> a, c, t, g, 未知 or other

<400> 135
ggccagtgaa ttcgagctcg gtacccgggg agaagtcatt taataaggcc actgttaaaa 60

nnnnnnnnag cttggcgtaa tcatggtcat agctgttt 98

<210> 136
<211> 25
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 136
gctgacatga agctgttgtg tgagg 25

<210> 137
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
引物

<400> 137
ggccagtgaa ttcgagctcg g 21

<210> 138
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
引物

<400> 138
caatttcaca caggaaacag ctatgacc 28

<210> 139
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
引物

<400> 139
cggggctggc ttaactatgc g 21

<210> 140
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
引物

<400> 140
gcccaatacg caaaccgcct 20

<210> 141
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
引物

<400> 141
ccatcccctt ctgtgaatgt 20

<210> 142
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的引物

<400> 142
tctccgtgtc tccaatctcc 20

<210> 143
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的引物

<400> 143
ctgggactca ggcgggtcac 20

<210> 144
<211> 25
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的引物

<400> 144
gctgacatga agctgttgtg tgagg 25

<210> 145
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 145
gagaagtcat ttaataaggc cact 24

<210> 146
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 146
gagaagtcat ttaataaggc ca 22

<210> 147
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 147
gagaagtcat ttaataaggc 20

<210> 148
<211> 18
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 148
gagaagtcat ttaataag 18

<210> 149
<211> 17
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 149
gagaagtcat ttaataa 17

<210> 150
<211> 16
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 150
gagaagtcat ttaata 16

<210> 151
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 151
gataagtcat ttaataaggc cact 24

<210> 152
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 152
gagaaggcat ttaataaggc cact 24

<210> 153
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 153
gagaagtcat gtaataaggc cact 24

<210> 154
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 154

gagaagtcat ttaagaaggc cact 24

<210> 155
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 155
gagaagtcat ttaataagtc cact 24

<210> 156
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 156
gagaagtcat ttaataaggc caat 24

<210> 157
<211> 42
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 157
atttctactg ttgtagatga gaagtcattt aataaggcca ct 42

<210> 158
<211> 41
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 158
tttctactgt tgtagatgag aagtcattta ataaggccac t 41

<210> 159
<211> 40
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 159
ttctactgtt gtagatgaga agtcatttaa taaggccact 40

<210> 160
<211> 39
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 160
tctactgttg tagatgagaa gtcatttaat aaggccact 39

<210> 161
<211> 35
<212> DNA

<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 161
ctgttgtaga tgagaagtca tttaataagg ccact 35

<210> 162
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 162
tgtagatgag aagtcattta ataaggccac t 31

<210> 163
<211> 19
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 163
aatttctgct gttgcagat 19

<210> 164
<211> 19
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 164
aatttccact gttgtggat 19

<210> 165
<211> 19
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 165
aattcctact gttgtaggt 19

<210> 166
<211> 19
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 166
aatttatact gttgtagat 19

<210> 167
<211> 19
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 167
aatttcgact gttgtagat 19

<210> 168
<211> 19
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 168
aatttctagt gttgtagat 19

<210> 169
<211> 19
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 169
aatttctact attgtagat 19

<210> 170
<211> 19
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 170
aatttctact gctgtagat 19

<210> 171
<211> 18
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 171
aatttctact ttgtagat 18

<210> 172
<211> 17
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 172
aatttctact tgtagat 17

<210> 173
<211> 19
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 173
aatttctact tttgtagaa 19

<210> 174
<211> 19

<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 174
aatttctact tttgtagac 19

<210> 175
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 175
taatttctac tgttgtagat 20

<210> 176
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 176
cctcactcct gctcggtgaa ttt 23

<210> 177
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 177
taatttctac tgttgtagat 20

<210> 178
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 178
aggagtgttc agtctccgtg aac 23

<210> 179
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 179
taatttctac tgttgtagat 20

<210> 180
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 180
ctgatggtcc atgtctgtta ctc 23

<210> 181
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 181
taatttctac tgttgtagat 20

<210> 182
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 182
tttcccttca gctaaaataa agg 23

<210> 183
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 183
taatttctac taagtgtaga t 21

<210> 184
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 184
cctcactcct gctcggtgaa ttt 23

<210> 185
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 185
taatttctac taagtgtaga t 21

<210> 186
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 186
aggagtgttc agtctccgtg aac 23

<210> 187

<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 187
taatttctac taagtgtaga t 21

<210> 188
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 188
ctgatggtcc atgtctgtta ctc 23

<210> 189
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 189
taatttctac taagtgtaga t 21

<210> 190
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 190
tttcccttca gctaaaataa agg 23

<210> 191
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 191
tcactcctgc tcggtgaatt 20

<210> 192
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 192
aaccctctgg ggaccgtttg 20

<210> 193
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的

寡核苷酸

<400> 193
agtacgttaa tgtttcctga 20

<210> 194

<400> 194
000

<210> 195
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 195
taatttctac tgttgtagat 20

<210> 196
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 196
agaaatgcat ggttctcatg c 21

<210> 197
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 197
aaaattacct agtaattagg t 21

<210> 198
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 198
ggatttctac ttttgtagat 20

<210> 199
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 199
aaatttctac ttttgtagat 20

<210> 200
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的

寡核苷酸

<400> 200
cgcgcccacg cggggcgcga c 21

<210> 201
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 201
taatttctac tcttgtagat 20

<210> 202
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 202
gaatttctac tattgtagat 20

<210> 203
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 203
gaatctctac tctttgtaga t 21

<210> 204
<211> 19
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 204
taatttctac tttgtagat 19

<210> 205
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 205
aaatttctac tgtttgtaga t 21

<210> 206
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 206
gaatttctac ttttgtagat 20

<210> 207
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 207
taatttctac taagtgtaga t 21

<210> 208
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 208
taatttctac tattgtagat 20

<210> 209
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 209
taatttctac ttcggtagat 20

<210> 210
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 210
taatttctac tattgtagat 20

<210> 211
<211> 6569
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的多核苷酸

<400> 211
catcaaggaa ttggttctaa gcttatagaa gcaatgatta aggaagccaa aaaaaataat 60

attgatgcaa tatttgtctt aggtcatcca agttattatc caaaatttgg ttttaaacca 120

gccacagaat atcagataaa atgtgaatat gatgtcccag cggatgtttt tatggtacta 180

gatttgtcag ctaaactagc tagtttaaaa ggacaaactg tctactatgc cgatgagttt 240

ggcaaaattt tttagatcta caaaattata aactaaataa agattcttat aataacttta 300

tatataatcg aaatgtagag aattttataa ggagtcttta tcatgtcaat ttatcaagaa 360

tttgttaata aatatagttt aagtaaaact ctaagatttg agttaatccc acagggtaaa 420

acacttgaaa acataaaagc aagaggtttg attttagatg atgagaaaag agctaaagac 480

tacaaaaagg ctaaacaaat aattgataaa tatcatcagt tttttataga ggagatatta 540

agttcggttt gtattagcga agatttatta caaaactatt ctgatgttta ttttaaactt 600

aaaaagagtg atgatgataa tctacaaaaa gattttaaaa gtgcaaaaga tacgataaag 660

aaacaaatat ctgaatatat aaaggactca gagaaattta agaatttgtt taatcaaaac 720

cttatcgatg ctaaaaaagg gcaagagtca gatttaattc tatggctaaa gcaatctaag 780

gataatggta tagaactatt taaagccaat agtgatatca cagatataga tgaggcgtta 840

gaaataatca aatcttttaa aggttggaca acttatttta agggttttca tgaaaatага 900

aaaaatgttt atagtagcaa tgatattcct acatctatta tttataggat agtagatgat 960

aatttgccta aatttctaga aaataaagct aagtatgaga gtttaaaaga caaagctcca 1020

gaagctataa actatgaaca aattaaaaaa gatttggcag aagagctaac ctttgatatt 1080

gactacaaaa catctgaagt taatcaaaga gttttttcac ttgatgaagt ttttgagata 1140

gcaaacttta ataattatct aaatcaaagt ggtattacta aatttaatac tattattggt 1200

ggtaaatttg taaatggtga aaatacaaag agaaaaggta taaatgaata tataaatcta 1260

tactcacagc aaataaatga taaaacactc aaaaaatata aaatgagtgt tttatttaag 1320

caaattttaa gtgatacaga atctaaatct tttgtaattg ataagttaga agatgatagt 1380

gatgtagtta caacgatgca aagtttttat gagcaaatag cagcttttaa aacagtagaa 1440

gaaaaatcta ttaaagaaac actatcttta ttatttgatg atttaaaagc tcaaaaactt 1500

gatttgagta aaatttattt taaaaatgat aaatctctta ctgatctatc acaacaagtt 1560

tttgatgatt atagtgttat tggtacagcg gtactagaat atataactca acaaatagca 1620

cctaaaaatc ttgataaccc tagtaagaaa gagcaagaat taatagccaa aaaaactgaa 1680

aaagcaaaat acttatctct agaaactata aagcttgcct tagaagaatt taataagcat 1740

agagatatag ataaacagtg taggtttgaa gaaatacttg caaactttgc ggctattccg 1800

atgatatttg atgaaatagc tcaaaacaaa gacaatttgg cacagatatc tatcaaatat 1860

caaaatcaag gtaaaaaaga cctacttcaa gctagtgcgg aagatgatgt taaagctatc 1920

aaggatcttt tagatcaaac taataatctc ttacataaac taaaaatatt tcatattagt 1980

cagtcagaag ataaggcaaa tattttagac aaggatgagc atttttatct agtatttgag 2040

gagtgctact ttgagctagc gaatatagtg cctctttata acaaaattag aaactatata 2100

actcaaaagc catatagtga tgagaaattt aagctcaatt ttgagaactc gactttggct 2160

aatggttggg ataaaaataa agagcctgac aatacggcaa ttttatttat caaagatgat 2220

aaatattatc tgggtgtgat gaataagaaa aataacaaaa tatttgatga taaagctatc 2280

aaagaaaata aaggcgaggg ttataaaaaa attgtttata aacttttacc tggcgcaaat 2340

aaaatgttac ctaaggtttt cttttctgct aaatctataa aattttataa tcctagtgaa 2400

gatatactta gaataagaaa tcattccaca catacaaaaa atggtagtcc tcaaaaagga 2460

tatgaaaaat ttgagtttaa tattgaagat tgccgaaaat ttatagattt ttataaacag 2520

tctataagta agcatccgga gtggaaagat tttggattta gattttctga tactcaaaga 2580

tataattcta tagatgaatt ttatagagaa gttgaaaatc aaggctacaa actaactttt 2640

gaaaatatat cagagagcta tattgatagc gtagttaatc agggtaaatt gtacctattc 2700

caaatctata ataaagattt ttcagcttat agcaaagggc gaccaaatct acatacttta 2760

tattggaaag cgctgtttga tgagagaaat cttcaagatg tggtttataa gctaaatggt 2820

gaggcagagc ttttttatcg taaacaatca atacctaaaa aaatcactca cccagctaaa 2880

gaggcaatag ctaataaaaa caaagataat cctaaaaaag agagtgtttt tgaatatgat 2940

ttaatcaaag ataaacgctt tactgaagat aagtttttct ttcactgtcc tattacaatc 3000

aattttaaat ctagtggagc taataagttt aatgatgaaa tcaatttatt gctaaaagaa 3060

aaagcaaatg atgttcatat attaagtata gatagaggtg aaagacattt agcttactat 3120

actttggtag atggtaaagg caatatcatc aaacaagata ctttcaacat cattggtaat 3180

gatagaatga aaacaaacta ccatgataag cttgctgcaa tagagaaaga tagggattca 3240

gctaggaaag actggaaaaa gataaataac atcaaagaga tgaaagaggg ctatctatct 3300

caggtagttc atgaaatagc taagctagtt atagagtata atgctattgt ggtttttgag 3360

gatttaaatt ttggatttaa aagagggcgt ttcaaggtag agaagcaggt ctatcaaaag 3420

ttagaaaaaa tgctaattga gaaactaaac tatctagttt tcaaagataa tgagtttgat 3480

aaaactgggg gagtgcttag agcttatcag ctaacagcac cttttgagac ttttaaaaag 3540

atgggtaaac aaacaggtat tatctactat gtaccagctg gttttacttc aaaaatttgt 3600

cctgtaactg gttttgtaaa tcagttatat cctaagtatg aaagtgtcag caaatctcaa 3660

gagttcttta gtaagtttga caagatttgt tataaccttg ataagggcta ttttgagttt 3720

agttttgatt ataaaaactt tggtgacaag gctgccaaag gcaagtggac tatagctagc 3780

tttgggagta gattgattaa ctttagaaat tcagataaaa atcataattg ggatactcga 3840

gaagtttatc caactaaaga gttggagaaa ttgctaaaag attattctat cgaatatggg 3900

catggcgaat gtatcaaagc agctatttgc ggtgagagcg acaaaaagtt ttttgctaag 3960

ctaactagtg tcctaaatac tatcttacaa atgcgtaact caaaaacagg tactgagtta 4020

gattatctaa tttcaccagt agcagatgta aatggcaatt tctttgattc gcgacaggcg 4080

ccaaaaaata tgcctcaaga tgctgatgcc aatggtgctt atcatattgg gctaaaaggt 4140

ctgatgctac taggtaggat caaaaataat caagagggca aaaaactcaa tttggttatc 4200

aaaaatgaag agtattttga gttcgtgcag aataggaata actaattcat tcaagaatat 4260

attaccctgt cagtttagcg actattacct ctttaataat ttgcagggga attattttag 4320

taatagtaat atacacaaga gttattgatt atatggaaaa ttatatttag ataacatggt 4380

taaatgattt tatattctgt ccttactcga tatatttgca taatatctat agtaatgcct 4440

cagatactac atactattca tctagccaaa caaaagggcg cgatgctcat aaaagtatcg 4500

ataaaggaat ctatagtacc aaaaaagatg acctgatcgg tatcgatgtt attaaccata 4560

aatatggttt ggttggtaaa attgatgttt ttcataaaga taagggctta cttgtggaga 4620

gaaaaaggca aatcaagact atctatgatg gctataaata tcagctttat gcgcaatatt 4680

tttgtctcca agagatgggc tatgatgtca aagccattaa attttattcg atggttgata 4740

ataaatcata cccaatagct ataccaactt cagctgagtt agaaaagttt gaaaaacata 4800

ttcaaacaat caagcaatat aatccaatgg ataactcatt taggcaaaat attgaaaagt 4860

gtaaattttg tatatatgca aacttatgtg ataaaacgga cttgtagatt atgtttagta 4920

aaaatgatat tgaatcaaag aatatagttt ttgttaatat ttttgatgga gtgaaactta 4980

gtctatcatt ggggaatata gttataaaag ataaagaaac tgatgaggtg aaaactaagc 5040

tttctgttca taaagttctt gcattgttta tcgtaggtaa tatgacgatg acctcgcaac 5100

ttttagagac ctgtaagaaa aatgctatac agctagtttt tatgaaaaat agctttagac 5160

catatctatg ttttggtgat attgctgagg ctaatttttt agctagatat aagcaatata 5220

gtgtagttga gcaagatata agtttagcaa ggatttttat aacatcaaag atacgcaatc 5280

aacataactt agtcaaaagc ctaagagata aaactccaga gcagcaagag atagtcaaaa 5340

agaataaaca gctaatagca gagttagaaa atacaacaag cctagcggag ctaatgggta 5400

tagagggcaa tgttgccaaa aatttcttca aaggattcta tggacattta gatagttggc 5460

aagggcgcaa acctagaata aaacaggatc catataatgt tgttttagac ttgggctata 5520

gtatgttgtt taattttgta gagtgttttt tgcgacttttt tggctttgat ttatacaagg 5580

gcttttgtca tcagacttgg tataagcgta aatccctagt ttgtgacttt gttgagccat 5640

ttagatgtat agtggataac caagttagaa aatcatggaa tctcgggcaa ttttctgtag 5700

aggattttgg ttgcaaaaat gagcagtttt atataaaaaa agataaaaca aaagactact 5760

caaaaatact ttttgccgag attatcagct acaagctaga gatatttgaa tatgtaagag 5820

aattttatcg tgcctttatg cgaggcaaag aaattgcaga gtatccaata ttttgttatg 5880

aaactaggag ggtgtatgtt gatagtcagt tatgatttta gtaataataa agtacgtgca 5940

aagtttgcca aatttctaga aagttatggt gtacgtttac aatattcggt atttgagctc 6000

aaatatagca agagaatgtt agacttgatt ttagctgaga tagaaaataa ctatgtacca 6060

ctatttacaa atgctgatag tgttttaatc tttaatgctc cagataaaga tgtgataaaa 6120

tatggttatg cgattcatag agaacaagag gttgttttta tagactaaaa attgcaaacc 6180

ttagtcttta tgttaaaata actactaagt tcttagagat atttaaaaat atgactgttg 6240

ttatatatca aaatgctaaa aaaatcatag attttaggtc tttttttgct gatttaggca 6300

aaaacgggtc taagaacttt aaataatttc tactgttgta gatgagaagt catttaataa 6360

ggccactgtt aaaagtctaa gaactttaaa taatttctac tgttgtagat gctactattc 6420

ctgtgccttc agataattca gtctaagaac tttaaataat ttctactgtt gtagatgtct 6480

agagcctttt gtattagtag ccggtctaag aactttaaat aatttctact gttgtagatt 6540

agcgatttat gaaggtcatt tttttgtct 6569

<210> 212
<211> 4170
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 212
tttacacttt atgcttccgg ctcgtatgtt aggaggtctt tatcatgtca atttatcaag 60

aatttgttaa taaatatagt ttaagtaaaa ctctaagatt tgagttaatc ccacagggta 120

aaacacttga aaacataaaa gcaagaggtt tgattttaga tgatgagaaa agagctaaag 180

actacaaaaa ggctaaacaa ataattgata aatatcatca gttttttata gaggagatat 240

taagttcggt ttgtattagc gaagatttat tacaaaacta ttctgatgtt tattttaaac 300

ttaaaaagag tgatgatgat aatctacaaa aagattttaa aagtgcaaaa gatacgataa 360

agaaacaaat atctgaatat ataaaggact cagagaaatt taagaatttg tttaatcaaa 420

accttatcga tgctaaaaaa gggcaagagt cagatttaat tctatggcta aagcaatcta 480

aggataatgg tatagaacta tttaaagcca atagtgatat cacagatata gatgaggcgt 540

tagaaataat caaatctttt aaaggttgga caacttattt taagggtttt catgaaaata 600

gaaaaaatgt ttatagtagc aatgatattc ctacatctat tatttatagg atagtagatg 660

ataatttgcc taaatttcta gaaaataaag ctaagtatga gagtttaaaa gacaaagctc 720

cagaagctat aaactatgaa caaattaaaa aagatttggc agaagagcta acctttgata 780

ttgactacaa aacatctgaa gttaatcaaa gagtttttc acttgatgaa gtttttgaga 840

tagcaaactt taataattat ctaaatcaaa gtggtattac taaatttaat actattattg 900

gtggtaaatt tgtaaatggt gaaaatacaa agagaaaagg tataaatgaa tatataaatc 960

tatactcaca gcaaataaat gataaaacac tcaaaaaata taaaatgagt gttttattta 1020

agcaaatttt aagtgataca gaatctaaat cttttgtaat tgataagtta gaagatgata 1080

gtgatgtagt tacaacgatg caaagttttt atgagcaaat agcagctttt aaaacagtag 1140

aagaaaaatc tattaaagaa acactatctt tattatttga tgatttaaaa gctcaaaaac 1200

ttgatttgag taaaatttat tttaaaaatg ataaatctct tactgatcta tcacaacaag 1260

tttttgatga ttatagtgtt attggtacag cggtactaga atatataact caacaaatag 1320

cacctaaaaa tcttgataac cctagtaaga aagagcaaga attaatagcc aaaaaaactg 1380

aaaaagcaaa atacttatct ctagaaacta taaagcttgc cttagaagaa tttaataagc 1440

atagagatat agataaacag tgtaggtttg aagaaatact tgcaaacttt gcggctattc 1500

cgatgatatt tgatgaaata gctcaaaaca aagacaattt ggcacagata tctatcaaat 1560

atcaaaatca aggtaaaaaa gacctacttc aagctagtgc ggaagatgat gttaaagcta 1620

tcaaggatct tttagatcaa actaataatc tcttacataa actaaaaata tttcatatta 1680

gtcagtcaga agataaggca aatattttag acaaggatga gcatttttat ctagtatttg 1740

aggagtgcta ctttgagcta gcgaatatag tgcctcttta taacaaaatt agaaactata 1800

taactcaaaa gccatatagt gatgagaaat ttaagctcaa ttttgagaac tcgactttgg 1860

ctaatggttg ggataaaaat aaagagcctg acaatacggc aattttattt atcaaagatg 1920

ataaatatta tctgggtgtg atgaataaga aaaataacaa aatatttgat gataaagcta 1980

tcaaagaaaa taaaggcgag ggttataaaa aaattgttta taaactttta cctggcgcaa 2040

ataaaatgtt acctaaggtt ttcttttctg ctaaatctat aaaattttat aatcctagtg 2100

aagatatact tagaataaga aatcattcca cacatacaaa aaatggtagt cctcaaaaag 2160

gatatgaaaa atttgagttt aatattgaag attgccgaaa atttatagat ttttataaac 2220

agtctataag taagcatccg gagtggaaag attttggatt tagattttct gatactcaaa 2280

gatataattc tatagatgaa ttttatagag aagttgaaaa tcaaggctac aaactaactt 2340

ttgaaaatat atcagagagc tatattgata gcgtagttaa tcagggtaaa ttgtacctat 2400

tccaaatcta taataaagat ttttcagctt atagcaaagg gcgaccaaat ctacatactt 2460

tatattggaa agcgctgttt gatgagagaa atcttcaaga tgtggtttat aagctaaatg 2520

gtgaggcaga gctttttttat cgtaaacaat caatacctaa aaaaatcact cacccagcta 2580

aagaggcaat agctaataaa aacaaagata atcctaaaaa agagagtgtt tttgaatatg 2640

atttaatcaa agataaacgc tttactgaag ataagttttt ctttcactgt cctattacaa 2700

tcaattttaa atctagtgga gctaataagt ttaatgatga aatcaattta ttgctaaaag 2760

aaaaagcaaa tgatgttcat atattaagta tagatagagg tgaaagacat ttagcttact 2820

atactttggt agatggtaaa ggcaatatca tcaaacaaga tactttcaac atcattggta 2880

atgatagaat gaaaacaaac taccatgata agcttgctgc aatagagaaa gatagggatt 2940

cagctaggaa agactggaaa aagataaata acatcaaaga gatgaaagag ggctatctat 3000

ctcaggtagt tcatgaaata gctaagctag ttatagagta taatgctatt gtggtttttg 3060

aggatttaaa ttttggattt aaaagagggc gtttcaaggt agagaagcag gtctatcaaa 3120

agttagaaaa aatgctaatt gagaaactaa actatctagt tttcaaagat aatgagtttg 3180

ataaaactgg gggagtgctt agagcttatc agctaacagc acctttttgag acttttaaaa 3240

agatgggtaa acaaacaggt attatctact atgtaccagc tggttttact tcaaaaattt 3300

gtcctgtaac tggttttgta aatcagttat atcctaagta tgaaagtgtc agcaaatctc 3360

aagagttctt tagtaagttt gacaagattt gttataacct tgataagggc tattttgagt 3420

ttagttttga ttataaaaac tttggtgaca aggctgccaa aggcaagtgg actatagcta 3480

gctttgggag tagattgatt aactttagaa attcagataa aaatcataat tgggatactc 3540

gagaagttta tccaactaaa gagttggaga aattgctaaa agattattct atcgaatatg 3600

ggcatggcga atgtatcaaa gcagctattt gcggtgagag cgacaaaaag ttttttgcta 3660

agctaactag tgtcctaaat actatcttac aaatgcgtaa ctcaaaaaca ggtactgagt 3720

tagattatct aatttcacca gtagcagatg taaatggcaa tttctttgat tcgcgacagg 3780

cgccaaaaaa tatgcctcaa gatgctgatg ccaatggtgc ttatcatatt gggctaaaag 3840

gtctgatgct actaggtagg atcaaaaata atcaagaggg caaaaaactc aatttggtta 3900

tcaaaaatga agagtatttt gagttcgtgc agaataggaa taactaattg acagctagct 3960

cagtcctagg tataatgcta gcgctgattt aggcaaaaac gggtctaaga actttaaata 4020

atttctactg ttgtagatga gaagtcattt aataaggcca ctgttaaaag tctaagaact 4080

ttaaataatt tctactgttg tagatgctac tattcctgtg ccttcagata attcagtcta 4140

agaactttaa ataatttcta ctgttgtaga 4170

<210> 213
<211> 4613
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 213
ctgtctacta tgccgatgag tttggcaaaa ttttttagat ctacaaaatt ataaactaaa 60

taaagattct tataataact ttatatataa tcgaaatgta gagaatttta taaggagtct 120

ttatcatgtc aatttatcaa gaatttgtta ataaatatag tttaagtaaa actctaagat 180

ttgagttaat cccacagggt aaaacacttg aaaacataaa agcaagaggt ttgattttag 240

atgatgagaa aagagctaaa gactacaaaa aggctaaaca aataattgat aaatatcatc 300

agtttttttat agaggagata ttaagttcgg tttgtattag cgaagattta ttacaaaact 360

attctgatgt ttattttaaa cttaaaaaga gtgatgatga taatctacaa aaagatttta 420

aaagtgcaaa agatacgata aagaaacaaa tatctgaata tataaaggac tcagagaaat 480

ttaagaattt gtttaatcaa aaccttatcg atgctaaaaa agggcaagag tcagatttaa 540

ttctatggct aaagcaatct aaggataatg gtatagaact atttaaagcc aatagtgata 600

tcacagatat agatgaggcg ttagaaataa tcaaatcttt taaaggttgg acaacttatt 660

ttaagggttt tcatgaaaat agaaaaaatg tttatagtag caatgatatt cctacatcta 720

ttatttatag gatagtagat gataatttgc ctaaatttct agaaaataaa gctaagtatg 780

agagtttaaa agacaaagct ccagaagcta taaactatga acaaattaaa aaagatttgg 840

cagaagagct aacctttgat attgactaca aaacatctga agttaatcaa agagtttttt 900

cacttgatga agtttttgag atagcaaact ttaataatta tctaaatcaa agtggtatta 960

ctaaatttaa tactattatt ggtggtaaat ttgtaaatgg tgaaaataca aagagaaaag 1020

gtataaatga atatataaat ctatactcac agcaaataaa tgataaaaca ctcaaaaaat 1080

ataaaatgag tgttttattt aagcaaattt taagtgatac agaatctaaa tcttttgtaa 1140

ttgataagtt agaagatgat agtgatgtag ttacaacgat gcaaagtttt tatgagcaaa 1200

tagcagcttt taaaacagta gaagaaaaat ctattaaaga aacactatct ttattatttg 1260

atgatttaaa agctcaaaaa cttgatttga gtaaaattta ttttaaaaat gataaatctc 1320

ttactgatct atcacaacaa gtttttgatg attatagtgt tattggtaca gcggtactag 1380

aatatataac tcaacaaata gcacctaaaa atcttgataa ccctagtaag aaagagcaag 1440

aattaatagc caaaaaaact gaaaaagcaa aatacttatc tctagaaact ataaagcttg 1500

ccttagaaga atttaataag catagagata tagataaaca gtgtaggttt gaagaaatac 1560

ttgcaaactt tgcggctatt ccgatgatat ttgatgaaat agctcaaaac aaagacaatt 1620

tggcacagat atctatcaaa tatcaaaatc aaggtaaaaa agacctactt caagctagtg 1680

cggaagatga tgttaaagct atcaaggatc ttttagatca aactaataat ctcttacata 1740

aactaaaaat atttcatatt agtcagtcag aagataaggc aaatatttta gacaaggatg 1800

agcattttta tctagtattt gaggagtgct actttgagct agcgaatata gtgcctcttt 1860
ataacaaaat tagaaactat ataactcaaa agccatatag tgatgagaaa tttaagctca 1920
attttgagaa ctcgactttg gctaatggtt gggataaaaa taaagagcct gacaatacgg 1980
caattttatt tatcaaagat gataaatatt atctgggtgt gatgaataag aaaaataaca 2040
aaatatttga tgataaagct atcaaagaaa ataaaggcga gggttataaa aaaattgttt 2100
ataaactttt acctggcgca aataaaatgt tacctaaggt tttcttttct gctaaatcta 2160
taaaatttta taatcctagt gaagatatac ttagaataag aaatcattcc acacatacaa 2220
aaaatggtag tcctcaaaaa ggatatgaaa aatttgagtt taatattgaa gattgccgaa 2280
aatttataga ttttttataaa cagtctataa gtaagcatcc ggagtggaaa gattttggat 2340
ttagattttc tgatactcaa agatataatt ctatagatga attttataga gaagttgaaa 2400
atcaaggcta caaactaact tttgaaaata tatcagagag ctatattgat agcgtagtta 2460
atcagggtaa attgtaccta ttccaaatct ataataaaga tttttcagct tatagcaaag 2520
ggcgaccaaa tctacatact ttatattgga aagcgctgtt tgatgagaga aatcttcaag 2580
atgtggttta taagctaaat ggtgaggcag agctttttta tcgtaaacaa tcaataccta 2640
aaaaaatcac tcacccagct aaagaggcaa tagctaataa aaacaaagat aatcctaaaa 2700
aagagagtgt ttttgaatat gatttaatca aagataaacg ctttactgaa gataagtttt 2760
tctttcactg tcctattaca atcaatttta aatctagtgg agctaataag tttaatgatg 2820
aaatcaattt attgctaaaa gaaaaagcaa atgatgttca tatattaagt atagatagag 2880
gtgaaagaca tttagcttac tatactttgg tagatggtaa aggcaatatc atcaaacaag 2940
atactttcaa catcattggt aatgatagaa tgaaaacaaa ctaccatgat aagcttgctg 3000
caatagagaa agatagggat tcagctagga aagactggaa aaagataaat aacatcaaag 3060
agatgaaaga gggctatcta tctcaggtag ttcatgaaat agctaagcta gttatagagt 3120
ataatgctat tgtggttttt gaggatttaa attttggatt taaaagaggg cgtttcaagg 3180
tagagaagca ggtctatcaa aagttagaaa aaatgctaat tgagaaacta aactatctag 3240
ttttcaaaga taatgagttt gataaaactg ggggagtgct tagagcttat cagctaacag 3300
caccttttga gacttttaaa aagatgggta aacaaacagg tattatctac tatgtaccag 3360
ctggttttac ttcaaaaatt tgtcctgtaa ctggttttgt aaatcagtta tatcctaagt 3420
atgaaagtgt cagcaaatct caagagttct ttagtaagtt tgacaagatt tgttataacc 3480
ttgataaggg ctattttgag tttagttttg attataaaaa ctttggtgac aaggctgcca 3540
aaggcaagtg gactatagct agctttggga gtagattgat taactttaga aattcagata 3600
aaaatcataa ttgggatact cgagaagttt atccaactaa agagttggag aaattgctaa 3660
aagattattc tatcgaatat gggcatggcg aatgtatcaa agcagctatt tgcggtgaga 3720
gcgacaaaaa gttttttgct aagctaacta gtgtcctaaa tactatctta caaatgcgta 3780
actcaaaaac aggtactgag ttagattatc taatttcacc agtagcagat gtaaatggca 3840
atttctttga ttcgcgacag gcgccaaaaa atatgcctca agatgctgat gccaatggtg 3900
cttatcatat tgggctaaaa ggtctgatgc tactaggtag gatcaaaaat aatcaagagg 3960
gcaaaaaact caatttggtt atcaaaaatg aagagtattt tgagttcgtg cagaatagga 4020
ataactaatt cattcaagaa tatattaccc tgtcagttta gcgactatta cctctttaat 4080
aatttgcagg ggaattattt tagtaatagt aatatacaca agagttattg attatatgga 4140
aaattatatt tagataacat ggttaaatga ttttatattc tgtccttact cgatatattt 4200
tttatagact aaaaattgca aaccttagtc tttatgttaa aataactact aagttcttag 4260
agatatttaa aaatatgact gttgttatat atcaaaatgc taaaaaaatc atagatttta 4320

ggtctttttt tgctgattta ggcaaaaacg ggtctaagaa ctttaaataa tttctactgt 4380

tgtagatgag aagtcattta ataaggccac tgttaaaagt ctaagaactt taaataattt 4440

ctactgttgt agatgctact attcctgtgc cttcagataa ttcagtctaa gaactttaaa 4500

taatttctac tgttgtagat gtctagagcc ttttgtatta gtagccggtc taagaacttt 4560

aaataatttc tactgttgta gattagcgat ttatgaaggt catttttttg tct 4613

<210> 214
<211> 4035
<212> DNA
<213> 土拉熱弗朗西絲菌

<400> 214
atgagcatct accaggagtt cgtcaacaag tattcactga gtaagacact gcggttcgag 60

ctgatcccac agggcaagac actggagaac atcaaggccc gaggcctgat tctggacgat 120

gagaagcggg caaaagacta taagaaagcc aagcagatca ttgataaata ccaccagttc 180

tttatcgagg aaattctgag ctccgtgtgc atcagtgagg atctgctgca gaattactca 240

gacgtgtact tcaagctgaa gaagagcgac gatgacaacc tgcagaagga cttcaagtcc 300

gccaaggaca ccatcaagaa acagattagc gagtacatca aggactccga aaagtttaaa 360

aatctgttca accagaatct gatcgatgct aagaaaggcc aggagtccga cctgatcctg 420

tggctgaaac agtctaagga caatgggatt gaactgttca aggctaactc cgatatcact 480

gatattgacg aggcactgga aatcatcaag agcttcaagg gatggaccac atactttaaa 540

ggcttccacg agaaccgcaa gaacgtgtac tccagcaacg acattcctac ctccatcatc 600

taccgaatcg tcgatgacaa tctgccaaag ttcctggaga acaaggccaa atatgaatct 660

ctgaaggaca aagctcccga ggcaattaat tacgaacaga tcaagaaaga tctggctgag 720

gaactgacat tcgatatcga ctataagact agcgaggtga accagagggt cttttccctg 780

gacgaggtgt ttgaaatcgc caatttcaac aattacctga accagtccgg cattactaaa 840

ttcaatacca tcattggcgg gaagtttgtg aacggggaga ataccaagcg caagggaatt 900

aacgaataca tcaatctgta tagccagcag atcaacgaca aaactctgaa gaaatacaag 960

atgtctgtgc tgttcaaaca gatcctgagt gataccgagt ccaagtcttt tgtcattgat 1020

aaactggaag atgactcaga cgtggtcact accatgcaga gcttttatga gcagatcgcc 1080

gctttcaaga cagtggagga aaaatctatt aaggaaactc tgagtctgct gttcgatgac 1140

ctgaaagccc agaagctgga cctgagtaag atctacttca aaaacgataa gagtctgaca 1200

gacctgtcac agcaggtgtt tgatgactat tccgtgattg ggaccgccgt cctggagtac 1260

attacacagc agatcgctcc aaagaacctg gataatccct ctaagaaaga gcaggaactg 1320

atcgctaaga aaaccgagaa ggcaaaatat ctgagtctgg aaacaattaa gctggcactg 1380

gaggagttca acaagcacag ggatattgac aaacagtgcc gctttgagga aatcctggcc 1440

aacttcgcag ccatccccat gattttgat gagatcgccc agaacaaaga caatctggct 1500

cagatcagta ttaagtacca gaaccagggc aagaaagacc tgctgcaggc ttcagcagaa 1560

gatgacgtga aagccatcaa ggatctgctg gaccagacca acaatctgct gcacaagctg 1620

aaaatcttcc atattagtca gtcagaggat aaggctaata tcctggataa agacgaacac 1680

ttctacctgg tgttcgagga atgttacttc gagctggcaa acattgtccc cctgtataac 1740

aagattagga actacatcac acagaagcct tactctgacg agaagtttaa actgaacttc 1800

gaaaatagta ccctggccaa cgggtgggat aagaacaagg agcctgacaa cacagctatc 1860

ctgttcatca aggatgacaa gtactatctg ggagtgatga ataagaaaaa caataagatc 1920

ttcgatgaca aagccattaa ggagaacaaa ggggaaggat acaagaaaat cgtgtataag 1980

ctgctgcccg gcgcaaataa gatgctgcct aaggtgttct tcagcgccaa gagtatcaaa 2040

ttctacaacc catccgagga catcctgcgg attagaaatc actcaacaca tactaagaac 2100

gggagccccc agaagggata tgagaaattt gagttcaaca tcgaggattg caggaagttt 2160

attgacttct acaagcagag catctccaaa caccctgaat ggaaggattt tggcttccgg 2220

ttttccgaca cacagagata taactctatc gacgagttct accgcgaggt ggaaaatcag 2280

gggtataagc tgacttttga gaacatttct gaaagttaca tcgacagcgt ggtcaatcag 2340

ggaaagctgt acctgttcca gatctataac aaagattttt cagcatacag caagggcaga 2400

ccaaacctgc atacactgta ctggaaggcc ctgttcgatg agaggaatct gcaggacgtg 2460

gtctataaac tgaacggaga ggccgaactg ttttaccgga agcagtctat tcctaagaaa 2520

atcactcacc cagctaagga ggccatcgct aacaagaaca aggacaatcc taagaaagag 2580

agcgtgttcg aatacgatct gattaaggac aagcggttca ccgaagataa gttctttttc 2640

cattgtccaa tcaccattaa cttcaagtca agcggcgcta acaagttcaa cgacgagatc 2700

aatctgctgc tgaaggaaaa agcaaacgat gtgcacatcc tgagcattga ccgaggagag 2760

cggcatctgg cctactatac cctggtggat ggcaaaggga atatcattaa gcaggataca 2820

ttcaacatca ttggcaatga ccggatgaaa accaactacc acgataaact ggctgcaatc 2880

gagaaggata gagactcagc taggaaggac tggaagaaaa tcaacaacat taaggagatg 2940

aaggaaggct atctgagcca ggtggtccat gagattgcaa agctggtcat cgaatacaat 3000

gccattgtgg tgttcgagga tctgaacttc ggctttaaga gggggcgctt taaggtggaa 3060

aaacaggtct atcagaagct ggagaaaatg ctgatcgaaa agctgaatta cctggtgttt 3120

aaagataacg agttcgacaa gaccggaggc gtcctgagag cctaccagct gacagctccc 3180

tttgaaactt tcaagaaaat gggaaaacag acaggcatca tctactatgt gccagccgga 3240

ttcacttcca agatctgccc cgtgaccggc tttgtcaacc agctgtaccc taaatatgag 3300

tcagtgagca agtcccagga atttttcagc aagttcgata agatctgtta taatctggac 3360

aaggggtact tcgagttttc cttcgattac aagaacttcg gcgacaaggc cgctaagggg 3420

aaatggacca ttgcctcctt cggatctcgc ctgatcaact ttcgaaattc cgataaaaac 3480

cacaattggg acactaggga ggtgtaccca accaaggagc tggaaaagct gctgaaagac 3540

tactctatcg agtatggaca tggcgaatgc atcaaggcag ccatctgtgg cgagagtgat 3600

aagaaatttt tcgccaagct gacctcagtg ctgaatacaa tcctgcagat gcggaactca 3660

aagaccggga cagaactgga ctatctgatt agccccgtgg ctgatgtcaa cggaaacttc 3720

ttcgacagca gacaggcacc caaaaatatg cctcaggatg cagacgccaa cggggcctac 3780

cacatcgggc tgaagggact gatgctgctg ggccggatca agaacaatca ggaggggaag 3840

aagctgaacc tggtcattaa gaacgaggaa tacttcgagt ttgtccagaa tagaaataac 3900

aaaaggccgg cggccacgaa aaaggccggc caggcaaaaa agaaaaaggg atcctaccca 3960

tacgatgttc cagattacgc ttatccctac gacgtgcctg attatgcata cccatatgat 4020

gtccccgact atgcc 4035

<210> 215
<211> 3834
<212> DNA
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 215
atggattacg gcaacggcca gtttgagcgg agagcccccc tgaccaagac aatcaccctg 60

cgcctgaagc ctatcggcga gacacgggag acaatccgcg agcagaagct gctggagcag 120

gacgccgcct tcagaaagct ggtggagaca gtgaccccta tcgtggacga ttgtatcagg 180

aagatcgccg ataacgccct gtgccacttt ggcaccgagt atgacttcag ctgtctgggc 240

aacgccatct ctaagaatga cagcaaggcc atcaagaagg agacagagaa ggtggagaag 300

ctgctggcca aggtgctgac cgagaatctg ccagatggcc tgcgcaaggt gaacgacatc 360

aattccgccg cctttatcca ggatacactg acctctttcg tgcaggacga tgccgacaag 420

cgggtgctga tccaggagct gaagggcaag accgtgctga tgcagcggtt cctgaccaca 480

cggatcacag ccctgaccgt gtggctgccc gacagagtgt tcgagaactt taatatcttc 540

atcgagaacg ccgagaagat gagaatcctg ctggactccc ctctgaatga gaagatcatg 600

aagtttgacc cagatgccga gcagtacgcc tctctggagt tctatggcca gtgcctgtct 660

cagaaggaca tcgatagcta caacctgatc atctccggca tctatgccga cgatgaggtg 720

aagaaccctg gcatcaatga gatcgtgaag gagtacaatc agcagatccg gggcgacaag 780

gatgagtccc cactgcccaa gctgaagaag ctgcacaagc agatcctgat gccagtggag 840

aaggccttct ttgtgcgcgt gctgtctaac gacagcgatg cccggagcat cctggagaag 900

atcctgaagg acacagagat gctgccctcc aagatcatcg aggccatgaa ggaggcagat 960

gcaggcgaca tcgccgtgta cggcagccgg ctgcacgagc tgagccacgt gatctacggc 1020

gatcacggca agctgtccca gatcatctat gacaaggagt ccaagaggat ctctgagctg 1080

atggagacac tgtctccaaa ggagcgcaag gagagcaaga agcggctgga gggcctggag 1140

gagcacatca gaaagtctac atacaccttc gacgagctga acaggtatgc cgagaagaat 1200

gtgatggcag catacatcgc agcagtggag gagtcttgtg ccgagatcat gagaaaggag 1260

aaggatctga ggaccctgct gagcaaggag gacgtgaaga tccggggcaa cagacacaat 1320

acactgatcg tgaagaacta ctttaatgcc tggaccgtgt tccggaacct gatcagaatc 1380

ctgaggcgca agtccgaggc cgagatcgac tctgacttct acgatgtgct ggacgattcc 1440

gtggaggtgc tgtctctgac atacaagggc gagaatctgt gccgcagcta tatcaccaag 1500

aagatcggct ccgacctgaa gcccgagatc gccacatacg gcagcgccct gaggcctaac 1560

agccgctggt ggtccccagg agagaagttt aatgtgaagt tccacaccat cgtgcggaga 1620

gatggccggc tgtactattt catcctgccc aagggcgcca agcctgtgga gctggaggac 1680

atggatggcg acatcgagtg tctgcagatg agaaagatcc ctaacccaac aatctttctg 1740

cccaagctgg tgttcaagga ccctgaggcc ttctttaggg ataatccaga ggccgacgag 1800

ttcgtgtttc tgagcggcat gaaggccccc gtgacaatca ccagagagac atacgaggcc 1860

tacaggtata agctgtatac cgtgggcaag ctgcgcgatg gcgaggtgtc cgaagaggag 1920

tacaagcggg ccctgctgca ggtgctgacc gcctacaagg agtttctgga gaacagaatg 1980

atctatgccg acctgaattt cggctttaag gatctggagg agtataagga cagctccgag 2040

tttatcaagc aggtggagac acacaacacc ttcatgtgct gggccaaggt gtctagctcc 2100

cagctggacg atctggtgaa gtctggcaac ggcctgctgt tcgagatctg gagcgagcgc 2160

ctggagtcct actataagta cggcaatgag aaggtgctgc ggggctatga gggcgtgctg 2220

ctgagcatcc tgaaggatga gaacctggtg tccatgcgga ccctgctgaa cagccggccc 2280

atgctggtgt accggccaaa ggagtctagc aagcctatgg tggtgcaccg ggatggcagc 2340

agagtggtgg acaggtttga taaggacggc aagtacatcc cccctgaggt gcacgacgag 2400

ctgtatcgct tctttaacaa tctgctgatc aaggagaagc tgggcgagaa ggcccggaag 2460

atcctggaca acaagaaggt gaaggtgaag gtgctggaga gcgagagagt gaagtggtcc 2520

aagttctacg atgagcagtt tgccgtgacc ttcagcgtga agaagaacgc cgattgtctg 2580

gacaccacaa aggacctgaa tgccgaagtg atggagcagt atagcgagtc caacagactg 2640

atcctgatca ggaataccac agatatcctg tactatctgg tgctggacaa gaatggcaag 2700

gtgctgaagc agagatccct gaacatcatc aatgacggcg ccagggatgt ggactggaag 2760

gagaggttcc gccaggtgac aaaggataga aacgagggct acaatgagtg ggattattcc 2820

aggacctcta acgacctgaa ggaggtgtac ctgaattatg ccctgaagga gatcgccgag 2880

gccgtgatcg agtacaacgc catcctgatc atcgagaaga tgtctaatgc ctttaaggac 2940

aagtatagct tcctggacga cgtgaccttc aagggcttcg agacaaagct gctggccaag 3000

ctgagcgatc tgcactttag gggcatcaag gacggcgagc catgttcctt cacaaacccc 3060

ctgcagctgt gccagaacga ttctaataag atcctgcagg acggcgtgat ctttatggtg 3120

ccaaattcta tgacacggag cctggacccc gacaccggct tcatctttgc catcaacgac 3180

cacaatatca ggaccaagaa ggccaagctg aactttctga gcaagttcga tcagctgaag 3240

gtgtcctctg agggctgcct gatcatgaag tacagcggcg attccctgcc tacacacaac 3300

accgacaatc gcgtgtggaa ctgctgttgc aatcacccaa tcacaaacta tgaccgggag 3360

acaaagaagg tggagttcat cgaggagccc gtggaggagc tgtcccgcgt gctggaggag 3420

aatggcatcg agacagacac cgagctgaac aagctgaatg agcgggagaa cgtgcctggc 3480

aaggtggtgg atgccatcta ctctctggtg ctgaattatc tgcgcggcac agtgagcgga 3540

gtggcaggac agagggccgt gtactatagc cctgtgaccg gcaagaagta cgatatctcc 3600

tttatccagg ccatgaacct gaataggaag tgtgactact ataggatcgg ctccaaggag 3660

aggggagagt ggaccgattt cgtggcccag ctgatcaaca aaaggccggc ggccacgaaa 3720

aaggccggcc aggcaaaaaa gaaaaaggga tcctacccat acgatgttcc agattacgct 3780

tatccctacg acgtgcctga ttatgcatac ccatatgatg tccccgacta tgcc 3834

<210> 216
<211> 4035
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 216
atgagcatct accaggagtt cgtcaacaag tattcactga gtaagacact gcggttcgag 60

ctgatcccac agggcaagac actggagaac atcaaggccc gaggcctgat tctggacgat 120

gagaagcggg caaaagacta taagaaagcc aagcagatca ttgataaata ccaccagttc 180

tttatcgagg aaattctgag ctccgtgtgc atcagtgagg atctgctgca gaattactca 240

gacgtgtact tcaagctgaa gaagagcgac gatgacaacc tgcagaagga cttcaagtcc 300

gccaaggaca ccatcaagaa acagattagc gagtacatca aggactccga aaagtttaaa 360

aatctgttca accagaatct gatcgatgct aagaaaggcc aggagtccga cctgatcctg 420

tggctgaaac agtctaagga caatgggatt gaactgttca aggctaactc cgatatcact 480

gatattgacg aggcactgga aatcatcaag agcttcaagg gatggaccac atactttaaa 540

ggcttccacg agaaccgcaa gaacgtgtac tccagcaacg acattcctac ctccatcatc 600

taccgaatcg tcgatgacaa tctgccaaag ttcctggaga acaaggccaa atatgaatct 660

ctgaaggaca aagctcccga ggcaattaat tacgaacaga tcaagaaaga tctggctgag 720

gaactgacat tcgatatcga ctataagact agcgaggtga accagagggt cttttccctg 780

gacgaggtgt ttgaaatcgc caatttcaac aattacctga accagtccgg cattactaaa 840

ttcaatacca tcattggcgg gaagtttgtg aacggggaga ataccaagcg caagggaatt 900

aacgaataca tcaatctgta tagccagcag atcaacgaca aaactctgaa gaaatacaag 960

atgtctgtgc tgttcaaaca gatcctgagt gataccgagt ccaagtcttt tgtcattgat 1020

aaactggaag atgactcaga cgtggtcact accatgcaga gcttttatga gcagatcgcc 1080

gctttcaaga cagtggagga aaaatctatt aaggaaactc tgagtctgct gttcgatgac 1140

ctgaaagccc agaagctgga cctgagtaag atctacttca aaaacgataa gagtctgaca 1200

gacctgtcac agcaggtgtt tgatgactat tccgtgattg ggaccgccgt cctggagtac 1260

attacacagc agatcgctcc aaagaacctg gataatccct ctaagaaaga gcaggaactg 1320

atcgctaaga aaaccgagaa ggcaaaatat ctgagtctgg aaacaattaa gctggcactg 1380

gaggagttca acaagcacag ggatattgac aaacagtgcc gctttgagga aatcctggcc 1440

aacttcgcag ccatccccat gatttttgat gagatcgccc agaacaaaga caatctggct 1500

cagatcagta ttaagtacca gaaccagggc aagaaagacc tgctgcaggc ttcagcagaa 1560

gatgacgtga aagccatcaa ggatctgctg gaccagacca acaatctgct gcacaagctg 1620

aaaatcttcc atattagtca gtcagaggat aaggctaata tcctggataa agacgaacac 1680

ttctacctgg tgttcgagga atgttacttc gagctggcaa acattgtccc cctgtataac 1740

aagattagga actacatcac acagaagcct tactctgacg agaagtttaa actgaacttc 1800

gaaaatagta ccctggccaa cgggtgggat aagaacaagg agcctgacaa cacagctatc 1860

ctgttcatca aggatgacaa gtactatctg ggagtgatga ataagaaaaa caataagatc 1920

ttcgatgaca aagccattaa ggagaacaaa ggggaaggat acaagaaaat cgtgtataag 1980

ctgctgcccg gcgcaaataa gatgctgcct aaggtgttct tcagcgccaa gagtatcaaa 2040

ttctacaacc catccgagga catcctgcgg attagaaatc actcaacaca tactaagaac 2100

gggagccccc agaagggata tgagaaattt gagttcaaca tcgaggattg caggaagttt 2160

attgacttct acaagcagag catctccaaa caccctgaat ggaaggattt tggcttccgg 2220

ttttccgaca cacagagata taactctatc gacgagttct accgcgaggt ggaaaatcag 2280

gggtataagc tgacttttga gaacatttct gaaagttaca tcgacagcgt ggtcaatcag 2340

ggaaagctgt acctgttcca gatctataac aaagattttt cagcatacag caagggcaga 2400

ccaaacctgc atacactgta ctggaaggcc ctgttcgatg agaggaatct gcaggacgtg 2460

gtctataaac tgaacggaga ggccgaactg ttttaccgga agcagtctat tcctaagaaa 2520

atcactcacc cagctaagga ggccatcgct aacaagaaca aggacaatcc taagaaagag 2580

agcgtgttcg aatacgatct gattaaggac aagcggttca ccgaagataa gttctttttc 2640

cattgtccaa tcaccattaa cttcaagtca agcggcgcta acaagttcaa cgacgagatc 2700

aatctgctgc tgaaggaaaa agcaaacgat gtgcacatcc tgagcattga ccgaggagag 2760

cggcatctgg cctactatac cctggtggat ggcaaaggga atatcattaa gcaggataca 2820

ttcaacatca ttggcaatga ccggatgaaa accaactacc acgataaact ggctgcaatc 2880

gagaaggata gagactcagc taggaaggac tggaagaaaa tcaacaacat taaggagatg 2940

aaggaaggct atctgagcca ggtggtccat gagattgcaa agctggtcat cgaatacaat 3000

gccattgtgg tgttcgagga tctgaacttc ggctttaaga gggggcgctt taaggtggaa 3060

aaacaggtct atcagaagct ggagaaaatg ctgatcgaaa agctgaatta cctggtgttt 3120

aaagataacg agttcgacaa gaccggaggc gtcctgagag cctaccagct gacagctccc 3180

tttgaaactt tcaagaaaat gggaaaacag acaggcatca tctactatgt gccagccgga 3240

ttcacttcca agatctgccc cgtgaccggc tttgtcaacc agctgtaccc taaatatgag 3300

tcagtgagca agtcccagga atttttcagc aagttcgata agatctgtta taatctggac 3360

aaggggtact tcgagttttc cttcgattac aagaacttcg gcgacaaggc cgctaagggg 3420

aaatggacca ttgcctcctt cggatctcgc ctgatcaact ttcgaaattc cgataaaaac 3480

cacaattggg acactaggga ggtgtaccca accaaggagc tggaaaagct gctgaaagac 3540

tactctatcg agtatggaca tggcgaatgc atcaaggcag ccatctgtgg cgagagtgat 3600

aagaaatttt tcgccaagct gacctcagtg ctgaatacaa tcctgcagat gcggaactca 3660

aagaccggga cagaactgga ctatctgatt agccccgtgg ctgatgtcaa cggaaacttc 3720

ttcgacagca gacaggcacc caaaaatatg cctcaggatg cagacgccaa cggggcctac 3780

cacatcgggc tgaagggact gatgctgctg ggccggatca agaacaatca ggaggggaag 3840

aagctgaacc tggtcattaa gaacgaggaa tacttcgagt ttgtccagaa tagaaataac 3900

aaaaggccgg cggccacgaa aaaggccggc caggcaaaaa agaaaaaggg atcctaccca 3960

tacgatgttc cagattacgc ttatccctac gacgtgcctg attatgcata cccatatgat 4020

gtccccgact atgcc 4035

<210> 217
<211> 4575
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 217
atgtccaact tctttaagaa tttcaccaac ctgtatgagc tgtccaagac actgaggttt 60

gagctgaagc ccgtgggcga caccctgaca aacatgaagg accacctgga gtacgatgag 120

aagctgcaga ccttcctgaa ggatcagaat atcgacgatg cctatcaggc cctgaagcct 180

cagttcgacg agatccacga ggagtttatc acagattctc tggagagcaa gaaggccaag 240

gagatcgact tctccgagta cctggatctg tttcaggaga agaaggagct gaacgactct 300

gagaagaagc tgcgcaacaa gatcggcgag acattcaaca aggccggcga gaagtggaag 360

aaggagaagt accctcagta tgagtggaag aagggctcca agatcgccaa tggcgccgac 420

atcctgtctt gccaggatat gctgcagttt atcaagtata agaacccaga ggatgagaag 480

atcaagaatt acatcgacga tacactgaag ggcttcttta cctatttcgg cggctttaat 540

cagaacaggg ccaactacta tgagacaaag aaggaggcct ccaccgcagt ggcaacaagg 600

atcgtgcacg agaacctgcc aaagttctgt gacaatgtga tccagtttaa gcacatcatc 660

aagcggaaga aggatggcac cgtggagaaa accgagagaa agaccgagta cctgaacgcc 720

taccagtatc tgaagaacaa taacaagatc acacagatca aggacgccga gacagagaag 780

atgatcgagt ctacacccat cgccgagaag atcttcgacg tgtactactt cagcagctgc 840

ctgagccaga agcagatcga ggagtacaac cggatcatcg gccactataa tctgctgatc 900

aacctgtata accaggccaa gagatctgag ggcaagcacc tgagcgccaa cgagaagaag 960

tataaggacc tgcctaagtt caagaccctg tataagcaga tcggctgcgg caagaagaag 1020

gacctgtttt acacaatcaa gtgtgatacc gaggaggagg ccaataagtc ccggaacgag 1080

ggcaaggagt cccactctgt ggaggagatc atcaacaagg cccaggaggc catcaataag 1140

tacttcaagt ctaataacga ctgtgagaat atcaacaccg tgcccgactt catcaactat 1200

atcctgacaa aggagaatta cgagggcgtg tattggagca aggccgccat gaacaccatc 1260

tccgacaagt acttcgccaa ttatcacgac ctgcaggata gactgaagga ggccaaggtg 1320

tttcagaagg ccgataagaa gtccgaggac gatatcaaga tcccagaggc catcgagctg 1380

tctggcctgt tcggcgtgct ggacagcctg gccgattggc agaccacact gtttaagtct 1440

agcatcctga gcaacgagga caagctgaag atcatcacag attcccagac cccctctgag 1500

gccctgctga agatgatctt caatgacatc gagaagaaca tggagtcctt tctgaaggag 1560

acaaacgata tcatcaccct gaagaagtat aagggcaata aggagggcac cgagaagatc 1620

aagcagtggt tcgactatac actggccatc aaccggatgc tgaagtactt tctggtgaag 1680

gagaataaga tcaagggcaa ctccctggat accaatatct ctgaggccct gaaaaccctg 1740

atctacagcg acgatgccga gtggttcaag tggtacgacg ccctgagaaa ctatctgacc 1800

cagaagcctc aggatgaggc caaggagaat aagctgaagc tgaatttcga caacccatct 1860

ctggccggcg gctgggatgt gaacaaggag tgcagcaatt tttgcgtgat cctgaaggac 1920

aagaacgaga agaagtacct ggccatcatg aagaagggcg agaataccct gttccagaag 1980

gagtggacag agggccgggg caagaacctg acaaagaagt ctaatccact gttcgagatc 2040

aataactgcg agatcctgag caagatggag tatgactttt gggccgacgt gagcaagatg 2100

atccccaagt gtagcaccca gctgaaggcc gtggtgaacc acttcaagca gtccgacaat 2160

gagttcatct ttcctatcgg ctacaaggtg acaagcggcg agaagtttag ggaggagtgc 2220

aagatctcca agcaggactt cgagctgaat aacaaggtgt ttaataagaa cgagctgagc 2280

gtgaccgcca tgcgctacga tctgtcctct acacaggaga agcagtatat caaggccttc 2340

cagaaggagt actgggagct gctgtttaag caggagaagc gggacaccaa gctgacaaat 2400

aacgagatct tcaacgagtg gatcaatttt tgcaacaaga agtatagcga gctgctgtcc 2460

tgggagagaa agtacaagga tgccctgacc aattggatca acttctgtaa gtactttctg 2520

agcaagtatc ccaagaccac actgttcaac tactctttta aggagagcga gaattataac 2580

tccctggacg agttctaccg ggacgtggat atctgttctt acaagctgaa tatcaacacc 2640

acaatcaata agagcatcct ggatagactg gtggaggagg gcaagctgta cctgtttgag 2700

atcaagaatc aggacagcaa cgatggcaag tccatcggcc acaagaataa cctgcacacc 2760

atctactgga acgccatctt cgagaatttt gacaacaggc ctaagctgaa tggcgaggcc 2820

gagatcttct atcgcaaggc catctccaag gataagctgg gcatcgtgaa gggcaagaaa 2880

accaagaacg gcaccgagat catcaagaat tacagattca gcaaggagaa gtttatcctg 2940

cacgtgccaa tcaccctgaa cttctgctcc aataacgagt atgtgaatga catcgtgaac 3000

acaaagttct acaatttttc caacctgcac tttctgggca tcgatagggg cgagaagcac 3060

ctggcctact attctctggt gaataagaac ggcgagatcg tggaccaggg cacactgaac 3120

ctgcctttca ccgacaagga tggcaatcag cgcagcatca agaaggagaa gtacttttat 3180

aacaagcagg aggacaagtg ggaggccaag gaggtggatt gttggaatta taacgacctg 3240

ctggatgcca tggcctctaa ccgggacatg gccagaaaga attggcagag gatcggcacc 3300

atcaaggagg ccaagaacgg ctacgtgagc ctggtcatca ggaagatcgc cgatctggcc 3360

gtgaataacg agcgccccgc cttcatcgtg ctggaggacc tgaatacagg ctttaagcgg 3420

tccagacaga agatcgataa gagcgtgtac cagaagttcg agctggccct ggccaagaag 3480

ctgaactttc tggtggacaa gaatgccaag cgcgatgaga tcggctcccc tacaaaggcc 3540

ctgcagctga ccccccctgt gaataactac ggcgacattg agaacaagaa gcaggccggc 3600

atcatgctgt atacccgggc caattatacc tctcagacag atccagccac aggctggaga 3660

aagaccatct atctgaaggc cggccccgag gagacaacat acaagaagga cggcaagatc 3720

aagaacaaga gcgtgaagga ccagatcatc gagacattca ccgatatcgg ctttgacggc 3780

aaggattact atttcgagta cgacaagggc gagtttgtgg atgagaaaac cggcgagatc 3840

aagcccaaga agtggcggct gtactccggc gagaatggca agtccctgga caggttccgc 3900

ggagagaggg agaaggataa gtatgagtgg aagatcgaca agatcgatat cgtgaagatc 3960

ctggacgatc tgttcgtgaa ttttgacaag aacatcagcc tgctgaagca gctgaaggag 4020

ggcgtggagc tgacccggaa taacgagcac ggcacaggcg agtccctgag attcgccatc 4080

aacctgatcc agcagatccg gaataccggc aataacgaga gagacaacga tttcatcctg 4140

tccccagtga gggacgagaa tggcaagcac tttgactctc gcgagtactg ggataaggag 4200

acaaagggcg agaagatcag catgcccagc tccggcgatg ccaatggcgc cttcaacatc 4260

gcccggaagg gcatcatcat gaacgcccac atcctggcca atagcgactc caaggatctg 4320

tccctgttcg tgtctgacga ggagtgggat ctgcacctga ataacaagac cgagtggaag 4380

aagcagctga acatcttttc tagcaggaag gccatggcca agcgcaagaa gaaaaggccg 4440

gcggccacga aaaaggccgg ccaggcaaaa aagaaaaagg gatcctaccc atacgatgtt 4500

ccagattacg cttatcccta cgacgtgcct gattatgcat acccatatga tgtccccgac 4560

tatgcctaag aattc 4575

<210> 218
<211> 4200
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 218
atggagaaca tcttcgacca gtttatcggc aagtacagcc tgtccaagac cctgagattc 60

gagctgaagc ccgtgggcaa gacagaggac ttcctgaaga tcaacaaggt gtttgagaag 120

gatcagacca tcgacgatag ctacaatcag gccaagttct attttgattc cctgcaccag 180

aagtttatcg acgccgccct ggcctccgat aagacatccg agctgtcttt ccagaacttt 240

gccgacgtgc tggagaagca gaataagatc atcctggata agaagagaga gatgggcgcc 300

ctgaggaagc gcgacaagaa cgccgtgggc atcgataggc tgcagaagga gatcaatgac 360

gccgaggata tcatccagaa ggagaaggag aagatctaca aggacgtgcg caccctgttc 420

gataacgagg ccgagtcttg gaaaacctac tatcaggagc gggaggtgga cggcaagaag 480

atcaccttca gcaaggccga cctgaagcag aagggcgccg attttctgac agccgccggc 540

atcctgaagg tgctgaagta tgagttcccc gaggagaagg agaaggagtt tcaggccaag 600

aaccagccct ccctgttcgt ggaggagaag gagaatcctg gccagaagag gtacatcttc 660

gactcttttg ataagttcgc cggctatctg accaagtttc agcagacaaa gaagaatctg 720

tacgcagcag acggcaccag cacagcagtg gccacccgca tcgccgataa ctttatcatc 780

ttccaccaga ataccaaggt gttccgggac aagtacaaga acaatcacac agacctgggc 840

ttcgatgagg agaacatctt tgagatcgag aggtataaga attgcctgct gcagcgcgag 900

atcgagcaca tcaagaatga gaatagctac aacaagatca tcggccggat caataagaag 960

atcaaggagt atcgggacca gaaggccaag gataccaagc tgacaaagtc cgacttccct 1020

ttctttaaga acctggataa gcagatcctg ggcgaggtgg agaaggagaa gcagctgatc 1080

gagaaaaccc gggagaaaac cgaggaggac gtgctgatcg agcggttcaa ggagttcatc 1140

gagaacaatg aggagaggtt caccgccgcc aagaagctga tgaatgcctt ctgtaacggc 1200

gagtttgagt ccgagtacga gggcatctat ctgaagaata aggccatcaa cacaatctcc 1260

cggagatggt tcgtgtctga cagagatttt gagctgaagc tgcctcagca gaagtccaag 1320

aacaagtctg agaagaatga gccaaaggtg aagaagttca tctccatcgc cgagatcaag 1380

aacgccgtgg aggagctgga cggcgatatc tttaaggccg tgttctacga caagaagatc 1440

atcgcccagg gcggctctaa gctggagcag ttcctggtca tctggaagta cgagtttgag 1500

tatctgttcc gggacatcga gagagagaac ggcgagaagc tgctgggcta tgatagctgc 1560

ctgaagatcg ccaagcagct gggcatcttc ccacaggaga aggaggcccg cgagaaggca 1620

accgccgtga tcaagaatta cgccgacgcc ggcctgggca tcttccagat gatgaagtat 1680

ttttctctgg acgataagga tcggaagaac acccccggcc agctgagcac aaatttctac 1740

gccgagtatg acggctacta caaggatttc gagtttatca agtactacaa cgagtttagg 1800

aacttcatca ccaagaagcc tttcgacgag gataagatca agctgaactt tgagaatggc 1860

gccctgctga agggctggga cgagaacaag gagtacgatt tcatgggcgt gatcctgaag 1920

aaggagggcc gcctgtatct gggcatcatg cacaagaacc accggaagct gtttcagtcc 1980

atgggcaatg ccaagggcga caacgccaat agataccaga agatgatcta taagcagatc 2040

gccgacgcct ctaaggatgt gcccaggctg ctgctgacca gcaagaaggc catggagaag 2100

ttcaagcctt cccaggagat cctgagaatc aagaaggaga aaaccttcaa gcgggagagc 2160

aagaactttt ccctgagaga tctgcacgcc ctgatcgagt actataggaa ctgcatccct 2220

cagtacagca attggtcctt ttatgacttc cagtttcagg ataccggcaa gtaccagaat 2280

atcaaggagt tcacagacga tgtgcagaag tacggctata agatctcctt tcgcgacatc 2340

gacgatgagt atatcaatca ggccctgaac gagggcaaga tgtacctgtt cgaggtggtg 2400

aacaaggata tctataacac caagaatggc tccaagaatc tgcacacact gtactttgag 2460

cacatcctgt ctgccgagaa cctgaatgac ccagtgttca agctgtctgg catggccgag 2520

atctttcagc ggcagcccag cgtgaacgaa agagagaaga tcaccacaca gaagaatcag 2580

tgtatcctgg acaagggcga tagagcctac aagtataggc gctacaccga gaagaagatc 2640

atgttccaca tgagcctggt gctgaacaca ggcaagggcg agatcaagca ggtgcagttt 2700

aataagatca tcaaccagag gatcagctcc tctgacaacg agatgagggt gaatgtgatc 2760

ggcatcgatc gcggcgagaa gaacctgctg tactatagcg tggtgaagca gaatggcgag 2820

atcatcgagc aggcctccct gaacgagatc aatggcgtga actaccggga caagctgatc 2880

gagagggaga aggagcgcct gaagaaccgg cagagctgga agcctgtggt gaagatcaag 2940

gatctgaaga agggctacat ctcccacgtg atccacaaga tctgccagct gatcgagaag 3000

tattctgcca tcgtggtgct ggaggacctg aatatgagat tcaagcagat caggggagga 3060

atcgagcgga gcgtgtacca gcagttcgag aaggccctga tcgataagct gggctatctg 3120

gtgtttaagg acaacaggga tctgagggca ccaggaggcg tgctgaatgg ctaccagctg 3180

tctgcccccc ttgtgagctt cgagaagatg cgcaagcaga ccggcatcct gttctacaca 3240

caggccgagt ataccagcaa gacagaccca atcaccggct ttcggaagaa cgtgtatatc 3300

tctaatagcg cctccctgga taagatcaag gaggccgtga agaagttcga cgccatcggc 3360

tgggatggca aggagcagtc ttacttcttt aagtacaacc cttacaacct ggccgacgag 3420

aagtataaga actctaccgt gagcaaggag tgggccatct ttgccagcgc cccaagaatc 3480

cggagacaga agggcgagga cggctactgg aagtatgata gggtgaaagt gaatgaggag 3540

ttcgagaagc tgctgaaggt ctggaatttt gtgaacccaa aggccacaga tatcaagcag 3600

gagatcatca agaaggagaa ggcaggcgac ctgcagggag agaaggagct ggatggccgg 3660

ctgagaaact tttggcactc tttcatctac ctgtttaacc tggtgctgga gctgcgcaat 3720

tctttcagcc tgcagatcaa gatcaaggca ggagaagtga tcgcagtgga cgagggcgtg 3780

gacttcatcg ccagcccagt gaagcccttc tttaccacac ccaaccctta catcccctcc 3840

aacctgtgct ggctggccgt ggagaatgca gacgcaaacg gagcctataa tatcgccagg 3900

aagggcgtga tgatcctgaa gaagatccgc gagcacgcca agaaggaccc cgagttcaag 3960

aagctgccaa acctgtttat cagcaatgca gagtgggacg aggcagcccg ggattggggc 4020

aagtacgcag gcaccacagc cctgaacctg gaccacaaaa ggccggcggc cacgaaaaag 4080

gccggccagg caaaaaagaa aaagggatcc tacccatacg atgttccaga ttacgcttat 4140

ccctacgacg tgcctgatta tgcataccca tatgatgtcc ccgactatgc ctaagaattc 4200

<210> 219
<211> 3894
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 219
atgcagaccc tgtttgagaa cttcacaaat cagtacccag tgtccaagac cctgcgcttt 60

gagctgatcc cccagggcaa gacaaaggac ttcatcgagc agaagggcct gctgaagaag 120

gatgaggacc gggccgagaa gtataagaag gtgaagaaca tcatcgatga gtaccacaag 180

gacttcatcg agaagtctct gaatggcctg aagctggacg gcctggagaa gtacaagacc 240

ctgtatctga agcaggagaa ggacgataag gataagaagg cctttgacaa ggagaaggag 300

aacctgcgca agcagatcgc caatgccttc cggaacaatg agaagtttaa gacactgttc 360

gccaaggagc tgatcaagaa cgatctgatg tctttcgcct gcgaggagga caagaagaat 420

gtgaaggagt ttgaggcctt caccacatac ttcaccggct tccaccagaa ccgcgccaat 480

atgtacgtgg ccgatgagaa gagaacagcc atcgccagca ggctgatcca cgagaacctg 540

ccaaagttta tcgacaatat caagatcttc gagaagatga agaaggaggc ccccgagctg 600

ctgtctcctt tcaaccagac cctgaaggat atgaaggacg tgatcaaggg caccacactg 660

gaggagatct ttagcctgga ttatttcaac aagaccctga cacagagcgg catcgacatc 720

tacaattccg tgatcggcgg cagaacccct gaggagggca agacaaagat caagggcctg 780

aacgagtaca tcaataccga cttcaaccag aagcagacag acaagaagaa gcggcagcca 840

aagttcaagc agctgtataa gcagatcctg agcgataggc agagcctgtc ctttatcgcc 900

gaggccttca agaacgacac cgagatcctg gaggccatcg agaagtttta cgtgaatgag 960

ctgctgcact tcagcaatga gggcaagtcc acaaacgtgc tggacgccat caagaatgcc 1020

gtgtctaacc tggagagctt taacctgacc aagatgtatt tccgctccgg cgcctctctg 1080

acagacgtga gccggaaggt gtttggcgag tggagcatca tcaatagagc cctggacaac 1140

tactatgcca ccacatatcc aatcaagccc agagagaagt ctgagaagta cgaggagagg 1200

aaggagaagt ggctgaagca ggacttcaac gtgagcctga tccagaccgc catcgatgag 1260

tacgacaacg agacagtgaa gggcaagaac agcggcaaag tgatcgccga ttattttgcc 1320

aagttctgcg acgataagga gacagacctg atccagaagg tgaacgaggg ctacatcgcc 1380

gtgaaggatc tgctgaatac accctgtcct gagaacgaga agctgggcag caataaggac 1440

caggtgaagc agatcaaggc ctttatggat tctatcatgg acatcatgca cttcgtgcgc 1500

cccctgagcc tgaaggatac cgacaaggag aaggatgaga cattctactc cctgttcaca 1560

cctctgtacg accacctgac ccagacaatc gccctgtata acaaggtgcg gaactatctg 1620

acccagaagc cttacagcac agagaagatc aagctgaact tcgagaacag caccctgctg 1680

ggcggctggg atctgaataa ggagacagac aacacagcca tcatcctgag gaaggataac 1740

ctgtactatc tgggcatcat ggacaagagg cacaatcgca tctttcggaa cgtgcccaag 1800

gccgataaga aggacttctg ctacgagaag atggtgtata agctgctgcc tggcgccaac 1860

aagatgctgc caaaggtgtt cttttctcag agcagaatcc aggagtttac cccttccgcc 1920

aagctgctgg agaactacgc caatgagaca cacaagaagg gcgataattt caacctgaat 1980

cactgtcaca agctgatcga tttctttaag gactctatca acaagcacga ggattggaag 2040

aatttcgact ttaggttcag cgccacctcc acctacgccg acctgagcgg cttttaccac 2100

gaggtggagc accagggcta caagatctct tttcagagcg tggccgattc cttcatcgac 2160

gatctggtga acgagggcaa gctgtacctg ttccagatct ataataagga cttttcccca 2220

ttctctaagg gcaagcccaa cctgcacacc ctgtactgga agatgctgtt tgatgagaac 2280

aatctgaagg acgtggtgta taagctgaat ggcgaggccg aggtgttcta ccgcaagaag 2340

agcattgccg agaagaacac cacaatccac aaggccaatg agtccatcat caacaagaat 2400

cctgataacc caaaggccac cagcaccttc aactatgata tcgtgaagga caagagatac 2460

accatcgaca agtttcagtt ccacatccca atcacaatga actttaaggc cgagggcatc 2520

ttcaacatga atcagagggt gaatcagttc ctgaaggcca atcccgatat caacatcatc 2580

ggcatcgaca gaggcgagag gcacctgctg tactatgccc tgatcaacca gaagggcaag 2640

atcctgaagc aggataccct gaatgtgatc gccaacgaga agcagaaggt ggactaccac 2700

aatctgctgg ataagaagga gggcgaccgc gcaaccgcaa ggcaggagtg gggcgtgatc 2760

gagacaatca aggagctgaa ggagggctat ctgtcccagg tcatccacaa gctgaccgat 2820

ctgatgatcg agaacaatgc catcatcgtg atggaggacc tgaactttgg cttcaagcgg 2880

ggcagacaga aggtggagaa gcaggtgtat cagaagtttg agaagatgct gatcgataag 2940

ctgaattacc tggtggacaa gaataagaag gcaaacgagc tgggaggcct gctgaacgca 3000

ttccagctgg ccaataagtt tgagtccttc cagaagatgg gcaagcagaa cggctttatc 3060

ttctacgtgc ccgcctggaa tacctctaag acagatcctg ccaccggctt tatcgacttc 3120

ctgaagcccc gctatgagaa cctgaatcag gccaaggatt tctttgagaa gtttgactct 3180

atccggctga acagcaaggc cgattacttt gagttcgcct ttgacttcaa gaatttcacc 3240

gagaaggccg atggcggcag aaccaagtgg acagtgtgca ccacaaacga ggacagatat 3300

gcctggaata gggccctgaa caataacagg ggcagccagg agaagtacga catcacagcc 3360

gagctgaagt ccctgttcga tggcaaggtg gactataagt ctggcaagga tctgaagcag 3420

cagatcgcca gccaggagtc cgccgacttc tttaaggccc tgatgaagaa cctgtccatc 3480

accctgtctc tgagacacaa taacggcgag aagggcgata atgagcagga ctacatcctg 3540

tcccctgtgg ccgattctaa gggccgcttc tttgactccc ggaaggccga cgatgacatg 3600

ccaaagaatg ccgacgccaa cggcgcctat cacatcgccc tgaagggcct gtggtgtctg 3660

gagcagatca gcaagaccga tgacctgaag aaggtgaagc tggccatctc caacaaggag 3720

tggctggagt tcgtgcagac actgaagggc aaaaggccgg cggccacgaa aaaggccggc 3780

caggcaaaaa agaaaaaggg atcctaccca tacgatgttc cagattacgc ttatccctac 3840

gacgtgcctg attatgcata cccatatgat gtccccgact atgcctaaga attc 3894

<210> 220
<211> 4065
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 220
atgacacagt tcgagggctt taccaacctg tatcaggtga gcaagacact gcggtttgag 60

ctgatcccac agggcaagac cctgaagcac atccaggagc agggcttcat cgaggaggac 120

aaggcccgca atgatcacta caaggagctg aagcccatca tcgatcggat ctacaagacc 180

tatgccgacc agtgcctgca gctggtgcag ctggattggg agaacctgag cgccgccatc 240

gactcctata gaaaggagaa aaccgaggag acaaggaacg ccctgatcga ggagcaggcc 300

acatatcgca atgccatcca cgactacttc atcggccgga cagacaacct gaccgatgcc 360

atcaataaga gacacgccga gatctacaag ggcctgttca aggccgagct gtttaatggc 420

aaggtgctga agcagctggg caccgtgacc acaaccgagc acgagaacgc cctgctgcgg 480

agcttcgaca agtttacaac ctacttctcc ggcttttatg agaacaggaa gaacgtgttc 540

agcgccgagg atatcagcac agccatccca caccgcatcg tgcaggacaa cttccccaag 600

tttaaggaga attgtcacat cttcacacgc ctgatcaccg ccgtgcccag cctgcgggag 660

cactttgaga acgtgaagaa ggccatcggc atcttcgtga gcacctccat cgaggaggtg 720

ttttccttcc ctttttataa ccagctgctg acacagaccc agatcgacct gtataaccag 780

ctgctgggag gaatctctcg ggaggcaggc accgagaaga tcaagggcct gaacgaggtg 840

ctgaatctgg ccatccagaa gaatgatgag acagcccaca tcatcgcctc cctgccacac 900

agattcatcc ccctgtttaa gcagatcctg tccgatagga acaccctgtc tttcatcctg 960

gaggagttta agagcgacga ggaagtgatc cagtccttct gcaagtacaa gacactgctg 1020

agaaacgaga acgtgctgga gacagccgag gccctgttta acgagctgaa cagcatcgac 1080

ctgacacaca tcttcatcag ccacaagaag ctggagacaa tcagcagcgc cctgtgcgac 1140

cactgggata cactgaggaa tgccctgtat gagcggagaa tctccgagct gacaggcaag 1200

atcaccaagt ctgccaagga gaaggtgcag cgcagcctga agcacgagga tatcaacctg 1260

caggagatca tctctgccgc aggcaaggag ctgagcgagg ccttcaagca gaaaaccagc 1320

gagatcctgt cccacgcaca cgccgccctg gatcagccac tgcctacaac cctgaagaag 1380

caggaggaga aggagatcct gaagtctcag ctggacagcc tgctgggcct gtaccacctg 1440

ctggactggt ttgccgtgga tgagtccaac gaggtggacc ccgagttctc tgcccggctg 1500

accggcatca agctggagat ggagccttct ctgagcttct acaacaaggc cagaaattat 1560

gccaccaaga agccctactc cgtggagaag ttcaagctga actttcagat gcctacactg 1620

gcctctggct gggacgtgaa taaggagaag aacaatggcg ccatcctgtt tgtgaagaac 1680

ggcctgtact atctgggcat catgccaaag cagaagggca ggtataaggc cctgagcttc 1740

gagcccacag agaaaaccag cgagggcttt gataagatgt actatgacta cttccctgat 1800

gccgccaaga tgatcccaaa gtgcagcacc cagctgaagg ccgtgacagc ccactttcag 1860

acccacacaa cccccatcct gctgtccaac aatttcatcg agcctctgga gatcacaaag 1920

gagatctacg acctgaacaa tcctgagaag gagccaaaga agtttcagac agcctacgcc 1980

aagaaaaccg gcgaccagaa gggctacaga gaggccctgt gcaagtggat cgacttcaca 2040

agggattttc tgtccaagta taccaagaca acctctatcg atctgtctag cctgcggcca 2100

tcctctcagt ataaggacct gggcgagtac tatgccgagc tgaatcccct gctgtaccac 2160

atcagcttcc agagaatcgc cgagaaggag atcatggatg ccgtggagac aggcaagctg 2220

tacctgttcc agatctataa caaggacttt gccaagggcc accacggcaa gcctaatctg 2280

cacacactgt attggaccgg cctgttttct ccagagaacc tggccaagac aagcatcaag 2340

ctgaatggcc aggccgagct gttctaccgc cctaagtcca ggatgaagag gatggcacac 2400

cggctgggag agaagatgct gaacaagaag ctgaaggatc agaaaaccccc aatccccgac 2460

accctgtacc aggagctgta cgactatgtg aatcacagac tgtcccacga cctgtctgat 2520

gaggccaggg ccctgctgcc caacgtgatc accaaggagg tgtctcacga gatcatcaag 2580

gataggcgct ttaccagcga caagttcttt ttccacgtgc ctatcacact gaactatcag 2640

gccgccaatt ccccatctaa gttcaaccag agggtgaatg cctacctgaa ggagcacccc 2700

gagacaccta tcatcggcat cgatcggggc gagagaaacc tgatctatat cacagtgatc 2760

gactccaccg gcaagatcct ggagcagcgg agcctgaaca ccatccagca gtttgattac 2820

cagaagaagc tggacaacag ggagaaggag agggtggcag caaggcaggc ctggtctgtg 2880

gtgggcacaa tcaaggatct gaagcagggc tatctgagcc aggtcatcca cgagatcgtg 2940

gacctgatga tccactacca ggccgtggtg gtgctggaga acctgaattt cggctttaag 3000

agcaagagga ccggcatcgc cgagaaggcc gtgtaccagc agttcgagaa gatgctgatc 3060

gataagctga attgcctggt gctgaaggac tatccagcag agaaagtggg aggcgtgctg 3120

aacccatacc agctgacaga ccagttcacc tcctttgcca agatgggcac ccagtctggc 3180

ttcctgtttt acgtgcctgc cccatataca tctaagatcg atcccctgac cggcttcgtg 3240

gacccccttcg tgtggaaaac catcaagaat cacgagagcc gcaagcactt cctggagggc 3300

ttcgactttc tgcactacga cgtgaaaacc ggcgacttca tcctgcactt taagatgaac 3360

agaaatctgt ccttccagag gggcctgccc ggctttatgc ctgcatggga tatcgtgttc 3420

gagaagaacg agacacagtt tgacgccaag ggcacccctt tcatcgccgg caagagaatc 3480

gtgccagtga tcgagaatca cagattcacc ggcagatacc gggacctgta tcctgccaac 3540

gagctgatcg ccctgctgga ggagaagggc atcgtgttca gggatggctc caacatcctg 3600

ccaaagctgc tggagaatga cgattctcac gccatcgaca ccatggtggc cctgatccgc 3660

agcgtgctgc agatgcggaa ctccaatgcc gccacaggcg aggactatat caacagcccc 3720

gtgcgcgatc tgaatggcgt gtgcttcgac tcccggtttc agaacccaga gtggcccatg 3780

gacgccgatg ccaatggcgc ctaccacatc gccctgaagg gccagctgct gctgaatcac 3840

ctgaaggaga gcaaggatct gaagctgcag aacggcatct ccaatcagga ctggctggcc 3900

tacatccagg agctgcgcaa caaaaggccg gcggccacga aaaaggccgg ccaggcaaaa 3960

aagaaaaagg gatcctaccc atacgatgtt ccagattacg cttatcccta cgacgtgcct 4020

gattatgcat acccatatga tgtccccgac tatgcctaag aattc 4065

<210> 221
<211> 3762
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 221
atgtactatg agtccctgac caagcagtac cccgtgtcta agacaatccg gaatgagctg 60

atccctatcg gcaagacact ggataacatc cgccagaaca atatcctgga gagcgacgtg 120

aagcggaagc agaactacga gcacgtgaag ggcatcctgg atgagtatca caagcagctg 180

atcaacgagg ccctggacaa ttgcaccctg ccatccctga agatcgccgc cgagatctac 240

ctgaagaatc agaaggaggt gtctgacaga gaggatttca acaagacaca ggacctgctg 300

aggaaggagg tggtggagaa gctgaaggcc cacgagaact ttaccaagat cggcaagaag 360

gacatcctgg atctgctgga gaagctgcct tccatctctg aggacgatta caatgccctg 420

gagagcttcc gcaactttta cacctatttc acatcctaca acaaggtgcg ggagaatctg 480

tattctgata aggagaagag ctccacagtg gcctacagac tgatcaacga gaatttccca 540

aagtttctgg acaatgtgaa gagctatagg tttgtgaaaa ccgcaggcat cctggcagat 600

ggcctgggag aggaggagca ggactccctg ttcatcgtgg agacattcaa caagaccctg 660

acacaggacg gcatcgatac ctacaattct caagtgggca agatcaactc tagcatcaat 720

ctgtataacc agaagaatca gaaggccaat ggcttcagaa agatccccaa gatgaagatg 780

ctgtataagc agatcctgtc cgatagggag gagtctttca tcgacgagtt tcagagcgat 840

gaggtgctga tcgacaacgt ggagtcttat ggcagcgtgc tgatcgagtc tctgaagtcc 900

tctaaggtga gcgccttctt tgatgccctg agagagtcta agggcaagaa cgtgtacgtg 960

aagaatgacc tggccaagac agccatgagc aacatcgtgt tcgagaattg gaggaccttt 1020

gacgatctgc tgaaccagga gtacgacctg gccaacgaga acaagaagaa ggacgataag 1080

tatttcgaga agcgccagaa ggagctgaag aagaataaga gctactccct ggagcacctg 1140

tgcaacctgt ccgaggattc ttgtaacctg atcgagaatt atatccacca gatctccgac 1200

gatatcgaga atatcatcat caacaatgag acattcctgc gcatcgtgat caatgagcac 1260

gacaggtccc gcaagctggc caagaaccgg aaggccgtga aggccatcaa ggactttctg 1320

gattctatca aggtgctgga gcgggagctg aagctgatca acagctccgg ccaggagctg 1380

gagaaggatc tgatcgtgta ctctgcccac gaggagctgc tggtggagct gaagcaggtg 1440

gacagcctgt ataacatgac cagaaattat ctgacaaaga agcctttctc taccgagaag 1500

gtgaagctga actttaatcg cagcacactg ctgaacggct gggatcggaa taaggagaca 1560

gacaacctgg gcgtgctgct gctgaaggac ggcaagtact atctgggcat catgaacaca 1620

agcgccaata aggccttcgt gaatcccccc gtggccaaga ccgagaaggt gtttaagaag 1680

gtggattaca agctgctgcc agtgcccaac cagatgctgc caaaggtgtt ctttgccaag 1740

agcaatatcg acttctataa cccctctagc gagatctact ccaattataa gaagggcacc 1800

cacaagaagg gcaatatgtt ttccctggag gattgtcaca acctgatcga cttctttaag 1860

gagtctatca gcaagcacga ggactggagc aagttcggct ttaagttcag cgatacagcc 1920

tcctacaacg acatctccga gttctatcgc gaggtggaga agcagggcta caagctgacc 1980

tatacagaca tcgatgagac atacatcaat gatctgatcg agcggaacga gctgtacctg 2040

ttccagatct ataataagga ctttagcatg tactccaagg gcaagctgaa cctgcacaca 2100

ctgtatttca tgatgctgtt tgatcagcgc aatatcgacg acgtggtgta taagctgaac 2160

ggagaggcag aggtgttcta taggccagcc tccatctctg aggacgagct gatcatccac 2220

aaggccggcg aggagatcaa gaacaagaat cctaaccggg ccagaaccaa ggagacaagc 2280

accttcagct acgacatcgt gaaggataag cggtatagca aggataagtt taccctgcac 2340

atccccatca caatgaactt cggcgtggat gaggtgaagc ggttcaacga cgccgtgaac 2400

agcgccatcc ggatcgatga gaatgtgaac gtgatcggca tcgaccgggg cgagagaaat 2460

ctgctgtacg tggtggtcat cgactctaag ggcaacatcc tggagcagat ctccctgaac 2520

tctatcatca ataaggagta cgacatcgag acagattatc acgcactgct ggatgagagg 2580

gagggcggca gagataaggc ccggaaggac tggaacaccg tggagaatat cagggacctg 2640

aaggccggct acctgagcca ggtggtgaac gtggtggcca agctggtgct gaagtataat 2700

gccatcatct gcctggagga cctgaacttt ggcttcaaga ggggccgcca gaaggtggag 2760

aagcaggtgt accagaagtt cgagaagatg ctgatcgata agctgaatta cctggtcatc 2820

gacaagagcc gcgagcagac atcccctaag gagctgggag gcgccctgaa cgcactgcag 2880

ctgacctcta agttcaagag ctttaaggag ctgggcaagc agtccggcgt gatctactat 2940

gtgcctgcct acctgacctc taagatcgat ccaaccacag gcttcgccaa tctgttttat 3000

atgaagtgtg agaacgtgga gaagtccaag agattctttg acggctttga tttcatcagg 3060

ttcaacgccc tggagaacgt gttcgagttc ggctttgact accggagctt cacccagagg 3120

gcctgcggca tcaattccaa gtggaccgtg tgcaccaacg gcgagcgcat catcaagtat 3180

cggaatccag ataagaacaa tatgttcgac gagaaggtgg tggtggtgac cgatgagatg 3240

aagaacctgt ttgagcagta caagatcccc tatgaggatg gcagaaatgt gaaggacatg 3300

atcatcagca acgaggaggc cgagttctac cggagactgt ataggctgct gcagcagacc 3360

ctgcagatga gaaacagcac ctccgacggc acaagggatt acatcatctc ccctgtgaag 3420

aataagagag aggcctactt caacagcgag ctgtccgacg gctctgtgcc aaaggacgcc 3480

gatgccaacg gcgcctacaa tatcgccaga aagggcctgt gggtgctgga gcagatcagg 3540

cagaagagcg agggcgagaa gatcaatctg gccatgacca acgccgagtg gctggagtat 3600

gcccagacac acctgctgaa aaggccggcg gccacgaaaa aggccggcca ggcaaaaaag 3660

aaaaagggat cctacccata cgatgttcca gattacgctt atccctacga cgtgcctgat 3720

tatgcatacc catatgatgt ccccgactat gcctaagaat tc 3762

<210> 222
<211> 3858
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 222
atgaacaatt acgacgagtt caccaagctg tatcctatcc agaaaaccat ccggtttgag 60

ctgaagccac agggcagaac catggagcac ctggagacat tcaacttctt tgaggaggac 120

cgggatagag ccgagaagta taagatcctg aaggaggcca tcgacgagta ccacaagaag 180

tttatcgatg agcacctgac caatatgtcc ctggattgga actctctgaa gcagatcagc 240

gagaagtact ataagagcag ggaggagaag gacaagaagg tgttcctgtc cgagcagaag 300

aggatgcgcc aggagatcgt gtctgagttt aagaaggacg atcgcttcaa ggacctgttt 360

tccaagaagc tgttctctga gctgctgaag gaggagatct acaagaaggg caaccaccag 420

gagatcgacg ccctgaagag cttcgataag ttttccggct atttcatcgg cctgcacgag 480

aataggaaga acatgtactc cgacggcgat gagatcaccg ccatctccaa tcgcatcgtg 540

aatgagaact tccccaagtt tctggataac ctgcagaagt accaggaggc caggaagaag 600

tatcctgagt ggatcatcaa ggccgagagc gccctggtgg cccacaatat caagatggac 660

gaggtgttct ccctggagta ctttaataag gtgctgaacc aggagggcat ccagcggtac 720

aacctggccc tgggcggcta tgtgaccaag agcggcgaga agatgatggg cctgaatgat 780

gccctgaacc tggcccacca gtccgagaag agctccaagg gcagaatcca catgaccccc 840

ctgttcaagc agatcctgtc cgagaaggag tccttctctt acatccccga cgtgtttaca 900

gaggattctc agctgctgcc tagcatcggc ggcttctttg cccagatcga gaatgacaag 960

gatggcaaca tcttcgaccg ggcccctggag ctgatctcta gctacgccga gtatgatacc 1020

gagcggatct atatcagaca ggccgacatc aatagagtgt ccaacgtgat ctttggagag 1080

tggggcaccc tgggaggcct gatgagggag tacaaggccg actctatcaa tgatatcaac 1140

ctggagcgca catgcaagaa ggtggacaag tggctggatt ctaaggagtt tgccctgagc 1200

gatgtgctgg aggccatcaa gaggaccggc aacaatgacg ccttcaacga gtatatctcc 1260

aagatgcgga cagccagaga gaagatcgat gccgcccgca aggagatgaa gttcatcagc 1320

gagaagatct ccggcgatga ggagtctatc cacatcatca agaccctgct ggacagcgtg 1380

cagcagttcc tgcacttctt taatctgttt aaggcaaggc aggacatccc actggatgga 1440

gccttctacg ccgagtttga cgaggtgcac agcaagctgt ttgccatcgt gcccctgtat 1500

aacaaggtgc ggaactatct gaccaagaac aatctgaaca caaagaagat caagctgaat 1560

ttcaagaacc ctacactggc caatggctgg gaccagaaca aggtgtacga ttatgcctcc 1620

ctgatctttc tgcgggacgg caattactat ctgggcatca tcaatcctaa gagaaagaag 1680

aacatcaagt tcgagcaggg ctctggcaac ggccccttct accggaagat ggtgtataag 1740

cagatccccg gccctaataa gaacctgcca agagtgttcc tgacctccac aaagggcaag 1800

aaggagtata agccctctaa ggagatcatc gagggctacg aggccgacaa gcacatcagg 1860

ggcgataagt tcgacctgga tttttgtcac aagctgatcg atttctttaa ggagtccatc 1920

gagaagcaca aggactggtc taagttcaac ttctacttca gcccaaccga gagctatggc 1980

gacatctctg agttctacct ggatgtggag aagcagggct atcgcatgca ctttgagaat 2040

atcagcgccg agacaatcga cgagtatgtg gagaagggcg atctgtttct gttccagatc 2100

tacaacaagg attttgtgaa ggccgccacc ggcaagaagg acatgcacac aatctactgg 2160

aatgccgcct tcagccccga gaacctgcag gacgtggtgg tgaagctgaa cggcgaggcc 2220

gagctgtttt atagggacaa gtccgatatc aaggagatcg tgcaccgcga gggcgagatc 2280

ctggtgaata ggacctacaa cggccgcaca ccagtgcccg acaagatcca caagaagctg 2340

accgattatc acaatggccg gacaaaggac ctgggcgagg ccaaggagta cctggataag 2400

gtgagatact tcaaggccca ctatgacatc accaaggatc ggagatacct gaacgacaag 2460

atctatttcc acgtgcctct gaccctgaac ttcaaggcca acggcaagaa gaatctgaac 2520

aagatggtca tcgagaagtt cctgtccgat gagaaggccc acatcatcgg catcgacagg 2580

ggcgagcgca atctgctgta ctattccatc atcgacaggt ctggcaagat catcgatcag 2640

cagagcctga atgtgatcga cggctttgat tatcgggaga agctgaacca gagagagatc 2700

gagatgaagg atgcccgcca gtcttggaac gccatcggca agatcaagga cctgaaggag 2760

ggctacctga gcaaggccgt gcacgagatc accaagatgg ccatccagta taatgccatc 2820

gtggtcatgg aggagctgaa ctacggcttc aagcggggcc ggttcaaggt ggagaagcag 2880

atctatcaga agttcgagaa tatgctgatc gataagatga actacctggt gtttaaggac 2940

gcacctgatg agtccccagg aggcgtgctg aatgcctacc agctgacaaa cccactggag 3000

tctttcgcca agctgggcaa gcagaccggc atcctgtttt acgtgccagc cgcctataca 3060

tccaagatcg accccaccac aggcttcgtg aatctgttta acacctcctc taagacaaac 3120

gcccaggagc ggaaggagtt cctgcagaag tttgagagca tctcctattc tgccaaggat 3180

ggcggcatct ttgccttcgc ctttgactac agaaagttcg gcaccagcaa gacagatcac 3240

aagaacgtgt ggaccgccta tacaaacggc gagaggatgc gctacatcaa ggagaagaag 3300

cggaatgagc tgtttgaccc ttctaaggag atcaaggagg ccctgaccag ctccggcatc 3360

aagtacgatg gcggccagaa catcctgcca gacatcctga ggagcaacaa taacggcctg 3420

atctacacaa tgtattctag cttcatcgcc gccatccaga tgcgcgtgta cgacggcaag 3480

gaggattata tcatcagccc catcaagaac tccaagggcg agttctttag gaccgacccc 3540

aagaggcgcg agctgcctat cgacgccgat gccaatggcg cctacaacat cgccctgagg 3600

ggagagctga caatgagggc aatcgcagag aagttcgacc ctgatagcga gaagatggcc 3660

aagctggagc tgaagcacaa ggattggttc gagtttatgc agaccagagg cgacaaaagg 3720

ccggcggcca cgaaaaaggc cggccaggca aaaaagaaaa agggatccta cccatacgat 3780

gttccagatt acgcttatcc ctacgacgtg cctgattatg catacccata tgatgtcccc 3840

gactatgcct aagaattc 3858

<210> 223
<211> 3990
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 223
atgaacggca ataggtccat cgtgtaccgc gagttcgtgg gcgtgatccc cgtggccaag 60

accctgagga atgagctgcg ccctgtgggc cacacacagg agcacatcat ccagaacggc 120

ctgatccagg aggacgagct gcggcaggag aagagcaccg agctgaagaa catcatggac 180

gattactata gagagtacat cgataagtct ctgagcggcg tgaccgacct ggacttcacc 240

ctgctgttcg agctgatgaa cctggtgcag agctcccccct ccaaggacaa taagaaggcc 300

ctggagaagg agcagtctaa gatgagggag cagatctgca cccacctgca gtccgactct 360

aactacaaga atatctttaa cgccaagctg ctgaaggaga tcctgcctga tttcatcaag 420

aactacaatc agtatgacgt gaaggataag gccggcaagc tggagacact ggccctgttt 480

aatggcttca gcacatactt taccgacttc tttgagaaga ggaagaacgt gttcaccaag 540

gaggccgtga gcacatccat cgcctaccgc atcgtgcacg agaactccct gatcttcctg 600

gccaatatga cctcttataa gaagatcagc gagaaggccc tggatgagat cgaagtgatc 660

gagaagaaca atcaggacaa gatgggcgat tgggagctga atcagatctt taaccctgac 720

ttctacaata tggtgctgat ccagtccggc atcgacttct acaacgagat ctgcggcgtg 780

gtgaatgccc acatgaacct gtactgtcag cagaccaaga acaattataa cctgttcaag 840

atgcggaagc tgcacaagca gatcctggcc tacaccagca ccagcttcga ggtgcccaag 900

atgttcgagg acgatatgag cgtgtataac gccgtgaacg ccttcatcga cgagacagag 960

aagggcaaca tcatcggcaa gctgaaggat atcgtgaata agtacgacga gctggatgag 1020

aagagaatct atatcagcaa ggacttttac gagacactga gctgcttcat gtccggcaac 1080

tggaatctga tcacaggctg cgtggagaac ttctacgatg agaacatcca cgccaagggc 1140

aagtccaagg aggagaaggt gaagaaggcc gtgaaggagg acaagtacaa gtctatcaat 1200

gacgtgaacg atctggtgga gaagtatatc gatgagaagg agaggaatga gttcaagaac 1260

agcaatgcca agcagtacat ccgcgagatc tccaacatca tcaccgacac agagacagcc 1320

cacctggagt atgacgatca catctctctg atcgagagcg aggagaaggc cgacgagatg 1380

aagaagcggc tggatatgta tatgaacatg taccactggg ccaaggcctt tatcgtggac 1440

gaggtgctgg acagagatga gatgttctac agcgatatcg acgatatcta taatatcctg 1500

gagaacatcg tgccactgta taatcgggtg agaaactacg tgacccagaa gccctacaac 1560

tctaagaaga tcaagctgaa tttccagagc cctacactgg ccaatggctg gtcccagtct 1620

aaggagttcg acaacaatgc catcatcctg atcagagata acaagtacta tctggccatc 1680

ttcaatgcca agaacaagcc agacaagaag atcatccagg gcaactccga taagaagaac 1740

gacaacgatt acaagaagat ggtgtataac ctgctgccag gcgccaacaa gatgctgccc 1800

aaggtgtttc tgtctaagaa gggcatcgag acattcaagc cctccgacta tatcatctct 1860

ggctacaacg cccacaagca catcaagaca agcgagaatt ttgatatctc cttctgtcgg 1920

gacctgatcg attacttcaa gaacagcatc gagaagcacg ccgagtggag aaagtatgag 1980

ttcaagtttt ccgccaccga cagctactcc gatatctctg agttctatcg ggaggtggag 2040

atgcagggct acagaatcga ctggacatat atcagcgagg ccgacatcaa caagctggat 2100

gaggagggca agatctatct gtttcagatc tacaataagg atttcgccga gaacagcacc 2160

ggcaaggaga atctgcacac aatgtacttt aagaacatct tctccgagga gaatctgaag 2220

gacatcatca tcaagctgaa cggccaggcc gagctgtttt atcggagagc ctctgtgaag 2280

aatcccgtga agcacaagaa ggatagcgtg ctggtgaaca agacctacaa gaatcagctg 2340

gacaacggcg acgtggtgag aatccccatc cctgacgata tctataacga gatctacaag 2400

atgtataatg gctacatcaa ggagtccgac ctgtctgagg ccgccaagga gtacctggat 2460

aaggtggagg tgaggaccgc ccagaaggac atcgtgaagg attaccgcta tacagtggac 2520

aagtacttca tccacacacc tatcaccatc aactataagg tgaccgcccg caacaatgtg 2580

aatgatatgg tggtgaagta catcgcccag aacgacgata tccacgtgat cggcatcgac 2640

cggggcgaga gaaacctgat ctacatctcc gtgatcgatt ctcacggcaa catcgtgaag 2700

cagaaatcct acaacatcct gaacaactac gactacaaga agaagctggt ggagaaggag 2760

aaaacccggg agtacgccag aaagaactgg aagagcatcg gcaatatcaa ggagctgaag 2820

gagggctata tctccggcgt ggtgcacgag atcgccatgc tgatcgtgga gtacaacgcc 2880

atcatcgcca tggaggacct gaattatggc tttaagaggg gccgcttcaa ggtggagcgg 2940

caggtgtacc agaagtttga gagcatgctg atcaataagc tgaactattt cgccagcaag 3000

gagaagtccg tggacgagcc aggaggcctg ctgaagggct atcagctgac ctacgtgccc 3060

gataatatca agaacctggg caagcagtgc ggcgtgatct tttacgtgcc tgccgccttc 3120

accagcaaga tcgacccatc cacaggcttt atctctgcct tcaactttaa gtctatcagc 3180

acaaatgcct ctcggaagca gttctttatg cagtttgacg agatcagata ctgtgccgag 3240

aaggatatgt tcagctttgg cttcgactac aacaacttcg atacctacaa catcacaatg 3300

ggcaagacac agtggaccgt gtatacaaac ggcgagagac tgcagtctga gttcaacaat 3360

gccaggcgca ccggcaagac aaagagcatc aatctgacag agacaatcaa gctgctgctg 3420

gaggacaatg agatcaacta cgccgacggc cacgatatca ggatcgatat ggagaagatg 3480

gacgaggata agaagagcga gttctttgcc cagctgctga gcctgtataa gctgaccgtg 3540

cagatgcgca attcctatac agaggccgag gagcaggaga acggcatctc ttacgacaag 3600

atcatcagcc ctgtgatcaa tgatgagggc gagttctttg actccgataa ctataaggag 3660

tctgacgata aggagtgcaa gatgccaaag gacgccgatg ccaacggcgc ctactgtatc 3720

gccctgaagg gcctgtatga ggtgctgaag atcaagagcg agtggaccga ggacggcttt 3780

gataggaatt gcctgaagct gccacacgca gagtggctgg acttcatcca gaacaagcgg 3840

tacgagaaaa ggccggcggc cacgaaaaag gccggccagg caaaaaagaa aaagggatcc 3900

tacccatacg atgttccaga ttacgcttat ccctacgacg tgcctgatta tgcataccca 3960

tatgatgtcc ccgactatgc ctaagaattc 3990

<210> 224
<211> 4263
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 224
atgctgttcc aggactttac ccacctgtat ccactgtcca agacagtgag atttgagctg 60

aagcccatcg ataggaccct ggagcacatc cacgccaaga acttcctgtc tcaggacgag 120

acaatggccg atatgcacca gaaggtgaaa gtgatcctgg acgattacca ccgcgacttc 180

atcgccgata tgatgggcga ggtgaagctg accaagctgg ccgagttcta tgacgtgtac 240

ctgaagtttc ggaagaaccc aaaggacgat gagctgcaga agcagctgaa ggatctgcag 300

gccgtgctga gaaaggagat cgtgaagccc atcggcaatg gcggcaagta taaggccggc 360

tacgacaggc tgttcggcgc caagctgttt aaggacggca aggagctggg cgatctggcc 420

aagttcgtga tcgcacagga gggagagagc tccccaaagc tggcccacct ggcccacttc 480

gagaagtttt ccacctattt cacaggcttt cacgataacc ggaagaatat gtattctgac 540

gaggataagc acaccgccat cgcctaccgc ctgatccacg agaacctgcc ccggtttatc 600

gacaatctgc agatcctgac cacaatcaag cagaagcact ctgccctgta cgatcagatc 660

atcaacgagc tgaccgccag cggcctggac gtgtctctgg ccagccacct ggatggctat 720

cacaagctgc tgacacagga gggcatcacc gcctacaata cactgctggg aggaatctcc 780

ggagaggcag gctctcctaa gatccagggc atcaacgagc tgatcaattc tcaccacaac 840

cagcactgcc acaagagcga gaaatcgcc aagctgaggc cactgcacaa gcagatcctg 900

tccgacggca tgagcgtgtc cttcctgccc tctaagtttg ccgacgatag cgagatgtgc 960

caggccgtga acgagttcta tcgccactac gccgacgtgt tcgccaaggt gcagagcctg 1020

ttcgacggct ttgacgatca ccagaaggat ggcatctacg tggagcacaa gaacctgaat 1080

gagctgtcca agcaggcctt cggcgacttt gcactgctgg gacgcgtgct ggacggatac 1140

tatgtggatg tggtgaatcc agagttcaac gagcggtttg ccaaggccaa gaccgacaat 1200

gccaaggcca agctgacaaa ggagaaggat aagttcatca agggcgtgca ctccctggcc 1260

tctctggagc aggccatcga gcactatacc gcaaggcacg acgatgagag cgtgcaggca 1320

ggcaagctgg gacagtactt caagcacggc ctggccggag tggacaaccc catccagaag 1380

atccacaaca atcacagcac catcaagggc tttctggaga gggagcgccc tgcaggagag 1440

agagccctgc caaagatcaa gtccggcaag aatcctgaga tgacacagct gaggcagctg 1500

aaggagctgc tggataacgc cctgaatgtg gcccacttcg ccaagctgct gaccacaaag 1560

accacactgg acaatcagga tggcaacttc tatggcgagt ttggcgtgct gtacgacgag 1620

ctggccaaga tccccaccct gtataacaag gtgagagatt acctgagcca gaagcctttc 1680

tccaccgaga agtacaagct gaactttggc aatccaacac tgctgaatgg ctgggacctg 1740

aacaaggaga aggataattt cggcgtgatc ctgcagaagg acggctgcta ctatctggcc 1800

ctgctggaca aggcccacaa gaaggtgttt gataacgccc ctaatacagg caagagcatc 1860

tatcagaaga tgatctataa gtacctggag gtgaggaagc agttccccaa ggtgttcttt 1920

tccaaggagg ccatcgccat caactaccac ccttctaagg agctggtgga gatcaaggac 1980

aagggccggc agagatccga cgatgagcgc ctgaagctgt atcggtttat cctggagtgt 2040

ctgaagatcc accctaagta cgataagaag ttcgagggcg ccatcggcga catccagctg 2100

tttaagaagg ataagaaggg cagagaggtg ccaatcagcg agaaggacct gttcgataag 2160

atcaacggca tcttttctag caagcctaag ctggagatgg aggacttctt tatcggcgag 2220

ttcaagaggt ataacccaag ccaggacctg gtggatcagt ataatatcta caagaagatc 2280

gactccaacg ataatcgcaa gaaggagaat ttctacaaca atcaccccaa gtttaagaag 2340

gatctggtgc ggtactatta cgagtctatg tgcaagcacg aggagtggga ggagagcttc 2400

gagttttcca agaagctgca ggacatcggc tgttacgtgg atgtgaacga gctgtttacc 2460

gagatcgaga cacggagact gaattataag atctccttct gcaacatcaa tgccgactac 2520

atcgatgagc tggtggagca gggccagctg tatctgttcc agatctacaa caaggacttt 2580

tccccaaagg cccacggcaa gcccaatctg cacaccctgt acttcaaggc cctgttttct 2640

gaggacaacc tggccgatcc tatctataag ctgaatggcg aggcccagat cttctacaga 2700

aaggcctccc tggacatgaa cgagacaaca atccacaggg ccggcgaggt gctggagaac 2760

aagaatcccg ataatcctaa gaagagacag ttcgtgtacg acatcatcaa ggataagagg 2820

tacacacagg acaagttcat gctgcacgtg ccaatcacca tgaactttgg cgtgcagggc 2880

atgacaatca aggagttcaa taagaaggtg aaccagtcta tccagcagta tgacgaggtg 2940

aacgtgatcg gcatcgatcg gggcgagaga cacctgctgt acctgaccgt gatcaatagc 3000

aagggcgaga tcctggagca gtgttccctg aacgacatca ccacagcctc tgccaatggc 3060

acacagatga ccacacctta ccacaagatc ctggataaga gggagatcga gcgcctgaac 3120

gcccgggtgg gatggggcga gatcgagaca atcaaggagc tgaagtctgg ctatctgagc 3180

cacgtggtgc accagatcag ccagctgatg ctgaagtaca acgccatcgt ggtgctggag 3240

gacctgaatt tcggctttaa gaggggccgc tttaaggtgg agaagcagat ctatcagaac 3300

ttcgagaatg ccctgatcaa gaagctgaac cacctggtgc tgaaggacaa ggccgacgat 3360

gagatcggct cttacaagaa tgccctgcag ctgaccaaca atttcacaga tctgaagagc 3420

atcggcaagc agaccggctt cctgttttat gtgcccgcct ggaacacctc taagatcgac 3480

cctgagacag gctttgtgga tctgctgaag ccaagatacg agaacatcgc ccagagccag 3540

gccttctttg gcaagttcga caagatctgc tataatgccg acaaggatta cttcgagttt 3600

cacatcgact acgccaagtt taccgataag gccaagaata gccgccagat ctggacaatc 3660

tgttcccacg gcgacaagcg gtacgtgtac gataagacag ccaaccagaa taagggcgcc 3720

gccaagggca tcaacgtgaa tgatgagctg aagtccctgt tcgcccgcca ccacatcaac 3780

gagaagcagc ccaacctggt catggacatc tgccagaaca atgataagga gtttcacaag 3840

tctctgatgt acctgctgaa aaccctgctg gccctgcggt acagcaacgc ctcctctgac 3900

gaggatttca tcctgtcccc cgtggcaaac gacgagggcg tgttctttaa tagcgccctg 3960

gccgacgata cacagcctca gaatgccgat gccaacggcg cctaccacat cgccctgaag 4020

ggcctgtggc tgctgaatga gctgaagaac tccgacgatc tgaacaaggt gaagctggcc 4080

atcgacaatc agacctggct gaatttcgcc cagaacagga aaaggccggc ggccacgaaa 4140

aaggccggcc aggcaaaaaa gaaaaaggga tcctacccat acgatgttcc agattacgct 4200

tatccctacg acgtgcctga ttatgcatac ccatatgatg tccccgacta tgcctaagaa 4260

ttc 4263

<210> 225
<211> 3933
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 225
atggaggact attccggctt tgtgaacatc tactctatcc agaaaaccct gaggttcgag 60

ctgaagccag tgggcaagac actggagcac atcgagaaga agggcttcct gaagaaggac 120

aagatccggg ccgaggatta caaggccgtg aagaagatca tcgataagta ccacagagcc 180

tatatcgagg aggtgtttga ttccgtgctg caccagaaga agaagaagga caagacccgc 240

ttttctacac agttcatcaa ggagatcaag gagttcagcg agctgtacta taagaccgag 300

aagaacatcc ccgacaagga gaggctggag gccctgagcg agaagctgcg caagatgctg 360

gtgggcgcct ttaagggcga gttctccgag gaggtggccg agaagtataa gaacctgttt 420

tctaaggagc tgatcaggaa tgagatcgag aagttctgcg agacagacga ggagcgcaag 480

caggtgtcta acttcaagag cttcaccaca tactttaccg gcttccactc caacaggcag 540

aatatctatt ccgacgagaa gaagtctaca gccatcggct accgcatcat ccaccagaac 600

ctgcctaagt tcctggataa tctgaagatc atcgagtcca tccagcggcg gttcaaggac 660

ttcccatggt ctgatctgaa gaagaacctg aagaagatcg ataagaatat caagctgacc 720

gagtacttca gcatcgacgg cttcgtgaac gtgctgaatc agaagggcat cgatgcctac 780

aacacaatcc tgggcggcaa gtccgaggag tctggcgaga agatccaggg cctgaacgag 840

tacatcaatc tgtatcggca gaagaacaat atcgacagaa agaacctgcc caatgtgaag 900

atcctgttta agcagatcct gggcgatagg gagacaaaga gctttatccc tgaggccttc 960

ccagacgatc agtccgtgct gaactctatc acagagttcg ccaagtacct gaagctggat 1020

aagaagaaga agagcatcat cgccgagctg aagaagtttc tgagctcctt caatcgctac 1080

gagctggacg gcatctatct ggccaacgat aatagcctgg cctctatcag caccttcctg 1140

tttgacgatt ggtcctttat caagaagtcc gtgtctttca agtatgacga gtccgtgggc 1200

gaccccaaga agaagatcaa gtctcccctg aagtacgaga aggagaagga gaagtggctg 1260

aagcagaagt actatacaat ctctttcctg aacgatgcca tcgagagcta ttccaagtct 1320

caggacgaga agagggtgaa gatccgcctg gaggcctact ttgccgagtt caagagcaag 1380

gacgatgcca agaagcagtt cgacctgctg gagaggatcg aggaggccta tgccatcgtg 1440

gagcctctgc tgggagcaga gtacccaagg gaccgcaacc tgaaggccga taagaaggaa 1500

gtgggcaaga tcaaggactt cctggatagc atcaagtccc tgcagttctt tctgaagcct 1560

ctgctgtccg ccgagatctt tgacgagaag gatctgggct tctacaatca gctggagggc 1620

tactatgagg agatcgattc tatcggccac ctgtataaca aggtgcggaa ttatctgacc 1680

ggcaagatct acagcaagga gaagtttaag ctgaacttcg agaacagcac cctgctgaag 1740

ggctgggacg agaaccggga ggtggccaat ctgtgcgtga tcttcagaga ggaccagaag 1800

tactatctgg gcgtgatgga taaggagaac aataccatcc tgtccgacat ccccaaggtg 1860

aagcctaacg agctgtttta cgagaagatg gtgtataagc tgatccccac acctcacatg 1920

cagctgcccc ggatcatctt ctctagcgac aacctgtcta tctataatcc tagcaagtcc 1980

atcctgaaga tcagagaggc caagagcttt aaggagggca agaacttcaa gctgaaggac 2040

tgtcacaagt ttatcgattt ctacaaggag tctatcagca agaatgagga ctggagcaga 2100

ttcgacttca agttcagcaa gaccagcagc tacgagaaca tcagcgagtt ttaccgggag 2160

gtggagagac agggctataa cctggacttc aagaaggtgt ctaagttcta catcgacagc 2220

ctggtggagg atggcaagct gtacctgttc cagatctata acaaggactt ttctatcttc 2280

agcaagggca agcccaatct gcacaccatc tattttcggt ccctgttctc taaggagaac 2340

ctgaaggacg tgtgcctgaa gctgaatggc gaggccgaga tgttctttcg gaagaagtcc 2400

atcaactacg atgagaagaa gaagcgggag ggccaccacc ccgagctgtt tgagaagctg 2460

aagtatccta tcctgaagga caagagatac agcgaggata agtttcagtt ccacctgccc 2520

atcagcctga acttcaagtc caaggagcgg ctgaacttta atctgaaagt gaatgagttc 2580

ctgaagagaa acaaggacat caatatcatc ggcatcgatc ggggcgagag aaacctgctg 2640

tacctggtca tgatcaatca gaagggcgag atcctgaagc agaccctgct ggacagcatg 2700

cagtccggca agggccggcc tgagatcaac tacaaggaga agctgcagga gaaggagatc 2760

gagagggata aggcccgcaa gagctggggc acagtggaga atatcaagga gctgaaggag 2820

ggctatctgt ctatcgtgat ccaccagatc agcaagctga tggtggagaa caatgccatc 2880

gtggtgctgg aggacctgaa catcggcttt aagcggggca gacagaaggt ggagcggcag 2940

gtgtaccaga agttcgagaa gatgctgatc gataagctga actttctggt gttcaaggag 3000

aataagccaa ccgagccagg aggcgtgctg aaggcctatc agctgacaga cgagtttcag 3060

tctttcgaga agctgagcaa gcagaccggc tttctgttct acgtgccaag ctggaacacc 3120

tccaagatcg accccagaac aggctttatc gatttcctgc accctgccta cgagaatatc 3180

gagaaggcca agcagtggat caacaagttt gattccatca ggttcaattc taagatggac 3240

tggtttgagt tcaccgccga tacacgcaag ttttccgaga acctgatgct gggcaagaat 3300

cgggtgtggg tcatctgcac cacaaatgtg gagcggtact tcaccagcaa gaccgccaac 3360

agctccatcc agtacaatag catccagatc accgagaagc tgaaggagct gtttgtggac 3420

atccctttca gcaacggcca ggatctgaag ccagagatcc tgaggaagaa tgacgccgtg 3480

ttctttaaga gcctgctgtt ttacatcaag accacactgt ccctgcgcca gaacaatggc 3540

aagaagggcg aggaggagaa ggacttcatc ctgagcccag tggtggattc caagggccgg 3600

ttctttaact ctctggaggc cagcgacgat gagcccaagg acgccgatgc caatggcgcc 3660

taccacatcg ccctgaaggg cctgatgaac ctgctggtgc tgaatgagac aaaggaggag 3720

aacctgagca gaccaaagtg gaagatcaag aataaggact ggctggagtt cgtgtgggag 3780

aggaaccgca aaaggccggc ggccacgaaa aaggccggcc aggcaaaaaa gaaaaaggga 3840

tcctacccat acgatgttcc agattacgct tatccctacg acgtgcctga ttatgcatac 3900

ccatatgatg tccccgacta tgcctaagaa ttc 3933

<210> 226
<211> 3828
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 226
atgagcaagc tggagaagtt tacaaactgc tactccctgt ctaagaccct gaggttcaag 60

gccatccctg tgggcaagac ccaggagaac atcgacaata agcggctgct ggtggaggac 120

gagaagagag ccgaggatta taagggcgtg aagaagctgc tggatcgcta ctatctgtct 180

tttatcaacg acgtgctgca cagcatcaag ctgaagaatc tgaacaatta catcagcctg 240

ttccggaaga aaaccagaac cgagaaggag aataaggagc tggagaacct ggagatcaat 300

ctgcggaagg agatcgccaa ggccttcaag ggcaacgagg gctacaagtc cctgtttaag 360

aaggatatca tcgagacaat cctgccagag ttcctggacg ataaggacga gatcgccctg 420

gtgaacagct tcaatggctt taccacagcc ttcaccggct tctttgataa cagagagaat 480

atgttttccg aggaggccaa gagcacatcc atcgccttca ggtgtatcaa cgagaatctg 540

acccgctaca tctctaatat ggacatcttc gagaaggtgg acgccatctt tgataagcac 600

gaggtgcagg agatcaagga gaagatcctg aacagcgact atgatgtgga ggatttcttt 660

gagggcgagt tctttaactt tgtgctgaca caggagggca tcgacgtgta taacgccatc 720

atcggcggct tcgtgaccga gagcggcgag aagatcaagg gcctgaacga gtacatcaac 780

ctgtataatc agaaaaccaa gcagaagctg cctaagttta agccactgta taagcaggtg 840

ctgagcgatc gggagtctct gagcttctac ggcgagggct atacatccga tgaggaggtg 900

ctggaggtgt ttagaaacac cctgaacaag aacagcgaga tcttcagctc catcaagaag 960

ctggagaagc tgttcaagaa ttttgacgag tactctagcg ccggcatctt tgtgaagaac 1020

ggccccgcca tcagcacaat ctccaaggat atcttcggcg agtggaacgt gatccgggac 1080

aagtggaatg ccgagtatga cgatatccac ctgaagaaga aggccgtggt gaccgagaag 1140

tacgaggacg atcggagaaa gtccttcaag aagatcggct ccttttctct ggagcagctg 1200

caggagtacg ccgacgccga tctgtctgtg gtggagaagc tgaaggagat catcatccag 1260

aaggtggatg agatctacaa ggtgtatggc tcctctgaga agctgttcga cgccgatttt 1320

gtgctggaga agagcctgaa gaagaacgac gccgtggtgg ccatcatgaa ggacctgctg 1380

gattctgtga agagcttcga gaattacatc aaggccttct ttggcgaggg caaggagaca 1440

aacagggacg agtccttcta tggcgatttt gtgctggcct acgacatcct gctgaaggtg 1500

gaccacatct acgatgccat ccgcaattat gtgacccaga agccctactc taaggataag 1560

ttcaagctgt attttcagaa ccctcagttc atgggcggct gggacaagga taaggagaca 1620

gactatcggg ccaccatcct gagatacggc tccaagtact atctggccat catggataag 1680

aagtacgcca agtgcctgca gaagatcgac aaggacgatg tgaacggcaa ttacgagaag 1740

atcaactata agctgctgcc cggccctaat aagatgctgc caaaggtgtt cttttctaag 1800

aagtggatgg cctactataa ccccagcgag gacatccaga agatctacaa gaatggcaca 1860

ttcaagaagg gcgatatgtt taacctgaat gactgtcaca agctgatcga cttctttaag 1920

gatagcatct cccggtatcc aaagtggtcc aatgcctacg atttcaactt ttctgagaca 1980

gagaagtata aggacatcgc cggcttttac agagaggtgg aggagcaggg ctataaggtg 2040

agcttcgagt ctgccagcaa gaaggaggtg gataagctgg tggaggaggg caagctgtat 2100

atgttccaga tctataacaa ggacttttcc gataagtctc acggcacacc caatctgcac 2160

accatgtact tcaagctgct gtttgacgag aacaatcacg gacagatcag gctgagcgga 2220

ggagcagagc tgttcatgag gcgcgcctcc ctgaagaagg aggagctggt ggtgcaccca 2280

gccaactccc ctatcgccaa caagaatcca gataatccca agaaaaccac aaccctgtcc 2340

tacgacgtgt ataaggataa gaggttttct gaggaccagt acgagctgca catcccaatc 2400

gccatcaata agtgccccaa gaacatcttc aagatcaata cagaggtgcg cgtgctgctg 2460

aagcacgacg ataaccccta tgtgatcggc atcgataggg gcgagcgcaa tctgctgtat 2520

atcgtggtgg tggacggcaa gggcaacatc gtggagcagt attccctgaa cgagatcatc 2580

aacaacttca acggcatcag gatcaagaca gattaccact ctctgctgga caagaaggag 2640

aaggagaggt tcgaggcccg ccagaactgg acctccatcg agaatatcaa ggagctgaag 2700

gccggctata tctctcaggt ggtgcacaag atctgcgagc tggtggagaa gtacgatgcc 2760

gtgatcgccc tggaggacct gaactctggc tttaagaata gccgcgtgaa ggtggagaag 2820

caggtgtatc agaagttcga gaagatgctg atcgataagc tgaactacat ggtggacaag 2880

aagtctaatc cttgtgcaac aggcggcgcc ctgaagggct atcagatcac caataagttc 2940

gagagcttta agtccatgtc tacccagaac ggcttcatct tttacatccc tgcctggctg 3000

acatccaaga tcgatccatc taccggcttt gtgaacctgc tgaaaaccaa gtataccagc 3060

atcgccgatt ccaagaagtt catcagctcc tttgacagga tcatgtacgt gcccgaggag 3120

gatctgttcg agtttgccct ggactataag aacttctctc gcacagacgc cgattacatc 3180

aagaagtgga agctgtactc ctacggcaac cggatcagaa tcttccggaa tcctaagaag 3240

aacaacgtgt tcgactggga ggaggtgtgc ctgaccagcg cctataagga gctgttcaac 3300

aagtacggca tcaattatca gcagggcgat atcagagccc tgctgtgcga gcagtccgac 3360

aaggccttct actctagctt tatggccctg atgagcctga tgctgcagat gcggaacagc 3420

atcacaggcc gcaccgacgt ggattttctg atcagccctg tgaagaactc cgacggcatc 3480

ttctacgata gccggaacta tgaggcccag gagaatgcca tcctgccaaa gaacgccgac 3540

gccaatggcg cctataacat cgccagaaag gtgctgtggg ccatcggcca gttcaagaag 3600

gccgaggacg agaagctgga taaggtgaag atcgccatct ctaacaagga gtggctggag 3660

tacgcccaga ccagcgtgaa gcacaaaagg ccggcggcca cgaaaaaggc cggccaggca 3720

aaaaagaaaa agggatccta cccatacgat gttccagatt acgcttatcc ctacgacgtg 3780

cctgattatg catacccata tgatgtcccc gactatgcct aagaattc 3828

<210> 227
<211> 3924
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 227
atggacagcc tgaaggattt caccaacctg taccccgtgt ccaagacact gcggtttgag 60

ctgaagcctg tgggcaagac cctggagaat atcgagaagg ccggcatcct gaaggaggat 120

gagcacagag ccgagagcta ccggagagtg aagaagatca tcgatacata tcacaaggtg 180

ttcatcgaca gctccctgga gaacatggcc aagatgggca tcgagaatga gatcaaggcc 240

atgctgcagt ccttttgcga gctgtataag aaggaccaca ggaccgaggg agaggacaag 300

gccctggata agatcagggc cgtgctgagg ggcctgatcg tgggagcctt caccggcgtg 360

tgcggccggc gggagaacac agtgcagaat gagaagtatg agagcctgtt taaggagaag 420

ctgatcaagg agatcctgcc agatttcgtg ctgtctacag aggccgagtc cctgcccttt 480

tctgtggagg aggccaccag aagcctgaag gagttcgact cctttacatc ttacttcgcc 540

ggcttttatg agaaccggaa gaatatctac tctaccaagc cccagagcac agccatcgcc 600

tatagactga tccacgagaa cctgcctaag ttcatcgata atatcctggt gtttcagaag 660

atcaaggagc caatcgccaa ggagctggag cacatcaggg cagacttcag cgccggcggc 720

tacatcaaga aggatgagcg cctggaggac atcttttccc tgaactacta tatccacgtg 780

ctgtctcagg ccggcatcga gaagtacaat gccctgatcg gcaagatcgt gaccgagggc 840

gatggcgaga tgaagggcct gaacgagcac atcaacctgt ataatcagca gaggggccgc 900

gaggaccggc tgccactgtt cagacccctg tataagcaga tcctgtctga tagggagcag 960

ctgtcctatc tgccagagtc tttcgagaag gacgaggagc tgctgagggc cctgaaggag 1020

ttttacgatc acatcgcaga ggacatcctg ggaaggaccc agcagctgat gacaagcatc 1080

tccgagtacg atctgtcccg gatctatgtg agaaacgata gccagctgac cgacatctcc 1140

aagaagatgc tgggcgattg gaatgccatc tacatggccc gggagagagc ctatgaccac 1200

gagcaggccc ccaagcgcat cacagccaag tacgagaggg accgcatcaa ggccctgaag 1260

ggcgaggagt ctatcagcct ggccaacctg aacagctgca tcgccttcct ggacaacgtg 1320

agggattgtc gcgtggacac ctatctgtct acactgggac agaaggaggg acctcacggc 1380

ctgagcaacc tggtggagaa cgtgttcgcc tcctaccacg aggccgagca gctgctgtct 1440

tttccctatc ctgaggagaa caatctgatc caggacaagg ataacgtggt gctgatcaag 1500

aacctgctgg ataatatcag cgacctgcag aggttcctga agccactgtg gggcatgggc 1560

gatgagcccg acaaggatga gaggttttac ggcgagtaca attatatcag gggcgccctg 1620

gaccaggtca tccctctgta taacaaggtg cggaattatc tgacccgcaa gccatactcc 1680

acacgcaagg tgaagctgaa cttcggcaat agccagctgc tgtccggctg ggataggaac 1740

aaggagaagg acaattcttg cgtgatcctg cgcaagggcc agaacttcta cctggccatc 1800

atgaacaatc ggcacaagcg gagcttcgag aataagatgc tgcccgagta taaggagggc 1860

gagccttact tcgagaagat ggattataag tttctgccag accccaacaa gatgctgccc 1920

aaggtgttcc tgtctaagaa gggcatcgag atctacaagc ctagcccaaa gctgctggag 1980

cagtatggcc acggcaccca caagaagggc gataccttca gcatggacga tctgcacgag 2040

ctgatcgact tctttaagca ctccatcgag gcccacgagg attggaagca gttcggcttt 2100

aagttcagcg acaccgccac atacgagaac gtgagcagct tctaccggga ggtggaggac 2160

cagggctaca agctgtcttt tagaaaggtg tccgagtctt acgtgtatag cctgatcgat 2220

cagggcaagc tgtacctgtt ccagatctat aacaaggact ttagcccttg ttccaagggc 2280

accccaaatc tgcacacact gtactggcgg atgctgttcg atgagagaaa cctggccgac 2340

gtgatctata agctggatgg caaggccgag atcttctttc gggagaagtc cctgaagaat 2400

gaccacccaa cccaccctgc aggcaagccc atcaagaaga agagccggca gaagaagggc 2460

gaggagagcc tgttcgagta cgatctggtg aaggaccgga gatataccat ggataagttt 2520

cagttccacg tgccaatcac aatgaacttt aagtgctctg ccggcagcaa ggtgaacgac 2580

atggtgaatg cccacatcag ggaggccaag gacatgcacg tgatcggcat cgataggggc 2640

gagcgcaatc tgctgtatat ctgcgtgatc gacagccgcg gcaccatcct ggatcagatc 2700

tccctgaaca caatcaatga catcgattat cacgatctgc tggagtccag ggacaaggat 2760

cgccagcagg agcacaggaa ctggcagacc atcgagggca tcaaggagct gaagcagggc 2820

tacctgtctc aggccgtgca ccgcatcgcc gagctgatgg tggcctataa ggccgtggtg 2880

gccctggagg acctgaacat gggcttcaag cggggcagac agaaggtgga gagcagcgtg 2940

taccagcagt ttgagaagca gctgatcgac aagctgaatt atctggtgga taagaagaag 3000

cggcccgagg acatcggagg cctgctgaga gcctaccagt tcaccgcccc tttcaagagc 3060

tttaaggaga tgggcaagca gaacggcttt ctgttctata tccctgcctg gaacacatcc 3120

aatatcgacc caaccacagg cttcgtgaac ctgtttcacg tgcagtacga gaatgtggat 3180

aaggccaaga gcttctttca gaagttcgac agcatctcct acaaccctaa gaaggattgg 3240

tttgagttcg cctttgacta taagaacttc accaagaagg ccgagggctc taggagcatg 3300

tggattctgt gcacccacgg ctcccggatc aagaacttca gaaattctca gaagaatggc 3360

cagtgggata gcgaggagtt tgccctgacc gaggccttca agtccctgtt tgtgcggtac 3420

gagatcgatt ataccgccga cctgaaaacc gccatcgtgg acgagaagca gaaggatttc 3480

tttgtggacc tgctgaagct gttcaagctg accgtgcaga tgagaaactc ctggaaggag 3540

aaggacctgg attacctgat ctctccagtg gccggcgccg atggcaggtt ctttgacaca 3600

cgcgagggca ataagagcct gcccaaggac gcagatgcaa acggagccta taatatcgcc 3660

ctgaagggcc tgtgggcact gaggcagatc agacagacct ccgagggcgg caagctgaag 3720

ctggccatct ctaacaagga gtggctgcag tttgtgcagg agagatccta cgagaaggac 3780

aaaaggccgg cggccacgaa aaaggccggc caggcaaaaa agaaaaaggg atcctaccca 3840

tacgatgttc cagattacgc ttatccctac gacgtgcctg attatgcata cccatatgat 3900

gtccccgact atgcctaaga attc 3924

<210> 228
<211> 4113
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 228
atggagaact atcaggagtt caccaacctg tttcagctga ataagacact gagattcgag 60

ctgaagccca tcggcaagac ctgcgagctg ctggaggagg gcaagatctt cgccagcggc 120

tcctttctgg agaaggacaa ggtgagggcc gataacgtga gctacgtgaa gaaggagatc 180

gacaagaagc acaagatctt tatcgaggag acactgagct ccttctctat cagcaacgat 240

ctgctgaagc agtactttga ctgctataat gagctgaagg ccttcaagaa ggactgtaag 300

agcgatgagg aggaggtgaa gaaaaccgcc ctgcgcaaca agtgtacctc catccagagg 360

gccatgcgcg aggccatctc tcaggccttt ctgaagagcc cccagaagaa gctgctggcc 420

atcaagaacc tgatcgagaa cgtgttcaag gccgacgaga atgtgcagca cttctccgag 480

tttaccagct atttctccgg ctttgagaca aacagagaga atttctactc tgacgaggag 540

aagtccacat ctatcgccta taggctggtg cacgataacc tgcctatctt catcaagaac 600

atctacatct tcgagaagct gaaggagcag ttcgacgcca agaccctgag cgagatcttc 660

gagaactaca agctgtatgt ggccggctct agcctggatg aggtgttctc cctggagtac 720

tttaacaata ccctgacaca gaagggcatc gacaactata atgccgtgat cggcaagatc 780

gtgaaggagg ataagcagga gatccagggc ctgaacgagc acatcaacct gtataatcag 840

aagcacaagg accggagact gcccttcttt atctccctga agaagcagat cctgtccgat 900

cgggaggccc tgtcttggct gcctgacatg ttcaagaatg attctgaagt gatcaaggcc 960

ctgaagggct tctacatcga ggacggcttt gagaacaatg tgctgacacc tctggccacc 1020

ctgctgtcct ctctggataa gtacaacctg aatggcatct ttatccgcaa caatgaggcc 1080

ctgagctccc tgtcccagaa cgtgtatcgg aatttttcta tcgacgaggc catcgatgcc 1140

aacgccgagc tgcagacctt caacaattac gagctgatcg ccaatgccct gcgcgccaag 1200

atcaagaagg agacaaagca gggccggaag tctttcgaga agtacgagga gtatatcgat 1260

aagaaggtga aggccatcga cagcctgtcc atccaggaga tcaacgagct ggtggagaat 1320

tacgtgagcg agtttaactc taatagcggc aacatgccaa gaaaggtgga ggactacttc 1380

agcctgatga ggaagggcga cttcggctcc aacgatctga tcgaaaatat caagaccaag 1440

ctgagcgccg cagagaagct gctgggcaca aagtaccagg agacagccaa ggacatcttc 1500

aagaaggatg agaactccaa gctgatcaag gagctgctgg acgccaccaa gcagttccag 1560

cactttatca agccactgct gggcacaggc gaggaggcag atcgggacct ggtgttctac 1620

ggcgattttc tgcccctgta tgagaagttt gaggagctga ccctgctgta taacaaggtg 1680

cggaatagac tgacacagaa gccctattcc aaggacaaga tccgcctgtg cttcaacaag 1740

cctaagctga tgacaggctg ggtggattcc aagaccgaga agtctgacaa cggcacacag 1800

tacggcggct atctgtttcg gaagaagaat gagatcggcg agtacgatta ttttctgggc 1860

atctctagca aggcccagct gttcagaaag aacgaggccg tgatcggcga ctacgagagg 1920

ctggattact atcagccaaa ggccaatacc atctacggct ctgcctatga gggcgagaac 1980

agctacaagg aggacaagaa gcggctgaac aaagtgatca tcgcctatat cgagcagatc 2040

aagcagacaa acatcaagaa gtctatcatc gagtccatct ctaagtatcc taatatcagc 2100

gacgatgaca aggtgacccc atcctctctg ctggagaaga tcaagaaggt gtctatcgac 2160

agctacaacg gcatcctgtc cttcaagtct tttcagagcg tgaacaagga agtgatcgat 2220

aacctgctga aaaccatcag ccccctgaag aacaaggccg agtttctgga cctgatcaat 2280

aaggattatc agatcttcac cgaggtgcag gccgtgatcg acgagatctg caagcagaaa 2340

accttcatct actttccaat ctccaacgtg gagctggaga aggagatggg cgataaggac 2400

aagcccctgt gcctgttcca gatcagcaat aaggatctgt ccttcgccaa gacctttagc 2460

gccaacctgc ggaagaagag aggcgccgag aatctgcaca caatgctgtt taaggccctg 2520

atggagggca accaggataa tctggacctg ggctctggcg ccatcttcta cagagccaag 2580

agcctggacg gcaacaagcc cacacaccct gccaatgagg ccatcaagtg taggaacgtg 2640

gccaataagg ataaggtgtc cctgttcacc tacgacatct ataagaacag gcgctacatg 2700

gagaataagt tcctgtttca cctgagcatc gtgcagaact ataaggccgc caatgactcc 2760

gcccagctga acagctccgc caccgagtat atcagaaagg ccgatgacct gcacatcatc 2820

ggcatcgata ggggcgagcg caatctgctg tactattccg tgatcgatat gaagggcaac 2880

atcgtggagc aggactctct gaatatcatc aggaacaatg acctggagac agattaccac 2940

gacctgctgg ataagaggga gaaggagcgc aaggccaacc ggcagaattg ggaggccgtg 3000

gagggcatca aggacctgaa gaagggctac ctgagccagg ccgtgcacca gatcgcccag 3060

ctgatgctga agtataacgc catcatcgcc ctggaggatc tgggccagat gtttgtgacc 3120

cgcggccaga agatcgagaa ggccgtgtac cagcagttcg agaagagcct ggtggataag 3180

ctgtcctacc tggtggacaa gaagcggcct tataatgagc tgggcggcat cctgaaggcc 3240

taccagctgg cctctagcat caccaagaac aattctgaca agcagaacgg cttcctgttt 3300

tatgtgccag cctggaatac aagcaagatc gatcccgtga ccggctttac agacctgctg 3360

cggcccaagg ccatgaccat caaggaggcc caggacttct ttggcgcctt cgataacatc 3420

tcttacaatg acaagggcta tttcgagttt gagacaaact acgacaagtt taagatcaga 3480

atgaagagcg cccagaccag gtggacaatc tgcaccttcg gcaatcggat caagagaaag 3540

aaggataaga actactggaa ttatgaggag gtggagctga ccgaggagtt caagaagctg 3600

tttaaggaca gcaacatcga ttacgagaac tgtaatctga aggaggagat ccagaacaag 3660

gacaatcgca agttctttga tgacctgatc aagctgctgc agctgacact gcagatgcgg 3720

aactccgatg acaagggcaa tgattatatc atctctcctg tggccaacgc cgagggccag 3780

ttctttgact cccgcaatgg cgataagaag ctgccactgg atgcagacgc aaacggagcc 3840

tacaatatcg cccgcaaggg cctgtggaac atccggcaga tcaagcagac caagaacgac 3900

aagaagctga atctgagcat ctcctctaca gagtggctgg atttcgtgcg ggagaagcct 3960

tacctgaaga aaaggccggc ggccacgaaa aaggccggcc aggcaaaaaa gaaaaaggga 4020

tcctacccat acgatgttcc agattacgct tatccctacg acgtgcctga ttatgcatac 4080

ccatatgatg tccccgacta tgcctaagaa ttc 4113

<210> 229
<211> 3882
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 229
atgaaaaccc agcacttctt tgaggacttc acaagcctgt actctctgag caagaccatc 60

cggtttgagc tgaagccaat cggcaagacc ctggagaaca tcaagaagaa tggcctgatc 120

cggagagatg agcagagact ggacgattac gagaagctga agaaagtgat cgacgagtat 180

cacgaggatt tcatcgccaa catcctgagc tccttttcct tctctgagga gatcctgcag 240

tcctacatcc agaatctgag cgagtccgag gccagggcca agatcgagaa aaccatgcgc 300

gacacactgg ccaaggcctt ctctgaggat gagaggtaca agagcatctt taagaaggag 360

ctggtgaaga aggacatccc cgtgtggtgc cctgcctata agagcctgtg caagaagttc 420

gataacttta ccacatctct ggtgcccttc cacgagaaca ggaagaacct gtataccagc 480

aatgagatca cagcctctat cccttatcgc atcgtgcacg tgaacctgcc aaagtttatc 540

cagaatatcg aggccctgtg cgagctgcag aagaagatgg gcgccgacct gtacctggag 600

atgatggaga acctgcgcaa cgtgtggccc agcttcgtga aaaccccaga cgacctgtgc 660

aacctgaaaa cctataatca cctgatggtg cagtctagca tcagcgagta caacaggttt 720

gtgggcggct attccaccga ggacggcaca aagcaccagg gcatcaacga gtggatcaat 780

atctacagac agaggaataa ggagatgcgc ctgcctggcc tggtgttcct gcacaagcag 840

atcctggcca aggtggactc ctctagcttc atcagcgata cactggagaa cgacgatcag 900

gtgttttgcg tgctgagaca gttcaggaag ctgttttgga ataccgtgtc ctctaaggag 960

gacgatgccg cctccctgaa ggacctgttc tgtggcctgt ctggctatga ccctgaggcc 1020

atctacgtga gcgatgccca cctggccaca atctccaaga acatctttga cagatggaat 1080

tacatctccg atgccatcag gcgcaagacc gaggtgctga tgccacggaa gaaggagagc 1140

gtggagagat atgccgagaa gatctccaag cagatcaaga agagacagtc ttacagcctg 1200

gccgagctgg acgatctgct ggcccactat agcgaggagt ccctgcccgc aggcttctct 1260

ctgctgagct actttacatc tctgggcggc cagaagtatc tggtgagcga cggcgaagtg 1320

atcctgtacg aggagggcag caacatctgg gacgaggtgc tgatcgcctt cagggatctg 1380

caggtcatcc tggacaagga cttcaccgag aagaagctgg gcaaggatga ggaggccgtg 1440

tctgtgatca agaaggccct ggacagcgcc ctgcgcctgc ggaagttctt tgatctgctg 1500

tccggcacag gcgcagagat caggagagac agctccttct atgccctgta taccgaccgg 1560

atggataagc tgaagggcct gctgaagatg tatgataagg tgagaaacta cctgaccaag 1620

aagccttatt ccatcgagaa gttcaagctg cactttgaca acccatccct gctgtctggc 1680

tgggataaga ataaggagct gaacaatctg tctgtgatct tccggcagaa cggctactat 1740

tacctgggca tcatgacacc caagggcaag aatctgttca agaccctgcc taagctgggc 1800

gccgaggaga tgttttatga gaagatggag tacaagcaga tcgccgagcc tatgctgatg 1860

ctgccaaagg tgttctttcc caagaaaacc aagccagcct tcgccccaga ccagagcgtg 1920

gtggatatct acaacaagaa aaccttcaag acaggccaga agggctttaa taagaaggac 1980

ctgtaccggc tgatcgactt ctacaaggag gccctgacag tgcacgagtg gaagctgttt 2040

aacttctcct tttctccaac cgagcagtat cggaatatcg gcgagttctt tgacgaggtg 2100

agagagcagg cctacaaggt gtccatggtg aacgtgcccg cctcttatat cgacgaggcc 2160

gtggagaacg gcaagctgta tctgttccag atctacaata aggacttcag cccctactcc 2220

aagggcatcc ctaacctgca cacactgtat tggaaggccc tgttcagcga gcagaatcag 2280

agccgggtgt ataagctgtg cggaggagga gagctgtttt atagaaaggc cagcctgcac 2340

atgcaggaca ccacagtgca ccccaagggc atctctatcc acaagaagaa cctgaataag 2400

aagggcgaga caagcctgtt caactacgac ctggtgaagg ataagaggtt taccgaggac 2460

aagttctttt tccacgtgcc tatctctatc aactacaaga ataagaagat caccaacgtg 2520

aatcagatgg tgcgcgatta tatcgcccag aacgacgatc tgcagatcat cggcatcgac 2580

cgcggcgagc ggaatctgct gtatatcagc cggatcgata caaggggcaa cctgctggag 2640

cagttcagcc tgaatgtgat cgagtccgac aagggcgatc tgagaaccga ctatcagaag 2700

atcctgggcg atcgcgagca ggagcggctg aggcgccggc aggagtggaa gtctatcgag 2760

agcatcaagg acctgaagga tggctacatg agccaggtgg tgcacaagat ctgtaacatg 2820

gtggtggagc acaaggccat cgtggtgctg gagaacctga atctgagctt catgaagggc 2880

aggaagaagg tggagaagtc cgtgtacgag aagtttgagc gcatgctggt ggacaagctg 2940

aactatctgg tggtggataa gaagaacctg tccaatgagc caggaggcct gtatgcagca 3000

taccagctga ccaatccact gttctctttt gaggagctgc acagataccc ccagagcggc 3060

atcctgtttt tcgtggaccc atggaacacc tctctgacag atcccagcac aggcttcgtg 3120

aatctgctgg gcagaatcaa ctacaccaat gtgggcgacg cccgcaagtt tttcgatcgg 3180

tttaacgcca tcagatatga cggcaagggc aatatcctgt tcgacctgga tctgtccaga 3240

tttgatgtga gggtggagac acagaggaag ctgtggacac tgaccacatt cggctctcgc 3300

atcgccaaat ccaagaagtc tggcaagtgg atggtggagc ggatcgagaa cctgagcctg 3360

tgctttctgg agctgttcga gcagtttaat atcggctaca gagtggagaa ggacctgaag 3420

aaggccatcc tgagccagga taggaaggag ttctatgtgc gcctgatcta cctgtttaac 3480

ctgatgatgc agatccggaa cagcgacggc gaggaggatt atatcctgtc tcccgccctg 3540

aacgagaaga atctgcagtt cgacagcagg ctgatcgagg ccaaggatct gcctgtggac 3600

gcagatgcaa acggagcata caatgtggcc cgcaagggcc tgatggtggt gcagagaatc 3660

aagaggggcg accacgagtc catccacagg atcggaaggg cacagtggct gagatatgtg 3720

caggagggca tcgtggagaa aaggccggcg gccacgaaaa aggccggcca ggcaaaaaag 3780

aaaaagggat cctacccata cgatgttcca gattacgctt atccctacga cgtgcctgat 3840

tatgcatacc catatgatgt ccccgactat gcctaagaat tc 3882

<210> 230
<211> 1345
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 230
Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys

```
            20                  25                  30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
        35                  40                  45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
    50                  55                  60

Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65                  70                  75                  80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
                85                  90                  95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
            100                 105                 110

Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
        115                 120                 125

Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
    130                 135                 140

Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145                 150                 155                 160

Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
                165                 170                 175

Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
            180                 185                 190

Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
        195                 200                 205

Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
    210                 215                 220

Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225                 230                 235                 240

Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
                245                 250                 255

Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
            260                 265                 270

Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
        275                 280                 285

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
    290                 295                 300

Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305                 310                 315                 320

Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
                325                 330                 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
            340                 345                 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
```

355 360 365

Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
370 375 380

Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385 390 395 400

Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
405 410 415

Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
420 425 430

Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
435 440 445

Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
450 455 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465 470 475 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
485 490 495

Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
500 505 510

Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
515 520 525

Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
530 535 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545 550 555 560

Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
565 570 575

Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
580 585 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
595 600 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
610 615 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625 630 635 640

Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
645 650 655

Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
660 665 670

Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
675 680 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln

690 695 700

Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
705 710 715 720

Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp
725 730 735

Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
740 745 750

Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
755 760 765

Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
770 775 780

Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
785 790 795 800

Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
805 810 815

Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
820 825 830

Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
835 840 845

Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu
850 855 860

Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
865 870 875 880

His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
885 890 895

Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
900 905 910

Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
915 920 925

Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
930 935 940

Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945 950 955 960

Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
965 970 975

Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
980 985 990

Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu
995 1000 1005

Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val
1010 1015 1020

Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu

1025 1030 1035

Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg
1040 1045 1050

Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly
1055 1060 1065

Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser
1070 1075 1080

Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys
1085 1090 1095

Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp
1100 1105 1110

Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe
1115 1120 1125

Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr
1130 1135 1140

Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp
1145 1150 1155

Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu
1160 1165 1170

Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly
1175 1180 1185

Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe
1190 1195 1200

Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg
1205 1210 1215

Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val
1220 1225 1230

Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1235 1240 1245

Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly
1250 1255 1260

Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu
1265 1270 1275

Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1280 1285 1290

Phe Val Gln Asn Arg Asn Asn Lys Arg Pro Ala Ala Thr Lys Lys
1295 1300 1305

Ala Gly Gln Ala Lys Lys Lys Lys Gly Ser Tyr Pro Tyr Asp Val
1310 1315 1320

Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro
1325 1330 1335

Tyr Asp Val Pro Asp Tyr Ala

1340 1345

<210> 231
<211> 1278
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 231
Met Asp Tyr Gly Asn Gly Gln Phe Glu Arg Arg Ala Pro Leu Thr Lys
1 5 10 15

Thr Ile Thr Leu Arg Leu Lys Pro Ile Gly Glu Thr Arg Glu Thr Ile
20 25 30

Arg Glu Gln Lys Leu Leu Glu Gln Asp Ala Ala Phe Arg Lys Leu Val
35 40 45

Glu Thr Val Thr Pro Ile Val Asp Asp Cys Ile Arg Lys Ile Ala Asp
50 55 60

Asn Ala Leu Cys His Phe Gly Thr Glu Tyr Asp Phe Ser Cys Leu Gly
65 70 75 80

Asn Ala Ile Ser Lys Asn Asp Ser Lys Ala Ile Lys Lys Glu Thr Glu
85 90 95

Lys Val Glu Lys Leu Leu Ala Lys Val Leu Thr Glu Asn Leu Pro Asp
100 105 110

Gly Leu Arg Lys Val Asn Asp Ile Asn Ser Ala Ala Phe Ile Gln Asp
115 120 125

Thr Leu Thr Ser Phe Val Gln Asp Asp Ala Asp Lys Arg Val Leu Ile
130 135 140

Gln Glu Leu Lys Gly Lys Thr Val Leu Met Gln Arg Phe Leu Thr Thr
145 150 155 160

Arg Ile Thr Ala Leu Thr Val Trp Leu Pro Asp Arg Val Phe Glu Asn
165 170 175

Phe Asn Ile Phe Ile Glu Asn Ala Glu Lys Met Arg Ile Leu Leu Asp
180 185 190

Ser Pro Leu Asn Glu Lys Ile Met Lys Phe Asp Pro Asp Ala Glu Gln
195 200 205

Tyr Ala Ser Leu Glu Phe Tyr Gly Gln Cys Leu Ser Gln Lys Asp Ile
210 215 220

Asp Ser Tyr Asn Leu Ile Ile Ser Gly Ile Tyr Ala Asp Asp Glu Val
225 230 235 240

Lys Asn Pro Gly Ile Asn Glu Ile Val Lys Glu Tyr Asn Gln Gln Ile
245 250 255

Arg Gly Asp Lys Asp Glu Ser Pro Leu Pro Lys Leu Lys Lys Leu His
260 265 270

Lys Gln Ile Leu Met Pro Val Glu Lys Ala Phe Phe Val Arg Val Leu
275 280 285

```
Ser Asn Asp Ser Asp Ala Arg Ser Ile Leu Glu Lys Ile Leu Lys Asp
    290                 295                 300

Thr Glu Met Leu Pro Ser Lys Ile Ile Glu Ala Met Lys Glu Ala Asp
305                 310                 315                 320

Ala Gly Asp Ile Ala Val Tyr Gly Ser Arg Leu His Glu Leu Ser His
                325                 330                 335

Val Ile Tyr Gly Asp His Gly Lys Leu Ser Gln Ile Ile Tyr Asp Lys
            340                 345                 350

Glu Ser Lys Arg Ile Ser Glu Leu Met Glu Thr Leu Ser Pro Lys Glu
        355                 360                 365

Arg Lys Glu Ser Lys Lys Arg Leu Glu Gly Leu Glu Glu His Ile Arg
    370                 375                 380

Lys Ser Thr Tyr Thr Phe Asp Glu Leu Asn Arg Tyr Ala Glu Lys Asn
385                 390                 395                 400

Val Met Ala Ala Tyr Ile Ala Ala Val Glu Glu Ser Cys Ala Glu Ile
                405                 410                 415

Met Arg Lys Glu Lys Asp Leu Arg Thr Leu Leu Ser Lys Glu Asp Val
            420                 425                 430

Lys Ile Arg Gly Asn Arg His Asn Thr Leu Ile Val Lys Asn Tyr Phe
        435                 440                 445

Asn Ala Trp Thr Val Phe Arg Asn Leu Ile Arg Ile Leu Arg Arg Lys
    450                 455                 460

Ser Glu Ala Glu Ile Asp Ser Asp Phe Tyr Asp Val Leu Asp Asp Ser
465                 470                 475                 480

Val Glu Val Leu Ser Leu Thr Tyr Lys Gly Glu Asn Leu Cys Arg Ser
                485                 490                 495

Tyr Ile Thr Lys Lys Ile Gly Ser Asp Leu Lys Pro Glu Ile Ala Thr
            500                 505                 510

Tyr Gly Ser Ala Leu Arg Pro Asn Ser Arg Trp Trp Ser Pro Gly Glu
        515                 520                 525

Lys Phe Asn Val Lys Phe His Thr Ile Val Arg Arg Asp Gly Arg Leu
    530                 535                 540

Tyr Tyr Phe Ile Leu Pro Lys Gly Ala Lys Pro Val Glu Leu Glu Asp
545                 550                 555                 560

Met Asp Gly Asp Ile Glu Cys Leu Gln Met Arg Lys Ile Pro Asn Pro
                565                 570                 575

Thr Ile Phe Leu Pro Lys Leu Val Phe Lys Asp Pro Glu Ala Phe Phe
            580                 585                 590

Arg Asp Asn Pro Glu Ala Asp Glu Phe Val Phe Leu Ser Gly Met Lys
        595                 600                 605

Ala Pro Val Thr Ile Thr Arg Glu Thr Tyr Glu Ala Tyr Arg Tyr Lys
    610                 615                 620
```

```
Leu Tyr Thr Val Gly Lys Leu Arg Asp Gly Glu Val Ser Glu Glu Glu
625                 630                 635                 640

Tyr Lys Arg Ala Leu Leu Gln Val Leu Thr Ala Tyr Lys Glu Phe Leu
                645                 650                 655

Glu Asn Arg Met Ile Tyr Ala Asp Leu Asn Phe Gly Phe Lys Asp Leu
            660                 665                 670

Glu Glu Tyr Lys Asp Ser Ser Glu Phe Ile Lys Gln Val Glu Thr His
        675                 680                 685

Asn Thr Phe Met Cys Trp Ala Lys Val Ser Ser Ser Gln Leu Asp Asp
    690                 695                 700

Leu Val Lys Ser Gly Asn Gly Leu Leu Phe Glu Ile Trp Ser Glu Arg
705                 710                 715                 720

Leu Glu Ser Tyr Tyr Lys Tyr Gly Asn Glu Lys Val Leu Arg Gly Tyr
                725                 730                 735

Glu Gly Val Leu Leu Ser Ile Leu Lys Asp Glu Asn Leu Val Ser Met
            740                 745                 750

Arg Thr Leu Leu Asn Ser Arg Pro Met Leu Val Tyr Arg Pro Lys Glu
        755                 760                 765

Ser Ser Lys Pro Met Val Val His Arg Asp Gly Ser Arg Val Val Asp
    770                 775                 780

Arg Phe Asp Lys Asp Gly Lys Tyr Ile Pro Pro Glu Val His Asp Glu
785                 790                 795                 800

Leu Tyr Arg Phe Phe Asn Asn Leu Leu Ile Lys Glu Lys Leu Gly Glu
                805                 810                 815

Lys Ala Arg Lys Ile Leu Asp Asn Lys Lys Val Lys Val Lys Val Leu
            820                 825                 830

Glu Ser Glu Arg Val Lys Trp Ser Lys Phe Tyr Asp Glu Gln Phe Ala
        835                 840                 845

Val Thr Phe Ser Val Lys Lys Asn Ala Asp Cys Leu Asp Thr Thr Lys
    850                 855                 860

Asp Leu Asn Ala Glu Val Met Glu Gln Tyr Ser Glu Ser Asn Arg Leu
865                 870                 875                 880

Ile Leu Ile Arg Asn Thr Thr Asp Ile Leu Tyr Tyr Leu Val Leu Asp
                885                 890                 895

Lys Asn Gly Lys Val Leu Lys Gln Arg Ser Leu Asn Ile Ile Asn Asp
            900                 905                 910

Gly Ala Arg Asp Val Asp Trp Lys Glu Arg Phe Arg Gln Val Thr Lys
        915                 920                 925

Asp Arg Asn Glu Gly Tyr Asn Glu Trp Asp Tyr Ser Arg Thr Ser Asn
    930                 935                 940

Asp Leu Lys Glu Val Tyr Leu Asn Tyr Ala Leu Lys Glu Ile Ala Glu
945                 950                 955                 960
```

Ala Val Ile Glu Tyr Asn Ala Ile Leu Ile Ile Glu Lys Met Ser Asn
965 970 975

Ala Phe Lys Asp Lys Tyr Ser Phe Leu Asp Asp Val Thr Phe Lys Gly
980 985 990

Phe Glu Thr Lys Leu Leu Ala Lys Leu Ser Asp Leu His Phe Arg Gly
995 1000 1005

Ile Lys Asp Gly Glu Pro Cys Ser Phe Thr Asn Pro Leu Gln Leu
1010 1015 1020

Cys Gln Asn Asp Ser Asn Lys Ile Leu Gln Asp Gly Val Ile Phe
1025 1030 1035

Met Val Pro Asn Ser Met Thr Arg Ser Leu Asp Pro Asp Thr Gly
1040 1045 1050

Phe Ile Phe Ala Ile Asn Asp His Asn Ile Arg Thr Lys Lys Ala
1055 1060 1065

Lys Leu Asn Phe Leu Ser Lys Phe Asp Gln Leu Lys Val Ser Ser
1070 1075 1080

Glu Gly Cys Leu Ile Met Lys Tyr Ser Gly Asp Ser Leu Pro Thr
1085 1090 1095

His Asn Thr Asp Asn Arg Val Trp Asn Cys Cys Cys Asn His Pro
1100 1105 1110

Ile Thr Asn Tyr Asp Arg Glu Thr Lys Lys Val Glu Phe Ile Glu
1115 1120 1125

Glu Pro Val Glu Glu Leu Ser Arg Val Leu Glu Glu Asn Gly Ile
1130 1135 1140

Glu Thr Asp Thr Glu Leu Asn Lys Leu Asn Glu Arg Glu Asn Val
1145 1150 1155

Pro Gly Lys Val Val Asp Ala Ile Tyr Ser Leu Val Leu Asn Tyr
1160 1165 1170

Leu Arg Gly Thr Val Ser Gly Val Ala Gly Gln Arg Ala Val Tyr
1175 1180 1185

Tyr Ser Pro Val Thr Gly Lys Lys Tyr Asp Ile Ser Phe Ile Gln
1190 1195 1200

Ala Met Asn Leu Asn Arg Lys Cys Asp Tyr Tyr Arg Ile Gly Ser
1205 1210 1215

Lys Glu Arg Gly Glu Trp Thr Asp Phe Val Ala Gln Leu Ile Asn
1220 1225 1230

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys
1235 1240 1245

Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr
1250 1255 1260

Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1265 1270 1275

<210> 232
<211> 1286
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 232

```
Met Leu Leu Tyr Glu Asn Tyr Thr Lys Arg Asn Gln Ile Thr Lys Ser
1               5                  10                  15

Leu Arg Leu Glu Leu Arg Pro Gln Gly Lys Thr Leu Arg Asn Ile Lys
            20                  25                  30

Glu Leu Asn Leu Leu Glu Gln Asp Lys Ala Ile Tyr Ala Leu Leu Glu
        35                  40                  45

Arg Leu Lys Pro Val Ile Asp Glu Gly Ile Lys Asp Ile Ala Arg Asp
    50                  55                  60

Thr Leu Lys Asn Cys Glu Leu Ser Phe Glu Lys Leu Tyr Glu His Phe
65                  70                  75                  80

Leu Ser Gly Asp Lys Lys Ala Tyr Ala Lys Glu Ser Glu Arg Leu Lys
                85                  90                  95

Lys Glu Ile Val Lys Thr Leu Ile Lys Asn Leu Pro Glu Gly Ile Gly
            100                 105                 110

Lys Ile Ser Glu Ile Asn Ser Ala Lys Tyr Leu Asn Gly Val Leu Tyr
        115                 120                 125

Asp Phe Ile Asp Lys Thr His Lys Asp Ser Glu Glu Lys Gln Asn Ile
    130                 135                 140

Leu Ser Asp Ile Leu Glu Thr Lys Gly Tyr Leu Ala Leu Phe Ser Lys
145                 150                 155                 160

Phe Leu Thr Ser Arg Ile Thr Thr Leu Glu Gln Ser Met Pro Lys Arg
                165                 170                 175

Val Ile Glu Asn Phe Glu Ile Tyr Ala Ala Asn Ile Pro Lys Met Gln
            180                 185                 190

Asp Ala Leu Glu Arg Gly Ala Val Ser Phe Ala Ile Glu Tyr Glu Ser
        195                 200                 205

Ile Cys Ser Val Asp Tyr Tyr Asn Gln Ile Leu Ser Gln Glu Asp Ile
    210                 215                 220

Asp Ser Tyr Asn Arg Leu Ile Ser Gly Ile Met Asp Glu Asp Gly Ala
225                 230                 235                 240

Lys Glu Lys Gly Ile Asn Gln Thr Ile Ser Glu Lys Asn Ile Lys Ile
                245                 250                 255

Lys Ser Glu His Leu Glu Glu Lys Pro Phe Arg Ile Leu Lys Gln Leu
            260                 265                 270

His Lys Gln Ile Leu Glu Glu Arg Glu Lys Ala Phe Thr Ile Asp His
        275                 280                 285

Ile Asp Ser Asp Glu Glu Val Val Gln Val Thr Lys Glu Ala Phe Glu
```

290 295 300

Gln Thr Lys Glu Gln Trp Glu Asn Ile Lys Lys Ile Asn Gly Phe Tyr
305 310 315 320

Ala Lys Asp Pro Gly Asp Ile Thr Leu Phe Ile Val Val Gly Pro Asn
325 330 335

Gln Thr His Val Leu Ser Gln Leu Ile Tyr Gly Glu His Asp Arg Ile
340 345 350

Arg Leu Leu Leu Glu Glu Tyr Glu Lys Asn Thr Leu Glu Val Leu Pro
355 360 365

Arg Arg Thr Lys Ser Glu Lys Ala Arg Tyr Asp Lys Phe Val Asn Ala
370 375 380

Val Pro Lys Lys Val Ala Lys Glu Ser His Thr Phe Asp Gly Leu Gln
385 390 395 400

Lys Met Thr Gly Asp Asp Arg Leu Phe Ile Leu Tyr Arg Asp Glu Leu
405 410 415

Ala Arg Asn Tyr Met Arg Ile Lys Glu Ala Tyr Gly Thr Phe Glu Arg
420 425 430

Asp Ile Leu Lys Ser Arg Arg Gly Ile Lys Gly Asn Arg Asp Val Gln
435 440 445

Glu Ser Leu Val Ser Phe Tyr Asp Glu Leu Thr Lys Phe Arg Ser Ala
450 455 460

Leu Arg Ile Ile Asn Ser Gly Asn Asp Glu Lys Ala Asp Pro Ile Phe
465 470 475 480

Tyr Asn Thr Phe Asp Gly Ile Phe Glu Lys Ala Asn Arg Thr Tyr Lys
485 490 495

Ala Glu Asn Leu Cys Arg Asn Tyr Val Thr Lys Ser Pro Ala Asp Asp
500 505 510

Ala Arg Ile Met Ala Ser Cys Leu Gly Thr Pro Ala Arg Leu Arg Thr
515 520 525

His Trp Trp Asn Gly Glu Glu Asn Phe Ala Ile Asn Asp Val Ala Met
530 535 540

Ile Arg Arg Gly Asp Glu Tyr Tyr Tyr Phe Val Leu Thr Pro Asp Val
545 550 555 560

Lys Pro Val Asp Leu Lys Thr Lys Asp Glu Thr Asp Ala Gln Ile Phe
565 570 575

Val Gln Arg Lys Gly Ala Lys Ser Phe Leu Gly Leu Pro Lys Ala Leu
580 585 590

Phe Lys Cys Ile Leu Glu Pro Tyr Phe Glu Ser Pro Glu His Lys Asn
595 600 605

Asp Lys Asn Cys Val Ile Glu Glu Tyr Val Ser Lys Pro Leu Thr Ile
610 615 620

Asp Arg Arg Ala Tyr Asp Ile Phe Lys Asn Gly Thr Phe Lys Lys Thr

625 630 635 640

Asn Ile Gly Ile Asp Gly Leu Thr Glu Glu Lys Phe Lys Asp Asp Cys
645 650 655

Arg Tyr Leu Ile Asp Val Tyr Lys Glu Phe Ile Ala Val Tyr Thr Arg
660 665 670

Tyr Ser Cys Phe Asn Met Ser Gly Leu Lys Arg Ala Asp Glu Tyr Asn
675 680 685

Asp Ile Gly Glu Phe Phe Ser Asp Val Asp Thr Arg Leu Cys Thr Met
690 695 700

Glu Trp Ile Pro Val Ser Phe Glu Arg Ile Asn Asp Met Val Asp Lys
705 710 715 720

Lys Glu Gly Leu Leu Phe Leu Val Arg Ser Met Phe Leu Tyr Asn Arg
725 730 735

Pro Arg Lys Pro Tyr Glu Arg Thr Phe Ile Gln Leu Phe Ser Asp Ser
740 745 750

Asn Met Glu His Thr Ser Met Leu Leu Asn Ser Arg Ala Met Ile Gln
755 760 765

Tyr Arg Ala Ala Ser Leu Pro Arg Arg Val Thr His Lys Lys Gly Ser
770 775 780

Ile Leu Val Ala Leu Arg Asp Ser Asn Gly Glu His Ile Pro Met His
785 790 795 800

Ile Arg Glu Ala Ile Tyr Lys Met Lys Asn Asn Phe Asp Ile Ser Ser
805 810 815

Glu Asp Phe Ile Met Ala Lys Ala Tyr Leu Ala Glu His Asp Val Ala
820 825 830

Ile Lys Lys Ala Asn Glu Asp Ile Ile Arg Asn Arg Arg Tyr Thr Glu
835 840 845

Asp Lys Phe Phe Leu Ser Leu Ser Tyr Thr Lys Asn Ala Asp Ile Ser
850 855 860

Ala Arg Thr Leu Asp Tyr Ile Asn Asp Lys Val Glu Glu Asp Thr Gln
865 870 875 880

Asp Ser Arg Met Ala Val Ile Val Thr Arg Asn Leu Lys Asp Leu Thr
885 890 895

Tyr Val Ala Val Val Asp Glu Lys Asn Asn Val Leu Glu Glu Lys Ser
900 905 910

Leu Asn Glu Ile Asp Gly Val Asn Tyr Arg Glu Leu Leu Lys Glu Arg
915 920 925

Thr Lys Ile Lys Tyr His Asp Lys Thr Arg Leu Trp Gln Tyr Asp Val
930 935 940

Ser Ser Lys Gly Leu Lys Glu Ala Tyr Val Glu Leu Ala Val Thr Gln
945 950 955 960

Ile Ser Lys Leu Ala Thr Lys Tyr Asn Ala Val Val Val Val Glu Ser

965 970 975

Met Ser Ser Thr Phe Lys Asp Lys Phe Ser Phe Leu Asp Glu Gln Ile
980 985 990

Phe Lys Ala Phe Glu Ala Arg Leu Cys Ala Arg Met Ser Asp Leu Ser
995 1000 1005

Phe Asn Thr Ile Lys Glu Gly Glu Ala Gly Ser Ile Ser Asn Pro
1010 1015 1020

Ile Gln Val Ser Asn Asn Asn Gly Asn Ser Tyr Gln Asp Gly Val
1025 1030 1035

Ile Tyr Phe Leu Asn Asn Ala Tyr Thr Arg Thr Leu Cys Pro Asp
1040 1045 1050

Thr Gly Phe Val Asp Val Phe Asp Lys Thr Arg Leu Ile Thr Met
1055 1060 1065

Gln Ser Lys Arg Gln Phe Phe Ala Lys Met Lys Asp Ile Arg Ile
1070 1075 1080

Asp Asp Gly Glu Met Leu Phe Thr Phe Asn Leu Glu Glu Tyr Pro
1085 1090 1095

Thr Lys Arg Leu Leu Asp Arg Lys Glu Trp Thr Val Lys Ile Ala
1100 1105 1110

Gly Asp Gly Ser Tyr Phe Asp Lys Asp Lys Gly Glu Tyr Val Tyr
1115 1120 1125

Val Asn Asp Ile Val Arg Glu Gln Ile Ile Pro Ala Leu Leu Glu
1130 1135 1140

Asp Lys Ala Val Phe Asp Gly Asn Met Ala Glu Lys Phe Leu Asp
1145 1150 1155

Lys Thr Ala Ile Ser Gly Lys Ser Val Glu Leu Ile Tyr Lys Trp
1160 1165 1170

Phe Ala Asn Ala Leu Tyr Gly Ile Ile Thr Lys Lys Asp Gly Glu
1175 1180 1185

Lys Ile Tyr Arg Ser Pro Ile Thr Gly Thr Glu Ile Asp Val Ser
1190 1195 1200

Lys Asn Thr Thr Tyr Asn Phe Gly Lys Lys Phe Met Phe Lys Gln
1205 1210 1215

Glu Tyr Arg Gly Asp Gly Asp Phe Leu Asp Ala Phe Leu Asn Tyr
1220 1225 1230

Met Gln Ala Gln Asp Ile Ala Val Lys Arg Pro Ala Ala Thr Lys
1235 1240 1245

Lys Ala Gly Gln Ala Lys Lys Lys Lys Gly Ser Tyr Pro Tyr Asp
1250 1255 1260

Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr
1265 1270 1275

Pro Tyr Asp Val Pro Asp Tyr Ala

1280 1285

<210> 233
<211> 1522
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 233

```
Met Ser Asn Phe Phe Lys Asn Phe Thr Asn Leu Tyr Glu Leu Ser Lys
1               5                  10                  15

Thr Leu Arg Phe Glu Leu Lys Pro Val Gly Asp Thr Leu Thr Asn Met
            20                  25                  30

Lys Asp His Leu Glu Tyr Asp Glu Lys Leu Gln Thr Phe Leu Lys Asp
        35                  40                  45

Gln Asn Ile Asp Asp Ala Tyr Gln Ala Leu Lys Pro Gln Phe Asp Glu
    50                  55                  60

Ile His Glu Glu Phe Ile Thr Asp Ser Leu Glu Ser Lys Lys Ala Lys
65                  70                  75                  80

Glu Ile Asp Phe Ser Glu Tyr Leu Asp Leu Phe Gln Glu Lys Lys Glu
                85                  90                  95

Leu Asn Asp Ser Glu Lys Lys Leu Arg Asn Lys Ile Gly Glu Thr Phe
            100                 105                 110

Asn Lys Ala Gly Glu Lys Trp Lys Lys Glu Lys Tyr Pro Gln Tyr Glu
        115                 120                 125

Trp Lys Lys Gly Ser Lys Ile Ala Asn Gly Ala Asp Ile Leu Ser Cys
    130                 135                 140

Gln Asp Met Leu Gln Phe Ile Lys Tyr Lys Asn Pro Glu Asp Glu Lys
145                 150                 155                 160

Ile Lys Asn Tyr Ile Asp Asp Thr Leu Lys Gly Phe Phe Thr Tyr Phe
                165                 170                 175

Gly Gly Phe Asn Gln Asn Arg Ala Asn Tyr Tyr Glu Thr Lys Lys Glu
            180                 185                 190

Ala Ser Thr Ala Val Ala Thr Arg Ile Val His Glu Asn Leu Pro Lys
        195                 200                 205

Phe Cys Asp Asn Val Ile Gln Phe Lys His Ile Ile Lys Arg Lys Lys
    210                 215                 220

Asp Gly Thr Val Glu Lys Thr Glu Arg Lys Thr Glu Tyr Leu Asn Ala
225                 230                 235                 240

Tyr Gln Tyr Leu Lys Asn Asn Asn Lys Ile Thr Gln Ile Lys Asp Ala
                245                 250                 255

Glu Thr Glu Lys Met Ile Glu Ser Thr Pro Ile Ala Glu Lys Ile Phe
            260                 265                 270

Asp Val Tyr Tyr Phe Ser Ser Cys Leu Ser Gln Lys Gln Ile Glu Glu
        275                 280                 285
```

```
Tyr Asn Arg Ile Ile Gly His Tyr Asn Leu Leu Ile Asn Leu Tyr Asn
    290                 295                 300

Gln Ala Lys Arg Ser Glu Gly Lys His Leu Ser Ala Asn Glu Lys Lys
305                 310                 315                 320

Tyr Lys Asp Leu Pro Lys Phe Lys Thr Leu Tyr Lys Gln Ile Gly Cys
                325                 330                 335

Gly Lys Lys Lys Asp Leu Phe Tyr Thr Ile Lys Cys Asp Thr Glu Glu
            340                 345                 350

Glu Ala Asn Lys Ser Arg Asn Glu Gly Lys Glu Ser His Ser Val Glu
        355                 360                 365

Glu Ile Ile Asn Lys Ala Gln Glu Ala Ile Asn Lys Tyr Phe Lys Ser
    370                 375                 380

Asn Asn Asp Cys Glu Asn Ile Asn Thr Val Pro Asp Phe Ile Asn Tyr
385                 390                 395                 400

Ile Leu Thr Lys Glu Asn Tyr Glu Gly Val Tyr Trp Ser Lys Ala Ala
                405                 410                 415

Met Asn Thr Ile Ser Asp Lys Tyr Phe Ala Asn Tyr His Asp Leu Gln
            420                 425                 430

Asp Arg Leu Lys Glu Ala Lys Val Phe Gln Lys Ala Asp Lys Lys Ser
        435                 440                 445

Glu Asp Asp Ile Lys Ile Pro Glu Ala Ile Glu Leu Ser Gly Leu Phe
    450                 455                 460

Gly Val Leu Asp Ser Leu Ala Asp Trp Gln Thr Thr Leu Phe Lys Ser
465                 470                 475                 480

Ser Ile Leu Ser Asn Glu Asp Lys Leu Lys Ile Ile Thr Asp Ser Gln
                485                 490                 495

Thr Pro Ser Glu Ala Leu Leu Lys Met Ile Phe Asn Asp Ile Glu Lys
            500                 505                 510

Asn Met Glu Ser Phe Leu Lys Glu Thr Asn Asp Ile Ile Thr Leu Lys
        515                 520                 525

Lys Tyr Lys Gly Asn Lys Glu Gly Thr Glu Lys Ile Lys Gln Trp Phe
    530                 535                 540

Asp Tyr Thr Leu Ala Ile Asn Arg Met Leu Lys Tyr Phe Leu Val Lys
545                 550                 555                 560

Glu Asn Lys Ile Lys Gly Asn Ser Leu Asp Thr Asn Ile Ser Glu Ala
                565                 570                 575

Leu Lys Thr Leu Ile Tyr Ser Asp Asp Ala Glu Trp Phe Lys Trp Tyr
            580                 585                 590

Asp Ala Leu Arg Asn Tyr Leu Thr Gln Lys Pro Gln Asp Glu Ala Lys
        595                 600                 605

Glu Asn Lys Leu Lys Leu Asn Phe Asp Asn Pro Ser Leu Ala Gly Gly
    610                 615                 620
```

```
Trp Asp Val Asn Lys Glu Cys Ser Asn Phe Cys Val Ile Leu Lys Asp
625                 630                 635                 640

Lys Asn Glu Lys Lys Tyr Leu Ala Ile Met Lys Lys Gly Glu Asn Thr
                645                 650                 655

Leu Phe Gln Lys Glu Trp Thr Glu Gly Arg Gly Lys Asn Leu Thr Lys
            660                 665                 670

Lys Ser Asn Pro Leu Phe Glu Ile Asn Asn Cys Glu Ile Leu Ser Lys
        675                 680                 685

Met Glu Tyr Asp Phe Trp Ala Asp Val Ser Lys Met Ile Pro Lys Cys
    690                 695                 700

Ser Thr Gln Leu Lys Ala Val Val Asn His Phe Lys Gln Ser Asp Asn
705                 710                 715                 720

Glu Phe Ile Phe Pro Ile Gly Tyr Lys Val Thr Ser Gly Glu Lys Phe
                725                 730                 735

Arg Glu Glu Cys Lys Ile Ser Lys Gln Asp Phe Glu Leu Asn Asn Lys
            740                 745                 750

Val Phe Asn Lys Asn Glu Leu Ser Val Thr Ala Met Arg Tyr Asp Leu
        755                 760                 765

Ser Ser Thr Gln Glu Lys Gln Tyr Ile Lys Ala Phe Gln Lys Glu Tyr
    770                 775                 780

Trp Glu Leu Leu Phe Lys Gln Glu Lys Arg Asp Thr Lys Leu Thr Asn
785                 790                 795                 800

Asn Glu Ile Phe Asn Glu Trp Ile Asn Phe Cys Asn Lys Lys Tyr Ser
                805                 810                 815

Glu Leu Leu Ser Trp Glu Arg Lys Tyr Lys Asp Ala Leu Thr Asn Trp
            820                 825                 830

Ile Asn Phe Cys Lys Tyr Phe Leu Ser Lys Tyr Pro Lys Thr Thr Leu
        835                 840                 845

Phe Asn Tyr Ser Phe Lys Glu Ser Glu Asn Tyr Asn Ser Leu Asp Glu
    850                 855                 860

Phe Tyr Arg Asp Val Asp Ile Cys Ser Tyr Lys Leu Asn Ile Asn Thr
865                 870                 875                 880

Thr Ile Asn Lys Ser Ile Leu Asp Arg Leu Val Glu Glu Gly Lys Leu
                885                 890                 895

Tyr Leu Phe Glu Ile Lys Asn Gln Asp Ser Asn Asp Gly Lys Ser Ile
            900                 905                 910

Gly His Lys Asn Asn Leu His Thr Ile Tyr Trp Asn Ala Ile Phe Glu
        915                 920                 925

Asn Phe Asp Asn Arg Pro Lys Leu Asn Gly Glu Ala Glu Ile Phe Tyr
    930                 935                 940

Arg Lys Ala Ile Ser Lys Asp Lys Leu Gly Ile Val Lys Gly Lys Lys
945                 950                 955                 960
```

```
Thr Lys Asn Gly Thr Glu Ile Ile Lys Asn Tyr Arg Phe Ser Lys Glu
                965                 970                 975

Lys Phe Ile Leu His Val Pro Ile Thr Leu Asn Phe Cys Ser Asn Asn
            980                 985                 990

Glu Tyr Val Asn Asp Ile Val Asn Thr Lys Phe Tyr Asn Phe Ser Asn
        995                 1000                1005

Leu His  Phe Leu Gly Ile Asp  Arg Gly Glu Lys His  Leu Ala Tyr
    1010                1015                1020

Tyr Ser  Leu Val Asn Lys Asn  Gly Glu Ile Val Asp  Gln Gly Thr
    1025                1030                1035

Leu Asn  Leu Pro Phe Thr Asp  Lys Asp Gly Asn Gln  Arg Ser Ile
    1040                1045                1050

Lys Lys  Glu Lys Tyr Phe Tyr  Asn Lys Gln Glu Asp  Lys Trp Glu
    1055                1060                1065

Ala Lys  Glu Val Asp Cys Trp  Asn Tyr Asn Asp Leu  Leu Asp Ala
    1070                1075                1080

Met Ala  Ser Asn Arg Asp Met  Ala Arg Lys Asn Trp  Gln Arg Ile
    1085                1090                1095

Gly Thr  Ile Lys Glu Ala Lys  Asn Gly Tyr Val Ser  Leu Val Ile
    1100                1105                1110

Arg Lys  Ile Ala Asp Leu Ala  Val Asn Asn Glu Arg  Pro Ala Phe
    1115                1120                1125

Ile Val  Leu Glu Asp Leu Asn  Thr Gly Phe Lys Arg  Ser Arg Gln
    1130                1135                1140

Lys Ile  Asp Lys Ser Val Tyr  Gln Lys Phe Glu Leu  Ala Leu Ala
    1145                1150                1155

Lys Lys  Leu Asn Phe Leu Val  Asp Lys Asn Ala Lys  Arg Asp Glu
    1160                1165                1170

Ile Gly  Ser Pro Thr Lys Ala  Leu Gln Leu Thr Pro  Pro Val Asn
    1175                1180                1185

Asn Tyr  Gly Asp Ile Glu Asn  Lys Lys Gln Ala Gly  Ile Met Leu
    1190                1195                1200

Tyr Thr  Arg Ala Asn Tyr Thr  Ser Gln Thr Asp Pro  Ala Thr Gly
    1205                1210                1215

Trp Arg  Lys Thr Ile Tyr Leu  Lys Ala Gly Pro Glu  Glu Thr Thr
    1220                1225                1230

Tyr Lys  Lys Asp Gly Lys Ile  Lys Asn Lys Ser Val  Lys Asp Gln
    1235                1240                1245

Ile Ile  Glu Thr Phe Thr Asp  Ile Gly Phe Asp Gly  Lys Asp Tyr
    1250                1255                1260

Tyr Phe  Glu Tyr Asp Lys Gly  Glu Phe Val Asp Glu  Lys Thr Gly
    1265                1270                1275
```

Glu Ile Lys Pro Lys Lys Trp Arg Leu Tyr Ser Gly Glu Asn Gly
1280 1285 1290

Lys Ser Leu Asp Arg Phe Arg Gly Glu Arg Glu Lys Asp Lys Tyr
1295 1300 1305

Glu Trp Lys Ile Asp Lys Ile Asp Ile Val Lys Ile Leu Asp Asp
1310 1315 1320

Leu Phe Val Asn Phe Asp Lys Asn Ile Ser Leu Leu Lys Gln Leu
1325 1330 1335

Lys Glu Gly Val Glu Leu Thr Arg Asn Asn Glu His Gly Thr Gly
1340 1345 1350

Glu Ser Leu Arg Phe Ala Ile Asn Leu Ile Gln Gln Ile Arg Asn
1355 1360 1365

Thr Gly Asn Asn Glu Arg Asp Asn Asp Phe Ile Leu Ser Pro Val
1370 1375 1380

Arg Asp Glu Asn Gly Lys His Phe Asp Ser Arg Glu Tyr Trp Asp
1385 1390 1395

Lys Glu Thr Lys Gly Glu Lys Ile Ser Met Pro Ser Ser Gly Asp
1400 1405 1410

Ala Asn Gly Ala Phe Asn Ile Ala Arg Lys Gly Ile Ile Met Asn
1415 1420 1425

Ala His Ile Leu Ala Asn Ser Asp Ser Lys Asp Leu Ser Leu Phe
1430 1435 1440

Val Ser Asp Glu Glu Trp Asp Leu His Leu Asn Asn Lys Thr Glu
1445 1450 1455

Trp Lys Lys Gln Leu Asn Ile Phe Ser Ser Arg Lys Ala Met Ala
1460 1465 1470

Lys Arg Lys Lys Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln
1475 1480 1485

Ala Lys Lys Lys Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr
1490 1495 1500

Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val
1505 1510 1515

Pro Asp Tyr Ala
1520

<210> 234
<211> 1397
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 234
Met Glu Asn Ile Phe Asp Gln Phe Ile Gly Lys Tyr Ser Leu Ser Lys
1 5 10 15

Thr Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr Glu Asp Phe Leu

```
           20                  25                  30

Lys Ile Asn Lys Val Phe Glu Lys Asp Gln Thr Ile Asp Asp Ser Tyr
        35                  40                  45

Asn Gln Ala Lys Phe Tyr Phe Asp Ser Leu His Gln Lys Phe Ile Asp
    50                  55                  60

Ala Ala Leu Ala Ser Asp Lys Thr Ser Glu Leu Ser Phe Gln Asn Phe
65                  70                  75                  80

Ala Asp Val Leu Glu Lys Gln Asn Lys Ile Ile Leu Asp Lys Lys Arg
                85                  90                  95

Glu Met Gly Ala Leu Arg Lys Arg Asp Lys Asn Ala Val Gly Ile Asp
            100                 105                 110

Arg Leu Gln Lys Glu Ile Asn Asp Ala Glu Asp Ile Ile Gln Lys Glu
        115                 120                 125

Lys Glu Lys Ile Tyr Lys Asp Val Arg Thr Leu Phe Asp Asn Glu Ala
    130                 135                 140

Glu Ser Trp Lys Thr Tyr Tyr Gln Glu Arg Glu Val Asp Gly Lys Lys
145                 150                 155                 160

Ile Thr Phe Ser Lys Ala Asp Leu Lys Gln Lys Gly Ala Asp Phe Leu
                165                 170                 175

Thr Ala Ala Gly Ile Leu Lys Val Leu Lys Tyr Glu Phe Pro Glu Glu
            180                 185                 190

Lys Glu Lys Glu Phe Gln Ala Lys Asn Gln Pro Ser Leu Phe Val Glu
        195                 200                 205

Glu Lys Glu Asn Pro Gly Gln Lys Arg Tyr Ile Phe Asp Ser Phe Asp
    210                 215                 220

Lys Phe Ala Gly Tyr Leu Thr Lys Phe Gln Gln Thr Lys Lys Asn Leu
225                 230                 235                 240

Tyr Ala Ala Asp Gly Thr Ser Thr Ala Val Ala Thr Arg Ile Ala Asp
                245                 250                 255

Asn Phe Ile Ile Phe His Gln Asn Thr Lys Val Phe Arg Asp Lys Tyr
            260                 265                 270

Lys Asn Asn His Thr Asp Leu Gly Phe Asp Glu Glu Asn Ile Phe Glu
        275                 280                 285

Ile Glu Arg Tyr Lys Asn Cys Leu Leu Gln Arg Glu Ile Glu His Ile
    290                 295                 300

Lys Asn Glu Asn Ser Tyr Asn Lys Ile Ile Gly Arg Ile Asn Lys Lys
305                 310                 315                 320

Ile Lys Glu Tyr Arg Asp Gln Lys Ala Lys Asp Thr Lys Leu Thr Lys
                325                 330                 335

Ser Asp Phe Pro Phe Phe Lys Asn Leu Asp Lys Gln Ile Leu Gly Glu
            340                 345                 350

Val Glu Lys Glu Lys Gln Leu Ile Glu Lys Thr Arg Glu Lys Thr Glu
```

355 360 365

Glu Asp Val Leu Ile Glu Arg Phe Lys Glu Phe Ile Glu Asn Asn Glu
370 375 380

Glu Arg Phe Thr Ala Ala Lys Lys Leu Met Asn Ala Phe Cys Asn Gly
385 390 395 400

Glu Phe Glu Ser Glu Tyr Glu Gly Ile Tyr Leu Lys Asn Lys Ala Ile
405 410 415

Asn Thr Ile Ser Arg Arg Trp Phe Val Ser Asp Arg Asp Phe Glu Leu
420 425 430

Lys Leu Pro Gln Gln Lys Ser Lys Asn Lys Ser Glu Lys Asn Glu Pro
435 440 445

Lys Val Lys Lys Phe Ile Ser Ile Ala Glu Ile Lys Asn Ala Val Glu
450 455 460

Glu Leu Asp Gly Asp Ile Phe Lys Ala Val Phe Tyr Asp Lys Lys Ile
465 470 475 480

Ile Ala Gln Gly Gly Ser Lys Leu Glu Gln Phe Leu Val Ile Trp Lys
485 490 495

Tyr Glu Phe Glu Tyr Leu Phe Arg Asp Ile Glu Arg Glu Asn Gly Glu
500 505 510

Lys Leu Leu Gly Tyr Asp Ser Cys Leu Lys Ile Ala Lys Gln Leu Gly
515 520 525

Ile Phe Pro Gln Glu Lys Glu Ala Arg Glu Lys Ala Thr Ala Val Ile
530 535 540

Lys Asn Tyr Ala Asp Ala Gly Leu Gly Ile Phe Gln Met Met Lys Tyr
545 550 555 560

Phe Ser Leu Asp Asp Lys Asp Arg Lys Asn Thr Pro Gly Gln Leu Ser
565 570 575

Thr Asn Phe Tyr Ala Glu Tyr Asp Gly Tyr Tyr Lys Asp Phe Glu Phe
580 585 590

Ile Lys Tyr Tyr Asn Glu Phe Arg Asn Phe Ile Thr Lys Lys Pro Phe
595 600 605

Asp Glu Asp Lys Ile Lys Leu Asn Phe Glu Asn Gly Ala Leu Leu Lys
610 615 620

Gly Trp Asp Glu Asn Lys Glu Tyr Asp Phe Met Gly Val Ile Leu Lys
625 630 635 640

Lys Glu Gly Arg Leu Tyr Leu Gly Ile Met His Lys Asn His Arg Lys
645 650 655

Leu Phe Gln Ser Met Gly Asn Ala Lys Gly Asp Asn Ala Asn Arg Tyr
660 665 670

Gln Lys Met Ile Tyr Lys Gln Ile Ala Asp Ala Ser Lys Asp Val Pro
675 680 685

Arg Leu Leu Leu Thr Ser Lys Lys Ala Met Glu Lys Phe Lys Pro Ser

690 695 700

Gln Glu Ile Leu Arg Ile Lys Lys Glu Lys Thr Phe Lys Arg Glu Ser
705 710 715 720

Lys Asn Phe Ser Leu Arg Asp Leu His Ala Leu Ile Glu Tyr Tyr Arg
725 730 735

Asn Cys Ile Pro Gln Tyr Ser Asn Trp Ser Phe Tyr Asp Phe Gln Phe
740 745 750

Gln Asp Thr Gly Lys Tyr Gln Asn Ile Lys Glu Phe Thr Asp Asp Val
755 760 765

Gln Lys Tyr Gly Tyr Lys Ile Ser Phe Arg Asp Ile Asp Asp Glu Tyr
770 775 780

Ile Asn Gln Ala Leu Asn Glu Gly Lys Met Tyr Leu Phe Glu Val Val
785 790 795 800

Asn Lys Asp Ile Tyr Asn Thr Lys Asn Gly Ser Lys Asn Leu His Thr
805 810 815

Leu Tyr Phe Glu His Ile Leu Ser Ala Glu Asn Leu Asn Asp Pro Val
820 825 830

Phe Lys Leu Ser Gly Met Ala Glu Ile Phe Gln Arg Gln Pro Ser Val
835 840 845

Asn Glu Arg Glu Lys Ile Thr Thr Gln Lys Asn Gln Cys Ile Leu Asp
850 855 860

Lys Gly Asp Arg Ala Tyr Lys Tyr Arg Arg Tyr Thr Glu Lys Lys Ile
865 870 875 880

Met Phe His Met Ser Leu Val Leu Asn Thr Gly Lys Gly Glu Ile Lys
885 890 895

Gln Val Gln Phe Asn Lys Ile Ile Asn Gln Arg Ile Ser Ser Ser Asp
900 905 910

Asn Glu Met Arg Val Asn Val Ile Gly Ile Asp Arg Gly Glu Lys Asn
915 920 925

Leu Leu Tyr Tyr Ser Val Val Lys Gln Asn Gly Glu Ile Ile Glu Gln
930 935 940

Ala Ser Leu Asn Glu Ile Asn Gly Val Asn Tyr Arg Asp Lys Leu Ile
945 950 955 960

Glu Arg Glu Lys Glu Arg Leu Lys Asn Arg Gln Ser Trp Lys Pro Val
965 970 975

Val Lys Ile Lys Asp Leu Lys Lys Gly Tyr Ile Ser His Val Ile His
980 985 990

Lys Ile Cys Gln Leu Ile Glu Lys Tyr Ser Ala Ile Val Val Leu Glu
995 1000 1005

Asp Leu Asn Met Arg Phe Lys Gln Ile Arg Gly Gly Ile Glu Arg
1010 1015 1020

Ser Val Tyr Gln Gln Phe Glu Lys Ala Leu Ile Asp Lys Leu Gly

1025 1030 1035

Tyr Leu Val Phe Lys Asp Asn Arg Asp Leu Arg Ala Pro Gly Gly
1040 1045 1050

Val Leu Asn Gly Tyr Gln Leu Ser Ala Pro Phe Val Ser Phe Glu
1055 1060 1065

Lys Met Arg Lys Gln Thr Gly Ile Leu Phe Tyr Thr Gln Ala Glu
1070 1075 1080

Tyr Thr Ser Lys Thr Asp Pro Ile Thr Gly Phe Arg Lys Asn Val
1085 1090 1095

Tyr Ile Ser Asn Ser Ala Ser Leu Asp Lys Ile Lys Glu Ala Val
1100 1105 1110

Lys Lys Phe Asp Ala Ile Gly Trp Asp Gly Lys Glu Gln Ser Tyr
1115 1120 1125

Phe Phe Lys Tyr Asn Pro Tyr Asn Leu Ala Asp Glu Lys Tyr Lys
1130 1135 1140

Asn Ser Thr Val Ser Lys Glu Trp Ala Ile Phe Ala Ser Ala Pro
1145 1150 1155

Arg Ile Arg Arg Gln Lys Gly Glu Asp Gly Tyr Trp Lys Tyr Asp
1160 1165 1170

Arg Val Lys Val Asn Glu Glu Phe Glu Lys Leu Leu Lys Val Trp
1175 1180 1185

Asn Phe Val Asn Pro Lys Ala Thr Asp Ile Lys Gln Glu Ile Ile
1190 1195 1200

Lys Lys Glu Lys Ala Gly Asp Leu Gln Gly Glu Lys Glu Leu Asp
1205 1210 1215

Gly Arg Leu Arg Asn Phe Trp His Ser Phe Ile Tyr Leu Phe Asn
1220 1225 1230

Leu Val Leu Glu Leu Arg Asn Ser Phe Ser Leu Gln Ile Lys Ile
1235 1240 1245

Lys Ala Gly Glu Val Ile Ala Val Asp Glu Gly Val Asp Phe Ile
1250 1255 1260

Ala Ser Pro Val Lys Pro Phe Phe Thr Thr Pro Asn Pro Tyr Ile
1265 1270 1275

Pro Ser Asn Leu Cys Trp Leu Ala Val Glu Asn Ala Asp Ala Asn
1280 1285 1290

Gly Ala Tyr Asn Ile Ala Arg Lys Gly Val Met Ile Leu Lys Lys
1295 1300 1305

Ile Arg Glu His Ala Lys Lys Asp Pro Glu Phe Lys Lys Leu Pro
1310 1315 1320

Asn Leu Phe Ile Ser Asn Ala Glu Trp Asp Glu Ala Ala Arg Asp
1325 1330 1335

Trp Gly Lys Tyr Ala Gly Thr Thr Ala Leu Asn Leu Asp His Lys

1340 1345 1350

Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys
1355 1360 1365

Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp
1370 1375 1380

Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1385 1390 1395

<210> 235
<211> 1295
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 235
Met Gln Thr Leu Phe Glu Asn Phe Thr Asn Gln Tyr Pro Val Ser Lys
1 5 10 15

Thr Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Lys Asp Phe Ile
20 25 30

Glu Gln Lys Gly Leu Leu Lys Lys Asp Glu Asp Arg Ala Glu Lys Tyr
35 40 45

Lys Lys Val Lys Asn Ile Ile Asp Glu Tyr His Lys Asp Phe Ile Glu
50 55 60

Lys Ser Leu Asn Gly Leu Lys Leu Asp Gly Leu Glu Lys Tyr Lys Thr
65 70 75 80

Leu Tyr Leu Lys Gln Glu Lys Asp Asp Lys Asp Lys Lys Ala Phe Asp
85 90 95

Lys Glu Lys Glu Asn Leu Arg Lys Gln Ile Ala Asn Ala Phe Arg Asn
100 105 110

Asn Glu Lys Phe Lys Thr Leu Phe Ala Lys Glu Leu Ile Lys Asn Asp
115 120 125

Leu Met Ser Phe Ala Cys Glu Glu Asp Lys Lys Asn Val Lys Glu Phe
130 135 140

Glu Ala Phe Thr Thr Tyr Phe Thr Gly Phe His Gln Asn Arg Ala Asn
145 150 155 160

Met Tyr Val Ala Asp Glu Lys Arg Thr Ala Ile Ala Ser Arg Leu Ile
165 170 175

His Glu Asn Leu Pro Lys Phe Ile Asp Asn Ile Lys Ile Phe Glu Lys
180 185 190

Met Lys Lys Glu Ala Pro Glu Leu Leu Ser Pro Phe Asn Gln Thr Leu
195 200 205

Lys Asp Met Lys Asp Val Ile Lys Gly Thr Thr Leu Glu Glu Ile Phe
210 215 220

Ser Leu Asp Tyr Phe Asn Lys Thr Leu Thr Gln Ser Gly Ile Asp Ile
225 230 235 240

Tyr Asn Ser Val Ile Gly Gly Arg Thr Pro Glu Glu Gly Lys Thr Lys
245 250 255

Ile Lys Gly Leu Asn Glu Tyr Ile Asn Thr Asp Phe Asn Gln Lys Gln
260 265 270

Thr Asp Lys Lys Lys Arg Gln Pro Lys Phe Lys Gln Leu Tyr Lys Gln
275 280 285

Ile Leu Ser Asp Arg Gln Ser Leu Ser Phe Ile Ala Glu Ala Phe Lys
290 295 300

Asn Asp Thr Glu Ile Leu Glu Ala Ile Glu Lys Phe Tyr Val Asn Glu
305 310 315 320

Leu Leu His Phe Ser Asn Glu Gly Lys Ser Thr Asn Val Leu Asp Ala
325 330 335

Ile Lys Asn Ala Val Ser Asn Leu Glu Ser Phe Asn Leu Thr Lys Met
340 345 350

Tyr Phe Arg Ser Gly Ala Ser Leu Thr Asp Val Ser Arg Lys Val Phe
355 360 365

Gly Glu Trp Ser Ile Ile Asn Arg Ala Leu Asp Asn Tyr Tyr Ala Thr
370 375 380

Thr Tyr Pro Ile Lys Pro Arg Glu Lys Ser Glu Lys Tyr Glu Glu Arg
385 390 395 400

Lys Glu Lys Trp Leu Lys Gln Asp Phe Asn Val Ser Leu Ile Gln Thr
405 410 415

Ala Ile Asp Glu Tyr Asp Asn Glu Thr Val Lys Gly Lys Asn Ser Gly
420 425 430

Lys Val Ile Ala Asp Tyr Phe Ala Lys Phe Cys Asp Asp Lys Glu Thr
435 440 445

Asp Leu Ile Gln Lys Val Asn Glu Gly Tyr Ile Ala Val Lys Asp Leu
450 455 460

Leu Asn Thr Pro Cys Pro Glu Asn Glu Lys Leu Gly Ser Asn Lys Asp
465 470 475 480

Gln Val Lys Gln Ile Lys Ala Phe Met Asp Ser Ile Met Asp Ile Met
485 490 495

His Phe Val Arg Pro Leu Ser Leu Lys Asp Thr Asp Lys Glu Lys Asp
500 505 510

Glu Thr Phe Tyr Ser Leu Phe Thr Pro Leu Tyr Asp His Leu Thr Gln
515 520 525

Thr Ile Ala Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Gln Lys Pro
530 535 540

Tyr Ser Thr Glu Lys Ile Lys Leu Asn Phe Glu Asn Ser Thr Leu Leu
545 550 555 560

Gly Gly Trp Asp Leu Asn Lys Glu Thr Asp Asn Thr Ala Ile Ile Leu
565 570 575

Arg Lys Asp Asn Leu Tyr Tyr Leu Gly Ile Met Asp Lys Arg His Asn
580 585 590

Arg Ile Phe Arg Asn Val Pro Lys Ala Asp Lys Lys Asp Phe Cys Tyr
595 600 605

Glu Lys Met Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro
610 615 620

Lys Val Phe Phe Ser Gln Ser Arg Ile Gln Glu Phe Thr Pro Ser Ala
625 630 635 640

Lys Leu Leu Glu Asn Tyr Ala Asn Glu Thr His Lys Lys Gly Asp Asn
645 650 655

Phe Asn Leu Asn His Cys His Lys Leu Ile Asp Phe Phe Lys Asp Ser
660 665 670

Ile Asn Lys His Glu Asp Trp Lys Asn Phe Asp Phe Arg Phe Ser Ala
675 680 685

Thr Ser Thr Tyr Ala Asp Leu Ser Gly Phe Tyr His Glu Val Glu His
690 695 700

Gln Gly Tyr Lys Ile Ser Phe Gln Ser Val Ala Asp Ser Phe Ile Asp
705 710 715 720

Asp Leu Val Asn Glu Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys
725 730 735

Asp Phe Ser Pro Phe Ser Lys Gly Lys Pro Asn Leu His Thr Leu Tyr
740 745 750

Trp Lys Met Leu Phe Asp Glu Asn Asn Leu Lys Asp Val Val Tyr Lys
755 760 765

Leu Asn Gly Glu Ala Glu Val Phe Tyr Arg Lys Lys Ser Ile Ala Glu
770 775 780

Lys Asn Thr Thr Ile His Lys Ala Asn Glu Ser Ile Ile Asn Lys Asn
785 790 795 800

Pro Asp Asn Pro Lys Ala Thr Ser Thr Phe Asn Tyr Asp Ile Val Lys
805 810 815

Asp Lys Arg Tyr Thr Ile Asp Lys Phe Gln Phe His Ile Pro Ile Thr
820 825 830

Met Asn Phe Lys Ala Glu Gly Ile Phe Asn Met Asn Gln Arg Val Asn
835 840 845

Gln Phe Leu Lys Ala Asn Pro Asp Ile Asn Ile Ile Gly Ile Asp Arg
850 855 860

Gly Glu Arg His Leu Leu Tyr Tyr Ala Leu Ile Asn Gln Lys Gly Lys
865 870 875 880

Ile Leu Lys Gln Asp Thr Leu Asn Val Ile Ala Asn Glu Lys Gln Lys
885 890 895

Val Asp Tyr His Asn Leu Leu Asp Lys Lys Glu Gly Asp Arg Ala Thr
900 905 910

```
Ala Arg Gln Glu Trp Gly Val Ile Glu Thr Ile Lys Glu Leu Lys Glu
        915                 920                 925

Gly Tyr Leu Ser Gln Val Ile His Lys Leu Thr Asp Leu Met Ile Glu
    930                 935                 940

Asn Asn Ala Ile Ile Val Met Glu Asp Leu Asn Phe Gly Phe Lys Arg
945                 950                 955                 960

Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu Lys Met
                965                 970                 975

Leu Ile Asp Lys Leu Asn Tyr Leu Val Asp Lys Asn Lys Lys Ala Asn
            980                 985                 990

Glu Leu Gly Gly Leu Leu Asn Ala Phe Gln Leu Ala Asn Lys Phe Glu
        995                 1000                1005

Ser Phe  Gln Lys Met Gly Lys  Gln Asn Gly Phe Ile  Phe Tyr Val
    1010                1015                1020

Pro Ala  Trp Asn Thr Ser Lys  Thr Asp Pro Ala Thr  Gly Phe Ile
    1025                1030                1035

Asp Phe  Leu Lys Pro Arg Tyr  Glu Asn Leu Asn Gln  Ala Lys Asp
    1040                1045                1050

Phe Phe  Glu Lys Phe Asp Ser  Ile Arg Leu Asn Ser  Lys Ala Asp
    1055                1060                1065

Tyr Phe  Glu Phe Ala Phe Asp  Phe Lys Asn Phe Thr  Glu Lys Ala
    1070                1075                1080

Asp Gly  Gly Arg Thr Lys Trp  Thr Val Cys Thr Thr  Asn Glu Asp
    1085                1090                1095

Arg Tyr  Ala Trp Asn Arg Ala  Leu Asn Asn Asn Arg  Gly Ser Gln
    1100                1105                1110

Glu Lys  Tyr Asp Ile Thr Ala  Glu Leu Lys Ser Leu  Phe Asp Gly
    1115                1120                1125

Lys Val  Asp Tyr Lys Ser Gly  Lys Asp Leu Lys Gln  Gln Ile Ala
    1130                1135                1140

Ser Gln  Glu Ser Ala Asp Phe  Phe Lys Ala Leu Met  Lys Asn Leu
    1145                1150                1155

Ser Ile  Thr Leu Ser Leu Arg  His Asn Asn Gly Glu  Lys Gly Asp
    1160                1165                1170

Asn Glu  Gln Asp Tyr Ile Leu  Ser Pro Val Ala Asp  Ser Lys Gly
    1175                1180                1185

Arg Phe  Phe Asp Ser Arg Lys  Ala Asp Asp Asp Met  Pro Lys Asn
    1190                1195                1200

Ala Asp  Ala Asn Gly Ala Tyr  His Ile Ala Leu Lys  Gly Leu Trp
    1205                1210                1215

Cys Leu  Glu Gln Ile Ser Lys  Thr Asp Asp Leu Lys  Lys Val Lys
    1220                1225                1230
```

Leu Ala Ile Ser Asn Lys Glu Trp Leu Glu Phe Val Gln Thr Leu
1235 1240 1245

Lys Gly Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys
1250 1255 1260

Lys Lys Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr
1265 1270 1275

Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp
1280 1285 1290

Tyr Ala
1295

<210> 236
<211> 1352
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 236
Met Thr Gln Phe Glu Gly Phe Thr Asn Leu Tyr Gln Val Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Lys His Ile Gln
20 25 30

Glu Gln Gly Phe Ile Glu Glu Asp Lys Ala Arg Asn Asp His Tyr Lys
35 40 45

Glu Leu Lys Pro Ile Ile Asp Arg Ile Tyr Lys Thr Tyr Ala Asp Gln
50 55 60

Cys Leu Gln Leu Val Gln Leu Asp Trp Glu Asn Leu Ser Ala Ala Ile
65 70 75 80

Asp Ser Tyr Arg Lys Glu Lys Thr Glu Glu Thr Arg Asn Ala Leu Ile
85 90 95

Glu Glu Gln Ala Thr Tyr Arg Asn Ala Ile His Asp Tyr Phe Ile Gly
100 105 110

Arg Thr Asp Asn Leu Thr Asp Ala Ile Asn Lys Arg His Ala Glu Ile
115 120 125

Tyr Lys Gly Leu Phe Lys Ala Glu Leu Phe Asn Gly Lys Val Leu Lys
130 135 140

Gln Leu Gly Thr Val Thr Thr Thr Glu His Glu Asn Ala Leu Leu Arg
145 150 155 160

Ser Phe Asp Lys Phe Thr Thr Tyr Phe Ser Gly Phe Tyr Glu Asn Arg
165 170 175

Lys Asn Val Phe Ser Ala Glu Asp Ile Ser Thr Ala Ile Pro His Arg
180 185 190

Ile Val Gln Asp Asn Phe Pro Lys Phe Lys Glu Asn Cys His Ile Phe
195 200 205

Thr Arg Leu Ile Thr Ala Val Pro Ser Leu Arg Glu His Phe Glu Asn

210 215 220

Val Lys Lys Ala Ile Gly Ile Phe Val Ser Thr Ser Ile Glu Glu Val
225 230 235 240

Phe Ser Phe Pro Phe Tyr Asn Gln Leu Leu Thr Gln Thr Gln Ile Asp
245 250 255

Leu Tyr Asn Gln Leu Leu Gly Gly Ile Ser Arg Glu Ala Gly Thr Glu
260 265 270

Lys Ile Lys Gly Leu Asn Glu Val Leu Asn Leu Ala Ile Gln Lys Asn
275 280 285

Asp Glu Thr Ala His Ile Ile Ala Ser Leu Pro His Arg Phe Ile Pro
290 295 300

Leu Phe Lys Gln Ile Leu Ser Asp Arg Asn Thr Leu Ser Phe Ile Leu
305 310 315 320

Glu Glu Phe Lys Ser Asp Glu Glu Val Ile Gln Ser Phe Cys Lys Tyr
325 330 335

Lys Thr Leu Leu Arg Asn Glu Asn Val Leu Glu Thr Ala Glu Ala Leu
340 345 350

Phe Asn Glu Leu Asn Ser Ile Asp Leu Thr His Ile Phe Ile Ser His
355 360 365

Lys Lys Leu Glu Thr Ile Ser Ser Ala Leu Cys Asp His Trp Asp Thr
370 375 380

Leu Arg Asn Ala Leu Tyr Glu Arg Arg Ile Ser Glu Leu Thr Gly Lys
385 390 395 400

Ile Thr Lys Ser Ala Lys Glu Lys Val Gln Arg Ser Leu Lys His Glu
405 410 415

Asp Ile Asn Leu Gln Glu Ile Ile Ser Ala Ala Gly Lys Glu Leu Ser
420 425 430

Glu Ala Phe Lys Gln Lys Thr Ser Glu Ile Leu Ser His Ala His Ala
435 440 445

Ala Leu Asp Gln Pro Leu Pro Thr Thr Leu Lys Lys Gln Glu Glu Lys
450 455 460

Glu Ile Leu Lys Ser Gln Leu Asp Ser Leu Leu Gly Leu Tyr His Leu
465 470 475 480

Leu Asp Trp Phe Ala Val Asp Glu Ser Asn Glu Val Asp Pro Glu Phe
485 490 495

Ser Ala Arg Leu Thr Gly Ile Lys Leu Glu Met Glu Pro Ser Leu Ser
500 505 510

Phe Tyr Asn Lys Ala Arg Asn Tyr Ala Thr Lys Lys Pro Tyr Ser Val
515 520 525

Glu Lys Phe Lys Leu Asn Phe Gln Met Pro Thr Leu Ala Ser Gly Trp
530 535 540

Asp Val Asn Lys Glu Lys Asn Asn Gly Ala Ile Leu Phe Val Lys Asn

545 550 555 560

Gly Leu Tyr Tyr Leu Gly Ile Met Pro Lys Gln Lys Gly Arg Tyr Lys
565 570 575

Ala Leu Ser Phe Glu Pro Thr Glu Lys Thr Ser Glu Gly Phe Asp Lys
580 585 590

Met Tyr Tyr Asp Tyr Phe Pro Asp Ala Ala Lys Met Ile Pro Lys Cys
595 600 605

Ser Thr Gln Leu Lys Ala Val Thr Ala His Phe Gln Thr His Thr Thr
610 615 620

Pro Ile Leu Leu Ser Asn Asn Phe Ile Glu Pro Leu Glu Ile Thr Lys
625 630 635 640

Glu Ile Tyr Asp Leu Asn Asn Pro Glu Lys Glu Pro Lys Lys Phe Gln
645 650 655

Thr Ala Tyr Ala Lys Lys Thr Gly Asp Gln Lys Gly Tyr Arg Glu Ala
660 665 670

Leu Cys Lys Trp Ile Asp Phe Thr Arg Asp Phe Leu Ser Lys Tyr Thr
675 680 685

Lys Thr Thr Ser Ile Asp Leu Ser Ser Leu Arg Pro Ser Ser Gln Tyr
690 695 700

Lys Asp Leu Gly Glu Tyr Tyr Ala Glu Leu Asn Pro Leu Leu Tyr His
705 710 715 720

Ile Ser Phe Gln Arg Ile Ala Glu Lys Glu Ile Met Asp Ala Val Glu
725 730 735

Thr Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ala Lys
740 745 750

Gly His His Gly Lys Pro Asn Leu His Thr Leu Tyr Trp Thr Gly Leu
755 760 765

Phe Ser Pro Glu Asn Leu Ala Lys Thr Ser Ile Lys Leu Asn Gly Gln
770 775 780

Ala Glu Leu Phe Tyr Arg Pro Lys Ser Arg Met Lys Arg Met Ala His
785 790 795 800

Arg Leu Gly Glu Lys Met Leu Asn Lys Lys Leu Lys Asp Gln Lys Thr
805 810 815

Pro Ile Pro Asp Thr Leu Tyr Gln Glu Leu Tyr Asp Tyr Val Asn His
820 825 830

Arg Leu Ser His Asp Leu Ser Asp Glu Ala Arg Ala Leu Leu Pro Asn
835 840 845

Val Ile Thr Lys Glu Val Ser His Glu Ile Ile Lys Asp Arg Arg Phe
850 855 860

Thr Ser Asp Lys Phe Phe Phe His Val Pro Ile Thr Leu Asn Tyr Gln
865 870 875 880

Ala Ala Asn Ser Pro Ser Lys Phe Asn Gln Arg Val Asn Ala Tyr Leu

885 890 895

Lys Glu His Pro Glu Thr Pro Ile Ile Gly Ile Asp Arg Gly Glu Arg
900 905 910

Asn Leu Ile Tyr Ile Thr Val Ile Asp Ser Thr Gly Lys Ile Leu Glu
915 920 925

Gln Arg Ser Leu Asn Thr Ile Gln Gln Phe Asp Tyr Gln Lys Lys Leu
930 935 940

Asp Asn Arg Glu Lys Glu Arg Val Ala Ala Arg Gln Ala Trp Ser Val
945 950 955 960

Val Gly Thr Ile Lys Asp Leu Lys Gln Gly Tyr Leu Ser Gln Val Ile
965 970 975

His Glu Ile Val Asp Leu Met Ile His Tyr Gln Ala Val Val Val Leu
980 985 990

Glu Asn Leu Asn Phe Gly Phe Lys Ser Lys Arg Thr Gly Ile Ala Glu
995 1000 1005

Lys Ala Val Tyr Gln Gln Phe Glu Lys Met Leu Ile Asp Lys Leu
1010 1015 1020

Asn Cys Leu Val Leu Lys Asp Tyr Pro Ala Glu Lys Val Gly Gly
1025 1030 1035

Val Leu Asn Pro Tyr Gln Leu Thr Asp Gln Phe Thr Ser Phe Ala
1040 1045 1050

Lys Met Gly Thr Gln Ser Gly Phe Leu Phe Tyr Val Pro Ala Pro
1055 1060 1065

Tyr Thr Ser Lys Ile Asp Pro Leu Thr Gly Phe Val Asp Pro Phe
1070 1075 1080

Val Trp Lys Thr Ile Lys Asn His Glu Ser Arg Lys His Phe Leu
1085 1090 1095

Glu Gly Phe Asp Phe Leu His Tyr Asp Val Lys Thr Gly Asp Phe
1100 1105 1110

Ile Leu His Phe Lys Met Asn Arg Asn Leu Ser Phe Gln Arg Gly
1115 1120 1125

Leu Pro Gly Phe Met Pro Ala Trp Asp Ile Val Phe Glu Lys Asn
1130 1135 1140

Glu Thr Gln Phe Asp Ala Lys Gly Thr Pro Phe Ile Ala Gly Lys
1145 1150 1155

Arg Ile Val Pro Val Ile Glu Asn His Arg Phe Thr Gly Arg Tyr
1160 1165 1170

Arg Asp Leu Tyr Pro Ala Asn Glu Leu Ile Ala Leu Leu Glu Glu
1175 1180 1185

Lys Gly Ile Val Phe Arg Asp Gly Ser Asn Ile Leu Pro Lys Leu
1190 1195 1200

Leu Glu Asn Asp Asp Ser His Ala Ile Asp Thr Met Val Ala Leu

1205 1210 1215

Ile Arg Ser Val Leu Gln Met Arg Asn Ser Asn Ala Ala Thr Gly
1220 1225 1230

Glu Asp Tyr Ile Asn Ser Pro Val Arg Asp Leu Asn Gly Val Cys
1235 1240 1245

Phe Asp Ser Arg Phe Gln Asn Pro Glu Trp Pro Met Asp Ala Asp
1250 1255 1260

Ala Asn Gly Ala Tyr His Ile Ala Leu Lys Gly Gln Leu Leu Leu
1265 1270 1275

Asn His Leu Lys Glu Ser Lys Asp Leu Lys Leu Gln Asn Gly Ile
1280 1285 1290

Ser Asn Gln Asp Trp Leu Ala Tyr Ile Gln Glu Leu Arg Asn Lys
1295 1300 1305

Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys
1310 1315 1320

Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp
1325 1330 1335

Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1340 1345 1350

<210> 237
<211> 1251
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 237
Met Tyr Tyr Glu Ser Leu Thr Lys Gln Tyr Pro Val Ser Lys Thr Ile
1 5 10 15

Arg Asn Glu Leu Ile Pro Ile Gly Lys Thr Leu Asp Asn Ile Arg Gln
20 25 30

Asn Asn Ile Leu Glu Ser Asp Val Lys Arg Lys Gln Asn Tyr Glu His
35 40 45

Val Lys Gly Ile Leu Asp Glu Tyr His Lys Gln Leu Ile Asn Glu Ala
50 55 60

Leu Asp Asn Cys Thr Leu Pro Ser Leu Lys Ile Ala Ala Glu Ile Tyr
65 70 75 80

Leu Lys Asn Gln Lys Glu Val Ser Asp Arg Glu Asp Phe Asn Lys Thr
85 90 95

Gln Asp Leu Leu Arg Lys Glu Val Val Glu Lys Leu Lys Ala His Glu
100 105 110

Asn Phe Thr Lys Ile Gly Lys Lys Asp Ile Leu Asp Leu Leu Glu Lys
115 120 125

Leu Pro Ser Ile Ser Glu Asp Asp Tyr Asn Ala Leu Glu Ser Phe Arg
130 135 140

```
Asn Phe Tyr Thr Tyr Phe Thr Ser Tyr Asn Lys Val Arg Glu Asn Leu
145                 150                 155                 160

Tyr Ser Asp Lys Glu Lys Ser Ser Thr Val Ala Tyr Arg Leu Ile Asn
                165                 170                 175

Glu Asn Phe Pro Lys Phe Leu Asp Asn Val Lys Ser Tyr Arg Phe Val
            180                 185                 190

Lys Thr Ala Gly Ile Leu Ala Asp Gly Leu Gly Glu Glu Glu Gln Asp
        195                 200                 205

Ser Leu Phe Ile Val Glu Thr Phe Asn Lys Thr Leu Thr Gln Asp Gly
    210                 215                 220

Ile Asp Thr Tyr Asn Ser Gln Val Gly Lys Ile Asn Ser Ser Ile Asn
225                 230                 235                 240

Leu Tyr Asn Gln Lys Asn Gln Lys Ala Asn Gly Phe Arg Lys Ile Pro
                245                 250                 255

Lys Met Lys Met Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Glu Ser
            260                 265                 270

Phe Ile Asp Glu Phe Gln Ser Asp Glu Val Leu Ile Asp Asn Val Glu
        275                 280                 285

Ser Tyr Gly Ser Val Leu Ile Glu Ser Leu Lys Ser Ser Lys Val Ser
    290                 295                 300

Ala Phe Phe Asp Ala Leu Arg Glu Ser Lys Gly Lys Asn Val Tyr Val
305                 310                 315                 320

Lys Asn Asp Leu Ala Lys Thr Ala Met Ser Asn Ile Val Phe Glu Asn
                325                 330                 335

Trp Arg Thr Phe Asp Asp Leu Leu Asn Gln Glu Tyr Asp Leu Ala Asn
            340                 345                 350

Glu Asn Lys Lys Lys Asp Asp Lys Tyr Phe Glu Lys Arg Gln Lys Glu
        355                 360                 365

Leu Lys Lys Asn Lys Ser Tyr Ser Leu Glu His Leu Cys Asn Leu Ser
    370                 375                 380

Glu Asp Ser Cys Asn Leu Ile Glu Asn Tyr Ile His Gln Ile Ser Asp
385                 390                 395                 400

Asp Ile Glu Asn Ile Ile Ile Asn Asn Glu Thr Phe Leu Arg Ile Val
                405                 410                 415

Ile Asn Glu His Asp Arg Ser Arg Lys Leu Ala Lys Asn Arg Lys Ala
            420                 425                 430

Val Lys Ala Ile Lys Asp Phe Leu Asp Ser Ile Lys Val Leu Glu Arg
        435                 440                 445

Glu Leu Lys Leu Ile Asn Ser Ser Gly Gln Glu Leu Glu Lys Asp Leu
    450                 455                 460

Ile Val Tyr Ser Ala His Glu Glu Leu Leu Val Glu Leu Lys Gln Val
465                 470                 475                 480
```

```
Asp Ser Leu Tyr Asn Met Thr Arg Asn Tyr Leu Thr Lys Lys Pro Phe
                485                 490                 495

Ser Thr Glu Lys Val Lys Leu Asn Phe Asn Arg Ser Thr Leu Leu Asn
            500                 505                 510

Gly Trp Asp Arg Asn Lys Glu Thr Asp Asn Leu Gly Val Leu Leu Leu
        515                 520                 525

Lys Asp Gly Lys Tyr Tyr Leu Gly Ile Met Asn Thr Ser Ala Asn Lys
    530                 535                 540

Ala Phe Val Asn Pro Pro Val Ala Lys Thr Glu Lys Val Phe Lys Lys
545                 550                 555                 560

Val Asp Tyr Lys Leu Leu Pro Val Pro Asn Gln Met Leu Pro Lys Val
                565                 570                 575

Phe Phe Ala Lys Ser Asn Ile Asp Phe Tyr Asn Pro Ser Ser Glu Ile
            580                 585                 590

Tyr Ser Asn Tyr Lys Lys Gly Thr His Lys Lys Gly Asn Met Phe Ser
        595                 600                 605

Leu Glu Asp Cys His Asn Leu Ile Asp Phe Phe Lys Glu Ser Ile Ser
    610                 615                 620

Lys His Glu Asp Trp Ser Lys Phe Gly Phe Lys Phe Ser Asp Thr Ala
625                 630                 635                 640

Ser Tyr Asn Asp Ile Ser Glu Phe Tyr Arg Glu Val Glu Lys Gln Gly
                645                 650                 655

Tyr Lys Leu Thr Tyr Thr Asp Ile Asp Glu Thr Tyr Ile Asn Asp Leu
            660                 665                 670

Ile Glu Arg Asn Glu Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe
        675                 680                 685

Ser Met Tyr Ser Lys Gly Lys Leu Asn Leu His Thr Leu Tyr Phe Met
    690                 695                 700

Met Leu Phe Asp Gln Arg Asn Ile Asp Asp Val Val Tyr Lys Leu Asn
705                 710                 715                 720

Gly Glu Ala Glu Val Phe Tyr Arg Pro Ala Ser Ile Ser Glu Asp Glu
                725                 730                 735

Leu Ile Ile His Lys Ala Gly Glu Glu Ile Lys Asn Lys Asn Pro Asn
            740                 745                 750

Arg Ala Arg Thr Lys Glu Thr Ser Thr Phe Ser Tyr Asp Ile Val Lys
        755                 760                 765

Asp Lys Arg Tyr Ser Lys Asp Lys Phe Thr Leu His Ile Pro Ile Thr
    770                 775                 780

Met Asn Phe Gly Val Asp Glu Val Lys Arg Phe Asn Asp Ala Val Asn
785                 790                 795                 800

Ser Ala Ile Arg Ile Asp Glu Asn Val Asn Val Ile Gly Ile Asp Arg
                805                 810                 815
```

Gly Glu Arg Asn Leu Leu Tyr Val Val Val Ile Asp Ser Lys Gly Asn
820 825 830

Ile Leu Glu Gln Ile Ser Leu Asn Ser Ile Ile Asn Lys Glu Tyr Asp
835 840 845

Ile Glu Thr Asp Tyr His Ala Leu Leu Asp Glu Arg Glu Gly Gly Arg
850 855 860

Asp Lys Ala Arg Lys Asp Trp Asn Thr Val Glu Asn Ile Arg Asp Leu
865 870 875 880

Lys Ala Gly Tyr Leu Ser Gln Val Val Asn Val Val Ala Lys Leu Val
885 890 895

Leu Lys Tyr Asn Ala Ile Ile Cys Leu Glu Asp Leu Asn Phe Gly Phe
900 905 910

Lys Arg Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu
915 920 925

Lys Met Leu Ile Asp Lys Leu Asn Tyr Leu Val Ile Asp Lys Ser Arg
930 935 940

Glu Gln Thr Ser Pro Lys Glu Leu Gly Gly Ala Leu Asn Ala Leu Gln
945 950 955 960

Leu Thr Ser Lys Phe Lys Ser Phe Lys Glu Leu Gly Lys Gln Ser Gly
965 970 975

Val Ile Tyr Tyr Val Pro Ala Tyr Leu Thr Ser Lys Ile Asp Pro Thr
980 985 990

Thr Gly Phe Ala Asn Leu Phe Tyr Met Lys Cys Glu Asn Val Glu Lys
995 1000 1005

Ser Lys Arg Phe Phe Asp Gly Phe Asp Phe Ile Arg Phe Asn Ala
1010 1015 1020

Leu Glu Asn Val Phe Glu Phe Gly Phe Asp Tyr Arg Ser Phe Thr
1025 1030 1035

Gln Arg Ala Cys Gly Ile Asn Ser Lys Trp Thr Val Cys Thr Asn
1040 1045 1050

Gly Glu Arg Ile Ile Lys Tyr Arg Asn Pro Asp Lys Asn Asn Met
1055 1060 1065

Phe Asp Glu Lys Val Val Val Val Thr Asp Glu Met Lys Asn Leu
1070 1075 1080

Phe Glu Gln Tyr Lys Ile Pro Tyr Glu Asp Gly Arg Asn Val Lys
1085 1090 1095

Asp Met Ile Ile Ser Asn Glu Glu Ala Glu Phe Tyr Arg Arg Leu
1100 1105 1110

Tyr Arg Leu Leu Gln Gln Thr Leu Gln Met Arg Asn Ser Thr Ser
1115 1120 1125

Asp Gly Thr Arg Asp Tyr Ile Ile Ser Pro Val Lys Asn Lys Arg
1130 1135 1140

Glu Ala Tyr Phe Asn Ser Glu Leu Ser Asp Gly Ser Val Pro Lys
1145 1150 1155

Asp Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala Arg Lys Gly Leu
1160 1165 1170

Trp Val Leu Glu Gln Ile Arg Gln Lys Ser Glu Gly Glu Lys Ile
1175 1180 1185

Asn Leu Ala Met Thr Asn Ala Glu Trp Leu Glu Tyr Ala Gln Thr
1190 1195 1200

His Leu Leu Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala
1205 1210 1215

Lys Lys Lys Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1220 1225 1230

Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro
1235 1240 1245

Asp Tyr Ala
1250

<210> 238
<211> 1283
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 238
Met Asn Asn Tyr Asp Glu Phe Thr Lys Leu Tyr Pro Ile Gln Lys Thr
1 5 10 15

Ile Arg Phe Glu Leu Lys Pro Gln Gly Arg Thr Met Glu His Leu Glu
20 25 30

Thr Phe Asn Phe Phe Glu Glu Asp Arg Asp Arg Ala Glu Lys Tyr Lys
35 40 45

Ile Leu Lys Glu Ala Ile Asp Glu Tyr His Lys Lys Phe Ile Asp Glu
50 55 60

His Leu Thr Asn Met Ser Leu Asp Trp Asn Ser Leu Lys Gln Ile Ser
65 70 75 80

Glu Lys Tyr Tyr Lys Ser Arg Glu Glu Lys Asp Lys Lys Val Phe Leu
85 90 95

Ser Glu Gln Lys Arg Met Arg Gln Glu Ile Val Ser Glu Phe Lys Lys
100 105 110

Asp Asp Arg Phe Lys Asp Leu Phe Ser Lys Lys Leu Phe Ser Glu Leu
115 120 125

Leu Lys Glu Glu Ile Tyr Lys Lys Gly Asn His Gln Glu Ile Asp Ala
130 135 140

Leu Lys Ser Phe Asp Lys Phe Ser Gly Tyr Phe Ile Gly Leu His Glu
145 150 155 160

Asn Arg Lys Asn Met Tyr Ser Asp Gly Asp Glu Ile Thr Ala Ile Ser

165 170 175

Asn Arg Ile Val Asn Glu Asn Phe Pro Lys Phe Leu Asp Asn Leu Gln
180 185 190

Lys Tyr Gln Glu Ala Arg Lys Lys Tyr Pro Glu Trp Ile Ile Lys Ala
195 200 205

Glu Ser Ala Leu Val Ala His Asn Ile Lys Met Asp Glu Val Phe Ser
210 215 220

Leu Glu Tyr Phe Asn Lys Val Leu Asn Gln Glu Gly Ile Gln Arg Tyr
225 230 235 240

Asn Leu Ala Leu Gly Gly Tyr Val Thr Lys Ser Gly Glu Lys Met Met
245 250 255

Gly Leu Asn Asp Ala Leu Asn Leu Ala His Gln Ser Glu Lys Ser Ser
260 265 270

Lys Gly Arg Ile His Met Thr Pro Leu Phe Lys Gln Ile Leu Ser Glu
275 280 285

Lys Glu Ser Phe Ser Tyr Ile Pro Asp Val Phe Thr Glu Asp Ser Gln
290 295 300

Leu Leu Pro Ser Ile Gly Gly Phe Phe Ala Gln Ile Glu Asn Asp Lys
305 310 315 320

Asp Gly Asn Ile Phe Asp Arg Ala Leu Glu Leu Ile Ser Ser Tyr Ala
325 330 335

Glu Tyr Asp Thr Glu Arg Ile Tyr Ile Arg Gln Ala Asp Ile Asn Arg
340 345 350

Val Ser Asn Val Ile Phe Gly Glu Trp Gly Thr Leu Gly Gly Leu Met
355 360 365

Arg Glu Tyr Lys Ala Asp Ser Ile Asn Asp Ile Asn Leu Glu Arg Thr
370 375 380

Cys Lys Lys Val Asp Lys Trp Leu Asp Ser Lys Glu Phe Ala Leu Ser
385 390 395 400

Asp Val Leu Glu Ala Ile Lys Arg Thr Gly Asn Asn Asp Ala Phe Asn
405 410 415

Glu Tyr Ile Ser Lys Met Arg Thr Ala Arg Glu Lys Ile Asp Ala Ala
420 425 430

Arg Lys Glu Met Lys Phe Ile Ser Glu Lys Ile Ser Gly Asp Glu Glu
435 440 445

Ser Ile His Ile Ile Lys Thr Leu Leu Asp Ser Val Gln Gln Phe Leu
450 455 460

His Phe Phe Asn Leu Phe Lys Ala Arg Gln Asp Ile Pro Leu Asp Gly
465 470 475 480

Ala Phe Tyr Ala Glu Phe Asp Glu Val His Ser Lys Leu Phe Ala Ile
485 490 495

Val Pro Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Lys Asn Asn Leu

500 505 510

Asn Thr Lys Lys Ile Lys Leu Asn Phe Lys Asn Pro Thr Leu Ala Asn
515 520 525

Gly Trp Asp Gln Asn Lys Val Tyr Asp Tyr Ala Ser Leu Ile Phe Leu
530 535 540

Arg Asp Gly Asn Tyr Tyr Leu Gly Ile Ile Asn Pro Lys Arg Lys Lys
545 550 555 560

Asn Ile Lys Phe Glu Gln Gly Ser Gly Asn Gly Pro Phe Tyr Arg Lys
565 570 575

Met Val Tyr Lys Gln Ile Pro Gly Pro Asn Lys Asn Leu Pro Arg Val
580 585 590

Phe Leu Thr Ser Thr Lys Gly Lys Lys Glu Tyr Lys Pro Ser Lys Glu
595 600 605

Ile Ile Glu Gly Tyr Glu Ala Asp Lys His Ile Arg Gly Asp Lys Phe
610 615 620

Asp Leu Asp Phe Cys His Lys Leu Ile Asp Phe Phe Lys Glu Ser Ile
625 630 635 640

Glu Lys His Lys Asp Trp Ser Lys Phe Asn Phe Tyr Phe Ser Pro Thr
645 650 655

Glu Ser Tyr Gly Asp Ile Ser Glu Phe Tyr Leu Asp Val Glu Lys Gln
660 665 670

Gly Tyr Arg Met His Phe Glu Asn Ile Ser Ala Glu Thr Ile Asp Glu
675 680 685

Tyr Val Glu Lys Gly Asp Leu Phe Leu Phe Gln Ile Tyr Asn Lys Asp
690 695 700

Phe Val Lys Ala Ala Thr Gly Lys Lys Asp Met His Thr Ile Tyr Trp
705 710 715 720

Asn Ala Ala Phe Ser Pro Glu Asn Leu Gln Asp Val Val Val Lys Leu
725 730 735

Asn Gly Glu Ala Glu Leu Phe Tyr Arg Asp Lys Ser Asp Ile Lys Glu
740 745 750

Ile Val His Arg Glu Gly Glu Ile Leu Val Asn Arg Thr Tyr Asn Gly
755 760 765

Arg Thr Pro Val Pro Asp Lys Ile His Lys Lys Leu Thr Asp Tyr His
770 775 780

Asn Gly Arg Thr Lys Asp Leu Gly Glu Ala Lys Glu Tyr Leu Asp Lys
785 790 795 800

Val Arg Tyr Phe Lys Ala His Tyr Asp Ile Thr Lys Asp Arg Arg Tyr
805 810 815

Leu Asn Asp Lys Ile Tyr Phe His Val Pro Leu Thr Leu Asn Phe Lys
820 825 830

Ala Asn Gly Lys Lys Asn Leu Asn Lys Met Val Ile Glu Lys Phe Leu

835 840 845

Ser Asp Glu Lys Ala His Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn
850 855 860

Leu Leu Tyr Tyr Ser Ile Ile Asp Arg Ser Gly Lys Ile Ile Asp Gln
865 870 875 880

Gln Ser Leu Asn Val Ile Asp Gly Phe Asp Tyr Arg Glu Lys Leu Asn
885 890 895

Gln Arg Glu Ile Glu Met Lys Asp Ala Arg Gln Ser Trp Asn Ala Ile
900 905 910

Gly Lys Ile Lys Asp Leu Lys Glu Gly Tyr Leu Ser Lys Ala Val His
915 920 925

Glu Ile Thr Lys Met Ala Ile Gln Tyr Asn Ala Ile Val Val Met Glu
930 935 940

Glu Leu Asn Tyr Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln
945 950 955 960

Ile Tyr Gln Lys Phe Glu Asn Met Leu Ile Asp Lys Met Asn Tyr Leu
965 970 975

Val Phe Lys Asp Ala Pro Asp Glu Ser Pro Gly Gly Val Leu Asn Ala
980 985 990

Tyr Gln Leu Thr Asn Pro Leu Glu Ser Phe Ala Lys Leu Gly Lys Gln
995 1000 1005

Thr Gly Ile Leu Phe Tyr Val Pro Ala Ala Tyr Thr Ser Lys Ile
1010 1015 1020

Asp Pro Thr Thr Gly Phe Val Asn Leu Phe Asn Thr Ser Ser Lys
1025 1030 1035

Thr Asn Ala Gln Glu Arg Lys Glu Phe Leu Gln Lys Phe Glu Ser
1040 1045 1050

Ile Ser Tyr Ser Ala Lys Asp Gly Gly Ile Phe Ala Phe Ala Phe
1055 1060 1065

Asp Tyr Arg Lys Phe Gly Thr Ser Lys Thr Asp His Lys Asn Val
1070 1075 1080

Trp Thr Ala Tyr Thr Asn Gly Glu Arg Met Arg Tyr Ile Lys Glu
1085 1090 1095

Lys Lys Arg Asn Glu Leu Phe Asp Pro Ser Lys Glu Ile Lys Glu
1100 1105 1110

Ala Leu Thr Ser Ser Gly Ile Lys Tyr Asp Gly Gly Gln Asn Ile
1115 1120 1125

Leu Pro Asp Ile Leu Arg Ser Asn Asn Asn Gly Leu Ile Tyr Thr
1130 1135 1140

Met Tyr Ser Ser Phe Ile Ala Ala Ile Gln Met Arg Val Tyr Asp
1145 1150 1155

Gly Lys Glu Asp Tyr Ile Ile Ser Pro Ile Lys Asn Ser Lys Gly

1160 1165 1170

Glu Phe Phe Arg Thr Asp Pro Lys Arg Arg Glu Leu Pro Ile Asp
1175 1180 1185

Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala Leu Arg Gly Glu Leu
1190 1195 1200

Thr Met Arg Ala Ile Ala Glu Lys Phe Asp Pro Asp Ser Glu Lys
1205 1210 1215

Met Ala Lys Leu Glu Leu Lys His Lys Asp Trp Phe Glu Phe Met
1220 1225 1230

Gln Thr Arg Gly Asp Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly
1235 1240 1245

Gln Ala Lys Lys Lys Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp
1250 1255 1260

Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp
1265 1270 1275

Val Pro Asp Tyr Ala
1280

<210> 239
<211> 1327
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 239
Met Asn Gly Asn Arg Ser Ile Val Tyr Arg Glu Phe Val Gly Val Ile
1 5 10 15

Pro Val Ala Lys Thr Leu Arg Asn Glu Leu Arg Pro Val Gly His Thr
20 25 30

Gln Glu His Ile Ile Gln Asn Gly Leu Ile Gln Glu Asp Glu Leu Arg
35 40 45

Gln Glu Lys Ser Thr Glu Leu Lys Asn Ile Met Asp Asp Tyr Tyr Arg
50 55 60

Glu Tyr Ile Asp Lys Ser Leu Ser Gly Val Thr Asp Leu Asp Phe Thr
65 70 75 80

Leu Leu Phe Glu Leu Met Asn Leu Val Gln Ser Ser Pro Ser Lys Asp
85 90 95

Asn Lys Lys Ala Leu Glu Lys Glu Gln Ser Lys Met Arg Glu Gln Ile
100 105 110

Cys Thr His Leu Gln Ser Asp Ser Asn Tyr Lys Asn Ile Phe Asn Ala
115 120 125

Lys Leu Leu Lys Glu Ile Leu Pro Asp Phe Ile Lys Asn Tyr Asn Gln
130 135 140

Tyr Asp Val Lys Asp Lys Ala Gly Lys Leu Glu Thr Leu Ala Leu Phe
145 150 155 160

```
Asn Gly Phe Ser Thr Tyr Phe Thr Asp Phe Phe Glu Lys Arg Lys Asn
                165                 170                 175

Val Phe Thr Lys Glu Ala Val Ser Thr Ser Ile Ala Tyr Arg Ile Val
            180                 185                 190

His Glu Asn Ser Leu Ile Phe Leu Ala Asn Met Thr Ser Tyr Lys Lys
        195                 200                 205

Ile Ser Glu Lys Ala Leu Asp Glu Ile Glu Val Ile Glu Lys Asn Asn
    210                 215                 220

Gln Asp Lys Met Gly Asp Trp Glu Leu Asn Gln Ile Phe Asn Pro Asp
225                 230                 235                 240

Phe Tyr Asn Met Val Leu Ile Gln Ser Gly Ile Asp Phe Tyr Asn Glu
                245                 250                 255

Ile Cys Gly Val Val Asn Ala His Met Asn Leu Tyr Cys Gln Gln Thr
            260                 265                 270

Lys Asn Asn Tyr Asn Leu Phe Lys Met Arg Lys Leu His Lys Gln Ile
        275                 280                 285

Leu Ala Tyr Thr Ser Thr Ser Phe Glu Val Pro Lys Met Phe Glu Asp
    290                 295                 300

Asp Met Ser Val Tyr Asn Ala Val Asn Ala Phe Ile Asp Glu Thr Glu
305                 310                 315                 320

Lys Gly Asn Ile Ile Gly Lys Leu Lys Asp Ile Val Asn Lys Tyr Asp
                325                 330                 335

Glu Leu Asp Glu Lys Arg Ile Tyr Ile Ser Lys Asp Phe Tyr Glu Thr
            340                 345                 350

Leu Ser Cys Phe Met Ser Gly Asn Trp Asn Leu Ile Thr Gly Cys Val
        355                 360                 365

Glu Asn Phe Tyr Asp Glu Asn Ile His Ala Lys Gly Lys Ser Lys Glu
    370                 375                 380

Glu Lys Val Lys Lys Ala Val Lys Glu Asp Lys Tyr Lys Ser Ile Asn
385                 390                 395                 400

Asp Val Asn Asp Leu Val Glu Lys Tyr Ile Asp Glu Lys Glu Arg Asn
                405                 410                 415

Glu Phe Lys Asn Ser Asn Ala Lys Gln Tyr Ile Arg Glu Ile Ser Asn
            420                 425                 430

Ile Ile Thr Asp Thr Glu Thr Ala His Leu Glu Tyr Asp Asp His Ile
        435                 440                 445

Ser Leu Ile Glu Ser Glu Glu Lys Ala Asp Glu Met Lys Lys Arg Leu
    450                 455                 460

Asp Met Tyr Met Asn Met Tyr His Trp Ala Lys Ala Phe Ile Val Asp
465                 470                 475                 480

Glu Val Leu Asp Arg Asp Glu Met Phe Tyr Ser Asp Ile Asp Asp Ile
                485                 490                 495
```

Tyr Asn Ile Leu Glu Asn Ile Val Pro Leu Tyr Asn Arg Val Arg Asn
500 505 510

Tyr Val Thr Gln Lys Pro Tyr Asn Ser Lys Lys Ile Lys Leu Asn Phe
515 520 525

Gln Ser Pro Thr Leu Ala Asn Gly Trp Ser Gln Ser Lys Glu Phe Asp
530 535 540

Asn Asn Ala Ile Ile Leu Ile Arg Asp Asn Lys Tyr Tyr Leu Ala Ile
545 550 555 560

Phe Asn Ala Lys Asn Lys Pro Asp Lys Lys Ile Ile Gln Gly Asn Ser
565 570 575

Asp Lys Lys Asn Asp Asn Asp Tyr Lys Lys Met Val Tyr Asn Leu Leu
580 585 590

Pro Gly Ala Asn Lys Met Leu Pro Lys Val Phe Leu Ser Lys Lys Gly
595 600 605

Ile Glu Thr Phe Lys Pro Ser Asp Tyr Ile Ile Ser Gly Tyr Asn Ala
610 615 620

His Lys His Ile Lys Thr Ser Glu Asn Phe Asp Ile Ser Phe Cys Arg
625 630 635 640

Asp Leu Ile Asp Tyr Phe Lys Asn Ser Ile Glu Lys His Ala Glu Trp
645 650 655

Arg Lys Tyr Glu Phe Lys Phe Ser Ala Thr Asp Ser Tyr Ser Asp Ile
660 665 670

Ser Glu Phe Tyr Arg Glu Val Glu Met Gln Gly Tyr Arg Ile Asp Trp
675 680 685

Thr Tyr Ile Ser Glu Ala Asp Ile Asn Lys Leu Asp Glu Glu Gly Lys
690 695 700

Ile Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ala Glu Asn Ser Thr
705 710 715 720

Gly Lys Glu Asn Leu His Thr Met Tyr Phe Lys Asn Ile Phe Ser Glu
725 730 735

Glu Asn Leu Lys Asp Ile Ile Ile Lys Leu Asn Gly Gln Ala Glu Leu
740 745 750

Phe Tyr Arg Arg Ala Ser Val Lys Asn Pro Val Lys His Lys Lys Asp
755 760 765

Ser Val Leu Val Asn Lys Thr Tyr Lys Asn Gln Leu Asp Asn Gly Asp
770 775 780

Val Val Arg Ile Pro Ile Pro Asp Asp Ile Tyr Asn Glu Ile Tyr Lys
785 790 795 800

Met Tyr Asn Gly Tyr Ile Lys Glu Ser Asp Leu Ser Glu Ala Ala Lys
805 810 815

Glu Tyr Leu Asp Lys Val Glu Val Arg Thr Ala Gln Lys Asp Ile Val
820 825 830

Lys Asp Tyr Arg Tyr Thr Val Asp Lys Tyr Phe Ile His Thr Pro Ile
835 840 845

Thr Ile Asn Tyr Lys Val Thr Ala Arg Asn Asn Val Asn Asp Met Val
850 855 860

Val Lys Tyr Ile Ala Gln Asn Asp Asp Ile His Val Ile Gly Ile Asp
865 870 875 880

Arg Gly Glu Arg Asn Leu Ile Tyr Ile Ser Val Ile Asp Ser His Gly
885 890 895

Asn Ile Val Lys Gln Lys Ser Tyr Asn Ile Leu Asn Asn Tyr Asp Tyr
900 905 910

Lys Lys Lys Leu Val Glu Lys Glu Lys Thr Arg Glu Tyr Ala Arg Lys
915 920 925

Asn Trp Lys Ser Ile Gly Asn Ile Lys Glu Leu Lys Glu Gly Tyr Ile
930 935 940

Ser Gly Val Val His Glu Ile Ala Met Leu Ile Val Glu Tyr Asn Ala
945 950 955 960

Ile Ile Ala Met Glu Asp Leu Asn Tyr Gly Phe Lys Arg Gly Arg Phe
965 970 975

Lys Val Glu Arg Gln Val Tyr Gln Lys Phe Glu Ser Met Leu Ile Asn
980 985 990

Lys Leu Asn Tyr Phe Ala Ser Lys Glu Lys Ser Val Asp Glu Pro Gly
995 1000 1005

Gly Leu Leu Lys Gly Tyr Gln Leu Thr Tyr Val Pro Asp Asn Ile
1010 1015 1020

Lys Asn Leu Gly Lys Gln Cys Gly Val Ile Phe Tyr Val Pro Ala
1025 1030 1035

Ala Phe Thr Ser Lys Ile Asp Pro Ser Thr Gly Phe Ile Ser Ala
1040 1045 1050

Phe Asn Phe Lys Ser Ile Ser Thr Asn Ala Ser Arg Lys Gln Phe
1055 1060 1065

Phe Met Gln Phe Asp Glu Ile Arg Tyr Cys Ala Glu Lys Asp Met
1070 1075 1080

Phe Ser Phe Gly Phe Asp Tyr Asn Asn Phe Asp Thr Tyr Asn Ile
1085 1090 1095

Thr Met Gly Lys Thr Gln Trp Thr Val Tyr Thr Asn Gly Glu Arg
1100 1105 1110

Leu Gln Ser Glu Phe Asn Asn Ala Arg Arg Thr Gly Lys Thr Lys
1115 1120 1125

Ser Ile Asn Leu Thr Glu Thr Ile Lys Leu Leu Leu Glu Asp Asn
1130 1135 1140

Glu Ile Asn Tyr Ala Asp Gly His Asp Ile Arg Ile Asp Met Glu
1145 1150 1155

Lys Met Asp Glu Asp Lys Lys Ser Glu Phe Phe Ala Gln Leu Leu
1160 1165 1170

Ser Leu Tyr Lys Leu Thr Val Gln Met Arg Asn Ser Tyr Thr Glu
1175 1180 1185

Ala Glu Glu Gln Glu Asn Gly Ile Ser Tyr Asp Lys Ile Ile Ser
1190 1195 1200

Pro Val Ile Asn Asp Glu Gly Glu Phe Phe Asp Ser Asp Asn Tyr
1205 1210 1215

Lys Glu Ser Asp Asp Lys Glu Cys Lys Met Pro Lys Asp Ala Asp
1220 1225 1230

Ala Asn Gly Ala Tyr Cys Ile Ala Leu Lys Gly Leu Tyr Glu Val
1235 1240 1245

Leu Lys Ile Lys Ser Glu Trp Thr Glu Asp Gly Phe Asp Arg Asn
1250 1255 1260

Cys Leu Lys Leu Pro His Ala Glu Trp Leu Asp Phe Ile Gln Asn
1265 1270 1275

Lys Arg Tyr Glu Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln
1280 1285 1290

Ala Lys Lys Lys Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr
1295 1300 1305

Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val
1310 1315 1320

Pro Asp Tyr Ala
1325

<210> 240
<211> 1418
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 240
Met Leu Phe Gln Asp Phe Thr His Leu Tyr Pro Leu Ser Lys Thr Val
1 5 10 15

Arg Phe Glu Leu Lys Pro Ile Asp Arg Thr Leu Glu His Ile His Ala
20 25 30

Lys Asn Phe Leu Ser Gln Asp Glu Thr Met Ala Asp Met His Gln Lys
35 40 45

Val Lys Val Ile Leu Asp Asp Tyr His Arg Asp Phe Ile Ala Asp Met
50 55 60

Met Gly Glu Val Lys Leu Thr Lys Leu Ala Glu Phe Tyr Asp Val Tyr
65 70 75 80

Leu Lys Phe Arg Lys Asn Pro Lys Asp Asp Glu Leu Gln Lys Gln Leu
85 90 95

Lys Asp Leu Gln Ala Val Leu Arg Lys Glu Ile Val Lys Pro Ile Gly

100 105 110

Asn Gly Gly Lys Tyr Lys Ala Gly Tyr Asp Arg Leu Phe Gly Ala Lys
115 120 125

Leu Phe Lys Asp Gly Lys Glu Leu Gly Asp Leu Ala Lys Phe Val Ile
130 135 140

Ala Gln Glu Gly Glu Ser Ser Pro Lys Leu Ala His Leu Ala His Phe
145 150 155 160

Glu Lys Phe Ser Thr Tyr Phe Thr Gly Phe His Asp Asn Arg Lys Asn
165 170 175

Met Tyr Ser Asp Glu Asp Lys His Thr Ala Ile Ala Tyr Arg Leu Ile
180 185 190

His Glu Asn Leu Pro Arg Phe Ile Asp Asn Leu Gln Ile Leu Thr Thr
195 200 205

Ile Lys Gln Lys His Ser Ala Leu Tyr Asp Gln Ile Ile Asn Glu Leu
210 215 220

Thr Ala Ser Gly Leu Asp Val Ser Leu Ala Ser His Leu Asp Gly Tyr
225 230 235 240

His Lys Leu Leu Thr Gln Glu Gly Ile Thr Ala Tyr Asn Thr Leu Leu
245 250 255

Gly Gly Ile Ser Gly Glu Ala Gly Ser Pro Lys Ile Gln Gly Ile Asn
260 265 270

Glu Leu Ile Asn Ser His His Asn Gln His Cys His Lys Ser Glu Arg
275 280 285

Ile Ala Lys Leu Arg Pro Leu His Lys Gln Ile Leu Ser Asp Gly Met
290 295 300

Ser Val Ser Phe Leu Pro Ser Lys Phe Ala Asp Asp Ser Glu Met Cys
305 310 315 320

Gln Ala Val Asn Glu Phe Tyr Arg His Tyr Ala Asp Val Phe Ala Lys
325 330 335

Val Gln Ser Leu Phe Asp Gly Phe Asp Asp His Gln Lys Asp Gly Ile
340 345 350

Tyr Val Glu His Lys Asn Leu Asn Glu Leu Ser Lys Gln Ala Phe Gly
355 360 365

Asp Phe Ala Leu Leu Gly Arg Val Leu Asp Gly Tyr Tyr Val Asp Val
370 375 380

Val Asn Pro Glu Phe Asn Glu Arg Phe Ala Lys Ala Lys Thr Asp Asn
385 390 395 400

Ala Lys Ala Lys Leu Thr Lys Glu Lys Asp Lys Phe Ile Lys Gly Val
405 410 415

His Ser Leu Ala Ser Leu Glu Gln Ala Ile Glu His Tyr Thr Ala Arg
420 425 430

His Asp Asp Glu Ser Val Gln Ala Gly Lys Leu Gly Gln Tyr Phe Lys

435 440 445

His Gly Leu Ala Gly Val Asp Asn Pro Ile Gln Lys Ile His Asn Asn
450 455 460

His Ser Thr Ile Lys Gly Phe Leu Glu Arg Glu Arg Pro Ala Gly Glu
465 470 475 480

Arg Ala Leu Pro Lys Ile Lys Ser Gly Lys Asn Pro Glu Met Thr Gln
485 490 495

Leu Arg Gln Leu Lys Glu Leu Leu Asp Asn Ala Leu Asn Val Ala His
500 505 510

Phe Ala Lys Leu Leu Thr Thr Lys Thr Thr Leu Asp Asn Gln Asp Gly
515 520 525

Asn Phe Tyr Gly Glu Phe Gly Val Leu Tyr Asp Glu Leu Ala Lys Ile
530 535 540

Pro Thr Leu Tyr Asn Lys Val Arg Asp Tyr Leu Ser Gln Lys Pro Phe
545 550 555 560

Ser Thr Glu Lys Tyr Lys Leu Asn Phe Gly Asn Pro Thr Leu Leu Asn
565 570 575

Gly Trp Asp Leu Asn Lys Glu Lys Asp Asn Phe Gly Val Ile Leu Gln
580 585 590

Lys Asp Gly Cys Tyr Tyr Leu Ala Leu Leu Asp Lys Ala His Lys Lys
595 600 605

Val Phe Asp Asn Ala Pro Asn Thr Gly Lys Ser Ile Tyr Gln Lys Met
610 615 620

Ile Tyr Lys Tyr Leu Glu Val Arg Lys Gln Phe Pro Lys Val Phe Phe
625 630 635 640

Ser Lys Glu Ala Ile Ala Ile Asn Tyr His Pro Ser Lys Glu Leu Val
645 650 655

Glu Ile Lys Asp Lys Gly Arg Gln Arg Ser Asp Asp Glu Arg Leu Lys
660 665 670

Leu Tyr Arg Phe Ile Leu Glu Cys Leu Lys Ile His Pro Lys Tyr Asp
675 680 685

Lys Lys Phe Glu Gly Ala Ile Gly Asp Ile Gln Leu Phe Lys Lys Asp
690 695 700

Lys Lys Gly Arg Glu Val Pro Ile Ser Glu Lys Asp Leu Phe Asp Lys
705 710 715 720

Ile Asn Gly Ile Phe Ser Ser Lys Pro Lys Leu Glu Met Glu Asp Phe
725 730 735

Phe Ile Gly Glu Phe Lys Arg Tyr Asn Pro Ser Gln Asp Leu Val Asp
740 745 750

Gln Tyr Asn Ile Tyr Lys Lys Ile Asp Ser Asn Asp Asn Arg Lys Lys
755 760 765

Glu Asn Phe Tyr Asn Asn His Pro Lys Phe Lys Lys Asp Leu Val Arg

770 775 780

Tyr Tyr Tyr Glu Ser Met Cys Lys His Glu Glu Trp Glu Glu Ser Phe
785 790 795 800

Glu Phe Ser Lys Lys Leu Gln Asp Ile Gly Cys Tyr Val Asp Val Asn
805 810 815

Glu Leu Phe Thr Glu Ile Glu Thr Arg Arg Leu Asn Tyr Lys Ile Ser
820 825 830

Phe Cys Asn Ile Asn Ala Asp Tyr Ile Asp Glu Leu Val Glu Gln Gly
835 840 845

Gln Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Pro Lys Ala
850 855 860

His Gly Lys Pro Asn Leu His Thr Leu Tyr Phe Lys Ala Leu Phe Ser
865 870 875 880

Glu Asp Asn Leu Ala Asp Pro Ile Tyr Lys Leu Asn Gly Glu Ala Gln
885 890 895

Ile Phe Tyr Arg Lys Ala Ser Leu Asp Met Asn Glu Thr Thr Ile His
900 905 910

Arg Ala Gly Glu Val Leu Glu Asn Lys Asn Pro Asp Asn Pro Lys Lys
915 920 925

Arg Gln Phe Val Tyr Asp Ile Ile Lys Asp Lys Arg Tyr Thr Gln Asp
930 935 940

Lys Phe Met Leu His Val Pro Ile Thr Met Asn Phe Gly Val Gln Gly
945 950 955 960

Met Thr Ile Lys Glu Phe Asn Lys Lys Val Asn Gln Ser Ile Gln Gln
965 970 975

Tyr Asp Glu Val Asn Val Ile Gly Ile Asp Arg Gly Glu Arg His Leu
980 985 990

Leu Tyr Leu Thr Val Ile Asn Ser Lys Gly Glu Ile Leu Glu Gln Cys
995 1000 1005

Ser Leu Asn Asp Ile Thr Thr Ala Ser Ala Asn Gly Thr Gln Met
1010 1015 1020

Thr Thr Pro Tyr His Lys Ile Leu Asp Lys Arg Glu Ile Glu Arg
1025 1030 1035

Leu Asn Ala Arg Val Gly Trp Gly Glu Ile Glu Thr Ile Lys Glu
1040 1045 1050

Leu Lys Ser Gly Tyr Leu Ser His Val Val His Gln Ile Ser Gln
1055 1060 1065

Leu Met Leu Lys Tyr Asn Ala Ile Val Val Leu Glu Asp Leu Asn
1070 1075 1080

Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Ile Tyr
1085 1090 1095

Gln Asn Phe Glu Asn Ala Leu Ile Lys Lys Leu Asn His Leu Val

1100 1105 1110

Leu Lys Asp Lys Ala Asp Asp Glu Ile Gly Ser Tyr Lys Asn Ala
1115 1120 1125

Leu Gln Leu Thr Asn Asn Phe Thr Asp Leu Lys Ser Ile Gly Lys
1130 1135 1140

Gln Thr Gly Phe Leu Phe Tyr Val Pro Ala Trp Asn Thr Ser Lys
1145 1150 1155

Ile Asp Pro Glu Thr Gly Phe Val Asp Leu Leu Lys Pro Arg Tyr
1160 1165 1170

Glu Asn Ile Ala Gln Ser Gln Ala Phe Phe Gly Lys Phe Asp Lys
1175 1180 1185

Ile Cys Tyr Asn Ala Asp Lys Asp Tyr Phe Glu Phe His Ile Asp
1190 1195 1200

Tyr Ala Lys Phe Thr Asp Lys Ala Lys Asn Ser Arg Gln Ile Trp
1205 1210 1215

Thr Ile Cys Ser His Gly Asp Lys Arg Tyr Val Tyr Asp Lys Thr
1220 1225 1230

Ala Asn Gln Asn Lys Gly Ala Ala Lys Gly Ile Asn Val Asn Asp
1235 1240 1245

Glu Leu Lys Ser Leu Phe Ala Arg His His Ile Asn Glu Lys Gln
1250 1255 1260

Pro Asn Leu Val Met Asp Ile Cys Gln Asn Asn Asp Lys Glu Phe
1265 1270 1275

His Lys Ser Leu Met Tyr Leu Leu Lys Thr Leu Leu Ala Leu Arg
1280 1285 1290

Tyr Ser Asn Ala Ser Ser Asp Glu Asp Phe Ile Leu Ser Pro Val
1295 1300 1305

Ala Asn Asp Glu Gly Val Phe Phe Asn Ser Ala Leu Ala Asp Asp
1310 1315 1320

Thr Gln Pro Gln Asn Ala Asp Ala Asn Gly Ala Tyr His Ile Ala
1325 1330 1335

Leu Lys Gly Leu Trp Leu Leu Asn Glu Leu Lys Asn Ser Asp Asp
1340 1345 1350

Leu Asn Lys Val Lys Leu Ala Ile Asp Asn Gln Thr Trp Leu Asn
1355 1360 1365

Phe Ala Gln Asn Arg Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly
1370 1375 1380

Gln Ala Lys Lys Lys Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp
1385 1390 1395

Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp
1400 1405 1410

Val Pro Asp Tyr Ala

1415

<210> 241
<211> 1308
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 241
Met Glu Asp Tyr Ser Gly Phe Val Asn Ile Tyr Ser Ile Gln Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr Leu Glu His Ile Glu
20 25 30

Lys Lys Gly Phe Leu Lys Lys Asp Lys Ile Arg Ala Glu Asp Tyr Lys
35 40 45

Ala Val Lys Lys Ile Ile Asp Lys Tyr His Arg Ala Tyr Ile Glu Glu
50 55 60

Val Phe Asp Ser Val Leu His Gln Lys Lys Lys Lys Asp Lys Thr Arg
65 70 75 80

Phe Ser Thr Gln Phe Ile Lys Glu Ile Lys Glu Phe Ser Glu Leu Tyr
85 90 95

Tyr Lys Thr Glu Lys Asn Ile Pro Asp Lys Glu Arg Leu Glu Ala Leu
100 105 110

Ser Glu Lys Leu Arg Lys Met Leu Val Gly Ala Phe Lys Gly Glu Phe
115 120 125

Ser Glu Glu Val Ala Glu Lys Tyr Lys Asn Leu Phe Ser Lys Glu Leu
130 135 140

Ile Arg Asn Glu Ile Glu Lys Phe Cys Glu Thr Asp Glu Glu Arg Lys
145 150 155 160

Gln Val Ser Asn Phe Lys Ser Phe Thr Thr Tyr Phe Thr Gly Phe His
165 170 175

Ser Asn Arg Gln Asn Ile Tyr Ser Asp Glu Lys Lys Ser Thr Ala Ile
180 185 190

Gly Tyr Arg Ile Ile His Gln Asn Leu Pro Lys Phe Leu Asp Asn Leu
195 200 205

Lys Ile Ile Glu Ser Ile Gln Arg Arg Phe Lys Asp Phe Pro Trp Ser
210 215 220

Asp Leu Lys Lys Asn Leu Lys Lys Ile Asp Lys Asn Ile Lys Leu Thr
225 230 235 240

Glu Tyr Phe Ser Ile Asp Gly Phe Val Asn Val Leu Asn Gln Lys Gly
245 250 255

Ile Asp Ala Tyr Asn Thr Ile Leu Gly Gly Lys Ser Glu Glu Ser Gly
260 265 270

Glu Lys Ile Gln Gly Leu Asn Glu Tyr Ile Asn Leu Tyr Arg Gln Lys
275 280 285

```
Asn Asn Ile Asp Arg Lys Asn Leu Pro Asn Val Lys Ile Leu Phe Lys
    290                 295                 300

Gln Ile Leu Gly Asp Arg Glu Thr Lys Ser Phe Ile Pro Glu Ala Phe
305                 310                 315                 320

Pro Asp Asp Gln Ser Val Leu Asn Ser Ile Thr Glu Phe Ala Lys Tyr
                325                 330                 335

Leu Lys Leu Asp Lys Lys Lys Lys Ser Ile Ile Ala Glu Leu Lys Lys
            340                 345                 350

Phe Leu Ser Ser Phe Asn Arg Tyr Glu Leu Asp Gly Ile Tyr Leu Ala
        355                 360                 365

Asn Asp Asn Ser Leu Ala Ser Ile Ser Thr Phe Leu Phe Asp Asp Trp
    370                 375                 380

Ser Phe Ile Lys Lys Ser Val Ser Phe Lys Tyr Asp Glu Ser Val Gly
385                 390                 395                 400

Asp Pro Lys Lys Lys Ile Lys Ser Pro Leu Lys Tyr Glu Lys Glu Lys
                405                 410                 415

Glu Lys Trp Leu Lys Gln Lys Tyr Tyr Thr Ile Ser Phe Leu Asn Asp
            420                 425                 430

Ala Ile Glu Ser Tyr Ser Lys Ser Gln Asp Glu Lys Arg Val Lys Ile
        435                 440                 445

Arg Leu Glu Ala Tyr Phe Ala Glu Phe Lys Ser Lys Asp Asp Ala Lys
    450                 455                 460

Lys Gln Phe Asp Leu Leu Glu Arg Ile Glu Glu Ala Tyr Ala Ile Val
465                 470                 475                 480

Glu Pro Leu Leu Gly Ala Glu Tyr Pro Arg Asp Arg Asn Leu Lys Ala
                485                 490                 495

Asp Lys Lys Glu Val Gly Lys Ile Lys Asp Phe Leu Asp Ser Ile Lys
            500                 505                 510

Ser Leu Gln Phe Phe Leu Lys Pro Leu Leu Ser Ala Glu Ile Phe Asp
        515                 520                 525

Glu Lys Asp Leu Gly Phe Tyr Asn Gln Leu Glu Gly Tyr Tyr Glu Glu
    530                 535                 540

Ile Asp Ser Ile Gly His Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr
545                 550                 555                 560

Gly Lys Ile Tyr Ser Lys Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser
                565                 570                 575

Thr Leu Leu Lys Gly Trp Asp Glu Asn Arg Glu Val Ala Asn Leu Cys
            580                 585                 590

Val Ile Phe Arg Glu Asp Gln Lys Tyr Tyr Leu Gly Val Met Asp Lys
        595                 600                 605

Glu Asn Asn Thr Ile Leu Ser Asp Ile Pro Lys Val Lys Pro Asn Glu
    610                 615                 620
```

```
Leu Phe Tyr Glu Lys Met Val Tyr Lys Leu Ile Pro Thr Pro His Met
625                 630                 635                 640

Gln Leu Pro Arg Ile Ile Phe Ser Ser Asp Asn Leu Ser Ile Tyr Asn
                645                 650                 655

Pro Ser Lys Ser Ile Leu Lys Ile Arg Glu Ala Lys Ser Phe Lys Glu
            660                 665                 670

Gly Lys Asn Phe Lys Leu Lys Asp Cys His Lys Phe Ile Asp Phe Tyr
        675                 680                 685

Lys Glu Ser Ile Ser Lys Asn Glu Asp Trp Ser Arg Phe Asp Phe Lys
    690                 695                 700

Phe Ser Lys Thr Ser Ser Tyr Glu Asn Ile Ser Glu Phe Tyr Arg Glu
705                 710                 715                 720

Val Glu Arg Gln Gly Tyr Asn Leu Asp Phe Lys Lys Val Ser Lys Phe
                725                 730                 735

Tyr Ile Asp Ser Leu Val Glu Asp Gly Lys Leu Tyr Leu Phe Gln Ile
            740                 745                 750

Tyr Asn Lys Asp Phe Ser Ile Phe Ser Lys Gly Lys Pro Asn Leu His
        755                 760                 765

Thr Ile Tyr Phe Arg Ser Leu Phe Ser Lys Glu Asn Leu Lys Asp Val
    770                 775                 780

Cys Leu Lys Leu Asn Gly Glu Ala Glu Met Phe Phe Arg Lys Lys Ser
785                 790                 795                 800

Ile Asn Tyr Asp Glu Lys Lys Lys Arg Glu Gly His His Pro Glu Leu
                805                 810                 815

Phe Glu Lys Leu Lys Tyr Pro Ile Leu Lys Asp Lys Arg Tyr Ser Glu
            820                 825                 830

Asp Lys Phe Gln Phe His Leu Pro Ile Ser Leu Asn Phe Lys Ser Lys
        835                 840                 845

Glu Arg Leu Asn Phe Asn Leu Lys Val Asn Glu Phe Leu Lys Arg Asn
    850                 855                 860

Lys Asp Ile Asn Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu Leu
865                 870                 875                 880

Tyr Leu Val Met Ile Asn Gln Lys Gly Glu Ile Leu Lys Gln Thr Leu
                885                 890                 895

Leu Asp Ser Met Gln Ser Gly Lys Gly Arg Pro Glu Ile Asn Tyr Lys
            900                 905                 910

Glu Lys Leu Gln Glu Lys Glu Ile Glu Arg Asp Lys Ala Arg Lys Ser
        915                 920                 925

Trp Gly Thr Val Glu Asn Ile Lys Glu Leu Lys Glu Gly Tyr Leu Ser
    930                 935                 940

Ile Val Ile His Gln Ile Ser Lys Leu Met Val Glu Asn Asn Ala Ile
945                 950                 955                 960
```

Val Val Leu Glu Asp Leu Asn Ile Gly Phe Lys Arg Gly Arg Gln Lys
965 970 975

Val Glu Arg Gln Val Tyr Gln Lys Phe Glu Lys Met Leu Ile Asp Lys
980 985 990

Leu Asn Phe Leu Val Phe Lys Glu Asn Lys Pro Thr Glu Pro Gly Gly
995 1000 1005

Val Leu Lys Ala Tyr Gln Leu Thr Asp Glu Phe Gln Ser Phe Glu
1010 1015 1020

Lys Leu Ser Lys Gln Thr Gly Phe Leu Phe Tyr Val Pro Ser Trp
1025 1030 1035

Asn Thr Ser Lys Ile Asp Pro Arg Thr Gly Phe Ile Asp Phe Leu
1040 1045 1050

His Pro Ala Tyr Glu Asn Ile Glu Lys Ala Lys Gln Trp Ile Asn
1055 1060 1065

Lys Phe Asp Ser Ile Arg Phe Asn Ser Lys Met Asp Trp Phe Glu
1070 1075 1080

Phe Thr Ala Asp Thr Arg Lys Phe Ser Glu Asn Leu Met Leu Gly
1085 1090 1095

Lys Asn Arg Val Trp Val Ile Cys Thr Thr Asn Val Glu Arg Tyr
1100 1105 1110

Phe Thr Ser Lys Thr Ala Asn Ser Ser Ile Gln Tyr Asn Ser Ile
1115 1120 1125

Gln Ile Thr Glu Lys Leu Lys Glu Leu Phe Val Asp Ile Pro Phe
1130 1135 1140

Ser Asn Gly Gln Asp Leu Lys Pro Glu Ile Leu Arg Lys Asn Asp
1145 1150 1155

Ala Val Phe Phe Lys Ser Leu Leu Phe Tyr Ile Lys Thr Thr Leu
1160 1165 1170

Ser Leu Arg Gln Asn Asn Gly Lys Lys Gly Glu Glu Glu Lys Asp
1175 1180 1185

Phe Ile Leu Ser Pro Val Val Asp Ser Lys Gly Arg Phe Phe Asn
1190 1195 1200

Ser Leu Glu Ala Ser Asp Asp Glu Pro Lys Asp Ala Asp Ala Asn
1205 1210 1215

Gly Ala Tyr His Ile Ala Leu Lys Gly Leu Met Asn Leu Leu Val
1220 1225 1230

Leu Asn Glu Thr Lys Glu Glu Asn Leu Ser Arg Pro Lys Trp Lys
1235 1240 1245

Ile Lys Asn Lys Asp Trp Leu Glu Phe Val Trp Glu Arg Asn Arg
1250 1255 1260

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys
1265 1270 1275

Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr
1280 1285 1290

Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1295 1300 1305

<210> 242
<211> 1273
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 242
Met Ser Lys Leu Glu Lys Phe Thr Asn Cys Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Arg Phe Lys Ala Ile Pro Val Gly Lys Thr Gln Glu Asn Ile Asp
20 25 30

Asn Lys Arg Leu Leu Val Glu Asp Glu Lys Arg Ala Glu Asp Tyr Lys
35 40 45

Gly Val Lys Lys Leu Leu Asp Arg Tyr Tyr Leu Ser Phe Ile Asn Asp
50 55 60

Val Leu His Ser Ile Lys Leu Lys Asn Leu Asn Asn Tyr Ile Ser Leu
65 70 75 80

Phe Arg Lys Lys Thr Arg Thr Glu Lys Glu Asn Lys Glu Leu Glu Asn
85 90 95

Leu Glu Ile Asn Leu Arg Lys Glu Ile Ala Lys Ala Phe Lys Gly Asn
100 105 110

Glu Gly Tyr Lys Ser Leu Phe Lys Lys Asp Ile Ile Glu Thr Ile Leu
115 120 125

Pro Glu Phe Leu Asp Asp Lys Asp Glu Ile Ala Leu Val Asn Ser Phe
130 135 140

Asn Gly Phe Thr Thr Ala Phe Thr Gly Phe Phe Asp Asn Arg Glu Asn
145 150 155 160

Met Phe Ser Glu Glu Ala Lys Ser Thr Ser Ile Ala Phe Arg Cys Ile
165 170 175

Asn Glu Asn Leu Thr Arg Tyr Ile Ser Asn Met Asp Ile Phe Glu Lys
180 185 190

Val Asp Ala Ile Phe Asp Lys His Glu Val Gln Glu Ile Lys Glu Lys
195 200 205

Ile Leu Asn Ser Asp Tyr Asp Val Glu Asp Phe Phe Glu Gly Glu Phe
210 215 220

Phe Asn Phe Val Leu Thr Gln Glu Gly Ile Asp Val Tyr Asn Ala Ile
225 230 235 240

Ile Gly Gly Phe Val Thr Glu Ser Gly Glu Lys Ile Lys Gly Leu Asn
245 250 255

Glu Tyr Ile Asn Leu Tyr Asn Gln Lys Thr Lys Gln Lys Leu Pro Lys

260 265 270

Phe Lys Pro Leu Tyr Lys Gln Val Leu Ser Asp Arg Glu Ser Leu Ser
275 280 285

Phe Tyr Gly Glu Gly Tyr Thr Ser Asp Glu Glu Val Leu Glu Val Phe
290 295 300

Arg Asn Thr Leu Asn Lys Asn Ser Glu Ile Phe Ser Ser Ile Lys Lys
305 310 315 320

Leu Glu Lys Leu Phe Lys Asn Phe Asp Glu Tyr Ser Ser Ala Gly Ile
325 330 335

Phe Val Lys Asn Gly Pro Ala Ile Ser Thr Ile Ser Lys Asp Ile Phe
340 345 350

Gly Glu Trp Asn Val Ile Arg Asp Lys Trp Asn Ala Glu Tyr Asp Asp
355 360 365

Ile His Leu Lys Lys Lys Ala Val Val Thr Glu Lys Tyr Glu Asp Asp
370 375 380

Arg Arg Lys Ser Phe Lys Lys Ile Gly Ser Phe Ser Leu Glu Gln Leu
385 390 395 400

Gln Glu Tyr Ala Asp Ala Asp Leu Ser Val Val Glu Lys Leu Lys Glu
405 410 415

Ile Ile Ile Gln Lys Val Asp Glu Ile Tyr Lys Val Tyr Gly Ser Ser
420 425 430

Glu Lys Leu Phe Asp Ala Asp Phe Val Leu Glu Lys Ser Leu Lys Lys
435 440 445

Asn Asp Ala Val Val Ala Ile Met Lys Asp Leu Leu Asp Ser Val Lys
450 455 460

Ser Phe Glu Asn Tyr Ile Lys Ala Phe Phe Gly Glu Gly Lys Glu Thr
465 470 475 480

Asn Arg Asp Glu Ser Phe Tyr Gly Asp Phe Val Leu Ala Tyr Asp Ile
485 490 495

Leu Leu Lys Val Asp His Ile Tyr Asp Ala Ile Arg Asn Tyr Val Thr
500 505 510

Gln Lys Pro Tyr Ser Lys Asp Lys Phe Lys Leu Tyr Phe Gln Asn Pro
515 520 525

Gln Phe Met Gly Gly Trp Asp Lys Asp Lys Glu Thr Asp Tyr Arg Ala
530 535 540

Thr Ile Leu Arg Tyr Gly Ser Lys Tyr Tyr Leu Ala Ile Met Asp Lys
545 550 555 560

Lys Tyr Ala Lys Cys Leu Gln Lys Ile Asp Lys Asp Asp Val Asn Gly
565 570 575

Asn Tyr Glu Lys Ile Asn Tyr Lys Leu Leu Pro Gly Pro Asn Lys Met
580 585 590

Leu Pro Lys Val Phe Phe Ser Lys Lys Trp Met Ala Tyr Tyr Asn Pro

595 600 605

Ser Glu Asp Ile Gln Lys Ile Tyr Lys Asn Gly Thr Phe Lys Lys Gly
610 615 620

Asp Met Phe Asn Leu Asn Asp Cys His Lys Leu Ile Asp Phe Phe Lys
625 630 635 640

Asp Ser Ile Ser Arg Tyr Pro Lys Trp Ser Asn Ala Tyr Asp Phe Asn
645 650 655

Phe Ser Glu Thr Glu Lys Tyr Lys Asp Ile Ala Gly Phe Tyr Arg Glu
660 665 670

Val Glu Glu Gln Gly Tyr Lys Val Ser Phe Glu Ser Ala Ser Lys Lys
675 680 685

Glu Val Asp Lys Leu Val Glu Glu Gly Lys Leu Tyr Met Phe Gln Ile
690 695 700

Tyr Asn Lys Asp Phe Ser Asp Lys Ser His Gly Thr Pro Asn Leu His
705 710 715 720

Thr Met Tyr Phe Lys Leu Leu Phe Asp Glu Asn Asn His Gly Gln Ile
725 730 735

Arg Leu Ser Gly Gly Ala Glu Leu Phe Met Arg Arg Ala Ser Leu Lys
740 745 750

Lys Glu Glu Leu Val Val His Pro Ala Asn Ser Pro Ile Ala Asn Lys
755 760 765

Asn Pro Asp Asn Pro Lys Lys Thr Thr Thr Leu Ser Tyr Asp Val Tyr
770 775 780

Lys Asp Lys Arg Phe Ser Glu Asp Gln Tyr Glu Leu His Ile Pro Ile
785 790 795 800

Ala Ile Asn Lys Cys Pro Lys Asn Ile Phe Lys Ile Asn Thr Glu Val
805 810 815

Arg Val Leu Leu Lys His Asp Asp Asn Pro Tyr Val Ile Gly Ile Asp
820 825 830

Arg Gly Glu Arg Asn Leu Leu Tyr Ile Val Val Val Asp Gly Lys Gly
835 840 845

Asn Ile Val Glu Gln Tyr Ser Leu Asn Glu Ile Ile Asn Asn Phe Asn
850 855 860

Gly Ile Arg Ile Lys Thr Asp Tyr His Ser Leu Leu Asp Lys Lys Glu
865 870 875 880

Lys Glu Arg Phe Glu Ala Arg Gln Asn Trp Thr Ser Ile Glu Asn Ile
885 890 895

Lys Glu Leu Lys Ala Gly Tyr Ile Ser Gln Val Val His Lys Ile Cys
900 905 910

Glu Leu Val Glu Lys Tyr Asp Ala Val Ile Ala Leu Glu Asp Leu Asn
915 920 925

Ser Gly Phe Lys Asn Ser Arg Val Lys Val Glu Lys Gln Val Tyr Gln

930 935 940

Lys Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr Met Val Asp Lys
945 950 955 960

Lys Ser Asn Pro Cys Ala Thr Gly Gly Ala Leu Lys Gly Tyr Gln Ile
965 970 975

Thr Asn Lys Phe Glu Ser Phe Lys Ser Met Ser Thr Gln Asn Gly Phe
980 985 990

Ile Phe Tyr Ile Pro Ala Trp Leu Thr Ser Lys Ile Asp Pro Ser Thr
995 1000 1005

Gly Phe Val Asn Leu Leu Lys Thr Lys Tyr Thr Ser Ile Ala Asp
1010 1015 1020

Ser Lys Lys Phe Ile Ser Ser Phe Asp Arg Ile Met Tyr Val Pro
1025 1030 1035

Glu Glu Asp Leu Phe Glu Phe Ala Leu Asp Tyr Lys Asn Phe Ser
1040 1045 1050

Arg Thr Asp Ala Asp Tyr Ile Lys Lys Trp Lys Leu Tyr Ser Tyr
1055 1060 1065

Gly Asn Arg Ile Arg Ile Phe Arg Asn Pro Lys Lys Asn Asn Val
1070 1075 1080

Phe Asp Trp Glu Glu Val Cys Leu Thr Ser Ala Tyr Lys Glu Leu
1085 1090 1095

Phe Asn Lys Tyr Gly Ile Asn Tyr Gln Gln Gly Asp Ile Arg Ala
1100 1105 1110

Leu Leu Cys Glu Gln Ser Asp Lys Ala Phe Tyr Ser Ser Phe Met
1115 1120 1125

Ala Leu Met Ser Leu Met Leu Gln Met Arg Asn Ser Ile Thr Gly
1130 1135 1140

Arg Thr Asp Val Asp Phe Leu Ile Ser Pro Val Lys Asn Ser Asp
1145 1150 1155

Gly Ile Phe Tyr Asp Ser Arg Asn Tyr Glu Ala Gln Glu Asn Ala
1160 1165 1170

Ile Leu Pro Lys Asn Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala
1175 1180 1185

Arg Lys Val Leu Trp Ala Ile Gly Gln Phe Lys Lys Ala Glu Asp
1190 1195 1200

Glu Lys Leu Asp Lys Val Lys Ile Ala Ile Ser Asn Lys Glu Trp
1205 1210 1215

Leu Glu Tyr Ala Gln Thr Ser Val Lys His Lys Arg Pro Ala Ala
1220 1225 1230

Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys Gly Ser Tyr Pro
1235 1240 1245

Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr

1250 1255 1260

Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1265 1270

<210> 243
<211> 1305
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 243
Met Asp Ser Leu Lys Asp Phe Thr Asn Leu Tyr Pro Val Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr Leu Glu Asn Ile Glu
20 25 30

Lys Ala Gly Ile Leu Lys Glu Asp Glu His Arg Ala Glu Ser Tyr Arg
35 40 45

Arg Val Lys Lys Ile Ile Asp Thr Tyr His Lys Val Phe Ile Asp Ser
50 55 60

Ser Leu Glu Asn Met Ala Lys Met Gly Ile Glu Asn Glu Ile Lys Ala
65 70 75 80

Met Leu Gln Ser Phe Cys Glu Leu Tyr Lys Lys Asp His Arg Thr Glu
85 90 95

Gly Glu Asp Lys Ala Leu Asp Lys Ile Arg Ala Val Leu Arg Gly Leu
100 105 110

Ile Val Gly Ala Phe Thr Gly Val Cys Gly Arg Arg Glu Asn Thr Val
115 120 125

Gln Asn Glu Lys Tyr Glu Ser Leu Phe Lys Glu Lys Leu Ile Lys Glu
130 135 140

Ile Leu Pro Asp Phe Val Leu Ser Thr Glu Ala Glu Ser Leu Pro Phe
145 150 155 160

Ser Val Glu Glu Ala Thr Arg Ser Leu Lys Glu Phe Asp Ser Phe Thr
165 170 175

Ser Tyr Phe Ala Gly Phe Tyr Glu Asn Arg Lys Asn Ile Tyr Ser Thr
180 185 190

Lys Pro Gln Ser Thr Ala Ile Ala Tyr Arg Leu Ile His Glu Asn Leu
195 200 205

Pro Lys Phe Ile Asp Asn Ile Leu Val Phe Gln Lys Ile Lys Glu Pro
210 215 220

Ile Ala Lys Glu Leu Glu His Ile Arg Ala Asp Phe Ser Ala Gly Gly
225 230 235 240

Tyr Ile Lys Lys Asp Glu Arg Leu Glu Asp Ile Phe Ser Leu Asn Tyr
245 250 255

Tyr Ile His Val Leu Ser Gln Ala Gly Ile Glu Lys Tyr Asn Ala Leu
260 265 270

Ile Gly Lys Ile Val Thr Glu Gly Asp Gly Glu Met Lys Gly Leu Asn
275 280 285

Glu His Ile Asn Leu Tyr Asn Gln Gln Arg Gly Arg Glu Asp Arg Leu
290 295 300

Pro Leu Phe Arg Pro Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Gln
305 310 315 320

Leu Ser Tyr Leu Pro Glu Ser Phe Glu Lys Asp Glu Glu Leu Leu Arg
325 330 335

Ala Leu Lys Glu Phe Tyr Asp His Ile Ala Glu Asp Ile Leu Gly Arg
340 345 350

Thr Gln Gln Leu Met Thr Ser Ile Ser Glu Tyr Asp Leu Ser Arg Ile
355 360 365

Tyr Val Arg Asn Asp Ser Gln Leu Thr Asp Ile Ser Lys Lys Met Leu
370 375 380

Gly Asp Trp Asn Ala Ile Tyr Met Ala Arg Glu Arg Ala Tyr Asp His
385 390 395 400

Glu Gln Ala Pro Lys Arg Ile Thr Ala Lys Tyr Glu Arg Asp Arg Ile
405 410 415

Lys Ala Leu Lys Gly Glu Glu Ser Ile Ser Leu Ala Asn Leu Asn Ser
420 425 430

Cys Ile Ala Phe Leu Asp Asn Val Arg Asp Cys Arg Val Asp Thr Tyr
435 440 445

Leu Ser Thr Leu Gly Gln Lys Glu Gly Pro His Gly Leu Ser Asn Leu
450 455 460

Val Glu Asn Val Phe Ala Ser Tyr His Glu Ala Glu Gln Leu Leu Ser
465 470 475 480

Phe Pro Tyr Pro Glu Glu Asn Asn Leu Ile Gln Asp Lys Asp Asn Val
485 490 495

Val Leu Ile Lys Asn Leu Leu Asp Asn Ile Ser Asp Leu Gln Arg Phe
500 505 510

Leu Lys Pro Leu Trp Gly Met Gly Asp Glu Pro Asp Lys Asp Glu Arg
515 520 525

Phe Tyr Gly Glu Tyr Asn Tyr Ile Arg Gly Ala Leu Asp Gln Val Ile
530 535 540

Pro Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Arg Lys Pro Tyr Ser
545 550 555 560

Thr Arg Lys Val Lys Leu Asn Phe Gly Asn Ser Gln Leu Leu Ser Gly
565 570 575

Trp Asp Arg Asn Lys Glu Lys Asp Asn Ser Cys Val Ile Leu Arg Lys
580 585 590

Gly Gln Asn Phe Tyr Leu Ala Ile Met Asn Asn Arg His Lys Arg Ser
595 600 605

Phe Glu Asn Lys Met Leu Pro Glu Tyr Lys Glu Gly Glu Pro Tyr Phe
610 615 620

Glu Lys Met Asp Tyr Lys Phe Leu Pro Asp Pro Asn Lys Met Leu Pro
625 630 635 640

Lys Val Phe Leu Ser Lys Lys Gly Ile Glu Ile Tyr Lys Pro Ser Pro
645 650 655

Lys Leu Leu Glu Gln Tyr Gly His Gly Thr His Lys Lys Gly Asp Thr
660 665 670

Phe Ser Met Asp Asp Leu His Glu Leu Ile Asp Phe Phe Lys His Ser
675 680 685

Ile Glu Ala His Glu Asp Trp Lys Gln Phe Gly Phe Lys Phe Ser Asp
690 695 700

Thr Ala Thr Tyr Glu Asn Val Ser Ser Phe Tyr Arg Glu Val Glu Asp
705 710 715 720

Gln Gly Tyr Lys Leu Ser Phe Arg Lys Val Ser Glu Ser Tyr Val Tyr
725 730 735

Ser Leu Ile Asp Gln Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys
740 745 750

Asp Phe Ser Pro Cys Ser Lys Gly Thr Pro Asn Leu His Thr Leu Tyr
755 760 765

Trp Arg Met Leu Phe Asp Glu Arg Asn Leu Ala Asp Val Ile Tyr Lys
770 775 780

Leu Asp Gly Lys Ala Glu Ile Phe Phe Arg Glu Lys Ser Leu Lys Asn
785 790 795 800

Asp His Pro Thr His Pro Ala Gly Lys Pro Ile Lys Lys Lys Ser Arg
805 810 815

Gln Lys Lys Gly Glu Glu Ser Leu Phe Glu Tyr Asp Leu Val Lys Asp
820 825 830

Arg Arg Tyr Thr Met Asp Lys Phe Gln Phe His Val Pro Ile Thr Met
835 840 845

Asn Phe Lys Cys Ser Ala Gly Ser Lys Val Asn Asp Met Val Asn Ala
850 855 860

His Ile Arg Glu Ala Lys Asp Met His Val Ile Gly Ile Asp Arg Gly
865 870 875 880

Glu Arg Asn Leu Leu Tyr Ile Cys Val Ile Asp Ser Arg Gly Thr Ile
885 890 895

Leu Asp Gln Ile Ser Leu Asn Thr Ile Asn Asp Ile Asp Tyr His Asp
900 905 910

Leu Leu Glu Ser Arg Asp Lys Asp Arg Gln Gln Glu His Arg Asn Trp
915 920 925

Gln Thr Ile Glu Gly Ile Lys Glu Leu Lys Gln Gly Tyr Leu Ser Gln
930 935 940

```
Ala Val His Arg Ile Ala Glu Leu Met Val Ala Tyr Lys Ala Val Val
945                 950                 955                 960

Ala Leu Glu Asp Leu Asn Met Gly Phe Lys Arg Gly Arg Gln Lys Val
                965                 970                 975

Glu Ser Ser Val Tyr Gln Gln Phe Glu Lys Gln Leu Ile Asp Lys Leu
            980                 985                 990

Asn Tyr Leu Val Asp Lys Lys Lys Arg Pro Glu Asp Ile Gly Gly Leu
        995                1000                1005

Leu Arg  Ala Tyr Gln Phe Thr  Ala Pro Phe Lys Ser  Phe Lys Glu
    1010                1015                1020

Met Gly  Lys Gln Asn Gly Phe  Leu Phe Tyr Ile Pro  Ala Trp Asn
    1025                1030                1035

Thr Ser  Asn Ile Asp Pro Thr  Thr Gly Phe Val Asn  Leu Phe His
    1040                1045                1050

Val Gln  Tyr Glu Asn Val Asp  Lys Ala Lys Ser Phe  Phe Gln Lys
    1055                1060                1065

Phe Asp  Ser Ile Ser Tyr Asn  Pro Lys Lys Asp Trp  Phe Glu Phe
    1070                1075                1080

Ala Phe  Asp Tyr Lys Asn Phe  Thr Lys Lys Ala Glu  Gly Ser Arg
    1085                1090                1095

Ser Met  Trp Ile Leu Cys Thr  His Gly Ser Arg Ile  Lys Asn Phe
    1100                1105                1110

Arg Asn  Ser Gln Lys Asn Gly  Gln Trp Asp Ser Glu  Glu Phe Ala
    1115                1120                1125

Leu Thr  Glu Ala Phe Lys Ser  Leu Phe Val Arg Tyr  Glu Ile Asp
    1130                1135                1140

Tyr Thr  Ala Asp Leu Lys Thr  Ala Ile Val Asp Glu  Lys Gln Lys
    1145                1150                1155

Asp Phe  Phe Val Asp Leu Leu  Lys Leu Phe Lys Leu  Thr Val Gln
    1160                1165                1170

Met Arg  Asn Ser Trp Lys Glu  Lys Asp Leu Asp Tyr  Leu Ile Ser
    1175                1180                1185

Pro Val  Ala Gly Ala Asp Gly  Arg Phe Phe Asp Thr  Arg Glu Gly
    1190                1195                1200

Asn Lys  Ser Leu Pro Lys Asp  Ala Asp Ala Asn Gly  Ala Tyr Asn
    1205                1210                1215

Ile Ala  Leu Lys Gly Leu Trp  Ala Leu Arg Gln Ile  Arg Gln Thr
    1220                1225                1230

Ser Glu  Gly Gly Lys Leu Lys  Leu Ala Ile Ser Asn  Lys Glu Trp
    1235                1240                1245

Leu Gln  Phe Val Gln Glu Arg  Ser Tyr Glu Lys Asp  Lys Arg Pro
    1250                1255                1260
```

Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys Gly Ser
1265 1270 1275

Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro
1280 1285 1290

Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1295 1300 1305

<210> 244
<211> 1368
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 244
Met Glu Asn Tyr Gln Glu Phe Thr Asn Leu Phe Gln Leu Asn Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Cys Glu Leu Leu Glu
20 25 30

Glu Gly Lys Ile Phe Ala Ser Gly Ser Phe Leu Glu Lys Asp Lys Val
35 40 45

Arg Ala Asp Asn Val Ser Tyr Val Lys Lys Glu Ile Asp Lys Lys His
50 55 60

Lys Ile Phe Ile Glu Glu Thr Leu Ser Ser Phe Ser Ile Ser Asn Asp
65 70 75 80

Leu Leu Lys Gln Tyr Phe Asp Cys Tyr Asn Glu Leu Lys Ala Phe Lys
85 90 95

Lys Asp Cys Lys Ser Asp Glu Glu Glu Val Lys Lys Thr Ala Leu Arg
100 105 110

Asn Lys Cys Thr Ser Ile Gln Arg Ala Met Arg Glu Ala Ile Ser Gln
115 120 125

Ala Phe Leu Lys Ser Pro Gln Lys Lys Leu Leu Ala Ile Lys Asn Leu
130 135 140

Ile Glu Asn Val Phe Lys Ala Asp Glu Asn Val Gln His Phe Ser Glu
145 150 155 160

Phe Thr Ser Tyr Phe Ser Gly Phe Glu Thr Asn Arg Glu Asn Phe Tyr
165 170 175

Ser Asp Glu Glu Lys Ser Thr Ser Ile Ala Tyr Arg Leu Val His Asp
180 185 190

Asn Leu Pro Ile Phe Ile Lys Asn Ile Tyr Ile Phe Glu Lys Leu Lys
195 200 205

Glu Gln Phe Asp Ala Lys Thr Leu Ser Glu Ile Phe Glu Asn Tyr Lys
210 215 220

Leu Tyr Val Ala Gly Ser Ser Leu Asp Glu Val Phe Ser Leu Glu Tyr
225 230 235 240

Phe Asn Asn Thr Leu Thr Gln Lys Gly Ile Asp Asn Tyr Asn Ala Val

245 250 255

Ile Gly Lys Ile Val Lys Glu Asp Lys Gln Glu Ile Gln Gly Leu Asn
260 265 270

Glu His Ile Asn Leu Tyr Asn Gln Lys His Lys Asp Arg Arg Leu Pro
275 280 285

Phe Phe Ile Ser Leu Lys Lys Gln Ile Leu Ser Asp Arg Glu Ala Leu
290 295 300

Ser Trp Leu Pro Asp Met Phe Lys Asn Asp Ser Glu Val Ile Lys Ala
305 310 315 320

Leu Lys Gly Phe Tyr Ile Glu Asp Gly Phe Glu Asn Asn Val Leu Thr
325 330 335

Pro Leu Ala Thr Leu Leu Ser Ser Leu Asp Lys Tyr Asn Leu Asn Gly
340 345 350

Ile Phe Ile Arg Asn Asn Glu Ala Leu Ser Ser Leu Ser Gln Asn Val
355 360 365

Tyr Arg Asn Phe Ser Ile Asp Glu Ala Ile Asp Ala Asn Ala Glu Leu
370 375 380

Gln Thr Phe Asn Asn Tyr Glu Leu Ile Ala Asn Ala Leu Arg Ala Lys
385 390 395 400

Ile Lys Lys Glu Thr Lys Gln Gly Arg Lys Ser Phe Glu Lys Tyr Glu
405 410 415

Glu Tyr Ile Asp Lys Lys Val Lys Ala Ile Asp Ser Leu Ser Ile Gln
420 425 430

Glu Ile Asn Glu Leu Val Glu Asn Tyr Val Ser Glu Phe Asn Ser Asn
435 440 445

Ser Gly Asn Met Pro Arg Lys Val Glu Asp Tyr Phe Ser Leu Met Arg
450 455 460

Lys Gly Asp Phe Gly Ser Asn Asp Leu Ile Glu Asn Ile Lys Thr Lys
465 470 475 480

Leu Ser Ala Ala Glu Lys Leu Leu Gly Thr Lys Tyr Gln Glu Thr Ala
485 490 495

Lys Asp Ile Phe Lys Lys Asp Glu Asn Ser Lys Leu Ile Lys Glu Leu
500 505 510

Leu Asp Ala Thr Lys Gln Phe Gln His Phe Ile Lys Pro Leu Leu Gly
515 520 525

Thr Gly Glu Glu Ala Asp Arg Asp Leu Val Phe Tyr Gly Asp Phe Leu
530 535 540

Pro Leu Tyr Glu Lys Phe Glu Glu Leu Thr Leu Leu Tyr Asn Lys Val
545 550 555 560

Arg Asn Arg Leu Thr Gln Lys Pro Tyr Ser Lys Asp Lys Ile Arg Leu
565 570 575

Cys Phe Asn Lys Pro Lys Leu Met Thr Gly Trp Val Asp Ser Lys Thr

580 585 590

Glu Lys Ser Asp Asn Gly Thr Gln Tyr Gly Gly Tyr Leu Phe Arg Lys
595 600 605

Lys Asn Glu Ile Gly Glu Tyr Asp Tyr Phe Leu Gly Ile Ser Ser Lys
610 615 620

Ala Gln Leu Phe Arg Lys Asn Glu Ala Val Ile Gly Asp Tyr Glu Arg
625 630 635 640

Leu Asp Tyr Tyr Gln Pro Lys Ala Asn Thr Ile Tyr Gly Ser Ala Tyr
645 650 655

Glu Gly Glu Asn Ser Tyr Lys Glu Asp Lys Lys Arg Leu Asn Lys Val
660 665 670

Ile Ile Ala Tyr Ile Glu Gln Ile Lys Gln Thr Asn Ile Lys Lys Ser
675 680 685

Ile Ile Glu Ser Ile Ser Lys Tyr Pro Asn Ile Ser Asp Asp Asp Lys
690 695 700

Val Thr Pro Ser Ser Leu Leu Glu Lys Ile Lys Lys Val Ser Ile Asp
705 710 715 720

Ser Tyr Asn Gly Ile Leu Ser Phe Lys Ser Phe Gln Ser Val Asn Lys
725 730 735

Glu Val Ile Asp Asn Leu Leu Lys Thr Ile Ser Pro Leu Lys Asn Lys
740 745 750

Ala Glu Phe Leu Asp Leu Ile Asn Lys Asp Tyr Gln Ile Phe Thr Glu
755 760 765

Val Gln Ala Val Ile Asp Glu Ile Cys Lys Gln Lys Thr Phe Ile Tyr
770 775 780

Phe Pro Ile Ser Asn Val Glu Leu Glu Lys Glu Met Gly Asp Lys Asp
785 790 795 800

Lys Pro Leu Cys Leu Phe Gln Ile Ser Asn Lys Asp Leu Ser Phe Ala
805 810 815

Lys Thr Phe Ser Ala Asn Leu Arg Lys Lys Arg Gly Ala Glu Asn Leu
820 825 830

His Thr Met Leu Phe Lys Ala Leu Met Glu Gly Asn Gln Asp Asn Leu
835 840 845

Asp Leu Gly Ser Gly Ala Ile Phe Tyr Arg Ala Lys Ser Leu Asp Gly
850 855 860

Asn Lys Pro Thr His Pro Ala Asn Glu Ala Ile Lys Cys Arg Asn Val
865 870 875 880

Ala Asn Lys Asp Lys Val Ser Leu Phe Thr Tyr Asp Ile Tyr Lys Asn
885 890 895

Arg Arg Tyr Met Glu Asn Lys Phe Leu Phe His Leu Ser Ile Val Gln
900 905 910

Asn Tyr Lys Ala Ala Asn Asp Ser Ala Gln Leu Asn Ser Ser Ala Thr

915 920 925

Glu Tyr Ile Arg Lys Ala Asp Asp Leu His Ile Ile Gly Ile Asp Arg
930 935 940

Gly Glu Arg Asn Leu Leu Tyr Tyr Ser Val Ile Asp Met Lys Gly Asn
945 950 955 960

Ile Val Glu Gln Asp Ser Leu Asn Ile Ile Arg Asn Asn Asp Leu Glu
965 970 975

Thr Asp Tyr His Asp Leu Leu Asp Lys Arg Glu Lys Glu Arg Lys Ala
980 985 990

Asn Arg Gln Asn Trp Glu Ala Val Glu Gly Ile Lys Asp Leu Lys Lys
995 1000 1005

Gly Tyr Leu Ser Gln Ala Val His Gln Ile Ala Gln Leu Met Leu
1010 1015 1020

Lys Tyr Asn Ala Ile Ile Ala Leu Glu Asp Leu Gly Gln Met Phe
1025 1030 1035

Val Thr Arg Gly Gln Lys Ile Glu Lys Ala Val Tyr Gln Gln Phe
1040 1045 1050

Glu Lys Ser Leu Val Asp Lys Leu Ser Tyr Leu Val Asp Lys Lys
1055 1060 1065

Arg Pro Tyr Asn Glu Leu Gly Gly Ile Leu Lys Ala Tyr Gln Leu
1070 1075 1080

Ala Ser Ser Ile Thr Lys Asn Asn Ser Asp Lys Gln Asn Gly Phe
1085 1090 1095

Leu Phe Tyr Val Pro Ala Trp Asn Thr Ser Lys Ile Asp Pro Val
1100 1105 1110

Thr Gly Phe Thr Asp Leu Leu Arg Pro Lys Ala Met Thr Ile Lys
1115 1120 1125

Glu Ala Gln Asp Phe Phe Gly Ala Phe Asp Asn Ile Ser Tyr Asn
1130 1135 1140

Asp Lys Gly Tyr Phe Glu Phe Glu Thr Asn Tyr Asp Lys Phe Lys
1145 1150 1155

Ile Arg Met Lys Ser Ala Gln Thr Arg Trp Thr Ile Cys Thr Phe
1160 1165 1170

Gly Asn Arg Ile Lys Arg Lys Lys Asp Lys Asn Tyr Trp Asn Tyr
1175 1180 1185

Glu Glu Val Glu Leu Thr Glu Glu Phe Lys Lys Leu Phe Lys Asp
1190 1195 1200

Ser Asn Ile Asp Tyr Glu Asn Cys Asn Leu Lys Glu Glu Ile Gln
1205 1210 1215

Asn Lys Asp Asn Arg Lys Phe Phe Asp Asp Leu Ile Lys Leu Leu
1220 1225 1230

Gln Leu Thr Leu Gln Met Arg Asn Ser Asp Asp Lys Gly Asn Asp

1235 1240 1245

Tyr Ile Ile Ser Pro Val Ala Asn Ala Glu Gly Gln Phe Phe Asp
1250 1255 1260

Ser Arg Asn Gly Asp Lys Lys Leu Pro Leu Asp Ala Asp Ala Asn
1265 1270 1275

Gly Ala Tyr Asn Ile Ala Arg Lys Gly Leu Trp Asn Ile Arg Gln
1280 1285 1290

Ile Lys Gln Thr Lys Asn Asp Lys Lys Leu Asn Leu Ser Ile Ser
1295 1300 1305

Ser Thr Glu Trp Leu Asp Phe Val Arg Glu Lys Pro Tyr Leu Lys
1310 1315 1320

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys
1325 1330 1335

Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr
1340 1345 1350

Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1355 1360 1365

<210> 245
<211> 1291
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 245
Met Lys Thr Gln His Phe Phe Glu Asp Phe Thr Ser Leu Tyr Ser Leu
1 5 10 15

Ser Lys Thr Ile Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Leu Glu
20 25 30

Asn Ile Lys Lys Asn Gly Leu Ile Arg Arg Asp Glu Gln Arg Leu Asp
35 40 45

Asp Tyr Glu Lys Leu Lys Lys Val Ile Asp Glu Tyr His Glu Asp Phe
50 55 60

Ile Ala Asn Ile Leu Ser Ser Phe Ser Phe Ser Glu Glu Ile Leu Gln
65 70 75 80

Ser Tyr Ile Gln Asn Leu Ser Glu Ser Glu Ala Arg Ala Lys Ile Glu
85 90 95

Lys Thr Met Arg Asp Thr Leu Ala Lys Ala Phe Ser Glu Asp Glu Arg
100 105 110

Tyr Lys Ser Ile Phe Lys Lys Glu Leu Val Lys Lys Asp Ile Pro Val
115 120 125

Trp Cys Pro Ala Tyr Lys Ser Leu Cys Lys Lys Phe Asp Asn Phe Thr
130 135 140

Thr Ser Leu Val Pro Phe His Glu Asn Arg Lys Asn Leu Tyr Thr Ser
145 150 155 160

```
Asn Glu Ile Thr Ala Ser Ile Pro Tyr Arg Ile Val His Val Asn Leu
                165                 170                 175

Pro Lys Phe Ile Gln Asn Ile Glu Ala Leu Cys Glu Leu Gln Lys Lys
            180                 185                 190

Met Gly Ala Asp Leu Tyr Leu Glu Met Met Glu Asn Leu Arg Asn Val
        195                 200                 205

Trp Pro Ser Phe Val Lys Thr Pro Asp Asp Leu Cys Asn Leu Lys Thr
    210                 215                 220

Tyr Asn His Leu Met Val Gln Ser Ser Ile Ser Glu Tyr Asn Arg Phe
225                 230                 235                 240

Val Gly Gly Tyr Ser Thr Glu Asp Gly Thr Lys His Gln Gly Ile Asn
                245                 250                 255

Glu Trp Ile Asn Ile Tyr Arg Gln Arg Asn Lys Glu Met Arg Leu Pro
            260                 265                 270

Gly Leu Val Phe Leu His Lys Gln Ile Leu Ala Lys Val Asp Ser Ser
        275                 280                 285

Ser Phe Ile Ser Asp Thr Leu Glu Asn Asp Asp Gln Val Phe Cys Val
    290                 295                 300

Leu Arg Gln Phe Arg Lys Leu Phe Trp Asn Thr Val Ser Ser Lys Glu
305                 310                 315                 320

Asp Asp Ala Ala Ser Leu Lys Asp Leu Phe Cys Gly Leu Ser Gly Tyr
                325                 330                 335

Asp Pro Glu Ala Ile Tyr Val Ser Asp Ala His Leu Ala Thr Ile Ser
            340                 345                 350

Lys Asn Ile Phe Asp Arg Trp Asn Tyr Ile Ser Asp Ala Ile Arg Arg
        355                 360                 365

Lys Thr Glu Val Leu Met Pro Arg Lys Lys Glu Ser Val Glu Arg Tyr
    370                 375                 380

Ala Glu Lys Ile Ser Lys Gln Ile Lys Lys Arg Gln Ser Tyr Ser Leu
385                 390                 395                 400

Ala Glu Leu Asp Asp Leu Leu Ala His Tyr Ser Glu Glu Ser Leu Pro
                405                 410                 415

Ala Gly Phe Ser Leu Leu Ser Tyr Phe Thr Ser Leu Gly Gly Gln Lys
            420                 425                 430

Tyr Leu Val Ser Asp Gly Glu Val Ile Leu Tyr Glu Glu Gly Ser Asn
        435                 440                 445

Ile Trp Asp Glu Val Leu Ile Ala Phe Arg Asp Leu Gln Val Ile Leu
    450                 455                 460

Asp Lys Asp Phe Thr Glu Lys Lys Leu Gly Lys Asp Glu Glu Ala Val
465                 470                 475                 480

Ser Val Ile Lys Lys Ala Leu Asp Ser Ala Leu Arg Leu Arg Lys Phe
                485                 490                 495
```

```
Phe Asp Leu Leu Ser Gly Thr Gly Ala Glu Ile Arg Arg Asp Ser Ser
            500                 505                 510

Phe Tyr Ala Leu Tyr Thr Asp Arg Met Asp Lys Leu Lys Gly Leu Leu
        515                 520                 525

Lys Met Tyr Asp Lys Val Arg Asn Tyr Leu Thr Lys Lys Pro Tyr Ser
    530                 535                 540

Ile Glu Lys Phe Lys Leu His Phe Asp Asn Pro Ser Leu Leu Ser Gly
545                 550                 555                 560

Trp Asp Lys Asn Lys Glu Leu Asn Asn Leu Ser Val Ile Phe Arg Gln
                565                 570                 575

Asn Gly Tyr Tyr Tyr Leu Gly Ile Met Thr Pro Lys Gly Lys Asn Leu
            580                 585                 590

Phe Lys Thr Leu Pro Lys Leu Gly Ala Glu Glu Met Phe Tyr Glu Lys
        595                 600                 605

Met Glu Tyr Lys Gln Ile Ala Glu Pro Met Leu Met Leu Pro Lys Val
    610                 615                 620

Phe Phe Pro Lys Lys Thr Lys Pro Ala Phe Ala Pro Asp Gln Ser Val
625                 630                 635                 640

Val Asp Ile Tyr Asn Lys Lys Thr Phe Lys Thr Gly Gln Lys Gly Phe
                645                 650                 655

Asn Lys Lys Asp Leu Tyr Arg Leu Ile Asp Phe Tyr Lys Glu Ala Leu
            660                 665                 670

Thr Val His Glu Trp Lys Leu Phe Asn Phe Ser Phe Ser Pro Thr Glu
        675                 680                 685

Gln Tyr Arg Asn Ile Gly Glu Phe Phe Asp Glu Val Arg Glu Gln Ala
    690                 695                 700

Tyr Lys Val Ser Met Val Asn Val Pro Ala Ser Tyr Ile Asp Glu Ala
705                 710                 715                 720

Val Glu Asn Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe
                725                 730                 735

Ser Pro Tyr Ser Lys Gly Ile Pro Asn Leu His Thr Leu Tyr Trp Lys
            740                 745                 750

Ala Leu Phe Ser Glu Gln Asn Gln Ser Arg Val Tyr Lys Leu Cys Gly
        755                 760                 765

Gly Gly Glu Leu Phe Tyr Arg Lys Ala Ser Leu His Met Gln Asp Thr
    770                 775                 780

Thr Val His Pro Lys Gly Ile Ser Ile His Lys Lys Asn Leu Asn Lys
785                 790                 795                 800

Lys Gly Glu Thr Ser Leu Phe Asn Tyr Asp Leu Val Lys Asp Lys Arg
                805                 810                 815

Phe Thr Glu Asp Lys Phe Phe Phe His Val Pro Ile Ser Ile Asn Tyr
            820                 825                 830
```

```
Lys Asn Lys Lys Ile Thr Asn Val Asn Gln Met Val Arg Asp Tyr Ile
        835                 840                 845

Ala Gln Asn Asp Asp Leu Gln Ile Ile Gly Ile Asp Arg Gly Glu Arg
    850                 855                 860

Asn Leu Leu Tyr Ile Ser Arg Ile Asp Thr Arg Gly Asn Leu Leu Glu
865                 870                 875                 880

Gln Phe Ser Leu Asn Val Ile Glu Ser Asp Lys Gly Asp Leu Arg Thr
                885                 890                 895

Asp Tyr Gln Lys Ile Leu Gly Asp Arg Glu Gln Glu Arg Leu Arg Arg
            900                 905                 910

Arg Gln Glu Trp Lys Ser Ile Glu Ser Ile Lys Asp Leu Lys Asp Gly
        915                 920                 925

Tyr Met Ser Gln Val Val His Lys Ile Cys Asn Met Val Val Glu His
    930                 935                 940

Lys Ala Ile Val Val Leu Glu Asn Leu Asn Leu Ser Phe Met Lys Gly
945                 950                 955                 960

Arg Lys Lys Val Glu Lys Ser Val Tyr Glu Lys Phe Glu Arg Met Leu
                965                 970                 975

Val Asp Lys Leu Asn Tyr Leu Val Val Asp Lys Lys Asn Leu Ser Asn
            980                 985                 990

Glu Pro Gly Gly Leu Tyr Ala Ala Tyr Gln Leu Thr Asn Pro Leu Phe
        995                 1000                1005

Ser Phe  Glu Glu Leu His Arg  Tyr Pro Gln Ser Gly  Ile Leu Phe
    1010                1015                1020

Phe Val  Asp Pro Trp Asn Thr  Ser Leu Thr Asp Pro  Ser Thr Gly
    1025                1030                1035

Phe Val  Asn Leu Leu Gly Arg  Ile Asn Tyr Thr Asn  Val Gly Asp
    1040                1045                1050

Ala Arg  Lys Phe Phe Asp Arg  Phe Asn Ala Ile Arg  Tyr Asp Gly
    1055                1060                1065

Lys Gly  Asn Ile Leu Phe Asp  Leu Asp Leu Ser Arg  Phe Asp Val
    1070                1075                1080

Arg Val  Glu Thr Gln Arg Lys  Leu Trp Thr Leu Thr  Thr Phe Gly
    1085                1090                1095

Ser Arg  Ile Ala Lys Ser Lys  Lys Ser Gly Lys Trp  Met Val Glu
    1100                1105                1110

Arg Ile  Glu Asn Leu Ser Leu  Cys Phe Leu Glu Leu  Phe Glu Gln
    1115                1120                1125

Phe Asn  Ile Gly Tyr Arg Val  Glu Lys Asp Leu Lys  Lys Ala Ile
    1130                1135                1140

Leu Ser  Gln Asp Arg Lys Glu  Phe Tyr Val Arg Leu  Ile Tyr Leu
    1145                1150                1155
```

Phe Asn Leu Met Met Gln Ile Arg Asn Ser Asp Gly Glu Glu Asp
1160 1165 1170

Tyr Ile Leu Ser Pro Ala Leu Asn Glu Lys Asn Leu Gln Phe Asp
1175 1180 1185

Ser Arg Leu Ile Glu Ala Lys Asp Leu Pro Val Asp Ala Asp Ala
1190 1195 1200

Asn Gly Ala Tyr Asn Val Ala Arg Lys Gly Leu Met Val Val Gln
1205 1210 1215

Arg Ile Lys Arg Gly Asp His Glu Ser Ile His Arg Ile Gly Arg
1220 1225 1230

Ala Gln Trp Leu Arg Tyr Val Gln Glu Gly Ile Val Glu Lys Arg
1235 1240 1245

Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys Gly
1250 1255 1260

Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val
1265 1270 1275

Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1280 1285 1290

<210> 246
<211> 18
<212> PRT
<213> 酸土脂環酸芽孢桿菌（Alicyclobacillus acidoterrestris）

<400> 246
Leu Arg Val Met Ser Val Asp Leu Gly Leu Arg Thr Ser Ala Ser Ile
1 5 10 15

Ser Val

<210> 247
<211> 22
<212> PRT
<213> 酸土脂環酸芽孢桿菌

<400> 247
Gln Arg Thr Leu Arg Gln Leu Arg Thr Gln Leu Ala Tyr Leu Arg Leu
1 5 10 15

Leu Val Arg Cys Gly Ser
20

<210> 248
<211> 12
<212> PRT
<213> 酸土脂環酸芽孢桿菌

<400> 248
Cys Gln Leu Ile Leu Leu Glu Glu Leu Ser Glu Tyr
1 5 10

<210> 249
<211> 16
<212> PRT
<213> 酸土脂環酸芽孢桿菌

<400> 249
His Gln Ile His Ala Asp Leu Asn Ala Ala Gln Asn Leu Gln Gln Arg
1 5 10 15

<210> 250
<211> 18
<212> PRT
<213> 洋蔥脂環酸芽孢桿菌（Alicyclobacillus contaminans）

<400> 250
Val Arg Val Met Ser Val Asp Leu Gly Val Arg Tyr Gly Ala Ala Ile
1 5 10 15

Ser Val

<210> 251
<211> 22
<212> PRT
<213> 洋蔥脂環酸芽孢桿菌

<400> 251
Lys Gln Ala Leu Ala Ala Ile Arg Ala Glu Met Ser Ile Leu Arg Lys
1 5 10 15

Trp Leu Arg Val Ser Gln
20

<210> 252
<211> 12
<212> PRT
<213> 洋蔥脂環酸芽孢桿菌

<400> 252
Cys Asp Leu Ile Leu Phe Glu Asp Leu Ser Arg Tyr
1 5 10

<210> 253
<211> 16
<212> PRT
<213> 洋蔥脂環酸芽孢桿菌

<400> 253
Lys Cys Val His Ala Asp Ile Asn Ala Ala His Asn Leu Gln Arg Arg
1 5 10 15

<210> 254
<211> 18
<212> PRT
<213> 意外脱硫弧菌（Desulfovibrio inopinatus）

<400> 254
Leu Arg Val Leu Ser Val Asp Leu Gly Met Arg Thr Phe Ala Ser Cys
1 5 10 15

Ser Val

<210> 255
<211> 22
<212> PRT
<213> 意外脱硫弧菌

<400> 255
Arg Ala Glu Ile Tyr Ala Leu Lys Arg Asp Ile Gln Arg Leu Lys Ser
1 5 10 15

Leu Leu Arg Leu Gly Glu
20

<210> 256
<211> 12
<212> PRT
<213> 意外脱硫弧菌

<400> 256
Cys Gln Leu Ile Leu Phe Glu Asp Leu Ala Arg Tyr

1 5 10

<210> 257
<211> 16
<212> PRT
<213> 意外脱硫弧菌

<400> 257
Cys Val Ile His Ala Asp Met Asn Ala Ala Gln Asn Leu Gln Arg Arg
1 5 10 15

<210> 258
<211> 18
<212> PRT
<213> thiodismutans脫硫鹽鹼桿菌

<400> 258
Leu Arg Val Leu Ser Val Asp Leu Gly Val Arg Ser Phe Ala Ala Cys
1 5 10 15

Ser Val

<210> 259
<211> 22
<212> PRT
<213> thiodismutans脫硫鹽鹼桿菌

<400> 259
Met Glu Glu Leu Arg Ser Leu Asn Gly Asp Ile Arg Arg Leu Lys Ala
1 5 10 15

Ile Leu Arg Leu Ser Val
20

<210> 260
<211> 12
<212> PRT
<213> thiodismutans脫硫鹽鹼桿菌

<400> 260
Cys Arg Leu Ile Leu Phe Glu Asp Leu Ala Arg Tyr
1 5 10

<210> 261
<211> 16
<212> PRT
<213> thiodismutans脫硫鹽鹼桿菌

<400> 261
His Val Ile His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln Arg Arg
1 5 10 15

<210> 262
<211> 18
<212> PRT
<213> 熱生腫塊芽孢桿菌（Tuberibacillus calidus）

<400> 262
Leu Arg Val Met Ser Val Asp Leu Gly Gln Arg Gln Ala Ala Ala Ile
1 5 10 15

Ser Ile

<210> 263
<211> 22
<212> PRT
<213> 熱生腫塊芽孢桿菌

<400> 263
Asp Gln Ala Ile Arg Asp Leu Ser Arg Lys Leu Lys Phe Leu Lys Asn
1 5 10 15

Val Leu Asn Met Gln Lys
            20

<210> 264
<211> 12
<212> PRT
<213> 熱生腫塊芽孢桿菌

<400> 264
Cys Gln Leu Val Leu Phe Glu Asp Leu Ser Arg Tyr
1   5   10

<210> 265
<211> 16
<212> PRT
<213> 熱生腫塊芽孢桿菌

<400> 265
Val Ile Thr His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln Lys Arg
1   5   10   15

<210> 266
<211> 18
<212> PRT
<213> 嗜熱澱粉芽孢桿菌

<400> 266
Leu Arg Val Met Ser Ile Asp Leu Gly Gln Arg Gln Ala Ala Ala Ala
1   5   10   15

Ser Ile

<210> 267
<211> 22
<212> PRT
<213> 嗜熱澱粉芽孢桿菌

<400> 267
Glu Asp Asn Leu Lys Leu Met Asn Gln Lys Leu Asn Phe Leu Arg Asn
1   5   10   15

Val Leu His Phe Gln Gln
            20

<210> 268
<211> 12
<212> PRT
<213> 嗜熱澱粉芽孢桿菌

<400> 268
Cys Gln Ile Ile Leu Phe Glu Asp Leu Ser Asn Tyr
1   5   10

<210> 269
<211> 16
<212> PRT
<213> 嗜熱澱粉芽孢桿菌

<400> 269
Val Thr Thr His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln Lys Arg
1   5   10   15

<210> 270
<211> 18
<212> PRT
<213> 芽孢桿菌屬某種

<400> 270
Phe Arg Val Met Ser Ile Asp Leu Gly Leu Arg Ala Ala Ala Ala Thr
1   5   10   15

Ser Ile

<210> 271
<211> 22
<212> PRT
<213> 芽孢桿菌屬某種

<400> 271
Phe Gln Leu His Gln Arg Val Lys Phe Gln Ile Arg Val Leu Ala Gln
1 5 10 15

Ile Met Arg Met Ala Asn
20

<210> 272
<211> 12
<212> PRT
<213> 芽孢桿菌屬某種

<400> 272
Cys Gln Val Ile Leu Phe Glu Asn Leu Ser Gln Tyr
1 5 10

<210> 273
<211> 16
<212> PRT
<213> 芽孢桿菌屬某種

<400> 273
Val Phe Leu Gln Ala Asp Ile Asn Ala Ala His Asn Leu Gln Lys Arg
1 5 10 15

<210> 274
<211> 18
<212> PRT
<213> 結瘤甲基桿菌（Methylobacterium nodulans）

<400> 274
Leu Arg Val Leu Ser Ile Asp Leu Gly Val Arg Ser Phe Ala Thr Cys
1 5 10 15

Ser Val

<210> 275
<211> 22
<212> PRT
<213> 結瘤甲基桿菌

<400> 275
Asp Ala Glu Leu Arg Gln Leu Arg Gly Gly Leu Asn Arg His Arg Gln
1 5 10 15

Leu Leu Arg Ala Ala Thr
20

<210> 276
<211> 12
<212> PRT
<213> 結瘤甲基桿菌

<400> 276
Cys His Val Ile Leu Phe Glu Asp Leu Ser Arg Tyr
1 5 10

<210> 277
<211> 16
<212> PRT
<213> 結瘤甲基桿菌

<400> 277
Ser Arg Ile His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln Arg Arg
1 5 10 15

<210> 278
<211> 18
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選alvus甲烷嗜甲基菌序列

<400> 278
Leu Lys Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu Ile Tyr Val
1 5 10 15

Thr Met

<210> 279
<211> 22
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選alvus甲烷嗜甲基菌序列

<400> 279
Arg Lys Ala Leu Asp Val Arg Glu Tyr Asp Asn Lys Glu Ala Arg Arg
1 5 10 15

Asn Trp Thr Lys Val Glu
20

<210> 280
<211> 13
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選alvus甲烷嗜甲基菌序列

<400> 280
Asn Ala Ile Ile Val Met Glu Asp Leu Asn His Gly Phe
1 5 10

<210> 281
<211> 16
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選alvus甲烷嗜甲基菌序列

<400> 281
Leu Pro Gln Asp Ser Asp Ala Asn Gly Ala Tyr Asn Ile Ala Leu Lys
1 5 10 15

<210> 282
<211> 18
<212> PRT
<213> 窮氏互養菌

<400> 282
Val Asn Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu Val Tyr Val
1 5 10 15

Ser Leu

<210> 283
<211> 22
<212> PRT
<213> 窮氏互養菌

<400> 283
His Ala Lys Leu Asn Gln Lys Glu Lys Glu Arg Asp Thr Ala Arg Lys
1 5 10 15

Ser Trp Lys Thr Ile Gly
20

<210> 284
<211> 13
<212> PRT
<213> 窮氏互養菌

<400> 284
Asn Ala Val Ile Val Met Glu Asp Leu Asn Ile Gly Phe
1 5 10

<210> 285
<211> 16
<212> PRT
<213> 窮氏互養菌

<400> 285
Leu Pro Ile Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Ala Leu Lys
1 5 10 15

<210> 286
<211> 18
<212> PRT
<213> 未知

<220>
<223> 未知的描述:
毛螺旋菌科細菌序列

<400> 286
Pro Tyr Val Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu Leu Tyr Ile
1 5 10 15

Val Val

<210> 287
<211> 22
<212> PRT
<213> 未知

<220>
<223> 未知的描述:
毛螺旋菌科細菌序列

<400> 287
His Ser Leu Leu Asp Lys Lys Glu Lys Glu Arg Phe Glu Ala Arg Gln
1 5 10 15

Asn Trp Thr Ser Ile Glu
20

<210> 288
<211> 13
<212> PRT
<213> 未知

<220>
<223> 未知的描述:
毛螺旋菌科細菌序列

<400> 288
Asp Ala Val Ile Ala Leu Glu Asp Leu Asn Ser Gly Phe
1 5 10

<210> 289
<211> 16
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 289
Leu Pro Lys Asn Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala Arg Lys
1 5 10 15

<210> 290
<211> 18
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 290
Val His Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr
1 5 10 15

Thr Leu

<210> 291
<211> 22
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 291
His Asp Lys Leu Ala Ala Ile Glu Lys Asp Arg Asp Ser Ala Arg Lys
1 5 10 15

Asp Trp Lys Lys Ile Asn
20

<210> 292
<211> 13
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 292
Asn Ala Ile Val Val Phe Glu Asp Leu Asn Phe Gly Phe
1 5 10

<210> 293
<211> 16
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 293
Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly Leu Lys
1 5 10 15

<210> 294
<211> 18
<212> PRT
<213> 山羊莫拉氏菌

<400> 294
Val Asn Val Ile Gly Ile Asp Arg Gly Glu Arg His Leu Leu Tyr Leu
1 5 10 15

Thr Val

<210> 295
<211> 22
<212> PRT
<213> 山羊莫拉氏菌

<400> 295
His Lys Ile Leu Asp Lys Arg Glu Ile Glu Arg Leu Asn Ala Arg Val
1 5 10 15

Gly Trp Gly Glu Ile Glu
20

<210> 296
<211> 13
<212> PRT
<213> 山羊莫拉氏菌

<400> 296
Asn Ala Ile Val Val Leu Glu Asp Leu Asn Phe Gly Phe
1 5 10

<210> 297
<211> 16
<212> PRT
<213> 山羊莫拉氏菌

<400> 297
Gln Pro Gln Asn Ala Asp Ala Asn Gly Ala Tyr His Ile Ala Leu Lys
1 5 10 15

<210> 298
<211> 18
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 298
Met His Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu Ile Tyr Leu
1 5 10 15

Cys Met

<210> 299
<211> 22
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 299
His Gln Leu Leu Lys Thr Arg Glu Asp Glu Asn Lys Ser Ala Arg Gln
1 5 10 15

Ser Trp Gln Thr Ile His
20

<210> 300
<211> 13
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 300
Asn Ala Ile Val Val Leu Glu Asp Leu Asn Phe Gly Phe
1 5 10

<210> 301
<211> 16
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 301
Met Pro Leu Asp Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala Arg Lys
1 5 10 15

<210> 302
<211> 18
<212> PRT
<213> 易北普雷沃菌

<400> 302
Thr His Ile Ile Gly Ile Asp Arg Gly Glu Arg His Leu Leu Tyr Leu
1 5 10 15

Ser Leu

<210> 303
<211> 22
<212> PRT
<213> 易北普雷沃菌

<400> 303
His Asn Leu Leu Glu Lys Arg Glu Lys Glu Arg Thr Glu Ala Arg His
1 5 10 15

Ser Trp Ser Ser Ile Glu
20

<210> 304
<211> 13
<212> PRT
<213> 易北普雷沃菌

<400> 304
Asn Ala Ile Val Val Leu Glu Asp Leu Asn Gly Gly Phe
1 5 10

<210> 305
<211> 16
<212> PRT
<213> 易北普雷沃菌

<400> 305
Phe Pro Glu Asn Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala Arg Lys
1 5 10 15

<210> 306
<211> 18
<212> PRT
<213> 密斯氏菌屬某種

<400> 306
Ile Asn Ile Ile Gly Ile Asp Arg Gly Glu Arg His Leu Leu Tyr Tyr
1 5 10 15

Ala Leu

<210> 307
<211> 22
<212> PRT
<213> 密斯氏菌屬某種

<400> 307
His Asn Leu Leu Asp Lys Lys Glu Gly Asp Arg Ala Thr Ala Arg Gln
1 5 10 15

Glu Trp Gly Val Ile Glu
20

<210> 308
<211> 13
<212> PRT
<213> 密斯氏菌屬某種

<400> 308
Asn Ala Ile Ile Val Met Glu Asp Leu Asn Phe Gly Phe

1 5 10

<210> 309
<211> 16
<212> PRT
<213> 密斯氏菌屬某種

<400> 309
Met Pro Lys Asn Ala Asp Ala Asn Gly Ala Tyr His Ile Ala Leu Lys
1 5 10 15

<210> 310
<211> 18
<212> PRT
<213> 狗口腔卟啉單胞菌

<400> 310
Met His Val Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu Leu Tyr Ile
1 5 10 15

Cys Val

<210> 311
<211> 22
<212> PRT
<213> 狗口腔卟啉單胞菌

<400> 311
His Asp Leu Leu Glu Ser Arg Asp Lys Asp Arg Gln Gln Glu Arg Arg
1 5 10 15

Asn Trp Gln Thr Ile Glu
20

<210> 312
<211> 13
<212> PRT
<213> 狗口腔卟啉單胞菌

<400> 312
Lys Ala Val Val Ala Leu Glu Asp Leu Asn Met Gly Phe
1 5 10

<210> 313
<211> 16
<212> PRT
<213> 狗口腔卟啉單胞菌

<400> 313
Leu Pro Lys Asp Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala Leu Lys
1 5 10 15

<210> 314
<211> 18
<212> PRT
<213> 硫磺礦硫化葉菌（Sulfolobus solfataricus）

<400> 314
Gly Lys Val Val Ala Ile Asp Val Gly Val Glu Lys Leu Leu Ile Thr
1 5 10 15

Ser Asp

<210> 315
<211> 23
<212> PRT
<213> 硫磺礦硫化葉菌

<400> 315
Val Lys His Ile His Arg Glu Leu Ser Arg Lys Lys Phe Leu Ser Asn
1 5 10 15

Asn Trp Phe Lys Ala Lys Val
20

<210> 316
<211> 13
<212> PRT
<213> 硫磺礦硫化葉菌

<400> 316
Tyr Asp Val Val Val Met Glu Gly Ile His Ala Lys Gln
1 5 10

<210> 317
<211> 16
<212> PRT
<213> 硫磺礦硫化葉菌

<400> 317
Trp Ile Ala Asp Arg Asp Tyr Asn Ala Ser Leu Asn Ile Leu Arg Gly
1 5 10 15

<210> 318
<211> 18
<212> PRT
<213> 念珠藻屬某種（Nostoc sp.）

<400> 318
Leu Lys Thr Ile Gly Leu Asp Val Gly Leu Asn His Phe Leu Thr Asp
1 5 10 15

Ser Glu

<210> 319
<211> 23
<212> PRT
<213> 念珠藻屬某種

<400> 319
Leu Lys Arg Leu Gln Arg Arg Leu Ser Lys Thr Lys Lys Gly Ser Asn
1 5 10 15

Asn Arg Val Lys Ala Arg Asn
20

<210> 320
<211> 13
<212> PRT
<213> 念珠藻屬某種

<400> 320
Ser Asp Leu Val Ala Tyr Glu Asp Leu Gln Val Arg Asn
1 5 10

<210> 321
<211> 16
<212> PRT
<213> 念珠藻屬某種

<400> 321
His Ile Gln Asp Arg Asp Trp Asn Ala Ala Arg Asn Ile Leu Glu Leu
1 5 10 15

<210> 322
<211> 18
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選加爾加亞硝化球菌（Candidatus Nitrososphaera gargensis）序列

<400> 322
Ala Lys Pro Val Gly Ile Asp Val Gly Ile Ala Lys Phe Cys His His

1 5 10 15

Ser Asp

<210> 323
<211> 23
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選加爾加亞硝化球菌序列

<400> 323
Leu Arg Arg Ala His Arg Arg Val Ser Arg Arg Gln Ile Gly Ser Asn
1 5 10 15

Asn Arg Lys Lys Ala Lys Arg
20

<210> 324
<211> 13
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選加爾加亞硝化球菌序列

<400> 324
Tyr Asp Leu Ile Phe Leu Glu Arg Leu Arg Val Met Asn
1 5 10

<210> 325
<211> 16
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選加爾加亞硝化球菌序列

<400> 325
Ala Ile Leu Asp Arg Asp Tyr Asn Ser Ala Ile Asn Ile Leu Lys Arg
1 5 10 15

<210> 326
<211> 18
<212> PRT
<213> 幽門螺桿菌（Helicobacter pylori）

<400> 326
Lys Lys Ala Val Gly Leu Asp Met Gly Leu Arg Thr Leu Ile Val Thr
1 5 10 15

Ser Asp

<210> 327
<211> 23
<212> PRT
<213> 幽門螺桿菌

<400> 327
Leu Thr Lys Ala Gln Arg Arg Leu Ser Lys Lys Val Lys Asp Ser Asn
1 5 10 15

Asn Arg Lys Lys Gln Ala Lys
20

<210> 328
<211> 13
<212> PRT

<213> 幽門螺桿菌

<400> 328
Tyr Asp Leu Ile Gly Val Glu Thr Leu Asn Val Lys Ala
1 5 10

<210> 329
<211> 16
<212> PRT
<213> 幽門螺桿菌

<400> 329
Thr Thr His His Arg Asp Tyr Asn Ala Ser Val Asn Ile Arg Asn Tyr
1 5 10 15

<210> 330
<211> 18
<212> PRT
<213> 海濱屈撓桿菌（Flexibacter litoralis）

<400> 330
Asn Gln Ala Val Gly Ile Asp Met Gly Ile Thr Phe Phe Cys Ile Asp
1 5 10 15

Ser Asn

<210> 331
<211> 23
<212> PRT
<213> 海濱屈撓桿菌

<400> 331
Leu Arg Ile Ala Asn Arg Ser Leu Ser Arg Lys Lys Lys Phe Ser Asn
1 5 10 15

Gly Trp Tyr Lys Lys Lys Val
20

<210> 332
<211> 13
<212> PRT
<213> 海濱屈撓桿菌

<400> 332
Asn Ser Leu Val Val Val Glu Asp Leu Lys Val Lys Asn
1 5 10

<210> 333
<211> 16
<212> PRT
<213> 海濱屈撓桿菌

<400> 333
His Glu Thr Asn Ala Asp Glu Asn Ala Ser Lys Asn Ile Leu Ser Glu
1 5 10 15

<210> 334
<211> 18
<212> PRT
<213> 大腸桿菌（Escherichia coli）

<400> 334
Ala Ser Met Val Gly Leu Asp Ala Gly Val Ala Lys Leu Ala Thr Leu
1 5 10 15

Ser Asp

<210> 335
<211> 23
<212> PRT
<213> 大腸桿菌

<400> 335
Leu Ala Arg Leu Gln Arg Gln Leu Ser Arg Lys Val Lys Phe Ser Asn
1 5 10 15

Asn Trp Gln Lys Gln Lys Arg
20

<210> 336
<211> 13
<212> PRT
<213> 大腸桿菌

<400> 336
His Ala Met Ile Val Ile Glu Asp Leu Lys Val Ser Asn
1 5 10

<210> 337
<211> 16
<212> PRT
<213> 大腸桿菌

<400> 337
Tyr Thr Ala Asn Ala Asp Val Asn Gly Ala Arg Asn Ile Leu Ala Ala
1 5 10 15

<210> 338
<211> 18
<212> PRT
<213> 肉毒桿菌（Clostridium botulinum）

<400> 338
Asn Lys Lys Val Gly Ile Asp Val Gly Leu Lys Glu Phe Ala Thr Thr
1 5 10 15

Ser Asp

<210> 339
<211> 23
<212> PRT
<213> 肉毒桿菌

<400> 339
Leu Ala Lys Leu Gln Lys Asp Leu Ser Arg Lys Lys Lys Asn Ser Asn
1 5 10 15

Asn Arg Lys Lys Ala Arg Leu
20

<210> 340
<211> 13
<212> PRT
<213> 肉毒桿菌

<400> 340
Asn Gln Ala Ile Val Ile Glu Asn Leu Lys Val Ser Asn
1 5 10

<210> 341
<211> 16
<212> PRT
<213> 肉毒桿菌

<400> 341
Met Ile Met Asp Arg Asp Leu Asn Ala Ser Lys Asn Leu Leu Asn Leu
1 5 10 15

<210> 342
<211> 18
<212> PRT
<213> 胺基酸球菌屬某種

<400> 342
Met Tyr Tyr Leu Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Tyr Ala

1 5 10 15

Val Thr

<210> 343
<211> 22
<212> PRT
<213> 胺基酸球菌屬某種

<400> 343
Ala Glu Arg Arg Ser Phe Arg Thr Ser Arg Arg Arg Leu Asp Arg Arg
1 5 10 15

Gln Gln Arg Val Lys Leu
20

<210> 344
<211> 13
<212> PRT
<213> 胺基酸球菌屬某種

<400> 344
Pro Lys Arg Ile Phe Ile Glu Met Ala Arg Asp Gly Glu
1 5 10

<210> 345
<211> 16
<212> PRT
<213> 胺基酸球菌屬某種

<400> 345
Leu His His Ala Lys Asp Ala Phe Leu Ala Ile Val Thr Gly Asn Val
1 5 10 15

<210> 346
<211> 18
<212> PRT
<213> 貓糞球菌（Coprococcus catus）

<400> 346
Glu Tyr Phe Leu Gly Leu Asp Met Gly Thr Gly Ser Leu Gly Trp Ala
1 5 10 15

Val Thr

<210> 347
<211> 22
<212> PRT
<213> 貓糞球菌

<400> 347
Glu Glu Arg Arg Met Phe Arg Thr Ala Arg Arg Arg Leu Asp Arg Arg
1 5 10 15

Asn Trp Arg Ile Gln Val
20

<210> 348
<211> 13
<212> PRT
<213> 貓糞球菌

<400> 348
Pro Lys Arg Val Phe Val Glu Met Ala Arg Glu Lys Gln
1 5 10

<210> 349
<211> 16
<212> PRT
<213> 貓糞球菌

<400> 349
Leu His His Ala Lys Asp Ala Tyr Leu Asn Ile Val Val Gly Asn Ala
1 5 10 15

<210> 350
<211> 18
<212> PRT
<213> 齒垢密螺旋體（Treponema denticola）

<400> 350
Asp Tyr Phe Leu Gly Leu Asp Val Gly Thr Gly Ser Val Gly Trp Ala
1 5 10 15

Val Thr

<210> 351
<211> 22
<212> PRT
<213> 齒垢密螺旋體

<400> 351
Glu Val Arg Arg Leu His Arg Gly Ala Arg Arg Arg Ile Glu Arg Arg
1 5 10 15

Lys Lys Arg Ile Lys Leu
20

<210> 352
<211> 13
<212> PRT
<213> 齒垢密螺旋體

<400> 352
Pro Lys Lys Ile Phe Ile Glu Met Ala Lys Gly Ala Glu
1 5 10

<210> 353
<211> 16
<212> PRT
<213> 齒垢密螺旋體

<400> 353
Phe His His Ala His Asp Ala Tyr Leu Asn Ile Val Val Gly Asn Val
1 5 10 15

<210> 354
<211> 18
<212> PRT
<213> 運動支原體（Mycoplasma mobile）

<400> 354
Lys Val Val Leu Gly Leu Asp Leu Gly Ile Ala Ser Val Gly Trp Cys
1 5 10 15

Leu Thr

<210> 355
<211> 22
<212> PRT
<213> 運動支原體

<400> 355
Glu Thr Arg Arg Lys Lys Arg Gly Gln Arg Arg Arg Asn Arg Arg Leu
1 5 10 15

Phe Thr Arg Lys Arg Asp
20

<210> 356
<211> 13
<212> PRT

<213> 運動支原體

<400> 356
Ile Glu Lys Ile Val Val Glu Val Thr Arg Ser Ser Asn
1 5 10

<210> 357
<211> 16
<212> PRT
<213> 運動支原體

<400> 357
Gly His His Ala Glu Asp Ala Tyr Phe Ile Thr Ile Ile Ser Gln Tyr
1 5 10 15

<210> 358
<211> 18
<212> PRT
<213> 嗜熱鏈球菌（Streptococcus thermophilus）

<400> 358
Asp Leu Val Leu Gly Leu Asp Ile Gly Ile Gly Ser Val Gly Val Gly
1 5 10 15

Ile Leu

<210> 359
<211> 22
<212> PRT
<213> 嗜熱鏈球菌

<400> 359
Leu Val Arg Arg Thr Asn Arg Gln Gly Arg Arg Leu Thr Arg Arg Lys
1 5 10 15

Lys His Arg Ile Val Arg
20

<210> 360
<211> 13
<212> PRT
<213> 嗜熱鏈球菌

<400> 360
Phe Asp Asn Ile Val Ile Glu Met Ala Arg Glu Thr Asn
1 5 10

<210> 361
<211> 16
<212> PRT
<213> 嗜熱鏈球菌

<400> 361
His His His Ala Val Asp Ala Leu Ile Ile Ala Ala Ser Ser Gln Leu
1 5 10 15

<210> 362
<211> 18
<212> PRT
<213> 空腸彎麴菌（Campylobacter jejuni）

<400> 362
Ala Arg Ile Leu Ala Phe Asp Ile Gly Ile Ser Ser Ile Gly Trp Ala
1 5 10 15

Phe Ser

<210> 363
<211> 22
<212> PRT
<213> 空腸彎麴菌

<400> 363
Leu Pro Arg Arg Leu Ala Arg Ser Ala Arg Lys Arg Leu Ala Arg Arg
1 5 10 15

Lys Ala Arg Leu Asn His
20

<210> 364
<211> 13
<212> PRT
<213> 空腸彎麴菌

<400> 364
Val His Lys Ile Asn Ile Glu Leu Ala Arg Glu Val Gly
1 5 10

<210> 365
<211> 16
<212> PRT
<213> 空腸彎麴菌

<400> 365
Leu His His Ala Ile Asp Ala Val Ile Ile Ala Tyr Ala Asn Asn Ser
1 5 10 15

<210> 366
<211> 18
<212> PRT
<213> 產氣莢膜梭菌（Clostridium perfringens）

<400> 366
Asn Tyr Ala Leu Gly Leu Asp Ile Gly Ile Thr Ser Val Gly Trp Ala
1 5 10 15

Val Ile

<210> 367
<211> 22
<212> PRT
<213> 產氣莢膜梭菌

<400> 367
Leu Pro Arg Arg Leu Ala Arg Gly Arg Arg Arg Leu Leu Arg Arg Lys
1 5 10 15

Ala Tyr Arg Val Glu Arg
20

<210> 368
<211> 13
<212> PRT
<213> 產氣莢膜梭菌

<400> 368
Pro Val Arg Ile Asn Ile Glu Leu Ala Arg Asp Leu Ala
1 5 10

<210> 369
<211> 16
<212> PRT
<213> 產氣莢膜梭菌

<400> 369
Lys His His Ala Leu Asp Ala Ala Val Val Gly Val Thr Thr Gln Gly
1 5 10 15

<210> 370
<211> 18
<212> PRT
<213> 嗜粘蛋白阿克曼氏菌（Akkermansia muciniphila）

<400> 370
Ser Leu Thr Phe Ser Phe Asp Ile Gly Tyr Ala Ser Ile Gly Trp Ala

1 5 10 15

Val Ile

<210> 371
<211> 22
<212> PRT
<213> 嗜粘蛋白阿克曼氏菌

<400> 371
Phe Lys Arg Arg Glu Tyr Arg Arg Leu Arg Arg Asn Ile Arg Ser Arg
1 5 10 15

Arg Val Arg Ile Glu Arg
20

<210> 372
<211> 13
<212> PRT
<213> 嗜粘蛋白阿克曼氏菌

<400> 372
Ile Ser Arg Val Cys Val Glu Val Gly Lys Glu Leu Thr
1 5 10

<210> 373
<211> 16
<212> PRT
<213> 嗜粘蛋白阿克曼氏菌

<400> 373
Leu His His Ala Leu Asp Ala Cys Val Leu Gly Leu Ile Pro Tyr Ile
1 5 10 15

<210> 374
<211> 18
<212> PRT
<213> 長雙歧桿菌（Bifidobacterium longum）

<400> 374
Arg Tyr Arg Ile Gly Ile Asp Val Gly Leu Asn Ser Val Gly Leu Ala
1 5 10 15

Ala Val

<210> 375
<211> 22
<212> PRT
<213> 長雙歧桿菌

<400> 375
Asn Met Ser Gly Val Ala Arg Arg Thr Arg Arg Met Arg Arg Arg Lys
1 5 10 15

Arg Glu Arg Leu His Lys
20

<210> 376
<211> 13
<212> PRT
<213> 長雙歧桿菌

<400> 376
Pro Val Ser Val Asn Ile Glu His Val Arg Ser Ser Phe
1 5 10

<210> 377
<211> 16
<212> PRT
<213> 長雙歧桿菌

<400> 377
Arg His His Ala Val Asp Ala Ser Val Ile Ala Met Met Asn Thr Ala
1 5 10 15

<210> 378
<211> 18
<212> PRT
<213> 產琥珀酸沃廉菌（Wolinella succinogenes）

<400> 378
Val Ser Pro Ile Ser Val Asp Leu Gly Gly Lys Asn Thr Gly Phe Phe
1 5 10 15

Ser Phe

<210> 379
<211> 22
<212> PRT
<213> 產琥珀酸沃廉菌

<400> 379
Val Gly Arg Arg Ser Lys Arg His Ser Lys Arg Asn Asn Leu Arg Asn
1 5 10 15

Lys Leu Val Lys Arg Leu
20

<210> 380
<211> 13
<212> PRT
<213> 產琥珀酸沃廉菌

<400> 380
Lys Val Pro Ile Ile Leu Glu Gln Asn Ala Phe Glu Tyr
1 5 10

<210> 381
<211> 16
<212> PRT
<213> 產琥珀酸沃廉菌

<400> 381
Ser Ser His Ala Ile Asp Ala Val Met Ala Phe Val Ala Arg Tyr Gln
1 5 10 15

<210> 382
<211> 18
<212> PRT
<213> 退伍軍人嗜肺病菌（Legionella pneumophila）

<400> 382
Leu Ser Pro Ile Gly Ile Asp Leu Gly Gly Lys Phe Thr Gly Val Cys
1 5 10 15

Leu Ser

<210> 383
<211> 22
<212> PRT
<213> 退伍軍人嗜肺病菌

<400> 383
Ala Gln Arg Arg Ala Thr Arg His Arg Val Arg Asn Lys Lys Arg Asn
1 5 10 15

Gln Phe Val Lys Arg Val
20

<210> 384
<211> 13
<212> PRT

<213> 退伍軍人嗜肺病菌

<400> 384
Leu Ile Pro Ile Tyr Leu Glu Gln Asn Arg Phe Glu Phe
1 5 10

<210> 385
<211> 15
<212> PRT
<213> 退伍軍人嗜肺病菌

<400> 385
Pro Ser His Ala Ile Asp Ala Thr Leu Thr Met Ser Ile Gly Leu
1 5 10 15

<210> 386
<211> 18
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 386
Ile Leu Pro Ile Ala Ile Asp Leu Gly Val Lys Asn Thr Gly Val Phe
1 5 10 15

Ser Ala

<210> 387
<211> 22
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 387
Asn Asn Arg Thr Ala Arg Arg His Gln Arg Arg Gly Ile Asp Arg Lys
1 5 10 15

Gln Leu Val Lys Arg Leu
20

<210> 388
<211> 13
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 388
His Ile Pro Ile Ile Thr Glu Ser Asn Ala Phe Glu Phe
1 5 10

<210> 389
<211> 16
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 389
Tyr Ser His Leu Ile Asp Ala Met Leu Ala Phe Cys Ile Ala Ala Asp
1 5 10 15

<210> 390
<211> 18
<212> PRT
<213> 釀膿鏈球菌（Streptococcus pyogenes）

<400> 390
Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala
1 5 10 15

Val Ile

<210> 391
<211> 22
<212> PRT
<213> 釀膿鏈球菌

<400> 391
Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg
1 5 10 15

Lys Asn Arg Ile Cys Tyr
20

<210> 392
<211> 13
<212> PRT
<213> 釀膿鏈球菌

<400> 392
Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
1 5 10

<210> 393
<211> 16
<212> PRT
<213> 釀膿鏈球菌

<400> 393
Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala
1 5 10 15

<210> 394
<211> 18
<212> PRT
<213> 德氏乳酸桿菌（Lactobacillus delbrueckii）

<400> 394
Lys Val Ser Leu Gly Val Asp Thr Gly Gln Arg His Ile Gly Phe Ala
1 5 10 15

Ile Val

<210> 395
<211> 23
<212> PRT
<213> 德氏乳酸桿菌

<400> 395
Tyr Thr Arg Lys Ile Tyr Arg Arg Ser Lys Arg Asn Arg Lys Thr Arg
1 5 10 15

Tyr Arg Gln Ala Arg Phe Leu
20

<210> 396
<211> 13
<212> PRT
<213> 德氏乳酸桿菌

<400> 396
Asn Pro Asp Leu His Ile Glu Val Gly Lys Phe Asp Met
1 5 10

<210> 397
<211> 15
<212> PRT
<213> 德氏乳酸桿菌

<400> 397
Lys Gly His Phe Met Asp Ala Ile Ala Ile Ser Gly Ile Lys Pro
1 5 10 15

<210> 398
<211> 18
<212> PRT
<213> evestigatum甲烷鹽菌

<400> 398
Pro Val Val Ala Gly Met Asp Ser Gly Ser Lys His Ile Gly Cys Ala

1 5 10 15

Ala Val

<210> 399
<211> 23
<212> PRT
<213> evestigatum甲烷鹽菌

<400> 399
Lys Asp Arg Ala Asp Tyr Arg Arg Asn Arg Arg Ser Arg Lys Thr Arg
1 5 10 15

Tyr Arg Lys Pro Arg Phe Asp
20

<210> 400
<211> 13
<212> PRT
<213> evestigatum甲烷鹽菌

<400> 400
Val Lys Lys Trp Ile Val Glu Thr Ala Ser Phe Asp Ile
1 5 10

<210> 401
<211> 15
<212> PRT
<213> evestigatum甲烷鹽菌

<400> 401
Lys Thr His Tyr Asn Asp Ala Val Ala Ile Cys Cys Asp Glu Asn
1 5 10 15

<210> 402
<211> 18
<212> PRT
<213> 肉毒桿菌

<400> 402
Pro Ile Thr Leu Gly Ile Asp Ser Gly Tyr Leu Asn Ile Gly Phe Ser
1 5 10 15

Ala Ile

<210> 403
<211> 22
<212> PRT
<213> 肉毒桿菌

<400> 403
Lys Glu Lys Ala Met Tyr Arg Arg Gln Arg Arg Ser Arg Leu Arg Tyr
1 5 10 15

Arg Lys Pro Arg Phe Asn
20

<210> 404
<211> 13
<212> PRT
<213> 肉毒桿菌

<400> 404
Ile Thr Asn Ile Ile Ile Glu Val Ala Asn Phe Asp Thr
1 5 10

<210> 405
<211> 15
<212> PRT
<213> 肉毒桿菌

<400> 405
Lys Thr His Tyr Asn Asp Ala Phe Cys Ile Ala Gly Ser Ser Asn
1 5 10 15

<210> 406
<211> 18
<212> PRT
<213> 喜熱油地芽孢桿菌（Geobacillus thermoleovorans）

<400> 406
Pro Val Ser Leu Gly Val Asp Met Gly Thr Arg His Val Gly Ile Ser
1 5 10 15

Ala Thr

<210> 407
<211> 23
<212> PRT
<213> 喜熱油地芽孢桿菌

<400> 407
Ala Ile Arg Arg Gln Phe Arg Arg Ser Arg Arg Asn Arg Lys Thr Arg
1 5 10 15

Tyr Arg Glu Ala Arg Phe Leu
20

<210> 408
<211> 13
<212> PRT
<213> 喜熱油地芽孢桿菌

<400> 408
Val Thr Ser Val Thr Ile Glu Val Ala Ala Phe Asp Thr
1 5 10

<210> 409
<211> 15
<212> PRT
<213> 喜熱油地芽孢桿菌

<400> 409
Lys Ser His Met Val Asp Ala Arg Cys Ile Ser Gly Asn Pro Leu
1 5 10 15

<210> 410
<211> 18
<212> PRT
<213> 登斯制氨菌（Ammonifex degensii）

<400> 410
Ser Leu Arg Ala Lys Val Asp Asp Gly Ser Arg Tyr Val Gly Ile Ala
1 5 10 15

Leu Val

<210> 411
<211> 23
<212> PRT
<213> 登斯制氨菌

<400> 411
Thr Leu Arg Arg Glu Tyr Arg Arg Gly Arg Arg Tyr Arg Ile Val Arg
1 5 10 15

His Arg Pro Cys Arg Asn Arg
20

<210> 412
<211> 13
<212> PRT

<213> 登斯制氨菌

<400> 412
Ile Ser Gly Val Asp Val Glu Leu Val Ser Ser Gly Val
1 5 10

<210> 413
<211> 15
<212> PRT
<213> 登斯制氨菌

<400> 413
Lys Ser His Thr Asn Asp Ala Leu Ser Leu Phe Leu Pro Gly Gly
1 5 10 15

<210> 414
<211> 18
<212> PRT
<213> 叢毛單胞菌某種（Polaromonas sp.）

<400> 414
Pro Leu Arg Ile Lys Leu Asp Pro Gly Ser Lys Thr Thr Gly Val Ala
1 5 10 15

Leu Val

<210> 415
<211> 22
<212> PRT
<213> 叢毛單胞菌某種

<400> 415
Thr Ala Arg Arg Gln Met Arg Arg Arg Arg Arg Ser Asn Leu Arg Cys
1 5 10 15

Arg Ala Pro Arg Phe Leu
20

<210> 416
<211> 13
<212> PRT
<213> 叢毛單胞菌某種

<400> 416
Val Arg Ala Ile Ser Ser Glu Leu Val Arg Phe Asp Met
1 5 10

<210> 417
<211> 15
<212> PRT
<213> 叢毛單胞菌某種

<400> 417
Lys Thr His Ala Leu Asp Ala Ala Cys Val Gly Gln Val Arg Phe
1 5 10 15

<210> 418
<211> 18
<212> PRT
<213> 多變魚腥藻菌（Anabaena variabilis）

<400> 418
Asp Leu Arg Ile Lys Leu Asp Pro Gly Ala Lys Ile Thr Gly Ile Ala
1 5 10 15

Leu Val

<210> 419
<211> 23
<212> PRT
<213> 多變魚腥藻菌

<400> 419
Ile Ser Arg Arg Gln Leu Arg Arg Thr Arg Arg Asn Arg Lys Thr Arg
1 5 10 15

Tyr Arg Lys Pro Arg Phe Leu
20

<210> 420
<211> 13
<212> PRT
<213> 多變魚腥藻菌

<400> 420
Ile Thr Ala Ile Ser Thr Glu Leu Val Lys Phe Asp Met
1 5 10

<210> 421
<211> 15
<212> PRT
<213> 多變魚腥藻菌

<400> 421
Lys Thr His Trp Leu Asp Ala Ala Cys Val Gly Gln Ser Thr Pro
1 5 10 15

<210> 422
<211> 18
<212> PRT
<213> 念珠藻屬某種

<400> 422
Pro Leu Arg Leu Lys Phe Asp Pro Gly Ala Lys Tyr Thr Gly Ile Ala
1 5 10 15

Leu Val

<210> 423
<211> 23
<212> PRT
<213> 念珠藻屬某種

<400> 423
Thr Ser Arg Arg Gln Leu Arg Arg Ser Arg Arg Ser Arg Lys Thr Arg
1 5 10 15

Tyr Arg Gln Pro Arg Phe Phe
20

<210> 424
<211> 13
<212> PRT
<213> 念珠藻屬某種

<400> 424
Ile Thr Ala Ile Ser Gln Glu Leu Val Lys Phe Asp Thr
1 5 10

<210> 425
<211> 15
<212> PRT
<213> 念珠藻屬某種

<400> 425
Lys Ser His Trp Leu Asp Ala Cys Cys Val Gly Ala Ser Thr Pro
1 5 10 15

<210> 426
<211> 18
<212> PRT
<213> 嗜熱棲熱菌（Thermus thermophilus）

<400> 426
Met Val Val Ala Gly Ile Asp Pro Gly Ile Thr His Leu Gly Leu Gly

1 5 10 15

Val Val

<210> 427
<211> 13
<212> PRT
<213> 嗜熱棲熱菌

<400> 427
Pro Glu Ala Val Ala Val Glu Glu Gln Phe Phe Tyr Arg
1 5 10

<210> 428
<211> 15
<212> PRT
<213> 嗜熱棲熱菌

<400> 428
Pro Ser His Leu Ala Asp Ala Leu Ala Ile Ala Leu Thr His Ala
1 5 10 15

<210> 429
<211> 10
<212> PRT
<213> 乳酸乳球菌（Lactococcus lactis）

<400> 429
Phe Leu Val Asn His Asn Tyr Tyr Ser Phe
1 5 10

<210> 430
<211> 16
<212> PRT
<213> 乳酸乳球菌

<400> 430
Leu Gln Lys Phe Thr Gly Asp Ile Glu Asn Leu Val Lys Ala Ser Leu
1 5 10 15

<210> 431
<211> 9
<212> PRT
<213> 乳酸乳球菌

<400> 431
Val Ile Val Pro Glu Leu Thr Phe Gly
1 5

<210> 432
<211> 15
<212> PRT
<213> 乳酸乳球菌

<400> 432
Trp Ile Arg Ala Gly Trp Phe Ile Arg Asn Arg Ser Ala His Tyr
1 5 10 15

<210> 433
<211> 13
<212> PRT
<213> 乳酸乳球菌

<400> 433
Asn Lys Asp Leu Phe Ala Phe Met Leu Ser Ile Lys Gln
1 5 10

<210> 434
<211> 10
<212> PRT
<213> 乳酸乳球菌

<400> 434
Phe Leu His Lys Asn Ser Tyr Phe Arg Phe

1 5 10

<210> 435
<211> 16
<212> PRT
<213> 乳酸乳球菌

<400> 435
Leu Phe Ile Phe Ser Thr Arg Leu Glu Ile Phe Trp Lys Lys Lys Ile
1 5 10 15

<210> 436
<211> 9
<212> PRT
<213> 乳酸乳球菌

<400> 436
Ala Leu Val Glu Glu Leu Thr Phe Gly
1 5

<210> 437
<211> 15
<212> PRT
<213> 乳酸乳球菌

<400> 437
Trp Met Asn Val Val Arg Leu Tyr Arg Asn Lys Ser Ala His Gly
1 5 10 15

<210> 438
<211> 13
<212> PRT
<213> 乳酸乳球菌

<400> 438
Lys Ser Tyr Leu Tyr Gly Ala Leu Tyr Val Phe Lys His
1 5 10

<210> 439
<211> 10
<212> PRT
<213> 白喉棒桿菌（Corynebacterium diphtheriae）

<400> 439
Leu Leu Ala Gln Leu Asn Tyr Tyr Arg Leu
1 5 10

<210> 440
<211> 16
<212> PRT
<213> 白喉棒桿菌

<400> 440
Val Phe Ile Glu Leu Asp Arg Val Glu Leu Ala Ile Gln Thr Arg Leu
1 5 10 15

<210> 441
<211> 9
<212> PRT
<213> 白喉棒桿菌

<400> 441
Ala Ala Val Glu Val Met Asp Trp Gly
1 5

<210> 442
<211> 15
<212> PRT
<213> 白喉棒桿菌

<400> 442
Trp Leu Lys Ser Leu Asn Ile Leu Arg Asn Tyr Ala Ala His His
1 5 10 15

<210> 443

<211> 13
<212> PRT
<213> 白喉棒桿菌

<400> 443
Gly Gln Leu Ser Met Ile Gln Tyr Leu His His Gln Leu
1 5 10

<210> 444
<211> 12
<212> PRT
<213> 波羅的海希瓦氏菌（Shewanella baltica）

<400> 444
Met Leu Ile Glu Asn Asp Leu Asp Gly Ile Glu Asn
1 5 10

<210> 445
<211> 16
<212> PRT
<213> 波羅的海希瓦氏菌

<400> 445
Asn Tyr Gln Leu Phe Tyr Phe Leu Glu Lys Thr Ile Arg Asn Gln Ile
1 5 10 15

<210> 446
<211> 15
<212> PRT
<213> 波羅的海希瓦氏菌

<400> 446
Val Met Phe Asn Leu Asn Thr Leu Arg Asn Pro Ile Ala His Cys
1 5 10 15

<210> 447
<211> 13
<212> PRT
<213> 波羅的海希瓦氏菌

<400> 447
Asp Glu Lys Leu Arg Leu Glu Ile Ser Leu Arg Asp Trp
1 5 10

<210> 448
<211> 12
<212> PRT
<213> 乳酸乳球菌

<400> 448
Leu Arg Glu Ile Asn Ile Lys Ala Ser Lys Ser Arg
1 5 10

<210> 449
<211> 16
<212> PRT
<213> 乳酸乳球菌

<400> 449
Leu Leu Pro Leu Leu His Lys Tyr Glu Trp Ser Leu Arg Lys Leu Ile
1 5 10 15

<210> 450
<211> 9
<212> PRT
<213> 乳酸乳球菌

<400> 450
Tyr Asp Phe Glu Glu Tyr Leu Phe Gly
1 5

<210> 451
<211> 15
<212> PRT
<213> 乳酸乳球菌

<400> 451
Asp Met Arg Leu Ile Arg Asp Gly Arg Asn Ile Val Gly His Asn
1 5 10 15

<210> 452
<211> 13
<212> PRT
<213> 乳酸乳球菌

<400> 452
Leu Ser Lys Gly Leu Lys Lys Tyr Ile Lys Lys Leu Asp
1 5 10

<210> 453
<211> 12
<212> PRT
<213> 伯米吉地桿菌（Geobacter bemidjiensis）

<400> 453
Arg Leu Pro Leu Thr Ser His Ile Gln Lys Gln Asp
1 5 10

<210> 454
<211> 16
<212> PRT
<213> 伯米吉地桿菌

<400> 454
Ile Tyr Pro Lys Leu Asn Arg Ile Glu Asn Arg Leu Arg His Tyr Leu
1 5 10 15

<210> 455
<211> 9
<212> PRT
<213> 伯米吉地桿菌

<400> 455
Phe Glu Leu Gly Lys Ile Val Tyr Ala
1 5

<210> 456
<211> 15
<212> PRT
<213> 伯米吉地桿菌

<400> 456
Lys Trp Ile Arg Leu Glu Glu Ile Arg His Lys Val Ala His Asn
1 5 10 15

<210> 457
<211> 13
<212> PRT
<213> 伯米吉地桿菌

<400> 457
Ala Asn Glu Tyr Ile Asp Ser Leu Gln Ser Ile Ile Asp
1 5 10

<210> 458
<211> 12
<212> PRT
<213> 腸道沙門氏菌（Salmonella enterica）

<400> 458
Phe Val Thr Ser Leu Glu His Leu Arg Gln Gln Gln
1 5 10

<210> 459
<211> 16
<212> PRT
<213> 腸道沙門氏菌

<400> 459
Ala Gln Arg Gln Leu Arg Ala Ile Glu Leu Thr Leu Lys Ala Leu Ile
1 5 10 15

<210> 460
<211> 9
<212> PRT
<213> 腸道沙門氏菌

<400> 460
Asn His Tyr Leu Lys Gln His Phe Gly
1 5

<210> 461
<211> 15
<212> PRT
<213> 腸道沙門氏菌

<400> 461
Phe Leu Asp Asp Cys Arg Leu Ala Arg Asn Glu Val Ile Ala Arg
1 5 10 15

<210> 462
<211> 13
<212> PRT
<213> 腸道沙門氏菌

<400> 462
Leu Met Leu Leu Asn Val Gln Tyr Gln Gln Ile Val Arg
1 5 10

<210> 463
<211> 12
<212> PRT
<213> 弗氏志賀菌（Shigella flexneri）

<400> 463
Phe Leu Trp Gln Leu Glu Tyr Leu Arg Glu Lys Gln
1 5 10

<210> 464
<211> 16
<212> PRT
<213> 弗氏志賀菌

<400> 464
Ser Leu Gln Gln Val Arg Ala Leu Glu Leu Thr Ile Arg Ser Leu Ile
1 5 10 15

<210> 465
<211> 9
<212> PRT
<213> 弗氏志賀菌

<400> 465
Leu Glu His Leu Asn Lys Leu Phe Gly
1 5

<210> 466
<211> 15
<212> PRT
<213> 弗氏志賀菌

<400> 466
Phe Leu Asp Asp Ile Arg Val Ile Arg Asn Arg Leu Ala His His
1 5 10 15

<210> 467
<211> 13
<212> PRT
<213> 弗氏志賀菌

<400> 467
Thr Thr Leu Val Asn Tyr Tyr Tyr Arg Glu Ile Thr Glu
1 5 10

<210> 468
<211> 16
<212> PRT

<213> 阿維鏈黴菌（Streptomyces avermitilis）

<400> 468
Ala Tyr Ile Trp Leu Asn Leu Val Glu Gln Arg Leu Arg Ala Val Val
1 5 10 15

<210> 469
<211> 9
<212> PRT
<213> 阿維鏈黴菌

<400> 469
Asn Val Leu Ser Phe Leu Thr Leu Pro
1 5

<210> 470
<211> 11
<212> PRT
<213> 阿維鏈黴菌

<400> 470
Leu Glu Val Thr Arg Asn Val Val Ser Arg Asn
1 5 10

<210> 471
<211> 13
<212> PRT
<213> 阿維鏈黴菌

<400> 471
Arg Tyr Gly Asp Val Val Gly Val His Pro Asp Arg Val
1 5 10

<210> 472
<211> 10
<212> PRT
<213> 幽門螺桿菌

<400> 472
Ser Ile Ser Val Leu His Tyr Asp Tyr Leu
1 5 10

<210> 473
<211> 16
<212> PRT
<213> 幽門螺桿菌

<400> 473
Leu Phe Leu Trp Ile His Phe Phe Glu Thr Ala Leu Arg Ser Lys Met
1 5 10 15

<210> 474
<211> 9
<212> PRT
<213> 幽門螺桿菌

<400> 474
Gln Ile Leu Asn Leu Phe Thr Leu Gly
1 5

<210> 475
<211> 15
<212> PRT
<213> 幽門螺桿菌

<400> 475
Thr Phe Ser Leu Ile Arg Lys Ala Arg Asn Asp Leu Phe His Asn
1 5 10 15

<210> 476
<211> 13
<212> PRT
<213> 幽門螺桿菌

<400> 476
Thr Leu Lys Leu Glu Arg Ala Ile Phe Phe Lys Thr Ile

1 5 10

<210> 477
<211> 11
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選神奇甲基產氧菌（Candidatus Methylomirabilis oxyfera）序列

<400> 477
Gly Pro Pro Glu Tyr Tyr Tyr Arg Leu Cys Arg
1 5 10

<210> 478
<211> 16
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選神奇甲基產氧菌序列

<400> 478
Ala Asp Ser Lys Leu Lys Asp Thr Val Ser Glu Met Arg Lys Phe Ile
1 5 10 15

<210> 479
<211> 15
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選神奇甲基產氧菌序列

<400> 479
Trp Met Asn Arg Ile Asn Glu Leu Arg Arg Ile Pro Ala His Pro
1 5 10 15

<210> 480
<211> 13
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選神奇甲基產氧菌序列

<400> 480
Asp Phe Glu Tyr Ile Asp Phe Ile Tyr Asp Glu Leu Met
1 5 10

<210> 481
<211> 12
<212> PRT
<213> 食芳烴新鞘胺醇菌

<400> 481
Thr Ala Val Lys Gln Gln Ser Phe Gly Met Glu Ala
1 5 10

<210> 482
<211> 16
<212> PRT
<213> 食芳烴新鞘胺醇菌

<400> 482
Ala Ala Ala Lys Val Thr Gln Ile His Lys Lys Leu Phe Asn Tyr Val
1 5 10 15

<210> 483
<211> 15
<212> PRT
<213> 食芳烴新鞘胺醇菌

<400> 483
Trp Ile Lys Val Leu Asn Asp Ile Arg Gln Tyr Thr Ala His Pro
1 5 10 15

<210> 484
<211> 13
<212> PRT
<213> 食芳烃新鞘胺醇菌

<400> 484
Gln Val Ser Phe Val Asn Glu Val Tyr Glu Lys Val Glu
1 5 10

<210> 485
<211> 11
<212> PRT
<213> 按蚊伊莉莎白菌（Elizabethkingia anophelis）

<400> 485
Gly Glu Ile Lys Tyr Trp Arg Thr Phe Gln Lys
1 5 10

<210> 486
<211> 16
<212> PRT
<213> 按蚊伊莉莎白菌

<400> 486
Ala Ile Ala Tyr Ile Arg Asp Ile Glu Thr Glu Phe Lys Ser Asp Phe
1 5 10 15

<210> 487
<211> 15
<212> PRT
<213> 按蚊伊莉莎白菌

<400> 487
Trp Met Val Lys Leu Glu Arg Ile Arg Asn Gln Asn Phe His Ser
1 5 10 15

<210> 488
<211> 13
<212> PRT
<213> 按蚊伊莉莎白菌

<400> 488
Glu Leu Ser Phe Leu Glu Glu Leu His Asp Trp Ile Tyr
1 5 10

<210> 489
<211> 12
<212> PRT
<213> 大腸桿菌

<400> 489
Phe Ser Ala Leu Pro Arg Ile Ile Glu Tyr Ala Tyr
1 5 10

<210> 490
<211> 16
<212> PRT
<213> 大腸桿菌

<400> 490
Pro Phe Leu Leu Leu Ser Glu Ile Glu Asn His Ile Arg Lys Leu Ile
1 5 10 15

<210> 491
<211> 9
<212> PRT
<213> 大腸桿菌

<400> 491
Glu Ser Val Ala Asp Leu Thr Phe Gly
1 5

<210> 492
<211> 15
<212> PRT
<213> 大腸桿菌

<400> 492
Glu Leu Asp Lys Val Arg Ile Ile Arg Asn Asp Val Met His Phe
1 5 10 15

<210> 493
<211> 13
<212> PRT
<213> 大腸桿菌

<400> 493
Asn His Glu Leu Leu His Asn Phe Val Arg Phe Ile His
1 5 10

<210> 494
<211> 12
<212> PRT
<213> 死海鹽盒菌（Haloarcula marismortui）

<400> 494
Phe Glu Leu Phe Asp Thr Leu Ala Glu Asp Asp Tyr
1 5 10

<210> 495
<211> 16
<212> PRT
<213> 死海鹽盒菌

<400> 495
Pro Phe Leu Gln Ile Gly Glu Ile Glu Glu Ser Leu Arg His Leu Phe
1 5 10 15

<210> 496
<211> 9
<212> PRT
<213> 死海鹽盒菌

<400> 496
Asp Arg Pro Glu Asp Phe Ser Phe Asp
1 5

<210> 497
<211> 15
<212> PRT
<213> 死海鹽盒菌

<400> 497
Leu Leu Glu Asp Ile Arg Glu Thr Arg Asn Ala Leu Leu His Phe
1 5 10 15

<210> 498
<211> 13
<212> PRT
<213> 死海鹽盒菌

<400> 498
Asp Arg Asp Gln Leu Asp Met Ala His Gly Tyr Phe Thr
1 5 10

<210> 499
<211> 12
<212> PRT
<213> 念珠藻屬某種

<400> 499
Met Lys Leu Leu Pro Ile Leu Gln Gln Asn Pro Arg
1 5 10

<210> 500
<211> 15
<212> PRT

<213> 念珠藻屬某種

<400> 500
Phe Gly Leu Val Thr Leu Leu Glu Met Asn Leu Leu Arg Leu Val
1 5 10 15

<210> 501
<211> 9
<212> PRT
<213> 念珠藻屬某種

<400> 501
Asp Leu Leu Asp Tyr Leu Gln Phe Cys
1 5

<210> 502
<211> 15
<212> PRT
<213> 念珠藻屬某種

<400> 502
Phe Leu Lys Ser Ala Glu Gln Leu Arg Asn Arg Leu Ala His Ala
1 5 10 15

<210> 503
<211> 13
<212> PRT
<213> 念珠藻屬某種

<400> 503
Ser Trp Asn Asp Leu Ile Ser Leu Ala Glu Ala Met Glu
1 5 10

<210> 504
<211> 10
<212> PRT
<213> 自養黃桿菌（Xanthobacter autotrophicus）

<400> 504
Val Phe Glu Gly Met Glu Leu Leu Pro Ala
1 5 10

<210> 505
<211> 14
<212> PRT
<213> 自養黃桿菌

<400> 505
Ala Leu Ile Pro Phe Val Glu Lys Arg Leu Glu Thr Ser Leu
1 5 10

<210> 506
<211> 8
<212> PRT
<213> 自養黃桿菌

<400> 506
Glu Ala Phe Lys Ala Val Leu Gly
1 5

<210> 507
<211> 15
<212> PRT
<213> 自養黃桿菌

<400> 507
Leu Val Asn Glu Leu Gly Asp Val Arg Asn Lys Leu Ser His Asn
1 5 10 15

<210> 508
<211> 13
<212> PRT
<213> 自養黃桿菌

<400> 508
Tyr Asp Asp Ala Glu Arg Ala Leu Asp Thr Met Arg Arg

1 5 10

<210> 509
<211> 10
<212> PRT
<213> 亨氏產甲烷螺菌（Methanospirillum hungatei）

<400> 509
Val Gly Arg Ala Met Asp Gln Leu Lys Thr
1 5 10

<210> 510
<211> 14
<212> PRT
<213> 亨氏產甲烷螺菌

<400> 510
Gly Leu Met Arg Phe Val Glu Arg Glu Met Lys Ser Ala Tyr
1 5 10

<210> 511
<211> 8
<212> PRT
<213> 亨氏產甲烷螺菌

<400> 511
Lys Val Phe Ser Gln Ile Leu Gly
1 5

<210> 512
<211> 15
<212> PRT
<213> 亨氏產甲烷螺菌

<400> 512
Leu Val Ser Glu Leu Arg Glu Thr Arg Asn Gln Trp Ala His Gln
1 5 10 15

<210> 513
<211> 13
<212> PRT
<213> 亨氏產甲烷螺菌

<400> 513
Thr Asn Asp Thr Leu Arg Ala Leu Asp Ser Thr Ala Arg
1 5 10

<210> 514
<211> 10
<212> PRT
<213> 玫瑰彎菌屬某種（Roseiflexus sp.）

<400> 514
Ile Gly Lys Ala Leu Asp Leu Leu Arg Gln
1 5 10

<210> 515
<211> 14
<212> PRT
<213> 玫瑰彎菌屬某種

<400> 515
Gly Leu Gln Pro Phe Ile Glu Arg Glu Leu Gln Asn His Tyr
1 5 10

<210> 516
<211> 8
<212> PRT
<213> 玫瑰彎菌屬某種

<400> 516
Asp Val Phe Arg Lys Thr Leu Gly
1 5

<210> 517

<211> 15
<212> PRT
<213> 玫瑰彎菌屬某種

<400> 517
Leu Val Ser Glu Leu Arg Glu Trp Arg Asn Lys Trp Ala His Gln
1 5 10 15

<210> 518
<211> 13
<212> PRT
<213> 玫瑰彎菌屬某種

<400> 518
Thr Asp Asp Thr Tyr Arg Val Leu Asp Ser Ala Ala Arg
1 5 10

<210> 519
<211> 10
<212> PRT
<213> 約氏瘧原蟲（Plasmodium yoelii）

<400> 519
Ile Leu Asn Ile Phe His Ile Leu Ser Ala
1 5 10

<210> 520
<211> 14
<212> PRT
<213> 約氏瘧原蟲

<400> 520
His Leu Ser Pro Ile Ile Glu Gln Ile Met Glu Met Glu Tyr
1 5 10

<210> 521
<211> 7
<212> PRT
<213> 約氏瘧原蟲

<400> 521
Asp Ile Phe Glu Asn Arg Ile
1 5

<210> 522
<211> 15
<212> PRT
<213> 約氏瘧原蟲

<400> 522
Ile Leu Glu Asn Leu Gln Lys Ala Ser Ile Phe Trp Ala Asn Gln
1 5 10 15

<210> 523
<211> 13
<212> PRT
<213> 約氏瘧原蟲

<400> 523
Glu Phe Phe Leu Ser Asn Leu Val Ser Ser Tyr Phe Phe
1 5 10

<210> 524
<211> 10
<212> PRT
<213> 小泰累爾氏梨漿蟲（Theileria parva）

<400> 524
Val Val Met Ile Phe Gln Cys Val Cys Asp
1 5 10

<210> 525
<211> 14
<212> PRT
<213> 小泰累爾氏梨漿蟲

<400> 525
Ala Phe Gln Pro Phe Ile Ser Lys Cys Met Leu Lys Lys Phe
1 5 10

<210> 526
<211> 7
<212> PRT
<213> 小泰累爾氏梨漿蟲

<400> 526
Asp Ile Phe Glu Gln Val Leu
1 5

<210> 527
<211> 15
<212> PRT
<213> 小泰累爾氏梨漿蟲

<400> 527
His Leu Asn Thr Ile Gln Thr Ala Ser Ile Tyr Trp Ala Asn Gln
1 5 10 15

<210> 528
<211> 8
<212> PRT
<213> 小泰累爾氏梨漿蟲

<400> 528
Asn Tyr Gly Lys Cys Arg Lys Ile
1 5

<210> 529
<211> 10
<212> PRT
<213> 蚤狀蚤（Daphnia pulex）

<400> 529
Ser Ser Lys Glu Ser Ala Ala Ile Ala Ile
1 5 10

<210> 530
<211> 14
<212> PRT
<213> 蚤狀蚤

<400> 530
Gly His Ile Val Phe Asp Thr Phe Leu Glu Asp Val Ala Pro
1 5 10

<210> 531
<211> 8
<212> PRT
<213> 蚤狀蚤

<400> 531
Asp Cys Phe Ile Ile Pro Pro Gly
1 5

<210> 532
<211> 15
<212> PRT
<213> 蚤狀蚤

<400> 532
Ile Leu Glu Arg Ala Met Asp Gly Arg His Ala Val Ser His His
1 5 10 15

<210> 533
<211> 13
<212> PRT
<213> 蚤狀蚤

<400> 533
Trp Glu Gln His Leu Lys Asp Tyr Val Tyr Ile Leu Thr
1 5 10

<210> 534
<211> 10
<212> PRT
<213> 智人

<400> 534
Ala Gly His Cys Leu Leu Leu Leu Arg Ser
1 5 10

<210> 535
<211> 16
<212> PRT
<213> 智人

<400> 535
Cys Leu Gln Gly Phe Val Gly Arg Glu Val Leu Ser Phe His Arg Gly
1 5 10 15

<210> 536
<211> 15
<212> PRT
<213> 智人

<400> 536
Lys Val Thr Glu Val Ile Lys Cys Arg Asn Glu Ile Met His Ser
1 5 10 15

<210> 537
<211> 13
<212> PRT
<213> 智人

<400> 537
Ser Ser Thr Trp Leu Arg Asp Phe Gln Met Lys Ile Gln
1 5 10

<210> 538
<211> 10
<212> PRT
<213> 佛羅里達文昌魚（Branchiostoma floridae）

<400> 538
Val Gly Ile Ala Leu Leu Thr Thr Arg Asp
1 5 10

<210> 539
<211> 16
<212> PRT
<213> 佛羅里達文昌魚

<400> 539
Gly Leu Thr Asn Val Thr Glu Gln Ala Ala Lys Glu Leu Gln Ala Glu
1 5 10 15

<210> 540
<211> 15
<212> PRT
<213> 佛羅里達文昌魚

<400> 540
Pro Leu Lys Asn Val Ile Glu Val Arg Asn Lys Thr Met His Ser
1 5 10 15

<210> 541
<211> 13
<212> PRT
<213> 佛羅里達文昌魚

<400> 541
Asp Arg Gln Thr Phe Asn Glu Tyr Met Asp Lys Met Glu
1 5 10

<210> 542
<211> 10
<212> PRT

<213> 智人

<400> 542
Val Ser Asp Leu Glu Lys Ser Leu Gly Thr
1 5 10

<210> 543
<211> 14
<212> PRT
<213> 智人

<400> 543
Gly Leu Ser Ser Ile Leu Glu Thr Glu Met Lys Ile Ala Phe
1 5 10

<210> 544
<211> 8
<212> PRT
<213> 智人

<400> 544
Lys His Trp Leu Ala Val Phe Gly
1 5

<210> 545
<211> 16
<212> PRT
<213> 智人

<400> 545
Thr Ile Glu Ser Leu Tyr Lys Asn Leu Arg Lys Ala Asn Lys Ala Val
1 5 10 15

<210> 546
<211> 13
<212> PRT
<213> 智人

<400> 546
Ser Arg Ser Leu Leu His Ala Phe Ser Thr Arg Ser Asn
1 5 10

<210> 547
<211> 10
<212> PRT
<213> 綠色鞭毛藻（Ostreococcus lucimarinus）

<400> 547
Met Glu Arg Leu Met Met Val Leu Asp His
1 5 10

<210> 548
<211> 14
<212> PRT
<213> 綠色鞭毛藻

<400> 548
Val Leu Ala Ile Val Leu Glu Gly Gly Leu Arg Ala Glu Phe
1 5 10

<210> 549
<211> 8
<212> PRT
<213> 綠色鞭毛藻

<400> 549
Ala Asn Trp Gly Ser Leu Phe Ser
1 5

<210> 550
<211> 14
<212> PRT
<213> 綠色鞭毛藻

<400> 550
Glu Ile Glu Val Leu Leu Asp Ala Ala Ile Arg Gln Arg Lys

1 5 10

<210> 551
<211> 13
<212> PRT
<213> 綠色鞭毛藻

<400> 551
Ala Arg Asp Val Ser Ser Ala Ala Val Ala Leu Leu Asn
1 5 10

<210> 552
<211> 10
<212> PRT
<213> 佛羅里達文昌魚

<400> 552
Leu Cys Gly Met Lys Thr Leu Leu Lys Ala
1 5 10

<210> 553
<211> 14
<212> PRT
<213> 佛羅里達文昌魚

<400> 553
Val Leu Ala Val Val Leu Glu Thr Glu Met Lys Ala Val Phe
1 5 10

<210> 554
<211> 8
<212> PRT
<213> 佛羅里達文昌魚

<400> 554
Lys His Trp Ile Ala Val Phe Gly
1 5

<210> 555
<211> 17
<212> PRT
<213> 佛羅里達文昌魚

<400> 555
His Leu Asp Ser Leu Val Lys His Phe Thr Arg Gly Arg Ser Tyr Gly
1 5 10 15

Val

<210> 556
<211> 13
<212> PRT
<213> 佛羅里達文昌魚

<400> 556
Ala Leu Gln Leu Val Arg Gln Leu His Asn His Ser Thr
1 5 10

<210> 557
<211> 10
<212> PRT
<213> 銅綠微囊藻（Microcystis aeruginosa）

<400> 557
Leu Asn Trp Leu Asp Gln Leu His Asp Asp
1 5 10

<210> 558
<211> 16
<212> PRT
<213> 銅綠微囊藻

<400> 558
Leu Ile Glu Leu Cys Gly Trp Ile Glu Glu Thr Met Asp Asp Ile Val

1 5 10 15

<210> 559
<211> 9
<212> PRT
<213> 銅綠微囊藻

<400> 559
Phe Arg Lys Met Leu Met Met Val Ile
1 5

<210> 560
<211> 15
<212> PRT
<213> 銅綠微囊藻

<400> 560
Tyr Leu Gly Asn Leu Lys Asp Ser Arg Asn Arg Ala Ala His Thr
1 5 10 15

<210> 561
<211> 13
<212> PRT
<213> 銅綠微囊藻

<400> 561
Phe Asp Lys Ile Tyr Gly Leu Leu Lys Glu Leu Asp Ala
1 5 10

<210> 562
<211> 12
<212> PRT
<213> 乳酸乳球菌

<400> 562
Leu Ser Glu Leu His Glu Phe Ile Lys Lys Leu Asn
1 5 10

<210> 563
<211> 16
<212> PRT
<213> 乳酸乳球菌

<400> 563
Val Ile Arg Ser Cys Gly Ile Ile Glu Gln Leu Thr Lys Thr Leu Ile
1 5 10 15

<210> 564
<211> 9
<212> PRT
<213> 乳酸乳球菌

<400> 564
Ile Asn Gly Leu Ile Asp Thr Phe Asp
1 5

<210> 565
<211> 15
<212> PRT
<213> 乳酸乳球菌

<400> 565
His Ile Asp Ser Leu Arg Gln Leu Arg Asn Ser Ile Ala His Gly
1 5 10 15

<210> 566
<211> 13
<212> PRT
<213> 乳酸乳球菌

<400> 566
Met Gly Tyr Phe Asp Ser Cys Ile Ile Leu Met Phe Arg
1 5 10

<210> 567

<211> 12
<212> PRT
<213> 弗蘭克氏菌屬某種（Frankia sp.）

<400> 567
Leu Ser Glu Leu Ala Ala Leu Val Gln Asp Gln Ala
1 5 10

<210> 568
<211> 16
<212> PRT
<213> 弗蘭克氏菌屬某種

<400> 568
Val Ile Arg Ser Cys Gly Tyr Leu Glu Gln Thr Val Ala Gly Thr Phe
1 5 10 15

<210> 569
<211> 9
<212> PRT
<213> 弗蘭克氏菌屬某種

<400> 569
Leu Glu Thr Leu Ala Gly Arg Phe Asp
1 5

<210> 570
<211> 15
<212> PRT
<213> 弗蘭克氏菌屬某種

<400> 570
Glu Leu Ala Thr Leu Val Asp Arg Arg Asn Arg Ile Ala His Gly
1 5 10 15

<210> 571
<211> 13
<212> PRT
<213> 弗蘭克氏菌屬某種

<400> 571
Leu Glu Leu His Arg Val Ala Cys Glu Ala Ala Asp Trp
1 5 10

<210> 572
<211> 10
<212> PRT
<213> 腦膜炎奈瑟氏菌（Neisseria meningitidis）

<400> 572
Cys Cys Ser Ile Phe Ser Asp Phe Arg Met
1 5 10

<210> 573
<211> 16
<212> PRT
<213> 腦膜炎奈瑟氏菌

<400> 573
Leu Phe His Val Val Ser Ile Phe Glu Ile Val Leu Arg Asn Lys Ile
1 5 10 15

<210> 574
<211> 9
<212> PRT
<213> 腦膜炎奈瑟氏菌

<400> 574
Gln Leu Val Ala Gly Leu Gly Phe Gly
1 5

<210> 575
<211> 15
<212> PRT
<213> 腦膜炎奈瑟氏菌

<400> 575
Glu Leu Ser Asn Ile Asn Lys Phe Arg Asn Arg Leu Ala His His
1 5 10 15

<210> 576
<211> 13
<212> PRT
<213> 腦膜炎奈瑟氏菌

<400> 576
Asp Val Asp Thr Ala Ser Val Phe Ser His Phe Ser Asp
1 5 10

<210> 577
<211> 10
<212> PRT
<213> 丁香假單胞菌（Pseudomonas syringae）

<400> 577
Leu Glu Lys His Phe Ser Ser Ala Arg Leu
1 5 10

<210> 578
<211> 16
<212> PRT
<213> 丁香假單胞菌

<400> 578
Met Met Pro Met Leu Ser Val Leu Glu Ile Ala Leu Lys Asn Gly Ile
1 5 10 15

<210> 579
<211> 9
<212> PRT
<213> 丁香假單胞菌

<400> 579
Lys Ile Val Ala Glu Leu Ala Phe Gly
1 5

<210> 580
<211> 15
<212> PRT
<213> 丁香假單胞菌

<400> 580
Ala Leu Asn Leu Ile Arg Asn Leu Arg Asn Arg Val Phe His His
1 5 10 15

<210> 581
<211> 13
<212> PRT
<213> 丁香假單胞菌

<400> 581
Asp Pro Gln Leu Val Pro Trp Leu Ala Gln Tyr Asp Arg
1 5 10

<210> 582
<211> 10
<212> PRT
<213> uraniireducens地桿菌（Geobacter uraniireducens）

<400> 582
Leu Arg Arg Ala Ile Ser His Glu Arg Leu
1 5 10

<210> 583
<211> 16
<212> PRT
<213> uraniireducens地桿菌

<400> 583
Leu Tyr Thr Pro Leu Gln Cys Leu Glu Val Cys Leu Arg Asn Ser Ile
1 5 10 15

<210> 584
<211> 9
<212> PRT
<213> uraniireducens地桿菌

<400> 584
Arg Ile Ile Pro Glu Leu Thr Phe Gly
1 5

<210> 585
<211> 15
<212> PRT
<213> uraniireducens地桿菌

<400> 585
Arg Phe Asn His Ile Arg Thr Leu Arg Asn Arg Ile Phe His His
1 5 10 15

<210> 586
<211> 13
<212> PRT
<213> uraniireducens地桿菌

<400> 586
Asn Pro Ala Met Met Thr Phe Val Glu Pro Phe Asp Ser
1 5 10

<210> 587
<211> 16
<212> PRT
<213> kujiense硫黏菌（Sulfuricurvum kujiense）

<400> 587
Glu Glu Lys Ser Glu Phe Ile Arg Glu Phe Phe Lys Arg Thr Leu His
1 5 10 15

<210> 588
<211> 9
<212> PRT
<213> kujiense硫黏菌

<400> 588
Thr Gln Thr Ile Asn Ser Phe Leu Gly
1 5

<210> 589
<211> 14
<212> PRT
<213> kujiense硫黏菌

<400> 589
Phe Arg Asn Tyr Leu Lys Arg Leu Arg Asn Ala Val Ser His
1 5 10

<210> 590
<211> 13
<212> PRT
<213> kujiense硫黏菌

<400> 590
Val Asn Leu Leu Ile Thr Leu Leu Ser Arg Asn Ile Leu
1 5 10

<210> 591
<211> 13
<212> PRT
<213> 嗜堿脫硫桿菌（Dethiobacter alkaliphilus）

<400> 591
Gln Val Val Glu Lys Asp Phe Val Ala Arg Thr Met His
1 5 10

<210> 592
<211> 9
<212> PRT

<213> 嗜堿脫硫桿菌

<400> 592
Thr Leu Leu Ile Asn Cys Leu Leu Gly
1 5

<210> 593
<211> 14
<212> PRT
<213> 嗜堿脫硫桿菌

<400> 593
Ala Ser Arg Phe Leu Gln Cys Met Arg Asn Ser Val Ala His
1 5 10

<210> 594
<211> 10
<212> PRT
<213> 嗜堿脫硫桿菌

<400> 594
Leu Ala Thr Lys Leu Ala Gln Tyr Val Gln
1 5 10

<210> 595
<211> 13
<212> PRT
<213> 肺炎克雷白氏桿菌（Klebsiella pneumoniae）

<400> 595
Ser Asp Phe Glu Thr Asp Phe Val Gln Arg Thr Leu Ala
1 5 10

<210> 596
<211> 9
<212> PRT
<213> 肺炎克雷白氏桿菌

<400> 596
Thr Leu Thr Leu Asn Cys Leu Leu Gly
1 5

<210> 597
<211> 14
<212> PRT
<213> 肺炎克雷白氏桿菌

<400> 597
Leu Arg Gln Leu Ile His Lys Met Arg Asn Ser Val Ala His
1 5 10

<210> 598
<211> 13
<212> PRT
<213> 肺炎克雷白氏桿菌

<400> 598
Leu Leu Pro Phe Leu Lys Tyr Tyr Ala Thr Leu Leu Leu
1 5 10

<210> 599
<211> 10
<212> PRT
<213> 乾酪乳酸桿菌（Lactobacillus casei）

<400> 599
Lys Ile Asp Arg Glu Met Phe Trp Arg Arg
1 5 10

<210> 600
<211> 16
<212> PRT
<213> 乾酪乳酸桿菌

<400> 600
Tyr Leu Leu Leu Tyr Ser Ser Trp Glu Gly Phe Ile Arg Ser Ile Ala

1 5 10 15

<210> 601
<211> 9
<212> PRT
<213> 乾酪乳酸桿菌

<400> 601
Leu Ala Arg Ile Val Ser Val Leu Asp
1 5

<210> 602
<211> 14
<212> PRT
<213> 乾酪乳酸桿菌

<400> 602
Asp Arg Asp Leu Leu Lys Val Arg Asn Glu Ile Ala His Gly
1 5 10

<210> 603
<211> 13
<212> PRT
<213> 乾酪乳酸桿菌

<400> 603
Thr Val Ser His Val Leu Glu Met Met Asp Leu Phe Ser
1 5 10

<210> 604
<211> 10
<212> PRT
<213> 柄桿菌屬某種（Caulobacter sp.）

<400> 604
Asp Leu Asp Ala Ala Arg Leu Arg Arg Ala
1 5 10

<210> 605
<211> 16
<212> PRT
<213> 柄桿菌屬某種

<400> 605
Ile Val Leu Ala Tyr Ser His Trp Glu Gly Phe Tyr Asn Glu Cys Ile
1 5 10 15

<210> 606
<211> 9
<212> PRT
<213> 柄桿菌屬某種

<400> 606
Leu Lys Glu Asn Phe Arg Ile Leu Gly
1 5

<210> 607
<211> 14
<212> PRT
<213> 柄桿菌屬某種

<400> 607
Asn Lys Glu Leu Val Gly Trp Arg His Ser Ile Ala His Gly
1 5 10

<210> 608
<211> 13
<212> PRT
<213> 柄桿菌屬某種

<400> 608
His Ile Ile Leu Thr Asn Ser Leu Leu Leu Thr Leu Ser
1 5 10

<210> 609

<211> 10
<212> PRT
<213> 銅綠微囊藻

<400> 609
Asn Leu Asp Glu Asp Met Ala Trp Arg Ile
1 5 10

<210> 610
<211> 16
<212> PRT
<213> 銅綠微囊藻

<400> 610
Ile Thr Thr Leu Tyr Ala His Trp Glu Gly Phe Ile Lys Tyr Ala Ala
1 5 10 15

<210> 611
<211> 9
<212> PRT
<213> 銅綠微囊藻

<400> 611
Phe Thr Asp Ile Cys Thr Ile Leu Gly
1 5

<210> 612
<211> 14
<212> PRT
<213> 銅綠微囊藻

<400> 612
Asp Glu Gln Leu Leu Thr Gln Arg Asn Lys Ile Ala His Gly
1 5 10

<210> 613
<211> 13
<212> PRT
<213> 銅綠微囊藻

<400> 613
Thr Tyr Asn Leu Val Ile Lys Leu Ile Arg Asp Phe Lys
1 5 10

<210> 614
<211> 11
<212> PRT
<213> 阿拉伯芥（Arabidopsis thaliana）

<400> 614
Pro Trp Leu Ser Trp Glu Glu Trp Asp Ser Val
1 5 10

<210> 615
<211> 15
<212> PRT
<213> 阿拉伯芥

<400> 615
Gly Ser Leu Pro Ala Pro Val Asp Val Thr Cys Ser Leu Ile Glu
1 5 10 15

<210> 616
<211> 9
<212> PRT
<213> 阿拉伯芥

<400> 616
Ile Ala Asp Ala Ala Arg Ala Ile Gly
1 5

<210> 617
<211> 15
<212> PRT
<213> 阿拉伯芥

<400> 617
Ile Pro Arg Lys Leu Ile Asp Leu Arg His Glu Gly Ser His Arg
1 5 10 15

<210> 618
<211> 13
<212> PRT
<213> 阿拉伯芥

<400> 618
Ala Ala Asp Glu Ala Leu Glu Trp Leu Lys Ser Tyr Tyr
1 5 10

<210> 619
<211> 11
<212> PRT
<213> 智人

<400> 619
Ala Trp Leu Ser Arg Ala Glu Trp Asp Gln Val
1 5 10

<210> 620
<211> 16
<212> PRT
<213> 智人

<400> 620
Gly Asn Glu Leu Pro Leu Ala Val Ala Ser Thr Ala Asp Leu Ile Arg
1 5 10 15

<210> 621
<211> 9
<212> PRT
<213> 智人

<400> 621
Leu Lys Cys Leu Ala Gln Glu Val Asn
1 5

<210> 622
<211> 15
<212> PRT
<213> 智人

<400> 622
Ile Pro Asp Trp Ile Val Asp Leu Arg His Glu Leu Thr His Lys
1 5 10 15

<210> 623
<211> 13
<212> PRT
<213> 智人

<400> 623
Gly Cys Tyr Phe Val Leu Asp Trp Leu Gln Lys Thr Tyr
1 5 10

<210> 624
<211> 11
<212> PRT
<213> 釀酒酵母（Saccharomyces cerevisiae）

<400> 624
Pro Trp Arg Asp Phe Ala Glu Leu Glu Glu Leu
1 5 10

<210> 625
<211> 16
<212> PRT
<213> 釀酒酵母

<400> 625
Ser Gln Tyr Leu Pro His Val Val Asp Ser Thr Ala Gln Ile Thr Cys
1 5 10 15

<210> 626
<211> 9
<212> PRT
<213> 釀酒酵母

<400> 626
Leu His Thr Leu Ala Ala Lys Ile Gly
1 5

<210> 627
<211> 15
<212> PRT
<213> 釀酒酵母

<400> 627
Leu Pro Ser Trp Phe Val Asp Leu Arg His Trp Gly Thr His Glu
1 5 10 15

<210> 628
<211> 13
<212> PRT
<213> 釀酒酵母

<400> 628
Ala Ala Asn Glu Ala Leu Ser Trp Leu Tyr Asp His Tyr
1 5 10

<210> 629
<211> 11
<212> PRT
<213> 肺炎鏈球菌（Streptococcus pneumoniae）

<400> 629
Ser Lys Pro Cys Ile Glu Ala Glu Asn Met Ile
1 5 10

<210> 630
<211> 16
<212> PRT
<213> 肺炎鏈球菌

<400> 630
Ala Phe Met Ala Arg Arg Ala Leu Glu Gln Ala Val His Trp Ile Tyr
1 5 10 15

<210> 631
<211> 8
<212> PRT
<213> 肺炎鏈球菌

<400> 631
Ser Ser Leu Val Trp Asp Asp Asp
1 5

<210> 632
<211> 15
<212> PRT
<213> 肺炎鏈球菌

<400> 632
Gln Ile Val Leu Leu Ile Arg Trp Gly Asn His Ala Ala His Gly
1 5 10 15

<210> 633
<211> 13
<212> PRT
<213> 肺炎鏈球菌

<400> 633
Ala Leu His His Leu Tyr Gln Phe Val Asn Phe Ile Asp
1 5 10

<210> 634
<211> 11
<212> PRT

<213> 銅綠微囊藻

<400> 634
Tyr Asp His Ala Ser Gln Ala Glu Gly Leu Val
1 5 10

<210> 635
<211> 16
<212> PRT
<213> 銅綠微囊藻

<400> 635
Cys Phe Tyr Thr Arg Phe Val Leu Glu Gln Met Val Cys Trp Leu Tyr
1 5 10 15

<210> 636
<211> 8
<212> PRT
<213> 銅綠微囊藻

<400> 636
Gly Ala Leu Ile His Glu Gln Thr
1 5

<210> 637
<211> 15
<212> PRT
<213> 銅綠微囊藻

<400> 637
Lys Ile Arg Thr Ile His Lys Val Gly Asn Asn Ala Ala His Asp
1 5 10 15

<210> 638
<211> 13
<212> PRT
<213> 銅綠微囊藻

<400> 638
Leu Ile Glu Glu Leu Phe His Leu Thr Tyr Trp Leu Val
1 5 10

<210> 639
<211> 11
<212> PRT
<213> 大腸桿菌

<400> 639
Tyr Ala Ile Ala Cys Ala Ala Glu Asn Asn Tyr
1 5 10

<210> 640
<211> 16
<212> PRT
<213> 大腸桿菌

<400> 640
Leu Ile Lys Met Arg Met Phe Gly Glu Ala Thr Ala Lys His Leu Gly
1 5 10 15

<210> 641
<211> 8
<212> PRT
<213> 大腸桿菌

<400> 641
His Asp Leu Leu Arg Glu Leu Gly
1 5

<210> 642
<211> 15
<212> PRT
<213> 大腸桿菌

<400> 642
Val Phe His Lys Leu Arg Arg Ile Gly Asn Gln Ala Val His Glu

1 5 10 15

<210> 643
<211> 13
<212> PRT
<213> 大腸桿菌

<400> 643
Cys Leu Arg Leu Gly Phe Arg Leu Ala Val Trp Tyr Tyr
1 5 10

<210> 644
<211> 11
<212> PRT
<213> 大豆慢生根瘤菌（Bradyrhizobium japonicum）

<400> 644
Val Gln Lys Leu Ile Lys Ala Ser Gln Leu Ala
1 5 10

<210> 645
<211> 16
<212> PRT
<213> 大豆慢生根瘤菌

<400> 645
Leu Thr Glu Val Arg Arg Ala Met Lys Ala Ala Ala Asp Leu Phe Trp
1 5 10 15

<210> 646
<211> 10
<212> PRT
<213> 大豆慢生根瘤菌

<400> 646
Leu Asn Arg Leu Gln Glu Phe Ala Arg Val
1 5 10

<210> 647
<211> 13
<212> PRT
<213> 大豆慢生根瘤菌

<400> 647
Arg Arg Leu Asn Asp Leu Ala Ser Lys Gly Val His Ala
1 5 10

<210> 648
<211> 13
<212> PRT
<213> 大豆慢生根瘤菌

<400> 648
Ala Glu Ala Arg Gln Gly Leu Val Gly Leu Tyr Phe Phe
1 5 10

<210> 649
<211> 11
<212> PRT
<213> 邁氏鉤端螺旋體（Leptospira meyeri）

<400> 649
Leu Pro Lys Phe Ser Ala Ile Tyr Ser Asn Leu
1 5 10

<210> 650
<211> 16
<212> PRT
<213> 邁氏鉤端螺旋體

<400> 650
Val His Ser Cys Arg Arg Leu Leu Gln Ser Val Ala Asp Lys Leu Met
1 5 10 15

<210> 651

<211> 10
<212> PRT
<213> 邁氏鉤端螺旋體

<400> 651
Ile Asn Arg Leu Ile Tyr Tyr Ile Glu Thr
1 5 10

<210> 652
<211> 13
<212> PRT
<213> 邁氏鉤端螺旋體

<400> 652
Asp Ser Val Phe Gln Ala Ser Gln Lys Gly Ser His Ser
1 5 10

<210> 653
<211> 13
<212> PRT
<213> 邁氏鉤端螺旋體

<400> 653
Gln Glu Ala Asp Arg Tyr Val Ile His Thr Phe Leu Leu
1 5 10

<210> 654
<211> 11
<212> PRT
<213> coprosuis擬桿菌

<400> 654
Val Val Asp Asp Arg Asp Phe Ser Leu Leu Ala
1 5 10

<210> 655
<211> 16
<212> PRT
<213> coprosuis擬桿菌

<400> 655
Leu Asp Arg Leu His Thr Tyr Val Ile Lys Phe Ile Arg Gln Leu Cys
1 5 10 15

<210> 656
<211> 9
<212> PRT
<213> coprosuis擬桿菌

<400> 656
Phe Gly Lys Tyr Val Lys Phe Ile Val
1 5

<210> 657
<211> 16
<212> PRT
<213> coprosuis擬桿菌

<400> 657
Ile Glu Ala Phe Asn Asp Ile Arg Asn Asn Lys Ser Phe Ala His Asp
1 5 10 15

<210> 658
<211> 13
<212> PRT
<213> coprosuis擬桿菌

<400> 658
Tyr Ala Glu Ser Val Leu Ile Phe Asn Asn Val Thr Asn
1 5 10

<210> 659
<211> 10
<212> PRT
<213> 大腸桿菌

<400> 659
Asn Val Asn Glu Asn Ile Tyr Gln Ala Leu
1 5 10

<210> 660
<211> 16
<212> PRT
<213> 大腸桿菌

<400> 660
Tyr Asp Arg Val His Thr Ala Leu His Ala Ser Leu Arg Gln Met Cys
1 5 10 15

<210> 661
<211> 9
<212> PRT
<213> 大腸桿菌

<400> 661
Leu Ser Leu Ile Thr Ala His Leu Lys
1 5

<210> 662
<211> 16
<212> PRT
<213> 大腸桿菌

<400> 662
Leu His Gly Ile Asn Asn Leu Arg Asn Asn Tyr Ser Met Ala His Pro
1 5 10 15

<210> 663
<211> 13
<212> PRT
<213> 大腸桿菌

<400> 663
Glu Ala Asp Ala Arg Phe Ala Ile Asn Leu Val Arg Ser
1 5 10

<210> 664
<211> 11
<212> PRT
<213> 乳酸乳球菌

<400> 664
Ile Met Asn Ile Gly Tyr Val Glu Lys Ile Leu
1 5 10

<210> 665
<211> 16
<212> PRT
<213> 乳酸乳球菌

<400> 665
Val Thr Lys Ser Arg Thr Ile Ile Glu Thr Val Phe Ile Ala Ile Leu
1 5 10 15

<210> 666
<211> 9
<212> PRT
<213> 乳酸乳球菌

<400> 666
Arg Ser Leu Val Asn Lys Thr Leu Gly
1 5

<210> 667
<211> 16
<212> PRT
<213> 乳酸乳球菌

<400> 667
Val Asp Ser Ile Thr Thr Met Arg Asn Ile Asn Ser Asp Ser His Gly
1 5 10 15

<210> 668
<211> 13
<212> PRT
<213> 乳酸乳球菌

<400> 668
Glu Ala Glu Ala Glu Leu Ile Leu Asn Ser Ala Val Asn
1 5 10

<210> 669
<211> 11
<212> PRT
<213> 吲哚嗜胨菌（Peptoniphilus indolicus）

<400> 669
Phe Leu Tyr Leu Lys Thr Leu Lys Asn Lys Glu
1 5 10

<210> 670
<211> 16
<212> PRT
<213> 吲哚嗜胨菌

<400> 670
Arg Gly Ile Thr Pro Leu Val Thr Glu Leu Phe Ile Leu Ile Ile Asp
1 5 10 15

<210> 671
<211> 9
<212> PRT
<213> 吲哚嗜胨菌

<400> 671
Leu Ile Glu Ile Ile Lys Asn Glu Arg
1 5

<210> 672
<211> 15
<212> PRT
<213> 吲哚嗜胨菌

<400> 672
Ile Arg Asp Val Glu Gly Lys Leu Arg Asn Arg Ala Ala His Glu
1 5 10 15

<210> 673
<211> 13
<212> PRT
<213> 吲哚嗜胨菌

<400> 673
Gly Asn Asn His Tyr Asp Ser Tyr Asp Leu Met Asn Lys
1 5 10

<210> 674
<211> 11
<212> PRT
<213> 結核分枝桿菌（Mycobacterium tuberculosis）

<400> 674
Ile Ser Ala Leu Ala Leu Leu Ala Lys Arg Glu
1 5 10

<210> 675
<211> 16
<212> PRT
<213> 結核分枝桿菌

<400> 675
Arg Ser Ala Thr Pro Ala Ile Thr Ile Val Leu Arg Ala Ala Val Ala
1 5 10 15

<210> 676
<211> 9
<212> PRT

<213> 結核分枝桿菌

<400> 676
Trp Leu Ala Leu Leu Arg Gln Phe Ala
1 5

<210> 677
<211> 15
<212> PRT
<213> 結核分枝桿菌

<400> 677
Leu Gly Arg Phe Glu Ser Arg Val Arg Asn Thr Ala Ala His Glu
1 5 10 15

<210> 678
<211> 11
<212> PRT
<213> 結核分枝桿菌

<400> 678
Ala Asp Leu Thr Leu Tyr Asp Arg Leu Asn Asp
1 5 10

<210> 679
<211> 12
<212> PRT
<213> 嗜熱鏈球菌

<400> 679
Tyr Leu Met Ile Asp Val Leu Lys Glu Arg Glu His
1 5 10

<210> 680
<211> 16
<212> PRT
<213> 嗜熱鏈球菌

<400> 680
Ile Glu Glu Ile Ile Lys Lys Asp His Glu Gly Leu Ile Val Phe Asp
1 5 10 15

<210> 681
<211> 9
<212> PRT
<213> 嗜熱鏈球菌

<400> 681
Tyr Leu Asn Ile Leu Glu Phe Tyr Glu
1 5

<210> 682
<211> 14
<212> PRT
<213> 嗜熱鏈球菌

<400> 682
Ile Leu Ser Leu Asn Gly Glu Arg Asn Lys Val Ala His Gly
1 5 10

<210> 683
<211> 13
<212> PRT
<213> 嗜熱鏈球菌

<400> 683
Asp Ser Ser Tyr Phe Asn Tyr Tyr Asp Lys Gln Asn Lys
1 5 10

<210> 684
<211> 10
<212> PRT
<213> 集胞藻屬某種（Synechocystis sp.）

<400> 684
Leu Ile Ser Val Val Ala Phe Arg Leu Gly

1 5 10

<210> 685
<211> 16
<212> PRT
<213> 集胞藻屬某種

<400> 685
Ile Leu Asp His Arg Lys Gln Ile Asn Phe Ala Leu Asn Asn Gly Gly
1 5 10 15

<210> 686
<211> 10
<212> PRT
<213> 集胞藻屬某種

<400> 686
Thr Glu Ile Arg Asn Asp Leu Ala His Cys
1 5 10

<210> 687
<211> 12
<212> PRT
<213> 集胞藻屬某種

<400> 687
Asn Lys Ile Phe Pro Gln Leu Glu Glu Ile Ala Asn
1 5 10

<210> 688
<211> 5
<212> PRT
<213> 詹氏甲烷球菌（Methanocaldococcus jannaschii）

<400> 688
Lys Asn Thr Leu Phe
1 5

<210> 689
<211> 14
<212> PRT
<213> 詹氏甲烷球菌

<400> 689
Lys Glu Asn Pro Asn Ser Gln Tyr Ile Lys Asn Glu Ile Ser
1 5 10

<210> 690
<211> 15
<212> PRT
<213> 詹氏甲烷球菌

<400> 690
Glu Asn Ile Asp Lys Phe Lys Ile Arg Asn Phe Leu Ala His Ala
1 5 10 15

<210> 691
<211> 13
<212> PRT
<213> 詹氏甲烷球菌

<400> 691
Ser Glu Lys Thr Ser Leu Arg Tyr Asn Lys Asn Tyr Ile
1 5 10

<210> 692
<211> 11
<212> PRT
<213> 強烈火球菌（Pyrococcus furiosus）

<400> 692
Ser Lys Ile Phe Glu Ser Leu Pro Arg Ile Gly
1 5 10

<210> 693

<211> 15
<212> PRT
<213> 強烈火球菌

<400> 693
Arg Gln Val Glu Trp Leu Arg Asn Leu Val Tyr Gly Arg Leu Trp
1 5 10 15

<210> 694
<211> 15
<212> PRT
<213> 強烈火球菌

<400> 694
Thr Ile Glu Ser Pro Asn Val Val Arg Asn Phe Ile Ala His Ser
1 5 10 15

<210> 695
<211> 13
<212> PRT
<213> 強烈火球菌

<400> 695
Asp Lys Glu Lys Ala Ala Asn Leu Ala Tyr Glu Ala Leu
1 5 10

<210> 696
<211> 5
<212> PRT
<213> 硫磺礦硫化葉菌

<400> 696
Ala Glu Thr Tyr Ala
1 5

<210> 697
<211> 13
<212> PRT
<213> 硫磺礦硫化葉菌

<400> 697
Asp Lys Val Thr Arg Ala Ile Ile Glu Asn Glu Val Asp
1 5 10

<210> 698
<211> 13
<212> PRT
<213> 硫磺礦硫化葉菌

<400> 698
Gly Lys Gly Phe Asp Lys Arg Ile Leu Tyr Ala His Gly
1 5 10

<210> 699
<211> 10
<212> PRT
<213> 硫磺礦硫化葉菌

<400> 699
Asp Lys Ile Asp Glu Ile Glu Arg Gln Ile
1 5 10

<210> 700
<211> 12
<212> PRT
<213> 食油脫硫球菌（Desulfococcus oleovorans）

<400> 700
Phe Ala Asn Ala Glu Arg Arg Phe Asp Glu Gly Lys
1 5 10

<210> 701
<211> 16
<212> PRT
<213> 食油脫硫球菌

<400> 701
Val Leu Arg Leu Tyr Arg Ile Val Glu Met Ala Gly Gln Gln Arg Leu
1 5 10 15

<210> 702
<211> 9
<212> PRT
<213> 食油脫硫球菌

<400> 702
Gly Tyr Ser Leu Leu Lys Glu Met Gly
1 5

<210> 703
<211> 17
<212> PRT
<213> 食油脫硫球菌

<400> 703
Ser Phe Leu Lys Ile Gln Asp Ser Arg Asn His Ser Phe Leu Ala His
1 5 10 15

Gly

<210> 704
<211> 13
<212> PRT
<213> 食油脫硫球菌

<400> 704
Tyr Met Ser Leu Arg Asp Phe Ile Val Ser Leu Asn Ile
1 5 10

<210> 705
<211> 12
<212> PRT
<213> 毛樣顫綠菌（Oscillochloris trichoides）

<400> 705
Leu Arg Asn Ala Glu Arg Arg Ala Ala Arg Ala Arg
1 5 10

<210> 706
<211> 16
<212> PRT
<213> 毛樣顫綠菌

<400> 706
Val Ala Arg Leu Tyr Arg Ala Thr Glu Leu Phe Ala Gln Ile Arg Leu
1 5 10 15

<210> 707
<211> 9
<212> PRT
<213> 毛樣顫綠菌

<400> 707
Ser Tyr Ala Leu Leu Gly Lys Leu Asp
1 5

<210> 708
<211> 17
<212> PRT
<213> 毛樣顫綠菌

<400> 708
Pro Leu Asn Asn Ala Leu Thr Arg Arg Asn Gln Ser Ile Leu Ala His
1 5 10 15

Gly

<210> 709

<211> 13
<212> PRT
<213> 毛樣顫綠菌

<400> 709
Tyr His Asp Leu Ala Ser His Leu Tyr Thr Leu Ile Asn
1 5 10

<210> 710
<211> 12
<212> PRT
<213> 智人

<400> 710
Phe Pro Glu Ile Phe Asp Ala Leu Glu Ser Leu Gln
1 5 10

<210> 711
<211> 14
<212> PRT
<213> 智人

<400> 711
Lys Leu Thr Ser Cys Leu Glu Arg Ala Leu Gly Asp Val Phe
1 5 10

<210> 712
<211> 9
<212> PRT
<213> 智人

<400> 712
Ser Glu Glu Leu Ala Gln Val Phe Ser
1 5

<210> 713
<211> 15
<212> PRT
<213> 智人

<400> 713
Gly Ser Pro Cys Gly Leu Asn Leu Arg Asn Val Leu Trp His Gly
1 5 10 15

<210> 714
<211> 13
<212> PRT
<213> 智人

<400> 714
Tyr Cys Ser Met Met Ile Leu Leu Thr Ala Gly Leu Gly
1 5 10

<210> 715
<211> 12
<212> PRT
<213> 溶組織內阿米巴（Entamoeba histolytica）

<400> 715
Trp Phe Glu Ser Phe Gln Glu Ile Ile Gln Thr Pro
1 5 10

<210> 716
<211> 14
<212> PRT
<213> 溶組織內阿米巴

<400> 716
Leu Leu Ser Val Gln Phe Asn Val His Leu Lys Asp Asn Ile
1 5 10

<210> 717
<211> 9
<212> PRT
<213> 溶組織內阿米巴

<400> 717
Lys Met Tyr Glu Glu His Thr Val Pro
1 5

<210> 718
<211> 15
<212> PRT
<213> 溶組織內阿米巴

<400> 718
Gly Pro Pro Thr Gly Leu Asn Leu Arg Asn Leu Leu Trp His Gly
1 5 10 15

<210> 719
<211> 13
<212> PRT
<213> 溶組織內阿米巴

<400> 719
His Ile Cys Leu Leu Ile Ile Leu Tyr Gln Thr Ile Gln
1 5 10

<210> 720
<211> 12
<212> PRT
<213> 金黃色葡萄球菌（Staphylococcus aureus）

<400> 720
Ile Glu His Gly Ile Ser Arg Phe Leu Glu Lys Asp
1 5 10

<210> 721
<211> 14
<212> PRT
<213> 金黃色葡萄球菌

<400> 721
Ile Leu Val Pro Gln Phe Glu Ser Thr Val Arg Arg Met Phe
1 5 10

<210> 722
<211> 9
<212> PRT
<213> 金黃色葡萄球菌

<400> 722
Arg Asp Asp Val Lys Ser Thr Leu Gly
1 5

<210> 723
<211> 15
<212> PRT
<213> 金黃色葡萄球菌

<400> 723
Val Glu Gln Ser Gly Leu Asn Leu Arg Asn Glu Ile Ala His Gly
1 5 10 15

<210> 724
<211> 12
<212> PRT
<213> 金黃色葡萄球菌

<400> 724
Lys Cys Ile Leu Val Ile Tyr Leu Phe Leu Ile Leu
1 5 10

<210> 725
<211> 12
<212> PRT
<213> 藍杆藻屬某種（Cyanothece sp.）

<400> 725
Leu Leu Lys Gly Ile Gln Ala Tyr Leu Glu Glu Asp
1 5 10

<210> 726
<211> 14
<212> PRT
<213> 藍杆藻屬某種

<400> 726
Leu Leu Ile Pro Gln Ile Glu Ala Ala Ile Arg Asn Leu Val
1 5 10

<210> 727
<211> 9
<212> PRT
<213> 藍杆藻屬某種

<400> 727
Ser Glu Gln Val Lys Gln Ser Leu Gly
1 5

<210> 728
<211> 15
<212> PRT
<213> 藍杆藻屬某種

<400> 728
Thr Asp Gln Arg Gly Trp Asn Val Arg Asn Asn Val Cys His Gly
1 5 10 15

<210> 729
<211> 12
<212> PRT
<213> 藍杆藻屬某種

<400> 729
Leu Thr Glu Arg Leu Ile His Ile Leu Leu Ile Leu
1 5 10

<210> 730
<211> 9
<212> PRT
<213> 硫磺礦硫化葉菌

<400> 730
Ile Ser Thr Ser Ala Glu Val Tyr Tyr
1 5

<210> 731
<211> 16
<212> PRT
<213> 硫磺礦硫化葉菌

<400> 731
Cys Glu Lys Tyr Tyr Lys Ala Ala Glu Glu Ala Ile Lys Leu Leu Val
1 5 10 15

<210> 732
<211> 8
<212> PRT
<213> 硫磺礦硫化葉菌

<400> 732
Lys Leu Leu Arg Ser Asn Asn Thr
1 5

<210> 733
<211> 15
<212> PRT
<213> 硫磺礦硫化葉菌

<400> 733
Leu Trp Lys Ser Ala Trp Thr Leu His Val Glu Gly Phe His Glu
1 5 10 15

<210> 734
<211> 13
<212> PRT

<213> 硫磺礦硫化葉菌

<400> 734
Leu Lys Glu Asp Val Arg Lys Leu Val Ile Phe Ala Val
1 5 10

<210> 735
<211> 9
<212> PRT
<213> 需氧耐超高溫熱棒菌（Pyrobaculum aerophilum）

<400> 735
Tyr Ala Glu Ala Ala Arg Glu Leu Leu
1 5

<210> 736
<211> 16
<212> PRT
<213> 需氧耐超高溫熱棒菌

<400> 736
Ser Glu Lys Ala Trp Gly Ala Ala Ala Leu Ala Val Lys Ala Tyr Ala
1 5 10 15

<210> 737
<211> 7
<212> PRT
<213> 需氧耐超高溫熱棒菌

<400> 737
Lys Ile Ala Gly Glu Leu Gly
1 5

<210> 738
<211> 14
<212> PRT
<213> 需氧耐超高溫熱棒菌

<400> 738
Ala Trp Ala Gln Ala Asn Ala Met His Ile Asn Phe Tyr Glu
1 5 10

<210> 739
<211> 13
<212> PRT
<213> 需氧耐超高溫熱棒菌

<400> 739
Ala Leu Lys Lys Val Ser Arg Leu Val Glu Glu Leu Thr
1 5 10

<210> 740
<211> 11
<212> PRT
<213> 智人

<400> 740
Arg Arg Trp Leu Arg Gln Ala Arg Ala Asn Phe
1 5 10

<210> 741
<211> 16
<212> PRT
<213> 智人

<400> 741
Asn Glu Trp Val Cys Phe Lys Cys Tyr Leu Ser Thr Lys Leu Ala Leu
1 5 10 15

<210> 742
<211> 8
<212> PRT
<213> 智人

<400> 742
Ala Gln Lys Ile Glu Glu Tyr Ser

1 5

<210> 743
<211> 15
<212> PRT
<213> 智人

<400> 743
Val His Thr Leu Glu Ala Tyr Gly Val Asp Ser Leu Lys Thr Arg
1 5 10 15

<210> 744
<211> 13
<212> PRT
<213> 智人

<220>
<221> MOD_RES
<222> (2)..(2)
<223>任何胺基酸

<400> 744
Val Xaa Glu Cys Thr Ala Cys Ile Ile Ile Lys Leu Glu
1 5 10

<210> 745
<211> 11
<212> PRT
<213> 流感嗜血桿菌（Haemophilus influenzae）

<400> 745
Lys Leu Asn Leu Asn Val Leu Asp Ala Ala Phe
1 5 10

<210> 746
<211> 16
<212> PRT
<213> 流感嗜血桿菌

<220>
<221> MOD_RES
<222> (14)..(15)
<223>任何胺基酸

<400> 746
Ile Gln Lys Phe Glu Phe Val Tyr Glu Leu Ser Leu Lys Xaa Xaa Lys
1 5 10 15

<210> 747
<211> 8
<212> PRT
<213> 流感嗜血桿菌

<400> 747
Leu Arg Glu Ala Leu Arg Phe Gly
1 5

<210> 748
<211> 15
<212> PRT
<213> 流感嗜血桿菌

<220>
<221> MOD_RES
<222> (8)..(8)
<223>任何胺基酸

<400> 748
Lys Trp Val Ala Tyr Arg Asp Xaa Arg Asn Ile Thr Ser His Thr
1 5 10 15

<210> 749
<211> 13
<212> PRT

<213> 流感嗜血桿菌

<400> 749
Asp Phe Leu Ile Glu Ser Ser Phe Leu Leu Glu Gln Leu
1 5 10

<210> 750
<211> 6
<212> PRT
<213> 嗜熱棲熱菌

<220>
<221> MOD_RES
<222> (1)..(1)
<223>任何胺基酸

<400> 750
Xaa Ala Glu Lys Ala Leu
1 5

<210> 751
<211> 16
<212> PRT
<213> 嗜熱棲熱菌

<400> 751
Ile Gln Arg Phe Glu Tyr Thr Phe Glu Ala Phe Trp Lys Ala Leu Gln
1 5 10 15

<210> 752
<211> 8
<212> PRT
<213> 嗜熱棲熱菌

<400> 752
Ile Arg Leu Ala Arg Glu Val Gly
1 5

<210> 753
<211> 15
<212> PRT
<213> 嗜熱棲熱菌

<220>
<221> MOD_RES
<222> (5)..(5)
<223>任何胺基酸

<400> 753
Leu Ala Leu Gly Xaa Val Asp Asp Arg Ser Leu Thr Val His Thr
1 5 10 15

<210> 754
<211> 13
<212> PRT
<213> 嗜熱棲熱菌

<220>
<221> MOD_RES
<222> (12)..(12)
<223>任何胺基酸

<400> 754
Ile Phe Arg Arg Leu Pro Asp Tyr Ala Arg Leu Xaa Glu
1 5 10

<210> 755
<211> 11
<212> PRT
<213> 馬紅球菌（Rhodococcus equi）

<400> 755
Val Asn Leu Leu Arg Arg Ala Asp Gly Leu Leu
1 5 10

<210> 756
<211> 16
<212> PRT
<213> 馬紅球菌

<400> 756
Phe Cys Ala Ala Tyr Val Gly Ala Leu Arg Gly Ala Ala Ala Val Leu
1 5 10 15

<210> 757
<211> 8
<212> PRT
<213> 馬紅球菌

<400> 757
Trp Val Leu Met Ala Arg Ala Glu
1 5

<210> 758
<211> 15
<212> PRT
<213> 馬紅球菌

<400> 758
Tyr Phe Ala Gly Tyr Ser Gly Leu Arg Ala Asp Leu Glu Ala Gly
1 5 10 15

<210> 759
<211> 13
<212> PRT
<213> 馬紅球菌

<400> 759
Asp Ala Glu Glu Val Asp Gly Phe Tyr Ala Glu Val Gly
1 5 10

<210> 760
<211> 11
<212> PRT
<213> 阿維鏈黴菌

<400> 760
Leu Asp Leu Leu Ala Gln Ala Arg Ala Gly Leu
1 5 10

<210> 761
<211> 16
<212> PRT
<213> 阿維鏈黴菌

<400> 761
Tyr Ala Thr Ala His Leu Ala Ala Leu Arg Thr Ala Ala Ala Val Leu
1 5 10 15

<210> 762
<211> 8
<212> PRT
<213> 阿維鏈黴菌

<400> 762
Trp Glu Val Leu Pro Glu Ile Ala
1 5

<210> 763
<211> 15
<212> PRT
<213> 阿維鏈黴菌

<400> 763
Leu Phe Ala Ser Gly Ala Gly Arg Arg Ala Arg Ala Glu Ala Gly
1 5 10 15

<210> 764
<211> 13
<212> PRT

<213> 阿維鏈黴菌

<400> 764
Ser Asn Arg Asp Ala Asp Asp Leu Ile Arg Asp Val Ala
1 5 10

<210> 765
<211> 11
<212> PRT
<213> 金黃色葡萄球菌

<400> 765
Ala Leu Ile Val Glu Glu Leu Phe Glu Tyr Ala
1 5 10

<210> 766
<211> 16
<212> PRT
<213> 金黃色葡萄球菌

<400> 766
Pro Ser Leu Thr Val Gln Val Ala Met Ala Gly Ala Met Leu Ile Gly
1 5 10 15

<210> 767
<211> 8
<212> PRT
<213> 金黃色葡萄球菌

<400> 767
Thr Glu Ala Val Lys Gln Ser Asp
1 5

<210> 768
<211> 10
<212> PRT
<213> 金黃色葡萄球菌

<400> 768
His Leu Cys Gln Phe Val Met Ser Gly Gln
1 5 10

<210> 769
<211> 13
<212> PRT
<213> 金黃色葡萄球菌

<400> 769
Ser Glu Lys Leu Leu Glu Ser Leu Glu Asn Phe Trp Asn
1 5 10

<210> 770
<211> 11
<212> PRT
<213> 屎腸球菌（Enterococcus faecium）

<400> 770
Asn Phe Leu Leu Cys Asn Phe Ser Asn Leu Trp
1 5 10

<210> 771
<211> 16
<212> PRT
<213> 屎腸球菌

<400> 771
Leu Glu Leu Leu Ser Gln Leu Gln Lys Asn Thr Leu Gln Leu Ile Arg
1 5 10 15

<210> 772
<211> 10
<212> PRT
<213> 屎腸球菌

<400> 772
Lys Lys Phe Ala Lys Thr Thr Ala Arg Leu

1 5 10

<210> 773
<211> 13
<212> PRT
<213> 屎腸球菌

<400> 773
Lys Val Glu Leu Phe Glu Ala Tyr Lys Asn Ser Leu Leu
1 5 10

<210> 774
<211> 9
<212> PRT
<213> 大腸桿菌

<400> 774
Gly Val Tyr Ala Asn Glu Leu Arg Ala
1 5

<210> 775
<211> 16
<212> PRT
<213> 大腸桿菌

<400> 775
Gly Gly Ile Arg Glu Ile Glu Phe Ile Val Gln Val Phe Gln Leu Ile
1 5 10 15

<210> 776
<211> 8
<212> PRT
<213> 大腸桿菌

<400> 776
Thr Leu Ser Ala Ile Ala Glu Leu
1 5

<210> 777
<211> 15
<212> PRT
<213> 大腸桿菌

<400> 777
Glu Gln Leu Arg Val Ala Tyr Leu Phe Leu Arg Arg Leu Glu Asn
1 5 10 15

<210> 778
<211> 13
<212> PRT
<213> 大腸桿菌

<400> 778
Leu Thr Gly His Met Thr Asn Val Arg Arg Val Phe Asn
1 5 10

<210> 779
<211> 11
<212> PRT
<213> 白蟻塞巴魯德菌（Sebaldella termitidis）

<400> 779
Ser Arg Cys Met Lys Ile Ala Gln Ser Gly Gln
1 5 10

<210> 780
<211> 16
<212> PRT
<213> 白蟻塞巴魯德菌

<400> 780
Ile Ala Glu Ala Glu Phe Ile Asn Glu Ser Ile Tyr Met Ile Tyr Leu
1 5 10 15

<210> 781

<211> 8
<212> PRT
<213> 白蟻塞巴魯德菌

<400> 781
Lys Asp Met Gln Phe Leu Pro Ile
1 5

<210> 782
<211> 14
<212> PRT
<213> 白蟻塞巴魯德菌

<400> 782
Asn Leu Leu Asn Asn Leu Ile Ser Ile Gln Asn Ser Glu Lys
1 5 10

<210> 783
<211> 13
<212> PRT
<213> 白蟻塞巴魯德菌

<400> 783
Ala Glu Lys Ile Cys Gly Leu Ile Ile Asn Glu Leu Lys
1 5 10

<210> 784
<211> 10
<212> PRT
<213> 天藍色鏈黴菌（Streptomyces coelicolor）

<400> 784
Ala Arg Leu Asp Ala Tyr Ala Asn Ser His
1 5 10

<210> 785
<211> 16
<212> PRT
<213> 天藍色鏈黴菌

<400> 785
Leu Asp Ala Ala Asp Ser Ile Gly Phe Leu Leu Glu Leu Leu Phe Ala
1 5 10 15

<210> 786
<211> 8
<212> PRT
<213> 天藍色鏈黴菌

<400> 786
Trp Glu Leu Asp Arg Phe Pro Leu
1 5

<210> 787
<211> 14
<212> PRT
<213> 天藍色鏈黴菌

<400> 787
Glu Leu Leu Ala Thr Leu Gly Arg Ile Thr Gly Ala Gly Gly
1 5 10

<210> 788
<211> 13
<212> PRT
<213> 天藍色鏈黴菌

<400> 788
Gln Arg Glu Leu Phe Gly Arg Val Glu Ala Ala Ala Arg
1 5 10

<210> 789
<211> 11
<212> PRT
<213> 嗜冷黃桿菌（Flavobacterium psychrophilum）

<400> 789
Tyr Ser Ile Tyr Lys Asn Ala Arg Gln Leu Arg
1 5 10

<210> 790
<211> 16
<212> PRT
<213> 嗜冷黄桿菌

<400> 790
Thr Ser Leu Leu Ile Leu Ser Ser Glu Glu Val Ile Lys Ser Ile Leu
1 5 10 15

<210> 791
<211> 8
<212> PRT
<213> 嗜冷黄桿菌

<400> 791
Gln Leu Ile Glu Leu Ser Ile Gly
1 5

<210> 792
<211> 15
<212> PRT
<213> 嗜冷黄桿菌

<400> 792
Lys Leu Thr Glu Phe Asp Asp Lys Lys Asn Gln Gly Phe Tyr Val
1 5 10 15

<210> 793
<211> 13
<212> PRT
<213> 嗜冷黄桿菌

<400> 793
Lys Thr Glu Phe Thr Glu Thr Lys Val Val Val Asp Arg
1 5 10

<210> 794
<211> 10
<212> PRT
<213> 乳酸乳球菌

<400> 794
Lys Cys Ile Asp His Ile Ser Val Leu Ile
1 5 10

<210> 795
<211> 16
<212> PRT
<213> 乳酸乳球菌

<400> 795
Thr Phe Ile Ser Ile Thr Ile Ile Glu Glu Val Gly Lys Thr His Ile
1 5 10 15

<210> 796
<211> 8
<212> PRT
<213> 乳酸乳球菌

<400> 796
Ser Leu Pro Thr Ile Lys Met Gly
1 5

<210> 797
<211> 14
<212> PRT
<213> 乳酸乳球菌

<400> 797
Thr Gly Glu Leu Ile Ser Ile Arg Glu Ser Ser Leu Tyr Ala
1 5 10

<210> 798
<211> 13
<212> PRT
<213> 乳酸乳球菌

<400> 798
Lys Glu Gln Ser Arg Ala Leu Leu Leu Tyr Ala Ile Glu
1 5 10

<210> 799
<211> 11
<212> PRT
<213> 惡臭假單胞菌（Pseudomonas putida）

<400> 799
Asp Ala Leu Leu Thr Asn Ala Ala Ser Leu Ile
1 5 10

<210> 800
<211> 16
<212> PRT
<213> 惡臭假單胞菌

<400> 800
Phe Ala Leu Ala His Leu Ala Arg Glu Glu Ile Ala Lys Thr Leu Met
1 5 10 15

<210> 801
<211> 8
<212> PRT
<213> 惡臭假單胞菌

<400> 801
Thr Ile Asn Ser Ile Val Phe Cys
1 5

<210> 802
<211> 12
<212> PRT
<213> 惡臭假單胞菌

<400> 802
Phe Arg Asn Asp Leu Lys Asn Asn Ser Leu Tyr Val
1 5 10

<210> 803
<211> 13
<212> PRT
<213> 惡臭假單胞菌

<400> 803
Ala Glu Arg Ala Leu Arg Thr Ile Thr Leu Ala Trp Asp
1 5 10

<210> 804
<211> 11
<212> PRT
<213> 生痰新月形單胞菌（Selenomonas sputigena）

<400> 804
Gln Ile Ala Tyr Tyr Leu Tyr Phe Met Tyr Leu
1 5 10

<210> 805
<211> 16
<212> PRT
<213> 生痰新月形單胞菌

<400> 805
Met Thr Ser Phe Ala Tyr Tyr Lys Ser Tyr Phe Asp Arg Val Thr Ala
1 5 10 15

<210> 806
<211> 10
<212> PRT

<213> 生痰新月形單胞菌

<400> 806
Arg Leu Cys Glu Phe Tyr Glu Glu Phe Asp
1 5 10

<210> 807
<211> 17
<212> PRT
<213> 生痰新月形單胞菌

<400> 807
Ile Ile Asp Lys Ala Gln Ala Leu Arg Tyr Ala Asn Pro Leu Thr His
1 5 10 15

Ser

<210> 808
<211> 13
<212> PRT
<213> 生痰新月形單胞菌

<400> 808
Ile Arg Glu Leu Ser Thr Leu Leu Asp Arg Tyr Ile Ala
1 5 10

<210> 809
<211> 11
<212> PRT
<213> 瑞士乳酸桿菌（Lactobacillus helveticus）

<400> 809
Trp Ile Ser Tyr Tyr Leu Tyr Phe Glu Ser Ile
1 5 10

<210> 810
<211> 16
<212> PRT
<213> 瑞士乳酸桿菌

<400> 810
Leu Thr Ser Tyr Ala Phe Phe Lys Asn Tyr Phe Asp Arg Thr Thr Ala
1 5 10 15

<210> 811
<211> 10
<212> PRT
<213> 瑞士乳酸桿菌

<400> 811
Gln Leu Gln Lys Val Tyr Arg Ile Leu Asn
1 5 10

<210> 812
<211> 17
<212> PRT
<213> 瑞士乳酸桿菌

<400> 812
Ile Ile Ser Lys Ala Asn Asp Leu Arg Asn Asn Asn Pro Leu Ser His
1 5 10 15

Ala

<210> 813
<211> 13
<212> PRT
<213> 瑞士乳酸桿菌

<400> 813
Ile Ala Thr Met Arg Ser Leu Phe Lys Leu Leu Val Glu
1 5 10

<210> 814
<211> 11
<212> PRT
<213> 乳酸乳球菌

<400> 814
Lys Ile Leu Asn Phe Ile Tyr Phe Arg Ala Lys
1 5 10

<210> 815
<211> 16
<212> PRT
<213> 乳酸乳球菌

<400> 815
Leu Glu Ser Phe Ala Tyr Tyr Lys Asn Tyr Phe Asp Arg Phe Val Ala
1 5 10 15

<210> 816
<211> 10
<212> PRT
<213> 乳酸乳球菌

<400> 816
Lys Leu Ile Asp Gly Leu Lys Gln Leu Asn
1 5 10

<210> 817
<211> 17
<212> PRT
<213> 乳酸乳球菌

<400> 817
Ile Ile Asn Glu Ala His Lys Ile Arg Asn Ser Asn Pro Val Ser His
1 5 10 15

Ser

<210> 818
<211> 13
<212> PRT
<213> 乳酸乳球菌

<400> 818
Leu Asn Asp Leu Lys Ile Ile Ile Glu Gln Leu Ser Thr
1 5 10

<210> 819
<211> 11
<212> PRT
<213> 假單胞菌屬某種（Pseudomonas sp.）

<400> 819
Lys Trp Leu Phe Ile Asp Gln Met Val Asp Leu
1 5 10

<210> 820
<211> 12
<212> PRT
<213> 假單胞菌屬某種

<400> 820
Phe Lys Phe Arg Glu Ile Arg Ile Glu Tyr Ser Gln
1 5 10

<210> 821
<211> 9
<212> PRT
<213> 假單胞菌屬某種

<400> 821
Tyr Glu Tyr Ala Gln Glu Ile Arg Ser
1 5

<210> 822
<211> 14
<212> PRT
<213> 假單胞菌屬某種

<400> 822
Arg Lys Ile Pro Asp Phe Arg Gly Lys Tyr Ala Ala His Ile
1 5 10

<210> 823
<211> 13
<212> PRT
<213> 假單胞菌屬某種

<400> 823
Lys Ala Leu Glu Phe Tyr Asn Trp Ile His Ser Asn Glu
1 5 10

<210> 824
<211> 11
<212> PRT
<213> 副霍亂弧菌（Vibrio paracholerae）

<400> 824
Glu Glu Ile Leu Ser Gly Leu Ile Gly Asp Leu
1 5 10

<210> 825
<211> 12
<212> PRT
<213> 副霍亂弧菌

<400> 825
Arg Lys Tyr Val Glu Leu Asn Gln Lys Tyr Gly Lys
1 5 10

<210> 826
<211> 10
<212> PRT
<213> 副霍亂弧菌

<400> 826
Gly Val Tyr Asn Asn Glu Ile Asn Lys Asn
1 5 10

<210> 827
<211> 14
<212> PRT
<213> 副霍亂弧菌

<400> 827
Thr Ala Ile Lys Lys Leu Arg Asn His Cys Val Ala His Val
1 5 10

<210> 828
<211> 13
<212> PRT
<213> 副霍亂弧菌

<400> 828
Phe Ala Asp Glu Phe Leu Asp Trp Ile Cys Pro Asp Asn
1 5 10

<210> 829
<211> 11
<212> PRT
<213> 大腸桿菌

<400> 829
Thr Met Ala Asp His Met Val Asn Glu Ala Trp
1 5 10

<210> 830
<211> 16
<212> PRT

<213> 大腸桿菌

<400> 830
Phe Asn Leu Ile Leu Gln Ser Ile Glu Phe Arg Leu Lys Gly Leu Ile
1 5 10 15

<210> 831
<211> 10
<212> PRT
<213> 大腸桿菌

<400> 831
Lys Val Tyr Asn Thr Phe Ala Ser Lys Ser
1 5 10

<210> 832
<211> 14
<212> PRT
<213> 大腸桿菌

<400> 832
Trp Phe Asn Ser Met Arg Ile Leu Arg Asn Arg Phe Met His
1 5 10

<210> 833
<211> 13
<212> PRT
<213> 大腸桿菌

<400> 833
Asp Ile Met Pro Glu Leu Ile Phe Thr Ser Val Val Arg
1 5 10

<210> 834
<211> 11
<212> PRT
<213> 硫還原地桿菌（Geobacter sulfurreducens）

<400> 834
Leu Asn Tyr Glu Ala Leu Tyr Val Lys Ser Lys
1 5 10

<210> 835
<211> 16
<212> PRT
<213> 硫還原地桿菌

<400> 835
Gln Leu Trp Ala Ser Met Ala Leu Glu Leu Leu Ala Lys Ser Ser Leu
1 5 10 15

<210> 836
<211> 9
<212> PRT
<213> 硫還原地桿菌

<400> 836
Gln Arg Leu Gly His Ile Ser Lys Leu
1 5

<210> 837
<211> 15
<212> PRT
<213> 硫還原地桿菌

<400> 837
Phe Cys Glu Gln Leu Ser Leu Arg Arg Asn Ser Glu Ile His Ser
1 5 10 15

<210> 838
<211> 13
<212> PRT
<213> 硫還原地桿菌

<400> 838
Asp Ala Trp Glu Val Lys Tyr Trp Tyr Ala Ile Glu Val

1 5 10

<210> 839
<211> 11
<212> PRT
<213> 天藍色鏈黴菌

<400> 839
Asp Val Ser Tyr Thr Pro Val Ser Asn Gly Met
1 5 10

<210> 840
<211> 16
<212> PRT
<213> 天藍色鏈黴菌

<400> 840
Val Leu His Leu Gln Ala Ala Thr Glu Val Leu Leu Lys Ala Arg Leu
1 5 10 15

<210> 841
<211> 10
<212> PRT
<213> 天藍色鏈黴菌

<400> 841
Asp Arg Leu Arg Asp Ile Ala Arg Leu Asp
1 5 10

<210> 842
<211> 15
<212> PRT
<213> 天藍色鏈黴菌

<400> 842
Arg Ile Lys Glu Pro Gly Glu Ser Arg Asn Ala Leu Gln His Tyr
1 5 10 15

<210> 843
<211> 13
<212> PRT
<213> 天藍色鏈黴菌

<400> 843
Tyr Ala Ile Glu Ser Arg Ala Ala Arg Val Leu Asp Phe
1 5 10

<210> 844
<211> 10
<212> PRT
<213> 問號鉤端螺旋體（Leptospira interrogans）

<400> 844
Cys Thr Arg Leu Tyr Asn Gln Ile Leu Glu
1 5 10

<210> 845
<211> 16
<212> PRT
<213> 問號鉤端螺旋體

<400> 845
Tyr Thr Lys Leu Phe Asn Ile Leu Asp Lys Val Ala Ala Ile Val Tyr
1 5 10 15

<210> 846
<211> 6
<212> PRT
<213> 問號鉤端螺旋體

<400> 846
Phe Pro Ser Thr Phe Gly
1 5

<210> 847

<211> 13
<212> PRT
<213> 問號鉤端螺旋體

<400> 847
His His Leu Arg Val Arg Arg Asn Asn Ile Val His Trp
1 5 10

<210> 848
<211> 13
<212> PRT
<213> 問號鉤端螺旋體

<400> 848
Glu Glu Asp Val Gln Arg Leu Phe Leu Ile Ser Lys Ala
1 5 10

<210> 849
<211> 10
<212> PRT
<213> 鮑氏志賀氏菌（Shigella boydii）

<400> 849
Met Glu Met Val Leu Asn Arg Leu Lys Ser
1 5 10

<210> 850
<211> 16
<212> PRT
<213> 鮑氏志賀氏菌

<400> 850
Phe Arg Leu Cys Phe Gly Ile Leu Asp Lys Ile Ala Val Ala Ile Cys
1 5 10 15

<210> 851
<211> 8
<212> PRT
<213> 鮑氏志賀氏菌

<400> 851
Pro Gln Lys Asn Ile Tyr Phe Gln
1 5

<210> 852
<211> 15
<212> PRT
<213> 鮑氏志賀氏菌

<400> 852
Glu Leu Ala Phe Tyr Lys Glu Trp Arg Asn Gly Leu Glu His Lys
1 5 10 15

<210> 853
<211> 13
<212> PRT
<213> 鮑氏志賀氏菌

<400> 853
Ile His His Phe Glu His Leu Leu Gln Ile Thr Arg Ser
1 5 10

<210> 854
<211> 10
<212> PRT
<213> 糞腸球菌（Enterococcus faecalis）

<400> 854
Phe Tyr Ser Leu Phe Asn Gln Ile Lys Gln
1 5 10

<210> 855
<211> 16
<212> PRT
<213> 糞腸球菌

<400> 855
Tyr Arg Ser Val Tyr Ser Ile Phe Asp Lys Ile Ala Tyr Phe Leu Asn
1 5 10 15

<210> 856
<211> 8
<212> PRT
<213> 糞腸球菌

<400> 856
Pro Lys Asn Leu Ile Thr Phe His
1 5

<210> 857
<211> 15
<212> PRT
<213> 糞腸球菌

<400> 857
Asn Leu Glu Lys Ile Ala Glu Ile Arg Asn Ala Met Glu His Lys
1 5 10 15

<210> 858
<211> 13
<212> PRT
<213> 糞腸球菌

<400> 858
Glu Lys Ile Thr Leu Glu Leu Phe Lys Leu Thr Arg Glu
1 5 10

<210> 859
<211> 10
<212> PRT
<213> 熱纖維瘤胃梭菌（Ruminiclostridium thermocellum）

<400> 859
Phe Asn Asn Arg Ala Phe Asp Leu Ile Val
1 5 10

<210> 860
<211> 16
<212> PRT
<213> 熱纖維瘤胃梭菌

<400> 860
Tyr Thr Arg Phe Glu Gly Leu Ile Asp Thr Ile Tyr His Ile Ile Asn
1 5 10 15

<210> 861
<211> 7
<212> PRT
<213> 熱纖維瘤胃梭菌

<400> 861
Lys Pro Ser Ser Glu Phe Arg
1 5

<210> 862
<211> 15
<212> PRT
<213> 熱纖維瘤胃梭菌

<400> 862
Val Tyr Lys Lys Ile Asn Lys Phe Arg Asn Asn Ile Val His Asn
1 5 10 15

<210> 863
<211> 13
<212> PRT
<213> 熱纖維瘤胃梭菌

<400> 863
Tyr Thr Thr Ser Thr Glu Phe Leu Asn Asn Ile Lys Asp
1 5 10

<210> 864
<211> 10
<212> PRT
<213> 蠟樣芽胞桿菌（Bacillus cereus）

<400> 864
Leu Asn Asn Arg Ile Phe Gln Leu Asp Leu
1 5 10

<210> 865
<211> 16
<212> PRT
<213> 蠟樣芽胞桿菌

<400> 865
Phe Pro Lys Ala Phe Thr Ala Leu Asp Leu Leu Ala His Leu Leu Phe
1 5 10 15

<210> 866
<211> 7
<212> PRT
<213> 蠟樣芽胞桿菌

<400> 866
Lys Thr Glu Lys Lys Ile Lys
1 5

<210> 867
<211> 15
<212> PRT
<213> 蠟樣芽胞桿菌

<400> 867
Glu Phe Gln Lys Ala Ser Lys Val Arg Asn Asp Ile Ile His Asn
1 5 10 15

<210> 868
<211> 13
<212> PRT
<213> 蠟樣芽胞桿菌

<400> 868
Tyr Thr Pro Ser Lys Glu Ile Leu Asn Ile Ala Arg Gly
1 5 10

<210> 869
<211> 10
<212> PRT
<213> 丁香假單胞菌

<400> 869
Glu Tyr Leu Arg Cys Lys Asp Ala Phe Glu
1 5 10

<210> 870
<211> 16
<212> PRT
<213> 丁香假單胞菌

<400> 870
Ser Ser Phe Ile His His Leu Tyr Glu Leu Tyr Met Ala Leu Phe Ala
1 5 10 15

<210> 871
<211> 8
<212> PRT
<213> 丁香假單胞菌

<400> 871
Ser Ile Asp Arg Gly Ala Val Ser
1 5

<210> 872
<211> 16
<212> PRT

<213> 丁香假單胞菌

<400> 872
Phe Gly Pro Ala Phe Arg Ser Met Arg Asn Lys Ile Ala Gly His Val
1 5 10 15

<210> 873
<211> 13
<212> PRT
<213> 丁香假單胞菌

<400> 873
Val Lys Leu Thr Glu Phe Phe Gln Lys Tyr His Pro Tyr
1 5 10

<210> 874
<211> 10
<212> PRT
<213> 食異生素伯克氏菌（Burkholderia xenovorans）

<400> 874
Glu Tyr Leu Arg Cys Asp Asp Ala Leu His
1 5 10

<210> 875
<211> 16
<212> PRT
<213> 食異生素伯克氏菌

<400> 875
Ala Arg Phe Ile His His Leu Tyr Glu Phe Asn Ile Ala Cys Ala Gln
1 5 10 15

<210> 876
<211> 8
<212> PRT
<213> 食異生素伯克氏菌

<400> 876
Arg Val Arg Arg Gln Ala Tyr Asn
1 5

<210> 877
<211> 16
<212> PRT
<213> 食異生素伯克氏菌

<400> 877
Phe Ala Lys Ala Phe Arg Thr Ala Arg Asn Thr Thr Asn Gly His Ala
1 5 10 15

<210> 878
<211> 13
<212> PRT
<213> 食異生素伯克氏菌

<400> 878
Leu Asn Leu Ser Asp Phe Phe Thr Arg Tyr His Arg Phe
1 5 10

<210> 879
<211> 10
<212> PRT
<213> 銅綠微囊藻

<400> 879
Glu His Leu Asp Cys Glu Leu Trp Glu Arg
1 5 10

<210> 880
<211> 16
<212> PRT
<213> 銅綠微囊藻

<400> 880
Ile Arg Asn Ala Thr Val Ile Leu Glu Asp Arg Met Arg Lys Leu Gly

1 5 10 15

<210> 881
<211> 8
<212> PRT
<213> 銅綠微囊藻

<400> 881
Gly Ile Val Asn Leu Ile Phe Gly
1 5

<210> 882
<211> 15
<212> PRT
<213> 銅綠微囊藻

<400> 882
Tyr Ser Gly Thr Met Lys Ile Phe Arg Asn Arg Tyr Ala His Arg
1 5 10 15

<210> 883
<211> 13
<212> PRT
<213> 銅綠微囊藻

<400> 883
Ile Ile Val Phe Ile Asp Leu Leu Leu Lys Met Leu Asp
1 5 10

<210> 884
<211> 11
<212> PRT
<213> 副溶血性弧菌（Vibrio parahaemolyticus）

<400> 884
Ser Arg Asn Val His Pro Asp Val Leu Lys Tyr
1 5 10

<210> 885
<211> 16
<212> PRT
<213> 副溶血性弧菌

<400> 885
Val Phe Glu Ala Thr Lys Ser Val Ala Asp Lys Ile Arg Asn Lys Thr
1 5 10 15

<210> 886
<211> 8
<212> PRT
<213> 副溶血性弧菌

<400> 886
Val Leu Val Asp Glu Ala Phe Ser
1 5

<210> 887
<211> 15
<212> PRT
<213> 副溶血性弧菌

<400> 887
Leu Lys Gly Leu Phe Gly Thr Phe Arg Asn Thr Thr Ala His Ala
1 5 10 15

<210> 888
<211> 13
<212> PRT
<213> 副溶血性弧菌

<400> 888
Ile Leu Ser Met Val Ser Leu Val His Arg Arg Leu Asp
1 5 10

<210> 889

<211> 11
<212> PRT
<213> 乳酸乳球菌

<400> 889
Ala Leu Glu Leu His Ser Glu Val Thr Lys Tyr
1 5 10

<210> 890
<211> 16
<212> PRT
<213> 乳酸乳球菌

<400> 890
Val Phe Glu Ser Cys Lys Gly Leu Phe Asp Arg Ile Arg Leu Ile Ser
1 5 10 15

<210> 891
<211> 8
<212> PRT
<213> 乳酸乳球菌

<400> 891
Thr Leu Ile Asn Gln Ala Phe Asn
1 5

<210> 892
<211> 15
<212> PRT
<213> 乳酸乳球菌

<400> 892
Ile Lys Thr Cys Leu Tyr Leu Tyr Arg Asn His Gln Ala His Val
1 5 10 15

<210> 893
<211> 13
<212> PRT
<213> 乳酸乳球菌

<400> 893
Gly Leu Met Ser Ile Ser Leu Ala His Glu Leu Leu Asp
1 5 10

<210> 894
<211> 11
<212> PRT
<213> 星狀海葵（Nematostella vectensis）

<400> 894
Ser Thr Thr Leu Thr Thr Phe Leu Asn Leu His
1 5 10

<210> 895
<211> 16
<212> PRT
<213> 星狀海葵

<400> 895
Glu Asp Tyr Asp Ile Thr Leu Leu Thr Cys Leu Leu Arg Asn Ile Cys
1 5 10 15

<210> 896
<211> 8
<212> PRT
<213> 星狀海葵

<400> 896
Asp Lys Leu Pro Pro Ala Tyr Asp
1 5

<210> 897
<211> 15
<212> PRT
<213> 星狀海葵

<400> 897
Val Val Arg Leu Arg His Tyr Arg Asn Asp Leu Tyr Ala His Ile
1 5 10 15

<210> 898
<211> 13
<212> PRT
<213> 星狀海葵

<400> 898
Trp Ala Asp Ile Ser Ala Ala Leu Leu Ser Leu Gly Gly
1 5 10

<210> 899
<211> 11
<212> PRT
<213> 佛羅里達文昌魚

<400> 899
Pro Pro Ser Leu Pro Ala Gln Leu Lys Lys His
1 5 10

<210> 900
<211> 16
<212> PRT
<213> 佛羅里達文昌魚

<400> 900
Glu Glu Phe Asp Ile Ser Leu Leu Leu Leu Leu Leu Lys Glu Leu Val
1 5 10 15

<210> 901
<211> 8
<212> PRT
<213> 佛羅里達文昌魚

<400> 901
Gly Arg Asp Ala Pro Tyr Ser Asp
1 5

<210> 902
<211> 13
<212> PRT
<213> 佛羅里達文昌魚

<400> 902
Lys Leu Gly Gln Phe Arg Asn Lys Asn Tyr Gly His Ile
1 5 10

<210> 903
<211> 13
<212> PRT
<213> 佛羅里達文昌魚

<400> 903
Trp Asp Glu Leu Thr Glu Ile Leu Val Asp Leu Gly Gly
1 5 10

<210> 904
<211> 11
<212> PRT
<213> 智人

<400> 904
Pro Pro Leu Leu Lys Lys Glu Leu Leu Ile His
1 5 10

<210> 905
<211> 16
<212> PRT
<213> 智人

<400> 905
Lys Gln Phe Asp Leu Cys Leu Leu Leu Ala Leu Ile Lys His Leu Asn
1 5 10 15

<210> 906
<211> 8
<212> PRT
<213> 智人

<400> 906
Asn Met Glu Pro Pro Ser Ser Asp
1 5

<210> 907
<211> 15
<212> PRT
<213> 智人

<400> 907
Ile Leu Arg Leu Cys Lys Tyr Arg Asp Ile Leu Leu Ser Glu Ile
1 5 10 15

<210> 908
<211> 13
<212> PRT
<213> 智人

<400> 908
Trp Lys Lys Val Ser Asp Ile Leu Leu Arg Leu Gly Met
1 5 10

<210> 909
<211> 11
<212> PRT
<213> 大腸桿菌

<400> 909
Val Thr Ala Glu Lys Leu Leu Val Ser Gly Leu
1 5 10

<210> 910
<211> 16
<212> PRT
<213> 大腸桿菌

<400> 910
Leu Tyr Pro Glu Leu Arg Thr Ile Glu Gly Val Leu Lys Ser Lys Met
1 5 10 15

<210> 911
<211> 8
<212> PRT
<213> 大腸桿菌

<400> 911
Tyr Ile Leu Lys Pro Gln Phe Ala
1 5

<210> 912
<211> 14
<212> PRT
<213> 大腸桿菌

<400> 912
Ala Tyr Thr Phe Phe Asn Val Glu Arg His Ser Leu Phe His
1 5 10

<210> 913
<211> 13
<212> PRT
<213> 大腸桿菌

<400> 913
Met Ile Ser Asp Met Ala Arg Leu Met Gly Lys Ala Thr
1 5 10

<210> 914
<211> 11
<212> PRT

<213> 深海發光桿菌（Photobacterium profundum）

<400> 914
Asp Thr Tyr Arg Ser Leu Leu Ser Ser Ser Tyr
1 5 10

<210> 915
<211> 16
<212> PRT
<213> 深海發光桿菌

<400> 915
Ile Tyr Pro Asp Leu Arg Val Leu Glu Gly Val Ile Lys Glu Ala Met
1 5 10 15

<210> 916
<211> 8
<212> PRT
<213> 深海發光桿菌

<400> 916
Thr Glu Leu Lys Thr Glu Tyr Asn
1 5

<210> 917
<211> 14
<212> PRT
<213> 深海發光桿菌

<400> 917
Cys Tyr Ala Tyr Phe Lys Ala His Arg His Ser Leu Phe His
1 5 10

<210> 918
<211> 13
<212> PRT
<213> 深海發光桿菌

<400> 918
Thr Thr Asp Thr Ile Gly Glu Val Met Gln Met Ser Glu
1 5 10

<210> 919
<211> 11
<212> PRT
<213> 熱葡糖苷酶地芽胞桿菌（Geobacillus thermoglucosidasius）

<400> 919
Leu Tyr Asp Arg Asp Arg Ile Glu Ala Ser Glu
1 5 10

<210> 920
<211> 16
<212> PRT
<213> 熱葡糖苷酶地芽胞桿菌

<400> 920
Val Ser Gly Thr Leu Arg Ala Phe Glu Gly Phe Phe Lys Lys Leu Leu
1 5 10 15

<210> 921
<211> 8
<212> PRT
<213> 熱葡糖苷酶地芽胞桿菌

<400> 921
Asp Ile Ser Glu Lys Val Phe Asn
1 5

<210> 922
<211> 14
<212> PRT
<213> 熱葡糖苷酶地芽胞桿菌

<400> 922
Met Leu Asn His Met Ser Gln Asp Arg Asn Pro Tyr Ser His

1 5 10

<210> 923
<211> 13
<212> PRT
<213> 熱葡糖苷酶地芽胞桿菌

<400> 923
Pro Leu Arg Thr Leu Asn Gln Ala Ile Ser Leu His Asn
1 5 10

<210> 924
<211> 12
<212> PRT
<213> 船蛆桿菌（Teredinibacter turnerae）

<400> 924
Cys Arg Ser Ile Arg Lys Leu Leu Asn Met Asn Ala
1 5 10

<210> 925
<211> 16
<212> PRT
<213> 船蛆桿菌

<400> 925
Ser Tyr Pro Leu Ile Tyr Glu Ile Glu Asn Leu Val Arg Lys Leu Ile
1 5 10 15

<210> 926
<211> 9
<212> PRT
<213> 船蛆桿菌

<400> 926
Ile Gln Leu Ser Asn Phe Leu Phe Asp
1 5

<210> 927
<211> 15
<212> PRT
<213> 船蛆桿菌

<400> 927
Arg Trp Gly Lys Leu Tyr Lys Leu Arg Cys Lys Ile Ala His Asn
1 5 10 15

<210> 928
<211> 13
<212> PRT
<213> 船蛆桿菌

<400> 928
Thr Thr Lys Leu Val Glu Glu Val Lys Leu Lys Ile Leu
1 5 10

<210> 929
<211> 5
<212> PRT
<213> 海沼甲燒球菌甲烷（Methanococcus maripaludis）

<400> 929
Phe Arg Leu Met Tyr
1 5

<210> 930
<211> 16
<212> PRT
<213> 海沼甲燒球菌甲烷

<400> 930
Phe Leu Asp Ser Val Leu Ala Leu Glu Ile Tyr His Thr Leu Lys Phe
1 5 10 15

<210> 931

<211> 10
<212> PRT
<213> 海沼甲燒球菌甲烷

<400> 931
Phe Ile Asn Lys Met Lys Asp Val Phe Asn
1 5 10

<210> 932
<211> 15
<212> PRT
<213> 海沼甲燒球菌甲烷

<400> 932
Ile Cys Arg Ile Ile Arg Asp Thr Arg Asn Lys Leu Val His Asp
1 5 10 15

<210> 933
<211> 13
<212> PRT
<213> 海沼甲燒球菌甲烷

<400> 933
Pro Tyr Phe Leu Ile Glu Leu Leu Lys Asn Ile Phe Lys
1 5 10

<210> 934
<211> 11
<212> PRT
<213> 食五環芳烴新鞘氨醇菌（Novosphingobium pentaromativorans）

<400> 934
Val His Arg Ala Leu Ser Trp Leu Arg Arg Ala
1 5 10

<210> 935
<211> 14
<212> PRT
<213> 食五環芳烴新鞘氨醇菌

<400> 935
Phe Ile Leu Leu Trp Ile Gly Phe Asn Ala Ala Tyr Ala Gly
1 5 10

<210> 936
<211> 10
<212> PRT
<213> 食五環芳烴新鞘氨醇菌

<400> 936
Glu Arg Ser Arg Thr Ala Ile Asn Tyr Ala
1 5 10

<210> 937
<211> 15
<212> PRT
<213> 食五環芳烴新鞘氨醇菌

<400> 937
Leu Phe Asp Arg Leu Tyr Val Leu Arg Asn Gln Leu Val His Gly
1 5 10 15

<210> 938
<211> 13
<212> PRT
<213> 食五環芳烴新鞘氨醇菌

<400> 938
Arg Asp Gln Val Arg Asp Gly Ala Ser Leu Leu Gly Cys
1 5 10

<210> 939
<211> 12
<212> PRT
<213> 氯鉻綠菌（Chlorobium chlorochromatii）

<400> 939
Ile Met Glu Gln Arg Lys Ala Ile Leu Glu Pro Leu
1 5 10

<210> 940
<211> 16
<212> PRT
<213> 氯鉻綠菌

<400> 940
Ala Val Ala Tyr Asn His Phe Val Pro Leu Leu Ala Gln Asp Leu Ile
1 5 10 15

<210> 941
<211> 6
<212> PRT
<213> 氯鉻綠菌

<400> 941
Lys Ile Ser Asn Lys Lys
1 5

<210> 942
<211> 15
<212> PRT
<213> 氯鉻綠菌

<400> 942
Ser Glu Lys Leu Lys Thr Phe Arg Asp Lys Tyr Tyr Ala His Leu
1 5 10 15

<210> 943
<211> 13
<212> PRT
<213> 氯鉻綠菌

<400> 943
Phe Leu Gly Ile His Arg Lys Ser Ala Asn Glu Met Trp
1 5 10

<210> 944
<211> 12
<212> PRT
<213> 乳酸乳球菌

<400> 944
Asp Ala Tyr Asn Lys Leu Ile Leu Leu Lys Gln Tyr
1 5 10

<210> 945
<211> 15
<212> PRT
<213> 乳酸乳球菌

<400> 945
Phe Phe Tyr Asn Asn Leu Leu Asp Ser Leu Val Ile Ala Ile Phe
1 5 10 15

<210> 946
<211> 6
<212> PRT
<213> 乳酸乳球菌

<400> 946
Asn Tyr Thr Asn Phe Pro
1 5

<210> 947
<211> 15
<212> PRT
<213> 乳酸乳球菌

<400> 947
Leu Glu Tyr Leu Tyr Ala Gln Arg Asn Lys Ile Tyr Val His Asn
1 5 10 15

<210> 948
<211> 13
<212> PRT
<213> 乳酸乳球菌

<400> 948
Asn Tyr Ala Trp Glu Pro Thr Asn Ile Asn Asp Trp Glu
1 5 10

<210> 949
<211> 16
<212> PRT
<213> 大腸桿菌

<400> 949
Glu Ser Val Ile Ala His Met Asn Glu Leu Leu Ile Ala Leu Ser Asp
1 5 10 15

<210> 950
<211> 15
<212> PRT
<213> 大腸桿菌

<400> 950
Arg Tyr Thr Gln Gln Gln Arg Leu Arg Thr Ala Ile Ala His His
1 5 10 15

<210> 951
<211> 12
<212> PRT
<213> 大腸桿菌

<400> 951
Glu Ala Arg His Glu Gln Leu Thr Lys Gly Gly Thr
1 5 10

<210> 952
<211> 16
<212> PRT
<213> 阪崎腸桿菌（Cronobacter sakazakii）

<400> 952
Gln His Val Ile Ala Pro Met Asn Glu Leu Leu Ile Ala Leu Ser Asp
1 5 10 15

<210> 953
<211> 15
<212> PRT
<213> 阪崎腸桿菌

<400> 953
Arg Tyr Asp Leu Gln Gln Gln Leu Arg Thr Ala Ile Ala His His
1 5 10 15

<210> 954
<211> 13
<212> PRT
<213> 阪崎腸桿菌

<400> 954
Ala Ala Glu Arg Leu Ala Glu Leu Thr Arg Gly Gly Thr
1 5 10

<210> 955
<211> 16
<212> PRT
<213> 智人

<400> 955
Glu Ser Arg Tyr Arg Thr Leu Arg Asn Val Gly Asn Glu Ser Asp Ile
1 5 10 15

<210> 956
<211> 10
<212> PRT

<213> 智人

<400> 956
Leu Gln Pro Gly Pro Ser Glu His Ser Lys
1 5 10

<210> 957
<211> 14
<212> PRT
<213> 智人

<400> 957
Val Gly Asp Leu Leu Lys Phe Ile Arg Asn Leu Gly Glu His
1 5 10

<210> 958
<211> 13
<212> PRT
<213> 智人

<400> 958
Ile Gly Asp Pro Ser Leu Tyr Phe Gln Lys Thr Phe Pro
1 5 10

<210> 959
<211> 16
<212> PRT
<213> 阿拉伯芥

<400> 959
Glu Met Arg Leu Ser Phe Leu Arg Asp Ala Ser Asp Arg Val Glu Leu
1 5 10 15

<210> 960
<211> 10
<212> PRT
<213> 阿拉伯芥

<400> 960
Met Glu Ser Thr Ala Pro Val Ala Ile Gly
1 5 10

<210> 961
<211> 15
<212> PRT
<213> 阿拉伯芥

<400> 961
Ile Arg Asp Leu Leu Arg Val Ile Arg Asn Lys Leu Asn His His
1 5 10 15

<210> 962
<211> 13
<212> PRT
<213> 阿拉伯芥

<400> 962
Pro Glu Gly Phe Asp Glu Tyr Phe Ala Val Arg Phe Pro
1 5 10

<210> 963
<211> 12
<212> PRT
<213> 幽門螺桿菌

<400> 963
Tyr Glu Leu Leu Trp Gln Glu Val Ile Arg Ala Lys
1 5 10

<210> 964
<211> 15
<212> PRT
<213> 幽門螺桿菌

<400> 964
Trp Val Ser Leu Gln Asn Val Met Arg Arg Ile Ile Glu Tyr Tyr

1 5 10 15

<210> 965
<211> 5
<212> PRT
<213> 幽門螺桿菌

<400> 965
Phe Arg Ile Leu Gly
1 5

<210> 966
<211> 17
<212> PRT
<213> 幽門螺桿菌

<400> 966
Lys Gln Val Phe Ser Ser Phe Ile Ser Trp Phe Asn Asp Gly Ser His
1 5 10 15

Gly

<210> 967
<211> 13
<212> PRT
<213> 幽門螺桿菌

<400> 967
Ile Glu Thr Tyr Leu Lys Val Phe Glu Asn Ile Phe Lys
1 5 10

<210> 968
<211> 12
<212> PRT
<213> 變異鏈球菌（Streptococcus mutans）

<400> 968
His Leu Met Leu Val Asp Glu Leu Lys Lys Ala Ile
1 5 10

<210> 969
<211> 15
<212> PRT
<213> 變異鏈球菌

<400> 969
Glu Lys Tyr His Phe Asn Leu Leu Arg Asn Leu Leu Glu Lys Thr
1 5 10 15

<210> 970
<211> 5
<212> PRT
<213> 變異鏈球菌

<400> 970
Ala Thr Phe Leu Gly
1 5

<210> 971
<211> 14
<212> PRT
<213> 變異鏈球菌

<400> 971
Pro Ala Pro Tyr Ile Arg Arg Ile Asn Leu His Ser His Ser
1 5 10

<210> 972
<211> 13
<212> PRT
<213> 變異鏈球菌

<400> 972
Lys Lys Val Leu Glu Arg Val Phe Asn Gln Phe Leu Gln

1 5 10

<210> 973
<211> 12
<212> PRT
<213> 大腸桿菌

<400> 973
His Leu His Leu Lys Gln Thr Ile Glu Gln Ala Ile
1 5 10

<210> 974
<211> 15
<212> PRT
<213> 大腸桿菌

<400> 974
Glu Arg Tyr His Phe Thr Leu Leu Arg Asn Leu Tyr Glu Lys Thr
1 5 10 15

<210> 975
<211> 5
<212> PRT
<213> 大腸桿菌

<400> 975
Ala Ser Phe Leu Gly
1 5

<210> 976
<211> 13
<212> PRT
<213> 大腸桿菌

<400> 976
Leu Tyr Leu Ser Arg Ile Ile Asn Phe Thr Ser His Ser
1 5 10

<210> 977
<211> 13
<212> PRT
<213> 大腸桿菌

<400> 977
Lys Ala Thr Val Lys Leu Leu Leu Asp His Leu Lys Asn
1 5 10

<210> 978
<211> 12
<212> PRT
<213> 脆弱擬桿菌（Bacteroides fragilis）

<400> 978
Lys Glu Ile Glu Glu Glu Arg Thr Val Gln Asn Ile
1 5 10

<210> 979
<211> 16
<212> PRT
<213> 脆弱擬桿菌

<400> 979
Thr Ser Phe Gly Glu Val Thr Glu Glu Tyr His Asp Glu Leu Tyr Ser
1 5 10 15

<210> 980
<211> 8
<212> PRT
<213> 脆弱擬桿菌

<400> 980
Tyr Ile Lys Glu Leu Ser Asn Gly
1 5

<210> 981

<211> 15
<212> PRT
<213> 脆弱擬桿菌

<400> 981
Gln Lys Thr Leu Thr Glu Lys Ile Arg His Gln Ile His His Pro
1 5 10 15

<210> 982
<211> 13
<212> PRT
<213> 脆弱擬桿菌

<400> 982
Glu Thr Glu Ile Arg Gln Ser Ile Glu Asp Met Arg Ala
1 5 10

<210> 983
<211> 12
<212> PRT
<213> 鞭毛甲基小桿菌（Methylobacillus flagellatus）

<400> 983
Ser Asn Gln Ile Pro Thr Arg Val Ser Pro Val Leu
1 5 10

<210> 984
<211> 16
<212> PRT
<213> 鞭毛甲基小桿菌

<400> 984
Ser Ala Phe Gly Glu Ala Ser Tyr Glu Tyr His Asn Glu Leu Tyr Gly
1 5 10 15

<210> 985
<211> 8
<212> PRT
<213> 鞭毛甲基小桿菌

<400> 985
Tyr Asn Arg Leu Arg Arg Asp Gly
1 5

<210> 986
<211> 15
<212> PRT
<213> 鞭毛甲基小桿菌

<400> 986
Gln Val Ile Leu Thr Glu Tyr Ile Arg His Gln Ile His His Pro
1 5 10 15

<210> 987
<211> 13
<212> PRT
<213> 鞭毛甲基小桿菌

<400> 987
Thr Ala Glu Leu Thr Glu Ser Ile Glu Thr Met Arg Leu
1 5 10

<210> 988
<211> 12
<212> PRT
<213> 人彎曲桿菌（Campylobacter hominis）

<400> 988
Lys Asp Gly Glu Gln Lys Lys Glu Val Lys Asn Val
1 5 10

<210> 989
<211> 16
<212> PRT
<213> 人彎曲桿菌

<400> 989
Met Ala Phe Gly Glu Ile Thr Glu Glu Tyr His Asn Glu Leu Tyr Gly
1 5 10 15

<210> 990
<211> 8
<212> PRT
<213> 人彎曲桿菌

<400> 990
Tyr Lys Lys Leu Lys Lys Asp Gly
1 5

<210> 991
<211> 15
<212> PRT
<213> 人彎曲桿菌

<400> 991
Lys Leu Thr Leu Thr Glu Tyr Ile Arg His Gln Ile His His Pro
1 5 10 15

<210> 992
<211> 13
<212> PRT
<213> 人彎曲桿菌

<400> 992
Leu Ser Glu Leu Lys Asp Ser Ile Glu Met Met Arg Asn
1 5 10

<210> 993
<211> 15
<212> PRT
<213> 智人

<400> 993
Ile Pro Asp Trp Ile Val Asp Leu Arg His Glu Leu Thr His Lys
1 5 10 15

<210> 994
<211> 15
<212> PRT
<213> 結核分枝桿菌

<400> 994
Leu Gly Arg Phe Glu Ser Arg Val Arg Asn Thr Ala Ala His Glu
1 5 10 15

<210> 995
<211> 15
<212> PRT
<213> 乳酸乳球菌

<400> 995
Trp Ile Arg Ala Gly Trp Phe Ile Arg Asn Arg Ser Ala His Tyr
1 5 10 15

<210> 996
<211> 15
<212> PRT
<213> 嗜熱棲熱菌

<220>
<221> MOD_RES
<222> (5)..(5)
<223>任何胺基酸

<400> 996
Leu Ala Leu Gly Xaa Val Asp Asp Arg Ser Leu Thr Val His Thr
1 5 10 15

<210> 997
<211> 22
<212> PRT

<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 997
Leu Lys Ser Met Leu Tyr Ser Met Arg Asn Ser Ser Phe His Phe Ser
1 5 10 15

Thr Glu Asn Val Asp Asn
20

<210> 998
<211> 22
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 998
Leu Lys Asp Val Ile Tyr Ser Met Arg Asn Asp Ser Phe His Tyr Ala
1 5 10 15

Thr Glu Asn His Asn Asn
20

<210> 999
<211> 22
<212> PRT
<213> 丁酸梭菌（Clostridium aminophilum）

<400> 999
Leu Arg Lys Ala Ile Tyr Ser Leu Arg Asn Glu Thr Phe His Phe Thr
1 5 10 15

Thr Leu Asn Lys Gly Ser
20

<210> 1000
<211> 22
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 1000
Ile Ile Gln Ile Ile Tyr Ser Leu Arg Asn Lys Ser Phe His Phe Lys
1 5 10 15

Thr Tyr Asp His Gly Asp
20

<210> 1001
<211> 22
<212> PRT
<213> 雞肉桿菌（Carnobacterium gallinarum）

<400> 1001
Leu Arg Gly Ser Val Gln Gln Ile Arg Asn Glu Ile Phe His Ser Phe
1 5 10 15

Asp Lys Asn Gln Lys Phe
20

<210> 1002
<211> 21
<212> PRT
<213> 雞肉桿菌

<400> 1002
Ile Arg Gly Ala Val Gln Arg Val Arg Asn Gln Ile Phe His Gln Gln
1 5 10 15

Ile Asn Lys Arg His
20

<210> 1003
<211> 20
<212> PRT
<213> propionicigenes帕魯迪菌（Paludibacter propionicigenes）

<400> 1003
Ile Arg Gly Ala Val Gln Gln Ile Arg Asn Asn Val Asn His Tyr Lys
1 5 10 15

Lys Asp Ala Leu
20

<210> 1004
<211> 20
<212> PRT
<213> 斯氏李斯特菌

<400> 1004
Leu Arg Gly Ala Ile Ala Pro Ile Arg Asn Glu Ile Ile His Leu Lys
1 5 10 15

Lys His Ser Trp
20

<210> 1005
<211> 20
<212> PRT
<213> 威氏李斯特菌（Listeria weihenstephanensis）

<400> 1005
Ile Arg Gly Ser Ile Gln Gln Ile Arg Asn Glu Val Tyr His Cys Lys
1 5 10 15

Lys His Ser Trp
20

<210> 1006
<211> 20
<212> PRT
<213> 紐約恩施李斯特菌（Listeria newyorkensis）

<400> 1006
Ile Arg Gly Ser Ile Gln Gln Ile Arg Asn Glu Val Tyr His Cys Lys
1 5 10 15

Lys His Ser Trp
20

<210> 1007
<211> 21
<212> PRT
<213> 韋德纖毛菌

<400> 1007
Ile Ser Tyr Ser Ile Tyr Asn Val Arg Asn Gly Val Gly His Phe Asn
1 5 10 15

Lys Leu Ile Leu Gly
20

<210> 1008
<211> 21
<212> PRT
<213> 韋德纖毛菌

<400> 1008
Met Leu Asn Ala Ile Thr Ser Ile Arg His Arg Val Val His Tyr Asn
1 5 10 15

Met Asn Thr Asn Ser
20

<210> 1009
<211> 21
<212> PRT
<213> 韋德纖毛菌

<400> 1009
Ile Asp Glu Ala Ile Ser Ser Ile Arg His Gly Ile Val His Phe Asn
1 5 10 15

Leu Glu Leu Glu Gly
20

<210> 1010
<211> 22
<212> PRT
<213> 荚膜紅細菌（Rhodobacter capsulatus）

<400> 1010
Leu Leu Arg Tyr Leu Arg Gly Cys Arg Asn Gln Thr Phe His Leu Gly
1 5 10 15

Ala Arg Ala Gly Phe Leu
20

<210> 1011
<211> 21
<212> PRT
<213> 頰纖毛菌（Leptotrichia buccalis）

<400> 1011
Ile Asp Glu Ala Ile Ser Ser Ile Arg His Gly Ile Val His Phe Asn
1 5 10 15

Leu Glu Leu Glu Gly
20

<210> 1012
<211> 21
<212> PRT
<213> 纖毛菌屬某種（Leptotrichia sp.）

<400> 1012
Ile Asp Glu Ala Ile Ser Ser Ile Arg His Gly Ile Val His Phe Asn
1 5 10 15

Leu Glu Leu Glu Gly
20

<210> 1013
<211> 16
<212> PRT
<213> 纖毛菌屬某種

<400> 1013
Phe Gln Lys Glu Gly Tyr Leu Leu Arg Asn Lys Ile Leu His Asn Ser
1 5 10 15

<210> 1014
<211> 15
<212> PRT
<213> 沙氏纖毛菌

<400> 1014
Phe Thr Lys Ile Gly Thr Asn Glu Arg Asn Arg Ile Leu His Ala
1 5 10 15

<210> 1015
<211> 14
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 1015
Phe Arg Asn Glu Ile Asp His Phe His Tyr Phe Tyr Asp Arg
1 5 10

<210> 1016
<211> 14
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 1016
Leu Arg Asn Tyr Ile Glu His Phe Arg Tyr Tyr Ser Ser Phe
1 5 10

<210> 1017
<211> 14
<212> PRT
<213> 丁酸梭菌

<400> 1017
Val Arg Lys Tyr Val Asp His Phe Lys Tyr Tyr Ala Thr Ser
1 5 10

<210> 1018
<211> 14
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 1018
Leu Arg Lys Tyr Val Asp His Phe Lys Tyr Tyr Ala Tyr Gly
1 5 10

<210> 1019
<211> 14
<212> PRT
<213> 雞肉桿菌

<400> 1019
Ile Arg Asn Gln Thr Ala His Leu Ser Val Leu Gln Leu Glu
1 5 10

<210> 1020
<211> 14
<212> PRT
<213> 雞肉桿菌

<400> 1020
Ile Arg Asn Asn Ile Ala His Leu His Val Leu Arg Asn Asp
1 5 10

<210> 1021
<211> 14
<212> PRT
<213> propionicigenes帕魯迪菌

<400> 1021
Ile Arg Asn His Ile Ala His Phe Asn Tyr Leu Thr Lys Asp
1 5 10

<210> 1022

<211> 14
<212> PRT
<213> 斯氏李斯特菌

<400> 1022
Lys Arg Asn Asn Ile Ser His Phe Asn Tyr Leu Asn Gly Gln
1 5 10

<210> 1023
<211> 14
<212> PRT
<213> 威氏李斯特菌

<400> 1023
Ala Arg Asn His Ile Ala His Leu Asn Tyr Leu Ser Leu Lys
1 5 10

<210> 1024
<211> 14
<212> PRT
<213> 紐約恩施李斯特菌

<400> 1024
Ala Arg Asn His Ile Ala His Leu Asn Tyr Leu Ser Leu Lys
1 5 10

<210> 1025
<211> 14
<212> PRT
<213> 韋德纖毛菌

<400> 1025
Phe Arg Asn Tyr Ile Ala His Phe Leu His Leu His Thr Lys
1 5 10

<210> 1026
<211> 14
<212> PRT
<213> 韋德纖毛菌

<400> 1026
Ile Arg Asn Tyr Ile Ala His Phe Asn Tyr Ile Pro Asp Ala
1 5 10

<210> 1027
<211> 14
<212> PRT
<213> 韋德纖毛菌

<400> 1027
Ile Arg Asn Tyr Ile Ala His Phe Asn Tyr Ile Pro His Ala
1 5 10

<210> 1028
<211> 14
<212> PRT
<213> 莢膜紅細菌

<400> 1028
Thr Arg Lys Asp Leu Ala His Phe Asn Val Leu Asp Arg Ala
1 5 10

<210> 1029
<211> 14
<212> PRT
<213> 頰纖毛菌

<400> 1029
Ile Arg Asn Tyr Ile Ala His Phe Asn Tyr Ile Pro His Ala
1 5 10

<210> 1030
<211> 14
<212> PRT
<213> 纖毛菌屬某種

<400> 1030
Ile Arg Asn Tyr Ile Ala His Phe Asn Tyr Ile Pro Asn Ala
1 5 10

<210> 1031
<211> 14
<212> PRT
<213> 纖毛菌屬某種

<400> 1031
Ile Arg Asn Tyr Ile Ser His Phe Tyr Ile Val Arg Asn Pro
1 5 10

<210> 1032
<211> 14
<212> PRT
<213> 沙氏纖毛菌

<400> 1032
Ile Arg Asn Tyr Ile Ser His Phe Tyr Ile Val Arg Asn Pro
1 5 10

<210> 1033
<211> 1200
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
共有序列

<220>
<221> MOD_RES
<222> (7)..(7)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (11)..(13)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (18)..(18)
<223> Ile or Leu

<220>
<221> MOD_RES
<222> (20)..(21)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (39)..(39)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (75)..(75)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (93)..(93)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (100)..(100)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (145)..(145)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (149)..(149)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (154)..(154)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (216)..(216)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (236)..(236)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (278)..(278)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (315)..(315)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (317)..(317)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (319)..(319)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (370)..(370)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (407)..(407)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (419)..(419)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (428)..(428)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (445)..(445)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (455)..(455)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (563)..(563)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (717)..(717)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (795)..(796)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (800)..(800)

<223>任何胺基酸

<220>
<221> MOD_RES
<222> (942)..(942)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (971)..(971)
<223> Ile or Leu

<220>
<221> MOD_RES
<222> (1045)..(1045)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (1050)..(1050)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (1063)..(1064)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (1117)..(1118)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (1132)..(1132)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (1138)..(1138)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (1194)..(1194)
<223>任何胺基酸

<220>
<221> MOD_RES
<222> (1196)..(1196)
<223>任何胺基酸

<400> 1033
Met Leu Phe Phe Met Ser Xaa Asp Ile Thr Xaa Xaa Xaa Asn Met Met
1 5 10 15

Met Xaa Tyr Xaa Xaa Phe Thr Asn Leu Tyr Pro Leu Ser Lys Thr Leu
20 25 30

Arg Phe Glu Leu Lys Pro Xaa Gly Lys Thr Leu Glu Asn Ile Glu Lys
35 40 45

Asn Gly Leu Leu Glu Lys Asp Glu Lys Arg Ala Glu Asp Tyr Lys Lys
50 55 60

Val Lys Lys Ile Ile Asp Glu Tyr His Lys Xaa Phe Ile Glu Glu Ala
65 70 75 80

Leu Ser Ser Val Lys Leu Ser Glu Leu Glu Glu Tyr Xaa Asp Leu Tyr
85 90 95

Leu Lys Lys Xaa Lys Asp Asp Lys Asp Lys Lys Ala Leu Lys Lys Glu
100 105 110

Gln Asp Lys Leu Arg Lys Glu Ile Val Lys Ala Phe Lys Ala Asp Glu
115 120 125

```
Lys Tyr Lys Lys Leu Phe Lys Lys Glu Leu Ile Leu Lys Asp Leu Pro
    130                 135                 140

Xaa Phe Val Lys Xaa Glu Glu Asp Lys Xaa Leu Leu Lys Ser Phe Lys
145                 150                 155                 160

Gly Phe Thr Thr Tyr Phe Thr Gly Phe His Glu Asn Arg Lys Asn Met
                165                 170                 175

Tyr Ser Asp Glu Asp Lys Ser Thr Ala Ile Ala Tyr Arg Ile Ile His
            180                 185                 190

Glu Asn Leu Pro Lys Phe Ile Asp Asn Ile Lys Ile Phe Glu Lys Ile
        195                 200                 205

Lys Glu Lys Ala Glu Leu Leu Xaa Glx Ile Glu Leu Glu Glu Ile Phe
    210                 215                 220

Ser Leu Asp Tyr Tyr Asn Asn Val Leu Thr Gln Xaa Gly Ile Asp Lys
225                 230                 235                 240

Tyr Asn Ala Ile Ile Gly Gly Ile Ser Thr Glu Asp Gly Lys Ile Lys
                245                 250                 255

Gly Leu Asn Glu Tyr Ile Asn Leu Tyr Asn Gln Gln Asn Lys Asp Lys
            260                 265                 270

Lys Lys Leu Pro Lys Xaa Lys Pro Leu Tyr Lys Gln Ile Leu Ser Asp
        275                 280                 285

Arg Glu Ser Leu Ser Phe Leu Pro Glu Lys Phe Glu Asp Asp Glu Glu
    290                 295                 300

Val Leu Asp Ala Ile Lys Glu Phe Tyr Asp Xaa Ile Xaa Glu Xaa Ile
305                 310                 315                 320

Leu Glu Lys Leu Lys Leu Leu Phe Asp Asn Leu Ser Glu Tyr Asp Leu
                325                 330                 335

Ser Lys Ile Tyr Ile Lys Asn Asp Ala Leu Thr Thr Ile Ser Gln Lys
            340                 345                 350

Val Phe Gly Asp Trp Ser Val Ile Gly Asp Ala Leu Glu Glu Tyr Tyr
        355                 360                 365

Asp Xaa Pro Lys Lys Lys Lys Glu Lys Tyr Glu Glu Lys Arg Lys Lys
    370                 375                 380

Lys Leu Lys Lys Lys Lys Ser Phe Ser Leu Glu Glu Leu Asn Glu Leu
385                 390                 395                 400

Leu Glu Glu Arg Ile Glu Xaa Tyr Phe Ala Thr Leu Gly Asp Leu Ile
                405                 410                 415

Glu Asn Xaa Tyr Leu Ala Ala Glu Lys Leu Leu Xaa Thr Glu Tyr Pro
            420                 425                 430

Glu Glu Lys Asn Leu Lys Lys Asp Lys Glu Ala Val Xaa Ala Ile Lys
        435                 440                 445

Asp Leu Leu Asp Ser Ile Xaa Asx Leu Gln His Phe Leu Lys Pro Leu
    450                 455                 460
```

```
Leu Gly Lys Gly Asp Glu Leu Asp Lys Asp Glu Asn Phe Tyr Gly Glu
465                 470                 475                 480

Phe Glu Glu Leu Tyr Glu Glu Leu Asp Glu Ile Ile Pro Leu Tyr Asn
                485                 490                 495

Lys Val Arg Asn Tyr Leu Thr Gln Lys Pro Tyr Ser Thr Glu Lys Ile
            500                 505                 510

Lys Leu Asn Phe Glu Asn Pro Thr Leu Leu Asn Gly Trp Asp Lys Asn
        515                 520                 525

Lys Glu Lys Asp Asn Leu Ala Ile Ile Leu Arg Lys Asp Gly Lys Tyr
    530                 535                 540

Tyr Leu Gly Ile Met Asn Lys Lys His Asn Lys Ile Phe Asp Asn Lys
545                 550                 555                 560

Pro Lys Xaa Cys Tyr Glu Lys Met Val Tyr Lys Leu Leu Pro Gly Pro
                565                 570                 575

Asn Lys Met Leu Pro Lys Val Phe Phe Ser Lys Lys Asn Ile Lys Glu
            580                 585                 590

Tyr Asn Pro Ser Glu Glu Ile Leu Glu Ile Tyr Lys Lys Gly Thr His
        595                 600                 605

Lys Lys Gly Asp Asn Phe Ser Leu Lys Asp Cys His Lys Leu Ile Asp
    610                 615                 620

Phe Phe Lys Glu Ser Ile Glu Lys His Glu Asp Trp Lys Lys Phe Gly
625                 630                 635                 640

Phe Lys Phe Ser Asp Thr Glu Ser Tyr Asn Asp Ile Ser Glu Phe Tyr
                645                 650                 655

Arg Glu Val Glu Lys Gln Gly Tyr Lys Leu Ser Phe Thr Lys Ile Ser
            660                 665                 670

Glu Ser Tyr Ile Asp Ser Leu Val Glu Glu Gly Lys Leu Tyr Leu Phe
        675                 680                 685

Gln Ile Tyr Asn Lys Asp Phe Ser Pro Tyr Ser Lys Gly Lys Pro Asn
    690                 695                 700

Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Ser Glu Xaa Asn Leu Lys
705                 710                 715                 720

Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Ile Phe Tyr Arg Lys
                725                 730                 735

Ala Ser Ile Lys Lys Lys Lys Ile Thr His Lys Ala Gly Glu Pro Ile
            740                 745                 750

Lys Asn Lys Asn Asn Pro Lys Lys Glu Ser Lys Phe Glu Tyr Asp Ile
        755                 760                 765

Ile Lys Asp Lys Arg Tyr Thr Glu Asp Lys Phe Phe Phe His Val Pro
    770                 775                 780

Ile Thr Met Asn Phe Lys Ala Ser Gly Asn Xaa Xaa Phe Asn Asp Xaa
785                 790                 795                 800
```

```
Val Asn Glu Tyr Ile Arg Glu Asn Asp Asp Val His Ile Ile Gly Ile
                805                 810                 815

Asp Arg Gly Glu Arg Asn Leu Leu Tyr Tyr Ser Val Ile Asp Ser Lys
            820                 825                 830

Gly Asn Ile Leu Glu Gln Asp Ser Leu Asn Ile Ile Asn Asn Lys Thr
        835                 840                 845

Asp Tyr His Asp Leu Leu Asp Glu Arg Glu Lys Glu Arg Asp Lys Ala
    850                 855                 860

Arg Lys Asn Trp Gln Thr Ile Glu Asn Ile Lys Glu Leu Lys Glu Gly
865                 870                 875                 880

Tyr Leu Ser Gln Val Val His Lys Ile Ala Lys Leu Met Ile Lys Tyr
                885                 890                 895

Asn Ala Ile Val Val Leu Glu Asp Leu Asn Phe Gly Phe Lys Arg Gly
            900                 905                 910

Arg Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu Lys Met Leu
        915                 920                 925

Ile Asp Lys Leu Asn Tyr Leu Val Asp Lys Asp Lys Glu Xaa Asp Glu
    930                 935                 940

Ile Gly Gly Leu Leu Asn Ala Tyr Gln Leu Thr Asn Pro Phe Glu Ser
945                 950                 955                 960

Phe Lys Lys Leu Gly Lys Gln Thr Gly Phe Xaa Phe Tyr Val Pro Ala
                965                 970                 975

Trp Asn Thr Ser Lys Ile Asp Pro Thr Thr Gly Phe Val Asn Leu Phe
            980                 985                 990

Tyr Pro Lys Tyr Glu Asn Val Asp Lys Ala Lys Glu Phe Phe Ser Lys
        995                 1000                1005

Phe Asp  Ser Ile Arg Tyr Asn Asp Lys Asp Tyr  Phe Glu Phe Ala
    1010                1015                1020

Phe Asp  Tyr Lys Asn Phe Thr Lys Lys Ala Glu  Gly Ser Arg Thr
    1025                1030                1035

Lys Trp  Thr Ile Cys Thr Xaa Gly Glu Arg Ile  Xaa Asn Phe Arg
    1040                1045                1050

Asn Ser  Asp Lys Asn Asn Lys Trp Asp Xaa Xaa  Glu Ile Asp Leu
    1055                1060                1065

Thr Glu  Glu Leu Lys Glu Leu Phe Lys Asp Tyr  Gly Ile Asn Tyr
    1070                1075                1080

Gly Glu  Asp Leu Lys Glu Ala Ile Cys Ser Glu  Asp Asp Lys Asp
    1085                1090                1095

Phe Phe  Lys Ser Leu Leu Tyr Leu Leu Lys Leu  Thr Leu Gln Met
    1100                1105                1110

Arg Asn  Ser Xaa Xaa Asp Asp Tyr Ile Ile Ser  Pro Val Ala Asn
    1115                1120                1125
```

Asp Asn Gly Xaa Phe Phe Asp Ser Arg Xaa Ala Lys Leu Pro Lys
1130 1135 1140

Asp Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala Arg Lys Gly Leu
1145 1150 1155

Trp Leu Leu Glu Gln Ile Lys Asn Thr Asp Glu Gly Lys Lys Ala
1160 1165 1170

Leu Ile Ser Asn Lys Glu Trp Leu Glu Phe Ala Gln Asn Arg Pro
1175 1180 1185

Tyr Leu Lys Asp Ala Xaa Ala Xaa Asn Lys Lys His
1190 1195 1200

<210> 1034
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1034
gugccacuuc ucagaucgcu cgcucaguga uccgac 36

<210> 1035
<211> 105
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1035
gucagaacac ugagcgagcg uucuuuuuga gaagcucaac gggcuuugcc accuggaaag 60

uggccauugg cacacccguu gaaaaaauuc uguccucuag acaga 105

<210> 1036
<211> 105
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1036
gucagaacac ugagcgagcg uucuuuuuga gaagcucaac gggcuuugcc accuggaaag 60

uggccauugg cacacccguu gaaaaaauuc uguccucuag acaga 105

<210> 1037
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1037
gugccacuuc ucagaucgcu cgcucaguga uccgac 36

<210> 1038
<211> 37
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的

寡核苷酸

<400> 1038
gugccaauca cccaacacug accaagcuug ccgagac 37

<210> 1039
<211> 64
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1039
cuuggggaaa gcuaggcaag uuuuggauga uaagaaauaa ucaugucaca aggagggagu 60

uuuu 64

<210> 1040
<211> 64
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1040
cuuggggaaa gcuaggcaag uuuuggauga uaagaaauaa ucaugucaca aggagggagu 60

uuuu 64

<210> 1041
<211> 37
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1041
gugccaauca cccaacacug accaagcuug ccgagac 37

<210> 1042
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1042
gccgcagcga augccguuuc acgaaucguc aggcgg 36

<210> 1043
<211> 75
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1043
gcuggagacg uuuuuugaaa cggcgagugc ugcggauagc gaguuucucu uggggaggcg 60

cucgcggcca cuuuu 75

<210> 1044
<211> 75
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的

寡核苷酸

<400> 1044
gcuggagacg uuuuuugaaa cggcgagugc ugcggauagc gaguuucucu uggggaggcg 60

cucgcggcca cuuuu 75

<210> 1045
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1045
gccgcagcga augccguuuc acgaaucguc aggcgg 36

<210> 1046
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1046
guccaagaaa aaagaaauga uacgaggcau uagcac 36

<210> 1047
<211> 107
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1047
cuggacgaug ucucuuuuau uucuuuuuuc uuggaucuga guacgagcac ccacauugga 60

cauuucgcau ggugggugcu cguacuauag guaaaacaaa ccuuuuu 107

<210> 1048
<211> 107
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1048
cuggacgaug ucucuuuuau uucuuuuuuc uuggaucuga guacgagcac ccacauugga 60

cauuucgcau ggugggugcu cguacuauag guaaaacaaa ccuuuuu 107

<210> 1049
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1049
guccaagaaa aaagaaauga uacgaggcau uagcac 36

<210> 1050
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的

寡核苷酸

<400> 1050
guucgaaagc uuaguggaaa gcuucguccu uagcac 36

<210> 1051
<211> 69
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1051
cacggauaau cacgacuuuc cacuaagcuu ucgaauuuua ugaugcgagc auccucucag 60

gucaaaaaa 69

<210> 1052
<211> 69
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1052
cacggauaau cacgacuuuc cacuaagcuu ucgaauuuua ugaugcgagc auccucucag 60

gucaaaaaa 69

<210> 1053
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1053
guucgaaagc uuaguggaaa gcuucguggu uagcac 36

<210> 1054
<211> 35
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1054
guauugagaa aagccagaua uaguuggcaa uagac 35

<210> 1055
<211> 62
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1055
auauuuugau ucccauuuau gguuauuuac cauaaauggg aaucaacuaa aaaauauuuu 60

uu 62

<210> 1056
<211> 35
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的

寡核苷酸

<400> 1056
guauugagaa aagccagaua uaguuggcaa uagac 35

<210> 1057
<211> 62
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1057
auauuuugau ucccauuuau gguuauuuac cauaaauggg aaucaacuaa aaaauauuuu 60

uu 62

<210> 1058
<211> 35
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1058
guugaugaga agagcccaag auagagggca auaac 35

<210> 1059
<211> 35
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1059
gcuggagaag auagcccaag aaagagggca auaac 35

<210> 1060
<211> 78
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1060
auuauuacca uuuugguugg aaugcuauua uaaaggauca uucgauuauu accucuaccu 60

cccuucccac gauuucuu 78

<210> 1061
<211> 78
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1061
attattacca ttttggttgg aatgctatta taaaggatca ttcgattatt acctctacct 60

cccttcccac gatttctt 78

<210> 1062
<211> 35
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的

寡核苷酸

<400> 1062
guugaugaga agagcccaag auagagggca auaac 35

<210> 1063
<211> 78
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1063
attattacca ttttggttgg aatgctatta taaaggatca ttcgattatt acctctacct 60

cccttcccac gatttctt 78

<210> 1064
<211> 35
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1064
gcuggagaag auagcccaag aaagagggca auaac 35

<210> 1065
<211> 35
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1065
guuuggagaa cagcccgaua uagagggcaa uagac 35

<210> 1066
<211> 81
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1066
gucuuacgac cucaguauua ggaagauuuc aaccaagaaa acuuaguuuc aggcuuaaug 60

aucgagucau gcagccaaag u 81

<210> 1067
<211> 81
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1067
gucuuacgac cucaguauua ggaagauuuc aaccaagaaa acuuaguuuc aggcuuaaug 60

aucgagucau gcagccaaag u 81

<210> 1068
<211> 35
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的

寡核苷酸

<400> 1068
guuuggagaa cagcccgaua uagagggcaa uagac 35

<210> 1069
<211> 35
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1069
guuuugagaa uagcccgaca uagagggcaa uagac 35

<210> 1070
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1070
guuaugaaaa cagcccgaca uagagggcaa uagaca 36

<210> 1071
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1071
guuauagucc ucuuacauuu agagguaguc uuuaau 36

<210> 1072
<211> 98
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1072
ucuuaagaac uucucuaccu gaaguuggau uauaaaugac ucuugcucuc auagauaucc 60

uccuuugaaa auauacacug ccgauuaauu accguuuu 98

<210> 1073
<211> 98
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1073
ucuuaagaac uucucuaccu gaaguuggau uauaaaugac ucuugcucuc auagauaucc 60

uccuuugaaa auauacacug ccgauuaauu accguuuu 98

<210> 1074
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1074
guuauagucc ucuuacauuu agagguaguc uuuaau 36

<210> 1075
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1075
guuauagucc ucuuacauuu agagguaguu uauauu 36

<210> 1076
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1076
guuauagucc ccuuacauuu aggguaguc uuuaau 36

<210> 1077
<211> 102
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1077
aauauaaauu cucccuaaau auaagagaau aauaacucaa ucucuucauu cguauuuugu 60

cuaguuaaga uaaguaccac caaauacaau caauccaaaa aa 102

<210> 1078
<211> 102
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1078
aauauaaauu cucccuaaau auaagagaau aauaacucaa ucucuucauu cguauuuugu 60

cuaguuaaga uaaguaccac caaauacaau caauccaaaa aa 102

<210> 1079
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1079
guuauagucc ucuuacauuu agagguaguu uauauu 36

<210> 1080
<211> 102
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1080
aauauaaauu cucccuaaau auaagagaau aauaacucaa ucucuucauu cguauuuugu 60

cuaguuaaga uaaguaccac caaauacaau caauccaaaa aa 102

<210> 1081
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1081
guuauagucc ccuuacauuu agggguaguc uuuaau 36

<210> 1082
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1082
guuguaguuc ccuucaauuu ugggauaauc cacaag 36

<210> 1083
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1083
guuuuagucc ucuuucauau agagguaguc ucuuac 36

<210> 1084
<211> 99
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1084
augaaaagag gacuaaaacu gaaagaggac uaaaacacca gauguggaua acuauauuag 60

uggcuauuaa aaauucgucg auauuagaga ggaaacuuu 99

<210> 1085
<211> 99
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1085
augaaaagag gacuaaaacu gaaagaggac uaaaacacca gauguggaua acuauauuag 60

uggcuauuaa aaauucgucg auauuagaga ggaaacuuu 99

<210> 1086
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1086
guuuuagucc ucuuucauau agagguaguc ucuuac 36

<210> 1087
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1087
guuuuagacc ucuucuauuu ugagguacuc uaaauc 36

<210> 1088
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1088
guuuuagucc ucuuuuguuu ugagguacuc uaaauc 36

<210> 1089
<211> 147
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1089
aagucagcgc acaacaaaga agaugacgaa caaaaucucu cgccaucuuc uuaaaauuau 60

uugccacaca gccaacauua uaagcguuaa aaccagcacc augaguacau uucacccaac 120

aaucagaauc cccguuucuc cguuuuu 147

<210> 1090
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1090
guuuuagucc ucuuuuguuu ugagguacuc uaaauc 36

<210> 1091
<211> 147
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1091
aagucagcgc acaacaaaga agaugacgaa caaaaucucu cgccaucuuc uuaaaauuau 60

uugccacaca gccaacauua uaagcguuaa aaccagcacc augaguacau uucacccaac 120

aaucagaauc cccguuucuc cguuuuu 147

<210> 1092
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1092
guuuuagauc ccuucguuuu ugggguuauc uauauc 36

<210> 1093
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1093
guuuuagucc ccuucguuuu ugggguaguc uaaauc 36

<210> 1094
<211> 113
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1094
gauuuagagc accccaaaag uaaugaaaau uugcaauuaa auaaggaaua uuaaaaaaau 60

gugauuuuaa aaaaauugaa gaaauuaaau gaaaaauugu ccaaguaaaa aaa 113

<210> 1095
<211> 70
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1095
auuuagauua ccccuuuaau uuauuuuacc auauuuuucu cauaaugcaa acuaauauuc 60

caaaauuuuu 70

<210> 1096
<211> 113
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1096
gauuuagagc accccaaaag uaaugaaaau uugcaauuaa auaaggaaua uuaaaaaaau 60

gugauuuuaa aaaaauugaa gaaauuaaau gaaaaauugu ccaaguaaaa aaa 113

<210> 1097
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1097
guuuuagucc ccuucguuuu ugggguaguc uaaauc 36

<210> 1098
<211> 70
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1098
auuuagauua ccccuuuaau uuauuuuacc auauuuuucu cauaaugcaa acuaauauuc 60

caaaauuuuu 70

<210> 1099
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1099
guuuuagucc ccuucguuuu ugggguaguc uaaauc 36

<210> 1100
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1100
guuuuagucc ccuucgauau uggggugguc uauauc 36

<210> 1101
<211> 95
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1101
auugaugugg uauacuaaaa auggaaaauu guauuuuuga uuagaaagau guaaaauuga 60

uuuaauuuaa aaauauuuua uuagauuaaa guaga 95

<210> 1102
<211> 95
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1102
auugaugugg uauacuaaaa auggaaaauu guauuuuuga uuagaaagau guaaaauuga 60

uuuaauuuaa aaauauuuua uuagauuaaa guaga 95

<210> 1103
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1103
guuuuagucc ccuucgauau uggggugguc uauauc 36

<210> 1104
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1104
guucaguccg ccgucgucuu ggcggugaug ugaggc 36

<210> 1105
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1105
guucaguccg ccgucauuuu ggcggugaug ugcucc 36

<210> 1106
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1106
guucaguccg ccgucgucuu ggcggugaug ugaggc 36

<210> 1107
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1107
guucaguccg ccgucauuuu ggcggugaug ugcucc 36

<210> 1108
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1108
guucaguccg ccgucgucuu ggcggugaug ugaggc 36

<210> 1109
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1109
guucaguccg ccgucauuuu ggcggugaug ugcucc 36

<210> 1110
<211> 1300
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 1110
Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
20 25 30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys

35 40 45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
50 55 60

Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65 70 75 80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
85 90 95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
100 105 110

Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
115 120 125

Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
130 135 140

Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145 150 155 160

Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
165 170 175

Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190

Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205

Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
210 215 220

Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240

Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
245 250 255

Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
260 265 270

Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
275 280 285

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300

Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320

Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
325 330 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
340 345 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
355 360 365

Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln

370 375 380

Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385 390 395 400

Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
405 410 415

Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
420 425 430

Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
435 440 445

Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
450 455 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465 470 475 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
485 490 495

Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
500 505 510

Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
515 520 525

Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
530 535 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545 550 555 560

Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
565 570 575

Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
580 585 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
595 600 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
610 615 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625 630 635 640

Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
645 650 655

Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
660 665 670

Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
675 680 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln
690 695 700

Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe

705 710 715 720

Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp
725 730 735

Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
740 745 750

Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
755 760 765

Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
770 775 780

Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
785 790 795 800

Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
805 810 815

Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
820 825 830

Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
835 840 845

Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu
850 855 860

Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
865 870 875 880

His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
885 890 895

Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
900 905 910

Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
915 920 925

Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
930 935 940

Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945 950 955 960

Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
965 970 975

Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
980 985 990

Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu
995 1000 1005

Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val
1010 1015 1020

Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu
1025 1030 1035

Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg

1040 1045 1050

Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly
1055 1060 1065

Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser
1070 1075 1080

Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys
1085 1090 1095

Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp
1100 1105 1110

Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe
1115 1120 1125

Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr
1130 1135 1140

Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp
1145 1150 1155

Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu
1160 1165 1170

Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly
1175 1180 1185

Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe
1190 1195 1200

Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg
1205 1210 1215

Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val
1220 1225 1230

Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1235 1240 1245

Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly
1250 1255 1260

Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu
1265 1270 1275

Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1280 1285 1290

Phe Val Gln Asn Arg Asn Asn
1295 1300

<210> 1111
<211> 1477
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選佩萊格裡尼細菌序列

<400> 1111
Met Ser Asn Phe Phe Lys Asn Phe Thr Asn Leu Tyr Glu Leu Ser Lys
1 5 10 15

```
Thr Leu Arg Phe Glu Leu Lys Pro Val Gly Asp Thr Leu Thr Asn Met
            20                  25                  30

Lys Asp His Leu Glu Tyr Asp Glu Lys Leu Gln Thr Phe Leu Lys Asp
        35                  40                  45

Gln Asn Ile Asp Asp Ala Tyr Gln Ala Leu Lys Pro Gln Phe Asp Glu
    50                  55                  60

Ile His Glu Glu Phe Ile Thr Asp Ser Leu Glu Ser Lys Lys Ala Lys
65                  70                  75                  80

Glu Ile Asp Phe Ser Glu Tyr Leu Asp Leu Phe Gln Glu Lys Lys Glu
                85                  90                  95

Leu Asn Asp Ser Glu Lys Lys Leu Arg Asn Lys Ile Gly Glu Thr Phe
            100                 105                 110

Asn Lys Ala Gly Glu Lys Trp Lys Lys Glu Lys Tyr Pro Gln Tyr Glu
        115                 120                 125

Trp Lys Lys Gly Ser Lys Ile Ala Asn Gly Ala Asp Ile Leu Ser Cys
    130                 135                 140

Gln Asp Met Leu Gln Phe Ile Lys Tyr Lys Asn Pro Glu Asp Glu Lys
145                 150                 155                 160

Ile Lys Asn Tyr Ile Asp Asp Thr Leu Lys Gly Phe Phe Thr Tyr Phe
                165                 170                 175

Gly Gly Phe Asn Gln Asn Arg Ala Asn Tyr Tyr Glu Thr Lys Lys Glu
            180                 185                 190

Ala Ser Thr Ala Val Ala Thr Arg Ile Val His Glu Asn Leu Pro Lys
        195                 200                 205

Phe Cys Asp Asn Val Ile Gln Phe Lys His Ile Ile Lys Arg Lys Lys
    210                 215                 220

Asp Gly Thr Val Glu Lys Thr Glu Arg Lys Thr Glu Tyr Leu Asn Ala
225                 230                 235                 240

Tyr Gln Tyr Leu Lys Asn Asn Asn Lys Ile Thr Gln Ile Lys Asp Ala
                245                 250                 255

Glu Thr Glu Lys Met Ile Glu Ser Thr Pro Ile Ala Glu Lys Ile Phe
            260                 265                 270

Asp Val Tyr Tyr Phe Ser Ser Cys Leu Ser Gln Lys Gln Ile Glu Glu
        275                 280                 285

Tyr Asn Arg Ile Ile Gly His Tyr Asn Leu Leu Ile Asn Leu Tyr Asn
    290                 295                 300

Gln Ala Lys Arg Ser Glu Gly Lys His Leu Ser Ala Asn Glu Lys Lys
305                 310                 315                 320

Tyr Lys Asp Leu Pro Lys Phe Lys Thr Leu Tyr Lys Gln Ile Gly Cys
                325                 330                 335

Gly Lys Lys Lys Asp Leu Phe Tyr Thr Ile Lys Cys Asp Thr Glu Glu
            340                 345                 350
```

```
Glu Ala Asn Lys Ser Arg Asn Glu Gly Lys Glu Ser His Ser Val Glu
        355                 360                 365

Glu Ile Ile Asn Lys Ala Gln Glu Ala Ile Asn Lys Tyr Phe Lys Ser
    370                 375                 380

Asn Asn Asp Cys Glu Asn Ile Asn Thr Val Pro Asp Phe Ile Asn Tyr
385                 390                 395                 400

Ile Leu Thr Lys Glu Asn Tyr Glu Gly Val Tyr Trp Ser Lys Ala Ala
                405                 410                 415

Met Asn Thr Ile Ser Asp Lys Tyr Phe Ala Asn Tyr His Asp Leu Gln
            420                 425                 430

Asp Arg Leu Lys Glu Ala Lys Val Phe Gln Lys Ala Asp Lys Lys Ser
        435                 440                 445

Glu Asp Asp Ile Lys Ile Pro Glu Ala Ile Glu Leu Ser Gly Leu Phe
    450                 455                 460

Gly Val Leu Asp Ser Leu Ala Asp Trp Gln Thr Thr Leu Phe Lys Ser
465                 470                 475                 480

Ser Ile Leu Ser Asn Glu Asp Lys Leu Lys Ile Ile Thr Asp Ser Gln
                485                 490                 495

Thr Pro Ser Glu Ala Leu Leu Lys Met Ile Phe Asn Asp Ile Glu Lys
            500                 505                 510

Asn Met Glu Ser Phe Leu Lys Glu Thr Asn Asp Ile Ile Thr Leu Lys
        515                 520                 525

Lys Tyr Lys Gly Asn Lys Glu Gly Thr Glu Lys Ile Lys Gln Trp Phe
    530                 535                 540

Asp Tyr Thr Leu Ala Ile Asn Arg Met Leu Lys Tyr Phe Leu Val Lys
545                 550                 555                 560

Glu Asn Lys Ile Lys Gly Asn Ser Leu Asp Thr Asn Ile Ser Glu Ala
                565                 570                 575

Leu Lys Thr Leu Ile Tyr Ser Asp Asp Ala Glu Trp Phe Lys Trp Tyr
            580                 585                 590

Asp Ala Leu Arg Asn Tyr Leu Thr Gln Lys Pro Gln Asp Glu Ala Lys
        595                 600                 605

Glu Asn Lys Leu Lys Leu Asn Phe Asp Asn Pro Ser Leu Ala Gly Gly
    610                 615                 620

Trp Asp Val Asn Lys Glu Cys Ser Asn Phe Cys Val Ile Leu Lys Asp
625                 630                 635                 640

Lys Asn Glu Lys Lys Tyr Leu Ala Ile Met Lys Lys Gly Glu Asn Thr
                645                 650                 655

Leu Phe Gln Lys Glu Trp Thr Glu Gly Arg Gly Lys Asn Leu Thr Lys
            660                 665                 670

Lys Ser Asn Pro Leu Phe Glu Ile Asn Asn Cys Glu Ile Leu Ser Lys
        675                 680                 685
```

```
Met Glu Tyr Asp Phe Trp Ala Asp Val Ser Lys Met Ile Pro Lys Cys
    690                 695                 700

Ser Thr Gln Leu Lys Ala Val Val Asn His Phe Lys Gln Ser Asp Asn
705                 710                 715                 720

Glu Phe Ile Phe Pro Ile Gly Tyr Lys Val Thr Ser Gly Glu Lys Phe
                725                 730                 735

Arg Glu Glu Cys Lys Ile Ser Lys Gln Asp Phe Glu Leu Asn Asn Lys
            740                 745                 750

Val Phe Asn Lys Asn Glu Leu Ser Val Thr Ala Met Arg Tyr Asp Leu
        755                 760                 765

Ser Ser Thr Gln Glu Lys Gln Tyr Ile Lys Ala Phe Gln Lys Glu Tyr
    770                 775                 780

Trp Glu Leu Leu Phe Lys Gln Glu Lys Arg Asp Thr Lys Leu Thr Asn
785                 790                 795                 800

Asn Glu Ile Phe Asn Glu Trp Ile Asn Phe Cys Asn Lys Lys Tyr Ser
                805                 810                 815

Glu Leu Leu Ser Trp Glu Arg Lys Tyr Lys Asp Ala Leu Thr Asn Trp
            820                 825                 830

Ile Asn Phe Cys Lys Tyr Phe Leu Ser Lys Tyr Pro Lys Thr Thr Leu
        835                 840                 845

Phe Asn Tyr Ser Phe Lys Glu Ser Glu Asn Tyr Asn Ser Leu Asp Glu
    850                 855                 860

Phe Tyr Arg Asp Val Asp Ile Cys Ser Tyr Lys Leu Asn Ile Asn Thr
865                 870                 875                 880

Thr Ile Asn Lys Ser Ile Leu Asp Arg Leu Val Glu Glu Gly Lys Leu
                885                 890                 895

Tyr Leu Phe Glu Ile Lys Asn Gln Asp Ser Asn Asp Gly Lys Ser Ile
            900                 905                 910

Gly His Lys Asn Asn Leu His Thr Ile Tyr Trp Asn Ala Ile Phe Glu
        915                 920                 925

Asn Phe Asp Asn Arg Pro Lys Leu Asn Gly Glu Ala Glu Ile Phe Tyr
    930                 935                 940

Arg Lys Ala Ile Ser Lys Asp Lys Leu Gly Ile Val Lys Gly Lys Lys
945                 950                 955                 960

Thr Lys Asn Gly Thr Glu Ile Ile Lys Asn Tyr Arg Phe Ser Lys Glu
                965                 970                 975

Lys Phe Ile Leu His Val Pro Ile Thr Leu Asn Phe Cys Ser Asn Asn
            980                 985                 990

Glu Tyr Val Asn Asp Ile Val Asn Thr Lys Phe Tyr Asn Phe Ser Asn
        995                 1000                1005

Leu His Phe Leu Gly Ile Asp Arg Gly Glu Lys His Leu Ala Tyr
    1010                1015                1020
```

Tyr Ser Leu Val Asn Lys Asn Gly Glu Ile Val Asp Gln Gly Thr
1025 1030 1035

Leu Asn Leu Pro Phe Thr Asp Lys Asp Gly Asn Gln Arg Ser Ile
1040 1045 1050

Lys Lys Glu Lys Tyr Phe Tyr Asn Lys Gln Glu Asp Lys Trp Glu
1055 1060 1065

Ala Lys Glu Val Asp Cys Trp Asn Tyr Asn Asp Leu Leu Asp Ala
1070 1075 1080

Met Ala Ser Asn Arg Asp Met Ala Arg Lys Asn Trp Gln Arg Ile
1085 1090 1095

Gly Thr Ile Lys Glu Ala Lys Asn Gly Tyr Val Ser Leu Val Ile
1100 1105 1110

Arg Lys Ile Ala Asp Leu Ala Val Asn Asn Glu Arg Pro Ala Phe
1115 1120 1125

Ile Val Leu Glu Asp Leu Asn Thr Gly Phe Lys Arg Ser Arg Gln
1130 1135 1140

Lys Ile Asp Lys Ser Val Tyr Gln Lys Phe Glu Leu Ala Leu Ala
1145 1150 1155

Lys Lys Leu Asn Phe Leu Val Asp Lys Asn Ala Lys Arg Asp Glu
1160 1165 1170

Ile Gly Ser Pro Thr Lys Ala Leu Gln Leu Thr Pro Pro Val Asn
1175 1180 1185

Asn Tyr Gly Asp Ile Glu Asn Lys Lys Gln Ala Gly Ile Met Leu
1190 1195 1200

Tyr Thr Arg Ala Asn Tyr Thr Ser Gln Thr Asp Pro Ala Thr Gly
1205 1210 1215

Trp Arg Lys Thr Ile Tyr Leu Lys Ala Gly Pro Glu Glu Thr Thr
1220 1225 1230

Tyr Lys Lys Asp Gly Lys Ile Lys Asn Lys Ser Val Lys Asp Gln
1235 1240 1245

Ile Ile Glu Thr Phe Thr Asp Ile Gly Phe Asp Gly Lys Asp Tyr
1250 1255 1260

Tyr Phe Glu Tyr Asp Lys Gly Glu Phe Val Asp Glu Lys Thr Gly
1265 1270 1275

Glu Ile Lys Pro Lys Lys Trp Arg Leu Tyr Ser Gly Glu Asn Gly
1280 1285 1290

Lys Ser Leu Asp Arg Phe Arg Gly Glu Arg Glu Lys Asp Lys Tyr
1295 1300 1305

Glu Trp Lys Ile Asp Lys Ile Asp Ile Val Lys Ile Leu Asp Asp
1310 1315 1320

Leu Phe Val Asn Phe Asp Lys Asn Ile Ser Leu Leu Lys Gln Leu
1325 1330 1335

Lys Glu Gly Val Glu Leu Thr Arg Asn Asn Glu His Gly Thr Gly
1340 1345 1350

Glu Ser Leu Arg Phe Ala Ile Asn Leu Ile Gln Gln Ile Arg Asn
1355 1360 1365

Thr Gly Asn Asn Glu Arg Asp Asn Asp Phe Ile Leu Ser Pro Val
1370 1375 1380

Arg Asp Glu Asn Gly Lys His Phe Asp Ser Arg Glu Tyr Trp Asp
1385 1390 1395

Lys Glu Thr Lys Gly Glu Lys Ile Ser Met Pro Ser Ser Gly Asp
1400 1405 1410

Ala Asn Gly Ala Phe Asn Ile Ala Arg Lys Gly Ile Ile Met Asn
1415 1420 1425

Ala His Ile Leu Ala Asn Ser Asp Ser Lys Asp Leu Ser Leu Phe
1430 1435 1440

Val Ser Asp Glu Glu Trp Asp Leu His Leu Asn Asn Lys Thr Glu
1445 1450 1455

Trp Lys Lys Gln Leu Asn Ile Phe Ser Ser Arg Lys Ala Met Ala
1460 1465 1470

Lys Arg Lys Lys
1475

<210> 1112
<211> 1477
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選佩萊格裡尼細菌序列

<400> 1112
Met Ser Asn Phe Phe Lys Asn Phe Thr Asn Leu Tyr Glu Leu Ser Lys
1 5 10 15

Thr Leu Arg Phe Glu Leu Lys Pro Val Gly Asp Thr Leu Thr Asn Met
20 25 30

Lys Asp His Leu Glu Tyr Asp Glu Lys Leu Gln Thr Phe Leu Lys Asp
35 40 45

Gln Asn Ile Asp Asp Ala Tyr Gln Ala Leu Lys Pro Gln Phe Asp Glu
50 55 60

Ile His Glu Glu Phe Ile Thr Asp Ser Leu Glu Ser Lys Lys Ala Lys
65 70 75 80

Glu Ile Asp Phe Ser Glu Tyr Leu Asp Leu Phe Gln Glu Lys Lys Glu
85 90 95

Leu Asn Asp Ser Glu Lys Lys Leu Arg Asn Lys Ile Gly Glu Thr Phe
100 105 110

Asn Lys Ala Gly Glu Lys Trp Lys Lys Glu Lys Tyr Pro Gln Tyr Glu
115 120 125

Trp Lys Lys Gly Ser Lys Ile Ala Asn Gly Ala Asp Ile Leu Ser Cys

130 135 140

Gln Asp Met Leu Gln Phe Ile Lys Tyr Lys Asn Pro Glu Asp Glu Lys
145 150 155 160

Ile Lys Asn Tyr Ile Asp Asp Thr Leu Lys Gly Phe Phe Thr Tyr Phe
165 170 175

Gly Gly Phe Asn Gln Asn Arg Ala Asn Tyr Tyr Glu Thr Lys Lys Glu
180 185 190

Ala Ser Thr Ala Val Ala Thr Arg Ile Val His Glu Asn Leu Pro Lys
195 200 205

Phe Cys Asp Asn Val Ile Gln Phe Lys His Ile Ile Lys Arg Lys Lys
210 215 220

Asp Gly Thr Val Glu Lys Thr Glu Arg Lys Thr Glu Tyr Leu Asn Ala
225 230 235 240

Tyr Gln Tyr Leu Lys Asn Asn Asn Lys Ile Thr Gln Ile Lys Asp Ala
245 250 255

Glu Thr Glu Lys Met Ile Glu Ser Thr Pro Ile Ala Glu Lys Ile Phe
260 265 270

Asp Val Tyr Tyr Phe Ser Ser Cys Leu Ser Gln Lys Gln Ile Glu Glu
275 280 285

Tyr Asn Arg Ile Ile Gly His Tyr Asn Leu Leu Ile Asn Leu Tyr Asn
290 295 300

Gln Ala Lys Arg Ser Glu Gly Lys His Leu Ser Ala Asn Glu Lys Lys
305 310 315 320

Tyr Lys Asp Leu Pro Lys Phe Lys Thr Leu Tyr Lys Gln Ile Gly Cys
325 330 335

Gly Lys Lys Lys Asp Leu Phe Tyr Thr Ile Lys Cys Asp Thr Glu Glu
340 345 350

Glu Ala Asn Lys Ser Arg Asn Glu Gly Lys Glu Ser His Ser Val Glu
355 360 365

Glu Ile Ile Asn Lys Ala Gln Glu Ala Ile Asn Lys Tyr Phe Lys Ser
370 375 380

Asn Asn Asp Cys Glu Asn Ile Asn Thr Val Pro Asp Phe Ile Asn Tyr
385 390 395 400

Ile Leu Thr Lys Glu Asn Tyr Glu Gly Val Tyr Trp Ser Lys Ala Ala
405 410 415

Met Asn Thr Ile Ser Asp Lys Tyr Phe Ala Asn Tyr His Asp Leu Gln
420 425 430

Asp Arg Leu Lys Glu Ala Lys Val Phe Gln Lys Ala Asp Lys Lys Ser
435 440 445

Glu Asp Asp Ile Lys Ile Pro Glu Ala Ile Glu Leu Ser Gly Leu Phe
450 455 460

Gly Val Leu Asp Ser Leu Ala Asp Trp Gln Thr Thr Leu Phe Lys Ser

```
465                 470                 475                 480

Ser Ile Leu Ser Asn Glu Asp Lys Leu Lys Ile Ile Thr Asp Ser Gln
                485                 490                 495

Thr Pro Ser Glu Ala Leu Leu Lys Met Ile Phe Asn Asp Ile Glu Lys
            500                 505                 510

Asn Met Glu Ser Phe Leu Lys Glu Thr Asn Asp Ile Ile Thr Leu Lys
        515                 520                 525

Lys Tyr Lys Gly Asn Lys Glu Gly Thr Glu Lys Ile Lys Gln Trp Phe
    530                 535                 540

Asp Tyr Thr Leu Ala Ile Asn Arg Met Leu Lys Tyr Phe Leu Val Lys
545                 550                 555                 560

Glu Asn Lys Ile Lys Gly Asn Ser Leu Asp Thr Asn Ile Ser Glu Ala
                565                 570                 575

Leu Lys Thr Leu Ile Tyr Ser Asp Asp Ala Glu Trp Phe Lys Trp Tyr
            580                 585                 590

Asp Ala Leu Arg Asn Tyr Leu Thr Gln Lys Pro Gln Asp Glu Ala Lys
        595                 600                 605

Glu Asn Lys Leu Lys Leu Asn Phe Asp Asn Pro Ser Leu Ala Gly Gly
    610                 615                 620

Trp Asp Val Asn Lys Glu Cys Ser Asn Phe Cys Val Ile Leu Lys Asp
625                 630                 635                 640

Lys Asn Glu Lys Lys Tyr Leu Ala Ile Met Lys Lys Gly Glu Asn Thr
                645                 650                 655

Leu Phe Gln Lys Glu Trp Thr Glu Gly Arg Gly Lys Asn Leu Thr Lys
            660                 665                 670

Lys Ser Asn Pro Leu Phe Glu Ile Asn Asn Cys Glu Ile Leu Ser Lys
        675                 680                 685

Met Glu Tyr Asp Phe Trp Ala Asp Val Ser Lys Met Ile Pro Lys Cys
    690                 695                 700

Ser Thr Gln Leu Lys Ala Val Val Asn His Phe Lys Gln Ser Asp Asn
705                 710                 715                 720

Glu Phe Ile Phe Pro Ile Gly Tyr Lys Val Thr Ser Gly Glu Lys Phe
                725                 730                 735

Arg Glu Glu Cys Lys Ile Ser Lys Gln Asp Phe Glu Leu Asn Asn Lys
            740                 745                 750

Val Phe Asn Lys Asn Glu Leu Ser Val Thr Ala Met Arg Tyr Asp Leu
        755                 760                 765

Ser Ser Thr Gln Glu Lys Gln Tyr Ile Lys Ala Phe Gln Lys Glu Tyr
    770                 775                 780

Trp Glu Leu Leu Phe Lys Gln Glu Lys Arg Asp Thr Lys Leu Thr Asn
785                 790                 795                 800

Asn Glu Ile Phe Asn Glu Trp Ile Asn Phe Cys Asn Lys Lys Tyr Ser
```

805 810 815

Glu Leu Leu Ser Trp Glu Arg Lys Tyr Lys Asp Ala Leu Thr Asn Trp
820 825 830

Ile Asn Phe Cys Lys Tyr Phe Leu Ser Lys Tyr Pro Lys Thr Thr Leu
835 840 845

Phe Asn Tyr Ser Phe Lys Glu Ser Glu Asn Tyr Asn Ser Leu Asp Glu
850 855 860

Phe Tyr Arg Asp Val Asp Ile Cys Ser Tyr Lys Leu Asn Ile Asn Thr
865 870 875 880

Thr Ile Asn Lys Ser Ile Leu Asp Arg Leu Val Glu Glu Gly Lys Leu
885 890 895

Tyr Leu Phe Glu Ile Lys Asn Gln Asp Ser Asn Asp Gly Lys Ser Ile
900 905 910

Gly His Lys Asn Asn Leu His Thr Ile Tyr Trp Asn Ala Ile Phe Glu
915 920 925

Asn Phe Asp Asn Arg Pro Lys Leu Asn Gly Glu Ala Glu Ile Phe Tyr
930 935 940

Arg Lys Ala Ile Ser Lys Asp Lys Leu Gly Ile Val Lys Gly Lys Lys
945 950 955 960

Thr Lys Asn Gly Thr Glu Ile Ile Lys Asn Tyr Arg Phe Ser Lys Glu
965 970 975

Lys Phe Ile Leu His Val Pro Ile Thr Leu Asn Phe Cys Ser Asn Asn
980 985 990

Glu Tyr Val Asn Asp Ile Val Asn Thr Lys Phe Tyr Asn Phe Ser Asn
995 1000 1005

Leu His Phe Leu Gly Ile Asp Arg Gly Glu Lys His Leu Ala Tyr
1010 1015 1020

Tyr Ser Leu Val Asn Lys Asn Gly Glu Ile Val Asp Gln Gly Thr
1025 1030 1035

Leu Asn Leu Pro Phe Thr Asp Lys Asp Gly Asn Gln Arg Ser Ile
1040 1045 1050

Lys Lys Glu Lys Tyr Phe Tyr Asn Lys Gln Glu Asp Lys Trp Glu
1055 1060 1065

Ala Lys Glu Val Asp Cys Trp Asn Tyr Asn Asp Leu Leu Asp Ala
1070 1075 1080

Met Ala Ser Asn Arg Asp Met Ala Arg Lys Asn Trp Gln Arg Ile
1085 1090 1095

Gly Thr Ile Lys Glu Ala Lys Asn Gly Tyr Val Ser Leu Val Ile
1100 1105 1110

Arg Lys Ile Ala Asp Leu Ala Val Asn Asn Glu Arg Pro Ala Phe
1115 1120 1125

Ile Val Leu Glu Asp Leu Asn Thr Gly Phe Lys Arg Ser Arg Gln

1130 1135 1140

Lys Ile Asp Lys Ser Val Tyr Gln Lys Phe Glu Leu Ala Leu Ala
1145 1150 1155

Lys Lys Leu Asn Phe Leu Val Asp Lys Asn Ala Lys Arg Asp Glu
1160 1165 1170

Ile Gly Ser Pro Thr Lys Ala Leu Gln Leu Thr Pro Pro Val Asn
1175 1180 1185

Asn Tyr Gly Asp Ile Glu Asn Lys Lys Gln Ala Gly Ile Met Leu
1190 1195 1200

Tyr Thr Arg Ala Asn Tyr Thr Ser Gln Thr Asp Pro Ala Thr Gly
1205 1210 1215

Trp Arg Lys Thr Ile Tyr Leu Lys Ala Gly Pro Glu Glu Thr Thr
1220 1225 1230

Tyr Lys Lys Asp Gly Lys Ile Lys Asn Lys Ser Val Lys Asp Gln
1235 1240 1245

Ile Ile Glu Thr Phe Thr Asp Ile Gly Phe Asp Gly Lys Asp Tyr
1250 1255 1260

Tyr Phe Glu Tyr Asp Lys Gly Glu Phe Val Asp Glu Lys Thr Gly
1265 1270 1275

Glu Ile Lys Pro Lys Lys Trp Arg Leu Tyr Ser Gly Glu Asn Gly
1280 1285 1290

Lys Ser Leu Asp Arg Phe Arg Gly Glu Arg Glu Lys Asp Lys Tyr
1295 1300 1305

Glu Trp Lys Ile Asp Lys Ile Asp Ile Val Lys Ile Leu Asp Asp
1310 1315 1320

Leu Phe Val Asn Phe Asp Lys Asn Ile Ser Leu Leu Lys Gln Leu
1325 1330 1335

Lys Glu Gly Val Glu Leu Thr Arg Asn Asn Glu His Gly Thr Gly
1340 1345 1350

Glu Ser Leu Arg Phe Ala Ile Asn Leu Ile Gln Gln Ile Arg Asn
1355 1360 1365

Thr Gly Asn Asn Glu Arg Asp Asn Asp Phe Ile Leu Ser Pro Val
1370 1375 1380

Arg Asp Glu Asn Gly Lys His Phe Asp Ser Arg Glu Tyr Trp Asp
1385 1390 1395

Lys Glu Thr Lys Gly Glu Lys Ile Ser Met Pro Ser Ser Gly Asp
1400 1405 1410

Ala Asn Gly Ala Phe Asn Ile Ala Arg Lys Gly Ile Ile Met Asn
1415 1420 1425

Ala His Ile Leu Ala Asn Ser Asp Ser Lys Asp Leu Ser Leu Phe
1430 1435 1440

Val Ser Asp Glu Glu Trp Asp Leu His Leu Asn Asn Lys Thr Glu

1445 1450 1455

Trp Lys Lys Gln Leu Asn Ile Phe Ser Ser Arg Lys Ala Met Ala
1460 1465 1470

Lys Arg Lys Lys
1475

<210> 1113
<211> 1477
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選佩萊格裡尼細菌序列

<400> 1113
Met Ser Asn Phe Phe Lys Asn Phe Thr Asn Leu Tyr Glu Leu Ser Lys
1 5 10 15

Thr Leu Arg Phe Glu Leu Lys Pro Val Gly Asp Thr Leu Thr Asn Met
20 25 30

Lys Asp His Leu Glu Tyr Asp Glu Lys Leu Gln Thr Phe Leu Lys Asp
35 40 45

Gln Asn Ile Asp Asp Ala Tyr Gln Ala Leu Lys Pro Gln Phe Asp Glu
50 55 60

Ile His Glu Glu Phe Ile Thr Asp Ser Leu Glu Ser Lys Lys Ala Lys
65 70 75 80

Glu Ile Asp Phe Ser Glu Tyr Leu Asp Leu Phe Gln Glu Lys Lys Glu
85 90 95

Leu Asn Asp Ser Glu Lys Lys Leu Arg Asn Lys Ile Gly Glu Thr Phe
100 105 110

Asn Lys Ala Gly Glu Lys Trp Lys Lys Glu Lys Tyr Pro Gln Tyr Glu
115 120 125

Trp Lys Lys Gly Ser Lys Ile Ala Asn Gly Ala Asp Ile Leu Ser Cys
130 135 140

Gln Asp Met Leu Gln Phe Ile Lys Tyr Lys Asn Pro Glu Asp Glu Lys
145 150 155 160

Ile Lys Asn Tyr Ile Asp Asp Thr Leu Lys Gly Phe Phe Thr Tyr Phe
165 170 175

Gly Gly Phe Asn Gln Asn Arg Ala Asn Tyr Tyr Glu Thr Lys Lys Glu
180 185 190

Ala Ser Thr Ala Val Ala Thr Arg Ile Val His Glu Asn Leu Pro Lys
195 200 205

Phe Cys Asp Asn Val Ile Gln Phe Lys His Ile Ile Lys Arg Lys Lys
210 215 220

Asp Gly Thr Val Glu Lys Thr Glu Arg Lys Thr Glu Tyr Leu Asn Ala
225 230 235 240

Tyr Gln Tyr Leu Lys Asn Asn Asn Lys Ile Thr Gln Ile Lys Asp Ala
245 250 255

```
Glu Thr Glu Lys Met Ile Glu Ser Thr Pro Ile Ala Glu Lys Ile Phe
            260                 265                 270

Asp Val Tyr Tyr Phe Ser Ser Cys Leu Ser Gln Lys Gln Ile Glu Glu
        275                 280                 285

Tyr Asn Arg Ile Ile Gly His Tyr Asn Leu Leu Ile Asn Leu Tyr Asn
    290                 295                 300

Gln Ala Lys Arg Ser Glu Gly Lys His Leu Ser Ala Asn Glu Lys Lys
305                 310                 315                 320

Tyr Lys Asp Leu Pro Lys Phe Lys Thr Leu Tyr Lys Gln Ile Gly Cys
                325                 330                 335

Gly Lys Lys Lys Asp Leu Phe Tyr Thr Ile Lys Cys Asp Thr Glu Glu
            340                 345                 350

Glu Ala Asn Lys Ser Arg Asn Glu Gly Lys Glu Ser His Ser Val Glu
        355                 360                 365

Glu Ile Ile Asn Lys Ala Gln Glu Ala Ile Asn Lys Tyr Phe Lys Ser
    370                 375                 380

Asn Asn Asp Cys Glu Asn Ile Asn Thr Val Pro Asp Phe Ile Asn Tyr
385                 390                 395                 400

Ile Leu Thr Lys Glu Asn Tyr Glu Gly Val Tyr Trp Ser Lys Ala Ala
                405                 410                 415

Met Asn Thr Ile Ser Asp Lys Tyr Phe Ala Asn Tyr His Asp Leu Gln
            420                 425                 430

Asp Arg Leu Lys Glu Ala Lys Val Phe Gln Lys Ala Asp Lys Lys Ser
        435                 440                 445

Glu Asp Asp Ile Lys Ile Pro Glu Ala Ile Glu Leu Ser Gly Leu Phe
    450                 455                 460

Gly Val Leu Asp Ser Leu Ala Asp Trp Gln Thr Thr Leu Phe Lys Ser
465                 470                 475                 480

Ser Ile Leu Ser Asn Glu Asp Lys Leu Lys Ile Ile Thr Asp Ser Gln
                485                 490                 495

Thr Pro Ser Glu Ala Leu Leu Lys Met Ile Phe Asn Asp Ile Glu Lys
            500                 505                 510

Asn Met Glu Ser Phe Leu Lys Glu Thr Asn Asp Ile Ile Thr Leu Lys
        515                 520                 525

Lys Tyr Lys Gly Asn Lys Glu Gly Thr Glu Lys Ile Lys Gln Trp Phe
    530                 535                 540

Asp Tyr Thr Leu Ala Ile Asn Arg Met Leu Lys Tyr Phe Leu Val Lys
545                 550                 555                 560

Glu Asn Lys Ile Lys Gly Asn Ser Leu Asp Thr Asn Ile Ser Glu Ala
                565                 570                 575

Leu Lys Thr Leu Ile Tyr Ser Asp Asp Ala Glu Trp Phe Lys Trp Tyr
            580                 585                 590
```

```
Asp Ala Leu Arg Asn Tyr Leu Thr Gln Lys Pro Gln Asp Glu Ala Lys
        595                 600                 605

Glu Asn Lys Leu Lys Leu Asn Phe Asp Asn Pro Ser Leu Ala Gly Gly
    610                 615                 620

Trp Asp Val Asn Lys Glu Cys Ser Asn Phe Cys Val Ile Leu Lys Asp
625                 630                 635                 640

Lys Asn Glu Lys Lys Tyr Leu Ala Ile Met Lys Lys Gly Glu Asn Thr
                645                 650                 655

Leu Phe Gln Lys Glu Trp Thr Glu Gly Arg Gly Lys Asn Leu Thr Lys
            660                 665                 670

Lys Ser Asn Pro Leu Phe Glu Ile Asn Asn Cys Glu Ile Leu Ser Lys
        675                 680                 685

Met Glu Tyr Asp Phe Trp Ala Asp Val Ser Lys Met Ile Pro Lys Cys
    690                 695                 700

Ser Thr Gln Leu Lys Ala Val Val Asn His Phe Lys Gln Ser Asp Asn
705                 710                 715                 720

Glu Phe Ile Phe Pro Ile Gly Tyr Lys Val Thr Ser Gly Glu Lys Phe
                725                 730                 735

Arg Glu Glu Cys Lys Ile Ser Lys Gln Asp Phe Glu Leu Asn Asn Lys
            740                 745                 750

Val Phe Asn Lys Asn Glu Leu Ser Val Thr Ala Met Arg Tyr Asp Leu
        755                 760                 765

Ser Ser Thr Gln Glu Lys Gln Tyr Ile Lys Ala Phe Gln Lys Glu Tyr
    770                 775                 780

Trp Glu Leu Leu Phe Lys Gln Glu Lys Arg Asp Thr Lys Leu Thr Asn
785                 790                 795                 800

Asn Glu Ile Phe Asn Glu Trp Ile Asn Phe Cys Asn Lys Lys Tyr Ser
                805                 810                 815

Glu Leu Leu Ser Trp Glu Arg Lys Tyr Lys Asp Ala Leu Thr Asn Trp
            820                 825                 830

Ile Asn Phe Cys Lys Tyr Phe Leu Ser Lys Tyr Pro Lys Thr Thr Leu
        835                 840                 845

Phe Asn Tyr Ser Phe Lys Glu Ser Glu Asn Tyr Asn Ser Leu Asp Glu
    850                 855                 860

Phe Tyr Arg Asp Val Asp Ile Cys Ser Tyr Lys Leu Asn Ile Asn Thr
865                 870                 875                 880

Thr Ile Asn Lys Ser Ile Leu Asp Arg Leu Val Glu Glu Gly Lys Leu
                885                 890                 895

Tyr Leu Phe Glu Ile Lys Asn Gln Asp Ser Asn Asp Gly Lys Ser Ile
            900                 905                 910

Gly His Lys Asn Asn Leu His Thr Ile Tyr Trp Asn Ala Ile Phe Glu
        915                 920                 925
```

```
Asn Phe Asp Asn Arg Pro Lys Leu Asn Gly Glu Ala Glu Ile Phe Tyr
    930                 935                 940

Arg Lys Ala Ile Ser Lys Asp Lys Leu Gly Ile Val Lys Gly Lys Lys
945                 950                 955                 960

Thr Lys Asn Gly Thr Glu Ile Ile Lys Asn Tyr Arg Phe Ser Lys Glu
                965                 970                 975

Lys Phe Ile Leu His Val Pro Ile Thr Leu Asn Phe Cys Ser Asn Asn
            980                 985                 990

Glu Tyr Val Asn Asp Ile Val Asn Thr Lys Phe Tyr Asn Phe Ser Asn
        995                 1000                1005

Leu His  Phe Leu Gly Ile Asp  Arg Gly Glu Lys His  Leu Ala Tyr
    1010                1015                1020

Tyr Ser  Leu Val Asn Lys Asn  Gly Glu Ile Val Asp  Gln Gly Thr
    1025                1030                1035

Leu Asn  Leu Pro Phe Thr Asp  Lys Asp Gly Asn Gln  Arg Ser Ile
    1040                1045                1050

Lys Lys  Glu Lys Tyr Phe Tyr  Asn Lys Gln Glu Asp  Lys Trp Glu
    1055                1060                1065

Ala Lys  Glu Val Asp Cys Trp  Asn Tyr Asn Asp Leu  Leu Asp Ala
    1070                1075                1080

Met Ala  Ser Asn Arg Asp Met  Ala Arg Lys Asn Trp  Gln Arg Ile
    1085                1090                1095

Gly Thr  Ile Lys Glu Ala Lys  Asn Gly Tyr Val Ser  Leu Val Ile
    1100                1105                1110

Arg Lys  Ile Ala Asp Leu Ala  Val Asn Asn Glu Arg  Pro Ala Phe
    1115                1120                1125

Ile Val  Leu Glu Asp Leu Asn  Thr Gly Phe Lys Arg  Ser Arg Gln
    1130                1135                1140

Lys Ile  Asp Lys Ser Val Tyr  Gln Lys Phe Glu Leu  Ala Leu Ala
    1145                1150                1155

Lys Lys  Leu Asn Phe Leu Val  Asp Lys Asn Ala Lys  Arg Asp Glu
    1160                1165                1170

Ile Gly  Ser Pro Thr Lys Ala  Leu Gln Leu Thr Pro  Pro Val Asn
    1175                1180                1185

Asn Tyr  Gly Asp Ile Glu Asn  Lys Lys Gln Ala Gly  Ile Met Leu
    1190                1195                1200

Tyr Thr  Arg Ala Asn Tyr Thr  Ser Gln Thr Asp Pro  Ala Thr Gly
    1205                1210                1215

Trp Arg  Lys Thr Ile Tyr Leu  Lys Ala Gly Pro Glu  Glu Thr Thr
    1220                1225                1230

Tyr Lys  Lys Asp Gly Lys Ile  Lys Asn Lys Ser Val  Lys Asp Gln
    1235                1240                1245
```

Ile Ile Glu Thr Phe Thr Asp Ile Gly Phe Asp Gly Lys Asp Tyr
1250 1255 1260

Tyr Phe Glu Tyr Asp Lys Gly Glu Phe Val Asp Glu Lys Thr Gly
1265 1270 1275

Glu Ile Lys Pro Lys Lys Trp Arg Leu Tyr Ser Gly Glu Asn Gly
1280 1285 1290

Lys Ser Leu Asp Arg Phe Arg Gly Glu Arg Glu Lys Asp Lys Tyr
1295 1300 1305

Glu Trp Lys Ile Asp Lys Ile Asp Ile Val Lys Ile Leu Asp Asp
1310 1315 1320

Leu Phe Val Asn Phe Asp Lys Asn Ile Ser Leu Leu Lys Gln Leu
1325 1330 1335

Lys Glu Gly Val Glu Leu Thr Arg Asn Asn Glu His Gly Thr Gly
1340 1345 1350

Glu Ser Leu Arg Phe Ala Ile Asn Leu Ile Gln Gln Ile Arg Asn
1355 1360 1365

Thr Gly Asn Asn Glu Arg Asp Asn Asp Phe Ile Leu Ser Pro Val
1370 1375 1380

Arg Asp Glu Asn Gly Lys His Phe Asp Ser Arg Glu Tyr Trp Asp
1385 1390 1395

Lys Glu Thr Lys Gly Glu Lys Ile Ser Met Pro Ser Ser Gly Asp
1400 1405 1410

Ala Asn Gly Ala Phe Asn Ile Ala Arg Lys Gly Ile Ile Met Asn
1415 1420 1425

Ala His Ile Leu Ala Asn Ser Asp Ser Lys Asp Leu Ser Leu Phe
1430 1435 1440

Val Ser Asp Glu Glu Trp Asp Leu His Leu Asn Asn Lys Thr Glu
1445 1450 1455

Trp Lys Lys Gln Leu Asn Ile Phe Ser Ser Arg Lys Ala Met Ala
1460 1465 1470

Lys Arg Lys Lys
1475

<210> 1114
<211> 1403
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選佛爾扣菌細菌序列

<400> 1114
Met Leu Phe Phe Met Ser Thr Asp Ile Thr Asn Lys Pro Arg Glu Lys
1 5 10 15

Gly Val Phe Asp Asn Phe Thr Asn Leu Tyr Glu Phe Ser Lys Thr Leu
20 25 30

Thr Phe Gly Leu Ile Pro Leu Lys Trp Asp Asp Asn Lys Lys Met Ile

```
        35                  40                  45

Val Glu Asp Glu Asp Phe Ser Val Leu Arg Lys Tyr Gly Val Ile Glu
    50                  55                  60

Glu Asp Lys Arg Ile Ala Glu Ser Ile Lys Ile Ala Lys Phe Tyr Leu
65                  70                  75                  80

Asn Ile Leu His Arg Glu Leu Ile Gly Lys Val Leu Gly Ser Leu Lys
                85                  90                  95

Phe Glu Lys Lys Asn Leu Glu Asn Tyr Asp Arg Leu Leu Gly Glu Ile
            100                 105                 110

Glu Lys Asn Asn Lys Asn Glu Asn Ile Ser Glu Asp Lys Lys Lys Glu
        115                 120                 125

Ile Arg Lys Asn Phe Lys Lys Glu Leu Ser Ile Ala Gln Asp Ile Leu
    130                 135                 140

Leu Lys Lys Val Gly Glu Val Phe Glu Ser Asn Gly Ser Gly Ile Leu
145                 150                 155                 160

Ser Ser Lys Asn Cys Leu Asp Glu Leu Thr Lys Arg Phe Thr Arg Gln
                165                 170                 175

Glu Val Asp Lys Leu Arg Arg Glu Asn Lys Asp Ile Gly Val Glu Tyr
            180                 185                 190

Pro Asp Val Ala Tyr Arg Glu Lys Asp Gly Lys Glu Glu Thr Lys Ser
        195                 200                 205

Phe Phe Ala Met Asp Val Gly Tyr Leu Asp Asp Phe His Lys Asn Arg
    210                 215                 220

Lys Gln Leu Tyr Ser Val Lys Gly Lys Lys Asn Ser Leu Gly Arg Arg
225                 230                 235                 240

Ile Leu Asp Asn Phe Glu Ile Phe Cys Lys Asn Lys Lys Leu Tyr Glu
                245                 250                 255

Lys Tyr Lys Asn Leu Asp Ile Asp Phe Ser Glu Ile Glu Arg Asn Phe
            260                 265                 270

Asn Leu Thr Leu Glu Lys Val Phe Asp Phe Asp Asn Tyr Asn Glu Arg
        275                 280                 285

Leu Thr Gln Glu Gly Leu Asp Glu Tyr Ala Lys Ile Leu Gly Gly Glu
    290                 295                 300

Ser Asn Lys Gln Glu Arg Thr Ala Asn Ile His Gly Leu Asn Gln Ile
305                 310                 315                 320

Ile Asn Leu Tyr Ile Gln Lys Lys Gln Ser Glu Gln Lys Ala Glu Gln
                325                 330                 335

Lys Glu Thr Gly Lys Lys Lys Ile Lys Phe Asn Lys Lys Asp Tyr Pro
            340                 345                 350

Thr Phe Thr Cys Leu Gln Lys Gln Ile Leu Ser Gln Val Phe Arg Lys
        355                 360                 365

Glu Ile Ile Ile Glu Ser Asp Arg Asp Leu Ile Arg Glu Leu Lys Phe
```

370 375 380

Phe Val Glu Glu Ser Lys Glu Lys Val Asp Lys Ala Arg Gly Ile Ile
385 390 395 400

Glu Phe Leu Leu Asn His Glu Glu Asn Asp Ile Asp Leu Ala Met Val
405 410 415

Tyr Leu Pro Lys Ser Lys Ile Asn Ser Phe Val Tyr Lys Val Phe Lys
420 425 430

Glu Pro Gln Asp Phe Leu Ser Val Phe Gln Asp Gly Ala Ser Asn Leu
435 440 445

Asp Phe Val Ser Phe Asp Lys Ile Lys Thr His Leu Glu Asn Asn Lys
450 455 460

Leu Thr Tyr Lys Ile Phe Phe Lys Thr Leu Ile Lys Glu Asn His Asp
465 470 475 480

Phe Glu Ser Phe Leu Ile Leu Leu Gln Gln Glu Ile Asp Leu Leu Ile
485 490 495

Asp Gly Gly Glu Thr Val Thr Leu Gly Gly Lys Lys Glu Ser Ile Thr
500 505 510

Ser Leu Asp Glu Lys Lys Asn Arg Leu Lys Glu Lys Leu Gly Trp Phe
515 520 525

Glu Gly Lys Val Arg Glu Asn Glu Lys Met Lys Asp Glu Glu Glu Gly
530 535 540

Glu Phe Cys Ser Thr Val Leu Ala Tyr Ser Gln Ala Val Leu Asn Ile
545 550 555 560

Thr Lys Arg Ala Glu Ile Phe Trp Leu Asn Glu Lys Gln Asp Ala Lys
565 570 575

Val Gly Glu Asp Asn Lys Asp Met Ile Phe Tyr Lys Lys Phe Asp Glu
580 585 590

Phe Ala Asp Asp Gly Phe Ala Pro Phe Phe Tyr Phe Asp Lys Phe Gly
595 600 605

Asn Tyr Leu Lys Arg Arg Ser Arg Asn Thr Thr Lys Glu Ile Lys Leu
610 615 620

His Phe Gly Asn Asp Asp Leu Leu Glu Gly Trp Asp Met Asn Lys Glu
625 630 635 640

Pro Glu Tyr Trp Ser Phe Ile Leu Arg Asp Arg Asn Gln Tyr Tyr Leu
645 650 655

Gly Ile Gly Lys Lys Asp Gly Glu Ile Phe His Lys Lys Leu Gly Asn
660 665 670

Ser Val Glu Ala Val Lys Glu Ala Tyr Glu Leu Glu Asn Glu Ala Asp
675 680 685

Phe Tyr Glu Lys Ile Asp Tyr Lys Gln Leu Asn Ile Asp Arg Phe Glu
690 695 700

Gly Ile Ala Phe Pro Lys Lys Thr Lys Thr Glu Glu Ala Phe Arg Gln

705 710 715 720

Val Cys Lys Lys Arg Ala Asp Glu Phe Leu Gly Gly Asp Thr Tyr Glu
725 730 735

Phe Lys Ile Leu Leu Ala Ile Lys Lys Glu Tyr Asp Asp Phe Lys Ala
740 745 750

Arg Arg Gln Lys Glu Lys Asp Trp Asp Ser Lys Phe Ser Lys Glu Lys
755 760 765

Met Ser Lys Leu Ile Glu Tyr Tyr Ile Thr Cys Leu Gly Lys Arg Asp
770 775 780

Asp Trp Lys Arg Phe Asn Leu Asn Phe Arg Gln Pro Lys Glu Tyr Glu
785 790 795 800

Asp Arg Ser Asp Phe Val Arg His Ile Gln Arg Gln Ala Tyr Trp Ile
805 810 815

Asp Pro Arg Lys Val Ser Lys Asp Tyr Val Asp Lys Lys Val Ala Glu
820 825 830

Gly Glu Met Phe Leu Phe Lys Val His Asn Lys Asp Phe Tyr Asp Phe
835 840 845

Glu Arg Lys Ser Glu Asp Lys Lys Asn His Thr Ala Asn Leu Phe Thr
850 855 860

Gln Tyr Leu Leu Glu Leu Phe Ser Cys Glu Asn Ile Lys Asn Ile Lys
865 870 875 880

Ser Lys Asp Leu Ile Glu Ser Ile Phe Glu Leu Asp Gly Lys Ala Glu
885 890 895

Ile Arg Phe Arg Pro Lys Thr Asp Asp Val Lys Leu Lys Ile Tyr Gln
900 905 910

Lys Lys Gly Lys Asp Val Thr Tyr Ala Asp Lys Arg Asp Gly Asn Lys
915 920 925

Glu Lys Glu Val Ile Gln His Arg Arg Phe Ala Lys Asp Ala Leu Thr
930 935 940

Leu His Leu Lys Ile Arg Leu Asn Phe Gly Lys His Val Asn Leu Phe
945 950 955 960

Asp Phe Asn Lys Leu Val Asn Thr Glu Leu Phe Ala Lys Val Pro Val
965 970 975

Lys Ile Leu Gly Met Asp Arg Gly Glu Asn Asn Leu Ile Tyr Tyr Cys
980 985 990

Phe Leu Asp Glu His Gly Glu Ile Glu Asn Gly Lys Cys Gly Ser Leu
995 1000 1005

Asn Arg Val Gly Glu Gln Ile Ile Thr Leu Glu Asp Asp Lys Lys
1010 1015 1020

Val Lys Glu Pro Val Asp Tyr Phe Gln Leu Leu Val Asp Arg Glu
1025 1030 1035

Gly Gln Arg Asp Trp Glu Gln Lys Asn Trp Gln Lys Met Thr Arg

1040 1045 1050

Ile Lys Asp Leu Lys Lys Ala Tyr Leu Gly Asn Val Val Ser Trp
1055 1060 1065

Ile Ser Lys Glu Met Leu Ser Gly Ile Lys Glu Gly Val Val Thr
1070 1075 1080

Ile Gly Val Leu Glu Asp Leu Asn Ser Asn Phe Lys Arg Thr Arg
1085 1090 1095

Phe Phe Arg Glu Arg Gln Val Tyr Gln Gly Phe Glu Lys Ala Leu
1100 1105 1110

Val Asn Lys Leu Gly Tyr Leu Val Asp Lys Lys Tyr Asp Asn Tyr
1115 1120 1125

Arg Asn Val Tyr Gln Phe Ala Pro Ile Val Asp Ser Val Glu Glu
1130 1135 1140

Met Glu Lys Asn Lys Gln Ile Gly Thr Leu Val Tyr Val Pro Ala
1145 1150 1155

Ser Tyr Thr Ser Lys Ile Cys Pro His Pro Lys Cys Gly Trp Arg
1160 1165 1170

Glu Arg Leu Tyr Met Lys Asn Ser Ala Ser Lys Glu Lys Ile Val
1175 1180 1185

Gly Leu Leu Lys Ser Asp Gly Ile Lys Ile Ser Tyr Asp Gln Lys
1190 1195 1200

Asn Asp Arg Phe Tyr Phe Glu Tyr Gln Trp Glu Gln Glu His Lys
1205 1210 1215

Ser Asp Gly Lys Lys Lys Lys Tyr Ser Gly Val Asp Lys Val Phe
1220 1225 1230

Ser Asn Val Ser Arg Met Arg Trp Asp Val Glu Gln Lys Lys Ser
1235 1240 1245

Ile Asp Phe Val Asp Gly Thr Asp Gly Ser Ile Thr Asn Lys Leu
1250 1255 1260

Lys Ser Leu Leu Lys Gly Lys Gly Ile Glu Leu Asp Asn Ile Asn
1265 1270 1275

Gln Gln Ile Val Asn Gln Gln Lys Glu Leu Gly Val Glu Phe Phe
1280 1285 1290

Gln Ser Ile Ile Phe Tyr Phe Asn Leu Ile Met Gln Ile Arg Asn
1295 1300 1305

Tyr Asp Lys Glu Lys Ser Gly Ser Glu Ala Asp Tyr Ile Gln Cys
1310 1315 1320

Pro Ser Cys Leu Phe Asp Ser Arg Lys Pro Glu Met Asn Gly Lys
1325 1330 1335

Leu Ser Ala Ile Thr Asn Gly Asp Ala Asn Gly Ala Tyr Asn Ile
1340 1345 1350

Ala Arg Lys Gly Phe Met Gln Leu Cys Arg Ile Arg Glu Asn Pro

1355 1360 1365

Gln Glu Pro Met Lys Leu Ile Thr Asn Arg Glu Trp Asp Glu Ala
1370 1375 1380

Val Arg Glu Trp Asp Ile Tyr Ser Ala Ala Gln Lys Ile Pro Val
1385 1390 1395

Leu Ser Glu Glu Asn
1400

<210> 1115
<211> 1403
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選佛爾扣菌細菌序列

<400> 1115
Met Leu Phe Phe Met Ser Thr Asp Ile Thr Asn Lys Pro Arg Glu Lys
1 5 10 15

Gly Val Phe Asp Asn Phe Thr Asn Leu Tyr Glu Phe Ser Lys Thr Leu
20 25 30

Thr Phe Gly Leu Ile Pro Leu Lys Trp Asp Asp Asn Lys Lys Met Ile
35 40 45

Val Glu Asp Glu Asp Phe Ser Val Leu Arg Lys Tyr Gly Val Ile Glu
50 55 60

Glu Asp Lys Arg Ile Ala Glu Ser Ile Lys Ile Ala Lys Phe Tyr Leu
65 70 75 80

Asn Ile Leu His Arg Glu Leu Ile Gly Lys Val Leu Gly Ser Leu Lys
85 90 95

Phe Glu Lys Lys Asn Leu Glu Asn Tyr Asp Arg Leu Leu Gly Glu Ile
100 105 110

Glu Lys Asn Asn Lys Asn Glu Asn Ile Ser Glu Asp Lys Lys Lys Glu
115 120 125

Ile Arg Lys Asn Phe Lys Lys Glu Leu Ser Ile Ala Gln Asp Ile Leu
130 135 140

Leu Lys Lys Val Gly Glu Val Phe Glu Ser Asn Gly Ser Gly Ile Leu
145 150 155 160

Ser Ser Lys Asn Cys Leu Asp Glu Leu Thr Lys Arg Phe Thr Arg Gln
165 170 175

Glu Val Asp Lys Leu Arg Arg Glu Asn Lys Asp Ile Gly Val Glu Tyr
180 185 190

Pro Asp Val Ala Tyr Arg Glu Lys Asp Gly Lys Glu Glu Thr Lys Ser
195 200 205

Phe Phe Ala Met Asp Val Gly Tyr Leu Asp Asp Phe His Lys Asn Arg
210 215 220

Lys Gln Leu Tyr Ser Val Lys Gly Lys Lys Asn Ser Leu Gly Arg Arg
225 230 235 240

```
Ile Leu Asp Asn Phe Glu Ile Phe Cys Lys Asn Lys Lys Leu Tyr Glu
                245                 250                 255

Lys Tyr Lys Asn Leu Asp Ile Asp Phe Ser Glu Ile Glu Arg Asn Phe
            260                 265                 270

Asn Leu Thr Leu Glu Lys Val Phe Asp Phe Asp Asn Tyr Asn Glu Arg
        275                 280                 285

Leu Thr Gln Glu Gly Leu Asp Glu Tyr Ala Lys Ile Leu Gly Gly Glu
    290                 295                 300

Ser Asn Lys Gln Glu Arg Thr Ala Asn Ile His Gly Leu Asn Gln Ile
305                 310                 315                 320

Ile Asn Leu Tyr Ile Gln Lys Lys Gln Ser Glu Gln Lys Ala Glu Gln
                325                 330                 335

Lys Glu Thr Gly Lys Lys Lys Ile Lys Phe Asn Lys Lys Asp Tyr Pro
            340                 345                 350

Thr Phe Thr Cys Leu Gln Lys Gln Ile Leu Ser Gln Val Phe Arg Lys
        355                 360                 365

Glu Ile Ile Ile Glu Ser Asp Arg Asp Leu Ile Arg Glu Leu Lys Phe
    370                 375                 380

Phe Val Glu Glu Ser Lys Glu Lys Val Asp Lys Ala Arg Gly Ile Ile
385                 390                 395                 400

Glu Phe Leu Leu Asn His Glu Glu Asn Asp Ile Asp Leu Ala Met Val
                405                 410                 415

Tyr Leu Pro Lys Ser Lys Ile Asn Ser Phe Val Tyr Lys Val Phe Lys
            420                 425                 430

Glu Pro Gln Asp Phe Leu Ser Val Phe Gln Asp Gly Ala Ser Asn Leu
        435                 440                 445

Asp Phe Val Ser Phe Asp Lys Ile Lys Thr His Leu Glu Asn Asn Lys
    450                 455                 460

Leu Thr Tyr Lys Ile Phe Phe Lys Thr Leu Ile Lys Glu Asn His Asp
465                 470                 475                 480

Phe Glu Ser Phe Leu Ile Leu Leu Gln Gln Glu Ile Asp Leu Leu Ile
                485                 490                 495

Asp Gly Gly Glu Thr Val Thr Leu Gly Gly Lys Lys Glu Ser Ile Thr
            500                 505                 510

Ser Leu Asp Glu Lys Lys Asn Arg Leu Lys Glu Lys Leu Gly Trp Phe
        515                 520                 525

Glu Gly Lys Val Arg Glu Asn Glu Lys Met Lys Asp Glu Glu Glu Gly
    530                 535                 540

Glu Phe Cys Ser Thr Val Leu Ala Tyr Ser Gln Ala Val Leu Asn Ile
545                 550                 555                 560

Thr Lys Arg Ala Glu Ile Phe Trp Leu Asn Glu Lys Gln Asp Ala Lys
                565                 570                 575
```

```
Val Gly Glu Asp Asn Lys Asp Met Ile Phe Tyr Lys Lys Phe Asp Glu
            580                 585                 590

Phe Ala Asp Asp Gly Phe Ala Pro Phe Phe Tyr Phe Asp Lys Phe Gly
        595                 600                 605

Asn Tyr Leu Lys Arg Arg Ser Arg Asn Thr Thr Lys Glu Ile Lys Leu
    610                 615                 620

His Phe Gly Asn Asp Asp Leu Leu Glu Gly Trp Asp Met Asn Lys Glu
625                 630                 635                 640

Pro Glu Tyr Trp Ser Phe Ile Leu Arg Asp Arg Asn Gln Tyr Tyr Leu
                645                 650                 655

Gly Ile Gly Lys Lys Asp Gly Glu Ile Phe His Lys Lys Leu Gly Asn
            660                 665                 670

Ser Val Glu Ala Val Lys Glu Ala Tyr Glu Leu Glu Asn Glu Ala Asp
        675                 680                 685

Phe Tyr Glu Lys Ile Asp Tyr Lys Gln Leu Asn Ile Asp Arg Phe Glu
    690                 695                 700

Gly Ile Ala Phe Pro Lys Lys Thr Lys Thr Glu Glu Ala Phe Arg Gln
705                 710                 715                 720

Val Cys Lys Lys Arg Ala Asp Glu Phe Leu Gly Gly Asp Thr Tyr Glu
                725                 730                 735

Phe Lys Ile Leu Leu Ala Ile Lys Lys Glu Tyr Asp Asp Phe Lys Ala
            740                 745                 750

Arg Arg Gln Lys Glu Lys Asp Trp Asp Ser Lys Phe Ser Lys Glu Lys
        755                 760                 765

Met Ser Lys Leu Ile Glu Tyr Tyr Ile Thr Cys Leu Gly Lys Arg Asp
    770                 775                 780

Asp Trp Lys Arg Phe Asn Leu Asn Phe Arg Gln Pro Lys Glu Tyr Glu
785                 790                 795                 800

Asp Arg Ser Asp Phe Val Arg His Ile Gln Arg Gln Ala Tyr Trp Ile
                805                 810                 815

Asp Pro Arg Lys Val Ser Lys Asp Tyr Val Asp Lys Lys Val Ala Glu
            820                 825                 830

Gly Glu Met Phe Leu Phe Lys Val His Asn Lys Asp Phe Tyr Asp Phe
        835                 840                 845

Glu Arg Lys Ser Glu Asp Lys Lys Asn His Thr Ala Asn Leu Phe Thr
    850                 855                 860

Gln Tyr Leu Leu Glu Leu Phe Ser Cys Glu Asn Ile Lys Asn Ile Lys
865                 870                 875                 880

Ser Lys Asp Leu Ile Glu Ser Ile Phe Glu Leu Asp Gly Lys Ala Glu
                885                 890                 895

Ile Arg Phe Arg Pro Lys Thr Asp Asp Val Lys Leu Lys Ile Tyr Gln
            900                 905                 910
```

```
Lys Lys Gly Lys Asp Val Thr Tyr Ala Asp Lys Arg Asp Gly Asn Lys
        915                 920                 925

Glu Lys Glu Val Ile Gln His Arg Arg Phe Ala Lys Asp Ala Leu Thr
    930                 935                 940

Leu His Leu Lys Ile Arg Leu Asn Phe Gly Lys His Val Asn Leu Phe
945                 950                 955                 960

Asp Phe Asn Lys Leu Val Asn Thr Glu Leu Phe Ala Lys Val Pro Val
                965                 970                 975

Lys Ile Leu Gly Met Asp Arg Gly Glu Asn Asn Leu Ile Tyr Tyr Cys
            980                 985                 990

Phe Leu Asp Glu His Gly Glu Ile Glu Asn Gly Lys Cys Gly Ser Leu
        995                 1000                1005

Asn Arg  Val Gly Glu Gln Ile  Ile Thr Leu Glu Asp  Asp Lys Lys
    1010                1015                1020

Val Lys  Glu Pro Val Asp Tyr  Phe Gln Leu Leu Val  Asp Arg Glu
    1025                1030                1035

Gly Gln  Arg Asp Trp Glu Gln  Lys Asn Trp Gln Lys  Met Thr Arg
    1040                1045                1050

Ile Lys  Asp Leu Lys Lys Ala  Tyr Leu Gly Asn Val  Val Ser Trp
    1055                1060                1065

Ile Ser  Lys Glu Met Leu Ser  Gly Ile Lys Glu Gly  Val Val Thr
    1070                1075                1080

Ile Gly  Val Leu Glu Asp Leu  Asn Ser Asn Phe Lys  Arg Thr Arg
    1085                1090                1095

Phe Phe  Arg Glu Arg Gln Val  Tyr Gln Gly Phe Glu  Lys Ala Leu
    1100                1105                1110

Val Asn  Lys Leu Gly Tyr Leu  Val Asp Lys Lys Tyr  Asp Asn Tyr
    1115                1120                1125

Arg Asn  Val Tyr Gln Phe Ala  Pro Ile Val Asp Ser  Val Glu Glu
    1130                1135                1140

Met Glu  Lys Asn Lys Gln Ile  Gly Thr Leu Val Tyr  Val Pro Ala
    1145                1150                1155

Ser Tyr  Thr Ser Lys Ile Cys  Pro His Pro Lys Cys  Gly Trp Arg
    1160                1165                1170

Glu Arg  Leu Tyr Met Lys Asn  Ser Ala Ser Lys Glu  Lys Ile Val
    1175                1180                1185

Gly Leu  Leu Lys Ser Asp Gly  Ile Lys Ile Ser Tyr  Asp Gln Lys
    1190                1195                1200

Asn Asp  Arg Phe Tyr Phe Glu  Tyr Gln Trp Glu Gln  Glu His Lys
    1205                1210                1215

Ser Asp  Gly Lys Lys Lys Lys  Tyr Ser Gly Val Asp  Lys Val Phe
    1220                1225                1230
```

Ser Asn Val Ser Arg Met Arg Trp Asp Val Glu Gln Lys Lys Ser
1235 1240 1245

Ile Asp Phe Val Asp Gly Thr Asp Gly Ser Ile Thr Asn Lys Leu
1250 1255 1260

Lys Ser Leu Leu Lys Gly Lys Gly Ile Glu Leu Asp Asn Ile Asn
1265 1270 1275

Gln Gln Ile Val Asn Gln Gln Lys Glu Leu Gly Val Glu Phe Phe
1280 1285 1290

Gln Ser Ile Ile Phe Tyr Phe Asn Leu Ile Met Gln Ile Arg Asn
1295 1300 1305

Tyr Asp Lys Glu Lys Ser Gly Ser Glu Ala Asp Tyr Ile Gln Cys
1310 1315 1320

Pro Ser Cys Leu Phe Asp Ser Arg Lys Pro Glu Met Asn Gly Lys
1325 1330 1335

Leu Ser Ala Ile Thr Asn Gly Asp Ala Asn Gly Ala Tyr Asn Ile
1340 1345 1350

Ala Arg Lys Gly Phe Met Gln Leu Cys Arg Ile Arg Glu Asn Pro
1355 1360 1365

Gln Glu Pro Met Lys Leu Ile Thr Asn Arg Glu Trp Asp Glu Ala
1370 1375 1380

Val Arg Glu Trp Asp Ile Tyr Ser Ala Ala Gln Lys Ile Pro Val
1385 1390 1395

Leu Ser Glu Glu Asn
1400

<210> 1116
<211> 1352
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
Parcubacteria群細菌序列

<400> 1116
Met Glu Asn Ile Phe Asp Gln Phe Ile Gly Lys Tyr Ser Leu Ser Lys
1 5 10 15

Thr Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr Glu Asp Phe Leu
20 25 30

Lys Ile Asn Lys Val Phe Glu Lys Asp Gln Thr Ile Asp Asp Ser Tyr
35 40 45

Asn Gln Ala Lys Phe Tyr Phe Asp Ser Leu His Gln Lys Phe Ile Asp
50 55 60

Ala Ala Leu Ala Ser Asp Lys Thr Ser Glu Leu Ser Phe Gln Asn Phe
65 70 75 80

Ala Asp Val Leu Glu Lys Gln Asn Lys Ile Ile Leu Asp Lys Lys Arg
85 90 95

Glu Met Gly Ala Leu Arg Lys Arg Asp Lys Asn Ala Val Gly Ile Asp

100 105 110

Arg Leu Gln Lys Glu Ile Asn Asp Ala Glu Asp Ile Ile Gln Lys Glu
115 120 125

Lys Glu Lys Ile Tyr Lys Asp Val Arg Thr Leu Phe Asp Asn Glu Ala
130 135 140

Glu Ser Trp Lys Thr Tyr Tyr Gln Glu Arg Glu Val Asp Gly Lys Lys
145 150 155 160

Ile Thr Phe Ser Lys Ala Asp Leu Lys Gln Lys Gly Ala Asp Phe Leu
165 170 175

Thr Ala Ala Gly Ile Leu Lys Val Leu Lys Tyr Glu Phe Pro Glu Glu
180 185 190

Lys Glu Lys Glu Phe Gln Ala Lys Asn Gln Pro Ser Leu Phe Val Glu
195 200 205

Glu Lys Glu Asn Pro Gly Gln Lys Arg Tyr Ile Phe Asp Ser Phe Asp
210 215 220

Lys Phe Ala Gly Tyr Leu Thr Lys Phe Gln Gln Thr Lys Lys Asn Leu
225 230 235 240

Tyr Ala Ala Asp Gly Thr Ser Thr Ala Val Ala Thr Arg Ile Ala Asp
245 250 255

Asn Phe Ile Ile Phe His Gln Asn Thr Lys Val Phe Arg Asp Lys Tyr
260 265 270

Lys Asn Asn His Thr Asp Leu Gly Phe Asp Glu Glu Asn Ile Phe Glu
275 280 285

Ile Glu Arg Tyr Lys Asn Cys Leu Leu Gln Arg Glu Ile Glu His Ile
290 295 300

Lys Asn Glu Asn Ser Tyr Asn Lys Ile Ile Gly Arg Ile Asn Lys Lys
305 310 315 320

Ile Lys Glu Tyr Arg Asp Gln Lys Ala Lys Asp Thr Lys Leu Thr Lys
325 330 335

Ser Asp Phe Pro Phe Phe Lys Asn Leu Asp Lys Gln Ile Leu Gly Glu
340 345 350

Val Glu Lys Glu Lys Gln Leu Ile Glu Lys Thr Arg Glu Lys Thr Glu
355 360 365

Glu Asp Val Leu Ile Glu Arg Phe Lys Glu Phe Ile Glu Asn Asn Glu
370 375 380

Glu Arg Phe Thr Ala Ala Lys Lys Leu Met Asn Ala Phe Cys Asn Gly
385 390 395 400

Glu Phe Glu Ser Glu Tyr Glu Gly Ile Tyr Leu Lys Asn Lys Ala Ile
405 410 415

Asn Thr Ile Ser Arg Arg Trp Phe Val Ser Asp Arg Asp Phe Glu Leu
420 425 430

Lys Leu Pro Gln Gln Lys Ser Lys Asn Lys Ser Glu Lys Asn Glu Pro

435 440 445

Lys Val Lys Lys Phe Ile Ser Ile Ala Glu Ile Lys Asn Ala Val Glu
450 455 460

Glu Leu Asp Gly Asp Ile Phe Lys Ala Val Phe Tyr Asp Lys Lys Ile
465 470 475 480

Ile Ala Gln Gly Gly Ser Lys Leu Glu Gln Phe Leu Val Ile Trp Lys
485 490 495

Tyr Glu Phe Glu Tyr Leu Phe Arg Asp Ile Glu Arg Glu Asn Gly Glu
500 505 510

Lys Leu Leu Gly Tyr Asp Ser Cys Leu Lys Ile Ala Lys Gln Leu Gly
515 520 525

Ile Phe Pro Gln Glu Lys Glu Ala Arg Glu Lys Ala Thr Ala Val Ile
530 535 540

Lys Asn Tyr Ala Asp Ala Gly Leu Gly Ile Phe Gln Met Met Lys Tyr
545 550 555 560

Phe Ser Leu Asp Asp Lys Asp Arg Lys Asn Thr Pro Gly Gln Leu Ser
565 570 575

Thr Asn Phe Tyr Ala Glu Tyr Asp Gly Tyr Tyr Lys Asp Phe Glu Phe
580 585 590

Ile Lys Tyr Tyr Asn Glu Phe Arg Asn Phe Ile Thr Lys Lys Pro Phe
595 600 605

Asp Glu Asp Lys Ile Lys Leu Asn Phe Glu Asn Gly Ala Leu Leu Lys
610 615 620

Gly Trp Asp Glu Asn Lys Glu Tyr Asp Phe Met Gly Val Ile Leu Lys
625 630 635 640

Lys Glu Gly Arg Leu Tyr Leu Gly Ile Met His Lys Asn His Arg Lys
645 650 655

Leu Phe Gln Ser Met Gly Asn Ala Lys Gly Asp Asn Ala Asn Arg Tyr
660 665 670

Gln Lys Met Ile Tyr Lys Gln Ile Ala Asp Ala Ser Lys Asp Val Pro
675 680 685

Arg Leu Leu Leu Thr Ser Lys Lys Ala Met Glu Lys Phe Lys Pro Ser
690 695 700

Gln Glu Ile Leu Arg Ile Lys Lys Glu Lys Thr Phe Lys Arg Glu Ser
705 710 715 720

Lys Asn Phe Ser Leu Arg Asp Leu His Ala Leu Ile Glu Tyr Tyr Arg
725 730 735

Asn Cys Ile Pro Gln Tyr Ser Asn Trp Ser Phe Tyr Asp Phe Gln Phe
740 745 750

Gln Asp Thr Gly Lys Tyr Gln Asn Ile Lys Glu Phe Thr Asp Asp Val
755 760 765

Gln Lys Tyr Gly Tyr Lys Ile Ser Phe Arg Asp Ile Asp Asp Glu Tyr

770 775 780

Ile Asn Gln Ala Leu Asn Glu Gly Lys Met Tyr Leu Phe Glu Val Val
785 790 795 800

Asn Lys Asp Ile Tyr Asn Thr Lys Asn Gly Ser Lys Asn Leu His Thr
805 810 815

Leu Tyr Phe Glu His Ile Leu Ser Ala Glu Asn Leu Asn Asp Pro Val
820 825 830

Phe Lys Leu Ser Gly Met Ala Glu Ile Phe Gln Arg Gln Pro Ser Val
835 840 845

Asn Glu Arg Glu Lys Ile Thr Thr Gln Lys Asn Gln Cys Ile Leu Asp
850 855 860

Lys Gly Asp Arg Ala Tyr Lys Tyr Arg Arg Tyr Thr Glu Lys Lys Ile
865 870 875 880

Met Phe His Met Ser Leu Val Leu Asn Thr Gly Lys Gly Glu Ile Lys
885 890 895

Gln Val Gln Phe Asn Lys Ile Ile Asn Gln Arg Ile Ser Ser Ser Asp
900 905 910

Asn Glu Met Arg Val Asn Val Ile Gly Ile Asp Arg Gly Glu Lys Asn
915 920 925

Leu Leu Tyr Tyr Ser Val Val Lys Gln Asn Gly Glu Ile Ile Glu Gln
930 935 940

Ala Ser Leu Asn Glu Ile Asn Gly Val Asn Tyr Arg Asp Lys Leu Ile
945 950 955 960

Glu Arg Glu Lys Glu Arg Leu Lys Asn Arg Gln Ser Trp Lys Pro Val
965 970 975

Val Lys Ile Lys Asp Leu Lys Lys Gly Tyr Ile Ser His Val Ile His
980 985 990

Lys Ile Cys Gln Leu Ile Glu Lys Tyr Ser Ala Ile Val Val Leu Glu
995 1000 1005

Asp Leu Asn Met Arg Phe Lys Gln Ile Arg Gly Gly Ile Glu Arg
1010 1015 1020

Ser Val Tyr Gln Gln Phe Glu Lys Ala Leu Ile Asp Lys Leu Gly
1025 1030 1035

Tyr Leu Val Phe Lys Asp Asn Arg Asp Leu Arg Ala Pro Gly Gly
1040 1045 1050

Val Leu Asn Gly Tyr Gln Leu Ser Ala Pro Phe Val Ser Phe Glu
1055 1060 1065

Lys Met Arg Lys Gln Thr Gly Ile Leu Phe Tyr Thr Gln Ala Glu
1070 1075 1080

Tyr Thr Ser Lys Thr Asp Pro Ile Thr Gly Phe Arg Lys Asn Val
1085 1090 1095

Tyr Ile Ser Asn Ser Ala Ser Leu Asp Lys Ile Lys Glu Ala Val

1100 1105 1110

Lys Lys Phe Asp Ala Ile Gly Trp Asp Gly Lys Glu Gln Ser Tyr
1115 1120 1125

Phe Phe Lys Tyr Asn Pro Tyr Asn Leu Ala Asp Glu Lys Tyr Lys
1130 1135 1140

Asn Ser Thr Val Ser Lys Glu Trp Ala Ile Phe Ala Ser Ala Pro
1145 1150 1155

Arg Ile Arg Arg Gln Lys Gly Glu Asp Gly Tyr Trp Lys Tyr Asp
1160 1165 1170

Arg Val Lys Val Asn Glu Glu Phe Glu Lys Leu Leu Lys Val Trp
1175 1180 1185

Asn Phe Val Asn Pro Lys Ala Thr Asp Ile Lys Gln Glu Ile Ile
1190 1195 1200

Lys Lys Glu Lys Ala Gly Asp Leu Gln Gly Glu Lys Glu Leu Asp
1205 1210 1215

Gly Arg Leu Arg Asn Phe Trp His Ser Phe Ile Tyr Leu Phe Asn
1220 1225 1230

Leu Val Leu Glu Leu Arg Asn Ser Phe Ser Leu Gln Ile Lys Ile
1235 1240 1245

Lys Ala Gly Glu Val Ile Ala Val Asp Glu Gly Val Asp Phe Ile
1250 1255 1260

Ala Ser Pro Val Lys Pro Phe Phe Thr Thr Pro Asn Pro Tyr Ile
1265 1270 1275

Pro Ser Asn Leu Cys Trp Leu Ala Val Glu Asn Ala Asp Ala Asn
1280 1285 1290

Gly Ala Tyr Asn Ile Ala Arg Lys Gly Val Met Ile Leu Lys Lys
1295 1300 1305

Ile Arg Glu His Ala Lys Lys Asp Pro Glu Phe Lys Lys Leu Pro
1310 1315 1320

Asn Leu Phe Ile Ser Asn Ala Glu Trp Asp Glu Ala Ala Arg Asp
1325 1330 1335

Trp Gly Lys Tyr Ala Gly Thr Thr Ala Leu Asn Leu Asp His
1340 1345 1350

<210> 1117
<211> 1331
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
Parcubacteria群細菌序列

<400> 1117
Met Lys Pro Val Gly Lys Thr Glu Asp Phe Leu Lys Ile Asn Lys Val
1 5 10 15

Phe Glu Lys Asp Gln Thr Ile Asp Asp Ser Tyr Asn Gln Ala Lys Phe
20 25 30

```
Tyr Phe Asp Ser Leu His Gln Lys Phe Ile Asp Ala Ala Leu Ala Ser
        35                  40                  45

Asp Lys Thr Ser Glu Leu Ser Phe Gln Asn Phe Ala Asp Val Leu Glu
    50                  55                  60

Lys Gln Asn Lys Ile Ile Leu Asp Lys Lys Arg Glu Met Gly Ala Leu
65                  70                  75                  80

Arg Lys Arg Asp Lys Asn Ala Val Gly Ile Asp Arg Leu Gln Lys Glu
                85                  90                  95

Ile Asn Asp Ala Glu Asp Ile Ile Gln Lys Glu Lys Glu Lys Ile Tyr
            100                 105                 110

Lys Asp Val Arg Thr Leu Phe Asp Asn Glu Ala Glu Ser Trp Lys Thr
        115                 120                 125

Tyr Tyr Gln Glu Arg Glu Val Asp Gly Lys Lys Ile Thr Phe Ser Lys
    130                 135                 140

Ala Asp Leu Lys Gln Lys Gly Ala Asp Phe Leu Thr Ala Ala Gly Ile
145                 150                 155                 160

Leu Lys Val Leu Lys Tyr Glu Phe Pro Glu Glu Lys Glu Lys Glu Phe
                165                 170                 175

Gln Ala Lys Asn Gln Pro Ser Leu Phe Val Glu Glu Lys Glu Asn Pro
            180                 185                 190

Gly Gln Lys Arg Tyr Ile Phe Asp Ser Phe Asp Lys Phe Ala Gly Tyr
        195                 200                 205

Leu Thr Lys Phe Gln Gln Thr Lys Lys Asn Leu Tyr Ala Ala Asp Gly
    210                 215                 220

Thr Ser Thr Ala Val Ala Thr Arg Ile Ala Asp Asn Phe Ile Ile Phe
225                 230                 235                 240

His Gln Asn Thr Lys Val Phe Arg Asp Lys Tyr Lys Asn Asn His Thr
                245                 250                 255

Asp Leu Gly Phe Asp Glu Glu Asn Ile Phe Glu Ile Glu Arg Tyr Lys
            260                 265                 270

Asn Cys Leu Leu Gln Arg Glu Ile Glu His Ile Lys Asn Glu Asn Ser
        275                 280                 285

Tyr Asn Lys Ile Ile Gly Arg Ile Asn Lys Lys Ile Lys Glu Tyr Arg
    290                 295                 300

Asp Gln Lys Ala Lys Asp Thr Lys Leu Thr Lys Ser Asp Phe Pro Phe
305                 310                 315                 320

Phe Lys Asn Leu Asp Lys Gln Ile Leu Gly Glu Val Glu Lys Glu Lys
                325                 330                 335

Gln Leu Ile Glu Lys Thr Arg Glu Lys Thr Glu Glu Asp Val Leu Ile
            340                 345                 350

Glu Arg Phe Lys Glu Phe Ile Glu Asn Asn Glu Glu Arg Phe Thr Ala
        355                 360                 365
```

```
Ala Lys Lys Leu Met Asn Ala Phe Cys Asn Gly Glu Phe Glu Ser Glu
    370                 375                 380

Tyr Glu Gly Ile Tyr Leu Lys Asn Lys Ala Ile Asn Thr Ile Ser Arg
385                 390                 395                 400

Arg Trp Phe Val Ser Asp Arg Asp Phe Glu Leu Lys Leu Pro Gln Gln
                405                 410                 415

Lys Ser Lys Asn Lys Ser Glu Lys Asn Glu Pro Lys Val Lys Lys Phe
            420                 425                 430

Ile Ser Ile Ala Glu Ile Lys Asn Ala Val Glu Glu Leu Asp Gly Asp
        435                 440                 445

Ile Phe Lys Ala Val Phe Tyr Asp Lys Lys Ile Ile Ala Gln Gly Gly
    450                 455                 460

Ser Lys Leu Glu Gln Phe Leu Val Ile Trp Lys Tyr Glu Phe Glu Tyr
465                 470                 475                 480

Leu Phe Arg Asp Ile Glu Arg Glu Asn Gly Glu Lys Leu Leu Gly Tyr
                485                 490                 495

Asp Ser Cys Leu Lys Ile Ala Lys Gln Leu Gly Ile Phe Pro Gln Glu
            500                 505                 510

Lys Glu Ala Arg Glu Lys Ala Thr Ala Val Ile Lys Asn Tyr Ala Asp
        515                 520                 525

Ala Gly Leu Gly Ile Phe Gln Met Met Lys Tyr Phe Ser Leu Asp Asp
    530                 535                 540

Lys Asp Arg Lys Asn Thr Pro Gly Gln Leu Ser Thr Asn Phe Tyr Ala
545                 550                 555                 560

Glu Tyr Asp Gly Tyr Tyr Lys Asp Phe Glu Phe Ile Lys Tyr Tyr Asn
                565                 570                 575

Glu Phe Arg Asn Phe Ile Thr Lys Lys Pro Phe Asp Glu Asp Lys Ile
            580                 585                 590

Lys Leu Asn Phe Glu Asn Gly Ala Leu Leu Lys Gly Trp Asp Glu Asn
        595                 600                 605

Lys Glu Tyr Asp Phe Met Gly Val Ile Leu Lys Lys Glu Gly Arg Leu
    610                 615                 620

Tyr Leu Gly Ile Met His Lys Asn His Arg Lys Leu Phe Gln Ser Met
625                 630                 635                 640

Gly Asn Ala Lys Gly Asp Asn Ala Asn Arg Tyr Gln Lys Met Ile Tyr
                645                 650                 655

Lys Gln Ile Ala Asp Ala Ser Lys Asp Val Pro Arg Leu Leu Leu Thr
            660                 665                 670

Ser Lys Lys Ala Met Glu Lys Phe Lys Pro Ser Gln Glu Ile Leu Arg
        675                 680                 685

Ile Lys Lys Glu Lys Thr Phe Lys Arg Glu Ser Lys Asn Phe Ser Leu
    690                 695                 700
```

```
Arg Asp Leu His Ala Leu Ile Glu Tyr Tyr Arg Asn Cys Ile Pro Gln
705                 710                 715                 720

Tyr Ser Asn Trp Ser Phe Tyr Asp Phe Gln Phe Gln Asp Thr Gly Lys
                725                 730                 735

Tyr Gln Asn Ile Lys Glu Phe Thr Asp Asp Val Gln Lys Tyr Gly Tyr
            740                 745                 750

Lys Ile Ser Phe Arg Asp Ile Asp Asp Glu Tyr Ile Asn Gln Ala Leu
        755                 760                 765

Asn Glu Gly Lys Met Tyr Leu Phe Glu Val Val Asn Lys Asp Ile Tyr
    770                 775                 780

Asn Thr Lys Asn Gly Ser Lys Asn Leu His Thr Leu Tyr Phe Glu His
785                 790                 795                 800

Ile Leu Ser Ala Glu Asn Leu Asn Asp Pro Val Phe Lys Leu Ser Gly
                805                 810                 815

Met Ala Glu Ile Phe Gln Arg Gln Pro Ser Val Asn Glu Arg Glu Lys
            820                 825                 830

Ile Thr Thr Gln Lys Asn Gln Cys Ile Leu Asp Lys Gly Asp Arg Ala
        835                 840                 845

Tyr Lys Tyr Arg Arg Tyr Thr Glu Lys Lys Ile Met Phe His Met Ser
    850                 855                 860

Leu Val Leu Asn Thr Gly Lys Gly Glu Ile Lys Gln Val Gln Phe Asn
865                 870                 875                 880

Lys Ile Ile Asn Gln Arg Ile Ser Ser Ser Asp Asn Glu Met Arg Val
                885                 890                 895

Asn Val Ile Gly Ile Asp Arg Gly Glu Lys Asn Leu Leu Tyr Tyr Ser
            900                 905                 910

Val Val Lys Gln Asn Gly Glu Ile Ile Glu Gln Ala Ser Leu Asn Glu
        915                 920                 925

Ile Asn Gly Val Asn Tyr Arg Asp Lys Leu Ile Glu Arg Glu Lys Glu
    930                 935                 940

Arg Leu Lys Asn Arg Gln Ser Trp Lys Pro Val Val Lys Ile Lys Asp
945                 950                 955                 960

Leu Lys Lys Gly Tyr Ile Ser His Val Ile His Lys Ile Cys Gln Leu
                965                 970                 975

Ile Glu Lys Tyr Ser Ala Ile Val Val Leu Glu Asp Leu Asn Met Arg
            980                 985                 990

Phe Lys Gln Ile Arg Gly Gly Ile Glu Arg Ser Val Tyr Gln Gln Phe
        995                 1000                1005

Glu Lys Ala Leu Ile Asp Lys Leu Gly Tyr Leu Val Phe Lys Asp
    1010                1015                1020

Asn Arg Asp Leu Arg Ala Pro Gly Gly Val Leu Asn Gly Tyr Gln
    1025                1030                1035
```

```
Leu Ser  Ala Pro Phe Val Ser  Phe Glu Lys Met Arg  Lys Gln Thr
    1040                1045                1050

Gly Ile  Leu Phe Tyr Thr Gln  Ala Glu Tyr Thr Ser  Lys Thr Asp
    1055                1060                1065

Pro Ile  Thr Gly Phe Arg Lys  Asn Val Tyr Ile Ser  Asn Ser Ala
    1070                1075                1080

Ser Leu  Asp Lys Ile Lys Glu  Ala Val Lys Lys Phe  Asp Ala Ile
    1085                1090                1095

Gly Trp  Asp Gly Lys Glu Gln  Ser Tyr Phe Phe Lys  Tyr Asn Pro
    1100                1105                1110

Tyr Asn  Leu Ala Asp Glu Lys  Tyr Lys Asn Ser Thr  Val Ser Lys
    1115                1120                1125

Glu Trp  Ala Ile Phe Ala Ser  Ala Pro Arg Ile Arg  Arg Gln Lys
    1130                1135                1140

Gly Glu  Asp Gly Tyr Trp Lys  Tyr Asp Arg Val Lys  Val Asn Glu
    1145                1150                1155

Glu Phe  Glu Lys Leu Leu Lys  Val Trp Asn Phe Val  Asn Pro Lys
    1160                1165                1170

Ala Thr  Asp Ile Lys Gln Glu  Ile Ile Lys Lys Glu  Lys Ala Gly
    1175                1180                1185

Asp Leu  Gln Gly Glu Lys Glu  Leu Asp Gly Arg Leu  Arg Asn Phe
    1190                1195                1200

Trp His  Ser Phe Ile Tyr Leu  Phe Asn Leu Val Leu  Glu Leu Arg
    1205                1210                1215

Asn Ser  Phe Ser Leu Gln Ile  Lys Ile Lys Ala Gly  Glu Val Ile
    1220                1225                1230

Ala Val  Asp Glu Gly Val Asp  Phe Ile Ala Ser Pro  Val Lys Pro
    1235                1240                1245

Phe Phe  Thr Thr Pro Asn Pro  Tyr Ile Pro Ser Asn  Leu Cys Trp
    1250                1255                1260

Leu Ala  Val Glu Asn Ala Asp  Ala Asn Gly Ala Tyr  Asn Ile Ala
    1265                1270                1275

Arg Lys  Gly Val Met Ile Leu  Lys Lys Ile Arg Glu  His Ala Lys
    1280                1285                1290

Lys Asp  Pro Glu Phe Lys Lys  Leu Pro Asn Leu Phe  Ile Ser Asn
    1295                1300                1305

Ala Glu  Trp Asp Glu Ala Ala  Arg Asp Trp Gly Lys  Tyr Ala Gly
    1310                1315                1320

Thr Thr  Ala Leu Asn Leu Asp  His
    1325                1330
```

<210> 1118
<211> 1285
<212> PRT

<213> 未知

<220>
<223> 未知的描述：
候選羅斯曼菌細菌序列

<400> 1118
Met Lys Ser Phe Asp Ser Phe Thr Asn Leu Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Lys Phe Glu Met Arg Pro Val Gly Asn Thr Gln Lys Met Leu Asp
20 25 30

Asn Ala Gly Val Phe Glu Lys Asp Lys Leu Ile Gln Lys Lys Tyr Gly
35 40 45

Lys Thr Lys Pro Tyr Phe Asp Arg Leu His Arg Glu Phe Ile Glu Glu
50 55 60

Ala Leu Thr Gly Val Glu Leu Ile Gly Leu Asp Glu Asn Phe Arg Thr
65 70 75 80

Leu Val Asp Trp Gln Lys Asp Lys Lys Asn Asn Val Ala Met Lys Ala
85 90 95

Tyr Glu Asn Ser Leu Gln Arg Leu Arg Thr Glu Ile Gly Lys Ile Phe
100 105 110

Asn Leu Lys Ala Glu Asp Trp Val Lys Asn Lys Tyr Pro Ile Leu Gly
115 120 125

Leu Lys Asn Lys Asn Thr Asp Ile Leu Phe Glu Glu Ala Val Phe Gly
130 135 140

Ile Leu Lys Ala Arg Tyr Gly Glu Glu Lys Asp Thr Phe Ile Glu Val
145 150 155 160

Glu Glu Ile Asp Lys Thr Gly Lys Ser Lys Ile Asn Gln Ile Ser Ile
165 170 175

Phe Asp Ser Trp Lys Gly Phe Thr Gly Tyr Phe Lys Lys Phe Phe Glu
180 185 190

Thr Arg Lys Asn Phe Tyr Lys Asn Asp Gly Thr Ser Thr Ala Ile Ala
195 200 205

Thr Arg Ile Ile Asp Gln Asn Leu Lys Arg Phe Ile Asp Asn Leu Ser
210 215 220

Ile Val Glu Ser Val Arg Gln Lys Val Asp Leu Ala Glu Thr Glu Lys
225 230 235 240

Ser Phe Ser Ile Ser Leu Ser Gln Phe Phe Ser Ile Asp Phe Tyr Asn
245 250 255

Lys Cys Leu Leu Gln Asp Gly Ile Asp Tyr Tyr Asn Lys Ile Ile Gly
260 265 270

Gly Glu Thr Leu Lys Asn Gly Glu Lys Leu Ile Gly Leu Asn Glu Leu
275 280 285

Ile Asn Gln Tyr Arg Gln Asn Asn Lys Asp Gln Lys Ile Pro Phe Phe
290 295 300

Lys Leu Leu Asp Lys Gln Ile Leu Ser Glu Lys Ile Leu Phe Leu Asp

305 310 315 320

Glu Ile Lys Asn Asp Thr Glu Leu Ile Glu Ala Leu Ser Gln Phe Ala
325 330 335

Lys Thr Ala Glu Glu Lys Thr Lys Ile Val Lys Lys Leu Phe Ala Asp
340 345 350

Phe Val Glu Asn Asn Ser Lys Tyr Asp Leu Ala Gln Ile Tyr Ile Ser
355 360 365

Gln Glu Ala Phe Asn Thr Ile Ser Asn Lys Trp Thr Ser Glu Thr Glu
370 375 380

Thr Phe Ala Lys Tyr Leu Phe Glu Ala Met Lys Ser Gly Lys Leu Ala
385 390 395 400

Lys Tyr Glu Lys Lys Asp Asn Ser Tyr Lys Phe Pro Asp Phe Ile Ala
405 410 415

Leu Ser Gln Met Lys Ser Ala Leu Leu Ser Ile Ser Leu Glu Gly His
420 425 430

Phe Trp Lys Glu Lys Tyr Tyr Lys Ile Ser Lys Phe Gln Glu Lys Thr
435 440 445

Asn Trp Glu Gln Phe Leu Ala Ile Phe Leu Tyr Glu Phe Asn Ser Leu
450 455 460

Phe Ser Asp Lys Ile Asn Thr Lys Asp Gly Glu Thr Lys Gln Val Gly
465 470 475 480

Tyr Tyr Leu Phe Ala Lys Asp Leu His Asn Leu Ile Leu Ser Glu Gln
485 490 495

Ile Asp Ile Pro Lys Asp Ser Lys Val Thr Ile Lys Asp Phe Ala Asp
500 505 510

Ser Val Leu Thr Ile Tyr Gln Met Ala Lys Tyr Phe Ala Val Glu Lys
515 520 525

Lys Arg Ala Trp Leu Ala Glu Tyr Glu Leu Asp Ser Phe Tyr Thr Gln
530 535 540

Pro Asp Thr Gly Tyr Leu Gln Phe Tyr Asp Asn Ala Tyr Glu Asp Ile
545 550 555 560

Val Gln Val Tyr Asn Lys Leu Arg Asn Tyr Leu Thr Lys Lys Pro Tyr
565 570 575

Ser Glu Glu Lys Trp Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn
580 585 590

Gly Trp Asp Lys Asn Lys Glu Ser Asp Asn Ser Ala Val Ile Leu Gln
595 600 605

Lys Gly Gly Lys Tyr Tyr Leu Gly Leu Ile Thr Lys Gly His Asn Lys
610 615 620

Ile Phe Asp Asp Arg Phe Gln Glu Lys Phe Ile Val Gly Ile Glu Gly
625 630 635 640

Gly Lys Tyr Glu Lys Ile Val Tyr Lys Phe Phe Pro Asp Gln Ala Lys

645 650 655

Met Phe Pro Lys Val Cys Phe Ser Ala Lys Gly Leu Glu Phe Phe Arg
660 665 670

Pro Ser Glu Glu Ile Leu Arg Ile Tyr Asn Asn Ala Glu Phe Lys Lys
675 680 685

Gly Glu Thr Tyr Ser Ile Asp Ser Met Gln Lys Leu Ile Asp Phe Tyr
690 695 700

Lys Asp Cys Leu Thr Lys Tyr Glu Gly Trp Ala Cys Tyr Thr Phe Arg
705 710 715 720

His Leu Lys Pro Thr Glu Glu Tyr Gln Asn Asn Ile Gly Glu Phe Phe
725 730 735

Arg Asp Val Ala Glu Asp Gly Tyr Arg Ile Asp Phe Gln Gly Ile Ser
740 745 750

Asp Gln Tyr Ile His Glu Lys Asn Glu Lys Gly Glu Leu His Leu Phe
755 760 765

Glu Ile His Asn Lys Asp Trp Asn Leu Asp Lys Ala Arg Asp Gly Lys
770 775 780

Ser Lys Thr Thr Gln Lys Asn Leu His Thr Leu Tyr Phe Glu Ser Leu
785 790 795 800

Phe Ser Asn Asp Asn Val Val Gln Asn Phe Pro Ile Lys Leu Asn Gly
805 810 815

Gln Ala Glu Ile Phe Tyr Arg Pro Lys Thr Glu Lys Asp Lys Leu Glu
820 825 830

Ser Lys Lys Asp Lys Lys Gly Asn Lys Val Ile Asp His Lys Arg Tyr
835 840 845

Ser Glu Asn Lys Ile Phe Phe His Val Pro Leu Thr Leu Asn Arg Thr
850 855 860

Lys Asn Asp Ser Tyr Arg Phe Asn Ala Gln Ile Asn Asn Phe Leu Ala
865 870 875 880

Asn Asn Lys Asp Ile Asn Ile Ile Gly Val Asp Arg Gly Glu Lys His
885 890 895

Leu Val Tyr Tyr Ser Val Ile Thr Gln Ala Ser Asp Ile Leu Glu Ser
900 905 910

Gly Ser Leu Asn Glu Leu Asn Gly Val Asn Tyr Ala Glu Lys Leu Gly
915 920 925

Lys Lys Ala Glu Asn Arg Glu Gln Ala Arg Arg Asp Trp Gln Asp Val
930 935 940

Gln Gly Ile Lys Asp Leu Lys Lys Gly Tyr Ile Ser Gln Val Val Arg
945 950 955 960

Lys Leu Ala Asp Leu Ala Ile Lys His Asn Ala Ile Ile Ile Leu Glu
965 970 975

Asp Leu Asn Met Arg Phe Lys Gln Val Arg Gly Gly Ile Glu Lys Ser

980 985 990

Ile Tyr Gln Gln Leu Glu Lys Ala Leu Ile Asp Lys Leu Ser Phe Leu
995 1000 1005

Val Asp Lys Gly Glu Lys Asn Pro Glu Gln Ala Gly His Leu Leu
1010 1015 1020

Lys Ala Tyr Gln Leu Ser Ala Pro Phe Glu Thr Phe Gln Lys Met
1025 1030 1035

Gly Lys Gln Thr Gly Ile Ile Phe Tyr Thr Gln Ala Ser Tyr Thr
1040 1045 1050

Ser Lys Ser Asp Pro Val Thr Gly Trp Arg Pro His Leu Tyr Leu
1055 1060 1065

Lys Tyr Phe Ser Ala Lys Lys Ala Lys Asp Asp Ile Ala Lys Phe
1070 1075 1080

Thr Lys Ile Glu Phe Val Asn Asp Arg Phe Glu Leu Thr Tyr Asp
1085 1090 1095

Ile Lys Asp Phe Gln Gln Ala Lys Glu Tyr Pro Asn Lys Thr Val
1100 1105 1110

Trp Lys Val Cys Ser Asn Val Glu Arg Phe Arg Trp Asp Lys Asn
1115 1120 1125

Leu Asn Gln Asn Lys Gly Gly Tyr Thr His Tyr Thr Asn Ile Thr
1130 1135 1140

Glu Asn Ile Gln Glu Leu Phe Thr Lys Tyr Gly Ile Asp Ile Thr
1145 1150 1155

Lys Asp Leu Leu Thr Gln Ile Ser Thr Ile Asp Glu Lys Gln Asn
1160 1165 1170

Thr Ser Phe Phe Arg Asp Phe Ile Phe Tyr Phe Asn Leu Ile Cys
1175 1180 1185

Gln Ile Arg Asn Thr Asp Asp Ser Glu Ile Ala Lys Lys Asn Gly
1190 1195 1200

Lys Asp Asp Phe Ile Leu Ser Pro Val Glu Pro Phe Phe Asp Ser
1205 1210 1215

Arg Lys Asp Asn Gly Asn Lys Leu Pro Glu Asn Gly Asp Asp Asn
1220 1225 1230

Gly Ala Tyr Asn Ile Ala Arg Lys Gly Ile Val Ile Leu Asn Lys
1235 1240 1245

Ile Ser Gln Tyr Ser Glu Lys Asn Glu Asn Cys Glu Lys Met Lys
1250 1255 1260

Trp Gly Asp Leu Tyr Val Ser Asn Ile Asp Trp Asp Asn Phe Val
1265 1270 1275

Thr Gln Ala Asn Ala Arg His
1280 1285

<210> 1119

<211> 1247
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
非培養的細菌序列

<400> 1119

```
Met Phe Lys Gly Asp Ala Phe Thr Gly Leu Tyr Glu Val Gln Lys Thr
1               5                  10                  15

Leu Arg Phe Glu Leu Val Pro Ile Gly Leu Thr Gln Ser Tyr Leu Glu
            20                  25                  30

Asn Asp Trp Val Ile Gln Lys Asp Lys Glu Val Glu Glu Asn Tyr Gly
        35                  40                  45

Lys Ile Lys Ala Tyr Phe Asp Leu Ile His Lys Glu Phe Val Arg Gln
    50                  55                  60

Ser Leu Glu Asn Ala Trp Leu Cys Gln Leu Asp Asp Phe Tyr Glu Lys
65                  70                  75                  80

Tyr Ile Glu Leu His Asn Ser Leu Glu Thr Arg Lys Asp Lys Asn Leu
                85                  90                  95

Ala Lys Gln Phe Glu Lys Val Met Lys Ser Leu Lys Lys Glu Phe Val
            100                 105                 110

Ser Phe Phe Asp Ala Lys Trp Asn Glu Trp Lys Gln Lys Phe Ser Phe
        115                 120                 125

Leu Lys Lys Trp Trp Ile Asp Val Leu Asn Glu Lys Glu Val Leu Asp
    130                 135                 140

Leu Met Ala Glu Phe Tyr Pro Asp Glu Lys Glu Leu Phe Asp Lys Phe
145                 150                 155                 160

Asp Lys Phe Phe Thr Tyr Phe Ser Asn Phe Lys Glu Ser Arg Lys Asn
                165                 170                 175

Phe Tyr Ala Asp Asp Gly Arg Ala Trp Ala Ile Ala Thr Arg Ala Ile
            180                 185                 190

Asp Glu Asn Leu Ile Thr Phe Ile Lys Asn Ile Glu Asp Phe Lys Lys
        195                 200                 205

Leu Asn Ser Ser Phe Arg Glu Phe Val Asn Asp Asn Phe Ser Glu Glu
    210                 215                 220

Asp Lys Gln Ile Phe Glu Ile Asp Phe Tyr Asn Asn Cys Leu Leu Gln
225                 230                 235                 240

Pro Trp Ile Asp Lys Tyr Asn Lys Ile Val Trp Trp Tyr Ser Leu Glu
                245                 250                 255

Asn Trp Glu Lys Val Gln Trp Leu Asn Glu Lys Ile Asn Asn Phe Lys
            260                 265                 270

Gln Asn Gln Asn Lys Ser Asn Ser Lys Asp Leu Lys Phe Pro Arg Met
        275                 280                 285

Lys Leu Leu Tyr Lys Gln Ile Leu Gly Asp Lys Glu Lys Lys Val Tyr
    290                 295                 300
```

```
Ile Asp Glu Ile Arg Asp Asp Lys Asn Leu Ile Asp Leu Ile Asp Asn
305                 310                 315                 320

Ser Lys Arg Arg Asn Gln Ile Lys Ile Asp Asn Ala Asn Asp Ile Ile
                325                 330                 335

Asn Asp Phe Ile Asn Asn Asn Ala Lys Phe Glu Leu Asp Lys Ile Tyr
            340                 345                 350

Leu Thr Arg Gln Ser Ile Asn Thr Ile Ser Ser Lys Tyr Phe Ser Ser
        355                 360                 365

Trp Asp Tyr Ile Arg Trp Tyr Phe Trp Thr Gly Glu Leu Gln Glu Phe
    370                 375                 380

Val Ser Phe Tyr Asp Leu Lys Glu Thr Phe Trp Lys Ile Glu Tyr Glu
385                 390                 395                 400

Thr Leu Glu Asn Ile Phe Lys Asp Cys Tyr Val Lys Gly Ile Asn Thr
                405                 410                 415

Glu Ser Gln Asn Asn Ile Val Phe Glu Thr Gln Gly Ile Tyr Glu Asn
            420                 425                 430

Phe Leu Asn Ile Phe Lys Phe Glu Phe Asn Gln Asn Ile Ser Gln Ile
        435                 440                 445

Ser Leu Leu Glu Trp Glu Leu Asp Lys Ile Gln Asn Glu Asp Ile Lys
    450                 455                 460

Lys Asn Glu Lys Gln Val Glu Val Ile Lys Asn Tyr Phe Asp Ser Val
465                 470                 475                 480

Met Ser Val Tyr Lys Met Thr Lys Tyr Phe Ser Leu Glu Lys Trp Lys
                485                 490                 495

Lys Arg Val Glu Leu Asp Thr Asp Asn Asn Phe Tyr Asn Asp Phe Asn
            500                 505                 510

Glu Tyr Leu Glu Gly Phe Glu Ile Trp Lys Asp Tyr Asn Leu Val Arg
        515                 520                 525

Asn Tyr Ile Thr Lys Lys Gln Val Asn Thr Asp Lys Ile Lys Leu Asn
    530                 535                 540

Phe Asp Asn Ser Gln Phe Leu Thr Trp Trp Asp Lys Asp Lys Glu Asn
545                 550                 555                 560

Glu Arg Leu Gly Ile Ile Leu Arg Arg Glu Trp Lys Tyr Tyr Leu Trp
                565                 570                 575

Ile Leu Lys Lys Trp Asn Thr Leu Asn Phe Gly Asp Tyr Leu Gln Lys
            580                 585                 590

Glu Trp Glu Ile Phe Tyr Glu Lys Met Asn Tyr Lys Gln Leu Asn Asn
        595                 600                 605

Val Tyr Arg Gln Leu Pro Arg Leu Leu Phe Pro Leu Thr Lys Lys Leu
    610                 615                 620

Asn Glu Leu Lys Trp Asp Glu Leu Lys Lys Tyr Leu Ser Lys Tyr Ile
625                 630                 635                 640
```

Gln Asn Phe Trp Tyr Asn Glu Glu Ile Ala Gln Ile Lys Ile Glu Phe
645 650 655

Asp Ile Phe Gln Glu Ser Lys Glu Lys Trp Glu Lys Phe Asp Ile Asp
660 665 670

Lys Leu Arg Lys Leu Ile Glu Tyr Tyr Lys Lys Trp Val Leu Ala Leu
675 680 685

Tyr Ser Asp Leu Tyr Asp Leu Glu Phe Ile Lys Tyr Lys Asn Tyr Asp
690 695 700

Asp Leu Ser Ile Phe Tyr Ser Asp Val Glu Lys Lys Met Tyr Asn Leu
705 710 715 720

Asn Phe Thr Lys Ile Asp Lys Ser Leu Ile Asp Gly Lys Val Lys Ser
725 730 735

Trp Glu Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Glu Ser
740 745 750

Lys Lys Glu Trp Ser Thr Glu Asn Ile His Thr Lys Tyr Phe Lys Leu
755 760 765

Leu Phe Asn Glu Lys Asn Leu Gln Asn Leu Val Val Lys Leu Ser Trp
770 775 780

Trp Ala Asp Ile Phe Phe Arg Asp Lys Thr Glu Asn Leu Lys Phe Lys
785 790 795 800

Lys Asp Lys Asn Gly Gln Glu Ile Leu Asp His Arg Arg Phe Ser Gln
805 810 815

Asp Lys Ile Met Phe His Ile Ser Ile Thr Leu Asn Ala Asn Cys Trp
820 825 830

Asp Lys Tyr Trp Phe Asn Gln Tyr Val Asn Glu Tyr Met Asn Lys Glu
835 840 845

Arg Asp Ile Lys Ile Ile Trp Ile Asp Arg Trp Glu Lys His Leu Ala
850 855 860

Tyr Tyr Cys Val Ile Asp Lys Ser Trp Lys Ile Phe Asn Asn Glu Ile
865 870 875 880

Trp Thr Leu Asn Glu Leu Asn Trp Val Asn Tyr Leu Glu Lys Leu Glu
885 890 895

Lys Ile Glu Ser Ser Arg Lys Asp Ser Arg Ile Ser Trp Trp Glu Ile
900 905 910

Glu Asn Ile Lys Glu Leu Lys Asn Gly Tyr Ile Ser Gln Val Ile Asn
915 920 925

Lys Leu Thr Glu Leu Ile Val Lys Tyr Asn Ala Ile Ile Val Phe Glu
930 935 940

Asp Leu Asn Ile Trp Phe Lys Arg Trp Arg Gln Lys Ile Glu Lys Gln
945 950 955 960

Ile Tyr Gln Lys Leu Glu Leu Ala Leu Ala Lys Lys Leu Asn Tyr Leu
965 970 975

Thr Gln Lys Asp Lys Lys Asp Asp Glu Ile Leu Trp Asn Leu Lys Ala
980 985 990

Leu Gln Leu Val Pro Lys Val Asn Asp Tyr Gln Asp Ile Trp Asn Tyr
995 1000 1005

Lys Gln Ser Trp Ile Met Phe Tyr Val Arg Ala Asn Tyr Thr Ser
1010 1015 1020

Val Thr Cys Pro Asn Cys Trp Leu Arg Lys Asn Leu Tyr Ile Ser
1025 1030 1035

Asn Ser Ala Thr Lys Glu Asn Gln Lys Lys Ser Leu Asn Ser Ile
1040 1045 1050

Ala Ile Lys Tyr Asn Asp Trp Lys Phe Ser Phe Ser Tyr Glu Ile
1055 1060 1065

Asp Asp Lys Ser Trp Lys Gln Lys Gln Ser Leu Asn Lys Lys Lys
1070 1075 1080

Phe Ile Val Tyr Ser Asp Ile Glu Arg Phe Val Tyr Ser Pro Leu
1085 1090 1095

Glu Lys Leu Thr Lys Val Ile Asp Val Asn Lys Lys Leu Leu Glu
1100 1105 1110

Leu Phe Arg Asp Phe Asn Leu Ser Leu Asp Ile Asn Lys Gln Ile
1115 1120 1125

Gln Glu Lys Asp Leu Asp Ser Val Phe Phe Lys Ser Leu Thr His
1130 1135 1140

Leu Phe Asn Leu Ile Leu Gln Leu Arg Asn Ser Asp Ser Lys Asp
1145 1150 1155

Asn Lys Asp Tyr Ile Ser Cys Pro Ser Cys Tyr Tyr His Ser Asn
1160 1165 1170

Asn Trp Leu Gln Trp Phe Glu Phe Asn Trp Asp Ala Asn Trp Ala
1175 1180 1185

Tyr Asn Ile Ala Arg Lys Gly Ile Ile Leu Leu Asp Arg Ile Arg
1190 1195 1200

Lys Asn Gln Glu Lys Pro Asp Leu Tyr Val Ser Asp Ile Asp Trp
1205 1210 1215

Asp Asn Phe Val Gln Ser Asn Gln Phe Pro Asn Thr Ile Ile Pro
1220 1225 1230

Ile Gln Asn Ile Glu Lys Gln Val Pro Leu Asn Ile Lys Ile
1235 1240 1245

<210> 1120
<211> 1219
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選分裂WS6細菌序列

<400> 1120
Met Lys Asn Val Phe Gly Gly Phe Thr Asn Leu Tyr Ser Leu Thr Lys

```
1               5                   10                  15

Thr Leu Arg Phe Glu Leu Lys Pro Thr Ser Lys Thr Gln Lys Leu Met
            20                  25                  30

Lys Arg Asn Asn Val Ile Gln Thr Asp Glu Glu Ile Asp Lys Leu Tyr
        35                  40                  45

His Asp Glu Met Lys Pro Ile Leu Asp Glu Ile His Arg Arg Phe Ile
    50                  55                  60

Asn Asp Ala Leu Ala Gln Lys Ile Phe Ile Ser Ala Ser Leu Asp Asn
65                  70                  75                  80

Phe Leu Lys Val Val Lys Asn Tyr Lys Val Glu Ser Ala Lys Lys Asn
                85                  90                  95

Ile Lys Gln Asn Gln Val Lys Leu Leu Gln Lys Glu Ile Thr Ile Lys
            100                 105                 110

Thr Leu Gly Leu Arg Arg Glu Val Val Ser Gly Phe Ile Thr Val Ser
        115                 120                 125

Lys Lys Trp Lys Asp Lys Tyr Val Gly Leu Gly Ile Lys Leu Lys Gly
    130                 135                 140

Asp Gly Tyr Lys Val Leu Thr Glu Gln Ala Val Leu Asp Ile Leu Lys
145                 150                 155                 160

Ile Glu Phe Pro Asn Lys Ala Lys Tyr Ile Asp Lys Phe Arg Gly Phe
                165                 170                 175

Trp Thr Tyr Phe Ser Gly Phe Asn Glu Asn Arg Lys Asn Tyr Tyr Ser
            180                 185                 190

Glu Glu Asp Lys Ala Thr Ser Ile Ala Asn Arg Ile Val Asn Glu Asn
        195                 200                 205

Leu Ser Arg Tyr Ile Asp Asn Ile Ile Ala Phe Glu Glu Ile Leu Gln
    210                 215                 220

Lys Ile Pro Asn Leu Lys Lys Phe Lys Gln Asp Leu Asp Ile Thr Ser
225                 230                 235                 240

Tyr Asn Tyr Tyr Leu Asn Gln Ala Gly Ile Asp Lys Tyr Asn Lys Ile
                245                 250                 255

Ile Gly Gly Tyr Ile Val Asp Lys Asp Lys Lys Ile Gln Gly Ile Asn
            260                 265                 270

Glu Lys Val Asn Leu Tyr Thr Gln Gln Thr Lys Lys Lys Leu Pro Lys
        275                 280                 285

Leu Lys Phe Leu Phe Lys Gln Ile Gly Ser Glu Arg Lys Gly Phe Gly
    290                 295                 300

Ile Phe Glu Ile Lys Glu Gly Lys Glu Trp Glu Gln Leu Gly Asp Leu
305                 310                 315                 320

Phe Lys Leu Gln Arg Thr Lys Ile Asn Ser Asn Gly Arg Glu Lys Gly
                325                 330                 335

Leu Phe Asp Ser Leu Arg Thr Met Tyr Arg Glu Phe Phe Asp Glu Ile
```

```
            340                 345                 350

Lys Arg Asp Ser Asn Ser Gln Ala Arg Tyr Ser Leu Asp Lys Ile Tyr
        355                 360                 365

Phe Asn Lys Ala Ser Val Asn Thr Ile Ser Asn Ser Trp Phe Thr Asn
    370                 375                 380

Trp Asn Lys Phe Ala Glu Leu Leu Asn Ile Lys Glu Asp Lys Lys Asn
385                 390                 395                 400

Gly Glu Lys Lys Ile Pro Glu Gln Ile Ser Ile Glu Asp Ile Lys Asp
                405                 410                 415

Ser Leu Ser Ile Ile Pro Lys Glu Asn Leu Glu Glu Leu Phe Lys Leu
            420                 425                 430

Thr Asn Arg Glu Lys His Asp Arg Thr Arg Phe Phe Gly Ser Asn Ala
        435                 440                 445

Trp Val Thr Phe Leu Asn Ile Trp Gln Asn Glu Ile Glu Glu Ser Phe
    450                 455                 460

Asn Lys Leu Glu Glu Lys Glu Lys Asp Phe Lys Lys Asn Ala Ala Ile
465                 470                 475                 480

Lys Phe Gln Lys Asn Asn Leu Val Gln Lys Asn Tyr Ile Lys Glu Val
                485                 490                 495

Cys Asp Arg Met Leu Ala Ile Glu Arg Met Ala Lys Tyr His Leu Pro
            500                 505                 510

Lys Asp Ser Asn Leu Ser Arg Glu Glu Asp Phe Tyr Trp Ile Ile Asp
        515                 520                 525

Asn Leu Ser Glu Gln Arg Glu Ile Tyr Lys Tyr Tyr Asn Ala Phe Arg
    530                 535                 540

Asn Tyr Ile Ser Lys Lys Pro Tyr Asn Lys Ser Lys Met Lys Leu Asn
545                 550                 555                 560

Phe Glu Asn Gly Asn Leu Leu Gly Gly Trp Ser Asp Gly Gln Glu Arg
                565                 570                 575

Asn Lys Ala Gly Val Ile Leu Arg Asn Gly Asn Lys Tyr Tyr Leu Gly
            580                 585                 590

Val Leu Ile Asn Arg Gly Ile Phe Arg Thr Asp Lys Ile Asn Asn Glu
        595                 600                 605

Ile Tyr Arg Thr Gly Ser Ser Lys Trp Glu Arg Leu Ile Leu Ser Asn
    610                 615                 620

Leu Lys Phe Gln Thr Leu Ala Gly Lys Gly Phe Leu Gly Lys His Gly
625                 630                 635                 640

Val Ser Tyr Gly Asn Met Asn Pro Glu Lys Ser Val Pro Ser Leu Gln
                645                 650                 655

Lys Phe Ile Arg Glu Asn Tyr Leu Lys Lys Tyr Pro Gln Leu Thr Glu
            660                 665                 670

Val Ser Asn Thr Lys Phe Leu Ser Lys Lys Asp Phe Asp Ala Ala Ile
```

675 680 685

Lys Glu Ala Leu Lys Glu Cys Phe Thr Met Asn Phe Ile Asn Ile Ala
690 695 700

Glu Asn Lys Leu Leu Glu Ala Glu Asp Lys Gly Asp Leu Tyr Leu Phe
705 710 715 720

Glu Ile Thr Asn Lys Asp Phe Ser Gly Lys Lys Ser Lys Gly Tyr Glu
725 730 735

Lys Gly Lys Asp Asn Ile His Thr Ile Tyr Trp Lys Tyr Leu Phe Ser
740 745 750

Glu Ser Asn Cys Lys Ser Pro Ile Ile Gly Leu Asn Gly Gly Ala Glu
755 760 765

Ile Phe Phe Arg Glu Gly Gln Lys Asp Lys Leu His Thr Lys Leu Asp
770 775 780

Lys Lys Gly Lys Lys Val Phe Asp Ala Lys Arg Tyr Ser Glu Asp Lys
785 790 795 800

Leu Phe Phe His Val Ser Ile Thr Ile Asn Tyr Gly Lys Pro Lys Asn
805 810 815

Ile Lys Phe Arg Asp Ile Ile Asn Gln Leu Ile Thr Ser Met Asn Val
820 825 830

Asn Ile Ile Gly Ile Asp Arg Gly Glu Lys His Leu Leu Tyr Tyr Ser
835 840 845

Val Ile Asp Ser Asn Gly Ile Ile Leu Lys Gln Gly Ser Leu Asn Lys
850 855 860

Ile Arg Val Gly Asp Lys Glu Val Asp Phe Asn Lys Lys Leu Thr Glu
865 870 875 880

Arg Ala Asn Glu Met Lys Lys Ala Arg Gln Ser Trp Glu Gln Ile Gly
885 890 895

Asn Ile Lys Asn Phe Lys Glu Gly Tyr Leu Ser Gln Ala Ile His Glu
900 905 910

Ile Tyr Gln Leu Met Ile Lys Tyr Asn Ala Ile Ile Val Leu Glu Asp
915 920 925

Leu Asn Thr Glu Phe Lys Ala Lys Arg Leu Ser Lys Val Glu Lys Ser
930 935 940

Val Tyr Lys Lys Phe Glu Leu Lys Leu Ala Arg Lys Leu Asn His Leu
945 950 955 960

Ile Leu Lys Asp Arg Asn Thr Asn Glu Ile Gly Gly Val Leu Lys Ala
965 970 975

Tyr Gln Leu Thr Pro Thr Ile Gly Gly Gly Asp Val Ser Lys Phe Glu
980 985 990

Lys Ala Lys Gln Trp Gly Met Met Phe Tyr Val Arg Ala Asn Tyr Thr
995 1000 1005

Ser Thr Thr Asp Pro Val Thr Gly Trp Arg Lys His Leu Tyr Ile

1010 1015 1020

Ser Asn Phe Ser Asn Asn Ser Val Ile Lys Ser Phe Phe Asp Pro
1025 1030 1035

Thr Asn Arg Asp Thr Gly Ile Glu Ile Phe Tyr Ser Gly Lys Tyr
1040 1045 1050

Arg Ser Trp Gly Phe Arg Tyr Val Gln Lys Glu Thr Gly Lys Lys
1055 1060 1065

Trp Glu Leu Phe Ala Thr Lys Glu Leu Glu Arg Phe Lys Tyr Asn
1070 1075 1080

Gln Thr Thr Lys Leu Cys Glu Lys Ile Asn Leu Tyr Asp Lys Phe
1085 1090 1095

Glu Glu Leu Phe Lys Gly Ile Asp Lys Ser Ala Asp Ile Tyr Ser
1100 1105 1110

Gln Leu Cys Asn Val Leu Asp Phe Arg Trp Lys Ser Leu Val Tyr
1115 1120 1125

Leu Trp Asn Leu Leu Asn Gln Ile Arg Asn Val Asp Lys Asn Ala
1130 1135 1140

Glu Gly Asn Lys Asn Asp Phe Ile Gln Ser Pro Val Tyr Pro Phe
1145 1150 1155

Phe Asp Ser Arg Lys Thr Asp Gly Lys Thr Glu Pro Ile Asn Gly
1160 1165 1170

Asp Ala Asn Gly Ala Leu Asn Ile Ala Arg Lys Gly Leu Met Leu
1175 1180 1185

Val Glu Arg Ile Lys Asn Asn Pro Glu Lys Tyr Glu Gln Leu Ile
1190 1195 1200

Arg Asp Thr Glu Trp Asp Ala Trp Ile Gln Asn Phe Asn Lys Val
1205 1210 1215

Asn

<210> 1121
<211> 1295
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 1121
Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
20 25 30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
35 40 45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
50 55 60

Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65 70 75 80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
85 90 95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
100 105 110

Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
115 120 125

Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
130 135 140

Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145 150 155 160

Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
165 170 175

Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190

Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205

Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
210 215 220

Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240

Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
245 250 255

Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
260 265 270

Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
275 280 285

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300

Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320

Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
325 330 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
340 345 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
355 360 365

Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
370 375 380

Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385 390 395 400

Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
405 410 415

```
Val Leu Glu Tyr Ile Thr Gln Gln Val Ala Pro Lys Asn Leu Asp Asn
            420                 425                 430

Pro Ser Lys Lys Glu Gln Asp Leu Ile Ala Lys Lys Thr Glu Lys Ala
        435                 440                 445

Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
    450                 455                 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465                 470                 475                 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
                485                 490                 495

Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
            500                 505                 510

Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
        515                 520                 525

Leu Leu Asp Gln Thr Asn Asn Leu Leu His Arg Leu Lys Ile Phe His
    530                 535                 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545                 550                 555                 560

Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
                565                 570                 575

Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
            580                 585                 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
        595                 600                 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
    610                 615                 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625                 630                 635                 640

Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
                645                 650                 655

Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
            660                 665                 670

Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
        675                 680                 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Asn Pro Gln
    690                 695                 700

Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe Ile Asp Phe Tyr Lys
705                 710                 715                 720

Glu Ser Ile Ser Lys His Pro Glu Trp Lys Asp Phe Gly Phe Arg Phe
                725                 730                 735

Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu Phe Tyr Arg Glu Val
            740                 745                 750
```

```
Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn Ile Ser Glu Ser Tyr
        755                 760                 765

Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr Leu Phe Gln Ile Tyr
    770                 775                 780

Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg Pro Asn Leu His Thr
785                 790                 795                 800

Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn Leu Gln Asp Val Val
                805                 810                 815

Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr Arg Lys Lys Ser Ile
            820                 825                 830

Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala Ile Ala Asn Lys Asn
        835                 840                 845

Lys Asp Asn Pro Lys Lys Glu Ser Phe Phe Glu Tyr Asp Leu Ile Lys
    850                 855                 860

Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe His Cys Pro Ile Thr
865                 870                 875                 880

Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe Asn Asp Glu Ile Asn
                885                 890                 895

Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His Ile Leu Ser Ile Asp
            900                 905                 910

Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu Val Asp Gly Lys Gly
        915                 920                 925

Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile Gly Asn Asp Arg Met
    930                 935                 940

Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile Glu Lys Asp Arg Asp
945                 950                 955                 960

Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn Ile Lys Glu Met Lys
                965                 970                 975

Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile Ala Lys Leu Val Ile
            980                 985                 990

Glu His Asn Ala Ile Val Val Phe Glu Asp Leu Asn Phe Gly Phe Lys
        995                 1000                1005

Arg Gly  Arg Phe Lys Val Glu Lys Gln Val Tyr  Gln Lys Leu Glu
    1010                1015                1020

Lys Met  Leu Ile Glu Lys Leu Asn Tyr Leu Val  Phe Lys Asp Asn
    1025                1030                1035

Glu Phe  Asp Lys Thr Gly Gly Val Leu Arg Ala  Tyr Gln Leu Thr
    1040                1045                1050

Ala Pro  Phe Glu Thr Phe Lys Lys Met Gly Lys  Gln Thr Gly Ile
    1055                1060                1065

Ile Tyr  Tyr Val Pro Ala Gly Phe Thr Ser Lys  Ile Cys Pro Val
    1070                1075                1080
```

Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys Tyr Glu Ser Val Ser
1085 1090 1095

Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp Lys Ile Cys Tyr Asn
1100 1105 1110

Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe Asp Tyr Lys Asn Phe
1115 1120 1125

Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr Ile Ala Ser Phe Gly
1130 1135 1140

Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp Lys Asn His Asn Trp
1145 1150 1155

Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu Leu Glu Lys Leu Leu
1160 1165 1170

Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly Glu Cys Ile Lys Ala
1175 1180 1185

Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe Phe Ala Lys Leu Thr
1190 1195 1200

Ser Ile Leu Asn Thr Ile Leu Gln Met Arg Asn Ser Lys Thr Gly
1205 1210 1215

Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val Ala Asp Val Asn Gly
1220 1225 1230

Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys Asn Met Pro Gln Asp
1235 1240 1245

Ala Asp Ala Asn Gly Ala Tyr His Ile Gly Leu Lys Gly Leu Met
1250 1255 1260

Leu Leu Asp Arg Ile Lys Asn Asn Gln Glu Gly Lys Lys Leu Asn
1265 1270 1275

Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu Phe Val Gln Asn Arg
1280 1285 1290

Asn Asn
1295

<210> 1122
<211> 1300
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 1122
Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
20 25 30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
35 40 45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
50 55 60

Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser

65 70 75 80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
85 90 95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
100 105 110

Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
115 120 125

Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
130 135 140

Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145 150 155 160

Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
165 170 175

Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190

Asp Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205

Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
210 215 220

Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240

Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
245 250 255

Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
260 265 270

Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
275 280 285

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300

Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320

Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
325 330 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
340 345 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
355 360 365

Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
370 375 380

Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385 390 395 400

Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala

405 410 415

Val Leu Glu Tyr Ile Thr Gln Gln Val Ala Pro Lys Asn Leu Asp Asn
420 425 430

Pro Ser Lys Lys Glu Gln Asp Leu Ile Ala Lys Lys Thr Glu Lys Ala
435 440 445

Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
450 455 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465 470 475 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
485 490 495

Asp Asn Leu Ala Gln Ile Ser Leu Lys Tyr Gln Asn Gln Gly Lys Lys
500 505 510

Asp Leu Leu Gln Ala Ser Ala Glu Glu Asp Val Lys Ala Ile Lys Asp
515 520 525

Leu Leu Asp Gln Thr Asn Asn Leu Leu His Arg Leu Lys Ile Phe His
530 535 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545 550 555 560

Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
565 570 575

Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
580 585 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
595 600 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
610 615 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625 630 635 640

Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
645 650 655

Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
660 665 670

Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
675 680 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Asn Pro Gln
690 695 700

Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
705 710 715 720

Ile Asp Phe Tyr Lys Glu Ser Ile Ser Lys His Pro Glu Trp Lys Asp
725 730 735

Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu

740 745 750

Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
755 760 765

Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
770 775 780

Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
785 790 795 800

Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
805 810 815

Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
820 825 830

Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
835 840 845

Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu
850 855 860

Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
865 870 875 880

His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
885 890 895

Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
900 905 910

Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
915 920 925

Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
930 935 940

Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945 950 955 960

Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
965 970 975

Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
980 985 990

Ala Lys Leu Val Ile Glu His Asn Ala Ile Val Val Phe Glu Asp Leu
995 1000 1005

Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val
1010 1015 1020

Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu
1025 1030 1035

Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg
1040 1045 1050

Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly
1055 1060 1065

Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser

1070 1075 1080

Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys
1085 1090 1095

Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp
1100 1105 1110

Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe
1115 1120 1125

Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr
1130 1135 1140

Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp
1145 1150 1155

Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu
1160 1165 1170

Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly
1175 1180 1185

Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe
1190 1195 1200

Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg
1205 1210 1215

Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val
1220 1225 1230

Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1235 1240 1245

Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly
1250 1255 1260

Leu Lys Gly Leu Met Leu Leu Asp Arg Ile Lys Asn Asn Gln Glu
1265 1270 1275

Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1280 1285 1290

Phe Val Gln Asn Arg Asn Asn
1295 1300

<210> 1123
<211> 1300
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 1123
Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
20 25 30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
35 40 45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
50 55 60

```
Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65                  70                  75                  80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
                85                  90                  95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
            100                 105                 110

Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
        115                 120                 125

Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
    130                 135                 140

Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145                 150                 155                 160

Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
                165                 170                 175

Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
            180                 185                 190

Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
        195                 200                 205

Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
    210                 215                 220

Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225                 230                 235                 240

Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
                245                 250                 255

Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
            260                 265                 270

Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
        275                 280                 285

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
    290                 295                 300

Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305                 310                 315                 320

Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
                325                 330                 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
            340                 345                 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
        355                 360                 365

Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
    370                 375                 380

Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385                 390                 395                 400
```

```
Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
                405                 410                 415

Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
            420                 425                 430

Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
        435                 440                 445

Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
    450                 455                 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465                 470                 475                 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
                485                 490                 495

Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
            500                 505                 510

Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
        515                 520                 525

Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
    530                 535                 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545                 550                 555                 560

Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
                565                 570                 575

Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
            580                 585                 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
        595                 600                 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
    610                 615                 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625                 630                 635                 640

Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
                645                 650                 655

Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
            660                 665                 670

Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
        675                 680                 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln
    690                 695                 700

Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
705                 710                 715                 720

Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp
                725                 730                 735
```

```
Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
            740                 745                 750

Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
        755                 760                 765

Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
    770                 775                 780

Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
785                 790                 795                 800

Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
                805                 810                 815

Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
            820                 825                 830

Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
        835                 840                 845

Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu
    850                 855                 860

Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
865                 870                 875                 880

His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
                885                 890                 895

Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
            900                 905                 910

Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
        915                 920                 925

Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
    930                 935                 940

Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945                 950                 955                 960

Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
                965                 970                 975

Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
            980                 985                 990

Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu
        995                 1000                1005

Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val
    1010                1015                1020

Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu
    1025                1030                1035

Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg
    1040                1045                1050

Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly
    1055                1060                1065
```

```
Lys Gln  Thr Gly Ile Ile Tyr  Tyr Val Pro Ala Gly  Phe Thr Ser
    1070                1075                1080

Lys Ile  Cys Pro Val Thr Gly  Phe Val Asn Gln Leu  Tyr Pro Lys
    1085                1090                1095

Tyr Glu  Ser Val Ser Lys Ser  Gln Glu Phe Phe Ser  Lys Phe Asp
    1100                1105                1110

Lys Ile  Cys Tyr Asn Leu Asp  Lys Gly Tyr Phe Glu  Phe Ser Phe
    1115                1120                1125

Asp Tyr  Lys Asn Phe Gly Asp  Lys Ala Ala Lys Gly  Lys Trp Thr
    1130                1135                1140

Ile Ala  Ser Phe Gly Ser Arg  Leu Ile Asn Phe Arg  Asn Ser Asp
    1145                1150                1155

Lys Asn  His Asn Trp Asp Thr  Arg Glu Val Tyr Pro  Thr Lys Glu
    1160                1165                1170

Leu Glu  Lys Leu Leu Lys Asp  Tyr Ser Ile Glu Tyr  Gly His Gly
    1175                1180                1185

Glu Cys  Ile Lys Ala Ala Ile  Cys Gly Glu Ser Asp  Lys Lys Phe
    1190                1195                1200

Phe Ala  Lys Leu Thr Ser Val  Leu Asn Thr Ile Leu  Gln Met Arg
    1205                1210                1215

Asn Ser  Lys Thr Gly Thr Glu  Leu Asp Tyr Leu Ile  Ser Pro Val
    1220                1225                1230

Ala Asp  Val Asn Gly Asn Phe  Phe Asp Ser Arg Gln  Ala Pro Lys
    1235                1240                1245

Asn Met  Pro Gln Asp Ala Asp  Ala Asn Gly Ala Tyr  His Ile Gly
    1250                1255                1260

Leu Lys  Gly Leu Met Leu Leu  Gly Arg Ile Lys Asn  Asn Gln Glu
    1265                1270                1275

Gly Lys  Lys Leu Asn Leu Val  Ile Lys Asn Glu Glu  Tyr Phe Glu
    1280                1285                1290

Phe Val  Gln Asn Arg Asn Asn
    1295                1300
```

<210> 1124
<211> 1307
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 1124

```
Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1               5                  10                  15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
            20                  25                  30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
        35                  40                  45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
```

50 55 60

Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65 70 75 80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
85 90 95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Lys Tyr
100 105 110

Ile Asn Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
115 120 125

Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
130 135 140

Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145 150 155 160

Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
165 170 175

Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190

Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205

Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
210 215 220

Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240

Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
245 250 255

Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
260 265 270

Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
275 280 285

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300

Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320

Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
325 330 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
340 345 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
355 360 365

Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
370 375 380

Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr

385 390 395 400

Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
405 410 415

Val Leu Glu Tyr Ile Thr Gln Gln Val Ala Pro Lys Asn Leu Asp Asn
420 425 430

Pro Ser Lys Lys Glu Gln Asp Leu Ile Ala Lys Lys Thr Glu Lys Ala
435 440 445

Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
450 455 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ser
465 470 475 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
485 490 495

Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
500 505 510

Asp Leu Leu Gln Ala Ser Ala Glu Glu Asp Val Lys Ala Ile Lys Asp
515 520 525

Leu Leu Asp Gln Thr Asn Asn Leu Leu His Arg Leu Lys Ile Phe His
530 535 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545 550 555 560

Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
565 570 575

Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
580 585 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Ser Gly
595 600 605

Trp Asp Lys Asn Lys Glu Ser Ala Asn Thr Ala Ile Leu Phe Ile Lys
610 615 620

Asp Asp Lys Tyr Tyr Leu Gly Ile Met Asp Lys Lys His Asn Lys Ile
625 630 635 640

Phe Ser Asp Lys Ala Ile Glu Glu Asn Lys Gly Glu Gly Tyr Lys Lys
645 650 655

Ile Val Tyr Lys Gln Ile Ala Asp Ala Ser Lys Asp Ile Gln Asn Leu
660 665 670

Met Ile Ile Asp Gly Lys Thr Val Cys Lys Lys Gly Arg Lys Asp Arg
675 680 685

Asn Gly Val Asn Arg Gln Leu Leu Ser Leu Lys Arg Lys His Leu Pro
690 695 700

Glu Asn Ile Tyr Arg Ile Lys Glu Thr Lys Ser Tyr Leu Lys Asn Glu
705 710 715 720

Ala Arg Phe Ser Arg Lys Asp Leu Tyr Asp Phe Ile Asp Tyr Tyr Lys

725 730 735

Asp Arg Leu Asp Tyr Tyr Asp Phe Glu Phe Glu Leu Lys Pro Ser Asn
740 745 750

Glu Tyr Ser Asp Phe Asn Asp Phe Thr Asn His Ile Gly Ser Gln Gly
755 760 765

Tyr Lys Leu Thr Phe Glu Asn Ile Ser Gln Asp Tyr Ile Asn Ser Leu
770 775 780

Val Asn Glu Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Ser Lys Asp Phe
785 790 795 800

Ser Ala Tyr Ser Lys Gly Arg Pro Asn Leu His Thr Leu Tyr Trp Lys
805 810 815

Ala Leu Phe Asp Glu Arg Asn Leu Gln Asp Val Val Tyr Lys Leu Asn
820 825 830

Gly Glu Ala Glu Leu Phe Tyr Arg Lys Gln Ser Ile Pro Lys Lys Ile
835 840 845

Thr His Pro Ala Lys Glu Thr Ile Ala Asn Lys Asn Lys Asp Asn Pro
850 855 860

Lys Lys Glu Ser Val Phe Glu Tyr Asp Leu Ile Lys Asp Lys Arg Phe
865 870 875 880

Thr Glu Asp Lys Phe Phe Phe His Cys Pro Ile Thr Ile Asn Phe Lys
885 890 895

Ser Ser Gly Ala Asn Lys Phe Asn Asp Glu Ile Asn Leu Leu Leu Lys
900 905 910

Glu Lys Ala Asn Asp Val His Ile Leu Ser Ile Asp Arg Gly Glu Arg
915 920 925

His Leu Ala Tyr Tyr Thr Leu Val Asp Gly Lys Gly Asn Ile Ile Lys
930 935 940

Gln Asp Asn Phe Asn Ile Ile Gly Asn Asp Arg Met Lys Thr Asn Tyr
945 950 955 960

His Asp Lys Leu Ala Ala Ile Glu Lys Asp Arg Asp Ser Ala Arg Lys
965 970 975

Asp Trp Lys Lys Ile Asn Asn Ile Lys Glu Met Lys Glu Gly Tyr Leu
980 985 990

Ser Gln Val Val His Glu Ile Ala Lys Leu Val Ile Glu Tyr Asn Ala
995 1000 1005

Ile Val Val Phe Glu Asp Leu Asn Phe Gly Phe Lys Arg Gly Arg
1010 1015 1020

Phe Lys Val Glu Lys Gln Val Tyr Gln Lys Leu Glu Lys Met Leu
1025 1030 1035

Ile Glu Lys Leu Asn Tyr Leu Val Phe Lys Asp Asn Glu Phe Asp
1040 1045 1050

Lys Thr Gly Gly Val Leu Arg Ala Tyr Gln Leu Thr Ala Pro Phe

1055 1060 1065

Glu Thr Phe Lys Lys Met Gly Lys Gln Thr Gly Ile Ile Tyr Tyr
1070 1075 1080

Val Pro Ala Gly Phe Thr Ser Lys Ile Cys Pro Val Thr Gly Phe
1085 1090 1095

Val Asn Gln Leu Tyr Pro Lys Tyr Glu Ser Val Ser Lys Ser Gln
1100 1105 1110

Glu Phe Phe Ser Lys Phe Asp Lys Ile Cys Tyr Asn Leu Asp Lys
1115 1120 1125

Gly Tyr Phe Glu Phe Ser Phe Asp Tyr Lys Asn Phe Gly Asp Lys
1130 1135 1140

Ala Ala Lys Gly Lys Trp Thr Ile Ala Ser Phe Gly Ser Arg Leu
1145 1150 1155

Ile Asn Phe Arg Asn Ser Asp Lys Asn His Asn Trp Asp Thr Arg
1160 1165 1170

Glu Val Tyr Pro Thr Lys Glu Leu Glu Lys Leu Leu Lys Asp Tyr
1175 1180 1185

Ser Ile Glu Tyr Gly His Gly Glu Cys Ile Lys Ala Ala Ile Cys
1190 1195 1200

Gly Glu Ser Asp Lys Lys Phe Phe Ala Lys Leu Thr Ser Val Leu
1205 1210 1215

Asn Thr Ile Leu Gln Met Arg Asn Ser Lys Thr Gly Thr Glu Leu
1220 1225 1230

Asp Tyr Leu Ile Ser Pro Val Ala Asp Val Asn Gly Asn Phe Phe
1235 1240 1245

Asp Ser Arg Gln Ala Pro Lys Asn Met Pro Gln Asp Ala Asp Ala
1250 1255 1260

Asn Gly Ala Tyr His Ile Gly Leu Lys Gly Leu Met Leu Leu Asp
1265 1270 1275

Arg Ile Lys Asn Asn Gln Glu Gly Lys Lys Leu Asn Leu Val Ile
1280 1285 1290

Lys Asn Glu Glu Tyr Phe Glu Phe Val Gln Asn Arg Asn Asn
1295 1300 1305

<210> 1125
<211> 1323
<212> PRT
<213> 解糖腺普雷沃菌

<400> 1125
Met Glu Asn Tyr Gln Glu Phe Thr Asn Leu Phe Gln Leu Asn Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Cys Glu Leu Leu Glu
20 25 30

Glu Gly Lys Ile Phe Ala Ser Gly Ser Phe Leu Glu Lys Asp Lys Val
35 40 45

```
Arg Ala Asp Asn Val Ser Tyr Val Lys Lys Glu Ile Asp Lys Lys His
    50                  55                  60

Lys Ile Phe Ile Glu Glu Thr Leu Ser Ser Phe Ser Ile Ser Asn Asp
65                  70                  75                  80

Leu Leu Lys Gln Tyr Phe Asp Cys Tyr Asn Glu Leu Lys Ala Phe Lys
                85                  90                  95

Lys Asp Cys Lys Ser Asp Glu Glu Glu Val Lys Lys Thr Ala Leu Arg
            100                 105                 110

Asn Lys Cys Thr Ser Ile Gln Arg Ala Met Arg Glu Ala Ile Ser Gln
        115                 120                 125

Ala Phe Leu Lys Ser Pro Gln Lys Lys Leu Leu Ala Ile Lys Asn Leu
    130                 135                 140

Ile Glu Asn Val Phe Lys Ala Asp Glu Asn Val Gln His Phe Ser Glu
145                 150                 155                 160

Phe Thr Ser Tyr Phe Ser Gly Phe Glu Thr Asn Arg Glu Asn Phe Tyr
                165                 170                 175

Ser Asp Glu Glu Lys Ser Thr Ser Ile Ala Tyr Arg Leu Val His Asp
            180                 185                 190

Asn Leu Pro Ile Phe Ile Lys Asn Ile Tyr Ile Phe Glu Lys Leu Lys
        195                 200                 205

Glu Gln Phe Asp Ala Lys Thr Leu Ser Glu Ile Phe Glu Asn Tyr Lys
    210                 215                 220

Leu Tyr Val Ala Gly Ser Ser Leu Asp Glu Val Phe Ser Leu Glu Tyr
225                 230                 235                 240

Phe Asn Asn Thr Leu Thr Gln Lys Gly Ile Asp Asn Tyr Asn Ala Val
                245                 250                 255

Ile Gly Lys Ile Val Lys Glu Asp Lys Gln Glu Ile Gln Gly Leu Asn
            260                 265                 270

Glu His Ile Asn Leu Tyr Asn Gln Lys His Lys Asp Arg Arg Leu Pro
        275                 280                 285

Phe Phe Ile Ser Leu Lys Lys Gln Ile Leu Ser Asp Arg Glu Ala Leu
    290                 295                 300

Ser Trp Leu Pro Asp Met Phe Lys Asn Asp Ser Glu Val Ile Lys Ala
305                 310                 315                 320

Leu Lys Gly Phe Tyr Ile Glu Asp Gly Phe Glu Asn Asn Val Leu Thr
                325                 330                 335

Pro Leu Ala Thr Leu Leu Ser Ser Leu Asp Lys Tyr Asn Leu Asn Gly
            340                 345                 350

Ile Phe Ile Arg Asn Asn Glu Ala Leu Ser Ser Leu Ser Gln Asn Val
        355                 360                 365

Tyr Arg Asn Phe Ser Ile Asp Glu Ala Ile Asp Ala Asn Ala Glu Leu
    370                 375                 380
```

```
Gln Thr Phe Asn Asn Tyr Glu Leu Ile Ala Asn Ala Leu Arg Ala Lys
385                 390                 395                 400

Ile Lys Lys Glu Thr Lys Gln Gly Arg Lys Ser Phe Glu Lys Tyr Glu
                405                 410                 415

Glu Tyr Ile Asp Lys Lys Val Lys Ala Ile Asp Ser Leu Ser Ile Gln
            420                 425                 430

Glu Ile Asn Glu Leu Val Glu Asn Tyr Val Ser Glu Phe Asn Ser Asn
        435                 440                 445

Ser Gly Asn Met Pro Arg Lys Val Glu Asp Tyr Phe Ser Leu Met Arg
    450                 455                 460

Lys Gly Asp Phe Gly Ser Asn Asp Leu Ile Glu Asn Ile Lys Thr Lys
465                 470                 475                 480

Leu Ser Ala Ala Glu Lys Leu Leu Gly Thr Lys Tyr Gln Glu Thr Ala
                485                 490                 495

Lys Asp Ile Phe Lys Lys Asp Glu Asn Ser Lys Leu Ile Lys Glu Leu
            500                 505                 510

Leu Asp Ala Thr Lys Gln Phe Gln His Phe Ile Lys Pro Leu Leu Gly
        515                 520                 525

Thr Gly Glu Glu Ala Asp Arg Asp Leu Val Phe Tyr Gly Asp Phe Leu
    530                 535                 540

Pro Leu Tyr Glu Lys Phe Glu Glu Leu Thr Leu Leu Tyr Asn Lys Val
545                 550                 555                 560

Arg Asn Arg Leu Thr Gln Lys Pro Tyr Ser Lys Asp Lys Ile Arg Leu
                565                 570                 575

Cys Phe Asn Lys Pro Lys Leu Met Thr Gly Trp Val Asp Ser Lys Thr
            580                 585                 590

Glu Lys Ser Asp Asn Gly Thr Gln Tyr Gly Gly Tyr Leu Phe Arg Lys
        595                 600                 605

Lys Asn Glu Ile Gly Glu Tyr Asp Tyr Phe Leu Gly Ile Ser Ser Lys
    610                 615                 620

Ala Gln Leu Phe Arg Lys Asn Glu Ala Val Ile Gly Asp Tyr Glu Arg
625                 630                 635                 640

Leu Asp Tyr Tyr Gln Pro Lys Ala Asn Thr Ile Tyr Gly Ser Ala Tyr
                645                 650                 655

Glu Gly Glu Asn Ser Tyr Lys Glu Asp Lys Lys Arg Leu Asn Lys Val
            660                 665                 670

Ile Ile Ala Tyr Ile Glu Gln Ile Lys Gln Thr Asn Ile Lys Lys Ser
        675                 680                 685

Ile Ile Glu Ser Ile Ser Lys Tyr Pro Asn Ile Ser Asp Asp Asp Lys
    690                 695                 700

Val Thr Pro Ser Ser Leu Leu Glu Lys Ile Lys Lys Val Ser Ile Asp
705                 710                 715                 720
```

```
Ser Tyr Asn Gly Ile Leu Ser Phe Lys Ser Phe Gln Ser Val Asn Lys
                725                 730                 735

Glu Val Ile Asp Asn Leu Leu Lys Thr Ile Ser Pro Leu Lys Asn Lys
            740                 745                 750

Ala Glu Phe Leu Asp Leu Ile Asn Lys Asp Tyr Gln Ile Phe Thr Glu
        755                 760                 765

Val Gln Ala Val Ile Asp Glu Ile Cys Lys Gln Lys Thr Phe Ile Tyr
    770                 775                 780

Phe Pro Ile Ser Asn Val Glu Leu Glu Lys Glu Met Gly Asp Lys Asp
785                 790                 795                 800

Lys Pro Leu Cys Leu Phe Gln Ile Ser Asn Lys Asp Leu Ser Phe Ala
                805                 810                 815

Lys Thr Phe Ser Ala Asn Leu Arg Lys Lys Arg Gly Ala Glu Asn Leu
            820                 825                 830

His Thr Met Leu Phe Lys Ala Leu Met Glu Gly Asn Gln Asp Asn Leu
        835                 840                 845

Asp Leu Gly Ser Gly Ala Ile Phe Tyr Arg Ala Lys Ser Leu Asp Gly
    850                 855                 860

Asn Lys Pro Thr His Pro Ala Asn Glu Ala Ile Lys Cys Arg Asn Val
865                 870                 875                 880

Ala Asn Lys Asp Lys Val Ser Leu Phe Thr Tyr Asp Ile Tyr Lys Asn
                885                 890                 895

Arg Arg Tyr Met Glu Asn Lys Phe Leu Phe His Leu Ser Ile Val Gln
            900                 905                 910

Asn Tyr Lys Ala Ala Asn Asp Ser Ala Gln Leu Asn Ser Ser Ala Thr
        915                 920                 925

Glu Tyr Ile Arg Lys Ala Asp Asp Leu His Ile Ile Gly Ile Asp Arg
    930                 935                 940

Gly Glu Arg Asn Leu Leu Tyr Tyr Ser Val Ile Asp Met Lys Gly Asn
945                 950                 955                 960

Ile Val Glu Gln Asp Ser Leu Asn Ile Ile Arg Asn Asn Asp Leu Glu
                965                 970                 975

Thr Asp Tyr His Asp Leu Leu Asp Lys Arg Glu Lys Glu Arg Lys Ala
            980                 985                 990

Asn Arg Gln Asn Trp Glu Ala Val Glu Gly Ile Lys Asp Leu Lys Lys
        995                 1000                1005

Gly Tyr Leu Ser Gln Ala Val His Gln Ile Ala Gln Leu Met Leu
    1010                1015                1020

Lys Tyr Asn Ala Ile Ile Ala Leu Glu Asp Leu Gly Gln Met Phe
    1025                1030                1035

Val Thr Arg Gly Gln Lys Ile Glu Lys Ala Val Tyr Gln Gln Phe
    1040                1045                1050
```

Glu Lys Ser Leu Val Asp Lys Leu Ser Tyr Leu Val Asp Lys Lys
1055 1060 1065

Arg Pro Tyr Asn Glu Leu Gly Gly Ile Leu Lys Ala Tyr Gln Leu
1070 1075 1080

Ala Ser Ser Ile Thr Lys Asn Asn Ser Asp Lys Gln Asn Gly Phe
1085 1090 1095

Leu Phe Tyr Val Pro Ala Trp Asn Thr Ser Lys Ile Asp Pro Val
1100 1105 1110

Thr Gly Phe Thr Asp Leu Leu Arg Pro Lys Ala Met Thr Ile Lys
1115 1120 1125

Glu Ala Gln Asp Phe Phe Gly Ala Phe Asp Asn Ile Ser Tyr Asn
1130 1135 1140

Asp Lys Gly Tyr Phe Glu Phe Glu Thr Asn Tyr Asp Lys Phe Lys
1145 1150 1155

Ile Arg Met Lys Ser Ala Gln Thr Arg Trp Thr Ile Cys Thr Phe
1160 1165 1170

Gly Asn Arg Ile Lys Arg Lys Lys Asp Lys Asn Tyr Trp Asn Tyr
1175 1180 1185

Glu Glu Val Glu Leu Thr Glu Glu Phe Lys Lys Leu Phe Lys Asp
1190 1195 1200

Ser Asn Ile Asp Tyr Glu Asn Cys Asn Leu Lys Glu Glu Ile Gln
1205 1210 1215

Asn Lys Asp Asn Arg Lys Phe Phe Asp Asp Leu Ile Lys Leu Leu
1220 1225 1230

Gln Leu Thr Leu Gln Met Arg Asn Ser Asp Asp Lys Gly Asn Asp
1235 1240 1245

Tyr Ile Ile Ser Pro Val Ala Asn Ala Glu Gly Gln Phe Phe Asp
1250 1255 1260

Ser Arg Asn Gly Asp Lys Lys Leu Pro Leu Asp Ala Asp Ala Asn
1265 1270 1275

Gly Ala Tyr Asn Ile Ala Arg Lys Gly Leu Trp Asn Ile Arg Gln
1280 1285 1290

Ile Lys Gln Thr Lys Asn Asp Lys Lys Leu Asn Leu Ser Ile Ser
1295 1300 1305

Ser Thr Glu Trp Leu Asp Phe Val Arg Glu Lys Pro Tyr Leu Lys
1310 1315 1320

<210> 1126
<211> 1323
<212> PRT
<213> 解糖腺普雷沃菌

<400> 1126
Met Glu Asn Tyr Gln Glu Phe Thr Asn Leu Phe Gln Leu Asn Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Cys Glu Leu Leu Glu

```
           20                  25                  30

Glu Gly Lys Ile Phe Ala Ser Gly Ser Phe Leu Glu Lys Asp Lys Val
        35                  40                  45

Arg Ala Asp Asn Val Ser Tyr Val Lys Lys Glu Ile Asp Lys Lys His
    50                  55                  60

Lys Ile Phe Ile Glu Glu Thr Leu Ser Ser Phe Ser Ile Ser Asn Asp
65                  70                  75                  80

Leu Leu Lys Gln Tyr Phe Asp Cys Tyr Asn Glu Leu Lys Ala Phe Lys
                85                  90                  95

Lys Asp Cys Lys Ser Asp Glu Glu Glu Val Lys Lys Thr Ala Leu Arg
            100                 105                 110

Asn Lys Cys Thr Ser Ile Gln Arg Ala Met Arg Glu Ala Ile Ser Gln
        115                 120                 125

Ala Phe Leu Lys Ser Pro Gln Lys Lys Leu Leu Ala Ile Lys Asn Leu
    130                 135                 140

Ile Glu Asn Val Phe Lys Ala Asp Glu Asn Val Gln His Phe Ser Glu
145                 150                 155                 160

Phe Thr Ser Tyr Phe Ser Gly Phe Glu Thr Asn Arg Glu Asn Phe Tyr
                165                 170                 175

Ser Asp Glu Glu Lys Ser Thr Ser Ile Ala Tyr Arg Leu Val His Asp
            180                 185                 190

Asn Leu Pro Ile Phe Ile Lys Asn Ile Tyr Ile Phe Glu Lys Leu Lys
        195                 200                 205

Glu Gln Phe Asp Ala Lys Thr Leu Ser Glu Ile Phe Glu Asn Tyr Lys
    210                 215                 220

Leu Tyr Val Ala Gly Ser Ser Leu Asp Glu Val Phe Ser Leu Glu Tyr
225                 230                 235                 240

Phe Asn Asn Thr Leu Thr Gln Lys Gly Ile Asp Asn Tyr Asn Ala Val
                245                 250                 255

Ile Gly Lys Ile Val Lys Glu Asp Lys Gln Glu Ile Gln Gly Leu Asn
            260                 265                 270

Glu His Ile Asn Leu Tyr Asn Gln Lys His Lys Asp Arg Arg Leu Pro
        275                 280                 285

Phe Phe Ile Ser Leu Lys Lys Gln Ile Leu Ser Asp Arg Glu Ala Leu
    290                 295                 300

Ser Trp Leu Pro Asp Met Phe Lys Asn Asp Ser Glu Val Ile Lys Ala
305                 310                 315                 320

Leu Lys Gly Phe Tyr Ile Glu Asp Gly Phe Glu Asn Asn Val Leu Thr
                325                 330                 335

Pro Leu Ala Thr Leu Leu Ser Ser Leu Asp Lys Tyr Asn Leu Asn Gly
            340                 345                 350

Ile Phe Ile Arg Asn Asn Glu Ala Leu Ser Ser Leu Ser Gln Asn Val
```

355 360 365

Tyr Arg Asn Phe Ser Ile Asp Glu Ala Ile Asp Ala Asn Ala Glu Leu
370 375 380

Gln Thr Phe Asn Asn Tyr Glu Leu Ile Ala Asn Ala Leu Arg Ala Lys
385 390 395 400

Ile Lys Lys Glu Thr Lys Gln Gly Arg Lys Ser Phe Glu Lys Tyr Glu
405 410 415

Glu Tyr Ile Asp Lys Lys Val Lys Ala Ile Asp Ser Leu Ser Ile Gln
420 425 430

Glu Ile Asn Glu Leu Val Glu Asn Tyr Val Ser Glu Phe Asn Ser Asn
435 440 445

Ser Gly Asn Met Pro Arg Lys Val Glu Asp Tyr Phe Ser Leu Met Arg
450 455 460

Lys Gly Asp Phe Gly Ser Asn Asp Leu Ile Glu Asn Ile Lys Thr Lys
465 470 475 480

Leu Ser Ala Ala Glu Lys Leu Leu Gly Thr Lys Tyr Gln Glu Thr Ala
485 490 495

Lys Asp Ile Phe Lys Lys Asp Glu Asn Ser Lys Leu Ile Lys Glu Leu
500 505 510

Leu Asp Ala Thr Lys Gln Phe Gln His Phe Ile Lys Pro Leu Leu Gly
515 520 525

Thr Gly Glu Glu Ala Asp Arg Asp Leu Val Phe Tyr Gly Asp Phe Leu
530 535 540

Pro Leu Tyr Glu Lys Phe Glu Glu Leu Thr Leu Leu Tyr Asn Lys Val
545 550 555 560

Arg Asn Arg Leu Thr Gln Lys Pro Tyr Ser Lys Asp Lys Ile Arg Leu
565 570 575

Cys Phe Asn Lys Pro Lys Leu Met Thr Gly Trp Val Asp Ser Lys Thr
580 585 590

Glu Lys Ser Asp Asn Gly Thr Gln Tyr Gly Gly Tyr Leu Phe Arg Lys
595 600 605

Lys Asn Glu Ile Gly Glu Tyr Asp Tyr Phe Leu Gly Ile Ser Ser Lys
610 615 620

Ala Gln Leu Phe Arg Lys Asn Glu Ala Val Ile Gly Asp Tyr Glu Arg
625 630 635 640

Leu Asp Tyr Tyr Gln Pro Lys Ala Asn Thr Ile Tyr Gly Ser Ala Tyr
645 650 655

Glu Gly Glu Asn Ser Tyr Lys Glu Asp Lys Lys Arg Leu Asn Lys Val
660 665 670

Ile Ile Ala Tyr Ile Glu Gln Ile Lys Gln Thr Asn Ile Lys Lys Ser
675 680 685

Ile Ile Glu Ser Ile Ser Lys Tyr Pro Asn Ile Ser Asp Asp Asp Lys

690 695 700

Val Thr Pro Ser Ser Leu Leu Glu Lys Ile Lys Lys Val Ser Ile Asp
705 710 715 720

Ser Tyr Asn Gly Ile Leu Ser Phe Lys Ser Phe Gln Ser Val Asn Lys
725 730 735

Glu Val Ile Asp Asn Leu Leu Lys Thr Ile Ser Pro Leu Lys Asn Lys
740 745 750

Ala Glu Phe Leu Asp Leu Ile Asn Lys Asp Tyr Gln Ile Phe Thr Glu
755 760 765

Val Gln Ala Val Ile Asp Glu Ile Cys Lys Gln Lys Thr Phe Ile Tyr
770 775 780

Phe Pro Ile Ser Asn Val Glu Leu Glu Lys Glu Met Gly Asp Lys Asp
785 790 795 800

Lys Pro Leu Cys Leu Phe Gln Ile Ser Asn Lys Asp Leu Ser Phe Ala
805 810 815

Lys Thr Phe Ser Ala Asn Leu Arg Lys Lys Arg Gly Ala Glu Asn Leu
820 825 830

His Thr Met Leu Phe Lys Ala Leu Met Glu Gly Asn Gln Asp Asn Leu
835 840 845

Asp Leu Gly Ser Gly Ala Ile Phe Tyr Arg Ala Lys Ser Leu Asp Gly
850 855 860

Asn Lys Pro Thr His Pro Ala Asn Glu Ala Ile Lys Cys Arg Asn Val
865 870 875 880

Ala Asn Lys Asp Lys Val Ser Leu Phe Thr Tyr Asp Ile Tyr Lys Asn
885 890 895

Arg Arg Tyr Met Glu Asn Lys Phe Leu Phe His Leu Ser Ile Val Gln
900 905 910

Asn Tyr Lys Ala Ala Asn Asp Ser Ala Gln Leu Asn Ser Ser Ala Thr
915 920 925

Glu Tyr Ile Arg Lys Ala Asp Asp Leu His Ile Ile Gly Ile Asp Arg
930 935 940

Gly Glu Arg Asn Leu Leu Tyr Tyr Ser Val Ile Asp Met Lys Gly Asn
945 950 955 960

Ile Val Glu Gln Asp Ser Leu Asn Ile Ile Arg Asn Asn Asp Leu Glu
965 970 975

Thr Asp Tyr His Asp Leu Leu Asp Lys Arg Glu Lys Glu Arg Lys Ala
980 985 990

Asn Arg Gln Asn Trp Glu Ala Val Glu Gly Ile Lys Asp Leu Lys Lys
995 1000 1005

Gly Tyr Leu Ser Gln Ala Val His Gln Ile Ala Gln Leu Met Leu
1010 1015 1020

Lys Tyr Asn Ala Ile Ile Ala Leu Glu Asp Leu Gly Gln Met Phe

1025 1030 1035

Val Thr Arg Gly Gln Lys Ile Glu Lys Ala Val Tyr Gln Gln Phe
1040 1045 1050

Glu Lys Ser Leu Val Asp Lys Leu Ser Tyr Leu Val Asp Lys Lys
1055 1060 1065

Arg Pro Tyr Asn Glu Leu Gly Gly Ile Leu Lys Ala Tyr Gln Leu
1070 1075 1080

Ala Ser Ser Ile Thr Lys Asn Asn Ser Asp Lys Gln Asn Gly Phe
1085 1090 1095

Leu Phe Tyr Val Pro Ala Trp Asn Thr Ser Lys Ile Asp Pro Val
1100 1105 1110

Thr Gly Phe Thr Asp Leu Leu Arg Pro Lys Ala Met Thr Ile Lys
1115 1120 1125

Glu Ala Gln Asp Phe Phe Gly Ala Phe Asp Asn Ile Ser Tyr Asn
1130 1135 1140

Asp Lys Gly Tyr Phe Glu Phe Glu Thr Asn Tyr Asp Lys Phe Lys
1145 1150 1155

Ile Arg Met Lys Ser Ala Gln Thr Arg Trp Thr Ile Cys Thr Phe
1160 1165 1170

Gly Asn Arg Ile Lys Arg Lys Lys Asp Lys Asn Tyr Trp Asn Tyr
1175 1180 1185

Glu Glu Val Glu Leu Thr Glu Glu Phe Lys Lys Leu Phe Lys Asp
1190 1195 1200

Ser Asn Ile Asp Tyr Glu Asn Cys Asn Leu Lys Glu Glu Ile Gln
1205 1210 1215

Asn Lys Asp Asn Arg Lys Phe Phe Asp Asp Leu Ile Lys Leu Leu
1220 1225 1230

Gln Leu Thr Leu Gln Met Arg Asn Ser Asp Asp Lys Gly Asn Asp
1235 1240 1245

Tyr Ile Ile Ser Pro Val Ala Asn Ala Glu Gly Gln Phe Phe Asp
1250 1255 1260

Ser Arg Asn Gly Asp Lys Lys Leu Pro Leu Asp Ala Asp Ala Asn
1265 1270 1275

Gly Ala Tyr Asn Ile Ala Arg Lys Gly Leu Trp Asn Ile Arg Gln
1280 1285 1290

Ile Lys Gln Thr Lys Asn Asp Lys Lys Leu Asn Leu Ser Ile Ser
1295 1300 1305

Ser Thr Glu Trp Leu Asp Phe Val Arg Glu Lys Pro Tyr Leu Lys
1310 1315 1320

<210> 1127
<211> 1246
<212> PRT
<213> 獼猴卟啉單胞菌

<400> 1127

```
Met Lys Thr Gln His Phe Phe Glu Asp Phe Thr Ser Leu Tyr Ser Leu
1               5                  10                  15

Ser Lys Thr Ile Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Leu Glu
            20                  25                  30

Asn Ile Lys Lys Asn Gly Leu Ile Arg Arg Asp Glu Gln Arg Leu Asp
        35                  40                  45

Asp Tyr Glu Lys Leu Lys Lys Val Ile Asp Glu Tyr His Glu Asp Phe
    50                  55                  60

Ile Ala Asn Ile Leu Ser Ser Phe Ser Phe Ser Glu Glu Ile Leu Gln
65                  70                  75                  80

Ser Tyr Ile Gln Asn Leu Ser Glu Ser Glu Ala Arg Ala Lys Ile Glu
                85                  90                  95

Lys Thr Met Arg Asp Thr Leu Ala Lys Ala Phe Ser Glu Asp Glu Arg
            100                 105                 110

Tyr Lys Ser Ile Phe Lys Lys Glu Leu Val Lys Lys Asp Ile Pro Val
        115                 120                 125

Trp Cys Pro Ala Tyr Lys Ser Leu Cys Lys Lys Phe Asp Asn Phe Thr
    130                 135                 140

Thr Ser Leu Val Pro Phe His Glu Asn Arg Lys Asn Leu Tyr Thr Ser
145                 150                 155                 160

Asn Glu Ile Thr Ala Ser Ile Pro Tyr Arg Ile Val His Val Asn Leu
                165                 170                 175

Pro Lys Phe Ile Gln Asn Ile Glu Ala Leu Cys Glu Leu Gln Lys Lys
            180                 185                 190

Met Gly Ala Asp Leu Tyr Leu Glu Met Met Glu Asn Leu Arg Asn Val
        195                 200                 205

Trp Pro Ser Phe Val Lys Thr Pro Asp Asp Leu Cys Asn Leu Lys Thr
    210                 215                 220

Tyr Asn His Leu Met Val Gln Ser Ser Ile Ser Glu Tyr Asn Arg Phe
225                 230                 235                 240

Val Gly Gly Tyr Ser Thr Glu Asp Gly Thr Lys His Gln Gly Ile Asn
                245                 250                 255

Glu Trp Ile Asn Ile Tyr Arg Gln Arg Asn Lys Glu Met Arg Leu Pro
            260                 265                 270

Gly Leu Val Phe Leu His Lys Gln Ile Leu Ala Lys Val Asp Ser Ser
        275                 280                 285

Ser Phe Ile Ser Asp Thr Leu Glu Asn Asp Asp Gln Val Phe Cys Val
    290                 295                 300

Leu Arg Gln Phe Arg Lys Leu Phe Trp Asn Thr Val Ser Ser Lys Glu
305                 310                 315                 320

Asp Asp Ala Ala Ser Leu Lys Asp Leu Phe Cys Gly Leu Ser Gly Tyr
                325                 330                 335
```

```
Asp Pro Glu Ala Ile Tyr Val Ser Asp Ala His Leu Ala Thr Ile Ser
            340                 345                 350

Lys Asn Ile Phe Asp Arg Trp Asn Tyr Ile Ser Asp Ala Ile Arg Arg
        355                 360                 365

Lys Thr Glu Val Leu Met Pro Arg Lys Lys Glu Ser Val Glu Arg Tyr
    370                 375                 380

Ala Glu Lys Ile Ser Lys Gln Ile Lys Lys Arg Gln Ser Tyr Ser Leu
385                 390                 395                 400

Ala Glu Leu Asp Asp Leu Leu Ala His Tyr Ser Glu Glu Ser Leu Pro
                405                 410                 415

Ala Gly Phe Ser Leu Leu Ser Tyr Phe Thr Ser Leu Gly Gly Gln Lys
            420                 425                 430

Tyr Leu Val Ser Asp Gly Glu Val Ile Leu Tyr Glu Glu Gly Ser Asn
        435                 440                 445

Ile Trp Asp Glu Val Leu Ile Ala Phe Arg Asp Leu Gln Val Ile Leu
    450                 455                 460

Asp Lys Asp Phe Thr Glu Lys Lys Leu Gly Lys Asp Glu Glu Ala Val
465                 470                 475                 480

Ser Val Ile Lys Lys Ala Leu Asp Ser Ala Leu Arg Leu Arg Lys Phe
                485                 490                 495

Phe Asp Leu Leu Ser Gly Thr Gly Ala Glu Ile Arg Arg Asp Ser Ser
            500                 505                 510

Phe Tyr Ala Leu Tyr Thr Asp Arg Met Asp Lys Leu Lys Gly Leu Leu
        515                 520                 525

Lys Met Tyr Asp Lys Val Arg Asn Tyr Leu Thr Lys Lys Pro Tyr Ser
    530                 535                 540

Ile Glu Lys Phe Lys Leu His Phe Asp Asn Pro Ser Leu Leu Ser Gly
545                 550                 555                 560

Trp Asp Lys Asn Lys Glu Leu Asn Asn Leu Ser Val Ile Phe Arg Gln
                565                 570                 575

Asn Gly Tyr Tyr Tyr Leu Gly Ile Met Thr Pro Lys Gly Lys Asn Leu
            580                 585                 590

Phe Lys Thr Leu Pro Lys Leu Gly Ala Glu Glu Met Phe Tyr Glu Lys
        595                 600                 605

Met Glu Tyr Lys Gln Ile Ala Glu Pro Met Leu Met Leu Pro Lys Val
    610                 615                 620

Phe Phe Pro Lys Lys Thr Lys Pro Ala Phe Ala Pro Asp Gln Ser Val
625                 630                 635                 640

Val Asp Ile Tyr Asn Lys Lys Thr Phe Lys Thr Gly Gln Lys Gly Phe
                645                 650                 655

Asn Lys Lys Asp Leu Tyr Arg Leu Ile Asp Phe Tyr Lys Glu Ala Leu
            660                 665                 670
```

```
Thr Val His Glu Trp Lys Leu Phe Asn Phe Ser Phe Ser Pro Thr Glu
        675                 680                 685

Gln Tyr Arg Asn Ile Gly Glu Phe Phe Asp Glu Val Arg Glu Gln Ala
    690                 695                 700

Tyr Lys Val Ser Met Val Asn Val Pro Ala Ser Tyr Ile Asp Glu Ala
705                 710                 715                 720

Val Glu Asn Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe
                725                 730                 735

Ser Pro Tyr Ser Lys Gly Ile Pro Asn Leu His Thr Leu Tyr Trp Lys
            740                 745                 750

Ala Leu Phe Ser Glu Gln Asn Gln Ser Arg Val Tyr Lys Leu Cys Gly
        755                 760                 765

Gly Gly Glu Leu Phe Tyr Arg Lys Ala Ser Leu His Met Gln Asp Thr
    770                 775                 780

Thr Val His Pro Lys Gly Ile Ser Ile His Lys Lys Asn Leu Asn Lys
785                 790                 795                 800

Lys Gly Glu Thr Ser Leu Phe Asn Tyr Asp Leu Val Lys Asp Lys Arg
                805                 810                 815

Phe Thr Glu Asp Lys Phe Phe Phe His Val Pro Ile Ser Ile Asn Tyr
            820                 825                 830

Lys Asn Lys Lys Ile Thr Asn Val Asn Gln Met Val Arg Asp Tyr Ile
        835                 840                 845

Ala Gln Asn Asp Asp Leu Gln Ile Ile Gly Ile Asp Arg Gly Glu Arg
    850                 855                 860

Asn Leu Leu Tyr Ile Ser Arg Ile Asp Thr Arg Gly Asn Leu Leu Glu
865                 870                 875                 880

Gln Phe Ser Leu Asn Val Ile Glu Ser Asp Lys Gly Asp Leu Arg Thr
                885                 890                 895

Asp Tyr Gln Lys Ile Leu Gly Asp Arg Glu Gln Glu Arg Leu Arg Arg
            900                 905                 910

Arg Gln Glu Trp Lys Ser Ile Glu Ser Ile Lys Asp Leu Lys Asp Gly
        915                 920                 925

Tyr Met Ser Gln Val Val His Lys Ile Cys Asn Met Val Val Glu His
    930                 935                 940

Lys Ala Ile Val Val Leu Glu Asn Leu Asn Leu Ser Phe Met Lys Gly
945                 950                 955                 960

Arg Lys Lys Val Glu Lys Ser Val Tyr Glu Lys Phe Glu Arg Met Leu
                965                 970                 975

Val Asp Lys Leu Asn Tyr Leu Val Val Asp Lys Lys Asn Leu Ser Asn
            980                 985                 990

Glu Pro Gly Gly Leu Tyr Ala Ala Tyr Gln Leu Thr Asn Pro Leu Phe
        995                 1000                1005
```

Ser Phe Glu Glu Leu His Arg Tyr Pro Gln Ser Gly Ile Leu Phe
1010 1015 1020

Phe Val Asp Pro Trp Asn Thr Ser Leu Thr Asp Pro Ser Thr Gly
1025 1030 1035

Phe Val Asn Leu Leu Gly Arg Ile Asn Tyr Thr Asn Val Gly Asp
1040 1045 1050

Ala Arg Lys Phe Phe Asp Arg Phe Asn Ala Ile Arg Tyr Asp Gly
1055 1060 1065

Lys Gly Asn Ile Leu Phe Asp Leu Asp Leu Ser Arg Phe Asp Val
1070 1075 1080

Arg Val Glu Thr Gln Arg Lys Leu Trp Thr Leu Thr Thr Phe Gly
1085 1090 1095

Ser Arg Ile Ala Lys Ser Lys Lys Ser Gly Lys Trp Met Val Glu
1100 1105 1110

Arg Ile Glu Asn Leu Ser Leu Cys Phe Leu Glu Leu Phe Glu Gln
1115 1120 1125

Phe Asn Ile Gly Tyr Arg Val Glu Lys Asp Leu Lys Lys Ala Ile
1130 1135 1140

Leu Ser Gln Asp Arg Lys Glu Phe Tyr Val Arg Leu Ile Tyr Leu
1145 1150 1155

Phe Asn Leu Met Met Gln Ile Arg Asn Ser Asp Gly Glu Glu Asp
1160 1165 1170

Tyr Ile Leu Ser Pro Ala Leu Asn Glu Lys Asn Leu Gln Phe Asp
1175 1180 1185

Ser Arg Leu Ile Glu Ala Lys Asp Leu Pro Val Asp Ala Asp Ala
1190 1195 1200

Asn Gly Ala Tyr Asn Val Ala Arg Lys Gly Leu Met Val Val Gln
1205 1210 1215

Arg Ile Lys Arg Gly Asp His Glu Ser Ile His Arg Ile Gly Arg
1220 1225 1230

Ala Gln Trp Leu Arg Tyr Val Gln Glu Gly Ile Val Glu
1235 1240 1245

<210> 1128
<211> 1257
<212> PRT
<213> 布氏普雷沃菌

<400> 1128
Met Gln Ile Asn Asn Leu Lys Ile Ile Tyr Met Lys Phe Thr Asp Phe
1 5 10 15

Thr Gly Leu Tyr Ser Leu Ser Lys Thr Leu Arg Phe Glu Leu Lys Pro
20 25 30

Ile Gly Lys Thr Leu Glu Asn Ile Lys Lys Ala Gly Leu Leu Glu Gln
35 40 45

Asp Gln His Arg Ala Asp Ser Tyr Lys Lys Val Lys Lys Ile Ile Asp

50 55 60

Glu Tyr His Lys Ala Phe Ile Glu Lys Ser Leu Ser Asn Phe Glu Leu
65 70 75 80

Lys Tyr Gln Ser Glu Asp Lys Leu Asp Ser Leu Glu Glu Tyr Leu Met
85 90 95

Tyr Tyr Ser Met Lys Arg Ile Glu Lys Thr Glu Lys Asp Lys Phe Ala
100 105 110

Lys Ile Gln Asp Asn Leu Arg Lys Gln Ile Ala Asp His Leu Lys Gly
115 120 125

Asp Glu Ser Tyr Lys Thr Ile Phe Ser Lys Asp Leu Ile Arg Lys Asn
130 135 140

Leu Pro Asp Phe Val Lys Ser Asp Glu Glu Arg Thr Leu Ile Lys Glu
145 150 155 160

Phe Lys Asp Phe Thr Thr Tyr Phe Lys Gly Phe Tyr Glu Asn Arg Glu
165 170 175

Asn Met Tyr Ser Ala Glu Asp Lys Ser Thr Ala Ile Ser His Arg Ile
180 185 190

Ile His Glu Asn Leu Pro Lys Phe Val Asp Asn Ile Asn Ala Phe Ser
195 200 205

Lys Ile Ile Leu Ile Pro Glu Leu Arg Glu Lys Leu Asn Gln Ile Tyr
210 215 220

Gln Asp Phe Glu Glu Tyr Leu Asn Val Glu Ser Ile Asp Glu Ile Phe
225 230 235 240

His Leu Asp Tyr Phe Ser Met Val Met Thr Gln Lys Gln Ile Glu Val
245 250 255

Tyr Asn Ala Ile Ile Gly Gly Lys Ser Thr Asn Asp Lys Lys Ile Gln
260 265 270

Gly Leu Asn Glu Tyr Ile Asn Leu Tyr Asn Gln Lys His Lys Asp Cys
275 280 285

Lys Leu Pro Lys Leu Lys Leu Leu Phe Lys Gln Ile Leu Ser Asp Arg
290 295 300

Ile Ala Ile Ser Trp Leu Pro Asp Asn Phe Lys Asp Asp Gln Glu Ala
305 310 315 320

Leu Asp Ser Ile Asp Thr Cys Tyr Lys Asn Leu Leu Asn Asp Gly Asn
325 330 335

Val Leu Gly Glu Gly Asn Leu Lys Leu Leu Leu Glu Asn Ile Asp Thr
340 345 350

Tyr Asn Leu Lys Gly Ile Phe Ile Arg Asn Asp Leu Gln Leu Thr Asp
355 360 365

Ile Ser Gln Lys Met Tyr Ala Ser Trp Asn Val Ile Gln Asp Ala Val
370 375 380

Ile Leu Asp Leu Lys Lys Gln Val Ser Arg Lys Lys Lys Glu Ser Ala

385 390 395 400

Glu Asp Tyr Asn Asp Arg Leu Lys Lys Leu Tyr Thr Ser Gln Glu Ser
405 410 415

Phe Ser Ile Gln Tyr Leu Asn Asp Cys Leu Arg Ala Tyr Gly Lys Thr
420 425 430

Glu Asn Ile Gln Asp Tyr Phe Ala Lys Leu Gly Ala Val Asn Asn Glu
435 440 445

His Glu Gln Thr Ile Asn Leu Phe Ala Gln Val Arg Asn Ala Tyr Thr
450 455 460

Ser Val Gln Ala Ile Leu Thr Thr Pro Tyr Pro Glu Asn Ala Asn Leu
465 470 475 480

Ala Gln Asp Lys Glu Thr Val Ala Leu Ile Lys Asn Leu Leu Asp Ser
485 490 495

Leu Lys Arg Leu Gln Arg Phe Ile Lys Pro Leu Leu Gly Lys Gly Asp
500 505 510

Glu Ser Asp Lys Asp Glu Arg Phe Tyr Gly Asp Phe Thr Pro Leu Trp
515 520 525

Glu Thr Leu Asn Gln Ile Thr Pro Leu Tyr Asn Met Val Arg Asn Tyr
530 535 540

Met Thr Arg Lys Pro Tyr Ser Gln Glu Lys Ile Lys Leu Asn Phe Glu
545 550 555 560

Asn Ser Thr Leu Leu Gly Gly Trp Asp Leu Asn Lys Glu His Asp Asn
565 570 575

Thr Ala Ile Ile Leu Arg Lys Asn Gly Leu Tyr Tyr Leu Ala Ile Met
580 585 590

Lys Lys Ser Ala Asn Lys Ile Phe Asp Lys Asp Lys Leu Asp Asn Ser
595 600 605

Gly Asp Cys Tyr Glu Lys Met Val Tyr Lys Leu Leu Pro Gly Ala Asn
610 615 620

Lys Met Leu Pro Lys Val Phe Phe Ser Lys Ser Arg Ile Asp Glu Phe
625 630 635 640

Lys Pro Ser Glu Asn Ile Ile Glu Asn Tyr Lys Lys Gly Thr His Lys
645 650 655

Lys Gly Ala Asn Phe Asn Leu Ala Asp Cys His Asn Leu Ile Asp Phe
660 665 670

Phe Lys Ser Ser Ile Ser Lys His Glu Asp Trp Ser Lys Phe Asn Phe
675 680 685

His Phe Ser Asp Thr Ser Ser Tyr Glu Asp Leu Ser Asp Phe Tyr Arg
690 695 700

Glu Val Glu Gln Gln Gly Tyr Ser Ile Ser Phe Cys Asp Val Ser Val
705 710 715 720

Glu Tyr Ile Asn Lys Met Val Glu Lys Gly Asp Leu Tyr Leu Phe Gln

725 730 735

Ile Tyr Asn Lys Asp Phe Ser Glu Phe Ser Lys Gly Thr Pro Asn Met
740 745 750

His Thr Leu Tyr Trp Asn Ser Leu Phe Ser Lys Glu Asn Leu Asn Asn
755 760 765

Ile Ile Tyr Lys Leu Asn Gly Gln Ala Glu Ile Phe Phe Arg Lys Lys
770 775 780

Ser Leu Asn Tyr Lys Arg Pro Thr His Pro Ala His Gln Ala Ile Lys
785 790 795 800

Asn Lys Asn Lys Cys Asn Glu Lys Lys Glu Ser Ile Phe Asp Tyr Asp
805 810 815

Leu Val Lys Asp Lys Arg Tyr Thr Val Asp Lys Phe Gln Phe His Val
820 825 830

Pro Ile Thr Met Asn Phe Lys Ser Thr Gly Asn Thr Asn Ile Asn Gln
835 840 845

Gln Val Ile Asp Tyr Leu Arg Thr Glu Asp Asp Thr His Ile Ile Gly
850 855 860

Ile Asp Arg Gly Glu Arg His Leu Leu Tyr Leu Val Val Ile Asp Ser
865 870 875 880

His Gly Lys Ile Val Glu Gln Phe Thr Leu Asn Glu Ile Val Asn Glu
885 890 895

Tyr Gly Gly Asn Ile Tyr Arg Thr Asn Tyr His Asp Leu Leu Asp Thr
900 905 910

Arg Glu Gln Asn Arg Glu Lys Ala Arg Glu Ser Trp Gln Thr Ile Glu
915 920 925

Asn Ile Lys Glu Leu Lys Glu Gly Tyr Ile Ser Gln Val Ile His Lys
930 935 940

Ile Thr Asp Leu Met Gln Lys Tyr His Ala Val Val Val Leu Glu Asp
945 950 955 960

Leu Asn Met Gly Phe Met Arg Gly Arg Gln Lys Val Glu Lys Gln Val
965 970 975

Tyr Gln Lys Phe Glu Glu Met Leu Ile Asn Lys Leu Asn Tyr Leu Val
980 985 990

Asn Lys Lys Ala Asp Gln Asn Ser Ala Gly Gly Leu Leu His Ala Tyr
995 1000 1005

Gln Leu Thr Ser Lys Phe Glu Ser Phe Gln Lys Leu Gly Lys Gln
1010 1015 1020

Ser Gly Phe Leu Phe Tyr Ile Pro Ala Trp Asn Thr Ser Lys Ile
1025 1030 1035

Asp Pro Val Thr Gly Phe Val Asn Leu Phe Asp Thr Arg Tyr Glu
1040 1045 1050

Ser Ile Asp Lys Ala Lys Ala Phe Phe Gly Lys Phe Asp Ser Ile

1055 1060 1065

Arg Tyr Asn Ala Asp Lys Asp Trp Phe Glu Phe Ala Phe Asp Tyr
1070 1075 1080

Asn Asn Phe Thr Thr Lys Ala Glu Gly Thr Arg Thr Asn Trp Thr
1085 1090 1095

Ile Cys Thr Tyr Gly Ser Arg Ile Arg Thr Phe Arg Asn Gln Ala
1100 1105 1110

Lys Asn Ser Gln Trp Asp Asn Glu Glu Ile Asp Leu Thr Lys Ala
1115 1120 1125

Tyr Lys Ala Phe Phe Ala Lys His Gly Ile Asn Ile Tyr Asp Asn
1130 1135 1140

Ile Lys Glu Ala Ile Ala Met Glu Thr Glu Lys Ser Phe Phe Glu
1145 1150 1155

Asp Leu Leu His Leu Leu Lys Leu Thr Leu Gln Met Arg Asn Ser
1160 1165 1170

Ile Thr Gly Thr Thr Thr Asp Tyr Leu Ile Ser Pro Val His Asp
1175 1180 1185

Ser Lys Gly Asn Phe Tyr Asp Ser Arg Ile Cys Asp Asn Ser Leu
1190 1195 1200

Pro Ala Asn Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala Arg Lys
1205 1210 1215

Gly Leu Met Leu Ile Gln Gln Ile Lys Asp Ser Thr Ser Ser Asn
1220 1225 1230

Arg Phe Lys Phe Ser Pro Ile Thr Asn Lys Asp Trp Leu Ile Phe
1235 1240 1245

Ala Gln Glu Lys Pro Tyr Leu Asn Asp
1250 1255

<210> 1129
<211> 1247
<212> PRT
<213> 布氏普雷沃菌

<400> 1129
Met Lys Phe Thr Asp Phe Thr Gly Leu Tyr Ser Leu Ser Lys Thr Leu
1 5 10 15

Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Leu Glu Asn Ile Lys Lys
20 25 30

Ala Gly Leu Leu Glu Gln Asp Gln His Arg Ala Asp Ser Tyr Lys Lys
35 40 45

Val Lys Lys Ile Ile Asp Glu Tyr His Lys Ala Phe Ile Glu Lys Ser
50 55 60

Leu Ser Asn Phe Glu Leu Lys Tyr Gln Ser Glu Asp Lys Leu Asp Ser
65 70 75 80

Leu Glu Glu Tyr Leu Met Tyr Tyr Ser Met Lys Arg Ile Glu Lys Thr
85 90 95

```
Glu Lys Asp Lys Phe Ala Lys Ile Gln Asp Asn Leu Arg Lys Gln Ile
            100                 105                 110

Ala Asp His Leu Lys Gly Asp Glu Ser Tyr Lys Thr Ile Phe Ser Lys
        115                 120                 125

Asp Leu Ile Arg Lys Asn Leu Pro Asp Phe Val Lys Ser Asp Glu Glu
    130                 135                 140

Arg Thr Leu Ile Lys Glu Phe Lys Asp Phe Thr Thr Tyr Phe Lys Gly
145                 150                 155                 160

Phe Tyr Glu Asn Arg Glu Asn Met Tyr Ser Ala Glu Asp Lys Ser Thr
                165                 170                 175

Ala Ile Ser His Arg Ile Ile His Glu Asn Leu Pro Lys Phe Val Asp
            180                 185                 190

Asn Ile Asn Ala Phe Ser Lys Ile Ile Leu Ile Pro Glu Leu Arg Glu
        195                 200                 205

Lys Leu Asn Gln Ile Tyr Gln Asp Phe Glu Glu Tyr Leu Asn Val Glu
    210                 215                 220

Ser Ile Asp Glu Ile Phe His Leu Asp Tyr Phe Ser Met Val Met Thr
225                 230                 235                 240

Gln Lys Gln Ile Glu Val Tyr Asn Ala Ile Ile Gly Gly Lys Ser Thr
                245                 250                 255

Asn Asp Lys Lys Ile Gln Gly Leu Asn Glu Tyr Ile Asn Leu Tyr Asn
            260                 265                 270

Gln Lys His Lys Asp Cys Lys Leu Pro Lys Leu Lys Leu Leu Phe Lys
        275                 280                 285

Gln Ile Leu Ser Asp Arg Ile Ala Ile Ser Trp Leu Pro Asp Asn Phe
    290                 295                 300

Lys Asp Asp Gln Glu Ala Leu Asp Ser Ile Asp Thr Cys Tyr Lys Asn
305                 310                 315                 320

Leu Leu Asn Asp Gly Asn Val Leu Gly Glu Gly Asn Leu Lys Leu Leu
                325                 330                 335

Leu Glu Asn Ile Asp Thr Tyr Asn Leu Lys Gly Ile Phe Ile Arg Asn
            340                 345                 350

Asp Leu Gln Leu Thr Asp Ile Ser Gln Lys Met Tyr Ala Ser Trp Asn
        355                 360                 365

Val Ile Gln Asp Ala Val Ile Leu Asp Leu Lys Lys Gln Val Ser Arg
    370                 375                 380

Lys Lys Lys Glu Ser Ala Glu Asp Tyr Asn Asp Arg Leu Lys Lys Leu
385                 390                 395                 400

Tyr Thr Ser Gln Glu Ser Phe Ser Ile Gln Tyr Leu Asn Asp Cys Leu
                405                 410                 415

Arg Ala Tyr Gly Lys Thr Glu Asn Ile Gln Asp Tyr Phe Ala Lys Leu
            420                 425                 430
```

```
Gly Ala Val Asn Asn Glu His Glu Gln Thr Ile Asn Leu Phe Ala Gln
        435                 440                 445

Val Arg Asn Ala Tyr Thr Ser Val Gln Ala Ile Leu Thr Thr Pro Tyr
    450                 455                 460

Pro Glu Asn Ala Asn Leu Ala Gln Asp Lys Glu Thr Val Ala Leu Ile
465                 470                 475                 480

Lys Asn Leu Leu Asp Ser Leu Lys Arg Leu Gln Arg Phe Ile Lys Pro
                485                 490                 495

Leu Leu Gly Lys Gly Asp Glu Ser Asp Lys Asp Glu Arg Phe Tyr Gly
            500                 505                 510

Asp Phe Thr Pro Leu Trp Glu Thr Leu Asn Gln Ile Thr Pro Leu Tyr
        515                 520                 525

Asn Met Val Arg Asn Tyr Met Thr Arg Lys Pro Tyr Ser Gln Glu Lys
    530                 535                 540

Ile Lys Leu Asn Phe Glu Asn Ser Thr Leu Leu Gly Gly Trp Asp Leu
545                 550                 555                 560

Asn Lys Glu His Asp Asn Thr Ala Ile Ile Leu Arg Lys Asn Gly Leu
                565                 570                 575

Tyr Tyr Leu Ala Ile Met Lys Lys Ser Ala Asn Lys Ile Phe Asp Lys
            580                 585                 590

Asp Lys Leu Asp Asn Ser Gly Asp Cys Tyr Glu Lys Met Val Tyr Lys
        595                 600                 605

Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val Phe Phe Ser Lys
    610                 615                 620

Ser Arg Ile Asp Glu Phe Lys Pro Ser Glu Asn Ile Ile Glu Asn Tyr
625                 630                 635                 640

Lys Lys Gly Thr His Lys Lys Gly Ala Asn Phe Asn Leu Ala Asp Cys
                645                 650                 655

His Asn Leu Ile Asp Phe Phe Lys Ser Ser Ile Ser Lys His Glu Asp
            660                 665                 670

Trp Ser Lys Phe Asn Phe His Phe Ser Asp Thr Ser Ser Tyr Glu Asp
        675                 680                 685

Leu Ser Asp Phe Tyr Arg Glu Val Glu Gln Gln Gly Tyr Ser Ile Ser
    690                 695                 700

Phe Cys Asp Val Ser Val Glu Tyr Ile Asn Lys Met Val Glu Lys Gly
705                 710                 715                 720

Asp Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Glu Phe Ser
                725                 730                 735

Lys Gly Thr Pro Asn Met His Thr Leu Tyr Trp Asn Ser Leu Phe Ser
            740                 745                 750

Lys Glu Asn Leu Asn Asn Ile Ile Tyr Lys Leu Asn Gly Gln Ala Glu
        755                 760                 765
```

```
Ile Phe Phe Arg Lys Lys Ser Leu Asn Tyr Lys Arg Pro Thr His Pro
    770                 775                 780

Ala His Gln Ala Ile Lys Asn Lys Asn Lys Cys Asn Glu Lys Lys Glu
785                 790                 795                 800

Ser Ile Phe Asp Tyr Asp Leu Val Lys Asp Lys Arg Tyr Thr Val Asp
                805                 810                 815

Lys Phe Gln Phe His Val Pro Ile Thr Met Asn Phe Lys Ser Thr Gly
            820                 825                 830

Asn Thr Asn Ile Asn Gln Gln Val Ile Asp Tyr Leu Arg Thr Glu Asp
        835                 840                 845

Asp Thr His Ile Ile Gly Ile Asp Arg Gly Glu Arg His Leu Leu Tyr
    850                 855                 860

Leu Val Val Ile Asp Ser His Gly Lys Ile Val Glu Gln Phe Thr Leu
865                 870                 875                 880

Asn Glu Ile Val Asn Glu Tyr Gly Gly Asn Ile Tyr Arg Thr Asn Tyr
                885                 890                 895

His Asp Leu Leu Asp Thr Arg Glu Gln Asn Arg Glu Lys Ala Arg Glu
            900                 905                 910

Ser Trp Gln Thr Ile Glu Asn Ile Lys Glu Leu Lys Glu Gly Tyr Ile
        915                 920                 925

Ser Gln Val Ile His Lys Ile Thr Asp Leu Met Gln Lys Tyr His Ala
    930                 935                 940

Val Val Val Leu Glu Asp Leu Asn Met Gly Phe Met Arg Gly Arg Gln
945                 950                 955                 960

Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu Glu Met Leu Ile Asn
                965                 970                 975

Lys Leu Asn Tyr Leu Val Asn Lys Lys Ala Asp Gln Asn Ser Ala Gly
            980                 985                 990

Gly Leu Leu His Ala Tyr Gln Leu Thr Ser Lys Phe Glu Ser Phe Gln
        995                 1000                1005

Lys Leu  Gly Lys Gln Ser Gly Phe Leu Phe Tyr  Ile Pro Ala Trp
    1010                1015                1020

Asn Thr  Ser Lys Ile Asp Pro Val Thr Gly Phe  Val Asn Leu Phe
    1025                1030                1035

Asp Thr  Arg Tyr Glu Ser Ile Asp Lys Ala Lys  Ala Phe Phe Gly
    1040                1045                1050

Lys Phe  Asp Ser Ile Arg Tyr Asn Ala Asp Lys  Asp Trp Phe Glu
    1055                1060                1065

Phe Ala  Phe Asp Tyr Asn Asn Phe Thr Thr Lys  Ala Glu Gly Thr
    1070                1075                1080

Arg Thr  Asn Trp Thr Ile Cys Thr Tyr Gly Ser  Arg Ile Arg Thr
    1085                1090                1095
```

```
Phe Arg  Asn Gln Ala Lys Asn  Ser Gln Trp Asp Asn  Glu Glu Ile
    1100                1105                1110

Asp Leu  Thr Lys Ala Tyr Lys  Ala Phe Phe Ala Lys  His Gly Ile
    1115                1120                1125

Asn Ile  Tyr Asp Asn Ile Lys  Glu Ala Ile Ala Met  Glu Thr Glu
    1130                1135                1140

Lys Ser  Phe Phe Glu Asp Leu  Leu His Leu Leu Lys  Leu Thr Leu
    1145                1150                1155

Gln Met  Arg Asn Ser Ile Thr  Gly Thr Thr Thr Asp  Tyr Leu Ile
    1160                1165                1170

Ser Pro  Val His Asp Ser Lys  Gly Asn Phe Tyr Asp  Ser Arg Ile
    1175                1180                1185

Cys Asp  Asn Ser Leu Pro Ala  Asn Ala Asp Ala Asn  Gly Ala Tyr
    1190                1195                1200

Asn Ile  Ala Arg Lys Gly Leu  Met Leu Ile Gln Gln  Ile Lys Asp
    1205                1210                1215

Ser Thr  Ser Ser Asn Arg Phe  Lys Phe Ser Pro Ile  Thr Asn Lys
    1220                1225                1230

Asp Trp  Leu Ile Phe Ala Gln  Glu Lys Pro Tyr Leu  Asn Asp
    1235                1240                1245
```

<210> 1130
<211> 1253
<212> PRT
<213> 易北普雷沃菌

<400> 1130

```
Met Asn Ile Lys Asn Phe Thr Gly Leu Tyr Pro Leu Ser Lys Thr Leu
1               5                  10                  15

Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Lys Glu Asn Ile Glu Lys
            20                  25                  30

Asn Gly Ile Leu Thr Lys Asp Glu Gln Arg Ala Lys Asp Tyr Leu Ile
        35                  40                  45

Val Lys Gly Phe Ile Asp Glu Tyr His Lys Gln Phe Ile Lys Asp Arg
    50                  55                  60

Leu Trp Asp Phe Lys Leu Pro Leu Glu Ser Glu Gly Glu Lys Asn Ser
65                  70                  75                  80

Leu Glu Glu Tyr Gln Glu Leu Tyr Glu Leu Thr Lys Arg Asn Asp Ala
                85                  90                  95

Gln Glu Ala Asp Phe Thr Glu Ile Lys Asp Asn Leu Arg Ser Ser Ile
            100                 105                 110

Thr Glu Gln Leu Thr Lys Ser Gly Ser Ala Tyr Asp Arg Ile Phe Lys
        115                 120                 125

Lys Glu Phe Ile Arg Glu Asp Leu Val Asn Phe Leu Glu Asp Glu Lys
    130                 135                 140

Asp Lys Asn Ile Val Lys Gln Phe Glu Asp Phe Thr Thr Tyr Phe Thr
```

145 150 155 160

Gly Phe Tyr Glu Asn Arg Lys Asn Met Tyr Ser Ser Glu Glu Lys Ser
165 170 175

Thr Ala Ile Ala Tyr Arg Leu Ile His Gln Asn Leu Pro Lys Phe Met
180 185 190

Asp Asn Met Arg Ser Phe Ala Lys Ile Ala Asn Ser Ser Val Ser Glu
195 200 205

His Phe Ser Asp Ile Tyr Glu Ser Trp Lys Glu Tyr Leu Asn Val Asn
210 215 220

Ser Ile Glu Glu Ile Phe Gln Leu Asp Tyr Phe Ser Glu Thr Leu Thr
225 230 235 240

Gln Pro His Ile Glu Val Tyr Asn Tyr Ile Ile Gly Lys Lys Val Leu
245 250 255

Glu Asp Gly Thr Glu Ile Lys Gly Ile Asn Glu Tyr Val Asn Leu Tyr
260 265 270

Asn Gln Gln Gln Lys Asp Lys Ser Lys Arg Leu Pro Phe Leu Val Pro
275 280 285

Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Lys Leu Ser Trp Ile Ala
290 295 300

Glu Glu Phe Asp Ser Asp Lys Lys Met Leu Ser Ala Ile Thr Glu Ser
305 310 315 320

Tyr Asn His Leu His Asn Val Leu Met Gly Asn Glu Asn Glu Ser Leu
325 330 335

Arg Asn Leu Leu Leu Asn Ile Lys Asp Tyr Asn Leu Glu Lys Ile Asn
340 345 350

Ile Thr Asn Asp Leu Ser Leu Thr Glu Ile Ser Gln Asn Leu Phe Gly
355 360 365

Arg Tyr Asp Val Phe Thr Asn Gly Ile Lys Asn Lys Leu Arg Val Leu
370 375 380

Thr Pro Arg Lys Lys Lys Glu Thr Asp Glu Asn Phe Glu Asp Arg Ile
385 390 395 400

Asn Lys Ile Phe Lys Thr Gln Lys Ser Phe Ser Ile Ala Phe Leu Asn
405 410 415

Lys Leu Pro Gln Pro Glu Met Glu Asp Gly Lys Pro Arg Asn Ile Glu
420 425 430

Asp Tyr Phe Ile Thr Gln Gly Ala Ile Asn Thr Lys Ser Ile Gln Lys
435 440 445

Glu Asp Ile Phe Ala Gln Ile Glu Asn Ala Tyr Glu Asp Ala Gln Val
450 455 460

Phe Leu Gln Ile Lys Asp Thr Asp Asn Lys Leu Ser Gln Asn Lys Thr
465 470 475 480

Ala Val Glu Lys Ile Lys Thr Leu Leu Asp Ala Leu Lys Glu Leu Gln

485 490 495

His Phe Ile Lys Pro Leu Leu Gly Ser Gly Glu Glu Asn Glu Lys Asp
500 505 510

Glu Leu Phe Tyr Gly Ser Phe Leu Ala Ile Trp Asp Glu Leu Asp Thr
515 520 525

Ile Thr Pro Leu Tyr Asn Lys Val Arg Asn Trp Leu Thr Arg Lys Pro
530 535 540

Tyr Ser Thr Glu Lys Ile Lys Leu Asn Phe Asp Asn Ala Gln Leu Leu
545 550 555 560

Gly Gly Trp Asp Val Asn Lys Glu His Asp Cys Ala Gly Ile Leu Leu
565 570 575

Arg Lys Asn Asp Ser Tyr Tyr Leu Gly Ile Ile Asn Lys Lys Thr Asn
580 585 590

His Ile Phe Asp Thr Asp Ile Thr Pro Ser Asp Gly Glu Cys Tyr Asp
595 600 605

Lys Ile Asp Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys
610 615 620

Val Phe Phe Ser Lys Ser Arg Ile Lys Glu Phe Glu Pro Ser Glu Ala
625 630 635 640

Ile Ile Asn Cys Tyr Lys Lys Gly Thr His Lys Lys Gly Lys Asn Phe
645 650 655

Asn Leu Thr Asp Cys His Arg Leu Ile Asn Phe Phe Lys Thr Ser Ile
660 665 670

Glu Lys His Glu Asp Trp Ser Lys Phe Gly Phe Lys Phe Ser Asp Thr
675 680 685

Glu Thr Tyr Glu Asp Ile Ser Gly Phe Tyr Arg Glu Val Glu Gln Gln
690 695 700

Gly Tyr Arg Leu Thr Ser His Pro Val Ser Ala Ser Tyr Ile His Ser
705 710 715 720

Leu Val Lys Glu Gly Lys Leu Tyr Leu Phe Gln Ile Trp Asn Lys Asp
725 730 735

Phe Ser Gln Phe Ser Lys Gly Thr Pro Asn Leu His Thr Leu Tyr Trp
740 745 750

Lys Met Leu Phe Asp Lys Arg Asn Leu Ser Asp Val Val Tyr Lys Leu
755 760 765

Asn Gly Gln Ala Glu Val Phe Tyr Arg Lys Ser Ser Ile Glu His Gln
770 775 780

Asn Arg Ile Ile His Pro Ala Gln His Pro Ile Thr Asn Lys Asn Glu
785 790 795 800

Leu Asn Lys Lys His Thr Ser Thr Phe Lys Tyr Asp Ile Ile Lys Asp
805 810 815

Arg Arg Tyr Thr Val Asp Lys Phe Gln Phe His Val Pro Ile Thr Ile

820 825 830

Asn Phe Lys Ala Thr Gly Gln Asn Asn Ile Asn Pro Ile Val Gln Glu
835 840 845

Val Ile Arg Gln Asn Gly Ile Thr His Ile Ile Gly Ile Asp Arg Gly
850 855 860

Glu Arg His Leu Leu Tyr Leu Ser Leu Ile Asp Leu Lys Gly Asn Ile
865 870 875 880

Ile Lys Gln Met Thr Leu Asn Glu Ile Ile Asn Glu Tyr Lys Gly Val
885 890 895

Thr Tyr Lys Thr Asn Tyr His Asn Leu Leu Glu Lys Arg Glu Lys Glu
900 905 910

Arg Thr Glu Ala Arg His Ser Trp Ser Ser Ile Glu Ser Ile Lys Glu
915 920 925

Leu Lys Asp Gly Tyr Met Ser Gln Val Ile His Lys Ile Thr Asp Met
930 935 940

Met Val Lys Tyr Asn Ala Ile Val Val Leu Glu Asp Leu Asn Gly Gly
945 950 955 960

Phe Met Arg Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe
965 970 975

Glu Lys Lys Leu Ile Asp Lys Leu Asn Tyr Leu Val Asp Lys Lys Leu
980 985 990

Asp Ala Asn Glu Val Gly Gly Val Leu Asn Ala Tyr Gln Leu Thr Asn
995 1000 1005

Lys Phe Glu Ser Phe Lys Lys Ile Gly Lys Gln Ser Gly Phe Leu
1010 1015 1020

Phe Tyr Ile Pro Ala Trp Asn Thr Ser Lys Ile Asp Pro Ile Thr
1025 1030 1035

Gly Phe Val Asn Leu Phe Asn Thr Arg Tyr Glu Ser Ile Lys Glu
1040 1045 1050

Thr Lys Val Phe Trp Ser Lys Phe Asp Ile Ile Arg Tyr Asn Lys
1055 1060 1065

Glu Lys Asn Trp Phe Glu Phe Val Phe Asp Tyr Asn Thr Phe Thr
1070 1075 1080

Thr Lys Ala Glu Gly Thr Arg Thr Lys Trp Thr Leu Cys Thr His
1085 1090 1095

Gly Thr Arg Ile Gln Thr Phe Arg Asn Pro Glu Lys Asn Ala Gln
1100 1105 1110

Trp Asp Asn Lys Glu Ile Asn Leu Thr Glu Ser Phe Lys Ala Leu
1115 1120 1125

Phe Glu Lys Tyr Lys Ile Asp Ile Thr Ser Asn Leu Lys Glu Ser
1130 1135 1140

Ile Met Gln Glu Thr Glu Lys Lys Phe Phe Gln Glu Leu His Asn

1145 1150 1155

Leu Leu His Leu Thr Leu Gln Met Arg Asn Ser Val Thr Gly Thr
1160 1165 1170

Asp Ile Asp Tyr Leu Ile Ser Pro Val Ala Asp Glu Asp Gly Asn
1175 1180 1185

Phe Tyr Asp Ser Arg Ile Asn Gly Lys Asn Phe Pro Glu Asn Ala
1190 1195 1200

Asp Ala Asn Gly Ala Tyr Asn Ile Ala Arg Lys Gly Leu Met Leu
1205 1210 1215

Ile Arg Gln Ile Lys Gln Ala Asp Pro Gln Lys Lys Phe Lys Phe
1220 1225 1230

Glu Thr Ile Thr Asn Lys Asp Trp Leu Lys Phe Ala Gln Asp Lys
1235 1240 1245

Pro Tyr Leu Lys Asp
1250

<210> 1131
<211> 1262
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
口腔類擬桿菌274序列

<400> 1131
Met Arg Lys Phe Asn Glu Phe Val Gly Leu Tyr Pro Ile Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Leu Glu His Ile Gln
20 25 30

Arg Asn Lys Leu Leu Glu His Asp Ala Val Arg Ala Asp Asp Tyr Val
35 40 45

Lys Val Lys Lys Ile Ile Asp Lys Tyr His Lys Cys Leu Ile Asp Glu
50 55 60

Ala Leu Ser Gly Phe Thr Phe Asp Thr Glu Ala Asp Gly Arg Ser Asn
65 70 75 80

Asn Ser Leu Ser Glu Tyr Tyr Leu Tyr Tyr Asn Leu Lys Lys Arg Asn
85 90 95

Glu Gln Glu Gln Lys Thr Phe Lys Thr Ile Gln Asn Asn Leu Arg Lys
100 105 110

Gln Ile Val Asn Lys Leu Thr Gln Ser Glu Lys Tyr Lys Arg Ile Asp
115 120 125

Lys Lys Glu Leu Ile Thr Thr Asp Leu Pro Asp Phe Leu Thr Asn Glu
130 135 140

Ser Glu Lys Glu Leu Val Glu Lys Phe Lys Asn Phe Thr Thr Tyr Phe
145 150 155 160

Thr Glu Phe His Lys Asn Arg Lys Asn Met Tyr Ser Lys Glu Glu Lys
165 170 175

```
Ser Thr Ala Ile Ala Phe Arg Leu Ile Asn Glu Asn Leu Pro Lys Phe
            180                 185                 190

Val Asp Asn Ile Ala Ala Phe Glu Lys Val Val Ser Ser Pro Leu Ala
        195                 200                 205

Glu Lys Ile Asn Ala Leu Tyr Glu Asp Phe Lys Glu Tyr Leu Asn Val
    210                 215                 220

Glu Glu Ile Ser Arg Val Phe Arg Leu Asp Tyr Tyr Asp Glu Leu Leu
225                 230                 235                 240

Thr Gln Lys Gln Ile Asp Leu Tyr Asn Ala Ile Val Gly Gly Arg Thr
                245                 250                 255

Glu Glu Asp Asn Lys Ile Gln Ile Lys Gly Leu Asn Gln Tyr Ile Asn
            260                 265                 270

Glu Tyr Asn Gln Gln Gln Thr Asp Arg Ser Asn Arg Leu Pro Lys Leu
        275                 280                 285

Lys Pro Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Ser Val Ser Trp
    290                 295                 300

Leu Pro Pro Lys Phe Asp Ser Asp Lys Asn Leu Leu Ile Lys Ile Lys
305                 310                 315                 320

Glu Cys Tyr Asp Ala Leu Ser Glu Lys Glu Lys Val Phe Asp Lys Leu
                325                 330                 335

Glu Ser Ile Leu Lys Ser Leu Ser Thr Tyr Asp Leu Ser Lys Ile Tyr
            340                 345                 350

Ile Ser Asn Asp Ser Gln Leu Ser Tyr Ile Ser Gln Lys Met Phe Gly
        355                 360                 365

Arg Trp Asp Ile Ile Ser Lys Ala Ile Arg Glu Asp Cys Ala Lys Arg
    370                 375                 380

Asn Pro Gln Lys Ser Arg Glu Ser Leu Glu Lys Phe Ala Glu Arg Ile
385                 390                 395                 400

Asp Lys Lys Leu Lys Thr Ile Asp Ser Ile Ser Ile Gly Asp Val Asp
                405                 410                 415

Glu Cys Leu Ala Gln Leu Gly Glu Thr Tyr Val Lys Arg Val Glu Asp
            420                 425                 430

Tyr Phe Val Ala Met Gly Glu Ser Glu Ile Asp Asp Glu Gln Thr Asp
        435                 440                 445

Thr Thr Ser Phe Lys Lys Asn Ile Glu Gly Ala Tyr Glu Ser Val Lys
    450                 455                 460

Glu Leu Leu Asn Asn Ala Asp Asn Ile Thr Asp Asn Asn Leu Met Gln
465                 470                 475                 480

Asp Lys Gly Asn Val Glu Lys Ile Lys Thr Leu Leu Asp Ala Ile Lys
                485                 490                 495

Asp Leu Gln Arg Phe Ile Lys Pro Leu Leu Gly Lys Gly Asp Glu Ala
            500                 505                 510
```

Asp Lys Asp Gly Val Phe Tyr Gly Glu Phe Thr Ser Leu Trp Thr Lys
515 520 525

Leu Asp Gln Val Thr Pro Leu Tyr Asn Met Val Arg Asn Tyr Leu Thr
530 535 540

Ser Lys Pro Tyr Ser Thr Lys Lys Ile Lys Leu Asn Phe Glu Asn Ser
545 550 555 560

Thr Leu Met Asp Gly Trp Asp Leu Asn Lys Glu Pro Asp Asn Thr Thr
565 570 575

Val Ile Phe Cys Lys Asp Gly Leu Tyr Tyr Leu Gly Ile Met Gly Lys
580 585 590

Lys Tyr Asn Arg Val Phe Val Asp Arg Glu Asp Leu Pro His Asp Gly
595 600 605

Glu Cys Tyr Asp Lys Met Glu Tyr Lys Leu Leu Pro Gly Ala Asn Lys
610 615 620

Met Leu Pro Lys Val Phe Phe Ser Glu Thr Gly Ile Gln Arg Phe Leu
625 630 635 640

Pro Ser Glu Glu Leu Leu Gly Lys Tyr Glu Arg Gly Thr His Lys Lys
645 650 655

Gly Ala Gly Phe Asp Leu Gly Asp Cys Arg Ala Leu Ile Asp Phe Phe
660 665 670

Lys Lys Ser Ile Glu Arg His Asp Asp Trp Lys Lys Phe Asp Phe Lys
675 680 685

Phe Ser Asp Thr Ser Thr Tyr Gln Asp Ile Ser Glu Phe Tyr Arg Glu
690 695 700

Val Glu Gln Gln Gly Tyr Lys Met Ser Phe Arg Lys Val Ser Val Asp
705 710 715 720

Tyr Ile Lys Ser Leu Val Glu Glu Gly Lys Leu Tyr Leu Phe Gln Ile
725 730 735

Tyr Asn Lys Asp Phe Ser Ala His Ser Lys Gly Thr Pro Asn Met His
740 745 750

Thr Leu Tyr Trp Lys Met Leu Phe Asp Glu Glu Asn Leu Lys Asp Val
755 760 765

Val Tyr Lys Leu Asn Gly Glu Ala Glu Val Phe Phe Arg Lys Ser Ser
770 775 780

Ile Thr Val Gln Ser Pro Thr His Pro Ala Asn Ser Pro Ile Lys Asn
785 790 795 800

Lys Asn Lys Asp Asn Gln Lys Lys Glu Ser Lys Phe Glu Tyr Asp Leu
805 810 815

Ile Lys Asp Arg Arg Tyr Thr Val Asp Lys Phe Leu Phe His Val Pro
820 825 830

Ile Thr Met Asn Phe Lys Ser Val Gly Gly Ser Asn Ile Asn Gln Leu
835 840 845

```
Val Lys Arg His Ile Arg Ser Ala Thr Asp Leu His Ile Ile Gly Ile
    850                 855                 860

Asp Arg Gly Glu Arg His Leu Leu Tyr Leu Thr Val Ile Asp Ser Arg
865                 870                 875                 880

Gly Asn Ile Lys Glu Gln Phe Ser Leu Asn Glu Ile Val Asn Glu Tyr
                885                 890                 895

Asn Gly Asn Thr Tyr Arg Thr Asp Tyr His Glu Leu Leu Asp Thr Arg
            900                 905                 910

Glu Gly Glu Arg Thr Glu Ala Arg Arg Asn Trp Gln Thr Ile Gln Asn
        915                 920                 925

Ile Arg Glu Leu Lys Glu Gly Tyr Leu Ser Gln Val Ile His Lys Ile
    930                 935                 940

Ser Glu Leu Ala Ile Lys Tyr Asn Ala Val Ile Val Leu Glu Asp Leu
945                 950                 955                 960

Asn Phe Gly Phe Met Arg Ser Arg Gln Lys Val Glu Lys Gln Val Tyr
                965                 970                 975

Gln Lys Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr Leu Val Asp
            980                 985                 990

Lys Lys Lys Pro Val Ala Glu Thr Gly Gly Leu Leu Arg Ala Tyr Gln
        995                 1000                1005

Leu Thr  Gly Glu Phe Glu Ser  Phe Lys Thr Leu Gly  Lys Gln Ser
    1010                1015                1020

Gly Ile  Leu Phe Tyr Val Pro  Ala Trp Asn Thr Ser  Lys Ile Asp
    1025                1030                1035

Pro Val  Thr Gly Phe Val Asn  Leu Phe Asp Thr His  Tyr Glu Asn
    1040                1045                1050

Ile Glu  Lys Ala Lys Val Phe  Phe Asp Lys Phe Lys  Ser Ile Arg
    1055                1060                1065

Tyr Asn  Ser Asp Lys Asp Trp  Phe Glu Phe Val Val  Asp Asp Tyr
    1070                1075                1080

Thr Arg  Phe Ser Pro Lys Ala  Glu Gly Thr Arg Arg  Asp Trp Thr
    1085                1090                1095

Ile Cys  Thr Gln Gly Lys Arg  Ile Gln Ile Cys Arg  Asn His Gln
    1100                1105                1110

Arg Asn  Asn Glu Trp Glu Gly  Gln Glu Ile Asp Leu  Thr Lys Ala
    1115                1120                1125

Phe Lys  Glu His Phe Glu Ala  Tyr Gly Val Asp Ile  Ser Lys Asp
    1130                1135                1140

Leu Arg  Glu Gln Ile Asn Thr  Gln Asn Lys Lys Glu  Phe Phe Glu
    1145                1150                1155

Glu Leu  Leu Arg Leu Leu Arg  Leu Thr Leu Gln Met  Arg Asn Ser
    1160                1165                1170
```

Met Pro Ser Ser Asp Ile Asp Tyr Leu Ile Ser Pro Val Ala Asn
1175 1180 1185

Asp Thr Gly Cys Phe Phe Asp Ser Arg Lys Gln Ala Glu Leu Lys
1190 1195 1200

Glu Asn Ala Val Leu Pro Met Asn Ala Asp Ala Asn Gly Ala Tyr
1205 1210 1215

Asn Ile Ala Arg Lys Gly Leu Leu Ala Ile Arg Lys Met Lys Gln
1220 1225 1230

Glu Glu Asn Asp Ser Ala Lys Ile Ser Leu Ala Ile Ser Asn Lys
1235 1240 1245

Glu Trp Leu Lys Phe Ala Gln Thr Lys Pro Tyr Leu Glu Asp
1250 1255 1260

<210> 1132
<211> 1264
<212> PRT
<213> 短普雷沃菌

<400> 1132
Met Lys Gln Phe Thr Asn Leu Tyr Gln Leu Ser Lys Thr Leu Arg Phe
1 5 10 15

Glu Leu Lys Pro Ile Gly Lys Thr Leu Glu His Ile Asn Ala Asn Gly
20 25 30

Phe Ile Asp Asn Asp Ala His Arg Ala Glu Ser Tyr Lys Lys Val Lys
35 40 45

Lys Leu Ile Asp Asp Tyr His Lys Asp Tyr Ile Glu Asn Val Leu Asn
50 55 60

Asn Phe Lys Leu Asn Gly Glu Tyr Leu Gln Ala Tyr Phe Asp Leu Tyr
65 70 75 80

Ser Gln Asp Thr Lys Asp Lys Gln Phe Lys Asp Ile Gln Asp Lys Leu
85 90 95

Arg Lys Ser Ile Ala Ser Ala Leu Lys Gly Asp Asp Arg Tyr Lys Thr
100 105 110

Ile Asp Lys Lys Glu Leu Ile Arg Gln Asp Met Lys Thr Phe Leu Lys
115 120 125

Lys Asp Thr Asp Lys Ala Leu Leu Asp Glu Phe Tyr Glu Phe Thr Thr
130 135 140

Tyr Phe Thr Gly Tyr His Glu Asn Arg Lys Asn Met Tyr Ser Asp Glu
145 150 155 160

Ala Lys Ser Thr Ala Ile Ala Tyr Arg Leu Ile His Asp Asn Leu Pro
165 170 175

Lys Phe Ile Asp Asn Ile Ala Val Phe Lys Lys Ile Ala Asn Thr Ser
180 185 190

Val Ala Asp Asn Phe Ser Thr Ile Tyr Lys Asn Phe Glu Glu Tyr Leu
195 200 205

Asn Val Asn Ser Ile Asp Glu Ile Phe Ser Leu Asp Tyr Tyr Asn Ile

210 215 220

Val Leu Thr Gln Thr Gln Ile Glu Val Tyr Asn Ser Ile Ile Gly Gly
225 230 235 240

Arg Thr Leu Glu Asp Asp Thr Lys Ile Gln Gly Ile Asn Glu Phe Val
245 250 255

Asn Leu Tyr Asn Gln Gln Leu Ala Asn Lys Lys Asp Arg Leu Pro Lys
260 265 270

Leu Lys Pro Leu Phe Lys Gln Ile Leu Ser Asp Arg Val Gln Leu Ser
275 280 285

Trp Leu Gln Glu Glu Phe Asn Thr Gly Ala Asp Val Leu Asn Ala Val
290 295 300

Lys Glu Tyr Cys Thr Ser Tyr Phe Asp Asn Val Glu Glu Ser Val Lys
305 310 315 320

Val Leu Leu Thr Gly Ile Ser Asp Tyr Asp Leu Ser Lys Ile Tyr Ile
325 330 335

Thr Asn Asp Leu Ala Leu Thr Asp Val Ser Gln Arg Met Phe Gly Glu
340 345 350

Trp Ser Ile Ile Pro Asn Ala Ile Glu Gln Arg Leu Arg Ser Asp Asn
355 360 365

Pro Lys Lys Thr Asn Glu Lys Glu Glu Lys Tyr Ser Asp Arg Ile Ser
370 375 380

Lys Leu Lys Lys Leu Pro Lys Ser Tyr Ser Leu Gly Tyr Ile Asn Glu
385 390 395 400

Cys Ile Ser Glu Leu Asn Gly Ile Asp Ile Ala Asp Tyr Tyr Ala Thr
405 410 415

Leu Gly Ala Ile Asn Thr Glu Ser Lys Gln Glu Pro Ser Ile Pro Thr
420 425 430

Ser Ile Gln Val His Tyr Asn Ala Leu Lys Pro Ile Leu Asp Thr Asp
435 440 445

Tyr Pro Arg Glu Lys Asn Leu Ser Gln Asp Lys Leu Thr Val Met Gln
450 455 460

Leu Lys Asp Leu Leu Asp Asp Phe Lys Ala Leu Gln His Phe Ile Lys
465 470 475 480

Pro Leu Leu Gly Asn Gly Asp Glu Ala Glu Lys Asp Glu Lys Phe Tyr
485 490 495

Gly Glu Leu Met Gln Leu Trp Glu Val Ile Asp Ser Ile Thr Pro Leu
500 505 510

Tyr Asn Lys Val Arg Asn Tyr Cys Thr Arg Lys Pro Phe Ser Thr Glu
515 520 525

Lys Ile Lys Val Asn Phe Glu Asn Ala Gln Leu Leu Asp Gly Trp Asp
530 535 540

Glu Asn Lys Glu Ser Thr Asn Ala Ser Ile Ile Leu Arg Lys Asn Gly

545 550 555 560

Met Tyr Tyr Leu Gly Ile Met Lys Lys Glu Tyr Arg Asn Ile Leu Thr
565 570 575

Lys Pro Met Pro Ser Asp Gly Asp Cys Tyr Asp Lys Val Val Tyr Lys
580 585 590

Phe Phe Lys Asp Ile Thr Thr Met Val Pro Lys Cys Thr Thr Gln Met
595 600 605

Lys Ser Val Lys Glu His Phe Ser Asn Ser Asn Asp Asp Tyr Thr Leu
610 615 620

Phe Glu Lys Asp Lys Phe Ile Ala Pro Val Val Ile Thr Lys Glu Ile
625 630 635 640

Phe Asp Leu Asn Asn Val Leu Tyr Asn Gly Val Lys Lys Phe Gln Ile
645 650 655

Gly Tyr Leu Asn Asn Thr Gly Asp Ser Phe Gly Tyr Asn His Ala Val
660 665 670

Glu Ile Trp Lys Ser Phe Cys Leu Lys Phe Leu Lys Ala Tyr Lys Ser
675 680 685

Thr Ser Ile Tyr Asp Phe Ser Ser Ile Glu Lys Asn Ile Gly Cys Tyr
690 695 700

Asn Asp Leu Asn Ser Phe Tyr Gly Ala Val Asn Leu Leu Leu Tyr Asn
705 710 715 720

Leu Thr Tyr Arg Lys Val Ser Val Asp Tyr Ile His Gln Leu Val Asp
725 730 735

Glu Asp Lys Met Tyr Leu Phe Met Ile Tyr Asn Lys Asp Phe Ser Thr
740 745 750

Tyr Ser Lys Gly Thr Pro Asn Met His Thr Leu Tyr Trp Lys Met Leu
755 760 765

Phe Asp Glu Ser Asn Leu Asn Asp Val Val Tyr Lys Leu Asn Gly Gln
770 775 780

Ala Glu Val Phe Tyr Arg Lys Lys Ser Ile Thr Tyr Gln His Pro Thr
785 790 795 800

His Pro Ala Asn Lys Pro Ile Asp Asn Lys Asn Val Asn Asn Pro Lys
805 810 815

Lys Gln Ser Asn Phe Glu Tyr Asp Leu Ile Lys Asp Lys Arg Tyr Thr
820 825 830

Val Asp Lys Phe Met Phe His Val Pro Ile Thr Leu Asn Phe Lys Gly
835 840 845

Met Gly Asn Gly Asp Ile Asn Met Gln Val Arg Glu Tyr Ile Lys Thr
850 855 860

Thr Asp Asp Leu His Phe Ile Gly Ile Asp Arg Gly Glu Arg His Leu
865 870 875 880

Leu Tyr Ile Cys Val Ile Asn Gly Lys Gly Glu Ile Val Glu Gln Tyr

885 890 895

Ser Leu Asn Glu Ile Val Asn Asn Tyr Lys Gly Thr Glu Tyr Lys Thr
900 905 910

Asp Tyr His Thr Leu Leu Ser Glu Arg Asp Lys Lys Arg Lys Glu Glu
915 920 925

Arg Ser Ser Trp Gln Thr Ile Glu Gly Ile Lys Glu Leu Lys Ser Gly
930 935 940

Tyr Leu Ser Gln Val Ile His Lys Ile Thr Gln Leu Met Ile Lys Tyr
945 950 955 960

Asn Ala Ile Val Leu Leu Glu Asp Leu Asn Met Gly Phe Lys Arg Gly
965 970 975

Arg Gln Lys Val Glu Ser Ser Val Tyr Gln Gln Phe Glu Lys Ala Leu
980 985 990

Ile Asp Lys Leu Asn Tyr Leu Val Asp Lys Asn Lys Asp Ala Asn Glu
995 1000 1005

Ile Gly Gly Leu Leu His Ala Tyr Gln Leu Thr Asn Asp Pro Lys
1010 1015 1020

Leu Pro Asn Lys Asn Ser Lys Gln Ser Gly Phe Leu Phe Tyr Val
1025 1030 1035

Pro Ala Trp Asn Thr Ser Lys Ile Asp Pro Val Thr Gly Phe Val
1040 1045 1050

Asn Leu Leu Asp Thr Arg Tyr Glu Asn Val Ala Lys Ala Gln Ala
1055 1060 1065

Phe Phe Lys Lys Phe Asp Ser Ile Arg Tyr Asn Lys Glu Tyr Asp
1070 1075 1080

Arg Phe Glu Phe Lys Phe Asp Tyr Ser Asn Phe Thr Ala Lys Ala
1085 1090 1095

Glu Asp Thr Arg Thr Gln Trp Thr Leu Cys Thr Tyr Gly Thr Arg
1100 1105 1110

Ile Glu Thr Phe Arg Asn Ala Glu Lys Asn Ser Asn Trp Asp Ser
1115 1120 1125

Arg Glu Ile Asp Leu Thr Thr Glu Trp Lys Thr Leu Phe Thr Gln
1130 1135 1140

His Asn Ile Pro Leu Asn Ala Asn Leu Lys Glu Ala Ile Leu Leu
1145 1150 1155

Gln Ala Asn Lys Asn Phe Tyr Thr Asp Ile Leu His Leu Met Lys
1160 1165 1170

Leu Thr Leu Gln Met Arg Asn Ser Val Thr Gly Thr Asp Ile Asp
1175 1180 1185

Tyr Met Val Ser Pro Val Ala Asn Glu Cys Gly Glu Phe Phe Asp
1190 1195 1200

Ser Arg Lys Val Lys Glu Gly Leu Pro Val Asn Ala Asp Ala Asn

1205 1210 1215

Gly Ala Tyr Asn Ile Ala Arg Lys Gly Leu Trp Leu Ala Gln Gln
1220 1225 1230

Ile Lys Asn Ala Asn Asp Leu Ser Asp Val Lys Leu Ala Ile Thr
1235 1240 1245

Asn Lys Glu Trp Leu Gln Phe Ala Gln Lys Lys Gln Tyr Leu Lys
1250 1255 1260

Asp

<210> 1133
<211> 1260
<212> PRT
<213> 狗口腔卟啉單胞菌

<400> 1133
Met Asp Ser Leu Lys Asp Phe Thr Asn Leu Tyr Pro Val Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr Leu Glu Asn Ile Glu
20 25 30

Lys Ala Gly Ile Leu Lys Glu Asp Glu His Arg Ala Glu Ser Tyr Arg
35 40 45

Arg Val Lys Lys Ile Ile Asp Thr Tyr His Lys Val Phe Ile Asp Ser
50 55 60

Ser Leu Glu Asn Met Ala Lys Met Gly Ile Glu Asn Glu Ile Lys Ala
65 70 75 80

Met Leu Gln Ser Phe Cys Glu Leu Tyr Lys Lys Asp His Arg Thr Glu
85 90 95

Gly Glu Asp Lys Ala Leu Asp Lys Ile Arg Ala Val Leu Arg Gly Leu
100 105 110

Ile Val Gly Ala Phe Thr Gly Val Cys Gly Arg Arg Glu Asn Thr Val
115 120 125

Gln Asn Glu Lys Tyr Glu Ser Leu Phe Lys Glu Lys Leu Ile Lys Glu
130 135 140

Ile Leu Pro Asp Phe Val Leu Ser Thr Glu Ala Glu Ser Leu Pro Phe
145 150 155 160

Ser Val Glu Glu Ala Thr Arg Ser Leu Lys Glu Phe Asp Ser Phe Thr
165 170 175

Ser Tyr Phe Ala Gly Phe Tyr Glu Asn Arg Lys Asn Ile Tyr Ser Thr
180 185 190

Lys Pro Gln Ser Thr Ala Ile Ala Tyr Arg Leu Ile His Glu Asn Leu
195 200 205

Pro Lys Phe Ile Asp Asn Ile Leu Val Phe Gln Lys Ile Lys Glu Pro
210 215 220

Ile Ala Lys Glu Leu Glu His Ile Arg Ala Asp Phe Ser Ala Gly Gly
225 230 235 240

```
Tyr Ile Lys Lys Asp Glu Arg Leu Glu Asp Ile Phe Ser Leu Asn Tyr
                245                 250                 255

Tyr Ile His Val Leu Ser Gln Ala Gly Ile Glu Lys Tyr Asn Ala Leu
            260                 265                 270

Ile Gly Lys Ile Val Thr Glu Gly Asp Gly Glu Met Lys Gly Leu Asn
        275                 280                 285

Glu His Ile Asn Leu Tyr Asn Gln Gln Arg Gly Arg Glu Asp Arg Leu
    290                 295                 300

Pro Leu Phe Arg Pro Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Gln
305                 310                 315                 320

Leu Ser Tyr Leu Pro Glu Ser Phe Glu Lys Asp Glu Glu Leu Leu Arg
                325                 330                 335

Ala Leu Lys Glu Phe Tyr Asp His Ile Ala Glu Asp Ile Leu Gly Arg
            340                 345                 350

Thr Gln Gln Leu Met Thr Ser Ile Ser Glu Tyr Asp Leu Ser Arg Ile
        355                 360                 365

Tyr Val Arg Asn Asp Ser Gln Leu Thr Asp Ile Ser Lys Lys Met Leu
    370                 375                 380

Gly Asp Trp Asn Ala Ile Tyr Met Ala Arg Glu Arg Ala Tyr Asp His
385                 390                 395                 400

Glu Gln Ala Pro Lys Arg Ile Thr Ala Lys Tyr Glu Arg Asp Arg Ile
                405                 410                 415

Lys Ala Leu Lys Gly Glu Glu Ser Ile Ser Leu Ala Asn Leu Asn Ser
            420                 425                 430

Cys Ile Ala Phe Leu Asp Asn Val Arg Asp Cys Arg Val Asp Thr Tyr
        435                 440                 445

Leu Ser Thr Leu Gly Gln Lys Glu Gly Pro His Gly Leu Ser Asn Leu
    450                 455                 460

Val Glu Asn Val Phe Ala Ser Tyr His Glu Ala Glu Gln Leu Leu Ser
465                 470                 475                 480

Phe Pro Tyr Pro Glu Glu Asn Asn Leu Ile Gln Asp Lys Asp Asn Val
                485                 490                 495

Val Leu Ile Lys Asn Leu Leu Asp Asn Ile Ser Asp Leu Gln Arg Phe
            500                 505                 510

Leu Lys Pro Leu Trp Gly Met Gly Asp Glu Pro Asp Lys Asp Glu Arg
        515                 520                 525

Phe Tyr Gly Glu Tyr Asn Tyr Ile Arg Gly Ala Leu Asp Gln Val Ile
    530                 535                 540

Pro Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Arg Lys Pro Tyr Ser
545                 550                 555                 560

Thr Arg Lys Val Lys Leu Asn Phe Gly Asn Ser Gln Leu Leu Ser Gly
                565                 570                 575
```

```
Trp Asp Arg Asn Lys Glu Lys Asp Asn Ser Cys Val Ile Leu Arg Lys
            580                 585                 590

Gly Gln Asn Phe Tyr Leu Ala Ile Met Asn Asn Arg His Lys Arg Ser
        595                 600                 605

Phe Glu Asn Lys Val Leu Pro Glu Tyr Lys Glu Gly Glu Pro Tyr Phe
    610                 615                 620

Glu Lys Met Asp Tyr Lys Phe Leu Pro Asp Pro Asn Lys Met Leu Pro
625                 630                 635                 640

Lys Val Phe Leu Ser Lys Lys Gly Ile Glu Ile Tyr Lys Pro Ser Pro
                645                 650                 655

Lys Leu Leu Glu Gln Tyr Gly His Gly Thr His Lys Lys Gly Asp Thr
            660                 665                 670

Phe Ser Met Asp Asp Leu His Glu Leu Ile Asp Phe Phe Lys His Ser
        675                 680                 685

Ile Glu Ala His Glu Asp Trp Lys Gln Phe Gly Phe Lys Phe Ser Asp
    690                 695                 700

Thr Ala Thr Tyr Glu Asn Val Ser Ser Phe Tyr Arg Glu Val Glu Asp
705                 710                 715                 720

Gln Gly Tyr Lys Leu Ser Phe Arg Lys Val Ser Glu Ser Tyr Val Tyr
                725                 730                 735

Ser Leu Ile Asp Gln Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys
            740                 745                 750

Asp Phe Ser Pro Cys Ser Lys Gly Thr Pro Asn Leu His Thr Leu Tyr
        755                 760                 765

Trp Arg Met Leu Phe Asp Glu Arg Asn Leu Ala Asp Val Ile Tyr Lys
    770                 775                 780

Leu Asp Gly Lys Ala Glu Ile Phe Phe Arg Glu Lys Ser Leu Lys Asn
785                 790                 795                 800

Asp His Pro Thr His Pro Ala Gly Lys Pro Ile Lys Lys Lys Ser Arg
                805                 810                 815

Gln Lys Lys Gly Glu Glu Ser Leu Phe Glu Tyr Asp Leu Val Lys Asp
            820                 825                 830

Arg His Tyr Thr Met Asp Lys Phe Gln Phe His Val Pro Ile Thr Met
        835                 840                 845

Asn Phe Lys Cys Ser Ala Gly Ser Lys Val Asn Asp Met Val Asn Ala
    850                 855                 860

His Ile Arg Glu Ala Lys Asp Met His Val Ile Gly Ile Asp Arg Gly
865                 870                 875                 880

Glu Arg Asn Leu Leu Tyr Ile Cys Val Ile Asp Ser Arg Gly Thr Ile
                885                 890                 895

Leu Asp Gln Ile Ser Leu Asn Thr Ile Asn Asp Ile Asp Tyr His Asp
            900                 905                 910
```

```
Leu Leu Glu Ser Arg Asp Lys Asp Arg Gln Gln Glu Arg Arg Asn Trp
        915                 920                 925

Gln Thr Ile Glu Gly Ile Lys Glu Leu Lys Gln Gly Tyr Leu Ser Gln
    930                 935                 940

Ala Val His Arg Ile Ala Glu Leu Met Val Ala Tyr Lys Ala Val Val
945                 950                 955                 960

Ala Leu Glu Asp Leu Asn Met Gly Phe Lys Arg Gly Arg Gln Lys Val
                965                 970                 975

Glu Ser Ser Val Tyr Gln Gln Phe Glu Lys Gln Leu Ile Asp Lys Leu
            980                 985                 990

Asn Tyr Leu Val Asp Lys Lys Lys Arg Pro Glu Asp Ile Gly Gly Leu
        995                 1000                1005

Leu Arg  Ala Tyr Gln Phe Thr  Ala Pro Phe Lys Ser  Phe Lys Glu
    1010                1015                1020

Met Gly  Lys Gln Asn Gly Phe  Leu Phe Tyr Ile Pro  Ala Trp Asn
    1025                1030                1035

Thr Ser  Asn Ile Asp Pro Thr  Thr Gly Phe Val Asn  Leu Phe His
    1040                1045                1050

Ala Gln  Tyr Glu Asn Val Asp  Lys Ala Lys Ser Phe  Phe Gln Lys
    1055                1060                1065

Phe Asp  Ser Ile Ser Tyr Asn  Pro Lys Lys Asp Trp  Phe Glu Phe
    1070                1075                1080

Ala Phe  Asp Tyr Lys Asn Phe  Thr Lys Lys Ala Glu  Gly Ser Arg
    1085                1090                1095

Ser Met  Trp Ile Leu Cys Thr  His Gly Ser Arg Ile  Lys Asn Phe
    1100                1105                1110

Arg Asn  Ser Gln Lys Asn Gly  Gln Trp Asp Ser Glu  Glu Phe Ala
    1115                1120                1125

Leu Thr  Glu Ala Phe Lys Ser  Leu Phe Val Arg Tyr  Glu Ile Asp
    1130                1135                1140

Tyr Thr  Ala Asp Leu Lys Thr  Ala Ile Val Asp Glu  Lys Gln Lys
    1145                1150                1155

Asp Phe  Phe Val Asp Leu Leu  Lys Leu Phe Lys Leu  Thr Val Gln
    1160                1165                1170

Met Arg  Asn Ser Trp Lys Glu  Lys Asp Leu Asp Tyr  Leu Ile Ser
    1175                1180                1185

Pro Val  Ala Gly Ala Asp Gly  Arg Phe Phe Asp Thr  Arg Glu Gly
    1190                1195                1200

Asn Lys  Ser Leu Pro Lys Asp  Ala Asp Ala Asn Gly  Ala Tyr Asn
    1205                1210                1215

Ile Ala  Leu Lys Gly Leu Trp  Ala Leu Arg Gln Ile  Arg Gln Thr
    1220                1225                1230
```

Ser Glu Gly Gly Lys Leu Lys Leu Ala Ile Ser Asn Lys Glu Trp
1235 1240 1245

Leu Gln Phe Val Gln Glu Arg Ser Tyr Glu Lys Asp
1250 1255 1260

<210> 1134
<211> 1260
<212> PRT
<213> 狗口腔卟啉單胞菌

<400> 1134
Met Asp Ser Leu Lys Asp Phe Thr Asn Leu Tyr Pro Val Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr Leu Glu Asn Ile Glu
20 25 30

Lys Ala Gly Ile Leu Lys Glu Asp Glu His Arg Ala Glu Ser Tyr Arg
35 40 45

Arg Val Lys Lys Ile Ile Asp Thr Tyr His Lys Val Phe Ile Asp Ser
50 55 60

Ser Leu Glu Asn Met Ala Lys Met Gly Ile Glu Asn Glu Ile Lys Ala
65 70 75 80

Met Leu Gln Ser Phe Cys Glu Leu Tyr Lys Lys Asp His Arg Thr Glu
85 90 95

Gly Glu Asp Lys Ala Leu Asp Lys Ile Arg Ala Val Leu Arg Gly Leu
100 105 110

Ile Val Gly Ala Phe Thr Gly Val Cys Gly Arg Arg Glu Asn Thr Val
115 120 125

Gln Asn Glu Lys Tyr Glu Ser Leu Phe Lys Glu Lys Leu Ile Lys Glu
130 135 140

Ile Leu Pro Asp Phe Val Leu Ser Thr Glu Ala Glu Ser Leu Pro Phe
145 150 155 160

Ser Val Glu Glu Ala Thr Arg Ser Leu Lys Glu Phe Asp Ser Phe Thr
165 170 175

Ser Tyr Phe Ala Gly Phe Tyr Glu Asn Arg Lys Asn Ile Tyr Ser Thr
180 185 190

Lys Pro Gln Ser Thr Ala Ile Ala Tyr Arg Leu Ile His Glu Asn Leu
195 200 205

Pro Lys Phe Ile Asp Asn Ile Leu Val Phe Gln Lys Ile Lys Glu Pro
210 215 220

Ile Ala Lys Glu Leu Glu His Ile Arg Ala Asp Phe Ser Ala Gly Gly
225 230 235 240

Tyr Ile Lys Lys Asp Glu Arg Leu Glu Asp Ile Phe Ser Leu Asn Tyr
245 250 255

Tyr Ile His Val Leu Ser Gln Ala Gly Ile Glu Lys Tyr Asn Ala Leu
260 265 270

Ile Gly Lys Ile Val Thr Glu Gly Asp Gly Glu Met Lys Gly Leu Asn

```
        275                 280                 285

Glu His Ile Asn Leu Tyr Asn Gln Gln Arg Gly Arg Glu Asp Arg Leu
    290                 295                 300

Pro Leu Phe Arg Pro Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Gln
305                 310                 315                 320

Leu Ser Tyr Leu Pro Glu Ser Phe Glu Lys Asp Glu Glu Leu Leu Arg
                325                 330                 335

Ala Leu Lys Glu Phe Tyr Asp His Ile Ala Glu Asp Ile Leu Gly Arg
            340                 345                 350

Thr Gln Gln Leu Met Thr Ser Ile Ser Glu Tyr Asp Leu Ser Arg Ile
        355                 360                 365

Tyr Val Arg Asn Asp Ser Gln Leu Thr Asp Ile Ser Lys Lys Met Leu
    370                 375                 380

Gly Asp Trp Asn Ala Ile Tyr Met Ala Arg Glu Arg Ala Tyr Asp His
385                 390                 395                 400

Glu Gln Ala Pro Lys Arg Ile Thr Ala Lys Tyr Glu Arg Asp Arg Ile
                405                 410                 415

Lys Ala Leu Lys Gly Glu Glu Ser Ile Ser Leu Ala Asn Leu Asn Ser
            420                 425                 430

Cys Ile Ala Phe Leu Asp Asn Val Arg Asp Cys Arg Val Asp Thr Tyr
        435                 440                 445

Leu Ser Thr Leu Gly Gln Lys Glu Gly Pro His Gly Leu Ser Asn Leu
    450                 455                 460

Val Glu Asn Val Phe Ala Ser Tyr His Glu Ala Glu Gln Leu Leu Ser
465                 470                 475                 480

Phe Pro Tyr Pro Glu Glu Asn Asn Leu Ile Gln Asp Lys Asp Asn Val
                485                 490                 495

Val Leu Ile Lys Asn Leu Leu Asp Asn Ile Ser Asp Leu Gln Arg Phe
            500                 505                 510

Leu Lys Pro Leu Trp Gly Met Gly Asp Glu Pro Asp Lys Asp Glu Arg
        515                 520                 525

Phe Tyr Gly Glu Tyr Asn Tyr Ile Arg Gly Ala Leu Asp Gln Val Ile
    530                 535                 540

Pro Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Arg Lys Pro Tyr Ser
545                 550                 555                 560

Thr Arg Lys Val Lys Leu Asn Phe Gly Asn Ser Gln Leu Leu Ser Gly
                565                 570                 575

Trp Asp Arg Asn Lys Glu Lys Asp Asn Ser Cys Val Ile Leu Arg Lys
            580                 585                 590

Gly Gln Asn Phe Tyr Leu Ala Ile Met Asn Asn Arg His Lys Arg Ser
        595                 600                 605

Phe Glu Asn Lys Val Leu Pro Glu Tyr Lys Glu Gly Glu Pro Tyr Phe
```

610 615 620

Glu Lys Met Asp Tyr Lys Phe Leu Pro Asp Pro Asn Lys Met Leu Pro
625 630 635 640

Lys Val Phe Leu Ser Lys Lys Gly Ile Glu Ile Tyr Lys Pro Ser Pro
645 650 655

Lys Leu Leu Glu Gln Tyr Gly His Gly Thr His Lys Lys Gly Asp Thr
660 665 670

Phe Ser Met Asp Asp Leu His Glu Leu Ile Asp Phe Phe Lys His Ser
675 680 685

Ile Glu Ala His Glu Asp Trp Lys Gln Phe Gly Phe Lys Phe Ser Asp
690 695 700

Thr Ala Thr Tyr Glu Asn Val Ser Ser Phe Tyr Arg Glu Val Glu Asp
705 710 715 720

Gln Gly Tyr Lys Leu Ser Phe Arg Lys Val Ser Glu Ser Tyr Val Tyr
725 730 735

Ser Leu Ile Asp Gln Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys
740 745 750

Asp Phe Ser Pro Cys Ser Lys Gly Thr Pro Asn Leu His Thr Leu Tyr
755 760 765

Trp Arg Met Leu Phe Asp Glu Arg Asn Leu Ala Asp Val Ile Tyr Lys
770 775 780

Leu Asp Gly Lys Ala Glu Ile Phe Phe Arg Glu Lys Ser Leu Lys Asn
785 790 795 800

Asp His Pro Thr His Pro Ala Gly Lys Pro Ile Lys Lys Lys Ser Arg
805 810 815

Gln Lys Lys Gly Glu Glu Ser Leu Phe Glu Tyr Asp Leu Val Lys Asp
820 825 830

Arg Arg Tyr Thr Met Asp Lys Phe Gln Phe His Val Pro Ile Thr Met
835 840 845

Asn Phe Lys Cys Ser Ala Gly Ser Lys Val Asn Asp Met Val Asn Ala
850 855 860

His Ile Arg Glu Ala Lys Asp Met His Val Ile Gly Ile Asp Arg Gly
865 870 875 880

Glu Arg Asn Leu Leu Tyr Ile Cys Val Ile Asp Ser Arg Gly Thr Ile
885 890 895

Leu Asp Gln Ile Ser Leu Asn Thr Ile Asn Asp Ile Asp Tyr His Asp
900 905 910

Leu Leu Glu Ser Arg Asp Lys Asp Arg Gln Gln Glu Arg Arg Asn Trp
915 920 925

Gln Thr Ile Glu Gly Ile Lys Glu Leu Lys Gln Gly Tyr Leu Ser Gln
930 935 940

Ala Val His Arg Ile Ala Glu Leu Met Val Ala Tyr Lys Ala Val Val

945 950 955 960

Ala Leu Glu Asp Leu Asn Met Gly Phe Lys Arg Gly Arg Gln Lys Val
965 970 975

Glu Ser Ser Val Tyr Gln Gln Phe Glu Lys Gln Leu Ile Asp Lys Leu
980 985 990

Asn Tyr Leu Val Asp Lys Lys Lys Arg Pro Glu Asp Ile Gly Gly Leu
995 1000 1005

Leu Arg Ala Tyr Gln Phe Thr Ala Pro Phe Lys Ser Phe Lys Glu
1010 1015 1020

Met Gly Lys Gln Asn Gly Phe Leu Phe Tyr Ile Pro Ala Trp Asn
1025 1030 1035

Thr Ser Asn Ile Asp Pro Thr Thr Gly Phe Val Asn Leu Phe His
1040 1045 1050

Ala Gln Tyr Glu Asn Val Asp Lys Ala Lys Ser Phe Phe Gln Lys
1055 1060 1065

Phe Asp Ser Ile Ser Tyr Asn Pro Lys Lys Asp Trp Phe Glu Phe
1070 1075 1080

Ala Phe Asp Tyr Lys Asn Phe Thr Lys Lys Ala Glu Gly Ser Arg
1085 1090 1095

Ser Met Trp Ile Leu Cys Thr His Gly Ser Arg Ile Lys Asn Phe
1100 1105 1110

Arg Asn Ser Gln Lys Asn Gly Gln Trp Asp Ser Glu Glu Phe Ala
1115 1120 1125

Leu Thr Glu Ala Phe Lys Ser Leu Phe Val Arg Tyr Glu Ile Asp
1130 1135 1140

Tyr Thr Ala Asp Leu Lys Thr Ala Ile Val Asp Glu Lys Gln Lys
1145 1150 1155

Asp Phe Phe Val Asp Leu Leu Lys Leu Phe Lys Leu Thr Val Gln
1160 1165 1170

Met Arg Asn Ser Trp Lys Glu Lys Asp Leu Asp Tyr Leu Ile Ser
1175 1180 1185

Pro Val Ala Gly Ala Asp Gly Arg Phe Phe Asp Thr Arg Glu Gly
1190 1195 1200

Asn Lys Ser Leu Pro Lys Asp Ala Asp Ala Asn Gly Ala Tyr Asn
1205 1210 1215

Ile Ala Leu Lys Gly Leu Trp Ala Leu Arg Gln Ile Arg Gln Thr
1220 1225 1230

Ser Glu Gly Gly Lys Leu Lys Leu Ala Ile Ser Asn Lys Glu Trp
1235 1240 1245

Leu Gln Phe Val Gln Glu Arg Ser Tyr Glu Lys Asp
1250 1255 1260

<210> 1135

<211> 1260
<212> PRT
<213> 狗口腔卟啉單胞菌

<400> 1135

```
Met Asp Ser Leu Lys Asp Phe Thr Asn Leu Tyr Pro Val Ser Lys Thr
1               5                  10                  15

Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr Leu Glu Asn Ile Glu
            20                  25                  30

Lys Ala Gly Ile Leu Lys Glu Asp Glu His Arg Ala Glu Ser Tyr Arg
        35                  40                  45

Arg Val Lys Lys Ile Ile Asp Thr Tyr His Lys Val Phe Ile Asp Ser
    50                  55                  60

Ser Leu Glu Asn Met Ala Lys Met Gly Ile Glu Asn Glu Ile Lys Ala
65                  70                  75                  80

Met Leu Gln Ser Phe Cys Glu Leu Tyr Lys Lys Asp His Arg Thr Glu
                85                  90                  95

Gly Glu Asp Lys Ala Leu Asp Lys Ile Arg Ala Val Leu Arg Gly Leu
            100                 105                 110

Ile Val Gly Ala Phe Thr Gly Val Cys Gly Arg Arg Glu Asn Thr Val
        115                 120                 125

Gln Asn Glu Lys Tyr Glu Ser Leu Phe Lys Glu Lys Leu Ile Lys Glu
    130                 135                 140

Ile Leu Pro Asp Phe Val Leu Ser Thr Glu Ala Glu Ser Leu Pro Phe
145                 150                 155                 160

Ser Val Glu Glu Ala Thr Arg Ser Leu Lys Glu Phe Asp Ser Phe Thr
                165                 170                 175

Ser Tyr Phe Ala Gly Phe Tyr Glu Asn Arg Lys Asn Ile Tyr Ser Thr
            180                 185                 190

Lys Pro Gln Ser Thr Ala Ile Ala Tyr Arg Leu Ile His Glu Asn Leu
        195                 200                 205

Pro Lys Phe Ile Asp Asn Ile Leu Val Phe Gln Lys Ile Lys Glu Pro
    210                 215                 220

Ile Ala Lys Glu Leu Glu His Ile Arg Ala Asp Phe Ser Ala Gly Gly
225                 230                 235                 240

Tyr Ile Lys Lys Asp Glu Arg Leu Glu Asp Ile Phe Ser Leu Asn Tyr
                245                 250                 255

Tyr Ile His Val Leu Ser Gln Ala Gly Ile Glu Lys Tyr Asn Ala Leu
            260                 265                 270

Ile Gly Lys Ile Val Thr Glu Gly Asp Gly Glu Met Lys Gly Leu Asn
        275                 280                 285

Glu His Ile Asn Leu Tyr Asn Gln Gln Arg Gly Arg Glu Asp Arg Leu
    290                 295                 300

Pro Leu Phe Arg Pro Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Gln
305                 310                 315                 320
```

Leu Ser Tyr Leu Pro Glu Ser Phe Glu Lys Asp Glu Glu Leu Leu Arg
325 330 335

Ala Leu Lys Glu Phe Tyr Asp His Ile Ala Glu Asp Ile Leu Gly Arg
340 345 350

Thr Gln Gln Leu Met Thr Ser Ile Ser Glu Tyr Asp Leu Ser Arg Ile
355 360 365

Tyr Val Arg Asn Asp Ser Gln Leu Thr Asp Ile Ser Lys Lys Met Leu
370 375 380

Gly Asp Trp Asn Ala Ile Tyr Met Ala Arg Glu Arg Ala Tyr Asp His
385 390 395 400

Glu Gln Ala Pro Lys Arg Ile Thr Ala Lys Tyr Glu Arg Asp Arg Ile
405 410 415

Lys Ala Leu Lys Gly Glu Glu Ser Ile Ser Leu Ala Asn Leu Asn Ser
420 425 430

Cys Ile Ala Phe Leu Asp Asn Val Arg Asp Cys Arg Val Asp Thr Tyr
435 440 445

Leu Ser Thr Leu Gly Gln Lys Glu Gly Pro His Gly Leu Ser Asn Leu
450 455 460

Val Glu Asn Val Phe Ala Ser Tyr His Glu Ala Glu Gln Leu Leu Ser
465 470 475 480

Phe Pro Tyr Pro Glu Glu Asn Asn Leu Ile Gln Asp Lys Asp Asn Val
485 490 495

Val Leu Ile Lys Asn Leu Leu Asp Asn Ile Ser Asp Leu Gln Arg Phe
500 505 510

Leu Lys Pro Leu Trp Gly Met Gly Asp Glu Pro Asp Lys Asp Glu Arg
515 520 525

Phe Tyr Gly Glu Tyr Asn Tyr Ile Arg Gly Ala Leu Asp Gln Val Ile
530 535 540

Pro Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Arg Lys Pro Tyr Ser
545 550 555 560

Thr Arg Lys Val Lys Leu Asn Phe Gly Asn Ser Gln Leu Leu Ser Gly
565 570 575

Trp Asp Arg Asn Lys Glu Lys Asp Asn Ser Cys Val Ile Leu Arg Lys
580 585 590

Gly Gln Asn Phe Tyr Leu Ala Ile Met Asn Asn Arg His Lys Arg Ser
595 600 605

Phe Glu Asn Lys Met Leu Pro Glu Tyr Lys Glu Gly Glu Pro Tyr Phe
610 615 620

Glu Lys Met Asp Tyr Lys Phe Leu Pro Asp Pro Asn Lys Met Leu Pro
625 630 635 640

Lys Val Phe Leu Ser Lys Lys Gly Ile Glu Ile Tyr Lys Pro Ser Pro
645 650 655

Lys Leu Leu Glu Gln Tyr Gly His Gly Thr His Lys Lys Gly Asp Thr
660 665 670

Phe Ser Met Asp Asp Leu His Glu Leu Ile Asp Phe Phe Lys His Ser
675 680 685

Ile Glu Ala His Glu Asp Trp Lys Gln Phe Gly Phe Lys Phe Ser Asp
690 695 700

Thr Ala Thr Tyr Glu Asn Val Ser Ser Phe Tyr Arg Glu Val Glu Asp
705 710 715 720

Gln Gly Tyr Lys Leu Ser Phe Arg Lys Val Ser Glu Ser Tyr Val Tyr
725 730 735

Ser Leu Ile Asp Gln Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys
740 745 750

Asp Phe Ser Pro Cys Ser Lys Gly Thr Pro Asn Leu His Thr Leu Tyr
755 760 765

Trp Arg Met Leu Phe Asp Glu Arg Asn Leu Ala Asp Val Ile Tyr Lys
770 775 780

Leu Asp Gly Lys Ala Glu Ile Phe Phe Arg Glu Lys Ser Leu Lys Asn
785 790 795 800

Asp His Pro Thr His Pro Ala Gly Lys Pro Ile Lys Lys Lys Ser Arg
805 810 815

Gln Lys Lys Gly Glu Glu Ser Leu Phe Glu Tyr Asp Leu Val Lys Asp
820 825 830

Arg Arg Tyr Thr Met Asp Lys Phe Gln Phe His Val Pro Ile Thr Met
835 840 845

Asn Phe Lys Cys Ser Ala Gly Ser Lys Val Asn Asp Met Val Asn Ala
850 855 860

His Ile Arg Glu Ala Lys Asp Met His Val Ile Gly Ile Asp Arg Gly
865 870 875 880

Glu Arg Asn Leu Leu Tyr Ile Cys Val Ile Asp Ser Arg Gly Thr Ile
885 890 895

Leu Asp Gln Ile Ser Leu Asn Thr Ile Asn Asp Ile Asp Tyr His Asp
900 905 910

Leu Leu Glu Ser Arg Asp Lys Asp Arg Gln Gln Glu His Arg Asn Trp
915 920 925

Gln Thr Ile Glu Gly Ile Lys Glu Leu Lys Gln Gly Tyr Leu Ser Gln
930 935 940

Ala Val His Arg Ile Ala Glu Leu Met Val Ala Tyr Lys Ala Val Val
945 950 955 960

Ala Leu Glu Asp Leu Asn Met Gly Phe Lys Arg Gly Arg Gln Lys Val
965 970 975

Glu Ser Ser Val Tyr Gln Gln Phe Glu Lys Gln Leu Ile Asp Lys Leu
980 985 990

```
Asn Tyr Leu Val Asp Lys Lys Lys Arg Pro Glu Asp Ile Gly Gly Leu
        995                 1000                1005

Leu Arg  Ala Tyr Gln Phe Thr  Ala Pro Phe Lys Ser  Phe Lys Glu
    1010                1015                1020

Met Gly  Lys Gln Asn Gly Phe  Leu Phe Tyr Ile Pro  Ala Trp Asn
    1025                1030                1035

Thr Ser  Asn Ile Asp Pro Thr  Thr Gly Phe Val Asn  Leu Phe His
    1040                1045                1050

Val Gln  Tyr Glu Asn Val Asp  Lys Ala Lys Ser Phe  Phe Gln Lys
    1055                1060                1065

Phe Asp  Ser Ile Ser Tyr Asn  Pro Lys Lys Asp Trp  Phe Glu Phe
    1070                1075                1080

Ala Phe  Asp Tyr Lys Asn Phe  Thr Lys Lys Ala Glu  Gly Ser Arg
    1085                1090                1095

Ser Met  Trp Ile Leu Cys Thr  His Gly Ser Arg Ile  Lys Asn Phe
    1100                1105                1110

Arg Asn  Ser Gln Lys Asn Gly  Gln Trp Asp Ser Glu  Glu Phe Ala
    1115                1120                1125

Leu Thr  Glu Ala Phe Lys Ser  Leu Phe Val Arg Tyr  Glu Ile Asp
    1130                1135                1140

Tyr Thr  Ala Asp Leu Lys Thr  Ala Ile Val Asp Glu  Lys Gln Lys
    1145                1150                1155

Asp Phe  Phe Val Asp Leu Leu  Lys Leu Phe Lys Leu  Thr Val Gln
    1160                1165                1170

Met Arg  Asn Ser Trp Lys Glu  Lys Asp Leu Asp Tyr  Leu Ile Ser
    1175                1180                1185

Pro Val  Ala Gly Ala Asp Gly  Arg Phe Phe Asp Thr  Arg Glu Gly
    1190                1195                1200

Asn Lys  Ser Leu Pro Lys Asp  Ala Asp Ala Asn Gly  Ala Tyr Asn
    1205                1210                1215

Ile Ala  Leu Lys Gly Leu Trp  Ala Leu Arg Gln Ile  Arg Gln Thr
    1220                1225                1230

Ser Glu  Gly Gly Lys Leu Lys  Leu Ala Ile Ser Asn  Lys Glu Trp
    1235                1240                1245

Leu Gln  Phe Val Gln Glu Arg  Ser Tyr Glu Lys Asp
    1250                1255                1260
```

<210> 1136
<211> 1262
<212> PRT
<213> 狗口腔卟啉單胞菌

<400> 1136

```
Met Pro Trp Ile Asp Leu Lys Asp Phe Thr Asn Leu Tyr Pro Val Ser
1               5                  10                  15

Lys Thr Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr Leu Glu Asn
```

```
           20                  25                  30

Ile Glu Lys Ala Gly Ile Leu Lys Glu Asp Glu His Arg Ala Glu Ser
        35                  40                  45

Tyr Arg Arg Val Lys Lys Ile Ile Asp Thr Tyr His Lys Val Phe Ile
    50                  55                  60

Asp Ser Ser Leu Glu Asn Met Ala Lys Met Gly Ile Glu Asn Glu Ile
65                  70                  75                  80

Lys Ala Met Leu Gln Ser Phe Cys Glu Leu Tyr Lys Lys Asp His Arg
                85                  90                  95

Thr Glu Gly Glu Asp Lys Ala Leu Asp Lys Ile Arg Ala Val Leu Arg
            100                 105                 110

Gly Leu Ile Val Gly Ala Phe Thr Gly Val Cys Gly Arg Arg Glu Asn
        115                 120                 125

Thr Val Gln Asn Glu Lys Tyr Glu Ser Leu Phe Lys Glu Lys Leu Ile
    130                 135                 140

Lys Glu Ile Leu Pro Asp Phe Val Leu Ser Thr Glu Ala Glu Ser Leu
145                 150                 155                 160

Pro Phe Ser Val Glu Glu Ala Thr Arg Ser Leu Lys Glu Phe Asp Ser
                165                 170                 175

Phe Thr Ser Tyr Phe Ala Gly Phe Tyr Glu Asn Arg Lys Asn Ile Tyr
            180                 185                 190

Ser Thr Lys Pro Gln Ser Thr Ala Ile Ala Tyr Arg Leu Ile His Glu
        195                 200                 205

Asn Leu Pro Lys Phe Ile Asp Asn Ile Leu Val Phe Gln Lys Ile Lys
    210                 215                 220

Glu Pro Ile Ala Lys Glu Leu Glu His Ile Arg Ala Asp Phe Ser Ala
225                 230                 235                 240

Gly Gly Tyr Ile Lys Lys Asp Glu Arg Leu Glu Asp Ile Phe Ser Leu
                245                 250                 255

Asn Tyr Tyr Ile His Val Leu Ser Gln Ala Gly Ile Glu Lys Tyr Asn
            260                 265                 270

Ala Leu Ile Gly Lys Ile Val Thr Glu Gly Asp Gly Glu Met Lys Gly
        275                 280                 285

Leu Asn Glu His Ile Asn Leu Tyr Asn Gln Gln Arg Gly Arg Glu Asp
    290                 295                 300

Arg Leu Pro Leu Phe Arg Pro Leu Tyr Lys Gln Ile Leu Ser Asp Arg
305                 310                 315                 320

Glu Gln Leu Ser Tyr Leu Pro Glu Ser Phe Glu Lys Asp Glu Glu Leu
                325                 330                 335

Leu Arg Ala Leu Lys Glu Phe Tyr Asp His Ile Ala Glu Asp Ile Leu
            340                 345                 350

Gly Arg Thr Gln Gln Leu Met Thr Ser Ile Ser Glu Tyr Asp Leu Ser
```

```
        355                 360                 365

Arg Ile Tyr Val Arg Asn Asp Ser Gln Leu Thr Asp Ile Ser Lys Lys
    370                 375                 380

Met Leu Gly Asp Trp Asn Ala Ile Tyr Met Ala Arg Glu Arg Ala Tyr
385                 390                 395                 400

Asp His Glu Gln Ala Pro Lys Arg Ile Thr Ala Lys Tyr Glu Arg Asp
                405                 410                 415

Arg Ile Lys Ala Leu Lys Gly Glu Glu Ser Ile Ser Leu Ala Asn Leu
            420                 425                 430

Asn Ser Cys Ile Ala Phe Leu Asp Asn Val Arg Asp Cys Arg Val Asp
        435                 440                 445

Thr Tyr Leu Ser Thr Leu Gly Gln Lys Glu Gly Pro His Gly Leu Ser
    450                 455                 460

Asn Leu Val Glu Asn Val Phe Ala Ser Tyr His Glu Ala Glu Gln Leu
465                 470                 475                 480

Leu Ser Phe Pro Tyr Pro Glu Glu Asn Asn Leu Ile Gln Asp Lys Asp
                485                 490                 495

Asn Val Val Leu Ile Lys Asn Leu Leu Asp Asn Ile Ser Asp Leu Gln
            500                 505                 510

Arg Phe Leu Lys Pro Leu Trp Gly Met Gly Asp Glu Pro Asp Lys Asp
        515                 520                 525

Glu Arg Phe Tyr Gly Glu Tyr Asn Tyr Ile Arg Gly Ala Leu Asp Gln
    530                 535                 540

Val Ile Pro Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Arg Lys Pro
545                 550                 555                 560

Tyr Ser Thr Arg Lys Val Lys Leu Asn Phe Gly Asn Ser Gln Leu Leu
                565                 570                 575

Ser Gly Trp Asp Arg Asn Lys Glu Lys Asp Asn Ser Cys Val Ile Leu
            580                 585                 590

Arg Lys Gly Gln Asn Phe Tyr Leu Ala Ile Met Asn Asn Arg His Lys
        595                 600                 605

Arg Ser Phe Glu Asn Lys Val Leu Pro Glu Tyr Lys Glu Gly Glu Pro
    610                 615                 620

Tyr Phe Glu Lys Met Asp Tyr Lys Phe Leu Pro Asp Pro Asn Lys Met
625                 630                 635                 640

Leu Pro Lys Val Phe Leu Ser Lys Lys Gly Ile Glu Ile Tyr Glu Pro
                645                 650                 655

Ser Pro Lys Leu Leu Glu Gln Tyr Gly His Gly Thr His Lys Lys Gly
            660                 665                 670

Asp Thr Phe Ser Met Asp Asp Leu His Glu Leu Ile Asp Phe Phe Lys
        675                 680                 685

His Ser Ile Glu Ala His Glu Asp Trp Lys Gln Phe Gly Phe Lys Phe
```

690 695 700

Ser Asp Thr Ala Thr Tyr Glu Asn Val Ser Ser Phe Tyr Arg Glu Val
705 710 715 720

Glu Asp Gln Gly Tyr Lys Leu Ser Phe Arg Lys Val Ser Glu Ser Tyr
725 730 735

Val Tyr Ser Leu Ile Asp Gln Gly Lys Leu Tyr Leu Phe Gln Ile Tyr
740 745 750

Asn Lys Asp Phe Ser Pro Cys Ser Lys Gly Thr Pro Asn Leu His Thr
755 760 765

Leu Tyr Trp Arg Met Leu Phe Asp Glu Arg Asn Leu Ala Asp Val Ile
770 775 780

Tyr Lys Leu Asp Gly Lys Ala Glu Ile Phe Phe Arg Glu Lys Ser Leu
785 790 795 800

Lys Asn Asp His Pro Thr His Pro Ala Gly Lys Pro Ile Lys Lys Lys
805 810 815

Ser Arg Gln Lys Lys Gly Glu Glu Ser Leu Phe Glu Tyr Asp Leu Val
820 825 830

Lys Asp Arg Arg Tyr Thr Met Asp Lys Phe Gln Phe His Val Pro Ile
835 840 845

Thr Met Asn Phe Lys Cys Ser Ala Gly Ser Lys Val Asn Asp Met Val
850 855 860

Asn Ala His Ile Arg Glu Ala Lys Asp Met His Val Ile Gly Ile Asp
865 870 875 880

Arg Gly Glu Arg Asn Leu Leu Tyr Ile Cys Val Ile Asp Ser Arg Gly
885 890 895

Thr Ile Leu Asp Gln Ile Ser Leu Asn Thr Ile Asn Asp Ile Asp Tyr
900 905 910

His Asp Leu Leu Glu Ser Arg Asp Lys Asp Arg Gln Gln Glu Arg Arg
915 920 925

Asn Trp Gln Thr Ile Glu Gly Ile Lys Glu Leu Lys Gln Gly Tyr Leu
930 935 940

Ser Gln Ala Val His Arg Ile Ala Glu Leu Met Val Ala Tyr Lys Ala
945 950 955 960

Val Val Ala Leu Glu Asp Leu Asn Met Gly Phe Lys Arg Gly Arg Gln
965 970 975

Lys Val Glu Ser Ser Val Tyr Gln Gln Phe Glu Lys Gln Leu Ile Asp
980 985 990

Lys Leu Asn Tyr Leu Val Asp Lys Lys Lys Arg Pro Glu Asp Ile Gly
995 1000 1005

Gly Leu Leu Arg Ala Tyr Gln Phe Thr Ala Pro Phe Lys Ser Phe
1010 1015 1020

Lys Glu Met Gly Lys Gln Asn Gly Phe Leu Phe Tyr Ile Pro Ala

1025 1030 1035

Trp Asn Thr Ser Asn Ile Asp Pro Thr Thr Gly Phe Val Asn Leu
1040 1045 1050

Phe His Ala Gln Tyr Glu Asn Val Asp Lys Ala Lys Ser Phe Phe
1055 1060 1065

Gln Lys Phe Asp Ser Ile Ser Tyr Asn Pro Lys Lys Asp Trp Phe
1070 1075 1080

Glu Phe Ala Phe Asp Tyr Lys Asn Phe Thr Lys Lys Ala Glu Gly
1085 1090 1095

Ser Arg Ser Met Trp Ile Leu Cys Thr His Gly Ser Arg Ile Lys
1100 1105 1110

Asn Phe Arg Asn Ser Gln Lys Asn Gly Gln Trp Asp Ser Glu Glu
1115 1120 1125

Phe Ala Leu Thr Glu Ala Phe Lys Ser Leu Phe Val Arg Tyr Glu
1130 1135 1140

Ile Asp Tyr Thr Ala Asp Leu Lys Thr Ala Ile Val Asp Glu Lys
1145 1150 1155

Gln Lys Asp Phe Phe Val Asp Leu Leu Lys Leu Phe Lys Leu Thr
1160 1165 1170

Val Gln Met Arg Asn Ser Trp Lys Glu Lys Asp Leu Asp Tyr Leu
1175 1180 1185

Ile Ser Pro Val Ala Gly Ala Asp Gly Arg Phe Phe Asp Thr Arg
1190 1195 1200

Glu Gly Asn Lys Ser Leu Pro Lys Asp Ala Asp Ala Asn Gly Ala
1205 1210 1215

Tyr Asn Ile Ala Leu Lys Gly Leu Trp Ala Leu Arg Gln Ile Arg
1220 1225 1230

Gln Thr Ser Glu Gly Gly Lys Leu Lys Leu Ala Ile Ser Asn Lys
1235 1240 1245

Glu Trp Leu Gln Phe Val Gln Glu Arg Ser Tyr Glu Lys Asp
1250 1255 1260

<210> 1137
<211> 1224
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 1137
Met Gly Leu Tyr Asp Gly Phe Val Asn Arg Tyr Ser Val Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Arg Thr Arg Glu Tyr Ile Glu
20 25 30

Thr Asn Gly Ile Leu Ser Asp Asp Glu Glu Arg Ala Lys Asp Tyr Lys
35 40 45

```
Thr Ile Lys Arg Leu Ile Asp Glu Tyr His Lys Asp Tyr Ile Ser Arg
    50                  55                  60

Cys Leu Lys Asn Val Asn Ile Ser Cys Leu Glu Glu Tyr Tyr His Leu
65                  70                  75                  80

Tyr Asn Ser Ser Asn Arg Asp Lys Arg His Glu Glu Leu Asp Ala Leu
                85                  90                  95

Ser Asp Gln Met Arg Gly Glu Ile Ala Ser Phe Leu Thr Gly Asn Asp
            100                 105                 110

Glu Tyr Lys Glu Gln Lys Ser Arg Asp Ile Ile Ile Asn Glu Arg Ile
        115                 120                 125

Ile Asn Phe Ala Ser Thr Asp Glu Glu Leu Ala Ala Val Lys Arg Phe
    130                 135                 140

Arg Lys Phe Thr Ser Tyr Phe Thr Gly Phe Phe Thr Asn Arg Glu Asn
145                 150                 155                 160

Met Tyr Ser Ala Glu Lys Lys Ser Thr Ala Ile Ala His Arg Ile Ile
                165                 170                 175

Asp Val Asn Leu Pro Lys Tyr Val Asp Asn Ile Lys Ala Phe Asn Thr
            180                 185                 190

Ala Ile Glu Ala Gly Val Phe Asp Ile Ala Glu Phe Glu Ser Asn Phe
        195                 200                 205

Lys Ala Ile Thr Asp Glu His Glu Val Ser Asp Leu Leu Asp Ile Thr
    210                 215                 220

Lys Tyr Ser Arg Phe Ile Arg Asn Glu Asp Ile Ile Ile Tyr Asn Thr
225                 230                 235                 240

Leu Leu Gly Gly Ile Ser Met Lys Asp Glu Lys Ile Gln Gly Leu Asn
                245                 250                 255

Glu Leu Ile Asn Leu His Asn Gln Lys His Pro Gly Lys Lys Val Pro
            260                 265                 270

Leu Leu Lys Val Leu Tyr Lys Gln Ile Leu Gly Asp Ser Gln Thr His
        275                 280                 285

Ser Phe Val Asp Asp Gln Phe Glu Asp Asp Gln Gln Val Ile Asn Ala
    290                 295                 300

Val Lys Ala Val Thr Asp Thr Phe Ser Glu Thr Leu Leu Gly Ser Leu
305                 310                 315                 320

Lys Ile Ile Ile Asn Asn Ile Gly His Tyr Asp Leu Asp Arg Ile Tyr
                325                 330                 335

Ile Lys Ala Gly Gln Asp Ile Thr Thr Leu Ser Lys Arg Ala Leu Asn
            340                 345                 350

Asp Trp His Ile Ile Thr Glu Cys Leu Glu Ser Glu Tyr Asp Asp Lys
        355                 360                 365

Phe Pro Lys Asn Lys Lys Ser Asp Thr Tyr Glu Glu Met Arg Asn Arg
    370                 375                 380
```

```
Tyr Val Lys Ser Phe Lys Ser Phe Ser Ile Gly Arg Leu Asn Ser Leu
385                 390                 395                 400

Val Thr Thr Tyr Thr Glu Gln Ala Cys Phe Leu Glu Asn Tyr Leu Gly
                405                 410                 415

Ser Phe Gly Gly Asp Thr Asp Lys Asn Cys Leu Thr Asp Phe Thr Asn
            420                 425                 430

Ser Leu Met Glu Val Glu His Leu Leu Asn Ser Glu Tyr Pro Val Thr
        435                 440                 445

Asn Arg Leu Ile Thr Asp Tyr Glu Ser Val Arg Ile Leu Lys Arg Leu
    450                 455                 460

Leu Asp Ser Glu Met Glu Val Ile His Phe Leu Lys Pro Leu Leu Gly
465                 470                 475                 480

Asn Gly Asn Glu Ser Asp Lys Asp Leu Val Phe Tyr Gly Glu Phe Glu
                485                 490                 495

Ala Glu Tyr Glu Lys Leu Leu Pro Val Ile Lys Val Tyr Asn Arg Val
            500                 505                 510

Arg Asn Tyr Leu Thr Arg Lys Pro Phe Ser Thr Glu Lys Ile Lys Leu
        515                 520                 525

Asn Phe Asn Ser Pro Thr Leu Leu Cys Gly Trp Ser Gln Ser Lys Glu
    530                 535                 540

Lys Glu Tyr Met Gly Val Ile Leu Arg Lys Asp Gly Gln Tyr Tyr Leu
545                 550                 555                 560

Gly Ile Met Thr Pro Ser Asn Lys Lys Ile Phe Ser Glu Ala Pro Lys
                565                 570                 575

Pro Asp Glu Asp Cys Tyr Glu Lys Met Val Leu Arg Tyr Ile Pro His
            580                 585                 590

Pro Tyr Gln Met Leu Pro Lys Val Phe Phe Ser Lys Ser Asn Ile Ala
        595                 600                 605

Phe Phe Asn Pro Ser Asp Glu Ile Leu Arg Ile Lys Lys Gln Glu Ser
    610                 615                 620

Phe Lys Lys Gly Lys Ser Phe Asn Arg Asp Asp Cys His Lys Phe Ile
625                 630                 635                 640

Asp Phe Tyr Lys Asp Ser Ile Asn Arg His Glu Glu Trp Arg Lys Phe
                645                 650                 655

Asn Phe Lys Phe Ser Asp Thr Asp Ser Tyr Glu Asp Ile Ser Arg Phe
            660                 665                 670

Tyr Lys Glu Val Glu Asn Gln Ala Phe Ser Met Ser Phe Thr Lys Ile
        675                 680                 685

Pro Thr Val Tyr Ile Asp Ser Leu Val Asp Glu Gly Lys Leu Tyr Leu
    690                 695                 700

Phe Lys Leu His Asn Lys Asp Phe Ser Glu His Ser Lys Gly Lys Pro
705                 710                 715                 720
```

```
Asn Leu His Thr Val Tyr Trp Asn Ala Leu Phe Ser Glu Tyr Asn Leu
                725                 730                 735

Gln Asn Thr Val Tyr Gln Leu Asn Gly Ser Ala Glu Ile Phe Phe Arg
            740                 745                 750

Lys Ala Ser Ile Pro Glu Asn Glu Arg Val Ile His Lys Lys Asn Val
        755                 760                 765

Pro Ile Thr Arg Lys Val Ala Glu Leu Asn Gly Lys Lys Glu Val Ser
    770                 775                 780

Val Phe Pro Tyr Asp Ile Ile Lys Asn Arg Arg Tyr Thr Val Asp Lys
785                 790                 795                 800

Phe Gln Phe His Val Pro Leu Lys Met Asn Phe Lys Ala Asp Glu Lys
                805                 810                 815

Lys Arg Ile Asn Asp Asp Val Ile Glu Ala Ile Arg Ser Asn Lys Gly
            820                 825                 830

Ile His Val Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu Leu Tyr Leu
        835                 840                 845

Ser Leu Ile Asn Glu Glu Gly Arg Ile Ile Glu Gln Arg Ser Leu Asn
    850                 855                 860

Ile Ile Asp Ser Gly Glu Gly His Thr Gln Asn Tyr Arg Asp Leu Leu
865                 870                 875                 880

Asp Ser Arg Glu Lys Asp Arg Glu Lys Ala Arg Glu Asn Trp Gln Glu
                885                 890                 895

Ile Gln Glu Ile Lys Asp Leu Lys Thr Gly Tyr Leu Ser Gln Ala Ile
            900                 905                 910

His Thr Ile Thr Lys Trp Met Lys Glu Tyr Asn Ala Ile Ile Val Leu
        915                 920                 925

Glu Asp Leu Asn Asp Arg Phe Thr Asn Gly Arg Lys Lys Val Glu Lys
    930                 935                 940

Gln Val Tyr Gln Lys Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr
945                 950                 955                 960

Tyr Val Asp Lys Asp Glu Glu Phe Asp Arg Met Gly Gly Thr His Arg
                965                 970                 975

Ala Leu Gln Leu Thr Glu Lys Phe Glu Ser Phe Gln Lys Leu Gly Arg
            980                 985                 990

Gln Thr Gly Phe Ile Phe Tyr Val Pro Ala Trp Asn Thr Ser Lys Leu
        995                 1000                1005

Asp Pro Thr Thr Gly Phe Val Asp Leu Leu Tyr Pro Lys Tyr Lys
    1010                1015                1020

Ser Val Asp Ala Thr Lys Asp Phe Ile Lys Lys Phe Asp Phe Ile
    1025                1030                1035

Arg Phe Asn Ser Glu Lys Asn Tyr Phe Glu Phe Gly Leu His Tyr
    1040                1045                1050
```

Ser Asn Phe Thr Glu Arg Ala Ile Gly Cys Arg Asp Glu Trp Ile
1055 1060 1065

Leu Cys Ser Tyr Gly Asn Arg Ile Val Asn Phe Arg Asn Ala Ala
1070 1075 1080

Lys Asn Asn Ser Trp Asp Tyr Lys Glu Ile Asp Ile Thr Lys Gln
1085 1090 1095

Leu Leu Asp Leu Phe Glu Lys Asn Gly Ile Asp Val Lys Gln Glu
1100 1105 1110

Asn Leu Ile Asp Ser Ile Cys Glu Met Lys Asp Lys Pro Phe Phe
1115 1120 1125

Lys Ser Leu Ile Ala Asn Ile Lys Leu Ile Leu Gln Ile Arg Asn
1130 1135 1140

Ser Ala Ser Gly Thr Asp Ile Asp Tyr Met Ile Ser Pro Ala Met
1145 1150 1155

Asn Asp Arg Gly Glu Phe Phe Asp Thr Arg Lys Gly Leu Gln Gln
1160 1165 1170

Leu Pro Leu Asp Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala Lys
1175 1180 1185

Lys Gly Leu Trp Ile Val Asp Gln Ile Arg Asn Thr Thr Gly Asn
1190 1195 1200

Asn Val Lys Met Ala Met Ser Asn Arg Glu Trp Met His Phe Ala
1205 1210 1215

Gln Glu Ser Arg Leu Ala
1220

<210> 1138
<211> 1318
<212> PRT
<213> 嗜鰓黃桿菌

<400> 1138
Met Thr Asn Lys Phe Thr Asn Gln Tyr Ser Leu Ser Lys Thr Leu Arg
1 5 10 15

Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Phe Ile Gln Glu Lys
20 25 30

Gly Leu Leu Ser Gln Asp Lys Gln Arg Ala Glu Ser Tyr Gln Glu Met
35 40 45

Lys Lys Thr Ile Asp Lys Phe His Lys Tyr Phe Ile Asp Leu Ala Leu
50 55 60

Ser Asn Ala Lys Leu Thr His Leu Glu Thr Tyr Leu Glu Leu Tyr Asn
65 70 75 80

Lys Ser Ala Glu Thr Lys Lys Glu Gln Lys Phe Lys Asp Asp Leu Lys
85 90 95

Lys Val Gln Asp Asn Leu Arg Lys Glu Ile Val Lys Ser Phe Ser Asp
100 105 110

Gly Asp Ala Lys Ser Ile Phe Ala Ile Leu Asp Lys Lys Glu Leu Ile

115 120 125

Thr Val Glu Leu Glu Lys Trp Phe Glu Asn Asn Glu Gln Lys Asp Ile
130 135 140

Tyr Phe Asp Glu Lys Phe Lys Thr Phe Thr Thr Tyr Phe Thr Gly Phe
145 150 155 160

His Gln Asn Arg Lys Asn Met Tyr Ser Val Glu Pro Asn Ser Thr Ala
165 170 175

Ile Ala Tyr Arg Leu Ile His Glu Asn Leu Pro Lys Phe Leu Glu Asn
180 185 190

Ala Lys Ala Phe Glu Lys Ile Lys Gln Val Glu Ser Leu Gln Val Asn
195 200 205

Phe Arg Glu Leu Met Gly Glu Phe Gly Asp Glu Gly Leu Ile Phe Val
210 215 220

Asn Glu Leu Glu Glu Met Phe Gln Ile Asn Tyr Tyr Asn Asp Val Leu
225 230 235 240

Ser Gln Asn Gly Ile Thr Ile Tyr Asn Ser Ile Ile Ser Gly Phe Thr
245 250 255

Lys Asn Asp Ile Lys Tyr Lys Gly Leu Asn Glu Tyr Ile Asn Asn Tyr
260 265 270

Asn Gln Thr Lys Asp Lys Lys Asp Arg Leu Pro Lys Leu Lys Gln Leu
275 280 285

Tyr Lys Gln Ile Leu Ser Asp Arg Ile Ser Leu Ser Phe Leu Pro Asp
290 295 300

Ala Phe Thr Asp Gly Lys Gln Val Leu Lys Ala Ile Phe Asp Phe Tyr
305 310 315 320

Lys Ile Asn Leu Leu Ser Tyr Thr Ile Glu Gly Gln Glu Glu Ser Gln
325 330 335

Asn Leu Leu Leu Leu Ile Arg Gln Thr Ile Glu Asn Leu Ser Ser Phe
340 345 350

Asp Thr Gln Lys Ile Tyr Leu Lys Asn Asp Thr His Leu Thr Thr Ile
355 360 365

Ser Gln Gln Val Phe Gly Asp Phe Ser Val Phe Ser Thr Ala Leu Asn
370 375 380

Tyr Trp Tyr Glu Thr Lys Val Asn Pro Lys Phe Glu Thr Glu Tyr Ser
385 390 395 400

Lys Ala Asn Glu Lys Lys Arg Glu Ile Leu Asp Lys Ala Lys Ala Val
405 410 415

Phe Thr Lys Gln Asp Tyr Phe Ser Ile Ala Phe Leu Gln Glu Val Leu
420 425 430

Ser Glu Tyr Ile Leu Thr Leu Asp His Thr Ser Asp Ile Val Lys Lys
435 440 445

His Ser Ser Asn Cys Ile Ala Asp Tyr Phe Lys Asn His Phe Val Ala

```
    450                 455                 460

Lys Lys Glu Asn Glu Thr Asp Lys Thr Phe Asp Phe Ile Ala Asn Ile
465                 470                 475                 480

Thr Ala Lys Tyr Gln Cys Ile Gln Gly Ile Leu Glu Asn Ala Asp Gln
                485                 490                 495

Tyr Glu Asp Glu Leu Lys Gln Asp Gln Lys Leu Ile Asp Asn Leu Lys
            500                 505                 510

Phe Phe Leu Asp Ala Ile Leu Glu Leu Leu His Phe Ile Lys Pro Leu
        515                 520                 525

His Leu Lys Ser Glu Ser Ile Thr Glu Lys Asp Thr Ala Phe Tyr Asp
    530                 535                 540

Val Phe Glu Asn Tyr Tyr Glu Ala Leu Ser Leu Leu Thr Pro Leu Tyr
545                 550                 555                 560

Asn Met Val Arg Asn Tyr Val Thr Gln Lys Pro Tyr Ser Thr Glu Lys
                565                 570                 575

Ile Lys Leu Asn Phe Glu Asn Ala Gln Leu Leu Asn Gly Trp Asp Ala
            580                 585                 590

Asn Lys Glu Gly Asp Tyr Leu Thr Thr Ile Leu Lys Lys Asp Gly Asn
        595                 600                 605

Tyr Phe Leu Ala Ile Met Asp Lys Lys His Asn Lys Ala Phe Gln Lys
    610                 615                 620

Phe Pro Glu Gly Lys Glu Asn Tyr Glu Lys Met Val Tyr Lys Leu Leu
625                 630                 635                 640

Pro Gly Val Asn Lys Met Leu Pro Lys Val Phe Phe Ser Asn Lys Asn
                645                 650                 655

Ile Ala Tyr Phe Asn Pro Ser Lys Glu Leu Leu Glu Asn Tyr Lys Lys
            660                 665                 670

Glu Thr His Lys Lys Gly Asp Thr Phe Asn Leu Glu His Cys His Thr
        675                 680                 685

Leu Ile Asp Phe Phe Lys Asp Ser Leu Asn Lys His Glu Asp Trp Lys
    690                 695                 700

Tyr Phe Asp Phe Gln Phe Ser Glu Thr Lys Ser Tyr Gln Asp Leu Ser
705                 710                 715                 720

Gly Phe Tyr Arg Glu Val Glu His Gln Gly Tyr Lys Ile Asn Phe Lys
                725                 730                 735

Asn Ile Asp Ser Glu Tyr Ile Asp Gly Leu Val Asn Glu Gly Lys Leu
            740                 745                 750

Phe Leu Phe Gln Ile Tyr Ser Lys Asp Phe Ser Pro Phe Ser Lys Gly
        755                 760                 765

Lys Pro Asn Met His Thr Leu Tyr Trp Lys Ala Leu Phe Glu Glu Gln
    770                 775                 780

Asn Leu Gln Asn Val Ile Tyr Lys Leu Asn Gly Gln Ala Glu Ile Phe
```

785 790 795 800

Phe Arg Lys Ala Ser Ile Lys Pro Lys Asn Ile Ile Leu His Lys Lys
805 810 815

Lys Ile Lys Ile Ala Lys Lys His Phe Ile Asp Lys Lys Thr Lys Thr
820 825 830

Ser Glu Ile Val Pro Val Gln Thr Ile Lys Asn Leu Asn Met Tyr Tyr
835 840 845

Gln Gly Lys Ile Ser Glu Lys Glu Leu Thr Gln Asp Asp Leu Arg Tyr
850 855 860

Ile Asp Asn Phe Ser Ile Phe Asn Glu Lys Asn Lys Thr Ile Asp Ile
865 870 875 880

Ile Lys Asp Lys Arg Phe Thr Val Asp Lys Phe Gln Phe His Val Pro
885 890 895

Ile Thr Met Asn Phe Lys Ala Thr Gly Gly Ser Tyr Ile Asn Gln Thr
900 905 910

Val Leu Glu Tyr Leu Gln Asn Asn Pro Glu Val Lys Ile Ile Gly Leu
915 920 925

Asp Arg Gly Glu Arg His Leu Val Tyr Leu Thr Leu Ile Asp Gln Gln
930 935 940

Gly Asn Ile Leu Lys Gln Glu Ser Leu Asn Thr Ile Thr Asp Ser Lys
945 950 955 960

Ile Ser Thr Pro Tyr His Lys Leu Leu Asp Asn Lys Glu Asn Glu Arg
965 970 975

Asp Leu Ala Arg Lys Asn Trp Gly Thr Val Glu Asn Ile Lys Glu Leu
980 985 990

Lys Glu Gly Tyr Ile Ser Gln Val Val His Lys Ile Ala Thr Leu Met
995 1000 1005

Leu Glu Glu Asn Ala Ile Val Val Met Glu Asp Leu Asn Phe Gly
1010 1015 1020

Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Ile Tyr Gln Lys
1025 1030 1035

Leu Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr Leu Val Leu Lys
1040 1045 1050

Asp Lys Gln Pro Gln Glu Leu Gly Gly Leu Tyr Asn Ala Leu Gln
1055 1060 1065

Leu Thr Asn Lys Phe Glu Ser Phe Gln Lys Met Gly Lys Gln Ser
1070 1075 1080

Gly Phe Leu Phe Tyr Val Pro Ala Trp Asn Thr Ser Lys Ile Asp
1085 1090 1095

Pro Thr Thr Gly Phe Val Asn Tyr Phe Tyr Thr Lys Tyr Glu Asn
1100 1105 1110

Val Asp Lys Ala Lys Ala Phe Phe Glu Lys Phe Glu Ala Ile Arg

1115 1120 1125

Phe Asn Ala Glu Lys Lys Tyr Phe Glu Phe Glu Val Lys Lys Tyr
1130 1135 1140

Ser Asp Phe Asn Pro Lys Ala Glu Gly Thr Gln Gln Ala Trp Thr
1145 1150 1155

Ile Cys Thr Tyr Gly Glu Arg Ile Glu Thr Lys Arg Gln Lys Asp
1160 1165 1170

Gln Asn Asn Lys Phe Val Ser Thr Pro Ile Asn Leu Thr Glu Lys
1175 1180 1185

Ile Glu Asp Phe Leu Gly Lys Asn Gln Ile Val Tyr Gly Asp Gly
1190 1195 1200

Asn Cys Ile Lys Ser Gln Ile Ala Ser Lys Asp Asp Lys Ala Phe
1205 1210 1215

Phe Glu Thr Leu Leu Tyr Trp Phe Lys Met Thr Leu Gln Met Arg
1220 1225 1230

Asn Ser Glu Thr Arg Thr Asp Ile Asp Tyr Leu Ile Ser Pro Val
1235 1240 1245

Met Asn Asp Asn Gly Thr Phe Tyr Asn Ser Arg Asp Tyr Glu Lys
1250 1255 1260

Leu Glu Asn Pro Thr Leu Pro Lys Asp Ala Asp Ala Asn Gly Ala
1265 1270 1275

Tyr His Ile Ala Lys Lys Gly Leu Met Leu Leu Asn Lys Ile Asp
1280 1285 1290

Gln Ala Asp Leu Thr Lys Lys Val Asp Leu Ser Ile Ser Asn Arg
1295 1300 1305

Asp Trp Leu Gln Phe Val Gln Lys Asn Lys
1310 1315

<210> 1139
<211> 1273
<212> PRT
<213> 黃桿菌屬某種

<400> 1139
Met Lys Asn Phe Ser Asn Leu Tyr Gln Val Ser Lys Thr Val Arg Phe
1 5 10 15

Glu Leu Lys Pro Ile Gly Asn Thr Leu Glu Asn Ile Lys Asn Lys Ser
20 25 30

Leu Leu Lys Asn Asp Ser Ile Arg Ala Glu Ser Tyr Gln Lys Met Lys
35 40 45

Lys Thr Ile Asp Glu Phe His Lys Tyr Phe Ile Asp Leu Ala Leu Asn
50 55 60

Asn Lys Lys Leu Ser Tyr Leu Asn Glu Tyr Ile Ala Leu Tyr Thr Gln
65 70 75 80

Ser Ala Glu Ala Lys Lys Glu Asp Lys Phe Lys Ala Asp Phe Lys Lys
85 90 95

```
Val Gln Asp Asn Leu Arg Lys Glu Ile Val Ser Ser Phe Thr Glu Gly
            100                 105                 110

Glu Ala Lys Ala Ile Phe Ser Val Leu Asp Lys Lys Glu Leu Ile Thr
        115                 120                 125

Ile Glu Leu Glu Lys Trp Lys Asn Glu Asn Asn Leu Ala Val Tyr Leu
    130                 135                 140

Asp Glu Ser Phe Lys Ser Phe Thr Thr Tyr Phe Thr Gly Phe His Gln
145                 150                 155                 160

Asn Arg Lys Asn Met Tyr Ser Ala Glu Ala Asn Ser Thr Ala Ile Ala
                165                 170                 175

Tyr Arg Leu Ile His Glu Asn Leu Pro Lys Phe Ile Glu Asn Ser Lys
            180                 185                 190

Ala Phe Glu Lys Ser Ser Gln Ile Ala Glu Leu Gln Pro Lys Ile Glu
        195                 200                 205

Lys Leu Tyr Lys Glu Phe Glu Ala Tyr Leu Asn Val Asn Ser Ile Ser
    210                 215                 220

Glu Leu Phe Glu Ile Asp Tyr Phe Asn Glu Val Leu Thr Gln Lys Gly
225                 230                 235                 240

Ile Thr Val Tyr Asn Asn Ile Ile Gly Gly Arg Thr Ala Thr Glu Gly
                245                 250                 255

Lys Gln Lys Ile Gln Gly Leu Asn Glu Ile Ile Asn Leu Tyr Asn Gln
            260                 265                 270

Thr Lys Pro Lys Asn Glu Arg Leu Pro Lys Leu Lys Gln Leu Tyr Lys
        275                 280                 285

Gln Ile Leu Ser Asp Arg Ile Ser Leu Ser Phe Leu Pro Asp Ala Phe
    290                 295                 300

Thr Glu Gly Lys Gln Val Leu Lys Ala Val Phe Glu Phe Tyr Lys Ile
305                 310                 315                 320

Asn Leu Leu Ser Tyr Lys Gln Asp Gly Val Glu Glu Ser Gln Asn Leu
                325                 330                 335

Leu Glu Leu Ile Gln Gln Val Val Lys Asn Leu Gly Asn Gln Asp Val
            340                 345                 350

Asn Lys Ile Tyr Leu Lys Asn Asp Thr Ser Leu Thr Thr Ile Ala Gln
        355                 360                 365

Gln Leu Phe Gly Asp Phe Ser Val Phe Ser Ala Ala Leu Gln Tyr Arg
    370                 375                 380

Tyr Glu Thr Val Val Asn Pro Lys Tyr Thr Ala Glu Tyr Gln Lys Ala
385                 390                 395                 400

Asn Glu Ala Lys Gln Glu Lys Leu Asp Lys Glu Lys Ile Lys Phe Val
                405                 410                 415

Lys Gln Asp Tyr Phe Ser Ile Ala Phe Leu Gln Glu Val Val Ala Asp
            420                 425                 430
```

Tyr Val Lys Thr Leu Asp Glu Asn Leu Asp Trp Lys Gln Lys Tyr Thr
435 440 445

Pro Ser Cys Ile Ala Asp Tyr Phe Thr Thr His Phe Ile Ala Lys Lys
450 455 460

Glu Asn Glu Ala Asp Lys Thr Phe Asn Phe Ile Ala Asn Ile Lys Ala
465 470 475 480

Lys Tyr Gln Cys Ile Gln Gly Ile Leu Glu Gln Ala Asp Asp Tyr Glu
485 490 495

Asp Glu Leu Lys Gln Asp Gln Lys Leu Ile Asp Asn Ile Lys Phe Phe
500 505 510

Leu Asp Ala Ile Leu Glu Val Val His Phe Ile Lys Pro Leu His Leu
515 520 525

Lys Ser Glu Ser Ile Thr Glu Lys Asp Asn Ala Phe Tyr Asp Val Phe
530 535 540

Glu Asn Tyr Tyr Glu Ala Leu Asn Val Val Thr Pro Leu Tyr Asn Met
545 550 555 560

Val Arg Asn Tyr Val Thr Gln Lys Pro Tyr Ser Thr Glu Lys Ile Lys
565 570 575

Leu Asn Phe Glu Asn Ala Gln Leu Leu Asn Gly Trp Asp Ala Asn Lys
580 585 590

Glu Lys Asp Tyr Leu Thr Thr Ile Leu Lys Arg Asp Gly Asn Tyr Phe
595 600 605

Leu Ala Ile Met Asp Lys Lys His Asn Lys Thr Phe Gln Gln Phe Thr
610 615 620

Glu Asp Asp Glu Asn Tyr Glu Lys Ile Val Tyr Lys Leu Leu Pro Gly
625 630 635 640

Val Asn Lys Met Leu Pro Lys Val Phe Phe Ser Asn Lys Asn Ile Ala
645 650 655

Phe Phe Asn Pro Ser Lys Glu Ile Leu Asp Asn Tyr Lys Asn Asn Thr
660 665 670

His Lys Lys Gly Ala Thr Phe Asn Leu Lys Asp Cys His Ala Leu Ile
675 680 685

Asp Phe Phe Lys Asp Ser Leu Asn Lys His Glu Asp Trp Lys Tyr Phe
690 695 700

Asp Phe Gln Phe Ser Glu Thr Lys Thr Tyr Gln Asp Leu Ser Gly Phe
705 710 715 720

Tyr Lys Glu Val Glu His Gln Gly Tyr Lys Ile Asn Phe Lys Lys Val
725 730 735

Ser Val Ser Gln Ile Asp Thr Leu Ile Glu Glu Gly Lys Met Tyr Leu
740 745 750

Phe Gln Ile Tyr Asn Lys Asp Phe Ser Pro Tyr Ala Lys Gly Lys Pro
755 760 765

```
Asn Met His Thr Leu Tyr Trp Lys Ala Leu Phe Glu Thr Gln Asn Leu
    770                 775                 780

Glu Asn Val Ile Tyr Lys Leu Asn Gly Gln Ala Glu Ile Phe Phe Arg
785                 790                 795                 800

Lys Ala Ser Ile Lys Lys Lys Asn Ile Ile Thr His Lys Ala His Gln
                805                 810                 815

Pro Ile Ala Ala Lys Asn Pro Leu Thr Pro Thr Ala Lys Asn Thr Phe
            820                 825                 830

Ala Tyr Asp Leu Ile Lys Asp Lys Arg Tyr Thr Val Asp Lys Phe Gln
        835                 840                 845

Phe His Val Pro Ile Thr Met Asn Phe Lys Ala Thr Gly Asn Ser Tyr
    850                 855                 860

Ile Asn Gln Asp Val Leu Ala Tyr Leu Lys Asp Asn Pro Glu Val Asn
865                 870                 875                 880

Ile Ile Gly Leu Asp Arg Gly Glu Arg His Leu Val Tyr Leu Thr Leu
                885                 890                 895

Ile Asp Gln Lys Gly Thr Ile Leu Leu Gln Glu Ser Leu Asn Val Ile
            900                 905                 910

Gln Asp Glu Lys Thr His Thr Pro Tyr His Thr Leu Leu Asp Asn Lys
        915                 920                 925

Glu Ile Ala Arg Asp Lys Ala Arg Lys Asn Trp Gly Ser Ile Glu Ser
    930                 935                 940

Ile Lys Glu Leu Lys Glu Gly Tyr Ile Ser Gln Val Val His Lys Ile
945                 950                 955                 960

Thr Lys Met Met Ile Glu His Asn Ala Ile Val Val Met Glu Asp Leu
                965                 970                 975

Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Ile Tyr
            980                 985                 990

Gln Lys Leu Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr Leu Val Leu
        995                 1000                1005

Lys Asp  Lys Gln Pro His Glu Leu Gly Gly Leu  Tyr Asn Ala Leu
    1010                1015                1020

Gln Leu  Thr Asn Lys Phe Glu Ser Phe Gln Lys  Met Gly Lys Gln
    1025                1030                1035

Ser Gly  Phe Leu Phe Tyr Val Pro Ala Trp Asn  Thr Ser Lys Ile
    1040                1045                1050

Asp Pro  Thr Thr Gly Phe Val Asn Tyr Phe Tyr  Thr Lys Tyr Glu
    1055                1060                1065

Asn Val  Glu Lys Ala Lys Thr Phe Phe Ser Lys  Phe Asp Ser Ile
    1070                1075                1080

Leu Tyr  Asn Lys Thr Lys Gly Tyr Phe Glu Phe  Val Val Lys Asn
    1085                1090                1095
```

Tyr Ser Asp Phe Asn Pro Lys Ala Ala Asp Thr Arg Gln Glu Trp
1100 1105 1110

Thr Ile Cys Thr His Gly Glu Arg Ile Glu Thr Lys Arg Gln Lys
1115 1120 1125

Glu Gln Asn Asn Asn Phe Val Ser Thr Thr Ile Gln Leu Thr Glu
1130 1135 1140

Gln Phe Val Asn Phe Phe Glu Lys Val Gly Leu Asp Leu Ser Lys
1145 1150 1155

Glu Leu Lys Thr Gln Leu Ile Ala Gln Asn Glu Lys Ser Phe Phe
1160 1165 1170

Glu Glu Leu Phe His Leu Leu Lys Leu Thr Leu Gln Met Arg Asn
1175 1180 1185

Ser Glu Ser His Thr Glu Ile Asp Tyr Leu Ile Ser Pro Val Ala
1190 1195 1200

Asn Glu Lys Gly Ile Phe Tyr Asp Ser Arg Lys Ala Thr Ala Ser
1205 1210 1215

Leu Pro Ile Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Ala Lys
1220 1225 1230

Lys Gly Leu Trp Ile Met Glu Gln Ile Asn Lys Thr Asn Ser Glu
1235 1240 1245

Asp Asp Leu Lys Lys Val Lys Leu Ala Ile Ser Asn Arg Glu Trp
1250 1255 1260

Leu Gln Tyr Val Gln Gln Val Gln Lys Lys
1265 1270

<210> 1140
<211> 1264
<212> PRT
<213> 密斯氏菌屬某種

<400> 1140
Met Gln Thr Leu Phe Glu Asn Phe Thr Asn Gln Tyr Pro Val Ser Lys
1 5 10 15

Thr Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Lys Asp Phe Ile
20 25 30

Glu Gln Lys Gly Leu Leu Lys Lys Asp Glu Asp Arg Ala Glu Lys Tyr
35 40 45

Lys Lys Val Lys Asn Ile Ile Asp Glu Tyr His Lys Asp Phe Ile Glu
50 55 60

Lys Ser Leu Asn Gly Leu Lys Leu Asp Gly Leu Glu Lys Tyr Lys Thr
65 70 75 80

Leu Tyr Leu Lys Gln Glu Lys Asp Asp Lys Asp Lys Lys Ala Phe Asp
85 90 95

Lys Glu Lys Glu Asn Leu Arg Lys Gln Ile Ala Asn Ala Phe Arg Asn
100 105 110

Asn Glu Lys Phe Lys Thr Leu Phe Ala Lys Glu Leu Ile Lys Asn Asp

115 120 125

Leu Met Ser Phe Ala Cys Glu Glu Asp Lys Lys Asn Val Lys Glu Phe
130 135 140

Glu Ala Phe Thr Thr Tyr Phe Thr Gly Phe His Gln Asn Arg Ala Asn
145 150 155 160

Met Tyr Val Ala Asp Glu Lys Arg Thr Ala Ile Ala Ser Arg Leu Ile
165 170 175

His Glu Asn Leu Pro Lys Phe Ile Asp Asn Ile Lys Ile Phe Glu Lys
180 185 190

Met Lys Lys Glu Ala Pro Glu Leu Leu Ser Pro Phe Asn Gln Thr Leu
195 200 205

Lys Asp Met Lys Asp Val Ile Lys Gly Thr Thr Leu Glu Glu Ile Phe
210 215 220

Ser Leu Asp Tyr Phe Asn Lys Thr Leu Thr Gln Ser Gly Ile Asp Ile
225 230 235 240

Tyr Asn Ser Val Ile Gly Gly Arg Thr Pro Glu Glu Gly Lys Thr Lys
245 250 255

Ile Lys Gly Leu Asn Glu Tyr Ile Asn Thr Tyr Phe Thr Gly Phe His
260 265 270

Gln Asn Arg Ala Asn Met Tyr Val Asp Phe Asn Gln Lys Gln Thr Asp
275 280 285

Lys Lys Lys Arg Gln Pro Lys Phe Lys Gln Leu Tyr Lys Gln Ile Leu
290 295 300

Ser Asp Arg Gln Ser Leu Ser Phe Ile Ala Glu Ala Phe Lys Asn Asp
305 310 315 320

Thr Glu Ile Leu Glu Ala Ile Glu Lys Phe Tyr Val Asn Glu Leu Leu
325 330 335

His Phe Ser Asn Glu Gly Lys Ser Thr Asn Val Leu Asp Ala Ile Lys
340 345 350

Asn Ala Val Ser Asn Leu Glu Ser Phe Asn Leu Thr Lys Met Tyr Phe
355 360 365

Arg Ser Gly Ala Ser Leu Thr Asp Val Ser Arg Lys Val Phe Gly Glu
370 375 380

Trp Ser Ile Ile Asn Arg Ala Leu Asp Asn Tyr Tyr Ala Thr Thr Tyr
385 390 395 400

Pro Ile Lys Pro Arg Glu Lys Ser Glu Lys Tyr Glu Glu Arg Lys Glu
405 410 415

Lys Trp Leu Lys Gln Asp Phe Asn Val Ser Leu Ile Gln Thr Ala Ile
420 425 430

Asp Glu Tyr Asp Asn Glu Thr Val Lys Gly Lys Asn Ser Gly Lys Val
435 440 445

Ile Ala Asp Tyr Phe Ala Lys Phe Cys Asp Asp Lys Glu Thr Asp Leu

```
    450                 455                 460

Ile Gln Lys Val Asn Glu Gly Tyr Ile Ala Val Lys Asp Leu Leu Asn
465                 470                 475                 480

Thr Pro Cys Pro Glu Asn Glu Lys Leu Gly Ser Asn Lys Asp Gln Val
                485                 490                 495

Lys Gln Ile Lys Ala Phe Met Asp Ser Ile Met Asp Ile Met His Phe
            500                 505                 510

Val Arg Pro Leu Ser Leu Lys Asp Thr Asp Lys Glu Lys Asp Glu Thr
        515                 520                 525

Phe Tyr Ser Leu Phe Thr Pro Leu Tyr Asp His Leu Thr Gln Thr Ile
    530                 535                 540

Ala Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Gln Lys Pro Tyr Ser
545                 550                 555                 560

Thr Glu Lys Ile Lys Leu Asn Phe Glu Asn Ser Thr Leu Leu Gly Gly
                565                 570                 575

Trp Asp Leu Asn Lys Glu Thr Asp Asn Thr Ala Ile Ile Leu Arg Lys
            580                 585                 590

Asp Asn Leu Tyr Tyr Leu Gly Ile Met Asp Lys Arg His Asn Arg Ile
        595                 600                 605

Phe Arg Asn Val Pro Lys Ala Asp Lys Lys Asp Phe Cys Tyr Glu Lys
    610                 615                 620

Met Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
625                 630                 635                 640

Phe Phe Ser Gln Ser Arg Ile Gln Glu Phe Thr Pro Ser Ala Lys Leu
                645                 650                 655

Leu Glu Asn Tyr Ala Asn Glu Thr His Lys Lys Gly Asp Asn Phe Asn
            660                 665                 670

Leu Asn His Cys His Lys Leu Ile Asp Phe Phe Lys Asp Ser Ile Asn
        675                 680                 685

Lys His Glu Asp Trp Lys Asn Phe Asp Phe Arg Phe Ser Ala Thr Ser
    690                 695                 700

Thr Tyr Ala Asp Leu Ser Gly Phe Tyr His Glu Val Glu His Gln Gly
705                 710                 715                 720

Tyr Lys Ile Ser Phe Gln Ser Val Ala Asp Ser Phe Ile Asp Asp Leu
                725                 730                 735

Val Asn Glu Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe
            740                 745                 750

Ser Pro Phe Ser Lys Gly Lys Pro Asn Leu His Thr Leu Tyr Trp Lys
        755                 760                 765

Met Leu Phe Asp Glu Asn Asn Leu Lys Asp Val Val Tyr Lys Leu Asn
    770                 775                 780

Gly Glu Ala Glu Val Phe Tyr Arg Lys Lys Ser Ile Ala Glu Lys Asn
```

785 790 795 800

Thr Thr Ile His Lys Ala Asn Glu Ser Ile Ile Asn Lys Asn Pro Asp
805 810 815

Asn Pro Lys Ala Thr Ser Thr Phe Asn Tyr Asp Ile Val Lys Asp Lys
820 825 830

Arg Tyr Thr Ile Asp Lys Phe Gln Phe His Ile Pro Ile Thr Met Asn
835 840 845

Phe Lys Ala Glu Gly Ile Phe Asn Met Asn Gln Arg Val Asn Gln Phe
850 855 860

Leu Lys Ala Asn Pro Asp Ile Asn Ile Ile Gly Ile Asp Arg Gly Glu
865 870 875 880

Arg His Leu Leu Tyr Tyr Ala Leu Ile Asn Gln Lys Gly Lys Ile Leu
885 890 895

Lys Gln Asp Thr Leu Asn Val Ile Ala Asn Glu Lys Gln Lys Val Asp
900 905 910

Tyr His Asn Leu Leu Asp Lys Lys Glu Gly Asp Arg Ala Thr Ala Arg
915 920 925

Gln Glu Trp Gly Val Ile Glu Thr Ile Lys Glu Leu Lys Glu Gly Tyr
930 935 940

Leu Ser Gln Val Ile His Lys Leu Thr Asp Leu Met Ile Glu Asn Asn
945 950 955 960

Ala Ile Ile Val Met Glu Asp Leu Asn Phe Gly Phe Lys Arg Gly Arg
965 970 975

Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu Lys Met Leu Ile
980 985 990

Asp Lys Leu Asn Tyr Leu Val Asp Lys Asn Lys Lys Ala Asn Glu Leu
995 1000 1005

Gly Gly Leu Leu Asn Ala Phe Gln Leu Ala Asn Lys Phe Glu Ser
1010 1015 1020

Phe Gln Lys Met Gly Lys Gln Asn Gly Phe Ile Phe Tyr Val Pro
1025 1030 1035

Ala Trp Asn Thr Ser Lys Thr Asp Pro Ala Thr Gly Phe Ile Asp
1040 1045 1050

Phe Leu Lys Pro Arg Tyr Glu Asn Leu Asn Gln Ala Lys Asp Phe
1055 1060 1065

Phe Glu Lys Phe Asp Ser Ile Arg Leu Asn Ser Lys Ala Asp Tyr
1070 1075 1080

Phe Glu Phe Ala Phe Asp Phe Lys Asn Phe Thr Glu Lys Ala Asp
1085 1090 1095

Gly Gly Arg Thr Lys Trp Thr Val Cys Thr Thr Asn Glu Asp Arg
1100 1105 1110

Tyr Ala Trp Asn Arg Ala Leu Asn Asn Asn Arg Gly Ser Gln Glu

1115 1120 1125

Lys Tyr Asp Ile Thr Ala Glu Leu Lys Ser Leu Phe Asp Gly Lys
1130 1135 1140

Val Asp Tyr Lys Ser Gly Lys Asp Leu Lys Gln Gln Ile Ala Ser
1145 1150 1155

Gln Glu Ser Ala Asp Phe Phe Lys Ala Leu Met Lys Asn Leu Ser
1160 1165 1170

Ile Thr Leu Ser Leu Arg His Asn Asn Gly Glu Lys Gly Asp Asn
1175 1180 1185

Glu Gln Asp Tyr Ile Leu Ser Pro Val Ala Asp Ser Lys Gly Arg
1190 1195 1200

Phe Phe Asp Ser Arg Lys Ala Asp Asp Asp Met Pro Lys Asn Ala
1205 1210 1215

Asp Ala Asn Gly Ala Tyr His Ile Ala Leu Lys Gly Leu Trp Cys
1220 1225 1230

Leu Glu Gln Ile Ser Lys Thr Asp Asp Leu Lys Lys Val Lys Leu
1235 1240 1245

Ala Ile Ser Asn Lys Glu Trp Leu Glu Phe Val Gln Thr Leu Lys
1250 1255 1260

Gly

<210> 1141
<211> 1250
<212> PRT
<213> 密斯氏菌屬某種

<400> 1141
Met Gln Thr Leu Phe Glu Asn Phe Thr Asn Gln Tyr Pro Val Ser Lys
1 5 10 15

Thr Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Lys Asp Phe Ile
20 25 30

Glu Gln Lys Gly Leu Leu Lys Lys Asp Glu Asp Arg Ala Glu Lys Tyr
35 40 45

Lys Lys Val Lys Asn Ile Ile Asp Glu Tyr His Lys Asp Phe Ile Glu
50 55 60

Lys Ser Leu Asn Gly Leu Lys Leu Asp Gly Leu Glu Glu Tyr Lys Thr
65 70 75 80

Leu Tyr Leu Lys Gln Glu Lys Asp Asp Lys Asp Lys Lys Ala Phe Asp
85 90 95

Lys Glu Lys Glu Asn Leu Arg Lys Gln Ile Ala Asn Ala Phe Arg Asn
100 105 110

Asn Glu Lys Phe Lys Thr Leu Phe Ala Lys Glu Leu Ile Lys Asn Asp
115 120 125

Leu Met Ser Phe Ala Cys Glu Glu Asp Lys Lys Asn Val Lys Glu Phe
130 135 140

```
Glu Ala Phe Thr Thr Ala Asp Glu Lys Arg Thr Ala Ile Ala Ser Arg
145                 150                 155                 160

Leu Ile His Glu Asn Leu Pro Lys Phe Ile Asp Asn Ile Lys Ile Phe
                165                 170                 175

Glu Lys Met Lys Lys Glu Ala Pro Glu Leu Leu Ser Pro Phe Asn Gln
            180                 185                 190

Thr Leu Lys Asp Met Lys Asp Val Ile Lys Gly Thr Thr Leu Glu Glu
        195                 200                 205

Ile Phe Ser Leu Asp Tyr Phe Asn Lys Thr Leu Thr Gln Ser Gly Ile
    210                 215                 220

Asp Ile Tyr Asn Ser Val Ile Gly Gly Arg Thr Pro Glu Glu Gly Lys
225                 230                 235                 240

Thr Lys Ile Lys Gly Leu Asn Glu Tyr Ile Asn Thr Tyr Phe Thr Gly
                245                 250                 255

Phe His Ser Asn Arg Gln Asn Ile Tyr Ser Asp Phe Asn Gln Lys Gln
            260                 265                 270

Thr Asp Lys Lys Lys Arg Gln Pro Lys Phe Lys Gln Leu Tyr Lys Gln
        275                 280                 285

Ile Leu Ser Asp Arg Gln Ser Leu Ser Phe Ile Ala Glu Ala Phe Lys
    290                 295                 300

Asn Asp Thr Glu Ile Leu Glu Ala Ile Glu Lys Phe Tyr Val Asn Glu
305                 310                 315                 320

Leu Leu His Phe Ser Asn Glu Gly Lys Ser Thr Asn Val Leu Asp Ala
                325                 330                 335

Ile Lys Asn Ala Val Ser Asn Leu Glu Ser Phe Asn Leu Thr Lys Ile
            340                 345                 350

Tyr Phe Arg Ser Gly Thr Ser Leu Thr Asp Val Ser Arg Lys Val Phe
        355                 360                 365

Gly Glu Trp Ser Ile Ile Asn Arg Ala Leu Asp Asn Tyr Tyr Ala Thr
    370                 375                 380

Thr Tyr Pro Ile Lys Pro Arg Glu Lys Ser Glu Lys Tyr Glu Glu Arg
385                 390                 395                 400

Lys Glu Lys Trp Leu Lys Gln Asp Phe Asn Val Ser Leu Ile Gln Thr
                405                 410                 415

Ala Ile Asp Glu Tyr Asp Asn Glu Thr Val Lys Gly Lys Asn Ser Gly
            420                 425                 430

Lys Val Ile Val Asp Tyr Phe Ala Lys Phe Cys Asp Asp Lys Glu Thr
        435                 440                 445

Asp Leu Ile Gln Lys Val Asn Glu Gly Tyr Ile Ala Val Lys Asp Leu
    450                 455                 460

Leu Asn Thr Pro Tyr Pro Glu Asn Glu Lys Leu Gly Ser Asn Lys Asp
465                 470                 475                 480
```

```
Gln Val Lys Gln Ile Lys Ala Phe Met Asp Ser Ile Met Asp Ile Met
                485                 490                 495

His Phe Val Arg Pro Leu Ser Leu Lys Asp Thr Asp Lys Glu Lys Asp
            500                 505                 510

Glu Thr Phe Tyr Ser Leu Phe Thr Pro Leu Tyr Asp His Leu Thr Gln
        515                 520                 525

Thr Ile Ala Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Gln Lys Pro
    530                 535                 540

Tyr Ser Thr Glu Lys Ile Lys Leu Asn Phe Glu Asn Ser Thr Leu Leu
545                 550                 555                 560

Gly Gly Trp Asp Leu Asn Lys Glu Thr Asp Asn Thr Ala Ile Ile Leu
                565                 570                 575

Arg Lys Glu Asn Leu Tyr Tyr Leu Gly Ile Met Asp Lys Arg His Asn
            580                 585                 590

Arg Ile Phe Arg Asn Val Pro Lys Ala Asp Lys Lys Asp Ser Cys Tyr
        595                 600                 605

Glu Lys Met Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro
    610                 615                 620

Lys Val Phe Phe Ser Gln Ser Arg Ile Gln Glu Phe Thr Pro Ser Ala
625                 630                 635                 640

Lys Leu Leu Glu Asn Tyr Glu Asn Glu Thr His Lys Lys Gly Asp Asn
                645                 650                 655

Phe Asn Leu Asn His Cys His Gln Leu Ile Asp Phe Phe Lys Asp Ser
            660                 665                 670

Ile Asn Lys His Glu Asp Trp Lys Asn Phe Asp Phe Arg Phe Ser Ala
        675                 680                 685

Thr Ser Thr Tyr Ala Asp Leu Ser Gly Phe Tyr His Glu Val Glu His
    690                 695                 700

Gln Gly Tyr Lys Ile Ser Phe Gln Ser Ile Ala Asp Ser Phe Ile Asp
705                 710                 715                 720

Asp Leu Val Asn Glu Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys
                725                 730                 735

Asp Phe Ser Pro Phe Ser Lys Gly Lys Pro Asn Leu His Thr Leu Tyr
            740                 745                 750

Trp Lys Met Leu Phe Asp Glu Asn Asn Leu Lys Asp Val Val Tyr Lys
        755                 760                 765

Leu Asn Gly Glu Ala Glu Val Phe Tyr Arg Lys Lys Ser Ile Ala Glu
    770                 775                 780

Lys Asn Thr Thr Ile His Lys Ala Asn Glu Ser Ile Ile Asn Lys Asn
785                 790                 795                 800

Pro Asp Asn Pro Lys Ala Thr Ser Thr Phe Asn Tyr Asp Ile Val Lys
                805                 810                 815
```

```
Asp Lys Arg Tyr Thr Ile Asp Lys Phe Gln Phe His Val Pro Ile Thr
            820                 825                 830

Met Asn Phe Lys Ala Glu Gly Ile Phe Asn Met Asn Gln Arg Val Asn
        835                 840                 845

Gln Phe Leu Lys Ala Asn Pro Asp Ile Asn Ile Ile Gly Ile Asp Arg
    850                 855                 860

Gly Glu Arg His Leu Leu Tyr Tyr Thr Leu Ile Asn Gln Lys Gly Lys
865                 870                 875                 880

Ile Leu Lys Gln Asp Thr Leu Asn Val Ile Ala Asn Glu Lys Gln Lys
                885                 890                 895

Val Asp Tyr His Asn Leu Leu Asp Lys Lys Glu Gly Asp Arg Ala Thr
            900                 905                 910

Ala Arg Gln Glu Trp Gly Val Ile Glu Thr Ile Lys Glu Leu Lys Glu
        915                 920                 925

Gly Tyr Leu Ser Gln Val Ile His Lys Leu Thr Asp Leu Met Ile Glu
    930                 935                 940

Asn Asn Ala Ile Ile Val Met Glu Asp Leu Asn Phe Gly Phe Lys Arg
945                 950                 955                 960

Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu Lys Met
                965                 970                 975

Leu Ile Asp Lys Leu Asn Tyr Leu Val Asp Lys Asn Lys Lys Ala Asn
            980                 985                 990

Glu Leu Gly Gly Leu Leu Asn Ala Phe Gln Leu Ala Asn Lys Phe Glu
        995                 1000                1005

Ser Phe  Gln Lys Met Gly Lys Gln Asn Gly Phe  Ile Phe Tyr Val
    1010                1015                1020

Pro Ala  Trp Asn Thr Ser Lys Thr Asp Pro Ala  Thr Gly Phe Ile
    1025                1030                1035

Asp Phe  Leu Lys Pro Arg Tyr Glu Asn Leu Lys  Gln Ala Lys Asp
    1040                1045                1050

Phe Phe  Glu Lys Phe Asp Ser Ile Arg Leu Asn  Ser Lys Ala Asp
    1055                1060                1065

Tyr Phe  Glu Phe Ala Phe Asp Phe Lys Asn Phe  Thr Gly Lys Ala
    1070                1075                1080

Asp Gly  Gly Arg Thr Lys Trp Thr Val Cys Thr  Thr Asn Glu Asp
    1085                1090                1095

Arg Tyr  Ala Trp Asn Arg Ala Leu Asn Asn Asn  Arg Gly Ser Gln
    1100                1105                1110

Glu Lys  Tyr Asp Ile Thr Ala Glu Leu Lys Ser  Leu Phe Asp Gly
    1115                1120                1125

Lys Val  Asp Tyr Lys Ser Gly Lys Asp Leu Lys  Gln Gln Ile Ala
    1130                1135                1140
```

Ser Gln Glu Leu Ala Asp Phe Phe Arg Thr Leu Met Lys Tyr Leu
1145 1150 1155

Ser Val Thr Leu Ser Leu Arg His Asn Asn Gly Glu Lys Gly Glu
1160 1165 1170

Thr Glu Gln Asp Tyr Ile Leu Ser Pro Val Ala Asp Ser Met Gly
1175 1180 1185

Lys Phe Phe Asp Ser Arg Lys Ala Gly Asp Asp Met Pro Lys Asn
1190 1195 1200

Ala Asp Ala Asn Gly Ala Tyr His Ile Ala Leu Lys Gly Leu Trp
1205 1210 1215

Cys Leu Glu Gln Ile Ser Lys Thr Asp Asp Leu Lys Lys Val Lys
1220 1225 1230

Leu Ala Ile Ser Asn Lys Glu Trp Leu Glu Phe Met Gln Thr Leu
1235 1240 1245

Lys Gly
1250

<210> 1142
<211> 1249
<212> PRT
<213> 稻田鉤端螺旋體

<400> 1142
Met Glu Asp Tyr Ser Gly Phe Val Asn Ile Tyr Ser Ile Gln Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr Leu Glu His Ile Glu
20 25 30

Lys Lys Gly Phe Leu Lys Lys Asp Lys Ile Arg Ala Glu Asp Tyr Lys
35 40 45

Ala Val Lys Lys Ile Ile Asp Lys Tyr His Arg Ala Tyr Ile Glu Glu
50 55 60

Val Phe Asp Ser Val Leu His Gln Lys Lys Lys Lys Asp Lys Thr Arg
65 70 75 80

Phe Ser Thr Gln Phe Ile Lys Glu Ile Lys Glu Phe Ser Glu Leu Tyr
85 90 95

Tyr Lys Thr Glu Lys Asn Ile Pro Asp Lys Glu Arg Leu Glu Ala Leu
100 105 110

Ser Glu Lys Leu Arg Lys Met Leu Val Gly Ala Phe Lys Gly Glu Phe
115 120 125

Ser Glu Glu Val Ala Glu Lys Tyr Lys Asn Leu Phe Ser Lys Glu Leu
130 135 140

Ile Arg Asn Glu Ile Glu Lys Phe Cys Glu Thr Asp Glu Glu Arg Lys
145 150 155 160

Gln Val Ser Asn Phe Lys Ser Phe Thr Thr Asp Glu Lys Lys Ser Thr
165 170 175

Ala Ile Gly Tyr Arg Ile Ile His Gln Asn Leu Pro Lys Phe Leu Asp

180 185 190

Asn Leu Lys Ile Ile Glu Ser Ile Gln Arg Arg Phe Lys Asp Phe Pro
195 200 205

Trp Ser Asp Leu Lys Lys Asn Leu Lys Lys Ile Asp Lys Asn Ile Lys
210 215 220

Leu Thr Glu Tyr Phe Ser Ile Asp Gly Phe Val Asn Val Leu Asn Gln
225 230 235 240

Lys Gly Ile Asp Ala Tyr Asn Thr Ile Leu Gly Gly Lys Ser Glu Glu
245 250 255

Ser Gly Glu Lys Ile Gln Gly Leu Asn Glu Tyr Ile Asn Leu Tyr Arg
260 265 270

Gln Lys Asn Asn Ile Asp Arg Lys Asn Leu Pro Asn Val Lys Ile Leu
275 280 285

Phe Lys Gln Ile Leu Gly Asp Arg Glu Thr Lys Ser Phe Ile Pro Glu
290 295 300

Ala Phe Pro Asp Asp Gln Ser Val Leu Asn Ser Ile Thr Glu Phe Ala
305 310 315 320

Lys Tyr Leu Lys Leu Asp Lys Lys Lys Lys Ser Ile Ile Ala Glu Leu
325 330 335

Lys Lys Phe Leu Ser Ser Phe Asn Arg Tyr Glu Leu Asp Gly Ile Tyr
340 345 350

Leu Ala Asn Asp Asn Ser Leu Ala Ser Ile Ser Thr Phe Leu Phe Asp
355 360 365

Asp Trp Ser Phe Ile Lys Lys Ser Val Ser Phe Lys Tyr Asp Glu Ser
370 375 380

Val Gly Asp Pro Lys Lys Lys Ile Lys Ser Pro Leu Lys Tyr Glu Lys
385 390 395 400

Glu Lys Glu Lys Trp Leu Lys Gln Lys Tyr Tyr Thr Ile Ser Phe Leu
405 410 415

Asn Asp Ala Ile Glu Ser Tyr Ser Lys Ser Gln Asp Glu Lys Arg Val
420 425 430

Lys Ile Arg Leu Glu Ala Tyr Phe Ala Glu Phe Lys Ser Lys Asp Asp
435 440 445

Ala Lys Lys Gln Phe Asp Leu Leu Glu Arg Ile Glu Glu Ala Tyr Ala
450 455 460

Ile Val Glu Pro Leu Leu Gly Ala Glu Tyr Pro Arg Asp Arg Asn Leu
465 470 475 480

Lys Ala Asp Lys Lys Glu Val Gly Lys Ile Lys Asp Phe Leu Asp Ser
485 490 495

Ile Lys Ser Leu Gln Phe Phe Leu Lys Pro Leu Leu Ser Ala Glu Ile
500 505 510

Phe Asp Glu Lys Asp Leu Gly Phe Tyr Asn Gln Leu Glu Gly Tyr Tyr

515 520 525

Glu Glu Ile Asp Ser Ile Gly His Leu Tyr Asn Lys Val Arg Asn Tyr
530 535 540

Leu Thr Gly Lys Ile Tyr Ser Lys Glu Lys Phe Lys Leu Asn Phe Glu
545 550 555 560

Asn Ser Thr Leu Leu Lys Gly Trp Asp Glu Asn Arg Glu Val Ala Asn
565 570 575

Leu Cys Val Ile Phe Arg Glu Asp Gln Lys Tyr Tyr Leu Gly Val Met
580 585 590

Asp Lys Glu Asn Asn Thr Ile Leu Ser Asp Ile Pro Lys Val Lys Pro
595 600 605

Asn Glu Leu Phe Tyr Glu Lys Met Val Tyr Lys Leu Ile Pro Thr Pro
610 615 620

His Met Gln Leu Pro Arg Ile Ile Phe Ser Ser Asp Asn Leu Ser Ile
625 630 635 640

Tyr Asn Pro Ser Lys Ser Ile Leu Lys Ile Arg Glu Ala Lys Ser Phe
645 650 655

Lys Glu Gly Lys Asn Phe Lys Leu Lys Asp Cys His Lys Phe Ile Asp
660 665 670

Phe Tyr Lys Glu Ser Ile Ser Lys Asn Glu Asp Trp Ser Arg Phe Asp
675 680 685

Phe Lys Phe Ser Lys Thr Ser Ser Tyr Glu Asn Ile Ser Glu Phe Tyr
690 695 700

Arg Glu Val Glu Arg Gln Gly Tyr Asn Leu Asp Phe Lys Lys Val Ser
705 710 715 720

Lys Phe Tyr Ile Asp Ser Leu Val Glu Asp Gly Lys Leu Tyr Leu Phe
725 730 735

Gln Ile Tyr Asn Lys Asp Phe Ser Ile Phe Ser Lys Gly Lys Pro Asn
740 745 750

Leu His Thr Ile Tyr Phe Arg Ser Leu Phe Ser Lys Glu Asn Leu Lys
755 760 765

Asp Val Cys Leu Lys Leu Asn Gly Glu Ala Glu Met Phe Phe Arg Lys
770 775 780

Lys Ser Ile Asn Tyr Asp Glu Lys Lys Lys Arg Glu Gly His His Pro
785 790 795 800

Glu Leu Phe Glu Lys Leu Lys Tyr Pro Ile Leu Lys Asp Lys Arg Tyr
805 810 815

Ser Glu Asp Lys Phe Gln Phe His Leu Pro Ile Ser Leu Asn Phe Lys
820 825 830

Ser Lys Glu Arg Leu Asn Phe Asn Leu Lys Val Asn Glu Phe Leu Lys
835 840 845

Arg Asn Lys Asp Ile Asn Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn

850 855 860

Leu Leu Tyr Leu Val Met Ile Asn Gln Lys Gly Glu Ile Leu Lys Gln
865 870 875 880

Thr Leu Leu Asp Ser Met Gln Ser Gly Lys Gly Arg Pro Glu Ile Asn
885 890 895

Tyr Lys Glu Lys Leu Gln Glu Lys Glu Ile Glu Arg Asp Lys Ala Arg
900 905 910

Lys Ser Trp Gly Thr Val Glu Asn Ile Lys Glu Leu Lys Glu Gly Tyr
915 920 925

Leu Ser Ile Val Ile His Gln Ile Ser Lys Leu Met Val Glu Asn Asn
930 935 940

Ala Ile Val Val Leu Glu Asp Leu Asn Ile Gly Phe Lys Arg Gly Arg
945 950 955 960

Gln Lys Val Glu Arg Gln Val Tyr Gln Lys Phe Glu Lys Met Leu Ile
965 970 975

Asp Lys Leu Asn Phe Leu Val Phe Lys Glu Asn Lys Pro Thr Glu Pro
980 985 990

Gly Gly Val Leu Lys Ala Tyr Gln Leu Thr Asp Glu Phe Gln Ser Phe
995 1000 1005

Glu Lys Leu Ser Lys Gln Thr Gly Phe Leu Phe Tyr Val Pro Ser
1010 1015 1020

Trp Asn Thr Ser Lys Ile Asp Pro Arg Thr Gly Phe Ile Asp Phe
1025 1030 1035

Leu His Pro Ala Tyr Glu Asn Ile Glu Lys Ala Lys Gln Trp Ile
1040 1045 1050

Asn Lys Phe Asp Ser Ile Arg Phe Asn Ser Lys Met Asp Trp Phe
1055 1060 1065

Glu Phe Thr Ala Asp Thr Arg Lys Phe Ser Glu Asn Leu Met Leu
1070 1075 1080

Gly Lys Asn Arg Val Trp Val Ile Cys Thr Thr Asn Val Glu Arg
1085 1090 1095

Tyr Phe Thr Ser Lys Thr Ala Asn Ser Ser Ile Gln Tyr Asn Ser
1100 1105 1110

Ile Gln Ile Thr Glu Lys Leu Lys Glu Leu Phe Val Asp Ile Pro
1115 1120 1125

Phe Ser Asn Gly Gln Asp Leu Lys Pro Glu Ile Leu Arg Lys Asn
1130 1135 1140

Asp Ala Val Phe Phe Lys Ser Leu Leu Phe Tyr Ile Lys Thr Thr
1145 1150 1155

Leu Ser Leu Arg Gln Asn Asn Gly Lys Lys Gly Glu Glu Glu Lys
1160 1165 1170

Asp Phe Ile Leu Ser Pro Val Val Asp Ser Lys Gly Arg Phe Phe

1175 1180 1185

Asn Ser Leu Glu Ala Ser Asp Asp Glu Pro Lys Asp Ala Asp Ala
1190 1195 1200

Asn Gly Ala Tyr His Ile Ala Leu Lys Gly Leu Met Asn Leu Leu
1205 1210 1215

Val Leu Asn Glu Thr Lys Glu Glu Asn Leu Ser Arg Pro Lys Trp
1220 1225 1230

Lys Ile Lys Asn Lys Asp Trp Leu Glu Phe Val Trp Glu Arg Asn
1235 1240 1245

Arg

<210> 1143
<211> 1230
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 1143
Met His Glu Asn Asn Gly Lys Ile Ala Asp Asn Phe Ile Gly Ile Tyr
1 5 10 15

Pro Val Ser Lys Thr Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr
20 25 30

Gln Glu Tyr Ile Glu Lys His Gly Ile Leu Asp Glu Asp Leu Lys Arg
35 40 45

Ala Gly Asp Tyr Lys Ser Val Lys Lys Ile Ile Asp Ala Tyr His Lys
50 55 60

Tyr Phe Ile Asp Glu Ala Leu Asn Gly Ile Gln Leu Asp Gly Leu Lys
65 70 75 80

Asn Tyr Tyr Glu Leu Tyr Glu Lys Lys Arg Asp Asn Asn Glu Glu Lys
85 90 95

Glu Phe Gln Lys Ile Gln Met Ser Leu Arg Lys Gln Ile Val Lys Arg
100 105 110

Phe Ser Glu His Pro Gln Tyr Lys Tyr Leu Phe Lys Lys Glu Leu Ile
115 120 125

Lys Asn Val Leu Pro Glu Phe Thr Lys Asp Asn Ala Glu Glu Gln Thr
130 135 140

Leu Val Lys Ser Phe Gln Glu Phe Thr Thr Tyr Phe Glu Gly Phe His
145 150 155 160

Gln Asn Arg Lys Asn Met Tyr Ser Asp Glu Glu Lys Ser Thr Ala Ile
165 170 175

Ala Tyr Arg Val Val His Gln Asn Leu Pro Lys Tyr Ile Asp Asn Met
180 185 190

Arg Ile Phe Ser Met Ile Leu Asn Thr Asp Ile Arg Ser Asp Leu Thr
195 200 205

```
Glu Leu Phe Asn Asn Leu Lys Thr Lys Met Asp Ile Thr Ile Val Glu
    210                 215                 220

Glu Tyr Phe Ala Ile Asp Gly Phe Asn Lys Val Val Asn Gln Lys Gly
225                 230                 235                 240

Ile Asp Val Tyr Asn Thr Ile Leu Gly Ala Phe Ser Thr Asp Asp Asn
                245                 250                 255

Thr Lys Ile Lys Gly Leu Asn Glu Tyr Ile Asn Leu Tyr Asn Gln Lys
            260                 265                 270

Asn Lys Ala Lys Leu Pro Lys Leu Lys Pro Leu Phe Lys Gln Ile Leu
        275                 280                 285

Ser Asp Arg Asp Lys Ile Ser Phe Ile Pro Glu Gln Phe Asp Ser Asp
    290                 295                 300

Thr Glu Val Leu Glu Ala Val Asp Met Phe Tyr Asn Arg Leu Leu Gln
305                 310                 315                 320

Phe Val Ile Glu Asn Glu Gly Gln Ile Thr Ile Ser Lys Leu Leu Thr
                325                 330                 335

Asn Phe Ser Ala Tyr Asp Leu Asn Lys Ile Tyr Val Lys Asn Asp Thr
            340                 345                 350

Thr Ile Ser Ala Ile Ser Asn Asp Leu Phe Asp Asp Trp Ser Tyr Ile
        355                 360                 365

Ser Lys Ala Val Arg Glu Asn Tyr Asp Ser Glu Asn Val Asp Lys Asn
    370                 375                 380

Lys Arg Ala Ala Ala Tyr Glu Glu Lys Lys Glu Lys Ala Leu Ser Lys
385                 390                 395                 400

Ile Lys Met Tyr Ser Ile Glu Glu Leu Asn Phe Phe Val Lys Lys Tyr
                405                 410                 415

Ser Cys Asn Glu Cys His Ile Glu Gly Tyr Phe Glu Arg Arg Ile Leu
            420                 425                 430

Glu Ile Leu Asp Lys Met Arg Tyr Ala Tyr Glu Ser Cys Lys Ile Leu
        435                 440                 445

His Asp Lys Gly Leu Ile Asn Asn Ile Ser Leu Cys Gln Asp Arg Gln
    450                 455                 460

Ala Ile Ser Glu Leu Lys Asp Phe Leu Asp Ser Ile Lys Glu Val Gln
465                 470                 475                 480

Trp Leu Leu Lys Pro Leu Met Ile Gly Gln Glu Gln Ala Asp Lys Glu
                485                 490                 495

Glu Ala Phe Tyr Thr Glu Leu Leu Arg Ile Trp Glu Glu Leu Glu Pro
            500                 505                 510

Ile Thr Leu Leu Tyr Asn Lys Val Arg Asn Tyr Val Thr Lys Lys Pro
        515                 520                 525

Tyr Thr Leu Glu Lys Val Lys Leu Asn Phe Tyr Lys Ser Thr Leu Leu
    530                 535                 540
```

```
Asp Gly Trp Asp Lys Asn Lys Glu Lys Asp Asn Leu Gly Ile Ile Leu
545                 550                 555                 560

Leu Lys Asp Gly Gln Tyr Tyr Leu Gly Ile Met Asn Arg Arg Asn Asn
                565                 570                 575

Lys Ile Ala Asp Asp Ala Pro Leu Ala Lys Thr Asp Asn Val Tyr Arg
            580                 585                 590

Lys Met Glu Tyr Lys Leu Leu Thr Lys Val Ser Ala Asn Leu Pro Arg
        595                 600                 605

Ile Phe Leu Lys Asp Lys Tyr Asn Pro Ser Glu Glu Met Leu Glu Lys
    610                 615                 620

Tyr Glu Lys Gly Thr His Leu Lys Gly Glu Asn Phe Cys Ile Asp Asp
625                 630                 635                 640

Cys Arg Glu Leu Ile Asp Phe Phe Lys Lys Gly Ile Lys Gln Tyr Glu
                645                 650                 655

Asp Trp Gly Gln Phe Asp Phe Lys Phe Ser Asp Thr Glu Ser Tyr Asp
            660                 665                 670

Asp Ile Ser Ala Phe Tyr Lys Glu Val Glu His Gln Gly Tyr Lys Ile
        675                 680                 685

Thr Phe Arg Asp Ile Asp Glu Thr Tyr Ile Asp Ser Leu Val Asn Glu
    690                 695                 700

Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Pro Tyr
705                 710                 715                 720

Ser Lys Gly Thr Lys Asn Leu His Thr Leu Tyr Trp Glu Met Leu Phe
                725                 730                 735

Ser Gln Gln Asn Leu Gln Asn Ile Val Tyr Lys Leu Asn Gly Asn Ala
            740                 745                 750

Glu Ile Phe Tyr Arg Lys Ala Ser Ile Asn Gln Lys Asp Val Val Val
        755                 760                 765

His Lys Ala Asp Leu Pro Ile Lys Asn Lys Asp Pro Gln Asn Ser Lys
    770                 775                 780

Lys Glu Ser Met Phe Asp Tyr Asp Ile Ile Lys Asp Lys Arg Phe Thr
785                 790                 795                 800

Cys Asp Lys Tyr Gln Phe His Val Pro Ile Thr Met Asn Phe Lys Ala
                805                 810                 815

Leu Gly Glu Asn His Phe Asn Arg Lys Val Asn Arg Leu Ile His Asp
            820                 825                 830

Ala Glu Asn Met His Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu
        835                 840                 845

Ile Tyr Leu Cys Met Ile Asp Met Lys Gly Asn Ile Val Lys Gln Ile
    850                 855                 860

Ser Leu Asn Glu Ile Ile Ser Tyr Asp Lys Asn Lys Leu Glu His Lys
865                 870                 875                 880
```

```
Arg Asn Tyr His Gln Leu Leu Lys Thr Arg Glu Asp Glu Asn Lys Ser
                885                 890                 895

Ala Arg Gln Ser Trp Gln Thr Ile His Thr Ile Lys Glu Leu Lys Glu
            900                 905                 910

Gly Tyr Leu Ser Gln Val Ile His Val Ile Thr Asp Leu Met Val Glu
        915                 920                 925

Tyr Asn Ala Ile Val Val Leu Glu Asp Leu Asn Phe Gly Phe Lys Gln
    930                 935                 940

Gly Arg Gln Lys Phe Glu Arg Gln Val Tyr Gln Lys Phe Glu Lys Met
945                 950                 955                 960

Leu Ile Asp Lys Leu Asn Tyr Leu Val Asp Lys Ser Lys Gly Met Asp
                965                 970                 975

Glu Asp Gly Gly Leu Leu His Ala Tyr Gln Leu Thr Asp Glu Phe Lys
            980                 985                 990

Ser Phe Lys Gln Leu Gly Lys Gln Ser Gly Phe Leu Tyr Tyr Ile Pro
        995                 1000                1005

Ala Trp  Asn Thr Ser Lys Leu  Asp Pro Thr Thr Gly  Phe Val Asn
    1010                1015                1020

Leu Phe  Tyr Thr Lys Tyr Glu  Ser Val Glu Lys Ser  Lys Glu Phe
    1025                1030                1035

Ile Asn  Asn Phe Thr Ser Ile  Leu Tyr Asn Gln Glu  Arg Glu Tyr
    1040                1045                1050

Phe Glu  Phe Leu Phe Asp Tyr  Ser Ala Phe Thr Ser  Lys Ala Glu
    1055                1060                1065

Gly Ser  Arg Leu Lys Trp Thr  Val Cys Ser Lys Gly  Glu Arg Val
    1070                1075                1080

Glu Thr  Tyr Arg Asn Pro Lys  Lys Asn Asn Glu Trp  Asp Thr Gln
    1085                1090                1095

Lys Ile  Asp Leu Thr Phe Glu  Leu Lys Lys Leu Phe  Asn Asp Tyr
    1100                1105                1110

Ser Ile  Ser Leu Leu Asp Gly  Asp Leu Arg Glu Gln  Met Gly Lys
    1115                1120                1125

Ile Asp  Lys Ala Asp Phe Tyr  Lys Lys Phe Met Lys  Leu Phe Ala
    1130                1135                1140

Leu Ile  Val Gln Met Arg Asn  Ser Asp Glu Arg Glu  Asp Lys Leu
    1145                1150                1155

Ile Ser  Pro Val Leu Asn Lys  Tyr Gly Ala Phe Phe  Glu Thr Gly
    1160                1165                1170

Lys Asn  Glu Arg Met Pro Leu  Asp Ala Asp Ala Asn  Gly Ala Tyr
    1175                1180                1185

Asn Ile  Ala Arg Lys Gly Leu  Trp Ile Ile Glu Lys  Ile Lys Asn
    1190                1195                1200
```

```
Thr Asp Val Glu Gln Leu Asp Lys Val Lys Leu Thr Ile Ser Asn
    1205                1210                1215

Lys Glu Trp Leu Gln Tyr Ala Gln Glu His Ile Leu
    1220                1225                1230
```

<210> 1144
<211> 1206
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 1144

```
Met Tyr Tyr Glu Ser Leu Thr Lys Gln Tyr Pro Val Ser Lys Thr Ile
1               5                   10                  15

Arg Asn Glu Leu Ile Pro Ile Gly Lys Thr Leu Asp Asn Ile Arg Gln
            20                  25                  30

Asn Asn Ile Leu Glu Ser Asp Val Lys Arg Lys Gln Asn Tyr Glu His
        35                  40                  45

Val Lys Gly Ile Leu Asp Glu Tyr His Lys Gln Leu Ile Asn Glu Ala
    50                  55                  60

Leu Asp Asn Cys Thr Leu Pro Ser Leu Lys Ile Ala Ala Glu Ile Tyr
65                  70                  75                  80

Leu Lys Asn Gln Lys Glu Val Ser Asp Arg Glu Asp Phe Asn Lys Thr
                85                  90                  95

Gln Asp Leu Leu Arg Lys Glu Val Val Glu Lys Leu Lys Ala His Glu
            100                 105                 110

Asn Phe Thr Lys Ile Gly Lys Lys Asp Ile Leu Asp Leu Leu Glu Lys
        115                 120                 125

Leu Pro Ser Ile Ser Glu Asp Asp Tyr Asn Ala Leu Glu Ser Phe Arg
    130                 135                 140

Asn Phe Tyr Thr Tyr Phe Thr Ser Tyr Asn Lys Val Arg Glu Asn Leu
145                 150                 155                 160

Tyr Ser Asp Lys Glu Lys Ser Ser Thr Val Ala Tyr Arg Leu Ile Asn
                165                 170                 175

Glu Asn Phe Pro Lys Phe Leu Asp Asn Val Lys Ser Tyr Arg Phe Val
            180                 185                 190

Lys Thr Ala Gly Ile Leu Ala Asp Gly Leu Gly Glu Glu Glu Gln Asp
        195                 200                 205

Ser Leu Phe Ile Val Glu Thr Phe Asn Lys Thr Leu Thr Gln Asp Gly
    210                 215                 220

Ile Asp Thr Tyr Asn Ser Gln Val Gly Lys Ile Asn Ser Ser Ile Asn
225                 230                 235                 240

Leu Tyr Asn Gln Lys Asn Gln Lys Ala Asn Gly Phe Arg Lys Ile Pro
                245                 250                 255

Lys Met Lys Met Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Glu Ser
```

260 265 270

Phe Ile Asp Glu Phe Gln Ser Asp Glu Val Leu Ile Asp Asn Val Glu
275 280 285

Ser Tyr Gly Ser Val Leu Ile Glu Ser Leu Lys Ser Ser Lys Val Ser
290 295 300

Ala Phe Phe Asp Ala Leu Arg Glu Ser Lys Gly Lys Asn Val Tyr Val
305 310 315 320

Lys Asn Asp Leu Ala Lys Thr Ala Met Ser Asn Ile Val Phe Glu Asn
325 330 335

Trp Arg Thr Phe Asp Asp Leu Leu Asn Gln Glu Tyr Asp Leu Ala Asn
340 345 350

Glu Asn Lys Lys Lys Asp Asp Lys Tyr Phe Glu Lys Arg Gln Lys Glu
355 360 365

Leu Lys Lys Asn Lys Ser Tyr Ser Leu Glu His Leu Cys Asn Leu Ser
370 375 380

Glu Asp Ser Cys Asn Leu Ile Glu Asn Tyr Ile His Gln Ile Ser Asp
385 390 395 400

Asp Ile Glu Asn Ile Ile Ile Asn Asn Glu Thr Phe Leu Arg Ile Val
405 410 415

Ile Asn Glu His Asp Arg Ser Arg Lys Leu Ala Lys Asn Arg Lys Ala
420 425 430

Val Lys Ala Ile Lys Asp Phe Leu Asp Ser Ile Lys Val Leu Glu Arg
435 440 445

Glu Leu Lys Leu Ile Asn Ser Ser Gly Gln Glu Leu Glu Lys Asp Leu
450 455 460

Ile Val Tyr Ser Ala His Glu Glu Leu Leu Val Glu Leu Lys Gln Val
465 470 475 480

Asp Ser Leu Tyr Asn Met Thr Arg Asn Tyr Leu Thr Lys Lys Pro Phe
485 490 495

Ser Thr Glu Lys Val Lys Leu Asn Phe Asn Arg Ser Thr Leu Leu Asn
500 505 510

Gly Trp Asp Arg Asn Lys Glu Thr Asp Asn Leu Gly Val Leu Leu Leu
515 520 525

Lys Asp Gly Lys Tyr Tyr Leu Gly Ile Met Asn Thr Ser Ala Asn Lys
530 535 540

Ala Phe Val Asn Pro Pro Val Ala Lys Thr Glu Lys Val Phe Lys Lys
545 550 555 560

Val Asp Tyr Lys Leu Leu Pro Val Pro Asn Gln Met Leu Pro Lys Val
565 570 575

Phe Phe Ala Lys Ser Asn Ile Asp Phe Tyr Asn Pro Ser Ser Glu Ile
580 585 590

Tyr Ser Asn Tyr Lys Lys Gly Thr His Lys Lys Gly Asn Met Phe Ser

595 600 605

Leu Glu Asp Cys His Asn Leu Ile Asp Phe Phe Lys Glu Ser Ile Ser
610 615 620

Lys His Glu Asp Trp Ser Lys Phe Gly Phe Lys Phe Ser Asp Thr Ala
625 630 635 640

Ser Tyr Asn Asp Ile Ser Glu Phe Tyr Arg Glu Val Glu Lys Gln Gly
645 650 655

Tyr Lys Leu Thr Tyr Thr Asp Ile Asp Glu Thr Tyr Ile Asn Asp Leu
660 665 670

Ile Glu Arg Asn Glu Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe
675 680 685

Ser Met Tyr Ser Lys Gly Lys Leu Asn Leu His Thr Leu Tyr Phe Met
690 695 700

Met Leu Phe Asp Gln Arg Asn Ile Asp Asp Val Val Tyr Lys Leu Asn
705 710 715 720

Gly Glu Ala Glu Val Phe Tyr Arg Pro Ala Ser Ile Ser Glu Asp Glu
725 730 735

Leu Ile Ile His Lys Ala Gly Glu Glu Ile Lys Asn Lys Asn Pro Asn
740 745 750

Arg Ala Arg Thr Lys Glu Thr Ser Thr Phe Ser Tyr Asp Ile Val Lys
755 760 765

Asp Lys Arg Tyr Ser Lys Asp Lys Phe Thr Leu His Ile Pro Ile Thr
770 775 780

Met Asn Phe Gly Val Asp Glu Val Lys Arg Phe Asn Asp Ala Val Asn
785 790 795 800

Ser Ala Ile Arg Ile Asp Glu Asn Val Asn Val Ile Gly Ile Asp Arg
805 810 815

Gly Glu Arg Asn Leu Leu Tyr Val Val Val Ile Asp Ser Lys Gly Asn
820 825 830

Ile Leu Glu Gln Ile Ser Leu Asn Ser Ile Ile Asn Lys Glu Tyr Asp
835 840 845

Ile Glu Thr Asp Tyr His Ala Leu Leu Asp Glu Arg Glu Gly Gly Arg
850 855 860

Asp Lys Ala Arg Lys Asp Trp Asn Thr Val Glu Asn Ile Arg Asp Leu
865 870 875 880

Lys Ala Gly Tyr Leu Ser Gln Val Val Asn Val Val Ala Lys Leu Val
885 890 895

Leu Lys Tyr Asn Ala Ile Ile Cys Leu Glu Asp Leu Asn Phe Gly Phe
900 905 910

Lys Arg Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu
915 920 925

Lys Met Leu Ile Asp Lys Leu Asn Tyr Leu Val Ile Asp Lys Ser Arg

930 935 940

Glu Gln Thr Ser Pro Lys Glu Leu Gly Gly Ala Leu Asn Ala Leu Gln
945 950 955 960

Leu Thr Ser Lys Phe Lys Ser Phe Lys Glu Leu Gly Lys Gln Ser Gly
965 970 975

Val Ile Tyr Tyr Val Pro Ala Tyr Leu Thr Ser Lys Ile Asp Pro Thr
980 985 990

Thr Gly Phe Ala Asn Leu Phe Tyr Met Lys Cys Glu Asn Val Glu Lys
995 1000 1005

Ser Lys Arg Phe Phe Asp Gly Phe Asp Phe Ile Arg Phe Asn Ala
1010 1015 1020

Leu Glu Asn Val Phe Glu Phe Gly Phe Asp Tyr Arg Ser Phe Thr
1025 1030 1035

Gln Arg Ala Cys Gly Ile Asn Ser Lys Trp Thr Val Cys Thr Asn
1040 1045 1050

Gly Glu Arg Ile Ile Lys Tyr Arg Asn Pro Asp Lys Asn Asn Met
1055 1060 1065

Phe Asp Glu Lys Val Val Val Val Thr Asp Glu Met Lys Asn Leu
1070 1075 1080

Phe Glu Gln Tyr Lys Ile Pro Tyr Glu Asp Gly Arg Asn Val Lys
1085 1090 1095

Asp Met Ile Ile Ser Asn Glu Glu Ala Glu Phe Tyr Arg Arg Leu
1100 1105 1110

Tyr Arg Leu Leu Gln Gln Thr Leu Gln Met Arg Asn Ser Thr Ser
1115 1120 1125

Asp Gly Thr Arg Asp Tyr Ile Ile Ser Pro Val Lys Asn Lys Arg
1130 1135 1140

Glu Ala Tyr Phe Asn Ser Glu Leu Ser Asp Gly Ser Val Pro Lys
1145 1150 1155

Asp Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala Arg Lys Gly Leu
1160 1165 1170

Trp Val Leu Glu Gln Ile Arg Gln Lys Ser Glu Gly Glu Lys Ile
1175 1180 1185

Asn Leu Ala Met Thr Asn Ala Glu Trp Leu Glu Tyr Ala Gln Thr
1190 1195 1200

His Leu Leu
1205

<210> 1145
<211> 1206
<212> PRT
<213> 丁酸弧菌屬某種

<400> 1145
Met Tyr Tyr Gln Asn Leu Thr Lys Lys Tyr Pro Val Ser Lys Thr Ile
1 5 10 15

```
Arg Asn Glu Leu Ile Pro Ile Gly Lys Thr Leu Glu Asn Ile Arg Lys
            20                  25                  30

Asn Asn Ile Leu Glu Ser Asp Val Lys Arg Lys Gln Asp Tyr Glu His
        35                  40                  45

Val Lys Gly Ile Met Asp Glu Tyr His Lys Gln Leu Ile Asn Glu Ala
    50                  55                  60

Leu Asp Asn Tyr Met Leu Pro Ser Leu Asn Gln Ala Ala Glu Ile Tyr
65                  70                  75                  80

Leu Lys Lys His Val Asp Val Glu Asp Arg Glu Glu Phe Lys Lys Thr
                85                  90                  95

Gln Asp Leu Leu Arg Arg Glu Val Thr Gly Arg Leu Lys Glu His Glu
            100                 105                 110

Asn Tyr Thr Lys Ile Gly Lys Lys Asp Ile Leu Asp Leu Leu Glu Lys
        115                 120                 125

Leu Pro Ser Ile Ser Glu Glu Asp Tyr Asn Ala Leu Glu Ser Phe Arg
    130                 135                 140

Asn Phe Tyr Thr Tyr Phe Thr Ser Tyr Asn Lys Val Arg Glu Asn Leu
145                 150                 155                 160

Tyr Ser Asp Glu Glu Lys Ser Ser Thr Val Ala Tyr Arg Leu Ile Asn
                165                 170                 175

Glu Asn Leu Pro Lys Phe Leu Asp Asn Ile Lys Ser Tyr Ala Phe Val
            180                 185                 190

Lys Ala Ala Gly Val Leu Ala Asp Cys Ile Glu Glu Glu Glu Gln Asp
        195                 200                 205

Ala Leu Phe Met Val Glu Thr Phe Asn Met Thr Leu Thr Gln Glu Gly
    210                 215                 220

Ile Asp Met Tyr Asn Tyr Gln Ile Gly Lys Val Asn Ser Ala Ile Asn
225                 230                 235                 240

Leu Tyr Asn Gln Lys Asn His Lys Val Glu Glu Phe Lys Lys Ile Pro
                245                 250                 255

Lys Met Lys Val Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Glu Val
            260                 265                 270

Phe Ile Gly Glu Phe Lys Asp Asp Glu Thr Leu Leu Ser Ser Ile Gly
        275                 280                 285

Ala Tyr Gly Asn Val Leu Met Thr Tyr Leu Lys Ser Glu Lys Ile Asn
    290                 295                 300

Ile Phe Phe Asp Ala Leu Arg Glu Ser Glu Gly Lys Asn Val Tyr Val
305                 310                 315                 320

Lys Asn Asp Leu Ser Lys Thr Thr Met Ser Asn Ile Val Phe Gly Ser
                325                 330                 335

Trp Ser Ala Phe Asp Glu Leu Leu Asn Gln Glu Tyr Asp Leu Ala Asn
            340                 345                 350
```

```
Glu Asn Lys Lys Lys Asp Asp Lys Tyr Phe Glu Lys Arg Gln Lys Glu
        355                 360                 365

Leu Lys Lys Asn Lys Ser Tyr Thr Leu Glu Gln Met Ser Asn Leu Ser
    370                 375                 380

Lys Glu Asp Ile Ser Pro Ile Glu Asn Tyr Ile Glu Arg Ile Ser Glu
385                 390                 395                 400

Asp Ile Glu Lys Ile Cys Ile Tyr Asn Gly Glu Phe Glu Lys Ile Val
                405                 410                 415

Val Asn Glu His Asp Ser Ser Arg Lys Leu Ser Lys Asn Ile Lys Ala
            420                 425                 430

Val Lys Val Ile Lys Asp Tyr Leu Asp Ser Ile Lys Glu Leu Glu His
        435                 440                 445

Asp Ile Lys Leu Ile Asn Gly Ser Gly Gln Glu Leu Glu Lys Asn Leu
    450                 455                 460

Val Val Tyr Val Gly Gln Glu Glu Ala Leu Glu Gln Leu Arg Pro Val
465                 470                 475                 480

Asp Ser Leu Tyr Asn Leu Thr Arg Asn Tyr Leu Thr Lys Lys Pro Phe
                485                 490                 495

Ser Thr Glu Lys Val Lys Leu Asn Phe Asn Lys Ser Thr Leu Leu Asn
            500                 505                 510

Gly Trp Asp Lys Asn Lys Glu Thr Asp Asn Leu Gly Ile Leu Phe Phe
        515                 520                 525

Lys Asp Gly Lys Tyr Tyr Leu Gly Ile Met Asn Thr Thr Ala Asn Lys
    530                 535                 540

Ala Phe Val Asn Pro Pro Ala Ala Lys Thr Glu Asn Val Phe Lys Lys
545                 550                 555                 560

Val Asp Tyr Lys Leu Leu Pro Gly Ser Asn Lys Met Leu Pro Lys Val
                565                 570                 575

Phe Phe Ala Lys Ser Asn Ile Gly Tyr Tyr Asn Pro Ser Thr Glu Leu
            580                 585                 590

Tyr Ser Asn Tyr Lys Lys Gly Thr His Lys Lys Gly Pro Ser Phe Ser
        595                 600                 605

Ile Asp Asp Cys His Asn Leu Ile Asp Phe Phe Lys Glu Ser Ile Lys
    610                 615                 620

Lys His Glu Asp Trp Ser Lys Phe Gly Phe Glu Phe Ser Asp Thr Ala
625                 630                 635                 640

Asp Tyr Arg Asp Ile Ser Glu Phe Tyr Arg Glu Val Glu Lys Gln Gly
                645                 650                 655

Tyr Lys Leu Thr Phe Thr Asp Ile Asp Glu Ser Tyr Ile Asn Asp Leu
            660                 665                 670

Ile Glu Lys Asn Glu Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe
        675                 680                 685
```

```
Ser Glu Tyr Ser Lys Gly Lys Leu Asn Leu His Thr Leu Tyr Phe Met
    690                 695                 700

Met Leu Phe Asp Gln Arg Asn Leu Asp Asn Val Val Tyr Lys Leu Asn
705                 710                 715                 720

Gly Glu Ala Glu Val Phe Tyr Arg Pro Ala Ser Ile Ala Glu Asn Glu
                725                 730                 735

Leu Val Ile His Lys Ala Gly Glu Gly Ile Lys Asn Lys Asn Pro Asn
            740                 745                 750

Arg Ala Lys Val Lys Glu Thr Ser Thr Phe Ser Tyr Asp Ile Val Lys
        755                 760                 765

Asp Lys Arg Tyr Ser Lys Tyr Lys Phe Thr Leu His Ile Pro Ile Thr
    770                 775                 780

Met Asn Phe Gly Val Asp Glu Val Arg Arg Phe Asn Asp Val Ile Asn
785                 790                 795                 800

Asn Ala Leu Arg Thr Asp Asp Asn Val Asn Val Ile Gly Ile Asp Arg
                805                 810                 815

Gly Glu Arg Asn Leu Leu Tyr Val Val Val Ile Asn Ser Glu Gly Lys
            820                 825                 830

Ile Leu Glu Gln Ile Ser Leu Asn Ser Ile Ile Asn Lys Glu Tyr Asp
        835                 840                 845

Ile Glu Thr Asn Tyr His Ala Leu Leu Asp Glu Arg Glu Asp Asp Arg
    850                 855                 860

Asn Lys Ala Arg Lys Asp Trp Asn Thr Ile Glu Asn Ile Lys Glu Leu
865                 870                 875                 880

Lys Thr Gly Tyr Leu Ser Gln Val Val Asn Val Val Ala Lys Leu Val
                885                 890                 895

Leu Lys Tyr Asn Ala Ile Ile Cys Leu Glu Asp Leu Asn Phe Gly Phe
            900                 905                 910

Lys Arg Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu
        915                 920                 925

Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu Val Ile Asp Lys Ser Arg
    930                 935                 940

Glu Gln Val Ser Pro Glu Lys Met Gly Gly Ala Leu Asn Ala Leu Gln
945                 950                 955                 960

Leu Thr Ser Lys Phe Lys Ser Phe Ala Glu Leu Gly Lys Gln Ser Gly
                965                 970                 975

Ile Ile Tyr Tyr Val Pro Ala Tyr Leu Thr Ser Lys Ile Asp Pro Thr
            980                 985                 990

Thr Gly Phe Val Asn Leu Phe Tyr Ile Lys Tyr Glu Asn Ile Glu Lys
        995                 1000                1005

Ala Lys Gln Phe Phe Asp Gly Phe Asp Phe Ile Arg Phe Asn Lys
    1010                1015                1020
```

Lys Asp Asp Met Phe Glu Phe Ser Phe Asp Tyr Lys Ser Phe Thr
1025 1030 1035

Gln Lys Ala Cys Gly Ile Arg Ser Lys Trp Ile Val Tyr Thr Asn
1040 1045 1050

Gly Glu Arg Ile Ile Lys Tyr Pro Asn Pro Glu Lys Asn Asn Leu
1055 1060 1065

Phe Asp Glu Lys Val Ile Asn Val Thr Asp Glu Ile Lys Gly Leu
1070 1075 1080

Phe Lys Gln Tyr Arg Ile Pro Tyr Glu Asn Gly Glu Asp Ile Lys
1085 1090 1095

Glu Ile Ile Ile Ser Lys Ala Glu Ala Asp Phe Tyr Lys Arg Leu
1100 1105 1110

Phe Arg Leu Leu His Gln Thr Leu Gln Met Arg Asn Ser Thr Ser
1115 1120 1125

Asp Gly Thr Arg Asp Tyr Ile Ile Ser Pro Val Lys Asn Asp Arg
1130 1135 1140

Gly Glu Phe Phe Cys Ser Glu Phe Ser Glu Gly Thr Met Pro Lys
1145 1150 1155

Asp Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala Arg Lys Gly Leu
1160 1165 1170

Trp Val Leu Glu Gln Ile Arg Gln Lys Asp Glu Gly Glu Lys Val
1175 1180 1185

Asn Leu Ser Met Thr Asn Ala Glu Trp Leu Lys Tyr Ala Gln Leu
1190 1195 1200

His Leu Leu
1205

<210> 1146
<211> 1205
<212> PRT
<213> 口小桿菌屬某種

<400> 1146
Met Tyr Tyr Asp Gly Leu Thr Lys Gln Tyr Ala Leu Ser Lys Thr Ile
1 5 10 15

Arg Asn Glu Leu Val Pro Ile Gly Lys Thr Leu Asp Asn Ile Lys Lys
20 25 30

Asn Arg Ile Leu Glu Ala Asp Ile Lys Arg Lys Ser Asp Tyr Glu His
35 40 45

Val Lys Lys Leu Met Asp Met Tyr His Lys Lys Ile Ile Asn Glu Ala
50 55 60

Leu Asp Asn Phe Lys Leu Ser Val Leu Glu Asp Ala Ala Asp Ile Tyr
65 70 75 80

Phe Asn Lys Gln Asn Asp Glu Arg Asp Ile Asp Ala Phe Leu Lys Ile
85 90 95

Gln Asp Lys Leu Arg Lys Glu Ile Val Glu Gln Leu Lys Gly His Thr

100 105 110

Asp Tyr Ser Lys Val Gly Asn Lys Asp Phe Leu Gly Leu Leu Lys Ala
115 120 125

Ala Ser Thr Glu Glu Asp Arg Ile Leu Ile Glu Ser Phe Asp Asn Phe
130 135 140

Tyr Thr Tyr Phe Thr Ser Tyr Asn Lys Val Arg Ser Asn Leu Tyr Ser
145 150 155 160

Ala Glu Asp Lys Ser Ser Thr Val Ala Tyr Arg Leu Ile Asn Glu Asn
165 170 175

Leu Pro Lys Phe Phe Asp Asn Ile Lys Ala Tyr Arg Thr Val Arg Asn
180 185 190

Ala Gly Val Ile Ser Gly Asp Met Ser Ile Val Glu Gln Asp Glu Leu
195 200 205

Phe Glu Val Asp Thr Phe Asn His Thr Leu Thr Gln Tyr Gly Ile Asp
210 215 220

Thr Tyr Asn His Met Ile Gly Gln Leu Asn Ser Ala Ile Asn Leu Tyr
225 230 235 240

Asn Gln Lys Met His Gly Ala Gly Ser Phe Lys Lys Leu Pro Lys Met
245 250 255

Lys Glu Leu Tyr Lys Gln Leu Leu Thr Glu Arg Glu Glu Glu Phe Ile
260 265 270

Glu Glu Tyr Thr Asp Asp Glu Val Leu Ile Thr Ser Val His Asn Tyr
275 280 285

Val Ser Tyr Leu Ile Asp Tyr Leu Asn Ser Asp Lys Val Glu Ser Phe
290 295 300

Phe Asp Thr Leu Arg Lys Ser Asp Gly Lys Glu Val Phe Ile Lys Asn
305 310 315 320

Asp Val Ser Lys Thr Thr Met Ser Asn Ile Leu Phe Asp Asn Trp Ser
325 330 335

Thr Ile Asp Asp Leu Ile Asn His Glu Tyr Asp Ser Ala Pro Glu Asn
340 345 350

Val Lys Lys Thr Lys Asp Asp Lys Tyr Phe Glu Lys Arg Gln Lys Asp
355 360 365

Leu Lys Lys Asn Lys Ser Tyr Ser Leu Ser Lys Ile Ala Ala Leu Cys
370 375 380

Arg Asp Thr Thr Ile Leu Glu Lys Tyr Ile Arg Arg Leu Val Asp Asp
385 390 395 400

Ile Glu Lys Ile Tyr Thr Ser Asn Asn Val Phe Ser Asp Ile Val Leu
405 410 415

Ser Lys His Asp Arg Ser Lys Lys Leu Ser Lys Asn Thr Asn Ala Val
420 425 430

Gln Ala Ile Lys Asn Met Leu Asp Ser Ile Lys Asp Phe Glu His Asp

435 440 445

Val Met Leu Ile Asn Gly Ser Gly Gln Glu Ile Lys Lys Asn Leu Asn
450 455 460

Val Tyr Ser Glu Gln Glu Ala Leu Ala Gly Ile Leu Arg Gln Val Asp
465 470 475 480

His Ile Tyr Asn Leu Thr Arg Asn Tyr Leu Thr Lys Lys Pro Phe Ser
485 490 495

Thr Glu Lys Ile Lys Leu Asn Phe Asn Arg Pro Thr Phe Leu Asp Gly
500 505 510

Trp Asp Lys Asn Lys Glu Glu Ala Asn Leu Gly Ile Leu Leu Ile Lys
515 520 525

Asp Asn Arg Tyr Tyr Leu Gly Ile Met Asn Thr Ser Ser Asn Lys Ala
530 535 540

Phe Val Asn Pro Pro Lys Ala Ile Ser Asn Asp Ile Tyr Lys Lys Val
545 550 555 560

Asp Tyr Lys Leu Leu Pro Gly Pro Asn Lys Met Leu Pro Lys Val Phe
565 570 575

Phe Ala Thr Lys Asn Ile Ala Tyr Tyr Ala Pro Ser Glu Glu Leu Leu
580 585 590

Ser Lys Tyr Arg Lys Gly Thr His Lys Lys Gly Asp Ser Phe Ser Ile
595 600 605

Asp Asp Cys Arg Asn Leu Ile Asp Phe Phe Lys Ser Ser Ile Asn Lys
610 615 620

Asn Thr Asp Trp Ser Thr Phe Gly Phe Asn Phe Ser Asp Thr Asn Ser
625 630 635 640

Tyr Asn Asp Ile Ser Asp Phe Tyr Arg Glu Val Glu Lys Gln Gly Tyr
645 650 655

Lys Leu Ser Phe Thr Asp Ile Asp Ala Cys Tyr Ile Lys Asp Leu Val
660 665 670

Asp Asn Asn Glu Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser
675 680 685

Pro Tyr Ser Lys Gly Lys Leu Asn Leu His Thr Leu Tyr Phe Lys Met
690 695 700

Leu Phe Asp Gln Arg Asn Leu Asp Asn Val Val Tyr Lys Leu Asn Gly
705 710 715 720

Glu Ala Glu Val Phe Tyr Arg Pro Ala Ser Ile Glu Ser Asp Glu Gln
725 730 735

Ile Ile His Lys Ser Gly Gln Asn Ile Lys Asn Lys Asn Gln Lys Arg
740 745 750

Ser Asn Cys Lys Lys Thr Ser Thr Phe Asp Tyr Asp Ile Val Lys Asp
755 760 765

Arg Arg Tyr Cys Lys Asp Lys Phe Met Leu His Leu Pro Ile Thr Val

770 775 780

Asn Phe Gly Thr Asn Glu Ser Gly Lys Phe Asn Glu Leu Val Asn Asn
785 790 795 800

Ala Ile Arg Ala Asp Lys Asp Val Asn Val Ile Gly Ile Asp Arg Gly
805 810 815

Glu Arg Asn Leu Leu Tyr Val Val Val Val Asp Pro Cys Gly Lys Ile
820 825 830

Ile Glu Gln Ile Ser Leu Asn Thr Ile Val Asp Lys Glu Tyr Asp Ile
835 840 845

Glu Thr Asp Tyr His Gln Leu Leu Asp Glu Lys Glu Gly Ser Arg Asp
850 855 860

Lys Ala Arg Lys Asp Trp Asn Thr Ile Glu Asn Ile Lys Glu Leu Lys
865 870 875 880

Glu Gly Tyr Leu Ser Gln Val Val Asn Ile Ile Ala Lys Leu Val Leu
885 890 895

Lys Tyr Asp Ala Ile Ile Cys Leu Glu Asp Leu Asn Phe Gly Phe Lys
900 905 910

Arg Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu Lys
915 920 925

Met Leu Ile Asp Lys Met Asn Tyr Leu Val Leu Asp Lys Ser Arg Lys
930 935 940

Gln Glu Ser Pro Gln Lys Pro Gly Gly Ala Leu Asn Ala Leu Gln Leu
945 950 955 960

Thr Ser Ala Phe Lys Ser Phe Lys Glu Leu Gly Lys Gln Thr Gly Ile
965 970 975

Ile Tyr Tyr Val Pro Ala Tyr Leu Thr Ser Lys Ile Asp Pro Thr Thr
980 985 990

Gly Phe Ala Asn Leu Phe Tyr Ile Lys Tyr Glu Ser Val Asp Lys Ala
995 1000 1005

Arg Asp Phe Phe Ser Lys Phe Asp Phe Ile Arg Tyr Asn Gln Met
1010 1015 1020

Asp Asn Tyr Phe Glu Phe Gly Phe Asp Tyr Lys Ser Phe Thr Glu
1025 1030 1035

Arg Ala Ser Gly Cys Lys Ser Lys Trp Ile Ala Cys Thr Asn Gly
1040 1045 1050

Glu Arg Ile Val Lys Tyr Arg Asn Ser Asp Lys Asn Asn Ser Phe
1055 1060 1065

Asp Asp Lys Thr Val Ile Leu Thr Asp Glu Tyr Arg Ser Leu Phe
1070 1075 1080

Asp Lys Tyr Leu Gln Asn Tyr Ile Asp Glu Asp Asp Leu Lys Asp
1085 1090 1095

Gln Ile Leu Gln Ile Asp Ser Ala Asp Phe Tyr Lys Asn Leu Ile

1100 1105 1110

Lys Leu Phe Gln Leu Thr Leu Gln Met Arg Asn Ser Ser Ser Asp
1115 1120 1125

Gly Lys Arg Asp Tyr Ile Ile Ser Pro Val Lys Asn Tyr Arg Glu
1130 1135 1140

Glu Phe Phe Cys Ser Glu Phe Ser Asp Asp Thr Phe Pro Arg Asp
1145 1150 1155

Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala Arg Lys Gly Leu Trp
1160 1165 1170

Val Ile Lys Gln Ile Arg Glu Thr Lys Ser Gly Thr Lys Ile Asn
1175 1180 1185

Leu Ala Met Ser Asn Ser Glu Trp Leu Glu Tyr Ala Gln Cys Asn
1190 1195 1200

Leu Leu
1205

<210> 1147
<211> 1205
<212> PRT
<213> 瘤胃假丁酸弧菌

<400> 1147
Met Tyr Tyr Gln Asn Leu Thr Lys Met Tyr Pro Ile Ser Lys Thr Leu
1 5 10 15

Arg Asn Glu Leu Ile Pro Val Gly Lys Thr Leu Glu Asn Ile Arg Lys
20 25 30

Asn Gly Ile Leu Glu Ala Asp Ile Gln Arg Lys Ala Asp Tyr Glu His
35 40 45

Val Lys Lys Leu Met Asp Asn Tyr His Lys Gln Leu Ile Asn Glu Ala
50 55 60

Leu Gln Gly Val His Leu Ser Asp Leu Ser Asp Ala Tyr Asp Leu Tyr
65 70 75 80

Phe Asn Leu Ser Lys Glu Lys Asn Ser Val Asp Ala Phe Ser Lys Cys
85 90 95

Gln Asp Lys Leu Arg Lys Glu Ile Val Ser Leu Leu Lys Asn His Glu
100 105 110

Asn Phe Pro Lys Ile Gly Asn Lys Glu Ile Ile Lys Leu Leu Gln Ser
115 120 125

Leu Tyr Asp Asn Asp Thr Asp Tyr Lys Ala Leu Asp Ser Phe Ser Asn
130 135 140

Phe Tyr Thr Tyr Phe Ser Ser Tyr Asn Glu Val Arg Lys Asn Leu Tyr
145 150 155 160

Ser Asp Glu Glu Lys Ser Ser Thr Val Ala Tyr Arg Leu Ile Asn Glu
165 170 175

Asn Leu Pro Lys Phe Leu Asp Asn Ile Lys Ala Tyr Ala Ile Ala Lys
180 185 190

```
Lys Ala Gly Val Arg Ala Glu Gly Leu Ser Glu Glu Asp Gln Asp Cys
        195                 200                 205

Leu Phe Ile Ile Glu Thr Phe Glu Arg Thr Leu Thr Gln Asp Gly Ile
    210                 215                 220

Asp Asn Tyr Asn Ala Ala Ile Gly Lys Leu Asn Thr Ala Ile Asn Leu
225                 230                 235                 240

Phe Asn Gln Gln Asn Lys Lys Gln Glu Gly Phe Arg Lys Val Pro Gln
                245                 250                 255

Met Lys Cys Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Glu Ala Phe
            260                 265                 270

Ile Asp Glu Phe Ser Asp Asp Glu Asp Leu Ile Thr Asn Ile Glu Ser
        275                 280                 285

Phe Ala Glu Asn Met Asn Val Phe Leu Asn Ser Glu Ile Ile Thr Asp
    290                 295                 300

Phe Lys Ile Ala Leu Val Glu Ser Asp Gly Ser Leu Val Tyr Ile Lys
305                 310                 315                 320

Asn Asp Val Ser Lys Thr Ser Phe Ser Asn Ile Val Phe Gly Ser Trp
                325                 330                 335

Asn Ala Ile Asp Glu Lys Leu Ser Asp Glu Tyr Asp Leu Ala Asn Ser
            340                 345                 350

Lys Lys Lys Lys Asp Glu Lys Tyr Tyr Glu Lys Arg Gln Lys Glu Leu
        355                 360                 365

Lys Lys Asn Lys Ser Tyr Asp Leu Glu Thr Ile Ile Gly Leu Phe Asp
    370                 375                 380

Asp Asn Ser Asp Val Ile Gly Lys Tyr Ile Glu Lys Leu Glu Ser Asp
385                 390                 395                 400

Ile Thr Ala Ile Ala Glu Ala Lys Asn Asp Phe Asp Glu Ile Val Leu
                405                 410                 415

Arg Lys His Asp Lys Asn Lys Ser Leu Arg Lys Asn Thr Asn Ala Val
            420                 425                 430

Glu Ala Ile Lys Ser Tyr Leu Asp Thr Val Lys Asp Phe Glu Arg Asp
        435                 440                 445

Ile Lys Leu Ile Asn Gly Ser Gly Gln Glu Val Glu Lys Asn Leu Val
    450                 455                 460

Val Tyr Ala Glu Gln Glu Asn Ile Leu Ala Glu Ile Lys Asn Val Asp
465                 470                 475                 480

Ser Leu Tyr Asn Met Ser Arg Asn Tyr Leu Thr Gln Lys Pro Phe Ser
                485                 490                 495

Thr Glu Lys Phe Lys Leu Asn Phe Asn Arg Ala Thr Leu Leu Asn Gly
            500                 505                 510

Trp Asp Lys Asn Lys Glu Thr Asp Asn Leu Gly Ile Leu Phe Glu Lys
        515                 520                 525
```

```
Asp Gly Met Tyr Tyr Leu Gly Ile Met Asn Thr Lys Ala Asn Lys Ile
    530                 535                 540

Phe Val Asn Ile Pro Lys Ala Thr Ser Asn Asp Val Tyr His Lys Val
545                 550                 555                 560

Asn Tyr Lys Leu Leu Pro Gly Pro Asn Lys Met Leu Pro Lys Val Phe
                565                 570                 575

Phe Ala Gln Ser Asn Leu Asp Tyr Tyr Lys Pro Ser Glu Glu Leu Leu
            580                 585                 590

Ala Lys Tyr Lys Ala Gly Thr His Lys Lys Gly Asp Asn Phe Ser Leu
        595                 600                 605

Glu Asp Cys His Ala Leu Ile Asp Phe Phe Lys Ala Ser Ile Glu Lys
    610                 615                 620

His Pro Asp Trp Ser Ser Phe Gly Phe Glu Phe Ser Glu Thr Cys Thr
625                 630                 635                 640

Tyr Glu Asp Leu Ser Gly Phe Tyr Arg Glu Val Glu Lys Gln Gly Tyr
                645                 650                 655

Lys Ile Thr Tyr Thr Asp Val Asp Ala Asp Tyr Ile Thr Ser Leu Val
            660                 665                 670

Glu Arg Asp Glu Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser
        675                 680                 685

Pro Tyr Ser Lys Gly Asn Leu Asn Leu His Thr Ile Tyr Leu Gln Met
    690                 695                 700

Leu Phe Asp Gln Arg Asn Leu Asn Asn Val Val Tyr Lys Leu Asn Gly
705                 710                 715                 720

Glu Ala Glu Val Phe Tyr Arg Pro Ala Ser Ile Asn Asp Glu Glu Val
                725                 730                 735

Ile Ile His Lys Ala Gly Glu Glu Ile Lys Asn Lys Asn Ser Lys Arg
            740                 745                 750

Ala Val Asp Lys Pro Thr Ser Lys Phe Gly Tyr Asp Ile Ile Lys Asp
        755                 760                 765

Arg Arg Tyr Ser Lys Asp Lys Phe Met Leu His Ile Pro Val Thr Met
    770                 775                 780

Asn Phe Gly Val Asp Glu Thr Arg Arg Phe Asn Asp Val Val Asn Asp
785                 790                 795                 800

Ala Leu Arg Asn Asp Glu Lys Val Arg Val Ile Gly Ile Asp Arg Gly
                805                 810                 815

Glu Arg Asn Leu Leu Tyr Val Val Val Val Asp Thr Asp Gly Thr Ile
            820                 825                 830

Leu Glu Gln Ile Ser Leu Asn Ser Ile Ile Asn Asn Glu Tyr Ser Ile
        835                 840                 845

Glu Thr Asp Tyr His Lys Leu Leu Asp Glu Lys Glu Gly Asp Arg Asp
    850                 855                 860
```

```
Arg Ala Arg Lys Asn Trp Thr Thr Ile Glu Asn Ile Lys Glu Leu Lys
865                 870                 875                 880

Glu Gly Tyr Leu Ser Gln Val Val Asn Val Ile Ala Lys Leu Val Leu
                885                 890                 895

Lys Tyr Asn Ala Ile Ile Cys Leu Glu Asp Leu Asn Phe Gly Phe Lys
            900                 905                 910

Arg Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu Lys
        915                 920                 925

Met Leu Ile Asp Lys Leu Asn Tyr Leu Val Ile Asp Lys Ser Arg Lys
    930                 935                 940

Gln Asp Lys Pro Glu Glu Phe Gly Gly Ala Leu Asn Ala Leu Gln Leu
945                 950                 955                 960

Thr Ser Lys Phe Thr Ser Phe Lys Asp Met Gly Lys Gln Thr Gly Ile
                965                 970                 975

Ile Tyr Tyr Val Pro Ala Tyr Leu Thr Ser Lys Ile Asp Pro Thr Thr
            980                 985                 990

Gly Phe Ala Asn Leu Phe Tyr Val Lys Tyr Glu Asn Val Glu Lys Ala
        995                 1000                1005

Lys Glu  Phe Phe Ser Arg Phe  Asp Ser Ile Ser Tyr  Asn Asn Glu
    1010                1015                1020

Ser Gly  Tyr Phe Glu Phe Ala  Phe Asp Tyr Lys Lys  Phe Thr Asp
    1025                1030                1035

Arg Ala  Cys Gly Ala Arg Ser  Gln Trp Thr Val Cys  Thr Tyr Gly
    1040                1045                1050

Glu Arg  Ile Ile Lys Phe Arg  Asn Thr Glu Lys Asn  Asn Ser Phe
    1055                1060                1065

Asp Asp  Lys Thr Ile Val Leu  Ser Glu Glu Phe Lys  Glu Leu Phe
    1070                1075                1080

Ser Ile  Tyr Gly Ile Ser Tyr  Glu Asp Gly Ala Glu  Leu Lys Asn
    1085                1090                1095

Lys Ile  Met Ser Val Asp Glu  Ala Asp Phe Phe Arg  Ser Leu Thr
    1100                1105                1110

Arg Leu  Phe Gln Gln Thr Met  Gln Met Arg Asn Ser  Ser Asn Asp
    1115                1120                1125

Val Thr  Arg Asp Tyr Ile Ile  Ser Pro Ile Met Asn  Asp Arg Gly
    1130                1135                1140

Glu Phe  Phe Asn Ser Glu Ala  Cys Asp Ala Ser Lys  Pro Lys Asp
    1145                1150                1155

Ala Asp  Ala Asn Gly Ala Phe  Asn Ile Ala Arg Lys  Gly Leu Trp
    1160                1165                1170

Val Leu  Glu Gln Ile Arg Asn  Thr Pro Ser Gly Asp  Lys Leu Asn
    1175                1180                1185
```

Leu Ala Met Ser Asn Ala Glu Trp Leu Glu Tyr Ala Gln Arg Asn
1190 1195 1200

Gln Ile
1205

<210> 1148
<211> 1231
<212> PRT
<213> 溶纖維丁酸弧菌

<400> 1148
Met Tyr Tyr Glu Ser Leu Thr Lys Leu Tyr Pro Ile Lys Lys Thr Ile
1 5 10 15

Arg Asn Glu Leu Val Pro Ile Gly Lys Thr Leu Glu Asn Ile Lys Lys
20 25 30

Asn Asn Ile Leu Glu Ala Asp Glu Asp Arg Lys Ile Ala Tyr Ile Arg
35 40 45

Val Lys Ala Ile Met Asp Asp Tyr His Lys Arg Leu Ile Asn Glu Ala
50 55 60

Leu Ser Gly Phe Ala Leu Ile Asp Leu Asp Lys Ala Ala Asn Leu Tyr
65 70 75 80

Leu Ser Arg Ser Lys Ser Ala Asp Asp Ile Glu Ser Phe Ser Arg Phe
85 90 95

Gln Asp Lys Leu Arg Lys Ala Ile Ala Lys Arg Leu Arg Glu His Glu
100 105 110

Asn Phe Gly Lys Ile Gly Asn Lys Asp Ile Ile Pro Leu Leu Gln Lys
115 120 125

Leu Ser Glu Asn Glu Asp Asp Tyr Asn Ala Leu Glu Ser Phe Lys Asn
130 135 140

Phe Tyr Thr Tyr Phe Glu Ser Tyr Asn Asp Val Arg Leu Asn Leu Tyr
145 150 155 160

Ser Asp Lys Glu Lys Ser Ser Thr Val Ala Tyr Arg Leu Ile Asn Glu
165 170 175

Asn Leu Pro Arg Phe Leu Asp Asn Ile Arg Ala Tyr Asp Ala Val Gln
180 185 190

Lys Ala Gly Ile Thr Ser Glu Glu Leu Ser Ser Glu Ala Gln Asp Gly
195 200 205

Leu Phe Leu Val Asn Thr Phe Asn Asn Val Leu Ile Gln Asp Gly Ile
210 215 220

Asn Thr Tyr Asn Glu Asp Ile Gly Lys Leu Asn Val Ala Ile Asn Leu
225 230 235 240

Tyr Asn Gln Lys Asn Ala Ser Val Gln Gly Phe Arg Lys Val Pro Lys
245 250 255

Met Lys Val Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Glu Ser Phe
260 265 270

Ile Asp Glu Phe Glu Ser Asp Thr Glu Leu Leu Asp Ser Leu Glu Ser

275 280 285

His Tyr Ala Asn Leu Ala Lys Tyr Phe Gly Ser Asn Lys Val Gln Leu
290 295 300

Leu Phe Thr Ala Leu Arg Glu Ser Lys Gly Val Asn Val Tyr Val Lys
305 310 315 320

Asn Asp Ile Ala Lys Thr Ser Phe Ser Asn Val Val Phe Gly Ser Trp
325 330 335

Ser Arg Ile Asp Glu Leu Ile Asn Gly Glu Tyr Asp Asp Asn Asn Asn
340 345 350

Arg Lys Lys Asp Glu Lys Tyr Tyr Asp Lys Arg Gln Lys Glu Leu Lys
355 360 365

Lys Asn Lys Ser Tyr Thr Ile Glu Lys Ile Ile Thr Leu Ser Thr Glu
370 375 380

Asp Val Asp Val Ile Gly Lys Tyr Ile Glu Lys Leu Glu Ser Asp Ile
385 390 395 400

Asp Asp Ile Arg Phe Lys Gly Lys Asn Phe Tyr Glu Ala Val Leu Cys
405 410 415

Gly His Asp Arg Ser Lys Lys Leu Ser Lys Asn Lys Gly Ala Val Glu
420 425 430

Ala Ile Lys Gly Tyr Leu Asp Ser Val Lys Asp Phe Glu Arg Asp Leu
435 440 445

Lys Leu Ile Asn Gly Ser Gly Gln Glu Leu Glu Lys Asn Leu Val Val
450 455 460

Tyr Gly Glu Gln Glu Ala Val Leu Ser Glu Leu Ser Gly Ile Asp Ser
465 470 475 480

Leu Tyr Asn Met Thr Arg Asn Tyr Leu Thr Lys Lys Pro Phe Ser Thr
485 490 495

Glu Lys Ile Lys Leu Asn Phe Asn Lys Pro Thr Phe Leu Asp Gly Trp
500 505 510

Asp Tyr Gly Asn Glu Glu Ala Tyr Leu Gly Phe Phe Met Ile Lys Glu
515 520 525

Gly Asn Tyr Phe Leu Ala Val Met Asp Ala Asn Trp Asn Lys Glu Phe
530 535 540

Arg Asn Ile Pro Ser Val Asp Lys Ser Asp Cys Tyr Lys Lys Val Ile
545 550 555 560

Tyr Lys Gln Ile Ser Ser Pro Glu Lys Ser Ile Gln Asn Leu Met Val
565 570 575

Ile Asp Gly Lys Thr Val Lys Lys Asn Gly Arg Lys Glu Lys Glu Gly
580 585 590

Ile His Ser Gly Glu Asn Leu Ile Leu Glu Glu Leu Lys Asn Thr Tyr
595 600 605

Leu Pro Lys Lys Ile Asn Asp Ile Arg Lys Arg Arg Ser Tyr Leu Asn

610 615 620

Gly Asp Thr Phe Ser Lys Lys Asp Leu Thr Glu Phe Ile Gly Tyr Tyr
625 630 635 640

Lys Gln Arg Val Ile Glu Tyr Tyr Asn Gly Tyr Ser Phe Tyr Phe Lys
645 650 655

Ser Asp Asp Asp Tyr Ala Ser Phe Lys Glu Phe Gln Glu Asp Val Gly
660 665 670

Arg Gln Ala Tyr Gln Ile Ser Tyr Val Asp Val Pro Val Ser Phe Val
675 680 685

Asp Asp Leu Ile Asn Ser Gly Lys Leu Tyr Leu Phe Arg Val Tyr Asn
690 695 700

Lys Asp Phe Ser Glu Tyr Ser Lys Gly Arg Leu Asn Leu His Thr Leu
705 710 715 720

Tyr Phe Lys Met Leu Phe Asp Glu Arg Asn Leu Lys Asn Val Val Tyr
725 730 735

Lys Leu Asn Gly Gln Ala Glu Val Phe Tyr Arg Pro Ser Ser Ile Lys
740 745 750

Lys Glu Glu Leu Ile Val His Arg Ala Gly Glu Glu Ile Lys Asn Lys
755 760 765

Asn Pro Lys Arg Ala Ala Gln Lys Pro Thr Arg Arg Leu Asp Tyr Asp
770 775 780

Ile Val Lys Asp Arg Arg Tyr Ser Gln Asp Lys Phe Met Leu His Thr
785 790 795 800

Ser Ile Ile Met Asn Phe Gly Ala Glu Glu Asn Val Ser Phe Asn Asp
805 810 815

Ile Val Asn Gly Val Leu Arg Asn Glu Asp Lys Val Asn Val Ile Gly
820 825 830

Ile Asp Arg Gly Glu Arg Asn Leu Leu Tyr Val Val Val Ile Asp Pro
835 840 845

Glu Gly Lys Ile Leu Glu Gln Arg Ser Leu Asn Cys Ile Thr Asp Ser
850 855 860

Asn Leu Asp Ile Glu Thr Asp Tyr His Arg Leu Leu Asp Glu Lys Glu
865 870 875 880

Ser Asp Arg Lys Ile Ala Arg Arg Asp Trp Thr Thr Ile Glu Asn Ile
885 890 895

Lys Glu Leu Lys Ala Gly Tyr Leu Ser Gln Val Val His Ile Val Ala
900 905 910

Glu Leu Val Leu Lys Tyr Asn Ala Ile Ile Cys Leu Glu Asp Leu Asn
915 920 925

Phe Gly Phe Lys Arg Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln
930 935 940

Lys Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr Leu Val Met Asp

945 950 955 960

Lys Ser Arg Glu Gln Leu Ser Pro Glu Lys Ile Ser Gly Ala Leu Asn
965 970 975

Ala Leu Gln Leu Thr Pro Asp Phe Lys Ser Phe Lys Val Leu Gly Lys
980 985 990

Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Tyr Leu Thr Ser Lys Ile
995 1000 1005

Asp Pro Met Thr Gly Phe Ala Asn Leu Phe Tyr Val Lys Tyr Glu
1010 1015 1020

Asn Val Asp Lys Ala Lys Glu Phe Phe Ser Lys Phe Asp Ser Ile
1025 1030 1035

Lys Tyr Asn Lys Asp Gly Lys Asn Trp Asn Thr Lys Gly Tyr Phe
1040 1045 1050

Glu Phe Ala Phe Asp Tyr Lys Lys Phe Thr Asp Arg Ala Tyr Gly
1055 1060 1065

Arg Val Ser Glu Trp Thr Val Cys Thr Val Gly Glu Arg Ile Ile
1070 1075 1080

Lys Phe Lys Asn Lys Glu Lys Asn Asn Ser Tyr Asp Asp Lys Val
1085 1090 1095

Ile Asp Leu Thr Asn Ser Leu Lys Glu Leu Phe Asp Ser Tyr Lys
1100 1105 1110

Val Thr Tyr Glu Ser Glu Val Asp Leu Lys Asp Ala Ile Leu Ala
1115 1120 1125

Ile Asp Asp Pro Ala Phe Tyr Arg Asp Leu Thr Arg Arg Leu Gln
1130 1135 1140

Gln Thr Leu Gln Met Arg Asn Ser Ser Cys Asp Gly Ser Arg Asp
1145 1150 1155

Tyr Ile Ile Ser Pro Val Lys Asn Ser Lys Gly Glu Phe Phe Cys
1160 1165 1170

Ser Asp Asn Asn Asp Asp Thr Thr Pro Asn Asp Ala Asp Ala Asn
1175 1180 1185

Gly Ala Phe Asn Ile Ala Arg Lys Gly Leu Trp Val Leu Asn Glu
1190 1195 1200

Ile Arg Asn Ser Glu Glu Gly Ser Lys Ile Asn Leu Ala Met Ser
1205 1210 1215

Asn Ala Gln Trp Leu Glu Tyr Ala Gln Asp Asn Thr Ile
1220 1225 1230

<210> 1149
<211> 1206
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 1149

```
Met Glu Asn Tyr Tyr Asp Ser Leu Thr Arg Gln Tyr Pro Val Thr Lys
1               5                  10                  15

Thr Ile Arg Gln Glu Leu Lys Pro Val Gly Lys Thr Leu Glu Asn Ile
            20                  25                  30

Lys Asn Ala Glu Ile Ile Glu Ala Asp Lys Gln Lys Lys Glu Ala Tyr
        35                  40                  45

Val Lys Val Lys Glu Leu Met Asp Glu Phe His Lys Ser Ile Ile Glu
    50                  55                  60

Lys Ser Leu Val Gly Ile Lys Leu Asp Gly Leu Ser Glu Phe Glu Lys
65                  70                  75                  80

Leu Tyr Lys Ile Lys Thr Lys Thr Asp Glu Asp Lys Asn Arg Ile Ser
                85                  90                  95

Glu Leu Phe Tyr Tyr Met Arg Lys Gln Ile Ala Asp Ala Leu Lys Asn
            100                 105                 110

Ser Arg Asp Tyr Gly Tyr Val Asp Asn Lys Asp Leu Ile Glu Lys Ile
        115                 120                 125

Leu Pro Glu Arg Val Lys Asp Glu Asn Ser Leu Asn Ala Leu Ser Cys
    130                 135                 140

Phe Lys Gly Phe Thr Thr Tyr Phe Thr Asp Tyr Tyr Lys Asn Arg Lys
145                 150                 155                 160

Asn Ile Tyr Ser Asp Glu Glu Lys His Ser Thr Val Gly Tyr Arg Cys
                165                 170                 175

Ile Asn Glu Asn Leu Leu Ile Phe Met Ser Asn Ile Glu Val Tyr Gln
            180                 185                 190

Ile Tyr Lys Lys Ala Asn Ile Lys Asn Asp Asn Tyr Asp Glu Glu Thr
        195                 200                 205

Leu Asp Lys Thr Phe Met Ile Glu Ser Phe Asn Glu Cys Leu Thr Gln
    210                 215                 220

Ser Gly Val Glu Ala Tyr Asn Ser Val Val Ala Ser Ile Lys Thr Ala
225                 230                 235                 240

Thr Asn Leu Tyr Ile Gln Lys Asn Asn Lys Glu Glu Asn Phe Val Arg
                245                 250                 255

Val Pro Lys Met Lys Val Leu Phe Lys Gln Ile Leu Ser Asp Arg Thr
            260                 265                 270

Ser Leu Phe Asp Gly Leu Ile Ile Glu Ser Asp Asp Glu Leu Leu Asp
        275                 280                 285

Lys Leu Cys Ser Phe Ser Ala Glu Val Asp Lys Phe Leu Pro Ile Asn
    290                 295                 300

Ile Asp Arg Tyr Ile Lys Thr Leu Met Asp Ser Asn Asn Gly Thr Gly
305                 310                 315                 320

Ile Tyr Val Lys Asn Asp Ser Ser Leu Thr Thr Leu Ser Asn Tyr Leu
                325                 330                 335
```

```
Thr Asp Ser Trp Ser Ser Ile Arg Asn Ala Phe Asn Glu Asn Tyr Asp
            340                 345                 350

Ala Lys Tyr Thr Gly Lys Val Asn Asp Lys Tyr Glu Glu Lys Arg Glu
        355                 360                 365

Lys Ala Tyr Lys Ser Asn Asp Ser Phe Glu Leu Asn Tyr Ile Gln Asn
    370                 375                 380

Leu Leu Gly Ile Asn Val Ile Asp Lys Tyr Ile Glu Arg Ile Asn Phe
385                 390                 395                 400

Asp Ile Lys Glu Ile Cys Glu Ala Tyr Lys Glu Met Thr Lys Asn Cys
                405                 410                 415

Phe Glu Asp His Asp Lys Thr Lys Lys Leu Gln Lys Asn Ile Lys Ala
            420                 425                 430

Val Ala Ser Ile Lys Ser Tyr Leu Asp Ser Leu Lys Asn Ile Glu Arg
        435                 440                 445

Asp Ile Lys Leu Leu Asn Gly Thr Gly Leu Glu Ser Arg Asn Glu Phe
    450                 455                 460

Phe Tyr Gly Glu Gln Ser Thr Val Leu Glu Glu Ile Thr Lys Val Asp
465                 470                 475                 480

Glu Leu Tyr Asn Ile Thr Arg Asn Tyr Leu Thr Lys Lys Pro Phe Ser
                485                 490                 495

Thr Glu Lys Met Lys Leu Asn Phe Asn Asn Pro Gln Leu Leu Gly Gly
            500                 505                 510

Trp Asp Val Asn Lys Glu Arg Asp Cys Tyr Gly Val Ile Leu Ile Lys
        515                 520                 525

Asp Asn Asn Tyr Tyr Leu Gly Ile Met Asp Lys Ser Ala Asn Lys Ser
    530                 535                 540

Phe Leu Asn Ile Lys Glu Ser Lys Asn Glu Asn Ala Tyr Lys Lys Val
545                 550                 555                 560

Asn Cys Lys Leu Leu Pro Gly Pro Asn Lys Met Phe Pro Lys Val Phe
                565                 570                 575

Phe Ala Lys Ser Asn Ile Asp Tyr Tyr Asp Pro Thr His Glu Ile Lys
            580                 585                 590

Lys Leu Tyr Asp Lys Gly Thr Phe Lys Lys Gly Asn Ser Phe Asn Leu
        595                 600                 605

Glu Asp Cys His Lys Leu Ile Asp Phe Tyr Lys Glu Ser Ile Lys Lys
    610                 615                 620

Asn Asp Asp Trp Lys Asn Phe Asn Phe Asn Phe Ser Asp Thr Lys Asp
625                 630                 635                 640

Tyr Glu Asp Ile Ser Gly Phe Phe Arg Glu Val Glu Ala Gln Asn Tyr
                645                 650                 655

Lys Ile Thr Tyr Thr Asn Val Ser Cys Asp Phe Ile Glu Ser Leu Val
            660                 665                 670
```

```
Asp Glu Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser
        675                 680                 685

Glu Tyr Ala Thr Gly Asn Leu Asn Leu His Thr Leu Tyr Leu Lys Met
    690                 695                 700

Leu Phe Asp Glu Arg Asn Leu Lys Asp Leu Cys Ile Lys Met Asn Gly
705                 710                 715                 720

Glu Ala Glu Val Phe Tyr Arg Pro Ala Ser Ile Leu Asp Glu Asp Lys
                725                 730                 735

Val Val His Lys Ala Asn Gln Lys Ile Thr Asn Lys Asn Thr Asn Ser
            740                 745                 750

Lys Lys Lys Glu Ser Ile Phe Ser Tyr Asp Ile Val Lys Asp Lys Arg
        755                 760                 765

Tyr Thr Val Asp Lys Phe Phe Ile His Leu Pro Ile Thr Leu Asn Tyr
    770                 775                 780

Lys Glu Gln Asn Val Ser Arg Phe Asn Asp Tyr Ile Arg Glu Ile Leu
785                 790                 795                 800

Lys Lys Ser Lys Asn Ile Arg Val Ile Gly Ile Asp Arg Gly Glu Arg
                805                 810                 815

Asn Leu Leu Tyr Val Val Val Cys Asp Ser Asp Gly Ser Ile Leu Tyr
            820                 825                 830

Gln Arg Ser Ile Asn Glu Ile Val Ser Gly Ser His Lys Thr Asp Tyr
        835                 840                 845

His Lys Leu Leu Asp Asn Lys Glu Lys Glu Arg Leu Ser Ser Arg Arg
    850                 855                 860

Asp Trp Lys Thr Ile Glu Asn Ile Lys Asp Leu Lys Ala Gly Tyr Met
865                 870                 875                 880

Ser Gln Val Val Asn Glu Ile Tyr Asn Leu Ile Leu Lys Tyr Asn Ala
                885                 890                 895

Ile Val Val Leu Glu Asp Leu Asn Ile Gly Phe Lys Asn Gly Arg Lys
            900                 905                 910

Lys Val Glu Lys Gln Val Tyr Gln Asn Phe Glu Lys Ala Leu Ile Asp
        915                 920                 925

Lys Leu Asn Tyr Leu Cys Ile Asp Lys Thr Arg Glu Gln Leu Ser Pro
    930                 935                 940

Ser Ser Pro Gly Gly Val Leu Asn Ala Tyr Gln Leu Thr Ala Lys Phe
945                 950                 955                 960

Glu Ser Phe Glu Lys Ile Gly Lys Gln Thr Gly Cys Ile Phe Tyr Val
                965                 970                 975

Pro Ala Tyr Leu Thr Ser Gln Ile Asp Pro Thr Thr Gly Phe Val Asn
            980                 985                 990

Leu Phe Tyr Gln Lys Asp Thr Ser Lys Gln Gly Leu Gln Leu Phe Phe
        995                 1000                1005
```

```
Arg Lys  Phe Lys Lys Ile Asn  Phe Asp Lys Val Ala  Ser Asn Phe
    1010                 1015                 1020

Glu Phe  Val Phe Asp Tyr Asn  Asp Phe Thr Asn Lys  Ala Glu Gly
    1025                 1030                 1035

Thr Lys  Thr Asn Trp Thr Ile  Ser Thr Gln Gly Thr  Arg Ile Ala
    1040                 1045                 1050

Lys Tyr  Arg Ser Asp Asp Ala  Asn Gly Lys Trp Ile  Ser Arg Thr
    1055                 1060                 1065

Val His  Pro Thr Asp Ile Ile  Lys Glu Ala Leu Asn  Arg Glu Lys
    1070                 1075                 1080

Ile Asn  Tyr Asn Asp Gly His  Asp Leu Ile Asp Glu  Ile Val Ser
    1085                 1090                 1095

Ile Glu  Lys Ser Ala Val Leu  Lys Glu Ile Tyr Tyr  Gly Phe Lys
    1100                 1105                 1110

Leu Thr  Leu Gln Leu Arg Asn  Ser Thr Leu Ala Asn  Glu Glu Glu
    1115                 1120                 1125

Gln Glu  Asp Tyr Ile Ile Ser  Pro Val Lys Asn Ser  Ser Gly Asn
    1130                 1135                 1140

Tyr Phe  Asp Ser Arg Ile Thr  Ser Lys Glu Leu Pro  Cys Asp Ala
    1145                 1150                 1155

Asp Ala  Asn Gly Ala Tyr Asn  Ile Ala Arg Lys Gly  Leu Trp Ala
    1160                 1165                 1170

Leu Glu  Gln Ile Arg Asn Ser  Glu Asn Val Ser Lys  Val Lys Leu
    1175                 1180                 1185

Ala Ile  Ser Asn Lys Glu Trp  Phe Glu Tyr Thr Gln  Asn Asn Ile
    1190                 1195                 1200

Pro Ser  Leu
    1205
```

<210> 1150
<211> 1228
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 1150

```
Met Ser Lys Leu Glu Lys Phe Thr Asn Cys Tyr Ser Leu Ser Lys Thr
1               5                  10                  15

Leu Arg Phe Lys Ala Ile Pro Val Gly Lys Thr Gln Glu Asn Ile Asp
            20                  25                  30

Asn Lys Arg Leu Leu Val Glu Asp Glu Lys Arg Ala Glu Asp Tyr Lys
        35                  40                  45

Gly Val Lys Lys Leu Leu Asp Arg Tyr Tyr Leu Ser Phe Ile Asn Asp
    50                  55                  60

Val Leu His Ser Ile Lys Leu Lys Asn Leu Asn Asn Tyr Ile Ser Leu
```

65 70 75 80

Phe Arg Lys Lys Thr Arg Thr Glu Lys Glu Asn Lys Glu Leu Glu Asn
85 90 95

Leu Glu Ile Asn Leu Arg Lys Glu Ile Ala Lys Ala Phe Lys Gly Asn
100 105 110

Glu Gly Tyr Lys Ser Leu Phe Lys Lys Asp Ile Ile Glu Thr Ile Leu
115 120 125

Pro Glu Phe Leu Asp Asp Lys Asp Glu Ile Ala Leu Val Asn Ser Phe
130 135 140

Asn Gly Phe Thr Thr Ala Phe Thr Gly Phe Phe Asp Asn Arg Glu Asn
145 150 155 160

Met Phe Ser Glu Glu Ala Lys Ser Thr Ser Ile Ala Phe Arg Cys Ile
165 170 175

Asn Glu Asn Leu Thr Arg Tyr Ile Ser Asn Met Asp Ile Phe Glu Lys
180 185 190

Val Asp Ala Ile Phe Asp Lys His Glu Val Gln Glu Ile Lys Glu Lys
195 200 205

Ile Leu Asn Ser Asp Tyr Asp Val Glu Asp Phe Phe Glu Gly Glu Phe
210 215 220

Phe Asn Phe Val Leu Thr Gln Glu Gly Ile Asp Val Tyr Asn Ala Ile
225 230 235 240

Ile Gly Gly Phe Val Thr Glu Ser Gly Glu Lys Ile Lys Gly Leu Asn
245 250 255

Glu Tyr Ile Asn Leu Tyr Asn Gln Lys Thr Lys Gln Lys Leu Pro Lys
260 265 270

Phe Lys Pro Leu Tyr Lys Gln Val Leu Ser Asp Arg Glu Ser Leu Ser
275 280 285

Phe Tyr Gly Glu Gly Tyr Thr Ser Asp Glu Glu Val Leu Glu Val Phe
290 295 300

Arg Asn Thr Leu Asn Lys Asn Ser Glu Ile Phe Ser Ser Ile Lys Lys
305 310 315 320

Leu Glu Lys Leu Phe Lys Asn Phe Asp Glu Tyr Ser Ser Ala Gly Ile
325 330 335

Phe Val Lys Asn Gly Pro Ala Ile Ser Thr Ile Ser Lys Asp Ile Phe
340 345 350

Gly Glu Trp Asn Val Ile Arg Asp Lys Trp Asn Ala Glu Tyr Asp Asp
355 360 365

Ile His Leu Lys Lys Lys Ala Val Val Thr Glu Lys Tyr Glu Asp Asp
370 375 380

Arg Arg Lys Ser Phe Lys Lys Ile Gly Ser Phe Ser Leu Glu Gln Leu
385 390 395 400

Gln Glu Tyr Ala Asp Ala Asp Leu Ser Val Val Glu Lys Leu Lys Glu

405 410 415

Ile Ile Ile Gln Lys Val Asp Glu Ile Tyr Lys Val Tyr Gly Ser Ser
420 425 430

Glu Lys Leu Phe Asp Ala Asp Phe Val Leu Glu Lys Ser Leu Lys Lys
435 440 445

Asn Asp Ala Val Val Ala Ile Met Lys Asp Leu Leu Asp Ser Val Lys
450 455 460

Ser Phe Glu Asn Tyr Ile Lys Ala Phe Phe Gly Glu Gly Lys Glu Thr
465 470 475 480

Asn Arg Asp Glu Ser Phe Tyr Gly Asp Phe Val Leu Ala Tyr Asp Ile
485 490 495

Leu Leu Lys Val Asp His Ile Tyr Asp Ala Ile Arg Asn Tyr Val Thr
500 505 510

Gln Lys Pro Tyr Ser Lys Asp Lys Phe Lys Leu Tyr Phe Gln Asn Pro
515 520 525

Gln Phe Met Gly Gly Trp Asp Lys Asp Lys Glu Thr Asp Tyr Arg Ala
530 535 540

Thr Ile Leu Arg Tyr Gly Ser Lys Tyr Tyr Leu Ala Ile Met Asp Lys
545 550 555 560

Lys Tyr Ala Lys Cys Leu Gln Lys Ile Asp Lys Asp Asp Val Asn Gly
565 570 575

Asn Tyr Glu Lys Ile Asn Tyr Lys Leu Leu Pro Gly Pro Asn Lys Met
580 585 590

Leu Pro Lys Val Phe Phe Ser Lys Lys Trp Met Ala Tyr Tyr Asn Pro
595 600 605

Ser Glu Asp Ile Gln Lys Ile Tyr Lys Asn Gly Thr Phe Lys Lys Gly
610 615 620

Asp Met Phe Asn Leu Asn Asp Cys His Lys Leu Ile Asp Phe Phe Lys
625 630 635 640

Asp Ser Ile Ser Arg Tyr Pro Lys Trp Ser Asn Ala Tyr Asp Phe Asn
645 650 655

Phe Ser Glu Thr Glu Lys Tyr Lys Asp Ile Ala Gly Phe Tyr Arg Glu
660 665 670

Val Glu Glu Gln Gly Tyr Lys Val Ser Phe Glu Ser Ala Ser Lys Lys
675 680 685

Glu Val Asp Lys Leu Val Glu Glu Gly Lys Leu Tyr Met Phe Gln Ile
690 695 700

Tyr Asn Lys Asp Phe Ser Asp Lys Ser His Gly Thr Pro Asn Leu His
705 710 715 720

Thr Met Tyr Phe Lys Leu Leu Phe Asp Glu Asn Asn His Gly Gln Ile
725 730 735

Arg Leu Ser Gly Gly Ala Glu Leu Phe Met Arg Arg Ala Ser Leu Lys

740 745 750

Lys Glu Glu Leu Val Val His Pro Ala Asn Ser Pro Ile Ala Asn Lys
755 760 765

Asn Pro Asp Asn Pro Lys Lys Thr Thr Thr Leu Ser Tyr Asp Val Tyr
770 775 780

Lys Asp Lys Arg Phe Ser Glu Asp Gln Tyr Glu Leu His Ile Pro Ile
785 790 795 800

Ala Ile Asn Lys Cys Pro Lys Asn Ile Phe Lys Ile Asn Thr Glu Val
805 810 815

Arg Val Leu Leu Lys His Asp Asp Asn Pro Tyr Val Ile Gly Ile Asp
820 825 830

Arg Gly Glu Arg Asn Leu Leu Tyr Ile Val Val Val Asp Gly Lys Gly
835 840 845

Asn Ile Val Glu Gln Tyr Ser Leu Asn Glu Ile Ile Asn Asn Phe Asn
850 855 860

Gly Ile Arg Ile Lys Thr Asp Tyr His Ser Leu Leu Asp Lys Lys Glu
865 870 875 880

Lys Glu Arg Phe Glu Ala Arg Gln Asn Trp Thr Ser Ile Glu Asn Ile
885 890 895

Lys Glu Leu Lys Ala Gly Tyr Ile Ser Gln Val Val His Lys Ile Cys
900 905 910

Glu Leu Val Glu Lys Tyr Asp Ala Val Ile Ala Leu Glu Asp Leu Asn
915 920 925

Ser Gly Phe Lys Asn Ser Arg Val Lys Val Glu Lys Gln Val Tyr Gln
930 935 940

Lys Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr Met Val Asp Lys
945 950 955 960

Lys Ser Asn Pro Cys Ala Thr Gly Gly Ala Leu Lys Gly Tyr Gln Ile
965 970 975

Thr Asn Lys Phe Glu Ser Phe Lys Ser Met Ser Thr Gln Asn Gly Phe
980 985 990

Ile Phe Tyr Ile Pro Ala Trp Leu Thr Ser Lys Ile Asp Pro Ser Thr
995 1000 1005

Gly Phe Val Asn Leu Leu Lys Thr Lys Tyr Thr Ser Ile Ala Asp
1010 1015 1020

Ser Lys Lys Phe Ile Ser Ser Phe Asp Arg Ile Met Tyr Val Pro
1025 1030 1035

Glu Glu Asp Leu Phe Glu Phe Ala Leu Asp Tyr Lys Asn Phe Ser
1040 1045 1050

Arg Thr Asp Ala Asp Tyr Ile Lys Lys Trp Lys Leu Tyr Ser Tyr
1055 1060 1065

Gly Asn Arg Ile Arg Ile Phe Arg Asn Pro Lys Lys Asn Asn Val

1070 1075 1080

Phe Asp Trp Glu Glu Val Cys Leu Thr Ser Ala Tyr Lys Glu Leu
1085 1090 1095

Phe Asn Lys Tyr Gly Ile Asn Tyr Gln Gln Gly Asp Ile Arg Ala
1100 1105 1110

Leu Leu Cys Glu Gln Ser Asp Lys Ala Phe Tyr Ser Ser Phe Met
1115 1120 1125

Ala Leu Met Ser Leu Met Leu Gln Met Arg Asn Ser Ile Thr Gly
1130 1135 1140

Arg Thr Asp Val Asp Phe Leu Ile Ser Pro Val Lys Asn Ser Asp
1145 1150 1155

Gly Ile Phe Tyr Asp Ser Arg Asn Tyr Glu Ala Gln Glu Asn Ala
1160 1165 1170

Ile Leu Pro Lys Asn Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala
1175 1180 1185

Arg Lys Val Leu Trp Ala Ile Gly Gln Phe Lys Lys Ala Glu Asp
1190 1195 1200

Glu Lys Leu Asp Lys Val Lys Ile Ala Ile Ser Asn Lys Glu Trp
1205 1210 1215

Leu Glu Tyr Ala Gln Thr Ser Val Lys His
1220 1225

<210> 1151
<211> 1373
<212> PRT
<213> 牛莫拉氏菌

<400> 1151
Met Leu Phe Gln Asp Phe Thr His Leu Tyr Pro Leu Ser Lys Thr Val
1 5 10 15

Arg Phe Glu Leu Lys Pro Ile Asp Arg Thr Leu Glu His Ile His Ala
20 25 30

Lys Asn Phe Leu Ser Gln Asp Glu Thr Met Ala Asp Met His Gln Lys
35 40 45

Val Lys Val Ile Leu Asp Asp Tyr His Arg Asp Phe Ile Ala Asp Met
50 55 60

Met Gly Glu Val Lys Leu Thr Lys Leu Ala Glu Phe Tyr Asp Val Tyr
65 70 75 80

Leu Lys Phe Arg Lys Asn Pro Lys Asp Asp Glu Leu Gln Lys Gln Leu
85 90 95

Lys Asp Leu Gln Ala Val Leu Arg Lys Glu Ile Val Lys Pro Ile Gly
100 105 110

Asn Gly Gly Lys Tyr Lys Ala Gly Tyr Asp Arg Leu Phe Gly Ala Lys
115 120 125

Leu Phe Lys Asp Gly Lys Glu Leu Gly Asp Leu Ala Lys Phe Val Ile
130 135 140

```
Ala Gln Glu Gly Glu Ser Ser Pro Lys Leu Ala His Leu Ala His Phe
145                 150                 155                 160

Glu Lys Phe Ser Thr Tyr Phe Thr Gly Phe His Asp Asn Arg Lys Asn
                165                 170                 175

Met Tyr Ser Asp Glu Asp Lys His Thr Ala Ile Ala Tyr Arg Leu Ile
            180                 185                 190

His Glu Asn Leu Pro Arg Phe Ile Asp Asn Leu Gln Ile Leu Thr Thr
        195                 200                 205

Ile Lys Gln Lys His Ser Ala Leu Tyr Asp Gln Ile Ile Asn Glu Leu
    210                 215                 220

Thr Ala Ser Gly Leu Asp Val Ser Leu Ala Ser His Leu Asp Gly Tyr
225                 230                 235                 240

His Lys Leu Leu Thr Gln Glu Gly Ile Thr Ala Tyr Asn Thr Leu Leu
                245                 250                 255

Gly Gly Ile Ser Gly Glu Ala Gly Ser Pro Lys Ile Gln Gly Ile Asn
            260                 265                 270

Glu Leu Ile Asn Ser His His Asn Gln His Cys His Lys Ser Glu Arg
        275                 280                 285

Ile Ala Lys Leu Arg Pro Leu His Lys Gln Ile Leu Ser Asp Gly Met
    290                 295                 300

Ser Val Ser Phe Leu Pro Ser Lys Phe Ala Asp Asp Ser Glu Met Cys
305                 310                 315                 320

Gln Ala Val Asn Glu Phe Tyr Arg His Tyr Ala Asp Val Phe Ala Lys
                325                 330                 335

Val Gln Ser Leu Phe Asp Gly Phe Asp Asp His Gln Lys Asp Gly Ile
            340                 345                 350

Tyr Val Glu His Lys Asn Leu Asn Glu Leu Ser Lys Gln Ala Phe Gly
        355                 360                 365

Asp Phe Ala Leu Leu Gly Arg Val Leu Asp Gly Tyr Tyr Val Asp Val
    370                 375                 380

Val Asn Pro Glu Phe Asn Glu Arg Phe Ala Lys Ala Lys Thr Asp Asn
385                 390                 395                 400

Ala Lys Ala Lys Leu Thr Lys Glu Lys Asp Lys Phe Ile Lys Gly Val
                405                 410                 415

His Ser Leu Ala Ser Leu Glu Gln Ala Ile Glu His Tyr Thr Ala Arg
            420                 425                 430

His Asp Asp Glu Ser Val Gln Ala Gly Lys Leu Gly Gln Tyr Phe Lys
        435                 440                 445

His Gly Leu Ala Gly Val Asp Asn Pro Ile Gln Lys Ile His Asn Asn
    450                 455                 460

His Ser Thr Ile Lys Gly Phe Leu Glu Arg Glu Arg Pro Ala Gly Glu
465                 470                 475                 480
```

```
Arg Ala Leu Pro Lys Ile Lys Ser Gly Lys Asn Pro Glu Met Thr Gln
                485                 490                 495

Leu Arg Gln Leu Lys Glu Leu Leu Asp Asn Ala Leu Asn Val Ala His
            500                 505                 510

Phe Ala Lys Leu Leu Thr Thr Lys Thr Thr Leu Asp Asn Gln Asp Gly
        515                 520                 525

Asn Phe Tyr Gly Glu Phe Gly Val Leu Tyr Asp Glu Leu Ala Lys Ile
    530                 535                 540

Pro Thr Leu Tyr Asn Lys Val Arg Asp Tyr Leu Ser Gln Lys Pro Phe
545                 550                 555                 560

Ser Thr Glu Lys Tyr Lys Leu Asn Phe Gly Asn Pro Thr Leu Leu Asn
                565                 570                 575

Gly Trp Asp Leu Asn Lys Glu Lys Asp Asn Phe Gly Val Ile Leu Gln
            580                 585                 590

Lys Asp Gly Cys Tyr Tyr Leu Ala Leu Leu Asp Lys Ala His Lys Lys
        595                 600                 605

Val Phe Asp Asn Ala Pro Asn Thr Gly Lys Ser Ile Tyr Gln Lys Met
    610                 615                 620

Ile Tyr Lys Tyr Leu Glu Val Arg Lys Gln Phe Pro Lys Val Phe Phe
625                 630                 635                 640

Ser Lys Glu Ala Ile Ala Ile Asn Tyr His Pro Ser Lys Glu Leu Val
                645                 650                 655

Glu Ile Lys Asp Lys Gly Arg Gln Arg Ser Asp Asp Glu Arg Leu Lys
            660                 665                 670

Leu Tyr Arg Phe Ile Leu Glu Cys Leu Lys Ile His Pro Lys Tyr Asp
        675                 680                 685

Lys Lys Phe Glu Gly Ala Ile Gly Asp Ile Gln Leu Phe Lys Lys Asp
    690                 695                 700

Lys Lys Gly Arg Glu Val Pro Ile Ser Glu Lys Asp Leu Phe Asp Lys
705                 710                 715                 720

Ile Asn Gly Ile Phe Ser Ser Lys Pro Lys Leu Glu Met Glu Asp Phe
                725                 730                 735

Phe Ile Gly Glu Phe Lys Arg Tyr Asn Pro Ser Gln Asp Leu Val Asp
            740                 745                 750

Gln Tyr Asn Ile Tyr Lys Lys Ile Asp Ser Asn Asp Asn Arg Lys Lys
        755                 760                 765

Glu Asn Phe Tyr Asn Asn His Pro Lys Phe Lys Lys Asp Leu Val Arg
    770                 775                 780

Tyr Tyr Tyr Glu Ser Met Cys Lys His Glu Glu Trp Glu Glu Ser Phe
785                 790                 795                 800

Glu Phe Ser Lys Lys Leu Gln Asp Ile Gly Cys Tyr Val Asp Val Asn
                805                 810                 815
```

```
Glu Leu Phe Thr Glu Ile Glu Thr Arg Arg Leu Asn Tyr Lys Ile Ser
            820                 825                 830

Phe Cys Asn Ile Asn Ala Asp Tyr Ile Asp Glu Leu Val Glu Gln Gly
        835                 840                 845

Gln Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Pro Lys Ala
    850                 855                 860

His Gly Lys Pro Asn Leu His Thr Leu Tyr Phe Lys Ala Leu Phe Ser
865                 870                 875                 880

Glu Asp Asn Leu Ala Asp Pro Ile Tyr Lys Leu Asn Gly Glu Ala Gln
                885                 890                 895

Ile Phe Tyr Arg Lys Ala Ser Leu Asp Met Asn Glu Thr Thr Ile His
            900                 905                 910

Arg Ala Gly Glu Val Leu Glu Asn Lys Asn Pro Asp Asn Pro Lys Lys
        915                 920                 925

Arg Gln Phe Val Tyr Asp Ile Ile Lys Asp Lys Arg Tyr Thr Gln Asp
    930                 935                 940

Lys Phe Met Leu His Val Pro Ile Thr Met Asn Phe Gly Val Gln Gly
945                 950                 955                 960

Met Thr Ile Lys Glu Phe Asn Lys Lys Val Asn Gln Ser Ile Gln Gln
                965                 970                 975

Tyr Asp Glu Val Asn Val Ile Gly Ile Asp Arg Gly Glu Arg His Leu
            980                 985                 990

Leu Tyr Leu Thr Val Ile Asn Ser Lys Gly Glu Ile Leu Glu Gln Cys
        995                 1000                1005

Ser Leu  Asn Asp Ile Thr Thr Ala Ser Ala Asn  Gly Thr Gln Met
    1010                1015                1020

Thr Thr  Pro Tyr His Lys Ile Leu Asp Lys Arg  Glu Ile Glu Arg
    1025                1030                1035

Leu Asn  Ala Arg Val Gly Trp Gly Glu Ile Glu  Thr Ile Lys Glu
    1040                1045                1050

Leu Lys  Ser Gly Tyr Leu Ser His Val Val His  Gln Ile Ser Gln
    1055                1060                1065

Leu Met  Leu Lys Tyr Asn Ala Ile Val Val Leu  Glu Asp Leu Asn
    1070                1075                1080

Phe Gly  Phe Lys Arg Gly Arg Phe Lys Val Glu  Lys Gln Ile Tyr
    1085                1090                1095

Gln Asn  Phe Glu Asn Ala Leu Ile Lys Lys Leu  Asn His Leu Val
    1100                1105                1110

Leu Lys  Asp Lys Ala Asp Asp Glu Ile Gly Ser  Tyr Lys Asn Ala
    1115                1120                1125

Leu Gln  Leu Thr Asn Asn Phe Thr Asp Leu Lys  Ser Ile Gly Lys
    1130                1135                1140
```

```
Gln Thr  Gly Phe Leu Phe Tyr  Val Pro Ala Trp Asn  Thr Ser Lys
    1145                 1150                 1155

Ile Asp  Pro Glu Thr Gly Phe  Val Asp Leu Leu Lys  Pro Arg Tyr
    1160                 1165                 1170

Glu Asn  Ile Ala Gln Ser Gln  Ala Phe Phe Gly Lys  Phe Asp Lys
    1175                 1180                 1185

Ile Cys  Tyr Asn Ala Asp Lys  Asp Tyr Phe Glu Phe  His Ile Asp
    1190                 1195                 1200

Tyr Ala  Lys Phe Thr Asp Lys  Ala Lys Asn Ser Arg  Gln Ile Trp
    1205                 1210                 1215

Thr Ile  Cys Ser His Gly Asp  Lys Arg Tyr Val Tyr  Asp Lys Thr
    1220                 1225                 1230

Ala Asn  Gln Asn Lys Gly Ala  Ala Lys Gly Ile Asn  Val Asn Asp
    1235                 1240                 1245

Glu Leu  Lys Ser Leu Phe Ala  Arg His His Ile Asn  Glu Lys Gln
    1250                 1255                 1260

Pro Asn  Leu Val Met Asp Ile  Cys Gln Asn Asn Asp  Lys Glu Phe
    1265                 1270                 1275

His Lys  Ser Leu Met Tyr Leu  Leu Lys Thr Leu Leu  Ala Leu Arg
    1280                 1285                 1290

Tyr Ser  Asn Ala Ser Ser Asp  Glu Asp Phe Ile Leu  Ser Pro Val
    1295                 1300                 1305

Ala Asn  Asp Glu Gly Val Phe  Phe Asn Ser Ala Leu  Ala Asp Asp
    1310                 1315                 1320

Thr Gln  Pro Gln Asn Ala Asp  Ala Asn Gly Ala Tyr  His Ile Ala
    1325                 1330                 1335

Leu Lys  Gly Leu Trp Leu Leu  Asn Glu Leu Lys Asn  Ser Asp Asp
    1340                 1345                 1350

Leu Asn  Lys Val Lys Leu Ala  Ile Asp Asn Gln Thr  Trp Leu Asn
    1355                 1360                 1365

Phe Ala  Gln Asn Arg
    1370
```

<210> 1152
<211> 1373
<212> PRT
<213> 牛莫拉氏菌

<400> 1152

```
Met Leu Phe Gln Asp Phe Thr His Leu Tyr Pro Leu Ser Lys Thr Val
1               5                  10                  15

Arg Phe Glu Leu Lys Pro Ile Asp Arg Thr Leu Glu His Ile His Ala
            20                  25                  30

Lys Asn Phe Leu Ser Gln Asp Glu Thr Met Ala Asp Met His Gln Lys
        35                  40                  45

Val Lys Val Ile Leu Asp Asp Tyr His Arg Asp Phe Ile Ala Asp Met
```

50 55 60

Met Gly Glu Val Lys Leu Thr Lys Leu Ala Glu Phe Tyr Asp Val Tyr
65 70 75 80

Leu Lys Phe Arg Lys Asn Pro Lys Asp Asp Glu Leu Gln Lys Gln Leu
85 90 95

Lys Asp Leu Gln Ala Val Leu Arg Lys Glu Ile Val Lys Pro Ile Gly
100 105 110

Asn Gly Gly Lys Tyr Lys Ala Gly Tyr Asp Arg Leu Phe Gly Ala Lys
115 120 125

Leu Phe Lys Asp Gly Lys Glu Leu Gly Asp Leu Ala Lys Phe Val Ile
130 135 140

Ala Gln Glu Gly Glu Ser Ser Pro Lys Leu Ala His Leu Ala His Phe
145 150 155 160

Glu Lys Phe Ser Thr Tyr Phe Thr Gly Phe His Asp Asn Arg Lys Asn
165 170 175

Met Tyr Ser Asp Glu Asp Lys His Thr Ala Ile Ala Tyr Arg Leu Ile
180 185 190

His Glu Asn Leu Pro Arg Phe Ile Asp Asn Leu Gln Ile Leu Thr Thr
195 200 205

Ile Lys Gln Lys His Ser Ala Leu Tyr Asp Gln Ile Ile Asn Glu Leu
210 215 220

Thr Ala Ser Gly Leu Asp Val Ser Leu Ala Ser His Leu Asp Gly Tyr
225 230 235 240

His Lys Leu Leu Thr Gln Glu Gly Ile Thr Ala Tyr Asn Thr Leu Leu
245 250 255

Gly Gly Ile Ser Gly Glu Ala Gly Ser Pro Lys Ile Gln Gly Ile Asn
260 265 270

Glu Leu Ile Asn Ser His His Asn Gln His Cys His Lys Ser Glu Arg
275 280 285

Ile Ala Lys Leu Arg Pro Leu His Lys Gln Ile Leu Ser Asp Gly Met
290 295 300

Ser Val Ser Phe Leu Pro Ser Lys Phe Ala Asp Asp Ser Glu Met Cys
305 310 315 320

Gln Ala Val Asn Glu Phe Tyr Arg His Tyr Ala Asp Val Phe Ala Lys
325 330 335

Val Gln Ser Leu Phe Asp Gly Phe Asp Asp His Gln Lys Asp Gly Ile
340 345 350

Tyr Val Glu His Lys Asn Leu Asn Glu Leu Ser Lys Gln Ala Phe Gly
355 360 365

Asp Phe Ala Leu Leu Gly Arg Val Leu Asp Gly Tyr Tyr Val Asp Val
370 375 380

Val Asn Pro Glu Phe Asn Glu Arg Phe Ala Lys Ala Lys Thr Asp Asn

385 390 395 400

Ala Lys Ala Lys Leu Thr Lys Glu Lys Asp Lys Phe Ile Lys Gly Val
405 410 415

His Ser Leu Ala Ser Leu Glu Gln Ala Ile Glu His Tyr Thr Ala Arg
420 425 430

His Asp Asp Glu Ser Val Gln Ala Gly Lys Leu Gly Gln Tyr Phe Lys
435 440 445

His Gly Leu Ala Gly Val Asp Asn Pro Ile Gln Lys Ile His Asn Asn
450 455 460

His Ser Thr Ile Lys Gly Phe Leu Glu Arg Glu Arg Pro Ala Gly Glu
465 470 475 480

Arg Ala Leu Pro Lys Ile Lys Ser Gly Lys Asn Pro Glu Met Thr Gln
485 490 495

Leu Arg Gln Leu Lys Glu Leu Leu Asp Asn Ala Leu Asn Val Ala His
500 505 510

Phe Ala Lys Leu Leu Thr Thr Lys Thr Thr Leu Asp Asn Gln Asp Gly
515 520 525

Asn Phe Tyr Gly Glu Phe Gly Val Leu Tyr Asp Glu Leu Ala Lys Ile
530 535 540

Pro Thr Leu Tyr Asn Lys Val Arg Asp Tyr Leu Ser Gln Lys Pro Phe
545 550 555 560

Ser Thr Glu Lys Tyr Lys Leu Asn Phe Gly Asn Pro Thr Leu Leu Asn
565 570 575

Gly Trp Asp Leu Asn Lys Glu Lys Asp Asn Phe Gly Val Ile Leu Gln
580 585 590

Lys Asp Gly Cys Tyr Tyr Leu Ala Leu Leu Asp Lys Ala His Lys Lys
595 600 605

Val Phe Asp Asn Ala Pro Asn Thr Gly Lys Ser Ile Tyr Gln Lys Met
610 615 620

Ile Tyr Lys Tyr Leu Glu Val Arg Lys Gln Phe Pro Lys Val Phe Phe
625 630 635 640

Ser Lys Glu Ala Ile Ala Ile Asn Tyr His Pro Ser Lys Glu Leu Val
645 650 655

Glu Ile Lys Asp Lys Gly Arg Gln Arg Ser Asp Asp Glu Arg Leu Lys
660 665 670

Leu Tyr Arg Phe Ile Leu Glu Cys Leu Lys Ile His Pro Lys Tyr Asp
675 680 685

Lys Lys Phe Glu Gly Ala Ile Gly Asp Ile Gln Leu Phe Lys Lys Asp
690 695 700

Lys Lys Gly Arg Glu Val Pro Ile Ser Glu Lys Asp Leu Phe Asp Lys
705 710 715 720

Ile Asn Gly Ile Phe Ser Ser Lys Pro Lys Leu Glu Met Glu Asp Phe

725 730 735

Phe Ile Gly Glu Phe Lys Arg Tyr Asn Pro Ser Gln Asp Leu Val Asp
740 745 750

Gln Tyr Asn Ile Tyr Lys Lys Ile Asp Ser Asn Asp Asn Arg Lys Lys
755 760 765

Glu Asn Phe Tyr Asn Asn His Pro Lys Phe Lys Lys Asp Leu Val Arg
770 775 780

Tyr Tyr Tyr Glu Ser Met Cys Lys His Glu Glu Trp Glu Glu Ser Phe
785 790 795 800

Glu Phe Ser Lys Lys Leu Gln Asp Ile Gly Cys Tyr Val Asp Val Asn
805 810 815

Glu Leu Phe Thr Glu Ile Glu Thr Arg Arg Leu Asn Tyr Lys Ile Ser
820 825 830

Phe Cys Asn Ile Asn Ala Asp Tyr Ile Asp Glu Leu Val Glu Gln Gly
835 840 845

Gln Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Pro Lys Ala
850 855 860

His Gly Lys Pro Asn Leu His Thr Leu Tyr Phe Lys Ala Leu Phe Ser
865 870 875 880

Glu Asp Asn Leu Ala Asp Pro Ile Tyr Lys Leu Asn Gly Glu Ala Gln
885 890 895

Ile Phe Tyr Arg Lys Ala Ser Leu Asp Met Asn Glu Thr Thr Ile His
900 905 910

Arg Ala Gly Glu Val Leu Glu Asn Lys Asn Pro Asp Asn Pro Lys Lys
915 920 925

Arg Gln Phe Val Tyr Asp Ile Ile Lys Asp Lys Arg Tyr Thr Gln Asp
930 935 940

Lys Phe Met Leu His Val Pro Ile Thr Met Asn Phe Gly Val Gln Gly
945 950 955 960

Met Thr Ile Lys Glu Phe Asn Lys Lys Val Asn Gln Ser Ile Gln Gln
965 970 975

Tyr Asp Glu Val Asn Val Ile Gly Ile Asp Arg Gly Glu Arg His Leu
980 985 990

Leu Tyr Leu Thr Val Ile Asn Ser Lys Gly Glu Ile Leu Glu Gln Cys
995 1000 1005

Ser Leu Asn Asp Ile Thr Thr Ala Ser Ala Asn Gly Thr Gln Met
1010 1015 1020

Thr Thr Pro Tyr His Lys Ile Leu Asp Lys Arg Glu Ile Glu Arg
1025 1030 1035

Leu Asn Ala Arg Val Gly Trp Gly Glu Ile Glu Thr Ile Lys Glu
1040 1045 1050

Leu Lys Ser Gly Tyr Leu Ser His Val Val His Gln Ile Ser Gln

1055 1060 1065

Leu Met Leu Lys Tyr Asn Ala Ile Val Val Leu Glu Asp Leu Asn
1070 1075 1080

Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Ile Tyr
1085 1090 1095

Gln Asn Phe Glu Asn Ala Leu Ile Lys Lys Leu Asn His Leu Val
1100 1105 1110

Leu Lys Asp Lys Ala Asp Asp Glu Ile Gly Ser Tyr Lys Asn Ala
1115 1120 1125

Leu Gln Leu Thr Asn Asn Phe Thr Asp Leu Lys Ser Ile Gly Lys
1130 1135 1140

Gln Thr Gly Phe Leu Phe Tyr Val Pro Ala Trp Asn Thr Ser Lys
1145 1150 1155

Ile Asp Pro Glu Thr Gly Phe Val Asp Leu Leu Lys Pro Arg Tyr
1160 1165 1170

Glu Asn Ile Ala Gln Ser Gln Ala Phe Phe Gly Lys Phe Asp Lys
1175 1180 1185

Ile Cys Tyr Asn Ala Asp Lys Asp Tyr Phe Glu Phe His Ile Asp
1190 1195 1200

Tyr Ala Lys Phe Thr Asp Lys Ala Lys Asn Ser Arg Gln Ile Trp
1205 1210 1215

Thr Ile Cys Ser His Gly Asp Lys Arg Tyr Val Tyr Asp Lys Thr
1220 1225 1230

Ala Asn Gln Asn Lys Gly Ala Ala Lys Gly Ile Asn Val Asn Asp
1235 1240 1245

Glu Leu Lys Ser Leu Phe Ala Arg His His Ile Asn Glu Lys Gln
1250 1255 1260

Pro Asn Leu Val Met Asp Ile Cys Gln Asn Asn Asp Lys Glu Phe
1265 1270 1275

His Lys Ser Leu Met Tyr Leu Leu Lys Thr Leu Leu Ala Leu Arg
1280 1285 1290

Tyr Ser Asn Ala Ser Ser Asp Glu Asp Phe Ile Leu Ser Pro Val
1295 1300 1305

Ala Asn Asp Glu Gly Val Phe Phe Asn Ser Ala Leu Ala Asp Asp
1310 1315 1320

Thr Gln Pro Gln Asn Ala Asp Ala Asn Gly Ala Tyr His Ile Ala
1325 1330 1335

Leu Lys Gly Leu Trp Leu Leu Asn Glu Leu Lys Asn Ser Asp Asp
1340 1345 1350

Leu Asn Lys Val Lys Leu Ala Ile Asp Asn Gln Thr Trp Leu Asn
1355 1360 1365

Phe Ala Gln Asn Arg

1370

<210> 1153
<211> 1264
<212> PRT
<213> 山羊莫拉氏菌

<400> 1153

```
Met Leu Phe Gln Asp Phe Thr His Leu Tyr Pro Leu Ser Lys Thr Met
1               5                  10                  15

Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Leu Glu His Ile His Ala
            20                  25                  30

Lys Asn Phe Leu Ser Gln Asp Glu Thr Met Ala Asp Met Tyr Gln Lys
        35                  40                  45

Val Lys Ala Ile Leu Asp Asp Tyr His Arg Asp Phe Ile Ala Asp Met
    50                  55                  60

Met Gly Glu Val Lys Leu Thr Lys Leu Ala Glu Phe Tyr Asp Val Tyr
65                  70                  75                  80

Leu Lys Phe Arg Lys Asn Pro Lys Asp Asp Gly Leu Gln Lys Gln Leu
                85                  90                  95

Lys Asp Leu Gln Ala Val Leu Arg Lys Glu Ile Val Lys Pro Ile Gly
            100                 105                 110

Asn Gly Gly Lys Tyr Lys Ala Gly Tyr Asp Arg Leu Phe Gly Ala Lys
        115                 120                 125

Leu Phe Lys Asp Gly Lys Glu Leu Gly Asp Leu Ala Lys Phe Val Ile
    130                 135                 140

Ala Gln Glu Gly Glu Ser Ser Pro Lys Leu Ala His Leu Ala His Phe
145                 150                 155                 160

Glu Lys Phe Ser Thr Tyr Phe Thr Gly Phe His Asp Asn Arg Lys Asn
                165                 170                 175

Met Tyr Ser Asp Glu Asp Lys His Thr Ala Ile Thr Tyr Arg Leu Ile
            180                 185                 190

His Glu Asn Leu Pro Arg Phe Ile Asp Asn Leu Gln Ile Leu Ala Thr
        195                 200                 205

Ile Lys Gln Lys His Ser Ala Leu Tyr Asp Gln Ile Ile Asn Glu Leu
    210                 215                 220

Thr Ala Ser Gly Leu Asp Val Ser Leu Ala Ser His Leu Asp Gly Tyr
225                 230                 235                 240

His Lys Leu Leu Thr Gln Glu Gly Ile Thr Ala Tyr Asn Thr Leu Leu
                245                 250                 255

Gly Gly Ile Ser Gly Glu Ala Gly Ser Arg Lys Ile Gln Gly Ile Asn
            260                 265                 270

Glu Leu Ile Asn Ser His His Asn Gln His Cys His Lys Ser Glu Arg
        275                 280                 285

Ile Ala Lys Leu Arg Pro Leu His Lys Gln Ile Leu Ser Asp Gly Met
    290                 295                 300
```

```
Gly Val Ser Phe Leu Pro Ser Lys Phe Ala Asp Asp Ser Glu Met Cys
305                 310                 315                 320

Gln Ala Val Asn Glu Phe Tyr Arg His Tyr Ala Asp Val Phe Ala Lys
                325                 330                 335

Val Gln Ser Leu Phe Asp Gly Phe Asp Asp His Gln Lys Asp Gly Ile
            340                 345                 350

Tyr Val Glu His Lys Asn Leu Asn Glu Leu Ser Lys Gln Ala Phe Gly
        355                 360                 365

Asp Phe Ala Leu Leu Gly Arg Val Leu Asp Gly Tyr Tyr Val Asp Val
    370                 375                 380

Val Asn Pro Glu Phe Asn Glu Arg Phe Ala Lys Ala Lys Thr Asp Asn
385                 390                 395                 400

Ala Lys Ala Lys Leu Thr Lys Glu Lys Asp Lys Phe Ile Lys Gly Val
                405                 410                 415

His Ser Leu Ala Ser Leu Glu Gln Ala Ile Glu His Tyr Thr Ala Arg
            420                 425                 430

His Asp Asp Glu Ser Val Gln Ala Gly Lys Leu Gly Gln Tyr Phe Lys
        435                 440                 445

His Gly Leu Ala Gly Val Asp Asn Pro Ile Gln Lys Ile His Asn Asn
    450                 455                 460

His Ser Thr Ile Lys Gly Phe Leu Glu Arg Glu Arg Pro Ala Gly Glu
465                 470                 475                 480

Arg Ala Leu Pro Lys Ile Lys Ser Gly Lys Asn Pro Glu Met Thr Gln
                485                 490                 495

Leu Arg Gln Leu Lys Glu Leu Leu Asp Asn Ala Leu Asn Val Ala His
            500                 505                 510

Phe Ala Lys Leu Leu Thr Thr Lys Thr Thr Leu Asp Asn Gln Asp Gly
        515                 520                 525

Asn Phe Tyr Gly Glu Phe Gly Ala Leu Tyr Asp Glu Leu Ala Lys Ile
    530                 535                 540

Pro Thr Leu Tyr Asn Lys Val Arg Asp Tyr Leu Ser Gln Lys Pro Phe
545                 550                 555                 560

Ser Thr Glu Lys Tyr Lys Leu Asn Phe Gly Asn Pro Thr Leu Leu Asn
                565                 570                 575

Gly Trp Asp Leu Asn Lys Glu Lys Asp Asn Phe Gly Ile Ile Leu Gln
            580                 585                 590

Lys Asp Gly Cys Tyr Tyr Leu Ala Leu Leu Asp Lys Ala His Lys Lys
        595                 600                 605

Val Phe Asp Asn Ala Pro Asn Thr Gly Lys Asn Val Tyr Gln Lys Met
    610                 615                 620

Ile Tyr Lys Leu Leu Pro Gly Pro Asn Lys Met Leu Pro Lys Val Phe
625                 630                 635                 640
```

```
Phe Ala Lys Ser Asn Leu Asp Tyr Tyr Asn Pro Ser Ala Glu Leu Leu
                645                 650                 655

Asp Lys Tyr Ala Gln Gly Thr His Lys Lys Gly Asn Asn Phe Asn Leu
            660                 665                 670

Lys Asp Cys His Ala Leu Ile Asp Phe Phe Lys Ala Gly Ile Asn Lys
        675                 680                 685

His Pro Glu Trp Gln His Phe Gly Phe Lys Phe Ser Pro Thr Ser Ser
    690                 695                 700

Tyr Gln Asp Leu Ser Asp Phe Tyr Arg Glu Val Glu Pro Gln Gly Tyr
705                 710                 715                 720

Gln Val Lys Phe Val Asp Ile Asn Ala Asp Tyr Ile Asn Glu Leu Val
                725                 730                 735

Glu Gln Gly Gln Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser
            740                 745                 750

Pro Lys Ala His Gly Lys Pro Asn Leu His Thr Leu Tyr Phe Lys Ala
        755                 760                 765

Leu Phe Ser Lys Asp Asn Leu Ala Asn Pro Ile Tyr Lys Leu Asn Gly
    770                 775                 780

Glu Ala Gln Ile Phe Tyr Arg Lys Ala Ser Leu Asp Met Asn Glu Thr
785                 790                 795                 800

Thr Ile His Arg Ala Gly Glu Val Leu Glu Asn Lys Asn Pro Asp Asn
                805                 810                 815

Pro Lys Lys Arg Gln Phe Val Tyr Asp Ile Ile Lys Asp Lys Arg Tyr
            820                 825                 830

Thr Gln Asp Lys Phe Met Leu His Val Pro Ile Thr Met Asn Phe Gly
        835                 840                 845

Val Gln Gly Met Thr Ile Lys Glu Phe Asn Lys Lys Val Asn Gln Ser
    850                 855                 860

Ile Gln Gln Tyr Asp Glu Val Asn Val Ile Gly Ile Asp Arg Gly Glu
865                 870                 875                 880

Arg His Leu Leu Tyr Leu Thr Val Ile Asn Ser Lys Gly Glu Ile Leu
                885                 890                 895

Glu Gln Arg Ser Leu Asn Asp Ile Thr Thr Ala Ser Ala Asn Gly Thr
            900                 905                 910

Gln Met Thr Thr Pro Tyr His Lys Ile Leu Asp Lys Arg Glu Ile Glu
        915                 920                 925

Arg Leu Asn Ala Arg Val Gly Trp Gly Glu Ile Glu Thr Ile Lys Glu
    930                 935                 940

Leu Lys Ser Gly Tyr Leu Ser His Val Val His Gln Ile Ser Gln Leu
945                 950                 955                 960

Met Leu Lys Tyr Asn Ala Ile Val Val Leu Glu Asp Leu Asn Phe Gly
                965                 970                 975
```

```
Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Ile Tyr Gln Asn Phe
            980                 985                 990

Glu Asn Ala Leu Ile Lys Lys Leu Asn His Leu Val Leu Lys Asp Glu
        995                1000                1005

Ala Asp  Asp Glu Ile Gly Ser  Tyr Lys Asn Ala Leu  Gln Leu Thr
    1010                1015                1020

Asn Asn  Phe Thr Asp Leu Lys  Ser Ile Gly Lys Gln  Thr Gly Phe
    1025                1030                1035

Leu Phe  Tyr Val Pro Ala Trp  Asn Thr Ser Lys Ile  Asp Pro Glu
    1040                1045                1050

Thr Gly  Phe Val Asp Leu Leu  Lys Pro Arg Tyr Glu  Asn Ile Ala
    1055                1060                1065

Gln Ser  Gln Ala Phe Phe Gly  Lys Phe Asp Lys Ile  Cys Tyr Asn
    1070                1075                1080

Ala Asp  Lys Asp Tyr Phe Glu  Phe His Ile Asp Tyr  Ala Lys Phe
    1085                1090                1095

Thr Asp  Lys Ala Lys Asn Ser  Arg Gln Ile Trp Lys  Ile Cys Ser
    1100                1105                1110

His Gly  Asp Lys Arg Tyr Val  Tyr Asp Lys Thr Ala  Asn Gln Asn
    1115                1120                1125

Lys Gly  Ala Thr Lys Gly Ile  Asn Val Asn Asp Glu  Leu Lys Ser
    1130                1135                1140

Leu Phe  Ala Arg His His Ile  Asn Asp Lys Gln Pro  Asn Leu Val
    1145                1150                1155

Met Asp  Ile Cys Gln Asn Asn  Asp Lys Glu Phe His  Lys Ser Leu
    1160                1165                1170

Ile Tyr  Leu Leu Lys Thr Leu  Leu Ala Leu Arg Tyr  Ser Asn Ala
    1175                1180                1185

Ser Ser  Asp Glu Asp Phe Ile  Leu Ser Pro Val Ala  Asn Asp Glu
    1190                1195                1200

Gly Met  Phe Phe Asn Ser Ala  Leu Ala Asp Asp Thr  Gln Pro Gln
    1205                1210                1215

Asn Ala  Asp Ala Asn Gly Ala  Tyr His Ile Ala Leu  Lys Gly Leu
    1220                1225                1230

Trp Val  Leu Glu Gln Ile Lys  Asn Ser Asp Asp Leu  Asn Lys Val
    1235                1240                1245

Lys Leu  Ala Ile Asp Asn Gln  Thr Trp Leu Asn Phe  Ala Gln Asn
    1250                1255                1260

Arg
```

<210> 1154
<211> 1235
<212> PRT

<213> 厭氧弧菌屬某種

<400> 1154

```
Met Val Ala Phe Ile Asp Glu Phe Val Gly Gln Tyr Pro Val Ser Lys
1               5                  10                  15

Thr Leu Arg Phe Glu Ala Arg Pro Val Pro Glu Thr Lys Lys Trp Leu
            20                  25                  30

Glu Ser Asp Gln Cys Ser Val Leu Phe Asn Asp Gln Lys Arg Asn Glu
        35                  40                  45

Tyr Tyr Gly Val Leu Lys Glu Leu Leu Asp Asp Tyr Tyr Arg Ala Tyr
    50                  55                  60

Ile Glu Asp Ala Leu Thr Ser Phe Thr Leu Asp Lys Ala Leu Leu Glu
65                  70                  75                  80

Asn Ala Tyr Asp Leu Tyr Cys Asn Arg Asp Thr Asn Ala Phe Ser Ser
                85                  90                  95

Cys Cys Glu Lys Leu Arg Lys Asp Leu Val Lys Ala Phe Gly Asn Leu
            100                 105                 110

Lys Asp Tyr Leu Leu Gly Ser Asp Gln Leu Lys Asp Leu Val Lys Leu
        115                 120                 125

Lys Ala Lys Val Asp Ala Pro Ala Gly Lys Gly Lys Lys Lys Ile Glu
    130                 135                 140

Val Asp Ser Arg Leu Ile Asn Trp Leu Asn Asn Asn Ala Lys Tyr Ser
145                 150                 155                 160

Ala Glu Asp Arg Glu Lys Tyr Ile Lys Ala Ile Glu Ser Phe Glu Gly
                165                 170                 175

Phe Val Thr Tyr Leu Thr Asn Tyr Lys Gln Ala Arg Glu Asn Met Phe
            180                 185                 190

Ser Ser Glu Asp Lys Ser Thr Ala Ile Ala Phe Arg Val Ile Asp Gln
        195                 200                 205

Asn Met Val Thr Tyr Phe Gly Asn Ile Arg Ile Tyr Glu Lys Ile Lys
    210                 215                 220

Ala Lys Tyr Pro Glu Leu Tyr Ser Ala Leu Lys Gly Phe Glu Lys Phe
225                 230                 235                 240

Phe Ser Pro Thr Ala Tyr Ser Glu Ile Leu Ser Gln Ser Lys Ile Asp
                245                 250                 255

Glu Tyr Asn Tyr Gln Cys Ile Gly Arg Pro Ile Asp Asp Ala Asp Phe
            260                 265                 270

Lys Gly Val Asn Ser Leu Ile Asn Glu Tyr Arg Gln Lys Asn Gly Ile
        275                 280                 285

Lys Ala Arg Glu Leu Pro Val Met Ser Met Leu Tyr Lys Gln Ile Leu
    290                 295                 300

Ser Asp Arg Asp Asn Ser Phe Met Ser Glu Val Ile Asn Arg Asn Glu
305                 310                 315                 320

Glu Ala Ile Glu Cys Ala Lys Asn Gly Tyr Lys Val Ser Tyr Ala Leu
```

325 330 335

Phe Asn Glu Leu Leu Gln Leu Tyr Lys Lys Ile Phe Thr Glu Asp Asn
340 345 350

Tyr Gly Asn Ile Tyr Val Lys Thr Gln Pro Leu Thr Glu Leu Ser Gln
355 360 365

Ala Leu Phe Gly Asp Trp Ser Ile Leu Arg Asn Ala Leu Asp Asn Gly
370 375 380

Lys Tyr Asp Lys Asp Ile Ile Asn Leu Ala Glu Leu Glu Lys Tyr Phe
385 390 395 400

Ser Glu Tyr Cys Lys Val Leu Asp Ala Asp Asp Ala Ala Lys Ile Gln
405 410 415

Asp Lys Phe Asn Leu Lys Asp Tyr Phe Ile Gln Lys Asn Ala Leu Asp
420 425 430

Ala Thr Leu Pro Asp Leu Asp Lys Ile Thr Gln Tyr Lys Pro His Leu
435 440 445

Asp Ala Met Leu Gln Ala Ile Arg Lys Tyr Lys Leu Phe Ser Met Tyr
450 455 460

Asn Gly Arg Lys Lys Met Asp Val Pro Glu Asn Gly Ile Asp Phe Ser
465 470 475 480

Asn Glu Phe Asn Ala Ile Tyr Asp Lys Leu Ser Glu Phe Ser Ile Leu
485 490 495

Tyr Asp Arg Ile Arg Asn Phe Ala Thr Lys Lys Pro Tyr Ser Asp Glu
500 505 510

Lys Met Lys Leu Ser Phe Asn Met Pro Thr Met Leu Ala Gly Trp Asp
515 520 525

Tyr Asn Asn Glu Thr Ala Asn Gly Cys Phe Leu Phe Ile Lys Asp Gly
530 535 540

Lys Tyr Phe Leu Gly Val Ala Asp Ser Lys Ser Lys Asn Ile Phe Asp
545 550 555 560

Phe Lys Lys Asn Pro His Leu Leu Asp Lys Tyr Ser Ser Lys Asp Ile
565 570 575

Tyr Tyr Lys Val Lys Tyr Lys Gln Val Ser Gly Ser Ala Lys Met Leu
580 585 590

Pro Lys Val Val Phe Ala Gly Ser Asn Glu Lys Ile Phe Gly His Leu
595 600 605

Ile Ser Lys Arg Ile Leu Glu Ile Arg Glu Lys Lys Leu Tyr Thr Ala
610 615 620

Ala Ala Gly Asp Arg Lys Ala Val Ala Glu Trp Ile Asp Phe Met Lys
625 630 635 640

Ser Ala Ile Ala Ile His Pro Glu Trp Asn Glu Tyr Phe Lys Phe Lys
645 650 655

Phe Lys Asn Thr Ala Glu Tyr Asp Asn Ala Asn Lys Phe Tyr Glu Asp

660 665 670

Ile Asp Lys Gln Thr Tyr Ser Leu Glu Lys Val Glu Ile Pro Thr Glu
675 680 685

Tyr Ile Asp Glu Met Val Ser Gln His Lys Leu Tyr Leu Phe Gln Leu
690 695 700

Tyr Thr Lys Asp Phe Ser Asp Lys Lys Lys Lys Lys Gly Thr Asp Asn
705 710 715 720

Leu His Thr Met Tyr Trp His Gly Val Phe Ser Asp Glu Asn Leu Lys
725 730 735

Ala Val Thr Glu Gly Thr Gln Pro Ile Ile Lys Leu Asn Gly Glu Ala
740 745 750

Glu Met Phe Met Arg Asn Pro Ser Ile Glu Phe Gln Val Thr His Glu
755 760 765

His Asn Lys Pro Ile Ala Asn Lys Asn Pro Leu Asn Thr Lys Lys Glu
770 775 780

Ser Val Phe Asn Tyr Asp Leu Ile Lys Asp Lys Arg Tyr Thr Glu Arg
785 790 795 800

Lys Phe Tyr Phe His Cys Pro Ile Thr Leu Asn Phe Arg Ala Asp Lys
805 810 815

Pro Ile Lys Tyr Asn Glu Lys Ile Asn Arg Phe Val Glu Asn Asn Pro
820 825 830

Asp Val Cys Ile Ile Gly Ile Asp Arg Gly Glu Arg His Leu Leu Tyr
835 840 845

Tyr Thr Val Ile Asn Gln Thr Gly Asp Ile Leu Glu Gln Gly Ser Leu
850 855 860

Asn Lys Ile Ser Gly Ser Tyr Thr Asn Asp Lys Gly Glu Lys Val Asn
865 870 875 880

Lys Glu Thr Asp Tyr His Asp Leu Leu Asp Arg Lys Glu Lys Gly Lys
885 890 895

His Val Ala Gln Gln Ala Trp Glu Thr Ile Glu Asn Ile Lys Glu Leu
900 905 910

Lys Ala Gly Tyr Leu Ser Gln Val Val Tyr Lys Leu Thr Gln Leu Met
915 920 925

Leu Gln Tyr Asn Ala Val Ile Val Leu Glu Asn Leu Asn Val Gly Phe
930 935 940

Lys Arg Gly Arg Thr Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu
945 950 955 960

Lys Ala Met Ile Asp Lys Leu Asn Tyr Leu Val Phe Lys Asp Arg Gly
965 970 975

Tyr Glu Met Asn Gly Ser Tyr Ala Lys Gly Leu Gln Leu Thr Asp Lys
980 985 990

Phe Glu Ser Phe Asp Lys Ile Gly Lys Gln Thr Gly Cys Ile Tyr Tyr

995 1000 1005

Val Ile Pro Ser Tyr Thr Ser His Ile Asp Pro Lys Thr Gly Phe
1010 1015 1020

Val Asn Leu Leu Asn Ala Lys Leu Arg Tyr Glu Asn Ile Thr Lys
1025 1030 1035

Ala Gln Asp Thr Ile Arg Lys Phe Asp Ser Ile Ser Tyr Asn Ala
1040 1045 1050

Lys Ala Asp Tyr Phe Glu Phe Ala Phe Asp Tyr Arg Ser Phe Gly
1055 1060 1065

Val Asp Met Ala Arg Asn Glu Trp Val Val Cys Thr Cys Gly Asp
1070 1075 1080

Leu Arg Trp Glu Tyr Ser Ala Lys Thr Arg Glu Thr Lys Ala Tyr
1085 1090 1095

Ser Val Thr Asp Arg Leu Lys Glu Leu Phe Lys Ala His Gly Ile
1100 1105 1110

Asp Tyr Val Gly Gly Glu Asn Leu Val Ser His Ile Thr Glu Val
1115 1120 1125

Ala Asp Lys His Phe Leu Ser Thr Leu Leu Phe Tyr Leu Arg Leu
1130 1135 1140

Val Leu Lys Met Arg Tyr Thr Val Ser Gly Thr Glu Asn Glu Asn
1145 1150 1155

Asp Phe Ile Leu Ser Pro Val Glu Tyr Ala Pro Gly Lys Phe Phe
1160 1165 1170

Asp Ser Arg Glu Ala Thr Ser Thr Glu Pro Met Asn Ala Asp Ala
1175 1180 1185

Asn Gly Ala Tyr His Ile Ala Leu Lys Gly Leu Met Thr Ile Arg
1190 1195 1200

Gly Ile Glu Asp Gly Lys Leu His Asn Tyr Gly Lys Gly Gly Glu
1205 1210 1215

Asn Ala Ala Trp Phe Lys Phe Met Gln Asn Gln Glu Tyr Lys Asn
1220 1225 1230

Asn Gly
1235

<210> 1155
<211> 1154
<212> PRT
<213> Proteocatella sphenisci

<400> 1155
Met Glu Asn Phe Lys Asn Leu Tyr Pro Ile Asn Lys Thr Leu Arg Phe
1 5 10 15

Glu Leu Arg Pro Tyr Gly Lys Thr Leu Glu Asn Phe Lys Lys Ser Gly
20 25 30

Leu Leu Glu Lys Asp Ala Phe Lys Ala Asn Ser Arg Arg Ser Met Gln
35 40 45

```
Ala Ile Ile Asp Glu Lys Phe Lys Glu Thr Ile Glu Glu Arg Leu Lys
    50                  55                  60

Tyr Thr Glu Phe Ser Glu Cys Asp Leu Gly Asn Met Thr Ser Lys Asp
65                  70                  75                  80

Lys Lys Ile Thr Asp Lys Ala Ala Thr Asn Leu Lys Lys Gln Val Ile
                85                  90                  95

Leu Ser Phe Asp Asp Glu Ile Phe Asn Asn Tyr Leu Lys Pro Asp Lys
            100                 105                 110

Asn Ile Asp Ala Leu Phe Lys Asn Asp Pro Ser Asn Pro Val Ile Ser
        115                 120                 125

Thr Phe Lys Gly Phe Thr Thr Tyr Phe Val Asn Phe Phe Glu Ile Arg
    130                 135                 140

Lys His Ile Phe Lys Gly Glu Ser Ser Gly Ser Met Ala Tyr Arg Ile
145                 150                 155                 160

Ile Asp Glu Asn Leu Thr Thr Tyr Leu Asn Asn Ile Glu Lys Ile Lys
                165                 170                 175

Lys Leu Pro Glu Glu Leu Lys Ser Gln Leu Glu Gly Ile Asp Gln Ile
            180                 185                 190

Asp Lys Leu Asn Asn Tyr Asn Glu Phe Ile Thr Gln Ser Gly Ile Thr
        195                 200                 205

His Tyr Asn Glu Ile Ile Gly Gly Ile Ser Lys Ser Glu Asn Val Lys
    210                 215                 220

Ile Gln Gly Ile Asn Glu Gly Ile Asn Leu Tyr Cys Gln Lys Asn Lys
225                 230                 235                 240

Val Lys Leu Pro Arg Leu Thr Pro Leu Tyr Lys Met Ile Leu Ser Asp
                245                 250                 255

Arg Val Ser Asn Ser Phe Val Leu Asp Thr Ile Glu Asn Asp Thr Glu
            260                 265                 270

Leu Ile Glu Met Ile Ser Asp Leu Ile Asn Lys Thr Glu Ile Ser Gln
        275                 280                 285

Asp Val Ile Met Ser Asp Ile Gln Asn Ile Phe Ile Lys Tyr Lys Gln
    290                 295                 300

Leu Gly Asn Leu Pro Gly Ile Ser Tyr Ser Ser Ile Val Asn Ala Ile
305                 310                 315                 320

Cys Ser Asp Tyr Asp Asn Asn Phe Gly Asp Gly Lys Arg Lys Lys Ser
                325                 330                 335

Tyr Glu Asn Asp Arg Lys Lys His Leu Glu Thr Asn Val Tyr Ser Ile
            340                 345                 350

Asn Tyr Ile Ser Glu Leu Leu Thr Asp Thr Asp Val Ser Ser Asn Ile
        355                 360                 365

Lys Met Arg Tyr Lys Glu Leu Glu Gln Asn Tyr Gln Val Cys Lys Glu
    370                 375                 380
```

```
Asn Phe Asn Ala Thr Asn Trp Met Asn Ile Lys Asn Ile Lys Gln Ser
385                 390                 395                 400

Glu Lys Thr Asn Leu Ile Lys Asp Leu Leu Asp Ile Leu Lys Ser Ile
                405                 410                 415

Gln Arg Phe Tyr Asp Leu Phe Asp Ile Val Asp Glu Asp Lys Asn Pro
            420                 425                 430

Ser Ala Glu Phe Tyr Thr Trp Leu Ser Lys Asn Ala Glu Lys Leu Asp
        435                 440                 445

Phe Glu Phe Asn Ser Val Tyr Asn Lys Ser Arg Asn Tyr Leu Thr Arg
    450                 455                 460

Lys Gln Tyr Ser Asp Lys Lys Ile Lys Leu Asn Phe Asp Ser Pro Thr
465                 470                 475                 480

Leu Ala Lys Gly Trp Asp Ala Asn Lys Glu Ile Asp Asn Ser Thr Ile
                485                 490                 495

Ile Met Arg Lys Phe Asn Asn Asp Arg Gly Asp Tyr Asp Tyr Phe Leu
            500                 505                 510

Gly Ile Trp Asn Lys Ser Thr Pro Ala Asn Glu Lys Ile Ile Pro Leu
        515                 520                 525

Glu Asp Asn Gly Leu Phe Glu Lys Met Gln Tyr Lys Leu Tyr Pro Asp
    530                 535                 540

Pro Ser Lys Met Leu Pro Lys Gln Phe Leu Ser Lys Ile Trp Lys Ala
545                 550                 555                 560

Lys His Pro Thr Thr Pro Glu Phe Asp Lys Lys Tyr Lys Glu Gly Arg
                565                 570                 575

His Lys Lys Gly Pro Asp Phe Glu Lys Glu Phe Leu His Glu Leu Ile
            580                 585                 590

Asp Cys Phe Lys His Gly Leu Val Asn His Asp Glu Lys Tyr Gln Asp
        595                 600                 605

Val Phe Gly Phe Asn Leu Arg Asn Thr Glu Asp Tyr Asn Ser Tyr Thr
    610                 615                 620

Glu Phe Leu Glu Asp Val Glu Arg Cys Asn Tyr Asn Leu Ser Phe Asn
625                 630                 635                 640

Lys Ile Ala Asp Thr Ser Asn Leu Ile Asn Asp Gly Lys Leu Tyr Val
                645                 650                 655

Phe Gln Ile Trp Ser Lys Asp Phe Ser Ile Asp Ser Lys Gly Thr Lys
            660                 665                 670

Asn Leu Asn Thr Ile Tyr Phe Glu Ser Leu Phe Ser Glu Glu Asn Met
        675                 680                 685

Ile Glu Lys Met Phe Lys Leu Ser Gly Glu Ala Glu Ile Phe Tyr Arg
    690                 695                 700

Pro Ala Ser Leu Asn Tyr Cys Glu Asp Ile Ile Lys Lys Gly His His
705                 710                 715                 720
```

```
His Ala Glu Leu Lys Asp Lys Phe Asp Tyr Pro Ile Ile Lys Asp Lys
                725                 730                 735

Arg Tyr Ser Gln Asp Lys Phe Phe Phe His Val Pro Met Val Ile Asn
            740                 745                 750

Tyr Lys Ser Glu Lys Leu Asn Ser Lys Ser Leu Asn Asn Arg Thr Asn
        755                 760                 765

Glu Asn Leu Gly Gln Phe Thr His Ile Ile Gly Ile Asp Arg Gly Glu
    770                 775                 780

Arg His Leu Ile Tyr Leu Thr Val Val Asp Val Ser Thr Gly Glu Ile
785                 790                 795                 800

Val Glu Gln Lys His Leu Asp Glu Ile Ile Asn Thr Asp Thr Lys Gly
                805                 810                 815

Val Glu His Lys Thr His Tyr Leu Asn Lys Leu Glu Glu Lys Ser Lys
            820                 825                 830

Thr Arg Asp Asn Glu Arg Lys Ser Trp Glu Ala Ile Glu Thr Ile Lys
        835                 840                 845

Glu Leu Lys Glu Gly Tyr Ile Ser His Val Ile Asn Glu Ile Gln Lys
    850                 855                 860

Leu Gln Glu Lys Tyr Asn Ala Leu Ile Val Met Glu Asn Leu Asn Tyr
865                 870                 875                 880

Gly Phe Lys Asn Ser Arg Ile Lys Val Glu Lys Gln Val Tyr Gln Lys
                885                 890                 895

Phe Glu Thr Ala Leu Ile Lys Lys Phe Asn Tyr Ile Ile Asp Lys Lys
            900                 905                 910

Asp Pro Glu Thr Tyr Ile His Gly Tyr Gln Leu Thr Asn Pro Ile Thr
        915                 920                 925

Thr Leu Asp Lys Ile Gly Asn Gln Ser Gly Ile Val Leu Tyr Ile Pro
    930                 935                 940

Ala Trp Asn Thr Ser Lys Ile Asp Pro Val Thr Gly Phe Val Asn Leu
945                 950                 955                 960

Leu Tyr Ala Asp Asp Leu Lys Tyr Lys Asn Gln Glu Gln Ala Lys Ser
                965                 970                 975

Phe Ile Gln Lys Ile Asp Asn Ile Tyr Phe Glu Asn Gly Glu Phe Lys
            980                 985                 990

Phe Asp Ile Asp Phe Ser Lys Trp Asn Asn Arg Tyr Ser Ile Ser Lys
        995                 1000                1005

Thr Lys Trp Thr Leu Thr Ser Tyr Gly Thr Arg Ile Gln Thr Phe
    1010                1015                1020

Arg Asn Pro Gln Lys Asn Asn Lys Trp Asp Ser Ala Glu Tyr Asp
    1025                1030                1035

Leu Thr Glu Glu Phe Lys Leu Ile Leu Asn Ile Asp Gly Thr Leu
    1040                1045                1050
```

Lys Ser Gln Asp Val Glu Thr Tyr Lys Lys Phe Met Ser Leu Phe
1055 1060 1065

Lys Leu Met Leu Gln Leu Arg Asn Ser Val Thr Gly Thr Asp Ile
1070 1075 1080

Asp Tyr Met Ile Ser Pro Val Thr Asp Lys Thr Gly Thr His Phe
1085 1090 1095

Asp Ser Arg Glu Asn Ile Lys Asn Leu Pro Ala Asp Ala Asp Ala
1100 1105 1110

Asn Gly Ala Tyr Asn Ile Ala Arg Lys Gly Ile Met Ala Ile Glu
1115 1120 1125

Asn Ile Met Asn Gly Ile Ser Asp Pro Leu Lys Ile Ser Asn Glu
1130 1135 1140

Asp Tyr Leu Lys Tyr Ile Gln Asn Gln Gln Glu
1145 1150

<210> 1156
<211> 1282
<212> PRT
<213> 挑剔真細菌

<400> 1156
Met Asn Gly Asn Arg Ser Ile Val Tyr Arg Glu Phe Val Gly Val Thr
1 5 10 15

Pro Val Ala Lys Thr Leu Arg Asn Glu Leu Arg Pro Val Gly His Thr
20 25 30

Gln Glu His Ile Ile Gln Asn Gly Leu Ile Gln Glu Asp Glu Leu Arg
35 40 45

Gln Glu Lys Ser Thr Glu Leu Lys Asn Ile Met Asp Asp Tyr Tyr Arg
50 55 60

Glu Tyr Ile Asp Lys Ser Leu Ser Gly Leu Thr Asp Leu Asp Phe Thr
65 70 75 80

Leu Leu Phe Glu Leu Met Asn Ser Val Gln Ser Ser Leu Ser Lys Asp
85 90 95

Asn Lys Lys Ala Leu Glu Lys Glu His Asn Lys Met Arg Glu Gln Ile
100 105 110

Cys Thr His Leu Gln Ser Asp Ser Asp Tyr Lys Asn Met Phe Asn Ala
115 120 125

Lys Leu Phe Lys Glu Ile Leu Pro Asp Phe Ile Lys Asn Tyr Asn Gln
130 135 140

Tyr Asp Val Lys Asp Lys Ala Gly Lys Leu Glu Thr Leu Ala Leu Phe
145 150 155 160

Asn Gly Phe Ser Thr Tyr Phe Thr Asp Phe Phe Glu Lys Arg Lys Asn
165 170 175

Val Phe Thr Lys Glu Ala Val Ser Thr Ser Ile Ala Tyr Arg Ile Val
180 185 190

His Glu Asn Ser Leu Ile Phe Leu Ala Asn Met Thr Ser Tyr Lys Lys

195 200 205

Ile Ser Glu Lys Ala Leu Asp Glu Ile Glu Val Ile Glu Lys Asn Asn
210 215 220

Gln Asp Lys Met Gly Asp Trp Glu Leu Asn Gln Ile Phe Asn Pro Asp
225 230 235 240

Phe Tyr Asn Met Val Leu Ile Gln Ser Gly Ile Asp Phe Tyr Asn Glu
245 250 255

Ile Cys Gly Val Val Asn Ala His Met Asn Leu Tyr Cys Gln Gln Thr
260 265 270

Lys Asn Asn Tyr Asn Leu Phe Lys Met Arg Lys Leu His Lys Gln Ile
275 280 285

Leu Ala Tyr Thr Ser Thr Ser Phe Glu Val Pro Lys Met Phe Glu Asp
290 295 300

Asp Met Ser Val Tyr Asn Ala Val Asn Ala Phe Ile Asp Glu Thr Glu
305 310 315 320

Lys Gly Asn Ile Ile Gly Lys Leu Lys Asp Ile Val Asn Lys Tyr Asp
325 330 335

Glu Leu Asp Glu Lys Arg Ile Tyr Ile Ser Lys Asp Phe Tyr Glu Thr
340 345 350

Leu Ser Cys Phe Met Ser Gly Asn Trp Asn Leu Ile Thr Gly Cys Val
355 360 365

Glu Asn Phe Tyr Asp Glu Asn Ile His Ala Lys Gly Lys Ser Lys Glu
370 375 380

Glu Lys Val Lys Lys Ala Val Lys Glu Asp Lys Tyr Lys Ser Ile Asn
385 390 395 400

Asp Val Asn Asp Leu Val Glu Lys Tyr Ile Asp Glu Lys Glu Arg Asn
405 410 415

Glu Phe Lys Asn Ser Asn Ala Lys Gln Tyr Ile Arg Glu Ile Ser Asn
420 425 430

Ile Ile Thr Asp Thr Glu Thr Ala His Leu Glu Tyr Asp Glu His Ile
435 440 445

Ser Leu Ile Glu Ser Glu Glu Lys Ala Asp Glu Ile Lys Lys Arg Leu
450 455 460

Asp Met Tyr Met Asn Met Tyr His Trp Val Lys Ala Phe Ile Val Asp
465 470 475 480

Glu Val Leu Asp Arg Asp Glu Met Phe Tyr Ser Asp Ile Asp Asp Ile
485 490 495

Tyr Asn Ile Leu Glu Asn Ile Val Pro Leu Tyr Asn Arg Val Arg Asn
500 505 510

Tyr Val Thr Gln Lys Pro Tyr Thr Ser Lys Lys Ile Lys Leu Asn Phe
515 520 525

Gln Ser Pro Thr Leu Ala Asn Gly Trp Ser Gln Ser Lys Glu Phe Asp

530 535 540

Asn Asn Ala Ile Ile Leu Ile Arg Asp Asn Lys Tyr Tyr Leu Ala Ile
545 550 555 560

Phe Asn Ala Lys Asn Lys Pro Asp Lys Lys Ile Ile Gln Gly Asn Ser
565 570 575

Asp Lys Lys Asn Asp Asn Asp Tyr Lys Lys Met Val Tyr Asn Leu Leu
580 585 590

Pro Gly Ala Asn Lys Met Leu Pro Lys Val Phe Leu Ser Lys Lys Gly
595 600 605

Ile Glu Thr Phe Lys Pro Ser Asp Tyr Ile Ile Ser Gly Tyr Asn Ala
610 615 620

His Lys His Ile Lys Thr Ser Glu Asn Phe Asp Ile Ser Phe Cys Arg
625 630 635 640

Asp Leu Ile Asp Tyr Phe Lys Asn Ser Ile Glu Lys His Ala Glu Trp
645 650 655

Arg Lys Tyr Glu Phe Lys Phe Ser Ala Thr Asp Ser Tyr Asn Asp Ile
660 665 670

Ser Glu Phe Tyr Arg Glu Val Glu Met Gln Gly Tyr Arg Ile Asp Trp
675 680 685

Thr Tyr Ile Ser Glu Ala Asp Ile Asn Lys Leu Asp Glu Glu Gly Lys
690 695 700

Ile Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ala Glu Asn Ser Thr
705 710 715 720

Gly Lys Glu Asn Leu His Thr Met Tyr Phe Lys Asn Ile Phe Ser Glu
725 730 735

Glu Asn Leu Lys Asn Ile Val Ile Lys Leu Asn Gly Gln Ala Glu Leu
740 745 750

Phe Tyr Arg Lys Ala Ser Val Lys Asn Pro Val Lys His Lys Lys Asp
755 760 765

Ser Val Leu Val Asn Lys Thr Tyr Lys Asn Gln Leu Asp Asn Gly Asp
770 775 780

Val Val Arg Ile Pro Ile Pro Asp Asp Ile Tyr Asn Glu Ile Tyr Lys
785 790 795 800

Met Tyr Asn Gly Tyr Ile Lys Glu Ser Asp Leu Ser Glu Ala Ala Lys
805 810 815

Glu Tyr Leu Asp Lys Val Glu Val Arg Thr Ala Gln Lys Asp Ile Val
820 825 830

Lys Asp Tyr Arg Tyr Thr Val Asp Lys Tyr Phe Ile His Thr Pro Ile
835 840 845

Thr Ile Asn Tyr Lys Val Thr Ala Arg Asn Asn Val Asn Asp Met Ala
850 855 860

Val Lys Tyr Ile Ala Gln Asn Asp Asp Ile His Val Ile Gly Ile Asp

865 870 875 880

Arg Gly Glu Arg Asn Leu Ile Tyr Ile Ser Val Ile Asp Ser His Gly
885 890 895

Asn Ile Val Lys Gln Lys Ser Tyr Asn Ile Leu Asn Asn Tyr Asp Tyr
900 905 910

Lys Lys Lys Leu Val Glu Lys Glu Lys Thr Arg Glu Tyr Ala Arg Lys
915 920 925

Asn Trp Lys Ser Ile Gly Asn Ile Lys Glu Leu Lys Glu Gly Tyr Ile
930 935 940

Ser Gly Val Val His Glu Ile Ala Met Leu Met Val Glu Tyr Asn Ala
945 950 955 960

Ile Ile Ala Met Glu Asp Leu Asn Tyr Gly Phe Lys Arg Gly Arg Phe
965 970 975

Lys Val Glu Arg Gln Val Tyr Gln Lys Phe Glu Ser Met Leu Ile Asn
980 985 990

Lys Leu Asn Tyr Phe Ala Ser Lys Gly Lys Ser Val Asp Glu Pro Gly
995 1000 1005

Gly Leu Leu Lys Gly Tyr Gln Leu Thr Tyr Val Pro Asp Asn Ile
1010 1015 1020

Lys Asn Leu Gly Lys Gln Cys Gly Val Ile Phe Tyr Val Pro Ala
1025 1030 1035

Ala Phe Thr Ser Lys Ile Asp Pro Ser Thr Gly Phe Ile Ser Ala
1040 1045 1050

Phe Asn Phe Lys Ser Ile Ser Thr Asn Ala Ser Arg Lys Gln Phe
1055 1060 1065

Phe Met Gln Phe Asp Glu Ile Arg Tyr Cys Ala Glu Lys Asp Met
1070 1075 1080

Phe Ser Phe Gly Phe Asp Tyr Asn Asn Phe Asp Thr Tyr Asn Ile
1085 1090 1095

Thr Met Gly Lys Thr Gln Trp Thr Val Tyr Thr Asn Gly Glu Arg
1100 1105 1110

Leu Gln Ser Glu Phe Asn Asn Ala Arg Arg Thr Gly Lys Thr Lys
1115 1120 1125

Ser Ile Asn Leu Thr Glu Thr Ile Lys Leu Leu Leu Glu Asp Asn
1130 1135 1140

Glu Ile Asn Tyr Ala Asp Gly His Asp Val Arg Ile Asp Met Glu
1145 1150 1155

Lys Met Tyr Glu Asp Lys Asn Ser Glu Phe Phe Ala Gln Leu Leu
1160 1165 1170

Ser Leu Tyr Lys Leu Thr Val Gln Met Arg Asn Ser Tyr Thr Glu
1175 1180 1185

Ala Glu Glu Gln Glu Lys Gly Ile Ser Tyr Asp Lys Ile Ile Ser

1190 1195 1200

Pro Val Ile Asn Asp Glu Gly Glu Phe Phe Asp Ser Asp Asn Tyr
1205 1210 1215

Lys Glu Ser Asp Asp Lys Glu Cys Lys Met Pro Lys Asp Ala Asp
1220 1225 1230

Ala Asn Gly Ala Tyr Cys Ile Ala Leu Lys Gly Leu Tyr Glu Val
1235 1240 1245

Leu Lys Ile Lys Ser Glu Trp Thr Glu Asp Gly Phe Asp Arg Asn
1250 1255 1260

Cys Leu Lys Leu Pro His Ala Glu Trp Leu Asp Phe Ile Gln Asn
1265 1270 1275

Lys Arg Tyr Glu
1280

<210> 1157
<211> 1282
<212> PRT
<213> 挑剔真細菌

<400> 1157
Met Asn Gly Asn Arg Ser Ile Val Tyr Arg Glu Phe Val Gly Val Ile
1 5 10 15

Pro Val Ala Lys Thr Leu Arg Asn Glu Leu Arg Pro Val Gly His Thr
20 25 30

Gln Glu His Ile Ile Gln Asn Gly Leu Ile Gln Glu Asp Glu Leu Arg
35 40 45

Gln Glu Lys Ser Thr Glu Leu Lys Asn Ile Met Asp Asp Tyr Tyr Arg
50 55 60

Glu Tyr Ile Asp Lys Ser Leu Ser Gly Val Thr Asp Leu Asp Phe Thr
65 70 75 80

Leu Leu Phe Glu Leu Met Asn Leu Val Gln Ser Ser Pro Ser Lys Asp
85 90 95

Asn Lys Lys Ala Leu Glu Lys Glu Gln Ser Lys Met Arg Glu Gln Ile
100 105 110

Cys Thr His Leu Gln Ser Asp Ser Asn Tyr Lys Asn Ile Phe Asn Ala
115 120 125

Lys Leu Leu Lys Glu Ile Leu Pro Asp Phe Ile Lys Asn Tyr Asn Gln
130 135 140

Tyr Asp Val Lys Asp Lys Ala Gly Lys Leu Glu Thr Leu Ala Leu Phe
145 150 155 160

Asn Gly Phe Ser Thr Tyr Phe Thr Asp Phe Phe Glu Lys Arg Lys Asn
165 170 175

Val Phe Thr Lys Glu Ala Val Ser Thr Ser Ile Ala Tyr Arg Ile Val
180 185 190

His Glu Asn Ser Leu Ile Phe Leu Ala Asn Met Thr Ser Tyr Lys Lys
195 200 205

```
Ile Ser Glu Lys Ala Leu Asp Glu Ile Glu Val Ile Glu Lys Asn Asn
    210                 215                 220

Gln Asp Lys Met Gly Asp Trp Glu Leu Asn Gln Ile Phe Asn Pro Asp
225                 230                 235                 240

Phe Tyr Asn Met Val Leu Ile Gln Ser Gly Ile Asp Phe Tyr Asn Glu
                245                 250                 255

Ile Cys Gly Val Val Asn Ala His Met Asn Leu Tyr Cys Gln Gln Thr
            260                 265                 270

Lys Asn Asn Tyr Asn Leu Phe Lys Met Arg Lys Leu His Lys Gln Ile
        275                 280                 285

Leu Ala Tyr Thr Ser Thr Ser Phe Glu Val Pro Lys Met Phe Glu Asp
    290                 295                 300

Asp Met Ser Val Tyr Asn Ala Val Asn Ala Phe Ile Asp Glu Thr Glu
305                 310                 315                 320

Lys Gly Asn Ile Ile Gly Lys Leu Lys Asp Ile Val Asn Lys Tyr Asp
                325                 330                 335

Glu Leu Asp Glu Lys Arg Ile Tyr Ile Ser Lys Asp Phe Tyr Glu Thr
            340                 345                 350

Leu Ser Cys Phe Met Ser Gly Asn Trp Asn Leu Ile Thr Gly Cys Val
        355                 360                 365

Glu Asn Phe Tyr Asp Glu Asn Ile His Ala Lys Gly Lys Ser Lys Glu
    370                 375                 380

Glu Lys Val Lys Lys Ala Val Lys Glu Asp Lys Tyr Lys Ser Ile Asn
385                 390                 395                 400

Asp Val Asn Asp Leu Val Glu Lys Tyr Ile Asp Glu Lys Glu Arg Asn
                405                 410                 415

Glu Phe Lys Asn Ser Asn Ala Lys Gln Tyr Ile Arg Glu Ile Ser Asn
            420                 425                 430

Ile Ile Thr Asp Thr Glu Thr Ala His Leu Glu Tyr Asp Asp His Ile
        435                 440                 445

Ser Leu Ile Glu Ser Glu Glu Lys Ala Asp Glu Met Lys Lys Arg Leu
    450                 455                 460

Asp Met Tyr Met Asn Met Tyr His Trp Ala Lys Ala Phe Ile Val Asp
465                 470                 475                 480

Glu Val Leu Asp Arg Asp Glu Met Phe Tyr Ser Asp Ile Asp Asp Ile
                485                 490                 495

Tyr Asn Ile Leu Glu Asn Ile Val Pro Leu Tyr Asn Arg Val Arg Asn
            500                 505                 510

Tyr Val Thr Gln Lys Pro Tyr Asn Ser Lys Lys Ile Lys Leu Asn Phe
        515                 520                 525

Gln Ser Pro Thr Leu Ala Asn Gly Trp Ser Gln Ser Lys Glu Phe Asp
    530                 535                 540
```

```
Asn Asn Ala Ile Ile Leu Ile Arg Asp Asn Lys Tyr Tyr Leu Ala Ile
545                 550                 555                 560

Phe Asn Ala Lys Asn Lys Pro Asp Lys Lys Ile Ile Gln Gly Asn Ser
                565                 570                 575

Asp Lys Lys Asn Asp Asn Asp Tyr Lys Lys Met Val Tyr Asn Leu Leu
            580                 585                 590

Pro Gly Ala Asn Lys Met Leu Pro Lys Val Phe Leu Ser Lys Lys Gly
        595                 600                 605

Ile Glu Thr Phe Lys Pro Ser Asp Tyr Ile Ile Ser Gly Tyr Asn Ala
    610                 615                 620

His Lys His Ile Lys Thr Ser Glu Asn Phe Asp Ile Ser Phe Cys Arg
625                 630                 635                 640

Asp Leu Ile Asp Tyr Phe Lys Asn Ser Ile Glu Lys His Ala Glu Trp
                645                 650                 655

Arg Lys Tyr Glu Phe Lys Phe Ser Ala Thr Asp Ser Tyr Ser Asp Ile
            660                 665                 670

Ser Glu Phe Tyr Arg Glu Val Glu Met Gln Gly Tyr Arg Ile Asp Trp
        675                 680                 685

Thr Tyr Ile Ser Glu Ala Asp Ile Asn Lys Leu Asp Glu Glu Gly Lys
    690                 695                 700

Ile Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ala Glu Asn Ser Thr
705                 710                 715                 720

Gly Lys Glu Asn Leu His Thr Met Tyr Phe Lys Asn Ile Phe Ser Glu
                725                 730                 735

Glu Asn Leu Lys Asp Ile Ile Ile Lys Leu Asn Gly Gln Ala Glu Leu
            740                 745                 750

Phe Tyr Arg Arg Ala Ser Val Lys Asn Pro Val Lys His Lys Lys Asp
        755                 760                 765

Ser Val Leu Val Asn Lys Thr Tyr Lys Asn Gln Leu Asp Asn Gly Asp
    770                 775                 780

Val Val Arg Ile Pro Ile Pro Asp Asp Ile Tyr Asn Glu Ile Tyr Lys
785                 790                 795                 800

Met Tyr Asn Gly Tyr Ile Lys Glu Ser Asp Leu Ser Glu Ala Ala Lys
                805                 810                 815

Glu Tyr Leu Asp Lys Val Glu Val Arg Thr Ala Gln Lys Asp Ile Val
            820                 825                 830

Lys Asp Tyr Arg Tyr Thr Val Asp Lys Tyr Phe Ile His Thr Pro Ile
        835                 840                 845

Thr Ile Asn Tyr Lys Val Thr Ala Arg Asn Asn Val Asn Asp Met Val
    850                 855                 860

Val Lys Tyr Ile Ala Gln Asn Asp Asp Ile His Val Ile Gly Ile Asp
865                 870                 875                 880
```

```
Arg Gly Glu Arg Asn Leu Ile Tyr Ile Ser Val Ile Asp Ser His Gly
                885                 890                 895

Asn Ile Val Lys Gln Lys Ser Tyr Asn Ile Leu Asn Asn Tyr Asp Tyr
            900                 905                 910

Lys Lys Lys Leu Val Glu Lys Glu Lys Thr Arg Glu Tyr Ala Arg Lys
        915                 920                 925

Asn Trp Lys Ser Ile Gly Asn Ile Lys Glu Leu Lys Glu Gly Tyr Ile
    930                 935                 940

Ser Gly Val Val His Glu Ile Ala Met Leu Ile Val Glu Tyr Asn Ala
945                 950                 955                 960

Ile Ile Ala Met Glu Asp Leu Asn Tyr Gly Phe Lys Arg Gly Arg Phe
                965                 970                 975

Lys Val Glu Arg Gln Val Tyr Gln Lys Phe Glu Ser Met Leu Ile Asn
            980                 985                 990

Lys Leu Asn Tyr Phe Ala Ser Lys Glu Lys Ser Val Asp Glu Pro Gly
        995                 1000                1005

Gly Leu  Leu Lys Gly Tyr Gln  Leu Thr Tyr Val Pro  Asp Asn Ile
    1010                1015                1020

Lys Asn  Leu Gly Lys Gln Cys  Gly Val Ile Phe Tyr  Val Pro Ala
    1025                1030                1035

Ala Phe  Thr Ser Lys Ile Asp  Pro Ser Thr Gly Phe  Ile Ser Ala
    1040                1045                1050

Phe Asn  Phe Lys Ser Ile Ser  Thr Asn Ala Ser Arg  Lys Gln Phe
    1055                1060                1065

Phe Met  Gln Phe Asp Glu Ile  Arg Tyr Cys Ala Glu  Lys Asp Met
    1070                1075                1080

Phe Ser  Phe Gly Phe Asp Tyr  Asn Asn Phe Asp Thr  Tyr Asn Ile
    1085                1090                1095

Thr Met  Gly Lys Thr Gln Trp  Thr Val Tyr Thr Asn  Gly Glu Arg
    1100                1105                1110

Leu Gln  Ser Glu Phe Asn Asn  Ala Arg Arg Thr Gly  Lys Thr Lys
    1115                1120                1125

Ser Ile  Asn Leu Thr Glu Thr  Ile Lys Leu Leu Leu  Glu Asp Asn
    1130                1135                1140

Glu Ile  Asn Tyr Ala Asp Gly  His Asp Ile Arg Ile  Asp Met Glu
    1145                1150                1155

Lys Met  Asp Glu Asp Lys Lys  Ser Glu Phe Phe Ala  Gln Leu Leu
    1160                1165                1170

Ser Leu  Tyr Lys Leu Thr Val  Gln Met Arg Asn Ser  Tyr Thr Glu
    1175                1180                1185

Ala Glu  Glu Gln Glu Asn Gly  Ile Ser Tyr Asp Lys  Ile Ile Ser
    1190                1195                1200
```

Pro Val Ile Asn Asp Glu Gly Glu Phe Phe Asp Ser Asp Asn Tyr
1205 1210 1215

Lys Glu Ser Asp Asp Lys Glu Cys Lys Met Pro Lys Asp Ala Asp
1220 1225 1230

Ala Asn Gly Ala Tyr Cys Ile Ala Leu Lys Gly Leu Tyr Glu Val
1235 1240 1245

Leu Lys Ile Lys Ser Glu Trp Thr Glu Asp Gly Phe Asp Arg Asn
1250 1255 1260

Cys Leu Lys Leu Pro His Ala Glu Trp Leu Asp Phe Ile Gln Asn
1265 1270 1275

Lys Arg Tyr Glu
1280

<210> 1158
<211> 1305
<212> PRT
<213> 真細菌某種

<400> 1158
Met Asn Lys Ala Ala Asp Asn Tyr Thr Gly Gly Asn Tyr Asp Glu Phe
1 5 10 15

Ile Ala Leu Ser Lys Val Gln Lys Thr Leu Arg Asn Glu Leu Lys Pro
20 25 30

Thr Pro Phe Thr Ala Glu His Ile Lys Gln Arg Gly Ile Ile Ser Glu
35 40 45

Asp Glu Tyr Arg Ala Gln Gln Ser Leu Glu Leu Lys Lys Ile Ala Asp
50 55 60

Glu Tyr Tyr Arg Asn Tyr Ile Thr His Lys Leu Asn Asp Ile Asn Asn
65 70 75 80

Leu Asp Phe Tyr Asn Leu Phe Asp Ala Ile Glu Glu Lys Tyr Lys Lys
85 90 95

Asn Asp Lys Asp Asn Arg Asp Lys Leu Asp Leu Val Glu Lys Ser Lys
100 105 110

Arg Gly Glu Ile Ala Lys Met Leu Ser Ala Asp Asp Asn Phe Lys Ser
115 120 125

Met Phe Glu Ala Lys Leu Ile Thr Lys Leu Leu Pro Asp Tyr Val Glu
130 135 140

Arg Asn Tyr Thr Gly Glu Asp Lys Glu Lys Ala Leu Glu Thr Leu Ala
145 150 155 160

Leu Phe Lys Gly Phe Thr Thr Tyr Phe Lys Gly Tyr Phe Lys Thr Arg
165 170 175

Lys Asn Met Phe Ser Gly Glu Gly Gly Ala Ser Ser Ile Cys His Arg
180 185 190

Ile Val Asn Val Asn Ala Ser Ile Phe Tyr Asp Asn Leu Lys Thr Phe
195 200 205

Met Arg Ile Gln Glu Lys Ala Gly Asp Glu Ile Ala Leu Ile Glu Glu

210 215 220

Glu Leu Thr Glu Lys Leu Asp Gly Trp Arg Leu Glu His Ile Phe Ser
225 230 235 240

Arg Asp Tyr Tyr Asn Glu Val Leu Ala Gln Lys Gly Ile Asp Tyr Tyr
245 250 255

Asn Gln Ile Cys Gly Asp Ile Asn Lys His Met Asn Leu Tyr Cys Gln
260 265 270

Gln Asn Lys Phe Lys Ala Asn Ile Phe Lys Met Met Lys Ile Gln Lys
275 280 285

Gln Ile Met Gly Ile Ser Glu Lys Ala Phe Glu Ile Pro Pro Met Tyr
290 295 300

Gln Asn Asp Glu Glu Val Tyr Ala Ser Phe Asn Glu Phe Ile Ser Arg
305 310 315 320

Leu Glu Glu Val Lys Leu Thr Asp Arg Leu Ile Asn Ile Leu Gln Asn
325 330 335

Ile Asn Ile Tyr Asn Thr Ala Lys Ile Tyr Ile Asn Ala Arg Tyr Tyr
340 345 350

Thr Asn Val Ser Ser Tyr Val Tyr Gly Gly Trp Gly Val Ile Asp Ser
355 360 365

Ala Ile Glu Arg Tyr Leu Tyr Asn Thr Ile Ala Gly Lys Gly Gln Ser
370 375 380

Lys Val Lys Lys Ile Glu Asn Ala Lys Lys Asp Asn Lys Phe Met Ser
385 390 395 400

Val Lys Glu Leu Asp Ser Ile Val Ala Glu Tyr Glu Pro Asp Tyr Phe
405 410 415

Asn Ala Pro Tyr Ile Asp Asp Asp Asp Asn Ala Val Lys Ala Phe Gly
420 425 430

Gly Gln Gly Val Leu Gly Tyr Phe Asn Lys Met Ser Glu Leu Leu Ala
435 440 445

Asp Val Ser Leu Tyr Thr Ile Asp Tyr Asn Ser Asp Asp Ser Leu Ile
450 455 460

Glu Asn Lys Glu Ser Ala Leu Arg Ile Lys Lys Gln Leu Asp Asp Ile
465 470 475 480

Met Ser Leu Tyr His Trp Leu Gln Thr Phe Ile Ile Asp Glu Val Val
485 490 495

Glu Lys Asp Asn Ala Phe Tyr Ala Glu Leu Glu Asp Ile Cys Cys Glu
500 505 510

Leu Glu Asn Val Val Thr Leu Tyr Asp Arg Ile Arg Asn Tyr Val Thr
515 520 525

Lys Lys Pro Tyr Ser Thr Gln Lys Phe Lys Leu Asn Phe Ala Ser Pro
530 535 540

Thr Leu Ala Ala Gly Trp Ser Arg Ser Lys Glu Phe Asp Asn Asn Ala

```
545                 550                 555                 560

Ile Ile Leu Leu Arg Asn Asn Lys Tyr Tyr Ile Ala Ile Phe Asn Val
                565                 570                 575

Asn Asn Lys Pro Asp Lys Gln Ile Ile Lys Gly Ser Glu Glu Gln Arg
            580                 585                 590

Leu Ser Thr Asp Tyr Lys Lys Met Val Tyr Asn Leu Leu Pro Gly Pro
        595                 600                 605

Asn Lys Met Leu Pro Lys Val Phe Ile Lys Ser Asp Thr Gly Lys Arg
    610                 615                 620

Asp Tyr Asn Pro Ser Ser Tyr Ile Leu Glu Gly Tyr Glu Lys Asn Arg
625                 630                 635                 640

His Ile Lys Ser Ser Gly Asn Phe Asp Ile Asn Tyr Cys His Asp Leu
                645                 650                 655

Ile Asp Tyr Tyr Lys Ala Cys Ile Asn Lys His Pro Glu Trp Lys Asn
            660                 665                 670

Tyr Gly Phe Lys Phe Lys Glu Thr Asn Gln Tyr Asn Asp Ile Gly Gln
        675                 680                 685

Phe Tyr Lys Asp Val Glu Lys Gln Gly Tyr Ser Ile Ser Trp Ala Tyr
    690                 695                 700

Ile Ser Glu Glu Asp Ile Asn Lys Leu Asp Glu Glu Gly Lys Ile Tyr
705                 710                 715                 720

Leu Phe Glu Ile Tyr Asn Lys Asp Leu Ser Ala His Ser Thr Gly Arg
                725                 730                 735

Asp Asn Leu His Thr Met Tyr Leu Lys Asn Ile Phe Ser Glu Asp Asn
            740                 745                 750

Leu Lys Asn Ile Cys Ile Glu Leu Asn Gly Glu Ala Glu Leu Phe Tyr
        755                 760                 765

Arg Lys Ser Ser Met Lys Ser Asn Ile Thr His Lys Lys Asp Thr Ile
    770                 775                 780

Leu Val Asn Lys Thr Tyr Ile Asn Glu Thr Gly Val Arg Val Ser Leu
785                 790                 795                 800

Ser Asp Glu Asp Tyr Met Lys Val Tyr Asn Tyr Tyr Asn Asn Asn Tyr
                805                 810                 815

Val Ile Asp Thr Glu Asn Asp Lys Asn Leu Ile Asp Ile Ile Glu Lys
            820                 825                 830

Ile Gly His Arg Lys Ser Lys Ile Asp Ile Val Lys Asp Lys Arg Tyr
        835                 840                 845

Thr Glu Asp Lys Tyr Phe Leu Tyr Leu Pro Ile Thr Ile Asn Tyr Gly
    850                 855                 860

Ile Glu Asp Glu Asn Val Asn Ser Lys Ile Ile Glu Tyr Ile Ala Lys
865                 870                 875                 880

Gln Asp Asn Met Asn Val Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu
```

885 890 895

Ile Tyr Ile Ser Val Ile Asp Asn Lys Gly Asn Ile Ile Glu Gln Lys
900 905 910

Ser Phe Asn Leu Val Asn Asn Tyr Asp Tyr Lys Asn Lys Leu Lys Asn
915 920 925

Met Glu Lys Thr Arg Asp Asn Ala Arg Lys Asn Trp Gln Glu Ile Gly
930 935 940

Lys Ile Lys Asp Val Lys Ser Gly Tyr Leu Ser Gly Val Ile Ser Lys
945 950 955 960

Ile Ala Arg Met Val Ile Asp Tyr Asn Ala Ile Ile Val Met Glu Asp
965 970 975

Leu Asn Lys Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Arg Gln Val
980 985 990

Tyr Gln Lys Phe Glu Asn Met Leu Ile Ser Lys Leu Asn Tyr Leu Val
995 1000 1005

Phe Lys Glu Arg Lys Ala Asp Glu Asn Gly Gly Ile Leu Arg Gly
1010 1015 1020

Tyr Gln Leu Thr Tyr Ile Pro Lys Ser Ile Lys Asn Val Gly Lys
1025 1030 1035

Gln Cys Gly Cys Ile Phe Tyr Val Pro Ala Ala Tyr Thr Ser Lys
1040 1045 1050

Ile Asp Pro Ala Thr Gly Phe Ile Asn Ile Phe Asp Phe Lys Lys
1055 1060 1065

Tyr Ser Gly Ser Gly Ile Asn Ala Lys Val Lys Asp Lys Lys Glu
1070 1075 1080

Phe Leu Met Ser Met Asn Ser Ile Arg Tyr Ile Asn Glu Cys Ser
1085 1090 1095

Glu Glu Tyr Glu Lys Ile Gly His Arg Glu Leu Phe Ala Phe Ser
1100 1105 1110

Phe Asp Tyr Asn Asn Phe Lys Thr Tyr Asn Val Ser Ser Pro Val
1115 1120 1125

Asn Glu Trp Thr Ala Tyr Thr Tyr Gly Glu Arg Ile Lys Lys Leu
1130 1135 1140

Tyr Lys Asp Gly Arg Trp Leu Arg Ser Glu Val Leu Asn Leu Thr
1145 1150 1155

Glu Asn Leu Ile Lys Leu Met Glu Gln Tyr Asn Ile Glu Tyr Lys
1160 1165 1170

Asp Gly His Asp Ile Arg Glu Asp Ile Ser His Met Asp Glu Thr
1175 1180 1185

Arg Asn Ala Asp Phe Ile Cys Ser Leu Phe Glu Glu Leu Lys Tyr
1190 1195 1200

Thr Val Gln Leu Arg Asn Ser Lys Ser Glu Ala Glu Asp Glu Asn

1205 1210 1215

Tyr Asp Arg Leu Val Ser Pro Ile Leu Asn Ser Ser Asn Gly Phe
1220 1225 1230

Tyr Asp Ser Ser Asp Tyr Met Glu Asn Glu Asn Asn Thr Thr His
1235 1240 1245

Thr Met Pro Lys Asp Ala Asp Ala Asn Gly Ala Tyr Cys Ile Ala
1250 1255 1260

Leu Lys Gly Leu Tyr Glu Ile Asn Lys Ile Lys Gln Asn Trp Ser
1265 1270 1275

Asp Asp Lys Lys Phe Lys Glu Asn Glu Leu Tyr Ile Asn Val Thr
1280 1285 1290

Glu Trp Leu Asp Tyr Ile Gln Asn Arg Arg Phe Glu
1295 1300 1305

<210> 1159
<211> 1259
<212> PRT
<213> 窮氏互養菌

<400> 1159
Met Ala Asn Ser Leu Lys Asp Phe Thr Asn Ile Tyr Gln Leu Ser Lys
1 5 10 15

Thr Leu Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Glu Glu His Ile
20 25 30

Asn Arg Lys Leu Ile Ile Met His Asp Glu Lys Arg Gly Glu Asp Tyr
35 40 45

Lys Ser Val Thr Lys Leu Ile Asp Asp Tyr His Arg Lys Phe Ile His
50 55 60

Glu Thr Leu Asp Pro Ala His Phe Asp Trp Asn Pro Leu Ala Glu Ala
65 70 75 80

Leu Ile Gln Ser Gly Ser Lys Asn Asn Lys Ala Leu Pro Ala Glu Gln
85 90 95

Lys Glu Met Arg Glu Lys Ile Ile Ser Met Phe Thr Ser Gln Ala Val
100 105 110

Tyr Lys Lys Leu Phe Lys Lys Glu Leu Phe Ser Glu Leu Leu Pro Glu
115 120 125

Met Ile Lys Ser Glu Leu Val Ser Asp Leu Glu Lys Gln Ala Gln Leu
130 135 140

Asp Ala Val Lys Ser Phe Asp Lys Phe Ser Thr Tyr Phe Thr Gly Phe
145 150 155 160

His Glu Asn Arg Lys Asn Ile Tyr Ser Lys Lys Asp Thr Ser Thr Ser
165 170 175

Ile Ala Phe Arg Ile Val His Gln Asn Phe Pro Lys Phe Leu Ala Asn
180 185 190

Val Arg Ala Tyr Thr Leu Ile Lys Glu Arg Ala Pro Glu Val Ile Asp
195 200 205

```
Lys Ala Gln Lys Glu Leu Ser Gly Ile Leu Gly Gly Lys Thr Leu Asp
    210                 215                 220

Asp Ile Phe Ser Ile Glu Ser Phe Asn Asn Val Leu Thr Gln Asp Lys
225                 230                 235                 240

Ile Asp Tyr Tyr Asn Gln Ile Ile Gly Gly Val Ser Gly Lys Ala Gly
                245                 250                 255

Asp Lys Lys Leu Arg Gly Val Asn Glu Phe Ser Asn Leu Tyr Arg Gln
            260                 265                 270

Gln His Pro Glu Val Ala Ser Leu Arg Ile Lys Met Val Pro Leu Tyr
        275                 280                 285

Lys Gln Ile Leu Ser Asp Arg Thr Thr Leu Ser Phe Val Pro Glu Ala
    290                 295                 300

Leu Lys Asp Asp Glu Gln Ala Ile Asn Ala Val Asp Gly Leu Arg Ser
305                 310                 315                 320

Glu Leu Glu Arg Asn Asp Ile Phe Asn Arg Ile Lys Arg Leu Phe Gly
                325                 330                 335

Lys Asn Asn Leu Tyr Ser Leu Asp Lys Ile Trp Ile Lys Asn Ser Ser
            340                 345                 350

Ile Ser Ala Phe Ser Asn Glu Leu Phe Lys Asn Trp Ser Phe Ile Glu
        355                 360                 365

Asp Ala Leu Lys Glu Phe Lys Glu Asn Glu Phe Asn Gly Ala Arg Ser
    370                 375                 380

Ala Gly Lys Lys Ala Glu Lys Trp Leu Lys Ser Lys Tyr Phe Ser Phe
385                 390                 395                 400

Ala Asp Ile Asp Ala Ala Val Lys Ser Tyr Ser Glu Gln Val Ser Ala
                405                 410                 415

Asp Ile Ser Ser Ala Pro Ser Ala Ser Tyr Phe Ala Lys Phe Thr Asn
            420                 425                 430

Leu Ile Glu Thr Ala Ala Glu Asn Gly Arg Lys Phe Ser Tyr Phe Ala
        435                 440                 445

Ala Glu Ser Lys Ala Phe Arg Gly Asp Asp Gly Lys Thr Glu Ile Ile
    450                 455                 460

Lys Ala Tyr Leu Asp Ser Leu Asn Asp Ile Leu His Cys Leu Lys Pro
465                 470                 475                 480

Phe Glu Thr Glu Asp Ile Ser Asp Ile Asp Thr Glu Phe Tyr Ser Ala
                485                 490                 495

Phe Ala Glu Ile Tyr Asp Ser Val Lys Asp Val Ile Pro Val Tyr Asn
            500                 505                 510

Ala Val Arg Asn Tyr Thr Thr Gln Lys Pro Phe Ser Thr Glu Lys Phe
        515                 520                 525

Lys Leu Asn Phe Glu Asn Pro Ala Leu Ala Lys Gly Trp Asp Lys Asn
    530                 535                 540
```

```
Lys Glu Gln Asn Asn Thr Ala Ile Ile Leu Met Lys Asp Gly Lys Tyr
545                 550                 555                 560

Tyr Leu Gly Val Ile Asp Lys Asn Asn Lys Leu Arg Ala Asp Asp Leu
                565                 570                 575

Ala Asp Asp Gly Ser Ala Tyr Gly Tyr Met Lys Met Asn Tyr Lys Phe
            580                 585                 590

Ile Pro Thr Pro His Met Glu Leu Pro Lys Val Phe Leu Pro Lys Arg
        595                 600                 605

Ala Pro Lys Arg Tyr Asn Pro Ser Arg Glu Ile Leu Leu Ile Lys Glu
    610                 615                 620

Asn Lys Thr Phe Ile Lys Asp Lys Asn Phe Asn Arg Thr Asp Cys His
625                 630                 635                 640

Lys Leu Ile Asp Phe Phe Lys Asp Ser Ile Asn Lys His Lys Asp Trp
                645                 650                 655

Arg Thr Phe Gly Phe Asp Phe Ser Asp Thr Asp Ser Tyr Glu Asp Ile
            660                 665                 670

Ser Asp Phe Tyr Met Glu Val Gln Asp Gln Gly Tyr Lys Leu Thr Phe
        675                 680                 685

Thr Arg Leu Ser Ala Glu Lys Ile Asp Lys Trp Val Glu Glu Gly Arg
    690                 695                 700

Leu Phe Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ala Asp Gly Ala Gln
705                 710                 715                 720

Gly Ser Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Ile Phe Ser Glu
                725                 730                 735

Glu Asn Leu Lys Asp Val Val Leu Lys Leu Asn Gly Glu Ala Glu Leu
            740                 745                 750

Phe Phe Arg Arg Lys Ser Ile Asp Lys Pro Ala Val His Ala Lys Gly
        755                 760                 765

Ser Met Lys Val Asn Arg Arg Asp Ile Asp Gly Asn Pro Ile Asp Glu
    770                 775                 780

Gly Thr Tyr Val Glu Ile Cys Gly Tyr Ala Asn Gly Lys Arg Asp Met
785                 790                 795                 800

Ala Ser Leu Asn Ala Gly Ala Arg Gly Leu Ile Glu Ser Gly Leu Val
                805                 810                 815

Arg Ile Thr Glu Val Lys His Glu Leu Val Lys Asp Lys Arg Tyr Thr
            820                 825                 830

Ile Asp Lys Tyr Phe Phe His Val Pro Phe Thr Ile Asn Phe Lys Ala
        835                 840                 845

Gln Gly Gln Gly Asn Ile Asn Ser Asp Val Asn Leu Phe Leu Arg Asn
    850                 855                 860

Asn Lys Asp Val Asn Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu
865                 870                 875                 880
```

```
Val Tyr Val Ser Leu Ile Asp Arg Asp Gly His Ile Lys Leu Gln Lys
                885                 890                 895

Asp Phe Asn Ile Ile Gly Gly Met Asp Tyr His Ala Lys Leu Asn Gln
            900                 905                 910

Lys Glu Lys Glu Arg Asp Thr Ala Arg Lys Ser Trp Lys Thr Ile Gly
        915                 920                 925

Thr Ile Lys Glu Leu Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu
    930                 935                 940

Ile Val Arg Leu Ala Val Asp Asn Asn Ala Val Ile Val Met Glu Asp
945                 950                 955                 960

Leu Asn Ile Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val
                965                 970                 975

Tyr Gln Lys Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr Leu Val
            980                 985                 990

Phe Lys Asp Ala Gly Tyr Asp Ala Pro Cys Gly Ile Leu Lys Gly Leu
        995                 1000                1005

Gln Leu  Thr Glu Lys Phe Glu  Ser Phe Thr Lys Leu  Gly Lys Gln
    1010                1015                1020

Cys Gly  Ile Ile Phe Tyr Ile  Pro Ala Gly Tyr Thr  Ser Lys Ile
    1025                1030                1035

Asp Pro  Thr Thr Gly Phe Val  Asn Leu Phe Asn Ile  Asn Asp Val
    1040                1045                1050

Ser Ser  Lys Glu Lys Gln Lys  Asp Phe Ile Gly Lys  Leu Asp Ser
    1055                1060                1065

Ile Arg  Phe Asp Ala Lys Arg  Asp Met Phe Thr Phe  Glu Phe Asp
    1070                1075                1080

Tyr Asp  Lys Phe Arg Thr Tyr  Gln Thr Ser Tyr Arg  Lys Lys Trp
    1085                1090                1095

Ala Val  Trp Thr Asn Gly Lys  Arg Ile Val Arg Glu  Lys Asp Lys
    1100                1105                1110

Asp Gly  Lys Phe Arg Met Asn  Asp Arg Leu Leu Thr  Glu Asp Met
    1115                1120                1125

Lys Asn  Ile Leu Asn Lys Tyr  Ala Leu Ala Tyr Lys  Ala Gly Glu
    1130                1135                1140

Asp Ile  Leu Pro Asp Val Ile  Ser Arg Asp Lys Ser  Leu Ala Ser
    1145                1150                1155

Glu Ile  Phe Tyr Val Phe Lys  Asn Thr Leu Gln Met  Arg Asn Ser
    1160                1165                1170

Lys Arg  Asp Thr Gly Glu Asp  Phe Ile Ile Ser Pro  Val Leu Asn
    1175                1180                1185

Ala Lys  Gly Arg Phe Phe Asp  Ser Arg Lys Thr Asp  Ala Ala Leu
    1190                1195                1200
```

Pro Ile Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Ala Leu Lys
1205 1210 1215

Gly Ser Leu Val Leu Asp Ala Ile Asp Glu Lys Leu Lys Glu Asp
1220 1225 1230

Gly Arg Ile Asp Tyr Lys Asp Met Ala Val Ser Asn Pro Lys Trp
1235 1240 1245

Phe Glu Phe Met Gln Thr Arg Lys Phe Asp Phe
1250 1255

<210> 1160
<211> 1238
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選白蟻甲烷枝原體序列

<400> 1160
Met Asn Asn Tyr Asp Glu Phe Thr Lys Leu Tyr Pro Ile Gln Lys Thr
1 5 10 15

Ile Arg Phe Glu Leu Lys Pro Gln Gly Arg Thr Met Glu His Leu Glu
20 25 30

Thr Phe Asn Phe Phe Glu Glu Asp Arg Asp Arg Ala Glu Lys Tyr Lys
35 40 45

Ile Leu Lys Glu Ala Ile Asp Glu Tyr His Lys Lys Phe Ile Asp Glu
50 55 60

His Leu Thr Asn Met Ser Leu Asp Trp Asn Ser Leu Lys Gln Ile Ser
65 70 75 80

Glu Lys Tyr Tyr Lys Ser Arg Glu Glu Lys Asp Lys Lys Val Phe Leu
85 90 95

Ser Glu Gln Lys Arg Met Arg Gln Glu Ile Val Ser Glu Phe Lys Lys
100 105 110

Asp Asp Arg Phe Lys Asp Leu Phe Ser Lys Lys Leu Phe Ser Glu Leu
115 120 125

Leu Lys Glu Glu Ile Tyr Lys Lys Gly Asn His Gln Glu Ile Asp Ala
130 135 140

Leu Lys Ser Phe Asp Lys Phe Ser Gly Tyr Phe Ile Gly Leu His Glu
145 150 155 160

Asn Arg Lys Asn Met Tyr Ser Asp Gly Asp Glu Ile Thr Ala Ile Ser
165 170 175

Asn Arg Ile Val Asn Glu Asn Phe Pro Lys Phe Leu Asp Asn Leu Gln
180 185 190

Lys Tyr Gln Glu Ala Arg Lys Lys Tyr Pro Glu Trp Ile Ile Lys Ala
195 200 205

Glu Ser Ala Leu Val Ala His Asn Ile Lys Met Asp Glu Val Phe Ser
210 215 220

Leu Glu Tyr Phe Asn Lys Val Leu Asn Gln Glu Gly Ile Gln Arg Tyr

225 230 235 240

Asn Leu Ala Leu Gly Gly Tyr Val Thr Lys Ser Gly Glu Lys Met Met
245 250 255

Gly Leu Asn Asp Ala Leu Asn Leu Ala His Gln Ser Glu Lys Ser Ser
260 265 270

Lys Gly Arg Ile His Met Thr Pro Leu Phe Lys Gln Ile Leu Ser Glu
275 280 285

Lys Glu Ser Phe Ser Tyr Ile Pro Asp Val Phe Thr Glu Asp Ser Gln
290 295 300

Leu Leu Pro Ser Ile Gly Gly Phe Phe Ala Gln Ile Glu Asn Asp Lys
305 310 315 320

Asp Gly Asn Ile Phe Asp Arg Ala Leu Glu Leu Ile Ser Ser Tyr Ala
325 330 335

Glu Tyr Asp Thr Glu Arg Ile Tyr Ile Arg Gln Ala Asp Ile Asn Arg
340 345 350

Val Ser Asn Val Ile Phe Gly Glu Trp Gly Thr Leu Gly Gly Leu Met
355 360 365

Arg Glu Tyr Lys Ala Asp Ser Ile Asn Asp Ile Asn Leu Glu Arg Thr
370 375 380

Cys Lys Lys Val Asp Lys Trp Leu Asp Ser Lys Glu Phe Ala Leu Ser
385 390 395 400

Asp Val Leu Glu Ala Ile Lys Arg Thr Gly Asn Asn Asp Ala Phe Asn
405 410 415

Glu Tyr Ile Ser Lys Met Arg Thr Ala Arg Glu Lys Ile Asp Ala Ala
420 425 430

Arg Lys Glu Met Lys Phe Ile Ser Glu Lys Ile Ser Gly Asp Glu Glu
435 440 445

Ser Ile His Ile Ile Lys Thr Leu Leu Asp Ser Val Gln Gln Phe Leu
450 455 460

His Phe Phe Asn Leu Phe Lys Ala Arg Gln Asp Ile Pro Leu Asp Gly
465 470 475 480

Ala Phe Tyr Ala Glu Phe Asp Glu Val His Ser Lys Leu Phe Ala Ile
485 490 495

Val Pro Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Lys Asn Asn Leu
500 505 510

Asn Thr Lys Lys Ile Lys Leu Asn Phe Lys Asn Pro Thr Leu Ala Asn
515 520 525

Gly Trp Asp Gln Asn Lys Val Tyr Asp Tyr Ala Ser Leu Ile Phe Leu
530 535 540

Arg Asp Gly Asn Tyr Tyr Leu Gly Ile Ile Asn Pro Lys Arg Lys Lys
545 550 555 560

Asn Ile Lys Phe Glu Gln Gly Ser Gly Asn Gly Pro Phe Tyr Arg Lys

565 570 575

Met Val Tyr Lys Gln Ile Pro Gly Pro Asn Lys Asn Leu Pro Arg Val
580 585 590

Phe Leu Thr Ser Thr Lys Gly Lys Lys Glu Tyr Lys Pro Ser Lys Glu
595 600 605

Ile Ile Glu Gly Tyr Glu Ala Asp Lys His Ile Arg Gly Asp Lys Phe
610 615 620

Asp Leu Asp Phe Cys His Lys Leu Ile Asp Phe Phe Lys Glu Ser Ile
625 630 635 640

Glu Lys His Lys Asp Trp Ser Lys Phe Asn Phe Tyr Phe Ser Pro Thr
645 650 655

Glu Ser Tyr Gly Asp Ile Ser Glu Phe Tyr Leu Asp Val Glu Lys Gln
660 665 670

Gly Tyr Arg Met His Phe Glu Asn Ile Ser Ala Glu Thr Ile Asp Glu
675 680 685

Tyr Val Glu Lys Gly Asp Leu Phe Leu Phe Gln Ile Tyr Asn Lys Asp
690 695 700

Phe Val Lys Ala Ala Thr Gly Lys Lys Asp Met His Thr Ile Tyr Trp
705 710 715 720

Asn Ala Ala Phe Ser Pro Glu Asn Leu Gln Asp Val Val Val Lys Leu
725 730 735

Asn Gly Glu Ala Glu Leu Phe Tyr Arg Asp Lys Ser Asp Ile Lys Glu
740 745 750

Ile Val His Arg Glu Gly Glu Ile Leu Val Asn Arg Thr Tyr Asn Gly
755 760 765

Arg Thr Pro Val Pro Asp Lys Ile His Lys Lys Leu Thr Asp Tyr His
770 775 780

Asn Gly Arg Thr Lys Asp Leu Gly Glu Ala Lys Glu Tyr Leu Asp Lys
785 790 795 800

Val Arg Tyr Phe Lys Ala His Tyr Asp Ile Thr Lys Asp Arg Arg Tyr
805 810 815

Leu Asn Asp Lys Ile Tyr Phe His Val Pro Leu Thr Leu Asn Phe Lys
820 825 830

Ala Asn Gly Lys Lys Asn Leu Asn Lys Met Val Ile Glu Lys Phe Leu
835 840 845

Ser Asp Glu Lys Ala His Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn
850 855 860

Leu Leu Tyr Tyr Ser Ile Ile Asp Arg Ser Gly Lys Ile Ile Asp Gln
865 870 875 880

Gln Ser Leu Asn Val Ile Asp Gly Phe Asp Tyr Arg Glu Lys Leu Asn
885 890 895

Gln Arg Glu Ile Glu Met Lys Asp Ala Arg Gln Ser Trp Asn Ala Ile

900 905 910

Gly Lys Ile Lys Asp Leu Lys Glu Gly Tyr Leu Ser Lys Ala Val His
915 920 925

Glu Ile Thr Lys Met Ala Ile Gln Tyr Asn Ala Ile Val Val Met Glu
930 935 940

Glu Leu Asn Tyr Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln
945 950 955 960

Ile Tyr Gln Lys Phe Glu Asn Met Leu Ile Asp Lys Met Asn Tyr Leu
965 970 975

Val Phe Lys Asp Ala Pro Asp Glu Ser Pro Gly Gly Val Leu Asn Ala
980 985 990

Tyr Gln Leu Thr Asn Pro Leu Glu Ser Phe Ala Lys Leu Gly Lys Gln
995 1000 1005

Thr Gly Ile Leu Phe Tyr Val Pro Ala Ala Tyr Thr Ser Lys Ile
1010 1015 1020

Asp Pro Thr Thr Gly Phe Val Asn Leu Phe Asn Thr Ser Ser Lys
1025 1030 1035

Thr Asn Ala Gln Glu Arg Lys Glu Phe Leu Gln Lys Phe Glu Ser
1040 1045 1050

Ile Ser Tyr Ser Ala Lys Asp Gly Gly Ile Phe Ala Phe Ala Phe
1055 1060 1065

Asp Tyr Arg Lys Phe Gly Thr Ser Lys Thr Asp His Lys Asn Val
1070 1075 1080

Trp Thr Ala Tyr Thr Asn Gly Glu Arg Met Arg Tyr Ile Lys Glu
1085 1090 1095

Lys Lys Arg Asn Glu Leu Phe Asp Pro Ser Lys Glu Ile Lys Glu
1100 1105 1110

Ala Leu Thr Ser Ser Gly Ile Lys Tyr Asp Gly Gly Gln Asn Ile
1115 1120 1125

Leu Pro Asp Ile Leu Arg Ser Asn Asn Asn Gly Leu Ile Tyr Thr
1130 1135 1140

Met Tyr Ser Ser Phe Ile Ala Ala Ile Gln Met Arg Val Tyr Asp
1145 1150 1155

Gly Lys Glu Asp Tyr Ile Ile Ser Pro Ile Lys Asn Ser Lys Gly
1160 1165 1170

Glu Phe Phe Arg Thr Asp Pro Lys Arg Arg Glu Leu Pro Ile Asp
1175 1180 1185

Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala Leu Arg Gly Glu Leu
1190 1195 1200

Thr Met Arg Ala Ile Ala Glu Lys Phe Asp Pro Asp Ser Glu Lys
1205 1210 1215

Met Ala Lys Leu Glu Leu Lys His Lys Asp Trp Phe Glu Phe Met

1220 1225 1230

Gln Thr Arg Gly Asp
1235

<210> 1161
<211> 1227
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選alvus甲烷嗜甲基菌序列

<400> 1161
Met Asp Ala Lys Glu Phe Thr Gly Gln Tyr Pro Leu Ser Lys Thr Leu
1 5 10 15

Arg Phe Glu Leu Arg Pro Ile Gly Arg Thr Trp Asp Asn Leu Glu Ala
20 25 30

Ser Gly Tyr Leu Ala Glu Asp Arg His Arg Ala Glu Cys Tyr Pro Arg
35 40 45

Ala Lys Glu Leu Leu Asp Asp Asn His Arg Ala Phe Leu Asn Arg Val
50 55 60

Leu Pro Gln Ile Asp Met Asp Trp His Pro Ile Ala Glu Ala Phe Cys
65 70 75 80

Lys Val His Lys Asn Pro Gly Asn Lys Glu Leu Ala Gln Asp Tyr Asn
85 90 95

Leu Gln Leu Ser Lys Arg Arg Lys Glu Ile Ser Ala Tyr Leu Gln Asp
100 105 110

Ala Asp Gly Tyr Lys Gly Leu Phe Ala Lys Pro Ala Leu Asp Glu Ala
115 120 125

Met Lys Ile Ala Lys Glu Asn Gly Asn Glu Ser Asp Ile Glu Val Leu
130 135 140

Glu Ala Phe Asn Gly Phe Ser Val Tyr Phe Thr Gly Tyr His Glu Ser
145 150 155 160

Arg Glu Asn Ile Tyr Ser Asp Glu Asp Met Val Ser Val Ala Tyr Arg
165 170 175

Ile Thr Glu Asp Asn Phe Pro Arg Phe Val Ser Asn Ala Leu Ile Phe
180 185 190

Asp Lys Leu Asn Glu Ser His Pro Asp Ile Ile Ser Glu Val Ser Gly
195 200 205

Asn Leu Gly Val Asp Asp Ile Gly Lys Tyr Phe Asp Val Ser Asn Tyr
210 215 220

Asn Asn Phe Leu Ser Gln Ala Gly Ile Asp Asp Tyr Asn His Ile Ile
225 230 235 240

Gly Gly His Thr Thr Glu Asp Gly Leu Ile Gln Ala Phe Asn Val Val
245 250 255

Leu Asn Leu Arg His Gln Lys Asp Pro Gly Phe Glu Lys Ile Gln Phe
260 265 270

```
Lys Gln Leu Tyr Lys Gln Ile Leu Ser Val Arg Thr Ser Lys Ser Tyr
        275                 280                 285

Ile Pro Lys Gln Phe Asp Asn Ser Lys Glu Met Val Asp Cys Ile Cys
    290                 295                 300

Asp Tyr Val Ser Lys Ile Glu Lys Ser Glu Thr Val Glu Arg Ala Leu
305                 310                 315                 320

Lys Leu Val Arg Asn Ile Ser Ser Phe Asp Leu Arg Gly Ile Phe Val
                325                 330                 335

Asn Lys Lys Asn Leu Arg Ile Leu Ser Asn Lys Leu Ile Gly Asp Trp
            340                 345                 350

Asp Ala Ile Glu Thr Ala Leu Met His Ser Ser Ser Ser Glu Asn Asp
        355                 360                 365

Lys Lys Ser Val Tyr Asp Ser Ala Glu Ala Phe Thr Leu Asp Asp Ile
    370                 375                 380

Phe Ser Ser Val Lys Lys Phe Ser Asp Ala Ser Ala Glu Asp Ile Gly
385                 390                 395                 400

Asn Arg Ala Glu Asp Ile Cys Arg Val Ile Ser Glu Thr Ala Pro Phe
                405                 410                 415

Ile Asn Asp Leu Arg Ala Val Asp Leu Asp Ser Leu Asn Asp Asp Gly
            420                 425                 430

Tyr Glu Ala Ala Val Ser Lys Ile Arg Glu Ser Leu Glu Pro Tyr Met
        435                 440                 445

Asp Leu Phe His Glu Leu Glu Ile Phe Ser Val Gly Asp Glu Phe Pro
    450                 455                 460

Lys Cys Ala Ala Phe Tyr Ser Glu Leu Glu Glu Val Ser Glu Gln Leu
465                 470                 475                 480

Ile Glu Ile Ile Pro Leu Phe Asn Lys Ala Arg Ser Phe Cys Thr Arg
                485                 490                 495

Lys Arg Tyr Ser Thr Asp Lys Ile Lys Val Asn Leu Lys Phe Pro Thr
            500                 505                 510

Leu Ala Asp Gly Trp Asp Leu Asn Lys Glu Arg Asp Asn Lys Ala Ala
        515                 520                 525

Ile Leu Arg Lys Asp Gly Lys Tyr Tyr Leu Ala Ile Leu Asp Met Lys
    530                 535                 540

Lys Asp Leu Ser Ser Ile Arg Thr Ser Asp Glu Asp Glu Ser Ser Phe
545                 550                 555                 560

Glu Lys Met Glu Tyr Lys Leu Leu Pro Ser Pro Val Lys Met Leu Pro
                565                 570                 575

Lys Ile Phe Val Lys Ser Lys Ala Ala Lys Glu Lys Tyr Gly Leu Thr
            580                 585                 590

Asp Arg Met Leu Glu Cys Tyr Asp Lys Gly Met His Lys Ser Gly Ser
        595                 600                 605
```

Ala Phe Asp Leu Gly Phe Cys His Glu Leu Ile Asp Tyr Tyr Lys Arg
610 615 620

Cys Ile Ala Glu Tyr Pro Gly Trp Asp Val Phe Asp Phe Lys Phe Arg
625 630 635 640

Glu Thr Ser Asp Tyr Gly Ser Met Lys Glu Phe Asn Glu Asp Val Ala
645 650 655

Gly Ala Gly Tyr Tyr Met Ser Leu Arg Lys Ile Pro Cys Ser Glu Val
660 665 670

Tyr Arg Leu Leu Asp Glu Lys Ser Ile Tyr Leu Phe Gln Ile Tyr Asn
675 680 685

Lys Asp Tyr Ser Glu Asn Ala His Gly Asn Lys Asn Met His Thr Met
690 695 700

Tyr Trp Glu Gly Leu Phe Ser Pro Gln Asn Leu Glu Ser Pro Val Phe
705 710 715 720

Lys Leu Ser Gly Gly Ala Glu Leu Phe Phe Arg Lys Ser Ser Ile Pro
725 730 735

Asn Asp Ala Lys Thr Val His Pro Lys Gly Ser Val Leu Val Pro Arg
740 745 750

Asn Asp Val Asn Gly Arg Arg Ile Pro Asp Ser Ile Tyr Arg Glu Leu
755 760 765

Thr Arg Tyr Phe Asn Arg Gly Asp Cys Arg Ile Ser Asp Glu Ala Lys
770 775 780

Ser Tyr Leu Asp Lys Val Lys Thr Lys Lys Ala Asp His Asp Ile Val
785 790 795 800

Lys Asp Arg Arg Phe Thr Val Asp Lys Met Met Phe His Val Pro Ile
805 810 815

Ala Met Asn Phe Lys Ala Ile Ser Lys Pro Asn Leu Asn Lys Lys Val
820 825 830

Ile Asp Gly Ile Ile Asp Asp Gln Asp Leu Lys Ile Ile Gly Ile Asp
835 840 845

Arg Gly Glu Arg Asn Leu Ile Tyr Val Thr Met Val Asp Arg Lys Gly
850 855 860

Asn Ile Leu Tyr Gln Asp Ser Leu Asn Ile Leu Asn Gly Tyr Asp Tyr
865 870 875 880

Arg Lys Ala Leu Asp Val Arg Glu Tyr Asp Asn Lys Glu Ala Arg Arg
885 890 895

Asn Trp Thr Lys Val Glu Gly Ile Arg Lys Met Lys Glu Gly Tyr Leu
900 905 910

Ser Leu Ala Val Ser Lys Leu Ala Asp Met Ile Ile Glu Asn Asn Ala
915 920 925

Ile Ile Val Met Glu Asp Leu Asn His Gly Phe Lys Ala Gly Arg Ser
930 935 940

```
Lys Ile Glu Lys Gln Val Tyr Gln Lys Phe Glu Ser Met Leu Ile Asn
945                 950                 955                 960

Lys Leu Gly Tyr Met Val Leu Lys Asp Lys Ser Ile Asp Gln Ser Gly
                965                 970                 975

Gly Ala Leu His Gly Tyr Gln Leu Ala Asn His Val Thr Thr Leu Ala
            980                 985                 990

Ser Val Gly Lys Gln Cys Gly Val Ile Phe Tyr Ile Pro Ala Ala Phe
        995                1000                1005

Thr Ser  Lys Ile Asp Pro Thr  Thr Gly Phe Ala Asp  Leu Phe Ala
    1010                1015                1020

Leu Ser  Asn Val Lys Asn Val  Ala Ser Met Arg Glu  Phe Phe Ser
    1025                1030                1035

Lys Met  Lys Ser Val Ile Tyr  Asp Lys Ala Glu Gly  Lys Phe Ala
    1040                1045                1050

Phe Thr  Phe Asp Tyr Leu Asp  Tyr Asn Val Lys Ser  Glu Cys Gly
    1055                1060                1065

Arg Thr  Leu Trp Thr Val Tyr  Thr Val Gly Glu Arg  Phe Thr Tyr
    1070                1075                1080

Ser Arg  Val Asn Arg Glu Tyr  Val Arg Lys Val Pro  Thr Asp Ile
    1085                1090                1095

Ile Tyr  Asp Ala Leu Gln Lys  Ala Gly Ile Ser Val  Glu Gly Asp
    1100                1105                1110

Leu Arg  Asp Arg Ile Ala Glu  Ser Asp Gly Asp Thr  Leu Lys Ser
    1115                1120                1125

Ile Phe  Tyr Ala Phe Lys Tyr  Ala Leu Asp Met Arg  Val Glu Asn
    1130                1135                1140

Arg Glu  Glu Asp Tyr Ile Gln  Ser Pro Val Lys Asn  Ala Ser Gly
    1145                1150                1155

Glu Phe  Phe Cys Ser Lys Asn  Ala Gly Lys Ser Leu  Pro Gln Asp
    1160                1165                1170

Ser Asp  Ala Asn Gly Ala Tyr  Asn Ile Ala Leu Lys  Gly Ile Leu
    1175                1180                1185

Gln Leu  Arg Met Leu Ser Glu  Gln Tyr Asp Pro Asn  Ala Glu Ser
    1190                1195                1200

Ile Arg  Leu Pro Leu Ile Thr  Asn Lys Ala Trp Leu  Thr Phe Met
    1205                1210                1215

Gln Ser  Gly Met Lys Thr Trp  Lys Asn
    1220                1225
```

<210> 1162
<211> 1334
<212> PRT
<213> 溶糊精琥珀酸弧菌

<400> 1162

```
Met Ser Ser Leu Thr Lys Phe Thr Asn Lys Tyr Ser Lys Gln Leu Thr
```

```
1               5                   10                  15

Ile Lys Asn Glu Leu Ile Pro Val Gly Lys Thr Leu Glu Asn Ile Lys
            20                  25                  30

Glu Asn Gly Leu Ile Asp Gly Asp Glu Gln Leu Asn Glu Asn Tyr Gln
        35                  40                  45

Lys Ala Lys Ile Ile Val Asp Asp Phe Leu Arg Asp Phe Ile Asn Lys
    50                  55                  60

Ala Leu Asn Asn Thr Gln Ile Gly Asn Trp Arg Glu Leu Ala Asp Ala
65                  70                  75                  80

Leu Asn Lys Glu Asp Glu Asp Asn Ile Glu Lys Leu Gln Asp Lys Ile
                85                  90                  95

Arg Gly Ile Ile Val Ser Lys Phe Glu Thr Phe Asp Leu Phe Ser Ser
            100                 105                 110

Tyr Ser Ile Lys Lys Asp Glu Lys Ile Ile Asp Asp Asp Asn Asp Val
        115                 120                 125

Glu Glu Glu Glu Leu Asp Leu Gly Lys Lys Thr Ser Ser Phe Lys Tyr
    130                 135                 140

Ile Phe Lys Lys Asn Leu Phe Lys Leu Val Leu Pro Ser Tyr Leu Lys
145                 150                 155                 160

Thr Thr Asn Gln Asp Lys Leu Lys Ile Ile Ser Ser Phe Asp Asn Phe
                165                 170                 175

Ser Thr Tyr Phe Arg Gly Phe Phe Glu Asn Arg Lys Asn Ile Phe Thr
            180                 185                 190

Lys Lys Pro Ile Ser Thr Ser Ile Ala Tyr Arg Ile Val His Asp Asn
        195                 200                 205

Phe Pro Lys Phe Leu Asp Asn Ile Arg Cys Phe Asn Val Trp Gln Thr
    210                 215                 220

Glu Cys Pro Gln Leu Ile Val Lys Ala Asp Asn Tyr Leu Lys Ser Lys
225                 230                 235                 240

Asn Val Ile Ala Lys Asp Lys Ser Leu Ala Asn Tyr Phe Thr Val Gly
                245                 250                 255

Ala Tyr Asp Tyr Phe Leu Ser Gln Asn Gly Ile Asp Phe Tyr Asn Asn
            260                 265                 270

Ile Ile Gly Gly Leu Pro Ala Phe Ala Gly His Glu Lys Ile Gln Gly
        275                 280                 285

Leu Asn Glu Phe Ile Asn Gln Glu Cys Gln Lys Asp Ser Glu Leu Lys
    290                 295                 300

Ser Lys Leu Lys Asn Arg His Ala Phe Lys Met Ala Val Leu Phe Lys
305                 310                 315                 320

Gln Ile Leu Ser Asp Arg Glu Lys Ser Phe Val Ile Asp Glu Phe Glu
                325                 330                 335

Ser Asp Ala Gln Val Ile Asp Ala Val Lys Asn Phe Tyr Ala Glu Gln
```

340 345 350

Cys Lys Asp Asn Asn Val Ile Phe Asn Leu Leu Asn Leu Ile Lys Asn
355 360 365

Ile Ala Phe Leu Ser Asp Asp Glu Leu Asp Gly Ile Phe Ile Glu Gly
370 375 380

Lys Tyr Leu Ser Ser Val Ser Gln Lys Leu Tyr Ser Asp Trp Ser Lys
385 390 395 400

Leu Arg Asn Asp Ile Glu Asp Ser Ala Asn Ser Lys Gln Gly Asn Lys
405 410 415

Glu Leu Ala Lys Lys Ile Lys Thr Asn Lys Gly Asp Val Glu Lys Ala
420 425 430

Ile Ser Lys Tyr Glu Phe Ser Leu Ser Glu Leu Asn Ser Ile Val His
435 440 445

Asp Asn Thr Lys Phe Ser Asp Leu Leu Ser Cys Thr Leu His Lys Val
450 455 460

Ala Ser Glu Lys Leu Val Lys Val Asn Glu Gly Asp Trp Pro Lys His
465 470 475 480

Leu Lys Asn Asn Glu Glu Lys Gln Lys Ile Lys Glu Pro Leu Asp Ala
485 490 495

Leu Leu Glu Ile Tyr Asn Thr Leu Leu Ile Phe Asn Cys Lys Ser Phe
500 505 510

Asn Lys Asn Gly Asn Phe Tyr Val Asp Tyr Asp Arg Cys Ile Asn Glu
515 520 525

Leu Ser Ser Val Val Tyr Leu Tyr Asn Lys Thr Arg Asn Tyr Cys Thr
530 535 540

Lys Lys Pro Tyr Asn Thr Asp Lys Phe Lys Leu Asn Phe Asn Ser Pro
545 550 555 560

Gln Leu Gly Glu Gly Phe Ser Lys Ser Lys Glu Asn Asp Cys Leu Thr
565 570 575

Leu Leu Phe Lys Lys Asp Asp Asn Tyr Tyr Val Gly Ile Ile Arg Lys
580 585 590

Gly Ala Lys Ile Asn Phe Asp Asp Thr Gln Ala Ile Ala Asp Asn Thr
595 600 605

Asp Asn Cys Ile Phe Lys Met Asn Tyr Phe Leu Leu Lys Asp Ala Lys
610 615 620

Lys Phe Ile Pro Lys Cys Ser Ile Gln Leu Lys Glu Val Lys Ala His
625 630 635 640

Phe Lys Lys Ser Glu Asp Asp Tyr Ile Leu Ser Asp Lys Glu Lys Phe
645 650 655

Ala Ser Pro Leu Val Ile Lys Lys Ser Thr Phe Leu Leu Ala Thr Ala
660 665 670

His Val Lys Gly Lys Lys Gly Asn Ile Lys Lys Phe Gln Lys Glu Tyr

675 680 685

Ser Lys Glu Asn Pro Thr Glu Tyr Arg Asn Ser Leu Asn Glu Trp Ile
690 695 700

Ala Phe Cys Lys Glu Phe Leu Lys Thr Tyr Lys Ala Ala Thr Ile Phe
705 710 715 720

Asp Ile Thr Thr Leu Lys Lys Ala Glu Glu Tyr Ala Asp Ile Val Glu
725 730 735

Phe Tyr Lys Asp Val Asp Asn Leu Cys Tyr Lys Leu Glu Phe Cys Pro
740 745 750

Ile Lys Thr Ser Phe Ile Glu Asn Leu Ile Asp Asn Gly Asp Leu Tyr
755 760 765

Leu Phe Arg Ile Asn Asn Lys Asp Phe Ser Ser Lys Ser Thr Gly Thr
770 775 780

Lys Asn Leu His Thr Leu Tyr Leu Gln Ala Ile Phe Asp Glu Arg Asn
785 790 795 800

Leu Asn Asn Pro Thr Ile Met Leu Asn Gly Gly Ala Glu Leu Phe Tyr
805 810 815

Arg Lys Glu Ser Ile Glu Gln Lys Asn Arg Ile Thr His Lys Ala Gly
820 825 830

Ser Ile Leu Val Asn Lys Val Cys Lys Asp Gly Thr Ser Leu Asp Asp
835 840 845

Lys Ile Arg Asn Glu Ile Tyr Gln Tyr Glu Asn Lys Phe Ile Asp Thr
850 855 860

Leu Ser Asp Glu Ala Lys Lys Val Leu Pro Asn Val Ile Lys Lys Glu
865 870 875 880

Ala Thr His Asp Ile Thr Lys Asp Lys Arg Phe Thr Ser Asp Lys Phe
885 890 895

Phe Phe His Cys Pro Leu Thr Ile Asn Tyr Lys Glu Gly Asp Thr Lys
900 905 910

Gln Phe Asn Asn Glu Val Leu Ser Phe Leu Arg Gly Asn Pro Asp Ile
915 920 925

Asn Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu Ile Tyr Val Thr
930 935 940

Val Ile Asn Gln Lys Gly Glu Ile Leu Asp Ser Val Ser Phe Asn Thr
945 950 955 960

Val Thr Asn Lys Ser Ser Lys Ile Glu Gln Thr Val Asp Tyr Glu Glu
965 970 975

Lys Leu Ala Val Arg Glu Lys Glu Arg Ile Glu Ala Lys Arg Ser Trp
980 985 990

Asp Ser Ile Ser Lys Ile Ala Thr Leu Lys Glu Gly Tyr Leu Ser Ala
995 1000 1005

Ile Val His Glu Ile Cys Leu Leu Met Ile Lys His Asn Ala Ile

1010 1015 1020

Val Val Leu Glu Asn Leu Asn Ala Gly Phe Lys Arg Ile Arg Gly
1025 1030 1035

Gly Leu Ser Glu Lys Ser Val Tyr Gln Lys Phe Glu Lys Met Leu
1040 1045 1050

Ile Asn Lys Leu Asn Tyr Phe Val Ser Lys Lys Glu Ser Asp Trp
1055 1060 1065

Asn Lys Pro Ser Gly Leu Leu Asn Gly Leu Gln Leu Ser Asp Gln
1070 1075 1080

Phe Glu Ser Phe Glu Lys Leu Gly Ile Gln Ser Gly Phe Ile Phe
1085 1090 1095

Tyr Val Pro Ala Ala Tyr Thr Ser Lys Ile Asp Pro Thr Thr Gly
1100 1105 1110

Phe Ala Asn Val Leu Asn Leu Ser Lys Val Arg Asn Val Asp Ala
1115 1120 1125

Ile Lys Ser Phe Phe Ser Asn Phe Asn Glu Ile Ser Tyr Ser Lys
1130 1135 1140

Lys Glu Ala Leu Phe Lys Phe Ser Phe Asp Leu Asp Ser Leu Ser
1145 1150 1155

Lys Lys Gly Phe Ser Ser Phe Val Lys Phe Ser Lys Ser Lys Trp
1160 1165 1170

Asn Val Tyr Thr Phe Gly Glu Arg Ile Ile Lys Pro Lys Asn Lys
1175 1180 1185

Gln Gly Tyr Arg Glu Asp Lys Arg Ile Asn Leu Thr Phe Glu Met
1190 1195 1200

Lys Lys Leu Leu Asn Glu Tyr Lys Val Ser Phe Asp Leu Glu Asn
1205 1210 1215

Asn Leu Ile Pro Asn Leu Thr Ser Ala Asn Leu Lys Asp Thr Phe
1220 1225 1230

Trp Lys Glu Leu Phe Phe Ile Phe Lys Thr Thr Leu Gln Leu Arg
1235 1240 1245

Asn Ser Val Thr Asn Gly Lys Glu Asp Val Leu Ile Ser Pro Val
1250 1255 1260

Lys Asn Ala Lys Gly Glu Phe Phe Val Ser Gly Thr His Asn Lys
1265 1270 1275

Thr Leu Pro Gln Asp Cys Asp Ala Asn Gly Ala Tyr His Ile Ala
1280 1285 1290

Leu Lys Gly Leu Met Ile Leu Glu Arg Asn Asn Leu Val Arg Glu
1295 1300 1305

Glu Lys Asp Thr Lys Lys Ile Met Ala Ile Ser Asn Val Asp Trp
1310 1315 1320

Phe Glu Tyr Val Gln Lys Arg Arg Gly Val Leu

1325 1330

<210> 1163
<211> 1307
<212> PRT
<213> 胺基酸球菌屬某種

<400> 1163
Met Thr Gln Phe Glu Gly Phe Thr Asn Leu Tyr Gln Val Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Lys His Ile Gln
20 25 30

Glu Gln Gly Phe Ile Glu Glu Asp Lys Ala Arg Asn Asp His Tyr Lys
35 40 45

Glu Leu Lys Pro Ile Ile Asp Arg Ile Tyr Lys Thr Tyr Ala Asp Gln
50 55 60

Cys Leu Gln Leu Val Gln Leu Asp Trp Glu Asn Leu Ser Ala Ala Ile
65 70 75 80

Asp Ser Tyr Arg Lys Glu Lys Thr Glu Glu Thr Arg Asn Ala Leu Ile
85 90 95

Glu Glu Gln Ala Thr Tyr Arg Asn Ala Ile His Asp Tyr Phe Ile Gly
100 105 110

Arg Thr Asp Asn Leu Thr Asp Ala Ile Asn Lys Arg His Ala Glu Ile
115 120 125

Tyr Lys Gly Leu Phe Lys Ala Glu Leu Phe Asn Gly Lys Val Leu Lys
130 135 140

Gln Leu Gly Thr Val Thr Thr Thr Glu His Glu Asn Ala Leu Leu Arg
145 150 155 160

Ser Phe Asp Lys Phe Thr Thr Tyr Phe Ser Gly Phe Tyr Glu Asn Arg
165 170 175

Lys Asn Val Phe Ser Ala Glu Asp Ile Ser Thr Ala Ile Pro His Arg
180 185 190

Ile Val Gln Asp Asn Phe Pro Lys Phe Lys Glu Asn Cys His Ile Phe
195 200 205

Thr Arg Leu Ile Thr Ala Val Pro Ser Leu Arg Glu His Phe Glu Asn
210 215 220

Val Lys Lys Ala Ile Gly Ile Phe Val Ser Thr Ser Ile Glu Glu Val
225 230 235 240

Phe Ser Phe Pro Phe Tyr Asn Gln Leu Leu Thr Gln Thr Gln Ile Asp
245 250 255

Leu Tyr Asn Gln Leu Leu Gly Gly Ile Ser Arg Glu Ala Gly Thr Glu
260 265 270

Lys Ile Lys Gly Leu Asn Glu Val Leu Asn Leu Ala Ile Gln Lys Asn
275 280 285

Asp Glu Thr Ala His Ile Ile Ala Ser Leu Pro His Arg Phe Ile Pro
290 295 300

```
Leu Phe Lys Gln Ile Leu Ser Asp Arg Asn Thr Leu Ser Phe Ile Leu
305                 310                 315                 320

Glu Glu Phe Lys Ser Asp Glu Glu Val Ile Gln Ser Phe Cys Lys Tyr
                325                 330                 335

Lys Thr Leu Leu Arg Asn Glu Asn Val Leu Glu Thr Ala Glu Ala Leu
            340                 345                 350

Phe Asn Glu Leu Asn Ser Ile Asp Leu Thr His Ile Phe Ile Ser His
        355                 360                 365

Lys Lys Leu Glu Thr Ile Ser Ser Ala Leu Cys Asp His Trp Asp Thr
    370                 375                 380

Leu Arg Asn Ala Leu Tyr Glu Arg Arg Ile Ser Glu Leu Thr Gly Lys
385                 390                 395                 400

Ile Thr Lys Ser Ala Lys Glu Lys Val Gln Arg Ser Leu Lys His Glu
                405                 410                 415

Asp Ile Asn Leu Gln Glu Ile Ile Ser Ala Ala Gly Lys Glu Leu Ser
            420                 425                 430

Glu Ala Phe Lys Gln Lys Thr Ser Glu Ile Leu Ser His Ala His Ala
        435                 440                 445

Ala Leu Asp Gln Pro Leu Pro Thr Thr Leu Lys Lys Gln Glu Glu Lys
    450                 455                 460

Glu Ile Leu Lys Ser Gln Leu Asp Ser Leu Leu Gly Leu Tyr His Leu
465                 470                 475                 480

Leu Asp Trp Phe Ala Val Asp Glu Ser Asn Glu Val Asp Pro Glu Phe
                485                 490                 495

Ser Ala Arg Leu Thr Gly Ile Lys Leu Glu Met Glu Pro Ser Leu Ser
            500                 505                 510

Phe Tyr Asn Lys Ala Arg Asn Tyr Ala Thr Lys Lys Pro Tyr Ser Val
        515                 520                 525

Glu Lys Phe Lys Leu Asn Phe Gln Met Pro Thr Leu Ala Ser Gly Trp
    530                 535                 540

Asp Val Asn Lys Glu Lys Asn Asn Gly Ala Ile Leu Phe Val Lys Asn
545                 550                 555                 560

Gly Leu Tyr Tyr Leu Gly Ile Met Pro Lys Gln Lys Gly Arg Tyr Lys
                565                 570                 575

Ala Leu Ser Phe Glu Pro Thr Glu Lys Thr Ser Glu Gly Phe Asp Lys
            580                 585                 590

Met Tyr Tyr Asp Tyr Phe Pro Asp Ala Ala Lys Met Ile Pro Lys Cys
        595                 600                 605

Ser Thr Gln Leu Lys Ala Val Thr Ala His Phe Gln Thr His Thr Thr
    610                 615                 620

Pro Ile Leu Leu Ser Asn Asn Phe Ile Glu Pro Leu Glu Ile Thr Lys
625                 630                 635                 640
```

Glu Ile Tyr Asp Leu Asn Asn Pro Glu Lys Glu Pro Lys Lys Phe Gln
645 650 655

Thr Ala Tyr Ala Lys Lys Thr Gly Asp Gln Lys Gly Tyr Arg Glu Ala
660 665 670

Leu Cys Lys Trp Ile Asp Phe Thr Arg Asp Phe Leu Ser Lys Tyr Thr
675 680 685

Lys Thr Thr Ser Ile Asp Leu Ser Ser Leu Arg Pro Ser Ser Gln Tyr
690 695 700

Lys Asp Leu Gly Glu Tyr Tyr Ala Glu Leu Asn Pro Leu Leu Tyr His
705 710 715 720

Ile Ser Phe Gln Arg Ile Ala Glu Lys Glu Ile Met Asp Ala Val Glu
725 730 735

Thr Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ala Lys
740 745 750

Gly His His Gly Lys Pro Asn Leu His Thr Leu Tyr Trp Thr Gly Leu
755 760 765

Phe Ser Pro Glu Asn Leu Ala Lys Thr Ser Ile Lys Leu Asn Gly Gln
770 775 780

Ala Glu Leu Phe Tyr Arg Pro Lys Ser Arg Met Lys Arg Met Ala His
785 790 795 800

Arg Leu Gly Glu Lys Met Leu Asn Lys Lys Leu Lys Asp Gln Lys Thr
805 810 815

Pro Ile Pro Asp Thr Leu Tyr Gln Glu Leu Tyr Asp Tyr Val Asn His
820 825 830

Arg Leu Ser His Asp Leu Ser Asp Glu Ala Arg Ala Leu Leu Pro Asn
835 840 845

Val Ile Thr Lys Glu Val Ser His Glu Ile Ile Lys Asp Arg Arg Phe
850 855 860

Thr Ser Asp Lys Phe Phe Phe His Val Pro Ile Thr Leu Asn Tyr Gln
865 870 875 880

Ala Ala Asn Ser Pro Ser Lys Phe Asn Gln Arg Val Asn Ala Tyr Leu
885 890 895

Lys Glu His Pro Glu Thr Pro Ile Ile Gly Ile Asp Arg Gly Glu Arg
900 905 910

Asn Leu Ile Tyr Ile Thr Val Ile Asp Ser Thr Gly Lys Ile Leu Glu
915 920 925

Gln Arg Ser Leu Asn Thr Ile Gln Gln Phe Asp Tyr Gln Lys Lys Leu
930 935 940

Asp Asn Arg Glu Lys Glu Arg Val Ala Ala Arg Gln Ala Trp Ser Val
945 950 955 960

Val Gly Thr Ile Lys Asp Leu Lys Gln Gly Tyr Leu Ser Gln Val Ile
965 970 975

```
His Glu Ile Val Asp Leu Met Ile His Tyr Gln Ala Val Val Val Leu
           980                 985                 990

Glu Asn Leu Asn Phe Gly Phe Lys Ser Lys Arg Thr Gly Ile Ala Glu
       995                1000                1005

Lys Ala  Val Tyr Gln Gln Phe  Glu Lys Met Leu Ile  Asp Lys Leu
    1010                1015                1020

Asn Cys  Leu Val Leu Lys Asp  Tyr Pro Ala Glu Lys  Val Gly Gly
    1025                1030                1035

Val Leu  Asn Pro Tyr Gln Leu  Thr Asp Gln Phe Thr  Ser Phe Ala
    1040                1045                1050

Lys Met  Gly Thr Gln Ser Gly  Phe Leu Phe Tyr Val  Pro Ala Pro
    1055                1060                1065

Tyr Thr  Ser Lys Ile Asp Pro  Leu Thr Gly Phe Val  Asp Pro Phe
    1070                1075                1080

Val Trp  Lys Thr Ile Lys Asn  His Glu Ser Arg Lys  His Phe Leu
    1085                1090                1095

Glu Gly  Phe Asp Phe Leu His  Tyr Asp Val Lys Thr  Gly Asp Phe
    1100                1105                1110

Ile Leu  His Phe Lys Met Asn  Arg Asn Leu Ser Phe  Gln Arg Gly
    1115                1120                1125

Leu Pro  Gly Phe Met Pro Ala  Trp Asp Ile Val Phe  Glu Lys Asn
    1130                1135                1140

Glu Thr  Gln Phe Asp Ala Lys  Gly Thr Pro Phe Ile  Ala Gly Lys
    1145                1150                1155

Arg Ile  Val Pro Val Ile Glu  Asn His Arg Phe Thr  Gly Arg Tyr
    1160                1165                1170

Arg Asp  Leu Tyr Pro Ala Asn  Glu Leu Ile Ala Leu  Leu Glu Glu
    1175                1180                1185

Lys Gly  Ile Val Phe Arg Asp  Gly Ser Asn Ile Leu  Pro Lys Leu
    1190                1195                1200

Leu Glu  Asn Asp Asp Ser His  Ala Ile Asp Thr Met  Val Ala Leu
    1205                1210                1215

Ile Arg  Ser Val Leu Gln Met  Arg Asn Ser Asn Ala  Ala Thr Gly
    1220                1225                1230

Glu Asp  Tyr Ile Asn Ser Pro  Val Arg Asp Leu Asn  Gly Val Cys
    1235                1240                1245

Phe Asp  Ser Arg Phe Gln Asn  Pro Glu Trp Pro Met  Asp Ala Asp
    1250                1255                1260

Ala Asn  Gly Ala Tyr His Ile  Ala Leu Lys Gly Gln  Leu Leu Leu
    1265                1270                1275

Asn His  Leu Lys Glu Ser Lys  Asp Leu Lys Leu Gln  Asn Gly Ile
    1280                1285                1290
```

Ser Asn Gln Asp Trp Leu Ala Tyr Ile Gln Glu Leu Arg Asn
1295 1300 1305

<210> 1164
<211> 1310
<212> PRT
<213> 孔茲氏創傷球菌

<400> 1164
Met Phe Glu Lys Leu Ser Asn Ile Val Ser Ile Ser Lys Thr Ile Arg
1 5 10 15

Phe Lys Leu Ile Pro Val Gly Lys Thr Leu Glu Asn Ile Glu Lys Leu
20 25 30

Gly Lys Leu Glu Lys Asp Phe Glu Arg Ser Asp Phe Tyr Pro Ile Leu
35 40 45

Lys Asn Ile Ser Asp Asp Tyr Tyr Arg Gln Tyr Ile Lys Glu Lys Leu
50 55 60

Ser Asp Leu Asn Leu Asp Trp Gln Lys Leu Tyr Asp Ala His Glu Leu
65 70 75 80

Leu Asp Ser Ser Lys Lys Glu Ser Gln Lys Asn Leu Glu Met Ile Gln
85 90 95

Ala Gln Tyr Arg Lys Val Leu Phe Asn Ile Leu Ser Gly Glu Leu Asp
100 105 110

Lys Ser Gly Glu Lys Asn Ser Lys Asp Leu Ile Lys Asn Asn Lys Ala
115 120 125

Leu Tyr Gly Lys Leu Phe Lys Lys Gln Phe Ile Leu Glu Val Leu Pro
130 135 140

Asp Phe Val Asn Asn Asn Asp Ser Tyr Ser Glu Glu Asp Leu Glu Gly
145 150 155 160

Leu Asn Leu Tyr Ser Lys Phe Thr Thr Arg Leu Lys Asn Phe Trp Glu
165 170 175

Thr Arg Lys Asn Val Phe Thr Asp Lys Asp Ile Val Thr Ala Ile Pro
180 185 190

Phe Arg Ala Val Asn Glu Asn Phe Gly Phe Tyr Tyr Asp Asn Ile Lys
195 200 205

Ile Phe Asn Lys Asn Ile Glu Tyr Leu Glu Asn Lys Ile Pro Asn Leu
210 215 220

Glu Asn Glu Leu Lys Glu Ala Asp Ile Leu Asp Asp Asn Arg Ser Val
225 230 235 240

Lys Asp Tyr Phe Thr Pro Asn Gly Phe Asn Tyr Val Ile Thr Gln Asp
245 250 255

Gly Ile Asp Val Tyr Gln Ala Ile Arg Gly Gly Phe Thr Lys Glu Asn
260 265 270

Gly Glu Lys Val Gln Gly Ile Asn Glu Ile Leu Asn Leu Thr Gln Gln
275 280 285

Gln Leu Arg Arg Lys Pro Glu Thr Lys Asn Val Lys Leu Gly Val Leu

290 295 300

Thr Lys Leu Arg Lys Gln Ile Leu Glu Tyr Ser Glu Ser Thr Ser Phe
305 310 315 320

Leu Ile Asp Gln Ile Glu Asp Asp Asn Asp Leu Val Asp Arg Ile Asn
325 330 335

Lys Phe Asn Val Ser Phe Phe Glu Ser Thr Glu Val Ser Pro Ser Leu
340 345 350

Phe Glu Gln Ile Glu Arg Leu Tyr Asn Ala Leu Lys Ser Ile Lys Lys
355 360 365

Glu Glu Val Tyr Ile Asp Ala Arg Asn Thr Gln Lys Phe Ser Gln Met
370 375 380

Leu Phe Gly Gln Trp Asp Val Ile Arg Arg Gly Tyr Thr Val Lys Ile
385 390 395 400

Thr Glu Gly Ser Lys Glu Glu Lys Lys Lys Tyr Lys Glu Tyr Leu Glu
405 410 415

Leu Asp Glu Thr Ser Lys Ala Lys Arg Tyr Leu Asn Ile Arg Glu Ile
420 425 430

Glu Glu Leu Val Asn Leu Val Glu Gly Phe Glu Glu Val Asp Val Phe
435 440 445

Ser Val Leu Leu Glu Lys Phe Lys Met Asn Asn Ile Glu Arg Ser Glu
450 455 460

Phe Glu Ala Pro Ile Tyr Gly Ser Pro Ile Lys Leu Glu Ala Ile Lys
465 470 475 480

Glu Tyr Leu Glu Lys His Leu Glu Glu Tyr His Lys Trp Lys Leu Leu
485 490 495

Leu Ile Gly Asn Asp Asp Leu Asp Thr Asp Glu Thr Phe Tyr Pro Leu
500 505 510

Leu Asn Glu Val Ile Ser Asp Tyr Tyr Ile Ile Pro Leu Tyr Asn Leu
515 520 525

Thr Arg Asn Tyr Leu Thr Arg Lys His Ser Asp Lys Asp Lys Ile Lys
530 535 540

Val Asn Phe Asp Phe Pro Thr Leu Ala Asp Gly Trp Ser Glu Ser Lys
545 550 555 560

Ile Ser Asp Asn Arg Ser Ile Ile Leu Arg Lys Gly Gly Tyr Tyr Tyr
565 570 575

Leu Gly Ile Leu Ile Asp Asn Lys Leu Leu Ile Asn Lys Lys Asn Lys
580 585 590

Ser Lys Lys Ile Tyr Glu Ile Leu Ile Tyr Asn Gln Ile Pro Glu Phe
595 600 605

Ser Lys Ser Ile Pro Asn Tyr Pro Phe Thr Lys Lys Val Lys Glu His
610 615 620

Phe Lys Asn Asn Val Ser Asp Phe Gln Leu Ile Asp Gly Tyr Val Ser

625 630 635 640

Pro Leu Ile Ile Thr Lys Glu Ile Tyr Asp Ile Lys Lys Glu Lys Lys
645 650 655

Tyr Lys Lys Asp Phe Tyr Lys Asp Asn Asn Thr Asn Lys Asn Tyr Leu
660 665 670

Tyr Thr Ile Tyr Lys Trp Ile Glu Phe Cys Lys Gln Phe Leu Tyr Lys
675 680 685

Tyr Lys Gly Pro Asn Lys Glu Ser Tyr Lys Glu Met Tyr Asp Phe Ser
690 695 700

Thr Leu Lys Asp Thr Ser Leu Tyr Val Asn Leu Asn Asp Phe Tyr Ala
705 710 715 720

Asp Val Asn Ser Cys Ala Tyr Arg Val Leu Phe Asn Lys Ile Asp Glu
725 730 735

Asn Thr Ile Asp Asn Ala Val Glu Asp Gly Lys Leu Leu Leu Phe Gln
740 745 750

Ile Tyr Asn Lys Asp Phe Ser Pro Glu Ser Lys Gly Lys Lys Asn Leu
755 760 765

His Thr Leu Tyr Trp Leu Ser Met Phe Ser Glu Glu Asn Leu Arg Thr
770 775 780

Arg Lys Leu Lys Leu Asn Gly Gln Ala Glu Ile Phe Tyr Arg Lys Lys
785 790 795 800

Leu Glu Lys Lys Pro Ile Ile His Lys Glu Gly Ser Ile Leu Leu Asn
805 810 815

Lys Ile Asp Lys Glu Gly Asn Thr Ile Pro Glu Asn Ile Tyr His Glu
820 825 830

Cys Tyr Arg Tyr Leu Asn Lys Lys Ile Gly Arg Glu Asp Leu Ser Asp
835 840 845

Glu Ala Ile Ala Leu Phe Asn Lys Asp Val Leu Lys Tyr Lys Glu Ala
850 855 860

Arg Phe Asp Ile Ile Lys Asp Arg Arg Tyr Ser Glu Ser Gln Phe Phe
865 870 875 880

Phe His Val Pro Ile Thr Phe Asn Trp Asp Ile Lys Thr Asn Lys Asn
885 890 895

Val Asn Gln Ile Val Gln Gly Met Ile Lys Asp Gly Glu Ile Lys His
900 905 910

Ile Ile Gly Ile Asp Arg Gly Glu Arg His Leu Leu Tyr Tyr Ser Val
915 920 925

Ile Asp Leu Glu Gly Asn Ile Val Glu Gln Gly Ser Leu Asn Thr Leu
930 935 940

Glu Gln Asn Arg Phe Asp Asn Ser Thr Val Lys Val Asp Tyr Gln Asn
945 950 955 960

Lys Leu Arg Thr Arg Glu Glu Asp Arg Asp Arg Ala Arg Lys Asn Trp

965 970 975

Thr Asn Ile Asn Lys Ile Lys Glu Leu Lys Asp Gly Tyr Leu Ser His
980 985 990

Val Val His Lys Leu Ser Arg Leu Ile Ile Lys Tyr Glu Ala Ile Val
995 1000 1005

Ile Met Glu Asn Leu Asn Gln Gly Phe Lys Arg Gly Arg Phe Lys
1010 1015 1020

Val Glu Arg Gln Val Tyr Gln Lys Phe Glu Leu Ala Leu Met Asn
1025 1030 1035

Lys Leu Ser Ala Leu Ser Phe Lys Glu Lys Tyr Asp Glu Arg Lys
1040 1045 1050

Asn Leu Glu Pro Ser Gly Ile Leu Asn Pro Ile Gln Ala Cys Tyr
1055 1060 1065

Pro Val Asp Ala Tyr Gln Glu Leu Gln Gly Gln Asn Gly Ile Val
1070 1075 1080

Phe Tyr Leu Pro Ala Ala Tyr Thr Ser Val Ile Asp Pro Val Thr
1085 1090 1095

Gly Phe Thr Asn Leu Phe Arg Leu Lys Ser Ile Asn Ser Ser Lys
1100 1105 1110

Tyr Glu Glu Phe Ile Lys Lys Phe Lys Asn Ile Tyr Phe Asp Asn
1115 1120 1125

Glu Glu Glu Asp Phe Lys Phe Ile Phe Asn Tyr Lys Asp Phe Ala
1130 1135 1140

Lys Ala Asn Leu Val Ile Leu Asn Asn Ile Lys Ser Lys Asp Trp
1145 1150 1155

Lys Ile Ser Thr Arg Gly Glu Arg Ile Ser Tyr Asn Ser Lys Lys
1160 1165 1170

Lys Glu Tyr Phe Tyr Val Gln Pro Thr Glu Phe Leu Ile Asn Lys
1175 1180 1185

Leu Lys Glu Leu Asn Ile Asp Tyr Glu Asn Ile Asp Ile Ile Pro
1190 1195 1200

Leu Ile Asp Asn Leu Glu Glu Lys Ala Lys Arg Lys Ile Leu Lys
1205 1210 1215

Ala Leu Phe Asp Thr Phe Lys Tyr Ser Val Gln Leu Arg Asn Tyr
1220 1225 1230

Asp Phe Glu Asn Asp Tyr Ile Ile Ser Pro Thr Ala Asp Asp Asn
1235 1240 1245

Gly Asn Tyr Tyr Asn Ser Asn Glu Ile Asp Ile Asp Lys Thr Asn
1250 1255 1260

Leu Pro Asn Asn Gly Asp Ala Asn Gly Ala Phe Asn Ile Ala Arg
1265 1270 1275

Lys Gly Leu Leu Leu Lys Asp Arg Ile Val Asn Ser Asn Glu Ser

1280 1285 1290

Lys Val Asp Leu Lys Ile Lys Asn Glu Asp Trp Ile Asn Phe Ile
1295 1300 1305

Ile Ser
1310

<210> 1165
<211> 1255
<212> PRT
<213> 解脲丁酸弧菌

<400> 1165
Met Leu Leu Tyr Glu Asn Tyr Thr Lys Arg Asn Gln Ile Thr Lys Ser
1 5 10 15

Leu Arg Leu Glu Leu Arg Pro Gln Gly Lys Thr Leu Arg Asn Ile Lys
20 25 30

Glu Leu Asn Leu Leu Glu Gln Asp Lys Ala Ile Tyr Ala Leu Leu Glu
35 40 45

Arg Leu Lys Pro Val Ile Asp Glu Gly Ile Lys Asp Ile Ala Arg Asp
50 55 60

Thr Leu Lys Asn Cys Glu Leu Ser Phe Glu Lys Leu Tyr Glu His Phe
65 70 75 80

Leu Ser Gly Asp Lys Lys Ala Tyr Ala Lys Glu Ser Glu Arg Leu Lys
85 90 95

Lys Glu Ile Val Lys Thr Leu Ile Lys Asn Leu Pro Glu Gly Ile Gly
100 105 110

Lys Ile Ser Glu Ile Asn Ser Ala Lys Tyr Leu Asn Gly Val Leu Tyr
115 120 125

Asp Phe Ile Asp Lys Thr His Lys Asp Ser Glu Glu Lys Gln Asn Ile
130 135 140

Leu Ser Asp Ile Leu Glu Thr Lys Gly Tyr Leu Ala Leu Phe Ser Lys
145 150 155 160

Phe Leu Thr Ser Arg Ile Thr Thr Leu Glu Gln Ser Met Pro Lys Arg
165 170 175

Val Ile Glu Asn Phe Glu Ile Tyr Ala Ala Asn Ile Pro Lys Met Gln
180 185 190

Asp Ala Leu Glu Arg Gly Ala Val Ser Phe Ala Ile Glu Tyr Glu Ser
195 200 205

Ile Cys Ser Val Asp Tyr Tyr Asn Gln Ile Leu Ser Gln Glu Asp Ile
210 215 220

Asp Ser Tyr Asn Arg Leu Ile Ser Gly Ile Met Asp Glu Asp Gly Ala
225 230 235 240

Lys Glu Lys Gly Ile Asn Gln Thr Ile Ser Glu Leu Met Gln Arg Phe
245 250 255

Leu Thr Thr Arg Ile Thr Ala Leu Thr Lys Asn Ile Lys Ile Lys Ser
260 265 270

```
Glu His Leu Glu Glu Lys Pro Phe Arg Ile Leu Lys Gln Leu His Lys
        275                 280                 285

Gln Ile Leu Glu Glu Arg Glu Lys Ala Phe Thr Ile Asp His Ile Asp
    290                 295                 300

Ser Asp Glu Glu Val Val Gln Val Thr Lys Glu Ala Phe Glu Gln Thr
305                 310                 315                 320

Lys Glu Gln Trp Glu Asn Ile Lys Lys Ile Asn Gly Phe Tyr Ala Lys
                325                 330                 335

Asp Pro Gly Asp Ile Thr Leu Phe Ile Val Val Gly Pro Asn Gln Thr
            340                 345                 350

His Val Leu Ser Gln Leu Ile Tyr Gly Glu His Asp Arg Ile Arg Leu
        355                 360                 365

Leu Leu Glu Glu Tyr Glu Lys Asn Thr Leu Glu Val Leu Pro Arg Arg
    370                 375                 380

Thr Lys Ser Glu Lys Ala Arg Tyr Asp Lys Phe Val Asn Ala Val Pro
385                 390                 395                 400

Lys Lys Val Ala Lys Glu Ser His Thr Phe Asp Gly Leu Gln Lys Met
                405                 410                 415

Thr Gly Asp Asp Arg Leu Phe Ile Leu Tyr Arg Asp Glu Leu Ala Arg
            420                 425                 430

Asn Tyr Met Arg Ile Lys Glu Ala Tyr Gly Thr Phe Glu Arg Asp Ile
        435                 440                 445

Leu Lys Ser Arg Arg Gly Ile Lys Gly Asn Arg Asp Val Gln Glu Ser
    450                 455                 460

Leu Val Ser Phe Tyr Asp Glu Leu Thr Lys Phe Arg Ser Ala Leu Arg
465                 470                 475                 480

Ile Ile Asn Ser Gly Asn Asp Glu Lys Ala Asp Pro Ile Phe Tyr Asn
                485                 490                 495

Thr Phe Asp Gly Ile Phe Glu Lys Ala Asn Arg Thr Tyr Lys Ala Glu
            500                 505                 510

Asn Leu Cys Arg Asn Tyr Val Thr Lys Ser Pro Ala Asp Asp Ala Arg
        515                 520                 525

Ile Met Ala Ser Cys Leu Gly Thr Pro Ala Arg Leu Arg Thr His Trp
    530                 535                 540

Trp Asn Gly Glu Glu Asn Phe Ala Ile Asn Asp Val Ala Met Ile Arg
545                 550                 555                 560

Arg Gly Asp Glu Tyr Tyr Tyr Phe Val Leu Thr Pro Asp Val Lys Pro
                565                 570                 575

Val Asp Leu Lys Thr Lys Asp Glu Thr Asp Ala Gln Ile Phe Val Gln
            580                 585                 590

Arg Lys Gly Ala Lys Ser Phe Leu Gly Leu Pro Lys Ala Leu Phe Lys
        595                 600                 605
```

Cys Ile Leu Glu Pro Tyr Phe Glu Ser Pro Glu His Lys Asn Asp Lys
610 615 620

Asn Cys Val Ile Glu Glu Tyr Val Ser Lys Pro Leu Thr Ile Asp Arg
625 630 635 640

Arg Ala Tyr Asp Ile Phe Lys Asn Gly Thr Phe Lys Lys Thr Asn Ile
645 650 655

Gly Ile Asp Gly Leu Thr Glu Glu Lys Phe Lys Asp Asp Cys Arg Tyr
660 665 670

Leu Ile Asp Val Tyr Lys Glu Phe Ile Ala Val Tyr Thr Arg Tyr Ser
675 680 685

Cys Phe Asn Met Ser Gly Leu Lys Arg Ala Asp Glu Tyr Asn Asp Ile
690 695 700

Gly Glu Phe Phe Ser Asp Val Asp Thr Arg Leu Cys Thr Met Glu Trp
705 710 715 720

Ile Pro Val Ser Phe Glu Arg Ile Asn Asp Met Val Asp Lys Lys Glu
725 730 735

Gly Leu Leu Phe Leu Val Arg Ser Met Phe Leu Tyr Asn Arg Pro Arg
740 745 750

Lys Pro Tyr Glu Arg Thr Phe Ile Gln Leu Phe Ser Asp Ser Asn Met
755 760 765

Glu His Thr Ser Met Leu Leu Asn Ser Arg Ala Met Ile Gln Tyr Arg
770 775 780

Ala Ala Ser Leu Pro Arg Arg Val Thr His Lys Lys Gly Ser Ile Leu
785 790 795 800

Val Ala Leu Arg Asp Ser Asn Gly Glu His Ile Pro Met His Ile Arg
805 810 815

Glu Ala Ile Tyr Lys Met Lys Asn Asn Phe Asp Ile Ser Ser Glu Asp
820 825 830

Phe Ile Met Ala Lys Ala Tyr Leu Ala Glu His Asp Val Ala Ile Lys
835 840 845

Lys Ala Asn Glu Asp Ile Ile Arg Asn Arg Arg Tyr Thr Glu Asp Lys
850 855 860

Phe Phe Leu Ser Leu Ser Tyr Thr Lys Asn Ala Asp Ile Ser Ala Arg
865 870 875 880

Thr Leu Asp Tyr Ile Asn Asp Lys Val Glu Glu Asp Thr Gln Asp Ser
885 890 895

Arg Met Ala Val Ile Val Thr Arg Asn Leu Lys Asp Leu Thr Tyr Val
900 905 910

Ala Val Val Asp Glu Lys Asn Asn Val Leu Glu Glu Lys Ser Leu Asn
915 920 925

Glu Ile Asp Gly Val Asn Tyr Arg Glu Leu Leu Lys Glu Arg Thr Lys
930 935 940

```
Ile Lys Tyr His Asp Lys Thr Arg Leu Trp Gln Tyr Asp Val Ser Ser
945                 950                 955                 960

Lys Gly Leu Lys Glu Ala Tyr Val Glu Leu Ala Val Thr Gln Ile Ser
                965                 970                 975

Lys Leu Ala Thr Lys Tyr Asn Ala Val Val Val Val Glu Ser Met Ser
            980                 985                 990

Ser Thr Phe Lys Asp Lys Phe Ser Phe Leu Asp Glu Gln Ile Phe Lys
        995                1000                1005

Ala Phe  Glu Ala Arg Leu Cys  Ala Arg Met Ser Asp  Leu Ser Phe
    1010                1015                1020

Asn Thr  Ile Lys Glu Gly Glu  Ala Gly Ser Ile Ser  Asn Pro Ile
    1025                1030                1035

Gln Val  Ser Asn Asn Asn Gly  Asn Ser Tyr Gln Asp  Gly Val Ile
    1040                1045                1050

Tyr Phe  Leu Asn Asn Ala Tyr  Thr Arg Thr Leu Cys  Pro Asp Thr
    1055                1060                1065

Gly Phe  Val Asp Val Phe Asp  Lys Thr Arg Leu Ile  Thr Met Gln
    1070                1075                1080

Ser Lys  Arg Gln Phe Phe Ala  Lys Met Lys Asp Ile  Arg Ile Asp
    1085                1090                1095

Asp Gly  Glu Met Leu Phe Thr  Phe Asn Leu Glu Glu  Tyr Pro Thr
    1100                1105                1110

Lys Arg  Leu Leu Asp Arg Lys  Glu Trp Thr Val Lys  Ile Ala Gly
    1115                1120                1125

Asp Gly  Ser Tyr Phe Asp Lys  Asp Lys Gly Glu Tyr  Val Tyr Val
    1130                1135                1140

Asn Asp  Ile Val Arg Glu Gln  Ile Ile Pro Ala Leu  Leu Glu Asp
    1145                1150                1155

Lys Ala  Val Phe Asp Gly Asn  Met Ala Glu Lys Phe  Leu Asp Lys
    1160                1165                1170

Thr Ala  Ile Ser Gly Lys Ser  Val Glu Leu Ile Tyr  Lys Trp Phe
    1175                1180                1185

Ala Asn  Ala Leu Tyr Gly Ile  Ile Thr Lys Lys Asp  Gly Glu Lys
    1190                1195                1200

Ile Tyr  Arg Ser Pro Ile Thr  Gly Thr Glu Ile Asp  Val Ser Lys
    1205                1210                1215

Asn Thr  Thr Tyr Asn Phe Gly  Lys Lys Phe Met Phe  Lys Gln Glu
    1220                1225                1230

Tyr Arg  Gly Asp Gly Asp Phe  Leu Asp Ala Phe Leu  Asn Tyr Met
    1235                1240                1245

Gln Ala  Gln Asp Ile Ala Val
    1250                1255
```

<210> 1166
<211> 1219
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 1166

```
Met Asp Tyr Gly Asn Gly Gln Phe Glu Arg Arg Ala Pro Leu Thr Lys
1               5                  10                  15

Thr Ile Thr Leu Arg Leu Lys Pro Ile Gly Glu Thr Arg Glu Thr Ile
            20                  25                  30

Arg Glu Gln Lys Leu Leu Glu Gln Asp Ala Ala Phe Arg Lys Leu Val
        35                  40                  45

Glu Thr Val Thr Pro Ile Val Asp Asp Cys Ile Arg Lys Ile Ala Asp
    50                  55                  60

Asn Ala Leu Cys His Phe Gly Thr Glu Tyr Asp Phe Ser Cys Leu Gly
65                  70                  75                  80

Asn Ala Ile Ser Lys Asn Asp Ser Lys Ala Ile Lys Lys Glu Thr Glu
                85                  90                  95

Lys Val Glu Lys Leu Leu Ala Lys Val Leu Thr Glu Asn Leu Pro Asp
            100                 105                 110

Gly Leu Arg Lys Val Asn Asp Ile Asn Ser Ala Ala Phe Ile Gln Asp
        115                 120                 125

Thr Leu Thr Ser Phe Val Gln Asp Asp Ala Asp Lys Arg Val Leu Ile
    130                 135                 140

Gln Glu Leu Lys Gly Lys Thr Val Val Trp Leu Pro Asp Arg Val Phe
145                 150                 155                 160

Glu Asn Phe Asn Ile Phe Ile Glu Asn Ala Glu Lys Met Arg Ile Leu
                165                 170                 175

Leu Asp Ser Pro Leu Asn Glu Lys Ile Met Lys Phe Asp Pro Asp Ala
            180                 185                 190

Glu Gln Tyr Ala Ser Leu Glu Phe Tyr Gly Gln Cys Leu Ser Gln Lys
        195                 200                 205

Asp Ile Asp Ser Tyr Asn Leu Ile Ile Ser Gly Ile Tyr Ala Asp Asp
    210                 215                 220

Glu Val Lys Asn Pro Gly Ile Asn Glu Ile Val Lys Glu Tyr Asn Gln
225                 230                 235                 240

Gln Ile Arg Gly Asp Lys Asp Glu Ser Pro Leu Pro Lys Leu Lys Lys
                245                 250                 255

Leu His Lys Gln Ile Leu Met Pro Val Glu Lys Ala Phe Phe Val Arg
            260                 265                 270

Val Leu Ser Asn Asp Ser Asp Ala Arg Ser Ile Leu Glu Lys Ile Leu
        275                 280                 285

Lys Asp Thr Glu Met Leu Pro Ser Lys Ile Ile Glu Ala Met Lys Glu
```

290 295 300

Ala Asp Ala Gly Asp Ile Ala Val Tyr Gly Ser Arg Leu His Glu Leu
305 310 315 320

Ser His Val Ile Tyr Gly Asp His Gly Lys Leu Ser Gln Ile Ile Tyr
325 330 335

Asp Lys Glu Ser Lys Arg Ile Ser Glu Leu Met Glu Thr Leu Ser Pro
340 345 350

Lys Glu Arg Lys Glu Ser Lys Lys Arg Leu Glu Gly Leu Glu Glu His
355 360 365

Ile Arg Lys Ser Thr Tyr Thr Phe Asp Glu Leu Asn Arg Tyr Ala Glu
370 375 380

Lys Asn Val Met Ala Ala Tyr Ile Ala Ala Val Glu Glu Ser Cys Ala
385 390 395 400

Glu Ile Met Arg Lys Glu Lys Asp Leu Arg Thr Leu Leu Ser Lys Glu
405 410 415

Asp Val Lys Ile Arg Gly Asn Arg His Asn Thr Leu Ile Val Lys Asn
420 425 430

Tyr Phe Asn Ala Trp Thr Val Phe Arg Asn Leu Ile Arg Ile Leu Arg
435 440 445

Arg Lys Ser Glu Ala Glu Ile Asp Ser Asp Phe Tyr Asp Val Leu Asp
450 455 460

Asp Ser Val Glu Val Leu Ser Leu Thr Tyr Lys Gly Glu Asn Leu Cys
465 470 475 480

Arg Ser Tyr Ile Thr Lys Lys Ile Gly Ser Asp Leu Lys Pro Glu Ile
485 490 495

Ala Thr Tyr Gly Ser Ala Leu Arg Pro Asn Ser Arg Trp Trp Ser Pro
500 505 510

Gly Glu Lys Phe Asn Val Lys Phe His Thr Ile Val Arg Arg Asp Gly
515 520 525

Arg Leu Tyr Tyr Phe Ile Leu Pro Lys Gly Ala Lys Pro Val Glu Leu
530 535 540

Glu Asp Met Asp Gly Asp Ile Glu Cys Leu Gln Met Arg Lys Ile Pro
545 550 555 560

Asn Pro Thr Ile Phe Leu Pro Lys Leu Val Phe Lys Asp Pro Glu Ala
565 570 575

Phe Phe Arg Asp Asn Pro Glu Ala Asp Glu Phe Val Phe Leu Ser Gly
580 585 590

Met Lys Ala Pro Val Thr Ile Thr Arg Glu Thr Tyr Glu Ala Tyr Arg
595 600 605

Tyr Lys Leu Tyr Thr Val Gly Lys Leu Arg Asp Gly Glu Val Ser Glu
610 615 620

Glu Glu Tyr Lys Arg Ala Leu Leu Gln Val Leu Thr Ala Tyr Lys Glu

625 630 635 640

Phe Leu Glu Asn Arg Met Ile Tyr Ala Asp Leu Asn Phe Gly Phe Lys
645 650 655

Asp Leu Glu Glu Tyr Lys Asp Ser Ser Glu Phe Ile Lys Gln Val Glu
660 665 670

Thr His Asn Thr Phe Met Cys Trp Ala Lys Val Ser Ser Ser Gln Leu
675 680 685

Asp Asp Leu Val Lys Ser Gly Asn Gly Leu Leu Phe Glu Ile Trp Ser
690 695 700

Glu Arg Leu Glu Ser Tyr Tyr Lys Tyr Gly Asn Glu Lys Val Leu Arg
705 710 715 720

Gly Tyr Glu Gly Val Leu Leu Ser Ile Leu Lys Asp Glu Asn Leu Val
725 730 735

Ser Met Arg Thr Leu Leu Asn Ser Arg Pro Met Leu Val Tyr Arg Pro
740 745 750

Lys Glu Ser Ser Lys Pro Met Val Val His Arg Asp Gly Ser Arg Val
755 760 765

Val Asp Arg Phe Asp Lys Asp Gly Lys Tyr Ile Pro Pro Glu Val His
770 775 780

Asp Glu Leu Tyr Arg Phe Phe Asn Asn Leu Leu Ile Lys Glu Lys Leu
785 790 795 800

Gly Glu Lys Ala Arg Lys Ile Leu Asp Asn Lys Lys Val Lys Val Lys
805 810 815

Val Leu Glu Ser Glu Arg Val Lys Trp Ser Lys Phe Tyr Asp Glu Gln
820 825 830

Phe Ala Val Thr Phe Ser Val Lys Lys Asn Ala Asp Cys Leu Asp Thr
835 840 845

Thr Lys Asp Leu Asn Ala Glu Val Met Glu Gln Tyr Ser Glu Ser Asn
850 855 860

Arg Leu Ile Leu Ile Arg Asn Thr Thr Asp Ile Leu Tyr Tyr Leu Val
865 870 875 880

Leu Asp Lys Asn Gly Lys Val Leu Lys Gln Arg Ser Leu Asn Ile Ile
885 890 895

Asn Asp Gly Ala Arg Asp Val Asp Trp Lys Glu Arg Phe Arg Gln Val
900 905 910

Thr Lys Asp Arg Asn Glu Gly Tyr Asn Glu Trp Asp Tyr Ser Arg Thr
915 920 925

Ser Asn Asp Leu Lys Glu Val Tyr Leu Asn Tyr Ala Leu Lys Glu Ile
930 935 940

Ala Glu Ala Val Ile Glu Tyr Asn Ala Ile Leu Ile Ile Glu Lys Met
945 950 955 960

Ser Asn Ala Phe Lys Asp Lys Tyr Ser Phe Leu Asp Asp Val Thr Phe

965 970 975

Lys Gly Phe Glu Thr Lys Leu Leu Ala Lys Leu Ser Asp Leu His Phe
980 985 990

Arg Gly Ile Lys Asp Gly Glu Pro Cys Ser Phe Thr Asn Pro Leu Gln
995 1000 1005

Leu Cys Gln Asn Asp Ser Asn Lys Ile Leu Gln Asp Gly Val Ile
1010 1015 1020

Phe Met Val Pro Asn Ser Met Thr Arg Ser Leu Asp Pro Asp Thr
1025 1030 1035

Gly Phe Ile Phe Ala Ile Asn Asp His Asn Ile Arg Thr Lys Lys
1040 1045 1050

Ala Lys Leu Asn Phe Leu Ser Lys Phe Asp Gln Leu Lys Val Ser
1055 1060 1065

Ser Glu Gly Cys Leu Ile Met Lys Tyr Ser Gly Asp Ser Leu Pro
1070 1075 1080

Thr His Asn Thr Asp Asn Arg Val Trp Asn Cys Cys Cys Asn His
1085 1090 1095

Pro Ile Thr Asn Tyr Asp Arg Glu Thr Lys Lys Val Glu Phe Ile
1100 1105 1110

Glu Glu Pro Val Glu Glu Leu Ser Arg Val Leu Glu Glu Asn Gly
1115 1120 1125

Ile Glu Thr Asp Thr Glu Leu Asn Lys Leu Asn Glu Arg Glu Asn
1130 1135 1140

Val Pro Gly Lys Val Val Asp Ala Ile Tyr Ser Leu Val Leu Asn
1145 1150 1155

Tyr Leu Arg Gly Thr Val Ser Gly Val Ala Gly Gln Arg Ala Val
1160 1165 1170

Tyr Tyr Ser Pro Val Thr Gly Lys Lys Tyr Asp Ile Ser Phe Ile
1175 1180 1185

Gln Ala Met Asn Leu Asn Arg Lys Cys Asp Tyr Tyr Arg Ile Gly
1190 1195 1200

Ser Lys Glu Arg Gly Glu Trp Thr Asp Phe Val Ala Gln Leu Ile
1205 1210 1215

Asn

<210> 1167
<211> 10715
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1167
gaattccgga tgagcattca tcaggcgggc aagaatgtga ataaaggccg gataaaactt 60

gtgcttattt ttctttacgg tctttaaaaa ggccgtaata tccagctgaa cggtctggtt 120

ataggtacat tgagcaactg actgaaatgc ctcaaaatgt tctttacgat gccattggga 180

tatatcaacg gtggtatatc cagtgatttt tttctccatt ttagcttcct tagctcctga 240

aaatctcgat aactcaaaaa atacgcccgg tagtgatctt atttcattat ggtgaaagtt 300

ggaacctctt acgtgccgat caacgtctca ttttcgccaa aagttggccc agggcttccc 360

ggtatcaaca gggacaccag gatttattta ttctgcgaag tgatcttccg tcacaggtat 420

ttattcggcg caaagtgcgt cgggtgatgc tgccaactta ctgatttagt gtatgatggt 480

gtttttgagg tgctccagtg gcttctgttt ctatcagctg tccctcctgt tcagctactg 540

acggggtggt gcgtaacggc aaaagcaccg ccggacatca gcgctagcgg agtgtatact 600

ggcttactat gttggcactg atgagggtgt cagtgaagtg cttcatgtgg caggagaaaa 660

aaggctgcac cggtgcgtca gcagaatatg tgatacagga tatattccgc ttcctcgctc 720

actgactcgc tacgctcggt cgttcgactg cggcgagcgg aaatggctta cgaacggggc 780

ggagatttcc tggaagatgc caggaagata cttaacaggg aagtgagagg gccgcggcaa 840

agccgttttt ccataggctc cgccccccctg acaagcatca cgaaatctga cgctcaaatc 900

agtggtggcg aaacccgaca ggactataaa gataccaggc gtttcccccct ggcggctccc 960

tcgtgcgctc tcctgttcct gcctttcggt ttaccggtgt cattccgctg ttatggccgc 1020

gtttgtctca ttccacgcct gacactcagt tccgggtagg cagttcgctc caagctggac 1080

tgtatgcacg aacccccccgt tcagtccgac cgctgcgcct tatccggtaa ctatcgtctt 1140

gagtccaacc cggaaagaca tgcaaaagca ccactggcag cagccactgg taattgattt 1200

agaggagtta gtcttgaagt catgcgccgg ttaaggctaa actgaaagga caagttttgg 1260

tgactgcgct cctccaagcc agttacctcg gttcaaagag ttggtagctc agagaacctt 1320

cgaaaaaccg ccctgcaagg cggttttttc gttttcagag caagagatta cgcgcagacc 1380

aaaacgatct caagaagatc atcttattaa tcagataaaa tatttcatca aggaattggt 1440

tctaagctta tagaagcaat gattaaggaa gccaaaaaaa ataatattga tgcaatattt 1500

gtcttaggtc atccaagtta ttatccaaaa tttggtttta aaccagccac agaatatcag 1560

ataaaatgtg aatatgatgt cccagcggat gtttttatgg tactagattt gtcagctaaa 1620

ctagctagtt taaaaggaca aactgtctac tatgccgatg agtttggcaa aatttttttag 1680

atctacaaaa ttataaacta aataaagatt cttataataa ctttatatat aatcgaaatg 1740

tagagaattt tataaggagt ctttatcatg tcaatttatc aagaatttgt taataaatat 1800

agtttaagta aaactctaag atttgagtta atcccacagg gtaaaacact tgaaaacata 1860

aaagcaagag gtttgatttt agatgatgag aaaagagcta aagactacaa aaaggctaaa 1920

caaataattg ataaatatca tcagtttttt atagaggaga tattaagttc ggtttgtatt 1980

agcgaagatt tattacaaaa ctattctgat gtttatttta aacttaaaaa gagtgatgat 2040

gataatctac aaaaagattt taaaagtgca aaagatacga taaagaaaca aatatctgaa 2100

tatataaagg actcagagaa atttaagaat ttgtttaatc aaaaccttat cgatgctaaa 2160

aaagggcaag agtcagattt aattctatgg ctaaagcaat ctaaggataa tggtatagaa 2220

ctatttaaag ccaatagtga tatcacagat atagatgagg cgttagaaat aatcaaatct 2280

tttaaaggtt ggacaactta ttttaagggt tttcatgaaa atagaaaaaa tgtttatagt 2340

agcaatgata ttcctacatc tattatttat aggatagtag atgataattt gcctaaattt 2400

ctagaaaata aagctaagta tgagagttta aaagacaaag ctccagaagc tataaactat 2460

gaacaaatta aaaaagattt ggcagaagag ctaacctttg atattgacta caaaacatct 2520

gaagttaatc aaagagtttt ttcacttgat gaagtttttg agatagcaaa ctttaataat 2580

tatctaaatc aaagtggtat tactaaattt aatactatta ttggtggtaa atttgtaaat 2640

ggtgaaaata caaagagaaa aggtataaat gaatatataa atctatactc acagcaaata 2700
aatgataaaa cactcaaaaa atataaaatg agtgttttat ttaagcaaat tttaagtgat 2760
acagaatcta aatcttttgt aattgataag ttagaagatg atagtgatgt agttacaacg 2820
atgcaaagtt tttatgagca aatagcagct tttaaaacag tagaagaaaa atctattaaa 2880
gaaacactat ctttattatt tgatgattta aaagctcaaa aacttgattt gagtaaaatt 2940
tattttaaaa atgataaatc tcttactgat ctatcacaac aagtttttga tgattatagt 3000
gttattggta cagcggtact agaatatata actcaacaaa tagcacctaa aaatcttgat 3060
aaccctagta agaaagagca agaattaata gccaaaaaaa ctgaaaaagc aaaatactta 3120
tctctagaaa ctataaagct tgccttagaa gaatttaata agcatagaga tatagataaa 3180
cagtgtaggt ttgaagaaat acttgcaaac tttgcggcta ttccgatgat atttgatgaa 3240
atagctcaaa acaaagacaa tttggcacag atatctatca aatatcaaaa tcaaggtaaa 3300
aaagacctac ttcaagctag tgcggaagat gatgttaaag ctatcaagga tcttttagat 3360
caaactaata atctcttaca taaactaaaa atatttcata ttagtcagtc agaagataag 3420
gcaaatattt tagacaagga tgagcatttt tatctagtat ttgaggagtg ctactttgag 3480
ctagcgaata tagtgcctct ttataacaaa attagaaact atataactca aaagccatat 3540
agtgatgaga aatttaagct caattttgag aactcgactt tggctaatgg ttgggataaa 3600
aataaagagc ctgacaatac ggcaatttta tttatcaaag atgataaata ttatctgggt 3660
gtgatgaata agaaaaataa caaaatattt gatgataaag ctatcaaaga aaataaaggc 3720
gagggttata aaaaaattgt ttataaactt ttacctggcg caaataaaat gttacctaag 3780
gttttctttt ctgctaaatc tataaaattt tataatccta gtgaagatat acttagaata 3840
agaaatcatt ccacacatac aaaaaatggt agtcctcaaa aaggatatga aaaatttgag 3900
tttaatattg aagattgccg aaaatttata gatttttata aacagtctat aagtaagcat 3960
ccggagtgga aagattttgg atttagattt tctgatactc aaagatataa ttctatagat 4020
gaattttata gagaagttga aaatcaaggc tacaaactaa cttttgaaaa tatatcagag 4080
agctatattg atagcgtagt taatcagggt aaattgtacc tattccaaat ctataataaa 4140
gatttttcag cttatagcaa agggcgacca aatctacata ctttatattg gaaagcgctg 4200
tttgatgaga gaaatcttca agatgtggtt tataagctaa atggtgaggc agagcttttt 4260
tatcgtaaac aatcaatacc taaaaaaatc actcacccag ctaaagaggc aatagctaat 4320
aaaaacaaag ataatcctaa aaaagagagt gtttttgaat atgatttaat caaagataaa 4380
cgctttactg aagataagtt tttctttcac tgtcctatta caatcaattt taaatctagt 4440
ggagctaata agtttaatga tgaaatcaat ttattgctaa aagaaaaagc aaatgatgtt 4500
catatattaa gtatagatag aggtgaaaga catttagctt actatacttt ggtagatggt 4560
aaaggcaata tcatcaaaca agatactttc aacatcattg gtaatgatag aatgaaaaca 4620
aactaccatg ataagcttgc tgcaatagag aaagataggg attcagctag gaaagactgg 4680
aaaaagataa ataacatcaa agagatgaaa gagggctatc tatctcaggt agttcatgaa 4740
atagctaagc tagttataga gtataatgct attgtggttt ttgaggattt aaattttgga 4800
tttaaaagag ggcgtttcaa ggtagagaag caggtctatc aaaagttaga aaaaatgcta 4860
attgagaaac taaactatct agttttcaaa gataatgagt ttgataaaac tgggggagtg 4920
cttagagctt atcagctaac agcacctttt gagactttta aaaagatggg taaacaaaca 4980
ggtattatct actatgtacc agctggtttt acttcaaaaa tttgtcctgt aactggtttt 5040
gtaaatcagt tatatcctaa gtatgaaagt gtcagcaaat ctcaagagtt ctttagtaag 5100
tttgacaaga tttgttataa ccttgataag ggctattttg agtttagttt tgattataaa 5160

aactttggtg acaaggctgc caaaggcaag tggactatag ctagctttgg gagtagattg 5220
attaacttta gaaattcaga taaaaatcat aattgggata ctcgagaagt ttatccaact 5280
aaagagttgg agaaattgct aaaagattat tctatcgaat atgggcatgg cgaatgtatc 5340
aaagcagcta tttgcggtga gagcgacaaa aagtttttg ctaagctaac tagtgtccta 5400
aatactatct tacaaatgcg taactcaaaa acaggtactg agttagatta tctaatttca 5460
ccagtagcag atgtaaatgg caatttcttt gattcgcgac aggcgccaaa aaatatgcct 5520
caagatgctg atgccaatgg tgcttatcat attgggctaa aaggtctgat gctactaggt 5580
aggatcaaaa ataatcaaga gggcaaaaaa ctcaatttgg ttatcaaaaa tgaagagtat 5640
tttgagttcg tgcagaatag gaataactaa ttcattcaag aatatattac cctgtcagtt 5700
tagcgactat tacctcttta ataatttgca ggggaattat tttagtaata gtaatataca 5760
caagagttat tgattatatg gaaaattata tttagataac atggttaaat gattttatat 5820
tctgtcctta ctcgatatat ttgcataata tctatagtaa tgcctcagat actacatact 5880
attcatctag ccaaacaaaa gggcgcgatg ctcataaaag tatcgataaa ggaatctata 5940
gtaccaaaaa agatgacctg atcggtatcg atgttattaa ccataaatat ggtttggttg 6000
gtaaaattga tgtttttcat aaagataagg gcttacttgt ggagagaaaa aggcaaatca 6060
agactatcta tgatggctat aaatatcagc tttatgcgca atatttttgt ctccaagaga 6120
tgggctatga tgtcaaagcc attaaatttt attcgatggt tgataataaa tcatacccaa 6180
tagctatacc aacttcagct gagttagaaa agtttgaaaa acatattcaa acaatcaagc 6240
aatataatcc aatggataac tcatttaggc aaaatattga aaagtgtaaa ttttgtatat 6300
atgcaaactt atgtgataaa acggacttgt agattatgtt tagtaaaaat gatattgaat 6360
caaagaatat agtttttgtt aatatttttg atggagtgaa acttagtcta tcattgggga 6420
atatagttat aaaagataaa gaaactgatg aggtgaaaac taagctttct gttcataaag 6480
ttcttgcatt gtttatcgta ggtaatatga cgatgacctc gcaactttta gagacctgta 6540
agaaaaatgc tatacagcta gtttttatga aaaatagctt tagaccatat ctatgttttg 6600
gtgatattgc tgaggctaat tttttagcta gatataagca atatagtgta gttgagcaag 6660
atataagttt agcaaggatt tttataacat caaagatacg caatcaacat aacttagtca 6720
aaagcctaag agataaaact ccagagcagc aagagatagt caaaaagaat aaacagctaa 6780
tagcagagtt agaaaatata acaagcctag cggagctaat gggtatagag ggcaatgttg 6840
ccaaaaattt cttcaaagga ttctatggac atttagatag ttggcaaggg cgcaaaccta 6900
gaataaaaca ggatccatat aatgttgttt tagacttggg ctatagtatg ttgtttaatt 6960
ttgtagagtg ttttttgcga ctttttggct ttgatttata caagggcttt tgtcatcaga 7020
cttggtataa gcgtaaatcc ctagtttgtg actttgttga gccatttaga tgtatagtgg 7080
ataaccaagt tagaaaatca tggaatctcg ggcaattttc tgtagaggat tttggttgca 7140
aaaatgagca gttttatata aaaaaagata aaacaaaaga ctactcaaaa atacttttg 7200
ccgagattat cagctacaag ctagagatat ttgaatatgt aagagaattt tatcgtgcct 7260
ttatgcgagg caaagaaatt gcagagtatc caatattttg ttatgaaact aggagggtgt 7320
atgttgatag tcagttatga ttttagtaat aataaagtac gtgcaaagtt tgccaaattt 7380
ctagaaagtt atggtgtacg tttacaatat tcggtatttg agctcaaata tagcaagaga 7440
atgttagact tgattttagc tgagatagaa aataactatg taccactatt tacaaatgct 7500
gatagtgttt taatctttaa tgctccagat aaagatgtga taaaatatgg ttatgcgatt 7560
catagagaac aagaggttgt ttttatagac taaaaattgc aaaccttagt ctttatgtta 7620
aaataactac taagttctta gagatattta aaaatatgac tgttgttata tatcaaaatg 7680

ctaaaaaaat catagatttt aggtcttttt ttgctgattt aggcaaaaac gggtctaaga 7740

actttaaata atttctactg ttgtagatga gaagtcattt aataaggcca ctgttaaaag 7800

tctaagaact ttaaataatt tctactgttg tagatgctac tattcctgtg ccttcagata 7860

attcagtcta agaactttaa ataatttcta ctgttgtaga tgtctagagc cttttgtatt 7920

agtagccggt ctaagaactt taaataattt ctactgttgt agattagcga tttatgaagg 7980

tcatttttt gtctagcttt aatgcggtag tttatcacag ttaaattgct aacgcagtca 8040

ggcaccgtgt atgaaatcta acaatgcgct catcgtcatc ctcggcaccg tcaccctgga 8100

tgctgtaggc ataggcttgg ttatgccggt actgccgggc ctcttgcggg atatcgtcca 8160

ttccgacagc atcgccagtc actatggcgt gctgctagcg ctatatgcgt tgatgcaatt 8220

tctatgcgca cccgttctcg gagcactgtc cgaccgcttt ggccgccgcc cagtcctgct 8280

cgcttcgcta cttggagcca ctatcgacta cgcgatcatg gcgaccacac ccgtcctgtg 8340

gatcctctac gccggacgca tcgtggccgg catcaccggc gccacaggtg cggttgctgg 8400

cgcctatatc gccgacatca ccgatgggga agatcgggct cgccacttcg ggctcatgag 8460

cgcttgtttc ggcgtgggta tggtggcagg ccccgtggcc gggggactgt tgggcgccat 8520

ctccttgcat gcaccattcc ttgcggcggc ggtgctcaac ggcctcaacc tactactggg 8580

ctgcttccta atgcaggagt cgcataaggg agagcgtcga ccgatgccct tgagagcctt 8640

caacccagtc agctccttcc ggtgggcgcg gggcatgact atcgtcgccg cacttatgac 8700

tgtcttcttt atcatgcaac tcgtaggaca ggtgccggca gcgctctggg tcattttcgg 8760

cgaggaccgc tttcgctgga gcgcgacgat gatcggcctg tcgcttgcgg tattcggaat 8820

cttgcacgcc ctcgctcaag ccttcgtcac tggtcccgcc accaaacgtt tcggcgagaa 8880

gcaggccatt atcgccggca tggcggccga cgcgctgggc tacgtcttgc tggcgttcgc 8940

gacgcgaggc tggatggcct tccccattat gattcttctc gcttccggcg gcatcgggat 9000

gcccgcgttg caggccatgc tgtccaggca ggtagatgac gaccatcagg gacagcttca 9060

aggatcgctc gcggctctta ccagcctaac ttcgatcatt ggaccgctga tcgtcacggc 9120

gatttatgcc gcctcggcga gcacatggaa cgggttggca tggattgtag gcgccgccct 9180

ataccttgtc tgcctccccg cgttgcgtcg cggtgcatgg agccgggcca cctcgacctg 9240

aatggaagcc ggcggcacct cgctaacgga ttcaccactc caagaattgg agccaatcaa 9300

ttcttgcgga gaactgtgaa tgcgcaaacc aacccttggc agaacatatc catcgcgtcc 9360

gccatctcca gcagccgcac gcggcgcatc tcgggcagcg ttgggtcctg gccacgggtg 9420

cgcatgatcg tgctcctgtc gttgaggacc cggctaggct ggcggggttg ccttactggt 9480

tagcagaatg aatcaccgat acgcgagcga acgtgaagcg actgctgctg caaaacgtct 9540

gcgacctgag caacaacatg aatggtcttc ggtttccgtg tttcgtaaag tctggaaacg 9600

cggaagtccc ctacgtgctg ctgaagttgc ccgcaacaga gagtggaacc aaccggtgat 9660

accacgatac tatgactgag agtcaacgcc atgagcggcc tcatttctta ttctgagtta 9720

caacagtccg caccgctgtc cggtagctcc ttccggtggg cgcggggcat gactatcgtc 9780

gccgcactta tgactgtctt ctttatcatg caactcgtag gacaggtgcc ggcagcgccc 9840

aacagtcccc cggccacggg gcctgccacc atacccacgc cgaaacaagc gccctgcacc 9900

attatgttcc ggatctgcat cgcaggatgc tgctggctac cctgtggaac acctacatct 9960

gtattaacga agcgctaacc gtttttatca ggctctggga ggcagaataa atgatcatat 10020

cgtcaattat tacctccacg gggagagcct gagcaaactg gcctcaggca tttgagaagc 10080

acacggtcac actgcttccg gtagtcaata aaccggtaaa ccagcaatag acataagcgg 10140

ctatttaacg accctgccct gaaccgacga ccgggtcgaa tttgctttcg aatttctgcc 10200

attcatccgc ttattatcac ttattcaggc gtagcaccag gcgtttaagg gcaccaataa 10260

ctgccttaaa aaaattacgc cccgccctgc cactcatcgc agtactgttg taattcatta 10320

agcattctgc cgacatggaa gccatcacag acggcatgat gaacctgaat cgccagcggc 10380

atcagcacct tgtcgccttg cgtataatat ttgcccatgg tgaaaacggg ggcgaagaag 10440

ttgtccatat tggccacgtt taaatcaaaa ctggtgaaac tcacccaggg attggctgag 10500

acgaaaaaca tattctcaat aaacccttta gggaaatagg ccaggttttc accgtaacac 10560

gccacatctt gcgaatatat gtgtagaaac tgccggaaat cgtcgtggta ttcactccag 10620

agcgatgaaa acgtttcagt ttgctcatgg aaaacggtgt aacaagggtg aacactatcc 10680

catatcacca gctcaccgtc tttcattgcc atacg 10715

<210> 1168
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1168
gccctgcaag gcggtttttt 20

<210> 1169
<211> 32
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1169
catcaaggaa ttgagcttat agaagccata gc 32

<210> 1170
<211> 29
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1170
tgatgctcca agtgatgcaa tcatccaag 29

<210> 1171
<211> 29
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1171
aaccatgtga accagccaca taaaatgtg 29

<210> 1172
<211> 48
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1172
atgtcaattt atcaagaata aatatagttt aagtaaaatg tcagtaaa 48

<210> 1173
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1173
aggagatatt aagttcggtt tgtattagcg aag 33

<210> 1174
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1174
tcgctaatac aaaccgaact taatatctcc t 31

<210> 1175
<211> 35
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1175
taatactatt attggtggta aatttgtaaa tggtg 35

<210> 1176
<211> 26
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1176
tgaggagtgc tactttgagc tagcga 26

<210> 1177
<211> 44
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1177
gattataaaa accttaggat tataaaattt aaagaaaacc ttag 44

<210> 1178
<211> 41
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1178
tatgttcata tgtgtggaat gattttctaa gtatatcttc a 41

<210> 1179
<211> 40
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1179
tgaagatata cttagaaatc attccacaca tatgaacata 40

<210> 1180
<211> 25
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1180
ggtaaaggtg aggcagtatc gtaaa 25

<210> 1181
<211> 26
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1181
ctgggggagt gcttagagct tatcag 26

<210> 1182
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1182
tgccaatggt tcatattggg tggg 24

<210> 1183
<211> 35
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1183
ttagttattc actcaaaata cttagttaaa aatac 35

<210> 1184
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1184
atttttgtct ccaagagatg ggctatga 28

<210> 1185
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1185
taaaaacaga ggttctctac ccgatact 28

<210> 1186
<211> 18
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1186
ggcaagggca cctagaag 18

<210> 1187
<211> 29
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1187
ggttatgcga ttcatagaga acaagaggt 29

<210> 1188
<211> 29
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1188
agacaaaaaa atgaccttca taaatcgct 29

<210> 1189
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1189
tcatacacgg ttgactgcgt t 21

<210> 1190
<211> 3912
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<220>
<221> CDS
<222> (1)..(3909)

<400> 1190
gcc acc atg aac atc aaa aac ttt acc ggg ctc tac ccc ctc agc aaa 48
Ala Thr Met Asn Ile Lys Asn Phe Thr Gly Leu Tyr Pro Leu Ser Lys
1 5 10 15

act ttg cgc ttt gaa ctc aag cct att ggc aaa acc aag gaa aac atc 96
Thr Leu Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Lys Glu Asn Ile
20 25 30

gag aaa aat ggc atc ctg acc aag gac gag caa cgg gct aaa gac tac 144
Glu Lys Asn Gly Ile Leu Thr Lys Asp Glu Gln Arg Ala Lys Asp Tyr
35 40 45

ctc ata gtc aaa ggc ttt att gac gag tat cac aag cag ttc atc aaa 192
Leu Ile Val Lys Gly Phe Ile Asp Glu Tyr His Lys Gln Phe Ile Lys
50 55 60

```
gac agg ctt tgg gac ttt aaa ttg cct ctc gaa agt gag ggg gag aag        240
Asp Arg Leu Trp Asp Phe Lys Leu Pro Leu Glu Ser Glu Gly Glu Lys
65                  70                  75                  80

aac agt ctc gaa gaa tac cag gaa ctg tac gag ctc act aag cgc aac        288
Asn Ser Leu Glu Glu Tyr Gln Glu Leu Tyr Glu Leu Thr Lys Arg Asn
                85                  90                  95

gat gcc cag gag gcc gac ttc acc gag att aaa gat aac ctt cgc agc        336
Asp Ala Gln Glu Ala Asp Phe Thr Glu Ile Lys Asp Asn Leu Arg Ser
            100                 105                 110

tct att acc gaa cag ctc acg aag tct gga tct gcg tac gat cgg att        384
Ser Ile Thr Glu Gln Leu Thr Lys Ser Gly Ser Ala Tyr Asp Arg Ile
        115                 120                 125

ttt aaa aaa gag ttc att aga gaa gac ctg gtc aac ttc ctc gaa gat        432
Phe Lys Lys Glu Phe Ile Arg Glu Asp Leu Val Asn Phe Leu Glu Asp
    130                 135                 140

gaa aaa gat aaa aat atc gtg aaa cag ttc gag gac ttt act aca tat        480
Glu Lys Asp Lys Asn Ile Val Lys Gln Phe Glu Asp Phe Thr Thr Tyr
145                 150                 155                 160

ttt acg ggt ttt tat gaa aat agg aag aac atg tac tct agc gaa gag        528
Phe Thr Gly Phe Tyr Glu Asn Arg Lys Asn Met Tyr Ser Ser Glu Glu
                165                 170                 175

aag tcc acg gcc atc gca tac cgg ctt atc cat cag aat ctg cca aaa        576
Lys Ser Thr Ala Ile Ala Tyr Arg Leu Ile His Gln Asn Leu Pro Lys
            180                 185                 190

ttc atg gac aac atg aga agt ttt gcc aaa att gca aat tcc agt gtt        624
Phe Met Asp Asn Met Arg Ser Phe Ala Lys Ile Ala Asn Ser Ser Val
        195                 200                 205

tcc gag cac ttt agc gac atc tat gaa agc tgg aag gaa tat ctg aat        672
Ser Glu His Phe Ser Asp Ile Tyr Glu Ser Trp Lys Glu Tyr Leu Asn
    210                 215                 220

gta aat agc atc gag gaa atc ttc cag ctc gac tat ttt agc gaa acc        720
Val Asn Ser Ile Glu Glu Ile Phe Gln Leu Asp Tyr Phe Ser Glu Thr
225                 230                 235                 240

ttg act cag cca cat att gag gtg tat aac tat att atc ggg aag aaa        768
Leu Thr Gln Pro His Ile Glu Val Tyr Asn Tyr Ile Ile Gly Lys Lys
                245                 250                 255

gtc ctg gaa gac gga acc gag ata aag ggc atc aac gag tat gtg aac        816
Val Leu Glu Asp Gly Thr Glu Ile Lys Gly Ile Asn Glu Tyr Val Asn
            260                 265                 270

ctc tac aat cag cag cag aaa gat aag agt aaa cga ctg cct ttc ctg        864
Leu Tyr Asn Gln Gln Gln Lys Asp Lys Ser Lys Arg Leu Pro Phe Leu
        275                 280                 285

gtg cca ctg tat aag caa att ttg tct gat agg gaa aaa ctc tcc tgg        912
Val Pro Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Lys Leu Ser Trp
    290                 295                 300

att gct gaa gag ttc gac agc gac aag aag atg ctg agc gct atc acc        960
Ile Ala Glu Glu Phe Asp Ser Asp Lys Lys Met Leu Ser Ala Ile Thr
305                 310                 315                 320

gag tct tac aac cac ctg cac aac gtg ttg atg ggt aac gag aac gaa       1008
Glu Ser Tyr Asn His Leu His Asn Val Leu Met Gly Asn Glu Asn Glu
                325                 330                 335

agc ctg cga aat ctg ctg ctg aat att aag gac tat aac ctg gag aaa       1056
Ser Leu Arg Asn Leu Leu Leu Asn Ile Lys Asp Tyr Asn Leu Glu Lys
            340                 345                 350

att aat atc aca aac gac ttg tct ctc acc gaa atc tcc cag aat ctt       1104
Ile Asn Ile Thr Asn Asp Leu Ser Leu Thr Glu Ile Ser Gln Asn Leu
        355                 360                 365

ttt ggc cga tat gat gta ttc aca aat ggg atc aaa aac aag ctg aga       1152
Phe Gly Arg Tyr Asp Val Phe Thr Asn Gly Ile Lys Asn Lys Leu Arg
    370                 375                 380

gtg ttg act cca agg aag aaa aag gag acg gac gaa aat ttt gag gac       1200
Val Leu Thr Pro Arg Lys Lys Lys Glu Thr Asp Glu Asn Phe Glu Asp
385                 390                 395                 400
```

```
cgc att aac aaa att ttt aag acc cag aag tcc ttc agc atc gct ttt      1248
Arg Ile Asn Lys Ile Phe Lys Thr Gln Lys Ser Phe Ser Ile Ala Phe
                405                 410                 415

ctg aac aag ctg cct cag ccc gaa atg gag gat ggg aag ccc cgg aac      1296
Leu Asn Lys Leu Pro Gln Pro Glu Met Glu Asp Gly Lys Pro Arg Asn
            420                 425                 430

att gag gac tat ttc att aca cag ggg gcg att aac acc aaa tct ata      1344
Ile Glu Asp Tyr Phe Ile Thr Gln Gly Ala Ile Asn Thr Lys Ser Ile
        435                 440                 445

cag aaa gaa gat atc ttc gcc caa att gag aat gca tac gag gat gca      1392
Gln Lys Glu Asp Ile Phe Ala Gln Ile Glu Asn Ala Tyr Glu Asp Ala
    450                 455                 460

cag gtg ttc ctg caa att aag gac acc gac aac aaa ctt agc cag aac      1440
Gln Val Phe Leu Gln Ile Lys Asp Thr Asp Asn Lys Leu Ser Gln Asn
465                 470                 475                 480

aag acg gcg gtg gaa aag atc aaa act ttg ctg gac gcc ttg aag gaa      1488
Lys Thr Ala Val Glu Lys Ile Lys Thr Leu Leu Asp Ala Leu Lys Glu
                485                 490                 495

ctc cag cac ttc atc aaa ccg ctg ctg ggc tct ggg gag gag aac gag      1536
Leu Gln His Phe Ile Lys Pro Leu Leu Gly Ser Gly Glu Glu Asn Glu
            500                 505                 510

aaa gac gaa ctg ttc tac ggt tcc ttc ctg gcc atc tgg gac gaa ctg      1584
Lys Asp Glu Leu Phe Tyr Gly Ser Phe Leu Ala Ile Trp Asp Glu Leu
        515                 520                 525

gac acc att aca cca ctt tat aac aaa gtg aga aat tgg ctg acc cga      1632
Asp Thr Ile Thr Pro Leu Tyr Asn Lys Val Arg Asn Trp Leu Thr Arg
    530                 535                 540

aaa cca tat tca aca gaa aaa atc aaa ttg aat ttc gac aac gct cag      1680
Lys Pro Tyr Ser Thr Glu Lys Ile Lys Leu Asn Phe Asp Asn Ala Gln
545                 550                 555                 560

ctg ctg gga ggg tgg gat gtc aat aaa gaa cac gac tgt gca ggt atc      1728
Leu Leu Gly Gly Trp Asp Val Asn Lys Glu His Asp Cys Ala Gly Ile
                565                 570                 575

ttg ttg cgg aaa aac gat agc tac tat ctc gga att atc aat aag aaa      1776
Leu Leu Arg Lys Asn Asp Ser Tyr Tyr Leu Gly Ile Ile Asn Lys Lys
            580                 585                 590

acc aac cac atc ttt gat acg gat att acg cca tca gat ggc gag tgc      1824
Thr Asn His Ile Phe Asp Thr Asp Ile Thr Pro Ser Asp Gly Glu Cys
        595                 600                 605

tat gac aaa atc gac tac aag ctc ctt ccc ggg gcg aac aaa atg ctt      1872
Tyr Asp Lys Ile Asp Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu
    610                 615                 620

cca aag gtg ttt ttt agt aag tcc cga atc aaa gag ttc gag cca tca      1920
Pro Lys Val Phe Phe Ser Lys Ser Arg Ile Lys Glu Phe Glu Pro Ser
625                 630                 635                 640

gag gcc ata atc aat tgc tat aag aag ggg aca cac aaa aaa gga aaa      1968
Glu Ala Ile Ile Asn Cys Tyr Lys Lys Gly Thr His Lys Lys Gly Lys
                645                 650                 655

aac ttt aac ctg acg gac tgt cac cgc ctg atc aac ttt ttt aag acc      2016
Asn Phe Asn Leu Thr Asp Cys His Arg Leu Ile Asn Phe Phe Lys Thr
            660                 665                 670

tca atc gag aaa cac gag gat tgg tca aaa ttc gga ttc aag ttc tcc      2064
Ser Ile Glu Lys His Glu Asp Trp Ser Lys Phe Gly Phe Lys Phe Ser
        675                 680                 685

gat acc gaa acg tat gag gat att agc ggt ttt tat aga gag gtc gag      2112
Asp Thr Glu Thr Tyr Glu Asp Ile Ser Gly Phe Tyr Arg Glu Val Glu
    690                 695                 700

cag cag gga tac agg ctg acg agc cat cca gtc agt gcc agc tat ata      2160
Gln Gln Gly Tyr Arg Leu Thr Ser His Pro Val Ser Ala Ser Tyr Ile
705                 710                 715                 720

cat agt ctg gtc aag gaa gga aaa ctg tac ctc ttc caa atc tgg aac      2208
His Ser Leu Val Lys Glu Gly Lys Leu Tyr Leu Phe Gln Ile Trp Asn
                725                 730                 735
```

```
aag gac ttt tct caa ttc tcc aag ggg acc cct aac ttg cac act ctc      2256
Lys Asp Phe Ser Gln Phe Ser Lys Gly Thr Pro Asn Leu His Thr Leu
            740                 745                 750

tat tgg aag atg ctg ttt gac aaa cgg aat ctt agc gat gtg gtt tat      2304
Tyr Trp Lys Met Leu Phe Asp Lys Arg Asn Leu Ser Asp Val Val Tyr
        755                 760                 765

aag ctg aat ggc cag gct gaa gtg ttc tat aga aag agc tcc att gaa      2352
Lys Leu Asn Gly Gln Ala Glu Val Phe Tyr Arg Lys Ser Ser Ile Glu
    770                 775                 780

cac cag aac cga att atc cac ccc gct cag cat ccc atc aca aat aag      2400
His Gln Asn Arg Ile Ile His Pro Ala Gln His Pro Ile Thr Asn Lys
785                 790                 795                 800

aat gag ctt aac aaa aag cac act agc acc ttc aaa tac gat atc atc      2448
Asn Glu Leu Asn Lys Lys His Thr Ser Thr Phe Lys Tyr Asp Ile Ile
                805                 810                 815

aaa gat cgc aga tac acg gtg gat aaa ttc cag ttc cat gtg ccc att      2496
Lys Asp Arg Arg Tyr Thr Val Asp Lys Phe Gln Phe His Val Pro Ile
            820                 825                 830

act ata aat ttt aag gcg acc ggg cag aac aac atc aac cca atc gtc      2544
Thr Ile Asn Phe Lys Ala Thr Gly Gln Asn Asn Ile Asn Pro Ile Val
        835                 840                 845

caa gag gtg att cgc caa aac ggt atc acc cac atc ata ggc atc gat      2592
Gln Glu Val Ile Arg Gln Asn Gly Ile Thr His Ile Ile Gly Ile Asp
    850                 855                 860

cga ggt gaa cgc cat ctt ctg tac ctc tct ctc atc gat ttg aaa ggc      2640
Arg Gly Glu Arg His Leu Leu Tyr Leu Ser Leu Ile Asp Leu Lys Gly
865                 870                 875                 880

aac atc atc aag cag atg act ctc aac gaa att att aat gag tat aag      2688
Asn Ile Ile Lys Gln Met Thr Leu Asn Glu Ile Ile Asn Glu Tyr Lys
                885                 890                 895

ggt gtg acc tat aag acc aac tac cat aac ctc ctg gag aag agg gag      2736
Gly Val Thr Tyr Lys Thr Asn Tyr His Asn Leu Leu Glu Lys Arg Glu
            900                 905                 910

aag gag cgg acc gag gcc aga cac tcc tgg agt agt att gaa agc ata      2784
Lys Glu Arg Thr Glu Ala Arg His Ser Trp Ser Ser Ile Glu Ser Ile
        915                 920                 925

aaa gaa ctg aag gat gga tac atg tca cag gtg att cac aaa att acg      2832
Lys Glu Leu Lys Asp Gly Tyr Met Ser Gln Val Ile His Lys Ile Thr
    930                 935                 940

gac atg atg gtt aag tac aat gcg att gtg gtc ctg gag gac ctc aac      2880
Asp Met Met Val Lys Tyr Asn Ala Ile Val Val Leu Glu Asp Leu Asn
945                 950                 955                 960

ggg ggg ttt atg cga ggc cgc cag aag gtc gag aag cag gtg tac cag      2928
Gly Gly Phe Met Arg Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln
                965                 970                 975

aaa ttt gaa aaa aag ttg atc gac aag ctg aac tat ctc gtt gac aag      2976
Lys Phe Glu Lys Lys Leu Ile Asp Lys Leu Asn Tyr Leu Val Asp Lys
            980                 985                 990

aaa ctc gac gct aac gag gtc ggc  gga gta ctg aat gct tat cag ctg      3024
Lys Leu Asp Ala Asn Glu Val Gly  Gly Val Leu Asn Ala Tyr Gln Leu
        995                 1000                1005

acc aac  aag ttc gag tct ttc  aag aag att ggg aaa  caa agc gga      3069
Thr Asn  Lys Phe Glu Ser Phe  Lys Lys Ile Gly Lys  Gln Ser Gly
    1010                1015                1020

ttt ttg  ttc tac atc ccc gcc tgg aac aca agc aaa atc gat cct      3114
Phe Leu  Phe Tyr Ile Pro Ala Trp Asn Thr Ser Lys Ile Asp Pro
    1025                1030                1035

ata aca  ggg ttc gtt aat ctg ttc aac acc agg tac gag tct atc      3159
Ile Thr  Gly Phe Val Asn Leu Phe Asn Thr Arg Tyr Glu Ser Ile
    1040                1045                1050

aag gag  aca aaa gtt ttt tgg tct aag ttt gat  att atc cga tac      3204
Lys Glu  Thr Lys Val Phe Trp Ser Lys Phe Asp  Ile Ile Arg Tyr
    1055                1060                1065
```

```
aat aaa  gag aag aat tgg ttc  gag ttc gtc ttc gat  tac aat acc        3249
Asn Lys  Glu Lys Asn Trp Phe  Glu Phe Val Phe Asp  Tyr Asn Thr
    1070                 1075                 1080

ttt acg  act aaa gcg gag gga  aca cgc act aag tgg  act ctg tgc        3294
Phe Thr  Thr Lys Ala Glu Gly  Thr Arg Thr Lys Trp  Thr Leu Cys
    1085                 1090                 1095

acc cac  ggc act cgc atc cag  aca ttc cgg aac cca  gaa aag aat        3339
Thr His  Gly Thr Arg Ile Gln  Thr Phe Arg Asn Pro  Glu Lys Asn
    1100                 1105                 1110

gcc cag  tgg gac aat aaa gag  atc aat ttg act gag  tcc ttc aaa        3384
Ala Gln  Trp Asp Asn Lys Glu  Ile Asn Leu Thr Glu  Ser Phe Lys
    1115                 1120                 1125

gct ctg  ttt gaa aag tac aag  atc gat atc acc agt  aat ctc aag        3429
Ala Leu  Phe Glu Lys Tyr Lys  Ile Asp Ile Thr Ser  Asn Leu Lys
    1130                 1135                 1140

gaa tcc  atc atg cag gaa acc  gag aag aag ttc ttc  cag gaa ctg        3474
Glu Ser  Ile Met Gln Glu Thr  Glu Lys Lys Phe Phe  Gln Glu Leu
    1145                 1150                 1155

cat aat  ctg ctc cac ctg acc  ctg cag atg agg aat  agc gtt act        3519
His Asn  Leu Leu His Leu Thr  Leu Gln Met Arg Asn  Ser Val Thr
    1160                 1165                 1170

gga acc  gac ata gac tat ttg  atc agc ccc gtt gcc  gat gag gat        3564
Gly Thr  Asp Ile Asp Tyr Leu  Ile Ser Pro Val Ala  Asp Glu Asp
    1175                 1180                 1185

gga aat  ttc tat gat agt cgc  ata aat ggc aaa aat  ttt ccg gag        3609
Gly Asn  Phe Tyr Asp Ser Arg  Ile Asn Gly Lys Asn  Phe Pro Glu
    1190                 1195                 1200

aat gcc  gat gcc aat ggc gcg  tac aac atc gca cga  aag ggt ctg        3654
Asn Ala  Asp Ala Asn Gly Ala  Tyr Asn Ile Ala Arg  Lys Gly Leu
    1205                 1210                 1215

atg ctt  att cgg cag atc aag  caa gca gat cca cag  aag aaa ttc        3699
Met Leu  Ile Arg Gln Ile Lys  Gln Ala Asp Pro Gln  Lys Lys Phe
    1220                 1225                 1230

aag ttt  gag aca atc acc aat  aaa gac tgg ctg aaa  ttc gcc caa        3744
Lys Phe  Glu Thr Ile Thr Asn  Lys Asp Trp Leu Lys  Phe Ala Gln
    1235                 1240                 1245

gac aag  ccc tat ctt aaa gat  ggc agc ggg aaa agg  ccg gcg gcc        3789
Asp Lys  Pro Tyr Leu Lys Asp  Gly Ser Gly Lys Arg  Pro Ala Ala
    1250                 1255                 1260

acg aaa  aag gcc ggc cag gca  aaa aag aaa aag gga  tcc tac cca        3834
Thr Lys  Lys Ala Gly Gln Ala  Lys Lys Lys Lys Gly  Ser Tyr Pro
    1265                 1270                 1275

tac gat  gtt cca gat tac gct  tat ccc tac gac gtg  cct gat tat        3879
Tyr Asp  Val Pro Asp Tyr Ala  Tyr Pro Tyr Asp Val  Pro Asp Tyr
    1280                 1285                 1290

gca tac  cca tac gat gtc ccc  gac tat gcc taa                       3912
Ala Tyr  Pro Tyr Asp Val Pro  Asp Tyr Ala
    1295                 1300
```

<210> 1191
<211> 1303
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 1191

```
Ala Thr Met Asn Ile Lys Asn Phe Thr Gly Leu Tyr Pro Leu Ser Lys
1               5                  10                  15


Thr Leu Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Lys Glu Asn Ile
            20                  25                  30
```

```
Glu Lys Asn Gly Ile Leu Thr Lys Asp Glu Gln Arg Ala Lys Asp Tyr
        35                  40                  45

Leu Ile Val Lys Gly Phe Ile Asp Glu Tyr His Lys Gln Phe Ile Lys
    50                  55                  60

Asp Arg Leu Trp Asp Phe Lys Leu Pro Leu Glu Ser Glu Gly Glu Lys
65                  70                  75                  80

Asn Ser Leu Glu Glu Tyr Gln Glu Leu Tyr Glu Leu Thr Lys Arg Asn
                85                  90                  95

Asp Ala Gln Glu Ala Asp Phe Thr Glu Ile Lys Asp Asn Leu Arg Ser
            100                 105                 110

Ser Ile Thr Glu Gln Leu Thr Lys Ser Gly Ser Ala Tyr Asp Arg Ile
        115                 120                 125

Phe Lys Lys Glu Phe Ile Arg Glu Asp Leu Val Asn Phe Leu Glu Asp
    130                 135                 140

Glu Lys Asp Lys Asn Ile Val Lys Gln Phe Glu Asp Phe Thr Thr Tyr
145                 150                 155                 160

Phe Thr Gly Phe Tyr Glu Asn Arg Lys Asn Met Tyr Ser Ser Glu Glu
                165                 170                 175

Lys Ser Thr Ala Ile Ala Tyr Arg Leu Ile His Gln Asn Leu Pro Lys
            180                 185                 190

Phe Met Asp Asn Met Arg Ser Phe Ala Lys Ile Ala Asn Ser Ser Val
        195                 200                 205

Ser Glu His Phe Ser Asp Ile Tyr Glu Ser Trp Lys Glu Tyr Leu Asn
    210                 215                 220

Val Asn Ser Ile Glu Glu Ile Phe Gln Leu Asp Tyr Phe Ser Glu Thr
225                 230                 235                 240

Leu Thr Gln Pro His Ile Glu Val Tyr Asn Tyr Ile Ile Gly Lys Lys
                245                 250                 255

Val Leu Glu Asp Gly Thr Glu Ile Lys Gly Ile Asn Glu Tyr Val Asn
            260                 265                 270

Leu Tyr Asn Gln Gln Gln Lys Asp Lys Ser Lys Arg Leu Pro Phe Leu
        275                 280                 285

Val Pro Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Lys Leu Ser Trp
    290                 295                 300

Ile Ala Glu Glu Phe Asp Ser Asp Lys Lys Met Leu Ser Ala Ile Thr
305                 310                 315                 320

Glu Ser Tyr Asn His Leu His Asn Val Leu Met Gly Asn Glu Asn Glu
                325                 330                 335

Ser Leu Arg Asn Leu Leu Leu Asn Ile Lys Asp Tyr Asn Leu Glu Lys
            340                 345                 350

Ile Asn Ile Thr Asn Asp Leu Ser Leu Thr Glu Ile Ser Gln Asn Leu
        355                 360                 365
```

```
Phe Gly Arg Tyr Asp Val Phe Thr Asn Gly Ile Lys Asn Lys Leu Arg
    370                 375                 380

Val Leu Thr Pro Arg Lys Lys Lys Glu Thr Asp Glu Asn Phe Glu Asp
385                 390                 395                 400

Arg Ile Asn Lys Ile Phe Lys Thr Gln Lys Ser Phe Ser Ile Ala Phe
                405                 410                 415

Leu Asn Lys Leu Pro Gln Pro Glu Met Glu Asp Gly Lys Pro Arg Asn
            420                 425                 430

Ile Glu Asp Tyr Phe Ile Thr Gln Gly Ala Ile Asn Thr Lys Ser Ile
        435                 440                 445

Gln Lys Glu Asp Ile Phe Ala Gln Ile Glu Asn Ala Tyr Glu Asp Ala
    450                 455                 460

Gln Val Phe Leu Gln Ile Lys Asp Thr Asp Asn Lys Leu Ser Gln Asn
465                 470                 475                 480

Lys Thr Ala Val Glu Lys Ile Lys Thr Leu Leu Asp Ala Leu Lys Glu
                485                 490                 495

Leu Gln His Phe Ile Lys Pro Leu Leu Gly Ser Gly Glu Glu Asn Glu
            500                 505                 510

Lys Asp Glu Leu Phe Tyr Gly Ser Phe Leu Ala Ile Trp Asp Glu Leu
        515                 520                 525

Asp Thr Ile Thr Pro Leu Tyr Asn Lys Val Arg Asn Trp Leu Thr Arg
    530                 535                 540

Lys Pro Tyr Ser Thr Glu Lys Ile Lys Leu Asn Phe Asp Asn Ala Gln
545                 550                 555                 560

Leu Leu Gly Gly Trp Asp Val Asn Lys Glu His Asp Cys Ala Gly Ile
                565                 570                 575

Leu Leu Arg Lys Asn Asp Ser Tyr Tyr Leu Gly Ile Ile Asn Lys Lys
            580                 585                 590

Thr Asn His Ile Phe Asp Thr Asp Ile Thr Pro Ser Asp Gly Glu Cys
        595                 600                 605

Tyr Asp Lys Ile Asp Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu
    610                 615                 620

Pro Lys Val Phe Phe Ser Lys Ser Arg Ile Lys Glu Phe Glu Pro Ser
625                 630                 635                 640

Glu Ala Ile Ile Asn Cys Tyr Lys Lys Gly Thr His Lys Lys Gly Lys
                645                 650                 655

Asn Phe Asn Leu Thr Asp Cys His Arg Leu Ile Asn Phe Phe Lys Thr
            660                 665                 670

Ser Ile Glu Lys His Glu Asp Trp Ser Lys Phe Gly Phe Lys Phe Ser
        675                 680                 685

Asp Thr Glu Thr Tyr Glu Asp Ile Ser Gly Phe Tyr Arg Glu Val Glu
    690                 695                 700
```

```
Gln Gln Gly Tyr Arg Leu Thr Ser His Pro Val Ser Ala Ser Tyr Ile
705                 710                 715                 720

His Ser Leu Val Lys Glu Gly Lys Leu Tyr Leu Phe Gln Ile Trp Asn
                725                 730                 735

Lys Asp Phe Ser Gln Phe Ser Lys Gly Thr Pro Asn Leu His Thr Leu
            740                 745                 750

Tyr Trp Lys Met Leu Phe Asp Lys Arg Asn Leu Ser Asp Val Val Tyr
        755                 760                 765

Lys Leu Asn Gly Gln Ala Glu Val Phe Tyr Arg Lys Ser Ser Ile Glu
    770                 775                 780

His Gln Asn Arg Ile Ile His Pro Ala Gln His Pro Ile Thr Asn Lys
785                 790                 795                 800

Asn Glu Leu Asn Lys Lys His Thr Ser Thr Phe Lys Tyr Asp Ile Ile
                805                 810                 815

Lys Asp Arg Arg Tyr Thr Val Asp Lys Phe Gln Phe His Val Pro Ile
            820                 825                 830

Thr Ile Asn Phe Lys Ala Thr Gly Gln Asn Asn Ile Asn Pro Ile Val
        835                 840                 845

Gln Glu Val Ile Arg Gln Asn Gly Ile Thr His Ile Ile Gly Ile Asp
    850                 855                 860

Arg Gly Glu Arg His Leu Leu Tyr Leu Ser Leu Ile Asp Leu Lys Gly
865                 870                 875                 880

Asn Ile Ile Lys Gln Met Thr Leu Asn Glu Ile Ile Asn Glu Tyr Lys
                885                 890                 895

Gly Val Thr Tyr Lys Thr Asn Tyr His Asn Leu Leu Glu Lys Arg Glu
            900                 905                 910

Lys Glu Arg Thr Glu Ala Arg His Ser Trp Ser Ser Ile Glu Ser Ile
        915                 920                 925

Lys Glu Leu Lys Asp Gly Tyr Met Ser Gln Val Ile His Lys Ile Thr
    930                 935                 940

Asp Met Met Val Lys Tyr Asn Ala Ile Val Val Leu Glu Asp Leu Asn
945                 950                 955                 960

Gly Gly Phe Met Arg Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln
                965                 970                 975

Lys Phe Glu Lys Lys Leu Ile Asp Lys Leu Asn Tyr Leu Val Asp Lys
            980                 985                 990

Lys Leu Asp Ala Asn Glu Val Gly Gly Val Leu Asn Ala Tyr Gln Leu
        995                 1000                1005

Thr Asn Lys Phe Glu Ser Phe Lys Lys Ile Gly Lys Gln Ser Gly
    1010                1015                1020

Phe Leu Phe Tyr Ile Pro Ala Trp Asn Thr Ser Lys Ile Asp Pro
    1025                1030                1035
```

```
Ile Thr  Gly Phe Val Asn Leu  Phe Asn Thr Arg Tyr  Glu Ser Ile
    1040                1045                1050

Lys Glu  Thr Lys Val Phe Trp  Ser Lys Phe Asp Ile  Ile Arg Tyr
    1055                1060                1065

Asn Lys  Glu Lys Asn Trp Phe  Glu Phe Val Phe Asp  Tyr Asn Thr
    1070                1075                1080

Phe Thr  Thr Lys Ala Glu Gly  Thr Arg Thr Lys Trp  Thr Leu Cys
    1085                1090                1095

Thr His  Gly Thr Arg Ile Gln  Thr Phe Arg Asn Pro  Glu Lys Asn
    1100                1105                1110

Ala Gln  Trp Asp Asn Lys Glu  Ile Asn Leu Thr Glu  Ser Phe Lys
    1115                1120                1125

Ala Leu  Phe Glu Lys Tyr Lys  Ile Asp Ile Thr Ser  Asn Leu Lys
    1130                1135                1140

Glu Ser  Ile Met Gln Glu Thr  Glu Lys Lys Phe Phe  Gln Glu Leu
    1145                1150                1155

His Asn  Leu Leu His Leu Thr  Leu Gln Met Arg Asn  Ser Val Thr
    1160                1165                1170

Gly Thr  Asp Ile Asp Tyr Leu  Ile Ser Pro Val Ala  Asp Glu Asp
    1175                1180                1185

Gly Asn  Phe Tyr Asp Ser Arg  Ile Asn Gly Lys Asn  Phe Pro Glu
    1190                1195                1200

Asn Ala  Asp Ala Asn Gly Ala  Tyr Asn Ile Ala Arg  Lys Gly Leu
    1205                1210                1215

Met Leu  Ile Arg Gln Ile Lys  Gln Ala Asp Pro Gln  Lys Lys Phe
    1220                1225                1230

Lys Phe  Glu Thr Ile Thr Asn  Lys Asp Trp Leu Lys  Phe Ala Gln
    1235                1240                1245

Asp Lys  Pro Tyr Leu Lys Asp  Gly Ser Gly Lys Arg  Pro Ala Ala
    1250                1255                1260

Thr Lys  Lys Ala Gly Gln Ala  Lys Lys Lys Lys Gly  Ser Tyr Pro
    1265                1270                1275

Tyr Asp  Val Pro Asp Tyr Ala  Tyr Pro Tyr Asp Val  Pro Asp Tyr
    1280                1285                1290

Ala Tyr  Pro Tyr Asp Val Pro  Asp Tyr Ala
    1295                1300
```

<210> 1192
<211> 112
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<220>
<221> 修飾的鹼基

<222> (30)..(37)
<223> a、c、t、g、未知或未知

<400> 1192
ggccagtgaa ttcgagctcg gtacccgggn nnnnnnngag aagtcattta ataaggccac 60

tgttaaaaag cttggcgtaa tcatggtcat agctgtttcc tgtgtgaaat tg 112

<210> 1193
<211> 112
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<220>
<221> 修飾的鹼基
<222> (61)..(68)
<223> a、c、t、g、未知或未知

<400> 1193
ggccagtgaa ttcgagctcg gtacccgggg agaagtcatt taataaggcc actgttaaaa 60

nnnnnnnnag cttggcgtaa tcatggtcat agctgtttcc tgtgtgaaat tg 112

<210> 1194
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1194
ggccagtgaa ttcgagctcg g 21

<210> 1195
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1195
caatttcaca caggaaacag ctatgacc 28

<210> 1196
<211> 81
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1196
aaguucuuag agauauuuaa aaauaugacu guuguuauau aucaaaaugc uaaaaaaauc 60

auagauuuua ggucuuuuuu u 81

<210> 1197
<211> 111
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1197
aatttttag atctacaaaa ttataaacta aataaagatt cttataataa ctttatatat 60

aatcgaaatg tagagaattt tataaggagt ctttatcatg tcaatttatc a 111

<210> 1198
<211> 77
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1198
aaaattataa actaaataaa gattcttata ataactttat atataatcga aatgtagaga 60

attttataag gagtctt 77

<210> 1199
<211> 77
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1199
aaaattataa actaaataaa gattcttata ataactttat atataatcga aatgtagaga 60

attttataag gagtctt 77

<210> 1200
<211> 77
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1200
aaaattataa actaaataaa gattcttata ataactttat atataatcga aatgtagaga 60

attttataag gagtctt 77

<210> 1201
<211> 77
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1201
aaattataaa ctaaataaag attcttataa taactttata tataatcgaa atgtagagaa 60

ttttataagg agtcttt 77

<210> 1202
<211> 14
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1202
aatttttag atct 14

<210> 1203
<211> 77
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1203
aaaauuauaa acuaaauaaa gauucuuaua auaacuuuau auauaaucga auguagagaa 60

auuuuauaag gagucuu 77

<210> 1204
<211> 44
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<220>
<221> MOD_RES
<222> (16)..(16)
<223>任何胺基酸

<400> 1204
Pro Tyr Ser Lys Gly Lys Pro Asn Leu His Thr Leu Tyr Trp Lys Xaa
1 5 10 15

Leu Phe Ser Glu Arg Asn Leu Lys Asp Val Val Tyr Lys Leu Asn Gly
20 25 30

Glu Ala Glu Ile Phe Tyr Arg Lys Lys Ser Ile Lys
35 40

<210> 1205
<211> 77
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1205
actttaaata atttctactg ttgtagatga gaagtcattt aataaggcca ctgttaaaag 60

tctaagaact ttaaata 77

<210> 1206
<211> 77
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1206
ctactattcc tgtgccttca gataattcag tctaagaact ttaaataatt tctactgttg 60

tagatgtcta gagcctt 77

<210> 1207
<211> 103
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1207
gtctaagaac tttaaataat ttctactgtt gtagatgaga agtcatttaa taaggccact 60

gttaaaagtc taagaacttt aaataatttc tactgttgta gat 103

<210> 1208
<211> 149
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1208
ctgatttagg caaaaacggg tctaagaact ttaaataatt tctactgttg tagatgagaa 60

gtcatttaat aaggccactg ttaaaagtct aagaacttta aataatttct actgttgtag 120

atgctactat tcctgtgcct tcagataat 149

<210> 1209
<211> 34
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1209
ttagagaagt catttaataa ggccactgtt aaaa 34

<210> 1210
<211> 28
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1210
acuuuaaaua auuucuacug uuguagau 28

<210> 1211
<211> 35
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1211
gagaagucau uuaauaaggc cacuguuaaa aagcu 35

<210> 1212
<211> 17
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1212
gucuaagaac uuuaaau 17

<210> 1213
<211> 28
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1213
acuuuaaaua auuucuacug uuguagau 28

<210> 1214
<211> 31
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的

寡核苷酸

<400> 1214
gagaagucau uuaauaaggc cacuguuaaa a 31

<210> 1215
<211> 17
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1215
gucuaagaac uuuaaau 17

<210> 1216
<211> 28
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1216
acuuuaaaua auuucuacug uuguagau 28

<210> 1217
<211> 30
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1217
gagaagucau uuaauaaggc cacuguuaaa 30

<210> 1218
<211> 17
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1218
gucuaagaac uuuaaau 17

<210> 1219
<211> 28
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1219
acuuuaaaua auuucuacug uuguagau 28

<210> 1220
<211> 27
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1220
gagaagucau uuaauaaggc cacuguu 27

<210> 1221
<211> 17
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 1221
gucuaagaac uuuaaau 17

<210> 1222
<211> 28
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 1222
acuuuaaaua auuucuacug uuguagau 28

<210> 1223
<211> 24
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 1223
gagaagucau uuaauaaggc cacu 24

<210> 1224
<211> 17
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 1224
gucuaagaac uuuaaau 17

<210> 1225
<211> 28
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 1225
acuuuaaaua auuucuacug uuguagau 28

<210> 1226
<211> 20
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 1226
gagaagucau uuaauaaggc 20

<210> 1227
<211> 17
<212> RNA
<213> 人工序列

<220>

<223> 人工序列的描述：合成的
寡核苷酸

<400> 1227
gucuaagaac uuuaaau 17

<210> 1228
<211> 28
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1228
acuuuaaaua auuucuacug uuguagau 28

<210> 1229
<211> 17
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1229
gagaagucau uuaauaa 17

<210> 1230
<211> 17
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1230
gucuaagaac uuuaaau 17

<210> 1231
<211> 28
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1231
acuuuaaaua auuucuacug uuguagau 28

<210> 1232
<211> 14
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1232
gagaagucau uuaa 14

<210> 1233
<211> 17
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1233
gucuaagaac uuuaaau 17

<210> 1234
<211> 28
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1234
acuuuaaaua auuucuacug uuguagau 28

<210> 1235
<211> 27
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1235
agucauuuaa uaaggccacu guuaaaa 27

<210> 1236
<211> 17
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1236
gucuaagaac uuuaaau 17

<210> 1237
<211> 28
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1237
acuuuaaaua auuucuacug uuguagau 28

<210> 1238
<211> 24
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1238
cauuuaauaa ggccacuguu aaaa 24

<210> 1239
<211> 17
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1239
gucuaagaac uuuaaau 17

<210> 1240
<211> 28
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1240
acuuuaaaua auuucuacug uuguagau 28

<210> 1241
<211> 20
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1241
uaauaaggcc acuguuaaaa 20

<210> 1242
<211> 17
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1242
gucuaagaac uuuaaau 17

<210> 1243
<211> 28
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1243
acuuuaaaua auuucuacug uuguagau 28

<210> 1244
<211> 17
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1244
uaaggccacu guuaaaa 17

<210> 1245
<211> 17
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1245
gucuaagaac uuuaaau 17

<210> 1246
<211> 28
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1246
acuuuaaaua auuucuacug uuguagau 28

<210> 1247
<211> 14
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1247
ggccacuguu aaaa 14

<210> 1248
<211> 17
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1248
gucuaagaac uuuaaau 17

<210> 1249
<211> 103
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1249
gtctaagaac tttaaataat ttctactgtt gtagatgaga agtcatttaa taaggccact 60

gttaaaagtc taagaacttt aaataatttc tactgttgta gat 103

<210> 1250
<211> 3849
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1250
ggtaccatgg attacggcaa cggccagttt gagcggagag ccccccctgac caagacaatc 60

accctgcgcc tgaagcctat cggcgagaca cgggagacaa tccgcgagca gaagctgctg 120

gagcaggacg ccgccttcag aaagctggtg gagacagtga cccctatcgt ggacgattgt 180

atcaggaaga tcgccgataa cgccctgtgc cactttggca ccgagtatga cttcagctgt 240

ctgggcaacg ccatctctaa gaatgacagc aaggccatca agaaggagac agagaaggtg 300

gagaagctgc tggccaaggt gctgaccgag aatctgccag atggcctgcg caaggtgaac 360

gacatcaatt ccgccgcctt tatccaggat acactgacct ctttcgtgca ggacgatgcc 420

gacaagcggg tgctgatcca ggagctgaag ggcaagaccg tgctgatgca gcggttcctg 480

accacacgga tcacagccct gaccgtgtgg ctgcccgaca gagtgttcga gaactttaat 540

atcttcatcg agaacgccga gaagatgaga atcctgctgg actcccctct gaatgagaag 600

atcatgaagt ttgacccaga tgccgagcag tacgcctctc tggagttcta tggccagtgc 660

ctgtctcaga aggacatcga tagctacaac ctgatcatct ccggcatcta tgccgacgat 720

gaggtgaaga accctggcat caatgagatc gtgaaggagt acaatcagca gatccggggc 780

gacaaggatg agtccccact gcccaagctg aagaagctgc acaagcagat cctgatgcca 840

gtggagaagg ccttctttgt gcgcgtgctg tctaacgaca gcgatgcccg gagcatcctg 900

gagaagatcc tgaaggacac agagatgctg ccctccaaga tcatcgaggc catgaaggag 960

gcagatgcag gcgacatcgc cgtgtacggc agccggctgc acgagctgag ccacgtgatc 1020

tacggcgatc acggcaagct gtcccagatc atctatgaca aggagtccaa gaggatctct 1080

gagctgatgg agacactgtc tccaaaggag cgcaaggaga gcaagaagcg gctggagggc 1140

ctggaggagc acatcagaaa gtctacatac accttcgacg agctgaacag gtatgccgag 1200

aagaatgtga tggcagcata catcgcagca gtggaggagt cttgtgccga gatcatgaga 1260

aaggagaagg atctgaggac cctgctgagc aaggaggacg tgaagatccg gggcaacaga 1320

cacaatacac tgatcgtgaa gaactacttt aatgcctgga ccgtgttccg gaacctgatc 1380

agaatcctga ggcgcaagtc cgaggccgag atcgactctg acttctacga tgtgctggac 1440

gattccgtgg aggtgctgtc tctgacatac aagggcgaga atctgtgccg cagctatatc 1500

accaagaaga tcggctccga cctgaagccc gagatcgcca catacggcag cgccctgagg 1560

cctaacagcc gctggtggtc cccaggagag aagtttaatg tgaagttcca caccatcgtg 1620

cggagagatg gccggctgta ctatttcatc ctgcccaagg gcgccaagcc tgtggagctg 1680

gaggacatgg atggcgacat cgagtgtctg cagatgagaa agatccctaa cccaacaatc 1740

tttctgccca agctggtgtt caaggaccct gaggccttct ttagggataa tccagaggcc 1800

gacgagttcg tgtttctgag cggcatgaag gcccccgtga caatcaccag agagacatac 1860

gaggcctaca ggtataagct gtataccgtg ggcaagctgc gcgatggcga ggtgtccgaa 1920

gaggagtaca agcgggccct gctgcaggtg ctgaccgcct acaaggagtt tctggagaac 1980

agaatgatct atgccgacct gaatttcggc tttaaggatc tggaggagta taaggacagc 2040

tccgagttta tcaagcaggt ggagacacac aacaccttca tgtgctgggc caaggtgtct 2100

agctcccagc tggacgatct ggtgaagtct ggcaacggcc tgctgttcga gatctggagc 2160

gagcgcctgg agtcctacta taagtacggc aatgagaagg tgctgcgggg ctatgagggc 2220

gtgctgctga gcatcctgaa ggatgagaac ctggtgtcca tgcggaccct gctgaacagc 2280

cggcccatgc tggtgtaccg gccaaaggag tctagcaagc ctatggtggt gcaccgggat 2340

ggcagcagag tggtggacag gtttgataag gacggcaagt acatcccccc tgaggtgcac 2400

gacgagctgt atcgcttctt taacaatctg ctgatcaagg agaagctggg cgagaaggcc 2460

cggaagatcc tggacaacaa gaaggtgaag gtgaaggtgc tggagagcga gagagtgaag 2520

tggtccaagt tctacgatga gcagtttgcc gtgaccttca gcgtgaagaa gaacgccgat 2580

tgtctggaca ccacaaagga cctgaatgcc gaagtgatgg agcagtatag cgagtccaac 2640

agactgatcc tgatcaggaa taccacagat atcctgtact atctggtgct ggacaagaat 2700

ggcaaggtgc tgaagcagag atccctgaac atcatcaatg acggcgccag ggatgtggac 2760

tggaaggaga ggttccgcca ggtgacaaag gatagaaacg agggctacaa tgagtgggat 2820

tattccagga cctctaacga cctgaaggag gtgtacctga attatgccct gaaggagatc 2880

gccgaggccg tgatcgagta caacgccatc ctgatcatcg agaagatgtc taatgccttt 2940

aaggacaagt atagcttcct ggacgacgtg accttcaagg gcttcgagac aaagctgctg 3000

gccaagctga gcgatctgca ctttaggggc atcaaggacg gcgagccatg ttccttcaca 3060

aacccccctgc agctgtgcca gaacgattct aataagatcc tgcaggacgg cgtgatcttt 3120

atggtgccaa attctatgac acggagcctg gaccccgaca ccggcttcat ctttgccatc 3180

aacgaccaca atatcaggac caagaaggcc aagctgaact ttctgagcaa gttcgatcag 3240

ctgaaggtgt cctctgaggg ctgcctgatc atgaagtaca gcggcgattc cctgcctaca 3300

cacaacaccg acaatcgcgt gtggaactgc tgttgcaatc acccaatcac aaactatgac 3360

cgggagacaa agaaggtgga gttcatcgag gagcccgtgg aggagctgtc ccgcgtgctg 3420

gaggagaatg gcatcgagac agacaccgag ctgaacaagc tgaatgagcg ggagaacgtg 3480

cctggcaagg tggtggatgc catctactct ctggtgctga attatctgcg cggcacagtg 3540

agcggagtgg caggacagag ggccgtgtac tatagccctg tgaccggcaa gaagtacgat 3600

atctccttta tccaggccat gaacctgaat aggaagtgtg actactatag gatcggctcc 3660

aaggagaggg gagagtggac cgatttcgtg gcccagctga tcaacaaaag gccggcggcc 3720

acgaaaaagg ccggccaggc aaaaaagaaa aagggatcct acccatacga tgttccagat 3780

tacgcttatc cctacgacgt gcctgattat gcatacccat atgatgtccc cgactatgcc 3840

taagaattc 3849

<210> 1251
<211> 1233
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 1251
Met Asp Tyr Gly Asn Gly Gln Phe Glu Arg Arg Ala Pro Leu Thr Lys
1 5 10 15

Thr Ile Thr Leu Arg Leu Lys Pro Ile Gly Glu Thr Arg Glu Thr Ile
20 25 30

Arg Glu Gln Lys Leu Leu Glu Gln Asp Ala Ala Phe Arg Lys Leu Val
35 40 45

Glu Thr Val Thr Pro Ile Val Asp Asp Cys Ile Arg Lys Ile Ala Asp
50 55 60

Asn Ala Leu Cys His Phe Gly Thr Glu Tyr Asp Phe Ser Cys Leu Gly
65 70 75 80

Asn Ala Ile Ser Lys Asn Asp Ser Lys Ala Ile Lys Lys Glu Thr Glu
85 90 95

Lys Val Glu Lys Leu Leu Ala Lys Val Leu Thr Glu Asn Leu Pro Asp
100 105 110

Gly Leu Arg Lys Val Asn Asp Ile Asn Ser Ala Ala Phe Ile Gln Asp
115 120 125

Thr Leu Thr Ser Phe Val Gln Asp Asp Ala Asp Lys Arg Val Leu Ile
130 135 140

Gln Glu Leu Lys Gly Lys Thr Val Leu Met Gln Arg Phe Leu Thr Thr
145 150 155 160

Arg Ile Thr Ala Leu Thr Val Trp Leu Pro Asp Arg Val Phe Glu Asn
165 170 175

Phe Asn Ile Phe Ile Glu Asn Ala Glu Lys Met Arg Ile Leu Leu Asp
180 185 190

Ser Pro Leu Asn Glu Lys Ile Met Lys Phe Asp Pro Asp Ala Glu Gln
195 200 205

Tyr Ala Ser Leu Glu Phe Tyr Gly Gln Cys Leu Ser Gln Lys Asp Ile
210 215 220

Asp Ser Tyr Asn Leu Ile Ile Ser Gly Ile Tyr Ala Asp Asp Glu Val
225 230 235 240

Lys Asn Pro Gly Ile Asn Glu Ile Val Lys Glu Tyr Asn Gln Gln Ile
245 250 255

Arg Gly Asp Lys Asp Glu Ser Pro Leu Pro Lys Leu Lys Lys Leu His
260 265 270

Lys Gln Ile Leu Met Pro Val Glu Lys Ala Phe Phe Val Arg Val Leu
275 280 285

Ser Asn Asp Ser Asp Ala Arg Ser Ile Leu Glu Lys Ile Leu Lys Asp
290 295 300

Thr Glu Met Leu Pro Ser Lys Ile Ile Glu Ala Met Lys Glu Ala Asp
305 310 315 320

Ala Gly Asp Ile Ala Val Tyr Gly Ser Arg Leu His Glu Leu Ser His
325 330 335

Val Ile Tyr Gly Asp His Gly Lys Leu Ser Gln Ile Ile Tyr Asp Lys
340 345 350

Glu Ser Lys Arg Ile Ser Glu Leu Met Glu Thr Leu Ser Pro Lys Glu
355 360 365

Arg Lys Glu Ser Lys Lys Arg Leu Glu Gly Leu Glu Glu His Ile Arg
370 375 380

Lys Ser Thr Tyr Thr Phe Asp Glu Leu Asn Arg Tyr Ala Glu Lys Asn
385 390 395 400

Val Met Ala Ala Tyr Ile Ala Ala Val Glu Glu Ser Cys Ala Glu Ile
405 410 415

Met Arg Lys Glu Lys Asp Leu Arg Thr Leu Leu Ser Lys Glu Asp Val
420 425 430

Lys Ile Arg Gly Asn Arg His Asn Thr Leu Ile Val Lys Asn Tyr Phe
435 440 445

Asn Ala Trp Thr Val Phe Arg Asn Leu Ile Arg Ile Leu Arg Arg Lys
450 455 460

Ser Glu Ala Glu Ile Asp Ser Asp Phe Tyr Asp Val Leu Asp Asp Ser
465 470 475 480

Val Glu Val Leu Ser Leu Thr Tyr Lys Gly Glu Asn Leu Cys Arg Ser
485 490 495

Tyr Ile Thr Lys Lys Ile Gly Ser Asp Leu Lys Pro Glu Ile Ala Thr
500 505 510

Tyr Gly Ser Ala Leu Arg Pro Asn Ser Arg Trp Trp Ser Pro Gly Glu
515 520 525

Lys Phe Asn Val Lys Phe His Thr Ile Val Arg Arg Asp Gly Arg Leu
530 535 540

Tyr Tyr Phe Ile Leu Pro Lys Gly Ala Lys Pro Val Glu Leu Glu Asp
545 550 555 560

Met Asp Gly Asp Ile Glu Cys Leu Gln Met Arg Lys Ile Pro Asn Pro
565 570 575

Thr Ile Phe Leu Pro Lys Leu Val Phe Lys Asp Pro Glu Ala Phe Phe
580 585 590

Arg Asp Asn Pro Glu Ala Asp Glu Phe Val Phe Leu Ser Gly Met Lys
595 600 605

Ala Pro Val Thr Ile Thr Arg Glu Thr Tyr Glu Ala Tyr Arg Tyr Lys
610 615 620

Leu Tyr Thr Val Gly Lys Leu Arg Asp Gly Glu Val Ser Glu Glu Glu
625 630 635 640

Tyr Lys Arg Ala Leu Leu Gln Val Leu Thr Ala Tyr Lys Glu Phe Leu
645 650 655

Glu Asn Arg Met Ile Tyr Ala Asp Leu Asn Phe Gly Phe Lys Asp Leu
660 665 670

Glu Glu Tyr Lys Asp Ser Ser Glu Phe Ile Lys Gln Val Glu Thr His
675 680 685

Asn Thr Phe Met Cys Trp Ala Lys Val Ser Ser Ser Gln Leu Asp Asp
690 695 700

Leu Val Lys Ser Gly Asn Gly Leu Leu Phe Glu Ile Trp Ser Glu Arg
705 710 715 720

Leu Glu Ser Tyr Tyr Lys Tyr Gly Asn Glu Lys Val Leu Arg Gly Tyr
725 730 735

Glu Gly Val Leu Leu Ser Ile Leu Lys Asp Glu Asn Leu Val Ser Met
740 745 750

Arg Thr Leu Leu Asn Ser Arg Pro Met Leu Val Tyr Arg Pro Lys Glu
755 760 765

Ser Ser Lys Pro Met Val Val His Arg Asp Gly Ser Arg Val Val Asp
770 775 780

Arg Phe Asp Lys Asp Gly Lys Tyr Ile Pro Pro Glu Val His Asp Glu
785 790 795 800

Leu Tyr Arg Phe Phe Asn Asn Leu Leu Ile Lys Glu Lys Leu Gly Glu
805 810 815

Lys Ala Arg Lys Ile Leu Asp Asn Lys Lys Val Lys Val Lys Val Leu
820 825 830

Glu Ser Glu Arg Val Lys Trp Ser Lys Phe Tyr Asp Glu Gln Phe Ala
835 840 845

Val Thr Phe Ser Val Lys Lys Asn Ala Asp Cys Leu Asp Thr Thr Lys
850 855 860

Asp Leu Asn Ala Glu Val Met Glu Gln Tyr Ser Glu Ser Asn Arg Leu
865 870 875 880

Ile Leu Ile Arg Asn Thr Thr Asp Ile Leu Tyr Tyr Leu Val Leu Asp
885 890 895

Lys Asn Gly Lys Val Leu Lys Gln Arg Ser Leu Asn Ile Ile Asn Asp
900 905 910

```
Gly Ala Arg Asp Val Asp Trp Lys Glu Arg Phe Arg Gln Val Thr Lys
        915                 920                 925

Asp Arg Asn Glu Gly Tyr Asn Glu Trp Asp Tyr Ser Arg Thr Ser Asn
    930                 935                 940

Asp Leu Lys Glu Val Tyr Leu Asn Tyr Ala Leu Lys Glu Ile Ala Glu
945                 950                 955                 960

Ala Val Ile Glu Tyr Asn Ala Ile Leu Ile Ile Glu Lys Met Ser Asn
                965                 970                 975

Ala Phe Lys Asp Lys Tyr Ser Phe Leu Asp Asp Val Thr Phe Lys Gly
            980                 985                 990

Phe Glu Thr Lys Leu Leu Ala Lys Leu Ser Asp Leu His Phe Arg Gly
        995                 1000                1005

Ile Lys  Asp Gly Glu Pro Cys  Ser Phe Thr Asn Pro  Leu Gln Leu
    1010                1015                1020

Cys Gln  Asn Asp Ser Asn Lys  Ile Leu Gln Asp Gly  Val Ile Phe
    1025                1030                1035

Met Val  Pro Asn Ser Met Thr  Arg Ser Leu Asp Pro  Asp Thr Gly
    1040                1045                1050

Phe Ile  Phe Ala Ile Asn Asp  His Asn Ile Arg Thr  Lys Lys Ala
    1055                1060                1065

Lys Leu  Asn Phe Leu Ser Lys  Phe Asp Gln Leu Lys  Val Ser Ser
    1070                1075                1080

Glu Gly  Cys Leu Ile Met Lys  Tyr Ser Gly Asp Ser  Leu Pro Thr
    1085                1090                1095

His Asn  Thr Asp Asn Arg Val  Trp Asn Cys Cys Cys  Asn His Pro
    1100                1105                1110

Ile Thr  Asn Tyr Asp Arg Glu  Thr Lys Lys Val Glu  Phe Ile Glu
    1115                1120                1125

Glu Pro  Val Glu Glu Leu Ser  Arg Val Leu Glu Glu  Asn Gly Ile
    1130                1135                1140

Glu Thr  Asp Thr Glu Leu Asn  Lys Leu Asn Glu Arg  Glu Asn Val
    1145                1150                1155

Pro Gly  Lys Val Val Asp Ala  Ile Tyr Ser Leu Val  Leu Asn Tyr
    1160                1165                1170

Leu Arg  Gly Thr Val Ser Gly  Val Ala Gly Gln Arg  Ala Val Tyr
    1175                1180                1185

Tyr Ser  Pro Val Thr Gly Lys  Lys Tyr Asp Ile Ser  Phe Ile Gln
    1190                1195                1200

Ala Met  Asn Leu Asn Arg Lys  Cys Asp Tyr Tyr Arg  Ile Gly Ser
    1205                1210                1215

Lys Glu  Arg Gly Glu Trp Thr  Asp Phe Val Ala Gln  Leu Ile Asn
    1220                1225                1230
```

<210> 1252
<211> 3873
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1252
ggtaccatgc tgctgtatga gaactacacc aagcggaatc agatcacaaa gagcctgagg 60

ctggagctgc gccctcaggg caagaccctg agaaacatca aggagctgaa tctgctggag 120

caggacaagg ccatctacgc cctgctggag cggctgaagc cagtgatcga tgagggcatc 180

aaggacatcg ccagagatac cctgaagaac tgcgagctgt cttttgagaa gctgtacgag 240

cacttcctga gcggcgacaa gaaggcctat gccaaggagt ccgagcggct gaagaaggag 300

atcgtgaaaa ccctgatcaa gaacctgccc gagggcatcg gcaagatctc tgagatcaac 360

agcgccaagt atctgaatgg cgtgctgtac gacttcatcg ataagacaca caaggactct 420

gaggagaagc agaatatcct gagcgatatc ctggagacaa agggctacct ggccctgttc 480

tccaagtttc tgacatctcg gatcaccaca ctggagcagt ccatgcctaa gagagtgatc 540

gagaacttcg agatctatgc cgccaatatc ccaaagatgc aggacgccct ggagagggga 600

gccgtgtctt ttgccatcga gtacgagagc atctgttccg tggattacta taaccagatc 660

ctgtcccagg aggacatcga ttcttataat cgcctgatca gcggcatcat ggacgaggat 720

ggcgccaagg agaagggcat caaccagacc atctccgaga agaatatcaa gatcaagtct 780

gagcacctgg aggagaagcc cttcagaatc ctgaagcagc tgcacaagca gatcctggag 840

gagcgcgaga aggcctttac aatcgaccac atcgacagcg atgaggaggt ggtgcaggtg 900

accaaggagg ccttcgagca gacaaaggag cagtgggaga acatcaagaa gatcaatggc 960

ttctacgcca aggaccccgg cgatatcacc ctgtttatcg tggtgggccc taaccagaca 1020

cacgtgctgt cccagctgat ctacggcgag cacgaccgga tcagactgct gctggaggag 1080

tatgagaaga acaccctgga ggtgctgccc cggagaacaa agtctgagaa ggccagatac 1140

gataagttcg tgaatgccgt gcctaagaag gtggccaagg agagccacac cttcgacggc 1200

ctgcagaaga tgacaggcga cgatcggctg tttatcctgt atagagatga gctggcccgg 1260

aactacatga gaatcaagga ggcctatggc acctttgagc gggacatcct gaagagcagg 1320

cgcggcatca agggcaatcg ggacgtgcag gagtctctgg tgagctttta cgatgagctg 1380

acaaagttca ggagcgccct gcgcatcatc aattccggca acgacgagaa ggccgatcca 1440

atcttctata acacctttga tggcatcttc gagaaggcca ataggacata caaggccgag 1500

aatctgtgcc gcaactatgt gaccaagtct ccagccgacg atgccaggat catggccagc 1560

tgtctgggca ccccagcaag gctgcgcaca cactggtgga atggcgagga gaacttcgcc 1620

atcaatgacg tggccatgat ccggagaggc gatgagtact attactttgt gctgacccca 1680

gacgtgaagc ccgtggacct gaaaaccaag gacgagacag atgcccagat cttcgtgcag 1740

cgcaagggcg ccaagtcctt tctgggcctg ccaaaggccc tgttcaagtg catcctggag 1800

ccttactttg agtccccaga gcacaagaat gacaagaact gcgtgatcga ggagtacgtg 1860

tctaagcccc tgaccatcga caggcgcgcc tatgatatct ttaagaacgg caccttcaag 1920

aaaaccaata tcggcatcga cggcctgacc gaggagaagt tcaaggacga ttgccgctat 1980

ctgatcgacg tgtataagga gttcatcgcc gtgtatacaa ggtacagctg ttttaacatg 2040

tccggcctga agcgcgccga cgagtacaat gatatcggcg agttcttttc tgacgtggat 2100

accaggctgt gcacaatgga gtggattccc gtgagcttcg agcgcatcaa cgacatggtg 2160

gataagaagg agggcctgct gtttctggtg aggagcatgt tcctgtataa ccggcccaga 2220

aagccttatg agcgcacctt tatccagctg ttcagcgact ccaacatgga gcacacatct 2280

atgctgctga atagcagggc catgatccag tacagggcag cctccctgcc acggagagtg 2340

acccacaaga agggcagcat cctggtggcc ctgcgggatt ccaacggcga gcacatcccc 2400

atgcacatca gagaggccat ctacaagatg aagaacaatt ttgacatcag ctccgaggat 2460

ttcatcatgg ccaaggccta tctggccgag cacgacgtgg ccatcaagaa ggccaacgag 2520

gatatcatca ggaataggcg ctacaccgag gacaagttct ttctgtctct gagctatacc 2580

aagaacgccg atatcagcgc ccgcacactg gactacatca atgataaggt ggaggaggac 2640

acacaggatt ccaggatggc cgtgatcgtg acccgcaacc tgaaggacct gacatacgtg 2700

gccgtggtgg atgagaagaa caatgtgctg gaggagaaga gcctgaacga gatcgacggc 2760

gtgaattatc gggagctgct gaaggagaga accaagatca agtaccacga caagacacgg 2820

ctgtggcagt atgacgtgag cagcaagggc ctgaaggagg cctacgtgga gctggccgtg 2880

acccagatct ccaagctggc cacaaagtat aacgccgtgg tggtggtgga gtccatgtcc 2940

tctaccttca aggacaagtt ctcttttctg gatgagcaga tcttcaaggc ctttgaggcc 3000

cggctgtgcg ccagaatgtc cgacctgtct tttaatacaa tcaaggaggg cgaggccggc 3060

tccatctcta accccatcca ggtgtccaac aataacggca attcttatca ggacggcgtg 3120

atctacttcc tgaataacgc ctatacccgg accctgtgcc ctgataccgg ctttgtggac 3180

gtgttcgata agacccggct gatcacaatg cagtctaaga gacagttctt tgccaagatg 3240

aaggacatca gaatcgacga tggcgagatg ctgttcacct ttaacctgga ggagtaccct 3300

acaaagaggc tgctggaccg caaggagtgg accgtgaaga tcgccggcga tggctcctat 3360

ttcgacaagg ataagggcga gtacgtgtac gtgaacgaca tcgtgagaga gcagatcatc 3420

ccagccctgc tggaggacaa ggccgtgttc gatggcaata tggccgagaa gtttctggat 3480

aagaccgcca tcagcggcaa gtccgtggag ctgatctaca agtggttcgc caacgccctg 3540

tatggcatca tcacaaagaa ggacggcgag aagatctacc ggagccccat caccggcaca 3600

gagatcgacg tgagcaagaa caccacatac aacttcggca agaagttcat gttcaagcag 3660

gagtatagag gcgacggcga ttttctggac gccttcctga attacatgca ggcccaggat 3720

atcgccgtga aaaggccggc ggccacgaaa aaggccggcc aggcaaaaaa gaaaaaggga 3780

tcctacccat acgatgttcc agattacgct tatccctacg acgtgcctga ttatgcatac 3840

ccatatgatg tccccgacta tgcctaagaa ttc 3873

<210> 1253
<211> 1241
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 1253
Met Leu Leu Tyr Glu Asn Tyr Thr Lys Arg Asn Gln Ile Thr Lys Ser
1 5 10 15

Leu Arg Leu Glu Leu Arg Pro Gln Gly Lys Thr Leu Arg Asn Ile Lys
20 25 30

Glu Leu Asn Leu Leu Glu Gln Asp Lys Ala Ile Tyr Ala Leu Leu Glu
35 40 45

Arg Leu Lys Pro Val Ile Asp Glu Gly Ile Lys Asp Ile Ala Arg Asp
50 55 60

```
Thr Leu Lys Asn Cys Glu Leu Ser Phe Glu Lys Leu Tyr Glu His Phe
65                  70                  75                  80

Leu Ser Gly Asp Lys Lys Ala Tyr Ala Lys Glu Ser Glu Arg Leu Lys
                85                  90                  95

Lys Glu Ile Val Lys Thr Leu Ile Lys Asn Leu Pro Glu Gly Ile Gly
            100                 105                 110

Lys Ile Ser Glu Ile Asn Ser Ala Lys Tyr Leu Asn Gly Val Leu Tyr
        115                 120                 125

Asp Phe Ile Asp Lys Thr His Lys Asp Ser Glu Glu Lys Gln Asn Ile
    130                 135                 140

Leu Ser Asp Ile Leu Glu Thr Lys Gly Tyr Leu Ala Leu Phe Ser Lys
145                 150                 155                 160

Phe Leu Thr Ser Arg Ile Thr Thr Leu Glu Gln Ser Met Pro Lys Arg
                165                 170                 175

Val Ile Glu Asn Phe Glu Ile Tyr Ala Ala Asn Ile Pro Lys Met Gln
            180                 185                 190

Asp Ala Leu Glu Arg Gly Ala Val Ser Phe Ala Ile Glu Tyr Glu Ser
        195                 200                 205

Ile Cys Ser Val Asp Tyr Tyr Asn Gln Ile Leu Ser Gln Glu Asp Ile
    210                 215                 220

Asp Ser Tyr Asn Arg Leu Ile Ser Gly Ile Met Asp Glu Asp Gly Ala
225                 230                 235                 240

Lys Glu Lys Gly Ile Asn Gln Thr Ile Ser Glu Lys Asn Ile Lys Ile
                245                 250                 255

Lys Ser Glu His Leu Glu Glu Lys Pro Phe Arg Ile Leu Lys Gln Leu
            260                 265                 270

His Lys Gln Ile Leu Glu Glu Arg Glu Lys Ala Phe Thr Ile Asp His
        275                 280                 285

Ile Asp Ser Asp Glu Glu Val Val Gln Val Thr Lys Glu Ala Phe Glu
    290                 295                 300

Gln Thr Lys Glu Gln Trp Glu Asn Ile Lys Lys Ile Asn Gly Phe Tyr
305                 310                 315                 320

Ala Lys Asp Pro Gly Asp Ile Thr Leu Phe Ile Val Val Gly Pro Asn
                325                 330                 335

Gln Thr His Val Leu Ser Gln Leu Ile Tyr Gly Glu His Asp Arg Ile
            340                 345                 350

Arg Leu Leu Leu Glu Glu Tyr Glu Lys Asn Thr Leu Glu Val Leu Pro
        355                 360                 365

Arg Arg Thr Lys Ser Glu Lys Ala Arg Tyr Asp Lys Phe Val Asn Ala
    370                 375                 380

Val Pro Lys Lys Val Ala Lys Glu Ser His Thr Phe Asp Gly Leu Gln
385                 390                 395                 400
```

Lys Met Thr Gly Asp Asp Arg Leu Phe Ile Leu Tyr Arg Asp Glu Leu
405 410 415

Ala Arg Asn Tyr Met Arg Ile Lys Glu Ala Tyr Gly Thr Phe Glu Arg
420 425 430

Asp Ile Leu Lys Ser Arg Arg Gly Ile Lys Gly Asn Arg Asp Val Gln
435 440 445

Glu Ser Leu Val Ser Phe Tyr Asp Glu Leu Thr Lys Phe Arg Ser Ala
450 455 460

Leu Arg Ile Ile Asn Ser Gly Asn Asp Glu Lys Ala Asp Pro Ile Phe
465 470 475 480

Tyr Asn Thr Phe Asp Gly Ile Phe Glu Lys Ala Asn Arg Thr Tyr Lys
485 490 495

Ala Glu Asn Leu Cys Arg Asn Tyr Val Thr Lys Ser Pro Ala Asp Asp
500 505 510

Ala Arg Ile Met Ala Ser Cys Leu Gly Thr Pro Ala Arg Leu Arg Thr
515 520 525

His Trp Trp Asn Gly Glu Glu Asn Phe Ala Ile Asn Asp Val Ala Met
530 535 540

Ile Arg Arg Gly Asp Glu Tyr Tyr Tyr Phe Val Leu Thr Pro Asp Val
545 550 555 560

Lys Pro Val Asp Leu Lys Thr Lys Asp Glu Thr Asp Ala Gln Ile Phe
565 570 575

Val Gln Arg Lys Gly Ala Lys Ser Phe Leu Gly Leu Pro Lys Ala Leu
580 585 590

Phe Lys Cys Ile Leu Glu Pro Tyr Phe Glu Ser Pro Glu His Lys Asn
595 600 605

Asp Lys Asn Cys Val Ile Glu Glu Tyr Val Ser Lys Pro Leu Thr Ile
610 615 620

Asp Arg Arg Ala Tyr Asp Ile Phe Lys Asn Gly Thr Phe Lys Lys Thr
625 630 635 640

Asn Ile Gly Ile Asp Gly Leu Thr Glu Glu Lys Phe Lys Asp Asp Cys
645 650 655

Arg Tyr Leu Ile Asp Val Tyr Lys Glu Phe Ile Ala Val Tyr Thr Arg
660 665 670

Tyr Ser Cys Phe Asn Met Ser Gly Leu Lys Arg Ala Asp Glu Tyr Asn
675 680 685

Asp Ile Gly Glu Phe Phe Ser Asp Val Asp Thr Arg Leu Cys Thr Met
690 695 700

Glu Trp Ile Pro Val Ser Phe Glu Arg Ile Asn Asp Met Val Asp Lys
705 710 715 720

Lys Glu Gly Leu Leu Phe Leu Val Arg Ser Met Phe Leu Tyr Asn Arg
725 730 735

Pro Arg Lys Pro Tyr Glu Arg Thr Phe Ile Gln Leu Phe Ser Asp Ser
740 745 750

Asn Met Glu His Thr Ser Met Leu Leu Asn Ser Arg Ala Met Ile Gln
755 760 765

Tyr Arg Ala Ala Ser Leu Pro Arg Arg Val Thr His Lys Lys Gly Ser
770 775 780

Ile Leu Val Ala Leu Arg Asp Ser Asn Gly Glu His Ile Pro Met His
785 790 795 800

Ile Arg Glu Ala Ile Tyr Lys Met Lys Asn Asn Phe Asp Ile Ser Ser
805 810 815

Glu Asp Phe Ile Met Ala Lys Ala Tyr Leu Ala Glu His Asp Val Ala
820 825 830

Ile Lys Lys Ala Asn Glu Asp Ile Ile Arg Asn Arg Arg Tyr Thr Glu
835 840 845

Asp Lys Phe Phe Leu Ser Leu Ser Tyr Thr Lys Asn Ala Asp Ile Ser
850 855 860

Ala Arg Thr Leu Asp Tyr Ile Asn Asp Lys Val Glu Glu Asp Thr Gln
865 870 875 880

Asp Ser Arg Met Ala Val Ile Val Thr Arg Asn Leu Lys Asp Leu Thr
885 890 895

Tyr Val Ala Val Val Asp Glu Lys Asn Asn Val Leu Glu Glu Lys Ser
900 905 910

Leu Asn Glu Ile Asp Gly Val Asn Tyr Arg Glu Leu Leu Lys Glu Arg
915 920 925

Thr Lys Ile Lys Tyr His Asp Lys Thr Arg Leu Trp Gln Tyr Asp Val
930 935 940

Ser Ser Lys Gly Leu Lys Glu Ala Tyr Val Glu Leu Ala Val Thr Gln
945 950 955 960

Ile Ser Lys Leu Ala Thr Lys Tyr Asn Ala Val Val Val Val Glu Ser
965 970 975

Met Ser Ser Thr Phe Lys Asp Lys Phe Ser Phe Leu Asp Glu Gln Ile
980 985 990

Phe Lys Ala Phe Glu Ala Arg Leu Cys Ala Arg Met Ser Asp Leu Ser
995 1000 1005

Phe Asn Thr Ile Lys Glu Gly Glu Ala Gly Ser Ile Ser Asn Pro
1010 1015 1020

Ile Gln Val Ser Asn Asn Asn Gly Asn Ser Tyr Gln Asp Gly Val
1025 1030 1035

Ile Tyr Phe Leu Asn Asn Ala Tyr Thr Arg Thr Leu Cys Pro Asp
1040 1045 1050

Thr Gly Phe Val Asp Val Phe Asp Lys Thr Arg Leu Ile Thr Met
1055 1060 1065

Gln Ser Lys Arg Gln Phe Phe Ala Lys Met Lys Asp Ile Arg Ile
1070 1075 1080

Asp Asp Gly Glu Met Leu Phe Thr Phe Asn Leu Glu Glu Tyr Pro
1085 1090 1095

Thr Lys Arg Leu Leu Asp Arg Lys Glu Trp Thr Val Lys Ile Ala
1100 1105 1110

Gly Asp Gly Ser Tyr Phe Asp Lys Asp Lys Gly Glu Tyr Val Tyr
1115 1120 1125

Val Asn Asp Ile Val Arg Glu Gln Ile Ile Pro Ala Leu Leu Glu
1130 1135 1140

Asp Lys Ala Val Phe Asp Gly Asn Met Ala Glu Lys Phe Leu Asp
1145 1150 1155

Lys Thr Ala Ile Ser Gly Lys Ser Val Glu Leu Ile Tyr Lys Trp
1160 1165 1170

Phe Ala Asn Ala Leu Tyr Gly Ile Ile Thr Lys Lys Asp Gly Glu
1175 1180 1185

Lys Ile Tyr Arg Ser Pro Ile Thr Gly Thr Glu Ile Asp Val Ser
1190 1195 1200

Lys Asn Thr Thr Tyr Asn Phe Gly Lys Lys Phe Met Phe Lys Gln
1205 1210 1215

Glu Tyr Arg Gly Asp Gly Asp Phe Leu Asp Ala Phe Leu Asn Tyr
1220 1225 1230

Met Gln Ala Gln Asp Ile Ala Val
1235 1240

<210> 1254
<211> 4581
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1254
ggtaccatgt ccaacttctt taagaatttc accaacctgt atgagctgtc caagacactg 60

aggtttgagc tgaagcccgt gggcgacacc ctgacaaaca tgaaggacca cctggagtac 120

gatgagaagc tgcagacctt cctgaaggat cagaatatcg acgatgccta tcaggccctg 180

aagcctcagt tcgacgagat ccacgaggag tttatcacag attctctgga gagcaagaag 240

gccaaggaga tcgacttctc cgagtacctg gatctgtttc aggagaagaa ggagctgaac 300

gactctgaga agaagctgcg caacaagatc ggcgagacat tcaacaaggc cggcgagaag 360

tggaagaagg agaagtaccc tcagtatgag tggaagaagg gctccaagat cgccaatggc 420

gccgacatcc tgtcttgcca ggatatgctg cagtttatca agtataagaa cccagaggat 480

gagaagatca agaattacat cgacgataca ctgaagggct tctttaccta tttcggcggc 540

tttaatcaga acagggccaa ctactatgag acaaagaagg aggcctccac cgcagtggca 600

acaaggatcg tgcacgagaa cctgccaaag ttctgtgaca atgtgatcca gtttaagcac 660

atcatcaagc ggaagaagga tggcaccgtg gagaaaaccg agagaaagac cgagtacctg 720

aacgcctacc agtatctgaa gaacaataac aagatcacac agatcaagga cgccgagaca 780

gagaagatga tcgagtctac acccatcgcc gagaagatct tcgacgtgta ctacttcagc 840

agctgcctga gccagaagca gatcgaggag tacaaccgga tcatcggcca ctataatctg 900

ctgatcaacc tgtataacca ggccaagaga tctgagggca agcacctgag cgccaacgag 960

aagaagtata aggacctgcc taagttcaag accctgtata agcagatcgg ctgcggcaag 1020

aagaaggacc tgttttacac aatcaagtgt gataccgagg aggaggccaa taagtcccgg 1080

aacgagggca aggagtccca ctctgtggag gagatcatca acaaggccca ggaggccatc 1140

aataagtact tcaagtctaa taacgactgt gagaatatca acaccgtgcc cgacttcatc 1200

aactatatcc tgacaaagga gaattacgag ggcgtgtatt ggagcaaggc cgccatgaac 1260

accatctccg acaagtactt cgccaattat cacgacctgc aggatagact gaaggaggcc 1320

aaggtgtttc agaaggccga taagaagtcc gaggacgata tcaagatccc agaggccatc 1380

gagctgtctg gcctgttcgg cgtgctggac agcctggccg attggcagac cacactgttt 1440

aagtctagca tcctgagcaa cgaggacaag ctgaagatca tcacagattc ccagaccccc 1500

tctgaggccc tgctgaagat gatcttcaat gacatcgaga agaacatgga gtcctttctg 1560

aaggagacaa acgatatcat caccctgaag aagtataagg gcaataagga gggcaccgag 1620

aagatcaagc agtggttcga ctatacactg gccatcaacc ggatgctgaa gtactttctg 1680

gtgaaggaga ataagatcaa gggcaactcc ctggatacca atatctctga ggccctgaaa 1740

accctgatct acagcgacga tgccgagtgg ttcaagtggt acgacgccct gagaaactat 1800

ctgacccaga agcctcagga tgaggccaag gagaataagc tgaagctgaa tttcgacaac 1860

ccatctctgg ccggcggctg ggatgtgaac aaggagtgca gcaatttttg cgtgatcctg 1920

aaggacaaga acgagaagaa gtacctggcc atcatgaaga agggcgagaa taccctgttc 1980

cagaaggagt ggacagaggg ccggggcaag aacctgacaa agaagtctaa tccactgttc 2040

gagatcaata actgcgagat cctgagcaag atggagtatg acttttgggc cgacgtgagc 2100

aagatgatcc ccaagtgtag cacccagctg aaggccgtgg tgaaccactt caagcagtcc 2160

gacaatgagt tcatctttcc tatcggctac aaggtgacaa gcggcgagaa gtttagggag 2220

gagtgcaaga tctccaagca ggacttcgag ctgaataaca aggtgtttaa taagaacgag 2280

ctgagcgtga ccgccatgcg ctacgatctg tcctctacac aggagaagca gtatatcaag 2340

gccttccaga aggagtactg ggagctgctg tttaagcagg agaagcggga caccaagctg 2400

acaaataacg agatcttcaa cgagtggatc aatttttgca acaagaagta tagcgagctg 2460

ctgtcctggg agagaaagta caaggatgcc ctgaccaatt ggatcaactt ctgtaagtac 2520

tttctgagca agtatcccaa gaccacactg ttcaactact cttttaagga gagcgagaat 2580

tataactccc tggacgagtt ctaccgggac gtggatatct gttcttacaa gctgaatatc 2640

aacaccacaa tcaataagag catcctggat agactggtgg aggagggcaa gctgtacctg 2700

tttgagatca agaatcagga cagcaacgat ggcaagtcca tcggccacaa gaataacctg 2760

cacaccatct actggaacgc catcttcgag aattttgaca acaggcctaa gctgaatggc 2820

gaggccgaga tcttctatcg caaggccatc tccaaggata agctgggcat cgtgaagggc 2880

aagaaaacca agaacggcac cgagatcatc aagaattaca gattcagcaa ggagaagttt 2940

atcctgcacg tgccaatcac cctgaacttc tgctccaata acgagtatgt gaatgacatc 3000

gtgaacacaa agttctacaa tttttccaac ctgcactttc tgggcatcga taggggcgag 3060

aagcacctgg cctactattc tctggtgaat aagaacggcg agatcgtgga ccagggcaca 3120

ctgaacctgc ctttcaccga caaggatggc aatcagcgca gcatcaagaa ggagaagtac 3180

ttttataaca agcaggagga caagtgggag gccaaggagg tggattgttg gaattataac 3240

gacctgctgg atgccatggc ctctaaccgg gacatggcca gaaagaattg gcagaggatc 3300

ggcaccatca aggaggccaa gaacggctac gtgagcctgg tcatcaggaa gatcgccgat 3360

ctggccgtga ataacgagcg ccccgccttc atcgtgctgg aggacctgaa tacaggcttt 3420

aagcggtcca gacagaagat cgataagagc gtgtaccaga agttcgagct ggccctggcc 3480

aagaagctga actttctggt ggacaagaat gccaagcgcg atgagatcgg ctcccctaca 3540

aaggccctgc agctgacccc ccctgtgaat aactacggcg acattgagaa caagaagcag 3600

gccggcatca tgctgtatac ccgggccaat tatacctctc agacagatcc agccacaggc 3660

tggagaaaga ccatctatct gaaggccggc cccgaggaga caacatacaa gaaggacggc 3720

aagatcaaga acaagagcgt gaaggaccag atcatcgaga cattcaccga tatcggcttt 3780

gacggcaagg attactattt cgagtacgac aagggcgagt ttgtggatga gaaaaccggc 3840

gagatcaagc ccaagaagtg gcggctgtac tccggcgaga atggcaagtc cctggacagg 3900

ttccgcggag agagggagaa ggataagtat gagtggaaga tcgacaagat cgatatcgtg 3960

aagatcctgg acgatctgtt cgtgaatttt gacaagaaca tcagcctgct gaagcagctg 4020

aaggagggcg tggagctgac ccggaataac gagcacggca caggcgagtc cctgagattc 4080

gccatcaacc tgatccagca gatccggaat accggcaata acgagagaga caacgatttc 4140

atcctgtccc cagtgaggga cgagaatggc aagcactttg actctcgcga gtactgggat 4200

aaggagacaa agggcgagaa gatcagcatg cccagctccg gcgatgccaa tggcgccttc 4260

aacatcgccc ggaagggcat catcatgaac gcccacatcc tggccaatag cgactccaag 4320

gatctgtccc tgttcgtgtc tgacgaggag tgggatctgc acctgaataa caagaccgag 4380

tggaagaagc agctgaacat cttttctagc aggaaggcca tggccaagcg caagaagaaa 4440

aggccggcgg ccacgaaaaa ggccggccag gcaaaaaaga aaaagggatc ctacccatac 4500

gatgttccag attacgctta tccctacgac gtgcctgatt atgcataccc atatgatgtc 4560

cccgactatg cctaagaatt c 4581

<210> 1255
<211> 1477
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 1255
Met Ser Asn Phe Phe Lys Asn Phe Thr Asn Leu Tyr Glu Leu Ser Lys
1 5 10 15

Thr Leu Arg Phe Glu Leu Lys Pro Val Gly Asp Thr Leu Thr Asn Met
20 25 30

Lys Asp His Leu Glu Tyr Asp Glu Lys Leu Gln Thr Phe Leu Lys Asp
35 40 45

Gln Asn Ile Asp Asp Ala Tyr Gln Ala Leu Lys Pro Gln Phe Asp Glu
50 55 60

Ile His Glu Glu Phe Ile Thr Asp Ser Leu Glu Ser Lys Lys Ala Lys
65 70 75 80

Glu Ile Asp Phe Ser Glu Tyr Leu Asp Leu Phe Gln Glu Lys Lys Glu
85 90 95

Leu Asn Asp Ser Glu Lys Lys Leu Arg Asn Lys Ile Gly Glu Thr Phe
100 105 110

```
Asn Lys Ala Gly Glu Lys Trp Lys Lys Glu Lys Tyr Pro Gln Tyr Glu
        115                 120                 125

Trp Lys Lys Gly Ser Lys Ile Ala Asn Gly Ala Asp Ile Leu Ser Cys
    130                 135                 140

Gln Asp Met Leu Gln Phe Ile Lys Tyr Lys Asn Pro Glu Asp Glu Lys
145                 150                 155                 160

Ile Lys Asn Tyr Ile Asp Asp Thr Leu Lys Gly Phe Phe Thr Tyr Phe
                165                 170                 175

Gly Gly Phe Asn Gln Asn Arg Ala Asn Tyr Tyr Glu Thr Lys Lys Glu
            180                 185                 190

Ala Ser Thr Ala Val Ala Thr Arg Ile Val His Glu Asn Leu Pro Lys
        195                 200                 205

Phe Cys Asp Asn Val Ile Gln Phe Lys His Ile Ile Lys Arg Lys Lys
    210                 215                 220

Asp Gly Thr Val Glu Lys Thr Glu Arg Lys Thr Glu Tyr Leu Asn Ala
225                 230                 235                 240

Tyr Gln Tyr Leu Lys Asn Asn Asn Lys Ile Thr Gln Ile Lys Asp Ala
                245                 250                 255

Glu Thr Glu Lys Met Ile Glu Ser Thr Pro Ile Ala Glu Lys Ile Phe
            260                 265                 270

Asp Val Tyr Tyr Phe Ser Ser Cys Leu Ser Gln Lys Gln Ile Glu Glu
        275                 280                 285

Tyr Asn Arg Ile Ile Gly His Tyr Asn Leu Leu Ile Asn Leu Tyr Asn
    290                 295                 300

Gln Ala Lys Arg Ser Glu Gly Lys His Leu Ser Ala Asn Glu Lys Lys
305                 310                 315                 320

Tyr Lys Asp Leu Pro Lys Phe Lys Thr Leu Tyr Lys Gln Ile Gly Cys
                325                 330                 335

Gly Lys Lys Lys Asp Leu Phe Tyr Thr Ile Lys Cys Asp Thr Glu Glu
            340                 345                 350

Glu Ala Asn Lys Ser Arg Asn Glu Gly Lys Glu Ser His Ser Val Glu
        355                 360                 365

Glu Ile Ile Asn Lys Ala Gln Glu Ala Ile Asn Lys Tyr Phe Lys Ser
    370                 375                 380

Asn Asn Asp Cys Glu Asn Ile Asn Thr Val Pro Asp Phe Ile Asn Tyr
385                 390                 395                 400

Ile Leu Thr Lys Glu Asn Tyr Glu Gly Val Tyr Trp Ser Lys Ala Ala
                405                 410                 415

Met Asn Thr Ile Ser Asp Lys Tyr Phe Ala Asn Tyr His Asp Leu Gln
            420                 425                 430

Asp Arg Leu Lys Glu Ala Lys Val Phe Gln Lys Ala Asp Lys Lys Ser
        435                 440                 445
```

```
Glu Asp Asp Ile Lys Ile Pro Glu Ala Ile Glu Leu Ser Gly Leu Phe
    450                 455                 460

Gly Val Leu Asp Ser Leu Ala Asp Trp Gln Thr Thr Leu Phe Lys Ser
465                 470                 475                 480

Ser Ile Leu Ser Asn Glu Asp Lys Leu Lys Ile Ile Thr Asp Ser Gln
                485                 490                 495

Thr Pro Ser Glu Ala Leu Leu Lys Met Ile Phe Asn Asp Ile Glu Lys
            500                 505                 510

Asn Met Glu Ser Phe Leu Lys Glu Thr Asn Asp Ile Ile Thr Leu Lys
        515                 520                 525

Lys Tyr Lys Gly Asn Lys Glu Gly Thr Glu Lys Ile Lys Gln Trp Phe
    530                 535                 540

Asp Tyr Thr Leu Ala Ile Asn Arg Met Leu Lys Tyr Phe Leu Val Lys
545                 550                 555                 560

Glu Asn Lys Ile Lys Gly Asn Ser Leu Asp Thr Asn Ile Ser Glu Ala
                565                 570                 575

Leu Lys Thr Leu Ile Tyr Ser Asp Asp Ala Glu Trp Phe Lys Trp Tyr
            580                 585                 590

Asp Ala Leu Arg Asn Tyr Leu Thr Gln Lys Pro Gln Asp Glu Ala Lys
        595                 600                 605

Glu Asn Lys Leu Lys Leu Asn Phe Asp Asn Pro Ser Leu Ala Gly Gly
    610                 615                 620

Trp Asp Val Asn Lys Glu Cys Ser Asn Phe Cys Val Ile Leu Lys Asp
625                 630                 635                 640

Lys Asn Glu Lys Lys Tyr Leu Ala Ile Met Lys Lys Gly Glu Asn Thr
                645                 650                 655

Leu Phe Gln Lys Glu Trp Thr Glu Gly Arg Gly Lys Asn Leu Thr Lys
            660                 665                 670

Lys Ser Asn Pro Leu Phe Glu Ile Asn Asn Cys Glu Ile Leu Ser Lys
        675                 680                 685

Met Glu Tyr Asp Phe Trp Ala Asp Val Ser Lys Met Ile Pro Lys Cys
    690                 695                 700

Ser Thr Gln Leu Lys Ala Val Val Asn His Phe Lys Gln Ser Asp Asn
705                 710                 715                 720

Glu Phe Ile Phe Pro Ile Gly Tyr Lys Val Thr Ser Gly Glu Lys Phe
                725                 730                 735

Arg Glu Glu Cys Lys Ile Ser Lys Gln Asp Phe Glu Leu Asn Asn Lys
            740                 745                 750

Val Phe Asn Lys Asn Glu Leu Ser Val Thr Ala Met Arg Tyr Asp Leu
        755                 760                 765

Ser Ser Thr Gln Glu Lys Gln Tyr Ile Lys Ala Phe Gln Lys Glu Tyr
    770                 775                 780
```

```
Trp Glu Leu Leu Phe Lys Gln Glu Lys Arg Asp Thr Lys Leu Thr Asn
785                 790                 795                 800

Asn Glu Ile Phe Asn Glu Trp Ile Asn Phe Cys Asn Lys Lys Tyr Ser
                805                 810                 815

Glu Leu Leu Ser Trp Glu Arg Lys Tyr Lys Asp Ala Leu Thr Asn Trp
            820                 825                 830

Ile Asn Phe Cys Lys Tyr Phe Leu Ser Lys Tyr Pro Lys Thr Thr Leu
        835                 840                 845

Phe Asn Tyr Ser Phe Lys Glu Ser Glu Asn Tyr Asn Ser Leu Asp Glu
    850                 855                 860

Phe Tyr Arg Asp Val Asp Ile Cys Ser Tyr Lys Leu Asn Ile Asn Thr
865                 870                 875                 880

Thr Ile Asn Lys Ser Ile Leu Asp Arg Leu Val Glu Glu Gly Lys Leu
                885                 890                 895

Tyr Leu Phe Glu Ile Lys Asn Gln Asp Ser Asn Asp Gly Lys Ser Ile
            900                 905                 910

Gly His Lys Asn Asn Leu His Thr Ile Tyr Trp Asn Ala Ile Phe Glu
        915                 920                 925

Asn Phe Asp Asn Arg Pro Lys Leu Asn Gly Glu Ala Glu Ile Phe Tyr
    930                 935                 940

Arg Lys Ala Ile Ser Lys Asp Lys Leu Gly Ile Val Lys Gly Lys Lys
945                 950                 955                 960

Thr Lys Asn Gly Thr Glu Ile Ile Lys Asn Tyr Arg Phe Ser Lys Glu
                965                 970                 975

Lys Phe Ile Leu His Val Pro Ile Thr Leu Asn Phe Cys Ser Asn Asn
            980                 985                 990

Glu Tyr Val Asn Asp Ile Val Asn Thr Lys Phe Tyr Asn Phe Ser Asn
        995                 1000                1005

Leu His  Phe Leu Gly Ile Asp Arg Gly Glu Lys  His Leu Ala Tyr
    1010                1015                1020

Tyr Ser  Leu Val Asn Lys Asn Gly Glu Ile Val  Asp Gln Gly Thr
    1025                1030                1035

Leu Asn  Leu Pro Phe Thr Asp Lys Asp Gly Asn  Gln Arg Ser Ile
    1040                1045                1050

Lys Lys  Glu Lys Tyr Phe Tyr Asn Lys Gln Glu  Asp Lys Trp Glu
    1055                1060                1065

Ala Lys  Glu Val Asp Cys Trp Asn Tyr Asn Asp  Leu Leu Asp Ala
    1070                1075                1080

Met Ala  Ser Asn Arg Asp Met Ala Arg Lys Asn  Trp Gln Arg Ile
    1085                1090                1095

Gly Thr  Ile Lys Glu Ala Lys Asn Gly Tyr Val  Ser Leu Val Ile
    1100                1105                1110
```

```
Arg Lys  Ile Ala Asp Leu Ala  Val Asn Asn Glu Arg  Pro Ala Phe
    1115                 1120                 1125

Ile Val  Leu Glu Asp Leu Asn  Thr Gly Phe Lys Arg  Ser Arg Gln
    1130                 1135                 1140

Lys Ile  Asp Lys Ser Val Tyr  Gln Lys Phe Glu Leu  Ala Leu Ala
    1145                 1150                 1155

Lys Lys  Leu Asn Phe Leu Val  Asp Lys Asn Ala Lys  Arg Asp Glu
    1160                 1165                 1170

Ile Gly  Ser Pro Thr Lys Ala  Leu Gln Leu Thr Pro  Pro Val Asn
    1175                 1180                 1185

Asn Tyr  Gly Asp Ile Glu Asn  Lys Lys Gln Ala Gly  Ile Met Leu
    1190                 1195                 1200

Tyr Thr  Arg Ala Asn Tyr Thr  Ser Gln Thr Asp Pro  Ala Thr Gly
    1205                 1210                 1215

Trp Arg  Lys Thr Ile Tyr Leu  Lys Ala Gly Pro Glu  Glu Thr Thr
    1220                 1225                 1230

Tyr Lys  Lys Asp Gly Lys Ile  Lys Asn Lys Ser Val  Lys Asp Gln
    1235                 1240                 1245

Ile Ile  Glu Thr Phe Thr Asp  Ile Gly Phe Asp Gly  Lys Asp Tyr
    1250                 1255                 1260

Tyr Phe  Glu Tyr Asp Lys Gly  Glu Phe Val Asp Glu  Lys Thr Gly
    1265                 1270                 1275

Glu Ile  Lys Pro Lys Lys Trp  Arg Leu Tyr Ser Gly  Glu Asn Gly
    1280                 1285                 1290

Lys Ser  Leu Asp Arg Phe Arg  Gly Glu Arg Glu Lys  Asp Lys Tyr
    1295                 1300                 1305

Glu Trp  Lys Ile Asp Lys Ile  Asp Ile Val Lys Ile  Leu Asp Asp
    1310                 1315                 1320

Leu Phe  Val Asn Phe Asp Lys  Asn Ile Ser Leu Leu  Lys Gln Leu
    1325                 1330                 1335

Lys Glu  Gly Val Glu Leu Thr  Arg Asn Asn Glu His  Gly Thr Gly
    1340                 1345                 1350

Glu Ser  Leu Arg Phe Ala Ile  Asn Leu Ile Gln Gln  Ile Arg Asn
    1355                 1360                 1365

Thr Gly  Asn Asn Glu Arg Asp  Asn Asp Phe Ile Leu  Ser Pro Val
    1370                 1375                 1380

Arg Asp  Glu Asn Gly Lys His  Phe Asp Ser Arg Glu  Tyr Trp Asp
    1385                 1390                 1395

Lys Glu  Thr Lys Gly Glu Lys  Ile Ser Met Pro Ser  Ser Gly Asp
    1400                 1405                 1410

Ala Asn  Gly Ala Phe Asn Ile  Ala Arg Lys Gly Ile  Ile Met Asn
    1415                 1420                 1425
```

Ala His Ile Leu Ala Asn Ser Asp Ser Lys Asp Leu Ser Leu Phe
1430 1435 1440

Val Ser Asp Glu Glu Trp Asp Leu His Leu Asn Asn Lys Thr Glu
1445 1450 1455

Trp Lys Lys Gln Leu Asn Ile Phe Ser Ser Arg Lys Ala Met Ala
1460 1465 1470

Lys Arg Lys Lys
1475

<210> 1256
<211> 4206
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1256
ggtaccatgg agaacatctt cgaccagttt atcggcaagt acagcctgtc caagaccctg 60

agattcgagc tgaagcccgt gggcaagaca gaggacttcc tgaagatcaa caaggtgttt 120

gagaaggatc agaccatcga cgatagctac aatcaggcca agttctattt tgattccctg 180

caccagaagt ttatcgacgc cgccctggcc tccgataaga catccgagct gtctttccag 240

aactttgccg acgtgctgga gaagcagaat aagatcatcc tggataagaa gagagagatg 300

ggcgccctga ggaagcgcga caagaacgcc gtgggcatcg ataggctgca gaaggagatc 360

aatgacgccg aggatatcat ccagaaggag aaggagaaga tctacaagga cgtgcgcacc 420

ctgttcgata acgaggccga gtcttggaaa acctactatc aggagcggga ggtggacggc 480

aagaagatca ccttcagcaa ggccgacctg aagcagaagg gcgccgattt tctgacagcc 540

gccggcatcc tgaaggtgct gaagtatgag ttccccgagg agaaggagaa ggagtttcag 600

gccaagaacc agccctccct gttcgtggag gagaaggaga atcctggcca gaagaggtac 660

atcttcgact cttttgataa gttcgccggc tatctgacca agtttcagca gacaaagaag 720

aatctgtacg cagcagacgg caccagcaca gcagtggcca cccgcatcgc cgataacttt 780

atcatcttcc accagaatac caaggtgttc cgggacaagt acaagaacaa tcacacagac 840

ctgggcttcg atgaggagaa catctttgag atcgagaggt ataagaattg cctgctgcag 900

cgcgagatcg agcacatcaa gaatgagaat agctacaaca agatcatcgg ccggatcaat 960

aagaagatca aggagtatcg ggaccagaag gccaaggata ccaagctgac aaagtccgac 1020

ttcccttttct ttaagaacct ggataagcag atcctgggcg aggtggagaa ggagaagcag 1080

ctgatcgaga aaacccggga gaaaaccgag gaggacgtgc tgatcgagcg gttcaaggag 1140

ttcatcgaga acaatgagga gaggttcacc gccgccaaga agctgatgaa tgccttctgt 1200

aacggcgagt ttgagtccga gtacgagggc atctatctga agaataaggc catcaacaca 1260

atctcccgga gatggttcgt gtctgacaga gattttgagc tgaagctgcc tcagcagaag 1320

tccaagaaca agtctgagaa gaatgagcca aaggtgaaga agttcatctc catcgccgag 1380

atcaagaacg ccgtggagga gctggacggc gatatcttta aggccgtgtt ctacgacaag 1440

aagatcatcg cccagggcgg ctctaagctg gagcagttcc tggtcatctg gaagtacgag 1500

tttgagtatc tgttccggga catcgagaga gagaacggcg agaagctgct gggctatgat 1560

agctgcctga agatcgccaa gcagctgggc atcttcccac aggagaagga ggcccgcgag 1620

aaggcaaccg ccgtgatcaa gaattacgcc gacgccggcc tgggcatctt ccagatgatg 1680

aagtattttt ctctggacga taaggatcgg aagaacaccc ccggccagct gagcacaaat 1740

ttctacgccg agtatgacgg ctactacaag gatttcgagt ttatcaagta ctacaacgag 1800

tttaggaact tcatcaccaa gaagcctttc gacgaggata agatcaagct gaactttgag 1860

aatggcgccc tgctgaaggg ctgggacgag aacaaggagt acgatttcat gggcgtgatc 1920

ctgaagaagg agggccgcct gtatctgggc atcatgcaca agaaccaccg gaagctgttt 1980

cagtccatgg gcaatgccaa gggcgacaac gccaatagat accagaagat gatctataag 2040

cagatcgccg acgcctctaa ggatgtgccc aggctgctgc tgaccagcaa gaaggccatg 2100

gagaagttca agccttccca ggagatcctg agaatcaaga aggagaaaac cttcaagcgg 2160

gagagcaaga acttttccct gagagatctg cacgccctga tcgagtacta taggaactgc 2220

atccctcagt acagcaattg gtccttttat gacttccagt ttcaggatac cggcaagtac 2280

cagaatatca aggagttcac agacgatgtg cagaagtacg gctataagat ctcctttcgc 2340

gacatcgacg atgagtatat caatcaggcc ctgaacgagg gcaagatgta cctgttcgag 2400

gtggtgaaca aggatatcta taacaccaag aatggctcca agaatctgca cacactgtac 2460

tttgagcaca tcctgtctgc cgagaacctg aatgacccag tgttcaagct gtctggcatg 2520

gccgagatct ttcagcggca gcccagcgtg aacgaaagag agaagatcac cacacagaag 2580

aatcagtgta tcctggacaa gggcgataga gcctacaagt ataggcgcta caccgagaag 2640

aagatcatgt tccacatgag cctggtgctg aacacaggca agggcgagat caagcaggtg 2700

cagtttaata agatcatcaa ccagaggatc agctcctctg acaacgagat gagggtgaat 2760

gtgatcggca tcgatcgcgg cgagaagaac ctgctgtact atagcgtggt gaagcagaat 2820

ggcgagatca tcgagcaggc ctccctgaac gagatcaatg gcgtgaacta ccgggacaag 2880

ctgatcgaga gggagaagga gcgcctgaag aaccggcaga gctggaagcc tgtggtgaag 2940

atcaaggatc tgaagaaggg ctacatctcc cacgtgatcc acaagatctg ccagctgatc 3000

gagaagtatt ctgccatcgt ggtgctggag gacctgaata tgagattcaa gcagatcagg 3060

ggaggaatcg agcggagcgt gtaccagcag ttcgagaagg ccctgatcga taagctgggc 3120

tatctggtgt ttaaggacaa cagggatctg agggcaccag gaggcgtgct gaatggctac 3180

cagctgtctg ccccctttgt gagcttcgag aagatgcgca agcagaccgg catcctgttc 3240

tacacacagg ccgagtatac cagcaagaca gacccaatca ccggctttcg gaagaacgtg 3300

tatatctcta atagcgcctc cctggataag atcaaggagg ccgtgaagaa gttcgacgcc 3360

atcggctggg atggcaagga gcagtcttac ttctttaagt acaaccctta caacctggcc 3420

gacgagaagt ataagaactc taccgtgagc aaggagtggg ccatctttgc cagcgcccca 3480

agaatccgga gacagaaggg cgaggacggc tactggaagt atgatagggt gaaagtgaat 3540

gaggagttcg agaagctgct gaaggtctgg aattttgtga acccaaaggc cacagatatc 3600

aagcaggaga tcatcaagaa ggagaaggca ggcgacctgc agggagagaa ggagctggat 3660

ggccggctga gaaacttttg gcactctttc atctacctgt ttaacctggt gctggagctg 3720

cgcaattctt tcagcctgca gatcaagatc aaggcaggag aagtgatcgc agtggacgag 3780

ggcgtggact tcatcgccag cccagtgaag cccttcttta ccacacccaa cccttacatc 3840

ccctccaacc tgtgctggct ggccgtggag aatgcagacg caaacggagc ctataatatc 3900

gccaggaagg gcgtgatgat cctgaagaag atccgcgagc acgccaagaa ggaccccgag 3960

ttcaagaagc tgccaaacct gtttatcagc aatgcagagt gggacgaggc agcccgggat 4020

tggggcaagt acgcaggcac cacagccctg aacctggacc acaaaaggcc ggcggccacg 4080

aaaaaggccg gccaggcaaa aaagaaaaag ggatcctacc catacgatgt tccagattac 4140

gcttatccct acgacgtgcc tgattatgca tacccatatg atgtccccga ctatgcctaa 4200

gaattc 4206

<210> 1257
<211> 1352
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 1257

```
Met Glu Asn Ile Phe Asp Gln Phe Ile Gly Lys Tyr Ser Leu Ser Lys
1               5                  10                  15

Thr Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr Glu Asp Phe Leu
            20                  25                  30

Lys Ile Asn Lys Val Phe Glu Lys Asp Gln Thr Ile Asp Asp Ser Tyr
        35                  40                  45

Asn Gln Ala Lys Phe Tyr Phe Asp Ser Leu His Gln Lys Phe Ile Asp
    50                  55                  60

Ala Ala Leu Ala Ser Asp Lys Thr Ser Glu Leu Ser Phe Gln Asn Phe
65                  70                  75                  80

Ala Asp Val Leu Glu Lys Gln Asn Lys Ile Ile Leu Asp Lys Lys Arg
                85                  90                  95

Glu Met Gly Ala Leu Arg Lys Arg Asp Lys Asn Ala Val Gly Ile Asp
            100                 105                 110

Arg Leu Gln Lys Glu Ile Asn Asp Ala Glu Asp Ile Ile Gln Lys Glu
        115                 120                 125

Lys Glu Lys Ile Tyr Lys Asp Val Arg Thr Leu Phe Asp Asn Glu Ala
    130                 135                 140

Glu Ser Trp Lys Thr Tyr Tyr Gln Glu Arg Glu Val Asp Gly Lys Lys
145                 150                 155                 160

Ile Thr Phe Ser Lys Ala Asp Leu Lys Gln Lys Gly Ala Asp Phe Leu
                165                 170                 175

Thr Ala Ala Gly Ile Leu Lys Val Leu Lys Tyr Glu Phe Pro Glu Glu
            180                 185                 190

Lys Glu Lys Glu Phe Gln Ala Lys Asn Gln Pro Ser Leu Phe Val Glu
        195                 200                 205

Glu Lys Glu Asn Pro Gly Gln Lys Arg Tyr Ile Phe Asp Ser Phe Asp
    210                 215                 220

Lys Phe Ala Gly Tyr Leu Thr Lys Phe Gln Gln Thr Lys Lys Asn Leu
225                 230                 235                 240

Tyr Ala Ala Asp Gly Thr Ser Thr Ala Val Ala Thr Arg Ile Ala Asp
                245                 250                 255

Asn Phe Ile Ile Phe His Gln Asn Thr Lys Val Phe Arg Asp Lys Tyr
            260                 265                 270

Lys Asn Asn His Thr Asp Leu Gly Phe Asp Glu Glu Asn Ile Phe Glu
        275                 280                 285

Ile Glu Arg Tyr Lys Asn Cys Leu Leu Gln Arg Glu Ile Glu His Ile
```

290 295 300

Lys Asn Glu Asn Ser Tyr Asn Lys Ile Ile Gly Arg Ile Asn Lys Lys
305 310 315 320

Ile Lys Glu Tyr Arg Asp Gln Lys Ala Lys Asp Thr Lys Leu Thr Lys
325 330 335

Ser Asp Phe Pro Phe Phe Lys Asn Leu Asp Lys Gln Ile Leu Gly Glu
340 345 350

Val Glu Lys Glu Lys Gln Leu Ile Glu Lys Thr Arg Glu Lys Thr Glu
355 360 365

Glu Asp Val Leu Ile Glu Arg Phe Lys Glu Phe Ile Glu Asn Asn Glu
370 375 380

Glu Arg Phe Thr Ala Ala Lys Lys Leu Met Asn Ala Phe Cys Asn Gly
385 390 395 400

Glu Phe Glu Ser Glu Tyr Glu Gly Ile Tyr Leu Lys Asn Lys Ala Ile
405 410 415

Asn Thr Ile Ser Arg Arg Trp Phe Val Ser Asp Arg Asp Phe Glu Leu
420 425 430

Lys Leu Pro Gln Gln Lys Ser Lys Asn Lys Ser Glu Lys Asn Glu Pro
435 440 445

Lys Val Lys Lys Phe Ile Ser Ile Ala Glu Ile Lys Asn Ala Val Glu
450 455 460

Glu Leu Asp Gly Asp Ile Phe Lys Ala Val Phe Tyr Asp Lys Lys Ile
465 470 475 480

Ile Ala Gln Gly Gly Ser Lys Leu Glu Gln Phe Leu Val Ile Trp Lys
485 490 495

Tyr Glu Phe Glu Tyr Leu Phe Arg Asp Ile Glu Arg Glu Asn Gly Glu
500 505 510

Lys Leu Leu Gly Tyr Asp Ser Cys Leu Lys Ile Ala Lys Gln Leu Gly
515 520 525

Ile Phe Pro Gln Glu Lys Glu Ala Arg Glu Lys Ala Thr Ala Val Ile
530 535 540

Lys Asn Tyr Ala Asp Ala Gly Leu Gly Ile Phe Gln Met Met Lys Tyr
545 550 555 560

Phe Ser Leu Asp Asp Lys Asp Arg Lys Asn Thr Pro Gly Gln Leu Ser
565 570 575

Thr Asn Phe Tyr Ala Glu Tyr Asp Gly Tyr Tyr Lys Asp Phe Glu Phe
580 585 590

Ile Lys Tyr Tyr Asn Glu Phe Arg Asn Phe Ile Thr Lys Lys Pro Phe
595 600 605

Asp Glu Asp Lys Ile Lys Leu Asn Phe Glu Asn Gly Ala Leu Leu Lys
610 615 620

Gly Trp Asp Glu Asn Lys Glu Tyr Asp Phe Met Gly Val Ile Leu Lys

625 630 635 640

Lys Glu Gly Arg Leu Tyr Leu Gly Ile Met His Lys Asn His Arg Lys
645 650 655

Leu Phe Gln Ser Met Gly Asn Ala Lys Gly Asp Asn Ala Asn Arg Tyr
660 665 670

Gln Lys Met Ile Tyr Lys Gln Ile Ala Asp Ala Ser Lys Asp Val Pro
675 680 685

Arg Leu Leu Leu Thr Ser Lys Lys Ala Met Glu Lys Phe Lys Pro Ser
690 695 700

Gln Glu Ile Leu Arg Ile Lys Lys Glu Lys Thr Phe Lys Arg Glu Ser
705 710 715 720

Lys Asn Phe Ser Leu Arg Asp Leu His Ala Leu Ile Glu Tyr Tyr Arg
725 730 735

Asn Cys Ile Pro Gln Tyr Ser Asn Trp Ser Phe Tyr Asp Phe Gln Phe
740 745 750

Gln Asp Thr Gly Lys Tyr Gln Asn Ile Lys Glu Phe Thr Asp Asp Val
755 760 765

Gln Lys Tyr Gly Tyr Lys Ile Ser Phe Arg Asp Ile Asp Asp Glu Tyr
770 775 780

Ile Asn Gln Ala Leu Asn Glu Gly Lys Met Tyr Leu Phe Glu Val Val
785 790 795 800

Asn Lys Asp Ile Tyr Asn Thr Lys Asn Gly Ser Lys Asn Leu His Thr
805 810 815

Leu Tyr Phe Glu His Ile Leu Ser Ala Glu Asn Leu Asn Asp Pro Val
820 825 830

Phe Lys Leu Ser Gly Met Ala Glu Ile Phe Gln Arg Gln Pro Ser Val
835 840 845

Asn Glu Arg Glu Lys Ile Thr Thr Gln Lys Asn Gln Cys Ile Leu Asp
850 855 860

Lys Gly Asp Arg Ala Tyr Lys Tyr Arg Arg Tyr Thr Glu Lys Lys Ile
865 870 875 880

Met Phe His Met Ser Leu Val Leu Asn Thr Gly Lys Gly Glu Ile Lys
885 890 895

Gln Val Gln Phe Asn Lys Ile Ile Asn Gln Arg Ile Ser Ser Ser Asp
900 905 910

Asn Glu Met Arg Val Asn Val Ile Gly Ile Asp Arg Gly Glu Lys Asn
915 920 925

Leu Leu Tyr Tyr Ser Val Val Lys Gln Asn Gly Glu Ile Ile Glu Gln
930 935 940

Ala Ser Leu Asn Glu Ile Asn Gly Val Asn Tyr Arg Asp Lys Leu Ile
945 950 955 960

Glu Arg Glu Lys Glu Arg Leu Lys Asn Arg Gln Ser Trp Lys Pro Val

965 970 975

Val Lys Ile Lys Asp Leu Lys Lys Gly Tyr Ile Ser His Val Ile His
980 985 990

Lys Ile Cys Gln Leu Ile Glu Lys Tyr Ser Ala Ile Val Val Leu Glu
995 1000 1005

Asp Leu Asn Met Arg Phe Lys Gln Ile Arg Gly Gly Ile Glu Arg
1010 1015 1020

Ser Val Tyr Gln Gln Phe Glu Lys Ala Leu Ile Asp Lys Leu Gly
1025 1030 1035

Tyr Leu Val Phe Lys Asp Asn Arg Asp Leu Arg Ala Pro Gly Gly
1040 1045 1050

Val Leu Asn Gly Tyr Gln Leu Ser Ala Pro Phe Val Ser Phe Glu
1055 1060 1065

Lys Met Arg Lys Gln Thr Gly Ile Leu Phe Tyr Thr Gln Ala Glu
1070 1075 1080

Tyr Thr Ser Lys Thr Asp Pro Ile Thr Gly Phe Arg Lys Asn Val
1085 1090 1095

Tyr Ile Ser Asn Ser Ala Ser Leu Asp Lys Ile Lys Glu Ala Val
1100 1105 1110

Lys Lys Phe Asp Ala Ile Gly Trp Asp Gly Lys Glu Gln Ser Tyr
1115 1120 1125

Phe Phe Lys Tyr Asn Pro Tyr Asn Leu Ala Asp Glu Lys Tyr Lys
1130 1135 1140

Asn Ser Thr Val Ser Lys Glu Trp Ala Ile Phe Ala Ser Ala Pro
1145 1150 1155

Arg Ile Arg Arg Gln Lys Gly Glu Asp Gly Tyr Trp Lys Tyr Asp
1160 1165 1170

Arg Val Lys Val Asn Glu Glu Phe Glu Lys Leu Leu Lys Val Trp
1175 1180 1185

Asn Phe Val Asn Pro Lys Ala Thr Asp Ile Lys Gln Glu Ile Ile
1190 1195 1200

Lys Lys Glu Lys Ala Gly Asp Leu Gln Gly Glu Lys Glu Leu Asp
1205 1210 1215

Gly Arg Leu Arg Asn Phe Trp His Ser Phe Ile Tyr Leu Phe Asn
1220 1225 1230

Leu Val Leu Glu Leu Arg Asn Ser Phe Ser Leu Gln Ile Lys Ile
1235 1240 1245

Lys Ala Gly Glu Val Ile Ala Val Asp Glu Gly Val Asp Phe Ile
1250 1255 1260

Ala Ser Pro Val Lys Pro Phe Phe Thr Thr Pro Asn Pro Tyr Ile
1265 1270 1275

Pro Ser Asn Leu Cys Trp Leu Ala Val Glu Asn Ala Asp Ala Asn

1280 1285 1290

Gly Ala Tyr Asn Ile Ala Arg Lys Gly Val Met Ile Leu Lys Lys
1295 1300 1305

Ile Arg Glu His Ala Lys Lys Asp Pro Glu Phe Lys Lys Leu Pro
1310 1315 1320

Asn Leu Phe Ile Ser Asn Ala Glu Trp Asp Glu Ala Ala Arg Asp
1325 1330 1335

Trp Gly Lys Tyr Ala Gly Thr Thr Ala Leu Asn Leu Asp His
1340 1345 1350

<210> 1258
<211> 3900
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1258
ggtaccatgc agaccctgtt tgagaacttc acaaatcagt acccagtgtc caagaccctg 60

cgctttgagc tgatccccca gggcaagaca aaggacttca tcgagcagaa gggcctgctg 120

aagaaggatg aggaccgggc cgagaagtat aagaaggtga agaacatcat cgatgagtac 180

cacaaggact tcatcgagaa gtctctgaat ggcctgaagc tggacggcct ggagaagtac 240

aagaccctgt atctgaagca ggagaaggac gataaggata agaaggcctt tgacaaggag 300

aaggagaacc tgcgcaagca gatcgccaat gccttccgga acaatgagaa gtttaagaca 360

ctgttcgcca aggagctgat caagaacgat ctgatgtctt tcgcctgcga ggaggacaag 420

aagaatgtga aggagtttga ggccttcacc acatacttca ccggcttcca ccagaaccgc 480

gccaatatgt acgtggccga tgagaagaga acagccatcg ccagcaggct gatccacgag 540

aacctgccaa agtttatcga caatatcaag atcttcgaga agatgaagaa ggaggccccc 600

gagctgctgt ctcctttcaa ccagaccctg aaggatatga aggacgtgat caagggcacc 660

acactggagg agatctttag cctggattat ttcaacaaga ccctgacaca gagcggcatc 720

gacatctaca attccgtgat cggcggcaga acccctgagg agggcaagac aaagatcaag 780

ggcctgaacg agtacatcaa taccgacttc aaccagaagc agacagacaa gaagaagcgg 840

cagccaaagt tcaagcagct gtataagcag atcctgagcg ataggcagag cctgtccttt 900

atcgccgagg ccttcaagaa cgacaccgag atcctggagg ccatcgagaa gttttacgtg 960

aatgagctgc tgcacttcag caatgagggc aagtccacaa acgtgctgga cgccatcaag 1020

aatgccgtgt ctaacctgga gagctttaac ctgaccaaga tgtatttccg ctccggcgcc 1080

tctctgacag acgtgagccg gaaggtgttt ggcgagtgga gcatcatcaa tagagccctg 1140

gacaactact atgccaccac atatccaatc aagcccagag agaagtctga gaagtacgag 1200

gagaggaagg agaagtggct gaagcaggac ttcaacgtga gcctgatcca gaccgccatc 1260

gatgagtacg acaacgagac agtgaagggc aagaacagcg gcaaagtgat cgccgattat 1320

tttgccaagt tctgcgacga taaggagaca gacctgatcc agaaggtgaa cgagggctac 1380

atcgccgtga aggatctgct gaatacaccc tgtcctgaga acgagaagct gggcagcaat 1440

aaggaccagg tgaagcagat caaggccttt atggattcta tcatggacat catgcacttc 1500

gtgcgccccc tgagcctgaa ggataccgac aaggagaagg atgagacatt ctactccctg 1560

ttcacacctc tgtacgacca cctgacccag acaatcgccc tgtataacaa ggtgcggaac 1620

tatctgaccc agaagcctta cagcacagag aagatcaagc tgaacttcga gaacagcacc 1680

ctgctgggcg gctgggatct gaataaggag acagacaaca cagccatcat cctgaggaag 1740

gataacctgt actatctggg catcatggac aagaggcaca atcgcatctt tcggaacgtg 1800

cccaaggccg ataagaagga cttctgctac gagaagatgg tgtataagct gctgcctggc 1860

gccaacaaga tgctgccaaa ggtgttcttt tctcagagca gaatccagga gtttacccct 1920

tccgccaagc tgctggagaa ctacgccaat gagacacaca agaagggcga taatttcaac 1980

ctgaatcact gtcacaagct gatcgatttc tttaaggact ctatcaacaa gcacgaggat 2040

tggaagaatt tcgactttag gttcagcgcc acctccacct acgccgacct gagcggcttt 2100

taccacgagg tggagcacca gggctacaag atctcttttc agagcgtggc cgattccttc 2160

atcgacgatc tggtgaacga gggcaagctg tacctgttcc agatctataa taaggacttt 2220

tccccattct ctaagggcaa gcccaacctg cacaccctgt actggaagat gctgtttgat 2280

gagaacaatc tgaaggacgt ggtgtataag ctgaatggcg aggccgaggt gttctaccgc 2340

aagaagagca ttgccgagaa gaacaccaca atccacaagg ccaatgagtc catcatcaac 2400

aagaatcctg ataacccaaa ggccaccagc accttcaact atgatatcgt gaaggacaag 2460

agatacacca tcgacaagtt tcagttccac atcccaatca caatgaactt taaggccgag 2520

ggcatcttca acatgaatca gagggtgaat cagttcctga aggccaatcc cgatatcaac 2580

atcatcggca tcgacagagg cgagaggcac ctgctgtact atgccctgat caaccagaag 2640

ggcaagatcc tgaagcagga taccctgaat gtgatcgcca acgagaagca gaaggtggac 2700

taccacaatc tgctggataa gaaggagggc gaccgcgcaa ccgcaaggca ggagtggggc 2760

gtgatcgaga caatcaagga gctgaaggag ggctatctgt cccaggtcat ccacaagctg 2820

accgatctga tgatcgagaa caatgccatc atcgtgatgg aggacctgaa ctttggcttc 2880

aagcggggca gacagaaggt ggagaagcag gtgtatcaga agtttgagaa gatgctgatc 2940

gataagctga attacctggt ggacaagaat aagaaggcaa acgagctggg aggcctgctg 3000

aacgcattcc agctggccaa taagtttgag tccttccaga agatgggcaa gcagaacggc 3060

tttatcttct acgtgcccgc ctggaatacc tctaagacag atcctgccac cggctttatc 3120

gacttcctga agccccgcta tgagaacctg aatcaggcca aggatttctt tgagaagttt 3180

gactctatcc ggctgaacag caaggccgat tactttgagt tcgcctttga cttcaagaat 3240

ttcaccgaga aggccgatgg cggcagaacc aagtggacag tgtgcaccac aaacgaggac 3300

agatatgcct ggaatagggc cctgaacaat aacaggggca gccaggagaa gtacgacatc 3360

acagccgagc tgaagtccct gttcgatggc aaggtggact ataagtctgg caaggatctg 3420

aagcagcaga tcgccagcca ggagtccgcc gacttcttta aggccctgat gaagaacctg 3480

tccatcaccc tgtctctgag acacaataac ggcgagaagg gcgataatga gcaggactac 3540

atcctgtccc ctgtggccga ttctaagggc cgcttctttg actcccggaa ggccgacgat 3600

gacatgccaa agaatgccga cgccaacggc gcctatcaca tcgccctgaa gggcctgtgg 3660

tgtctggagc agatcagcaa gaccgatgac ctgaagaagg tgaagctggc catctccaac 3720

aaggagtggc tggagttcgt gcagacactg aagggcaaaa ggccggcggc cacgaaaaag 3780

gccggccagg caaaaaagaa aaagggatcc tacccatacg atgttccaga ttacgcttat 3840

ccctacgacg tgcctgatta tgcataccca tatgatgtcc ccgactatgc ctaagaattc 3900

<210> 1259
<211> 1250
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 1259

```
Met Gln Thr Leu Phe Glu Asn Phe Thr Asn Gln Tyr Pro Val Ser Lys
1               5                  10                  15

Thr Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Lys Asp Phe Ile
            20                  25                  30

Glu Gln Lys Gly Leu Leu Lys Lys Asp Glu Asp Arg Ala Glu Lys Tyr
        35                  40                  45

Lys Lys Val Lys Asn Ile Ile Asp Glu Tyr His Lys Asp Phe Ile Glu
    50                  55                  60

Lys Ser Leu Asn Gly Leu Lys Leu Asp Gly Leu Glu Lys Tyr Lys Thr
65                  70                  75                  80

Leu Tyr Leu Lys Gln Glu Lys Asp Asp Lys Asp Lys Lys Ala Phe Asp
                85                  90                  95

Lys Glu Lys Glu Asn Leu Arg Lys Gln Ile Ala Asn Ala Phe Arg Asn
            100                 105                 110

Asn Glu Lys Phe Lys Thr Leu Phe Ala Lys Glu Leu Ile Lys Asn Asp
        115                 120                 125

Leu Met Ser Phe Ala Cys Glu Glu Asp Lys Lys Asn Val Lys Glu Phe
    130                 135                 140

Glu Ala Phe Thr Thr Tyr Phe Thr Gly Phe His Gln Asn Arg Ala Asn
145                 150                 155                 160

Met Tyr Val Ala Asp Glu Lys Arg Thr Ala Ile Ala Ser Arg Leu Ile
                165                 170                 175

His Glu Asn Leu Pro Lys Phe Ile Asp Asn Ile Lys Ile Phe Glu Lys
            180                 185                 190

Met Lys Lys Glu Ala Pro Glu Leu Leu Ser Pro Phe Asn Gln Thr Leu
        195                 200                 205

Lys Asp Met Lys Asp Val Ile Lys Gly Thr Thr Leu Glu Glu Ile Phe
    210                 215                 220

Ser Leu Asp Tyr Phe Asn Lys Thr Leu Thr Gln Ser Gly Ile Asp Ile
225                 230                 235                 240

Tyr Asn Ser Val Ile Gly Gly Arg Thr Pro Glu Glu Gly Lys Thr Lys
                245                 250                 255

Ile Lys Gly Leu Asn Glu Tyr Ile Asn Thr Asp Phe Asn Gln Lys Gln
            260                 265                 270

Thr Asp Lys Lys Lys Arg Gln Pro Lys Phe Lys Gln Leu Tyr Lys Gln
        275                 280                 285

Ile Leu Ser Asp Arg Gln Ser Leu Ser Phe Ile Ala Glu Ala Phe Lys
    290                 295                 300

Asn Asp Thr Glu Ile Leu Glu Ala Ile Glu Lys Phe Tyr Val Asn Glu
305                 310                 315                 320

Leu Leu His Phe Ser Asn Glu Gly Lys Ser Thr Asn Val Leu Asp Ala
                325                 330                 335
```

```
Ile Lys Asn Ala Val Ser Asn Leu Glu Ser Phe Asn Leu Thr Lys Met
            340                 345                 350

Tyr Phe Arg Ser Gly Ala Ser Leu Thr Asp Val Ser Arg Lys Val Phe
        355                 360                 365

Gly Glu Trp Ser Ile Ile Asn Arg Ala Leu Asp Asn Tyr Tyr Ala Thr
    370                 375                 380

Thr Tyr Pro Ile Lys Pro Arg Glu Lys Ser Glu Lys Tyr Glu Glu Arg
385                 390                 395                 400

Lys Glu Lys Trp Leu Lys Gln Asp Phe Asn Val Ser Leu Ile Gln Thr
                405                 410                 415

Ala Ile Asp Glu Tyr Asp Asn Glu Thr Val Lys Gly Lys Asn Ser Gly
            420                 425                 430

Lys Val Ile Ala Asp Tyr Phe Ala Lys Phe Cys Asp Asp Lys Glu Thr
        435                 440                 445

Asp Leu Ile Gln Lys Val Asn Glu Gly Tyr Ile Ala Val Lys Asp Leu
    450                 455                 460

Leu Asn Thr Pro Cys Pro Glu Asn Glu Lys Leu Gly Ser Asn Lys Asp
465                 470                 475                 480

Gln Val Lys Gln Ile Lys Ala Phe Met Asp Ser Ile Met Asp Ile Met
                485                 490                 495

His Phe Val Arg Pro Leu Ser Leu Lys Asp Thr Asp Lys Glu Lys Asp
            500                 505                 510

Glu Thr Phe Tyr Ser Leu Phe Thr Pro Leu Tyr Asp His Leu Thr Gln
        515                 520                 525

Thr Ile Ala Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Gln Lys Pro
    530                 535                 540

Tyr Ser Thr Glu Lys Ile Lys Leu Asn Phe Glu Asn Ser Thr Leu Leu
545                 550                 555                 560

Gly Gly Trp Asp Leu Asn Lys Glu Thr Asp Asn Thr Ala Ile Ile Leu
                565                 570                 575

Arg Lys Asp Asn Leu Tyr Tyr Leu Gly Ile Met Asp Lys Arg His Asn
            580                 585                 590

Arg Ile Phe Arg Asn Val Pro Lys Ala Asp Lys Lys Asp Phe Cys Tyr
        595                 600                 605

Glu Lys Met Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro
    610                 615                 620

Lys Val Phe Phe Ser Gln Ser Arg Ile Gln Glu Phe Thr Pro Ser Ala
625                 630                 635                 640

Lys Leu Leu Glu Asn Tyr Ala Asn Glu Thr His Lys Lys Gly Asp Asn
                645                 650                 655

Phe Asn Leu Asn His Cys His Lys Leu Ile Asp Phe Phe Lys Asp Ser
            660                 665                 670
```

```
Ile Asn Lys His Glu Asp Trp Lys Asn Phe Asp Phe Arg Phe Ser Ala
        675                 680                 685

Thr Ser Thr Tyr Ala Asp Leu Ser Gly Phe Tyr His Glu Val Glu His
    690                 695                 700

Gln Gly Tyr Lys Ile Ser Phe Gln Ser Val Ala Asp Ser Phe Ile Asp
705                 710                 715                 720

Asp Leu Val Asn Glu Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys
                725                 730                 735

Asp Phe Ser Pro Phe Ser Lys Gly Lys Pro Asn Leu His Thr Leu Tyr
            740                 745                 750

Trp Lys Met Leu Phe Asp Glu Asn Asn Leu Lys Asp Val Val Tyr Lys
        755                 760                 765

Leu Asn Gly Glu Ala Glu Val Phe Tyr Arg Lys Lys Ser Ile Ala Glu
    770                 775                 780

Lys Asn Thr Thr Ile His Lys Ala Asn Glu Ser Ile Ile Asn Lys Asn
785                 790                 795                 800

Pro Asp Asn Pro Lys Ala Thr Ser Thr Phe Asn Tyr Asp Ile Val Lys
                805                 810                 815

Asp Lys Arg Tyr Thr Ile Asp Lys Phe Gln Phe His Ile Pro Ile Thr
            820                 825                 830

Met Asn Phe Lys Ala Glu Gly Ile Phe Asn Met Asn Gln Arg Val Asn
        835                 840                 845

Gln Phe Leu Lys Ala Asn Pro Asp Ile Asn Ile Ile Gly Ile Asp Arg
    850                 855                 860

Gly Glu Arg His Leu Leu Tyr Tyr Ala Leu Ile Asn Gln Lys Gly Lys
865                 870                 875                 880

Ile Leu Lys Gln Asp Thr Leu Asn Val Ile Ala Asn Glu Lys Gln Lys
                885                 890                 895

Val Asp Tyr His Asn Leu Leu Asp Lys Lys Glu Gly Asp Arg Ala Thr
            900                 905                 910

Ala Arg Gln Glu Trp Gly Val Ile Glu Thr Ile Lys Glu Leu Lys Glu
        915                 920                 925

Gly Tyr Leu Ser Gln Val Ile His Lys Leu Thr Asp Leu Met Ile Glu
    930                 935                 940

Asn Asn Ala Ile Ile Val Met Glu Asp Leu Asn Phe Gly Phe Lys Arg
945                 950                 955                 960

Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu Lys Met
                965                 970                 975

Leu Ile Asp Lys Leu Asn Tyr Leu Val Asp Lys Asn Lys Lys Ala Asn
            980                 985                 990

Glu Leu Gly Gly Leu Leu Asn Ala Phe Gln Leu Ala Asn Lys Phe Glu
        995                 1000                1005
```

Ser Phe Gln Lys Met Gly Lys Gln Asn Gly Phe Ile Phe Tyr Val
1010 1015 1020

Pro Ala Trp Asn Thr Ser Lys Thr Asp Pro Ala Thr Gly Phe Ile
1025 1030 1035

Asp Phe Leu Lys Pro Arg Tyr Glu Asn Leu Asn Gln Ala Lys Asp
1040 1045 1050

Phe Phe Glu Lys Phe Asp Ser Ile Arg Leu Asn Ser Lys Ala Asp
1055 1060 1065

Tyr Phe Glu Phe Ala Phe Asp Phe Lys Asn Phe Thr Glu Lys Ala
1070 1075 1080

Asp Gly Gly Arg Thr Lys Trp Thr Val Cys Thr Thr Asn Glu Asp
1085 1090 1095

Arg Tyr Ala Trp Asn Arg Ala Leu Asn Asn Asn Arg Gly Ser Gln
1100 1105 1110

Glu Lys Tyr Asp Ile Thr Ala Glu Leu Lys Ser Leu Phe Asp Gly
1115 1120 1125

Lys Val Asp Tyr Lys Ser Gly Lys Asp Leu Lys Gln Gln Ile Ala
1130 1135 1140

Ser Gln Glu Ser Ala Asp Phe Phe Lys Ala Leu Met Lys Asn Leu
1145 1150 1155

Ser Ile Thr Leu Ser Leu Arg His Asn Asn Gly Glu Lys Gly Asp
1160 1165 1170

Asn Glu Gln Asp Tyr Ile Leu Ser Pro Val Ala Asp Ser Lys Gly
1175 1180 1185

Arg Phe Phe Asp Ser Arg Lys Ala Asp Asp Asp Met Pro Lys Asn
1190 1195 1200

Ala Asp Ala Asn Gly Ala Tyr His Ile Ala Leu Lys Gly Leu Trp
1205 1210 1215

Cys Leu Glu Gln Ile Ser Lys Thr Asp Asp Leu Lys Lys Val Lys
1220 1225 1230

Leu Ala Ile Ser Asn Lys Glu Trp Leu Glu Phe Val Gln Thr Leu
1235 1240 1245

Lys Gly
1250

<210> 1260
<211> 4071
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1260
ggtaccatga cacagttcga gggctttacc aacctgtatc aggtgagcaa gacactgcgg 60

tttgagctga tcccacaggg caagaccctg aagcacatcc aggagcaggg cttcatcgag 120

gaggacaagg cccgcaatga tcactacaag gagctgaagc ccatcatcga tcggatctac 180

aagacctatg ccgaccagtg cctgcagctg gtgcagctgg attgggagaa cctgagcgcc 240

gccatcgact cctatagaaa ggagaaaacc gaggagacaa ggaacgccct gatcgaggag 300

caggccacat atcgcaatgc catccacgac tacttcatcg gccggacaga caacctgacc 360

gatgccatca ataagagaca cgccgagatc tacaagggcc tgttcaaggc cgagctgttt 420

aatggcaagg tgctgaagca gctgggcacc gtgaccacaa ccgagcacga gaacgccctg 480

ctgcggagct tcgacaagtt tacaacctac ttctccggct tttatgagaa caggaagaac 540

gtgttcagcg ccgaggatat cagcacagcc atcccacacc gcatcgtgca ggacaacttc 600

cccaagttta aggagaattg tcacatcttc acacgcctga tcaccgccgt gcccagcctg 660

cgggagcact ttgagaacgt gaagaaggcc atcggcatct tcgtgagcac ctccatcgag 720

gaggtgtttt ccttcccttt ttataaccag ctgctgacac agacccagat cgacctgtat 780

aaccagctgc tgggaggaat ctctcgggag gcaggcaccg agaagatcaa gggcctgaac 840

gaggtgctga atctggccat ccagaagaat gatgagacag cccacatcat cgcctccctg 900

ccacacagat tcatcccccc gtttaagcag atcctgtccg ataggaacac cctgtctttc 960

atcctggagg agtttaagag cgacgaggaa gtgatccagt ccttctgcaa gtacaagaca 1020

ctgctgagaa acgagaacgt gctggagaca gccgaggccc tgtttaacga gctgaacagc 1080

atcgacctga cacacatctt catcagccac aagaagctgg agacaatcag cagcgccctg 1140

tgcgaccact gggatacact gaggaatgcc ctgtatgagc ggagaatctc cgagctgaca 1200

ggcaagatca ccaagtctgc caaggagaag gtgcagcgca gcctgaagca cgaggatatc 1260

aacctgcagg agatcatctc tgccgcaggc aaggagctga gcgaggcctt caagcagaaa 1320

accagcgaga tcctgtccca cgcacacgcc gccctggatc agccactgcc tacaaccctg 1380

aagaagcagg aggagaagga gatcctgaag tctcagctgg acagcctgct gggcctgtac 1440

cacctgctgg actggtttgc cgtggatgag tccaacgagg tggaccccga gttctctgcc 1500

cggctgaccg gcatcaagct ggagatggag ccttctctga gcttctacaa caaggccaga 1560

aattatgcca ccaagaagcc ctactccgtg gagaagttca agctgaactt tcagatgcct 1620

acactggcct ctggctggga cgtgaataag gagaagaaca atggcgccat cctgtttgtg 1680

aagaacggcc tgtactatct gggcatcatg ccaaagcaga agggcaggta taaggccctg 1740

agcttcgagc ccacagagaa aaccagcgag ggctttgata agatgtacta tgactacttc 1800

cctgatgccg ccaagatgat cccaaagtgc agcacccagc tgaaggccgt gacagcccac 1860

tttcagaccc acacaacccc catcctgctg tccaacaatt tcatcgagcc tctggagatc 1920

acaaaggaga tctacgacct gaacaatcct gagaaggagc caaagaagtt tcagacagcc 1980

tacgccaaga aaaccggcga ccagaagggc tacagagagg ccctgtgcaa gtggatcgac 2040

ttcacaaggg attttctgtc caagtatacc aagacaacct ctatcgatct gtctagcctg 2100

cggccatcct ctcagtataa ggacctgggc gagtactatg ccgagctgaa tcccctgctg 2160

taccacatca gcttccagag aatcgccgag aaggagatca tggatgccgt ggagacaggc 2220

aagctgtacc tgttccagat ctataacaag gactttgcca agggccacca cggcaagcct 2280

aatctgcaca cactgtattg gaccggcctg ttttctccag agaacctggc caagacaagc 2340

atcaagctga atggccaggc cgagctgttc taccgcccta agtccaggat gaagaggatg 2400

gcacaccggc tgggagagaa gatgctgaac aagaagctga aggatcagaa aaccccaatc 2460

cccgacaccc tgtaccagga gctgtacgac tatgtgaatc acagactgtc ccacgacctg 2520

tctgatgagg ccagggccct gctgcccaac gtgatcacca aggaggtgtc tcacgagatc 2580

atcaaggata ggcgctttac cagcgacaag ttcttttttcc acgtgcctat cacactgaac 2640

tatcaggccg ccaattcccc atctaagttc aaccagaggg tgaatgccta cctgaaggag 2700

caccccgaga cacctatcat cggcatcgat cggggcgaga gaaacctgat ctatatcaca 2760

gtgatcgact ccaccggcaa gatcctggag cagcggagcc tgaacaccat ccagcagttt 2820

gattaccaga agaagctgga caacagggag aaggagaggg tggcagcaag gcaggcctgg 2880

tctgtggtgg gcacaatcaa ggatctgaag cagggctatc tgagccaggt catccacgag 2940

atcgtggacc tgatgatcca ctaccaggcc gtggtggtgc tggagaacct gaatttcggc 3000

tttaagagca agaggaccgg catcgccgag aaggccgtgt accagcagtt cgagaagatg 3060

ctgatcgata agctgaattg cctggtgctg aaggactatc cagcagagaa agtgggaggc 3120

gtgctgaacc cataccagct gacagaccag ttcacctcct ttgccaagat gggcacccag 3180

tctggcttcc tgttttacgt gcctgcccca tatacatcta agatcgatcc cctgaccggc 3240

ttcgtggacc ccttcgtgtg gaaaaccatc aagaatcacg agagccgcaa gcacttcctg 3300

gagggcttcg actttctgca ctacgacgtg aaaaccggcg acttcatcct gcactttaag 3360

atgaacagaa atctgtcctt ccagaggggc ctgcccggct ttatgcctgc atgggatatc 3420

gtgttcgaga agaacgagac acagtttgac gccaagggca ccccttttcat cgccggcaag 3480

agaatcgtgc cagtgatcga gaatcacaga ttcaccggca gataccggga cctgtatcct 3540

gccaacgagc tgatcgccct gctggaggag aagggcatcg tgttcaggga tggctccaac 3600

atcctgccaa agctgctgga gaatgacgat tctcacgcca tcgacaccat ggtggccctg 3660

atccgcagcg tgctgcagat gcggaactcc aatgccgcca caggcgagga ctatatcaac 3720

agccccgtgc gcgatctgaa tggcgtgtgc ttcgactccc ggtttcagaa cccagagtgg 3780

cccatggacg ccgatgccaa tggcgcctac cacatcgccc tgaagggcca gctgctgctg 3840

aatcacctga aggagagcaa ggatctgaag ctgcagaacg gcatctccaa tcaggactgg 3900

ctggcctaca tccaggagct gcgcaacaaa aggccggcgg ccacgaaaaa ggccggccag 3960

gcaaaaaaga aaaagggatc ctacccatac gatgttccag attacgctta tccctacgac 4020

gtgcctgatt atgcataccc atatgatgtc cccgactatg cctaagaatt c 4071

<210> 1261
<211> 1307
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 1261
Met Thr Gln Phe Glu Gly Phe Thr Asn Leu Tyr Gln Val Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Lys His Ile Gln
20 25 30

Glu Gln Gly Phe Ile Glu Glu Asp Lys Ala Arg Asn Asp His Tyr Lys
35 40 45

Glu Leu Lys Pro Ile Ile Asp Arg Ile Tyr Lys Thr Tyr Ala Asp Gln
50 55 60

Cys Leu Gln Leu Val Gln Leu Asp Trp Glu Asn Leu Ser Ala Ala Ile
65 70 75 80

Asp Ser Tyr Arg Lys Glu Lys Thr Glu Glu Thr Arg Asn Ala Leu Ile
85 90 95

Glu Glu Gln Ala Thr Tyr Arg Asn Ala Ile His Asp Tyr Phe Ile Gly

100 105 110

Arg Thr Asp Asn Leu Thr Asp Ala Ile Asn Lys Arg His Ala Glu Ile
115 120 125

Tyr Lys Gly Leu Phe Lys Ala Glu Leu Phe Asn Gly Lys Val Leu Lys
130 135 140

Gln Leu Gly Thr Val Thr Thr Thr Glu His Glu Asn Ala Leu Leu Arg
145 150 155 160

Ser Phe Asp Lys Phe Thr Thr Tyr Phe Ser Gly Phe Tyr Glu Asn Arg
165 170 175

Lys Asn Val Phe Ser Ala Glu Asp Ile Ser Thr Ala Ile Pro His Arg
180 185 190

Ile Val Gln Asp Asn Phe Pro Lys Phe Lys Glu Asn Cys His Ile Phe
195 200 205

Thr Arg Leu Ile Thr Ala Val Pro Ser Leu Arg Glu His Phe Glu Asn
210 215 220

Val Lys Lys Ala Ile Gly Ile Phe Val Ser Thr Ser Ile Glu Glu Val
225 230 235 240

Phe Ser Phe Pro Phe Tyr Asn Gln Leu Leu Thr Gln Thr Gln Ile Asp
245 250 255

Leu Tyr Asn Gln Leu Leu Gly Gly Ile Ser Arg Glu Ala Gly Thr Glu
260 265 270

Lys Ile Lys Gly Leu Asn Glu Val Leu Asn Leu Ala Ile Gln Lys Asn
275 280 285

Asp Glu Thr Ala His Ile Ile Ala Ser Leu Pro His Arg Phe Ile Pro
290 295 300

Leu Phe Lys Gln Ile Leu Ser Asp Arg Asn Thr Leu Ser Phe Ile Leu
305 310 315 320

Glu Glu Phe Lys Ser Asp Glu Glu Val Ile Gln Ser Phe Cys Lys Tyr
325 330 335

Lys Thr Leu Leu Arg Asn Glu Asn Val Leu Glu Thr Ala Glu Ala Leu
340 345 350

Phe Asn Glu Leu Asn Ser Ile Asp Leu Thr His Ile Phe Ile Ser His
355 360 365

Lys Lys Leu Glu Thr Ile Ser Ser Ala Leu Cys Asp His Trp Asp Thr
370 375 380

Leu Arg Asn Ala Leu Tyr Glu Arg Arg Ile Ser Glu Leu Thr Gly Lys
385 390 395 400

Ile Thr Lys Ser Ala Lys Glu Lys Val Gln Arg Ser Leu Lys His Glu
405 410 415

Asp Ile Asn Leu Gln Glu Ile Ile Ser Ala Ala Gly Lys Glu Leu Ser
420 425 430

Glu Ala Phe Lys Gln Lys Thr Ser Glu Ile Leu Ser His Ala His Ala

435 440 445

Ala Leu Asp Gln Pro Leu Pro Thr Thr Leu Lys Lys Gln Glu Glu Lys
450 455 460

Glu Ile Leu Lys Ser Gln Leu Asp Ser Leu Leu Gly Leu Tyr His Leu
465 470 475 480

Leu Asp Trp Phe Ala Val Asp Glu Ser Asn Glu Val Asp Pro Glu Phe
485 490 495

Ser Ala Arg Leu Thr Gly Ile Lys Leu Glu Met Glu Pro Ser Leu Ser
500 505 510

Phe Tyr Asn Lys Ala Arg Asn Tyr Ala Thr Lys Lys Pro Tyr Ser Val
515 520 525

Glu Lys Phe Lys Leu Asn Phe Gln Met Pro Thr Leu Ala Ser Gly Trp
530 535 540

Asp Val Asn Lys Glu Lys Asn Asn Gly Ala Ile Leu Phe Val Lys Asn
545 550 555 560

Gly Leu Tyr Tyr Leu Gly Ile Met Pro Lys Gln Lys Gly Arg Tyr Lys
565 570 575

Ala Leu Ser Phe Glu Pro Thr Glu Lys Thr Ser Glu Gly Phe Asp Lys
580 585 590

Met Tyr Tyr Asp Tyr Phe Pro Asp Ala Ala Lys Met Ile Pro Lys Cys
595 600 605

Ser Thr Gln Leu Lys Ala Val Thr Ala His Phe Gln Thr His Thr Thr
610 615 620

Pro Ile Leu Leu Ser Asn Asn Phe Ile Glu Pro Leu Glu Ile Thr Lys
625 630 635 640

Glu Ile Tyr Asp Leu Asn Asn Pro Glu Lys Glu Pro Lys Lys Phe Gln
645 650 655

Thr Ala Tyr Ala Lys Lys Thr Gly Asp Gln Lys Gly Tyr Arg Glu Ala
660 665 670

Leu Cys Lys Trp Ile Asp Phe Thr Arg Asp Phe Leu Ser Lys Tyr Thr
675 680 685

Lys Thr Thr Ser Ile Asp Leu Ser Ser Leu Arg Pro Ser Ser Gln Tyr
690 695 700

Lys Asp Leu Gly Glu Tyr Tyr Ala Glu Leu Asn Pro Leu Leu Tyr His
705 710 715 720

Ile Ser Phe Gln Arg Ile Ala Glu Lys Glu Ile Met Asp Ala Val Glu
725 730 735

Thr Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ala Lys
740 745 750

Gly His His Gly Lys Pro Asn Leu His Thr Leu Tyr Trp Thr Gly Leu
755 760 765

Phe Ser Pro Glu Asn Leu Ala Lys Thr Ser Ile Lys Leu Asn Gly Gln

770 775 780

Ala Glu Leu Phe Tyr Arg Pro Lys Ser Arg Met Lys Arg Met Ala His
785 790 795 800

Arg Leu Gly Glu Lys Met Leu Asn Lys Lys Leu Lys Asp Gln Lys Thr
805 810 815

Pro Ile Pro Asp Thr Leu Tyr Gln Glu Leu Tyr Asp Tyr Val Asn His
820 825 830

Arg Leu Ser His Asp Leu Ser Asp Glu Ala Arg Ala Leu Leu Pro Asn
835 840 845

Val Ile Thr Lys Glu Val Ser His Glu Ile Ile Lys Asp Arg Arg Phe
850 855 860

Thr Ser Asp Lys Phe Phe Phe His Val Pro Ile Thr Leu Asn Tyr Gln
865 870 875 880

Ala Ala Asn Ser Pro Ser Lys Phe Asn Gln Arg Val Asn Ala Tyr Leu
885 890 895

Lys Glu His Pro Glu Thr Pro Ile Ile Gly Ile Asp Arg Gly Glu Arg
900 905 910

Asn Leu Ile Tyr Ile Thr Val Ile Asp Ser Thr Gly Lys Ile Leu Glu
915 920 925

Gln Arg Ser Leu Asn Thr Ile Gln Gln Phe Asp Tyr Gln Lys Lys Leu
930 935 940

Asp Asn Arg Glu Lys Glu Arg Val Ala Ala Arg Gln Ala Trp Ser Val
945 950 955 960

Val Gly Thr Ile Lys Asp Leu Lys Gln Gly Tyr Leu Ser Gln Val Ile
965 970 975

His Glu Ile Val Asp Leu Met Ile His Tyr Gln Ala Val Val Val Leu
980 985 990

Glu Asn Leu Asn Phe Gly Phe Lys Ser Lys Arg Thr Gly Ile Ala Glu
995 1000 1005

Lys Ala Val Tyr Gln Gln Phe Glu Lys Met Leu Ile Asp Lys Leu
1010 1015 1020

Asn Cys Leu Val Leu Lys Asp Tyr Pro Ala Glu Lys Val Gly Gly
1025 1030 1035

Val Leu Asn Pro Tyr Gln Leu Thr Asp Gln Phe Thr Ser Phe Ala
1040 1045 1050

Lys Met Gly Thr Gln Ser Gly Phe Leu Phe Tyr Val Pro Ala Pro
1055 1060 1065

Tyr Thr Ser Lys Ile Asp Pro Leu Thr Gly Phe Val Asp Pro Phe
1070 1075 1080

Val Trp Lys Thr Ile Lys Asn His Glu Ser Arg Lys His Phe Leu
1085 1090 1095

Glu Gly Phe Asp Phe Leu His Tyr Asp Val Lys Thr Gly Asp Phe

1100 1105 1110

Ile Leu His Phe Lys Met Asn Arg Asn Leu Ser Phe Gln Arg Gly
1115 1120 1125

Leu Pro Gly Phe Met Pro Ala Trp Asp Ile Val Phe Glu Lys Asn
1130 1135 1140

Glu Thr Gln Phe Asp Ala Lys Gly Thr Pro Phe Ile Ala Gly Lys
1145 1150 1155

Arg Ile Val Pro Val Ile Glu Asn His Arg Phe Thr Gly Arg Tyr
1160 1165 1170

Arg Asp Leu Tyr Pro Ala Asn Glu Leu Ile Ala Leu Leu Glu Glu
1175 1180 1185

Lys Gly Ile Val Phe Arg Asp Gly Ser Asn Ile Leu Pro Lys Leu
1190 1195 1200

Leu Glu Asn Asp Asp Ser His Ala Ile Asp Thr Met Val Ala Leu
1205 1210 1215

Ile Arg Ser Val Leu Gln Met Arg Asn Ser Asn Ala Ala Thr Gly
1220 1225 1230

Glu Asp Tyr Ile Asn Ser Pro Val Arg Asp Leu Asn Gly Val Cys
1235 1240 1245

Phe Asp Ser Arg Phe Gln Asn Pro Glu Trp Pro Met Asp Ala Asp
1250 1255 1260

Ala Asn Gly Ala Tyr His Ile Ala Leu Lys Gly Gln Leu Leu Leu
1265 1270 1275

Asn His Leu Lys Glu Ser Lys Asp Leu Lys Leu Gln Asn Gly Ile
1280 1285 1290

Ser Asn Gln Asp Trp Leu Ala Tyr Ile Gln Glu Leu Arg Asn
1295 1300 1305

<210> 1262
<211> 3768
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1262
ggtaccatgt actatgagtc cctgaccaag cagtaccccg tgtctaagac aatccggaat 60

gagctgatcc ctatcggcaa gacactggat aacatccgcc agaacaatat cctggagagc 120

gacgtgaagc ggaagcagaa ctacgagcac gtgaagggca tcctggatga gtatcacaag 180

cagctgatca acgaggccct ggacaattgc accctgccat ccctgaagat cgccgccgag 240

atctacctga agaatcagaa ggaggtgtct gacagagagg atttcaacaa gacacaggac 300

ctgctgagga aggaggtggt ggagaagctg aaggcccacg agaactttac caagatcggc 360

aagaaggaca tcctggatct gctggagaag ctgccttcca tctctgagga cgattacaat 420

gccctggaga gcttccgcaa cttttacacc tatttcacat cctacaacaa ggtgcgggag 480

aatctgtatt ctgataagga gaagagctcc acagtggcct acagactgat caacgagaat 540

ttcccaaagt ttctggacaa tgtgaagagc tataggtttg tgaaaaccgc aggcatcctg 600

gcagatggcc tgggagagga ggagcaggac tccctgttca tcgtggagac attcaacaag 660

accctgacac aggacggcat cgatacctac aattctcaag tgggcaagat caactctagc 720

atcaatctgt ataaccagaa gaatcagaag gccaatggct tcagaaagat ccccaagatg 780

aagatgctgt ataagcagat cctgtccgat agggaggagt ctttcatcga cgagtttcag 840

agcgatgagg tgctgatcga caacgtggag tcttatggca gcgtgctgat cgagtctctg 900

aagtcctcta aggtgagcgc cttctttgat gccctgagag agtctaaggg caagaacgtg 960

tacgtgaaga atgacctggc caagacagcc atgagcaaca tcgtgttcga gaattggagg 1020

acctttgacg atctgctgaa ccaggagtac gacctggcca acgagaacaa gaagaaggac 1080

gataagtatt tcgagaagcg ccagaaggag ctgaagaaga ataagagcta ctccctggag 1140

cacctgtgca acctgtccga ggattcttgt aacctgatcg agaattatat ccaccagatc 1200

tccgacgata tcgagaatat catcatcaac aatgagacat tcctgcgcat cgtgatcaat 1260

gagcacgaca ggtcccgcaa gctggccaag aaccggaagg ccgtgaaggc catcaaggac 1320

tttctggatt ctatcaaggt gctggagcgg gagctgaagc tgatcaacag ctccggccag 1380

gagctggaga aggatctgat cgtgtactct gcccacgagg agctgctggt ggagctgaag 1440

caggtggaca gcctgtataa catgaccaga aattatctga caaagaagcc tttctctacc 1500

gagaaggtga agctgaactt taatcgcagc acactgctga acggctggga tcggaataag 1560

gagacagaca acctgggcgt gctgctgctg aaggacggca agtactatct gggcatcatg 1620

aacacaagcg ccaataaggc cttcgtgaat ccccctgtgg ccaagaccga gaaggtgttt 1680

aagaaggtgg attacaagct gctgccagtg cccaaccaga tgctgccaaa ggtgttcttt 1740

gccaagagca atatcgactt ctataacccc tctagcgaga tctactccaa ttataagaag 1800

ggcacccaca agaagggcaa tatgttttcc ctggaggatt gtcacaacct gatcgacttc 1860

tttaaggagt ctatcagcaa gcacgaggac tggagcaagt tcggctttaa gttcagcgat 1920

acagcctcct acaacgacat ctccgagttc tatcgcgagg tggagaagca gggctacaag 1980

ctgacctata cagacatcga tgagacatac atcaatgatc tgatcgagcg gaacgagctg 2040

tacctgttcc agatctataa taaggacttt agcatgtact ccaagggcaa gctgaacctg 2100

cacacactgt atttcatgat gctgtttgat cagcgcaata tcgacgacgt ggtgtataag 2160

ctgaacggag aggcagaggt gttctatagg ccagcctcca tctctgagga cgagctgatc 2220

atccacaagg ccggcgagga gatcaagaac aagaatccta accgggccag aaccaaggag 2280

acaagcacct tcagctacga catcgtgaag gataagcggt atagcaagga taagtttacc 2340

ctgcacatcc ccatcacaat gaacttcggc gtggatgagg tgaagcggtt caacgacgcc 2400

gtgaacagcg ccatccggat cgatgagaat gtgaacgtga tcggcatcga ccggggcgag 2460

agaaatctgc tgtacgtggt ggtcatcgac tctaagggca acatcctgga gcagatctcc 2520

ctgaactcta tcatcaataa ggagtacgac atcgagacag attatcacgc actgctggat 2580

gagagggagg gcggcagaga taaggcccgg aaggactgga acaccgtgga gaatatcagg 2640

gacctgaagg ccggctacct gagccaggtg gtgaacgtgg tggccaagct ggtgctgaag 2700

tataatgcca tcatctgcct ggaggacctg aactttggct tcaagagggg ccgccagaag 2760

gtggagaagc aggtgtacca gaagttcgag aagatgctga tcgataagct gaattacctg 2820

gtcatcgaca agagccgcga gcagacatcc cctaaggagc tgggaggcgc cctgaacgca 2880

ctgcagctga cctctaagtt caagagcttt aaggagctgg gcaagcagtc cggcgtgatc 2940

tactatgtgc ctgcctacct gacctctaag atcgatccaa ccacaggctt cgccaatctg 3000

ttttatatga agtgtgagaa cgtggagaag tccaagagat tctttgacgg ctttgatttc 3060

atcaggttca acgccctgga gaacgtgttc gagttcggct ttgactaccg gagcttcacc 3120

cagagggcct gcggcatcaa ttccaagtgg accgtgtgca ccaacggcga gcgcatcatc 3180

aagtatcgga atccagataa gaacaatatg ttcgacgaga aggtggtggt ggtgaccgat 3240

gagatgaaga acctgtttga gcagtacaag atcccctatg aggatggcag aaatgtgaag 3300

gacatgatca tcagcaacga ggaggccgag ttctaccgga gactgtatag gctgctgcag 3360

cagaccctgc agatgagaaa cagcacctcc gacggcacaa gggattacat catctcccct 3420

gtgaagaata agagagaggc ctacttcaac agcgagctgt ccgacggctc tgtgccaaag 3480

gacgccgatg ccaacggcgc ctacaatatc gccagaaagg gcctgtgggt gctggagcag 3540

atcaggcaga agagcgaggg cgagaagatc aatctggcca tgaccaacgc cgagtggctg 3600

gagtatgccc agacacacct gctgaaaagg ccggcggcca cgaaaaaggc cggccaggca 3660

aaaaagaaaa agggatccta cccatacgat gttccagatt acgcttatcc ctacgacgtg 3720

cctgattatg catacccata tgatgtcccc gactatgcct aagaattc 3768

<210> 1263
<211> 1206
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 1263

Met Tyr Tyr Glu Ser Leu Thr Lys Gln Tyr Pro Val Ser Lys Thr Ile
1 5 10 15

Arg Asn Glu Leu Ile Pro Ile Gly Lys Thr Leu Asp Asn Ile Arg Gln
20 25 30

Asn Asn Ile Leu Glu Ser Asp Val Lys Arg Lys Gln Asn Tyr Glu His
35 40 45

Val Lys Gly Ile Leu Asp Glu Tyr His Lys Gln Leu Ile Asn Glu Ala
50 55 60

Leu Asp Asn Cys Thr Leu Pro Ser Leu Lys Ile Ala Ala Glu Ile Tyr
65 70 75 80

Leu Lys Asn Gln Lys Glu Val Ser Asp Arg Glu Asp Phe Asn Lys Thr
85 90 95

Gln Asp Leu Leu Arg Lys Glu Val Val Glu Lys Leu Lys Ala His Glu
100 105 110

Asn Phe Thr Lys Ile Gly Lys Lys Asp Ile Leu Asp Leu Leu Glu Lys
115 120 125

Leu Pro Ser Ile Ser Glu Asp Asp Tyr Asn Ala Leu Glu Ser Phe Arg
130 135 140

Asn Phe Tyr Thr Tyr Phe Thr Ser Tyr Asn Lys Val Arg Glu Asn Leu
145 150 155 160

Tyr Ser Asp Lys Glu Lys Ser Ser Thr Val Ala Tyr Arg Leu Ile Asn
165 170 175

Glu Asn Phe Pro Lys Phe Leu Asp Asn Val Lys Ser Tyr Arg Phe Val
180 185 190

Lys Thr Ala Gly Ile Leu Ala Asp Gly Leu Gly Glu Glu Glu Gln Asp
195 200 205

```
Ser Leu Phe Ile Val Glu Thr Phe Asn Lys Thr Leu Thr Gln Asp Gly
    210                 215                 220

Ile Asp Thr Tyr Asn Ser Gln Val Gly Lys Ile Asn Ser Ser Ile Asn
225                 230                 235                 240

Leu Tyr Asn Gln Lys Asn Gln Lys Ala Asn Gly Phe Arg Lys Ile Pro
                245                 250                 255

Lys Met Lys Met Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Glu Ser
            260                 265                 270

Phe Ile Asp Glu Phe Gln Ser Asp Glu Val Leu Ile Asp Asn Val Glu
        275                 280                 285

Ser Tyr Gly Ser Val Leu Ile Glu Ser Leu Lys Ser Ser Lys Val Ser
    290                 295                 300

Ala Phe Phe Asp Ala Leu Arg Glu Ser Lys Gly Lys Asn Val Tyr Val
305                 310                 315                 320

Lys Asn Asp Leu Ala Lys Thr Ala Met Ser Asn Ile Val Phe Glu Asn
                325                 330                 335

Trp Arg Thr Phe Asp Asp Leu Leu Asn Gln Glu Tyr Asp Leu Ala Asn
            340                 345                 350

Glu Asn Lys Lys Lys Asp Asp Lys Tyr Phe Glu Lys Arg Gln Lys Glu
        355                 360                 365

Leu Lys Lys Asn Lys Ser Tyr Ser Leu Glu His Leu Cys Asn Leu Ser
    370                 375                 380

Glu Asp Ser Cys Asn Leu Ile Glu Asn Tyr Ile His Gln Ile Ser Asp
385                 390                 395                 400

Asp Ile Glu Asn Ile Ile Ile Asn Asn Glu Thr Phe Leu Arg Ile Val
                405                 410                 415

Ile Asn Glu His Asp Arg Ser Arg Lys Leu Ala Lys Asn Arg Lys Ala
            420                 425                 430

Val Lys Ala Ile Lys Asp Phe Leu Asp Ser Ile Lys Val Leu Glu Arg
        435                 440                 445

Glu Leu Lys Leu Ile Asn Ser Ser Gly Gln Glu Leu Glu Lys Asp Leu
    450                 455                 460

Ile Val Tyr Ser Ala His Glu Glu Leu Leu Val Glu Leu Lys Gln Val
465                 470                 475                 480

Asp Ser Leu Tyr Asn Met Thr Arg Asn Tyr Leu Thr Lys Lys Pro Phe
                485                 490                 495

Ser Thr Glu Lys Val Lys Leu Asn Phe Asn Arg Ser Thr Leu Leu Asn
            500                 505                 510

Gly Trp Asp Arg Asn Lys Glu Thr Asp Asn Leu Gly Val Leu Leu Leu
        515                 520                 525

Lys Asp Gly Lys Tyr Tyr Leu Gly Ile Met Asn Thr Ser Ala Asn Lys
    530                 535                 540
```

```
Ala Phe Val Asn Pro Pro Val Ala Lys Thr Glu Lys Val Phe Lys Lys
545                 550                 555                 560

Val Asp Tyr Lys Leu Leu Pro Val Pro Asn Gln Met Leu Pro Lys Val
                565                 570                 575

Phe Phe Ala Lys Ser Asn Ile Asp Phe Tyr Asn Pro Ser Ser Glu Ile
            580                 585                 590

Tyr Ser Asn Tyr Lys Lys Gly Thr His Lys Lys Gly Asn Met Phe Ser
        595                 600                 605

Leu Glu Asp Cys His Asn Leu Ile Asp Phe Phe Lys Glu Ser Ile Ser
    610                 615                 620

Lys His Glu Asp Trp Ser Lys Phe Gly Phe Lys Phe Ser Asp Thr Ala
625                 630                 635                 640

Ser Tyr Asn Asp Ile Ser Glu Phe Tyr Arg Glu Val Glu Lys Gln Gly
                645                 650                 655

Tyr Lys Leu Thr Tyr Thr Asp Ile Asp Glu Thr Tyr Ile Asn Asp Leu
            660                 665                 670

Ile Glu Arg Asn Glu Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe
        675                 680                 685

Ser Met Tyr Ser Lys Gly Lys Leu Asn Leu His Thr Leu Tyr Phe Met
    690                 695                 700

Met Leu Phe Asp Gln Arg Asn Ile Asp Asp Val Val Tyr Lys Leu Asn
705                 710                 715                 720

Gly Glu Ala Glu Val Phe Tyr Arg Pro Ala Ser Ile Ser Glu Asp Glu
                725                 730                 735

Leu Ile Ile His Lys Ala Gly Glu Glu Ile Lys Asn Lys Asn Pro Asn
            740                 745                 750

Arg Ala Arg Thr Lys Glu Thr Ser Thr Phe Ser Tyr Asp Ile Val Lys
        755                 760                 765

Asp Lys Arg Tyr Ser Lys Asp Lys Phe Thr Leu His Ile Pro Ile Thr
    770                 775                 780

Met Asn Phe Gly Val Asp Glu Val Lys Arg Phe Asn Asp Ala Val Asn
785                 790                 795                 800

Ser Ala Ile Arg Ile Asp Glu Asn Val Asn Val Ile Gly Ile Asp Arg
                805                 810                 815

Gly Glu Arg Asn Leu Leu Tyr Val Val Val Ile Asp Ser Lys Gly Asn
            820                 825                 830

Ile Leu Glu Gln Ile Ser Leu Asn Ser Ile Ile Asn Lys Glu Tyr Asp
        835                 840                 845

Ile Glu Thr Asp Tyr His Ala Leu Leu Asp Glu Arg Glu Gly Gly Arg
    850                 855                 860

Asp Lys Ala Arg Lys Asp Trp Asn Thr Val Glu Asn Ile Arg Asp Leu
865                 870                 875                 880
```

Lys Ala Gly Tyr Leu Ser Gln Val Val Asn Val Val Ala Lys Leu Val
885 890 895

Leu Lys Tyr Asn Ala Ile Ile Cys Leu Glu Asp Leu Asn Phe Gly Phe
900 905 910

Lys Arg Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu
915 920 925

Lys Met Leu Ile Asp Lys Leu Asn Tyr Leu Val Ile Asp Lys Ser Arg
930 935 940

Glu Gln Thr Ser Pro Lys Glu Leu Gly Gly Ala Leu Asn Ala Leu Gln
945 950 955 960

Leu Thr Ser Lys Phe Lys Ser Phe Lys Glu Leu Gly Lys Gln Ser Gly
965 970 975

Val Ile Tyr Tyr Val Pro Ala Tyr Leu Thr Ser Lys Ile Asp Pro Thr
980 985 990

Thr Gly Phe Ala Asn Leu Phe Tyr Met Lys Cys Glu Asn Val Glu Lys
995 1000 1005

Ser Lys Arg Phe Phe Asp Gly Phe Asp Phe Ile Arg Phe Asn Ala
1010 1015 1020

Leu Glu Asn Val Phe Glu Phe Gly Phe Asp Tyr Arg Ser Phe Thr
1025 1030 1035

Gln Arg Ala Cys Gly Ile Asn Ser Lys Trp Thr Val Cys Thr Asn
1040 1045 1050

Gly Glu Arg Ile Ile Lys Tyr Arg Asn Pro Asp Lys Asn Asn Met
1055 1060 1065

Phe Asp Glu Lys Val Val Val Val Thr Asp Glu Met Lys Asn Leu
1070 1075 1080

Phe Glu Gln Tyr Lys Ile Pro Tyr Glu Asp Gly Arg Asn Val Lys
1085 1090 1095

Asp Met Ile Ile Ser Asn Glu Glu Ala Glu Phe Tyr Arg Arg Leu
1100 1105 1110

Tyr Arg Leu Leu Gln Gln Thr Leu Gln Met Arg Asn Ser Thr Ser
1115 1120 1125

Asp Gly Thr Arg Asp Tyr Ile Ile Ser Pro Val Lys Asn Lys Arg
1130 1135 1140

Glu Ala Tyr Phe Asn Ser Glu Leu Ser Asp Gly Ser Val Pro Lys
1145 1150 1155

Asp Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala Arg Lys Gly Leu
1160 1165 1170

Trp Val Leu Glu Gln Ile Arg Gln Lys Ser Glu Gly Glu Lys Ile
1175 1180 1185

Asn Leu Ala Met Thr Asn Ala Glu Trp Leu Glu Tyr Ala Gln Thr
1190 1195 1200

His Leu Leu
1205

<210> 1264
<211> 3864
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1264
ggtaccatga acaattacga cgagttcacc aagctgtatc ctatccagaa aaccatccgg 60

tttgagctga agccacaggg cagaaccatg gagcacctgg agacattcaa cttctttgag 120

gaggaccggg atagagccga gaagtataag atcctgaagg aggccatcga cgagtaccac 180

aagaagttta tcgatgagca cctgaccaat atgtccctgg attggaactc tctgaagcag 240

atcagcgaga agtactataa gagcagggag gagaaggaca agaaggtgtt cctgtccgag 300

cagaagagga tgcgccagga gatcgtgtct gagtttaaga aggacgatcg cttcaaggac 360

ctgttttcca agaagctgtt ctctgagctg ctgaaggagg agatctacaa gaagggcaac 420

caccaggaga tcgacgccct gaagagcttc gataagtttt ccggctattt catcggcctg 480

cacgagaata ggaagaacat gtactccgac ggcgatgaga tcaccgccat ctccaatcgc 540

atcgtgaatg agaacttccc caagtttctg gataacctgc agaagtacca ggaggccagg 600

aagaagtatc ctgagtggat catcaaggcc gagagcgccc tggtggccca caatatcaag 660

atggacgagg tgttctccct ggagtacttt aataaggtgc tgaaccagga gggcatccag 720

cggtacaacc tggccctggg cggctatgtg accaagagcg gcgagaagat gatgggcctg 780

aatgatgccc tgaacctggc ccaccagtcc gagaagagct ccaagggcag aatccacatg 840

acccccctgt tcaagcagat cctgtccgag aaggagtcct tctcttacat ccccgacgtg 900

tttacagagg attctcagct gctgcctagc atcggcggct tctttgccca gatcgagaat 960

gacaaggatg gcaacatctt cgaccgggcc ctggagctga tctctagcta cgccgagtat 1020

gataccgagc ggatctatat cagacaggcc gacatcaata gagtgtccaa cgtgatcttt 1080

ggagagtggg gcaccctggg aggcctgatg agggagtaca aggccgactc tatcaatgat 1140

atcaacctgg agcgcacatg caagaaggtg gacaagtggc tggattctaa ggagtttgcc 1200

ctgagcgatg tgctggaggc catcaagagg accggcaaca atgacgcctt caacgagtat 1260

atctccaaga tgcggacagc cagagagaag atcgatgccg cccgcaagga gatgaagttc 1320

atcagcgaga agatctccgg cgatgaggag tctatccaca tcatcaagac cctgctggac 1380

agcgtgcagc agttcctgca cttctttaat ctgtttaagg caaggcagga catcccactg 1440

gatggagcct tctacgccga gtttgacgag gtgcacagca agctgtttgc catcgtgccc 1500

ctgtataaca aggtgcggaa ctatctgacc aagaacaatc tgaacacaaa gaagatcaag 1560

ctgaatttca agaaccctac actggccaat ggctgggacc agaacaaggt gtacgattat 1620

gcctccctga tctttctgcg ggacggcaat tactatctgg gcatcatcaa tcctaagaga 1680

aagaagaaca tcaagttcga gcagggctct ggcaacggcc ccttctaccg gaagatggtg 1740

tataagcaga tccccggccc taataagaac ctgccaagag tgttcctgac ctccacaaag 1800

ggcaagaagg agtataagcc ctctaaggag atcatcgagg gctacgaggc cgacaagcac 1860

atcaggggcg ataagttcga cctggatttt tgtcacaagc tgatcgattt ctttaaggag 1920

tccatcgaga agcacaagga ctggtctaag ttcaacttct acttcagccc aaccgagagc 1980

tatggcgaca tctctgagtt ctacctggat gtggagaagc agggctatcg catgcacttt 2040

gagaatatca gcgccgagac aatcgacgag tatgtggaga agggcgatct gtttctgttc 2100

cagatctaca acaaggattt tgtgaaggcc gccaccggca agaaggacat gcacacaatc 2160

tactggaatg ccgccttcag ccccgagaac ctgcaggacg tggtggtgaa gctgaacggc 2220

gaggccgagc tgttttatag ggacaagtcc gatatcaagg agatcgtgca ccgcgagggc 2280

gagatcctgg tgaataggac ctacaacggc cgcacaccag tgcccgacaa gatccacaag 2340

aagctgaccg attatcacaa tggccggaca aaggacctgg gcgaggccaa ggagtacctg 2400

gataaggtga gatacttcaa ggcccactat gacatcacca aggatcggag atacctgaac 2460

gacaagatct atttccacgt gcctctgacc ctgaacttca aggccaacgg caagaagaat 2520

ctgaacaaga tggtcatcga gaagttcctg tccgatgaga aggcccacat catcggcatc 2580

gacaggggcg agcgcaatct gctgtactat tccatcatcg acaggtctgg caagatcatc 2640

gatcagcaga gcctgaatgt gatcgacggc tttgattatc gggagaagct gaaccagaga 2700

gagatcgaga tgaaggatgc ccgccagtct tggaacgcca tcggcaagat caaggacctg 2760

aaggagggct acctgagcaa ggccgtgcac gagatcacca agatggccat ccagtataat 2820

gccatcgtgg tcatggagga gctgaactac ggcttcaagc ggggccggtt caaggtggag 2880

aagcagatct atcagaagtt cgagaatatg ctgatcgata agatgaacta cctggtgttt 2940

aaggacgcac ctgatgagtc cccaggaggc gtgctgaatg cctaccagct gacaaaccca 3000

ctggagtctt tcgccaagct gggcaagcag accggcatcc tgttttacgt gccagccgcc 3060

tatacatcca agatcgaccc caccacaggc ttcgtgaatc tgtttaacac ctcctctaag 3120

acaaacgccc aggagcggaa ggagttcctg cagaagtttg agagcatctc ctattctgcc 3180

aaggatggcg gcatctttgc cttcgccttt gactacagaa agttcggcac cagcaagaca 3240

gatcacaaga acgtgtggac cgcctataca aacggcgaga ggatgcgcta catcaaggag 3300

aagaagcgga atgagctgtt tgacccttct aaggagatca aggaggccct gaccagctcc 3360

ggcatcaagt acgatggcgg ccagaacatc ctgccagaca tcctgaggag caacaataac 3420

ggcctgatct acacaatgta ttctagcttc atcgccgcca tccagatgcg cgtgtacgac 3480

ggcaaggagg attatatcat cagccccatc aagaactcca agggcgagtt ctttaggacc 3540

gaccccaaga ggcgcgagct gcctatcgac gccgatgcca atggcgccta caacatcgcc 3600

ctgaggggag agctgacaat gagggcaatc gcagagaagt tcgaccctga tagcgagaag 3660

atggccaagc tggagctgaa gcacaaggat tggttcgagt ttatgcagac cagaggcgac 3720

aaaaggccgg cggccacgaa aaaggccggc caggcaaaaa agaaaaaggg atcctaccca 3780

tacgatgttc cagattacgc ttatccctac gacgtgcctg attatgcata cccatatgat 3840

gtccccgact atgcctaaga attc 3864

<210> 1265
<211> 1238
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 1265

Met Asn Asn Tyr Asp Glu Phe Thr Lys Leu Tyr Pro Ile Gln Lys Thr
1 5 10 15

Ile Arg Phe Glu Leu Lys Pro Gln Gly Arg Thr Met Glu His Leu Glu
20 25 30

Thr Phe Asn Phe Phe Glu Glu Asp Arg Asp Arg Ala Glu Lys Tyr Lys
35 40 45

```
Ile Leu Lys Glu Ala Ile Asp Glu Tyr His Lys Lys Phe Ile Asp Glu
    50                  55                  60

His Leu Thr Asn Met Ser Leu Asp Trp Asn Ser Leu Lys Gln Ile Ser
65                  70                  75                  80

Glu Lys Tyr Tyr Lys Ser Arg Glu Glu Lys Asp Lys Lys Val Phe Leu
                85                  90                  95

Ser Glu Gln Lys Arg Met Arg Gln Glu Ile Val Ser Glu Phe Lys Lys
            100                 105                 110

Asp Asp Arg Phe Lys Asp Leu Phe Ser Lys Lys Leu Phe Ser Glu Leu
        115                 120                 125

Leu Lys Glu Glu Ile Tyr Lys Lys Gly Asn His Gln Glu Ile Asp Ala
    130                 135                 140

Leu Lys Ser Phe Asp Lys Phe Ser Gly Tyr Phe Ile Gly Leu His Glu
145                 150                 155                 160

Asn Arg Lys Asn Met Tyr Ser Asp Gly Asp Glu Ile Thr Ala Ile Ser
                165                 170                 175

Asn Arg Ile Val Asn Glu Asn Phe Pro Lys Phe Leu Asp Asn Leu Gln
            180                 185                 190

Lys Tyr Gln Glu Ala Arg Lys Lys Tyr Pro Glu Trp Ile Ile Lys Ala
        195                 200                 205

Glu Ser Ala Leu Val Ala His Asn Ile Lys Met Asp Glu Val Phe Ser
    210                 215                 220

Leu Glu Tyr Phe Asn Lys Val Leu Asn Gln Glu Gly Ile Gln Arg Tyr
225                 230                 235                 240

Asn Leu Ala Leu Gly Gly Tyr Val Thr Lys Ser Gly Glu Lys Met Met
                245                 250                 255

Gly Leu Asn Asp Ala Leu Asn Leu Ala His Gln Ser Glu Lys Ser Ser
            260                 265                 270

Lys Gly Arg Ile His Met Thr Pro Leu Phe Lys Gln Ile Leu Ser Glu
        275                 280                 285

Lys Glu Ser Phe Ser Tyr Ile Pro Asp Val Phe Thr Glu Asp Ser Gln
    290                 295                 300

Leu Leu Pro Ser Ile Gly Gly Phe Phe Ala Gln Ile Glu Asn Asp Lys
305                 310                 315                 320

Asp Gly Asn Ile Phe Asp Arg Ala Leu Glu Leu Ile Ser Ser Tyr Ala
                325                 330                 335

Glu Tyr Asp Thr Glu Arg Ile Tyr Ile Arg Gln Ala Asp Ile Asn Arg
            340                 345                 350

Val Ser Asn Val Ile Phe Gly Glu Trp Gly Thr Leu Gly Gly Leu Met
        355                 360                 365

Arg Glu Tyr Lys Ala Asp Ser Ile Asn Asp Ile Asn Leu Glu Arg Thr
    370                 375                 380
```

```
Cys Lys Lys Val Asp Lys Trp Leu Asp Ser Lys Glu Phe Ala Leu Ser
385                 390                 395                 400

Asp Val Leu Glu Ala Ile Lys Arg Thr Gly Asn Asn Asp Ala Phe Asn
                405                 410                 415

Glu Tyr Ile Ser Lys Met Arg Thr Ala Arg Glu Lys Ile Asp Ala Ala
            420                 425                 430

Arg Lys Glu Met Lys Phe Ile Ser Glu Lys Ile Ser Gly Asp Glu Glu
        435                 440                 445

Ser Ile His Ile Ile Lys Thr Leu Leu Asp Ser Val Gln Gln Phe Leu
    450                 455                 460

His Phe Phe Asn Leu Phe Lys Ala Arg Gln Asp Ile Pro Leu Asp Gly
465                 470                 475                 480

Ala Phe Tyr Ala Glu Phe Asp Glu Val His Ser Lys Leu Phe Ala Ile
                485                 490                 495

Val Pro Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Lys Asn Asn Leu
            500                 505                 510

Asn Thr Lys Lys Ile Lys Leu Asn Phe Lys Asn Pro Thr Leu Ala Asn
        515                 520                 525

Gly Trp Asp Gln Asn Lys Val Tyr Asp Tyr Ala Ser Leu Ile Phe Leu
    530                 535                 540

Arg Asp Gly Asn Tyr Tyr Leu Gly Ile Ile Asn Pro Lys Arg Lys Lys
545                 550                 555                 560

Asn Ile Lys Phe Glu Gln Gly Ser Gly Asn Gly Pro Phe Tyr Arg Lys
                565                 570                 575

Met Val Tyr Lys Gln Ile Pro Gly Pro Asn Lys Asn Leu Pro Arg Val
            580                 585                 590

Phe Leu Thr Ser Thr Lys Gly Lys Lys Glu Tyr Lys Pro Ser Lys Glu
        595                 600                 605

Ile Ile Glu Gly Tyr Glu Ala Asp Lys His Ile Arg Gly Asp Lys Phe
    610                 615                 620

Asp Leu Asp Phe Cys His Lys Leu Ile Asp Phe Phe Lys Glu Ser Ile
625                 630                 635                 640

Glu Lys His Lys Asp Trp Ser Lys Phe Asn Phe Tyr Phe Ser Pro Thr
                645                 650                 655

Glu Ser Tyr Gly Asp Ile Ser Glu Phe Tyr Leu Asp Val Glu Lys Gln
            660                 665                 670

Gly Tyr Arg Met His Phe Glu Asn Ile Ser Ala Glu Thr Ile Asp Glu
        675                 680                 685

Tyr Val Glu Lys Gly Asp Leu Phe Leu Phe Gln Ile Tyr Asn Lys Asp
    690                 695                 700

Phe Val Lys Ala Ala Thr Gly Lys Lys Asp Met His Thr Ile Tyr Trp
705                 710                 715                 720
```

```
Asn Ala Ala Phe Ser Pro Glu Asn Leu Gln Asp Val Val Val Lys Leu
                725                 730                 735

Asn Gly Glu Ala Glu Leu Phe Tyr Arg Asp Lys Ser Asp Ile Lys Glu
            740                 745                 750

Ile Val His Arg Glu Gly Glu Ile Leu Val Asn Arg Thr Tyr Asn Gly
        755                 760                 765

Arg Thr Pro Val Pro Asp Lys Ile His Lys Lys Leu Thr Asp Tyr His
    770                 775                 780

Asn Gly Arg Thr Lys Asp Leu Gly Glu Ala Lys Glu Tyr Leu Asp Lys
785                 790                 795                 800

Val Arg Tyr Phe Lys Ala His Tyr Asp Ile Thr Lys Asp Arg Arg Tyr
                805                 810                 815

Leu Asn Asp Lys Ile Tyr Phe His Val Pro Leu Thr Leu Asn Phe Lys
            820                 825                 830

Ala Asn Gly Lys Lys Asn Leu Asn Lys Met Val Ile Glu Lys Phe Leu
        835                 840                 845

Ser Asp Glu Lys Ala His Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn
    850                 855                 860

Leu Leu Tyr Tyr Ser Ile Ile Asp Arg Ser Gly Lys Ile Ile Asp Gln
865                 870                 875                 880

Gln Ser Leu Asn Val Ile Asp Gly Phe Asp Tyr Arg Glu Lys Leu Asn
                885                 890                 895

Gln Arg Glu Ile Glu Met Lys Asp Ala Arg Gln Ser Trp Asn Ala Ile
            900                 905                 910

Gly Lys Ile Lys Asp Leu Lys Glu Gly Tyr Leu Ser Lys Ala Val His
        915                 920                 925

Glu Ile Thr Lys Met Ala Ile Gln Tyr Asn Ala Ile Val Val Met Glu
    930                 935                 940

Glu Leu Asn Tyr Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln
945                 950                 955                 960

Ile Tyr Gln Lys Phe Glu Asn Met Leu Ile Asp Lys Met Asn Tyr Leu
                965                 970                 975

Val Phe Lys Asp Ala Pro Asp Glu Ser Pro Gly Gly Val Leu Asn Ala
            980                 985                 990

Tyr Gln Leu Thr Asn Pro Leu Glu Ser Phe Ala Lys Leu Gly Lys Gln
        995                 1000                1005

Thr Gly Ile Leu Phe Tyr Val Pro Ala Ala Tyr Thr Ser Lys Ile
    1010                1015                1020

Asp Pro Thr Thr Gly Phe Val Asn Leu Phe Asn Thr Ser Ser Lys
    1025                1030                1035

Thr Asn Ala Gln Glu Arg Lys Glu Phe Leu Gln Lys Phe Glu Ser
    1040                1045                1050
```

Ile Ser Tyr Ser Ala Lys Asp Gly Gly Ile Phe Ala Phe Ala Phe
1055 1060 1065

Asp Tyr Arg Lys Phe Gly Thr Ser Lys Thr Asp His Lys Asn Val
1070 1075 1080

Trp Thr Ala Tyr Thr Asn Gly Glu Arg Met Arg Tyr Ile Lys Glu
1085 1090 1095

Lys Lys Arg Asn Glu Leu Phe Asp Pro Ser Lys Glu Ile Lys Glu
1100 1105 1110

Ala Leu Thr Ser Ser Gly Ile Lys Tyr Asp Gly Gly Gln Asn Ile
1115 1120 1125

Leu Pro Asp Ile Leu Arg Ser Asn Asn Asn Gly Leu Ile Tyr Thr
1130 1135 1140

Met Tyr Ser Ser Phe Ile Ala Ala Ile Gln Met Arg Val Tyr Asp
1145 1150 1155

Gly Lys Glu Asp Tyr Ile Ile Ser Pro Ile Lys Asn Ser Lys Gly
1160 1165 1170

Glu Phe Phe Arg Thr Asp Pro Lys Arg Arg Glu Leu Pro Ile Asp
1175 1180 1185

Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala Leu Arg Gly Glu Leu
1190 1195 1200

Thr Met Arg Ala Ile Ala Glu Lys Phe Asp Pro Asp Ser Glu Lys
1205 1210 1215

Met Ala Lys Leu Glu Leu Lys His Lys Asp Trp Phe Glu Phe Met
1220 1225 1230

Gln Thr Arg Gly Asp
1235

<210> 1266
<211> 3996
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1266
ggtaccatga acggcaatag gtccatcgtg taccgcgagt tcgtgggcgt gatccccgtg 60

gccaagaccc tgaggaatga gctgcgccct gtgggccaca cacaggagca catcatccag 120

aacggcctga tccaggagga cgagctgcgg caggagaaga gcaccgagct gaagaacatc 180

atggacgatt actatagaga gtacatcgat aagtctctga gcggcgtgac cgacctggac 240

ttcaccctgc tgttcgagct gatgaacctg gtgcagagct ccccctccaa ggacaataag 300

aaggccctgg agaaggagca gtctaagatg agggagcaga tctgcaccca cctgcagtcc 360

gactctaact acaagaatat ctttaacgcc aagctgctga aggagatcct gcctgatttc 420

atcaagaact acaatcagta tgacgtgaag gataaggccg gcaagctgga gacactggcc 480

ctgtttaatg gcttcagcac atactttacc gacttctttg agaagaggaa gaacgtgttc 540

accaaggagg ccgtgagcac atccatcgcc taccgcatcg tgcacgagaa ctccctgatc 600

ttcctggcca atatgacctc ttataagaag atcagcgaga aggccctgga tgagatcgaa 660

gtgatcgaga agaacaatca ggacaagatg ggcgattggg agctgaatca gatctttaac 720

cctgacttct acaatatggt gctgatccag tccggcatcg acttctacaa cgagatctgc 780

ggcgtggtga atgcccacat gaacctgtac tgtcagcaga ccaagaacaa ttataacctg 840

ttcaagatgc ggaagctgca caagcagatc ctggcctaca ccagcaccag cttcgaggtg 900

cccaagatgt tcgaggacga tatgagcgtg tataacgccg tgaacgcctt catcgacgag 960

acagagaagg gcaacatcat cggcaagctg aaggatatcg tgaataagta cgacgagctg 1020

gatgagaaga gaatctatat cagcaaggac ttttacgaga cactgagctg cttcatgtcc 1080

ggcaactgga atctgatcac aggctgcgtg gagaacttct acgatgagaa catccacgcc 1140

aagggcaagt ccaaggagga gaaggtgaag aaggccgtga aggaggacaa gtacaagtct 1200

atcaatgacg tgaacgatct ggtggagaag tatatcgatg agaaggagag gaatgagttc 1260

aagaacagca atgccaagca gtacatccgc gagatctcca acatcatcac cgacacagag 1320

acagcccacc tggagtatga cgatcacatc tctctgatcg agagcgagga gaaggccgac 1380

gagatgaaga agcggctgga tatgtatatg aacatgtacc actgggccaa ggcctttatc 1440

gtggacgagg tgctggacag agatgagatg ttctacagcg atatcgacga tatctataat 1500

atcctggaga acatcgtgcc actgtataat cgggtgagaa actacgtgac ccagaagccc 1560

tacaactcta agaagatcaa gctgaatttc cagagcccta cactggccaa tggctggtcc 1620

cagtctaagg agttcgacaa caatgccatc atcctgatca gagataacaa gtactatctg 1680

gccatcttca atgccaagaa caagccagac aagaagatca tccagggcaa ctccgataag 1740

aagaacgaca acgattacaa gaagatggtg tataacctgc tgccaggcgc caacaagatg 1800

ctgcccaagg tgtttctgtc taagaagggc atcgagacat tcaagccctc cgactatatc 1860

atctctggct acaacgccca caagcacatc aagacaagcg agaattttga tatctccttc 1920

tgtcgggacc tgatcgatta cttcaagaac agcatcgaga agcacgccga gtggagaaag 1980

tatgagttca agttttccgc caccgacagc tactccgata tctctgagtt ctatcgggag 2040

gtggagatgc agggctacag aatcgactgg acatatatca gcgaggccga catcaacaag 2100

ctggatgagg agggcaagat ctatctgttt cagatctaca ataaggattt cgccgagaac 2160

agcaccggca aggagaatct gcacacaatg tactttaaga acatcttctc cgaggagaat 2220

ctgaaggaca tcatcatcaa gctgaacggc caggccgagc tgttttatcg gagagcctct 2280

gtgaagaatc ccgtgaagca caagaaggat agcgtgctgg tgaacaagac ctacaagaat 2340

cagctggaca acggcgacgt ggtgagaatc cccatccctg acgatatcta taacgagatc 2400

tacaagatgt ataatggcta catcaaggag tccgacctgt ctgaggccgc caaggagtac 2460

ctggataagg tggaggtgag gaccgcccag aaggacatcg tgaaggatta ccgctataca 2520

gtggacaagt acttcatcca cacacctatc accatcaact ataaggtgac cgcccgcaac 2580

aatgtgaatg atatggtggt gaagtacatc gcccagaacg acgatatcca cgtgatcggc 2640

atcgaccggg gcgagagaaa cctgatctac atctccgtga tcgattctca cggcaacatc 2700

gtgaagcaga aatcctacaa catcctgaac aactacgact acaagaagaa gctggtggag 2760

aaggagaaaa cccgggagta cgccagaaag aactggaaga gcatcggcaa tatcaaggag 2820

ctgaaggagg gctatatctc cggcgtggtg cacgagatcg ccatgctgat cgtggagtac 2880

aacgccatca tcgccatgga ggacctgaat tatggcttta agaggggccg cttcaaggtg 2940

gagcggcagg tgtaccagaa gtttgagagc atgctgatca ataagctgaa ctatttcgcc 3000

agcaaggaga agtccgtgga cgagccagga ggcctgctga agggctatca gctgacctac 3060

gtgcccgata atatcaagaa cctgggcaag cagtgcggcg tgatctttta cgtgcctgcc 3120

gccttcacca gcaagatcga cccatccaca ggctttatct ctgccttcaa ctttaagtct 3180

atcagcacaa atgcctctcg gaagcagttc tttatgcagt ttgacgagat cagatactgt 3240

gccgagaagg atatgttcag ctttggcttc gactacaaca acttcgatac ctacaacatc 3300

acaatgggca agacacagtg gaccgtgtat acaaacggcg agagactgca gtctgagttc 3360

aacaatgcca ggcgcaccgg caagacaaag agcatcaatc tgacagagac aatcaagctg 3420

ctgctggagg acaatgagat caactacgcc gacggccacg atatcaggat cgatatggag 3480

aagatggacg aggataagaa gagcgagttc tttgcccagc tgctgagcct gtataagctg 3540

accgtgcaga tgcgcaattc ctatacagag gccgaggagc aggagaacgg catctcttac 3600

gacaagatca tcagccctgt gatcaatgat gagggcgagt tctttgactc cgataactat 3660

aaggagtctg acgataagga gtgcaagatg ccaaaggacg ccgatgccaa cggcgcctac 3720

tgtatcgccc tgaagggcct gtatgaggtg ctgaagatca agagcgagtg gaccgaggac 3780

ggctttgata ggaattgcct gaagctgcca cacgcagagt ggctggactt catccagaac 3840

aagcggtacg agaaaaggcc ggcggccacg aaaaaggccg gccaggcaaa aaagaaaaag 3900

ggatcctacc catacgatgt tccagattac gcttatccct acgacgtgcc tgattatgca 3960

tacccatatg atgtccccga ctatgcctaa gaattc 3996

<210> 1267
<211> 1282
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 1267

Met	Asn	Gly	Asn	Arg	Ser	Ile	Val	Tyr	Arg	Glu	Phe	Val	Gly	Val	Ile
1				5				10						15	
Pro	Val	Ala	Lys	Thr	Leu	Arg	Asn	Glu	Leu	Arg	Pro	Val	Gly	His	Thr
			20					25					30		
Gln	Glu	His	Ile	Ile	Gln	Asn	Gly	Leu	Ile	Gln	Glu	Asp	Glu	Leu	Arg
		35					40					45			
Gln	Glu	Lys	Ser	Thr	Glu	Leu	Lys	Asn	Ile	Met	Asp	Asp	Tyr	Tyr	Arg
	50					55					60				
Glu	Tyr	Ile	Asp	Lys	Ser	Leu	Ser	Gly	Val	Thr	Asp	Leu	Asp	Phe	Thr
65					70					75					80
Leu	Leu	Phe	Glu	Leu	Met	Asn	Leu	Val	Gln	Ser	Ser	Pro	Ser	Lys	Asp
				85					90					95	
Asn	Lys	Lys	Ala	Leu	Glu	Lys	Glu	Gln	Ser	Lys	Met	Arg	Glu	Gln	Ile
			100					105					110		
Cys	Thr	His	Leu	Gln	Ser	Asp	Ser	Asn	Tyr	Lys	Asn	Ile	Phe	Asn	Ala
		115					120					125			
Lys	Leu	Leu	Lys	Glu	Ile	Leu	Pro	Asp	Phe	Ile	Lys	Asn	Tyr	Asn	Gln
	130					135					140				
Tyr	Asp	Val	Lys	Asp	Lys	Ala	Gly	Lys	Leu	Glu	Thr	Leu	Ala	Leu	Phe
145					150					155					160
Asn	Gly	Phe	Ser	Thr	Tyr	Phe	Thr	Asp	Phe	Phe	Glu	Lys	Arg	Lys	Asn
				165					170					175	

```
Val Phe Thr Lys Glu Ala Val Ser Thr Ser Ile Ala Tyr Arg Ile Val
            180                 185                 190

His Glu Asn Ser Leu Ile Phe Leu Ala Asn Met Thr Ser Tyr Lys Lys
        195                 200                 205

Ile Ser Glu Lys Ala Leu Asp Glu Ile Glu Val Ile Glu Lys Asn Asn
    210                 215                 220

Gln Asp Lys Met Gly Asp Trp Glu Leu Asn Gln Ile Phe Asn Pro Asp
225                 230                 235                 240

Phe Tyr Asn Met Val Leu Ile Gln Ser Gly Ile Asp Phe Tyr Asn Glu
                245                 250                 255

Ile Cys Gly Val Val Asn Ala His Met Asn Leu Tyr Cys Gln Gln Thr
            260                 265                 270

Lys Asn Asn Tyr Asn Leu Phe Lys Met Arg Lys Leu His Lys Gln Ile
        275                 280                 285

Leu Ala Tyr Thr Ser Thr Ser Phe Glu Val Pro Lys Met Phe Glu Asp
    290                 295                 300

Asp Met Ser Val Tyr Asn Ala Val Asn Ala Phe Ile Asp Glu Thr Glu
305                 310                 315                 320

Lys Gly Asn Ile Ile Gly Lys Leu Lys Asp Ile Val Asn Lys Tyr Asp
                325                 330                 335

Glu Leu Asp Glu Lys Arg Ile Tyr Ile Ser Lys Asp Phe Tyr Glu Thr
            340                 345                 350

Leu Ser Cys Phe Met Ser Gly Asn Trp Asn Leu Ile Thr Gly Cys Val
        355                 360                 365

Glu Asn Phe Tyr Asp Glu Asn Ile His Ala Lys Gly Lys Ser Lys Glu
    370                 375                 380

Glu Lys Val Lys Lys Ala Val Lys Glu Asp Lys Tyr Lys Ser Ile Asn
385                 390                 395                 400

Asp Val Asn Asp Leu Val Glu Lys Tyr Ile Asp Glu Lys Glu Arg Asn
                405                 410                 415

Glu Phe Lys Asn Ser Asn Ala Lys Gln Tyr Ile Arg Glu Ile Ser Asn
            420                 425                 430

Ile Ile Thr Asp Thr Glu Thr Ala His Leu Glu Tyr Asp Asp His Ile
        435                 440                 445

Ser Leu Ile Glu Ser Glu Glu Lys Ala Asp Glu Met Lys Lys Arg Leu
    450                 455                 460

Asp Met Tyr Met Asn Met Tyr His Trp Ala Lys Ala Phe Ile Val Asp
465                 470                 475                 480

Glu Val Leu Asp Arg Asp Glu Met Phe Tyr Ser Asp Ile Asp Asp Ile
                485                 490                 495

Tyr Asn Ile Leu Glu Asn Ile Val Pro Leu Tyr Asn Arg Val Arg Asn
            500                 505                 510
```

Tyr Val Thr Gln Lys Pro Tyr Asn Ser Lys Lys Ile Lys Leu Asn Phe
515 520 525

Gln Ser Pro Thr Leu Ala Asn Gly Trp Ser Gln Ser Lys Glu Phe Asp
530 535 540

Asn Asn Ala Ile Ile Leu Ile Arg Asp Asn Lys Tyr Tyr Leu Ala Ile
545 550 555 560

Phe Asn Ala Lys Asn Lys Pro Asp Lys Lys Ile Ile Gln Gly Asn Ser
565 570 575

Asp Lys Lys Asn Asp Asn Asp Tyr Lys Lys Met Val Tyr Asn Leu Leu
580 585 590

Pro Gly Ala Asn Lys Met Leu Pro Lys Val Phe Leu Ser Lys Lys Gly
595 600 605

Ile Glu Thr Phe Lys Pro Ser Asp Tyr Ile Ile Ser Gly Tyr Asn Ala
610 615 620

His Lys His Ile Lys Thr Ser Glu Asn Phe Asp Ile Ser Phe Cys Arg
625 630 635 640

Asp Leu Ile Asp Tyr Phe Lys Asn Ser Ile Glu Lys His Ala Glu Trp
645 650 655

Arg Lys Tyr Glu Phe Lys Phe Ser Ala Thr Asp Ser Tyr Ser Asp Ile
660 665 670

Ser Glu Phe Tyr Arg Glu Val Glu Met Gln Gly Tyr Arg Ile Asp Trp
675 680 685

Thr Tyr Ile Ser Glu Ala Asp Ile Asn Lys Leu Asp Glu Glu Gly Lys
690 695 700

Ile Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ala Glu Asn Ser Thr
705 710 715 720

Gly Lys Glu Asn Leu His Thr Met Tyr Phe Lys Asn Ile Phe Ser Glu
725 730 735

Glu Asn Leu Lys Asp Ile Ile Ile Lys Leu Asn Gly Gln Ala Glu Leu
740 745 750

Phe Tyr Arg Arg Ala Ser Val Lys Asn Pro Val Lys His Lys Lys Asp
755 760 765

Ser Val Leu Val Asn Lys Thr Tyr Lys Asn Gln Leu Asp Asn Gly Asp
770 775 780

Val Val Arg Ile Pro Ile Pro Asp Asp Ile Tyr Asn Glu Ile Tyr Lys
785 790 795 800

Met Tyr Asn Gly Tyr Ile Lys Glu Ser Asp Leu Ser Glu Ala Ala Lys
805 810 815

Glu Tyr Leu Asp Lys Val Glu Val Arg Thr Ala Gln Lys Asp Ile Val
820 825 830

Lys Asp Tyr Arg Tyr Thr Val Asp Lys Tyr Phe Ile His Thr Pro Ile
835 840 845

```
Thr Ile Asn Tyr Lys Val Thr Ala Arg Asn Asn Val Asn Asp Met Val
    850                 855                 860

Val Lys Tyr Ile Ala Gln Asn Asp Asp Ile His Val Ile Gly Ile Asp
865                 870                 875                 880

Arg Gly Glu Arg Asn Leu Ile Tyr Ile Ser Val Ile Asp Ser His Gly
                885                 890                 895

Asn Ile Val Lys Gln Lys Ser Tyr Asn Ile Leu Asn Asn Tyr Asp Tyr
            900                 905                 910

Lys Lys Lys Leu Val Glu Lys Glu Lys Thr Arg Glu Tyr Ala Arg Lys
        915                 920                 925

Asn Trp Lys Ser Ile Gly Asn Ile Lys Glu Leu Lys Glu Gly Tyr Ile
    930                 935                 940

Ser Gly Val Val His Glu Ile Ala Met Leu Ile Val Glu Tyr Asn Ala
945                 950                 955                 960

Ile Ile Ala Met Glu Asp Leu Asn Tyr Gly Phe Lys Arg Gly Arg Phe
                965                 970                 975

Lys Val Glu Arg Gln Val Tyr Gln Lys Phe Glu Ser Met Leu Ile Asn
            980                 985                 990

Lys Leu Asn Tyr Phe Ala Ser Lys Glu Lys Ser Val Asp Glu Pro Gly
        995                 1000                1005

Gly Leu  Leu Lys Gly Tyr Gln  Leu Thr Tyr Val Pro  Asp Asn Ile
    1010                1015                1020

Lys Asn  Leu Gly Lys Gln Cys  Gly Val Ile Phe Tyr  Val Pro Ala
    1025                1030                1035

Ala Phe  Thr Ser Lys Ile Asp  Pro Ser Thr Gly Phe  Ile Ser Ala
    1040                1045                1050

Phe Asn  Phe Lys Ser Ile Ser  Thr Asn Ala Ser Arg  Lys Gln Phe
    1055                1060                1065

Phe Met  Gln Phe Asp Glu Ile  Arg Tyr Cys Ala Glu  Lys Asp Met
    1070                1075                1080

Phe Ser  Phe Gly Phe Asp Tyr  Asn Asn Phe Asp Thr  Tyr Asn Ile
    1085                1090                1095

Thr Met  Gly Lys Thr Gln Trp  Thr Val Tyr Thr Asn  Gly Glu Arg
    1100                1105                1110

Leu Gln  Ser Glu Phe Asn Asn  Ala Arg Arg Thr Gly  Lys Thr Lys
    1115                1120                1125

Ser Ile  Asn Leu Thr Glu Thr  Ile Lys Leu Leu Leu  Glu Asp Asn
    1130                1135                1140

Glu Ile  Asn Tyr Ala Asp Gly  His Asp Ile Arg Ile  Asp Met Glu
    1145                1150                1155

Lys Met  Asp Glu Asp Lys Lys  Ser Glu Phe Phe Ala  Gln Leu Leu
    1160                1165                1170
```

Ser Leu Tyr Lys Leu Thr Val Gln Met Arg Asn Ser Tyr Thr Glu
1175 1180 1185

Ala Glu Glu Gln Glu Asn Gly Ile Ser Tyr Asp Lys Ile Ile Ser
1190 1195 1200

Pro Val Ile Asn Asp Glu Gly Glu Phe Phe Asp Ser Asp Asn Tyr
1205 1210 1215

Lys Glu Ser Asp Asp Lys Glu Cys Lys Met Pro Lys Asp Ala Asp
1220 1225 1230

Ala Asn Gly Ala Tyr Cys Ile Ala Leu Lys Gly Leu Tyr Glu Val
1235 1240 1245

Leu Lys Ile Lys Ser Glu Trp Thr Glu Asp Gly Phe Asp Arg Asn
1250 1255 1260

Cys Leu Lys Leu Pro His Ala Glu Trp Leu Asp Phe Ile Gln Asn
1265 1270 1275

Lys Arg Tyr Glu
1280

<210> 1268
<211> 4269
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1268
ggtaccatgc tgttccagga ctttacccac ctgtatccac tgtccaagac agtgagattt 60

gagctgaagc ccatcgatag gaccctggag cacatccacg ccaagaactt cctgtctcag 120

gacgagacaa tggccgatat gcaccagaag gtgaaagtga tcctggacga ttaccaccgc 180

gacttcatcg ccgatatgat gggcgaggtg aagctgacca agctggccga gttctatgac 240

gtgtacctga agtttcggaa gaacccaaag gacgatgagc tgcagaagca gctgaaggat 300

ctgcaggccg tgctgagaaa ggagatcgtg aagcccatcg gcaatggcgg caagtataag 360

gccggctacg acaggctgtt cggcgccaag ctgtttaagg acggcaagga gctgggcgat 420

ctggccaagt tcgtgatcgc acaggaggga gagagctccc caaagctggc ccacctggcc 480

cacttcgaga agttttccac ctatttcaca ggctttcacg ataaccggaa gaatatgtat 540

tctgacgagg ataagcacac cgccatcgcc taccgcctga tccacgagaa cctgccccgg 600

tttatcgaca atctgcagat cctgaccaca atcaagcaga agcactctgc cctgtacgat 660

cagatcatca acgagctgac cgccagcggc ctggacgtgt ctctggccag ccacctggat 720

ggctatcaca agctgctgac acaggagggc atcaccgcct acaatacact gctgggagga 780

atctccggag aggcaggctc tcctaagatc cagggcatca acgagctgat caattctcac 840

cacaaccagc actgccacaa gagcgagaga atcgccaagc tgaggccact gcacaagcag 900

atcctgtccg acggcatgag cgtgtccttc ctgccctcta agtttgccga cgatagcgag 960

atgtgccagg ccgtgaacga gttctatcgc cactacgccg acgtgttcgc caaggtgcag 1020

agcctgttcg acggctttga cgatcaccag aaggatggca tctacgtgga gcacaagaac 1080

ctgaatgagc tgtccaagca ggccttcggc gactttgcac tgctgggacg cgtgctggac 1140

ggatactatg tggatgtggt gaatccagag ttcaacgagc ggtttgccaa ggccaagacc 1200

gacaatgcca aggccaagct gacaaaggag aaggataagt tcatcaaggg cgtgcactcc 1260

ctggcctctc tggagcaggc catcgagcac tataccgcaa ggcacgacga tgagagcgtg 1320

caggcaggca agctgggaca gtacttcaag cacggcctgg ccggagtgga caaccccatc 1380

cagaagatcc acaacaatca cagcaccatc aagggctttc tggagaggga gcgccctgca 1440

ggagagagag ccctgccaaa gatcaagtcc ggcaagaatc ctgagatgac acagctgagg 1500

cagctgaagg agctgctgga taacgccctg aatgtggccc acttcgccaa gctgctgacc 1560

acaaagacca cactggacaa tcaggatggc aacttctatg gcgagtttgg cgtgctgtac 1620

gacgagctgg ccaagatccc caccctgtat aacaaggtga gagattacct gagccagaag 1680

cctttctcca ccgagaagta caagctgaac tttggcaatc caacactgct gaatggctgg 1740

gacctgaaca aggagaagga taatttcggc gtgatcctgc agaaggacgg ctgctactat 1800

ctggccctgc tggacaaggc ccacaagaag gtgtttgata acgcccctaa tacaggcaag 1860

agcatctatc agaagatgat ctataagtac ctggaggtga ggaagcagtt ccccaaggtg 1920

ttcttttcca aggaggccat cgccatcaac taccacccтt ctaaggagct ggtggagatc 1980

aaggacaagg gccggcagag atccgacgat gagcgcctga agctgtatcg gtttatcctg 2040

gagtgtctga agatccaccc taagtacgat aagaagttcg agggcgccat cggcgacatc 2100

cagctgttta agaaggataa gaagggcaga gaggtgccaa tcagcgagaa ggacctgttc 2160

gataagatca acggcatctt ttctagcaag cctaagctgg agatggagga cttctttatc 2220

ggcgagttca agaggtataa cccaagccag gacctggtgg atcagtataa tatctacaag 2280

aagatcgact ccaacgataa tcgcaagaag gagaatttct acaacaatca ccccaagttt 2340

aagaaggatc tggtgcggta ctattacgag tctatgtgca agcacgagga gtgggaggag 2400

agcttcgagt tttccaagaa gctgcaggac atcggctgtt acgtggatgt gaacgagctg 2460

tttaccgaga tcgagacacg gagactgaat tataagatct ccttctgcaa catcaatgcc 2520

gactacatcg atgagctggt ggagcagggc cagctgtatc tgttccagat ctacaacaag 2580

gacttttccc caaaggccca cggcaagccc aatctgcaca ccctgtactt caaggccctg 2640

ttttctgagg acaacctggc cgatcctatc tataagctga atggcgaggc ccagatcttc 2700

tacagaaagg cctccctgga catgaacgag acaacaatcc acagggccgg cgaggtgctg 2760

gagaacaaga atcccgataa tcctaagaag agacagttcg tgtacgacat catcaaggat 2820

aagaggtaca cacaggacaa gttcatgctg cacgtgccaa tcaccatgaa ctttggcgtg 2880

cagggcatga caatcaagga gttcaataag aaggtgaacc agtctatcca gcagtatgac 2940

gaggtgaacg tgatcggcat cgatcggggc gagagacacc tgctgtacct gaccgtgatc 3000

aatagcaagg gcgagatcct ggagcagtgt tccctgaacg acatcaccac agcctctgcc 3060

aatggcacac agatgaccac accttaccac aagatcctgg ataagaggga gatcgagcgc 3120

ctgaacgccc gggtgggatg gggcgagatc gagacaatca aggagctgaa gtctggctat 3180

ctgagccacg tggtgcacca gatcagccag ctgatgctga agtacaacgc catcgtggtg 3240

ctggaggacc tgaatttcgg ctttaagagg ggccgcttta aggtggagaa gcagatctat 3300

cagaacttcg agaatgccct gatcaagaag ctgaaccacc tggtgctgaa ggacaaggcc 3360

gacgatgaga tcggctctta caagaatgcc ctgcagctga ccaacaattt cacagatctg 3420

aagagcatcg gcaagcagac cggcttcctg ttttatgtgc ccgcctggaa cacctctaag 3480

atcgaccctg agacaggctt tgtggatctg ctgaagccaa gatacgagaa catcgcccag 3540

agccaggcct tctttggcaa gttcgacaag atctgctata atgccgacaa ggattacttc 3600

gagtttcaca tcgactacgc caagtttacc gataaggcca agaatagccg ccagatctgg 3660

acaatctgtt cccacggcga caagcggtac gtgtacgata agacagccaa ccagaataag 3720

ggcgccgcca agggcatcaa cgtgaatgat gagctgaagt ccctgttcgc ccgccaccac 3780

atcaacgaga agcagcccaa cctggtcatg gacatctgcc agaacaatga taaggagttt 3840

cacaagtctc tgatgtacct gctgaaaacc ctgctggccc tgcggtacag caacgcctcc 3900

tctgacgagg atttcatcct gtcccccgtg gcaaacgacg agggcgtgtt ctttaatagc 3960

gccctggccg acgatacaca gcctcagaat gccgatgcca acggcgccta ccacatcgcc 4020

ctgaagggcc tgtggctgct gaatgagctg aagaactccg acgatctgaa caaggtgaag 4080

ctggccatcg acaatcagac ctggctgaat ttcgcccaga acaggaaaag gccggcggcc 4140

acgaaaaagg ccggccaggc aaaaaagaaa aagggatcct acccatacga tgttccagat 4200

tacgcttatc cctacgacgt gcctgattat gcatacccat atgatgtccc cgactatgcc 4260

taagaattc 4269

<210> 1269
<211> 1373
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 1269
Met Leu Phe Gln Asp Phe Thr His Leu Tyr Pro Leu Ser Lys Thr Val
1 5 10 15

Arg Phe Glu Leu Lys Pro Ile Asp Arg Thr Leu Glu His Ile His Ala
20 25 30

Lys Asn Phe Leu Ser Gln Asp Glu Thr Met Ala Asp Met His Gln Lys
35 40 45

Val Lys Val Ile Leu Asp Asp Tyr His Arg Asp Phe Ile Ala Asp Met
50 55 60

Met Gly Glu Val Lys Leu Thr Lys Leu Ala Glu Phe Tyr Asp Val Tyr
65 70 75 80

Leu Lys Phe Arg Lys Asn Pro Lys Asp Asp Glu Leu Gln Lys Gln Leu
85 90 95

Lys Asp Leu Gln Ala Val Leu Arg Lys Glu Ile Val Lys Pro Ile Gly
100 105 110

Asn Gly Gly Lys Tyr Lys Ala Gly Tyr Asp Arg Leu Phe Gly Ala Lys
115 120 125

Leu Phe Lys Asp Gly Lys Glu Leu Gly Asp Leu Ala Lys Phe Val Ile
130 135 140

Ala Gln Glu Gly Glu Ser Ser Pro Lys Leu Ala His Leu Ala His Phe
145 150 155 160

Glu Lys Phe Ser Thr Tyr Phe Thr Gly Phe His Asp Asn Arg Lys Asn
165 170 175

Met Tyr Ser Asp Glu Asp Lys His Thr Ala Ile Ala Tyr Arg Leu Ile
180 185 190

His Glu Asn Leu Pro Arg Phe Ile Asp Asn Leu Gln Ile Leu Thr Thr
195 200 205

Ile Lys Gln Lys His Ser Ala Leu Tyr Asp Gln Ile Ile Asn Glu Leu
210 215 220

```
Thr Ala Ser Gly Leu Asp Val Ser Leu Ala Ser His Leu Asp Gly Tyr
225                 230                 235                 240

His Lys Leu Leu Thr Gln Glu Gly Ile Thr Ala Tyr Asn Thr Leu Leu
                245                 250                 255

Gly Gly Ile Ser Gly Glu Ala Gly Ser Pro Lys Ile Gln Gly Ile Asn
            260                 265                 270

Glu Leu Ile Asn Ser His His Asn Gln His Cys His Lys Ser Glu Arg
        275                 280                 285

Ile Ala Lys Leu Arg Pro Leu His Lys Gln Ile Leu Ser Asp Gly Met
    290                 295                 300

Ser Val Ser Phe Leu Pro Ser Lys Phe Ala Asp Asp Ser Glu Met Cys
305                 310                 315                 320

Gln Ala Val Asn Glu Phe Tyr Arg His Tyr Ala Asp Val Phe Ala Lys
                325                 330                 335

Val Gln Ser Leu Phe Asp Gly Phe Asp Asp His Gln Lys Asp Gly Ile
            340                 345                 350

Tyr Val Glu His Lys Asn Leu Asn Glu Leu Ser Lys Gln Ala Phe Gly
        355                 360                 365

Asp Phe Ala Leu Leu Gly Arg Val Leu Asp Gly Tyr Tyr Val Asp Val
    370                 375                 380

Val Asn Pro Glu Phe Asn Glu Arg Phe Ala Lys Ala Lys Thr Asp Asn
385                 390                 395                 400

Ala Lys Ala Lys Leu Thr Lys Glu Lys Asp Lys Phe Ile Lys Gly Val
                405                 410                 415

His Ser Leu Ala Ser Leu Glu Gln Ala Ile Glu His Tyr Thr Ala Arg
            420                 425                 430

His Asp Asp Glu Ser Val Gln Ala Gly Lys Leu Gly Gln Tyr Phe Lys
        435                 440                 445

His Gly Leu Ala Gly Val Asp Asn Pro Ile Gln Lys Ile His Asn Asn
    450                 455                 460

His Ser Thr Ile Lys Gly Phe Leu Glu Arg Glu Arg Pro Ala Gly Glu
465                 470                 475                 480

Arg Ala Leu Pro Lys Ile Lys Ser Gly Lys Asn Pro Glu Met Thr Gln
                485                 490                 495

Leu Arg Gln Leu Lys Glu Leu Leu Asp Asn Ala Leu Asn Val Ala His
            500                 505                 510

Phe Ala Lys Leu Leu Thr Thr Lys Thr Thr Leu Asp Asn Gln Asp Gly
        515                 520                 525

Asn Phe Tyr Gly Glu Phe Gly Val Leu Tyr Asp Glu Leu Ala Lys Ile
    530                 535                 540

Pro Thr Leu Tyr Asn Lys Val Arg Asp Tyr Leu Ser Gln Lys Pro Phe
545                 550                 555                 560
```

Ser Thr Glu Lys Tyr Lys Leu Asn Phe Gly Asn Pro Thr Leu Leu Asn
565 570 575

Gly Trp Asp Leu Asn Lys Glu Lys Asp Asn Phe Gly Val Ile Leu Gln
580 585 590

Lys Asp Gly Cys Tyr Tyr Leu Ala Leu Leu Asp Lys Ala His Lys Lys
595 600 605

Val Phe Asp Asn Ala Pro Asn Thr Gly Lys Ser Ile Tyr Gln Lys Met
610 615 620

Ile Tyr Lys Tyr Leu Glu Val Arg Lys Gln Phe Pro Lys Val Phe Phe
625 630 635 640

Ser Lys Glu Ala Ile Ala Ile Asn Tyr His Pro Ser Lys Glu Leu Val
645 650 655

Glu Ile Lys Asp Lys Gly Arg Gln Arg Ser Asp Asp Glu Arg Leu Lys
660 665 670

Leu Tyr Arg Phe Ile Leu Glu Cys Leu Lys Ile His Pro Lys Tyr Asp
675 680 685

Lys Lys Phe Glu Gly Ala Ile Gly Asp Ile Gln Leu Phe Lys Lys Asp
690 695 700

Lys Lys Gly Arg Glu Val Pro Ile Ser Glu Lys Asp Leu Phe Asp Lys
705 710 715 720

Ile Asn Gly Ile Phe Ser Ser Lys Pro Lys Leu Glu Met Glu Asp Phe
725 730 735

Phe Ile Gly Glu Phe Lys Arg Tyr Asn Pro Ser Gln Asp Leu Val Asp
740 745 750

Gln Tyr Asn Ile Tyr Lys Lys Ile Asp Ser Asn Asp Asn Arg Lys Lys
755 760 765

Glu Asn Phe Tyr Asn Asn His Pro Lys Phe Lys Lys Asp Leu Val Arg
770 775 780

Tyr Tyr Tyr Glu Ser Met Cys Lys His Glu Glu Trp Glu Glu Ser Phe
785 790 795 800

Glu Phe Ser Lys Lys Leu Gln Asp Ile Gly Cys Tyr Val Asp Val Asn
805 810 815

Glu Leu Phe Thr Glu Ile Glu Thr Arg Arg Leu Asn Tyr Lys Ile Ser
820 825 830

Phe Cys Asn Ile Asn Ala Asp Tyr Ile Asp Glu Leu Val Glu Gln Gly
835 840 845

Gln Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Pro Lys Ala
850 855 860

His Gly Lys Pro Asn Leu His Thr Leu Tyr Phe Lys Ala Leu Phe Ser
865 870 875 880

Glu Asp Asn Leu Ala Asp Pro Ile Tyr Lys Leu Asn Gly Glu Ala Gln
885 890 895

```
Ile Phe Tyr Arg Lys Ala Ser Leu Asp Met Asn Glu Thr Thr Ile His
            900                 905                 910

Arg Ala Gly Glu Val Leu Glu Asn Lys Asn Pro Asp Asn Pro Lys Lys
        915                 920                 925

Arg Gln Phe Val Tyr Asp Ile Ile Lys Asp Lys Arg Tyr Thr Gln Asp
    930                 935                 940

Lys Phe Met Leu His Val Pro Ile Thr Met Asn Phe Gly Val Gln Gly
945                 950                 955                 960

Met Thr Ile Lys Glu Phe Asn Lys Lys Val Asn Gln Ser Ile Gln Gln
                965                 970                 975

Tyr Asp Glu Val Asn Val Ile Gly Ile Asp Arg Gly Glu Arg His Leu
            980                 985                 990

Leu Tyr Leu Thr Val Ile Asn Ser Lys Gly Glu Ile Leu Glu Gln Cys
        995                 1000                1005

Ser Leu  Asn Asp Ile Thr Thr  Ala Ser Ala Asn Gly  Thr Gln Met
    1010                1015                1020

Thr Thr  Pro Tyr His Lys Ile  Leu Asp Lys Arg Glu  Ile Glu Arg
    1025                1030                1035

Leu Asn  Ala Arg Val Gly Trp  Gly Glu Ile Glu Thr  Ile Lys Glu
    1040                1045                1050

Leu Lys  Ser Gly Tyr Leu Ser  His Val Val His Gln  Ile Ser Gln
    1055                1060                1065

Leu Met  Leu Lys Tyr Asn Ala  Ile Val Val Leu Glu  Asp Leu Asn
    1070                1075                1080

Phe Gly  Phe Lys Arg Gly Arg  Phe Lys Val Glu Lys  Gln Ile Tyr
    1085                1090                1095

Gln Asn  Phe Glu Asn Ala Leu  Ile Lys Lys Leu Asn  His Leu Val
    1100                1105                1110

Leu Lys  Asp Lys Ala Asp Asp  Glu Ile Gly Ser Tyr  Lys Asn Ala
    1115                1120                1125

Leu Gln  Leu Thr Asn Asn Phe  Thr Asp Leu Lys Ser  Ile Gly Lys
    1130                1135                1140

Gln Thr  Gly Phe Leu Phe Tyr  Val Pro Ala Trp Asn  Thr Ser Lys
    1145                1150                1155

Ile Asp  Pro Glu Thr Gly Phe  Val Asp Leu Leu Lys  Pro Arg Tyr
    1160                1165                1170

Glu Asn  Ile Ala Gln Ser Gln  Ala Phe Phe Gly Lys  Phe Asp Lys
    1175                1180                1185

Ile Cys  Tyr Asn Ala Asp Lys  Asp Tyr Phe Glu Phe  His Ile Asp
    1190                1195                1200

Tyr Ala  Lys Phe Thr Asp Lys  Ala Lys Asn Ser Arg  Gln Ile Trp
    1205                1210                1215
```

Thr Ile Cys Ser His Gly Asp Lys Arg Tyr Val Tyr Asp Lys Thr
1220 1225 1230

Ala Asn Gln Asn Lys Gly Ala Ala Lys Gly Ile Asn Val Asn Asp
1235 1240 1245

Glu Leu Lys Ser Leu Phe Ala Arg His His Ile Asn Glu Lys Gln
1250 1255 1260

Pro Asn Leu Val Met Asp Ile Cys Gln Asn Asn Asp Lys Glu Phe
1265 1270 1275

His Lys Ser Leu Met Tyr Leu Leu Lys Thr Leu Leu Ala Leu Arg
1280 1285 1290

Tyr Ser Asn Ala Ser Ser Asp Glu Asp Phe Ile Leu Ser Pro Val
1295 1300 1305

Ala Asn Asp Glu Gly Val Phe Phe Asn Ser Ala Leu Ala Asp Asp
1310 1315 1320

Thr Gln Pro Gln Asn Ala Asp Ala Asn Gly Ala Tyr His Ile Ala
1325 1330 1335

Leu Lys Gly Leu Trp Leu Leu Asn Glu Leu Lys Asn Ser Asp Asp
1340 1345 1350

Leu Asn Lys Val Lys Leu Ala Ile Asp Asn Gln Thr Trp Leu Asn
1355 1360 1365

Phe Ala Gln Asn Arg
1370

<210> 1270
<211> 3939
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1270
ggtaccatgg aggactattc cggctttgtg aacatctact ctatccagaa aaccctgagg 60

ttcgagctga agccagtggg caagacactg gagcacatcg agaagaaggg cttcctgaag 120

aaggacaaga tccgggccga ggattacaag gccgtgaaga agatcatcga taagtaccac 180

agagcctata tcgaggaggt gtttgattcc gtgctgcacc agaagaagaa gaaggacaag 240

acccgctttt ctacacagtt catcaaggag atcaaggagt tcagcgagct gtactataag 300

accgagaaga acatccccga caaggagagg ctggaggccc tgagcgagaa gctgcgcaag 360

atgctggtgg gcgcctttaa gggcgagttc tccgaggagg tggccgagaa gtataagaac 420

ctgttttcta aggagctgat caggaatgag atcgagaagt tctgcgagac agacgaggag 480

cgcaagcagg tgtctaactt caagagcttc accacatact ttaccggctt ccactccaac 540

aggcagaata tctattccga cgagaagaag tctacagcca tcggctaccg catcatccac 600

cagaacctgc ctaagttcct ggataatctg aagatcatcg agtccatcca gcggcggttc 660

aaggacttcc catggtctga tctgaagaag aacctgaaga agatcgataa gaatatcaag 720

ctgaccgagt acttcagcat cgacggcttc gtgaacgtgc tgaatcagaa gggcatcgat 780

gcctacaaca caatcctggg cggcaagtcc gaggagtctg gcgagaagat ccagggcctg 840

aacgagtaca tcaatctgta tcggcagaag aacaatatcg acagaaagaa cctgcccaat 900

gtgaagatcc tgtttaagca gatcctgggc gatagggaga caaagagctt tatccctgag 960

gccttcccag acgatcagtc cgtgctgaac tctatcacag agttcgccaa gtacctgaag 1020

ctggataaga agaagaagag catcatcgcc gagctgaaga agtttctgag ctccttcaat 1080

cgctacgagc tggacggcat ctatctggcc aacgataata gcctggcctc tatcagcacc 1140

ttcctgtttg acgattggtc ctttatcaag aagtccgtgt ctttcaagta tgacgagtcc 1200

gtgggcgacc ccaagaagaa gatcaagtct cccctgaagt acgagaagga gaaggagaag 1260

tggctgaagc agaagtacta tacaatctct ttcctgaacg atgccatcga gagctattcc 1320

aagtctcagg acgagaagag ggtgaagatc cgcctggagg cctactttgc cgagttcaag 1380

agcaaggacg atgccaagaa gcagttcgac ctgctggaga ggatcgagga ggcctatgcc 1440

atcgtggagc ctctgctggg agcagagtac ccaagggacc gcaacctgaa ggccgataag 1500

aaggaagtgg gcaagatcaa ggacttcctg gatagcatca agtccctgca gttctttctg 1560

aagcctctgc tgtccgccga gatctttgac gagaaggatc tgggcttcta caatcagctg 1620

gagggctact atgaggagat cgattctatc ggccacctgt ataacaaggt gcggaattat 1680

ctgaccggca agatctacag caaggagaag tttaagctga acttcgagaa cagcaccctg 1740

ctgaagggct gggacgagaa ccgggaggtg gccaatctgt gcgtgatctt cagagaggac 1800

cagaagtact atctgggcgt gatggataag gagaacaata ccatcctgtc cgacatcccc 1860

aaggtgaagc ctaacgagct gttttacgag aagatggtgt ataagctgat ccccacacct 1920

cacatgcagc tgccccggat catcttctct agcgacaacc tgtctatcta taatcctagc 1980

aagtccatcc tgaagatcag agaggccaag agctttaagg agggcaagaa cttcaagctg 2040

aaggactgtc acaagtttat cgatttctac aaggagtcta tcagcaagaa tgaggactgg 2100

agcagattcg acttcaagtt cagcaagacc agcagctacg agaacatcag cgagttttac 2160

cgggaggtgg agagacaggg ctataacctg gacttcaaga aggtgtctaa gttctacatc 2220

gacagcctgg tggaggatgg caagctgtac ctgttccaga tctataacaa ggacttttct 2280

atcttcagca agggcaagcc caatctgcac accatctatt ttcggtccct gttctctaag 2340

gagaacctga aggacgtgtg cctgaagctg aatggcgagg ccgagatgtt ctttcggaag 2400

aagtccatca actacgatga gaagaagaag cgggagggcc accaccccga gctgtttgag 2460

aagctgaagt atcctatcct gaaggacaag agatacagcg aggataagtt tcagttccac 2520

ctgcccatca gcctgaactt caagtccaag gagcggctga actttaatct gaaagtgaat 2580

gagttcctga agagaaacaa ggacatcaat atcatcggca tcgatcgggg cgagagaaac 2640

ctgctgtacc tggtcatgat caatcagaag ggcgagatcc tgaagcagac cctgctggac 2700

agcatgcagt ccggcaaggg ccggcctgag atcaactaca aggagaagct gcaggagaag 2760

gagatcgaga gggataaggc ccgcaagagc tggggcacag tggagaatat caaggagctg 2820

aaggagggct atctgtctat cgtgatccac cagatcagca agctgatggt ggagaacaat 2880

gccatcgtgg tgctggagga cctgaacatc ggctttaagc ggggcagaca gaaggtggag 2940

cggcaggtgt accagaagtt cgagaagatg ctgatcgata agctgaactt tctggtgttc 3000

aaggagaata agccaaccga gccaggaggc gtgctgaagg cctatcagct gacagacgag 3060

tttcagtctt tcgagaagct gagcaagcag accggctttc tgttctacgt gccaagctgg 3120

aacacctcca agatcgaccc cagaacaggc tttatcgatt tcctgcaccc tgcctacgag 3180

aatatcgaga aggccaagca gtggatcaac aagtttgatt ccatcaggtt caattctaag 3240

atggactggt ttgagttcac cgccgataca cgcaagtttt ccgagaacct gatgctgggc 3300

aagaatcggg tgtgggtcat ctgcaccaca aatgtggagc ggtacttcac cagcaagacc 3360

gccaacagct ccatccagta caatagcatc cagatcaccg agaagctgaa ggagctgttt 3420

gtggacatcc ctttcagcaa cggccaggat ctgaagccag agatcctgag gaagaatgac 3480

gccgtgttct ttaagagcct gctgttttac atcaagacca cactgtccct gcgccagaac 3540

aatggcaaga agggcgagga ggagaaggac ttcatcctga gcccagtggt ggattccaag 3600

ggccggttct ttaactctct ggaggccagc gacgatgagc ccaaggacgc cgatgccaat 3660

ggcgcctacc acatcgccct gaagggcctg atgaacctgc tggtgctgaa tgagacaaag 3720

gaggagaacc tgagcagacc aaagtggaag atcaagaata aggactggct ggagttcgtg 3780

tgggagagga accgcaaaag gccggcggcc acgaaaaagg ccggccaggc aaaaaagaaa 3840

aagggatcct acccatacga tgttccagat tacgcttatc cctacgacgt gcctgattat 3900

gcatacccat atgatgtccc cgactatgcc taagaattc 3939

<210> 1271
<211> 1263
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 1271
Met Glu Asp Tyr Ser Gly Phe Val Asn Ile Tyr Ser Ile Gln Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr Leu Glu His Ile Glu
20 25 30

Lys Lys Gly Phe Leu Lys Lys Asp Lys Ile Arg Ala Glu Asp Tyr Lys
35 40 45

Ala Val Lys Lys Ile Ile Asp Lys Tyr His Arg Ala Tyr Ile Glu Glu
50 55 60

Val Phe Asp Ser Val Leu His Gln Lys Lys Lys Lys Asp Lys Thr Arg
65 70 75 80

Phe Ser Thr Gln Phe Ile Lys Glu Ile Lys Glu Phe Ser Glu Leu Tyr
85 90 95

Tyr Lys Thr Glu Lys Asn Ile Pro Asp Lys Glu Arg Leu Glu Ala Leu
100 105 110

Ser Glu Lys Leu Arg Lys Met Leu Val Gly Ala Phe Lys Gly Glu Phe
115 120 125

Ser Glu Glu Val Ala Glu Lys Tyr Lys Asn Leu Phe Ser Lys Glu Leu
130 135 140

Ile Arg Asn Glu Ile Glu Lys Phe Cys Glu Thr Asp Glu Glu Arg Lys
145 150 155 160

Gln Val Ser Asn Phe Lys Ser Phe Thr Thr Tyr Phe Thr Gly Phe His
165 170 175

Ser Asn Arg Gln Asn Ile Tyr Ser Asp Glu Lys Lys Ser Thr Ala Ile
180 185 190

Gly Tyr Arg Ile Ile His Gln Asn Leu Pro Lys Phe Leu Asp Asn Leu
195 200 205

Lys Ile Ile Glu Ser Ile Gln Arg Arg Phe Lys Asp Phe Pro Trp Ser
210 215 220

```
Asp Leu Lys Lys Asn Leu Lys Lys Ile Asp Lys Asn Ile Lys Leu Thr
225                 230                 235                 240

Glu Tyr Phe Ser Ile Asp Gly Phe Val Asn Val Leu Asn Gln Lys Gly
                245                 250                 255

Ile Asp Ala Tyr Asn Thr Ile Leu Gly Gly Lys Ser Glu Glu Ser Gly
            260                 265                 270

Glu Lys Ile Gln Gly Leu Asn Glu Tyr Ile Asn Leu Tyr Arg Gln Lys
        275                 280                 285

Asn Asn Ile Asp Arg Lys Asn Leu Pro Asn Val Lys Ile Leu Phe Lys
    290                 295                 300

Gln Ile Leu Gly Asp Arg Glu Thr Lys Ser Phe Ile Pro Glu Ala Phe
305                 310                 315                 320

Pro Asp Asp Gln Ser Val Leu Asn Ser Ile Thr Glu Phe Ala Lys Tyr
                325                 330                 335

Leu Lys Leu Asp Lys Lys Lys Lys Ser Ile Ile Ala Glu Leu Lys Lys
            340                 345                 350

Phe Leu Ser Ser Phe Asn Arg Tyr Glu Leu Asp Gly Ile Tyr Leu Ala
        355                 360                 365

Asn Asp Asn Ser Leu Ala Ser Ile Ser Thr Phe Leu Phe Asp Asp Trp
    370                 375                 380

Ser Phe Ile Lys Lys Ser Val Ser Phe Lys Tyr Asp Glu Ser Val Gly
385                 390                 395                 400

Asp Pro Lys Lys Lys Ile Lys Ser Pro Leu Lys Tyr Glu Lys Glu Lys
                405                 410                 415

Glu Lys Trp Leu Lys Gln Lys Tyr Tyr Thr Ile Ser Phe Leu Asn Asp
            420                 425                 430

Ala Ile Glu Ser Tyr Ser Lys Ser Gln Asp Glu Lys Arg Val Lys Ile
        435                 440                 445

Arg Leu Glu Ala Tyr Phe Ala Glu Phe Lys Ser Lys Asp Asp Ala Lys
    450                 455                 460

Lys Gln Phe Asp Leu Leu Glu Arg Ile Glu Glu Ala Tyr Ala Ile Val
465                 470                 475                 480

Glu Pro Leu Leu Gly Ala Glu Tyr Pro Arg Asp Arg Asn Leu Lys Ala
                485                 490                 495

Asp Lys Lys Glu Val Gly Lys Ile Lys Asp Phe Leu Asp Ser Ile Lys
            500                 505                 510

Ser Leu Gln Phe Phe Leu Lys Pro Leu Leu Ser Ala Glu Ile Phe Asp
        515                 520                 525

Glu Lys Asp Leu Gly Phe Tyr Asn Gln Leu Glu Gly Tyr Tyr Glu Glu
    530                 535                 540

Ile Asp Ser Ile Gly His Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr
545                 550                 555                 560
```

Gly Lys Ile Tyr Ser Lys Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser
565 570 575

Thr Leu Leu Lys Gly Trp Asp Glu Asn Arg Glu Val Ala Asn Leu Cys
580 585 590

Val Ile Phe Arg Glu Asp Gln Lys Tyr Tyr Leu Gly Val Met Asp Lys
595 600 605

Glu Asn Asn Thr Ile Leu Ser Asp Ile Pro Lys Val Lys Pro Asn Glu
610 615 620

Leu Phe Tyr Glu Lys Met Val Tyr Lys Leu Ile Pro Thr Pro His Met
625 630 635 640

Gln Leu Pro Arg Ile Ile Phe Ser Ser Asp Asn Leu Ser Ile Tyr Asn
645 650 655

Pro Ser Lys Ser Ile Leu Lys Ile Arg Glu Ala Lys Ser Phe Lys Glu
660 665 670

Gly Lys Asn Phe Lys Leu Lys Asp Cys His Lys Phe Ile Asp Phe Tyr
675 680 685

Lys Glu Ser Ile Ser Lys Asn Glu Asp Trp Ser Arg Phe Asp Phe Lys
690 695 700

Phe Ser Lys Thr Ser Ser Tyr Glu Asn Ile Ser Glu Phe Tyr Arg Glu
705 710 715 720

Val Glu Arg Gln Gly Tyr Asn Leu Asp Phe Lys Lys Val Ser Lys Phe
725 730 735

Tyr Ile Asp Ser Leu Val Glu Asp Gly Lys Leu Tyr Leu Phe Gln Ile
740 745 750

Tyr Asn Lys Asp Phe Ser Ile Phe Ser Lys Gly Lys Pro Asn Leu His
755 760 765

Thr Ile Tyr Phe Arg Ser Leu Phe Ser Lys Glu Asn Leu Lys Asp Val
770 775 780

Cys Leu Lys Leu Asn Gly Glu Ala Glu Met Phe Phe Arg Lys Lys Ser
785 790 795 800

Ile Asn Tyr Asp Glu Lys Lys Lys Arg Glu Gly His His Pro Glu Leu
805 810 815

Phe Glu Lys Leu Lys Tyr Pro Ile Leu Lys Asp Lys Arg Tyr Ser Glu
820 825 830

Asp Lys Phe Gln Phe His Leu Pro Ile Ser Leu Asn Phe Lys Ser Lys
835 840 845

Glu Arg Leu Asn Phe Asn Leu Lys Val Asn Glu Phe Leu Lys Arg Asn
850 855 860

Lys Asp Ile Asn Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu Leu
865 870 875 880

Tyr Leu Val Met Ile Asn Gln Lys Gly Glu Ile Leu Lys Gln Thr Leu
885 890 895

```
Leu Asp Ser Met Gln Ser Gly Lys Gly Arg Pro Glu Ile Asn Tyr Lys
            900                 905                 910

Glu Lys Leu Gln Glu Lys Glu Ile Glu Arg Asp Lys Ala Arg Lys Ser
        915                 920                 925

Trp Gly Thr Val Glu Asn Ile Lys Glu Leu Lys Glu Gly Tyr Leu Ser
    930                 935                 940

Ile Val Ile His Gln Ile Ser Lys Leu Met Val Glu Asn Asn Ala Ile
945                 950                 955                 960

Val Val Leu Glu Asp Leu Asn Ile Gly Phe Lys Arg Gly Arg Gln Lys
                965                 970                 975

Val Glu Arg Gln Val Tyr Gln Lys Phe Glu Lys Met Leu Ile Asp Lys
            980                 985                 990

Leu Asn Phe Leu Val Phe Lys Glu Asn Lys Pro Thr Glu Pro Gly Gly
        995                 1000                1005

Val Leu  Lys Ala Tyr Gln Leu  Thr Asp Glu Phe Gln  Ser Phe Glu
    1010                1015                1020

Lys Leu  Ser Lys Gln Thr Gly  Phe Leu Phe Tyr Val  Pro Ser Trp
    1025                1030                1035

Asn Thr  Ser Lys Ile Asp Pro  Arg Thr Gly Phe Ile  Asp Phe Leu
    1040                1045                1050

His Pro  Ala Tyr Glu Asn Ile  Glu Lys Ala Lys Gln  Trp Ile Asn
    1055                1060                1065

Lys Phe  Asp Ser Ile Arg Phe  Asn Ser Lys Met Asp  Trp Phe Glu
    1070                1075                1080

Phe Thr  Ala Asp Thr Arg Lys  Phe Ser Glu Asn Leu  Met Leu Gly
    1085                1090                1095

Lys Asn  Arg Val Trp Val Ile  Cys Thr Thr Asn Val  Glu Arg Tyr
    1100                1105                1110

Phe Thr  Ser Lys Thr Ala Asn  Ser Ser Ile Gln Tyr  Asn Ser Ile
    1115                1120                1125

Gln Ile  Thr Glu Lys Leu Lys  Glu Leu Phe Val Asp  Ile Pro Phe
    1130                1135                1140

Ser Asn  Gly Gln Asp Leu Lys  Pro Glu Ile Leu Arg  Lys Asn Asp
    1145                1150                1155

Ala Val  Phe Phe Lys Ser Leu  Leu Phe Tyr Ile Lys  Thr Thr Leu
    1160                1165                1170

Ser Leu  Arg Gln Asn Asn Gly  Lys Lys Gly Glu Glu  Glu Lys Asp
    1175                1180                1185

Phe Ile  Leu Ser Pro Val Val  Asp Ser Lys Gly Arg  Phe Phe Asn
    1190                1195                1200

Ser Leu  Glu Ala Ser Asp Asp  Glu Pro Lys Asp Ala  Asp Ala Asn
    1205                1210                1215
```

Gly Ala Tyr His Ile Ala Leu Lys Gly Leu Met Asn Leu Leu Val
1220 1225 1230

Leu Asn Glu Thr Lys Glu Glu Asn Leu Ser Arg Pro Lys Trp Lys
1235 1240 1245

Ile Lys Asn Lys Asp Trp Leu Glu Phe Val Trp Glu Arg Asn Arg
1250 1255 1260

<210> 1272
<211> 3834
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1272
ggtaccatga gcaagctgga gaagtttaca aactgctact ccctgtctaa gaccctgagg 60

ttcaaggcca tccctgtggg caagacccag gagaacatcg acaataagcg gctgctggtg 120

gaggacgaga agagagccga ggattataag ggcgtgaaga agctgctgga tcgctactat 180

ctgtctttta tcaacgacgt gctgcacagc atcaagctga agaatctgaa caattacatc 240

agcctgttcc ggaagaaaac cagaaccgag aaggagaata aggagctgga gaacctggag 300

atcaatctgc ggaaggagat cgccaaggcc ttcaagggca acgagggcta caagtccctg 360

tttaagaagg atatcatcga gacaatcctg ccagagttcc tggacgataa ggacgagatc 420

gccctggtga acagcttcaa tggctttacc acagccttca ccggcttctt tgataacaga 480

gagaatatgt tttccgagga ggccaagagc acatccatcg ccttcaggtg tatcaacgag 540

aatctgaccc gctacatctc taatatggac atcttcgaga aggtggacgc catctttgat 600

aagcacgagg tgcaggagat caaggagaag atcctgaaca gcgactatga tgtggaggat 660

ttctttgagg gcgagttctt taactttgtg ctgacacagg agggcatcga cgtgtataac 720

gccatcatcg gcggcttcgt gaccgagagc ggcgagaaga tcaagggcct gaacgagtac 780

atcaacctgt ataatcagaa aaccaagcag aagctgccta agtttaagcc actgtataag 840

caggtgctga gcgatcggga gtctctgagc ttctacggcg agggctatac atccgatgag 900

gaggtgctgg aggtgtttag aaacaccctg aacaagaaca gcgagatctt cagctccatc 960

aagaagctgg agaagctgtt caagaatttt gacgagtact ctagcgccgg catctttgtg 1020

aagaacggcc ccgccatcag cacaatctcc aaggatatct tcggcgagtg gaacgtgatc 1080

cgggacaagt ggaatgccga gtatgacgat atccacctga agaagaaggc cgtggtgacc 1140

gagaagtacg aggacgatcg gagaaagtcc ttcaagaaga tcggctcctt ttctctggag 1200

cagctgcagg agtacgccga cgccgatctg tctgtggtgg agaagctgaa ggagatcatc 1260

atccagaagg tggatgagat ctacaaggtg tatggctcct ctgagaagct gttcgacgcc 1320

gattttgtgc tggagaagag cctgaagaag aacgacgccg tggtggccat catgaaggac 1380

ctgctggatt ctgtgaagag cttcgagaat tacatcaagg ccttctttgg cgagggcaag 1440

gagacaaaca gggacgagtc cttctatggc gattttgtgc tggcctacga catcctgctg 1500

aaggtggacc acatctacga tgccatccgc aattatgtga cccagaagcc ctactctaag 1560

gataagttca agctgtattt tcagaaccct cagttcatgg gcggctggga caaggataag 1620

gagacagact atcgggccac catcctgaga tacggctcca agtactatct ggccatcatg 1680

gataagaagt acgccaagtg cctgcagaag atcgacaagg acgatgtgaa cggcaattac 1740

gagaagatca actataagct gctgcccggc cctaataaga tgctgccaaa ggtgttcttt 1800

tctaagaagt ggatggccta ctataacccc agcgaggaca tccagaagat ctacaagaat 1860

ggcacattca agaagggcga tatgtttaac ctgaatgact gtcacaagct gatcgacttc 1920

tttaaggata gcatctcccg gtatccaaag tggtccaatg cctacgattt caacttttct 1980

gagacagaga agtataagga catcgccggc ttttacagag aggtggagga gcagggctat 2040

aaggtgagct tcgagtctgc cagcaagaag gaggtggata agctggtgga ggagggcaag 2100

ctgtatatgt tccagatcta taacaaggac ttttccgata agtctcacgg cacacccaat 2160

ctgcacacca tgtacttcaa gctgctgttt gacgagaaca atcacggaca gatcaggctg 2220

agcggaggag cagagctgtt catgaggcgc gcctccctga agaaggagga gctggtggtg 2280

cacccagcca actcccctat cgccaacaag aatccagata atcccaagaa aaccacaacc 2340

ctgtcctacg acgtgtataa ggataagagg ttttctgagg accagtacga gctgcacatc 2400

ccaatcgcca tcaataagtg ccccaagaac atcttcaaga tcaatacaga ggtgcgcgtg 2460

ctgctgaagc acgacgataa cccctatgtg atcggcatcg ataggggcga gcgcaatctg 2520

ctgtatatcg tggtggtgga cggcaagggc aacatcgtgg agcagtattc cctgaacgag 2580

atcatcaaca acttcaacgg catcaggatc aagacagatt accactctct gctggacaag 2640

aaggagaagg agaggttcga ggcccgccag aactggacct ccatcgagaa tatcaaggag 2700

ctgaaggccg gctatatctc tcaggtggtg cacaagatct gcgagctggt ggagaagtac 2760

gatgccgtga tcgccctgga ggacctgaac tctggcttta agaatagccg cgtgaaggtg 2820

gagaagcagg tgtatcagaa gttcgagaag atgctgatcg ataagctgaa ctacatggtg 2880

gacaagaagt ctaatccttg tgcaacaggc ggcgccctga agggctatca gatcaccaat 2940

aagttcgaga gctttaagtc catgtctacc cagaacggct tcatctttta catccctgcc 3000

tggctgacat ccaagatcga tccatctacc ggctttgtga acctgctgaa aaccaagtat 3060

accagcatcg ccgattccaa gaagttcatc agctcctttg acaggatcat gtacgtgccc 3120

gaggaggatc tgttcgagtt tgccctggac tataagaact tctctcgcac agacgccgat 3180

tacatcaaga agtggaagct gtactcctac ggcaaccgga tcagaatctt ccggaatcct 3240

aagaagaaca acgtgttcga ctgggaggag gtgtgcctga ccagcgccta taaggagctg 3300

ttcaacaagt acggcatcaa ttatcagcag ggcgatatca gagccctgct gtgcgagcag 3360

tccgacaagg ccttctactc tagctttatg gccctgatga gcctgatgct gcagatgcgg 3420

aacagcatca caggccgcac cgacgtggat tttctgatca gccctgtgaa gaactccgac 3480

ggcatcttct acgatagccg gaactatgag gcccaggaga atgccatcct gccaaagaac 3540

gccgacgcca atggcgccta taacatcgcc agaaaggtgc tgtgggccat cggccagttc 3600

aagaaggccg aggacgagaa gctggataag gtgaagatcg ccatctctaa caaggagtgg 3660

ctggagtacg cccagaccag cgtgaagcac aaaaggccgg cggccacgaa aaaggccggc 3720

caggcaaaaa agaaaaaggg atcctaccca tacgatgttc cagattacgc ttatccctac 3780

gacgtgcctg attatgcata cccatatgat gtccccgact atgcctaaga attc 3834

<210> 1273
<211> 1228
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 1273
Met Ser Lys Leu Glu Lys Phe Thr Asn Cys Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Arg Phe Lys Ala Ile Pro Val Gly Lys Thr Gln Glu Asn Ile Asp

```
            20                  25                  30

Asn Lys Arg Leu Leu Val Glu Asp Glu Lys Arg Ala Glu Asp Tyr Lys
        35                  40                  45

Gly Val Lys Lys Leu Leu Asp Arg Tyr Tyr Leu Ser Phe Ile Asn Asp
    50                  55                  60

Val Leu His Ser Ile Lys Leu Lys Asn Leu Asn Asn Tyr Ile Ser Leu
65                  70                  75                  80

Phe Arg Lys Lys Thr Arg Thr Glu Lys Glu Asn Lys Glu Leu Glu Asn
                85                  90                  95

Leu Glu Ile Asn Leu Arg Lys Glu Ile Ala Lys Ala Phe Lys Gly Asn
            100                 105                 110

Glu Gly Tyr Lys Ser Leu Phe Lys Lys Asp Ile Ile Glu Thr Ile Leu
        115                 120                 125

Pro Glu Phe Leu Asp Asp Lys Asp Glu Ile Ala Leu Val Asn Ser Phe
    130                 135                 140

Asn Gly Phe Thr Thr Ala Phe Thr Gly Phe Phe Asp Asn Arg Glu Asn
145                 150                 155                 160

Met Phe Ser Glu Glu Ala Lys Ser Thr Ser Ile Ala Phe Arg Cys Ile
                165                 170                 175

Asn Glu Asn Leu Thr Arg Tyr Ile Ser Asn Met Asp Ile Phe Glu Lys
            180                 185                 190

Val Asp Ala Ile Phe Asp Lys His Glu Val Gln Glu Ile Lys Glu Lys
        195                 200                 205

Ile Leu Asn Ser Asp Tyr Asp Val Glu Asp Phe Phe Glu Gly Glu Phe
    210                 215                 220

Phe Asn Phe Val Leu Thr Gln Glu Gly Ile Asp Val Tyr Asn Ala Ile
225                 230                 235                 240

Ile Gly Gly Phe Val Thr Glu Ser Gly Glu Lys Ile Lys Gly Leu Asn
                245                 250                 255

Glu Tyr Ile Asn Leu Tyr Asn Gln Lys Thr Lys Gln Lys Leu Pro Lys
            260                 265                 270

Phe Lys Pro Leu Tyr Lys Gln Val Leu Ser Asp Arg Glu Ser Leu Ser
        275                 280                 285

Phe Tyr Gly Glu Gly Tyr Thr Ser Asp Glu Glu Val Leu Glu Val Phe
    290                 295                 300

Arg Asn Thr Leu Asn Lys Asn Ser Glu Ile Phe Ser Ser Ile Lys Lys
305                 310                 315                 320

Leu Glu Lys Leu Phe Lys Asn Phe Asp Glu Tyr Ser Ser Ala Gly Ile
                325                 330                 335

Phe Val Lys Asn Gly Pro Ala Ile Ser Thr Ile Ser Lys Asp Ile Phe
            340                 345                 350

Gly Glu Trp Asn Val Ile Arg Asp Lys Trp Asn Ala Glu Tyr Asp Asp
```

```
        355                 360                 365

Ile His Leu Lys Lys Lys Ala Val Val Thr Glu Lys Tyr Glu Asp Asp
    370                 375                 380

Arg Arg Lys Ser Phe Lys Lys Ile Gly Ser Phe Ser Leu Glu Gln Leu
385                 390                 395                 400

Gln Glu Tyr Ala Asp Ala Asp Leu Ser Val Val Glu Lys Leu Lys Glu
                405                 410                 415

Ile Ile Ile Gln Lys Val Asp Glu Ile Tyr Lys Val Tyr Gly Ser Ser
            420                 425                 430

Glu Lys Leu Phe Asp Ala Asp Phe Val Leu Glu Lys Ser Leu Lys Lys
        435                 440                 445

Asn Asp Ala Val Val Ala Ile Met Lys Asp Leu Leu Asp Ser Val Lys
    450                 455                 460

Ser Phe Glu Asn Tyr Ile Lys Ala Phe Phe Gly Glu Gly Lys Glu Thr
465                 470                 475                 480

Asn Arg Asp Glu Ser Phe Tyr Gly Asp Phe Val Leu Ala Tyr Asp Ile
                485                 490                 495

Leu Leu Lys Val Asp His Ile Tyr Asp Ala Ile Arg Asn Tyr Val Thr
            500                 505                 510

Gln Lys Pro Tyr Ser Lys Asp Lys Phe Lys Leu Tyr Phe Gln Asn Pro
        515                 520                 525

Gln Phe Met Gly Gly Trp Asp Lys Asp Lys Glu Thr Asp Tyr Arg Ala
    530                 535                 540

Thr Ile Leu Arg Tyr Gly Ser Lys Tyr Tyr Leu Ala Ile Met Asp Lys
545                 550                 555                 560

Lys Tyr Ala Lys Cys Leu Gln Lys Ile Asp Lys Asp Asp Val Asn Gly
                565                 570                 575

Asn Tyr Glu Lys Ile Asn Tyr Lys Leu Leu Pro Gly Pro Asn Lys Met
            580                 585                 590

Leu Pro Lys Val Phe Phe Ser Lys Lys Trp Met Ala Tyr Tyr Asn Pro
        595                 600                 605

Ser Glu Asp Ile Gln Lys Ile Tyr Lys Asn Gly Thr Phe Lys Lys Gly
    610                 615                 620

Asp Met Phe Asn Leu Asn Asp Cys His Lys Leu Ile Asp Phe Phe Lys
625                 630                 635                 640

Asp Ser Ile Ser Arg Tyr Pro Lys Trp Ser Asn Ala Tyr Asp Phe Asn
                645                 650                 655

Phe Ser Glu Thr Glu Lys Tyr Lys Asp Ile Ala Gly Phe Tyr Arg Glu
            660                 665                 670

Val Glu Glu Gln Gly Tyr Lys Val Ser Phe Glu Ser Ala Ser Lys Lys
        675                 680                 685

Glu Val Asp Lys Leu Val Glu Glu Gly Lys Leu Tyr Met Phe Gln Ile
```

690 695 700

Tyr Asn Lys Asp Phe Ser Asp Lys Ser His Gly Thr Pro Asn Leu His
705 710 715 720

Thr Met Tyr Phe Lys Leu Leu Phe Asp Glu Asn Asn His Gly Gln Ile
725 730 735

Arg Leu Ser Gly Gly Ala Glu Leu Phe Met Arg Arg Ala Ser Leu Lys
740 745 750

Lys Glu Glu Leu Val Val His Pro Ala Asn Ser Pro Ile Ala Asn Lys
755 760 765

Asn Pro Asp Asn Pro Lys Lys Thr Thr Thr Leu Ser Tyr Asp Val Tyr
770 775 780

Lys Asp Lys Arg Phe Ser Glu Asp Gln Tyr Glu Leu His Ile Pro Ile
785 790 795 800

Ala Ile Asn Lys Cys Pro Lys Asn Ile Phe Lys Ile Asn Thr Glu Val
805 810 815

Arg Val Leu Leu Lys His Asp Asp Asn Pro Tyr Val Ile Gly Ile Asp
820 825 830

Arg Gly Glu Arg Asn Leu Leu Tyr Ile Val Val Val Asp Gly Lys Gly
835 840 845

Asn Ile Val Glu Gln Tyr Ser Leu Asn Glu Ile Ile Asn Asn Phe Asn
850 855 860

Gly Ile Arg Ile Lys Thr Asp Tyr His Ser Leu Leu Asp Lys Lys Glu
865 870 875 880

Lys Glu Arg Phe Glu Ala Arg Gln Asn Trp Thr Ser Ile Glu Asn Ile
885 890 895

Lys Glu Leu Lys Ala Gly Tyr Ile Ser Gln Val Val His Lys Ile Cys
900 905 910

Glu Leu Val Glu Lys Tyr Asp Ala Val Ile Ala Leu Glu Asp Leu Asn
915 920 925

Ser Gly Phe Lys Asn Ser Arg Val Lys Val Glu Lys Gln Val Tyr Gln
930 935 940

Lys Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr Met Val Asp Lys
945 950 955 960

Lys Ser Asn Pro Cys Ala Thr Gly Gly Ala Leu Lys Gly Tyr Gln Ile
965 970 975

Thr Asn Lys Phe Glu Ser Phe Lys Ser Met Ser Thr Gln Asn Gly Phe
980 985 990

Ile Phe Tyr Ile Pro Ala Trp Leu Thr Ser Lys Ile Asp Pro Ser Thr
995 1000 1005

Gly Phe Val Asn Leu Leu Lys Thr Lys Tyr Thr Ser Ile Ala Asp
1010 1015 1020

Ser Lys Lys Phe Ile Ser Ser Phe Asp Arg Ile Met Tyr Val Pro

1025 1030 1035

Glu Glu Asp Leu Phe Glu Phe Ala Leu Asp Tyr Lys Asn Phe Ser
1040 1045 1050

Arg Thr Asp Ala Asp Tyr Ile Lys Lys Trp Lys Leu Tyr Ser Tyr
1055 1060 1065

Gly Asn Arg Ile Arg Ile Phe Arg Asn Pro Lys Lys Asn Asn Val
1070 1075 1080

Phe Asp Trp Glu Glu Val Cys Leu Thr Ser Ala Tyr Lys Glu Leu
1085 1090 1095

Phe Asn Lys Tyr Gly Ile Asn Tyr Gln Gln Gly Asp Ile Arg Ala
1100 1105 1110

Leu Leu Cys Glu Gln Ser Asp Lys Ala Phe Tyr Ser Ser Phe Met
1115 1120 1125

Ala Leu Met Ser Leu Met Leu Gln Met Arg Asn Ser Ile Thr Gly
1130 1135 1140

Arg Thr Asp Val Asp Phe Leu Ile Ser Pro Val Lys Asn Ser Asp
1145 1150 1155

Gly Ile Phe Tyr Asp Ser Arg Asn Tyr Glu Ala Gln Glu Asn Ala
1160 1165 1170

Ile Leu Pro Lys Asn Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala
1175 1180 1185

Arg Lys Val Leu Trp Ala Ile Gly Gln Phe Lys Lys Ala Glu Asp
1190 1195 1200

Glu Lys Leu Asp Lys Val Lys Ile Ala Ile Ser Asn Lys Glu Trp
1205 1210 1215

Leu Glu Tyr Ala Gln Thr Ser Val Lys His
1220 1225

<210> 1274
<211> 3930
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1274
ggtaccatgg acagcctgaa ggatttcacc aacctgtacc ccgtgtccaa gacactgcgg 60

tttgagctga agcctgtggg caagaccctg gagaatatcg agaaggccgg catcctgaag 120

gaggatgagc acagagccga gagctaccgg agagtgaaga agatcatcga tacatatcac 180

aaggtgttca tcgacagctc cctggagaac atggccaaga tgggcatcga gaatgagatc 240

aaggccatgc tgcagtcctt ttgcgagctg tataagaagg accacaggac cgagggagag 300

gacaaggccc tggataagat cagggccgtg ctgaggggcc tgatcgtggg agccttcacc 360

ggcgtgtgcg gccggcggga gaacacagtg cagaatgaga agtatgagag cctgtttaag 420

gagaagctga tcaaggagat cctgccagat ttcgtgctgt ctacagaggc cgagtccctg 480

ccctttctg tggaggaggc caccagaagc ctgaaggagt tcgactcctt tacatcttac 540

ttcgccggct tttatgagaa ccggaagaat atctactcta ccaagcccca gagcacagcc 600

atcgcctata gactgatcca cgagaacctg cctaagttca tcgataatat cctggtgttt 660
cagaagatca aggagccaat cgccaaggag ctggagcaca tcagggcaga cttcagcgcc 720
ggcggctaca tcaagaagga tgagcgcctg gaggacatct tttccctgaa ctactatatc 780
cacgtgctgt ctcaggccgg catcgagaag tacaatgccc tgatcggcaa gatcgtgacc 840
gagggcgatg gcgagatgaa gggcctgaac gagcacatca acctgtataa tcagcagagg 900
ggccgcgagg accggctgcc actgttcaga cccctgtata agcagatcct gtctgatagg 960
gagcagctgt cctatctgcc agagtctttc gagaaggacg aggagctgct gagggccctg 1020
aaggagtttt acgatcacat cgcagaggac atcctgggaa ggacccagca gctgatgaca 1080
agcatctccg agtacgatct gtcccggatc tatgtgagaa acgatagcca gctgaccgac 1140
atctccaaga agatgctggg cgattggaat gccatctaca tggcccggga gagagcctat 1200
gaccacgagc aggcccccaa gcgcatcaca gccaagtacg agagggaccg catcaaggcc 1260
ctgaagggcg aggagtctat cagcctggcc aacctgaaca gctgcatcgc cttcctggac 1320
aacgtgaggg attgtcgcgt ggacacctat ctgtctacac tgggacagaa ggagggacct 1380
cacggcctga gcaacctggt ggagaacgtg ttcgcctcct accacgaggc cgagcagctg 1440
ctgtcttttc cctatcctga ggagaacaat ctgatccagg acaaggataa cgtggtgctg 1500
atcaagaacc tgctggataa tatcagcgac ctgcagaggt tcctgaagcc actgtggggc 1560
atgggcgatg agcccgacaa ggatgagagg ttttacggcg agtacaatta tatcaggggc 1620
gccctggacc aggtcatccc tctgtataac aaggtgcgga attatctgac ccgcaagcca 1680
tactccacac gcaaggtgaa gctgaacttc ggcaatagcc agctgctgtc cggctgggat 1740
aggaacaagg agaaggacaa ttcttgcgtg atcctgcgca agggccagaa cttctacctg 1800
gccatcatga acaatcggca caagcggagc ttcgagaata agatgctgcc cgagtataag 1860
gagggcgagc cttacttcga gaagatggat tataagtttc tgccagaccc caacaagatg 1920
ctgcccaagg tgttcctgtc taagaagggc atcgagatct acaagcctag cccaaagctg 1980
ctggagcagt atggccacgg cacccacaag aagggcgata ccttcagcat ggacgatctg 2040
cacgagctga tcgacttctt taagcactcc atcgaggccc acgaggattg gaagcagttc 2100
ggctttaagt tcagcgacac cgccacatac gagaacgtga gcagcttcta ccgggaggtg 2160
gaggaccagg gctacaagct gtcttttaga aaggtgtccg agtcttacgt gtatagcctg 2220
atcgatcagg gcaagctgta cctgttccag atctataaca aggactttag cccttgttcc 2280
aagggcaccc caaatctgca cacactgtac tggcggatgc tgttcgatga gagaaacctg 2340
gccgacgtga tctataagct ggatggcaag gccgagatct tctttcggga gaagtccctg 2400
aagaatgacc acccaaccca ccctgcaggc aagcccatca agaagaagag ccggcagaag 2460
aagggcgagg agagcctgtt cgagtacgat ctggtgaagg accggagata taccatggat 2520
aagtttcagt tccacgtgcc aatcacaatg aactttaagt gctctgccgg cagcaaggtg 2580
aacgacatgg tgaatgccca catcagggag gccaaggaca tgcacgtgat cggcatcgat 2640
aggggcgagc gcaatctgct gtatatctgc gtgatcgaca gccgcggcac catcctggat 2700
cagatctccc tgaacacaat caatgacatc gattatcacg atctgctgga gtccagggac 2760
aaggatcgcc agcaggagca caggaactgg cagaccatcg agggcatcaa ggagctgaag 2820
cagggctacc tgtctcaggc cgtgcaccgc atcgccgagc tgatggtggc ctataaggcc 2880
gtggtggccc tggaggacct gaacatgggc ttcaagcggg gcagacagaa ggtggagagc 2940
agcgtgtacc agcagtttga gaagcagctg atcgacaagc tgaattatct ggtggataag 3000
aagaagcggc ccgaggacat cggaggcctg ctgagagcct accagttcac cgccccttc 3060
aagagcttta aggagatggg caagcagaac ggctttctgt tctatatccc tgcctggaac 3120

acatccaata tcgacccaac cacaggcttc gtgaacctgt ttcacgtgca gtacgagaat 3180

gtggataagg ccaagagctt ctttcagaag ttcgacagca tctcctacaa ccctaagaag 3240

gattggtttg agttcgcctt tgactataag aacttcacca agaaggccga gggctctagg 3300

agcatgtgga ttctgtgcac ccacggctcc cggatcaaga acttcagaaa ttctcagaag 3360

aatggccagt gggatagcga ggagtttgcc ctgaccgagg ccttcaagtc cctgtttgtg 3420

cggtacgaga tcgattatac cgccgacctg aaaaccgcca tcgtggacga gaagcagaag 3480

gatttctttg tggacctgct gaagctgttc aagctgaccg tgcagatgag aaactcctgg 3540

aaggagaagg acctggatta cctgatctct ccagtggccg gcgccgatgg caggttcttt 3600

gacacacgcg agggcaataa gagcctgccc aaggacgcag atgcaaacgg agcctataat 3660

atcgccctga agggcctgtg ggcactgagg cagatcagac agacctccga gggcggcaag 3720

ctgaagctgg ccatctctaa caaggagtgg ctgcagtttg tgcaggagag atcctacgag 3780

aaggacaaaa ggccggcggc cacgaaaaag gccggccagg caaaaaagaa aaagggatcc 3840

tacccatacg atgttccaga ttacgcttat ccctacgacg tgcctgatta tgcataccca 3900

tatgatgtcc ccgactatgc ctaagaattc 3930

<210> 1275
<211> 1260
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 1275
Met Asp Ser Leu Lys Asp Phe Thr Asn Leu Tyr Pro Val Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr Leu Glu Asn Ile Glu
20 25 30

Lys Ala Gly Ile Leu Lys Glu Asp Glu His Arg Ala Glu Ser Tyr Arg
35 40 45

Arg Val Lys Lys Ile Ile Asp Thr Tyr His Lys Val Phe Ile Asp Ser
50 55 60

Ser Leu Glu Asn Met Ala Lys Met Gly Ile Glu Asn Glu Ile Lys Ala
65 70 75 80

Met Leu Gln Ser Phe Cys Glu Leu Tyr Lys Lys Asp His Arg Thr Glu
85 90 95

Gly Glu Asp Lys Ala Leu Asp Lys Ile Arg Ala Val Leu Arg Gly Leu
100 105 110

Ile Val Gly Ala Phe Thr Gly Val Cys Gly Arg Arg Glu Asn Thr Val
115 120 125

Gln Asn Glu Lys Tyr Glu Ser Leu Phe Lys Glu Lys Leu Ile Lys Glu
130 135 140

Ile Leu Pro Asp Phe Val Leu Ser Thr Glu Ala Glu Ser Leu Pro Phe
145 150 155 160

Ser Val Glu Glu Ala Thr Arg Ser Leu Lys Glu Phe Asp Ser Phe Thr
165 170 175

```
Ser Tyr Phe Ala Gly Phe Tyr Glu Asn Arg Lys Asn Ile Tyr Ser Thr
            180                 185                 190

Lys Pro Gln Ser Thr Ala Ile Ala Tyr Arg Leu Ile His Glu Asn Leu
        195                 200                 205

Pro Lys Phe Ile Asp Asn Ile Leu Val Phe Gln Lys Ile Lys Glu Pro
    210                 215                 220

Ile Ala Lys Glu Leu Glu His Ile Arg Ala Asp Phe Ser Ala Gly Gly
225                 230                 235                 240

Tyr Ile Lys Lys Asp Glu Arg Leu Glu Asp Ile Phe Ser Leu Asn Tyr
                245                 250                 255

Tyr Ile His Val Leu Ser Gln Ala Gly Ile Glu Lys Tyr Asn Ala Leu
            260                 265                 270

Ile Gly Lys Ile Val Thr Glu Gly Asp Gly Glu Met Lys Gly Leu Asn
        275                 280                 285

Glu His Ile Asn Leu Tyr Asn Gln Gln Arg Gly Arg Glu Asp Arg Leu
    290                 295                 300

Pro Leu Phe Arg Pro Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Gln
305                 310                 315                 320

Leu Ser Tyr Leu Pro Glu Ser Phe Glu Lys Asp Glu Glu Leu Leu Arg
                325                 330                 335

Ala Leu Lys Glu Phe Tyr Asp His Ile Ala Glu Asp Ile Leu Gly Arg
            340                 345                 350

Thr Gln Gln Leu Met Thr Ser Ile Ser Glu Tyr Asp Leu Ser Arg Ile
        355                 360                 365

Tyr Val Arg Asn Asp Ser Gln Leu Thr Asp Ile Ser Lys Lys Met Leu
    370                 375                 380

Gly Asp Trp Asn Ala Ile Tyr Met Ala Arg Glu Arg Ala Tyr Asp His
385                 390                 395                 400

Glu Gln Ala Pro Lys Arg Ile Thr Ala Lys Tyr Glu Arg Asp Arg Ile
                405                 410                 415

Lys Ala Leu Lys Gly Glu Glu Ser Ile Ser Leu Ala Asn Leu Asn Ser
            420                 425                 430

Cys Ile Ala Phe Leu Asp Asn Val Arg Asp Cys Arg Val Asp Thr Tyr
        435                 440                 445

Leu Ser Thr Leu Gly Gln Lys Glu Gly Pro His Gly Leu Ser Asn Leu
    450                 455                 460

Val Glu Asn Val Phe Ala Ser Tyr His Glu Ala Glu Gln Leu Leu Ser
465                 470                 475                 480

Phe Pro Tyr Pro Glu Glu Asn Asn Leu Ile Gln Asp Lys Asp Asn Val
                485                 490                 495

Val Leu Ile Lys Asn Leu Leu Asp Asn Ile Ser Asp Leu Gln Arg Phe
            500                 505                 510
```

```
Leu Lys Pro Leu Trp Gly Met Gly Asp Glu Pro Asp Lys Asp Glu Arg
        515                 520                 525

Phe Tyr Gly Glu Tyr Asn Tyr Ile Arg Gly Ala Leu Asp Gln Val Ile
    530                 535                 540

Pro Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Arg Lys Pro Tyr Ser
545                 550                 555                 560

Thr Arg Lys Val Lys Leu Asn Phe Gly Asn Ser Gln Leu Leu Ser Gly
                565                 570                 575

Trp Asp Arg Asn Lys Glu Lys Asp Asn Ser Cys Val Ile Leu Arg Lys
            580                 585                 590

Gly Gln Asn Phe Tyr Leu Ala Ile Met Asn Asn Arg His Lys Arg Ser
        595                 600                 605

Phe Glu Asn Lys Met Leu Pro Glu Tyr Lys Glu Gly Glu Pro Tyr Phe
    610                 615                 620

Glu Lys Met Asp Tyr Lys Phe Leu Pro Asp Pro Asn Lys Met Leu Pro
625                 630                 635                 640

Lys Val Phe Leu Ser Lys Lys Gly Ile Glu Ile Tyr Lys Pro Ser Pro
                645                 650                 655

Lys Leu Leu Glu Gln Tyr Gly His Gly Thr His Lys Lys Gly Asp Thr
            660                 665                 670

Phe Ser Met Asp Asp Leu His Glu Leu Ile Asp Phe Phe Lys His Ser
        675                 680                 685

Ile Glu Ala His Glu Asp Trp Lys Gln Phe Gly Phe Lys Phe Ser Asp
    690                 695                 700

Thr Ala Thr Tyr Glu Asn Val Ser Ser Phe Tyr Arg Glu Val Glu Asp
705                 710                 715                 720

Gln Gly Tyr Lys Leu Ser Phe Arg Lys Val Ser Glu Ser Tyr Val Tyr
                725                 730                 735

Ser Leu Ile Asp Gln Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys
            740                 745                 750

Asp Phe Ser Pro Cys Ser Lys Gly Thr Pro Asn Leu His Thr Leu Tyr
        755                 760                 765

Trp Arg Met Leu Phe Asp Glu Arg Asn Leu Ala Asp Val Ile Tyr Lys
    770                 775                 780

Leu Asp Gly Lys Ala Glu Ile Phe Phe Arg Glu Lys Ser Leu Lys Asn
785                 790                 795                 800

Asp His Pro Thr His Pro Ala Gly Lys Pro Ile Lys Lys Lys Ser Arg
                805                 810                 815

Gln Lys Lys Gly Glu Glu Ser Leu Phe Glu Tyr Asp Leu Val Lys Asp
            820                 825                 830

Arg Arg Tyr Thr Met Asp Lys Phe Gln Phe His Val Pro Ile Thr Met
        835                 840                 845
```

```
Asn Phe Lys Cys Ser Ala Gly Ser Lys Val Asn Asp Met Val Asn Ala
    850                 855                 860

His Ile Arg Glu Ala Lys Asp Met His Val Ile Gly Ile Asp Arg Gly
865                 870                 875                 880

Glu Arg Asn Leu Leu Tyr Ile Cys Val Ile Asp Ser Arg Gly Thr Ile
                885                 890                 895

Leu Asp Gln Ile Ser Leu Asn Thr Ile Asn Asp Ile Asp Tyr His Asp
            900                 905                 910

Leu Leu Glu Ser Arg Asp Lys Asp Arg Gln Gln Glu His Arg Asn Trp
        915                 920                 925

Gln Thr Ile Glu Gly Ile Lys Glu Leu Lys Gln Gly Tyr Leu Ser Gln
    930                 935                 940

Ala Val His Arg Ile Ala Glu Leu Met Val Ala Tyr Lys Ala Val Val
945                 950                 955                 960

Ala Leu Glu Asp Leu Asn Met Gly Phe Lys Arg Gly Arg Gln Lys Val
                965                 970                 975

Glu Ser Ser Val Tyr Gln Gln Phe Glu Lys Gln Leu Ile Asp Lys Leu
            980                 985                 990

Asn Tyr Leu Val Asp Lys Lys Lys Arg Pro Glu Asp Ile Gly Gly Leu
        995                 1000                1005

Leu Arg  Ala Tyr Gln Phe Thr  Ala Pro Phe Lys Ser  Phe Lys Glu
    1010                1015                1020

Met Gly  Lys Gln Asn Gly Phe  Leu Phe Tyr Ile Pro  Ala Trp Asn
    1025                1030                1035

Thr Ser  Asn Ile Asp Pro Thr  Thr Gly Phe Val Asn  Leu Phe His
    1040                1045                1050

Val Gln  Tyr Glu Asn Val Asp  Lys Ala Lys Ser Phe  Phe Gln Lys
    1055                1060                1065

Phe Asp  Ser Ile Ser Tyr Asn  Pro Lys Lys Asp Trp  Phe Glu Phe
    1070                1075                1080

Ala Phe  Asp Tyr Lys Asn Phe  Thr Lys Lys Ala Glu  Gly Ser Arg
    1085                1090                1095

Ser Met  Trp Ile Leu Cys Thr  His Gly Ser Arg Ile  Lys Asn Phe
    1100                1105                1110

Arg Asn  Ser Gln Lys Asn Gly  Gln Trp Asp Ser Glu  Glu Phe Ala
    1115                1120                1125

Leu Thr  Glu Ala Phe Lys Ser  Leu Phe Val Arg Tyr  Glu Ile Asp
    1130                1135                1140

Tyr Thr  Ala Asp Leu Lys Thr  Ala Ile Val Asp Glu  Lys Gln Lys
    1145                1150                1155

Asp Phe  Phe Val Asp Leu Leu  Lys Leu Phe Lys Leu  Thr Val Gln
    1160                1165                1170
```

Met Arg Asn Ser Trp Lys Glu Lys Asp Leu Asp Tyr Leu Ile Ser
1175 1180 1185

Pro Val Ala Gly Ala Asp Gly Arg Phe Phe Asp Thr Arg Glu Gly
1190 1195 1200

Asn Lys Ser Leu Pro Lys Asp Ala Asp Ala Asn Gly Ala Tyr Asn
1205 1210 1215

Ile Ala Leu Lys Gly Leu Trp Ala Leu Arg Gln Ile Arg Gln Thr
1220 1225 1230

Ser Glu Gly Gly Lys Leu Lys Leu Ala Ile Ser Asn Lys Glu Trp
1235 1240 1245

Leu Gln Phe Val Gln Glu Arg Ser Tyr Glu Lys Asp
1250 1255 1260

<210> 1276
<211> 4119
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1276
ggtaccatgg agaactatca ggagttcacc aacctgtttc agctgaataa gacactgaga 60

ttcgagctga agcccatcgg caagacctgc gagctgctgg aggagggcaa gatcttcgcc 120

agcggctcct ttctggagaa ggacaaggtg agggccgata acgtgagcta cgtgaagaag 180

gagatcgaca agaagcacaa gatctttatc gaggagacac tgagctcctt ctctatcagc 240

aacgatctgc tgaagcagta ctttgactgc tataatgagc tgaaggcctt caagaaggac 300

tgtaagagcg atgaggagga ggtgaagaaa accgccctgc gcaacaagtg tacctccatc 360

cagagggcca tgcgcgaggc catctctcag gcctttctga agagccccca gaagaagctg 420

ctggccatca agaacctgat cgagaacgtg ttcaaggccg acgagaatgt gcagcacttc 480

tccgagttta ccagctattt ctccggcttt gagacaaaca gagagaattt ctactctgac 540

gaggagaagt ccacatctat cgcctatagg ctggtgcacg ataacctgcc tatcttcatc 600

aagaacatct acatcttcga gaagctgaag gagcagttcg acgccaagac cctgagcgag 660

atcttcgaga actacaagct gtatgtggcc ggctctagcc tggatgaggt gttctccctg 720

gagtacttta acaataccct gacacagaag ggcatcgaca actataatgc cgtgatcggc 780

aagatcgtga aggaggataa gcaggagatc cagggcctga acgagcacat caacctgtat 840

aatcagaagc acaaggaccg gagactgccc ttctttatct ccctgaagaa gcagatcctg 900

tccgatcggg aggccctgtc ttggctgcct gacatgttca agaatgattc tgaagtgatc 960

aaggccctga agggcttcta catcgaggac ggctttgaga acaatgtgct gacacctctg 1020

gccaccctgc tgtcctctct ggataagtac aacctgaatg gcatctttat ccgcaacaat 1080

gaggccctga gctccctgtc ccagaacgtg tatcggaatt tttctatcga cgaggccatc 1140

gatgccaacg ccgagctgca gaccttcaac aattacgagc tgatcgccaa tgccctgcgc 1200

gccaagatca agaaggagac aaagcagggc cggaagtctt tcgagaagta cgaggagtat 1260

atcgataaga aggtgaaggc catcgacagc ctgtccatcc aggagatcaa cgagctggtg 1320

gagaattacg tgagcgagtt taactctaat agcggcaaca tgccaagaaa ggtggaggac 1380

tacttcagcc tgatgaggaa gggcgacttc ggctccaacg atctgatcga aaatatcaag 1440

accaagctga gcgccgcaga gaagctgctg ggcacaaagt accaggagac agccaaggac 1500

atcttcaaga aggatgagaa ctccaagctg atcaaggagc tgctggacgc caccaagcag 1560

ttccagcact ttatcaagcc actgctgggc acaggcgagg aggcagatcg ggacctggtg 1620

ttctacggcg attttctgcc cctgtatgag aagtttgagg agctgaccct gctgtataac 1680

aaggtgcgga atagactgac acagaagccc tattccaagg acaagatccg cctgtgcttc 1740

aacaagccta agctgatgac aggctgggtg gattccaaga ccgagaagtc tgacaacggc 1800

acacagtacg gcggctatct gtttcggaag aagaatgaga tcggcgagta cgattatttt 1860

ctgggcatct ctagcaaggc ccagctgttc agaaagaacg aggccgtgat cggcgactac 1920

gagaggctgg attactatca gccaaaggcc aataccatct acggctctgc ctatgagggc 1980

gagaacagct acaaggagga caagaagcgg ctgaacaaag tgatcatcgc ctatatcgag 2040

cagatcaagc agacaaacat caagaagtct atcatcgagt ccatctctaa gtatcctaat 2100

atcagcgacg atgacaaggt gaccccatcc tctctgctgg agaagatcaa gaaggtgtct 2160

atcgacagct acaacggcat cctgtccttc aagtcttttc agagcgtgaa caaggaagtg 2220

atcgataacc tgctgaaaac catcagcccc ctgaagaaca aggccgagtt tctggacctg 2280

atcaataagg attatcagat cttcaccgag gtgcaggccg tgatcgacga gatctgcaag 2340

cagaaaacct tcatctactt tccaatctcc aacgtggagc tggagaagga gatgggcgat 2400

aaggacaagc ccctgtgcct gttccagatc agcaataagg atctgtcctt cgccaagacc 2460

tttagcgcca acctgcggaa gaagagaggc gccgagaatc tgcacacaat gctgtttaag 2520

gccctgatgg agggcaacca ggataatctg gacctgggct ctggcgccat cttctacaga 2580

gccaagagcc tggacggcaa caagcccaca caccctgcca atgaggccat caagtgtagg 2640

aacgtggcca ataaggataa ggtgtccctg ttcacctacg acatctataa gaacaggcgc 2700

tacatggaga ataagttcct gtttcacctg agcatcgtgc agaactataa ggccgccaat 2760

gactccgccc agctgaacag ctccgccacc gagtatatca gaaaggccga tgacctgcac 2820

atcatcggca tcgatagggg cgagcgcaat ctgctgtact attccgtgat cgatatgaag 2880

ggcaacatcg tggagcagga ctctctgaat atcatcagga acaatgacct ggagacagat 2940

taccacgacc tgctggataa gagggagaag gagcgcaagg ccaaccggca gaattgggag 3000

gccgtggagg gcatcaagga cctgaagaag ggctacctga gccaggccgt gcaccagatc 3060

gcccagctga tgctgaagta taacgccatc atcgccctgg aggatctggg ccagatgttt 3120

gtgacccgcg gccagaagat cgagaaggcc gtgtaccagc agttcgagaa gagcctggtg 3180

gataagctgt cctacctggt ggacaagaag cggccttata atgagctggg cggcatcctg 3240

aaggcctacc agctggcctc tagcatcacc aagaacaatt ctgacaagca gaacggcttc 3300

ctgttttatg tgccagcctg gaatacaagc aagatcgatc ccgtgaccgg ctttacagac 3360

ctgctgcggc ccaaggccat gaccatcaag gaggcccagg acttctttgg cgccttcgat 3420

aacatctctt acaatgacaa gggctatttc gagtttgaga caaactacga caagtttaag 3480

atcagaatga agagcgccca gaccaggtgg acaatctgca ccttcggcaa tcggatcaag 3540

agaaagaagg ataagaacta ctggaattat gaggaggtgg agctgaccga ggagttcaag 3600

aagctgttta aggacagcaa catcgattac gagaactgta atctgaagga ggagatccag 3660

aacaaggaca atcgcaagtt ctttgatgac ctgatcaagc tgctgcagct gacactgcag 3720

atgcggaact ccgatgacaa gggcaatgat tatatcatct ctcctgtggc caacgccgag 3780

ggccagttct ttgactcccg caatggcgat aagaagctgc cactggatgc agacgcaaac 3840

ggagcctaca atatcgcccg caagggcctg tggaacatcc ggcagatcaa gcagaccaag 3900

aacgacaaga agctgaatct gagcatctcc tctacagagt ggctggattt cgtgcgggag 3960

aagccttacc tgaagaaaag gccggcggcc acgaaaaagg ccggccaggc aaaaaagaaa 4020

aagggatcct acccatacga tgttccagat tacgcttatc cctacgacgt gcctgattat 4080

gcatacccat atgatgtccc cgactatgcc taagaattc 4119

<210> 1277
<211> 1323
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 1277
Met Glu Asn Tyr Gln Glu Phe Thr Asn Leu Phe Gln Leu Asn Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Cys Glu Leu Leu Glu
20 25 30

Glu Gly Lys Ile Phe Ala Ser Gly Ser Phe Leu Glu Lys Asp Lys Val
35 40 45

Arg Ala Asp Asn Val Ser Tyr Val Lys Lys Glu Ile Asp Lys Lys His
50 55 60

Lys Ile Phe Ile Glu Glu Thr Leu Ser Ser Phe Ser Ile Ser Asn Asp
65 70 75 80

Leu Leu Lys Gln Tyr Phe Asp Cys Tyr Asn Glu Leu Lys Ala Phe Lys
85 90 95

Lys Asp Cys Lys Ser Asp Glu Glu Glu Val Lys Lys Thr Ala Leu Arg
100 105 110

Asn Lys Cys Thr Ser Ile Gln Arg Ala Met Arg Glu Ala Ile Ser Gln
115 120 125

Ala Phe Leu Lys Ser Pro Gln Lys Lys Leu Leu Ala Ile Lys Asn Leu
130 135 140

Ile Glu Asn Val Phe Lys Ala Asp Glu Asn Val Gln His Phe Ser Glu
145 150 155 160

Phe Thr Ser Tyr Phe Ser Gly Phe Glu Thr Asn Arg Glu Asn Phe Tyr
165 170 175

Ser Asp Glu Glu Lys Ser Thr Ser Ile Ala Tyr Arg Leu Val His Asp
180 185 190

Asn Leu Pro Ile Phe Ile Lys Asn Ile Tyr Ile Phe Glu Lys Leu Lys
195 200 205

Glu Gln Phe Asp Ala Lys Thr Leu Ser Glu Ile Phe Glu Asn Tyr Lys
210 215 220

Leu Tyr Val Ala Gly Ser Ser Leu Asp Glu Val Phe Ser Leu Glu Tyr
225 230 235 240

Phe Asn Asn Thr Leu Thr Gln Lys Gly Ile Asp Asn Tyr Asn Ala Val
245 250 255

Ile Gly Lys Ile Val Lys Glu Asp Lys Gln Glu Ile Gln Gly Leu Asn
260 265 270

Glu His Ile Asn Leu Tyr Asn Gln Lys His Lys Asp Arg Arg Leu Pro

275 280 285

Phe Phe Ile Ser Leu Lys Lys Gln Ile Leu Ser Asp Arg Glu Ala Leu
290 295 300

Ser Trp Leu Pro Asp Met Phe Lys Asn Asp Ser Glu Val Ile Lys Ala
305 310 315 320

Leu Lys Gly Phe Tyr Ile Glu Asp Gly Phe Glu Asn Asn Val Leu Thr
325 330 335

Pro Leu Ala Thr Leu Leu Ser Ser Leu Asp Lys Tyr Asn Leu Asn Gly
340 345 350

Ile Phe Ile Arg Asn Asn Glu Ala Leu Ser Ser Leu Ser Gln Asn Val
355 360 365

Tyr Arg Asn Phe Ser Ile Asp Glu Ala Ile Asp Ala Asn Ala Glu Leu
370 375 380

Gln Thr Phe Asn Asn Tyr Glu Leu Ile Ala Asn Ala Leu Arg Ala Lys
385 390 395 400

Ile Lys Lys Glu Thr Lys Gln Gly Arg Lys Ser Phe Glu Lys Tyr Glu
405 410 415

Glu Tyr Ile Asp Lys Lys Val Lys Ala Ile Asp Ser Leu Ser Ile Gln
420 425 430

Glu Ile Asn Glu Leu Val Glu Asn Tyr Val Ser Glu Phe Asn Ser Asn
435 440 445

Ser Gly Asn Met Pro Arg Lys Val Glu Asp Tyr Phe Ser Leu Met Arg
450 455 460

Lys Gly Asp Phe Gly Ser Asn Asp Leu Ile Glu Asn Ile Lys Thr Lys
465 470 475 480

Leu Ser Ala Ala Glu Lys Leu Leu Gly Thr Lys Tyr Gln Glu Thr Ala
485 490 495

Lys Asp Ile Phe Lys Lys Asp Glu Asn Ser Lys Leu Ile Lys Glu Leu
500 505 510

Leu Asp Ala Thr Lys Gln Phe Gln His Phe Ile Lys Pro Leu Leu Gly
515 520 525

Thr Gly Glu Glu Ala Asp Arg Asp Leu Val Phe Tyr Gly Asp Phe Leu
530 535 540

Pro Leu Tyr Glu Lys Phe Glu Glu Leu Thr Leu Leu Tyr Asn Lys Val
545 550 555 560

Arg Asn Arg Leu Thr Gln Lys Pro Tyr Ser Lys Asp Lys Ile Arg Leu
565 570 575

Cys Phe Asn Lys Pro Lys Leu Met Thr Gly Trp Val Asp Ser Lys Thr
580 585 590

Glu Lys Ser Asp Asn Gly Thr Gln Tyr Gly Gly Tyr Leu Phe Arg Lys
595 600 605

Lys Asn Glu Ile Gly Glu Tyr Asp Tyr Phe Leu Gly Ile Ser Ser Lys

610 615 620

Ala Gln Leu Phe Arg Lys Asn Glu Ala Val Ile Gly Asp Tyr Glu Arg
625 630 635 640

Leu Asp Tyr Tyr Gln Pro Lys Ala Asn Thr Ile Tyr Gly Ser Ala Tyr
645 650 655

Glu Gly Glu Asn Ser Tyr Lys Glu Asp Lys Lys Arg Leu Asn Lys Val
660 665 670

Ile Ile Ala Tyr Ile Glu Gln Ile Lys Gln Thr Asn Ile Lys Lys Ser
675 680 685

Ile Ile Glu Ser Ile Ser Lys Tyr Pro Asn Ile Ser Asp Asp Asp Lys
690 695 700

Val Thr Pro Ser Ser Leu Leu Glu Lys Ile Lys Lys Val Ser Ile Asp
705 710 715 720

Ser Tyr Asn Gly Ile Leu Ser Phe Lys Ser Phe Gln Ser Val Asn Lys
725 730 735

Glu Val Ile Asp Asn Leu Leu Lys Thr Ile Ser Pro Leu Lys Asn Lys
740 745 750

Ala Glu Phe Leu Asp Leu Ile Asn Lys Asp Tyr Gln Ile Phe Thr Glu
755 760 765

Val Gln Ala Val Ile Asp Glu Ile Cys Lys Gln Lys Thr Phe Ile Tyr
770 775 780

Phe Pro Ile Ser Asn Val Glu Leu Glu Lys Glu Met Gly Asp Lys Asp
785 790 795 800

Lys Pro Leu Cys Leu Phe Gln Ile Ser Asn Lys Asp Leu Ser Phe Ala
805 810 815

Lys Thr Phe Ser Ala Asn Leu Arg Lys Lys Arg Gly Ala Glu Asn Leu
820 825 830

His Thr Met Leu Phe Lys Ala Leu Met Glu Gly Asn Gln Asp Asn Leu
835 840 845

Asp Leu Gly Ser Gly Ala Ile Phe Tyr Arg Ala Lys Ser Leu Asp Gly
850 855 860

Asn Lys Pro Thr His Pro Ala Asn Glu Ala Ile Lys Cys Arg Asn Val
865 870 875 880

Ala Asn Lys Asp Lys Val Ser Leu Phe Thr Tyr Asp Ile Tyr Lys Asn
885 890 895

Arg Arg Tyr Met Glu Asn Lys Phe Leu Phe His Leu Ser Ile Val Gln
900 905 910

Asn Tyr Lys Ala Ala Asn Asp Ser Ala Gln Leu Asn Ser Ser Ala Thr
915 920 925

Glu Tyr Ile Arg Lys Ala Asp Asp Leu His Ile Ile Gly Ile Asp Arg
930 935 940

Gly Glu Arg Asn Leu Leu Tyr Tyr Ser Val Ile Asp Met Lys Gly Asn

945 950 955 960

Ile Val Glu Gln Asp Ser Leu Asn Ile Ile Arg Asn Asn Asp Leu Glu
965 970 975

Thr Asp Tyr His Asp Leu Leu Asp Lys Arg Glu Lys Glu Arg Lys Ala
980 985 990

Asn Arg Gln Asn Trp Glu Ala Val Glu Gly Ile Lys Asp Leu Lys Lys
995 1000 1005

Gly Tyr Leu Ser Gln Ala Val His Gln Ile Ala Gln Leu Met Leu
1010 1015 1020

Lys Tyr Asn Ala Ile Ile Ala Leu Glu Asp Leu Gly Gln Met Phe
1025 1030 1035

Val Thr Arg Gly Gln Lys Ile Glu Lys Ala Val Tyr Gln Gln Phe
1040 1045 1050

Glu Lys Ser Leu Val Asp Lys Leu Ser Tyr Leu Val Asp Lys Lys
1055 1060 1065

Arg Pro Tyr Asn Glu Leu Gly Gly Ile Leu Lys Ala Tyr Gln Leu
1070 1075 1080

Ala Ser Ser Ile Thr Lys Asn Asn Ser Asp Lys Gln Asn Gly Phe
1085 1090 1095

Leu Phe Tyr Val Pro Ala Trp Asn Thr Ser Lys Ile Asp Pro Val
1100 1105 1110

Thr Gly Phe Thr Asp Leu Leu Arg Pro Lys Ala Met Thr Ile Lys
1115 1120 1125

Glu Ala Gln Asp Phe Phe Gly Ala Phe Asp Asn Ile Ser Tyr Asn
1130 1135 1140

Asp Lys Gly Tyr Phe Glu Phe Glu Thr Asn Tyr Asp Lys Phe Lys
1145 1150 1155

Ile Arg Met Lys Ser Ala Gln Thr Arg Trp Thr Ile Cys Thr Phe
1160 1165 1170

Gly Asn Arg Ile Lys Arg Lys Lys Asp Lys Asn Tyr Trp Asn Tyr
1175 1180 1185

Glu Glu Val Glu Leu Thr Glu Glu Phe Lys Lys Leu Phe Lys Asp
1190 1195 1200

Ser Asn Ile Asp Tyr Glu Asn Cys Asn Leu Lys Glu Glu Ile Gln
1205 1210 1215

Asn Lys Asp Asn Arg Lys Phe Phe Asp Asp Leu Ile Lys Leu Leu
1220 1225 1230

Gln Leu Thr Leu Gln Met Arg Asn Ser Asp Asp Lys Gly Asn Asp
1235 1240 1245

Tyr Ile Ile Ser Pro Val Ala Asn Ala Glu Gly Gln Phe Phe Asp
1250 1255 1260

Ser Arg Asn Gly Asp Lys Lys Leu Pro Leu Asp Ala Asp Ala Asn

1265 1270 1275

Gly Ala Tyr Asn Ile Ala Arg Lys Gly Leu Trp Asn Ile Arg Gln
1280 1285 1290

Ile Lys Gln Thr Lys Asn Asp Lys Lys Leu Asn Leu Ser Ile Ser
1295 1300 1305

Ser Thr Glu Trp Leu Asp Phe Val Arg Glu Lys Pro Tyr Leu Lys
1310 1315 1320

<210> 1278
<211> 3888
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1278
ggtaccatga aaacccagca cttctttgag gacttcacaa gcctgtactc tctgagcaag 60

accatccggt ttgagctgaa gccaatcggc aagaccctgg agaacatcaa gaagaatggc 120

ctgatccgga gagatgagca gagactggac gattacgaga agctgaagaa agtgatcgac 180

gagtatcacg aggatttcat cgccaacatc ctgagctcct tttccttctc tgaggagatc 240

ctgcagtcct acatccagaa tctgagcgag tccgaggcca gggccaagat cgagaaaacc 300

atgcgcgaca cactggccaa ggccttctct gaggatgaga ggtacaagag catctttaag 360

aaggagctgg tgaagaagga catccccgtg tggtgccctg cctataagag cctgtgcaag 420

aagttcgata actttaccac atctctggtg cccttccacg agaacaggaa gaacctgtat 480

accagcaatg agatcacagc ctctatccct tatcgcatcg tgcacgtgaa cctgccaaag 540

tttatccaga atatcgaggc cctgtgcgag ctgcagaaga agatgggcgc cgacctgtac 600

ctggagatga tggagaacct gcgcaacgtg tggcccagct tcgtgaaaac cccagacgac 660

ctgtgcaacc tgaaaaccta taatcacctg atggtgcagt ctagcatcag cgagtacaac 720

aggtttgtgg gcggctattc caccgaggac ggcacaaagc accagggcat caacgagtgg 780

atcaatatct acagacagag gaataaggag atgcgcctgc ctggcctggt gttcctgcac 840

aagcagatcc tggccaaggt ggactcctct agcttcatca gcgatacact ggagaacgac 900

gatcaggtgt tttgcgtgct gagacagttc aggaagctgt tttggaatac cgtgtcctct 960

aaggaggacg atgccgcctc cctgaaggac ctgttctgtg gcctgtctgg ctatgaccct 1020

gaggccatct acgtgagcga tgcccacctg gccacaatct ccaagaacat ctttgacaga 1080

tggaattaca tctccgatgc catcaggcgc aagaccgagg tgctgatgcc acggaagaag 1140

gagagcgtgg agagatatgc cgagaagatc tccaagcaga tcaagaagag acagtcttac 1200

agcctggccg agctggacga tctgctggcc cactatagcg aggagtccct gcccgcaggc 1260

ttctctctgc tgagctactt tacatctctg ggcggccaga agtatctggt gagcgacggc 1320

gaagtgatcc tgtacgagga gggcagcaac atctgggacg aggtgctgat cgccttcagg 1380

gatctgcagg tcatcctgga caaggacttc accgagaaga agctgggcaa ggatgaggag 1440

gccgtgtctg tgatcaagaa ggccctggac agcgccctgc gcctgcggaa gttctttgat 1500

ctgctgtccg gcacaggcgc agagatcagg agagacagct ccttctatgc cctgtatacc 1560

gaccggatgg ataagctgaa gggcctgctg aagatgtatg ataaggtgag aaactacctg 1620

accaagaagc cttattccat cgagaagttc aagctgcact ttgacaaccc atccctgctg 1680

tctggctggg ataagaataa ggagctgaac aatctgtctg tgatcttccg gcagaacggc 1740

tactattacc tgggcatcat gacacccaag ggcaagaatc tgttcaagac cctgcctaag 1800

ctgggcgccg aggagatgtt ttatgagaag atggagtaca agcagatcgc cgagcctatg 1860

ctgatgctgc caaaggtgtt ctttcccaag aaaaccaagc cagccttcgc cccagaccag 1920

agcgtggtgg atatctacaa caagaaaacc ttcaagacag gccagaaggg ctttaataag 1980

aaggacctgt accggctgat cgacttctac aaggaggccc tgacagtgca cgagtggaag 2040

ctgtttaact tctccttttc tccaaccgag cagtatcgga atatcggcga gttctttgac 2100

gaggtgagag agcaggccta caaggtgtcc atggtgaacg tgcccgcctc ttatatcgac 2160

gaggccgtgg agaacggcaa gctgtatctg ttccagatct acaataagga cttcagcccc 2220

tactccaagg gcatccctaa cctgcacaca ctgtattgga aggccctgtt cagcgagcag 2280

aatcagagcc gggtgtataa gctgtgcgga ggaggagagc tgttttatag aaaggccagc 2340

ctgcacatgc aggacaccac agtgcacccc aagggcatct ctatccacaa gaagaacctg 2400

aataagaagg gcgagacaag cctgttcaac tacgacctgg tgaaggataa gaggtttacc 2460

gaggacaagt tctttttcca cgtgcctatc tctatcaact acaagaataa gaagatcacc 2520

aacgtgaatc agatggtgcg cgattatatc gcccagaacg acgatctgca gatcatcggc 2580

atcgaccgcg gcgagcggaa tctgctgtat atcagccgga tcgatacaag gggcaacctg 2640

ctggagcagt tcagcctgaa tgtgatcgag tccgacaagg gcgatctgag aaccgactat 2700

cagaagatcc tgggcgatcg cgagcaggag cggctgaggc gccggcagga gtggaagtct 2760

atcgagagca tcaaggacct gaaggatggc tacatgagcc aggtggtgca caagatctgt 2820

aacatggtgg tggagcacaa ggccatcgtg gtgctggaga acctgaatct gagcttcatg 2880

aagggcagga agaaggtgga gaagtccgtg tacgagaagt ttgagcgcat gctggtggac 2940

aagctgaact atctggtggt ggataagaag aacctgtcca atgagccagg aggcctgtat 3000

gcagcatacc agctgaccaa tccactgttc tcttttgagg agctgcacag atacccccag 3060

agcggcatcc tgtttttcgt ggacccatgg aacacctctc tgacagatcc cagcacaggc 3120

ttcgtgaatc tgctgggcag aatcaactac accaatgtgg gcgacgcccg caagtttttc 3180

gatcggttta acgccatcag atatgacggc aagggcaata tcctgttcga cctggatctg 3240

tccagatttg atgtgagggt ggagacacag aggaagctgt ggacactgac cacattcggc 3300

tctcgcatcg ccaaatccaa gaagtctggc aagtggatgg tggagcggat cgagaacctg 3360

agcctgtgct ttctggagct gttcgagcag tttaatatcg gctacagagt ggagaaggac 3420

ctgaagaagg ccatcctgag ccaggatagg aaggagttct atgtgcgcct gatctacctg 3480

tttaacctga tgatgcagat ccggaacagc gacggcgagg aggattatat cctgtctccc 3540

gccctgaacg agaagaatct gcagttcgac agcaggctga tcgaggccaa ggatctgcct 3600

gtggacgcag atgcaaacgg agcatacaat gtggcccgca agggcctgat ggtggtgcag 3660

agaatcaaga ggggcgacca cgagtccatc cacaggatcg gaagggcaca gtggctgaga 3720

tatgtgcagg agggcatcgt ggagaaaagg ccggcggcca cgaaaaaggc cggccaggca 3780

aaaaagaaaa agggatccta cccatacgat gttccagatt acgcttatcc ctacgacgtg 3840

cctgattatg catacccata tgatgtcccc gactatgcct aagaattc 3888

<210> 1279
<211> 1246
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 1279
Met Lys Thr Gln His Phe Phe Glu Asp Phe Thr Ser Leu Tyr Ser Leu
1 5 10 15

```
Ser Lys Thr Ile Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Leu Glu
            20                  25                  30

Asn Ile Lys Lys Asn Gly Leu Ile Arg Arg Asp Glu Gln Arg Leu Asp
        35                  40                  45

Asp Tyr Glu Lys Leu Lys Lys Val Ile Asp Glu Tyr His Glu Asp Phe
    50                  55                  60

Ile Ala Asn Ile Leu Ser Ser Phe Ser Phe Ser Glu Glu Ile Leu Gln
65                  70                  75                  80

Ser Tyr Ile Gln Asn Leu Ser Glu Ser Glu Ala Arg Ala Lys Ile Glu
                85                  90                  95

Lys Thr Met Arg Asp Thr Leu Ala Lys Ala Phe Ser Glu Asp Glu Arg
            100                 105                 110

Tyr Lys Ser Ile Phe Lys Lys Glu Leu Val Lys Lys Asp Ile Pro Val
        115                 120                 125

Trp Cys Pro Ala Tyr Lys Ser Leu Cys Lys Lys Phe Asp Asn Phe Thr
    130                 135                 140

Thr Ser Leu Val Pro Phe His Glu Asn Arg Lys Asn Leu Tyr Thr Ser
145                 150                 155                 160

Asn Glu Ile Thr Ala Ser Ile Pro Tyr Arg Ile Val His Val Asn Leu
                165                 170                 175

Pro Lys Phe Ile Gln Asn Ile Glu Ala Leu Cys Glu Leu Gln Lys Lys
            180                 185                 190

Met Gly Ala Asp Leu Tyr Leu Glu Met Met Glu Asn Leu Arg Asn Val
        195                 200                 205

Trp Pro Ser Phe Val Lys Thr Pro Asp Asp Leu Cys Asn Leu Lys Thr
    210                 215                 220

Tyr Asn His Leu Met Val Gln Ser Ser Ile Ser Glu Tyr Asn Arg Phe
225                 230                 235                 240

Val Gly Gly Tyr Ser Thr Glu Asp Gly Thr Lys His Gln Gly Ile Asn
                245                 250                 255

Glu Trp Ile Asn Ile Tyr Arg Gln Arg Asn Lys Glu Met Arg Leu Pro
            260                 265                 270

Gly Leu Val Phe Leu His Lys Gln Ile Leu Ala Lys Val Asp Ser Ser
        275                 280                 285

Ser Phe Ile Ser Asp Thr Leu Glu Asn Asp Asp Gln Val Phe Cys Val
    290                 295                 300

Leu Arg Gln Phe Arg Lys Leu Phe Trp Asn Thr Val Ser Ser Lys Glu
305                 310                 315                 320

Asp Asp Ala Ala Ser Leu Lys Asp Leu Phe Cys Gly Leu Ser Gly Tyr
                325                 330                 335

Asp Pro Glu Ala Ile Tyr Val Ser Asp Ala His Leu Ala Thr Ile Ser
            340                 345                 350
```

Lys Asn Ile Phe Asp Arg Trp Asn Tyr Ile Ser Asp Ala Ile Arg Arg
355 360 365

Lys Thr Glu Val Leu Met Pro Arg Lys Lys Glu Ser Val Glu Arg Tyr
370 375 380

Ala Glu Lys Ile Ser Lys Gln Ile Lys Lys Arg Gln Ser Tyr Ser Leu
385 390 395 400

Ala Glu Leu Asp Asp Leu Leu Ala His Tyr Ser Glu Glu Ser Leu Pro
405 410 415

Ala Gly Phe Ser Leu Leu Ser Tyr Phe Thr Ser Leu Gly Gly Gln Lys
420 425 430

Tyr Leu Val Ser Asp Gly Glu Val Ile Leu Tyr Glu Glu Gly Ser Asn
435 440 445

Ile Trp Asp Glu Val Leu Ile Ala Phe Arg Asp Leu Gln Val Ile Leu
450 455 460

Asp Lys Asp Phe Thr Glu Lys Lys Leu Gly Lys Asp Glu Glu Ala Val
465 470 475 480

Ser Val Ile Lys Lys Ala Leu Asp Ser Ala Leu Arg Leu Arg Lys Phe
485 490 495

Phe Asp Leu Leu Ser Gly Thr Gly Ala Glu Ile Arg Arg Asp Ser Ser
500 505 510

Phe Tyr Ala Leu Tyr Thr Asp Arg Met Asp Lys Leu Lys Gly Leu Leu
515 520 525

Lys Met Tyr Asp Lys Val Arg Asn Tyr Leu Thr Lys Lys Pro Tyr Ser
530 535 540

Ile Glu Lys Phe Lys Leu His Phe Asp Asn Pro Ser Leu Leu Ser Gly
545 550 555 560

Trp Asp Lys Asn Lys Glu Leu Asn Asn Leu Ser Val Ile Phe Arg Gln
565 570 575

Asn Gly Tyr Tyr Tyr Leu Gly Ile Met Thr Pro Lys Gly Lys Asn Leu
580 585 590

Phe Lys Thr Leu Pro Lys Leu Gly Ala Glu Glu Met Phe Tyr Glu Lys
595 600 605

Met Glu Tyr Lys Gln Ile Ala Glu Pro Met Leu Met Leu Pro Lys Val
610 615 620

Phe Phe Pro Lys Lys Thr Lys Pro Ala Phe Ala Pro Asp Gln Ser Val
625 630 635 640

Val Asp Ile Tyr Asn Lys Lys Thr Phe Lys Thr Gly Gln Lys Gly Phe
645 650 655

Asn Lys Lys Asp Leu Tyr Arg Leu Ile Asp Phe Tyr Lys Glu Ala Leu
660 665 670

Thr Val His Glu Trp Lys Leu Phe Asn Phe Ser Phe Ser Pro Thr Glu
675 680 685

Gln Tyr Arg Asn Ile Gly Glu Phe Phe Asp Glu Val Arg Glu Gln Ala
690 695 700

Tyr Lys Val Ser Met Val Asn Val Pro Ala Ser Tyr Ile Asp Glu Ala
705 710 715 720

Val Glu Asn Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe
725 730 735

Ser Pro Tyr Ser Lys Gly Ile Pro Asn Leu His Thr Leu Tyr Trp Lys
740 745 750

Ala Leu Phe Ser Glu Gln Asn Gln Ser Arg Val Tyr Lys Leu Cys Gly
755 760 765

Gly Gly Glu Leu Phe Tyr Arg Lys Ala Ser Leu His Met Gln Asp Thr
770 775 780

Thr Val His Pro Lys Gly Ile Ser Ile His Lys Lys Asn Leu Asn Lys
785 790 795 800

Lys Gly Glu Thr Ser Leu Phe Asn Tyr Asp Leu Val Lys Asp Lys Arg
805 810 815

Phe Thr Glu Asp Lys Phe Phe Phe His Val Pro Ile Ser Ile Asn Tyr
820 825 830

Lys Asn Lys Lys Ile Thr Asn Val Asn Gln Met Val Arg Asp Tyr Ile
835 840 845

Ala Gln Asn Asp Asp Leu Gln Ile Ile Gly Ile Asp Arg Gly Glu Arg
850 855 860

Asn Leu Leu Tyr Ile Ser Arg Ile Asp Thr Arg Gly Asn Leu Leu Glu
865 870 875 880

Gln Phe Ser Leu Asn Val Ile Glu Ser Asp Lys Gly Asp Leu Arg Thr
885 890 895

Asp Tyr Gln Lys Ile Leu Gly Asp Arg Glu Gln Glu Arg Leu Arg Arg
900 905 910

Arg Gln Glu Trp Lys Ser Ile Glu Ser Ile Lys Asp Leu Lys Asp Gly
915 920 925

Tyr Met Ser Gln Val Val His Lys Ile Cys Asn Met Val Val Glu His
930 935 940

Lys Ala Ile Val Val Leu Glu Asn Leu Asn Leu Ser Phe Met Lys Gly
945 950 955 960

Arg Lys Lys Val Glu Lys Ser Val Tyr Glu Lys Phe Glu Arg Met Leu
965 970 975

Val Asp Lys Leu Asn Tyr Leu Val Val Asp Lys Lys Asn Leu Ser Asn
980 985 990

Glu Pro Gly Gly Leu Tyr Ala Ala Tyr Gln Leu Thr Asn Pro Leu Phe
995 1000 1005

Ser Phe Glu Glu Leu His Arg Tyr Pro Gln Ser Gly Ile Leu Phe
1010 1015 1020

Phe Val Asp Pro Trp Asn Thr Ser Leu Thr Asp Pro Ser Thr Gly
1025 1030 1035

Phe Val Asn Leu Leu Gly Arg Ile Asn Tyr Thr Asn Val Gly Asp
1040 1045 1050

Ala Arg Lys Phe Phe Asp Arg Phe Asn Ala Ile Arg Tyr Asp Gly
1055 1060 1065

Lys Gly Asn Ile Leu Phe Asp Leu Asp Leu Ser Arg Phe Asp Val
1070 1075 1080

Arg Val Glu Thr Gln Arg Lys Leu Trp Thr Leu Thr Thr Phe Gly
1085 1090 1095

Ser Arg Ile Ala Lys Ser Lys Lys Ser Gly Lys Trp Met Val Glu
1100 1105 1110

Arg Ile Glu Asn Leu Ser Leu Cys Phe Leu Glu Leu Phe Glu Gln
1115 1120 1125

Phe Asn Ile Gly Tyr Arg Val Glu Lys Asp Leu Lys Lys Ala Ile
1130 1135 1140

Leu Ser Gln Asp Arg Lys Glu Phe Tyr Val Arg Leu Ile Tyr Leu
1145 1150 1155

Phe Asn Leu Met Met Gln Ile Arg Asn Ser Asp Gly Glu Glu Asp
1160 1165 1170

Tyr Ile Leu Ser Pro Ala Leu Asn Glu Lys Asn Leu Gln Phe Asp
1175 1180 1185

Ser Arg Leu Ile Glu Ala Lys Asp Leu Pro Val Asp Ala Asp Ala
1190 1195 1200

Asn Gly Ala Tyr Asn Val Ala Arg Lys Gly Leu Met Val Val Gln
1205 1210 1215

Arg Ile Lys Arg Gly Asp His Glu Ser Ile His Arg Ile Gly Arg
1220 1225 1230

Ala Gln Trp Leu Arg Tyr Val Gln Glu Gly Ile Val Glu
1235 1240 1245

<210> 1280
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1280
gucuaagaac uuuaaauaau uucuacuguu guagau 36

<210> 1281
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1281
gucuaagaac uuuaaauaau uucuacuguu guagau 36

<210> 1282
<211> 35
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1282
gcuauaaugc cuauauaauu ucuacuauug uagau 35

<210> 1283
<211> 35
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1283
gcuauaaugc cuauauaauu ucuacuauug uagau 35

<210> 1284
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1284
gccuauaagg cuuuaguaau uucuacuauu guagau 36

<210> 1285
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1285
guuuuggagu accuuagaaa ugcaugguuc ucaugc 36

<210> 1286
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1286
gucaauaaga cucauuuaau uucuacuucg guagau 36

<210> 1287
<211> 35
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1287
guuuuagaac cuuaaaaauu accuaguaau uaggu 35

<210> 1288
<211> 36

<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1288
gucuagguac ucucuuuaau uucuacuauu guagau 36

<210> 1289
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1289
guuuaaaagu ccuauuggau uucuacuuuu guagau 36

<210> 1290
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1290
guuucaaaga uuaaauaauu ucuacuaagu guagau 36

<210> 1291
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1291
gccaaauacc ucuauaaaau uucuacuuuu guagau 36

<210> 1292
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1292
cucuaaagag aggaaagaau uucuacuuuu guagau 36

<210> 1293
<211> 32
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1293
guuucaaucc acgcgcccac gcggggcgcg ac 32

<210> 1294
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1294
gucuaacgac cuuuuaaauu ucuacuguuu guagau 36

<210> 1295
<211> 35
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1295
gucaaaagac cuuuuuaauu ucuacucuug uagau 35

<210> 1296
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1296
guuugaauaa ccuuaaauaa uuucuacuuu guagau 36

<210> 1297
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1297
gcuuagaaca uuuaaagaau uucuacuauu guagau 36

<210> 1298
<211> 35
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1298
cucaaaacuc auucgaaucu cuacucuuug uagau 35

<210> 1299
<211> 35
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1299
cucaaaacuc auucgaaucu cuacucuuug uagau 35

<210> 1300
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1300
gcuuagaaca uuuaaagaau uucuacuauu guagau 36

<210> 1301

<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1301
guuugaauaa ccuuaaauaa uuucuacuuu guagau 36

<210> 1302
<211> 35
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1302
gucaaaagac cuuuuuaauu ucuacucuug uagau 35

<210> 1303
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1303
gucuaacgac cuuuuaaauu ucuacuguuu guagau 36

<210> 1304
<211> 32
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1304
guuucaaucc acgcgcccac gcggggcgcg ac 32

<210> 1305
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1305
cucuaaagag aggaaagaau uucuacuuuu guagau 36

<210> 1306
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1306
gccaaauacc ucuauaaaau uucuacuuuu guagau 36

<210> 1307
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的

寡核苷酸

<400> 1307
guuucaaaga uuaaauaauu ucuacuaagu guagau 36

<210> 1308
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1308
guuuaaaagu ccuauuggau uucuacuuuu guagau 36

<210> 1309
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1309
gucuagguac ucucuuuaau uucuacuauu guagau 36

<210> 1310
<211> 35
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1310
guuuuagaac cuuaaaaauu accuaguaau uaggu 35

<210> 1311
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1311
gucaauaaga cucauuuaau uucuacuucg guagau 36

<210> 1312
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1312
guuuuggagu accuuagaaa ugcaugguuc ucaugc 36

<210> 1313
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1313
gccuauaagg cuuuaguaau uucuacuauu guagau 36

<210> 1314
<211> 77
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1314
gggacuuuaa auaauuucua cuguuguaga uaggccccag uggcugcucu gggggccucc 60

gucuaagaac uuuaaau 77

<210> 1315
<211> 77
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1315
gggacuuuaa auaauuucua cuguuguaga uucaucugug ccccucccuc ccuggcccag 60

gucuaagaac uuuaaau 77

<210> 1316
<211> 77
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1316
gggacuuuaa auaauuucua cuguuguaga ugugguugcc cacccuaguc auuggaggug 60

gucuaagaac uuuaaau 77

<210> 1317
<211> 77
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1317
gggacuuuaa auaauuucua cuguuguaga uggggcccca ggccgggguc cccucugacc 60

gucuaagaac uuuaaau 77

<210> 1318
<211> 639
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1318
ccatcccctt ctgtgaatgt tagacccatg ggagcagctg gtcagagggg accccggcct 60

ggggccccta accctatgta gcctcagtct tcccatcagg ctctcagctc agcctgagtg 120

ttgaggcccc agtggctgct ctgggggcct cctgagtttc tcatctgtgc ccctccctcc 180

ctggcccagg tgaaggtgtg gttccagaac cggaggacaa agtacaaacg gcagaagctg 240

gaggaggaag ggcctgagtc cgagcagaag aagaagggct cccatcacat caaccggtgg 300

cgcattgcca cgaagcaggc caatggggag gacatcgatg tcacctccaa tgactagggt 360

gggcaaccac aaacccacga gggcagagtg ctgcttgctg ctggccaggc ccctgcgtgg 420

gcccaagctg gactctggcc actccctggc caggctttgg ggaggcctgg agtcatggcc 480

ccacagggct tgaagcccgg ggccgccatt gacagaggga caagcaatgg gctggctgag 540

gcctgggacc acttggcctt ctcctcggag agcctgcctg cctgggcggg cccgcccgcc 600

accgcagcct cccagctgct ctccgtgtct ccaatctcc 639

<210> 1319
<211> 106
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1319
ggggucuaag aacuuuaaau aauuucuacu guuguagaug agaagucauu uaauaaggcc 60

acuguuaaaa gucuaagaac uuuaaauaau uucuacuguu guagau 106

<210> 1320
<211> 100
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1320
ggggaacuuu aaauaauuuc uacuguugua gaugagaagu cauuuaauaa ggccacuguu 60

aaaagucuaa gaacuuuaaa uaauuucuac uguuguagau 100

<210> 1321
<211> 94
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1321
ggguaaauaa uuucuacugu uguagaugag aagucauuua auaaggccac uguuaaaagu 60

cuaagaacuu uaaauaauuu cuacuguugu agau 94

<210> 1322
<211> 88
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1322
gggauuucua cuguuguaga ugagaaguca uuuaauaagg ccacuguuaa aagucuaaga 60

acuuuaaaua auuucuacug uuguagau 88

<210> 1323
<211> 82
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1323
gggacuguug uagaugagaa gucauuuaau aaggccacug uuaaaagucu aagaacuuua 60

aauaauuucu acuguuguag au 82

<210> 1324
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1324
gucuaagaac uuuaaauaau uucuacuguu guagau 36

<210> 1325
<211> 79
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1325
gggactttaa ataatttcta ctgttgtaga ttagaagtca tttaataagg ccactgttaa 60

aagtctaaga actttaaat 79

<210> 1326
<211> 79
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1326
gggactttaa ataatttcta ctgttgtaga tgaggagtca tttaataagg ccactgttaa 60

aagtctaaga actttaaat 79

<210> 1327
<211> 79
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1327
gggactttaa ataatttcta ctgttgtaga tgagaagcca tttaataagg ccactgttaa 60

aagtctaaga actttaaat 79

<210> 1328
<211> 79
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1328
gggactttaa ataatttcta ctgttgtaga tgagaagtca attaataagg ccactgttaa 60

aagtctaaga actttaaat 79

<210> 1329
<211> 79
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1329
gggactttaa ataatttcta ctgttgtaga tgagaagtca ttttataagg ccactgttaa 60

aagtctaaga actttaaat 79

<210> 1330
<211> 79
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1330
gggactttaa ataatttcta ctgttgtaga tgagaagtca tttaatgagg ccactgttaa 60

aagtctaaga actttaaat 79

<210> 1331
<211> 79
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1331
gggactttaa ataatttcta ctgttgtaga tgagaagtca tttaataagc ccactgttaa 60

aagtctaaga actttaaat 79

<210> 1332
<211> 79
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1332
gggactttaa ataatttcta ctgttgtaga tgagaagtca tttaataagg cctctgttaa 60

aagtctaaga actttaaat 79

<210> 1333
<211> 79
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1333
gggactttaa ataatttcta ctgttgtaga tgagaagtca tttaataagg ccactattaa 60

aagtctaaga actttaaat 79

<210> 1334
<211> 79
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1334
gggactttaa ataatttcta ctgttgtaga tgagaagtca tttaataagg ccactgttta 60

aagtctaaga actttaaat 79

<210> 1335
<211> 79
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1335
gggactttaa ataatttcta ctgttgtaga tgagaagtca tttaataagg ccactgttaa 60

aggtctaaga actttaaat 79

<210> 1336
<211> 76
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1336
acuuuaaaua auuucuacug uuguagauga gaagucauuu aauaaggcca cuguuaaaag 60

ucuaagaacu uuaaau 76

<210> 1337
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1337
gucuaagaac uuuaaauaau uucuacuguu guagau 36

<210> 1338
<211> 35
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1338
gcuauaaugc cuauauaauu ucuacuauug uagau 35

<210> 1339
<211> 43
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>
<221> 修飾的鹼基
<222> (20)..(43)
<223> a、c、u、g、未知或其他

<220>
<221> misc_feature
<222> (40)..(43)
<223> 可以存在或可以不存在

<400> 1339
aauuucuacu guuguagaun nnnnnnnnnn nnnnnnnnnn nnn 43

<210> 1340
<211> 43
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>
<221> 修飾的鹼基
<222> (20)..(43)
<223> a、c、u、g、未知或其他

<400> 1340
aauuucuacu guuguagaun nnnnnnnnnn nnnnnnnnnn nnn 43

<210> 1341
<211> 43
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>
<221> 修飾的鹼基
<222> (20)..(43)
<223> a、c、u、g、未知或其他

<220>
<221> misc_feature
<222> (40)..(43)
<223> 可以存在或可以不存在

<400> 1341
aauuucuacu auuguagaun nnnnnnnnnn nnnnnnnnnn nnn 43

<210> 1342
<211> 43
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>
<221> 修飾的鹼基
<222> (20)..(43)
<223> a、c、u、g、未知或其他

<400> 1342
aauuucuacu auuguagaun nnnnnnnnnn nnnnnnnnnn nnn 43

<210> 1343
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1343
ttaataaggc cactgttaaa a 21

<210> 1344
<211> 14
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>
<221> 修飾的鹼基
<222> (14)..(14)
<223> a、c、t、g、未知或未知

<400> 1344
ttaataaggc cacn 14

<210> 1345
<211> 13
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1345
gccactgtta aaa 13

<210> 1346
<211> 43
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1346
aauuucuacu uguguagaug agaagucauu uaauaaggcc acu 43

<210> 1347
<211> 22
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1347
gagaagucau uuaauaaggc ca 22

<210> 1348
<211> 20
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1348
gagaagucau uuaauaaggc 20

<210> 1349
<211> 18
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1349
gagaagucau uuaauaag 18

<210> 1350
<211> 17
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1350
gagaagucau uuaauaa 17

<210> 1351
<211> 16
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1351
gagaagucau uuaaua 16

<210> 1352
<211> 15
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1352
gagaagucau uuaau 15

<210> 1353
<211> 68
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1353
gaattatctg aaggcacagg aatagtagca tctacaacag tagaaattat ttaaagttct 60

tagacttt 68

<210> 1354
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1354
atctacaaca gtagaaatta tttaaagttc ttagac 36

<210> 1355
<211> 42
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1355
tctgaaggca caggaatagt agcatctaca acagtagaaa tt 42

<210> 1356
<211> 42
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1356
tctgaaggca caggaatagt agcatctaca acagtagaaa tt 42

<210> 1357
<211> 42
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1357
tctgaaggca caggaatagt agcatctaca acagtagaaa tt 42

<210> 1358
<211> 13
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1358
atctacaaca gta 13

<210> 1359
<211> 13
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1359
atctacaaca gta 13

<210> 1360
<211> 13
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1360
atctacaaca gta 13

<210> 1361
<211> 13
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1361
atctacaaca gta 13

<210> 1362
<211> 13
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1362
atctacaaca gta 13

<210> 1363
<211> 13
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1363
atctacaaca gta 13

<210> 1364
<211> 77

<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1364
agacttttaa cagtggcctt attaaatgac ttctcatcta caacagtaga aattatttaa 60

agttcttaga cccgttt 77

<210> 1365
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1365
atctacaaca gtagaaatta tttaaagttc ttagac 36

<210> 1366
<211> 14
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1366
agacttttaa cagt 14

<210> 1367
<211> 18
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1367
aatgacttct catctaca 18

<210> 1368
<211> 19
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1368
atttaaagtt cttagaccc 19

<210> 1369
<211> 14
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1369
agacttttaa cagt 14

<210> 1370
<211> 24
<212> DNA
<213> 人工序列

<220>

<223> 人工序列的描述：合成的
寡核苷酸

<400> 1370
aatgacttct catctacaac agta 24

<210> 1371
<211> 17
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1371
agttcttaga cccgttt 17

<210> 1372
<211> 26
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1372
gacttttaac agtggcctta ttaaat 26

<210> 1373
<211> 40
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1373
ctacaacagt agaaattatt taaagttctt agacccgttt 40

<210> 1374
<211> 19
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1374
agacttttaa cagtggcct 19

<210> 1375
<211> 18
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1375
cttctcatct acaacagt 18

<210> 1376
<211> 17
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1376
agttcttaga cccgttt 17

<210> 1377
<211> 16
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1377
cttattaaat gacttc 16

<210> 1378
<211> 32
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1378
gtagaaatta tttaaagttc ttagacccgt tt 32

<210> 1379
<211> 35
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1379
agacttttaa cagtggcctt attaaatgac ttctc 35

<210> 1380
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1380
tagaaattat ttaaagttct tagacccgtt t 31

<210> 1381
<211> 13
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1381
agacttttaa cag 13

<210> 1382
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1382
aaatgacttc tgatctacaa cagt 24

<210> 1383
<211> 16
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1383
gttcttagac ccgttt 16

<210> 1384
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1384
agacttttaa cagtggcctt a 21

<210> 1385
<211> 18
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1385
tctcatctac aacagtag 18

<210> 1386
<211> 16
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1386
gttcttagac ccgttt 16

<210> 1387
<211> 10
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1387
agacttttaa 10

<210> 1388
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1388
attaaatgac ttctcatcta caac 24

<210> 1389
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1389
ttaaagttct tagacccgtt t 21

<210> 1390
<211> 14
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1390
agacttttaa cagt 14

<210> 1391
<211> 25
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1391
aatgacttct catctacaac agtag 25

<210> 1392
<211> 14
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1392
tcttagaccc gttt 14

<210> 1393
<211> 40
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1393
agacttttaa cagtggcctt attaaatgac ttctcatcta 40

<210> 1394
<211> 27
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1394
aattatttaa agttcttaga cccgttt 27

<210> 1395
<211> 16
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1395
agacttttaa cagtgg 16

<210> 1396
<211> 18
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1396
tgacttctca tctacaac 18

<210> 1397
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1397
ttaaagttct tagacccgtt t 21

<210> 1398
<211> 13
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1398
agacttttaa cag 13

<210> 1399
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1399
aaatgacttc tcatctacaa cagt 24

<210> 1400
<211> 14
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1400
tcttagaccc gttt 14

<210> 1401
<211> 32
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1401
atctacaaca gtagaaatta tttaaagttc tt 32

<210> 1402
<211> 65
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1402

ggtacccggg gatcctttag agaagtcatt taataaggcc actgttaaaa agcttggcgt 60

aatca 65

<210> 1403
<211> 43
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1403
aauuucuacu guuguagaug agaagucauu uaauaaggcc acu 43

<210> 1404
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1404
ttaataaggc cactgttaaa a 21

<210> 1405
<211> 14
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
引物

<220>
<221> 修飾的鹼基
<222> (1)..(1)
<223> a、c、t、g、未知或未知

<400> 1405
ngccactgtt aaaa 14

<210> 1406
<211> 14
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
引物

<220>
<221> 修飾的鹼基
<222> (14)..(14)
<223> a、c、t、g、未知或未知

<400> 1406
ttaataaggc cacn 14

<210> 1407
<211> 43
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1407
aauuucuacu guuguagaug agaagucauu uaauaaggcc acu 43

<210> 1408
<211> 22

<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 1408
gagaagucau uuaauaaggc ca 22

<210> 1409
<211> 20
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 1409
gagaagucau uuaauaaggc 20

<210> 1410
<211> 18
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 1410
gagaagucau uuaauaag 18

<210> 1411
<211> 17
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 1411
gagaagucau uuaauaa 17

<210> 1412
<211> 16
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 1412
gagaagucau uuaaua 16

<210> 1413
<211> 15
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 1413
gagaagucau uuaau 15

<210> 1414
<211> 43
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 1414
aauuucuacu guuguagaug agaagucauu uaauaaggcc acu 43

<210> 1415
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>
<223> 組合的DNA/RNA分子的描述：合成的
寡核苷酸

<400> 1415
gataagucau uuaauaaggc cacu 24

<210> 1416
<211> 24
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1416
gagaaggcau uuaauaaggc cacu 24

<210> 1417
<211> 24
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1417
gagaagucau guaauaaggc cacu 24

<210> 1418
<211> 24
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1418
gagaagucau uuaagaaggc cacu 24

<210> 1419
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>
<223> 組合的DNA/RNA分子的描述：合成的
寡核苷酸

<400> 1419
gagaagucau uuaauaagtc cacu 24

<210> 1420
<211> 24
<212> RNA
<213> 人工序列

<220>

<223> 人工序列的描述：合成的
寡核苷酸

<400> 1420
gagaagucau uuaauaaggc caau 24

<210> 1421
<211> 43
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1421
aauuucuacu guuguagaug agaagucauu uaauaaggcc acu 43

<210> 1422
<211> 19
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1422
aauuucugcu guugcagau 19

<210> 1423
<211> 19
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1423
aauuuccacu guuguggau 19

<210> 1424
<211> 19
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1424
aauuccuacu guuguaggu 19

<210> 1425
<211> 19
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1425
aauuuauacu guuguagau 19

<210> 1426
<211> 19
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1426
aauuucgacu guuguagau 19

<210> 1427
<211> 19
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 1427
aauuucuagu guuguagau 19

<210> 1428
<211> 19
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 1428
aauuucaucu auuguagau 19

<210> 1429
<211> 19
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 1429
aauuucaucu gcuguagau 19

<210> 1430
<211> 18
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 1430
aauuucaucu uuguagau 18

<210> 1431
<211> 17
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 1431
aauuucaucu uguagau 17

<210> 1432
<211> 19
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的寡核苷酸

<400> 1432
aauuucuacu uuuguagaa 19

<210> 1433
<211> 19
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1433
aauuucuacu uuuguagac 19

<210> 1434
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1434
gucuaagaac uuuaaauaau uucuacuguu guagau 36

<210> 1435
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1435
guuuuggagu accuuagaaa ugcaugguuc ucaugc 36

<210> 1436
<211> 35
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1436
guuuuagaac cuuaaaaauu accuaguaau uaggu 35

<210> 1437
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1437
guuuaaaagu ccuauuggau uucuacuuuu guagau 36

<210> 1438
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1438
gccaaauacc ucuauaaaau uucuacuuuu guagau 36

<210> 1439
<211> 32
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1439
guuucaaucc acgcgcccac gcggggcgcg ac 32

<210> 1440
<211> 35
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1440
gucaaaagac cuuuuuaauu ucuacucuug uagau 35

<210> 1441
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1441
gcuuagaaca uuuaaagaau uucuacuauu guagau 36

<210> 1442
<211> 35
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1442
cucaaaacuc auucgaaucu cuacucuuug uagau 35

<210> 1443
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1443
guuugaauaa ccuuaaauaa uuucuacuuu guagau 36

<210> 1444
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1444
gucuaacgac cuuuuaaauu ucuacuguuu guagau 36

<210> 1445
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1445
cucuaaagag aggaaagaau uucuacuuuu guagau 36

<210> 1446
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1446
guuucaaaga uuaaauaauu ucuacuaagu guagau 36

<210> 1447
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1447
gucuagguac ucucuuuaau uucuacuauu guagau 36

<210> 1448
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1448
gucaauaaga cucauuuaau uucuacuucg guagau 36

<210> 1449
<211> 36
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1449
gccuauaagg cuuuaguaau uucuacuauu guagau 36

<210> 1450
<211> 20
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1450
uaauuucuac uguuguagau 20

<210> 1451
<211> 20
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1451
gaaaugcaug guucucaugc 20

<210> 1452
<211> 20
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1452

aaauuaccua guaauuaggu 20

<210> 1453
<211> 32
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1453
guuucaaucc acgcgcccac gcggggcgcg ac 32

<210> 1454
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1454
aatgtttcct gatggtccat gtctgttact cgcctg 36

<210> 1455
<211> 23
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1455
cugauggucc augucuguua cuc 23

<210> 1456
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1456
aatgtttcct gactcgcctg 20

<210> 1457
<211> 12
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1457
aatgtttcct ga 12

<210> 1458
<211> 32
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1458
aatgtttcct gatggtccat gtctgtcgcc tg 32

<210> 1459
<211> 32
<212> DNA

<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1459
aatgtttcct gatggtccat gtctgtcgcc tg 32

<210> 1460
<211> 30
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1460
aatgtttcct gatggtccat gttgtgcctg 30

<210> 1461
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1461
aatgtttcct gatggtccat gtactcgcct g 31

<210> 1462
<211> 32
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1462
aatgtttcct gatggtccat gtctgtcgcc tg 32

<210> 1463
<211> 32
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1463
aatgtttcct gatggtccat gttactcgcc tg 32

<210> 1464
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1464
aatgtttcct gatggtccat gtctgttcgc ctg 33

<210> 1465
<211> 34
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1465
aatgtttcct gatggtccat gtcttactcg cctg 34

<210> 1466
<211> 27
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1466
tttccctcac tcctgctcgg tgaattt 27

<210> 1467
<211> 27
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1467
tttgaggagt gttcagtctc cgtgaac 27

<210> 1468
<211> 27
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1468
tttcctgatg gtccatgtct gttactc 27

<210> 1469
<211> 27
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1469
tttatttccc ttcagctaaa ataaagg 27

<210> 1470
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1470
tcactcctgc tcggtgaatt tgg 23

<210> 1471
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1471
aaccctctgg ggaccgtttg agg 23

<210> 1472
<211> 23

<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1472
agtacgttaa tgtttcctga tgg 23

<210> 1473
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1473
tttcccttca gctaaaataa agg 23

<210> 1474
<211> 35
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1474
tttggagaag tcatttaata aggccactgt taaaa 35

<210> 1475
<211> 35
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1475
tttagagaag tcatttaata aggccactgt taaaa 35

<210> 1476
<211> 14
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>
<221> 修飾的鹼基
<222> (1)..(1)
<223> a、c、t、g、未知或未知

<400> 1476
ngccactgtt aaaa 14

<210> 1477
<211> 14
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>
<221> 修飾的鹼基
<222> (1)..(1)
<223> a、c、t、g、未知或未知

<400> 1477
ngccactgtt aaaa 14

<210> 1478
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>
<221> 修飾的鹼基
<222> (28)..(28)
<223> a、c、t、g、未知或未知

<400> 1478
tttggagaag tcatttaata aggccacn 28

<210> 1479
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>
<221> 修飾的鹼基
<222> (28)..(28)
<223> a、c、t、g、未知或未知

<400> 1479
tttagagaag tcatttaata aggccacn 28

<210> 1480
<211> 35
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1480
tttcgagaag tcatttaata aggccactgt taaaa 35

<210> 1481
<211> 14
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>
<221> 修飾的鹼基
<222> (1)..(1)
<223> a、c、t、g、未知或未知

<400> 1481
ngccactgtt aaaa 14

<210> 1482
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>
<221> 修飾的鹼基
<222> (28)..(28)
<223> a、c、t、g、未知或未知

<400> 1482
tttcgagaag tcatttaata aggccacn 28

<210> 1483
<211> 35
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1483
tttggctact attcctgtgc cttcagataa ttcaa 35

<210> 1484
<211> 35
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1484
ttttgctact attcctgtgc cttcagataa ttcaa 35

<210> 1485
<211> 15
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>
<221> 修飾的鹼基
<222> (1)..(1)
<223> a、c、t、g、未知或未知

<400> 1485
nttcagataa ttcaa 15

<210> 1486
<211> 15
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>
<221> 修飾的鹼基
<222> (1)..(1)
<223> a、c、t、g、未知或未知

<400> 1486
nttcagataa ttcaa 15

<210> 1487
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>

<221> 修飾的鹼基
<222> (28)..(28)
<223> a、c、t、g、未知或未知

<400> 1487
tttggctact attcctgtgc cttcagan 28

<210> 1488
<211> 27
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1488
ttttgctact attcctgtgc cttcaga 27

<210> 1489
<211> 35
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1489
tttcgtctag agccttttgt attagtagcc gagct 35

<210> 1490
<211> 35
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1490
tttggtctag agccttttgt attagtagcc gagct 35

<210> 1491
<211> 15
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>
<221> 修飾的鹼基
<222> (1)..(1)
<223> a、c、t、g、未知或未知

<400> 1491
nttagtagcc gagct 15

<210> 1492
<211> 15
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>
<221> 修飾的鹼基
<222> (1)..(1)
<223> a、c、t、g、未知或未知

<400> 1492
nttagtagcc gagct 15

<210> 1493
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>
<221> 修飾的鹼基
<222> (28)..(28)
<223> a、c、t、g、未知或未知

<400> 1493
tttcgtctag agccttttgt attagtan 28

<210> 1494
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<220>
<221> 修飾的鹼基
<222> (28)..(28)
<223> a、c、t、g、未知或未知

<400> 1494
tttggtctag agccttttgt attagtan 28

<210> 1495
<211> 724
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 1495
Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
20 25 30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
35 40 45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
50 55 60

Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65 70 75 80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
85 90 95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
100 105 110

Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
115 120 125

Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
130 135 140

Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145 150 155 160

```
Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
                165                 170                 175

Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
            180                 185                 190

Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
        195                 200                 205

Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
    210                 215                 220

Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225                 230                 235                 240

Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
                245                 250                 255

Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
            260                 265                 270

Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
        275                 280                 285

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
    290                 295                 300

Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305                 310                 315                 320

Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
                325                 330                 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
            340                 345                 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
        355                 360                 365

Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
    370                 375                 380

Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385                 390                 395                 400

Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
                405                 410                 415

Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
            420                 425                 430

Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
        435                 440                 445

Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
    450                 455                 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465                 470                 475                 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
                485                 490                 495
```

```
Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
            500                 505                 510

Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
        515                 520                 525

Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
    530                 535                 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545                 550                 555                 560

Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
                565                 570                 575

Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
            580                 585                 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
        595                 600                 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
    610                 615                 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625                 630                 635                 640

Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
                645                 650                 655

Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
            660                 665                 670

Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
        675                 680                 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln
    690                 695                 700

Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
705                 710                 715                 720

Ile Asp Phe Tyr
```

<210> 1496
<211> 680
<212> PRT
<213> 胺基酸球菌屬某種

<400> 1496

```
Met Thr Gln Phe Glu Gly Phe Thr Asn Leu Tyr Gln Val Ser Lys Thr
1               5                   10                  15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Lys His Ile Gln
            20                  25                  30

Glu Gln Gly Phe Ile Glu Glu Asp Lys Ala Arg Asn Asp His Tyr Lys
        35                  40                  45

Glu Leu Lys Pro Ile Ile Asp Arg Ile Tyr Lys Thr Tyr Ala Asp Gln
    50                  55                  60

Cys Leu Gln Leu Val Gln Leu Asp Trp Glu Asn Leu Ser Ala Ala Ile
```

65 70 75 80

Asp Ser Tyr Arg Lys Glu Lys Thr Glu Glu Thr Arg Asn Ala Leu Ile
85 90 95

Glu Glu Gln Ala Thr Tyr Arg Asn Ala Ile His Asp Tyr Phe Ile Gly
100 105 110

Arg Thr Asp Asn Leu Thr Asp Ala Ile Asn Lys Arg His Ala Glu Ile
115 120 125

Tyr Lys Gly Leu Phe Lys Ala Glu Leu Phe Asn Gly Lys Val Leu Lys
130 135 140

Gln Leu Gly Thr Val Thr Thr Thr Glu His Glu Asn Ala Leu Leu Arg
145 150 155 160

Ser Phe Asp Lys Phe Thr Thr Tyr Phe Ser Gly Phe Tyr Glu Asn Arg
165 170 175

Lys Asn Val Phe Ser Ala Glu Asp Ile Ser Thr Ala Ile Pro His Arg
180 185 190

Ile Val Gln Asp Asn Phe Pro Lys Phe Lys Glu Asn Cys His Ile Phe
195 200 205

Thr Arg Leu Ile Thr Ala Val Pro Ser Leu Arg Glu His Phe Glu Asn
210 215 220

Val Lys Lys Ala Ile Gly Ile Phe Val Ser Thr Ser Ile Glu Glu Val
225 230 235 240

Phe Ser Phe Pro Phe Tyr Asn Gln Leu Leu Thr Gln Thr Gln Ile Asp
245 250 255

Leu Tyr Asn Gln Leu Leu Gly Gly Ile Ser Arg Glu Ala Gly Thr Glu
260 265 270

Lys Ile Lys Gly Leu Asn Glu Val Leu Asn Leu Ala Ile Gln Lys Asn
275 280 285

Asp Glu Thr Ala His Ile Ile Ala Ser Leu Pro His Arg Phe Ile Pro
290 295 300

Leu Phe Lys Gln Ile Leu Ser Asp Arg Asn Thr Leu Ser Phe Ile Leu
305 310 315 320

Glu Glu Phe Lys Ser Asp Glu Glu Val Ile Gln Ser Phe Cys Lys Tyr
325 330 335

Lys Thr Leu Leu Arg Asn Glu Asn Val Leu Glu Thr Ala Glu Ala Leu
340 345 350

Phe Asn Glu Leu Asn Ser Ile Asp Leu Thr His Ile Phe Ile Ser His
355 360 365

Lys Lys Leu Glu Thr Ile Ser Ser Ala Leu Cys Asp His Trp Asp Thr
370 375 380

Leu Arg Asn Ala Leu Tyr Glu Arg Arg Ile Ser Glu Leu Thr Gly Lys
385 390 395 400

Ile Thr Lys Ser Ala Lys Glu Lys Val Gln Arg Ser Leu Lys His Glu

```
                405                 410                 415

Asp Ile Asn Leu Gln Glu Ile Ile Ser Ala Ala Gly Lys Glu Leu Ser
            420                 425                 430

Glu Ala Phe Lys Gln Lys Thr Ser Glu Ile Leu Ser His Ala His Ala
        435                 440                 445

Ala Leu Asp Gln Pro Leu Pro Thr Thr Leu Lys Lys Gln Glu Glu Lys
    450                 455                 460

Glu Ile Leu Lys Ser Gln Leu Asp Ser Leu Leu Gly Leu Tyr His Leu
465                 470                 475                 480

Leu Asp Trp Phe Ala Val Asp Glu Ser Asn Glu Val Asp Pro Glu Phe
                485                 490                 495

Ser Ala Arg Leu Thr Gly Ile Lys Leu Glu Met Glu Pro Ser Leu Ser
            500                 505                 510

Phe Tyr Asn Lys Ala Arg Asn Tyr Ala Thr Lys Lys Pro Tyr Ser Val
        515                 520                 525

Glu Lys Phe Lys Leu Asn Phe Gln Met Pro Thr Leu Ala Ser Gly Trp
    530                 535                 540

Asp Val Asn Lys Glu Lys Asn Asn Gly Ala Ile Leu Phe Val Lys Asn
545                 550                 555                 560

Gly Leu Tyr Tyr Leu Gly Ile Met Pro Lys Gln Lys Gly Arg Tyr Lys
                565                 570                 575

Ala Leu Ser Phe Glu Pro Thr Glu Lys Thr Ser Glu Gly Phe Asp Lys
            580                 585                 590

Met Tyr Tyr Asp Tyr Phe Pro Asp Ala Ala Lys Met Ile Pro Lys Cys
        595                 600                 605

Ser Thr Gln Leu Lys Ala Val Thr Ala His Phe Gln Thr His Thr Thr
    610                 615                 620

Pro Ile Leu Leu Ser Asn Asn Phe Ile Glu Pro Leu Glu Ile Thr Lys
625                 630                 635                 640

Glu Ile Tyr Asp Leu Asn Asn Pro Glu Lys Glu Pro Lys Lys Phe Gln
                645                 650                 655

Thr Ala Tyr Ala Lys Lys Thr Gly Asp Gln Lys Gly Tyr Arg Glu Ala
            660                 665                 670

Leu Cys Lys Trp Ile Asp Phe Thr
        675                 680
```

<210> 1497
<211> 639
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 1497

```
Met Ser Lys Leu Glu Lys Phe Thr Asn Cys Tyr Ser Leu Ser Lys Thr
1               5                   10                  15
```

Leu Arg Phe Lys Ala Ile Pro Val Gly Lys Thr Gln Glu Asn Ile Asp
20 25 30

Asn Lys Arg Leu Leu Val Glu Asp Glu Lys Arg Ala Glu Asp Tyr Lys
35 40 45

Gly Val Lys Lys Leu Leu Asp Arg Tyr Tyr Leu Ser Phe Ile Asn Asp
50 55 60

Val Leu His Ser Ile Lys Leu Lys Asn Leu Asn Asn Tyr Ile Ser Leu
65 70 75 80

Phe Arg Lys Lys Thr Arg Thr Glu Lys Glu Asn Lys Glu Leu Glu Asn
85 90 95

Leu Glu Ile Asn Leu Arg Lys Glu Ile Ala Lys Ala Phe Lys Gly Asn
100 105 110

Glu Gly Tyr Lys Ser Leu Phe Lys Lys Asp Ile Ile Glu Thr Ile Leu
115 120 125

Pro Glu Phe Leu Asp Asp Lys Asp Glu Ile Ala Leu Val Asn Ser Phe
130 135 140

Asn Gly Phe Thr Thr Ala Phe Thr Gly Phe Phe Asp Asn Arg Glu Asn
145 150 155 160

Met Phe Ser Glu Glu Ala Lys Ser Thr Ser Ile Ala Phe Arg Cys Ile
165 170 175

Asn Glu Asn Leu Thr Arg Tyr Ile Ser Asn Met Asp Ile Phe Glu Lys
180 185 190

Val Asp Ala Ile Phe Asp Lys His Glu Val Gln Glu Ile Lys Glu Lys
195 200 205

Ile Leu Asn Ser Asp Tyr Asp Val Glu Asp Phe Phe Glu Gly Glu Phe
210 215 220

Phe Asn Phe Val Leu Thr Gln Glu Gly Ile Asp Val Tyr Asn Ala Ile
225 230 235 240

Ile Gly Gly Phe Val Thr Glu Ser Gly Glu Lys Ile Lys Gly Leu Asn
245 250 255

Glu Tyr Ile Asn Leu Tyr Asn Gln Lys Thr Lys Gln Lys Leu Pro Lys
260 265 270

Phe Lys Pro Leu Tyr Lys Gln Val Leu Ser Asp Arg Glu Ser Leu Ser
275 280 285

Phe Tyr Gly Glu Gly Tyr Thr Ser Asp Glu Glu Val Leu Glu Val Phe
290 295 300

Arg Asn Thr Leu Asn Lys Asn Ser Glu Ile Phe Ser Ser Ile Lys Lys
305 310 315 320

Leu Glu Lys Leu Phe Lys Asn Phe Asp Glu Tyr Ser Ser Ala Gly Ile
325 330 335

Phe Val Lys Asn Gly Pro Ala Ile Ser Thr Ile Ser Lys Asp Ile Phe
340 345 350

```
Gly Glu Trp Asn Val Ile Arg Asp Lys Trp Asn Ala Glu Tyr Asp Asp
        355                 360                 365

Ile His Leu Lys Lys Lys Ala Val Val Thr Glu Lys Tyr Glu Asp Asp
    370                 375                 380

Arg Arg Lys Ser Phe Lys Lys Ile Gly Ser Phe Ser Leu Glu Gln Leu
385                 390                 395                 400

Gln Glu Tyr Ala Asp Ala Asp Leu Ser Val Val Glu Lys Leu Lys Glu
                405                 410                 415

Ile Ile Ile Gln Lys Val Asp Glu Ile Tyr Lys Val Tyr Gly Ser Ser
            420                 425                 430

Glu Lys Leu Phe Asp Ala Asp Phe Val Leu Glu Lys Ser Leu Lys Lys
        435                 440                 445

Asn Asp Ala Val Val Ala Ile Met Lys Asp Leu Leu Asp Ser Val Lys
    450                 455                 460

Ser Phe Glu Asn Tyr Ile Lys Ala Phe Phe Gly Glu Gly Lys Glu Thr
465                 470                 475                 480

Asn Arg Asp Glu Ser Phe Tyr Gly Asp Phe Val Leu Ala Tyr Asp Ile
                485                 490                 495

Leu Leu Lys Val Asp His Ile Tyr Asp Ala Ile Arg Asn Tyr Val Thr
            500                 505                 510

Gln Lys Pro Tyr Ser Lys Asp Lys Phe Lys Leu Tyr Phe Gln Asn Pro
        515                 520                 525

Gln Phe Met Gly Gly Trp Asp Lys Asp Lys Glu Thr Asp Tyr Arg Ala
    530                 535                 540

Thr Ile Leu Arg Tyr Gly Ser Lys Tyr Tyr Leu Ala Ile Met Asp Lys
545                 550                 555                 560

Lys Tyr Ala Lys Cys Leu Gln Lys Ile Asp Lys Asp Asp Val Asn Gly
                565                 570                 575

Asn Tyr Glu Lys Ile Asn Tyr Lys Leu Leu Pro Gly Pro Asn Lys Met
            580                 585                 590

Leu Pro Lys Val Phe Phe Ser Lys Lys Trp Met Ala Tyr Tyr Asn Pro
        595                 600                 605

Ser Glu Asp Ile Gln Lys Ile Tyr Lys Asn Gly Thr Phe Lys Lys Gly
    610                 615                 620

Asp Met Phe Asn Leu Asn Asp Cys His Lys Leu Ile Asp Phe Phe
625                 630                 635
```

<210> 1498
<211> 44
<212> RNA
<213> 土拉熱弗朗西絲菌

<220>
<221> 修飾的鹼基
<222> (21)..(44)
<223> a、c、u、g、未知或其他

<400> 1498
uaauuucuac uguuguagau nnnnnnnnnn nnnnnnnnnn nnnn 44

<210> 1499
<211> 44
<212> RNA
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<220>
<221> 修飾的鹼基
<222> (21)..(44)
<223> a、c、u、g、未知或其他

<400> 1499
gaaaugcaug guucucaugc nnnnnnnnnn nnnnnnnnnn nnnn 44

<210> 1500
<211> 44
<212> RNA
<213> 解朊丁酸弧菌

<220>
<221> 修飾的鹼基
<222> (21)..(44)
<223> a、c、u、g、未知或其他

<400> 1500
aaauuaccua guaauuaggu nnnnnnnnnn nnnnnnnnnn nnnn 44

<210> 1501
<211> 44
<212> RNA
<213> 未知

<220>
<223> 未知的描述：
候選佩萊格裡尼細菌序列

<220>
<221> 修飾的鹼基
<222> (21)..(44)
<223> a、c、u、g、未知或其他

<400> 1501
ggauuucuac uuuuguagau nnnnnnnnnn nnnnnnnnnn nnnn 44

<210> 1502
<211> 44
<212> RNA
<213> 未知

<220>
<223> 未知的描述：
候選佩萊格裡尼細菌序列

<220>
<221> 修飾的鹼基
<222> (21)..(44)
<223> a、c、u、g、未知或其他

<400> 1502
aaauuucuac uuuuguagau nnnnnnnnnn nnnnnnnnnn nnnn 44

<210> 1503
<211> 44
<212> RNA
<213> 密斯氏菌屬某種

<220>
<221> 修飾的鹼基

<222> (21)..(44)
<223> a、c、u、g、未知或其他

<400> 1503
gcgcccacgc ggggcgcgac nnnnnnnnnn nnnnnnnnnn nnnn 44

<210> 1504
<211> 44
<212> RNA
<213> 胺基酸球菌屬某種

<220>
<221> 修飾的鹼基
<222> (21)..(44)
<223> a、c、u、g、未知或其他

<400> 1504
uaauuucuac ucuuguagau nnnnnnnnnn nnnnnnnnnn nnnn 44

<210> 1505
<211> 44
<212> RNA
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<220>
<221> 修飾的鹼基
<222> (21)..(44)
<223> a、c、u、g、未知或其他

<400> 1505
gaauuucuac uauuguagau nnnnnnnnnn nnnnnnnnnn nnnn 44

<210> 1506
<211> 45
<212> RNA
<213> 未知

<220>
<223> 未知的描述：
候選白蟻甲烷枝原體序列

<220>
<221> 修飾的鹼基
<222> (22)..(45)
<223> a、c、u、g、未知或其他

<400> 1506
gaaucucuac ucuuuguaga unnnnnnnnn nnnnnnnnnn nnnnn 45

<210> 1507
<211> 43
<212> RNA
<213> 挑剔真細菌

<220>
<221> 修飾的鹼基
<222> (20)..(43)
<223> a、c、u、g、未知或其他

<400> 1507
uaauuucuac uuuguagaun nnnnnnnnnn nnnnnnnnnn nnn 43

<210> 1508
<211> 44
<212> RNA
<213> 牛莫拉氏菌

<220>
<221> 修飾的鹼基
<222> (21)..(44)

<223> a、c、u、g、未知或其他

<400> 1508
aauuucuacu guuuguagau nnnnnnnnnn nnnnnnnnnn nnnn 44

<210> 1509
<211> 44
<212> RNA
<213> 稻田鉤端螺旋體

<220>
<221> 修飾的鹼基
<222> (21)..(44)
<223> a、c、u、g、未知或其他

<400> 1509
gaauuucuac uuuuguagau nnnnnnnnnn nnnnnnnnnn nnnn 44

<210> 1510
<211> 44
<212> RNA
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<220>
<221> 修飾的鹼基
<222> (21)..(44)
<223> a、c、u、g、未知或其他

<400> 1510
aauuucuacu aaguguagau nnnnnnnnnn nnnnnnnnnn nnnn 44

<210> 1511
<211> 44
<212> RNA
<213> 狗口腔卟啉單胞菌

<220>
<221> 修飾的鹼基
<222> (21)..(44)
<223> a、c、u、g、未知或其他

<400> 1511
uaauuucuac uauuguagau nnnnnnnnnn nnnnnnnnnn nnnn 44

<210> 1512
<211> 44
<212> RNA
<213> 解糖胨普雷沃菌

<220>
<221> 修飾的鹼基
<222> (21)..(44)
<223> a、c、u、g、未知或其他

<400> 1512
uaauuucuac uucgguagau nnnnnnnnnn nnnnnnnnnn nnnn 44

<210> 1513
<211> 44
<212> RNA
<213> 獼猴卟啉單胞菌

<220>
<221> 修飾的鹼基
<222> (21)..(44)
<223> a、c、u、g、未知或其他

<400> 1513
uaauuucuac uauuguagau nnnnnnnnnn nnnnnnnnnn nnnn 44

<210> 1514
<211> 35
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1514
ttttgagaag tcatttaata aggccactgt taaaa 35

<210> 1515
<211> 13
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
引物

<220>
<221> 修飾的鹼基
<222> (1)..(1)
<223> a、c、t、g、未知或未知

<400> 1515
nccactgtta aaa 13

<210> 1516
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
引物

<220>
<221> 修飾的鹼基
<222> (28)..(28)
<223> a、c、t、g、未知或未知

<400> 1516
ttttgagaag tcatttaata aggccacn 28

<210> 1517
<211> 35
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1517
ttttgagaag tcatttaata aggccactgt taaaa 35

<210> 1518
<211> 13
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
引物

<220>
<221> 修飾的鹼基
<222> (1)..(1)
<223> a、c、t、g、未知或未知

<400> 1518
nccactgtta aaa 13

<210> 1519
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
引物

<220>
<221> 修飾的鹼基
<222> (28)..(28)
<223> a、c、t、g、未知或未知

<400> 1519
ttttgagaag tcatttaata aggccacn 28

<210> 1520
<211> 1307
<212> PRT
<213> 胺基酸球菌屬某種

<400> 1520
Met Thr Gln Phe Glu Gly Phe Thr Asn Leu Tyr Gln Val Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Lys His Ile Gln
20 25 30

Glu Gln Gly Phe Ile Glu Glu Asp Lys Ala Arg Asn Asp His Tyr Lys
35 40 45

Glu Leu Lys Pro Ile Ile Asp Arg Ile Tyr Lys Thr Tyr Ala Asp Gln
50 55 60

Cys Leu Gln Leu Val Gln Leu Asp Trp Glu Asn Leu Ser Ala Ala Ile
65 70 75 80

Asp Ser Tyr Arg Lys Glu Lys Thr Glu Glu Thr Arg Asn Ala Leu Ile
85 90 95

Glu Glu Gln Ala Thr Tyr Arg Asn Ala Ile His Asp Tyr Phe Ile Gly
100 105 110

Arg Thr Asp Asn Leu Thr Asp Ala Ile Asn Lys Arg His Ala Glu Ile
115 120 125

Tyr Lys Gly Leu Phe Lys Ala Glu Leu Phe Asn Gly Lys Val Leu Lys
130 135 140

Gln Leu Gly Thr Val Thr Thr Thr Glu His Glu Asn Ala Leu Leu Arg
145 150 155 160

Ser Phe Asp Lys Phe Thr Thr Tyr Phe Ser Gly Phe Tyr Glu Asn Arg
165 170 175

Lys Asn Val Phe Ser Ala Glu Asp Ile Ser Thr Ala Ile Pro His Arg
180 185 190

Ile Val Gln Asp Asn Phe Pro Lys Phe Lys Glu Asn Cys His Ile Phe
195 200 205

Thr Arg Leu Ile Thr Ala Val Pro Ser Leu Arg Glu His Phe Glu Asn
210 215 220

Val Lys Lys Ala Ile Gly Ile Phe Val Ser Thr Ser Ile Glu Glu Val
225 230 235 240

```
Phe Ser Phe Pro Phe Tyr Asn Gln Leu Leu Thr Gln Thr Gln Ile Asp
                245                 250                 255

Leu Tyr Asn Gln Leu Leu Gly Gly Ile Ser Arg Glu Ala Gly Thr Glu
            260                 265                 270

Lys Ile Lys Gly Leu Asn Glu Val Leu Asn Leu Ala Ile Gln Lys Asn
        275                 280                 285

Asp Glu Thr Ala His Ile Ile Ala Ser Leu Pro His Arg Phe Ile Pro
    290                 295                 300

Leu Phe Lys Gln Ile Leu Ser Asp Arg Asn Thr Leu Ser Phe Ile Leu
305                 310                 315                 320

Glu Glu Phe Lys Ser Asp Glu Glu Val Ile Gln Ser Phe Cys Lys Tyr
                325                 330                 335

Lys Thr Leu Leu Arg Asn Glu Asn Val Leu Glu Thr Ala Glu Ala Leu
            340                 345                 350

Phe Asn Glu Leu Asn Ser Ile Asp Leu Thr His Ile Phe Ile Ser His
        355                 360                 365

Lys Lys Leu Glu Thr Ile Ser Ser Ala Leu Cys Asp His Trp Asp Thr
    370                 375                 380

Leu Arg Asn Ala Leu Tyr Glu Arg Arg Ile Ser Glu Leu Thr Gly Lys
385                 390                 395                 400

Ile Thr Lys Ser Ala Lys Glu Lys Val Gln Arg Ser Leu Lys His Glu
                405                 410                 415

Asp Ile Asn Leu Gln Glu Ile Ile Ser Ala Ala Gly Lys Glu Leu Ser
            420                 425                 430

Glu Ala Phe Lys Gln Lys Thr Ser Glu Ile Leu Ser His Ala His Ala
        435                 440                 445

Ala Leu Asp Gln Pro Leu Pro Thr Thr Leu Lys Lys Gln Glu Glu Lys
    450                 455                 460

Glu Ile Leu Lys Ser Gln Leu Asp Ser Leu Leu Gly Leu Tyr His Leu
465                 470                 475                 480

Leu Asp Trp Phe Ala Val Asp Glu Ser Asn Glu Val Asp Pro Glu Phe
                485                 490                 495

Ser Ala Arg Leu Thr Gly Ile Lys Leu Glu Met Glu Pro Ser Leu Ser
            500                 505                 510

Phe Tyr Asn Lys Ala Arg Asn Tyr Ala Thr Lys Lys Pro Tyr Ser Val
        515                 520                 525

Glu Lys Phe Lys Leu Asn Phe Gln Met Pro Thr Leu Ala Ser Gly Trp
    530                 535                 540

Asp Val Asn Lys Glu Lys Asn Asn Gly Ala Ile Leu Phe Val Lys Asn
545                 550                 555                 560

Gly Leu Tyr Tyr Leu Gly Ile Met Pro Lys Gln Lys Gly Arg Tyr Lys
                565                 570                 575
```

Ala Leu Ser Phe Glu Pro Thr Glu Lys Thr Ser Glu Gly Phe Asp Lys
580 585 590

Met Tyr Tyr Asp Tyr Phe Pro Asp Ala Ala Lys Met Ile Pro Lys Cys
595 600 605

Ser Thr Gln Leu Lys Ala Val Thr Ala His Phe Gln Thr His Thr Thr
610 615 620

Pro Ile Leu Leu Ser Asn Asn Phe Ile Glu Pro Leu Glu Ile Thr Lys
625 630 635 640

Glu Ile Tyr Asp Leu Asn Asn Pro Glu Lys Glu Pro Lys Lys Phe Gln
645 650 655

Thr Ala Tyr Ala Lys Lys Thr Gly Asp Gln Lys Gly Tyr Arg Glu Ala
660 665 670

Leu Cys Lys Trp Ile Asp Phe Thr Arg Asp Phe Leu Ser Lys Tyr Thr
675 680 685

Lys Thr Thr Ser Ile Asp Leu Ser Ser Leu Arg Pro Ser Ser Gln Tyr
690 695 700

Lys Asp Leu Gly Glu Tyr Tyr Ala Glu Leu Asn Pro Leu Leu Tyr His
705 710 715 720

Ile Ser Phe Gln Arg Ile Ala Glu Lys Glu Ile Met Asp Ala Val Glu
725 730 735

Thr Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ala Lys
740 745 750

Gly His His Gly Lys Pro Asn Leu His Thr Leu Tyr Trp Thr Gly Leu
755 760 765

Phe Ser Pro Glu Asn Leu Ala Lys Thr Ser Ile Lys Leu Asn Gly Gln
770 775 780

Ala Glu Leu Phe Tyr Arg Pro Lys Ser Arg Met Lys Arg Met Ala His
785 790 795 800

Arg Leu Gly Glu Lys Met Leu Asn Lys Lys Leu Lys Asp Gln Lys Thr
805 810 815

Pro Ile Pro Asp Thr Leu Tyr Gln Glu Leu Tyr Asp Tyr Val Asn His
820 825 830

Arg Leu Ser His Asp Leu Ser Asp Glu Ala Arg Ala Leu Leu Pro Asn
835 840 845

Val Ile Thr Lys Glu Val Ser His Glu Ile Ile Lys Asp Arg Arg Phe
850 855 860

Thr Ser Asp Lys Phe Phe Phe His Val Pro Ile Thr Leu Asn Tyr Gln
865 870 875 880

Ala Ala Asn Ser Pro Ser Lys Phe Asn Gln Arg Val Asn Ala Tyr Leu
885 890 895

Lys Glu His Pro Glu Thr Pro Ile Ile Gly Ile Asp Arg Gly Glu Arg
900 905 910

```
Asn Leu Ile Tyr Ile Thr Val Ile Asp Ser Thr Gly Lys Ile Leu Glu
        915                 920                 925

Gln Arg Ser Leu Asn Thr Ile Gln Gln Phe Asp Tyr Gln Lys Lys Leu
    930                 935                 940

Asp Asn Arg Glu Lys Glu Arg Val Ala Ala Arg Gln Ala Trp Ser Val
945                 950                 955                 960

Val Gly Thr Ile Lys Asp Leu Lys Gln Gly Tyr Leu Ser Gln Val Ile
                965                 970                 975

His Glu Ile Val Asp Leu Met Ile His Tyr Gln Ala Val Val Val Leu
            980                 985                 990

Glu Asn Leu Asn Phe Gly Phe Lys Ser Lys Arg Thr Gly Ile Ala Glu
        995                 1000                1005

Lys Ala  Val Tyr Gln Gln Phe  Glu Lys Met Leu Ile  Asp Lys Leu
    1010                1015                1020

Asn Cys  Leu Val Leu Lys Asp  Tyr Pro Ala Glu Lys  Val Gly Gly
    1025                1030                1035

Val Leu  Asn Pro Tyr Gln Leu  Thr Asp Gln Phe Thr  Ser Phe Ala
    1040                1045                1050

Lys Met  Gly Thr Gln Ser Gly  Phe Leu Phe Tyr Val  Pro Ala Pro
    1055                1060                1065

Tyr Thr  Ser Lys Ile Asp Pro  Leu Thr Gly Phe Val  Asp Pro Phe
    1070                1075                1080

Val Trp  Lys Thr Ile Lys Asn  His Glu Ser Arg Lys  His Phe Leu
    1085                1090                1095

Glu Gly  Phe Asp Phe Leu His  Tyr Asp Val Lys Thr  Gly Asp Phe
    1100                1105                1110

Ile Leu  His Phe Lys Met Asn  Arg Asn Leu Ser Phe  Gln Arg Gly
    1115                1120                1125

Leu Pro  Gly Phe Met Pro Ala  Trp Asp Ile Val Phe  Glu Lys Asn
    1130                1135                1140

Glu Thr  Gln Phe Asp Ala Lys  Gly Thr Pro Phe Ile  Ala Gly Lys
    1145                1150                1155

Arg Ile  Val Pro Val Ile Glu  Asn His Arg Phe Thr  Gly Arg Tyr
    1160                1165                1170

Arg Asp  Leu Tyr Pro Ala Asn  Glu Leu Ile Ala Leu  Leu Glu Glu
    1175                1180                1185

Lys Gly  Ile Val Phe Arg Asp  Gly Ser Asn Ile Leu  Pro Lys Leu
    1190                1195                1200

Leu Glu  Asn Asp Asp Ser His  Ala Ile Asp Thr Met  Val Ala Leu
    1205                1210                1215

Ile Arg  Ser Val Leu Gln Met  Arg Asn Ser Asn Ala  Ala Thr Gly
    1220                1225                1230
```

Glu Asp Tyr Ile Asn Ser Pro Val Arg Asp Leu Asn Gly Val Cys
1235 1240 1245

Phe Asp Ser Arg Phe Gln Asn Pro Glu Trp Pro Met Asp Ala Asp
1250 1255 1260

Ala Asn Gly Ala Tyr His Ile Ala Leu Lys Gly Gln Leu Leu Leu
1265 1270 1275

Asn His Leu Lys Glu Ser Lys Asp Leu Lys Leu Gln Asn Gly Ile
1280 1285 1290

Ser Asn Gln Asp Trp Leu Ala Tyr Ile Gln Glu Leu Arg Asn
1295 1300 1305

<210> 1521
<211> 365
<212> PRT
<213> 海棲熱袍菌（Thermotoga maritima）

<400> 1521
Met Arg Pro Glu Arg Leu Thr Val Arg Asn Phe Leu Gly Leu Lys Asn
1 5 10 15

Val Asp Ile Glu Phe Gln Ser Gly Ile Thr Val Val Glu Gly Pro Asn
20 25 30

Gly Ala Gly Lys Ser Ser Leu Phe Glu Ala Ile Ser Phe Ala Leu Phe
35 40 45

Gly Asn Gly Ile Arg Tyr Pro Asn Ser Tyr Asp Tyr Val Asn Arg Asn
50 55 60

Ala Val Asp Gly Thr Ala Arg Leu Val Phe Gln Phe Glu Arg Gly Gly
65 70 75 80

Lys Arg Tyr Glu Ile Ile Arg Glu Ile Asn Ala Leu Gln Arg Lys His
85 90 95

Asn Ala Lys Leu Ser Glu Ile Leu Glu Asn Gly Lys Lys Ala Ala Ile
100 105 110

Ala Ala Lys Pro Thr Ser Val Lys Gln Glu Val Glu Lys Ile Leu Gly
115 120 125

Ile Glu His Arg Thr Phe Ile Arg Thr Val Phe Leu Pro Gln Gly Glu
130 135 140

Ile Asp Lys Leu Leu Ile Ser Pro Pro Ser Glu Ile Thr Glu Ile Ile
145 150 155 160

Ser Asp Val Phe Gln Ser Lys Glu Thr Leu Glu Lys Leu Glu Lys Leu
165 170 175

Leu Lys Glu Lys Met Lys Lys Leu Glu Asn Glu Ile Ser Ser Gly Gly
180 185 190

Ala Gly Gly Ala Gly Gly Ser Leu Glu Lys Lys Leu Lys Glu Met Ser
195 200 205

Asp Glu Tyr Asn Asn Leu Asp Leu Leu Arg Lys Tyr Leu Phe Asp Lys
210 215 220

Ser Asn Phe Ser Arg Tyr Phe Thr Gly Arg Val Leu Glu Ala Val Leu

225 230 235 240

Lys Arg Thr Lys Ala Tyr Leu Asp Ile Leu Thr Asn Gly Arg Phe Asp
245 250 255

Ile Asp Phe Asp Asp Glu Lys Gly Gly Phe Ile Ile Lys Asp Trp Gly
260 265 270

Ile Glu Arg Pro Ala Arg Gly Leu Ser Gly Gly Glu Arg Ala Leu Ile
275 280 285

Ser Ile Ser Leu Ala Met Ser Leu Ala Glu Val Ala Ser Gly Arg Leu
290 295 300

Asp Ala Phe Phe Ile Asp Glu Gly Phe Ser Ser Leu Asp Thr Glu Asn
305 310 315 320

Lys Glu Lys Ile Ala Ser Val Leu Lys Glu Leu Glu Arg Leu Asn Lys
325 330 335

Val Ile Val Phe Ile Thr His Asp Arg Glu Phe Ser Glu Ala Phe Asp
340 345 350

Arg Lys Leu Arg Ile Thr Gly Gly Val Val Val Asn Glu
355 360 365

<210> 1522
<211> 1307
<212> PRT
<213> 胺基酸球菌屬某種

<400> 1522
Met Thr Gln Phe Glu Gly Phe Thr Asn Leu Tyr Gln Val Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Lys His Ile Gln
20 25 30

Glu Gln Gly Phe Ile Glu Glu Asp Lys Ala Arg Asn Asp His Tyr Lys
35 40 45

Glu Leu Lys Pro Ile Ile Asp Arg Ile Tyr Lys Thr Tyr Ala Asp Gln
50 55 60

Cys Leu Gln Leu Val Gln Leu Asp Trp Glu Asn Leu Ser Ala Ala Ile
65 70 75 80

Asp Ser Tyr Arg Lys Glu Lys Thr Glu Glu Thr Arg Asn Ala Leu Ile
85 90 95

Glu Glu Gln Ala Thr Tyr Arg Asn Ala Ile His Asp Tyr Phe Ile Gly
100 105 110

Arg Thr Asp Asn Leu Thr Asp Ala Ile Asn Lys Arg His Ala Glu Ile
115 120 125

Tyr Lys Gly Leu Phe Lys Ala Glu Leu Phe Asn Gly Lys Val Leu Lys
130 135 140

Gln Leu Gly Thr Val Thr Thr Thr Glu His Glu Asn Ala Leu Leu Arg
145 150 155 160

Ser Phe Asp Lys Phe Thr Thr Tyr Phe Ser Gly Phe Tyr Glu Asn Arg
165 170 175

```
Lys Asn Val Phe Ser Ala Glu Asp Ile Ser Thr Ala Ile Pro His Arg
            180                 185                 190

Ile Val Gln Asp Asn Phe Pro Lys Phe Lys Glu Asn Cys His Ile Phe
        195                 200                 205

Thr Arg Leu Ile Thr Ala Val Pro Ser Leu Arg Glu His Phe Glu Asn
    210                 215                 220

Val Lys Lys Ala Ile Gly Ile Phe Val Ser Thr Ser Ile Glu Glu Val
225                 230                 235                 240

Phe Ser Phe Pro Phe Tyr Asn Gln Leu Leu Thr Gln Thr Gln Ile Asp
                245                 250                 255

Leu Tyr Asn Gln Leu Leu Gly Gly Ile Ser Arg Glu Ala Gly Thr Glu
            260                 265                 270

Lys Ile Lys Gly Leu Asn Glu Val Leu Asn Leu Ala Ile Gln Lys Asn
        275                 280                 285

Asp Glu Thr Ala His Ile Ile Ala Ser Leu Pro His Arg Phe Ile Pro
    290                 295                 300

Leu Phe Lys Gln Ile Leu Ser Asp Arg Asn Thr Leu Ser Phe Ile Leu
305                 310                 315                 320

Glu Glu Phe Lys Ser Asp Glu Glu Val Ile Gln Ser Phe Cys Lys Tyr
                325                 330                 335

Lys Thr Leu Leu Arg Asn Glu Asn Val Leu Glu Thr Ala Glu Ala Leu
            340                 345                 350

Phe Asn Glu Leu Asn Ser Ile Asp Leu Thr His Ile Phe Ile Ser His
        355                 360                 365

Lys Lys Leu Glu Thr Ile Ser Ser Ala Leu Cys Asp His Trp Asp Thr
    370                 375                 380

Leu Arg Asn Ala Leu Tyr Glu Arg Arg Ile Ser Glu Leu Thr Gly Lys
385                 390                 395                 400

Ile Thr Lys Ser Ala Lys Glu Lys Val Gln Arg Ser Leu Lys His Glu
                405                 410                 415

Asp Ile Asn Leu Gln Glu Ile Ile Ser Ala Ala Gly Lys Glu Leu Ser
            420                 425                 430

Glu Ala Phe Lys Gln Lys Thr Ser Glu Ile Leu Ser His Ala His Ala
        435                 440                 445

Ala Leu Asp Gln Pro Leu Pro Thr Thr Leu Lys Lys Gln Glu Glu Lys
    450                 455                 460

Glu Ile Leu Lys Ser Gln Leu Asp Ser Leu Leu Gly Leu Tyr His Leu
465                 470                 475                 480

Leu Asp Trp Phe Ala Val Asp Glu Ser Asn Glu Val Asp Pro Glu Phe
                485                 490                 495

Ser Ala Arg Leu Thr Gly Ile Lys Leu Glu Met Glu Pro Ser Leu Ser
            500                 505                 510
```

```
Phe Tyr Asn Lys Ala Arg Asn Tyr Ala Thr Lys Lys Pro Tyr Ser Val
        515                 520                 525

Glu Lys Phe Lys Leu Asn Phe Gln Met Pro Thr Leu Ala Ser Gly Trp
    530                 535                 540

Asp Val Asn Lys Glu Lys Asn Asn Gly Ala Ile Leu Phe Val Lys Asn
545                 550                 555                 560

Gly Leu Tyr Tyr Leu Gly Ile Met Pro Lys Gln Lys Gly Arg Tyr Lys
                565                 570                 575

Ala Leu Ser Phe Glu Pro Thr Glu Lys Thr Ser Glu Gly Phe Asp Lys
            580                 585                 590

Met Tyr Tyr Asp Tyr Phe Pro Asp Ala Ala Lys Met Ile Pro Lys Cys
        595                 600                 605

Ser Thr Gln Leu Lys Ala Val Thr Ala His Phe Gln Thr His Thr Thr
    610                 615                 620

Pro Ile Leu Leu Ser Asn Asn Phe Ile Glu Pro Leu Glu Ile Thr Lys
625                 630                 635                 640

Glu Ile Tyr Asp Leu Asn Asn Pro Glu Lys Glu Pro Lys Lys Phe Gln
                645                 650                 655

Thr Ala Tyr Ala Lys Lys Thr Gly Asp Gln Lys Gly Tyr Arg Glu Ala
            660                 665                 670

Leu Cys Lys Trp Ile Asp Phe Thr Arg Asp Phe Leu Ser Lys Tyr Thr
        675                 680                 685

Lys Thr Thr Ser Ile Asp Leu Ser Ser Leu Arg Pro Ser Ser Gln Tyr
    690                 695                 700

Lys Asp Leu Gly Glu Tyr Tyr Ala Glu Leu Asn Pro Leu Leu Tyr His
705                 710                 715                 720

Ile Ser Phe Gln Arg Ile Ala Glu Lys Glu Ile Met Asp Ala Val Glu
                725                 730                 735

Thr Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ala Lys
            740                 745                 750

Gly His His Gly Lys Pro Asn Leu His Thr Leu Tyr Trp Thr Gly Leu
        755                 760                 765

Phe Ser Pro Glu Asn Leu Ala Lys Thr Ser Ile Lys Leu Asn Gly Gln
    770                 775                 780

Ala Glu Leu Phe Tyr Arg Pro Lys Ser Arg Met Lys Arg Met Ala His
785                 790                 795                 800

Arg Leu Gly Glu Lys Met Leu Asn Lys Lys Leu Lys Asp Gln Lys Thr
                805                 810                 815

Pro Ile Pro Asp Thr Leu Tyr Gln Glu Leu Tyr Asp Tyr Val Asn His
            820                 825                 830

Arg Leu Ser His Asp Leu Ser Asp Glu Ala Arg Ala Leu Leu Pro Asn
        835                 840                 845
```

```
Val Ile Thr Lys Glu Val Ser His Glu Ile Ile Lys Asp Arg Arg Phe
    850                 855                 860

Thr Ser Asp Lys Phe Phe Phe His Val Pro Ile Thr Leu Asn Tyr Gln
865                 870                 875                 880

Ala Ala Asn Ser Pro Ser Lys Phe Asn Gln Arg Val Asn Ala Tyr Leu
                885                 890                 895

Lys Glu His Pro Glu Thr Pro Ile Ile Gly Ile Asp Arg Gly Glu Arg
            900                 905                 910

Asn Leu Ile Tyr Ile Thr Val Ile Asp Ser Thr Gly Lys Ile Leu Glu
        915                 920                 925

Gln Arg Ser Leu Asn Thr Ile Gln Gln Phe Asp Tyr Gln Lys Lys Leu
    930                 935                 940

Asp Asn Arg Glu Lys Glu Arg Val Ala Ala Arg Gln Ala Trp Ser Val
945                 950                 955                 960

Val Gly Thr Ile Lys Asp Leu Lys Gln Gly Tyr Leu Ser Gln Val Ile
                965                 970                 975

His Glu Ile Val Asp Leu Met Ile His Tyr Gln Ala Val Val Val Leu
            980                 985                 990

Glu Asn Leu Asn Phe Gly Phe Lys Ser Lys Arg Thr Gly Ile Ala Glu
        995                 1000                1005

Lys Ala  Val Tyr Gln Gln Phe  Glu Lys Met Leu Ile  Asp Lys Leu
    1010                1015                1020

Asn Cys  Leu Val Leu Lys Asp  Tyr Pro Ala Glu Lys  Val Gly Gly
    1025                1030                1035

Val Leu  Asn Pro Tyr Gln Leu  Thr Asp Gln Phe Thr  Ser Phe Ala
    1040                1045                1050

Lys Met  Gly Thr Gln Ser Gly  Phe Leu Phe Tyr Val  Pro Ala Pro
    1055                1060                1065

Tyr Thr  Ser Lys Ile Asp Pro  Leu Thr Gly Phe Val  Asp Pro Phe
    1070                1075                1080

Val Trp  Lys Thr Ile Lys Asn  His Glu Ser Arg Lys  His Phe Leu
    1085                1090                1095

Glu Gly  Phe Asp Phe Leu His  Tyr Asp Val Lys Thr  Gly Asp Phe
    1100                1105                1110

Ile Leu  His Phe Lys Met Asn  Arg Asn Leu Ser Phe  Gln Arg Gly
    1115                1120                1125

Leu Pro  Gly Phe Met Pro Ala  Trp Asp Ile Val Phe  Glu Lys Asn
    1130                1135                1140

Glu Thr  Gln Phe Asp Ala Lys  Gly Thr Pro Phe Ile  Ala Gly Lys
    1145                1150                1155

Arg Ile  Val Pro Val Ile Glu  Asn His Arg Phe Thr  Gly Arg Tyr
    1160                1165                1170
```

Arg Asp Leu Tyr Pro Ala Asn Glu Leu Ile Ala Leu Leu Glu Glu
1175 1180 1185

Lys Gly Ile Val Phe Arg Asp Gly Ser Asn Ile Leu Pro Lys Leu
1190 1195 1200

Leu Glu Asn Asp Asp Ser His Ala Ile Asp Thr Met Val Ala Leu
1205 1210 1215

Ile Arg Ser Val Leu Gln Met Arg Asn Ser Asn Ala Ala Thr Gly
1220 1225 1230

Glu Asp Tyr Ile Asn Ser Pro Val Arg Asp Leu Asn Gly Val Cys
1235 1240 1245

Phe Asp Ser Arg Phe Gln Asn Pro Glu Trp Pro Met Asp Ala Asp
1250 1255 1260

Ala Asn Gly Ala Tyr His Ile Ala Leu Lys Gly Gln Leu Leu Leu
1265 1270 1275

Asn His Leu Lys Glu Ser Lys Asp Leu Lys Leu Gln Asn Gly Ile
1280 1285 1290

Ser Asn Gln Asp Trp Leu Ala Tyr Ile Gln Glu Leu Arg Asn
1295 1300 1305

<210> 1523
<211> 169
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 1523
Gly Pro His Met Val Val Ala Gly Ile Asp Pro Gly Ile Thr His Leu
1 5 10 15

Gly Leu Gly Val Val Ala Val Glu Gly Lys Gly Ala Leu Lys Ala Arg
20 25 30

Leu Leu His Gly Glu Val Val Lys Thr Ser Pro Gln Glu Pro Ala Lys
35 40 45

Glu Arg Val Gly Arg Ile His Ala Arg Val Leu Glu Val Leu His Arg
50 55 60

Phe Arg Pro Glu Ala Val Ala Val Gln Glu Gln Phe Phe Tyr Arg Gln
65 70 75 80

Asn Glu Leu Ala Tyr Lys Val Gly Trp Ala Leu Gly Ala Val Leu Val
85 90 95

Ala Ala Phe Glu Ala Gly Val Pro Val Tyr Ala Tyr Gly Pro Met Gln
100 105 110

Val Lys Gln Ala Leu Ala Gly His Gly His Ala Ala Lys Glu Glu Val
115 120 125

Ala Leu Met Val Arg Gly Ile Leu Gly Leu Lys Glu Ala Pro Arg Pro
130 135 140

Ser His Leu Ala Asp Ala Leu Ala Ile Ala Leu Thr His Ala Phe Tyr

145 150 155 160

Ala Arg Met Gly Thr Ala Lys Pro Leu
165

<210> 1524
<211> 1307
<212> PRT
<213> 胺基酸球菌屬某種

<400> 1524
Met Thr Gln Phe Glu Gly Phe Thr Asn Leu Tyr Gln Val Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Lys His Ile Gln
20 25 30

Glu Gln Gly Phe Ile Glu Glu Asp Lys Ala Arg Asn Asp His Tyr Lys
35 40 45

Glu Leu Lys Pro Ile Ile Asp Arg Ile Tyr Lys Thr Tyr Ala Asp Gln
50 55 60

Cys Leu Gln Leu Val Gln Leu Asp Trp Glu Asn Leu Ser Ala Ala Ile
65 70 75 80

Asp Ser Tyr Arg Lys Glu Lys Thr Glu Glu Thr Arg Asn Ala Leu Ile
85 90 95

Glu Glu Gln Ala Thr Tyr Arg Asn Ala Ile His Asp Tyr Phe Ile Gly
100 105 110

Arg Thr Asp Asn Leu Thr Asp Ala Ile Asn Lys Arg His Ala Glu Ile
115 120 125

Tyr Lys Gly Leu Phe Lys Ala Glu Leu Phe Asn Gly Lys Val Leu Lys
130 135 140

Gln Leu Gly Thr Val Thr Thr Thr Glu His Glu Asn Ala Leu Leu Arg
145 150 155 160

Ser Phe Asp Lys Phe Thr Thr Tyr Phe Ser Gly Phe Tyr Glu Asn Arg
165 170 175

Lys Asn Val Phe Ser Ala Glu Asp Ile Ser Thr Ala Ile Pro His Arg
180 185 190

Ile Val Gln Asp Asn Phe Pro Lys Phe Lys Glu Asn Cys His Ile Phe
195 200 205

Thr Arg Leu Ile Thr Ala Val Pro Ser Leu Arg Glu His Phe Glu Asn
210 215 220

Val Lys Lys Ala Ile Gly Ile Phe Val Ser Thr Ser Ile Glu Glu Val
225 230 235 240

Phe Ser Phe Pro Phe Tyr Asn Gln Leu Leu Thr Gln Thr Gln Ile Asp
245 250 255

Leu Tyr Asn Gln Leu Leu Gly Gly Ile Ser Arg Glu Ala Gly Thr Glu
260 265 270

Lys Ile Lys Gly Leu Asn Glu Val Leu Asn Leu Ala Ile Gln Lys Asn
275 280 285

```
Asp Glu Thr Ala His Ile Ile Ala Ser Leu Pro His Arg Phe Ile Pro
    290                 295                 300

Leu Phe Lys Gln Ile Leu Ser Asp Arg Asn Thr Leu Ser Phe Ile Leu
305                 310                 315                 320

Glu Glu Phe Lys Ser Asp Glu Glu Val Ile Gln Ser Phe Cys Lys Tyr
                325                 330                 335

Lys Thr Leu Leu Arg Asn Glu Asn Val Leu Glu Thr Ala Glu Ala Leu
            340                 345                 350

Phe Asn Glu Leu Asn Ser Ile Asp Leu Thr His Ile Phe Ile Ser His
        355                 360                 365

Lys Lys Leu Glu Thr Ile Ser Ser Ala Leu Cys Asp His Trp Asp Thr
    370                 375                 380

Leu Arg Asn Ala Leu Tyr Glu Arg Arg Ile Ser Glu Leu Thr Gly Lys
385                 390                 395                 400

Ile Thr Lys Ser Ala Lys Glu Lys Val Gln Arg Ser Leu Lys His Glu
                405                 410                 415

Asp Ile Asn Leu Gln Glu Ile Ile Ser Ala Ala Gly Lys Glu Leu Ser
            420                 425                 430

Glu Ala Phe Lys Gln Lys Thr Ser Glu Ile Leu Ser His Ala His Ala
        435                 440                 445

Ala Leu Asp Gln Pro Leu Pro Thr Thr Leu Lys Lys Gln Glu Glu Lys
    450                 455                 460

Glu Ile Leu Lys Ser Gln Leu Asp Ser Leu Leu Gly Leu Tyr His Leu
465                 470                 475                 480

Leu Asp Trp Phe Ala Val Asp Glu Ser Asn Glu Val Asp Pro Glu Phe
                485                 490                 495

Ser Ala Arg Leu Thr Gly Ile Lys Leu Glu Met Glu Pro Ser Leu Ser
            500                 505                 510

Phe Tyr Asn Lys Ala Arg Asn Tyr Ala Thr Lys Lys Pro Tyr Ser Val
        515                 520                 525

Glu Lys Phe Lys Leu Asn Phe Gln Met Pro Thr Leu Ala Ser Gly Trp
    530                 535                 540

Asp Val Asn Lys Glu Lys Asn Asn Gly Ala Ile Leu Phe Val Lys Asn
545                 550                 555                 560

Gly Leu Tyr Tyr Leu Gly Ile Met Pro Lys Gln Lys Gly Arg Tyr Lys
                565                 570                 575

Ala Leu Ser Phe Glu Pro Thr Glu Lys Thr Ser Glu Gly Phe Asp Lys
            580                 585                 590

Met Tyr Tyr Asp Tyr Phe Pro Asp Ala Ala Lys Met Ile Pro Lys Cys
        595                 600                 605

Ser Thr Gln Leu Lys Ala Val Thr Ala His Phe Gln Thr His Thr Thr
    610                 615                 620
```

```
Pro Ile Leu Leu Ser Asn Asn Phe Ile Glu Pro Leu Glu Ile Thr Lys
625                 630                 635                 640

Glu Ile Tyr Asp Leu Asn Asn Pro Glu Lys Glu Pro Lys Lys Phe Gln
                645                 650                 655

Thr Ala Tyr Ala Lys Lys Thr Gly Asp Gln Lys Gly Tyr Arg Glu Ala
            660                 665                 670

Leu Cys Lys Trp Ile Asp Phe Thr Arg Asp Phe Leu Ser Lys Tyr Thr
        675                 680                 685

Lys Thr Thr Ser Ile Asp Leu Ser Ser Leu Arg Pro Ser Ser Gln Tyr
    690                 695                 700

Lys Asp Leu Gly Glu Tyr Tyr Ala Glu Leu Asn Pro Leu Leu Tyr His
705                 710                 715                 720

Ile Ser Phe Gln Arg Ile Ala Glu Lys Glu Ile Met Asp Ala Val Glu
                725                 730                 735

Thr Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ala Lys
            740                 745                 750

Gly His His Gly Lys Pro Asn Leu His Thr Leu Tyr Trp Thr Gly Leu
        755                 760                 765

Phe Ser Pro Glu Asn Leu Ala Lys Thr Ser Ile Lys Leu Asn Gly Gln
    770                 775                 780

Ala Glu Leu Phe Tyr Arg Pro Lys Ser Arg Met Lys Arg Met Ala His
785                 790                 795                 800

Arg Leu Gly Glu Lys Met Leu Asn Lys Lys Leu Lys Asp Gln Lys Thr
                805                 810                 815

Pro Ile Pro Asp Thr Leu Tyr Gln Glu Leu Tyr Asp Tyr Val Asn His
            820                 825                 830

Arg Leu Ser His Asp Leu Ser Asp Glu Ala Arg Ala Leu Leu Pro Asn
        835                 840                 845

Val Ile Thr Lys Glu Val Ser His Glu Ile Ile Lys Asp Arg Arg Phe
    850                 855                 860

Thr Ser Asp Lys Phe Phe Phe His Val Pro Ile Thr Leu Asn Tyr Gln
865                 870                 875                 880

Ala Ala Asn Ser Pro Ser Lys Phe Asn Gln Arg Val Asn Ala Tyr Leu
                885                 890                 895

Lys Glu His Pro Glu Thr Pro Ile Ile Gly Ile Asp Arg Gly Glu Arg
            900                 905                 910

Asn Leu Ile Tyr Ile Thr Val Ile Asp Ser Thr Gly Lys Ile Leu Glu
        915                 920                 925

Gln Arg Ser Leu Asn Thr Ile Gln Gln Phe Asp Tyr Gln Lys Lys Leu
    930                 935                 940

Asp Asn Arg Glu Lys Glu Arg Val Ala Ala Arg Gln Ala Trp Ser Val
945                 950                 955                 960
```

```
Val Gly Thr Ile Lys Asp Leu Lys Gln Gly Tyr Leu Ser Gln Val Ile
                965                 970                 975

His Glu Ile Val Asp Leu Met Ile His Tyr Gln Ala Val Val Val Leu
            980                 985                 990

Glu Asn Leu Asn Phe Gly Phe Lys Ser Lys Arg Thr Gly Ile Ala Glu
        995                1000                1005

Lys Ala  Val Tyr Gln Gln Phe  Glu Lys Met Leu Ile  Asp Lys Leu
    1010                1015                1020

Asn Cys  Leu Val Leu Lys Asp  Tyr Pro Ala Glu Lys  Val Gly Gly
    1025                1030                1035

Val Leu  Asn Pro Tyr Gln Leu  Thr Asp Gln Phe Thr  Ser Phe Ala
    1040                1045                1050

Lys Met  Gly Thr Gln Ser Gly  Phe Leu Phe Tyr Val  Pro Ala Pro
    1055                1060                1065

Tyr Thr  Ser Lys Ile Asp Pro  Leu Thr Gly Phe Val  Asp Pro Phe
    1070                1075                1080

Val Trp  Lys Thr Ile Lys Asn  His Glu Ser Arg Lys  His Phe Leu
    1085                1090                1095

Glu Gly  Phe Asp Phe Leu His  Tyr Asp Val Lys Thr  Gly Asp Phe
    1100                1105                1110

Ile Leu  His Phe Lys Met Asn  Arg Asn Leu Ser Phe  Gln Arg Gly
    1115                1120                1125

Leu Pro  Gly Phe Met Pro Ala  Trp Asp Ile Val Phe  Glu Lys Asn
    1130                1135                1140

Glu Thr  Gln Phe Asp Ala Lys  Gly Thr Pro Phe Ile  Ala Gly Lys
    1145                1150                1155

Arg Ile  Val Pro Val Ile Glu  Asn His Arg Phe Thr  Gly Arg Tyr
    1160                1165                1170

Arg Asp  Leu Tyr Pro Ala Asn  Glu Leu Ile Ala Leu  Leu Glu Glu
    1175                1180                1185

Lys Gly  Ile Val Phe Arg Asp  Gly Ser Asn Ile Leu  Pro Lys Leu
    1190                1195                1200

Leu Glu  Asn Asp Asp Ser His  Ala Ile Asp Thr Met  Val Ala Leu
    1205                1210                1215

Ile Arg  Ser Val Leu Gln Met  Arg Asn Ser Asn Ala  Ala Thr Gly
    1220                1225                1230

Glu Asp  Tyr Ile Asn Ser Pro  Val Arg Asp Leu Asn  Gly Val Cys
    1235                1240                1245

Phe Asp  Ser Arg Phe Gln Asn  Pro Glu Trp Pro Met  Asp Ala Asp
    1250                1255                1260

Ala Asn  Gly Ala Tyr His Ile  Ala Leu Lys Gly Gln  Leu Leu Leu
    1265                1270                1275
```

Asn His Leu Lys Glu Ser Lys Asp Leu Lys Leu Gln Asn Gly Ile
1280 1285 1290

Ser Asn Gln Asp Trp Leu Ala Tyr Ile Gln Glu Leu Arg Asn
1295 1300 1305

<210> 1525
<211> 1300
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 1525
Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
20 25 30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
35 40 45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
50 55 60

Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65 70 75 80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
85 90 95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
100 105 110

Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
115 120 125

Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
130 135 140

Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145 150 155 160

Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
165 170 175

Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190

Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205

Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
210 215 220

Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240

Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
245 250 255

Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
260 265 270

Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys

275 280 285

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300

Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320

Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
325 330 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
340 345 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
355 360 365

Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
370 375 380

Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385 390 395 400

Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
405 410 415

Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
420 425 430

Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
435 440 445

Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
450 455 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465 470 475 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
485 490 495

Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
500 505 510

Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
515 520 525

Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
530 535 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545 550 555 560

Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
565 570 575

Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
580 585 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
595 600 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys

610 615 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625 630 635 640

Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
645 650 655

Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
660 665 670

Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
675 680 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln
690 695 700

Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
705 710 715 720

Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp
725 730 735

Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
740 745 750

Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
755 760 765

Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
770 775 780

Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
785 790 795 800

Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
805 810 815

Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
820 825 830

Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
835 840 845

Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu
850 855 860

Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
865 870 875 880

His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
885 890 895

Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
900 905 910

Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
915 920 925

Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
930 935 940

Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile

945 950 955 960

Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
965 970 975

Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
980 985 990

Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu
995 1000 1005

Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val
1010 1015 1020

Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu
1025 1030 1035

Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg
1040 1045 1050

Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly
1055 1060 1065

Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser
1070 1075 1080

Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys
1085 1090 1095

Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp
1100 1105 1110

Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe
1115 1120 1125

Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr
1130 1135 1140

Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp
1145 1150 1155

Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu
1160 1165 1170

Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly
1175 1180 1185

Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe
1190 1195 1200

Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg
1205 1210 1215

Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val
1220 1225 1230

Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1235 1240 1245

Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly
1250 1255 1260

Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu

1265 1270 1275

Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1280 1285 1290

Phe Val Gln Asn Arg Asn Asn
1295 1300

<210> 1526
<211> 119
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 1526
Phe Ala Lys Gly His His Gly Lys Pro Asn Leu His Thr Leu Tyr Trp
1 5 10 15

Thr Gly Leu Phe Ser Pro Glu Asn Leu Ala Lys Thr Ser Ile Lys Leu
20 25 30

Asn Gly Gln Ala Glu Leu Phe Tyr Arg Pro Lys Ser Arg Met Lys Arg
35 40 45

Met Ala His Arg Leu Gly Glu Lys Met Leu Asn Lys Lys Leu Lys Asp
50 55 60

Gln Lys Thr Pro Ile Pro Asp Thr Leu Tyr Gln Glu Leu Tyr Asp Tyr
65 70 75 80

Val Asn His Arg Leu Ser His Asp Leu Ser Asp Glu Ala Arg Ala Leu
85 90 95

Leu Pro Asn Val Ile Thr Lys Glu Val Ser His Glu Ile Ile Lys Asp
100 105 110

Arg Arg Phe Thr Ser Asp Lys
115

<210> 1527
<211> 118
<212> PRT
<213> 腸道羅斯氏菌（Roseburia intestinalis）

<400> 1527
Phe Arg Glu Gly Val Lys Asp Lys Pro Leu Ser Tyr Ile Cys Phe Glu
1 5 10 15

Val Leu Tyr Trp Thr Gly Met Arg Glu Gly Glu Leu Leu Ala Leu Ser
20 25 30

Pro Ala Asp Ile Asp Ile Asp Asn Lys Leu Ile Ser Ile Asn Arg Thr
35 40 45

Tyr Gln Arg Ile Gly Gly Lys Asp Val Phe Thr Ser Pro Lys Thr Arg
50 55 60

Lys Ser Lys Arg Thr Ile Pro Ile Pro Asp Phe Leu Cys Gln Glu Leu
65 70 75 80

Ser Asp Tyr Ile Gln Ser Arg Tyr Met Leu Asp Ala Asp Glu Arg Leu
85 90 95

Phe Pro Val Thr Lys Ser Tyr Leu Ser His Glu Met Ile Arg Gly Cys

100 105 110

Lys Ile Thr Gly Ala Lys
115

<210> 1528
<211> 44
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 1528
Ala Tyr Ser Lys Gly Arg Pro Asn Leu His Thr Leu Tyr Trp Lys Ala
1 5 10 15

Leu Phe Asp Glu Arg Asn Leu Gln Asp Val Val Tyr Lys Leu Asn Gly
20 25 30

Glu Ala Glu Leu Phe Tyr Arg Lys Gln Ser Ile Pro
35 40

<210> 1529
<211> 48
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 1529
Ser Tyr Tyr Lys Tyr Gly Asn Glu Lys Val Leu Arg Gly Tyr Glu Gly
1 5 10 15

Val Leu Leu Ser Ile Leu Lys Asp Glu Asn Leu Val Ser Met Arg Thr
20 25 30

Leu Leu Asn Ser Arg Pro Met Leu Val Tyr Arg Pro Lys Glu Ser Ser
35 40 45

<210> 1530
<211> 36
<212> PRT
<213> 解朊丁酸弧菌

<400> 1530
Pro Tyr Glu Arg Thr Phe Ile Gln Leu Phe Ser Asp Ser Asn Met Glu
1 5 10 15

His Thr Ser Met Leu Leu Asn Ser Arg Ala Met Ile Gln Tyr Arg Ala
20 25 30

Ala Ser Leu Pro
35

<210> 1531
<211> 59
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選佛爾扣菌細菌序列

<400> 1531
Asp Phe Glu Arg Lys Ser Glu Asp Lys Lys Asn His Thr Ala Asn Leu
1 5 10 15

Phe Thr Gln Tyr Leu Leu Glu Leu Phe Ser Cys Glu Asn Ile Lys Asn
20 25 30

Ile Lys Ser Lys Asp Leu Ile Glu Ser Ile Phe Glu Leu Asp Gly Lys

35 40 45

Ala Glu Ile Arg Phe Arg Pro Lys Thr Asp Asp
50 55

<210> 1532
<211> 43
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選佩萊格裡尼細菌序列

<400> 1532
Asp Gly Lys Ser Ile Gly His Lys Asn Asn Leu His Thr Ile Tyr Trp
1 5 10 15

Asn Ala Ile Phe Glu Asn Phe Asp Asn Arg Pro Lys Leu Asn Gly Glu
20 25 30

Ala Glu Ile Phe Tyr Arg Lys Ala Ile Ser Lys
35 40

<210> 1533
<211> 44
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選分裂WS6細菌序列

<400> 1533
Gly Lys Lys Ser Gly Lys Asp Asn Ile His Thr Ile Tyr Trp Lys Tyr
1 5 10 15

Leu Phe Ser Glu Ser Asn Cys Lys Ser Pro Ile Ile Gly Leu Asn Gly
20 25 30

Gly Ala Glu Ile Phe Phe Arg Glu Gly Gln Lys Asp
35 40

<210> 1534
<211> 46
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
非培養的細菌序列

<400> 1534
Glu Ser Lys Lys Glu Trp Ser Thr Glu Asn Ile His Thr Lys Tyr Phe
1 5 10 15

Lys Leu Leu Phe Asn Glu Lys Asn Leu Gln Asn Leu Val Val Lys Leu
20 25 30

Ser Trp Trp Ala Asp Ile Phe Phe Arg Asp Lys Thr Glu Asn
35 40 45

<210> 1535
<211> 52
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
Parcubacteria群細菌序列

<400> 1535
Leu Asp Lys Ala Arg Asp Gly Lys Ser Lys Thr Thr Gln Lys Asn Leu

1 5 10 15

His Thr Leu Tyr Phe Glu Ser Leu Phe Ser Asn Asp Asn Val Val Gln
20 25 30

Asn Phe Pro Ile Lys Leu Asn Gly Gln Ala Glu Ile Phe Tyr Arg Pro
35 40 45

Lys Thr Glu Lys
50

<210> 1536
<211> 44
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
Parcubacteria群細菌序列

<400> 1536
Asn Thr Lys Asn Gly Ser Lys Asn Leu His Thr Leu Tyr Phe Glu His
1 5 10 15

Ile Leu Ser Ala Glu Asn Leu Asn Asp Pro Val Phe Lys Leu Ser Gly
20 25 30

Met Ala Glu Ile Phe Gln Arg Gln Pro Ser Val Asn
35 40

<210> 1537
<211> 44
<212> PRT
<213> 孔茲氏創傷球菌

<400> 1537
Asn Thr Lys Asn Gly Ser Lys Asn Leu His Thr Leu Tyr Phe Glu His
1 5 10 15

Ile Leu Ser Ala Glu Asn Leu Asn Asp Pro Val Phe Lys Leu Ser Gly
20 25 30

Met Ala Glu Ile Phe Gln Arg Gln Pro Ser Val Asn
35 40

<210> 1538
<211> 44
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選alvus甲烷嗜甲基菌序列

<400> 1538
Pro Glu Ser Lys Gly Lys Lys Asn Leu His Thr Leu Tyr Trp Leu Ser
1 5 10 15

Met Phe Ser Glu Glu Asn Leu Arg Thr Arg Lys Leu Lys Leu Asn Gly
20 25 30

Gln Ala Glu Ile Phe Tyr Arg Lys Lys Leu Glu Lys
35 40

<210> 1539
<211> 44
<212> PRT
<213> 胺基酸球菌屬某種

<400> 1539
Glu Asn Ala His Gly Asn Lys Asn Met His Thr Met Tyr Trp Glu Gly

1 5 10 15

Leu Phe Ser Pro Gln Asn Leu Glu Ser Pro Val Phe Lys Leu Ser Gly
20 25 30

Gly Ala Glu Leu Phe Phe Arg Lys Ser Ser Ile Pro
35 40

<210> 1540
<211> 44
<212> PRT
<213> 溶糊精琥珀酸弧菌

<400> 1540
Lys Gly His His Gly Lys Pro Asn Leu His Thr Leu Tyr Trp Thr Gly
1 5 10 15

Leu Phe Ser Pro Glu Asn Leu Ala Lys Thr Ser Ile Lys Leu Asn Gly
20 25 30

Gln Ala Glu Leu Phe Tyr Arg Pro Lys Ser Arg Met
35 40

<210> 1541
<211> 44
<212> PRT
<213> 真細菌某種

<400> 1541
Ser Lys Ser Thr Gly Thr Lys Asn Leu His Thr Leu Tyr Leu Gln Ala
1 5 10 15

Ile Phe Asp Glu Arg Asn Leu Asn Asn Pro Thr Ile Met Leu Asn Gly
20 25 30

Gly Ala Glu Leu Phe Tyr Arg Lys Glu Ser Ile Glu
35 40

<210> 1542
<211> 44
<212> PRT
<213> 挑剔真細菌

<400> 1542
Ala His Ser Thr Gly Arg Asp Asn Leu His Thr Met Tyr Leu Lys Asn
1 5 10 15

Ile Phe Ser Glu Asp Asn Leu Lys Asn Ile Cys Ile Glu Leu Asn Gly
20 25 30

Glu Ala Glu Leu Phe Tyr Arg Lys Ser Ser Met Lys
35 40

<210> 1543
<211> 44
<212> PRT
<213> 挑剔真細菌

<400> 1543
Glu Asn Ser Thr Gly Lys Glu Asn Leu His Thr Met Tyr Phe Lys Asn
1 5 10 15

Ile Phe Ser Glu Glu Asn Leu Lys Asn Ile Val Ile Lys Leu Asn Gly
20 25 30

Gln Ala Glu Leu Phe Tyr Arg Lys Ala Ser Val Lys
35 40

<210> 1544

<211> 44
<212> PRT
<213> Proteocatella sphenisci

<400> 1544
Glu Asn Ser Thr Gly Lys Glu Asn Leu His Thr Met Tyr Phe Lys Asn
1 5 10 15

Ile Phe Ser Glu Glu Asn Leu Lys Asp Ile Ile Ile Lys Leu Asn Gly
20 25 30

Gln Ala Glu Leu Phe Tyr Arg Arg Ala Ser Val Lys
35 40

<210> 1545
<211> 44
<212> PRT
<213> 厭氧弧菌屬某種

<400> 1545
Ile Asp Ser Lys Gly Thr Lys Asn Leu Asn Thr Ile Tyr Phe Glu Ser
1 5 10 15

Leu Phe Ser Glu Glu Asn Met Ile Glu Lys Met Phe Lys Leu Ser Gly
20 25 30

Glu Ala Glu Ile Phe Tyr Arg Pro Ala Ser Leu Asn
35 40

<210> 1546
<211> 52
<212> PRT
<213> 解糖腺普雷沃菌

<400> 1546
Asp Lys Lys Lys Lys Lys Gly Thr Asp Asn Leu His Thr Met Tyr Trp
1 5 10 15

His Gly Val Phe Ser Asp Glu Asn Leu Lys Ala Val Thr Glu Gly Thr
20 25 30

Gln Pro Ile Ile Lys Leu Asn Gly Glu Ala Glu Met Phe Met Arg Asn
35 40 45

Pro Ser Ile Glu
50

<210> 1547
<211> 47
<212> PRT
<213> 獼猴卟啉單胞菌

<400> 1547
Lys Thr Phe Ser Ala Asn Leu Arg Lys Lys Arg Gly Ala Glu Asn Leu
1 5 10 15

His Thr Met Leu Phe Lys Ala Leu Met Glu Gly Asn Gln Asp Asn Leu
20 25 30

Asp Leu Gly Ser Gly Ala Ile Phe Tyr Arg Ala Lys Ser Leu Asp
35 40 45

<210> 1548
<211> 43
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
候選白蟻甲烷枝原體序列

<400> 1548
Pro Tyr Ser Lys Gly Ile Pro Asn Leu His Thr Leu Tyr Trp Lys Ala
1 5 10 15

Leu Phe Ser Glu Gln Asn Gln Ser Arg Val Tyr Lys Leu Cys Gly Gly
20 25 30

Gly Glu Leu Phe Tyr Arg Lys Ala Ser Leu His
35 40

<210> 1549
<211> 44
<212> PRT
<213> 窮氏互養菌

<400> 1549
Lys Ala Ala Thr Gly Lys Lys Asp Met His Thr Ile Tyr Trp Asn Ala
1 5 10 15

Ala Phe Ser Pro Glu Asn Leu Gln Asp Val Val Val Lys Leu Asn Gly
20 25 30

Glu Ala Glu Leu Phe Tyr Arg Asp Lys Ser Asp Ile
35 40

<210> 1550
<211> 44
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 1550
Asp Gly Ala Gln Gly Ser Pro Asn Leu His Thr Leu Tyr Trp Lys Ala
1 5 10 15

Ile Phe Ser Glu Glu Asn Leu Lys Asp Val Val Leu Lys Leu Asn Gly
20 25 30

Glu Ala Glu Leu Phe Phe Arg Arg Lys Ser Ile Asp
35 40

<210> 1551
<211> 42
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 1551
Asp Lys Ser His Gly Thr Pro Asn Leu His Thr Met Tyr Phe Lys Leu
1 5 10 15

Leu Phe Asp Glu Asn Asn His Gly Gln Ile Arg Leu Ser Gly Gly Ala
20 25 30

Glu Leu Phe Met Arg Arg Ala Ser Leu Lys
35 40

<210> 1552
<211> 44
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 1552
Glu His Ser Lys Gly Lys Pro Asn Leu His Thr Val Tyr Trp Asn Ala
1 5 10 15

Leu Phe Ser Glu Tyr Asn Leu Gln Asn Thr Val Tyr Gln Leu Asn Gly
20 25 30

Ser Ala Glu Ile Phe Phe Arg Lys Ala Ser Ile Pro
35 40

<210> 1553
<211> 44
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 1553
Ala Tyr Ser Lys Gly Arg Pro Asn Leu His Thr Leu Tyr Trp Lys Ala
1 5 10 15

Leu Phe Asp Glu Arg Asn Leu Gln Asp Val Val Tyr Lys Leu Asn Gly
20 25 30

Glu Ala Glu Leu Phe Tyr Arg Lys Gln Ser Ile Pro
35 40

<210> 1554
<211> 44
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 1554
Ala Tyr Ser Lys Gly Arg Pro Asn Leu His Thr Leu Tyr Trp Lys Ala
1 5 10 15

Leu Phe Asp Glu Arg Asn Leu Gln Asp Val Val Tyr Lys Leu Asn Gly
20 25 30

Glu Ala Glu Leu Phe Tyr Arg Lys Lys Ser Ile Pro
35 40

<210> 1555
<211> 44
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 1555
Ala Tyr Ser Lys Gly Arg Pro Asn Leu His Thr Leu Tyr Trp Lys Ala
1 5 10 15

Leu Phe Asp Glu Arg Asn Leu Gln Asp Val Val Tyr Lys Leu Asn Gly
20 25 30

Glu Ala Glu Leu Phe Tyr Arg Lys Gln Ser Ile Pro
35 40

<210> 1556
<211> 44
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 1556
Ala Tyr Ser Lys Gly Arg Pro Asn Leu His Thr Leu Tyr Trp Lys Ala
1 5 10 15

Leu Phe Asp Glu Arg Asn Leu Gln Asp Val Val Tyr Lys Leu Asn Gly
20 25 30

Glu Ala Glu Leu Phe Tyr Arg Lys Gln Ser Ile Pro
35 40

<210> 1557
<211> 44

<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 1557
Glu Tyr Ala Thr Gly Asn Leu Asn Leu His Thr Leu Tyr Leu Lys Met
1 5 10 15

Leu Phe Asp Glu Arg Asn Leu Lys Asp Leu Cys Ile Lys Met Asn Gly
20 25 30

Glu Ala Glu Val Phe Tyr Arg Pro Ala Ser Ile Leu
35 40

<210> 1558
<211> 44
<212> PRT
<213> 溶纖維丁酸弧菌

<400> 1558
Glu Tyr Ser Lys Gly Arg Leu Asn Leu His Thr Leu Tyr Phe Lys Met
1 5 10 15

Leu Phe Asp Glu Arg Asn Leu Lys Asn Val Val Tyr Lys Leu Asn Gly
20 25 30

Gln Ala Glu Val Phe Tyr Arg Pro Ser Ser Ile Lys
35 40

<210> 1559
<211> 44
<212> PRT
<213> 瘤胃假丁酸弧菌

<400> 1559
Pro Tyr Ser Lys Gly Asn Leu Asn Leu His Thr Ile Tyr Leu Gln Met
1 5 10 15

Leu Phe Asp Gln Arg Asn Leu Asn Asn Val Val Tyr Lys Leu Asn Gly
20 25 30

Glu Ala Glu Val Phe Tyr Arg Pro Ala Ser Ile Asn
35 40

<210> 1560
<211> 44
<212> PRT
<213> 口小桿菌屬某種

<400> 1560
Pro Tyr Ser Lys Gly Lys Leu Asn Leu His Thr Leu Tyr Phe Lys Met
1 5 10 15

Leu Phe Asp Gln Arg Asn Leu Asp Asn Val Val Tyr Lys Leu Asn Gly
20 25 30

Glu Ala Glu Val Phe Tyr Arg Pro Ala Ser Ile Glu
35 40

<210> 1561
<211> 44
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 1561

Met Tyr Ser Lys Gly Lys Leu Asn Leu His Thr Leu Tyr Phe Met Met
1 5 10 15

Leu Phe Asp Gln Arg Asn Ile Asp Asp Val Val Tyr Lys Leu Asn Gly
20 25 30

Glu Ala Glu Val Phe Tyr Arg Pro Ala Ser Ile Ser
35 40

<210> 1562
<211> 44
<212> PRT
<213> 丁酸弧菌屬某種

<400> 1562
Glu Tyr Ser Lys Gly Lys Leu Asn Leu His Thr Leu Tyr Phe Met Met
1 5 10 15

Leu Phe Asp Gln Arg Asn Leu Asp Asn Val Val Tyr Lys Leu Asn Gly
20 25 30

Glu Ala Glu Val Phe Tyr Arg Pro Ala Ser Ile Ala
35 40

<210> 1563
<211> 44
<212> PRT
<213> 山羊莫拉氏菌

<400> 1563
Pro Lys Ala His Gly Lys Pro Asn Leu His Thr Leu Tyr Phe Lys Ala
1 5 10 15

Leu Phe Ser Lys Asp Asn Leu Ala Asn Pro Ile Tyr Lys Leu Asn Gly
20 25 30

Glu Ala Gln Ile Phe Tyr Arg Lys Ala Ser Leu Asp
35 40

<210> 1564
<211> 44
<212> PRT
<213> 牛莫拉氏菌

<400> 1564
Pro Lys Ala His Gly Lys Pro Asn Leu His Thr Leu Tyr Phe Lys Ala
1 5 10 15

Leu Phe Ser Glu Asp Asn Leu Ala Asp Pro Ile Tyr Lys Leu Asn Gly
20 25 30

Glu Ala Gln Ile Phe Tyr Arg Lys Ala Ser Leu Asp
35 40

<210> 1565
<211> 44
<212> PRT
<213> 稻田鉤端螺旋體

<400> 1565
Ile Phe Ser Lys Gly Lys Pro Asn Leu His Thr Ile Tyr Phe Arg Ser
1 5 10 15

Leu Phe Ser Lys Glu Asn Leu Lys Asp Val Cys Leu Lys Leu Asn Gly
20 25 30

Glu Ala Glu Met Phe Phe Arg Lys Lys Ser Ile Asn
35 40

<210> 1566
<211> 44
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 1566
Pro Tyr Ser Lys Gly Thr Lys Asn Leu His Thr Leu Tyr Trp Glu Met
1 5 10 15

Leu Phe Ser Gln Gln Asn Leu Gln Asn Ile Val Tyr Lys Leu Asn Gly
20 25 30

Asn Ala Glu Ile Phe Tyr Arg Lys Ala Ser Ile Asn
35 40

<210> 1567
<211> 44
<212> PRT
<213> 密斯氏菌屬某種

<400> 1567
Pro Phe Ser Lys Gly Lys Pro Asn Leu His Thr Leu Tyr Trp Lys Met
1 5 10 15

Leu Phe Asp Glu Asn Asn Leu Lys Asp Val Val Tyr Lys Leu Asn Gly
20 25 30

Glu Ala Glu Val Phe Tyr Arg Lys Lys Ser Ile Ala
35 40

<210> 1568
<211> 44
<212> PRT
<213> 密斯氏菌屬某種

<400> 1568
Pro Phe Ser Lys Gly Lys Pro Asn Leu His Thr Leu Tyr Trp Lys Met
1 5 10 15

Leu Phe Asp Glu Asn Asn Leu Lys Asp Val Val Tyr Lys Leu Asn Gly
20 25 30

Glu Ala Glu Val Phe Tyr Arg Lys Lys Ser Ile Ala
35 40

<210> 1569
<211> 44
<212> PRT
<213> 黃桿菌屬某種

<400> 1569
Pro Tyr Ala Lys Gly Lys Pro Asn Met His Thr Leu Tyr Trp Lys Ala
1 5 10 15

Leu Phe Glu Thr Gln Asn Leu Glu Asn Val Ile Tyr Lys Leu Asn Gly
20 25 30

Gln Ala Glu Ile Phe Phe Arg Lys Ala Ser Ile Lys
35 40

<210> 1570
<211> 44
<212> PRT
<213> 嗜鰓黃桿菌

<400> 1570
Pro Phe Ser Lys Gly Lys Pro Asn Met His Thr Leu Tyr Trp Lys Ala
1 5 10 15

Leu Phe Glu Glu Gln Asn Leu Gln Asn Val Ile Tyr Lys Leu Asn Gly
20 25 30

Gln Ala Glu Ile Phe Phe Arg Lys Ala Ser Ile Lys
35 40

<210> 1571
<211> 44
<212> PRT
<213> 狗口腔卟啉單胞菌

<400> 1571
Pro Cys Ser Lys Gly Thr Pro Asn Leu His Thr Leu Tyr Trp Arg Met
1 5 10 15

Leu Phe Asp Glu Arg Asn Leu Ala Asp Val Ile Tyr Lys Leu Asp Gly
20 25 30

Lys Ala Glu Ile Phe Phe Arg Glu Lys Ser Leu Lys
35 40

<210> 1572
<211> 44
<212> PRT
<213> 狗口腔卟啉單胞菌

<400> 1572
Pro Cys Ser Lys Gly Thr Pro Asn Leu His Thr Leu Tyr Trp Arg Met
1 5 10 15

Leu Phe Asp Glu Arg Asn Leu Ala Asp Val Ile Tyr Lys Leu Asp Gly
20 25 30

Lys Ala Glu Ile Phe Phe Arg Glu Lys Ser Leu Lys
35 40

<210> 1573
<211> 44
<212> PRT
<213> 狗口腔卟啉單胞菌

<400> 1573
Pro Cys Ser Lys Gly Thr Pro Asn Leu His Thr Leu Tyr Trp Arg Met
1 5 10 15

Leu Phe Asp Glu Arg Asn Leu Ala Asp Val Ile Tyr Lys Leu Asp Gly
20 25 30

Lys Ala Glu Ile Phe Phe Arg Glu Lys Ser Leu Lys
35 40

<210> 1574
<211> 44
<212> PRT
<213> 狗口腔卟啉單胞菌

<400> 1574
Pro Cys Ser Lys Gly Thr Pro Asn Leu His Thr Leu Tyr Trp Arg Met
1 5 10 15

Leu Phe Asp Glu Arg Asn Leu Ala Asp Val Ile Tyr Lys Leu Asp Gly
20 25 30

Lys Ala Glu Ile Phe Phe Arg Glu Lys Ser Leu Lys
35 40

<210> 1575
<211> 44

<212> PRT
<213> 短普雷沃菌

<400> 1575
Thr Tyr Ser Lys Gly Thr Pro Asn Met His Thr Leu Tyr Trp Lys Met
1 5 10 15

Leu Phe Asp Glu Ser Asn Leu Asn Asp Val Val Tyr Lys Leu Asn Gly
20 25 30

Gln Ala Glu Val Phe Tyr Arg Lys Lys Ser Ile Thr
35 40

<210> 1576
<211> 44
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
口腔類擬桿菌274序列

<400> 1576
Ala His Ser Lys Gly Thr Pro Asn Met His Thr Leu Tyr Trp Lys Met
1 5 10 15

Leu Phe Asp Glu Glu Asn Leu Lys Asp Val Val Tyr Lys Leu Asn Gly
20 25 30

Glu Ala Glu Val Phe Phe Arg Lys Ser Ser Ile Thr
35 40

<210> 1577
<211> 44
<212> PRT
<213> 易北普雷沃菌

<400> 1577
Gln Phe Ser Lys Gly Thr Pro Asn Leu His Thr Leu Tyr Trp Lys Met
1 5 10 15

Leu Phe Asp Lys Arg Asn Leu Ser Asp Val Val Tyr Lys Leu Asn Gly
20 25 30

Gln Ala Glu Val Phe Tyr Arg Lys Ser Ser Ile Glu
35 40

<210> 1578
<211> 44
<212> PRT
<213> 布氏普雷沃菌

<400> 1578
Glu Phe Ser Lys Gly Thr Pro Asn Met His Thr Leu Tyr Trp Asn Ser
1 5 10 15

Leu Phe Ser Lys Glu Asn Leu Asn Asn Ile Ile Tyr Lys Leu Asn Gly
20 25 30

Gln Ala Glu Ile Phe Phe Arg Lys Lys Ser Leu Asn
35 40

<210> 1579
<211> 44
<212> PRT
<213> 布氏普雷沃菌

<400> 1579
Glu Phe Ser Lys Gly Thr Pro Asn Met His Thr Leu Tyr Trp Asn Ser
1 5 10 15

Leu Phe Ser Lys Glu Asn Leu Asn Asn Ile Ile Tyr Lys Leu Asn Gly
20 25 30

Gln Ala Glu Ile Phe Phe Arg Lys Lys Ser Leu Asn
35 40

<210> 1580
<211> 43
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
寡核苷酸

<400> 1580
aatttctact gttgtagatg agaagtcatt taataaggcc act 43

<210> 1581
<211> 468
<212> PRT
<213> 土拉熱弗朗西絲菌

<400> 1581
Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
1 5 10 15

Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu
20 25 30

Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
35 40 45

His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
50 55 60

Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
65 70 75 80

Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
85 90 95

Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
100 105 110

Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
115 120 125

Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
130 135 140

Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
145 150 155 160

Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu
165 170 175

Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val Tyr
180 185 190

Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu Val Phe
195 200 205

Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg Ala Tyr Gln
210 215 220

Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly Lys Gln Thr Gly

225 230 235 240

Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser Lys Ile Cys Pro Val
245 250 255

Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys Tyr Glu Ser Val Ser Lys
260 265 270

Ser Gln Glu Phe Phe Ser Lys Phe Asp Lys Ile Cys Tyr Asn Leu Asp
275 280 285

Lys Gly Tyr Phe Glu Phe Ser Phe Asp Tyr Lys Asn Phe Gly Asp Lys
290 295 300

Ala Ala Lys Gly Lys Trp Thr Ile Ala Ser Phe Gly Ser Arg Leu Ile
305 310 315 320

Asn Phe Arg Asn Ser Asp Lys Asn His Asn Trp Asp Thr Arg Glu Val
325 330 335

Tyr Pro Thr Lys Glu Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu
340 345 350

Tyr Gly His Gly Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp
355 360 365

Lys Lys Phe Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln
370 375 380

Met Arg Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro
385 390 395 400

Val Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
405 410 415

Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly Leu
420 425 430

Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu Gly Lys
435 440 445

Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu Phe Val Gln
450 455 460

Asn Arg Asn Asn
465

<210> 1582
<211> 518
<212> PRT
<213> 胺基酸球菌屬某種

<400> 1582
Arg Pro Lys Ser Arg Met Lys Arg Met Ala His Arg Leu Gly Glu Lys
1 5 10 15

Met Leu Asn Lys Lys Leu Lys Asp Gln Lys Thr Pro Ile Pro Asp Thr
20 25 30

Leu Tyr Gln Glu Leu Tyr Asp Tyr Val Asn His Arg Leu Ser His Asp
35 40 45

Leu Ser Asp Glu Ala Arg Ala Leu Leu Pro Asn Val Ile Thr Lys Glu
50 55 60

```
Val Ser His Glu Ile Ile Lys Asp Arg Arg Phe Thr Ser Asp Lys Phe
65                  70                  75                  80

Phe Phe His Val Pro Ile Thr Leu Asn Tyr Gln Ala Ala Asn Ser Pro
                85                  90                  95

Ser Lys Phe Asn Gln Arg Val Asn Ala Tyr Leu Lys Glu His Pro Glu
            100                 105                 110

Thr Pro Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu Ile Tyr Ile
        115                 120                 125

Thr Val Ile Asp Ser Thr Gly Lys Ile Leu Glu Gln Arg Ser Leu Asn
    130                 135                 140

Thr Ile Gln Gln Phe Asp Tyr Gln Lys Lys Leu Asp Asn Arg Glu Lys
145                 150                 155                 160

Glu Arg Val Ala Ala Arg Gln Ala Trp Ser Val Val Gly Thr Ile Lys
                165                 170                 175

Asp Leu Lys Gln Gly Tyr Leu Ser Gln Val Ile His Glu Ile Val Asp
            180                 185                 190

Leu Met Ile His Tyr Gln Ala Val Val Val Leu Glu Asn Leu Asn Phe
        195                 200                 205

Gly Phe Lys Ser Lys Arg Thr Gly Ile Ala Glu Lys Ala Val Tyr Gln
    210                 215                 220

Gln Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Cys Leu Val Leu Lys
225                 230                 235                 240

Asp Tyr Pro Ala Glu Lys Val Gly Gly Val Leu Asn Pro Tyr Gln Leu
                245                 250                 255

Thr Asp Gln Phe Thr Ser Phe Ala Lys Met Gly Thr Gln Ser Gly Phe
            260                 265                 270

Leu Phe Tyr Val Pro Ala Pro Tyr Thr Ser Lys Ile Asp Pro Leu Thr
        275                 280                 285

Gly Phe Val Asp Pro Phe Val Trp Lys Thr Ile Lys Asn His Glu Ser
    290                 295                 300

Arg Lys His Phe Leu Glu Gly Phe Asp Phe Leu His Tyr Asp Val Lys
305                 310                 315                 320

Thr Gly Asp Phe Ile Leu His Phe Lys Met Asn Arg Asn Leu Ser Phe
                325                 330                 335

Gln Arg Gly Leu Pro Gly Phe Met Pro Ala Trp Asp Ile Val Phe Glu
            340                 345                 350

Lys Asn Glu Thr Gln Phe Asp Ala Lys Gly Thr Pro Phe Ile Ala Gly
        355                 360                 365

Lys Arg Ile Val Pro Val Ile Glu Asn His Arg Phe Thr Gly Arg Tyr
    370                 375                 380

Arg Asp Leu Tyr Pro Ala Asn Glu Leu Ile Ala Leu Leu Glu Glu Lys
385                 390                 395                 400
```

Gly Ile Val Phe Arg Asp Gly Ser Asn Ile Leu Pro Lys Leu Leu Glu
405 410 415

Asn Asp Asp Ser His Ala Ile Asp Thr Met Val Ala Leu Ile Arg Ser
420 425 430

Val Leu Gln Met Arg Asn Ser Asn Ala Ala Thr Gly Glu Asp Tyr Ile
435 440 445

Asn Ser Pro Val Arg Asp Leu Asn Gly Val Cys Phe Asp Ser Arg Phe
450 455 460

Gln Asn Pro Glu Trp Pro Met Asp Ala Asp Ala Asn Gly Ala Tyr His
465 470 475 480

Ile Ala Leu Lys Gly Gln Leu Leu Leu Asn His Leu Lys Glu Ser Lys
485 490 495

Asp Leu Lys Leu Gln Asn Gly Ile Ser Asn Gln Asp Trp Leu Ala Tyr
500 505 510

Ile Gln Glu Leu Arg Asn
515

<210> 1583
<211> 482
<212> PRT
<213> 未知

<220>
<223> 未知的描述：
毛螺旋菌科細菌序列

<400> 1583
Arg Arg Ala Ser Leu Lys Lys Glu Glu Leu Val Val His Pro Ala Asn
1 5 10 15

Ser Pro Ile Ala Asn Lys Asn Pro Asp Asn Pro Lys Lys Thr Thr Thr
20 25 30

Leu Ser Tyr Asp Val Tyr Lys Asp Lys Arg Phe Ser Glu Asp Gln Tyr
35 40 45

Glu Leu His Ile Pro Ile Ala Ile Asn Lys Cys Pro Lys Asn Ile Phe
50 55 60

Lys Ile Asn Thr Glu Val Arg Val Leu Leu Lys His Asp Asp Asn Pro
65 70 75 80

Tyr Val Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu Leu Tyr Ile Val
85 90 95

Val Val Asp Gly Lys Gly Asn Ile Val Glu Gln Tyr Ser Leu Asn Glu
100 105 110

Ile Ile Asn Asn Phe Asn Gly Ile Arg Ile Lys Thr Asp Tyr His Ser
115 120 125

Leu Leu Asp Lys Lys Glu Lys Glu Arg Phe Glu Ala Arg Gln Asn Trp
130 135 140

Thr Ser Ile Glu Asn Ile Lys Glu Leu Lys Ala Gly Tyr Ile Ser Gln
145 150 155 160

Val Val His Lys Ile Cys Glu Leu Val Glu Lys Tyr Asp Ala Val Ile

```
                165                 170                 175

Ala Leu Glu Asp Leu Asn Ser Gly Phe Lys Asn Ser Arg Val Lys Val
            180                 185                 190

Glu Lys Gln Val Tyr Gln Lys Phe Glu Lys Met Leu Ile Asp Lys Leu
        195                 200                 205

Asn Tyr Met Val Asp Lys Lys Ser Asn Pro Cys Ala Thr Gly Gly Ala
    210                 215                 220

Leu Lys Gly Tyr Gln Ile Thr Asn Lys Phe Glu Ser Phe Lys Ser Met
225                 230                 235                 240

Ser Thr Gln Asn Gly Phe Ile Phe Tyr Ile Pro Ala Trp Leu Thr Ser
                245                 250                 255

Lys Ile Asp Pro Ser Thr Gly Phe Val Asn Leu Leu Lys Thr Lys Tyr
            260                 265                 270

Thr Ser Ile Ala Asp Ser Lys Lys Phe Ile Ser Ser Phe Asp Arg Ile
        275                 280                 285

Met Tyr Val Pro Glu Glu Asp Leu Phe Glu Phe Ala Leu Asp Tyr Lys
    290                 295                 300

Asn Phe Ser Arg Thr Asp Ala Asp Tyr Ile Lys Lys Trp Lys Leu Tyr
305                 310                 315                 320

Ser Tyr Gly Asn Arg Ile Arg Ile Phe Arg Asn Pro Lys Lys Asn Asn
                325                 330                 335

Val Phe Asp Trp Glu Glu Val Cys Leu Thr Ser Ala Tyr Lys Glu Leu
            340                 345                 350

Phe Asn Lys Tyr Gly Ile Asn Tyr Gln Gln Gly Asp Ile Arg Ala Leu
        355                 360                 365

Leu Cys Glu Gln Ser Asp Lys Ala Phe Tyr Ser Ser Phe Met Ala Leu
    370                 375                 380

Met Ser Leu Met Leu Gln Met Arg Asn Ser Ile Thr Gly Arg Thr Asp
385                 390                 395                 400

Val Asp Phe Leu Ile Ser Pro Val Lys Asn Ser Asp Gly Ile Phe Tyr
                405                 410                 415

Asp Ser Arg Asn Tyr Glu Ala Gln Glu Asn Ala Ile Leu Pro Lys Asn
            420                 425                 430

Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala Arg Lys Val Leu Trp Ala
        435                 440                 445

Ile Gly Gln Phe Lys Lys Ala Glu Asp Glu Lys Leu Asp Lys Val Lys
    450                 455                 460

Ile Ala Ile Ser Asn Lys Glu Trp Leu Glu Tyr Ala Gln Thr Ser Val
465                 470                 475                 480

Lys His
```

<210> 1584

<211> 5
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
肽

<400> 1584
Gly Gly Gly Gly Ser
1 5

<210> 1585
<211> 10
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
肽

<400> 1585
Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
1 5 10

<210> 1586
<211> 20
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
肽

<400> 1586
Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
1 5 10 15

Gly Gly Gly Ser
20

<210> 1587
<211> 25
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
肽

<400> 1587
Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
1 5 10 15

Gly Gly Gly Ser Gly Gly Gly Gly Ser
20 25

<210> 1588
<211> 35
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 1588
Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
1 5 10 15

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
20 25 30

Gly Gly Ser
35

<210> 1589
<211> 40
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 1589
Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
1 5 10 15

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
20 25 30

Gly Gly Ser Gly Gly Gly Gly Ser
35 40

<210> 1590
<211> 50
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 1590
Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
1 5 10 15

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
20 25 30

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly
35 40 45

Gly Ser
50

<210> 1591
<211> 55
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多肽

<400> 1591
Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
1 5 10 15

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
20 25 30

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly
35 40 45

Gly Ser Gly Gly Gly Gly Ser
50 55

<210> 1592
<211> 120
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成的
多核苷酸

<400> 1592
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 60

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 120

<210> 1593
<211> 4
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的肽

<400> 1593
Gly Gly Gly Ser
1

<210> 1594
<211> 12
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的肽

<400> 1594
Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser
1 5 10

<210> 1595
<211> 7
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成的肽

<400> 1595
Ala Glu Ala Ala Ala Lys Ala
1 5

第 111105280 號專利申請案
112 年 10 月 6 日修正替換頁

發明申請專利範圍

1. 一種包含 CRISPR-Cas 系統的組成物，包含

a) 包含 RuvC 核酸酶結構域但不包含 HNH 結構域的 V 型 Cas 蛋白，或編碼該 V 型 Cas 蛋白的多核苷酸，以及

b) V 型 CRISPR-Cas 系統指導物，或編碼該 V 型 CRISPR-Cas 系統指導物的多核苷酸，該 V 型 CRISPR-Cas 系統指導物為經工程化以在真核生物細胞之基因組內與原型間隔區相鄰模體（PAM）之靶序列雜交，其中該 V 型 CRISPR-Cas 系統指導物能與該 V 型 Cas 蛋白形成 CRISPR 複合物。

2. 如申請專利範圍第 1 項所述之組成物，其中該組成物包含該 V 型 CRISPR-Cas 系統指導物及該 V 型 Cas 蛋白。

3. 如申請專利範圍第 1 項所述之組成物，其中該組成物包含藉由該 V 型 CRISPR-Cas 系統指導物及該 V 型 Cas 蛋白所形成之 CRISPR 複合物。

4. 如申請專利範圍第 1 項所述之組成物，其中該組成物包含一種編碼該 V 型 CRISPR-Cas 系統指導物及該 V 型 Cas 蛋白的載體。

5. 如申請專利範圍第 4 項所述之組成物，其中該組成物包含一種編碼該 V 型 CRISPR-Cas 系統指導物及該 V 型 Cas 蛋白的病毒載體。

6. 如申請專利範圍第 5 項所述之組成物，其中該病毒載體

包含腺病毒載體、慢病毒載體、或腺伴隨病毒載體。

7. 如申請專利範圍第 1 項所述之組成物，其中該組成物包含編碼該 V 型 Cas 蛋白的 mRNA。

8. 如申請專利範圍第 7 項所述之組成物，其中該編碼該 V 型 Cas 蛋白的 mRNA 被包含在脂質奈米粒子、脂質體、外來體(exosome)、或微泡中。

9. 如申請專利範圍第 1 項所述之組成物，其中該 V 型 Cas 蛋白連接至少一個核定位信號。

10. 如申請專利範圍第 1 項所述之組成物，該 V 型 Cas 蛋白係來自土拉熱弗朗西絲菌 1、易北普雷沃菌、毛螺旋菌科細菌 MC2017 1、解朊丁酸弧菌、佩萊格裡尼菌科細菌 GW2011_GWA2_33_10、儉菌總門細菌 GW2011_GWC2_44_17、密斯氏菌屬某種 SCADC、胺基酸球菌屬某種 BV3L6、毛螺旋菌科細菌 MA2020、候選白蟻甲烷枝原體、挑剔真細菌、牛莫拉氏菌 237、稻田鉤端螺旋體、毛螺旋菌科細菌 ND2006、狗口腔卟啉單胞菌 3、解糖胨普雷沃菌或獼猴卟啉單胞菌。

11. 如申請專利範圍第 1 項所述之組成物，其中該 V 型 Cas 蛋白係新殺手法蘭西斯菌 U112 Cpf1(FnCpf1)、胺基酸球菌屬某種 BV3L6 Cpf1（AsCpf1）、或毛螺旋菌科細菌 ND2006（LbCpf1）。

12. 如申請專利範圍第 1 項所述之組成物，其中該 V 型 Cas 蛋白在催化結構域內包含一個突變。

13. 如申請專利範圍第 1 項所述之組成物，其中該 V 型 Cas 蛋白在催化結構域內包含一個突變且連接於異源功能結構域。

14. 如申請專利範圍第 13 項所述之組成物，其中該異源功能結構域具有如下之一或多個活性：修飾烷酶活性，去甲酶活性，轉錄啟動活性，轉錄阻遏活性，轉錄釋放因子活性，組蛋白修飾活性，核酸酶活性，單股 RNA 切割活性，雙股 RNA 切割活性，單股 DNA 切割活性，雙股 DNA 切割活性和核酸結合活性。

15. 如申請專利範圍第 1 項所述之組成物，其中該組成物包含或編碼多種 V 型 CRISPR-Cas 系統指導物，該多種 V 型 CRISPR-Cas 系統指導物之各者係工程化以與不同的靶序列雜交。

16. 如申請專利範圍第 1 項所述之組成物，其中該 V 型 CRISPR-Cas 系統指導物包含連接至同向重複序列之指導序列，其中該同向重複序列係藉由 SEQ ID NOs:195 至 210 中任一者所編碼或包含 SEQ ID NOs: 1434 至 1449 中任一者。

17. 如申請專利範圍第 1 項所述之組成物，其中該 V 型 CRISPR-Cas 系統指導物包含一種修飾核苷酸或一種非核苷酸組分。

18. 如申請專利範圍第 1 項所述之組成物，其中該 V 型 CRISPR-Cas 系統指導物包含一個化學修飾，其中該化學

修飾為核糖環中兩個碳原子間的亞甲基橋、硫代磷酸酯鍵、或於一個或多個末端核苷酸處摻入 2´-O-甲基(M)、2´-O-甲基 3´硫代磷酸酯（MS）或 2´-O-甲基 3´硫代 PACE（MSP）。

19. 如申請專利範圍第 1 項所述之組成物，其中該靶序列係鄰近於 T 富集之 PAM(T-rich PAM)。

20. 如申請專利範圍第 19 項所述之組成物，其中該 V 型 Cas 蛋白為 FnCpf1 且該 PAM 為 TTN，其中 N 係 A、C、G、或 T，或其中該 V 型 Cas 蛋白為 AsCpf1 或 LbCpf1 且該 PAM 序列為 TTTV，其中 V 係 A、C、或 G。

21. 如申請專利範圍第 1 項所述之組成物，其中該 CRISPR 複合物在靶序列產生股斷裂(strand break)。

22. 如申請專利範圍第 21 項所述之組成物，其中該股斷裂為具有 5'突出端的交錯雙股斷裂。

23. 如申請專利範圍第 21 項所述之組成物，其中該股斷裂為具有 4 或 5-nt 突出端的交錯雙股斷裂。

24. 如申請專利範圍第 21 項所述之組成物，其中該組成物復包含單股的寡核苷酸，該單股寡核苷酸作為雙股斷裂之同源定向修復的模板。

25. 如申請專利範圍第 21 項所述之組成物，其中該組成物復包含雙股寡核苷酸以嵌入雙股斷裂中。

26. 一種體外或離體之修飾真核生物細胞的方法，包含將申請專利範圍第 1 至 25 項中任一者之組成物遞送到真核

生物細胞，其中該鄰近真核生物細胞之基因組內原型間隔區相鄰模體（PAM）之靶序列係藉由 V 型 Cas 蛋白及 V 型 CRISPR-Cas 系統指導物所形成的 CRISPR 複合物修飾。

27. 如申請專利範圍第 26 項所述之方法，其中該真核生物細胞為哺乳類動物細胞或植物細胞。

28. 如申請專利範圍第 27 項所述之方法，其中該真核生物細胞為人類細胞。

29. 如申請專利範圍第 28 項所述之方法，其中該真核生物細胞為造血細胞或淋巴細胞。

30. 如申請專利範圍第 29 項所述之方法，其中該真核生物細胞為造血 CD34+幹/前驅細胞，自然殺手細胞、細胞毒性 T 淋巴細胞、調節型 T 淋巴細胞、或腫瘤浸潤淋巴細胞。

31. 一種藉由如申請專利範圍第 26 項之方法產生的工程化及非自然出現之真核生物細胞。

32. 一種如申請專利範圍第 1 至 25 項中任一者所述之組成物或如申請專利範圍第 31 項所述之細胞的用途，係用以製備用於基因疾病或病症的藥物。

33. 如申請專利範圍第 32 項所述之用途，其中該基因疾病或病症係血液疾病或病症。

34. 如申請專利範圍第 33 項所述之用途，其中該血液疾病或病症係鐮狀細胞性貧血、β-地中海貧血或血友病。

35. 如申請專利範圍第 32 項所述之用途，其中該基因疾病或

病症係眼科或眼疾病或病症。

36. 如申請專利範圍第 35 項所述之用途，其中該眼科或眼疾病或病症係萊伯氏先天性黑矇（Leber Congenital Amaurosis）（LCA）、Usher 症候群、色素性視網膜炎或原發性開角型青光眼。

37. 如申請專利範圍第 32 項所述之用途，其中該基因疾病或病症係肌肉疾病或病症。

38. 如申請專利範圍第 37 項所述之用途，其中該肌肉疾病或病症係囊性纖維化或杜氏肌營養不良。

39. 如申請專利範圍第 32 項所述之用途，其中該基因疾病或病症係癌症。

40. 如申請專利範圍第 39 項所述之用途，其中該 V 型 CRISPR-Cas 系統指導物與相關於 PD1、CTLA4、TRAC、TRBC、B2M、或 MHC2TA 之靶序列雜交。

圖式

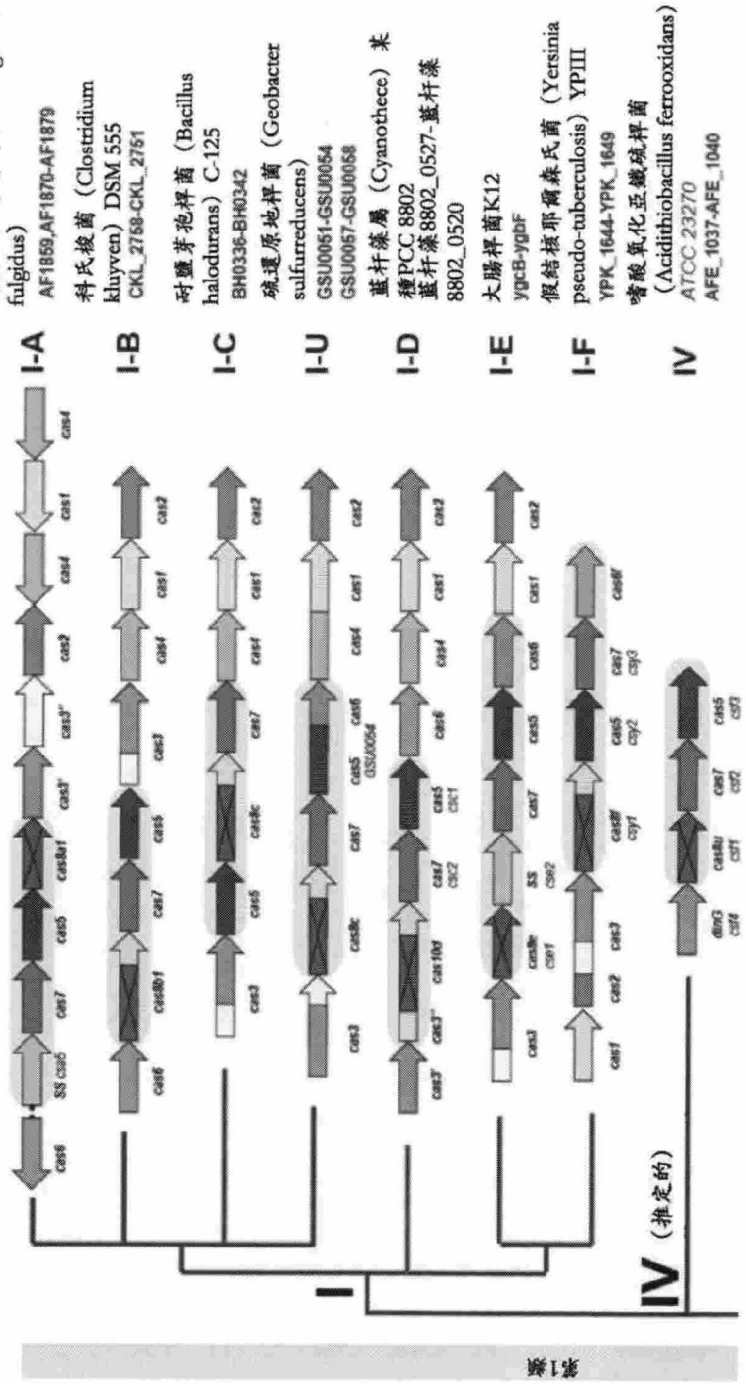

閃球古生球菌（Archaeoglobus fulgidus）
AF1859_AF1870-AF1879
科氏梭菌（Clostridium kluyven）DSM 555
CKL_2758-CKL_2751
耐鹽芽孢桿菌（Bacillus halodurans）C-125
BH0336-BH0342
硫還原地桿菌（Geobacter sulfurreducens）
GSU0051-GSU0054
GSU0057-GSU0058
藍杆藻屬（Cyanothece）菜種PCC 8802
藍杆藻8802_0527-藍杆藻8802_0520
大腸桿菌K12
ygcB-ygbF
假結核耶爾森氏菌（Yersinia pseudo-tuberculosis）YPIII
YPK_1644-YPK_1649
嗜酸氧化亞鐵硫桿菌（Acidithiobacillus ferrooxidans）ATCC 23270
AFE_1037-AFE_1040
I-A
I-B
I-C
I-U
I-D
I-E
I-F
IV
I
IV（推定的）
第1族

圖 1A

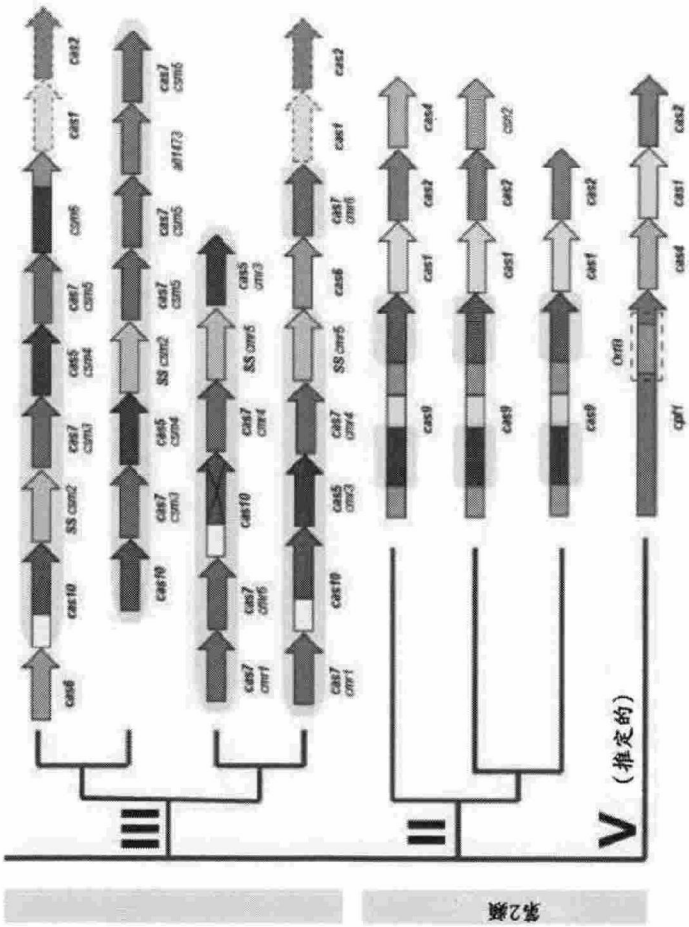

III-A 表皮葡萄球菌RP62A
SERP2463-SERP2455
III-D 集胞藻属某種PCC 6803
sll7067-sll7063
III-C 热自养嗜热甲烷杆菌菌株δH（Methanothermobacter thermautotrophicus Str. Delta H）
MTH328-MTH323
III-B 強烈火球菌（Pyrococcus furiosus）DSM 3638
PF1131-PF1124
II-B 巴黎退伍軍人嗜肺病菌菌株（Legionella pneumophila str. Paris）
lpp0160-lpp0163
II-A 嗜热链球菌CNRZ1066
str0657-str0660
II-C 乳糖奈瑟氏菌（Neisseria lactamica）020-06
NLA_17660-NLA_17680
V 弗朗西絲菌属cf.新殺手Fx1
FNFX1_1431-FNFX1_1428
III
II
V（推定的）
第2類

圖 1B

階段
表現
干擾
自我調整
相關的免疫
區塊
轉錄物切割
效應物模組crRNA和靶標結合
靶切割
間隔區插入
調節，休眠
輔助子，未知作用
I 型
Cas6
SS*
Cas7
Cas5
Cas8/LS
HD Cas3"
Cas3'
Cas1
Cas2
Cas4
II 型
RNAse III
Cas9
Cas1
Cas2
Cas4
Csn2
III 型
Cas6
SS
Cas7
Cas5
Cas10/LS
Cas1
Cas2
CARF
IV 型
LS
SS
Cas7
Cas5
DinG
V 型
Cpf1
Cas1
Cas2
Cas4

圖 2

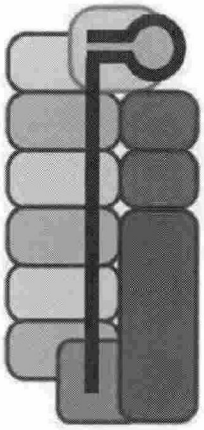


圖 3A

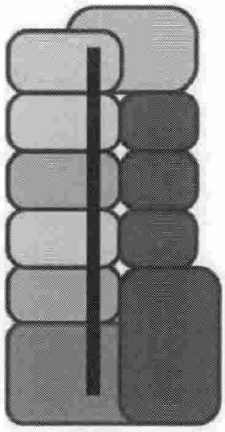


圖 3B

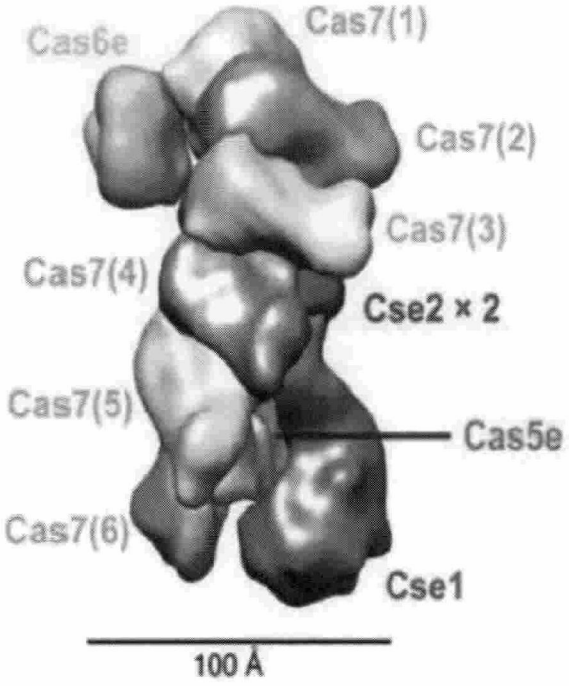

Cas6e
Cas7(1)
Cas7(2)
Cas7(3)
Cas7(4)
Cse2 × 2
Cas7(5)
Cas5e
Cas7(6)
Cse1
100 Å

圖 3C

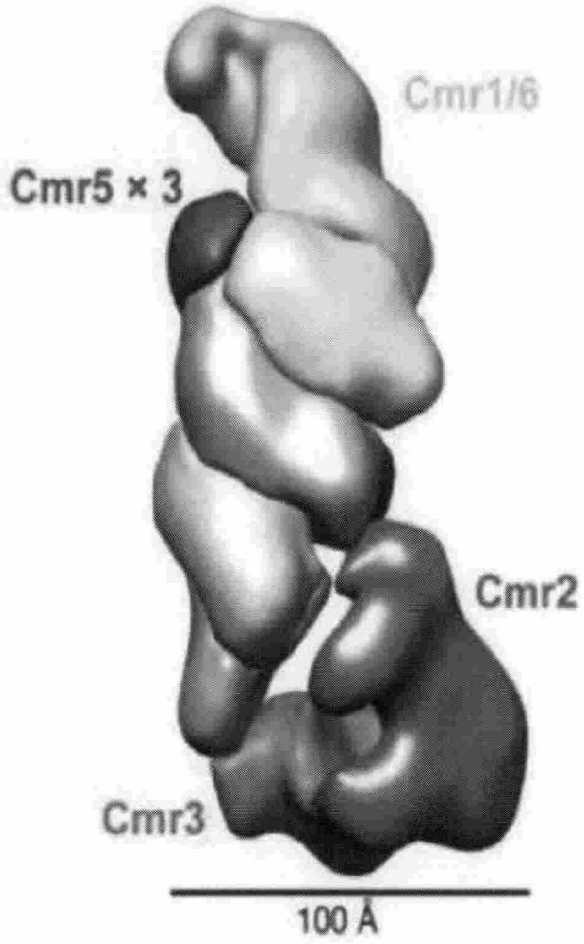


圖 3D

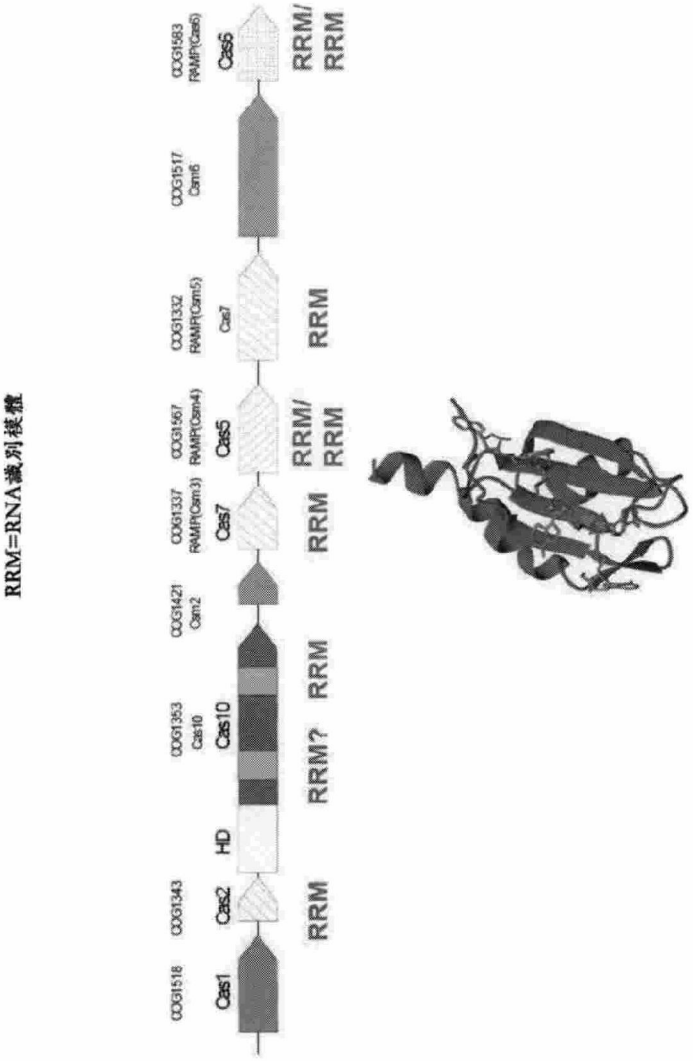

RRM=RNA識別模體
COG1518
Cas1
COG1343
Cas2
HD
COG1353
Cas10
COG1421
Csm2
COG1337
RAMP(Csm3)
Cas7
COG1567
RAMP(Csm4)
Cas5
COG1332
RAMP(Csm5)
Cas7
COG1517
Csm6
COG1583
RAMP(Cas6)
Cas6
RRM
RRM?
RRM
RRM
RRM/
RRM
RRM
RRM/
RRM

圖 4

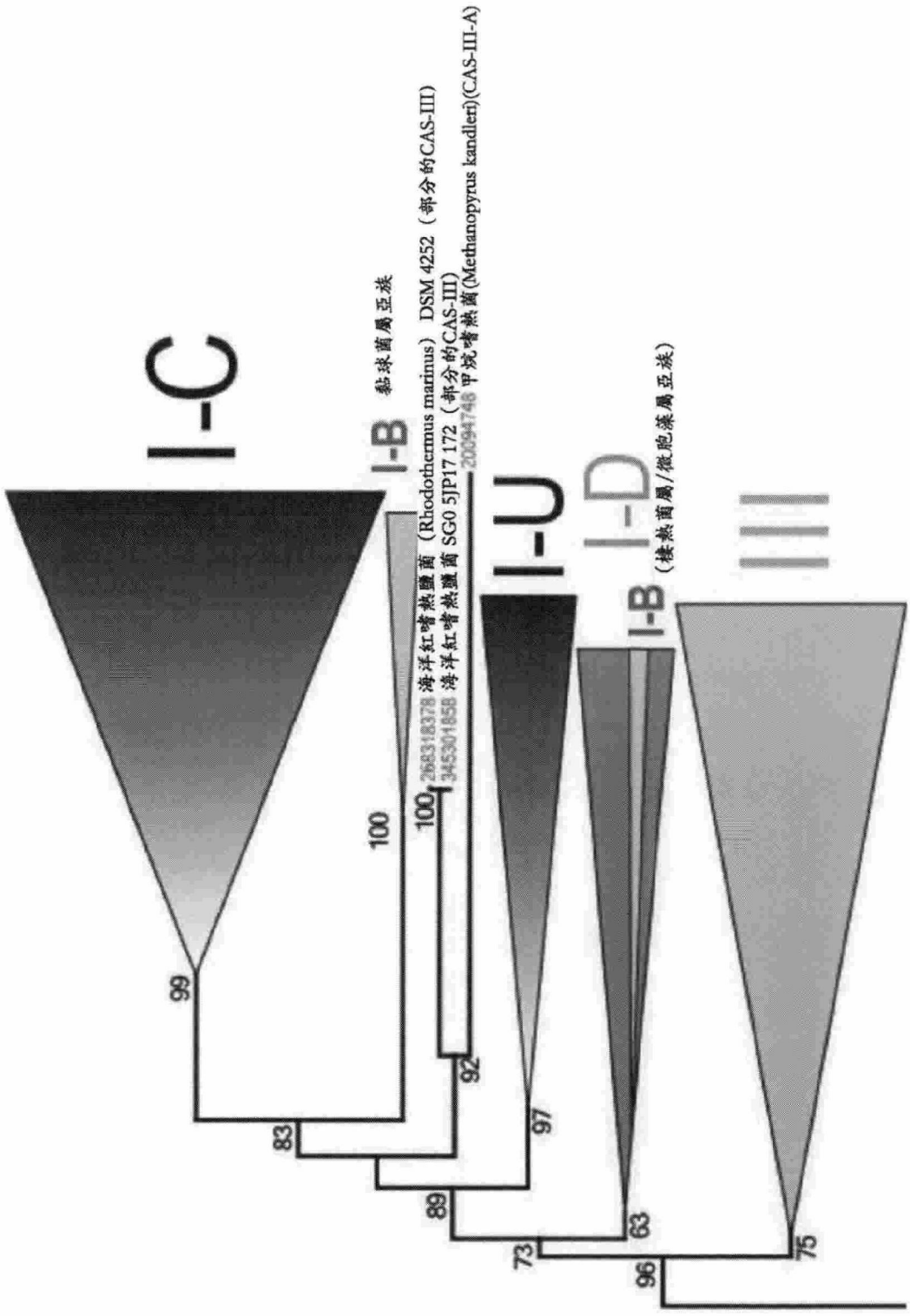

I-C
I-B 黏球菌屬亞族
2683183781 海洋紅嗜熱鹽菌（Rhodothermus marinus） DSM 4252（部分的CAS-III）
3453018581 海洋紅嗜熱鹽菌 SG0 5JP17 172（部分的CAS-III）
200944748 甲烷嗜熱菌(Methanopyrus kandleri)(CAS-III-A)
I-U
I-D
I-B（棲熱菌屬/微胞藻屬亞族）
III
99
100
100
83
92
89
97
73
63
96
75

圖 5A

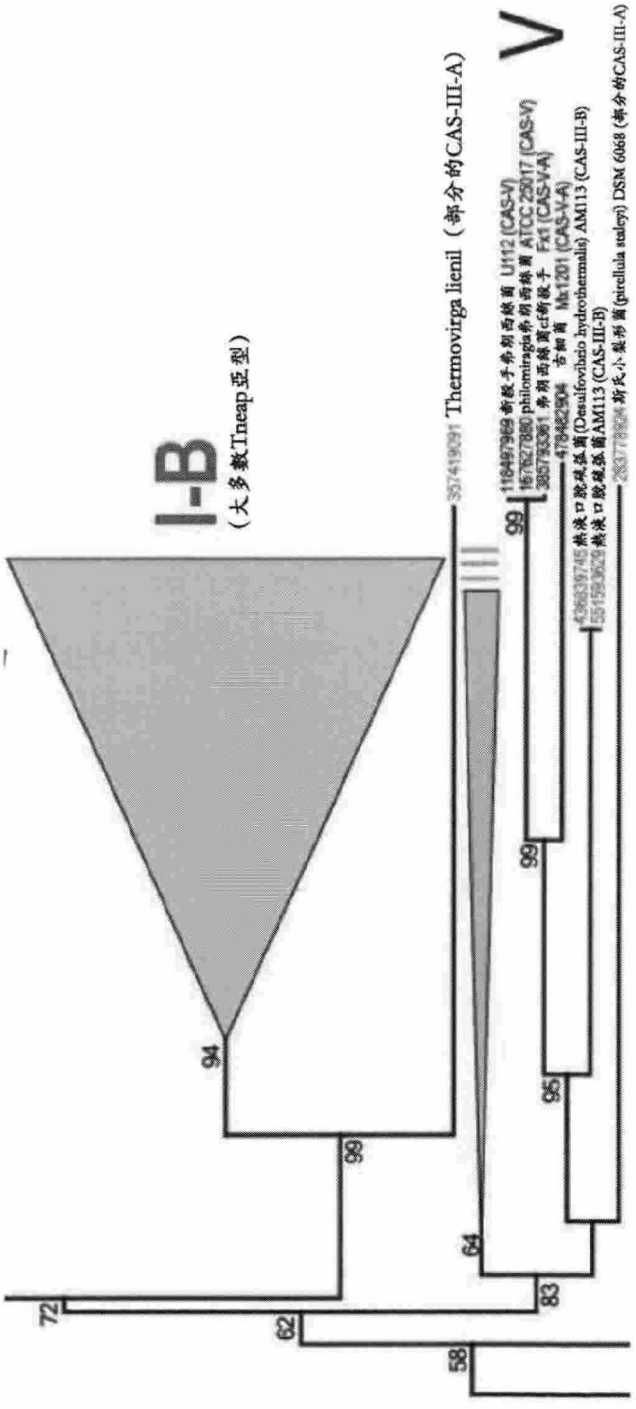

I-B
（大多數Tneap亞型）
V
Thermovirga lienil（部分的CAS-III-A）
94
99
64
99
95
83
72
62
58

圖 5B

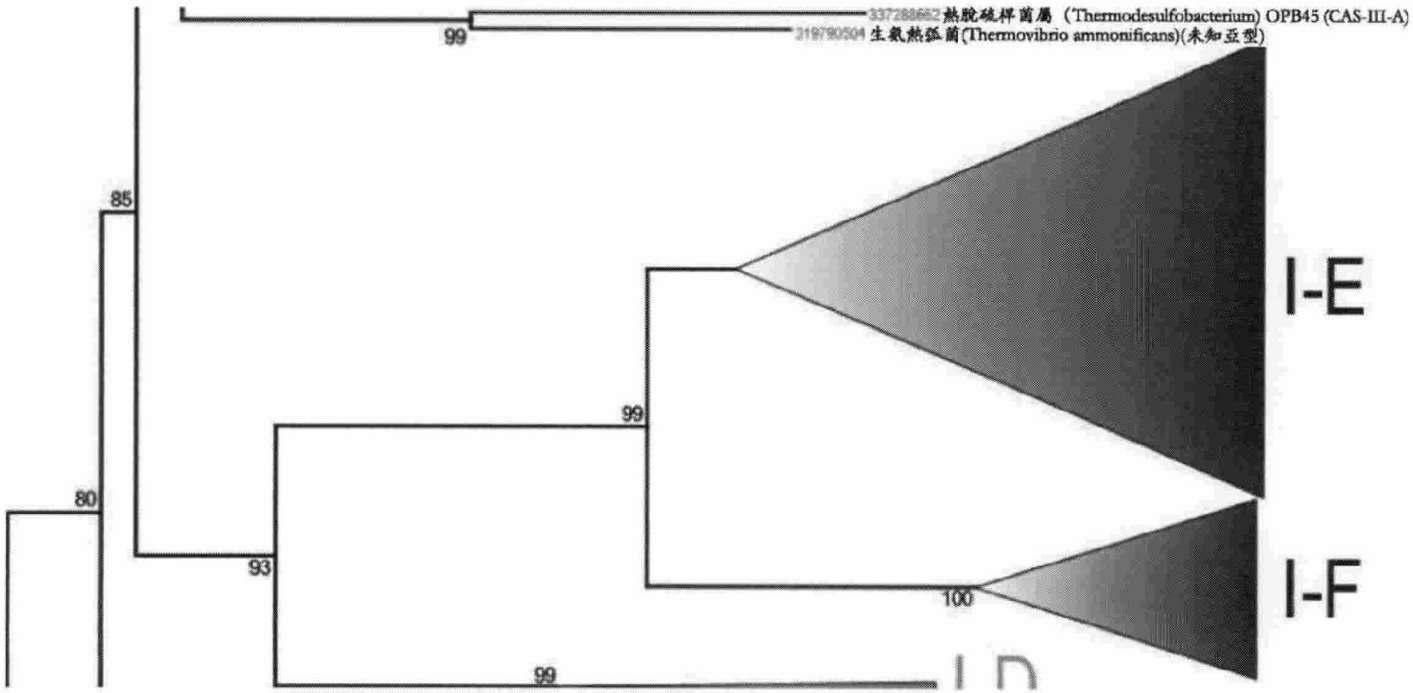

337288662 熱脫硫桿菌屬 (Thermodesulfobacterium) OPB45 (CAS-III-A)
319790504 生氨熱弧菌(Thermovibrio ammonificans)(未知亞型)
99
85
80
93
99
99
100
I-E
I-F

圖 5C

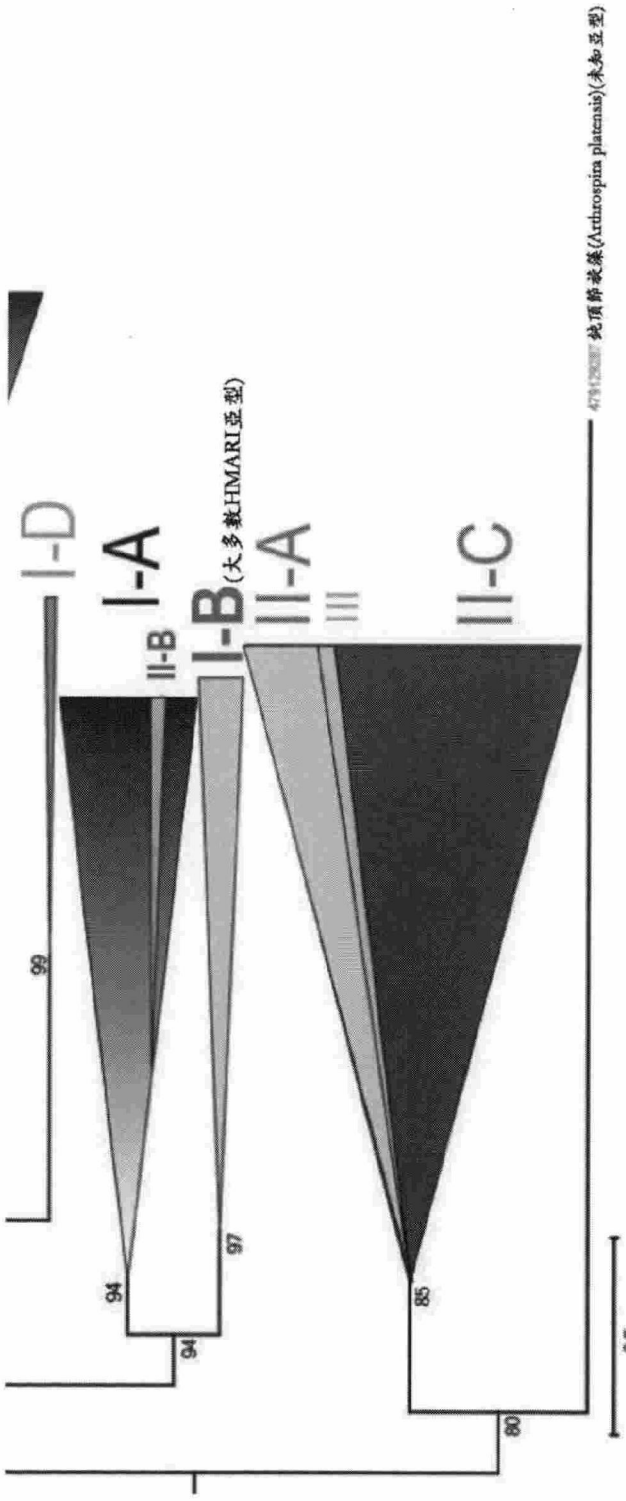
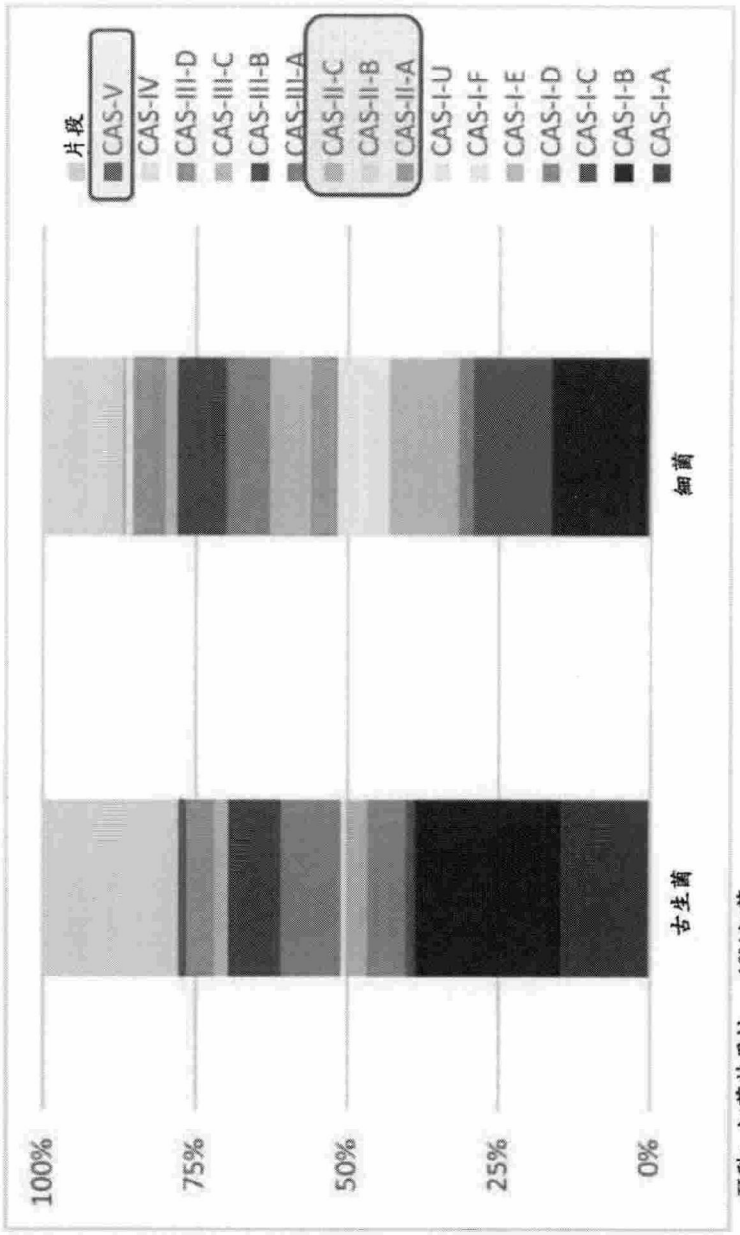


圖 5D



II型：細菌特異性，~10%細菌
V型：稀有的-若干種細菌，一種古生菌

圖 6

簡短的流程描述

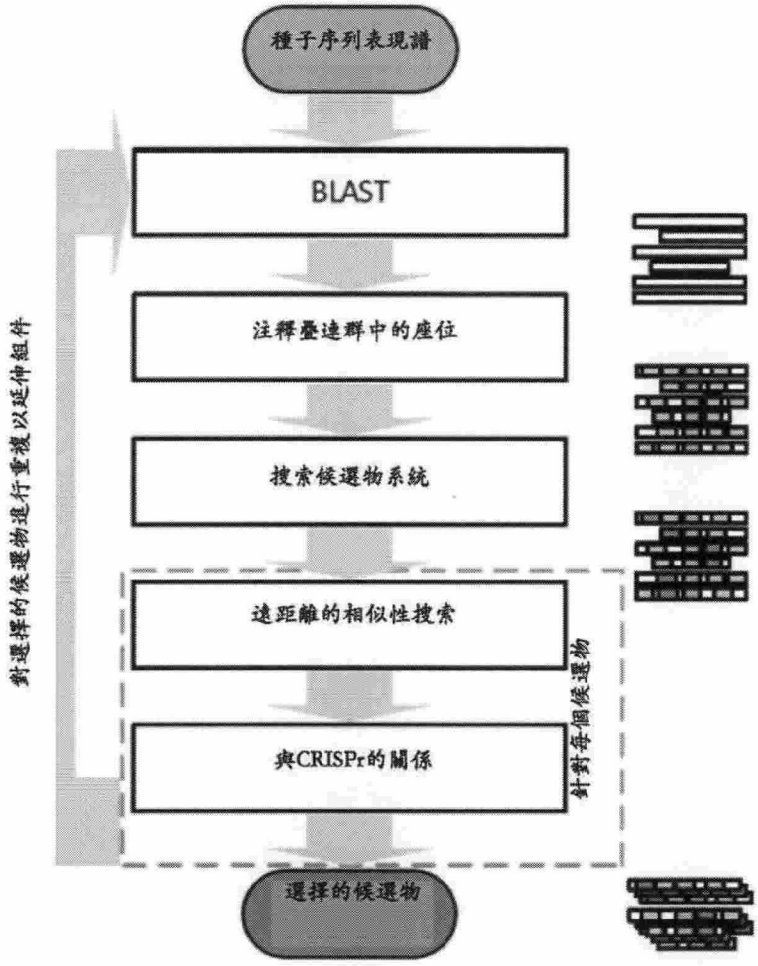

種子序列表現譜
BLAST
注釋疊連群中的座位
搜索候選物系統
遠距離的相似性搜索
與CRISPr的關係
針對每個候選物
選擇的候選物
針對選擇的候選物進行重複以延伸組件

圖 7

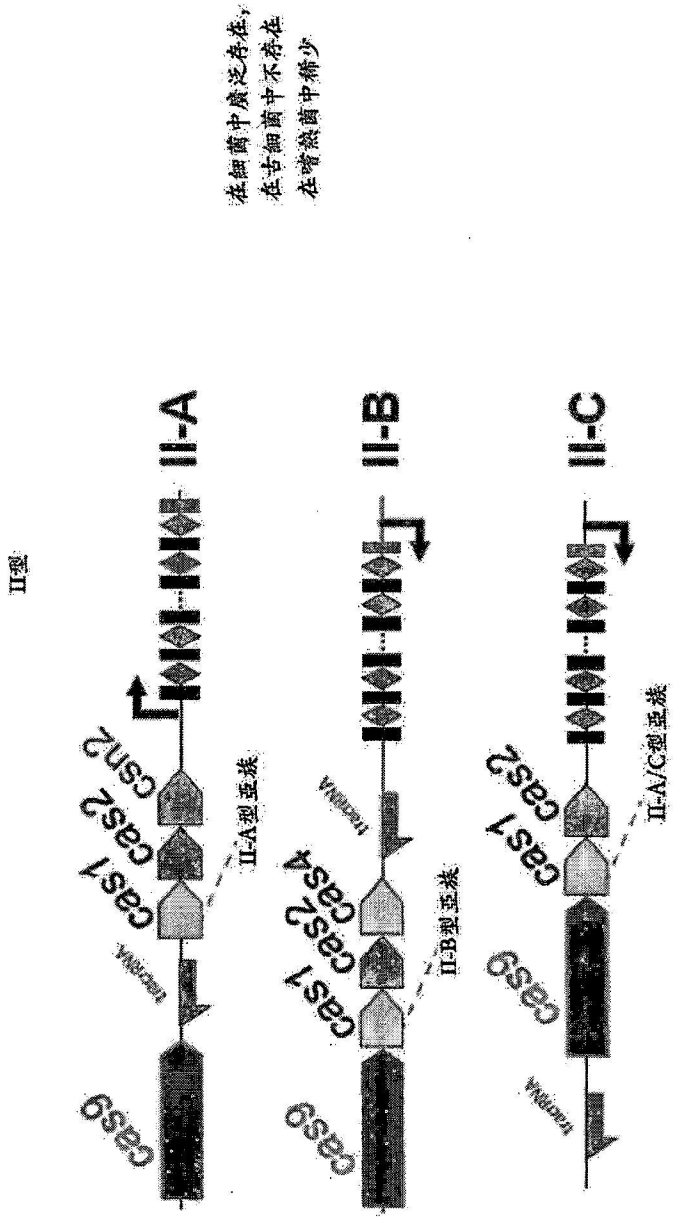
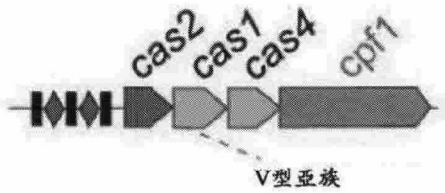

II型
II-A
II-B
II-C
cas9
cas1
cas2
csn2
tracrRNA
II-A型亞族
II-B型亞族
II-A/C型亞族
在細菌中廣泛存在，
在古細菌中不存在
在嗜熱菌中稀少

圖 8A



PreFran/型V-A?亞型

alvus甲烷嗜甲基菌Mx1201
犬口腔卟啉單胞菌
狗口腔卟啉單胞菌
獼猴卟啉單胞菌
布氏普雷沃菌
解糖胨普雷沃菌
口腔類擬桿菌Z74
孔兹刻傳球菌
胺基酸球菌屬某種BV3L6
philomiragia弗朗西絲氏菌
弗朗西絲菌屬cf.新殺手Fx1
土拉弗朗西絲菌U112
福田鉤端螺旋體
真細菌某種CAG76
挑剔真細菌ATCC 27750

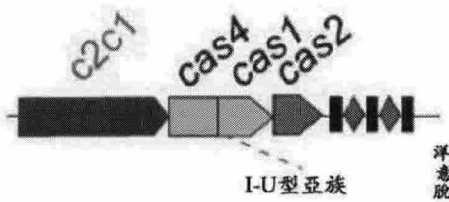
廣古菌門
擬桿菌門
擬桿菌門
擬桿菌門
擬桿菌門
擬桿菌門
擬桿菌門
擬桿菌門
梭菌目
Negativicutes
γ變形菌綱(Gamma)
γ變形菌綱
γ變形菌綱
螺旋體門
真桿菌科
真桿菌科

與cas1沒有聯繫

解脫丁酸弧菌B316
毛螺旋菌科細菌CDB1
嗜蝕黃桿菌FL-15

毛螺旋菌科
毛螺旋菌科
擬桿菌門

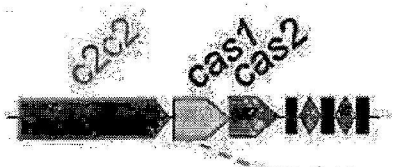
圖 8B



脂環酸芽孢桿菌/型V-B?亞型

酸土脂環酸芽孢桿菌ATCC 49025	芽孢桿菌綱(Bacilli)
洋蔥脂環酸芽孢桿菌(Alicyclobacillus contaminans)DSM 17975	芽孢桿菌綱
意外脫硫弧菌(Desulfovibrio inopinatus)DSM 10711	δ變形菌綱(Delta)
脫硫鹽鹼桿菌(Desulfonatronum thiodismutans)菌株MLF-1	δ變形菌綱
豐祐菌科細菌TAVS	疣微菌門
熱生腫塊芽胞桿菌(Tuberibacillus calidus)DSM 17572	芽孢桿菌綱
嗜熱澱粉芽孢桿菌菌株B4166	芽孢桿菌綱
短芽孢桿菌某種CF 112 PM108_疊連群_61.61	芽孢桿菌綱
芽孢桿菌屬某種NSP2.1	芽孢桿菌綱
Desulfatirhabdium butyrativorans DSM 18734	δ變形菌綱
土壤短芽孢桿菌(Brevibacillus agri)BAB-2500	芽孢桿菌綱
短芽孢桿菌屬某種CF112	芽孢桿菌綱
結瘤甲基桿菌(Methylobacterium nodutans)OPS 2060	α變形菌綱(Alpha)

圖 8C



II-A/C型亞族

纖毛菌屬/型VI亞型

毛螺旋菌科_細菌NK4A179	梭菌綱
Clostridiaridium某種	梭菌綱
雞肉桿菌	芽孢桿菌綱
propionicigene帕魯迪菌	擬桿菌門
斯氏李斯特菌	芽孢桿菌綱
韋德泰凡李斯特菌	芽孢桿菌綱
紐約恩施李斯特菌(Listeria_newyorkensis)	芽孢桿菌綱
頰纖毛菌	梭桿菌綱
纖毛菌屬某種口腔類_225	梭桿菌綱
韋德纖毛菌	梭桿菌綱
纖毛菌屬某種	梭桿菌綱
沙氏纖毛菌	梭桿菌綱

圖 8D

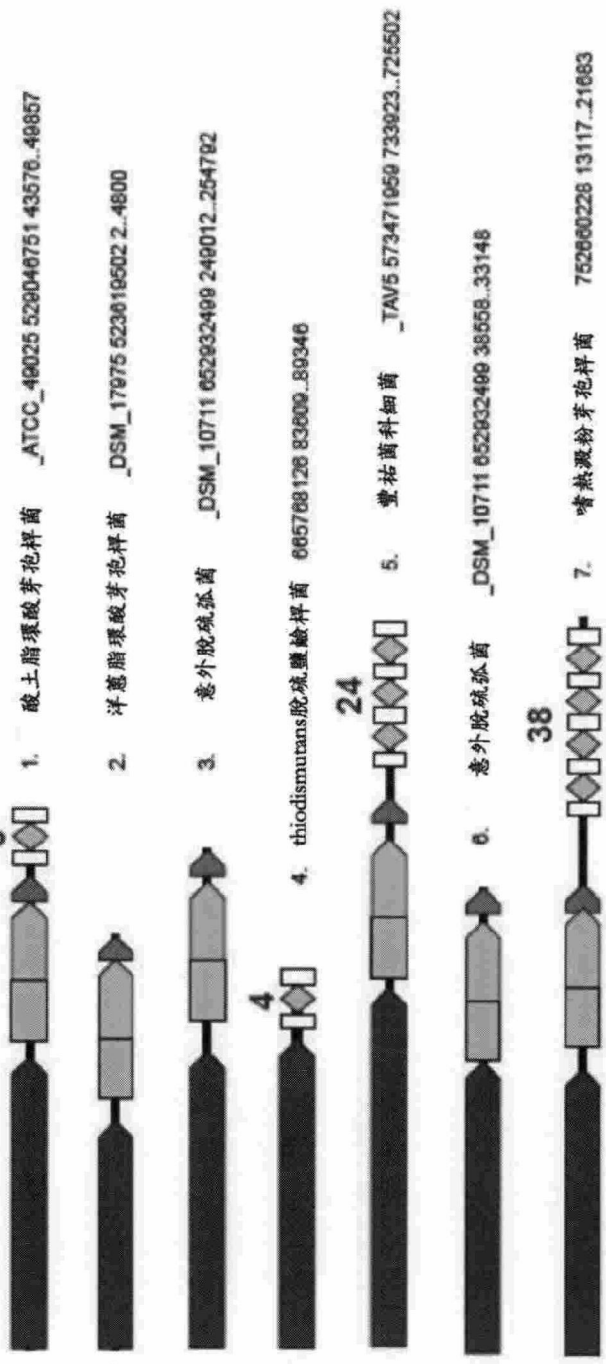

6
1. 酸土脂環酸芽孢桿菌 _ATCC_49025 5290467514 3576..49857
2. 洋蔥脂環酸芽孢桿菌 _DSM_17975 5236195022 2..4800
3. 意外脫硫弧菌 _DSM_10711 6529324001 249012..254792
4
4. thiodismutans脫硫鹽鹼桿菌 6657681261 83609..89346
24
5. 豐祐菌科細菌 _TAV5 5734719501 733923..725502
38
6. 意外脫硫弧菌 _DSM_10711 6529324001 38558..33148
7. 嗜熱源粉芽孢桿菌 7526602281 13117..21683
Cas4/Cas1
Cas2
C2c1
其他非Cas基因
CRISPR陣列

圖 9A

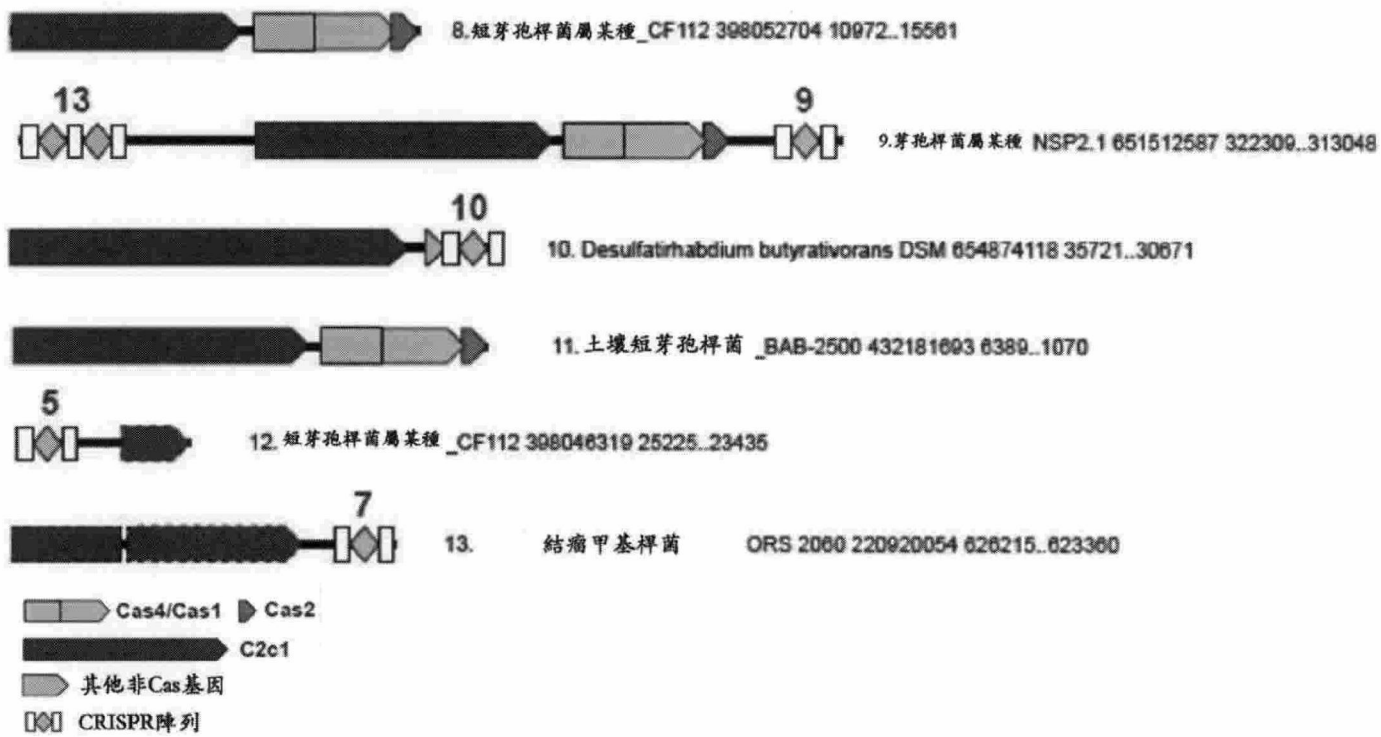

8.短芽孢桿菌屬某種_CF112 398052704 10972..15561
13
9
9.芽孢桿菌屬某種 NSP2.1 651512587 322309..313048
10
10. Desulfatirhabdium butyrativorans DSM 654874118 35721..30671
11. 土壤短芽孢桿菌 _BAB-2500 432181693 6389..1070
5
12. 短芽孢桿菌屬某種 _CF112 398046319 25225..23435
7
13. 結瘤甲基桿菌 ORS 2060 220920054 626215..623360
Cas4/Cas1
Cas2
C2c1
其他非Cas基因
CRISPR陣列

圖 9B

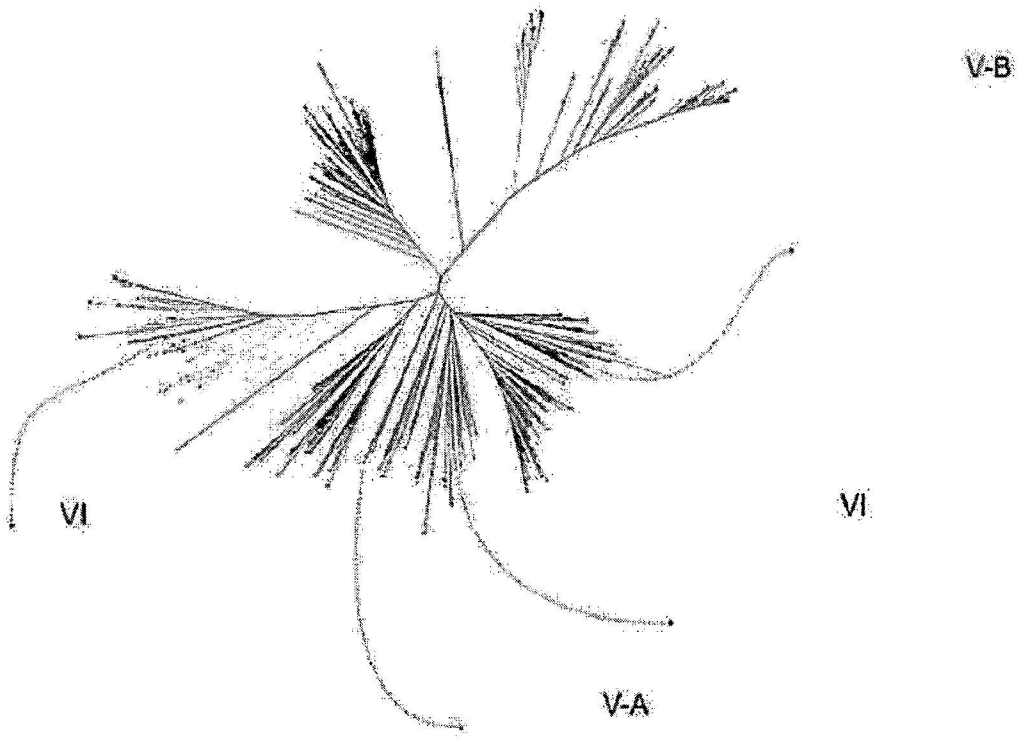


圖 10A

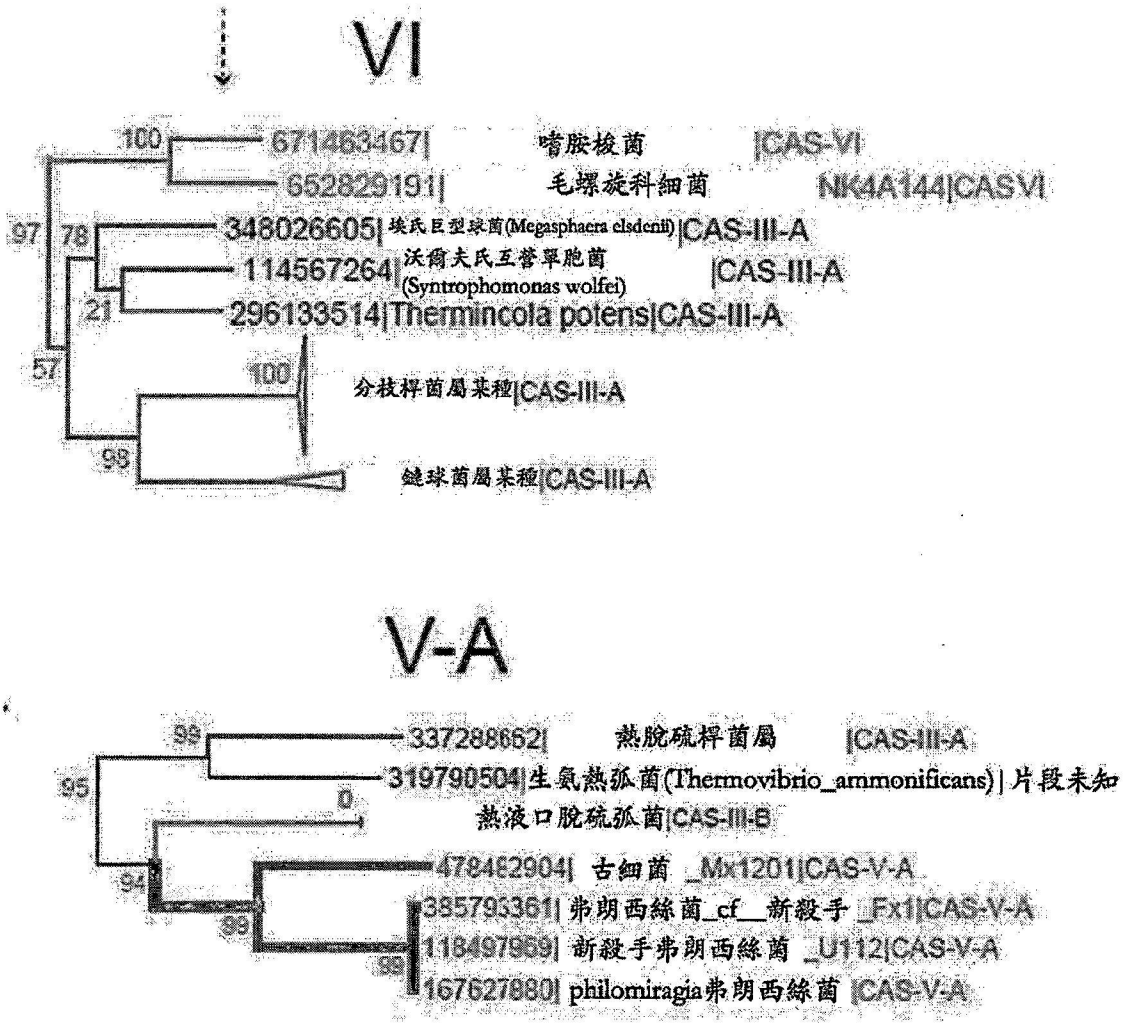

VI
100
671463467| 嗜胺梭菌 |CAS-VI
652829191| 毛螺旋科細菌 NK4A144|CASVI
97
78
348026605| 埃氏巨型球菌(Megasphaera elsdenii) |CAS-III-A
114567264| 沃爾夫氏互營單胞菌 (Syntrophomonas wolfei) |CAS-III-A
21
296133514|Thermincola potens|CAS-III-A
57
100
分枝桿菌屬某種|CAS-III-A
99
鏈球菌屬某種|CAS-III-A
V-A
99
337288662| 熱脫硫桿菌屬 |CAS-III-A
95
319790504|生氨熱弧菌(Thermovibrio_ammonificans)| 片段未知
0
熱液口脫硫弧菌|CAS-III-B
94
478462904| 古細菌 _Mx1201|CAS-V-A
99
385793361| 弗朗西絲菌_cf__新殺手 _Fx1|CAS-V-A
99
118497959| 新殺手弗朗西絲菌 _U112|CAS-V-A
167627980| philomiragia弗朗西絲菌 |CAS-V-A

圖 10B

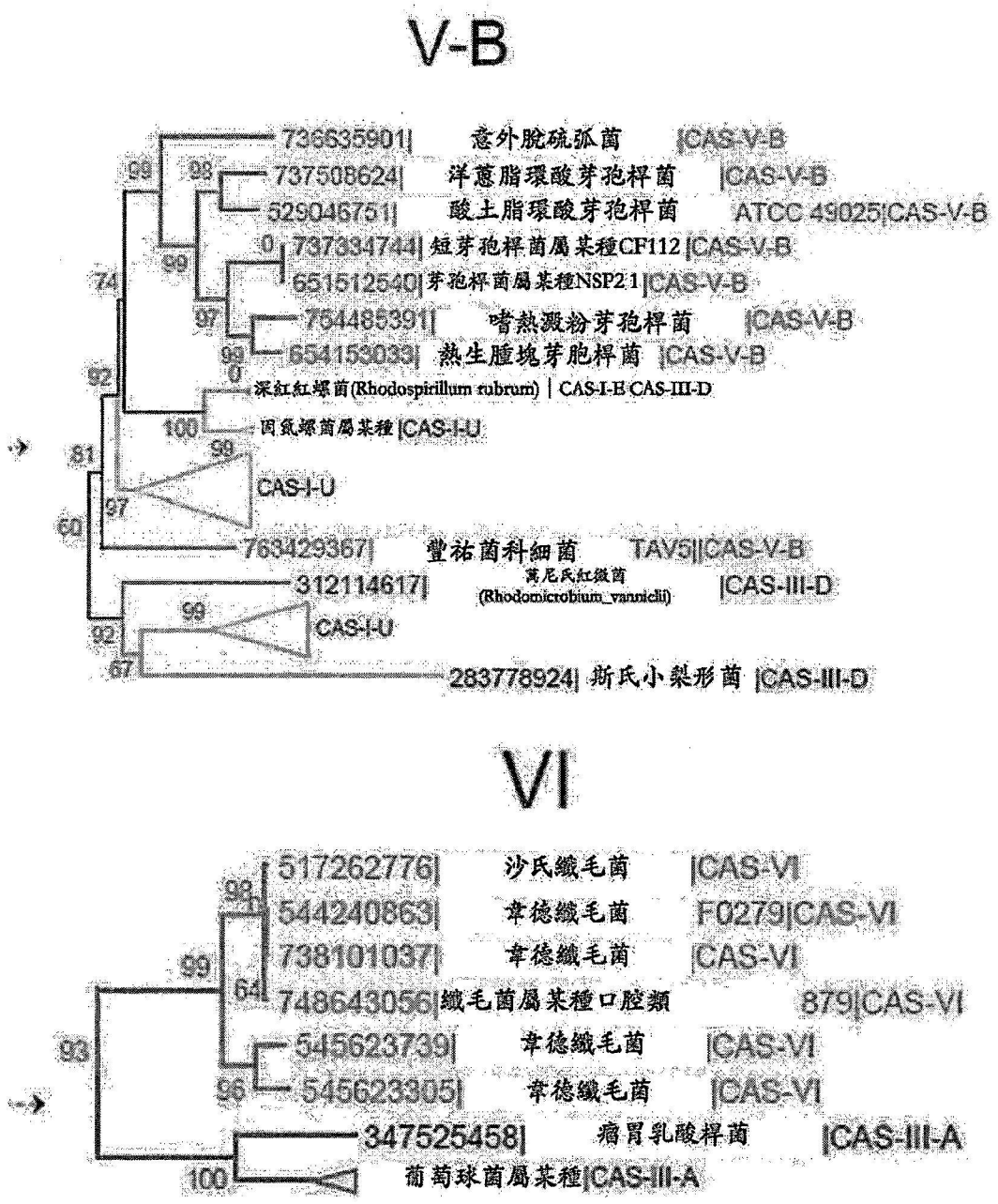

V-B
736635901| 意外脫硫弧菌 |CAS-V-B
737508624| 洋蔥脂環酸芽孢桿菌 |CAS-V-B
529046751| 酸土脂環酸芽孢桿菌 ATCC 49025|CAS-V-B
737334744| 短芽孢桿菌屬菜種CF112 |CAS-V-B
651512540| 芽孢桿菌屬菜種NSP2.1 |CAS-V-B
754485391| 嗜熱澱粉芽孢桿菌 |CAS-V-B
654153033| 熱生膿塊芽胞桿菌 |CAS-V-B
深紅紅螺菌(Rhodospirillum rubrum) | CAS-I-E CAS-III-D
固氮螺菌屬菜種 |CAS-I-U
CAS-I-U
763429367| 豐祐菌科細菌 TAV5||CAS-V-B
312114617| 萬尼氏紅微菌 (Rhodomicrobium_vannielii) |CAS-III-D
CAS-I-U
283778924| 斯氏小梨形菌 |CAS-III-D
VI
517262776| 沙氏纖毛菌 |CAS-VI
544240863| 韋德纖毛菌 F0279|CAS-VI
738101037| 韋德纖毛菌 |CAS-VI
748643056|纖毛菌屬菜種口腔類 879|CAS-VI
545623739| 韋德纖毛菌 |CAS-VI
545623305| 韋德纖毛菌 |CAS-VI
347525458| 瘤胃乳酸桿菌 |CAS-III-A
葡萄球菌屬菜種|CAS-III-A

圖 10C

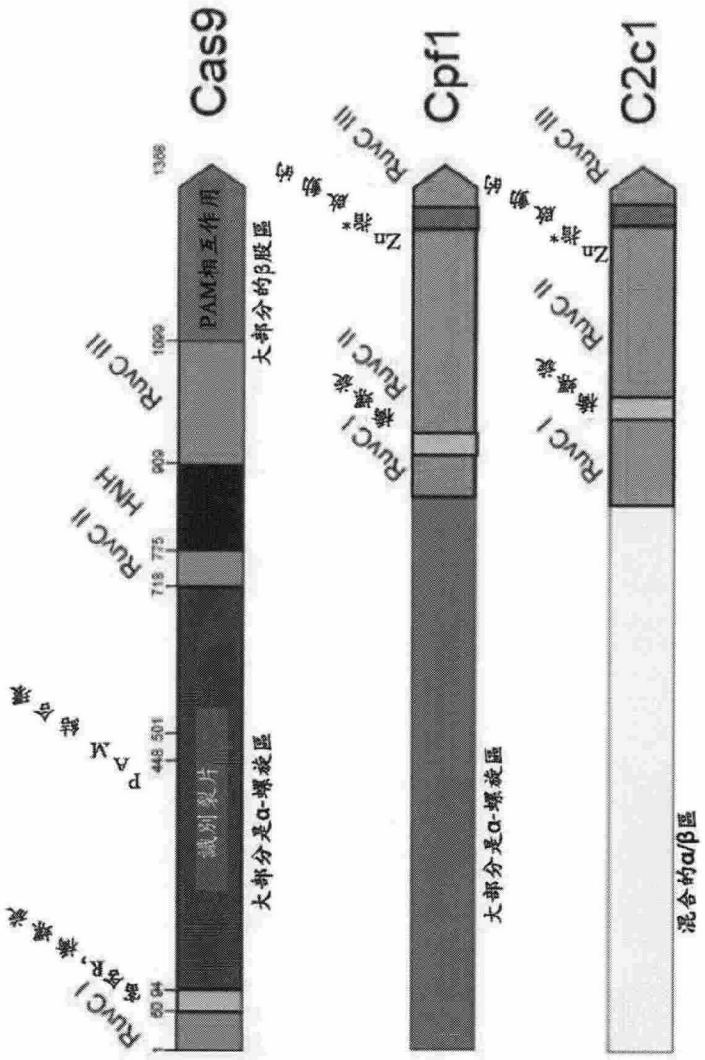

Cas9
Cpf1
C2c1
RuvC I
RuvC II
HNH
RuvC III
PAM
PAM相互作用
大部分是α-螺旋區
大部分的β股區
大部分是α-螺旋區
混合的α/β區
1
60
94
448
501
718
775
909
1099
1368

圖 11A

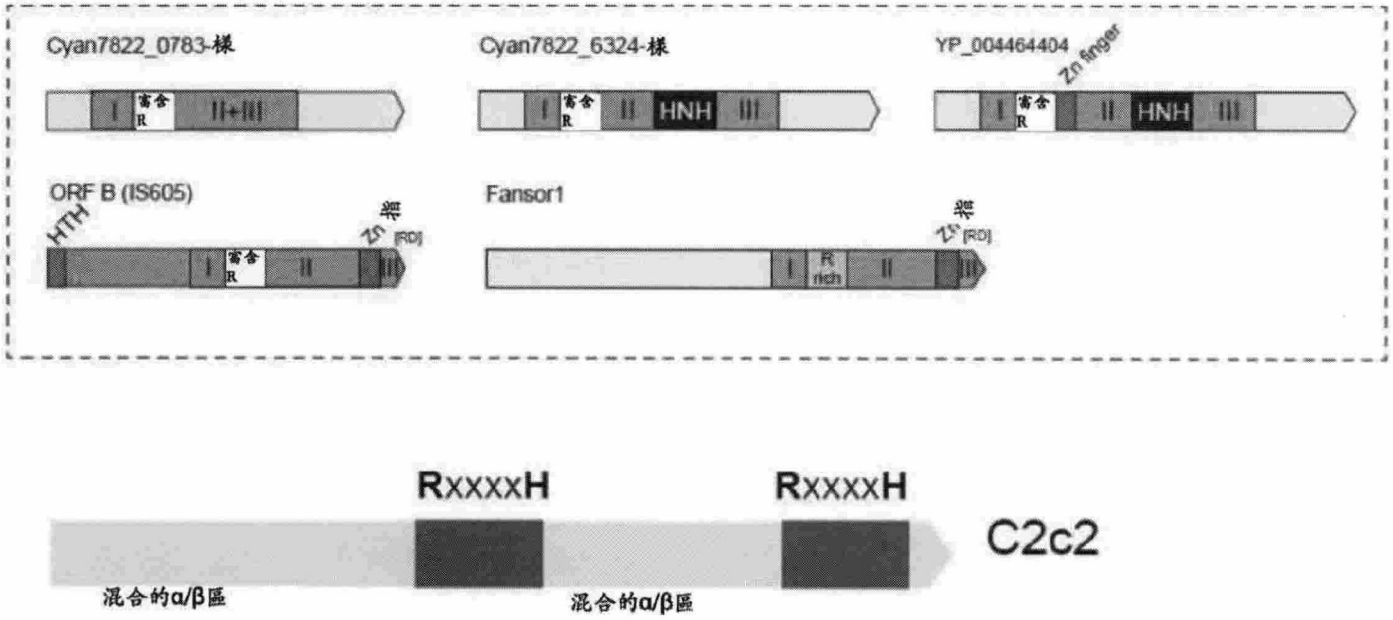

Cyan7822_0783-樣
I
富含R
II+III
Cyan7822_6324-樣
I
富含R
II
HNH
III
YP_004464404
Zn finger
I
富含R
II
HNH
III
ORF B (IS605)
HTH
I
富含R
II
Zn 指 [RD]
III
Fansor1
I
R rich
II
Zn 指 [RD]
III
RxxxxH
RxxxxH
C2c2
混合的α/β區
混合的α/β區

圖 11B

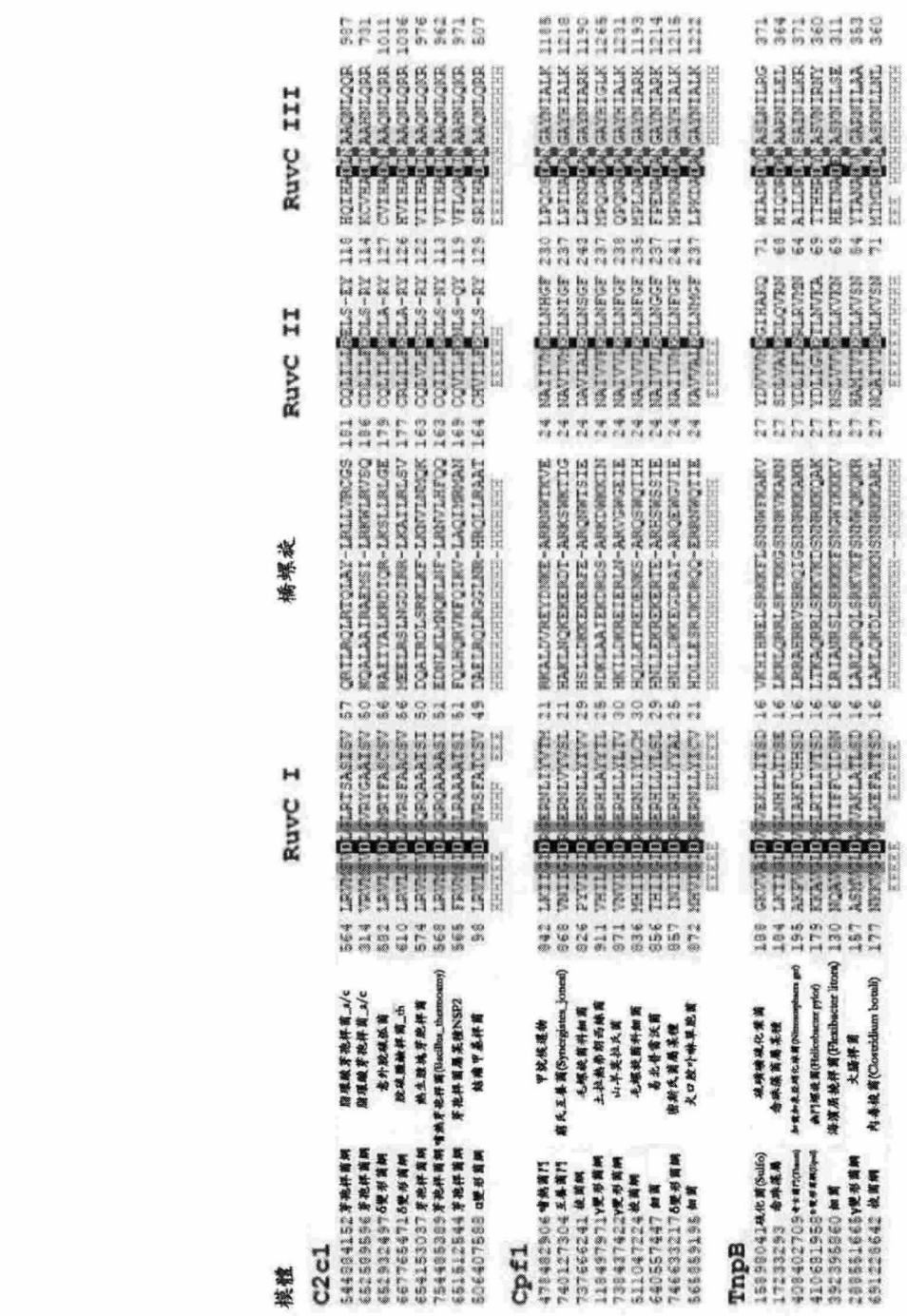


圖 12A

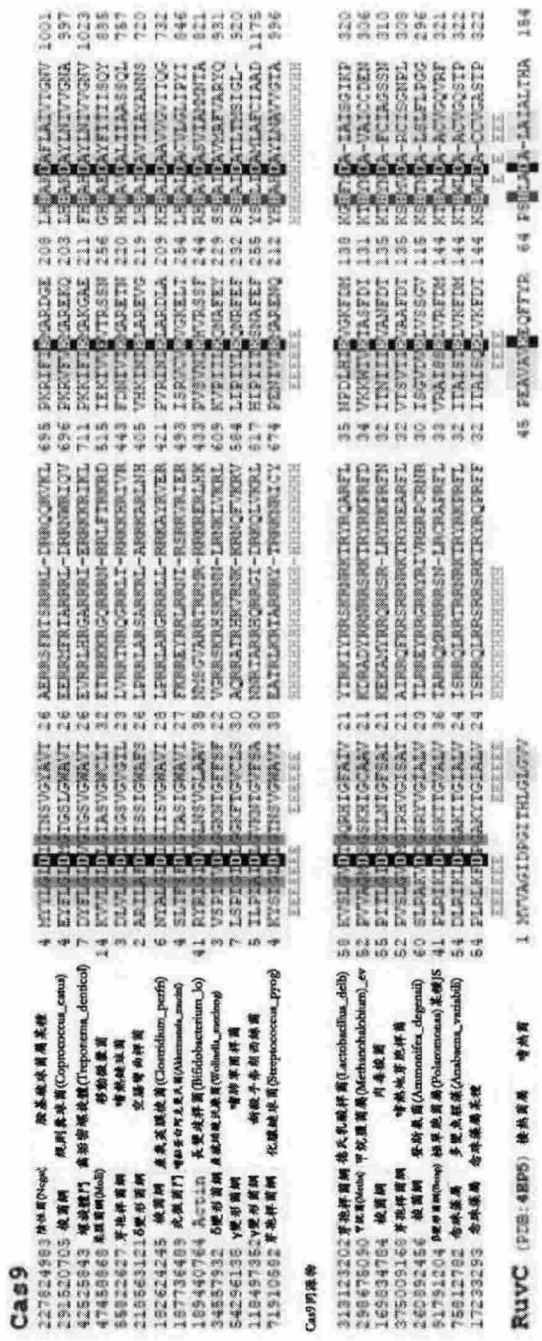


圖 12B


1000 nt
Cas1
Cas2
C2c2
其他非Cas基因
CRISPR陣列
10
RT
1. 毛螺旋科細菌 _MA2020 675284558 94827..87964
3
2. 毛螺旋科細菌 NK4A179 551041829 10913..5842
9
ST-PknB
3. 嗜胺梭菌 _DSM_10710 658749509 1521..14194
6
4. 毛螺旋科細菌 _NK4A144 523401370 11419..17600
9
5
5. 雞肉桿菌DSM _4847 675281379 168298..162241
3
7
6. 雞肉桿菌DSM _4847 675281379 2459220..2464096
26
7. propionicigene帕魯迪菌 _WB4 313202490 223908..229213
5
8. 血清型斯氏李斯特菌_12b菌株 289433373 1174057..1177790

圖 13A

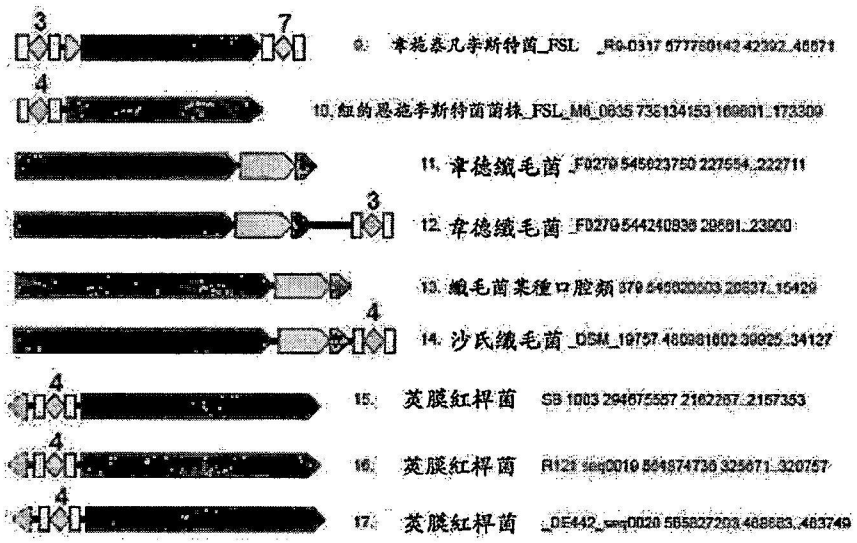

3
7
9. 韋施泰凡李斯特菌_FSL _R9-0317 577760142 42392..46571
4
10. 紐約恩施李斯特菌菌株_FSL_M6_0635 736134153 169601..173309
11. 韋德纖毛菌 _F0279 545623750 227554..222711
3
12. 韋德纖毛菌 _F0279 544240836 29661..23900
13. 纖毛菌某種口腔類 379 545620503 20637..15429
4
14. 沙氏纖毛菌 _DSM_19757 480981602 39925..34127
4
15. 英膜紅桿菌 SB 1003 294675557 2162267..2157353
4
16. 英膜紅桿菌 R121 seq0019 564874736 325671..320757
4
17. 英膜紅桿菌 _DE442_seq0020 565827203 468683..463749

圖 13B

推斷的二級結構

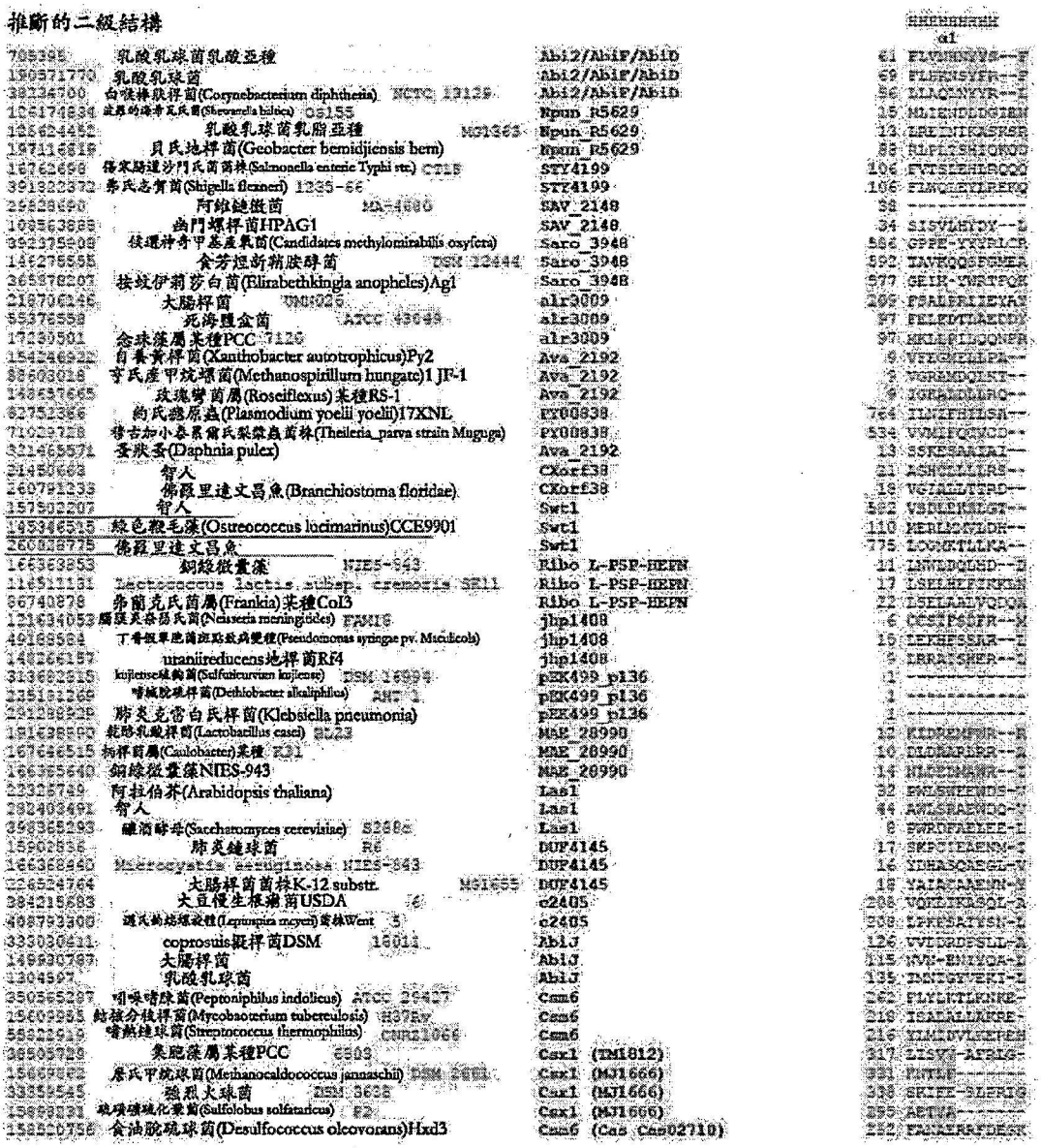


圖 14A

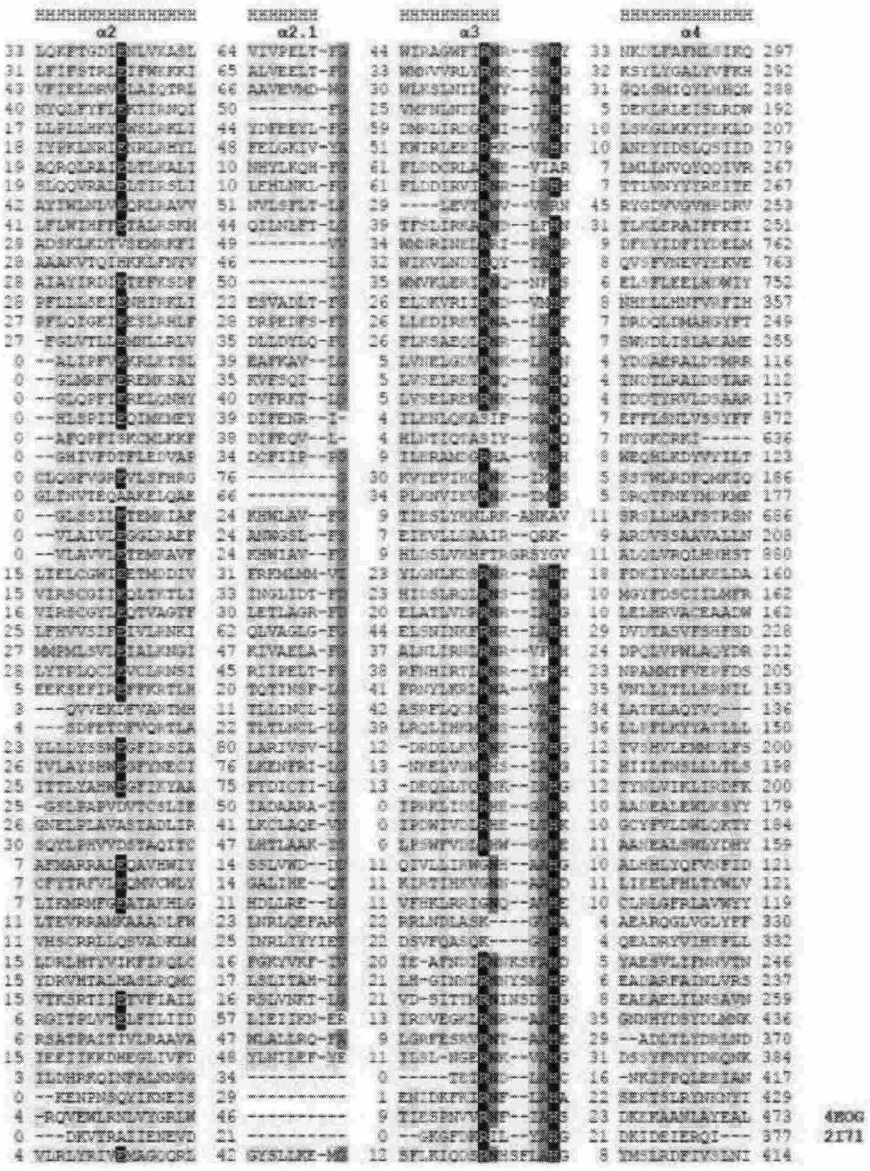
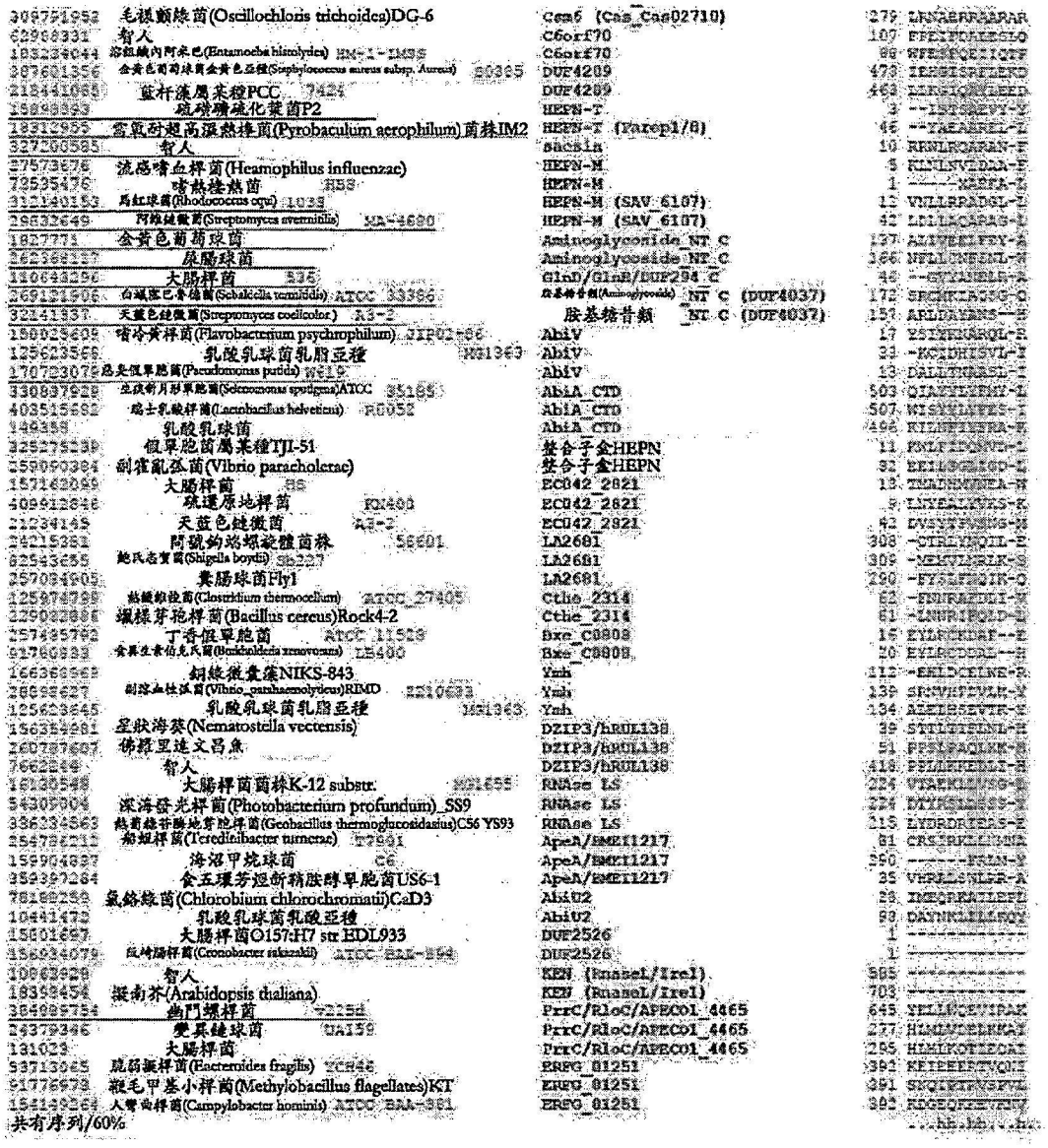


圖 14B


309791952 毛狀顫綠菌(Oscillochloris trichoides)DG-6
62966331 智人
183234044 溶組織內阿米巴(Entamoeba histolytica) HM-1:IMSS
387601356 金黃色葡萄球菌金黃色亞種(Staphylococcus aureus subsp. Aureus) 80385
218441065 藍杆藻屬菜種PCC 7424
18898893 硫磺礦硫化葉菌P2
18312955 需氧耐超高溫熱棒菌(Pyrobaculum aerophilum)菌株IM2
327200585 智人
27573676 流感嗜血桿菌(Heamophilus influenzae)
73595476 嗜熱栖熱菌 HB8
312140153 馬紅球菌(Rhodococcus equ) 103S
29832649 阿維鏈黴菌(Streptomyces avermitilis) MA-4680
1827771 金黃色葡萄球菌
262366137 屎腸球菌
110648296 大腸桿菌 536
269121906 白蟻腔已食繩菌(Schalcella teruiddis) ATCC 33386
32141837 天藍色鏈黴菌(Streptomyces coelicolor) A3-2
150029609 嗜冷黃桿菌(Flavobacterium psychrophilum) JIP02-86
125623568 乳酸乳球菌乳脂亞種 MG1363
170723078 惡臭假單胞菌(Pseudomonas putida) W619
330897928 亞淡新月狀單胞菌(Selenomonas sputigena)ATCC 35185
408515692 瑞士乳酸桿菌(Lactobacillus helveticus) R0052
148388 乳酸乳球菌
325275239 假單胞菌屬菜種TJI-51
259090384 副霍亂弧菌(Vibrio paracholerae)
157162099 大腸桿菌 HS
409912846 硫還原地桿菌 RM400
21234145 天藍色鏈黴菌 A3-2
24215381 閃號約翰螺旋體菌株 58601
82543655 鮑氏志賀菌(Shigella boydii) Sb227
257094905 糞腸球菌Fly1
125974795 熱纖梭菌(Clostridium thermocellum) ATCC 27405
229032884 蠟樣芽孢桿菌(Bacillus cereus)Rock4-2
257495792 丁香假單胞菌 ATCC 11528
91780833 食果生食伯克氏菌(Burkholderia xenovorans) LB400
166368962 銅綠微囊藻NIKS-843
28899627 副溶血性弧菌(Vibrio_parahaemolyticus)RIMD 2210633
125623645 乳酸乳球菌乳脂亞種 MG1363
156354981 星狀海葵(Nematostella vectensis)
260787607 佛羅里達文昌魚
7662249 智人
16120548 大腸桿菌菌株K-12 substr. MG1655
54309904 深海發光桿菌(Photobacterium profundum)_SS9
386334563 熱葡糖苷酶地芽胞桿菌(Geobacillus thermoglucosidasius)C56 YS93
254782212 船蛆桿菌(Teredinibacter turnerae) T7901
159904897 海沼甲烷球菌 C6
359397284 食五環芳烴新鞘胺醇單胞菌US6-1
78180259 氯銘綠菌(Chlorobium chlorochromatii)CaD3
10441472 乳酸乳球菌乳酸亞種
15801697 大腸桿菌O157:H7 str EDL933
156934079 丘神腸桿菌(Cronobacter sakazakii) ATCC BAA-894
10968928 智人
18398454 擬南芥(Arabidopsis thaliana)
384085754 幽門螺桿菌 v225d
24379346 變異鏈球菌 UA159
131023 大腸桿菌
93713045 脆弱擬桿菌(Bacteroides fragilis) YCH46
91776973 鞭毛甲基小桿菌(Methylobacillus flagellates)KT
154149264 人彎曲桿菌(Campylobacter hominis) ATCC BAA-381
共有序列/60%

Csm6 (Cas_Cas02710)
C6orf70
C6orf70
DUF4209
DUF4209
HEPN-T
HEPN-T (Parep1/8)
Sacsin
HEPN-M
HEPN-M
HEPN-M (SAV_6107)
HEPN-M (SAV_6107)
Aminoglycoside_NT_C
Aminoglycoside_NT_C
GlnD/GlnB/DUF294_C
胺基糖苷類(Aminoglycoside) NT_C (DUF4037)
胺基糖苷類 NT_C (DUF4037)
AbiV
AbiV
AbiV
AbiA_CTD
AbiA_CTD
AbiA_CTD
整合子金HEPN
整合子金HEPN
EC042_2821
EC042_2821
EC042_2821
LA2681
LA2681
LA2681
Cthe_2314
Cthe_2314
Bxe_C0808
Bxe_C0808
Ymh
Ymh
Ymh
DZIP3/hRUL138
DZIP3/hRUL138
DZIP3/hRUL138
RNAse LS
RNAse LS
RNAse LS
ApeA/BME11217
ApeA/BME11217
ApeA/BME11217
AbiU2
AbiU2
DUF2526
DUF2526
KEN (RnaseL/Ire1)
KEN (RnaseL/Ire1)
PrrC/RloC/APECO1_4465
PrrC/RloC/APECO1_4465
PrrC/RloC/APECO1_4465
ERFG_01251
ERFG_01251
ERFG_01251

圖 14C

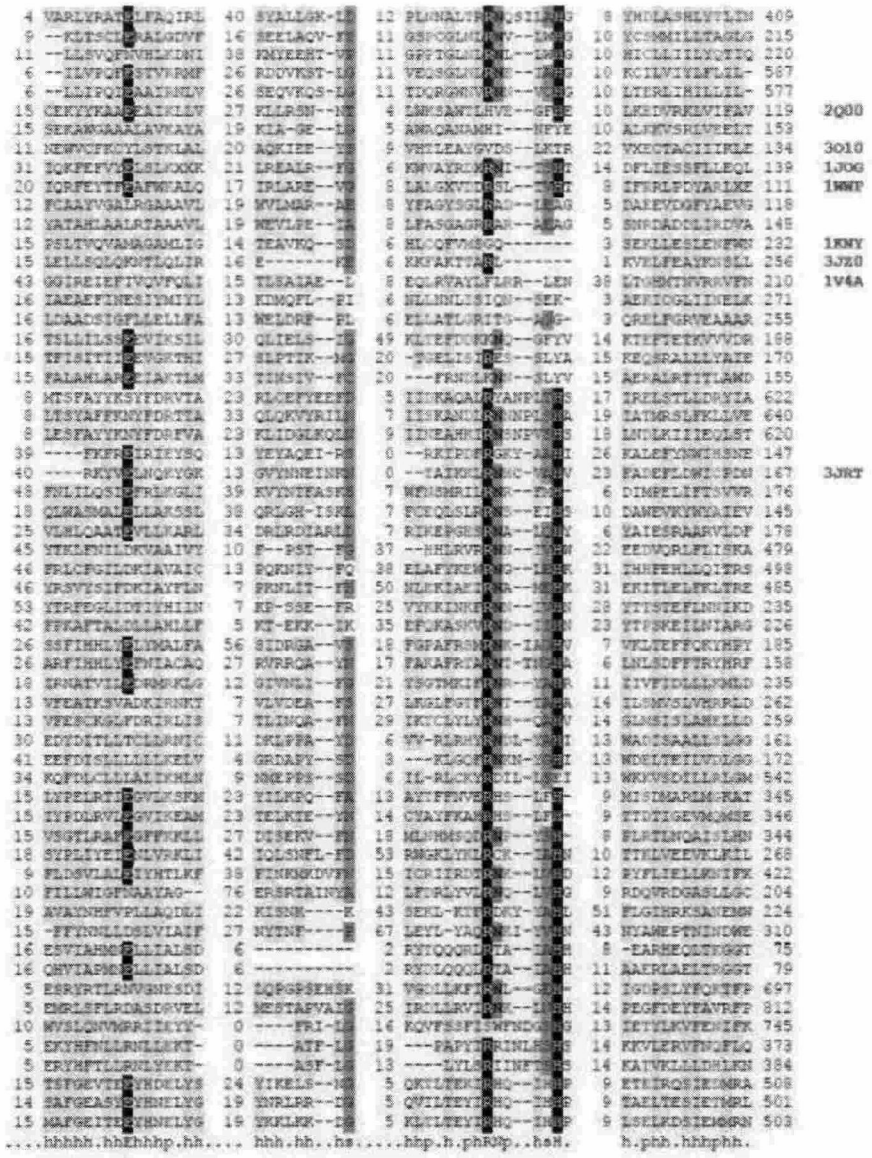


圖 14D

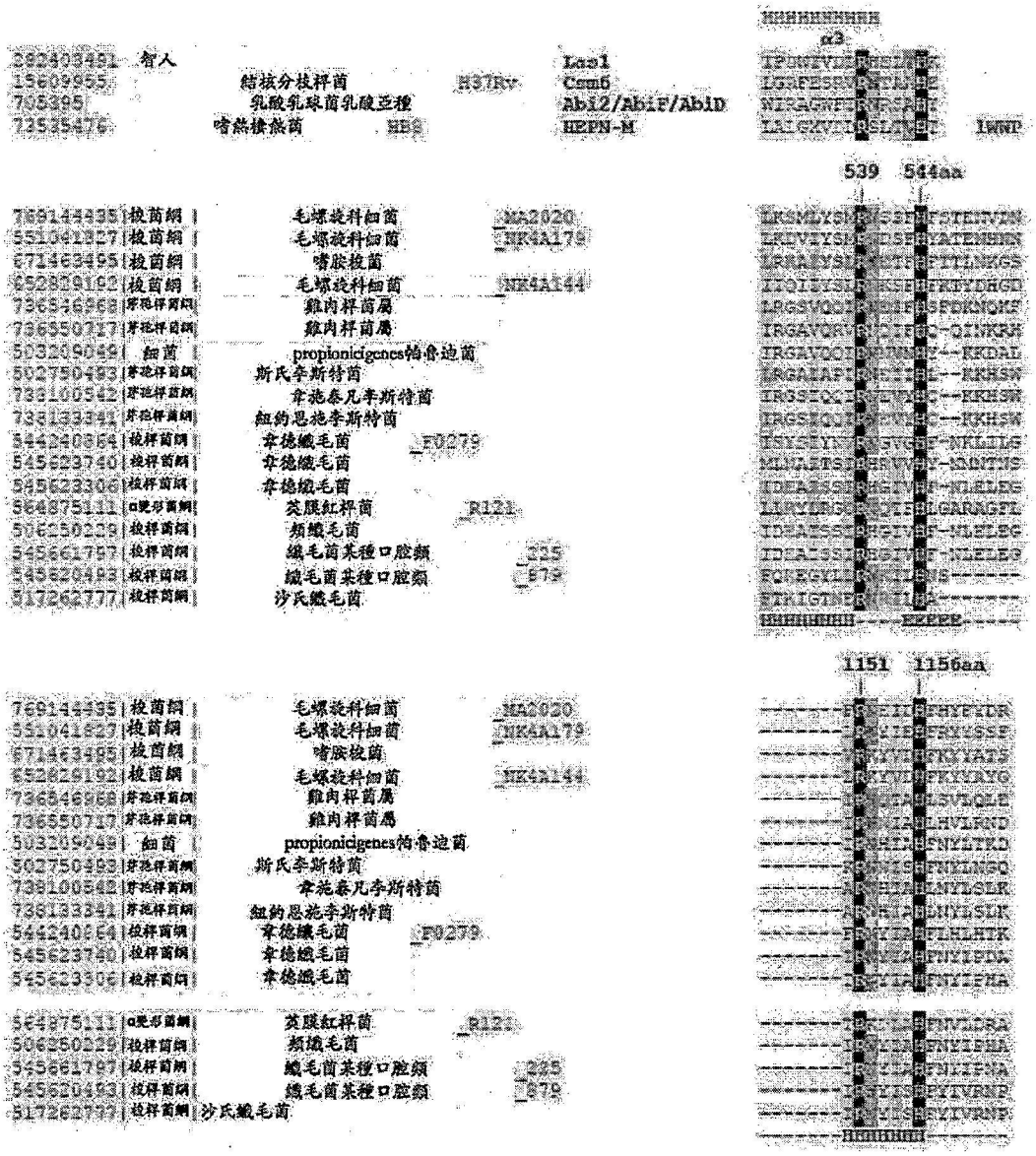

智人
結核分枝桿菌
H37Rv
Las1
Csm6
Abi2/AbiF/AbiD
HEPN-M
乳酸乳球菌乳酸亞種
嗜熱棲熱菌
HB8
α3
539
544aa
1151
1156aa
毛螺旋科細菌
MA2020
NK4A179
嗜胺梭菌
毛螺旋科細菌
NK4A144
雞肉桿菌屬
雞肉桿菌屬
propionicigenes帕魯迪菌
斯氏李斯特菌
韋施泰凡李斯特菌
紐約思施李斯特菌
韋德纖毛菌
F0279
韋德纖毛菌
韋德纖毛菌
英膜紅桿菌
R121
頰纖毛菌
纖毛菌某種口腔類
225
纖毛菌某種口腔類
879
沙氏纖毛菌

圖 14E

DR: GUGCCACUUCUCAGAUCGCUCGCUCAGUGAUCCGAC

tracrRNA:
GUCAGAACACUGAGCGAGCGUUCUUUUUGAGAAGCUCAACGGGCUUUGCCACCUGGAA
AGUGGCCAUUGGCACACCCGUUGAAAAAAUUCUGUCCUCUAGACAGA

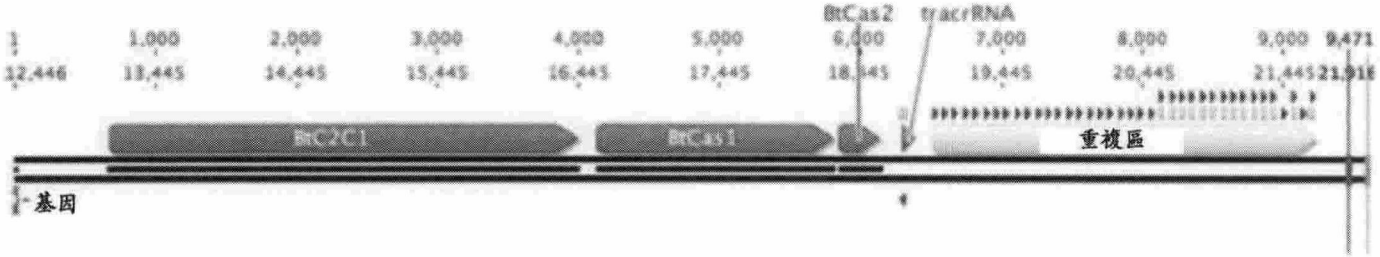
圖 15

重複區
tracrRNA
基因
DR: GUGCCAAUCACACCCAACACUGACCAAGCUUGCCGAGAC
tracrRNA:
CUUGGGGAAAGCUAGGCAAGUUUUGGAUGAUAAGAAAUAAUCAUGUCACAAGGAGGGA
GUUUUU

圖 16

1
735,109
1,000
734,110
2,000
733,110
3,000
732,110
4,000
731,110
5,000
730,110
6,000
729,110
7,000
728,110
8,048
727062
9,000
726,110
9,943
725,167
基因
ObCas1
ObCas4
ObCas2
gene
tracrRNA
重複區
基因
DR: GCCGCAGCGAAUGCCGUUUCACGAAUCGUCAGGCGG
tracrRNA:
GCUGGAGACGUUUUUUGAAACGGCGAGUGCUGCGGAUAGCGAGUUUCUCUUGGGGAGG
CGCUCGCGGCCACUUUU

圖 17



DR: GUCCAAGAAAAAAGAAAUGAUACGAGGCAUUAGCAC

tracrRNA:
CUGGACGAUGUCUCUUUUAUUUCUUUUUUCUUGGAUCUGAGUACGAGCACCCACAUUG
GACAUUUCGCAUGGUGGGUGCUCGUACUAUAGGUAAAACAAACCUUUUU

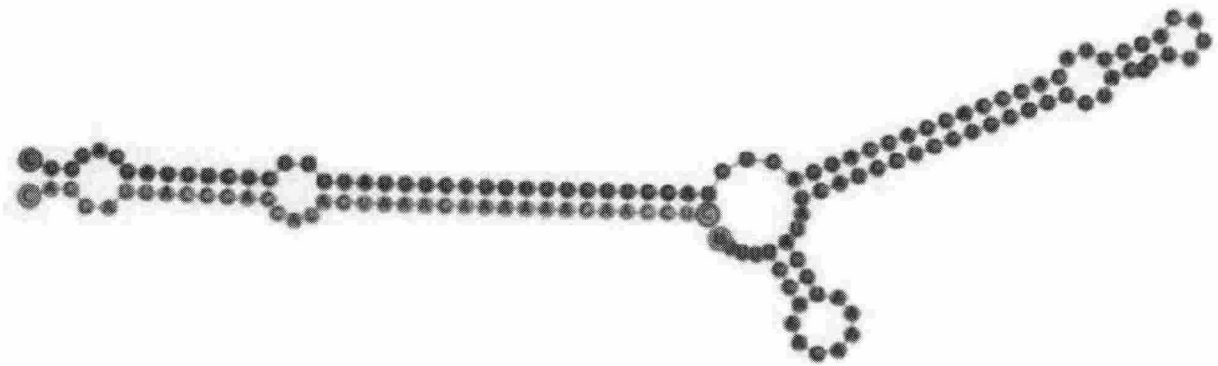
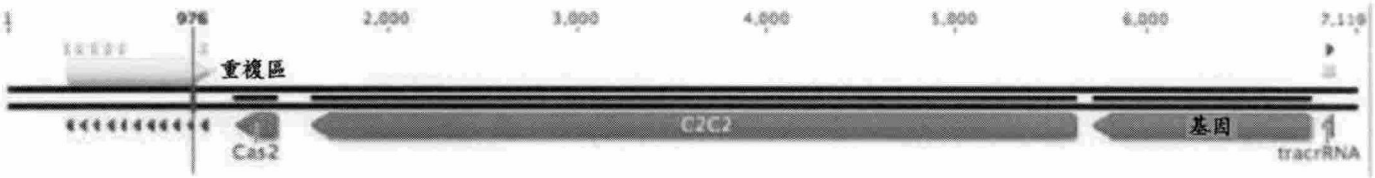


圖 18

DR: GUUCGAAAGCUUAGUGGAAAGCUUCGUGGUUAGCAC

tracrRNA:
CACGGAUAAUCACGACUUUCCACUAAGCUUUCGAAUUUUAUGAUGCGAGCAUCCUCUC
AGGUCAAAAAA

圖 19

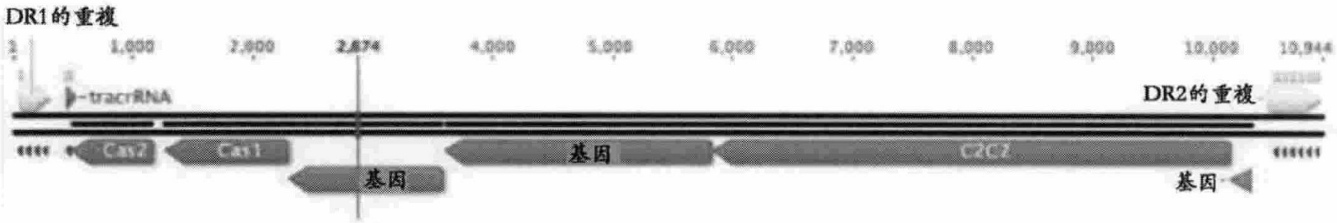


DR: GUAUUGAGAAAAGCCAGAUAUAGUUGGCAAUAGAC

tracrRNA:
AUAUUUUGAUUCCCAUUUAUGGUUAUUUACCAUAAAUGGGAAUCAACUAAAAAAUAUU
UUUU

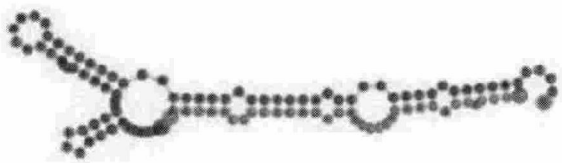


圖 20

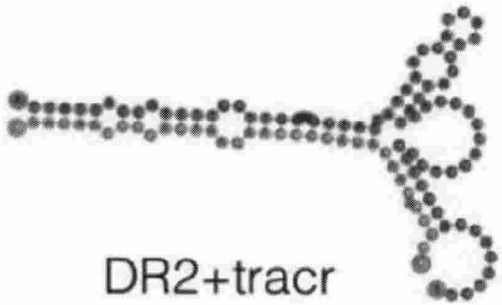


DR1: GUUGAUGAGAAGAGCCCAAGAUAGAGGGCAAUAAC
DR2: GCUGGAGAAGAUAGCCCAAGAAAGAGGGCAAUAAC

tracrRNA:
AUUAUUACCAUUUUGGUUGGAAUGCUAUUAUAAAGGAUCAUUCGAUUAUUACCUCUAC
CUCCCUUCCCACGAUUUCUU



DR1+tracr



DR2+tracr

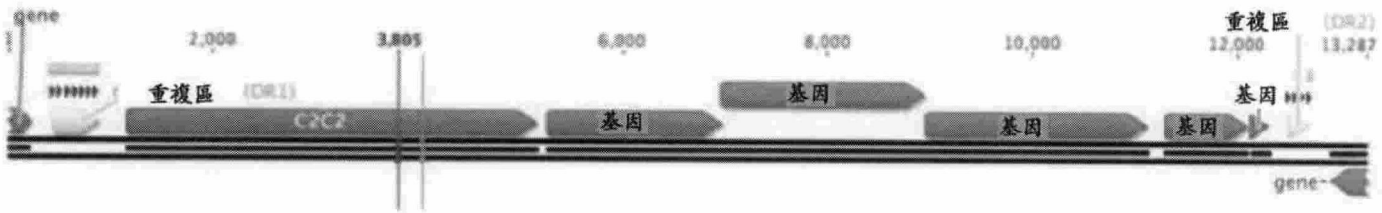
取向看起來奇怪，但是可能系統不同於II型

圖 21

DR: GUUUGGAGAACAGCCCGAUAUAGAGGGCAAUAGAC

tracrRNA:
GUCUUACGACCUCAGUAUUAGGAAGAUUUCAACCAAGAAAACUUAGUUUCAGGCUUAA
UGAUCGAGUCAUGCAGCCAAAGU

圖 22



DR1: GUUUUGAGAAUAGCCCGACAUAGAGGGCAAUAGAC
DR2: GUUAUGAAAACAGCCCGACAUAGAGGGCAAUAGACA

tracrRNA: ???

圖 23

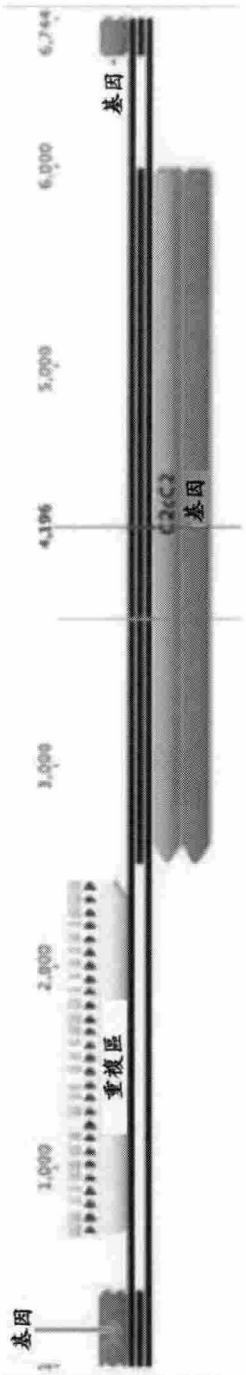
1
1,000
1,974
3,000
4,000
5,000
6,000
7,000
8,000
9,280
cas2
重複區
重複區
基因
基因
基因
基因
tracrRNA
DR: GUUAUAGUCCUCUUACAUUUAGAGGUAGUCUUUAAU
tracrRNA:
UCUUAAGAACUUCUCUACCUGAAGUUGGAUUAUAAAUGACUCUUGCUCUCAUAGAUAU
CCUCCUUUGAAAAUAUACACUGCCGAUUAAUUACCGUUUU

圖 24

DR1： GUUAUAGUCCUCUUACAUUUAGAGGUAGUUUAUAUU
DR2： GUUAUAGUCCCCUUACAUUUAGGGGUAGUCUUUAAU

tracrRNA:
AAUAUAAAUUCUCCCUAAAUAUAAGAGAAUAAUAACUCAAUCUCUUCAUUCGUAUUUU
GUCUAGUUAAGAUAAGUACCACCAAAUACAAUCAAUCCAAAAAA

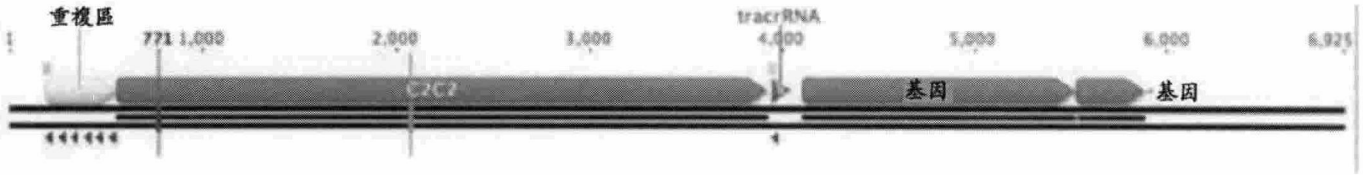
圖 25



DR: GUUGUAGUUCCCUUCAAUUUUGGGAUAAUCCACAAG

tracrRNA: ???

圖 26



DR: GUUUUAGUCCUCUUUCAUAUAGAGGUAGUCUCUUAC

tracrRNA:
AUGAAAAGAGGACUAAAACUGAAAGAGGACUAAAACACCAGAUGUGGAUAACUAUAUU
AGUGGCUAUUAAAAAUUCGUCGAUAUUAGAGAGGAAACUUU

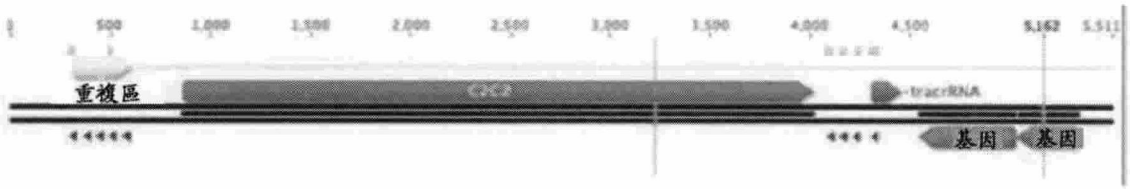


圖 27

DR: GUUUUAGACCUCUUCUAUUUUGAGGUACUCUAAAUC

tracrRNA: ???

圖 28



DR: GUUUUAGUCCUCUUUUGUUUUGAGGUACUCUAAAUC

tracrRNA:
AAGUCAGCGCACAACAAAGAAGAUGACGAACAAAAUCUCUCGCCAUCUUCUUAAAAUU
AUUUGCCACACAGCCAACAUUAUAAGCGUUAAAACCAGCACCAUGAGUACAUUUCACC
CAACAAUCAGAAUCCCCGUUUCUCCGUUUUU

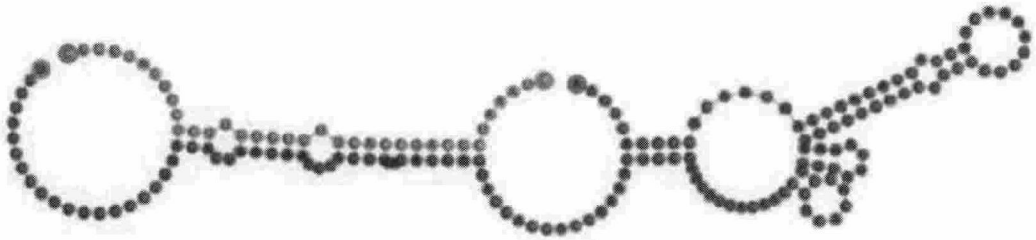
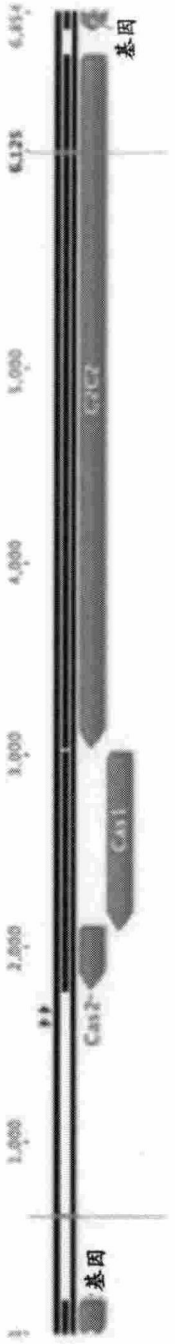


圖 29



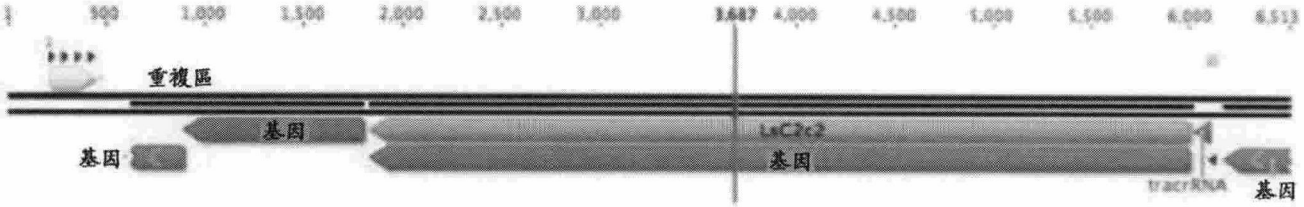
DR: GUUUUAGAUCCCUUCGUUUUUGGGGUUAUCUAUAUC

tracrRNA: ???

圖 30

組裝間隙
tracrRNA2
Cas2
Cas1
CRC2
基因
tracrRNA
基因
DR: GUUUUAGUCCCCUUCGUUUUUGGGGUAGUCUAAAUC
tracrRNA1:
GAUUUAGAGCACCCCAAAAGUAAUGAAAAUUUGCAAUUAAAUAAGGAAUAUUAAAAAA
AUGUGAUUUUAAAAAAAAUUGAAGAAAUUAAAAUGAAAAAAUUGUCCAAGUAAAAAAAA
tracrRNA2:
AUUUAGAUUACCCCUUUAAUUUAUUUUACCAUAUUUUUCUCAUAAUGCAAACUAAUAU
UCCAAAAUUUUU
DR+tracr1
DR+tracr2

圖 31



DR: GUUUUAGUCCCCUUCGAUAUUGGGGUGGUCUAUAUC

tracrRNA:
AUUGAUGUGGUAUACUAAAAAUGGAAAAUUGUAUUUUUGAUUAGAAAGAUGUAAAAUU
GAUUUAAUUUAAAAAUAUUUUAUUAGAUUAAAGUAGA

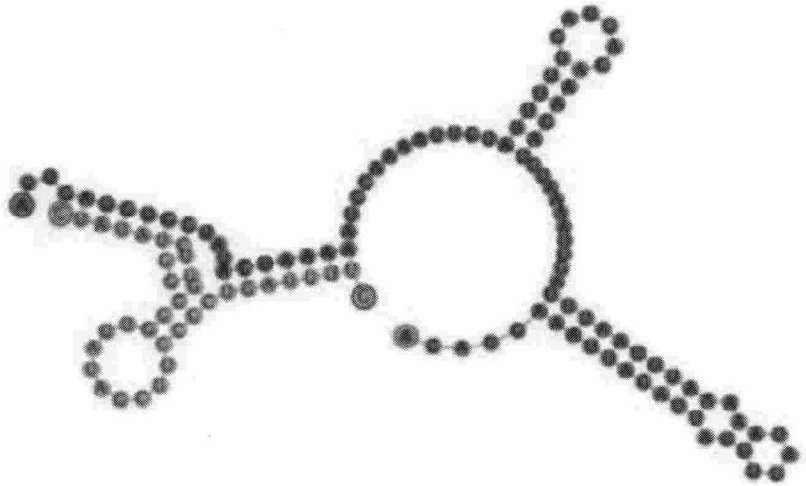
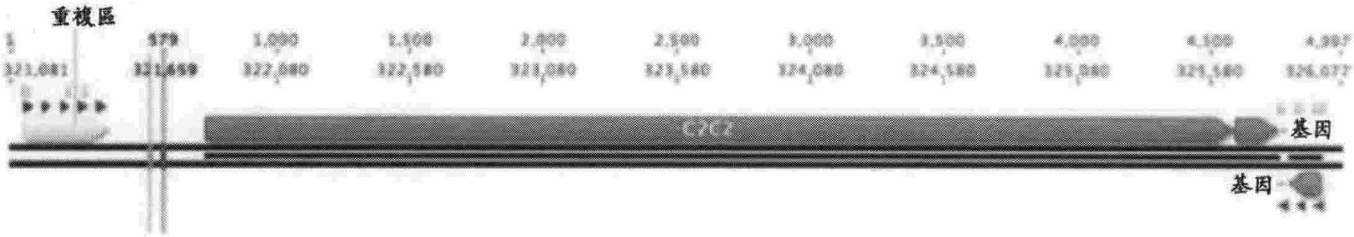


圖 32

DR1: GUUCAGUCCGCCGUCGUCUUGGCGGUGAUGUGAGGC
DR2: GUUCAGUCCGCCGUCAUUUUGGCGGUGAUGUGCUCC

tracrRNA: ???

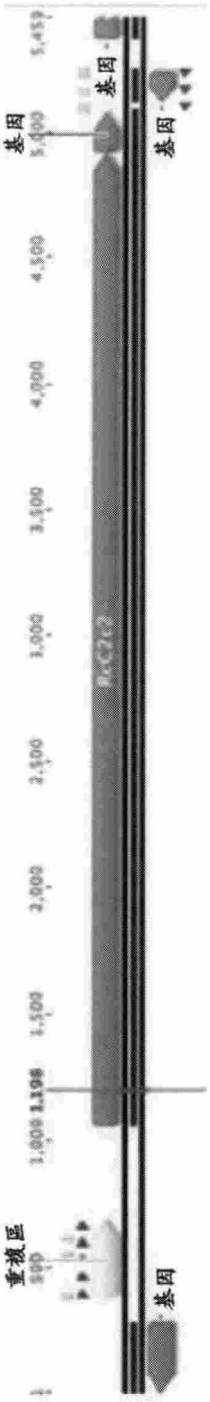
圖 33



DR1: GUUCAGUCCGCCGUCGUCUUGGCGGUGAUGUGAGGC
DR2: GUUCAGUCCGCCGUCAUUUUGGCGGUGAUGUGCUCC

tracrRNA: ???

圖 34



DR1: GUUCAGUCCGCCGUCGUCUUGGCGGUGAUGUGAGGC
DR2: GUUCAGUCCGCCGUCAUUUUGGCGGUGAUGUGCUCC

圖 35

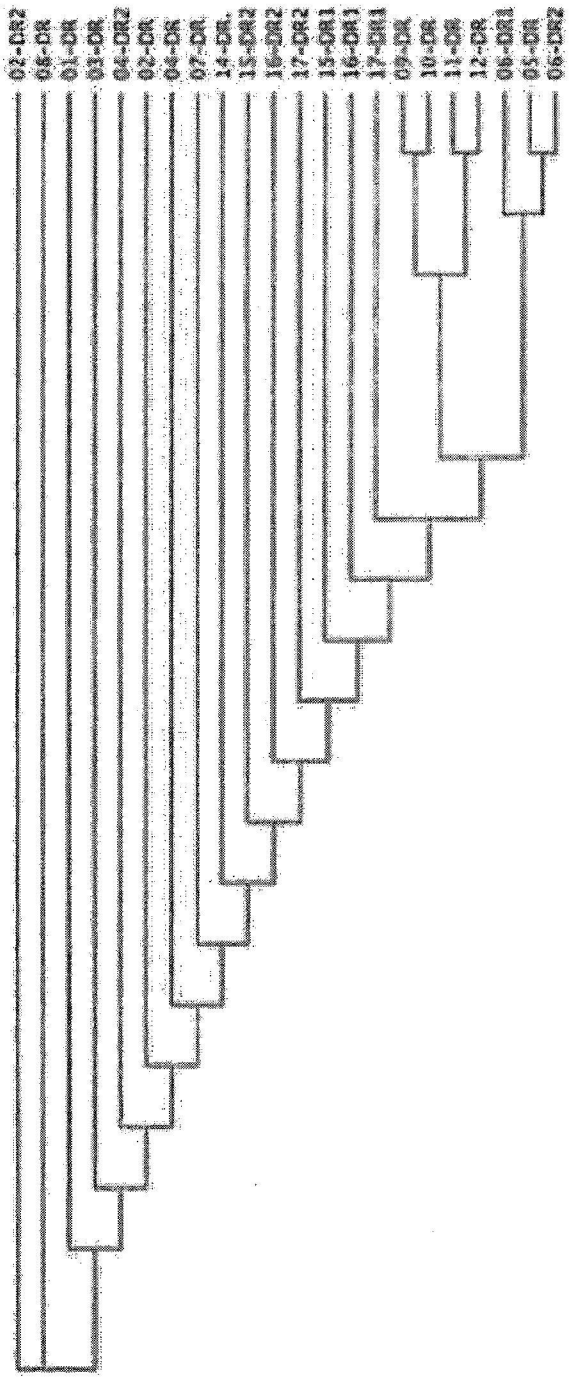

02-DR2
06-DR
01-DR
03-DR
08-DR2
02-DR
04-DR
07-DR
14-DR
13-DR2
10-DR2
17-DR2
15-DR1
16-DR1
17-DR1
09-DR
10-DR
11-DR
12-DR
08-DR1
05-DR
06-DR2

圖 36

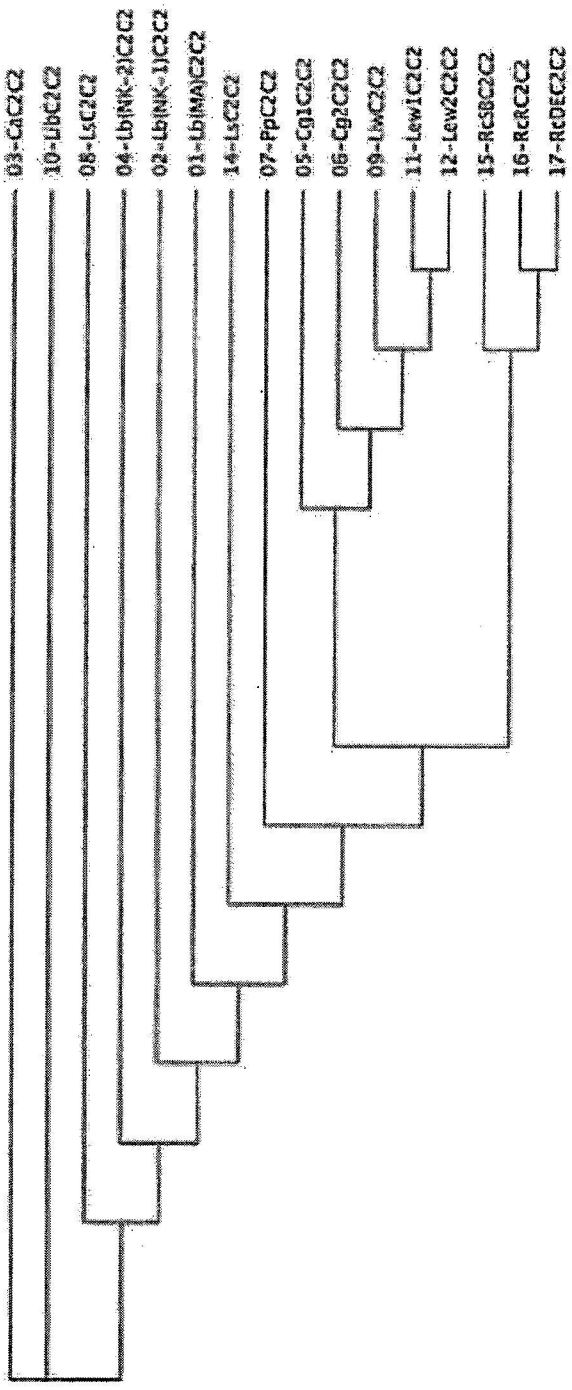

03-CaC2C2
10-LbC2C2
08-LsC2C2
04-LbiNK-2)C2C2
02-LbiNK-1)C2C2
01-LbiMA)C2C2
14-LsC2C2
07-PpC2C2
05-Cg1C2C2
06-Cp2C2C2
09-LwC2C2
11-LewiC2C2
12-Lew2C2C2
15-RcSBC2C2
16-RcRC2C2
17-RcDEC2C2

圖 37

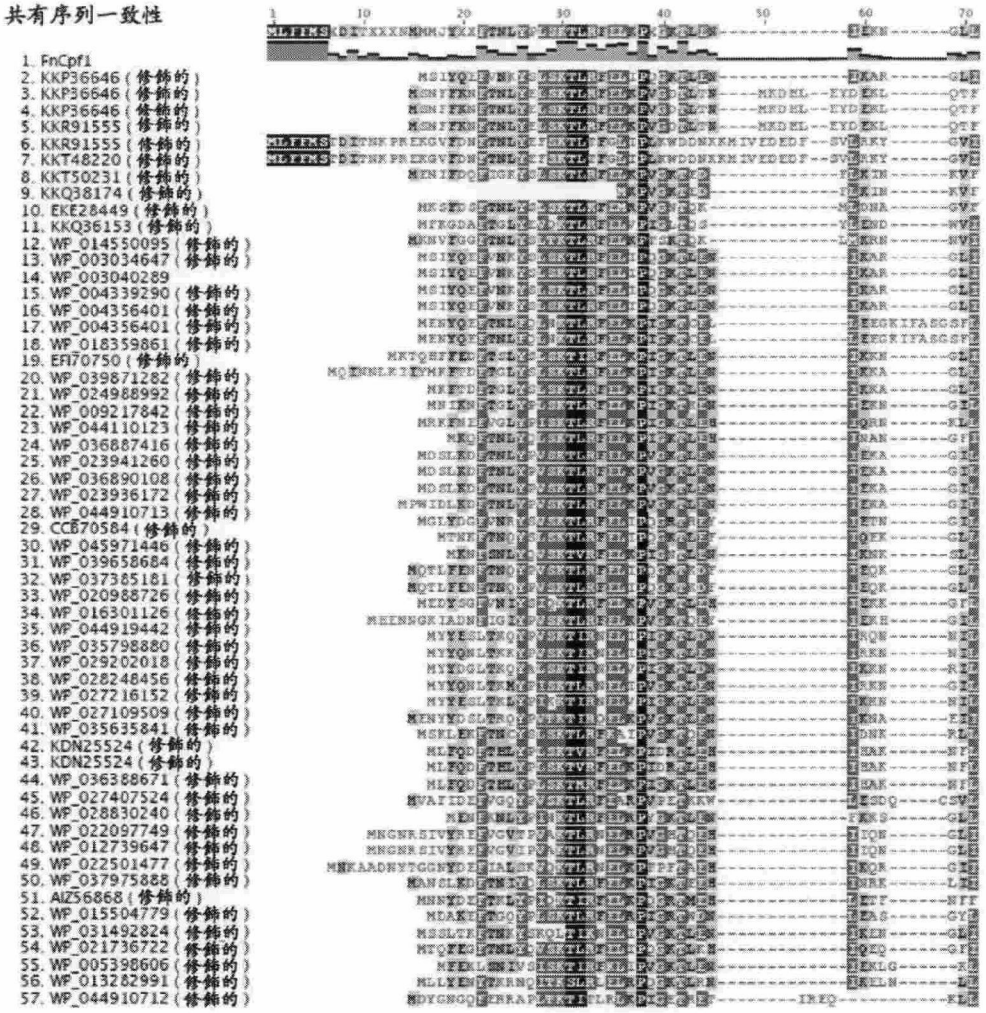

共有序列一致性
1. FnCpf1
2. KKP36646 (修飾的)
3. KKP36646 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKT50231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE28449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359861 (修飾的)
19. EFI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024988992 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890108 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044910713 (修飾的)
29. CCB70584 (修飾的)
30. WP_045971446 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037385181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301126 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975888 (修飾的)
51. AIZ56868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398606 (修飾的)
56. WP_013282991 (修飾的)
57. WP_044910712 (修飾的)

圖 38A

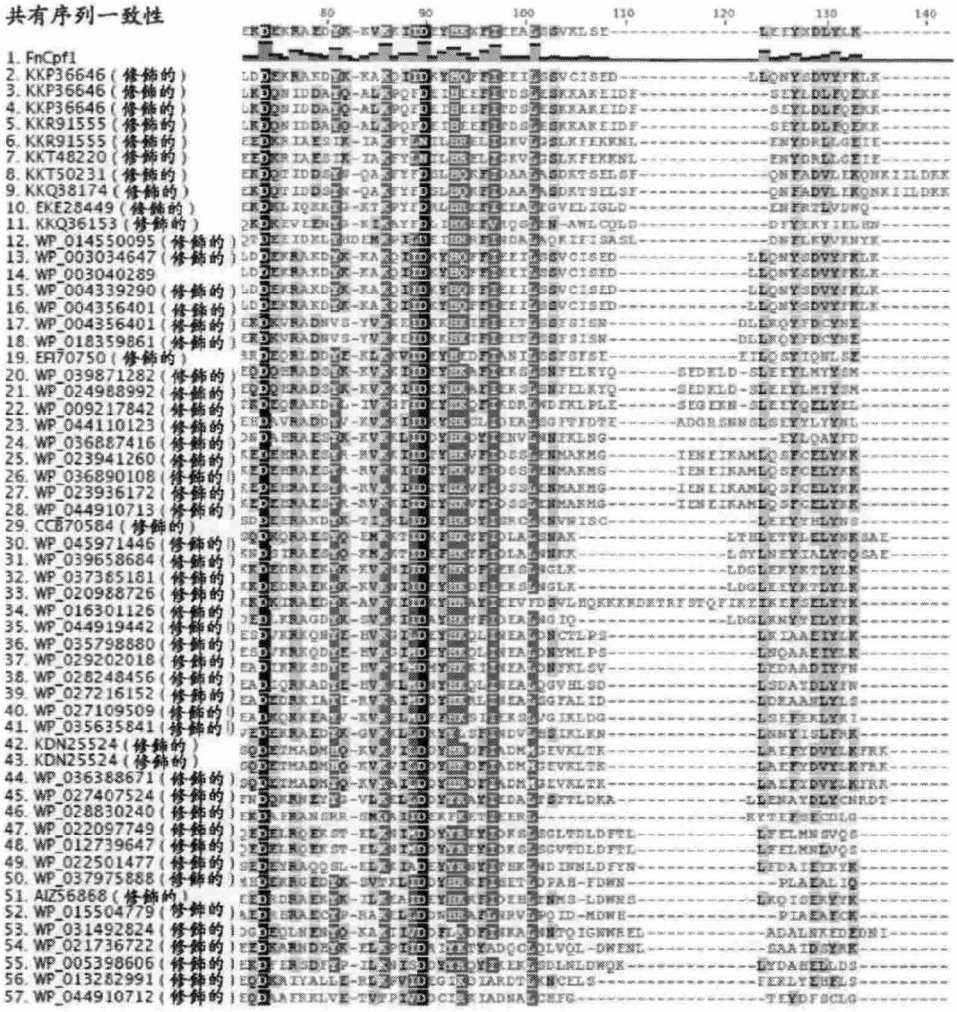

共有序列一致性
80
90
100
110
120
130
140
1. FnCpf1
2. KKP36646 (修飾的)
3. KKP36646 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKT50231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE28449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359861 (修飾的)
19. EFI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024988992 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890108 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044910713 (修飾的)
29. CCB70584 (修飾的)
30. WP_045971446 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037365181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301126 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975888 (修飾的)
51. AIZ56868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398606 (修飾的)
56. WP_013282991 (修飾的)
57. WP_044910712 (修飾的)

圖 38B

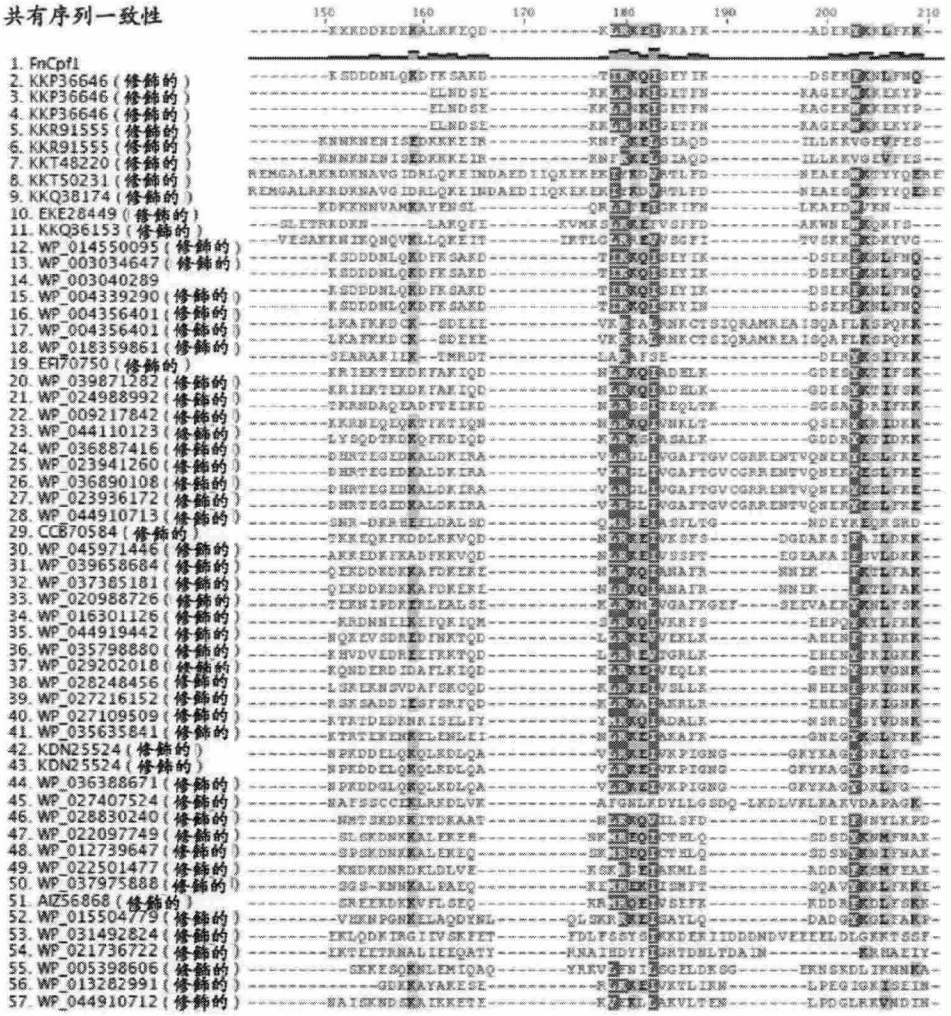

共有序列一致性
150
160
170
180
190
200
210
1. FnCpf1
2. KKP36646 (修飾的)
3. KKP36646 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKT50231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE28449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359861 (修飾的)
19. EFI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024988992 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890108 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044910713 (修飾的)
29. CCB70584 (修飾的)
30. WP_045971446 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037385181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301126 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975888 (修飾的)
51. AIZ56868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398606 (修飾的)
56. WP_013282991 (修飾的)
57. WP_044910712 (修飾的)

圖 38C

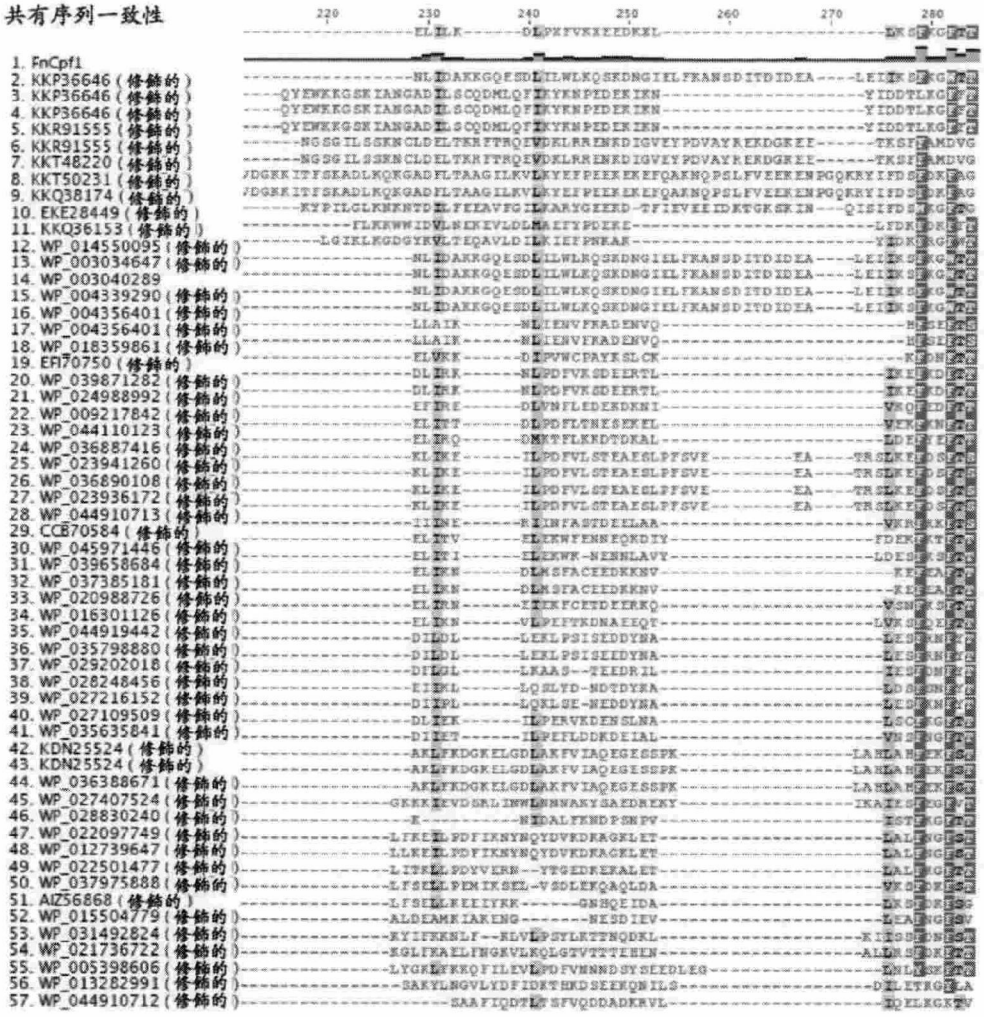

共有序列一致性
220
230
240
250
260
270
280
1. FnCpf1
2. KKP36646 (修飾的)
3. KKP36646 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKT50231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE28449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359861 (修飾的)
19. EFI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024988992 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890108 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044910713 (修飾的)
29. CCB70584 (修飾的)
30. WP_045971446 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037385181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301126 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975888 (修飾的)
51. AIZ56868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398606 (修飾的)
56. WP_013282991 (修飾的)
57. WP_044910712 (修飾的)

圖 38D

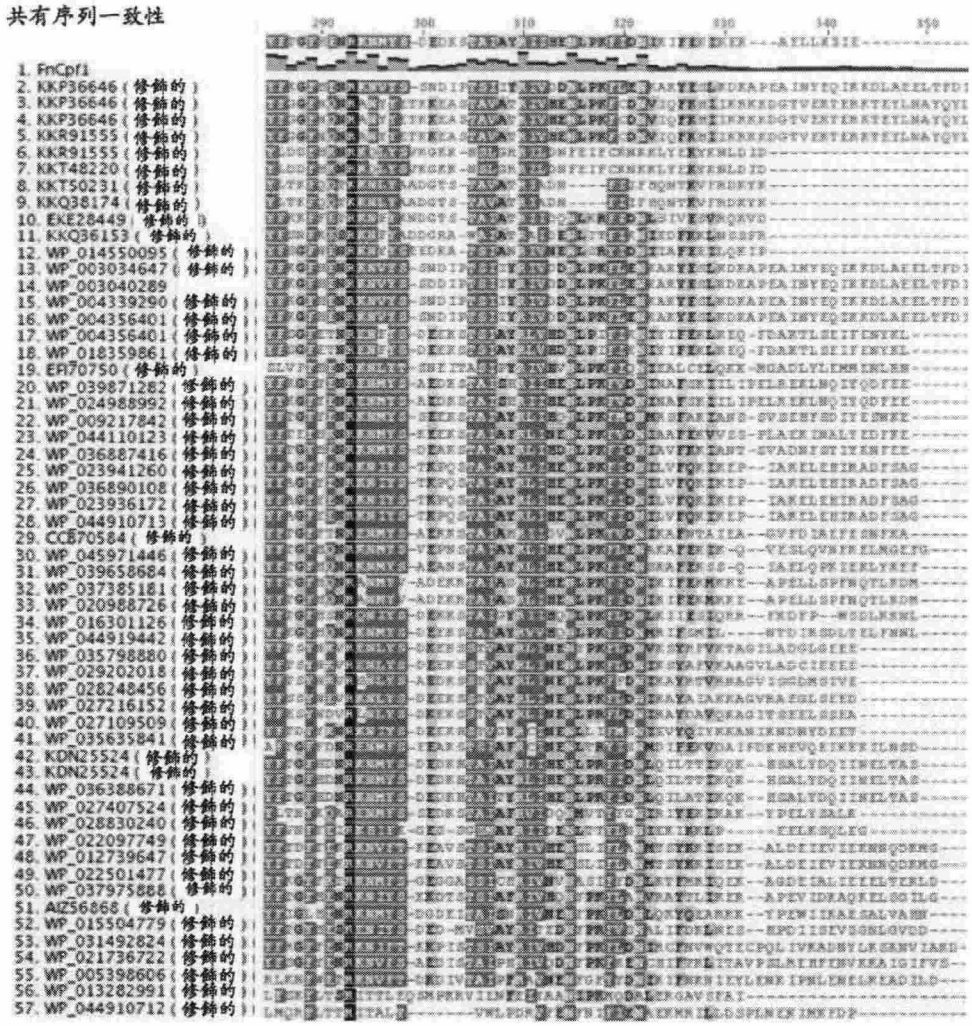

共有序列一致性
290
300
310
320
330
340
350
1. FnCpf1
2. KKP36646 (修飾的)
3. KKP36646 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKT50231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE28449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359861 (修飾的)
19. EFI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024988992 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890108 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044910713 (修飾的)
29. CCE70584 (修飾的)
30. WP_045971446 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037385181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301126 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975888 (修飾的)
51. AIZ56868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398606 (修飾的)
56. WP_013282991 (修飾的)
57. WP_044910712 (修飾的)

圖 38E

共有序列一致性

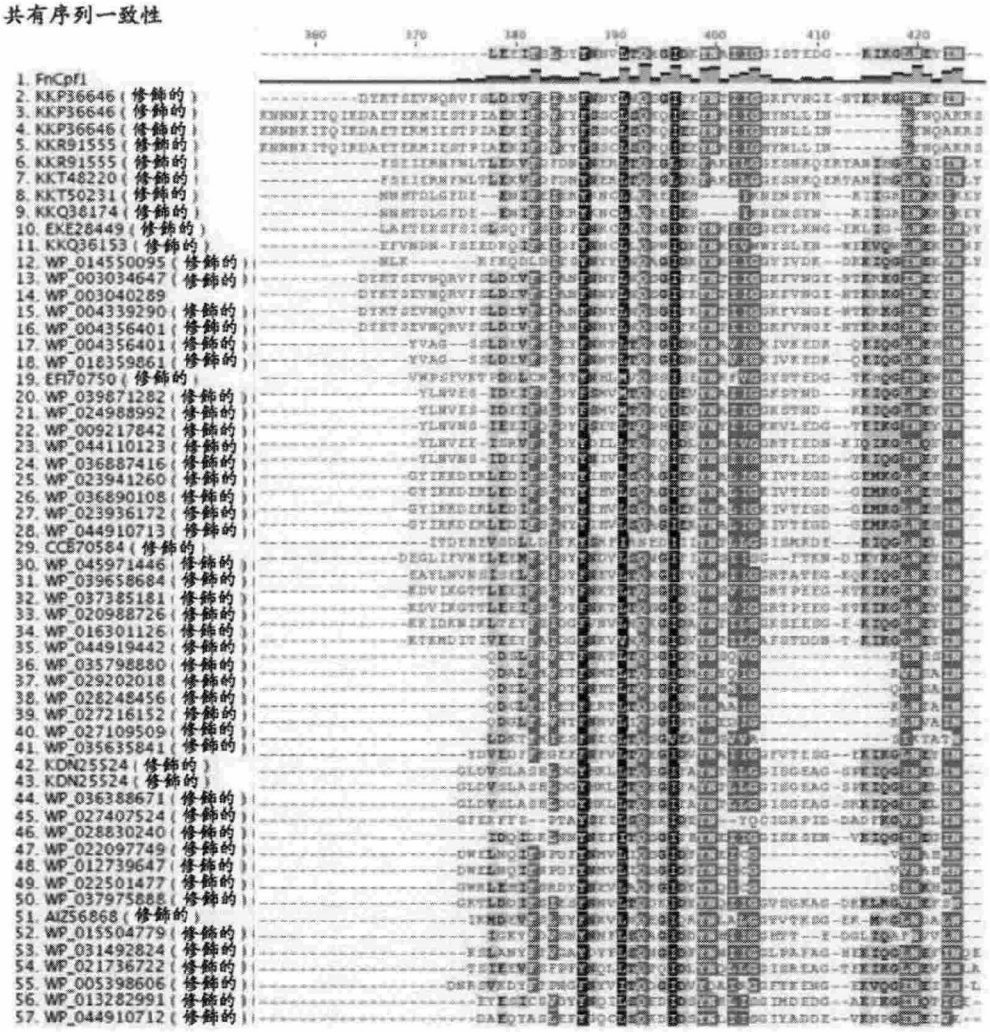


圖 38F

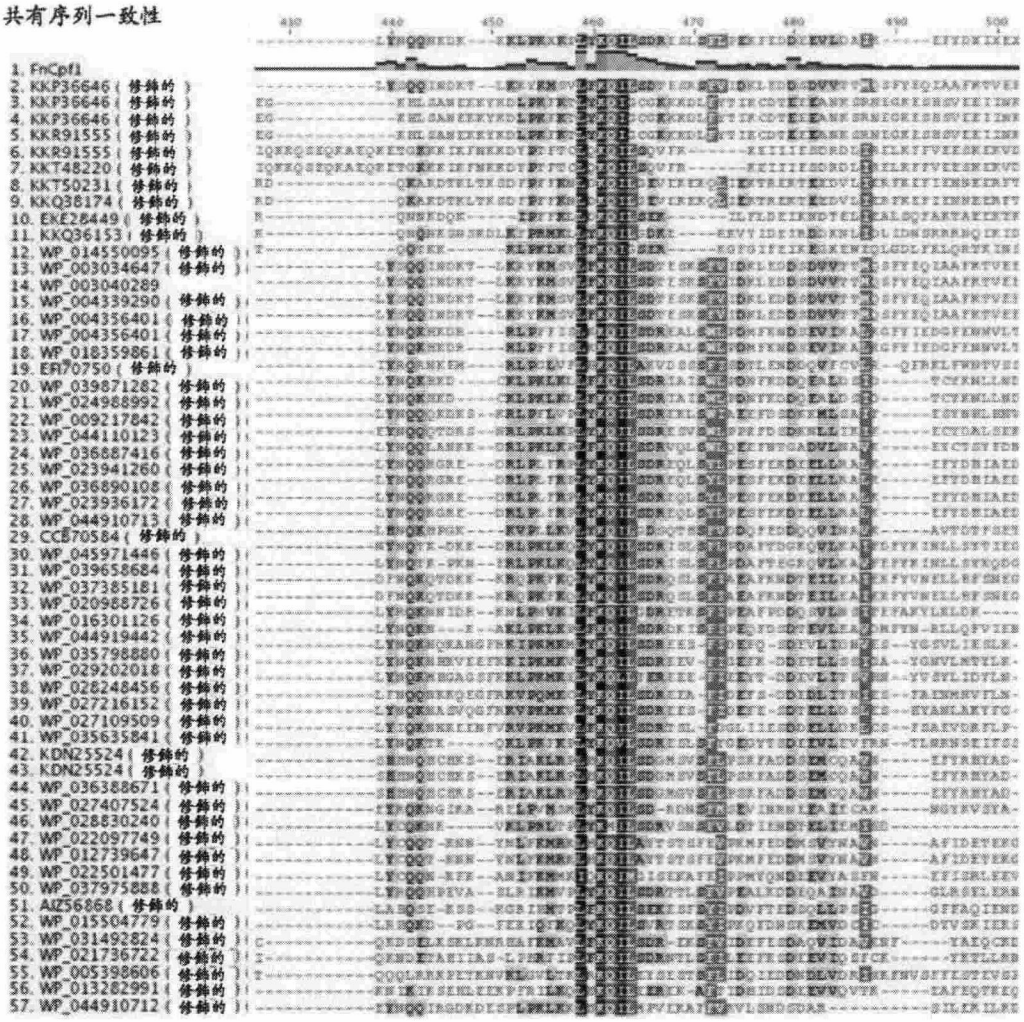

共有序列一致性
1. FnCpf1
2. KKP36646 (修飾的)
3. KKP36646 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKT50231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE28449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359861 (修飾的)
19. EFI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024988992 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890108 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044910713 (修飾的)
29. CCE70584 (修飾的)
30. WP_045971446 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037385181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301126 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975888 (修飾的)
51. AIZ56868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398606 (修飾的)
56. WP_013282991 (修飾的)
57. WP_044910712 (修飾的)

圖 38G

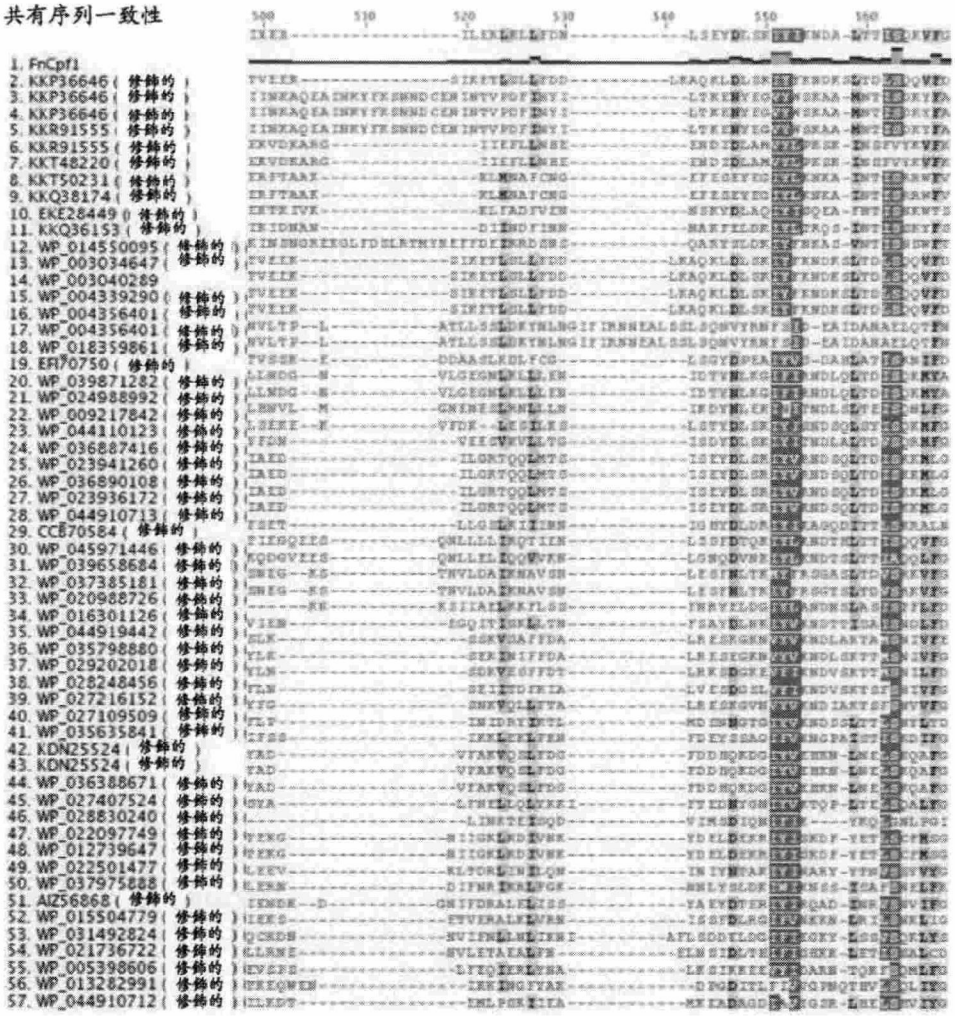

共有序列一致性
500
510
520
530
540
550
560
1. FnCpf1
2. KKP36646 (修飾的)
3. KKP36646 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKT50231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE28449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359861 (修飾的)
19. ERI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024988992 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890108 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044910713 (修飾的)
29. CCE70584 (修飾的)
30. WP_045971446 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037385181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301126 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975888 (修飾的)
51. AIZ56868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398606 (修飾的)
56. WP_013282991 (修飾的)
57. WP_044910712 (修飾的)

圖 38H

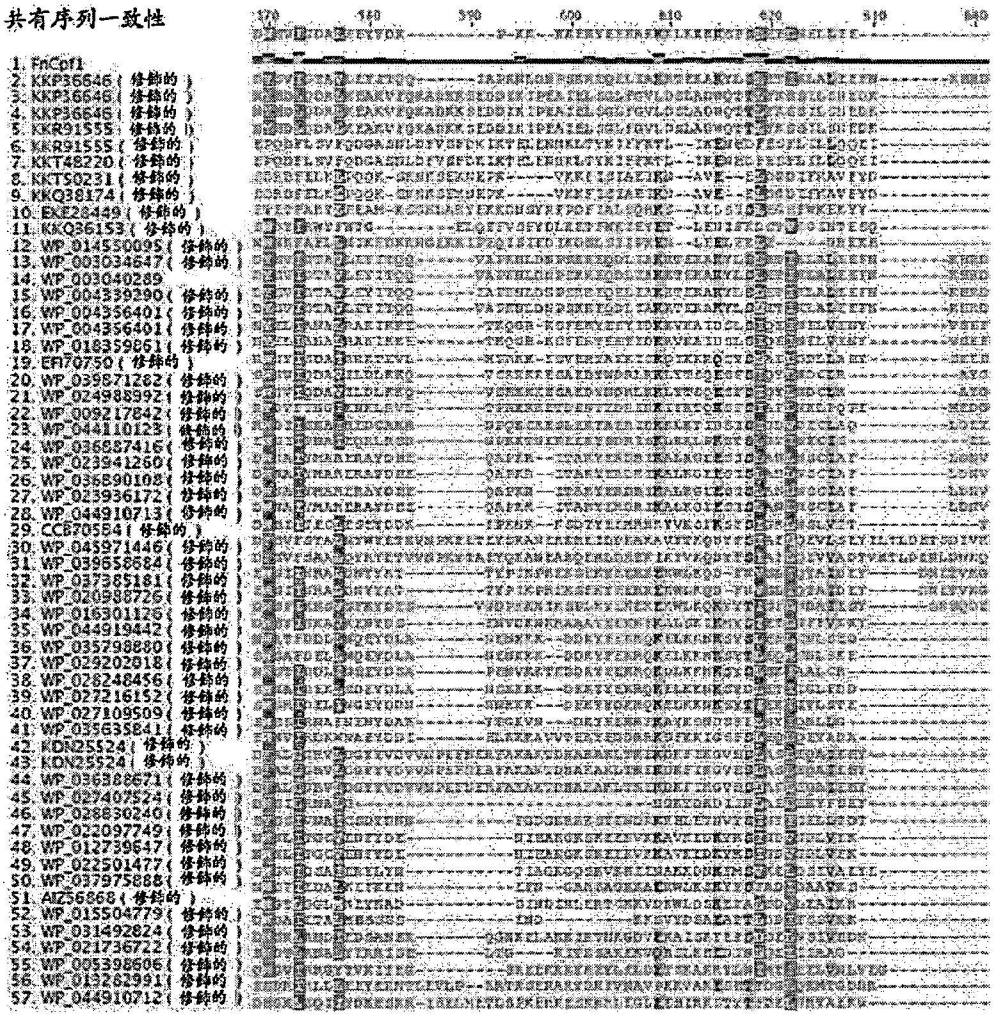

共有序列一致性
1. FnCpf1
2. KKP36646 (修飾的)
3. KKP36646 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKT50231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE28449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359861 (修飾的)
19. EFI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024988892 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890108 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044910713 (修飾的)
29. CCB70584 (修飾的)
30. WP_045971446 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037385181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301126 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975888 (修飾的)
51. AIZ56868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398606 (修飾的)
56. WP_013262991 (修飾的)
57. WP_044910712 (修飾的)

圖 38I

共有序列一致性

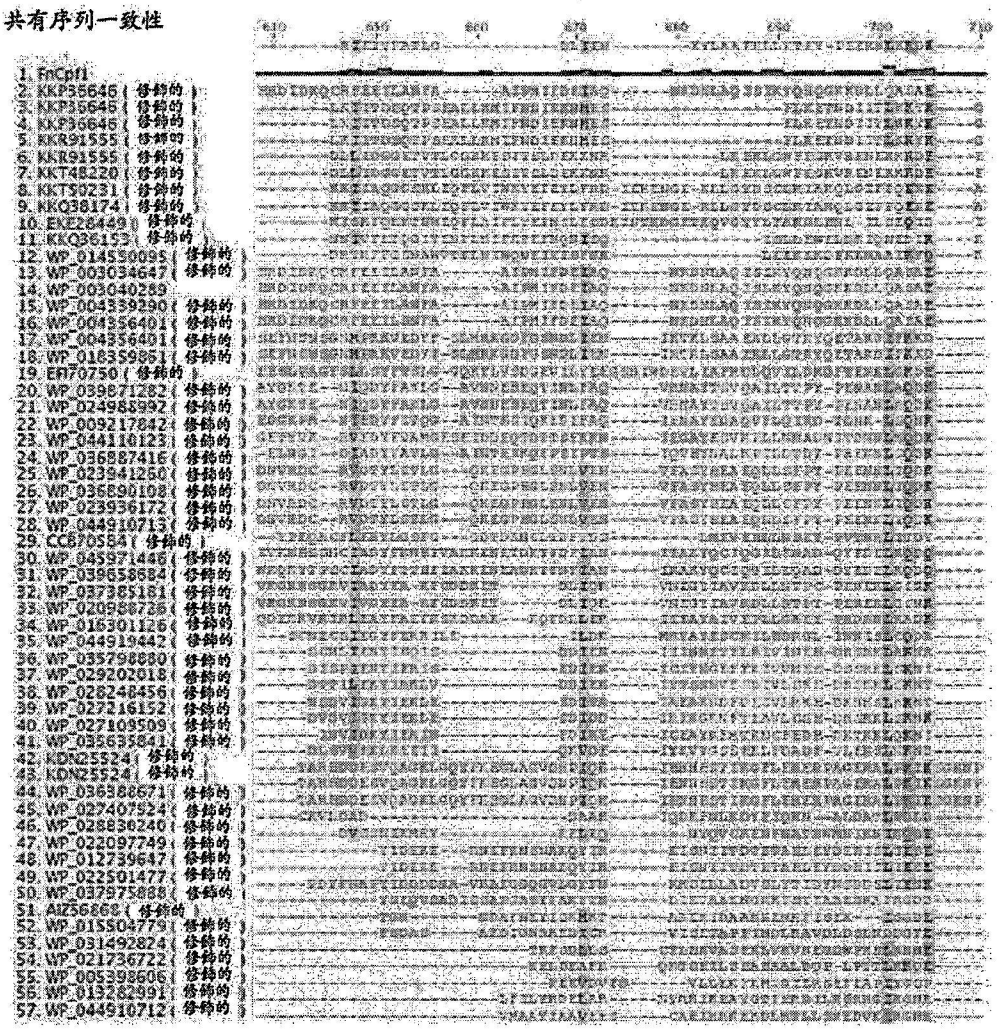


圖 38J

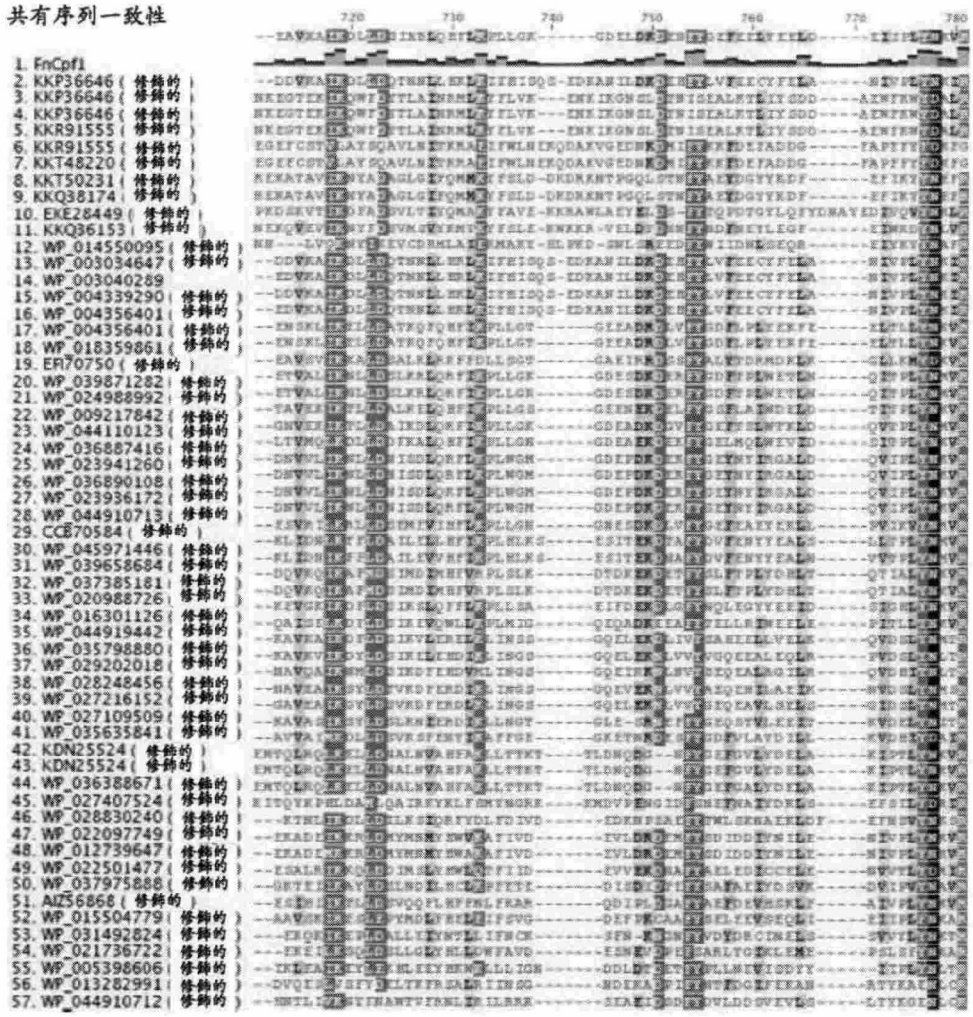

共有序列一致性
720
730
740
750
760
770
780
1. FnCpf1
2. KKP36646 (修飾的)
3. KKP36646 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKT50231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE28449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359861 (修飾的)
19. EFI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024988992 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890108 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044910713 (修飾的)
29. CCB70584 (修飾的)
30. WP_045971446 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037385181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301126 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975888 (修飾的)
51. AIZ56868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398606 (修飾的)
56. WP_013282991 (修飾的)
57. WP_044910712 (修飾的)

圖 38K

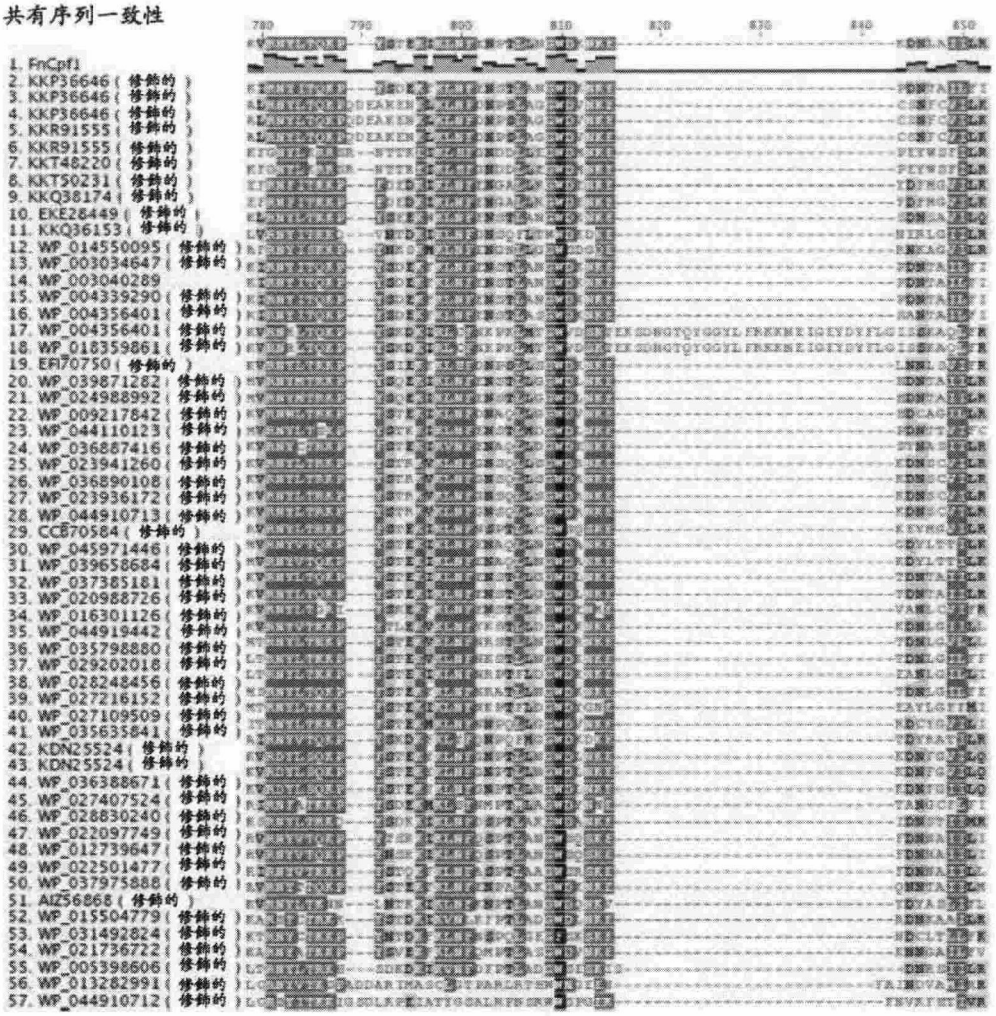

共有序列一致性
780
790
800
810
820
830
840
850
1. FnCpf1
2. KKP36646 (修飾的)
3. KKP36646 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKT50231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE28449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359861 (修飾的)
19. EFI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024988992 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890108 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044910713 (修飾的)
29. CCB70584 (修飾的)
30. WP_045971446 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037385181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301126 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975888 (修飾的)
51. AIZ56868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398606 (修飾的)
56. WP_013282991 (修飾的)
57. WP_044910712 (修飾的)

圖 38L

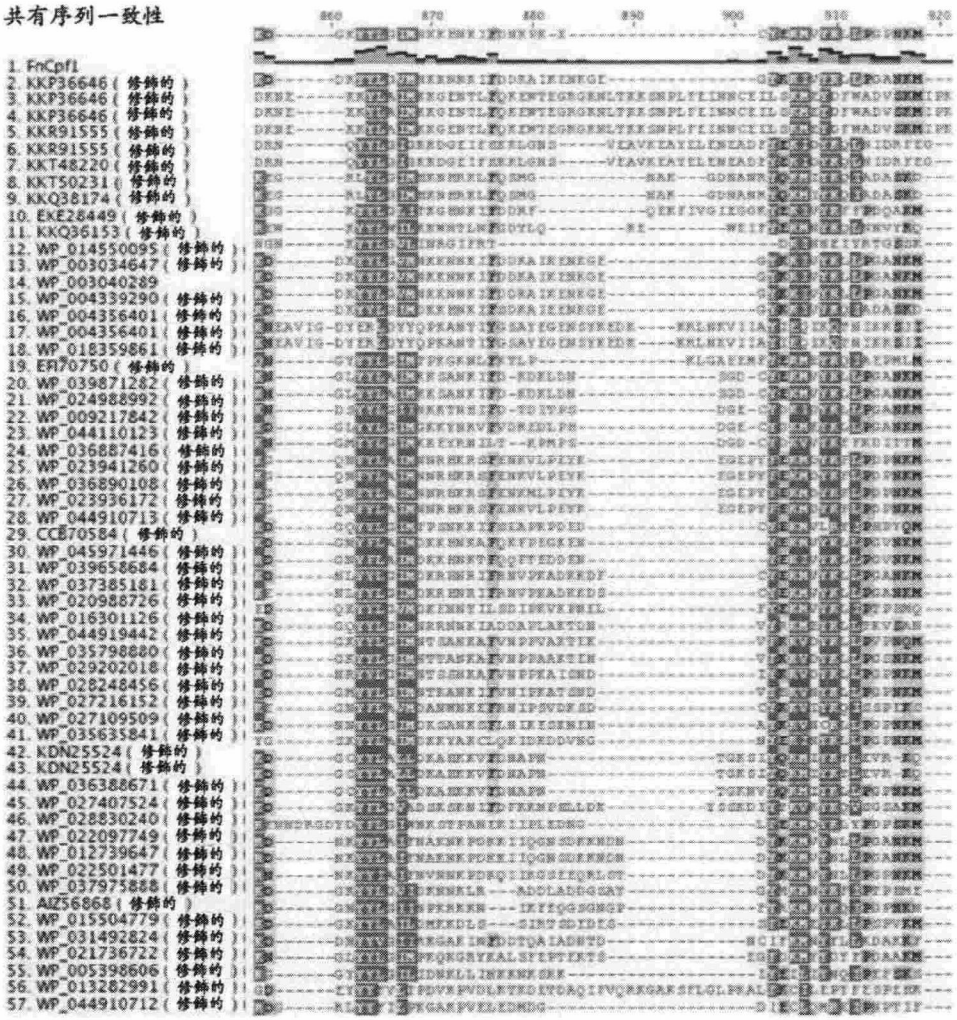

共有序列一致性
860
870
880
890
900
910
920
1. FnCpf1
2. KKP36646 (修飾的)
3. KKP36646 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKT50231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE28449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359861 (修飾的)
19. EFI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024988992 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890108 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044910713 (修飾的)
29. CCE70584 (修飾的)
30. WP_045971446 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037385181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301126 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975888 (修飾的)
51. AIZ56868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398606 (修飾的)
56. WP_013282991 (修飾的)
57. WP_044910712 (修飾的)

圖 38M

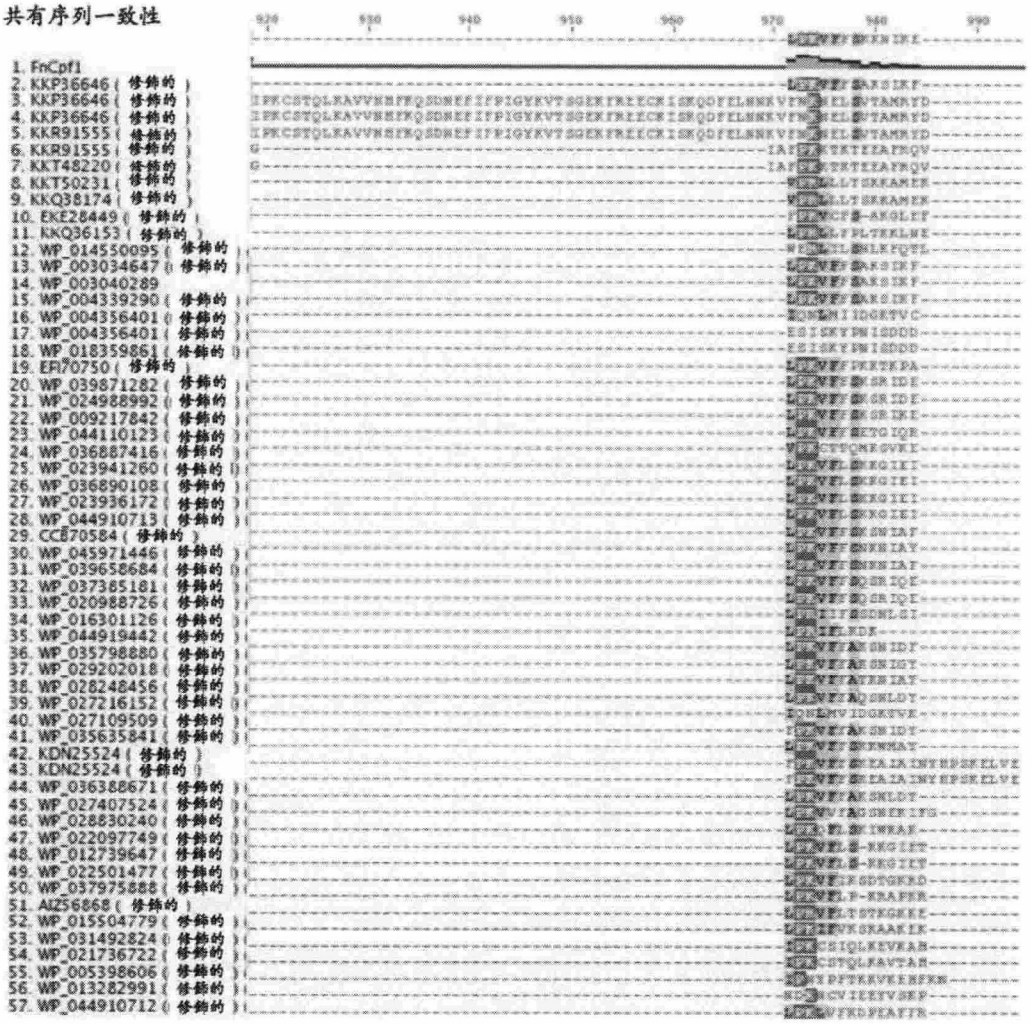

共有序列一致性
920
930
940
950
960
970
980
990
1. FnCpf1
2. KKP36646 (修飾的)
3. KKP36646 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKT50231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE28449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359851 (修飾的)
19. EFI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024988992 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890108 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044910713 (修飾的)
29. CCE70584 (修飾的)
30. WP_045971446 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037385181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301126 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975888 (修飾的)
51. AIZ56868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398606 (修飾的)
56. WP_013282991 (修飾的)
57. WP_044910712 (修飾的)

圖 38N

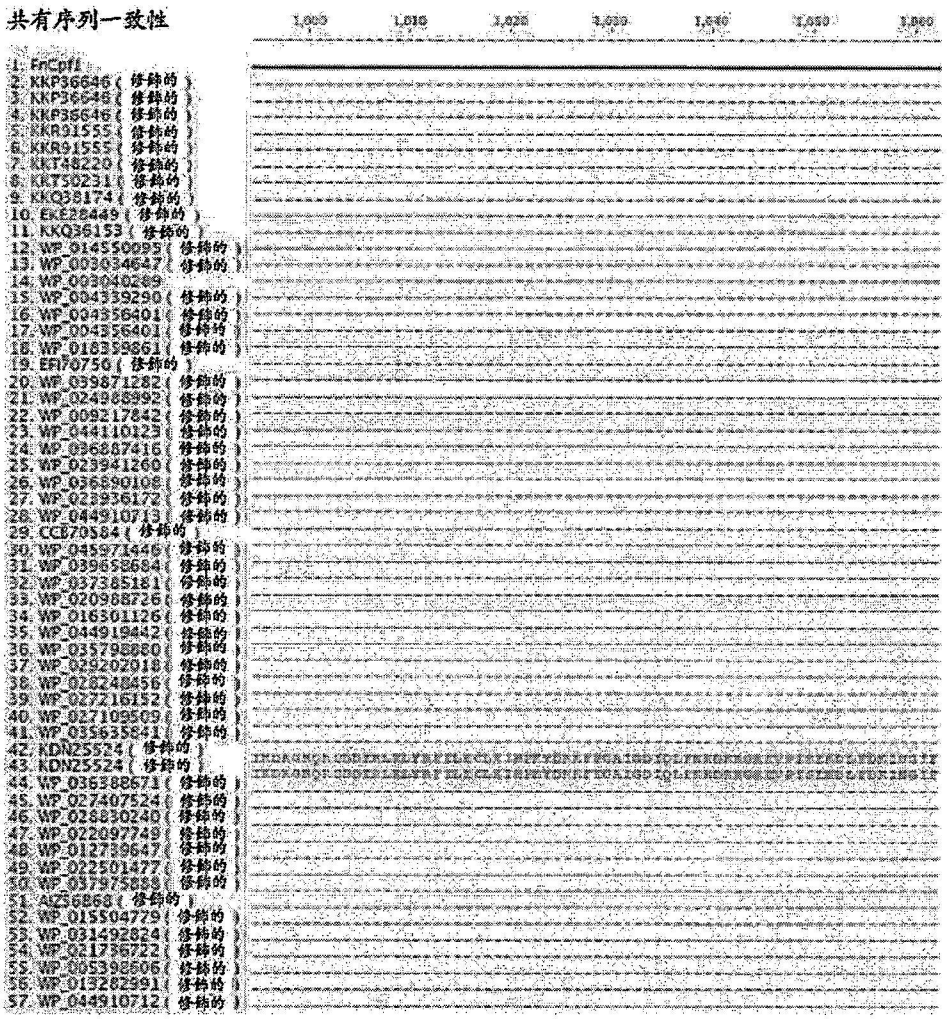

共有序列一致性
1,005
1,010
1,015
1,020
1,025
1,030
1,035
1,040
1,045
1,050
1,055
1,060
1. FnCpf1
2. KKP36646 (修飾的)
3. KKP36648 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKT50231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE28449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359861 (修飾的)
19. EFI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024868992 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890108 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044910713 (修飾的)
29. CCB70584 (修飾的)
30. WP_045971446 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037385181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301126 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975888 (修飾的)
51. AQZ36868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398506 (修飾的)
56. WP_013282991 (修飾的)
57. WP_044910712 (修飾的)

圖 38O

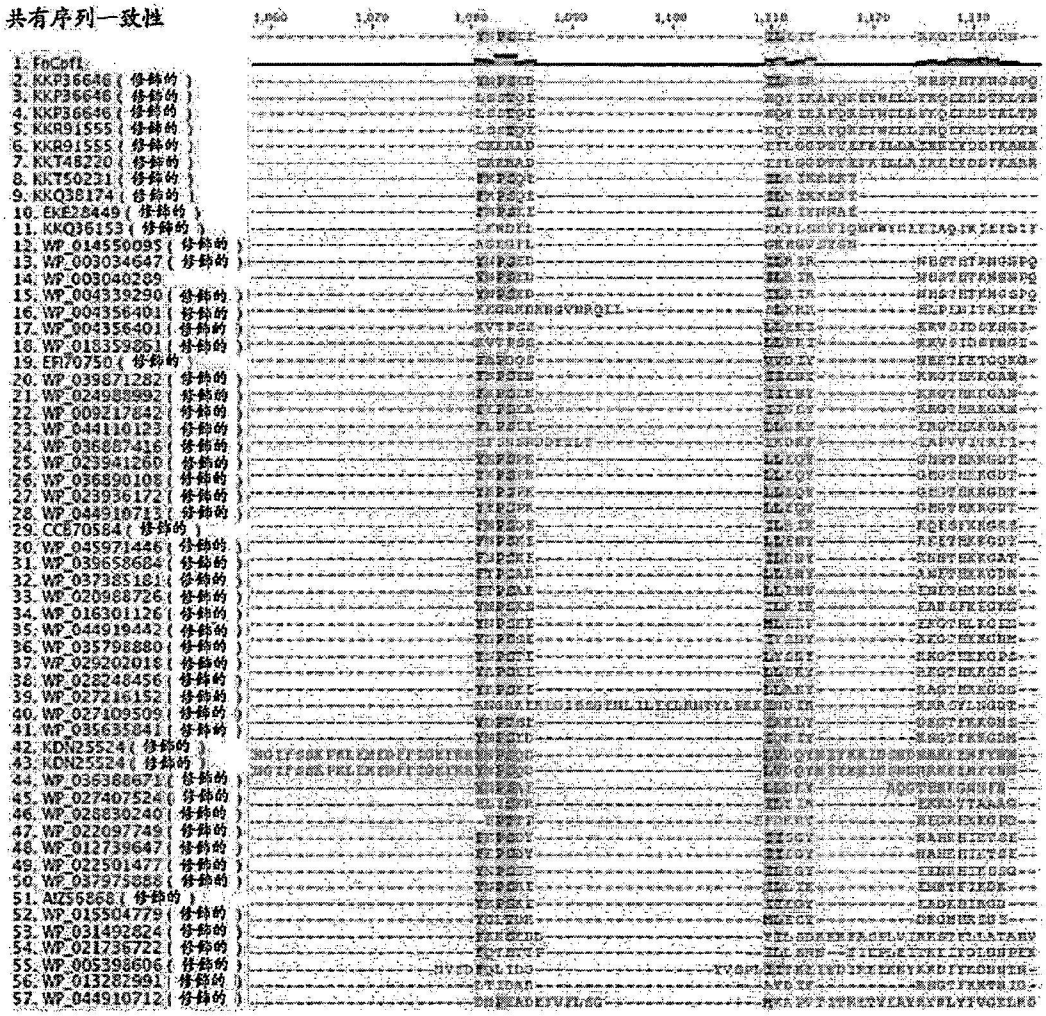

共有序列一致性
1. FoCbf1
2. KKP36646 (修飾的)
3. KKP36646 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKT50231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE28449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359851 (修飾的)
19. EFI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024988892 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890105 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044918713 (修飾的)
29. CCE70584 (修飾的)
30. WP_045971446 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037385181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301126 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975088 (修飾的)
51. AIZ56868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398606 (修飾的)
56. WP_013282591 (修飾的)
57. WP_044910712 (修飾的)

圖 38P

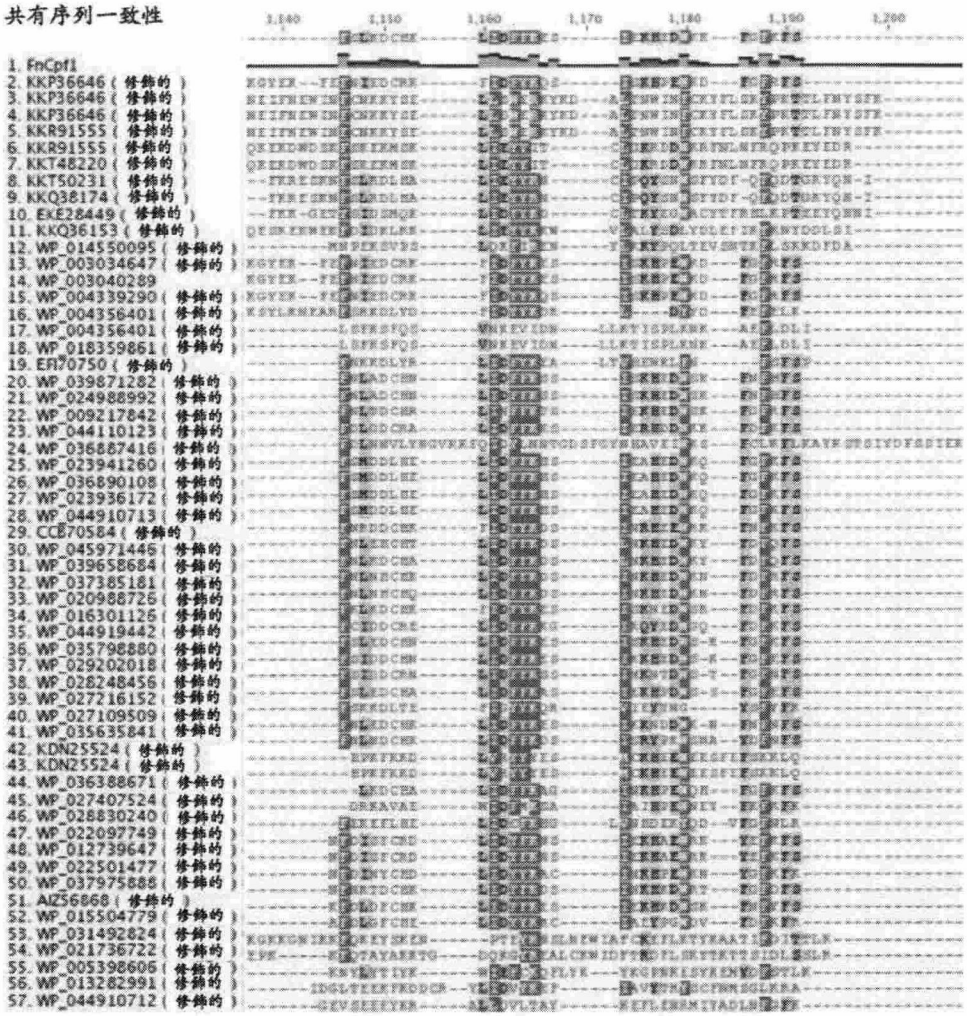

共有序列一致性
1,140
1,150
1,160
1,170
1,180
1,190
1,200
1. FnCpf1
2. KKP36646 (修飾的)
3. KKP36646 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKT50231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE28449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359861 (修飾的)
19. EFI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024988992 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890108 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044910713 (修飾的)
29. CCB70584 (修飾的)
30. WP_045971445 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037385181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301125 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975888 (修飾的)
51. AIZ56868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398606 (修飾的)
56. WP_013282991 (修飾的)
57. WP_044910712 (修飾的)

圖 38Q

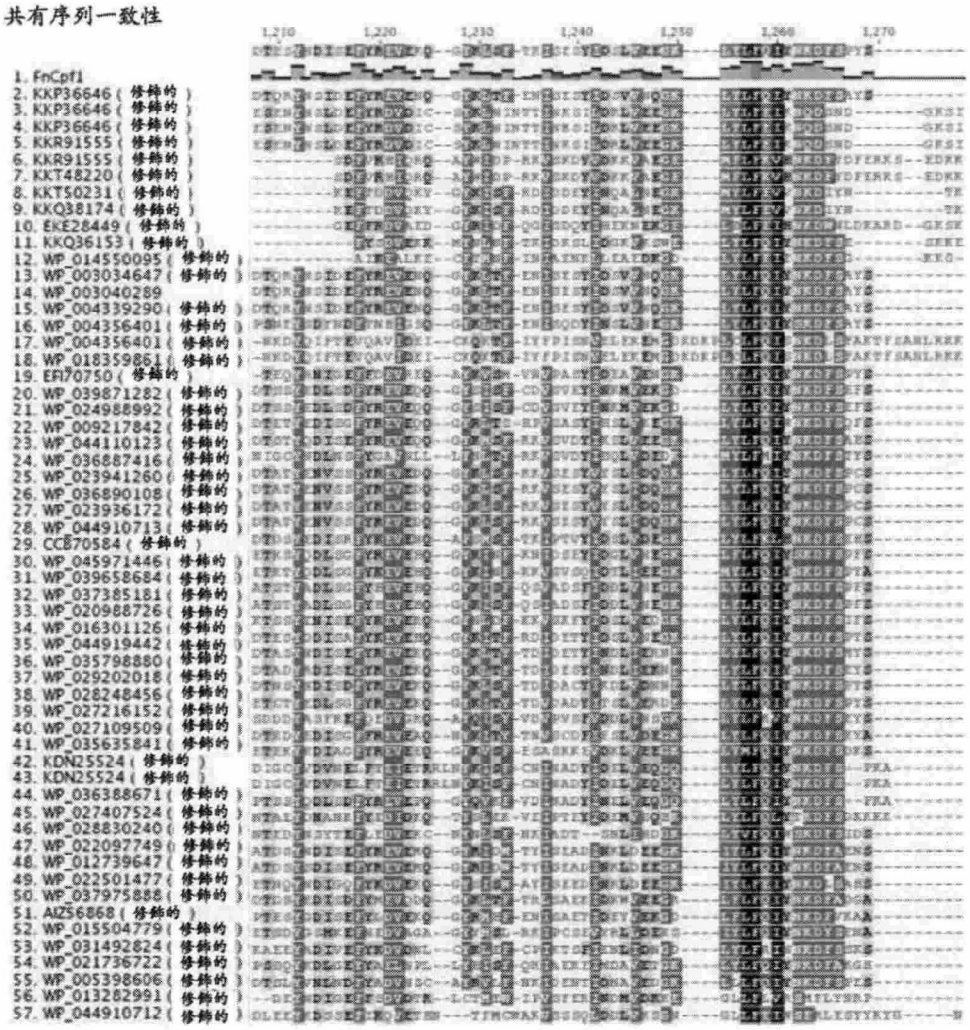

共有序列一致性
1,210
1,220
1,230
1,240
1,250
1,260
1,270
1. FnCpf1
2. KKP36646 (修飾的)
3. KKP36646 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKTS0231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE28449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359861 (修飾的)
19. EFI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024988992 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890108 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044910713 (修飾的)
29. CCE70584 (修飾的)
30. WP_045971446 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037385181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301126 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975888 (修飾的)
51. AIZ56868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398606 (修飾的)
56. WP_013282991 (修飾的)
57. WP_044910712 (修飾的)

圖 38R

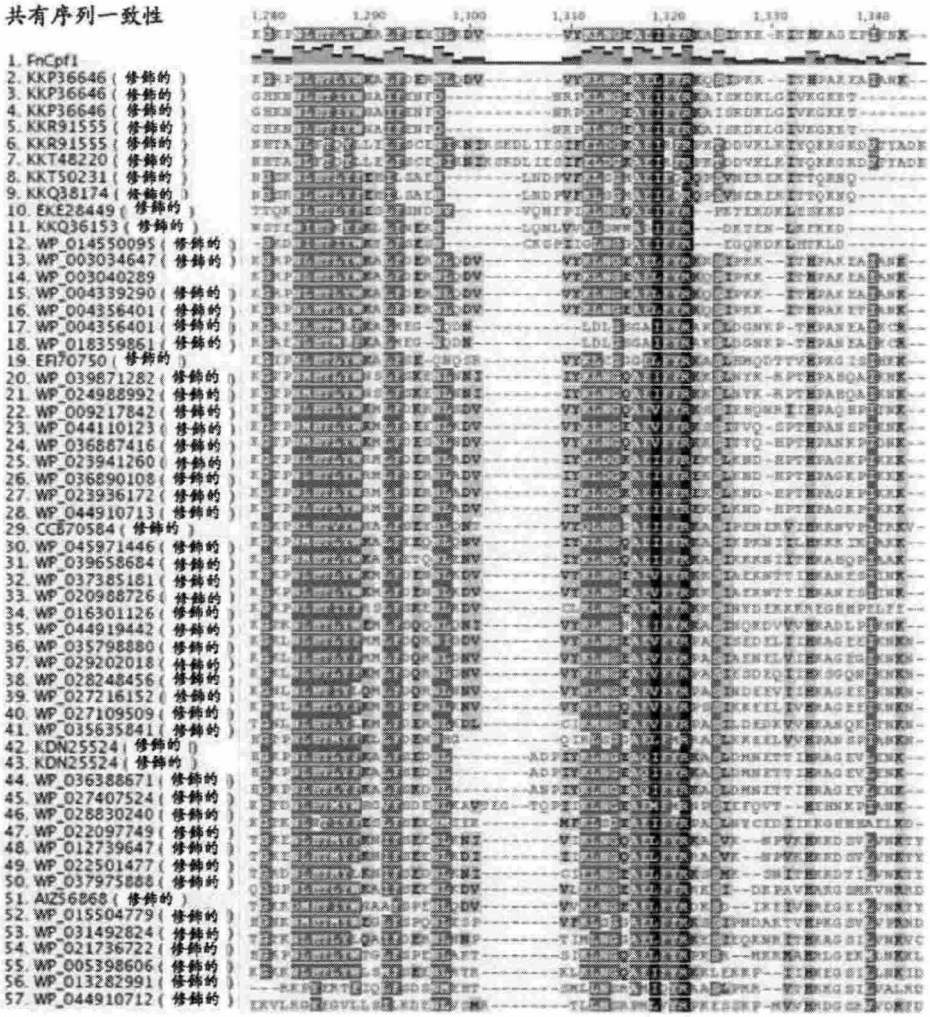

共有序列一致性
1,280
1,290
1,300
1,310
1,320
1,330
1,340
1. FnCpf1
2. KKP36646 (修飾的)
3. KKP36646 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKT50231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE28449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359861 (修飾的)
19. EFI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024988992 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890108 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044910713 (修飾的)
29. CCB70584 (修飾的)
30. WP_045971446 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037385181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301126 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975888 (修飾的)
51. AIZ56868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398606 (修飾的)
56. WP_013282991 (修飾的)
57. WP_044910712 (修飾的)

圖 38S

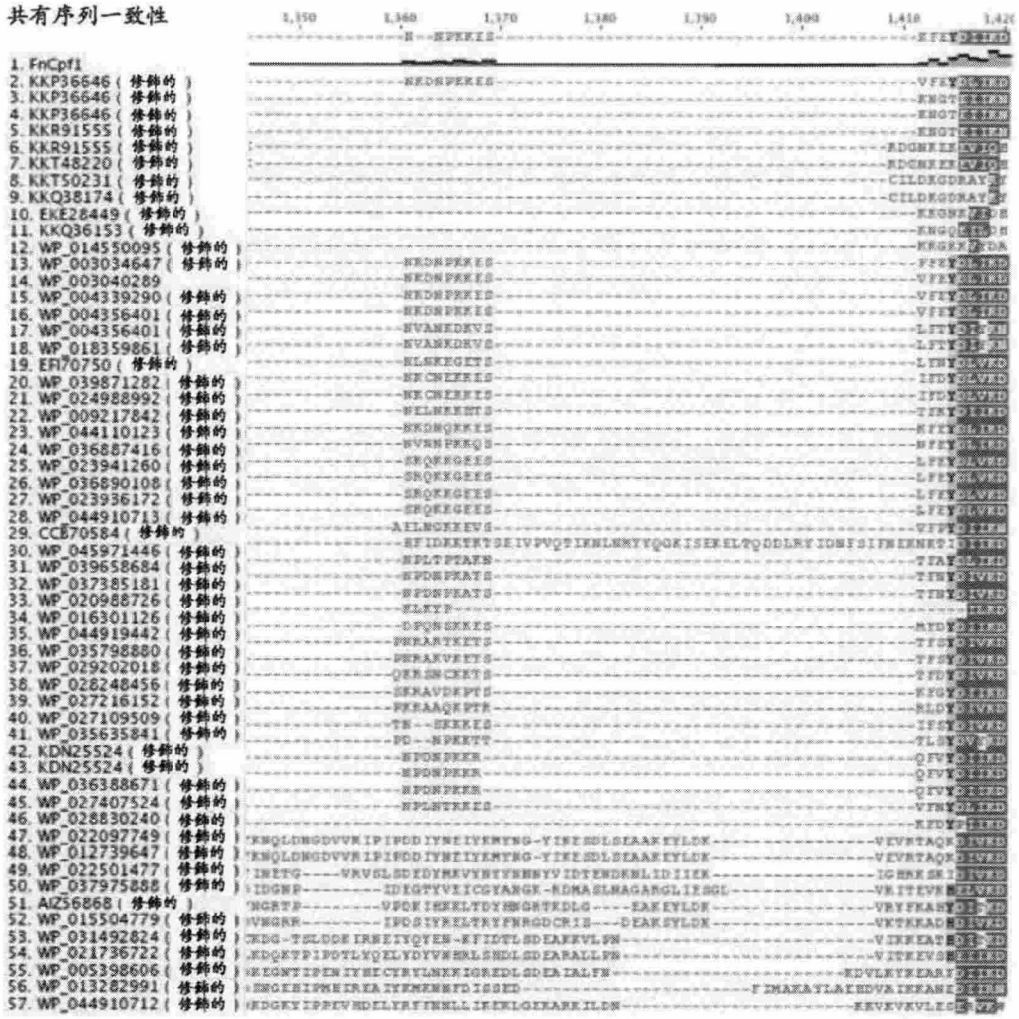

共有序列一致性
1,350
1,360
1,370
1,380
1,390
1,400
1,410
1. FnCpf1
2. KKP36646 (修飾的)
3. KKP36646 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKT50231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE28449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359861 (修飾的)
19. EFI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024988992 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890108 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044910713 (修飾的)
29. CCE70584 (修飾的)
30. WP_045971446 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037385181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301126 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975885 (修飾的)
51. AIZ36868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398606 (修飾的)
56. WP_013282991 (修飾的)
57. WP_044910712 (修飾的)

圖 38T

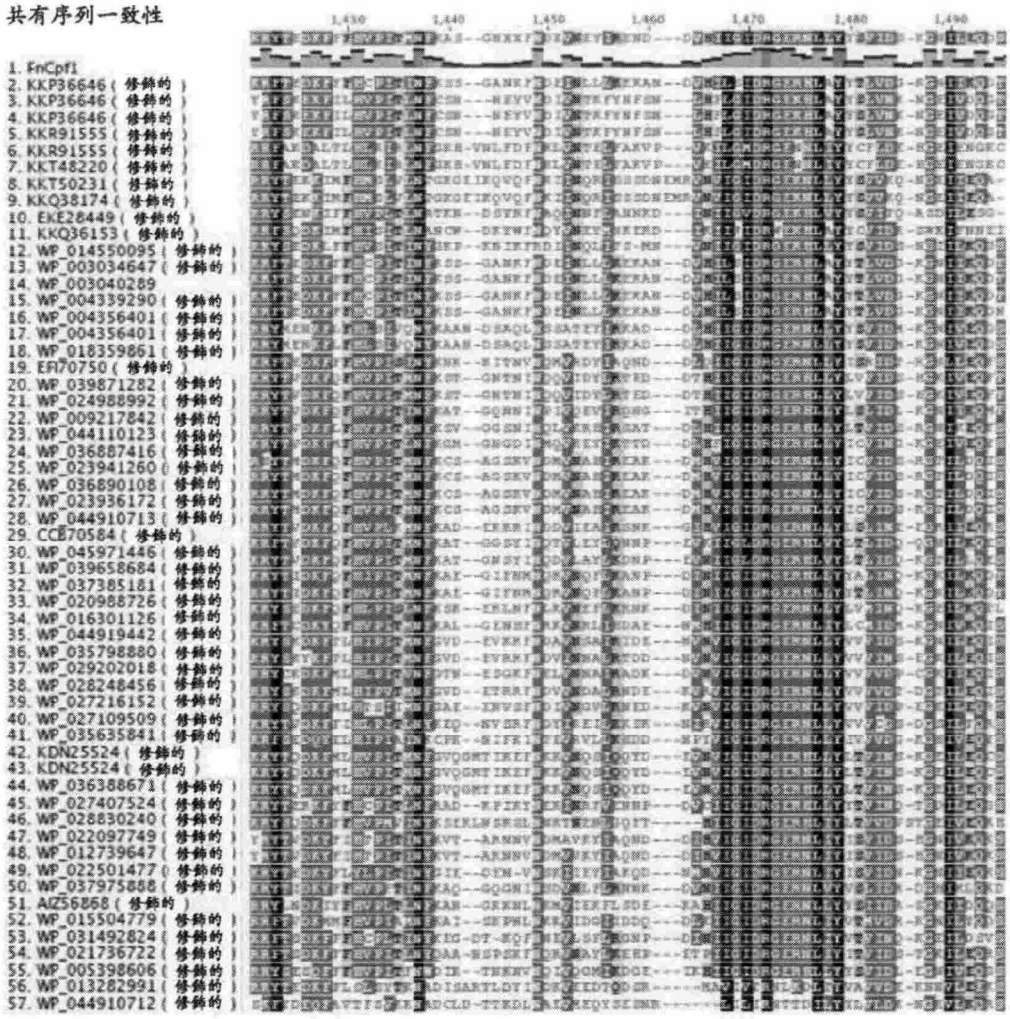

共有序列一致性
1,430
1,440
1,450
1,460
1,470
1,480
1,490
1. FnCpf1
2. KKP36646 (修飾的)
3. KKP36646 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKT50231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE28449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359861 (修飾的)
19. EFI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024988992 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890108 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044910713 (修飾的)
29. CCE70584 (修飾的)
30. WP_045971446 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037385181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301126 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975888 (修飾的)
51. AIZ56868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398606 (修飾的)
56. WP_013282991 (修飾的)
57. WP_044910712 (修飾的)

圖 38U

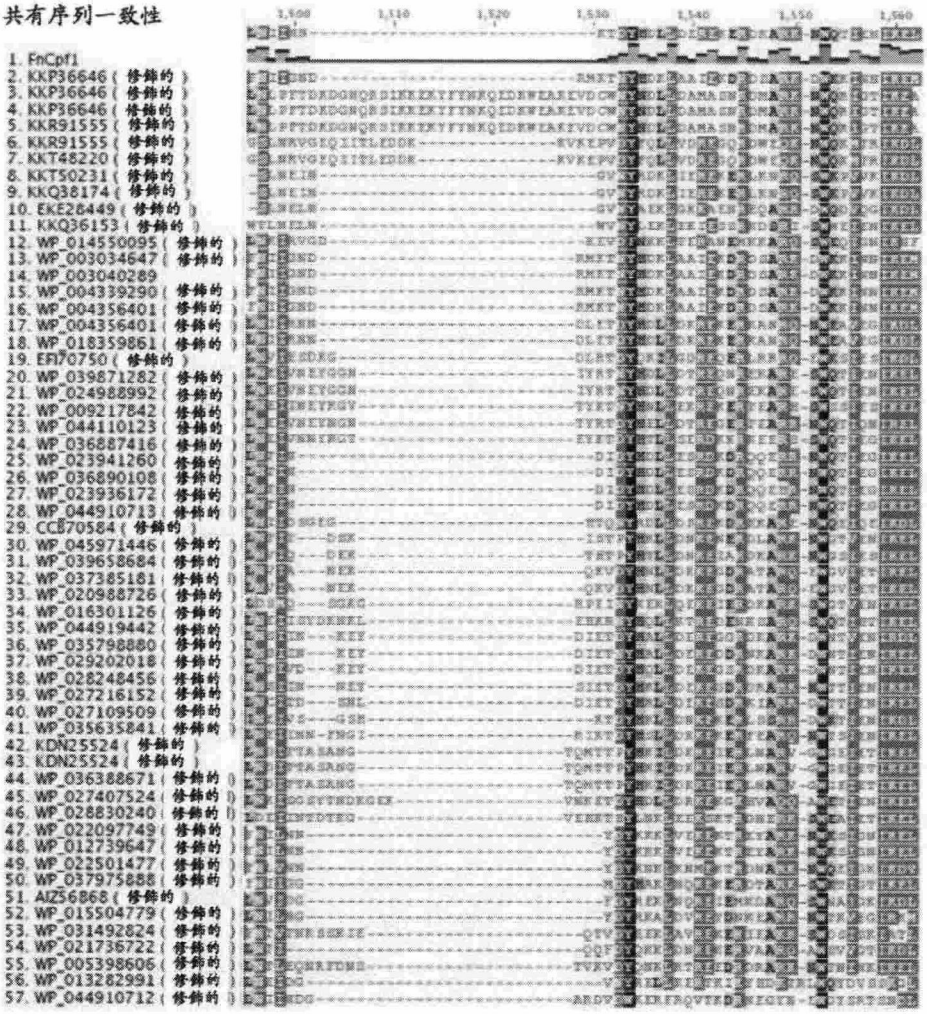

共有序列一致性
1,500
1,510
1,520
1,530
1,540
1,550
1,560
1. FnCpf1
2. KKP36646 (修飾的)
3. KKP36646 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKT50231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE26449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359861 (修飾的)
19. EFI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024988992 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890108 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044910713 (修飾的)
29. CCE70584 (修飾的)
30. WP_045971446 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037385181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301126 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975888 (修飾的)
51. AIZ56868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398606 (修飾的)
56. WP_013282991 (修飾的)
57. WP_044910712 (修飾的)

圖 38V

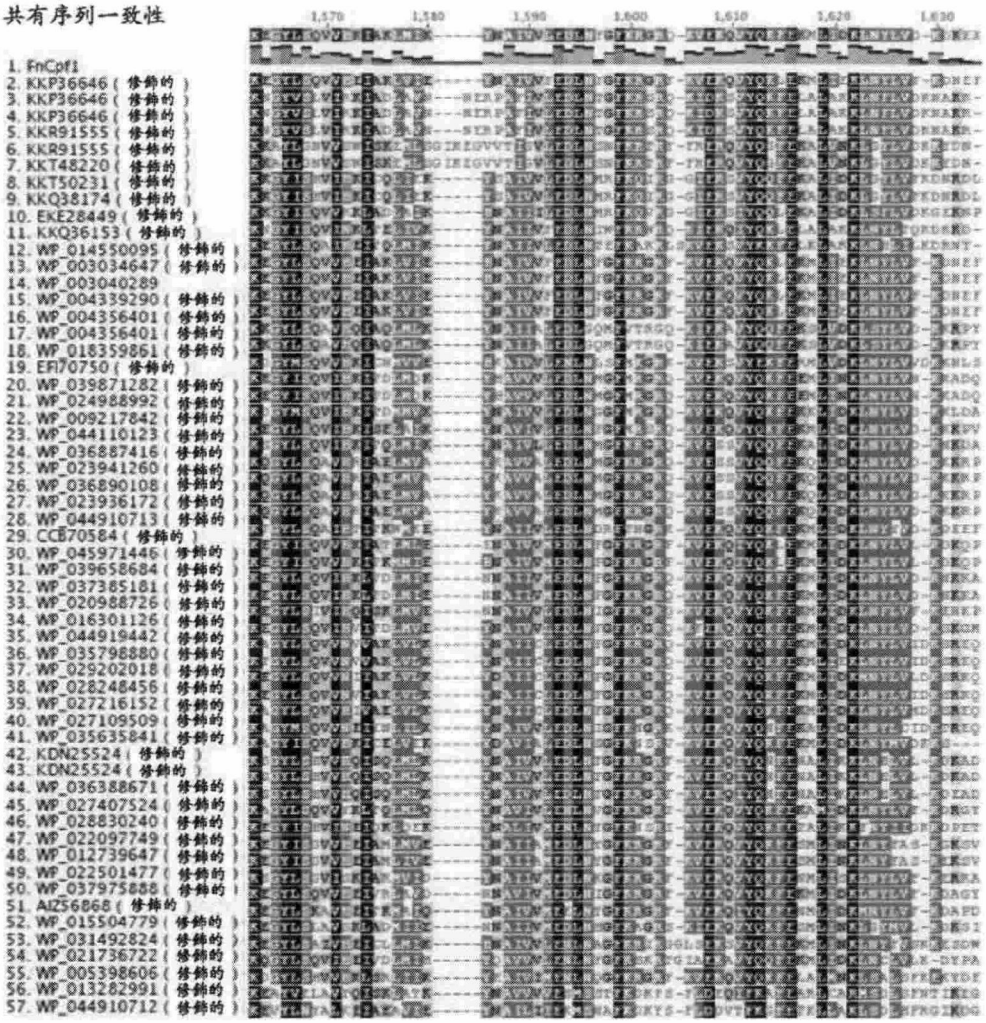

共有序列一致性
1,570
1,580
1,590
1,600
1,610
1,620
1,630
1. FnCpf1
2. KKP36646 (修飾的)
3. KKP36646 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKT50231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE28449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359861 (修飾的)
19. EFI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024988992 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890108 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044910713 (修飾的)
29. CCE70584 (修飾的)
30. WP_045971446 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037385181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301126 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975888 (修飾的)
51. AIZ56868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398606 (修飾的)
56. WP_013282991 (修飾的)
57. WP_044910712 (修飾的)

圖 38W

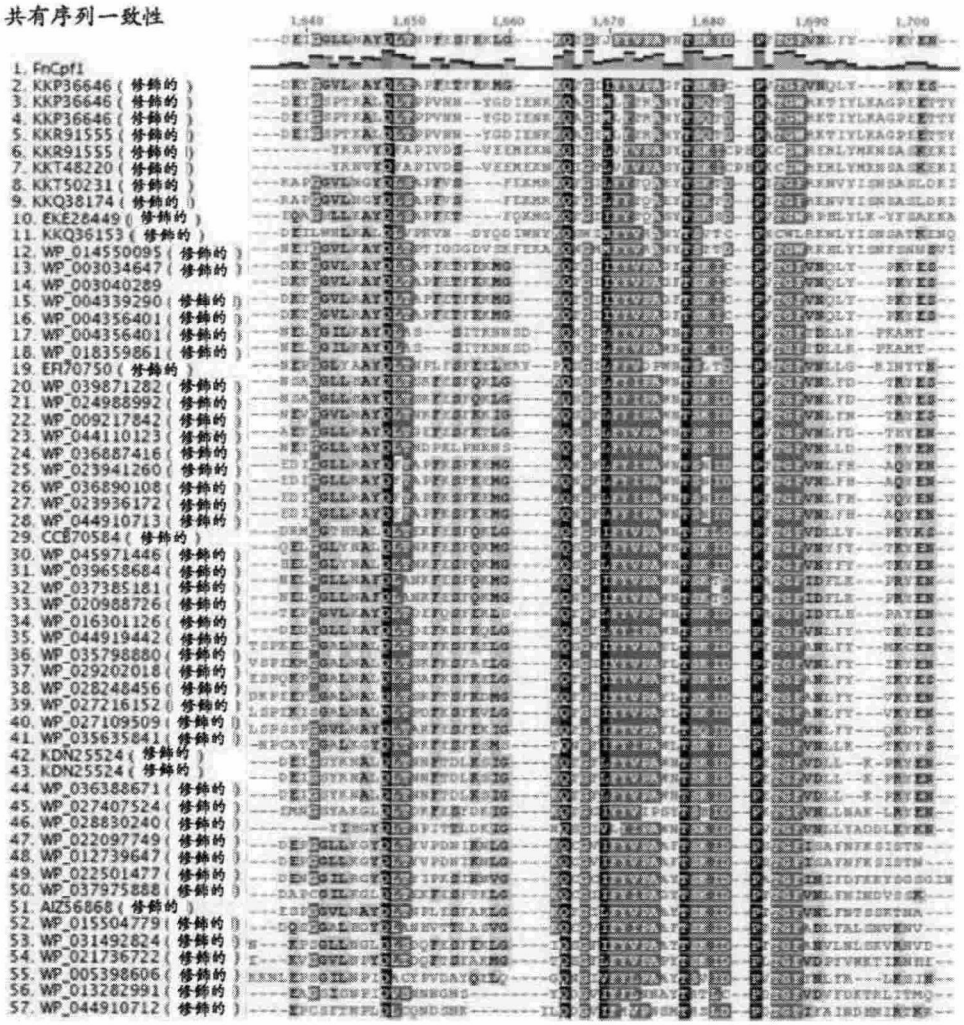

共有序列一致性
1,640
1,650
1,660
1,670
1,680
1,690
1,700
1. FnCpf1
2. KKP36646 (修飾的)
3. KKP36646 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKT50231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE28449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359861 (修飾的)
19. EFI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024988992 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890108 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044910713 (修飾的)
29. CCE70584 (修飾的)
30. WP_045971446 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037385181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301126 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975888 (修飾的)
51. AIZ56868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398606 (修飾的)
56. WP_013282991 (修飾的)
57. WP_044910712 (修飾的)

圖 38X

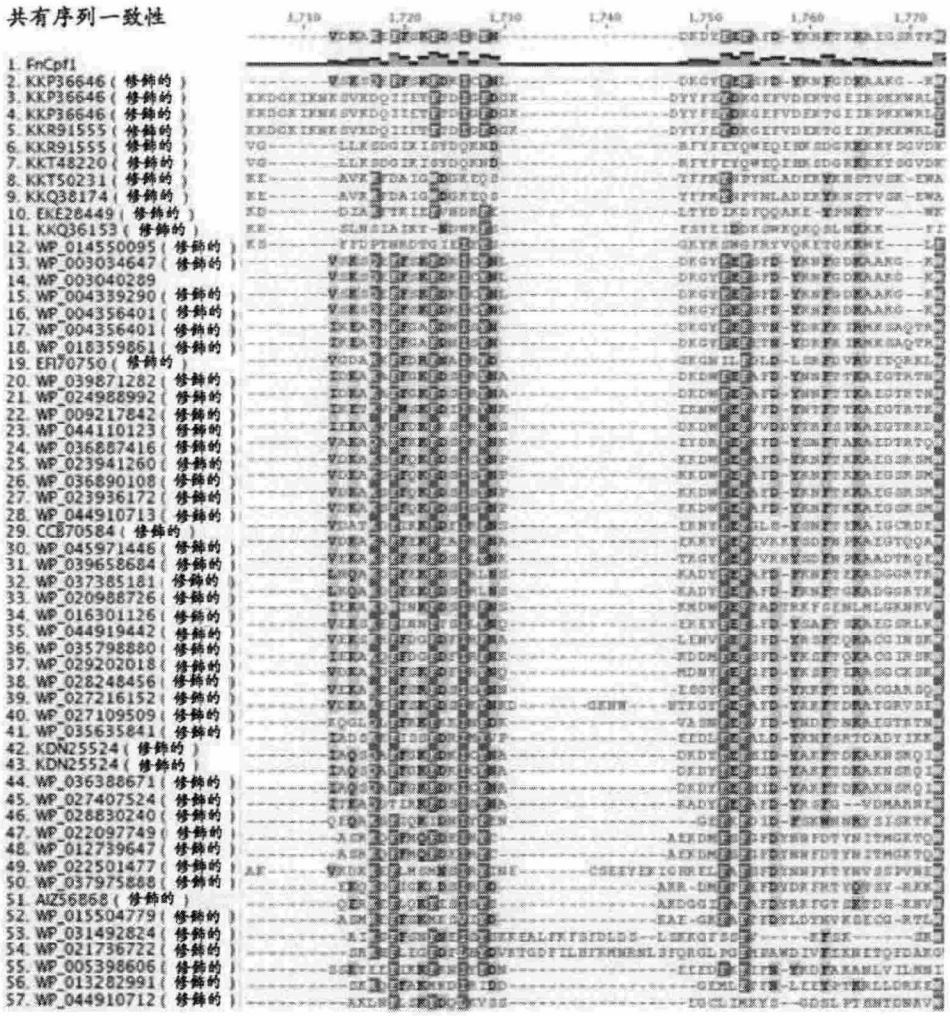

共有序列一致性
1,710
1,720
1,730
1,740
1,750
1,760
1,770
1. FnCpf1
2. KKP36646 (修飾的)
3. KKP36646 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKT50231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE28449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359861 (修飾的)
19. EFI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024988992 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890108 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044910713 (修飾的)
29. CCB70584 (修飾的)
30. WP_045971446 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037385181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301126 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975888 (修飾的)
51. AIZ56868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398606 (修飾的)
56. WP_013282991 (修飾的)
57. WP_044910712 (修飾的)

圖 38Y

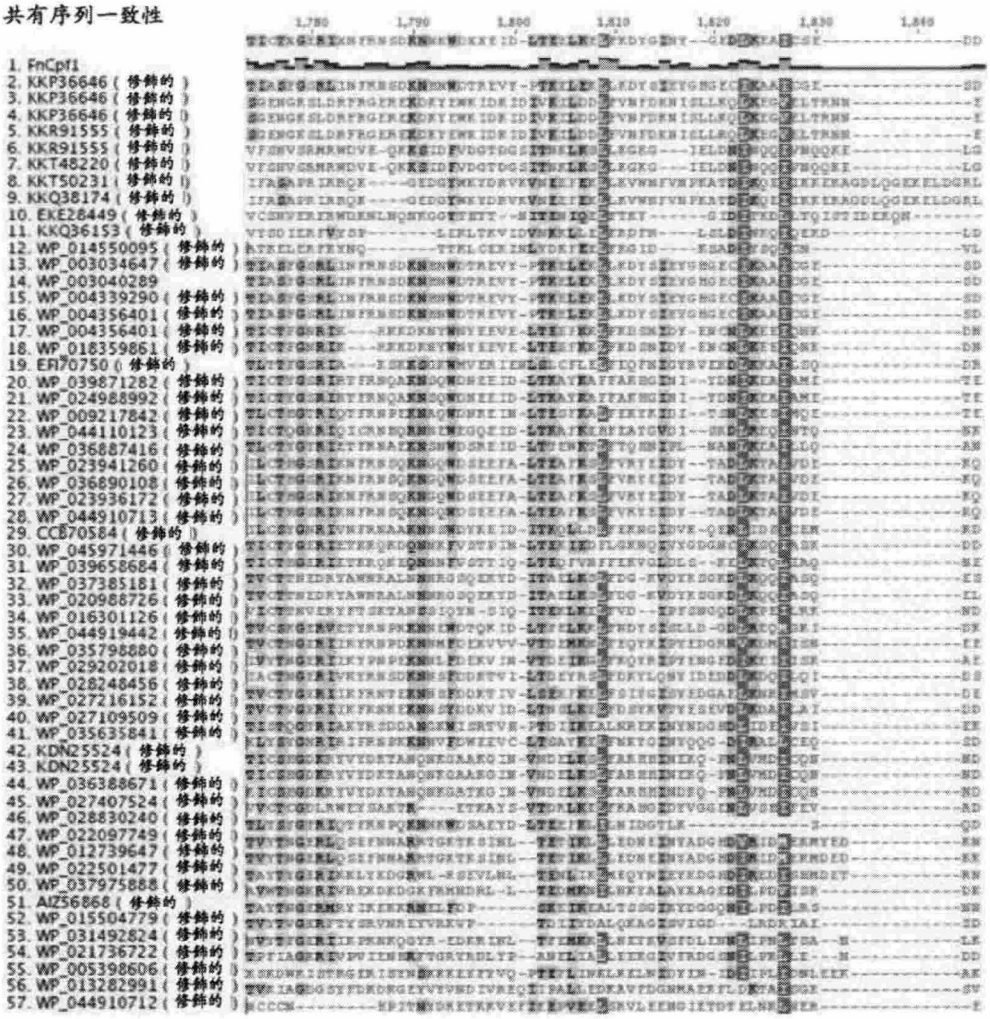

共有序列一致性
1,780
1,790
1,800
1,810
1,820
1,830
1,840
1. FnCpf1
2. KKP36646 (修飾的)
3. KKP36646 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKT50231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE28449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359861 (修飾的)
19. EFI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024988992 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890108 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044910713 (修飾的)
29. CCB70584 (修飾的)
30. WP_045971446 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037385181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301126 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975888 (修飾的)
51. AIZ56868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398606 (修飾的)
56. WP_013282991 (修飾的)
57. WP_044910712 (修飾的)

圖 38Z

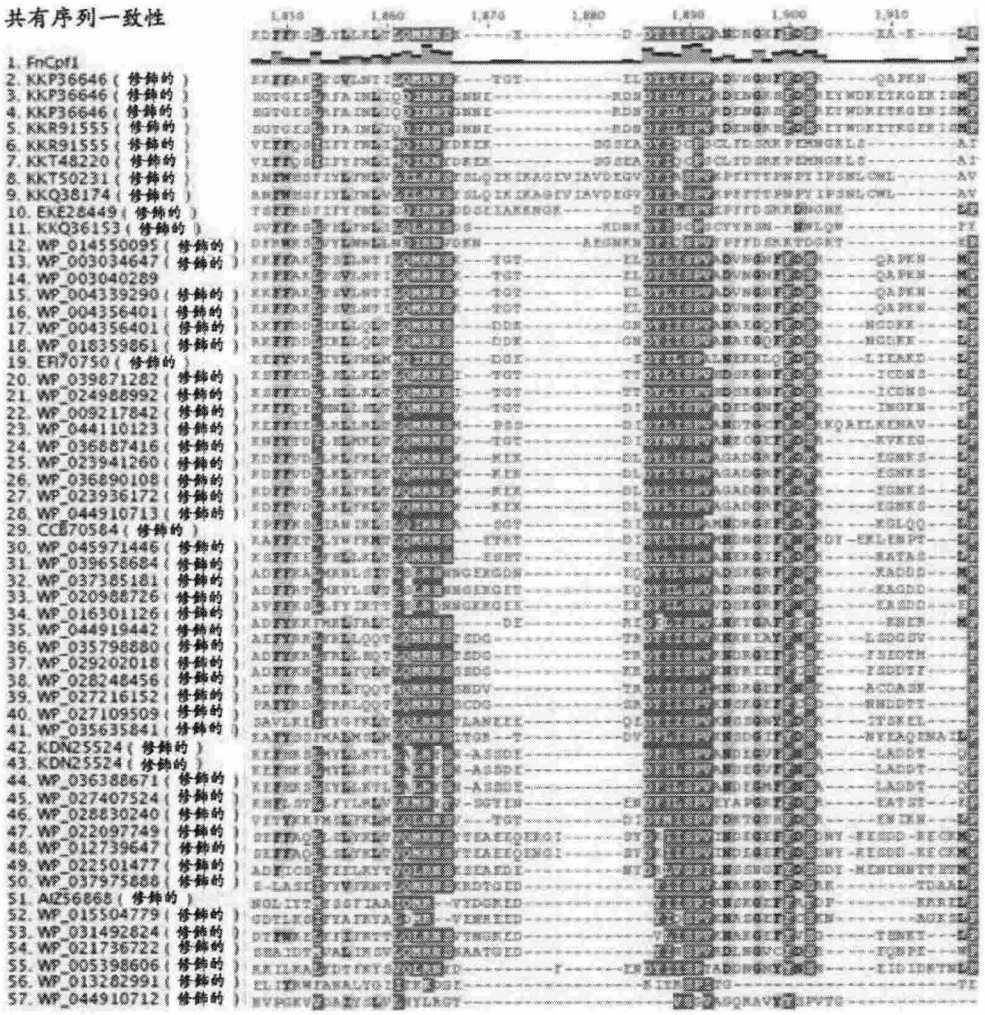

共有序列一致性
1. FnCpf1
2. KKP36646 (修飾的)
3. KKP36646 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKT50231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE28449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359861 (修飾的)
19. ERI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024988992 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890108 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044910713 (修飾的)
29. CCE70584 (修飾的)
30. WP_045971446 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037385181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301126 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975888 (修飾的)
51. AIZ56868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398606 (修飾的)
56. WP_013282991 (修飾的)
57. WP_044910712 (修飾的)

圖 38AA

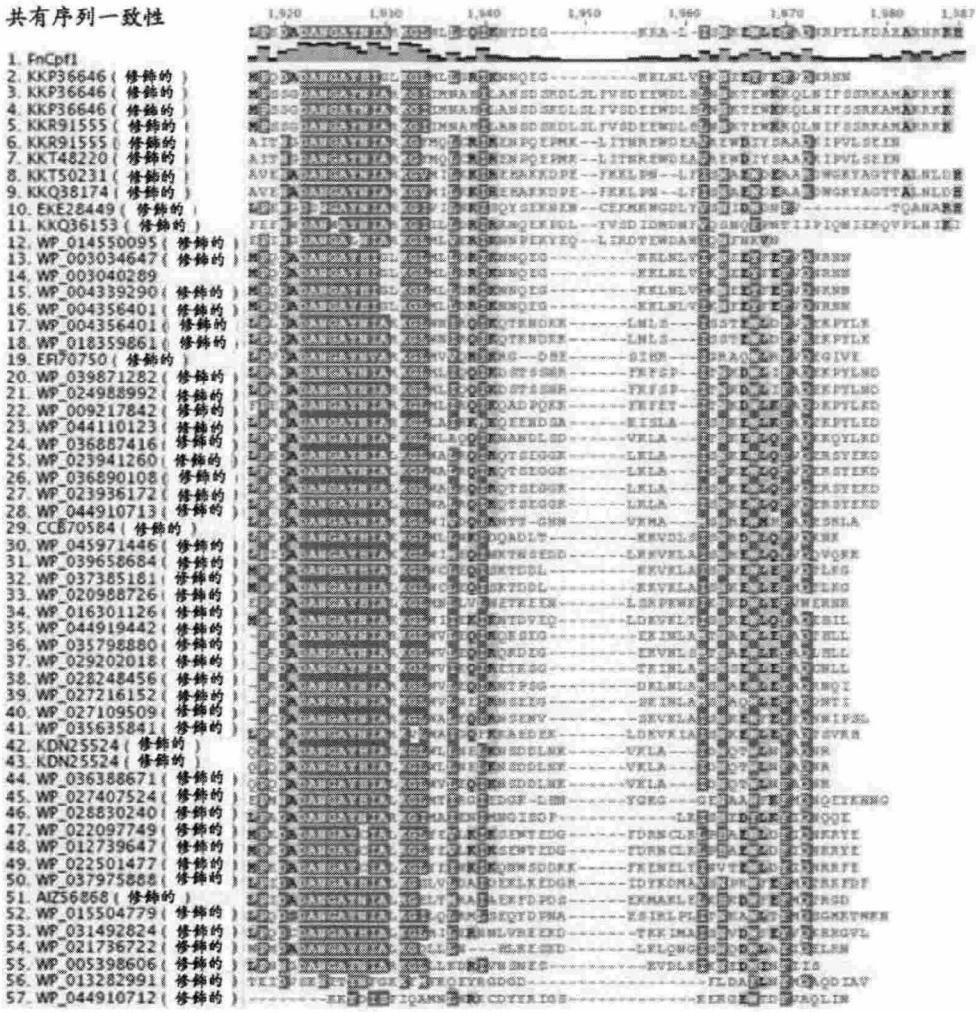

共有序列一致性
1. FnCpf1
2. KKP36646 (修飾的)
3. KKP36646 (修飾的)
4. KKP36646 (修飾的)
5. KKR91555 (修飾的)
6. KKR91555 (修飾的)
7. KKT48220 (修飾的)
8. KKT50231 (修飾的)
9. KKQ38174 (修飾的)
10. EKE28449 (修飾的)
11. KKQ36153 (修飾的)
12. WP_014550095 (修飾的)
13. WP_003034647 (修飾的)
14. WP_003040289
15. WP_004339290 (修飾的)
16. WP_004356401 (修飾的)
17. WP_004356401 (修飾的)
18. WP_018359861 (修飾的)
19. EFI70750 (修飾的)
20. WP_039871282 (修飾的)
21. WP_024988992 (修飾的)
22. WP_009217842 (修飾的)
23. WP_044110123 (修飾的)
24. WP_036887416 (修飾的)
25. WP_023941260 (修飾的)
26. WP_036890108 (修飾的)
27. WP_023936172 (修飾的)
28. WP_044910713 (修飾的)
29. CCE70584 (修飾的)
30. WP_045971446 (修飾的)
31. WP_039658684 (修飾的)
32. WP_037385181 (修飾的)
33. WP_020988726 (修飾的)
34. WP_016301126 (修飾的)
35. WP_044919442 (修飾的)
36. WP_035798880 (修飾的)
37. WP_029202018 (修飾的)
38. WP_028248456 (修飾的)
39. WP_027216152 (修飾的)
40. WP_027109509 (修飾的)
41. WP_035635841 (修飾的)
42. KDN25524 (修飾的)
43. KDN25524 (修飾的)
44. WP_036388671 (修飾的)
45. WP_027407524 (修飾的)
46. WP_028830240 (修飾的)
47. WP_022097749 (修飾的)
48. WP_012739647 (修飾的)
49. WP_022501477 (修飾的)
50. WP_037975888 (修飾的)
51. AIZ56868 (修飾的)
52. WP_015504779 (修飾的)
53. WP_031492824 (修飾的)
54. WP_021736722 (修飾的)
55. WP_005398606 (修飾的)
56. WP_013282991 (修飾的)
57. WP_044910712 (修飾的)

圖 38BB

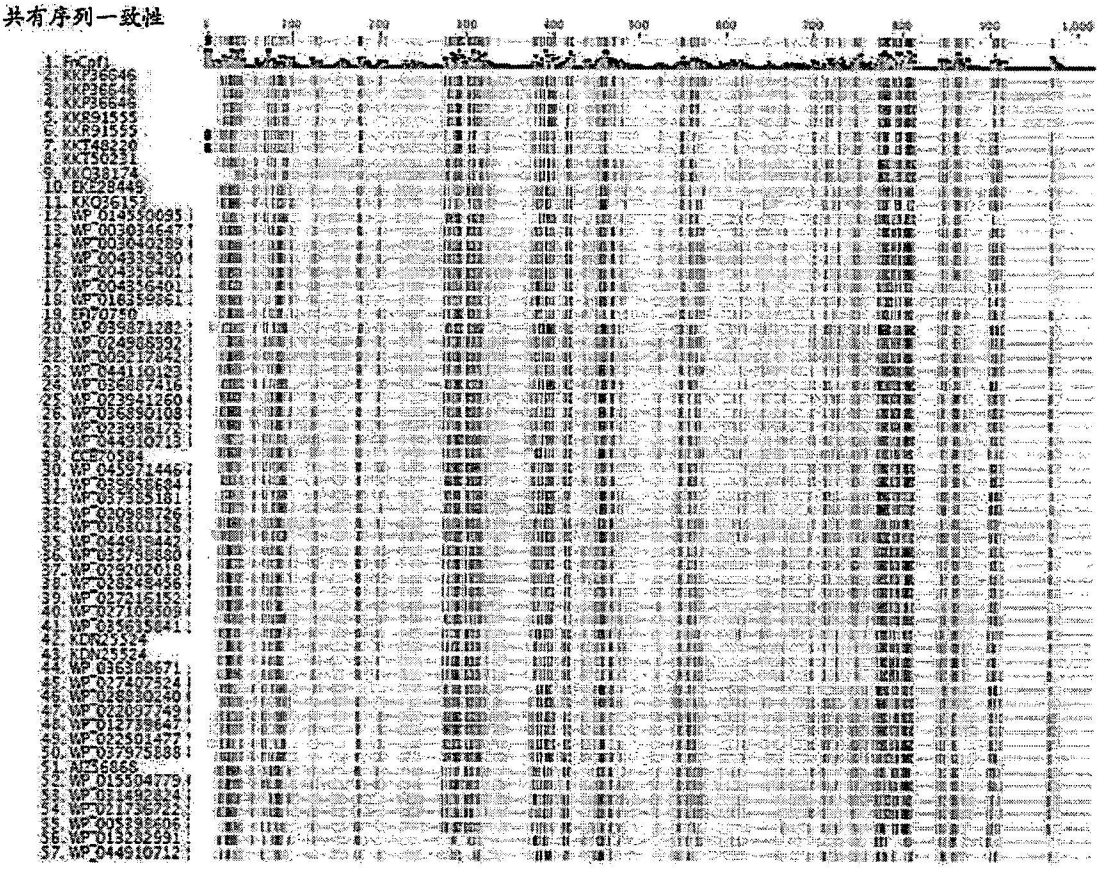


圖 39A

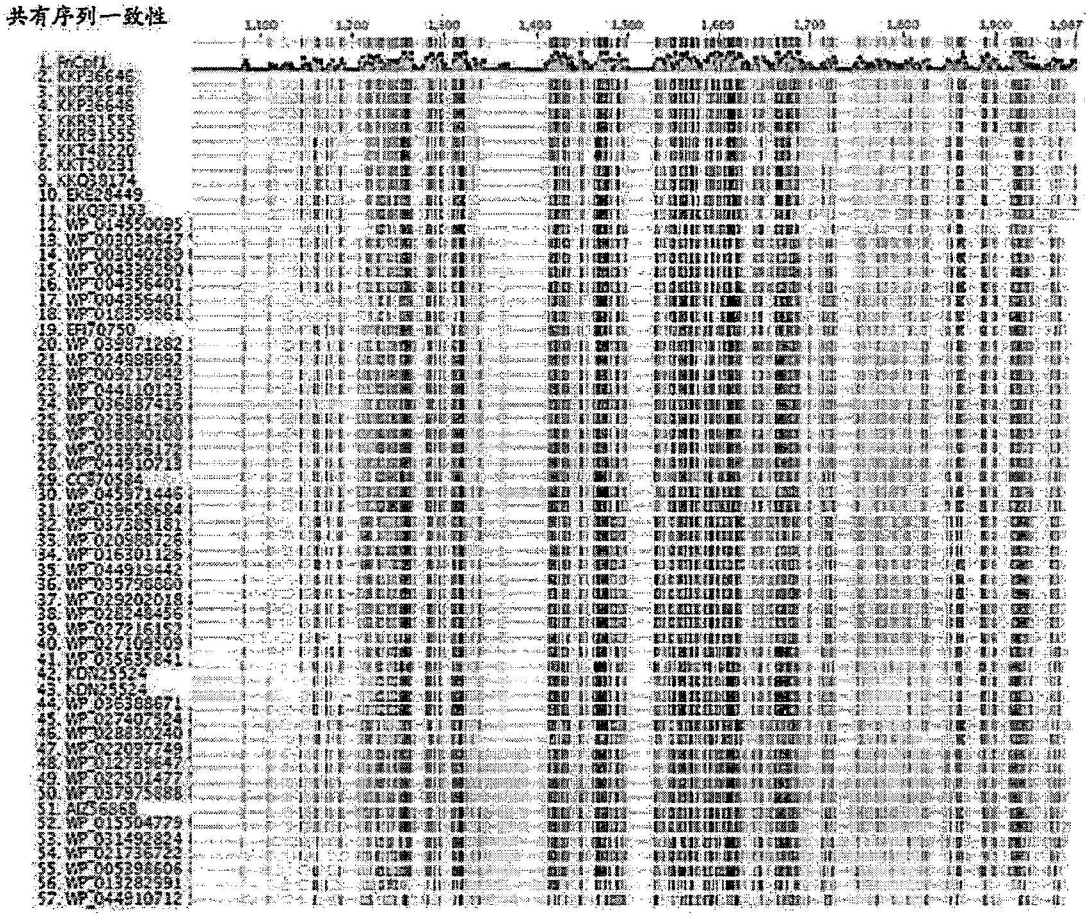

共有序列一致性
1.100
1.200
1.300
1.400
1.500
1.600
1.700
1.800
1.900
1.987
1. PnCpf1
2. KKP36646
3. KKP36646
4. KKP36646
5. KKR91555
6. KKR91555
7. KKT48220
8. KKT50231
9. KKQ38174
10. EKE28449
11. RKQ35153
12. WP_014550095
13. WP_003034647
14. WP_003040289
15. WP_004339290
16. WP_004356401
17. WP_004356401
18. WP_018359861
19. EFI70750
20. WP_039871282
21. WP_024988992
22. WP_009217842
23. WP_044110123
24. WP_036887416
25. WP_023941260
26. WP_036890108
27. WP_023936172
28. WP_044910713
29. CCS70584
30. WP_045971446
31. WP_039658684
32. WP_037385181
33. WP_020988726
34. WP_016301126
35. WP_044919442
36. WP_035798880
37. WP_029202018
38. WP_028248456
39. WP_027216152
40. WP_027109509
41. WP_035635841
42. KDN25524
43. KDN25524
44. WP_036388671
45. WP_027407524
46. WP_028830240
47. WP_022097749
48. WP_012739647
49. WP_022501477
50. WP_037975888
51. AIZ56868
52. WP_015504779
53. WP_031492824
54. WP_021736722
55. WP_005398606
56. WP_013282591
57. WP_044910712

圖 39B

pACYC184 FnCpf1座位(pY001)

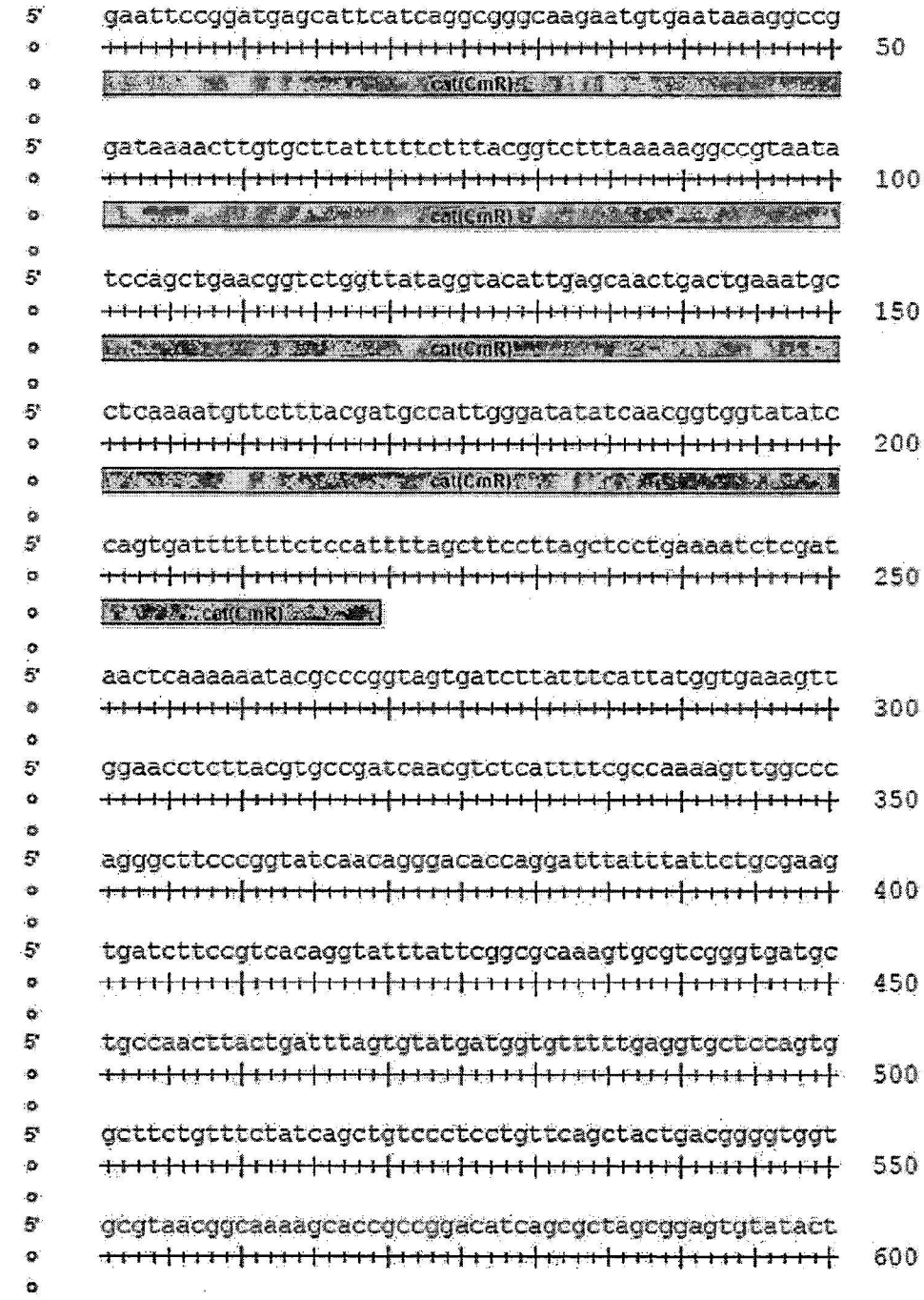


圖 40A

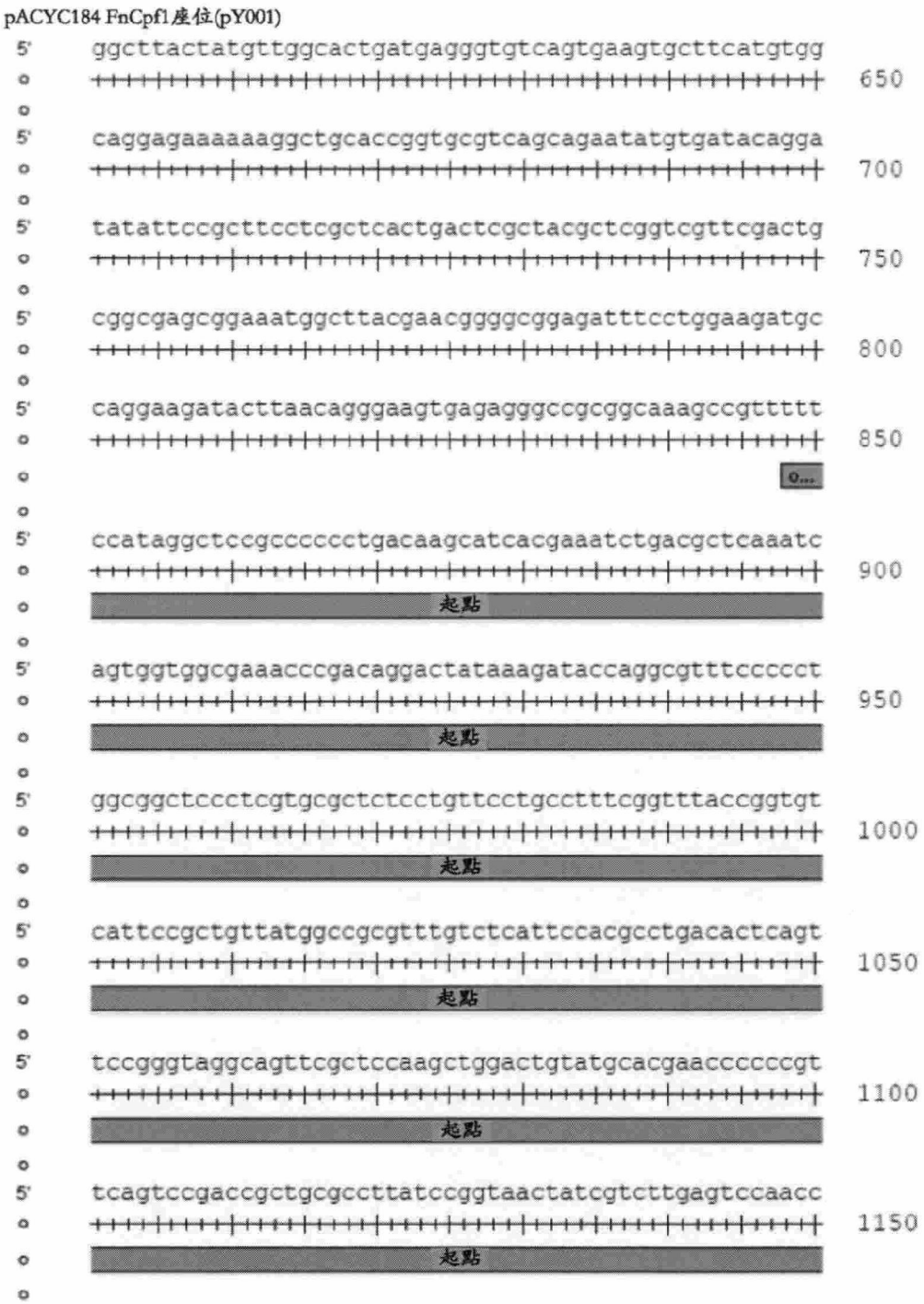

pACYC184 FnCpf1座位(pY001)
5' ggcttactatgttggcactgatgagggtgtcagtgaagtgcttcatgtgg
650
5' caggagaaaaaaggctgcaccggtgcgtcagcagaatatgtgatacagga
700
5' tatattccgcttcctcgctcactgactcgctacgctcggtcgttcgactg
750
5' cggcgagcggaaatggcttacgaacggggcggagatttcctggaagatgc
800
5' caggaagatacttaacagggaagtgagagggccgcggcaaagccgttttt
850
5' ccataggctccgccccctgacaagcatcacgaaatctgacgctcaaatc
900
起點
5' agtggtggcgaaacccgacaggactataaagataccaggcgtttccccct
950
起點
5' ggcggctccctcgtgcgctctcctgttcctgcctttcggtttaccggtgt
1000
起點
5' cattccgctgttatggccgcgtttgtctcattccacgcctgacactcagt
1050
起點
5' tccgggtaggcagttcgctccaagctggactgtatgcacgaacccccgt
1100
起點
5' tcagtccgaccgctgcgccttatccggtaactatcgtcttgagtccaacc
1150
起點

圖 40B

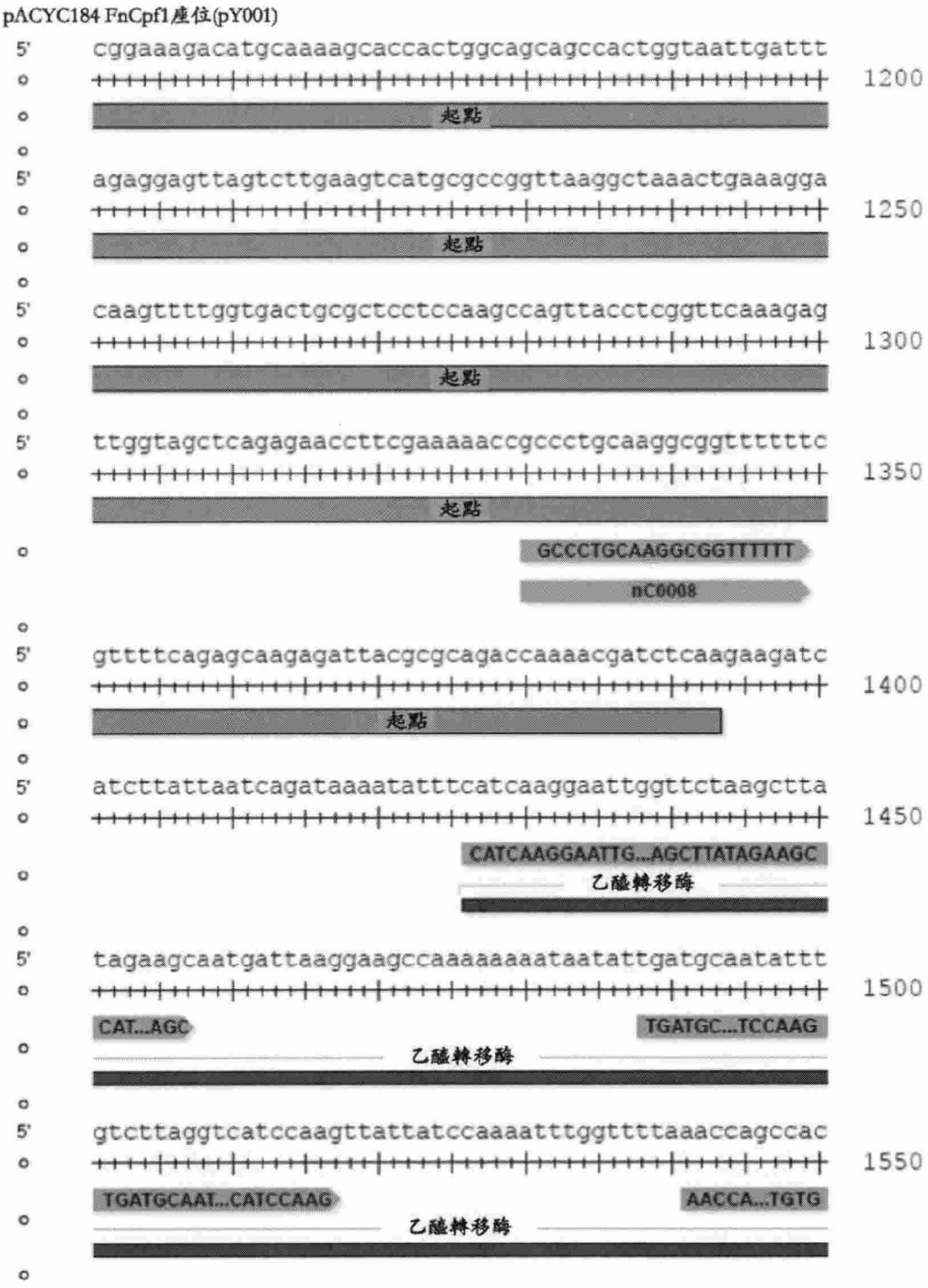

pACYC184 FnCpf1座位(pY001)
5'
cggaaagacatgcaaaagcaccactggcagcagccactggtaattgattt
1200
起點
5'
agaggagttagtcttgaagtcatgcgccggttaaggctaaactgaaagga
1250
起點
5'
caagttttggtgactgcgctcctccaagccagttacctcggttcaaagag
1300
起點
5'
ttggtagctcagagaaccttcgaaaaaccgccctgcaaggcggttttttc
1350
起點
GCCCTGCAAGGCGGTTTTTT
nC0008
5'
gttttcagagcaagagattacgcgcagaccaaaacgatctcaagaagatc
1400
起點
5'
atcttattaatcagataaaatatttcatcaaggaattggttctaagctta
1450
CATCAAGGAATTG...AGCTTATAGAAGC
乙醯轉移酶
5'
tagaagcaatgattaaggaagccaaaaaaaaataatattgatgcaatattt
1500
CAT...AGC
TGATGC...TCCAAG
乙醯轉移酶
5'
gtcttaggtcatccaagttattatccaaaatttggttttaaaccagccac
1550
TGATGCAAT...CATCCAAG
AACCA...TGTG
乙醯轉移酶

圖 40C

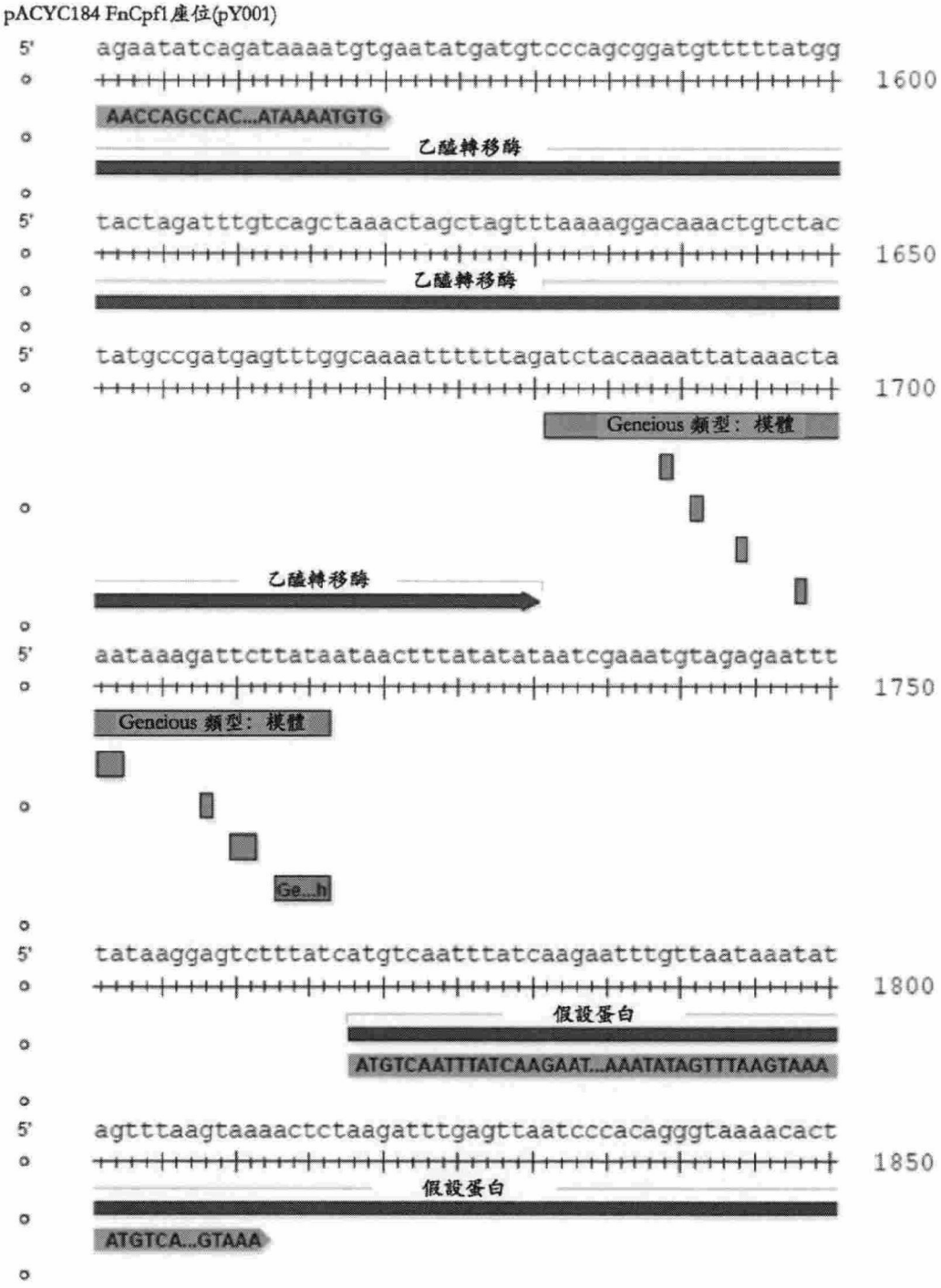

pACYC184 FnCpf1座位(pY001)
5' agaatatcagataaaatgtgaatatgatgtcccagcggatgtttttatgg
1600
AACCAGCCAC...ATAAAATGTG
乙醯轉移酶
5' tactagatttgtcagctaaactagctagtttaaaaggacaaactgtctac
1650
乙醯轉移酶
5' tatgccgatgagtttggcaaaatttttttagatctacaaaattataaacta
1700
Geneious 類型: 模體
乙醯轉移酶
5' aataaagattcttataataactttatatataatcgaaatgtagagaattt
1750
Geneious 類型: 模體
Ge...h
5' tataaggagtctttatcatgtcaatttatcaagaatttgttaataaatat
1800
假設蛋白
ATGTCAATTTATCAAGAAT...AAATATAGTTTAAGTAAA
5' agtttaagtaaaactctaagatttgagttaatcccacagggtaaaacact
1850
假設蛋白
ATGTCA...GTAAA

圖 40D

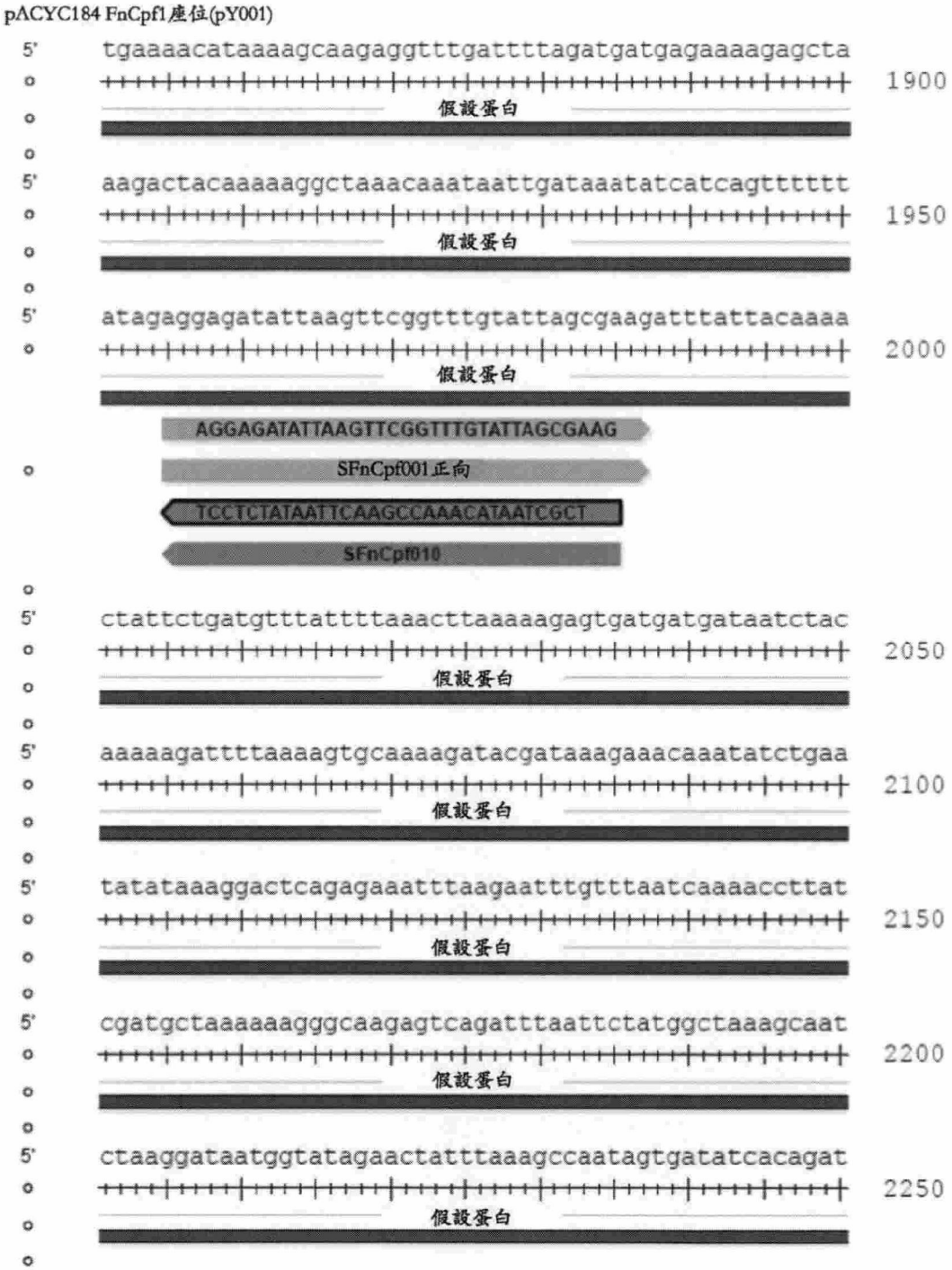

pACYC184 FnCpf1座位(pY001)
5' tgaaaacataaaagcaagaggtttgattttagatgatgagaaaagagcta
1900
假設蛋白
5' aagactacaaaaaggctaaacaaataattgataaatatcatcagtttttt
1950
假設蛋白
5' atagaggagatattaagttcggtttgtattagcgaagatttattacaaaa
2000
假設蛋白
AGGAGATATTAAGTTCGGTTTGTATTAGCGAAG
SFnCpf001正向
TCCTCTATAATTCAAGCCAAACATAATCGCT
SFnCpf010
5' ctattctgatgtttatttttaaacttaaaaaagagtgatgatgataatctac
2050
假設蛋白
5' aaaaagattttaaaagtgcaaaagatacgataaagaaacaaatatctgaa
2100
假設蛋白
5' tatataaaggactcagagaaatttaagaatttgtttaatcaaaaccttat
2150
假設蛋白
5' cgatgctaaaaaagggcaagagtcagatttaattctatggctaaagcaat
2200
假設蛋白
5' ctaaggataatggtatagaactatttaaagccaatagtgatatcacagat
2250
假設蛋白

圖 40E

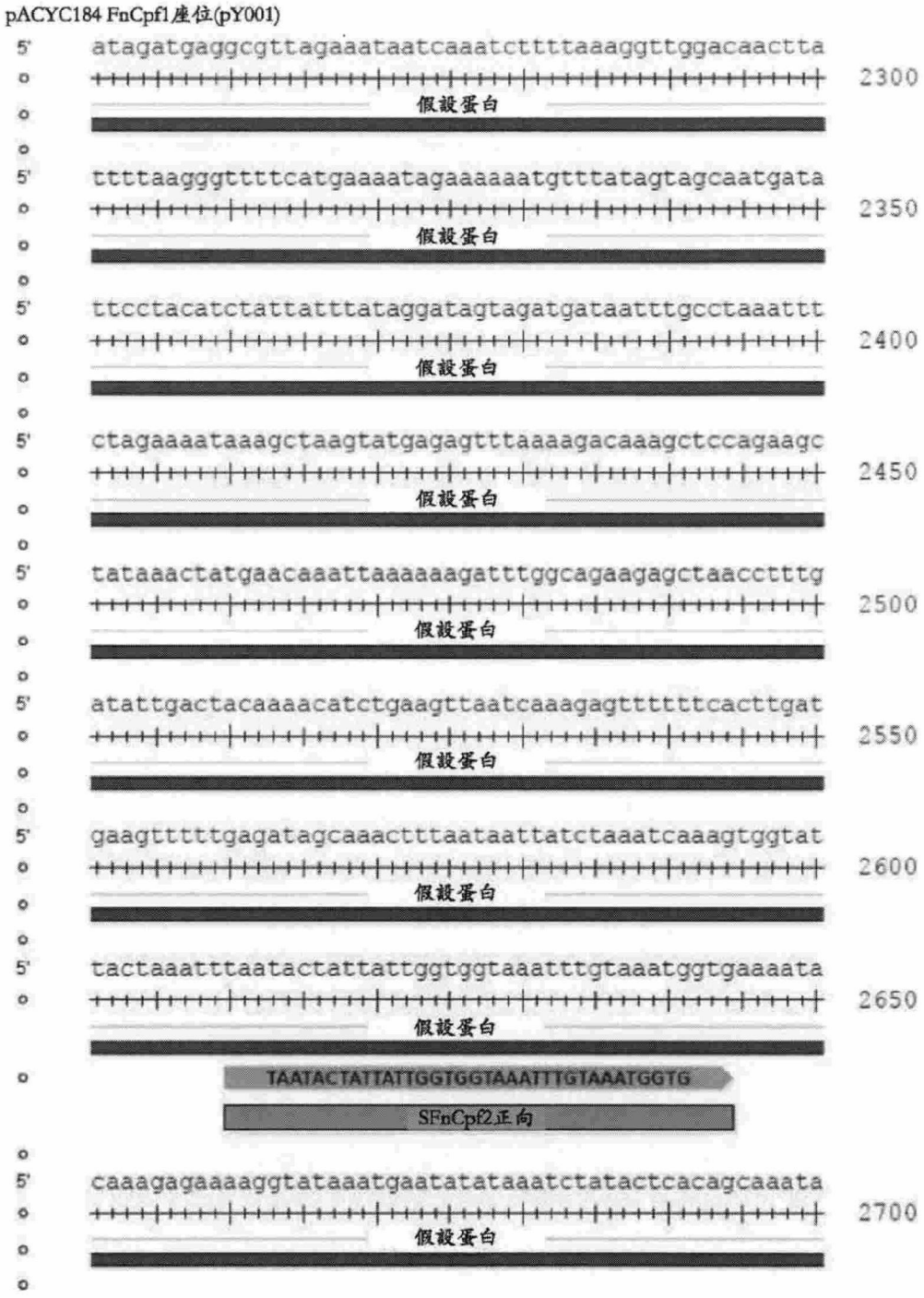

pACYC184 FnCpf1座位(pY001)
5' atagatgaggcgttagaaataatcaaatcttttaaaggttggacaactta
2300
假設蛋白
5' ttttaagggttttcatgaaaatagaaaaaatgtttatagtagcaatgata
2350
假設蛋白
5' ttcctacatctattatttataggatagtagatgataatttgcctaaattt
2400
假設蛋白
5' ctagaaaataaagctaagtatgagagtttaaaagacaaagctccagaagc
2450
假設蛋白
5' tataaactatgaacaaattaaaaaagatttggcagaagagctaacctttg
2500
假設蛋白
5' atattgactacaaaacatctgaagttaatcaaagagtttttcacttgat
2550
假設蛋白
5' gaagtttttgagatagcaaactttaataattatctaaatcaaagtggtat
2600
假設蛋白
5' tactaaatttaatactattattggtggtaaatttgtaaatggtgaaaata
2650
假設蛋白
TAATACTATTATTGGTGGTAAATTTGTAAATGGTG
SFnCpf2正向
5' caaagagaaaaggtataaatgaatatataaatctatactcacagcaaata
2700
假設蛋白

圖 40F

pACYC184 FnCpf1座位(pY001)

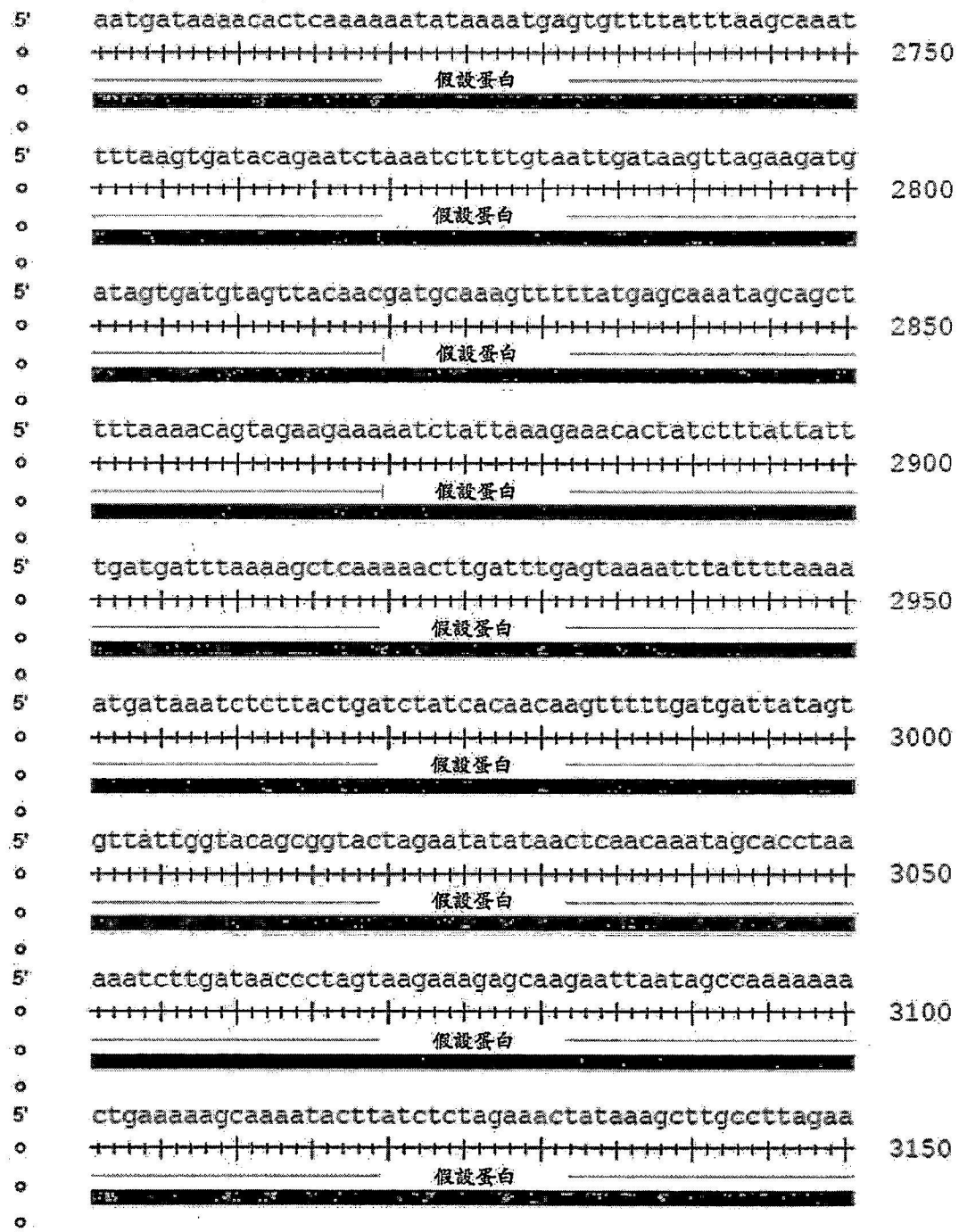


圖 40G

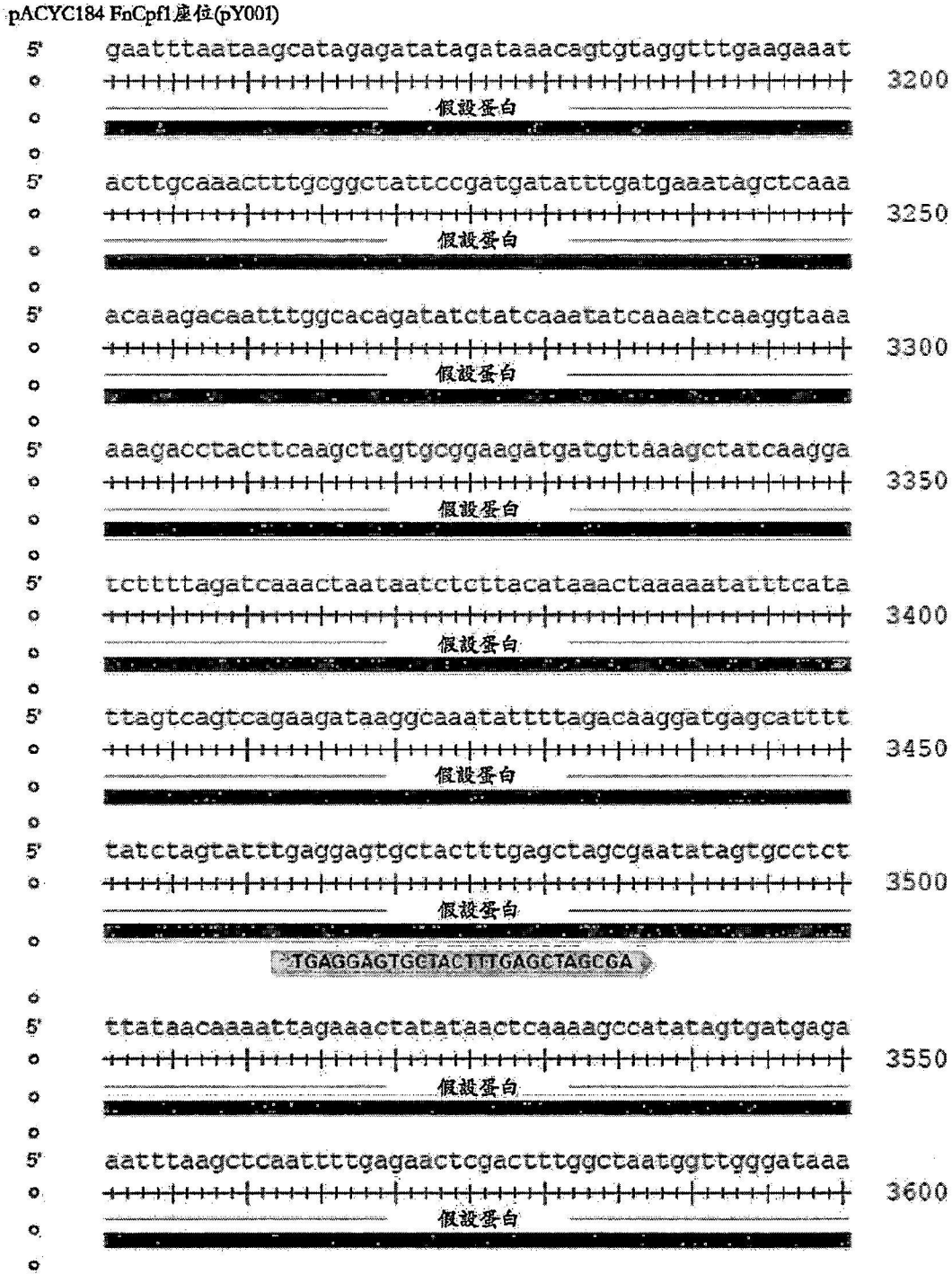

pACYC184 FnCpf1座位(pY001)
5' gaatttaataagcatagagatatagataaacagtgtaggtttgaagaaat
3200
假設蛋白
5' acttgcaaactttgcggctattccgatgatatttgatgaaatagctcaaa
3250
假設蛋白
5' acaaagacaatttggcacagatatctatcaaatatcaaaatcaaggtaaa
3300
假設蛋白
5' aaagacctacttcaagctagtgcggaagatgatgttaaagctatcaagga
3350
假設蛋白
5' tcttttagatcaaactaataatctcttacataaactaaaaatatttcata
3400
假設蛋白
5' ttagtcagtcagaagataaggcaaatattttagacaaggatgagcatttt
3450
假設蛋白
5' tatctagtatttgaggagtgctactttgagctagcgaatatagtgcctct
3500
假設蛋白
TGAGGAGTGCTACTTTGAGCTAGCGA
5' ttataacaaaattagaaactatataactcaaaagccatatagtgatgaga
3550
假設蛋白
5' aatttaagctcaattttgagaactcgactttggctaatggttgggataaa
3600
假設蛋白

圖 40H

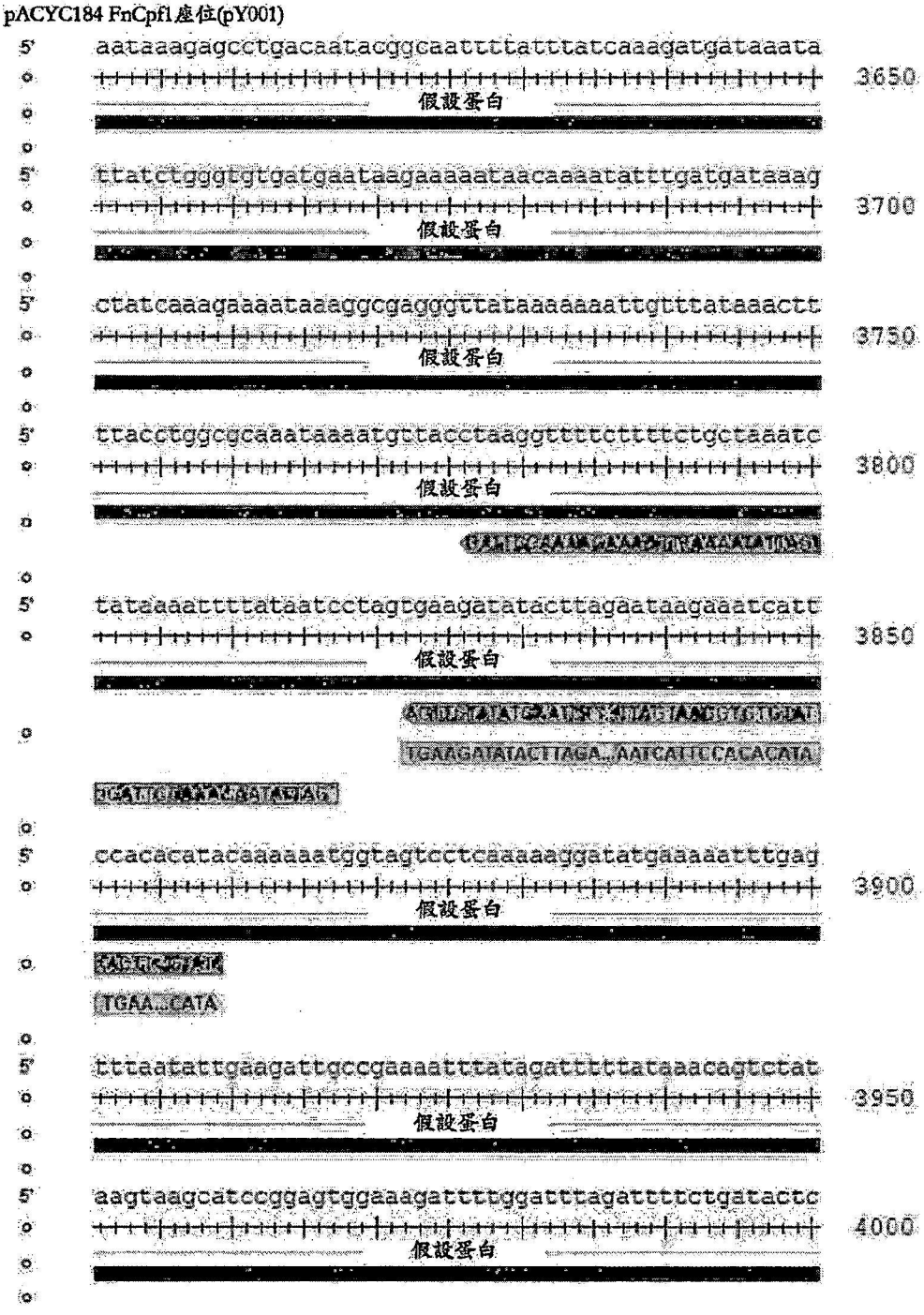

pACYC184 FnCpf1座位(pY001)
5'
aataaagagcctgacaatacggcaattttatttatcaaagatgataaata
3650
假設蛋白
ttatctgggtgtgatgaataagaaaaataacaaaatatttgatgataaag
3700
假設蛋白
ctatcaaagaaaataaggcgagggttataaaaaaattgtttataaactt
3750
假設蛋白
ttacctggcgcaaataaaatgttacctaaggttttcttttctgctaaatc
3800
假設蛋白
tataaaattttataatcctagtgaagatatacttagaataagaaatcatt
3850
假設蛋白
TGAAGATATACTTAGA...AATCATTCCACACATA
ccacacatacaaaaaatggtagtcctcaaaaaggatatgaaaaatttgag
3900
假設蛋白
TGAA...CATA
tttaatattgaagattgccgaaaatttatagattttataaacagtctat
3950
假設蛋白
aagtaagcatccggagtggaaagattttggatttagattttctgatactc
4000
假設蛋白

圖 40I

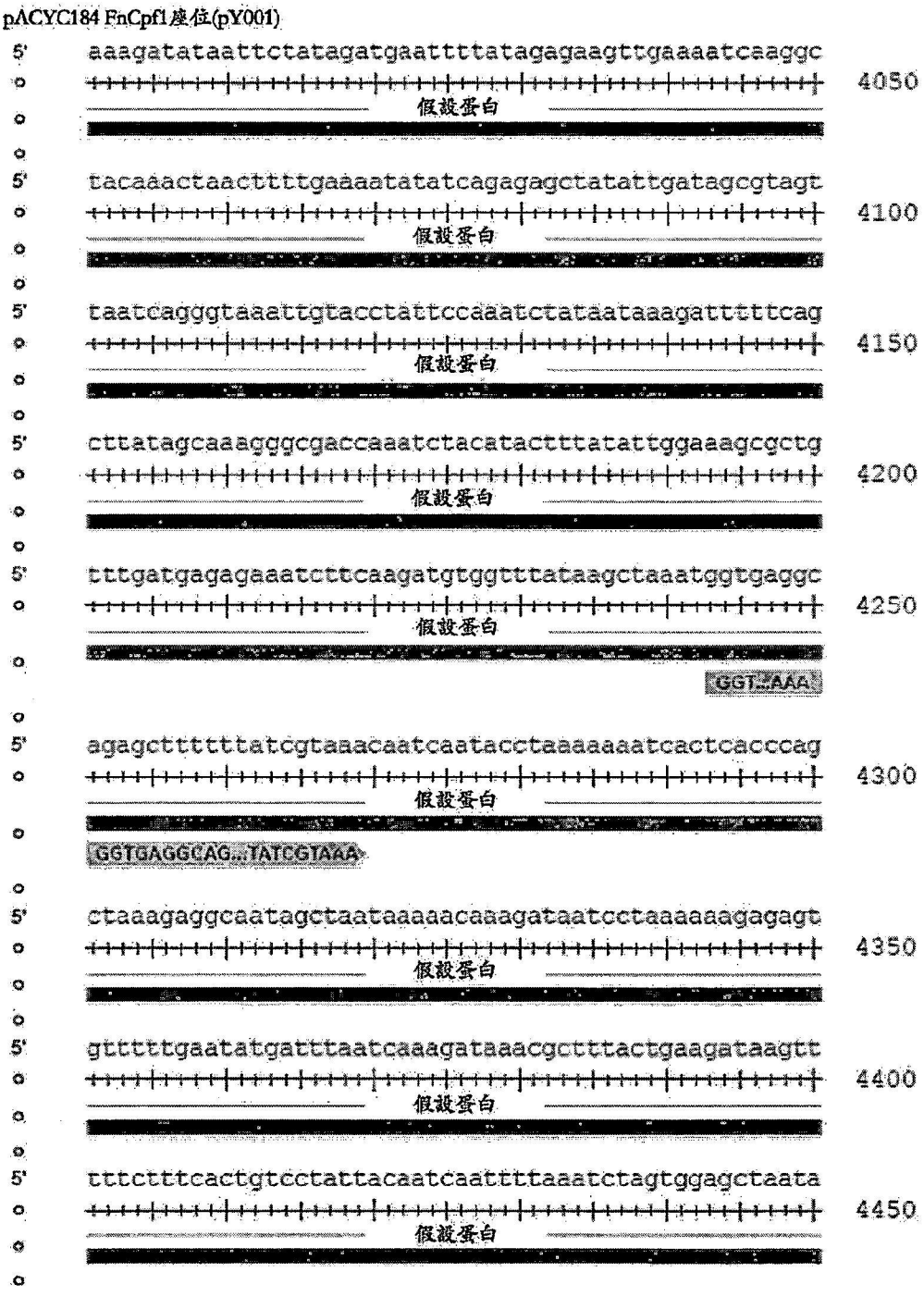

pACYC184 FnCpf1座位(pY001)
5'
aaagatataattctatagatgaattttatagagaagttgaaaatcaaggc
4050
假設蛋白
5'
tacaaactaacttttgaaaatatatcagagagctatattgatagcgtagt
4100
假設蛋白
5'
taatcagggtaaattgtacctattccaaatctataataaagatttttcag
4150
假設蛋白
5'
cttatagcaaagggcgaccaaatctacatactttatattggaaagcgctg
4200
假設蛋白
5'
tttgatgagagaaatcttcaagatgtggtttataagctaaatggtgaggc
4250
假設蛋白
GGT...AAA
5'
agagctttttatcgtaaacaatcaatacctaaaaaaatcactcacccag
4300
假設蛋白
GGTGAGGCAG...TATCGTAAA
5'
ctaaagaggcaatagctaataaaaacaaagataatcctaaaaaagagagt
4350
假設蛋白
5'
gtttttgaatatgatttaatcaaagataaacgctttactgaagataagtt
4400
假設蛋白
5'
tttctttcactgtcctattacaatcaattttaaatctagtggagctaata
4450
假設蛋白

圖 40J

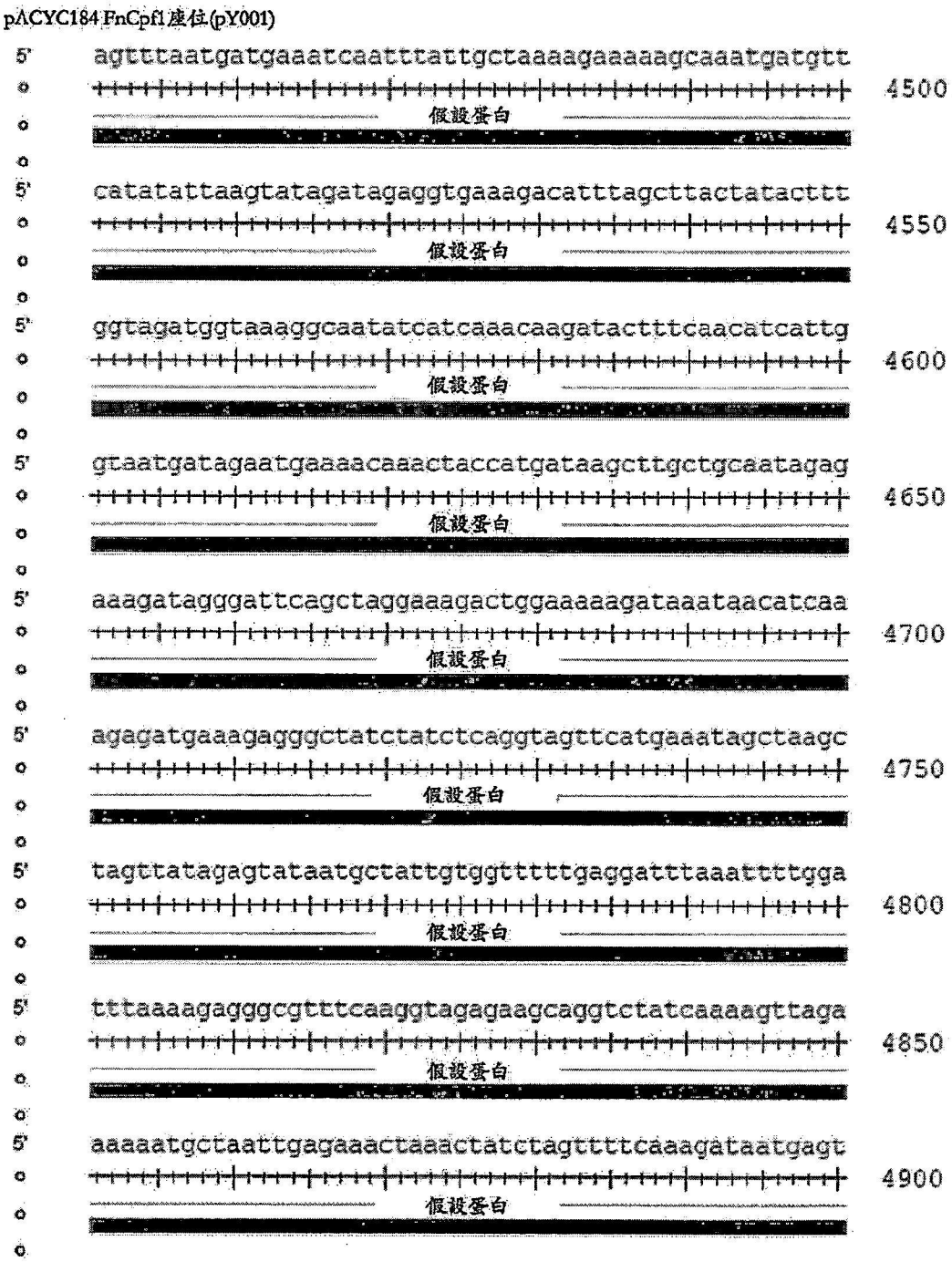

pACYC184 FnCpf1 座位(pY001)
5' agtttaatgatgaaatcaatttattgctaaaagaaaaagcaaatgatgtt
4500
假設蛋白
5' catatattaagtatagatagaggtgaaagacatttagcttactatacttt
4550
假設蛋白
5' ggtagatggtaaaggcaatatcatcaaacaagatactttcaacatcattg
4600
假設蛋白
5' gtaatgatagaatgaaaacaaactaccatgataagcttgctgcaatagag
4650
假設蛋白
5' aaagatagggattcagctaggaaagactggaaaaagataaataacatcaa
4700
假設蛋白
5' agagatgaaagagggctatctatctcaggtagttcatgaaatagctaagc
4750
假設蛋白
5' tagttatagagtataatgctattgtggtttttgaggatttaaattttgga
4800
假設蛋白
5' tttaaaagagggcgtttcaaggtagagaagcaggtctatcaaaagttaga
4850
假設蛋白
5' aaaaatgctaattgagaaactaaactatctagttttcaaagataatgagt
4900
假設蛋白

圖 40K

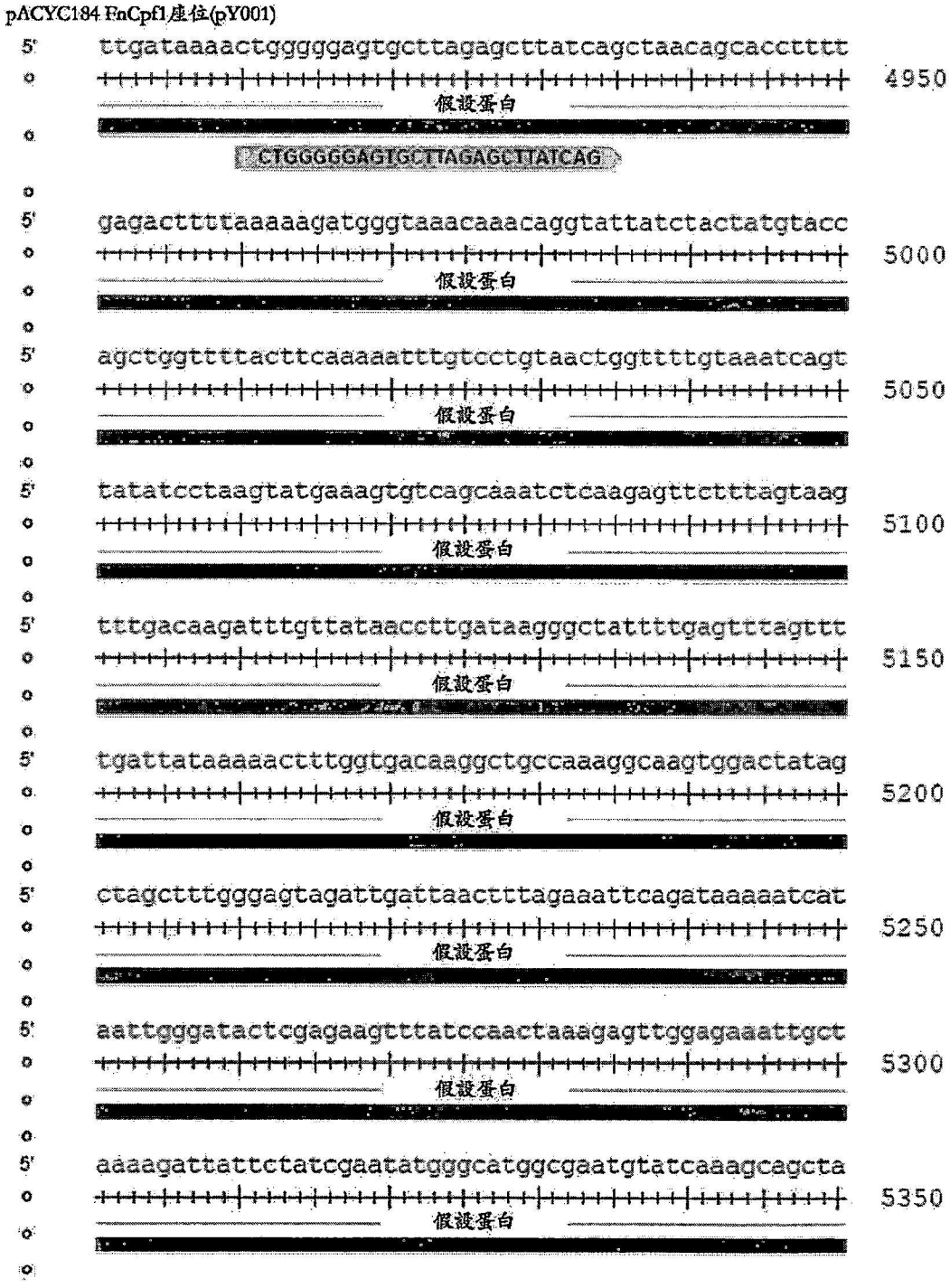

pACYC184 FnCpf1座位(pY001)
5'
ttgataaaactgggggagtgcttagagcttatcagctaacagcacctttt
4950
假設蛋白
CTGGGGGAGTGCTTAGAGCTTATCAG
gagacttttaaaaagatgggtaaacaaacaggtattatctactatgtacc
5000
假設蛋白
agctggttttacttcaaaaatttgtcctgtaactggttttgtaaatcagt
5050
假設蛋白
tatatcctaagtatgaaagtgtcagcaaatctcaagagttctttagtaag
5100
假設蛋白
tttgacaagatttgttataaccttgataagggctattttgagtttagttt
5150
假設蛋白
tgattataaaaactttggtgacaaggctgccaaaggcaagtggactatag
5200
假設蛋白
ctagctttgggagtagattgattaactttagaaattcagataaaaatcat
5250
假設蛋白
aattgggatactcgagaagtttatccaactaaagagttggagaaattgct
5300
假設蛋白
aaaagattattctatcgaatatgggcatggcgaatgtatcaaagcagcta
5350
假設蛋白

圖 40L

pACYC184-FnCpf1座位(pY001)

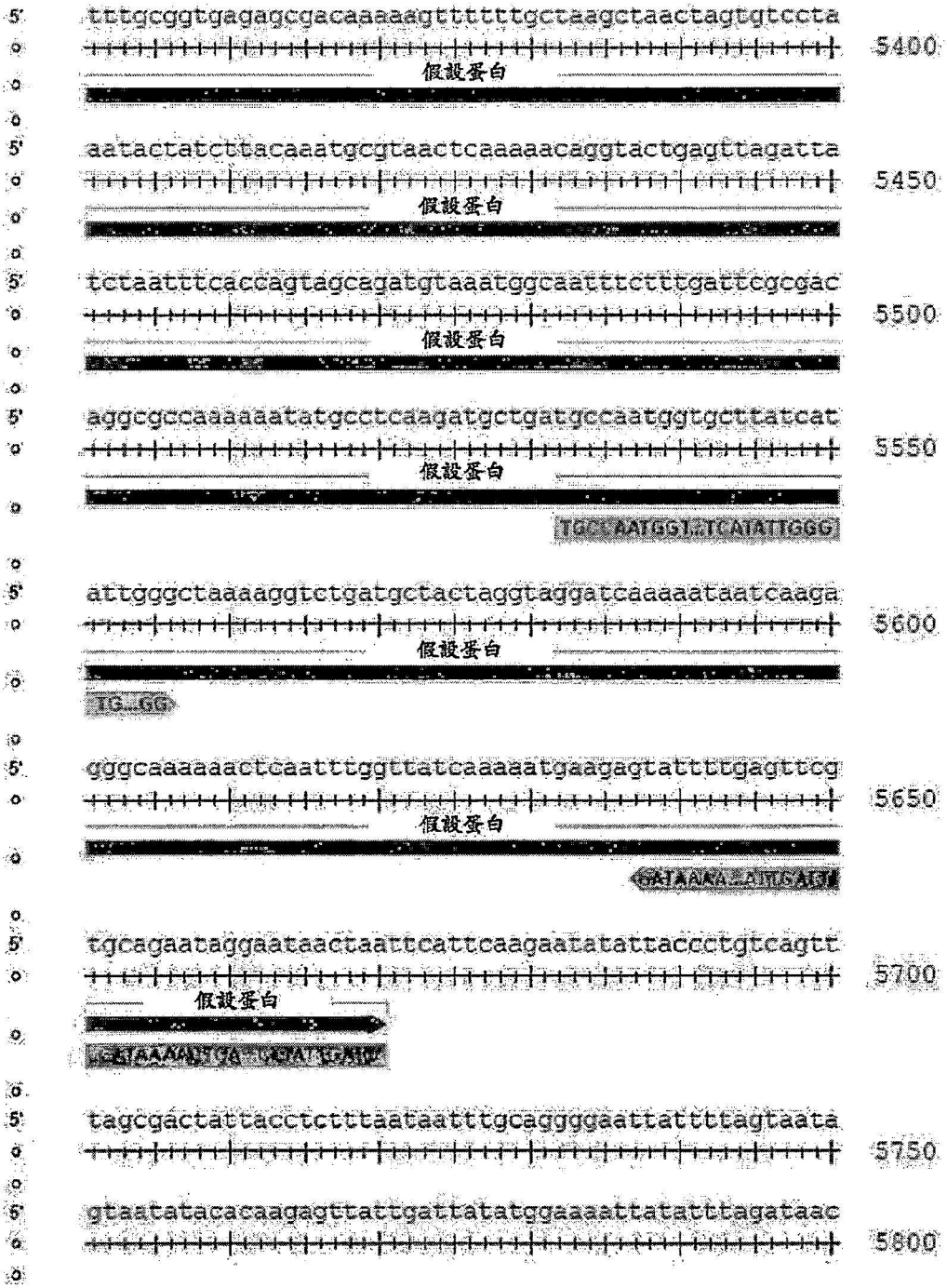


圖 40M

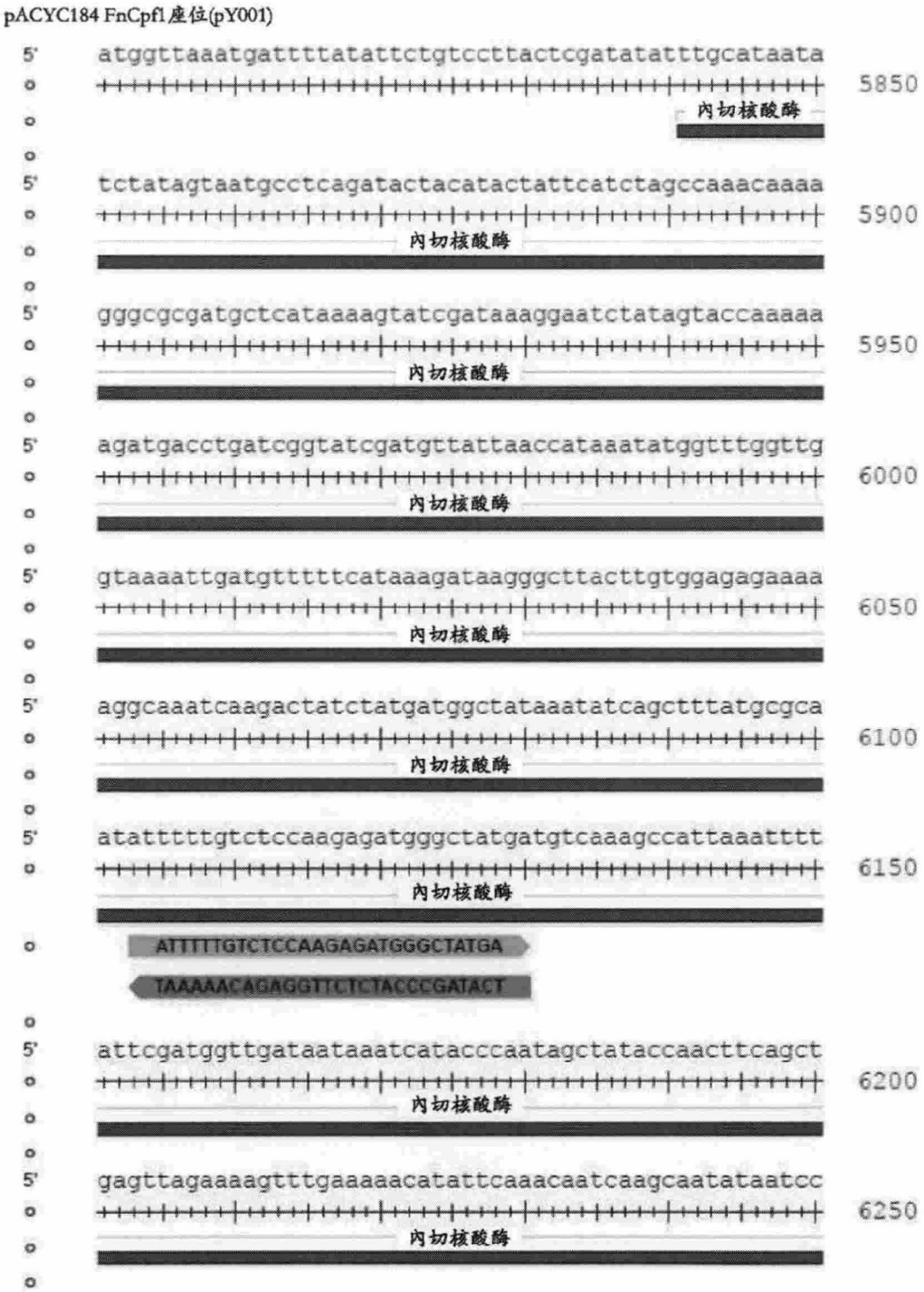

pACYC184 FnCpf1座位(pY001)
5'
atggttaaatgattttatattctgtccttactcgatatatttgcataata
5850
内切核酸酶
5'
tctatagtaatgcctcagatactacatactattcatctagccaaacaaaa
5900
内切核酸酶
5'
gggcgcgatgctcataaaagtatcgataaaggaatctatagtaccaaaaa
5950
内切核酸酶
5'
agatgacctgatcggtatcgatgttattaaccataaatatggtttggttg
6000
内切核酸酶
5'
gtaaaattgatgtttttcataaagataagggcttacttgtggagagaaaa
6050
内切核酸酶
5'
aggcaaatcaagactatctatgatggctataaatatcagctttatgcgca
6100
内切核酸酶
5'
atatttttgtctccaagagatgggctatgatgtcaaagccattaaatttt
6150
内切核酸酶
ATTTTTGTCTCCAAGAGATGGGCTATGA
TAAAAACAGAGGTTCTCTACCCGATACT
5'
attcgatggttgataataaatcatacccaatagctataccaacttcagct
6200
内切核酸酶
5'
gagttagaaaagtttgaaaaacatattcaaacaatcaagcaatataatcc
6250
内切核酸酶

圖 40N

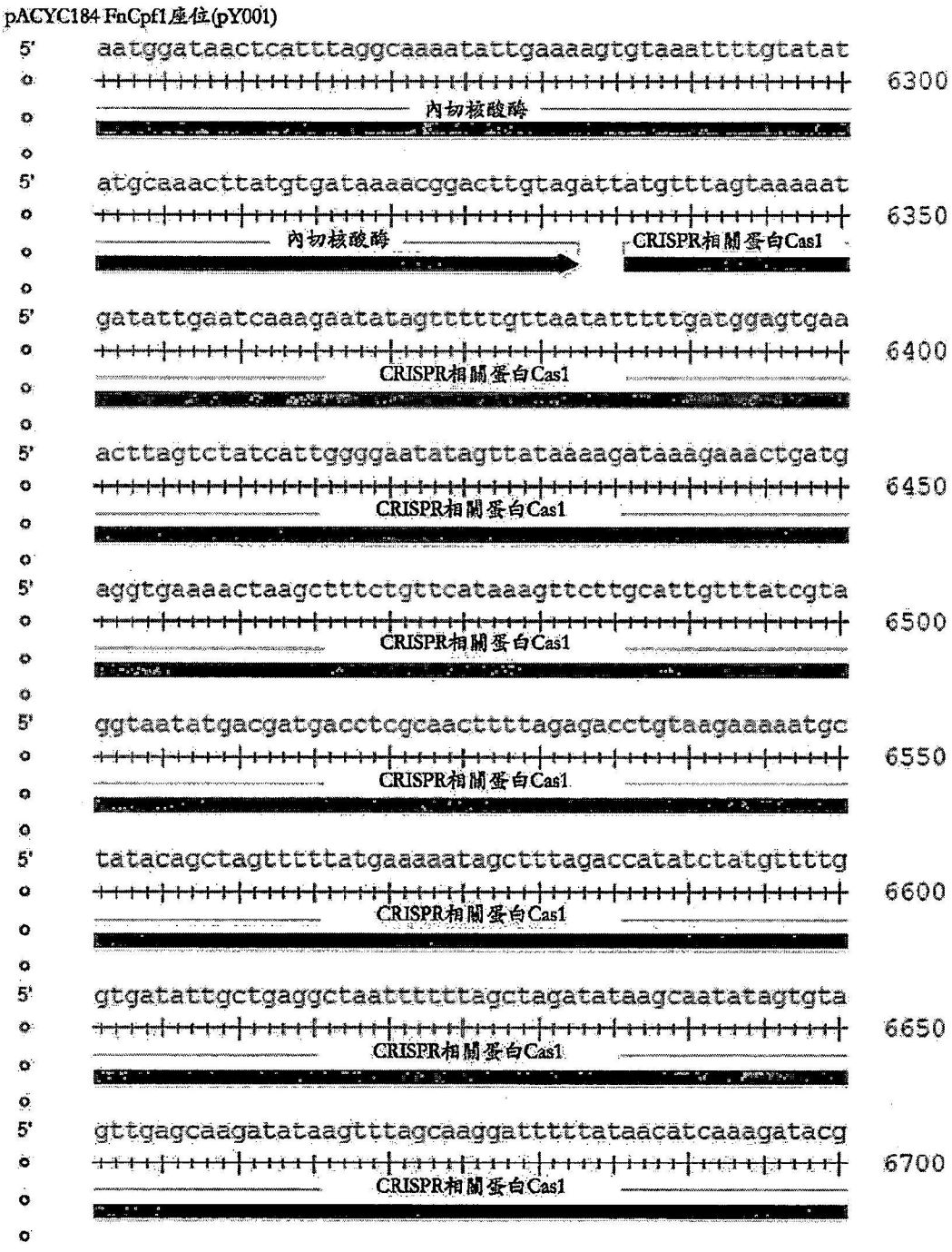

pACYC184-FnCpf1座位(pY001)
5' aatggataactcatttaggcaaaatattgaaaagtgtaaattttgtatat 6300
內切核酸酶
5' atgcaaacttatgtgataaaacggacttgtagattatgtttagtaaaaat 6350
內切核酸酶
CRISPR相關蛋白Cas1
5' gatattgaatcaaagaatatagtttttgttaatatttttgatggagtgaa 6400
CRISPR相關蛋白Cas1
5' acttagtctatcattggggaatatagttataaaagataaagaaactgatg 6450
CRISPR相關蛋白Cas1
5' aggtgaaaactaagctttctgttcataaagttcttgcattgtttatcgta 6500
CRISPR相關蛋白Cas1
5' ggtaatatgacgatgacctcgcaacttttagagacctgtaagaaaaatgc 6550
CRISPR相關蛋白Cas1
5' tatacagctagtttttatgaaaaatagctttagaccatatctatgttttg 6600
CRISPR相關蛋白Cas1
5' gtgatattgctgaggctaatttttttagctagatataagcaatatagtgta 6650
CRISPR相關蛋白Cas1
5' gttgagcaagatataagtttagcaaggatttttataacatcaaagatacg 6700
CRISPR相關蛋白Cas1

圖 40O

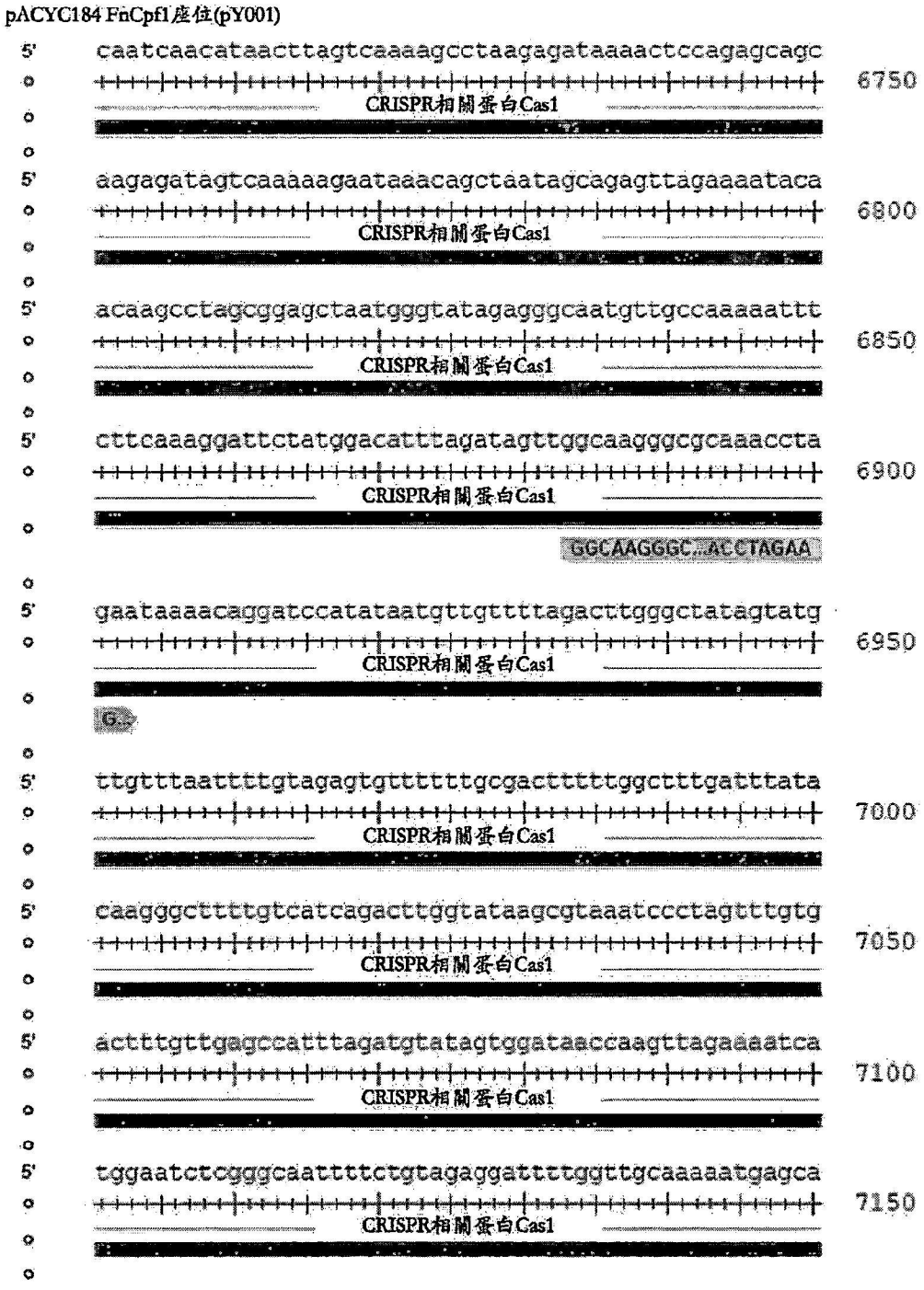

pACYC184 FnCpf1座位(pY001)
5' caatcaacataacttagtcaaaagcctaagagataaaactccagagcagc
6750
CRISPR相關蛋白Cas1
5' aagagatagtcaaaaagaataaacagctaatagcagagttagaaaataca
6800
CRISPR相關蛋白Cas1
5' acaagcctagcggagctaatgggtatagagggcaatgttgccaaaaattt
6850
CRISPR相關蛋白Cas1
5' cttcaaaggattctatggacatttagatagttggcaagggcgcaaaccta
6900
CRISPR相關蛋白Cas1
GGCAAGGGC ACCTAGAA
5' gaataaaacaggatccatataatgttgttttagacttgggctatagtatg
6950
CRISPR相關蛋白Cas1
G
5' ttgtttaattttgtagagtgtttttgcgacttttggctttgatttata
7000
CRISPR相關蛋白Cas1
5' caagggcttttgtcatcagacttggtataagcgtaaatccctagtttgtg
7050
CRISPR相關蛋白Cas1
5' actttgttgagccatttagatgtatagtggataaccaagttagaaaatca
7100
CRISPR相關蛋白Cas1
5' tggaatctcgggcaattttctgtagaggattttggttgcaaaaatgagca
7150
CRISPR相關蛋白Cas1

圖 40P

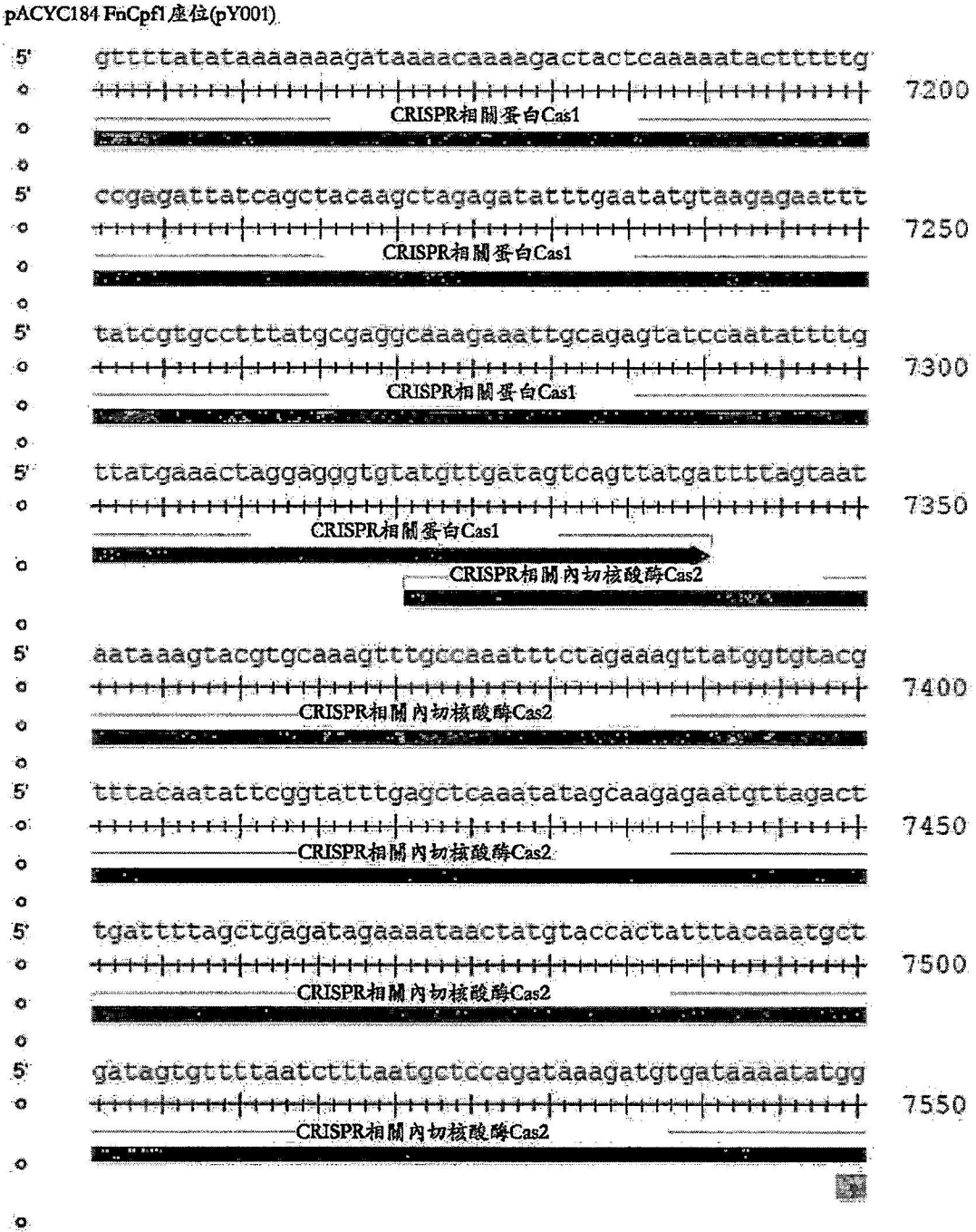

pACYC184 FnCpf1座位(pY001)
5' gttttatataaaaaaagataaaacaaaagactactcaaaaatacttttg
7200
CRISPR相關蛋白Cas1
5' ccgagattatcagctacaagctagagatatttgaatatgtaagagaattt
7250
CRISPR相關蛋白Cas1
5' tatcgtgcctttatgcgaggcaaagaaattgcagagtatccaatattttg
7300
CRISPR相關蛋白Cas1
5' ttatgaaactaggagggtgtatgttgatagtcagttatgattttagtaat
7350
CRISPR相關蛋白Cas1
CRISPR相關內切核酸酶Cas2
5' aataaagtacgtgcaaagtttgccaaatttctagaaagttatggtgtacg
7400
CRISPR相關內切核酸酶Cas2
5' tttacaatattcggtatttgagctcaaatatagcaagagaatgttagact
7450
CRISPR相關內切核酸酶Cas2
5' tgattttagctgagatagaaaataactatgtaccactatttacaaatgct
7500
CRISPR相關內切核酸酶Cas2
5' gatagtgttttaatctttaatgctccagataaagatgtgataaaatatgg
7550
CRISPR相關內切核酸酶Cas2

圖 40Q

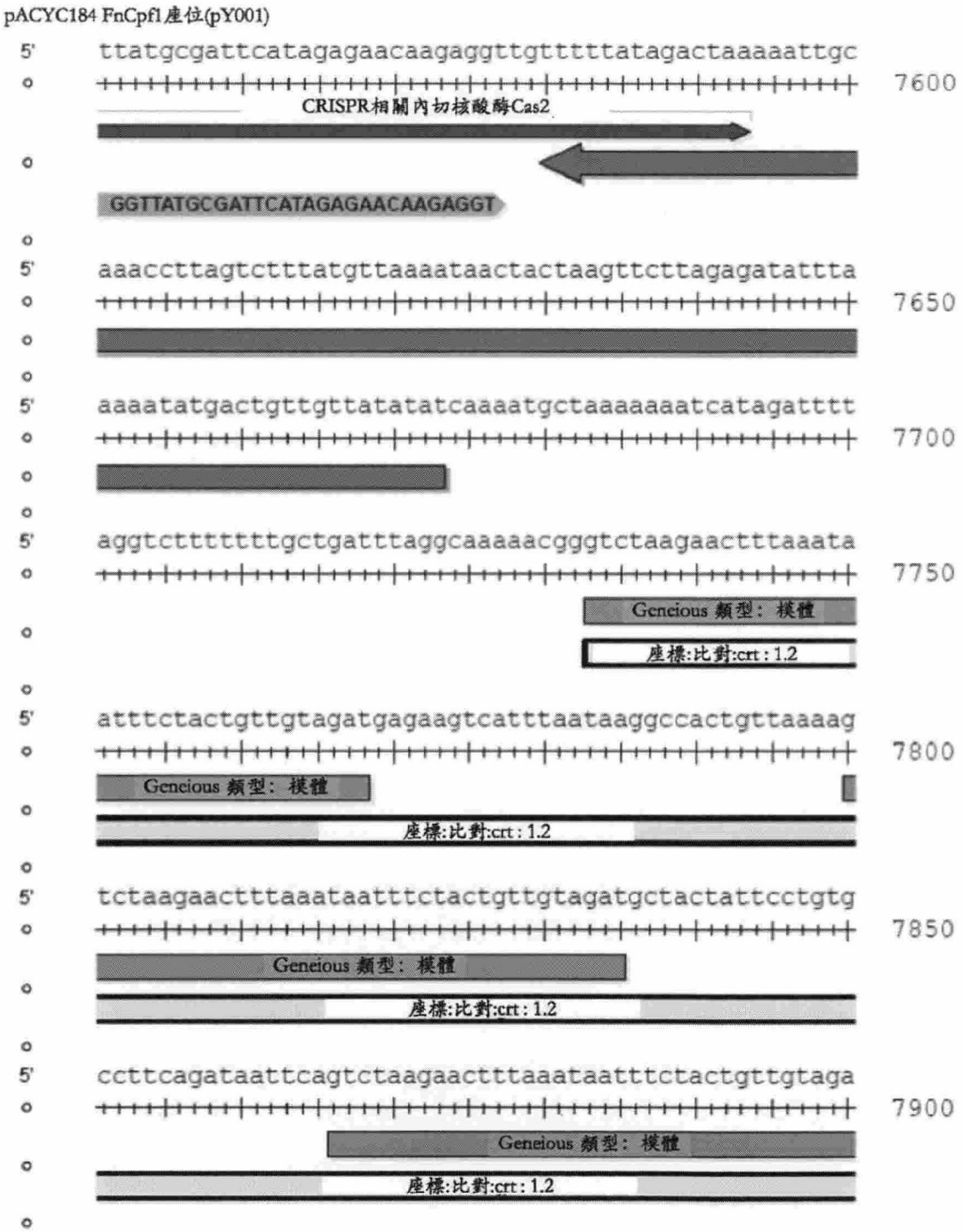

pACYC184 FnCpf1座位(pY001)
5'
ttatgcgattcatagagaacaagaggttgtttttatagactaaaaattgc
7600
CRISPR相關內切核酸酶Cas2
GGTTATGCGATTCATAGAGAACAAGAGGT
5'
aaaccttagtctttatgttaaaataactactaagttcttagagatattta
7650
5'
aaaatatgactgttgttatatatcaaaatgctaaaaaaatcatagatttt
7700
5'
aggtctttttttgctgatttaggcaaaaacgggtctaagaactttaaata
7750
Geneious 類型: 模體
座標:比對:crt : 1.2
5'
atttctactgttgtagatgagaagtcatttaataaggccactgttaaaag
7800
Geneious 類型: 模體
座標:比對:crt : 1.2
5'
tctaagaactttaaataatttctactgttgtagatgctactattcctgtg
7850
Geneious 類型: 模體
座標:比對:crt : 1.2
5'
ccttcagataattcagtctaagaactttaaataatttctactgttgtaga
7900
Geneious 類型: 模體
座標:比對:crt : 1.2

圖 40R

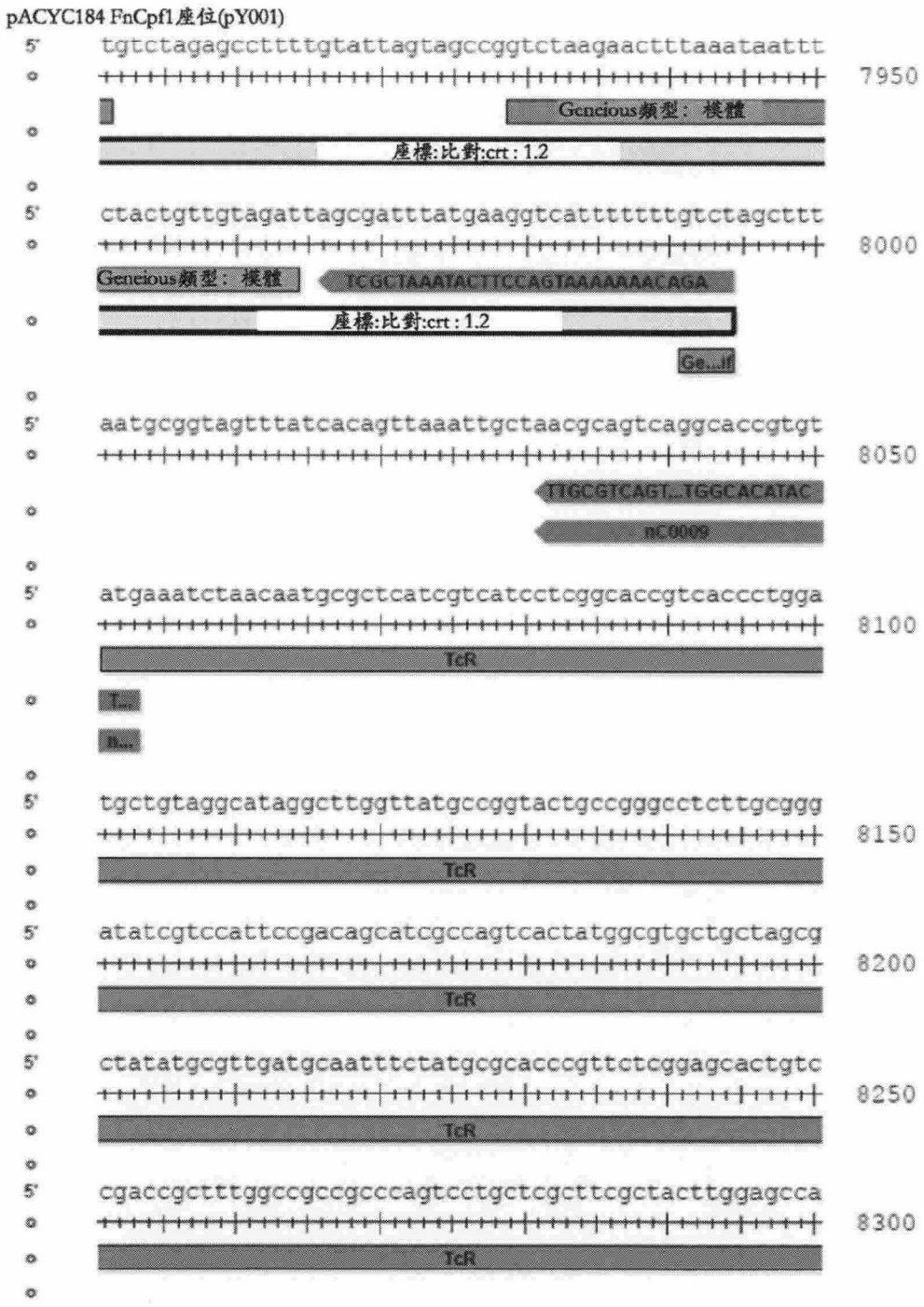

pACYC184 FnCpf1座位(pY001)
5' tgtctagagccttttgtattagtagccggtctaagaactttaaataattt
7950
Geneious頻型： 模體
座標:比對:crt : 1.2
5' ctactgttgtagattagcgatttatgaaggtcatttttttgtctagcttt
8000
Geneious頻型： 模體
TCGCTAAATACTTCCAGTAAAAAAACAGA
座標:比對:crt : 1.2
Ge...lf
5' aatgcggtagtttatcacagttaaattgctaacgcagtcaggcaccgtgt
8050
TTGCGTCAGT...TGGCACATAC
nC0009
5' atgaaatctaacaatgcgctcatcgtcatcctcggcaccgtcaccctgga
8100
TcR
T...
n...
5' tgctgtaggcataggcttggttatgccggtactgccgggcctcttgcggg
8150
TcR
5' atatcgtccattccgacagcatcgccagtcactatggcgtgctgctagcg
8200
TcR
5' ctatatgcgttgatgcaatttctatgcgcacccgttctcggagcactgtc
8250
TcR
5' cgaccgctttggccgccgcccagtcctgctcgcttcgctacttggagcca
8300
TcR

圖 40S

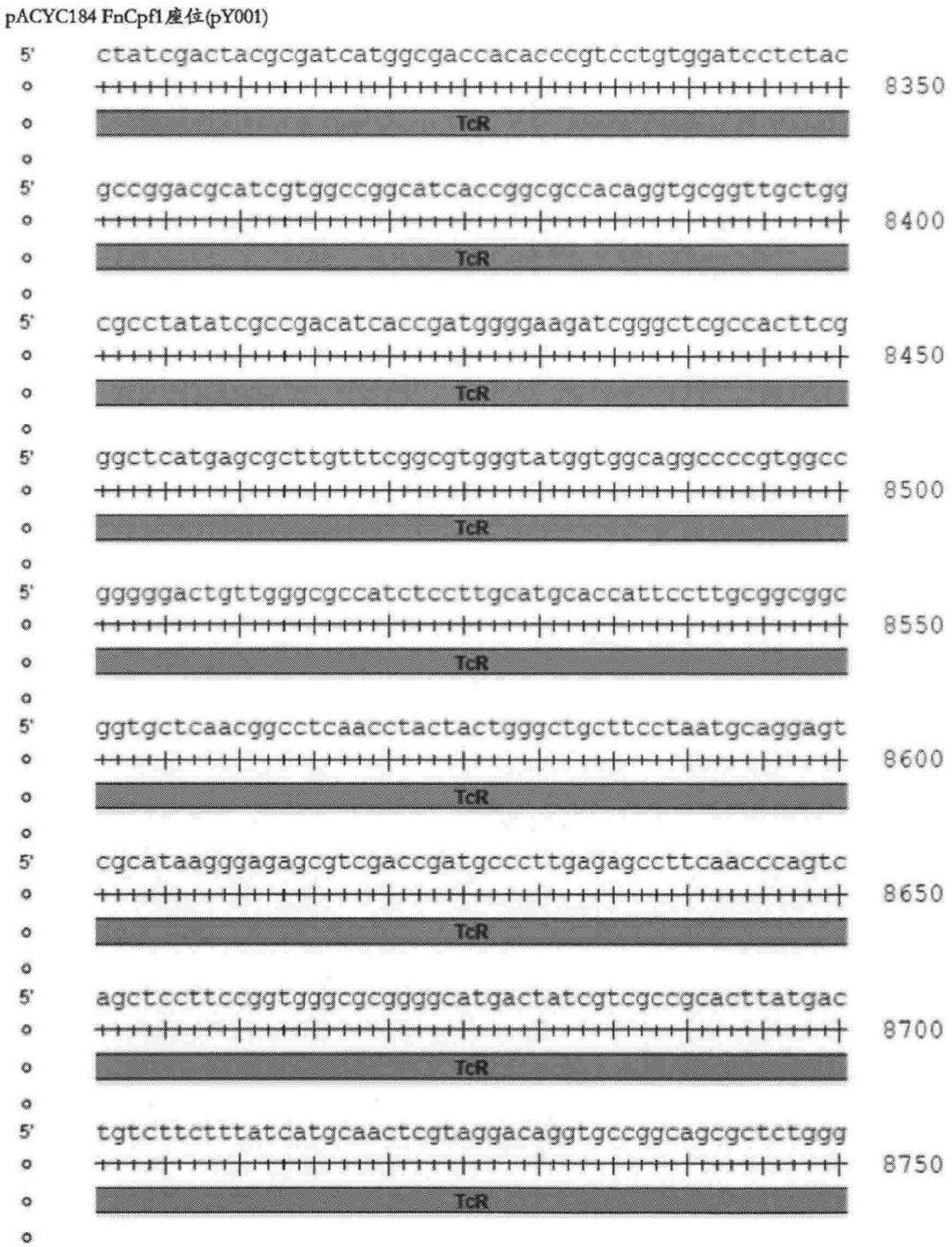

pACYC184 FnCpf1座位(pY001)
5' ctatcgactacgcgatcatggcgaccacacccgtcctgtggatcctctac
8350
TcR
5' gccggacgcatcgtggccggcatcaccggcgccacaggtgcggttgctgg
8400
TcR
5' cgcctatatcgccgacatcaccgatggggaagatcgggctcgccacttcg
8450
TcR
5' ggctcatgagcgcttgtttcggcgtgggtatggtggcaggccccgtggcc
8500
TcR
5' gggggactgttgggcgccatctccttgcatgcaccattccttgcggcggc
8550
TcR
5' ggtgctcaacggcctcaacctactactgggctgcttcctaatgcaggagt
8600
TcR
5' cgcataagggagagcgtcgaccgatgcccttgagagccttcaacccagtc
8650
TcR
5' agctccttccggtgggcgcggggcatgactatcgtcgccgcacttatgac
8700
TcR
5' tgtcttctttatcatgcaactcgtaggacaggtgccggcagcgctctggg
8750
TcR

圖 40T

pACYC184 FnCpf1座位(pY001)

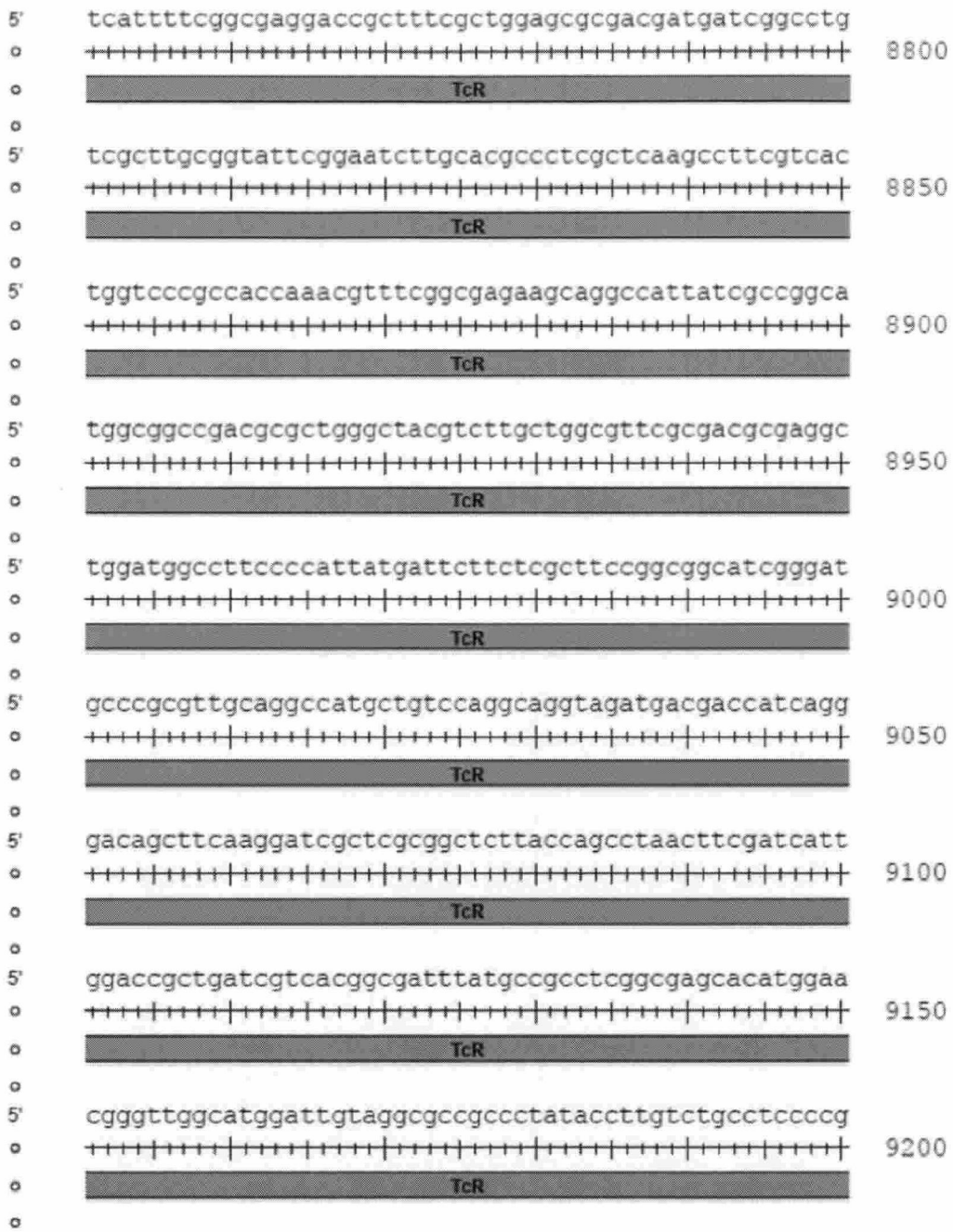


圖 40U

pACYC184 FnCpf1座位(pY001)

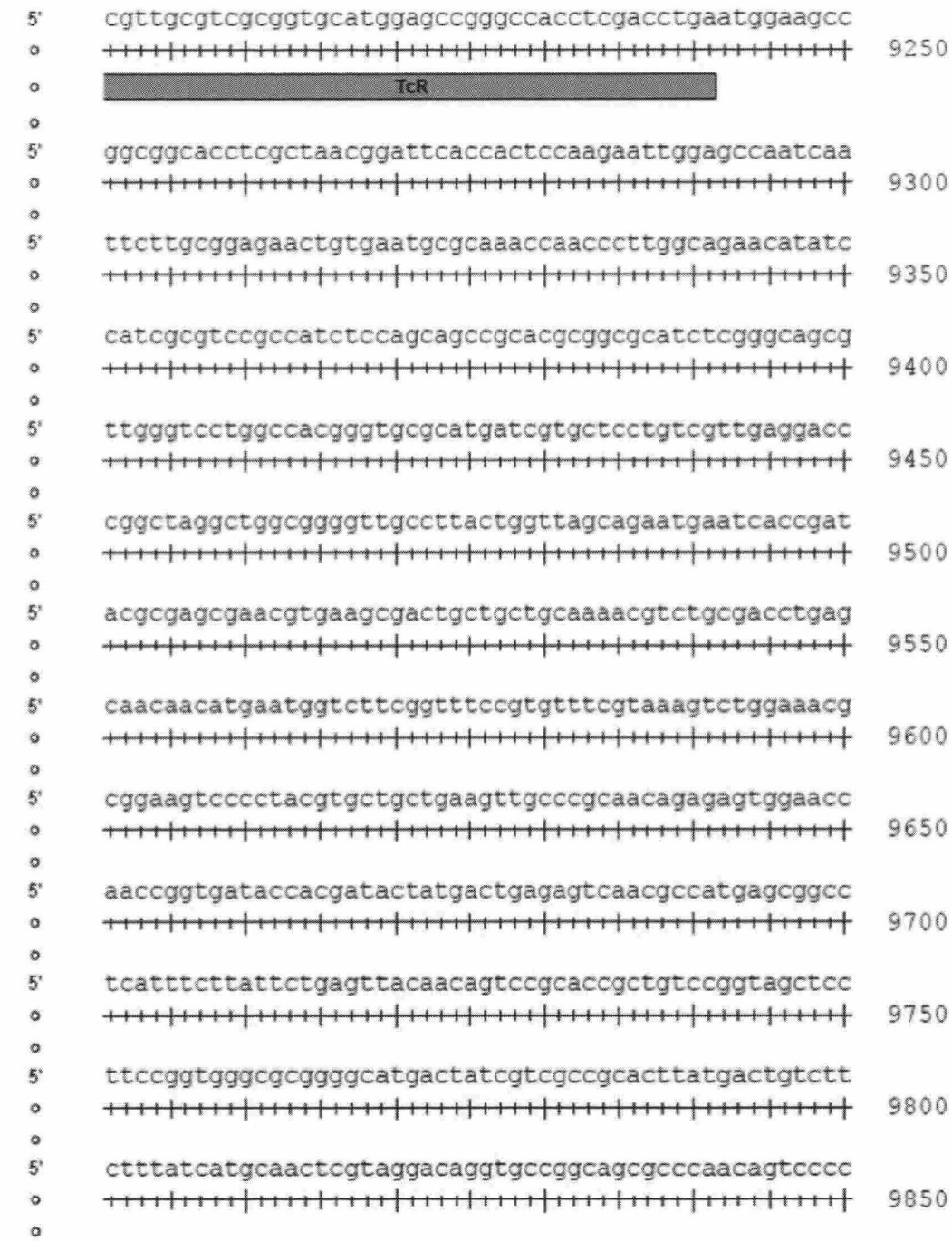


圖 40V

pACYC184 FnCpf1座位(pY001)

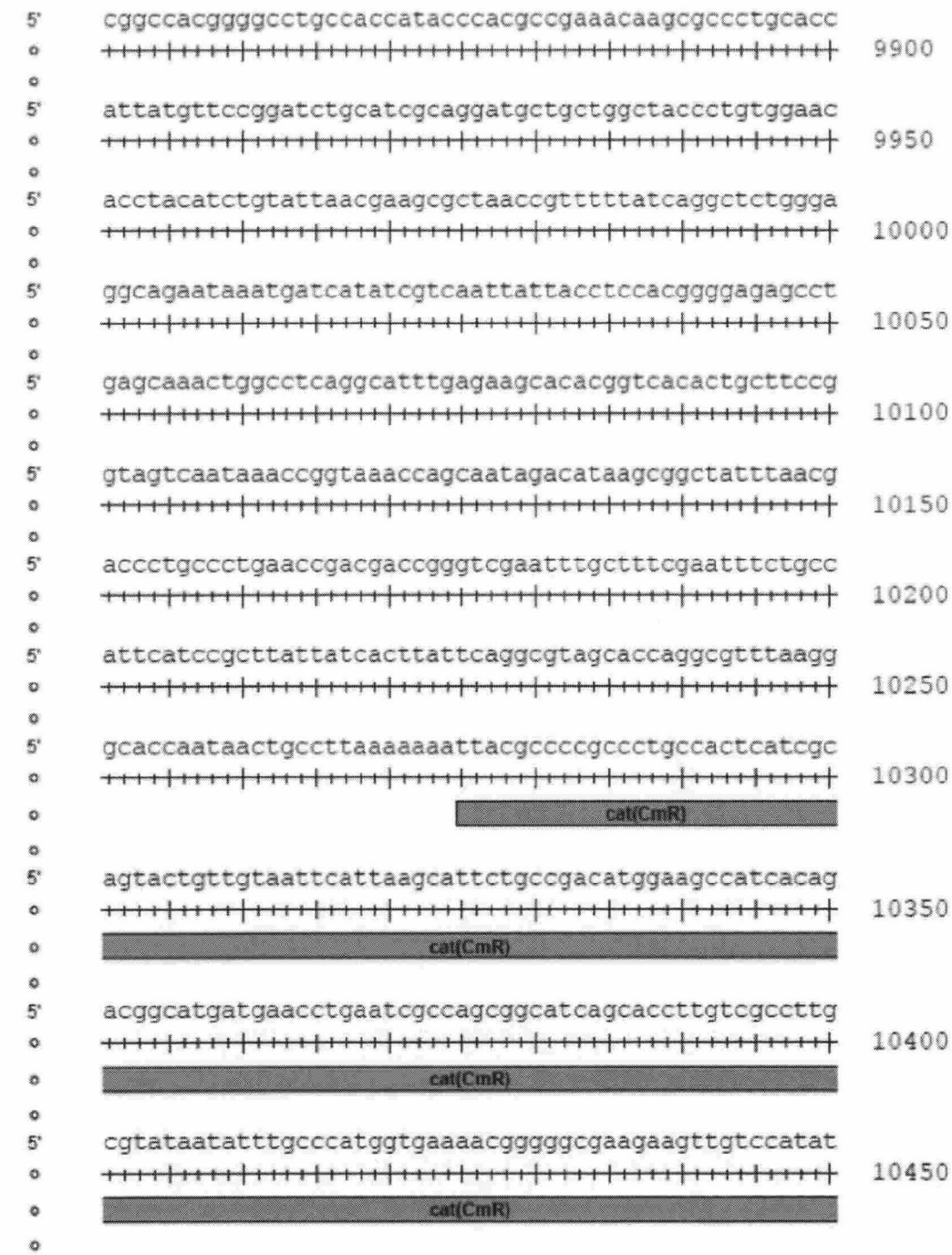


圖 40W

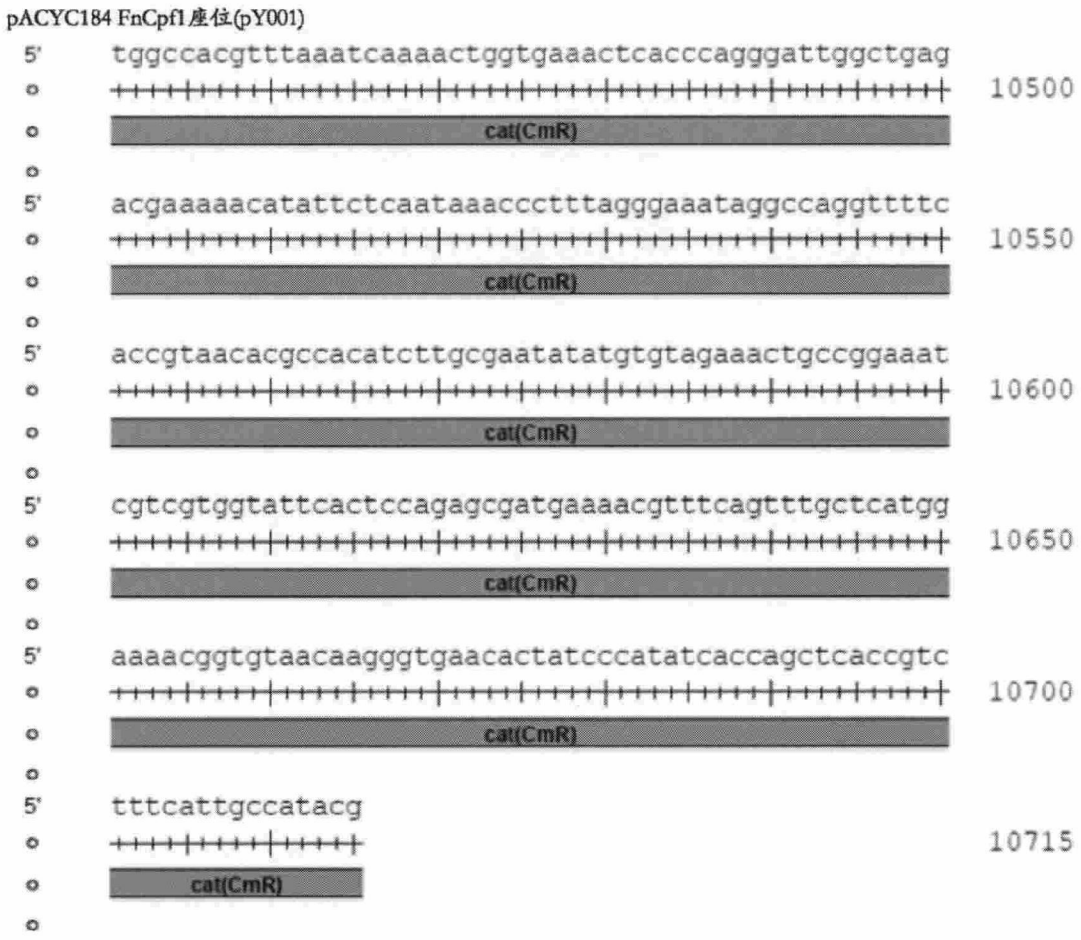

pACYC184 FnCpf1座位(pY001)
5'
tggccacgtttaaatcaaaactggtgaaactcacccagggattggctgag
10500
cat(CmR)
5'
acgaaaaacatattctcaataaacccttttagggaaataggccaggttttc
10550
cat(CmR)
5'
accgtaacacgccacatcttgcgaatatatgtgtagaaactgccggaaat
10600
cat(CmR)
5'
cgtcgtggtattcactccagagcgatgaaaacgtttcagtttgctcatgg
10650
cat(CmR)
5'
aaaacggtgtaacaagggtgaacactatcccatatcaccagctcaccgtc
10700
cat(CmR)
5'
tttcattgccatacg
10715
cat(CmR)

圖 40X

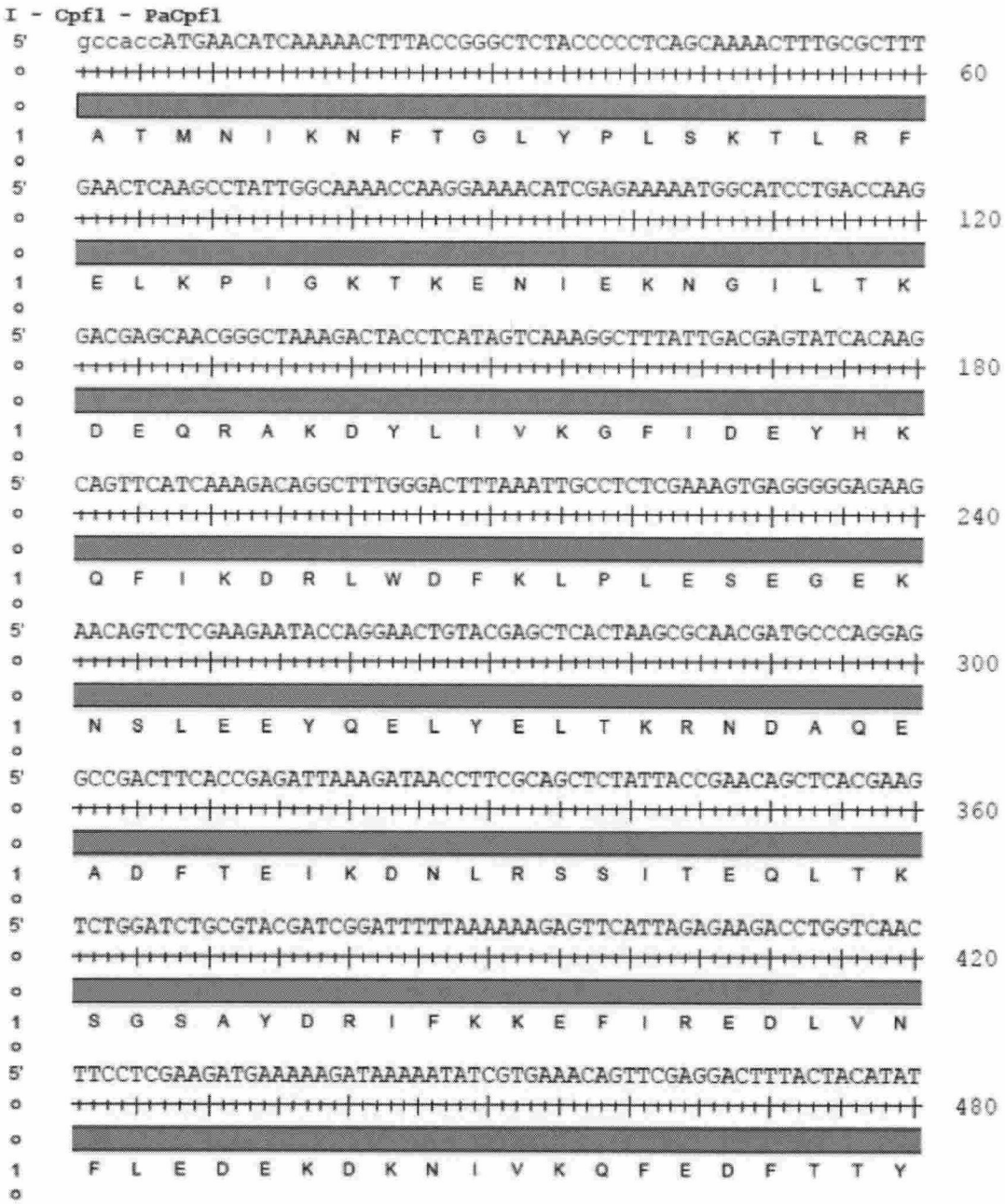

I - Cpf1 - PaCpf1
5' gccaccATGAACATCAAAAACTTTACCGGGCTCTACCCCCTCAGCAAAACTTTGCGCTTT 60
1 A T M N I K N F T G L Y P L S K T L R F
5' GAACTCAAGCCTATTGGCAAAACCAAGGAAAACATCGAGAAAAATGGCATCCTGACCAAG 120
1 E L K P I G K T K E N I E K N G I L T K
5' GACGAGCAACGGGCTAAAGACTACCTCATAGTCAAAGGCTTTATTGACGAGTATCACAAG 180
1 D E Q R A K D Y L I V K G F I D E Y H K
5' CAGTTCATCAAAGACAGGCTTTGGGACTTTAAATTGCCTCTCGAAAGTGAGGGGGAGAAG 240
1 Q F I K D R L W D F K L P L E S E G E K
5' AACAGTCTCGAAGAATACCAGGAACTGTACGAGCTCACTAAGCGCAACGATGCCCAGGAG 300
1 N S L E E Y Q E L Y E L T K R N D A Q E
5' GCCGACTTCACCGAGATTAAAGATAACCTTCGCAGCTCTATTACCGAACAGCTCACGAAG 360
1 A D F T E I K D N L R S S I T E Q L T K
5' TCTGGATCTGCGTACGATCGGATTTTTAAAAAAGAGTTCATTAGAGAAGACCTGGTCAAC 420
1 S G S A Y D R I F K K E F I R E D L V N
5' TTCCTCGAAGATGAAAAAGATAAAAATATCGTGAAACAGTTCGAGGACTTTACTACATAT 480
1 F L E D E K D K N I V K Q F E D F T T Y

圖 41A

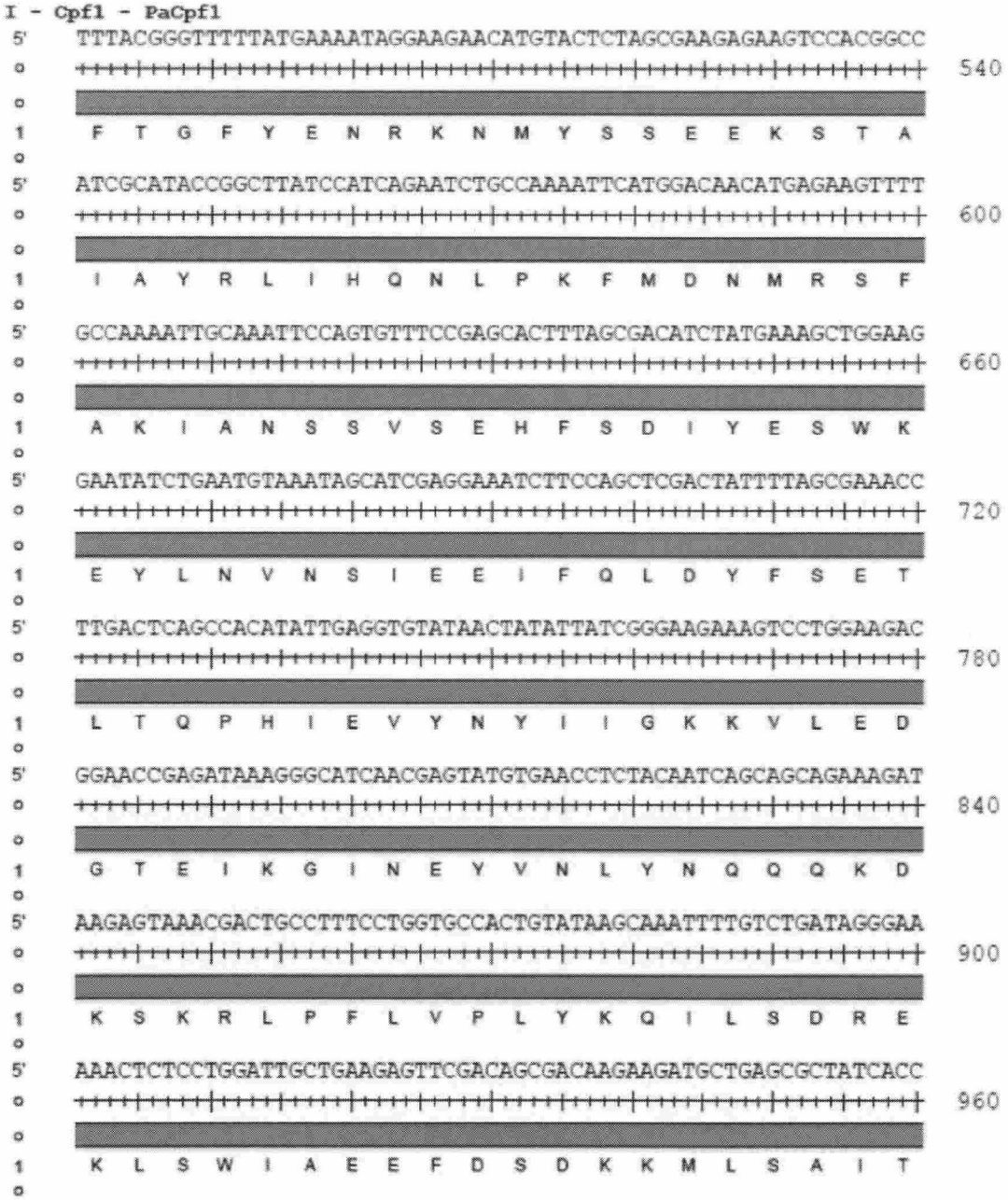

I - Cpf1 - PaCpf1
5' TTTACGGGTTTTTATGAAAATAGGAAGAACATGTACTCTAGCGAAGAGAAGTCCACGGCC
540
F T G F Y E N R K N M Y S S E E K S T A
5' ATCGCATACCGGCTTATCCATCAGAATCTGCCAAAATTCATGGACAACATGAGAAGTTTT
600
I A Y R L I H Q N L P K F M D N M R S F
5' GCCAAAATTGCAAATTCCAGTGTTTCCGAGCACTTTAGCGACATCTATGAAAGCTGGAAG
660
A K I A N S S V S E H F S D I Y E S W K
5' GAATATCTGAATGTAAATAGCATCGAGGAAATCTTCCAGCTCGACTATTTTAGCGAAACC
720
E Y L N V N S I E E I F Q L D Y F S E T
5' TTGACTCAGCCACATATTGAGGTGTATAACTATATTATCGGGAAGAAAGTCCTGGAAGAC
780
L T Q P H I E V Y N Y I I G K K V L E D
5' GGAACCGAGATAAAGGGCATCAACGAGTATGTGAACCTCTACAATCAGCAGCAGAAAGAT
840
G T E I K G I N E Y V N L Y N Q Q Q K D
5' AAGAGTAAACGACTGCCTTTCCTGGTGCCACTGTATAAGCAAATTTTGTCTGATAGGGAA
900
K S K R L P F L V P L Y K Q I L S D R E
5' AAACTCTCCTGGATTGCTGAAGAGTTCGACAGCGACAAGAAGATGCTGAGCGCTATCACC
960
K L S W I A E E F D S D K K M L S A I T

圖 41B

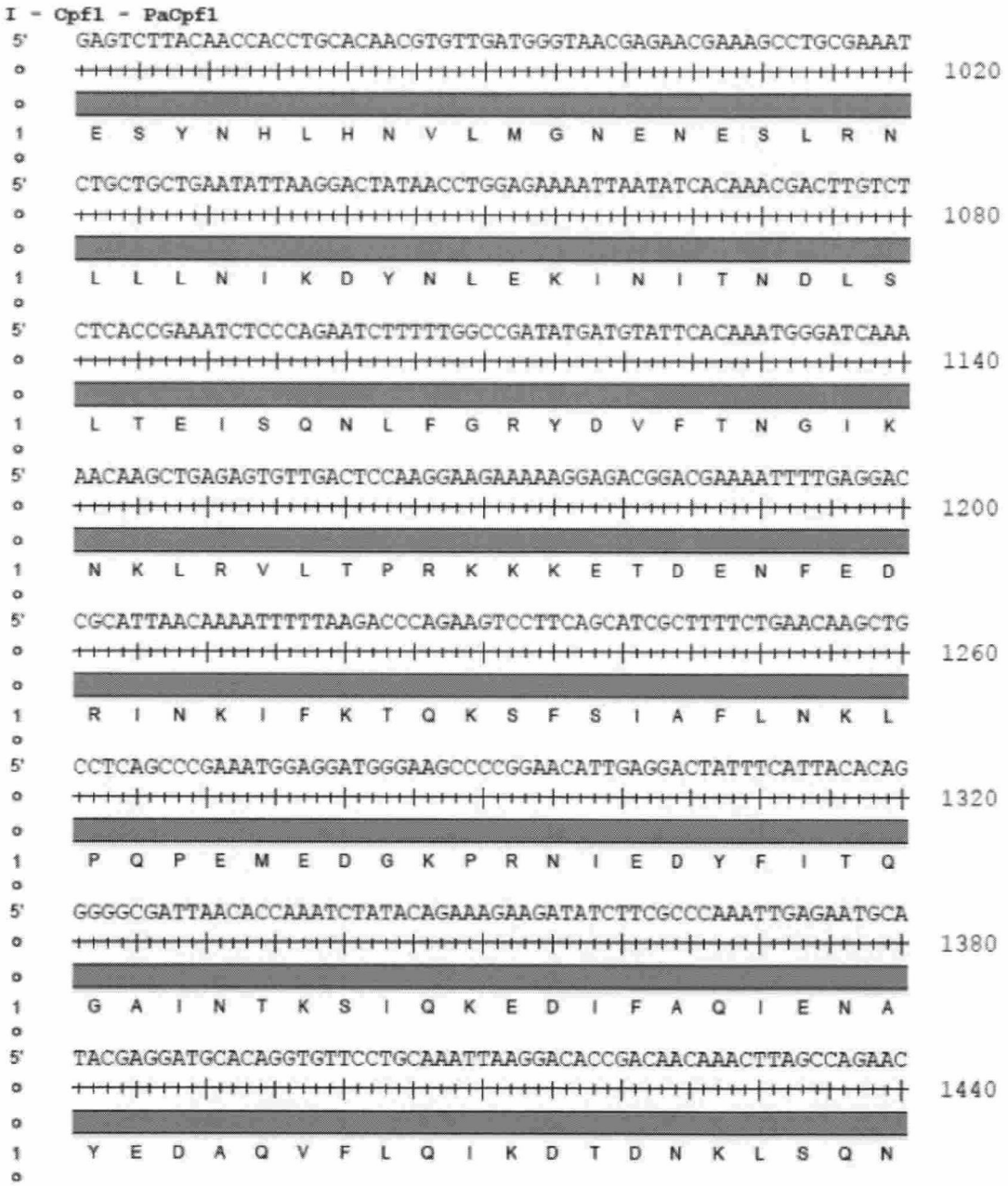

I - Cpf1 - PaCpf1
5' GAGTCTTACAACCACCTGCACAACGTGTTGATGGGTAACGAGAACGAAAGCCTGCGAAAT
1020
E S Y N H L H N V L M G N E N E S L R N
5' CTGCTGCTGAATATTAAGGACTATAACCTGGAGAAAATTAATATCACAAACGACTTGTCT
1080
L L L N I K D Y N L E K I N I T N D L S
5' CTCACCGAAATCTCCCAGAATCTTTTTGGCCGATATGATGTATTCACAAATGGGATCAAA
1140
L T E I S Q N L F G R Y D V F T N G I K
5' AACAAGCTGAGAGTGTTGACTCCAAGGAAGAAAAAGGAGACGGACGAAAATTTTGAGGAC
1200
N K L R V L T P R K K K E T D E N F E D
5' CGCATTAACAAAATTTTTAAGACCCAGAAGTCCTTCAGCATCGCTTTTCTGAACAAGCTG
1260
R I N K I F K T Q K S F S I A F L N K L
5' CCTCAGCCCGAAATGGAGGATGGGAAGCCCCGGAACATTGAGGACTATTTCATTACACAG
1320
P Q P E M E D G K P R N I E D Y F I T Q
5' GGGGCGATTAACACCAAATCTATACAGAAAGAAGATATCTTCGCCCAAATTGAGAATGCA
1380
G A I N T K S I Q K E D I F A Q I E N A
5' TACGAGGATGCACAGGTGTTCCTGCAAATTAAGGACACCGACAACAAACTTAGCCAGAAC
1440
Y E D A Q V F L Q I K D T D N K L S Q N

圖 41C

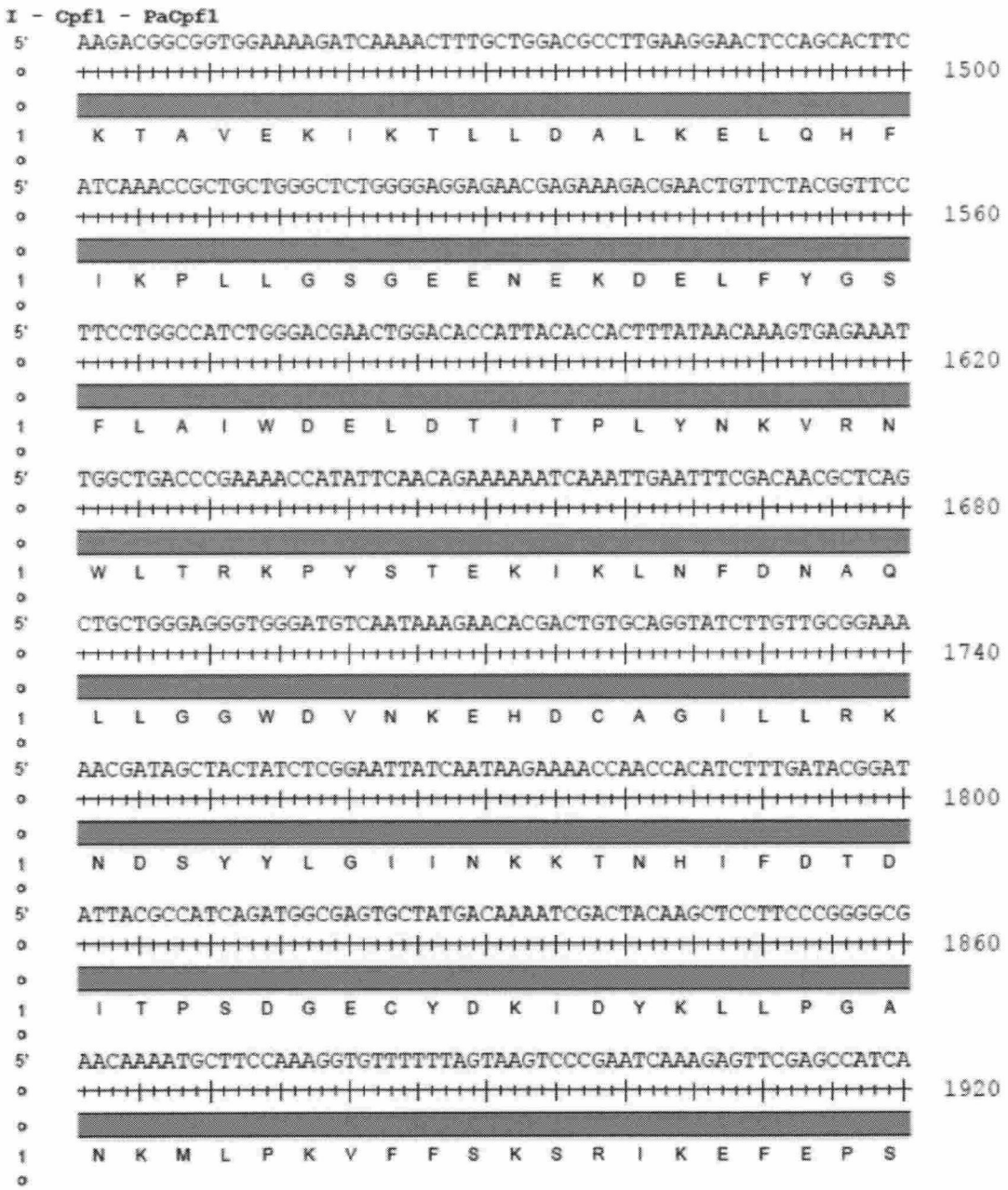

I - Cpf1 - PaCpf1
5' AAGACGGCGGTGGAAAAGATCAAAACTTTGCTGGACGCCTTGAAGGAACTCCAGCACTTC
1500
K T A V E K I K T L L D A L K E L Q H F
5' ATCAAACCGCTGCTGGGCTCTGGGGAGGAGAACGAGAAAGACGAACTGTTCTACGGTTCC
1560
I K P L L G S G E E N E K D E L F Y G S
5' TTCCTGGCCATCTGGGACGAACTGGACACCATTACACCACTTTATAACAAAGTGAGAAAT
1620
F L A I W D E L D T I T P L Y N K V R N
5' TGGCTGACCCGAAAACCATATTCAACAGAAAAAATCAAATTGAATTTCGACAACGCTCAG
1680
W L T R K P Y S T E K I K L N F D N A Q
5' CTGCTGGGAGGGTGGGATGTCAATAAAGAACACGACTGTGCAGGTATCTTGTTGCGGAAA
1740
L L G G W D V N K E H D C A G I L L R K
5' AACGATAGCTACTATCTCGGAATTATCAATAAGAAAACCAACCACATCTTTGATACGGAT
1800
N D S Y Y L G I I N K K T N H I F D T D
5' ATTACGCCATCAGATGGCGAGTGCTATGACAAAATCGACTACAAGCTCCTTCCCGGGGCG
1860
I T P S D G E C Y D K I D Y K L L P G A
5' AACAAAATGCTTCCAAAGGTGTTTTTTAGTAAGTCCCGAATCAAAGAGTTCGAGCCATCA
1920
N K M L P K V F F S K S R I K E F E P S

圖 41D

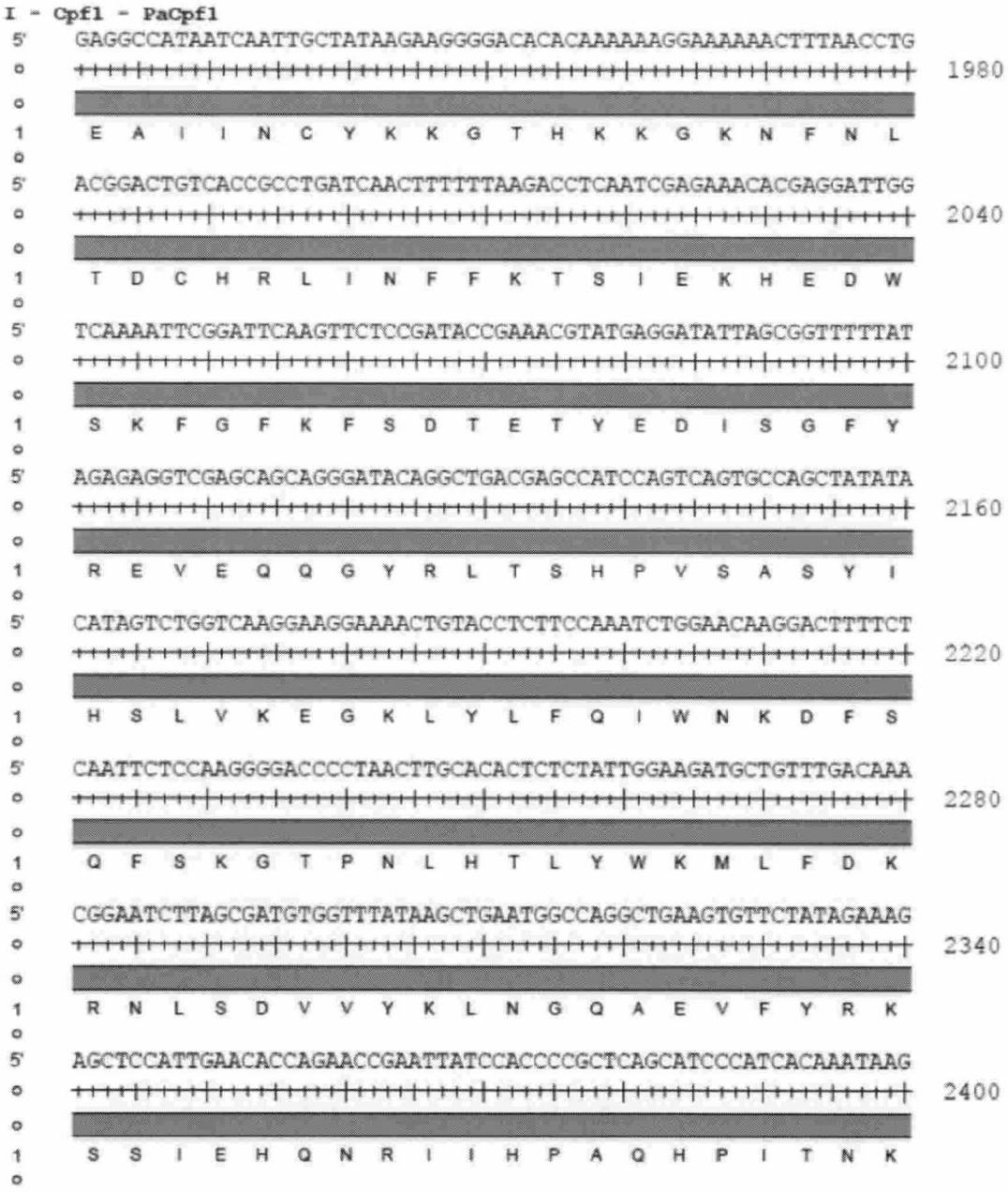

I - Cpf1 - PaCpf1
5' GAGGCCATAATCAATTGCTATAAGAAGGGGACACACAAAAAAGGAAAAAACTTTAACCTG
1980
E A I I N C Y K K G T H K K G K N F N L
5' ACGGACTGTCACCGCCTGATCAACTTTTTTAAGACCTCAATCGAGAAACACGAGGATTGG
2040
T D C H R L I N F F K T S I E K H E D W
5' TCAAAATTCGGATTCAAGTTCTCCGATACCGAAACGTATGAGGATATTAGCGGTTTTTAT
2100
S K F G F K F S D T E T Y E D I S G F Y
5' AGAGAGGTCGAGCAGCAGGGATACAGGCTGACGAGCCATCCAGTCAGTGCCAGCTATATA
2160
R E V E Q Q G Y R L T S H P V S A S Y I
5' CATAGTCTGGTCAAGGAAGGAAAACTGTACCTCTTCCAAATCTGGAACAAGGACTTTTCT
2220
H S L V K E G K L Y L F Q I W N K D F S
5' CAATTCTCCAAGGGGACCCCTAACTTGCACACTCTCTATTGGAAGATGCTGTTTGACAAA
2280
Q F S K G T P N L H T L Y W K M L F D K
5' CGGAATCTTAGCGATGTGGTTTATAAGCTGAATGGCCAGGCTGAAGTGTTCTATAGAAAG
2340
R N L S D V V Y K L N G Q A E V F Y R K
5' AGCTCCATTGAACACCAGAACCGAATTATCCACCCCGCTCAGCATCCCATCACAAATAAG
2400
S S I E H Q N R I I H P A Q H P I T N K

圖 41E

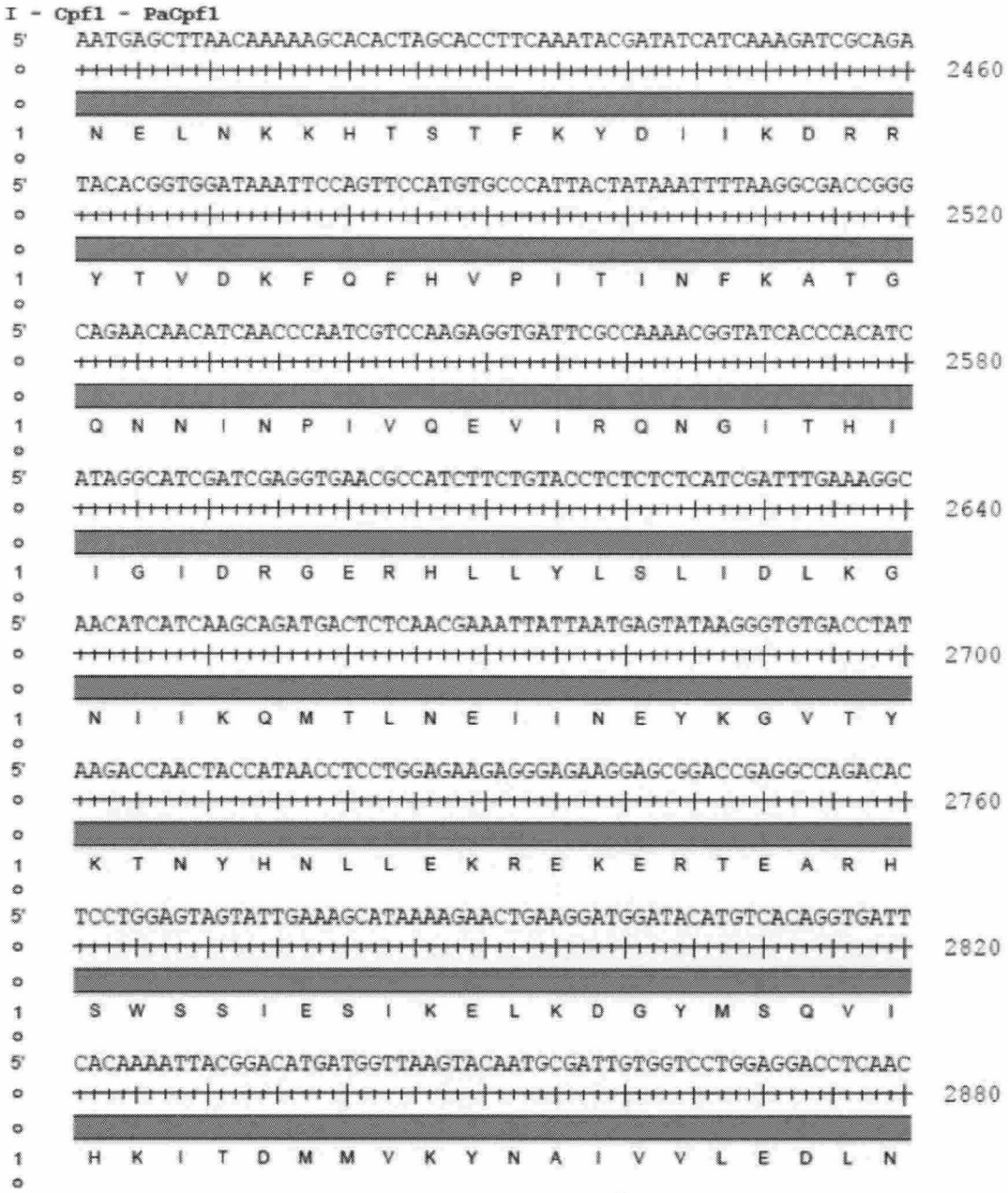

I - Cpf1 - PaCpf1
5' AATGAGCTTAACAAAAAGCACACTAGCACCTTCAAATACGATATCATCAAAGATCGCAGA
2460
1 N E L N K K H T S T F K Y D I I K D R R
5' TACACGGTGGATAAATTCCAGTTCCATGTGCCCATTACTATAAATTTTAAGGCGACCGGG
2520
1 Y T V D K F Q F H V P I T I N F K A T G
5' CAGAACAACATCAACCCAATCGTCCAAGAGGTGATTCGCCAAAACGGTATCACCCACATC
2580
1 Q N N I N P I V Q E V I R Q N G I T H I
5' ATAGGCATCGATCGAGGTGAACGCCATCTTCTGTACCTCTCTCTCATCGATTTGAAAGGC
2640
1 I G I D R G E R H L L Y L S L I D L K G
5' AACATCATCAAGCAGATGACTCTCAACGAAATTATTAATGAGTATAAGGGTGTGACCTAT
2700
1 N I I K Q M T L N E I I N E Y K G V T Y
5' AAGACCAACTACCATAACCTCCTGGAGAAGAGGGAGAAGGAGCGGACCGAGGCCAGACAC
2760
1 K T N Y H N L L E K R E K E R T E A R H
5' TCCTGGAGTAGTATTGAAAGCATAAAAGAACTGAAGGATGGATACATGTCACAGGTGATT
2820
1 S W S S I E S I K E L K D G Y M S Q V I
5' CACAAAATTACGGACATGATGGTTAAGTACAATGCGATTGTGGTCCTGGAGGACCTCAAC
2880
1 H K I T D M M V K Y N A I V V L E D L N

圖 41F

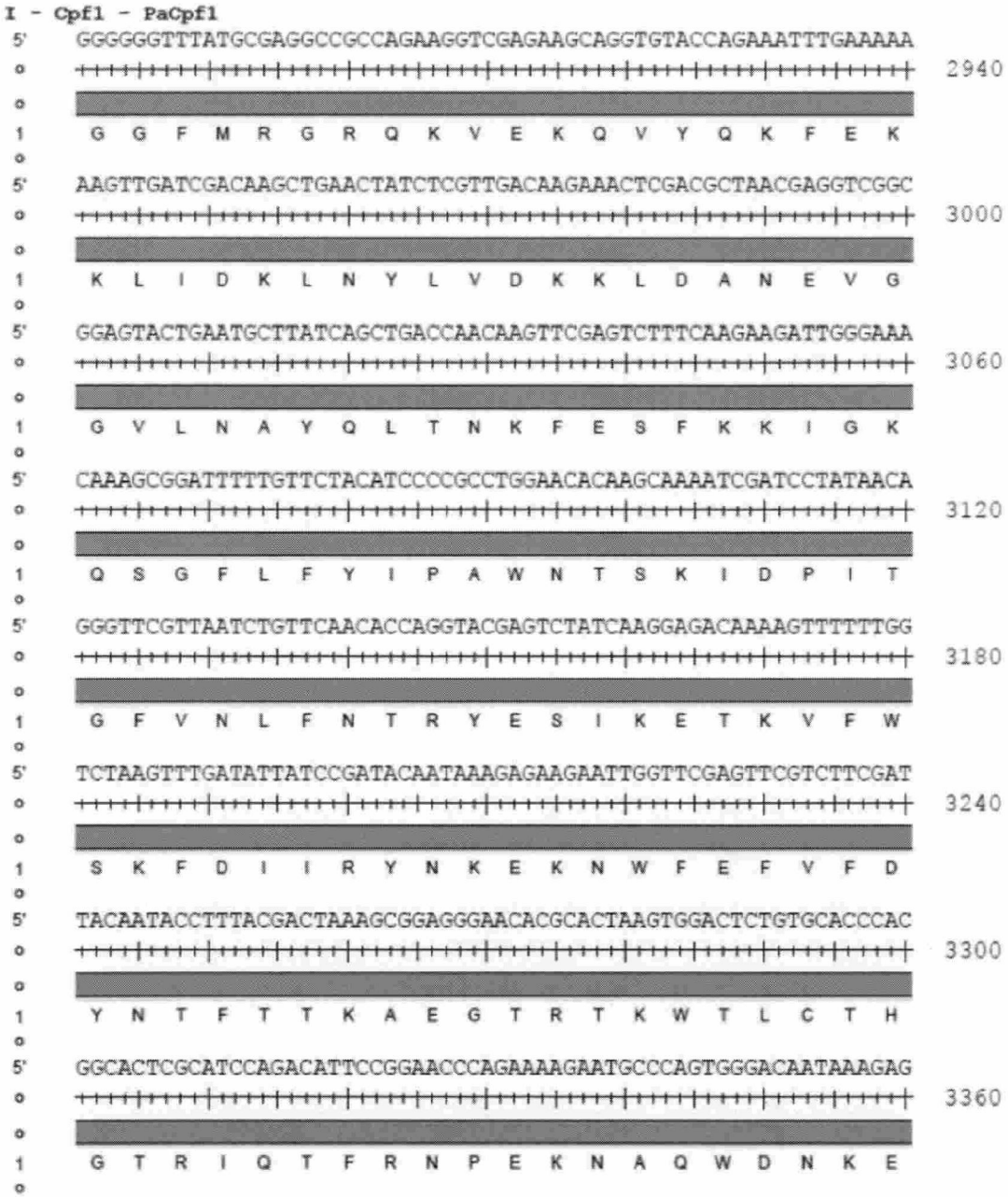

I - Cpf1 - PaCpf1
5' GGGGGGTTTATGCGAGGCCGCCAGAAGGTCGAGAAGCAGGTGTACCAGAAATTTGAAAAA
2940
1 G G F M R G R Q K V E K Q V Y Q K F E K
5' AAGTTGATCGACAAGCTGAACTATCTCGTTGACAAGAAACTCGACGCTAACGAGGTCGGC
3000
1 K L I D K L N Y L V D K K L D A N E V G
5' GGAGTACTGAATGCTTATCAGCTGACCAACAAGTTCGAGTCTTTCAAGAAGATTGGGAAA
3060
1 G V L N A Y Q L T N K F E S F K K I G K
5' CAAAGCGGATTTTTGTTCTACATCCCCGCCTGGAACACAAGCAAAATCGATCCTATAACA
3120
1 Q S G F L F Y I P A W N T S K I D P I T
5' GGGTTCGTTAATCTGTTCAACACCAGGTACGAGTCTATCAAGGAGACAAAAGTTTTTTGG
3180
1 G F V N L F N T R Y E S I K E T K V F W
5' TCTAAGTTTGATATTATCCGATACAATAAAGAGAAGAATTGGTTCGAGTTCGTCTTCGAT
3240
1 S K F D I I R Y N K E K N W F E F V F D
5' TACAATACCTTTACGACTAAAGCGGAGGGAACACGCACTAAGTGGACTCTGTGCACCCAC
3300
1 Y N T F T T K A E G T R T K W T L C T H
5' GGCACTCGCATCCAGACATTCCGGAACCCAGAAAAGAATGCCCAGTGGGACAATAAAGAG
3360
1 G T R I Q T F R N P E K N A Q W D N K E

圖 41G

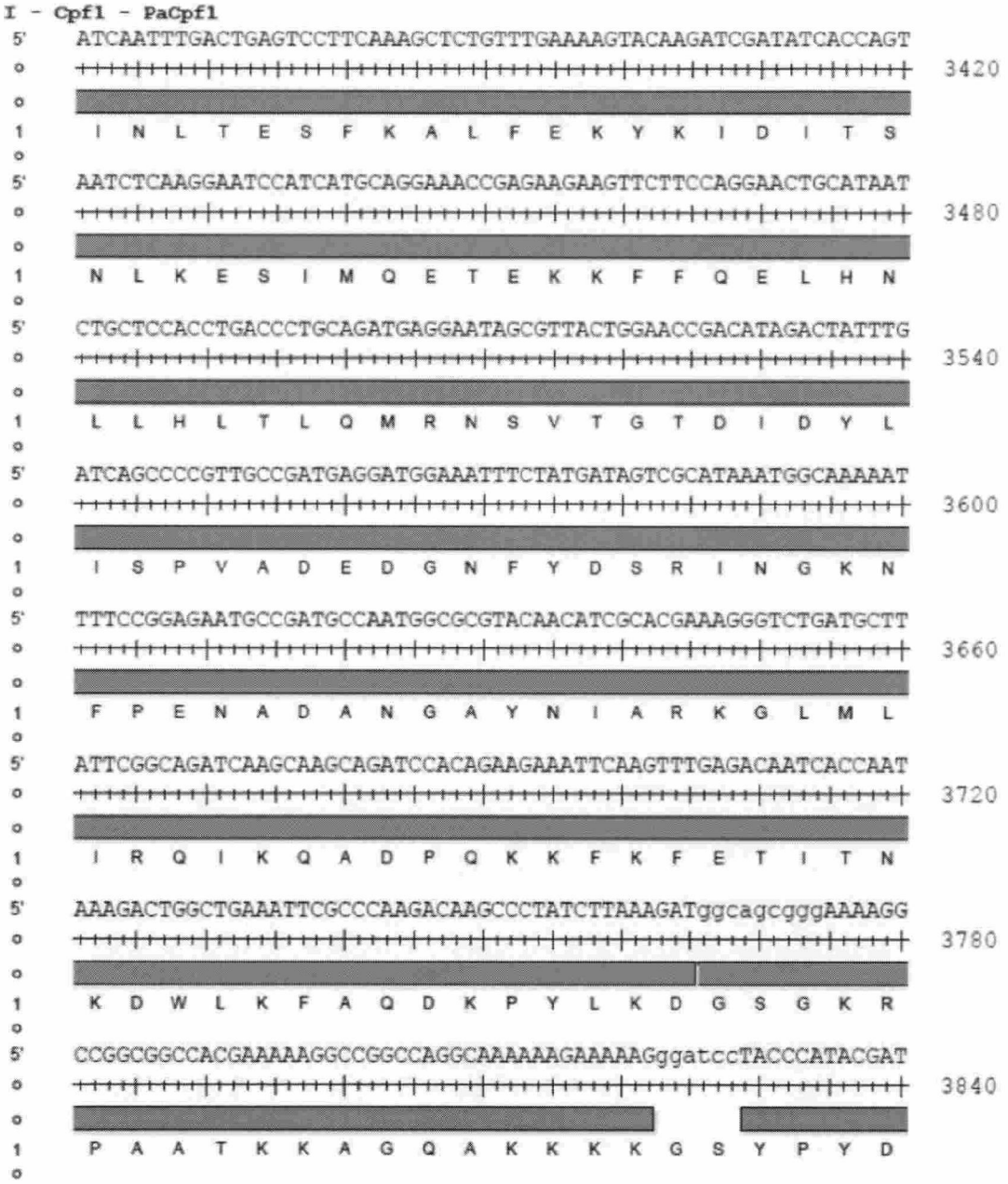

I - Cpf1 - PaCpf1
5' ATCAATTTGACTGAGTCCTTCAAAGCTCTGTTTGAAAAGTACAAGATCGATATCACCAGT
3420
I N L T E S F K A L F E K Y K I D I T S
5' AATCTCAAGGAATCCATCATGCAGGAAACCGAGAAGAAGTTCTTCCAGGAACTGCATAAT
3480
N L K E S I M Q E T E K K F F Q E L H N
5' CTGCTCCACCTGACCCTGCAGATGAGGAATAGCGTTACTGGAACCGACATAGACTATTTG
3540
L L H L T L Q M R N S V T G T D I D Y L
5' ATCAGCCCCGTTGCCGATGAGGATGGAAATTTCTATGATAGTCGCATAAATGGCAAAAAT
3600
I S P V A D E D G N F Y D S R I N G K N
5' TTTCCGGAGAATGCCGATGCCAATGGCGCGTACAACATCGCACGAAAGGGTCTGATGCTT
3660
F P E N A D A N G A Y N I A R K G L M L
5' ATTCGGCAGATCAAGCAAGCAGATCCACAGAAGAAATTCAAGTTTGAGACAATCACCAAT
3720
I R Q I K Q A D P Q K K F K F E T I T N
5' AAAGACTGGCTGAAATTCGCCCAAGACAAGCCCTATCTTAAAGATggcagcgggAAAAGG
3780
K D W L K F A Q D K P Y L K D G S G K R
5' CCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGggatccTACCCATACGAT
3840
P A A T K K A G Q A K K K K G S Y P Y D

圖 41H

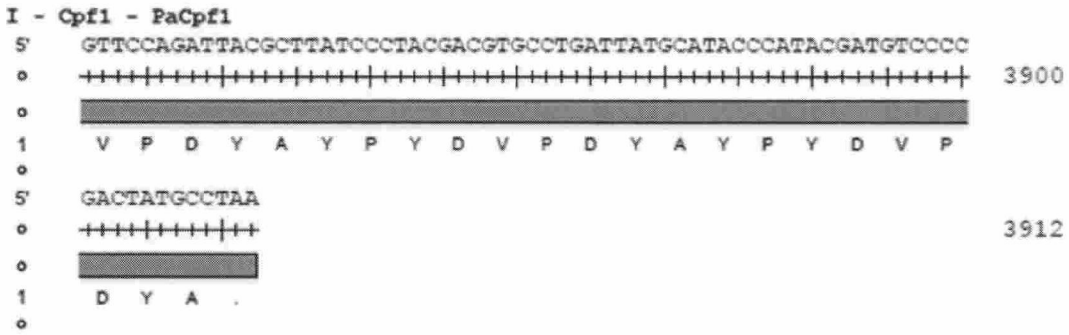

I - Cpf1 - PaCpf1
5'
GTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATACGATGTCCCC
3900
V P D Y A Y P Y D V P D Y A Y P Y D V P
5'
GACTATGCCTAA
3912
D Y A .

圖 41I

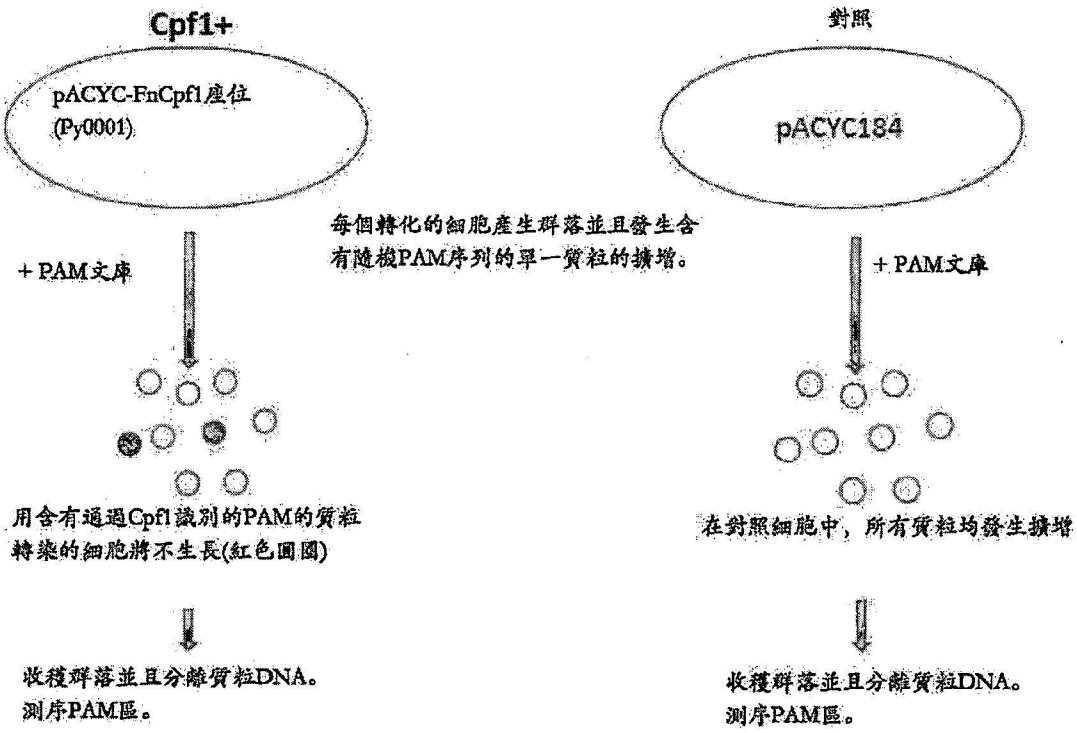

Cpf1+
pACYC-FnCpf1座位
(Py0001)
對照
pACYC184
每個轉化的細胞產生群落並且發生含有隨機PAM序列的單一質粒的擴增。
+ PAM文庫
+ PAM文庫
用含有通過Cpf1識別的PAM的質粒轉染的細胞將不生長(紅色圓圈)
在對照細胞中，所有質粒均發生擴增
收穫群落並且分離質粒DNA。
測序PAM區。
收穫群落並且分離質粒DNA。
測序PAM區。

圖 42

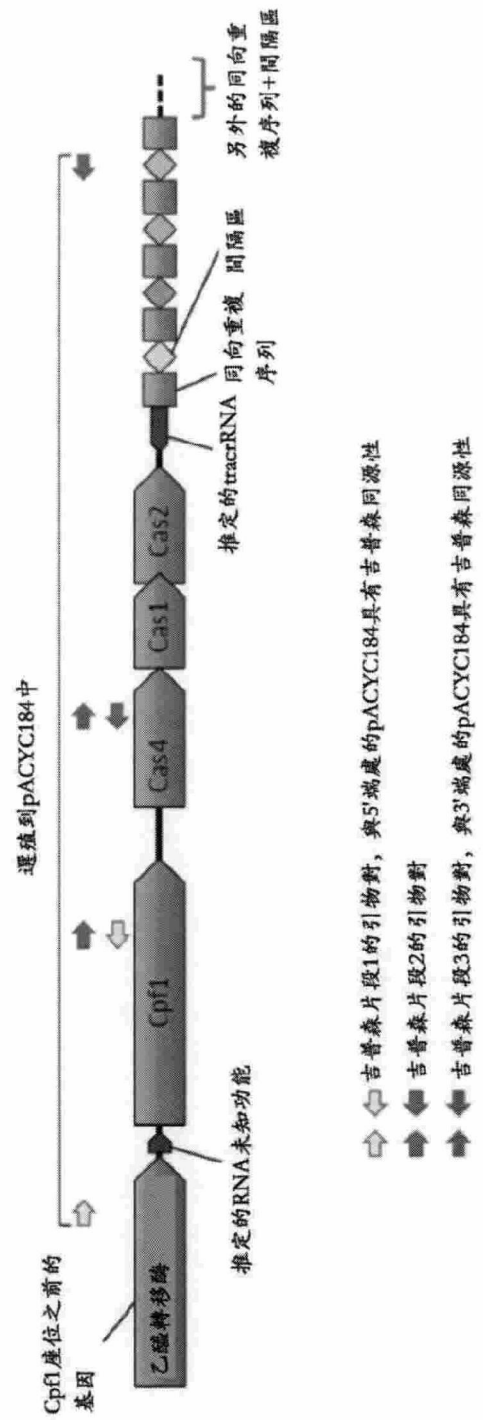

選殖到pACYC184中
Cpf1座位之前的基因
乙酸轉移酶
Cpf1
Cas4
Cas1
Cas2
推定的RNA未知功能
推定的tracrRNA
同向重複序列
間隔區
另外的同向重複序列+間隔區
吉普森片段1的引物對，與5'端處的pACYC184具有吉普森同源性
吉普森片段2的引物對
吉普森片段3的引物對，與3'端處的pACYC184具有吉普森同源性

圖 43

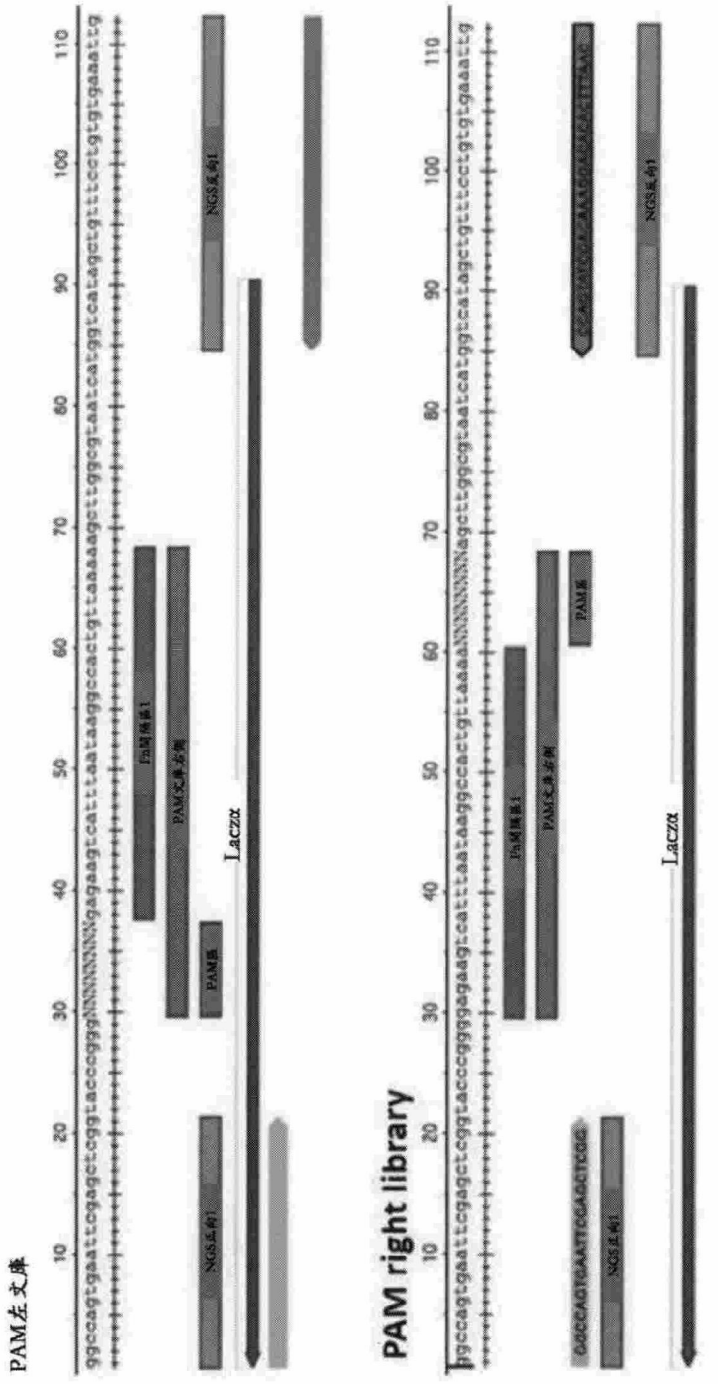


圖 44

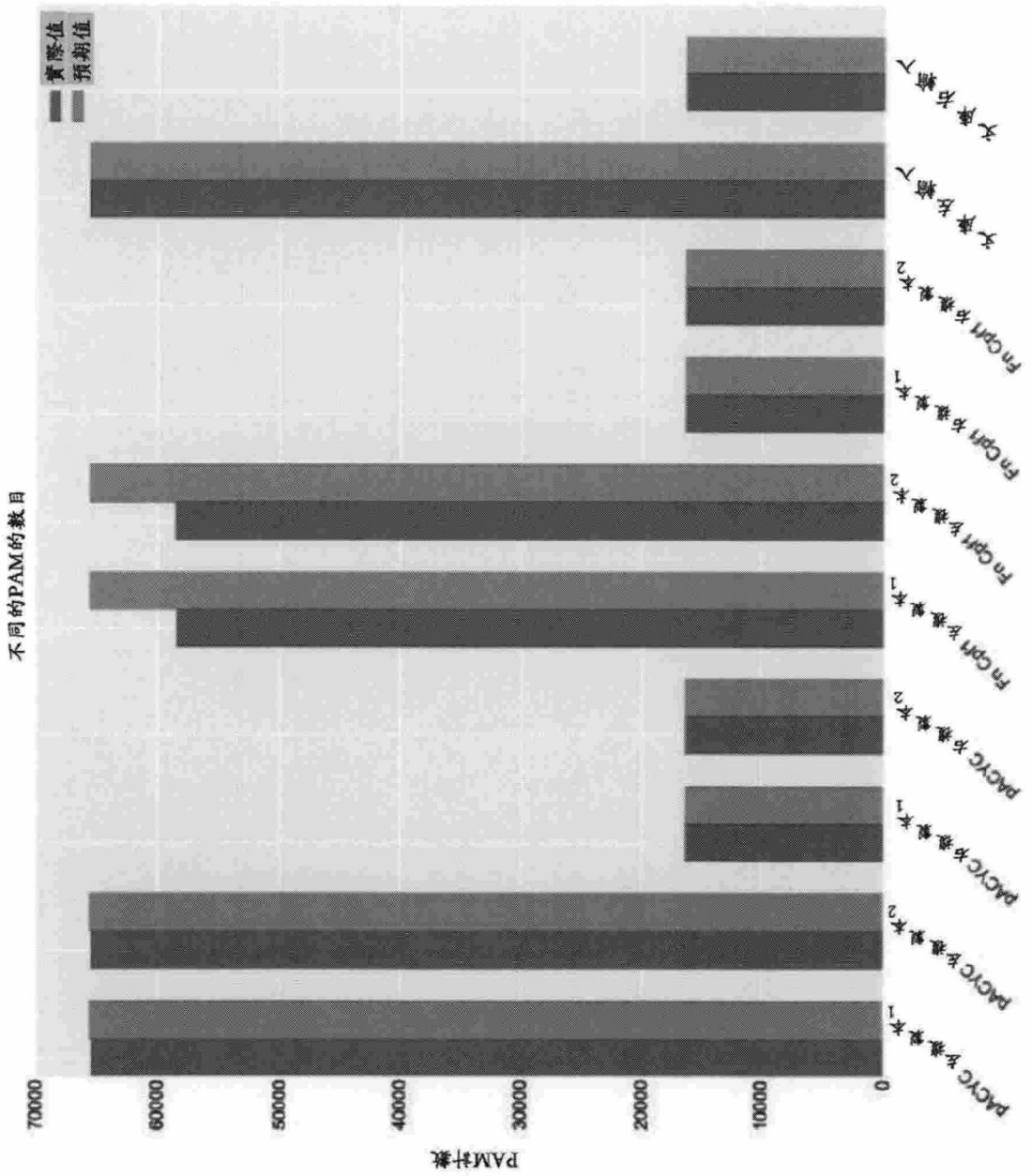

不同的PAM的數目
實際值
預期值
PAM計數
70000
60000
50000
40000
30000
20000
10000
0
PACYC 步進 集 K1
PACYC 步進 集 K2
PACYC 步進 集 K1
PACYC 步進 集 K2
Fn Cpf1 步進 集 K1
Fn Cpf1 步進 集 K2
Fn Cpf1 步進 集 K1
Fn Cpf1 步進 集 K2
文庫 步 插 入
文庫 步 插 入

圖 45A

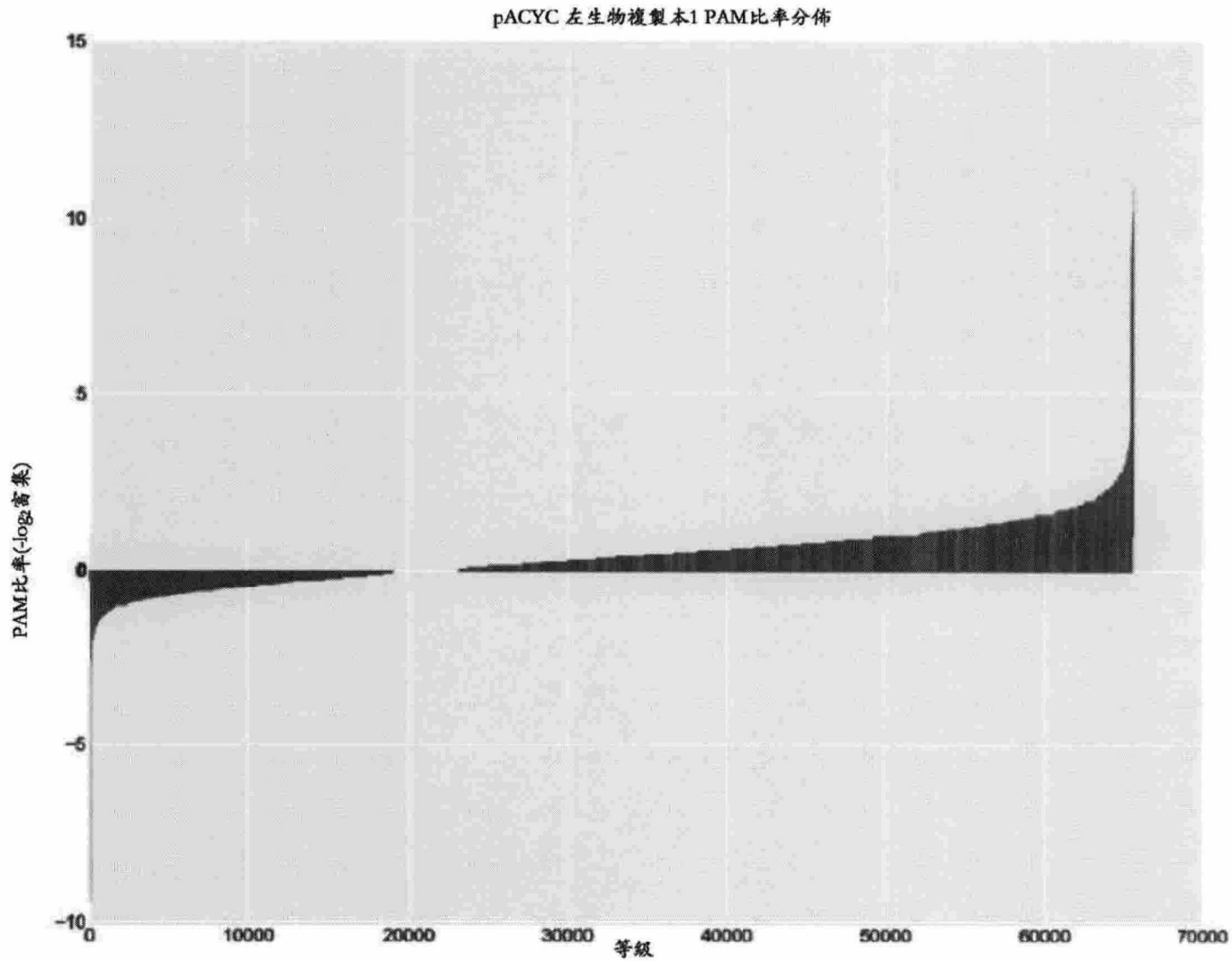

pACYC 左生物複製本1 PAM比率分佈
PAM比率(-log2富集)
15
10
5
0
-5
-10
0
10000
20000
30000
40000
50000
60000
70000
等級

圖 45B

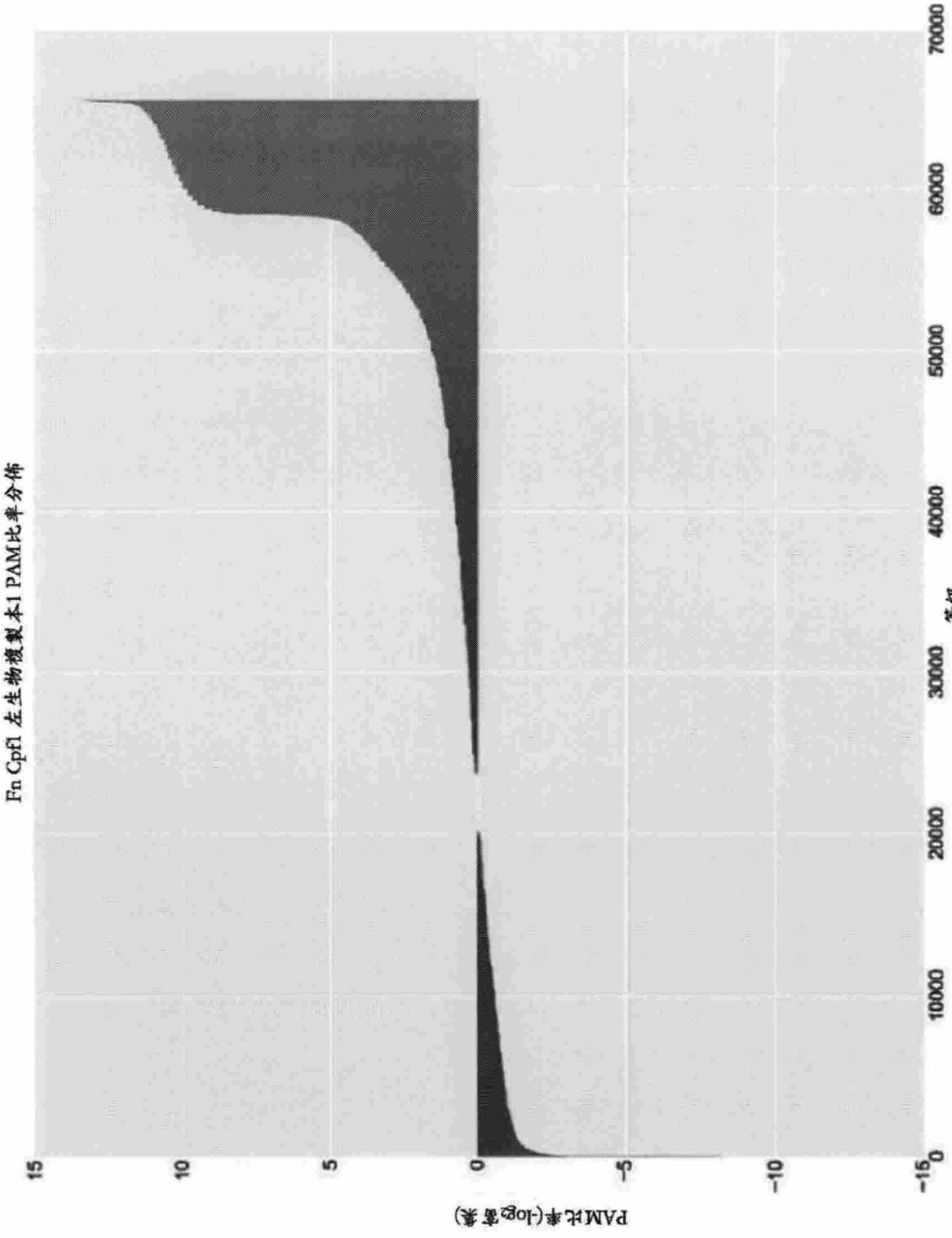

Fn Cpf1 左生物複製本1 PAM比率分佈
PAM比率(-log2富集)
15
10
5
0
-5
-10
-15
0
10000
20000
30000
40000
50000
60000
70000
等級

圖 45C

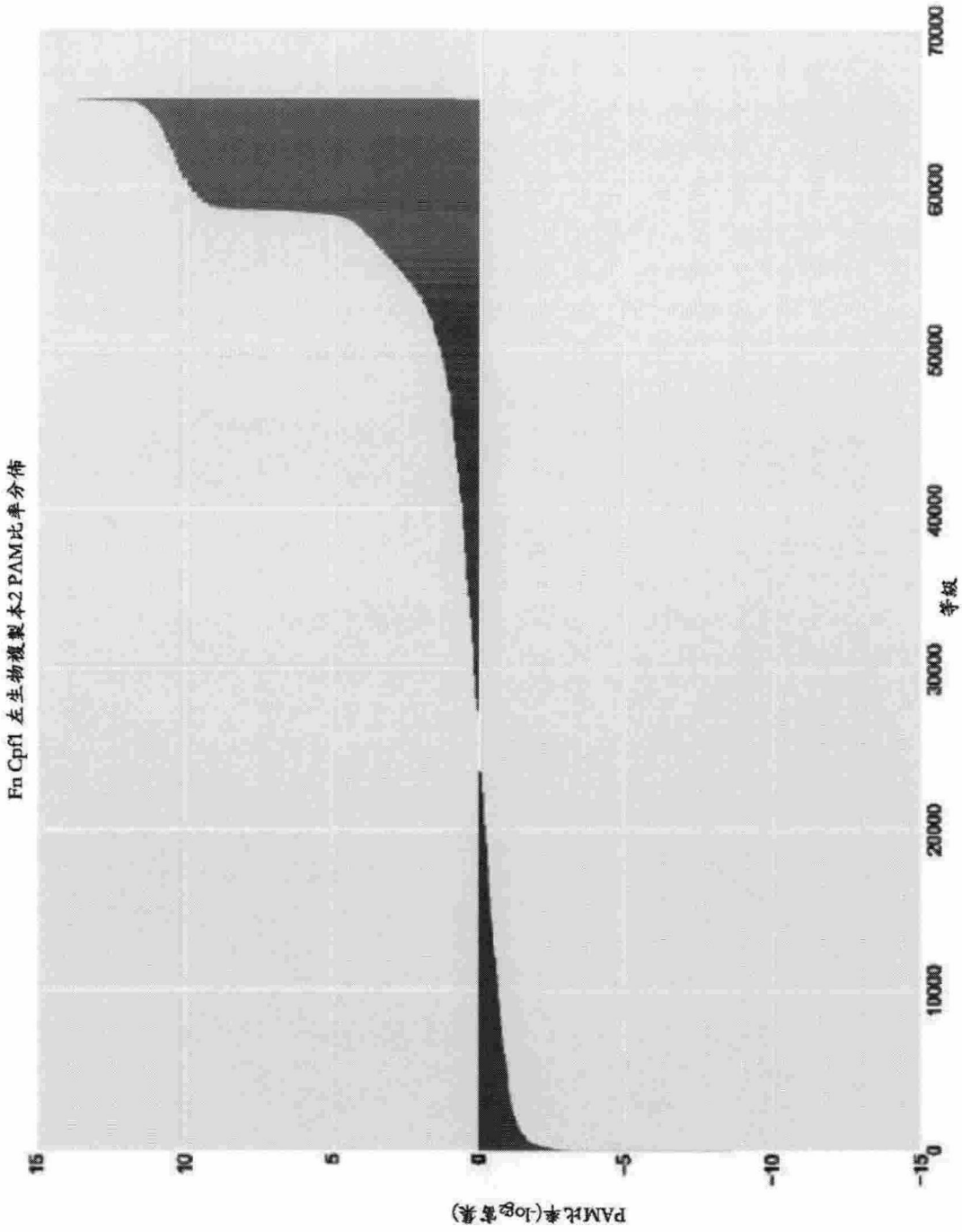

Fn Cpf1 左生物複製本2 PAM比率分佈
PAM比率(-log2富集)
15
10
5
0
-5
-10
-15
0
10000
20000
30000
40000
50000
60000
70000
等級

圖 45D

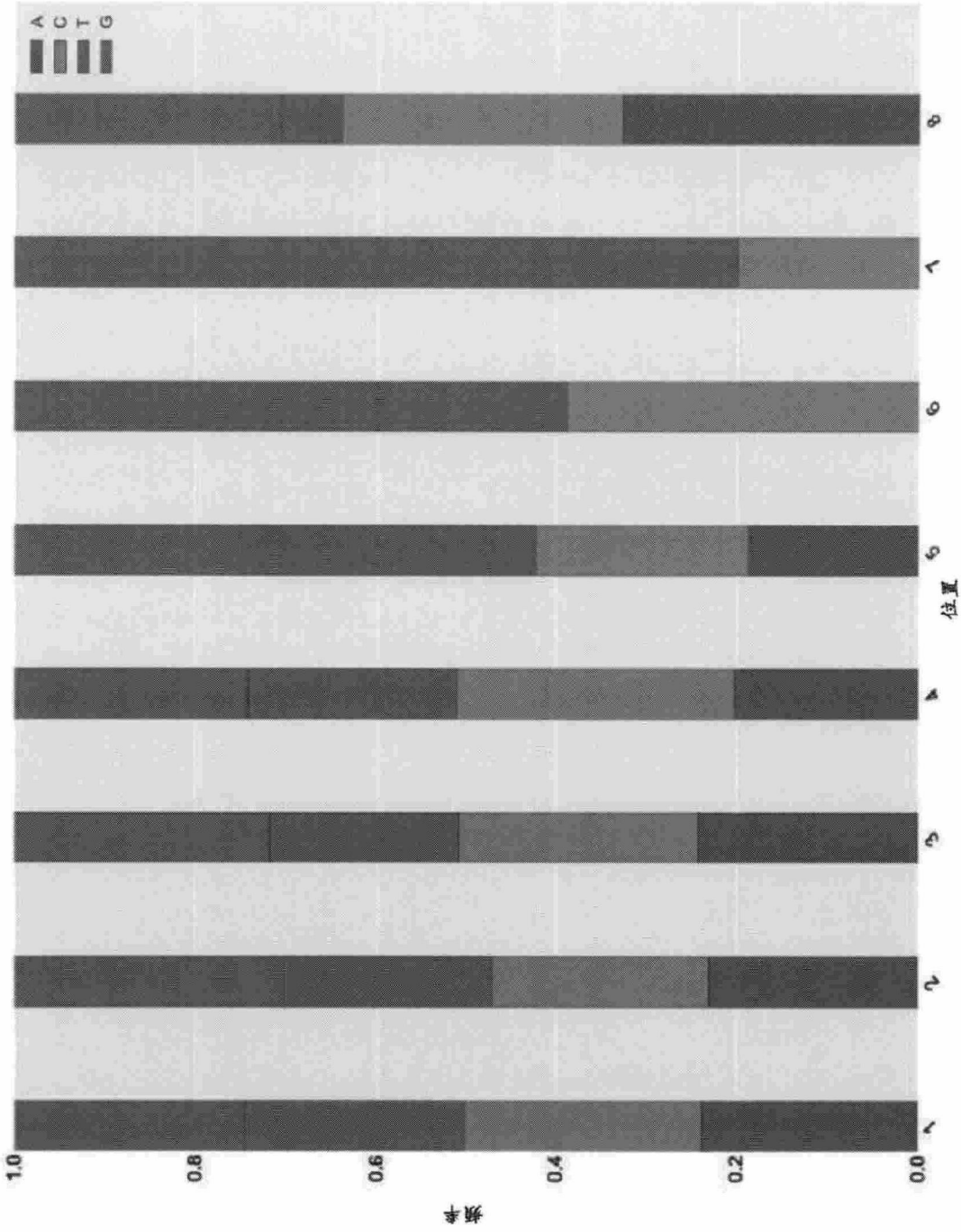


圖 45E

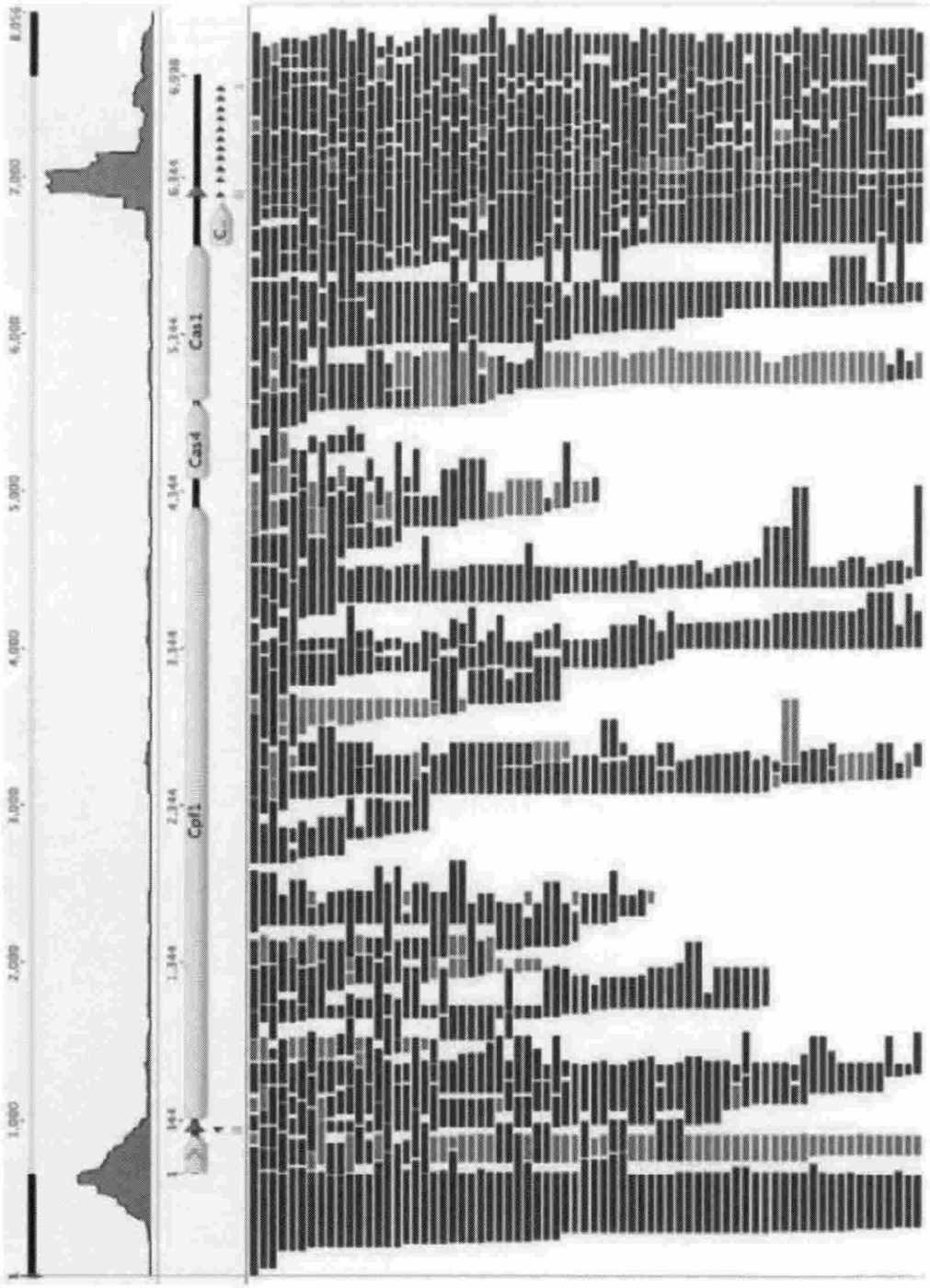


圖 46


6,700
6,800
6,900
7,000
7,100
7,200
7,300
7,400
7,500
7,600
6,044
6,144
6,244
6,344
6,444
6,544
6,644
6,744
6,844
Cas2

圖 47

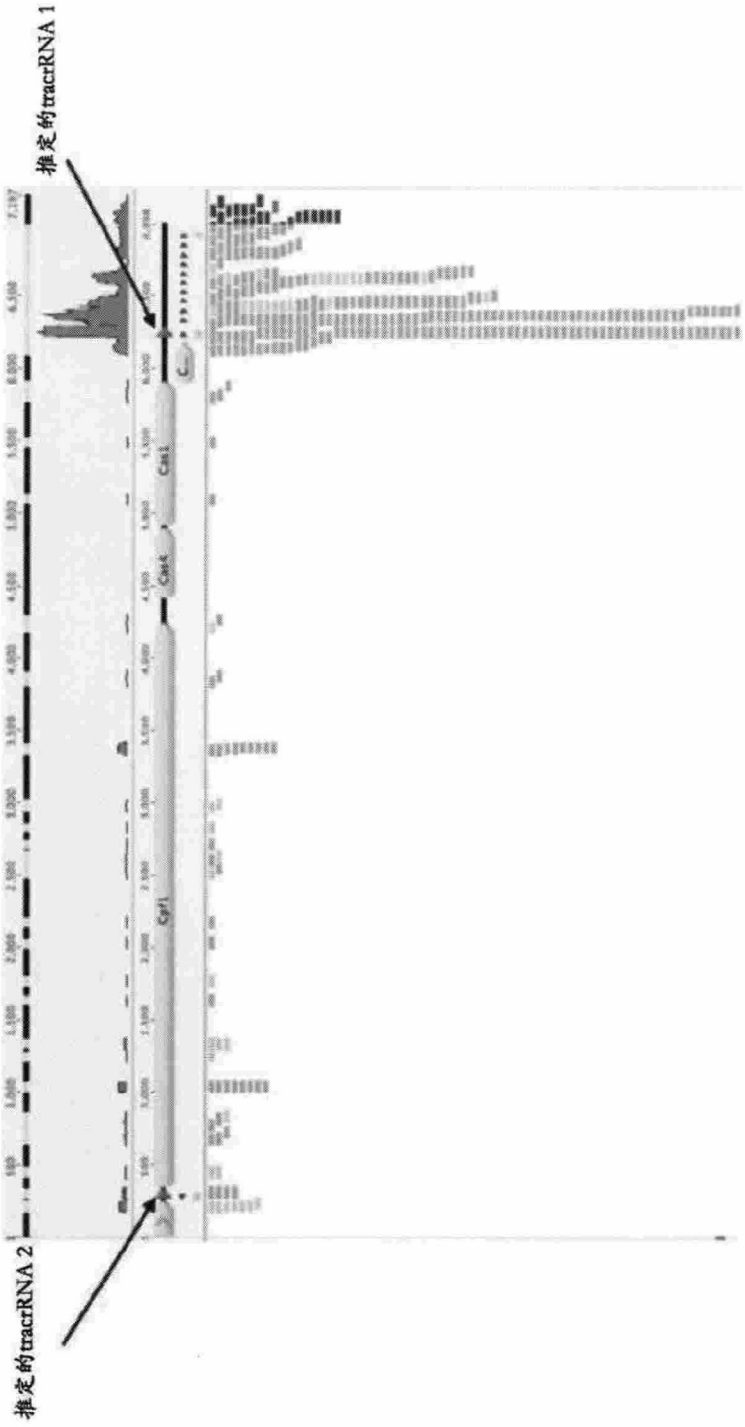


圖 48

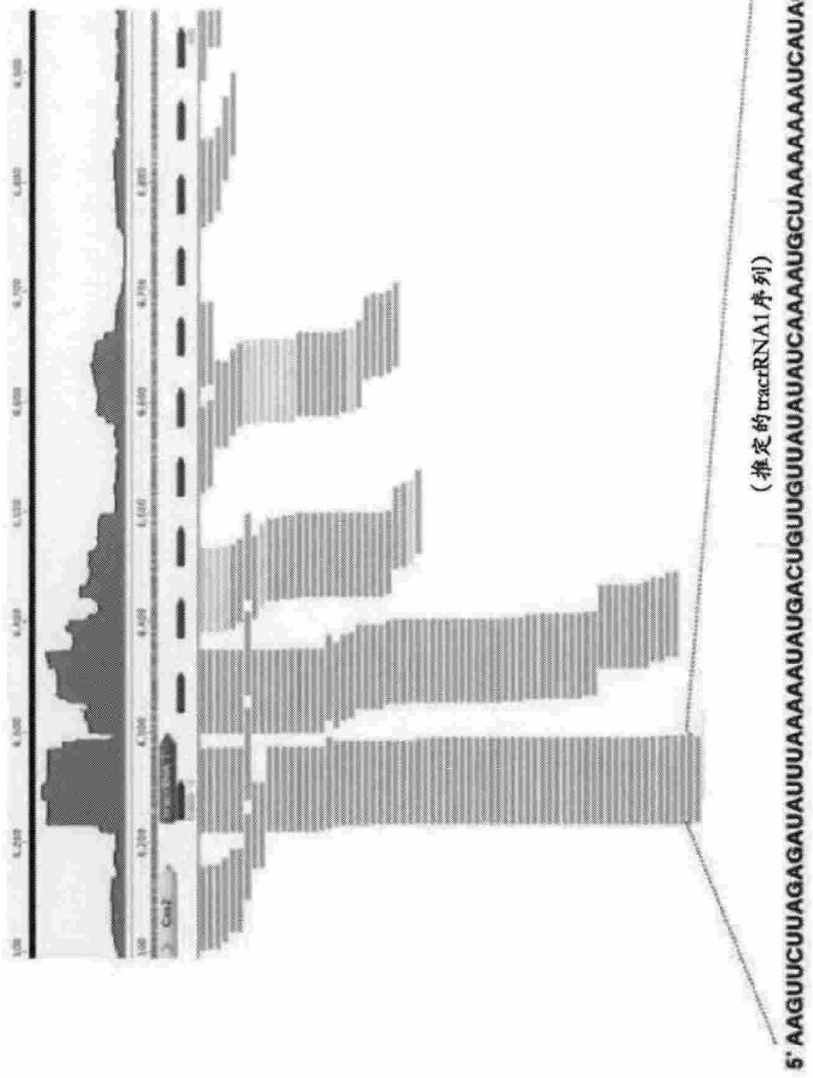

Cas2
5' AAGUUCUUAGAGAUAUUUAAAAAUAUGACUGUUGUUAUAUAUCAAAAUGCUAAAAAAAUCAUAGAUUUUAGGUCUUUUUUU 3'
(推定的tracrRNA1序列)

圖 49

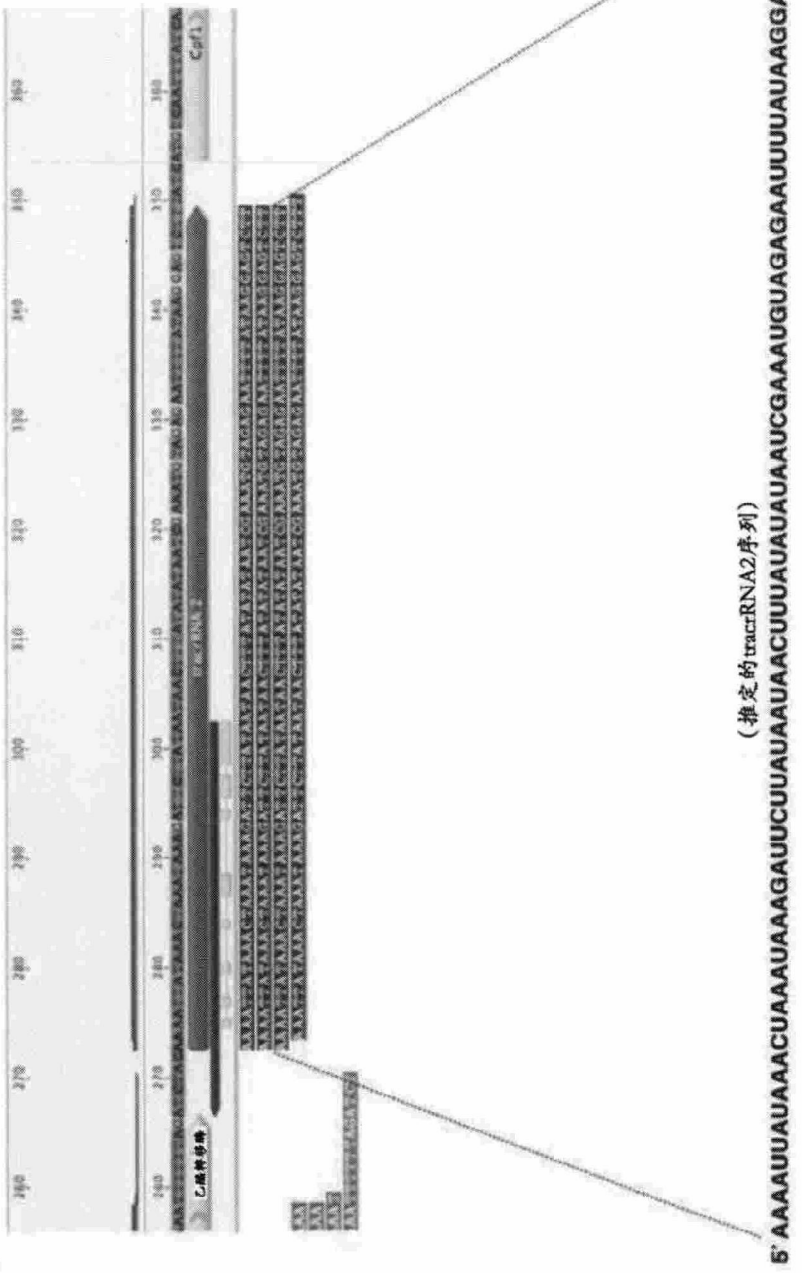

Cas1
(推定的tracrRNA2序列)
5' AAAAUUAUAAACUAAAUAAAGAUUCUUAUAAUAACUUUAUAUAAUAAUCGAAAUGUAGAGAAUUUAUAAGGAGUCUU 3'

圖 50

版本1: 5' ACTTTAAATAATTTCTACTGTTGTAGATGAGAAGTCATTTAATAAGGCCACTGTTAAAAGTCTAAGAACTTTAAATA 3'

版本2: 5' CTACTATTCCTGTGCCTTCAGATAATTCAGTCTAAGAACTTTAAATAATTTCTACTGTTGTAGATGTCTAGAGCCTT 3'

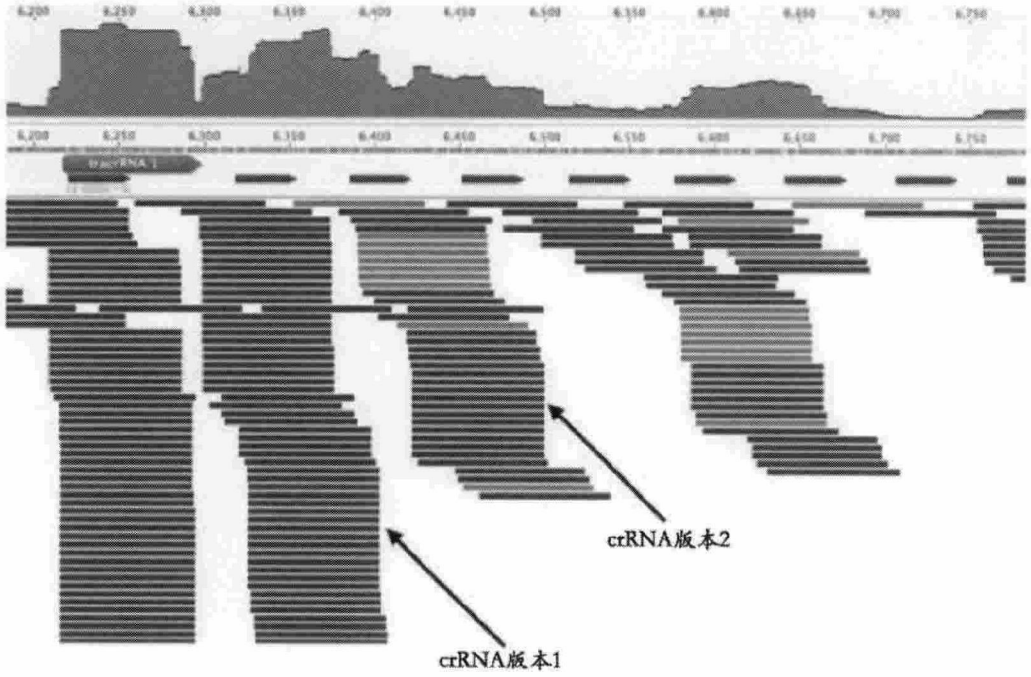


圖 51

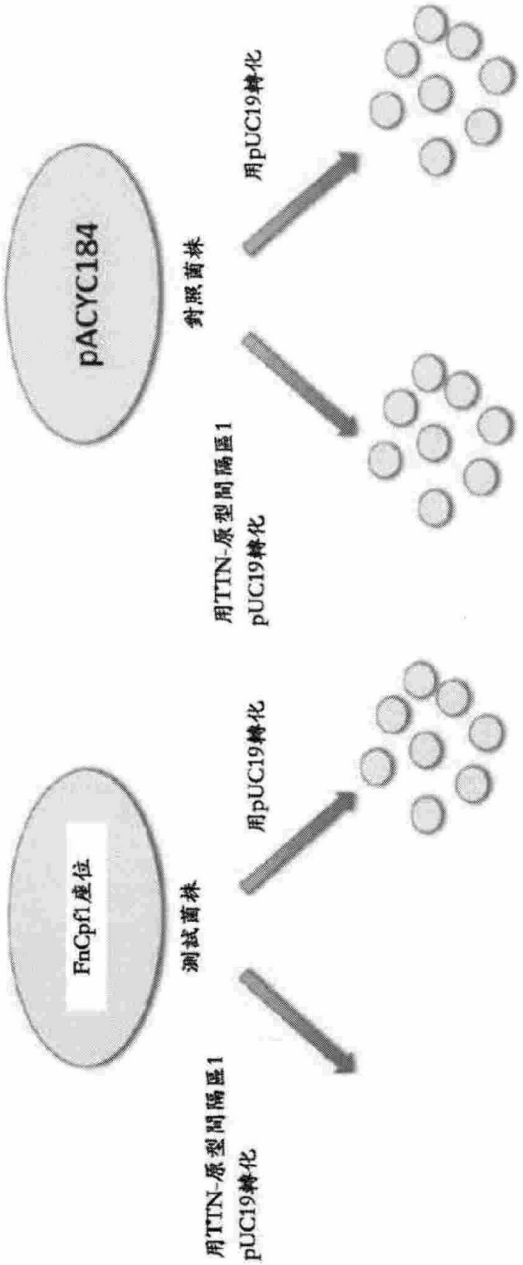

FnCpf1座位
pACYC184
測試菌株
對照菌株
用TTN-原型間隔區1
pUC19轉化
用pUC19轉化
用TTN-原型間隔區1
pUC19轉化
用pUC19轉化

圖 52

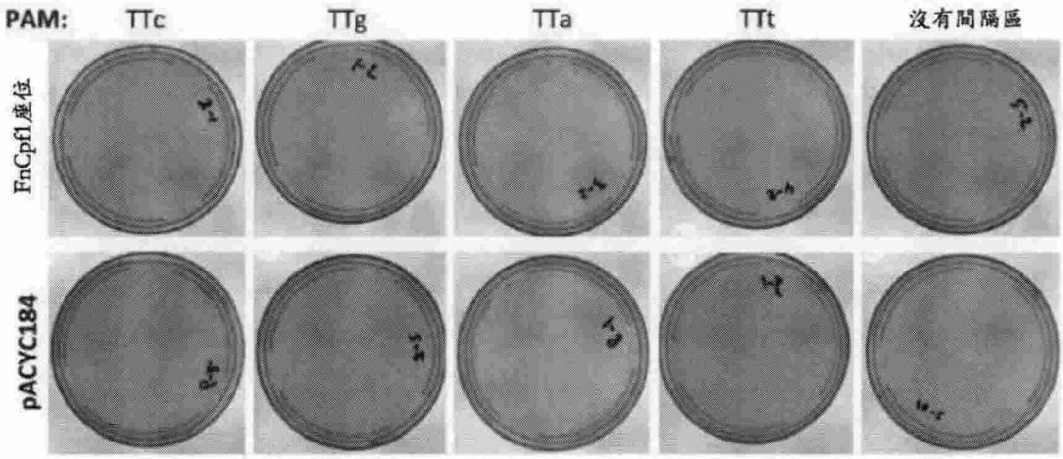


圖 53

推定的tracRNA
tracRNA號: 2
Csf1
3
Cas4
4
Cas1
Cas2
5
1
DR
間隔區
crRNA
正向引物
PAM
原型間隔區
pUC/原型間隔區
反向引物
BsaI

圖 54

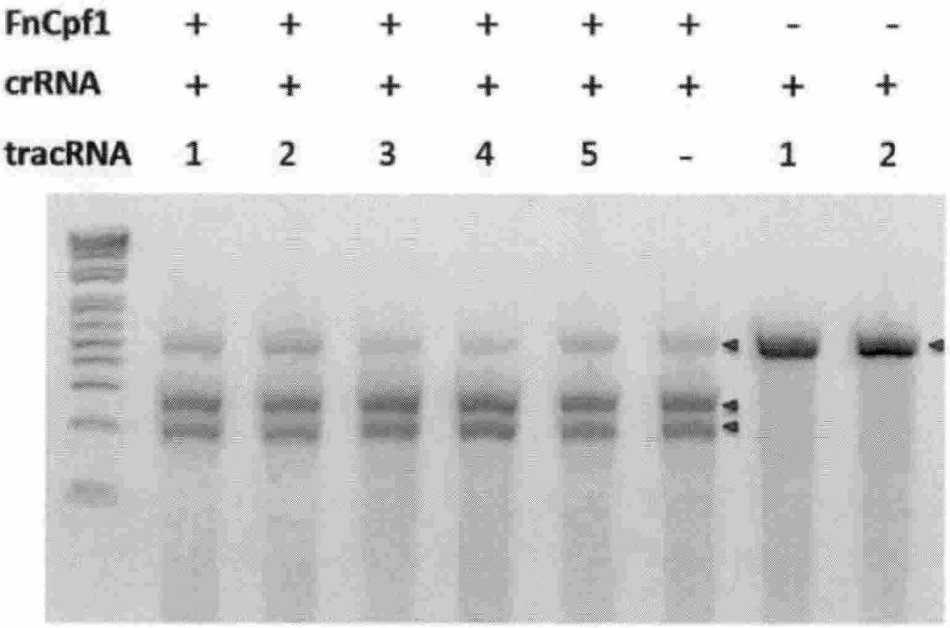

FnCpf1 + + + + + + - -
crRNA + + + + + + + +
tracRNA 1 2 3 4 5 - 1 2

圖 55

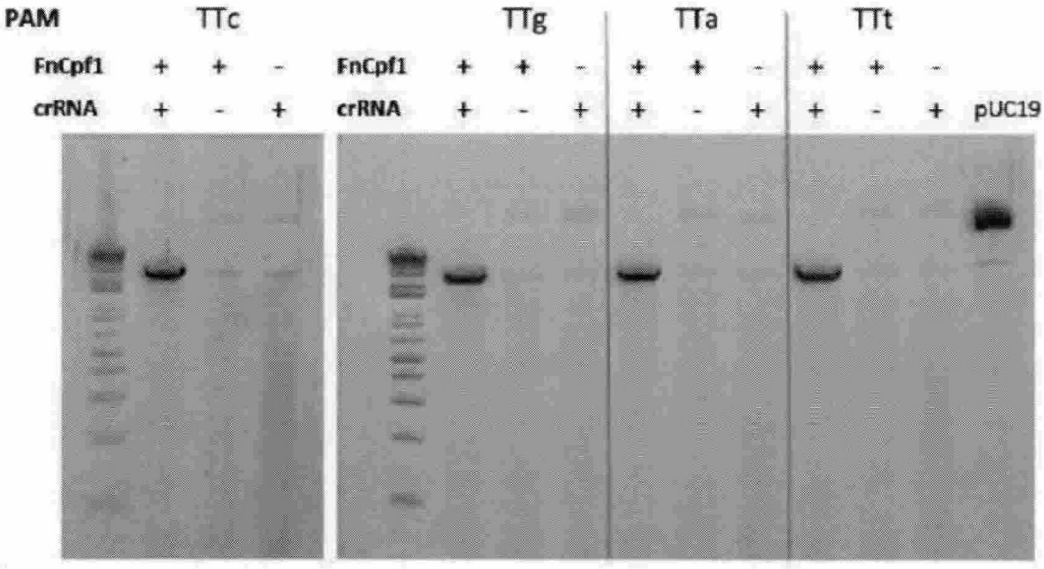
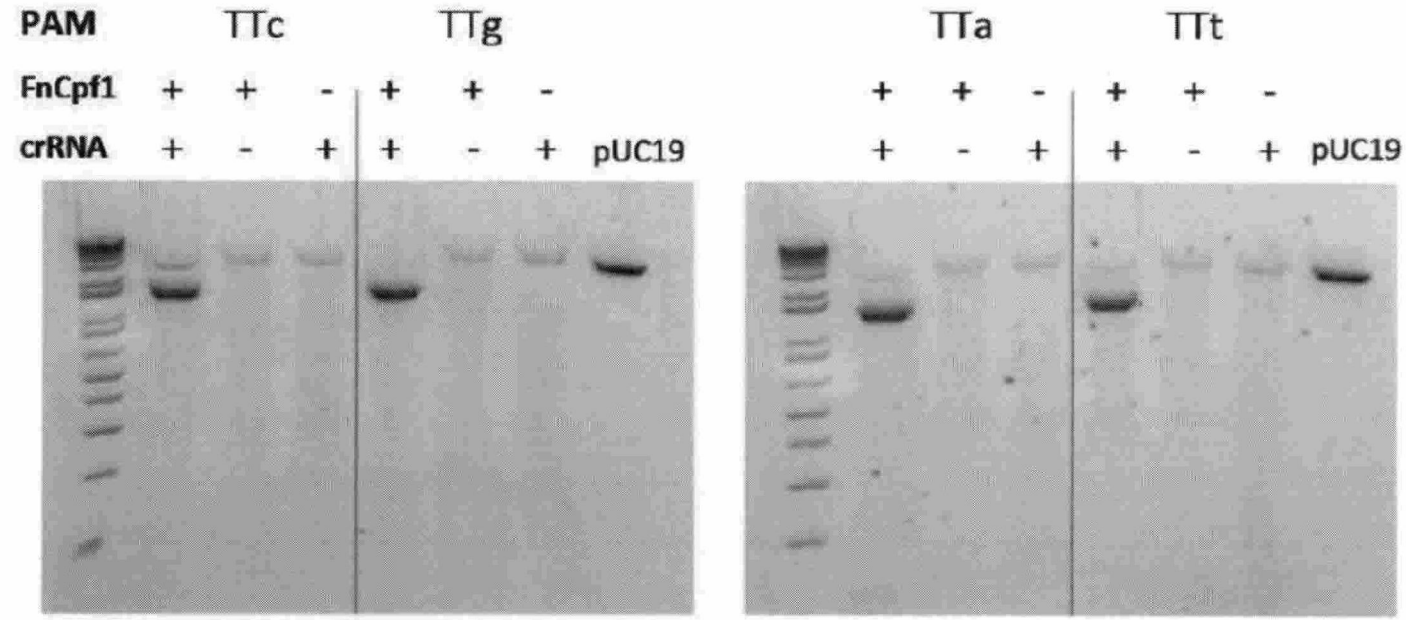
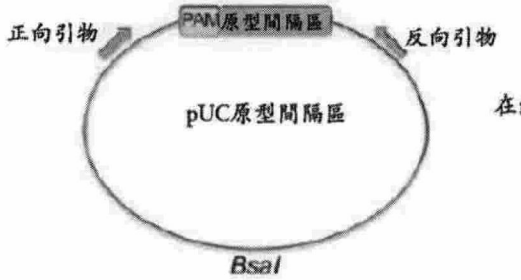

PAM
TTc
TTg
TTa
TTt
FnCpf1 + + -
crRNA + - +
pUC19

圖 56





在細胞裂解物中培養之後的Bsal消化

圖 57

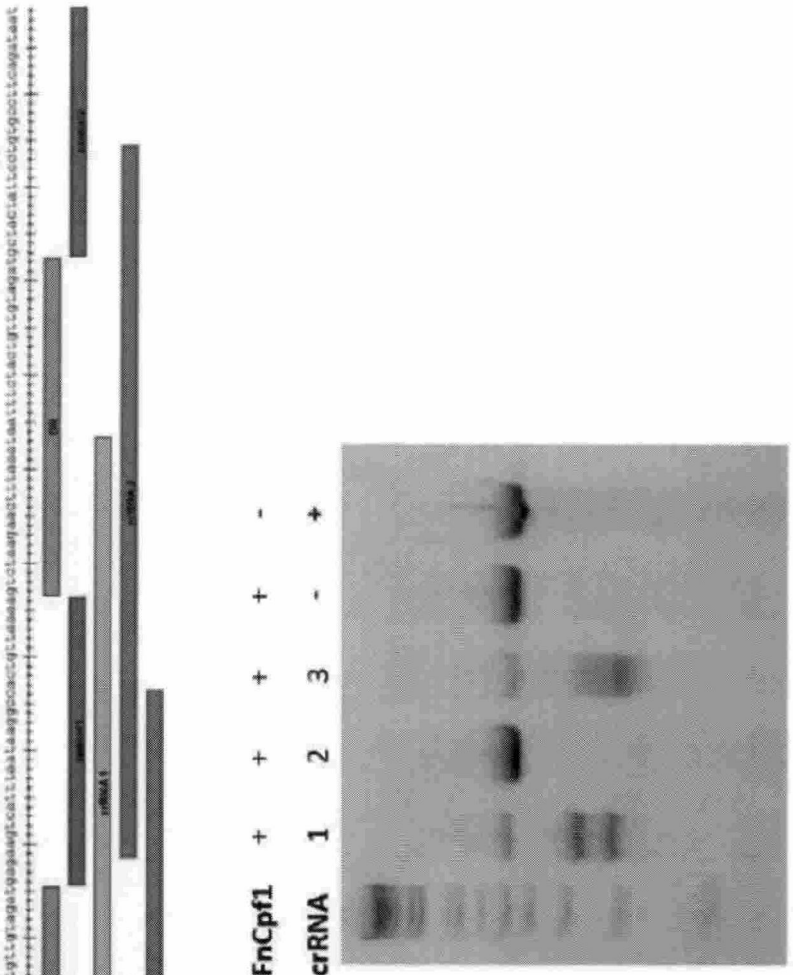
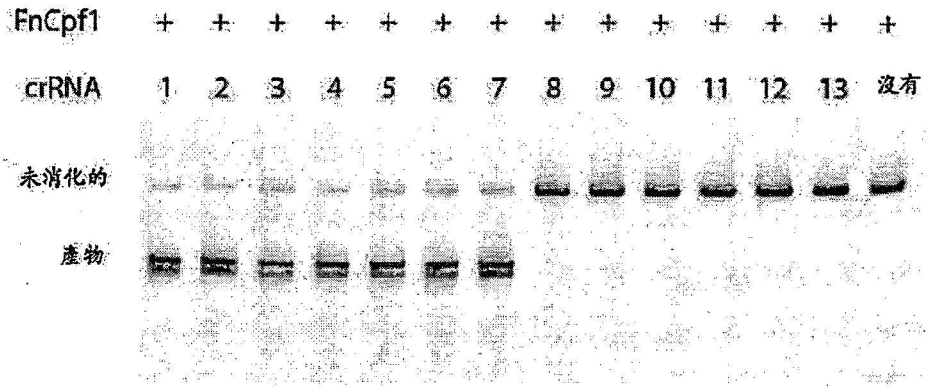


圖 58



crRNA			
1	acuuuaaauaauuucuacuguuguagau	gagaagucauuuaauaaggccacuguuaaaaagcu	gucuaagaaacuuuaaaau
2	acuuuaaauaauuucuacuguuguagau	gagaagucauuuaauaaggccacuguuaaaa	gucuaagaaacuuuaaau
3	acuuuaaauaauuucuacuguuguagau	gagaagucauuuaauaaggccacuguuaaa	gucuaagaaacuuuaaau
4	acuuuaaauaauuucuacuguuguagau	gagaagucauuuaauaaggccacuguu	gucuaagaaacuuuaaau
5	acuuuaaauaauuucuacuguuguagau	gagaagucauuuaauaaggccacu	gucuaagaaacuuuaaau
6	acuuuaaauaauuucuacuguuguagau	gagaagucauuuaauaaggcc	gucuaagaaacuuuaaau
7	acuuuaaauaauuucuacuguuguagau	gagaagucauuuaauaa	gucuaagaaacuuuaaau
8	acuuuaaauaauuucuacuguuguagau	gagaagucauuuaa	gucuaagaaacuuuaaau
9	acuuuaaauaauuucuacuguuguagau	agucauuuaauaaggccacuguuaaaa	gucuaagaaacuuuaaau
10	acuuuaaauaauuucuacuguuguagau	cauuuaauaaggccacuguuaaaa	gucuaagaaacuuuaaau
11	acuuuaaauaauuucuacuguuguagau	uaauaaggccacuguuaaaa	gucuaagaaacuuuaaau
12	acuuuaaauaauuucuacuguuguagau	uaaggccacuguuaaaa	gucuaagaaacuuuaaau
13	acuuuaaauaauuucuacuguuguagau	ggccacuguuaaaa	gucuaagaaacuuuaaau

圖 59

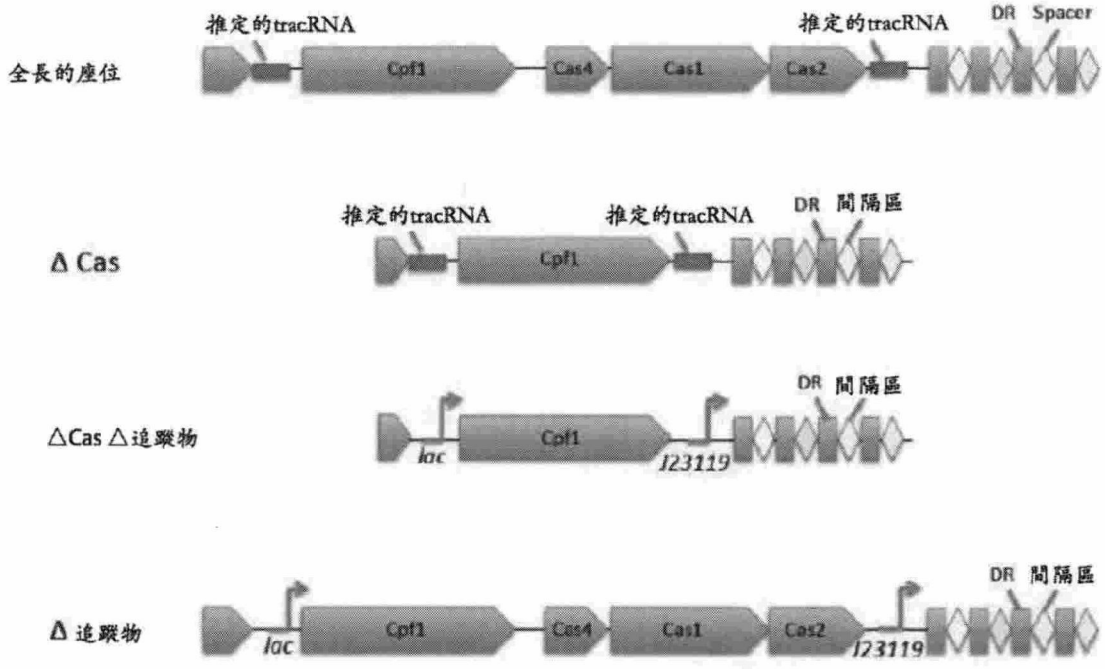

全長的座位
推定的tracRNA
Cpf1
Cas4
Cas1
Cas2
推定的tracRNA
DR
Spacer
Δ Cas
推定的tracRNA
推定的tracRNA
Cpf1
DR
間隔區
△Cas △追蹤物
lac
Cpf1
J23119
DR
間隔區
Δ 追蹤物
lac
Cpf1
Cas4
Cas1
Cas2
J23119
DR
間隔區

圖 60

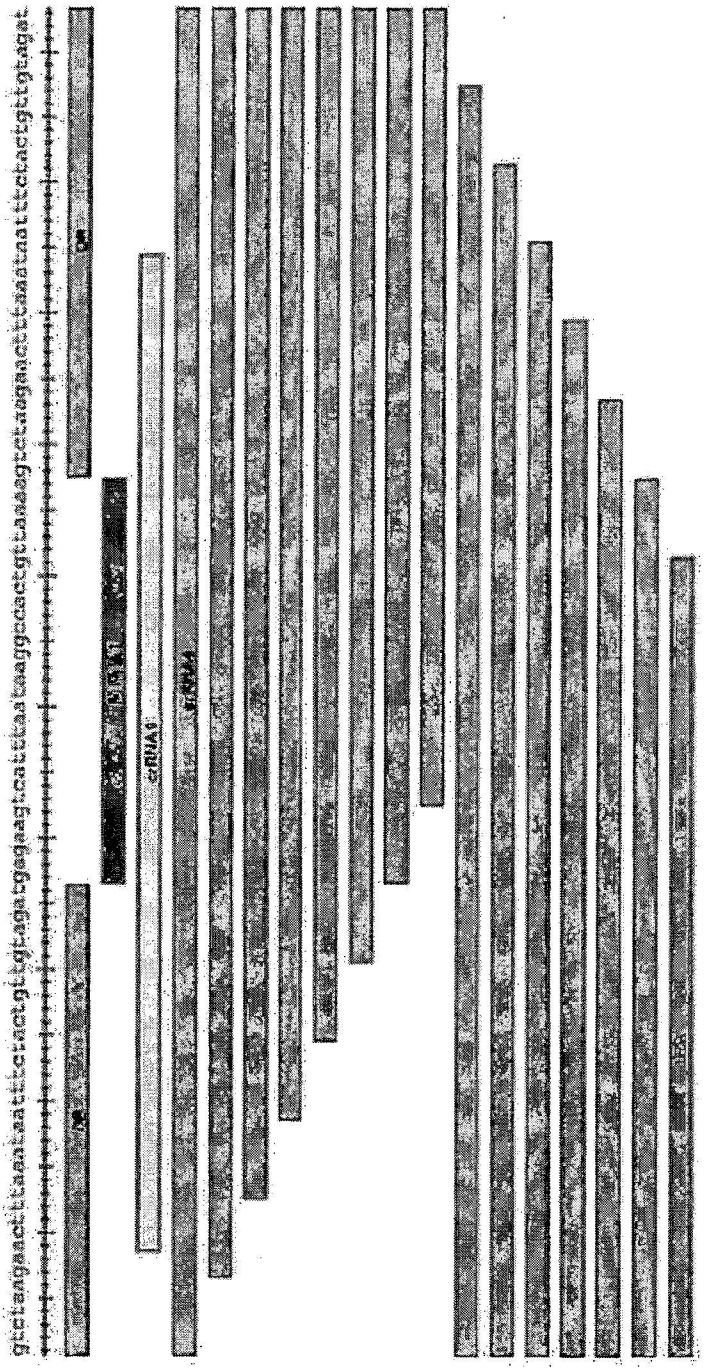


圖 61

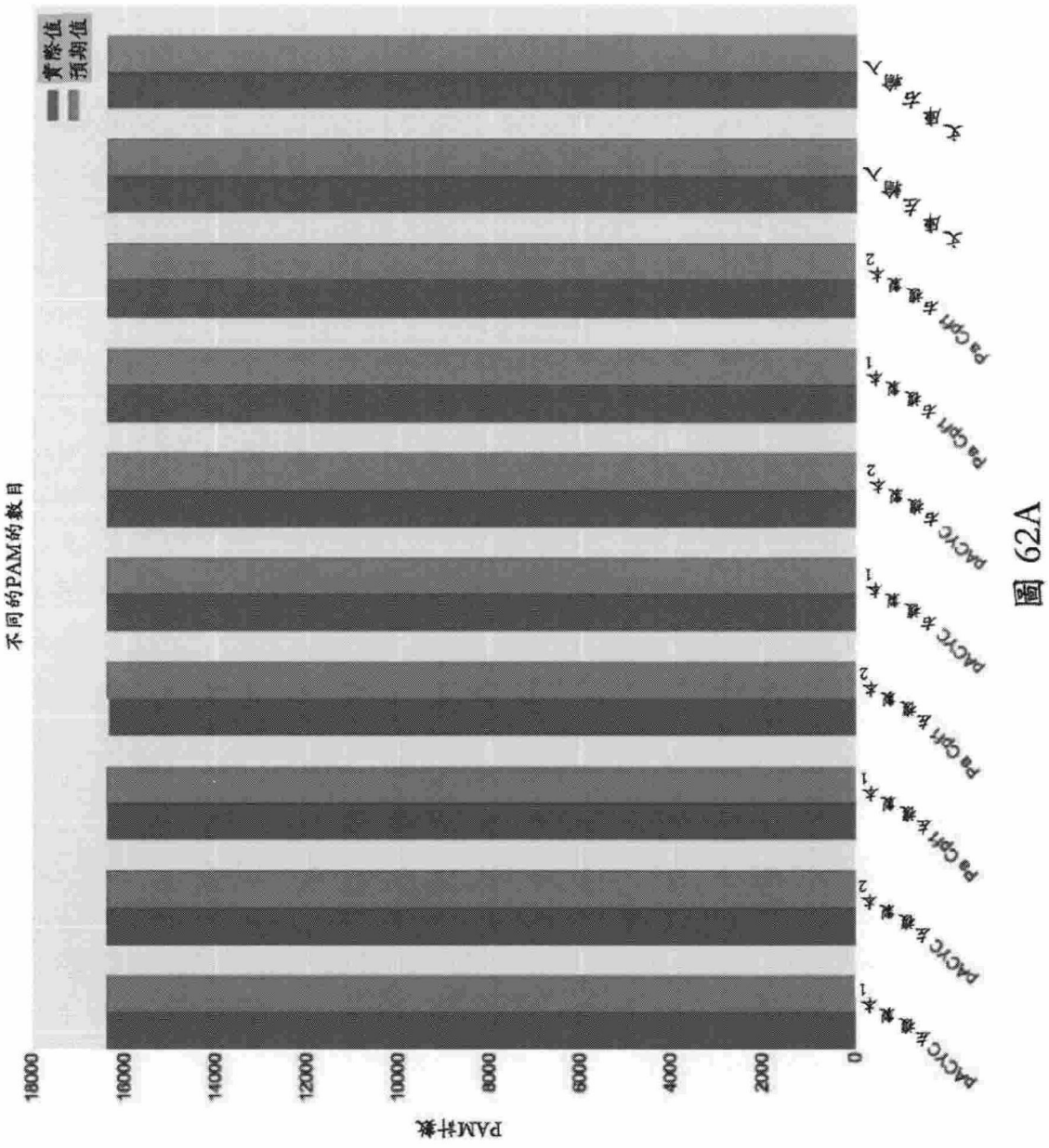

不同的PAM的數目
實際值
預期值
PAM計數
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
PACYC 步進數 K1
PACYC 步進數 K2
Pa Cpf1 步進數 K1
Pa Cpf1 步進數 K2
PACYC 步進數 K1
PACYC 步進數 K2
Pa Cpf1 步進數 K1
Pa Cpf1 步進數 K2
文庫步進 1
文庫步進 1

圖 62A

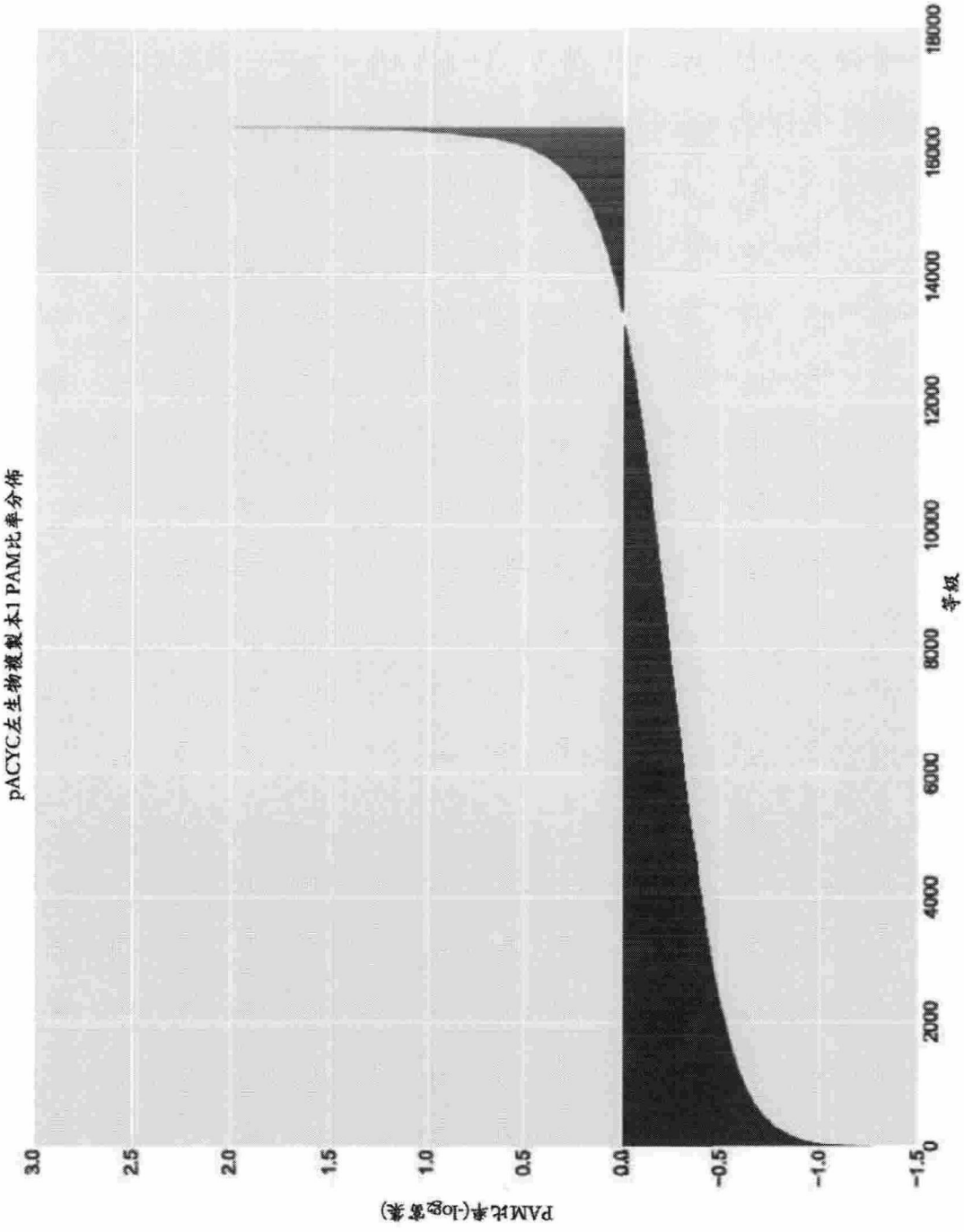

pACYC左生物複製本1 PAM比率分佈
PAM比率(-log2等級)
等級
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
0
2000
4000
6000
8000
10000
12000
14000
16000
18000

圖 62B

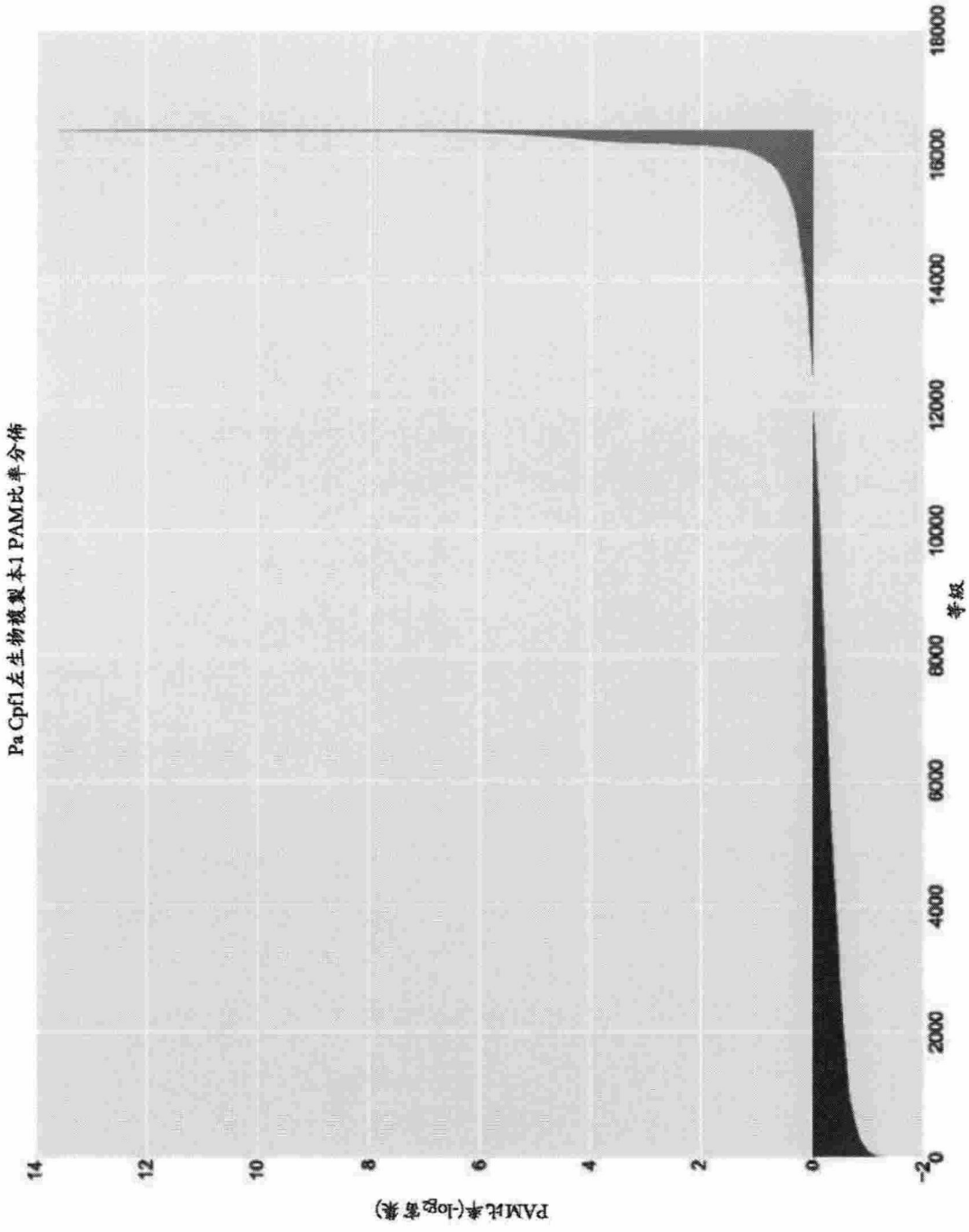

Pa Cpf1左生物複製本1 PAM比率分佈
PAM比率(-log2密集)
等級
14
12
10
8
6
4
2
0
-2
0
2000
4000
6000
8000
10000
12000
14000
16000
18000

圖 62C

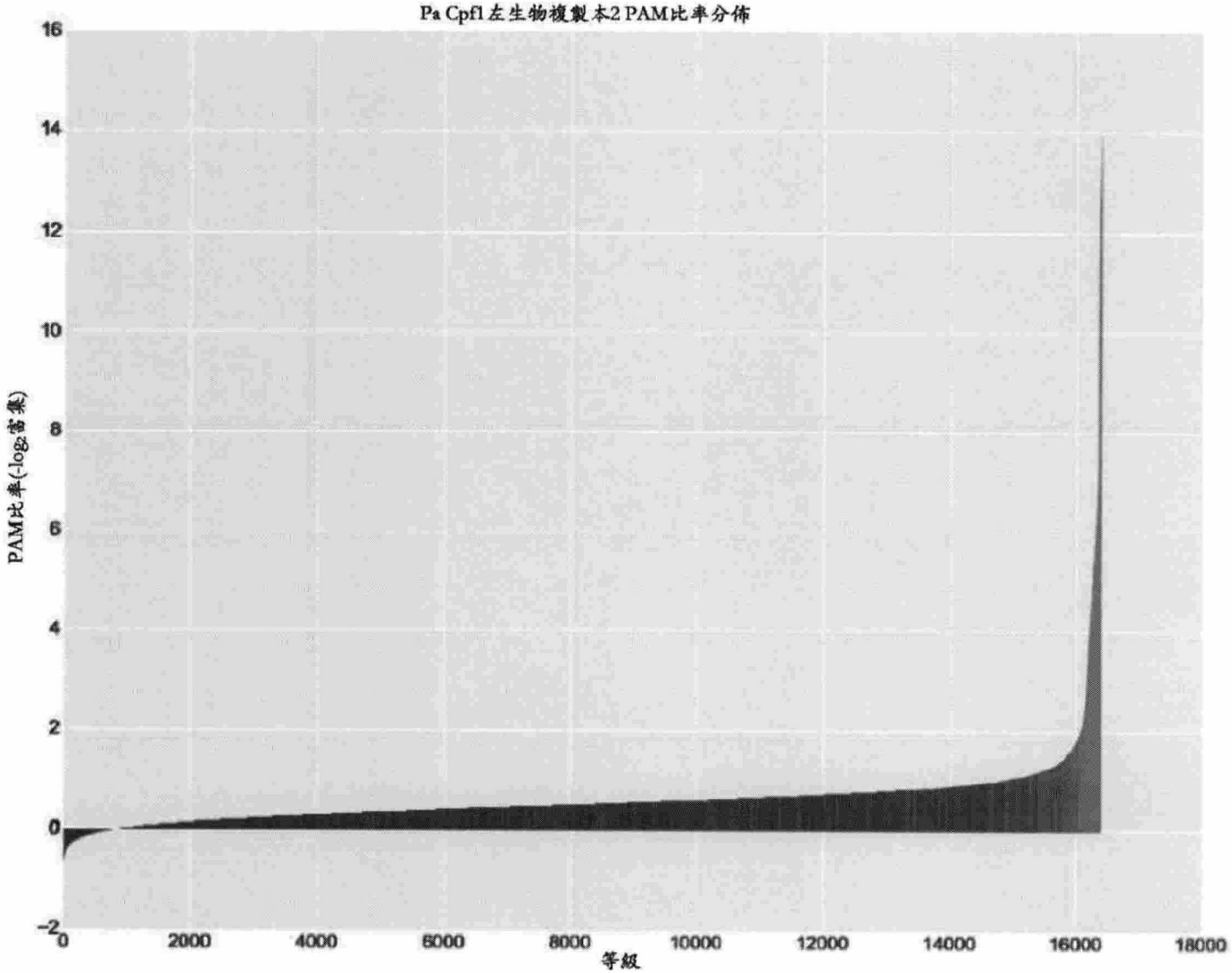

Pa Cpf1左生物複製本2 PAM比率分佈
PAM比率(-log2富集)
16
14
12
10
8
6
4
2
0
-2
0
2000
4000
6000
8000
10000
12000
14000
16000
18000
等級

圖 62D

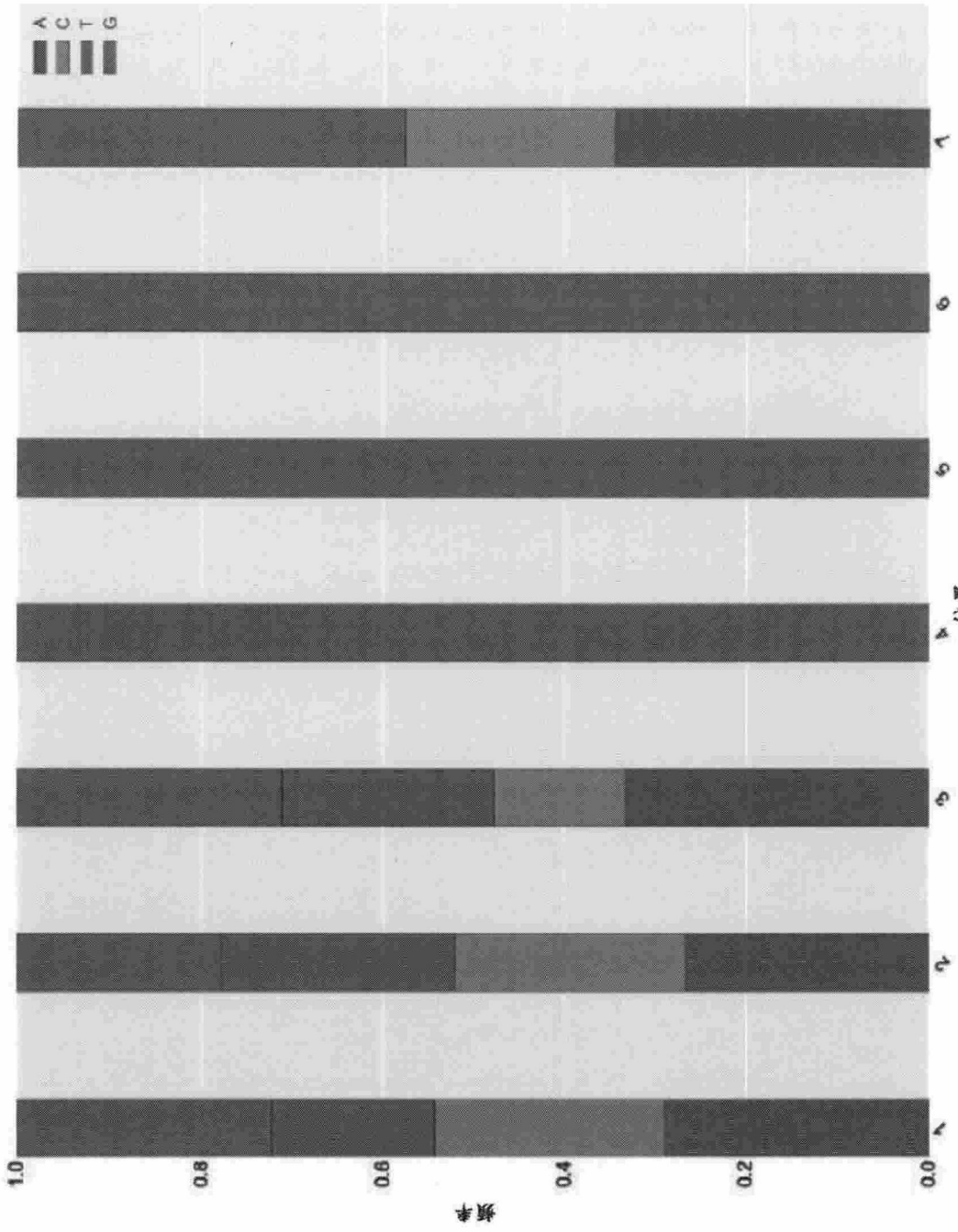


圖 62E

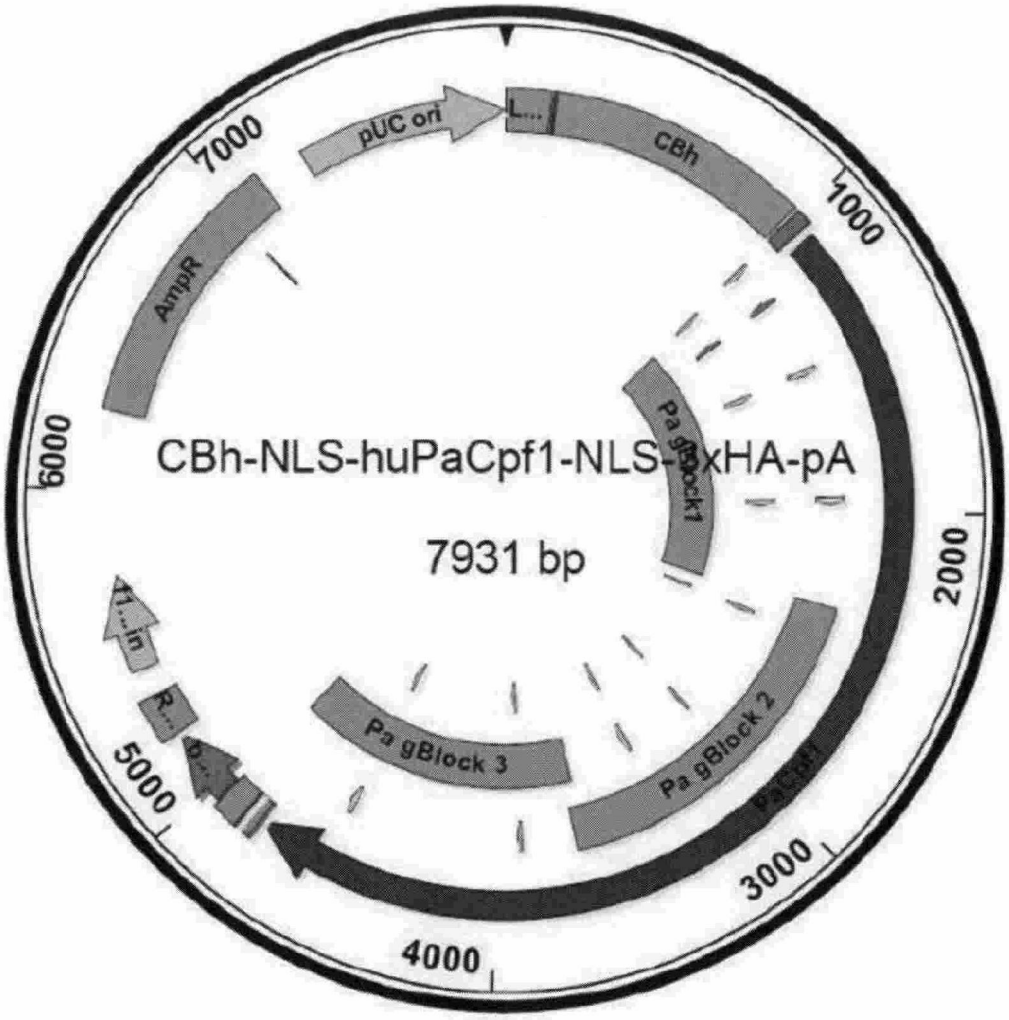

7000
1000
2000
3000
4000
5000
6000
pUC ori
L...
CBh
AmpR
Pa gBlock1
CBh-NLS-huPaCpf1-NLS-3xHA-pA
7931 bp
f1...in
R...
Pa gBlock 3
Pa gBlock 2
PaCpf1

圖 63

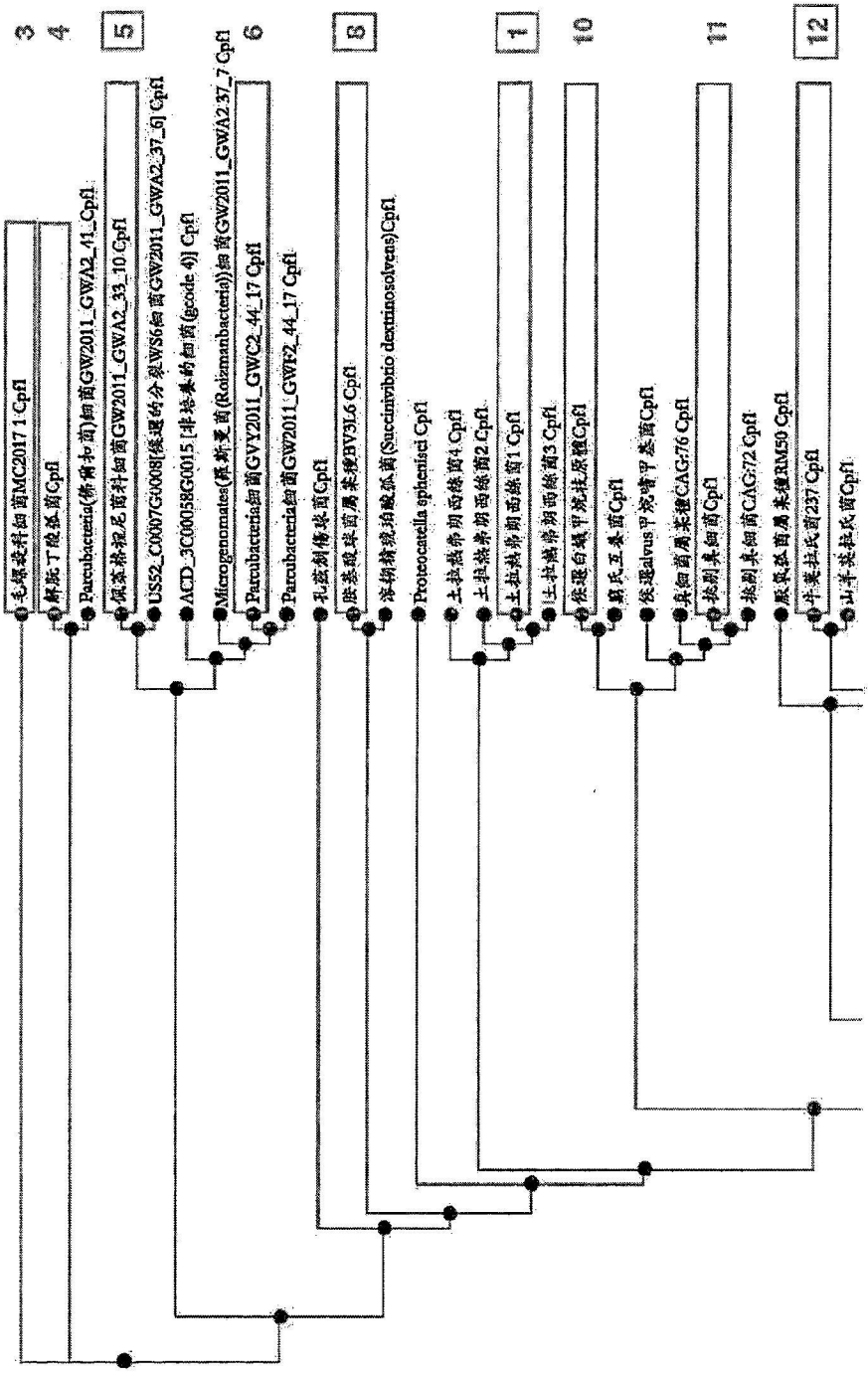

3
4
5
6
8
1
10
11
12
毛螺旋科細菌MC2017 1 Cpfl
解脂丁酸弧菌Cpfl
Parcubacteria(佛爾和菌)細菌GW2011_GWA2_41_Cpfl
佩萊格拉尼菌科細菌GW2011_GWA2_33_10 Cpfl
US52_C0007G0008[候選的分裂WS6細菌GW2011_GWA2_37_6] Cpfl
ACD_3C00058G0015 [非培養的細菌(gcode 4)] Cpfl
Microgenomates(蘇斯麥菌(Roizmanbacteria))細菌GW2011_GWA2 37_7 Cpfl
Parcubacteria細菌GVY2011_GWC2_44_17 Cpfl
Parcubacteria細菌GW2011_GWF2_44_17 Cpfl
孔茲劍傷球菌Cpfl
胺基酸球菌屬未種BV3L6 Cpfl
溶糊精琥珀酸弧菌(Succinivibrio dextrinosolvens)Cpfl
Proteocatella sphenisci Cpfl
土拉熱帝朗西綠菌4 Cpfl
土拉熱帝朗西綠菌2 Cpfl
土拉熱帝朗西綠菌1 Cpfl
土拉熱帝朗西綠菌3 Cpfl
候選白蟻甲烷桿原體Cpfl
鄧氏互養菌Cpfl
候選alvus甲烷嗜甲基菌Cpfl
真細菌屬未種CAG:76 Cpfl
挑剔真細菌Cpfl
挑剔真細菌CAG:72 Cpfl
脈氧弧菌屬未種RM50 Cpfl
牛莢拉氏菌237 Cpfl
山羊莢拉氏菌Cpfl

圖 64A

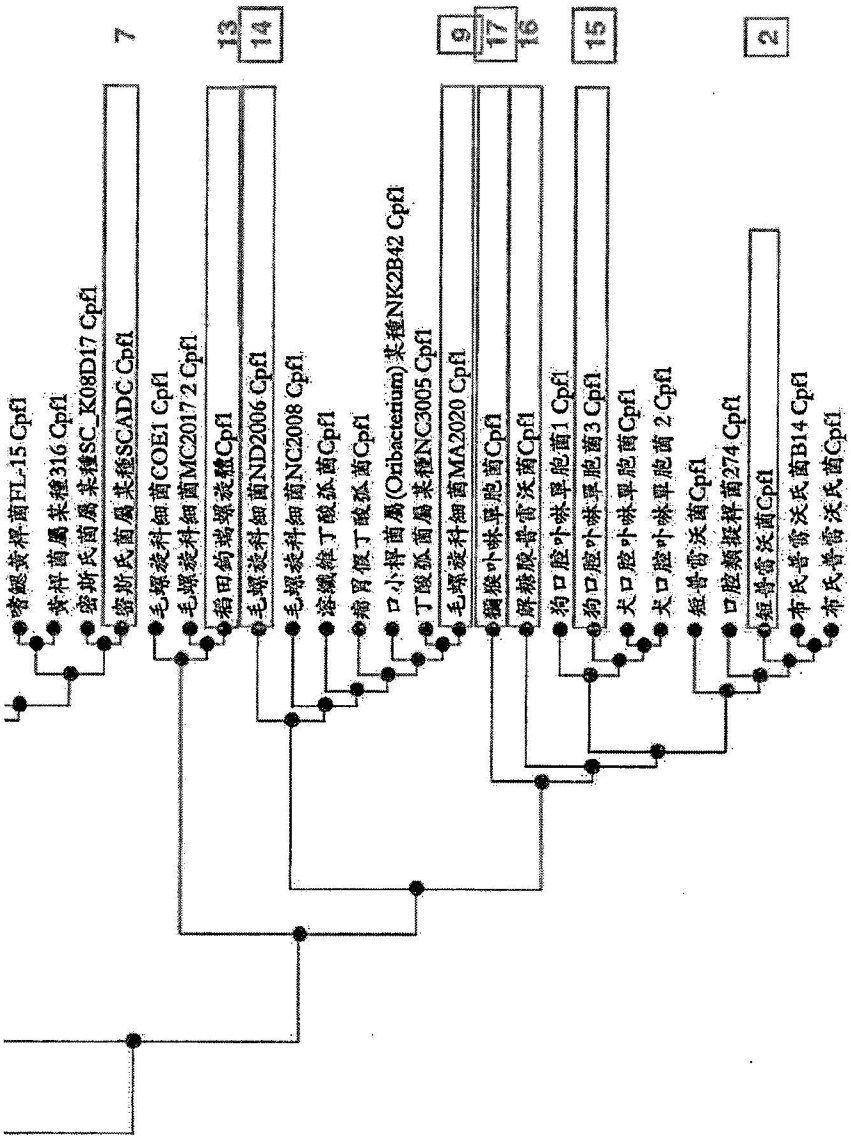

嗜鰓黃桿菌FL-15 Cpf1
黃桿菌屬菜種316 Cpf1
密斯氏菌屬菜種SC_K08D17 Cpf1
密斯氏菌屬菜種SCADC Cpf1
7
毛螺旋科細菌COE1 Cpf1
毛螺旋科細菌MC2017 2 Cpf1
稻田鉤端螺旋體Cpf1
13
毛螺旋科細菌ND2006 Cpf1
14
毛螺旋科細菌NC2008 Cpf1
溶纖維丁酸弧菌Cpf1
瘤胃假丁酸弧菌Cpf1
口小桿菌屬(Oribacterium)菜種NK2B42 Cpf1
丁酸弧菌屬菜種NC3005 Cpf1
毛螺旋科細菌MA2020 Cpf1
9
獼猴卟啉單胞菌Cpf1
17
解糖脫普雷沃菌Cpf1
16
狗口腔卟啉單胞菌1 Cpf1
狗口腔卟啉單胞菌3 Cpf1
15
犬口腔卟啉單胞菌Cpf1
犬口腔卟啉單胞菌 2 Cpf1
短普雷沃菌Cpf1
口腔類捩桿菌274 Cpf1
短普雷沃菌Cpf1
布氏普雷沃氏菌B14 Cpf1
布氏普雷沃氏菌Cpf1
2

圖 64B

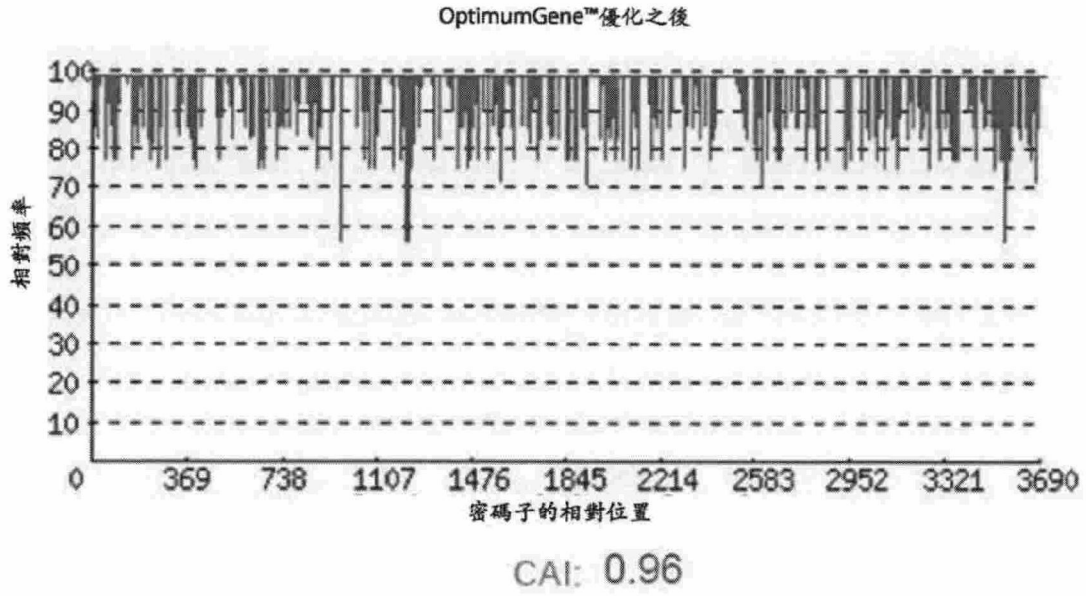

OptimumGene™優化之後
相對頻率
100
90
80
70
60
50
40
30
20
10
0
369
738
1107
1476
1845
2214
2583
2952
3321
3690
密碼子的相對位置
CAI: 0.96

圖 65A

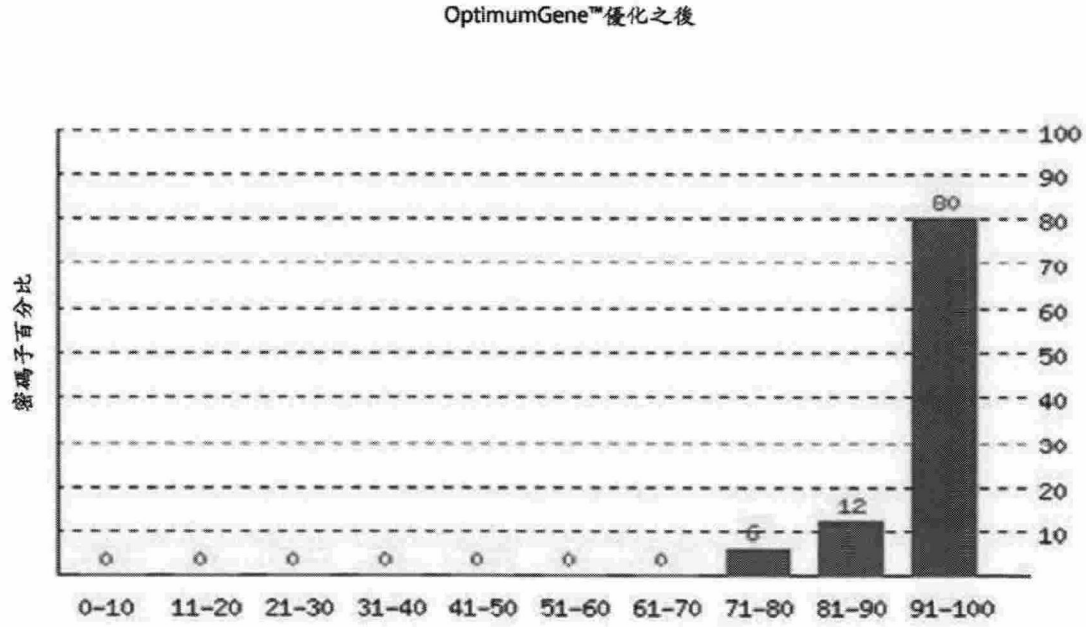

OptimumGene™優化之後
密碼子百分比
100
90
80
70
60
50
40
30
20
10
0
0
0
0
0
0
0
6
12
80
0-10
11-20
21-30
31-40
41-50
51-60
61-70
71-80
81-90
91-100

圖 65B

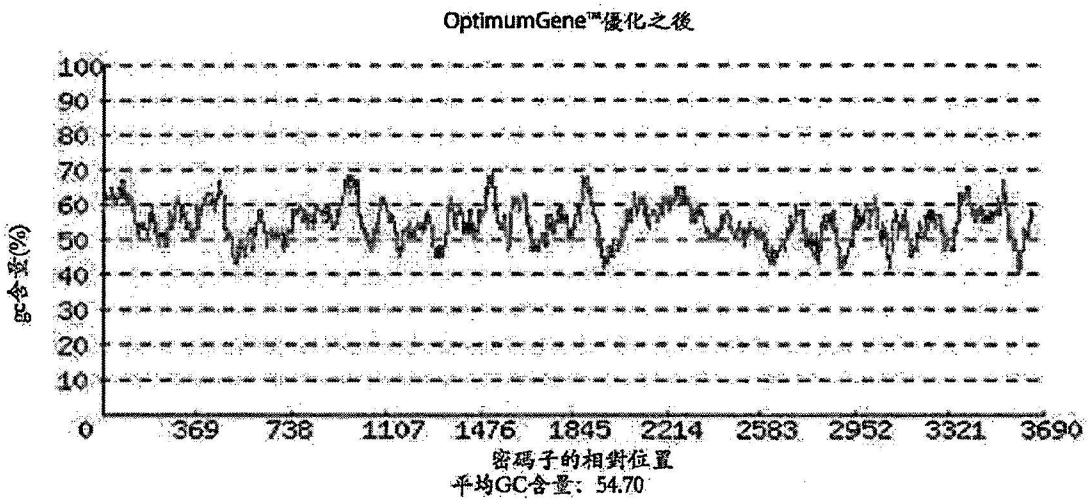

OptimumGene™優化之後
100
90
80
70
60
50
40
30
20
10
0
GC含量(%)
369
738
1107
1476
1845
2214
2583
2952
3321
3690
密碼子的相對位置
平均GC含量：54.70

圖 65C

限制性內切酶	優化的
*綠色：過濾位點；藍色：檢查位點(未過濾的)；紅色：保留位點	
BamHI(GGATCC)	1(3754)
EcoRI(GAATTC)	1(3844)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(3818)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

順式作用元件	優化的
剪接(GGTAAG)	0
剪接(GGTGAT)	0
PolyA(AATAAA)	0
PolyA(ATTAAA)	0
去穩定化(ATTTA)	0
PolyT(TTTTTT)	0
PolyA(AAAAAAA)	0

抗病毒模體	優化的
	0

圖 65D

優化之後

最大同向重複序列：大小：15 距離：2688 頻率：2

最大反向/重複序列：無

最大Dyad重複序列：大小：13 Tm：49.0 起始位置：1666,1908

圖 65E

SEQ ID NO: 2

ATGGATTACGGCAACGGCCAGTTTGAGCGGAGAGCCCCCCTGACCAAGACAATCACCCTGCGCCTGAAGCCTATC
GGCGAGACACGGGAGACAATCCGCGAGCAGAAGCTGCTGGAGCAGGACGCCGCCTTCAGAAAGCTGGTGGAGACA
GTGACCCCTATCGTGGACGATTGTATCAGGAAGATCGCCGATAACGCCCTGTGCCACTTTGGCACCGAGTATGAC
TTCAGCTGTCTGGGCAACGCCATCTCTAAGAATGACAGCAAGGCCATCAAGAAGGAGACAGAGAAGGTGGAGAAG
CTGCTGGCCAAGGTGCTGACCGAGAATCTGCCAGATGGCCTGCGCAAGGTGAACGACATCAATTCCGCCGCCTTT
ATCCAGGATACACTGACCTCTTTCGTGCAGGACGATGCCGACAAGCGGGTGCTGATCCAGGAGCTGAAGGGCAAG
ACCGTGCTGATGCAGCGGTTCCTGACCACACGGATCACAGCCCTGACCGTGTGGCTGCCCGACAGAGTGTTCGAG
AACTTTAATATCTTCATCGAGAACGCCGAGAAGATGAGAATCCTGCTGGACTCCCCTCTGAATGAGAAGATCATG
AAGTTTGACCCAGATGCCGAGCAGTACGCCTCTCTGGAGTTCTATGGCCAGTGCCTGTCTCAGAAGGACATCGAT
AGCTACAACCTGATCATCTCCGGCATCTATGCCGACGATGAGGTGAAGAACCCTGGCATCAATGAGATCGTGAAG
GAGTACAATCAGCAGATCCGGGGCGACAAGGATGAGTCCCCACTGCCCAAGCTGAAGAAGCTGCACAAGCAGATC
CTGATGCCAGTGGAGAAGGCCTTCTTTGTGCGCGTGCTGTCTAACGACAGCGATGCCCCGGAGCATCCTGGAGAAG
ATCCTGAAGGACACAGAGATGCTGCCCTCCAAGATCATCGAGGCCATGAAGGAGGCAGATGCAGGCGACATCGCC
GTGTACGGCAGCCGGCTGCACGAGCTGAGCCACGTGATCTACGGCGATCACGGCAAGCTGTCCCAGATCATCTAT
GACAAGGAGTCCAAGAGGATCTCTGAGCTGATGGAGACACTGTCTCCAAAGGAGCGCAAGGAGAGCAAGAAGCGG
CTGGAGGGCCTGGAGGAGCACATCAGAAAGTCTACATACACCTTCGACGAGCTGAACAGGTATGCCGAGAAGAAT
GTGATGGCAGCATACATCGCAGCAGTGGAGGAGTCTTGTGCCGAGATCATGAGAAAGGAGAAGGATCTGAGGACC
CTGCTGAGCAAGGAGGACGTGAAGATCCGGGGCAACAGACACAATACACTGATCGTGAAGAACTACTTTAATGCC
TGGACCGTGTTCCGGAACCTGATCAGAATCCTGAGGCGCAAGTCCGAGGCCGAGATCGACTCTGACTTCTACGAT
GTGCTGGACGATTCCGTGGAGGTGCTGTCTCTGACATACAAGGGCGAGAATCTGTGCCGCAGCTATATCACCAAG
AAGATCGGCTCCGACCTGAAGCCCGAGATCGCCACATACGGCAGCGCCCTGAGGCCTAACAGCCGCTGGTGGTCC
CCAGGAGAGAAGTTTAATGTGAAGTTCCACACCATCGTGCGGAGAGATGGCCGGCTGTACTATTTCATCCTGCCC
AAGGGCGCCAAGCCTGTGGAGCTGGAGGACATGGATGGCGACATCGAGTGTCTGCAGATGAGAAAGATCCCTAAC
CCAACAATCTTTCTGCCCAAGCTGGTGTTCAAGGACCCTGAGGCCTTCTTTAGGGATAATCCAGAGGCCGACGAG
TTCGTGTTTCTGAGCGGCATGAAGGCCCCCGTGACAATCACCAGAGAGACATACGAGGCCTACAGGTATAAGCTG
TATACCGTGGGCAAGCTGCGCGATGGCGAGGTGTCCGAAGAGGAGTACAAGCGGGCCCTGCTGCAGGTGCTGACC
GCCTACAAGGAGTTTCTGGAGAACAGAATGATCTATGCCGACCTGAATTTCGGCTTTAAGGATCTGGAGGAGTAT
AAGGACAGCTCCGAGTTTATCAAGCAGGTGGAGACACACAACACCTTCATGTGCTGGGCCAAGGTGTCTAGCTCC
CAGCTGGACGATCTGGTGAAGTCTGGCAACGGCCTGCTGTTCGAGATCTGGAGCGAGCGCCTGGAGTCCTACTAT
AAGTACGGCAATGAGAAGGTGCTGCGGGGCTATGAGGGCGTGCTGCTGAGCATCCTGAAGGATGAGAACCTGGTG
TCCATGCGGACCCTGCTGAACAGCCGGCCCATGCTGGTGTACCGGCCAAAGGAGTCTAGCAAGCCTATGGTGGTG
CACCGGGATGGCAGCAGAGTGGTGGACAGGTTTGATAAGGACGGCAAGTACATCCCCCCTGAGGTGCACGACGAG
CTGTATCGCTTCTTTAACAATCTGCTGATCAAGGAGAAGCTGGGCGAGAAGGCCCGGAAGATCCTGGACAACAAG
AAGGTGAAGGTGAAGGTGCTGGAGAGCGAGAGAGTGAAGTGGTCCAAGTTCTACGATGAGCAGTTTGCCGTGACC
TTCAGCGTGAAGAAGAACGCCGATTGTCTGGAGACCACAAAGGACCTGAATGCCGAAGTGATGGAGCAGTATAGC
GAGTCCAACAGACTGATCCTGATCAGGAATACCACAGATATCCTGTACTATCTGGTGCTGGACAAGAATGGCAAG
GTGCTGAAGCAGAGATCCCTGAACATCATCAATGACGGCGCCAGGGATGTGGACTGGAAGGAGAGGTTCCGCCAG
GTGACAAAGGATAGAAACGAGGGCTACAATGAGTGGGATTATTCCAGGACCTCTAACGACCTGAAGGAGGTGTAC

圖 65F

CTGAATTATGCCCTGAAGGAGATCGCCGAGGCCGTGATCGAGTACAACGCCATCCTGATCATCGAGAAGATGTCT
AATGCCTTTAAGGACAAGTATAGCTTCCTGGACGACGTGACCTTCAAGGGCTTCGAGACAAAGCTGCTGGCCAAG
CTGAGCGATCTGCACTTTAGGGGCATCAAGGACGGCGAGCCATGTTCCTTCACAAACCCCCTGCAGCTGTGCCAG
AACGATTCTAATAAGATCCTGCAGGACGGCGTGATCTTTATGGTGCCAAATTCTATGACACGGAGCCTGGACCCC
GACACCGGCTTCATCTTTGCCATCAACGACCACAATATCAGGACCAAGAAGGCCAAGCTGAACTTTCTGAGCAAG
TTCGATCAGCTGAAGGTGTCCTCTGAGGGCTGCCTGATCATGAAGTACAGCGGCGATTCCCTGCCTACACACAAC
ACCGACAATCGCGTGTGGAACTGCTGTTGCAATCACCCAATCACAAACTATGACCGGGAGACAAAGAAGGTGGAG
TTCATCGAGGAGCCCGTGGAGGAGCTGTCCCGCGTGCTGGAGGAGAATGGCATCGAGACAGACACCGAGCTGAAC
AAGCTGAATGAGCGGGAGAACGTGCCTGGCAAGGTGGTGGATGCCATCTACTCTCTGGTGCTGAATTATCTGCGC
GGCACAGTGAGCCGAGTGCAGGACAGAGGGCCGTGTACTATAGCCCTGTGACCGGCAAGAAGTACGATATCTCC
TTTATCCAGGCCATGAACCTGAATAGGAAGTGTGACTACTATAGGATCGGCTCCAAGGAGAGGGGAGAGTGGACC
GATTTCGTGGCCCAGCTGATCAAC
[illegible]
[illegible]

圖 65G

MDYGNGQFERRAPLTKTITLRLKPIGETRETIREQKLLEQDAAFRKLVETVTPIVDDCIR
KIADNALCHFGTEYDFSCLGNAISKNDSKAIKKETEKVEKLLAKVLTENLPDGLRKVNDI
NSAAFIQDTLTSFVQDDADKRVLIQELKGKTVLMQRFLTTRITALTVWLPDRVFENFNIF
IENAEKMRILLDSPLNEKIMKFDPDAEQYASLEFYGQCLSQKDIDSYNLIISGIYADDEV
KNPGINEIVKEYNQQIRGDKDESPLPKLKKLHKQILMPVEKAFFVRVLSNDSDARSILEK
ILKDTEMLPSKIIEAMKEADAGDIAVYGSRLHELSHVIYGDHGKLSQIIYDKESKRISEL
METLSPKERKESKKRLEGLEEHIRKSTYTFDELNRYAEKNVMAAYIAAVEESCAEIMRKE
KDLRTLLSKEDVKIRGNRHNTLIVKNYFNAWTVFRNLIRILRRKSEAEIDSDFYDVLDDS
VEVLSLTYKGENLCRSYITKKIGSDLKPEIATYGSALRPNSRWWSPGEKFNVKFHTIVRR
DGRLYYFILPKGAKPVELEDMDGDIECLQMRKIPNPTIFLPKLVFKDPEAFFRDNPEADE
FVFLSGMKAPVTITRETYEAYRYKLYTVGKLRDGEVSEEEYKRALLQVLTAYKEFLENRM
IYADLNFGFKDLEEYKDSSEFIKQVETHNTFMCWAKVSSSQLDDLVKSGNGLLFEIWSER
LESYYKYGNEKVLRGYEGVLLSILKDENLVSMRTLLNSRPMLVYRPKESSKPMVVHRDGS
RVVDRFDKDGKYIPPEVHDELYRFFNNLLIKEKLGEKARKILDNKKVKVKVLESERVKWS
KFYDEQFAVTFSVKKNADCLDTTKDLNAEVMEQYSESNRLILIRNTTDILYYLVLDKNGK
VLKQRSLNIINDGARDVDWKERFRQVTKDRNEGYNEWDYSRTSNDLKEVYLNYALKEIAE
AVIEYNAILIIEKMSNAFKDKYSFLDDVTFKGFETKLLAKLSDLHFRGIKDGEPCSFTNP
LQLCQNDSNKILQDGVIFMVPNSMTRSLDPDTGFIFAINDHNIRTKKAKLNFLSKFDQLK
VSSEGCLIMKYSGDSLPTHNTDNRVWNCCCNHPITNYDRETKKVEFIEEPVEELSRVLEE
NGIETDTELNKLNERENVPGKVVDAIYSLVLNYLRGTVSGVAGQRAVYYSPVTGKKYDIS
FIQAMNLNRKCDYYRIGSKERGEWIDFVAQLIN

圖 65H

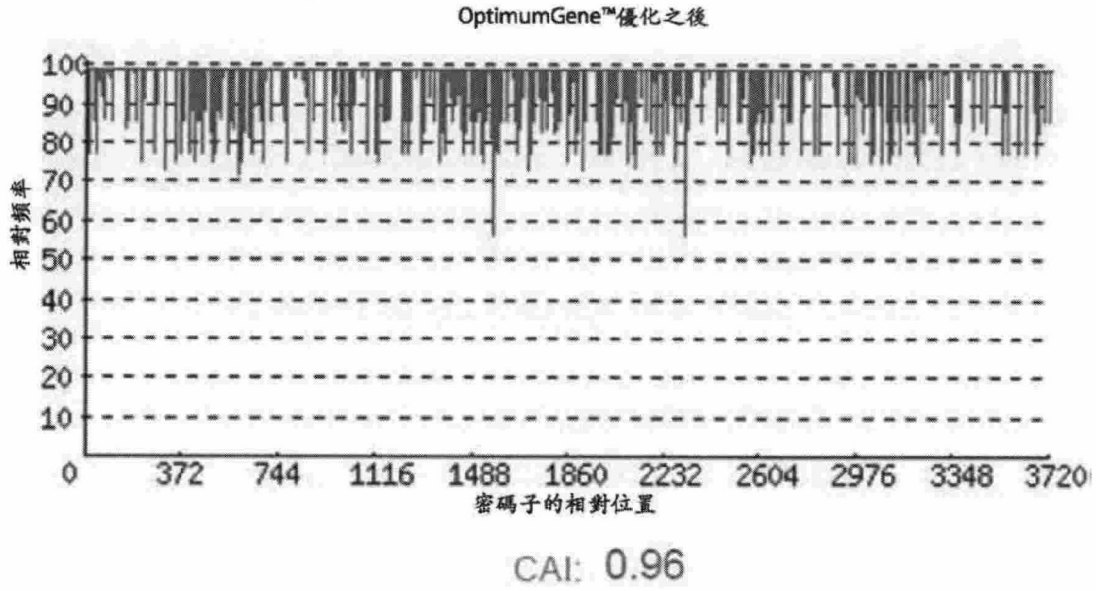

OptimumGene™優化之後
相對頻率
100
90
80
70
60
50
40
30
20
10
0
372
744
1116
1488
1860
2232
2604
2976
3348
3720
密碼子的相對位置
CAI: 0.96

圖 66A

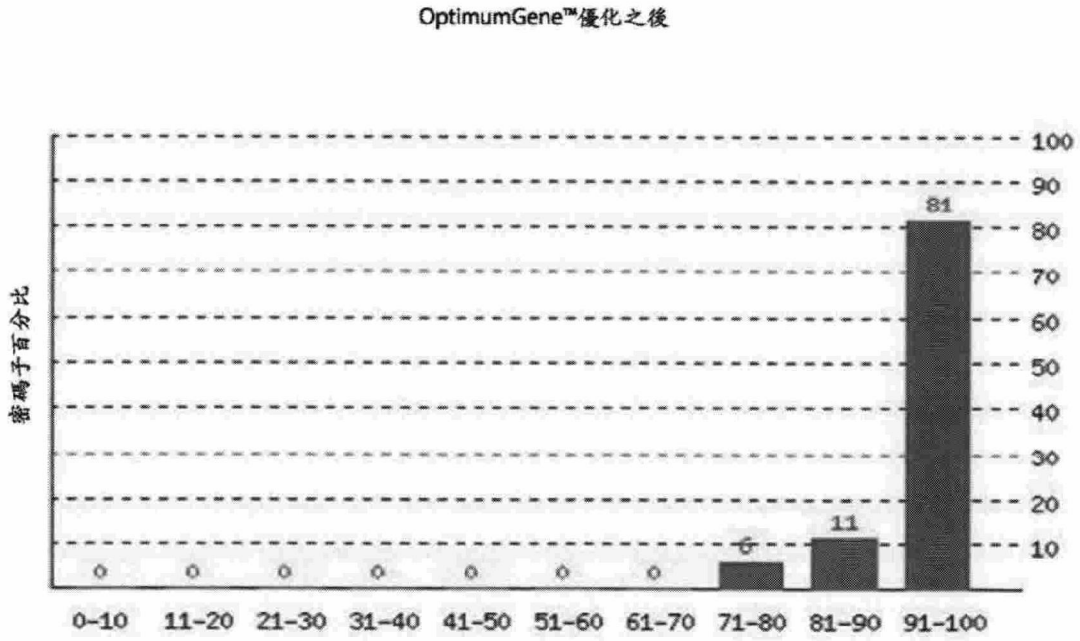

OptimumGene™優化之後
密碼子百分比
100
90
80
70
60
50
40
30
20
10
0
0
0
0
0
0
0
6
11
81
0-10
11-20
21-30
31-40
41-50
51-60
61-70
71-80
81-90
91-100

圖 66B

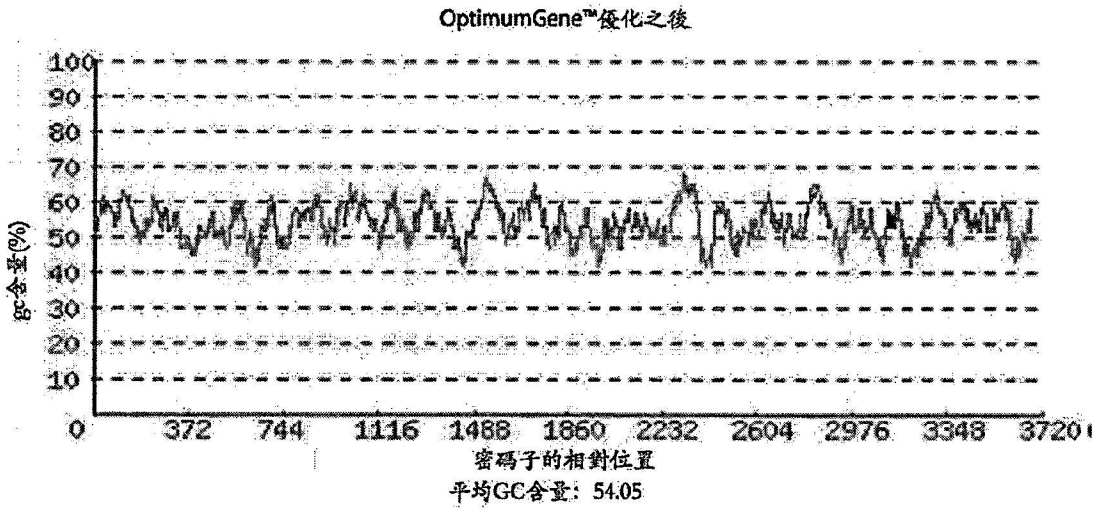

OptimumGene™優化之後
100
90
80
70
60
50
40
30
20
10
0
372
744
1116
1488
1860
2232
2604
2976
3348
3720
GC含量(%)
密碼子的相對位置
平均GC含量：54.05

圖 66C

限制性內切酶	優化的
*綠色：過濾位點；藍色：檢查位點(未過濾的)；紅色：保留位點	
BamHI(GGATCC)	1(3778)
EcoRI(GAATTC)	1(3868)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(3842)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

順式作用元件	優化的
剪接(GGTAAG)	0
剪接(GGTGAT)	0
PolyA(AATAAA)	0
PolyA(ATTAAA)	0
去穩定化(ATTTA)	0
PolyT(TTTTTT)	0
PolyA(AAAAAAA)	0

抗病毒模體	優化的
	0

圖 66D

優化之後

最大同向重複序列：大小：12 距離：3285 頻率：2

最大反向重複序列：大小：12 Tm：43.5 起始位置：1631,1547

最大Dyad重複序列：無

圖 66E

GATGCC

ATGCTGCTGTATGAGAACTACACCAAGCGGAATCAGATCACAAAGAGCCTGAGGCTGGAGCTGCCCCCTCAGGGC
AAGACCCTGAGAAACATCAAGGAGCTGAATCTGCTGGAGCAGGACAAGGCCATCTACGCCCTGCTGGAGCGGCTG
AACCCAGTGATCGATGAGGGCATCAAGGACATCGCCAGAGATACCCTGAAGAACTGGGAGCTGTCTTTTGAGAAG
CTGTACGAGCACTTCCTGAGCGGCGACAAGAAGGCCTATGCCAAGGAGTCCGAGCGGCTGAAGAAGGAGATCGTG
AAAACCCTGATCAAGAACCTGCCCGAGGGCATCGGCAAGATCTCTGAGATCAACAGCGCCAAGTATCTGAATGGC
GTGCTGTACGACTTCATCGATAAGACACACAAGGACTCTGAGGAGAAGCAGAATATCCTGAGCGATATCCTGGAG
ACAAAGGGCTACCTGGCCCTGTTCTCCAAGTTTCTGACATCTCGGATCACCACACTGGAGCAGTCCATGCCTAAG
AGAGTGATCGAGAACTTCGAGATCTATGCCGCCAATATCCCAAAGATGCAGGACGCCCTGGAGAGGGGAGCCGTG
TCTTTTGCCATCGAGTACGAGAGCATCTGTTCCGTGGATTACTATAACCAGATCCTGTCCCAGGAGGACATCGAT
TCTTATAATCGCCTGATCAGCGGCATCATGGACGAGGATGGCGCCAAGGAGAAGGGCATCAACCAGACCATCTCC
GAGAAGAATATCAAGATCAAGTCTGAGCACCTGGAGGAGAAGCCCTTCAGAATCCTGAAGCAGCTGCACAAGCAG
ATCCTGGAGGAGCGCGAGAAGGCCTTTACAATCGACCACATCGACAGCGATGAGGAGGTGGTGCAGGTGACCAAG
GAGGCCTTCGAGCAGACAAAGGAGCAGTGGGAGAACATCAAGAAGATCAATGGCTTCTACGCCAAGGACCCCGGC
GATATCACCCTGTTTATCGTGGTGGGCCCTAACCAGACACACGTGCTGTCCCAGCTGATCTACGGCGAGCACGAC
CGGATCAGACTGCTGCTGGAGGAGTATGAGAAGAACACCCTGGAGGTGCTGCCCCGGAGAACAAAGTCTGAGAAG
GCCAGATACGATAAGTTCGTGAATGCCGTGCCTAAGAAGGTGGCCAAGGAGAGCCACACCTTCGACGGCCTGCAG
AAGATGACAGGCGACGATCGGCTGTTTATCCTGTATAGAGATGAGCTGGCCCGGAACTACATGAGAATCAAGGAG
GCCTATGGCACCTTTGAGCGGGACATCCTGAAGAGCAGGCGCGGCATCAAGGGCAATCGGGACGTGCAGGAGTCT
CTGGTGAGCTTTTACGATGAGCTGACAAAGTTCAGGAGCGCCCTGCGCATCATCAATTCCGGCAACGACGAGAAG
GCCGATCCAATCTTCTATAACACCTTTGATGGCATCTTCGAGAAGGCCAATAGGACATACAAGGCCGAGAATCTG
TGCCGCAACTATGTGACCAAGTCTCCAGCCGACGATGCCAGGATCATGGCCAGCTGTCTGGGCACCCCAGCAAGG
CTGCGCACACACTGGTGGAATGGCGAGGAGAACTTCGCCATCAATGACGTGGCCATGATCCGGAGAGGCGATGAG
TACTATTACTTTGTGCTGACCCCAGACGTGAAGCCCGTGGACCTGAAAACCAAGGACGAGACAGATGCCCAGATC
TTCGTGCAGCGCAAGGGCGCCAAGTCCTTTCTGGGCCTGCCAAAGGCCCTGTTCAAGTGCATCCTGGAGCCTTAC
TTTGAGTCCCCAGAGCACAAGAATGACAAGAACTGCGTGATCGAGGAGTACGTGTCTAAGCCCCTGACCATCGAC
AGGCGCGCCTATGATATCTTTAAGAACGGCACCTTCAAGAAAACCAATATCGGCATCGACGGCCTGACCGAGGAG
AAGTTCAAGGACGATTGCCGCTATCTGATCGACGTGTATAACGAGTTCATCGCCCTGTATACAAGGTACAGCTGT
TTTAACATGTCCGGCCTGAAGCGCGCCGACGAGTACAATGATATCGGCGAGTTCTTTTCTGACGTGGATACCAGG
CTGTGCACAATGGAGTGGATTCCCGTGAGCTTCGAGCGCATCAACGACATGGTGGATAAGAAGGAGGGCCTGCTG
TTTCTGGTGAGGAGCATGTTCCTGTATAACCGGCCCAGAAAGCCTTATGAGCGCACCTTTATCCAGCTGTTCAGC
GACTCCAACATGGAGCACACATCTATGCTGCTGAATAGCAGGGCCATGATCCAGTACAGGGCAGCCTCCCTGCCA
CGGAGAGTGACCCACAAGAAGGGCAGCATCCTGGTGGCCCTGCGGGATTCCAACGGCGAGCACATCCCCATGCAC
ATCAGAGAGGCCATCTACAAGATGAAGAACAATTTTGACATCAGCTCCGAGGATTTCATCATGGCCAAGGCCTAT
CTGGCCGAGCACGACGTGGCCATCAAGAAGGCCAACGAGGATATCATCAGGAATAGGCGCTACACCGAGGACAAG
TTCTTTCTGTCTCTGAGCTATACCAAGAACGCCGATATCAGCGCCCGCACACTCGACTACATCAATGATAAGGTG
GACGAGGACACACAGGATTCCAGGATGGCCGTGATCGTGACCCGCAACCTGAAGGACCTGACATACGTGGCCGTG
GTGGATGAGAAGAACAATGTGCTGGAGGAGAAGAGCCTGAACGAGATCGACGGCGTGAATTATCGGGAGCTGCTG
AAGGAGAGAACCAAGATCAAGTACCACGACAAGACACGGCTGTGGCAGTATGACGTGAGCAGCAAGGGCCTGAAG

圖 66F

GAGGCCTACGTGGAGCTGGCCGTGACCCAGATCTCCAAGCTGGCCACAAAGTATAACGCCGTGGTGGTGGTGGAG
TCCATGTCCTCTACCTTCAAGGACAAGTTCTCTTTTCTGGATGAGCAGATCTTCAAGGCCTTTGAGGCCCGGCTG
TGCGCCAGAATGTCCGACCTGTCTTTTAATACAATCAAGGAGGGCGAGGCCGGCTCCATCTCTAACCCCATCCAG
GTGTCCAACAATAACGGCAATTCTTATCAGGACGGCGTGATCTACTTCCTGAATAACGCCTATACCCGGACCCTG
TGCCCTGATACCGGCTTTGTGGACGTGTTCGATAAGACCCGGCTGATCACAATGCAGTCTAAGAGACAGTTCTTT
GCCAAGATGAAGGACATCAGAATCGACGATGGCGAGATGCTGTTCACCTTTAACCTGGAGGAGTACCCTACAAAG
AGGCTGCTGGACCGCAAGGAGTGGACCGTGAAGATCGCCGGCGATGGCTCCTATTTCGACAAGGATAAGGGCGAG
TACGTGTACGTGAACGACATCGTGAGAGAGCAGATCATCCCAGCCCTGCTGGAGGACAAGGCCGTGTTCGATGGC
AATATGGCCGAGAAGTTTCTGGATAAGACCGCCATCAGCGGCAAGTCCGTGGAGCTGATCTACAAGTGGTTCGCC
AACGCCCTGTATGGCATCATCACAAAGAAGGACGGCGAGAAGATCTACCGGAGCCCCATCACCGGCACAGAGATC
GACGTGAGCAAGAACACCACATACAACTTCGGCAAGAAGTTCATGTTCAAGCAGGAGTATAGAGGCGACGGCGAT
TTTCTGGACGCCTTCCTGAATTACATGCAGGCCCAGGATATCGCCGTG
[illegible]
[illegible]

圖 66G

MLLYENYTKRNQITKSLRLELRPQGKTLRNIKELNLLEQDKAIYALLERLKPVIDEGIKD
IARDTLKNCELSFEKLYEHFLSGDKKAYAKESERLKKEIVKTLIKNLPEGIGKISEINSA
KYLNGVLYDFIDKTHKDSEEKQNILSDILETKGYLALFSKFLTSRITTLEQSMPKRVIEN
FEIYAANIPKMQDALERGAVSFAIEYESICSVDYYNQILSQEDIDSYNRLISGIMDEDGA
KEKGINQTISEKNIKIKSEHLEEKPFRILKQLHKQILEEREKAFTIDHIDSDEEVVQVTK
EAFEQTKEQWENIKKINGFYAKDPGDITLFIVVGPNQTHVLSQLIYGEHDRIRLLLEEYE
KNTLEVLPRRTKSEKARYDKFVNAVPKKVAKESHTFDGLQKMTGDDRLFILYRDELARNY
MRIKEAYGTFERDILKSRRGIKGNRDVQESLVSFYDELTKFRSALRIINSGNDEKADPIF
YNTFDGIFEKANRTYKAENLCRNYVTKSPADDARIMASCLGTPARLRTHWWNGEENFAIN
DVAMIRRGDEYYYFVLTPDVKPVDLKTKDETDAQIFVQRKGAKSFLGLPKALFKCILEPY
FESPEHKNDKNCVIEEYVSKPLTIDRRAYDIFKNGTFKKTNIGIDGLTEEKFKDDCRYLI
DVYKEFIAVYTRYSCFNMSGLKRADEYNDIGEFFSDVDTRLCTMEWIPVSFERINDMVDK
KEGLLFLVRSMFLYNRPRKPYERTFIQLPSDSNMEHTSMLLNSRAMIQYRAASLPRRVTH
KKGSILVALRDSNGEHIPMHIREAIYKMKNNFDISSEDFIMAKAYLAEHDVAIKKANEDI
IRNRRYTEDKFFLSLSYTKNADISARTLDYINDKVEEDTQDSRMAVIVTRNLKDLTYVAV
VDEKNNVLEEKSLNEIDGVNYRELLKERTKIKYHDKTRLWQYDVSSKGLKEAYVELAVTQ
ISKLATKYNAVVVVESMSSTFKDKFSFLDEQIFKAFEARLCARMSDLSFNTIKEGEAGSI
SNPIQVSNNNGNSYQDGVIYFLNNAYTRTLCPDTGFVDVFDKTRLITMQSKRQFFAKMKD
IRIDDGEMLFTFNLEEYPTKRLLDRKEWTVKIAGDGSYFDKDKGEYVYVNDIVREQIIPA
LLEDKAVFDGNMAEKFLDKTAISGKSVELIYKWFANALYGIITKKDGEKIYRSPITGTEI
DVSKNTTYNFGKKFMFKQEYRGDGDFLDAFLNYMQAQDIAV

圖 66H

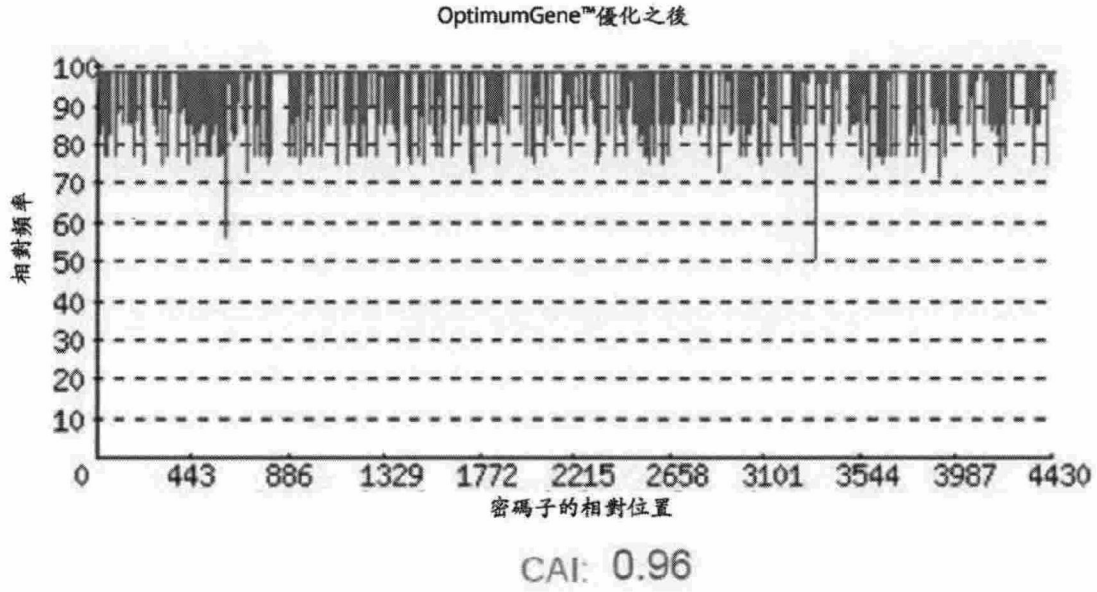

OptimumGene™優化之後
相對頻率
100
90
80
70
60
50
40
30
20
10
0
443
886
1329
1772
2215
2658
3101
3544
3987
4430
密碼子的相對位置
CAI: 0.96

圖 67A

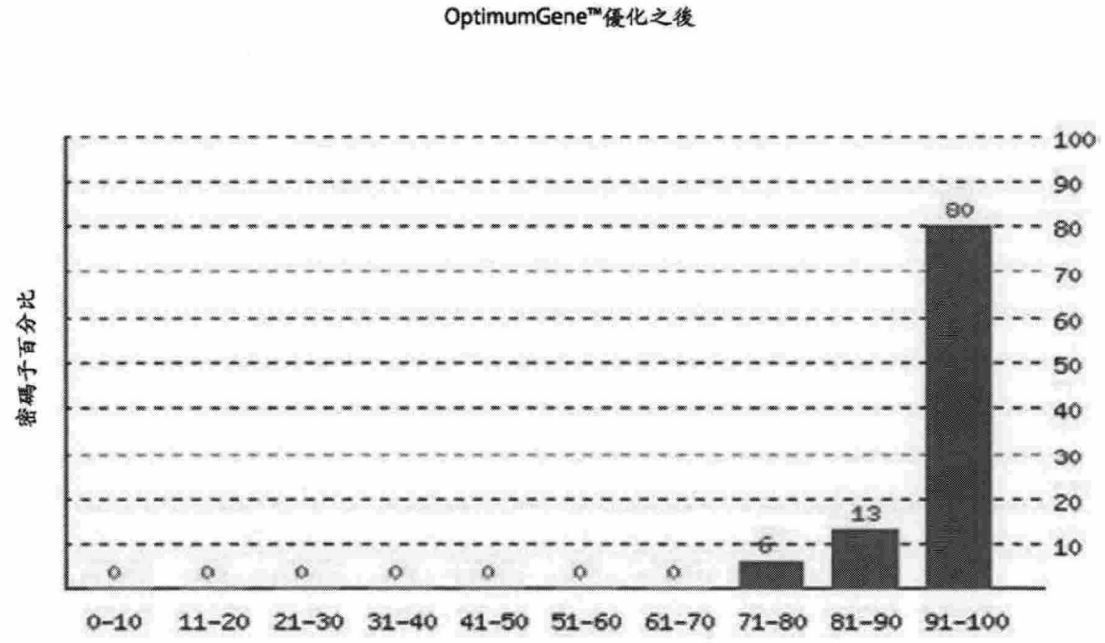

OptimumGene™優化之後
密碼子百分比
100
90
80
70
60
50
40
30
20
10
0
0
0
0
0
0
0
6
13
80
0-10
11-20
21-30
31-40
41-50
51-60
61-70
71-80
81-90
91-100

圖 67B

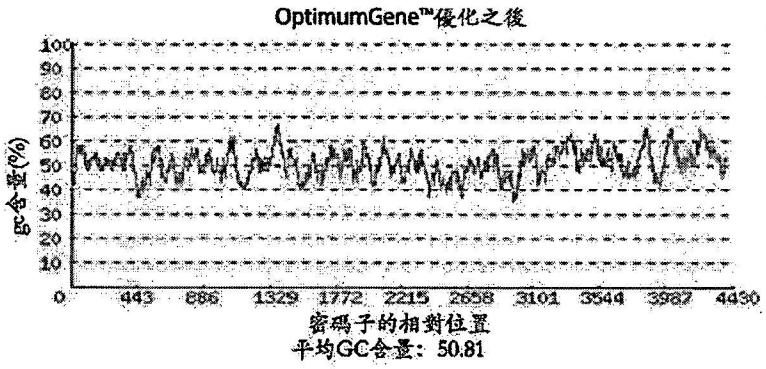

OptimumGene™優化之後
GC含量(%)
100
90
80
70
60
50
40
30
20
10
0
443
886
1329
1772
2215
2658
3101
3544
3987
4430
密碼子的相對位置
平均GC含量：50.81

圖 67C

限制性內切酶	優化的
綠色：過濾位點；藍色：檢查位點(未過濾的)；紅色：保留位點	
BamHI(GGATCC)	1(4486)
EcoRI(GAATTC)	1(4576)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(4550)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

順式作用元件	優化的
剪接(GGTAAG)	0
剪接(GGTGAT)	0
PolyA(AATAAA)	0
PolyA(ATTAAA)	0
去穩定化(ATTTA)	0
PolyT(TTTTTT)	0
PolyA(AAAAAAA)	0

抗病毒模體	優化的
	0

圖 67D

優化之後

最大同向重複序列：大小：15 距離：3543 頻率：2

最大反向重複序列：無

最大Dyad重複序列：無

圖 67E

GCTAGC

ATGTCCAACTTCTTTAAGAATTTCACCAACCTGTATGAGCTGTCCAAGACACTGAGGTTTGAGCTGAAGCCCGTG
GGCGACACCCTGACAAACATGAAGGACCACCTGGAGTACGATGAGAAGCTGCAGACCTTCCTGAAGGATCAGAAT
ATCGACGATGCCTATCAGGCCCTGAAGCCTCAGTTCGACGAGATCCACGAGGAGTTTATCACAGATTCTCTGGAG
AGCAAGAAGGCCAAGGAGATCGACTTCTCCGAGTACCTGGATCTGTTTCAGGAGAAGAAGGAGCTGAACGACTCT
GAGAAGAAGCTGCGCAACAAGATCGGCGAGACATTCAACAAGGCCGGCGAGAAGTGGAAGAAGGAGAAGTACCCT
CAGTATGAGTGGAAGAAGGGCTCCAAGATCGCCAATGGCGCCGACATCCTGTCTTGCCAGGATATGCTGCAGTTT
ATCAAGTATAAGAACCCAGAGGATGAGAAGATCAAGAATTACATCGACGATACACTGAAGGGCTTCTTTACCTAT
TTCGGCGGCTTTAATCAGAACAGGGCCAACTACTATGAGACAAAGAAGGAGGCCTCCACCGCAGTGGCAACAAGG
ATCGTGCACGAGAACCTGCCAAAGTTCTGTGACAATGTGATCCAGTTTAAGCACATCATCAAGCGGAAGAAGGAT
GGCACCGTGGAGAAAACCGAGAGAAAGACCGAGTACCTGAACGCCTACCAGTATCTGAAGAACAATAACAAGATC
ACACAGATCAAGGACGCCGAGACAGAGAAGATGATCGAGTCTACACCCATCGCCGAGAAGATCTTCGACGTGTAC
TACTTCAGCAGCTGCCTGAGCCAGAAGCAGATCGAGGAGTACAACCGGATCATCGGCCACTATAATCTGCTGATC
AACCTGTATAACCAGGCCAAGAGATCTGAGGGCAAGCACCTGAGCGCCAACGAGAAGAAGTATAAGGACCTGCCT
AAGTTCAAGACCCTGTATAAGCAGATCGGCTGCGGCAAGAAGAAGGACCTGTTTTACACAATCAAGTGTGATACC
GAGGAGGAGGCCAATAAGTCCCGGAACGAGGGCAAGGAGTCCCACTCTGTGGAGGAGATCATCAACAAGGCCCAG
GAGGCCATCAATAAGTACTTCAAGTCTAATAACGACTGTGAGAATATCAACACCGTGCCCGACTTCATCAACTAT
ATCCTGACAAAGGAGAATTACGAGGGCGTGTATTGGAGCAAGGCCGCCATGAACACCATCTCCGACAAGTACTTC
GCCAATTATCACGACCTGCAGGATAGACTGAAGGAGGCCAAGGTGTTTCAGAAGGCCGATAAGAAGTCCGAGGAC
GATATCAAGATCCCAGAGGCCATCGAGCTGTCTGGCCTGTTCGGCGTGCTGGACAGCCTGGCCGATTGGCAGACC
ACACTGTTTAAGTCTAGCATCCTGAGCAACGAGGACAAGCTGAAGATCATCACAGATTCCCAGACCCCTCTGAG
GCCCTGCTGAAGATGATCTTCAATGACATCGAGAAGAACATGGAGTCCTTTCTGAAGGAGACAAACGATATCATC
ACCCTGAAGAAGTATAAGGGCAATAAGGAGGGCACCGAGAAGATCAAGCAGTGGTTCGACTATACACTGGCCATC
AACCGGATGCTGAAGTACTTTCTGGTGAAGGAGAATAAGATCAAGGGCAACTCCCTGGATACCAATATCTCTGAG
GCCCTGAAAACCCTGATCTACAGCGACGATGCCGAGTGGTTCAAGTGGTACGACGCCCTGAGAAACTATCTGACC
CAGAAGCCTCAGGATGAGGCCAAGGAGAATAAGCTGAAGCTGAATTTCGACAACCCATCTCTGGCCGGCGGCTGG
GATGTGAACAAGGAGTGCAGCAATTTTTGCGTGATCCTGAAGGACAAGAACGAGAAGAAGTACCTGGCCATCATG
AAGAAGGGCGAGAATACCCTGTTCCAGAAGGAGTGGACAGAGGGCCGGGGCAAGAACCTGACAAAGAAGTCTAAT
CCACTGTTCGAGATCAATAACTGCGAGATCCTGAGCAAGATGGAGTATGACTTTTGGGCCGACGTGAGCAAGATG
ATCCCCAAGTGTAGCACCCAGCTGAAGGCCGTGGTGAACCACTTCAAGCAGTCCGACAATGAGTTCATCTTTCCT
ATCGGCTACAAGGTGACAAGCGGCGAGAAGTTTAGGGAGGAGTGCAAGATCTCCAAGCAGGACTTCGAGCTGAAT
AACAAGGTGTTTAATAAGAACGAGCTGAGCGTGACCGCCATGCGCTACGATCTGTCCTCTACACAGGAGAAGCAG
TATATCAAGGCCTTCCAGAAGGAGTACTGGGAGCTGCTGTTTAAGCAGGAGAAGCGGGACACCAAGCTGACAAAT
AACGAGATCTTCAACGAGTGGATCAATTTTTGCAACAAGAAGTATAGCGAGCTGCTGTCCTGGGAGAGAAAGTAC
AAGGATGCCCTGACCAATTGGATCAACTTCTGTAAGTACTTTCTGAGCAAGTATCCCAAGACCACACTGTTCAAC
TACTCTTTTAAGGAGAGCGAGAATTATAACTCCCTGGACGAGTTCTACCGGGACGTGGATATCTGTTCTTACAAG
CTGAATATCAACACCACAATCAATAAGAGCATCCTGGATAGACTGGTGGAGGAGGGCAAGCTGTACCTGTTTGAG
ATCAAGAATCAGGACAGCAACGATGGCAAGTCCATCGGCCACAAGAATAACCTGCACACCATCTACTGGAACGCC
ATCTTCGAGAATTTTGACAACAGGCCTAAGCTGAATGGCGAGGCCGAGATCTTCTATCGCAAGGCCATCTCCAAG

圖 67F

GATAAGCTGGGCATCGTGAAGGGCAAGAAAACCAAGAACGGCACCGAGATCATCAAGAATTACAGATTCAGCAAG
GAGAAGTTTATCCTGCACGTGCCAATCACCCTGAACTTCTGCTCCAATAACGAGTATGTGAATGACATCGTGAAC
ACAAAGTTCTACAATTTTTCCAACCTGCACTTTCTGGGCATCGATAGGGGCGAGAAGCACCTGGCCTACTATTCT
CTGGTGAATAAGAACGGCGAGATCGTGGACCAGGGCACACTGAACCTGCCTTTCACCGACAAGGATGGCAATCAG
CGCAGCATCAAGAAGGAGAAGTACTTTTATAACAAGCAGGAGGACAAGTGGGAGGCCAAGGAGGTGGATTGTTGG
AATTATAACGACCTGCTGGATGCCATGGCCTCTAACCGGGACATGGCCAGAAAGAATTGGCAGAGGATCGGCACC
ATCAAGGAGGCCAAGAACGGCTACGTGAGCCTGGTCATCAGGAAGATCGCCGATCTGGCCGTGAATAACGAGCGC
CCCGCCTTCATCGTGCTGGAGGACCTGAATACAGGCTTTAAGCGGTCCAGACAGAAGATCGATAAGAGCGTGTAC
CAGAAGTTCGAGCTGGCCCTGGCCAAGAAGCTGAACTTTCTGGTGGACAAGAATGCCAAGCGCGATGAGATCGGC
TCCCCTACAAAGGCCCTGCAGCTGACCCCCCCTGTGAATAACTACGGCGACATTGAGAACAAGAAGCAGGCCGGC
ATCATGCTGTATACCCGGGCCAATTATACCTCTCAGACAGATCCAGCCACAGGCTGGAGAAAGACCATCTATCTG
AAGGCCGGCCCCGAGGAGACAACATACAAGAAGGACGGCAAGATCAAGAACAAGAGCGTGAAGGACCAGATCATC
GAGACATTCACCGATATCGGCTTTGACGGCAAGGATTACTATTCGAGTACGACAAGGGCGAGTTTGTGGATGAG
AAAACCGGCGAGATCAAGCCCAAGAAGTGGCGGCTGTACTCCGGCGAGAATGGCAAGTCCCTGGACAGGTTCCGC
CGAGAGAGGGAGAAGGATAAGTATGAGTGGAAGATCGACAAGATCGATATCGTGAAGATCCTGGACGATCTGTTC
GTGAATTTTGACAAGAACATCAGCCTGCTGAAGCAGCTGAAGGAGGGCGTGGAGCTGACCCGGAATAACGAGCAC
GGCACAGGCGAGTCCCTGAGATTCGCCATCAACCTGATCCAGCAGATCCGGAATACCGGCAATAACGAGAGAGAC
AACGATTTCATCCTGTCCCCAGTGAGGGACGAGAATGGCAAGCACTTTGACTCTCGCGAGTACTGGGATAAGGAG
ACAAAGGGCGAGAAGATCAGCATGCCCAGCTCCGGCGATGCCAATGGCGCCTTCAACATCGCCCGGAAGGGCATC
ATCATGAACGCCCACATCCTGGCCAATAGCGACTCCAAGGATCTGTCCCTGTTCGTGTCTGACGAGGAGTGGGAT
CTGCACCTGAATAACAAGACCGAGTGGAAGAAGCAGCTGAACATCTTTTCTAGCAGGAAGGCCATGGCCAAGCGC
AAGAAG
AAAAGGCCGGCGGCCACCAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCCTACCCATACGATGTTCCAGAT
TACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCCTAAGAATTC

圖 67G

MSNFFKNFTNLYELSKTLRFELKPVGDTLTNMKDHLEYDEKLQTFLKDQNIDDAYQALKP
QFDEIHEEFITDSLESKKAKEIDFSEYLDLFQEKKELNDSEKKLRNKIGETFNKAGEKWK
KEKYPQYEWKKGSKIANGADILSCQDMLQFIKYKNPEDEKIKNYIDDTLKGFFTYFGGFN
QNRANYYETKKEASTAVATRIVHENLPKFCDNVIQFKHIIKRKKDGTVEKTERKTEYLNA
YQYLKNNNKITQIKDAETEKMIESTPIAEKIFDVYYFSSCLSQKQIEEYNRIIGHYNLLI
NLYNQAKRSEGKHLSANEKKYKDLPKFKTLYKQIGCGKKKDLFYTIKCDTEEEANKSRNE
GKESHSVEEIINKAQEAINKYFKSNNDCENINTVPDFINYILTKENYEGVYWSKAAMNTI
SDKYFANYHDLQDRLKEAKVFQKADKKSEDDIKIPEAIELSGLFGVLDSLADWQTTLFKS
SILSNEDKLKIITDSQTPSEALLKMIFNDIEKNMESFLKETNDIITLKKYKGNKEGTEKI
KQWFDYTLAINRMLKYFLVKENKIKGNSLDTNISEALKTLIYSDDAEWFKWYDALRNYLT
QKPQDEAKENKLKLNFDNPSLAGGWDVNKECSNFCVILKDKNEKKYLAIMKKGENTLFQK
EWTEGRGKNLTKKSNPLFEINNCEILSKMEYDFWADVSKMIPKCSTQLKAVVNHFKQSDN
EFIFPIGYKVTSGEKFREECKISKQDFELNNKVFNKNELSVTAMRYDLSSTQEKQYIKAF
QKEYWELLFKQEKRDTKLTNNEIFNEWINFCNKKYSELLSWERKYKDALTNWINFCKYFL
SKYPKTTLFNYSFKESENYNSLDEFYRDVDICSYKLNINTTINKSILDRLVEEGKLYLFE
IKNQDSNDGKSIGHKNNLHTIYWNAIFENFDNRPKLNGEAEIFYRKAISKDKLGIVKGKK
TKNGTEIIKNYRFSKEKFILHVPITLNFCSNNEYVNDIVNTKFYNFSNLHFLGIDRGEKH
LAYYSLVNKNGEIVDQGTLNLPFTDKDGNQRSIKKEKYFYNKQEDKWEAKEVDCWNYNDL
LDAMASNRDMARKNWQRIGTIKEAKNGYVSLVIRKIADLAVNNERPAFIVLEDLNTGFKR
SRQKIDKSVYQKFELALAKKLNFLVDKNAKRDEIGSPTKALQLTPPVNNYGDIENKKQAG
IMLYTRANYTSQTDPATGWRKTIYLKAGPEETTYKKDGKIKNKSVKDQIIETFTDIGFDG
KDYYFEYDKGEFVDEKTGEIKPKKWRLYSGENGKSLDRFRGEREKDKYEWKIDKIDIVKI
LDDLFVNFDKNISLLKQLKEGVELTRNNEHGTGESLRFAINLIQQIRNTGNNERDNDFIL
SPVRDENGKHFDSREYWDKETKGEKISMPSSGDANGAFNIARKGIIMNAHILANSDSKDL
SLFVSDEEWDLHLNNKTEWKKQLNIFSSRKAMAKRKK

圖 67H

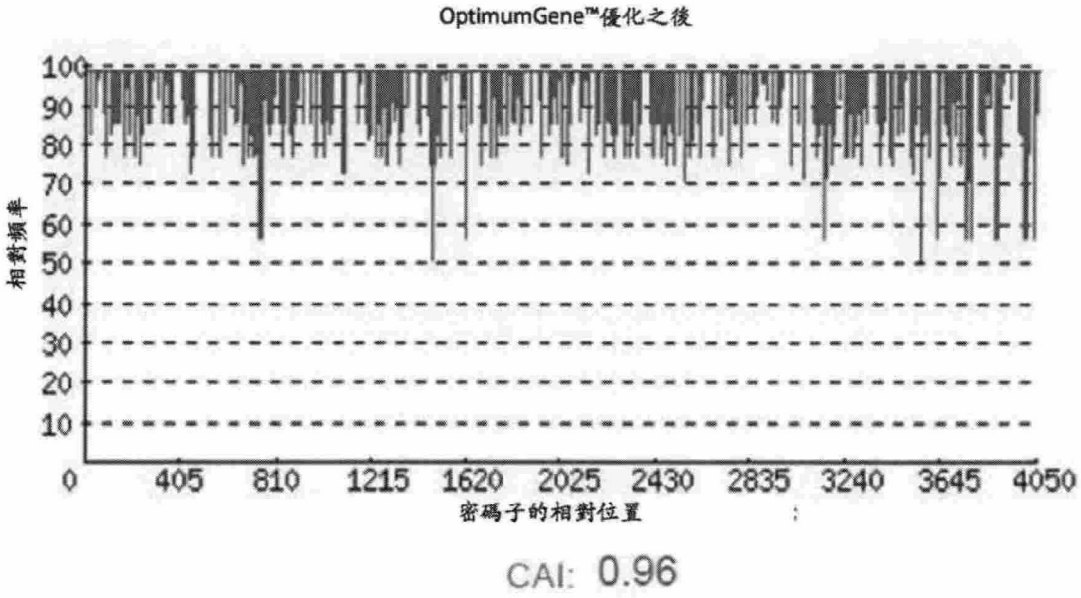

OptimumGene™優化之後
相對頻率
100
90
80
70
60
50
40
30
20
10
0
405
810
1215
1620
2025
2430
2835
3240
3645
4050
密碼子的相對位置
CAI: 0.96

圖 68A

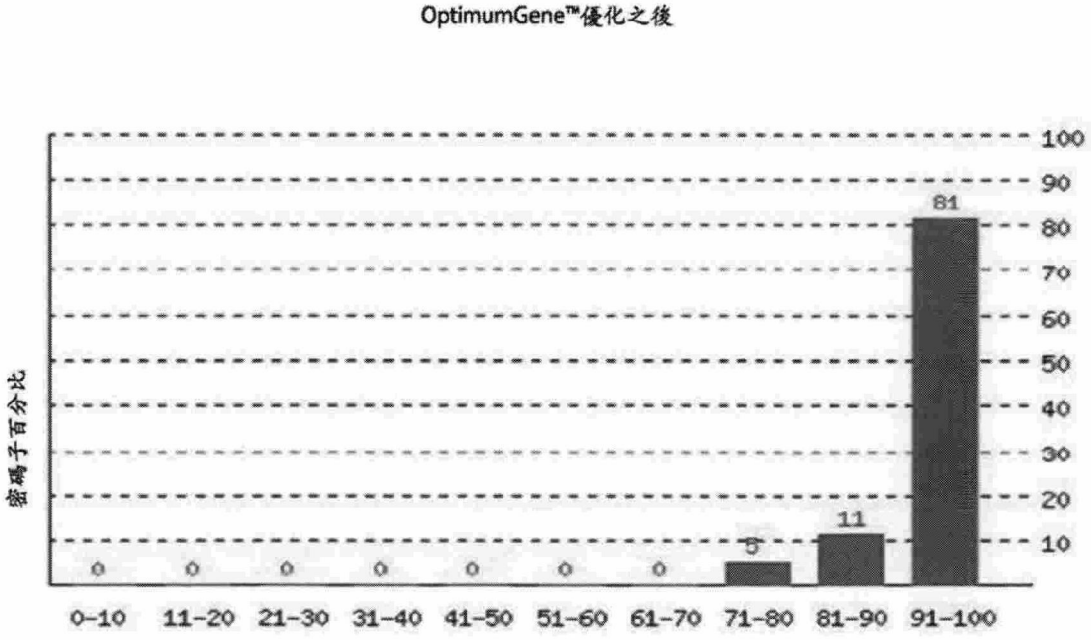

OptimumGene™優化之後
密碼子百分比
100
90
80
70
60
50
40
30
20
10
0
0
0
0
0
0
0
5
11
81
0-10
11-20
21-30
31-40
41-50
51-60
61-70
71-80
81-90
91-100

圖 68B

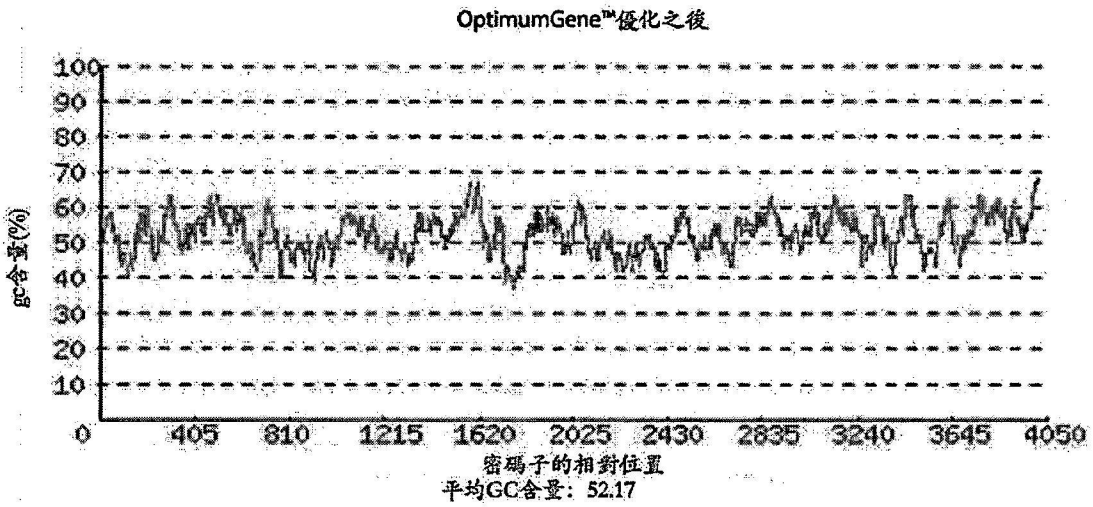

OptimumGene™優化之後
100
90
80
70
60
50
40
30
20
10
0
405
810
1215
1620
2025
2430
2835
3240
3645
4050
GC含量(%)
密碼子的相對位置
平均GC含量：52.17

圖 68C

限制性內切酶	優化的
* 綠色：過濾位點；藍色：檢查位點(未過濾的)；紅色：保留位點	
BamHI(GGATCC)	1(4111)
EcoRI(GAATTC)	1(4201)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(4175)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

順式作用元件	優化的
剪接(GGTAAG)	0
剪接(GGTGAT)	0
PolyA(AATAAA)	0
PolyA(ATTAAA)	0
去穩定化(ATTTA)	0
PolyT(TTTTTT)	0
PolyA(AAAAAAA)	0

抗病毒模體	優化的
	0

圖 68D

優化之後

最大同向重複序列：大小：14 距離：2001 頻率：2

最大反向重複序列：無

最大Dyad重複序列：無

圖 68E

GCTACC

ATGGAGAACATCTTCGACCAGTTTATCGGCAAGTACAGCCTGTCCAAGACCCTGAGATTCGAGCTGAAGCCCGTG
GGCAAGACAGAGGACTTCCTGAAGATCAACAAGGTGTTTGAGAAGGATCAGACCATCGACGATAGCTACAATCAG
GCCAAGTTCTATTTTGATTCCCTGCACCAGAAGTTTATCGACGCCGCCCTGGCCTCCGATAAGACATCCGAGCTG
TCTTTCCAGAACTTTGCCGACGTGCTGGAGAAGCAGAATAAGATCATCCTGGATAAGAAGAGAGAGATGGGCGCC
CTGAGGAAGCGCGACAAGAACGCCGTGGGCATCGATAGGCTGCAGAAGGAGATCAATGACGCCGAGGATATCATC
CAGAAGGAGAAGGAGAAGATCTACAAGGACGTGCGCACCCTGTTCGATAACGAGGCCGAGTCTTGGAAAACCTAC
TATCAGGAGCGGGAGGTGGACGGCAAGAAGATCACCTTCAGCAAGGCCGACCTGAAGCAGAAGGGCGCCGATTTT
CTGACAGCCGCCGGCATCCTGAAGGTGCTGAAGTATGAGTTCCCCGAGGAGAAGGAGAAGGAGTTTCAGGCCAAG
AACCAGCCCTCCCTGTTCGTGGAGGAGAAGGAGAATCCTGGCCAGAAGAGGTACATCTTCGACTCTTTTGATAAG
TTCGCCGGCTATCTGACCAAGTTTCAGCAGACAAAGAAGAATCTGTACGCAGCAGACGGCACCAGCACAGCAGTG
GCCACCCGCATCGCCGATAACTTTATCATCTTCCACCAGAATACCAAGGTGTTCCGGGACAAGTACAAGAACAAT
CACACAGACCTGGGCTTCGATGAGGAGAACATCTTTGAGATCGAGAGGTATAAGAATTGCCTGCTGCAGCGCGAG
ATCGAGCACATCAAGAATGAGAATAGCTACAACAAGATCATCGGCCGGATCAATAAGAAGATCAAGGAGTATCGG
GACCAGAAGGCCAAGGATACCAAGCTGACAAAGTCCGACTTCCCTTTCTTTAAGAACCTGGATAAGCAGATCCTG
GGCGAGGTGGAGAAGGAGAAGCAGCTGATCGAGAAAACCCGGGAGAAAACCGAGGAGGACGTGCTGATCGAGCGG
TTCAAGGAGTTCATCGAGAACAATGAGGAGAGGTTCACCGCCGCCAAGAAGCTGATGAATGCCTTCTGTAACGGC
GAGTTTGAGTCCGAGTACGAGGGCATCTATCTGAAGAATAAGGCCATCAACACAATCTCCCGGGAGATGGTTCGTG
TCTGACAGAGATTTTGAGCTGAAGCTGCCTCAGCAGAAGTCCAAGAACAAGTCTGAGAAGAATGAGCCAAAGGTG
AAGAAGTTCATCTCCATCGCCGAGATCAAGAACGCCGTGGAGGAGCTGGACGGCGATATCTTTAAGGCCGTGTTC
TACGACAAGAAGATCATCGCCCAGGGCGGCTCTAAGCTGGAGCAGTTCCTGGTCATCTGGAAGTACGAGTTTGAG
TATCTGTTCCGGGACATCGAGAGAGAGAACGGCGAGAAGCTGCTGGGCTATGATAGCTGCCTGAAGATCGCCAAG
CAGCTGGGCATCTTCCCACAGGAGAAGGAGGCCCGCGAGAAGGCAACCGCCGTGATCAAGAATTACGCCGACGCC
GGCCTGGGCATCTTCCAGATGATGAAGTATTTTTCTCTGGACGATAAGGATCGGAAGAACACCCCCGGCCAGCTG
AGCACAAATTTCTACGCCCAGTATGACGGCTACTACAACGATTTCGAGTTTATCAAGTACTACAACGAGTTTAGG
AACTTCATCACCAAGAAGCCTTTCGACGAGGATAAGATCAAGCTGAACTTTGAGAATGGCGCCCTGCTGAAGGGC
TGGGACGAGAACAAGGAGTACGATTTCATGGGCGTGATCCTGAAGAAGGAGGGCCGCCTGTATCTGGGCATCATG
CACAAGAACCACCGGAAGCTGTTCAGTCGATGGGCAATGCCAAGGGCGACAACGCCAATAGATACCAGAAGATG
ATCTATAAGCAGATCGCCGACGCCTCTAAGGATGTGCCCAGGCTGCTGCTGACCAGCAAGAAGGCCATGGAGAAG
TTCAAGCCTTCCCAGGAGATCCTGAGAATCAAGAAGGAGAAAACCTTCAAGCGGGAGAGCAAGAACTTTTCCCTG
AGAGATCTGCACGCCCTGATCGAGTACTATAGGAACTGCATCCCTCAGTACAGCAATTGGTCCTTTTATGACTTC
CAGTTTCAGGATACCGGCAAGTACCAGAATATCAAGGAGTTCACAGACGATGTGCAGAAGTACGGCTATAAGATC
TCCTTTCGCGACATCGACCATGAGTATATCAATCAGGCCCTGAACGAGGGCAAGATGTACCTGTTCGAGGTGGTG
AACAAGGATATCTATAACACCAAGAATGGCTCCAAGAATCTGCACACACTGTACTTTGAGCACATCCTGTCTGCC
GAGAACCTGAATGACCCAGTGTTCAAGCTGTCTGGCATGGCCGAGATCTTTCAGCGGCAGCCCAGCGTGAACGAA
AGAGAGAAGATCACCACACAGAAGAATCAGTGTATCCTGGACAAGGGCGATAGAGCCTACAAGTATAGGCGCTAC
ACCGAGAAGAAGATCATGTTCCACATGAGCCTGGTGCTGAACACAGGCAAGGGCGAGATCAAGCAGGTGCAGTTT
AATAAGATCATCAACCAGAGGATCAGCTCCTCTGACAACGAGATGAGGGTGAATGTGATCGGCATCGATCGCGGC
GAGAAGAACCTGCTGTACTATAGCGTGGTGAAGCAGAATGGCGAGATCATCGAGCAGGCCTCCCTGAACGAGATC

圖 68F

AATGGCGTGAACTACCGGGACAAGCTGATCGAGAGGGAGAAGGAGCGCCTGAAGAACCGGCAGAGCTGGAAGCCT
GTGGTGAAGATCAAGGATCTGAAGAAGGGCTACATCTCCCACGTGATCCACAAGATCTGCCAGCTGATCGAGAAG
TATTCTGCCATCGTGGTGCTGGAGGACCTGAATATGAGATTCAAGCAGATCAGGGGAGGAATCGAGCGGAGCGTG
TACCAGCAGTTCGAGAAGGCCCTGATCGATAAGCTGGGCTATCTGGTGTTTAAGGACAACAGGGATCTGAGGGCA
CCAGGAGGCGTGCTGAATGGCTACCAGCTGTCTGCCCCCTTTGTGAGCTTCGAGAAGATGCGCAAGCAGACCGGC
ATCCTGTTCTACACACAGGCCGAGTATACCAGCAAGACAGACCCAATCACCGGCTTTCGGAAGAACGTGTATATC
TCTAATAGCGCCTCCCTGGATAAGATCAAGGAGGCCGTGAAGAAGTTCGACGCCATCGGCTGGGATGGCAAGGAG
CAGTCTTACTTCTTTAAGTACAACCCTTACAACCTGGCCGACGAGAAGTATAAGAACTCTACCGTGAGCAAGGAG
TGGGCCATCTTTGCCAGCGCCCCAAGAATCCGGAGACAGAAGGGCGAGGACGGCTACTGGAAGTATGATAGGGTG
AAAGTGAATGAGGAGTTCGAGAAGCTGCTGAAGGTCTGGAATTTTGTGAACCCAAAGGCCACAGATATCAAGCAG
GAGATCATCAAGAAGGAGAACGCAGGCGACCTGCAGGGAGAGAAGGAGCTGGATGGCCGGCTGAGAAACTTTTGG
CACTCTTTCATCTACCTGTTTAACCTGGTGCTGGAGCTGCGCAATTCTTTCAGCCTGCAGATCAAGATCAAGGCA
GGAGAAGTGATCGCAGTGGACGAGGGCGTGGACTTCATCGCCAGCCCAGTGAAGCCCTTCTTTACCACACCCAAC
CCTTACATCCCCTCCAACCTGTGCTGGCTGGCCGTGGAGAATGCAGACGCAAACGGAGCCTATAATATCGCCAGG
AAGGGCGTGATGATCCTGAAGAAGATCCGCGAGCACGCCAAGAAGGACCCCGAGTTCAAGAAGCTGCCAAACCTG
TTTATCAGCAATGCAGAGTGGGACGAGGCAGCCCGGGATTGGGGCAAGTACGCAGGCACCACAGCCCTGAACCTG
GACCAC

AAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCCTACCCATACGATGTTCCAGAT
TACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCCCAAGAATTC

圖 68G

MENIFDQFIGKYSLSKTLRFELKPVGKTEDFLKINKVFEKDQTIDDSYNQAKFYFDSLHQ
KFIDAALASDKTSELSFQNFADVLEKQNKIILDKKREMGALRKRDKNAVGIDRLQKEIND
AEDIIQKEKEKIYKDVRTLFDNEAESWKTYYQEREVDGKKITFSKADLKQKGADFLTAAG
ILKVLKYEFPEEKEKEFQAKNQPSLFVEEKENPGQKRYIFDSFDKFAGYLTKFQQTKKNL
YAADGTSTAVATRIADNFIIFHQNTKVFRDKYKNNHTDLGFDEENIFEIERYKNCLLQRE
IEHIKNENSYNKIIGRINKKIKEYRDQKAKDTKLTKSDFPFFKNLDKQILGEVEKEKQLI
EKTREKTEEDVLIERFKEFIENNEERFTAAKKLMNAFCNGEFESEYEGIYLKNKAINTIS
RRWFVSDRDFELKLPQQKSKNKSEKNEPKVKKFISIAEIKNAVEELDGDIFKAVFYDKKI
IAQGGSKLEQFLVIWKYEFEYLFRDIERENGEKLLGYDSCLKIAKQLGIFPQEKEAREKA
TAVIKNYADAGLGIFQMMKYFSLDDKDRKNTPGQLSTNFYAEYDGYYKDFEFIKYYNEFR
NFITKKPFDEDKIKLNFENGALLKGWDENKEYDFMGVILKKEGRLYLGIMHKNHRKLFQS
MGNAKGDNANRYQKMIYKQIADASKDVPRLLLTSKKAMEKFKPSQEILRIKKEKTFKRES
KNFSLRDLHALIEYYRNCIPQYSNWSFYDFQFQDTGKYQNIKEFTDDVQKYGYKISFRDI
DDEYINQALNEGKMYLFEVVNKDIYNTKNGSKNLHTLYFEHILSAENLNDPVFKLSGMAE
IFQRQPSVNEREKITTQKNQCILDKGDRAYKYRRYTEKKIMFHMSLVLNTGKGEIKQVQF
NKIINQRISSSDNEMRVNVIGIDRGEKNLLYYSVVKQNGEIIEQASLNEINGVNYRDKLI
EREKERLKNRQSWKPVVKIKDLKKGYISHVIHKICQLIEKYSAIVVLEDLNMRFKQIRGG
IERSVYQQFEKALIDKLGYLVFKDNRDLRAPGGVLNGYQLSAPFVSFEKMRKQTGILFYT
QAEYTSKTDPITGFRKNVYISNSASLDKIKEAVKKFDAIGWDGKEQSYFFKYNPYNLADE
KYKNSTVSKEWAIFASAPRIRRQKGEDGYWKYDRVKVNEEFEKLLKVWNFVNPKATDIKQ
EIIKKEKAGDLQGEKELDGRLRNFWHSFIYLFNLVLELRNSFSLQIKIKAGEVIAVDEGV
DFIASPVKPFFTTPNPYIPSNLCWLAVENADANGAYNIARKGVMILKKIREHAKKDPEFK
KLPNLFISNAEWDEAARDWGKYAGTTALNLDH

圖 68H

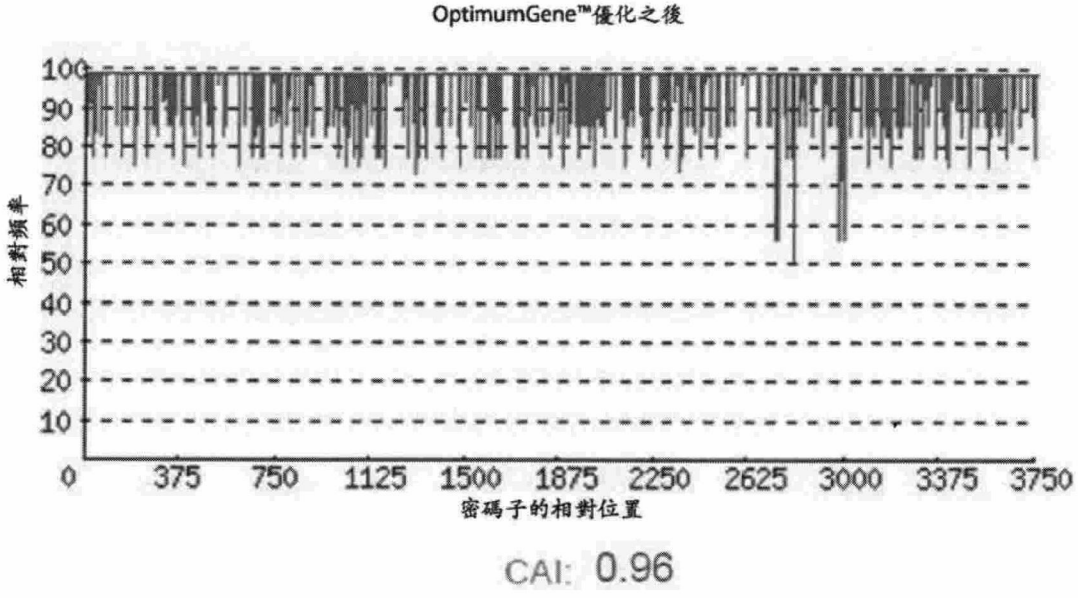

OptimumGene™優化之後
相對頻率
100
90
80
70
60
50
40
30
20
10
0
375
750
1125
1500
1875
2250
2625
3000
3375
3750
密碼子的相對位置
CAI: 0.96

圖 69A

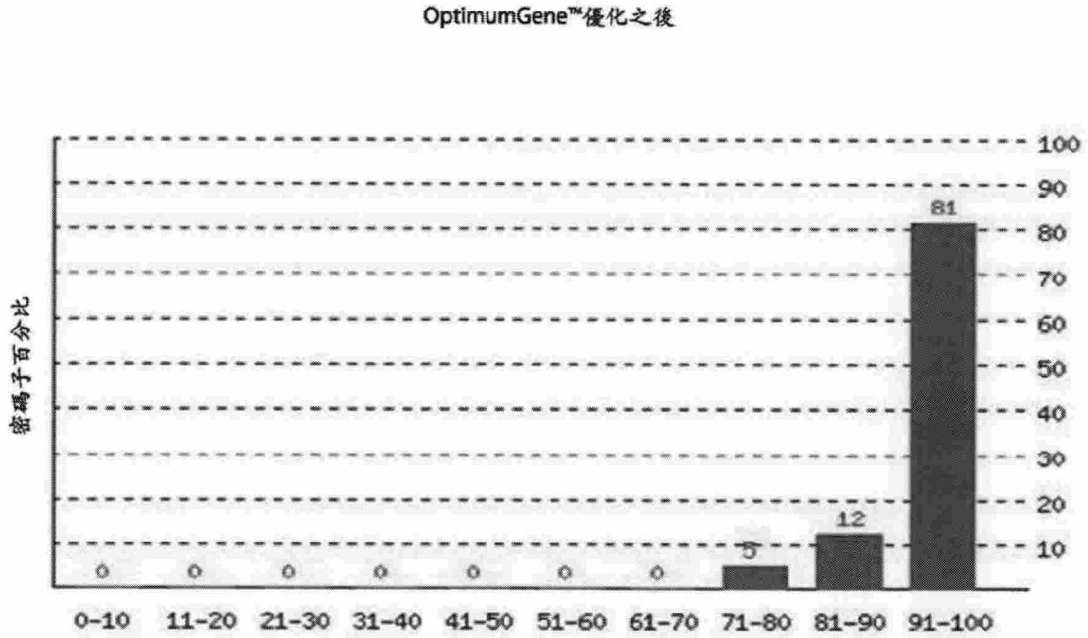

OptimumGene™優化之後
密碼子百分比
100
90
80
70
60
50
40
30
20
10
0
0
0
0
0
0
0
5
12
81
0-10
11-20
21-30
31-40
41-50
51-60
61-70
71-80
81-90
91-100

圖 69B

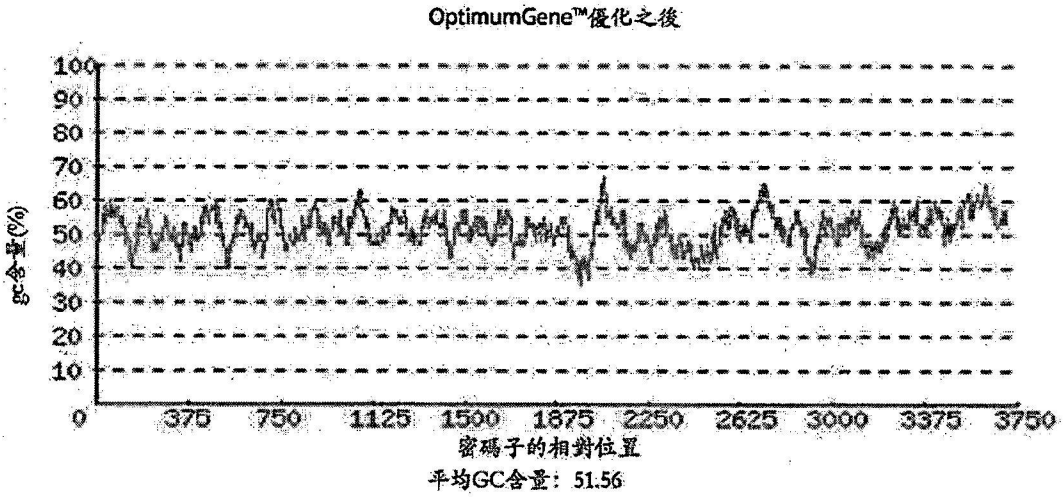

OptimumGene™優化之後
100
90
80
70
60
50
40
30
20
10
0
gc含量(%)
375
750
1125
1500
1875
2250
2625
3000
3375
3750
密碼子的相對位置
平均GC含量：51.56

圖 69C

限制性內切酶	優化的
*綠色：過濾位點；藍色：檢查位點(未過濾的)；紅色：保留位點	
BamHI(GGATCC)	1(3805)
EcoRI(GAATTC)	1(3895)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(3869)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

順式作用元件	優化的
剪接(GGTAAG)	0
剪接(GGTGAT)	0
PolyA(AATAAA)	0
PolyA(ATTAAA)	0
去穩定化(ATTTA)	0
PolyT(TTTTTT)	0
PolyA(AAAAAAA)	0

抗病毒模體	優化的
	0

圖 69D

優化之後

最大同向重複序列：大小：15 距離：93 頻率：2

最大反向重複序列：大小：12 Tm：39.3 起始位置：842,2867

最大Dyad重複序列：無

圖 69E

ATGCAGACCCTGTTTGAGAACTTCACAAATCAGTACCCAGTGTCCAAGACCCTGCGCTTTGAGCTGATCCCCCAG
GGCAAGACAAAGGACTTCATCGAGCAGAAGGGCCTGCTGAAGAAGGATGAGGACCGGGCCGAGAAGTATAAGAAG
GTGAAGAACATCATCGATGAGTACCACAAGGACTTCATCGAGAAGTCTCTGAATGGCCTGAAGCTGGACGGCCTG
GACAAGTACAAGACCCTGTATCTGAAGCAGGAGAAGGACGATAAGGATAAGAAGGGCCTTTGACAAGGAGAAGGAG
AACCTGCGCAAGCAGATCGCCAATGCCTTCCGGAACAATGAGAAGTTTAAGACACTGTTCGCCAAGGAGCTGATC
AAGAACGATCTGATGTCTTTCGCCTGCGAGGAGGACAAGAAGAATGTGAAGGAGTTTGAGGCCTTCACCACATAC
TTCACCGGCTTCCACCAGAACCGCGCCAATATGTACGTGGCCGATGAGAAGAGAACAGCCATCGCCAGCAGGCTG
ATCCACGAGAACCTGCCAAAGTTTATCGACAATATCAAGATCTTCGAGAAGATGAAGAAGGAGGCCCCCGAGCTG
CTGTCTCCTTTCAACCAGACCCTGAAGGATATGAAGGACGTGATCAAGGGCACCACACTGGAGGAGATCTTTAGC
CTGGATTATTTCAACAAGACCCTGACACAGAGCGGCATCGACATCTACAATTCCGTGATCGGCGGCAGAACCCCT
GACGAGGGCAAGACAAAGATCAAGGGCCTGAACGAGTACATCAATACCGACTTCAACCAGAAGCAGACAGACAAG
AAGAAGCGGCAGCCAAAGTTCAAGCAGCTGTATAAGCAGATCCTGAGCGATAGGCAGAGCCTGTCCTTTATCGCC
GAGGCCTTCAAGAACGACACCGAGATCCTGGAGGCCATCGAGAAGTTTACGTGAATGAGCTGCTGCACTTCAGC
AATGAGGGCAAGTCCACAAACGTGCTGGACGCCATCAAGAATGCCGTGTCTAACCTGGAGAGCTTTAACCTGACC
AAGATGTATTTCGGCTCCGGCGCCTCTCTGACAGACGTGAGCCGGAAGGTGTTTGGCGAGTGGAGCATCATCAAT
AGAGCCCTGGACAACTACTATGCCACCACATATCCAATCAAGCCCAGAGAGAAGTCTGAGAAGTACGAGGAGAGG
AAGGAGAAGTGGCTGAAGCAGGACTTCAACGTGAGCCTGATCCAGACCGCCATCGATGAGTACGACAACGAGACA
GTGAAGGGCAAGAACAGCGGCAAAGTGATCGCCGATTATTTTGCCAAGTTCTGCGACGATAAGGAGACAGACCTG
ATCCAGAAGGTGAACGAGGGCTACATCGCCGTGAAGGATCTGCTGAATACACCCTGTCCTGAGAACGAGAAGCTG
GGCAGCAATAAGGACCAGGTGAAGCAGATCAAGGCCTTTATGGATTCTATCATGGACATCATGCACTTCGTGCGC
CCCCTGAGCCTGAAGGATACCGACAAGGAGAAGGATGAGACATTCTACTCCCTGTTCACACCTCTGTACGACCAC
CTGACCCAGACAATCGCCCTGTATAACAAGGTGCGGAACTATCTGACCCAGAAGCCTTACAGCACAGAGAAGATC
AAGCTGAACTTCGAGAACAGCACCCTGCTGGGCGGCTGGGATCTGAATAAGGAGACAGACAACACAGCCATCATC
CTGAGGAAGGATAACCTGTACTATCTGGGCATCATGGACAACAGGCACAATCGCATCTTTCGGAACGTGCCCAAG
GCCGATAAGAAGGACTTCTGCTACGAGAAGATGGTGTATAAGCTGCTGCCTGGCGCCAACAAGATGCTGCCAAAG
GTGTTCTTTTCTCAGAGCAGAATCCAGGAGTTTACCCCTTCCGCCAAGCTGCTGGAGAACTACGCCAATGAGAGCA
CACAAGAAGGGCGATAATTTCAACCTGAATCACTGTCACAAGCTGATCGATTTCTTTAAGGACTCTATCAACAAG
CACGAGGATTGGAAGAATTTCGACTTTAGGTTCAGCGCCACCTCCACCTACGCCGACCTGAGCGGCTTTTACCAC
GAGGTGGAGCACCAGGGCTACAAGATCTCTTTTCAGAGCGTGGCCGATTCCTTCATCGACGATCTGGTGAACGAG
GGCAAGCTGTACCTGTTCCAGATCTATAATAAGGACTTTTCCCCATTCTCTAAGGGCAAGCCCAACCTGCACACC
CTGTACTGGAAGATGCTGTTTGATGAGAACAATCTGAAGGACGTGGTGTATAAGCTGAATGGCGAGGCCGAGGTG
TTCTACCGCAAGAAGAGCATTGCCGAGAAGAACACCACAATCCACAAGGCCAATGAGTCCATCATCAACAAGAAT
CCTGATAACCCAAAGGCCACCAGCACCTTCAACTATGATATCGTGAAGGACAAGAGATACACCATCGACAAGTTT
CAGTTCCACATCCCAATCACAATGAACTTTAAGGCCGAGGGCATCTTCAACATGAATCAGAGGGTGAATCAGTTC
CTGAAGGCCAATCCCGATATCAACATCATCGGCATCGACAGAGGCGAGAGGCACCTGCTGTACTATGCCCTGATC
AACCAGAAGGGCAAGATCCTGAAGCAGGATACCCTGAATGTCATCGCCAACGAGAAGCAGAAGGTGGACTACCAC
AATCTGCTGGATAAGAAGGAGGGCGACCGCGCAACCGCAAGGCAGGAGTGGGGCGTGATCGAGACAATCAAGGAG
CTGAAGGAGGGCTATCTGTCCCAGGTCATCCACAAGCTGACCGATCTGATGATCGAGAACAATGCCATCATCGTG

圖 69F

ATGGAGGACCTGAACTTTGGCTTCAAGCGGGGCAGACAGAAGGTGGAGAAGCAGGTGTATCAGAAGTTTGAGAAG
ATGCTGATCGATAAGCTGAATTACCTGGTGGACAAGAATAAGAAGGCAAACGAGCTGGGAGGCCTGCTGAACGCA
TTCCAGCTGGCCAATAAGTTTGAGTCCTTCCAGAAGATGGGCAAGCAGAACGGCTTTATCTTCTACGTGCCCGCC
TGGAATACCTCTAAGACAGATCCTGCCACCGGCTTTATCGACTTCCTGAAGCCCCGCTATGAGAACCTGAATCAG
GCCAAGGATTTCTTTGAGAAGTTTGACTCTATCCGGCTGAACAGCAAGGCCGATTACTTTGAGTTCGCCTTTGAC
TTCAAGAATTTCACCGAGAAGGCCGATGGCGGCAGAACCAAGTGGACAGTGTGCACCACAAACGAGGACAGATAT
GCCTGGAATAGGGCCCTGAACAATAACAGGGGCAGCCAGGAGAAGTACGACATCACAGCCGAGCTGAAGTCCCTG
TTCGATGGCAAGGTGGACTATAAGTCTGGCAAGGATCTGAAGCAGCAGATCGCCAGCCAGGAGTCCGCCGACTTC
TTTAAGGCCCTGATGAAGAACCTGTCCATCACCCTGTCTCTGAGACACAATAACGGCGAGAAGGGCGATAATGAG
CAGGACTACATCCTGTCCCCTGTGGCCGATTCTAAGGGCCGCTTCTTTGACTCCCGGAAGGCCGACGATGACATG
CCAAAGAATGCCGACGCCAACGGCGCCTATCACATCGCCCTGAAGGGCCTGTGGTGTCTGGAGCAGATCAGCAAG
ACCGATGACCTGAAGAAGGTGAAGCTGGCCATCTCCAACAAGGAGTGGCTGGAGTTCGTGCAGACACTGAAGGGC
AAAAGGCCTTCCGCCAGAAAAAGCCCGCCAAGCAAAAAAGAAAAAGGGATCCTACCCATACGATGTTCCAGAT
TACGCTTATTCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCGGACTATGCCTAAGAATTC

圖 69G

MQTLFENFTNQYPVSKTLRFELIPQGKTKDFIEQKGLLKKDEDRAEKYKKVKNIIDEYHK
DFIEKSLNGLKLDGLEKYKTLYLKQEKDDKDKKAFDKEKENLRKQIANAFRNNEKFKTLF
AKELIKNDLMSFACEEDKKNVKEFEAFTTYFTGFHQNRANMYVADEKRTAIASRLIHENL
PKFIDNIKIFEKMKKEAPELLSPFNQTLKDMKDVIKGTTLEEIFSLDYFNKTLTQSGIDI
YNSVIGGRTPEEGKTKIKGLNEYINTDFNQKQTDKKKRQPKFKQLYKQILSDRQSLSFIA
EAFKNDTEILEAIEKFYVNELLHFSNEGKSTNVLDAIKNAVSNLESFNLTKMYFRSGASL
TDVSRKVFGEWSIINRALDNYYATTYPIKPREKSEKYEERKEKWLKQDFNVSLIQTAIDE
YDNETVKGKNSGKVIADYFAKFCDDKETDLIQKVNEGYIAVKDLLNTPCPENEKLGSNKD
QVKQIKAFMDSIMDIMHFVRPLSLKDTDKEKDETFYSLFTPLYDHLTQTIALYNKVRNYL
TQKPYSTEKIKLNFENSTLLGGWDLNKETDNTAIILRKDNLYYLGIMDKRHNRIFRNVPK
ADKKDFCYEKMVYKLLPGANKMLPKVFFSQSRIQEFTPSAKLLENYANETHKKGDNFNLN
HCHKLIDFFKDSINKHEDWKNFDFRFSATSTYADLSGFYHEVEHQGYKISFQSVADSFID
DLVNEGKLYLFQIYNKDFSPFSKGKPNLHTLYWKMLFDENNLKDVVYKLNGEAEVFYRKK
SIAEKNTTIHKANESIINKNPDNPKATSTFNYDIVKDKRYTIDKFQFHIPITMNFKAEGI
FNMNQRVNQFLKANPDINIIGIDRGERHLLYYALINQKGKILKQDTLNVIANEKQKVDYH
NLLDKKEGDRATARQEWGVIETIKELKEGYLSQVIHKLTDLMIENNAIIVMEDLNFGFKR
GRQKVEKQVYQKFEKMLIDKLNYLVDKNKKANELGGLLNAFQLANKFESFQKMGKQNGFI
FYVPAWNTSKTDPATGFIDFLKPRYENLNQAKDFFEKFDSIRLNSKADYFEFAFDFKNFT
EKADGGRTKWTVCTTNEDRYAWNRALNNNRGSQEKYDITAELKSLFDGKVDYKSGKDLKQ
QIASQESADFFKALMKNLSITLSLRHNNGEKGDNEQDYILSPVADSKGRFFDSRKADDDM
PKNADANGAYHIALKGLWCLEQISKTDDLKKVKLAISNKEWLEFVQTLKG

圖 69H

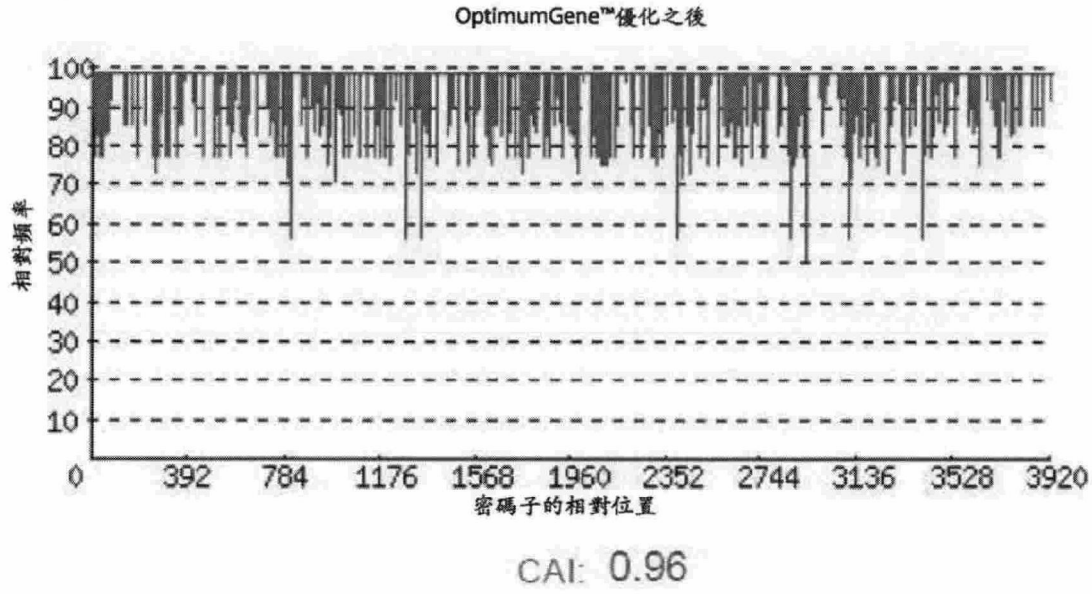

OptimumGene™優化之後
相對頻率
100
90
80
70
60
50
40
30
20
10
0
392
784
1176
1568
1960
2352
2744
3136
3528
3920
密碼子的相對位置
CAI: 0.96

圖 70A

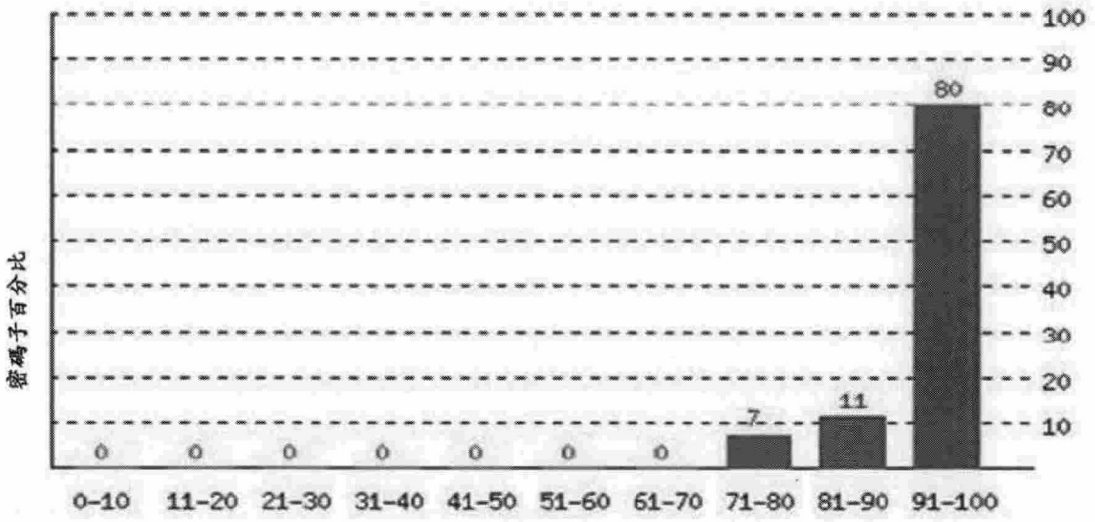

密碼子百分比
100
90
80
70
60
50
40
30
20
10
0
0
0
0
0
0
0
7
11
80
0-10
11-20
21-30
31-40
41-50
51-60
61-70
71-80
81-90
91-100

圖 70B

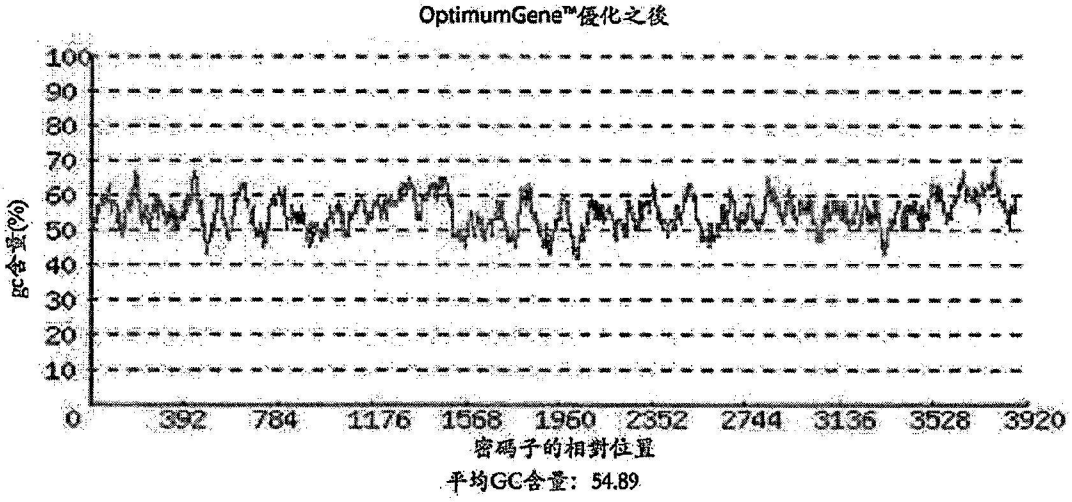

OptimumGene™優化之後
100
90
80
70
60
50
40
30
20
10
0
GC含量(%)
392
784
1176
1568
1960
2352
2744
3136
3528
3920
密碼子的相對位置
平均GC含量：54.89

圖 70C

限制性內切酶	優化的
綠色：過濾位點；藍色：檢查位點(未過濾的)；紅色：保留位點	
BamHI(GGATCC)	1(3976)
EcoRI(GAATTC)	1(4066)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(4040)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

順式作用元件	優化的
剪接(GGTAAG)	0
剪接(GGTGAT)	0
PolyA(AATAAA)	0
PolyA(ATTAAA)	0
去穩定化(ATTTA)	0
PolyT(TTTTTT)	0
PolyA(AAAAAAA)	0

抗病毒模體	優化的
	0

圖 70D

優化之後

最大同向重複序列：大小：15 距離：36 頻率：2

最大反向重複序列：無

最大Dyad重複序列：無

圖 70E

GGTACC

ATGACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAAGACACTGCGGTTTGAGCTGATCCCACAGGGC
AAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAGGACAAGGCCCGCAATGATCACTACAAGGAGCTG
AAGCCCATCATCGATCGGATCTACAAGACCTATGCCGACCAGTGCCTGCAGCTGGTGCAGCTCGATTGGGAGAAC
CTGAGCGCCGCCATCGACTCGTATAGAAAGGAGAAAACCGAGGAGACAAGGAACGCCCTGATCGAGGAGCAGGCC
ACATATCGCAATGCCATCCACGACTACTTCATCGGCCGGACAGACAACCTGACCGATGCCATCAATAAGAGACAC
GCCGAGATCTACAAGGGCCTGTTCAAGGCCGAGCTGTTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACC
ACAACCGAGCACGAGAACGCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTCTCCGGCTTTTATGAGAAC
AGGAAGAACGTGTTCAGCGCCGAGGATATCAGCACAGCCATCCCACACCGCATCGTGCAGGACAACTTCCCCAAG
TTTAAGGAGAATTGTCACATCTTCACACGCCTGATCACCGCCGTGCCCAGCCTGCGGGAGCACTTTGAGAACGTG
AAGAAGGCCATCGGCATCTTCGTGAGCACCTCCATCGAGGAGGTGTTTTCCTTCCCTTTTTATAACCAGCTGCTG
ACACAGACCCAGATCGACCTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAGGCAGGCACCGAGAAGATCAAG
GGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATGAGACAGCCCACATCATCGCCTCCCTGCCACAC
AGATTCATCCCCCTGTTTAAGCAGATCCTGTCCGATAGGAACACCCTGTCTTTCATCCTGGAGGAGTTTAAGAGC
CACCAGGAAGTGATCCAGTCCTTCTGCAAGTACAAGACACTGCTGAGAAACGAGAACGTGCTGGAGACAGCCGAG
GCCCTGTTTAACGAGCTGAACAGCATCGACCTGACACACATCTTCATCAGCCACAAGAAGCTGGAGACAATCAGC
AGCGCCCTGTGCGACCACTGGGATACACTGAGGAATGCCCTGTATGAGCGGAGAATCTCCGAGCTGACAGGCAAG
ATCACCAAGTCTGCCAAGGAGAAGGTGCAGCGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCATCTCT
GCCGCAGGCAAGGAGCTGAGCGAGGCCTTCAAGCAGAAAACCAGCGAGATCCTGTCCCACGCACACGCCGCCCTG
GATCAGCCACTGCCTACAACCCTGAAGAAGCAGGAGGAGAAGGAGATCCTGAAGTCTCAGCTGGACAGCCTGCTG
GGCCTGTACCACCTGCTGGACTGGTTTGCCGTGGATGAGTCCAACGAGGTGGACCCCGAGTTCTCTGCCCGGCTG
ACCGGCATCAAGCTGGAGATGGAGCCTTCTCTGAGCTTCTACAACAAGGCCAGAAATTATGCCACCAAGAAGCCC
TACTCCGTGGAGAAGTTCAAGCTGAACTTTCAGATGCCTACACTGGCCTCTGGCTGGGACGTGAATAAGGAGAAG
AACAATGGCGCCATCCTGTTTGTGAAGAACGGCCTGTACTATCTGGGCATCATGCCAAAGCAGAAGGGCAGGTAT
AAGGCCCTGAGCTTCGAGCCCACAGAGAAAACCAGCGAGGGCTTTGATAAGATGTACTATGACTACTTCCCTGAT
GCCGCCAAGATGATCCCAAAGTGCAGCACCCAGCTGAAGGCCGTGACAGCCCACTTTCAGACCCACACAACCCCC
ATCCTGCTGTCCAACAATTTCATCGAGCCTCTGGAGATCACAAAGGAGATCTACGACCTGAACAATCCTGAGAAG
GAGCCAAAGGAAGTTTCAGACAGCCTACGCCAAGAAAACCGGCGACCAGAAGGGCTACAGAGAGGCCCTGTGCAAG
TGGATCGACTTCACAAGGGATTTTCTGTCCAAGTATACCAAGACAACCTCTATCGATCTGTCTAGCCTGCGGCCA
TCCTCTCAGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATCCCCTGCTGTACCACATCAGCTTCCAGAGA
ATCGCCGAGAAGGAGATCATGGATGCCGTGGAGACAGGCAAGCTGTACCTGTTCCAGATCTATAACAAGGACTTT
GCCAAGGGCCACCACGGCAAGCCTAATCTGCACACACTGTATTGGACCGGCCTGTTTTCTCCAGAGAACCTGGCC
AAGACAAGCATCAAGCTGAATGGCCAGGCCGAGCTGTTCTACCGCCCTAAGTCCAGGATGAAGAGGATGGCACAC
CGGCTGGGAGAGAAGATGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCCGACACCCTGTACCAGGAG
CTGTACGACTATGTGAATCACAGACTGTCCCACGACCTGTCTGATGAGGCCAGGGCCCTGCTGCCCAACGTGATC
ACCAAGGAGGTGTCTCACGAGATCATCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTTCCACGTGCCTATC
ACACTGAACTATCAGGCCGCCAATTCCCCATCTAAGTTCAACCAGAGGGTGAATGCCTACCTGAAGGAGCACCCC
GAGACACCTATCATCGGCATCGATCGGGGCGAGAGAAACCTGATCTATATCACAGTGATCGACTCCACCGGCAAG
ATCCTGGAGCAGCGGAGCCTGAACACCATCCAGCAGTTTGATTACCAGAAGAAGCTGGACAACAGGGAGAAGGAG

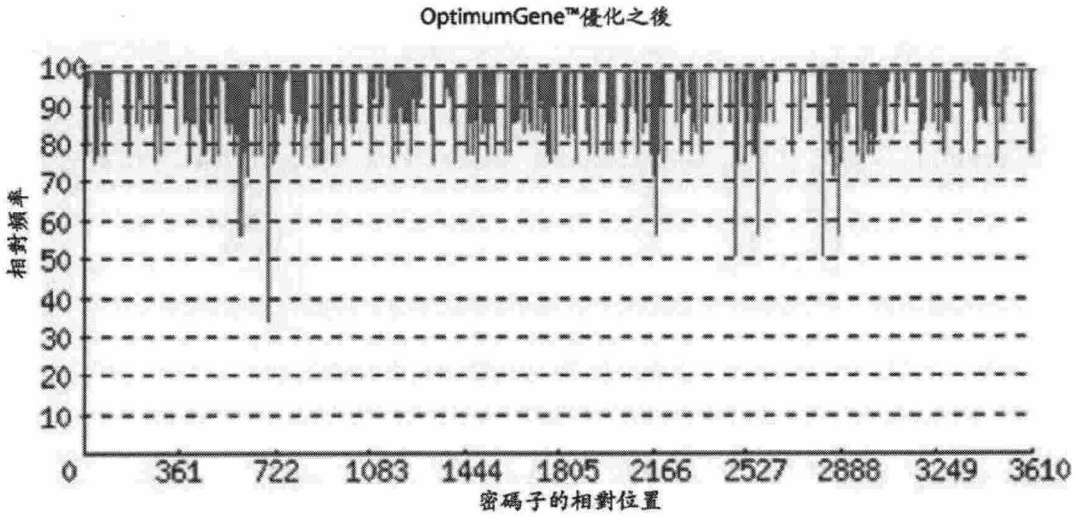
圖 70F

AGGGTGGCAGCAAGGCAGGCCTGGTCTGTGGTGGGCACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTC
ATCCACGAGATCGTGGACCTGATGATCCACTACCAGGCCGTGGTGGTGCTGGAGAACCTGAATTTCGGCTTTAAG
AGCAAGAGGACCGGCATCGCCGAGAAGGCCGTGTACCAGCAGTTCGAGAAGATGCTGATCGATAAGCTGAATTGC
CTGGTGCTGAAGGACTATCCAGCAGAGAAAGTGGGAGGCGTGCTGAACCCATACCAGCTGACAGACCAGTTCACC
TCCTTTGCCAAGATGGGCACCCAGTCTGGCTTCCTGTTTTACGTGCCTGCCCCATATACATCTAAGATCGATCCC
CTGACCGGCTTCGTGGACCCCTTCGTGTGGAAAACCATCAAGAATCACGAGAGCCGCAAGCACTTCCTGGAGGGC
TTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCATCCTGCACTTTAAGATGAACAGAAATCTGTCCTTC
CAGAGGGGCCTGCCCGGCTTTATGCCTGCATGGGATATCGTGTTCGAGAAGAACGAGACACAGTTTGACGCCAAG
GGCACCCCTTTCATCGCCGGCAAGAGAATCGTGCCAGTGATCGAGAATCACAGATTCACCGGCAGATACCGGGAC
CTGTATCCTGCCAACGAGCTGATCGCCCTGCTGGAGGAGAAGGGCATCGTGTTCAGGGATGGCTCCAACATCCTG
CCAAAGCTGCTGGAGAATGACGATTCTCACGCCATCGACACCATGGTGGCCCTGATCCGCAGCGTGCTGCAGATG
CGGAACTCCAATGCCGCCACAGGCGAGGACTATATCAACAGCCCCGTGCGCGATCTGAATGGCGTGTGCTTCGAC
TCCCGGTTTCAGAACCCAGAGTGGCCCATGGACGCCGATGCCAATGGCGCCTACCACATCGCCCTGAAGGGCCAG
CTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCTGAAGCTGCAGAACGGCATCTCCAATCAGGACTGGCTGGCC
TACATCCAGGAGCTGCGCAAC
AAAAGGCCGGCGCCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCCTACCCATACGATGTTCCAGAT
TACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATCCATATGTCCCTGACTATGCCTAAGAATTC

圖 70G

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKT
YADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDA
INKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVF
SAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEV
FSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPH
RFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSID
LTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINL
QEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHL
LDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTL
ASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPD
AAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYA
KKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYH
ISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIK
LNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSD
EARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHP
ETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQAWSV
VGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLI
DKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFV
DPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVF
EKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNIL
PKLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPM
DADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRN

圖 70H



CAI: 0.96

圖 71A

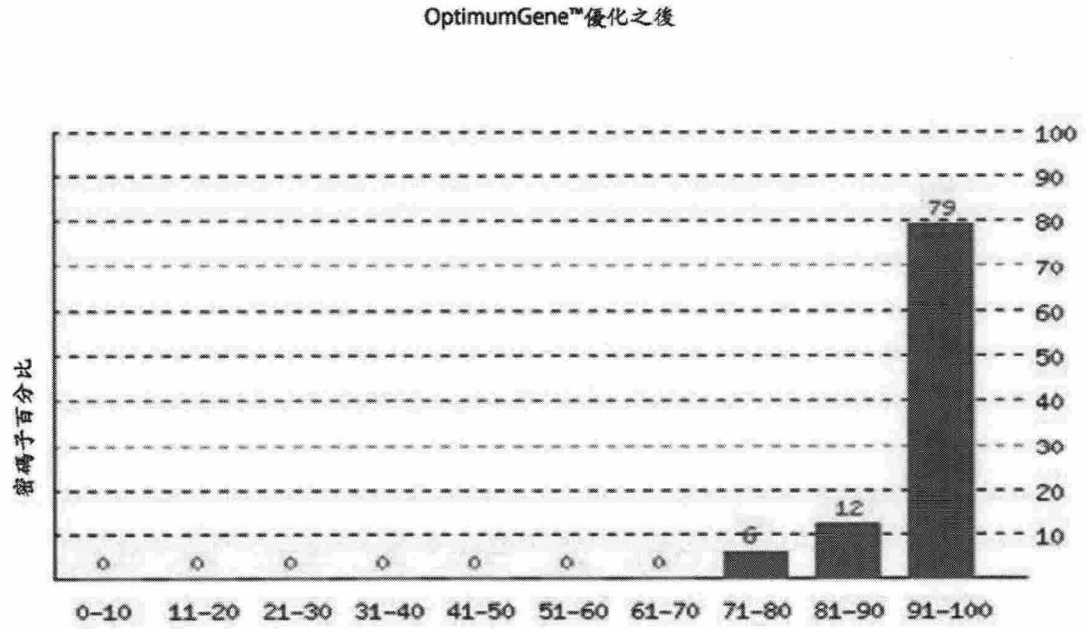


圖 71B

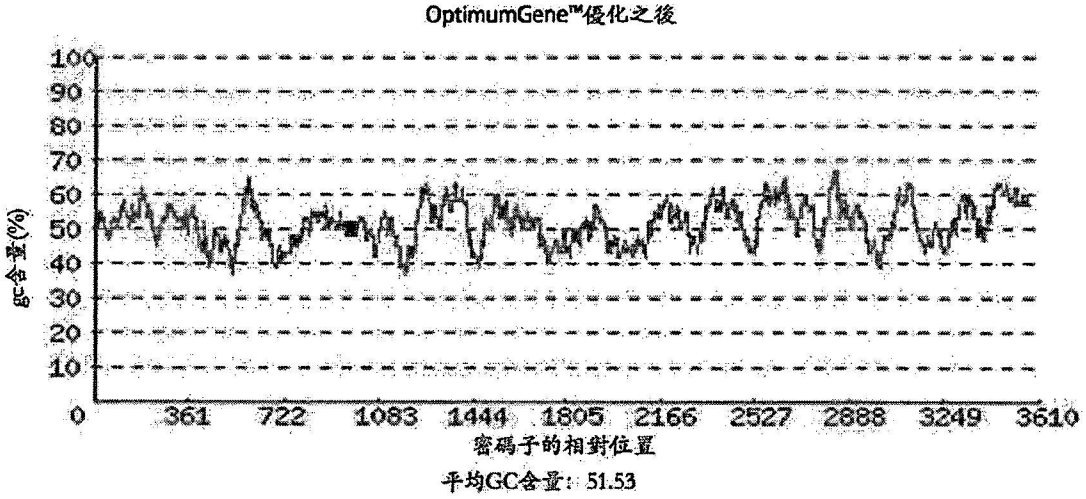

OptimumGene™優化之後
100
90
80
70
60
50
40
30
20
10
0
GC含量(%)
361
722
1083
1444
1805
2166
2527
2888
3249
3610
密碼子的相對位置
平均GC含量：51.53

圖 71C

限制性內切酶	優化的
綠色：過濾位點；藍色：檢查位點(未過濾的)；紅色：保留位點	
BamHI(GGATCC)	1(3673)
EcoRI(GAATTC)	1(3763)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(3737)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

順式作用元件	優化的
剪接(GGTAAG)	0
剪接(GGTGAT)	0
PolyA(AATAAA)	0
PolyA(ATTAAA)	0
去穩定化(ATTTA)	0
PolyT(TTTTTT)	0
PolyA(AAAAAAA)	0

抗病毒模體	優化的
	0

圖 71D

優化之後

最大同向重複序列：大小：17 距離：1800 頻率：2

最大反向重複序列：無

最大Dyad重複序列：無

圖 71E

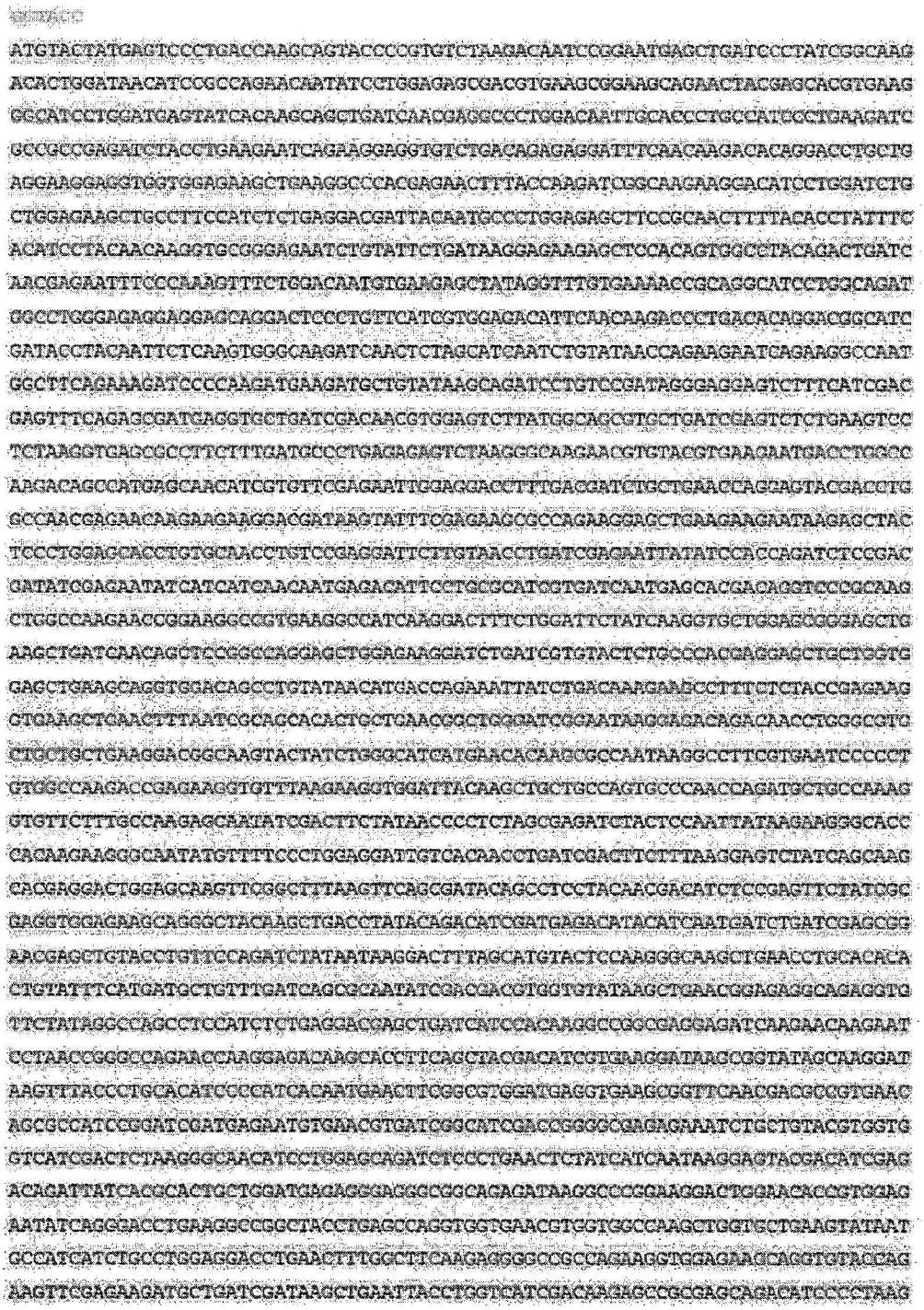

ATGTACTATGAGTCCCTGACCAAGCAGTACCCCGTGTCTAAGACAATCCGGAATGAGCTGATCCCTATCGGCAAG
ACACTGGATAACATCCGCCAGAACAATATCCTGGAGAGCGACGTGAAGCGGAAGCAGAACTACGAGCACGTGAAG
GCCATCCTGGATGAGTATCACAAGCAGCTGATCAACGAGGCCCTGGACAATTGCACCCTGCCATCGCTGAAGATC
GCCGCCGAGATCTACCTGAAGAATCAGAAGGAGGTGTCTGACAGAGAGGATTTCAACAAGACACAGGACCTGCTG
AGGAAGGAGGTGGTGGAGAAGCTGAAGGCCCACGAGAACTTTACCAAGATCGGCAAGAAGGACATCCTGGATCTG
CTGGAGAAGCTGCCTTCCATCTCTGAGGACGATTACAATGCCCTGGAGAGCTTCCGCAACTTTTACACCTATTTC
ACATCCTACAACAAGGTGCGGGAGAATCTGTATTCTGATAAGGAGAAGAGCTCCACAGTGGCCTACAGACTGATC
AACGAGAATTTCCCAAAGTTTCTGGACAATGTGAAGAGCTATAGGTTTGTGAAAACCGCAGGCATCCTGGCAGAT
GGCCTGGGAGAGGAGGAGCAGGACTCCCTGTTCATCGTGGAGACATTCAACAAGACCCTGACACAGGACGGCATC
GATACCTACAATTCTCAAGTGGGCAAGATCAACTCTAGCATCAATCTGTATAACCAGAAGAATCAGAAGGCCAAT
GGCTTCAGAAAGATCCCCAAGATGAAGATGCTGTATAAGCAGATCCTGTCCGATAGGGAGGAGTCTTTCATCGAC
GAGTTTCAGAGCGATGAGGTGCTGATCGACAACGTGGAGTCTTATGGCAGCGTGCTGATCGAGTCTCTGAAGTCC
TCTAAGGTGAGCGCCTTCTTTGATGCCCTGAGAGAGTCTAAGGGCAAGAACGTGTACGTGAAGAATGACCTGGCC
AAGACAGCCATGAGCAACATCGTGTTCGAGAATTGGAGGACCTTTGACGATCTGCTGAACCAGGAGTACGACCTG
GCCAACGAGAACAAGAAGAAGGACGATAAGTATTTCGAGAAGCGCCAGAAGGAGCTGAAGAAGAATAAGAGCTAC
TCCCTGGAGCACCTGTGCAACCTGTCCGAGGATTCTTGTAACCTGATCGAGAATTATATCCACCAGATCTCCGAC
GATATCGAGAATATCATCATCGAACAATGAGACATTCCTGCGCATCGTGATCAATGAGCACGACAGGTCCCGCAAG
CTGGCCAAGAACCGGAAGGCCGTGAAGGCCATCAAGGACTTTCTGGATTCTATCAAGGTGCTGGAGGGGGAGCTG
AAGCTGATCAACAGCTCCGGCCAGGAGCTGGAGAAGGATCTGATCGTGTACTCTGCCCACGAGGAGCTGCTGGTG
GAGCTGAAGCAGGTGGACAGCCTGTATAACATGACCAGAAATTATCTGACAAAGAAGCCTTTCTCTACCGAGAAG
GTGAAGCTGAACTTTAATCGCAGCACACTGCTGAACGGCTGGGATCGGAATAAGGAGACAGACAACCTGGGCGTG
CTGCTGCTGAAGGACGGCAAGTACTATCTGGGCATCATGAACACAAGCGCCAATAAGGCCTTCGTGAATCCCCCT
GTGGCCAAGACCGAGAAGGTGTTTAAGAAGGTGGATTACAAGCTGCTGCCAGTGCCCAACCAGATGCTGCCAAAG
GTGTTCTTTGCCAAGAGCAATATCGACTTCTATAACCCCTCTAGCGAGATCTACTCCAATTATAAGAAGGGCACC
CAGAAGAAGGGCAATATGTTTTCCCTGGAGGATTGTCACAACCTGATCGACTTCTTTAAGGAGTCTATCAGCAAG
CACGAGGACTGGAGCAAGTTCGGCTTTAAGTTCAGCGATACAGCCTCCTACAACGACATCTCCGAGTTCTATCGC
GAGGTGGAGAAGCAGGGCTACAAGCTGACCTATACAGACATCGATGAGACATACATCAATGATCTGATCGAGCGG
AACGAGCTGTACCTGTTCCAGATCTATAATAAGGACTTTAGCATGTACTCCAAGGGCAAGCTGAACCTGCACACA
CTGTATTTCATGATGCTGTTTGATCAGCGCAATATCGACGACGTGGTGTATAAGCTGAACGGAGAGGCAGAGGTG
TTCTATAGGCCAGCCTCCATCTCTGAGGACGAGCTGATCATCCACAAGGCCGGCGACGAGATCAAGAACAAGAAT
CCTAACGGGGCCAGAACCAAGGAGACAAGCACCTTCAGCTACGACATCGTGAAGGATAAGCGGTATAGCAAGGAT
AAGTTTACCCTGCACATCCCCATCACAATGAACTTCGGCGTGGATGAGGTGAAGCGGTTCAACGACGCCGTGAAC
AGCGCCATCCGGATCGATGAGAATGTGAACGTGATCGGCATCGACCGGGGCGAGAGAAATCTGCTGTACGTGGTG
GTCATCGACTCTAAGGGCAACATCCTGGAGCAGATCTCCCTGAACTCTATCATCAATAAGGAGTACGACATCGAG
ACAGATTATCACGCACTGCTGGATGAGAGGGAGGGCGGCAGAGATAAGGCCCGGAAGGACTGGAACACCGTGGAG
AATATCAGGGACCTGAAGGCCGGCTACCTGAGCCAGGTGGTGAACGTGGTGGCCAAGCTGGTGCTGAAGTATAAT
GCCATCATCTGCCTGGAGGACCTGAACTTTGGCTTCAAGAGGGGCCGCCAGAAGGTGGAGAAGCAGGTGTACCAG
AAGTTCGAGAAGATGCTGATCGATAAGCTGAATTACCTGGTCATCGACAAGAGCCGCGAGCAGACATCCCCTAAG

圖 71F

GAGCTGGGAGGCGCCCTGAACGCACTGCAGCTGACCTCTAAGTTCAAGAGCTTTAAGGAGCTGGGCAAGCAGTCC
GGCGTGATCTACTATGTGCCTGCCTACCTGACCTCTAAGATCGATCCAACCACAGGCTTCGCCAATCTGTTTTAT
ATGAAGTGTGAGAACGTGGAGAAGTCCAAGAGATTCTTTGACGGCTTTGATTTCATCAGGTTCAACGCCCTGGAG
AACGTGTTCGAGTTCGGCTTTGACTACCGGAGCTTCACCCAGAGGGCCTGCGGCATCAATTCCAAGTGGACCGTG
TGCACCAACGGCGAGCGCATCATCAAGTATCGGAATCCAGATAAGAACAATATGTTCGACGAGAAGGTGGTGGTG
GTGACCGATGAGATGAAGAACCTGTTTGAGCAGTACAAGATCCCCTATGAGGATGGCAGAAATGTGAAGGACATG
ATCATCAGCAACGAGGAGGCCGAGTTCTACCGGAGACTGTATAGGCTGCTGCAGCAGACCCTGCAGATGAGAAAC
AGCACCTCCGACGGCACAAGGGATTACATCATCTCCCCTGTGAAGAATAAGAGAGAGGCCTACTTCAACAGCGAG
CTGTCCGACGGCTCTGTGCCAAAGGACGCCGATGCCAACGGCGCCTACAATATCGCCAGAAAGGGCCTGTGGGTG
CTGGAGCAGATCAGGCAGAAGAGCGAGGGCGAGAAGATCAATCTGGCCATGACCAACGCCGAGTGGCTGGAGTAT
GCCCAGACACACCTGCTG



圖 71G

MYYESLTKQYPVSKTIRNELIPIGKTLDNIRQNNILESDVKRKQNYEHVKGILDEYHKQL
INEALDNCTLPSLKIAAEIYLKNQKEVSDREDFNKTQDLLRKEVVEKLKAHENFTKIGKK
DILDLLEKLPSISEDDYNALESFRNFYTYFTSYNKVRENLYSDKEKSSTVAYRLINENFP
KFLDNVKSYRFVKTAGILADGLGEEEQDSLFIVETFNKTLTQDGIDTYNSQVGKINSSIN
LYNQKNQKANGFRKIPKMKMLYKQILSDREESFIDEFQSDEVLIDNVESYGSVLIESLKS
SKVSAFFDALRESKGKNVYVKNDLAKTAMSNIVFENWRTFDDLLNQEYDLANENKKKDDK
YFEKRQKELKKNKSYSLEHLCNLSEDSCNLIENYIHQISDDIENIIINNETFLRIVINEH
DRSRKLAKNRKAVKAIKDFLDSIKVLERELKLINSSGQELEKDLIVYSAHEELLVELKQV
DSLYNMTRNYLTKKPFSTEKVKLNFNRSTLLNGWDRNKETDNLGVLLLKDGKYYLGIMNT
SANKAFVNPPVAKTEKVFKKVDYKLLPVPNQMLPKVFFAKSNIDFYNPSSEIYSNYKKGT
HKKGNMFSLEDCHNLIDFFKESISKHEDWSKFGFKFSDTASYNDISEFYREVEKQGYKLT
YTDIDETYINDLIERNELYLFQIYNKDFSMYSKGKLNLHTLYFMMLFDQRNIDDVVYKLN
GEAEVFYRPASISEDELIIHKAGEEIKNKNPNRARTKETSTFSYDIVKDKRYSKDKFTLH
IPITMNFGVDEVKRFNDAVNSAIRIDENVNVIGIDRGERNLLYVVVIDSKGNILEQISLN
SIINKEYDIETDYHALLDEREGGRDKARKDWNTVENIRDLKAGYLSQVVNVVAKLVLKYN
AIICLEDLNFGFKRGRQKVEKQVYQKFEKMLIDKLNYLVIDKSREQTSPKELGGALNALQ
LTSKFKSFKELGKQSGVIYYVPAYLTSKIDPTTGFANLFYMKCENVEKSKRFFDGFDFIR
FNALENVFEFGFDYRSFTQRACGINSKWTVCTNGERIIKYRNPDKNNMFDEKVVVVTDEM
KNLFEQYKIPYEDGRNVKDMIISNEEAEFYRRLYRLLQQTLQMRNSTSDGTRDYIISPVK
NKREAYFNSELSDGSVPKDADANGAYNIARKGLWVLEQIRQKSEGEKINLAMTNAEWLEY
AQTHLL

圖 71H

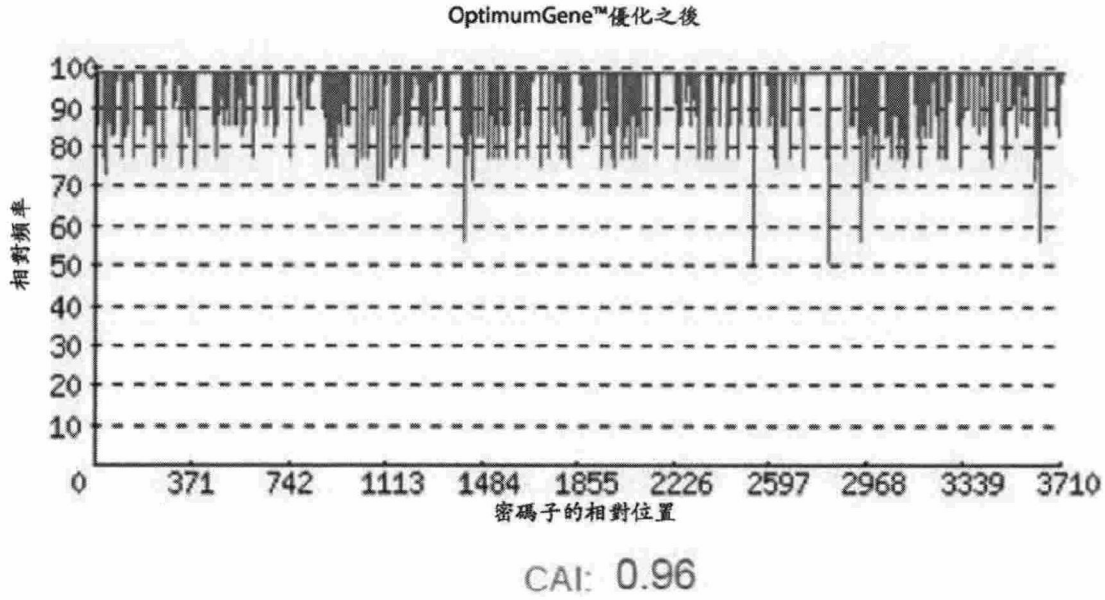

OptimumGene™優化之後
相對頻率
100
90
80
70
60
50
40
30
20
10
0
371
742
1113
1484
1855
2226
2597
2968
3339
3710
密碼子的相對位置
CAI: 0.96

圖 72A

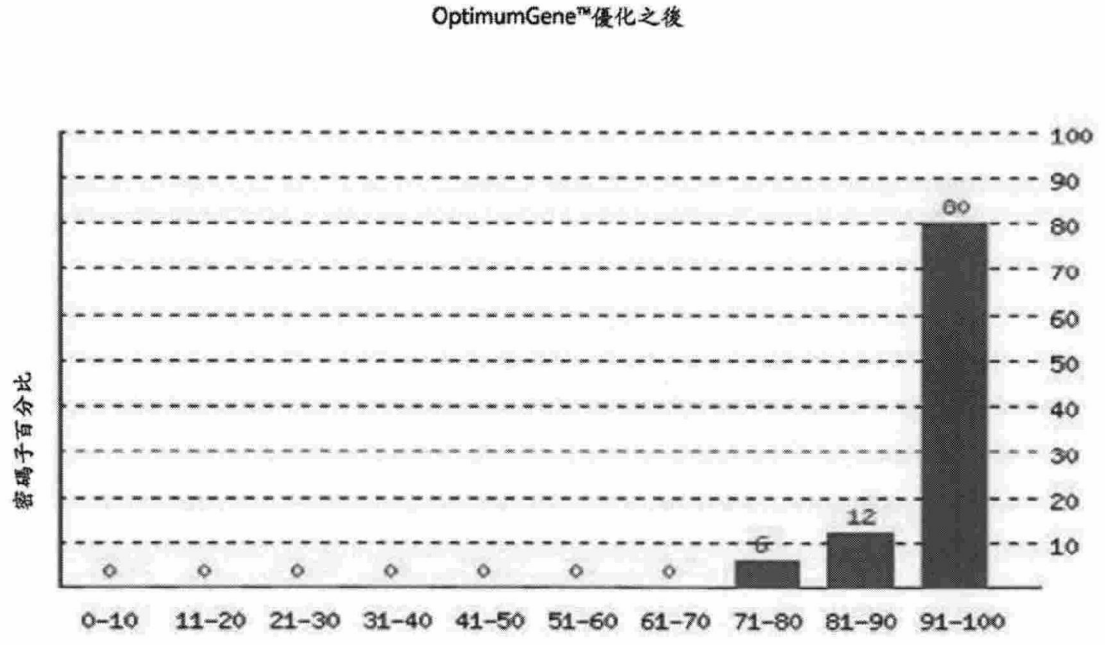

OptimumGene™優化之後
密碼子百分比
100
90
80
70
60
50
40
30
20
10
6
12
80
0-10
11-20
21-30
31-40
41-50
51-60
61-70
71-80
81-90
91-100

圖 72B

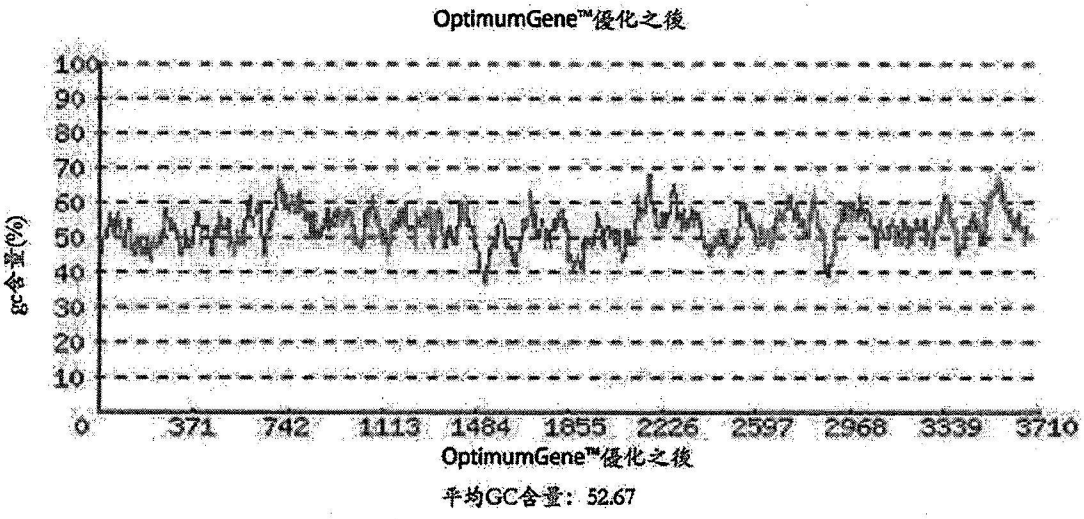

OptimumGene™優化之後
100
90
80
70
60
50
40
30
20
10
0
GC含量(%)
371
742
1113
1484
1855
2226
2597
2968
3339
3710
OptimumGene™優化之後
平均GC含量：52.67

圖 72C

限制性內切酶	優化的
*綠色：過濾位點；藍色：檢查位點(未過濾的)；紅色：保留位點	
BamHI(GGATCC)	1(3769)
EcoRI(GAATTC)	1(3859)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(3833)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

順式作用元件	優化的
剪接(GGTAAG)	0
剪接(GGTGAT)	0
PolyA(AATAAA)	0
PolyA(ATTAAA)	0
去穩定化(ATTTA)	0
PolyT(TTTTTT)	0
PolyA(AAAAAAA)	0

抗病毒模體	優化的
	0

圖 72D

優化之後

最大同向重複序列：大小：13 距離：1506 頻率：2

最大反向重複序列：無

最大Dyad重複序列：無

圖 72E

GTTACC

ATGAACAATTACGACGAGTTCACCAAGCTGTATCCTATCCAGAAAACCATCCGGTTTGAGCTGAAGCCACAGGGC
AGAACCATGGAGCACCTGGAGACATTCAACTTCTTTGAGCAGGACCGGGATAGAGCCGAGAAGTATAAGATCCTG
AAGGAGGCCATCGACGAGTACCACAAGAAGTTTATCGATGAGCACCTGACCAATATGTCCCTGGATTGGAACTCT
CTGAAGCAGATCAGCGAGAAGTACTATAAGAGCAGGGAGGAGAAGGACAAGAAGGTGTTCCTGTCCGAGCAGAAG
AGGATGCGCCAGGAGATCGTGTCTGAGTTTAAGAAGGACGATCGCTTCAAGGACCTGTTTTCCAAGAAGCTGTTC
TCTGAGCTGCTGAAGGAGGAGATCTACAAGAAGGGCAACCACCAGGAGATCGACGCCCTGAAGAGCTTCGATAAG
TTTTCCGGCTATTTCATCGGCCTGCACGAGAATAGGAAGAACATGTACTCCGACGGCGATGAGATCACCGCCATC
TCCAATCGCATCGTGAATGAGAACTTCCCCAAGTTTCTGGATAACCTGCAGAAGTACCAGGAGGCCAGGAAGAAG
TATCCTGAGTGGATCATCAAGGCCGAGAGCGCCCTGGTGGCCCACAATATCAAGATGGACGAGGTGTTCTCCCTG
GAGTACTTTAATAAGGTGCTGAACCAGGAGGGCATCCAGCGGTACAACCTGGCCCTGGGCGGCTATGTGACCAAG
AGCGGCGAGAAGATGATGGGCCTGAATGATGCCCTGAACCTGGCCCACCAGTCCGAGAAGAGCTCCAAGGGCAGA
ATCCACATGACCCCCCTGTTCAAGCAGATCCTGTCCGAGAAGGAGTCCTTCTCTTACATCCCCGACGTGTTTACA
GAGGATTCTCAGCTGCTGCCTAGCATCGGCGGCTTCTTTGCCCAGATCGAGAATGACAAGGATGGCAACATCTTC
GACCGGGCCCTGGAGCTGATCTCTAGCTACGCCGAGTATGATACCGAGCGGATCTATATCAGACAGGCCGACATC
AATAGAGTGTCCAACGTGATCTTTGGAGAGTGGGGCACCCTGGGAGGCCTGATGACGGAGTACAAGGCCGACTCT
ATCAATGATATCAACCTGGAGCGCACATGCAAGAAGGTGGACAAGTGGCTGGATTCTAAGGAGTTTGCCCTGAGC
GATGTGCTGGAGGCCATCAAGAGGACCGGCAACAATGACGCCTTCAACGAGTATATCTCCAAGATGCGGACAGCC
AGAGAGAAGATCGATGCCGCCCGCAAGGAGATGAAGTTCATCAGCGAGAAGATCTCCGGCCATGAGGAGTCTATC
CACATCATCAAGACCCTGCTGGACAGCGTGCAGCAGTTCCTGCACTTCTTTAATCTGTTTAAGGCAAGGCAGGAC
ATGCCACTGGATGGAGCCTTCTACGCCGAGTTTGACGAGGTGCACAGCAAGCTGTTTGCCATCGTGCCCCTGTAT
AACAAGGTGCGGAACTATCTGACCAAGAACAATCTGAACACAAAGAAGATCAAGCTGAATTTCAAGAACCCTACA
CTGGCCAATGGCTGGGACCAGAACAAGGTGTACGATTATGCCTCCCTGATCTTTCTGCGGGACGGCAATTACTAT
CTGGGCATCATCAATCCTAAGAGAAAGAAGAAGAACATCAAGTTCGAGCAGCGCTCTGGCAACGGCCCCTTCTACCGG
AAGATGGTGTATAAGCAGATCCCCGGCCCTAATAAGAACCTGCCAAGAGTGTTCCTGACCTGCACAAAGGGCAAG
AAGGAGTATAAGCCCTCTAAGGAGATCATCGAGGGCTACGAGGCCGACAAGCACATCAGGGGCGATAAGTTCGAC
CTGGATTTTTGTCACAAGCTGATCGATTTCTTTAAGGAGTCCATCGAGAAGCACAAGGACTGGTCTAAGTTCAAC
TTCTACTTCAGCCCAACCGAGAGCTATGGCGACATCTCTGAGTTCTACCTGGATGTGGAGAAGCAGGGCTATGGC
ATGCACTTTGAGAATATCAGCGCCGAGACAATCGACGAGTATGTGGAGAAGGGCGATCTGTTTCTGTTCCAGATC
TACAACAAGGATTTTGTGAAGGCCGCCACCGGCAAGAAGGACATGCACACAATCTACTGGAATGCCGCCTTCAGC
CCCGAGAACCTGCAGGACGTGGTGGTGAAGCTGAACGGCGAGGCCGAGCTGTTTTATAGGGACAAGTCCGATATC
AAGGAGATCGTGCACCGCGAGGGCGAGATCCTGGTGAATAGGACCTACAACGGCCGCACACCAGTGCCCGACAAG
ATCCACAAGAAGCTGACCGATTATCACAATGGCCGGACAAACGACCTGGGCGAGGCCAAGCAGTACCTGGATAAG
GTGAGATACTTCAAGGCCCACTATGACATCACCAAGGATCGGAGATACCTGAACGACAAGATCTATTTCCACGTG
CCTCTGACCCTGAACTTCAAGGCCAACGGCAAGAAGAATCTGAACAAGATGGTCATCGAGAAGTTCCTGTCCGAT
GAGAAGGCCCACATCATCGGCATCGACAGGGGCGAGCGCAATCTGCTGTACIATTCCATCATCGACAGGTCTGGC
AAGATCATCGATCAGCAGAGCCTGAATGTGATCGACGGCTTTGATTATCGGGAGAAGCTGAACCAGAGAGAGATC
GAGATGAAGGATGCCCGCCAGTCTTGGAACGCCATCGGCAAGATCAAGGACCTGAAGGAGGGCTACCTGAGCAAG
GCCGTGCACGAGATCACCAAGATGGCCATCCAGTATAATGCCATCGTGGTCATGGAGGAGCTGAACTACGGCTTC

圖 72F

AAGCGGGGCCGGTTCAAGGTGGAGAAGCAGATCTATCAGAAGTTCGAGAATATGCTGATCGATAAGATGAACTAC
CTGGTGTTTAAGGACGCACCTGATGAGTCCCCAGGAGGCGTGCTGAATGCCTACCAGCTGACAAACCCACTGGAG
TCTTTCGCCAAGCTGGGCAAGCAGACCGGCATCCTGTTTTACGTGCCAGCCGCCTATACATCCAAGATCGACCCC
ACCACAGGCTTCGTGAATCTGTTTAACACCTCCTCTAAGACAAACGCCCAGGAGCGGAAGGAGTTCCTGCAGAAG
TTTGAGAGCATCTCCTATTCTGCCAAGGATGGCGGCATCTTTGCCTTCGCCTTTGACTACAGAAAGTTCGGCACC
AGCAAGACAGATCACAAGAACGTGTGGACCGCCTATACAAACGGCGAGAGGATGCGCTACATCAAGGAGAAGAAG
CGGAATGAGCTGTTTGACCCTTCTAAGGAGATCAAGGAGGCCCTGACCAGCTCCGGCATCAAGTACGATGGCGGC
CAGAACATCCTGCCAGACATCCTGAGGAGCAACAATAACGGCCTGATCTACACAATGTATTCTAGCTTCATCGCC
GCCATCCAGATGCGCGTGTACGACGGCAAGGAGGATTATATCATCAGCCCCATCAAGAACTCCAAGGGCGAGTTC
TTTAGGACCGACCCCAAGAGGCGCGAGCTGCCTATCGACGCCGATGCCAATGGCGCCTACAACATCGCCCTGAGG
GGAGAGCTGACAATGAGGGCAATCGCAGAGAAGTTCGACCCTGATAGCGAGAAGATGGCCAAGCTGGAGCTGAAG
CACAAGGATTGGTTCGAGTTTATGCAGACCAGAGGCGAC

圖 72G

MNNYDEFTKLYPIQKTIRFELKPQGRTMEHLETFNFFEEDRDRAEKYKILKEAIDEYHKK
FIDEHLTNMSLDWNSLKQISEKYYKSREEKDKKVFLSEQKRMRQEIVSEFKKDDRFKDLF
SKKLFSELLKEEIYKKGNHQEIDALKSFDKFSGYFIGLHENRKNMYSDGDEITAISNRIV
NENFPKFLDNLQKYQEARKKYPEWIIKAESALVAHNIKMDEVFSLEYFNKVLNQEGIQRY
NLALGGYVTKSGEKMMGLNDALNLAHQSEKSSKGRIHMTPLFKQILSEKESFSYIPDVFT
EDSQLLPSIGGFFAQIENDKDGNIFDRALELISSYAEYDTERIYIRQADINRVSNVIFGE
WGTLGGLMREYKADSINDINLERTCKKVDKWLDSKEFALSDVLEAIKRTGNNDAFNEYIS
KMRTAREKIDAARKEMKFISEKISGDEESIHIIKTLLDSVQQFLHFFNLFKARQDIPLDG
AFYAEFDEVHSKLFAIVPLYNKVRNYLTKNNLNTKKIKLNFKNPTLANGWDQNKVYDYAS
LIFLRDGNYYLGIINPKRKKNIKFEQGSGNGPFYRKMVYKQIPGPNKNLPRVFLTSTKGK
KEYKPSKEIIEGYEADKHIRGDKFDLDFCHKLIDFFKESIEKHKDWSKFNFYFSPTESYG
DISEFYLDVEKQGYRMHFENISAETIDEYVEKGDLFLFQIYNKDFVKAATGKKDMHTIYW
NAAFSPENLQDVVVKLNGEAELFYRDKSDIKEIVHREGEILVNRTYNGRTPVPDKIHKKL
TDYHNGRTKDLGEAKEYLDKVRYFKAHYDITKDRRYLNDKIYFHVPLTLNFKANGKKNLN
KMVIEKFLSDEKAHIIGIDRGERNLLYYSIIDRSGKIIDQQSLNVIDGFDYREKLNQREI
EMKDARQSWNAIGKIKDLKEGYLSKAVHEITKMAIQYNAIVVMEELNYGFKRGRFKVEKQ
IYQKFENMLIDKMNYLVFKDAPDESPGGVLNAYQLTNPLESFAKLGKQTGILFYVPAAYT
SKIDPTTGFVNLFNTSSKTNAQERKEFLQKFESISYSAKDGGIFAFAFDYRKFGTSKTDH
KNVWTAYTNGERMRYIKEKKRNELFDPSKEIKEALTSSGIKYDGGQNILPDILRSNNNGL
IYTMYSSFIAAIQMRVYDGKEDYIISPIKNSKGEFFRTDPKRRELPIDADANGAYNIALR
GELTMRAIAEKFDPDSEKMAKLELKHKDWFEFMQTRGD

圖 72H

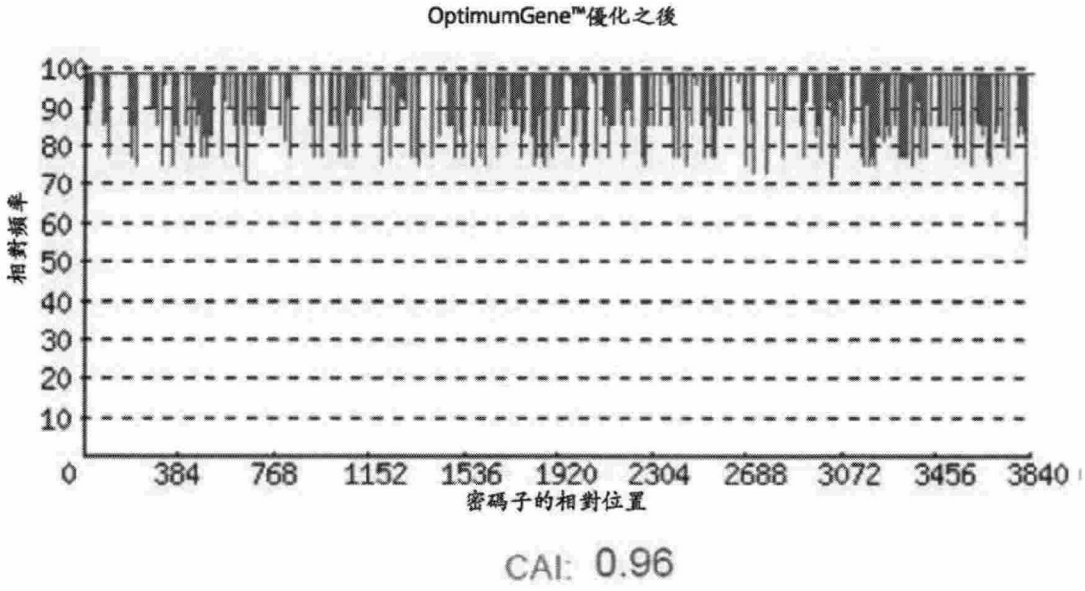

OptimumGene™優化之後
相對頻率
100
90
80
70
60
50
40
30
20
10
0
384
768
1152
1536
1920
2304
2688
3072
3456
3840
密碼子的相對位置
CAI: 0.96

圖 73A

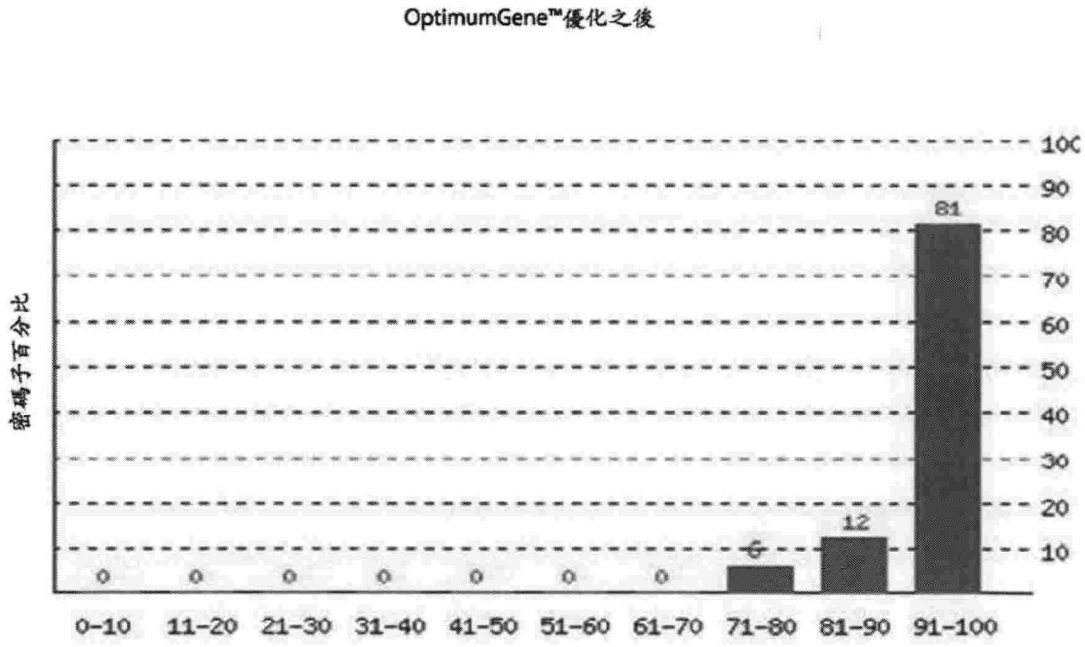

OptimumGene™優化之後
密碼子百分比
100
90
80
70
60
50
40
30
20
10
0
0
0
0
0
0
0
6
12
81
0-10
11-20
21-30
31-40
41-50
51-60
61-70
71-80
81-90
91-100

圖 73B

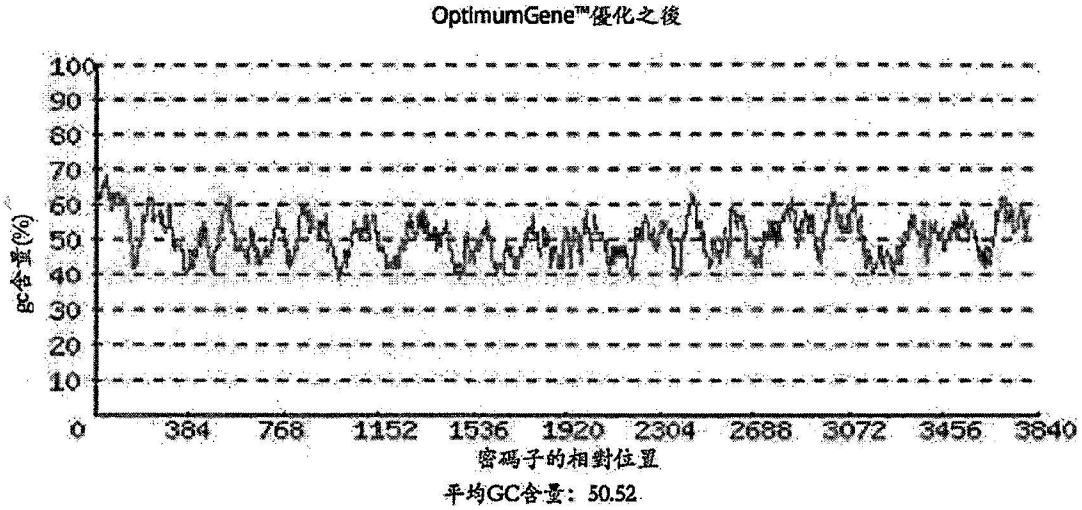

OptimumGene™優化之後
100
90
80
70
60
50
40
30
20
10
0
GC含量(%)
384
768
1152
1536
1920
2304
2688
3072
3456
3840
密碼子的相對位置
平均GC含量：50.52

圖 73C

限制性內切酶	優化的
* 綠色：過濾位點；藍色：檢查位點(未過濾的)；紅色：保留位點	
BamHI(GGATCC)	1(3901)
EcoRI(GAATTC)	1(3991)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(3965)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

順式作用元件	優化的
剪接(GGTAAG)	0
剪接(GGTGAT)	0
PolyA(AATAAA)	0
PolyA(ATTAAA)	0
去穩定化(ATTTA)	0
PolyT(TTTTTT)	0
PolyA(AAAAAAA)	0

抗病毒模體	優化的
	0

圖 73D

優化之後

最大同向重複序列：大小：14 距離：894 頻率：2

最大反向重複序列：無

最大Dyad重複序列：大小：13 Tm：40.1 起始位置：2322,3831

圖 73E

GGTTACC

ATGAACGGCAATAGGTCCATCGTGTACCGCGAGTTCGTGGGCGTGATCCCCGTGGCCAAGACCCTGAGGAATGAG
CTGCGCCCTGTGGGCCACACACAGGAGCACATCATCCAGAACGGCCTGATCCAGGAGGACGAGCTGCGGCAGGAG
AAGAGCACCGAGCTGAAGAACATCATGGACGATTACTATAGAGAGTACATCGATAAGTCTCTGAGCGGCGTGACC
GACCTGGACTTCACCCTGCTGTTCGAGCTGATGAACCTGGTGCAGAGCTCCCCCTCCAAGGACAATAAGAAGGCC
CTGGAGAAGGAGCAGTCTAAGATGAGGGAGCAGATCTGCACCCACCTGCAGTCCGACTCTAACTACAAGAATATC
TTTAACGCCAAGCTGCTGAAGGAGATCCTGCCTGATTTCATCAAGAACTACAATCAGTATGACGTGAAGGATAAG
GCCGGCAAGCTGGAGACACTGGCCCTGTTTAATGGCTTCAGCACATACTTTACCGACTTCTTTGAGAAGAGGAAG
AACGTGTTCACCAAGGAGGCCGTGAGCACATCCATCGCCTACCGCATCGTGCACGAGAACTCCCTGATCTTCCTG
GCCAATATGACCTCTTATAAGAAGATCAGCGAGAAGGCCCTGGATGAGATCGAAGTGATCGAGAAGAACAATCAG
GACAAGATGGGCGATTGGGAGCTGAATCAGATCTTTAACCCTGACTTCTACAATATGGTGCTGATCCAGTCCGGC
ATCGACTTCTACAACGAGATCTGCGGCGTGGTGAATGCCCACATGAACCTGTACTGTCAGCAGACCAAGAACAAT
TATAACCTGTTCAAGATGCGGAAGCTGCACAAGCAGATCCTGGCCTACACCAGCACCAGCTTCGAGGTGCCCAAG
ATGTTCGAGGACGATATGAGCGTGTATAACGCCGTGAACGCCTTCATCGACGAGACAGAGAAGGGCAACATCATC
GGCAAGCTGAAGGATATCGTGAATAAGTACGACGAGCTGGATGAGAAGAGAATCTATATCAGCAAGGACTTTTAC
GAGACACTGAGCTGCTTCATGTCCGGCAACTGGAATCTGATCACAGGCTGCGTGGAGAACTTCTACGATGAGAAC
ATCCACGCCAAGGGCAAGTCCAAGGAGGAGAAGGTGAAGAAGGCCGTGAAGGAGGACAAGTACAAGTCTATCAAT
GACGTGAACGATCTGGTGGAGAAGTATATCGATGAGAAGGAGAGGAATGAGTTCAAGAACAGCAATGCCAAGCAG
TACATCCGCGAGATCTCCAACATCATCACCGACACAGAGACAGCCCACCTGGAGTATGACGATCACATCTCTCTG
ATCGAGAGCGAGGAGAAGGCCGACGAGATGAAGAAGCGGCTGGATATGTATATGAACATGTACCACTGGGCCAAG
GCCTTTATCGTGGACGAGGTGCTGGACAGAGATGAGATGTTCTACAGCGATATCGACGATATCTATAATATCCTG
GAGAACATCGTGCCACTGTATAATCGGGTGAGAAACTACGTGACCCAGAAGCCCTACAACTCTAAGAAGATCAAG
CTGAATTTCCAGAGCCCTACACTGGCCAATGGCTGGTCCCAGTCTAAGGAGTTCGACAACAATGCCATCATCCTG
ATCAGAGATAACAAGTACTATCTGGCCATCTTCAATGCCAAGAACAAGCCAGACAAGAAGATCATCCAGGGCAAC
TCCGATAAGAAGAACGACAACGATTACAAGAAGATGGTGTATAACCTGCTGCCAGGCGCCAACAAGATGCTGCCC
AAGGTGTTTCTGTCTAAGAAGGGCATCGAGACATTCAAGCCCTCCGACTATATCATCTCTGGCTACAACGCCCAC
AAGCACATCAAGACAAGCGAGAATTTTGATATCTCCTTCTGTCGGGACCTGATCGATTACTTCAAGAACAGCATC
GAGAAGCACGCCGAGTGGAAGAAGTATGAGTTCAAGTTTTCCGCCACCGACAGCTACTCCGATATCTCTGAGTTC
TATCGGGAGGTGGAGATGCAGGGCTACAGAATCGACTGGACATATATCAGCGAGGCCGACATCAACAAGCTGGAT
GAGGAGGGCAAGATCTATCTGTTTCAGATCTACAATAAGGATTTCGCCGAGAACAGCACCGGCAAGGAGAATCTG
CACACAATGTACTTTAAGAACATCTTCTCCGAGGAGAATCTGAAGGACATCATCATCAAGCTGAACGGCCAGGCC
GAGCTGTTTTATCGGAGAGCCTCTGTGAAGAATCCCGTGAAGCACAAGAAGGATAGCGTGCTGGTGAACAAGACC
TACAAGAATCAGCTGGACAACGGCGACGTGGTGAGAATCCCCATCCCTGACGATATCTATAACGAGATCTACAAG
ATGTATAATGGCTACATCAAGGAGTCCGACCTGTCTGAGGCCGCCAAGGAGTACCTGGATAAGGTGGAGGTGAGG
ACCGCCCAGAAGGACATCGTGAAGGATTACCGCTATACAGTGGACAAGTACTTCATCCACACACCTATCACCATC
AACTATAAGGTGACCGCCCGCAACAATGTGAATGATATGGTGGTGAAGTACATCGCCCAGAACGACGATATCCAC
GTGATCGGCATCGACCGGGGCGAGAGAAACCTGATCTACATCTCCGTGATCGATTCTCACGGCAACATCGTGAAG
CAGAAATCCTACAACATCCTGAACAACTACGACTACAAGAAGAAGCTGGTGGAGAAGGAGAAAACCCGGGAGTAC
GCCAGAAAGAACTGGAAGAGCATCGGCAATATCAAGGAGCTGAAGGAGGGCTATATCTCCGGCGTGGTGCACGAG

圖 73F

ATCGCCATGCTGATCGTGGAGTACAACGCCATCATCGCCATGGAGGACCTGAATTATGGCTTTAAGAGGGGCCGC
TTCAAGGTGGAGCGGCAGGTGTACCAGAAGTTTGAGAGCATGCTGATCAATAAGCTGAACTATTTCGCCAGCAAG
GAGAAGTCCGTGGACGAGCCAGGAGGCCTGCTGAAGGGCTATCAGCTGACCTACGTGCCCGATAATATCAAGAAC
CTGGGCAAGCAGTGCGGCGTGATCTTTTACGTGCCTGCCGCCTTCACCAGCAAGATCGACCCATCCACAGGCTTT
ATCTCTGCCTTCAACTTTAAGTCTATCAGCACAAATGCCTCTCGGAAGCAGTTCTTTATGCAGTTTGACGAGATC
AGATACTGTGCCGAGAAGGATATGTTCAGCTTTGGCTTCGACTACAACAACTTCGATACCTACAACATCACAATG
GGCAAGACACAGTGGACCGTGTATACAAACGGCGAGAGACTGCAGTCTGAGTTCAACAATGCCAGGCGCACCGGC
AAGACAAAGAGCATCAATCTGACAGAGACAATCAAGCTGCTGCTGGAGGACAATGAGATCAACTACGCCGACGGC
CACGATATCAGGATCGATATGGAGAAGATGGACGAGGATAAGAAGAGCGAGTTCTTTGCCCAGCTGCTGAGCCTG
TATAAGCTGACCGTGCAGATGCGCAATTCCTATACAGAGGCCGAGGAGCAGGAGAACGGCATCTCTTACGACAAG
ATCATCAGCCCTGTGATCAATGATGAGGGCGAGTTCTTTGACTCCGATAACTATAAGGAGTCTGACGATAAGGAG
TGCAAGATGCCAAAGGACGCCGATGCCAACGGCGCCTACTGTATCGCCCTGAAGGGCCTGTATGAGGTGCTGAAG
ATCAAGAGCGAGTGGACCGAGGACGGCTTTGATAGGAATTGCCTGAAGCTGCCACACGCAGAGTGGCTGGACTTC
ATCCAGAACAAGCGGTACGAG
AAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCCTACCCATACGATGTTCCAGAT
TACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCCTAAGAATTC

圖 73G

MNGNRSIVYREFVGVIPVAKTLRNELRPVGHTQEHIIQNGLIQEDELRQEKSTELKNIMD
DYYREYIDKSLSGVTDLDFTLLFELMNLVQSSPSKDNKKALEKEQSKMREQICTHLQSDS
NYKNIFNAKLLKEILPDFIKNYNQYDVKDKAGKLETLALFNGFSTYFTDFFEKRKNVFTK
EAVSTSIAYRIVHENSLIFLANMTSYKKISEKALDEIEVIEKNNQDKMGDWELNQIFNPD
FYNMVLIQSGIDFYNEICGVVNAHMNLYCQQTKNNYNLFKMRKLHKQILAYTSTSFEVPK
MFEDDMSVYNAVNAFIDETEKGNIIGKLKDIVNKYDELDEKRIYISKDFYETLSCFMSGN
WNLITGCVENFYDENIHAKGKSKEEKVKKAVKEDKYKSINDVNDLVEKYIDEKERNEFKN
SNAKQYIREISNIITDTETAHLEYDDHISLIESEEKADEMKKRLDMYMNMYHWAKAFIVD
EVLDRDEMFYSDIDDIYNILENIVPLYNRVRNYVTQKPYNSKKIKLNFQSPTLANGWSQS
KEFDNNAIILIRDNKYYLAIFNAKNKPDKKIIQGNSDKKNDNDYKKMVYNLLPGANKMLP
KVFLSKKGIETFKPSDYIISGYNAHKHIKTSENFDISFCRDLIDYFKNSIEKHAEWRKYE
FKFSATDSYSDISEFYREVEMQGYRIDWTYISEADINKLDEEGKIYLFQIYNKDFAENST
GKENLHTMYFKNIFSEENLKDIIIKLNGQAELFYRRASVKNPVKHKKDSVLVNKTYKNQL
DNGDVVRIPIPDDIYNEIYKMYNGYIKESDLSEAAKEYLDKVEVRTAQKDIVKDYRYTVD
KYFIHTPITINYKVTARNNVNDMVVKYIAQNDDIHVIGIDRGERNLIYISVIDSHGNIVK
QKSYNILNNYDYKKKLVEKEKTREYARKNWKSIGNIKELKEGYISGVVHEIAMLIVEYNA
IIAMEDLNYGFKRGRFKVERQVYQKFESMLINKLNYFASKEKSVDEPGGLLKGYQLTYVP
DNIKNLGKQCGVIFYVPAAFTSKIDPSTGFISAFNFKSISTNASRKQFFMQFDEIRYCAE
KDMFSFGFDYNNFDTYNITMGKTQWTVYTNGERLQSEFNNARRTGKTKSINLTETIKLLL
EDNEINYADGHDIRIDMEKMDEDKKSEFFAQLLSLYKLTVQMRNSYTEAEEQENGISYDK
IISPVINDEGEFFDSDNYKESDDKECKMPKDADANGAYCIALKGLYEVLKIKSEWTEDGF
DRNCLKLPHAEWLDFIQNKRYE

圖 73H

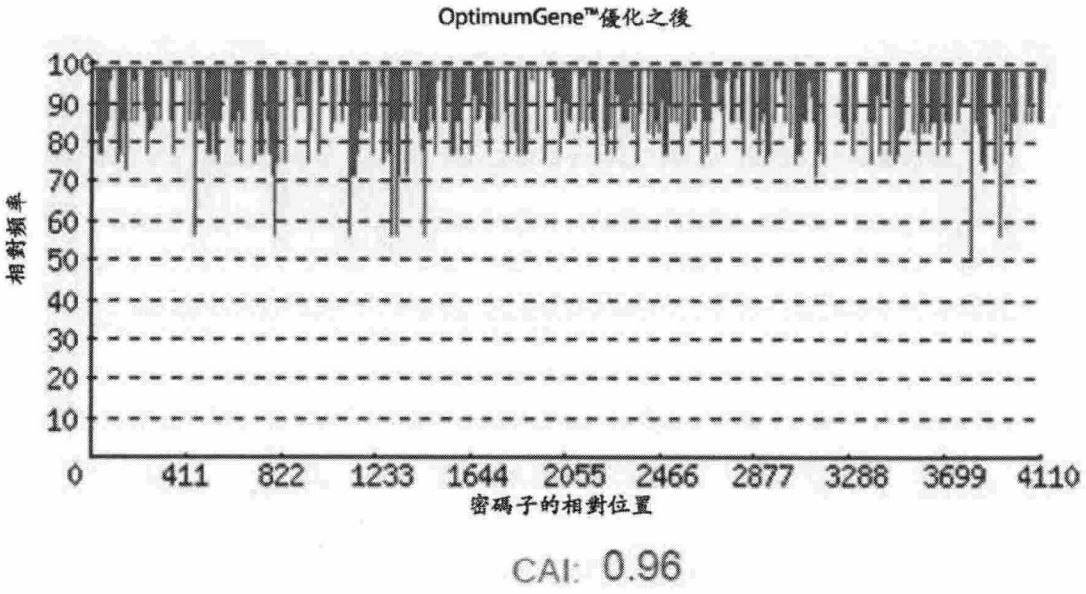

OptimumGene™優化之後
相對頻率
100
90
80
70
60
50
40
30
20
10
0
411
822
1233
1644
2055
2466
2877
3288
3699
4110
密碼子的相對位置
CAI: 0.96

圖 74A

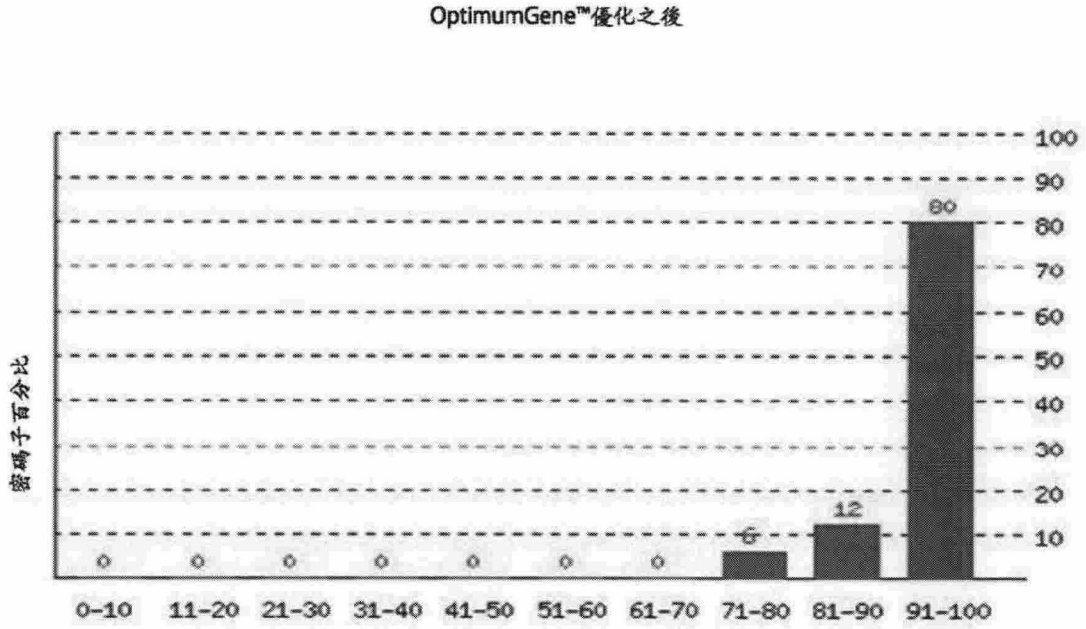
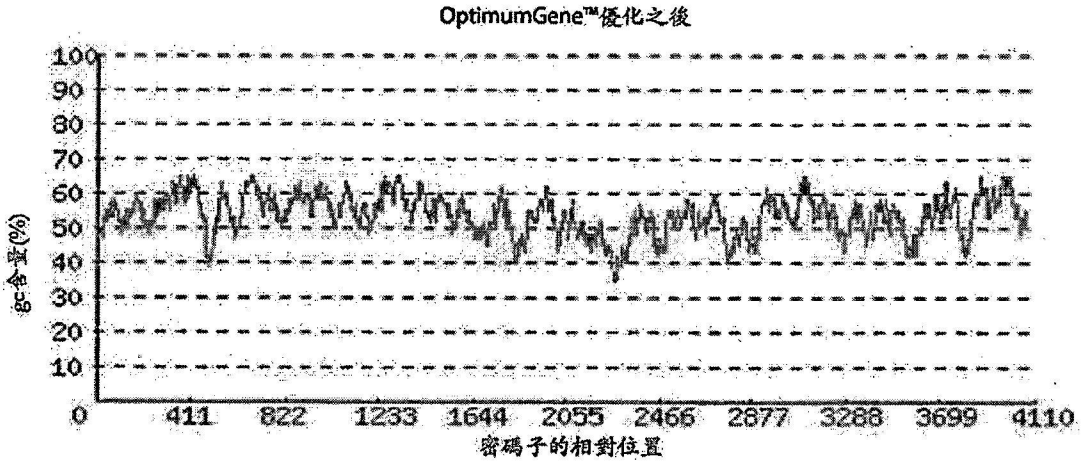

OptimumGene™優化之後
密碼子百分比
100
90
80
70
60
50
40
30
20
10
0
0
0
0
0
0
0
6
12
80
0-10
11-20
21-30
31-40
41-50
51-60
61-70
71-80
81-90
91-100

圖 74B



平均GC含量：53.58

圖 74C

限制性內切酶	優化的
綠色：過濾位點；藍色：檢查位點(未過濾的)；紅色：保留位點	
BamHI(GGATCC)	1(4174)
EcoRI(GAATTC)	1(4264)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(4238)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

順式作用元件	優化的
剪接(GGTAAG)	0
剪接(GGTGAT)	0
PolyA(AATAAA)	0
PolyA(ATTAAA)	0
去穩定化(ATTTA)	0
PolyT(TTTTTT)	0
PolyA(AAAAAAA)	0

抗病毒模體	優化的
	0

圖 74D

優化之後

最大同向重複序列：大小：14 距離：525 頻率：2

最大反向重複序列：無

最大Dyad重複序列：無

圖 74E

ATGCTGTTCCAGGACTTTACCCACCTGTATCCACTGTCCAAGACAGTGAGATTTGAGCTGAAGCCCATCGATACG
ACCCTGGAGCACATCCACGCCAAGAACTTCCTGTCTCAGGACGAGACAATGGCCGATATGCACCAGAAGGTGAAA
GTGATCCTGGACGATTACCACCGCGACTTCATCGCCGATATGATGGCGAGGTGAAGCTGACCAAGCTGGCCGAG
TTCTATGACGTGTACCTGAAGTTTCGGAAGAACCCAAAGGACGATGAGCTGCAGAAGCAGCTGAAGGATCTGCAG
GCCGTGCTGAGAAAGGAGATCGTGAAGCCCATCGGCAATGGCGGCAAGTATAAGGCCGGCTACGACAGGCTGTTC
GGCGCCAAGCTGTTTAAGGACGGCAAGGAGCTGGGCGATCTGGCCAAGTTCGTGATCGCACAGGAGGGAGAGAGC
TCCCCAAAGCTGGCCCACCTGGCCGACTTCGAGAAGTTTTCCACCTATTTCACAGGCTTTCACGATAACCGGAAG
AATATGTATTCTGACGAGGATAAGCACACCGCCATCGCCTACCGCCTGATCCACGAGAACCTGCCCCGGTTTATC
GACAATCTGCAGATCCTGACCACAATCAAGCAGAAGCACTCTGCCCTGTACGATCAGATCATCAACGAGCTGACC
GCCAGCGGCCTGGACGTGTCTCTGGCCAGCCACCTGGATGGCTATCACAAGCTGCTGACACAGGAGGGCATCACC
GCCTACAATACACTGCTGGGAGGAATCTCCGGAGAGGCAGGCTCTCCTAAGATCCAGGGCATCAACGAGCTGATC
AATTCTCACCACAACCAGCACTGCCACAAGAGCGAGAGGAATCGCCAAGCTGAGGCCACTGCACAAGCAGATCCTG
TCCGACGGCATGAGCGTGTCCTTCCTGCCCTCTAAGTTTGCCGACGATAGCGAGATGTGCCAGGCCGTGAACGAG
TTCTATCGCCACTACGCCGACGTGTTCGCCAAGGTGCAGAGCCTGTTCGACGCTTTGACGATCACCAGAAGGAT
GGCATCTACGTGGAGCACAACAACCTGAATGAGCTGTCCAAGCAGGCCTTCGGCGACTTTGCACTGCTGGGACGC
GTGCTGGACGGATACTATGTGGATGTGGTGAATCCAGAGTTCAACGAGCGGTTTGCCAAGGCCAAGACCGACAAT
GCCAAGGCCAAGCTGACAAAGGAGAAGGATAAGTTCATCAAGGGCGTGCACTCCCTGGCCTCTCTGGAGCAGGCC
ATCGAGCACTATACCGCCAAGGCACGACGATGAGAGCGTGCAGGCAGGCAAGCTGGGACAGTACTTCAAGCACGGC
CTGGCCGGAGTGGACAACCCCATCCAGAAGATCCACAACAATCACAGCACCATCAAGGGCTTTCTGGAGAGGGAG
CGCCCTGCAGGAGAGAGAGCCCTGCCAAAGATCAAGTCCGGCAAGAATCCTGAGATGACACAGCTGAGGCAGCTG
AAGGAGCTGCTGGATAACGCCCTGAATGTGGCCCACTTGGCCAAGCTGCTGACCACAAAGACCAGCACTGGACAAT
CAGGATGGCAACTTCTATGGCGAGTTTGGCGTGCTGTACGACGAGCTGGCCAAGATCCCCACCCTGTATAACAAG
GTGAGAGATTACCTGAGCCAGAAGCCTTTCTCCACCGAGAAGTACAAGCTGAACTTTGGCAATCCAACACTGCTG
AATGGCTGGGACCTGAACAAGGAGAAGGATAATTTCGGCGTGATCCTGCAGAAGGACGGCTGCTACTATCTGGCC
CTGCTGGACAAGGCCCACAAGAAGGTGTTTGATAACGCCCCTAATACAGGCAAGAGCATCTATCAGAAGATGATC
TATAAGTACCTGGAGGTGAGGAAGCAGTTCCCCAAGGTGTTCTTTTCCAAGGAGGCCATCGCCATCAACTACCAC
CCTTCTAAGGAGCTGGTGGAGATCAAGGACAAGGGCCGGCAGAGATCCGACGATGAGCGCCTGAAGCTGTATCGG
TTTATCCTGGAGTGTCTGAAGATCCACCCTAAGTACGATAAGAAGTTCGAGGGCGCCATCGGCGACATCCAGCTG
TTTAAGAAGGATAAGAAGGGCAGAGAGGTGCCAATCAGCGAGAAGGACCTGTTCGATAAGATCAACGGCATCTTT
TCTAGCAAGCCTAAGCTGGAGATGGAGGACTTCTTTATCGGCGAGTTCAAGAGGTATAACCCAAGCCAGGACCTG
GTGGATCAGTATAATATCTACAAGAAGATCGACTCCAACGATAATCGCAAGAAGGAGAATTTCTACAACAATCAC
CCCAAGTTTAAGAAGGATCTGGTGCGGTACTATTACGAGTCTATGTGCAAGCACGAGGAGTGGGAGGAGAGCTTC
GAGTTTTCCAAGAAGCTGCAGGACATCGGCTGTTACGTGGATGTGAACGAGCTGTTTACCGAGATCGAGACACGG
AGACTGAATTATAAGATCTCCTTCTGCAACATCAATGCCGACTACATCGATGAGCTGGTGGAGCAGGGCCAGCTG
TATCTGTTCCAGATCTACAACAAGGACTTTTCCCCAAAGGCCCACGGCAAGCCCAATCTGCACACCCTGTACTTC
AAGGCCCTGTTTTCTGAGGACAACCTGGCCGATCCTATCTATAAGCTGAATGGCGAGGCCGAGATCTTCTACAGA
AAGGCCTCCCTGGACATGAACGAGACAACAATCCACAGGGCCGGCGAGGTGCTGGAGAACAAGAATCCCGATAAT
CCTAAGAAGAGACAGTTCGTGTACGACATCATCAAGGATAAGAGGTACACACAGGACAAGTTCATGCTGCACGTG

圖 74F

CCAATCACCATGAACTTTGGCGTGCAGGGCATGACAATCAAGGAGTTCAATAAGAAGGTGAACCAGTCTATCCAG
CAGTATGACGAGGTGAACGTGATCGGCATCGATCGGGGCGAGAGACACCTGCTGTACCTGACCGTGATCAATAGC
AAGGGCGAGATCCTGGAGCAGTGTTCCCTGAACGACATCACCACAGCCTCTGCCAATGGCACACAGATGACCACA
CCTTACCACAAGATCCTGGATAAGAGGGAGATCGAGCGCCTGAACGCCCGGGTGGGATGGGGCGAGATCGAGACA
ATCAAGGAGCTGAAGTCTGGCTATCTGAGCCACGTGGTGCACCAGATCAGCCAGCTGATGCTGAAGTACAACGCC
ATCGTGGTGCTGGAGGACCTGAATTTCGGCTTTAAGAGGGGCCGCTTTAAGGTGGAGAAGCAGATCTATCAGAAC
TTCGAGAATGCCCTGATCAAGAAGCTGAACCACCTGGTGCTGAAGGACAAGGCCGACGATGAGATCGGCTCTTAC
AAGAATGCCCTGCAGCTGACCAACAATTTCACAGATCTGAAGAGCATCGGCAAGCAGACCGGCTTCCTGTTTTAT
GTGCCCGCCTGGAACACCTCTAAGATCGACCCTGAGACAGGCTTTGTGGATCTGCTGAAGCCAAGATACGAGAAC
ATCGCCCAGAGCCAGGCCTTCTTTGGCAAGTTCGACAAGATCTGCTATAATGCCGACAAGGATTACTTCGAGTTT
CACATCGACTACGCCAAGTTTACCGATAAGGCCAAGAATAGCCGCCAGATCTGGACAATCTGTTCCCACGGCGAC
AAGCGGTACGTGTACGATAAGACAGCCAACCAGAATAAGGGCGCCGCCAAGGGCATCAACGTGAATGATGAGCTG
AAGTCCCTGTTCGCCCGCCACCACATCAACGAGAAGCAGCCCAACCTGGTCATGGACATCTGCCAGAACAATGAT
AAGGAGTTTCACAAGTCTCTGATGTACCTGCTGAAAACCCTGCTGGCCCTGCGGTACAGCAACGCCTCCTCTGAC
GAGGATTTCATCCTGTCCCCCGTGGCAAACGACGAGGGCGTGTTCTTTAATAGCGCCCTGGCCGACGATACACAG
CCTCAGAATGCCGATGCCAACGGCGCCTACCACATCGCCCTGAAGGGCCTGTGGCTGCTGAATGAGCTGAAGAAC
TCCGACGATCTGAACAAGGTGAAGCTGGCCATCGACAATCAGACCTGGCTGAATTTCGCCCAGAACAGG
AAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCCTACCCATACGATGTTCCAGAT
TACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCCTACAAATTC

圖 74G

MLFQDFTHLYPLSKTVRFELKPIDRTLEHIHAKNFLSQDETMADMHQKVKVILDDYHRDF
IADMMGEVKLTKLAEFYDVYLKFRKNPKDDELQKQLKDLQAVLRKEIVKPIGNGGKYKAG
YDRLFGAKLFKDGKELGDLAKFVIAQEGESSPKLAHLAHFEKFSTYFTGFHDNRKNMYSD
EDKHTAIAYRLIHENLPRFIDNLQILTTIKQKHSALYDQIINELTASGLDVSLASHLDGY
HKLLTQEGITAYNTLLGGISGEAGSPKIQGINELINSHHNQHCHKSERIAKLRPLHKQIL
SDGMSVSFLPSKFADDSEMCQAVNEFYRHYADVFAKVQSLFDGFDDHQKDGIYVEHKNLN
ELSKQAFGDFALLGRVLDGYYVDVVNPEFNERFAKAKTDNAKAKLTKEKDKFIKGVHSLA
SLEQAIEHYTARHDDESVQAGKLGQYFKHGLAGVDNPIQKIHNNHSTIKGFLERERPAGE
RALPKIKSGKNPEMTQLRQLKELLDNALNVAHFAKLLTTKTTLDNQDGNFYGEFGVLYDE
LAKIPTLYNKVRDYLSQKPFSTEKYKLNFGNPTLLNGWDLNKEKDNFGVILQKDGCYYLA
LLDKAHKKVFDNAPNTGKSIYQKMIYKYLEVRKQFPKVFFSKEAIAINYHPSKELVEIKD
KGRQRSDDERLKLYRFILECLKIHPKYDKKFEGAIGDIQLFKKDKKGREVPISEKDLFDK
INGIFSSKPKLEMEDFFIGEFKRYNPSQDLVDQYNIYKKIDSNDNRKKENFYNNHPKFKK
DLVRYYYESMCKHEEWEESFEFSKKLQDIGCYVDVNELFTEIETRRLNYKISFCNINADY
IDELVEQGQLYLFQIYNKDFSPKAHGKPNLHTLYFKALFSEDNLADPIYKLNGEAQIFYR
KASLDMNETTIHRAGEVLENKNPDNPKKRQFVYDIIKDKRYTQDKFMLHVPITMNFGVQG
MTIKEFNKKVNQSIQQYDEVNVIGIDRGERHLLYLTVINSKGEILEQCSLNDITTASANG
TQMTTPYHKILDKREIERLNARVGWGEIETIKELKSGYLSHVVHQISQLMLKYNAIVVLE
DLNFGFKRGRFKVEKQIYQNFENALIKKLNHLVLKDKADDEIGSYKNALQLTNNFTDLKS
IGKQTGFLFYVPAWNTSKIDPETGFVDLLKPRYENIAQSQAFFGKFDKICYNADKDYFEF
HIDYAKFTDKAKNSRQIWTICSHGDKRYVYDKTANQNKGAAKGINVNDELKSLFARHHIN
EKQPNLVMDICQNNDKEFHKSLMYLLKTLLALRYSNASSDEDFILSPVANDEGVFFNSAL
ADDTQPQNADANGAYHIALKGLWLLNELKNSDDLNKVKLAIDNQTWLNFAQNR

圖 74H

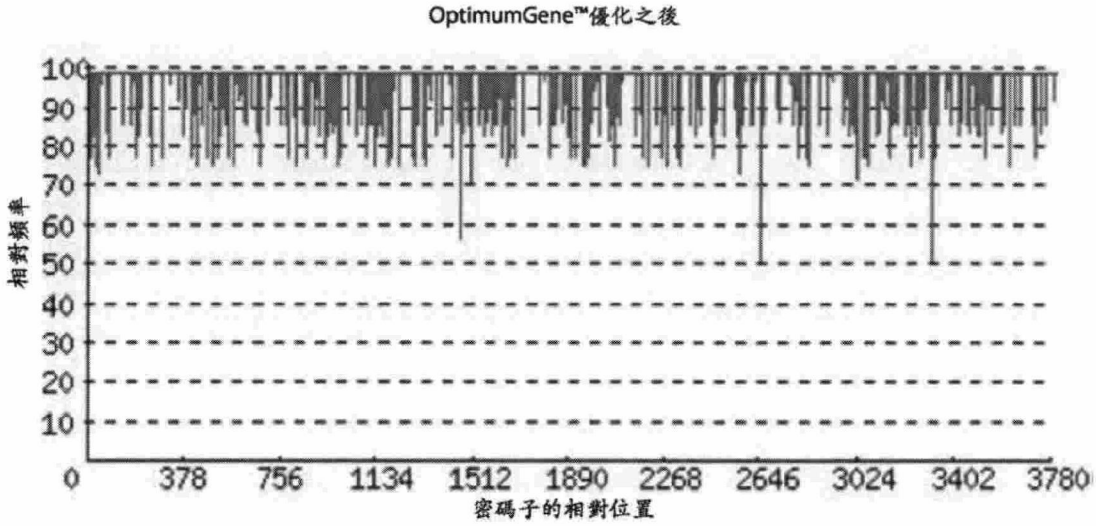

OptimumGene™優化之後
相對頻率
100
90
80
70
60
50
40
30
20
10
0
378
756
1134
1512
1890
2268
2646
3024
3402
3780
密碼子的相對位置
CAI: 0.96

圖 75A

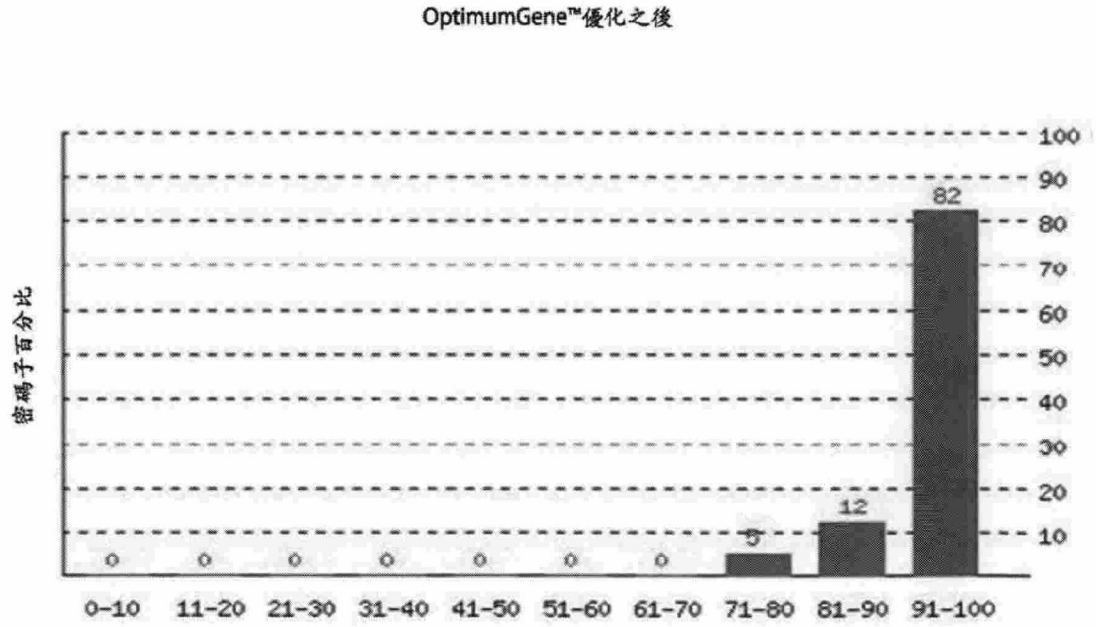

OptimumGene™優化之後
密碼子百分比
100
90
80
70
60
50
40
30
20
10
0
0
0
0
0
0
0
5
12
82
0-10
11-20
21-30
31-40
41-50
51-60
61-70
71-80
81-90
91-100

圖 75B

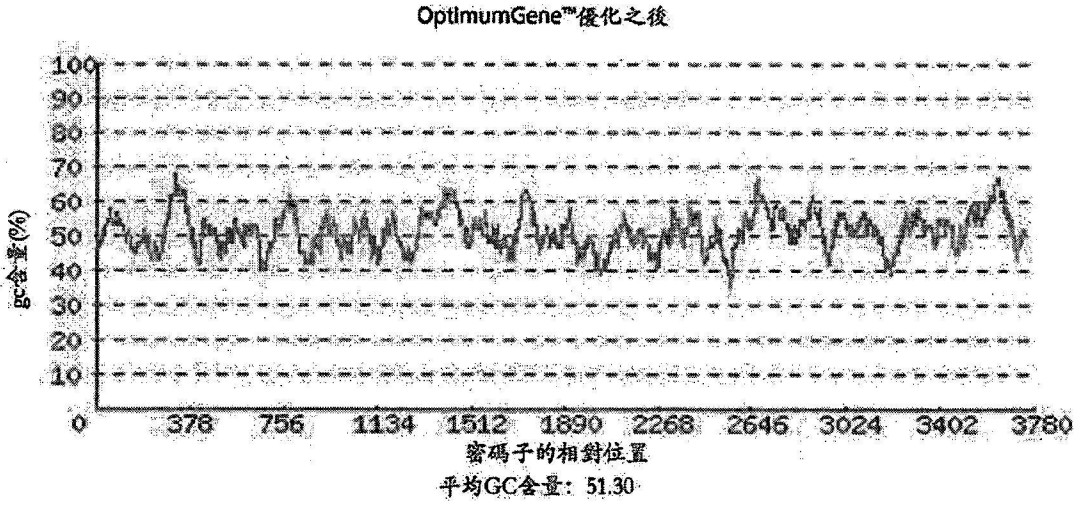

OptimumGene™優化之後
100
90
80
70
60
50
40
30
20
10
0
gc含量(%)
378
756
1134
1512
1890
2268
2646
3024
3402
3780
密碼子的相對位置
平均GC含量: 51.30

圖 75C

限制性內切酶	優化的
綠色：過濾位點；藍色：檢查位點(未過濾的)；紅色：保留位點	
BamHI(GGATCC)	1(3844)
EcoRI(GAATTC)	1(3934)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(3908)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

順式作用元件	優化的
剪接(GGTAAG)	0
剪接(GGTGAT)	0
PolyA(AATAAA)	0
PolyA(ATTAAA)	0
去穩定化(ATTTA)	0
PolyT(TTTTTT)	0
PolyA(AAAAAAA)	0

抗病毒模體	優化的
	0

圖 75D

優化之後

最大同向重複序列：大小：14 距離：111 頻率：2

最大反向重複序列：無

最大Dyad重複序列：大小：13 Tm：35.4 起始位置：1027,221

圖 75E

GGTACC

ATGGAGGACTATTCCGGCTTTGTGAACATCTACTCTATCCAGAAAACCCTGAGGTTCGAGCTGAAGCCAGTGGGC
AAGACACTGGAGCACATCGAGAAGAAGGGCTTCCTGAAGAAGGACAAGATCCGGGCCGAGGATTACAAGGCCGTG
AAGAAGATCATCGATAAGTACCACAGAGCCTATATCGAGGAGGTGTTTGATTCCGTGCTGCACCAGAAGAAGAAG
AAGGACAAGACCCGCTTTTCTACACAGTTCATCAAGGAGATCAAGGAGTTCAGCGAGCTGTACTATAAGACCGAG
AAGAACATCCCCGACAAGGAGAGGCTGGAGGCCCTGAGCGAGAAGCTGCGCAAGATGCTGGTGGGCGCCTTTAAG
GGCGAGTTCTCCGAGGAGGTGGCCGAGAAGTATAAGAACCTGTTTTCTAAGGAGCTGATCAGGAATGAGATCGAG
AAGTTCTGCGAGACAGACGAGGAGCGCAAGCAGGTGTCTAACTTCAAGAGCTTCACCACATACTTTACCGGCTTC
CACTCCAACAGGCAGAATATCTATTCCGACGAGAAGAAGTCTACAGCCATCGGCTACCGCATCATCCACCAGAAC
CTGCCTAAGTTCCTGGATAATCTGAAGATCATCGAGTCCATCCAGCGGCGGTTCAAGGACTTCCCATGGTCTGAT
CTGAAGAAGAACCTGAAGAAGATCGATAAGAATATCAAGCTGACCGAGTACTTCAGCATCGACGGCTTCGTGAAC
GTGCTGAATCAGAAGGGCATCGATGCCTACAACACAATCCTGGGCGGCAAGTCCGAGGAGTCTGGCGAGAAGATC
CAGGGCCTGAACGAGTACATCAATCTGTATCGGCAGAAGAACAATATCGACAGAAAGAACCTGCCCAATGTGAAG
ATCCTGTTTAAGCAGATCCTGGGCGATAGGGAGACAAAGAGCTTTATCCCTGAGGCCTTCCCAGACGATCAGTCC
GTGCTGAACTCTATCACAGAGTTCGCCAAGTACCTGAAGCTGGATAAGAAGAAGAAGAGCATCATCGCCGAGCTG
AAGAAGTTTCTGAGCTCCTTCAATCGCTACGAGCTGGACGGCATCTATCTGGCCAACGATAATAGCCTGGCCTCT
ATCAGCACCTTCCTGTTTGACGATTGGTCCTTTATCAAGAAGTCCGTGTCTTTCAAGTATGACGAGTCCGTGGGC
GACCCCAAGAAGAAGATCAAGTCTCCCCTGAAGTACGAGAAGGAGAAGGAGAAGTGGCTGAAGCAGAAGTACTAT
ACAATCTCTTTCCTGAACGATGCCATCGAGAGCTATTCCAAGTCTCAGGACGAGAAGAGGGTGAAGATCCGCCTG
GAGGCCTACTTTGCCGAGTTCAAGAGCAAGGACGATGCCAAGAAGCAGTTCGACCTGCTGGAGAGGATCGAGGAG
GCCTATGCCATCGTGGAGCCTCTGCTGGGAGCAGAGTACCCAAGGGACCGCAACCTGAAGGCCGATAAGAAGGAA
GTGGGCAAGATCAAGGACTTCCTGGATAGCATCAAGTCCCTGCACTTCTTTCTGAAGCCTCTGCTGTCCGCCGAG
ATCTTTGACGAGAAGGATCTGGGCTTCTACAATCAGCTGGAGGGCTACTATGAGGAGATCGATTCTATCGGCCAC
CTGTATAACAAGGTGCGGAATTATCTGACCGGCAAGATCTACAGCAAGGAGAAGTTTAAGCTGAACTTCGAGAAC
AGCACCCTGCTGAAGGGCTGGGACGAGAACCGGGAGGTGGCCAATCTGTGCGTGATCTTCAGAGAGGACCAGAAG
TACTATCTGGGCGTGATGGATAAGGAGAACAATACCATCCTGTCCGACATCCCCAAGGTGAAGCCTAACGAGCTG
TTTTACGAGAAGATGGTGTATAAGCTGATCCCCACACCTCACATGCAGCTGCCCCGGATCATCTTCTCTAGCGAC
AACCTGTCTATCTATAATCCTAGCAAGTCCATCCTGAAGATCAGAGACGCCAGAGCTTTAAGGAGGGCAAGAAC
TTCAAGCTGAAGGACTGTCACAAGTTTATCGATTCTACAAGGAGTCTATCAGCAAGAATGAGGACTGGAGCAGA
TTCGACTTCAAGTTCAGCAAGACCAGCAGCTACGAGAACATCAGCGAGTTTTACCGGGAGGTGGAGAGACAGGGC
TATAACCTGGACTTCAAGAAGGTGTCTAAGTTCTACATCGACAGCCTGGTGGAGGATGGCAAGCTGTACCTGTTC
CAGATCTATAACAAGGACTTTTCTATCTTCAGCAAGGGCAAGCCCAATCTGCACACCATCTATTTTCGGTCCCTG
TTCTCTAAGGAGAACCTGAAGGACGTGTGCCTGAAGCTGAATGGCGAGGCCGAGATGTTCTTTCGGAAGAAGTCC
ATCACTACGATGAGAAGAAGAAGCGGGAGGGCCACCACCCCGAGCTGTTTGAGAAGCTGAAGTATCCTATCCTG
AAGGACAAGAGATACAGCGAGGATAAGTTTCAGTTCCACCTGCCCATCAGCCTGAACTTCAAGTCCAAGGAGCGG
CTGAACTTTAATCTGAAAGTGAATGAGTTCCTGAAGAGAAACAACGACATCAATATCATCGGCATCGATCGGGGC
GAGACAAACCTGCTGTACCTGGTCATGATCAATCAGAAGGGCGAGATCCTGAAGCAGACCCTGCTGGACAGCATG
CAGTCCGGCAAGGGCCGGCCTGAGATCAACTACAAGGAGAAGCTGCAGGAGAAGGAGATCGAGAGGGATAAGGCC
CGCAAGAGCTGGGGCACAGTGGAGAATATCAAGGAGCTGAAGGAGGGCTATCTGTCTATCGTGATCCACCAGATC

圖 75F

AGCAAGCTGATGGTGGAGAACAATGCCATCGTGGTGCTGGAGGACCTGAACATCGGCTTTAAGCGGGGCAGACAG
AAGGTGGAGCGGCAGGTGTACCAGAAGTTCGAGAAGATGCTGATCGATAAGCTGAACTTTCTGGTGTTCAAGGAG
AATAAGCCAACCGAGCCAGGAGGCGTGCTGAAGGCCTATCAGCTGACAGACGAGTTTCAGTCTTTCGAGAAGCTG
AGCAAGCAGACCGGCTTTCTGTTCTACGTGCCAAGCTGGAACACCTCCAAGATCGACCCCAGAACAGGCTTTATC
GATTTCCTGCACCCTGCCTACGAGAATATCGAGAAGGCCAAGCAGTGGATCAACAAGTTTGATTCCATCAGGTTC
AATTCTAAGATGGACTGGTTTGAGTTCACCGCCGATACACGCAAGTTTTCCGAGAACCTGATGCTGGGCAAGAAT
CGGGTGTGGGTCATCTGCACCACAAATGTGGAGCGGTACTTCACCAGCAAGACCGCCAACAGCTCCATCCAGTAC
AATAGCATCCAGATCACCGAGAAGCTGAAGGAGCTGTTTGTGGACATCCCTTTCAGCAACGGCCAGGATCTGAAG
CCAGAGATCCTGAGGAAGAATGACGCCGTGTTCTTTAAGAGCCTGCTGTTTTACATCAAGACCACACTGTCCCTG
CGCCAGAACAATGGCAAGAAGGGCGAGGAGGAGAAGGACTTCATCCTGAGCCCAGTGGTGGATTCCAAGGGCCGG
TTCTTTAACTCTCTGGAGGCCAGCGACGATGAGCCCAAGGACGCCGATGCCAATGGCGCCTACCACATCGCCCTG
AAGGGCCTGATGAACCTGCTGGTGCTGAATGAGACAAAGGAGGAGAACCTGAGCAGACCAAAGTGGAAGATCAAG
AATAAGGACTGGCTGGAGTTCGTGTGGGAGAGGAACCGC

圖 75G

MEDYSGFVNIYSIQKTLRFELKPVGKTLEHIEKKGFLKKDKIRAEDYKAVKKIIDKYHRA
YIEEVFDSVLHQKKKKDKTRFSTQFIKEIKEFSELYYKTEKNIPDKERLEALSEKLRKML
VGAFKGEFSEEVAEKYKNLFSKELIRNEIEKFCETDEERKQVSNFKSFTTYFTGFHSNRQ
NIYSDEKKSTAIGYRIIHQNLPKFLDNLKIIESIQRRFKDFPWSDLKKNLKKIDKNIKLT
EYFSIDGFVNVLNQKGIDAYNTILGGKSEESGEKIQGLNEYINLYRQKNNIDRKNLPNVK
ILFKQILGDRETKSFIPEAFPDDQSVLNSITEFAKYLKLDKKKKSIIAELKKFLSSFNRY
ELDGIYLANDNSLASISTFLFDDWSFIKKSVSFKYDESVGDPKKKIKSPLKYEKEKEKWL
KQKYYTISFLNDAIESYSKSQDEKRVKIRLEAYFAEFKSKDDAKKQFDLLERIEEAYAIV
EPLLGAEYPRDRNLKADKKEVGKIKDFLDSIKSLQFFLKPLLSAEIFDEKDLGFYNQLEG
YYEEIDSIGHLYNKVRNYLTGKIYSKEKFKLNFENSTLLKGWDENREVANLCVIFREDQK
YYLGVMDKENNTILSDIPKVKPNELFYEKMVYKLIPTPHMQLPRIIFSSDNLSIYNPSKS
ILKIREAKSFKEGKNFKLKDCHKFIDFYKESISKNEDWSRFDFKFSKTSSYENISEFYRE
VERQGYNLDFKKVSKFYIDSLVEDGKLYLFQIYNKDFSIFSKGKPNLHTIYFRSLFSKEN
LKDVCLKLNGEAEMFFRKKSINYDEKKKREGHHPELFEKLKYPILKDKRYSEDKFQFHLP
ISLNFKSKERLNFNLKVNEFLKRNKDINIIGIDRGERNLLYLVMINQKGEILKQTLLDSM
QSGKGRPEINYKEKLQEKEIERDKARKSWGTVENIKELKEGYLSIVIHQISKLMVENNAI
VVLEDLNIGFKRGRQKVERQVYQKFEKMLIDKLNFLVFKENKPTEPGGVLKAYQLTDEFQ
SFEKLSKQTGFLFYVPSWNTSKIDPRTGFIDFLHPAYENIEKAKQWINKFDSIRFNSKMD
WFEFTADTRKFSENLMLGKNRVWVICTTNVERYFTSKTANSSIQYNSIQITEKLKELFVD
IPFSNGQDLKPEILRKNDAVFFKSLLFYIKTTLSLRQNNGKKGEEEKDFILSPVVDSKGR
FFNSLEASDDEPKDADANGAYHIALKGLMNLLVLNETKEENLSRPKWKIKNKDWLEFVWE
RNR

圖 75H

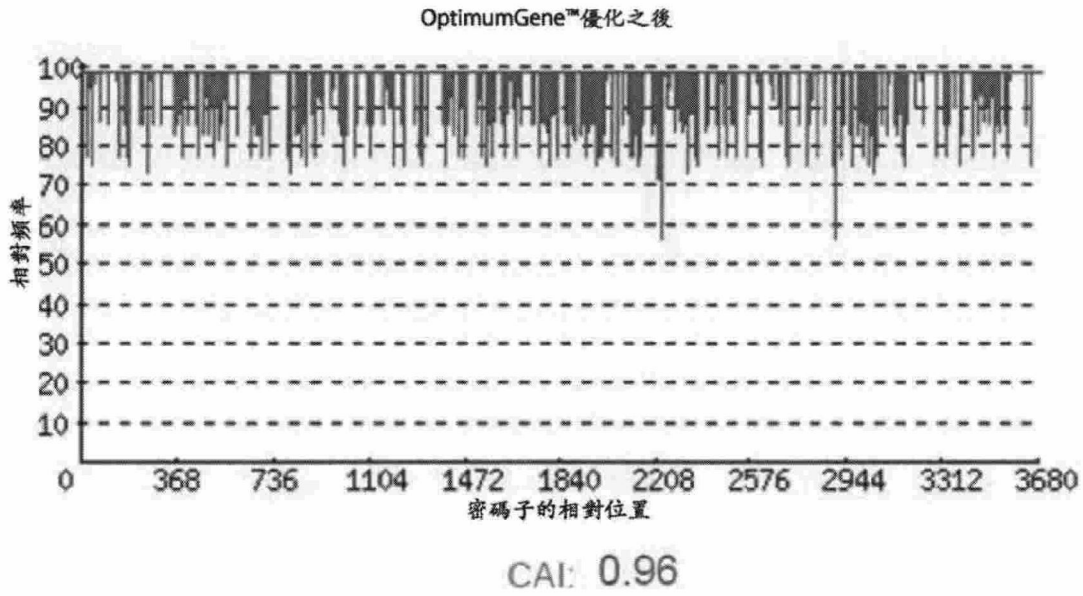

OptimumGene™優化之後
100
90
80
70
60
50
40
30
20
10
相對頻率
0
368
736
1104
1472
1840
2208
2576
2944
3312
3680
密碼子的相對位置
CAI: 0.96

圖 76A

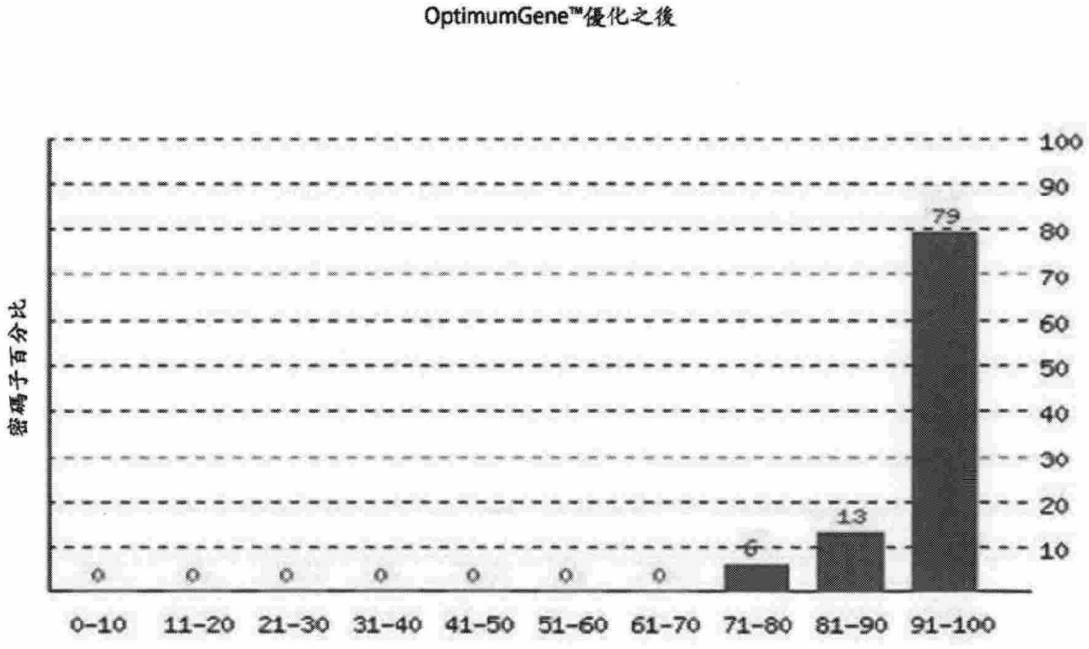

OptimumGene™優化之後
密碼子百分比
100
90
80
70
60
50
40
30
20
10
79
13
6
0
0
0
0
0
0
0
0-10
11-20
21-30
31-40
41-50
51-60
61-70
71-80
81-90
91-100

圖 76B

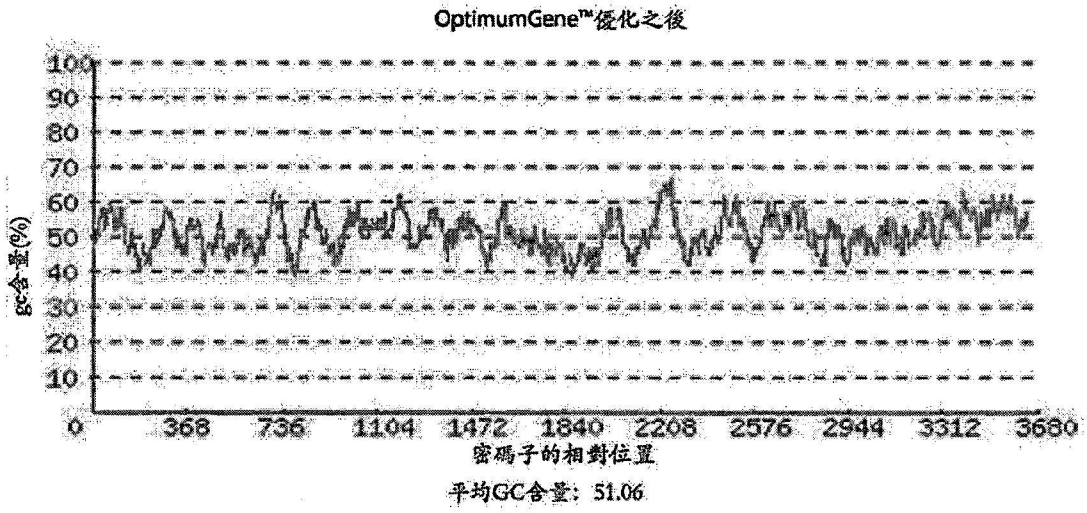

OptimumGene™優化之後
100
90
80
70
60
50
40
30
20
10
0
gc含量(%)
368
736
1104
1472
1840
2208
2576
2944
3312
3680
密碼子的相對位置
平均GC含量：51.06

圖 76C

限制性內切酶	優化的
*綠色：過濾位點；藍色：檢查位點(未過濾的)；紅色：保留位點	
BamHI(GGATCC)	1(3739)
EcoRI(GAATTC)	1(3829)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(3803)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

順式作用元件	優化的
剪接(GGTAAG)	0
剪接(GGTGAT)	0
PolyA(AATAAA)	0
PolyA(ATTAAA)	0
去穩定化(ATTTA)	0
PolyT(TTTTTT)	0
PolyA(AAAAAAA)	0

抗病毒模體	優化的
	0

圖 76D

優化之後

最大同向重複序列：大小：14 距離：150 頻率：2

最大反向重複序列：無

最大Dyad重複序列：無

圖 76E

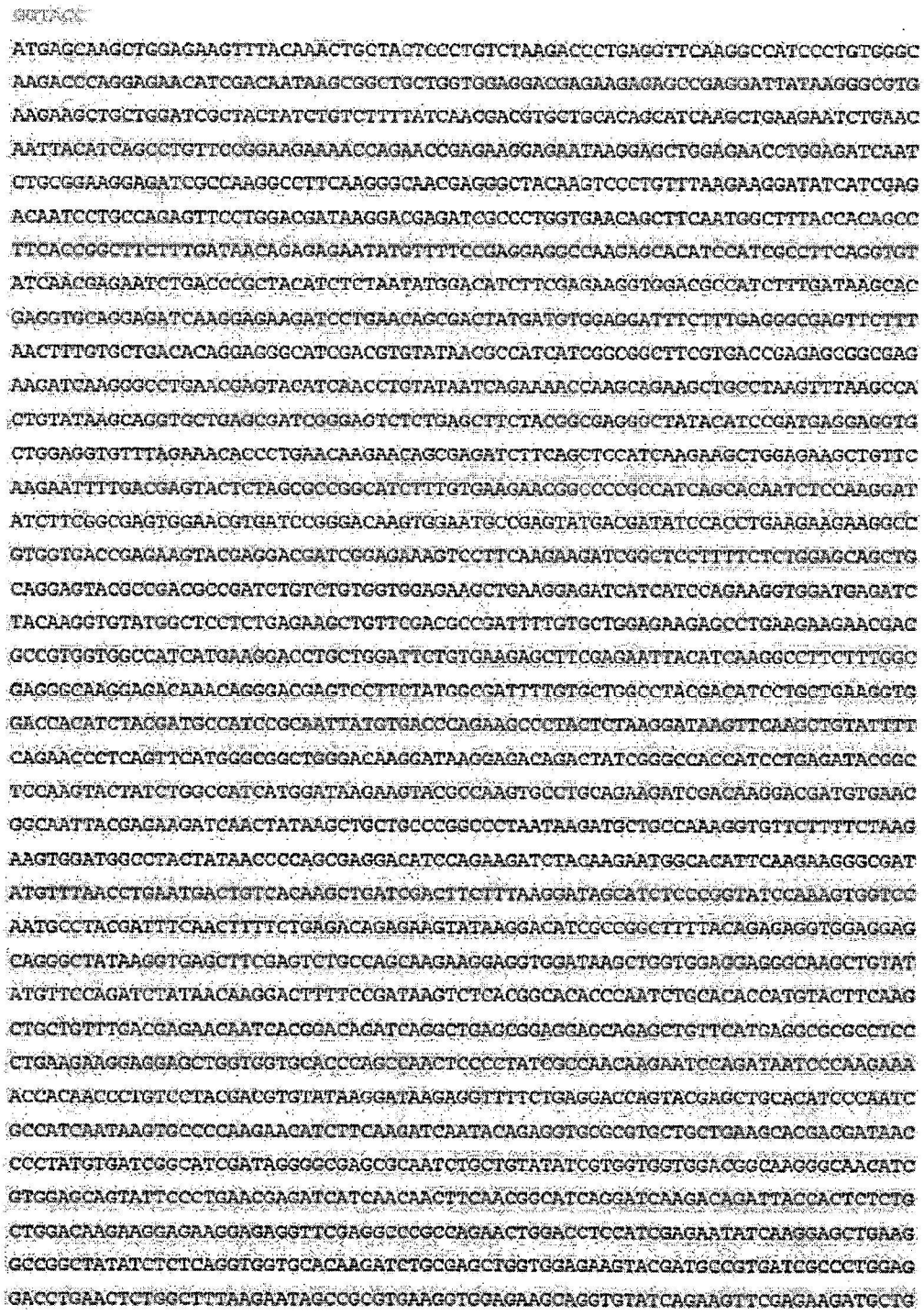

ATGAGCAAGCTGGAGAAGTTTACAAACTGCTACTCCCTGTCTAAGACCCTGAGGTTCAAGGCCATCCCTGTGGGC
AAGACCCAGGAGAACATCGACAATAAGCGGCTGCTGGTGGAGGACGAGAAGAGAGCCGAGGATTATAAGGGCGTG
AAGAAGCTGCTGGATCGCTACTATCTGTCTTTTATCAACGACGTGCTGCACAGCATCAAGCTGAAGAATCTGAAC
AATTACATCAGCCTGTTCCGGAAGAAAACCAGAACCGAGAAGGAGAATAAGGAGCTGGAGAACCTGGAGATCAAT
CTGCGGAAGGAGATCGCCAAGGCCTTCAAGGGCAACGAGGGCTACAAGTCCCTGTTTAAGAAGGATATCATCGAG
ACAATCCTGCCAGAGTTCCTGGACGATAAGGACGAGATCGCCCTGGTGAACAGCTTCAATGGCTTTACCACAGCC
TTCACCGGCTTCTTTGATAACAGAGAGAATATGTTTTCCGAGGAGGCCAAGAGCACATCCATCGCCTTCAGGTGT
ATCAACGAGAATCTGACCCGCTACATCTCTAATATGGACATCTTCGAGAAGGTGGACGCCATCTTTGATAAGCAC
GAGGTGCAGGAGATCAAGGAGAAGATCCTGAACAGCGACTATGATGTGGAGGATTTCTTTGAGGGCGAGTTCTTT
AACTTTGTGCTGACACAGGAGGGCATCGACGTGTATAACGCCATCATCGGCGGCTTCGTGACCGAGAGCGGCGAG
AAGATCAAGGGCCTGAACGAGTACATCAACCTGTATAATCAGAAAACCAAGCAGAAGCTGCCTAAGTTTAAGCCA
CTGTATAAGCAGGTGCTGAGCGATCGGGAGTCTCTGAGCTTCTACGGCGAGGGCTATACATCCGATGAGGAGGTG
CTGGAGGTGTTTAGAAACACCCTGAACAAGAACAGCGAGATCTTCAGCTCCATCAAGAAGCTGGAGAAGCTGTTC
AAGAATTTTGACGAGTACTCTAGCGCCGGCATCTTTGTGAAGAACGGCCCCGCCATCAGCACAATCTCCAAGGAT
ATCTTCGGCGAGTGGAACGTGATCCGGGACAAGTGGAATGCCGAGTATGACGATATCCACCTGAAGAAGAAGGCC
GTGGTGACCGAGAAGTACGAGGACGATCGGAGAAAGTCCTTCAAGAAGATCGGCTCCTTTTCTCTGGAGCAGCTG
CAGGAGTACGCCGACGCCGATCTGTCTGTGGTGGAGAAGCTGAAGGAGATCATCATCCAGAAGGTGGATGAGATC
TACAAGGTGTATGGCTCCTCTGAGAAGCTGTTCGACGCCGATTTTGTGCTGGAGAAGAGCCTGAAGAAGAACGAC
GCCGTGGTGGCCATCATGAAGGACCTGCTGGATTCTGTGAAGAGCTTCGAGAATTACATCAAGGCCTTCTTTGGC
GAGGGCAAGGAGACAAACAGGGACGAGTCCTTCTATGGCGATTTTGTGCTGGCCTACGACATCCTGCTGAAGGTG
GACCACATCTACGATGCCATCCGCAATTATGTGACCCAGAAGCCCTACTCTAAGGATAAGTTCAAGCTGTATTTT
CAGAACCCTCAGTTCATGGGCGGCTGGGACAAGGATAAGGAGACAGACTATCGGGCCACCATCCTGAGATACGGC
TCCAAGTACTATCTGGCCATCATGGATAAGAAGTACGCCAAGTGCCTGCAGAAGATCGACAAGGACGATGTGAAC
GGCAATTACGAGAAGATCAACTATAAGCTGCTGCCCGGCCCTAATAAGATGCTGCCAAAGGTGTTCTTTTCTAAG
AAGTGGATGGCCTACTATAACCCCAGCGAGGACATCCAGAAGATCTACAAGAATGGCACATTCAAGAAGGGCGAT
ATGTTTAACCTGAATGACTGTCACAAGCTGATCGACTTCTTTAAGGATAGCATCTCCCGGTATCCAAAGTGGTCC
AATGCCTACGATTTCAACTTTTCTGAGACAGAGAAGTATAAGGACATCGCCGGCTTTTACAGAGAGGTGGAGGAG
CAGGGCTATAAGGTGAGCTTCGAGTCTGCCAGCAAGAAGGAGGTGGATAAGCTGGTGGAGGAGGGCAAGCTGTAT
ATGTTCCAGATCTATAACAAGGACTTTTCCGATAAGTCTCACGGCACACCCAATCTGCACACCATGTACTTCAAG
CTGCTGTTTGACGAGAACAATCACGGACAGATCAGGCTGAGCGGAGGAGCAGAGCTGTTCATGAGGCGCGCCTCC
CTGAAGAAGGAGGAGCTGGTGGTGCACCCAGCCAACTCCCCTATCGCCAACAAGAATCCAGATAATCCCAAGAAA
ACCACAACCCTGTCCTACGACGTGTATAAGGATAAGAGGTTTTCTGAGGACCAGTACGAGCTGCACATCCCAATC
GCCATCAATAAGTGCCCCAAGAACATCTTCAAGATCAATACAGAGGTGCGCGTGCTGCTGAAGCACGACGATAAC
CCCTATGTGATCGGCATCGATAGGGGCGAGCGCAATCTGCTGTATATCGTGGTGGTGGACGGCAAGGGCAACATC
GTGGAGCAGTATTCCCTGAACGAGATCATCAACAACTTCAACGGCATCAGGATCAAGACAGATTACCACTCTCTG
CTGGACAAGAAGGAGAAGGAGAGGTTCGAGGCCCGCCAGAACTGGACCTCCATCGAGAATATCAAGGAGCTGAAG
GCCGGCTATATCTCTCAGGTGGTGCACAAGATCTGCGAGCTGGTGGAGAAGTACGATGCCGTGATCGCCCTGGAG
GACCTGAACTCTGGCTTTAAGAATAGCCGCGTGAAGGTGGAGAAGCAGGTGTATCAGAAGTTCGAGAAGATGCTG

圖 76F

ATCGATAAGCTGAACTACATGGTGGACAAGAAGTCTAATCCTTGTGCAACAGGCGGCGCCCTGAAGGGCTATCAG
ATCACCAATAAGTTCGAGAGCTTTAAGTCCATGTCTACCCAGAACGGCTTCATCTTTTACATCCCTGCCTGGCTG
ACATCCAAGATCGATCCATCTACCGGCTTTGTGAACCTGCTGAAAACCAAGTATACCAGCATCGCCGATTCCAAG
AAGTTCATCAGCTCCTTTGACAGGATCATGTACGTGCCCGAGGAGGATCTGTTCGAGTTTGCCCTGGACTATAAG
AACTTCTCTCGCACAGACGCCGATTACATCAAGAAGTGGAAGCTGTACTCCTACGGCAACCGGATCAGAATCTTC
CGGAATCCTAAGAAGAACAACGTGTTCGACTGGGAGGAGGTGTGCCTGACCAGCGCCTATAAGGAGCTGTTCAAC
AAGTACGGCATCAATTATCAGCAGGGCGATATCAGAGCCCTGCTGTGCGAGCAGTCCGACAAGGCCTTCTACTCT
AGCTTTATGGCCCTGATGAGCCTGATGCTGCAGATGCGGAACAGCATCACAGGCCGCACCGACGTGGATTTTCTG
ATCAGCCCTGTGAAGAACTCCGACGGCATCTTCTACGATAGCCGGAACTATGAGGCCCAGGAGAATGCCATCCTG
CCAAAGAACGCCGACGCCAATGGCGCCTATAACATCGCCAGAAAGGTGCTGTGGGCCATCGGCCAGTTCAAGAAG
GCCGAGGACGAGAAGCTGGATAAGGTGAAGATCGCCATCTCTAACAAGGAGTGGCTGGAGTACGCCCAGACCAGC
GTGAAGCAC

圖 76G

MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKKLLDRYYLS
FINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNEGYKSLFK
KDIIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRCINENL
TRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNAI
IGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRESLSFYGEGYTSDEEV
LEVFRNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRD
KWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSVVEKLKEIIIQ
KVDEIYKVYGSSEKLFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKET
NRDESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKET
DYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSK
KWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDSISRYPKWSNAYDFNFSET
EKYKDIAGFYREVEEQGYKVSFESASKKEVDKLVEEGKLYMFQIYNKDFSDKSHGTPNLH
TMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEELVVHPANSPIANKNPDNPKKTTTLS
YDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTEVRVLLKHDDNPYVIGIDRGERNLLY
IVVVDGKGNIVEQYSLNEIINNFNGIRIKTDYHSLLDKKEKERFEARQNWTSIENIKELK
AGYISQVVHKICELVEKYDAVIALEDLNSGFKNSRVKVEKQVYQKFEKMLIDKLNYMVDK
KSNPCATGGALKGYQITNKFESFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLKTKYTS
IADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKK
NNVFDWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNS
ITGRTDVDFLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIGQFKK
AEDEKLDKVKIAISNKEWLEYAQTSVKH

圖 76H

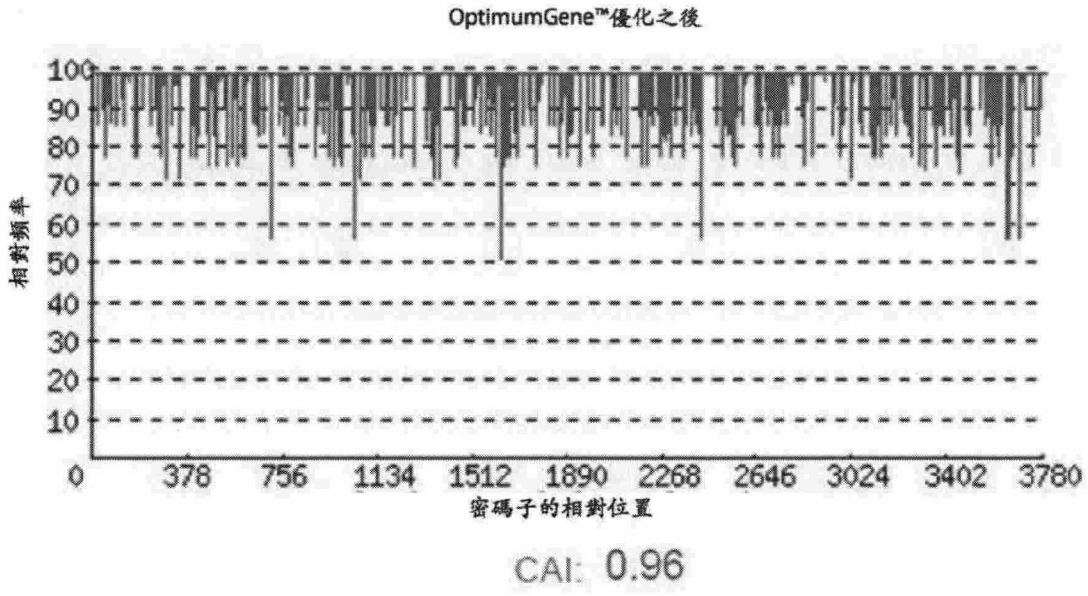

OptimumGene™優化之後
相對頻率
100
90
80
70
60
50
40
30
20
10
0
378
756
1134
1512
1890
2268
2646
3024
3402
3780
密碼子的相對位置
CAI: 0.96

圖 77A

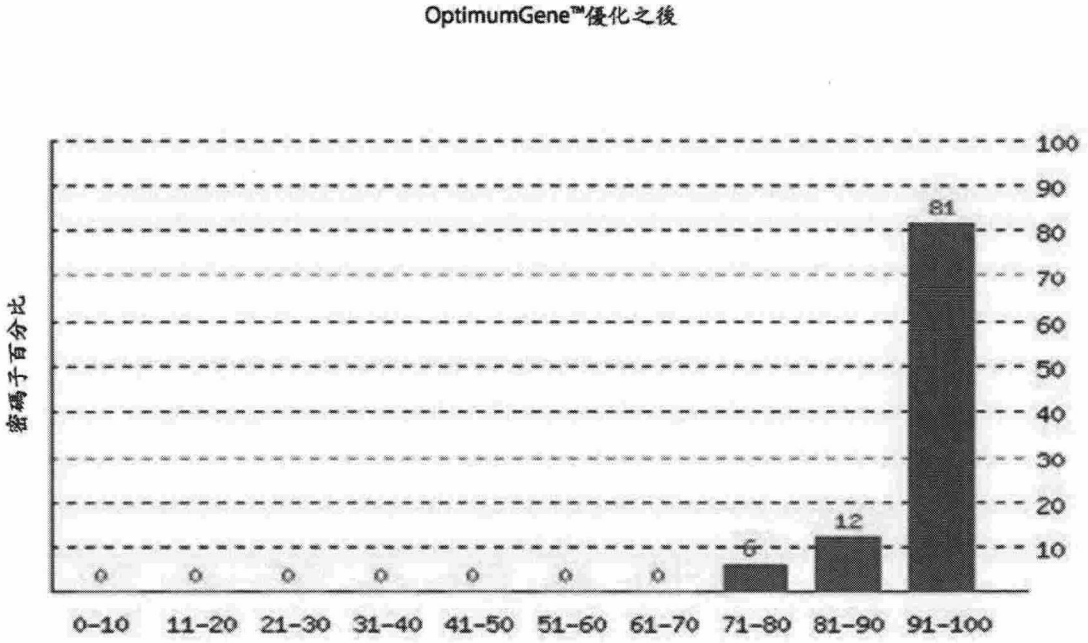

OptimumGene™優化之後
密碼子百分比
100
90
80
70
60
50
40
30
20
10
81
12
6
0
0-10
11-20
21-30
31-40
41-50
51-60
61-70
71-80
81-90
91-100

圖 77B

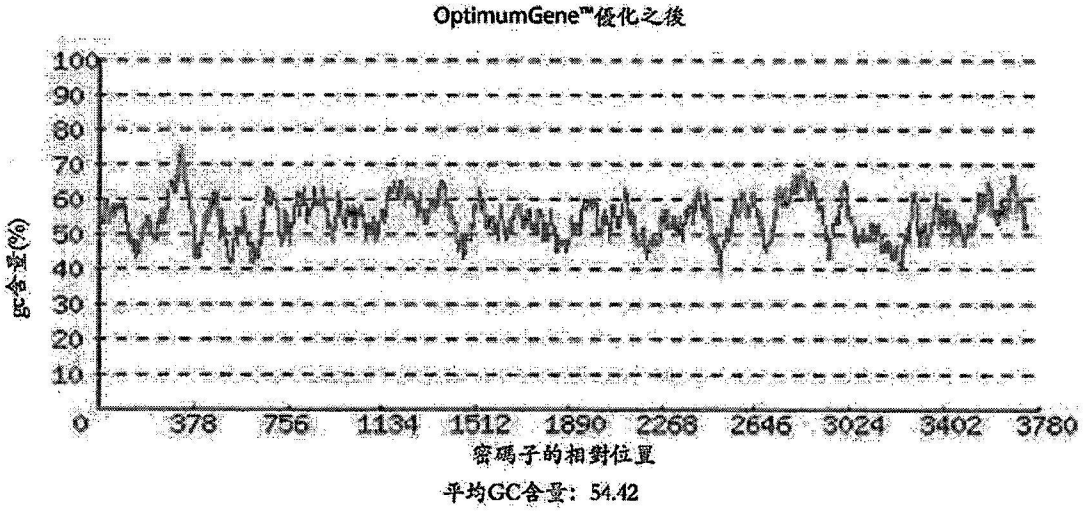

OptimumGene™優化之後
100
90
80
70
60
50
40
30
20
10
0
GC含量(%)
378
756
1134
1512
1890
2268
2646
3024
3402
3780
密碼子的相對位置
平均GC含量：54.42

圖 77C

限制性內切酶	優化的
綠色：過濾位點；藍色：檢查位點(未過濾的)；紅色：保留位點	
BamHI(GGATCC)	1(3835)
EcoRI(GAATTC)	1(3925)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(3899)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

順式作用元件	優化的
剪接(GGTAAG)	0
剪接(GGTGAT)	0
PolyA(AATAAA)	0
PolyA(ATTAAA)	0
去穩定化(ATTTA)	0
PolyT(TTTTTT)	0
PolyA(AAAAAAA)	0

抗病毒模體	優化的
	0

圖 77D

優化之後

最大同向重複序列：大小：13 距離：2043 頻率：2

最大反向重複序列：大小：13 Tm：41.3 起始位置：1469,2963

最大Dyad重複序列：無

圖 77E

GGTACC

ATGGACAGCCTGAAGGATTTCACCAACCTGTACCCCGTGTCCAAGACACTGCGGTTTGAGCTGAAGCCTGTGGGC
AAGACCCTGGAGAATATCGAGAAGGCCGGCATCCTGAAGGAGGATGAGCACAGAGCCGAGAGCTACCGGAGAGTG
AAGAAGATCATCGATACATATCACAAGGTGTTCATCGACAGCTCCCTGGAGAACATGGCCAAGATGGGCATCGAG
AATGAGATCAAGGCCATGCTGCAGTCCTTTTGCGAGCTGTATAAGAAGGACCACAGGACCGAGGCAGAGGACAAG
GCCCTGGATAAGATCAGGGCCGTGCTGAGGGGCCTGATCGTGGCAGCCTTCACCGGCGTGTGCGGCCGGCGGGAG
AACACAGTGCACAATGAGAAGTATGAGAGCCTGTTTAAGGAGAAGCTGATCAAGGAGATCCTGCCAGATTTCGTG
CTGTCTACAGAGGCCGAGTCCCTGCCCTTTTCTGTGGAGGAGGCCACCAGAAGCCTGAAGGAGTTCGACTCCTTT
ACATCTTACTTCGCCGGCTTTTATGAGAACCGGAAGAATATCTACTCTACCAAGCCCCAGAGCACAGCCATCGCC
TATAGACTGATCCACGAGAACCTGCCTAAGTTCATCGATAATATCCTGGTGTTTCAGAAGATCAAGGAGCCAATC
GCCAAGGAGCTGGAGCACATCAGGGCAGACTTCAGCGCCGGCGGCTACATCAAGAAGGATGAGCGCCTGGAGGAC
ATCTTTTCCCTGAACTACTATATCCACGTGCTGTCTCAGGCCGGCATCGAGAAGTACAATGCCCTGATCGGCAAG
ATCGTGACCGAGGGCGATGGCGAGATGAAGGGCCTGAACGAGCACATCAACCTGTATAATCAGCAGAGGGGCCGC
GAGGACCGGCTGCCACTGTTCAGACCCCTGTATAAGCAGATCCTGTCTGATAGGGAGCAGCTGTCCTATCTGCCA
GAGTCTTTCGAGAAGGACGAGGAGCTGCTGAGGGCCCTGAAGGAGTTTTACGATCACATCGCAGAGGACATCCTG
GGAAGGACCCAGCAGCTGATGACAAGCATCTCCGAGTACGATCTGTCCCGGATCTATGTGAGAAACGATAGCCAG
CTGACCGACATCTCCAAGAAGATGCTGGGCGATTGGAATGCCATCTACATGGCCCGGGAGAAGAGCCTATGACCAC
GAGCAGGCCCCCAAGCGCATCACAGCCAAGTACGAGAGGGACCGCATCAAGGCCCTGAAGGGCGAGCAGTCTATC
AGCCTGGCCAACCTGAACAGCTGCATCGCCTTCCTGGACAACGTGAGGGATTGTCGCGTGGACACCTATCTGTCT
ACACTGGGACAGAAGGAGGGACCTCACGGCCTGAGCAACCTGGTGGAGAACGTGTTCGCCTCCTACCACGAGGCC
GAGCAGCTGCTGTCTTTTCCCTATCCTGAGGAGAACAATCTGATCCAGGACAAGGATAACGTGGTGCTGATCAAG
AACCTGCTGGATAATATCAGCGACCTGCAGAGGTTCCTGAAGCCACTGTGGGGCATCGGCGATGAGCCCGACAAG
GATGAGAGGTTTTACGGCGAGTACAATTATATCAGGGGCGCCCTGGACCAGGTCATCCCTCTGTATAACAAGGTG
CGGAATTATCTGACCCGCAAGCCATACTCCACACGCAAGGTGAAGCTGAACTTCGGCAATAGCCAGCTGCTGTCC
GGCTGGGATAGGAACAAGGAGAAGGACAATTCTTGCGTGATCCTGCGCAAGGGCCAGAACTTCTACCTGGCCATC
ATGAACAATCGGCACAAGCGGAGCTTCGAGAATAAGATGCTGCCCGAGTATAAGGAGGGCGAGCCTTACTTCGAG
AAGATGGATTATAAGTTTCTGCCAGACCCCAACAAGATGCTGCCCAAGGTGTTCCTGTCTAAGAAGGGCATCGAG
ATCTACAAGCCTAGCCCAAAGCTGCTGGAGCAGTATGGCCACGGCACCCACAAGAAGGGCGATACCTTCAGCATG
GACGATCTGCACGAGCTGATCGACTTCTTTAAGCACTCCATCGAGGCCCACGAGGATTGGAAGCAGTTCGGCTTT
AAGTTCAGCGACACCGCCACATACGAGAACGTGAGCAGCTTCTACCGGGAGGTGGAGGACCAGGGCTACAAGCTG
TCTTTTAGAAAGGTGTCCGAGTCTTACGTGTATAGCCTGATCGATCAGGGCAAGCTGTACCTGTTCCAGATCTAT
AACAAGGACTTTAGCCCTTGTTCCAAGGGCACCCCAAATCTGCACACACTGTACTGGCGGATGCTGTTCGATGAG
AGAAACCTGGGCGACGTGATCTATAAGCTGGATGGCAAGGCCGAGATCTTCTTTCGGGAGAAGTCCCTGAAGAAT
GACCACCCACACCCACCCTGCAGGCAAGCCCATCAAGAAGGAAGAGCCGGCAGAAGAAGGGCGAGGAGAGCCTGTTC
GAGTACGATCTGGTGAAGGACCGGAGATATACCATGGATAAGTTTCAGTTCCACGTGCCAATCACAATGAACTTT
AAGTGCTCTGCCGGCAGCAAGGTGAACGACATGGTGAATGCCCACATCAGGGAGGCCAAGGACATGCACGTGATC
GGCATCGATAGGGGCGAGCGCAATCTGCTGTATATCTGCGTGATCGACAGCCGCGGCACCATCCTGGATCAGATC
TCCCTGAACACAATCAATGACATCGATTATCACGATCTGCTGGAGTCCAGGGACAAGGATCGCCAGCAGGAGCAC
AGGAACTGGCAGACCATCGAGGGCATCAAGGAGCTGAAGCAGGGCTACCTGTCTCAGGCCGTGCACCGCATCGCC

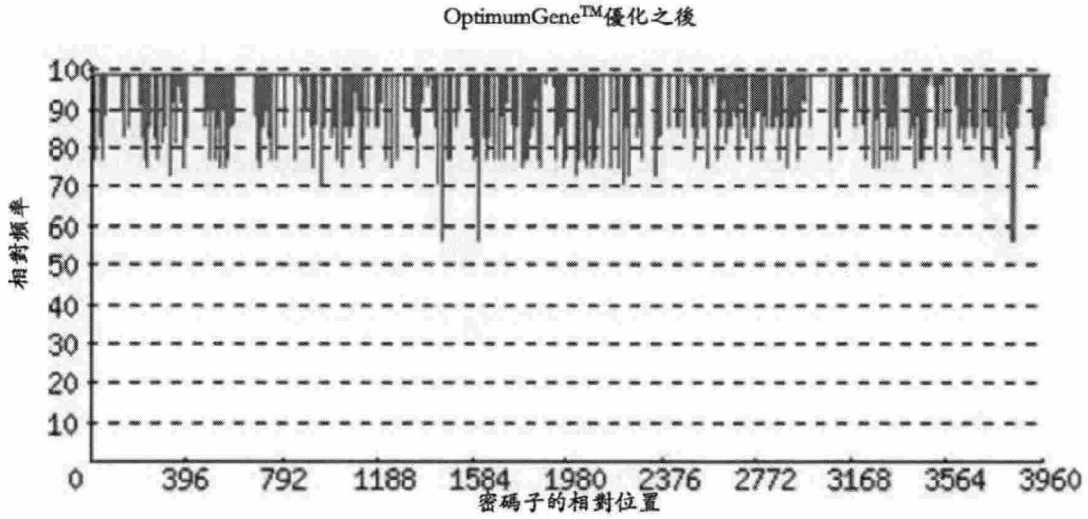
圖 77F

GAGCTGATGGTGGCCTATAAGGCCGTGGTGGCCCTGGAGGACCTGAACATGGGCTTCAAGCGGGGCAGACAGAAG
GTGGAGAGCAGCGTGTACCAGCAGTTTGAGAAGCAGCTGATCGACAAGCTGAATTATCTGGTGGATAAGAAGAAG
CGGCCCGAGGACATCGGAGGCCTGCTGAGAGCCTACCAGTTCACCGCCCCTTTCAAGAGCTTTAAGGAGATGGGC
AAGCAGAACGGCTTTCTGTTCTATATCCCTGCCTGGAACACATCCAATATCGACCCAACCACAGGCTTCGTGAAC
CTGTTTCACGTGCAGTACGAGAATGTGGATAAGGCCAAGAGCTTCTTTCAGAAGTTCGACAGCATCTCCTACAAG
CCTAAGAAGGATTGGTTTGAGTTCGCCTTTGACTATAAGAACTTCACCAAGAAGGCCGAGGGCTCTAGGAGCATG
TGGATTCTGTGCACCCACGGCTCCCGGATCAAGAACTTCAGAAATTCTCAGAAGAATGGCCAGTGGGATAGCGAG
GAGTTTGCCCTGACCGAGGCCTTCAAGTCCCTGTTTGTGCGGTACGAGATCGATTATACCGCCGACCTGAAAACC
GCCATCGTGGACGAGAAGCAGAAGGATTTCTTTGTGGACCTGCTGAAGCTGTTCAAGCTGACCGTGCAGATGAGA
AACTCCTGGAAGGAGAAGGACCTGGATTACCTGATCTCTCCAGTGGCCGGCGCCGATGGCAGGTTCTTTGACACA
CGCGAGGGCAATAAGAGCCTGCCCAAGGACGCAGATGCAAACGGAGCCTATAATATCGCCCTGAAGGGCCTGTGG
GCACTGAGGCAGATCAGACAGACCTCCGAGGGCGGCAAGCTGAAGCTGGCCATCTCTAACAAGGAGTGGCTGCAG
TTTGTGCAGGAGAGATCCTACGAGAAGGAC

圖 77G

MDSLKDFTNLYPVSKTLRFELKPVGKTLENIEKAGILKEDEHRAESYRRVKKIIDTYHKV
FIDSSLENMAKMGIENEIKAMLQSFCELYKKDHRTEGEDKALDKIRAVLRGLIVGAFTGV
CGRRENTVQNEKYESLFKEKLIKEILPDFVLSTEAESLPFSVEEATRSLKEFDSFTSYFA
GFYENRKNIYSTKPQSTAIAYRLIHENLPKFIDNILVFQKIKEPIAKELEHIRADFSAGG
YIKKDERLEDIFSLNYYIHVLSQAGIEKYNALIGKIVTEGDGEMKGLNEHINLYNQQRGR
EDRLPLFRPLYKQILSDREQLSYLPESFEKDEELLRALKEFYDHIAEDILGRTQQLMTSI
SEYDLSRIYVRNDSQLTDISKKMLGDWNAIYMARERAYDHEQAPKRITAKYERDRIKALK
GEESISLANLNSCIAFLDNVRDCRVDTYLSTLGQKEGPHGLSNLVENVFASYHEAEQLLS
FPYPEENNLIQDKDNVVLIKNLLDNISDLQRFLKPLWGMGDEPDKDERFYGEYNYIRGAL
DQVIPLYNKVRNYLTRKPYSTRKVKLNFGNSQLLSGWDRNKEKDNSCVILRKGQNFYLAI
MNNRHKRSFENKMLPEYKEGEPYFEKMDYKFLPDPNKMLPKVFLSKKGIEIYKPSPKLLE
QYGHGTHKKGDTFSMDDLHELIDFFKHSIEAHEDWKQFGFKFSDTATYENVSSFYREVED
QGYKLSFRKVSESYVYSLIDQGKLYLFQIYNKDFSPCSKGTPNLHTLYWRMLFDERNLAD
VIYKLDGKAEIFFREKSLKNDHPTHPAGKPIKKKSRQKKGEESLFEYDLVKDRRYTMDKF
QFHVPITMNFKCSAGSKVNDMVNAHIREAKDMHVIGIDRGERNLLYICVIDSRGTILDQI
SLNTINDIDYHDLLESRDKDRQQEHRNWQTIEGIKELKQGYLSQAVHRIAELMVAYKAVV
ALEDLNMGFKRGRQKVESSVYQQFEKQLIDKLNYLVDKKKRPEDIGGLLRAYQFTAPFKS
FKEMGKQNGFLFYIPAWNTSNIDPTTGFVNLFHVQYENVDKAKSFFQKFDSISYNPKKDW
FEFAFDYKNFTKKAEGSRSMWILCTHGSRIKNFRNSQKNGQWDSEEFALTEAFKSLFVRY
EIDYTADLKTAIVDEKQKDFFVDLLKLFKLTVQMRNSWKEKDLDYLISPVAGADGRFFDT
REGNKSLPKDADANGAYNIALKGLWALRQIRQTSEGGKLKLAISNKEWLQFVQERSYEKD

圖 77H



CAI: 0.96

圖 78A

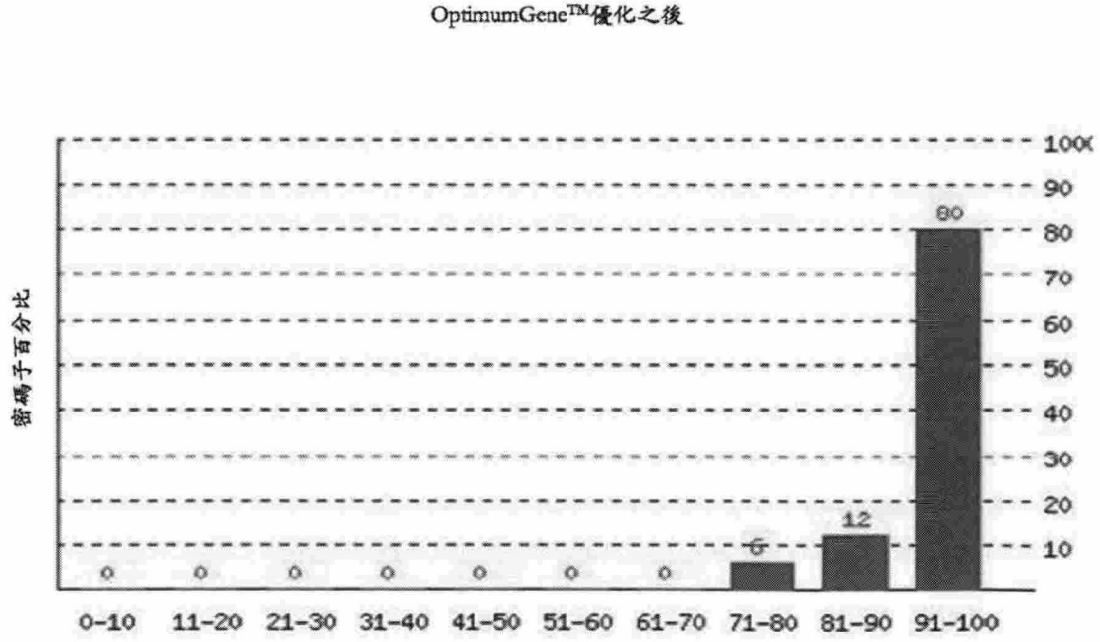


圖 78B

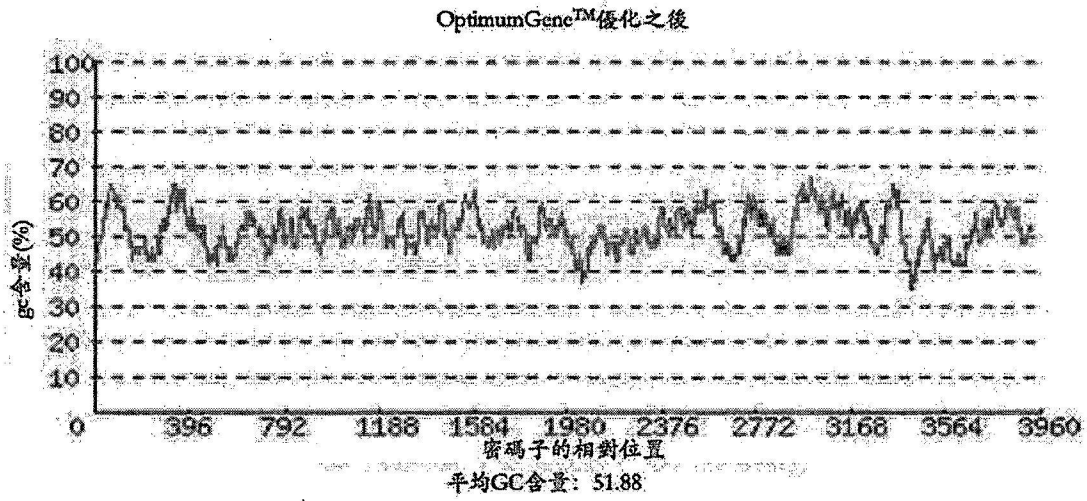

OptimumGene™優化之後
100
90
80
70
60
50
40
30
20
10
0
396
792
1188
1584
1980
2376
2772
3168
3564
3960
GC含量(%)
密碼子的相對位置
平均GC含量：51.88

圖 78C

限制性內切酶	優化的
綠色：過濾位點；藍色：檢查位點(未過濾的)；紅色：保留位點。	
BamHI(GGATCC)	1(4024)
EcoRI(GAATTC)	1(4114)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(4088)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

順式作用元件	優化的
剪接(GGTAAG)	0
剪接(GGTGAT)	0
PolyA(AATAAA)	0
PolyA(ATTAAA)	0
去穩定化(ATTTA)	0
PolyT(TTTTTT)	0
PolyA(AAAAAAA)	0

抗病毒模體	優化的
	0

圖 78D

優化之後

最大同向重複序列： 大小：15 距離：1842 頻率：2

最大反向重複序列：無

最大Dyad重複序列：無

圖 78E

ATGGAGAACTATCAGGAGTTCACCAACCTGTTTCAGCTGAATAAGACACTGAGATTCGAGCTGAAGCCCATCGGC
AAGACCTGCGAGCTGCTGGAGGAGGGCAAGATCTTCGCCAGCGGCTCCTTTCTGGAGAAGGACAAGGTGAGGGCC
GATAACGTGAGCTACGTGAAGAAGGAGATCGACAAGAAGCACAAGATCTTTATCGAGGAGACACTGAGCTCCTTC
TCTATCAGCAACGATCTGCTGAAGCAGTACTTTGACTGCTATAATGAGCTGAAGGCCTTCAAGAAGGACTGTAAG
AGCGATGAGGAGGAGGTGAAGAAAACCGCCCTGCGCAACAAGTGTACCTCCATCCAGAGGGCCATGCGCGAGGCC
ATCTCTCAGGCCTTTCTGAAGAGCCCCCAGAAGAAGCTGCTGGCCATCAAGAACCTGATCGAGAACGTGTTCAAG
GCCGACGAGAATGTGCAGCACTTCTCCGAGTTTACCAGCTATTTCTCCGGCTTTGAGAACAAACAGAGAGAATTTC
TACTCTGACGAGGAGAAGTCCACATCTATCGCCTATAGGCTGGTGCACGATAACCTGCCTATCTTCATCAAGAAC
ATCTACATCTTCGAGAAGCTGAAGGAGCAGTTCGACGCCAAGACCCTGAGCGAGATCTTCGAGAACTACAAGCTG
TATGTGGCCGGCTCTAGCCTGGATGAGGTGTTCTCCCTGGAGTACTTTAACAATACCCTGACACAGAAGGGCATC
GACAACTATAATGCCGTGATCGGCAAGATCGTGAAGGAGGATAAGCAGGAGATCCAGGGCCTGAACGAGCACATC
AACCTGTATAATCAGAAGCACAAGGACCCGGAGACTGCCCTTCTTTATCTCCCTGAAGAAGCAGATCCTGTCCGAT
CGGGAGGCCCTGTCTTGGCTGCCTGACATGTTCAAGAATGATTCTGAAGTGATCAAGGCCCTGAAGGGCTTCTAC
ATCGAGGACGGCTTTGAGAACAATGTGCTGACACCTCTGGCCACCCTGCTGTCCTCTCTGGATAAGTACAACCTG
AATGGCATCTTTATCCGCAACAATGAGGCCCTGAGCTCCCTGTCCCAGAACGTGTATCGGAATTTTTCTATCGAC
GAGGCCATCGATGCCAACGCCGAGCTGCAGACCTTCAACAATTACGAGCTGATCGCCAATGCCCTGCGCGCCAAG
ATCAAGAAGGAGACAAAGCAGGGCCGGAAGTCTTTCGAGAAGTACGAGGAGTATATCGATAAGAAGGTGAAGGCC
ATCGACAGCCTGTCCATCCAGGAGATCAACGAGCTGGTGGAGAATTACGTGAGCGAGTTTAACTCTAATAGCGGC
AACATGCCAAGAAAGGTGGAGGACTACTTCAGCCTGATGAGGAAGGGCGACTTCGGCTCCAACGATCTGATCGAA
AATATCAAGACCAAGCTGAGCGCCGCAGAGAAGCTGCTGGGCACAAAGTACCAGGAGACAGCCAAGGACATCTTC
AAGAAGGATGAGAACTCCAAGCTGATCAAGGAGCTGCTGGACGCCACCAAGCAGTTCCAGCACTTTATCAAGCCA
CTGCTGGGCACAGGCGAGGAGGCAGATCGGGACCTGGTGTTCTACGGCGATTTTCTGCCCCTGTATGAGAAGTTT
GAGGAGCTGACCCTGCTGTATAACAAGGTGCGGAATAGACTGACACAGAAGCCCTATTCCAAGGACAAGATCCGC
CTGTGCTTCAACAAGCCTAAGCTGATGACAGGCTGGGTGGATTCCAAGACCGAGAAGTCTGACAACGGCACACAG
TACGGCGGCTATCTGTTTCGGAAGAAGAATGAGATCGGCGAGTACGATTATTTTCTGGGCATCTCTAGCAAGGCC
CAGCTGTTCAGAAAGAACGAGGCCGTGATCGGCGACTACGAGAGGCTGGATTACTATCAGCCAAAGGCCAATACC
ATCTACGGCTCTGCCTATGAGGGCGAGAACAGCTACAAGGAGGACAAGAAGCGGCTGAACAAAGTGATCATCGCC
TATATCGAGCAGATCAAGCAGACAAACATCAAGAAGTCTATCATCGAGTCCATCTCTAAGTATCCTAATATCAGC
GACGATGACAAGGTGACCCCATCCTCTCTGCTGGAGAAGATCAAGAAGGTGTCTATCGACAGCTACAACGGCATC
CTGTCCTTCAAGTCTTTTCAGAGCGTGAACAAGGAAGTGATCGATAACCTGCTGAAAACCATCAGCCCCCTGAAG
AACAAGGCCGAGTTTCTGGACCTGATCAATAAGGATTATCAGATCTTCACCGAGGTGCAGGCCGTGATCGACGAG
ATCTGCAAGCAGAAAACCTTCATCTACTTTCCAATCTCCAACGTGGAGCTGGAGAAGGAGATGGGCGATAAGGAC
AAGCCCCTGTGCCTGTTCCAGATCAGCAATAAGGATCTGTCCTTCGCCAAGACCTTTAGCGCCAACCTGCGGAAG
AAGAGAGGCGCCGAGAATCTGCACACAATGCTGTTTAAGGCCCTGATGGAGGGCAACCAGGATAATCTGGACCTG
GGCTCTGGCGCCATCTTCTACAGAGCCAAGAGCCTGGACGGCAACAAGCCCACACACCCTGCCAATGAGGCCATC
AAGTGTAGGAACGTGGCCAATAAGGATAAGGTGTCCCTGTTCACCTACGACATCTATAAGAACAGGCGCTACATG
GAGAATAAGTTCCTGTTTCACCTGAGCATCGTGCAGAACTATAAGGCCGCCAATGACTCCGCCCAGCTGAACAGC
TCCGCCACCGAGTATATCAGAAGGCCGATGACCTGCACATCATCGGCATCGATAGGGGCGAGCGCAATCTGCTG

圖 78F

TACTATTCCGTGATCGATATGAAGGGCAACATCGTGGAGCAGGACTCTCTGAATATCATCAGGAACAATGACCTG
GAGACAGATTACCACGACCTGCTGGATAAGAGGGAGAAGGAGCGCAAGGCCAACCGGCAGAATTGGGAGGCCGTG
GAGGGCATCAAGGACCTGAAGAAGGGCTACCTGAGCCAGGCCGTGCACCAGATCGCCCAGCTGATGCTGAAGTAT
AACGCCATCATCGCCCTGGAGGATCTGGGCCAGATGTTTGTGACCCGCGGCCAGAAGATCGAGAAGGCCGTGTAC
CAGCAGTTCGAGAAGAGCCTGGTGGATAAGCTGTCCTACCTGGTGGACAAGAAGCGGCCTTATAATGAGCTGGGC
GGCATCCTGAAGGCCTACCAGCTGGCCTCTAGCATCACCAAGAACAATTCTGACAAGCAGAACGGCTTCCTGTTT
TATGTGCCAGCCTGGAATACAAGCAAGATCGATCCCGTGACCGGCTTTACAGACCTGCTGCGGCCCAAGGCCATG
ACCATCAAGGAGGCCCAGGACTTCTTTGGCGCCTTCGATAACATCTCTTACAATGACAAGGGCTATTTCGAGTTT
GAGACAAACTACGACAAGTTTAAGATCAGAATGAAGAGCGCCCAGACCAGGTGGACAATCTGCACCTTCGGCAAT
CCGATCAAGAGAAAGAAGGATAAGAACTACTCGAATTATGAGGAGGTGGAGCTGACCCAGCAGTTCAAGAAGCTG
TTTAAGGACAGCAACATCGATTACGAGAACTGTAATCTGAAGGAGGAGATCCAGAACAAGGACAATCGCAAGTTC
TTTGATGACCTGATCAAGCTGCTGCAGCTGACACTGCAGATGCGGAACTCCGATGACAAGGGCAATGATTATATC
ATCTCTCCTGTGGCCAACGCCGAGGGCCAGTTCTTTGACTCCCGCAATGGCGATAAGAAGCTGCCACTGGATGCA
GACGCAAACGGAGCCTACAATATCGCCCGCAAGGGCCTGTGGAACATCCGGCAGATCAAGCAGACCAAGAACGAC
AAGAAGCTGAATCTGAGCATCTCCTCTACAGAGTGGCTGGATTTCGTGCGGGAGAAGCCTTACCTGAAG
AAAAGGCCGGCGGCCAAACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCCTACCCATACGATGTTCCAGAT
TACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTTCCGGATTATGCCGAAGAATTC

圖 78G

MENYQEFTNLFQLNKTLRFELKPIGKTCELLEEGKIFASGSFLEKDKVRADNVSYVKKEI
DKKHKIFIEETLSSFSISNDLLKQYFDCYNELKAFKKDCKSDEEEVKKTALRNKCTSIQR
AMREAISQAFLKSPQKKLLAIKNLIENVFKADENVQHFSEFTSYFSGFETNRENFYSDEE
KSTSIAYRLVHDNLPIFIKNIYIFEKLKEQFDAKTLSEIFENYKLYVAGSSLDEVFSLEY
FNNTLTQKGIDNYNAVIGKIVKEDKQEIQGLNEHINLYNQKHKDRRLPFFISLKKQILSD
REALSWLPDMFKNDSEVIKALKGFYIEDGFENNVLTPLATLLSSLDKYNLNGIFIRNNEA
LSSLSQNVYRNFSIDEAIDANAELQTFNNYELIANALRAKIKKETKQGRKSFEKYEEYID
KKVKAIDSLSIQEINELVENYVSEFNSNSGNMPRKVEDYFSLMRKGDFGSNDLIENIKTK
LSAAEKLLGTKYQETAKDIFKKDENSKLIKELLDATKQFQHFIKPLLGTGEEADRDLVFY
GDFLPLYEKFEELTLLYNKVRNRLTQKPYSKDKIRLCFNKPKLMTGWVDSKTEKSDNGTQ
YGGYLFRKKNEIGEYDYFLGISSKAQLFRKNEAVIGDYERLDYYQPKANTIYGSAYEGEN
SYKEDKKRLNKVIIAYIEQIKQTNIKKSIIESISKYPNISDDDKVTPSSLLEKIKKVSID
SYNGILSPKSFQSVNKEVIDNLLKTISPLKNKAEFLDLINKDYQIFTEVQAVIDEICKQK
TFIYFPISNVELEKEMGDKDKPLCLFQISNKDLSFAKTFSANLRKKRGAENLHTMLFKAL
MEGNQDNLDLGSGAIFYRAKSLDGNKPTHPANEAIKCRNVANKDKVSLFTYDIYKNRRYM
ENKFLFHLSIVQNYKAANDSAQLNSSATEYIRKADDLHIIGIDRGERNLLYYSVIDMKGN
IVEQDSLNIIRNNDLETDYHDLLDKREKERKANRQNWEAVEGIKDLKKGYLSQAVHQIAQ
LMLKYNAIIALEDLGQMFVTRGQKIEKAVYQQFEKSLVDKLSYLVDKKRPYNELGGILKA
YQLASSITKNNSDKQNGFLFYVPAWNTSKIDPVTGFTDLLRPKAMTIKEAQDFFGAFDNI
SYNDKGYFEFETNYDKFKIRMKSAQTRWTICTFGNRIKRKKDKNYWNYEEVELTEEFKKL
FKDSNIDYENCNLKEEIQNKDNRKFFDDLIKLLQLTLQMRNSDDKGNDYIISPVANAEGQ
FFDSRNGDKKLPLDADANGAYNIARKGLWNIRQIKQTKNDKKLNLSISSTEWLDFVREKP
YLK

圖 78H

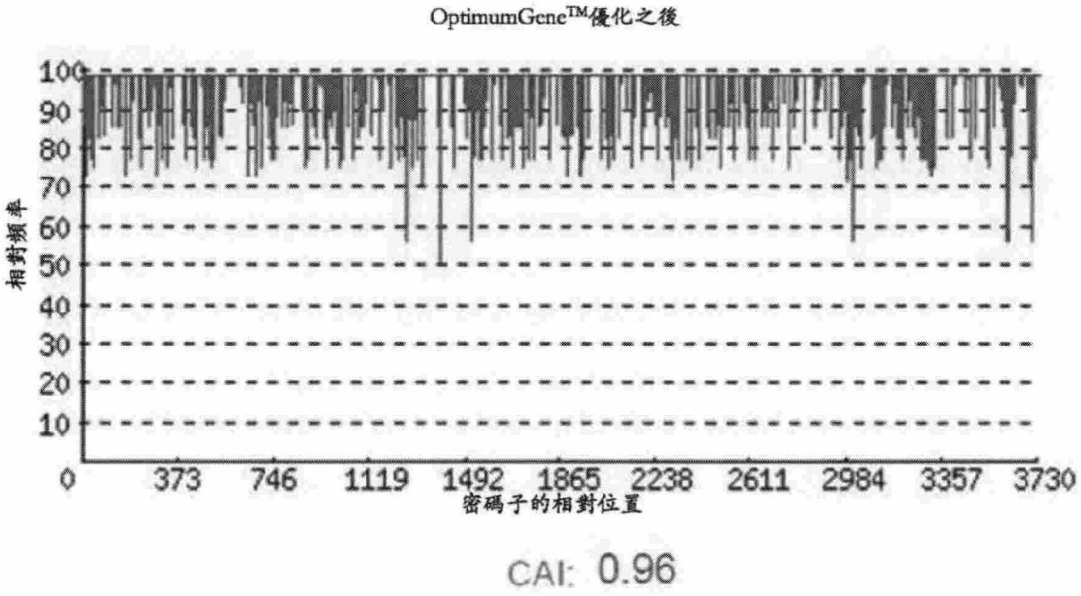

OptimumGene™優化之後
相對頻率
100
90
80
70
60
50
40
30
20
10
0
373
746
1119
1492
1865
2238
2611
2984
3357
3730
密碼子的相對位置
CAI: 0.96

圖 79A

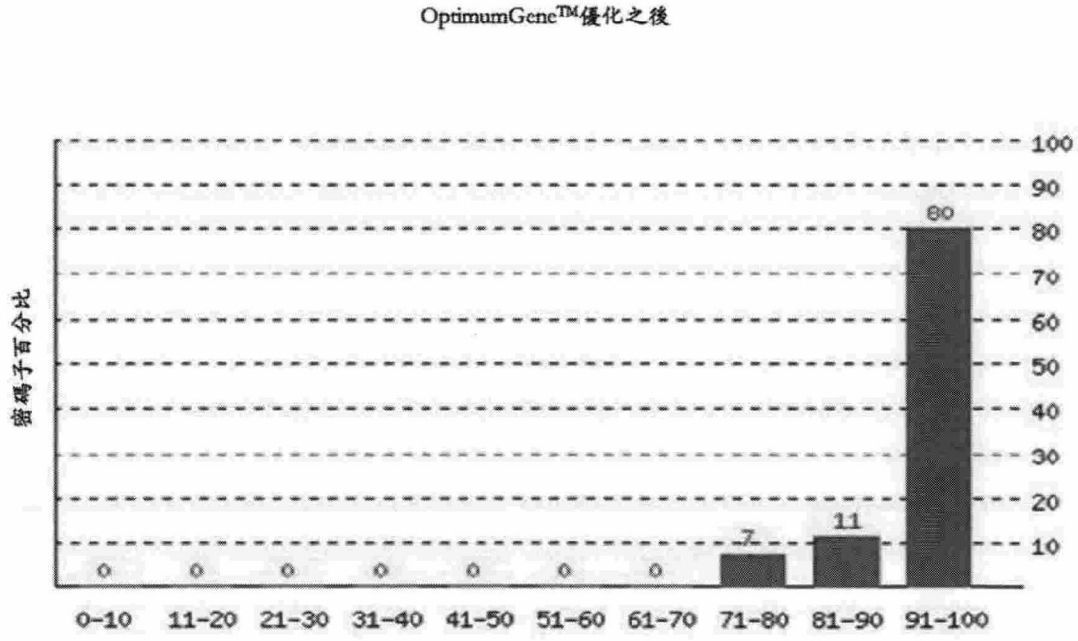

OptimumGene™優化之後
密碼子百分比
100
90
80
70
60
50
40
30
20
10
0
0
0
0
0
0
0
7
11
80
0-10
11-20
21-30
31-40
41-50
51-60
61-70
71-80
81-90
91-100

圖 79B

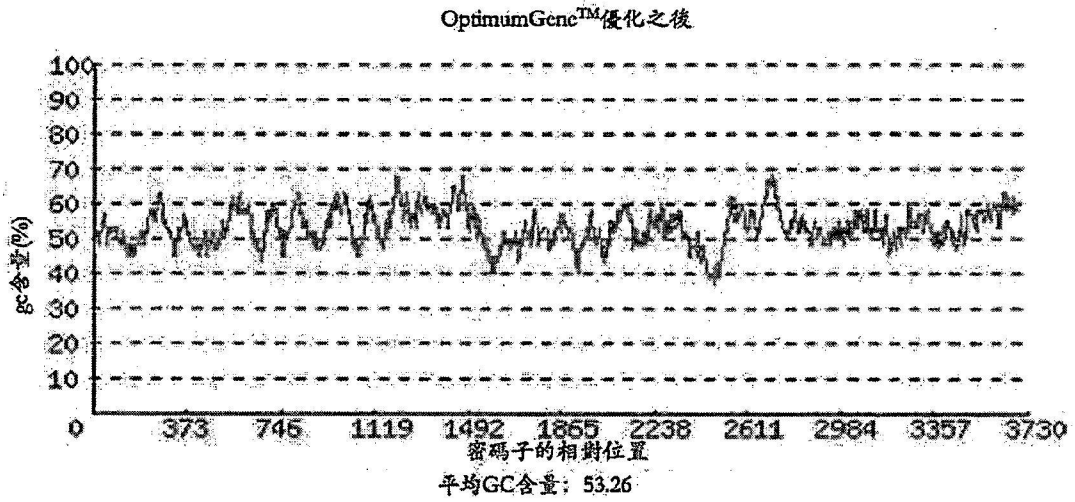

OptimumGene™優化之後
100
90
80
70
60
50
40
30
20
10
0
GC含量(%)
373
745
1119
1492
1865
2238
2611
2984
3357
3730
密碼子的相對位置
平均GC含量：53.26

圖 79C

限制性內切酶	優化的
* 綠色：過濾位點；藍色：檢查位點(未過濾的)；紅色：保留位點。	
BamHI(GGATCC)	1(3793)
EcoRI(GAATTC)	1(3883)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(3857)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

順式作用元件	優化的
剪接(GGTAAG)	0
剪接(GGTGAT)	0
PolyA(AATAAA)	0
PolyA(ATTAAA)	0
去穩定化(ATTTA)	0
PolyT(TTTTTT)	0
PolyA(AAAAAAA)	0

抗病毒模體	優化的
	0

圖 79D

優化之後

最大同向重複序列：大小：13 距離：99 頻率：2

最大反向重複序列：無

最大Dyad重複序列：無

圖 79E

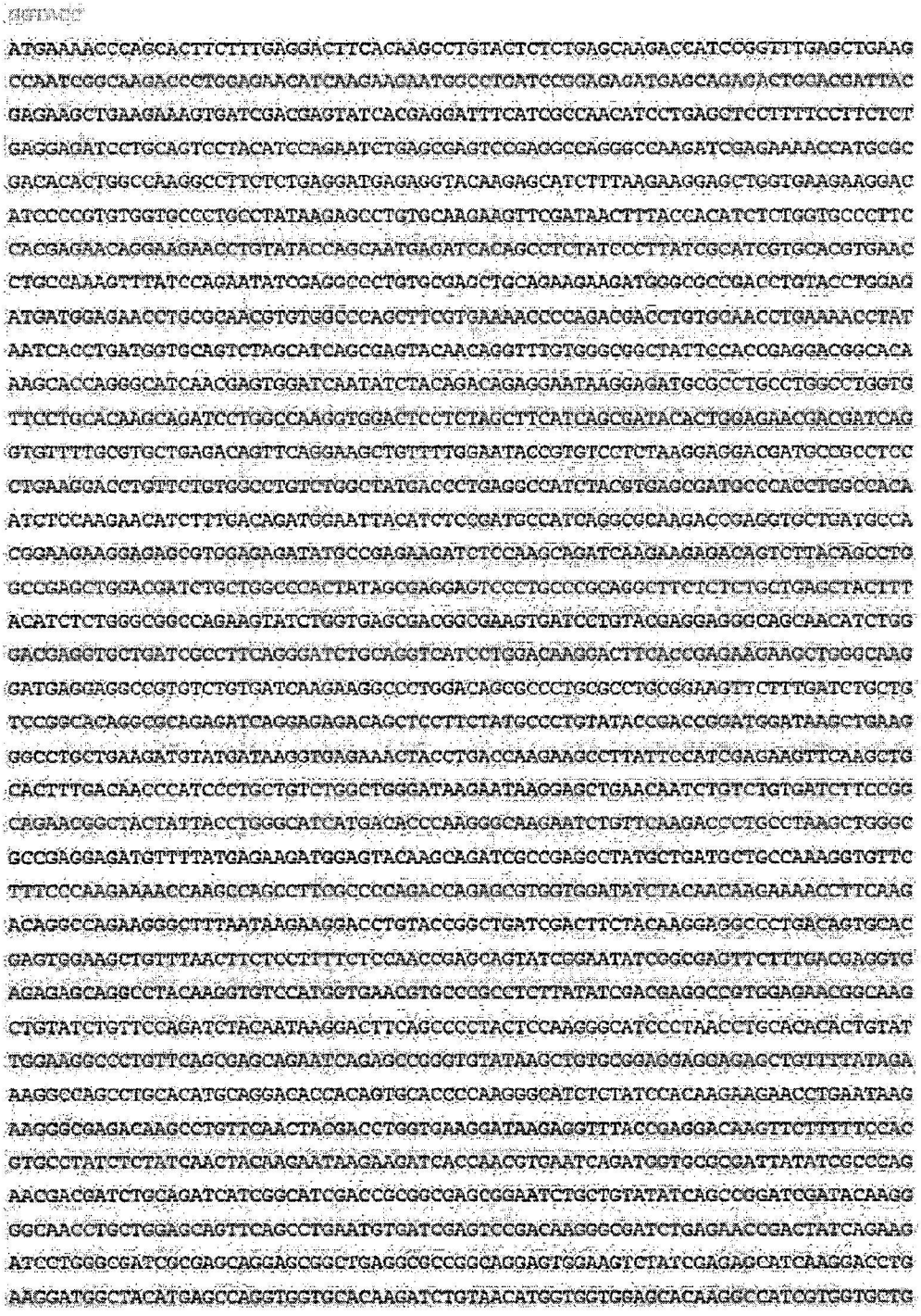

ATGAAAACCCAGCACTTCTTTGAGGACTTCACAAGCCTGTACTCTCTGAGCAAGACCATCCGGTTTGAGCTGAAG
CCAATCGGCAAGACCCTGGAGAACATCAAGAAGAATGGCCTGATCCGGAGACATGAGCAGAGACTGGACGATTAC
GAGAAGCTGAAGAAAGTGATCGACGAGTATCACGAGGATTTCATCGCCAACATCCTGAGCTCCTTTTCCTTCTCT
GAGGAGATCCTGCAGTCCTACATCCAGAATCTGAGCGAGTCCGAGGCCAGGGCCAAGATCGAGAAAACCATGCGC
GACACACTGGCCAAGGCCTTCTCTGAGGATGAGAGGTACAAGAGCATCTTTAAGAAGGAGCTGGTGAAGAAGGAC
ATCCCCGTGTGGTGCCCTGCCTATAAGAGCCTGTGCAAGAAGTTCGATAACTTTACCACATCTCTGGTGCCCTTC
CACGAGAACAGGAAGAACCTGTATACCAGCAATGAGATCACAGCCTCTATCCCTTATCGCATCGTGCACGTGAAC
CTGCCAAAGTTTATCCAGAATATCGAGGCCCTGTGCGAGCTGCAGAAGAAGATGGGCGCCGACCTGTACCTGGAG
ATGATGGAGAACCTGCGCAACGTGTGGCCCAGCTTCGTGAAAACCCCAGACGACCTGTGGAACCTGAAAACCTAT
AATCACCTGATGGTGCAGTCTAGCATCAGCGAGTACAACAGGTTTGTGGGCGGCTATTCCACCGAGGACGGCACA
AAGCACCAGGGCATCAACGAGTGGATCAATATCTACAGACAGAGGAATAAGGAGATGCGCCTGCCTGGCCTGGTG
TTCCTGCACAAGCAGATCCTGGCCAAGGTGGACTCCTCTAGCTTCATCAGCGATACACTGGAGAACGACGATCAG
GTGTTTTGCGTGCTGAGACAGTTCAGGAAGCTGTTTTGGAATACCGTGTCCTCTAAGGAGGACGATGCCGCCTCC
CTGAAGGACCTGTTCTGTGGCCTGTCTGGCTATGACCCTGAGGCCATCTACGTGAGCGATGCCCAGCTGGCCACA
ATCTCCAAGAACATCTTTGACAGATGGAATTACATCTCCGATGCCATCAGGCGCAAGACCGAGGTGCTGATGCCA
CGGAAGAAGGAGAGCGTGGAGAGATATGCCGAGAAGATCTCCAAGCAGATCAAGAAGAGACAGTCTTACAGCCTG
GCCGAGCTGGACGATCTGCTGGCCCACTATAGCGAGGAGTCCCTGCCCGCAGGCTTCTCTCTGCTGAGCTACTTT
ACATCTCTGGGCGGCCAGAAGTATCTGGTGAGCGACGGCGAAGTGATCCTGTACGAGGAGGGCAGCAACATCTGG
GACGAGGTGCTGATCGCCTTCAGGGATCTGCAGGTCATCCTGGACAAGGACTTCACCGAGAAGAAGCTGGGCAAG
GATGAGGAGGCCGTGTCTGTGATCAAGAAGGCCCTGGACAGCGCCCTGCGCCTGCGGAAGTTCTTTGATCTGCTG
TCCGGCACAGGCGCAGAGATCAGGAGAGACAGCTCCTTCTATGCCCTGTATACCGACCGGATGGATAAGCTGAAG
GGCCTGCTGAAGATGTATGATAAGGTGAGAAACTACCTGACCAAGAAGCCTTATTCCATCGAGAAGTTCAAGCTG
CACTTTGACAACCCATCCCTGCTGTCTGGCTGGGATAAGAATAAGGAGCTGAACAATCTGTCTGTGATCTTCCGG
CAGAACGGCTACTATTACCTGGGCATCATGAACACCCAAGGGCAAGAATCTGTTCAAGACCCTGCCTAAGCTGGGC
GCCGAGGAGATGTTTTATGAGAAGATGGAGTACAAGCAGATCGCCGAGCCTATGCTGATGCTGCCAAAGGTGTTC
TTTCCCAAGAAAACCAAGCCAGCCTTCGCCCCAGACCAGAGCGTGGTGGATATCTACAACAAGAAAACCTTCAAG
ACAGGCCAGAAGGGCTTTAATAAGAAGGACCTGTACCGGCTGATCGACTTCTACAAGGAGGCCCTGACAGTGCAC
GAGTGGAAGCTGTTTAACTTCTCCTTTTCTCCAACCGAGCAGTATCGGAATATCGGCGAGTTCTTTGACGAGGTG
AGAGAGCAGGCCTACAAGGTGTCCATGGTGAACGTGCCCGCCTCTTATATCGACGAGGCCGTGGAGAACGGCAAG
CTGTATCTGTTCCAGATCTACAATAAGGACTTCAGCCCCTACTCCAAGGGCATCCCTAACCTGCACACACTGTAT
TGGAAGGCCCTGTTCAGCGAGCAGAATCAGAGCCGGGTGTATAAGCTGTGCGGAGGAGGAGAGCTGTTTTATAGA
AAGGCCAGCCTGCACATGCAGGACACCACAGTGCACCCCAAGGGCATCTCTATCCACAAGAAGAACCTGAATAAG
AAGGGCGAGACAAGCCTGTTCAACTACGACCTGGTGAAGGATAAGAGGTTTACCGAGGACAAGTTCTTTTTCCAC
GTGCCTATCTCTATCAACTACAAGAATAAGAAGATCACCAACGTGAATCAGATGGTGCGCGATTATATCGCCCAG
AACGACGATCTGCAGATCATCGGCATCGACCGCGGCGAGCGGAATCTGCTGTATATCAGCCGGATCGATACAAGG
GGCAACCTGCTGGAGCAGTTCAGCCTGAATGTGATCGAGTCCGACAAGGGCGATCTGAGAACCGACTATCAGAAG
ATCCTGGGCGATCGCGAGCAGGAGCGGCTGAGGCGCCGGCAGGAGTGGAAGTCTATCGAGAGCATCAAGGACCTG
AAGGATGGCTACATGAGCCAGGTGGTGCACAAGATCTGTAACATGGTGGTGGAGCACAAGGCCATCGTGGTGCTG

圖 79F

GAGAACCTGAATCTGAGCTTCATGAAGGGCAGGAAGAAGGTGGAGAAGTCCGTGTACGAGAAGTTTGAGCGCATG
CTGGTGGACAAGCTGAACTATCTGGTGGTGGATAAGAAGAACCTGTCCAATGAGCCAGGAGGCCTGTATGCAGCA
TACCAGCTGACCAATCCACTGTTCTCTTTTGAGGAGCTGCACAGATACCCCCAGAGCGGCATCCTGTTTTTCGTG
GACCCATGGAACACCTCTCTGACAGATCCCAGCACAGGCTTCGTGAATCTGCTGGGCAGAATCAACTACACCAAT
GTGGGCGACGCCCGCAAGTTTTTCGATCGGTTTAACGCCATCAGATATGACGGCAAGGGCAATATCCTGTTCGAC
CTGGATCTGTCCAGATTTGATGTGAGGGTGGAGACACAGAGGAAGCTGTGGACACTGACCACATTCGGCTCTCGC
ATCGCCAAATCCAAGAAGTCTGGCAAGTGGATGGTGGAGCGGATCGAGAACCTGAGCCTGTGCTTTCTGGAGCTG
TTCGAGCAGTTTAATATCGGCTACAGAGTGGAGAAGGACCTGAAGAAGGCCATCCTGAGCCAGGATAGGAAGGAG
TTCTATGTGCGCCTGATCTACCTGTTTAACCTGATGATGCAGATCCGGAACAGCGACGGCGAGGAGGATTATATC
CTGTCTCCCGCCCTGAACGAGAAGAATCTGCAGTTCGACAGCAGGCTGATCGAGGCCAAGGATCTGCCTGTGGAC
GCAGATGCAAACGGAGCATACAATGTGGCCCGCAAGGGCCTGATGGTGGTGCAGAGAATCAAGAGGGGCGACCAC
GAGTCCATCCACAGGATCGGAAGGGCACAGTGGCTGAGATATGTGCAGGAGGGCATCGTGGAG

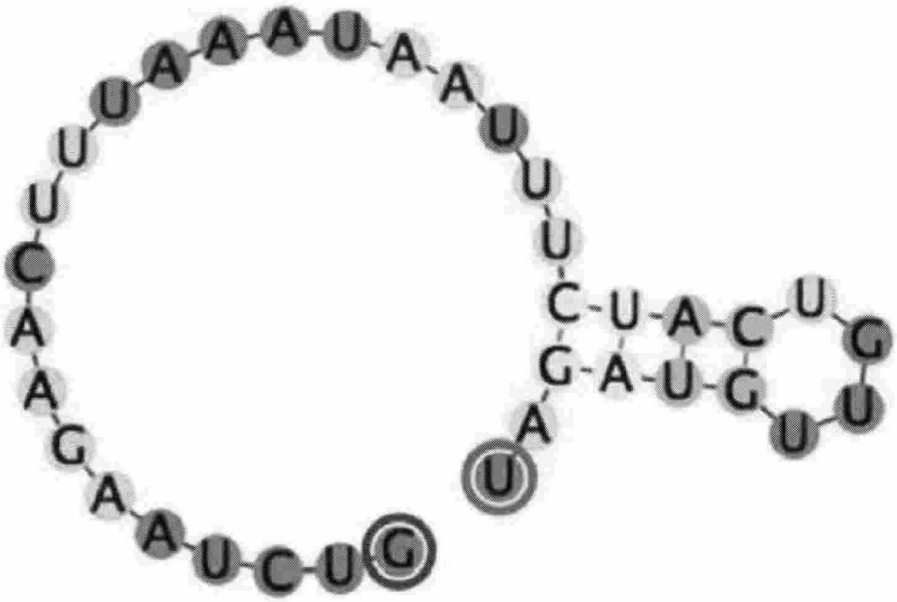


圖 79G

MKTQHFFEDFTSLYSLSKTIRFELKPIGKTLENIKKNGLIRRDEQRLDDYEKLKKVIDEY
HEDFIANILSSFSFSEEILQSYIQNLSESEARAKIEKTMRDTLAKAFSEDERYKSIFKKE
LVKKDIPVWCPAYKSLCKKFDNFTTSLVPFHENRKNLYTSNEITASIPYRIVHVNLPKFI
QNIEALCELQKKMGADLYLEMMENLRNVWPSFVKTPDDLCNLKTYNHLMVQSSISEYNRF
VGGYSTEDGTKHQGINEWINIYRQRNKEMRLPGLVFLHKQILAKVDSSSFISDTLENDDQ
VFCVLRQFRKLFWNTVSSKEDDAASLKDLFCGLSGYDPEAIYVSDAHLATISKNIFDRWN
YISDAIRRKTEVLMPRKKESVERYAEKISKQIKKRQSYSLAELDDLLAHYSEESLPAGFS
LLSYFTSLGGQKYLVSDGEVILYEEGSNIWDEVLIAFRDLQVILDKDFTEKKLGKDEEAV
SVIKKALDSALRLRKFFDLLSGTGAEIRRDSSFYALYTDRMDKLKGLLKMYDKVRNYLTK
KPYSIEKFKLHFDNPSLLSGWDKNKELNNLSVIFRQNGYYYLGIMTPKGKNLFKTLPKLG
AEEMFYEKMEYKQIAEPMLMLPKVFFPKKTKPAFAPDQSVVDIYNKKTFKTGQKGFNKKD
LYRLIDFYKEALTVHEWKLFNFSFSPTEQYRNIGEFFDEVREQAYKVSMVNVPASYIDEA
VENGKLYLFQIYNKDFSPYSKGIPNLHTLYWKALFSEQNQSRVYKLCGGGELFYRKASLH
MQDTTVHPKGISIHKKNLNKKGETSLFNYDLVKDKRFTEDKFFFHVPISINYKNKKITNV
NQMVRDYIAQNDDLQIIGIDRGERNLLYISRIDTRGNLLEQFSLNVIESDKGDLRTDYQK
ILGDREQERLRRRQEWKSIESIKDLKDGYMSQVVHKICNMVVEHKAIVVLENLNLSFMKG
RKKVEKSVYEKFERMLVDKLNYLVVDKKNLSNEPGGLYAAYQLTNPLFSFEELHRYPQSG
ILFFVDPWNTSLTDPSTGFVNLLGRINYTNVGDARKFFDRFNAIRYDGKGNILFDLDLSR
FDVRVETQRKLWTLTTFGSRIAKSKKSGKWMVERIENLSLCFLELFEQFNIGYRVEKDLK
KAILSQDRKEFYVRLIYLFNLMMQIRNSDGEEDYILSPALNEKNLQFDSRLIEAKDLPVD
ADANGAYNVARKGLMVVQRIKRGDHESIHRIGRAQWLRYVQEGIVE

圖 79H

>1 DR
GUCUAAGAACUUUAAAUAAUUUCUACUGUUGUAGAU



>2 DR
GCUAUAAUGCCUAUAUAAUUUCUACUAUUGUAGAU

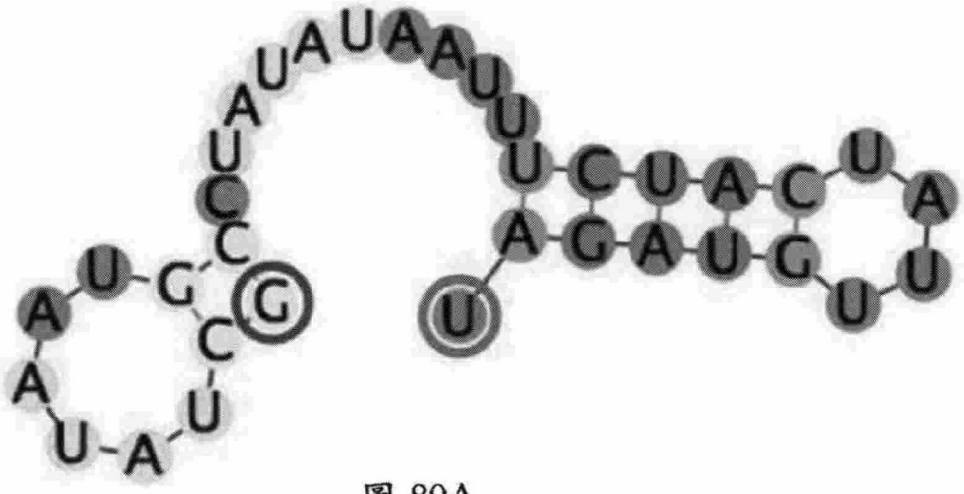
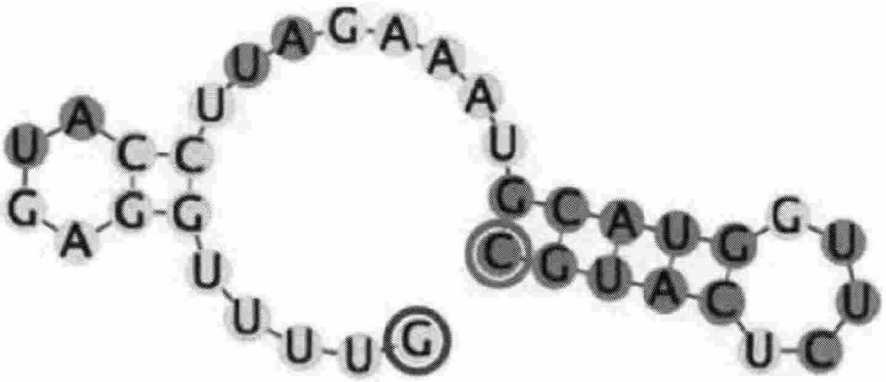


圖 80A

>3 DR
GCCUAUAAGGCUUUAGUAAUUUCUACUAUUGUAGAU



>4 DR
GUCAAUAAGACUCAUUUAAUUUCUACUUCGGUAGAU

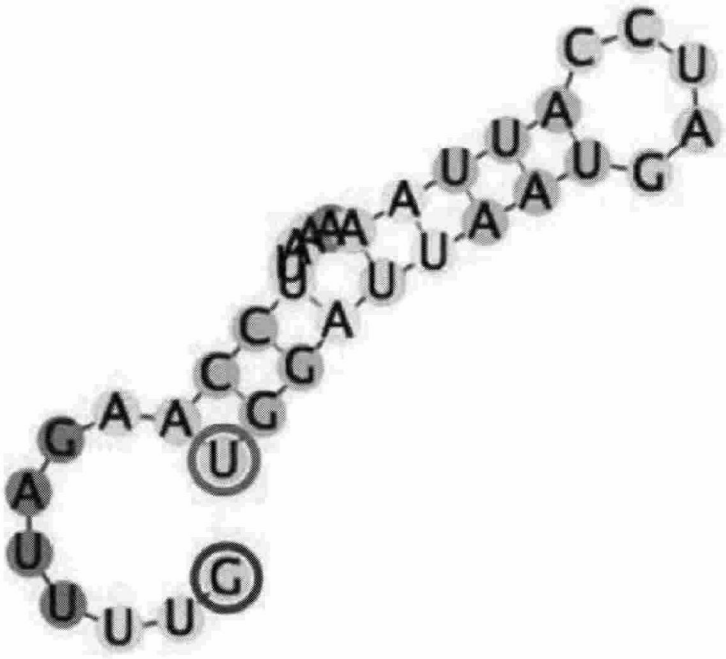
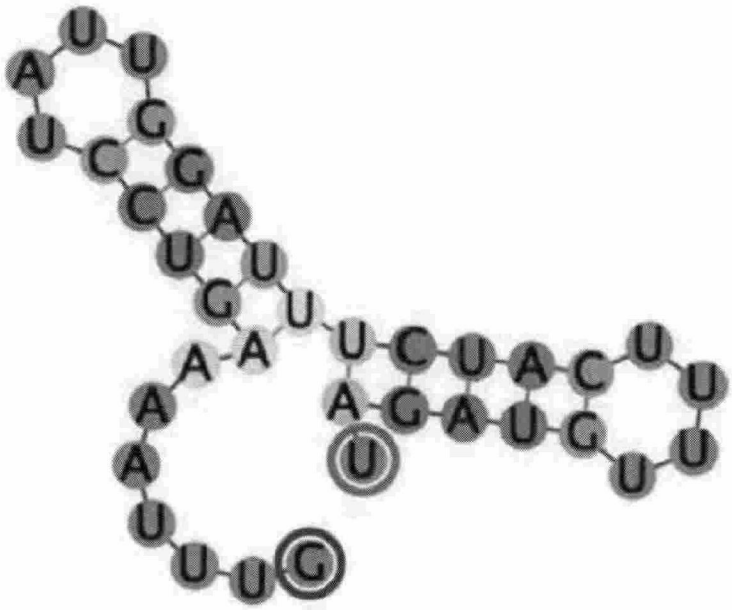


圖 80B

>5 DR
GUCUAGGUACUCUCUUUAAUUUCUACUAUUGUAGAU



>6 DR
GUUUCAAAGAUUAAAUAAUUUCUACUAAGUGUAGAU

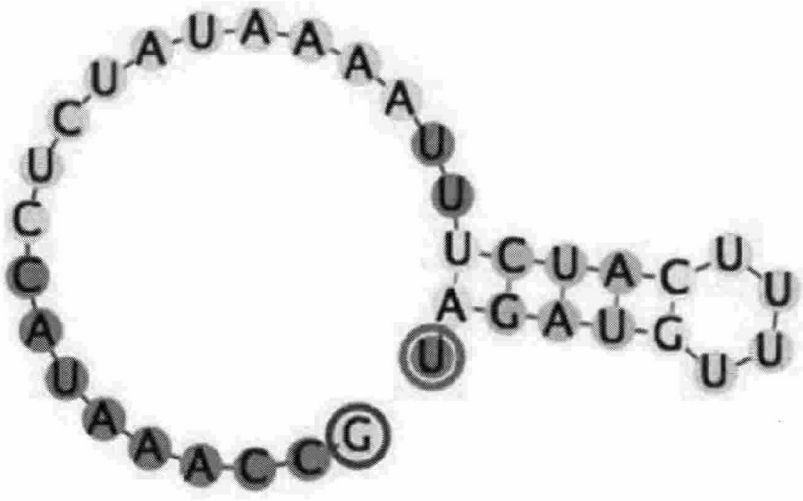
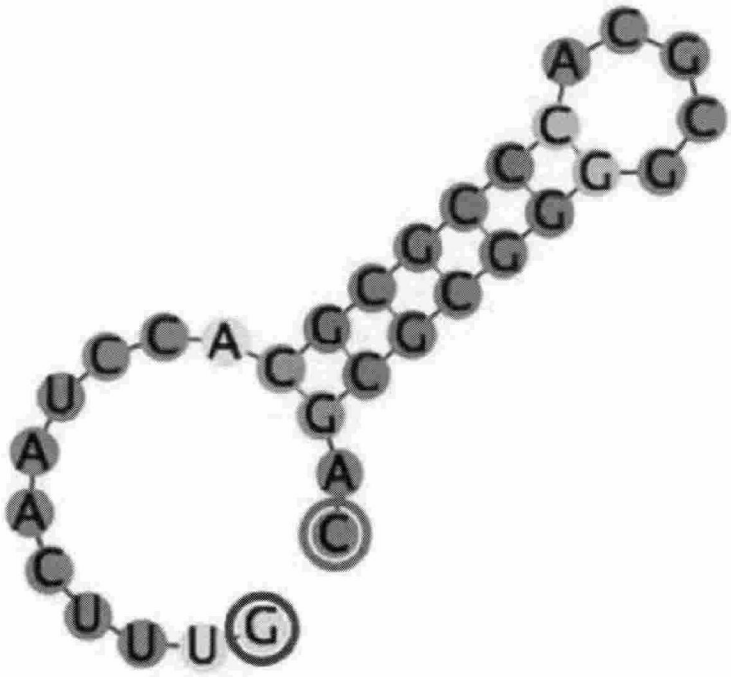


圖 80C

>7 DR
CUCUAAAGAGAGGAAAGAAUUUCUACUUUUGUAGAU



>8 DR
GUCUAACGACCUUUUAAAUUUCUACUGUUUGUAGAU

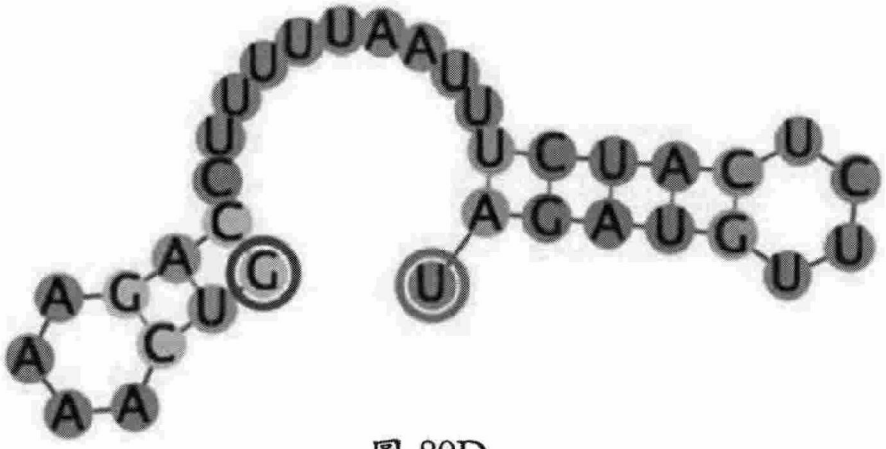
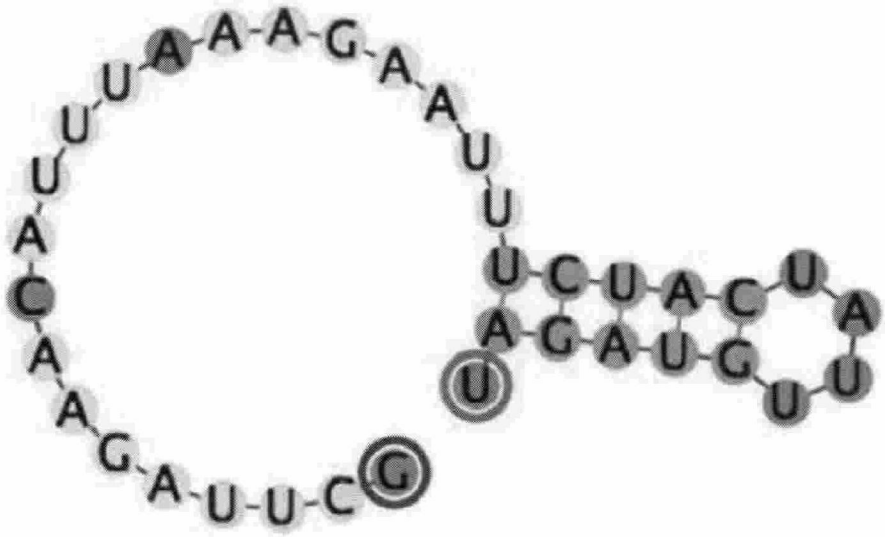


圖 80D

>9 DR
GUUUGAAUAACCUUAAAUAAUUUCUACUUUGUAGAU



>10 DR
CUCAAAACUCAUUCGAAUCUCUACUCUUUGUAGAU

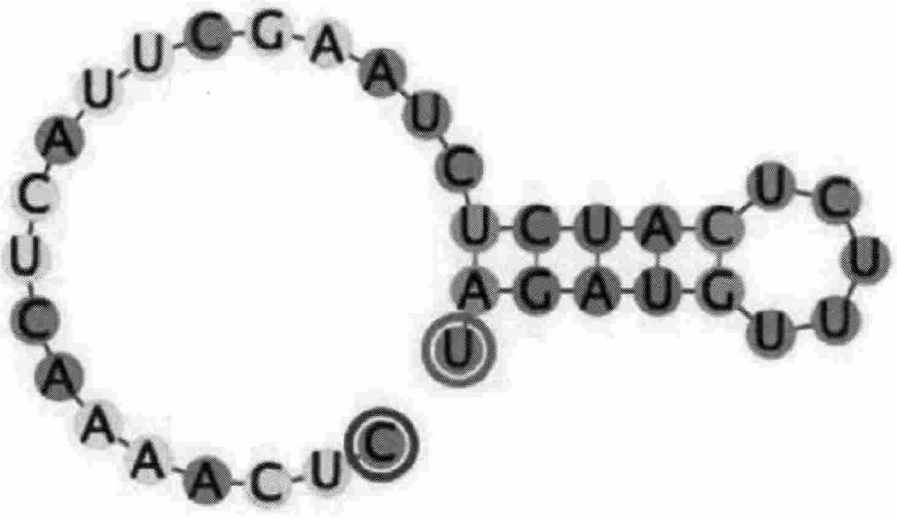
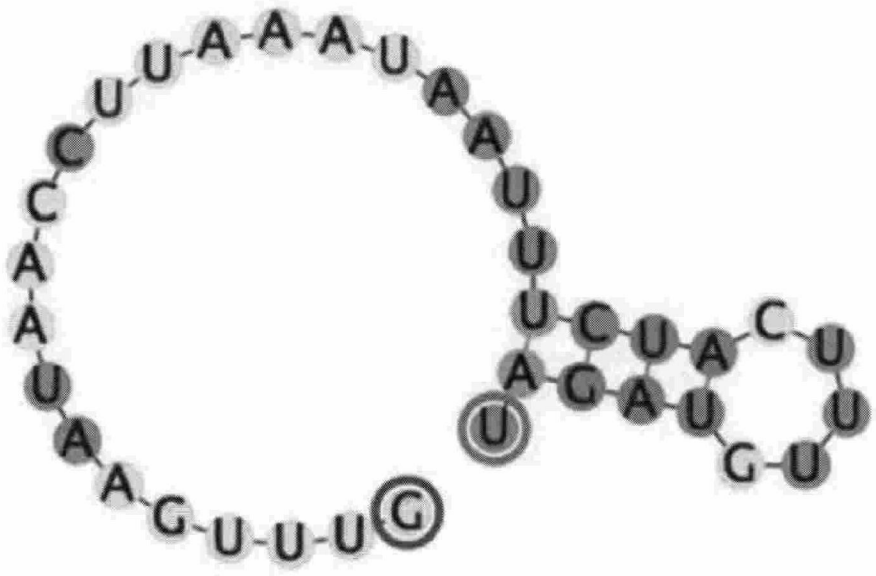


圖 80E

>11 DR
GCUUAGAACAUUUAAAGAAUUUCUACUAUUGUAGAU



>12 DR
GUCAAAAGACCUUUUUAAUUUCUACUCUUGUAGAU

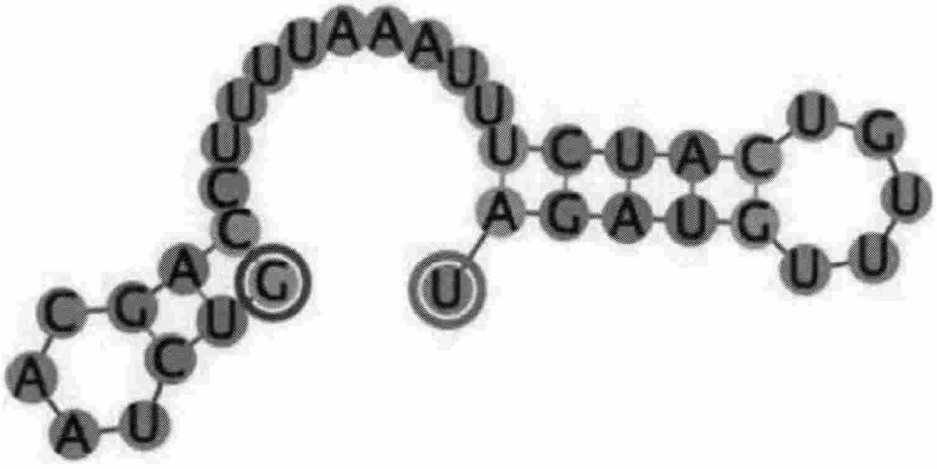
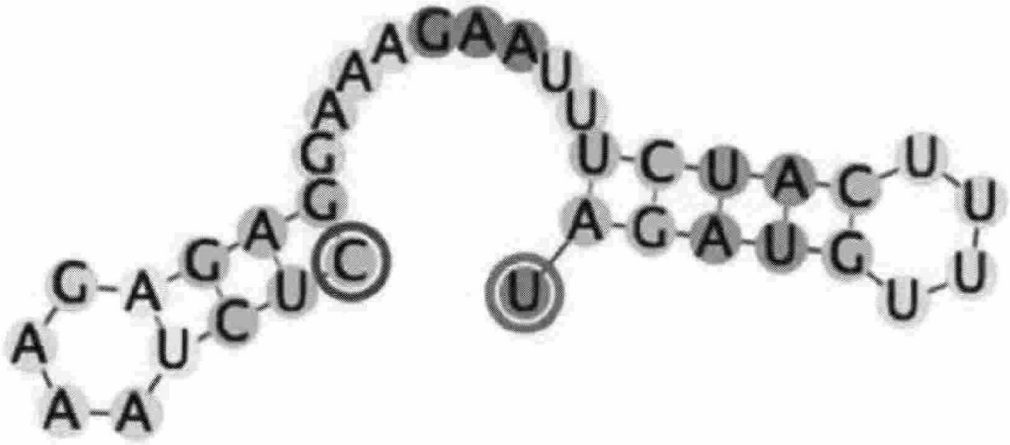


圖 80F

>13 DR
GUUUCAAUCCACGCGCCCACGCGGGGCGCGAC



>14 DR
GCCAAAUACCUCUAUAAAAUUUCUACUUUUGUAGAU

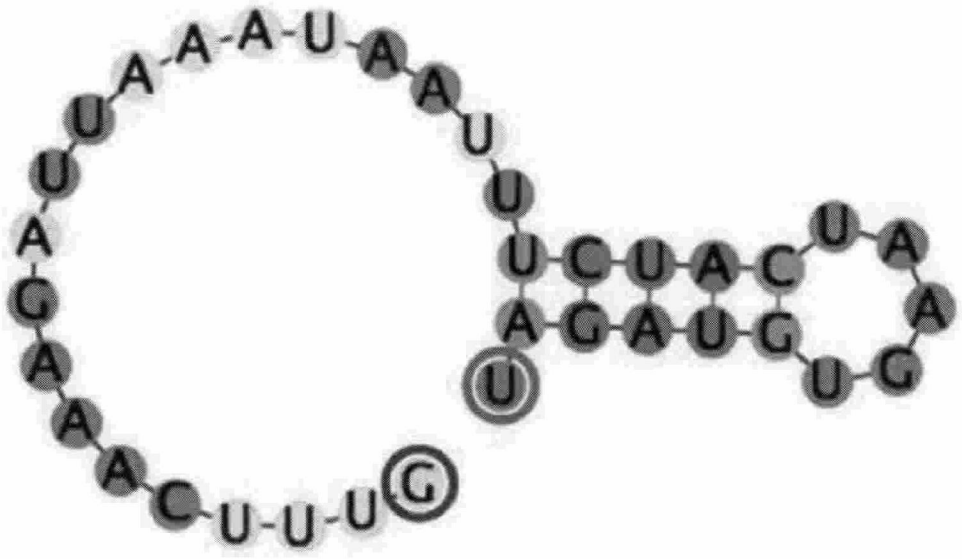
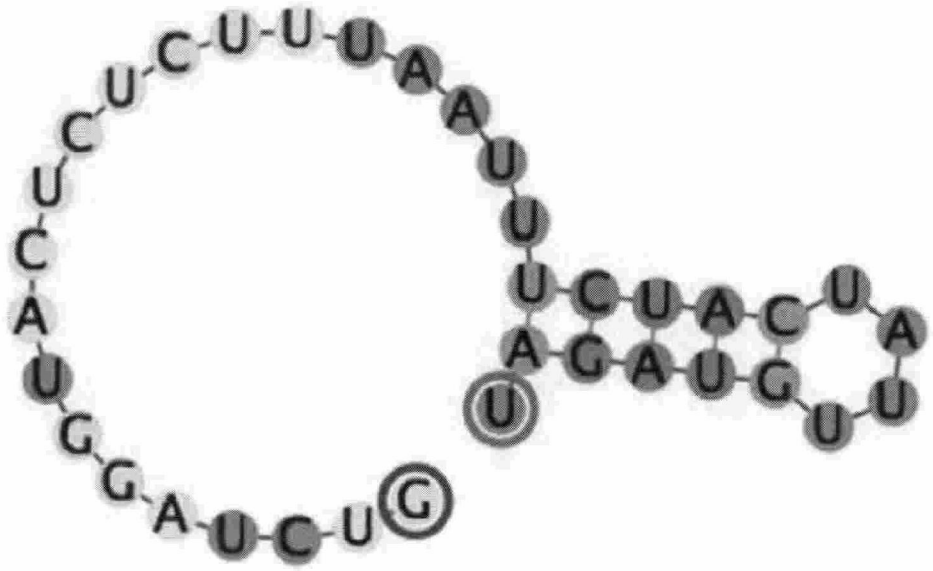


圖 80G

>15 DR
GUUUAAAAGUCCUAUUGGAUUUCUACUUUUGUAGAU



>16 DR
GUUUUAGAACCUUAAAAAUUACCUAGUAAUUAGGU

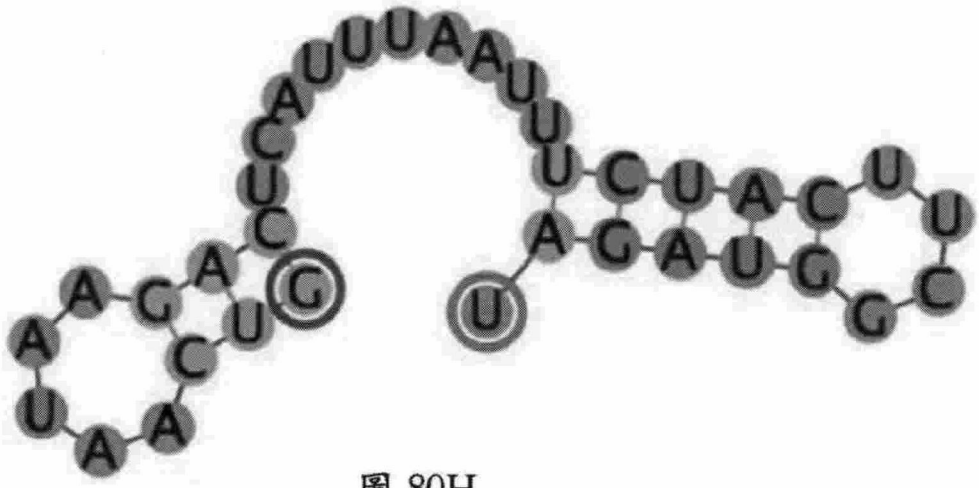


圖 80H

>17 DR
GUUUUGGAGUACCUUAGAAAUGCAUGGUUCUCAUGC

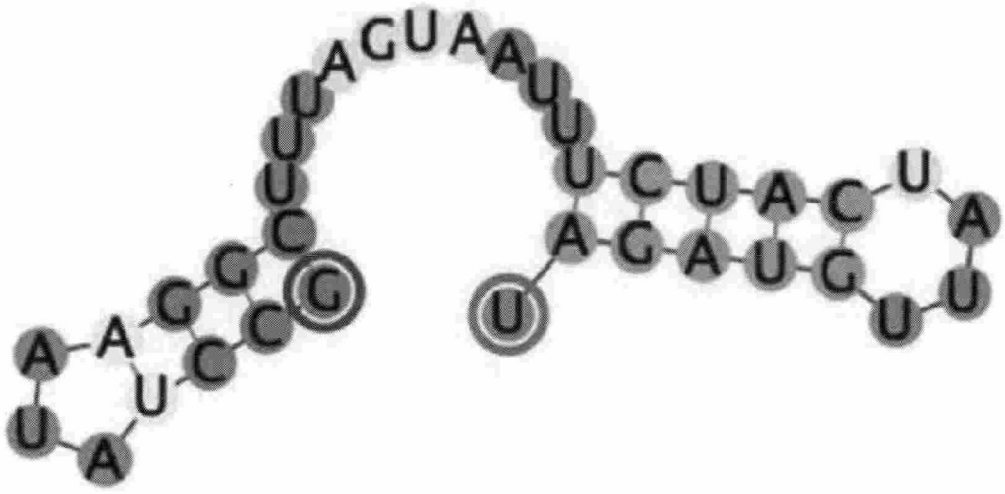
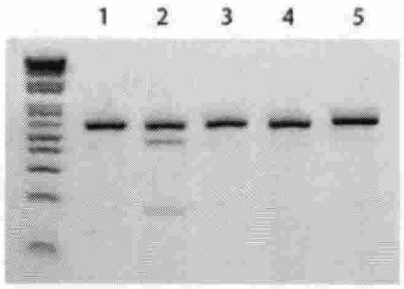


圖 80I



泳道	crRNA	DNA底物
1	Emx1 crRNA1	EMX1 擴增子
2	Emx1 crRNA2	EMX1 擴增子
3	Emx1 crRNA10	EMX1 擴增子
4	Emx1 crRNA13	EMX1 擴增子
5	無	EMX1 擴增子

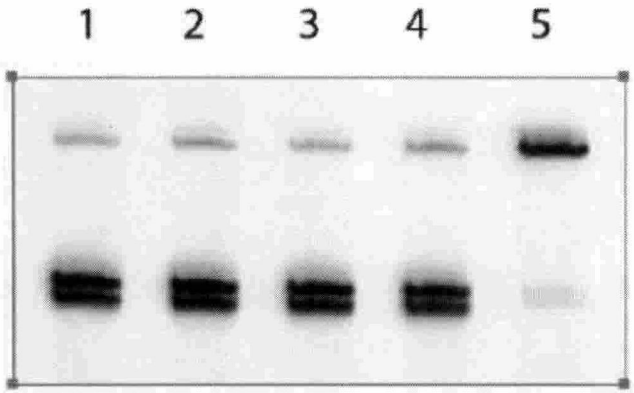
crRNA序列(5'至3')

crRNA1
GGGACUUUAAAUAAUUUCUACUGUUGUAGAUAGGCCCCAGUGGCUGCUCUGGGGGCCUCCGUCUAAGAACUU
UAAAU
crRNA2
GGGACUUUAAAUAAUUUCUACUGUUGUAGAUUCAUCUGUGCCCCUCCCUCCCCUGGCCCAGGUCUAAGAACUU
UAAAU
crRNA10
GGGACUUUAAAUAAUUUCUACUGUUGUAGAUGUGGUUGCCCACCCUAGUCAUUGGAGGUGGUCUAAGAACUU
UAAAU
crRNA13
GGGACUUUAAAUAAUUUCUACUGUUGUAGAUGGGGCCCCAGGCCGGGGUCCCCUCUGACCGUCUAAGAACUU
UAAAU

Emx1擴增子序列:

Ccatccccttctgtgaatgttagacccatgggagcagctggtcagaggggaccccggcctgggg
cccctaaccctatgtagcctcagtcttcccatcaggctctcagctcagcctgagtgttgaggcc
ccagtggctgctctgggggcctcctgagtttctcatctgtgcccctccctccctggcccaggtg
aaggtgtggttccagaaccggaggacaaagtacaaacggcagaagctggaggaggaagggcctg
agtccgagcagaagaagaagggctcccatcacatcaaccggtggcgcattgccacgaagcaggc
caatggggaggacatcgatgtcacctccaatgactagggtgggcaaccacaaacccacgagggc
agagtgctgcttgctgctggccaggcccctgcgtgggcccaagctggactctggccactccctg
gccaggctttggggaggcctggagtcatggccccacagggcttgaagcccggggccgccattga
cagagggacaagcaatgggctggctgaggcctgggaccacttggccttctcctcggagagcctg
cctgcctgggcgggcccgcccgccaccgcagcctcccagctgctctccgtgtctccaatctcc

圖 81



泳道	DNA			crRNA
1	pUC19-	間隔區	1(TTc)	deltaDR1
2	pUC19-	間隔區	1(TTc)	deltaDR2
3	pUC19-	間隔區	1(TTc)	deltaDR3
4	pUC19-	間隔區	1(TTc)	deltaDR4
5	pUC19-	間隔區	1(TTc)	deltaDR5

crRNA序列(5'至3'):

deltaDR1
GGGGUCUAAGAACUUUAAAUAAUUUCUACUGUUGUAGAUGAGAAGUCAUUUAAUAAGGCCACUGUUAAAAGUCUAAGAACUUUAAAUAAU
UUCUACUGUUGUAGAU
deltaDR2
GGGGAACUUUAAAUAAUUUCUACUGUUGUAGAUGAGAAGUCAUUUAAUAAGGCCACUGUUAAAAGUCUAAGAACUUUAAAUAAUUUCUACUGUUGUA
GAU
deltaDR3
GGGUAAAUAAUUUCUACUGUUGUAGAUGAGAAGUCAUUUAAUAAGGCCACUGUUAAAAGUCUAAGAACUUUAAAUAAUUUCUACUGUUGUAGAU
deltaDR4
GGGAUUUCUACUGUUGUAGAUGAGAAGUCAUUUAAUAAGGCCACUGUUAAAAGUCUAAGAACUUUAAAUAAUUUCUACUGUUGUAGAU
deltaDR5
GGGACUGUUGUAGAUGAGAAGUCAUUUAAUAAGGCCACUGUUAAAAGUCUAAGAACUUUAAAUAAUUUCUACUGUUGUAGAU

圖 82A

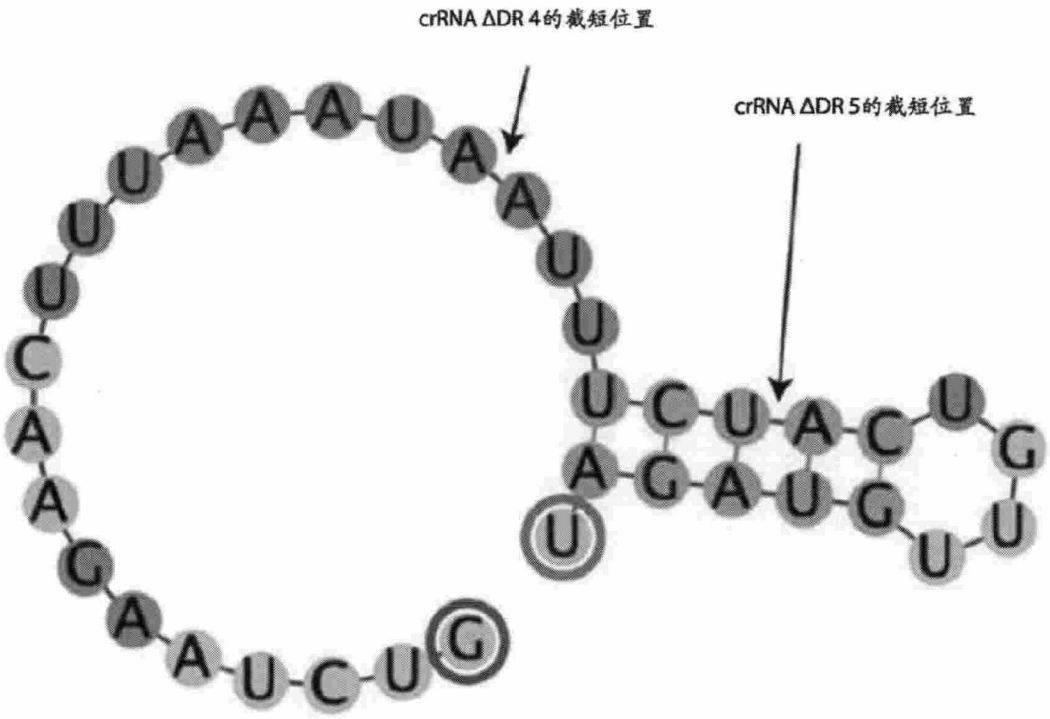
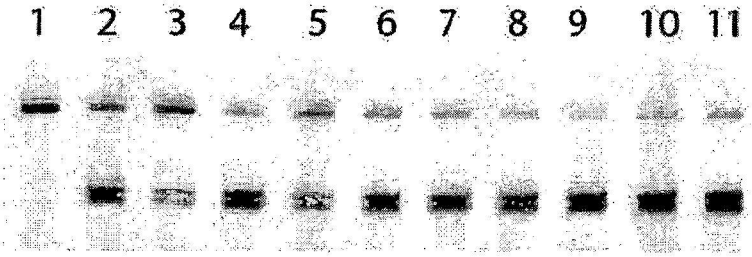


圖 82B

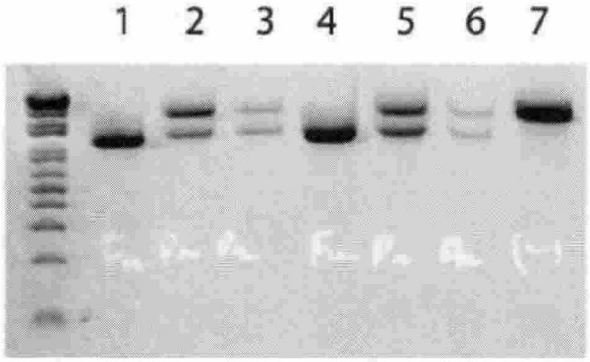


每個泳道由含有Cpf1的細胞裂解物、具有TTc原型間隔區的pUC19和相應的crRNA組成，如1-11指示。

crRNA序列(5'至3'，失配底部底線示出)

1	GGGACTTTAAATAATTTCTACTGTTGTAGATTAGAAGTCATTTAATAAGGCCACTGTTAAAAGTCTAAGAACTTTAAAT
2	GGGACTTTAAATAATTTCTACTGTTGTAGATGAGGAGTCATTTAATAAGGCCACTGTTAAAAGTCTAAGAACTTTAAAT
3	GGGACTTTAAATAATTTCTACTGTTGTAGATGAGAAGCCATTTAATAAGGCCACTGTTAAAAGTCTAAGAACTTTAAAT
4	GGGACTTTAAATAATTTCTACTGTTGTAGATGAGAAGTCAATTAATAAGGCCACTGTTAAAAGTCTAAGAACTTTAAAT
5	GGGACTTTAAATAATTTCTACTGTTGTAGATGAGAAGTCATTTTATAAGGCCACTGTTAAAAGTCTAAGAACTTTAAAT
6	GGGACTTTAAATAATTTCTACTGTTGTAGATGAGAAGTCATTTAATGAGGCCACTGTTAAAAGTCTAAGAACTTTAAAT
7	GGGACTTTAAATAATTTCTACTGTTGTAGATGAGAAGTCATTTAATAAGCCCACTGTTAAAAGTCTAAGAACTTTAAAT
8	GGGACTTTAAATAATTTCTACTGTTGTAGATGAGAAGTCATTTAATAAGGCCTCTGTTAAAAGTCTAAGAACTTTAAAT
9	GGGACTTTAAATAATTTCTACTGTTGTAGATGAGAAGTCATTTAATAAGGCCACTATTAAAAGTCTAAGAACTTTAAAT
10	GGGACTTTAAATAATTTCTACTGTTGTAGATGAGAAGTCATTTAATAAGGCCACTGTTTAAAGTCTAAGAACTTTAAAT
11	GGGACTTTAAATAATTTCTACTGTTGTAGATGAGAAGTCATTTAATAAGGCCACTGTTAAAGGTCTAAGAACTTTAAAT

圖 83



泳道	純化的蛋白Soln DNA底物	crRNA
1	FnCpf1 具有TTc原型間隔區的pUC	間隔區1-靶向crRNA
2	PaCpf1 具有TTc原型間隔區的pUC	間隔區1-靶向crRNA
1	PaCpf1 具有TTc原型間隔區的pUC	間隔區1-靶向crRNA
1	GST-FnCpf1 具有TTc原型間隔區的pUC	間隔區1-靶向crRNA
1	GST-PaCpf1 具有TTc原型間隔區的pUC	間隔區1-靶向crRNA
1	MBP-PaCpf1 具有TTc原型間隔區的pUC	間隔區1-靶向crRNA
1	具有TTc原型間隔區的空pUC	間隔區1-靶向crRNA

間隔區-靶向crRNA序列(5'至3'):

acuuuaaauaauuucuacuguuguagaugagaagucauuuaauaaggccacuguuaaaagucuaagaacu
uuaaau

圖 84

FnCpf1 DR二級結構(莖環突出顯示):

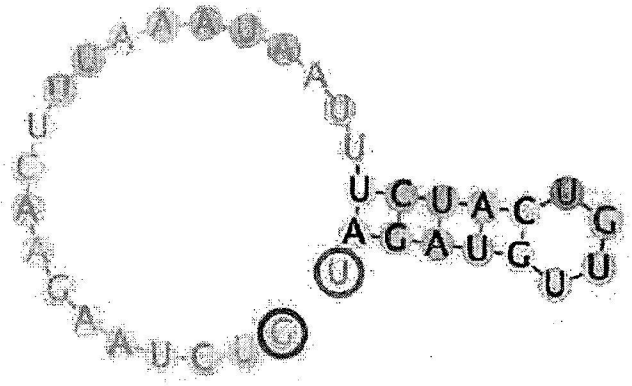


圖 85A

PaCpf1 DR二級結構(莖環突出顯示，除了環區中的單個鹼基不同之外都相同):

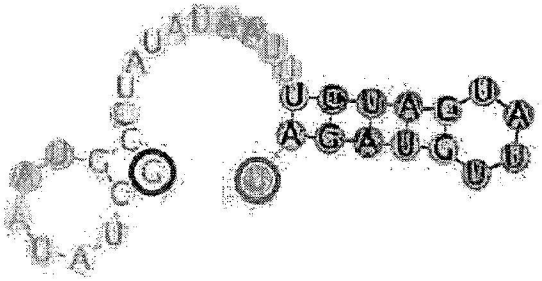


圖 85B

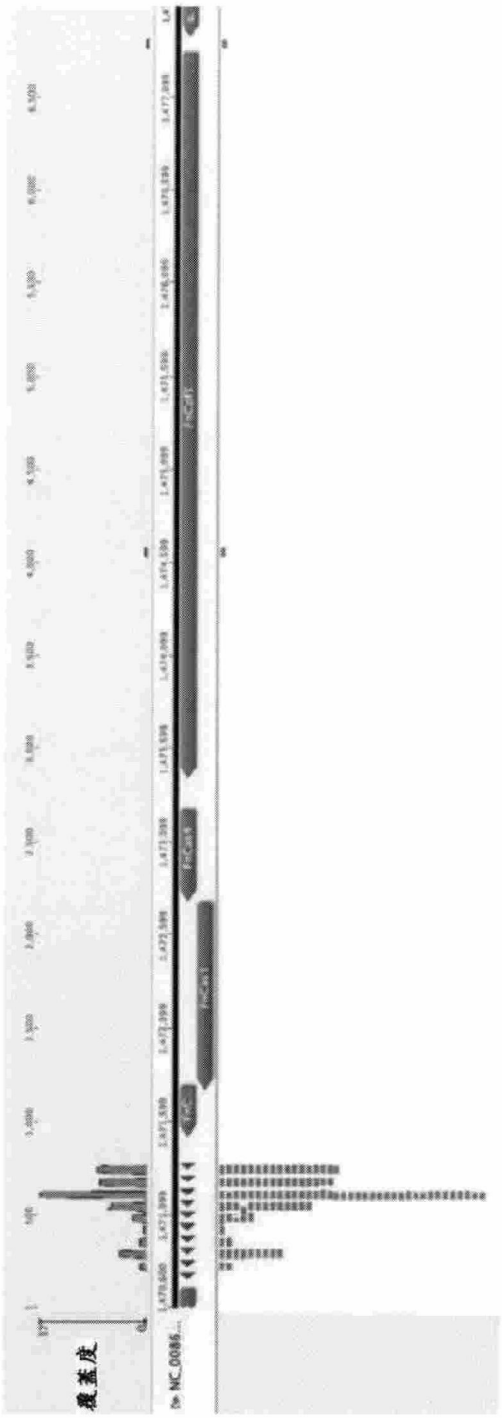


圖 86

FnCpf1的成熟crRNA序列具有以下形式：

5′-AAUUUCUACUGUUGUAGAUN$_{20-24}$-3′

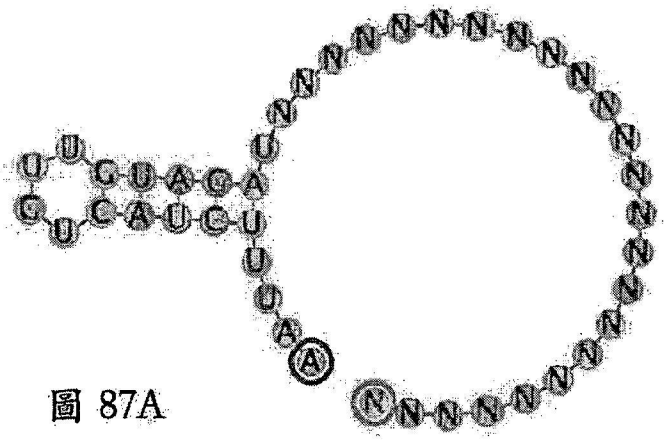


圖 87A

PaCpf1的成熟crRNA序列具有以下形式：

5′-AAUUUCUACUAUUGUAGAUN$_{20-24}$-3′

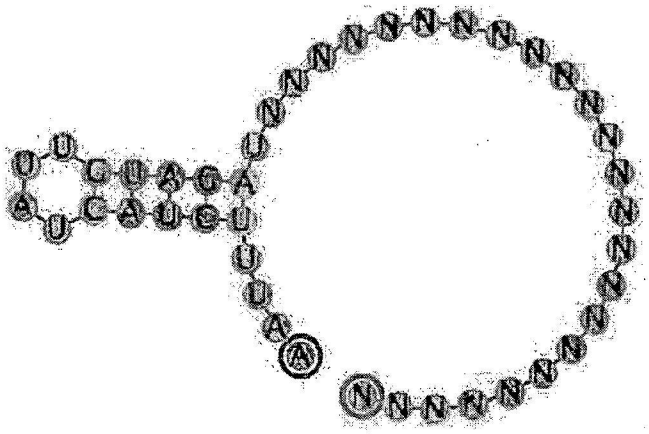


圖 87B

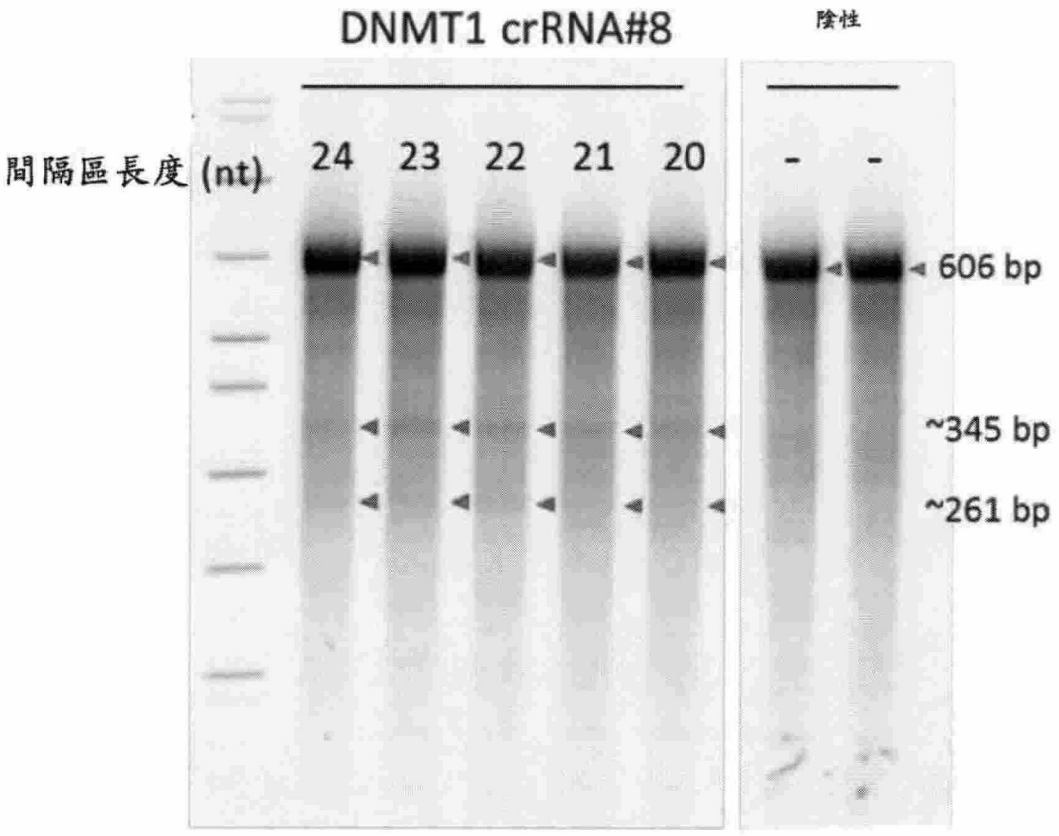


圖 88

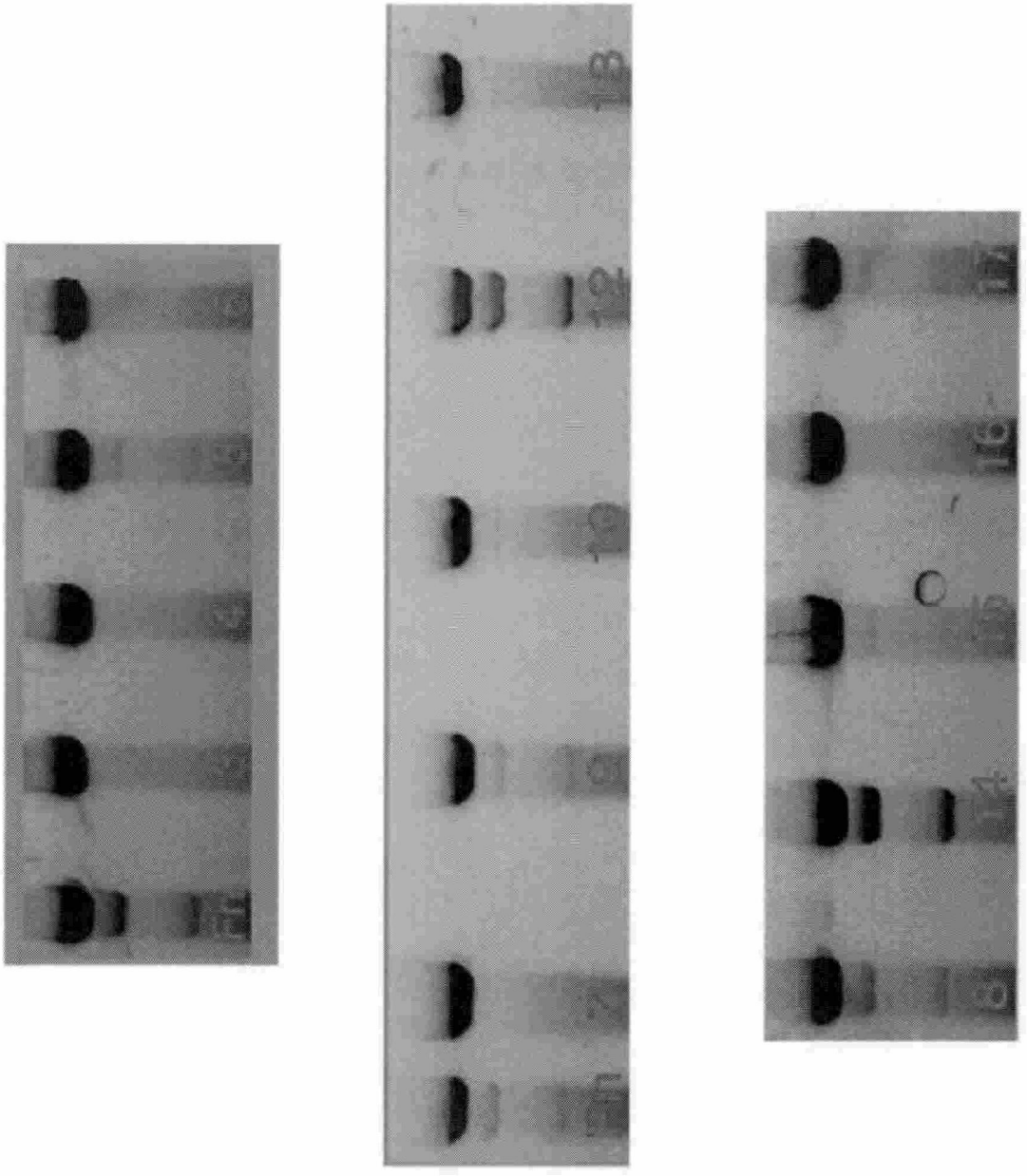


圖 89

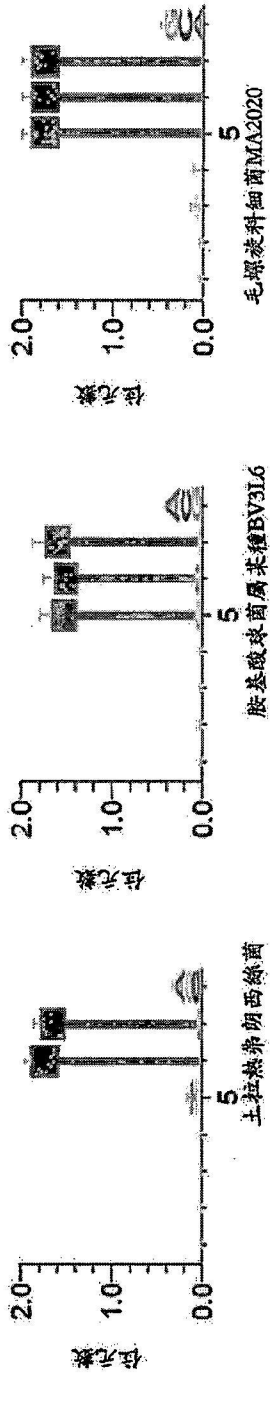


圖 90A

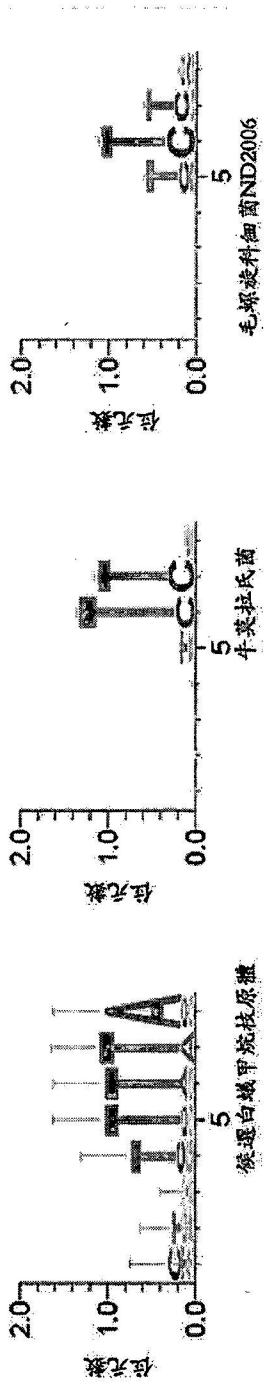


圖 90B

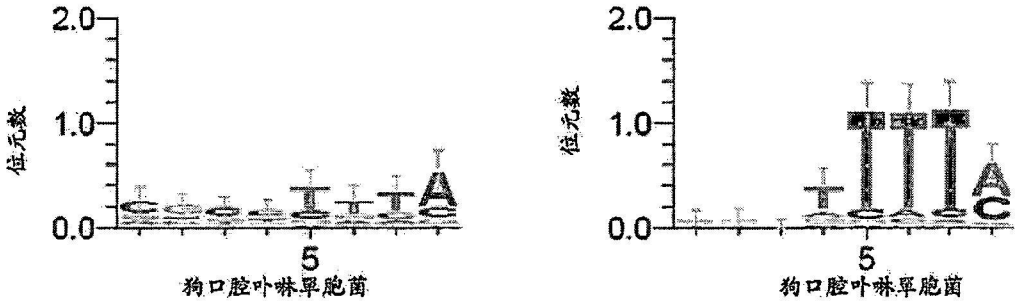


圖 90C

原型間隔區(1)

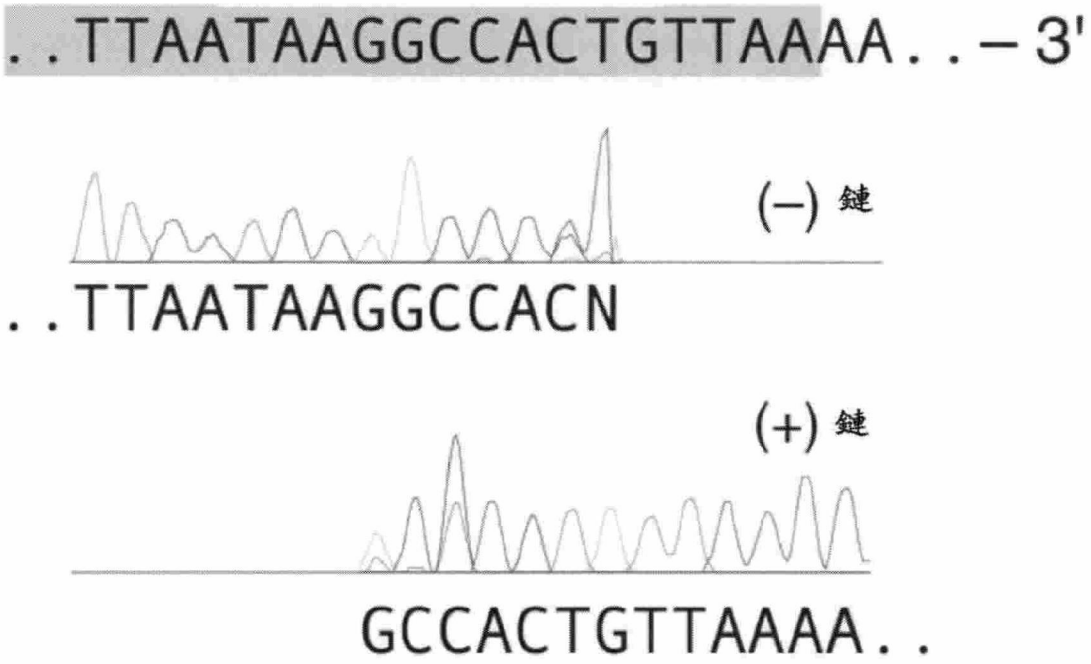

..TTAATAAGGCCACTGTTAAAA..–3'
(–) 鏈
..TTAATAAGGCCACN
(+) 鏈
GCCACTGTTAAAA..

圖 91

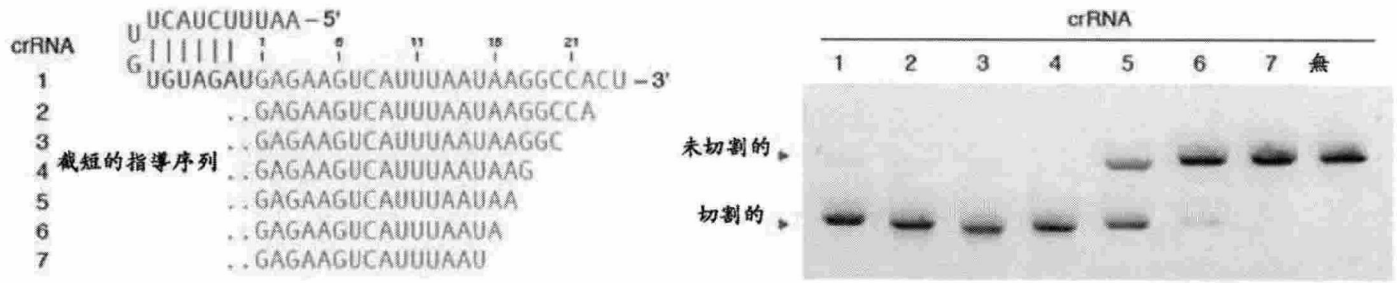

crRNA
U UCAUCUUUAA-5'
G UGUAGAUGAGAAGUCAUUUAAUAAGGCCACU-3'
1
2 ..GAGAAGUCAUUUAAUAAGGCCA
3 ..GAGAAGUCAUUUAAUAAGGC
4 截短的指導序列 ..GAGAAGUCAUUUAAUAAG
5 ..GAGAAGUCAUUUAAUAA
6 ..GAGAAGUCAUUUAAUA
7 ..GAGAAGUCAUUUAAU
crRNA
1 2 3 4 5 6 7 無
未切割的
切割的

圖 92

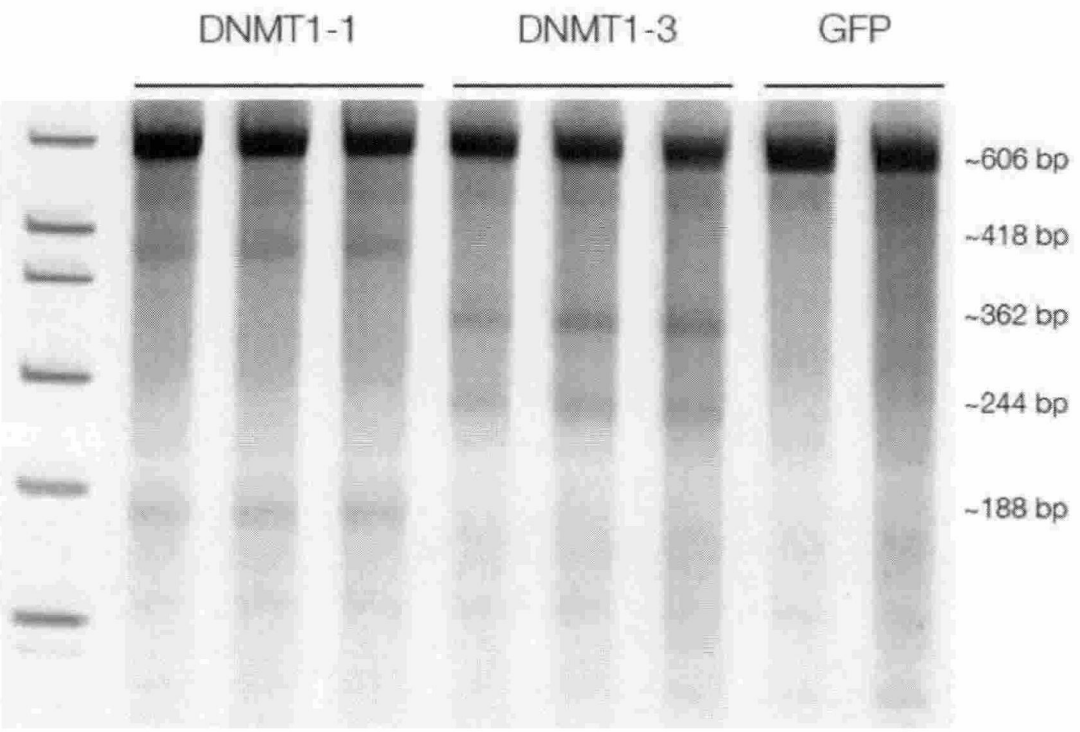


圖 93

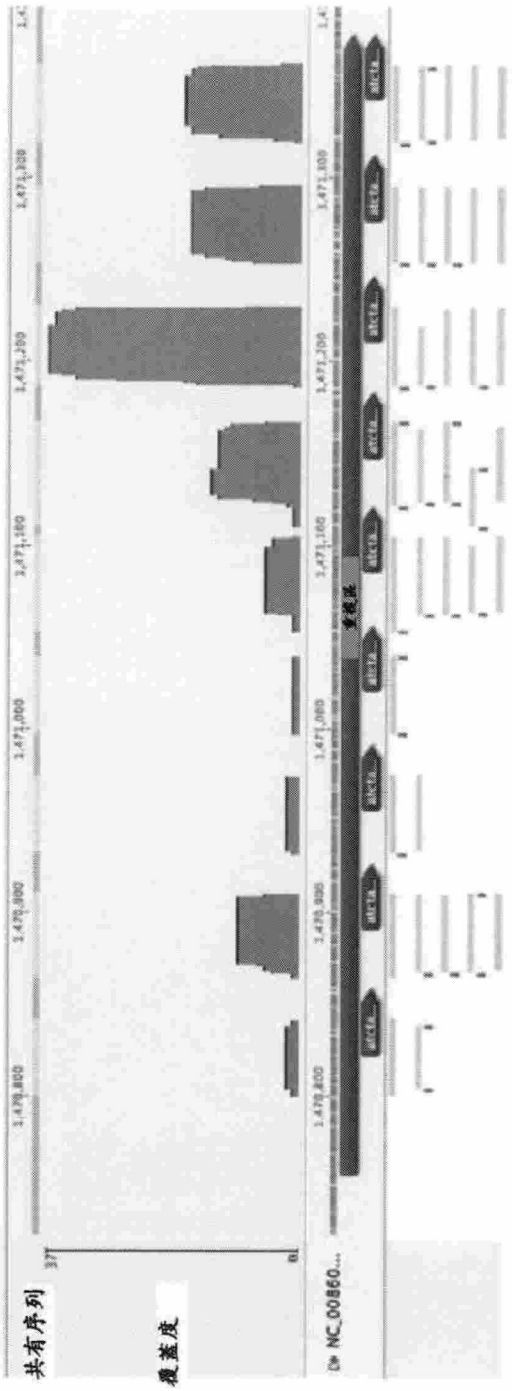


圖 94A

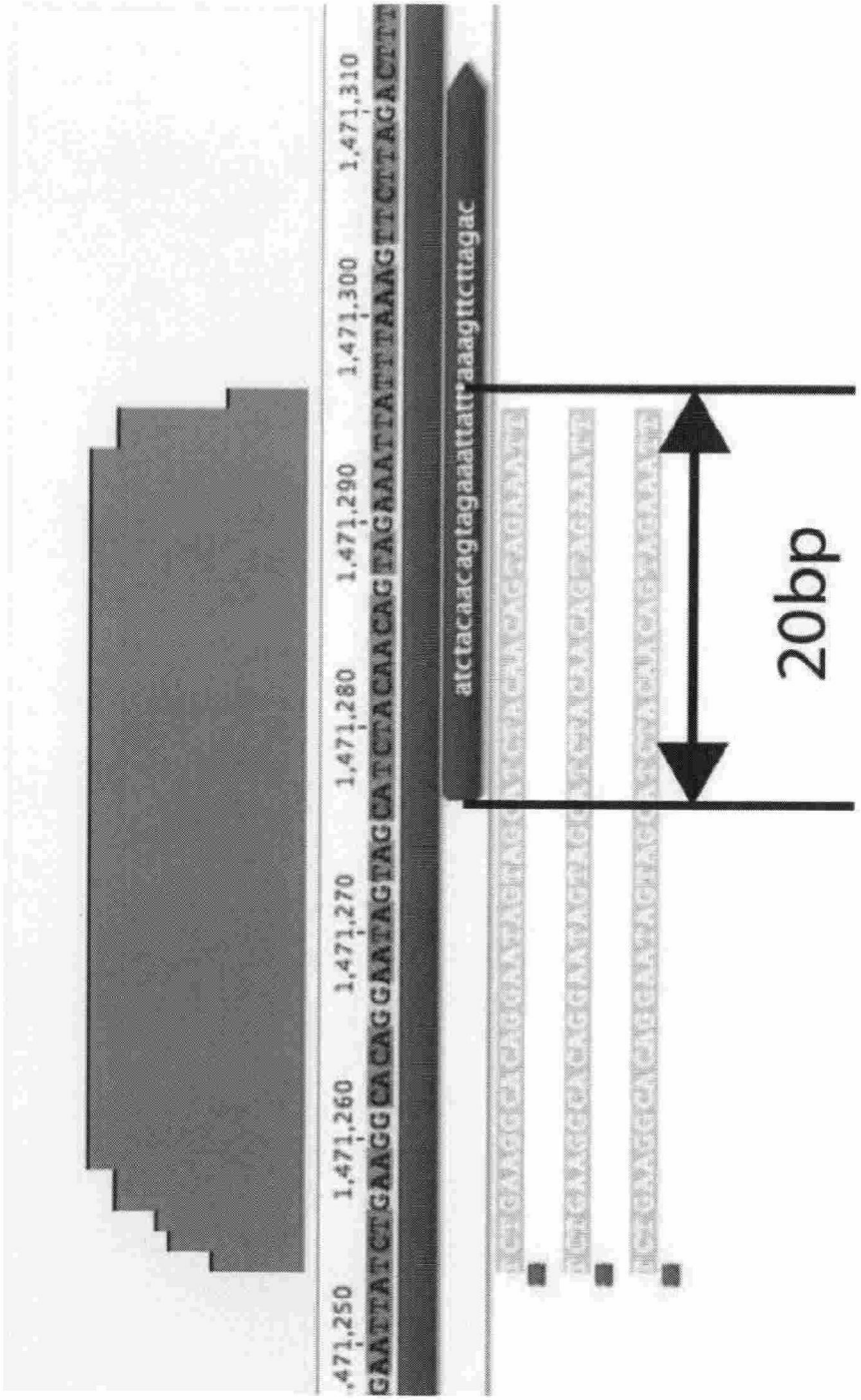

1,471,250
1,471,260
1,471,270
1,471,280
1,471,290
1,471,300
1,471,310
GAATTATCTGAAGGCACAGGAATAGTAGCATCTACAACAGTAGAAATTATTTAAGTTCTTAGACTTT
atctacaacagtagaaattatttaaagttcttagac
20bp

圖 94B

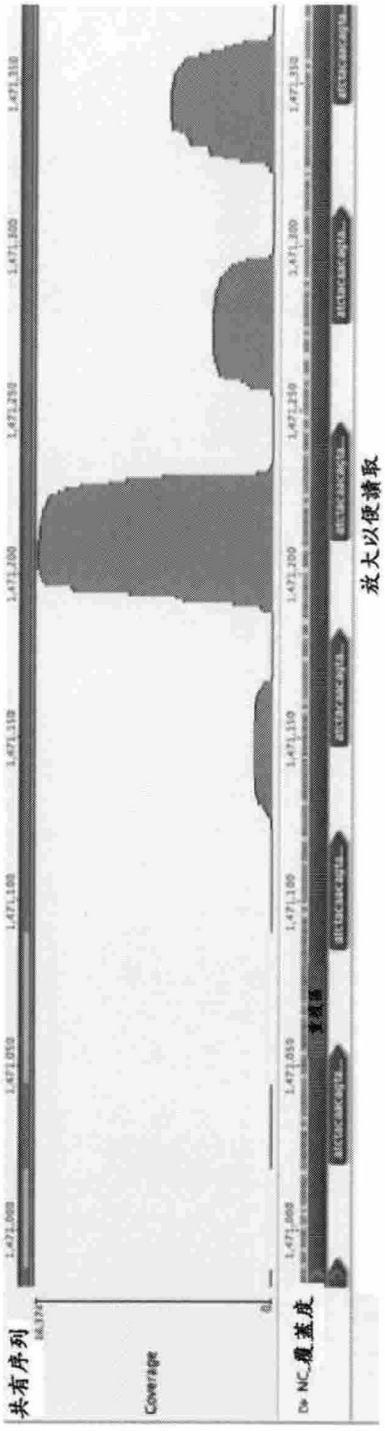


圖 94C

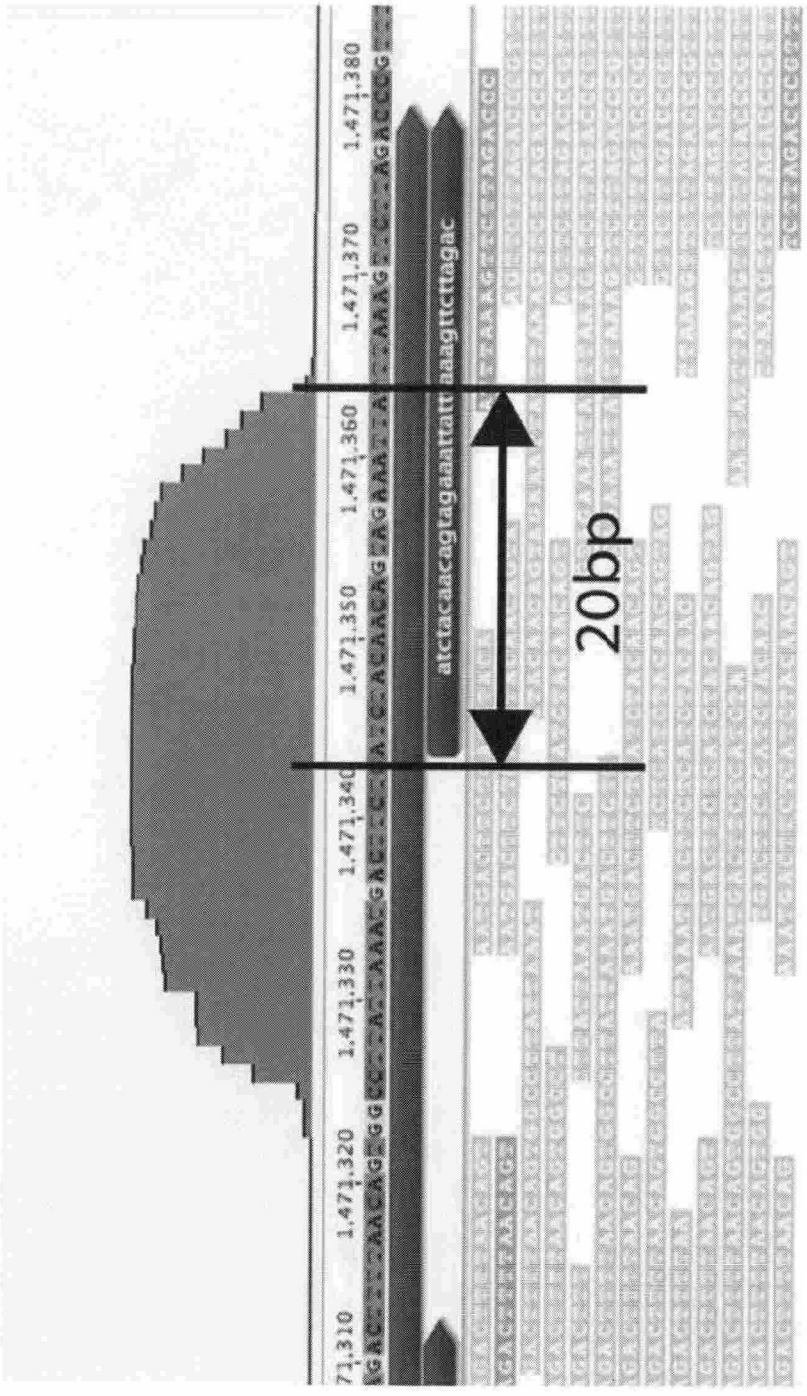

1,471,310
1,471,320
1,471,330
1,471,340
1,471,350
1,471,360
1,471,370
1,471,380
20bp

圖 94D

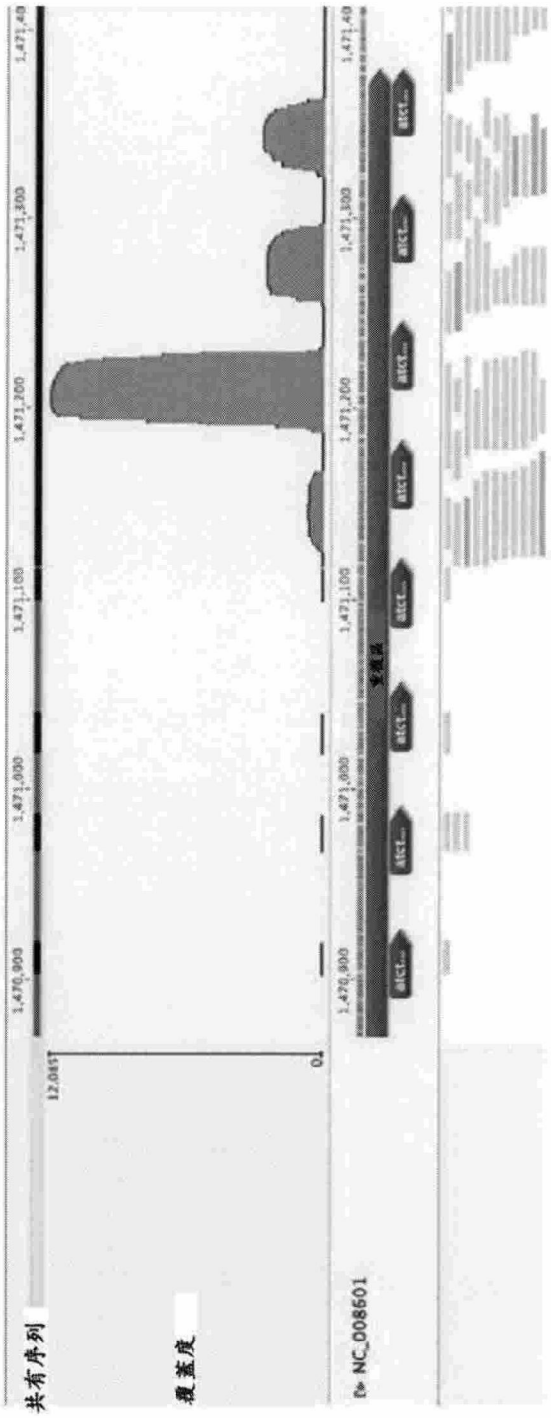


圖 94E

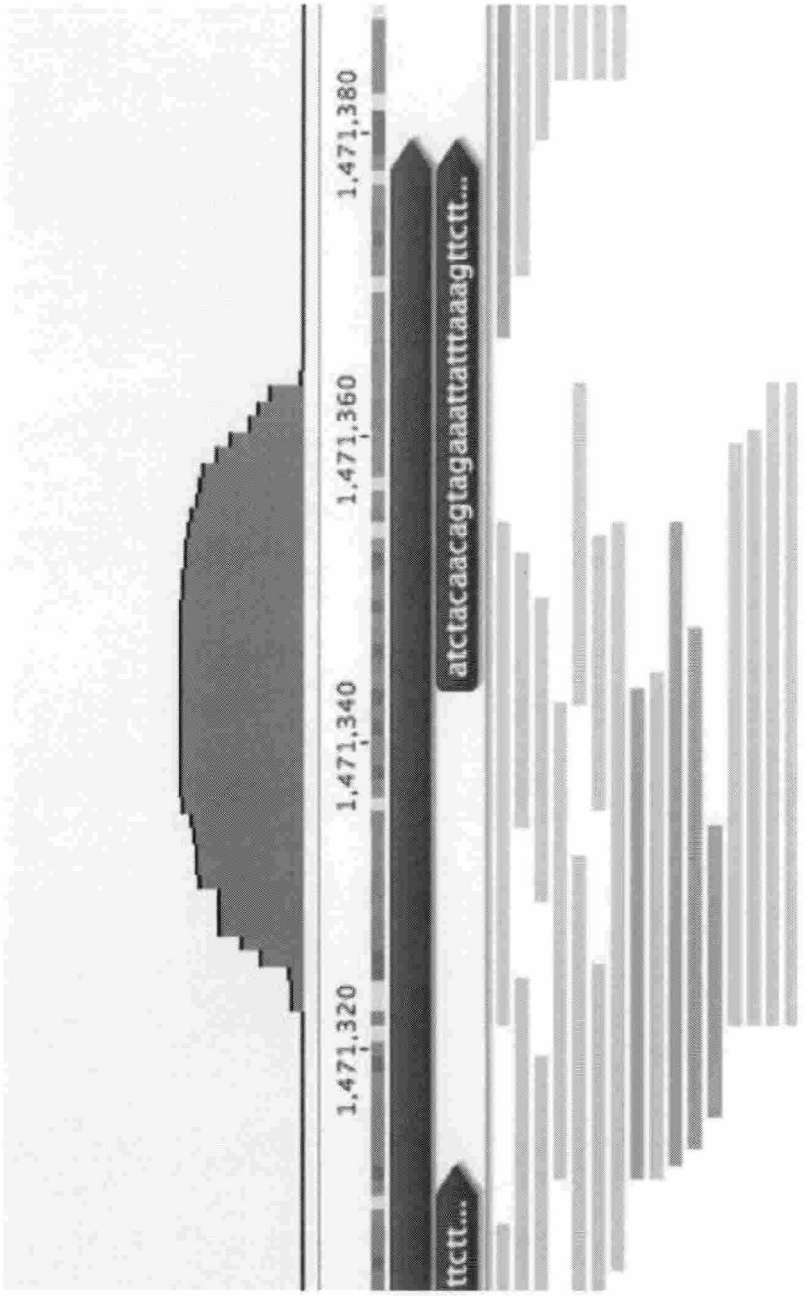

1,471,320
1,471,340
1,471,360
1,471,380
atctacaacagtagaaaattatttaaagttctt...
ttctt...

圖 94F

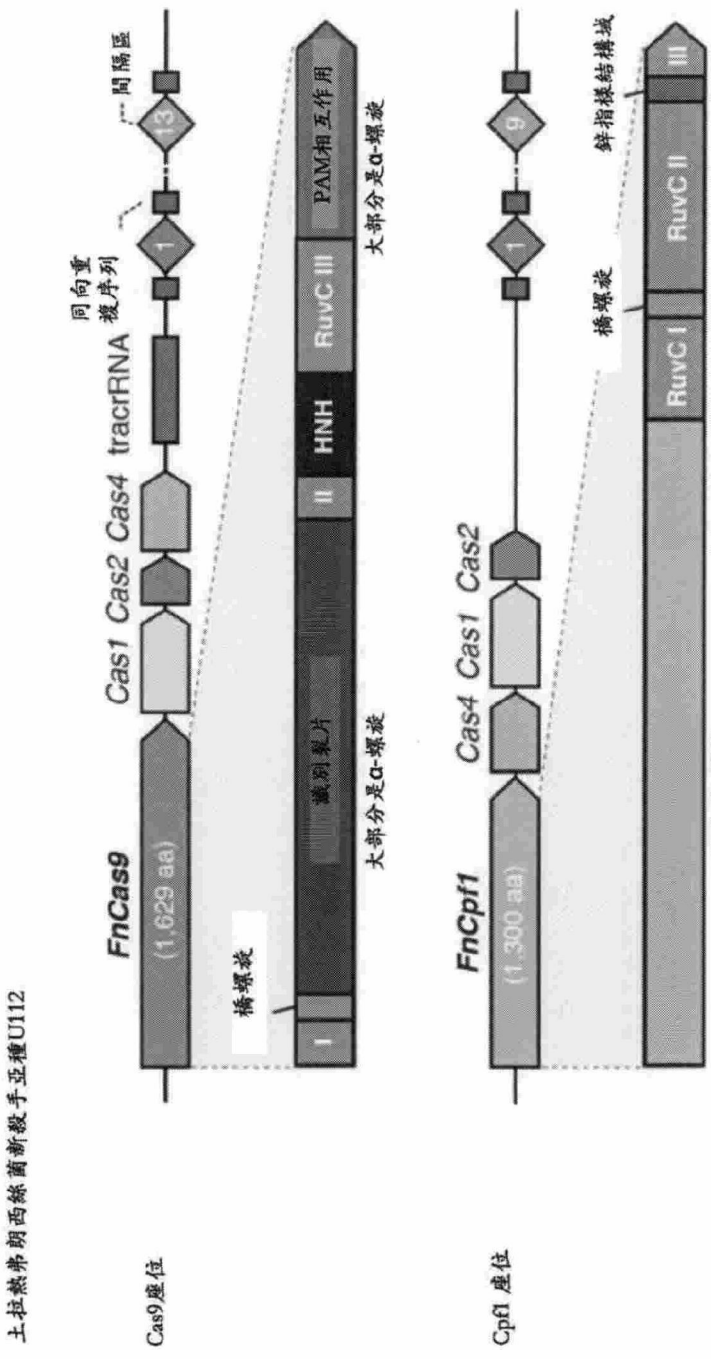

土拉熱弗朗西絲菌新殺手亞種U112
Cas9座位
FnCas9
(1,629 aa)
Cas1 Cas2 Cas4 tracrRNA
同向重複序列
間隔區
1
13
橋螺旋
1
藏匿裂片
II
HNH
RuvC III
PAM相互作用
大部分是α-螺旋
大部分是α-螺旋
Cpf1 座位
FnCpf1
(1,300 aa)
Cas4 Cas1 Cas2
1
9
橋螺旋
鋅指樣結構域
RuvC I
RuvC II
III
混合的α/β區

圖 95A

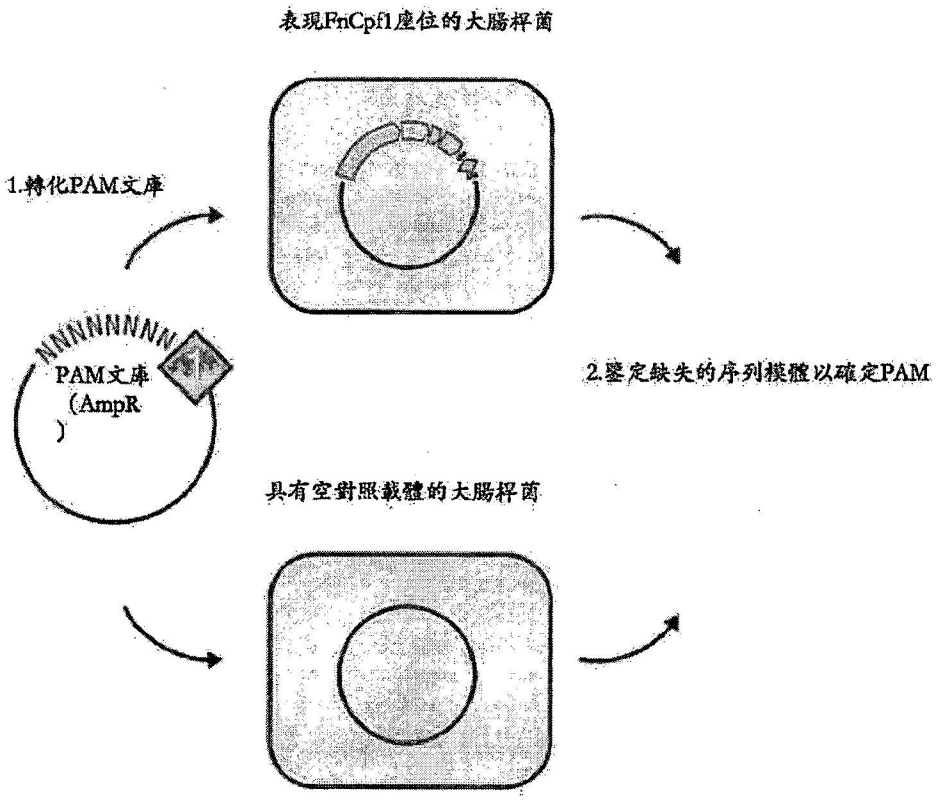

表現FnCpf1座位的大腸桿菌
1.轉化PAM文庫
NNNNNNNN
PAM文庫
(AmpR
)
2.鑒定缺失的序列模體以確定PAM
具有空對照載體的大腸桿菌

圖 95B

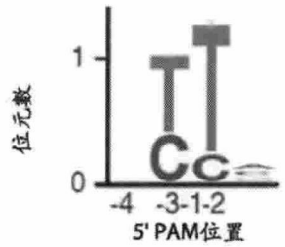


圖 95C

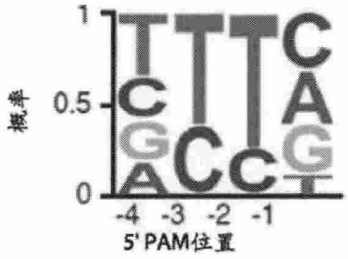


圖 95D

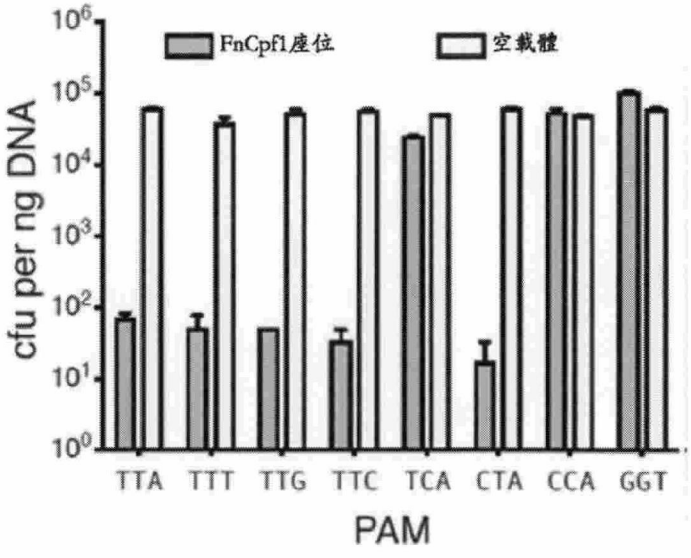


圖 95E

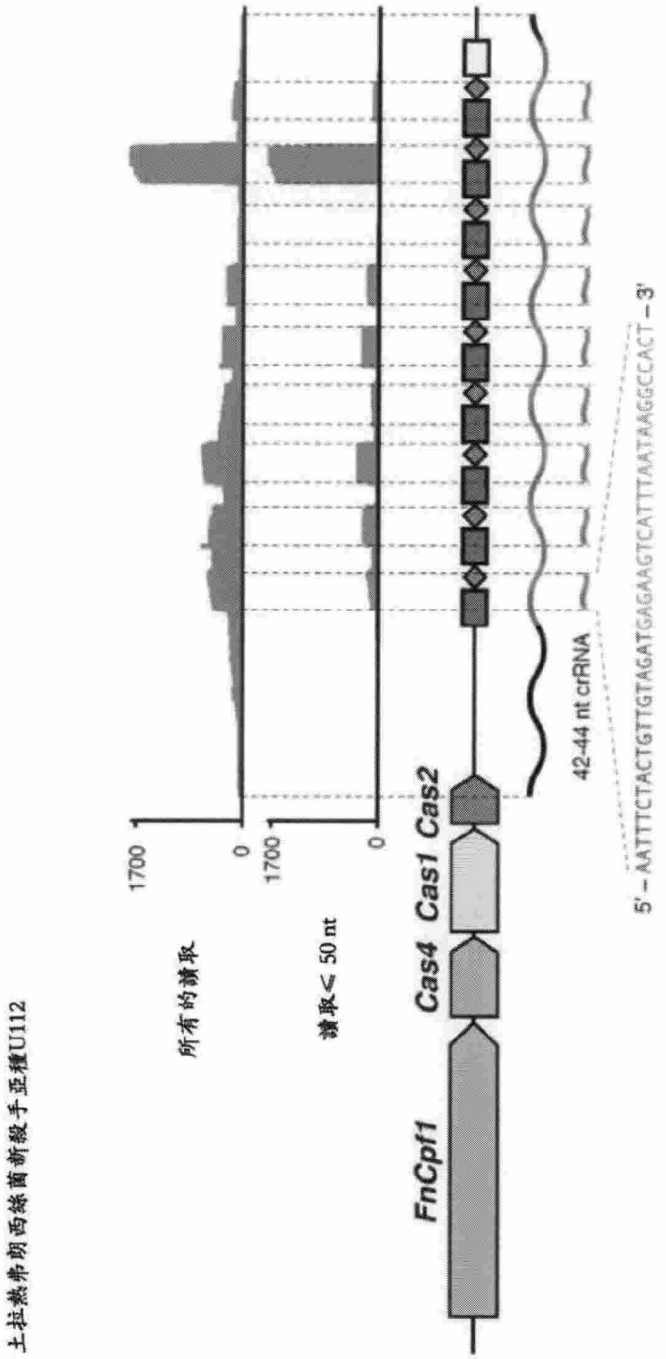

土拉热弗朗西丝菌新颖手亚型U112
所有的读取
1700
0
读取< 50 nt
1700
0
FnCpf1
Cas4
Cas1
Cas2
42-44 nt crRNA
5' – AATTTCTACTGTTGTAGATGAGAAGTCATTTAATAAGGCCACT – 3'

圖 96A

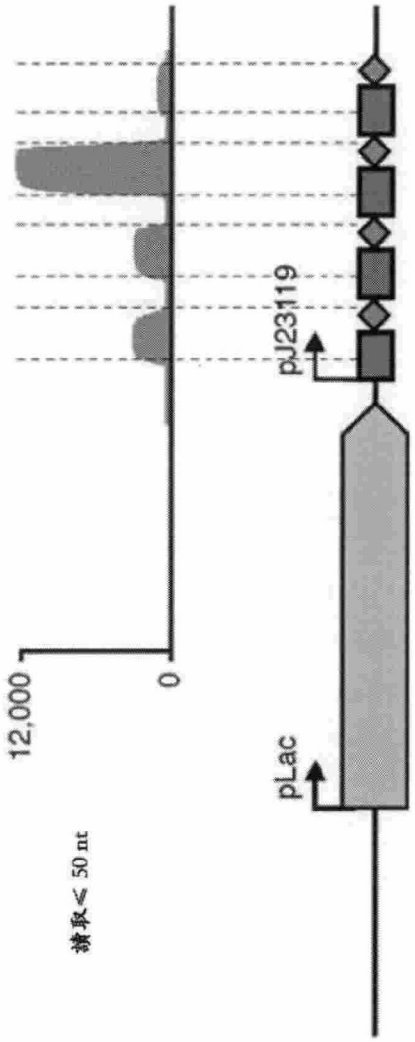


圖 96B

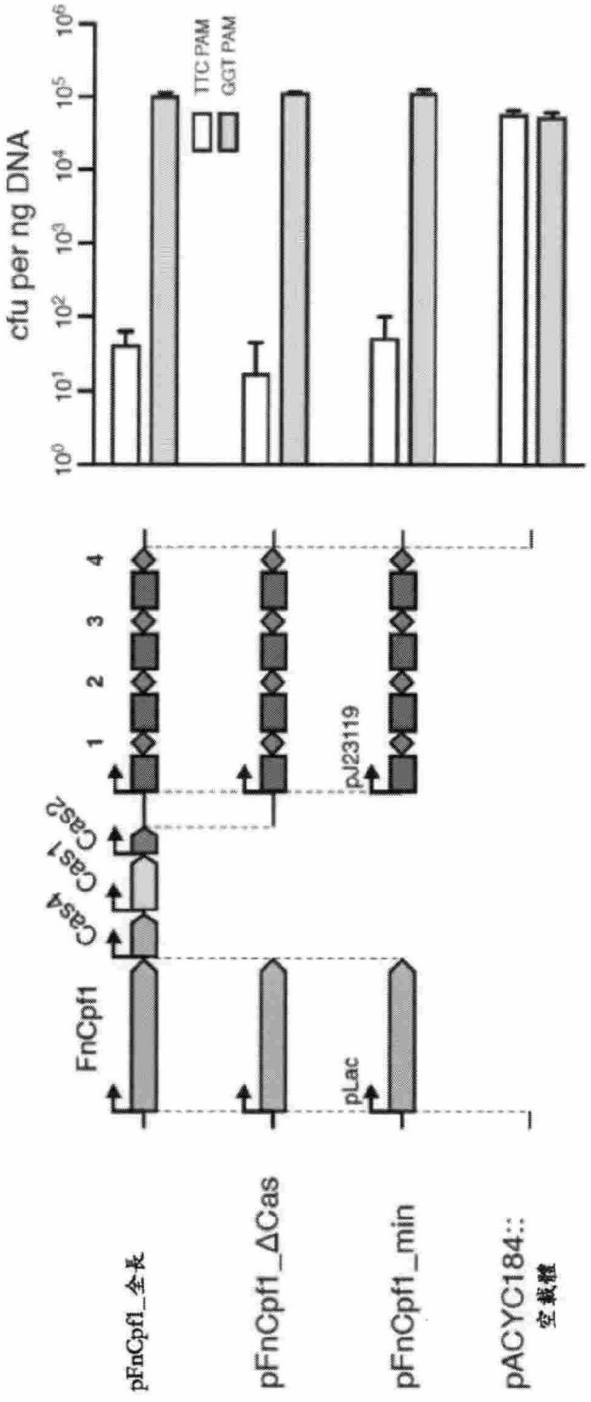

cfu per ng DNA
10^0
10^1
10^2
10^3
10^4
10^5
10^6
TTC PAM
GGT PAM
FnCpf1
Cas1
Cas2
Cas4
1
2
3
4
pLac
pJ23119
pFnCpf1_全长
pFnCpf1_ΔCas
pFnCpf1_min
pACYC184::
空载体

圖 96C

Cpf1
PAM
原型間隔區 (1)
5' – . . GGTACCCGGGGATCCTTTAGAGAAGTCATTTAATAAGGCCACTGTTAAAAAGCTTGGCGTAATCA . . – 3'
3' – . . CCATGGGCCCCTAGGAAATCTCTTCAGTAAATTATTCCGGTGACAATTTTTCGAACCGCATTAGT . . – 5'
DNA靶標
crRNA
U UGUAGAUGAGAAGUCAUUUAAUAAGGCCACU – 3'
G UCAUCUUUAA – 5'

圖 97A

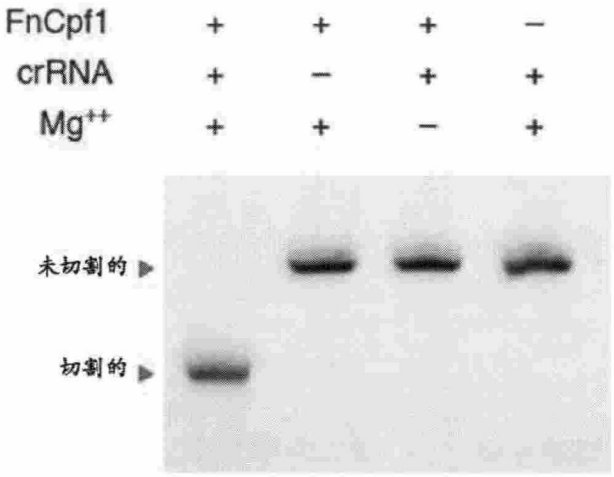

FnCpf1 + + + −
crRNA + − + +
Mg++ + + − +
未切割的
切割的

圖 97B

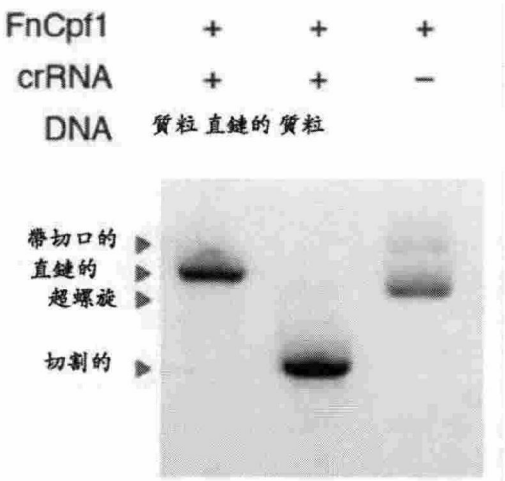

FnCpf1 + + +
crRNA + + −
DNA 質粒 直鏈的 質粒
帶切口的
直鏈的
超螺旋
切割的

圖 97C

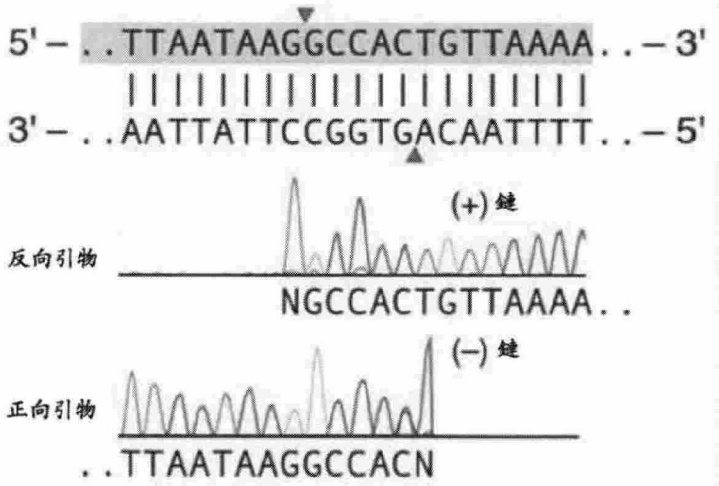

5' – . . TTAATAAGGCCACTGTTAAAA . . – 3'
3' – . . AATTATTCCGGTGACAATTTT . . – 5'
(+) 鏈
反向引物
NGCCACTGTTAAAA . .
(–) 鏈
正向引物
. . TTAATAAGGCCACN

圖 97D

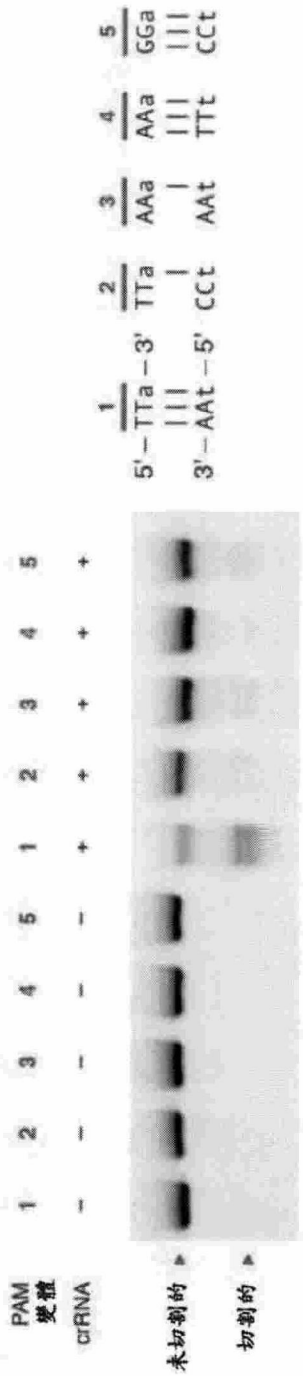

PAM
變體
crRNA
未切割的
切割的
1 2 3 4 5 1 2 3 4 5
– – – – – + + + + +
1 2 3 4 5
5' – TTa – 3'
3' – AAt – 5'
TTa
CCt
AAa
AAt
AAa
TTt
GGa
CCt

圖 97E

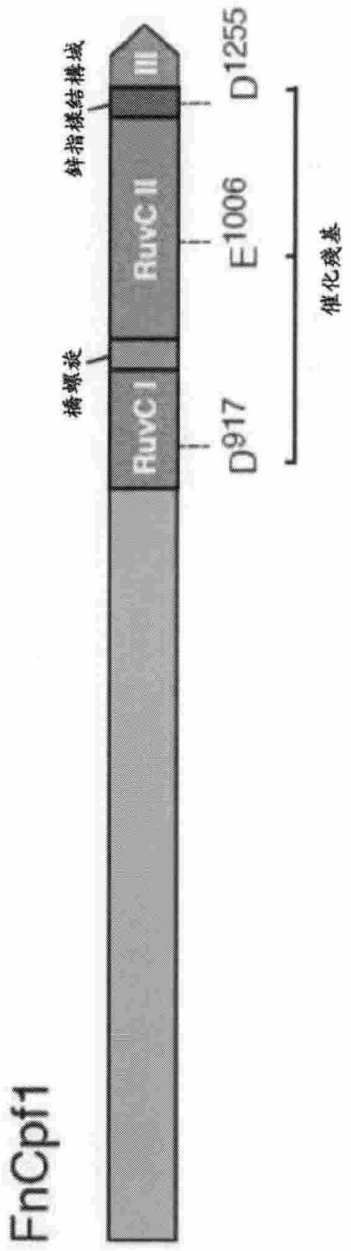


圖 98A

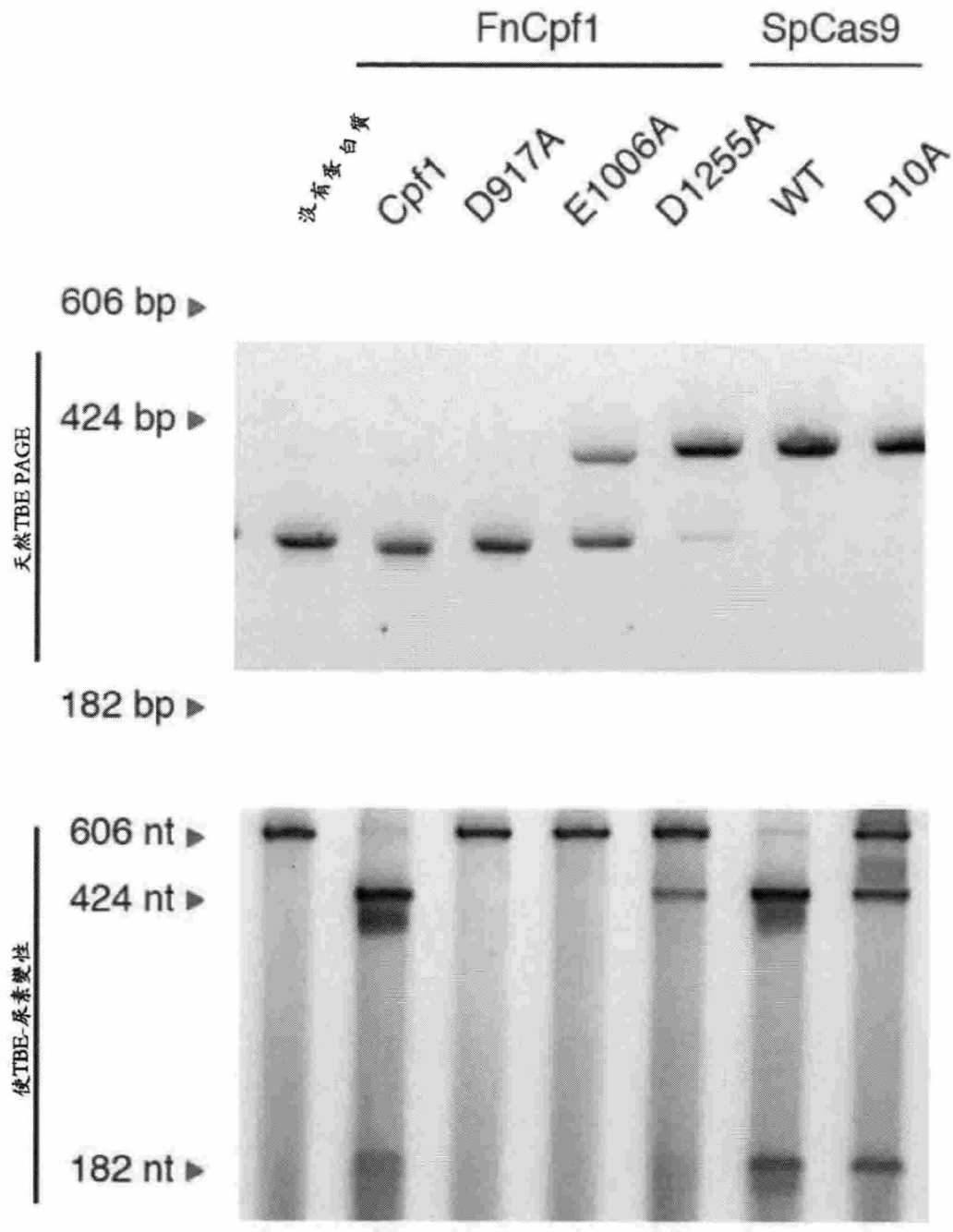
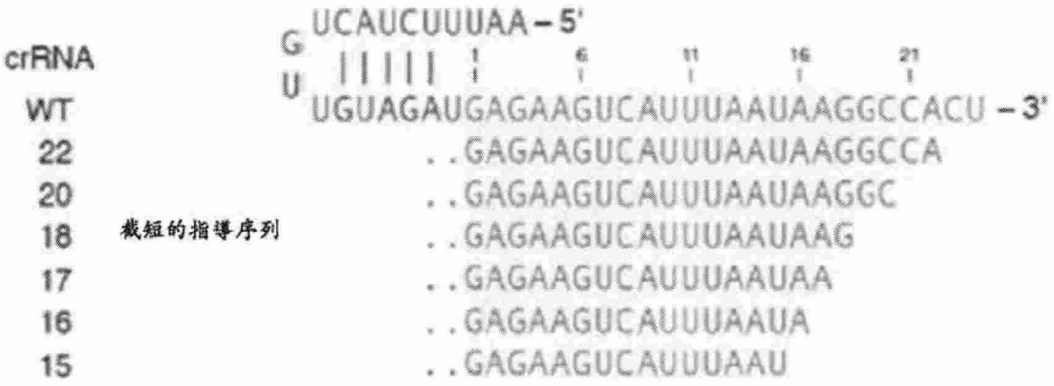

FnCpf1
SpCas9
沒有蛋白質
Cpf1
D917A
E1006A
D1255A
WT
D10A
606 bp
424 bp
182 bp
天然TBE PAGE
606 nt
424 nt
182 nt
使TBE-尿素變性

圖 98B


crRNA
WT
22
20
18
17
16
15
截短的指導序列
G
U
UCAUCUUUAA – 5'
1
6
11
16
21
UGUAGAUGAGAAGUCAUUUAAUAAGGCCACU – 3'
..GAGAAGUCAUUUAAUAAGGCCA
..GAGAAGUCAUUUAAUAAGGC
..GAGAAGUCAUUUAAUAAG
..GAGAAGUCAUUUAAUAA
..GAGAAGUCAUUUAAUA
..GAGAAGUCAUUUAAU

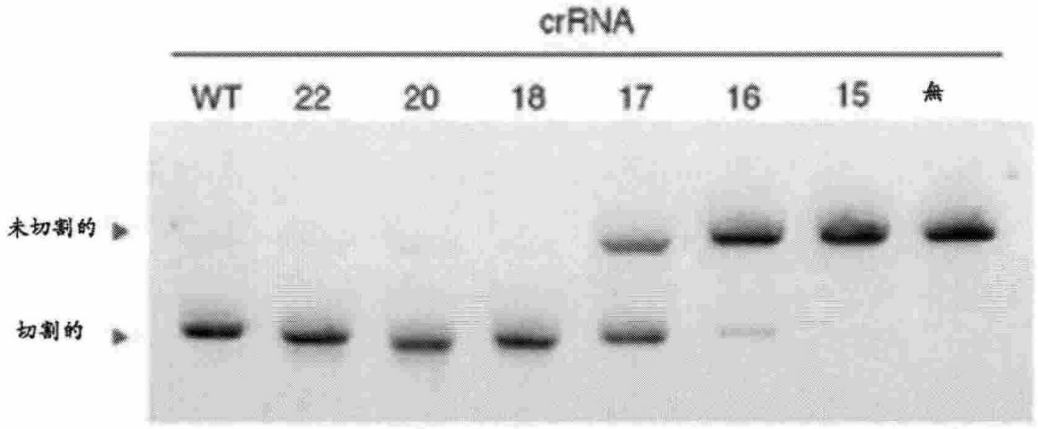
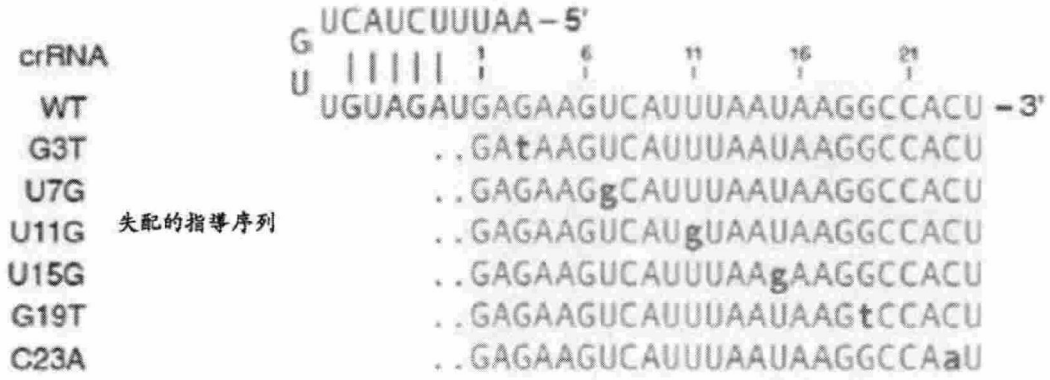

crRNA
WT
22
20
18
17
16
15
無
未切割的
切割的

圖 99A


crRNA
G
U
UCAUCUUUAA-5'
1
6
11
16
21
WT
UGUAGAUGAGAAGUCAUUUAAUAAGGCCACU-3'
G3T
..GAtAAGUCAUUUAAUAAGGCCACU
U7G
..GAGAAGgCAUUUAAUAAGGCCACU
U11G
失配的指導序列
..GAGAAGUCAUgUAAUAAGGCCACU
U15G
..GAGAAGUCAUUUAAgAAGGCCACU
G19T
..GAGAAGUCAUUUAAUAAGtCCACU
C23A
..GAGAAGUCAUUUAAUAAGGCCAaU

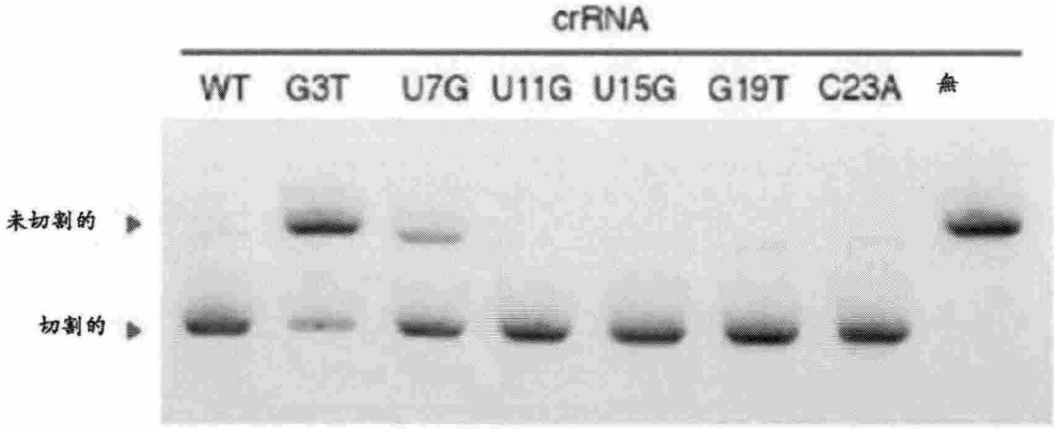

crRNA
WT
G3T
U7G
U11G
U15G
G19T
C23A
無
未切割的
切割的

圖 99B

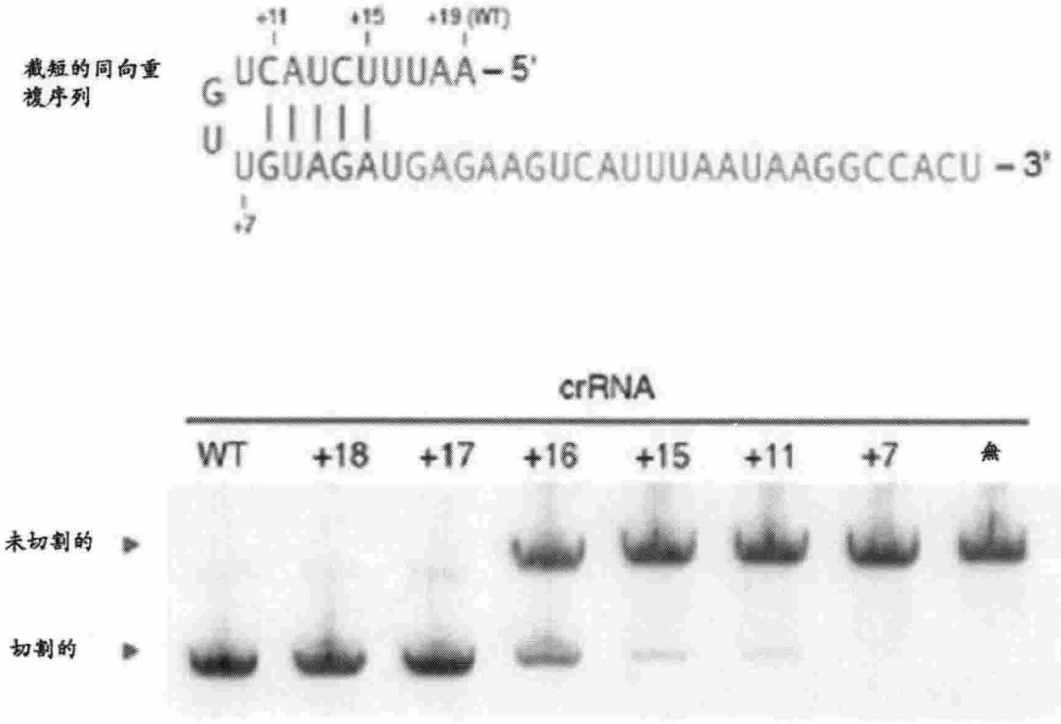

截短的同向重複序列
+11
+15
+19 (WT)
UCAUCUUUAA – 5'
G
U
UGUAGAUGAGAAGUCAUUUAAUAAGGCCACU – 3'
+7
crRNA
WT
+18
+17
+16
+15
+11
+7
無
未切割的
切割的

圖 99C

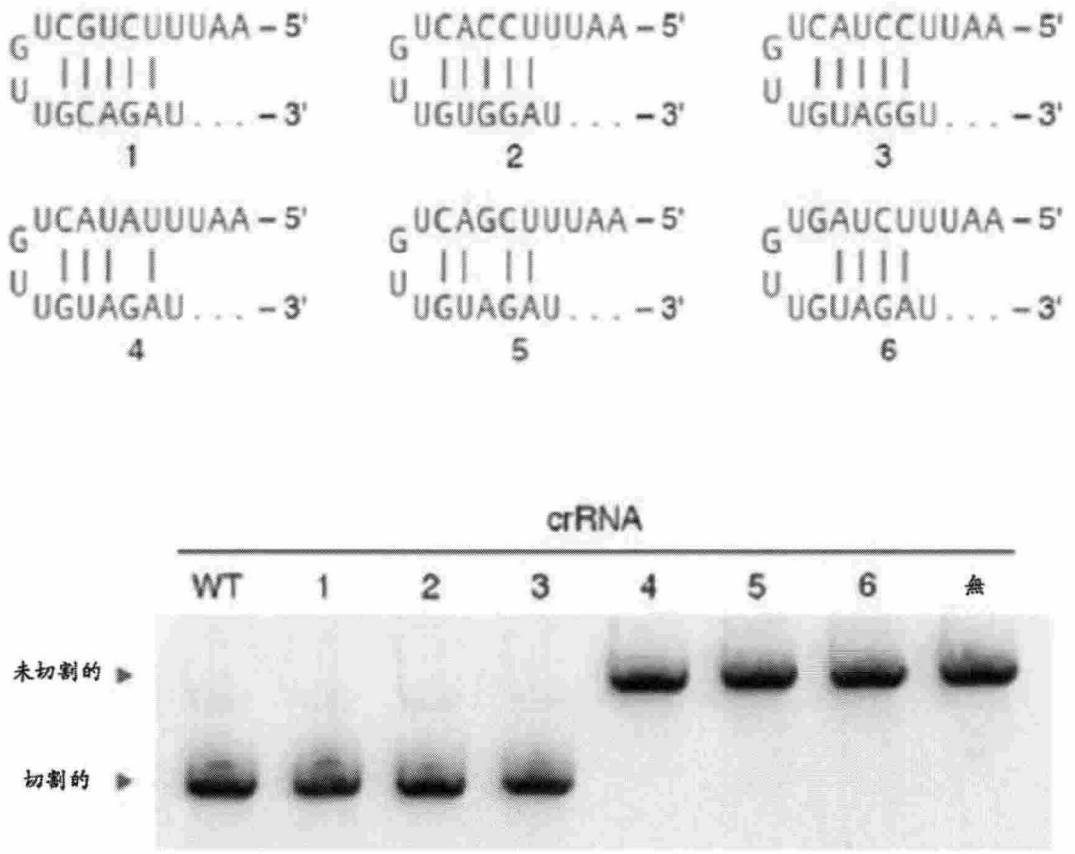

G UCGUCUUUAA – 5'
U UGCAGAU . . . – 3'
1
G UCACCUUUAA – 5'
U UGUGGAU . . . – 3'
2
G UCAUCCUUAA – 5'
U UGUAGGU . . . – 3'
3
G UCAUAUUUAA – 5'
U UGUAGAU . . . – 3'
4
G UCAGCUUUAA – 5'
U UGUAGAU . . . – 3'
5
G UGAUCUUUAA – 5'
U UGUAGAU . . . – 3'
6
crRNA
WT
1
2
3
4
5
6
無
未切割的
切割的

圖 99D

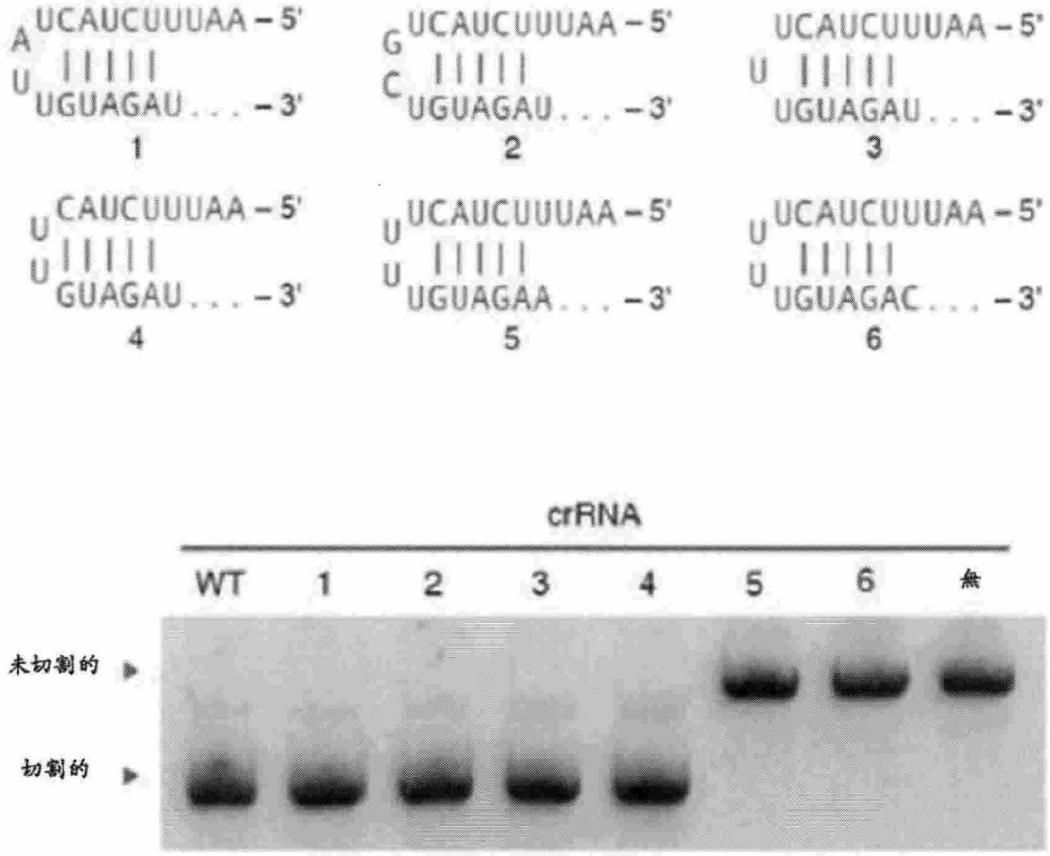

A UCAUCUUUAA – 5'
U |||||
UGUAGAU . . . – 3'
1
G UCAUCUUUAA – 5'
C |||||
UGUAGAU . . . – 3'
2
UCAUCUUUAA – 5'
U |||||
UGUAGAU . . . – 3'
3
U CAUCUUUAA – 5'
U |||||
GUAGAU . . . – 3'
4
U UCAUCUUUAA – 5'
U |||||
UGUAGAA . . . – 3'
5
U UCAUCUUUAA – 5'
U |||||
UGUAGAC . . . – 3'
6
crRNA
WT
1
2
3
4
5
6
無
未切割的
切割的

圖 99E

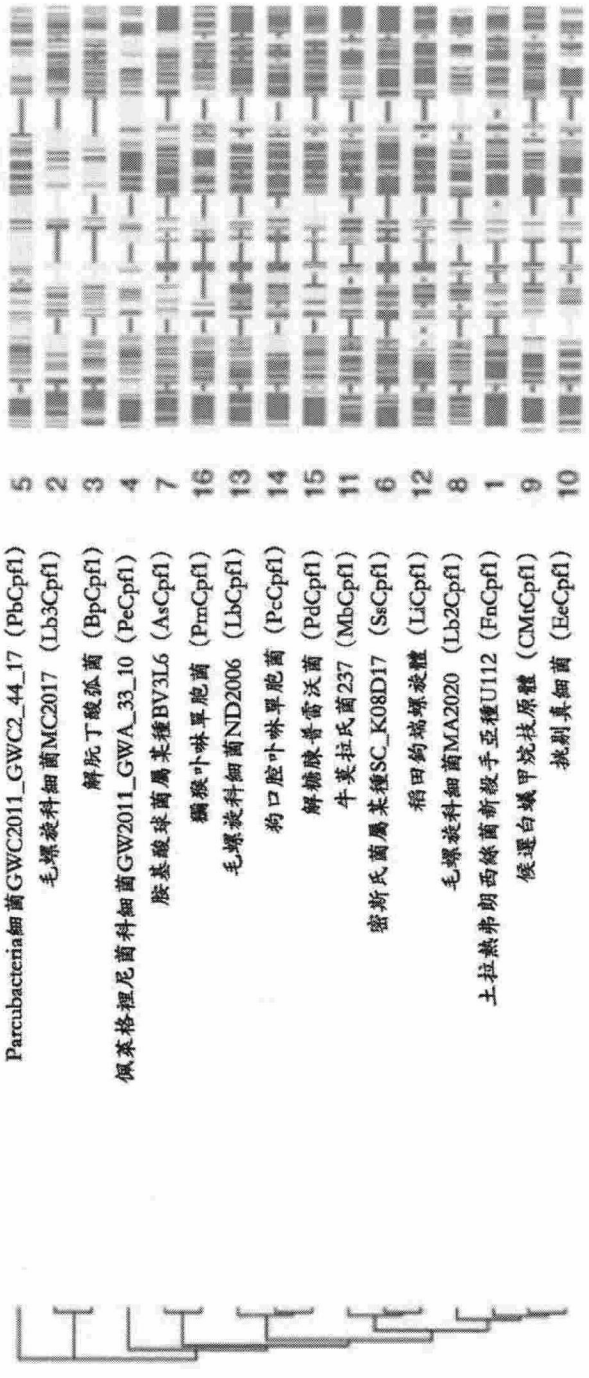

ID
Parcubacteria細菌GWC2011_GWC2_44_17（PbCpf1）
毛螺旋科細菌MC2017（Lb3Cpf1）
解阮丁酸弧菌（BpCpf1）
佩萊格裡尼菌科細菌GW2011_GWA_33_10（PeCpf1）
胺基酸球菌屬某種BV3L6（AsCpf1）
猶猴卟啉單胞菌（PmCpf1）
毛螺旋科細菌ND2006（LbCpf1）
約口腔卟啉單胞菌（PcCpf1）
解糖胨普雷沃菌（PdCpf1）
牛莫拉氏菌237（MbCpf1）
密斯氏菌屬某種SC_K08D17（SsCpf1）
稻田約瑞塚菌（LiCpf1）
毛螺旋科細菌MA2020（Lb2Cpf1）
土拉熱弗朗西絲菌新級手亞種U112（FnCpf1）
候選白堿甲烷拉原體（CMtCpf1）
挑剔真細菌（EeCpf1）
5
2
3
4
7
16
13
14
15
11
6
12
8
1
9
10

圖 100A

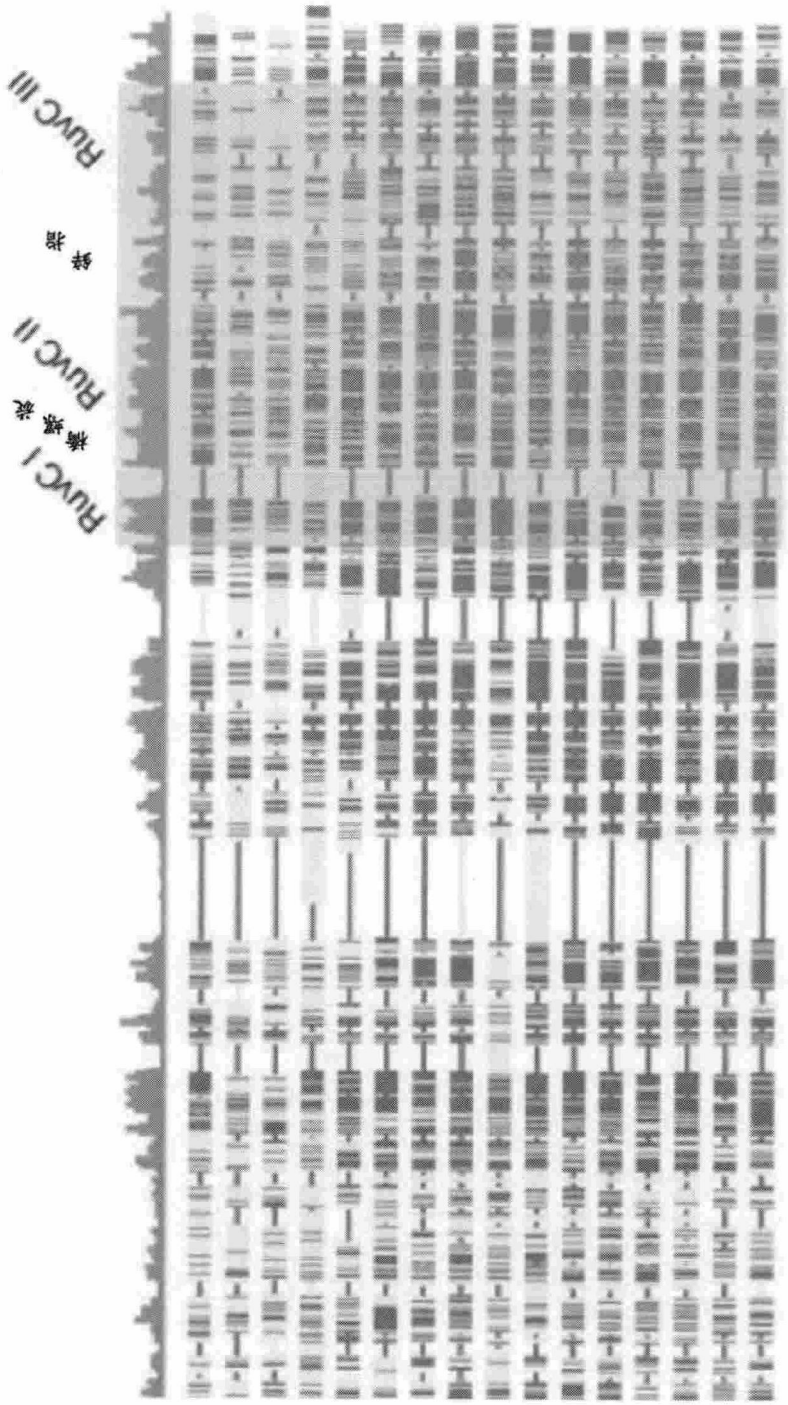


圖 100B

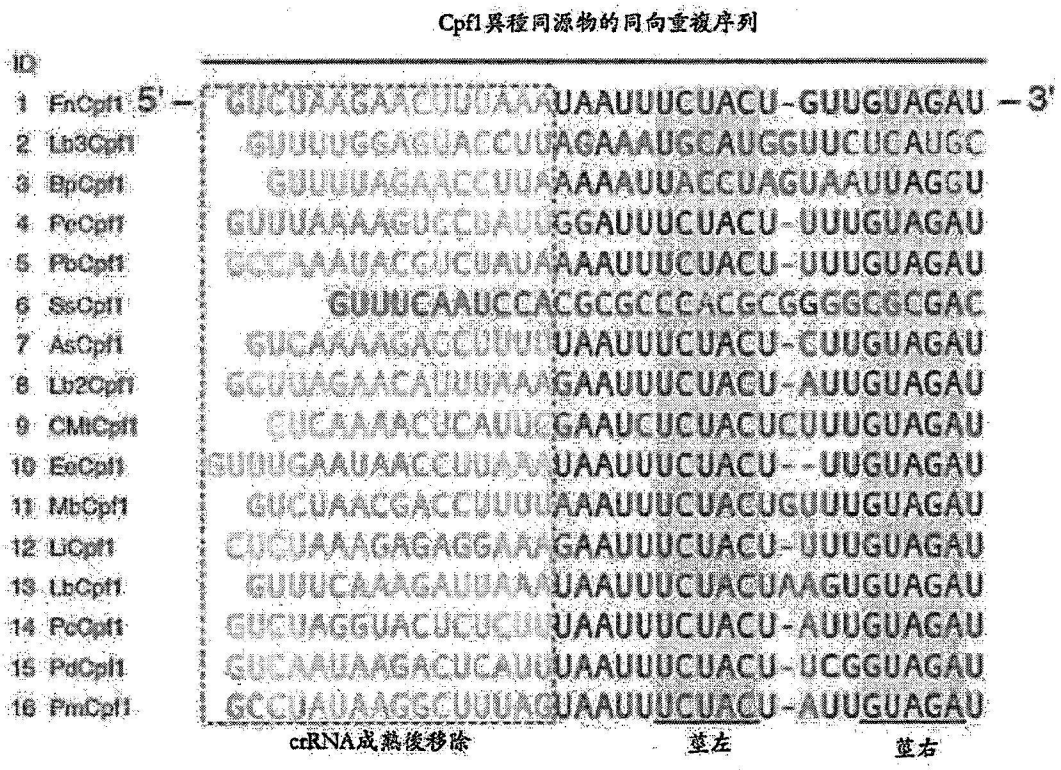

Cpf1異種同源物的同向重複序列
ID
1 FnCpf1 5'
2 Lb3Cpf1
3 BpCpf1
4 PeCpf1
5 PbCpf1
6 SsCpf1
7 AsCpf1
8 Lb2Cpf1
9 CMtCpf1
10 EeCpf1
11 MbCpf1
12 LiCpf1
13 LbCpf1
14 PcCpf1
15 PdCpf1
16 PmCpf1
3'
crRNA成熟後移除
莖左
莖右

圖 100C

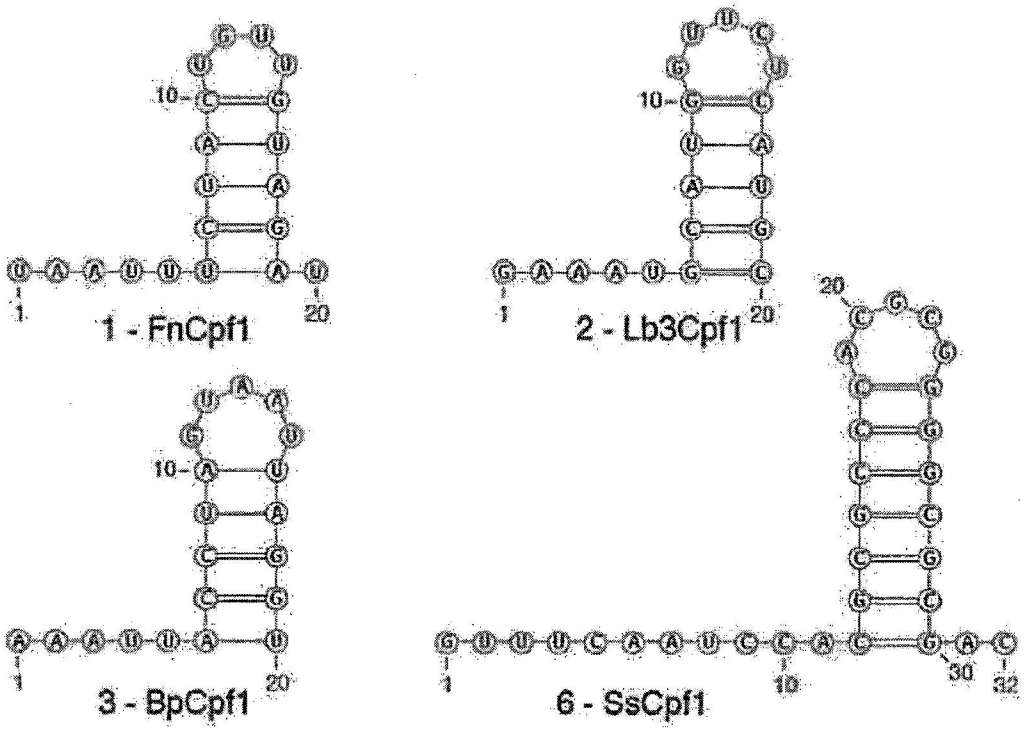


圖 100D

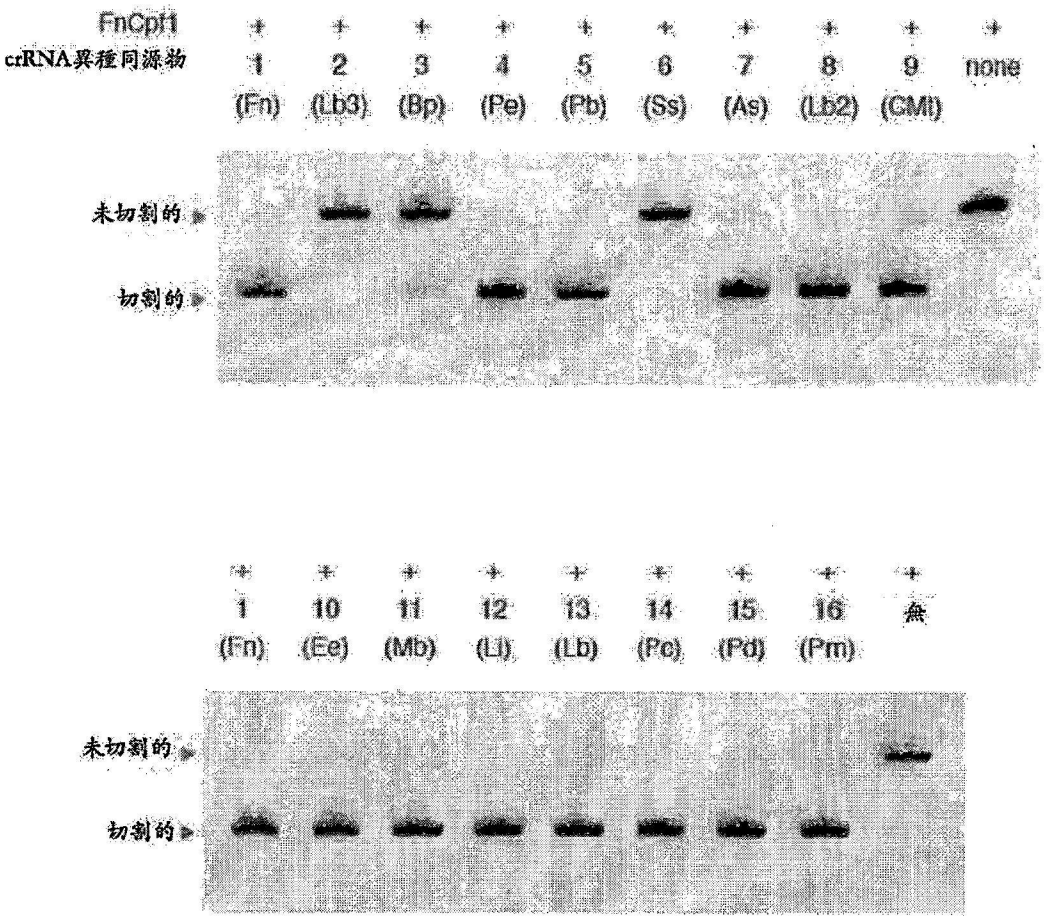

FnCpf1
crRNA異種同源物
1 (Fn)
2 (Lb3)
3 (Bp)
4 (Pe)
5 (Pb)
6 (Ss)
7 (As)
8 (Lb2)
9 (CMt)
none
未切割的
切割的
1 (Fn)
10 (Ee)
11 (Mb)
12 (Li)
13 (Lb)
14 (Pc)
15 (Pd)
16 (Pm)
無
未切割的
切割的

圖 100E

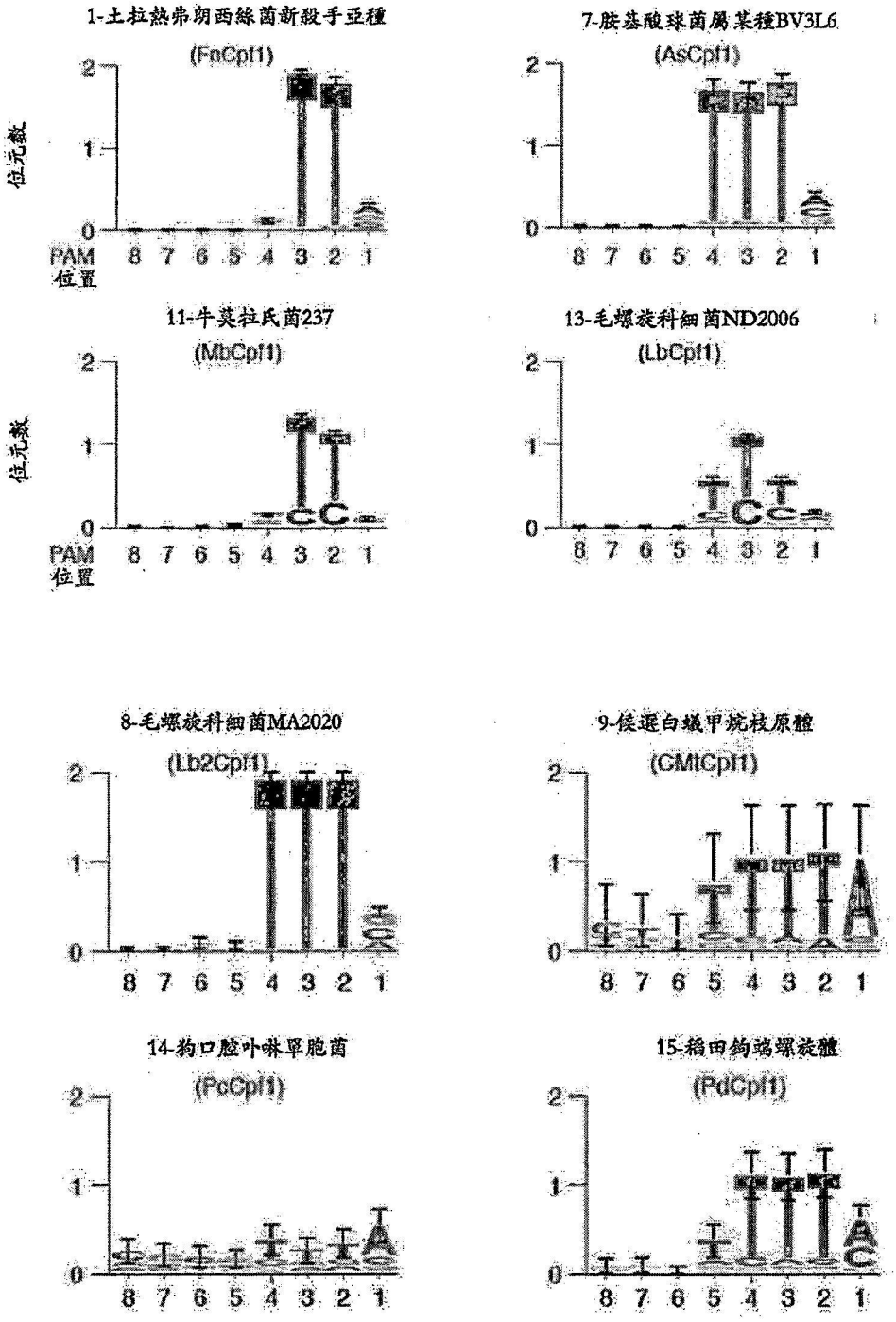
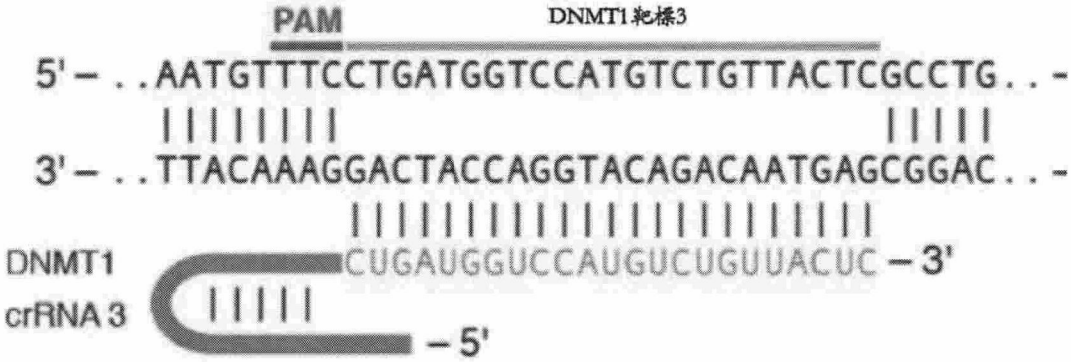

1-土拉熱弗朗西絲菌新殺手亞種
(FnCpf1)
7-胺基酸球菌屬某種BV3L6
(AsCpf1)
11-牛莫拉氏菌237
(MbCpf1)
13-毛螺旋科細菌ND2006
(LbCpf1)
8-毛螺旋科細菌MA2020
(Lb2Cpf1)
9-候選白蟻甲烷枝原體
(CMtCpf1)
14-狗口腔卟啉單胞菌
(PcCpf1)
15-稻田鉤端螺旋體
(PdCpf1)
位元數
PAM
位置
8 7 6 5 4 3 2 1
0
1
2

圖 100F


crRNA
密碼子優化的Cpf1
NLS
pCpf1

圖 101A


PAM
DNMT1靶標3
5'– . .AATGTTTCCTGATGGTCCATGTCTGTTACTCGCCTG. . -
3'– . .TTACAAAGGACTACCAGGTACAGACAATGAGCGGAC. . -
DNMT1
CUGAUGGUCCAUGUCUGUUACUC – 3'
crRNA 3
– 5'

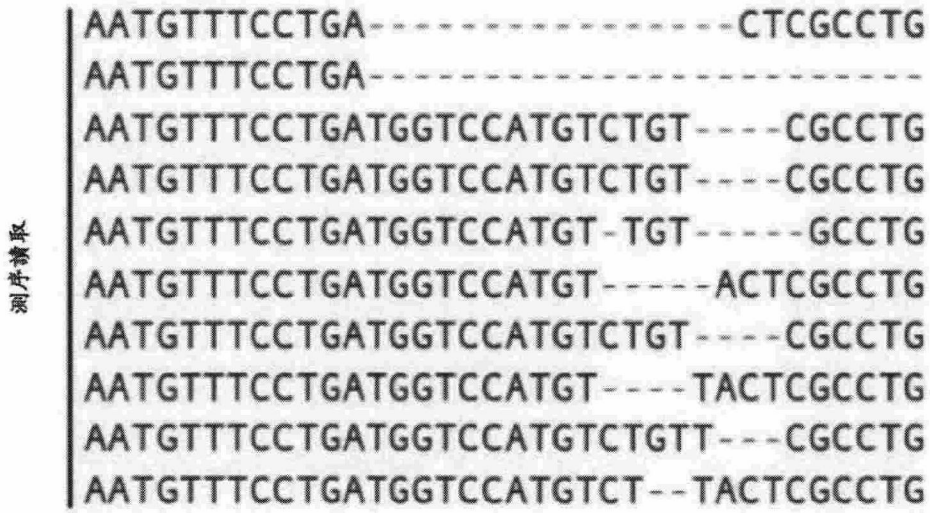

測序讀取
AATGTTTCCTGA----------CTCGCCTG
AATGTTTCCTGA----------
AATGTTTCCTGATGGTCCATGTCTGT----CGCCTG
AATGTTTCCTGATGGTCCATGTCTGT----CGCCTG
AATGTTTCCTGATGGTCCATGT-TGT-----GCCTG
AATGTTTCCTGATGGTCCATGT-----ACTCGCCTG
AATGTTTCCTGATGGTCCATGTCTGT----CGCCTG
AATGTTTCCTGATGGTCCATGT----TACTCGCCTG
AATGTTTCCTGATGGTCCATGTCTGTT---CGCCTG
AATGTTTCCTGATGGTCCATGTCT--TACTCGCCTG

圖 101B

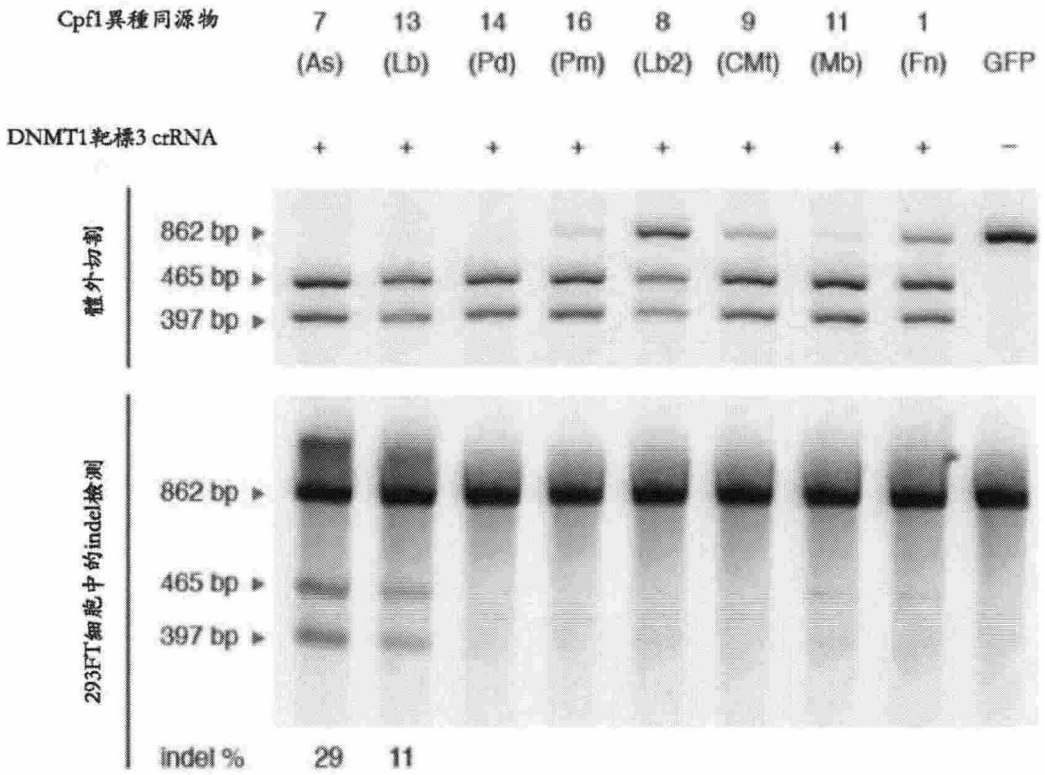

Cpf1異種同源物
7 (As)
13 (Lb)
14 (Pd)
16 (Pm)
8 (Lb2)
9 (CMt)
11 (Mb)
1 (Fn)
GFP
DNMT1靶標3 crRNA
+ + + + + + + + −
體外切割
862 bp
465 bp
397 bp
293FT細胞中的indel檢測
862 bp
465 bp
397 bp
indel %
29
11

圖 101C


PAM
指導序列
Cpf1
DNMT1
靶標
位點
1 TTTCCCTCACTCCTGCTCGGTGAATTT
2 TTTGAGGAGTGTTCAGTCTCCGTGAAC
3 TTTCCTGATGGTCCATGTCTGTTACTC
4 TTTATTTCCCTTCAGCTAAAATAAAGG

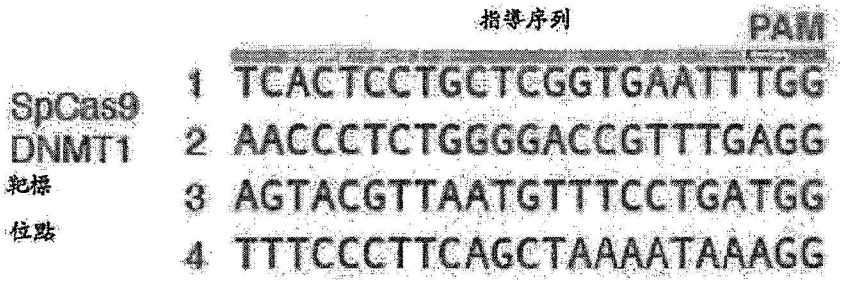

指導序列
PAM
SpCas9
DNMT1
靶標
位點
1 TCACTCCTGCTCGGTGAATTTGG
2 AACCCTCTGGGGACCGTTTGAGG
3 AGTACGTTAATGTTTCCTGATGG
4 TTTCCCTTCAGCTAAAATAAAGG

圖 101D

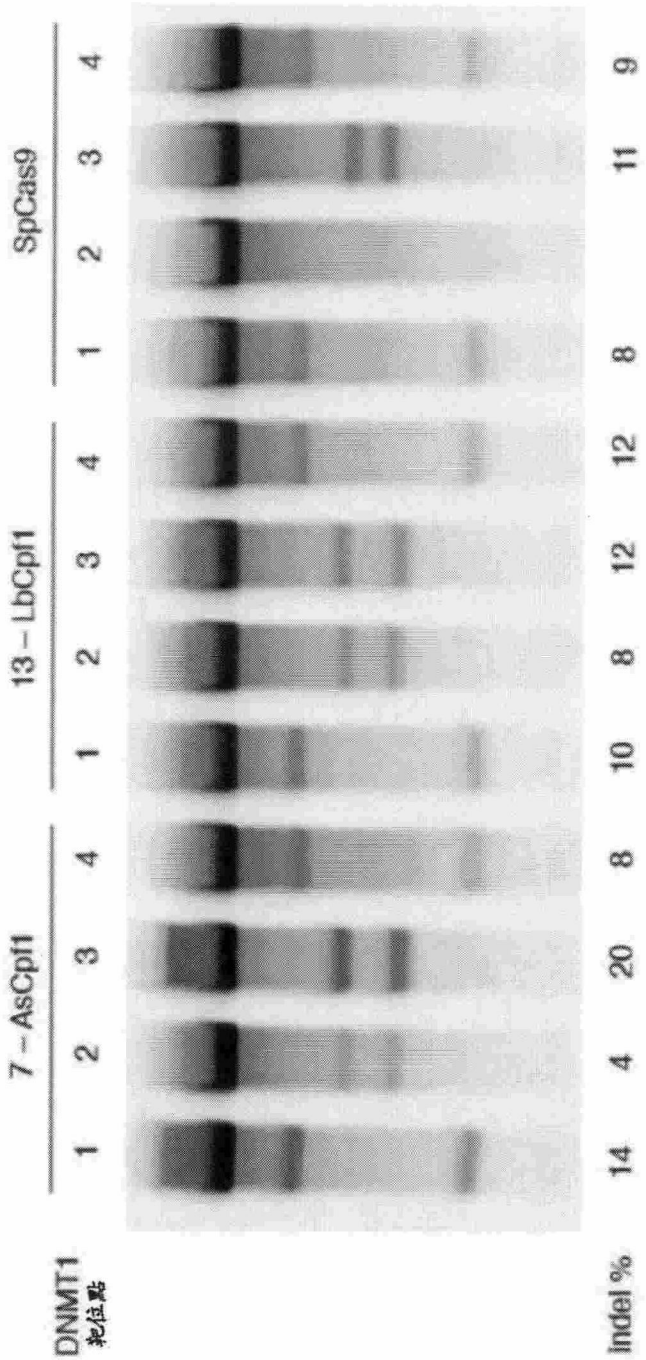

DNMT1
靶位點
7 – AsCpf1
1
2
3
4
13 – LbCpf1
1
2
3
4
SpCas9
1
2
3
4
indel %
14
4
20
8
10
8
12
12
8
11
9

圖 101E

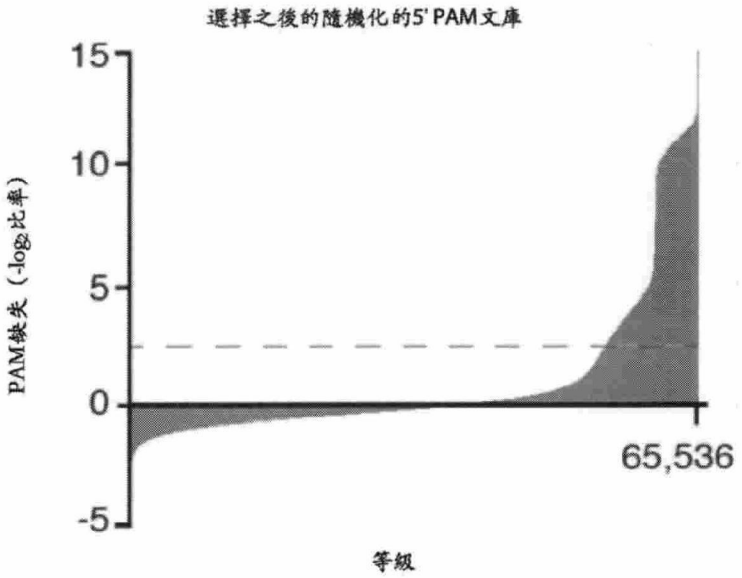

選擇之後的隨機化的5' PAM文庫
15
10
5
0
-5
PAM缺失（-log2比率）
65,536
等級

圖 102A

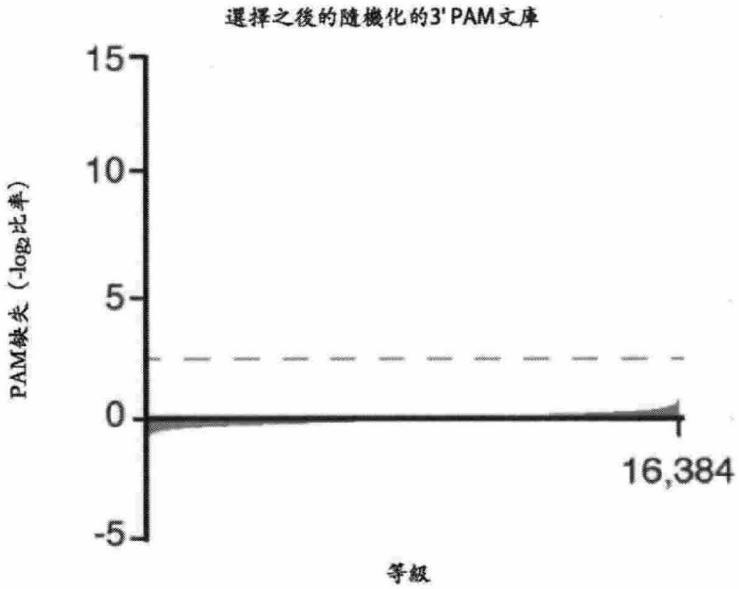

選擇之後的隨機化的3' PAM文庫
15
10
5
0
-5
PAM缺失（-log2比率）
16,384
等級

圖 102B

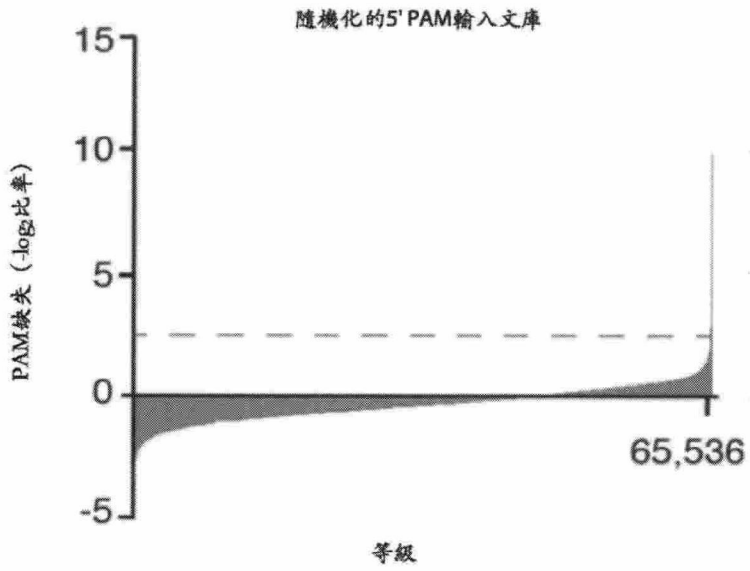

隨機化的5' PAM輸入文庫
15
10
5
0
-5
PAM缺失（-log2比率）
65,536
等級

圖 102C

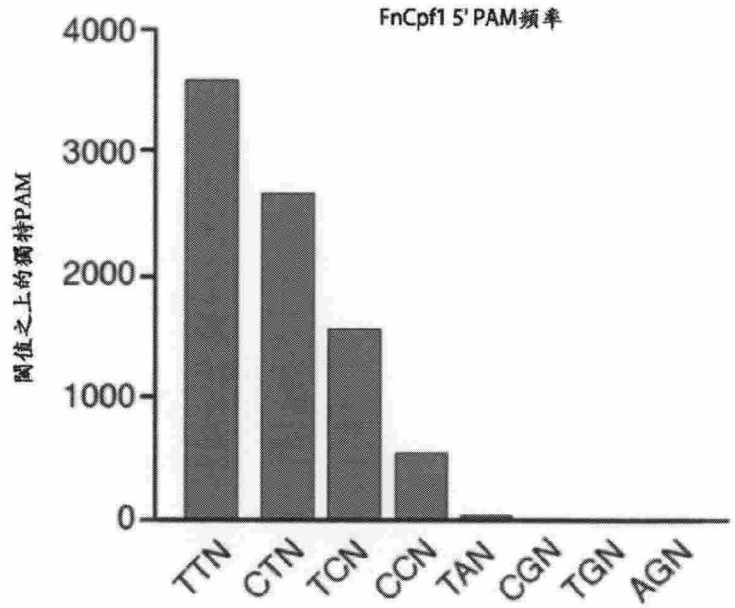

FnCpf1 5' PAM頻率
4000
3000
2000
1000
0
閾值之上的獨特PAM
TTN
CTN
TCN
CCN
TAN
CGN
TGN
AGN

圖 102D

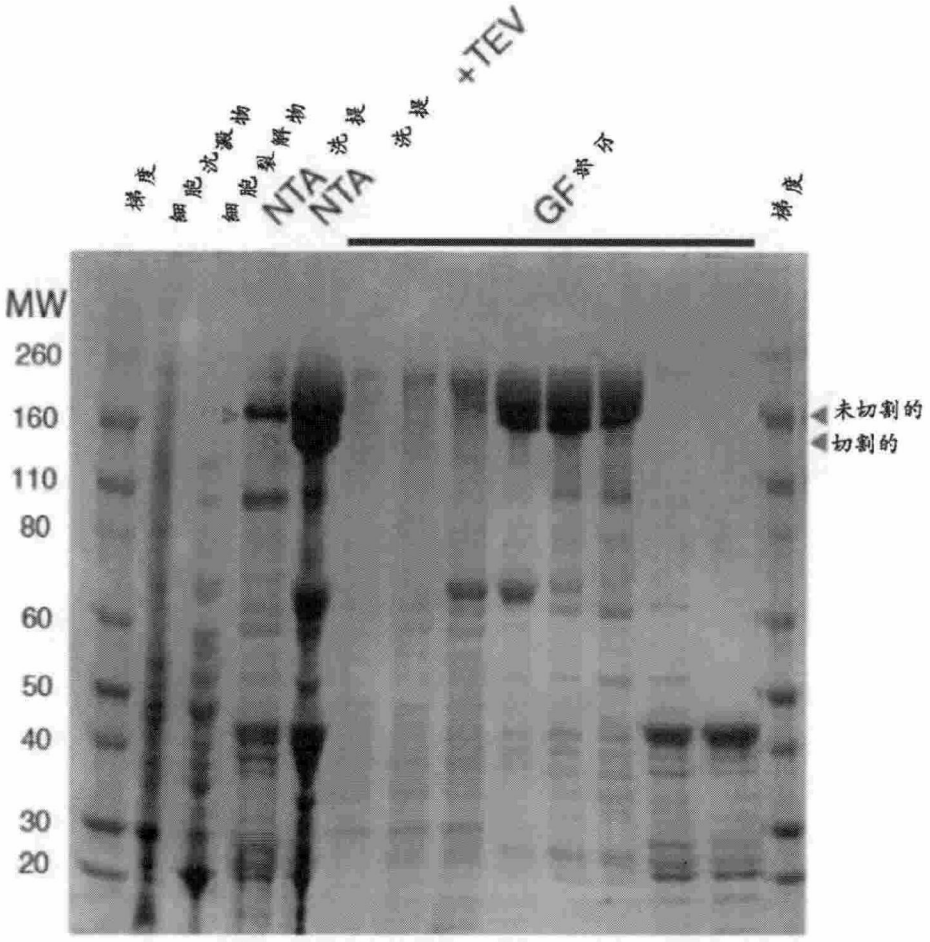

梯度
細胞沈澱物
細胞裂解物
NTA
NTA
洗提
洗提
+TEV
GF部分
梯度
MW
260
160
110
80
60
50
40
30
20
未切割的
切割的

圖 103A

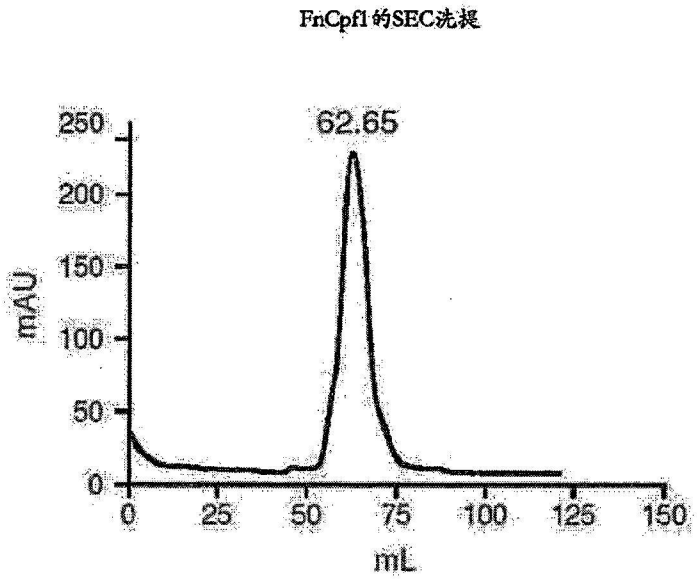

FnCpf1的SEC洗提
62.65
250
200
150
100
50
0
mAU
0
25
50
75
100
125
150
mL

圖 103B

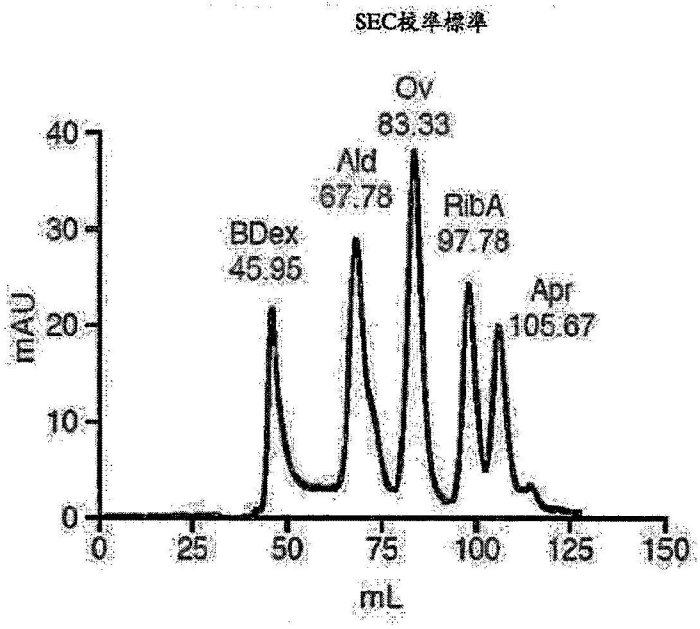

SEC校準標準
Ov
83.33
Ald
67.78
RibA
97.78
BDex
45.95
Apr
105.67
40
30
20
10
0
mAU
0
25
50
75
100
125
150
mL

圖 103C

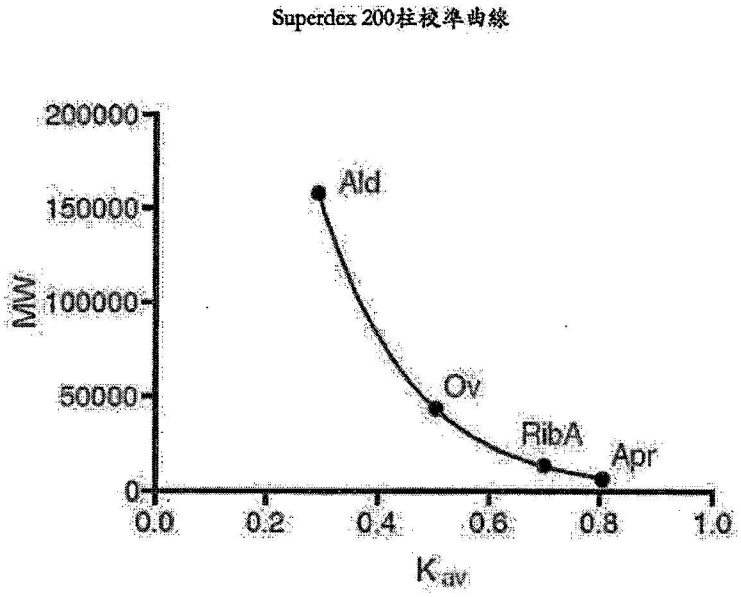

Superdex 200柱校準曲線
200000
150000
100000
50000
0
MW
Ald
Ov
RibA
Apr
0.0
0.2
0.4
0.6
0.8
1.0
K_{av}

圖 103D

FnCpf1
間隔區 1
PAM:
TTG
TTA
TTC
PAM
原型間隔區 (1)
5'– . . TTTGGAGAAGTCATTTAATAAGGCCACTGTTAAAA. . –3'
3'– . . AAACCTCTTCAGTAAATTATTCCGGTGACAATTTT. . –5'
(+)鏈
NGCCACTGTTAAAA. .
(–)鏈
. . TTTGGAGAAGTCATTTAATAAGGCCACN
PAM
原型間隔區 (1)
5'– . . TTTCGAGAAGTCATTTAATAAGGCCACTGTTAAAA. . –3'
3'– . . AAAGCTCTTCAGTAAATTATTCCGGTGACAATTTT. . –5'
(+)鏈
NGCCACTGTTAAAA. .
(–)鏈
. . TTTCGAGAAGTCATTTAATAAGGCCACN
PAM
原型間隔區 (1)
5'– . . TTTAGAGAAGTCATTTAATAAGGCCACTGTTAAAA. . –3'
3'– . . AAATCTCTTCAGTAAATTATTCCGGTGACAATTTT. . –5'
(+)鏈
NGCCACTGTTAAAA. .
(–)鏈
. . TTTAGAGAAGTCATTTAATAAGGCCACN

圖 104A

FnCpf1
間隔區 2
PAM:
TTG
TTT
PAM
原型間隔區 (2)
5'– . . TTTGGCTACTATTCCTGTGCCTTCAGATAATTCAA. . –3'
3'– . . AAACCGATGATAAGGACACGGAAGTCTATTAAGTT. . –5'
(+)鏈
NTTCAGATAATTCAA. . .
(–)鏈
. . TTTGGCTACTATTCCTGTGCCTTCAGAN
PAM
原型間隔區 (2)
5'– . . TTTTGCTACTATTCCTGTGCCTTCAGATAATTCAA. . –3'
3'– . . AAAACGATGATAAGGACACGGAAGTCTATTAAGTT. . –5'
(+)鏈
NTTCAGATAATTCAA. . .
(–)鏈
. . TTTTGCTACTATTCCTGTGCCTTCAGA

圖 104B

FnCpf1
間隔區 3
PAM:
TTC
TTG
PAM
原型間隔區 (3)
5'– . . TTTCGTCTAGAGCCTTTTGTATTAGTAGCCGAGCT . . – 3'
3'– . . AAAGCAGATCTCGGAAAACATAATCATCGGCTCGA . . – 5'
(+) 鏈
NTTAGTAGCCGAGCT . .
(–) 鏈
. . TTTCGTCTAGAGCCTTTTGTATTAGTAN
PAM
原型間隔區 (3)
5'– . . TTTGGTCTAGAGCCTTTTGTATTAGTAGCCGAGCT . . – 3'
3'– . . AAACCAGATCTCGGAAAACATAATCATCGGCTCGA . . – 5'
(+) 鏈
NTTAGTAGCCGAGCT . .
(–) 鏈
. . TTTGGTCTAGAGCCTTTTGTATTAGTAN

圖 104C

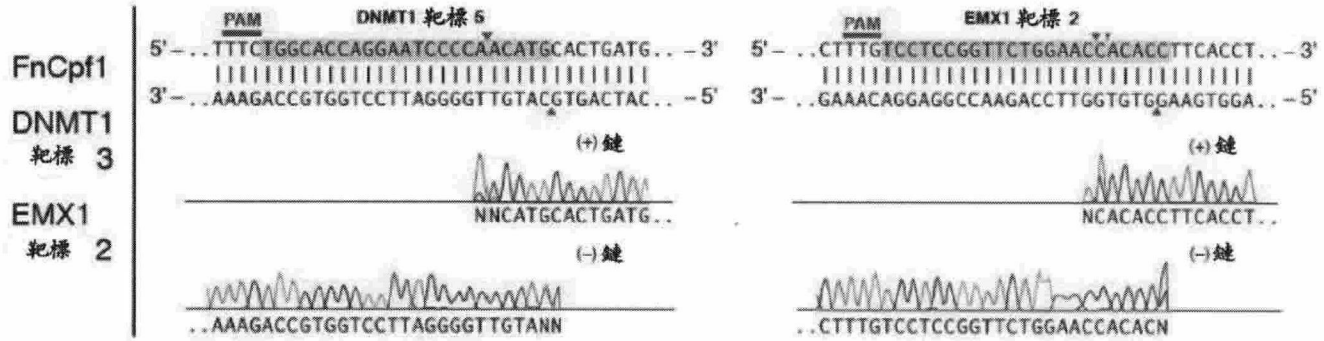

FnCpf1
DNMT1
靶標 3
EMX1
靶標 2
PAM
DNMT1 靶標 5
5'-..TTTCTGGCACCAGGAATCCCCAACATGCACTGATG..-3'
3'-..AAAGACCGTGGTCCTTAGGGGTTGTACGTGACTAC..-5'
(+)鏈
NNCATGCACTGATG..
(-)鏈
..AAAGACCGTGGTCCTTAGGGGTTGTANN
PAM
EMX1 靶標 2
5'-..CTTTGTCCTCCGGTTCTGGAACCACACCTTCACCT..-3'
3'-..GAAACAGGAGGCCAAGACCTTGGTGTGGAAGTGGA..-5'
(+)鏈
NCACACCTTCACCT..
(-)鏈
..CTTTGTCCTCCGGTTCTGGAACCACACN

圖 104D

cRNA
WT
G1T
G3T
A5C
U7G
A9C
U11G
A13C
U15G
A17C
G19T
C21A
C23A
無
未切割的
切割的
失配的指導
序列

圖 104E

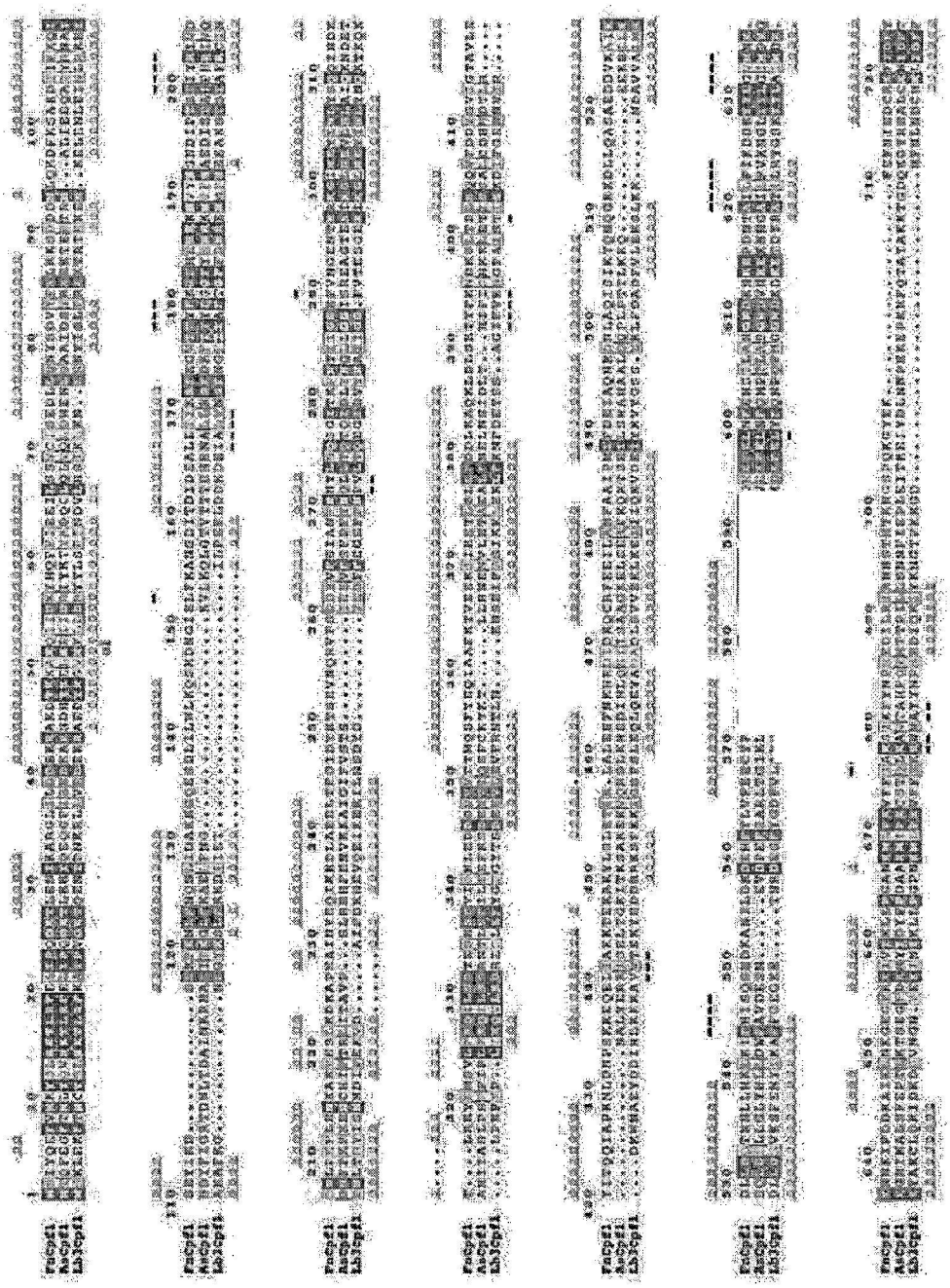


圖 105A

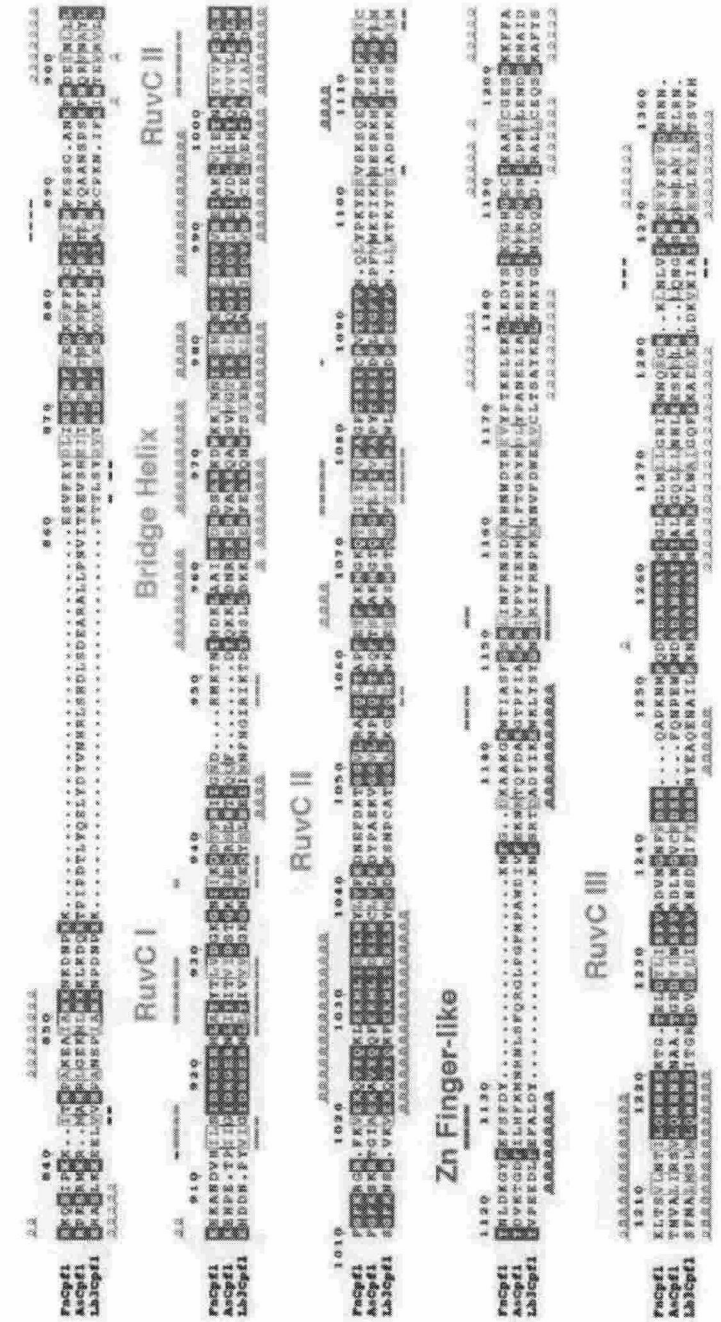
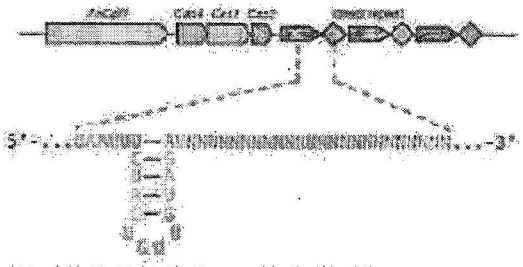
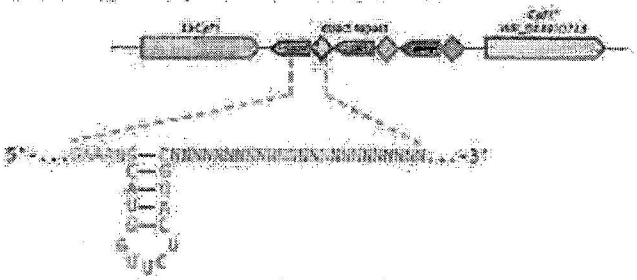


圖 105B

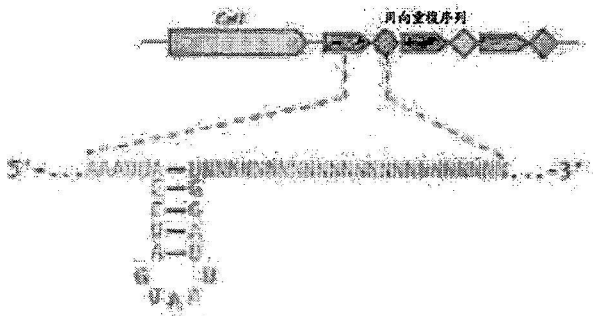
1-土拉熱弗朗西絲菌新殺手亞種U112(NC_008601)



2-毛螺旋菌科細菌MC2017(NZ_KL370607)



3-解朊丁酸弧菌(NC_014383)



4-佩萊格裡尼菌科細菌GW2011_GWA_33_10(LBOO01000015)

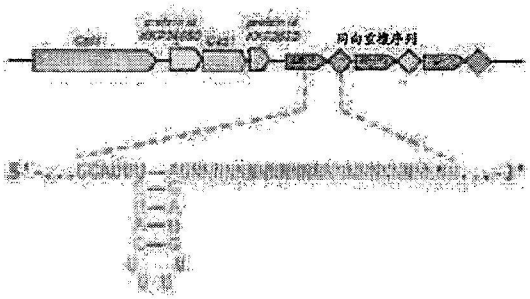
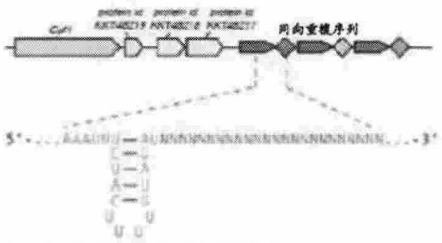
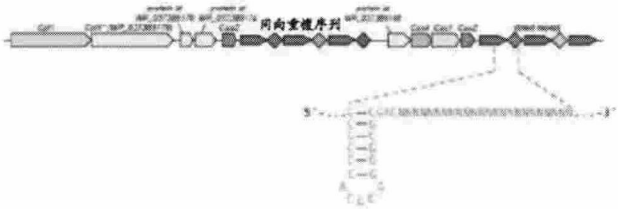


圖 106A

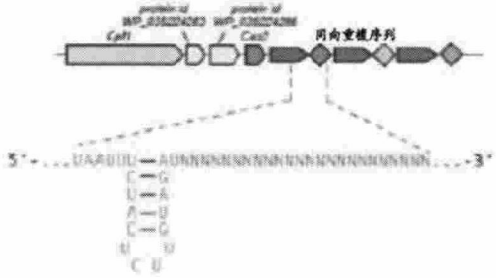
5-P.細菌GW2011_GWC2_44_17 *(LCIC01000001)*



6- *G. sp. GC_K08D17 (NZ_JQDO01000121)*



7- *A. sp. BV3L6 (NZ_AWUR01000016)*



8-毛螺旋菌MA2020 *(NZ_JQKK01000008)*

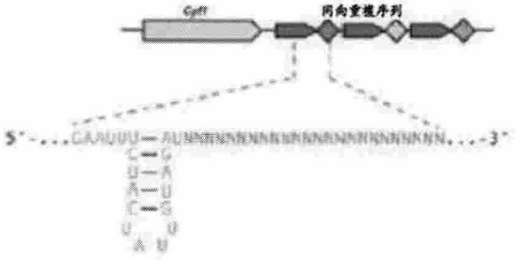
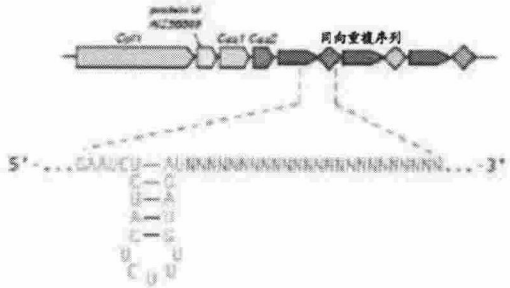
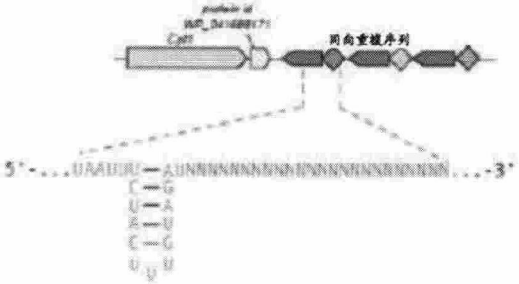


圖 106B

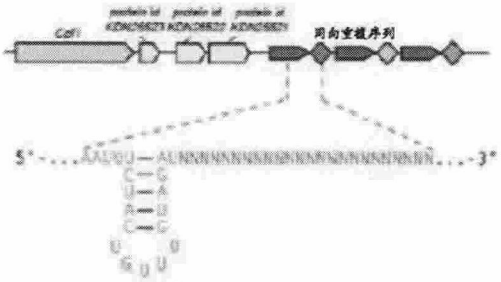
9-候選白蟻甲烷枝原體 *(CP010070)*



10-挑剔真細菌 *(NC_012778)*



11-牛莫拉氏菌237*(AOMT01000011)*



12-稻田鉤端螺旋體*(NZ_AHMM020000017)*

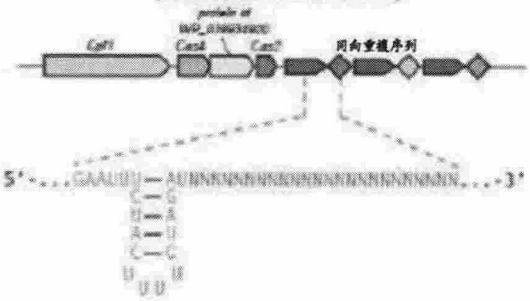
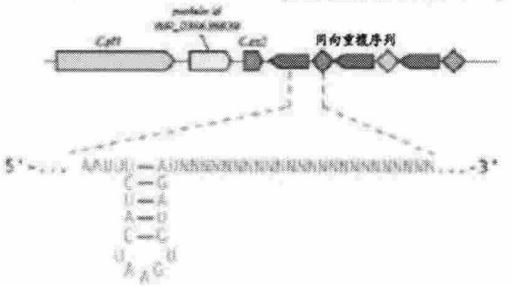
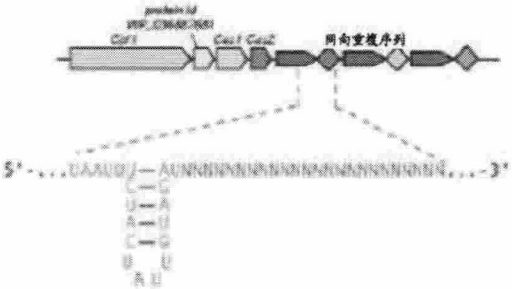


圖 106C

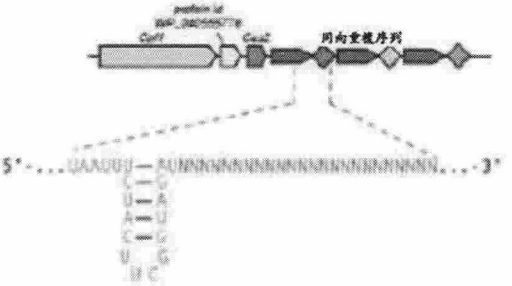
13-毛螺旋菌科細菌ND2006*(NZ_JNKS01000011)*



14-狗口腔卟啉單胞菌 *(NZ_JQJC01000021)*



15-解糖腺普雷沃菌 *(NZ_AEDO01000031)*



16-獼猴卟啉單胞菌 *(NZ_BAKQ01000001)*

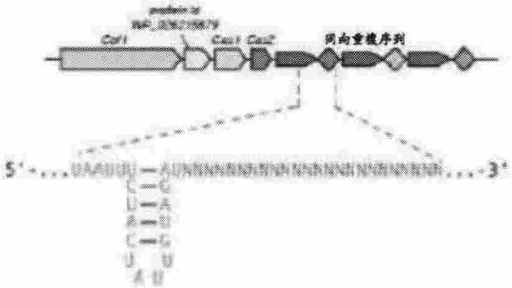


圖 106D

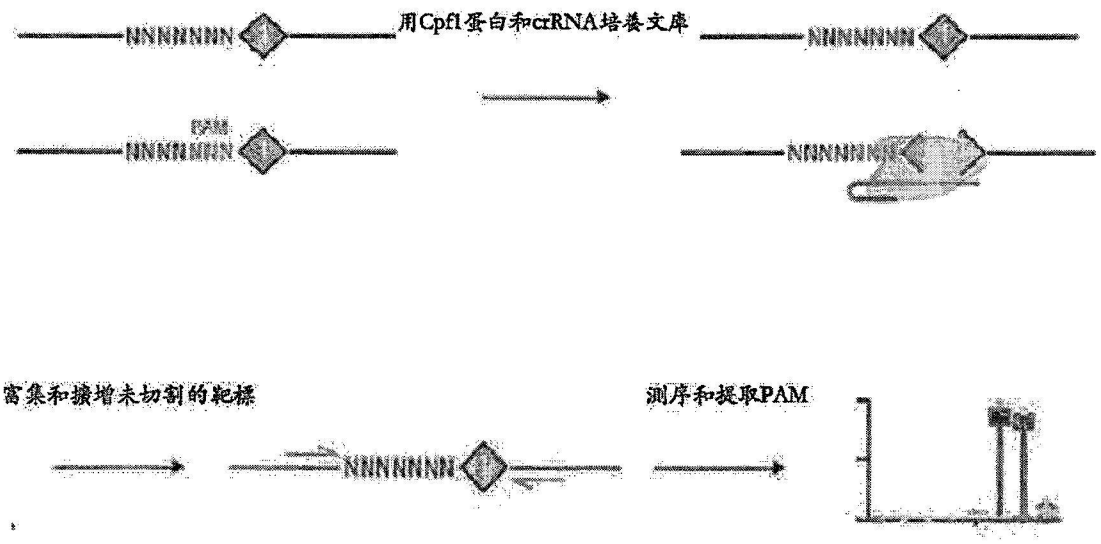


圖 107A

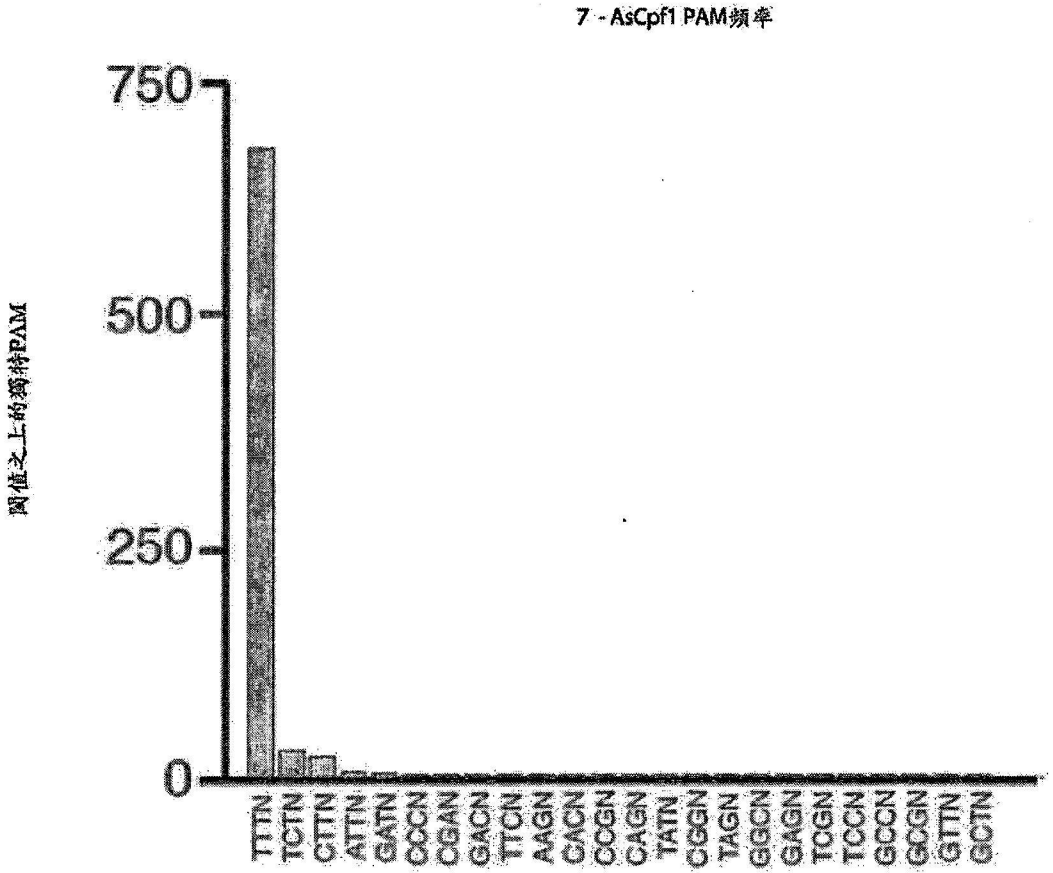

7 - AsCpf1 PAM頻率
閾值之上的獨特PAM
750
500
250
0
TTTN
TCTN
CTTN
ATTN
GATN
CCCN
CGAN
GACN
TTCN
AAGN
CACN
CCGN
CAGN
TATN
CGGN
TAGN
GGCN
GAGN
TCGN
TCCN
GCCN
GCGN
GTTN
GCTN

圖 107B

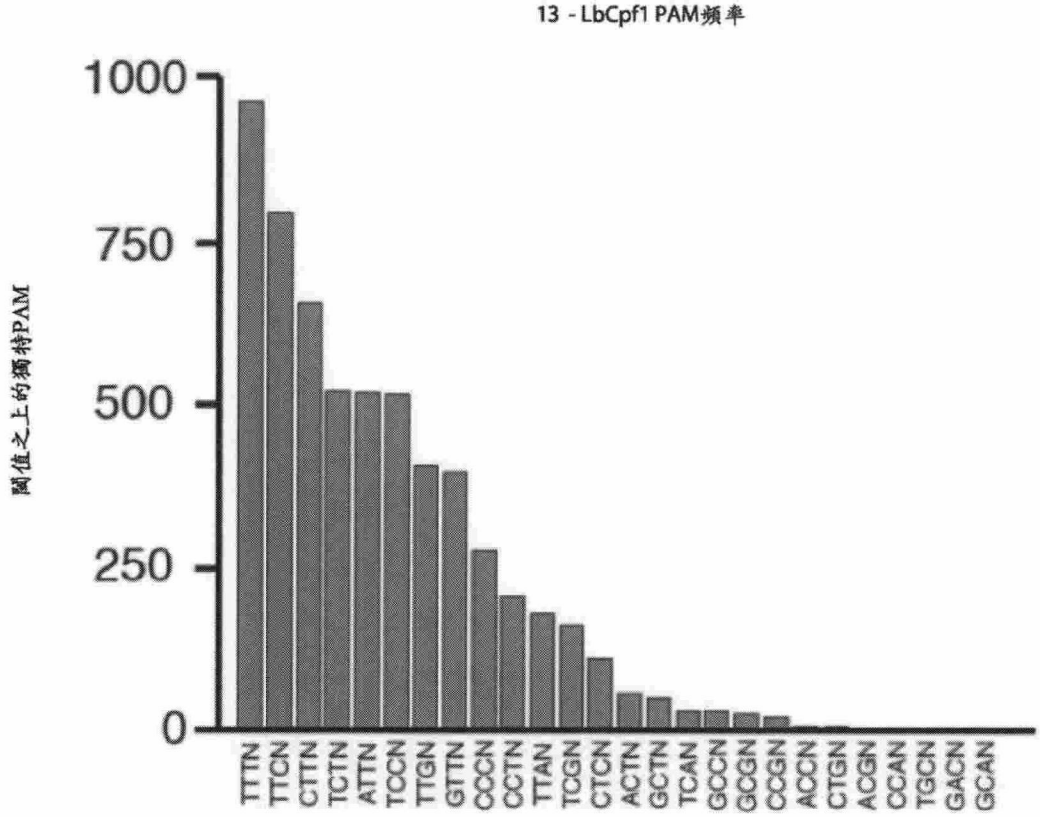

13 - LbCpf1 PAM頻率
閾值之上的獨特PAM
1000
750
500
250
0
TTTN
TTCN
CTTN
TCTN
ATTN
TCCN
TTGN
GTTN
CCCN
CCTN
TTAN
TCGN
CTCN
ACTN
GCTN
TCAN
GCCN
GCGN
CCGN
ACCN
CTGN
ACGN
CCAN
TGCN
GACN
GCAN

圖 107C

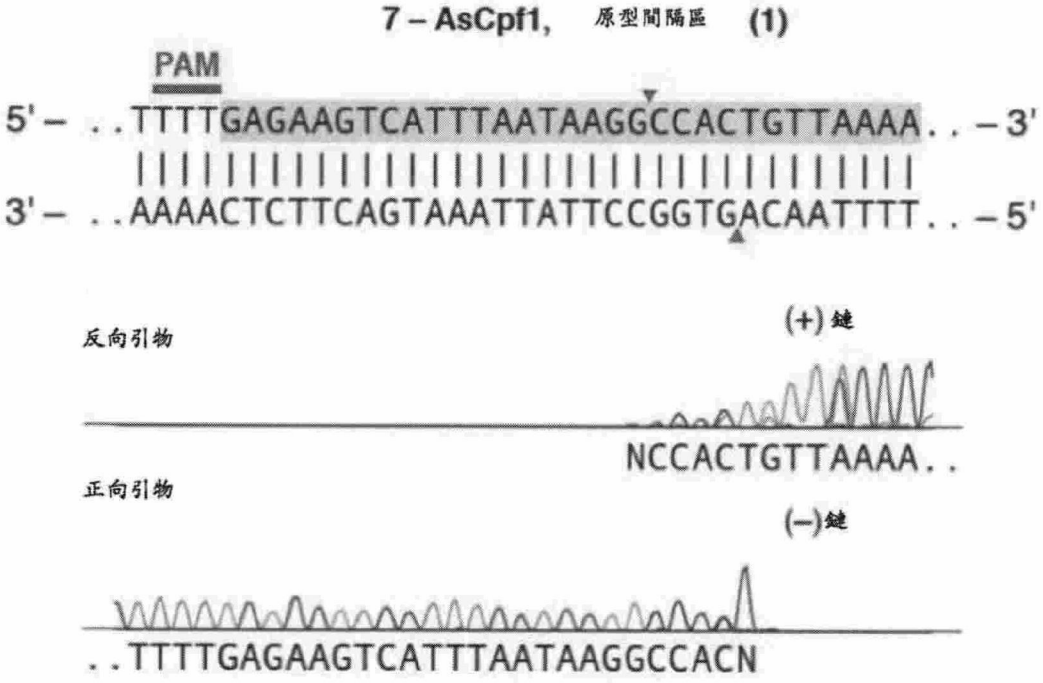

7 – AsCpf1, 原型間隔區 (1)
PAM
5' – . . TTTTGAGAAGTCATTTAATAAGGCCACTGTTAAAA . . – 3'
3' – . . AAAACTCTTCAGTAAATTATTCCGGTGACAATTTT . . – 5'
反向引物
(+) 鏈
NCCACTGTTAAAA . .
正向引物
(–) 鏈
. . TTTTGAGAAGTCATTTAATAAGGCCACN

圖 107D

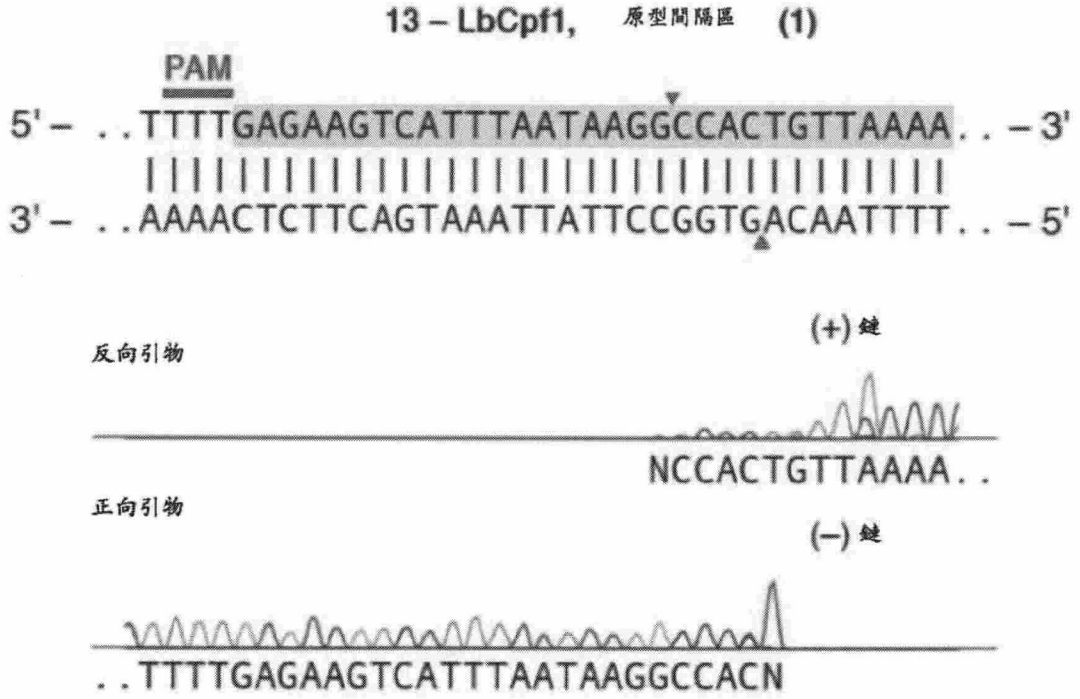

13 – LbCpf1, 原型間隔區 (1)
PAM
5' – . . TTTTGAGAAGTCATTTAATAAGGCCACTGTTAAAA . . – 3'
3' – . . AAAACTCTTCAGTAAATTATTCCGGTGACAATTTT . . – 5'
反向引物
(+) 鏈
NCCACTGTTAAAA . .
正向引物
(–) 鏈
. . TTTTGAGAAGTCATTTAATAAGGCCACN

圖 107E

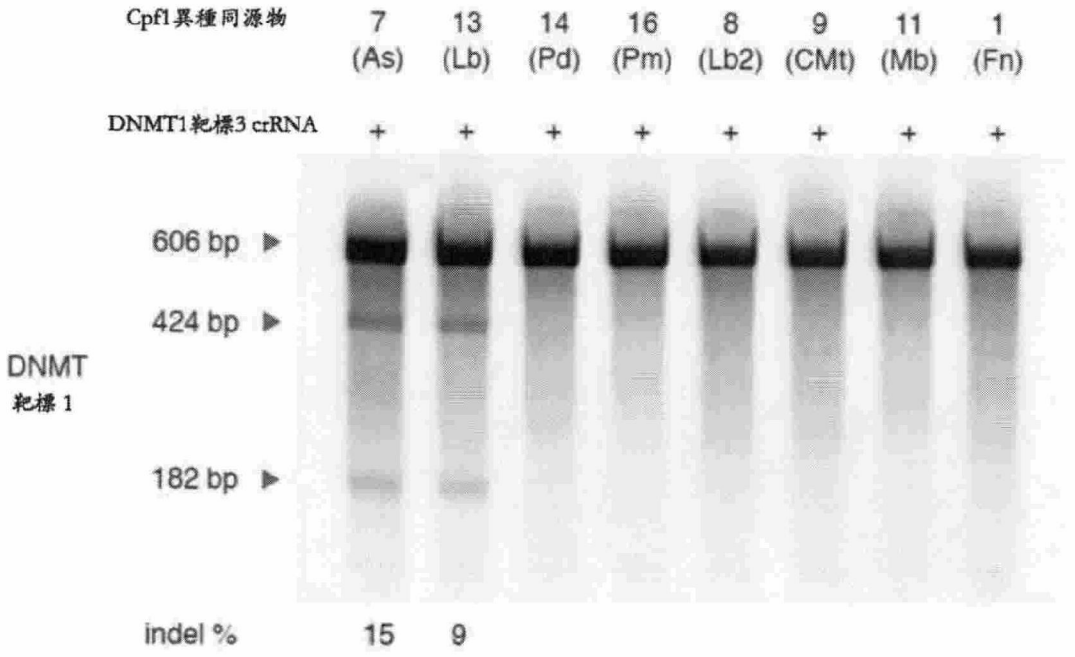

Cpf1異種同源物
7 (As)
13 (Lb)
14 (Pd)
16 (Pm)
8 (Lb2)
9 (CMt)
11 (Mb)
1 (Fn)
DNMT1靶標3 crRNA
+ + + + + + + +
606 bp
424 bp
DNMT 靶標 1
182 bp
indel %
15
9

圖 108A

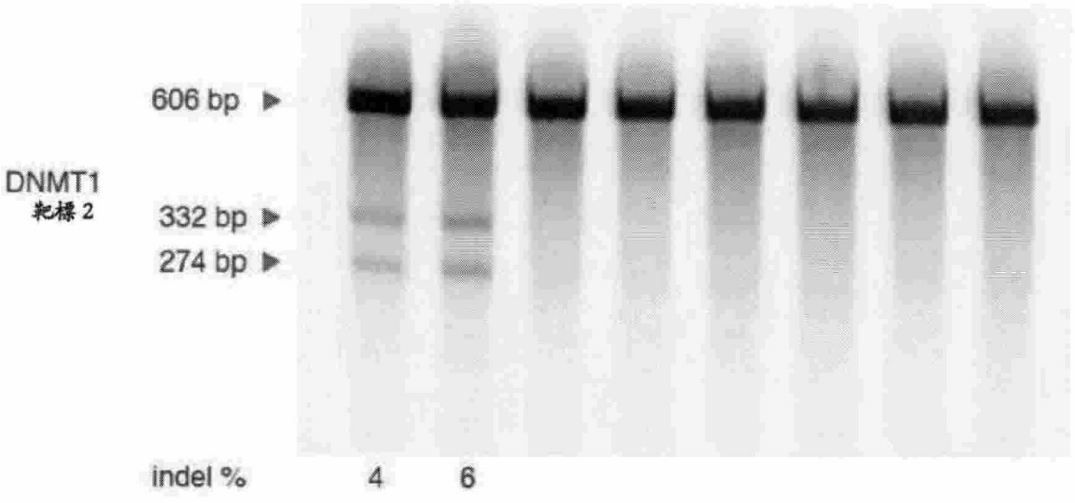

606 bp
DNMT1 靶標 2
332 bp
274 bp
indel %
4
6

圖 108B

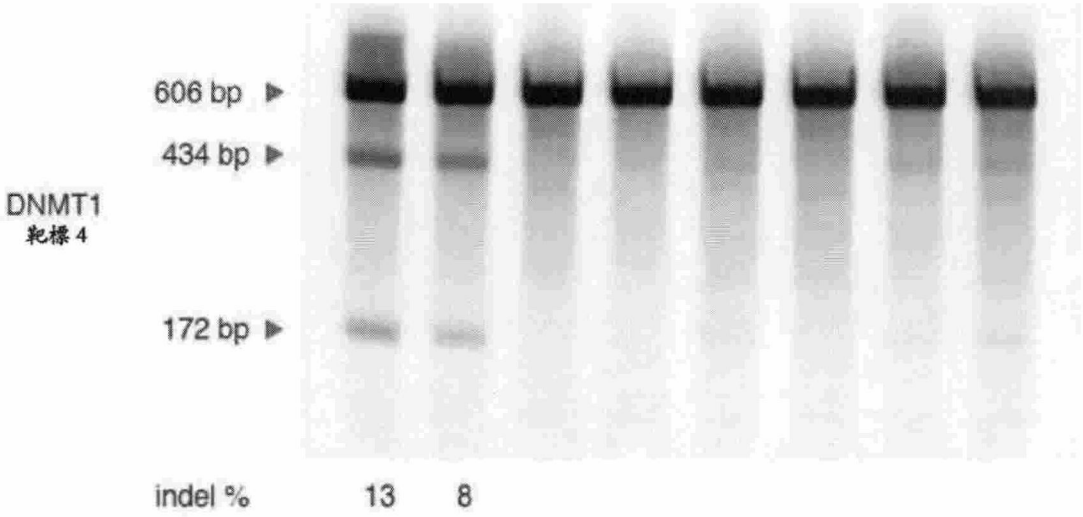


圖 108C

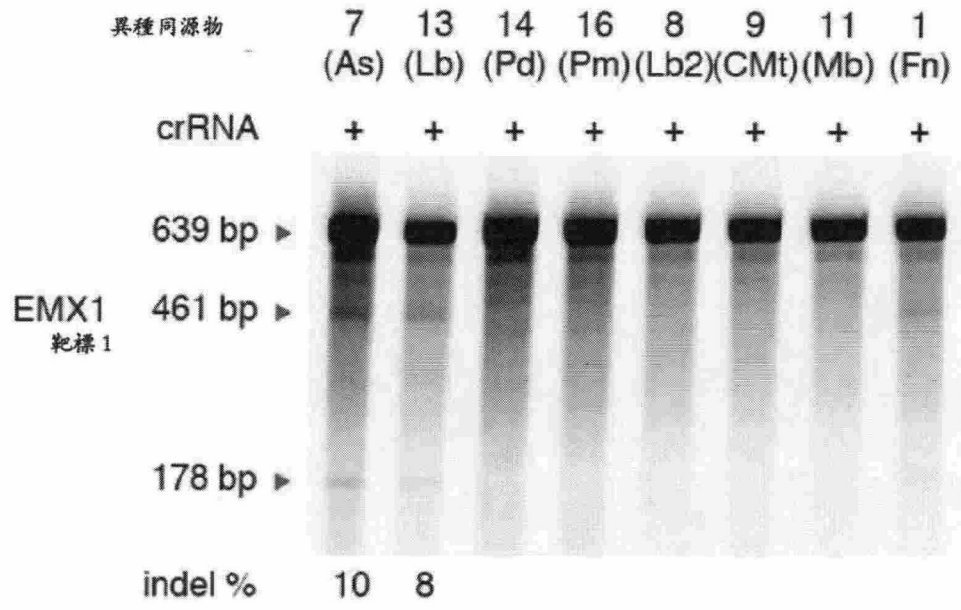

異種同源物
7 (As)
13 (Lb)
14 (Pd)
16 (Pm)
8 (Lb2)
9 (CMt)
11 (Mb)
1 (Fn)
crRNA
+ + + + + + + +
639 bp
EMX1
461 bp
靶標 1
178 bp
indel %
10 8

圖 108D

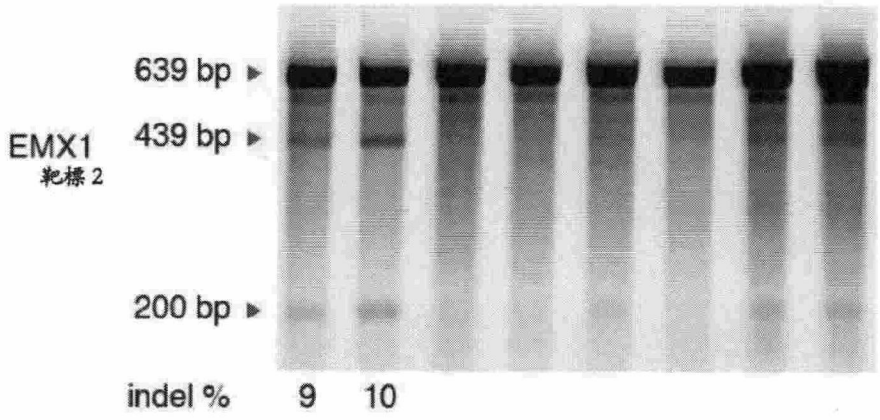

639 bp
EMX1
439 bp
靶標 2
200 bp
indel %
9 10

圖 108E

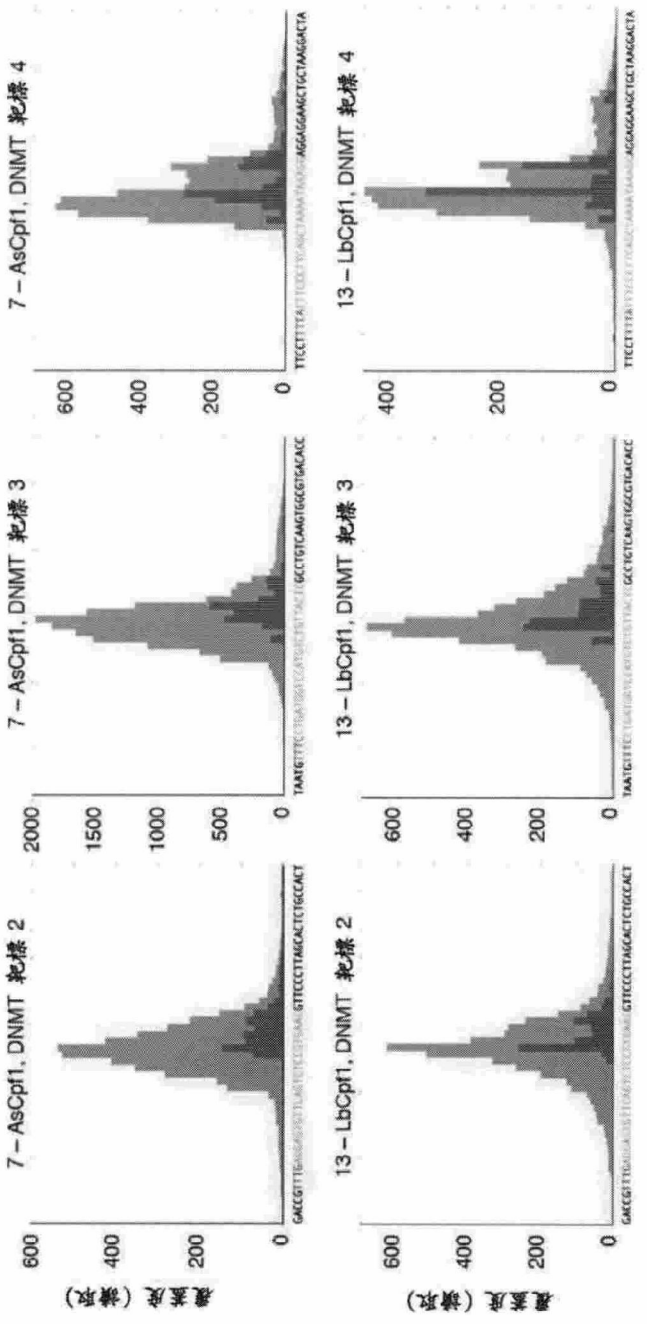

7 – AsCpf1, DNMT 靶標 2
7 – AsCpf1, DNMT 靶標 3
7 – AsCpf1, DNMT 靶標 4
13 – LbCpf1, DNMT 靶標 2
13 – LbCpf1, DNMT 靶標 3
13 – LbCpf1, DNMT 靶標 4
覆蓋度 (讀取)

圖 108F

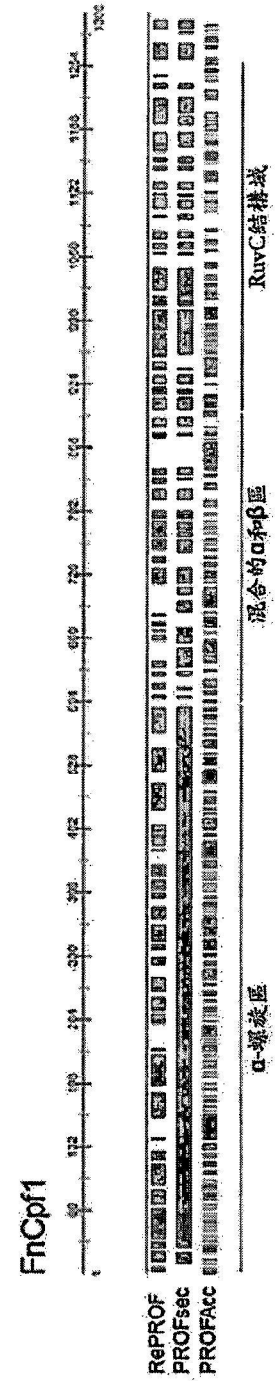


圖 109A

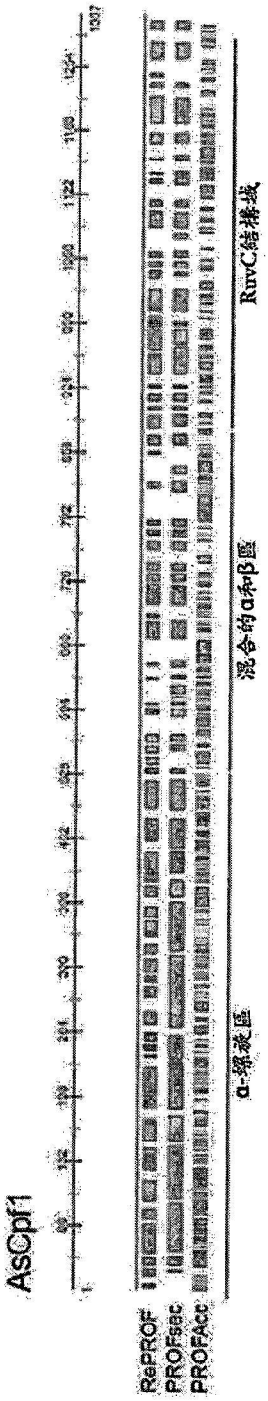


圖 109B

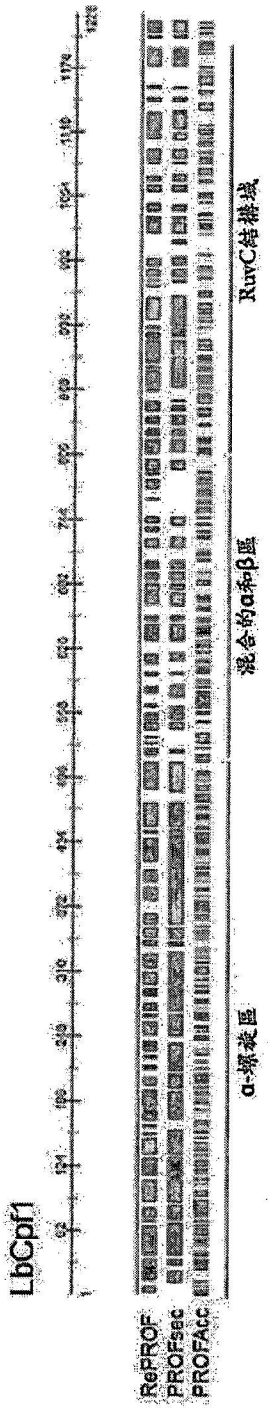


圖 109C

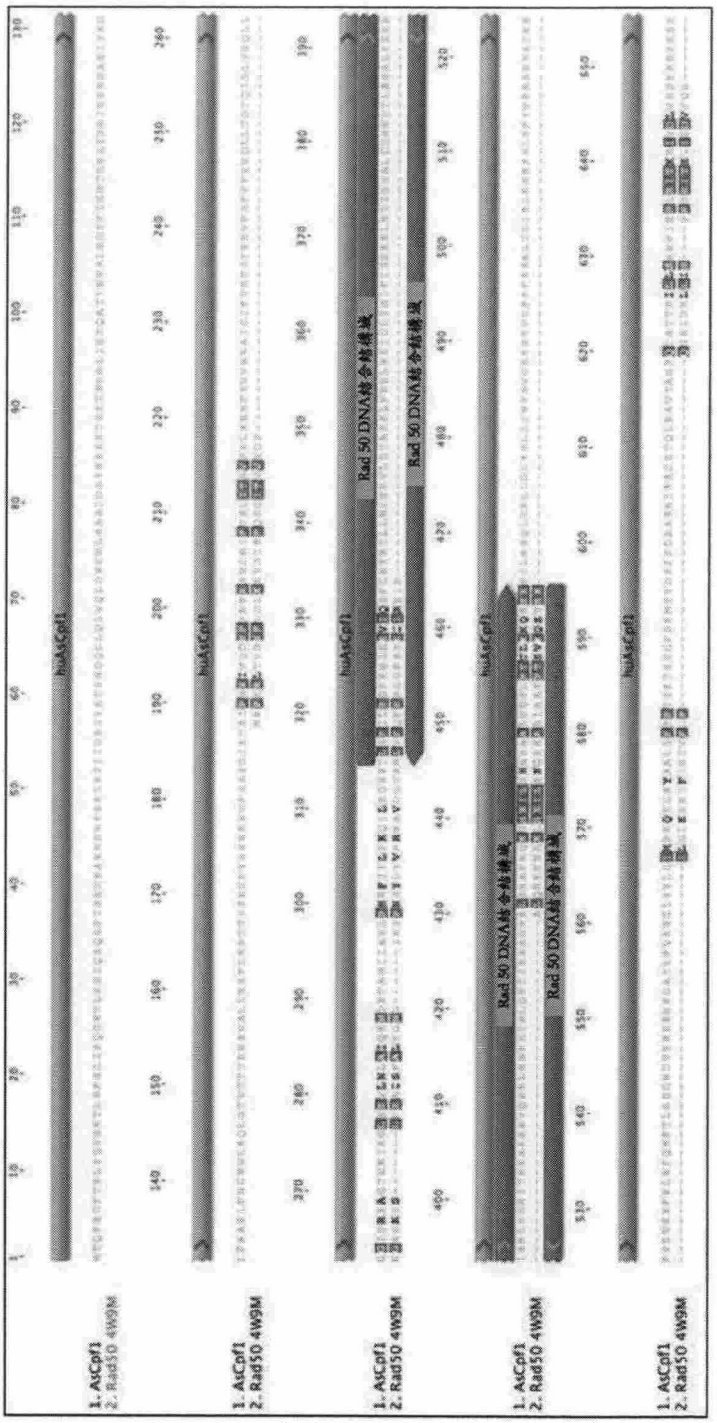


圖 110A

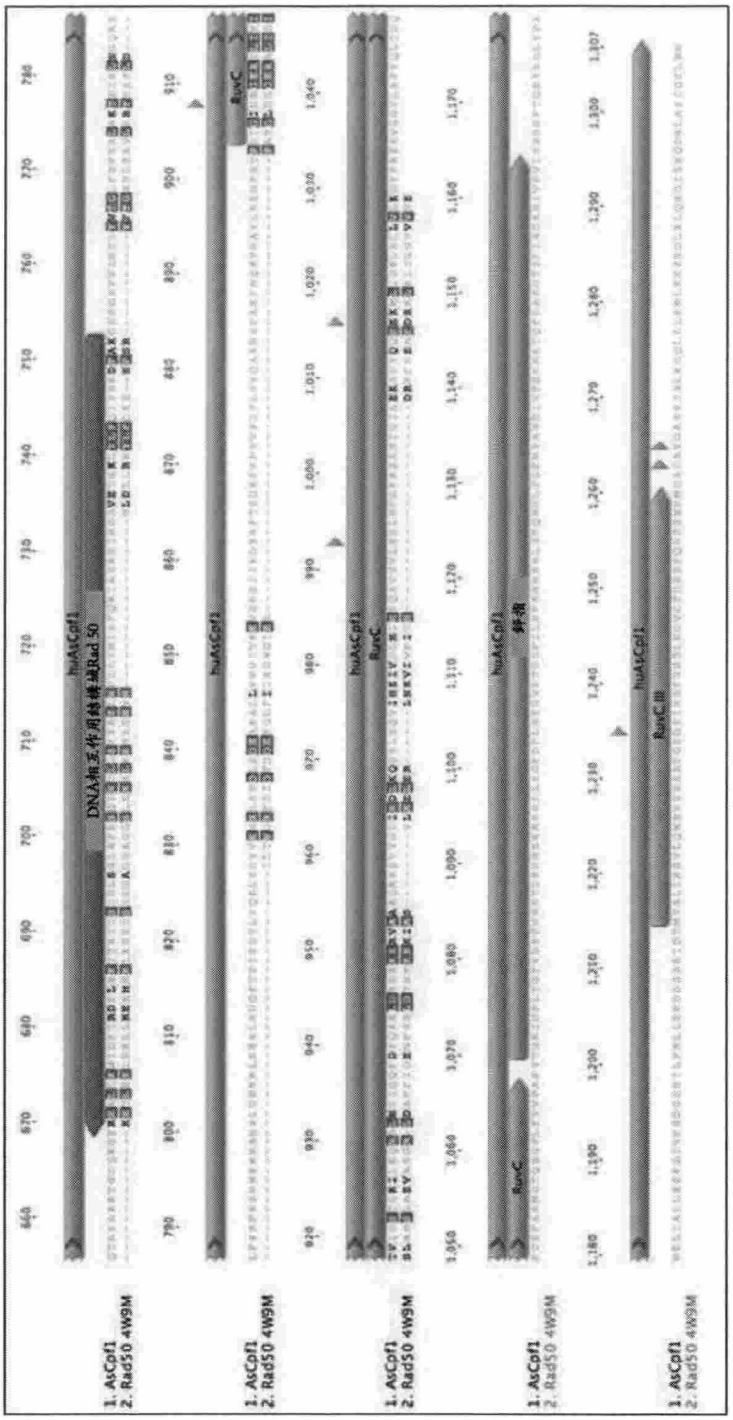

1. AsCpf1
2. Rad50 4W9M
DNA结合作用结构域Rad 50
huAsCpf1
RuvC
RuvC III

圖 110B



圖 110C

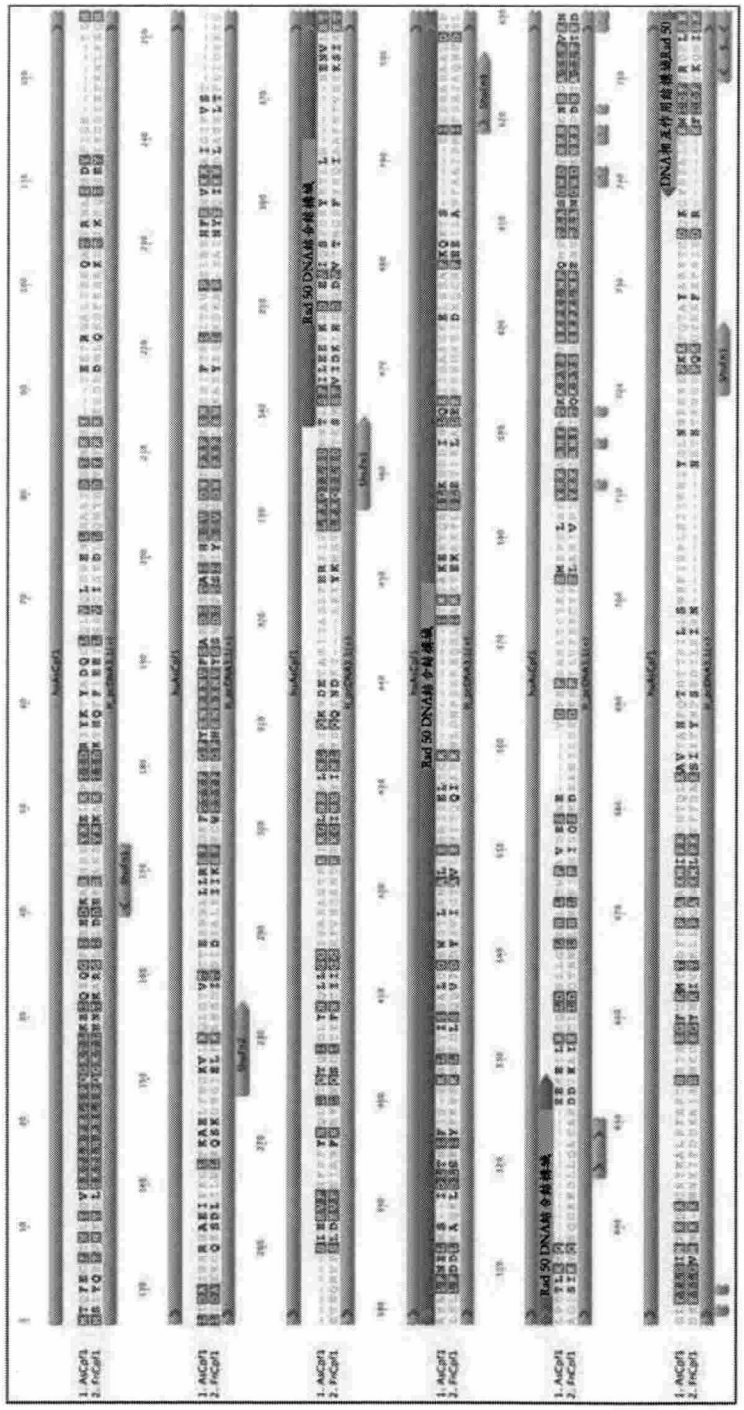


圖 110D

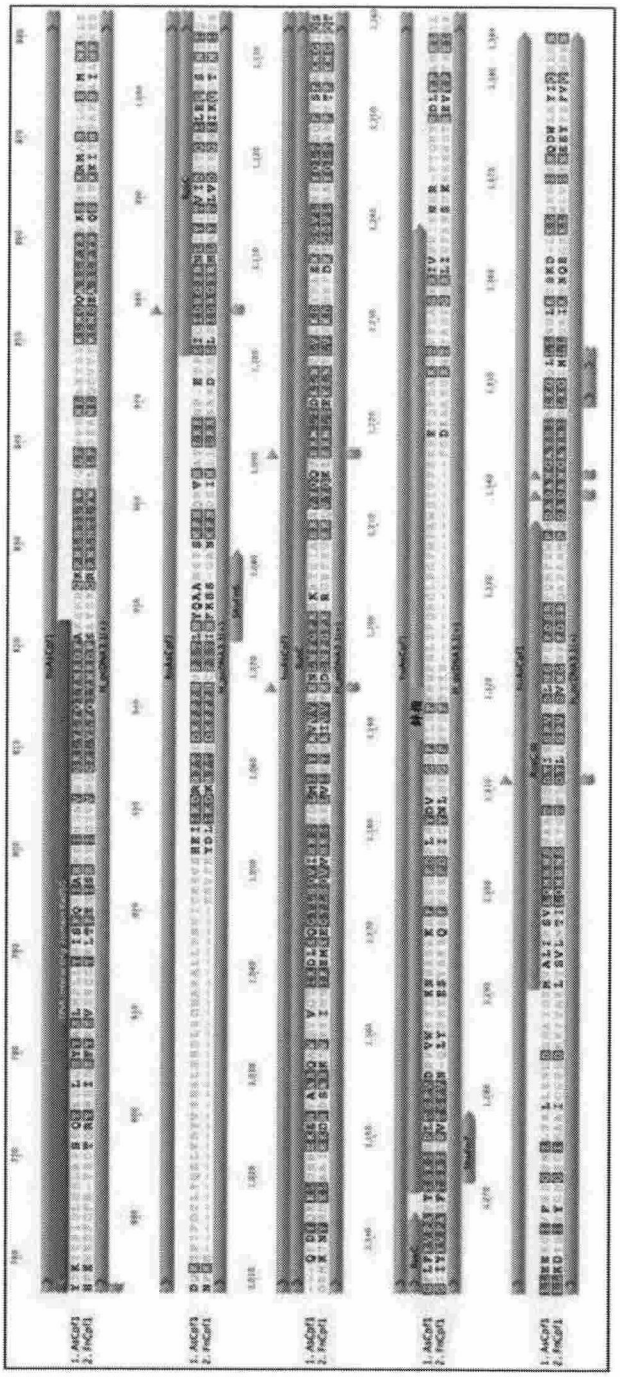


圖 110E

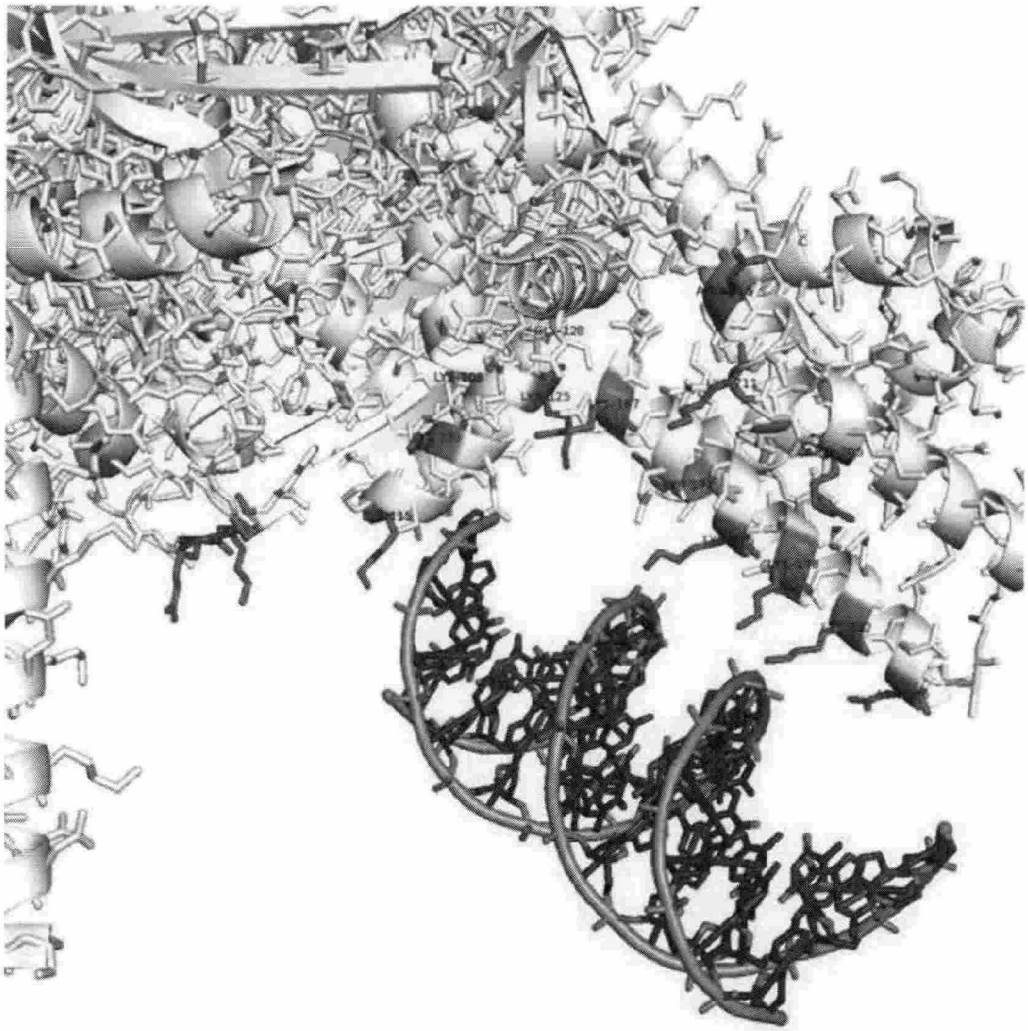


圖 111

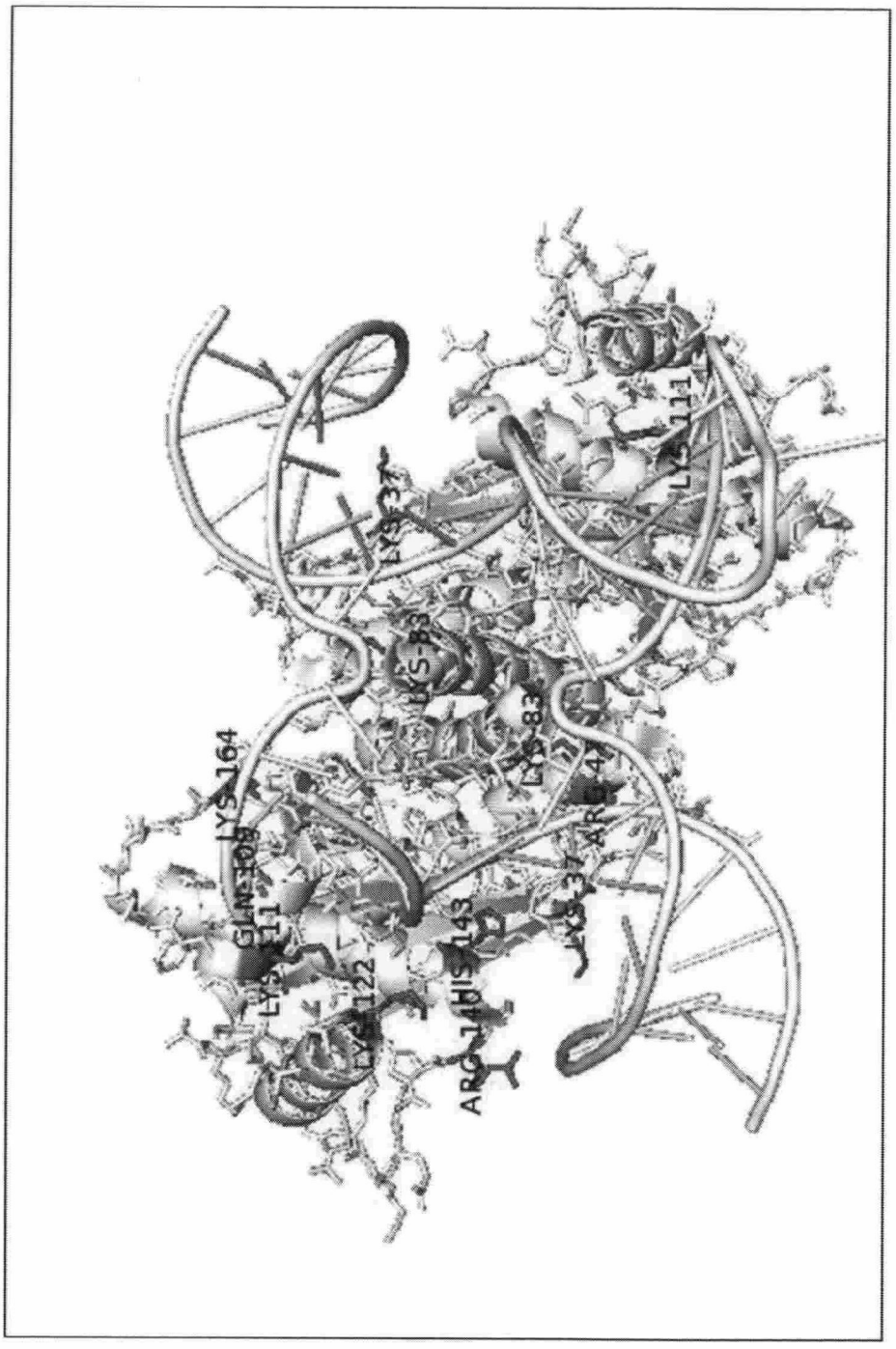

LYS-37
LYS-111
LYS-83
LYS-83
LYS-164
GLN-109
LYS-111
LYS-122
ARG-140
HIS-143
LYS-37
ARG-47

圖 112

位點特異性重組酶XerD[腸道葛氏菌M50/1]

序列ID: emb|CBL10059.1| 長度: 416 匹配數目: 1

範围1:175至292 GenPept 圖形 下一匹配 上一匹配

得分	期望值	方法	一致性	相似性	間隙
46.2位 (108)	0.14	組成性矩陣調整 (Compositional matrix adjust)。	42/127(33%)	59/127(46%)	17/127(13%)

```
Query  750  FAKGHHGKP----NLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEK  805
            F +G   KP        LYWTG+   E LA +   ++   +L       + R   R+G K
Sbjct  175  FREGVKDKPLSYICFEVLYWTGMREGELLALSPADIDIDNKLI-----SINRTYQRIGGK  229

Query  806  MLNKKLKDQKT----PIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKD  861
             +    K +K+    PIPD L QEL DY+  R   D +DE    L P V    +SHE+I+
Sbjct  230  DVFTSPKTRKSKRTIPIPDFLCQELSDYIQSRYMLD-ADE--RLFP-VTKSYLSHEMIRG  285

Query  862  RRFTSDK  868
             + T  K
Sbjct  286  CKITGAK  292
```

圖 113

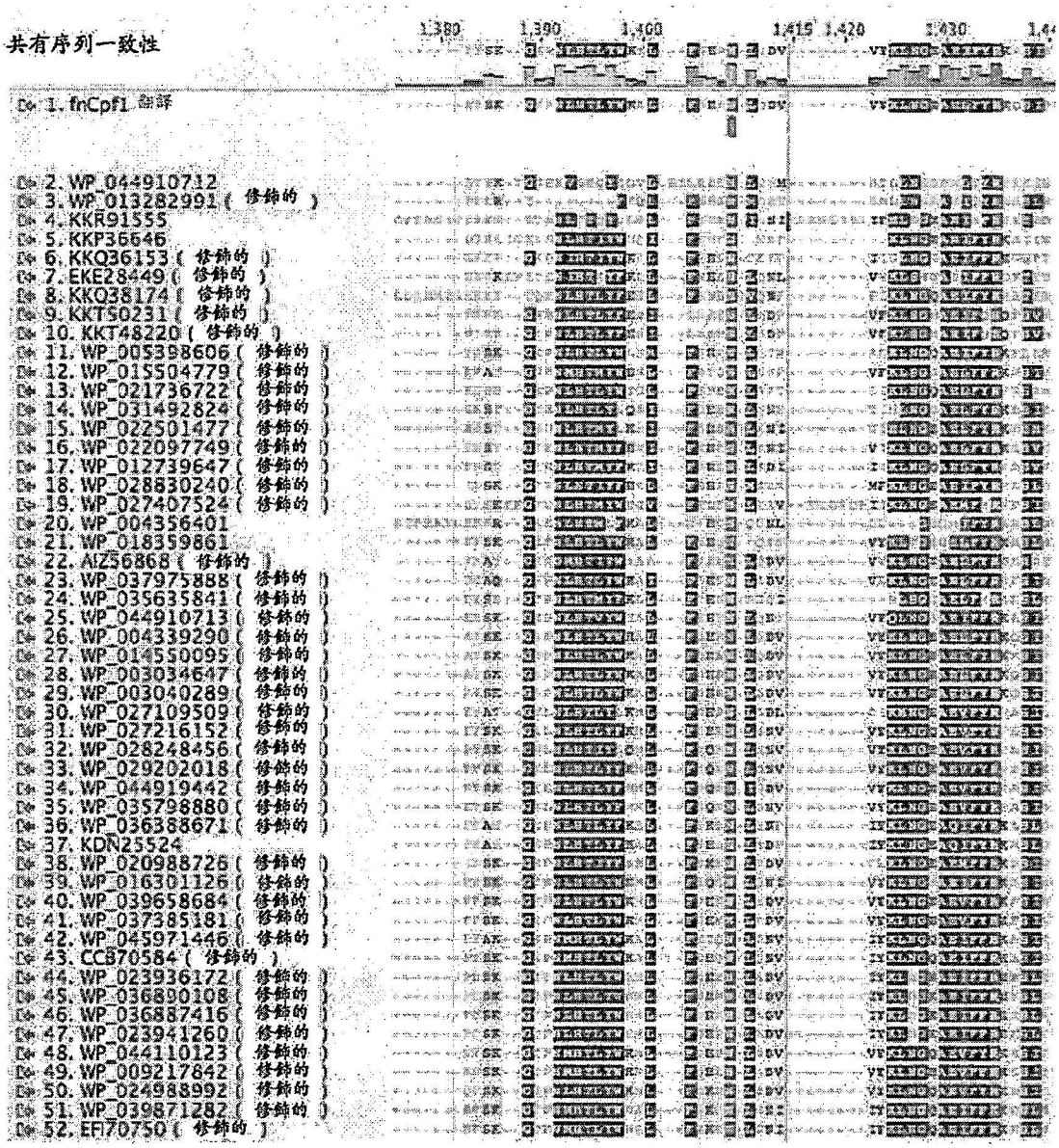

共有序列一致性
1,380
1,390
1,400
1,410
1,420
1,430
1. fnCpf1 翻譯
2. WP_044910712
3. WP_013282991 (修飾的)
4. KKR91555
5. KKP36646
6. KKQ36153 (修飾的)
7. EKE28449 (修飾的)
8. KKQ38174 (修飾的)
9. KKT50231 (修飾的)
10. KKT48220 (修飾的)
11. WP_005398606 (修飾的)
12. WP_015504779 (修飾的)
13. WP_021736722 (修飾的)
14. WP_031492824 (修飾的)
15. WP_022501477 (修飾的)
16. WP_022097749 (修飾的)
17. WP_012739647 (修飾的)
18. WP_028830240 (修飾的)
19. WP_027407524 (修飾的)
20. WP_004356401
21. WP_018359861
22. AIZ56868 (修飾的)
23. WP_037975888 (修飾的)
24. WP_035635841 (修飾的)
25. WP_044910713 (修飾的)
26. WP_004339290 (修飾的)
27. WP_014550095 (修飾的)
28. WP_003034647 (修飾的)
29. WP_003040289 (修飾的)
30. WP_027109509 (修飾的)
31. WP_027216152 (修飾的)
32. WP_028248456 (修飾的)
33. WP_029202018 (修飾的)
34. WP_044919442 (修飾的)
35. WP_035798880 (修飾的)
36. WP_036388671 (修飾的)
37. KDN25524
38. WP_020988726 (修飾的)
39. WP_016301126 (修飾的)
40. WP_039658684 (修飾的)
41. WP_037385181 (修飾的)
42. WP_045971446 (修飾的)
43. CCB70584 (修飾的)
44. WP_023936172 (修飾的)
45. WP_036890108 (修飾的)
46. WP_036887416 (修飾的)
47. WP_023941260 (修飾的)
48. WP_044110123 (修飾的)
49. WP_009217842 (修飾的)
50. WP_024988992 (修飾的)
51. WP_039871282 (修飾的)
52. EFI70750 (修飾的)

圖 114

攜帶針對AsCpf1的串聯指導序列的質粒

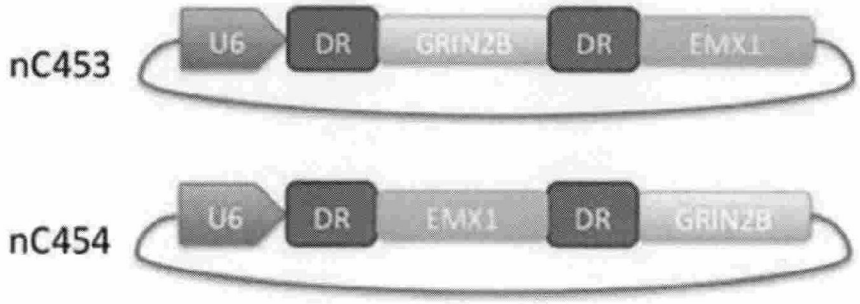


圖 115A

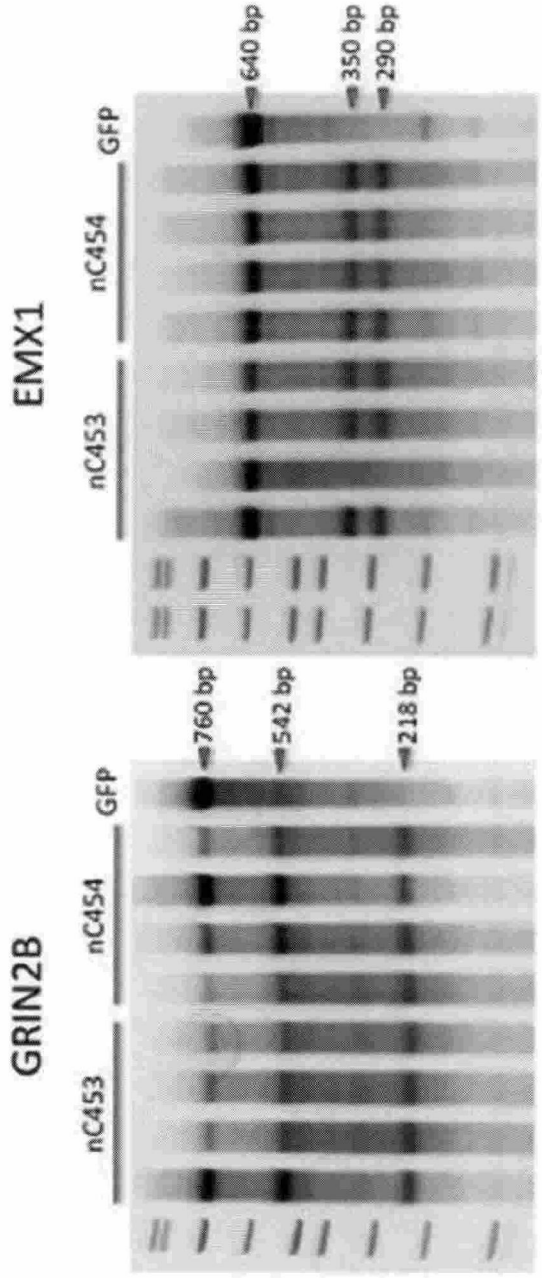

EMX1
nC453
nC454
GFP
640 bp
350 bp
290 bp
GRIN2B
nC453
nC454
GFP
760 bp
542 bp
218 bp

圖 115B

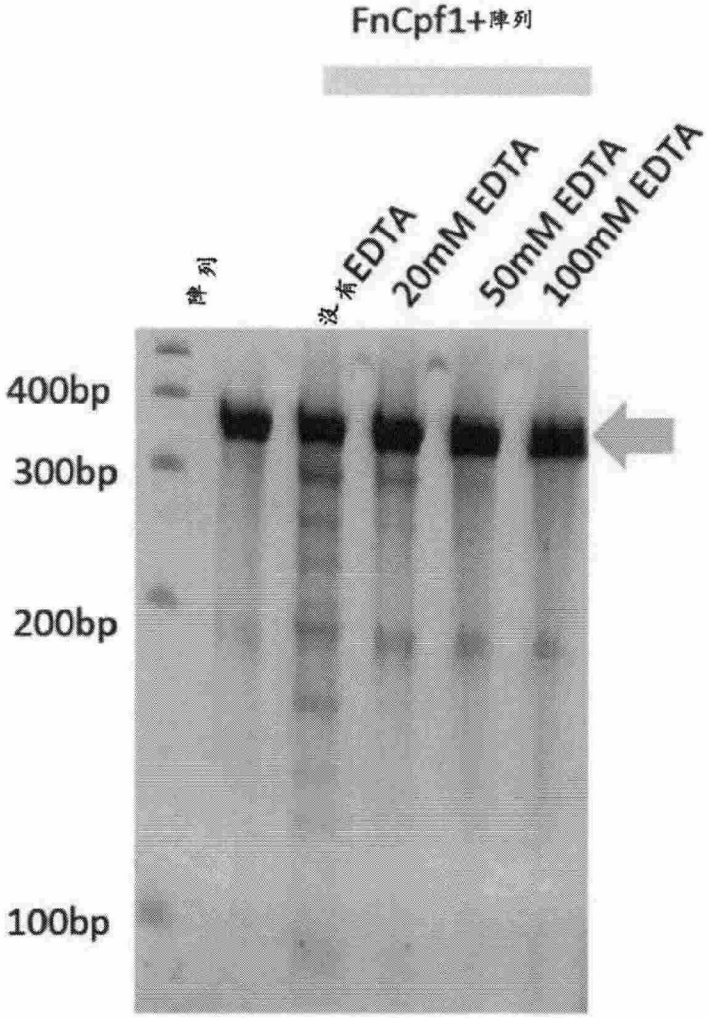

FnCpf1+陣列
陣列
沒有EDTA
20mM EDTA
50mM EDTA
100mM EDTA
400bp
300bp
200bp
100bp

圖 116