



(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 I426126 B

(45)公告日：中華民國 103 (2014) 年 02 月 11 日

(21)申請案號：099125328

(22)申請日：中華民國 91 (2002) 年 04 月 15 日

(51)Int. Cl. : **C12N1/20 (2006.01)** **C12N5/10 (2006.01)**
 C12N15/00 (2006.01) **C12N15/63 (2006.01)**
 C12N15/70 (2006.01) **C12P19/02 (2006.01)**
 C07H21/00 (2006.01)

(30)優先權：2001/04/16 美國 60/284,066
 2001/06/15 美國 60/298,796
 2001/09/18 美國 60/323,269

(71)申請人：D S M智慧財產有限公司 (荷蘭) DSM IP ASSETS B. V. (NL)
 荷蘭

(72)發明人：庫尼爾 傑瑞M KUNER, JERRY M. (US)；巴克雷 威廉R BARCLAY, WILLIAM R. (US)；曼茲 詹姆士G METZ, JAMES G. (US)；法拉特 詹姆士H FLATT, JAMES H. (US)

(74)代理人：惲軼群；陳文郎

(56)參考文獻：

EP 0823475A WO 9855625A

審查人員：林俊茂

申請專利範圍項數：36 項 圖式數：3 共 0 頁

(54)名稱

多不飽和脂肪酸 (P U F A) 聚乙醯合成酶系統及其用途 (二)

PUFA POLYKETIDE SYNTHASE SYSTEMS AND USES THEREOF

(57)摘要

本發明係一般性地關於一種分離自或起源自非細菌生物的多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統；關於其同源物；關於編碼此 PUFA PKS 系統的生物活性區段之分離的核酸分子及重組型核酸分子；關於包含 PUFA PKS 系統之基因改性的生物；關於製得及使用此系統來製造有興趣的生物活性分子之方法；及關於用來鑑別具有此 PUFA PKS 系統的新穎細菌及非細菌微生物之新穎的方法。

The invention generally relates to polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) systems isolated from or derived from non-bacterial organisms, to homologues thereof, to isolated nucleic acid molecules and recombinant nucleic acid molecules encoding biologically active domains of such a PUFA PKS system, to genetically modified organisms comprising PUFA PKS systems, to methods of making and using such systems for the production of bioactive molecules of interest, and to novel methods for identifying new bacterial and non-bacterial microorganisms having such a PUFA PKS system.

六、發明說明：

【發明所屬之技術領域】

發明領域

本發明係關於從微生物(包括真核生物，諸如破囊壺菌
5 (Thraustochytrid)微生物)來之多不飽和脂肪酸(PUFA)聚乙
醯合成酶(PKS)系統。更特別地是，本發明係關於一種編碼
非細菌的PUFA PKS系統之核酸；關於非細菌的PUFA PKS
系統；關於包含非細菌的PUFA PKS系統之經基因改性的生
物；及關於在本文中所揭示的非細菌PUFA PKS系統之製造
10 及使用方法。本發明亦關於一種鑑別包含PUFA PKS系統的
細菌與非細菌微生物之方法。

【先前技術】

發明背景

聚乙醯合成酶(PKS)系統一般已在技藝中熟知為源自
15 於脂肪酸合成酶(FAS)系統的酵素複合體，但是其經常高度
地改性以製造典型地會顯示出與脂肪酸有些微相似的特別
產物。發現者已嘗試開發已於文獻中描述的分屬於三種基
本型式之聚乙醯合成酶(PKS)系統，其典型地指為：型式II、
型式I及組合式(modular)。型式II系統之特徵為可分離的蛋
20 白質，其每個皆可進行一明顯的酵素反應。該些酵素會協
同地工作以製造出最後產物，且系統的每個各別的酵素在
製造最後產物中典型地會參與數次。此型式的系統以類似
於在植物及細菌中發現的脂肪酸合成酶(FAS)系統之方式
操作。型式I PKS系統類似於型式II系統，該些酵素使用迭

代(iterative)方式來製造最後產物。型式I與型式II在酵素活性上不同，其不與分離的蛋白質有關，而以較大的蛋白質區段(domain)進行。此系統與在動物及真菌中發現的型式I FAS系統類似。

5 比較型式I及II系統，在組合式PKS系統中，每個酵素區段在製造最後產物時僅使用一次。該些區段可在非常大的蛋白質中發現，且每個反應的產物會傳遞至在PKS蛋白質中的另一個區段。額外地，在上述描述的全部PKS系統中，若有碳-碳雙鍵併入最後產物，則其總是反式結構。

10 在上述描述的型式I及型式II PKS系統中，會在每次循環中進行相同的反應設定直到獲得最後產物。在生物合成程序期間並不允許引進獨特的反應。該組合式PKS系統則需要並不使用經濟的迭代反應之巨大蛋白質(即，對每個反應需要不同的區段)。額外地，如上述所描述，在先前描述
15 的全部PKS系統中，碳-碳雙鍵皆以反式結構引進。

 多不飽和脂肪酸類(PUFAs)為大部分真菌類中薄膜脂質的重要組分(羅利森(Lauritzen)等人，*Prog. Lipid Res.* 401(2001)；麥克康(McConn)等人，植物期刊(*Plant J.*)15，521(1998))，且為某些荷爾蒙及信號分子的前驅物(海勒(Heller)等人，藥物(*Drugs*)55，487(1998)；奎利爾門(Creelman)等人，*Annu. Rev. Plant Physiol. Plant Mol. Biol.* 48，355(1997))。熟知的PUFA合成途徑包括藉由延長及有
20 氧性去飽和反應加工來自脂肪酸合成酶(FAS)的飽和16：0或18：0脂肪酸(縮寫X：Y指為包含X個碳原子及Y個順式雙

鍵之醯基；PUFAs的雙鍵位置由雙鍵的系統性亞甲基中斷相對於脂肪酸鏈(T3或T6)的甲基碳而指出)(史柏雷求(Sprecher), *Curr. Opin. Clin. Nutr. Metab. Care* 2, 135(1999); 帕客-巴能斯(Parker-Barnes)等人, *Proc. Natl. Acad. Sci. USA* 97, 8284(2000); 仙克寧(Shanklin)等人, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49, 611(1998))。起始自乙醯基-CoA, DHA的合成需要約30個不同的酵素活性及幾乎70個反應(包括四個反覆的脂肪酸合成循環步驟)。聚乙醯合成(PKSs)會進行某些與FAS相同的反應(哈普伍得(Hopwood)等人, *Annu. Rev. Genet.* 24, 37(1990); 班特利(Bentley)等人, *Annu. Rev. Microbiol.* 53, 411(1999))及使用相同的小蛋白質(或區段)、醯基載體蛋白質(ACP)作為成長碳鏈的共價附著位置。但是,在這些酵素系統中,經常省略在FAS中看見的還原、脫水及還原之完全循環,所以可產生高度衍生的碳鏈,其典型地包含許多酮基及羥基基團和反式結構的碳-碳雙鍵。PKSs的線性產物經常會環化以形成複雜的生物化學物,其包括抗生素及許多其它二級產物(哈普伍得等人, (1990)前述; 班特利等人, (1999), 前述; 奇聽(Keating)等人, *Curr. Opin. Chem. Biol.* 3, 598(1999))。

非常長鏈之PUFAs(諸如廿二碳六烯酸(DHA; 22:6T3)及廿十碳五烯酸(EPA; 20:5T3))已報導可來自一些海洋細菌物種,包括志瓦菌 sp. (*Shewanella* sp.) (尼丘斯(Nichols)等人, *Curr. Op. Biotechnol.* 10, 240(1999); 葉日瓦(Yazawa), *Lipids* 31, S(1996); 狄龍(DeLong)等人, *Appl.*

Environ. Microbiol. 51, 730(1986))。從志瓦菌 sp. 菌株 SCRC2738 來的基因組片段(選殖為質體 pEPA)之分析導致鑑別出五個開放性讀碼區(Orfs)，總計 20Kb，此對在大腸桿菌中製造 EPA 來說是需要且足夠的(葉日瓦，(1996)，如前述)。一些預測的蛋白質區段為 FAS 酵素的同源物，然而其它區域則顯示出不與熟知功能的蛋白質同源。根據這些觀察及生物化學的研究，已建議在志瓦菌中的 PUFA 合成包括由 FAS 製造的 16-或 18-碳脂肪酸之延長，及利用不明瞭的有
5 氧性去飽和酶來嵌入雙鍵(瓦塔納貝(Watanabe)等人，*J. Biochem.* 122, 467(1997))。對此假設為錯誤的了解乃起始自由五個志瓦菌 Orfs 所譯出的蛋白質序列之再檢查。在五個 Orfs 中的至少 11 區域可識別為推斷的酵素區段(參見美茲(Metz)等人，*Science* 293: 290-293(2001))。當與在基因庫中的序列比較時，這些中的七個更強烈地與 PKS 蛋白質有關而超過與 FAS 蛋白質。包含在此族群中的為可推定地譯出
15 丙二酸單醯基-CoA: ACP 轉醯酶(MAT)、3-丙脂醯基-ACP 合成酶(KS)、3-丙脂醯基-ACP 還原酶(KR)、轉醯酶(AT)、磷酸泛醯醯基乙胺轉位酶、鏈長(或鏈起始)因子(CLF)之區段，及六個 ACP 區段之高度獨特的分子團(即，多於二團簇
20 的 ACP 區段尚未先前地報導有存在於 PKS 或 FAS 序列中)。但是，此三個區域與細菌的 FAS 蛋白質更高度地類似。這些之一則類似於近來描述的從肺炎鏈球菌(*Streptococcus pneumoniae*)來之抗二氯苯氧氣酚烯醯還原酶(ER)(喜斯(Heath)等人，*Nature* 406,145(2000))；使用 LALIGN 程式(矩

陣，BLOSUM50；間隔開口罰分(gap opening penalty)，-10；延長罰分(elongation penalty)-1)，比較ORF8胜肽與肺炎鏈球菌烯醯還原酶可指出有49%的類似性而超過386aa重疊。此二個區域為由*fabA*譯出的大腸桿菌FAS蛋白質之同源物，其可催化反-2-癸烯醯-ACP之合成及此產物至順-3-癸烯醯-ACP之可逆的異構化反應(喜斯等人，*J. Biol. Chem.*，271，27795(1996))。基於此，看起來似乎是在從志瓦菌來的EPA中，其雙鍵至少某些乃由在Orf7中之類FabA的區段催化之脫水酶-異構酶機構引進。

10 懷有pEPA質體而厭氧生長的大腸桿菌細胞會將EPA累積至與有氧培養基相同的程度(美茲等人，2001，前述)，此指出具氧依賴性的去飽和酶與EPA合成無關。當pEPA引進大腸桿菌(其不能合成單不飽和脂肪酸，但生長需要不飽和脂肪酸)的*fabB*⁻突變物時，所產生的細胞會遺失其脂肪酸營養缺陷性(auxotrophy)。它們亦會比其它含pEPA的菌株累積有更高程度的EPA，故建議EPA會與內生製造的用來轉染成甘油酯之單不飽和脂肪酸競爭。當含pEPA的大腸桿菌細胞於¹³C-醋酸酯的存在下生長時，從該些細胞來之經純化的EPA之¹³C-NMR分析資料可確定EPA的相同性，且提供此脂肪酸從乙醯基-CoA及丙二酸單醯基-CoA合成的證據(參見美茲等人，2001，前述)。從含pEPA的*fabB*⁻細胞來之無細胞均漿可合成從¹⁴C-丙二酸單醯基-CoA來的EPA及飽和脂肪酸二者。當該均漿分離成200,000xg的高速小粒與無薄膜上層液片段時，飽和脂肪酸合成局限在上層液，而與型

式II FAS酵素的可溶本質一致(馬格紐松(Magnuson)等人, *Microbiol. Rev.* 57, 522(1993))。EPA之合成已發現僅在高速小粒片段中, 此指示出EPA合成可在不依賴大腸桿菌FAS酵素或從細胞質片段來之可溶的中間體(諸如16:0-ACP)下發生。因為由志瓦菌EPA基因譯出的蛋白質不特別地為疏水的, EPA合成活性對此片段的限制可反映出其對與薄膜有關的醯基接收分子之需要。額外地, 與大腸桿菌FAS比較, EPA合成特別地具NADPH依賴性且不需要NADH。這些結果全部與編碼多官能性PKS(其可各自獨立地作用為FAS、鏈加長酶及去飽和酶活性, 以直接地合成EPA)的pEPA基因一致。類似地, 已在志瓦菌中鑑別的PUFA合成之PKS途徑在海洋細菌中很普遍。與志瓦菌基因團簇具有高同源性的基因已在深淵發光菌(*Photobacterium profundum*)(愛倫(Allen)等人, *Appli. Environ. Microbiol.* 65: 1710(1999))及在海洋莫瑞特菌(*Moritella marina*) (海洋弧菌(*Vibrio marinus*))(塔拿卡(Tanaka)等人, *Biotechnol. Lett.* 21: 939(1999))中鑑別出。

對志瓦菌進行生物化學及分子基因分析則可提供聚乙醯合成酶(其能合成從丙二酸單醯基-CoA來的PUFAs)具令人信服的證據。志瓦菌PKS的EPA合成之完整方法已建議。蛋白質區段之鑑別則類似於大腸桿菌FabA蛋白質, 且觀察到之厭氧發生的細菌EPA合成則可提供順式雙鍵之嵌入乃透過雙官能的脫水酶/2-反,3-順異構酶(DH/2,3I)作用而發生之機制的證據。在大腸桿菌中, 3-順式醯基中間體與丙

二酸單醯基-ACP之縮合需要特別的丙脂醯基-ACP合成酶，此可提供有二個KS存在於志瓦菌基因團中(在Orf5及Orf7)之原理。但是，PKS循環會以二個碳增加之方式來延長鏈，同時在EP產物中之雙鍵會發生在每第三個碳處。若在EPA之C-14及C-8處的雙鍵由2-反,2-順異構化反應(DH/2,2I)，接著將該順式雙鍵併入該延長的脂肪酸鏈而產生時，則可解決此分離問題。反式雙鍵酵素地轉染至順式結構而沒有鍵漂移已熟知會例如在類視色素循環中的11-順-視黃醛之合成中發生(珍(Jang)等人，*J. Biol. Chem.* 275, 28128(2000))。雖然此酵素功能尚未在志瓦菌PKS中鑑別，其可歸於未指定的蛋白質區段之一。

在志瓦菌及另一種海洋細菌(海洋弧菌)中的PUFA合成之PKS途徑，則詳細地描述在美國專利案號6,140,486中(從1998年6月4日申請的美國申請序號09/090,793發證，發表名稱為“藉由聚乙醯似的合成基因在植物中之表現來製造多不飽和脂肪酸”，其全文以參考之方式併於本文)。

多不飽和脂肪酸類(PUFAs)視為可有用地用在營養、醫藥、工業及其它用途上。從天然來源及化學合成來的PUFAs之廣泛供應並不足以應付商業需要。因為對從亞麻油酸(LA, 18:2)_{9,12}(在大部分的植物物種中常見)變換至更飽和及較長鏈PUFAs之脂肪酸合成上需要一些分離的去飽和酶及鏈加長酶酵素，在設計(engineering)用來表現PUFAs(諸如EPA及DHA)之植物宿主細胞時，則需要表現五或六個分離的酵素活性以獲得表現性(至少對EPA及DHA)。額外地，

為了製造有用量的此PUFAs，需要額外的設計成果，例如需向下調節酵素對基質的競爭、設計出具有較高的酵素活性(諸如藉由致突變)或標的(targeting)質體胞器的酵素。因此有興趣的是獲得與從可天然地製造這些脂肪酸的物種之

5 PUFAs生物合成有關的基因物質，及在可操控以允許製造出商業量的PUFAs之異種系統中單獨或組合地表現出該分離的物質。

在海洋細菌(諸如志瓦菌及海洋弧菌)中發現的PUFA PKS系統(參見美國專利案號6,140,486，同前所述)，對商業

10 的PUFA製造方法提供了新的對策。但是，這些海洋細菌有一些限制，而此限制最後終將限制其在商業上的有用程度。首先，雖然美國專利案號6,140,486揭示出海洋細菌PUFA PKS系統可使用來基因地改性植物，因為海洋細菌天然地生活及生長在冷的海洋環境中，故這些細菌的酵素系

15 統在大於30EC時並不會良好地作用。比較上，許多莊稼植物(其為有吸引力的使用PUFA PKS系統來基因操控的目標)具有在大於30EC的溫度下正常生長的條件，其範圍可到高於40EC。因此，海洋細菌PUFA PKS系統並不預期可容易地合適於在正常生長條件下的植物表現性。再者，海洋細菌

20 PUFAs PKS基因(其來自細菌來源)可能不會與真核宿主細胞的基因組相容，或至少需要明顯地適應以便在真核宿主中作用。額外地，熟知的海洋細菌PUFA PKS系統並不會直接地製造三酸甘油酯，然而想要的是能直接地製造三酸甘油酯，因為三酸甘油酯為一種微生物的脂質儲存產物，因此

可以在微生物/植物細胞中累積至非常高的程度(例如最高為細胞重量的80-85%)(與“結構”脂質產物(例如磷脂類)相反，其通常僅可低程度地累積(例如最大為少於細胞重量的10-15%))。

- 5 因此，在技藝中需要其它具有較大的彈性而可用於商業用途之PUFA PKS系統。

【發明內容】

發明概要

- 本發明的一個具體實施例係關於一種分離的核酸分子，其包含一種選自下列的核酸序列：(a)一核酸序列，其編碼選自於由下列所組成之群的胺基酸序列：序列辨識編號：2、序列辨識編號：4、序列辨識編號：6及其生物活性片段；(b)一核酸序列，其編碼選自於由下列所組成之群的胺基酸序列：序列辨識編號：8、序列辨識編號：10、序列辨識編號：13、序列辨識編號：18、序列辨識編號：20、序列辨識編號：22、序列辨識編號：24、序列辨識編號：26、序列辨識編號：28、序列辨識編號：30、序列辨識編號：32及其生物活性片段；(c)一核酸序列，其編碼具有與(a)之胺基酸序列的至少500個連續胺基酸有至少約60%相同之胺基酸序列，其中該胺基酸序列具有多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統的至少一個區段之生物活性；(d)一核酸序列，其編碼與(b)之胺基酸序列有至少約60%相同的胺基酸序列，其中該胺基酸序列具有多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統的至少一個區段之生物

活性；及(e)一核酸序列，其與核酸序列(a)、(b)、(c)或(d)完全地互補。在另一個觀點中，該核酸序列編碼的胺基酸序列與下列有至少約70%相同，或至少約80%相同，或至少約90%相同：(1)選自於由下列所組成之群的胺基酸序列之至少500個連續胺基酸：序列辨識編號：2、序列辨識編號：4及序列辨識編號：6；及/或(2)一核酸序列，其編碼與選自於由下列所組成之群的胺基酸序列有至少約70%相同的胺基酸序列：序列辨識編號：8、序列辨識編號：10、序列辨識編號：13、序列辨識編號：18、序列辨識編號：20、序列辨識編號：22、序列辨識編號：24、序列辨識編號：26、序列辨識編號：28、序列辨識編號：30及序列辨識編號：32。在一個較佳的具體實施例中，該核酸序列編碼一種選自於下列的胺基酸序列：序列辨識編號：2、序列辨識編號：4、序列辨識編號：6、序列辨識編號：8、序列辨識編號：10、序列辨識編號：13、序列辨識編號：18、序列辨識編號：20、序列辨識編號：22、序列辨識編號：24、序列辨識編號：26、序列辨識編號：28、序列辨識編號：30、序列辨識編號：32及/或其生物活性片段。在一個觀點中，該核酸序列選自於下列：序列辨識編號：1、序列辨識編號：3、序列辨識編號：5、序列辨識編號：7、序列辨識編號：9、序列辨識編號：12、序列辨識編號：17、序列辨識編號：19、序列辨識編號：21、序列辨識編號：23、序列辨識編號：25、序列辨識編號：27、序列辨識編號：29及序列辨識編號：31。

本發明的另一個具體實施例係關於一種包含如上所述之核酸分子的重組型核酸分子，其操作地連結到至少一種轉錄控制序列。在另一個具體實施例中，本發明係關於一種以上述直接描述之重組型核酸分子來轉染的重組細胞。

- 5 本發明的更另一個具體實施例係關於一種經基因改性的微生物，其中該微生物可表現出包含一種具有多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統的至少一個生物活性區段之PKS系統。該PUFA PKS系統的至少一個區段可由選自於下列的核酸序列譯出：(a)一核酸序列，其編碼從破
- 10 囊壺菌微生物來的多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統之至少一個區段；(b)一核酸序列，其編碼從一微生物(其可利用本發明之篩選方法來鑑別)來的PUFA PKS系統之至少一個區段；(c)一核酸序列，其編碼選自於由下列所組成之群的胺基酸序列：序列辨識編號：2、序列辨識編
- 15 號：4、序列辨識編號：6及其生物活性片段；(d)一核酸序列，其編碼選自於由下列所組成之群的胺基酸序列：序列辨識編號：8、序列辨識編號：10、序列辨識編號：13、序列辨識編號：18、序列辨識編號：20、序列辨識編號：22、序列辨識編號：24、序列辨識編號：26、序列辨識編號：
- 20 28、序列辨識編號：30、序列辨識編號：32及其生物活性片段；(e)一核酸序列，其編碼一胺基酸序列其與選自於由下列所組成之群的胺基酸序列：序列辨識編號：2、序列辨識編號：4及序列辨識編號：6之至少500個連續胺基酸有至少約60%相同，其中該胺基酸序列具有PUFA PKS系統的至

少一個區段之生物活性；及，(f)一核酸序列，其編碼與選
自於由下列所組成之群的胺基酸序列有至少約60%相同之
胺基酸序列：序列辨識編號：8、序列辨識編號：10、序列
辨識編號：13、序列辨識編號：18、序列辨識編號：20、
5 序列辨識編號：22、序列辨識編號：24、序列辨識編號：
26、序列辨識編號：28、序列辨識編號：30及序列辨識編
號：32，其中該胺基酸序列具有PUFA PKS系統的至少一個
區段之生物活性。在此具體實施例中，該微生物經基因地
改性以影響PKS系統的活性。本發明在上述(b)中提及的篩
10 選方法包括：(i)選擇一可製造至少一種PUFA的微生物；
及，(ii)鑑別從(i)來的微生物，其具有在發酵培養液中於溶
氧條件少於約飽和的5%下可製造增加的PUFAs之能力，而
可與在發酵培養液中於溶氧條件大於飽和的5%下由微生
物製造之PUFAs比較，更佳地為飽和的10%，更佳地大於飽
15 和的15%，更佳地大於飽和的20%。

在一個觀點中，該微生物可內生地表現包含該PUFA
PKS系統的至少一個區段之PKS系統，及其中該基因改性在
一編碼PUFA PKS系統的至少一個區段之核酸序列中。例
如，該基因改性可在一能譯出具有至少一種下列蛋白質的
20 生物活性之區段的核酸序列中：丙二酸單醯基-CoA：ACP
轉醯酶(MAT)、 β -酮基醯基-ACP合成酶(KS)、酮還原酶
(KR)、轉醯酶(AT)、類FabA的 β -羥基醯基-ACP脫水酶
(DH)、磷酸泛醯巯基乙胺轉位酶、鏈長因子(CLF)、醯基載
體蛋白質(ACP)、烯醯ACP-還原酶(ER)、可催化反-2-癸烯

醯-ACP合成的酵素、可催化反-2-癸烯醯-ACP至順-3-癸烯
 醯-ACP之可逆的異構化反應之酵素、及可催化順-3-癸烯醯
 -ACP至順-法生油酸的延長反應之酵素。在一個觀點中，該
 基因改性在一編碼選自於由下列所組成之群的胺基酸序列
 5 之核酸序列中：(a)一胺基酸序列，其與選自於由下列所組
 成之群的胺基酸序列：序列辨識編號：2、序列辨識編號：
 4及序列辨識編號：6之至少500個連續胺基酸有至少約70%
 相同，較佳地至少約80%相同及更佳地至少約90%相同；其
 中該胺基酸序列具有PUFA PKS系統的至少一個區段之生
 10 物活性；及，(b)一胺基酸序列，其與選自於由下列所組成
 之群的胺基酸序列有至少約70%相同：序列辨識編號：8、
 序列辨識編號：10、序列辨識編號：13、序列辨識編號：
 18、序列辨識編號：20、序列辨識編號：22、序列辨識編
 號：24、序列辨識編號：26、序列辨識編號：28、序列辨
 15 識編號：30，及序列辨識編號：32，較佳地至少約80%相
 同及更佳地至少約90%相同；其中該胺基酸序列具有PUFA
 PKS系統的至少一個區段之生物活性。

在一個觀點中，該經基因改性的微生物為一種破囊壺
 菌，其可包括(但是非為限制)選自於下列之屬的破囊壺菌：
 20 裂壺菌屬(*Schizochytrium*)及破囊壺菌屬。在另一個觀點
 中，該微生物已進一步地基因改性以重組地表現出至少一
 種編碼至少一個從細菌的PUFA PKS系統、從型式I PKS系
 統、從型式II PKS系統及/或從組合式PKS系統來之生物活
 性區段的核酸分子。

在此具體實施例的另一個觀點中，該微生物可內生地表現出一種包含PUFA PKS系統的至少一個生物活性區段之PUFA PKS系統，其中該基因改性包含選自於由下列所組成之群的重組型核酸分子之表現性：一種編碼至少一個從第二PKS系統來的生物活性區段之重組型核酸分子，及一種編碼能影響PUFA PKS系統活性之蛋白質的重組型核酸分子。較佳地，該重組型核酸分子包含上述描述的核酸序列之任何一種。

在此具體實施例的一個觀點中，該重組型核酸分子編碼一種磷酸泛醌巯基乙胺轉位酶。在另一個觀點中，該重組型核酸分子包含一種編碼至少一個從細菌的PUFA PKS系統、型式I PKS系統、型式II PKS系統及/或組合式PKS系統來之生物活性區段的核酸序列。

在此具體實施例的另一個觀點中，該微生物藉由以編碼多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統的至少一個區段之重組型核酸分子轉染而基因地改性。此重組型核酸分子可包括任何具有任何上述描述的核酸序列之重組型核酸分子。在一個觀點中，該微生物已進一步地基因改性以重組地表現出至少一種編碼至少一個從細菌的PUFA PKS系統、型式I PKS系統、型式II PKS系統或組合式PKS系統來的生物活性區段之核酸分子。

本發明的更另一個具體實施例係關於一種經基因改性的植物，其中該植物已基因地改性以重組地表現出包含多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統的至少一個

生物活性區段之PKS系統。該區段可由任何上述描述的核酸序列譯出。在一個觀點中，該植物已進一步地基因改性以重組地表現出至少一種編碼至少一個從細菌的PUFA PKS系統、型式I PKS系統、型式II PKS系統及組合式PKS

5 系統來的生物活性區段之核酸分子。

本發明的另一個具體實施例係關於一種鑑別具有多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統的微生物之方法。該方法包括的步驟有：(a)選擇一可製造至少一種PUFA的微生物；及，(b)鑑別從(a)來的微生物，其具有在發酵培養液中於溶氧條件少於約飽和的5%下可製造增加的PUFAs

10 之能力，如與在發酵培養液中溶氧條件大於飽和的5%下由微生物製造之PUFAs比較，更佳地為飽和的10%，更佳地大於飽和的15%及更佳地大於飽和的20%。可製造至少一種PUFA且具有在溶氧條件少於約飽和的5%下可製造增加的

15 PUFAs之能力的微生物可鑑別為包含PUFA PKS系統的候選生物。

在此具體實施例的一個觀點中，步驟(b)包括鑑別從(a)來的微生物，其具有在溶解的氧少於約飽和的2%之條件下製造增加的PUFAs之能力，及更佳地在溶解的氧少於約飽和的1%之條件下，甚至更佳地在溶解條件約0%飽和下。

20

在此具體實施例的另一個觀點中，該在(a)中選擇的微生物具有藉由吞噬作用吃掉細菌的能力。在另一個觀點中，該在(a)中選擇的微生物具有簡單的脂肪酸分佈(profile)。在另一個觀點中，該在(a)中選擇的微生物為一種

非細菌的微生物。在另一個觀點中，該在(a)中選擇的微生物為一種真核生物。在另一個觀點中，該在(a)中選擇的微生物為一種破囊壺菌目成員。在另一個觀點中，該在(a)中選擇的微生物具有可在溫度大於約15EC下製造PUFAs的能力，較佳地大於約20EC，更佳地大於約25EC及甚至更佳地大於約30EC。在另一個觀點中，該在(a)中選擇的微生物具有製造大於生物乾燥重量的5%之有興趣的生物活性化合物(例如，脂質)之能力，及更佳地大於生物乾燥重量的10%。在更另一個觀點中，該在(a)中選擇的微生物包括大於其總脂肪酸的30%之C14:0, C16:0及C16:1，同時亦可製造至少一種具有三個或更多不飽和鍵的長鏈脂肪酸，較佳地，該在(a)中選擇的微生物包括大於其總脂肪酸的40%之C14:0、C16:0及C16:1，同時亦可製至少一種具有三個或更多不飽和鍵的長鏈脂肪酸。在另一個觀點中，該在(a)中選擇的微生物包括大於其總脂肪酸的30%之C14:0, C16:0及C16:1，同時亦可製造至少一種具有四個或更多不飽和鍵的長鏈脂肪酸，更佳地同時亦可製造至少一種具有五個或更多不飽和鍵的長鏈脂肪酸。

在此具體實施例的另一個觀點中，該方法進一步包含步驟(c)偵測該生物是否包含PUFA PKS系統。在此觀點中，該偵測步驟可包括偵測一種在微生物中可與能譯出從破囊壺菌PUFA PKS系統來的胺基酸序列之核酸序列於嚴苛的條件下雜交之核酸序列。再者，該偵測步驟可包括偵測一種在該生物中可由從破囊壺菌 PUFA PKS系統來的核酸序

列來之寡核苷引子放大的核酸序列。

本發明的另一個具體實施例係關於一種可利用上述描述的篩選方法鑑別的微生物，其中該微生物已基因地改性以調節由PUFA PKS系統製造的分子。

5 本發明的更另一個具體實施例係關於一種由聚乙醯合成酶系統製造之生物活性分子的製造方法。該方法包括的步驟有在可有效地製造該生物活性分子的條件下培養一基因改性的生物(其可表現出包含多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統的至少一個生物活性區段之PKS系
10 統)。該PUFA PKS系統的區段可由任何上述描述的核酸序列譯出。

 在此具體實施例的一個觀點中，該生物可內生地表現出包含PUFA PKS系統的至少一個區段之PKS系統，且該基因改性在編碼PUFA PKS系統的至少一個區段之核酸序列
15 中。例如，該基因改性可改變至少一種由該內生的PKS系統所製造之產物，如與野生型生物比較。

 在此具體實施例的另一個觀點中，該生物可內生地表現出包含PUFA PKS系統的至少一個生物活性區段之PKS系統，且該基因改性包含以選自於由下列所組成之群的重
20 組型核酸分子來轉染該生物：一種編碼從第二PKS系統來的至少一個生物活性區段之重組型核酸分子，及一種編碼會影響PUFA PKS系統活性的蛋白質之重組型核酸分子。例如，該基因改性可改變至少一種由內生的PKS系統所製造之產物，如與野生型生物比較。

在此具體實施例的更另一個觀點中，該生物可藉由以能譯出多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統的至少一個區段之重組型核酸分子來轉染而基因改性。在另一個觀點中，該生物可製造一種與沒有經基因改性之天然發生的生物不同之多不飽和脂肪酸(PUFA)分佈。在另一個觀點中，該生物可內生地表現出一種非細菌的PUFA PKS系統，及其中該基因改性包含以從不同的PKS系統來之區段取代編碼該非細菌PUFA PKS系統的至少一個區段之核酸序列。

10 在更另一個觀點中，該生物可內生地表現出一非細菌的PUFA PKS系統，其已藉由以編碼能調節由PUFA PKS系統製造的脂肪酸之鏈長的蛋白質之重組型核酸分子來轉染該生物而改性。例如，該編碼能調節脂肪酸的鏈長之蛋白質的重組型核酸分子可取代能在非細菌的PUFA PKS系統中譯出鏈長因子之核酸序列。在另一個觀點中，該可調節由PUFA PKS系統製造之脂肪酸的鏈長之蛋白質為一種鏈長因子。在另一個觀點中，該可調節由PUFA PKS系統製造的蛋白質之脂肪酸的鏈長為一種能指導C20單元合成的鏈長因子。

20 在一個觀點中，該生物可表現出一種在選自於下列的區段中包含基因改性的非細菌PUFA PKS系統：一個編碼類FabA的 β -羥基醯基-ACP脫水酶(DH)區段之區段，及一個編碼3-丙脂醯基-ACP合成酶(KS)的區段，其中該改性可改變由PUFA PKS系統製造的長鏈脂肪酸比率(如與缺乏改性的

比較)。在一個觀點中，該改性包含以不具有異構化活性的DH區段取代在非細菌PUFA PKS系統中之類FabA的 β -羥基醯基-ACP脫水酶(DH)。在另一個觀點中，該改性可選自於由下列所組成之群：刪除全部或部分區段、以從不同生物

5 來的同源區段取代該區段、及該區段之突變。

在另一個觀點中，該生物可表現一PKS系統且該基因改性包括以從PUFA PKS系統來之類FabA的 β -羥基醯基-ACP脫水酶(DH)區段取代不具有異構化活性的DH區段。

在另一個觀點中，該生物可表現出一包含在烯醯-ACP還原酶(ER)區段中改性的非細菌PUFA PKS系統，其中該改性可造成製造出不同的化合物(如與缺乏改性的比較)。例如，該改性可選自於由下列所組成之群：刪除全部或部分的ER區段、以從不同生物來的ER區段取代該ER區段、及該ER區段之突變。

10

在一個觀點中，可由本方法製造的生物活性分子包括(但是非為限制)抗炎性配方、化學治療劑、活性賦形劑、骨質疏鬆藥物、抗抑鬱劑、抗驚厥劑、抗幽門桿菌(*Helicobacter pylori*)藥、用來治療神經變性疾病的藥物、用來治療肝退化疾病的藥物、抗生素及降低膽固醇配方。在一個觀點中，

15

20 該生物活性分子為一種多不飽和脂肪酸(PUFA)。在另一個觀點中，該生物活性分子為一種包含順式碳-碳雙鍵結構的分子。在另一個觀點中，該生物活性分子為一種在每第三個碳處包含雙鍵的分子。

在此具體實施例的一個觀點中，該生物為一種微生

物，在另一個觀點中，該生物為一種植物。

本發明的另一個具體實施例係關於一種製造與天然發生的植物不同之多不飽和脂肪酸(PUFA)分佈的植物之方法，其包括基因地改性該植物的細胞以表現出包含至少一
5 種具有編碼PUFA PKS系統的至少一個生物活性區段之核酸序列的重組型核酸分子之PKS系統。該PUFA PKS系統的區段可由上述描述的任何核酸序列譯出。

本發明的更另一個具體實施例係關於一種包含至少一種脂肪酸的最後產物之改性方法，其包括將一種由可表現
10 出至少一種具有編碼PUFA PKS系統的至少一個生物活性區段之核酸序列的重組型核酸分子之重組宿主細胞所製造的油加入至該最後產物。該PUFA PKS系統的區段可由上述描述的任何核酸序列譯出。在一個觀點中，該最後產物選自於由下列所組成之群：膳食補充品、食物產品、醫藥配
15 方、母乳化動物乳及幼兒配方。該醫藥配方可包括(但是非為限制)：抗炎性配方、化學治療劑、活性賦形劑、骨質疏鬆藥物、抗抑鬱劑、抗驚厥劑、抗幽門桿菌藥物、用來治療神經變性疾病的藥物、用來治療肝退化疾病的藥物、抗
20 生素及降低膽固醇配方。在一個觀點中，該最後產物可使用來治療選自於由下列所組成之群的症狀：慢性發炎、急性發炎、胃腸病、癌、惡病質、心臟再阻塞、神經變性病、肝退化病症、血脂質失調、骨質疏鬆、骨關節炎、自體免疫疾病、初期子癇、早產、與年齡有關的斑狀丘疹、肺病及過氧化物酶體病。

本發明的更另一個具體實施例係關於一種母乳化動物乳的製造方法，其包含以至少一種具有編碼PUFA PKS系統的至少一個生物活性區段之核酸序列的重組型核酸分子來基因地改性製乳動物的製乳細胞。該PUFA PKS系統的區段可由上述描述的任何核酸序列譯出。

本發明的更另一個具體實施例係關於一種重組微生物的製造方法，其包含基因地性微生物細胞以表現出至少一種具有編碼PUFA PKS系統的至少一個生物活性區段之核酸序列的重組型核酸分子。該PUFA PKS系統區段可由上述描述的任何核酸序列譯出。

本發明的更另一個具體實施例係關於一種已改性以表現出多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統的重組宿主細胞，其中該PKS可催化迭代及非迭代酵素反應二者。該PUFA PKS系統包括：(a)至少二個烯醯ACP-還原酶(ER)區段；(b)至少六個醯基載體蛋白質(ACP)區段；(c)至少二個 β -酮基醯基-ACP合成酶(KS)區段；(d)至少一個轉醯酶(AT)區段；(e)至少一個酮還原酶(KR)區段；(f)至少二個類FabA的 β -羥基醯基-ACP脫水酶(DH)區段；(g)至少一個鏈長因子(CLF)區段；及(h)至少一個丙二酸單醯基-CoA:ACP轉醯酶(MAT)區段。在一個觀點中，該PUFA PKS系統為一種真核PUFA PKS系統。在另一個觀點中，該PUFA PKS系統為一種藻類的PUFA PKS系統，較佳地為破囊壺菌目的PUFA PKS系統，其可包括(但是為非為限制)裂壺菌屬PUFA PKS系統或破囊壺菌屬PUFA PKS系統。

在此具體實施例中，該PUFA PKS系統可在原核宿主細胞或在真核宿主細胞中表現。在一個觀點中，該宿主細胞為一種植物細胞。此外，本發明的一個具體實施例為一種包括至少一種PUFA的產物之製造方法，其包括在有效以製造該產物之條件下生長包含此植物細胞的植物。該宿主細胞為一種微生物細胞，且於此實例中，本發明的一個具體實施例為一種製造包含至少一種PUFA的產物之方法，其包括在有效以製造該產物之條件下培養一種包含此微生物細胞的培養基。在一個觀點中，該PKS系統可催化三酸甘油脂的直接製造。

本發明的更另一個具體實施例係關於一種包含多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統的經基因改性之微生物，其中該PKS可催化迭代及非迭代酵素反應二者。該PUFA PKS系統包含：(a)至少二個烯醯ACP-還原酶(ER)區段；(b)至少六個醯基載體蛋白質(ACP)區段；(c)至少二個 β -酮基醯基-ACP合成酶(KS)區段；(d)至少一個轉醯酶(AT)區段；(e)至少一個酮還原酶(KR)區段；(f)至少二個類FabA的 β -羥基醯基-ACP脫水酶(DH)區段；(g)至少一個鏈長因子(CLF)區段；及(h)至少一個丙二酸單醯基-CoA：ACP轉醯酶(MAT)區段。該基因改性可影響PUFA PKS系統的活性。在此具體實施例的一個觀點中，該微生物為一種真核微生物。

本發明的更另一個具體實施例係關於一種已改性而可表現出非細菌的多不飽和脂肪酸(PUFA)聚乙醯合成酶

(PKS)系統之重組宿主細胞，其中該非細菌的PUFA PKS可催化迭代及非迭代酵素反應二者。該非細菌的PUFA PKS系統包含：(a)至少一個烯醯ACP-還原酶(ER)區段；(b)多重的醯基載體蛋白質(ACP)區段；(c)至少二個 β -酮基醯基-ACP合成酶(KS)區段；(d)至少一個轉醯酶(AT)區段；(e)至少一個酮還原酶(KR)區段；(f)至少二個類FabA的 β -羥基醯基-ACP脫水酶(DH)區段；(g)至少一個鏈長因子(CLF)區段；及(h)至少一個丙二酸單醯基-CoA:ACP轉醯酶(MAT)區段。

圖式簡單說明

10 第1圖為裂壺菌屬PUFA PKS系統的區段結構圖；

第2圖為從裂壺菌屬及志瓦菌來的PKS區段之比較圖；

第3圖為從裂壺菌屬來之PKS區段與從念珠藻(*Nostoc sp.*)(其產物為一種不包含任何雙鍵的長鏈脂肪酸)來之相關的PKS系統之比較圖。

15 發明之詳細說明

本發明通常係關於一種非細菌衍生的多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統；關於包含非細菌PUFA PKS系統之基因改性的生物；關於製造及使用此系統來製造有興趣的產物(包括生物活性分子)之方法；及關於用來鑑別具有此PUFA PKS系統的新穎真核微生物之新穎的方法。如於本文中所使用，PUFA PKS系統通常具有下列鑑別特徵：(1)其可製造PUFAs，作為該系統的天然產物；及(2)其包含數個組合成複合體的多官能性蛋白質，其可進行脂肪酸鏈的迭代加工和非迭代加工二者，包括在選擇的循環

中之反-順異構化反應及烯醯還原反應(例如，參見第1圖)。

更特別地，首先，形成本發明之基礎的PUFA PKS系統可產生多不飽和脂肪酸類(PUFAs)作為產物(即，內生地(天然地)包含此PKS系統的生物可使用此系統來製得

5 PUFAs)。於本文指出的PUFAs較佳地為具有至少16個碳的碳鏈長度之多不飽和脂肪酸類，更佳地為至少18個碳，更佳地至少20個碳及更佳地為22或更多個碳，而具有至少3或更多個雙鍵，較佳地4或更多，更佳地5或更多，甚至更佳地6或更多個雙鍵，其中全部的雙鍵為順式結構。本發明之

10 目標為經由基因操控或最後產物的操控而發現或創造可製造出想要的鏈長度且具有想要的雙鍵數目之多不飽和脂肪酸類的PKS系統。PUFAs的實例包括(但是非為限制)DHA(廿二碳六烯酸(C22:6, T-3))、DPA(廿二碳五烯酸(C22:5, T-6))及EPA(廿十碳五烯酸(C20:5, T-3))。

15 其次，於本文描述的PUFA PKS系統可併合迭代及非迭代反應二者，其可區別出該系統與先前描述的PKS系統(例如，型式I、型式II或組合式)。更特別地是，於本文中描述的PUFA PKS系統包括在每個循環期間顯示出有作用和僅有在某些循環期間會顯示有作用的那些之區段。此重要觀

20 點與顯示出和細菌的Fab A酵素同源之區段相關。例如，大腸桿菌的Fab A酵素已顯示出具有二種酵素活性。其擁有一種可將水分子(H₂O)從含羥基的碳鏈中萃取出來而在其碳鏈中遺留下反式雙鍵之脫水活性。此外，其具有將該反式雙鍵轉染至順式結構的異構酶活性。此異構化反應可藉由

相關的雙鍵位置漂移至毗連的碳而達成。在PKS(及FAS)系統中，該主要的碳鏈會以2個碳增加的方式延長。因此可預測製造這些PKS系統的PUFA產物所需要之延長反應數目。例如，為了製造DHA(C22:6，全部順式)需要10個延長反應。因為在最後產物僅有6個雙鍵，此意謂著在反應循環的某些期間，需保留雙鍵(如為順式異構物)，而在其他期間，需在下一個延長之前還原雙鍵。

在海洋細菌之PUFA PKS系統發現前(參見美國專利案號6,140,486)，PKS系統並不熟知具有此迭代及選擇性酵素反應之組合，且並不認為它們能夠製造順式碳-碳雙鍵結構。但是，由本發明描述的PUFA PKS系統具有引進順式雙鍵的能力和改變化循環中的反應序列之能力。

因此，本發明家建議使用PUFA PKS系統的這些特徵來製造一定範圍的無法由先前描述(型式II、型式I及組合式)之PKS系統製造的生物活性分子。這些生物活性分子包括(但是非為限制)多不飽和脂肪酸類(PUFAs)、抗生素或其它生物活性化合物，其有許多將在下列討論。例如，使用於本文描述的PUFA PKS基因結構知識，可使用任何數種方法來改變PUFA PKS基因，或將這些基因的部分與其它合成系統(包括其它PKS系統)結合，如此以製造新穎產物。此特別型式的系統之固有的可進行迭代及選擇性反應二者之能力將使此系統能夠產生產物，其若將類似的方法應用至其它PKS系統型式時並不會發現。

在一個具體實施例中，根據本發明之PUFA PKS系統包

含至少下列的生物活性區段：(a)至少二個烯醯ACP-還原酶(ER)區段；(b)至少六個醯基載體蛋白質(ACP)區段；(c)至少二個 β -酮基醯基-ACP合成酶(KS)區段；(d)至少一個轉醯酶(AT)區段；(e)至少一個酮還原酶(KR)區段；(f)至少二個類FabA的 β -羥基醯基-ACP脫水酶(DH)區段；(g)至少一個鏈長因子(CLF)區段；及(h)至少一個丙二酸單醯基-CoA:ACP轉醯酶(MAT)區段。這些區段的功能通常各別地在技藝中熟知而將在下列詳細以與本發明有關之PUFA PKS系統來描述。

10 在另一個具體實施例中，該PUFA PKS系統包含至少下列的生物活性區段：(a)至少一個烯醯ACP-還原酶(ER)區段；(b)多個醯基載體蛋白質(ACP)區段(至少四個，較佳地至少五個，更佳地至少六個，甚至更佳地七個、八個、九個或多於九個)；(c)至少二個 β -酮基醯基-ACP合成酶(KS)區段；(d)至少一個轉醯酶(AT)區段；(e)至少一個酮還原酶(KR)區段；(f)至少二個類FabA的 β -羥基醯基-ACP脫水酶(DH)區段；(g)至少一個鏈長因子(CLF)區段；及(h)至少一個丙二酸單醯基-CoA:ACP轉醯酶(MAT)區段。較佳地，
15 此PUFA PKS系統為一種非細菌的PUFA-PKS系統。

20 在一個具體實施例中，本發明之PUFA PKS系統為一種非細菌的PUFA PKS系統。換句話說，在一個具體實施例中，本發明之PUFA PKS系統分離自非為細菌的生物，或為一種衍生自從非細菌的生物(諸如真核生物或原始細菌(archaebacterium))來之PUFA PKS系統的同源化合物。真菌

類根據細胞的分化程度而分離自原核生物。具有較高分化的高等族群稱為真核生物。具有較少分化細胞的低等族群稱為原核生物。通常地，原核生物不具有核膜，在細胞分裂期間不具有細胞有絲分裂而僅具有一染色體，其細胞質包含70S核糖體，它們不具有任何粒腺體、內質網、葉綠體、溶酶體或高爾基氏體，其鞭毛(若存在)由單一微纖維組成。比較上，真菌類具有核膜，它們在細胞分裂期間具有細胞有絲分裂，它們具有許多染色體，其細胞質包括80S核糖體，它們具有粒腺體、內質網、葉綠體(在藻類中)、溶酶體及高爾基氏體，及其鞭毛(若存在)由許多微纖維組成。通常地，細菌為原核生物，然而藻類、真菌、原生生物、原生動物門及高等植物為真菌類。海洋細菌(例如，志瓦菌及海洋弧菌)的PUFA PKS系統不為本發明之基本，雖然本發明考慮到使用從這些細菌PUFA PKS系統來的區段與從本發明之非細菌PUFA PKS系統來的區段連接。例如，根據本發明，基因改性的生物可藉由合併非細菌的PUFA PKS官能性區段與細菌的PUFA PKS官能性區段，和PKS官能性區段或從其它PKS系統(型式I、型式II、組合式)或FAS系統來的蛋白質而製造。

裂壺菌屬為一種破囊壺菌海洋微生物，其可累積大量的甘油三酯而富含DHA及廿二碳五烯酸(DPA；22：5T-6)；例如，乾燥重量的30%之DHA+DPA(巴克萊(Barclay)等人，*J. Appl. Phycol.* 6,123(1994))。在藉由延長/去飽和途徑來合成20-及22-碳PUFAs之真菌類中，18-、20-及22-碳的中間體

庫(pools)相當大，所以使用 $[^{14}\text{C}]$ -醋酸酯的活體內標定實驗可顯示出清楚的前驅物-產物動力學以用來預測中間體(傑樂門 (Gellerman) 等人，*Biochim. Biophys. Acta* 573 : 23(1979))。再者，以放射性同位素示蹤而外源地提供至此生物的
5 中間體可轉染至最後PUFA產物。本發明家已顯示出 $[1-^{14}\text{C}]$ -醋酸酯會由裂壺菌屬細胞快速地取用而併入脂肪酸中，而且在最短的標定時間(1分鐘)處，在脂肪酸中回收的DHA包含31%的標籤物，而在 $[^{14}\text{C}]$ -醋酸酯併入10-15分鐘及連續培養生長24小時的期間，此百分比基本上保持無變化
10 (參見實例3)。類似地，遍及實驗DPA表示出10%的標籤物。並無證據顯示在16-或18-碳脂肪酸與22-碳多不飽和脂肪酸間有前驅物-產物關係。這些結果與從包含非常小量的(大概酵素-包圍)中間體庫之 $[^{14}\text{C}]$ -醋酸酯來的DHA快速合成一致。源自於裂壺菌屬培養物的無細胞均漿會將 $[1-^{14}\text{C}]$ -丙二
15 酸單醯基-CoA併入DHA、DPA及飽和脂肪酸。相同的生物合成活性可由100,000xg上層液部分保留，但是不呈現在薄膜小粒中。因此，在裂壺菌屬中的DHA及DPA合成不包括薄膜-結合的(membrane-bound)去飽和酶或脂肪酸延長酵素類，如描述於其它真菌類的那些(帕客-巴能斯等人，2000，
20 前述；仙克寧等人，1998，前述)。這些分級資料可與從志瓦菌酵素獲得的那些比較(參見美茲等人，2001，前述)，且指示出裂壺菌屬酵素使用不同的(可溶的)醯基接收分子，諸如CoA。

在共審查的美國申請序號09/231,899中，已建構出從裂

壺菌屬來的cDNA基因庫且已定序約8,000個隨機(random)選殖物(ESTs)。在此資料組中，僅有一種適度表現的基因(全部序列的0.3%)可鑑別為脂肪酸去飽和酶，雖然第二推定的去飽和酶可由單一選殖物(0.01%)表示。相較之下，具有與志瓦菌PKS基因的11個區段之8個類似的序列(顯示在第2圖)全部在0.2-0.5%的頻率處鑑別出。在美國申請序號09/231,899中，一些顯示出與志瓦菌PKS基因同源的cDNA選殖物已定序，且不同的選殖物可組合進入核酸序列以表示二個部分開放性讀碼區及一個完全開放性讀碼區。包含在美國申請序號09/231,899中描述的第一部分開放性讀碼區之cDNA序列的核苷酸390-4443(在其中表示為序列辨識編號：69)與於本文中表示為OrfA(序列辨識編號：1)的核苷酸4677-8730(加上終止密碼子(stop codon))序列相符(match)。包含在美國申請序號09/231,899中描述的第二部分開放性讀碼區之cDNA序列的核苷酸1-4876(在其中表示為序列辨識編號：71)與於本文中表示為OrfB(序列辨識編號：3)的核苷酸1311-6177(加上終止密碼子)序列相符。包含在美國申請序號09/231,899中描述的完全開放性讀碼區之cDNA序列的核苷酸145-4653(在其中表示為序列辨識編號：76及錯誤地標示為部分開放性讀碼區)與於本文中表示為OrfC(序列辨識編號：5)的序列之全部序列(加上終止密碼子)相符。

由本發明家進一步定序的cDNA及基因組選殖物可允許鑑別OrfA、OrfB及OrfC每個的全長基因組序列，及完全

鑑別該些與在志瓦菌中的那些同源之區段(參見第2圖)。應注意的是在裂壺菌屬中，由於在生物基因組中缺乏插入子，對本發明家的最佳知識來說，基因組DNA及cDNA相同。因此，關於從裂壺菌屬來的核苷序列可指為基因組DNA
5 或cDNA。根據裂壺菌屬PKS區段與志瓦菌的比較，明顯地，裂壺菌屬基因組編碼的蛋白質高度地與在志瓦菌中能催化EPA合成的蛋白質類似。在裂壺菌屬中的蛋白質可架構出一種可催化DHA及DPA合成的PUFA PKS系統。如於本文中詳細地討論，鑑別志瓦菌用之反應方法的簡單改性將
10 允許用於在裂壺菌屬中的DHA合成。原核生物志瓦菌與真核裂壺菌屬基因間之同源性建議，PUFA PKS已進行橫向的基因轉染。

第1圖為從裂壺菌屬PUFA PKS系統來的三個開放性讀碼區之圖示及包含此PUFA PKS系統的區段結構。如在下列
15 的實例1中所描述，每個開放性讀碼區的區段結構如下：

開放性讀碼區A(OrfA)：

OrfA的完整核苷酸序列於本文表示為序列辨識編號：
1。序列辨識編號：1之核苷酸4677-8730與在美國申請序號09/231,899中表示為序列辨識編號：69的核苷酸390-4443序
20 列相符合。因此，序列辨識編號：1之核苷酸1-4676顯示出不在美國申請序號09/231,899中揭示的額外序列。此序列辨識編號：1的新區段可在OrfA中譯出下列區段：(1)ORFA-KS區段；(2)ORFA-MAT區段；及(3)至少一部分的ACP區段區域(例如，至少ACP區段1-4)。應注意的是在美國申請序號

09/231,899中的序列辨識編號：69之核苷酸1-389與於本文揭示的序列辨識編號：1之位置4677上游的389核苷酸不相符。因此，在美國申請序號09/231,899中的序列辨識編號：69之位置1-389顯露出被錯誤地放置在鄰近於該序列的核苷酸390-4443。這些第一389個核苷酸大部分(約在位置5 60-389)與本發明的OrfA(序列辨識編號：1)之上游部分相符，因此，咸信錯誤是發生在美國申請序號09/231,899中製備cDNA構造的連續序列(contig)結果。在美國申請序號09/231,899中發生的排比錯誤區段在高重覆序列區段中10 (即，ACP區段，在下列討論)，其在組合從不同的cDNA選植物來之序列時大概會造成一些混淆。

OrfA為一8730核苷酸序列(不包括終止密碼子)，其編碼2910個胺基酸序列，於本文表示為序列辨識編號：2。在OrfA中有十二個區段：(a)一個 γ -酮基醯基-ACP合成酶(KS)區段；(b)一個丙二酸單醯基-CoA：ACP轉醯酶(MAT)區段；15 (c)九個醯基載體蛋白質(ACP)區段；及(d)一個酮還原酶(KR)區段。

OrfA的核苷酸序列已由基因銀行(GenBank)保存為編號AF378327(胺基酸序列編號AAK728879)。OrfA在標準BLAST搜尋中與熟知的序列比較(BLAST 2.0基本BLAST同源搜尋，以標準預設參數在全部6個開放性讀碼區中使用blastp做胺基酸搜尋，blastn做核酸搜尋及blastX做核酸搜尋及轉譯的胺基酸序列之搜尋，其中該查詢序列利用預設值來過濾低複雜性區域(描述在阿次朱耳(Altschul)，S.F.，馬

登(Madden), T.L., 史卡弗(Schäffer), A.A., 張(Zhang), J., 張, Z., 米勒(Miller), W.&李普門(Lipman), D.J.(1997)“有間隔的(Gapped)BLAST及PSI-BLAST：新一代蛋白質資料庫搜尋程式。”核酸(Nucleic Acids) Res. 25：3389-3402，其全文以參考之方式併於本文))。在核酸程度上，OrfA不與任何熟知的核苷酸序列明顯地同源。在胺基酸程度上，與ORFA最大程度同源的序列有：念珠藻7120異型細胞醣脂類合成酶(編號NC_003272)，其與ORFA有42%相同超過1001個胺基酸殘基；及海洋莫瑞特菌(海洋弧菌)ORF8(編號AB025342)，其與ORFA有40%相同超過993個胺基酸殘基。

在OrfA中的第一區段為KS區段，於本文亦指為ORFA-KS。此區段包含於跨越在序列辨識編號：1(OrfA)之約位置1至40間的起始點至在約序列辨識編號：1之位置1428及1500間的終點之核苷酸序列中。包含編碼ORFA-KS區段的序列之核苷酸序列於本文中表為序列辨識編號：7(序列辨識編號：1之位置1-1500)。包含KS區段的胺基酸序列跨越在序列辨識編號：2(ORFA)之約位置1及14間的起始點至在序列辨識編號：2之約位置476及500間的終點。包含該ORFA-KS區段的胺基酸序列於本文中表為序列辨識編號：8(序列辨識編號：2之位置1-500)。應注意的是ORFA-KS區段包括活性位置母題(motif)：DXAC*(^{*}醯基黏結位置C₂₁₅)。

根據本發明，具有3-酮基醯基-ACP合成酶(KS)生物活性(功能)的區段或蛋白質之特徵為酵素，其可進行FAS(及

PKS)延長反應循環的起始步驟。指定用於延長的醯基會利用硫酯鍵結在酵素的活性位置處連結至半胱胺酸殘基。在多步驟反應中，醯基-酵素與丙二酸單醯基-ACP進行縮合以形成-酮基醯基-ACP、CO₂及游離態酵素。KS在延長循環中扮演一關鍵角色，且在許多系統中已顯示出具有比反應循環的其它酵素大的基質特異性。例如，大腸桿菌具有三種可區別的KS酵素-每個其自身在生物的生理學上具有特別的角色(馬格紐松(Magnuson)等人，*Microbiol. Rev.* 57, 522(1993))。PUFA-PKS系統的二個KS區段在PUFA生物合成反應序列中具有可區別的角色。

KS已充分地標出具有酵素等級的特徵。許多已證實的KS基因之序列已熟知，該活性位置母題已經鑑別且數種結晶結構已測定。蛋白質(或蛋白質的區段)可容易地藉由與熟知的KS序列同源而鑑別為屬於KS酵素家族。

在OrfA中的第二區段為MAT區段，於本文中亦指為ORFA-MAT。此區段包含於跨越在序列辨識編號：1(OrfA)之約位置1723及1798間的起始點至序列辨識編號：1之約位置2805及3000間的終點之核苷酸序列中。包含編碼ORFA-MAT區段之序列的核苷酸序列於本文表示為序列辨識編號：9(序列辨識編號：1之位置1723-3000)。包括MAT區段之胺基酸序列跨越在序列辨識編號：2(ORFA)之約位置575及600間的起始點至序列辨識編號：2之約位置935及1000間的終點。包含該ORFA-MAT區段的胺基酸序列於本文中表示為序列辨識編號：10(序列辨識編號：2之位置

575-1000)。需注意的是該ORFA-MAT區段包括一活性位置母題：GHS**XG*(*醯基黏結位置S₇₀₆)，於本文表示為序列辨識編號：11。

根據本發明，具有丙二酸單醯基-CoA：ACP轉醯酶(MAT)生物活性(功能)之區段或蛋白質的特徵為其可將丙二酸單醯基部分從丙二酸單醯基-CoA轉移至ACP。除了活性位置母題(*GxSxG*)外，這些酵素在關鍵的位置處具有延長的母題®及Q胺基酸，此可鑑別其為MAT酵素(比較至裂壺菌屬OrfB的AT區段)。在某些PKS系統中(但是非為PUFA PKS區段)，MAT區段較佳地載有甲基-或乙基-丙二酸酯而至ACP基族(從相符合的CoA酯來)，因此會將分枝引進該線性碳鏈。MAT區段可由熟知的MAT序列其同源物及其延長的母題結構識別。

OrfA的區段3-11為九個銜接ACP區段，於本文中亦指為ORFA-ACP(在該序列中的第一區段為ORFA-ACP1、第二區段為ORFA-ACP2、第三區段為ORFA-ACP3等等)。第一ACP區段(ORFA-ACP1)包括於跨越從序列辨識編號：1(OrfA)之約位置3343至約位置3600的核苷酸序列中。包含編碼ORFA-ACP1區段之序列的核苷酸序列於本文表示為序列辨識編號：12(序列辨識編號：1之位置3343-3600)。包含第一ACP區段的胺基酸序列跨越從序列辨識編號：2之約位置1115至約位置1200。包含ORFA-ACP1區段之胺基酸序列於本文表示為序列辨識編號：13(序列辨識編號：2之位置1115-1200)。應注意的是ORFA-ACP1區段包括一活性位

置母題：LGIDS*(*泛醯巰基乙胺黏結母題S₁₁₅₇)，於本文由
序列辨識編號：14表示。

全部九個ACP區段的核苷酸及胺基酸序列已高度地保存，因此，每個區段的序列於本文中不由各別的序列識別
5 符(identifier)表示。但是，根據於本文中揭示的訊息，熟知此技藝之人士可容易地決定包含其它八個ACP區段每個的序列(參見下列討論)。

全部九個ACP區段一起跨越從序列辨識編號：1之約位置3283至約位置6288的OrfA區段，其與從序列辨識編號：2
10 之約1095至約2096的胺基酸位置相符合。含全部九個區段的全部ACP區段之核苷酸序列於本文表示為序列辨識編號：16。由序列辨識編號：16表示的區段包括在各別的ACP區段間之連結子(linker)部分。九個區段的重覆區間為序列辨識編號：16的約每330個核苷酸(在毗連的活性位置絲胺
15 酸間所測量之胺基酸的實際數目範圍從104至116個胺基酸)。九個ACP區段每個包括一泛醯巰基乙胺黏結母題LGIDS*(於本文中由序列辨識編號：14表示)，其中S*為泛醯巰基乙胺黏結位置絲胺酸(S)。該泛醯巰基乙胺黏結位置絲胺酸(S)位於接近每個ACP區段序列的中心。在ACP區段
20 區域的每個末端處及之間，每個ACP區段為一高度富含脯胺酸(P)及丙胺酸(A)之區域，其相信為一連結子區域。例如，在ACP區段1及2間的序列為：APAPVKAAAPAAPVASAPAPA，於本文中表示為序列辨識編號：15。九個ACP區段每個之活性位置絲胺酸殘基的場

所(即，泛醯巯基乙胺黏結位置)(與序列辨識編號：2的胺基
酸序列有關)如下：ACP1=S₁₁₅₇；ACP2=S₁₂₆₆；ACP3=S₁₃₇₇；
ACP4=S₁₄₈₈；ACP5=S₁₆₀₄；ACP6=S₁₇₁₅；ACP7=S₁₈₁₉；
ACP8=S₁₉₃₀；及ACP9=S₂₀₃₄。所提供的ACP區段之平均尺寸
5 約85個胺基酸(排除連結子)，及約110個胺基酸(包括連結
子)，且該活性位置絲胺酸約在該區段中心，熟知技藝之人
士可容易地測定九個ACP區段每個在OrfA中的位置。

根據本發明，具有醯基載體蛋白質(ACP)生物活性(功
能)的區段或蛋白質之特徵為小的多胜肽(典型地，80至100
10 個胺基酸長)，其功能為增長脂肪醯基鏈用之載體，其經由
硫酯連結至蛋白質之共價鍵結的共因子。它們在較大的蛋
白質中發生成為分離的單元或區段。ACPs可藉由將CoA的
磷酸泛醯巯基乙胺基部分轉移至高度保存的ACP絲胺酸殘
基，來從不活化的apo-形式轉換成具官能性的holo-形式。
15 醯基可藉由硫酯連結在磷酸泛醯巯基乙胺基部分的游離態
界標處附著至ACP。ACPs可利用放射性泛醯巯基乙胺標定
及利用熟知的ACPs序列同源物來鑑別。上述所提及的母題
(LGIDS*)之變異體的存在亦為ACP的識別標誌。

在OrfA中的區段12為一KR區段，於本文中亦指為
20 ORFA-KR。此區段包於在跨越序列辨識編號：1之約位置
6598的起始點至序列辨識編號：1之約位置8730的終點之核
苷酸序列中。包含編碼ORFA-KR區段的序列之核苷酸序列
於本文中表為序列辨識編號：17(序列辨識編號：1之位
置6598-8730)。包含KR區段的胺基酸序列跨越序列辨識編

號：2(ORFA)之約位置2200的起始點至序列辨識編號：2之約位置2910的終點。包含ORFA-KR區段之胺基酸序列於本文中
 5 表示為序列辨識編號：18(序列辨識編號：2之位置2200-2910)。在此KR區段中為一與短鏈醛-脫氫酶類(KR為此家族成員)同源的核心區域。此核心區域跨越從序列辨識
 10 編號：1之約位置7198至約位置7500，而與序列辨識編號：2之胺基酸位置2400-2500相符合。

根據本發明，具有酮還原酶活性的區段或蛋白質(亦指
 15 為3-丙脂醯基-ACP還原酶(KR)生物活性(功能))之特徵為其可催化ACP的3-酮基醯基形式之具吡啶-核苷酸-依賴性的還原反應。其為在*de novo*脂肪酸生物合成延長循環中的第一還原步驟，且反應經常在聚乙醯生物合成中進行。明顯的序列類似性可以烯醯ACP-還原酶類(ER)的一個家族、FAS的其它還原酶(但不是存在於PUFA PKS系統中的ER家
 20 族)及短鏈醇脫氫酶家族來觀察。在上指出之PUFA PKS區域的Pfam分析顯露出與在核心區段中的短鏈醇脫氫酶家族同源。相同區域的BLAST分析顯露出在核心區域中相配至熟知的KR酵素，且同源物的延長區域相配至從其它已特徵化的PUFA PKS系統來之區段。

20 開放性讀碼區B(OrfB)：

OrfB的完整核苷酸序列於本文中表示為序列辨識編號：3。序列辨識編號：3之核苷酸1311-6177與在美國申請序號09/231,899中表示為序列辨識編號：71的核苷酸1-4867序列相符合(在美國申請序號09/231,899中的cDNA序列在

終止密碼子外包含約345個額外的核苷酸，包括polyA尾部)。因此，序列辨識編號：1之核苷酸1-1310顯示出不在美國申請序號09/231,899中揭示之額外的序列。序列辨識編號：3的此新穎區域包含大部分由OrfB譯出的KS區段。

5 OrfB為一6177核苷酸序列(不包括終止密碼子)，其編碼2059個胺基酸序列，於本文中表示為序列辨識編號：4。在OrfB中有四個區段：(a)一個 β -酮基醯基-ACP合成酶(KS)區段；(b)一個鏈長因子(CLF)區段；(c)一個醯基轉位酶(AT)區段；及(d)一個烯醯ACP-還原酶(ER)區段。

10 OrfB的核苷酸序列已保存在基因銀行編號AF378328(胺基酸序列編號AAK728880)中。OrfB可以上所述的標準BLAST搜尋來與熟知的序列比較。在核酸程度上，OrfB並不與任何熟知的核苷酸序列明顯地同源。在胺基酸程度上，與ORFB具有最大的同源程度之序列有：志瓦
15 菌sp.假想蛋白質(編號U73935)，其與ORFB有53%的相同且超過458個胺基酸殘基；海洋莫瑞特菌(海洋弧菌)ORF11(編號AB025342)，其與ORFB有53%相同而超過460個胺基酸殘基；深淵發光菌歐美加(omega)-3多不飽和脂肪酸合成酶PfaD(編號AF409100)，其與ORFB有52%相同且超過457個
20 胺基酸殘基；及念珠藻7120假設蛋白質(編號NC_003272)，其與ORFB有53%的相同且超過430個胺基酸殘基。

在OrfB中的第一區段為KS區段，於本文中亦指為ORFB-KS。此區段包含於跨越在序列辨識編號：3(OrfB)之約位置1及43間的起始點至在序列辨識編號：3之約位置

1332及1350間的終點之核苷酸序列中。包含編碼ORFB-KS
區段的序列之核苷酸序列於本文中表示為序列辨識編號：
19(序列辨識編號：3之位置1-1350)。包含KS區段的胺基酸
序列跨越在序列辨識編號：4(ORFB)之約位置1及15間的起
5 始點至在序列辨識編號：4之約位置444及450間的終點。包
含ORFB-KS區段的胺基酸序列於本文中表示為序列辨識編
號：20(序列辨識編號：4之位置1-450)。應注意的是該
ORFB-KS區段包括一活性位置母題：DXAC*(*醯基黏結位
置C₁₉₆)。KS生物活性及鑑別具有此活性的蛋白質或區段之
10 方法已於上述描述。

在OrfB中的第二區段為CLF區段，於本文中亦指為
ORFB-CLF。此區段包含於跨越在序列辨識編號：3(OrfB)
之約位置1378及1402間的起始點至在序列辨識編號：3之約
位置2682及2700間的終點之核苷酸序列中。包含編碼
15 ORFB-CLF區段之序列的核苷酸序列於本文中表示為序列
辨識編號：21(序列辨識編號：3之位置1378-2700)。包含CLF
區段的胺基酸序列跨越在序列辨識編號：4(ORFB)之約位置
460及468間的起始點至在序列辨識編號：4之約位置894及
900間的終點。包含ORFB-CLF區段的胺基酸序列於本文中
20 表示為序列辨識編號：22(序列辨識編號：4之位置
460-900)。應注意的是該ORFB-CLF區段包含一沒有醯基-
黏結的半胱胺酸之KS活性位置母題。

根據本發明，基於下列原理將一區段或蛋白質指為鏈
長因子(CLF)。該CLF初始地描述為型式II(分解的酵素)PKS

系統的特徵，且假設在決定延長最後產物的循環數目(因此鏈長度)上佔有一席之地。CLF胺基酸序列顯示出與KS區段同源(且認為與KS蛋白質形成異種二聚體)，但是它們缺乏活性位置半胱胺酸。CLF在PKS系統中的角色現在有爭議。

- 5 新證據(C.畢山(Bisang)等人，自然**401**，502(1999))建議為在引導(提供欲延長的起始醃基)該PKS系統中的角色。在此角色中，CLF區段認為會脫羧基丙二酸酯(如為丙二酸單醃基-ACP)，因此形成可轉移至KS活性位置的醋酸酯基團。此醋酸酯因此作用為可進行起始延長(縮合)反應的‘引導’
- 10 分子。型式II CLF的同源物在某些組合式PKS系統中已鑑別為‘負載’區段。具有CLF序列特徵的區段可在現在鑑別的全部PUFA PKS系統中發現，且在每個實例中發現為多區段蛋白質的部分。

- 在OrfB中的第三區段為AT區段，於本文中亦指為
- 15 ORFB-AT。此區段包含於跨越在序列辨識編號：3(OrfB)之約位置2701及3598間的起始點至在序列辨識編號：3之約位置3975及4200間的終點之核苷酸序列中。包含編碼ORFB-AT區段之序列的核苷酸序列於本文中表示為序列辨識編號：23(序列辨識編號：3之位置2701-4200)。包含AT
- 20 區段的胺基酸序列跨越在序列辨識編號：4(ORFB)之約位置901及1200間的起始點至在序列辨識編號：4之約位置1325及1400間的終點。包含ORFB-AT區段的胺基酸序列於本文中表示為序列辨識編號：24(序列辨識編號：4之位置901-1400)。應注意的是該ORFB-AT區段包含一GxS*xG(*

醯基黏結位置S₁₁₄₀)的活性位置母題，其為轉醯酶(AT)蛋白質的特徵。

“轉醯酶”或“AT”指為一般的酵素種類，其可進行一些可區別的醯基轉移反應。該裂壺菌屬區段已顯示出與存在於其它現在已檢驗且與某些特定功能已鑑別之轉醯酶(例如至丙二酸單醯基-CoA：ACP轉醯酶，MAT)具有非常弱的同源性之PUFA PKS系統全部中的區段具有好的同源性。雖然與MAT具弱的同源性，此AT區段不相信可作用為MAT，因為其不具有此類酵素之延長的母題結構特徵(參見MAT區段描述，上述)。對此公告目的來說，在PUFA PKS系統中的AT區段之功能包括(但是非為限制)：將從ORFA ACP區段來的脂肪醯基轉移成水(即硫酸酯酶-釋放該脂肪醯基成為游離態脂肪酸)，將脂肪醯基至轉移受器(諸如CoA)，轉移在不同ACP區段當中的醯基，或轉移該脂肪醯基至親油性受器分子(例如至溶血磷脂酸(lysophosphadic acid))。

在OrfB中的第四個區段為ER區段，於本文中亦指為ORFB-ER。此區段包含於跨越序列辨識編號：3(OrfB)之約位置4648的起始點至序列辨識編號：3之約位置6177的終點之核苷酸序列中。包含編碼ORFB-ER區段之序列的核苷酸序列於本文中表示為序列辨識編號：25(序列辨識編號：3之位置4648-6177)。包含ER區段的胺基酸序列跨越序列辨識編號：4(ORFB)之約位置1550的起始點至序列辨識編號：4之約位置2059的終點。包含ORFB-ER區段的胺基酸序列於本文中表示為序列辨識編號：26(序列辨識編號：4之位置

1550-2059)。

根據本發明，此區段具有烯醯還原酶(ER)生物活性。ER酵素可還原在脂肪醯基-ACP中的反式雙鍵(藉由DH活性引進)，而造成完全飽和那些碳。在PUFA-PKS中的ER區段
5 顯示出與新特徵化的ER酵素家族同源(喜斯等人，自然406,145(2000))。喜斯及岩石(Rock)藉由選殖從肺炎鏈球菌來之有興趣的基因、純化由此基因表現的蛋白質並顯示出其在活體外試驗中具有ER活性來鑑別此新種類的ER酵
10 素。OrfB的裂壺菌屬ER區段序列顯示出與肺炎鏈球菌(*S. pneumoniae*)ER蛋白質同源。現在檢驗的PUFA PKS系統全部包含至少一個與裂壺菌屬ER區段具有非常高的序列同源性之區段。裂壺菌屬PUFA PKS系統包括二個ER區段(一個在OrfB上及一個在OrfC上)。

開放性讀碼區C(OrfC)：

15 OrfC的完整核苷酸序列於本文中表示為序列辨識編號：5。序列辨識編號：5之核苷酸1-4509(即，全部的開放性讀碼區序列，不包括終止密碼子)與在美國申請序號09/231,899中表示為序列辨識編號：76之核苷酸145-4653序列相符合(在美國申請序號09/231,899中的cDNA序列包括
20 OrfC的起始密碼子之約144個核苷酸上游及超出終止密碼子約110個核苷酸，包括polyA尾部)。OrfC為一4509核苷酸序列(不包括終止密碼子)，其編碼1503胺基酸序列，於本文中
中表示為序列辨識編號：6。在OrfC中有三個區段：(a)二個類FabA的 β -羥基醯基-ACP脫水酶(DH)區段；及(b)一個烯醯

ACP-還原酶(ER)區段。

OrfC 核苷酸序列已由基因銀行保存為編號 AF378329(胺基酸序列編號AAK728881)。OrfC可以上述的標準BLAST搜尋與熟知的序列比較。在核酸程度上，OrfC
 5 不與任何熟知的核苷酸序列有明顯的同源性。在胺基酸程度上(Blastp)，與ORFC具有最大程度的同源序列為：海洋莫瑞特菌(海洋弧菌)ORF11(編號ABO25342)，其與OrfC有45%的相同且超過514個胺基酸殘基；志瓦菌sp.假設蛋白質8(編號U73935)，其與ORFC有49%相同且超過447個胺基酸
 10 殘基；念珠藻假設蛋白質(編號NC_003272)，其與ORFC有49%相同而超過430個胺基酸殘基；及志瓦菌sp.假設蛋白質7(編號U73935)，其與ORFC有37%相同且超過930個胺基酸殘基。

在OrfC中的第一區段為DH區段，於本文中亦指為
 15 ORFC-DH1。此為在OrfC的二個DH區段中之一個，因此標示為DH1。此區段包含於跨越在序列辨識編號：5(OrfC)之約位置1及778間的起始點至在序列辨識編號：5之約位置1233及1350間的終點之核苷酸序列中。包含編碼ORFC-DH1區段的序列之核苷酸序列於本文中表示為序列
 20 辨識編號：27(序列辨識編號：5之位置1-1350)。包含DH1區段的胺基酸序列跨越在序列辨識編號：6(ORFC)之約位置1及260間的起始點至在序列辨識編號：6之約位置411及450間的終點。包含ORFC-DH1區段的胺基酸序列於本文中表示為序列辨識編號：28(序列辨識編號：6之位置1-450)。

在PUFA PKS系統中二個DH區段(參見在下列的DH2)的特徵已在前述部分描述。此種類的酵素可從 β -酮基醯基-ACP移除HOH且在碳鏈中留下反式雙鍵。該PUFA PKS系統的DH區段顯示出與其FAS系統有關的細菌DH酵素同源(而非其它PKS系統的DH區段)。細菌的DH、類FabA的DH之次小組具有順-反異構酶活性(喜斯等人, *J. Biol. Chem.*, 271, 27795(1996))。其與類FabA的DH同源, 此指示出一個或二個DH區段可反應用來在PUFA PKS產物中嵌入順式雙鍵。

10 在OrfC中的第二區段為DH區段, 於本文中亦指為ORFC-DH2。此為在OrfC的二個DH區段中之第二個, 因此標示為DH2。此區段包含於跨越在序列辨識編號: 5(OrfC)之約位置1351及2437間的起始點至在序列辨識編號: 5之約位置2607及2850間的終點之核苷酸序列中。包含編碼
15 ORFC-DH2區段之序列的核苷酸序列於本文中表示為序列辨識編號: 29(序列辨識編號: 5之位置1351-2850)。包含DH2區段的胺基酸序列跨越在序列辨識編號: 6(ORFC)之約位置451及813間的起始點至在序列辨識編號: 6之約位置869及950間的終點。包含該ORFC-DH2區段的胺基酸序列於本文
20 中表示為序列辨識編號: 30(序列辨識編號: 6之位置451-950)。該DH生物活性已於上述描述。

在OrfC中的第三區段為ER區段, 於本文中亦指為ORFC-ER。此區段包含於跨越序列辨識編號: 5(OrfC)之約位置2998的起始點至序列辨識編號: 5之約位置4509的終點

之核苷酸序列中。包含編碼ORFC-ER區段的序列之核苷酸序列於本文中表示為序列辨識編號：31(序列辨識編號：5之位置2998-4509)。包含ER區段的胺基酸序列跨越序列辨識編號：6(ORFC)之約位置1000的起始點至序列辨識編號：6之約位置1502的終點。包含ORFC-ER區段的胺基酸序列於本文中表示為序列辨識編號：32(序列辨識編號：6之位置1000-1502)。該ER生物活性已於上述描述。

本發明的一個具體實施例係關於一種包含從非細菌PUFA PKS系統來的核酸序列之分離的核酸分子、其同源化合物、其片段及/或與任何此核酸序列互補的核酸序列。在一個觀點中，本發明係關於一種包含選自於由下列所組成之群的核酸序列之分離的核酸分子：(a)一核酸序列，其編碼選自於由下列所組成之群的胺基酸序列：序列辨識編號：2、序列辨識編號：4、序列辨識編號：6及其生物活性片段；(b)一核酸序列，其編碼選自於由下列所組成之群的胺基酸序列：序列辨識編號：8、序列辨識編號：10、序列辨識編號：13、序列辨識編號：18、序列辨識編號：20、序列辨識編號：22、序列辨識編號：24、序列辨識編號：26、序列辨識編號：28、序列辨識編號：30、序列辨識編號：32及其生物活性片段；(c)一核酸序列，其編碼與(a)之胺基酸序列的至少500個連續胺基酸有至少約60%相同之胺基酸序列，其中該胺基酸序列具有多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統的至少一個區段之生物活性；(d)一核酸序列，其編碼與(b)的胺基酸序列有至少約60%

相同之胺基酸序列，其中該胺基酸序列具有多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統的至少一個區段之生物活性；或(e)一核酸序列，其可與(a)、(b)、(c)或(d)之核酸序列完全互補。在進一步的具體實施例中，包含編碼活性位置區段或上述描述的數個PUFA PKS區段之其它官能性母題的序列之核酸序列意欲包含在本發明中。

根據本發明，具有PUFA PKS系統的至少一個區段之生物活性的胺基酸序列為一種胺基酸序列，其具有於本文中詳細描述的PUFA PKS系統之至少一個區段的生物活性，其例示有裂壺菌屬PUFA PKS系統。在裂壺菌屬PUFA PKS系統中，不同區段的生物活性已在上文中詳細地描述。因此，本發明之分離的核酸分子編碼任何PUFA PKS開放性讀碼區的轉譯產物、PUFA PKS區段、其生物活性片段、或任何天然發生的PUFA PKS開放性讀碼區或具有生物活性的區段之同源化合物。所提供的蛋白質或區段之同源化合物為一種具有與天然發生的參考胺基酸序列(即，參考蛋白質或區段)不同之胺基酸序列的蛋白質或多肽，其中至少一個或少數個(但是非限制於一個或少數個)胺基酸已刪除(例如，蛋白質的截斷形式，諸如胜肽或片段)、嵌入、轉化、取代及/或衍生(例如，利用糖化作用、磷酸化作用、乙醯化作用、肉豆蔻醯化作用、異戊烯化作用、棕櫚酸化作用、醯胺化作用及/或糖基磷脂醯基肌醇的加成作用)。PUFA PKS蛋白質或區段的較佳同源物在下列有詳細的描述。應注意的是同源物可包括合成製造的同源物、所提供的蛋白質或區段

之天然發生的對等變異物、或從其它非衍生出參考序列的生物之生物來的同源序列。

通常地，蛋白質或區段的生物活性或生物作用指為任何藉由可歸因於天然發生的蛋白質或區段形式之蛋白質或區段所具有或進行的功能，其可在活體內(即，在蛋白質的天然生理環境中)或活體外(即，在實驗室條件下)測量或觀察到。PUFA PKS系統及構成PUFA PKS系統的各別蛋白質/區段之生物活性已詳細地描述於本文別處。蛋白質或區段之改性(諸如在下列討論的同源化合物或模倣物)可產生具有與天然發生的蛋白質或區段之相同生物活性的蛋白質或區段，或產生生物活性比天然發生的蛋白質或區段減少或增加之蛋白質或區段。會造成蛋白質或區段的表現性減少或活性減少之改性可指為該蛋白質或區段之失活(完全或部分)、向下調節或減少作用。類似地，會造成蛋白質或區段的表現性增加或活性增加之改性可指為該蛋白質或區段的放大、過度產生、活化、提高、向上調節或增加作用。PUFA PKS系統的官能性區段為一個能進行生物功能(即，具有生物活性)的區段(即，可為部分蛋白質的區段)。

根據本發明，該分離的核酸分子為一種已從其天然環境移出(即，其已接受人類操控)的核酸分子，而該天然環境為發現該天然核酸分子的基因組或染色體。就其本身而論，“分離的”不必反映出該核酸分子已被純化的程度，而是顯示出該分子不包含全部的基因組或全部的染色體(在其中發現天然核酸分子)。分離的核酸分子可包含一基因。

包含一基因之分離的核酸分子不為包含此基因的染色體片段，而是包括與基因有關的譯碼區域及調節區域，但是在相同的染色體上並無天然地發現其它基因。分離的核酸分子亦可包括一由額外的核酸嚙合(flanked)(即，在序列的5'及/或3'末端)之特定的核酸序列(即，為異種序列)，其中該額外的核酸在天然中正常地不會與該特定的核酸序列嚙合。分離的核酸分子可包括DNA、RNA(例如，mRNA)、或DNA或RNA之衍生物(例如，cDNA)。雖然措辭“核酸分子”主要指為生理的核酸分子，及措辭“核酸序列”主要指為在核酸分子上的核苷酸序列，二措辭可交替地使用，特別是關於能譯出一蛋白質或一蛋白質區域之核酸分子或核酸序列。

較佳地，本發明之分離的核酸分子可使用重組DNA技術(例如，聚合酶連鎖反應(PCR)放大、選殖)或化學合成來製造。分離的核酸分子包括天然的核酸分子及其同源化合物，包括(但是非為限制)天然的對等基因變異物及已以此方法嵌入、刪除、取代及/或轉化之經改性的核酸分子之核苷酸，此改性可在如於本文所述的PUFA PKS系統生物活性上提供想要的效應。蛋白質同源化合物(例如，可由核酸同源化合物譯出的蛋白質)已在上述詳細地討論。

核酸分子同源化合物可由熟知此技藝之人士使用一些方法來製造(參見，例如，山姆布魯克(Sambrook)等人，分子選殖：實驗室手冊，冷泉哈伯實驗室出版社，1989)。例如，核酸分子可使用多種技術來改性，包括(但是非為限制)

傳統的致突變技術及重組DNA技術，諸如定點突變、核酸分子的化學處理以引發突變、核酸片段的限制酵素切割、核酸片段的連接、核酸序列的選擇區域之PCR放大及/或致突變、寡核苷酸混合物之合成及連接混合物團以“建立”核酸分子之混合物及其組合。核酸分子同源化合物可藉由該核酸所譯出的蛋白質及/或藉由與野生型基因雜交而譯出的蛋白質之功能來篩選，而選自於一經改性的核酸之混合物。

本發明之核酸分子的最小尺寸為足以形成可與於本發明中有用的核酸分子之互補序列形成安定的混合物(例如，在中度、高度或非常高度嚴苛的條件下)之探針或寡核苷酸引子的尺寸，或足以譯出具有根據本發明之PUFA PKS系統的至少一個區段之生物活性的胺基酸序列之尺寸。如此，編碼此一蛋白質的核酸分子尺寸可依核酸組成物、在核酸分子與互補序列間之同源或相同百分比、和雜交條件本身(例如，溫度、鹽濃度及甲醯胺濃度)而定。使用作為寡核苷酸引子或作為探針的核酸分子之最小尺寸長度典型地至少約12至約15個核苷酸(若該核酸分子富含GC)及至少約15至約18個鹼基長(若它們富含AT)。除了實際的限制外，本發明之核酸分子的最大尺寸並無限制，其中該核酸分子可包括一序列，其足以譯出PUFA PKS系統區段的生物活性片段、PUFA PKS系統的全部區段、在PUFA PKS系統的開放性讀碼區(Orf)中之數個區段、PUFA PKS系統的全部Orf，或多於一個的PUFA PKS系統之Orf。

在本發明的一個具體實施例中，分離的核酸分子包括或基本地由選自於下列所組成之群的核酸序列所組成：序列辨識編號：2、序列辨識編號：4、序列辨識編號：6、序列辨識編號：8、序列辨識編號：10、序列辨識編號：13、
5 序列辨識編號：18、序列辨識編號：20、序列辨識編號：22、序列辨識編號：24、序列辨識編號：26、序列辨識編號：28、序列辨識編號：30、序列辨識編號：32或其生物活性片段。在一個觀點中，該核酸序列選自於由下列所組成之群：序列辨識編號：1、序列辨識編號：3、序列辨識
10 編號：5、序列辨識編號：7、序列辨識編號：9、序列辨識編號：12、序列辨識編號：17、序列辨識編號：19、序列辨識編號：21、序列辨識編號：23、序列辨識編號：25、序列辨識編號：27、序列辨識編號：29，及序列辨識編號：31。在本發明的一個具體實施例中，上述描述的任何PUFA
15 PKS胺基酸序列(和此序列的同源物)可從至少一個(最高約20)與所提供的胺基酸序列之每個C-及/或N-終端嚙合的額外異種胺基酸製造。所產生的蛋白質或多胜肽可指為“實質上由”所提供的胺基酸序列“組成”。根據本發明，該異種胺基酸為一種胺基酸序列，其經發現不天然地(即，不在天然
20 中(活體內)發現)與所提供的胺基酸序列嚙合；或已嚙合該天然發生的編碼所提供的胺基酸序列(如其在基因中發生般)之核酸序列的核苷酸，若此類核苷酸在天然發生的序列中由欲衍生出所提供的胺基酸序列之生物使用標準密碼子來轉譯出時，其將不由此核苷酸譯出。類似地，當使用於

本文中提及的核酸序列時，措辭“實質上由...組成”指為一種編碼所提供的胺基酸序列之核酸序列，該胺基酸序列可在能譯出所提供的胺基酸序列之核酸序列的每個5'及/或3'末端處，與至少一個(最高如約60)額外的異種核苷酸啣合。

- 5 該些異種核苷酸經發現不天然地(即，不在天然中(活體內)發現)與編碼所提供的胺基酸序列(如其在天然基因中發生般)之核酸序列啣合。

本發明亦包括一種分離的核酸分子，其包含一種編碼具有PUFA PKS系統的至少一個區段之生物活性的胺基酸序列之核酸序列。在一個觀點中，此核酸序列編碼裂壺菌屬PUFA PKS ORFs或區段的任何同源化合物，包括：序列辨識編號：2、序列辨識編號：4、序列辨識編號：6、序列辨識編號：8、序列辨識編號：10、序列辨識編號：13、序列辨識編號：18、序列辨識編號：20、序列辨識編號：22、
15 序列辨識編號：24、序列辨識編號：26、序列辨識編號：28、序列辨識編號：30或序列辨識編號：32，其中該同源化合物具有PUFA PKS系統的至少一個區段之生物活性，如先前於本文中所描述般。

在本發明的一個觀點中，由本發明包含的裂壺菌屬
20 PUFA PKS蛋白質或區段之同源化合物包括一胺基酸序列，其與選自於下列的胺基酸序列：序列辨識編號：2、序列辨識編號：4及序列辨識編號：6之至少500個連續胺基酸有至少約60%相同；其中該胺基酸序列具有PUFA PKS系統的至少一個區段之生物活性。在進一步觀點中，該同源化

合物的胺基酸序列與序列辨識編號：2、序列辨識編號：4
及序列辨識編號：6任何一種(或與序列辨識編號：6的最大
長度)之至少約600個連續胺基酸有至少約60%相同，更佳地
至少約700個連續胺基酸，更佳地至少約800個連續胺基
5 酸，更佳地至少約900個連續胺基酸，更佳地至少約1000個
連續胺基酸，更佳地至少約1100個連續胺基酸，更佳地至
少約1200個連續胺基酸，更佳地至少約1300個連續胺基
酸、更佳地至少約1400個連續胺基酸及更佳地至少約1500
個連續胺基酸。在進一步觀點中，該同源化合物的胺基酸
10 序列與序列辨識編號：2或序列辨識編號：4的任何一種(或
序列辨識編號：4的最大長度)之至少約1600個連續胺基酸
有至少約60%相同，更佳地至少約1700個連續胺基酸，更
佳地至少約1800個連續胺基酸，更佳地至少約1900個連續
胺基酸及更佳地至少約2000個連續胺基酸。在進一步觀點
15 中，該同源化合物的胺基酸序列與序列辨識編號：2的至少
約2100個連續胺基酸有至少約60%相同，更佳地至少約
2200個連續胺基酸，更佳地至少約2300個連續胺基酸，更
佳地至少約2400個連續胺基酸，更佳地至少約2500個連續
胺基酸，更佳地至少約2600個連續胺基酸，更佳地至少約
20 2700個連續胺基酸，更佳地至少約2800個連續胺基酸，甚
至更佳地為最大的長度。

在另一個觀點中，包含於本發明中之裂壺菌屬PUFA
PKS蛋白質或區段的同源化合物包括一與選自於下列的胺
基酸序列：序列辨識編號：2、序列辨識編號：4，或序列

辨識編號：6有至少約65%相同之胺基酸序列，更佳地至少約70%相同，更佳地至少約75%相同，更佳地至少約80%相同，更佳地至少85%相同，更佳地至少約90%相同，更佳地至少約95%相同，更佳地至少約96%相同，更佳地至少約97%相同，更佳地至少約98%相同，更佳地至少約99%相同，且超過在上段描述的任何連續胺基酸之長度，其中該胺基酸序列具有PUFA PKS系統的至少一個區段之生物活性。

在本發明的一個觀點中，包含於本發明中的裂壺菌屬PUFA PKS蛋白質或區段之同源化合物包括一與選自於下列的胺基酸序列有至少約60%相同之胺基酸序列：序列辨識編號：8、序列辨識編號：10、序列辨識編號：13、序列辨識編號：18、序列辨識編號：20、序列辨識編號：22、序列辨識編號：24、序列辨識編號：26、序列辨識編號：28、序列辨識編號：30或序列辨識編號：32，其中該胺基酸序列具有PUFA PKS系統的至少一個區段之生物活性。在進一步觀點中，該同源化合物的胺基酸序列與選自於下列的胺基酸序列有至少約65%相同：序列辨識編號：8、序列辨識編號：10、序列辨識編號：13、序列辨識編號：18、序列辨識編號：20、序列辨識編號：22、序列辨識編號：24、序列辨識編號：26、序列辨識編號：28、序列辨識編號：30、序列辨識編號：32，更佳地至少約70%相同，更佳地至少約75%相同，更佳地至少約80%相同，更佳地至少約85%相同，更佳地至少約90%相同，更佳地至少約95%相

同，更佳地至少約96%相同，更佳地至少約97%相同，更佳地至少約98%相同及更佳地至少約99%相同於，其中該胺基酸序列具有PUFA PKS系統的至少一個區段之生物活性。

根據本發明，與於本文中描述的核酸或胺基酸序列有關之名稱“連續的(contiguous)”或“連貫的(consecutive)”意謂著以未斷裂的序列連接。例如，對包含第二序列的30個連續(或連貫的)胺基酸之第一序列來說，此意謂著該第一序列包含一完整的30個胺基酸殘基序列，其與在第二序列中的30個胺基酸殘基之完整序列乃100%相同。類似地，對具有與第二序列“100%相同”之第一序列來說，此意謂著第一序列精確地與第二序列相符合且在核苷酸或胺基酸間並無間隔。

如於本文中所使用(除非另有指定)，所提及的相同百分比(%)指為同源體的評估，其可使用下列方法來進行：

(1)BLAST 2.0基本BLAST同源搜尋，全部使用標準預設參數在全部6個開放性讀碼區中以blastp做胺基酸搜尋、以blastn做核酸搜尋及blastX做核酸搜尋及已轉譯的胺基酸搜尋，其中該查詢序列可利用預設值過濾出低複雜性區域(描述在阿次朱耳，S.F.，瑪登，T.L.，史卡弗，A.A.，張，J.，張，Z.，米勒，W.&李普門，D.J.(1997)“有間隔的BLAST及PSI-BLAST：新一代蛋白質資料庫搜尋程式。” *Nucleic Acids Res.* 25：3389-3402，其全文以參考之方式併於本文)；(2)BLAST 2排比(alignment)(使用下列描述的參數)；(3)及/或具有標準預設參數的PSI-BLAST(位置特定的重複

(Position-Specific Iterated)BLAST)。應注意的是由於在 BLAST 2.0基本BLAST及BLAST 2間的標準參數上之某些差異，二種特定的序列其可使用BLAST 2程式認定為具有明顯的同源性，然而在BLAST 2.0基本BLAST中所進行的搜尋(使用序列之一作為查詢序列)可能在上部配對中不與第二序列同源。此外，PSI-BLAST提供一種自動的、容易使用的“分佈(profile)”搜尋版本，其為一種敏感的搜尋序列同源化合物的方法。該程式首先進行具間隔的BLAST資料庫搜尋。PSI-BLAST程式使用任何有意義的排比資訊回饋來構造出一位置特定的計分矩陣，而取代了下一輪資料庫搜尋需用之查詢序列。因此，可了解的是該相同百分比可使用這些程式的任何一種來決定。

二種特定的序列可使用BLAST 2序列來彼此排比，如描述在塔土梭伐(Tatusova)及瑪登，(1999)，“BLAST 2序列—一種新穎的、比較蛋白質及核苷酸序列用之工具”，*FEMS Microbiol Lett.* 174：247-250，其全文以參考之方式併於本文。BLAST 2序列排比使用BLAST 2.0演算法以blastp或blastn來進行，以在二種序列間進行具間隔的BLAST搜尋(BLAST 2.0)，以允許在產生的排比中引進間隔(刪除及嵌入)。於此，為了清楚的目的，可使用如下的標準預設參數來進行BLAST 2序列排比。

對blastn來說，使用0 BLOSUM62矩陣：

配對的獎勵(reward)=1

錯誤配對的罰分(penalty)=-2

開放間距(open gap)(5)及延伸間距(extension gap)(2)罰分

間距x_下降(dropoff)(50)預期(expect)(10)字形大小(word size)(11)過濾(filter)(開)

5 對blastp來說，使用0 BLOSUM62矩陣：

開放間距(11)及延伸間距(1)罰分

間距x_下降(50)預期(10)字形大小(3)過濾(開)。

在本發明的另一個具體實施例中，具有本發明之PUFA
PKS系統的至少一個區段之生物活性的胺基酸序列包括一
10 胺基酸序列，其足夠地類似於天然發生的PUFA PKS蛋白質
或多胜肽，而編碼該胺基酸序列的核酸序列能在中度、高
度或非常高度嚴苛的條件(描述在下列)下雜交至(即，與)編
碼天然發生的PUFA PKS蛋白質或多胜肽之核酸分子(即，
至編碼天然發生的PUFA PKS蛋白質或多胜肽之核酸股的
15 補體)。較佳地，具有本發明之PUFA PKS系統的至少一個
區段之生物活性的胺基酸序列可由一核酸序列譯出，其能
在中度、高度或非常高度嚴苛的條件下雜交至編碼包含由
任何下列胺基酸序列表示之蛋白質的核酸序列補體：序列
辨識編號：2、序列辨識編號：4、序列辨識編號：6、序列
20 辨識編號：8、序列辨識編號：10、序列辨識編號：13、序
列辨識編號：18、序列辨識編號：20、序列辨識編號：22、
序列辨識編號：24、序列辨識編號：26、序列辨識編號：
28、序列辨識編號：30或序列辨識編號：32。推衍出互補
序列的方法已由熟知此技藝之人士所熟知。應注意的是因

為胺基酸定序及核酸定序技藝並非完全無誤差，於本文中
所顯現的序列充其量僅代表本發明之明顯的PUFA PKS區
段及蛋白質序列。

如於本文中所使用，雜交條件指為一種標準雜交條
5 件，在此之下核酸分子可使用來鑑別類似的核酸分子。此
標準條件例如揭示在山姆布魯克等人，分子選殖：實驗室
手冊，冷泉哈伯實驗室出版社，1989，山姆布魯克等人，
同前所述，其全文以參考之方式併入本文(特別地參見第
9.31-9.62頁)。此外，計算適當的雜交及清洗條件以獲得核
10 苷酸的錯誤配對之變異程度在准許範圍的雜交之式子則揭
示在，例如，門寇斯(Meinkoth)等人，1984，Anal. Biochem.
138，267-284；門寇斯等人，同前所述，其全文以參考之
方式併入本文。

更特別地是，中度嚴苛的雜交及洗滌條件(如於本文指
15 出)指為可分離出具有與在雜交反應中使用來探測的核酸
分子之核酸序列至少約70%相同的核酸分子之條件(即，該
條件准許約30%或較少的核苷酸錯誤配對)。高度嚴苛的雜
交及洗滌條件(如於本文指出)指為可分離出具有與在雜交
反應中使用來探測的核酸分子之核酸序列至少約80%相同
20 的核酸分子之條件(即，該條件可准許約20%或較少的核
苷酸錯誤配對)。非常高度嚴苛的雜交及洗滌條件(如於本文
指出)指為可分離出具有與在雜交反應中使用來探測的核酸
分子之核酸序列至少約90%相同的核酸分子之條件(即，該
條件准許約10%或較少的核苷酸錯誤配對)。如上述所討

論，熟知技藝之人士可使用在門寇斯等人(同前所述)中的式
子來計算適當的雜交及清洗條件，以獲得這些特別的核苷
酸錯誤配對程度。這些條件將依欲形成的DNA：RNA或
DNA：DNA混雜體而變化。經計算的DNA：DNA混雜體之
5 5 融化溫度比DNA：RNA混雜體少10EC。在特別的具體實施
例中，DNA：DNA混雜體的嚴苛雜交條件包括：在6X
SSC(0.9M Na⁺)的離子強度下，在約20EC至約35EC間之溫
度(較不嚴苛)中雜交，更佳地，在約28EC至約40EC間(更嚴
苛)，甚至更佳地，在約35EC至約45EC間(甚至更嚴苛)，且
10 具適當的清洗條件。在特別的具體實施例中，DNA：RNA
混雜體的嚴苛雜交條件包括在6X SSC(0.9M Na⁺)的離子強
度下，在約30EC至約45EC間之溫度中雜交，更佳地，在約
38EC至約50EC間，甚至更佳地在約45EC至約55EC間，以
類似嚴苛的清洗條件。這些值根據大於約100個核苷酸之分
15 子、0%甲醯胺及約40%的G+C含量之融化溫度來計算。此
外，T_m可如山姆布魯克等人(前述，第9.31至9.62頁)所提而
經驗地計算。通常地，清洗條件應該儘可能地嚴苛，且應
該適當地選擇雜交條件。例如，雜交條件可包括鹽與溫度
20 條件(其低於特別的混雜體之經計算的T_m約20-25EC)之組
合；及清洗條件典型地包括鹽與溫度條件(其低於特別的混
雜體之經計算的T_m約12-20EC)之組合。合適使用於DNA：
DNA混雜體的雜交條件實例包括在6X SSC(50%甲醯胺)
中，在約42EC下雜交2-24小時；接著為洗滌步驟，其包括
在室溫下於約2X SSC中洗滌一次或多次；接著為在較高的

溫度及較低的離子強度下額外的洗滌(例如，在約0.1X-0.5X SSC中，於約37EC下清洗至少一次，接著在約0.1X-0.5X SSC中，於約68EC下清洗至少一次)。

5 本發明的另一個具體實施例包括一種重組型核酸分子，其包含一重組載體及一包含編碼具有如於本文中描述之PUFA PKS系統的至少一個區段之生物活性的胺基酸序列之核酸序列的核酸分子。此些核酸序列已在上文中詳細地描述。根據本發明，重組載體為一種經設計(即，製造)的核酸分子，其可使用來作為操控選擇一核酸序列及將此

10 一核酸序列引進宿主細胞之工具。該重組載體因此合適使用於選殖、定序及/或操作核酸序列之選擇，諸如可藉由將所選擇的核酸序列表現及/或傳遞進入宿主細胞以形成一重組細胞。此載體典型地包括異性核酸序列，其為一種經發現不天然地毗連至欲選殖或傳遞的核酸序列之核酸序

15 列，雖然該載體亦可包括天然發現可毗連至本發明的核酸分子或對本發明之核酸分子的表現有用(在下列詳細地討論)之控制核酸序列(例如，起動子、未轉譯的區段)。該載體可為RNA或DNA(原核的或真核的)，典型地為一質體。該載體可維持為一種染色體外的元素(例如，質體)或可併入

20 重組生物(例如，微生物或植物)的染色體。全部的載體可適當地保持在宿主細胞中，或在某些條件下，該質體DNA可刪除，而留下本發明之核酸分子。併入的核酸分子可由染色體的起動子控制、由本身或質體的起動子控制或由數個組合的起動子控制。核酸分子的單一或多重複製可併入染

色體。本發明之重組載體可包括至少一種選擇標誌。

在一個具體實施例中，在本發明的重組型核酸分子中所使用之重組載體為一種表現載體(expression vector)。如於本文中所使用，措辭“表現載體”習慣上指為一種載體，其合適用來製造一種編碼的產物(例如，有興趣的蛋白質)。在此具體實施例中，該編碼欲製造的產物(例如，PUFA PKS 區段)之核酸序列可嵌入重組載體以製造一重組型核酸分子。該編碼欲製造的蛋白質之核酸序列可以一方法嵌入載體，而將該核酸序列操作地連結至載體中的控制序列，而使其能夠在重組宿主細胞中轉錄及轉譯該核酸序列。

在另一個具體實施例中，在本發明的重組型核酸分子中所使用的重組載體為一種目標載體(targeting vector)。如於本文中所使用，措辭“目標載體”習慣上指為一種載體，其可使用來將特別的核酸分子傳遞進入重組宿主細胞，其中該核酸分子可使用來在宿主細胞或微生物中刪除或失活一內生基因(即，使用於目標基因的分裂或關閉(knock-out)技藝)。此載體亦在技藝中熟知為“關閉”載體。在此具體實施例的一個觀點中，部分載體(但是更典型地，核酸分子嵌入載體(即，嵌入))具有一核酸序列，其與在宿主細胞中的目標基因之核酸序列(即，其目標為刪除或失活的基因)同源。將該載體嵌入的核酸序列設計成黏結至目標基因，如此該目標基因及該嵌入可進行同源重組，藉此以刪除、失活或減弱內生目標基因(即，至少部分的內生目標基因經突變或刪除)。

典型地，該重組型核酸分子包含本發明的至少一種核酸分子而操作地連結至一個或多個轉錄控制序列。如於本文中所以使用，措辭“重組分子”或“重組型核酸分子”主要指為一種操作地連結至一轉錄控制序列的核酸分子或核酸序列，但是其可與措辭“核酸分子”交替使用，當此核酸分子為如本文中所以討論之重組分子時。根據本發明，措辭“操作地連結”指為將核酸分子以一方法連結至一轉錄控制序列，如此當該分子轉染 (transfect)(即，轉染、轉導 (transduced)、轉染、轉接 (conjugated) 或導入 (conducted)) 進入宿主細胞時能夠表現。轉錄控制序列為一種控制轉錄的起始、延長或終止的序列。特別重要的轉錄控制序列為可控制轉錄起始者，諸如起動子、增強子、操作子及抑制子序列。合適的轉錄控制序列包括任何可在已引進重組型核酸分子的宿主細胞或生物中作用之轉錄控制序列。

本發明之重組型核酸分子亦可包括額外的控制序列，諸如轉譯控制序列、複製起始源及其它可與重組細胞相容的控制序列。在一個具體實施例中，本發明之重組分子(包括併入宿主細胞染色體的那些)亦包括分泌訊號(即，信號部分核酸序列)，以使表現的蛋白質能夠從製造蛋白質的細胞中分泌出。合適的信號部分包括與欲表現的蛋白質天然地相關之信號部分或任何能指導根據本發明之蛋白質分泌的異種信號部分。在另一個具體實施例中，本發明之重組分子包含一領導序列 (leader sequence) 以使表現的蛋白質能夠傳遞至及嵌入宿主細胞的薄膜。合適的領導序列包括與蛋

白質天然地相關的領導序列，或任何能指導蛋白質傳達及嵌入至細胞薄膜的異種領導序列。

本發明家已發現裂壺菌屬PUFA PKS Orfs A及B在基因組中緊密地連結且在Orfs間之區域已定序。Orfs在相反的方向上定向，且4244個鹼基對隔開該起始(ATG)密碼子(即它們如下安排：3'OrfA5'-4244bp-5'OrfB3')。4244bp基因間的區域之檢驗並不顯露出有任何明瞭的Orfs(在BlastX搜尋上並無發現明顯的配對)。Orfs A及B二者在裂壺菌屬中有高度表現(至少在產生油期間)，此意味著活性起動子元素埋在此基因間區段中。這些基因元素相信具有作為二方向起動子序列而用於轉殖基因的應用之用途。例如，在一個較佳的具體實施例中，可選殖此區域、在每個末端處放置任何有興趣的基因、且將此架構引進裂壺菌屬(或某些其它起動子已顯示出可作用的宿主)。已預期的是該控制元素在適當的條件下將提供二種引進基因經協調的、高程度的表現。包含裂壺菌屬PUFA PKS控制元素(例如，起動子)的控制區域之完整核苷酸序列於本文中表示為序列辨識編號：36。

在一個類似的方法中，OrfC在裂壺菌屬中在產生油期間會高度表現，而控制元素預計存在於其起始密碼子的區域上游中。OrfC的基因組DNA上游區域已選殖且定序，而於本文中表示為(序列辨識編號：37)。此序列包括OrfC起始密碼子的3886 nt立即上游(immediately upstream)。此區域的檢驗並不顯露出任何明顯的Orfs(即，在BlastX搜尋中無發現明顯的配對)。咸信包含在此區域中的控制元素在適當

的條件下將提供放置在其後面的基因具高程度之表現性。額外地，在適當的條件下，表現程度可在A-B 基因間的區域(序列辨識編號：36)之控制下以基因協調。

5 因此，在一個具體實施例中，在本發明中有用的重組型核酸分子(如揭示於本文)可包括一包含在序列辨識編號：36及/或序列辨識編號：37中的PUFA PKS控制區域。此控制區域可包括序列辨識編號：36及/或序列辨識編號：37的任何部分(片段)，其具有至少基本的PUFA PKS轉錄活性。

10 本發明的一個或多個重組分子可使用來製造一種本發明之譯出產物(例如，PUFA PKS區段、蛋白質或系統)。在一個具體實施例中，該譯出產物可在有效以製造該蛋白質之條件下利用如於本文中所描述的表現一核酸分子而製造。製造譯出蛋白質的較佳方法為藉由以一個或多個重組
15 分子來轉染宿主細胞而形成一重組細胞。合適於轉染的宿主細胞包括(但是非為限制)任何細菌、真菌(例如，酵母菌)、昆蟲、植物或可轉染的動物細胞。該宿主細胞可為非轉染的細胞或已經以至少一種其它重組型核酸分子轉染的細胞。

20 根據本發明，名詞“轉染(transfection)”習慣上指為任何可將外源核酸分子(即，重組型核酸分子)嵌入細胞的方法。名詞“轉形(transformation)”可與名詞“轉染”交替地使用，當此名詞使用來指為將核酸分子引進微生物細胞(諸如藻類、細菌及酵母菌)時。在微生物系統中，名詞“轉形”使用

來描述一種由於微生物取得外源核酸而產生之遺傳改變，而基本上與名詞“轉染”同義。但是，在動物細胞中，轉形具有第二種意義，其可指為在培養基中細胞生長性質的改變，例如在它們變成癌之後。因此，為了避免混亂，名詞“轉染”較佳地使用在關於將外源核酸引進動物細胞，及名詞“轉染”於本文中將通常地使用來包括動物細胞、植物細胞之轉染及微生物細胞的轉染，至該名詞涉及將外源核酸引進細胞的程度。因此，轉染技術包括(但是非為限制)轉染、粒子轟擊、電穿孔、顯微注射法、微脂粒感染(lipofection)、吸附、感染及原生質體熔融。

需由熟知此技藝之人士了解的是可使用重組DNA技藝來改善轉移的核酸分子之表現性的控制，其藉由操控例如在宿主細胞中的核酸分子之複製數目、那些核酸分子的轉錄效率、所產生的轉錄之轉譯效率及後轉譯改性的效率。額外地，起動子序列可基因地設計以改善表現程度，如與天生的起動子比較。控制核酸分子之表現性有用的重組技術包括(但是非為限制)將核酸分子整合進入一個或多個宿主細胞染色體、將載體穩定序列加入至質體、取代或改性轉錄控制訊號(例如，起動子、操作子、增強子)、取代或改性轉譯控制訊號(例如，核糖體黏結位置、閃-達卡諾(Shine-Dalgarn)序列)、將核酸分子改性至與宿主細胞使用的密碼子相符合、及刪除會使轉錄不穩定的序列。

通常上述討論的關於重組型核酸分子及宿主細胞的轉移意欲應用至任何於本文中討論的重組型核酸分子，包括

那些編碼具有從PUFA PKS來的至少一個區段之生物活性的任何胺基酸序列、那些編碼從其它PKS系統來的胺基酸序列、及那些編碼其它蛋白質或區段。

5 本發明亦關於使用新穎的方法來鑑別具有PUFA PKS
系統的微生物，其在結構、區段組織及/或功能上類似於裂
壺菌屬的PUFA PKS系統。在一個具體實施例中，該微生物
為一種非細菌的微生物，較佳地，由此方法鑑別的微生物
為一種真核微生物。此外，本發明係關於藉由此方法鑑別
10 的微生物；及關於將這些微生物及從這些微生物來的PUFA
PKS系統根據本發明使用在不同的PUFA PKS系統應用中
(例如，基因地改性的生物及製造生物活性分子之方法)。於
本文中描述及說明的獨特篩選方法能夠快速地鑑別新穎的
包含PUFA PKS系統類似於本發明之裂壺菌屬的PUFA PKS
系統之微生物菌株。於本文中申請人已使用此方法來發現
15 及揭示出一包含的PUFA PKS系統與在裂壺菌屬中發現的
類似之破囊壺菌屬微生物。此發現將詳細地描述在下列實
例2中。

具有PUFA PKS系統類似於在裂壺菌屬中發現的微生物，諸如由本發明家發現且描述在實例2的破囊壺菌屬微生物，其可容易利用下列方法來鑑別/分離/篩選，各別地使用或以這些方法的任何組合。

通常地，鑑別具有多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統的非細菌微生物之方法包括第一步驟(a)選擇一可製造至少一種PUFA的微生物；及第二步驟(b)鑑別從(a)

來的微生物，其具有在發酵培養液中於溶氧條件少於約飽和的5%下可製造增加的PUFAs之能力，如與在發酵培養液中於溶氧條件大於飽和之5%下由該微生物製造之PUFAs比較，更佳地飽和的10%，更佳地大於飽和的15%及更佳地大於飽和的20%。可製造至少一種PUFA且具有在溶氧條件少於約飽和之5%下製造增加的PUFAs之能力的微生物，經鑑別為包含PUFA PKS系統的候選生物。繼鑑別出該微生物為包含PUFA PKS系統的強候選生物後，該方法可包括額外的步驟(c)偵測在步驟(b)中鑑別的生物是否包含PUFA PKS系統。

在本發明的一個具體實施例中，步驟(b)可藉由在低氧/缺氧條件及有氧條件下培養已選擇用於篩選方法的微生物而進行，且除了測量在該生物中的PUFA含量外，尚測量脂肪酸分佈和脂肪含量。藉由比較在低氧/缺氧條件下的結果與在有氧條件下的結果，該方法可提供一該測試微生物是否包含本發明之PUFA PKS系統的強烈跡象。此較佳的具體實施例詳細地描述在下列。

初始地，將欲檢驗PUFA PKS系統是否存在的微生物菌株培養在有氧條件下以引導細胞大數量產生(微生物的生物量)。至於鑑別方法的元素，然後將這些細胞放置在低氧或缺氧的培養條件下(例如，在培養基中溶解的氧少於約飽和的5%，更佳地少於約2%，甚至更佳地少於約1%及最佳地溶解的氧約0%飽和)讓其生長約另一個24-72小時。在此方法中，微生物應該在大於約15EC的溫度下培養，更佳地

大於約20EC，甚至更佳地大於約25EC及甚至更佳地大於30EC。低或缺氧的培養環境可在能引發此型式的大氣環境培養室中(因此在培養基中)容易地維持，或可藉由以直接地在培養燒瓶/容器其自身中引發低氧環境之方法培養該些細胞。

5 在一個較佳的培養方法中，可將微生物培養在燒瓶(shake flask)中，取代正常下會包含小量的培養基--少於約總容量的50%及通常少於約總容量的25%--以便當其在振動台上搖晃時保持培養基通氣，代替的是以培養基填充至大於約其容量的50%，更佳地大於約60%及最佳地大於約其容量的75%。高負載培養基的搖瓶可防止當其放置在振動台上時其會在燒瓶中非常良好地混合，而防止氧擴散進入培養基。因此當微生物生長時，它們會用盡存在於培養基中的氧而在搖瓶中天然地產生低或無氧環境。

15 在培養週期後，可獲得該些細胞及分析有興趣的生物活性化合物(例如，脂質)之含量，但是最特別地，為分析包含二個或更多不飽和鍵的化合物，更佳地三個或更多雙鍵及甚至更佳地四個或更多雙鍵。對脂質來說，那些擁有此些化合物在大於約微生物的乾燥重量之5%，更佳地大於約10%，更佳地大於約15%，及甚至更佳地大於約20%的菌株可鑑別如預測地包含上述描述之型式之新穎的PKS系統。對其它生物活性化合物來說，諸如以較小量合成的抗生素或化合物，那些擁有此化合物在大於約微生物的乾燥重量之0.5%，更佳地大於約0.1%，更佳地大於約0.25%，更佳地

大於約0.5%，更佳地大於約0.75%，更佳地大於約1%，更佳地大於約2.5%及更佳地大於約5%的菌株可鑑別如預測地包含上述描述的型式之新穎的PKS系統。

再者(或與此方法相關連)，包含如於本文所描述的新穎
5 PUFA PKS系統之預期的微生物菌株可藉由檢驗菌株(其可
藉由培養該生物或經由已公告或其它容易可獲得的來源而
獲得)的脂肪酸分佈來鑑別。若該微生物包含大於約總脂肪
酸(如為C14：0、C16：0及/或C16：1)的30%，更佳地大於
約40%，更佳地大於約45%及甚至更佳地大於約50%，同時
10 亦產生至少一種含三個或更多不飽和鍵的長鏈脂肪酸，更
佳地4個或更多雙鍵，更佳地5個或更多雙鍵及甚至更佳地6
個或更多雙鍵，則此微生物的菌株鑑別為一種類似的候選
生物而具有在本發明中描述的型式之新穎的PUFA PKS系
統。藉由在上述描述之低氧條件下篩選此生物，且確認其
15 會製造包含二個或更多不飽和鍵的生物活性分子將建議在
生物中存在有新穎的PUFA PKS系統，其可能進一步藉由分
析微生物的基因組而確認。

此方法的成功亦可藉由篩選已熟知包含C17：0及或
C17：1脂肪酸的真核菌株而提高(與上述描述的大百分比的
20 C14：0、C16：0及C16：1脂肪酸相關連)，因為C17：0及
C17：1脂肪酸為細菌(原核生物)系潛在的標誌，或可影響
脂肪酸製造系統。另一個用來鑑別包含新穎的PUFA PKS系
統之菌株的標誌為由該生物產生的簡單脂肪酸分佈。根據
本發明，“簡單脂肪酸分佈”定義為由菌株製造的8或較少的

脂肪酸，其程度大於總脂肪酸的10%。

5 使用這些方法或標誌(單獨地或較佳地以組合方式)的任何一種將可使熟知技藝之人士容易地鑑別出高度預測而包含在本發明中描述之型式之新穎的PUFA PKS系統之微生物菌株。

在一個結合上述描述之許多方法及標誌之較佳的具體實施例中，已發展出新穎的生物理論(biorational)篩選(使用燒瓶培養基)來偵測包含製造PUFA的PKS系統之微生物。此篩選系統如下進行：

10 將部分欲測試的菌株/微生物之培養基放置在具有50毫升的培養基(有氧處理)之250毫升裝有擋板的(baffled)燒瓶中，且將相同的菌株培養基之另一部分放置在具有200毫升的培養基(缺氧/低氧處理)之250毫升無裝有擋板的燒瓶中。可依欲評估的微生物之型式及菌株而使用不同的培養

15 基。將二燒瓶放置在200rpm的振動台上。在48-72小時之培養時間後，該培養物可藉由離心機而獲得，且經由氣相層析法分析該些細胞的脂肪酸甲基酯含量，以對每個培養物測量下列資料：(1)脂肪酸分佈；(2)PUFA含量；及(3)脂肪含量(估計為量總脂肪酸/細胞乾燥重量)。

20 然後，藉由詢問下列五個問題來分析這些資料(是/否)：
比較從低O₂/缺氧燒瓶來之資料與從有氧燒瓶來的資料：

(1)與有氧培養物比較，在低氧培養物中的DHA(或其它PUFA含量)(如為%FAME(脂肪酸甲基酯))是否保持在約相

同(或較佳地增加)的程度？

(2)在缺氧培養物中之C14:0+C16:0+C16:1是否大於約40%TFA？

5 (3)在缺氧培養物中是否有習知的氧依賴性鏈加長酶/去飽和酶途徑之非常些微(<1%如為FAME)(或無)的前驅物(C18:3n-3+C18:2n-6+C18:3n-6)？

(4)與有氧培養物比較，在低氧培養物中的脂肪含量(如為量總脂肪酸/細胞乾燥重量)是否增加？

10 (5)與有氧培養物比較，在低氧培養物中的DHA(或其它PUFA含量)是否增加(如為細胞乾燥重量%)？

若首先三個問題回答為是，此為菌株包含製造長鏈PUFAs用之PKS基因系統好的跡象。越多的問題回答為是(較佳地首先三個問題必需回答為是)，越強的跡象顯示菌株包含此PKS基因系統。若五個問題全部回答為是，則已有
15 非常強的跡象顯示該菌株包含製造長鏈PUFAs用之PKS基因系統。缺乏18:3n-3/18:2n-6/18:3n-6將指示出低氧條件將已關閉或抑制習知的PUFA合成途徑。高的14:0/16:0/16:1脂肪為一種細菌影響的脂肪酸合成分佈(存在有C17:0及17:1亦為此指示劑)及簡單脂肪酸分佈之初始指示劑。
20 在低氧條件下增加的PUFA合成及含PUFA的脂肪合成直接地象徵PUFA PKS系統，因為此系統不需要氧而可製得高不飽和脂肪酸。

最後地，在本發明的鑑別方法中，一旦鑑別出強的候選生物，較佳地篩選該微生物以偵測該微生物是否包含

PUFA PKS系統。例如，可篩選該微生物的基因組以偵測是否
5 否存在一種或多種編碼如於本文描述的PUFA PKS系統區
段之核酸序列。較佳地，此偵測步驟包括合適的核酸偵測
方法，諸如雜交、放大及或定序一種或多種在有興趣的微
生物中之核酸序列。在偵測方法中所使用的探針及/或引子
可起源於任何熟知的PUFA PKS系統，包括在美國專利案號
6,140,486中描述的海洋細菌PUFA PKS系統，或在美國申請
序號09/231,899及於本文中描述的破囊壺菌PUFA PKS系
10 統。一旦鑑別出新穎的PUFA PKS系統，從這些系統來的基
因物質亦可使用來偵測額外新穎的PUFA PKS系統。用於鑑
別及偵測序列目的之核酸雜交、放大及定序方法已在技藝
中熟知。使用這些偵測方法，可評估序列同源物及區段結
構(例如，不同的PUFA PKS官能至區段之存在、數目及/或
安排)，且與於本文中描述之熟知的PUFA PKS系統比較。

15 在某些具體實施例中，PUFA PKS系統可使用生物學試
驗來鑑別。例如，在美國申請序號09/231,899的實例7中描
述出關鍵的實驗結果，其使用相當熟知的某些型式之脂肪
酸合成系統的抑制劑(即，硫乳酸黴素(thiolactomycin))。本
發明家顯示出可特別地阻擋在裂壺菌屬的整個細胞中之
20 PUFAs合成而沒有阻斷短鏈飽和脂肪酸的合成。此結果的
意義如下：本發明家已從裂壺菌屬之cDNA序列分析熟知，
型式I脂肪酸合成酶系統存在於裂壺菌屬中。已熟知硫
乳酸黴素並不抑制型式I FAS系統，此與本發明家的資料一
致，即，飽和脂肪酸(在裂壺菌屬中主要為C14：0及C16：

0)之製造不會因硫乳酸黴素處理而抑制。在文獻或在本發明家擁有的資料中並無跡象顯示出硫乳酸黴素具有任何延長C14:0或C16:0脂肪酸或其去飽和(即轉染短鏈飽和脂肪酸至PUFAs藉由標準途徑)之抑制劑效應。因此，在裂壺菌屬中的PUFA製造被硫乳酸黴素阻擋之事實強烈地說明標準的PUFA合成途徑並不會在裂壺菌屬中製造PUFAs，而是包括了不同的合成途徑。再者，已經先前地測定的是志瓦菌PUFA PKS系統會由硫乳酸黴素抑制(注意本發明的PUFA PKS系統具有型式I及型式II系統二者之元素)，已熟知的是硫乳酸黴素為一種型式II FAS系統的抑制劑(諸如其在大腸桿菌中發現)。因此，此實驗指出裂壺菌屬可製造PUFAs，由於途徑不包含型式I FAS。可使用類似的原理及偵測步驟來偵測已使用於本文中所示的新穎篩選方法而鑑別出之微生物中的PUFA PKS系統。

此外，實例3顯示出額外的生化資料，其提供裂壺菌屬中的PUFAs不由標準途徑製造之證據(即，在C16:0及DHA間之前驅物產物動力學並不在整個細胞中觀察到及，活體外PUFA合成可從薄膜片段分離--標準的PUFA合成途徑之全部的脂肪酸去飽和酶(除了 $\delta 9$ 去飽和酶外，其嵌入該系列的第一個雙鍵)與細胞膜有關)。此型式的生物化學資料可使用來偵測在由上述描述的新穎篩選方法所鑑別出之微生物中的PUFA PKS活性。

可使用本發明之篩選/鑑別方法來篩選的較佳微生物菌株可選自於由下列所組成之群：細菌、藻類、真菌、原

生動物門或原生生物，但是最佳的是來自由藻類、真菌、原生動物門及原生生物組成的真核微生物。這些微生物較佳地能生長且製造出包含二個或多個不飽和鍵的生物活性化合物，在溫度大於約15EC，更佳地大於約20EC，甚至更佳地大於約25EC及最佳地大於約30EC。

在本發明的此方法之某些具體實施例中，可在能於溫度超過約20EC，較佳地超過約25EC及甚至更佳地超過約30EC下製造PUFAs之細菌中鑑別出新穎的細菌PUFA PKS系統。如先前地於本文中所描述，海洋細菌、志瓦菌及海洋弧菌(描述在美國專利案號6,140,486中)，不會在較高的溫度下製造PUFAs，此限制起源於這些細菌的PUFA PKS系統之有效度，特別地在田野條件下的植物應用中。因此，在一個具體實施例中，本發明之篩選方法可使用來鑑別出具有PUFA PKS系統而能在較高的溫度(例如，上述約20、25或30EC)下生長及製造PUFA的細菌。在此具體實施例中，可將真核生長的抑制劑諸如制真菌素(抗真菌劑)或放線菌酮(真核蛋白質合成的抑制劑)加入至瓊脂板中，該板使用來培養/選擇從水樣品/土壤樣品(從描述在下列的孳生地/棲所型式中收集)來的起始菌株。此方法將幫助選擇出沒有(或最小的)真核菌株污染之豐富的細菌菌株。此選擇方法(在高溫(例如30EC)下培養該些板，然後選擇該些會製造至少一種PUFA的菌株之組合)將初始地鑑別出具有PUFA PKS系統而可在高溫下操作的候選細菌菌株(與在先述技藝中的那些細菌菌株相反，其僅在溫度少於約20EC及更佳地低於約

SEC下會顯示出PUFA產物)。

用來收集較佳型式的微生物以用於根據本發明之PUFA PKS系統篩選的場所包括任何下列一種：低氧環境(或接近這些型式的低氧環境之場所，包括在動物的腸子
5 中，包括會消耗微生物或含微生物的食物之無脊椎動物(包括濾食型式生物))、低或無含氧的水生孳生地(包括淡水、鹽液及海洋)、及特別地在海洋中於-或接近-低氧環境(區域)。該些微生物菌株較佳地不為專性厭氧菌，而是可適應生活在有氧及低或缺氧的環境二者之菌株。包含有氧及低
10 氧或缺氧的環境二者之土壤環境亦為發現這些生物的優良環境，特別地在水生孳生地或臨時的水生孳生地等這些型式的土壤。

特別佳的微生物菌株為一選自於由下列所組成之群的菌株：藻類、真菌(包括酵母菌)、原生動物門或原生生物，
15 在其生命循環的部分期間能藉由諸如吞噬作用、吞噬的或內吞的能力等機制來消耗整個細菌細胞(噬菌(bacterivory))，及/或其生命循環具有以變形蟲階段或裸露的原生質存在之階段。此滋養方法將大大地增加將細菌的PKS系統轉移進入真核細胞的潛力，若錯誤發生且細菌細胞
20 或其DNA)不獲得消化反而官能性地併入真核細胞時。

能噬菌(特別地藉由吞噬作用或細胞攝粒作用)的微生物(除了破囊壺菌類(Thraustochytrids)成員外)之菌株可在下列微生物種類(包括但是不限制於此些實例屬)中發現：

在藻類及藻類似的微生物(包括真菌界)中：裸藻綱(例

如：屬眼蟲藻及袋鞭藻屬)；金黃藻綱(例如：棕鞭藻屬)；
 錐囊藻綱(Dinobryaceae)(例如：錐囊藻屬、扁金囊藻屬
 (*Platychnysis*)及定鞭藻類(*Chrysochromulina*))；雙鞭藻綱(包
 括：無色渦鞭毛藻屬(*Cryptothecodinium*)、裸甲藻屬
 5 (*Gymnodinium*)、多甲藻屬(*Peridinium*)、角藻屬(*Ceratium*)、
 螺溝藻屬(*Gyrodinium*)及銳鼻藻屬(*Oxyrrhis*))；褐鞭藻綱種
 類(例如：褐鞭藻屬(*Cryptomonas*)及紅鞭藻屬
 (*Rhodomonas*))；黃綠藻綱(例如：淺色盤藻屬
 (*Olisthodiscus*))(及包括發生變形蟲階段形式的藻類，如在
 10 氣根瘤藻屬(*Rhizochloridaceae*)鞭毛蟲(flagellates)，及巴卻
 絲毛藻(*Aphanochaete pascheri*)、毛枝布米勒藻(*Bumilleria*
stigeoclonium)及對生無隔藻(*Vaucheria geminata*)的游動孢
 子/配子中)；真點蟲徽菌綱(Eustigmatophyceae)；及芮尼斯
 菌綱(Prymnesiopyceae)(包括芮尼斯菌屬(*Prymnesium*)及雙
 15 頭菌屬(*Diacronema*))。

在菌類中，包括：原生單胞菌(Proteromonads)、乳白
 菌(Opalines)、發生胚菌(Developayella)、雙孢子菌
 (Diplophorys)、拉斑絲菌(Larbrinthulids)、破囊壺菌類、雙
 克斯菌(Bicosecids)、卵形菌、亞壺菌綱
 20 (*Hypochytridiomycetes*)、寇馬迅菌(Commatium)、網花球菌
 (*Reticulosphaera*)、毛單胞菌(*Pelagomonas*)、波球菌
 (*Pelapococcus*)、歐力克菌(*Ollicola*)、金黃球菌
 (*Aureococcus*)、巴馬菌(*Parmales*)、針晶菌
 (*Raphidiophytes*)、新芮菌(*Synurids*)、色根瘤菌

(Rhizochromulinales)、片內爾菌(Pedinellales)、隔盤菌(Dictyochales)、金銹頂菌(Chrysomeridales)、束隔金菌(Sarcinochrysidales)、亥烏拉菌(Hydrurales)、希柏代菌(Hibberdiales)及克姆林菌(Chromulinales)。

- 5 在真菌中：黏液菌綱(形成單核粘變形體(myxamoebae))--黏菌；集胞黏菌綱(Acrasiceae)包括集胞黏菌目(例如：沙拼那屬(*Sappinia*))；斑瘤黏菌綱(Guttulinaceae)(例如：斑瘤林黏菌屬(*Guttulinopsis*)、及斑瘤黏菌屬(*Guttulina*))；網柱黏菌綱(Dictysteliaceae)(例如：腐混黴菌屬(*Acrasis*)、網柱黏菌屬(*Dictyostelium*)、多峰地菌屬(*Polysphondylium*)及顆恩亞菌屬(*Coenonia*))；及藻狀菌綱(Phycomyceae)包括：壺菌目(Chytridiales)、安賽菌目(Ancylistales)、後膜囊菌目(Blastocladales)、單卵壺菌目(Monoblepharidales)、水黴菌目(Saprolegniales)、霜黴目(Peronosporales)、毛黴菌目(Mucorales)及蟲霉目(Entomophthorales)。
- 10
- 15

- 在原生動物中：具有能噬菌(包括吞噬作用)的生命階段之原生動物門菌株，其可選自於分類為纖毛蟲，鞭毛蟲或阿米巴之型式。原蟲纖毛蟲包括族群：Chonotrichs、
- 20 Colpodids、Cyrtophores、Haptorids、Karyorelicts、Oligohymenophora、Polyhymenophora (spirotrichs)、Prostomes及吸管蟲綱(Suctoria)。原蟲性鞭毛蟲包括 Biosoecids；Bodonids；Cercomonads；Chrysophytes(例如：Anthophysa屬、Chrysamoemba屬、Chrysosphaerella屬、

Dendromonas屬、Dinobryon屬、魚鱗藻屬(Mallomonas)、棕
 鞭藻屬、Paraphysomonas屬、Poterioochromonas屬、Spumella
 屬、Syncrypta屬、黃群藻屬(Synura)及黃鞭毛藻屬
 (Uroglena)；環鞭毛蟲；隱芽植物(Cryptophytes)(例如：冷
 5 單胞藻屬(Chilomonas)、隱藻屬(Cryptomonas)、青藻屬
 (Cyanomonas)及角單胞藻屬(Goniomonas)；
 Dinoflagellates；Diplomonads；裸藻(Euglenoids)；
 Heterolobosea；Pedinellids；Pelobionts；Phalansteriids；
 Pseudodendromonads；Spongomonads；及團藻目
 10 (Volvocales)(及其它鞭毛蟲，包括未指定的鞭毛蟲屬
 Artodiscus、Clautriavia、Helkesimastix、Kathablepharis及
 Multicilia)。似阿米巴的原生動物包括族群：Actinophryids；
 Centrohelids；Desmothoracids；Diplophryids；Eumamoebae；
 Heterolobosea；Leptomyxids；Nucleariid Filose amoebae；
 15 Pelebionts；Testate amoebae及Vampyrellids(及包括未指定的
 阿米巴類屬：Gymnophrys、Biomyxa、Microcometes、
 Reticulomyxa、Belonocystis、Elaeorhanis、Allelogromia、
 Gromia或Lieberkuhnia)。原蟲目包括下列：
 Percolomonadeae；Heterolobosea；Lyromonadea；
 20 Pseudociliata；Trichomonadea；Hypermastigea；
 Heteromiteae；Telonemea；Cyathobodonea；Ebridea；
 Ppytomyxea；Opalinae；Kinetomonadea；Hemimastigea；
 Protostelea；Myxagastrea；Dictyostelea；Choanomonadea；
 Apicomonadea；Eogregarinea；Neogregarinea；

Coelotrophaea ; Eucoccidea ; Haemosporea ; Piroplasmaea ;
 Spirotrichea ; Prostomatea ; Litostomatea ; Phyllopharyngea ;
 Nassophorea ; Oligohymenophorea ; Colpodea ; Karyorelicta ;
 Nucleochelea ; Centrochelea ; 棘針綱 (Acantharea) ;
 5 Sticholonchea ; 多泡綱 (Polycystinea) ; 濃彩綱 (Phaeodarea) ;
 Lobosea ; Filosea ; Athalamea ; Monothalamea ; Polythalamea ;
 Xenophyophorea ; Schizocladea ; Holosea ; Entamoebae ; 黏
 液孢子蟲 (Myxosporea) ; Actinomyxoea ; Halosporea ;
 Paramyxoea ; Rhombozoa 及 Orthonectea 。

10 本發明的一個較佳具體實施例包括編列在上述的微生物菌株，其已從編列在上述的較佳孳生地處收集。

本發明的一個具體實施例係關於使用上述描述的新穎 PUFA PKS 篩選方法鑑別的任何微生物；關於 PUFA PKS 基因及因此譯出的蛋白質；及關於以任何於本文中描述的方法使用此微生物及/或 PUFA PKS 基因及蛋白質(包括同源物及其片段)。特別地，本發明包括利用本發明之篩選方法鑑別的生物，其然後基因地改性以藉由該 PUFA PKS 系統調節生物活性分子的製造。

15

本發明的更另一個具體實施例係關於一種分離的核酸分子，其包含一編碼從破囊壺菌微生物來的多不飽和脂肪酸 (PUFA) 聚乙醯合成酶 (PKS) 系統之至少一個的生物活性區段或其生物活性片段之核酸序列。如上述討論，本發明家已成功地使用該方法來鑑別一種具有 PUFA PKS 系統的非細菌微生物，來鑑別其它包含 PUFA PKS 系統的破囊壺菌

20

目成員。三種此微生物的鑑別已描述在實例2。特別地，本發明家已使用本發明之篩選方法鑑別破囊壺菌23B(ATCC20892)為已高度預測包含PUFA PKS系統，接著偵測在破囊壺菌23B基因組中與於本文中揭示的裂壺菌屬

5 PUFA PKS基因雜交的序列。軟體裂壺菌屬(*Schizochytrium limacium*)(IFO32693)及優肯那菌(*Ulkenia*)(BP-5601)亦已鑑別為包含PUFA PKS系統的好候選生物。根據這些資料及在破囊壺菌目成員中的類似物，咸信許多其它破囊壺菌目PUFA PKS系統現在可容易地使用由本發明所提供的方法

10 及工具來鑑別。因此，破囊壺菌目PUFA PKS系統及其部分及/或同源物(例如，蛋白質、區段及其片段)、包含此系統及其部分及/或同源物的基因改性生物、及使用此微生物及PUFA PKS系統的方法皆包含在本發明中。

破囊壺菌類的分類已隨著發展而產生修正。分類理論學家將破囊壺菌類放置在藻類或藻類似的原生生物中。但是，因為分類上的不確定，對本發明之目標來說，在本發明中描述為破囊壺菌類之菌株較佳地視為(目：破囊壺菌目；科：破囊壺菌科(*Thraustochytriaceae*)；屬：破囊壺菌屬、裂壺菌屬、拉斑絲菌(*Labyrinthuloides*)或日本壺菌屬

15 (Japonochytrium))。分類的改變則總整理在下列。於本文中所揭示的某些單細胞微生物之菌株為破囊壺菌目成員。破囊壺菌類為一種具有分類歷史問題之海洋真核生物類。破囊壺菌類的分類位置問題已由摩斯(Moss)(1986)、班韋柏(Bahnweb)及傑寇(Jackle)(1986)及錢伯蘭(Chamberlain)及

20

摩斯(1988)所評論。根據本發明，措辭“破囊壺菌”、“破囊壺菌目微生物”及“破囊壺菌目的微生物”可交替地使用。

為了方便起見，破囊壺菌類首先由分類學者分配在藻菌綱(Phycomycetes)(似真菌的藻類)中之其它無色游動孢子真核生物類。但是，藻菌綱的名稱最後則從分類狀態去除，而破囊壺菌類則保留在卵菌綱(Oomycetes)(雙鞭毛游動孢子真菌)中。最初假定卵菌綱與不等鞭毛藻類有關，最後在廣泛範圍的超微結構及生化研究下(由巴爾(Barr)(1983)生物系統(biosystem))支持此假設。卵菌綱事實上由李大耳(Leedale)(李大耳，1974，Taxon 23:261-270)及其它藻類學家接受為不等鞭毛藻類部分。但是，由於其方便的異養本質情形，卵菌綱及破囊壺菌類已大部分由真菌學者(研究真菌的科學家)研究，而非由藻類學家(研究藻類的科學家)。

從另一個分類觀點來看，進化生物學家已發展出二種思考真核生物類如何發展的一般學派。理論之一提出由薄膜包圍的胞器為經由一系列的內共生現象之外生源(馬爾估里斯(Margulis)，1970，真核細胞之起源(Origin of Eukaryotic Cells)，耶魯大學出版社(Yale University Press)，紐哈文(New Haven))；例如，粒腺體起源於細菌的內共生體、葉綠體起源於藍藻類及鞭毛起源於螺旋體。另一個理論則建議由薄膜包圍的胞器乃經由自體過程從無薄膜包圍的原核生物原型系統中逐漸地發展形成(卡伐里爾(Cavalier)-史密斯(Smith)1975，自然(Lond.)256:462-468)。但是，此二組進化生物學家已將卵菌綱與破囊壺菌類移出

真菌，而將其與雜色藻類(chromophyte algae)放置在雜色藻界(Chromophyta)(卡伐里爾-史密斯，1981，生物系統14:461-481)(此界更近已擴大至包括其它原生生物，且此界的成員現在稱為菌類界(Stramenopiles))；或與全部的藻類
5 放置在原生生物界(馬爾估里斯及沙根(Sagan)，1985，生物系統18:141-147)。

隨著電子顯微鏡的發展，在二種破囊壺菌類的游動孢子之超微結構上的研究，破囊壺菌屬及裂壺菌屬(博金斯(Perkins)，1976，pp.279-312在“水生真菌學的最新發展
10 (Recent Advances in Aquatic Mycology)”(ed. E. B. G. 瓊斯(Jones))，約翰威利&桑斯(John Wiley & Sons)，紐約；卡日瑪(Kazama)，1980，*Can. J. Bot.* 58:2434-2446；巴爾，1981，生物系統14:359-370)已提供破囊壺菌科僅略微地與卵菌綱相關的好證據。額外地，可顯示5S核糖體RNA序列的相依
15 分析(一種多元統計資料形式)之基因資料指出破囊壺菌目明顯地為一種獨特的真核生物群，其可完全地從卵菌綱成員的真菌中分出，而最緊密地與紅及褐藻類相關(門內拉(Mannella)等人，1987，*Mol. Evol.* 24:228-235)。大部分的分類學者已同意將破囊壺菌類移出卵菌綱(巴克尼吉-蓋西
20 亞(Bartnicki-Garcia)，1987，pp.389-403在“真菌的進化生物學(Evolution Biology of Fungi)”中(eds. 雷納(Rayner)，A.D.M.，布拉希爾(Brasier)，C.M.&摩爾(Moore)，D.)，康橋(Cambridge)大學出版社，康橋)。

總而言之，使用卡伐里爾-史密斯的分類系統(1981，，

生物系統14：461-481，1983；卡伐里爾-史密斯，1993，*Microbiol Rev.* 57：953-994)，破囊壺菌類與雜色藻類乃分類在雜色藻界(真菌界)。此分類位置已更近由卡伐里爾-史密斯等人使用不等鞭毛藻門的18S rRNA識別標誌重申，以說明破囊壺菌類為chromists而非真菌(卡伐里爾-史密斯等人，1994，*Phil. Tran. Roy. Soc. 倫敦系列生物科學(London Series BioSciences)*346：387-397)。此將其放置在與真菌完全不同的界，其全部放置在狹義的真菌界(Eufungi)中。因此，將破囊壺菌類的分類位置總整理在下列：

- 10 界：雜色藻界(真菌界)
 門：不等鞭毛門(Heterokonta)
 目：破囊壺菌目
 科：破囊壺菌科
 屬：破囊壺菌屬、裂壺菌屬、拉斑絲菌屬或日本壺菌
 15 屬

某些早期的分類學者將破囊壺菌屬的少數原始成員(具有變形蟲生命階段的那些)分離出個別的屬稱為優肯那菌屬。但是，現在已熟知大部分(若非全部)的破囊壺菌類(包括破囊壺菌屬及裂壺菌屬)具有變形蟲階段，如此某些人不將優肯那菌屬視為有用的屬。如於本文中所使用，破囊壺菌屬將包括優肯那菌。

雖然在較高的門及界分類中處於不確定的分類位置，破囊壺菌類仍然為一種特殊及有特徵的族群，其成員仍然可分類在破囊壺菌目中。

多不飽和脂肪酸類(PUFAs)為較高的真菌類中之基本
薄膜組分及許多由脂質衍生的信號分子之前驅物。本發明
之PUFA PKS系統使用的PUFA合成途徑並不需要飽和脂肪
酸的去飽和及延長反應。該些途徑利用PUFA PKSs催化，
5 其與先前了解的PKSs在結構及機制二者上有區別。順式雙
鍵的產生建議包含位置-特定的異構酶類；這些酵素類相信
可有用地用於製造新穎的抗生素家族。

為了使用本發明之PUFA PKS系統來製造明顯高產率
的多種生物活性分子，生物(較佳地為微生物或植物)可經基
10 因地改性以影響PUFA PKS系統的活性。在一個觀點中，此
生物可內生地包含及表現出PUFA PKS系統，及該基因改性
可為一種或多種內生的PUFA PKS系統之官能性區段的基
因改性，藉此，該改性在PUFA PKS系統的活性上具有某些
影響。在另一個觀點中，此生物可內生地包含及表現出
15 PUFA PKS系統，及該基因改性可引進至少一種外生的核酸
序列(例如，重組型核酸分子)，其中該外源核酸序列編碼從
第二PKS系統來的至少一個生物活性區段或蛋白質，及/或
一種可影響該PUFA PKS系統的活性之蛋白質(例如，磷酸
泛醯巯基乙胺基轉位酶(PPTase)，在下列討論)。在更另一
20 個觀點中，該生物不必需內生地(天然地)包含一PUFA PKS
系統，而可基因地改性以引進至少一種能譯出具有PUFA
PKS系統的至少一個區段之生物活性的胺基酸序列之重組
型核酸分子。在此觀點中，PUFA PKS活性可由引進或增加
生物中的PUFA PKS活性而影響。與每個這些觀點相關的不

同具體實施例將在下列更詳細地討論。

因此，根據本發明，一個具體實施例係關於一種基因改性的微生物，其中該微生物可表現出一種包含多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統的至少一個生物活性區段之PKS系統。該PUFA PKS系統的至少一個區段可藉由選自於下列的核酸序列譯出：(a)一核酸序列，其編碼從破囊壺菌微生物來的多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統之至少一個區段；(b)一核酸序列，其編碼從由本發明之篩選方法所鑑別出的微生物來之PUFA PKS系統的至少一個區段；(c)一核酸序列，其編碼一胺基酸序列其與選自於由下列所組成之群的胺基酸序列：序列辨識編號：2、序列辨識編號：4及序列辨識編號：6之至少500個連續胺基酸有至少約60%相同；其中該胺基酸序列具有PUFA PKS系統的至少一個區段之生物活性；及，(d)一核酸序列，其編碼一與選自於由下列所組成之群的胺基酸序列有至少約60%相同的胺基酸序列：序列辨識編號：8、序列辨識編號：10、序列辨識編號：13、序列辨識編號：18、序列辨識編號：20、序列辨識編號：22、序列辨識編號：24、序列辨識編號：26、序列辨識編號：28、序列辨識編號：30及序列辨識編號：32；其中該胺基酸序列具有PUFA PKS系統的至少一個區段之生物活性。該基因改性可影響生物中的PKS系統活性。在部分(b)中提及的篩選方法已在上述中詳細地描述，且包括的步驟有：(a)選擇一可製造至少一種PUFA的微生物；及，(b)鑑別從(a)來的微生物，其具有在

發酵培養液中於溶氧條件少於約飽和的5%下可製造增加的PUFAs之能力，如與在發酵培養液中於溶氧條件大於約飽和的5%下由微生物製造之PUFAs比較，較佳地約10%，更佳地約15%，更佳地約飽和的20%。該經基因改性的微生物可包括任何一種或多種上述-鑑別的核酸序列，及/或任何裂壺菌屬PUFA PKS ORFs或區段的任何其它同源物，如上述詳細地描述。

如於本文中所使用，該基因改性的微生物可包括基因改性的細菌、原生生物、微藻類、真菌或其它微生物及特別地，於本文描述的破囊壺菌目(例如，破囊壺菌)的任何屬(例如，裂壺菌屬、破囊壺菌屬、日本壺菌屬、拉斑絲菌)。此基因改性的微生物具有一基因組，其已從其正常(即，野生型或天然發生)形式改性(即，突變或改變)，如此可獲得想要的結果(即，增加或改性PUFA PKS活性及/或使用該PKS系統來製造想要的產物)。微生物的基因改性可使用標準菌株成長及/或分子基因技術來達成。此類技術已在技藝中熟知且通常揭示可用於微生物，例如，在山姆布魯克等人，1989，分子選殖：實驗室手冊，冷泉哈伯實驗室出版社。山姆布魯克等人(同前所述)之參考資料其全文以參考方式併入本文。經基因改性的微生物可包括一種微生物，其中核酸分子已嵌入、刪除或改性(即，突變；例如，藉由核苷酸的嵌入、刪除、取代及/或反轉)，以此方法改性可在微生物中提供想要的效應。

可根據本發明來改性的較佳微生物宿主細胞包括(但

是非為限制)任何細菌、原生生物、微藻類、真菌或原生動物門。在一個觀點中，較佳的基因改性微生物包括(但是非為限制)任何破囊壺菌目微生物。可使用於本發明的特別佳宿主細胞可包括從下列屬來的微生物(但是非為限制):破囊壺菌屬、拉斑絲菌、日本壺菌屬及裂壺菌屬。在這些屬中的較佳物種包括(但是非為限制):任何裂壺菌屬物種，包括叢狀裂壺菌(*Schizochytrium aggregatum*)、軟體裂壺菌屬、微小(*minutum*)裂壺菌屬;任何破囊壺菌屬物種(包括先前的優肯那物種，諸如 *U. 維斯兒均 (visurgensis)*、*U. 阿米巴 (amoeboida)*、*U. 薩卡瑞那 (sarkariana)*、*U. 普分達 (profunda)*、*U. 輻射狀 (radiata)*、*U. 微小 (minuta)*及優肯菌 *sp. BP-5601*)，及包括紋狀破囊壺菌 (*Thraustochytrium strium*)、黃金破囊壺菌 (*Thraustochytrium aureum*)、粉紅破囊壺菌 (*Thraustochytrium roseum*)；及任何日本壺菌屬物種。特別佳的破囊壺菌目菌株包括(但是非為限制):裂壺菌 (S31)(ATCC20888)；裂壺菌 (S8)(ATCC20889)；裂壺菌 (LC-RM)(ATCC18915)；裂壺菌 (SR21)；叢狀裂壺菌 (Goldstein et Belsky)(ATCC28209)；微小裂壺菌屬 (Honda et Yokochi)(IFO32693)；破囊壺菌 (23B)(ATCC20891)；紋狀破囊壺菌 (Schneider) (ATCC24473)；黃金破囊壺菌 (Goldstein)(ATCC34304)；粉紅破囊壺菌 (Goldstein)(ATCC28210)；及日本壺菌 (L1)(ATCC28207)。其它合適的用來基因改性的宿主微生物實例包括(但是非為限制)酵母菌包括出芽酵母菌 (*Saccharomyces*

cerevisiae)、卡爾酵母菌(*Saccharomyces carlsbergensis*)；或其它酵母菌，諸如念珠菌屬、克魯維酵母菌屬(*Kluyveromyces*)；或其它真菌，例如，絲狀真菌，諸如麴菌屬、脈孢菌屬(*Neurospora*)、青黴菌屬(*Penicillium*)等等

5 細菌細胞亦可使用作為宿主。此包括大腸桿菌，其在發酵製程中 useful。再者，亦可使用諸如乳酸菌物種或桿菌屬物種等宿主作為宿主。

本發明的另一個具體實施例係關於一種基因改性的植物，其中該植物已基因地改性以重組地表現出一包含多不

10 飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統的至少一個生物活性區段之PKS系統。該區段可由選自於下列的核酸序列譯出：(a)一核酸序列，其編碼從破囊壺菌微生物來之多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統的至少一個區段；(b)一核酸序列，其編碼從藉由於本文中描述的篩選

15 及選擇方法所鑑別出的微生物來之PUFA PKS系統的至少一個區段(參見在上述基因改性的微生物之討論中的方法簡述)；(c)一核酸序列，其編碼選自於由下列所組成之群的胺基酸序列：序列辨識編號：2、序列辨識編號：4、序列辨識編號：6及其生物活性片段；(d)一核酸序列，其編碼選

20 自於由下列所組成之群的胺基酸序列：序列辨識編號：8、序列辨識編號：10、序列辨識編號：13、序列辨識編號：18、序列辨識編號：20、序列辨識編號：22、序列辨識編號：24、序列辨識編號：26、序列辨識編號：28、序列辨識編號：30、序列辨識編號：32及其生物活性片段；(e)一

核酸序列，其編碼一胺基酸序列其與選自於由下列所組成之群的胺基酸序列：序列辨識編號：2、序列辨識編號：4，及序列辨識編號：6的至少500個連續胺基酸有至少約60%相同；其中該胺基酸序列具有PUFA PKS系統的至少一個區段之生物活性；及/或(f)一核酸序列，其編碼一與選自於由下列所組成之群的胺基酸序列有至少約60%相同之胺基酸序列：序列辨識編號：8、序列辨識編號：10、序列辨識編號：13、序列辨識編號：18、序列辨識編號：20、序列辨識編號：22、序列辨識編號：24、序列辨識編號：26、序列辨識編號：28、序列辨識編號：30及序列辨識編號：32；其中該胺基酸序列具有PUFA PKS系統的至少一個區段之生物活性。該基因改性的植物可包括任何一種或多種上述鑑別的核酸序列，及/或任何裂壺菌屬PUFA PKS ORFs或區段的任何其它同源物，如在上述詳細地描述。

如於本文中所使用，該基因改性的植物可包括任何基因改性的植物(包括高等植物)，特別地，任何可消耗的植物或可有用用來製造本發明想要的生物活性分子之植物。此基因改性的植物已從其正常(即，野生型或天然發生)形式的基因組改性(即，突變或改變)，如此可獲得想要的結果(即，增加或改性PUFA PKS活性及/或使用該PKS系統製造想要的產物)。植物的基因改性可使用標準的菌株成長及/或分子基因技術來達成。將編碼想要的胺基酸序列之重組型核酸分子併入該基因組植物的製造轉殖基因植物之方法已在技藝中熟知。可根據本發明來基因改性的較佳植物較佳地為

一種合適於由動物(包括人類)消耗的植物。

可根據本發明之基因改性的較佳植物(即，植物宿主細胞)包括(但是非為限制)任何高等植物，及特別是可消耗的植物(包括莊稼植物)，特別是使用其油的植物。此類植物可
5 包括例如：油菜籽(canola)、大豆類、油菜籽(rapeseed)、亞麻子、玉米、紅花類、向日葵類及煙草。其它較佳的植物包括那些已熟知可製造出能使用作為醫藥劑、調味劑、營養品、機能性食物成分或化妝活性的試劑之化合物的植物，或可基因地設計以製造這些化合物/試劑的植物。

10 根據本發明，該基因改性的微生物或植物包括已使用重組技藝改性的微生物或植物。如於本文中所使用，會造成基因表現、基因功能或基因產物功能(即，可由基因譯出的蛋白質)減低的基因改性可指為基因的鈍化
15 (inactivation)(完全或部分)、刪除(deletion)、中斷(interruption)、妨礙(blockage)或向下調節。例如，會造成由此基因譯出的蛋白質之功能減低的基因之基因改性可為基因完全刪除(即，基因不存在，因此蛋白質不存在)；基因突變，而造成蛋白質不完全或不轉譯(例如，蛋白質不表現)；或基因突變，其會減少或徹底破壞蛋白質的天然功能
20 (例如，一可表現出具有減少或無酵素活性或作用之蛋白質)之結果。會造成基因表現或功能增加之基因改性可指為基因的放大、過度產生(overproduction)、過度表現(overexpression)、活化、提高、加入或向上調節。

根據本發明的微生物或植物之基因改性較佳地會影響

由植物表現的PKS系統之活性，不論該PKS系統是否為內生的及經基因改性的、將重組型核酸分子引進生物而內生的、或完全利用重組技藝提供的。根據本發明，“影響PKS系統的活性”包括任何基因改性，其可在由生物表現的PKS

5 系統中造成任何可偵測或可測量的改變或改性，如與缺乏基因改性的比較。在PKS系統中之可偵測的改變或改性可包括(但是非為限制)：將PKS系統活性引進一生物，如此該生物現在具有可測量/可偵測的PKS系統活性(即，該生物在基因改性之前不包括PKS系統)；將從不同的PKS系統(而非

10 由生物內生地表現的PKS系統)來之官能性區段引進生物，如此該PKS系統活性可改性(例如，將細菌的PUFA PKS區段或型式I PKS區段引進其內生地表現出非細菌的PUFA PKS系統之生物)；由該PKS系統所製造的生物活性分子之量的改變(例如，該系統可製造出更多(增加量)或較少(減少量)

15 所提供的產物，如與缺乏基因改性比較)；由PKS該系統所製造的生物活性分子的型式之改變(例如，該系統可製造出新穎或不同的產物，或由該系統天然地製造的產物的變異物)；及/或由該PKS系統所製造的多種生物活性分子之比率改變(例如，該系統可製造出不同比率的一種PUFA與另一

20 種PUFA；其可製造出一種完全不同的脂質分佈，如與缺乏基因改性的比較；或其可將多種PUFAs放置在甘油三酯的不同位置中，如與天然結構比較)。此基因改性包括任何型式的基因改性，特別地包括藉由重組技藝及藉由標準的致突變所製得之改性。

應注意的是，所提及的增加在PUFA PKS系統中官能性區段或蛋白質的活性指為在包含該區段或蛋白質的(或該區段或蛋白質欲引進的)生物中之任何基因改性，其可使該區段或蛋白質系統增加官能性；且其可包括該區段或蛋白質的較高活性(例如，特定的活性或活體內的酵素活性)；可減低該區段或蛋白質系統之抑制或降解；及可過度表現該區段或蛋白質。例如，可藉由使用能提供比天生的起動子具有更高表現程度的起動子而增加基因複製數目、增加表現程度；或基因可藉由基因工程或標準致突變來改變以增加由基因所譯出的區段或蛋白質之活性。

類似地，所提的減低在PUFA PKS系統中的官能性區段或蛋白質之活性指為在包含此區段或蛋白質(或該區段或蛋白質欲引進的)的生物中之任何基因改性，其會造成減低區段或蛋白質的官能性；其可包括減低該區段或蛋白質的活性；其可增加該區段或蛋白質的抑制或降解；及可減低或消除該區段或蛋白質的表現性。例如，本發明之區段或蛋白質作用可藉由阻礙或減低該區段或蛋白質的製造、“關閉”編碼該區段或蛋白質之基因或其部分、減低該區段或蛋白質的活性、或抑制該區段或蛋白質的活性而降低。阻礙或減低一區段或蛋白質的製造可包括將該能譯出該區段或蛋白質的基因放置在起動子(promotor)的控制下，其需要在生長培養基中存在一引導化合物。藉由建立此些條件，如此誘導物(inducer)會變成從培養基中減少，而編碼該區段或蛋白質的基因之表現性(因此，蛋白質合成)可關閉。阻礙或

減低該區段或蛋白質的活性亦可包括使用切除技藝方法(類似於在美國專利案號4,743,546中所描述,其以參考之方式併於本文)。為了使用此方法,編碼有興趣的蛋白質之基因可在特定的基因序列(其允許從基因組中切除特定的、經控制的基因)間選殖。切除可藉由例如偏移在培養物中的培養溫度(如在美國專利案號4,743,546中),或藉由某些其它物理或營養的訊號來引發。

在本發明的一個具體實施例中,該基因改性包括一種編碼具有如於本文描述的非細菌PUFA PKS系統之至少一個區段的生物活性之胺基酸序列的核酸序列之改性。此一改性可對一在內生地(天然地)表現的非細菌PUFA PKS系統中的胺基酸序列,藉此,天然地包含此系統的微生物可藉由例如標準的致突變、選擇技術及/或分子基因技術(包括基因工程技術)而基因地改性。基因工程技術可包括,例如,使用目標重組載體來刪除部分內生基因,或以一異種序列取代部分內生基因。可引進宿主基因組的異種序列之實例包括編碼從另一個PKS系統來的至少一種官能性區段之序列,諸如不同的非細菌PUFA PKS系統、細菌的PUFA PKS系統、型式I PKS系統、型式II PKS系統或組合式PKS系統。引進宿主的基因組之其它異種序列包括一種編碼不為PKS系統的區段之蛋白質或官能性區段的序列,但是其將影響內生的PKS系統之活性。例如,可將能譯出磷酸泛醯巰基乙胺基轉位酶的核酸分子引進宿主基因組(在下列討論)。在下列將詳細地討論可對內生的PUFA PKS系統製得的特定

改性。

在本發明的此具體實施例之另一個觀點中，該基因改性可包括：(1)引進一編碼具有非細菌的PUFA PKS系統之至少一個區段的生物活性之胺基酸序列的重組型核酸分子；

5 及/或(2)引進一編碼能影響PUFA PKS系統活性的蛋白質或官能性區段之重組型核酸分子進入宿主。該宿主可包括：(1)一不表現出任何PKS系統的宿主細胞，其中PKS系統的全部官能性區段可引進該宿主細胞，及其中至少一個官能性區段來自非細菌的PUFA PKS系統；(2)一可表現出一具有

10 非細菌的PUFA PKS系統之至少一個官能性區段的PKS系統(內生或重組)之宿主細胞，其中該引進的重組型核酸分子編碼至少一種額外的非細菌PUFA PKS區段功能，或另一種會影響宿主PKS系統的活性之蛋白質或區段；及(3)一可表現一非必需的PKS系統(內生或重組)之宿主細胞，其包括一

15 種從非細菌的PUFA PKS來之區段功能，其中該引進的重組型核酸分子包括一編碼非細菌的PUFA PKS系統之至少一個官能性區段之核酸序列。換句話說，本發意欲包含任何基因改性的生物(例如，微生物或植物)，其中該生物包含至少一種非細菌的PUFA PKS區段功能(內生性地或藉由重組

20 改性)，其中該基因改性在該非細菌的PUFA PKS區段功能上或在該PKS系統上具有可測量的效應，當該生物包含一官能性PKS系統時。

因此，當使用本發明之非細菌的PUFA PKS系統時，其例如可使用從破囊壺菌PUFA PKS系統來的基因，該基因混

合物可使用來延伸PUFA產物的範圍至包括EPA、DHA、ARA、GLA、SDA及其他，和製造出廣泛多種的生物活性分子，包括抗生素、其它醫藥化合物及其它想要的產物。獲得這些生物活性分子的方法包括不僅混合從不同生物來的基因，而且亦包含多種基因地改性於本文中揭示的非細菌PUFA PKS基因的方法。本發明之非細菌的PUFA PKS系統之基因基礎與區段結構的知識可提供做為設計能製造多種生物活性分子的新穎基因改性生物之基礎。雖然任何PKS區段及相關的基因之混合及改性已由本發明家預期到，以實例說明之，PUFA-PKS系統的不同可能操控則在下列關於基因改性及生物活性分子製造中討論。

例如，在一個具體實施例中，非細菌的PUFA-PKS系統產物(諸如由破囊壺菌類製造的那些)可藉由改性CLF(鏈長因子)區段而改變。此區段為型式II(分解的酵素類)PKS系統之特徵。其胺基酸序列顯示出與KS(酮基合成酶對)區段同源，但是其缺乏活性位置半胱胺酸。CLF可作用以決定最後產物的延長循環數目(因此鏈長度)。在本發明的此具體實施例中，使用現知的FAS及PKS合成狀態，可提供藉由直接地改性非細菌的PUFA-PKS系統來製造ARA的合理對策。在文獻中關於CLF在PKS系統中的功能尚有爭議(C.畢山等人，自然401，502(1999))，且已認知的是其它區段可與最後產物的鏈長度之決定有關。但是，可明瞭的是裂壺菌屬能製造DHA(C22：6，T-3)及DPA(C22：5，T-6)二者。在PUFA-PKS系統中，該順式雙鍵可在增加碳鏈合成期間引

進。因為T-3及T-6雙鍵的放置在分子合成中較早地發生，故不預計它們將影響隨後的最後-產物鏈長之決定。因此，不由理論所限制，本發明家相信將能引導C20單元(而非C22單元)之合成的因素(例如CLF)引進裂壺菌屬PUFA-PKS系統，可造成製造出EPA(C20:5, T-3)及ARA(C20:4, T-6)。例如，可將CLF利用在異種系統中，藉由直接地代替從EPA製造系統(諸如從發光菌屬來的)來的CLF進入裂壺菌屬基因組。然後，可分析所產生的轉化株之脂肪酸的分佈改變以鑑別出製造EPA及/或ARA的轉化株。

除了依賴異種系統(重組系統，諸如可引進植物)的發展外，CLF概念可利用在裂壺菌屬中(即，藉由裂壺菌屬基因組之改性)。轉染及同源重組已在裂壺菌屬中說明。此可利用來以OrfB之CLF建構一選殖物，而取代以從C20 PUFA-PKS系統來的CLF建構。標誌基因將嵌入編碼區域的下游。然後可轉染該野生型細胞，選擇該標誌顯型，然後篩選已併入新穎CLF者。再次，可分析這些在脂肪酸分佈上的任何影響以鑑別出可製造EPA及/或ARA的轉化株。若發現某些其它非與CLF相關的那些因素可影響最後產物的鏈長時，可使用類似的對策來改變那些因素。

另一個包含PUFA-PKS產物改變的較佳具體實施例包括3-羥基醯基-ACP脫水酶/酮基合成酶對之改性或取代。在順-法生油酸(C18:1, Δ^{11})於大腸桿菌中合成期間，順式雙鍵的產生相信對特定的DH醯素、3-羥基醯基-ACP脫水酶、*FabA*基因的產物具依賴性。此醯素可從3-酮基醯基-ACP移

出HOH而在碳鏈中留下反式雙鍵。DH、類FabA的次小組擁有順-反異構酶活性(喜斯等人, 1996, 前述)。細菌及非細菌的PUFA-PKS系統之新穎觀點為存在二個類FabA的DH區段。不由理論限制, 本發明家相信這些DH區段的一個或二者將具有順-反異構酶活性(DH區段的操控在下列更詳細地討論)。

另一個在大腸桿菌中合成不飽和脂肪酸的觀點為需要特別的KS酵素(β -丙脂醯基-ACP合成酶)、*FabB*基因的產物。此為進行脂肪酸縮合的酵素, 其會在活性位置處以丙二酸單醯基-ACP連結至半胱胺酸殘基(藉由硫-酯鍵結)。在多步驟反應中, 會釋放CO₂且直鏈以二個碳的方式延長。咸信僅有此KS可延長包含雙鍵的碳鏈。此延長僅當該雙鍵為順式結構時才發生; 若其於反式結構, 則該雙鍵會由烯醯-ACP還原酶(ER)在延長之前還原(喜斯等人, 1996, 前述)。

已特徵化的全部PUFA-PKS系統至此具有二個KS區段, 其中之一顯示出比另一個更與大腸桿菌類*FabB*的KS同源。再次, 不由理論限制, 本發明家相信在PUFA-PKS系統中, DH(類*FabA*)及KS(類*FabB*)酵素區段的特異性及交互作用可決定順式雙鍵在最後產物中的數目及位置。因為2-碳延長反應的數目大於存在於PUFA-PKS最後產物中的雙鍵數目, 可決定的是在某些延長循環中會發生完全還原。因此, DH及KS區段可使用來作為改變DHA/DPA比率或其它長鏈脂肪酸比率的目標。這些可藉由引進從其它系統來的同源區段或藉由致突變這些基因片段來改性及/或評估。

在另一個具體實施例中，ER(烯醯-ACP還原酶-一種酵素，其可還原在脂肪醯基-ACP中的反式雙鍵而產生全飽和碳)區段可經改性或取代以改變由PKS系統製造的產物型式。例如，本發明家熟知裂壺菌屬PUFA-PKS系統與先前描述5 的細菌系統(其有二個(而非一個)ER區段)不同。不由理論限制，本發明家相信這些ER區段會強烈地影響所產生的PKS製造產物。所產生的PKS產物可藉由分別地關閉各別的區段或藉由改性其核苷酸序列或藉由從其它生物來的ER區段取代而改變。

10 在另一個具體實施例中，可將能譯出不為PKS系統部分的蛋白質或區段(但是其會影響PKS系統)之核酸分子引進生物。例如，上述描述的全部PUFA PKS系統包括多重、銜接的ACP區段。ACP(為一種分離的蛋白質或為一種較大的蛋白質之區段)需要磷酸泛醯巯基乙胺輔因子之附著才15 可產生活性，holo-ACP。磷酸泛醯巯基乙胺附著至apo-ACP可藉由酵素(磷酸泛醯巯基乙胺基轉位酶(PPTase))的總科成員來進行(蘭巴拉特(Lambalot)R.H.等人，化學及生物學，3，923(1996))。

與其它PKS及FAS系統類似，本發明家推測存在於裂壺20 菌屬 ORFA蛋白質中的多重ACP區段之活化可藉由特定內生的PPTase來進行。編碼此假定的(presumed)PPTase之基因尚未在裂壺菌屬中鑑別。若此基因存在於裂壺菌屬中，可設想一些能使用的鑑別方法及選殖方法。這些可包括(但是將不限制至)：產生及部分定序從活化地生長的裂壺菌屬細

胞而製備之cDNA基因庫(注意，一個序列已於現在可獲得的裂壺菌屬cDNA基因庫組中鑑別，而顯示出與PPTase同源；但是，其顯示出為多區段FAS蛋白質的部分，如此無法譯出想要的OrfA特定PPTase)；使用存在於許多PPTase中的
5 胺基酸母題而在PCR反應中設計之降解寡核苷酸引子(以獲得一種可篩選基因組或cDNA基因庫的核酸探針分子)；以蛋白質-蛋白質交互作用為主的基因方法(例如酵母菌二種混合物系統)，其中用ORFA-ACP區段將使用作為“誘餌”以發現“目標”(即PPTase)；及純化及部分定序該酵素其自身，
10 作為一種產生篩選基因組或cDNA基因庫用之核酸探針工具。

亦可想像的是異種PPTase可活化該裂壺菌屬ORFA ACP區段。已顯示出某些PPTases具有寬廣的基質容忍度，例如枯草桿菌的sfp酵素(蘭巴拉特等人，前述)及輪枝鏈黴菌(*Streptomyces verticillus*)的svp酵素(山雀日(Sanchez)等人，2001，化學&生物學8：725-738)。這些酵素可經測試以看看它們是否能活化該裂壺菌屬ACP區段。同樣地，最近的公告描述在煙草中真菌PKS蛋白質的表現性(亞爾潘尼(Yalpani)等人，2001，植物細胞(*The Plant Cell*)13：
15 1401-1409)。已引進PKS系統之產物(其可由展青黴菌(*Penicillium patulum*)的6-甲基水楊酸合成酶基因譯出)已在轉殖基因的植物中偵測，雖然相符合的真菌PPTase並不存在於那些植物中。此建議內生的植物PPTase可識別及活化該真菌PKS ACP區段。與此觀察相關聯，本發明家已在阿
20

拉伯芥屬 (*Arabidopsis*) 的全部基因組資料庫 (其編碼 PPTases) 中鑑別出二種序列 (基因)。這些序列 (基因銀行編號 ; AAG51443 及 AAC05345) 現在編列為編碼“未知的蛋白質”。根據存在於已轉譯的蛋白質序列中之一些識別標誌母題 (包括 G(I/V)D 及 WxxKE(A/S)xxK (序列辨識編號 : 33)) ,

5 它們可鑑別為推定的 (putative) PPTases (列在蘭巴拉特等人中, 1996, 為全部 PPTases 的特徵)。此外, 這二種推定的蛋白質包括二種額外的母題, 其典型地在 PPTases 中發現而典型地與 PKS 及非核糖體的胜肽合成系統相關 ; 即,

10 FN(I/L/V)SHS (序列辨識編號 : 34) 及 (I/V/L)G(I/L/V)D(I/L/V) (序列辨識編號 : 35)。再者, 這些母題會發生在蛋白質序列中預計的相對位置處。似乎該阿拉伯芥屬基因的同源物存在於其它植物中, 諸如煙草。再次, 這些基因可經選殖及表現以看看它們編碼的該些酵素是否能活化該些裂壺菌屬

15 ORFA ACP 區段, 或再者, OrfA 可否直接地表現在轉殖基因的植物中 (其目標為質體或細胞質)。

另一個可識別 ORFA ACP 區段為基質的異種 PPTase 為 Het I 蛋白質念珠藻 PCC7120 (先前稱為魚腥藻 (*Anabaena sp.*) PCC7120)。如在美國專利案號 6,140,486 中所提到,

20 *Shewanella* 的一些 PUFA-PKS 基因顯示出與存在於 PKS 團簇 (其在念珠藻屬中發現) 中之蛋白質區段有高精度同源 (該專利的第 2 圖)。此念珠藻 PKS 系統與長鏈 (C26 或 C28) 羥基脂肪酸的合成有關, 其合適酯化成糖部分而形成部分的異型細胞細胞壁。這些念珠藻 PKS 區段亦與在裂壺菌屬 PKS 蛋白質

之Orfs B及C中發現的區段高度地同源(即相同物與在志瓦
 菌PKS蛋白質中發現的那些相符合)。直到非常最近，存在
 於基因銀行資料庫中的念珠藻PKS區段並無顯示出與裂壺
 菌屬OrfA的任何區段有高度的同源性(或同源的志瓦菌
 5 Orf5蛋白質)。但是，該念珠藻的完整基因組最近已定序，
 且結果現在僅可獲得PKS基因團簇的上游區域之序列。在
 此區域有三個Orfs顯示出與OrfA的區段同源(KS、MAT、
 ACP及KR)(參見第3圖)。包含在此組中的為二種ACP區段，
 此二者顯示出與ORFA ACP區段高度同源。在念珠藻PKS團
 10 簇的末端為編碼該Het I PPTase的基因。先前地，並不明瞭
 的是該Het I酵素可為誰的基質，但是在新近鑑別的Orf(Hgl
 E)團簇中存在的銜接ACP區段強烈地建議至本發明家其為
 那些ACPs。裂壺菌屬及念珠藻的ACP區段之同源物和在二
 蛋白質中區段的銜接安排使得Het I為一種很有可能的裂壺
 15 菌屬ORFA ACPs之異種活性的候選物種。本發明家相信首
 先要識別及考慮此念珠藻HetI PPTase的用途。

如在美茲等人(2001，前述)中指出，PUFA PKS系統的一
 種新穎特徵為存在二個脫水酶區段，二者皆顯示出與大
 腸桿菌的FabA蛋白質同源。隨著上述提及的新穎念珠藻
 20 PKS基因序列之可用度，現在可比較二種系統及其產物。
 在念珠藻團簇中的區段序列(從Hgl E至Het I)(其已由本發
 明家定義)為(參見第3圖)：

KS-MAT-2xACP·KR·KS·CLF-AT·ER(HetM·HetN)HetI

在裂壺菌屬PUFA-PKS Orfs A、B&C中，該序列

(OrfA-B-C)為：

KS-MAT-9xACP-KR KS-CLF-AT-ER DH-DH-ER

可看見該些區段序列的相符合度(亦有高的胺基酸序列同源性)。該念珠藻PKS系統的產物為一種不含雙鍵(順式或反式)的長鏈羥基脂肪酸(具有一個或二個羥基的C26或C28)。該裂壺菌屬PKS系統的產物為一種長鏈多不飽和脂肪酸(C22，含5或6個雙鍵，全部為順式)。在二個區段組間之明顯差異為在裂壺菌屬蛋白質中存在二個DH區段，僅有該些區段牽涉到DHA及DPA的順式雙鍵之形成(大概在念珠藻系統中的HetM及HetN與包含羥基有關，及亦包含一DH區段(其起源與在PUFA中發現的那些不同)。同樣地，在裂壺菌屬Orfs B及C中重覆的ER區段之角色並不熟知(該第二ER區段中不顯現其它已特徵化的PUFA PKS系統)。在此二組區段間之胺基酸序列同源物意謂著一種進化關係。可想像該PUFA PKS基因組起源自(以進化觀念)為祖先之念珠藻似的PKS基因組，而併入DH(類FabA)區段。DH區段的加入將造成在新穎的PKS最後產物結構中引進順式雙鍵。

裂壺菌屬與念珠藻PKS區段結構的比較和在裂壺菌屬與志瓦菌PUFA-PKS蛋白質間之區段組織比較，可說明天生與改變區段順序的能力和併入新區段與產生新的最後產物。此外，該些基因現在可於實驗室中操控以產生新穎的產物。從這些觀察可指出應該可以直接或隨機的方法連續地操控該些系統而影響最後產物。例如，在一個較佳的具體實施例中，可設想以PUFA-PKS系統的一個DH(類FabA)

區段來取代不具有異構化活性且會潛在地產生具順及反-
雙鍵混合物分子與DH區段。裂壺菌屬PUFA PKS系統的現
在產物為DHA及DPA(C22:5T6)。若操控該系統來製造C20
5 脂肪酸時，將預計該產物為EPA及ARA(C20:4T6)。此可
提供一種ARA的新來源。亦可取代從相關的會製造不同的
DHA至DPA比率的PUFA-PKS系統來之區段，例如可使用從
破囊壺菌屬23B來的基因(其PUFA PKS系統於本文中第一
次鑑別)。

額外地，可設想特別地改變在裂壺菌屬PUFA PKS系統
10 (其它已描述的PUFA PKS系統至此不具有二個ER區段)中
的一個ER區段(例如移除或失活)，以測量其在最後產物分
佈上的影響。類似的策略已試圖使用更複雜或較不複雜的
方法直接地用於PUFA-PKS蛋白質之每個可區別的區段。當
然並不限制單一區段的操作。最後地，可延伸該方法，以
15 混合從該PUFA-PKS系統及其它PKS或FAS系統來的區段
(例如，型式I、型式II、組合式)，而產生全部範圍的新穎最
後產物。例如，可將PUFA-PKSDH區段引進無法正常地將
順式雙鍵併入最後產物的系統。

此外，本發明包括的方法為基因地改性微生物或植物
20 細胞，其藉由：基因地改性在編碼具有根據本發明之非細
菌PUFA PKS系統的至少一個官能性區段之生物活性的胺
基酸序列之生物中的至少一種核酸序列；及/或表現出至少
一種包含一編碼此胺基酸序列的核酸序列之重組型核酸分
子。此些序列的不同具體實施例、基因地改性一種生物的

方法及特定的改性已在上述有詳細地描述。典型地，該方法可使用來製造一種特別的經基因改性而可製造特別的生物活性分子或許多分子之生物。

本發明的一個具體實施例係關於一種已經改性而可表現出多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統的重組宿主細胞，其中該PKS可催化迭代及非迭代酵素反應二者，及其中該PUFA PKS系統包含：(a)至少二個烯醯ACP-還原酶(ER)區段；(b)至少六個醯基載體蛋白質(ACP)區段；(c)至少二個 β -酮基醯基-ACP合成酶(KS)區段；(d)至少一個轉醯酶(AT)區段；(e)至少一個酮還原酶(KR)區段；(f)至少二個類FabA的 β -羥基醯基-ACP脫水酶(DH)區段；(g)至少一個鏈長因子(CLF)區段；及(h)至少一個丙二酸單醯基-CoA：ACP轉醯酶(MAT)區段。在一個具體實施例中，該PUFA PKS系統為一種真核PUFA PKS系統。在一個較佳的具體實施例中，該PUFA PKS系統為一種藻類的PUFA PKS系統。在更佳的具體實施例中，該PUFA PKS系統為一種破囊壺菌目PUFA PKS系統。此PUFA PKS系統可包括(但是非為限制)一裂壺菌屬PUFA PKS系統，及一破囊壺菌屬 PUFA PKS系統。在一個具體實施例中，該PUFA PKS系統可在一原核宿主細胞中表現。在另一個具體實施例中，該PUFA PKS系統可在一真核宿主細胞中表現。

本發明的另一個具體實施例係關於一種已經改性而可表現出一非細菌的PUFA PKS系統的重組宿主細胞，其中該PKS系統可催化迭代及非迭代酵素反應二者，及其中該非

細菌的PUFA PKS系統包含至少下列的生物活性區段：(a)至少一個烯醯ACP-還原酶(ER)區段；(b)多個醯基載體蛋白質(ACP)區段(至少四個)；(c)至少二個 β -酮基醯基-ACP合成酶(KS)區段；(d)至少一個轉醯酶(AT)區段；(e)至少一個酮還原酶(KR)區段；(f)至少二個類FabA的 β -羥基醯基-ACP脫水酶(DH)區段；(g)至少一個鏈長因子(CLF)區段；及(h)至少一個丙二酸單醯基-CoA：ACP轉醯酶(MAT)區段。

本發明的此具體實施例之一個觀點係關於一種包含至少一種PUFA的產物之製造方法，其包括在有效以製造該產物之條件下生長一種包含任何上述描述的重組宿主細胞之植物，其中該重組宿主細胞為一種植物細胞。本發明之此具體實施例的另一個觀點係關於一種包含至少一種PUFA的產物之製造方法，其包括在有效以製造該產物之條件下培養一種包含任何上述描述的重組宿主細胞之培養基，其中該宿主細胞為一種微生物細胞。在一個較佳的具體實施例中，在宿主細胞中的PKS系統可催化三酸甘油脂的直接製造。

本發明的另一個具體實施例係關於一種包含一種非細菌的多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統之微生物，其中該PKS可催化迭代及非迭代酵素反應二者，及其中該PUFA PKS系統包含：(a)至少二個烯醯ACP-還原酶(ER)區段；(b)至少六個醯基載體蛋白質(ACP)區段；(c)至少二個 β -酮基醯基-ACP合成酶(KS)區段；(d)至少一個轉醯酶(AT)區段；(e)至少一個酮還原酶(KR)區段；(f)至少二個

類FabA的 β -羥基醯基-ACP脫水酶(DH)區段；(g)至少一個鏈長因子(CLF)區段；及(h)至少一個丙二酸單醯基-CoA:ACP轉醯酶(MAT)區段。較佳地，該微生物為一種非細菌的微生物，及更佳地為一種真核微生物。

5 本發明的更另一個具體實施例係關於一種包含一種非細菌的多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統之微生物，其中該PKS可催化迭代及非迭代酵素反應二者，及其中該PUFA PKS系統包含：(a)至少一個烯醯ACP-還原酶(ER)區段；(b)多個醯基載體蛋白質(ACP)區段(至少四
10 個)；(c)至少二個 β -酮基醯基-ACP合成酶(KS)區段；(d)至少一個轉醯酶(AT)區段；(e)至少一個酮還原酶(KR)區段；(f)至少二個類FabA的 β -羥基醯基-ACP脫水酶(DH)區段；(g)至少一個鏈長因子(CLF)區段；及(h)至少一個丙二酸單醯基-CoA:ACP轉醯酶(MAT)區段。

15 在本發明的一個具體實施例中，經考量該致突變計劃可與選擇的篩選方法結合以獲得有興趣的生物活性分子。此將包括搜尋一定範圍的生物活性化合物之方法。此搜尋將不限制於含順式雙鍵的那些分子之製造。該些致突變方法可包括(但是非為限制)：化學致突變、基因改組、編碼特
20 定的酵素區段之基因的開關區域、或限制在那些基因的特定區域之致突變、和其它方法。

 例如，高生產量的致突變方法可使用來影響或最佳化製造想要的生物活性分子。一旦有效的模型系統已發展，則可以高生產量方法改性這些基因。這些技藝之使用可設

想在二個程度上。首先，若可對有興趣的產物(例如，ARA)之製造設計出一種能充分選擇的篩選，則可使用來試圖改變該系統以製造此產物(例如，在場所上，或在與其它策略協調上，諸如上述討論的那些)。額外地，若上述概述的策略可產生一組確定能製造有興趣的產物之基因時，然後可使用高生產量技藝來最佳化該系統。例如，若所引進的區段僅會在相當低的溫度下起作用時，可設計選擇方法以移除該限制。在本發明的一個具體實施例中，篩選方法可使用來鑑別具有類似於裂壺菌屬的PUFA PKS系統之新穎的PKS系統之額外的非細菌生物，如描述於本文(參見上述)。在此生物中鑑別的同源PKS系統可以類似於在本文中所描述之用於裂壺菌屬和用於基因物質的額外來源(從此可產生進一步改性及/或突變PKS系統用以在該微生物中、在另一種微生物中、或在高等植物中表現)的那些方法來使用，以製造多種化合物。

需了解的是許多基因改變(隨機或直接)(其可引進天生的(內生的、天然的)PKS系統)將造成酵素功能的鈍化。本發明一個較佳的具體實施例包括一種系統，其可僅選擇不會阻礙PKS系統製造產物的能力之那些改性。例如，大腸桿菌的FabB-菌株不能合成不飽和脂肪酸，且需要以可代替其正常不飽和脂肪酸的脂肪酸來補充該培養基用以生長(參見美茲等人，2001，前述)。但是，當該菌株以官能性PUFA-PKS系統(即可在大腸桿菌宿主中製造PUFA產物)轉染時可移除此需求(補充培養基)(參見美茲等人，2001，前

述，第2A圖)。該轉染的FabB-菌株現在需要一種官能性 PUFA-PKS系統(以製造不飽和脂肪酸)用於生長而沒有補充。在此實例的關鍵元素為廣泛範圍的不飽和脂肪酸之製造將足夠(甚至不飽和脂肪酸代替品，諸如枝鏈的脂肪酸)。

5 因此，在本發明的另一個較佳具體實施例中，可在本文揭示的一個或多個PUFA PKS基因中產生一大數量的突變，然後轉染該經適當改性的FabB-菌株(例如在一包含ER區段的表現建構中產生突變，及在分離的質體上轉染一具有其它基本區段的FabB-菌株，或併入染色體)，及僅選擇那些可
10 沒有補充培養基而可生長之轉化株(即，其仍然具有製造可補充FabB-缺陷之分子的能力)。可發展額外的篩選以尋找由在此活化PKS系統之選擇的次小組中所製造的特別化合物(例如使用GC用於脂肪酸)。可設想一些類似的選擇性篩選用於有興趣的生物活性分子。

15 如上所述，在本發明的一個具體實施例中，該經基因改性的微生物或植物包括一種微生物或植物，其具有提高的合成想要的生物活性分子(產物)之能力，或其具有新引進的合成特定產物(例如，合成特定的抗生素)之能力。根據本發明，“提高的合成”產物“的能力”指為以與合成產物相關
20 的途徑之任何提高或向上調節，如此該微生物或植物可製造一種增加量的產物(包括先前尚未有之產物的任何製造)，如與在相同條件下培養或生長該野生型微生物或植物比較。製造此基因改性的生物之方法已在上述詳細地描述。

本發明的一個具體實施例為一種製造想要的生物活性

分子(亦指為產物或化合物)之方法，其藉由生長或培養本發明之基因改性的微生物或植物(在上述詳細地描述)。此方法包括的步驟有各別地在發酵培養基中培養或在合適的環境(諸如土壤)中，生長一具有如先前於本文中描述及根據本發明之基因改性的微生物或植物。在一個較佳的具體實施例中，製造本發明之生物活性分子的方法包括之步驟有，在有效以製造該生物活性分子之條件下，培養一種可表現一包含多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統的至少一個生物活性區段之PKS系統的經基因改性之生物。在此較佳的觀點中，該PUFA PKS系統的至少一個區段可由選自於由下列所組成之群的核酸序列譯出：(a)一核酸序列，其編碼從破囊壺菌微生物來之多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統的至少一個區段；(b)一核酸序列，其編碼從藉由本發明的新穎篩選方法所鑑別(上述詳細地描述)之微生物來的PUFA PKS系統之至少一個區段；(c)一核酸序列，其編碼選自於由下列所組成之群的胺基酸序列：序列辨識編號：2、序列辨識編號：4、序列辨識編號：6及其生物活性片段，(d)一核酸序列，其編碼選自於由下列所組成之群的胺基酸序列：序列辨識編號：8、序列辨識編號：10、序列辨識編號：13、序列辨識編號：18、序列辨識編號：20、序列辨識編號：22、序列辨識編號：24、序列辨識編號：26、序列辨識編號：28、序列辨識編號：30、序列辨識編號：32及其生物活性片段；(e)一核酸序列，其編碼一胺基酸序列其與選自於由下列所組成之群的胺基酸序列：

序列辨識編號：2、序列辨識編號：4及序列辨識編號：6的至少500個連續胺基酸有至少約60%相同；其中該胺基酸序列具有PUFA PKS系統的至少一個區段之生物活性；及，(f)一核酸序列，其編碼一與選自於由下列所組成之群的胺基酸序列有至少約60%相同之胺基酸序列：序列辨識編號：8、序列辨識編號：10、序列辨識編號：13、序列辨識編號：18、序列辨識編號：20、序列辨識編號：22、序列辨識編號：24、序列辨識編號：26、序列辨識編號：28、序列辨識編號：30及序列辨識編號：32；其中該胺基酸序列具有PUFA PKS系統的至少一個區段之生物活性。在此方法的較佳觀點中，該生物經基因地改性以影響PKS系統的活性(上述詳細地描述)。可用於與本發明之PUFA PKS系統有關的基因改性之較佳的宿主細胞已於上述描述。

在製造本發明想要的生物活性化合物之方法中，該經基因改性的微生物可在有效以製造該生物活性化合物之條件下在合適的培養基中培養或生長。適當(或有效)的培養基指為任何培養基，當以其培養時本發明之經基因改性的微生物可製造想要的產物。此培養基典型地為一種包含可同化的碳、氮及磷酸鹽來源之水性培養基。此培養基亦可包含適當的鹽類、礦物類、金屬及其它滋養藥。本發明之微生物可培養在習知的發酵生物反應器中。該微生物可利用任何發酵製程培養，其包括(但是非為限制)批次、進料-批次、細胞再循環及連續發酵。用於根據本發明之有潛力的宿主微生物之較佳生長條件已在技藝中熟知。由該經基因

改性的微生物製造之想要的生物活性分子可從發酵培養基使用習知的分離及純化技術回收。例如，該發酵培養基可經過濾或離心以移出微生物、細胞碎物及其它微粒物質，且該產物可從無細胞的上層液利用習知的方法回收，諸如，例如，離子交換、層析、萃取、溶劑萃取、薄膜分離、電滲析、逆滲透、蒸餾、化學衍生作用及結晶。再者，可使用能產生想要的化合物(或其萃取物及不同的片段)之微生物，而沒有從該產物移除該微生物組分。

在製造本發明想要的生物活性化合物之方法中，將經基因改性的植物培養在發酵培養基中或在合適的培養基(諸如土壤)生長。適當(或有效)的發酵培養基已在上述詳細地討論。合適用於高等植物的生長培養基包括任何可用於植物的生長培養基，包括(但是非為限制)土壤、沙、任何其它可支持根生長的微粒培養基(例如蛭石、珍珠岩等等)；或水栽培，和合適的光、水及可最佳化生長高等植物的營養補充品。本發明之經基因改性的植物已設計可經由已根據本發明基因地改性的PKS系統之活性來製造出明顯量想要的產物。該些化合物可經由純化方法(其可從植物萃取化合物)回收。在一個較佳的具體實施例中，該化合物藉由收割該植物回收。在此具體實施例中，該植物可以其天然狀態消耗或進一步加工成可消耗的產物。

如上所述，在一個觀點中，於本發明中有用的基因改性微生物可內生地包含及表現出一PUFA PKS系統，及該基因改性可為該內生的PUFA PKS系統的一個或多個官能性

區段之基因改性，藉此該改性在PUFA PKS系統的活性上具有某些效應。在另一個觀點中，此生物可內生地包含及表現出一PUFA PKS系統，及該基因改性可為引進至少一種外源核酸序列(例如，重組型核酸分子)，其中該外源核酸序列編碼從第二PKS系統來的至少一個生物活性區段或蛋白質及/或一種會影響該PUFA PKS系統的活性之蛋白質(例如，磷酸泛醯巯基乙胺基轉位酶(PPTase)，在下列討論)。在更另一個觀點中，該生物不必內生地(天然地)包含一PUFA PKS系統，但是可基因地改性以引進至少一種編碼具有PUFA PKS系統的至少一個區段之生物活性的胺基酸序列之重組型核酸分子。在此觀點中，PUFA PKS活性會受在生物中引進或增加的PUFA PKS活性影響。與這些觀點每個有關的不同具體實施例已在上述詳細地討論。

在製造生物活性化合物的方法之一個具體實施例中，該基因改性可改變由內生PKS系統製造的至少一種產物，如與野生型生物比較。

在另一個具體實施例中，該生物可內生地表現出一包含PUFA PKS系統的至少一個生物活性區段之PKS系統，及該基因改性包括以一選自於由下列所組成之群的重組型核酸分子轉染該生物：一種重組型核酸分子，其編碼從第二PKS系統來的至少一個生物活性區段；及一種重組型核酸分子，其編碼一種會影響PUFA PKS系統活性之蛋白質。在此具體實施例中，該基因改性較佳地改變由該內生PKS系統製造的至少一種產物，如與野生型生物比較。第二PKS

系統可包括另一種PUFA PKS系統(細菌或非細菌的)、型式I PKS系統、型式II PKS系統及/或組合式PKS系統。會影響PKS系統活性的蛋白質實例已在上述描述(例如, PPTase)。

5 在另一個具體實施例中, 該生物可藉由以編碼多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統的至少一個區段之重組型核酸分子轉染而基因地改性。此重組型核酸分子已先前地於本文中詳細地描述。

在另一個具體實施例中, 該生物內生地表現出一種非細菌的PUFA PKS系統, 及該基因改性包括以從不同的PKS
10 系統來之區段取代編碼非細菌PUFA PKS系統的至少一個區段之核酸序列。在另一個具體實施例中, 該生物可內生地表現出一種非細菌的PUFA PKS系統, 其已藉由以一編碼一能調節由PUFA PKS系統製造的脂肪酸之鏈長度的蛋白質之重組型核酸分子轉染該生物而改性。在一個觀點中,
15 該編碼一能調節脂肪酸的鏈長度的蛋白質之重組型核酸分子可取代一能在非細菌的PUFA PKS系統中譯出鏈長因子之核酸序列。在另一個觀點中, 該可調節由PUFA PKS系統製造的脂肪酸之鏈長度的蛋白質為一種鏈長因子。在另一個觀點中, 該可調節由PUFA PKS系統製造的脂肪酸之鏈長
20 度的蛋白質為一種指導C20單元合成的鏈長因子。

在另一個具體實施例中, 該生物可表現出一種包含在選自於由下列所組成之群的區段中之基因改性的非細菌PUFA PKS系統: 一區段, 其編碼 β -羥基醯基-ACP脫水酶(DH); 及一區段, 其編碼 β -丙脂醯基ACP合成酶(KS), 其

中該改性可改變由PUFA PKS系統製造的長鏈脂肪酸之比率，如與缺乏改性的比較。在此具體實施例的一個觀點中，該改性可選自於由下列所組成之群：刪除該區段的全部或部分、以從不同生物來的同源區段取代該區段、及該區段的突變。

在另一個具體實施例中，該生物可表現出一種包含在烯醯-ACP還原酶(ER)區段中改性的非細菌PUFA PKS系統其，其中該改性可造成製造出不同的化合物，如與缺乏改性的比較。在此具體實施例的一個觀點中，該改性選自於由下列所組成之群：刪除該ER區段的全部或部分、以從不同生物來的ER區段取代該ER區段及該ER區段的突變。

在製造生物活性分子的方法之一個具體實施例中，該生物可製造一種與沒有基因改性的天然發生之生物不同的多不飽和脂肪酸(PUFA)分佈。

許多其它可有用地製造生物活性分子的基因改性將由熟知此技藝、所提供的本公告之人士所明瞭，且不同的其它改性已於本文中先前地討論。本發明預計任何與如於本文描述的PUFA PKS系統有關之基因改性，其可造成製造出想要的生物活性分子。

根據本發明，生物活性分子包括任何具有生物活性的分子(化合物、產物等等)，及其可由包含至少一種具有如於本文描述的非細菌PUFA PKS系統之至少一個官能性區段的生物活性之胺基酸序列的PKS系統製造。此生物活性分子可包括(但是非為限制)：多不飽和脂肪酸(PUFA)、抗炎

性配方、化學治療劑、活性賦形劑、骨質疏鬆藥物、抗抑鬱劑、抗驚厥劑、抗幽門桿菌藥物、用來治療神經變性疾病的藥物、用來治療肝退化疾病的藥物、抗生素及降低膽固醇配方。本發明之非細菌PUFA PKS系統的一個優點為此系統可引進順式碳-碳雙鍵結構的能力，及該分子在每個第三碳處包含一雙鍵。此能力可使用來製造多種化合物。

較佳地，有興趣的生物活性化合物可由該基因改性的微生物製造，其量大於約該微生物的乾燥重量之0.05%，較佳地大於約0.1%，更佳地大於約0.25%，更佳地大於約0.5%，更佳地大於約0.75%，更佳地大於約1%，更佳地大於約2.5%，更佳地大於約5%，更佳地大於約10%，更佳地大於約15%及甚至更佳地大於約20%。對脂質化合物來說，較佳地，此化合物的製造量為大於約該微生物的乾燥重量之5%。對其它生物活性化合物(諸如抗生素或小量合成的化合物)來說，擁有該微生物的乾燥重量之此化合物的那些菌株已鑑別如預測地包含上述描述的新穎PKS系統型式。在某些具體實施例中，特別的生物活性分子(化合物)由該微生物分泌而非累積。因此，此生物活性分子通常地可從培養基回收，且該產生的分子濃度將依微生物及培養基的尺寸而變化。

本發明的一個具體實施例係關於一種包含至少一種脂肪酸的最後產物之改性方法，其包括將一種由一可表現出至少一種包含一編碼PUFA PKS系統的至少一個生物活性區段之核酸序列的重組型核酸分子之重組宿主細胞所製造

的油加入至該最後產物。該PUFA PKS系統可為任何非細菌的PUFA PKS系統，較佳地為選自於下列之群：(a)一核酸序列，其編碼從破囊壺菌微生物來之多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統的至少一個區段；(b)一核酸序列，其編碼從藉由於本文揭示的新穎篩選方法而鑑別之微生物的PUFA PKS系統之至少一個區段；(c)一核酸序列，其編碼選自於由下列所組成之群的胺基酸序列：序列辨識編號：2、序列辨識編號：4、序列辨識編號：6及其生物活性片段；(d)一核酸序列，其編碼選自於由下列所組成之群的胺基酸序列：序列辨識編號：8、序列辨識編號：10、序列辨識編號：13、序列辨識編號：18、序列辨識編號：20、序列辨識編號：22、序列辨識編號：24、序列辨識編號：26、序列辨識編號：28、序列辨識編號：30、序列辨識編號：32及其生物活性片段；(e)一核酸序列，其編碼一胺基酸序列其與選自於由下列所組成之群的胺基酸序列：序列辨識編號：2、序列辨識編號：4及序列辨識編號：6之至少500個連續胺基酸有至少約60%相同；其中該胺基酸序列具有PUFA PKS系統的至少一個區段之生物活性；及，(f)一核酸序列，其編碼一與選自於由下列所組成之群的胺基酸序列有至少約60%相同的胺基酸序列：序列辨識編號：8、序列辨識編號：10、序列辨識編號：13、序列辨識編號：18、序列辨識編號：20、序列辨識編號：22、序列辨識編號：24、序列辨識編號：26、序列辨識編號：28、序列辨識編號：30及序列辨識編號：32；其中該胺基酸序列具有PUFA

PKS系統的至少一個區段之生物活性。這些核酸序列的變異已在上述詳細地描述。

較佳地，該最後產物選自於由下列所組成之群：食物、膳食補充品、醫藥配方、母乳化動物乳及幼兒配方。合適的醫藥配方包括(但是非為限制)抗炎性配方、化學治療劑、5 活性賦形劑、骨質疏鬆藥物、抗抑鬱劑、抗驚厥劑、抗幽門桿菌藥物、用來治療神經變性疾病的藥物、用來治療肝退化疾病的藥物、抗生素及降低膽固醇配方。在一個具體實施例中，該最後產物可使用來治療一種選自於由下列所組成之群的症狀：慢性炎症、急性發炎、胃腸病、癌、惡10 病質、心臟再阻塞、神經變性病、肝退化病症、血脂質失調、骨質疏鬆、骨關節炎、自體免疫疾病、初期子癩、早產、與年齡有關的斑狀丘疹、肺病及過氧化物酶體病。

合適的食物產品包括(但是非為限制)精製的烘烤製品、麵包及捲餅、穀類早餐、加工及未加工的乳酪、佐料(蕃15 茄醬、美乃滋等等)、奶製品(牛奶、酸乳酪(yogurt))、布丁類及明膠甜點類、含二氧化碳的飲料類、茶類、粉末化的飲料混合物、加工的魚產品、水果系列的飲料類、口香糖、硬糕餅、冷凍的奶製品、加工的肉產品、堅果仁及堅果仁20 系列的果醬(spreads)、麵糰、加工的家禽產品、肉汁類及調味汁類、馬鈴薯片及其它碎片或脆片、巧克力及其它糕餅、湯類及湯混合物、大豆系列產物(乳類、飲料類、乳脂狀類、增白劑類)、蔬菜油系列的醬類(spreads)及蔬菜系列飲料類。

本發明的更另一個具體實施例係關於一種製造母乳化

的動物乳之方法。此方法包括的步驟有以包含一編碼PUFA PKS系統的至少一個生物活性區段之核酸序列的至少一種重組型核酸分子，基因地改性該製乳動物的製乳細胞。該

5 該PUFA PKS系統的至少一個區段可由一選自於由下列所組成之群的核酸序列譯出：(a)一核酸序列，其編碼從破囊壺菌微生物來的多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統之至少一個區段；(b)一核酸序列，其編碼從藉由

10 先前於本文描述的新穎篩選方法所鑑別之微生物來的PUFA PKS系統之至少一個區段；(c)一核酸序列，其編碼選自於由下列所組成之群的胺基酸序列：序列辨識編號：2、序列辨識編號：4、序列辨識編號：6及其生物活性片段；

15 (d)一核酸序列，其編碼選自於由下列所組成之群的胺基酸序列：序列辨識編號：8、序列辨識編號：10、序列辨識編號：13、序列辨識編號：18、序列辨識編號：20、序列辨識編號：22、序列辨識編號：24、序列辨識編號：26、序列辨識編號：28、序列辨識編號：30、序列辨識編號：32及其生物活性片段；(e)一核酸序列，其編碼一胺基酸序列

20 其與選自於由下列所組成之群的胺基酸序列：序列辨識編號：2、序列辨識編號：4，及序列辨識編號：6的至少500個連續胺基酸有至少約60%相同；其中該胺基酸序列具有PUFA PKS系統的至少一個區段之生物活性；及/或(f)一核酸序列，其編碼一與選自於由下列所組成之群的胺基酸序列有至少約60%相同的胺基酸序列：序列辨識編號：8、序

列辨識編號：10、序列辨識編號：13、序列辨識編號：18、
 序列辨識編號：20、序列辨識編號：22、序列辨識編號：
 24、序列辨識編號：26、序列辨識編號：28、序列辨識編
 號：30及序列辨識編號：32；其中該胺基酸序列具有PUFA
 5 PKS系統的至少一個區段之生物活性。

基因地改性一宿主細胞及以製造一種經基因改性的非
 人類、製乳動物之方法已在技藝中熟知。可改性的宿主動
 物實例包括犬屬、羊、豬類、山羊類、犛牛等等，其經得
 起對轉殖基因表現群的快速擴張之基因操控及選殖。對動
 10 物來說，可採用PKS似的轉殖基因來在目標胞器、組織及
 體液中透過基因控制區域的改性而表現。特別有興趣的為
 在宿主動物的母乳中製造PUFAs。

【實施方式】

15 較佳實施例之詳細說明

下列實例僅提供闡明之目的而不意欲限制本發明之範
 圍。

實例

實例1

20 下列實例描述進一步分析從裂壺菌屬來的PKS相關序
 列。

本發明家已使用從PCT公告案號WO0042195及美國申
 請序號09/231,899來的、在實例8及9中概述之共通方法，來
 定序出包含在裂壺菌屬PUFA PKS系統中的全部三個開放

性讀碼區(Orfs)之全部長度的基因組DNA。在該裂壺菌屬PKS蛋白質中的生物活性區段在第1圖中圖解地描出。該裂壺菌屬PUFA PKS系統的區段結構更特別地描述於下。

開放性讀碼區A(OrfA)：

5 OrfA的完整核苷酸序列於本文中表示為序列辨識編號：1。OrfA為一8730核苷酸序列(不包括終止密碼子)，其編碼2910個胺基酸序列，其於本文中表示為序列辨識編號：2。在OrfA中有十二個區段：

(a)一個 β -酮基醯基-ACP合成酶(KS)區段；

10 (b)一個丙二酸單醯基-CoA：ACP轉醯酶(MAT)區段；

(c)九個醯基載體蛋白質(ACP)區段；

(d)一個酮還原酶(KR)區段。

包含在OrfA中的區段已根據下列測定：

(1)以Pfam程式產生分析(Pfam為一種蛋白質區段或保存的蛋白質區域之多重排比的資料庫)。該排比顯示出某些具有意含蛋白質的功能之進化的保存結構。從Pfam排比建立之曲線隱藏的馬克夫模型(hidden Markov models)(曲線HMMs)對自動地識別屬於存在的蛋白質家族之新穎的蛋白質(即使為弱的同源)非常有用。不像標準的成對排比方法
15 (例如BLAST、FASTA)，Pfam HMMs明智地處理多區段蛋白質。提供所使用的Pfam版本之參考資料有：貝特曼(Bateman)A、柏尼(Birney)E、西魯提(Cerruti)L、德彬(Durbin)R、愛特威勒(Etwiller)L、愛狄(Eddy)SR、葛利費茲-瓊斯(Griffiths-Jones)S、候威(Howe)KL、馬修爾
20

(Marshall)M、桑哈摩 (Sonnhammer)EL(2002) 核酸搜尋 (Nucleic Acids Research)30(1)：276280)；及/或

(2)與細菌的PUFA-PKS系統(例如，志瓦菌)做同源比較，使用BLAST 2.0基本BLAST同源搜尋，以標準預設參數
5 使用blastp做胺基酸搜尋，其中該查詢序列可利用預設的參數過濾出低複雜性區段(描述在阿次朱耳、S.F.、馬登、T.L.、史卡弗、A.A.、張、J.、張、Z.、米勒、W.&李普門，D.J.(1997)“有間隔的BLAST及PSI-BLAST：新一代蛋白質資料庫搜尋程式。” Nucleic Acids Res. 25：3389-3402，其全文以參考之
10 方式併於本文)。

所提供的用於各別區段之序列相信可包含能譯出一官能性區段的序列之最大長度，及可在Orf中包含額外的嚙合序列。

15 ORFA-KS

在OrfA中的第一區段為KS區段，於本文中亦指為ORFA-KS。此區段包含於跨越在序列辨識編號：1(OrfA)之約位置1及40間的起始點至在序列辨識編號：1之約位置1428及1500間的終點之核苷酸序列中。該包含編碼
20 ORFA-KS區段的序列之核苷酸序列於本文中表示為序列辨識編號：7(序列辨識編號：1之位置1-1500)。該包含KS區段的胺基酸序列跨越在序列辨識編號：2(ORFA)之約位置1及14間的起始點至在序列辨識編號：2之約位置476及500間的終點。該包含ORFA-KS區段的胺基酸序列於本文中表示

為序列辨識編號：8(序列辨識編號：2之位置1-500)。應注意的是該ORFA-KS區段包含一活性位置母題：DXAC*(^{*}醯基黏結位置C₂₁₅)。

ORFA-MAT

5 在OrfA中的第二區段為MAT區段，於本文中亦指為ORFA-MAT。此區段包含於跨越在序列辨識編號：1(OrfA)之約位置1723及1798間的起始點至序列辨識編號：1之約位置2805及3000間的終點之核苷酸序列中。該包含編碼ORFA-MAT區段之序列的核苷酸序列於本文中表示為序列

10 辨識編號：9(序列辨識編號：1之位置1723-3000)。該包含MAT區段之胺基酸序列跨越在序列辨識編號：2(ORFA)之約位置575及600間的起始點至序列辨識編號：2之約位置935及1000間的終點。該包含ORFA-MAT區段的胺基酸序列於本文中表示為序列辨識編號：10(序列辨識編號：2之位置

15 575-1000)。需注意的是該ORFA-MAT區段包含一活性位置母題：GHS*XG(^{*}醯基黏結位置S₇₀₆)，於本文中表示為序列辨識編號：11。

ORFA-ACP#1-9

20 OrfA之區段3-11有九個銜接ACP區段，於本文中亦指為ORFA-ACP(在序列中的第一區段為ORFA-ACP1、第二區段為ORFA-ACP2、第三區段為ORFA-ACP3等等)。第一ACP區段(ORFA-ACP1)包含於跨越從序列辨識編號：1(OrfA)之約位置3343至約位置3600的核苷酸序列中。該包含編碼ORFA-ACP1區段之序列的核苷酸序列於本文中表示為序

列辨識編號：12(序列辨識編號：1之位置3343-3600)。該包含第一ACP區段的胺基酸序列跨越從序列辨識編號：2之約位置1115至約位置1200。該包含ORFA-ACP1區段之胺基酸序列於本文中表示為序列辨識編號：13(序列辨識編號：2
 5 之位置1115-1200)。應注意的是該ORFA-ACP1區段包含一活性位置母題：LGIDS*(**泛醯巯基乙胺黏結母題S₁₁₅₇*)，於本文中由序列辨識編號：14表示。全部九個ACP區段的核苷酸及胺基酸序列已高度地保存，因此，每個區段的序列於本文中不由各別的序列識別符表示。但是，根據於本文
 10 中所揭示的訊息，熟知此技藝之人士可容易地決定出其它八個ACP區段每個的序列。該九個區段的重覆區間為核苷酸序列辨識編號：1的約110至約330。

九個ACP區段全部一起跨越從序列辨識編號：1之約位置3283至約位置6288的OrfA區段，其與序列辨識編號：2
 15 之約1095至約2096的胺基酸位置相符合。此區段包括在各別的ACP區段間之連結子部分。九個ACP區段每個包含一泛醯巯基乙胺黏結母題LGIDS*(於本文中由序列辨識編號：14表示)，其中*為泛醯巯基乙胺黏結位置絲胺酸S。在ACP區段區域的每個末端及在之間，每個ACP區段為一高
 20 度富含脯胺酸(P)及丙胺酸(A)的區域，其相信為一種連結子區段。例如，在ACP區段1及2間為序列：APAPVKAAAPAAPVASAPAPA，於本文中表示為序列辨識編號：15。

ORFA-KR

在 OrfA 中的區段 12 為一 KR 區段，於本文中亦指為 ORFA-KR。此區段包含於跨越序列辨識編號：1 之約位置 6598 的起始點至序列辨識編號：1 之約位置 8730 的終點之核苷酸序列中。該包含編碼 ORFA-KR 區段的序列之核苷酸序列於本文中表示為序列辨識編號：17 (序列辨識編號：1 之位置 6598-8730)。該包含 KR 區段的胺基酸序列跨越序列辨識編號：2 (ORFA) 之約位置 2200 的起始點至序列辨識編號：2 之約位置 2910 的終點。該包含 ORFA-KR 區段之胺基酸序列於本文中表示為序列辨識編號：18 (序列辨識編號：2 之位置 2200-2910)。在此 KR 區段中為一與短鏈醛-脫氫酶 (KR 為此家族成員) 同源之核心區域。此核心區域跨越從序列辨識編號：1 之約位置 7198 至約位置 7500，其與序列辨識編號：2 之胺基酸位置 2400-2500 相符合。

開放性讀碼區 B (OrfB)：

OrfB 的完整核苷酸序列於本文中表示為序列辨識編號：3。OrfB 為一 6177 核苷酸序列 (不包括終止密碼子)，其編碼 2059 個胺基酸序列，於本文中表示為序列辨識編號：4。在 OrfB 中有四個區段：

- (a) 一個 3-酮基醯基-ACP 合成酶 (KS) 區段；
- (b) 一個鏈長因子 (CLF) 區段；
- (c) 一個醯基轉位酶 (AT) 區段；及
- (d) 一個烯醯 ACP-還原酶 (ER) 區段。

該些包含在 OrfB 中的區段已根據下列測定：(1) 以 Pfam 程式產生分析，上述已描述；及/或 (2) 與細菌的 PUFA-PKS

系統(例如，志瓦菌)做同源比較，使用BLAST 2.0基本BLAST同源搜尋，亦已於上述描述。所提供的用於各別區段之序列相信包含編碼一官能性區段的序列之最大長度，及可在Orf中包含額外的嚙合序列。

5 ORFB-KS

在OrfB中的第一區段為KS區段，於本文中亦指為ORFB-KS。此區段包含於跨越在序列辨識編號：3(OrfB)之約位置1及43間的起始點至在序列辨識編號：3之約位置1332及1350間的終點之核苷酸序列中。該包含編碼

10 ORFB-KS區段的序列之核苷酸序列於本文中表示為序列辨識編號：19(序列辨識編號：3之位置1-1350)。該包含KS區段的胺基酸序列跨越在序列辨識編號：4(ORFB)之約位置1及15間的起始點至在序列辨識編號：4之約位置444及450間的終點。該包含ORFB-KS區段的胺基酸序列於本文中

15 表示為序列辨識編號：20(序列辨識編號：4之位置1-450)。應注意的是該ORFB-KS區段包含一活性位置母題：DXAC*(^{*}醯基黏結位置C₁₉₆)。

ORFB-CLF

在OrfB中的第二區段為CLF區段，於本文中亦指為

20 ORFB-CLF。此區段包含於跨越在序列辨識編號：3(OrfB)之約位置1378及1402間的起始點至在序列辨識編號：3之約位置2682及2700間的終點之核苷酸序列中。該包含編碼ORFB-CLF區段之序列的核苷酸序列於本文中表示為序列辨識編號：21(序列辨識編號：3之位置1378-2700)。該包含

CLF區段的胺基酸序列跨越在序列辨識編號：4(ORFB)之約位置460及468間的起始點至在序列辨識編號：4之約位置894及900間的終點。該包含ORFB-CLF區段的胺基酸序列於本文中
 5 表示為序列辨識編號：22(序列辨識編號：4之位置460-900)。應注意的是該ORFB-CLF區段包含一沒有醯基-黏結半胱胺酸的KS活性位置母題。

ORFB-AT

在OrfB中的第三區段為AT區段，於本文中亦指為ORFB-AT。此區段包含於跨越在序列辨識編號：3(OrfB)之約位置2701及3598間的起始點至在序列辨識編號：3之約位置3975及4200間的終點之核苷酸序列中。該包含編碼ORFB-AT區段之序列的核苷酸序列於本文中
 10 表示為序列辨識編號：23(序列辨識編號：3之位置2701-4200)。該包含AT區段的胺基酸序列跨越在序列辨識編號：4(ORFB)之約位置901及1200間的起始點至在序列辨識編號：4之約位置1325及1400間的終點。該包含ORFB-AT區段的胺基酸序列於本文中
 15 表示為序列辨識編號：24(序列辨識編號：4之位置901-1400)。應注意的是該ORFB-AT區段包含一GxS*xG的活性位置母題(*醯基黏結位置S₁₁₄₀)。

20 ORFB-ER

在OrfB中的第四個區段為ER區段，於本文中亦指為ORFB-ER。此區段包含於跨越序列辨識編號：3(OrfB)之約位置4648的起始點至序列辨識編號：3之約位置6177的終點之核苷酸序列中。該包含編碼ORFB-ER區段之序列的核苷

酸序列於本文中表示為序列辨識編號：25(序列辨識編號：3之位置4648-6177)。該包含ER區段的胺基酸序列跨越序列辨識編號：4(ORFB)之約位置1550的起始點至序列辨識編號：4之約位置2059的終點。該包含ORFB-ER區段的胺基酸

5 序列於本文中表示為序列辨識編號：26(序列辨識編號：4之位置1550-2059)。

開放性讀碼區C(OrfC)：

OrfC的完整核苷酸序列於本文中表示為序列辨識編號：5。OrfC為一4509核苷酸序列(不包括終止密碼子)，其

10 編碼1503胺基酸序列，於本文中表示為序列辨識編號：6。在OrfC中有三個區段：

- (a)二個類FabA的 β -羥基醯基-ACP脫水酶(DH)區段；
- (b)一個烯醯ACP-還原酶(ER)區段。

包含在ORFC中之區段已根據下列測定：(1)以Pfam程式產生分析，上述已描述；及/或(2)與細菌的PUFA-PKS系統(例如，志瓦菌)做同源比較，使用BLAST 2.0基本BLAST同源搜尋，亦已於上述描述。所提供的用於各別區段之序列相信包括編碼一官能性區段的序列之最大長度，及可在Orf中包括額外的嚙合序列。

15

20 ORFC-DH1

在OrfC中的第一區段為DH區段，於本文中亦指為ORFC-DH1。此為在OrfC的二個DH區段中之一個，因此標示為DH1。此區段包含於跨越在序列辨識編號：5(OrfC)之約位置1及778間的起始點至在序列辨識編號：5之約位置

1233 及 1350 間的終點之核苷酸序列中。該包含編碼 ORFC-DH1 區段的序列之核苷酸序列於本文中表示為序列辨識編號：27(序列辨識編號：5 之位置 1-1350)。該包含 DH1 區段的胺基酸序列跨越在序列辨識編號：6(ORFC)之約位置 1 及 260 間的起始點至在序列辨識編號：6 之約位置 411 及 450 間的終點。該包含 ORFC-DH1 區段的胺基酸序列於本文中

ORFC-DH2

在 OrfC 中的第二區段為 DH 區段，於本文中亦指為 ORFC-DH2。此為在 OrfC 的二個 DH 區段中之第二個，因此標示為 DH2。此區段包含於跨越在序列辨識編號：5(OrfC) 之約位置 1351 及 2437 間的起始點至在序列辨識編號：5 之約位置 2607 及 2850 間的終點之核苷酸序列中。該包含編碼 ORFC-DH2 區段之序列的核苷酸序列於本文中表示為序列辨識編號：29(序列辨識編號：5 之位置 1351-2850)。該包含 DH2 區段的胺基酸序列跨越在約序列辨識編號：6(ORFC) 之位置 451 及 813 間的起始點至在序列辨識編號：6 之約位置 869 及 950 間的終點。該包含 ORFC-DH2 區段的胺基酸序列於本文中

ORFC-ER

在 OrfC 中的第三區段為 ER 區段，於本文中亦指為 ORFC-ER。此區段包含於跨越序列辨識編號：5(OrfC) 之約位置 2998 的起始點至序列辨識編號：5 之約位置 4509 的終點

之核苷酸序列中。該包含編碼ORFC-ER區段的序列之核苷酸序列於本文中表示為序列辨識編號：31(序列辨識編號：5之位置2998-4509)。該包含ER區段的胺基酸序列跨越序列辨識編號：6(ORFC)之約位置1000的起始點至序列辨識編號：6之約位置1502的終點。該包含ORFC-ER區段的胺基酸序列於本文中表示為序列辨識編號：32(序列辨識編號：6之位置1000-1502)。

實例2

下列實例描述使用本發明之篩選方法來鑑別三種包含根據本發明之PUFA PKS系統的其它非細菌生物。

破囊壺菌23B(ATCC20892)根據描述在美國臨時專利申請序號60/298,796中的篩選方法培養，而於本文中詳細地描述。

已發展用來偵測包含製造PUFA的PKS系統之微生物生理理論篩選(使用燒瓶培養)如下：

將2毫升欲測試的菌株/微生物培養物放置在含50毫升培養基(有氧處理)且裝有擋板的250毫升燒瓶中，及將另一相同菌株的2毫升培養物放置在含200毫升培養基(缺氧處理)且無裝擋板的250毫升燒瓶中。將二燒瓶放置在200rpm的振動台上。在48-72小時的培養時間後，可利用離心機獲得該些培養物及分析細胞的脂肪酸甲基酯類，經由氣相層析法對每個培養物測量下列資料：(1)脂肪酸分佈；(2)PUFA含量；(3)脂肪含量(估計為總脂肪酸(TFA)量)。

然後詢問下列五個問題以分析這些資料：

選擇標準：低O₂/缺氧燒瓶對有氧燒瓶(是/否)

(1)與有氧培養物比較，在低氧培養物中的DHA(或其它PUFA含量)(如為%FAME(脂肪酸甲基酯))是否保持在約相同(或較佳地增加)的程度？

5 (2)在缺氧培養物中之C14：0+C16：0+C16：1是否大於約40%TFA？

(3)在缺氧培養物中是否有習知的氧依賴性鏈加長酶/去飽和酶途徑之非常些微(<1%如為FAME)(或無)的前驅物(C18：3n-3+C18：2n-6+C18：3n-6)？

10 (4)與有氧培養物比較，在低氧培養物中的脂肪含量(如為量總脂肪酸/細胞乾燥重量)是否增加？

(5)與有氧培養物比較，在低氧培養物中的DHA(或其它PUFA含量)是否增加(如為細胞乾燥重量%)？

15 若首先三個問題回答為是，此為菌株包含製造長鏈PUFAs用之PKS基因系統好的跡象。越多的問題回答為是(較佳地首先三個問題必需回答為是)，越強的跡象顯示菌株包含此PKS基因系統。若五個問題全部回答為是，則已有非常強的跡象顯示該菌株包含製造長鏈PUFAs用之PKS基因系統。

20 下列方法概述出上述輪廓，使用冷凍的破囊壺菌23B(ATCC20892)小玻璃瓶來接種(inoculate)包含50毫升RCA培養基的250毫升燒瓶。在25EC下在振動台(200rpm)上搖晃該培養物72小時。RCA培養基包含下列：

RCA培養基

去離子水

1000毫升

離夫結晶(Reef Crystals)®海鹽	40克/升
葡萄糖	20克/升
麩胺酸單鈉(MSG)	20克/升
酵母菌萃取物	1克/升
P11金屬*	5毫升/升
維他命混合物*	1毫升/升
pH	7.0

*P11金屬混合物及維他命混合物與在美國專利案號5,130,742中概述的那些相同，其全文以參考之方式併於本文。

然後，使用25毫升該72小時老的培養物來接種另一個
 5 包含50毫升的低氮RCA培養基(10克/升MSG取代20克/升)
 之250毫升燒瓶，及使用另一25毫升培養來接種包含175毫
 升的低氮RCA培養基之250毫升燒瓶。然後，在25EC下將此
 二燒瓶放置在振動台(200rpm)上72小時。然後可經由離心
 機獲得細胞，且利用冷凍乾燥法乾燥。使用標準氣相層析
 10 程序分析該些乾燥細胞的脂肪含量、脂肪酸分佈及含量(諸
 如在美國5,130,742中概述的那些)。

對破囊壺菌屬23B的篩選結果如下：

DHA(以FAME%)是否增加?	是(38->44%)
C14：0+C16：0+C16：1是否大於約40%TFA?	是(44%)
無C18：3(n-3)或C18：3(n-6)?	是(0%)
脂肪含量是否增加?	是(增加2倍)
DHA(或其它HUFA含量)是否增加?	是(增加2.3倍)

該些結果(特別是DHA含量(如為%FAME)在低氧之條
 15 件下明顯地增加)、條件強烈地顯示製造PUFA的PKS系統存
 在於此破囊壺菌屬的菌株中。

為了提供額外的資料以証實PUFA PKS系統存在，使用

從裂壺菌屬菌株 20888(一種菌株其已經測量包含製造 PUFA 的 PKS 系統(即, 上述描述的序列辨識編號 s: 1-32)) 來的 PKS 探針進行破囊壺菌屬 23B 之南方墨點分析。使用南方墨點技術來偵測破囊壺菌屬 23B 基因組 DNA(其與從 PKS

5 PUFA 合成基因來的雜交探針同源)片段。以 *Cla*I 或 *Kpn*I 限制核酸內切酶消化破囊壺菌屬 23B 基因組 DNA, 利用瓊脂糖凝膠電泳(0.7%瓊脂糖, 在標準三醋酸酯-EDTA 緩衝劑)分離, 及藉由毛細管轉移墨點化至思克雷趣 & 思酷兒-耐川 (Schleicher & Schuell Nytran) 增壓薄膜。使用二個經洋地黃

10 素(digoxigenin)標定的雜交探針-一個特定用於裂壺菌屬 PKS OrfB 之烯醯還原酶(ER)區域(OrfB 之核苷酸 5012-5511; 序列辨識編號: 3), 及另一個特定用於在裂壺菌屬 PKS OrfC 的開始處之保存區域(OrfC 之核苷酸 76-549; 序列辨識編號: 5)。

15 ORFB-ER 探針可在破囊壺菌屬 23B 基因組 DNA 中偵測一種約 13Kb 的 *Cla*I 片段及約 3.6Kb 的 *Kpn*I 片段。OrfC 探針可在破囊壺菌屬 23B 基因組 DNA 中偵測約 7.5kb 的 *Cla*I 片段及約 4.6kb 的 *Kpn*I 片段。

最後地, 重組基因庫(由從嵌入載體 λ FIX II(Stratagene) 的破囊壺菌屬 23B 基因組 DNA 之 DNA 片段組成)可使用經洋地黃素標定而與下列裂壺菌屬 20888 PUFA-PKS 基因部分相符合的探針來篩選: OrfA 之核苷酸 7385-7879(序列辨識編號: 1)、OrfB 之核苷酸 5012-5511(序列辨識編號: 3)及 OrfC 之核苷酸 76-549(序列辨識編號: 5)。這些探針每個可偵測

20

從破囊壺菌屬23B基因庫來的正斑(positive plaques)，此指示出在裂壺菌屬PUFA-PKS基因與破囊壺菌屬23B之基因間有廣大的同源性。

總而言之，這些結果說明其破囊壺菌屬23B基因組DNA
5 包含與從裂壺菌屬20888來的PKS基因同源之序列。

此破囊壺菌微生物包含於本文中而為一種這些基因的額外來源而可使用在上述的具體實施例中。

破囊壺菌屬23B(ATCC20892)在其脂肪酸分佈上明顯地與裂壺菌(ATCC20888)不同。破囊壺菌屬23B具有的
10 DHA：DPA(n-6)比率如14：1一般高，如與在裂壺菌屬(ATCC20888)僅有2-3：1比較。破囊壺菌屬23B亦具有較高程度的C20：5(n-3)。分析在破囊壺菌屬23B的PUFA PKS系統中的區段(與熟知的裂壺菌屬PUFA PKS系統比較)應該可提供我們如何改性這些區段以影響使用這些系統製造的
15 PUFA之比率及型式的關鍵訊息。

上述描述的篩選方法已使用來鑑別其它有潛力的包含PUFA PKS系統的候選菌株。已由本發明家鑑別具有PUFA PKS系統之二種額外的菌株為微小裂壺菌(SR21)光陽(Honda)&尤寇去(Yokochi)(IFO32693)及優肯那菌(BP-5601)。二
20 者可如上述般篩選，但是在N2培養基中(葡萄糖：60克/升； KH_2PO_4 ：4.0克/升；酵母菌萃取物：1.0克/升；玉米漿：1毫升/升； NH_4NO_3 ：1.0克/升；人造海鹽(離夫結晶)：20克/升；上述的濃度全部混合在去離子水中)。對裂壺菌屬及優肯那菌株二者來說，對上述討論的用於破囊壺菌屬23B之篩

選問題的頭先三個回答為是(裂壺菌屬-DHA%FAME32->41%(有氣vs缺氧); 58% 14:0/16:0/16:1; 0%前驅物)及(Ulkenia-DHA%FAME28->44%(有氣vs缺氧); 63% 14:0/16:0/16:1; 0%前驅物), 此指示出這些
 5 菌株為包含PUFA PKS系統的好候選菌株。對每種菌株來說, 最後的二個問題之答案為否: 在*S.微小*中脂肪從乾燥重量的61%減少至乾燥重量的22%, 且DHA從乾燥重量的21-9%; 及在優肯那菌中脂肪從乾燥重量的59%減少至21%, 且DHA從乾燥重量的16%至9%。這些破囊壺菌微生物於本文中亦主張為額外的基因來源而可用於上述的具體
 10 實施例中。

實例3

下列實例說明在裂壺菌屬中的DHA及DPA合成不包括
 15 薄膜-結合的去飽和酶或脂肪酸延長酵素, 如描述用於其它真菌類的那些(帕客-巴能斯等人, 2000, 前述; 仙克寧等人, 1998, 前述)。

裂壺菌屬可累積大量的甘油三酯而富含在DHA及廿二
 20 碳五烯酸(DPA; 22:5T6)中; 例如, 乾燥重量的30%之DHA+DPA。在利用延長/去飽和途徑合成20-及22-碳PUFAs的真菌類中, 18-、20-及22-碳中間體庫相當地大, 所以使用 $[^{14}\text{C}]$ -醋酸酯活體內標定實驗可顯露出清楚的前驅物-產物動力學而用來預測中間體。再者, 以放射性同位素示蹤的而外源地提供至此生物的中間體可轉染至該最後的

PUFA產物。

將[1-¹⁴C]醋酸酯供應至2-天老的培養物作為在零時間處的單一脈衝。然後，可利用離心獲得細胞樣品，且萃取脂質。此外，由該些細胞吸收的[1-¹⁴C]醋酸酯可藉由測量在離心之前及之後樣品的放射性而估計。起源於總細胞脂質的脂肪酸甲基酯類可利用AgNO₃-TLC分離(溶劑，己烷：二乙基醚：醋酸，70：30：2，以體積計)。脂肪酸帶的鑑定已由氣相層析法確認，及在其中的放射性可該閃爍計數來測量。結果顯示出[1-¹⁴C]-醋酸酯可由裂壺菌屬細胞快速地吸收且併入脂肪酸，而且在最短的標定時間(1分鐘)內在脂肪酸中回收包含31%標籤的DHA，且此百分比在[¹⁴C]-醋酸酯併入的10-15分鐘及隨後的24小時培養生長(資料無顯示)期間基本上保持未改變。類似地，遍及實驗DPA表示出10%的標籤。並無證據顯示在16-或18-碳脂肪酸與22-碳多不飽和脂肪酸間有前驅物-產物關係。這些結果與從包含非常小量的(或許有酵素-結合)中間體庫之[¹⁴C]-醋酸酯來的DHA快速合成一致。

其次，細胞在包含2mM的DTT、2mM的EDTA及10%的甘油之100mM磷酸鹽緩衝溶(pH7.2)中，藉由以玻璃微珠渦動而破裂。該無細胞均漿以100,000g離心1小時。將等液份的總均漿、小粒(H-S小粒)及上層液(H-S超級(super))片段培養在25EC下補充有20：M乙醯基-CoA、100：M[1-¹⁴C]丙二酸單醯基-CoA(0.9Gbq/莫耳)、2mM NADH及2mM NADPH的均化緩衝液中60分鐘。萃取分析物，及製備脂肪酸甲基

酯，且在以因勢吞影像器(Instantimager)(博克裝置(Packard Instruments)，美利登(Meriden)，CT)偵測放射性之前如上所述般分離。結果顯示出起源於裂壺菌屬的無細胞均漿培養物會將[1-¹⁴C]-丙二酸單醯基-CoA併入DHA、DPA及飽和脂肪酸(資料無顯示)。相同的生物合成活性由100,000xg上層液片段保留，而非不存在於薄膜小粒中。這些資料可與在細菌酵素分析期間所獲得那些比較(參見前述美茲等人，2001)，其可指示出使用不同的(可溶的)醯基受器分子。因此，在裂壺菌屬中的DHA及DPA合成不包括薄膜-結合的去飽和酶或脂肪酸延長酵素，如描述用於其它真菌類者。

雖然已詳細地描述本發明的不同具體實施例，可明瞭的是那些具體實施例將可由熟知此技藝之人士進行改性及調整。但是，需明確地了解，此改性及調整包含在下列提出之本發明的申請專利範圍中。

15 **【圖式簡單說明】**

第1圖為裂壺菌屬PUFA PKS系統的區段結構圖；

第2圖為從裂壺菌屬及志瓦菌來的PKS區段之比較圖；

第3圖為從裂壺菌屬來之PKS區段與從念珠藻(*Nostoc sp.*)(其產物為一種不包含任何雙鍵的長鏈脂肪酸)來之相關的PKS系統之比較圖。

20 **【主要元件符號說明】**

(無)

序列清單

序列表列 <110> Metz, James Barclay, William Flatt, James
 Kuner, Jerry <120> 多不飽和脂肪酸(PUFA)聚乙醯合成酶系統及其用途 <130>
 2997-29-PCT <150> 60/284,066 <151> 2001-04-16 <150> 60/298,796 <151> 2001-06-15 <150>
 60/323,269 <151> 2001-09-18 <160> 37 <170> PatentIn version 3.1 <210> 1 <211> 8730 <212>
 DNA <213> 裂壺菌屬 sp. <220> <221> CDS <222> (1)..(8730) <223>

<400> 1

atg gcg gcc cgt ctg cag gag caa aag gga ggc gag atg gat acc cgc 48
 Met Ala Ala Arg Leu Gln Glu Gln Lys Gly Gly Glu Met Asp Thr Arg
 1 5 10 15

att gcc atc atc ggc atg tgg gcc atc ctc ccc tgc ggc acg acc gtg 96
 Ile Ala Ile Ile Gly Met Ser Ala Ile Leu Pro Cys Gly Thr Thr Val
 20 25 30

cgc gag tgg ggc gag acc atc cgc gcc ggc atc gac tgc ctg tgg gat 144
 Arg Glu Ser Trp Glu Thr Ile Arg Ala Gly Ile Asp Cys Leu Ser Asp
 35 40 45

ctc ccc gag gac cgc gtc gac gtg acg gcg tac ttt gac ccc gtc aag 192
 Leu Pro Glu Asp Arg Val Asp Val Thr Ala Tyr Phe Asp Pro Val Lys
 50 55 60

acc acc aag gac aag atc tac tgc aag cgc ggt ggc ttc att ccc gag 240
 Thr Thr Lys Asp Lys Ile Tyr Cys Lys Arg Gly Gly Phe Ile Pro Glu
 65 70 75 80

tac gac ttt gac gcc cgc gag ttc gga ctc aac atg ttc cag atg gag 288
 Tyr Asp Phe Asp Ala Arg Glu Phe Gly Leu Asn Met Phe Gln Met Glu
 85 90 95

gac tgg gac gca aac cag acc atc tgg ctt ctc aag gtc aag gag gcc 336
 Asp Ser Asp Ala Asn Gln Thr Ile Ser Leu Leu Lys Val Lys Glu Ala
 100 105 110

ctc cag gac gcc ggc atc gac gcc ctc ggc aag gaa aag aag aac atc 384
 Leu Gln Asp Ala Gly Ile Asp Ala Leu Gly Lys Glu Lys Lys Asn Ile
 115 120 125

ggc tgc gtg ctc ggc att ggc ggc ggc caa aag tcc agc cac gag ttc 432
 Gly Cys Val Leu Gly Ile Gly Gly Gly Gln Lys Ser Ser His Glu Phe
 130 135 140

tac tgg cgc ctt aat tat gtt gtc gtg gag aag gtc ctc cgc aag atg 480
 Tyr Ser Arg Leu Asn Tyr Val Val Val Glu Lys Val Leu Arg Lys Met
 145 150 155 160

ggc atg ccc gag gag gac gtc aag gtc gcc gtc gaa aag tac aag gcc 528
 Gly Met Pro Glu Glu Asp Val Lys Val Ala Val Glu Lys Tyr Lys Ala
 165 170 175

aac ttc ccc gag tgg cgc ctc gac tcc ttc cct ggc ttc ctc ggc aac 576
 Asn Phe Pro Glu Trp Arg Leu Asp Ser Phe Pro Gly Phe Leu Gly Asn
 180 185 190

gtc acc gcc ggt cgc tgc acc aac acc ttc aac ctc gac ggc atg aac 624
 Val Thr Ala Gly Arg Cys Thr Asn Thr Phe Asn Leu Asp Gly Met Asn
 195 200 205

tgc gtt gtc gac gcc gca tgc gcc tcg tcc ctc atc gcc gtc aag gtc 672
 Cys Val Val Asp Ala Ala Cys Ala Ser Ser Leu Ile Ala Val Lys Val
 210 215 220

gcc atc gac gag ctg ctc tac ggt gac tgc gac atg atg gtc acc ggt 720
 Ala Ile Asp Glu Leu Leu Tyr Gly Asp Cys Asp Met Met Val Thr Gly
 225 230 235 240

gcc acc tgc acg gat aac tcc atc ggc atg tac atg gcc ttc tcc aag 768
 Ala Thr Cys Thr Asp Asn Ser Ile Gly Met Tyr Met Ala Phe Ser Lys
 245 250 255

acc ccc gtg ttc tcc acg gac ccc agc gtg cgc gcc tac gac gaa aag 816
 Thr Pro Val Phe Ser Thr Asp Pro Ser Val Arg Ala Tyr Asp Glu Lys
 260 265 270

aca aag ggc atg ctc atc ggc gag ggc tcc gcc atg ctc gtc ctc aag 864
 Thr Lys Gly Met Leu Ile Gly Glu Gly Ser Ala Met Leu Val Leu Lys
 275 280 285

gcg tac gcc gac gcc gtc cgc gac ggc gat gag atc cac gct gtt att 912
 Arg Tyr Ala Asp Ala Val Arg Asp Gly Asp Glu Ile His Ala Val Ile
 290 295 300

gcg ggc tgc gcc tcc tcc agt gat ggc aag gcc gcc ggc atc tac acg 960
 Arg Gly Cys Ala Ser Ser Ser Asp Gly Lys Ala Ala Gly Ile Tyr Thr
 305 310 315 320

ccc acc att tcg ggc cag gag gag gcc ctc cgc cgc gcc tac aac cgc 1008
 Pro Thr Ile Ser Gly Gln Glu Glu Ala Leu Arg Arg Ala Tyr Asn Arg
 325 330 335

gcc tgt gtc gac ccg gcc acc gtc act ctc gtc gag ggt cac gcc acc 1056
 Ala Cys Val Asp Pro Ala Thr Val Thr Leu Val Glu Gly His Gly Thr
 340 345 350

ggt act ccc gtt ggc gac cgc atc gag ctc acc gcc ttg cgc aac ctc 1104
 Gly Thr Pro Val Gly Asp Arg Ile Glu Leu Thr Ala Leu Arg Asn Leu
 355 360 365

ttt gac aag gcc tac ggc gag ggc aac acc gaa aag gtc gct gtg ggc 1152
 Phe Asp Lys Ala Tyr Gly Glu Gly Asn Thr Glu Lys Val Ala Val Gly
 370 375 380

agc atc aag tcc agc atc ggc cat ctc aag gcc gtc gcc ggt ctc gcc 1200
 Ser Ile Lys Ser Ser Ile Gly His Leu Lys Ala Val Ala Gly Leu Ala
 385 390 395 400

ggt atg atc aag gtc atc atg gcg ctc aag cac aag act ctc ccg ggc 1248

Gly Met Ile Lys Val Ile Met Ala Leu Lys His Lys Thr Leu Pro Gly
 405 410 415

acc atc aac gtc gac aac cca ccc aac ctc tac gac aac acg ccc atc 1296
 Thr Ile Asn Val Asp Asn Pro Pro Asn Leu Tyr Asp Asn Thr Pro Ile
 420 425 430

aac gag tcc tcg ctc tac att aac acc atg aac cgc ccc tgg ttc ccg 1344
 Asn Glu Ser Ser Leu Tyr Ile Asn Thr Met Asn Arg Pro Trp Phe Pro
 435 440 445

ccc cct ggt gtg ccc cgc cgc gcc ggc att tcg agc ttt ggc ttt ggt 1392
 Pro Pro Gly Val Pro Arg Arg Ala Gly Ile Ser Ser Phe Gly Phe Gly
 450 455 460

ggc gcc aac tac cac gcc gtc ctc gag gag gcc gag ccc gag cac acg 1440
 Gly Ala Asn Tyr His Ala Val Leu Glu Glu Ala Glu Pro Glu His Thr
 465 470 475 480

acc gcg tac cgc ctc aac aag cgc ccg cag ccc gtg ctc atg atg gcc 1488
 Thr Ala Tyr Arg Leu Asn Lys Arg Pro Gln Pro Val Leu Met Met Ala
 485 490 495

gcc acg ccc gcg gcc ctc cag tcg ctc tgc gag gcc cag ctc aag gag 1536
 Ala Thr Pro Ala Ala Leu Gln Ser Leu Cys Glu Ala Gln Leu Lys Glu
 500 505 510

ttc gag gcc gcc atc aag gag aac gag acc gtc aag aac acc gcc tac 1584
 Phe Glu Ala Ala Ile Lys Glu Asn Glu Thr Val Lys Asn Thr Ala Tyr
 515 520 525

atc aag tgc gtc aag ttc ggc gag cag ttc aaa ttc cct gcc tcc atc 1632
 Ile Lys Cys Val Lys Phe Gly Glu Gln Phe Lys Phe Pro Gly Ser Ile
 530 535 540

ccg gcc aca aac gcg cgc ctc ggc ttc ctc gtc aag gat get gag gat 1680
 Pro Ala Thr Asn Ala Arg Leu Gly Phe Leu Val Lys Asp Ala Glu Asp
 545 550 555 560

gcc tgc tcc acc ctc cgt gcc atc tgc gcc caa ttc gcc aag gat gtc 1728
 Ala Cys Ser Thr Leu Arg Ala Ile Cys Ala Gln Phe Ala Lys Asp Val
 565 570 575

acc aag gag gcc tgg cgc ctc ccc cgc gag gcc gtc agc ttc cgc gcc 1776
 Thr Lys Glu Ala Trp Arg Leu Pro Arg Glu Gly Val Ser Phe Arg Ala
 580 585 590

aag gcc atc gcc acc aac gcc gct gtc gcc gcg ctc ttc tcc gcc cag 1824
 Lys Gly Ile Ala Thr Asn Gly Ala Val Ala Ala Leu Phe Ser Gly Gln
 595 600 605

ggc gcg cag tac acg cac atg ttt agc gag gtg gcc atg aac tgg ccc 1872
 Gly Ala Gln Tyr Thr His Met Phe Ser Glu Val Ala Met Asn Trp Pro
 610 615 620

cag ttc cgc cag agc att gcc gcc atg gac gcc gcc cag tcc aag gtc 1920
 Gln Phe Arg Gln Ser Ile Ala Ala Met Asp Ala Ala Gln Ser Lys Val
 625 630 635 640

gct gga agc gac aag gac ttt gag cgc gtc tcc cag gtc ctc tac ccg 1968
 Ala Gly Ser Asp Lys Asp Phe Glu Arg Val Ser Gln Val Leu Tyr Pro
 645 650 655

cgc aag ccg tac gag cgt gag ccc gag cag aac ccc aag aag atc tcc 2016
 Arg Lys Pro Tyr Glu Arg Glu Pro Glu Gln Asn Pro Lys Lys Ile Ser
 660 665 670

ctc acc gcc tac tcg cag ccc tcg acc ctg gcc tgc gct ctc ggt gcc 2064
 Leu Thr Ala Tyr Ser Gln Pro Ser Thr Leu Ala Cys Ala Leu Gly Ala
 675 680 685

ttt gag atc ttc aag gag gcc ggc ttc acc ccg gac ttt gcc gcc ggc 2112
 Phe Glu Ile Phe Lys Glu Ala Gly Phe Thr Pro Asp Phe Ala Ala Gly
 690 695 700

cat tcg ctc ggt gag ttc gcc gcc ctc tac gcc gcg ggc tgc gtc gac 2160
 His Ser Leu Gly Glu Phe Ala Ala Leu Tyr Ala Ala Gly Cys Val Asp
 705 710 715 720

cgc gac gag ctc ttt gag ctt gtc tgc cgc cgc gcc cgc atc atg ggc 2208
 Arg Asp Glu Leu Phe Glu Leu Val Cys Arg Arg Ala Arg Ile Met Gly
 725 730 735

ggc aag gac gca ccg gcc acc ccc aag gga tgc atg gcc gcc gtc att 2256
 Gly Lys Asp Ala Pro Ala Thr Pro Lys Gly Cys Met Ala Ala Val Ile
 740 745 750

ggc ccc aac gcc gag aac atc aag gtc cag gcc gcc aac gtc tgg ctc 2304
 Gly Pro Asn Ala Glu Asn Ile Lys Val Gln Ala Ala Asn Val Trp Leu
 755 760 765

ggc aac tcc aac tcg cct tcg cag acc gtc atc acc ggc tcc gtc gaa 2352
 Gly Asn Ser Asn Ser Pro Ser Gln Thr Val Ile Thr Gly Ser Val Glu
 770 775 780

ggt atc cag gcc gag agc gcc cgc ctc cag aag gag ggc ttc cgc gtc 2400
 Gly Ile Gln Ala Glu Ser Ala Arg Leu Gln Lys Glu Gly Phe Arg Val
 785 790 795 800

gtg cct ctt gcc tgc gag agc gcc ttc cac tcg ccc cag atg gag aac 2448
 Val Pro Leu Ala Cys Glu Ser Ala Phe His Ser Pro Gln Met Glu Asn
 805 810 815

gcc tcg tcg gcc ttc aag gac gtc atc tcc aag gtc tcc ttc cgc acc 2496
 Ala Ser Ser Ala Phe Lys Asp Val Ile Ser Lys Val Ser Phe Arg Thr
 820 825 830

ccc aag gcc gag acc aag ctc ttc agc aac gtc tct ggc gag acc tac 2544

Pro Lys Ala Glu Thr Lys Leu Phe Ser Asn Val Ser Gly Glu Thr Tyr
835 840 845

ccc acg gac gcc cgc gag atg ctt acg cag cac atg acc agc agc gtc 2592
Pro Thr Asp Ala Arg Glu Met Leu Thr Gln His Met Thr Ser Ser Val
850 855 860

aag ttc ctc acc cag gtc cgc aac atg cac cag gcc ggt gcg cgc atc 2640
Lys Phe Leu Thr Gln Val Arg Asn Met His Gln Ala Gly Ala Arg Ile
865 870 875 880

ttt gtc gag ttc gga ccc aag cag gtg ctc tcc aag ctt gtc tcc gag 2688
Phe Val Glu Phe Gly Pro Lys Gln Val Leu Ser Lys Leu Val Ser Glu
885 890 895

acc ctc aag gat gac ccc tcg gtt gtc acc gtc tct gtc aac ccg gcc 2736
Thr Leu Lys Asp Asp Pro Ser Val Val Thr Val Ser Val Asn Pro Ala
900 905 910

tcg ggc acg gat tcg gac atc cag ctc cgc gac gcg gcc gtc cag ctc 2784
Ser Gly Thr Asp Ser Asp Ile Gln Leu Arg Asp Ala Ala Val Gln Leu
915 920 925

gtt gtc gct ggc gtc aac ctt cag ggc ttt gac aag tgg gac gcc ccc 2832
Val Val Ala Gly Val Asn Leu Gln Gly Phe Asp Lys Trp Asp Ala Pro
930 935 940

gat gcc acc cgc atg cag gcc atc aag aag aag cgc act acc ctc cgc 2880
Asp Ala Thr Arg Met Gln Ala Ile Lys Lys Lys Arg Thr Thr Leu Arg
945 950 955 960

ctt tcg gcc gcc acc tac gtc tcg gac aag acc aag aag gtc cgc gac 2928
Leu Ser Ala Ala Thr Tyr Val Ser Asp Lys Thr Lys Lys Val Arg Asp
965 970 975

gcc gcc atg aac gat ggc cgc tgc gtc acc tac ctc aag ggc gcc gca 2976
Ala Ala Met Asn Asp Gly Arg Cys Val Thr Tyr Leu Lys Gly Ala Ala
980 985 990

ccg ctc atc aag gcc ccg gag ccc gtt gtc gac gag gcc gcc aag cgc 3024
Pro Leu Ile Lys Ala Pro Glu Pro Val Val Asp Glu Ala Ala Lys Arg
995 1000 1005

gag gcc gag cgt ctc cag aag gag ctt cag gat gcc cag cgc cag 3069
Glu Ala Glu Arg Leu Gln Lys Glu Leu Gln Asp Ala Gln Arg Gln
1010 1015 1020

ctc gac gac gcc aag cgc gcc gcc gcc gag gcc aac tcc aag ctc 3114
Leu Asp Asp Ala Lys Arg Ala Ala Ala Glu Ala Asn Ser Lys Leu
1025 1030 1035

gcc gct gcc aag gag gag gcc aag acc gcc gct gct tcg gcc aag 3159
Ala Ala Ala Lys Glu Glu Ala Lys Thr Ala Ala Ala Ser Ala Lys
1040 1045 1050

ccc gca gtt gac act gct gtt gtc gaa aag cat cgt gcc atc ctc 3204
 Pro Ala Val Asp Thr Ala Val Val Glu Lys His Arg Ala Ile Leu
 1055 1060 1065

aag tcc atg ctc gcg gag ctc gat ggc tac gga tcg gtc gac gct 3249
 Lys Ser Met Leu Ala Glu Leu Asp Gly Tyr Gly Ser Val Asp Ala
 1070 1075 1080

tct tcc ctc cag cag cag cag cag cag cag acg gcc ccc gcc ccg 3294
 Ser Ser Leu Gln Gln Gln Gln Gln Gln Thr Ala Pro Ala Pro
 1085 1090 1095

gtc aag gct gct gcg cct gcc gcc ccc gtt gcc tcg gcc cct gcc 3339
 Val Lys Ala Ala Ala Pro Ala Ala Pro Val Ala Ser Ala Pro Ala
 1100 1105 1110

ccg gct gtc tcg aac gag ctt ctt gag aag gcc gag act gtc gtc 3384
 Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu Thr Val Val
 1115 1120 1125

atg gag gtc ctc gcc gcc aag acc ggc tac gag acc gac atg atc 3429
 Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile
 1130 1135 1140

gag gct gac atg gag ctc gag acc gag ctc ggc att gac tcc atc 3474
 Glu Ala Asp Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile
 1145 1150 1155

aag cgt gtc gag atc ctc tcc gag gtc cag gcc atg ctc aat gtc 3519
 Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met Leu Asn Val
 1160 1165 1170

gag gcc aag gat gtc gat gcc ctc agc cgc act cgc act gtt ggt 3564
 Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr Arg Thr Val Gly
 1175 1180 1185

gag gtt gtc aac gcc atg aag gcc gag atc gct ggc agc tct gcc 3609
 Glu Val Val Asn Ala Met Lys Ala Glu Ile Ala Gly Ser Ser Ala
 1190 1195 1200

ccg gcg cct gct gcc gct gct ccg gct ccg gcc aag gct gcc cct 3654
 Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Lys Ala Ala Pro
 1205 1210 1215

gcc gcc gct gcg cct gct gtc tcg aac gag ctt ctc gag aag gcc 3699
 Ala Ala Ala Ala Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala
 1220 1225 1230

gag acc gtc gtc atg gag gtc ctc gcc gcc aag act ggc tac gag 3744
 Gly Thr Val Val Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu
 1235 1240 1245

act gac atg atc gag tcc gac atg gag ctc gag act gag ctc gcc 3789

Thr Asp Met Ile Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly
 1250 1255 1260

 att gac tcc atc aag cgt gtc gag atc ctc tcc gag gtt cag gcc 3834
 Ile Asp Ser Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala
 1265 1270 1275

 atg ctc aac gtc gag gcc aag gac gtc gac gct ctc agc cgc act 3879
 Met Leu Asn Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr
 1280 1285 1290

 cgc act gtg ggt gag gtc gtc aac gcc atg aag gct gag atc gct 3924
 Arg Thr Val Gly Glu Val Val Asn Ala Met Lys Ala Glu Ile Ala
 1295 1300 1305

 ggt ggc tct gcc ccg gcg cct gcc gcc gct gcc cca ggt ccg gct 3969
 Gly Gly Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro Gly Pro Ala
 1310 1315 1320

 gct gcc gcc cct gcg cct gcc gcc gcc gcc cct gct gtc tcg aac 4014
 Ala Ala Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Val Ser Asn
 1325 1330 1335

 gag ctt ctt gag aag gcc gag acc gtc gtc atg gag gtc ctc gcc 4059
 Glu Leu Leu Glu Lys Ala Glu Thr Val Val Met Glu Val Leu Ala
 1340 1345 1350

 gcc aag act ggc tac gag act gac atg atc gag tcc gac atg gag 4104
 Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp Met Glu
 1355 1360 1365

 ctc gag acc gag ctc ggc att gac tcc atc aag cgt gtc gag att 4149
 Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu Ile
 1370 1375 1380

 ctc tcc gag gtc cag gcc atg ctc aac gtc gag gcc aag gac gtc 4194
 Leu Ser Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val
 1385 1390 1395

 gac gct ctc agc cgc acc cgc act gtt ggc gag gtc gtc gat gcc 4239
 Asp Ala Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val Asp Ala
 1400 1405 1410

 atg aag gcc gag atc gct ggt ggc tct gcc ccg gcg cct gcc gcc 4284
 Met Lys Ala Glu Ile Ala Gly Gly Ser Ala Pro Ala Pro Ala Ala
 1415 1420 1425

 gct gct cct gct ccg gct gct gcc gcc cct gcg cct gcc gcc cct 4329
 Ala Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Ala Pro
 1430 1435 1440

 ggc cct gct gtc tcg agc gag ctt ctc gag aag gcc gag act gtc 4374
 Ala Pro Ala Val Ser Ser Glu Leu Leu Glu Lys Ala Glu Thr Val
 1445 1450 1455

gtc atg gag gtc ctc gcc gcc aag act ggc tac gag act gac atg 4419
 Val Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met
 1460 1465 1470

atc gag tcc gac atg gag ctc gag acc gag ctc ggc att gac tcc 4464
 Ile Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser
 1475 1480 1485

atc aag cgt gtc gag att ctc tcc gag gtc cag gcc atg ctc aac 4509
 Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met Leu Asn
 1490 1495 1500

gtc gag gcc aag gac gtc gac gct ctc agc cgc acc cgc act gtt 4554
 Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr Arg Thr Val
 1505 1510 1515

ggc gag gtc gtc gat gcc atg aag gcc gag atc gct ggt ggc tct 4599
 Gly Glu Val Val Asp Ala Met Lys Ala Glu Ile Ala Gly Gly Ser
 1520 1525 1530

gcc ccg gcg cct gcc gcc gct gct cct gct ccg gct gct gcc gcc 4644
 Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Ala Ala Ala
 1535 1540 1545

cct gcg cct gcc gcc cct gcg cct gcc gcc cct gcg cct gct gtc 4689
 Pro Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Pro Ala Val
 1550 1555 1560

tcg agc gag ctt ctc gag aag gcc gag act gtc gtc atg gag gtc 4734
 Ser Ser Glu Leu Leu Glu Lys Ala Glu Thr Val Val Met Glu Val
 1565 1570 1575

ctc gcc gcc aag act ggc tac gag act gac atg att gag tcc gac 4779
 Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp
 1580 1585 1590

atg gag ctc gag acc gag ctc ggc att gac tcc atc aag cgt gtc 4824
 Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg Val
 1595 1600 1605

gag att ctc tcc gag gtt cag gcc atg ctc aac gtc gag gcc aag 4869
 Glu Ile Leu Ser Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys
 1610 1615 1620

gac gtc gac gct ctc agc cgc act cgc act gtt ggt gag gtc gtc 4914
 Asp Val Asp Ala Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val
 1625 1630 1635

gat gcc atg aag gct gag atc gct ggc agc tcc gcc tcg gcg cct 4959
 Asp Ala Met Lys Ala Glu Ile Ala Gly Ser Ser Ala Ser Ala Pro
 1640 1645 1650

gcc gcc gct gct cct gct ccg gct gct gcc gct cct gcg ccc gct 5004

Ala Ala Ala Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala
1655 1660 1665

gcc gcc gcc cct gct gtc tcg aac gag ctt ctc gag aaa gcc gag 5049
Ala Ala Ala Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu
1670 1675 1680

act gtc gtc atg gag gtc ctc gcc gcc aag act ggc tac gag act 5094
Thr Val Val Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr
1685 1690 1695

gac atg atc gag tcc gac atg gag ctc gag act gag ctc gcc att 5139
Asp Met Ile Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly Ile
1700 1705 1710

gac tcc atc aag cgt gtc gag atc ctc tcc gag gtt cag gcc atg 5184
Asp Ser Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met
1715 1720 1725

ctc aac gtc gag gcc aag gac gtc gat gcc ctc agc cgc acc cgc 5229
Leu Asn Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr Arg
1730 1735 1740

act gtt ggc gag gtt gtc gat gcc atg aag gcc gag atc gct ggt 5274
Thr Val Gly Glu Val Val Asp Ala Met Lys Ala Glu Ile Ala Gly
1745 1750 1755

ggc tct gcc ccg gcg cct gcc gcc gct gcc cct gct ccg gct gcc 5319
Gly Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Ala
1760 1765 1770

gcc gcc cct gct gtc tcg aac gag ctt ctc gag aag gcc gag act 5364
Ala Ala Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu Thr
1775 1780 1785

gtc gtc atg gag gtc ctc gcc gcc aag act ggc tac gag acc gac 5409
Val Val Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp
1790 1795 1800

atg atc gag tcc gac atg gag ctc gag acc gag ctc gcc att gac 5454
Met Ile Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly Ile Asp
1805 1810 1815

tcc atc aag cgt gtc gag att ctc tcc gag gtt cag gcc atg ctc 5499
Ser Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met Leu
1820 1825 1830

aac gtc gag gcc aag gac gtc gat gct ctc agc cgc act cgc act 5544
Asn Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr Arg Thr
1835 1840 1845

gtt ggc gag gtc gtc gat gcc atg aag gct gag atc gcc gcc agc 5589
Val Gly Glu Val Val Asp Ala Met Lys Ala Glu Ile Ala Gly Ser
1850 1855 1860

tcg gcc ccg gcg cct gcc gcc gct gct cct gct ccg gct gct gcc 5634
 Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Ala Ala
 1865 1870 1875

gct cct gcg ccc gct gcc gct gcc cct gct gtc tcg agc gag ctt 5679
 Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Val Ser Ser Glu Leu
 1880 1885 1890

ctc gag aag gcc gag acc gtc gtc atg gag gtc ctc gcc gcc aag 5724
 Leu Glu Lys Ala Glu Thr Val Val Met Glu Val Leu Ala Ala Lys
 1895 1900 1905

act gcc tac gag act gac atg att gag tcc gac atg gag ctc gag 5769
 Thr Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp Met Glu Leu Glu
 1910 1915 1920

act gag ctc gcc att gac tcc atc aag cgt gtc gag atc ctc tcc 5814
 Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu Ile Leu Ser
 1925 1930 1935

gag gtt cag gcc atg ctc aac gtc gag gcc aag gac gtc gat gcc 5859
 Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val Asp Ala
 1940 1945 1950

ctc agc cgc acc cgc act gtt ggc gag gtt gtc gat gcc atg aag 5904
 Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val Asp Ala Met Lys
 1955 1960 1965

gcc gag atc gct ggt ggc tct gcc ccg gcg cct gcc gcc gct gcc 5949
 Ala Glu Ile Ala Gly Gly Ser Ala Pro Ala Pro Ala Ala Ala Ala
 1970 1975 1980

cct gct ccg get gcc gcc gcc cct gct gtc tcg aac gag ctt ctt 5994
 Pro Ala Pro Ala Ala Ala Ala Pro Ala Val Ser Asn Glu Leu Leu
 1985 1990 1995

gag aag gcc gag acc gtc gtc atg gag gtc ctc gcc gcc aag act 6039
 Glu Lys Ala Glu Thr Val Val Met Glu Val Leu Ala Ala Lys Thr
 2000 2005 2010

ggc tac gag acc gac atg atc gag tcc gac atg gag ctc gag acc 6084
 Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp Met Glu Leu Glu Thr
 2015 2020 2025

gag ctc gcc att gac tcc atc aag cgt gtc gag att ctc tcc gag 6129
 Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu Ile Leu Ser Glu
 2030 2035 2040

gtt cag gcc atg ctc aac gtc gag gcc aag gac gtc gac gct ctc 6174
 Val Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val Asp Ala Leu
 2045 2050 2055

agc cgc act cgc act gtt ggc gag gtc gtc gat gcc atg aag gct 6219

Ser Arg Thr Arg Thr Val Gly Glu Val Val Asp Ala Met Lys Ala
2060 2065 2070

gag atc gct ggt ggc tct gcc ccg gcg cct gcc gcc gct gct cct 6264
Glu Ile Ala Gly Gly Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro
2075 2080 2085

gcc tcg gct ggc gcc gcg cct gcg gtc aag att gac tcg gtc cac 6309
Ala Ser Ala Gly Ala Ala Pro Ala Val Lys Ile Asp Ser Val His
2090 2095 2100

ggc gct gac tgt gat gat ctt tcc ctg atg cac gcc aag gtg gtt 6354
Gly Ala Asp Cys Asp Asp Leu Ser Leu Met His Ala Lys Val Val
2105 2110 2115

gac atc cgc cgc ccg gac gag ctg atc ctg gag cgc ccc gag aac 6399
Asp Ile Arg Arg Pro Asp Glu Leu Ile Leu Glu Arg Pro Glu Asn
2120 2125 2130

ccc gcc gtt ctg gtt gtc gat gac ggc agc gag ctg acc ctg gcc 6444
Arg Pro Val Leu Val Val Asp Asp Gly Ser Glu Leu Thr Leu Ala
2135 2140 2145

ctg gtc cgc gtc ctg ggc gcc tgc gcc gtt gtc ctg acc ttt gag 6489
Leu Val Arg Val Leu Gly Ala Cys Ala Val Val Leu Thr Phe Glu
2150 2155 2160

ggt ctg cag ctg gct cag cgc gct ggt gcc gct gcc atc cgc cac 6534
Gly Leu Gln Leu Ala Gln Arg Ala Gly Ala Ala Ala Ile Arg His
2165 2170 2175

gtg ctg gcc aag gat ctt tcc gcg gag agc gcc gag aag gcc atc 6579
Val Leu Ala Lys Asp Leu Ser Ala Glu Ser Ala Glu Lys Ala Ile
2180 2185 2190

aag gag gcc gag cag cgc ttt ggc gct ctg ggc ggc ttc atc tcg 6624
Lys Glu Ala Glu Gln Arg Phe Gly Ala Leu Gly Gly Phe Ile Ser
2195 2200 2205

cag cag gcg gag cgc ttc gag ccc gcc gaa atc ctg ggc ttc acg 6669
Gln Gln Ala Glu Arg Phe Glu Pro Ala Glu Ile Leu Gly Phe Thr
2210 2215 2220

ctc atg tgc gcc aag ttc gcc aag gct tcc ctg tgc acg gct gtg 6714
Leu Met Cys Ala Lys Phe Ala Lys Ala Ser Leu Cys Thr Ala Val
2225 2230 2235

gct ggc ggc cgc ccg gcc ttt atc ggt gtg gcg cgc ctt gac ggc 6759
Ala Gly Gly Arg Pro Ala Phe Ile Gly Val Ala Arg Leu Asp Gly
2240 2245 2250

gca ctg gga ttc act tcg cag ggc act tct gac gcg ctg aag cgt 6804
Arg Leu Gly Phe Thr Ser Gln Gly Thr Ser Asp Ala Leu Lys Arg
2255 2260 2265

gcc cag cgt ggt gcc atc ttt ggc ctc tgc aag acc atc ggc ctc 6849
 Ala Gln Arg Gly Ala Ile Phe Gly Leu Cys Lys Thr Ile Gly Leu
 2270 2275 2280

gag tgg tcc gag tct gac gtc ttt tcc cgc ggc gtg gac att gct 6894
 Glu Trp Ser Glu Ser Asp Val Phe Ser Arg Gly Val Asp Ile Ala
 2285 2290 2295

cag ggc atg cac ccc gag gat gcc gcc gtg cgc att gtg cgc gag 6939
 Gln Gly Met His Pro Glu Asp Ala Ala Val Ala Ile Val Arg Glu
 2300 2305 2310

atg cgc tgc gct gac att cgc att cgc gag gtc ggc att ggc gca 6984
 Met Ala Cys Ala Asp Ile Arg Ile Arg Glu Val Gly Ile Gly Ala
 2315 2320 2325

aac cag cag cgc tgc acg atc cgt gcc gcc aag ctc gag acc ggc 7029
 Asn Gln Gln Arg Cys Thr Ile Arg Ala Ala Lys Leu Glu Thr Gly
 2330 2335 2340

aac ccg cag cgc cag atc gcc aag gac gac gtg ctg ctc gtt tct 7074
 Asn Pro Gln Arg Gln Ile Ala Lys Asp Asp Val Leu Leu Val Ser
 2345 2350 2355

ggc ggc gct cgc ggc atc acg cct ctt tgc atc cgg gag atc acg 7119
 Gly Gly Ala Arg Gly Ile Thr Pro Leu Cys Ile Arg Glu Ile Thr
 2360 2365 2370

cgc cag atc gcg ggc ggc aag tac att ctg ctt ggc cgc agc aag 7164
 Arg Gln Ile Ala Gly Gly Lys Tyr Ile Leu Leu Gly Arg Ser Lys
 2375 2380 2385

gtc tct gcg agc gaa ccg gca tgg tgc gct ggc atc act gac gag 7209
 Val Ser Ala Ser Glu Pro Ala Trp Cys Ala Gly Ile Thr Asp Glu
 2390 2395 2400

aag gct gtg caa aag gct gct acc cag gag ctc aag cgc gcc ttt 7254
 Lys Ala Val Gln Lys Ala Ala Thr Gln Glu Leu Lys Arg Ala Phe
 2405 2410 2415

agc gct ggc gag ggc ccc aag ccc acg ccc cgc gct gtc act aag 7299
 Ser Ala Gly Glu Gly Pro Lys Pro Thr Pro Arg Ala Val Thr Lys
 2420 2425 2430

ctt gtg ggc tct gtt ctt ggc gct cgc gag gtg cgc agc tct att 7344
 Leu Val Gly Ser Val Leu Gly Ala Arg Glu Val Arg Ser Ser Ile
 2435 2440 2445

gct cgc att gaa gcg ctc ggc ggc aag gcc atc tac tcg tcg tgc 7389
 Ala Ala Ile Glu Ala Leu Gly Gly Lys Ala Ile Tyr Ser Ser Cys
 2450 2455 2460

gac gtg aac tct gcc gcc gac gtg gcc aag gcc gtg cgc gat gcc 7434

Asp Val Asn Ser Ala Ala Asp Val Ala Lys Ala Val Arg Asp Ala
 2465 2470 2475

gag tcc cag ctc ggt gcc cgc gtc tcc ggc atc gtt cat gcc tcc 7479
 Glu Ser Gln Leu Gly Ala Arg Val Ser Gly Ile Val His Ala Ser
 2480 2485 2490

ggc gtg ctc cgc gac cgt ctc atc gag aag aag ctc ccc gac gag 7524
 Gly Val Leu Arg Asp Arg Leu Ile Glu Lys Lys Leu Pro Asp Glu
 2495 2500 2505

ttc gac gcc gtc ttt ggc acc aag gtc acc ggt ctc gag aac ctc 7569
 Phe Asp Ala Val Phe Gly Thr Lys Val Thr Gly Leu Glu Asn Leu
 2510 2515 2520

ctc gcc gcc gtc gac cgc gcc aac ctc aag cac atg gtc ctc ttc 7614
 Leu Ala Ala Val Asp Arg Ala Asn Leu Lys His Met Val Leu Phe
 2525 2530 2535

agc tcc ctc gcc ggc ttc cac ggc aac gtc ggc cag tct gac tac 7659
 Ser Ser Leu Ala Gly Phe His Gly Asn Val Gly Gln Ser Asp Tyr
 2540 2545 2550

gcc atg gcc aac gag gcc ctt aac aag atg ggc ctc gag ctc gcc 7704
 Ala Met Ala Asn Glu Ala Leu Asn Lys Met Gly Leu Glu Leu Ala
 2555 2560 2565

aag gac gtc tcc gtc aag tcc atc tgc ttc ggt ccc tgg gac ggt 7749
 Lys Asp Val Ser Val Lys Ser Ile Cys Phe Gly Pro Trp Asp Gly
 2570 2575 2580

ggc atg gtg acg ccc cag ctc aag aag cag ttc cag gag atg ggc 7794
 Gly Met Val Thr Pro Gln Leu Lys Lys Gln Phe Gln Glu Met Gly
 2585 2590 2595

gtg cag atc atc ccc cgc gag ggc ggc gct gat acc gtg cgc cgc 7839
 Val Gln Ile Ile Pro Arg Glu Gly Gly Ala Asp Thr Val Ala Arg
 2600 2605 2610

atc gtg ctc gcc tcc tcc ccc gct gag atc ctt gtc ggc aac tgg 7884
 Ile Val Leu Gly Ser Ser Pro Ala Glu Ile Leu Val Gly Asn Trp
 2615 2620 2625

cgc acc ccc tcc aag aag gtc ggc tcc gac acc atc acc ctg cac 7929
 Arg Thr Pro Ser Lys Lys Val Gly Ser Asp Thr Ile Thr Leu His
 2630 2635 2640

cgc aag att tcc gcc aag tcc aac ccc ttc ctc gag gac cac gtc 7974
 Arg Lys Ile Ser Ala Lys Ser Asn Pro Phe Leu Glu Asp His Val
 2645 2650 2655

atc cag ggc cgc cgc gtg ctg ccc atg acg ctg gcc atc ggc tcc 8019
 Ile Gln Gly Arg Arg Val Leu Pro Met Thr Leu Ala Ile Gly Ser
 2660 2665 2670

ctc gcg gag acc tgc ctc ggc ctc ttc ccc ggc tac tcg ctc tgg 8064
 Leu Ala Glu Thr Cys Leu Gly Leu Phe Pro Gly Tyr Ser Leu Trp
 2675 2680 2685

gcc att gac gac gcc cag ctc ttc aag ggt gtc act gtc gac ggc 8109
 Ala Ile Asp Asp Ala Gln Leu Phe Lys Gly Val Thr Val Asp Gly
 2690 2695 2700

gac gtc aac tgc gag gtg acc ctc acc ccg tcg acg gcg ccc tcg 8154
 Asp Val Asn Cys Glu Val Thr Leu Thr Pro Ser Thr Ala Pro Ser
 2705 2710 2715

ggc cgc gtc aac gtc cag gcc acg ctc aag acc ttt tcc agc ggc 8199
 Gly Arg Val Asn Val Gln Ala Thr Leu Lys Thr Phe Ser Ser Gly
 2720 2725 2730

aag ctg gtc ccg gcc tac cgc gcc gtc atc gtg ctc tcc aac cag 8244
 Lys Leu Val Pro Ala Tyr Arg Ala Val Ile Val Leu Ser Asn Gln
 2735 2740 2745

ggc gcg ccc ccg gcc aac gcc acc atg cag ccg ccc tcg ctc gat 8289
 Gly Ala Pro Pro Ala Asn Ala Thr Met Gln Pro Pro Ser Leu Asp
 2750 2755 2760

gcc gat ccg gcg ctc cag gcc tcc gtc tac gac ggc aag acc ctc 8334
 Ala Asp Pro Ala Leu Gln Gly Ser Val Tyr Asp Gly Lys Thr Leu
 2765 2770 2775

ttc cac ggc ccg gcc ttc cgc ggc atc gat gac gtg ctc tcg tgc 8379
 Phe His Gly Pro Ala Phe Arg Gly Ile Asp Asp Val Leu Ser Cys
 2780 2785 2790

acc aag agc cag ctt gtg gcc aag tgc agc gct gtc ccc ggc tcc 8424
 Thr Lys Ser Gln Leu Val Ala Lys Cys Ser Ala Val Pro Gly Ser
 2795 2800 2805

gac gcc gct cgc ggc gag ttt gcc acg gac act gac gcc cat gac 8469
 Asp Ala Ala Arg Gly Glu Phe Ala Thr Asp Thr Asp Ala His Asp
 2810 2815 2820

ccc ttc gtg aac gac ctg gcc ttt cag gcc atg ctc gtc tgg gtg 8514
 Pro Phe Val Asn Asp Leu Ala Phe Gln Ala Met Leu Val Trp Val
 2825 2830 2835

cgc cgc acg ctc ggc cag gct gcg ctc ccc aac tcg atc cag cgc 8559
 Arg Arg Thr Leu Gly Gln Ala Ala Leu Pro Asn Ser Ile Gln Arg
 2840 2845 2850

atc gtc cag cac cgc ccg gtc ccg cag gac aag ccc ttc tac att 8604
 Ile Val Gln His Arg Pro Val Pro Gln Asp Lys Pro Phe Tyr Ile
 2855 2860 2865

acc ctc cgc tcc aac cag tcg ggc ggt cac tcc cag cac aag cac 8649

Thr Leu Arg Ser Asn Gln Ser Gly Gly His Ser Gln His Lys His
 2870 2875 2880

gcc ctt cag ttc cac aac gag cag ggc gat ctc ttc att gat gtc 8694
 Ala Leu Gln Phe His Asn Glu Gln Gly Asp Leu Phe Ile Asp Val
 2885 2890 2895

cag gct tcg gtc atc gcc acg gac agc ctt gcc ttc 8730
 Gln Ala Ser Val Ile Ala Thr Asp Ser Leu Ala Phe
 2900 2905 2910

<210> 2 <211> 2910 <212> PRT <213> 裂壺菌屬 sp. <400> 2

Met Ala Ala Arg Leu Gln Glu Gln Lys Gly Gly Glu Met Asp Thr Arg
 1 5 10 15

Ile Ala Ile Ile Gly Met Ser Ala Ile Leu Pro Cys Gly Thr Thr Val
 20 25 30

Arg Glu Ser Trp Glu Thr Ile Arg Ala Gly Ile Asp Cys Leu Ser Asp
 35 40 45

Leu Pro Glu Asp Arg Val Asp Val Thr Ala Tyr Phe Asp Pro Val Lys
 50 55 60

Thr Thr Lys Asp Lys Ile Tyr Cys Lys Arg Gly Gly Phe Ile Pro Glu
 65 70 75 80

Tyr Asp Phe Asp Ala Arg Glu Phe Gly Leu Asn Met Phe Gln Met Glu
 85 90 95

Asp Ser Asp Ala Asn Gln Thr Ile Ser Leu Leu Lys Val Lys Glu Ala
 100 105 110

Leu Gln Asp Ala Gly Ile Asp Ala Leu Gly Lys Glu Lys Lys Asn Ile
 115 120 125

Gly Cys Val Leu Gly Ile Gly Gly Gly Gln Lys Ser Ser His Glu Phe
 130 135 140

Tyr Ser Arg Leu Asn Tyr Val Val Val Glu Lys Val Leu Arg Lys Met
 145 150 155 160

Gly Met Pro Glu Glu Asp Val Lys Val Ala Val Glu Lys Tyr Lys Ala
 165 170 175

Asn Phe Pro Glu Trp Arg Leu Asp Ser Phe Pro Gly Phe Leu Gly Asn
 180 185 190

Val Thr Ala Gly Arg Cys Thr Asn Thr Phe Asn Leu Asp Gly Met Asn
 195 200 205

Cys Val Val Asp Ala Ala Cys Ala Ser Ser Leu Ile Ala Val Lys Val
 210 215 220

Ala Ile Asp Glu Leu Leu Tyr Gly Asp Cys Asp Met Met Val Thr Gly
 225 230 235 240

Ala Thr Cys Thr Asp Asn Ser Ile Gly Met Tyr Met Ala Phe Ser Lys
 245 250 255

Thr Pro Val Phe Ser Thr Asp Pro Ser Val Arg Ala Tyr Asp Glu Lys
 260 265 270

Thr Lys Gly Met Leu Ile Gly Glu Gly Ser Ala Met Leu Val Leu Lys
 275 280 285

Arg Tyr Ala Asp Ala Val Arg Asp Gly Asp Glu Ile His Ala Val Ile
 290 295 300

Arg Gly Cys Ala Ser Ser Ser Asp Gly Lys Ala Ala Gly Ile Tyr Thr
 305 310 315 320

Pro Thr Ile Ser Gly Gln Glu Glu Ala Leu Arg Arg Ala Tyr Asn Arg
 325 330 335

Ala Cys Val Asp Pro Ala Thr Val Thr Leu Val Glu Gly His Gly Thr
 340 345 350

Gly Thr Pro Val Gly Asp Arg Ile Glu Leu Thr Ala Leu Arg Asn Leu
 355 360 365

Phe Asp Lys Ala Tyr Gly Glu Gly Asn Thr Glu Lys Val Ala Val Gly
 370 375 380

Ser Ile Lys Ser Ser Ile Gly His Leu Lys Ala Val Ala Gly Leu Ala
385 390 395 400

Gly Met Ile Lys Val Ile Met Ala Leu Lys His Lys Thr Leu Pro Gly
405 410 415

Thr Ile Asn Val Asp Asn Pro Pro Asn Leu Tyr Asp Asn Thr Pro Ile
420 425 430

Asn Glu Ser Ser Leu Tyr Ile Asn Thr Met Asn Arg Pro Trp Phe Pro
435 440 445

Pro Pro Gly Val Pro Arg Arg Ala Gly Ile Ser Ser Phe Gly Phe Gly
450 455 460

Gly Ala Asn Tyr His Ala Val Leu Glu Glu Ala Glu Pro Glu His Thr
465 470 475 480

Thr Ala Tyr Arg Leu Asn Lys Arg Pro Gln Pro Val Leu Met Met Ala
485 490 495

Ala Thr Pro Ala Ala Leu Gln Ser Leu Cys Glu Ala Gln Leu Lys Glu
500 505 510

Phe Glu Ala Ala Ile Lys Glu Asn Glu Thr Val Lys Asn Thr Ala Tyr
515 520 525

Ile Lys Cys Val Lys Phe Gly Glu Gln Phe Lys Phe Pro Gly Ser Ile
530 535 540

Pro Ala Thr Asn Ala Arg Leu Gly Phe Leu Val Lys Asp Ala Glu Asp
545 550 555 560

Ala Cys Ser Thr Leu Arg Ala Ile Cys Ala Gln Phe Ala Lys Asp Val
565 570 575

Thr Lys Glu Ala Trp Arg Leu Pro Arg Glu Gly Val Ser Phe Arg Ala
580 585 590

Lys Gly Ile Ala Thr Asn Gly Ala Val Ala Ala Leu Phe Ser Gly Gln
 595 600 605

Gly Ala Gln Tyr Thr His Met Phe Ser Glu Val Ala Met Asn Trp Pro
 610 615 620

Gln Phe Arg Gln Ser Ile Ala Ala Met Asp Ala Ala Gln Ser Lys Val
 625 630 635 640

Ala Gly Ser Asp Lys Asp Phe Glu Arg Val Ser Gln Val Leu Tyr Pro
 645 650 655

Arg Lys Pro Tyr Glu Arg Glu Pro Glu Gln Asn Pro Lys Lys Ile Ser
 660 665 670

Leu Thr Ala Tyr Ser Gln Pro Ser Thr Leu Ala Cys Ala Leu Gly Ala
 675 680 685

Phe Glu Ile Phe Lys Glu Ala Gly Phe Thr Pro Asp Phe Ala Ala Gly
 690 695 700

His Ser Leu Gly Glu Phe Ala Ala Leu Tyr Ala Ala Gly Cys Val Asp
 705 710 715 720

Arg Asp Glu Leu Phe Glu Leu Val Cys Arg Arg Ala Arg Ile Met Gly
 725 730 735

Gly Lys Asp Ala Pro Ala Thr Pro Lys Gly Cys Met Ala Ala Val Ile
 740 745 750

Gly Pro Asn Ala Glu Asn Ile Lys Val Gln Ala Ala Asn Val Trp Leu
 755 760 765

Gly Asn Ser Asn Ser Pro Ser Gln Thr Val Ile Thr Gly Ser Val Glu
 770 775 780

Gly Ile Gln Ala Glu Ser Ala Arg Leu Gln Lys Glu Gly Phe Arg Val
 785 790 795 800

Val Pro Leu Ala Cys Glu Ser Ala Phe His Ser Pro Gln Met Glu Asn
 805 810 815

Ala Ser Ser Ala Phe Lys Asp Val Ile Ser Lys Val Ser Phe Arg Thr
 820 825 830

Pro Lys Ala Glu Thr Lys Leu Phe Ser Asn Val Ser Gly Glu Thr Tyr
 835 840 845

Pro Thr Asp Ala Arg Glu Met Leu Thr Gln His Met Thr Ser Ser Val
 850 855 860

Lys Phe Leu Thr Gln Val Arg Asn Met His Gln Ala Gly Ala Arg Ile
 865 870 875 880

Phe Val Glu Phe Gly Pro Lys Gln Val Leu Ser Lys Leu Val Ser Glu
 885 890 895

Thr Leu Lys Asp Asp Pro Ser Val Val Thr Val Ser Val Asn Pro Ala
 900 905 910

Ser Gly Thr Asp Ser Asp Ile Gln Leu Arg Asp Ala Ala Val Gln Leu
 915 920 925

Val Val Ala Gly Val Asn Leu Gln Gly Phe Asp Lys Trp Asp Ala Pro
 930 935 940

Asp Ala Thr Arg Met Gln Ala Ile Lys Lys Lys Arg Thr Thr Leu Arg
 945 950 955 960

Leu Ser Ala Ala Thr Tyr Val Ser Asp Lys Thr Lys Lys Val Arg Asp
 965 970 975

Ala Ala Met Asn Asp Gly Arg Cys Val Thr Tyr Leu Lys Gly Ala Ala
 980 985 990

Pro Leu Ile Lys Ala Pro Glu Pro Val Val Asp Glu Ala Ala Lys Arg
 995 1000 1005

Glu Ala Glu Arg Leu Gln Lys Glu Leu Gln Asp Ala Gln Arg Gln
 1010 1015 1020

Leu Asp Asp Ala Lys Arg Ala Ala Ala Glu Ala Asn Ser Lys Leu
 1025 1030 1035

Ala Ala Ala Lys Glu Glu Ala Lys Thr Ala Ala Ala Ser Ala Lys
 1040 1045 1050

Pro Ala Val Asp Thr Ala Val Val Glu Lys His Arg Ala Ile Leu
 1055 1060 1065

Lys Ser Met Leu Ala Glu Leu Asp Gly Tyr Gly Ser Val Asp Ala
 1070 1075 1080

Ser Ser Leu Gln Gln Gln Gln Gln Gln Gln Thr Ala Pro Ala Pro
 1085 1090 1095

Val Lys Ala Ala Ala Pro Ala Ala Pro Val Ala Ser Ala Pro Ala
 1100 1105 1110

Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu Thr Val Val
 1115 1120 1125

Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile
 1130 1135 1140

Glu Ala Asp Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile
 1145 1150 1155

Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met Leu Asn Val
 1160 1165 1170

Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr Arg Thr Val Gly
 1175 1180 1185

Glu Val Val Asn Ala Met Lys Ala Glu Ile Ala Gly Ser Ser Ala
 1190 1195 1200

Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Lys Ala Ala Pro
 1205 1210 1215

Ala Ala Ala Ala Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala
 1220 1225 1230

Glu Thr Val Val Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu
 1235 1240 1245

Thr Asp Met Ile Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly
 1250 1255 1260

Ile Asp Ser Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala
 1265 1270 1275

Met Leu Asn Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr
 1280 1285 1290

Arg Thr Val Gly Glu Val Val Asn Ala Met Lys Ala Glu Ile Ala
 1295 1300 1305

Gly Gly Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro Gly Pro Ala
 1310 1315 1320

Ala Ala Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Val Ser Asn
 1325 1330 1335

Glu Leu Leu Glu Lys Ala Glu Thr Val Val Met Glu Val Leu Ala
 1340 1345 1350

Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp Met Glu
 1355 1360 1365

Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu Ile
 1370 1375 1380

Leu Ser Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val
 1385 1390 1395

Asp Ala Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val Asp Ala
 1400 1405 1410

Met Lys Ala Glu Ile Ala Gly Gly Ser Ala Pro Ala Pro Ala Ala
 1415 1420 1425

Ala Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Ala Pro
 1430 1435 1440

Ala Pro Ala Val Ser Ser Glu Leu Leu Glu Lys Ala Glu Thr Val
 1445 1450 1455

Val Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met
 1460 1465 1470

Ile Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser
 1475 1480 1485

Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met Leu Asn
 1490 1495 1500

Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr Arg Thr Val
 1505 1510 1515

Gly Glu Val Val Asp Ala Met Lys Ala Glu Ile Ala Gly Gly Ser
 1520 1525 1530

Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Ala Ala Ala
 1535 1540 1545

Pro Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Pro Ala Val
 1550 1555 1560

Ser Ser Glu Leu Leu Glu Lys Ala Glu Thr Val Val Met Glu Val
 1565 1570 1575

Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp
 1580 1585 1590

Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg Val
 1595 1600 1605

Glu Ile Leu Ser Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys
 1610 1615 1620

Asp Val Asp Ala Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val
 1625 1630 1635

Asp Ala Met Lys Ala Glu Ile Ala Gly Ser Ser Ala Ser Ala Pro
 1640 1645 1650

Ala Ala Ala Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala
 1655 1660 1665

Ala Ala Ala Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu
 1670 1675 1680

Thr Val Val Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr
 1685 1690 1695

Asp Met Ile Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly Ile
 1700 1705 1710

Asp Ser Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met
 1715 1720 1725

Leu Asn Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr Arg
 1730 1735 1740

Thr Val Gly Glu Val Val Asp Ala Met Lys Ala Glu Ile Ala Gly
 1745 1750 1755

Gly Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Ala
 1760 1765 1770

Ala Ala Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu Thr
 1775 1780 1785

Val Val Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp
 1790 1795 1800

Met Ile Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly Ile Asp
 1805 1810 1815

Ser Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met Leu
 1820 1825 1830

Asn Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr Arg Thr
1835 1840 1845

Val Gly Glu Val Val Asp Ala Met Lys Ala Glu Ile Ala Gly Ser
1850 1855 1860

Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Ala Ala
1865 1870 1875

Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Val Ser Ser Glu Leu
1880 1885 1890

Leu Glu Lys Ala Glu Thr Val Val Met Glu Val Leu Ala Ala Lys
1895 1900 1905

Thr Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp Met Glu Leu Glu
1910 1915 1920

Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu Ile Leu Ser
1925 1930 1935

Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val Asp Ala
1940 1945 1950

Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val Asp Ala Met Lys
1955 1960 1965

Ala Glu Ile Ala Gly Gly Ser Ala Pro Ala Pro Ala Ala Ala Ala
1970 1975 1980

Pro Ala Pro Ala Ala Ala Ala Pro Ala Val Ser Asn Glu Leu Leu
1985 1990 1995

Glu Lys Ala Glu Thr Val Val Met Glu Val Leu Ala Ala Lys Thr
2000 2005 2010

Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp Met Glu Leu Glu Thr
2015 2020 2025

Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu Ile Leu Ser Glu
2030 2035 2040

Val Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val Asp Ala Leu
2045 2050 2055

Ser Arg Thr Arg Thr Val Gly Glu Val Val Asp Ala Met Lys Ala
2060 2065 2070

Glu Ile Ala Gly Gly Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro
2075 2080 2085

Ala Ser Ala Gly Ala Ala Pro Ala Val Lys Ile Asp Ser Val His
2090 2095 2100

Gly Ala Asp Cys Asp Asp Leu Ser Leu Met His Ala Lys Val Val
2105 2110 2115

Asp Ile Arg Arg Pro Asp Glu Leu Ile Leu Glu Arg Pro Glu Asn
2120 2125 2130

Arg Pro Val Leu Val Val Asp Asp Gly Ser Glu Leu Thr Leu Ala
2135 2140 2145

Leu Val Arg Val Leu Gly Ala Cys Ala Val Val Leu Thr Phe Glu
2150 2155 2160

Gly Leu Gln Leu Ala Gln Arg Ala Gly Ala Ala Ala Ile Arg His
2165 2170 2175

Val Leu Ala Lys Asp Leu Ser Ala Glu Ser Ala Glu Lys Ala Ile
2180 2185 2190

Lys Glu Ala Glu Gln Arg Phe Gly Ala Leu Gly Gly Phe Ile Ser
2195 2200 2205

Gln Gln Ala Glu Arg Phe Glu Pro Ala Glu Ile Leu Gly Phe Thr
2210 2215 2220

Leu Met Cys Ala Lys Phe Ala Lys Ala Ser Leu Cys Thr Ala Val
2225 2230 2235

Ala Gly Gly Arg Pro Ala Phe Ile Gly Val Ala Arg Leu Asp Gly
 2240 2245 2250

Arg Leu Gly Phe Thr Ser Gln Gly Thr Ser Asp Ala Leu Lys Arg
 2255 2260 2265

Ala Gln Arg Gly Ala Ile Phe Gly Leu Cys Lys Thr Ile Gly Leu
 2270 2275 2280

Glu Trp Ser Glu Ser Asp Val Phe Ser Arg Gly Val Asp Ile Ala
 2285 2290 2295

Gln Gly Met His Pro Glu Asp Ala Ala Val Ala Ile Val Arg Glu
 2300 2305 2310

Met Ala Cys Ala Asp Ile Arg Ile Arg Glu Val Gly Ile Gly Ala
 2315 2320 2325

Asn Gln Gln Arg Cys Thr Ile Arg Ala Ala Lys Leu Glu Thr Gly
 2330 2335 2340

Asn Pro Gln Arg Gln Ile Ala Lys Asp Asp Val Leu Leu Val Ser
 2345 2350 2355

Gly Gly Ala Arg Gly Ile Thr Pro Leu Cys Ile Arg Glu Ile Thr
 2360 2365 2370

Arg Gln Ile Ala Gly Gly Lys Tyr Ile Leu Leu Gly Arg Ser Lys
 2375 2380 2385

Val Ser Ala Ser Glu Pro Ala Trp Cys Ala Gly Ile Thr Asp Glu
 2390 2395 2400

Lys Ala Val Gln Lys Ala Ala Thr Gln Glu Leu Lys Arg Ala Phe
 2405 2410 2415

Ser Ala Gly Glu Gly Pro Lys Pro Thr Pro Arg Ala Val Thr Lys
 2420 2425 2430

Leu Val Gly Ser Val Leu Gly Ala Arg Glu Val Arg Ser Ser Ile
 2435 2440 2445

Ala Ala Ile Glu Ala Leu Gly Gly Lys Ala Ile Tyr Ser Ser Cys
2450 2455 2460

Asp Val Asn Ser Ala Ala Asp Val Ala Lys Ala Val Arg Asp Ala
2465 2470 2475

Glu Ser Gln Leu Gly Ala Arg Val Ser Gly Ile Val His Ala Ser
2480 2485 2490

Gly Val Leu Arg Asp Arg Leu Ile Glu Lys Lys Leu Pro Asp Glu
2495 2500 2505

Phe Asp Ala Val Phe Gly Thr Lys Val Thr Gly Leu Glu Asn Leu
2510 2515 2520

Leu Ala Ala Val Asp Arg Ala Asn Leu Lys His Met Val Leu Phe
2525 2530 2535

Ser Ser Leu Ala Gly Phe His Gly Asn Val Gly Gln Ser Asp Tyr
2540 2545 2550

Ala Met Ala Asn Glu Ala Leu Asn Lys Met Gly Leu Glu Leu Ala
2555 2560 2565

Lys Asp Val Ser Val Lys Ser Ile Cys Phe Gly Pro Trp Asp Gly
2570 2575 2580

Gly Met Val Thr Pro Gln Leu Lys Lys Gln Phe Gln Glu Met Gly
2585 2590 2595

Val Gln Ile Ile Pro Arg Glu Gly Gly Ala Asp Thr Val Ala Arg
2600 2605 2610

Ile Val Leu Gly Ser Ser Pro Ala Glu Ile Leu Val Gly Asn Trp
2615 2620 2625

Arg Thr Pro Ser Lys Lys Val Gly Ser Asp Thr Ile Thr Leu His
2630 2635 2640

Arg Lys Ile Ser Ala Lys Ser Asn Pro Phe Leu Glu Asp His Val
 2645 2650 2655

Ile Gln Gly Arg Arg Val Leu Pro Met Thr Leu Ala Ile Gly Ser
 2660 2665 2670

Leu Ala Glu Thr Cys Leu Gly Leu Phe Pro Gly Tyr Ser Leu Trp
 2675 2680 2685

Ala Ile Asp Asp Ala Gln Leu Phe Lys Gly Val Thr Val Asp Gly
 2690 2695 2700

Asp Val Asn Cys Glu Val Thr Leu Thr Pro Ser Thr Ala Pro Ser
 2705 2710 2715

Gly Arg Val Asn Val Gln Ala Thr Leu Lys Thr Phe Ser Ser Gly
 2720 2725 2730

Lys Leu Val Pro Ala Tyr Arg Ala Val Ile Val Leu Ser Asn Gln
 2735 2740 2745

Gly Ala Pro Pro Ala Asn Ala Thr Met Gln Pro Pro Ser Leu Asp
 2750 2755 2760

Ala Asp Pro Ala Leu Gln Gly Ser Val Tyr Asp Gly Lys Thr Leu
 2765 2770 2775

Phe His Gly Pro Ala Phe Arg Gly Ile Asp Asp Val Leu Ser Cys
 2780 2785 2790

Thr Lys Ser Gln Leu Val Ala Lys Cys Ser Ala Val Pro Gly Ser
 2795 2800 2805

Asp Ala Ala Arg Gly Glu Phe Ala Thr Asp Thr Asp Ala His Asp
 2810 2815 2820

Pro Phe Val Asn Asp Leu Ala Phe Gln Ala Met Leu Val Trp Val
 2825 2830 2835

Arg Arg Thr Leu Gly Gln Ala Ala Leu Pro Asn Ser Ile Gln Arg
 2840 2845 2850

Ile Val Gln His Arg Pro Val Pro Gln Asp Lys Pro Phe Tyr Ile
 2855 2860 2865

Thr Leu Arg Ser Asn Gln Ser Gly Gly His Ser Gln His Lys His
 2870 2875 2880

Ala Leu Gln Phe His Asn Glu Gln Gly Asp Leu Phe Ile Asp Val
 2885 2890 2895

Gln Ala Ser Val Ile Ala Thr Asp Ser Leu Ala Phe
 2900 2905 2910

<210> 3 <211> 6177 <212> DNA <213> 裂壺菌屬 sp. <220> <221> CDS <222> (1)..(6177)
 <223>

<400> 3
 atg gcc gct cgg aat gtg agc gcc gcg cat gag atg cac gat gaa aag 48
 Met Ala Ala Arg Asn Val Ser Ala Ala His Glu Met His Asp Glu Lys
 1 5 10 15

cgc atc gcc gtc gtc ggc atg gcc gtc cag tac gcc gga tgc aaa acc 96
 Arg Ile Ala Val Val Gly Met Ala Val Gln Tyr Ala Gly Cys Lys Thr
 20 25 30

aag gac gag ttc tgg gag gtg ctc atg aac ggc aag gtc gag tcc aag 144
 Lys Asp Glu Phe Trp Glu Val Leu Met Asn Gly Lys Val Glu Ser Lys
 35 40 45

gtg atc agc gac aaa cga ctc ggc tcc aac tac cgc gcc gag cac tac 192
 Val Ile Ser Asp Lys Arg Leu Gly Ser Asn Tyr Arg Ala Glu His Tyr
 50 55 60

aaa gca gag cgc agc aag tat gcc gac acc ttt tgc aac gaa acg tac 240
 Lys Ala Glu Arg Ser Lys Tyr Ala Asp Thr Phe Cys Asn Glu Thr Tyr
 65 70 75 80

ggc acc ctt gac gag aac gag atc gac aac gag cac gaa ctc ctc ctc 288
 Gly Thr Leu Asp Glu Asn Glu Ile Asp Asn Glu His Glu Leu Leu Leu
 85 90 95

aac ctc gcc aag cag gca ctc gca gag aca tcc gtc aaa gac tcg aca 336
 Asn Leu Ala Lys Gln Ala Leu Ala Glu Thr Ser Val Lys Asp Ser Thr
 100 105 110

cgc tgc ggc atc gtc agc ggc tgc ctc tcg ttc ccc atg gac aac ctc 384
 Arg Cys Gly Ile Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu
 115 120 125

cag ggt gaa ctc ctc aac gtg tac caa aac cat gtc gag aaa aag ctc 432
 Gln Gly Glu Leu Leu Asn Val Tyr Gln Asn His Val Glu Lys Lys Leu
 130 135 140

ggg gcc cgc gtc ttc aag gac gcc tcc cat tgg tcc gaa cgc gag cag 480
 Gly Ala Arg Val Phe Lys Asp Ala Ser His Trp Ser Glu Arg Glu Gln
 145 150 155 160

tcc aac aaa ccc gag gcc ggt gac cgc cgc atc ttc atg gac ccg gcc 528
 Ser Asn Lys Pro Glu Ala Gly Asp Arg Arg Ile Phe Met Asp Pro Ala
 165 170 175

tcc ttc gtc gcc gaa gaa ctc aac ctc ggc gcc ctt cac tac tcc gtc 576
 Ser Phe Val Ala Glu Glu Leu Asn Leu Gly Ala Leu His Tyr Ser Val
 180 185 190

gac gca gca tgc gcc acg gcg ctc tac gtg ctc cgc ctc gcg cag gat 624
 Asp Ala Ala Cys Ala Thr Ala Leu Tyr Val Leu Arg Leu Ala Gln Asp
 195 200 205

cat ctc gtc tcc ggc gcc gcc gac gtc atg ctc tgc ggt gcc acc tgc 672
 His Leu Val Ser Gly Ala Ala Asp Val Met Leu Cys Gly Ala Thr Cys
 210 215 220

ctg ccg gag ccc ttt ttc atc ctt tgg ggc ttt tcc acc ttc cag gcc 720
 Leu Pro Glu Pro Phe Phe Ile Leu Ser Gly Phe Ser Thr Phe Gln Ala
 225 230 235 240

atg ccc gtc ggc acg ggc cag aac gtg tcc atg ccg ctg cac aag gac 768
 Met Pro Val Gly Thr Gly Gln Asn Val Ser Met Pro Leu His Lys Asp
 245 250 255

agc cag ggc ctc acc ccg ggt gag ggc gcc tcc atc atg gtc ctc aag 816
 Ser Gln Gly Leu Thr Pro Gly Glu Gly Gly Ser Ile Met Val Leu Lys
 260 265 270

cgt ctc gat gat gcc atc cgc gac ggc gac cac att tac ggc acc ctt 864
 Arg Leu Asp Asp Ala Ile Arg Asp Gly Asp His Ile Tyr Gly Thr Leu
 275 280 285

ctc ggc gcc aat gtc agc aac tcc ggc aca ggt ctg ccc ctc aag ccc 912
 Leu Gly Ala Asn Val Ser Asn Ser Gly Thr Gly Leu Pro Leu Lys Pro
 290 295 300

ctt ctc ccc agc gag aaa aag tgc ctc atg gac acc tac acg cgc att 960
 Leu Leu Pro Ser Glu Lys Lys Cys Leu Met Asp Thr Tyr Thr Arg Ile
 305 310 315 320

aac gtg cac ccg cac aag att cag tac gtc gag tgc cac gcc acc ggc 1008
 Asn Val His Pro His Lys Ile Gln Tyr Val Glu Cys His Ala Thr Gly
 325 330 335

acg ccc cag ggt gat cgt gtg gaa atc gac gcc gtc aag gcc tgc ttt 1056
 Thr Pro Gln Gly Asp Arg Val Glu Ile Asp Ala Val Lys Ala Cys Phe

340	345	350	
gaa ggc aag gtc ccc cgt ttc ggt acc aca aag ggc aac ttt gga cac 1104			
Glu Gly Lys Val Pro Arg Phe Gly Thr Thr Lys Gly Asn Phe Gly His			
355	360	365	
acc cts gyc gca gcc ggc ttt gcc ggt atg tgc aag gtc ctc ctc tcc 1152			
Thr Xaa Xaa Ala Ala Gly Phe Ala Gly Met Cys Lys Val Leu Leu Ser			
370	375	380	
atg aag cat ggc atc atc ccg ccc acc ccg ggt atc gat gac gag acc 1200			
Met Lys His Gly Ile Ile Pro Pro Thr Pro Gly Ile Asp Asp Glu Thr			
385	390	395	400
aag atg gac cct ctc gtc gtc tcc ggt gag gcc atc cca tgg cca gag 1248			
Lys Met Asp Pro Leu Val Val Ser Gly Glu Ala Ile Pro Trp Pro Glu			
405	410	415	
acc aac ggc gag ccc aag cgc gcc ggt ctc tgc gcc ttt ggc ttt ggt 1296			
Thr Asn Gly Glu Pro Lys Arg Ala Gly Leu Ser Ala Phe Gly Phe Gly			
420	425	430	
ggc acc aac gcc cat gcc gtc ttt gag gag cat gac ccc tcc aac gcc 1344			
Gly Thr Asn Ala His Ala Val Phe Glu Glu His Asp Pro Ser Asn Ala			
435	440	445	
gcc tgc acg ggc cac gac tcc att tet gcg ctc tgc gcc cgc tgc ggc 1392			
Ala Cys Thr Gly His Asp Ser Ile Ser Ala Leu Ser Ala Arg Cys Gly			
450	455	460	
ggt gaa agc aac atg cgc atc gcc atc act ggt atg gac gcc acc ttt 1440			
Gly Glu Ser Asn Met Arg Ile Ala Ile Thr Gly Met Asp Ala Thr Phe			
465	470	475	480
ggc gct ctc aag gga ctc gac gcc ttc gag cgc gcc att tac acc ggc 1488			
Gly Ala Leu Lys Gly Leu Asp Ala Phe Glu Arg Ala Ile Tyr Thr Gly			
485	490	495	
ggt cac ggt gcc atc cca ctc cca gaa aag cgc tgg cgc ttt ctc ggc 1536			
Ala His Gly Ala Ile Pro Leu Pro Glu Lys Arg Trp Arg Phe Leu Gly			
500	505	510	
aag gac aag gac ttt ctt gac ctc tgc ggc gtc aag gcc acc ccg cac 1584			
Lys Asp Lys Asp Phe Leu Asp Leu Cys Gly Val Lys Ala Thr Pro His			
515	520	525	
ggc tgc tac att gaa gat gtt gag gtc gac ttc cag cgc ctc cgc acg 1632			
Gly Cys Tyr Ile Glu Asp Val Glu Val Asp Phe Gln Arg Leu Arg Thr			
530	535	540	
ccc atg acc cct gaa gac atg ctc ctc cct cag cag ctt ctg gcc gtc 1680			
Pro Met Thr Pro Glu Asp Met Leu Leu Pro Gln Gln Leu Leu Ala Val			
545	550	555	560

acc acc att gac cgc gcc atc ctc gac tcg gga atg aaa aag ggt ggc 1728
 Thr Thr Ile Asp Arg Ala Ile Leu Asp Ser Gly Met Lys Lys Gly Gly
 565 570 575

aat gtc gcc gtc ttt gtc ggc ctc ggc acc gac ctc gag ctc tac cgt 1776
 Asn Val Ala Val Phe Val Gly Leu Gly Thr Asp Leu Glu Leu Tyr Arg
 580 585 590

cac cgt gct cgc gtc get ctc aag gag cgc gtc cgc cct gaa gcc tcc 1824
 His Arg Ala Arg Val Ala Leu Lys Glu Arg Val Arg Pro Glu Ala Ser
 595 600 605

aag aag ctc aat gac atg atg cag tac att aac gac tgc ggc aca tcc 1872
 Lys Lys Leu Asn Asp Met Met Gln Tyr Ile Asn Asp Cys Gly Thr Ser
 610 615 620

aca tcg tac acc tcg tac att ggc aac ctc gtc gcc acg cgc gtc tcg 1920
 Thr Ser Tyr Thr Ser Tyr Ile Gly Asn Leu Val Ala Thr Arg Val Ser
 625 630 635 640

tcg cag tgg ggc ttc acg ggc ccc tcc ttt acg atc acc gag ggc aac 1968
 Ser Gln Trp Gly Phe Thr Gly Pro Ser Phe Thr Ile Thr Glu Gly Asn
 645 650 655

aac tcc gtc tac cgc tgc gcc gag ctc ggc aag tac ctc ctc gag acc 2016
 Asn Ser Val Tyr Arg Cys Ala Glu Leu Gly Lys Tyr Leu Leu Glu Thr
 660 665 670

ggc gag gtc gat ggc gtc gtc gtt gcg ggt gtc gat ctc tgc ggc agt 2064
 Gly Glu Val Asp Gly Val Val Val Ala Gly Val Asp Leu Cys Gly Ser
 675 680 685

gcc gaa aac ctt tac gtc aag tct cgc cgc ttc aag gtg tcc acc tcc 2112
 Ala Glu Asn Leu Tyr Val Lys Ser Arg Arg Phe Lys Val Ser Thr Ser
 690 695 700

gat acc ccg cgc gcc agc ttt gac gcc gcc gcc gat ggc tac ttt gtc 2160
 Asp Thr Pro Arg Ala Ser Phe Asp Ala Ala Ala Asp Gly Tyr Phe Val
 705 710 715 720

ggc gag ggc tgc ggt gcc ttt gtg ctc aag cgt gag act agc tgc acc 2208
 Gly Glu Gly Cys Gly Ala Phe Val Leu Lys Arg Glu Thr Ser Cys Thr
 725 730 735

aag gac gac cgt atc tac get tgc atg gat gcc atc gtc cct ggc aac 2256
 Lys Asp Asp Arg Ile Tyr Ala Cys Met Asp Ala Ile Val Pro Gly Asn
 740 745 750

gtc cct agc gcc tgc ttg cgc gag gcc ctc gac cag gcg cgc gtc aag 2304
 Val Pro Ser Ala Cys Leu Arg Glu Ala Leu Asp Gln Ala Arg Val Lys
 755 760 765

ccg ggc gat atc gag atg ctc gag ctc agc gcc gac tcc gcc cgc cac 2352
 Pro Gly Asp Ile Glu Met Leu Glu Leu Ser Ala Asp Ser Ala Arg His

770	775	780	
ctc aag gac ccg tcc gtc ctg ccc aag gag ctc act gcc gag gag gaa			2400
Leu Lys Asp Pro Ser Val Leu Pro Lys Glu Leu Thr Ala Glu Glu Glu			
785	790	795	800
atc ggc ggc ctt cag acg atc ctt cgt gac gat gac aag ctc ccg cgc			2448
Ile Gly Gly Leu Gln Thr Ile Leu Arg Asp Asp Asp Lys Leu Pro Arg			
805	810	815	
aac gtc gca acg ggc agt gtc aag gcc acc gtc ggt gac acc ggt tat			2496
Asn Val Ala Thr Gly Ser Val Lys Ala Thr Val Gly Asp Thr Gly Tyr			
820	825	830	
gcc tct ggt gct gcc agc ctc atc aag gct gcg ctt tgc atc tac aac			2544
Ala Ser Gly Ala Ala Ser Leu Ile Lys Ala Ala Leu Cys Ile Tyr Asn			
835	840	845	
cgc tac ctg ccc agc aac ggc gac gac tgg gat gaa ccc gcc cct gag			2592
Arg Tyr Leu Pro Ser Asn Gly Asp Asp Trp Asp Glu Pro Ala Pro Glu			
850	855	860	
gcg ccc tgg gac agc acc ctc ttt gcg tgc cag acc tcg cgc gct tgg			2640
Ala Pro Trp Asp Ser Thr Leu Phe Ala Cys Gln Thr Ser Arg Ala Trp			
865	870	875	880
ctc aag aac cct ggc gag cgt cgc tat gcg gcc gtc tcg ggc gtc tcc			2688
Leu Lys Asn Pro Gly Glu Arg Arg Tyr Ala Ala Val Ser Gly Val Ser			
885	890	895	
gag acg cgc tcg tgc tat tcc gtg ctc ctc tcc gaa gcc gag ggc cac			2736
Glu Thr Arg Ser Cys Tyr Ser Val Leu Leu Ser Glu Ala Glu Gly His			
900	905	910	
tac gag cgc gag aac cgc atc tcg ctc gac gag gag gcg ccc aag ctc			2784
Tyr Glu Arg Glu Asn Arg Ile Ser Leu Asp Glu Glu Ala Pro Lys Leu			
915	920	925	
att gtg ctt cgc gcc gac tcc cac gag gag atc ctt ggt cgc ctc gac			2832
Ile Val Leu Arg Ala Asp Ser His Glu Glu Ile Leu Gly Arg Leu Asp			
930	935	940	
aag atc cgc gag cgc ttc ttg cag ccc acg ggc gcc gcc ccg cgc gag			2880
Lys Ile Arg Glu Arg Phe Leu Gln Pro Thr Gly Ala Ala Pro Arg Glu			
945	950	955	960
tcc gag ctc aag gcg cag gcc cgc cgc atc ttc ctc gag ctc ctc ggc			2928
Ser Glu Leu Lys Ala Gln Ala Arg Arg Ile Phe Leu Glu Leu Leu Gly			
965	970	975	
gag acc ctt gcc cag gat gcc gct tct tca ggc tcg caa aag ccc ctc			2976
Glu Thr Leu Ala Gln Asp Ala Ala Ser Ser Gly Ser Gln Lys Pro Leu			
980	985	990	

gct ctc agc ctc gtc tcc acg ccc tcc aag ctc cag cgc gag gtc gag 3024
Ala Leu Ser Leu Val Ser Thr Pro Ser Lys Leu Gln Arg Glu Val Glu
995 1000 1005

ctc gcg gcc aag ggt atc ccg cgc tgc ctc aag atg cgc cgc gat 3069
Leu Ala Ala Lys Gly Ile Pro Arg Cys Leu Lys Met Arg Arg Asp
1010 1015 1020

tgg agc tcc cct gct ggc agc cgc tac gcg cct gag ccg ctc gcc 3114
Trp Ser Ser Pro Ala Gly Ser Arg Tyr Ala Pro Glu Pro Leu Ala
1025 1030 1035

agc gac cgc gtc gcc ttc atg tac ggc gaa ggt cgc agc cct tac 3159
Ser Asp Arg Val Ala Phe Met Tyr Gly Glu Gly Arg Ser Pro Tyr
1040 1045 1050

tac ggc atc acc caa gac att cac cgc att tgg ccc gaa ctc cac 3204
Tyr Gly Ile Thr Gln Asp Ile His Arg Ile Trp Pro Glu Leu His
1055 1060 1065

gag gtc atc aac gaa aag acg aac cgt ctc tgg gcc gaa ggc gac 3249
Glu Val Ile Asn Glu Lys Thr Asn Arg Leu Trp Ala Glu Gly Asp
1070 1075 1080

cgc tgg gtc atg ccg cgc gcc agc ttc aag tcg gag ctc gag agc 3294
Arg Trp Val Met Pro Arg Ala Ser Phe Lys Ser Glu Leu Glu Ser
1085 1090 1095

cag cag caa gag ttt gat cgc aac atg att gaa atg ttc cgt ctt 3339
Gln Gln Gln Glu Phe Asp Arg Asn Met Ile Glu Met Phe Arg Leu
1100 1105 1110

gga atc ctc acc tca att gcc ttc acc aat ctg gcg cgc gac gtt 3384
Gly Ile Leu Thr Ser Ile Ala Phe Thr Asn Leu Ala Arg Asp Val
1115 1120 1125

ctc aac atc acg ccc aag gcc gcc ttt ggc ctc agt ctt ggc gag 3429
Leu Asn Ile Thr Pro Lys Ala Ala Phe Gly Leu Ser Leu Gly Glu
1130 1135 1140

att tcc atg att ttt gcc ttt tcc aag aag aac ggt ctc atc tcc 3474
Ile Ser Met Ile Phe Ala Phe Ser Lys Lys Asn Gly Leu Ile Ser
1145 1150 1155

gac cag ctc acc aag gat ctt cgc gag tcc gac gtg tgg aac aag 3519
Asp Gln Leu Thr Lys Asp Leu Arg Glu Ser Asp Val Trp Asn Lys
1160 1165 1170

gct ctg gcc gtt gaa ttt aat gcg ctg cgc gag gcc tgg ggc att 3564
Ala Leu Ala Val Glu Phe Asn Ala Leu Arg Glu Ala Trp Gly Ile
1175 1180 1185

cca cag agt gtc ccc aag gac gag ttc tgg caa ggc tac att gtg 3609
Pro Gln Ser Val Pro Lys Asp Glu Phe Trp Gln Gly Tyr Ile Val

1190	1195	1200	
cgc ggc acc aag cag gat atc gag gcg gcc atc gcc ccg gac agc	3654		
Arg Gly Thr Lys Gln Asp Ile Glu Ala Ala Ile Ala Pro Asp Ser			
1205	1210	1215	
aag tac gtg cgc ctc acc atc atc aat gat gcc aac acc gcc ctc	3699		
Lys Tyr Val Arg Leu Thr Ile Ile Asn Asp Ala Asn Thr Ala Leu			
1220	1225	1230	
att agc ggc aag ccc gac gcc tgc aag gct gcg atc gcg cgt ctc	3744		
Ile Ser Gly Lys Pro Asp Ala Cys Lys Ala Ala Ile Ala Arg Leu			
1235	1240	1245	
ggg ggc aac att cct gcg ctt ccc gtg acc cag ggc atg tgc ggc	3789		
Gly Gly Asn Ile Pro Ala Leu Pro Val Thr Gln Gly Met Cys Gly			
1250	1255	1260	
cac tgc ccc gag gtg gga cct tat acc aag gat atc gcc aag atc	3834		
His Cys Pro Glu Val Gly Pro Tyr Thr Lys Asp Ile Ala Lys Ile			
1265	1270	1275	
cat gcc aac ctt gag ttc ccc gtt gtc gac ggc ctt gac ctc tgg	3879		
His Ala Asn Leu Glu Phe Pro Val Val Asp Gly Leu Asp Leu Trp			
1280	1285	1290	
acc aca atc aac cag aag cgc ctc gtg cca cgc gcc acg ggc gcc	3924		
Thr Thr Ile Asn Gln Lys Arg Leu Val Pro Arg Ala Thr Gly Ala			
1295	1300	1305	
aag gac gaa tgg gcc cct tct tcc ttt ggc gag tac gcc ggc cag	3969		
Lys Asp Glu Trp Ala Pro Ser Ser Phe Gly Glu Tyr Ala Gly Gln			
1310	1315	1320	
ctc tac gag aag cag gct aac ttc ccc caa atc gtc gag acc att	4014		
Leu Tyr Glu Lys Gln Ala Asn Phe Pro Gln Ile Val Glu Thr Ile			
1325	1330	1335	
tac aag caa aac tac gac gtc ttt gtc gag gtt ggg ccc aac aac	4059		
Tyr Lys Gln Asn Tyr Asp Val Phe Val Glu Val Gly Pro Asn Asn			
1340	1345	1350	
cac cgt agc acc gca gtg cgc acc acg ctt ggt ccc cag cgc aac	4104		
His Arg Ser Thr Ala Val Arg Thr Thr Leu Gly Pro Gln Arg Asn			
1355	1360	1365	
cac ctt gct ggc gcc atc gac aag cag aac gag gat gct tgg acg	4149		
His Leu Ala Gly Ala Ile Asp Lys Gln Asn Glu Asp Ala Trp Thr			
1370	1375	1380	
acc atc gtc aag ctt gtg gct tcg ctc aag gcc cac ctt gtt cct	4194		
Thr Ile Val Lys Leu Val Ala Ser Leu Lys Ala His Leu Val Pro			
1385	1390	1395	

ggc gtc acg atc tcg ccg ctg tac cac tcc aag ctt gtg gcg gag 4239
 Gly Val Thr Ile Ser Pro Leu Tyr His Ser Lys Leu Val Ala Glu
 1400 1405 1410

gct cag gct tgc tac gct gcg ctc tgc aag ggt gaa aag ccc aag 4284
 Ala Gln Ala Cys Tyr Ala Ala Leu Cys Lys Gly Glu Lys Pro Lys
 1415 1420 1425

aag aac aag ttt gtg gcg aag att cag ctc aac ggt cgc ttc aac 4329
 Lys Asn Lys Phe Val Arg Lys Ile Gln Leu Asn Gly Arg Phe Asn
 1430 1435 1440

agc aag gcg gac ccc atc tcc tcg gcc gat ctt gcc agc ttt ccg 4374
 Ser Lys Ala Asp Pro Ile Ser Ser Ala Asp Leu Ala Ser Phe Pro
 1445 1450 1455

cct gcg gac cct gcc att gaa gcc gcc atc tcg agc cgc atc atg 4419
 Pro Ala Asp Pro Ala Ile Glu Ala Ala Ile Ser Ser Arg Ile Met
 1460 1465 1470

aag cct gtc gct ccc aag ttc tac gcg cgt ctc aac att gac gag 4464
 Lys Pro Val Ala Pro Lys Phe Tyr Ala Arg Leu Asn Ile Asp Glu
 1475 1480 1485

cag gac gag acc cga gat ccg atc ctc aac aag gac aac gcg ccg 4509
 Gln Asp Glu Thr Arg Asp Pro Ile Leu Asn Lys Asp Asn Ala Pro
 1490 1495 1500

tct tct tct tct tct tct tct tct tct tct tct tct tct 4554
 Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser
 1505 1510 1515

ccg tcg cct gct cct tcg gcc ccc gtg caa aag aag gct gct ccc 4599
 Pro Ser Pro Ala Pro Ser Ala Pro Val Gln Lys Lys Ala Ala Pro
 1520 1525 1530

gcc gcg gag acc aag gct gtt gct tcg gct gac gca ctt cgc agt 4644
 Ala Ala Glu Thr Lys Ala Val Ala Ser Ala Asp Ala Leu Arg Ser
 1535 1540 1545

gcc ctg ctc gat ctc gac agt atg ctt gcg ctg agc tct gcc agt 4689
 Ala Leu Leu Asp Leu Asp Ser Met Leu Ala Leu Ser Ser Ala Ser
 1550 1555 1560

gcc tcc ggc aac ctt gtt gag act gcg cct agc gac gcc tcg gtc 4734
 Ala Ser Gly Asn Leu Val Glu Thr Ala Pro Ser Asp Ala Ser Val
 1565 1570 1575

att gtg ccg ccc tgc aac att gcg gat ctc ggc agc cgc gcc ttc 4779
 Ile Val Pro Pro Cys Asn Ile Ala Asp Leu Gly Ser Arg Ala Phe
 1580 1585 1590

atg aaa acg tac ggt gtt tcg gcg cct ctg tac acg ggc gcc atg 4824
 Met Lys Thr Tyr Gly Val Ser Ala Pro Leu Tyr Thr Gly Ala Met

1595	1600	1605	
gcc aag ggc att gcc tct gcg gac ctc gtc att gcc gcc ggc cgc 4869			
Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala Ala Gly Arg			
1610	1615	1620	
cag ggc atc ctt gcg tcc ttt ggc gcc ggc gga ctt ccc atg cag 4914			
Gln Gly Ile Leu Ala Ser Phe Gly Ala Gly Gly Leu Pro Met Gln			
1625	1630	1635	
gtt gtg cgt gag tcc atc gaa aag att cag gcc gcc ctg ccc aat 4959			
Val Val Arg Glu Ser Ile Glu Lys Ile Gln Ala Ala Leu Pro Asn			
1640	1645	1650	
ggc cgg tac gct gtc aac ctt atc cat tct ccc ttt gac agc aac 5004			
Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp Ser Asn			
1655	1660	1665	
ctc gaa aag ggc aat gtc gat ctc ttc ctc gag aag ggt gtc acc 5049			
Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys Gly Val Thr			
1670	1675	1680	
ttt gtc gag gcc tcg gcc ttt atg acg ctc acc cgg cag gtc gtg 5094			
Phe Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln Val Val			
1685	1690	1695	
cgg tac cgc gcg gct ggc ctc acg cgc aac gcc gac ggc tcg gtc 5139			
Arg Tyr Arg Ala Ala Gly Leu Thr Arg Asn Ala Asp Gly Ser Val			
1700	1705	1710	
aac atc cgc aac cgt atc att ggc aag gtc tcg cgc acc gag ctc 5184			
Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr Glu Leu			
1715	1720	1725	
gcc gag atg ttc atg cgt cct gcg ccc gag cac ctt ctt cag aag 5229			
Ala Glu Met Phe Met Arg Pro Ala Pro Glu His Leu Leu Gln Lys			
1730	1735	1740	
ctc att gct tcc ggc gag atc aac cag gag cag gcc gag ctc gcc 5274			
Leu Ile Ala Ser Gly Glu Ile Asn Gln Glu Gln Ala Glu Leu Ala			
1745	1750	1755	
cgc cgt gtt ccc gtc gct gac gac atc gcg gtc gaa gct gac tcg 5319			
Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala Asp Ser			
1760	1765	1770	
ggt ggc cac acc gac aac cgc ccc atc cac gtc att ctg ccc ctc 5364			
Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile Leu Pro Leu			
1775	1780	1785	
atc atc aac ctt cgc gac cgc ctt cac cgc gag tgc ggc tac cgg 5409			
Ile Ile Asn Leu Arg Asp Arg Leu His Arg Glu Cys Gly Tyr Pro			
1790	1795	1800	

gcc aac ctt cgc gtc cgt gtg ggc gcc ggc ggt ggc att ggg tgc 5454
Ala Asn Leu Arg Val Arg Val Gly Ala Gly Gly Gly Ile Gly Cys
1805 1810 1815

ccc cag gcg cgc ctg gcc acc ttc aac atg ggt gcc tcc ttt att 5499
Pro Gln Ala Ala Leu Ala Thr Phe Asn Met Gly Ala Ser Phe Ile
1820 1825 1830

gtc acc ggc acc gtg aac cag gtc gcc aag cag tcg ggc acg tgc 5544
Val Thr Gly Thr Val Asn Gln Val Ala Lys Gln Ser Gly Thr Cys
1835 1840 1845

gac aat gtg cgc aag cag ctc gcg aag gcc act tac tcg gac gta 5589
Asp Asn Val Arg Lys Gln Leu Ala Lys Ala Thr Tyr Ser Asp Val
1850 1855 1860

tgc atg gcc ccg gct gcc gac atg ttc gag gaa ggc gtc aag ctt 5634
Cys Met Ala Pro Ala Ala Asp Met Phe Glu Glu Gly Val Lys Leu
1865 1870 1875

cag gtc ctc aag aag gga acc atg ttt ccc tcg cgc gcc aac aag 5679
Gln Val Leu Lys Lys Gly Thr Met Phe Pro Ser Arg Ala Asn Lys
1880 1885 1890

ctc tac gag ctc ttt tgc aag tac gac tcg ttc gag tcc atg ccc 5724
Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe Glu Ser Met Pro
1895 1900 1905

ccc gca gag ctt gcg cgc gtc gag aag cgc atc ttc agc cgc gcg 5769
Pro Ala Glu Leu Ala Arg Val Glu Lys Arg Ile Phe Ser Arg Ala
1910 1915 1920

ctc gaa gag gtc tgg gac gag acc aaa aac ttt tac att aac cgt 5814
Leu Glu Glu Val Trp Asp Glu Thr Lys Asn Phe Tyr Ile Asn Arg
1925 1930 1935

ctt cac aac ccg gag aag atc cag cgc gcc gag cgc gac ccc aag 5859
Leu His Asn Pro Glu Lys Ile Gln Arg Ala Glu Arg Asp Pro Lys
1940 1945 1950

ctc aag atg tcg ctg tgc ttt cgc tgg tac ctg agc ctg gcg agc 5904
Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Ser Leu Ala Ser
1955 1960 1965

cgc tgg gcc aac act gga get tcc gat cgc gtc atg gac tac cag 5949
Arg Trp Ala Asn Thr Gly Ala Ser Asp Arg Val Met Asp Tyr Gln
1970 1975 1980

gtc tgg tgc ggt cct gcc att ggt tcc ttc aac gat ttc atc aag 5994
Val Trp Cys Gly Pro Ala Ile Gly Ser Phe Asn Asp Phe Ile Lys
1985 1990 1995

gga act tac ctt gat ccg gcc gtc gca aac gag tac ccg tgc gtc 6039
Gly Thr Tyr Leu Asp Pro Ala Val Ala Asn Glu Tyr Pro Cys Val

Arg Cys Gly Ile Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu
 115 120 125

Gln Gly Glu Leu Leu Asn Val Tyr Gln Asn His Val Glu Lys Lys Leu
 130 135 140

Gly Ala Arg Val Phe Lys Asp Ala Ser His Trp Ser Glu Arg Glu Gln
 145 150 155 160

Ser Asn Lys Pro Glu Ala Gly Asp Arg Arg Ile Phe Met Asp Pro Ala
 165 170 175

Ser Phe Val Ala Glu Glu Leu Asn Leu Gly Ala Leu His Tyr Ser Val
 180 185 190

Asp Ala Ala Cys Ala Thr Ala Leu Tyr Val Leu Arg Leu Ala Gln Asp
 195 200 205

His Leu Val Ser Gly Ala Ala Asp Val Met Leu Cys Gly Ala Thr Cys
 210 215 220

Leu Pro Glu Pro Phe Phe Ile Leu Ser Gly Phe Ser Thr Phe Gln Ala
 225 230 235 240

Met Pro Val Gly Thr Gly Gln Asn Val Ser Met Pro Leu His Lys Asp
 245 250 255

Ser Gln Gly Leu Thr Pro Gly Glu Gly Gly Ser Ile Met Val Leu Lys
 260 265 270

Arg Leu Asp Asp Ala Ile Arg Asp Gly Asp His Ile Tyr Gly Thr Leu
 275 280 285

Leu Gly Ala Asn Val Ser Asn Ser Gly Thr Gly Leu Pro Leu Lys Pro
 290 295 300

Leu Leu Pro Ser Glu Lys Lys Cys Leu Met Asp Thr Tyr Thr Arg Ile
 305 310 315 320

Asn Val His Pro His Lys Ile Gln Tyr Val Glu Cys His Ala Thr Gly
 325 330 335

Thr Pro Gln Gly Asp Arg Val Glu Ile Asp Ala Val Lys Ala Cys Phe
 340 345 350

Glu Gly Lys Val Pro Arg Phe Gly Thr Thr Lys Gly Asn Phe Gly His
 355 360 365

Thr Xaa Xaa Ala Ala Gly Phe Ala Gly Met Cys Lys Val Leu Leu Ser
 370 375 380

Met Lys His Gly Ile Ile Pro Pro Thr Pro Gly Ile Asp Asp Glu Thr
 385 390 395 400

Lys Met Asp Pro Leu Val Val Ser Gly Glu Ala Ile Pro Trp Pro Glu
 405 410 415

Thr Asn Gly Glu Pro Lys Arg Ala Gly Leu Ser Ala Phe Gly Phe Gly
 420 425 430

Gly Thr Asn Ala His Ala Val Phe Glu Glu His Asp Pro Ser Asn Ala
 435 440 445

Ala Cys Thr Gly His Asp Ser Ile Ser Ala Leu Ser Ala Arg Cys Gly
 450 455 460

Gly Glu Ser Asn Met Arg Ile Ala Ile Thr Gly Met Asp Ala Thr Phe
 465 470 475 480

Gly Ala Leu Lys Gly Leu Asp Ala Phe Glu Arg Ala Ile Tyr Thr Gly
 485 490 495

Ala His Gly Ala Ile Pro Leu Pro Glu Lys Arg Trp Arg Phe Leu Gly
 500 505 510

Lys Asp Lys Asp Phe Leu Asp Leu Cys Gly Val Lys Ala Thr Pro His
 515 520 525

Gly Cys Tyr Ile Glu Asp Val Glu Val Asp Phe Gln Arg Leu Arg Thr
 530 535 540

Pro Met Thr Pro Glu Asp Met Leu Leu Pro Gln Gln Leu Leu Ala Val
545 550 555 560

Thr Thr Ile Asp Arg Ala Ile Leu Asp Ser Gly Met Lys Lys Gly Gly
 565 570 575

Asn Val Ala Val Phe Val Gly Leu Gly Thr Asp Leu Glu Leu Tyr Arg
 580 585 590

His Arg Ala Arg Val Ala Leu Lys Glu Arg Val Arg Pro Glu Ala Ser
 595 600 605

Lys Lys Leu Asn Asp Met Met Gln Tyr Ile Asn Asp Cys Gly Thr Ser
 610 615 620

Thr Ser Tyr Thr Ser Tyr Ile Gly Asn Leu Val Ala Thr Arg Val Ser
625 630 635 640

Ser Gln Trp Gly Phe Thr Gly Pro Ser Phe Thr Ile Thr Glu Gly Asn
 645 650 655

Asn Ser Val Tyr Arg Cys Ala Glu Leu Gly Lys Tyr Leu Leu Glu Thr
 660 665 670

Gly Glu Val Asp Gly Val Val Val Ala Gly Val Asp Leu Cys Gly Ser
 675 680 685

Ala Glu Asn Leu Tyr Val Lys Ser Arg Arg Phe Lys Val Ser Thr Ser
 690 695 700

Asp Thr Pro Arg Ala Ser Phe Asp Ala Ala Ala Asp Gly Tyr Phe Val
705 710 715 720

Gly Glu Gly Cys Gly Ala Phe Val Leu Lys Arg Glu Thr Ser Cys Thr
 725 730 735

Lys Asp Asp Arg Ile Tyr Ala Cys Met Asp Ala Ile Val Pro Gly Asn
 740 745 750

Val Pro Ser Ala Cys Leu Arg Glu Ala Leu Asp Gln Ala Arg Val Lys
 755 760 765

Pro Gly Asp Ile Glu Met Leu Glu Leu Ser Ala Asp Ser Ala Arg His
770 775 780

Leu Lys Asp Pro Ser Val Leu Pro Lys Glu Leu Thr Ala Glu Glu Glu
785 790 795 800

Ile Gly Gly Leu Gln Thr Ile Leu Arg Asp Asp Asp Lys Leu Pro Arg
805 810 815

Asn Val Ala Thr Gly Ser Val Lys Ala Thr Val Gly Asp Thr Gly Tyr
820 825 830

Ala Ser Gly Ala Ala Ser Leu Ile Lys Ala Ala Leu Cys Ile Tyr Asn
835 840 845

Arg Tyr Leu Pro Ser Asn Gly Asp Asp Trp Asp Glu Pro Ala Pro Glu
850 855 860

Ala Pro Trp Asp Ser Thr Leu Phe Ala Cys Gln Thr Ser Arg Ala Trp
865 870 875 880

Leu Lys Asn Pro Gly Glu Arg Arg Tyr Ala Ala Val Ser Gly Val Ser
885 890 895

Glu Thr Arg Ser Cys Tyr Ser Val Leu Leu Ser Glu Ala Glu Gly His
900 905 910

Tyr Glu Arg Glu Asn Arg Ile Ser Leu Asp Glu Glu Ala Pro Lys Leu
915 920 925

Ile Val Leu Arg Ala Asp Ser His Glu Glu Ile Leu Gly Arg Leu Asp
930 935 940

Lys Ile Arg Glu Arg Phe Leu Gln Pro Thr Gly Ala Ala Pro Arg Glu
945 950 955 960

Ser Glu Leu Lys Ala Gln Ala Arg Arg Ile Phe Leu Glu Leu Leu Gly
965 970 975

Glu Thr Leu Ala Gln Asp Ala Ala Ser Ser Gly Ser Gln Lys Pro Leu
 980 985 990

Ala Leu Ser Leu Val Ser Thr Pro Ser Lys Leu Gln Arg Glu Val Glu
 995 1000 1005

Leu Ala Ala Lys Gly Ile Pro Arg Cys Leu Lys Met Arg Arg Asp
 1010 1015 1020

Trp Ser Ser Pro Ala Gly Ser Arg Tyr Ala Pro Glu Pro Leu Ala
 1025 1030 1035

Ser Asp Arg Val Ala Phe Met Tyr Gly Glu Gly Arg Ser Pro Tyr
 1040 1045 1050

Tyr Gly Ile Thr Gln Asp Ile His Arg Ile Trp Pro Glu Leu His
 1055 1060 1065

Glu Val Ile Asn Glu Lys Thr Asn Arg Leu Trp Ala Glu Gly Asp
 1070 1075 1080

Arg Trp Val Met Pro Arg Ala Ser Phe Lys Ser Glu Leu Glu Ser
 1085 1090 1095

Gln Gln Gln Glu Phe Asp Arg Asn Met Ile Glu Met Phe Arg Leu
 1100 1105 1110

Gly Ile Leu Thr Ser Ile Ala Phe Thr Asn Leu Ala Arg Asp Val
 1115 1120 1125

Leu Asn Ile Thr Pro Lys Ala Ala Phe Gly Leu Ser Leu Gly Glu
 1130 1135 1140

Ile Ser Met Ile Phe Ala Phe Ser Lys Lys Asn Gly Leu Ile Ser
 1145 1150 1155

Asp Gln Leu Thr Lys Asp Leu Arg Glu Ser Asp Val Trp Asn Lys
 1160 1165 1170

Ala Leu Ala Val Glu Phe Asn Ala Leu Arg Glu Ala Trp Gly Ile
 1175 1180 1185

Pro Gln Ser Val Pro Lys Asp Glu Phe Trp Gln Gly Tyr Ile Val
 1190 1195 1200

Arg Gly Thr Lys Gln Asp Ile Glu Ala Ala Ile Ala Pro Asp Ser
 1205 1210 1215

Lys Tyr Val Arg Leu Thr Ile Ile Asn Asp Ala Asn Thr Ala Leu
 1220 1225 1230

Ile Ser Gly Lys Pro Asp Ala Cys Lys Ala Ala Ile Ala Arg Leu
 1235 1240 1245

Gly Gly Asn Ile Pro Ala Leu Pro Val Thr Gln Gly Met Cys Gly
 1250 1255 1260

His Cys Pro Glu Val Gly Pro Tyr Thr Lys Asp Ile Ala Lys Ile
 1265 1270 1275

His Ala Asn Leu Glu Phe Pro Val Val Asp Gly Leu Asp Leu Trp
 1280 1285 1290

Thr Thr Ile Asn Gln Lys Arg Leu Val Pro Arg Ala Thr Gly Ala
 1295 1300 1305

Lys Asp Glu Trp Ala Pro Ser Ser Phe Gly Glu Tyr Ala Gly Gln
 1310 1315 1320

Leu Tyr Glu Lys Gln Ala Asn Phe Pro Gln Ile Val Glu Thr Ile
 1325 1330 1335

Tyr Lys Gln Asn Tyr Asp Val Phe Val Glu Val Gly Pro Asn Asn
 1340 1345 1350

His Arg Ser Thr Ala Val Arg Thr Thr Leu Gly Pro Gln Arg Asn
 1355 1360 1365

His Leu Ala Gly Ala Ile Asp Lys Gln Asn Glu Asp Ala Trp Thr
 1370 1375 1380

Thr Ile Val Lys Leu Val Ala Ser Leu Lys Ala His Leu Val Pro
 1385 1390 1395

Gly Val Thr Ile Ser Pro Leu Tyr His Ser Lys Leu Val Ala Glu
 1400 1405 1410

Ala Gln Ala Cys Tyr Ala Ala Leu Cys Lys Gly Glu Lys Pro Lys
 1415 1420 1425

Lys Asn Lys Phe Val Arg Lys Ile Gln Leu Asn Gly Arg Phe Asn
 1430 1435 1440

Ser Lys Ala Asp Pro Ile Ser Ser Ala Asp Leu Ala Ser Phe Pro
 1445 1450 1455

Pro Ala Asp Pro Ala Ile Glu Ala Ala Ile Ser Ser Arg Ile Met
 1460 1465 1470

Lys Pro Val Ala Pro Lys Phe Tyr Ala Arg Leu Asn Ile Asp Glu
 1475 1480 1485

Gln Asp Glu Thr Arg Asp Pro Ile Leu Asn Lys Asp Asn Ala Pro
 1490 1495 1500

Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser
 1505 1510 1515

Pro Ser Pro Ala Pro Ser Ala Pro Val Gln Lys Lys Ala Ala Pro
 1520 1525 1530

Ala Ala Glu Thr Lys Ala Val Ala Ser Ala Asp Ala Leu Arg Ser
 1535 1540 1545

Ala Leu Leu Asp Leu Asp Ser Met Leu Ala Leu Ser Ser Ala Ser
 1550 1555 1560

Ala Ser Gly Asn Leu Val Glu Thr Ala Pro Ser Asp Ala Ser Val
 1565 1570 1575

Ile Val Pro Pro Cys Asn Ile Ala Asp Leu Gly Ser Arg Ala Phe
 1580 1585 1590

Met Lys Thr Tyr Gly Val Ser Ala Pro Leu Tyr Thr Gly Ala Met
 1595 1600 1605

Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala Ala Gly Arg
 1610 1615 1620

Gln Gly Ile Leu Ala Ser Phe Gly Ala Gly Gly Leu Pro Met Gln
 1625 1630 1635

Val Val Arg Glu Ser Ile Glu Lys Ile Gln Ala Ala Leu Pro Asn
 1640 1645 1650

Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp Ser Asn
 1655 1660 1665

Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys Gly Val Thr
 1670 1675 1680

Phe Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln Val Val
 1685 1690 1695

Arg Tyr Arg Ala Ala Gly Leu Thr Arg Asn Ala Asp Gly Ser Val
 1700 1705 1710

Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr Glu Leu
 1715 1720 1725

Ala Glu Met Phe Met Arg Pro Ala Pro Glu His Leu Leu Gln Lys
 1730 1735 1740

Leu Ile Ala Ser Gly Glu Ile Asn Gln Glu Gln Ala Glu Leu Ala
 1745 1750 1755

Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala Asp Ser
 1760 1765 1770

Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile Leu Pro Leu
 1775 1780 1785

Ile Ile Asn Leu Arg Asp Arg Leu His Arg Glu Cys Gly Tyr Pro
 1790 1795 1800

Ala Asn Leu Arg Val Arg Val Gly Ala Gly Gly Gly Ile Gly Cys
 1805 1810 1815

Pro Gln Ala Ala Leu Ala Thr Phe Asn Met Gly Ala Ser Phe Ile
 1820 1825 1830

Val Thr Gly Thr Val Asn Gln Val Ala Lys Gln Ser Gly Thr Cys
 1835 1840 1845

Asp Asn Val Arg Lys Gln Leu Ala Lys Ala Thr Tyr Ser Asp Val
 1850 1855 1860

Cys Met Ala Pro Ala Ala Asp Met Phe Glu Glu Gly Val Lys Leu
 1865 1870 1875

Gln Val Leu Lys Lys Gly Thr Met Phe Pro Ser Arg Ala Asn Lys
 1880 1885 1890

Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe Glu Ser Met Pro
 1895 1900 1905

Pro Ala Glu Leu Ala Arg Val Glu Lys Arg Ile Phe Ser Arg Ala
 1910 1915 1920

Leu Glu Glu Val Trp Asp Glu Thr Lys Asn Phe Tyr Ile Asn Arg
 1925 1930 1935

Leu His Asn Pro Glu Lys Ile Gln Arg Ala Glu Arg Asp Pro Lys
 1940 1945 1950

Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Ser Leu Ala Ser
 1955 1960 1965

Arg Trp Ala Asn Thr Gly Ala Ser Asp Arg Val Met Asp Tyr Gln
 1970 1975 1980

Val Trp Cys Gly Pro Ala Ile Gly Ser Phe Asn Asp Phe Ile Lys
 1985 1990 1995

Gly Thr Tyr Leu Asp Pro Ala Val Ala Asn Glu Tyr Pro Cys Val
2000 2005 2010

Val Gln Ile Asn Lys Gln Ile Leu Arg Gly Ala Cys Phe Leu Arg
2015 2020 2025

Arg Leu Glu Ile Leu Arg Asn Ala Arg Leu Ser Asp Gly Ala Ala
2030 2035 2040

Ala Leu Val Ala Ser Ile Asp Asp Thr Tyr Val Pro Ala Glu Lys
2045 2050 2055

Leu

<210> 5 <211> 4509 <212> DNA <213> 裂壺菌屬 sp. <220> <221> CDS <222> (1)..(4509)
<223>

<400> 5

atg gcg ctc cgt gtc aag acg aac aag aag cca tgc tgg gag atg acc 48
Met Ala Leu Arg Val Lys Thr Asn Lys Lys Pro Cys Trp Glu Met Thr
1 5 10 15

aag gag gag ctg acc agc ggc aag acc gag gtg ttc aac tat gag gaa 96
Lys Glu Glu Leu Thr Ser Gly Lys Thr Glu Val Phe Asn Tyr Glu Glu
20 25 30

ctc ctc gag ttc gca gag ggc gac atc gcc aag gtc ttc gga ccc gag 144
Leu Leu Glu Phe Ala Glu Gly Asp Ile Ala Lys Val Phe Gly Pro Glu
35 40 45

ttc gcc gtc atc gac aag tac ccg cgc cgc gtg cgc ctg ccc gcc cgc 192
Phe Ala Val Ile Asp Lys Tyr Pro Arg Arg Val Arg Leu Pro Ala Arg
50 55 60

gag tac ctg ctc gtg acc cgc gtc acc ctc atg gac gcc gag gtc aac 240
Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Asn
65 70 75 80

aac tac cgc gtc ggc gcc cgc atg gtc acc gag tac gat ctc ccc gtc 288
Asn Tyr Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Leu Pro Val
85 90 95

aac gga gag ctc tcc gag ggc gga gac tgc ccc tgg gcc gtc ctg gtc 336
Asn Gly Glu Leu Ser Glu Gly Gly Asp Cys Pro Trp Ala Val Leu Val
100 105 110

gag agt ggc cag tgc gat ctc atg ctc atc tcc tac atg ggc att gac 384
 Glu Ser Gly Gln Cys Asp Leu Met Leu Ile Ser Tyr Met Gly Ile Asp
 115 120 125

ttc cag aac cag ggc gac cgc gtc tac cgc ctg ctc aac acc acg ctc 432
 Phe Gln Asn Gln Gly Asp Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu
 130 135 140

acc ttt tac ggc gtg gcc cac gag ggc gag acc ctc gag tac gac att 480
 Thr Phe Tyr Gly Val Ala His Glu Gly Glu Thr Leu Glu Tyr Asp Ile
 145 150 155 160

cgc gtc acc ggc ttc gcc aag cgt ctc gac ggc ggc atc tcc atg ttc 528
 Arg Val Thr Gly Phe Ala Lys Arg Leu Asp Gly Gly Ile Ser Met Phe
 165 170 175

ttc ttc gag tac gac tgc tac gtc aac ggc cgc ctc ctc atc gag atg 576
 Phe Phe Glu Tyr Asp Cys Tyr Val Asn Gly Arg Leu Leu Ile Glu Met
 180 185 190

cgc gat ggc tgc gcc ggc ttc ttc acc aac gag gag ctc gac gcc ggc 624
 Arg Asp Gly Cys Ala Gly Phe Phe Thr Asn Glu Glu Leu Asp Ala Gly
 195 200 205

aag ggc gtc gtc ttc acc cgc ggc gac ctc gcc gcc cgc gcc aag atc 672
 Lys Gly Val Val Phe Thr Arg Gly Asp Leu Ala Ala Arg Ala Lys Ile
 210 215 220

cca aag cag gac gtc tcc ccc tac gcc gtc gcc ccc tgc ctc cac aag 720
 Pro Lys Gln Asp Val Ser Pro Tyr Ala Val Ala Pro Cys Leu His Lys
 225 230 235 240

acc aag ctc aac gaa aag gag atg cag acc ctc gtc gac aag gac tgg 768
 Thr Lys Leu Asn Glu Lys Glu Met Gln Thr Leu Val Asp Lys Asp Trp
 245 250 255

gca tcc gtc ttt ggc tcc aag aac ggc atg ccg gaa atc aac tac aaa 816
 Ala Ser Val Phe Gly Ser Lys Asn Gly Met Pro Glu Ile Asn Tyr Lys
 260 265 270

ctc tgc gcg cgt aag atg ctc atg att gac cgc gtc acc agc att gac 864
 Leu Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr Ser Ile Asp
 275 280 285

cac aag ggc ggt gtc tac ggc ctc ggt cag ctc gtc ggt gaa aag atc 912
 His Lys Gly Gly Val Tyr Gly Leu Gly Gln Leu Val Gly Glu Lys Ile
 290 295 300

ctc gag cgc gac cac tgg tac ttt ccc tgc cac ttt gtc aag gat cag 960
 Leu Glu Arg Asp His Trp Tyr Phe Pro Cys His Phe Val Lys Asp Gln
 305 310 315 320

gtc atg gcc gga tcc ctc gtc tcc gac ggc tgc agc cag atg ctc aag 1008
 Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Met Leu Lys

325 330 335
 atg tac atg atc tgg ctc ggc ctc cac ctc acc acc gga ccc ttt gac 1056
 Met Tyr Met Ile Trp Leu Gly Leu His Leu Thr Thr Gly Pro Phe Asp
 340 345 350

 ttc cgc ccg gtc aac ggc cac ccc aac aag gtc cgc tgc cgc ggc caa 1104
 Phe Arg Pro Val Asn Gly His Pro Asn Lys Val Arg Cys Arg Gly Gln
 355 360 365

 atc tcc ccg cac aag ggc aag ctc gtc tac gtc atg gag atc aag gag 1152
 Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Glu
 370 375 380

 atg ggc ttc gac gag gac aac gac ccg tac gcc att gcc gac gtc aac 1200
 Met Gly Phe Asp Glu Asp Asn Asp Pro Tyr Ala Ile Ala Asp Val Asn
 385 390 395 400

 atc att gat gtc gac ttc gaa aag ggc cag gac ttt agc ctc gac cgc 1248
 Ile Ile Asp Val Asp Phe Glu Lys Gly Gln Asp Phe Ser Leu Asp Arg
 405 410 415

 atc agc gac tac ggc aag ggc gac ctc aac aag aag atc gtc gtc gac 1296
 Ile Ser Asp Tyr Gly Lys Gly Asp Leu Asn Lys Lys Ile Val Val Asp
 420 425 430

 ttt aag ggc atc gct ctc aag atg cag aag cgc tcc acc aac aag aac 1344
 Phe Lys Gly Ile Ala Leu Lys Met Gln Lys Arg Ser Thr Asn Lys Asn
 435 440 445

 ccc tcc aag gtt cag ccc gtc ttt gcc aac ggc gcc gcc act gtc ggc 1392
 Pro Ser Lys Val Gln Pro Val Phe Ala Asn Gly Ala Ala Thr Val Gly
 450 455 460

 ccc gag gcc tcc aag gct tcc tcc ggc gcc agc gcc agc gcc agc gcc 1440
 Pro Glu Ala Ser Lys Ala Ser Ser Gly Ala Ser Ala Ser Ala Ser Ala
 465 470 475 480

 gcc ccg gcc aag cct gcc ttc agc gcc gat gtt ctt gcg ccc aag ccc 1488
 Ala Pro Ala Lys Pro Ala Phe Ser Ala Asp Val Leu Ala Pro Lys Pro
 485 490 495

 gtt gcc ctt ccc gag cac atc ctc aag ggc gac gcc ctc gcc ccc aag 1536
 Val Ala Leu Pro Glu His Ile Leu Lys Gly Asp Ala Leu Ala Pro Lys
 500 505 510

 gag atg tcc tgg cac ccc atg gcc cgc atc ccg ggc aac ccg acg ccc 1584
 Glu Met Ser Trp His Pro Met Ala Arg Ile Pro Gly Asn Pro Thr Pro
 515 520 525

 tct ttt gcg ccc tgg gcc tac aag ccg cgc aac atc gcc ttt acg ccc 1632
 Ser Phe Ala Pro Ser Ala Tyr Lys Pro Arg Asn Ile Ala Phe Thr Pro
 530 535 540

ttc ccc ggc aac ccc aac gat aac gac cac acc ccg ggc aag atg ccg 1680
 Phe Pro Gly Asn Pro Asn Asp Asn Asp His Thr Pro Gly Lys Met Pro
 545 550 555 560

ctc acc tgg ttc aac atg gcc gag ttc atg gcc ggc aag gtc agc atg 1728
 Leu Thr Trp Phe Asn Met Ala Glu Phe Met Ala Gly Lys Val Ser Met
 565 570 575

tgc ctc ggc ccc gag ttc gcc aag ttc gac gac tcg aac acc agc cgc 1776
 Cys Leu Gly Pro Glu Phe Ala Lys Phe Asp Asp Ser Asn Thr Ser Arg
 580 585 590

agc ccc gct tgg gac ctc gct ctc gtc acc cgc gcc gtg tct gtg tct 1824
 Ser Pro Ala Trp Asp Leu Ala Leu Val Thr Arg Ala Val Ser Val Ser
 595 600 605

gac ctc aag cac gtc aac tac cgc aac atc gac ctc gac ccc tcc aag 1872
 Asp Leu Lys His Val Asn Tyr Arg Asn Ile Asp Leu Asp Pro Ser Lys
 610 615 620

ggc acc atg gtc ggc gag ttc gac tgc ccc gcg gac gcc tgg ttc tac 1920
 Gly Thr Met Val Gly Glu Phe Asp Cys Pro Ala Asp Ala Trp Phe Tyr
 625 630 635 640

aag ggc gcc tgc aac gat gcc cac atg ccg tac tcg atc ctc atg gag 1968
 Lys Gly Ala Cys Asn Asp Ala His Met Pro Tyr Ser Ile Leu Met Glu
 645 650 655

atc gcc ctc cag acc tcg ggt gtg ctc acc tcg gtg ctc aag gcg ccc 2016
 Ile Ala Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro
 660 665 670

ctg acc atg gag aag gac gac atc ctc ttc cgc aac ctc gac gcc aac 2064
 Leu Thr Met Glu Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Asn
 675 680 685

gcc gag ttc gtg cgc gcc gac ctc gac tac cgc ggc aag act atc cgc 2112
 Ala Glu Phe Val Arg Ala Asp Leu Asp Tyr Arg Gly Lys Thr Ile Arg
 690 695 700

aac gtc acc aag tgc act ggc tac agc atg ctc ggc gag atg ggc gtc 2160
 Asn Val Thr Lys Cys Thr Gly Tyr Ser Met Leu Gly Glu Met Gly Val
 705 710 715 720

cac cgc ttc acc ttt gag ctc tac gtc gat gat gtg ctc ttt tac aag 2208
 His Arg Phe Thr Phe Glu Leu Tyr Val Asp Asp Val Leu Phe Tyr Lys
 725 730 735

ggc tcg acc tcg ttc ggc tgg ttc gtg ccc gag gtc ttt gcc gcc cag 2256
 Gly Ser Thr Ser Phe Gly Trp Phe Val Pro Glu Val Phe Ala Ala Gln
 740 745 750

gcc ggc ctc gac aac ggc cgc aag tcg gag ccc tgg ttc att gag aac 2304
 Ala Gly Leu Asp Asn Gly Arg Lys Ser Glu Pro Trp Phe Ile Glu Asn

755 760 765
 aag gtt ccg gcc tcg cag gtc tcc tcc ttt gac gtg cgc ccc aac ggc 2352
 Lys Val Pro Ala Ser Gln Val Ser Ser Phe Asp Val Arg Pro Asn Gly
 770 775 780

 agc ggc cgc acc gcc atc ttc gcc aac gcc ccc agc ggc gcc cag ctc 2400
 Ser Gly Arg Thr Ala Ile Phe Ala Asn Ala Pro Ser Gly Ala Gln Leu
 785 790 795 800

 aac cgc cgc acg gac cag ggc cag tac ctc gac gcc gtc gac att gtc 2448
 Asn Arg Arg Thr Asp Gln Gly Gln Tyr Leu Asp Ala Val Asp Ile Val
 805 810 815

 tcc ggc agc ggc aag aag agc ctc ggc tac gcc cac ggt tcc aag acg 2496
 Ser Gly Ser Gly Lys Lys Ser Leu Gly Tyr Ala His Gly Ser Lys Thr
 820 825 830

 gtc aac ccg aac gac tgg ttc ttc tcg tgc cac ttt tgg ttt gac tcg 2544
 Val Asn Pro Asn Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Ser
 835 840 845

 gtc atg ccc gga agt ctc ggt gtc gag tcc atg ttc cag ctc gtc gag 2592
 Val Met Pro Gly Ser Leu Gly Val Glu Ser Met Phe Gln Leu Val Glu
 850 855 860

 gcc atc gcc gcc cac gag gat ctc gct ggc aaa gca cgg cat tgc caa 2640
 Ala Ile Ala Ala His Glu Asp Leu Ala Gly Lys Ala Arg His Cys Gln
 865 870 875 880

 ccc cac ctt tgt gca cgc ccc cgg gca aga tca agc tgg aag tac cgc 2688
 Pro His Leu Cys Ala Arg Pro Arg Ala Arg Ser Ser Trp Lys Tyr Arg
 885 890 895

 ggc cag ctc acg ccc aag agc aag aag atg gac tcg gag gtc cac atc 2736
 Gly Gln Leu Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile
 900 905 910

 gtg tcc gtg gac gcc cac gac ggc gtt gtc gac ctc gtc gcc gac ggc 2784
 Val Ser Val Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly
 915 920 925

 ttc ctc tgg gcc gac agc ctc cgc gtc tac tcg gtg agc aac att cgc 2832
 Phe Leu Trp Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg
 930 935 940

 gtg cgc atc gcc tcc ggt gag gcc cct gcc gcc gcc tcc tcc gcc gcc 2880
 Val Arg Ile Ala Ser Gly Glu Ala Pro Ala Ala Ala Ser Ser Ala Ala
 945 950 955 960

 tct gtg ggc tcc tcg gct tcg tcc gtc gag cgc acg cgc tcg agc ccc 2928
 Ser Val Gly Ser Ser Ala Ser Ser Val Glu Arg Thr Arg Ser Ser Pro
 965 970 975

gct gtc gcc tcc gcc ccg gcc cag acc atc gac ctc aag cag ctc aag 2976
Ala Val Ala Ser Gly Pro Ala Gln Thr Ile Asp Leu Lys Gln Leu Lys
980 985 990

acc gag ctc ctc gag ctc gat gcc ccg ctc tac ctc tcg cag gac ccg 3024
Thr Glu Leu Leu Glu Leu Asp Ala Pro Leu Tyr Leu Ser Gln Asp Pro
995 1000 1005

acc agc ggc cag ctc aag aag cac acc gac gtg gcc tcc ggc cag 3069
Thr Ser Gly Gln Leu Lys Lys His Thr Asp Val Ala Ser Gly Gln
1010 1015 1020

gcc acc atc gtg cag ccc tgc acg ctc ggc gac ctc ggt gac cgc 3114
Ala Thr Ile Val Gln Pro Cys Thr Leu Gly Asp Leu Gly Asp Arg
1025 1030 1035

tcc ttc atg gag acc tac gcc gtc gtc gcc ccg ctg tac acg gcc 3159
Ser Phe Met Glu Thr Tyr Gly Val Val Ala Pro Leu Tyr Thr Gly
1040 1045 1050

gcc atg gcc aag ggc att gcc tcg gcg gac ctc gtc atc gcc gcc 3204
Ala Met Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala Ala
1055 1060 1065

ggc aag cgc aag atc ctc gcc tcc ttt gcc gcc ggc ggc ctc ccc 3249
Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala Gly Gly Leu Pro
1070 1075 1080

atg cac cac gtg cgc gcc gcc ctc gag aag atc cag gcc gcc ctg 3294
Met His His Val Arg Ala Ala Leu Glu Lys Ile Gln Ala Ala Leu
1085 1090 1095

cct cag ggc ccc tac gcc gtc aac ctc atc cac tcg cct ttt gac 3339
Pro Gln Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp
1100 1105 1110

agc aac ctc gag aag ggc aac gtc gat ctc ttc ctc gag aag ggc 3384
Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys Gly
1115 1120 1125

gtc act gtg gtg gag gcc tcg gca ttc atg acc ctc acc ccg cag 3429
Val Thr Val Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln
1130 1135 1140

gtc gtg cgc tac cgc gcc gcc ggc ctc tcg cgc aac gcc gac ggt 3474
Val Val Arg Tyr Arg Ala Ala Gly Leu Ser Arg Asn Ala Asp Gly
1145 1150 1155

tcg gtc aac atc cgc aac cgc atc atc gcc aag gtc tcg cgc acc 3519
Ser Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr
1160 1165 1170

gag ctc gcc gag atg ttc atc cgc ccg gcc ccg gag cac ctc ctc 3564
Glu Leu Ala Glu Met Phe Ile Arg Pro Ala Pro Glu His Leu Leu

1175	1180	1185	
gag aag ctc atc gcc tcg ggc gag atc acc cag gag cag gcc gag 3609			
Glu Lys Leu Ile Ala Ser Gly Glu Ile Thr Gln Glu Gln Ala Glu			
1190	1195	1200	
ctc cgc cgc cgc gtt ccc gtc gcc gac gat atc get gtc gag get 3654			
Leu Ala Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala			
1205	1210	1215	
gac tcg ggc ggc cac acc gac aac cgc ccc atc cac gtc atc ctc 3699			
Asp Ser Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile Leu			
1220	1225	1230	
ccg ctc atc atc aac ctc cgc aac cgc ctg cac cgc gag tgc ggc 3744			
Pro Leu Ile Ile Asn Leu Arg Asn Arg Leu His Arg Glu Cys Gly			
1235	1240	1245	
tac ccc cgc cac ctc cgc gtc cgc gtt ggc gcc ggc ggt ggc gtc 3789			
Tyr Pro Ala His Leu Arg Val Arg Val Gly Ala Gly Gly Gly Val			
1250	1255	1260	
ggc tgc ccg cag gcc gcc gcc gcc ggc ctc acc atg ggc gcc gcc 3834			
Gly Cys Pro Gln Ala Ala Ala Ala Ala Leu Thr Met Gly Ala Ala			
1265	1270	1275	
ttc atc gtc acc ggc act gtc aac cag gtc gcc aag cag tcc ggc 3879			
Phe Ile Val Thr Gly Thr Val Asn Gln Val Ala Lys Gln Ser Gly			
1280	1285	1290	
acc tgc gac aac gtg cgc aag cag ctc tcg cag gcc acc tac tcg 3924			
Thr Cys Asp Asn Val Arg Lys Gln Leu Ser Gln Ala Thr Tyr Ser			
1295	1300	1305	
gat atc tgc atg gcc ccg gcc gcc gac atg ttc gag gag ggc gtc 3969			
Asp Ile Cys Met Ala Pro Ala Ala Asp Met Phe Glu Glu Gly Val			
1310	1315	1320	
aag ctc cag gtc ctc aag aag gga acc atg ttc ccc tcg cgc gcc 4014			
Lys Leu Gln Val Leu Lys Lys Gly Thr Met Phe Pro Ser Arg Ala			
1325	1330	1335	
aac aag ctc tac gag ctc ttt tgc aag tac gac tcc ttc gac tcc 4059			
Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe Asp Ser			
1340	1345	1350	
atg cct cct gcc gag ctc gag cgc atc gag aag cgt atc ttc aag 4104			
Met Pro Pro Ala Glu Leu Glu Arg Ile Glu Lys Arg Ile Phe Lys			
1355	1360	1365	
cgc gca ctc cag gag gtc tgg gag gag acc aag gac ttt tac att 4149			
Arg Ala Leu Gln Glu Val Trp Glu Glu Thr Lys Asp Phe Tyr Ile			
1370	1375	1380	

aac ggt ctc aag aac ccg gag aag atc cag cgc gcc gag cac gac 4194
 Asn Gly Leu Lys Asn Pro Glu Lys Ile Gln Arg Ala Glu His Asp
 1385 1390 1395

ccc aag ctc aag atg tgg ctc tgc ttc cgc tgg tac ctt ggt ctt 4239
 Pro Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu
 1400 1405 1410

gcc agc cgc tgg gcc aac atg ggc gcc ccg gac cgc gtc atg gac 4284
 Ala Ser Arg Trp Ala Asn Met Gly Ala Pro Asp Arg Val Met Asp
 1415 1420 1425

tac cag gtc tgg tgt ggc ccg gcc att ggc gcc ttc aac gac ttc 4329
 Tyr Gln Val Trp Cys Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe
 1430 1435 1440

atc aag ggc acc tac ctc gac ccc gct gtc tcc aac gag tac ccc 4374
 Ile Lys Gly Thr Tyr Leu Asp Pro Ala Val Ser Asn Glu Tyr Pro
 1445 1450 1455

tgt gtc gtc cag atc aac ctg caa atc ctc cgt ggt gcc tgc tac 4419
 Cys Val Val Gln Ile Asn Leu Gln Ile Leu Arg Gly Ala Cys Tyr
 1460 1465 1470

ctg cgc cgt ctc aac gcc ctg cgc aac gac ccg cgc att gac ctc 4464
 Leu Arg Arg Leu Asn Ala Leu Arg Asn Asp Pro Arg Ile Asp Leu
 1475 1480 1485

gag acc gag gat gct gcc ttt gtc tac gag ccc acc aac gcg ctc 4509
 Glu Thr Glu Asp Ala Ala Phe Val Tyr Glu Pro Thr Asn Ala Leu
 1490 1495 1500

<210> 6 <211> 1503 <212> PRT <213> 裂壺菌屬 sp. <400> 6

Met Ala Leu Arg Val Lys Thr Asn Lys Lys Pro Cys Trp Glu Met Thr
 1 5 10 15

Lys Glu Glu Leu Thr Ser Gly Lys Thr Glu Val Phe Asn Tyr Glu Glu
 20 25 30

Leu Leu Glu Phe Ala Glu Gly Asp Ile Ala Lys Val Phe Gly Pro Glu
 35 40 45

Phe Ala Val Ile Asp Lys Tyr Pro Arg Arg Val Arg Leu Pro Ala Arg
 50 55 60

Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Asn
 65 70 75 80

Asn Tyr Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Leu Pro Val
85 90 95

Asn Gly Glu Leu Ser Glu Gly Gly Asp Cys Pro Trp Ala Val Leu Val
100 105 110

Glu Ser Gly Gln Cys Asp Leu Met Leu Ile Ser Tyr Met Gly Ile Asp
115 120 125

Phe Gln Asn Gln Gly Asp Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu
130 135 140

Thr Phe Tyr Gly Val Ala His Glu Gly Glu Thr Leu Glu Tyr Asp Ile
145 150 155 160

Arg Val Thr Gly Phe Ala Lys Arg Leu Asp Gly Gly Ile Ser Met Phe
165 170 175

Phe Phe Glu Tyr Asp Cys Tyr Val Asn Gly Arg Leu Leu Ile Glu Met
180 185 190

Arg Asp Gly Cys Ala Gly Phe Phe Thr Asn Glu Glu Leu Asp Ala Gly
195 200 205

Lys Gly Val Val Phe Thr Arg Gly Asp Leu Ala Ala Arg Ala Lys Ile
210 215 220

Pro Lys Gln Asp Val Ser Pro Tyr Ala Val Ala Pro Cys Leu His Lys
225 230 235 240

Thr Lys Leu Asn Glu Lys Glu Met Gln Thr Leu Val Asp Lys Asp Trp
245 250 255

Ala Ser Val Phe Gly Ser Lys Asn Gly Met Pro Glu Ile Asn Tyr Lys
260 265 270

Leu Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr Ser Ile Asp
275 280 285

His Lys Gly Gly Val Tyr Gly Leu Gly Gln Leu Val Gly Glu Lys Ile

290 295 300
 Leu Glu Arg Asp His Trp Tyr Phe Pro Cys His Phe Val Lys Asp Gln
 305 310 315 320
 Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Met Leu Lys
 325 330 335
 Met Tyr Met Ile Trp Leu Gly Leu His Leu Thr Thr Gly Pro Phe Asp
 340 345 350
 Phe Arg Pro Val Asn Gly His Pro Asn Lys Val Arg Cys Arg Gly Gln
 355 360 365
 Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Glu
 370 375 380
 Met Gly Phe Asp Glu Asp Asn Asp Pro Tyr Ala Ile Ala Asp Val Asn
 385 390 395 400
 Ile Ile Asp Val Asp Phe Glu Lys Gly Gln Asp Phe Ser Leu Asp Arg
 405 410 415
 Ile Ser Asp Tyr Gly Lys Gly Asp Leu Asn Lys Lys Ile Val Val Asp
 420 425 430
 Phe Lys Gly Ile Ala Leu Lys Met Gln Lys Arg Ser Thr Asn Lys Asn
 435 440 445
 Pro Ser Lys Val Gln Pro Val Phe Ala Asn Gly Ala Ala Thr Val Gly
 450 455 460
 Pro Glu Ala Ser Lys Ala Ser Ser Gly Ala Ser Ala Ser Ala Ser Ala
 465 470 475 480
 Ala Pro Ala Lys Pro Ala Phe Ser Ala Asp Val Leu Ala Pro Lys Pro
 485 490 495
 Val Ala Leu Pro Glu His Ile Leu Lys Gly Asp Ala Leu Ala Pro Lys
 500 505 510

Glu Met Ser Trp His Pro Met Ala Arg Ile Pro Gly Asn Pro Thr Pro
 515 520 525

Ser Phe Ala Pro Ser Ala Tyr Lys Pro Arg Asn Ile Ala Phe Thr Pro
 530 535 540

Phe Pro Gly Asn Pro Asn Asp Asn Asp His Thr Pro Gly Lys Met Pro
 545 550 555 560

Leu Thr Trp Phe Asn Met Ala Glu Phe Met Ala Gly Lys Val Ser Met
 565 570 575

Cys Leu Gly Pro Glu Phe Ala Lys Phe Asp Asp Ser Asn Thr Ser Arg
 580 585 590

Ser Pro Ala Trp Asp Leu Ala Leu Val Thr Arg Ala Val Ser Val Ser
 595 600 605

Asp Leu Lys His Val Asn Tyr Arg Asn Ile Asp Leu Asp Pro Ser Lys
 610 615 620

Gly Thr Met Val Gly Glu Phe Asp Cys Pro Ala Asp Ala Trp Phe Tyr
 625 630 635 640

Lys Gly Ala Cys Asn Asp Ala His Met Pro Tyr Ser Ile Leu Met Glu
 645 650 655

Ile Ala Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro
 660 665 670

Leu Thr Met Glu Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Asn
 675 680 685

Ala Glu Phe Val Arg Ala Asp Leu Asp Tyr Arg Gly Lys Thr Ile Arg
 690 695 700

Asn Val Thr Lys Cys Thr Gly Tyr Ser Met Leu Gly Glu Met Gly Val
 705 710 715 720

His Arg Phe Thr Phe Glu Leu Tyr Val Asp Asp Val Leu Phe Tyr Lys

725 730 735

Gly Ser Thr Ser Phe Gly Trp Phe Val Pro Glu Val Phe Ala Ala Gln
740 745 750

Ala Gly Leu Asp Asn Gly Arg Lys Ser Glu Pro Trp Phe Ile Glu Asn
755 760 765

Lys Val Pro Ala Ser Gln Val Ser Ser Phe Asp Val Arg Pro Asn Gly
770 775 780

Ser Gly Arg Thr Ala Ile Phe Ala Asn Ala Pro Ser Gly Ala Gln Leu
785 790 795 800

Asn Arg Arg Thr Asp Gln Gly Gln Tyr Leu Asp Ala Val Asp Ile Val
805 810 815

Ser Gly Ser Gly Lys Lys Ser Leu Gly Tyr Ala His Gly Ser Lys Thr
820 825 830

Val Asn Pro Asn Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Ser
835 840 845

Val Met Pro Gly Ser Leu Gly Val Glu Ser Met Phe Gln Leu Val Glu
850 855 860

Ala Ile Ala Ala His Glu Asp Leu Ala Gly Lys Ala Arg His Cys Gln
865 870 875 880

Pro His Leu Cys Ala Arg Pro Arg Ala Arg Ser Ser Trp Lys Tyr Arg
885 890 895

Gly Gln Leu Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile
900 905 910

Val Ser Val Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly
915 920 925

Phe Leu Trp Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg
930 935 940

Val Arg Ile Ala Ser Gly Glu Ala Pro Ala Ala Ala Ser Ser Ala Ala
 945 950 955 960

Ser Val Gly Ser Ser Ala Ser Ser Val Glu Arg Thr Arg Ser Ser Pro
 965 970 975

Ala Val Ala Ser Gly Pro Ala Gln Thr Ile Asp Leu Lys Gln Leu Lys
 980 985 990

Thr Glu Leu Leu Glu Leu Asp Ala Pro Leu Tyr Leu Ser Gln Asp Pro
 995 1000 1005

Thr Ser Gly Gln Leu Lys Lys His Thr Asp Val Ala Ser Gly Gln
 1010 1015 1020

Ala Thr Ile Val Gln Pro Cys Thr Leu Gly Asp Leu Gly Asp Arg
 1025 1030 1035

Ser Phe Met Glu Thr Tyr Gly Val Val Ala Pro Leu Tyr Thr Gly
 1040 1045 1050

Ala Met Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala Ala
 1055 1060 1065

Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala Gly Gly Leu Pro
 1070 1075 1080

Met His His Val Arg Ala Ala Leu Glu Lys Ile Gln Ala Ala Leu
 1085 1090 1095

Pro Gln Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp
 1100 1105 1110

Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys Gly
 1115 1120 1125

Val Thr Val Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln
 1130 1135 1140

Val Val Arg Tyr Arg Ala Ala Gly Leu Ser Arg Asn Ala Asp Gly

1145	1150	1155
Ser Val	Asn Ile Arg Asn Arg	Ile Ile Gly Lys Val Ser Arg Thr
1160	1165	1170
Glu Leu	Ala Glu Met Phe Ile	Arg Pro Ala Pro Glu His Leu Leu
1175	1180	1185
Glu Lys	Leu Ile Ala Ser Gly	Glu Ile Thr Gln Glu Gln Ala Glu
1190	1195	1200
Leu Ala	Arg Arg Val Pro Val	Ala Asp Asp Ile Ala Val Glu Ala
1205	1210	1215
Asp Ser	Gly Gly His Thr Asp	Asn Arg Pro Ile His Val Ile Leu
1220	1225	1230
Pro Leu	Ile Ile Asn Leu Arg	Asn Arg Leu His Arg Glu Cys Gly
1235	1240	1245
Tyr Pro	Ala His Leu Arg Val	Arg Val Gly Ala Gly Gly Gly Val
1250	1255	1260
Gly Cys	Pro Gln Ala Ala Ala	Ala Ala Leu Thr Met Gly Ala Ala
1265	1270	1275
Phe Ile	Val Thr Gly Thr Val	Asn Gln Val Ala Lys Gln Ser Gly
1280	1285	1290
Thr Cys	Asp Asn Val Arg Lys	Gln Leu Ser Gln Ala Thr Tyr Ser
1295	1300	1305
Asp Ile	Cys Met Ala Pro Ala	Ala Asp Met Phe Glu Glu Gly Val
1310	1315	1320
Lys Leu	Gln Val Leu Lys Lys	Gly Thr Met Phe Pro Ser Arg Ala
1325	1330	1335
Asn Lys	Leu Tyr Glu Leu Phe	Cys Lys Tyr Asp Ser Phe Asp Ser
1340	1345	1350

Met Pro Pro Ala Glu Leu Glu Arg Ile Glu Lys Arg Ile Phe Lys
 1355 1360 1365

Arg Ala Leu Gln Glu Val Trp Glu Glu Thr Lys Asp Phe Tyr Ile
 1370 1375 1380

Asn Gly Leu Lys Asn Pro Glu Lys Ile Gln Arg Ala Glu His Asp
 1385 1390 1395

Pro Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu
 1400 1405 1410

Ala Ser Arg Trp Ala Asn Met Gly Ala Pro Asp Arg Val Met Asp
 1415 1420 1425

Tyr Gln Val Trp Cys Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe
 1430 1435 1440

Ile Lys Gly Thr Tyr Leu Asp Pro Ala Val Ser Asn Glu Tyr Pro
 1445 1450 1455

Cys Val Val Gln Ile Asn Leu Gln Ile Leu Arg Gly Ala Cys Tyr
 1460 1465 1470

Leu Arg Arg Leu Asn Ala Leu Arg Asn Asp Pro Arg Ile Asp Leu
 1475 1480 1485

Glu Thr Glu Asp Ala Ala Phe Val Tyr Glu Pro Thr Asn Ala Leu
 1490 1495 1500

<210> 7 <211> 600 <212> DNA <213> 裂壺菌屬 sp. <220> <221> CDS <222> (1)..(600)
 <223>

<400> 7

atg gcg gcc cgt ctg cag gag caa aag gga ggc gag atg gat acc cgc 48
 Met Ala Ala Arg Leu Gln Glu Gln Lys Gly Gly Glu Met Asp Thr Arg
 1 5 10 15

att gcc atc atc ggc atg tgg gcc atc ctg ccc tgc ggc acg acc gtg 96
 Ile Ala Ile Ile Gly Met Ser Ala Ile Leu Pro Cys Gly Thr Thr Val
 20 25 30

cgc gag tgg tgg gag acc atc cgc gcc ggc atc gac tgc ctg tgg gat 144

Arg Glu Ser Trp Glu Thr Ile Arg Ala Gly Ile Asp Cys Leu Ser Asp
 35 40 45

ctc ccc gag gac cgc gtc gac gtg acg gcg tac ttt gac ccc gtc aag 192
 Leu Pro Glu Asp Arg Val Asp Val Thr Ala Tyr Phe Asp Pro Val Lys
 50 55 60

acc acc aag gac aag atc tac tgc aag cgc ggt ggc ttc att ccc gag 240
 Thr Thr Lys Asp Lys Ile Tyr Cys Lys Arg Gly Gly Phe Ile Pro Glu
 65 70 75 80

tac gac ttt gac gcc cgc gag ttc gga ctc aac atg ttc cag atg gag 288
 Tyr Asp Phe Asp Ala Arg Glu Phe Gly Leu Asn Met Phe Gln Met Glu
 85 90 95

gac tcg gac gca aac cag acc atc tcg ctt ctc aag gtc aag gag gcc 336
 Asp Ser Asp Ala Asn Gln Thr Ile Ser Leu Leu Lys Val Lys Glu Ala
 100 105 110

ctc cag gac gcc ggc atc gac gcc ctc ggc aag gaa aag aag aac atc 384
 Leu Gln Asp Ala Gly Ile Asp Ala Leu Gly Lys Glu Lys Lys Asn Ile
 115 120 125

ggc tgc gtg ctc ggc att ggc ggc ggc caa aag tcc agc cac gag ttc 432
 Gly Cys Val Leu Gly Ile Gly Gly Gly Gln Lys Ser Ser His Glu Phe
 130 135 140

tac tcg cgc ctt aat tat gtt gtc gtg gag aag gtc ctc cgc aag atg 480
 Tyr Ser Arg Leu Asn Tyr Val Val Val Glu Lys Val Leu Arg Lys Met
 145 150 155 160

ggc atg ccc gag gag gac gtc aag gtc gcc gtc gaa aag tac aag gcc 528
 Gly Met Pro Glu Glu Asp Val Lys Val Ala Val Glu Lys Tyr Lys Ala
 165 170 175

aac ttc ccc gag tgg cgc ctc gac tcc ttc cct ggc ttc ctc ggc aac 576
 Asn Phe Pro Glu Trp Arg Leu Asp Ser Phe Pro Gly Phe Leu Gly Asn
 180 185 190

gtc acc gcc ggt cgc tgc acc aac 600
 Val Thr Ala Gly Arg Cys Thr Asn
 195 200

<210> 8 <211> 200 <212> PRT <213> 裂壺菌屬 sp. <400> 8

Met Ala Ala Arg Leu Gln Glu Gln Lys Gly Gly Glu Met Asp Thr Arg
 1 5 10 15

Ile Ala Ile Ile Gly Met Ser Ala Ile Leu Pro Cys Gly Thr Thr Val
 20 25 30

Arg Glu Ser Trp Glu Thr Ile Arg Ala Gly Ile Asp Cys Leu Ser Asp
35 40 45

Leu Pro Glu Asp Arg Val Asp Val Thr Ala Tyr Phe Asp Pro Val Lys
50 55 60

Thr Thr Lys Asp Lys Ile Tyr Cys Lys Arg Gly Gly Phe Ile Pro Glu
65 70 75 80

Tyr Asp Phe Asp Ala Arg Glu Phe Gly Leu Asn Met Phe Gln Met Glu
85 90 95

Asp Ser Asp Ala Asn Gln Thr Ile Ser Leu Leu Lys Val Lys Glu Ala
100 105 110

Leu Gln Asp Ala Gly Ile Asp Ala Leu Gly Lys Glu Lys Lys Asn Ile
115 120 125

Gly Cys Val Leu Gly Ile Gly Gly Gly Gln Lys Ser Ser His Glu Phe
130 135 140

Tyr Ser Arg Leu Asn Tyr Val Val Val Glu Lys Val Leu Arg Lys Met
145 150 155 160

Gly Met Pro Glu Glu Asp Val Lys Val Ala Val Glu Lys Tyr Lys Ala
165 170 175

Asn Phe Pro Glu Trp Arg Leu Asp Ser Phe Pro Gly Phe Leu Gly Asn
180 185 190

Val Thr Ala Gly Arg Cys Thr Asn
195 200

<210> 9 <211> 1278 <212> DNA <213> 裂壺菌屬 sp. <220> <221> CDS <222> (1)..(1278)
<223>

<400> 9

gat gtc acc aag gag gcc tgg cgc ctc ccc cgc gag ggc gtc agc ttc 48
Asp Val Thr Lys Glu Ala Trp Arg Leu Pro Arg Glu Gly Val Ser Phe
1 5 10 15

cgc gcc aag ggc atc gcc acc aac ggc gct gtc gcc cgc ctc ttc tcc 96
Arg Ala Lys Gly Ile Ala Thr Asn Gly Ala Val Ala Ala Leu Phe Ser

20 25 30
 ggc cag ggc gcg cag tac acg cac atg ttt agc gag gtg gcc atg aac 144
 Gly Gln Gly Ala Gln Tyr Thr His Met Phe Ser Glu Val Ala Met Asn
 35 40 45
 tgg ccc cag ttc cgc cag agc att gcc gcc atg gac gcc gcc cag tcc 192
 Trp Pro Gln Phe Arg Gln Ser Ile Ala Ala Met Asp Ala Ala Gln Ser
 50 55 60
 aag gtc gct gga agc gac aag gac ttt gag cgc gtc tcc cag gtc ctc 240
 Lys Val Ala Gly Ser Asp Lys Asp Phe Glu Arg Val Ser Gln Val Leu
 65 70 75 80
 tac ccg cgc aag ccg tac gag cgt gag ccc gag cag aac ccc aag aag 288
 Tyr Pro Arg Lys Pro Tyr Glu Arg Glu Pro Glu Gln Asn Pro Lys Lys
 85 90 95
 atc tcc ctc acc gcc tac tcg cag ccc tcg acc ctg gcc tgc gct ctc 336
 Ile Ser Leu Thr Ala Tyr Ser Gln Pro Ser Thr Leu Ala Cys Ala Leu
 100 105 110
 ggt gcc ttt gag atc ttc aag gag gcc ggc ttc acc ccg gac ttt gcc 384
 Gly Ala Phe Glu Ile Phe Lys Glu Ala Gly Phe Thr Pro Asp Phe Ala
 115 120 125
 gcc ggc cat tcg ctc ggt gag ttc gcc gcc ctc tac gcc gcg ggc tgc 432
 Ala Gly His Ser Leu Gly Glu Phe Ala Ala Leu Tyr Ala Ala Gly Cys
 130 135 140
 gtc gac cgc gac gag ctc ttt gag ctt gtc tgc cgc cgc gcc cgc atc 480
 Val Asp Arg Asp Glu Leu Phe Glu Leu Val Cys Arg Arg Ala Arg Ile
 145 150 155 160
 atg ggc ggc aag gac gca ccg gcc acc ccc aag gga tgc atg gcc gcc 528
 Met Gly Gly Lys Asp Ala Pro Ala Thr Pro Lys Gly Cys Met Ala Ala
 165 170 175
 gtc att gcc ccc aac gcc gag aac atc aag gtc cag gcc gcc aac gtc 576
 Val Ile Gly Pro Asn Ala Glu Asn Ile Lys Val Gln Ala Ala Asn Val
 180 185 190
 tgg ctc ggc aac tcc aac tcg cct tcg cag acc gtc atc acc ggc tcc 624
 Trp Leu Gly Asn Ser Asn Ser Pro Ser Gln Thr Val Ile Thr Gly Ser
 195 200 205
 gtc gaa ggt atc cag gcc gag agc gcc cgc ctc cag aag gag ggc ttc 672
 Val Glu Gly Ile Gln Ala Glu Ser Ala Arg Leu Gln Lys Glu Gly Phe
 210 215 220
 cgc gtc gtg cct ctt gcc tgc gag agc gcc ttc cac tcg ccc cag atg 720
 Arg Val Val Pro Leu Ala Cys Glu Ser Ala Phe His Ser Pro Gln Met
 225 230 235 240

gag aac gcc tcg tcg gcc ttc aag gac gtc atc tcc aag gtc tcc ttc 768
 Glu Asn Ala Ser Ser Ala Phe Lys Asp Val Ile Ser Lys Val Ser Phe
 245 250 255

cgc acc ccc aag gcc gag acc aag etc ttc agc aac gtc tct ggc gag 816
 Arg Thr Pro Lys Ala Glu Thr Lys Leu Phe Ser Asn Val Ser Gly Glu
 260 265 270

acc tac ccc acg gac gcc cgc gag atg ctt acg cag cac atg acc agc 864
 Thr Tyr Pro Thr Asp Ala Arg Glu Met Leu Thr Gln His Met Thr Ser
 275 280 285

agc gtc aag ttc ctc acc cag gtc cgc aac atg cac cag gcc ggt gcg 912
 Ser Val Lys Phe Leu Thr Gln Val Arg Asn Met His Gln Ala Gly Ala
 290 295 300

cgc atc ttt gtc gag ttc gga ccc aag cag gtg ctc tcc aag ctt gtc 960
 Arg Ile Phe Val Glu Phe Gly Pro Lys Gln Val Leu Ser Lys Leu Val
 305 310 315 320

tcc gag acc ctc aag gat gac ccc tcg gtt gtc acc gtc tct gtc aac 1008
 Ser Glu Thr Leu Lys Asp Asp Pro Ser Val Val Thr Val Ser Val Asn
 325 330 335

cgc gcc tcg ggc acg gat tcg gac atc cag ctc cgc gac gcg gcc gtc 1056
 Pro Ala Ser Gly Thr Asp Ser Asp Ile Gln Leu Arg Asp Ala Ala Val
 340 345 350

cag ctc gtt gtc gct ggc gtc aac ctt cag ggc ttt gac aag tgg gac 1104
 Gln Leu Val Val Ala Gly Val Asn Leu Gln Gly Phe Asp Lys Trp Asp
 355 360 365

gcc ccc gat gcc acc cgc atg cag gcc atc aag aag aag cgc act acc 1152
 Ala Pro Asp Ala Thr Arg Met Gln Ala Ile Lys Lys Lys Arg Thr Thr
 370 375 380

ctc cgc ctt tcg gcc gcc acc tac gtc tcg gac aag acc aag aag gtc 1200
 Leu Arg Leu Ser Ala Ala Thr Tyr Val Ser Asp Lys Thr Lys Lys Val
 385 390 395 400

cgc gac gcc gcc atg aac gat ggc cgc tgc gtc acc tac ctc aag gcc 1248
 Arg Asp Ala Ala Met Asn Asp Gly Arg Cys Val Thr Tyr Leu Lys Gly
 405 410 415

gcc gca ccg ctc atc aag gcc ccg gag ccc 1278
 Ala Ala Pro Leu Ile Lys Ala Pro Glu Pro
 420 425

<210> 10 <211> 426 <212> PRT <213> 裂壺菌屬 sp. <400> 10

Asp Val Thr Lys Glu Ala Trp Arg Leu Pro Arg Glu Gly Val Ser Phe
 1 5 10 15

Arg Ala Lys Gly Ile Ala Thr Asn Gly Ala Val Ala Ala Leu Phe Ser
 20 25 30

Gly Gln Gly Ala Gln Tyr Thr His Met Phe Ser Glu Val Ala Met Asn
 35 40 45

Trp Pro Gln Phe Arg Gln Ser Ile Ala Ala Met Asp Ala Ala Gln Ser
 50 55 60

Lys Val Ala Gly Ser Asp Lys Asp Phe Glu Arg Val Ser Gln Val Leu
 65 70 75 80

Tyr Pro Arg Lys Pro Tyr Glu Arg Glu Pro Glu Gln Asn Pro Lys Lys
 85 90 95

Ile Ser Leu Thr Ala Tyr Ser Gln Pro Ser Thr Leu Ala Cys Ala Leu
 100 105 110

Gly Ala Phe Glu Ile Phe Lys Glu Ala Gly Phe Thr Pro Asp Phe Ala
 115 120 125

Ala Gly His Ser Leu Gly Glu Phe Ala Ala Leu Tyr Ala Ala Gly Cys
 130 135 140

Val Asp Arg Asp Glu Leu Phe Glu Leu Val Cys Arg Arg Ala Arg Ile
 145 150 155 160

Met Gly Gly Lys Asp Ala Pro Ala Thr Pro Lys Gly Cys Met Ala Ala
 165 170 175

Val Ile Gly Pro Asn Ala Glu Asn Ile Lys Val Gln Ala Ala Asn Val
 180 185 190

Trp Leu Gly Asn Ser Asn Ser Pro Ser Gln Thr Val Ile Thr Gly Ser
 195 200 205

Val Glu Gly Ile Gln Ala Glu Ser Ala Arg Leu Gln Lys Glu Gly Phe
 210 215 220

Arg Val Val Pro Leu Ala Cys Glu Ser Ala Phe His Ser Pro Gln Met

225 230 235 240
 Glu Asn Ala Ser Ser Ala Phe Lys Asp Val Ile Ser Lys Val Ser Phe
 245 250 255
 Arg Thr Pro Lys Ala Glu Thr Lys Leu Phe Ser Asn Val Ser Gly Glu
 260 265 270
 Thr Tyr Pro Thr Asp Ala Arg Glu Met Leu Thr Gln His Met Thr Ser
 275 280 285
 Ser Val Lys Phe Leu Thr Gln Val Arg Asn Met His Gln Ala Gly Ala
 290 295 300
 Arg Ile Phe Val Glu Phe Gly Pro Lys Gln Val Leu Ser Lys Leu Val
 305 310 315 320
 Ser Glu Thr Leu Lys Asp Asp Pro Ser Val Val Thr Val Ser Val Asn
 325 330 335
 Pro Ala Ser Gly Thr Asp Ser Asp Ile Gln Leu Arg Asp Ala Ala Val
 340 345 350
 Gln Leu Val Val Ala Gly Val Asn Leu Gln Gly Phe Asp Lys Trp Asp
 355 360 365
 Ala Pro Asp Ala Thr Arg Met Gln Ala Ile Lys Lys Lys Arg Thr Thr
 370 375 380
 Leu Arg Leu Ser Ala Ala Thr Tyr Val Ser Asp Lys Thr Lys Lys Val
 385 390 395 400
 Arg Asp Ala Ala Met Asn Asp Gly Arg Cys Val Thr Tyr Leu Lys Gly
 405 410 415
 Ala Ala Pro Leu Ile Lys Ala Pro Glu Pro
 420 425

<210> 11 <211> 5 <212> PRT <213> 裂壺菌屬 sp. <220> <221> MISC_FEATURE <222> (4).
 (4) <223> X = 任何胺基酸

<400> 11

Gly His Ser Xaa Gly
1 5

<210> 12 <211> 258 <212> DNA <213> 裂壺菌屬 sp. <220> <221> CDS <222> (1)..(258)
<223>

<400> 12

gct gtc tcg aac gag ctt ctt gag aag gcc gag act gtc gtc atg gag 48
Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu Thr Val Val Met Glu
1 5 10 15

gtc ctc gcc gcc aag acc ggc tac gag acc gac atg atc gag gct gac 96
Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile Glu Ala Asp
20 25 30

atg gag ctc gag acc gag ctc ggc att gac tcc atc aag cgt gtc gag 144
Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu
35 40 45

atc ctc tcc gag gtc cag gcc atg ctc aat gtc gag gcc aag gat gtc 192
Ile Leu Ser Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val
50 55 60

gat gcc ctc agc cgc act cgc act gtt ggt gag gtt gtc aac gcc atg 240
Asp Ala Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val Asn Ala Met
65 70 75 80

aag gcc gag atc gct ggc 258
Lys Ala Glu Ile Ala Gly
85

<210> 13 <211> 86 <212> PRT <213> 裂壺菌屬 sp. <400> 13

Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu Thr Val Val Met Glu
1 5 10 15

Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile Glu Ala Asp
20 25 30

Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu
35 40 45

Ile Leu Ser Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val
50 55 60

Asp Ala Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val Asn Ala Met

65 70 75 80

Lys Ala Glu Ile Ala Gly
85

<210> 14 <211> 5 <212> PRT <213> 裂壺菌屬 sp. <400> 14

Leu Gly Ile Asp Ser
1 5

<210> 15 <211> 21 <212> PRT <213> 裂壺菌屬 sp. <400> 15

Ala Pro Ala Pro Val Lys Ala Ala Ala Pro Ala Ala Pro Val Ala Ser
1 5 10 15

Ala Pro Ala Pro Ala
20

<210> 16 <211> 3006 <212> DNA <213> 裂壺菌屬 sp. <400> 16
gcccccgccc cggtaaggc tgcctgcct gccgccccg ttgcctcggc cctcgcctcc 60

gctgtctcga acgagcttct tgagaaggcc gagactgtcg tcatggaggt cctcgcctcc 120

aagaccggct acgagaccga catgatcgag gctgacatgg agctcgagac cgagctcggc 180

attgactcca tcaagcgtgt cgagatcctc tccgaggtcc aggccatgct caatgtcgag 240

gccaaagatg tcatgacctc cagccgcact cgcactgttg gtgaggtgt caacgccatg 300

aagcccgaga tgcctggcag ctctgccccg gcgctcctg ccctcgtcc ggtccggcc 360

aagctgccc ctgccccgc tgcctcgtgt gtctcgaac agcttctcga gaagccgag 420

accgtcgtca tggaggtcct cggcccaag actggctacg agactgacat gatcgagtcc 480

gacatggagc tcgagactga gctcggcatt gactccatca agcgtgtcga gatcctctcc 540

gaggctcagg ccatgctcaa cgtcgaggcc aaggacgtcg acgtctcag ccgcactcgc 600

actgtgggtg aggtgtcaa cgcctgaag gctgagatcg ctggggctc tccccggcg 660

cctgcccggc ctgccccagg tccggctgct gccgcccctg cgctcggcc cgcgcccct 720

gctgtctcga acgagcttct tgagaaggcc gagaccgtcg tcatggaggt cctcgcctcc 780

aagactggct acgagactga catgatcgag tccgacatgg agctcgagac cgagctcggc 840

attgactcca tcaagcgtgt cgagattctc tccgaggtcc aggccatgct caactcgag 900

gccaaggacg tcgacgtct cagccgcacc cgcactgtg gcgaggcgt cgalccatg 960
aaggccgaga tcgctggtg cctgccccg gcgcctgcc ccgctgtcc tgcctggct 1020
gctgccccc ctgcctgc cgcctcgc cctgctct cgcgcgact tctcagaag 1080
gccgagactg tcgcatga ggtcctgcc gccaaactg gctacgagac tgacatgac 1140
gagtcgaca tggagctga gaccgagct ggcaatgact ccaalcaagc tctcagatt 1200
ctctccgagg tccaggccat gctcaacgc gaggccaagg acgtcagac tctcagccg 1260
acccgcactg ttggcgagg ctctgagcc atgaaggcc agatcgtgg tgctctgcc 1320
ccggcgctg ccggcgtg tctgtctc gctgctgcc cccctgcgc tgcggccct 1380
gcgcctgcc cccctgcgc tctgtctc agcgagctc tcgagaagg cgagactgc 1440
gtatggagg tctcggcc caagactgc tacgagact acatgattg gtccgacatg 1500
gagctcaga ccgagctcg caatgactc atcaagcgt tcgagattc ctccgaggt 1560
caggccatg tcaactga gcccaaggac gtcgacgctc tcagccgac tcgactgtt 1620
ggtgagctg tcgagccat gaaggctgag atcgtggca gctccgctc ggccctgcc 1680
gccgctgc ctctccgc tctgcccct cctcggccc ctgcccgc cctgtctc 1740
tcgaacgagc ttctcagaa agccgagact gtcgtcagc aggtcctgc cgccaagact 1800
ggctacgaga ctgacatg cgcgctcc atggagctc agactgagc cggcattgac 1860
tccalcaagc gtctcagat cctctccgag gttcagcca tctcaactg cgaagccaag 1920
gacgtcagc cctcagcc caccgcact gttggcagg ttctcagc catgaaggcc 1980
gagatcgtg gtgctctc cccggcct gccggcctg cccctgtcc ggtgccc 2040
gcccctgtg tctcgaaga gctctcagc aaggccgaga ctgtctcat ggagctctc 2100
gccccaaga ctggctacga gaccgacatg atcgagctc acatggagc cgagaccgag 2160
ctcggcattg actccalca gctgtcagc attctctc aggtcagc catgctaac 2220
gtcagggcca aggacgtga tctctcagc cgcactcga ctgtggcga ggtctctgac 2280
gccalgaagg ctgagatgc cggcagctc gcccggcgc ctgcccgc tctctctgct 2340
ccggctctg ccgctctc gccgctgc gctgcccctg ctgtctcag cgagctctc 2400
gagaaggccg agaccgtct catggaggtc ctgcccga agactggcta cgagactgac 2460
atgattgagc ccgacatga gctcagact gactcggca ttactccat caagcgtgc 2520

gagatcctct ccgaggttca ggccalgtct aacgtcgagg ccaaggacgt cgaigccctc 2580
 agccgcaccc gcactgttgg cgaggtgtc gaigccatga aggccagat cgtgtgtgc 2640
 tctgccccgg cgctgccgc cgtgccccct gctccggctg ccgcccccc tctgtctc 2700
 aacgagcttc ttgagaaggc cgagaccgtc gtcaltgagg toctgccgc caagactggc 2760
 tacgagaccg acalgtatga gtccgacatg gagctcgaga ccgagctcgg cattgactcc 2820
 atcaagcgtg tcgagattct ctccgaggtt cagcccatgc tcaacgtcga ggccaaggac 2880
 gtcgacgtc tcagccgcac tcgcaactgt ggcgagctc tcgatccat gaaggctgag 2940
 atcgtgtgtg gctctgcccc ggcgcccccc gccgctgtc ctgctcggc tggcggcgc 3000
 cctgcg 3006

<210> 17 <211> 2133 <212> DNA <213> 裂壺菌屬 sp. <220> <221> CDS <222> (1)..(2133)
<223>

<400> 17

ttt ggc gct ctc ggc ggc ttc atc tcg cag cag gcg gag cgc ttc gag 48
 Phe Gly Ala Leu Gly Gly Phe Ile Ser Gln Gln Ala Glu Arg Phe Glu
 1 5 10 15

ccc gcc gaa atc ctc ggc ttc acg ctc atg tgc gcc aag ttc gcc aag 96
 Pro Ala Glu Ile Leu Gly Phe Thr Leu Met Cys Ala Lys Phe Ala Lys
 20 25 30

gct tcc ctc tgc acg gct gtg gct ggc ggc cgc ccg gcc ttt atc ggt 144
 Ala Ser Leu Cys Thr Ala Val Ala Gly Gly Arg Pro Ala Phe Ile Gly
 35 40 45

gtg gcg cgc ctt gac ggc cgc ctc gga ttc act tcg cag ggc act tct 192
 Val Ala Arg Leu Asp Gly Arg Leu Gly Phe Thr Ser Gln Gly Thr Ser
 50 55 60

gac gcg ctc aag cgt gcc cag cgt ggt gcc atc ttt ggc ctc tgc aag 240
 Asp Ala Leu Lys Arg Ala Gln Arg Gly Ala Ile Phe Gly Leu Cys Lys
 65 70 75 80

acc atc ggc ctc gag tgg tcc gag tct gac gtc ttt tcc cgc ggc gtg 288
 Thr Ile Gly Leu Glu Trp Ser Glu Ser Asp Val Phe Ser Arg Gly Val
 85 90 95

gac att gct cag ggc atg cac ccc gag gat gcc gcc gtg gcg att gtg 336
 Asp Ile Ala Gln Gly Met His Pro Glu Asp Ala Ala Val Ala Ile Val
 100 105 110

cgc gag atg gcg tgc gct gac att cgc att cgc gag gtc ggc att ggc 384
 Arg Glu Met Ala Cys Ala Asp Ile Arg Ile Arg Glu Val Gly Ile Gly
 115 120 125

gca aac cag cag cgc tgc acg atc cgt gcc gcc aag ctc gag acc ggc 432
Ala Asn Gln Gln Arg Cys Thr Ile Arg Ala Ala Lys Leu Glu Thr Gly

130 135 140

aac ceg cag cgc cag atc gcc aag gac gac gtg ctg ctc gtt tct ggc 480
Asn Pro Gln Arg Gln Ile Ala Lys Asp Asp Val Leu Leu Val Ser Gly

145 150 155 160

ggc gct cgc ggc atc acg cct ctt tgc atc cgg gag atc acg cgc cag 528

Gly Ala Arg Gly Ile Thr Pro Leu Cys Ile Arg Glu Ile Thr Arg Gln

165 170 175

atc gcg ggc ggc aag tac att ctg ctt ggc cgc agc aag gtc tct gcg 576

Ile Ala Gly Gly Tyr Ile Leu Leu Gly Arg Ser Lys Val Ser Ala

180 185 190

agc gaa ceg gca tgg tgc gct ggc atc act gac gag aag gct gtg caa 624

Ser Glu Pro Ala Trp Cys Ala Gly Ile Thr Asp Glu Lys Ala Val Gln

195 200 205

aag gct gct acc cag gag ctc aag cgc gcc ttt agc gct ggc gag ggc 672

Lys Ala Ala Thr Gln Glu Leu Lys Arg Ala Phe Ser Ala Gly Glu Gly

210 215 220

ccc aag ccc acg ccc cgc gct gtc act aag ctt gtg ggc tct gtt ctt 720

Pro Lys Pro Thr Pro Arg Ala Val Thr Lys Leu Val Gly Ser Val Leu

225 230 235 240

ggc gct cgc gag gtg cgc agc tct att gct gcg att gaa gcg ctc ggc 768

Gly Ala Arg Glu Val Arg Ser Ser Ile Ala Ala Ile Glu Ala Leu Gly

245 250 255

ggc aag gcc atc tac tcg tcg tgc gac gtg aac tct gcc gcc gac gtg 816

Gly Lys Ala Ile Tyr Ser Ser Cys Asp Val Asn Ser Ala Ala Asp Val

260 265 270

gcc aag gcc gtg cgc gat gcc gag tcc cag ctc ggt gcc cgc gtc tcg 864

Ala Lys Ala Val Arg Asp Ala Glu Ser Gln Leu Gly Ala Arg Val Ser

275 280 285

ggc atc gtt cat gcc tcg ggc gtg ctc cgc gac cgt ctc atc gag aag 912

Gly Ile Val His Ala Ser Gly Val Leu Arg Asp Arg Leu Ile Glu Lys

290 295 300

aag ctc ccc gac gag ttc gac gcc gtc ttt ggc acc aag gtc acc ggt 960

Lys Leu Pro Asp Glu Phe Asp Ala Val Phe Gly Thr Lys Val Thr Gly

305 310 315 320

ctc gag aac ctc ctc gcc gcc gtc gac cgc gcc aac ctc aag cac atg 1008

Leu Glu Asn Leu Leu Ala Ala Val Asp Arg Ala Asn Leu Lys His Met

325 330 335

gtc ctc ttc agc tcg ctc gcc ggc ttc cac ggc aac gtc ggc cag tct 1056

Val Leu Phe Ser Ser Leu Ala Gly Phe His Gly Asn Val Gly Gln Ser
 340 345 350

gac tac gcc atg gcc aac gag gcc ctt aac aag atg ggc ctc gag ctc 1104
 Asp Tyr Ala Met Ala Asn Glu Ala Leu Asn Lys Met Gly Leu Glu Leu
 355 360 365

gcc aag gac gtc tcg gtc aag tcg atc tgc ttc ggt ccc tgg gac ggt 1152
 Ala Lys Asp Val Ser Val Lys Ser Ile Cys Phe Gly Pro Trp Asp Gly
 370 375 380

ggc atg gtg acg ccg cag ctc aag aag cag ttc cag gag atg ggc gtg 1200
 Gly Met Val Thr Pro Gln Leu Lys Lys Gln Phe Gln Glu Met Gly Val
 385 390 395 400

cag atc atc ccc cgc gag ggc ggc gct gat acc gtg gcg cgc atc gtg 1248
 Gln Ile Ile Pro Arg Glu Gly Gly Ala Asp Thr Val Ala Arg Ile Val
 405 410 415

ctc ggc tcc tcg ccg gct gag atc ctt gtc ggc aac tgg cgc acc ccg 1296
 Leu Gly Ser Ser Pro Ala Glu Ile Leu Val Gly Asn Trp Arg Thr Pro
 420 425 430

tcc aag aag gtc ggc tcg gac acc atc acc ctg cac cgc aag att tcc 1344
 Ser Lys Lys Val Gly Ser Asp Thr Ile Thr Leu His Arg Lys Ile Ser
 435 440 445

gcc aag tcc aac ccc ttc ctc gag gac cac gtc atc cag ggc cgc cgc 1392
 Ala Lys Ser Asn Pro Phe Leu Glu Asp His Val Ile Gln Gly Arg Arg
 450 455 460

gtg ctg ccc atg acg ctg gcc att ggc tcg ctc gcg gag acc tgc ctc 1440
 Val Leu Pro Met Thr Leu Ala Ile Gly Ser Leu Ala Glu Thr Cys Leu
 465 470 475 480

ggc ctc ttc ccc ggc tac tcg ctc tgg gcc att gac gac gcc cag ctc 1488
 Gly Leu Phe Pro Gly Tyr Ser Leu Trp Ala Ile Asp Asp Ala Gln Leu
 485 490 495

ttc aag ggt gtc act gtc gac ggc gac gtc aac tgc gag gtg acc ctc 1536
 Phe Lys Gly Val Thr Val Asp Gly Asp Val Asn Cys Glu Val Thr Leu
 500 505 510

acc ccg tcg acg gcg ccc tcg ggc cgc gtc aac gtc cag gcc acg ctc 1584
 Thr Pro Ser Thr Ala Pro Ser Gly Arg Val Asn Val Gln Ala Thr Leu
 515 520 525

aag acc ttt tcc agc ggc aag ctg gtc ccg gcc tac cgc gcc gtc atc 1632
 Lys Thr Phe Ser Ser Gly Lys Leu Val Pro Ala Tyr Arg Ala Val Ile
 530 535 540

gtg ctc tcc aac cag ggc gcg ccc ccg gcc aac gcc acc atg cag ccg 1680
 Val Leu Ser Asn Gln Gly Ala Pro Pro Ala Asn Ala Thr Met Gln Pro
 545 550 555 560

ccc tcg ctc gat gcc gat ccg gcg ctc cag ggc tcc gtc tac gac ggc 1728
 Pro Ser Leu Asp Ala Asp Pro Ala Leu Gln Gly Ser Val Tyr Asp Gly
 565 570 575

aag acc ctc ttc cac ggc ccg gcc ttc cgc ggc atc gat gac gtg ctc 1776
 Lys Thr Leu Phe His Gly Pro Ala Phe Arg Gly Ile Asp Asp Val Leu
 580 585 590

tcg tgc acc aag agc cag ctt gtg gcc aag tgc agc gct gtc ccc ggc 1824
 Ser Cys Thr Lys Ser Gln Leu Val Ala Lys Cys Ser Ala Val Pro Gly
 595 600 605

tcc gac gcc gct cgc ggc gag ttt gcc acg gac act gac gcc cat gac 1872
 Ser Asp Ala Ala Arg Gly Glu Phe Ala Thr Asp Thr Asp Ala His Asp
 610 615 620

ccc ttc gtg aac gac ctg gcc ttt cag gcc atg ctc gtc tgg gtg cgc 1920
 Pro Phe Val Asn Asp Leu Ala Phe Gln Ala Met Leu Val Trp Val Arg
 625 630 635 640

cgc acg ctc ggc cag gct gcg ctc ccc aac tcg atc cag cgc atc gtc 1968
 Arg Thr Leu Gly Gln Ala Ala Leu Pro Asn Ser Ile Gln Arg Ile Val
 645 650 655

cag cac cgc ccg gtc ccg cag gac aag ccc ttc tac att acc ctc cgc 2016
 Gln His Arg Pro Val Pro Gln Asp Lys Pro Phe Tyr Ile Thr Leu Arg
 660 665 670

tcc aac cag tcg ggc ggt cac tcc cag cac aag cac gcc ctt cag ttc 2064
 Ser Asn Gln Ser Gly Gly His Ser Gln His Lys His Ala Leu Gln Phe
 675 680 685

cac aac gag cag ggc gat ctc ttc att gat gtc cag gct tcg gtc atc 2112
 His Asn Glu Gln Gly Asp Leu Phe Ile Asp Val Gln Ala Ser Val Ile
 690 695 700

gcc acg gac agc ctt gcc ttc 2133
 Ala Thr Asp Ser Leu Ala Phe
 705 710

<210> 18 <211> 711 <212> PRT <213> 裂壺菌屬 sp. <400> 18

Phe Gly Ala Leu Gly Gly Phe Ile Ser Gln Gln Ala Glu Arg Phe Glu
 1 5 10 15

Pro Ala Glu Ile Leu Gly Phe Thr Leu Met Cys Ala Lys Phe Ala Lys
 20 25 30

Ala Ser Leu Cys Thr Ala Val Ala Gly Gly Arg Pro Ala Phe Ile Gly
 35 40 45

Val Ala Arg Leu Asp Gly Arg Leu Gly Phe Thr Ser Gln Gly Thr Ser
50 55 60

Asp Ala Leu Lys Arg Ala Gln Arg Gly Ala Ile Phe Gly Leu Cys Lys
65 70 75 80

Thr Ile Gly Leu Glu Trp Ser Glu Ser Asp Val Phe Ser Arg Gly Val
85 90 95

Asp Ile Ala Gln Gly Met His Pro Glu Asp Ala Ala Val Ala Ile Val
100 105 110

Arg Glu Met Ala Cys Ala Asp Ile Arg Ile Arg Glu Val Gly Ile Gly
115 120 125

Ala Asn Gln Gln Arg Cys Thr Ile Arg Ala Ala Lys Leu Glu Thr Gly
130 135 140

Asn Pro Gln Arg Gln Ile Ala Lys Asp Asp Val Leu Leu Val Ser Gly
145 150 155 160

Gly Ala Arg Gly Ile Thr Pro Leu Cys Ile Arg Glu Ile Thr Arg Gln
165 170 175

Ile Ala Gly Gly Lys Tyr Ile Leu Leu Gly Arg Ser Lys Val Ser Ala
180 185 190

Ser Glu Pro Ala Trp Cys Ala Gly Ile Thr Asp Glu Lys Ala Val Gln
195 200 205

Lys Ala Ala Thr Gln Glu Leu Lys Arg Ala Phe Ser Ala Gly Glu Gly
210 215 220

Pro Lys Pro Thr Pro Arg Ala Val Thr Lys Leu Val Gly Ser Val Leu
225 230 235 240

Gly Ala Arg Glu Val Arg Ser Ser Ile Ala Ala Ile Glu Ala Leu Gly
245 250 255

Gly Lys Ala Ile Tyr Ser Ser Cys Asp Val Asn Ser Ala Ala Asp Val
 260 265 270

Ala Lys Ala Val Arg Asp Ala Glu Ser Gln Leu Gly Ala Arg Val Ser
 275 280 285

Gly Ile Val His Ala Ser Gly Val Leu Arg Asp Arg Leu Ile Glu Lys
 290 295 300

Lys Leu Pro Asp Glu Phe Asp Ala Val Phe Gly Thr Lys Val Thr Gly
 305 310 315 320

Leu Glu Asn Leu Leu Ala Ala Val Asp Arg Ala Asn Leu Lys His Met
 325 330 335

Val Leu Phe Ser Ser Leu Ala Gly Phe His Gly Asn Val Gly Gln Ser
 340 345 350

Asp Tyr Ala Met Ala Asn Glu Ala Leu Asn Lys Met Gly Leu Glu Leu
 355 360 365

Ala Lys Asp Val Ser Val Lys Ser Ile Cys Phe Gly Pro Trp Asp Gly
 370 375 380

Gly Met Val Thr Pro Gln Leu Lys Lys Gln Phe Gln Glu Met Gly Val
 385 390 395 400

Gln Ile Ile Pro Arg Glu Gly Gly Ala Asp Thr Val Ala Arg Ile Val
 405 410 415

Leu Gly Ser Ser Pro Ala Glu Ile Leu Val Gly Asn Trp Arg Thr Pro
 420 425 430

Ser Lys Lys Val Gly Ser Asp Thr Ile Thr Leu His Arg Lys Ile Ser
 435 440 445

Ala Lys Ser Asn Pro Phe Leu Glu Asp His Val Ile Gln Gly Arg Arg
 450 455 460

Val Leu Pro Met Thr Leu Ala Ile Gly Ser Leu Ala Glu Thr Cys Leu
 465 470 475 480

Gly Leu Phe Pro Gly Tyr Ser Leu Trp Ala Ile Asp Asp Ala Gln Leu
485 490 495

Phe Lys Gly Val Thr Val Asp Gly Asp Val Asn Cys Glu Val Thr Leu
500 505 510

Thr Pro Ser Thr Ala Pro Ser Gly Arg Val Asn Val Gln Ala Thr Leu
515 520 525

Lys Thr Phe Ser Ser Gly Lys Leu Val Pro Ala Tyr Arg Ala Val Ile
530 535 540

Val Leu Ser Asn Gln Gly Ala Pro Pro Ala Asn Ala Thr Met Gln Pro
545 550 555 560

Pro Ser Leu Asp Ala Asp Pro Ala Leu Gln Gly Ser Val Tyr Asp Gly
565 570 575

Lys Thr Leu Phe His Gly Pro Ala Phe Arg Gly Ile Asp Asp Val Leu
580 585 590

Ser Cys Thr Lys Ser Gln Leu Val Ala Lys Cys Ser Ala Val Pro Gly
595 600 605

Ser Asp Ala Ala Arg Gly Glu Phe Ala Thr Asp Thr Asp Ala His Asp
610 615 620

Pro Phe Val Asn Asp Leu Ala Phe Gln Ala Met Leu Val Trp Val Arg
625 630 635 640

Arg Thr Leu Gly Gln Ala Ala Leu Pro Asn Ser Ile Gln Arg Ile Val
645 650 655

Gln His Arg Pro Val Pro Gln Asp Lys Pro Phe Tyr Ile Thr Leu Arg
660 665 670

Ser Asn Gln Ser Gly Gly His Ser Gln His Lys His Ala Leu Gln Phe
675 680 685

His Asn Glu Gln Gly Asp Leu Phe Ile Asp Val Gln Ala Ser Val Ile
 690 695 700

Ala Thr Asp Ser Leu Ala Phe
 705 710

<210> 19 <211> 1350 <212> DNA <213> 裂壺菌屬 sp. <220> <221> CDS <222> (1)..(1350)
 <223>

<400> 19
 atg gcc gct cgg aat gtg agc gcc gcg cat gag atg cac gat gaa aag 48
 Met Ala Ala Arg Asn Val Ser Ala Ala His Glu Met His Asp Glu Lys
 1 5 10 15

cgc atc gcc gtc gtc ggc atg gcc gtc cag tac gcc gga tgc aaa acc 96
 Arg Ile Ala Val Val Gly Met Ala Val Gln Tyr Ala Gly Cys Lys Thr
 20 25 30

aag gac gag ttc tgg gag gtg ctc atg aac ggc aag gtc gag tcc aag 144
 Lys Asp Glu Phe Trp Glu Val Leu Met Asn Gly Lys Val Glu Ser Lys
 35 40 45

gtg atc agc gac aaa cga ctc ggc tcc aac tac cgc gcc gag cac tac 192
 Val Ile Ser Asp Lys Arg Leu Gly Ser Asn Tyr Arg Ala Glu His Tyr
 50 55 60

aaa gca gag cgc agc aag tat gcc gac acc ttt tgc aac gaa acg tac 240
 Lys Ala Glu Arg Ser Lys Tyr Ala Asp Thr Phe Cys Asn Glu Thr Tyr
 65 70 75 80

ggc acc ctt gac gag aac gag atc gac aac gag cac gaa ctc ctc ctc 288
 Gly Thr Leu Asp Glu Asn Glu Ile Asp Asn Glu His Glu Leu Leu Leu
 85 90 95

aac ctc gcc aag cag gca ctc gca gag aca tcc gtc aaa gac tgg aca 336
 Asn Leu Ala Lys Gln Ala Leu Ala Glu Thr Ser Val Lys Asp Ser Thr
 100 105 110

cgc tgc gcc atc gtc agc gcc tgc ctc tgg ttc ccc atg gac aac ctc 384
 Arg Cys Gly Ile Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu
 115 120 125

cag ggt gaa ctc ctc aac gtg tac caa aac cat gtc gag aaa aag ctc 432
 Gln Gly Glu Leu Leu Asn Val Tyr Gln Asn His Val Glu Lys Lys Leu
 130 135 140

ggg gcc cgc gtc ttc aag gac gcc tcc cat tgg tcc gaa cgc gag cag 480
 Gly Ala Arg Val Phe Lys Asp Ala Ser His Trp Ser Glu Arg Glu Gln
 145 150 155 160

tcc aac aaa ccc gag gcc ggt gac cgc cgc atc ttc atg gac ccg gcc 528
 Ser Asn Lys Pro Glu Ala Gly Asp Arg Arg Ile Phe Met Asp Pro Ala

165	170	175	
tcc ttc gtc gcc gaa gaa ctc aac ctc ggc gcc ctt cac tac tcc gtc 576			
Ser Phe Val Ala Glu Glu Leu Asn Leu Gly Ala Leu His Tyr Ser Val			
180	185	190	
gac gca gca tgc gcc acg gcg ctc tac gtg ctc cgc ctc gcg cag gat 624			
Asp Ala Ala Cys Ala Thr Ala Leu Tyr Val Leu Arg Leu Ala Gln Asp			
195	200	205	
cat ctc gtc tcc ggc gcc gcc gac gtc atg ctc tgc ggt gcc acc tgc 672			
His Leu Val Ser Gly Ala Ala Asp Val Met Leu Cys Gly Ala Thr Cys			
210	215	220	
ctg ccg gag ccc ttt ttc atc ctt tgc ggc ttt tcc acc ttc cag gcc 720			
Leu Pro Glu Pro Phe Phe Ile Leu Ser Gly Phe Ser Thr Phe Gln Ala			
225	230	235	240
atg ccc gtc ggc acg ggc cag aac gtg tcc atg ccg ctg cac aag gac 768			
Met Pro Val Gly Thr Gly Gln Asn Val Ser Met Pro Leu His Lys Asp			
245	250	255	
agc cag ggc ctc acc ccg ggt gag ggc ggc tcc atc atg gtc ctc aag 816			
Ser Gln Gly Leu Thr Pro Gly Glu Gly Gly Ser Ile Met Val Leu Lys			
260	265	270	
cgt ctc gat gat gcc atc cgc gac ggc gac cac att tac ggc acc ctt 864			
Arg Leu Asp Asp Ala Ile Arg Asp Gly Asp His Ile Tyr Gly Thr Leu			
275	280	285	
ctc ggc gcc aat gtc agc aac tcc ggc aca ggt ctg ccc ctc aag ccc 912			
Leu Gly Ala Asn Val Ser Asn Ser Gly Thr Gly Leu Pro Leu Lys Pro			
290	295	300	
ctt ctc ccc agc gag aaa aag tgc ctc atg gac acc tac acg cgc att 960			
Leu Leu Pro Ser Glu Lys Lys Cys Leu Met Asp Thr Tyr Thr Arg Ile			
305	310	315	320
aac gtg cac ccg cac aag att cag tac gtc gag tgc cac gcc acc ggc 1008			
Asn Val His Pro His Lys Ile Gln Tyr Val Glu Cys His Ala Thr Gly			
325	330	335	
acg ccc cag ggt gat cgt gtg gaa atc gac gcc gtc aag gcc tgc ttt 1056			
Thr Pro Gln Gly Asp Arg Val Glu Ile Asp Ala Val Lys Ala Cys Phe			
340	345	350	
gaa ggc aag gtc ccc cgt ttc ggt acc aca aag ggc aac ttt gga cac 1104			
Glu Gly Lys Val Pro Arg Phe Gly Thr Thr Lys Gly Asn Phe Gly His			
355	360	365	
acc cts gyc gca gcc ggc ttt gcc ggt atg tgc aag gtc ctc ctc tcc 1152			
Thr Xaa Xaa Ala Ala Gly Phe Ala Gly Met Cys Lys Val Leu Leu Ser			
370	375	380	

atg aag cat ggc atc atc ccg ccc acc ccg ggt atc gat gac gag acc 1200
 Met Lys His Gly Ile Ile Pro Pro Thr Pro Gly Ile Asp Asp Glu Thr
 385 390 395 400

aag atg gac cct ctc gtc gtc tcc ggt gag gcc atc cca tgg cca gag 1248
 Lys Met Asp Pro Leu Val Val Ser Gly Glu Ala Ile Pro Trp Pro Glu
 405 410 415

acc aac ggc gag ccc aag cgc gcc ggt ctc tcg gcc ttt ggc ttt ggt 1296
 Thr Asn Gly Glu Pro Lys Arg Ala Gly Leu Ser Ala Phe Gly Phe Gly
 420 425 430

ggc acc aac gcc cat gcc gtc ttt gag gag cat gac ccc tcc aac gcc 1344
 Gly Thr Asn Ala His Ala Val Phe Glu Glu His Asp Pro Ser Asn Ala
 435 440 445

gcc tgc 1350
 Ala Cys
 450

<210> 20 <211> 450 <212> PRT <213> 裂壺菌屬 sp. <220> <221> misc_特徵 <222> (370)..
 (370) <223> 在位置370處的'Xaa'代表Leu.

<220> <221> misc_特徵 <222> (371)..(371) <223> 在位置371處的'Xaa'代表Ala或Val.

<400> 20

Met Ala Ala Arg Asn Val Ser Ala Ala His Glu Met His Asp Glu Lys
 1 5 10 15

Arg Ile Ala Val Val Gly Met Ala Val Gln Tyr Ala Gly Cys Lys Thr
 20 25 30

Lys Asp Glu Phe Trp Glu Val Leu Met Asn Gly Lys Val Glu Ser Lys
 35 40 45

Val Ile Ser Asp Lys Arg Leu Gly Ser Asn Tyr Arg Ala Glu His Tyr
 50 55 60

Lys Ala Glu Arg Ser Lys Tyr Ala Asp Thr Phe Cys Asn Glu Thr Tyr
 65 70 75 80

Gly Thr Leu Asp Glu Asn Glu Ile Asp Asn Glu His Glu Leu Leu Leu
 85 90 95

Asn Leu Ala Lys Gln Ala Leu Ala Glu Thr Ser Val Lys Asp Ser Thr
 100 105 110

Arg Cys Gly Ile Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu
115 120 125

Gln Gly Glu Leu Leu Asn Val Tyr Gln Asn His Val Glu Lys Lys Leu
130 135 140

Gly Ala Arg Val Phe Lys Asp Ala Ser His Trp Ser Glu Arg Glu Gln
145 150 155 160

Ser Asn Lys Pro Glu Ala Gly Asp Arg Arg Ile Phe Met Asp Pro Ala
165 170 175

Ser Phe Val Ala Glu Glu Leu Asn Leu Gly Ala Leu His Tyr Ser Val
180 185 190

Asp Ala Ala Cys Ala Thr Ala Leu Tyr Val Leu Arg Leu Ala Gln Asp
195 200 205

His Leu Val Ser Gly Ala Ala Asp Val Met Leu Cys Gly Ala Thr Cys
210 215 220

Leu Pro Glu Pro Phe Phe Ile Leu Ser Gly Phe Ser Thr Phe Gln Ala
225 230 235 240

Met Pro Val Gly Thr Gly Gln Asn Val Ser Met Pro Leu His Lys Asp
245 250 255

Ser Gln Gly Leu Thr Pro Gly Glu Gly Gly Ser Ile Met Val Leu Lys
260 265 270

Arg Leu Asp Asp Ala Ile Arg Asp Gly Asp His Ile Tyr Gly Thr Leu
275 280 285

Leu Gly Ala Asn Val Ser Asn Ser Gly Thr Gly Leu Pro Leu Lys Pro
290 295 300

Leu Leu Pro Ser Glu Lys Lys Cys Leu Met Asp Thr Tyr Thr Arg Ile
305 310 315 320

Asn Val His Pro His Lys Ile Gln Tyr Val Glu Cys His Ala Thr Gly
 325 330 335

Thr Pro Gln Gly Asp Arg Val Glu Ile Asp Ala Val Lys Ala Cys Phe
 340 345 350

Glu Gly Lys Val Pro Arg Phe Gly Thr Thr Lys Gly Asn Phe Gly His
 355 360 365

Thr Xaa Xaa Ala Ala Gly Phe Ala Gly Met Cys Lys Val Leu Leu Ser
 370 375 380

Met Lys His Gly Ile Ile Pro Pro Thr Pro Gly Ile Asp Asp Glu Thr
 385 390 395 400

Lys Met Asp Pro Leu Val Val Ser Gly Glu Ala Ile Pro Trp Pro Glu
 405 410 415

Thr Asn Gly Glu Pro Lys Arg Ala Gly Leu Ser Ala Phe Gly Phe Gly
 420 425 430

Gly Thr Asn Ala His Ala Val Phe Glu Glu His Asp Pro Ser Asn Ala
 435 440 445

Ala Cys
 450

<210> 21 <211> 1323 <212> DNA <213> 裂壺菌屬 sp. <220> <221> CDS <222> (1)..(1323)
 <223>

<400> 21
 tcg gcc cgc tgc ggc ggt gaa agc aac atg cgc atc gcc atc act ggt 48
 Ser Ala Arg Cys Gly Gly Glu Ser Asn Met Arg Ile Ala Ile Thr Gly
 1 5 10 15

atg gac gcc acc ttt ggc gct ctc aag gga ctc gac gcc ttc gag cgc 96
 Met Asp Ala Thr Phe Gly Ala Leu Lys Gly Leu Asp Ala Phe Glu Arg
 20 25 30

gcc att tac acc ggc gct cac ggt gcc atc cca ctc cca gaa aag cgc 144
 Ala Ile Tyr Thr Gly Ala His Gly Ala Ile Pro Leu Pro Glu Lys Arg
 35 40 45

tgg cgc ttt ctc ggc aag gac aag gac ttt ctt gac ctc tgc ggc gtc 192
 Trp Arg Phe Leu Gly Lys Asp Lys Asp Phe Leu Asp Leu Cys Gly Val

50	55	60	
aag gcc acc ccg cac ggc tgc tac att gaa gat gtt gag gtc gac ttc	240		
Lys Ala Thr Pro His Gly Cys Tyr Ile Glu Asp Val Glu Val Asp Phe			
65	70	75	80
cag cgc ctc cgc acg ccc atg acc cct gaa gac atg ctc ctc cct cag	288		
Gln Arg Leu Arg Thr Pro Met Thr Pro Glu Asp Met Leu Leu Pro Gln			
85	90	95	
cag ctt ctg gcc gtc acc acc att gac cgc gcc atc ctc gac tcg gga	336		
Gln Leu Leu Ala Val Thr Thr Ile Asp Arg Ala Ile Leu Asp Ser Gly			
100	105	110	
atg aaa aag ggt ggc aat gtc gcc gtc ttt gtc ggc ctc ggc acc gac	384		
Met Lys Lys Gly Gly Asn Val Ala Val Phe Val Gly Leu Gly Thr Asp			
115	120	125	
ctc gag ctc tac cgt cac cgt gct cgc gtc gct ctc aag gag cgc gtc	432		
Leu Glu Leu Tyr Arg His Arg Ala Arg Val Ala Leu Lys Glu Arg Val			
130	135	140	
cgc cct gaa gcc tcc aag aag ctc aat gac atg atg cag tac att aac	480		
Arg Pro Glu Ala Ser Lys Lys Leu Asn Asp Met Met Gln Tyr Ile Asn			
145	150	155	160
gac tgc ggc aca tcc aca tcg tac acc tcg tac att ggc aac ctc gtc	528		
Asp Cys Gly Thr Ser Thr Ser Tyr Thr Ser Tyr Ile Gly Asn Leu Val			
165	170	175	
gcc acg cgc gtc tcg tcg cag tgg ggc ttc acg ggc ccc tcc ttt acg	576		
Ala Thr Arg Val Ser Ser Gln Trp Gly Phe Thr Gly Pro Ser Phe Thr			
180	185	190	
atc acc gag ggc aac aac tcc gtc tac cgc tgc gcc gag ctc ggc aag	624		
Ile Thr Glu Gly Asn Asn Ser Val Tyr Arg Cys Ala Glu Leu Gly Lys			
195	200	205	
tac ctc ctc gag acc ggc gag gtc gat ggc gtc gtc gtt gcg ggt gtc	672		
Tyr Leu Leu Glu Thr Gly Glu Val Asp Gly Val Val Val Ala Gly Val			
210	215	220	
gat ctc tgc ggc agt gcc gaa aac ctt tac gtc aag tct cgc cgc ttc	720		
Asp Leu Cys Gly Ser Ala Glu Asn Leu Tyr Val Lys Ser Arg Arg Phe			
225	230	235	240
aag gtg tcc acc tcc gat acc ccg cgc gcc agc ttt gac gcc gcc gcc	768		
Lys Val Ser Thr Ser Asp Thr Pro Arg Ala Ser Phe Asp Ala Ala Ala			
245	250	255	
gat ggc tac ttt gtc ggc gag ggc tgc ggt gcc ttt gtg ctc aag cgt	816		
Asp Gly Tyr Phe Val Gly Glu Gly Cys Gly Ala Phe Val Leu Lys Arg			
260	265	270	

gag act agc tgc acc aag gac gac cgt atc tac gct tgc atg gat gcc 864
 Glu Thr Ser Cys Thr Lys Asp Asp Arg Ile Tyr Ala Cys Met Asp Ala
 275 280 285

atc gtc cct ggc aac gtc cct agc gcc tgc ttg cgc gag gcc ctc gac 912
 Ile Val Pro Gly Asn Val Pro Ser Ala Cys Leu Arg Glu Ala Leu Asp
 290 295 300

cag gcg cgc gtc aag ccg ggc gat atc gag atg ctc gag ctc agc gcc 960
 Gln Ala Arg Val Lys Pro Gly Asp Ile Glu Met Leu Glu Leu Ser Ala
 305 310 315 320

gac tcc gcc cgc cac ctc aag gac ccg tcc gtc ctg ccc aag gag ctc 1008
 Asp Ser Ala Arg His Leu Lys Asp Pro Ser Val Leu Pro Lys Glu Leu
 325 330 335

act gcc gag gag gaa atc ggc ggc ctt cag acg atc ctt cgt gac gat 1056
 Thr Ala Glu Glu Ile Gly Gly Leu Gln Thr Ile Leu Arg Asp Asp
 340 345 350

gac aag ctc ccg cgc aac gtc gca acg ggc agt gtc aag gcc acc gtc 1104
 Asp Lys Leu Pro Arg Asn Val Ala Thr Gly Ser Val Lys Ala Thr Val
 355 360 365

ggt gac acc ggt tat gcc tct ggt gct gcc agc ctc atc aag gct gcg 1152
 Gly Asp Thr Gly Tyr Ala Ser Gly Ala Ala Ser Leu Ile Lys Ala Ala
 370 375 380

ctt tgc atc tac aac cgc tac ctg ccc agc aac ggc gac gac tgg gat 1200
 Leu Cys Ile Tyr Asn Arg Tyr Leu Pro Ser Asn Gly Asp Asp Trp Asp
 385 390 395 400

gaa ccc gcc cct gag gcg ccc tgg gac agc acc ctc ttt gcg tgc cag 1248
 Glu Pro Ala Pro Glu Ala Pro Trp Asp Ser Thr Leu Phe Ala Cys Gln
 405 410 415

acc tcg cgc gct tgg ctc aag aac cct ggc gag cgt cgc tat gcg gcc 1296
 Thr Ser Arg Ala Trp Leu Lys Asn Pro Gly Glu Arg Arg Tyr Ala Ala
 420 425 430

gtc tcg ggc gtc tcc gag acg cgc tcg 1323
 Val Ser Gly Val Ser Glu Thr Arg Ser
 435 440

<210> 22 <211> 441 <212> PRT <213> 裂壺菌屬 sp. <400> 22

Ser Ala Arg Cys Gly Gly Glu Ser Asn Met Arg Ile Ala Ile Thr Gly
 1 5 10 15

Met Asp Ala Thr Phe Gly Ala Leu Lys Gly Leu Asp Ala Phe Glu Arg
 20 25 30

Ala Ile Tyr Thr Gly Ala His Gly Ala Ile Pro Leu Pro Glu Lys Arg
35 40 45

Trp Arg Phe Leu Gly Lys Asp Lys Asp Phe Leu Asp Leu Cys Gly Val
50 55 60

Lys Ala Thr Pro His Gly Cys Tyr Ile Glu Asp Val Glu Val Asp Phe
65 70 75 80

Gln Arg Leu Arg Thr Pro Met Thr Pro Glu Asp Met Leu Leu Pro Gln
85 90 95

Gln Leu Leu Ala Val Thr Thr Ile Asp Arg Ala Ile Leu Asp Ser Gly
100 105 110

Met Lys Lys Gly Gly Asn Val Ala Val Phe Val Gly Leu Gly Thr Asp
115 120 125

Leu Glu Leu Tyr Arg His Arg Ala Arg Val Ala Leu Lys Glu Arg Val
130 135 140

Arg Pro Glu Ala Ser Lys Lys Leu Asn Asp Met Met Gln Tyr Ile Asn
145 150 155 160

Asp Cys Gly Thr Ser Thr Ser Tyr Thr Ser Tyr Ile Gly Asn Leu Val
165 170 175

Ala Thr Arg Val Ser Ser Gln Trp Gly Phe Thr Gly Pro Ser Phe Thr
180 185 190

Ile Thr Glu Gly Asn Asn Ser Val Tyr Arg Cys Ala Glu Leu Gly Lys
195 200 205

Tyr Leu Leu Glu Thr Gly Glu Val Asp Gly Val Val Val Ala Gly Val
210 215 220

Asp Leu Cys Gly Ser Ala Glu Asn Leu Tyr Val Lys Ser Arg Arg Phe
225 230 235 240

Lys Val Ser Thr Ser Asp Thr Pro Arg Ala Ser Phe Asp Ala Ala Ala

245 250 255

Asp Gly Tyr Phe Val Gly Glu Gly Cys Gly Ala Phe Val Leu Lys Arg
260 265 270

Glu Thr Ser Cys Thr Lys Asp Asp Arg Ile Tyr Ala Cys Met Asp Ala
275 280 285

Ile Val Pro Gly Asn Val Pro Ser Ala Cys Leu Arg Glu Ala Leu Asp
290 295 300

Gln Ala Arg Val Lys Pro Gly Asp Ile Glu Met Leu Glu Leu Ser Ala
305 310 315 320

Asp Ser Ala Arg His Leu Lys Asp Pro Ser Val Leu Pro Lys Glu Leu
325 330 335

Thr Ala Glu Glu Glu Ile Gly Gly Leu Gln Thr Ile Leu Arg Asp Asp
340 345 350

Asp Lys Leu Pro Arg Asn Val Ala Thr Gly Ser Val Lys Ala Thr Val
355 360 365

Gly Asp Thr Gly Tyr Ala Ser Gly Ala Ala Ser Leu Ile Lys Ala Ala
370 375 380

Leu Cys Ile Tyr Asn Arg Tyr Leu Pro Ser Asn Gly Asp Asp Trp Asp
385 390 395 400

Glu Pro Ala Pro Glu Ala Pro Trp Asp Ser Thr Leu Phe Ala Cys Gln
405 410 415

Thr Ser Arg Ala Trp Leu Lys Asn Pro Gly Glu Arg Arg Tyr Ala Ala
420 425 430

Val Ser Gly Val Ser Glu Thr Arg Ser
435 440

<210> 23 <211> 1500 <212> DNA <213> 裂壺菌屬 sp. <220> <221> CDS <222> (1)..(1500)
<223>

<400> 23

tgc tat tcc gtg ctc ctc tcc gaa gcc gag ggc cac tac gag cgc gag 48
 Cys Tyr Ser Val Leu Leu Ser Glu Ala Glu Gly His Tyr Glu Arg Glu
 1 5 10 15

aac cgc atc tgc ctc gac gag gag gcg ccc aag ctc att gtg ctt cgc 96
 Asn Arg Ile Ser Leu Asp Glu Glu Ala Pro Lys Leu Ile Val Leu Arg
 20 25 30

gcc gac tcc cac gag gag atc ctt ggt cgc ctc gac aag atc cgc gag 144
 Ala Asp Ser His Glu Glu Ile Leu Gly Arg Leu Asp Lys Ile Arg Glu
 35 40 45

cgc ttc ttg cag ccc acg ggc gcc gcc ccg cgc gag tcc gag ctc aag 192
 Arg Phe Leu Gln Pro Thr Gly Ala Ala Pro Arg Glu Ser Glu Leu Lys
 50 55 60

gcg cag gcc cgc cgc atc ttc ctc gag ctc ctc ggc gag acc ctt gcc 240
 Ala Gln Ala Arg Arg Ile Phe Leu Glu Leu Leu Gly Glu Thr Leu Ala
 65 70 75 80

cag gat gcc gct tct tca ggc tgc caa aag ccc ctc gct ctc agc ctc 288
 Gln Asp Ala Ala Ser Ser Gly Ser Gln Lys Pro Leu Ala Leu Ser Leu
 85 90 95

gtc tcc acg ccc tcc aag ctc cag cgc gag gtc gag ctc gcg gcc aag 336
 Val Ser Thr Pro Ser Lys Leu Gln Arg Glu Val Glu Leu Ala Ala Lys
 100 105 110

ggt atc ccg cgc tgc ctc aag atg cgc cgc gat tgg agc tcc cct gct 384
 Gly Ile Pro Arg Cys Leu Lys Met Arg Arg Asp Trp Ser Ser Pro Ala
 115 120 125

ggc agc cgc tac gcg cct gag ccg ctc gcc agc gac cgc gtc gcc ttc 432
 Gly Ser Arg Tyr Ala Pro Glu Pro Leu Ala Ser Asp Arg Val Ala Phe
 130 135 140

atg tac ggc gaa ggt cgc agc cct tac tac ggc atc acc caa gac att 480
 Met Tyr Gly Glu Gly Arg Ser Pro Tyr Tyr Gly Ile Thr Gln Asp Ile
 145 150 155 160

cac cgc att tgg ccc gaa ctc cac gag gtc atc aac gaa aag acg aac 528
 His Arg Ile Trp Pro Glu Leu His Glu Val Ile Asn Glu Lys Thr Asn
 165 170 175

cgt ctc tgg gcc gaa ggc gac cgc tgg gtc atg ccg cgc gcc agc ttc 576
 Arg Leu Trp Ala Glu Gly Asp Arg Trp Val Met Pro Arg Ala Ser Phe
 180 185 190

aag tgc gag ctc gag agc cag cag caa gag ttt gat cgc aac atg att 624
 Lys Ser Glu Leu Glu Ser Gln Gln Gln Glu Phe Asp Arg Asn Met Ile
 195 200 205

gaa atg ttc cgt ctt gga atc ctc acc tca att gcc ttc acc aat ctg 672

Glu Met Phe Arg Leu Gly Ile Leu Thr Ser Ile Ala Phe Thr Asn Leu
 210 215 220

gcg cgc gac gtt ctc aac atc acg ccc aag gcc gcc ttt ggc ctc agt 720
 Ala Arg Asp Val Leu Asn Ile Thr Pro Lys Ala Ala Phe Gly Leu Ser
 225 230 235 240

ctt ggc gag att tcc atg att ttt gcc ttt tcc aag aag aac ggt ctc 768
 Leu Gly Glu Ile Ser Met Ile Phe Ala Phe Ser Lys Lys Asn Gly Leu
 245 250 255

atc tcc gac cag ctc acc aag gat ctt cgc gag tcc gac gtg tgg aac 816
 Ile Ser Asp Gln Leu Thr Lys Asp Leu Arg Glu Ser Asp Val Trp Asn
 260 265 270

aag gct ctg gcc gtt gaa ttt aat gcg ctg cgc gag gcc tgg ggc att 864
 Lys Ala Leu Ala Val Glu Phe Asn Ala Leu Arg Glu Ala Trp Gly Ile
 275 280 285

cca cag agt gtc ccc aag gac gag ttc tgg caa ggc tac att gtg cgc 912
 Pro Gln Ser Val Pro Lys Asp Glu Phe Trp Gln Gly Tyr Ile Val Arg
 290 295 300

ggc acc aag cag gat atc gag gcg gcc atc gcc ccg gac agc aag tac 960
 Gly Thr Lys Gln Asp Ile Glu Ala Ala Ile Ala Pro Asp Ser Lys Tyr
 305 310 315 320

gtg cgc ctc acc atc atc aat gat gcc aac acc gcc ctc att agc ggc 1008
 Val Arg Leu Thr Ile Ile Asn Asp Ala Asn Thr Ala Leu Ile Ser Gly
 325 330 335

aag ccc gac gcc tgc aag get gcg atc gcg cgt ctc ggt ggc aac att 1056
 Lys Pro Asp Ala Cys Lys Ala Ala Ile Ala Arg Leu Gly Gly Asn Ile
 340 345 350

ccr gcg ctt ccc gtg acc cag ggc atg tgc ggc cac tgc ccc gag gtg 1104
 Pro Ala Leu Pro Val Thr Gln Gly Met Cys Gly His Cys Pro Glu Val
 355 360 365

gga cct tat acc aag gat atc gcc aag atc cat gcc aac ctt gag ttc 1152
 Gly Pro Tyr Thr Lys Asp Ile Ala Lys Ile His Ala Asn Leu Glu Phe
 370 375 380

ccc gtt gtc gac ggc ctt gac ctc tgg acc aca atc aac cag aag cgc 1200
 Pro Val Val Asp Gly Leu Asp Leu Trp Thr Thr Ile Asn Gln Lys Arg
 385 390 395 400

ctc gtg cca cgc gcc acg ggc gcc aag gac gaa tgg gcc cct tct tcc 1248
 Leu Val Pro Arg Ala Thr Gly Ala Lys Asp Glu Trp Ala Pro Ser Ser
 405 410 415

ttt ggc gag tac gcc ggc cag ctc tac gag aag cag gct aac ttc ccc 1296
 Phe Gly Glu Tyr Ala Gly Gln Leu Tyr Glu Lys Gln Ala Asn Phe Pro
 420 425 430

caa atc gtc gag acc att tac aag caa aac tac gac gtc ttt gtc gag 1344
 Gln Ile Val Glu Thr Ile Tyr Lys Gln Asn Tyr Asp Val Phe Val Glu
 435 440 445

gtt ggg ccc aac aac cac cgt agc acc gca gtg cgc acc acg ctt ggt 1392
 Val Gly Pro Asn Asn His Arg Ser Thr Ala Val Arg Thr Thr Leu Gly
 450 455 460

ccc cag cgc aac cac ctt gct ggc gcc atc gac aag cag aac gag gat 1440
 Pro Gln Arg Asn His Leu Ala Gly Ala Ile Asp Lys Gln Asn Glu Asp
 465 470 475 480

gct tgg acg acc atc gtc aag ctt gtg gct tcg ctc aag gcc cac ctt 1488
 Ala Trp Thr Thr Ile Val Lys Leu Val Ala Ser Leu Lys Ala His Leu
 485 490 495

gtt cct ggc gtc 1500
 Val Pro Gly Val
 500

<210> 24 <211> 500 <212> PRT <213> 裂壺菌屬 sp. <400> 24

Cys Tyr Ser Val Leu Leu Ser Glu Ala Glu Gly His Tyr Glu Arg Glu
 1 5 10 15

Asn Arg Ile Ser Leu Asp Glu Glu Ala Pro Lys Leu Ile Val Leu Arg
 20 25 30

Ala Asp Ser His Glu Glu Ile Leu Gly Arg Leu Asp Lys Ile Arg Glu
 35 40 45

Arg Phe Leu Gln Pro Thr Gly Ala Ala Pro Arg Glu Ser Glu Leu Lys
 50 55 60

Ala Gln Ala Arg Arg Ile Phe Leu Glu Leu Leu Gly Glu Thr Leu Ala
 65 70 75 80

Gln Asp Ala Ala Ser Ser Gly Ser Gln Lys Pro Leu Ala Leu Ser Leu
 85 90 95

Val Ser Thr Pro Ser Lys Leu Gln Arg Glu Val Glu Leu Ala Ala Lys
 100 105 110

Gly Ile Pro Arg Cys Leu Lys Met Arg Arg Asp Trp Ser Ser Pro Ala
 115 120 125

Gly Ser Arg Tyr Ala Pro Glu Pro Leu Ala Ser Asp Arg Val Ala Phe
 130 135 140

Met Tyr Gly Glu Gly Arg Ser Pro Tyr Tyr Gly Ile Thr Gln Asp Ile
 145 150 155 160

His Arg Ile Trp Pro Glu Leu His Glu Val Ile Asn Glu Lys Thr Asn
 165 170 175

Arg Leu Trp Ala Glu Gly Asp Arg Trp Val Met Pro Arg Ala Ser Phe
 180 185 190

Lys Ser Glu Leu Glu Ser Gln Gln Gln Glu Phe Asp Arg Asn Met Ile
 195 200 205

Glu Met Phe Arg Leu Gly Ile Leu Thr Ser Ile Ala Phe Thr Asn Leu
 210 215 220

Ala Arg Asp Val Leu Asn Ile Thr Pro Lys Ala Ala Phe Gly Leu Ser
 225 230 235 240

Leu Gly Glu Ile Ser Met Ile Phe Ala Phe Ser Lys Lys Asn Gly Leu
 245 250 255

Ile Ser Asp Gln Leu Thr Lys Asp Leu Arg Glu Ser Asp Val Trp Asn
 260 265 270

Lys Ala Leu Ala Val Glu Phe Asn Ala Leu Arg Glu Ala Trp Gly Ile
 275 280 285

Pro Gln Ser Val Pro Lys Asp Glu Phe Trp Gln Gly Tyr Ile Val Arg
 290 295 300

Gly Thr Lys Gln Asp Ile Glu Ala Ala Ile Ala Pro Asp Ser Lys Tyr
 305 310 315 320

Val Arg Leu Thr Ile Ile Asn Asp Ala Asn Thr Ala Leu Ile Ser Gly
 325 330 335

Lys Pro Asp Ala Cys Lys Ala Ala Ile Ala Arg Leu Gly Gly Asn Ile
 340 345 350

Pro Ala Leu Pro Val Thr Gln Gly Met Cys Gly His Cys Pro Glu Val
 355 360 365

Gly Pro Tyr Thr Lys Asp Ile Ala Lys Ile His Ala Asn Leu Glu Phe
 370 375 380

Pro Val Val Asp Gly Leu Asp Leu Trp Thr Thr Ile Asn Gln Lys Arg
 385 390 395 400

Leu Val Pro Arg Ala Thr Gly Ala Lys Asp Glu Trp Ala Pro Ser Ser
 405 410 415

Phe Gly Glu Tyr Ala Gly Gln Leu Tyr Glu Lys Gln Ala Asn Phe Pro
 420 425 430

Gln Ile Val Glu Thr Ile Tyr Lys Gln Asn Tyr Asp Val Phe Val Glu
 435 440 445

Val Gly Pro Asn Asn His Arg Ser Thr Ala Val Arg Thr Thr Leu Gly
 450 455 460

Pro Gln Arg Asn His Leu Ala Gly Ala Ile Asp Lys Gln Asn Glu Asp
 465 470 475 480

Ala Trp Thr Thr Ile Val Lys Leu Val Ala Ser Leu Lys Ala His Leu
 485 490 495

Val Pro Gly Val
 500

<210> 25 <211> 1530 <212> DNA <213> 裂壺菌屬 sp. <220> <221> CDS <222> (1)..(1530)
 <223>

<400> 25

ctg ctc gat ctc gac agt atg ctt gcg ctg agc tct gcc agt gcc tcc 48
 Leu Leu Asp Leu Asp Ser Met Leu Ala Leu Ser Ser Ala Ser Ala Ser
 1 5 10 15

ggc aac ctt gtt gag act gcg cct agc gac gcc tcg gtc att gtg ccg 96
 Gly Asn Leu Val Glu Thr Ala Pro Ser Asp Ala Ser Val Ile Val Pro

20	25	30	
ccc tgc aac att gcg gat ctc ggc agc cgc gcc ttc atg aaa acg tac	144		
Pro Cys Asn Ile Ala Asp Leu Gly Ser Arg Ala Phe Met Lys Thr Tyr			
35	40	45	
ggt gtt tgg gcg cct ctg tac acg ggc gcc atg gcc aag ggc att gcc	192		
Gly Val Ser Ala Pro Leu Tyr Thr Gly Ala Met Ala Lys Gly Ile Ala			
50	55	60	
tct gcg gac ctc gtc att gcc gcc ggc cgc cag ggc atc ctt gcg tcc	240		
Ser Ala Asp Leu Val Ile Ala Ala Gly Arg Gln Gly Ile Leu Ala Ser			
65	70	75	80
ttt ggc gcc ggc gga ctt ccc atg cag gtt gtg cgt gag tcc atc gaa	288		
Phe Gly Ala Gly Gly Leu Pro Met Gln Val Val Arg Glu Ser Ile Glu			
85	90	95	
aag att cag gcc gcc ctg ccc aat ggc ccg tac gct gtc aac ctt atc	336		
Lys Ile Gln Ala Ala Leu Pro Asn Gly Pro Tyr Ala Val Asn Leu Ile			
100	105	110	
cat tct ccc ttt gac agc aac ctc gaa aag ggc aat gtc gat ctc ttc	384		
His Ser Pro Phe Asp Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe			
115	120	125	
ctc gag aag ggt gtc acc ttt gtc gag gcc tgg gcc ttt atg acg ctc	432		
Leu Glu Lys Gly Val Thr Phe Val Glu Ala Ser Ala Phe Met Thr Leu			
130	135	140	
acc ccg cag gtc gtg cgg tac cgc gcg gct ggc ctc acg cgc aac gcc	480		
Thr Pro Gln Val Val Arg Tyr Arg Ala Ala Gly Leu Thr Arg Asn Ala			
145	150	155	160
gac ggc tgg gtc aac atc cgc aac cgt atc att ggc aag gtc tgg cgc	528		
Asp Gly Ser Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg			
165	170	175	
acc gag ctc gcc gag atg ttc atg cgt cct gcg ccc gag cac ctt ctt	576		
Thr Glu Leu Ala Glu Met Phe Met Arg Pro Ala Pro Glu His Leu Leu			
180	185	190	
cag aag ctc att gct tcc ggc gag atc aac cag gag cag gcc gag ctc	624		
Gln Lys Leu Ile Ala Ser Gly Glu Ile Asn Gln Glu Gln Ala Glu Leu			
195	200	205	
gcc cgc cgt gtt ccc gtc gct gac gac atc gcg gtc gaa gct gac tgg	672		
Ala Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala Asp Ser			
210	215	220	
ggt ggc cac acc gac aac cgc ccc atc cac gtc att ctg ccc ctc atc	720		
Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile Leu Pro Leu Ile			
225	230	235	240

atc aac ctt cgc gac cgc ctt cac cgc gag tgc ggc tac ccg gcc aac 768
 Ile Asn Leu Arg Asp Arg Leu His Arg Glu Cys Gly Tyr Pro Ala Asn
 245 250 255

ctt cgc gtc cgt gtg ggc gcc ggc ggt ggc att ggg tgc ccc cag gcg 816
 Leu Arg Val Arg Val Gly Ala Gly Gly Gly Ile Gly Cys Pro Gln Ala
 260 265 270

gcg ctg gcc acc ttc aac atg ggt gcc tcc ttt att gtc acc ggc acc 864
 Ala Leu Ala Thr Phe Asn Met Gly Ala Ser Phe Ile Val Thr Gly Thr
 275 280 285

gtg aac cag gtc gcc aag cag tgc ggc acg tgc gac aat gtg cgc aag 912
 Val Asn Gln Val Ala Lys Gln Ser Gly Thr Cys Asp Asn Val Arg Lys
 290 295 300

cag ctc gcg aag gcc act tac tgc gac gta tgc atg gcc ccg gct gcc 960
 Gln Leu Ala Lys Ala Thr Tyr Ser Asp Val Cys Met Ala Pro Ala Ala
 305 310 315 320

gac atg ttc gag gaa ggc gtc aag ctt cag gtc ctc aag aag gga acc 1008
 Asp Met Phe Glu Glu Gly Val Lys Leu Gln Val Leu Lys Lys Gly Thr
 325 330 335

atg ttt ccc tgc cgc gcc aac aag ctc tac gag ctc ttt tgc aag tac 1056
 Met Phe Pro Ser Arg Ala Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr
 340 345 350

gac tgc ttc gag tcc atg ccc ccc gca gag ctt gcg cgc gtc gag aag 1104
 Asp Ser Phe Glu Ser Met Pro Pro Ala Glu Leu Ala Arg Val Glu Lys
 355 360 365

cgc atc ttc agc cgc gcg ctc gaa gag gtc tgg gac gag acc aaa aac 1152
 Arg Ile Phe Ser Arg Ala Leu Glu Glu Val Trp Asp Glu Thr Lys Asn
 370 375 380

ttt tac att aac cgt ctt cac aac ccg gag aag atc cag cgc gcc gag 1200
 Phe Tyr Ile Asn Arg Leu His Asn Pro Glu Lys Ile Gln Arg Ala Glu
 385 390 395 400

cgc gac ccc aag ctc aag atg tgc ctg ttc cgc tgg tac ctg agc 1248
 Arg Asp Pro Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Ser
 405 410 415

ctg gcg agc cgc tgg gcc aac act gga gct tcc gat cgc gtc atg gac 1296
 Leu Ala Ser Arg Trp Ala Asn Thr Gly Ala Ser Asp Arg Val Met Asp
 420 425 430

tac cag gtc tgg tgc ggt cct gcc att ggt tcc ttc aac gat ttc atc 1344
 Tyr Gln Val Trp Cys Gly Pro Ala Ile Gly Ser Phe Asn Asp Phe Ile
 435 440 445

aag gga act tac ctt gat ccg gcc gtc gca aac gag tac ccg tgc gtc 1392
 Lys Gly Thr Tyr Leu Asp Pro Ala Val Ala Asn Glu Tyr Pro Cys Val

450 455 460
 gtt cag att aac aag cag atc ctt cgt gga gcg tgc ttc ttg cgc cgt 1440
 Val Gln Ile Asn Lys Gln Ile Leu Arg Gly Ala Cys Phe Leu Arg Arg
 465 470 475 480
 ctc gaa att ctg cgc aac gca cgc ctt tcc gat ggc gct gcc gct ctt 1488
 Leu Glu Ile Leu Arg Asn Ala Arg Leu Ser Asp Gly Ala Ala Ala Leu
 485 490 495
 gtg gcc agc atc gat gac aca tac gtc ccg gcc gag aag ctg 1530
 Val Ala Ser Ile Asp Asp Thr Tyr Val Pro Ala Glu Lys Leu
 500 505 510

<210> 26 <211> 510 <212> PRT <213> 裂壺菌屬 sp. <400> 26

Leu Leu Asp Leu Asp Ser Met Leu Ala Leu Ser Ser Ala Ser Ala Ser
 1 5 10 15

Gly Asn Leu Val Glu Thr Ala Pro Ser Asp Ala Ser Val Ile Val Pro
 20 25 30

Pro Cys Asn Ile Ala Asp Leu Gly Ser Arg Ala Phe Met Lys Thr Tyr
 35 40 45

Gly Val Ser Ala Pro Leu Tyr Thr Gly Ala Met Ala Lys Gly Ile Ala
 50 55 60

Ser Ala Asp Leu Val Ile Ala Ala Gly Arg Gln Gly Ile Leu Ala Ser
 65 70 75 80

Phe Gly Ala Gly Gly Leu Pro Met Gln Val Val Arg Glu Ser Ile Glu
 85 90 95

Lys Ile Gln Ala Ala Leu Pro Asn Gly Pro Tyr Ala Val Asn Leu Ile
 100 105 110

His Ser Pro Phe Asp Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe
 115 120 125

Leu Glu Lys Gly Val Thr Phe Val Glu Ala Ser Ala Phe Met Thr Leu
 130 135 140

Thr Pro Gln Val Val Arg Tyr Arg Ala Ala Gly Leu Thr Arg Asn Ala

145 150 155 160
 Asp Gly Ser Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg
 165 170 175
 Thr Glu Leu Ala Glu Met Phe Met Arg Pro Ala Pro Glu His Leu Leu
 180 185 190
 Gln Lys Leu Ile Ala Ser Gly Glu Ile Asn Gln Glu Gln Ala Glu Leu
 195 200 205
 Ala Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala Asp Ser
 210 215 220
 Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile Leu Pro Leu Ile
 225 230 235 240
 Ile Asn Leu Arg Asp Arg Leu His Arg Glu Cys Gly Tyr Pro Ala Asn
 245 250 255
 Leu Arg Val Arg Val Gly Ala Gly Gly Gly Ile Gly Cys Pro Gln Ala
 260 265 270
 Ala Leu Ala Thr Phe Asn Met Gly Ala Ser Phe Ile Val Thr Gly Thr
 275 280 285
 Val Asn Gln Val Ala Lys Gln Ser Gly Thr Cys Asp Asn Val Arg Lys
 290 295 300
 Gln Leu Ala Lys Ala Thr Tyr Ser Asp Val Cys Met Ala Pro Ala Ala
 305 310 315 320
 Asp Met Phe Glu Glu Gly Val Lys Leu Gln Val Leu Lys Lys Gly Thr
 325 330 335
 Met Phe Pro Ser Arg Ala Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr
 340 345 350
 Asp Ser Phe Glu Ser Met Pro Pro Ala Glu Leu Ala Arg Val Glu Lys
 355 360 365

Arg Ile Phe Ser Arg Ala Leu Glu Glu Val Trp Asp Glu Thr Lys Asn
 370 375 380

Phe Tyr Ile Asn Arg Leu His Asn Pro Glu Lys Ile Gln Arg Ala Glu
 385 390 395 400

Arg Asp Pro Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Ser
 405 410 415

Leu Ala Ser Arg Trp Ala Asn Thr Gly Ala Ser Asp Arg Val Met Asp
 420 425 430

Tyr Gln Val Trp Cys Gly Pro Ala Ile Gly Ser Phe Asn Asp Phe Ile
 435 440 445

Lys Gly Thr Tyr Leu Asp Pro Ala Val Ala Asn Glu Tyr Pro Cys Val
 450 455 460

Val Gln Ile Asn Lys Gln Ile Leu Arg Gly Ala Cys Phe Leu Arg Arg
 465 470 475 480

Leu Glu Ile Leu Arg Asn Ala Arg Leu Ser Asp Gly Ala Ala Ala Leu
 485 490 495

Val Ala Ser Ile Asp Asp Thr Tyr Val Pro Ala Glu Lys Leu
 500 505 510

<210> 27 <211> 4512 <212> DNA <213> 裂壺菌屬 sp. <220> <221> CDS <222> (1)..(4512)
 <223>

<400> 27

atg gcg ctc cgt gtc aag acg aac aag aag cca tgc tgg gag atg acc 48
 Met Ala Leu Arg Val Lys Thr Asn Lys Lys Pro Cys Trp Glu Met Thr
 1 5 10 15

aag gag gag ctg acc agc ggc aag acc gag gtg ttc aac tat gag gaa 96
 Lys Glu Glu Leu Thr Ser Gly Lys Thr Glu Val Phe Asn Tyr Glu Glu
 20 25 30

ctc ctc gag ttc gca gag ggc gac atc gcc aag gtc ttc gga ccc gag 144
 Leu Leu Glu Phe Ala Glu Gly Asp Ile Ala Lys Val Phe Gly Pro Glu
 35 40 45

ttc gcc gtc atc gac aag tac ccg cgc cgc gtg cgc ctg ccc gcc cgc 192

Phe Ala Val Ile Asp Lys Tyr Pro Arg Arg Val Arg Leu Pro Ala Arg
50 55 60

gag tac ctg ctc gtc acc cgc gtc acc ctc atg gac gcc gag gtc aac 240
Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Asn
65 70 75 80

aac tac cgc gtc ggc gcc cgc atg gtc acc gag tac gat ctc ccc gtc 288
Asn Tyr Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Leu Pro Val
85 90 95

aac gga gag ctc tcc gag ggc gga gac tgc ccc tgg gcc gtc ctg gtc 336
Asn Gly Glu Leu Ser Glu Gly Gly Asp Cys Pro Trp Ala Val Leu Val
100 105 110

gag agt ggc cag tgc gat ctc atg ctc atc tcc tac atg ggc att gac 384
Glu Ser Gly Gln Cys Asp Leu Met Leu Ile Ser Tyr Met Gly Ile Asp
115 120 125

ttc cag aac cag ggc gac cgc gtc tac cgc ctg ctc aac acc acg ctc 432
Phe Gln Asn Gln Gly Asp Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu
130 135 140

acc ttt tac ggc gtc gcc cac gag ggc gag acc ctc gag tac gac att 480
Thr Phe Tyr Gly Val Ala His Glu Gly Glu Thr Leu Glu Tyr Asp Ile
145 150 155 160

cgc gtc acc ggc ttc gcc aag cgt ctc gac ggc ggc atc tcc atg ttc 528
Arg Val Thr Gly Phe Ala Lys Arg Leu Asp Gly Gly Ile Ser Met Phe
165 170 175

ttc ttc gag tac gac tgc tac gtc aac ggc cgc ctc ctc atc gag atg 576
Phe Phe Glu Tyr Asp Cys Tyr Val Asn Gly Arg Leu Leu Ile Glu Met
180 185 190

cgc gat ggc tgc gcc ggc ttc ttc acc aac gag gag ctc gac gcc ggc 624
Arg Asp Gly Cys Ala Gly Phe Phe Thr Asn Glu Glu Leu Asp Ala Gly
195 200 205

aag ggc gtc gtc ttc acc cgc ggc gac ctc gcc gcc cgc gcc aag atc 672
Lys Gly Val Val Phe Thr Arg Gly Asp Leu Ala Ala Arg Ala Lys Ile
210 215 220

cca aag cag gac gtc tcc ccc tac gcc gtc gcc ccc tgc ctc cac aag 720
Pro Lys Gln Asp Val Ser Pro Tyr Ala Val Ala Pro Cys Leu His Lys
225 230 235 240

acc aag ctc aac gaa aag gag atg cag acc ctc gtc gac aag gac tgg 768
Thr Lys Leu Asn Glu Lys Glu Met Gln Thr Leu Val Asp Lys Asp Trp
245 250 255

gca tcc gtc ttt ggc tcc aag aac ggc atg cgg gaa atc aac tac aaa 816
Ala Ser Val Phe Gly Ser Lys Asn Gly Met Pro Glu Ile Asn Tyr Lys
260 265 270

ctc tgc gcg cgt aag atg ctc atg att gac cgc gtc acc agc att gac 864
 Leu Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr Ser Ile Asp
 275 280 285

cac aag ggc ggt gtc tac ggc ctc ggt cag ctc gtc ggt gaa aag atc 912
 His Lys Gly Gly Val Tyr Gly Leu Gly Gln Leu Val Gly Glu Lys Ile
 290 295 300

ctc gag cgc gac cac tgg tac ttt ccc tgc cac ttt gtc aag gat cag 960
 Leu Glu Arg Asp His Trp Tyr Phe Pro Cys His Phe Val Lys Asp Gln
 305 310 315 320

gtc atg gcc gga tcc ctc gtc tcc gac ggc tgc agc cag atg ctc aag 1008
 Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Met Leu Lys
 325 330 335

atg tac atg atc tgg ctc ggc ctc cac ctc acc acc gga ccc ttt gac 1056
 Met Tyr Met Ile Trp Leu Gly Leu His Leu Thr Thr Gly Pro Phe Asp
 340 345 350

ttc cgc ccg gtc aac ggc cac ccc aac aag gtc cgc tgc cgc ggc caa 1104
 Phe Arg Pro Val Asn Gly His Pro Asn Lys Val Arg Cys Arg Gly Gln
 355 360 365

atc tcc ccg cac aag ggc aag ctc gtc tac gtc atg gag atc aag gag 1152
 Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Glu
 370 375 380

atg ggc ttc gac gag gac aac gac ccg tac gcc att gcc gac gtc aac 1200
 Met Gly Phe Asp Glu Asp Asn Asp Pro Tyr Ala Ile Ala Asp Val Asn
 385 390 395 400

atc att gat gtc gac ttc gaa aag ggc cag gac ttt agc ctc gac cgc 1248
 Ile Ile Asp Val Asp Phe Glu Lys Gly Gln Asp Phe Ser Leu Asp Arg
 405 410 415

atc agc gac tac ggc aag ggc gac ctc aac aag aag atc gtc gtc gac 1296
 Ile Ser Asp Tyr Gly Lys Gly Asp Leu Asn Lys Lys Ile Val Val Asp
 420 425 430

ttt aag ggc atc gct ctc aag atg cag aag cgc tcc acc aac aag aac 1344
 Phe Lys Gly Ile Ala Leu Lys Met Gln Lys Arg Ser Thr Asn Lys Asn
 435 440 445

ccc tcc aag gtt cag ccc gtc ttt gcc aac ggc gcc gcc act gtc ggc 1392
 Pro Ser Lys Val Gln Pro Val Phe Ala Asn Gly Ala Ala Thr Val Gly
 450 455 460

ccc gag gcc tcc aag gct tcc tcc ggc gcc agc gcc agc gcc gcc 1440
 Pro Glu Ala Ser Lys Ala Ser Ser Gly Ala Ser Ala Ser Ala Ser Ala
 465 470 475 480

gcc ccg gcc aag cct gcc ttc agc gcc gal gtt ctt gcg ccc aag ccc 1488

Ala Pro Ala Lys Pro Ala Phe Ser Ala Asp Val Leu Ala Pro Lys Pro
485 490 495

gtt gcc ctt occ gag cac atc ctc aag ggc gac gcc ctc gcc ccc aag 1536
Val Ala Leu Pro Glu His Ile Leu Lys Gly Asp Ala Leu Ala Pro Lys
500 505 510

gag atg toc tgg cac ccc atg gcc cgc atc ccg ggc aac ccg acg ccc 1584
Ser Phe Met Ser Trp His Pro Met Ala Arg Ile Pro Gly Asn Pro Thr Pro
515 520 525

tct ttt gcg ccc tcg gcc tac aag ccg cgc aac atc gcc ttt acg ccc 1632
Ser Phe Ala Pro Ser Ala Tyr Lys Pro Arg Asn Ile Ala Phe Thr Pro
530 535 540

ttc ccc ggc aac ccc aac gat aac gac cac acc ccg ggc aag atg ccg 1680
Phe Pro Gly Asn Pro Asn Asp Asn Asp His Thr Pro Gly Lys Met Pro
545 550 555 560

ctc acc tgg ttc aac atg gcc gag ttc atg gcc ggc aag gtc agc atg 1728
Leu Thr Trp Phe Asn Met Ala Glu Phe Met Ala Gly Lys Val Ser Met
565 570 575

tgc ctc ggc ccc gag ttc gcc aag ttc gac gac tcg aac acc agc cgc 1776
Cys Leu Gly Pro Glu Phe Ala Lys Phe Asp Asp Ser Asn Thr Ser Arg
580 585 590

agc ccc gct tgg gac ctc get ctc gtc acc cgc gcc gtg tct gtg tct 1824
Ser Pro Ala Trp Asp Leu Ala Leu Val Thr Arg Ala Val Ser Val Ser
595 600 605

gac ctc aag cac gtc aac tac cgc aac atc gac ctc gac ccc tcc aag 1872
Asp Leu Lys His Val Asn Tyr Arg Asn Ile Asp Leu Asp Pro Ser Lys
610 615 620

ggt acc atg gtc ggc gag ttc gac tgc ccc gcg gac gcc tgg ttc tac 1920
Gly Thr Met Val Gly Glu Phe Asp Cys Pro Ala Asp Ala Trp Phe Tyr
625 630 635 640

aag ggc gcc tgc aac gat gcc cac atg ccg tac tcg atc ctc aig gag 1968
Lys Gly Ala Cys Asn Asp Ala His Met Pro Tyr Ser Ile Leu Met Glu
645 650 655

atc gcc ctc cag acc tcg ggt gtg ctc acc tcg gtg ctc aag gcg ccc 2016
Ile Ala Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro
660 665 670

ctg acc atg gag aag gac gac atc ctc ttc cgc aac ctc gac gcc aac 2064
Leu Thr Met Glu Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Asn
675 680 685

gcc gag ttc gtg cgc gcc gac ctc gac tac cgc ggc aag act atc cgc 2112
Ala Glu Phe Val Arg Ala Asp Leu Asp Tyr Arg Gly Lys Thr Ile Arg
690 695 700

aac gtc acc aag tgc act ggc tac agc atg ctc ggc gag atg ggc gtc 2160
 Asn Val Thr Lys Cys Thr Gly Tyr Ser Met Leu Gly Glu Met Gly Val
 705 710 715 720

cac cgc ttc acc ttt gag ctc tac gtc gat gat gtg ctc ttt tac aag 2208
 His Arg Phe Thr Phe Glu Leu Tyr Val Asp Asp Val Leu Phe Tyr Lys
 725 730 735

ggc tcg acc tcg ttc ggc tgg ttc gtg ccc gag gtc ttt gcc gcc cag 2256
 Gly Ser Thr Ser Phe Gly Trp Phe Val Pro Glu Val Phe Ala Ala Gln
 740 745 750

gcc ggc ctc gac aac ggc cgc aag tcg gag ccc tgg ttc att gag aac 2304
 Ala Gly Leu Asp Asn Gly Arg Lys Ser Glu Pro Trp Phe Ile Glu Asn
 755 760 765

aag gtt ccg gcc tcg cag gtc tcc tcc ttt gac gtg cgc ccc aac ggc 2352
 Lys Val Pro Ala Ser Gln Val Ser Ser Phe Asp Val Arg Pro Asn Gly
 770 775 780

agc ggc cgc acc gcc atc ttc gcc aac gcc ccc agc ggc gcc cag ctc 2400
 Ser Gly Arg Thr Ala Ile Phe Ala Asn Ala Pro Ser Gly Ala Gln Leu
 785 790 795 800

aac cgc cgc acg gac cag ggc cag tac ctc gac gcc gtc gac att gtc 2448
 Asn Arg Arg Thr Asp Gln Gly Gln Tyr Leu Asp Ala Val Asp Ile Val
 805 810 815

tcc ggc agc ggc aag aag agc ctc ggc tac gcc cac ggt tcc aag acg 2496
 Ser Gly Ser Gly Lys Lys Ser Leu Gly Tyr Ala His Gly Ser Lys Thr
 820 825 830

gtc aac ccg aac gac tgg ttc ttc tcg tgc cac ttt tgg ttt gac tcg 2544
 Val Asn Pro Asn Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Ser
 835 840 845

gtc atg ccc gga agt ctc ggt gtc gag tcc atg ttc cag ctc gtc gag 2592
 Val Met Pro Gly Ser Leu Gly Val Glu Ser Met Phe Gln Leu Val Glu
 850 855 860

gcc atc gcc gcc cac gag gat ctc gct ggc aaa gca cgg cat tgc caa 2640
 Ala Ile Ala Ala His Glu Asp Leu Ala Gly Lys Ala Arg His Cys Gln
 865 870 875 880

ccc cac ctt tgt gca cgc ccc cgg gca aga tca agc tgg aag tac cgc 2688
 Pro His Leu Cys Ala Arg Pro Arg Ala Arg Ser Ser Trp Lys Tyr Arg
 885 890 895

ggc cag ctc acg ccc aag agc aag aag atg gac tcg gag gtc cac atc 2736
 Gly Gln Leu Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile
 900 905 910

gtg tcc gtg gac gcc cac gac ggc gtt gtc gac ctc gtc gcc gac ggc 2784

Val Ser Val Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly
915 920 925

ttc ctc tgg gcc gac agc ctc cgc gtc tac tcg gtg agc aac att cgc 2832
Phe Leu Trp Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg
930 935 940

gtg cgc atc gcc tcc ggt gag gcc cct gcc gcc gcc tcc tcc gcc gcc 2880
Val Arg Ile Ala Ser Gly Glu Ala Pro Ala Ala Ala Ser Ser Ala Ala
945 950 955 960

tct gtg ggc tcc tcg gct tcg tcc gtc gag cgc acg cgc tcg agc ccc 2928
Ser Val Gly Ser Ser Ala Ser Ser Val Glu Arg Thr Arg Ser Ser Pro
965 970 975

gct gtc gcc tcc ggc ccg gcc cag acc atc gac ctc aag cag ctc aag 2976
Ala Val Ala Ser Gly Pro Ala Gln Thr Ile Asp Leu Lys Gln Leu Lys
980 985 990

acc gag ctc ctc gag ctc gat gcc ccg ctc tac ctc tcg cag gac ccg 3024
Thr Glu Leu Leu Glu Leu Asp Ala Pro Leu Tyr Leu Ser Gln Asp Pro
995 1000 1005

acc agc ggc cag ctc aag aag cac acc gac gtg gcc tcc ggc cag 3069
Thr Ser Gly Gln Leu Lys Lys His Thr Asp Val Ala Ser Gly Gln
1010 1015 1020

gcc acc atc gtg cag ccc tgc acg ctc ggc gac ctc ggt gac cgc 3114
Ala Thr Ile Val Gln Pro Cys Thr Leu Gly Asp Leu Gly Asp Arg
1025 1030 1035

tcc ttc atg gag acc tac ggc gtc gtc gcc ccg ctg tac acg ggc 3159
Ser Phe Met Glu Thr Tyr Gly Val Val Ala Pro Leu Tyr Thr Gly
1040 1045 1050

gcc atg gcc aag ggc att gcc tcg gcg gac ctc gtc atc gcc gcc 3204
Ala Met Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala Ala
1055 1060 1065

ggc aag cgc aag atc ctc ggc tcc ttt ggc gcc ggc ggc ctc ccc 3249
Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala Gly Gly Leu Pro
1070 1075 1080

atg cac cac gtg cgc gcc gcc ctc gag aag atc cag gcc gcc ctg 3294
Met His His Val Arg Ala Ala Leu Glu Lys Ile Gln Ala Ala Leu
1085 1090 1095

cct cag ggc ccc tac gcc gtc aac ctc atc cac tcg cct ttt gac 3339
Pro Gln Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp
1100 1105 1110

agc aac ctc gag aag ggc aac gtc gat ctc ttc ctc gag aag ggc 3384
Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys Gly
1115 1120 1125

gtc act gtg gtg gag gcc tcg gca ttc atg acc ctc acc ccg cag 3429
 Val Thr Val Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln
 1130 1135 1140

gtc gtg cgc tac cgc gcc gcc ggc ctc tcg cgc aac gcc gac ggt 3474
 Val Val Arg Tyr Arg Ala Ala Gly Leu Ser Arg Asn Ala Asp Gly
 1145 1150 1155

tcg gtc aac atc cgc aac cgc atc atc ggc aag gtc tcg cgc acc 3519
 Ser Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr
 1160 1165 1170

gag ctc gcc gag atg ttc atc cgc ccg gcc ccg gag cac ctc ctc 3564
 Glu Leu Ala Glu Met Phe Ile Arg Pro Ala Pro Glu His Leu Leu
 1175 1180 1185

gag aag ctc atc gcc tcg ggc gag atc acc cag gag cag gcc gag 3609
 Glu Lys Leu Ile Ala Ser Gly Glu Ile Thr Gln Glu Gln Ala Glu
 1190 1195 1200

ctc gcg cgc cgc gtt ccc gtc gcc gac gat atc gct gtc gag gct 3654
 Leu Ala Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala
 1205 1210 1215

gac tcg ggc ggc cac acc gac aac cgc ccc atc cac gtc atc ctc 3699
 Asp Ser Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile Leu
 1220 1225 1230

ccg ctc atc atc aac ctc cgc aac cgc ctg cac cgc gag tgc ggc 3744
 Pro Leu Ile Ile Asn Leu Arg Asn Arg Leu His Arg Glu Cys Gly
 1235 1240 1245

tac ccc gcg cac ctc cgc gtc cgc gtt ggc gcc ggc ggt ggc gtc 3789
 Tyr Pro Ala His Leu Arg Val Arg Val Gly Ala Gly Gly Gly Val
 1250 1255 1260

ggc tgc ccg cag gcc gcc gcc gcc gcg ctc acc atg ggc gcc gcc 3834
 Gly Cys Pro Gln Ala Ala Ala Ala Ala Leu Thr Met Gly Ala Ala
 1265 1270 1275

ttc atc gtc acc ggc act gtc aac cag gtc gcc aag cag tcc ggc 3879
 Phe Ile Val Thr Gly Thr Val Asn Gln Val Ala Lys Gln Ser Gly
 1280 1285 1290

acc tgc gac aac gtg cgc aag cag ctc tcg cag gcc acc tac tcg 3924
 Thr Cys Asp Asn Val Arg Lys Gln Leu Ser Gln Ala Thr Tyr Ser
 1295 1300 1305

gat atc tgc atg gcc ccg gcc gcc gac atg ttc gag gag ggc gtc 3969
 Asp Ile Cys Met Ala Pro Ala Ala Asp Met Phe Glu Glu Gly Val
 1310 1315 1320

aag ctc cag gtc ctc aag aag gga acc atg ttc ccc tcg cgc gcc 4014

Lys Leu Gln Val Leu Lys Lys Gly Thr Met Phe Pro Ser Arg Ala
 1325 1330 1335

aac aag ctc tac gag ctc ttt tgc aag tac gac tcc ttc gac tcc 4059
 Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe Asp Ser
 1340 1345 1350

atg cct cct gcc gag ctc gag cgc atc gag aag cgt atc ttc aag 4104
 Met Pro Pro Ala Glu Leu Glu Arg Ile Glu Lys Arg Ile Phe Lys
 1355 1360 1365

cgc gca ctc cag gag gtc tgg gag gag acc aag gac ttt tac att 4149
 Arg Ala Leu Gln Glu Val Trp Glu Glu Thr Lys Asp Phe Tyr Ile
 1370 1375 1380

aac ggt ctc aag aac ccg gag aag atc cag cgc gcc gag cac gac 4194
 Asn Gly Leu Lys Asn Pro Glu Lys Ile Gln Arg Ala Glu His Asp
 1385 1390 1395

ccc aag ctc aag atg tgc ctc tgc ttc cgc tgg tac ctt ggt ctt 4239
 Pro Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu
 1400 1405 1410

gcc agc cgc tgg gcc aac atg ggc gcc ccg gac cgc gtc atg gac 4284
 Ala Ser Arg Trp Ala Asn Met Gly Ala Pro Asp Arg Val Met Asp
 1415 1420 1425

tac cag gtc tgg tgt ggc ccg gcc att ggc gcc ttc aac gac ttc 4329
 Tyr Gln Val Trp Cys Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe
 1430 1435 1440

atc aag ggc acc tac ctc gac ccc gct gtc tcc aac gag tac ccc 4374
 Ile Lys Gly Thr Tyr Leu Asp Pro Ala Val Ser Asn Glu Tyr Pro
 1445 1450 1455

tgt gtc gtc cag atc aac ctg caa atc ctc cgt ggt gcc tgc tac 4419
 Cys Val Val Gln Ile Asn Leu Gln Ile Leu Arg Gly Ala Cys Tyr
 1460 1465 1470

ctg cgc cgt ctc aac gcc ctg cgc aac gac ccg cgc att gac ctc 4464
 Leu Arg Arg Leu Asn Ala Leu Arg Asn Asp Pro Arg Ile Asp Leu
 1475 1480 1485

gag acc gag gat gct gcc ttt gtc tac gag ccc acc aac gcg ctc 4509
 Glu Thr Glu Asp Ala Ala Phe Val Tyr Glu Pro Thr Asn Ala Leu
 1490 1495 1500

taa 4512

<210> 28 <211> 1503 <212> PRT <213> 裂壺菌屬 sp. <400> 28

Met Ala Leu Arg Val Lys Thr Asn Lys Lys Pro Cys Trp Glu Met Thr
 1 5 10 15

Lys Glu Glu Leu Thr Ser Gly Lys Thr Glu Val Phe Asn Tyr Glu Glu
 20 25 30

Leu Leu Glu Phe Ala Glu Gly Asp Ile Ala Lys Val Phe Gly Pro Glu
 35 40 45

Phe Ala Val Ile Asp Lys Tyr Pro Arg Arg Val Arg Leu Pro Ala Arg
 50 55 60

Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Asn
 65 70 75 80

Asn Tyr Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Leu Pro Val
 85 90 95

Asn Gly Glu Leu Ser Glu Gly Gly Asp Cys Pro Trp Ala Val Leu Val
 100 105 110

Glu Ser Gly Gln Cys Asp Leu Met Leu Ile Ser Tyr Met Gly Ile Asp
 115 120 125

Phe Gln Asn Gln Gly Asp Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu
 130 135 140

Thr Phe Tyr Gly Val Ala His Glu Gly Glu Thr Leu Glu Tyr Asp Ile
 145 150 155 160

Arg Val Thr Gly Phe Ala Lys Arg Leu Asp Gly Gly Ile Ser Met Phe
 165 170 175

Phe Phe Glu Tyr Asp Cys Tyr Val Asn Gly Arg Leu Leu Ile Glu Met
 180 185 190

Arg Asp Gly Cys Ala Gly Phe Phe Thr Asn Glu Glu Leu Asp Ala Gly
 195 200 205

Lys Gly Val Val Phe Thr Arg Gly Asp Leu Ala Ala Arg Ala Lys Ile
 210 215 220

Pro Lys Gln Asp Val Ser Pro Tyr Ala Val Ala Pro Cys Leu His Lys
225 230 235 240

Thr Lys Leu Asn Glu Lys Glu Met Gln Thr Leu Val Asp Lys Asp Trp
 245 250 255

Ala Ser Val Phe Gly Ser Lys Asn Gly Met Pro Glu Ile Asn Tyr Lys
 260 265 270

Leu Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr Ser Ile Asp
 275 280 285

His Lys Gly Gly Val Tyr Gly Leu Gly Gln Leu Val Gly Glu Lys Ile
 290 295 300

Leu Glu Arg Asp His Trp Tyr Phe Pro Cys His Phe Val Lys Asp Gln
305 310 315 320

Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Met Leu Lys
 325 330 335

Met Tyr Met Ile Trp Leu Gly Leu His Leu Thr Thr Gly Pro Phe Asp
 340 345 350

Phe Arg Pro Val Asn Gly His Pro Asn Lys Val Arg Cys Arg Gly Gln
 355 360 365

Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Glu
 370 375 380

Met Gly Phe Asp Glu Asp Asn Asp Pro Tyr Ala Ile Ala Asp Val Asn
385 390 395 400

Ile Ile Asp Val Asp Phe Glu Lys Gly Gln Asp Phe Ser Leu Asp Arg
 405 410 415

Ile Ser Asp Tyr Gly Lys Gly Asp Leu Asn Lys Lys Ile Val Val Asp
 420 425 430

Phe Lys Gly Ile Ala Leu Lys Met Gln Lys Arg Ser Thr Asn Lys Asn
 435 440 445

Pro Ser Lys Val Gln Pro Val Phe Ala Asn Gly Ala Ala Thr Val Gly
 450 455 460

Pro Glu Ala Ser Lys Ala Ser Ser Gly Ala Ser Ala Ser Ala Ser Ala
 465 470 475 480

Ala Pro Ala Lys Pro Ala Phe Ser Ala Asp Val Leu Ala Pro Lys Pro
 485 490 495

Val Ala Leu Pro Glu His Ile Leu Lys Gly Asp Ala Leu Ala Pro Lys
 500 505 510

Glu Met Ser Trp His Pro Met Ala Arg Ile Pro Gly Asn Pro Thr Pro
 515 520 525

Ser Phe Ala Pro Ser Ala Tyr Lys Pro Arg Asn Ile Ala Phe Thr Pro
 530 535 540

Phe Pro Gly Asn Pro Asn Asp Asn Asp His Thr Pro Gly Lys Met Pro
 545 550 555 560

Leu Thr Trp Phe Asn Met Ala Glu Phe Met Ala Gly Lys Val Ser Met
 565 570 575

Cys Leu Gly Pro Glu Phe Ala Lys Phe Asp Asp Ser Asn Thr Ser Arg
 580 585 590

Ser Pro Ala Trp Asp Leu Ala Leu Val Thr Arg Ala Val Ser Val Ser
 595 600 605

Asp Leu Lys His Val Asn Tyr Arg Asn Ile Asp Leu Asp Pro Ser Lys
 610 615 620

Gly Thr Met Val Gly Glu Phe Asp Cys Pro Ala Asp Ala Trp Phe Tyr
 625 630 635 640

Lys Gly Ala Cys Asn Asp Ala His Met Pro Tyr Ser Ile Leu Met Glu
 645 650 655

Ile Ala Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro
660 665 670

Leu Thr Met Glu Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Asn
675 680 685

Ala Glu Phe Val Arg Ala Asp Leu Asp Tyr Arg Gly Lys Thr Ile Arg
690 695 700

Asn Val Thr Lys Cys Thr Gly Tyr Ser Met Leu Gly Glu Met Gly Val
705 710 715 720

His Arg Phe Thr Phe Glu Leu Tyr Val Asp Asp Val Leu Phe Tyr Lys
725 730 735

Gly Ser Thr Ser Phe Gly Trp Phe Val Pro Glu Val Phe Ala Ala Gln
740 745 750

Ala Gly Leu Asp Asn Gly Arg Lys Ser Glu Pro Trp Phe Ile Glu Asn
755 760 765

Lys Val Pro Ala Ser Gln Val Ser Ser Phe Asp Val Arg Pro Asn Gly
770 775 780

Ser Gly Arg Thr Ala Ile Phe Ala Asn Ala Pro Ser Gly Ala Gln Leu
785 790 795 800

Asn Arg Arg Thr Asp Gln Gly Gln Tyr Leu Asp Ala Val Asp Ile Val
805 810 815

Ser Gly Ser Gly Lys Lys Ser Leu Gly Tyr Ala His Gly Ser Lys Thr
820 825 830

Val Asn Pro Asn Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Ser
835 840 845

Val Met Pro Gly Ser Leu Gly Val Glu Ser Met Phe Gln Leu Val Glu
850 855 860

Ala Ile Ala Ala His Glu Asp Leu Ala Gly Lys Ala Arg His Cys Gln
865 870 875 880

Pro His Leu Cys Ala Arg Pro Arg Ala Arg Ser Ser Trp Lys Tyr Arg
 885 890 895

Gly Gln Leu Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile
 900 905 910

Val Ser Val Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly
 915 920 925

Phe Leu Trp Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg
 930 935 940

Val Arg Ile Ala Ser Gly Glu Ala Pro Ala Ala Ala Ser Ser Ala Ala
 945 950 955 960

Ser Val Gly Ser Ser Ala Ser Ser Val Glu Arg Thr Arg Ser Ser Pro
 965 970 975

Ala Val Ala Ser Gly Pro Ala Gln Thr Ile Asp Leu Lys Gln Leu Lys
 980 985 990

Thr Glu Leu Leu Glu Leu Asp Ala Pro Leu Tyr Leu Ser Gln Asp Pro
 995 1000 1005

Thr Ser Gly Gln Leu Lys Lys His Thr Asp Val Ala Ser Gly Gln
 1010 1015 1020

Ala Thr Ile Val Gln Pro Cys Thr Leu Gly Asp Leu Gly Asp Arg
 1025 1030 1035

Ser Phe Met Glu Thr Tyr Gly Val Val Ala Pro Leu Tyr Thr Gly
 1040 1045 1050

Ala Met Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala Ala
 1055 1060 1065

Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala Gly Gly Leu Pro
 1070 1075 1080

Met His His Val Arg Ala Ala Leu Glu Lys Ile Gln Ala Ala Leu
 1085 1090 1095

Pro Gln Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp
 1100 1105 1110

Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys Gly
 1115 1120 1125

Val Thr Val Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln
 1130 1135 1140

Val Val Arg Tyr Arg Ala Ala Gly Leu Ser Arg Asn Ala Asp Gly
 1145 1150 1155

Ser Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr
 1160 1165 1170

Glu Leu Ala Glu Met Phe Ile Arg Pro Ala Pro Glu His Leu Leu
 1175 1180 1185

Glu Lys Leu Ile Ala Ser Gly Glu Ile Thr Gln Glu Gln Ala Glu
 1190 1195 1200

Leu Ala Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala
 1205 1210 1215

Asp Ser Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile Leu
 1220 1225 1230

Pro Leu Ile Ile Asn Leu Arg Asn Arg Leu His Arg Glu Cys Gly
 1235 1240 1245

Tyr Pro Ala His Leu Arg Val Arg Val Gly Ala Gly Gly Gly Val
 1250 1255 1260

Gly Cys Pro Gln Ala Ala Ala Ala Ala Leu Thr Met Gly Ala Ala
 1265 1270 1275

Phe Ile Val Thr Gly Thr Val Asn Gln Val Ala Lys Gln Ser Gly
 1280 1285 1290

Thr Cys Asp Asn Val Arg Lys Gln Leu Ser Gln Ala Thr Tyr Ser
 1295 1300 1305

Asp Ile Cys Met Ala Pro Ala Ala Asp Met Phe Glu Glu Gly Val
 1310 1315 1320

Lys Leu Gln Val Leu Lys Lys Gly Thr Met Phe Pro Ser Arg Ala
 1325 1330 1335

Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe Asp Ser
 1340 1345 1350

Met Pro Pro Ala Glu Leu Glu Arg Ile Glu Lys Arg Ile Phe Lys
 1355 1360 1365

Arg Ala Leu Gln Glu Val Trp Glu Glu Thr Lys Asp Phe Tyr Ile
 1370 1375 1380

Asn Gly Leu Lys Asn Pro Glu Lys Ile Gln Arg Ala Glu His Asp
 1385 1390 1395

Pro Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu
 1400 1405 1410

Ala Ser Arg Trp Ala Asn Met Gly Ala Pro Asp Arg Val Met Asp
 1415 1420 1425

Tyr Gln Val Trp Cys Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe
 1430 1435 1440

Ile Lys Gly Thr Tyr Leu Asp Pro Ala Val Ser Asn Glu Tyr Pro
 1445 1450 1455

Cys Val Val Gln Ile Asn Leu Gln Ile Leu Arg Gly Ala Cys Tyr
 1460 1465 1470

Leu Arg Arg Leu Asn Ala Leu Arg Asn Asp Pro Arg Ile Asp Leu
 1475 1480 1485

Glu Thr Glu Asp Ala Ala Phe Val Tyr Glu Pro Thr Asn Ala Leu
 1490 1495 1500

<210> 29 <211> 1500 <212> DNA <213> 裂壺菌屬 sp. <220> <221> CDS <222> (1)..(1500)
 <223>

<400> 29

aag gtt cag ccc gtc ttt gcc aac ggc gcc gcc act gtc ggc ccc gag 48
 Lys Val Gln Pro Val Phe Ala Asn Gly Ala Ala Thr Val Gly Pro Glu
 1 5 10 15

gcc tcc aag gct tcc tcc ggc gcc agc gcc agc gcc agc gcc gcc ccg 96
 Ala Ser Lys Ala Ser Ser Gly Ala Ser Ala Ser Ala Ser Ala Ala Pro
 20 25 30

gcc aag cct gcc ttc agc gcc gat gtt ctt gcg ccc aag ccc gtt gcc 144
 Ala Lys Pro Ala Phe Ser Ala Asp Val Leu Ala Pro Lys Pro Val Ala
 35 40 45

ctt ccc gag cac atc ctc aag ggc gac gcc ctc gcc ccc aag gag atg 192
 Leu Pro Glu His Ile Leu Lys Gly Asp Ala Leu Ala Pro Lys Glu Met
 50 55 60

tcc tgg cac ccc atg gcc cgc atc ccg ggc aac ccg acg ccc tct ttt 240
 Ser Trp His Pro Met Ala Arg Ile Pro Gly Asn Pro Thr Pro Ser Phe
 65 70 75 80

gcg ccc tcg gcc tac aag ccg cgc aac atc gcc ttt acg ccc ttc ccc 288
 Ala Pro Ser Ala Tyr Lys Pro Arg Asn Ile Ala Phe Thr Pro Phe Pro
 85 90 95

ggc aac ccc aac gat aac gac cac acc cgg ggc aag atg cgg ctc acc 336
 Gly Asn Pro Asn Asp Asn Asp His Thr Pro Gly Lys Met Pro Leu Thr
 100 105 110

tgg ttc aac atg gcc gag ttc atg gcc ggc aag gtc agc atg tgc ctc 384
 Trp Phe Asn Met Ala Glu Phe Met Ala Gly Lys Val Ser Met Cys Leu
 115 120 125

ggc ccc gag ttc gcc aag ttc gac gac tcg aac acc agc cgc agc ccc 432
 Gly Pro Glu Phe Ala Lys Phe Asp Asp Ser Asn Thr Ser Arg Ser Pro
 130 135 140

gct tgg gac ctc gct ctc gtc acc cgc gcc gtg tct gtg tct gac ctc 480
 Ala Trp Asp Leu Ala Leu Val Thr Arg Ala Val Ser Val Ser Asp Leu
 145 150 155 160

aag cac gtc aac tac cgc aac atc gac ctc gac ccc tcc aag ggt acc 528
 Lys His Val Asn Tyr Arg Asn Ile Asp Leu Asp Pro Ser Lys Gly Thr
 165 170 175

atg gtc ggc gag ttc gac tgc ccc gcg gac gcc tgg ttc tac aag ggc 576
 Met Val Gly Glu Phe Asp Cys Pro Ala Asp Ala Trp Phe Tyr Lys Gly

180 185 190
 gcc tgc aac gat gcc cac atg ccg tac tcg atc ctc atg gag atc gcc 624
 Ala Cys Asn Asp Ala His Met Pro Tyr Ser Ile Leu Met Glu Ile Ala
 195 200 205

 ctc cag acc tcg ggt gtg ctc acc tcg gtg ctc aag gcg ccc ctg acc 672
 Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro Leu Thr
 210 215 220

 atg gag aag gac gac atc ctc ttc cgc aac ctc gac gcc aac gcc gag 720
 Met Glu Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Asn Ala Glu
 225 230 235 240

 ttc gtg cgc gcc gac ctc gac tac cgc ggc aag act atc cgc aac gtc 768
 Phe Val Arg Ala Asp Leu Asp Tyr Arg Gly Lys Thr Ile Arg Asn Val
 245 250 255

 acc aag tgc act ggc tac agc atg ctc ggc gag atg ggc gtc cac cgc 816
 Thr Lys Cys Thr Gly Tyr Ser Met Leu Gly Glu Met Gly Val His Arg
 260 265 270

 ttc acc ttt gag ctc tac gtc gat gat gtg ctc ttt tac aag ggc tcg 864
 Phe Thr Phe Glu Leu Tyr Val Asp Asp Val Leu Phe Tyr Lys Gly Ser
 275 280 285

 acc tcg ttc ggc tgg ttc gtg ccc gag gtc ttt gcc gcc cag gcc ggc 912
 Thr Ser Phe Gly Trp Phe Val Pro Glu Val Phe Ala Ala Gln Ala Gly
 290 295 300

 ctc gac aac ggc cgc aag tcg gag ccc tgg ttc att gag aac aag gtt 960
 Leu Asp Asn Gly Arg Lys Ser Glu Pro Trp Phe Ile Glu Asn Lys Val
 305 310 315 320

 ccg gcc tcg cag gtc tcc tcc ttt gac gtg cgc ccc aac ggc agc ggc 1008
 Pro Ala Ser Gln Val Ser Ser Phe Asp Val Arg Pro Asn Gly Ser Gly
 325 330 335

 cgc acc gcc atc ttc gcc aac gcc ccc agc ggc gcc cag ctc aac cgc 1056
 Arg Thr Ala Ile Phe Ala Asn Ala Pro Ser Gly Ala Gln Leu Asn Arg
 340 345 350

 cgc acg gac cag ggc cag tac ctc gac gcc gtc gac att gtc tcc ggc 1104
 Arg Thr Asp Gln Gly Gln Tyr Leu Asp Ala Val Asp Ile Val Ser Gly
 355 360 365

 agc ggc aag aag agc ctc ggc tac gcc cac ggt tcc aag acg gtc aac 1152
 Ser Gly Lys Lys Ser Leu Gly Tyr Ala His Gly Ser Lys Thr Val Asn
 370 375 380

 ccg aac gac tgg ttc ttc tcg tgc cac ttt tgg ttt gac tcg gtc atg 1200
 Pro Asn Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Ser Val Met
 385 390 395 400

ccc gga agt ctc ggt gtc gag tcc atg ttc cag ctc gtc gag gcc atc 1248
 Pro Gly Ser Leu Gly Val Glu Ser Met Phe Gln Leu Val Glu Ala Ile
 405 410 415

gcc gcc cac gag gat ctc gct ggc aaa gca cgg cat tgc caa ccc cac 1296
 Ala Ala His Glu Asp Leu Ala Gly Lys Ala Arg His Cys Gln Pro His
 420 425 430

ctt tgt gca cgc ccc cgg gca aga tca agc tgg aag tac cgc ggc cag 1344
 Leu Cys Ala Arg Pro Arg Ala Arg Ser Ser Trp Lys Tyr Arg Gly Gln
 435 440 445

ctc acg ccc aag agc aag aag atg gac tcg gag gtc cac atc gtg tcc 1392
 Leu Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile Val Ser
 450 455 460

gtg gac gcc cac gac ggc gtt gtc gac ctc gtc gcc gac ggc ttc ctc 1440
 Val Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly Phe Leu
 465 470 475 480

tgg gcc gac agc ctc cgc gtc tac tcg gtg agc aac att cgc gtg cgc 1488
 Trp Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg Val Arg
 485 490 495

atc gcc tcc ggt 1500
 Ile Ala Ser Gly
 500

<210> 30 <211> 500 <212> PRT <213> 裂壺菌屬 sp. <400> 30

Lys Val Gln Pro Val Phe Ala Asn Gly Ala Ala Thr Val Gly Pro Glu
 1 5 10 15

Ala Ser Lys Ala Ser Ser Gly Ala Ser Ala Ser Ala Ser Ala Ala Pro
 20 25 30

Ala Lys Pro Ala Phe Ser Ala Asp Val Leu Ala Pro Lys Pro Val Ala
 35 40 45

Leu Pro Glu His Ile Leu Lys Gly Asp Ala Leu Ala Pro Lys Glu Met
 50 55 60

Ser Trp His Pro Met Ala Arg Ile Pro Gly Asn Pro Thr Pro Ser Phe
 65 70 75 80

Ala Pro Ser Ala Tyr Lys Pro Arg Asn Ile Ala Phe Thr Pro Phe Pro
 85 90 95

Gly Asn Pro Asn Asp Asn Asp His Thr Pro Gly Lys Met Pro Leu Thr
 100 105 110

Trp Phe Asn Met Ala Glu Phe Met Ala Gly Lys Val Ser Met Cys Leu
 115 120 125

Gly Pro Glu Phe Ala Lys Phe Asp Asp Ser Asn Thr Ser Arg Ser Pro
 130 135 140

Ala Trp Asp Leu Ala Leu Val Thr Arg Ala Val Ser Val Ser Asp Leu
 145 150 155 160

Lys His Val Asn Tyr Arg Asn Ile Asp Leu Asp Pro Ser Lys Gly Thr
 165 170 175

Met Val Gly Glu Phe Asp Cys Pro Ala Asp Ala Trp Phe Tyr Lys Gly
 180 185 190

Ala Cys Asn Asp Ala His Met Pro Tyr Ser Ile Leu Met Glu Ile Ala
 195 200 205

Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro Leu Thr
 210 215 220

Met Glu Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Asn Ala Glu
 225 230 235 240

Phe Val Arg Ala Asp Leu Asp Tyr Arg Gly Lys Thr Ile Arg Asn Val
 245 250 255

Thr Lys Cys Thr Gly Tyr Ser Met Leu Gly Glu Met Gly Val His Arg
 260 265 270

Phe Thr Phe Glu Leu Tyr Val Asp Asp Val Leu Phe Tyr Lys Gly Ser
 275 280 285

Thr Ser Phe Gly Trp Phe Val Pro Glu Val Phe Ala Ala Gln Ala Gly
 290 295 300

Leu Asp Asn Gly Arg Lys Ser Glu Pro Trp Phe Ile Glu Asn Lys Val

305 310 315 320

Pro Ala Ser Gln Val Ser Ser Phe Asp Val Arg Pro Asn Gly Ser Gly
 325 330 335

Arg Thr Ala Ile Phe Ala Asn Ala Pro Ser Gly Ala Gln Leu Asn Arg
 340 345 350

Arg Thr Asp Gln Gly Gln Tyr Leu Asp Ala Val Asp Ile Val Ser Gly
 355 360 365

Ser Gly Lys Lys Ser Leu Gly Tyr Ala His Gly Ser Lys Thr Val Asn
 370 375 380

Pro Asn Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Ser Val Met
 385 390 395 400

Pro Gly Ser Leu Gly Val Glu Ser Met Phe Gln Leu Val Glu Ala Ile
 405 410 415

Ala Ala His Glu Asp Leu Ala Gly Lys Ala Arg His Cys Gln Pro His
 420 425 430

Leu Cys Ala Arg Pro Arg Ala Arg Ser Ser Trp Lys Tyr Arg Gly Gln
 435 440 445

Leu Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile Val Ser
 450 455 460

Val Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly Phe Leu
 465 470 475 480

Trp Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg Val Arg
 485 490 495

Ile Ala Ser Gly
 500

<210> 31 <211> 1512 <212> DNA <213> 裂壺菌屬 sp. <220> <221> CDS <222> (1)..(1512)
<223>

<400> 31
 gcc ccg ctc tac ctc tcg cag gac ccg acc agc ggc cag ctc aag aag 48
 Ala Pro Leu Tyr Leu Ser Gln Asp Pro Thr Ser Gly Gln Leu Lys Lys
 1 5 10 15

 cac acc gac gtg gcc tcc ggc cag gcc acc atc gtg cag ccc tgc acg 96
 His Thr Asp Val Ala Ser Gly Gln Ala Thr Ile Val Gln Pro Cys Thr
 20 25 30

 ctc ggc gac ctc ggt gac cgc tcc ttc atg gag acc tac ggc gtc gtc 144
 Leu Gly Asp Leu Gly Asp Arg Ser Phe Met Glu Thr Tyr Gly Val Val
 35 40 45

 gcc ccg ctg tac acg ggc gcc atg gcc aag ggc att gcc tcg gcg gac 192
 Ala Pro Leu Tyr Thr Gly Ala Met Ala Lys Gly Ile Ala Ser Ala Asp
 50 55 60

 ctc gtc atc gcc gcc ggc aag cgc aag atc ctc ggc tcc ttt ggc gcc 240
 Leu Val Ile Ala Ala Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala
 65 70 75 80

 ggc ggc ctc ccc atg cac cac gtg cgc gcc gcc ctc gag aag atc cag 288
 Gly Gly Leu Pro Met His His Val Arg Ala Ala Leu Glu Lys Ile Gln
 85 90 95

 gcc gcc ctg cct cag ggc ccc tac gcc gtc aac ctc atc cac tcg cct 336
 Ala Ala Leu Pro Gln Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro
 100 105 110

 ttt gac agc aac ctc gag aag ggc aac gtc gat ctc ttc ctc gag aag 384
 Phe Asp Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys
 115 120 125

 ggc gtc act gtg gtg gag gcc tcg gca ttc atg acc ctc acc ccg cag 432
 Gly Val Thr Val Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln
 130 135 140

 gtc gtg cgc tac cgc gcc gcc ggc ctc tcg cgc aac gcc gac ggt tcg 480
 Val Val Arg Tyr Arg Ala Ala Gly Leu Ser Arg Asn Ala Asp Gly Ser
 145 150 155 160

 gtc aac atc cgc aac cgc atc atc gcc aag gtc tcg cgc acc gag ctc 528
 Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr Glu Leu
 165 170 175

 gcc gag atg ttc atc cgc ccg gcc ccg gag cac ctc ctc gag aag ctc 576
 Ala Glu Met Phe Ile Arg Pro Ala Pro Glu His Leu Leu Glu Lys Leu
 180 185 190

 atc gcc tcg ggc gag atc acc cag gag cag gcc gag ctc gcg cgc cgc 624
 Ile Ala Ser Gly Glu Ile Thr Gln Glu Gln Ala Glu Leu Ala Arg Arg
 195 200 205

 gtt ccc gtc gcc gac gat atc gct gtc gag gct gac tcg ggc gcc cac 672



Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala Asp Ser Gly Gly His
 210 215 220
 acc gac aac cgc ccc atc cac gtc atc ctc ccg ctc atc atc aac ctc 720
 Thr Asp Asn Arg Pro Ile His Val Ile Leu Pro Leu Ile Ile Asn Leu
 225 230 235 240
 cgc aac cgc ctg cac cgc gag tgc ggc tac ccc gcg cac ctc cgc gtc 768
 Arg Asn Arg Leu His Arg Glu Cys Gly Tyr Pro Ala His Leu Arg Val
 245 250 255
 cgc gtt ggc gcc ggc ggt ggc gtc ggc tgc ccg cag gcc gcc gcc gcc 816
 Arg Val Gly Ala Gly Gly Gly Val Gly Cys Pro Gln Ala Ala Ala Ala
 260 265 270
 ggc ctc acc atg ggc gcc gcc ttc atc gtc acc ggc act gtc aac cag 864
 Ala Leu Thr Met Gly Ala Ala Phe Ile Val Thr Gly Thr Val Asn Gln
 275 280 285
 gtc gcc aag cag tcc ggc acc tgc gac aac gtg cgc aag cag ctc tcg 912
 Val Ala Lys Gln Ser Gly Thr Cys Asp Asn Val Arg Lys Gln Leu Ser
 290 295 300
 cag gcc acc tac tcg gat atc tgc atg gcc ccg gcc gcc gac atg ttc 960
 Gln Ala Thr Tyr Ser Asp Ile Cys Met Ala Pro Ala Ala Asp Met Phe
 305 310 315 320
 gag gag ggc gtc aag ctc cag gtc ctc aag aag gga acc atg ttc ccc 1008
 Glu Glu Gly Val Lys Leu Gln Val Leu Lys Lys Gly Thr Met Phe Pro
 325 330 335
 tcg cgc gcc aac aag ctc tac gag ctc ttt tgc aag tac gac tcc ttc 1056
 Ser Arg Ala Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe
 340 345 350
 gac tcc atg cct cct gcc gag ctc gag cgc atc gag aag cgt atc ttc 1104
 Asp Ser Met Pro Pro Ala Glu Leu Glu Arg Ile Glu Lys Arg Ile Phe
 355 360 365
 aag cgc gca ctc cag gag gtc tgg gag gag acc aag gac ttt tac att 1152
 Lys Arg Ala Leu Gln Glu Val Trp Glu Glu Thr Lys Asp Phe Tyr Ile
 370 375 380
 aac ggt ctc aag aac ccg gag aag atc cag cgc gcc gag cac gac ccc 1200
 Asn Gly Leu Lys Asn Pro Glu Lys Ile Gln Arg Ala Glu His Asp Pro
 385 390 395 400
 aag ctc aag atg tcg ctc tgc ttc cgc tgg tac ctt ggt ctt gcc agc 1248
 Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu Ala Ser
 405 410 415
 cgc tgg gcc aac atg ggc gcc ccg gac cgc gtc atg gac tac cag gtc 1296
 Arg Trp Ala Asn Met Gly Ala Pro Asp Arg Val Met Asp Tyr Gln Val
 420 425 430

tgg tgt ggc ccg gcc att ggc gcc ttc aac gac ttc atc aag ggc acc 1344
 Trp Cys Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe Ile Lys Gly Thr
 435 440 445

tac ctc gac ccc gct gtc tcc aac gag tac ccc tgt gtc gtc cag atc 1392
 Tyr Leu Asp Pro Ala Val Ser Asn Glu Tyr Pro Cys Val Val Gln Ile
 450 455 460

aac ctg caa atc ctc cgt ggt gcc tgc tac ctg cgc cgt ctc aac gcc 1440
 Asn Leu Gln Ile Leu Arg Gly Ala Cys Tyr Leu Arg Arg Leu Asn Ala
 465 470 475 480

ctg cgc aac gac ccg cgc att gac ctc gag acc gag gat gct gcc ttt 1488
 Leu Arg Asn Asp Pro Arg Ile Asp Leu Glu Thr Glu Asp Ala Ala Phe
 485 490 495

gtc tac gag ccc acc aac gcg ctc 1512
 Val Tyr Glu Pro Thr Asn Ala Leu
 500

<210> 32 <211> 504 <212> PRT <213> 裂壺菌屬 sp. <400> 32

Ala Pro Leu Tyr Leu Ser Gln Asp Pro Thr Ser Gly Gln Leu Lys Lys
 1 5 10 15

His Thr Asp Val Ala Ser Gly Gln Ala Thr Ile Val Gln Pro Cys Thr
 20 25 30

Leu Gly Asp Leu Gly Asp Arg Ser Phe Met Glu Thr Tyr Gly Val Val
 35 40 45

Ala Pro Leu Tyr Thr Gly Ala Met Ala Lys Gly Ile Ala Ser Ala Asp
 50 55 60

Leu Val Ile Ala Ala Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala
 65 70 75 80

Gly Gly Leu Pro Met His His Val Arg Ala Ala Leu Glu Lys Ile Gln
 85 90 95

Ala Ala Leu Pro Gln Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro
 100 105 110

Phe Asp Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys
 115 120 125

Gly Val Thr Val Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln
130 135 140

Val Val Arg Tyr Arg Ala Ala Gly Leu Ser Arg Asn Ala Asp Gly Ser
145 150 155 160

Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr Glu Leu
165 170 175

Ala Glu Met Phe Ile Arg Pro Ala Pro Glu His Leu Leu Glu Lys Leu
180 185 190

Ile Ala Ser Gly Glu Ile Thr Gln Glu Gln Ala Glu Leu Ala Arg Arg
195 200 205

Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala Asp Ser Gly Gly His
210 215 220

Thr Asp Asn Arg Pro Ile His Val Ile Leu Pro Leu Ile Ile Asn Leu
225 230 235 240

Arg Asn Arg Leu His Arg Glu Cys Gly Tyr Pro Ala His Leu Arg Val
245 250 255

Arg Val Gly Ala Gly Gly Gly Val Gly Cys Pro Gln Ala Ala Ala Ala
260 265 270

Ala Leu Thr Met Gly Ala Ala Phe Ile Val Thr Gly Thr Val Asn Gln
275 280 285

Val Ala Lys Gln Ser Gly Thr Cys Asp Asn Val Arg Lys Gln Leu Ser
290 295 300

Gln Ala Thr Tyr Ser Asp Ile Cys Met Ala Pro Ala Ala Asp Met Phe
305 310 315 320

Glu Glu Gly Val Lys Leu Gln Val Leu Lys Lys Gly Thr Met Phe Pro
325 330 335

Ser Arg Ala Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe
 340 345 350

Asp Ser Met Pro Pro Ala Glu Leu Glu Arg Ile Glu Lys Arg Ile Phe
 355 360 365

Lys Arg Ala Leu Gln Glu Val Trp Glu Glu Thr Lys Asp Phe Tyr Ile
 370 375 380

Asn Gly Leu Lys Asn Pro Glu Lys Ile Gln Arg Ala Glu His Asp Pro
 385 390 395 400

Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu Ala Ser
 405 410 415

Arg Trp Ala Asn Met Gly Ala Pro Asp Arg Val Met Asp Tyr Gln Val
 420 425 430

Trp Cys Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe Ile Lys Gly Thr
 435 440 445

Tyr Leu Asp Pro Ala Val Ser Asn Glu Tyr Pro Cys Val Val Gln Ile
 450 455 460

Asn Leu Gln Ile Leu Arg Gly Ala Cys Tyr Leu Arg Arg Leu Asn Ala
 465 470 475 480

Leu Arg Asn Asp Pro Arg Ile Asp Leu Glu Thr Glu Asp Ala Ala Phe
 485 490 495

Val Tyr Glu Pro Thr Asn Ala Leu
 500

<210> 33 <211> 9 <212> PRT <213> <220> <223>

<220> <221> MISC_FEATURE <222> (2)..(3) <223> x =

<220> <221> MISC_FEATURE <222> (6)..(6) <223> x = A S

<220> <221> MISC_FEATURE <222> (7)..(8) <223> x =

<400> 33

Trp Xaa Xaa Lys Glu Xaa Xaa Xaa Lys

1 5

<210> 34 <211> 6 <212> PRT <213> 人造序列 <220> <223> 母題

<220> <221> MISC_特徵 <222> (3)..(3) <223> x = I 或 L 或 V

<400> 34

Phe Asn Xaa Ser His Ser

1 5

<210> 35 <211> 5 <212> PRT <213> 人造序列 <220> <223> 母題

<220> <221> MISC_特徵 <222> (1)..(5) <223> x = I 或 L 或 V

<400> 35

Xaa Gly Xaa Asp Xaa

1 5

<210> 36 <211> 4244 <212> DNA <213> 裂壺菌屬 sp. <400> 36
tttctctc tcgagctgt gctgctgctg ctgctgctgc tgcctcctg ctggttcga 60

cgctcgttc atcaagcgt cgtcgtcctg accgatcggg gcgtgcgtgc gtgcgtgagt 120

ctgttgcca ggcagccgca ggctgctgt ctgttggtgt agttttacc tcggggttcg 180

gggtctgct gctcccgt cccgcccgcc gcccccgta tccaccgcc tcgctccgc 240

ccatcgggcc tcgctctc gcgccgacg catcgcgcgc atcgcagca tcatgtgcc 300

acgcacgggg ggcgcgcgc cccgcgtccc ccgccgccgc cgtcgtcgtc tggcgtgcc 360

gtcgcgcgcc tcttcttc cctcgcctcc tcttctccc gagccccct gcttcttc 420

gccccgcag cggcgcgcag gaagcgagga gagcggggag gagagaagaa aagaaaagaa 480

aagaaaagaa aataacagcg ccgtctcgcg cagacgcgcg cggccgcgtg cgagcgcgcg 540

tgatggggt tctcgtggcg cggctgcggc ctggcccggc ctgcctttg aggtgcaggc 600

tttgggagag aagagtggga cgcggagaag ataagatggt gccatggcgc aggacggaga 660

ggttgtgaa acttctcga gcggcacagg cgatggcgag agaccgacag ctgccggcgc 720

ggaggggatg gatactccc gagcctggca tggacgagct gcccgcgcgg atctggctgg 780

ccgcgcgcgc gtgggtccgg aggcgcgagg ttggtttct tcalactga taccataggg 840

tattcatct tcttcccag gaaggaagca agtcacatag agtatcacta gcctaagat 900

ggactctatg ttttagggca cgtcggagca gaaggcgcga gcgattcga tgcgagcga 960
agatacagca cagagacctt gccggcgacg cggatgcagg cgagcacgca cgcaccgcac 1020
gcacggcagc ggtgcacgcg ctctcggca gatgcacggt tctgcgccg gccttiacat 1080
ttttgattt taggtggtg gcctgccact ttgaacalca tccacaagtc aacgcagcat 1140
caagaggcaa gcaagtacat acatccattc gaattcaagt tcaagagacg cagcaacagc 1200
cgccgctccg ctcaagctgc agctagctgg ctgacagggc tcgctggctg tagtggaaaa 1260
ttcattcac tttctgcat ccgcggccag caggcccga cgcacgtct ctcgtttgt 1320
tgttcttcg tgcgtgcgtg cgtgcctcc agctgcctg ctaatctcc gcgcgatcca 1380
acgaccctcg gtcgtgccg caagcgaac ccgacgccga cctggccaat gccgcaagaa 1440
tgctaagcgc gcagcaatgc tgagagtaat ctccagccca ccaagtcatt atcgtgccc 1500
aagtctccat cgcagccaca ttcaggcttt ctctctct cctccctct cttctgccc 1560
ggagagaagg aaagaccgcg cggccgccg tctgcgctg tgacgggctg tccgtttaa 1620
gccctcttag acagttccta ggtgccgggc gccgccgcg ctccgtcga ggcacacgta 1680
ggcggccacg ggtcccccc gcacctcca cacctcttc ccccgcagcc ggaccgcgcg 1740
ccgtctcctt acgcactcg cgggccgcc gcccgcaac ccgagcgcgt gctgtggcg 1800
ccgtctccg gccgctgg aggtctccc cgcgccgcg tactccgggt cctgtgctg 1860
acgtacttaa tattaacagt gggacctcgc acaggacctg acggcagcac agacgtgcc 1920
gcctcgaic gctgggagc caggcgagc atcccggcg ggcgccgac cggggaggct 1980
gccccggcg ctctccggc cggcgccgc atcaggcgga tgacgcaaga gccctgcag 2040
tcgtcgtc cgggagcgc agcgcggcg cagcgtgcc aagctccgc cctctgctg 2100
tggtgcatg cctgcctgcc tgctgcctg cgtgcgtcg tgcgtcgtg ccttcgtcg 2160
tgctgcctt cgtgcgtcg tgcgtgagtg cggcggaaga gggatcatg gaggatcaat 2220
caccgccgc acctcgactt ttgaagaagc cgcgatcga tgcgatcga tgcgatcga 2280
cgcgataccg tgcgaggcta cgaagcgagt ctggccggcc gtcatacaac gcacgtttc 2340
gagaaggagg gctggcgag gcgtgcatg cggcgacct tgcaacgcg gcgtctcgtg 2400
gctggcaag gtgcctggag gatctaaca tcgtctat gatgctatag ctgtgctgat 2460
ccccgtcca ttccaccag tctgtcctg ccgcctgacc tgccttggc tticctcaa 2520

gttctctcc gccgggctt caggaccgag acgagacctg cagctgcagc tagactcgcg 2580
ctcgtcgcg gaggaticgc cggocgccgg gccggacggg actcgcgagg tcacacggcc 2640
gccggcgatc gcgalggtg tctgacgta ctcgtcgtg gcagccgtac gtcagcgagc 2700
ccgcctcctg attgtgatt cgttagtgg ttgtggtg attgtgat taamttt 2760
gttcgtaggc ttggtatag ctaatagtt agttatact ggtcctctc ggtgctgatt 2820
tagctgact tgggtccaca ccactgccc tctactgtga atggatcaat ggacgcacga 2880
cggcccgacg aaagtgcgcg agtgagtaa cctaagcaac gccggtctc agaggggacg 2940
cacgccctcc gtcgcagta gtcagacag gcagaaaagc gtcctagga ccacgcagc 3000
acgcacgcac gcacgcagc ccgcacgcac gtcctctcc tcgctgect atttttag 3060
gttctctcc gcacggcct acctctcgt cctcgcctc gccgcaccag gcggcagcag 3120
cgatacctgc cggtgccc cccgtcacgc gtcagccgc agctcagccc agccgcgagc 3180
tagggttgt tcgtctgaa ttgttgatt tgattgatt tgattgac cgatccgac 3240
cgaictgac tgattgctt tctttgctt tcttccctc ccggcgcgga ccaagcgtc 3300
gtcgcgcgc cgcagctcc ctctctcc cagccctct tctctcccg cctctcgcgc 3360
aagcacgcag ctccgcccc gcatccggc ggtcggcgg tcgacgacc cgctgccgc 3420
tctctctg gcgggctt tctccatcg cactcttc tctccatac gtcctactac 3480
gtacatacat actgccgct tctctctt ccagcgcgc gcggcgga gctgcgacg 3540
tctcggcc cgcgggcc gcgcgcgc ccgccgcc ccgcgtcga ggccctcgc 3600
gccgccccc ctccgtccg ctccgagcc gcgagagggc gcggcgcg cgatggatg 3660
atggatgat ggaaggatg atggattg tgatcgat gcggcgcag gccggagatg 3720
agcaggagc agcgcgcg cgccgcagc ggatcgcag gccctcgtc gcctcgcgc 3780
cgtgccgc cccgcctgc gagcctgc gcgagcgag cgagcgagc agcgggctt 3840
tcttctc gcgcgccgt tggcctcgt tctctgtc ttgcgtagc ggccgcgg 3900
tggaagatg ctattcaat cgaccattc acgcacgc tccggcgc agagaaggcc 3960
gaggaggag agcaagcaa ccaaaagct tcgcgtcgc ggctcggc tcgagcgtc 4020
tcggagag agtcttgcgc gaccaccgg cagcagcagc agcagcagc gcgtctcga 4080
gcacgagc gcacgagc acgagcagc gattcagc aagaggac acacggtgt 4140

cagcgcttag ctgctcgal acagaaagag gcgggtggg cgtaaaaaa aaggagcacg 4200

caagccgcca gccagccagc tagctagcca gctgacctgc caaa 4244

<210> 37 <211> 3886 <212> DNA <213> 裂壺菌屬 sp. <220> <221> misc_特徴 <222>
(2115)..(2115) <223> n = a, c, g, 或 t

<400> 37
gacttgatt gccaaactct ggattgicga ttccgatgaa tcgagctctt tgtgtcgag 60
ctctggcttg ccgagcttc agaaatagac aaaattgccg agttcctgat tgcggggctc 120
tcgattgcca aggtctggtg gattctcga cctcgtatt tcaaaactt ggtcgtctcg 180
tcggattctt tcctgattg tttgtcaag accttgagat tgtgcaaac ctgatcgtt 240
gacaaacctt tgatgcacag cagccttca tcacgctcag ctctgtcat tgattatatt 300
ccccctgaca gccaacacct tgatgcaggg tctcaacctt gattttgga ggccatcalt 360
agcatcacgc cccggcactc acctcaaca ttcgacagcc aacgctttt ttcttcgac 420
taggaltcga gaataaaagc aggtcaccac gaccgtaggc caacgcgaca accatggaaa 480
taaagtgaca acgaacgact tgcaagttta aatgtaaaga gcagcaattg cccgcccaca 540
gacaaatgaa agcaggcgcc gagtcttatt tgaggaggtg ggcctgtggc aatggcgcaa 600
agaaaatcaa ggacaaggag agcaggttac gtaccggtat actggtatcc gtacatggat 660
ggttcttggc aagttgacgg gatgtgtcgc agtgaccgtg gtatgtaacg aaagagccgc 720
aagggcaagg aaagcaagag aatgcagact ttccacagc atggatgggt ccgacagctg 780
ccgcatgatg aaacgtgta ttacacctgg cacgtggtgg cgcacgcgcc cacatatgat 840
cgcggcgcg ggtgtattat acatttccc cctcaggtct actgccaacc ctccatcgt 900
cgctcgtcgc aacgacgcaa gcccttcgca tcgtgcagcc tcttctggt aaggcaagag 960
ctaaacccaa acctaaagca aagaacatt ttacctctct ctctctccca ttggtcgcgt 1020
gcgtccgcc gctcgtcct cctcctgcca gtgtcgcgcc ctaactccc ccttcccctc 1080
ctcccctct cctcccctct cctctgccac cgcccctctc tccgcgctgc gtgcggtgct 1140
gccctggacc aatggcatgc tgctgcacgc tcggcggatg acgcaagccg cttcgcaatt 1200
tccggatcag atctcggcgg ggcgtcgcgc gcggggtcac tgcggacctg ccgcgcccc 1260
tgctcttctt acatccatca tgcctccaa acctccgct cctccacgca cgtacgcacg 1320

cccgtctgca cgcgcgcart gccgtgcga aagcaagcgc ccgccgcgc cccggcgacg 1380
ggaaggcggc cgcggtctcc ctccgcggtt gccctgctcc cgcgcggggc tggcgggca 1440
gcagaaggcg ggtggcggcg gcggcttccg tcttctcag cggcctacgt cggcgcggc 1500
gcgcgagact acgcalgcc ttgcgtcag cgtcgcagg tagccgcgc gggcctagcg 1560
ttccgttg cggcgccct aagccccgg cgcgcacggt attgccgca taccgtacgg 1620
ccaagaccgc cgcagagctc ggccctctc gccccagcca gccagcagcg cagcggagga 1680
agagcgcga ggcgcggcgg gaggcggcc gcggagcagc gcagagcggg gcggagcagc 1740
gcggagcaga acggcagac tcggagcggg caggcgggc agagctttg ggttaagga 1800
ccgggttacc ggcgaagtga gcggctcgg gtagcggctg tggaggggt gagtacgca 1860
gcacgatcg agcgagagag agacgctgcc gcgaalcaag aaggtaggcg cgtcgcgag 1920
cgcgcggcg gagcggagcg agggagagg agaggagag agaggaggg agacgtgcc 1980
gcggcgggc ctggcctggc ctggttggc ttggtcagc cggcctgtc cgagcgtca 2040
gctggagtg ggtgattca ttggattt ctttgttt tgtttctc tcttccgg 2100
aaagtgtgg ccgncgggt tctttgtt tgattctc aaaagtitt ggttgggt 2160
ctctcttg gctctctc agcggctcc gccacgcc cggcctctc tctctctc 2220
tctctctc tctcgtcg talactag tactttga tactacata catcccgc 2280
gccgtcccg cgaggttg ctccctgg agcaatcga tccgatcga tccgatcga 2340
cgcgacgca cgcgagtc tggtcgcg tgtgctgtg gcttcttc ttacttct 2400
tcgagctc ccgtttct cttctctc cacccacca ccaacgaaag aagatcgcc 2460
ccggcacgcc gctgagaagg gctggcggc atgacggc gcgcgccgc tcccgtt 2520
gcctcgtc ctgctctc tctctctc gctctctc ctgctctc tctctctc 2580
gcgcgagc tttccagc gccggcgt ctggcctc ccgttccag tccggtga 2640
gagatcgaat gagagataa ctggatgat tcatcagg ataatgaa gatggtga 2700
tgcctttc ctttccag tccacagc gaagcaggc cgcgtgaac tcccctac 2760
cgcatactc tgcctgat cgcctgat gcacgacgc ctgccggga gccacagc 2820
ggcgacagg cggccagca gccaggcagc cagccaggca gccaccagag ggccagag 2880
cgcgcctc gcacgcgc cagtgcgc alcgtcga gtgcagact gattcccc 2940

cgcgatctc cgcgagcccg aaacgaagag cgccgtacgg gcccalccta gcgtcgcctc 3000
gcaccgcac gcacgcac gcgttccta gagagtagta ctgcacgaag gcaccattc 3060
cgcgctctc ttcggcgcga tcgaggcccc cggcgcgcg acgatcgcgg cggccgcggc 3120
gctggcggcg gccctggcgc tcgcgctggc ggccgcgcg ggcgtctggc cctggcgcgc 3180
gcgggcggcg caggaggagc ggcagcggct gctcggcc agagaagagc gcgcccggcc 3240
cggggaggga cggggaggag aaggagaagg cgcgcaaggc ggccccgaaa gagaagacc 3300
tggacttgaa cgcgaagaag aagaagaagg agaagaagt gaagaagaag aagaagaag 3360
agaggaagt gaagaagac aggagcaggc gcgtccaag gcgcgtctc ttccggaggc 3420
gcgtccagc tgcggcgcg gggcgggctg cggggcggc gcgggcgcg gtgcgggcag 3480
aggggacgc cgcgcggagg cggaggggc cgcgcggag ccctgctgc tgcgggcgc 3540
ccggcccga ggtgtggcgc gcgcgacgac ggaggcgcg acgccagcg ccgacgacg 3600
aaggccggc gcgtcggcg gcggaaggc ccgcgcggag caggggcgg agcaggacaa 3660
ggcgcaggag caggagcagg gccgggagc ggagcggag cgggcgcgg agcccaggc 3720
agaaccaat cgagatccag agcgcgcaga ggccggccg gagcccagc ccgcgccga 3780
gatcactagt accgtcggc aatcacagca gcagcagcag cagcagcag agcagcagca 3840
gcagcagcag ccacgagagg gagataaaga aaaagcggca gagacg 3886

公告本 發明專利說明書

雙面影印

(本說明書格式、順序，請勿任意更動，※記號部分請勿填寫；)

※申請案號： 99125328

※申請日： 91.4.15 ※IPC 分類

原申請案號：由第 091107633 號申請案分割。

C12N	1/20.	2006.01
C12N	5/10.	2006.01
C12N	15/00	2006.01
C12N	15/63	2006.01
C12N	15/70	2006.01
C12P	19/02.	2006.01
C07H	21/00	2006.01

分割案

一、發明名稱：(中文/英文)

多不飽和脂肪酸(PUFA)聚乙醯合成酶系統及其用途(二)/PUFA
POLYKETIDE SYNTHASE SYSTEMS AND USES THEREOF

二、中文發明摘要：

本發明係一般性地關於一種分離自或起源自非細菌生物的多不飽和脂肪酸(PUFA)聚乙醯合成酶(PKS)系統；關於其同源物；關於編碼此 PUFA PKS 系統的生物活性區段之分離的核酸分子及重組型核酸分子；關於包含 PUFA PKS 系統之基因改性的生物；關於製得及使用此系統來製造有興趣的生物活性分子之方法；及關於用來鑑別具有此 PUFA PKS 系統的新穎細菌及非細菌微生物之新穎的方法。

三、英文發明摘要：

The invention generally relates to polyunsaturated fatty acid (PUFA) polyketide synthase (PKS) systems isolated from or derived from non-bacterial organisms, to homologues thereof, to isolated nucleic acid molecules and recombinant nucleic acid molecules encoding biologically active domains of such a PUFA PKS system, to genetically modified organisms comprising PUFA PKS systems, to methods of making and using such systems for the production of bioactive molecules of interest, and to novel methods for identifying new bacterial and non-bacterial microorganisms having such a PUFA PKS system.

第 1 圖

307 kDa



KS MAT 9X ACP KR

Orf A

225 kDa



KS CLF AT ER

Orf B

166 kDa

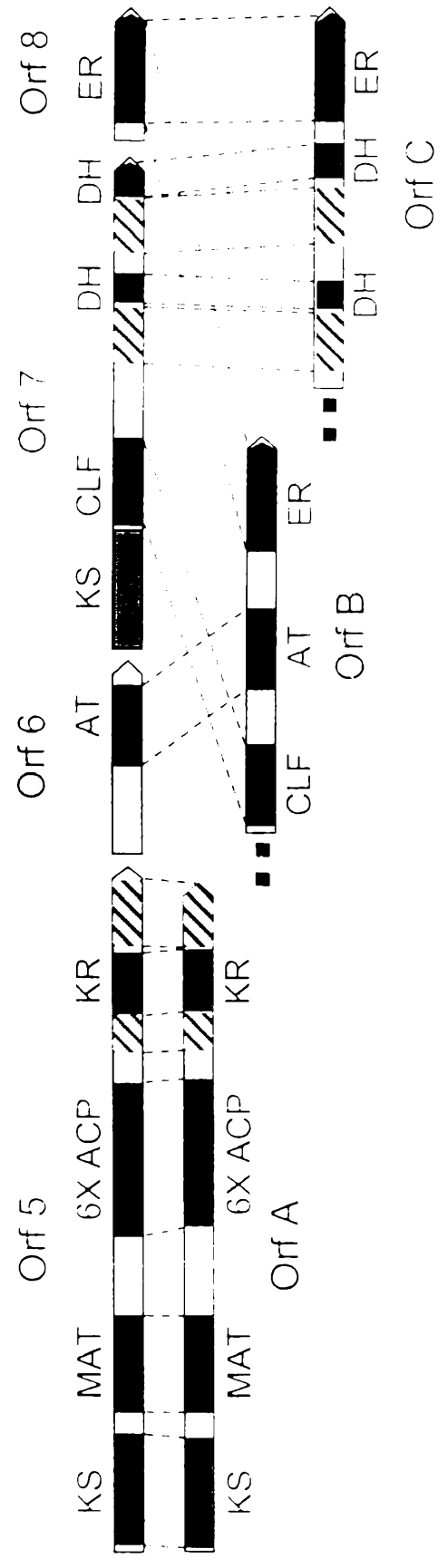


DH DH ER

Orf C

第 2 圖

PKS Orfs 區段的比較：
裂壺菌屬 vs *Shewanella*

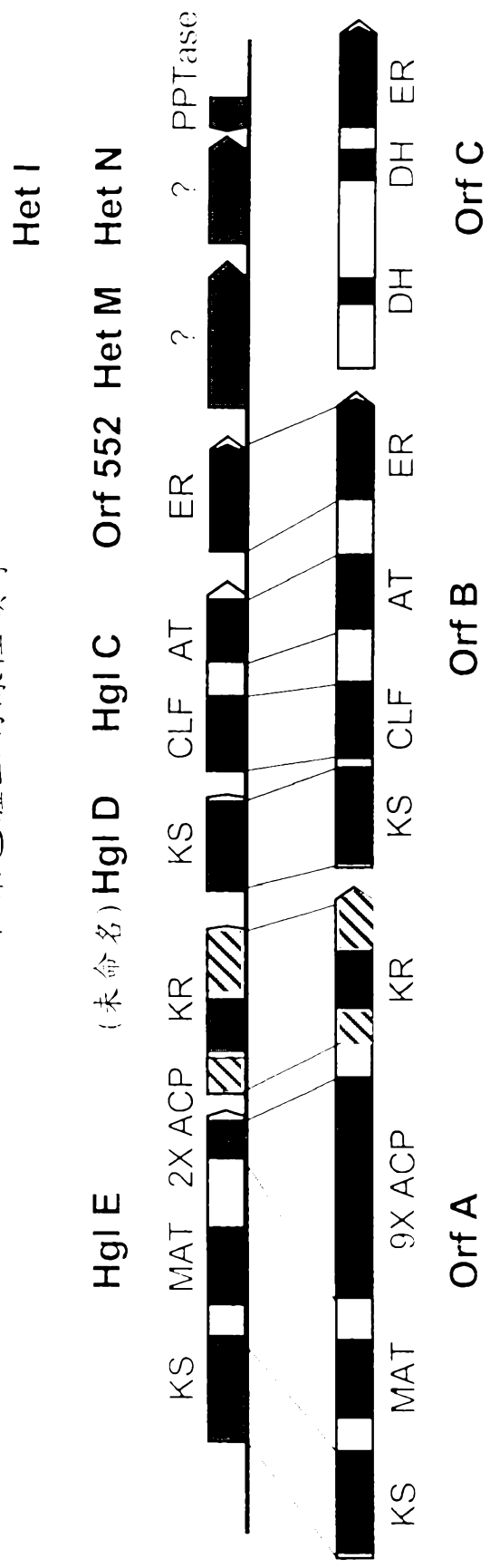


氨基酸序列同源性



第 3 圖

念球藻(7120)vs 裂壺菌屬
在 Hgl/Het 團簇中的念球藻 ORFs
在染色體上的線性順序



裂壺菌屬 PUFA ORFs(分離的 Orfs)

相同度%

= 20-35%
 = 35-50%
 > 50%

四、指定代表圖：

(一)本案指定代表圖為：第 () 圖。(無)

(二)本代表圖之元件符號簡單說明：

五、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

(無)

(02)年9月11日修(訂)正本
P1-8

公告本

七、申請專利範圍：

1. 一種分離的核酸分子，其選自於由下列所組成之組群：
 - a. 一核酸分子，其包含一核酸序列，該核酸序列編碼一胺基酸序列，該胺基酸序列與序列辨識編號：4 至少有95%相同性，其中該胺基酸序列具有 β -酮基醯基-ACP合成酶(KS)活性、鏈長因子(CLF)活性、轉醯酶(AT)活性以及烯醯ACP-還原酶(ER)活性；
 - b. 一核酸分子，其包含一核酸序列，該核酸序列編碼一胺基酸序列，該胺基酸序列與序列辨識編號：20 至少有95%相同性，且該胺基酸序列具有 β -酮基醯基-ACP合成酶(KS)活性；以及
 - c. 一核酸分子，其基本上由一核酸序列所構成，該核酸序列編碼一與下列胺基酸序列具有至少95%相同性之胺基酸序列：
 - (i) 序列辨識編號：22，且具有CLF活性；
 - (ii) 序列辨識編號：24，且具有AT活性；或
 - (iii) 序列辨識編號：26，且具有ER活性。
2. 一種分離的核酸分子，其選自於由下列所組成之組群：
 - a. 一核酸分子，其包含一核酸序列，該核酸序列編碼一胺基酸序列，該胺基酸序列與序列辨識編號：4 至少有96%相同性，其中該胺基酸序列具有 β -酮基醯基-ACP合成酶(KS)活性、鏈長因子(CLF)活性、轉醯酶(AT)活性以及烯醯ACP-還原酶(ER)活性；
 - b. 一核酸分子，其包含一核酸序列，該核酸序列編碼

一胺基酸序列，該胺基酸序列與序列辨識編號：20
至少有96%相同性，且該胺基酸序列具有 β -酮基醯基
-ACP合成酶(KS)活性；以及

5 c.一核酸分子，其基本上由一核酸序列所構成，該核酸
序列編碼一與下列胺基酸序列具有至少96%相同性之
胺基酸序列：

(i)序列辨識編號：22，且具有CLF活性；

(ii)序列辨識編號：24，且具有AT活性；或

(iii)序列辨識編號：26，且具有ER活性。

10 3. 如申請專利範圍第1項之分離的核酸分子，其中該核酸
分子選自於由下列所組成之組群：

15 a.一核酸分子，其包含一核酸序列，該核酸序列編碼
一胺基酸序列，該胺基酸序列與序列辨識編號：4至
至少有97%相同性，其中該胺基酸序列具有 β -酮基醯基
-ACP合成酶(KS)活性、鏈長因子(CLF)活性、轉醯酶
(AT)活性以及烯醯ACP-還原酶(ER)活性；

20 b.一核酸分子，其包含一核酸序列，該核酸序列編碼
一胺基酸序列，該胺基酸序列與序列辨識編號：20
至少有97%相同性，且該胺基酸序列具有 β -酮基醯基
-ACP合成酶(KS)活性；以及

c.一核酸分子，其基本上由一核酸序列所構成，該核
酸序列編碼一與下列胺基酸序列具有至少97%相同
性之胺基酸序列：

(i)序列辨識編號：22，且具有CLF活性；

(ii) 序列辨識編號：24，且具有 AT 活性；或

(iii) 序列辨識編號：26，且具有 ER 活性。

4. 如申請專利範圍第 1 項之分離的核酸分子，其中該核酸分子係選自於由下列所組成之組群：

5 a. 一核酸分子，其包含一核酸序列，該核酸序列編碼一胺基酸序列，該胺基酸序列與序列辨識編號：4 至少有 98% 相同性，其中該胺基酸序列具有 β -酮基醯基-ACP 合成酶(KS)活性、鏈長因子(CLF)活性、轉醯酶(AT)活性以及烯醯 ACP-還原酶(ER)活性；

10 b. 一核酸分子，其包含一核酸序列，該核酸序列編碼一胺基酸序列，該胺基酸序列與序列辨識編號：20 至少有 98% 相同性，且該胺基酸序列具有 β -酮基醯基-ACP 合成酶(KS)活性；以及

15 c. 一核酸分子，其基本上由一核酸序列所構成，該核酸序列編碼一與下列胺基酸序列具有至少 98% 相同性之胺基酸序列：

(i) 序列辨識編號：22，且具有 CLF 活性；

(ii) 序列辨識編號：24，且具有 AT 活性；或

(iii) 序列辨識編號：26，且具有 ER 活性。

20 5. 如申請專利範圍第 1 項之分離的核酸分子，其中該核酸分子係選自於由下列所組成之組群：

a. 一核酸分子，其包含一核酸序列，該核酸序列編碼一胺基酸序列，該胺基酸序列與序列辨識編號：4 至少有 99% 相同性，其中該胺基酸序列具有 β -酮基醯基

-ACP合成酶(KS)活性、鏈長因子(CLF)活性、轉醯酶(AT)活性以及烯醯ACP-還原酶(ER)活性；

5 b. 一核酸分子，其包含一核酸序列，該核酸序列編碼一胺基酸序列，該胺基酸序列與序列辨識編號：20至少有99%相同性，且該胺基酸序列具有 β -酮基醯基-ACP合成酶(KS)活性；以及

c. 一核酸分子，其基本上由一核酸序列所構成，該核酸序列編碼一與下列胺基酸序列具有至少99%相同性之胺基酸序列：

10 (i) 序列辨識編號：22，且具有CLF活性；

(ii) 序列辨識編號：24，且具有AT活性；或

(iii) 序列辨識編號：26，且具有ER活性。

15 6. 如申請專利範圍第1項之分離的核酸分子，其中該核酸分子由一選自於由下列所組成之組群的核酸序列所構成：序列辨識編號：3、序列辨識編號：19、序列辨識編號：21、序列辨識編號：23、序列辨識編號：25。

20 7. 如申請專利範圍第1項之分離的核酸分子，其中該核酸分子包含一核酸序列，該核酸序列編碼一胺基酸序列，該胺基酸序列與序列辨識編號：4至少有95%相同性，其中該胺基酸序列具有 β -酮基醯基-ACP合成酶(KS)活性、鏈長因子(CLF)活性、轉醯酶(AT)活性以及烯醯ACP-還原酶(ER)活性。

8. 如申請專利範圍第1項之分離的核酸分子，其中該核酸分子包含一核酸序列，該核酸序列編碼一胺基酸序

列，該胺基酸序列與序列辨識編號：4 至少有 98% 相同性，其中該胺基酸序列具有 β -酮基醯基-ACP 合成酶 (KS) 活性、鏈長因子 (CLF) 活性、轉醯酶 (AT) 活性以及烯醯 ACP-還原酶 (ER) 活性。

- 5 9. 如申請專利範圍第 1 項之分離的核酸分子，其中該核酸分子包含一編碼序列辨識編號：4 之胺基酸序列的核酸序列。
- 10 10. 如申請專利範圍第 1 項之分離的核酸分子，其中該核酸分子包含一核酸序列，該核酸序列編碼一胺基酸序列，該胺基酸序列與序列辨識編號：20 至少有 98% 相同性，且具有 β -酮基醯基-ACP 合成酶 (KS) 活性。
11. 如申請專利範圍第 1 項之分離的核酸分子，其中該核酸分子包含一編碼序列辨識編號：20 之胺基酸序列的核酸序列。
- 15 12. 如申請專利範圍第 1 項之分離的核酸分子，其中該核酸分子係由一核酸序列所構成，該核酸序列編碼一胺基酸序列，該胺基酸序列與序列辨識編號：22 至少有 98% 相同性，且具有 CLF 活性。
- 20 13. 如申請專利範圍第 1 項之分離的核酸分子，其中該核酸分子係由一編碼序列辨識編號：22 之胺基酸序列的核酸序列所構成。
14. 如申請專利範圍第 1 項之分離的核酸分子，其中該核酸分子係由一核酸序列所構成，該核酸序列編碼一胺基酸序列，該胺基酸序列與序列辨識編號：24 至少有 98%

相同性，且具有AT活性。

15. 如申請專利範圍第1項之分離的核酸分子，其中該核酸分子係由一編碼序列辨識編號：24之胺基酸序列的核酸序列所構成。
- 5 16. 如申請專利範圍第1項之分離的核酸分子，其中該核酸分子係由一核酸序列所構成，該核酸序列編碼一胺基酸序列，該胺基酸序列與序列辨識編號：26至少有98%相同性，且具有ER活性。
17. 如申請專利範圍第1項之分離的核酸分子，其中該核酸
10 分子係由一編碼序列辨識編號：26之胺基酸序列的核酸序列所構成。
18. 如申請專利範圍第1項之分離的核酸分子，其中該核酸分子係為一裂壺菌(*Schizochytrium*)核酸分子。
19. 如申請專利範圍第1項之分離的核酸分子，其中該核酸
15 分子係為一裂壺菌ATCC20888核酸分子。
20. 一種重組型核酸分子，其包含如申請專利範圍第1項至第19項中任一項之核酸分子，該核酸分子被可操作地連結到至少一個轉錄控制序列。
21. 一種重組型細胞，其以如申請專利範圍第20項之重組
20 型核酸分子所轉染。
22. 如申請專利範圍第21項之重組型細胞，其中該重組型細胞係為一微生物。
23. 如申請專利範圍第21項之重組型細胞，其中該重組型細胞係為一植物細胞。

24. 一種分離的核酸分子，其基本上係由一核酸序列構成，該核酸序列係完全互補於如申請專利範圍第1至19項中任一項之核酸分子。
25. 一種基因改性的微藻，其中該微藻已被一包含如申請專利範圍第20項之核酸分子的重組型核酸分子轉型，且表現該包含該如申請專利範圍第20項之核酸分子的重組型核酸分子。
26. 如申請專利範圍第25項之基因改性微藻，其中該微藻已被一重組型核酸分子轉型，該重組型核酸分子包含一核酸序列，該核酸序列編碼一蛋白質，該蛋白質具有 β -酮基醯基-ACP合成酶(KS)活性、鏈長因子(CLF)活性、轉醯酶(AT)活性以及烯醯ACP-還原酶(ER)活性，其中該蛋白質包含一與序列辨識編號：4至少有95%相同性之胺基酸序列。
27. 如申請專利範圍第26項之基因改性微藻，其中該核酸序列係為一裂壺菌(*Schizochytrium*)核酸序列。
28. 如申請專利範圍第26項之基因改性微藻，其中該核酸序列係為一裂壺菌ATCC20888核酸序列。
29. 如申請專利範圍第25項之基因改性微藻，其中該微藻可內生地表現出一PUFA PKS系統。
30. 如申請專利範圍第25項至第29項任一項之基因改性微藻，其中該微藻已被進一步以一重組型核酸分子轉型，該重組型核酸分子係編碼一磷酸泛醯巯基乙胺轉位酶。
31. 如申請專利範圍第25項之基因改性微藻，其中該微藻

已被進一步以一核酸序列轉型，該核酸序列編碼源自於型式I迭代(iterative) (類真菌) PKS系統的至少一個生物活性區段。

32. 如申請專利範圍第25項之基因改性微藻，其中該微藻
5 已被進一步以一核酸序列轉型，該核酸序列編碼源自於型式II PKS系統的至少一個生物活性蛋白質。
33. 如申請專利範圍第25項之基因改性微藻，其中該微藻已被進一步以一核酸序列轉型，該核酸序列編碼源自於型式I組合式PKS系統的至少一個生物活性區段。
- 10 34. 如申請專利範圍第25項之基因改性微藻，其中該微藻係為破囊壺菌目(*Thraustochytriales*)。
35. 如申請專利範圍第34項之基因改性微藻，其中該微藻係為裂壺菌 (*Schizochytrium*) 或破囊壺菌 (*Thraustochytrium*)。
- 15 36. 一種生產重組型微生物的方法，其包含：
- a). 以至少一如申請專利範圍第1至19項任一項之核酸分子基因改性微生物細胞；以及
 - b). 在允許該至少一核酸分子表現的條件下培養該細胞。