I875705

(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 I875705 B**

(45)公告日： 中華民國 114 (2025) 年 03 月 11 日

(21)申請案號：108128641 (22)申請日： 中華民國 108 (2019) 年 08 月 12 日

(51)Int. Cl.： *A61K48/00 (2006.01)* *C12N15/12 (2006.01)*
C07K14/755 (2006.01) *C12N15/86 (2006.01)*
C12N7/00 (2006.01) *A61K9/00 (2006.01)*
A61P3/00 (2006.01) *A61P1/16 (2006.01)*
A61P7/04 (2006.01)

(30)優先權：2018/08/09 美國 62/716,826

(71)申請人：美商百歐維拉提夫治療公司 (美國) BIOVERATIV THERAPEUTICS INC. (US)
美國

(72)發明人：劉 彤瑤 LIU, TONGYAO (US)；賽瑞吉 阿列希 SEREGIN, ALEXEY (RU)；彼得斯 羅伯特 PETERS, ROBERT T. (US)；劉佳云 LIU, JIAYUN (CN)；查卡斯 飛利浦 ZAKAS, PHILIP (US)；查格 道格拉斯 DRAGER, DOUGLAS (US)；帕塔羅約 懷特 蘇珊娜 PATARROYO-WHITE, SUSANNAH (US)

(74)代理人：何愛文；王仁君

(56)參考文獻：

TW	201629225A	CN	107531774A
US	2016/0102297A1	WO	2006/033756A2
WO	2008/016391A2		

審查人員：吳思瑩

申請專利範圍項數：22 項 圖式數：8 共 570 頁

(54)名稱

核酸分子及其用於非病毒基因療法的用途

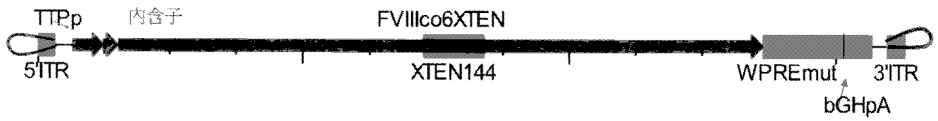
(57)摘要

本公開文本提供了核酸分子，其包含第一反向末端重複體(ITR)、第二 ITR 和編碼靶序列的基因盒。在一些實施例中，靶序列編碼 miRNA 和/或治療性蛋白質。在某些實施例中，治療性蛋白質包含凝血因子、生長因子、激素、細胞激素、抗體、其片段及其組合。在一些實施例中，第一 ITR 和/或第二 ITR 是非腺相關病毒(AAV)的 ITR。本公開文本還提供了治療個體的肝臟代謝障礙的方法，其包括將核酸分子或由其編碼的多肽投予個體。

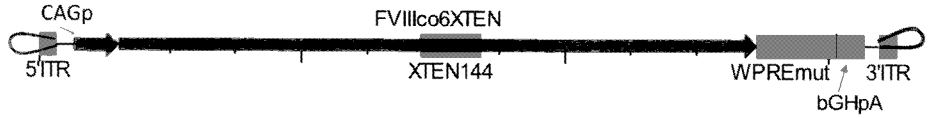
The present disclosure provides nucleic acid molecules comprising a first inverted terminal repeat (ITR), a second ITR, and a genetic cassette encoding a target sequence. In some embodiments, the target sequence encodes a miRNA and/or a therapeutic protein. In certain embodiments, the therapeutic protein comprises a clotting factor, a growth factor, a hormone, a cytokine, an antibody, a fragment thereof, and a combination thereof. In some embodiments, the first ITR and/or the second ITR is an ITR of a non-adeno-associated virus (AAV). The present disclosure also provides methods of treating a metabolic disorder of the liver in a subject comprising administering to the subject the nucleic acid molecule or a polypeptide encoded thereby.

指定代表圖：

【圖1A】


TTRp
內含子
FVIIIco6XTEN
5'ITR
XTEN144
WPREmut
3'ITR
bGHpA

【圖1B】


CAGp
FVIIIco6XTEN
5'ITR
XTEN144
WPREmut
3'ITR
bGHpA

I875705

【發明摘要】

【中文發明名稱】

核酸分子及其用於非病毒基因療法的用途

【英文發明名稱】

NUCLEIC ACID MOLECULES AND USES THEREOF FOR NON-VIRAL GENE THERAPY

【中文】

本公開文本提供了核酸分子，其包含第一反向末端重複體（ITR）、第二ITR和編碼靶序列的基因盒。在一些實施例中，靶序列編碼miRNA和/或治療性蛋白質。在某些實施例中，治療性蛋白質包含凝血因子、生長因子、激素、細胞激素、抗體、其片段及其組合。在一些實施例中，第一ITR和/或第二ITR是非腺相關病毒（AAV）的ITR。本公開文本還提供了治療個體的肝臟代謝障礙的方法，其包括將核酸分子或由其編碼的多肽投予個體。

【英文】

The present disclosure provides nucleic acid molecules comprising a first inverted terminal repeat (ITR), a second ITR, and a genetic cassette encoding a target sequence. In some embodiments, the target sequence encodes a miRNA and/or a therapeutic protein. In certain embodiments, the therapeutic protein comprises a clotting factor, a growth factor, a hormone, a cytokine, an antibody, a fragment thereof, and a combination thereof. In some embodiments, the first ITR and/or the second ITR is an ITR of a non-adeno-associated virus (AAV). The present disclosure also provides methods of treating a metabolic disorder of the liver in a subject comprising administering to the subject the nucleic acid molecule or a polypeptide encoded thereby.

【指定代表圖】

圖1A、圖1B

【代表圖之符號簡單說明】

無

【特徵化學式】

無

【發明說明書】

【中文發明名稱】

核酸分子及其用於非病毒基因療法的用途

【英文發明名稱】

NUCLEIC ACID MOLECULES AND USES THEREOF FOR NON-VIRAL GENE THERAPY

【相關申請案】

【0001】 本申請案主張2018年8月9日提交的美國臨時專利申請序列號62/716,826的優先權，將其全部公開內容通過引用特此併入本文。

【參照電子提交的序列表】

【0002】 以ASCII文字檔電子提交的序列表（名稱：SA9-465TW_SL_ST25.txt；大小：460,648字節；創建日期：2019年8月8日）的內容通過引用以其整體併入本文。

【先前技術】

【0003】 基因療法提供了治療多種疾病的持久手段的潛力。過去，許多基因療法治療通常依賴於使用病毒。有多種可以選擇用於這個目的的病毒劑，每種病毒劑具有不同特性，這使得它們更適於或更不適於基因療法。Zhou等人, Adv Drug Deliv Rev. 106(Pt A):3-26, 2016。然而，一些病毒載體的不期望特性（包括其免疫原性譜或其引起癌症的傾向）已經導致臨床安全性問題，並且直到最近，限制其在某些應用（例如，疫苗和溶瘤策略）中的臨床使用。Cotter等人, Front Biosci. 10:1098-105 (2005)。

【0004】 腺相關病毒（AAV）是一種最常研究的基因療法載體。AAV是包圍並保護大約4.8千鹼基（kb）的小的單鏈DNA基因組的蛋白質外殼。Naso等人, BioDrugs, 31(4): 317-334, 2017。AAV屬於細小病毒家族並且依賴於與其他病毒（主要是腺病毒）共感染以複製。*同上*。其單鏈基因組含

有三個基因 Rep（複製）、Cap（衣殼）和 aap（組裝）。同上。這些編碼序列側接有基因組複製和包裝所需的反向末端重複體（ITR）。*同上*。這兩個順式作用 AAV ITR 的長度為大約 145 個核苷酸，具有中斷的回文序列，所述回文序列可以折疊成在 DNA 複製的起始期間起引子作用的 T 形髮夾結構。

【0005】 然而，使用常規 AAV 作為基因遞送載體具有某些缺點。一個主要缺點與 AAV 的約 4.5 kb 異源 DNA 的有限病毒包裝能力有關。（Dong 等人, Hum Gene Ther. 7(17): 2101-12, 1996）。另外，投予 AAV 載體可以誘導人的免疫應答。雖然已經顯示 AAV 的免疫原性低於一些其他病毒（即腺病毒），但是衣殼蛋白可以觸發人免疫系統的多種組分。參見 Naso 等人, 2017。AAV 是人群中的常見病毒，並且大多數人已經暴露於 AAV，因此大多數人已經產生針對其先前所暴露的特定變異體的免疫應答。這種預先存在的適應性應答可以包括中和抗體（NAb）和 T 細胞，它們可以降低隨後用 AAV 再感染的臨床功效和/或已經被轉導的細胞的消除，這可能使具有預先存在的抗 AAV 免疫力的患者沒有資格接受基於 AVV 的基因療法治療。此外，有證據表明，AAV ITR 的 T 形髮夾環易受結合 AAV ITR 的 T 形髮夾結構的宿主細胞蛋白質/蛋白質複合物的抑制。參見例如，Zhou 等人, *Scientific Reports* 7:5432（2017 年 7 月 14 日）。

【0006】 因此，本領域需要在體外和體內環境中有效且持久地表現靶序列（例如，治療性蛋白質和/或 miRNA），同時避免現有 AAV 載體技術的一些無意的結果和限制。

【發明內容】

【0007】 在某些態樣中，提供了核酸分子，所述核酸分子包含側接于基因盒的第一反向末端重複體（ITR）和第二 ITR，所述基因盒包含異源多核苷酸序列，其中第一 ITR 和/或第二 ITR 包含與 SEQ ID NO: 180、181、183、184、185、186、187 或 188 中所述的核苷酸序列至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同的核苷酸序列或其功能衍生物。

【0008】 在某些例示性實施例中，第一 ITR 包含 SEQ ID NO: 180 中所述的核苷酸序列，並且第二 ITR 包含 SEQ ID NO: 181 中所述的核苷酸序列。在某些例示性實施例中，第一 ITR 包含 SEQ ID NO: 183 中所述的核苷酸序列，並且第二 ITR 包含 SEQ ID NO: 184 中所述的核苷酸序列。在某些例示性實施例中，第一 ITR 包含 SEQ ID NO: 185 中所述的核苷酸序列，並且第二 ITR 包含 SEQ ID NO: 186 中所述的核苷酸序列。在某些例示性實施例中，第一 ITR 包含 SEQ ID NO: 187 中所述的核苷酸序列，並且第二 ITR 包含 SEQ ID NO: 188 中所述的核苷酸序列。

【0009】 在某些例示性實施例中，第一 ITR 和/或第二 ITR 由 SEQ ID NO: 180、181、183、184、185、186、187 或 188 中所述的核苷酸序列組成。在某些例示性實施例中，第一 ITR 和第二 ITR 是彼此的反向補體。

【0010】 在某些例示性實施例中，核酸分子還包含啟動子。在某些例示性實施例中，啟動子是組織特異性啟動子。在某些例示性實施例中，啟動子驅動異源多核苷酸序列在選自由下列組成之群組的器官中的表現：肌肉、中樞神經系統（CNS）、眼睛、肝臟、心臟、腎臟、胰腺、肺、皮膚、膀胱、泌尿道或其任何組合。在某些例示性實施例中，啟動子驅動異源多核苷酸序列在以下細胞中的表現：肝細胞、內皮細胞、心肌細胞、骨骼肌細胞、竇狀細胞、傳入神經元、傳出神經元、中間神經元、神經膠質細胞、星形膠質細胞、少突膠質細胞、小神經膠質細胞、室管膜細胞、肺上皮細胞、許旺細胞（Schwann cell）、衛星細胞、感光細胞、視網膜神經節細胞或其任何組合。在某些例示性實施例中，啟動子定位於異源多核苷酸序列的 5'。在某些例示性實施例中，啟動子選自小鼠甲狀腺素啟動子（mTTR）、內源人類 VIII 因子啟動子（F8）、人類 α-1-抗胰蛋白酶啟動子（hAAT）、人類白蛋白最小啟動子、小鼠白蛋白啟動子、三重四脯胺酸（TTP）啟動子、CASI 啟動子、CAG 啟動子、巨細胞病毒（CMV）啟動子、α1-抗胰蛋白酶（AAT）、肌肉肌酸激酶（MCK）、肌球蛋白重鏈 α（αMHC）、肌紅蛋白（MB）、結蛋白（DES）、SPc5-12、2R5Sc5-12、dMCK、tMCK 和磷酸甘油酸激酶（PGK）啟動子。

【0011】 在某些例示性實施例中，異源多核苷酸序列還包含內含子序列。在某些例示性實施例中，內含子序列定位於異源多核苷酸序列的 5'。在某些例示性實施例中，內含子序列定位於啟動子的 3'。在某些例示性實施例中，內含子序列包含合成內含子序列。在某些例示性實施例中，內含子序列包含 SEQ ID NO: 115 或 192。

【0012】 在某些例示性實施例中，基因盒還包含轉錄後調控元件。在某些例示性實施例中，轉錄後調控元件定位於異源多核苷酸序列的 3'。在某些例示性實施例中，轉錄後調控元件包含突變的土撥鼠肝炎病毒轉錄後調控元件（WPRE）、微小 RNA 結合位點、DNA 核靶向序列或其任何組合。在某些例示性實施例中，微小 RNA 結合位點包含與 miR142-3p 的結合位點。

【0013】 在某些例示性實施例中，基因盒還包含 3'UTR 多聚(A)尾序列。在某些例示性實施例中，3'UTR 多聚(A)尾序列選自 bGH 多聚(A)、肌動蛋白多聚(A)、血紅蛋白多聚(A)及其任何組合。在某些例示性實施例中，3'UTR 多聚(A)尾序列包含 bGH 多聚(A)。

【0014】 在某些例示性實施例中，基因盒還包含強化子序列。在某些例示性實施例中，強化子序列定位於第一 ITR 與第二 ITR 之間。

【0015】 在某些例示性實施例中，核酸分子從 5'至 3'包含：第一 ITR、基因盒和第二 ITR；其中基因盒包含組織特異性啟動子序列、內含子序列、異源多核苷酸序列、轉錄後調控元件和 3'UTR 多聚(A)尾序列。在某些例示性實施例中，基因盒從 5'至 3'包含：組織特異性啟動子序列、內含子序列、異源多核苷酸序列、轉錄後調控元件和 3'UTR 多聚(A)尾序列。在某些例示性實施例中，組織特異性啟動子序列包含 TTT 啟動子；內含子是合成內含子；轉錄後調控元件包含 WPRE；並且 3'UTR 多聚(A)尾序列包含 bGHpA。

【0016】 在某些例示性實施例中，基因盒包含單鏈核酸。在某些例示性實施例中，基因盒包含雙鏈核酸。

【0017】 在某些例示性實施例中，異源多核苷酸序列編碼凝血因子、生長因子、激素、細胞激素、抗體、其片段或其任何組合。

【0018】 在某些例示性實施例中，異源多核苷酸序列編碼選自由下列組成之群組的生長因子：腎上腺髓質素（AM）、血管生成素（Ang）、自

分泌運動因子、骨形態發生蛋白（BMP）（例如，BMP2、BMP4、BMP5、BMP7）、睫狀神經營養因子家族成員（例如，睫狀神經營養因子（CNTF）、白血病抑制因子（LIF）、白介素-6（IL-6））、集落刺激因子（例如，巨噬細胞集落刺激因子（m-CSF）、粒細胞集落刺激因子（G-CSF）、粒細胞巨噬細胞集落刺激因子（GM-CSF））、表皮生長因子（EGF）、肝配蛋白(ephrin)（例如，肝配蛋白 A1、肝配蛋白 A2、肝配蛋白 A3、肝配蛋白 A4、肝配蛋白 A5、肝配蛋白 B1、肝配蛋白 B2、肝配蛋白 B3）、促紅細胞生成素（EPO）、成纖維細胞生長因子（FGF）（例如，FGF1、FGF2、FGF3、FGF4、FGF5、FGF6、FGF7、FGF8、FGF9、FGF10、FGF11、FGF12、FGF13、FGF14、FGF15、FGF16、FGF17、FGF18、FGF19、FGF20、FGF21、FGF22、FGF23）、胎牛促體素（FBS）、GDNF 家族成員（例如，神經膠質細胞株衍生的神經營養因子（GDNF）、神經秩蛋白(neurturin)、普塞芬蛋白（persephin）、阿特敏蛋白（artemin））、生長分化因子-9（GDF9）、肝細胞生長因子（HGF）、肝腫瘤衍生的生長因子（HDGF）、胰島素、類胰島素生長因子（例如，類胰島素生長因子-1（IGF-1）或 IGF-2、白介素（IL）（例如，IL-1、IL-2、IL-3、IL-4、IL-5、IL-6、IL-7）、角質形成細胞生長因子（KGF）、遷移刺激因子（MSF）、巨噬細胞刺激蛋白（MSP 或類肝細胞生長因子蛋白（HGFLP））、肌生成抑制蛋白（GDF-8）、神經調節蛋白（例如，神經調節蛋白 1（NRG1）、NRG2、NRG3、NRG4）、神經營養蛋白（例如，腦衍生的神經營養因子（BDNF）、神經生長因子（NGF）、神經營養蛋白-3（NT-3）、NT-4、胎盤生長因子（PGF）、血小板衍生的生長因子（PDGF）、腎胺酶（renalase）（RNLS）、T 細胞生長因子（TCGF）、血小板生成素（TPO）、轉化生長因子（例如，轉化生長因子 α（TGF-α）、TGF-β、腫瘤壞死因子-α（TNF-α）和血管內皮生長因子（VEGF）及其任何組合。

【0019】 在某些例示性實施例中，異源多核苷酸序列編碼激素。

【0020】 在某些例示性實施例中，異源多核苷酸序列編碼細胞激素。

【0021】 在某些例示性實施例中，異源多核苷酸序列編碼抗體或其片段。

【0022】 在某些例示性實施例中，異源多核苷酸序列編碼選自由下列組成之群組的基因：X-關聯性(X-linked)肌營養不良蛋白、MTM1（肌微管素）、酪胺酸羥化酶、AADC、環水解酶、SMN1、FXN（粒線體型共濟失調蛋白（frataxin））、GUCY2D、RS1、CFH、HTRA、ARMS、CFB/CC2、CNGA/CNGB、Prf65、ARSA、PSAP、IDUA（MPS I）、IDS（MPS II）、PAH、GAA（酸性α-葡糖苷酶）及其任何組合。

【0023】 在某些例示性實施例中，異源多核苷酸序列編碼微小RNA（miRNA）。在某些例示性實施例中，miRNA下調選自由下列組成之群組的靶基因的表現：SOD1、HTT、RHO及其任何組合。

【0024】 在某些例示性實施例中，異源多核苷酸序列編碼選自由下列組成之群組的凝血因子：I因子（FI）、II因子（FII）、III因子（FIII）、IV因子（FVI）、V因子（FV）、VI因子（FVI）、VII因子（FVII）、VIII因子（FVIII）、IX因子（FIX）、X因子（FX）、XI因子（FXI）、XII因子（FXII）、XIII因子（FVIII）、血管性血友病因子（Von Willebrand factor，VWF）、前微血管增滲素、高分子量激肽原、纖連蛋白、抗凝血酶III、肝素輔因子II、蛋白質C、蛋白質S、蛋白質Z、蛋白質Z相關蛋白酶抑制劑（ZPI）、纖維蛋白溶酶原、α2-抗纖維蛋白溶酶、組織纖維蛋白溶酶原活化劑（tPA）、尿激酶、纖維蛋白溶酶原活化物抑制劑-1（PAI-1）、纖維蛋白溶酶原活化物抑制劑-2（PAI2）及其任何組合。

【0025】 在某些例示性實施例中，凝血因子是FVIII。在某些例示性實施例中，FVIII包含全長成熟FVIII。在某些例示性實施例中，FVIII包含與具有SEQ ID NO: 106的胺基酸序列至少約70%、至少約75%、至少約80%、至少約85%、至少約90%、至少約95%、至少約96%、至少約97%、至少約98%、至少約99%或100%相同的胺基酸序列。

【0026】 在某些例示性實施例中，FVIII包含A1結構域、A2結構域、A3結構域、C1結構域、C2結構域和部分或無B結構域。在某些例示性實施例中，FVIII包含與SEQ ID NO:109的胺基酸序列至少約70%、至少約75%、至少約80%、至少約85%、至少約90%、至少約95%、至少約96%、至少約97%、至少約98%、至少約99%或100%相同的胺基酸序列。

【0027】 在某些例示性實施例中，凝血因子包含異源部分。在某些例示性實施例中，異源部分選自白蛋白或其片段、免疫球蛋白 Fc 區、人絨毛膜促性腺激素的 β 亞基的 C 末端肽（CTP）、PAS 序列、HAP 序列、轉鐵蛋白或其片段、白蛋白結合部分、其衍生物或其任何組合。在某些例示性實施例中，異源部分連接至 FVIII 的 N 末端或 C 末端或者插入 FVIII 的兩個胺基酸之間。在某些例示性實施例中，異源部分在選自表 4 中所列插入位點的一個或多個插入位點插入兩個胺基酸之間。

【0028】 在某些例示性實施例中，FVIII 還包含 A1 結構域、A2 結構域、C1 結構域、C2 結構域、任選的 B 結構域和異源部分，其中異源部分緊鄰對應於成熟 FVIII（SEQ ID NO: 106）的胺基酸 745 的下游插入。

【0029】 在某些例示性實施例中，FVIII 還包含 FcRn 結合配偶體 (partner)。在某些例示性實施例中，FcRn 結合配偶體包含免疫球蛋白恒定結構域的 Fc 區。

【0030】 在某些例示性實施例中，編碼 FVIII 的核酸序列經密碼子優化。在某些例示性實施例中，編碼 FVIII 的核酸序列經密碼子優化以供在人體內表現。

【0031】 在某些例示性實施例中，編碼 FVIII 的核酸序列包含與 SEQ ID NO: 107 的核苷酸序列至少約 60%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或約 100%相同的核苷酸序列。

【0032】 在某些例示性實施例中，編碼 FVIII 的核酸序列包含與 SEQ ID NO: 71 的核苷酸序列至少約 60%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或約 100%相同的核苷酸序列。

【0033】 在某些例示性實施例中，異源多核苷酸序列經密碼子優化。在某些例示性實施例中，異源多核苷酸序列經密碼子優化以供在人體內表現。

【0034】 在某些例示性實施例中，核酸分子以遞送劑配製。在某些例示性實施例中，遞送劑包含脂質奈米顆粒。在某些例示性實施例中，遞送劑選自脂質體、非脂質聚合分子和內體(endosome)及其任何組合。

【0035】 在某些例示性實施例中，核酸分子經配製用於靜脈內、經皮、真皮內、皮下、經肺或口服遞送或其任何組合。在某些例示性實施例中，核酸分子經配製用於靜脈內遞送。

【0036】 在某些態樣中，提供了包含如本文所述的核酸分子的載體。

【0037】 在某些態樣中，提供了包含如本文所述的核酸分子的宿主細胞。

【0038】 在某些態樣中，提供了醫藥組合物，其包含如本文所述的核酸分子或載體以及醫藥上可接受的賦形劑。

【0039】 在某些態樣中，提供了醫藥組合物，其包含如本文所述的宿主細胞以及醫藥上可接受的賦形劑。

【0040】 在某些態樣中，提供了套組，其包含如本文所述的核酸分子以及用於將核酸分子投予有需要的個體的說明書。

【0041】 在某些態樣中，提供了用於產生如本文所述的核酸分子的桿狀病毒系統。

【0042】 在某些例示性實施例中，在昆蟲細胞中產生如本文所述的核酸分子。

【0043】 在某些態樣中，提供了包含如本文所述的核酸分子的奈米顆粒遞送系統。

【0044】 在某些態樣中，提供了產生多肽的方法，其包括在適宜條件下培養如本文所述的宿主細胞和回收多肽。

【0045】 在某些態樣中，提供了產生具有凝血活性的多肽的方法，其包括：在適宜條件下培養如本文所述的宿主細胞和回收具有凝血活性的多肽。

【0046】 在某些態樣中，提供了在有需要的個體體內表現異源多核苷酸序列的方法，其包括將如本文所述的核酸分子、如本文所述的載體或如本文所述的醫藥組合物投予個體。

【0047】 在某些態樣中，提供了在有需要的個體體內表現凝血因子的方法，其包括將如本文所述的核酸分子、如本文所述的載體、如本文所述的多肽或如本文所述的醫藥組合物投予個體。

【0048】 在某些態樣中，提供了治療有需要的個體的疾病或障礙的方法，其包括將如本文所述的核酸分子、如本文所述的載體或如本文所述的醫藥組合物投予個體。

【0049】 在某些態樣中，提供了治療患有凝血因子缺乏的個體的方法，其包括將如本文所述的核酸分子、如本文所述的載體、如本文所述的多肽或如本文所述的醫藥組合物投予個體。

【0050】 在某些態樣中，提供了治療有需要的個體的凝血因子缺乏的方法，其包括將如本文所述的核酸分子、如本文所述的載體、如本文所述的多肽或如本文所述的醫藥組合物投予個體。

【0051】 在某些例示性實施例中，核酸分子是靜脈內、經皮、真皮內、皮下、口服、經肺或其任何組合地投予。在某些例示性實施例中，核酸分子是靜脈內投予。

【0052】 在某些例示性實施例中，所述方法還包括將第二藥劑投予個體。

【0053】 在某些例示性實施例中，個體是哺乳動物。在某些例示性實施例中，個體是人。

【0054】 在某些例示性實施例中，相對於投予前個體體內的 FVIII 活性，將核酸分子投予個體導致增加的 FVIII 活性，其中 FVIII 活性增加至少約 2 倍、至少約 3 倍、至少約 4 倍、至少約 5 倍、至少約 6 倍、至少約 7 倍、至少約 8 倍、至少約 9 倍、至少約 10 倍、至少約 11 倍、至少約 12 倍、至少約 13 倍、至少約 14 倍、至少約 15 倍、至少約 20 倍、至少約 25 倍、至少約 30 倍、至少約 35 倍、至少約 40 倍、至少約 50 倍、至少約 60 倍、至少約 70 倍、至少約 80 倍、至少約 90 倍或至少約 100 倍。

【0055】 在某些例示性實施例中，個體患有出血障礙。在某些例示性實施例中，出血障礙是血友病。在某些例示性實施例中，出血障礙是血友病 A。

【0056】 在某些態樣中，提供了治療有需要的個體的出血障礙的方法，其包括將核酸分子投予個體，所述核酸分子包含側接于基因盒的第一反向末端重複體（ITR）和第二 ITR，所述基因盒包含編碼凝血因子的異源多核苷酸序列，其中第一 ITR 和/或第二 ITR 包含與 SEQ ID NO: 180、181、183、184、185、186、187 或 188 中所述的核苷酸序列至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同的核苷酸序列或其功能衍生物。

【0057】 在某些態樣中，提供了治療有需要的個體的血友病 A 的方法，其包括將核酸分子投予個體，所述核酸分子包含側接于基因盒的第一反向末端重複體（ITR）和第二 ITR，所述基因盒包含編碼 VIII 因子（FVIII）的異源多核苷酸序列，其中第一 ITR 和/或第二 ITR 包含與 SEQ ID NO: 180、181、183、184、185、186、187 或 188 中所述的核苷酸序列至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同的核苷酸序列或其功能衍生物。

【0058】 在某些態樣中，提供了治療有需要的個體的肝臟代謝障礙的方法，其包括將核酸分子投予個體，所述核酸分子包含側接于基因盒的第一反向末端重複體（ITR）和第二 ITR，所述基因盒包含編碼在個體體內缺乏的肝臟相關代謝酶的異源多核苷酸序列，其中第一 ITR 和/或第二 ITR 是非腺相關病毒（非 AAV）的 ITR。

【0059】 在某些例示性實施例中，第一ITR和/或第二ITR包含與SEQ ID NO: 180、181、183、184、185、186、187 或 188 中所述的核苷酸序列至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同的核苷酸序列或其功能衍生物。

【0060】 在某些態樣中，提供了治療有需要的個體的肝臟代謝障礙的方法，其包括將核酸分子投予個體，所述核酸分子包含側接于基因盒的第一反向末端重複體（ITR）和第二 ITR，所述基因盒包含編碼在個體體內缺乏的肝臟相關代謝酶的異源多核苷酸序列，其中第一 ITR 和/或第二 ITR 包

含與 SEQ ID NO: 180、181、183、184、185、186、187 或 188 中所述的核苷酸序列至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同的核苷酸序列或其功能衍生物。

【0061】 在某些例示性實施例中，基因盒包含單鏈核酸。在某些例示性實施例中，基因盒包含雙鏈核酸。

【0062】 在某些例示性實施例中，肝臟代謝障礙選自由苯丙酮尿（PKU）、尿素循環疾病、溶酶體貯積障礙和糖原貯積病組成之群組。在某些例示性實施例中，肝臟代謝障礙是苯丙酮尿（PKU）。

【0063】 在某些例示性實施例中，核酸分子是靜脈內、經皮、真皮內、皮下、口服、經肺或其任何組合地投予。在某些例示性實施例中，核酸分子是靜脈內投予。

【0064】 在某些例示性實施例中，所述方法還包括將第二藥劑投予個體。

【0065】 在某些例示性實施例中，個體是哺乳動物。在某些例示性實施例中，個體是人類。

【0066】 在某些態樣中，提供了治療有需要的個體的苯丙酮尿（PKU）的方法，其包括將核酸分子投予個體，所述核酸分子包含側接于基因盒的第一反向末端重複體（ITR）和第二 ITR，所述基因盒包含編碼苯丙胺酸羥化酶的異源多核苷酸序列，其中第一 ITR 和/或第二 ITR 包含與 SEQ ID NO: 180、181、183、184、185、186、187 或 188 中所述的核苷酸序列至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同的核苷酸序列或其功能衍生物。

【0067】 在某些例示性實施例中，基因盒包含單鏈核酸。在某些例示性實施例中，基因盒包含雙鏈核酸。

【0068】 在某些例示性實施例中，核酸分子是用遞送劑配製。在某些例示性實施例中，遞送劑包含脂質奈米顆粒。

【0069】 在某些態樣中，提供了選殖核酸分子的方法，其包括將能夠形成複雜二級結構的核酸分子插入適宜載體中，和將所得載體引入包含SbcCD複合物中的破壞的細菌宿主菌株中。

【0070】 在某些例示性實施例中，SbcCD複合物中的破壞包含SbcC基因和/或SbcD基因中的基因破壞。在某些例示性實施例中，SbcCD複合物中的破壞包含SbcC基因中的基因破壞。在某些例示性實施例中，SbcCD複合物中的破壞包含SbcD基因中的基因破壞。

【0071】 在某些例示性實施例中，核酸分子包含第一反向末端重複體（ITR）和第二ITR，其中第一和/或第二ITR是非腺相關病毒（非AAV）ITR。

【0072】 在某些例示性實施例中，第一ITR和/或第二ITR包含與SEQ ID NO: 180、181、183、184、185、186、187或188中所述的核苷酸序列至少約75%、至少約80%、至少約85%、至少約90%、至少約95%、至少約96%、至少約97%、至少約98%、至少約99%或100%相同的核苷酸序列或其功能衍生物。

【0073】 在某些例示性實施例中，核酸分子還包含基因盒，其中基因盒側接有第一ITR和第二ITR。

【0074】 在某些例示性實施例中，基因盒包含異源多核苷酸序列。

【0075】 在某些例示性實施例中，適宜載體是低拷貝載體。在某些例示性實施例中，適宜載體是pBR322。

【0076】 在某些例示性實施例中，細菌宿主菌株不能拆分十字形DNA結構。

【0077】 在某些例示性實施例中，細菌宿主菌株是PMC103，其包含基因型 *sbcC*、*recD*、*mcrA*、*ΔmcrBCF*。在某些例示性實施例中，細菌宿主菌株是PMC107，其包含基因型 *recBC*、*recJ*、*sbcBC*、*mcrA*、*ΔmcrBCF*。在某些例示性實施例中，細菌宿主菌株是SURE，其包含基因型 *recB*、*recJ*、*sbcC*、*mcrA*、*ΔmcrBCF*、*umuC*、*uvrC*。

【0078】 在某些態樣中，提供了選殖核酸分子的方法，其包括將能夠形成複雜二級結構的核酸分子插入適宜載體中，和將所得載體引入包含

SbcCD 複合物中的破壞的細菌宿主菌株中，其中核酸分子包含第一反向末端重複體(ITR)和第二 ITR，其中第一 ITR 和/或第二 ITR 包含與 SEQ ID NO: 180、181、183、184、185、186、187 或 188 中所述的核苷酸序列至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同的核苷酸序列或其功能衍生物。

【圖式簡單說明】

【0079】 **圖 1A-1B** 是單鏈凝血因子（例如，FVIII）表現盒的示意圖。顯示了來自非 AAV 的 5' ITR（在 ssDNA 結構的末端具有髮夾環）、來自非 AAV 的 3' ITR（具有髮夾環）、啟動子序列（例如，TTPp 或 CAGp）和轉基因序列（例如，在 B 結構域內插入 XTEN144 的 FVIIIco6XTEN 序列）的位置。例示性表現盒還顯示了其他可能的元件，例如內含子序列、WPREmut 序列和 bGHpA 序列。

【0080】 **圖 1C-1F** 是用於製備單鏈凝血因子表現盒（如圖 1A-1B 中所示的盒）的質體的示意圖，其中盒的 ITR 衍生自 AAV2（圖 1C）、B19（圖 1D）、GPV（圖 1E），或者是野生型 B19 ITR 序列（圖 1F）。用 PvuII（在 PvuII 位點）（圖 1C）或 LguI（在 LguI 位點）（圖 1D-1F）消化包含如此處所示的 ssFVIII 表現盒的質體建構體，以精確釋放包含 ITR 和表現盒的序列。使雙鏈 DNA 在 95°C 下熱變性以產生 ssDNA，然後在 4°C 下培育以允許 ITR 結構形成。

【0081】 **圖 2A** 是種系發生樹(phylogenetic tree)，其說明瞭各種細小病毒家族成員之間的關係。B19、AAV-2 和 GPV 用輪廓框標記。

【0082】 **圖 2B** 是包括髮夾結構在內的各種盒的示意圖。

【0083】 **圖 3A 和 3B** 是 B19、GPV 和 AAV2 的 ITR（圖 3A）以及 B19 和 GPV 的 ITR（圖 3B）的比對。灰色陰影顯示了同源性。

【0084】 **圖 4A-4C** 顯示了在 Hem A 小鼠體內在通過流體力學注射（HDI）投予單鏈 FVIII-AAV 裸 DNA（ssAAV-FVIII；圖 1C）、ssDNA-B19 FVIII（圖 1D）或 ssDNA-GPV FVIII（圖 1E）後的 FVIII 血漿活性。在用 50 μg/小鼠（圖 4C）、20 μg/小鼠（圖 4A 和 4B）、10 μg/小鼠（圖 4A、4B

和 4C）或 5 μg/小鼠（圖 4A）的 ssDNA 的單次 HDI 處理的小鼠體內，在 24 小時、3 天、2 周、3 周、1 個月、2 個月、3 個月、4 個月、5 個月和 6 個月時，量測血漿樣本中的 FVIII 活性（作為人體內正常生理水準的百分比）。給出 5 μg/小鼠的質體 DNA 的 HDI 作為對照（圖 4A、4B 和 4C）。

【0085】 **圖 5** 顯示了在單次流體力學注射等摩爾量的單鏈裸 DNA（ssAAV-FVIII，圖 1A）、含有 ITR 序列的雙鏈 AAV-FVIII DNA（dsDNA）、不含 ITR 序列的雙鏈 FVIII DNA（無 ITR 的 dsDNA）或不含 ITR 或細菌序列的環化雙鏈 FVIII DNA（微環）後，血友病 A 小鼠血漿中的 FVIII 活性。dsDNA 是通過用 PvuII 對 AAV-FVIII 質體（圖 5）進行酶切割來產生，但沒有進行熱變性。無 ITR 的 dsDNA 是通過用 AflII 對 AAV-FVIII 質體（圖 5）進行酶切割來產生並且隨後進行純化。微環 DNA 是通過在 AflII 位點連接無 ITR DNA 的 dsDNA 來產生。經 3 個月或 4 個月收集小鼠血漿並且通過產色活性測定來確定 FVIII。

【0086】 **圖 6** 顯示了在流體力學注射 30 μg 單鏈裸 FVIII-DNA（圖 1A、圖 1D-1F）後，血友病 A 小鼠血漿中的 FVIII 活性。每週收集血漿持續 7 周並且通過產色測定來確定 FVIII 活性。在 35 天（描繪為黑色箭頭）後，通過流體力學注射向接受 FVIII-B19d135 和 FVIII-GPVd162 ssDNA 的小鼠再次投予 30 μg。

【0087】 **圖 7A** 是單鏈鼠類苯丙胺酸羥化酶（例如，PAH）表現盒的示意圖。顯示了來自非 AAV 的 5' ITR（在 ssDNA 結構的末端具有髮夾環）、來自非 AAV 的 3' ITR（具有髮夾環）、啟動子序列（例如，CAGp）和轉基因序列（例如，3xFLAG_mPAH 序列）的位置。例示性表現盒還顯示了其他可能的元件，例如 WPREmut 序列和 bGHpA 序列。

【0088】 **圖 7B-7D** 顯示了在通過流體力學注射單次投予含有鼠類 PAH cDNA 和非 AAV ITR B19d135 或 GPVd162 的單鏈 DNA 之前（第 0 天）和之後，苯丙酮尿（PKU）小鼠體內苯丙胺酸（Phe）的血漿濃度。在 ssDNA 投予後第 3、7、14、28、42 和 56 天收集血漿。殘餘苯丙胺酸水準顯示為以 μg/ml 計的濃度（圖 7B-7C）或顯示為投予前的百分比（圖 7D）。水平線描繪了投予前的基線 Phe 水準。

【0089】 **圖 7E** 顯示了來自使用含有鼠類 PAH 轉基因和 B19d135 或 GPVd165 ITR 的 ssDNA 處理的 PKU 小鼠的肝臟裂解物的西方免疫印跡。在處理後第 81 天收集肝臟，並提取蛋白質裂解物。每個孔代表單隻動物。使用 M2 抗 FLAG 抗體偵檢 FLAG 標記的鼠類 PAH 蛋白，並且包括 GAPDH 上樣對照以供比較。

【0090】 **圖 8A-B** 顯示了在使用囊封(encapsulated)FVIII-AAV DNA（圖 1A-1C）的脂質奈米顆粒轉導後，Huh7 細胞上清液中的 FVIII 活性水準。將在 CAGp 啟動子（圖 1B）下的質體 FVIII-AAV 以等於 3 的胺與磷酸（NP）比率囊封，並且以通過 picogreen 測定確定的不同濃度應用於 Huh7 細胞（圖 8A）。將在 TTPp 啟動子（圖 1A）下的質體、雙鏈線性（ds）和單鏈（ss）AAV-FVIII 以等於 2 的 NP 比率也囊封於脂質奈米顆粒中，並且用於以不同 DNA 濃度轉導 Huh7 細胞（圖 8B）。與人類 FACT 血漿標準品相比，通過產色(chromogenic)活性測定來量測 FVIII。

【實施方式】

【0091】 本公開文本描述了類質體核酸分子，其包含第一反向末端重複體（ITR）、第二 ITR 和基因盒（例如，編碼靶序列（本文中也稱作異源多核苷酸序列），例如治療性蛋白質或 miRNA），其中第一 ITR 和/或第二 ITR 是非腺相關病毒的 ITR（例如，第一 ITR 和/或第二 ITR 來自非 AAV）。在一些實施例中，基因盒編碼治療性蛋白質，例如，靶序列編碼治療性蛋白質。在一些實施例中，治療性蛋白質包含選自由下列組成之群組的蛋白質：凝血因子、生長因子、激素、細胞激素、抗體、其片段或其組合。在一些實施例中，基因盒編碼 X-關聯性肌營養不良蛋白、MTM1（肌微管素）、酪胺酸羥化酶、AADC、環水解酶、SMN1、FXN（粒線體型共濟失調蛋白）、GUCY2D、RS1、CFH、HTRA、ARMS、CFB/CC2、CNGA/CNGB、Prf65、ARSA、PSAP、IDUA（MPS I）、IDS（MPS II）、PAH、GAA（酸性 α-葡糖苷酶）或其任何組合。

【0092】 在一些實施例中，治療性蛋白質包含凝血因子。在一個特定實施例中，治療性蛋白質包含 FVIII 或 FIX 蛋白。

【0093】 在一些實施例中，基因盒編碼 miRNA。在某些實施例中，miRNA 下調選自由下列組成之群組的靶基因的表現：SOD1、HTT、RHO 或其任何組合。

【0094】 在某些實施例中，非 AAV 選自病毒科細小病毒科（*Parvoviridae*）的成員及其任何組合。本公開文本還涉及在有需要的個體體內表現治療性蛋白質（例如，凝血因子（例如，FVIII））的方法，其包括將包含第一反向末端重複體（ITR）、第二 ITR 和基因盒（例如，編碼治療性蛋白質或 miRNA）的核酸分子投予個體，其中第一 ITR 和/或第二 ITR 是非腺相關病毒（非 AAV）的 ITR。在某些實施例中，本公開文本描述了分離的核酸分子，其包含與選自 SEQ ID NO: 113 和 120 的核苷酸序列具有序列同源性的核苷酸序列。

【0095】 在某些實施例中，本公開文本提供了核酸分子，所述核酸分子包含側接于基因盒的第一反向末端重複體（ITR）和第二 ITR，所述基因盒包含異源多核苷酸序列，其中第一和/或第二 ITR 衍生自細小病毒 B19 或鵝細小病毒（GPV）。

【0096】 本公開文本的例示性建構體說明於附圖和序列表中。為了提供對說明書和申請專利範圍的清晰理解，下文提供了以下定義。

I. 定義

【0097】 應注意，術語「一個/一種（a）」或「一個/一種（an）」實體是指一個/一種或多個/多種所述實體：例如，「一個核苷酸序列」應理解為代表一個或多個核苷酸序列。類似地，「一種治療性蛋白質」和「一種 miRNA」應理解為分別代表一種或多種治療性蛋白質和一種或多種 miRNA。因此，術語「一個/一種（a）」（或「一個/一種（an）」）、「一個/一種或多個/多種」和「至少一個/一種」在本文中可互換使用。

【0098】 術語「約」在本文中用於意指大約、大致、大約地或在...左右。在術語「約」結合數字範圍使用時，其通過擴展所述數值上下的邊界來修飾所述範圍。通常，術語「約」在本文中用於向上或向下（較高或較低）以 10%的差異修飾高於和低於所述值的數值。

【0099】 同樣如本文所用，「和/或」是指並且涵蓋一個或多個相關列示項目的任何和所有可能組合，以及在替代方案（「或」）中解釋時組合的缺少。

【0100】 「核酸」、「核酸分子」、「核苷酸」、「一個或多個核苷酸的序列」和「多核苷酸」可互換使用並且是指呈單鏈形式或雙鏈螺旋的核糖核苷（腺苷、鳥苷、尿苷或胞苷；「RNA 分子」）或脫氧核糖核苷（脫氧腺苷、脫氧鳥苷、脫氧胸苷或脫氧胞苷；「DNA 分子」）的磷酸酯聚合形式或其任何磷酸酯類似物，如硫代磷酸酯和硫代酸酯。單鏈核酸序列是指單鏈 DNA（ssDNA）或單鏈 RNA（ssRNA）。雙鏈 DNA-DNA、DNA-RNA 和 RNA-RNA 螺旋是可能的。術語核酸分子、特別是 DNA 或 RNA 分子僅僅是指分子的一級和二級結構，並且不將其限制為任何特定三級形式。因此，這個術語包括尤其在線性或環狀 DNA 分子（例如，限制性片段）、質體、超螺旋化 DNA 和染色體中發現的雙鏈 DNA。在討論特定雙鏈 DNA 分子的結構時，在本文中可以根據常規習慣描述序列，僅沿 DNA 的非轉錄鏈（即，具有與 mRNA 同源的序列的鏈）以 5’至 3’方向給出序列。「重組 DNA 分子」是已經經歷分子生物學操縱的 DNA 分子。DNA 包括但不限於 cDNA、基因組 DNA、質體 DNA、合成 DNA 和半合成 DNA。本公開文本的「核酸組合物」包含如本文所述的一種或多種核酸。

【0101】 如本文所用，「反向末端重複體」（或「ITR」）是指位於單鏈核酸序列的 5'端或 3'端的核酸子序列(subsequence)，其包含一組核苷酸（初始序列），之後下游是其反向補體，即回文序列。初始序列與反向補體之間的插入核苷酸序列可以具有任何長度，包括零。在一個實施例中，可用於本公開文本的 ITR 包含一個或多個「回文序列」。ITR 可以具有任何數目的功能。在一些實施例中，本文所述的 ITR 形成髮夾結構。在一些實施例中，ITR 形成 T 形髮夾結構。在一些實施例中，ITR 形成非 T 形髮夾結構，例如 U 形髮夾結構。在一些實施例中，ITR 促進核酸分子在細胞的細胞核中的長期存活。在一些實施例中，ITR 促進核酸分子在細胞的細胞核中的永久存活（例如，持續細胞的整個壽命）。在一些實施例中，ITR 促進核酸分子在細胞的細胞核中的穩定性。在一些實施例中，ITR 促進核酸分子在細胞的細胞

核中的保留。在一些實施例中，ITR 促進核酸分子在細胞的細胞核中的持久性。在一些實施例中，ITR 抑制或防止核酸分子在細胞的細胞核中的降解。

【0102】 在一個實施例中，初始序列和/或反向補體包含約 2-600 個核苷酸、約 2-550 個核苷酸、約 2-500 個核苷酸、約 2-450 個核苷酸、約 2-400 個核苷酸、約 2-350 個核苷酸、約 2-300 個核苷酸或約 2-250 個核苷酸。在一些實施例中，初始序列和/或反向補體包含約 5-600 個核苷酸、約 10-600 個核苷酸、約 15-600 個核苷酸、約 20-600 個核苷酸、約 25-600 個核苷酸、約 30-600 個核苷酸、約 35-600 個核苷酸、約 40-600 個核苷酸、約 45-600 個核苷酸、約 50-600 個核苷酸、約 60-600 個核苷酸、約 70-600 個核苷酸、約 80-600 個核苷酸、約 90-600 個核苷酸、約 100-600 個核苷酸、約 150-600 個核苷酸、約 200-600 個核苷酸、約 300-600 個核苷酸、約 350-600 個核苷酸、約 400-600 個核苷酸、約 450-600 個核苷酸、約 500-600 個核苷酸或約 550-600 個核苷酸。在一些實施例中，初始序列和/或反向補體包含約 5-550 個核苷酸、約 5 至 500 個核苷酸、約 5-450 個核苷酸、約 5 至 400 個核苷酸、約 5-350 個核苷酸、約 5 至 300 個核苷酸或約 5-250 個核苷酸。在一些實施例中，初始序列和/或反向補體包含約 10-550 個核苷酸、約 15-500 個核苷酸、約 20-450 個核苷酸、約 25-400 個核苷酸、約 30-350 個核苷酸、約 35-300 個核苷酸或約 40-250 個核苷酸。在某些實施例中，初始序列和/或反向補體包含約 225 個核苷酸、約 250 個核苷酸、約 275 個核苷酸、約 300 個核苷酸、約 325 個核苷酸、約 350 個核苷酸、約 375 個核苷酸、約 400 個核苷酸、約 425 個核苷酸、約 450 個核苷酸、約 475 個核苷酸、約 500 個核苷酸、約 525 個核苷酸、約 550 個核苷酸、約 575 個核苷酸或約 600 個核苷酸。在特定實施例中，初始序列和/或反向補體包含約 400 個核苷酸。

【0103】 在其他實施例中，初始序列和/或反向補體包含約 2-200 個核苷酸、約 5-200 個核苷酸、約 10-200 個核苷酸、約 20-200 個核苷酸、約 30-200 個核苷酸、約 40-200 個核苷酸、約 50-200 個核苷酸、約 60-200 個核苷酸、約 70-200 個核苷酸、約 80-200 個核苷酸、約 90-200 個核苷酸、約 100-200 個核苷酸、約 125-200 個核苷酸、約 150-200 個核苷酸或約 175-200 個核苷酸。在其他實施例中，初始序列和/或反向補體包含約 2-150 個核苷

酸、約 5-150 個核苷酸、約 10-150 個核苷酸、約 20-150 個核苷酸、約 30-150 個核苷酸、約 40-150 個核苷酸、約 50-150 個核苷酸、約 75-150 個核苷酸、約 100-150 個核苷酸或約 125-150 個核苷酸。在其他實施例中，初始序列和/或反向補體包含約 2-100 個核苷酸、約 5-100 個核苷酸、約 10-100 個核苷酸、約 20-100 個核苷酸、約 30-100 個核苷酸、約 40-100 個核苷酸、約 50-100 個核苷酸或約 75-100 個核苷酸。在其他實施例中，初始序列和/或反向補體包含約 2-50 個核苷酸、約 10-50 個核苷酸、約 20-50 個核苷酸、約 30-50 個核苷酸、約 40-50 個核苷酸、約 3-30 個核苷酸、約 4-20 個核苷酸或約 5-10 個核苷酸。在另一個實施例中，初始序列和/或反向補體由 2 個核苷酸、3 個核苷酸、4 個核苷酸、5 個核苷酸、6 個核苷酸、7 個核苷酸、8 個核苷酸、9 個核苷酸、10 個核苷酸、11 個核苷酸、12 個核苷酸、13 個核苷酸、14 個核苷酸、15 個核苷酸、16 個核苷酸、17 個核苷酸、18 個核苷酸、19 個核苷酸或 20 個核苷酸組成。在其他實施例中，初始序列與反向補體之間的插入核苷酸是（例如，由以下組成）0 個核苷酸、1 個核苷酸、2 個核苷酸、3 個核苷酸、4 個核苷酸、5 個核苷酸、6 個核苷酸、7 個核苷酸、8 個核苷酸、9 個核苷酸、10 個核苷酸、11 個核苷酸、12 個核苷酸、13 個核苷酸、14 個核苷酸、15 個核苷酸、16 個核苷酸、17 個核苷酸、18 個核苷酸、19 個核苷酸或 20 個核苷酸。

【0104】 因此，如本文所用的「ITR」可以折疊回自身並形成雙鏈區段。例如，在折疊形成雙螺旋時，序列 GATCXXXXGATC 包含 GATC 的初始序列及其補體（3'CTAG5'）。在一些實施例中，ITR 包含初始序列與反向補體之間的連續回文序列（例如，GATCGATC）。在一些實施例中，ITR 包含初始序列與反向補體之間的中斷的回文序列（例如，GATCXXXXGATC）。在一些實施例中，連續或中斷的回文序列的互補部分彼此相互作用以形成「髮夾環」結構。如本文所用，在單鏈核苷酸分子上的至少兩個互補序列鹼基配對形成雙鏈部分時，產生「髮夾環」結構。在一些實施例中，僅 ITR 的一部分形成髮夾環。在其他實施例中，整個 ITR 形成髮夾環。

【0105】 在本公開文本中，至少一個 ITR 是非腺病毒相關病毒（非 AAV）的 ITR。在某些實施例中，ITR 是病毒科細小病毒科的非 AAV 成員的 ITR。在一些實施例中，ITR 是依賴病毒屬（*Dependovirus*）或紅病毒屬（*Erythrovirus*）的非 AAV 成員的 ITR。在特定實施例中，ITR 是以下病毒的 ITR：鵝細小病毒（GPV）、番鴨細小病毒（MDPV）或紅病毒屬細小病毒 B19（也稱為細小病毒 B19、靈長類紅細小病毒 1、B19 病毒和紅病毒）。在某些實施例中，兩個 ITR 中的一個 ITR 是 AAV 的 ITR。在其他實施例中，建構體中兩個 ITR 中的一個 ITR 是選自由下列組成之群組的 AAV 血清型的 ITR：血清型 1、2、3、4、5、6、7、8、9、10、11 及其任何組合。在一個特定實施例中，ITR 衍生自 AAV 血清型 2，例如 AAV 血清型 2 的 ITR。

【0106】 在本公開文本的某些態樣中，核酸分子包含兩個 ITR 5' ITR 和 3' ITR，其中 5' ITR 位於核酸分子的 5'末端，並且 3' ITR 位於核酸分子的 3'末端。5' ITR 和 3' ITR 可以衍生自相同病毒或不同病毒。在某些實施例中，5' ITR 衍生自 AAV 並且 3' ITR 並非衍生自 AAV 病毒（例如，非 AAV）。在一些實施例中，3' ITR 衍生自 AAV 並且 5' ITR 並非衍生自 AAV 病毒（例如，非 AAV）。在其他實施例中，5' ITR 並非衍生自 AAV 病毒（例如，非 AAV），並且 3' ITR 衍生自相同或不同的非 AAV 病毒。

【0107】 如本文所用的術語「細小病毒」涵蓋細小病毒科，包括但不限於自主複製的細小病毒和依賴病毒。自主細小病毒包括例如以下屬的成員：博卡病毒屬（*Bocavirus*）、依賴病毒屬、紅病毒屬、阿留申病毒屬（Amdovirus）、細小病毒屬（*Parvovirus*）、濃核病毒屬（*Densovirus*）、重複病毒屬（*Iteravirus*）、康特拉病毒屬（*Contravirus*）、禽細小病毒屬（*Aveparvovirus*）、反芻類細小病毒屬（*Copiparvovirus*）、原細小病毒屬（*Protoparvovirus*）、四型細小病毒屬（*Tetraparvovirus*）、雙義濃核病毒屬（*Ambidensovirus*）、短濃核病毒屬（*Brevidensovirus*）、肝胰濃核病毒屬（*Hepandensovirus*）和對蝦濃核病毒屬（*Penstyldensovirus*）。

【0108】 例示性自主細小病毒包括但不限於豬細小病毒、小鼠微小病毒、犬細小病毒、水貂腸炎病毒、牛細小病毒、雞細小病毒、貓瘟病毒、貓細小病毒、鵝細小病毒、H1 細小病毒、番鴨細小病毒、蛇細小病毒和 B19

病毒。其他自主細小病毒是業內熟習此項技術者已知的。參見例如，FIELDS等人 VIROLOGY, 第 2 卷, 第 69 章（第 4 版, Lippincott-Raven Publishers）。

【0109】 如本文所用的術語「非 AAV」涵蓋來自細小病毒科的核酸、蛋白質和病毒，不包括細小病毒科的任何腺相關病毒（AAV）。「非 AAV」包括但不限於以下屬的自主複製成員：博卡病毒屬、依賴病毒屬、紅病毒屬、阿留申病毒屬、細小病毒屬、濃核病毒屬、重複病毒屬、康特拉病毒屬、禽細小病毒屬、反芻類細小病毒屬、原細小病毒屬、四型細小病毒屬、雙義濃核病毒屬、短濃核病毒屬、肝胰濃核病毒屬和對蝦濃核病毒屬。

【0110】 如本文所用，術語「腺相關病毒」（AAV）包括但不限於 AAV 1 型、AAV 2 型、AAV 3 型（包括 3A 型和 3B 型）、AAV 4 型、AAV 5 型、AAV 6 型、AAV 7 型、AAV 8 型、AAV 9 型、AAV 10 型、AAV 11 型、AAV 12 型、AAV 13 型、蛇 AAV、禽 AAV、牛 AAV、犬 AAV、馬 AAV、綿羊 AAV、山羊 AAV、蝦 AAV、Gao 等人（J. Virol. 78:6381 (2004)）和 Moris 等人（Virol. 33:375 (2004)）所披露的那些 AAV 血清型和進化枝以及目前已知或將來發現的任何其他 AAV。參見例如，FIELDS 等人 VIROLOGY, 第 2 卷, 第 69 章（第 4 版, Lippincott-Raven Publishers）。

【0111】 如本文所用，術語「衍生自」是指組分從指定分子或生物體分離或使用指定分子或生物體製備，或者資訊（例如，胺基酸或核酸序列）來自指定分子或生物體。例如，衍生自第二核酸序列（例如，ITR）的核酸序列（例如，ITR）可以包括與第二核酸序列的核苷酸序列相同或基本上相似的核苷酸序列。在核苷酸或多肽的情況下，衍生的物質可以通過例如天然存在的突變誘發、人工定向突變誘發或人工隨機突變誘發來獲得。用於衍生核苷酸或多肽的突變誘發可以是有意定向的或有意隨機的，或者是各自的混合。突變誘發核苷酸或多肽以產生衍生自第一核苷酸或多肽的不同核苷酸或多肽可以是隨機事件（例如，由於聚合酶失真(infidelity)而引起），並且所衍生的核苷酸或多肽的鑑定可以通過適當的篩選方法（例如，如本文所討論）來進行。多肽的突變誘發通常需要操縱編碼多肽的多核苷酸。在一些實施例中，衍生自第二核苷酸或胺基酸序列的核苷酸或胺基酸序列分別與第二核苷酸或胺基酸序列具有至少 50%、至少 51%、至少 52%、至

少 53%、至少 54%、至少 55%、至少 56%、至少 57%、至少 58%、至少 59%、至少 60%、至少 61%、至少 62%、至少 63%、至少 64%、至少 65%、至少 66%、至少 67%、至少 68%、至少 69%、至少 70%、至少 71%、至少 72%、至少 73%、至少 74%、至少 75%、至少 76%、至少 77%、至少 78%、至少 79%、至少 80%、至少 81%、至少 82%、至少 83%、至少 84%、至少 85%、至少 86%、至少 87%、至少 88%、至少 89%、至少 90%、至少 91%、至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%、至少 99%或 100%的序列同一性，其中第一核苷酸或胺基酸序列保留第二核苷酸或胺基酸序列的生物學活性。在其他實施例中，衍生自非 AAV（或 AAV）的 ITR 的 ITR 與非 AAV ITR（或分別地，AAV ITR）至少 90%相同，其中非 AAV（或 AAV）ITR 保留非 AAV ITR（或分別地，AAV ITR）的功能特性。在一些實施例中，衍生自非 AAV（或 AAV）的 ITR 的 ITR 與非 AAV ITR（或分別地，AAV ITR）至少 80%相同，其中非 AAV（或 AAV）ITR 保留非 AAV ITR（或分別地，AAV ITR）的功能特性。在一些實施例中，衍生自非 AAV（或 AAV）的 ITR 的 ITR 與非 AAV ITR（或分別地，AAV ITR）至少 70%相同，其中非 AAV（或 AAV）ITR 保留非 AAV ITR（或分別地，AAV ITR）的功能特性。在一些實施例中，衍生自非 AAV（或 AAV）的 ITR 的 ITR 與非 AAV ITR（或分別地，AAV ITR）至少 60%相同，其中非 AAV（或 AAV）ITR 保留非 AAV ITR（或分別地，AAV ITR）的功能特性。在一些實施例中，衍生自非 AAV（或 AAV）的 ITR 的 ITR 與非 AAV ITR（或分別地，AAV ITR）至少 50%相同，其中非 AAV（或 AAV）ITR 保留非 AAV ITR（或分別地，AAV ITR）的功能特性。

【0112】 在某些實施例中，衍生自非 AAV（或 AAV）的 ITR 的 ITR 包含非 AAV（或 AAV）的 ITR 的片段或由所述片段組成。在一些實施例中，衍生自非 AAV（或 AAV）的 ITR 的 ITR 包含非 AAV（或 AAV）的 ITR 的片段或由所述片段組成，其中所述片段包含至少約 5 個核苷酸、至少約 10 個核苷酸、至少約 15 個核苷酸、至少約 20 個核苷酸、至少約 25 個核苷酸、至少約 30 個核苷酸、至少約 35 個核苷酸、至少約 40 個核苷酸、至少約 45 個核苷酸、至少約 50 個核苷酸、至少約 55 個核苷酸、至少約 60 個核

苷酸、至少約 65 個核苷酸、至少約 70 個核苷酸、至少約 75 個核苷酸、至少約 80 個核苷酸、至少約 85 個核苷酸、至少約 90 個核苷酸、至少約 95 個核苷酸、至少約 100 個核苷酸、至少約 125 個核苷酸、至少約 150 個核苷酸、至少約 175 個核苷酸、至少約 200 個核苷酸、至少約 225 個核苷酸、至少約 250 個核苷酸、至少約 275 個核苷酸、至少約 300 個核苷酸、至少約 325 個核苷酸、至少約 350 個核苷酸、至少約 375 個核苷酸、至少約 400 個核苷酸、至少約 425 個核苷酸、至少約 450 個核苷酸、至少約 475 個核苷酸、至少約 500 個核苷酸、至少約 525 個核苷酸、至少約 550 個核苷酸、至少約 575 個核苷酸或至少約 600 個核苷酸；其中衍生自非 AAV（或 AAV）的 ITR 的 ITR 保留非 AAV ITR（或分別地，AAV ITR）的功能特性。在某些實施例中，衍生自非 AAV（或 AAV）的 ITR 的 ITR 包含非 AAV（或 AAV）的 ITR 的片段或由所述片段組成，其中所述片段包含至少約 129 個核苷酸，並且其中衍生自非 AAV（或 AAV）的 ITR 的 ITR 保留非 AAV ITR（或分別地，AAV ITR）的功能特性。在某些實施例中，衍生自非 AAV（或 AAV）的 ITR 的 ITR 包含非 AAV（或 AAV）的 ITR 的片段或由所述片段組成，其中所述片段包含至少約 102 個核苷酸，並且其中衍生自非 AAV（或 AAV）的 ITR 的 ITR 保留非 AAV ITR（或分別地，AAV ITR）的功能特性。

【0113】 在一些實施例中，衍生自非 AAV（或 AAV）的 ITR 的 ITR 包含非 AAV（或 AAV）的 ITR 的片段或由所述片段組成，其中所述片段包含非 AAV（或 AAV）的 ITR 的長度的至少約 5%、至少約 10%、至少約 15%、至少約 20%、至少約 25%、至少約 30%、至少約 35%、至少約 40%、至少約 45%、至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%。

【0114】 在某些實施例中，在正確比對時，衍生自第二核苷酸或胺基酸序列的核苷酸或胺基酸序列分別與第二核苷酸或胺基酸序列的同源部分具有至少 50%、至少 51%、至少 52%、至少 53%、至少 54%、至少 55%、至少 56%、至少 57%、至少 58%、至少 59%、至少 60%、至少 61%、至少 62%、至少 63%、至少 64%、至少 65%、至少 66%、至少 67%、至少 68%、

至少 69%、至少 70%、至少 71%、至少 72%、至少 73%、至少 74%、至少 75%、至少 76%、至少 77%、至少 78%、至少 79%、至少 80%、至少 81%、至少 82%、至少 83%、至少 84%、至少 85%、至少 86%、至少 87%、至少 88%、至少 89%、至少 90%、至少 91%、至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%、至少 99%或 100%的序列同一性，其中第一核苷酸或胺基酸序列保留第二核苷酸或胺基酸序列的生物學活性。在其他實施例中，在正確比對時，衍生自非 AAV（或 AAV）的 ITR 的 ITR 與非 AAV ITR（或分別地，AAV ITR）的同源部分至少 90%相同，其中第一核苷酸或胺基酸序列保留第二核苷酸或胺基酸序列的生物學活性。在一些實施例中，在正確比對時，衍生自非 AAV（或 AAV）的 ITR 的 ITR 與非 AAV ITR（或分別地，AAV ITR）的同源部分至少 80%相同，其中第一核苷酸或胺基酸序列保留第二核苷酸或胺基酸序列的生物學活性。在一些實施例中，在正確比對時，衍生自非 AAV（或 AAV）的 ITR 的 ITR 與非 AAV ITR（或分別地，AAV ITR）的同源部分至少 70%相同，其中第一核苷酸或胺基酸序列保留第二核苷酸或胺基酸序列的生物學活性。在一些實施例中，在正確比對時，衍生自非 AAV（或 AAV）的 ITR 的 ITR 與非 AAV ITR（或分別地，AAV ITR）的同源部分至少 60%相同，其中第一核苷酸或胺基酸序列保留第二核苷酸或胺基酸序列的生物學活性。在一些實施例中，在正確比對時，衍生自非 AAV（或 AAV）的 ITR 的 ITR 與非 AAV ITR（或分別地，AAV ITR）的同源部分至少 50%相同，其中第一核苷酸或胺基酸序列保留第二核苷酸或胺基酸序列的生物學活性。

【0115】 「無衣殼（capsid-free）」或「少衣殼（capsid-less）」載體或核酸分子是指不具有衣殼的載體建構體。在一些實施例中，少衣殼載體或核酸分子不含編碼例如 AAV Rep 蛋白的序列。

【0116】 如本文所用，「編碼區」或「編碼序列」是多核苷酸中由可轉譯成胺基酸的密碼子組成的部分。雖然「終止密碼子」（TAG、TGA 或 TAA）通常不轉譯成胺基酸，但是可以認為其是編碼區的一部分，但是任何側接序列（例如啟動子、核糖體結合位點、轉錄終止子、內含子等）不是編碼

區的一部分。編碼區的邊界通常由 5'末端的起始密碼子（編碼所得多肽的胺基末端）和 3'末端的轉譯終止密碼子（編碼所得多肽的羧基末端）來決定。兩個或更多個編碼區可以存在於單一多核苷酸建構體中（例如，在單一載體上），或者存在於單獨的多核苷酸建構體中（例如，在單獨的（不同的）載體上）。然後，結果就是單一載體可以僅含有單個編碼區，或者包含兩個或更多個編碼區。

【0117】 哺乳動物細胞分泌的某些蛋白質與分泌信號肽相關，一旦生長中的蛋白質鏈跨越糙面內質網的輸出已經起始所述分泌信號肽便從成熟蛋白質切割下來。業內熟習此項技術者知道，信號肽通常融合至多肽的 N 末端，並且是從完整或“全長”多肽切割下來以產生多肽的分泌或“成熟”形式。在某些實施例中，天然信號肽或所述序列的保留指導多肽分泌的能力的功能衍生物與所述多肽可操作地相關。可替代地，可以使用異源哺乳動物信號肽（例如，人組織纖維蛋白溶酶原活化劑（TPA）或小鼠 ß-葡糖醛酸糖苷酶信號肽）或其功能衍生物。

【0118】 術語「下游」是指核苷酸序列位於參考核苷酸序列的 3'。在某些實施例中，下游核苷酸序列是指轉錄起點之後的序列。例如，基因的轉譯起始密碼子位於轉錄起始位點的下游。

【0119】 術語「上游」是指核苷酸序列位於參考核苷酸序列的 5'。在某些實施例中，上游核苷酸序列是指位於編碼區或轉錄起點的 5'側的序列。例如，大多數啟動子位於轉錄起始位點的上游。

【0120】 如本文所用，術語「基因調控區」或「調控區」是指位於編碼區的上游（5'非編碼序列）、內部或下游（3'非編碼序列），並且影響相關編碼區的轉錄、RNA 加工、穩定性或轉譯的核苷酸序列。調控區可以包括啟動子、轉譯前導序列、內含子、多聚腺苷酸化識別序列、RNA 加工位點、效應子結合位點或莖環結構。如果意圖在真核細胞中表現編碼區，則多聚腺苷酸化信號和轉錄終止序列通常將位於編碼序列的 3'。

【0121】 編碼產物（例如，miRNA 或基因產物（例如，多肽，如治療性蛋白質））的多核苷酸可以包括與一個或多個編碼區可操作地締合的啟動子和/或其他表現（例如，轉錄或轉譯）控制元件。在可操作締合中，

以將基因產物的表現置於一個或多個調控區的影響或控制下的方式將基因產物（例如，多肽）的編碼區與所述一個或多個調控區締合。例如，如果啟動子功能的誘導引起編碼由編碼區所編碼的基因產物的 mRNA 的轉錄，並且如果啟動子與編碼區之間的連接的性質不干擾啟動子指導基因產物表現的能力或者不干擾 DNA 範本被轉錄的能力，則編碼區和啟動子“可操作地締合”。除了啟動子以外的其他表現控制元件（例如，強化子、操縱子、阻遏物和轉錄終止信號）也可以與編碼區可操作地締合以指導基因產物表現。

【0122】 「轉錄控制序列」是指提供編碼序列在宿主細胞中的表現的 DNA 調控序列，如啟動子、強化子、終止子等。多種轉錄控制區是業內熟習此項技術者已知的。這些轉錄控制區包括但不限於作用於脊椎動物細胞中的轉錄控制區，例如但不限於來自巨細胞病毒（立即早期啟動子，與內含子-A 結合）、猿猴病毒 40（早期啟動子）和逆轉錄病毒（如勞斯肉瘤病毒）的啟動子和強化子區段。其他轉錄控制區包括衍生自脊椎動物基因的那些控制區，如肌動蛋白、熱休克蛋白、牛生長激素和兔 ß-珠蛋白，以及能夠控制真核細胞中的基因表現的其他序列。其他適宜轉錄控制區包括組織特異性啟動子和強化子以及淋巴因子誘導型啟動子（例如，可由干擾素或白介素誘導的啟動子）。

【0123】 類似地，多種轉譯控制元件是業內熟習此項技術者已知的。這些轉譯控制元件包括但不限於核糖體結合位點、轉譯起始和終止密碼子以及衍生自微小核糖核酸病毒的元件（特別是內部核糖體進入位點，或 IRES，也稱為 CITE 序列）。

【0124】 如本文所用的術語「表現」是指多核苷酸產生基因產物（例如，RNA 或多肽）的過程。其包括但不限於多核苷酸轉錄為信使 RNA（mRNA）、轉移 RNA（tRNA）、小髮夾 RNA（shRNA）、小干擾 RNA（siRNA）或任何其他 RNA 產物，以及 mRNA 轉譯為多肽。表現產生“基因產物”。如本文所用，基因產物可以是核酸（例如，通過基因轉錄產生的信使 RNA）或從轉錄物轉譯的多肽。本文所述的基因產物還包括具有轉錄後修飾（例如，多聚腺苷酸化或剪接）的核酸，或者具有轉譯後修飾（例

如，甲基化、糖基化、脂質添加、與其他蛋白質亞單元締合或蛋白質水解切割）的多肽。如本文所用，術語「產量」是指通過基因表現產生的多肽的量。

【0125】 「載體」是指用於將核酸選殖和/或轉移至宿主細胞中的任何媒劑。載體可以是複製子，另一核酸區段可以與其連接以實現所連接區段的複製。「複製子」是指在體內起自主複製單元的作用（即，能夠在其自身控制下複製）的任何基因元件（例如，質體、噬菌體、粘粒、染色體、病毒）。術語「載體」包括在體外、離體或在體內將核酸引入細胞中的媒劑。業內已知並使用大量載體，包括例如質體、經修飾真核病毒或經修飾細菌病毒。將多核苷酸插入適宜載體中可以通過將適當多核苷酸片段連接至具有互補粘性末端的所選載體中來完成。

【0126】 可以對載體進行工程化以編碼可選標記或報告子，其提供對已經併入載體的細胞的選擇或鑑定。可選標記或報告子的表現允許鑑定和/或選擇併入並表現含於載體上的其他編碼區的宿主細胞。業內已知並使用的可選標記基因的例子包括：提供針對安比西林、鏈黴素、慶大黴素、卡那黴素、潮黴素、雙丙胺磷除草劑、磺醯胺等的抗性的基因；和用作表型標記的基因，即花色苷調控基因、異戊烷基轉移酶基因等。業內已知並使用的報告子的例子包括：螢光素酶（Luc）、綠色螢光蛋白（GFP）、氯黴素乙醯轉移酶（CAT）、β-半乳糖苷酶（LacZ）、β-葡糖醛酸糖苷酶（Gus）等。還可以將可選標記視為報告子(reporter)。

【0127】 如本文所用的術語「宿主細胞」是指例如可以或已經用作ssDNA 或載體的接受者的微生物、酵母細胞、昆蟲細胞和哺乳動物細胞。所述術語包括已經轉導的初始細胞的後代。因此，如本文所用的「宿主細胞」通常是指已經用外源 DNA 序列轉導的細胞。應理解，由於天然、偶然或刻意突變，單個親代細胞的後代的形態或基因組或總 DNA 補體可能不一定與初始親代完全相同。在一些實施例中，宿主細胞可以是體外宿主細胞。

【0128】 術語「可選標記」是指能夠基於標記基因的效應（即，對抗生素的抗性、對除草劑的抗性、比色標記、酶、螢光標記等）加以選擇的鑑定因子（通常是抗生素或化學抗性基因），其中所述效應用於跟蹤所關注

核酸的遺傳和/或鑑定已經遺傳所關注核酸的細胞或生物體。業內已知並使用的可選標記基因的例子包括：提供針對安比西林、鏈黴素、慶大黴素、卡那黴素、潮黴素、雙丙胺磷除草劑、磺醯胺等的抗性的基因；和用作表型標記的基因，即花色苷調控基因、異戊烷基轉移酶基因等。

【0129】 術語「報告子基因」是指編碼能夠基於報告子基因的效應加以鑑定的鑑定因子的核酸，其中所述效應用於跟蹤所關注核酸的遺傳性、鑑定已經遺傳所關注核酸的細胞或生物體和/或量測基因表現誘導或轉錄。業內已知並使用的報告子基因的例子包括：螢光素酶（Luc）、綠色螢光蛋白（GFP）、氯黴素乙醯轉移酶（CAT）、β-半乳糖苷酶（LacZ）、β-葡糖醛酸糖苷酶（Gus）等。也可以將可選標記基因視為報告子基因。

【0130】 「啟動子」與「啟動子序列」可互換使用並且是指能夠控制編碼序列或功能 RNA 的表現的 DNA 序列。通常，編碼序列位於啟動子序列的 3'。啟動子可以整體衍生自天然基因，或者由衍生自在自然界中發現的不同啟動子的不同元件構成，或者甚至包含合成 DNA 區段。業內熟習此項技術者應理解，不同的啟動子可以指導基因在不同組織或細胞類型中、或在不同發育階段、或響應不同環境或生理條件而表現。使基因在大多數細胞類型中在大多數時間表現的啟動子一般被稱為「組成型啟動子」。使基因在特定細胞類型中表現的啟動子一般被稱為「細胞特異性啟動子」或「組織特異性啟動子」。使基因在發育或細胞分化的特定階段表現的啟動子一般被稱為「發育特異性啟動子」或「細胞分化特異性啟動子」。在用誘導啟動子的試劑、生物分子、化學品、配位體、光等暴露或處理細胞後被誘導並且使基因表現的啟動子一般被稱為「誘導型啟動子」或「可調型啟動子」。還認識到，由於在大多數情況下，尚未完全界定調控序列的確切邊界，所以不同長度的 DNA 片段可以具有相同的啟動子活性。

【0131】 啟動子序列通常在其 3'末端以轉錄起始位點為界，並向上游（5'方向）延伸以包括以高於背景的可偵檢水準起始轉錄所必需的最小數目的鹼基或元件。在啟動子序列內將發現轉錄起始位點（例如通過用核酸酶 S1 標示（mapping）方便地界定）以及負責 RNA 聚合酶的結合的蛋白質結合結構域（共有序列）。

【0132】 在一些實施例中，核酸分子包含組織特異性啟動子。在某些實施例中，組織特異性啟動子驅動治療性蛋白質（例如，凝血因子）在肝臟中（例如，在肝細胞和/或內皮細胞中）的表現。在特定實施例中，啟動子選自小鼠甲狀腺素啟動子（mTTR）、內源人類 VIII 因子啟動子（F8）、人類 α-1-抗胰蛋白酶啟動子（hAAT）、人類白蛋白最小啟動子、小鼠白蛋白啟動子、三重四脯胺酸（TTP）啟動子、CASI 啟動子、CAG 啟動子、巨細胞病毒（CMV）啟動子、磷酸甘油酸激酶（PGK）啟動子及其任何組合。在一些實施例中，啟動子選自肝臟特異性啟動子（例如，α1-抗胰蛋白酶（AAT））、肌肉特異性啟動子（例如，肌肉肌酸激酶（MCK）、肌球蛋白重鏈 α（αMHC）、肌紅蛋白（MB）和結蛋白（DES））、合成啟動子（例如，SPc5-12、2R5Sc5-12、dMCK 和 tMCK）及其任何組合。在一個特定實施例中，啟動子包含 TTP 啟動子。

【0133】 術語「限制性內切核酸酶」與「限制酶」可互換使用並且是指在雙鏈 DNA 內的特定核苷酸序列內結合並切割的酶。

【0134】 術語「質體」是指通常攜帶並非細胞的中心代謝的一部分的基因，並且通常呈環狀雙鏈 DNA 分子的形式的染色體外元件。此類元件可以是衍生自任何來源的單鏈或雙鏈 DNA 或 RNA 的線性、環狀或超螺旋化的自主複製序列、基因組整合序列、噬菌體或核苷酸序列，其中多個核苷酸序列已經接合或重組至獨特建構體中，所述建構體能夠將用於所選基因產物的啟動子片段和 DNA 序列以及適當的 3'非轉譯序列引入細胞中。

【0135】 可以使用的真核病毒載體包括但不限於腺病毒載體、逆轉錄病毒載體、腺相關病毒載體、痘病毒（例如，牛痘病毒載體）、桿狀病毒載體或皰疹病毒載體。非病毒載體包括質體、脂質體、帶電脂質（細胞轉染劑）、DNA-蛋白質複合物和生物聚合物。

【0136】 「選殖載體」是指「複製子」，其是單位長度的連續複製的核酸並且其包含複製起點，如質體、噬菌體或粘粒，另一核酸區段可以與所述複製子連接以實現所連接區段的複製。某些選殖載體能夠在一種細胞類型（例如，細菌）中複製，並在另一細胞類型（例如，真核細胞）中表現。

選殖載體通常包含一個或多個可以用於選擇包含載體的細胞的序列和/或一個或多個用於插入所關注核酸序列的多重選殖位點。

【0137】 術語「表現載體」是指設計為使得所插入核酸序列在插入宿主細胞中之後能夠表現的媒劑。將所插入核酸序列以如上所述的與調控區可操作地締合地放置。

【0138】 通過業內熟知的方法將載體引入宿主細胞中，所述方法是例如轉染、電穿孔、顯微注射、轉導、細胞融合、DEAE 葡聚糖、磷酸鈣沈澱、脂轉染（溶酶體融合）、使用基因槍或 DNA 載體轉運蛋白。如本文所用，「培養（culture）」、「培養（to culture）」和「培養（culturing）」意指在允許細胞生長或分裂的體外條件下培育細胞或使細胞維持存活狀態。如本文所用，「所培養細胞」意指在體外繁殖的細胞。

【0139】 如本文所用，術語「多肽」旨在涵蓋單數「多肽」以及複數「多肽」，並且是指由通過醯胺鍵（也稱為肽鍵）線性連接的單體（胺基酸）構成的分子。術語「多肽」是指兩個或更多個胺基酸的任何一條或多條鏈，並且不涉及產物的具體長度。因此，肽、二肽、三肽、寡肽、「蛋白質」、「胺基酸鏈」或用於指兩個或更多個胺基酸的一條或多條鏈的任何其他術語包括於「多肽」的定義內，並且術語「多肽」可以代替這些術語中的任一個術語或與其互換使用。術語「多肽」還意圖指多肽的表現後修飾的產物，所述修飾包括但不限於糖基化、乙醯化、磷酸化、醯胺化、通過已知的保護/阻斷基團衍生、蛋白質水解切割或通過非天然存在的胺基酸修飾。多肽可以從天然生物來源衍生或通過重組技術產生，但不一定是從指定的核酸序列轉譯。其能以任何方式產生，包括通過化學合成產生。

【0140】 術語「胺基酸」包括丙胺酸（Ala 或 A）；精胺酸（Arg 或 R）；天門冬醯胺酸（Asn 或 N）；天門冬胺酸（Asp 或 D）；半胱胺酸（Cys 或 C）；麩醯胺酸（Gln 或 Q）；麩胺酸（Glu 或 E）；甘胺酸（Gly 或 G）；組胺酸（His 或 H）；異白胺酸（Ile 或 I）；白胺酸（Leu 或 L）；離胺酸（Lys 或 K）；甲硫胺酸（Met 或 M）；苯丙胺酸（Phe 或 F）；脯胺酸（Pro 或 P）；絲胺酸（Ser 或 S）；蘇胺酸（Thr 或 T）；色胺酸（Trp 或 W）；酪胺酸（Tyr 或 Y）；和撷胺酸（Val 或 V）。非傳統胺基酸也在本公開文

本的範圍內，並且包括正白胺酸、鳥胺酸、正撷胺酸、高絲胺酸和其他胺基酸殘基類似物，如 Ellman 等人 Meth. Enzym. 202:301-336 (1991)中所述的那些。為了產生此類非天然存在的胺基酸殘基，可以使用 Noren 等人 Science 244:182 (1989)和 Ellman 等人（同上）的程式。簡單來說，這些程式涉及用非天然存在的胺基酸殘基化學啟動抑制劑 tRNA，之後進行 RNA 的體外轉錄和轉譯。引入非傳統胺基酸也可以使用業內已知的肽化學來實現。如本文所用，術語「極性胺基酸」包括具有零個淨電荷，但是在其側鏈的不同部分具有非零部分電荷的胺基酸（例如，M、F、W、S、Y、N、Q、C）。這些胺基酸可以參與疏水相互作用和靜電相互作用。如本文所用，術語「帶電胺基酸」包括在其側鏈上具有非零淨電荷的胺基酸（例如，R、K、H、E、D）。這些胺基酸可以參與疏水相互作用和靜電相互作用。

【0141】 本公開文本中還包括多肽的片段或變異體及其任何組合。術語「片段」或「變異體」在涉及本公開文本的多肽結合結構域或結合分子時包括保留參考多肽的至少一些特性（例如，對 FcRn 結合結構域或 Fc 變異體的 FcRn 結合親和性、對 FVIII 變異體的凝結活性或對 VWF 片段的 FVIII 結合活性）的任何多肽。除了本文中其他地方討論的特定抗體片段以外，多肽片段包括蛋白質水解片段以及缺失片段，但是不包括天然存在的全長多肽（或成熟多肽）。本公開文本的多肽結合結構域或結合分子的變異體包括如上所述的片段，以及由於胺基酸取代、缺失或插入而具有改變的胺基酸序列的多肽。變異體可以是天然存在的或非天然存在的。非天然存在的變異體可以使用業內已知的突變誘發技術來產生。變異體多肽可以包含保守或非保守胺基酸取代、缺失或添加。

【0142】 「保守胺基酸取代」是用具有類似側鏈的胺基酸殘基替代胺基酸殘基的取代。業內已經定義具有類似側鏈的胺基酸殘基的家族，包括鹼性側鏈（例如，離胺酸、精胺酸、組胺酸）、酸性側鏈（例如，天門冬胺酸、麩胺酸）、不帶電極性側鏈（例如，甘胺酸、天門冬醯胺酸、麩醯胺酸、絲胺酸、蘇胺酸、酪胺酸、半胱胺酸）、非極性側鏈（例如，丙胺酸、撷胺酸、白胺酸、異白胺酸、脯胺酸、苯丙胺酸、甲硫胺酸、色胺酸）、β-分支側鏈（例如，蘇胺酸、撷胺酸、異白胺酸）和芳香族側鏈（例如，酪

胺酸、苯丙胺酸、色胺酸、組胺酸）。因此，如果用來自相同側鏈家族的另一胺基酸替代多肽中的胺基酸，則將取代視為保守的。在另一個實施例中，胺基酸串可以用結構上類似但側鏈家族成員的順序和/或組成不同的串來保守地替代。

【0143】 如業內已知的術語「同一性百分比」是兩個或更多個多肽序列或者兩個或更多個多核苷酸序列之間的關係，如通過比較序列所確定。在業內，根據具體情況，「同一性」也意指多肽或多核苷酸序列之間的序列關聯性程度，如通過此類序列的串之間的匹配所確定。「同一性」可以通過已知方法容易地計算，所述方法包括但不限於以下文獻中所述的那些方法：*Computational Molecular Biology* (Lesk, A. M.編輯) Oxford University Press, New York (1988)；*Biocomputing: Informatics and Genome Projects* (Smith, D. W.編輯) Academic Press, New York (1993)；*Computer Analysis of Sequence Data, Part I* (Griffin, A. M.和 Griffin, H. G.編輯) Humana Press, New Jersey (1994)；*Sequence Analysis in Molecular Biology* (von Heinje, G.編輯) Academic Press (1987)；和 *Sequence Analysis Primer* (Gribskov, M.和 Devereux, J.編輯) Stockton Press, New York (1991)。確定同一性的較佳方法設計為給出所測試序列之間的最佳匹配。確定同一性的方法被編纂於可公開獲得的電腦程式中。序列比對和同一性百分比計算可以使用序列分析軟體來進行，所述序列分析軟體是例如 LASERGENE 生物資訊學計算套件的 Megalign 程式（DNASTAR Inc.，威斯康辛州麥迪遜市）、GCG 程式套件（Wisconsin Package 9.0 版，Genetics Computer Group（GCG），威斯康辛州麥迪遜市）、BLASTP、BLASTN、BLASTX（Altschul 等人, *J. Mol. Biol. 215*:403 (1990)）和 DNASTAR（DNASTAR, Inc. 1228 S. Park St. 威斯康辛州麥迪遜市 53715 USA）。在本申請的上下文中將理解，如果使用序列分析軟體進行分析，除非另外指定，否則分析結果將基於所參考程式的「缺省值」。如本文所用的「缺省值」將意指在首次初始化時最初用軟體加載的任一組值或參數。出於確定本公開文本的治療性蛋白質（例如，凝血因子）序列與參考序列之間的同一性百分比的目的，僅使用參考序列中對應於本公開文本的治療性蛋白質（例如，凝血因子）序列中的核苷酸的核苷酸來計算同一性百分比。

例如，在比較含有 B 結構域的全長 FVIII 核苷酸序列與本公開文本的優化的 B 結構域缺失（BDD）FVIII 核苷酸序列時，將使用包括 A1、A2、A3、C1 和 C2 結構域的比對部分來計算同一性百分比。全長 FVIII 序列的編碼 B 結構域的部分（其將在比對中產生大的「空位」）中的核苷酸將不會被計為錯配。另外，在確定本公開文本的優化的 BDD FVIII 序列或其指定部分（例如，SEQ ID NO: 3 的核苷酸 58-2277 和 2320-4374）與參考序列之間的同一性百分比時，將通過比對、用匹配核苷酸數除以如本文所述的優化的 BDD-FVIII 序列或其指定部分的完整序列中的核苷酸總數來計算同一性百分比。

【0144】 如本文所用，對應於本公開文本的特定序列中的核苷酸的核苷酸是通過比對本公開文本的序列以最大化與參考序列的同一性來鑑定。用於鑑定參考序列中的等效胺基酸的編號是基於用於鑑定本公開文本的序列中的相應胺基酸的編號。

【0145】 「融合」或「嵌合」蛋白包含連接至第二胺基酸序列的第一胺基酸序列，所述第一胺基酸序列在自然界中並非天然地與所述第二胺基酸序列連接。可以使通常存在於單獨蛋白質中的胺基酸序列在融合多肽中聚在一起，或者可以將通常存在於相同蛋白質中的胺基酸序列以新排列置於融合多肽（例如，本公開文本的 VIII 因子結構域與 Ig Fc 結構域的融合物）中。例如通過化學合成或通過產生並轉譯其中以所需關係編碼肽區的多核苷酸來產生融合蛋白。嵌合蛋白還可以包含通過共價非肽鍵或非共價鍵與第一胺基酸序列締合的第二胺基酸序列。

【0146】 如本文所用，術語「插入位點」是指多肽或其片段、變異體或衍生物中的如下位置，其緊鄰可以插入異源部分的位置的上游。將「插入位點」指定為編號，所述編號是胺基酸在參考序列中的編號。例如，FVIII 中的「插入位點」是指成熟天然 FVIII（SEQ ID NO: 15）中插入位點所對應的胺基酸序列的編號，其緊鄰插入位置的 N 末端。例如，片語「a3 包含在對應於 SEQ ID NO: 15 的胺基酸 1656 的插入位元點之異源部分」指示，異源部分位於對應於 SEQ ID NO: 15 的胺基酸 1656 和胺基酸 1657 的兩個胺基酸之間。

【0147】 如本文所用的片語「緊鄰胺基酸的下游」是指與胺基酸的末端羧基緊緊相鄰的位置。類似地，片語「緊鄰胺基酸的上游」是指與胺基酸的末端胺基緊緊相鄰的位置。

【0148】 如本文所用的術語「插入（inserted）」、「插入（is inserted）」、「插入（inserted into）」或語法相關術語涉及相對於親代多肽中的類似位置，異源部分在多肽（例如，凝血因子）中的位置。例如，在某些實施例中，「插入」等涉及相對于天然成熟人類 FVIII 中的類似位置，異源部分在重組 FVIII 多肽中的位置。如本文所用，所述術語是指多肽的特徵，並且並非指示、暗示或意味著製備多肽的任何方法或工藝。

【0149】 如本文所用，術語「半衰期」是指特定多肽在體內的生物半衰期。半衰期可以表示為投予個體的數量的一半被從動物的循環和/或其他組織中清除所需的時間。在將給定多肽的清除曲線構建為時間函數時，所述曲線通常是二相的，具有快速的 α 相和較長的 β 相。α 相通常代表所投予的 Fc 多肽在血管內與血管外空間之間的平衡，並且部分取決於多肽的大小。β 相通常代表多肽在血管內空間中的分解代謝。在一些實施例中，治療性蛋白質（例如，凝血因子，例如 FVIII）和包含所述治療性蛋白質的嵌合蛋白是單相的，並且因此不具有 α 相，而是只具有單一 β 相。因此，在某些實施例中，如本文所用的術語半衰期是指多肽在 β 相中的半衰期。

【0150】 如本文所用的術語「連接」是指第一胺基酸序列或核苷酸序列分別與第二胺基酸序列或核苷酸序列共價或非共價接合。第一胺基酸或核苷酸序列可以與第二胺基酸或核苷酸序列直接接合或並置，或者可替代地，插入序列可以將第一序列共價接合至第二序列。術語“連接”不僅意指第一胺基酸序列與第二胺基酸序列在 C 末端或 N 末端融合，還包括將完整的第一胺基酸序列（或第二胺基酸序列）插入第二胺基酸序列（或分別地，第一胺基酸序列）中的任兩個胺基酸之間。在一個實施例中，第一胺基酸序列可以通過肽鍵或連接子連接至第二胺基酸序列。第一核苷酸序列可以通過磷酸二酯鍵或連接子連接至第二核苷酸序列。連接子可以是肽或多肽（對於多肽鏈）或者核苷酸或核苷酸鏈（對於核苷酸鏈）或者任何化

學部分（對於多肽和多核苷酸鏈二者）。術語「連接」還通過連字元（-）來指示。

【0151】 如本文所用的止血意指停止或減慢出血（bleeding）或出血症（hemorrhage）；或者停止或減慢血液流經血管或身體部分。

【0152】 如本文所用的止血障礙意指遺傳上的遺傳性或獲得性病症，特徵在於由於形成纖維蛋白凝塊的能力受損或無能，傾向於自發地或由於創傷而出血。此類障礙的例子包括血友病。三種主要形式是血友病 A（VIII 因子缺乏）、血友病 B（IX 因子缺乏或「克雷司馬斯病」）和血友病 C（XI 因子缺乏，輕度出血傾向）。其他止血障礙包括例如血管性血友病、XI 因子缺乏（PTA 缺乏）、XII 因子缺乏、纖維蛋白原、凝血酶原、V 因子、VII 因子、X 因子或 XIII 因子的缺乏或結構異常、GPIb 缺陷或缺乏的巨血小板綜合征。vWF 受體 GPIb 可能缺陷並導致缺少初級凝塊形成（初級止血）和增加的出血傾向、以及 Glanzman 和 Naegeli 的血小板無力症（Glanzmann 血小板無力症）。在肝功能衰竭（急性和慢性形式）中，肝臟的凝結因子產生不足；這可能增加出血風險。

【0153】 可以預防性地使用本公開文本的分離的核酸分子、分離的多肽或包含分離核酸分子的載體。如本文所用，術語「預防性治療」是指在出血發作之前投予分子。在一個實施例中，需要通用止血劑的個體正在進行或即將進行手術。本公開文本的多核苷酸、多肽或載體可以作為預防藥在手術之前或之後投予。本公開文本的多核苷酸、多肽或載體可以在手術期間或之後投予以控制急性出血發作。手術可以包括但不限於肝移植、肝切除、牙科手術或幹細胞移植。

【0154】 本公開文本的分離的核酸分子、分離的多肽或載體也用於按需要治療。術語「按需要治療」是指響應出血發作的症狀或者在可能引起出血的活動之前投予分離的核酸分子、分離的多肽或載體。在一個態樣中，可以在出血開始時（如在損傷後）或在預期要出血時（如在手術前），將按需要治療投予個體。在另一個態樣中，可以在增加出血風險的活動（如接觸運動）之前投予按需要治療。

【0155】 如本文所用，術語「急性出血」是指不論潛在原因的出血發作。例如，個體可能患有創傷、尿毒症、遺傳性出血障礙（例如，VII因子缺乏）、血小板障礙或由於產生針對凝血因子的抗體而產生的抗性。

【0156】 如本文所用的治療（treat）、治療（treatment）、治療（treating）是指例如疾病或病症的嚴重程度的降低；病程持續時間的減少；與疾病或病症相關的一種或多種症狀的改善；向患有疾病或病症的個體提供有益作用，但不一定治癒疾病或病症；或者預防與疾病或病症相關的一種或多種症狀。在一個實施例中，術語「治療（treating）」或「治療（treatment）」意指在個體體內通過投予本公開文本的分離的核酸分子、分離的多肽或載體維持例如 FVIII 谷水準為至少約 1 IU/dL、2 IU/dL、3 IU/dL、4 IU/dL、5 IU/dL、6 IU/dL、7 IU/dL、8 IU/dL、9 IU/dL、10 IU/dL、11 IU/dL、12 IU/dL、13 IU/dL、14 IU/dL、15 IU/dL、16 IU/dL、17 IU/dL、18 IU/dL、19 IU/dL、20 IU/dL、25 IU/dL、30 IU/dL、35 IU/dL、40 IU/dL、45 IU/dL、50 IU/dL、55 IU/dL、60 IU/dL、65 IU/dL、70 IU/dL、75 IU/dL、80 IU/dL、85 IU/dL、90 IU/dL、95 IU/dL、100 IU/dL、105 IU/dL、110 IU/dL、115 IU/dL、120 IU/dL、125 IU/dL、130 IU/dL、135 IU/dL、140 IU/dL、145 IU/dL 或 150 IU/dL。在另一個實施例中，治療（treating）或治療（treatment）意指維持 FVIII 谷水準介於約 1 與約 150 IU/dL 之間、約 1 與約 125 IU/dL 之間、約 1 與約 100 IU/dL 之間、約 1 與約 90 IU/dL 之間、約 1 與約 85 IU/dL 之間、約 1 與約 80 IU/dL 之間、約 1 與約 75 IU/dL 之間、約 1 與約 70 IU/dL 之間、約 1 與約 65 IU/dL 之間、約 1 與約 60 IU/dL 之間、約 1 與約 55 IU/dL 之間、約 1 與約 50 IU/dL 之間、約 1 與約 45 IU/dL 之間、約 1 與約 40 IU/dL 之間、約 1 與約 35 IU/dL 之間、約 1 與約 30 IU/dL 之間、約 1 與約 25 IU/dL 之間、約 25 與約 125 IU/dL 之間、約 50 與約 100 IU/dL 之間、約 50 與約 75 IU/dL 之間、約 75 與約 100 IU/dL 之間、約 1 與約 20 IU/dL 之間、約 2 與約 20 IU/dL 之間、約 3 與約 20 IU/dL 之間、約 4 與約 20 IU/dL 之間、約 5 與約 20 IU/dL 之間、約 6 與約 20 IU/dL 之間、約 7 與約 20 IU/dL 之間、約 8 與約 20 IU/dL 之間、約 9 與約 20 IU/dL 之間或約 10 與約 20 IU/dL 之間。疾病或病症的治療（treatment）或治療（treating）還可以包括維持個體體內的 FVIII 活性為以下水準：與非

血友病個體體內的 FVIII 活性的至少約 1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、100%、105%、110%、115%、120%、125%、130%、135%、140%、145%或 150%相當。治療所需的最低谷水準可以通過一種或多種已知方法來量測，並且可以針對每個個人加以調節（增加或減小）。

【0157】 如本文所用，「投予」意指通過醫藥上可接受的途徑將本公開文本的醫藥上可接受的核酸分子、從所述核酸分子表現的多肽或包含所述核酸分子的載體投予個體。投予途徑可以是靜脈內的，例如靜脈內注射和靜脈內輸注。其他投予途徑包括例如皮下、肌內、口服、經鼻和經肺投予。核酸分子、多肽和載體可以作為包含至少一種賦形劑的醫藥組合物的一部分來投予。

【0158】 如本文所用的術語「醫藥上可接受的」是指分子實體和組合物在投予人時在生理上可耐受並且通常不會產生毒性或過敏反應或類似的不良反應（如胃部不適、頭暈等）。任選地，如本文所用，術語「醫藥上可接受的」意指由聯邦政府或州政府的管理機構批准或者在美國藥典或其他公認藥典中列示用於動物，更特別是用於人類。

【0159】 如本文所用，片語「有需要的個體」包括將受益於本公開文本的核酸分子、多肽或載體的投予例如以改良止血的個體（如哺乳動物個體）。在一個實施例中，個體包括但不限於患有血友病的個體。在另一個實施例中，個體包括但不限於已經產生針對治療性蛋白質（例如，凝血因子，例如 FVIII）的抑制劑，並且因此需要繞路療法的個體。個體可以是成年人或未成年人（例如，小於 12 歲）。

【0160】 如本文所用，術語「治療性蛋白質」是指業內已知可以投予個體的任何多肽。在一些實施例中，治療性蛋白質包含選自由下列組成之群組的蛋白質：凝血因子、生長因子、抗體、其功能片段或其組合。如本文所用，術語"凝血因子"是指天然存在的或重組產生的分子或其類似物，其預防個體的出血發作或縮短出血發作的持續時間。換句話說，其意指具有促凝血活性（即，負責將纖維蛋白原轉化為不溶性纖維蛋白的網狀物，從而

引起血液凝結或凝血）的分子。如本文所用的「凝血因子」包括啟動的凝血因子、其酶原或可啟動的凝血因子。「可啟動的凝血因子」是呈無活性形式（例如，呈其酶原形式）的凝血因子，其能夠被轉化為活性形式。術語「凝血因子」包括但不限於I因子（FI）、II因子（FII）、V因子（FV）、FVII、FVIII、FIX、X因子（FX）、XI因子（FXI）、XII因子（FXII）、XIII因子（FXIII）、血管性血友病因子（VWF）、前微血管增滲素、高分子量激肽原、纖連蛋白、抗凝血酶III、肝素輔因子II、蛋白質C、蛋白質S、蛋白質Z、蛋白質Z相關蛋白酶抑制劑（ZPI）、纖維蛋白溶酶原、α2-抗纖維蛋白溶酶、組織纖維蛋白溶酶原活化劑（tPA）、尿激酶、纖維蛋白溶酶原活化物抑制劑-1（PAI-1）、纖維蛋白溶酶原活化物抑制劑-2（PAI2）、其酶原、其活化形式或其任何組合。

【0161】 如本文所用的凝血活性意指參與生化反應級聯的能力，所述級聯以形成纖維蛋白凝塊告終和/或降低出血症或出血發作的嚴重程度、持續時間或頻率。

【0162】 如本文所用，「生長因子」包括業內已知的包括細胞激素和激素在內的任何生長因子。在一些實施例中，生長因子選自腎上腺髓質素（AM）、血管生成素（Ang）、自分泌運動因子、骨形態發生蛋白（BMP）（例如，BMP2、BMP4、BMP5、BMP7）、睫狀神經營養因子家族成員（例如，睫狀神經營養因子（CNTF）、白血病抑制因子（LIF）、白介素-6（IL-6））、集落刺激因子（例如，巨噬細胞集落刺激因子（m-CSF）、粒細胞集落刺激因子（G-CSF）、粒細胞巨噬細胞集落刺激因子（GM-CSF））、表皮生長因子（EGF）、肝配蛋白（例如，肝配蛋白A1、肝配蛋白A2、肝配蛋白A3、肝配蛋白A4、肝配蛋白A5、肝配蛋白B1、肝配蛋白B2、肝配蛋白B3）、促紅細胞生成素（EPO）、成纖維細胞生長因子（FGF）（例如，FGF1、FGF2、FGF3、FGF4、FGF5、FGF6、FGF7、FGF8、FGF9、FGF10、FGF11、FGF12、FGF13、FGF14、FGF15、FGF16、FGF17、FGF18、FGF19、FGF20、FGF21、FGF22、FGF23）、胎牛促體素（FBS）、GDNF家族成員（例如，神經膠質細胞株衍生的神經營養因子（GDNF）、神經秩蛋白、普塞芬蛋白、阿特敏蛋白）、生長分化因子-9（GDF9）、肝細胞生長因子

（HGF）、肝腫瘤衍生的生長因子（HDGF）、胰島素、類胰島素生長因子（例如，類胰島素生長因子-1（IGF-1）或IGF-2、白介素（IL）（例如，IL-1、IL-2、IL-3、IL-4、IL-5、IL-6、IL-7）、角質形成細胞生長因子（KGF）、遷移刺激因子（MSF）、巨噬細胞刺激蛋白（MSP或類肝細胞生長因子蛋白（HGFLP））、肌生成抑制蛋白（GDF-8）、神經調節蛋白（例如，神經調節蛋白1（NRG1）、NRG2、NRG3、NRG4）、神經營養蛋白（例如，腦衍生的神經營養因子（BDNF）、神經生長因子（NGF）、神經營養蛋白-3（NT-3）、NT-4、胎盤生長因子（PGF）、血小板衍生的生長因子（PDGF）、腎胺酶（RNLS）、T細胞生長因子（TCGF）、血小板生成素（TPO）、轉化生長因子（例如，轉化生長因子α（TGF-α）、TGF-β、腫瘤壞死因子-α（TNF-α）和血管內皮生長因子（VEGF）。

【0163】 在一些實施例中，治療性蛋白質由選自由下列組成之群組的基因編碼：X-關聯性肌營養不良蛋白、MTM1（肌微管素）、酪胺酸羥化酶、AADC、環水解酶、SMN1、FXN（粒線體型共濟失調蛋白）、GUCY2D、RS1、CFH、HTRA、ARMS、CFB/CC2、CNGA/CNGB、Prf65、ARSA、PSAP、IDUA（MPS I）、IDS（MPS II）、PAH、GAA（酸性α-葡糖苷酶）或其任何組合。

【0164】 如本文所用，術語「異源的」或「外源的」是指此類分子通常在給定背景下（例如，在細胞中或在多肽中）未發現。例如，可以將外源或異源分子引入細胞中並且僅在例如通過轉染或其他形式的遺傳工程化操縱細胞之後存在，或者異源胺基酸序列可以存在于其天然不存在的蛋白質中。

【0165】 如本文所用，術語「異源核苷酸序列」是指不與給定多核苷酸序列一起天然存在的核苷酸序列。在一個實施例中，異源核苷酸序列編碼能夠延長治療性蛋白質（例如，凝血因子，例如FVIII）的半衰期的多肽。在另一個實施例中，異源核苷酸序列編碼增加治療性蛋白質（例如，凝血因子，例如FVIII）的流體力學半徑的多肽。在其他實施例中，異源核苷酸序列編碼改良治療性蛋白質的一種或多種藥物代謝動力學特性但不顯著影響其生物學活性或功能（例如，促凝血活性）的多肽。在一些實施例中，治療性蛋白質通過連接子與異源核苷酸序列編碼的多肽連接或相連。異源

核苷酸序列編碼的多肽部分的非限制性例子尤其包括免疫球蛋白恒定區或其部分、白蛋白或其片段、白蛋白結合部分、轉鐵蛋白、美國專利申請號20100292130的PAS多肽、HAP序列、轉鐵蛋白或其片段、人絨毛膜促性腺激素的β亞基的C末端肽(CTP)、白蛋白結合小分子、XTEN序列、FcRn結合部分(例如,完整Fc區或其結合至FcRn的部分)、單鏈Fc區(ScFc區,例如如US 2008/0260738、WO 2008/012543或WO 2008/1439545中所述)、聚甘胺酸連接子、聚絲胺酸連接子、選自甘胺酸(G)、丙胺酸(A)、絲胺酸(S)、蘇胺酸(T)、麩胺酸(E)和脯胺酸(P)的兩種類型胺基酸的6-40個胺基酸的肽和短多肽(具有從小於50%至大於50%變化的二級結構程度)、或其兩個或更多個組合。在一些實施例中,異源核苷酸序列編碼的多肽連接至非多肽部分。非多肽部分的非限制性例子包括聚乙二醇(PEG)、白蛋白結合小分子、聚唾液酸、羥乙基澱粉(HES)、其衍生物或其任何組合。

【0166】 如本文所用,術語「Fc區」定義為多肽中對應于天然Ig的Fc區的部分,即如通過其兩條重鏈的相應Fc結構域的二聚締合所形成。天然Fc區與另一Fc區形成同二聚體。相比之下,如本文所用的術語「基因融合的Fc區」或「單鏈Fc區」(scFc區)是指合成二聚Fc區,其包括基因連接於單條多肽鏈內(即,編碼於單個連續基因序列中)的Fc結構域。

【0167】 在一個實施例中,「Fc區」是指單條Ig重鏈的如下部分,其始於緊鄰木瓜蛋白酶切割位點(即,IgG中的殘基216,以重鏈恒定區的第一個殘基為114)上游的鉸鏈區並終於抗體的C末端。因此,完整Fc結構域至少包含鉸鏈結構域、CH2結構域和CH3結構域。

【0168】 根據Ig同種型,Ig恒定區的Fc區可以包括CH2、CH3和CH4結構域以及鉸鏈區。包含Ig的Fc區的嵌合蛋白賦予嵌合蛋白若干種所需特性,包括增加的穩定性、增加的血清半衰期(參見Capon等人, 1989, *Nature* 337:525)以及與Fc受體(如新生Fc受體(FcRn))的結合(美國專利號6,086,875、6,485,726、6,030,613;WO 03/077834;US2003-0235536A1),將上述文獻通過引用以其整體併入本文。

【0169】 「參考核苷酸序列」在本文中用作與本公開文本的核苷酸序列的比較時，是與本公開文本的核苷酸序列基本上相同的多核苷酸序列，但是所述參考序列未經優化。例如，由 SEQ ID NO: 1 的密碼子優化的 BDD FVIII 和在其 3'端連接至 SEQ ID NO: 1 的編碼單鏈 Fc 區的異源核苷酸序列組成的核酸分子的參考核苷酸序列是由 SEQ ID NO: 16 的初始（或「親代」）BDD FVIII 和在其 3'端連接至 SEQ ID NO: 16 的編碼單鏈 Fc 區的相同異源核苷酸序列組成的核酸分子。

【0170】 如本文所用，關於核苷酸序列的術語「優化」是指編碼多肽的多核苷酸序列，其中所述多核苷酸序列已經突變以增強所述多核苷酸序列的特性。在一些實施例中，進行優化以增加轉錄水準、增加轉譯水準、增加穩態 mRNA 水準、增加或減少調控蛋白（如通用轉錄因子）的結合、增加或減少剪接或增加多核苷酸序列產生的多肽的產量。可以對多核苷酸序列進行以使其優化的變化的例子包括密碼子優化、G/C 含量優化、去除重複序列、去除富含 AT 的元件、去除隱蔽剪接位點、去除阻遏轉錄或轉譯的順式作用元件、添加或去除多聚 T 或多聚 A 序列、在轉錄起始位點周圍添加增強轉錄的序列（如 Kozak 共有序列）、去除可以形成莖環結構的序列、去除去穩定序列及其兩個或更多個組合。

II. 核酸分子

【0171】 本公開文本涉及編碼靶序列的類質體無衣殼的核酸分子，其中靶序列編碼治療性蛋白質或可以調節靶蛋白（例如，miRNA）的表現的基因。衣殼是病毒的蛋白質外殼，包封病毒的遺傳物質。已知衣殼通過保護病毒基因組、將基因組遞送至主體和與主體相互作用來輔助病毒粒子的功能。然而，尤其在用於基因療法中時，病毒衣殼可以是限制載體的包裝能力和/或誘導免疫應答方面的因素。

【0172】 AAV 載體已經作為一種更常見類型的基因療法載體出現。然而，衣殼的存在限制了 AAV 載體在基因療法中的效用。具體地，衣殼自身可能將載體中包括的轉基因的大小限制到低至小於 4.5 kb。即使在添加調控元件之前，可用于基因療法中的多種治療性蛋白質也可以輕易超過這個大小。

【0173】 此外，構成衣殼的蛋白質可能作為個體的免疫系統可靶向的抗原。AAV在一般群體中非常常見，大多數人在其一生中已經暴露於AAV。因此，大多數潛在的基因療法接受者可能已經產生過針對AAV的免疫應答，並且因此更有可能排斥所述療法。

【0174】 本公開文本的某些態樣旨在克服AAV載體的這些缺陷。具體地，本公開文本的某些態樣涉及核酸分子，其包含第一ITR、第二ITR和基因盒（例如，編碼治療性蛋白質和/或miRNA）。在一些實施例中，第一ITR和第二ITR側接於包含異源多核苷酸序列的基因盒。在一些實施例中，核酸分子不包含編碼衣殼蛋白、複製蛋白和/或組裝蛋白的基因。在一些實施例中，基因盒編碼治療性蛋白質。在一些實施例中，治療性蛋白質包含凝血因子。在一些實施例中，基因盒編碼miRNA。在某些實施例中，基因盒定位於第一ITR與第二ITR之間。在一些實施例中，核酸分子還包含一個或多個非編碼區。在某些實施例中，所述一個或多個非編碼區包含啟動子序列、內含子、轉錄後調控元件、3'UTR多聚(A)序列或其任何組合。

【0175】 在一個實施例中，基因盒是單鏈核酸。在另一個實施例中，基因盒是雙鏈核酸。

【0176】 在一個實施例中，核酸分子包含：

(a) 作為細小病毒科的非AAV家族成員的ITR（例如，B19或GPV ITR）的第一ITR；

(b) 組織特異性啟動子序列，例如TTP啟動子；

(c) 內含子，例如合成內含子；

(d) 編碼miRNA或治療性蛋白質（例如，凝血因子）的核苷酸；

(e) 轉錄後調控元件，例如WPRE；

(f) 3'UTR多聚(A)尾序列，例如bGHpA；

(g) 作為細小病毒科的非AAV家族成員的ITR（例如，B19或GPV ITR）的第二ITR。

【0177】 在一個實施例中，核酸分子包含：

(a) 作為細小病毒科的非AAV家族成員的ITR的第一ITR；

(b) 組織特異性啟動子序列，例如TTP啟動子；

(c) 內含子，例如合成內含子；

(d) 編碼miRNA的核苷酸，其中所述miRNA下調選自由下列組成之群組的靶基因的表現：SOD1、HTT、RHO及其任何組合；

(e) 轉錄後調控元件，例如WPRE；

(f) 3'UTR多聚(A)尾序列，例如bGHpA；

(g) 作為細小病毒科的非AAV家族成員的ITR的第二ITR。

【0178】 在一個實施例中，核酸分子包含：

(a) 作為細小病毒科的非AAV家族成員的ITR的第一ITR；

(b) 組織特異性啟動子序列，例如TTP啟動子；

(c) 內含子，例如合成內含子；

(d) 編碼X-關聯性肌營養不良蛋白、MTM1（肌微管素）、酪胺酸羥化酶、AADC、環水解酶、SMN1、FXN（粒線體型共濟失調蛋白）、GUCY2D、RS1、CFH、HTRA、ARMS、CFB/CC2、CNGA/CNGB、Prf65、ARSA、PSAP、IDUA（MPS I）、IDS（MPS II）、PAH、GAA（酸性α-葡糖苷酶）或其任何組合的核苷酸；

(e) 轉錄後調控元件，例如WPRE；

(f) 3'UTR多聚(A)尾序列，例如bGHpA；

(g) 作為細小病毒科的非AAV家族成員的ITR的第二ITR。

【0179】 在一個實施例中，核酸分子包含：

(a) 作為AAV（例如，AAV血清型2基因組）的ITR的第一ITR；

(b) 組織特異性啟動子序列，例如TTP啟動子；

(c) 內含子，例如合成內含子；

(d) 編碼FVIII的核苷酸；其中所述核苷酸與選自SEQ ID NO: 1-14或SEQ ID NO: 71的核苷酸序列具有至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性，其中所述核苷酸編碼的FVIII保留FVIII活性；

(e) 轉錄後調控元件，例如WPRE；

(f) 3'UTR多聚(A)尾序列，例如bGHpA；和

(g) 作為AAV（例如，AAV血清型2基因組）的ITR的第二ITR。

【0180】 在一個實施例中，核酸分子包含：

(a) 作為AAV（例如，AAV血清型2基因組）的ITR的第一ITR；

(b) 組織特異性啟動子序列，例如TTP啟動子；

(c) 內含子，例如合成內含子；

(d) 編碼miRNA的核苷酸，其中所述miRNA下調靶基因的表現，所述靶基因是例如SOD1、HTT、RHO及其任何組合；

(f) 3'UTR多聚(A)尾序列，例如bGHpA；和

(g) 作為AAV（例如，AAV血清型2基因組）的ITR的第二ITR。

【0181】 在一個實施例中，核酸分子包含：

(a) 作為AAV（例如，AAV血清型2基因組）的ITR的第一ITR；

(b) 組織特異性啟動子序列，例如TTP啟動子；

(c) 內含子，例如合成內含子；

(d) 編碼X-關聯性肌營養不良蛋白、MTM1（肌微管素）、酪胺酸羥化酶、AADC、環水解酶、SMN1、FXN（粒線體型共濟失調蛋白）、GUCY2D、RS1、CFH、HTRA、ARMS、CFB/CC2、CNGA/CNGB、Prf65、ARSA、PSAP、IDUA（MPS I）、IDS（MPS II）、PAH、GAA（酸性α-葡糖苷酶）或其任何組合的核苷酸；

(f) 3'UTR多聚(A)尾序列，例如bGHpA；和

(g) 作為AAV（例如，AAV血清型2基因組）的ITR的第二ITR。

【0182】 在另一個實施例中，核酸分子包含：

(a) 第一ITR；

(b) 組織特異性啟動子序列，例如TTP啟動子；

(c) 內含子，例如合成內含子；

(d) 編碼miRNA或治療性蛋白質（例如，凝血因子）的核苷酸；

(e) 轉錄後調控元件，例如WPRE；

(f) 3'UTR多聚(A)尾序列，例如bGHpA；和

(g) 第二ITR，

其中第一ITR或第二ITR中的一個是細小病毒科的非AAV家族成員的ITR，並且另一個ITR是AAV（例如，AAV血清型2基因組）的ITR。

【0183】 在另一個實施例中，核酸分子包含：

(a) 帶有SEQ ID NO: 111中所述的AAV2 5' ITR序列的5' ITR；

(b) 組織特異性啟動子序列，例如TTP啟動子；

(c) 內含子，例如合成內含子；

(d) 編碼FVIII（例如，FVIIIco6XTEN）的異源多核苷酸序列；

(e) 轉錄後調控元件，例如WPRE；

(f) 3'UTR多聚(A)尾序列，例如bGHpA；和/或

(g) 帶有SEQ ID NO: 124中所述的AAV2 3' ITR序列的3' ITR。

【0184】 在另一個實施例中，核酸分子包含：

(a) 帶有SEQ ID NO: 111中所述的AAV2 5' ITR序列的5' ITR；

(b) 組織特異性啟動子序列，例如CAG啟動子；

(c) 內含子，例如合成內含子；

(d) 編碼FVIII（例如，FVIIIco6XTEN）的異源多核苷酸序列；

(e) 轉錄後調控元件，例如WPRE；

(f) 3'UTR多聚(A)尾序列，例如bGHpA；和/或

(g) 帶有SEQ ID NO: 193中所述的AAV2 3' ITR序列的3' ITR。

【0185】 在另一個實施例中，核酸分子包含：

(a) 第一ITR；

(b) 組織特異性啟動子序列，TTP啟動子；

(c) 內含子，例如合成內含子；

(d) 編碼miRNA或治療性蛋白質（例如，凝血因子）的核苷酸；

(e) 轉錄後調控元件，例如WPRE；

(f) 3'UTR多聚(A)尾序列，例如bGHpA；和

(g) 第二ITR，

其中第一ITR是合成ITR，第二ITR是合成ITR，或者第一ITR和第二ITR二者都是合成ITR。

【0186】 在另一個實施例中，核酸分子包含：

(a) 第一B19 ITR；

(b) 組織特異性啟動子序列，例如TTP啟動子；

(c) 內含子，例如合成內含子；

(d) 編碼選自由下列組成之群組的治療性蛋白質的異源多核苷酸序列：凝血因子、生長因子、激素、細胞激素、抗體、其片段及其組合；

(e) 轉錄後調控元件，例如WPRE；

(f) 3'UTR多聚(A)尾序列，例如bGHpA；和/或

(g) 第二B19 ITR。

【0187】 在另一個實施例中，核酸分子包含：

(a) 第一GPV ITR；

(b) 組織特異性啟動子序列，例如TTP啟動子；

(c) 內含子，例如合成內含子；

(d) 編碼選自由下列組成之群組的治療性蛋白質的異源多核苷酸序列：凝血因子、生長因子、激素、細胞激素、抗體、其片段及其組合；

(e) 轉錄後調控元件，例如WPRE；

(f) 3'UTR多聚(A)尾序列，例如bGHpA；和/或

(g) 第二GPV ITR。

【0188】 在另一個實施例中，核酸分子包含：

(a) 第一B19 ITR；

(b) 泛在啟動子序列，例如CAG啟動子；

(c) 內含子，例如合成內含子；

(d) 編碼選自由下列組成之群組的治療性蛋白質的異源多核苷酸序列：凝血因子、生長因子、激素、細胞激素、抗體、其片段及其組合；

(e) 轉錄後調控元件，例如WPRE；

(f) 3'UTR多聚(A)尾序列，例如bGHpA；和/或

(g) 第二B19 ITR。

【0189】 在另一個實施例中，核酸分子包含：

(a) 第一GPV ITR；

(b) 泛在啟動子序列，例如CAG啟動子；

(c) 內含子，例如合成內含子；

(d) 編碼選自由下列組成之群組的治療性蛋白質的異源多核苷酸序列：

凝血因子、生長因子、激素、細胞激素、抗體、其片段及其組合；

(e) 轉錄後調控元件，例如WPRE；

(f) 3'UTR多聚(A)尾序列，例如bGHpA；和/或

(g) 第二GPV ITR。

【0190】 在另一個實施例中，核酸分子包含：

(a) 第一B19 ITR；

(b) 組織特異性啟動子序列，例如TTP啟動子；

(c) 內含子，例如合成內含子；

(d) 編碼苯丙胺酸羥化酶（PAH）的異源多核苷酸序列；

(e) 轉錄後調控元件，例如WPRE；

(f) 3'UTR多聚(A)尾序列，例如bGHpA；和/或

(g) 第二B19 ITR。

【0191】 在另一個實施例中，核酸分子包含：

(a) 第一GPV ITR；

(b) 組織特異性啟動子序列，例如TTP啟動子；

(c) 內含子，例如合成內含子；

(d) 編碼苯丙胺酸羥化酶（PAH）的異源多核苷酸序列；

(e) 轉錄後調控元件，例如WPRE；

(f) 3'UTR多聚(A)尾序列，例如bGHpA；和/或

(g) 第二GPV ITR。

【0192】 在另一個實施例中，核酸分子包含：

(a) 帶有SEQ ID NO: 180中所述的B19d135 5' ITR序列的5' ITR；

(b) 組織特異性啟動子序列，例如TTP啟動子；

(c) 內含子，例如合成內含子；

(d) 編碼FVIII（例如，FVIIIco6XTEN）的異源多核苷酸序列；

(e) 轉錄後調控元件，例如WPRE；

(f) 3'UTR多聚(A)尾序列，例如bGHpA；和/或

(g) 帶有SEQ ID NO: 181中所述的B19d135 3' ITR序列的3' ITR。

【0193】 在另一個實施例中，核酸分子包含：

(a) 帶有SEQ ID NO: 183中所述的GPVd162 5' ITR序列的5' ITR；

(b) 組織特異性啟動子序列，例如TTP啟動子；

(c) 內含子，例如合成內含子；

(d) 編碼FVIII（例如，FVIIIco6XTEN）的異源多核苷酸序列；

(e) 轉錄後調控元件，例如WPRE；

(f) 3'UTR多聚(A)尾序列，例如bGHpA；和/或

(g) 帶有SEQ ID NO: 184中所述的GPVd162 3' ITR序列的3' ITR。

【0194】 在另一個實施例中，核酸分子包含：

(a) 帶有SEQ ID NO: 185中所述的全長B19 5' ITR序列的5' ITR；

(b) 組織特異性啟動子序列，例如TTP啟動子；

(c) 內含子，例如合成內含子；

(d) 編碼FVIII（例如，FVIIIco6XTEN）的異源多核苷酸序列；

(e) 轉錄後調控元件，例如WPRE；

(f) 3'UTR多聚(A)尾序列，例如bGHpA；和/或

(g) 帶有SEQ ID NO: 186中所述的全長B19 3' ITR序列的3' ITR。

【0195】 在另一個實施例中，核酸分子包含：

(a) 帶有SEQ ID NO: 187中所述的全長GPV 5' ITR序列的5' ITR；

(b) 組織特異性啟動子序列，例如TTP啟動子；

(c) 內含子，例如合成內含子；

(d) 編碼FVIII（例如，FVIIIco6XTEN）的異源多核苷酸序列；

(e) 轉錄後調控元件，例如WPRE；

(f) 3'UTR多聚(A)尾序列，例如bGHpA；和/或

(g) 帶有SEQ ID NO: 188中所述的全長GPV 3' ITR序列的3' ITR。

【0196】 在另一個實施例中，核酸分子包含：

(a) 帶有SEQ ID NO: 180中所述的B19d135 5' ITR序列的5' ITR；

(b) 組織特異性啟動子序列，例如CAG啟動子；

(c) 內含子，例如合成內含子；

(d) 編碼PAH的異源多核苷酸序列；

(e) 轉錄後調控元件，例如WPRE；

(f) 3'UTR多聚(A)尾序列，例如bGHpA；和/或

(g) 帶有SEQ ID NO: 181中所述的B19d135 3' ITR序列的3' ITR。

【0197】 在另一個實施例中，核酸分子包含：

(a) 帶有SEQ ID NO: 183中所述的GPVd162 5' ITR序列的5' ITR；

(b) 組織特異性啟動子序列，例如CAG啟動子；

(c) 內含子，例如合成內含子；

(d) 編碼PAH的異源多核苷酸序列；

(e) 轉錄後調控元件，例如WPRE；

(f) 3'UTR多聚(A)尾序列，例如bGHpA；和/或

(g) 帶有SEQ ID NO: 184中所述的GPVd162 3' ITR序列的3' ITR。

【0198】 在另一個實施例中，核酸分子包含：

(a) 帶有SEQ ID NO: 185中所述的全長B19 5' ITR序列的5' ITR；

(b) 組織特異性啟動子序列，例如CAG啟動子；

(c) 內含子，例如合成內含子；

(d) 編碼PAH的異源多核苷酸序列；

(e) 轉錄後調控元件，例如WPRE；

(f) 3'UTR多聚(A)尾序列，例如bGHpA；和/或

(g) 帶有SEQ ID NO: 186中所述的全長B19 3' ITR序列的3' ITR。

【0199】 在另一個實施例中，核酸分子包含：

(a) 帶有SEQ ID NO: 187中所述的全長GPV 5' ITR序列的5' ITR；

(b) 組織特異性啟動子序列，例如CAG啟動子；

(c) 內含子，例如合成內含子；

(d) 編碼PAH的異源多核苷酸序列；

(e) 轉錄後調控元件，例如WPRE；

(f) 3'UTR多聚(A)尾序列，例如bGHpA；和/或

(g) 帶有SEQ ID NO: 188中所述的全長GPV 3' ITR序列的3' ITR。

A. 反向末端重複體

【0200】 本公開文本的某些態樣涉及核酸分子，其包含第一 ITR（例如，5' ITR）和第二 ITR（例如，3' ITR）。通常，ITR 參與細小病毒（例如，

AAV)從原核質體進行 DNA 複製和拯救或切除(Samulski 等人, 1983, 1987；Senapathy 等人, 1984；Gottlieb 和 Muzyczka, 1988）。另外，ITR 似乎是 AAV 原病毒整合和將 AAV DNA 包裝至病毒粒子中所需的最小序列(McLaughlin 等人, 1988；Samulski 等人, 1989）。這些元件是細小病毒基因組的有效操縱必需的。假設，對於 ITR 功能必不可少的最小限定元件是 Rep 結合位點（例如，RBS；對於 AAV2，GCGCGCTCGCTCGCTC（SEQ ID NO: 104））和末端解鏈位點（例如，TRS；對於 AAV2，AGTTGG（SEQ ID NO: 105））加上允許髮夾形成的可變回文序列。回文核苷酸區域通常作為 DNA 複製的起點以及作為病毒的包裝信號以順式一起作用。ITR 中的互補序列在 DNA 複製期間折疊成髮夾結構。在一些實施例中，ITR 折疊成髮夾 T 形結構。在其他實施例中，ITR 折疊成非 T 形髮夾結構，例如折疊成 U 形髮夾結構。數據表明，AAV ITR 的 T 形髮夾結構可以抑制由 ITR 側接的轉基因的表現。參見例如，Zhou 等人, *Scientific Reports* 7:5432（2017 年 7 月 14 日）。通過利用不形成 T 形髮夾結構的 ITR，可以避免這種形式的抑制。因此，在某些態樣中，與包含形成 T 形髮夾的 AAV ITR 的多核苷酸相比，包含非 AAV ITR 的多核苷酸具有改良的轉基因表現。

【0201】 在一些實施例中，ITR 包含天然存在的 ITR，例如，ITR 包含細小病毒 ITR 的全部或一部分。在一些實施例中，ITR 包含合成序列。在一個實施例中，第一 ITR 或第二 ITR 包含合成序列。在另一個實施例中，第一 ITR 和第二 ITR 各自包含合成序列。在一些實施例中，第一 ITR 或第二 ITR 包含天然存在的序列。在另一個實施例中，第一 ITR 和第二 ITR 各自包含天然存在的序列。

【0202】 在一些實施例中，ITR 包含天然存在的 ITR 的一部分（例如，截短的 ITR）或由所述部分組成。在一些實施例中，ITR 包含天然存在的 ITR 的片段或由所述片段組成，其中所述片段包含至少約 5 個核苷酸、至少約 10 個核苷酸、至少約 15 個核苷酸、至少約 20 個核苷酸、至少約 25 個核苷酸、至少約 30 個核苷酸、至少約 35 個核苷酸、至少約 40 個核苷酸、至少約 45 個核苷酸、至少約 50 個核苷酸、至少約 55 個核苷酸、至少約 60 個核苷酸、至少約 65 個核苷酸、至少約 70 個核苷酸、至少約 75 個核苷酸、

至少約 80 個核苷酸、至少約 85 個核苷酸、至少約 90 個核苷酸、至少約 95 個核苷酸、至少約 100 個核苷酸、至少約 125 個核苷酸、至少約 150 個核苷酸、至少約 175 個核苷酸、至少約 200 個核苷酸、至少約 225 個核苷酸、至少約 250 個核苷酸、至少約 275 個核苷酸、至少約 300 個核苷酸、至少約 325 個核苷酸、至少約 350 個核苷酸、至少約 375 個核苷酸、至少約 400 個核苷酸、至少約 425 個核苷酸、至少約 450 個核苷酸、至少約 475 個核苷酸、至少約 500 個核苷酸、至少約 525 個核苷酸、至少約 550 個核苷酸、至少約 575 個核苷酸或至少約 600 個核苷酸；其中 ITR 保留天然存在的 ITR 的功能特性。在某些實施例中，ITR 包含天然存在的 ITR 的片段或由所述片段組成，其中所述片段包含至少約 129 個核苷酸；其中 ITR 保留天然存在的 ITR 的功能特性。在某些實施例中，ITR 包含天然存在的 ITR 的片段或由所述片段組成，其中所述片段包含至少約 102 個核苷酸；其中 ITR 保留天然存在的 ITR 的功能特性。

【0203】 在一些實施例中，ITR 包含天然存在的 ITR 的一部分或由所述部分組成，其中所述片段包含天然存在的 ITR 的長度的至少約 5%、至少約 10%、至少約 15%、至少約 20%、至少約 25%、至少約 30%、至少約 35%、至少約 40%、至少約 45%、至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%；其中所述片段保留天然存在的 ITR 的功能特性。

【0204】 在某些實施例中，ITR 包含如下序列或由如下序列組成，所述序列在正確比對時與天然存在的 ITR 的同源部分具有至少 50%、至少 51%、至少 52%、至少 53%、至少 54%、至少 55%、至少 56%、至少 57%、至少 58%、至少 59%、至少 60%、至少 61%、至少 62%、至少 63%、至少 64%、至少 65%、至少 66%、至少 67%、至少 68%、至少 69%、至少 70%、至少 71%、至少 72%、至少 73%、至少 74%、至少 75%、至少 76%、至少 77%、至少 78%、至少 79%、至少 80%、至少 81%、至少 82%、至少 83%、至少 84%、至少 85%、至少 86%、至少 87%、至少 88%、至少 89%、至少 90%、至少 91%、至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、

至少 97%、至少 98%、至少 99%或 100%的序列同一性；其中 ITR 保留天然存在的 ITR 的功能特性。在其他實施例中，ITR 包含如下序列或由如下序列組成，所述序列在正確比對時與天然存在的 ITR 的同源部分具有至少 90%的序列同一性；其中 ITR 保留天然存在的 ITR 的功能特性。在一些實施例中，ITR 包含如下序列或由如下序列組成，所述序列在正確比對時與天然存在的 ITR 的同源部分具有至少 80%的序列同一性；其中 ITR 保留天然存在的 ITR 的功能特性。在一些實施例中，ITR 包含如下序列或由如下序列組成，所述序列在正確比對時與天然存在的 ITR 的同源部分具有至少 70%的序列同一性；其中 ITR 保留天然存在的 ITR 的功能特性。在一些實施例中，ITR 包含如下序列或由如下序列組成，所述序列在正確比對時與天然存在的 ITR 的同源部分具有至少 60%的序列同一性；其中 ITR 保留天然存在的 ITR 的功能特性。在一些實施例中，ITR 包含如下序列或由如下序列組成，所述序列在正確比對時與天然存在的 ITR 的同源部分具有至少 50%的序列同一性；其中 ITR 保留天然存在的 ITR 的功能特性。

【0205】 在一些實施例中，ITR 包含來自 AAV 基因組的 ITR。在一些實施例中，ITR 是選自由下列組成之群組的 AAV 基因組的 ITR：AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV11 及其任何組合。在一個特定實施例中，ITR 是 AAV2 基因組的 ITR。在另一個實施例中，ITR 是經遺傳工程化以在其 5´和 3´端包括衍生自一個或多個 AAV 基因組的 ITR 的合成序列。

【0206】 在一些實施例中，ITR 並非衍生自 AAV 基因組。在一些實施例中，ITR 是非 AAV 的 ITR。在一些實施例中，ITR 是來自選自但不限於以下的病毒科細小病毒科的非 AAV 基因組的 ITR：博卡病毒屬、依賴病毒屬、紅病毒屬、阿留申病毒屬、細小病毒屬、濃核病毒屬、重複病毒屬、康特拉病毒屬、禽細小病毒屬、反芻類細小病毒屬、原細小病毒屬、四型細小病毒屬、雙義濃核病毒屬、短濃核病毒屬、肝胰濃核病毒屬、對蝦濃核病毒屬及其任何組合。在某些實施例中，ITR 衍生自紅病毒屬細小病毒 B19（人病毒）。在另一個實施例中，ITR 衍生自番鴨細小病毒（MDPV）株。在某些實施例中，MDPV 株被減弱，例如 MDPV 株 FZ91-30。在其他

實施例中，MDPV株是病原性的，例如MDPV株YY。在一些實施例中，ITR衍生自豬細小病毒，例如豬細小病毒U44978。在一些實施例中，ITR衍生自小鼠微小病毒，例如小鼠微小病毒U34256。在一些實施例中，ITR衍生自犬細小病毒，例如犬細小病毒M19296。在一些實施例中，ITR衍生自水貂腸炎病毒，例如水貂腸炎病毒D00765。在一些實施例中，ITR衍生自依賴細小病毒（*Dependoparvovirus*）。在一個實施例中，依賴細小病毒是依賴病毒屬鵝細小病毒（GPV）株。在一個具體實施例中，GPV株被減弱，例如GPV株82-0321V。在另一個具體實施例中，GPV株是病原性的，例如GPV株B。

【0207】 核酸分子的第一ITR和第二ITR可以衍生自相同基因組(例如，衍生自相同病毒的基因組），或衍生自不同基因組（例如，衍生自兩種或更多種不同病毒基因組的基因組）。在某些實施例中，第一ITR和第二ITR衍生自相同的AAV基因組。在一個具體實施例中，存在於本發明的核酸分子中的這兩個ITR是相同的，並且特別地可以是AAV2 ITR。在其他實施例中，第一ITR衍生自AAV基因組並且第二ITR並非衍生自AAV基因組（例如，非AAV基因組）。在其他實施例中，第一ITR並非衍生自AAV基因組（例如，非AAV基因組）並且第二ITR衍生自AAV基因組。在仍其他實施例中，第一ITR和第二ITR二者都並非衍生自AAV基因組(例如，非AAV基因組）。在一個特定實施例中，第一ITR和第二ITR是相同的。

【0208】 在一些實施例中，第一ITR衍生自AAV基因組，並且第二ITR衍生自選自由下列組成之群組的基因組：博卡病毒屬、依賴病毒屬、紅病毒屬、阿留申病毒屬、細小病毒屬、濃核病毒屬、重複病毒屬、康特拉病毒屬、禽細小病毒屬、反芻類細小病毒屬、原細小病毒屬、四型細小病毒屬、雙義濃核病毒屬、短濃核病毒屬、肝胰濃核病毒屬、對蝦濃核病毒屬及其任何組合。在其他實施例中，第二ITR衍生自AAV基因組，並且第一ITR衍生自選自由下列組成之群組的基因組：博卡病毒屬、依賴病毒屬、紅病毒屬、阿留申病毒屬、細小病毒屬、濃核病毒屬、重複病毒屬、康特拉病毒屬、禽細小病毒屬、反芻類細小病毒屬、原細小病毒屬、四型細小

病毒屬、雙義濃核病毒屬、短濃核病毒屬、肝胰濃核病毒屬、對蝦濃核病毒屬及其任何組合。在其他實施例中，第一 ITR 和第二 ITR 衍生自選自由下列組成之群組的基因組：博卡病毒屬、依賴病毒屬、紅病毒屬、阿留申病毒屬、細小病毒屬、濃核病毒屬、重複病毒屬、康特拉病毒屬、禽細小病毒屬、反芻類細小病毒屬、原細小病毒屬、四型細小病毒屬、雙義濃核病毒屬、短濃核病毒屬、肝胰濃核病毒屬、對蝦濃核病毒屬及其任何組合，其中第一 ITR 和第二 ITR 衍生自相同的基因組。在其他實施例中，第一 ITR 和第二 ITR 衍生自選自由下列組成之群組的基因組：博卡病毒屬、依賴病毒屬、紅病毒屬、阿留申病毒屬、細小病毒屬、濃核病毒屬、重複病毒屬、康特拉病毒屬、禽細小病毒屬、反芻類細小病毒屬、原細小病毒屬、四型細小病毒屬、雙義濃核病毒屬、短濃核病毒屬、肝胰濃核病毒屬、對蝦濃核病毒屬及其任何組合，其中第一 ITR 和第二 ITR 衍生自不同的基因組。

【0209】 在一些實施例中，第一 ITR 衍生自 AAV 基因組，並且第二 ITR 衍生自紅病毒屬細小病毒 B19（人病毒）。在其他實施例中，第二 ITR 衍生自 AAV 基因組，並且第一 ITR 衍生自紅病毒屬細小病毒 B19（人病毒）。

【0210】 在某些實施例中，第一 ITR 和/或第二 ITR 包含衍生自 B19 的 ITR 的全部或一部分，或由所述 ITR 的全部或一部分組成。在一些實施例中，第一 ITR 和/或第二 ITR 包含如下核苷酸序列或由如下核苷酸序列組成，所述核苷酸序列與選自 SEQ ID NO: 167、168、169、170 和 171 的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同，其中第一 ITR 和/或第二 ITR 保留衍生出它的 B19 ITR 的功能特性。在一些實施例中，第一 ITR 和/或第二 ITR 包含如下核苷酸序列或由如下核苷酸序列組成，所述核苷酸序列與選自 SEQ ID NO: 167、168、169、170 和 171 的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%

或 100%相同，其中第一 ITR 和/或第二 ITR 能夠形成髮夾結構。在某些實施例中，髮夾結構不包含 T 形髮夾。

【0211】 在一些實施例中，第一 ITR 和/或第二 ITR 包含選自 SEQ ID NO: 167、168、169、170 和 171 的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 和/或第二 ITR 包含 SEQ ID NO: 167 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 和/或第二 ITR 包含 SEQ ID NO: 168 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 和/或第二 ITR 包含 SEQ ID NO: 169 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 和/或第二 ITR 包含 SEQ ID NO: 170 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 和/或第二 ITR 包含 SEQ ID NO: 171 中所述的核苷酸序列或由所述核苷酸序列組成。

表1.樣本細小病毒ITR序列。

細小病毒	ITR ID	描述	長度（nt）	序列
B19	wt	基因庫：KY940273.1	383	CCAAATCAGATGCCGCCGGTCGCCGCCGGT AGGCGGGACTTCCGGTACAAGATGGCGGA CAATTACGTCATTTCCTGTGACGTCATTTCC TGTGACGTCACTTCCGGTGGGCGGGACTTC CGGAATTAGGGTTGGCTCTGGGCCAGCTTG CTTGGGGTTGCCTTGACACTAAGACAAGCG GCGCGCCGCTTGATCTTAGTGGCACGTCAA CCCCAAGCGCTGGCCCAGAGCCAACCCTA ATTCCGGAAGTCCCGCCCACCGGAAGTGA CGTCACAGGAAATGACGTCACAGGAAATG ACGTAATTGTCCGCCATCTTGTACCGGAAG TCCCGCCTACCGGCGGCGACCGGCGGCATC

				TGATTTGGTGTCTTCTTTTAAATTTT（SEQ ID NO: 167）
	d135	不包括前135個核苷酸	248	CTCTGGGCCAGCTTGCTTGGGGTTGCCTTGACACTAAGACAAGCGGCGCGCCGCTTGATCTTAGTGGCACGTCAACCCCAAGCGCTGGCCCAGAGCCAACCCTAATTCCGGAAGTCCCGCCCACCGGAAGTGACGTCACAGGAAATGACGTCACAGGAAATGACGTAATTGTCCGCCATCTTGTACCGGAAGTCCCGCCTACCGGCGGCGACCGGCGGCATCTGATTTGGTGTCTTCTTTTAAATTTT（SEQ ID NO: 168）
	v1	基於與AAV2的比較的最小序列	129	CGGCGCGCCGCTTGATCTTAGTGGCACGTCAACCAGCGCTGGCCCAGAGCCAACCCTAATTCCGGAAGTCCTCAGTCCGCCATCTTGCCCGCCTACCGGCGGCGACCGGCGGCATCATTTGGTGTTCTT（SEQ ID NO: 169）
	v2	不包括前135個核苷酸以及回文序列中相應的互補的135個核苷酸	113	CTCTGGGCCAGCTTGCTTGGGGTTGCCTTGACACTAAGACAAGCGGCGCGCCGCTTGATCTTAGTGGCACGTCAACCCCAAGCGCTGGCCCAGAGTGTCTTCTTTTAAATTTT（SEQ ID NO: 170）
	v3	基於與GPV的比較的最	340	CAAATCAGATGCCGCCGGTCGCCGCCGGTAGGCGGGACTTCCGGTACAAGATGGCGGA

		小序列		CAATTACGTCATTTCCTGTGACGTATTTCCT GTGACGTACTTCCGGTGGCGGGACTTCCGG AATTTTGGCTCTGGGCCAGCTTGCTTGGGG TTGCCTTGACCAAGCGCGCGCCGCTTGATC ACCCCAAGCGCTGGCCCAGAGCCACCTAA CCGGAAGTCCCCCCACCGGAAGTGACGTC ACAGGAAAGACGTCACAGGAAGTAATTGT CCGCCATCTTGTACCGGAAGTCCCGCACCG GCGGCGACCGGCGGCATCTGATTTGGTGTC TTCTTTTAAATTTT（**SEQ ID NO: 171**）
GPV	wt	基因庫： U25749.1	444	CTCATTGGAGGGTTCGTTCGTTCGAACCAG CCAATCAGGGGAGGGGGAAGTGACGCAAG TTCCGGTCACATGCTTCCGGTGACGCACAT CCGGTGACGTAGTTCCGGTCACGTGCTTCC TGTCACGTGTTTCCGGTCACGTGACTTCCG GTCATGTGACTTCCGGTGACGTGTTTCCGG CTGTTAGGTTGACCACGCGCATGCCGCGCG GTCAGCCCAATAGTTAAGCCGGAAACACG TCACCGGAAGTCACATGACCGGAAGTCAC GTGACCGGAAACACGTGACAGGAAGCACG TGACCGGAACTACGTCACCGGATGTGCGTC ACCGGAAGCATGTGACCGGAACTTGCGTC ACTTCCCCCTCCCCTGATTGGCTGGTTCGA ACGAACGAACCCTCCAATGAGACTCAAGG ACAAGAGGATATTTTGCGCGCCAGGAAGT

				G（**SEQ ID NO: 172**）
	d162	不包括前162個核苷酸	282	CGGTGACGTGTTTCCGGCTGTTAGGTTGAC CACGCGCATGCCGCGCGGTCAGCCCAATA GTTAAGCCGGAAACACGTCACCGGAAGTC ACATGACCGGAAGTCACGTGACCGGAAAC ACGTGACAGGAAGCACGTGACCGGAACTA CGTCACCGGATGTGCGTCACCGGAAGCAT GTGACCGGAACTTGCGTCACTTCCCCCTCC CCTGATTGGCTGGTTCGAACGAACGAACCC TCCAATGAGACTCAAGGACAAGAGGATAT TTTGCGCGCCAGGAAGTG（**SEQ ID NO: 173**）
	v1	基於與AAV2的比較的最小序列	145	TTGACCACGCGCATGCCGCGCGGTCAGCCC AATAGTTAAGCCGGGTGACCACACGTGAC AGGAAGCACGGGATGTGCGTCACCGGAAG CAGTGACCGGGCTGGTTCGAACGAACGAA CCCTCCAACTCAAGGACAAGAGGATATT （**SEQ ID NO: 174**）
	v2	不包括前162個核苷酸以及回文序列中相應的互補的162個核苷酸	120	CGGTGACGTGTTTCCGGCTGTTAGGTTGAC CACGCGCATGCCGCGCGGTCAGCCCAATA GTTAAGCCGGAAACACGTCACCGACTCAA GGACAAGAGGATATTTTGCGCGCCAGGAA GTG（**SEQ ID NO: 175**）
	v3	基於與B19	102	GGGAACAATCAGGGGAAGTGACCGGTGAC

		的比較的最小序列		GTCATGTAACTTGCGTCACTTCCCGTTCGA ACGAACGAACGAGACTCAAGGACAAGAGG CGCGCCAGGAAGTG（**SEQ ID NO: 176**）
AAV2	wt	基因庫：NC_001401.2	145	TTGGCCACTCCCTCTCTGCGCGCTCGCTCG CTCACTGAGGCCGGGCGACCAAAGGTCGC CCGACGCCCGGGCTTTGCCCGGGCGGCCTC AGTGAGCGAGCGAGCGCGCAGAGAGGGAG TGGCCAACTCCATCACTAGGGGTTCCT （**SEQ ID NO: 177**）
	GTx	用於GTx載體中	130	CTGCGCGCTCGCTCGCTCACTGAGGCCGCC CGGGCAAAGCCCGGGCGTCGGGCGACCTT TGGTCGCCCGGCCTCAGTGAGCGAGCGAG CGCGCAGAGAGGGAGTGGCCAACTCCATC ACTAGGGGTTCCT（**SEQ ID NO: 178**）

【0212】 在某些實施例中，第一 ITR 和/或第二 ITR 包含與 SEQ ID NO: 167 中所述的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同的核苷酸序列。在某些實施例中，第一 ITR 和/或第二 ITR 由 SEQ ID NO: 167 組成。在某些實施例中，第一 ITR 和/或第二 ITR 包含與 SEQ ID NO: 168 中所述的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同的核苷酸序列。在某些實施例中，第一 ITR 和/或第二 ITR 由 SEQ ID NO: 168 組成。在某些實施例中，第一 ITR 和/或第二 ITR 包含核苷酸序列，其中所述核苷酸序列包含 SEQ ID NO: 169

中所述的最小核苷酸序列，並且其中所述核苷酸序列與 SEQ ID NO: 167 中所述的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同，保留衍生出它的 B19 ITR 的功能特性。在一些實施例中，第一 ITR 和/或第二 ITR 包含核苷酸序列，其中所述核苷酸序列包含 SEQ ID NO: 169 中所述的最小核苷酸序列，並且其中所述核苷酸序列與 SEQ ID NO: 167 中所述的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同，其中第一 ITR 和/或第二 ITR 能夠形成髮夾結構。在某些實施例中，髮夾結構不包含 T 形髮夾。

【0213】 在某些實施例中，第一 ITR 和/或第二 ITR 包含衍生自 B19 的 ITR 的全部或一部分，或由所述 ITR 的全部或一部分組成。在一些實施例中，第二 ITR 是第一 ITR 的反向補體。在一些實施例中，第一 ITR 是第二 ITR 的反向補體。在一些實施例中，第一 ITR 和/或第二 ITR 包含如下核苷酸序列或其功能衍生物，或由如下核苷酸序列或其功能衍生物組成，所述核苷酸序列與選自 SEQ ID NO: 180、181、185 和 186 的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同。在一些實施例中，功能衍生物保留衍生出它的 B19 ITR 的功能特性。在一些實施例中，第一 ITR 和/或第二 ITR 包含如下核苷酸序列或其功能衍生物，或由如下核苷酸序列或其功能衍生物組成，所述核苷酸序列與選自 SEQ ID NO: 180、181、185 和 186 的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同。在一些實施例中，功能衍生物能夠形成髮夾結構。在某些實施例中，髮夾結構不包含 T 形髮夾。

【0214】 在某些實施例中，第一 ITR 和/或第二 ITR 包含與 SEQ ID NO: 180 中所述的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同的核苷酸序列。在某些實施例中，第一 ITR 和/或第二 ITR 由 SEQ ID NO: 180 組成。在某些實施例中，第一 ITR 和/或第二 ITR 包含與 SEQ ID NO: 181 中所述的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同的核苷酸序列。在某些實施例中，第一 ITR 和/或第二 ITR 由 SEQ ID NO: 181 組成。在某些實施例中，第一 ITR 和/或第二 ITR 包含與 SEQ ID NO: 185 中所述的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同的核苷酸序列。在某些實施例中，第一 ITR 和/或第二 ITR 由 SEQ ID NO: 185 組成。在某些實施例中，第一 ITR 和/或第二 ITR 包含與 SEQ ID NO: 186 中所述的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同的核苷酸序列。在某些實施例中，第一 ITR 和/或第二 ITR 由 SEQ ID NO: 186 組成。

【0215】 在一些實施例中，第一 ITR 和/或第二 ITR 包含選自 SEQ ID NO: 180、181、185 和 186 的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 包含 SEQ ID NO: 180 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 包含 SEQ ID NO: 181 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 包含 SEQ ID NO: 185 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 包含 SEQ ID NO: 186 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第二 ITR 包含 SEQ ID NO: 180 中所述的核苷酸

序列或由所述核苷酸序列組成。在一些實施例中，第二 ITR 包含 SEQ ID NO: 181 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第二 ITR 包含 SEQ ID NO: 185 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第二 ITR 包含 SEQ ID NO: 186 中所述的核苷酸序列或由所述核苷酸序列組成。

【0216】 在一些實施例中，第一 ITR 包含 SEQ ID NO:180 中所述的核苷酸序列或由所述核苷酸序列組成，並且第二 ITR 包含 SEQ ID NO: 181 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 包含 SEQ ID NO:181 中所述的核苷酸序列或由所述核苷酸序列組成，並且第二 ITR 包含 SEQ ID NO: 180 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 包含 SEQ ID NO:185 中所述的核苷酸序列或由所述核苷酸序列組成，並且第二 ITR 包含 SEQ ID NO: 186 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 包含 SEQ ID NO:186 中所述的核苷酸序列或由所述核苷酸序列組成，並且第二 ITR 包含 SEQ ID NO: 185 中所述的核苷酸序列或由所述核苷酸序列組成。

【0217】 在一些實施例中，第一 ITR 衍生自 AAV 基因組，並且第二 ITR 衍生自 GPV。在其他實施例中，第二 ITR 衍生自 AAV 基因組，並且第一 ITR 衍生自 GPV。

【0218】 在某些實施例中，第一 ITR 和/或第二 ITR 包含與 SEQ ID NO: 172 中所述的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同的核苷酸序列。在某些實施例中，第一 ITR 和/或第二 ITR 由 SEQ ID NO: 172 組成。在某些實施例中，第一 ITR 和/或第二 ITR 包含與 SEQ ID NO: 173 中所述的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同的核苷酸序列。在某些實施例中，第一 ITR 和/或第二 ITR 由 SEQ ID NO: 173 組成。在某些實施例中，第一 ITR

和/或第二 ITR 包含衍生自 GPV 的 ITR 的全部或一部分，或由所述 ITR 的全部或一部分組成。在一些實施例中，第一 ITR 和/或第二 ITR 包含如下核苷酸序列或由如下核苷酸序列組成，所述核苷酸序列與選自 SEQ ID NO: 172、173、174、175 和 176 的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同，其中第一 ITR 和/或第二 ITR 保留衍生出它的 GPV ITR 的功能特性。在一些實施例中，第一 ITR 和/或第二 ITR 包含衍生自 GPV 的 ITR 的全部或一部分，或由所述 ITR 的全部或一部分組成。在一些實施例中，第一 ITR 和/或第二 ITR 包含如下核苷酸序列或由如下核苷酸序列組成，所述核苷酸序列與選自 SEQ ID NO: 172、173、174、175 和 176 的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同，其中第一 ITR 和/或第二 ITR 能夠形成髮夾結構。在某些實施例中，髮夾結構不包含 T 形髮夾。在一些實施例中，第一 ITR 和/或第二 ITR 包含選自 SEQ ID NO: 172、173、174、175 和 176 的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 和/或第二 ITR 包含 SEQ ID NO: 172 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 和/或第二 ITR 包含 SEQ ID NO: 173 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 和/或第二 ITR 包含 SEQ ID NO: 174 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 和/或第二 ITR 包含 SEQ ID NO: 175 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 和/或第二 ITR 包含 SEQ ID NO: 176 中所述的核苷酸序列或由所述核苷酸序列組成。

【0219】 在某些實施例中，第一 ITR 和/或第二 ITR 包含核苷酸序列，其中所述核苷酸序列包含 SEQ ID NO: 174 中所述的最小核苷酸序列，並且其中所述核苷酸序列與 SEQ ID NO: 172 中所述的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少

約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同，其中第一 ITR 和/或第二 ITR 保留衍生出它的 GPV ITR 的功能特性。在一些實施例中，第一 ITR 和/或第二 ITR 包含核苷酸序列，其中所述核苷酸序列包含 SEQ ID NO: 174 中所述的最小核苷酸序列，並且其中所述核苷酸序列與 SEQ ID NO: 172 中所述的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同，其中第一 ITR 和/或第二 ITR 能夠形成髮夾結構。在某些實施例中，髮夾結構不包含 T 形髮夾。

【0220】 在某些實施例中，第一 ITR 和/或第二 ITR 包含核苷酸序列，其中所述核苷酸序列包含 SEQ ID NO: 176 中所述的最小核苷酸序列，並且其中所述核苷酸序列與 SEQ ID NO: 172 中所述的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同，其中第一 ITR 和/或第二 ITR 保留衍生出它的 GPV ITR 的功能特性。在一些實施例中，第一 ITR 和/或第二 ITR 包含核苷酸序列，其中所述核苷酸序列包含 SEQ ID NO: 176 中所述的最小核苷酸序列，並且其中所述核苷酸序列與 SEQ ID NO: 172 中所述的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同，其中第一 ITR 和/或第二 ITR 能夠形成髮夾結構。在某些實施例中，髮夾結構不包含 T 形髮夾。

【0221】 在某些實施例中，第一 ITR 和/或第二 ITR 包含衍生自 GPV 的 ITR 的全部或一部分，或由所述 ITR 的全部或一部分組成。在一些實施例中，第二 ITR 是第一 ITR 的反向補體。在一些實施例中，第一 ITR 是第二 ITR 的反向補體。在一些實施例中，第一 ITR 和/或第二 ITR 包含如下核苷酸序列或其功能衍生物，或由如下核苷酸序列或其功能衍生物組成，所

述核苷酸序列與選自 SEQ ID NO: 183、184、187 和 188 的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同。在一些實施例中，功能衍生物保留衍生出它的 GPV ITR 的功能特性。在一些實施例中，第一 ITR 和/或第二 ITR 包含如下核苷酸序列或其功能衍生物，或由如下核苷酸序列或其功能衍生物組成，所述核苷酸序列與選自 SEQ ID NO: 183、184、187 和 188 的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同。在一些實施例中，功能衍生物能夠形成髮夾結構。在某些實施例中，髮夾結構不包含 T 形髮夾。

【0222】 在某些實施例中，第一 ITR 和/或第二 ITR 包含與 SEQ ID NO: 183 中所述的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同的核苷酸序列。在某些實施例中，第一 ITR 和/或第二 ITR 由 SEQ ID NO: 183 組成。在某些實施例中，第一 ITR 和/或第二 ITR 包含與 SEQ ID NO: 184 中所述的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同的核苷酸序列。在某些實施例中，第一 ITR 和/或第二 ITR 由 SEQ ID NO: 184 組成。在某些實施例中，第一 ITR 和/或第二 ITR 包含與 SEQ ID NO: 187 中所述的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同的核苷酸序列。在某些實施例中，第一 ITR 和/或第二 ITR 由 SEQ ID NO: 187 組成。在某些實施例中，第一 ITR 和/或第二 ITR 包含與 SEQ ID NO: 188 中所述的核苷酸序列至

少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同的核苷酸序列。在某些實施例中，第一 ITR 和/或第二 ITR 由 SEQ ID NO: 188 組成。

【0223】 在一些實施例中，第一 ITR 和/或第二 ITR 包含選自 SEQ ID NO: 183、184、187 和 188 的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 包含 SEQ ID NO: 183 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 包含 SEQ ID NO: 184 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 包含 SEQ ID NO: 187 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 包含 SEQ ID NO: 188 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第二 ITR 包含 SEQ ID NO: 183 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第二 ITR 包含 SEQ ID NO: 184 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第二 ITR 包含 SEQ ID NO: 187 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第二 ITR 包含 SEQ ID NO: 188 中所述的核苷酸序列或由所述核苷酸序列組成。

【0224】 在一些實施例中，第一 ITR 包含 SEQ ID NO:183 中所述的核苷酸序列或由所述核苷酸序列組成，並且第二 ITR 包含 SEQ ID NO: 184 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 包含 SEQ ID NO:184 中所述的核苷酸序列或由所述核苷酸序列組成，並且第二 ITR 包含 SEQ ID NO: 183 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 包含 SEQ ID NO:187 中所述的核苷酸序列或由所述核苷酸序列組成，並且第二 ITR 包含 SEQ ID NO: 188 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 包含 SEQ ID NO:188 中所述的核苷酸序列或由所述核苷酸序列組成，並且第二 ITR 包含 SEQ ID NO: 187 中所述的核苷酸序列或由所述核苷酸序列組成。

【0225】 在某些實施例中，第一 ITR 或第二 ITR 中的一個包含衍生自 AAV2 的 ITR 的全部或一部分或由所述 ITR 的全部或一部分組成。在一

些實施例中，第一 ITR 或第二 ITR 包含如下核苷酸序列或由如下核苷酸序列組成，所述核苷酸序列與 SEQ ID NO: 177 或 178 中所述的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同，其中第一 ITR 和/或第二 ITR 保留衍生出它的 AAV2 ITR 的功能特性。在一些實施例中，第一 ITR 或第二 ITR 包含如下核苷酸序列或由如下核苷酸序列組成，所述核苷酸序列與 SEQ ID NO: 177 或 178 中所述的核苷酸序列至少約 50%、至少約 55%、至少約 60%、至少約 65%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同，其中第一 ITR 和/或第二 ITR 能夠形成髮夾結構。在某些實施例中，髮夾結構不包含 T 形髮夾。在一些實施例中，第一 ITR 和/或第二 ITR 包含 SEQ ID NO: 177 或 178 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 和/或第二 ITR 包含 SEQ ID NO: 177 中所述的核苷酸序列或由所述核苷酸序列組成。在一些實施例中，第一 ITR 和/或第二 ITR 包含 SEQ ID NO: 178 中所述的核苷酸序列或由所述核苷酸序列組成。

【0226】 在一些實施例中，第一 ITR 衍生自 AAV 基因組，並且第二 ITR 衍生自番鴨細小病毒（MDPV）株。在其他實施例中，第二 ITR 衍生自 AAV 基因組，並且第一 ITR 衍生自番鴨細小病毒（MDPV）株。在某些實施例中，MDPV 株被減弱，例如 MDPV 株 FZ91-30。在其他實施例中，MDPV 株是病原性的，例如 MDPV 株 YY。

【0227】 在一些實施例中，第一 ITR 衍生自 AAV 基因組，並且第二 ITR 衍生自依賴細小病毒。在一些實施例中，第二 ITR 衍生自 AAV 基因組，並且第一 ITR 衍生自依賴細小病毒。在其他實施例中，第一 ITR 衍生自 AAV 基因組，並且第二 ITR 衍生自依賴病毒屬鵝細小病毒（GPV）株。在其他實施例中，第二 ITR 衍生自 AAV 基因組，並且第一 ITR 衍生自依賴病毒屬 GPV 株。在某些實施例中，GPV 株被減弱，例如 GPV 株 82-0321V。在其他實施例中，GPV 株是病原性的，例如 GPV 株 B。

【0228】 在某些實施例中，第一 ITR 衍生自 AAV 基因組，並且第二 ITR 衍生自選自由下列組成之群組的基因組：豬細小病毒，例如豬細小病毒株 U44978；小鼠微小病毒，例如小鼠微小病毒株 U34256；犬細小病毒，例如犬細小病毒株 M19296；水貂腸炎病毒，例如水貂腸炎病毒株 D00765；及其任何組合。在其他實施例中，第二 ITR 衍生自 AAV 基因組，並且第一 ITR 衍生自選自由下列組成之群組的基因組：豬細小病毒，例如豬細小病毒株 U44978；小鼠微小病毒，例如小鼠微小病毒株 U34256；犬細小病毒，例如犬細小病毒株 M19296；水貂腸炎病毒，例如水貂腸炎病毒株 D00765；及其任何組合。

【0229】 在另一個特定實施例中，ITR 是經遺傳工程化以在其 5´和 3´端包括並非衍生自 AAV 基因組的 ITR 的合成序列。在另一個特定實施例中，ITR 是經遺傳工程化以在其 5´和 3´端包括衍生自一個或多個非 AAV 基因組的 ITR 的合成序列。存在於本發明的核酸分子中的這兩個 ITR 可以是相同或不同的非 AAV 基因組。具體地，ITR 可以衍生自相同的非 AAV 基因組。在一個具體實施例中，存在於本發明的核酸分子中的這兩個 ITR 是相同的，並且特別地可以是 AAV2 ITR。

【0230】 在一些實施例中，ITR 序列包含一個或多個回文序列。本文所公開的 ITR 的回文序列包括但不限於天然回文序列（即，在自然界中發現的序列）、合成序列（即，未在自然界中發現的序列）（如偽回文序列）及其組合或經修飾形式。“偽回文序列”是包括不完美的回文序列在內的回文 DNA 序列，其與形成二級結構的天然 AAV 或非 AAV 回文序列中的序列共用小於 80%（包括小於 70%、60%、50%、40%、30%、20%、10%或 5%或無）核酸序列同一性。天然回文序列可以從本文所公開的任何基因組獲得或衍生。合成回文序列可以基於本文所公開的任何基因組。

【0231】 回文序列可以是連續或中斷的。在一些實施例中，回文序列是中斷的，其中回文序列包含第二序列的插入。在一些實施例中，第二序列包含啟動子、強化子、整合酶的整合位點（例如，Cre 或 Flp 重組酶的位點）、基因產物的開放閱讀框或其組合。

【0232】 在一些實施例中，ITR 形成髮夾環結構。在一個實施例中，第一 ITR 形成髮夾結構。在另一個實施例中，第二 ITR 形成髮夾結構。仍然在另一個實施例中，第一 ITR 和第二 ITR 二者都形成髮夾結構。在一些實施例中，第一 ITR 和/或第二 ITR 不形成 T 形髮夾結構。在某些實施例中，第一 ITR 和/或第二 ITR 形成非 T 形髮夾結構。在一些實施例中，非 T 形髮夾結構包含 U 形髮夾結構。

【0233】 在一些實施例中，本文所述的核酸分子中的 ITR 可以是轉錄啟動的 ITR。轉錄啟動的 ITR 可以包含已經通過包括至少一個轉錄活性元件而被轉錄啟動的野生型 ITR 的全部或一部分。多種類型的轉錄活性元件適用於這個背景。在一些實施例中，轉錄活性元件是組成型轉錄活性元件。組成型轉錄活性元件提供持續水準的基因轉錄，並且在需要持續表現轉基因時是較佳的。在其他實施例中，轉錄活性元件是誘導型轉錄活性元件。誘導型轉錄活性元件通常在誘導物（或誘導條件）不存在下展現低活性，並且在誘導物存在下（或切換至誘導條件）上調。當僅在某些時間或在某些位置需要表現時或當需要使用誘導劑逐步增加表現水準時，誘導型轉錄活性元件可能是較佳的。轉錄活性元件也可以具有組織特異性；也就是說，其僅在某些組織或細胞類型中展現活性。

【0234】 能以多種方式將轉錄活性元件併入 ITR 中。在一些實施例中，將轉錄活性元件併入 ITR 的任何部分的 5´ 或 ITR 的任何部分的 3´。在其他實施例中，轉錄啟動的 ITR 的轉錄活性元件位於兩個 ITR 序列之間。如果轉錄活性元件包含兩個或更多個必須被間隔開的元件，則那些元件可以與 ITR 的部分交替。在一些實施例中，ITR 的髮夾結構缺失並用轉錄元件的反向重複來替代。這後一種排列將產生模擬結構中的缺失部分的髮夾。多個串聯轉錄活性元件也可以存在於轉錄啟動的 ITR 中，並且這些元件可以相鄰或間隔開。另外，可以將蛋白質結合位點（例如，Rep 結合位點）引入轉錄啟動的 ITR 的轉錄活性元件中。轉錄活性元件可以包含使得能夠通過 RNA 聚合酶受控轉錄 DNA 以形成 RNA 的任何序列，並且可以包含例如如下文所定義的轉錄活性元件。

【0235】 轉錄啟動的 ITR 向具有相對受限的核苷酸序列長度的核酸分子提供轉錄啟動和 ITR 功能二者，這有效地使可以攜帶並從核酸分子表現的轉基因的長度最大化。將轉錄活性元件併入 ITR 中能以多種方式來完成。對 ITR 序列和轉錄活性元件的序列需求的比較可以提供對編碼 ITR 內的元件的方式的瞭解。例如，可以通過在複製轉錄活性元件的功能元件的 ITR 序列中引入特定變化將轉錄活性添加至 ITR。業內存在多種技術可有效添加、缺失和/或改變特定位點的具體核苷酸序列（參見例如，Deng 和 Nickoloff (1992) Anal. Biochem. 200:81-88）。產生轉錄啟動的 ITR 的另一方式涉及在 ITR 中的所需位置引入限制位點。另外，可以使用業內已知的方法將多個轉錄啟動元件併入轉錄啟動的 ITR 中。

【0236】 通過說明，可以藉由包括一個或多個諸如以下等轉錄活性元件來產生轉錄啟動的 ITR：TATA 盒、GC 盒、CCAAT 盒、Sp1 位點、Inr 區、CRE（cAMP 調控元件）位點、ATF-1/CRE 位點、APBβ 盒、APBα 盒、CArG 盒、CCAC 盒或如業內已知參與轉錄的任何其他元件。

【0237】 本公開文本的多個態樣提供了選殖本文所述的核酸分子的方法，其包括將能夠形成複雜二級結構的核酸分子插入適宜載體中，和將所得載體引入適宜細菌宿主菌株中。如業內已知，核酸的複雜二級結構（例如，長回文區）可能不穩定並且難以在細菌宿主菌株中選殖。例如，本公開文本的包含第一 ITR 和第二 ITR（例如，非 AAV 細小病毒 ITR，例如 B19 或 GPV ITR）的核酸分子可能難以使用常規方法來選殖。長 DNA 回文序列抑制 DNA 複製並且在大腸桿菌（*E. coli*）、芽孢桿菌屬（*Bacillus*）、鏈球菌屬（*Steptococcus*）、鏈黴菌屬（*Streptomyces*）、釀酒酵母（*S. cerevisiae*）、小鼠和人的基因組中不穩定。這些效應是由於通過鏈內鹼基配對形成髮夾或十字形結構所致。在大腸桿菌中，在 SbcC 或 SbcD 突變體中可以顯著克服對 DNA 複製的抑制。SbcD 是核酸酶亞基，並且 SbcC 是 SbcCD 複合物的 ATP 酶亞基。大腸桿菌 SbcCD 複合物是負責防止長回文序列複製的外切核酸酶複合物。SbcCD 複合物是具有 ATP 依賴性雙鏈 DNA 外切核酸酶活性和 ATP 非依賴性單鏈 DNA 內切核酸酶活性的核複合物。SbcCD 可以識別 DNA 回文序列並通過攻擊所產生的髮夾結構來瓦解複製叉。

【0238】 在某些實施例中，適宜細菌宿主菌株不能拆分十字形 DNA 結構。在某些實施例中，適宜細菌宿主菌株包含 SbcCD 複合物中的破壞。在一些實施例中，SbcCD 複合物中的破壞包含 SbcC 基因和/或 SbcD 基因中的基因破壞。在某些實施例中，SbcCD 複合物中的破壞包含 SbcC 基因中的基因破壞。包含 SbcC 基因中的基因破壞的各種細菌宿主菌株是業內已知的。例如但不限於，細菌宿主菌株 PMC103 包含基因型 *sbcC*、*recD*、*mcrA*、*ΔmcrBCF*；細菌宿主菌株 PMC107 包含基因型 *recBC*、*recJ*、*sbcBC*、*mcrA*、*ΔmcrBCF*；並且細菌宿主菌株 SURE 包含基因型 *recB*、*recJ*、*sbcC*、*mcrA*、*ΔmcrBCF*、*umuC*、*uvrC*。因此，在一些實施例中，選殖本文所述的核酸分子的方法包括將能夠形成複雜二級結構的核酸分子插入適宜載體中，和將所得載體引入主體菌株 PMC103、PMC107 或 SURE 中。在某些實施例中，選殖本文所述的核酸分子的方法包括將能夠形成複雜二級結構的核酸分子插入適宜載體中，和將所得載體引入主體菌株 PMC103 中。

【0239】 適宜載體是業內已知的並且描述于本文中的其他地方。在某些實施例中，用於本公開文本的選殖方法中的適宜載體是低拷貝載體。在某些實施例中，用於本公開文本的選殖方法中的適宜載體是 pBR322。

【0240】 因此，本公開文本提供了選殖核酸分子的方法，其包括將能夠形成複雜二級結構的核酸分子插入適宜載體中，和將所得載體引入包含 SbcCD 複合物中的破壞的細菌宿主菌株中，其中核酸分子包含第一反向末端重複體(ITR)和第二 ITR，其中第一 ITR 和/或第二 ITR 包含與 SEQ ID NO: 180、181、183、184、185、186、187 或 188 中所述的核苷酸序列至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同的核苷酸序列或其功能衍生物。

B. 治療性蛋白質

【0241】 本公開文本的某些態樣涉及包含第一 ITR、第二 ITR 和編碼靶序列的基因盒的核酸分子，其中靶序列編碼治療性蛋白質。在一些實施例中，基因盒編碼一種治療性蛋白質。在一些實施例中，基因盒編碼多於一種治療性蛋白質。在一些實施例中，基因盒編碼相同治療性蛋白質的兩

個或更多個拷貝。在一些實施例中，基因盒編碼相同治療性蛋白質的兩種或更多種變異體。在一些實施例中，基因盒編碼兩種或更多種不同的治療性蛋白質。

【0242】 本公開文本的某些實施例涉及包含第一 ITR、第二 ITR 和編碼治療性蛋白質的基因盒的核酸分子，其中治療性蛋白質包含凝血因子。在一些實施例中，凝血因子選自 FI、FII、FIII、FIV、FV、FVI、FVII、FVIII、FIX、FX、FXI、FXII、FXIII、VWF、前微血管增滲素、高分子量激肽原、纖連蛋白、抗凝血酶 III、肝素輔因子 II、蛋白質 C、蛋白質 S、蛋白質 Z、蛋白質 Z 相關蛋白酶抑制劑（ZPI）、纖維蛋白溶酶原、α2-抗纖維蛋白溶酶、組織纖維蛋白溶酶原活化劑（tPA）、尿激酶、纖維蛋白溶酶原活化物抑制劑-1（PAI-1）、纖維蛋白溶酶原活化物抑制劑-2（PAI2）、其任何酶原、其任何活性形式及其任何組合。在一個實施例中，凝血因子包含 FVIII 或其變異體或片段。在另一個實施例中，凝血因子包含 FIX 或其變異體或片段。在另一個實施例中，凝血因子包含 FVII 或其變異體或片段。在另一個實施例中，凝血因子包含 VWF 或其變異體或片段。

1. 凝血因子

【0243】 在一些實施例中，核酸分子包含第一 ITR、第二 ITR 和編碼靶序列的基因盒，其中靶序列編碼治療性蛋白質，其中治療性蛋白質包含 VIII 因子多肽。除非另外指定，否則如本文所用的「VIII 因子」在整個本申請中縮寫為「FVIII」，意指具有其在凝結中的正常作用的功能性 FVIII 多肽。因此，術語 FVIII 包括功能性變異體多肽。「FVIII 蛋白」可與 FVIII 多肽（或蛋白質）或 FVIII 互換使用。FVIII 功能的例子包括但不限於啟動凝結的能力、作為 IX 因子的輔因子作用的能力或在 Ca^{2+}和磷脂存在下與 IX 因子形成因子 X 酶（tenase）複合物的能力，所述複合物然後將 X 因子轉化為活化形式 Xa。FVIII 蛋白可以是人、豬、犬、大鼠或鼠類 FVIII 蛋白。另外，來自人和其他物種的 FVIII 之間的比較已經鑑定出可能為功能所需的保守殘基（Cameron 等人, *Thromb. Haemost.* 79:317-22 (1998)；US 6,251,632）。全長多肽和多核苷酸序列是已知的，許多功能片段、突變體和經修飾形式也是已知的。多個 FVIII 胺基酸和核苷酸序列披露於例如以下文獻中：美國公

開號 2015/0158929 A1、2014/0308280 A1 和 2014/0370035 A1 以及國際公開號 WO 2015/106052 A1。FVIII 多肽包括例如全長 FVIII、在 N 末端缺少 Met 的全長 FVIII、成熟 FVIII（缺少信號序列）、在 N 末端具有額外 Met 的成熟 FVIII 和/或完全或部分缺失 B 結構域的 FVIII。FVIII 變異體包括部分或完全缺失的 B 結構域缺失。

a. FVIII和編碼FVIII蛋白的多核苷酸序列

【0244】 在一些實施例中，核酸分子包含第一 ITR、第二 ITR 和編碼靶序列的基因盒，其中靶序列編碼治療性蛋白質，其中治療性蛋白質包含 VIII 因子多肽。除非另外指定，否則如本文所用的「VIII 因子」在整個本申請中縮寫為「FVIII」，意指具有其在凝結中的正常作用的功能性 FVIII 多肽。因此，術語 FVIII 包括功能性變異體多肽。「FVIII 蛋白」可與 FVIII 多肽（或蛋白質）或 FVIII 互換使用。FVIII 功能的例子包括但不限於啟動凝結的能力、作為 IX 因子的輔因子作用的能力或在 Ca^{2+}和磷脂存在下與 IX 因子形成因子 X 酶複合物的能力，所述複合物然後將 X 因子轉化為活化形式 Xa。FVIII 蛋白可以是人、豬、犬、大鼠或鼠類 FVIII 蛋白。另外，來自人和其他物種的 FVIII 之間的比較已經鑑定出可能為功能所需的保守殘基（Cameron 等人, *Thromb. Haemost.* 79:317-22 (1998)；US 6,251,632）。全長多肽和多核苷酸序列是已知的，許多功能片段、突變體和經修飾形式也是已知的。多個 FVIII 胺基酸和核苷酸序列披露於例如以下文獻中：美國公開號 2015/0158929 A1、2014/0308280 A1 和 2014/0370035 A1 以及國際公開號 WO 2015/106052 A1。FVIII 多肽包括例如全長 FVIII、在 N 末端缺少 Met 的全長 FVIII、成熟 FVIII（缺少信號序列）、在 N 末端具有額外 Met 的成熟 FVIII 和/或完全或部分缺失 B 結構域的 FVIII。FVIII 變異體包括部分或完全缺失的 B 結構域缺失。

【0245】 本文所用的嵌合蛋白中的 FVIII 部分具有 FVIII 活性。FVIII 活性可以通過業內任何已知方法來量測。多種測試可用於評估凝結系統的功能：啟動部分促凝血酶原激酶時間（aPTT）測試、產色測定、ROTEM 測定、凝血酶原時間（PT）測試（也用於確定 INR）、纖維蛋白原測試（通常通過克勞斯法進行）、血小板計數、血小板功能測試（通常通過 PFA-100

進行）、TCT、出血時間、混合測試（如果將患者的血漿與正常血漿混合，是否矯正異常）、凝結因子測定、抗磷脂抗體、D-二聚體、遺傳測試（例如，V因子 Leiden、凝血酶原突變 G20210A）、稀釋魯塞爾蝰蛇毒時間（dRVVT）、雜項血小板功能測試、凝血彈性描記法（TEG 或 Sonoclot）、凝血彈性量測法（TEM®，例如 ROTEM®）或優球蛋白溶解時間（ELT）。

【0246】 aPTT 測試是量測「固有」（也稱為接觸啟動路徑）和常見凝結路徑二者的功效的表現指示物。這個測試一般用於量測市售重組凝血因子（例如，FVIII）的凝血活性。其與量測外在路徑的凝血酶原時間（PT）結合使用。

【0247】 ROTEM 分析提供關於止血的整體動力學的信息：凝血時間、凝塊形成、凝塊穩定性和溶解。凝血彈性量測法中的不同參數依賴於血漿凝結系統的活性、血小板功能、纖維蛋白溶解或影響這些相互作用的許多因素。這個分析可以提供次級止血的全面見解。

【0248】 產色測定機制是基於血液凝結級聯的原理，其中在啟動的IX因子、磷脂和鈣離子存在下，啟動的 FVIII 加速 X 因子轉化為 X 因子 a。X 因子 a 活性是通過水解對 X 因子 a 具有特異性的對硝基苯胺（pNA）底物來評估。在 405 nM 下量測的對硝基苯胺的初始釋放速率與 X 因子 a 活性成正比，並且因此與樣本中的 FVIII 活性成正比。

【0249】 產色測定得到了國際血栓形成與止血學會（ISTH）的科學標準化委員會（SSC）的 FVIII 和 IX 因子小組委員會的推薦。從 1994 以來，產色測定也已經是歐洲藥典中用於指定 FVIII 濃縮物效力的參考方法。因此，在一個實施例中，包含 FVIII 的嵌合多肽具有與包含成熟 FVIII 或 BDD FVIII 的嵌合多肽（例如，ADVATE®、REFACTO®或 ELOCTATE®）相當的 FVIII 活性。

【0250】 在另一個實施例中，本公開文本的包含 FVIII 的嵌合蛋白的 X 因子 a 產生速率與包含成熟 FVIII 或 BDD FVIII 的嵌合蛋白（例如，ADVATE®、REFACTO®或 ELOCTATE®）相當。

【0251】 為了將 X 因子啟動為 X 因子 a，啟動的 IX 因子（IX 因子 a）在 Ca^{2+}、膜磷脂和 FVIII 輔因子存在下水解 X 因子中的一個精胺酸-異白胺

酸鍵以形成 X 因子 a。因此，FVIII 與 IX 因子的相互作用在凝結路徑中具有關鍵作用。在某些實施例中，包含 FVIII 的嵌合多肽能以與包含成熟 FVIII 序列或 BDD FVIII 的嵌合多肽（例如，ADVATE®、REFACTO®或 ELOCTATE®）相當的速率與 IX 因子 a 相互作用。

【0252】 另外，FVIII 結合至血管性血友病因子，同時在循環中無活性。FVIII 在不結合至 VWF 時快速降解，並且通過凝血酶的作用從 VWF 釋放。在一些實施例中，包含 FVIII 的嵌合多肽以與包含成熟 FVIII 序列或 BDD FVIII 的嵌合多肽（例如，ADVATE®、REFACTO®或 ELOCTATE®）相當的水準結合至血管性血友病因子。

【0253】 可以在鈣和磷脂存在下通過啟動的蛋白質 C 使 FVIII 失活。啟動的蛋白質 C 在 A1 結構域中的精胺酸 336 之後切割 FVIII 重鏈，從而破壞 X 因子底物相互作用位點，並在 A2 結構域中的精胺酸 562 之後切割，從而增強 A2 結構域的解離並且破壞與 IX 因子 a 的相互作用位點。這種切割還對切 A2 結構域（43 kDa）並產生 A2-N（18 kDa）和 A2-C（25 kDa）結構域。因此，啟動的蛋白質 C 可以催化重鏈中的多個切割位點。在一個實施例中，通過啟動的蛋白質 C 以與包含成熟 FVIII 序列或 BDD FVIII 的嵌合多肽（例如，ADVATE®、REFACTO®或 ELOCTATE®）相當的水準使包含 FVIII 的嵌合多肽失活。

【0254】 在其他實施例中，包含 FVIII 的嵌合蛋白具有與包含成熟 FVIII 序列或 BDD FVIII 的嵌合多肽（例如，ADVATE®、REFACTO®或 ELOCTATE®）相當的體內 FVIII 活性。在一個特定實施例中，在 HemA 小鼠尾靜脈橫斷模型中，包含 FVIII 的嵌合多肽能夠以與包含成熟 FVIII 序列或 BDD FVIII 的嵌合多肽（例如，ADVATE®、REFACTO®或 ELOCTATE®）相當的水準保護 HemA 小鼠。

【0255】 如本文所用的 FVIII 的「B 結構域」與業內已知的 B 結構域相同，所述 B 結構域是通過成熟人類 FVIII 的內部胺基酸序列同一性和凝血酶的蛋白質水解切割位點（例如，殘基 Ser741-Arg1648）來定義。其他人類 FVIII 結構域是相對於成熟人類 FVIII，通過以下胺基酸殘基來定義：成熟 FVIII 的 A1，殘基 Ala1-Arg372；A2，殘基 Ser373-Arg740；A3，殘基

Ser1690-Ile2032；C1，殘基 Arg2033-Asn2172；C2，殘基 Ser2173-Tyr2332。除非另外指示，否則在不提及任何 SEQ ID 編號的情況下，本文中使用的序列殘基編號對應於不含信號肽序列（19 個胺基酸）的 FVIII 序列。A3-C1-C2 序列（也稱為 FVIII 重鏈）包括殘基 Ser1690-Tyr2332。其餘序列（殘基 Glu1649-Arg1689）通常被稱為 FVIII 輕鏈啟動肽。豬、小鼠和犬 FVIII 的所有結構域（包括 B 結構域）的邊界的位置也是業內已知的。在一個實施例中，FVIII 的 B 結構域缺失（「B 結構域缺失的 FVIII」或「BDD FVIII」）。BDD FVIII 的例子是 REFACTO®（重組 BDD FVIII）。在一個特定實施例中，B 結構域缺失的 FVIII 變異體包含成熟 FVIII 的胺基酸殘基 746 至 1648 的缺失。

【0256】 「B 結構域缺失的 FVIII」可以具有以下文獻中披露的完全或部分缺失：美國專利號 6,316,226、6,346,513、7,041,635、5,789,203、6,060,447、5,595,886、6,228,620、5,972,885、6,048,720、5,543,502、5,610,278、5,171,844、5,112,950、4,868,112 和 6,458,563 以及國際公開號 WO 2015106052 A1（PCT/US2015/010738）。在一些實施例中，用於本公開文本的方法中的 B 結構域缺失的 FVIII 序列包含美國專利號 6,316,226（也在 US 6,346,513 中）的第 4 欄第 4 行至第 5 欄第 28 行和實例 1-5 中披露的缺失中的任何一個。在另一個實施例中，B 結構域缺失的 VIII 因子是 S743/Q1638 B 結構域缺失的 VIII 因子（SQ BDD FVIII）（例如，具有胺基酸 744 至胺基酸 1637 的缺失的 VIII 因子，例如具有成熟 FVIII 的胺基酸 1-743 和胺基酸 1638-2332 的 VIII 因子）。在一些實施例中，用於本公開文本的方法中的 B 結構域缺失的 FVIII 具有美國專利號 5,789,203（還有 US 6,060,447、US 5,595,886 和 US 6,228,620）的第 2 欄第 26-51 行和實例 5-8 中披露的缺失。在一些實施例中，B 結構域缺失的 VIII 因子具有以下文獻中所述的缺失：美國專利號 5,972,885 的第 1 欄第 25 行至第 2 欄第 40 行；美國專利號 6,048,720 的第 6 欄第 1-22 行和實例 1；美國專利號 5,543,502 的第 2 欄第 17-46 行；美國專利號 5,171,844 的第 4 欄第 22 行至第 5 欄第 36 行；美國專利號 5,112,950 的第 2 欄第 55-68 行、圖 2 和實例 1；美國專利號 4,868,112 的第 2 欄第 2 行至第 19 欄第 21 行和表 2；美國專利號 7,041,635 的第 2 欄第 1 行至第 3 欄第

19行、第3欄第40行至第4欄第67行、第7欄第43行至第8欄第26行和第11欄第5行至第13欄第39行;或美國專利號6,458,563的第4欄第25-53行。在一些實施例中，B結構域缺失的FVIII具有大部分B結構域的缺失，但是仍含有將初級轉譯產物在體內蛋白質水解加工為兩條多肽鏈所必需的B結構域的胺基末端序列，如WO 91/09122中所披露。在一些實施例中，B結構域缺失的FVIII構建為具有胺基酸747-1638的缺失，即實際上完全缺失B結構域。Hoeben R.C.等人 *J. Biol. Chem.* 265 (13): 7318-7323 (1990)。B結構域缺失的VIII因子還可以含有FVIII的胺基酸771–1666或胺基酸868-1562的缺失。Meulien P.等人 *Protein Eng.* 2(4): 301-6 (1988)。作為本發明的一部分的其他B結構域缺失包括：胺基酸982至1562或760至1639（Toole等人, *Proc. Natl. Acad. Sci. U.S.A.* (1986) 83, 5939-5942）、797至1562（Eaton,等人 *Biochemistry* (1986) 25:8343-8347）、741至1646（Kaufman（PCT公開申請號WO 87/04187））、747-1560（Sarver等人, *DNA* (1987) 6:553-564）、741至1648（Pasek（PCT申請號88/00831））或者816至1598或741至1648（Lagner（Behring Inst. Mitt. (1988)第82期:16-25、EP 295597））的缺失。在一個特定實施例中，B結構域缺失的FVIII包含成熟FVIII的胺基酸殘基746至1648的缺失。在另一個實施例中，B結構域缺失的FVIII包含成熟FVIII的胺基酸殘基745至1648的缺失。在一些實施例中，BDD FVIII包含含有對應于成熟全長FVIII的胺基酸765至1652中的缺失的單鏈FVIII（也稱為rVIII-SingleChain和AFSTYLA®）。參見美國專利號7,041,635。

【0257】 在其他實施例中，BDD FVIII包括含有B結構域的如下片段的FVIII多肽，所述片段保留一個或多個N連接糖基化位點，例如對應于全長FVIII序列的胺基酸序列的殘基757、784、828、900、963或任選地943。B結構域片段的例子包括B結構域的226個胺基酸或163個胺基酸，如Miao, H.Z.等人, *Blood* 103(a): 3412-3419 (2004)；Kasuda, A等人, *J. Thromb. Haemost.* 6: 1352-1359 (2008)和 Pipe, S.W.等人, *J. Thromb. Haemost.* 9: 2235-2242 (2011)中所披露（即，保留B結構域的前226個胺基酸或163個胺基酸）。在仍其他實施例中，BDD FVIII還包含在殘基309處的點突變（從Phe至Ser）以改良BDD FVIII蛋白的表現。參見Miao, H.Z.等人, Blood 103(a):

3412-3419 (2004)。在仍其他實施例中，BDD FVIII 包括含有 B 結構域的一部分，但是不含一個或多個弗林蛋白酶切割位點（例如，Arg1313 和 Arg 1648）的 FVIII 多肽。參見 Pipe, S.W.等人, *J. Thromb. Haemost.* 9: 2235-2242 (2011)。在一些實施例中，BDD FVIII 包含含有對應于成熟全長 FVIII 的胺基酸 765 至 1652 中的缺失的單鏈 FVIII（也稱為 rVIII-SingleChain 和 AFSTYLA®）。參見美國專利號 7,041,635。可以在任何 FVIII 序列中產生前述每一種缺失。

【0258】 已知眾多種功能 FVIII 變異體，如上文和下文所討論。另外，已經在血友病患者體內鑑定出 FVIII 中的數百種非功能性突變，並且已經確定，這些突變對 FVIII 功能的影響更多歸因於其在 FVIII 的三維結構內的位置而不是取代的性質（Cutler 等人, *Hum. Mutat. 19*:274-8 (2002)），將所述文獻通過引用以其整體併入本文。另外，來自人和其他物種的 FVIII 之間的比較已經鑑定出可能為功能所需的保守殘基（Cameron 等人, Thromb. Haemost. 79:317-22 (1998)；US 6,251,632），將所述文獻通過引用以其整體併入本文。

【0259】 在一些實施例中，FVIII 多肽包含其 FVIII 變異體或片段，其中其 FVIII 變異體或片段具有 FVIII 活性。在一些實施例中，基因盒編碼全長 FVIII 多肽。在其他實施例中，基因盒編碼 B 結構域缺失的（BDD）FVIII 多肽，其中缺失 FVIII 的 B 結構域的全部或一部分。在一個特定實施例中，基因盒編碼包含與 SEQ ID NO: 106、107、109、110、111 或 112 具有至少約 80%、至少約 85%、至少約 86%、至少約 87%、至少約 88%、至少約 89%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性的胺基酸序列的多肽。在一些實施例中，基因盒編碼具有 SEQ ID NO: 17 的胺基酸序列的多肽或其片段。在一些實施例中，基因盒編碼具有 SEQ ID NO: 106 的胺基酸序列的多肽或其片段。在一些實施例中，基因盒包含與 SEQ ID NO: 107 具有至少 85%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性的核苷酸序列。在一些實施例中，基因盒編碼具有 SEQ ID NO: 109 的胺基酸序列的多肽或其片段。在一些實施例中，基因盒包含與 SEQ ID NO: 16 具有至少

85%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性的核苷酸序列。在一些實施例中，基因盒包含與 SEQ ID NO: 109 具有至少 85%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性的核苷酸序列。

【0260】 在一些實施例中，本公開文本的基因盒編碼包含信號肽的 FVIII 多肽或其片段。在其他實施例中，基因盒編碼缺少信號肽的 FVIII 多肽。在一些實施例中，信號肽包含 SEQ ID NO: 17 的胺基酸 1-19。

【0261】 在一些實施例中，基因盒包含編碼 FVIII 多肽的核苷酸序列，其中核苷酸序列經密碼子優化。在某些實施例中，基因盒包含國際申請號 PCT/US2017/015879 中披露的核苷酸序列，將所述申請通過引用以其整體併入本文。在一些實施例中，基因盒包含編碼 FVIII 多肽的核苷酸序列，其中核苷酸序列經密碼子優化。在某些實施例中，基因盒包含與選自 SEQ ID NO: 1-14 的核苷酸序列具有至少 85%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性的核苷酸序列。在一些實施例中，基因盒包含與 SEQ ID NO: 71 具有至少 85%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性的核苷酸序列。在一些實施例中，基因盒包含與 SEQ ID NO: 19 具有至少 85%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性的核苷酸序列。

i. 編碼FVIII多肽的密碼子優化的核苷酸序列

【0262】 在一些實施例中，本公開文本的核酸分子包含第一 ITR、第二 ITR 和編碼靶序列的基因盒，其中靶序列編碼治療性蛋白質，其中第一 ITR 和第二 ITR 衍生自 AAV 基因組，並且其中基因盒包含編碼 FVIII 多肽的密碼子優化的核苷酸序列。在一些實施例中，密碼子優化的核苷酸序列編碼全長 FVIII 多肽。在其他實施例中，密碼子優化的核苷酸序列編碼 B 結構域缺失的（BDD）FVIII 多肽，其中缺失 FVIII 的 B 結構域的全部或一部分。在一個特定實施例中，密碼子優化的核苷酸序列編碼包含與 SEQ ID NO: 17 具有至少約 80%、至少約 85%、至少約 86%、至少約 87%、至少約 88%、至少約 89%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、

至少約 98%或至少約 99%序列同一性的胺基酸序列的多肽或其片段。在一個實施例中，密碼子優化的核苷酸序列編碼具有 SEQ ID NO: 17 的胺基酸序列的多肽或其片段。

【0263】 在一些實施例中，密碼子優化的核苷酸序列編碼包含信號肽的 FVIII 多肽或其片段。在其他實施例中，密碼子優化的序列編碼缺少信號肽的 FVIII 多肽。在一些實施例中，信號肽包含 SEQ ID NO: 17 的胺基酸 1-19。

【0264】 在一些實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含如下核苷酸序列，其包含編碼 FVIII 多肽的 N 末端部分的第一核酸序列和編碼 FVIII 多肽的 C 末端部分的第二核酸序列；其中第一核酸序列與 (i) SEQ ID NO: 3 的核苷酸 58-1791 或 (ii) SEQ ID NO: 4 的核苷酸 58-1791 具有至少 85%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性；並且其中 N 末端部分和 C 末端部分一起具有 FVIII 多肽活性。在一個特定實施例中，第一核酸序列與 SEQ ID NO: 3 的核苷酸 58-1791 具有至少 85%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性。在另一個實施例中，第一核酸序列與 SEQ ID NO: 4 的核苷酸 58-1791 具有至少 85%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性。在其他實施例中，第一核苷酸序列包含 SEQ ID NO: 3 的核苷酸 58-1791 或 SEQ ID NO: 4 的核苷酸 58-1791。

【0265】 在其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含如下核苷酸序列，其包含編碼 FVIII 多肽的 N 末端部分的第一核酸序列和編碼 FVIII 多肽的 C 末端部分的第二核酸序列；其中第一核酸序列與 (i) SEQ ID NO: 3 的核苷酸 1-1791 或 (ii) SEQ ID NO: 4 的核苷酸 1-1791 具有至少 85%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性；並且其中 N 末端部分和 C 末端部分一起具有 FVIII 多肽活性。在一個實施例中，第一核苷酸序列包含 SEQ ID NO: 3 的核苷酸 1-1791 或 SEQ ID NO: 4 的核苷酸 1-1791。在另一個實施例中，第二核苷酸序列與 SEQ ID NO: 3 的核苷酸 1792-4374 或 SEQ ID NO: 4 的核苷酸 1792-4374 具

有至少 60%、至少 70%、至少 80%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性。在一個特定實施例中，第二核苷酸序列包含 SEQ ID NO: 3 的核苷酸 1792-4374 或 SEQ ID NO: 4 的核苷酸 1792-4374。在仍另一個實施例中，第二核苷酸序列與 SEQ ID NO: 3 的核苷酸 1792-2277 和 2320-4374 或 SEQ ID NO: 4 的核苷酸 1792-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 3 的核苷酸 1792-4374 或 SEQ ID NO: 4 的核苷酸 1792-4374）具有至少 60%、至少 70%、至少 80%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性。在一個特定實施例中，第二核苷酸序列包含 SEQ ID NO: 3 的核苷酸 1792-2277 和 2320-4374 或 SEQ ID NO: 4 的核苷酸 1792-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 3 的核苷酸 1792-4374 或 SEQ ID NO: 4 的核苷酸 1792-4374）。

【0266】 在一些實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含如下核苷酸序列，其包含編碼 FVIII 多肽的 N 末端部分的第一核酸序列和編碼 FVIII 多肽的 C 末端部分的第二核酸序列；其中第二核酸序列與 (i) SEQ ID NO: 5 的核苷酸 1792-4374 或 (ii) SEQ ID NO: 6 的核苷酸 1792-4374 具有至少 85%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性；並且其中 N 末端部分和 C 末端部分一起具有 FVIII 多肽活性。在某些實施例中，第二核酸序列與 SEQ ID NO: 5 的核苷酸 1792-4374 具有至少 85%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性。在其他實施例中，第二核酸序列與 SEQ ID NO: 6 的核苷酸 1792-4374 具有至少 85%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性。在一個特定實施例中，第二核酸序列包含 SEQ ID NO: 5 的核苷酸 1792-4374 或 SEQ ID NO: 6 的核苷酸 1792-4374。在一些實施例中，上文所列示的與第二核酸序列連接的第一核酸序列與 SEQ ID NO: 5 的核苷酸 58-1791 或 SEQ ID NO: 6 的核苷酸 58-1791 具有至少 60%、至少 70%、至少 80%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性。在其他實施例中，上文所列示的與第二核酸序列連接的第一核酸序列與 SEQ ID NO: 5 的核苷

酸 1-1791 或 SEQ ID NO: 6 的核苷酸 1-1791 具有至少 60%、至少 70%、至少 80%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性。

【0267】 在其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含如下核苷酸序列，其包含編碼 FVIII 多肽的 N 末端部分的第一核酸序列和編碼 FVIII 多肽的 C 末端部分的第二核酸序列；其中第二核酸序列與 (i) SEQ ID NO: 5 的核苷酸 1792-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 5 的核苷酸 1792-4374）或 (ii) SEQ ID NO: 6 的核苷酸 1792-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 6 的核苷酸 1792-4374）具有至少 85%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性；並且其中 N 末端部分和 C 末端部分一起具有 FVIII 多肽活性。在某些實施例中，第二核酸序列與 SEQ ID NO: 5 的核苷酸 1792-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 5 的核苷酸 1792-4374）具有至少 85%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性。在其他實施例中，第二核酸序列與 SEQ ID NO: 6 的核苷酸 1792-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 6 的核苷酸 1792-4374）具有至少 85%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性。在一個特定實施例中，第二核酸序列包含 SEQ ID NO: 5 的核苷酸 1792-2277 和 2320-4374 或 SEQ ID NO: 6 的核苷酸 1792-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 5 的核苷酸 1792-4374 或 SEQ ID NO: 6 的核苷酸 1792-4374）。在一些實施例中，上文所列示的與第二核酸序列連接的第一核酸序列與 SEQ ID NO: 5 的核苷酸 58-1791 或 SEQ ID NO: 6 的核苷酸 58-1791 具有至少 60%、至少 70%、至少 80%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性。在其他實施例中，上文所列示的與第二核酸序列連接的第一核酸序列與 SEQ ID NO: 5 的核苷酸 1-1791 或 SEQ ID NO: 6 的

核苷酸 1-1791 具有至少 60%、至少 70%、至少 80%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性。

【0268】 在一些實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含如下核苷酸序列，其包含編碼 FVIII 多肽的 N 末端部分的第一核酸序列和編碼 FVIII 多肽的 C 末端部分的第二核酸序列；其中第一核酸序列與 (i) SEQ ID NO: 1 的核苷酸 58-1791、(ii) SEQ ID NO: 2 的核苷酸 58-1791、(iii) SEQ ID NO: 70 的核苷酸 58-1791 或 (iv) SEQ ID NO: 71 的核苷酸 58-1791 具有至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性；並且其中 N 末端部分和 C 末端部分一起具有 FVIII 多肽活性。在其他實施例中，第一核苷酸序列包含 SEQ ID NO: 1 的核苷酸 58-1791、SEQ ID NO: 2 的核苷酸 58-1791、(iii) SEQ ID NO: 70 的核苷酸 58-1791 或 (iv) SEQ ID NO: 71 的核苷酸 58-1791。

【0269】 在其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含如下核苷酸序列，其包含編碼 FVIII 多肽的 N 末端部分的第一核酸序列和編碼 FVIII 多肽的 C 末端部分的第二核酸序列；其中第一核酸序列與 (i) SEQ ID NO: 1 的核苷酸 1-1791、(ii) SEQ ID NO: 2 的核苷酸 1-1791、(iii) SEQ ID NO: 70 的核苷酸 1-1791 或 (iv) SEQ ID NO: 71 的核苷酸 1-1791 具有至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性；並且其中 N 末端部分和 C 末端部分一起具有 FVIII 多肽活性。在一個實施例中，第一核苷酸序列包含 SEQ ID NO: 1 的核苷酸 1-1791、SEQ ID NO: 2 的核苷酸 1-1791、(iii) SEQ ID NO: 70 的核苷酸 1-1791 或 (iv) SEQ ID NO: 71 的核苷酸 1-1791。在另一個實施例中，與第一核苷酸序列連接的第二核苷酸序列與 SEQ ID NO: 1 的核苷酸 1792-4374、SEQ ID NO: 2 的核苷酸 1792-4374、(iii) SEQ ID NO: 70 的核苷酸 1792-4374 或 (iv) SEQ ID NO: 71 的核苷酸 1792-4374 具有至少 60%、至少 70%、至少 80%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性。在一個特定實施例中，與第一核苷酸序列連接的第二核苷酸序列包含 (i) SEQ ID NO: 1 的核苷酸 1792-4374、(ii) SEQ ID NO: 2 的核苷酸 1792-4374、(iii) SEQ ID NO: 70 的核苷酸 1792-4374 或 (iv) SEQ ID NO: 71 的核苷酸 1792-4374。

在其他實施例中，與第一核苷酸序列連接的第二核苷酸序列與 (i) SEQ ID NO: 1 的核苷酸 1792-2277 和 2320-4374、(ii) SEQ ID NO: 2 的核苷酸 1792-2277 和 2320-4374、(iii) SEQ ID NO: 70 的核苷酸 1792-2277 和 2320-4374 或 (iv) SEQ ID NO: 71 的核苷酸 1792-2277 和 2320-4374 具有至少 60%、至少 70%、至少 80%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性。在一個實施例中，第二核苷酸序列包含 (i) SEQ ID NO: 1 的核苷酸 1792-2277 和 2320-4374、(ii) SEQ ID NO: 2 的核苷酸 1792-2277 和 2320-4374、(iii) SEQ ID NO: 70 的核苷酸 1792-2277 和 2320-4374 或 (iv) SEQ ID NO: 71 的核苷酸 1792-2277 和 2320-4374。

【0270】 在另一個實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含如下核苷酸序列，其包含編碼 FVIII 多肽的 N 末端部分的第一核酸序列和編碼 FVIII 多肽的 C 末端部分的第二核酸序列；其中第二核酸序列與 (i) SEQ ID NO: 1 的核苷酸 1792-4374、(ii) SEQ ID NO: 2 的核苷酸 1792-4374、(iii) SEQ ID NO: 70 的核苷酸 1792-4374 或 (iv) SEQ ID NO: 71 的核苷酸 1792-4374 具有至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性；並且其中 N 末端部分和 C 末端部分一起具有 FVIII 多肽活性。在一個特定實施例中，第二核酸序列包含 (i) SEQ ID NO: 1 的核苷酸 1792-4374、(ii) SEQ ID NO: 2 的核苷酸 1792-4374、(iii) SEQ ID NO: 70 的核苷酸 1792-4374 或 (iv) SEQ ID NO: 71 的核苷酸 1792-4374。在一些實施例中，編碼 FVIII 多肽的密碼子優化的序列包含如下核苷酸序列，其包含編碼 FVIII 多肽的 N 末端部分的第一核酸序列和編碼 FVIII 多肽的 C 末端部分的第二核酸序列；其中第二核酸序列與 (i) SEQ ID NO: 1 的核苷酸 1792-2277 和 2320-4374、(ii) SEQ ID NO: 2 的核苷酸 1792-2277 和 2320-4374、(iii) SEQ ID NO: 70 的核苷酸 1792-2277 和 2320-4374 或 (iv) SEQ ID NO: 71 的核苷酸 1792-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 1 的核苷酸 1792-4374、SEQ ID NO: 2 的核苷酸 1792-4374、SEQ ID NO: 70 的核苷酸 1792-4374 或 SEQ ID NO: 71 的核苷酸 1792-4374）具有至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性；並且其中 N 末端部分和 C 末端部分一起

具有 FVIII 多肽活性。在一個實施例中，第二核酸序列包含 (i) SEQ ID NO: 1 的核苷酸 1792-2277 和 2320-4374、(ii) SEQ ID NO: 2 的核苷酸 1792-2277 和 2320-4374、(iii) SEQ ID NO: 70 的核苷酸 1792-2277 和 2320-4374 或 (iv) SEQ ID NO: 71 的核苷酸 1792-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 1 的核苷酸 1792-4374、SEQ ID NO: 2 的核苷酸 1792-4374、SEQ ID NO: 70 的核苷酸 1792-4374 或 SEQ ID NO: 71 的核苷酸 1792-4374）。

【0271】 在一些實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含編碼具有 FVIII 活性的多肽的核苷酸序列，其中所述核苷酸序列包含與 SEQ ID NO: 1 的核苷酸 58 至 4374 具有至少 90%、至少 91%、至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性的核酸序列。在其他實施例中，核苷酸序列包含與 SEQ ID NO: 1 的核苷酸 58-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 1 的核苷酸 58-4374）具有至少 90%、至少 91%、至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性的核酸序列。在其他實施例中，核酸序列與 SEQ ID NO: 1 具有至少 90%、至少 91%、至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性。在其他實施例中，核苷酸序列包含 SEQ ID NO: 1 的核苷酸 58-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 1 的核苷酸 58-4374）或 SEQ ID NO: 1 的核苷酸 58 至 4374。在仍其他實施例中，核苷酸序列包含 SEQ ID NO: 1 的核苷酸 1-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 1 的核苷酸 1-4374）或 SEQ ID NO: 1 的核苷酸 1 至 4374。

【0272】 在一些實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含編碼具有 FVIII 活性的多肽的核苷酸序列，其中所述核苷酸序列包含與 SEQ ID NO: 2 的核苷酸 58 至 4374 具有至少 94%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性的核酸序列。在其他實施例中，核苷酸序列包含與 SEQ ID NO: 2 的核苷酸 58-2277 和 2320-4374 具有至少

94%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性的核酸序列。在其他實施例中，核酸序列與 SEQ ID NO: 2 具有至少 94%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性。在其他實施例中，核苷酸序列包含 SEQ ID NO: 2 的核苷酸 58-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 2 的核苷酸 58-4374）或 SEQ ID NO: 2 的核苷酸 58 至 4374。在仍其他實施例中，核苷酸序列包含 SEQ ID NO: 2 的核苷酸 1-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 2 的核苷酸 1-4374）或 SEQ ID NO: 2 的核苷酸 1 至 4374。

【0273】 在一些實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含編碼具有 FVIII 活性的多肽的核苷酸序列，其中所述核苷酸序列包含與 SEQ ID NO: 70 的核苷酸 58 至 4374 具有至少 85%、至少 86%、至少 87%、至少 88%、至少 89%、至少 90%、至少 91%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性的核酸序列。在其他實施例中，核苷酸序列包含與 SEQ ID NO: 70 的核苷酸 58-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 70 的核苷酸 58-4374）具有至少 85%、至少 86%、至少 87%、至少 88%、至少 89%、至少 90%、至少 91%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性的核酸序列。在其他實施例中，核酸序列與 SEQ ID NO: 70 具有至少 85%、至少 86%、至少 87%、至少 88%、至少 89%、至少 90%、至少 91%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性。在其他實施例中，核苷酸序列包含 SEQ ID NO: 70 的核苷酸 58-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 70 的核苷酸 58-4374）或 SEQ ID NO: 70 的核苷酸 58 至 4374。在仍其他實施例中，核苷酸序列包含 SEQ ID NO: 70 的核苷酸 1-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 70 的核苷酸 1-4374）或 SEQ ID NO: 70 的核苷酸 1 至 4374。

【0274】 在一些實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含編碼具有 FVIII 活性的多肽的核苷酸序列，其中所述核苷酸序列包含

與 SEQ ID NO: 71 的核苷酸 58 至 4374 具有至少 85%、至少 86%、至少 87%、至少 88%、至少 89%、至少 90%、至少 91%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性的核酸序列。在其他實施例中，核苷酸序列包含與 SEQ ID NO: 71 的核苷酸 58-2277 和 2320-4374（即，不含編碼B結構域或B結構域片段的核苷酸的SEQ ID NO: 71 的核苷酸 58-4374）具有至少 85%、至少 86%、至少 87%、至少 88%、至少 89%、至少 90%、至少 91%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性的核酸序列。在其他實施例中，核酸序列與 SEQ ID NO: 71 具有至少 85%、至少 86%、至少 87%、至少 88%、至少 89%、至少 90%、至少 91%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性。在其他實施例中，核苷酸序列包含 SEQ ID NO: 71 的核苷酸 58-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 71 的核苷酸 58-4374）或 SEQ ID NO: 71 的核苷酸 58 至 4374。在仍其他實施例中，核苷酸序列包含 SEQ ID NO: 71 的核苷酸 1-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 71 的核苷酸 1-4374）或 SEQ ID NO: 71 的核苷酸 1 至 4374。

【0275】 在一些實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含編碼具有 FVIII 活性的多肽的核苷酸序列，其中所述核苷酸序列包含與 SEQ ID NO: 3 的核苷酸 58 至 4374 具有至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性的核酸序列。在其他實施例中，核苷酸序列包含與 SEQ ID NO: 3 的核苷酸 58-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 3 的核苷酸 58-4374）具有至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性的核酸序列。在某些實施例中，核酸序列與 SEQ ID NO: 3 具有至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性。在一些實施例中，核苷酸序列包含 SEQ ID NO: 3 的核苷酸 58-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 3 的核苷酸 58-4374）或 SEQ ID NO: 3 的核苷酸 58 至 4374。在仍其他實施例中，核

苷酸序列包含 SEQ ID NO: 3 的核苷酸 58-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 3 的核苷酸 1-4374）或 SEQ ID NO: 3 的核苷酸 1 至 4374。

【0276】 在一些實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含編碼具有 FVIII 活性的多肽的核苷酸序列，其中所述核苷酸序列包含與 SEQ ID NO: 4 的核苷酸 58 至 4374 具有至少 91%、至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性的核酸序列。在其他實施例中，核苷酸序列包含與 SEQ ID NO: 4 的核苷酸 58-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 4 的核苷酸 58-4374）具有至少 91%、至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性的核酸序列。在其他實施例中，核酸序列與 SEQ ID NO: 4 具有至少 91%、至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性。在其他實施例中，核苷酸序列包含 SEQ ID NO: 4 的核苷酸 58-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 4 的核苷酸 58-4374）或 SEQ ID NO: 4 的核苷酸 58 至 4374。在仍其他實施例中，核苷酸序列包含 SEQ ID NO: 4 的核苷酸 1-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 4 的核苷酸 1-4374）或 SEQ ID NO: 4 的核苷酸 1 至 4374。

【0277】 在一些實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含編碼具有 FVIII 活性的多肽的核苷酸序列，其中所述核苷酸序列包含與 SEQ ID NO: 5 的核苷酸 58 至 4374 具有至少 89%、至少 90%、至少 91%、至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性的核酸序列。在其他實施例中，核苷酸序列包含與 SEQ ID NO: 5 的核苷酸 58-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 5 的核苷酸 58-4374）具有至少 89%、至少 90%、至少 91%、至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性的核酸序列。在某些實施例中，核酸序列與 SEQ ID NO: 5 具有至少 89%、至少 90%、至少 91%、

至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性。在一些實施例中，核苷酸序列包含 SEQ ID NO: 5 的核苷酸 58-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 5 的核苷酸 58-4374）或 SEQ ID NO: 5 的核苷酸 58 至 4374。在仍其他實施例中，核苷酸序列包含 SEQ ID NO: 5 的核苷酸 1-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 5 的核苷酸 1-4374）或 SEQ ID NO: 5 的核苷酸 1 至 4374。

【0278】 在一些實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含編碼具有 FVIII 活性的多肽的核苷酸序列，其中所述核苷酸序列包含與 SEQ ID NO: 6 的核苷酸 58 至 4374 具有至少 89%、至少 90%、至少 91%、至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性的核酸序列。在其他實施例中，核苷酸序列包含與 SEQ ID NO: 6 的核苷酸 58-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 6 的核苷酸 58-4374）具有至少 89%、至少 90%、至少 91%、至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性的核酸序列。在某些實施例中，核酸序列與 SEQ ID NO: 6 具有至少 89%、至少 90%、至少 91%、至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%序列同一性。在一些實施例中，核苷酸序列包含 SEQ ID NO: 6 的核苷酸 58-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 6 的核苷酸 58-4374）或 SEQ ID NO: 6 的核苷酸 58 至 4374。在仍其他實施例中，核苷酸序列包含 SEQ ID NO: 6 的核苷酸 1-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 6 的核苷酸 1-4374）或 SEQ ID NO: 6 的核苷酸 1 至 4374。

【0279】 在一些實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含編碼信號肽的核酸序列。在某些實施例中，信號肽是 FVIII 信號肽。在一些實施例中，編碼信號肽的核酸序列經密碼子優化。在一個特定實施例中，編碼信號肽的核酸序列與 (i) SEQ ID NO: 1 的核苷酸 1 至 57；(ii) SEQ ID NO: 2 的核苷酸 1 至 57；(iii) SEQ ID NO: 3 的核苷酸 1 至 57；(iv) SEQ ID

NO: 4 的核苷酸 1 至 57；(v) SEQ ID NO: 5 的核苷酸 1 至 57；(vi) SEQ ID NO: 6 的核苷酸 1 至 57；(vii) SEQ ID NO: 70 的核苷酸 1 至 57；(viii) SEQ ID NO: 71 的核苷酸 1 至 57；或 (ix) SEQ ID NO: 68 的核苷酸 1 至 57 具有至少 60%、至少 70%、至少 80%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%、至少 99%或 100%序列同一性。

【0280】 SEQ ID NO: 1-6、70 和 71 是起始或「親代」或「野生型」FVIII 核苷酸序列 SEQ ID NO: 16 的優化形式。SEQ ID NO: 16 編碼 B 結構域缺失的人類 FVIII。雖然 SEQ ID NO: 1-6、70 和 71 衍生自 FVIII 的特定 B 結構域缺失形式（SEQ ID NO: 16），但是應理解，本公開文本還包括編碼 FVIII 的其他形式的核酸的優化形式。例如，FVIII 的其他形式可以包括全長 FVIII、FVIII 的其他 B 結構域缺失（本文所述）或保留 FVIII 活性的其他 FVIII 片段。

【0281】 在一個實施例中，基因盒包含 FVIII 建構體，其包括如表 2A-2F 中所列示的多核苷酸序列。在一個實施例中，基因盒包含 FVIII 建構體，其包括表 2A 中所述的多核苷酸序列。在一個實施例中，基因盒包含 FVIII 建構體，其包括表 2B 中所述的多核苷酸序列。在一個實施例中，基因盒包含 FVIII 建構體，其包括表 2C 中所述的多核苷酸序列。在一個實施例中，基因盒包含 FVIII 建構體，其包括表 2D 中所述的多核苷酸序列。在一個實施例中，基因盒包含 FVIII 建構體，其包括表 2E 中所述的多核苷酸序列。在一個實施例中，基因盒包含 FVIII 建構體，其包括表 2F 中所述的多核苷酸序列。

【0282】 在某些實施例中，分離的核酸分子包含與 SEQ ID NO: 179、182、189 或 194 的核苷酸序列具有至少約 60%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或約 100%序列同一性的核苷酸序列。在一些實施例中，分離的核酸分子包含與 SEQ ID NO: 179 的核苷酸序列具有至少約 60%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或約 100%序列同一性的核苷酸序列。在一

些實施例中，分離的核酸分子包含與 SEQ ID NO: 182 的核苷酸序列具有至少約 60%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或約 100%序列同一性的核苷酸序列。在一些實施例中，分離的核酸分子包含與 SEQ ID NO: 189 的核苷酸序列具有至少約 60%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或約 100%序列同一性的核苷酸序列。在一些實施例中，分離的核酸分子包含與 SEQ ID NO: 194 的核苷酸序列具有至少約 60%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或約 100%序列同一性的核苷酸序列。在一些實施例中，分離的核酸分子保留表現功能性 FVIII 蛋白的能力。

表2A：例示性AAV-FVIII建構體（核苷酸1-6526；SEQ ID NO: 110）

描述	序列
5'ITR（5'端AAV2反向末端重複體）（SEQ ID NO: 111）	1 -- CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCG GGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGG GAGTGGCCAACTCCATCACTAGGGGTTCCT – 130
質體骨架序列（PBS）-1（SEQ ID NO: 112）	131 -- GCGGCAATTCAGTCGATAACTATAACGGTCCTAAGGTAGCGATTTAAATA CGCGCTCTCTTAAGGTAGCCCCGGGACGCGTCAATTGAGATCTGGATCCG GTACCGAATTCGCGGCCGCCTCGACGACTAGCGTTTAATTAA – 272
TTPp（肝臟特異性啟動子）（SEQ ID NO: 113）	273 -- ACGCGTGTCTGTCTGCACATTTCGTAGAGCGAGTGTTCCGATACTCTAATC TCCCTAGGCAAGGTTCATATTTGTGTAGGTTACTTATTCTCCTTTTGTTGAC TAAGTCAATAATCAGAATCAGCAGGTTTGGAGTCAGCTTGGCAGGGATCA

	GCAGCCTGGGTTGGAAGGAGGGGGTATAAAAGCCCCTTCACCAGGAGAA GCCGTCACACAGATCCACAAGCTCCTG – 501
PBS-2（SEQ ID NO: 114）	502 – AG – 503
合成內含子（SEQ ID NO: 115）	504 – GTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGC CCTTGCGTGCCTTGAATTACTGACACTGACATCCACTTTTTCTTTTTCTCCA CAG –609
PBS-3（SEQ ID NO: 116）	610 – CTAGCGCCACC – 620
FVIIIco6XTEN（SEQ ID NO: 117）（含有XTEN144的密碼子優化 的FVIII形式6的開放閱讀框；XTEN序列通過加雙下劃線（SEQ ID NO: 118）標記）	621 – ATGCAGATTGAGCTGTCCACTTGTTTCTTCCTGTGCCTCCTGCGCTTCTGTT TCTCCGCCACTCGCCGGTACTACCTTGGAGCCGTGGAGCTTTCATGGGACT ACATGCAGAGCGACCTGGGCGAACTCCCCGTGGATGCCAGATTCCCCCCC CGCGTGCCAAAGTCCTTCCCCTTTAACACCTCCGTGGTGTACAAGAAAACC CTCTTTGTCGAGTTCACTGACCACCTGTTCAACATCGCCAAGCCGCGCCCA CCTTGGATGGGCCTCCTGGGACCGACCATTCAAGCTGAAGTGTACGACAC CGTGGTGATCACCCTGAAGAACATGGCGTCCCACCCCGTGTCCCTGCATG CGGTCGGAGTGTCCTACTGGAAGGCCTCCGAAGGAGCTGAGTACGACGAC CAGACTAGCCAGCGGGAAAAGGAGGACGATAAAGTGTTCCCGGGCGGCT CGCATACTTACGTGTGGCAAGTCCTGAAGGAAAACGGACCTATGGCATCC GATCCTCTGTGCCTGACTTACTCCTACCTTTCCCATGTGGACCTCGTGAAG GACCTGAACAGCGGGCTGATTGGTGCACTTCTCGTGTGCCGCGAAGGTTC GCTCGCTAAGGAAAAGACCCAGACCCTCCATAAGTTCATCCTTTTGTTCGC TGTGTTCGATGAAGGAAAGTCATGGCATTCCGAAACTAAGAACTCGCTGA

	TGCAGGACCGGGATGCCGCCTCAGCCCGCGCCTGGCCTAAAATGCATACA
	GTCAACGGATACGTGAATCGGTCACTGCCCGGGCTCATCGGTTGTCACAG
	AAAGTCCGTGTACTGGCACGTCATCGGCATGGGCACTACGCCTGAAGTGC
	ACTCCATCTTCCTGGAAGGGCACACCTTCCTCGTGCGCAACCACCGCCAG
	GCCTCTCTGGAAATCTCCCCGATTACCTTTCTGACCGCCCAGACTCTGCTC
	ATGGACCTGGGGCAGTTCCTTCTCTTCTGCCACATCTCCAGCCATCAGCAC
	GACGGAATGGAGGCCTACGTGAAGGTGGACTCATGCCCGGAAGAACCTC
	AGTTGCGGATGAAGAACAACGAGGAGGCCGAGGACTATGACGACGATTT
	GACTGACTCCGAGATGGACGTCGTGCGGTTCGATGACGACAACAGCCCCA
	GCTTCATCCAGATTCGCAGCGTGGCCAAGAAGCACCCCAAAACCTGGGTG
	CACTACATCGCGGCCGAGGAAGAAGATTGGGACTACGCCCCGTTGGTGCT
	GGCACCCGATGACCGGTCGTACAAGTCCCAGTATCTGAACAATGGTCCGC
	AGCGGATTGGCAGAAAGTACAAGAAAGTGCGGTTCATGGCGTACACTGAC
	GAAACGTTTAAGACCCGGGAGGCCATTCAACATGAGAGCGGCATTCTGGG
	ACCACTGCTGTACGGAGAGGTCGGCGATACCCTGCTCATCATCTTCAAAA
	ACCAGGCCTCCCGGCCTTACAACATCTACCCTCACGGAATCACCGACGTG
	CGGCCACTCTACTCGCGGCGCCTGCCGAAGGGCGTCAAGCACCTGAAAGA
	CTTCCCTATCCTGCCGGGCGAAATCTTCAAGTATAAGTGGACCGTCACCGT
	GGAGGACGGGCCCACCAAGAGCGATCCTAGGTGTCTGACTCGGTACTACT
	CCAGCTTCGTGAACATGGAACGGGACCTGGCATCGGGACTCATTGGACCG
	CTGCTGATCTGCTACAAAGAGTCGGTGGATCAACGCGGCAACCAGATCAT
	GTCCGACAAGCGCAACGTGATCCTGTTCTCCGTGTTTGATGAAAACAGAT
	CCTGGTACCTCACTGAAAACATCCAGAGGTTCCTCCCAAACCCCGCAGGA
	GTGCAACTGGAGGACCCTGAGTTTCAGGCCTCGAATATCATGCACTCGAT
	TAACGGTTACGTGTTCGACTCGCTGCAACTGAGCGTGTGCCTCCATGAAGT

	CGCTTACTGGTACATTCTGTCCATCGGCGCCCAGACTGACTTCCTGAGCGT
	GTTCTTTTCCGGTTACACCTTTAAGCACAAGATGGTGTACGAAGATACCCT
	GACCCTGTTCCCTTTCTCCGGCGAAACGGTGTTCATGTCGATGGAGAACCC
	GGGTCTGTGGATTCTGGGATGCCACAACAGCGACTTTCGGAACCGCGGAA
	TGACTGCCCTGCTGAAGGTGTCCTCATGCGACAAGAACACCGGAGACTAC
	TACGAGGACTCCTACGAGGATATCTCAGCCTACCTCCTGTCCAAGAACAA
	CGCGATCGAGCCGCGCAGCTTCAGCCAGAACGGCGCGCCAACATCAGAG
	AGCGCCACCCCTGAAAGTGGTCCCGGGAGCGAGCCAGCCACATCTGGGTC
	GGAAACGCCAGGCACAAGTGAGTCTGCAACTCCCGAGTCCGGACCTGGCT
	CCGAGCCTGCCACTAGCGGCTCCGAGACTCCGGGAACTTCCGAGAGCGCT
	ACACCAGAAAGCGGACCCGGAACCAGTACCGAACCTAGCGAGGGCTCTG
	CTCCGGGCAGCCCAGCCGGCTCTCCTACATCCACGGAGGAGGGCACTTCC
	GAATCCGCCACCCCGGAGTCAGGGCCAGGATCTGAACCCGCTACCTCAGG
	CAGTGAGACGCCAGGAACGAGCGAGTCCGCTACACCGGAGAGTGGGCCA
	GGGAGCCCTGCTGGATCTCCTACGTCCACTGAGGAAGGGTCACCAGCGGG
	CTCGCCCACCAGCACTGAAGAAGGTGCCTCGAGCCCGCCTGTGCTGAAGA
	GGCACCAGCGAGAAATTACCCGGACCACCCTCCAATCGGATCAGGAGGA
	AATCGACTACGACGACACCATCTCGGTGGAAATGAAGAAGGAAGATTTCG
	ATATCTACGACGAGGACGAAAATCAGTCCCCTCGCTCATTCCAAAAGAAA
	ACTAGACACTACTTTATCGCCGCGGTGGAAAGACTGTGGGACTATGGAAT
	GTCATCCAGCCCTCACGTCCTTCGGAACCGGGCCCAGAGCGGATCGGTGC
	CTCAGTTCAAGAAAGTGGTGTTCCAGGAGTTCACCGACGGCAGCTTCACC
	CAGCCGCTGTACCGGGGAGAACTGAACGAACACCTGGGCCTGCTCGGTCC
	CTACATCCGCGCGGAAGTGGAGGATAACATCATGGTGACCTTCCGTAACC
	AAGCATCCAGACCTTACTCCTTCTATTCCTCCCTGATCTCATACGAGGAGG

	ACCAGCGCCAAGGCGCCGAGCCCCGCAAGAACTTCGTCAAGCCCAACGA
	GACTAAGACCTACTTCTGGAAGGTCCAACACCATATGGCCCCGACCAAGG
	ATGAGTTTGACTGCAAGGCCTGGGCCTACTTCTCCGACGTGGACCTTGAG
	AAGGATGTCCATTCCGGCCTGATCGGGCCGCTGCTCGTGTGTCACACCAA
	CACCCTGAACCCAGCGCATGGACGCCAGGTCACCGTCCAGGAGTTTGCTC
	TGTTCTTCACCATTTTTGACGAAACTAAGTCCTGGTACTTCACCGAGAATA
	TGGAGCGAAACTGTAGAGCGCCCTGCAATATCCAGATGGAAGATCCGACT
	TTCAAGGAGAACTATAGATTCCACGCCATCAACGGGTACATCATGGATAC
	TCTGCCGGGGCTGGTCATGGCCCAGGATCAGAGGATTCGGTGGTACTTGC
	TGTCAATGGGATCGAACGAAAACATTCACTCCATTCACTTCTCCGGTCACG
	TGTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCGCTGTACAATCTG
	TACCCCGGGGTGTTCGAAACTGTGGAGATGCTGCCGTCCAAGGCCGGCAT
	CTGGAGAGTGGAGTGCCTGATCGGAGAGCACCTCCACGCGGGGATGTCCA
	CCCTCTTCCTGGTGTACTCGAATAAGTGCCAGACCCCGCTGGGCATGGCCT
	CGGGCCACATCAGAGACTTCCAGATCACAGCAAGCGGACAATACGGCCA
	ATGGGCGCCGAAGCTGGCCCGCTTGCACTACTCCGGATCGATCAACGCAT
	GGTCCACCAAGGAACCGTTCTCGTGGATTAAGGTGGACCTCCTGGCCCCT
	ATGATTATCCACGGAATTAAGACCCAGGGCGCCAGGCAGAAGTTCTCCTC
	CCTGTACATCTCGCAATTCATCATCATGTACAGCCTGGACGGGAAGAAGT
	GGCAGACTTACAGGGGAAACTCCACCGGCACCCTGATGGTCTTTTTCGGC
	AACGTGGATTCCTCCGGCATTAAGCACAACATCTTCAACCCACCGATCAT
	AGCCAGATATATTAGGCTCCACCCCACTCACTACTCAATCCGCTCAACTCT
	TCGGATGGAACTCATGGGGTGCGACCTGAACTCCTGCTCCATGCCGTTGG
	GGATGGAATCAAAGGCTATTAGCGACGCCCAGATCACCGCGAGCTCCTAC
	TTCACTAACATGTTCGCCACCTGGAGCCCCTCCAAGGCCAGGCTGCACTTG

	CAGGGACGGTCAAATGCCTGGCGGCCGCAAGTGAACAATCCGAAGGAAT GGCTTCAAGTGGATTTCCAAAAGACCATGAAAGTGACCGGAGTCACCACC CAGGGAGTGAAGTCCCTTCTGACCTCGATGTATGTGAAGGAGTTCCTGATT AGCAGCAGCCAGGACGGGCACCAGTGGACCCTGTTCTTCCAAAACGGAAA GGTCAAGGTGTTCCAGGGGAACCAGGACTCGTTCACACCCGTGGTGAACT CCCTGGACCCCCCACTGCTGACGCGGTACTTGAGGATTCATCCTCAGTCCT GGGTCCATCAGATTGCATTGCGAATGGAAGTCCTGGGCTGCGAGGCCCAG GACCTGTACTGA – 5444
PBS-4（SEQ ID NO: 119）	5445 -- ATCAGCCTGAGCTCGCTGA – 5463
WPRE（突變的土撥鼠肝炎病毒轉錄後調控元件）（SEQ ID NO: 120）	5464 -- TCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCT TAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTG TATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAAT CCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTG GCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTG CCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGC CACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTC GGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCT TTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCT TCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCC TGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGA GTCGGATCTCCCTTTGGGCCGCCTCCCCGCTG – 6058
PBS-5（SEQ ID NO: 121）	6059 -- ATCAGCCT – 6066

bGHpA（牛生長激素多聚腺苷酸化信號）（SEQ ID NO: 122）	6067 -- CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCC TTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGA GGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGG GGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCAT GCTGGGGA –6277
PBS-6（SEQ ID NO: 123）	6278 -- TGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACGGGCTCGAGAAGCT TCTAGATATCCTCTCTTAAGGTAGCATCGAGATTTAAATTAGGGATAACAG GGTAATGGCGCGGGCCGC – 6396
3'ITR（3'端AAV2反向末端重複體）（SEQ ID NO: 124）	6397 -- AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGC TCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGGCTTTGCCCGG GCGGCCTCAGTGAGCGAGCGAGCGCGCAG – 6526

表2B：帶有B19d135 ITR的例示性B19-FVIII建構體（核苷酸1-6762；SEQ ID NO: 179）

描述	序列
5'ITR（SEQ ID NO: 180）	1 – CTCTGGGCCAGCTTGCTTGGGGTTGCCTTGACACTAAGACAAGCGGCGCG CCGCTTGATCTTAGTGGCACGTCAACCCCAAGCGCTGGCCCAGAGCCAAC CCTAATTCCGGAAGTCCCGCCCACCGGAAGTGACGTCACAGGAAATGACG TCACAGGAAATGACGTAATTGTCCGCCATCTTGTACCGGAAGTCCCGCCT ACCGGCGGCGACCGGCGGCATCTGATTTGGTGTCTTCTTTTAAATTTT –

	248
TTPp（肝臟特異性啟動子）（SEQ ID NO: 113）	391 – ACGCGTGTCTGTCTGCACATTTCGTAGAGCGAGTGTTCCGATACTCTAATC TCCCTAGGCAAGGTTCATATTTGTGTAGGTTACTTATTCTCCTTTTGTTGAC TAAGTCAATAATCAGAATCAGCAGGTTTGGAGTCAGCTTGGCAGGGATCA GCAGCCTGGGTTGGAAGGAGGGGGTATAAAAGCCCCTTCACCAGGAGAA GCCGTCACACAGATCCACAAGCTCCTG – 619
合成內含子（SEQ ID NO: 115）	622 – GTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGC CCTTGCGTGCCTTGAATTACTGACACTGACATCCACTTTTTCTTTTTCTCCA CAG --727
FVIIIco6XTEN（SEQ ID NO: 117）（含有XTEN144的密碼子優化的FVIII形式6的開放閱讀框；XTEN序列通過加<u>雙下劃線（SEQ ID NO: 118）</u>標記）	739 – ATGCAGATTGAGCTGTCCACTTGTTTCTTCCTGTGCCTCCTGCGCTTCTGTT TCTCCGCCACTCGCCGGTACTACCTTGGAGCCGTGGAGCTTTCATGGGACT ACATGCAGAGCGACCTGGGCGAACTCCCCGTGGATGCCAGATTCCCCCCC CGCGTGCCAAAGTCCTTCCCCTTTAACACCTCCGTGGTGTACAAGAAAAC CCTCTTTGTCGAGTTCACTGACCACCTGTTCAACATCGCCAAGCCGCGCCC ACCTTGGATGGGCCTCCTGGGACCGACCATTCAAGCTGAAGTGTACGACA CCGTGGTGATCACCCTGAAGAACATGGCGTCCCACCCCGTGTCCCTGCAT GCGGTCGGAGTGTCCTACTGGAAGGCCTCCGAAGGAGCTGAGTACGACGA CCAGACTAGCCAGCGGGAAAAGGAGGACGATAAAGTGTTCCCGGGCGGC TCGCATACTTACGTGTGGCAAGTCCTGAAGGAAAACGGACCTATGGCATC CGATCCTCTGTGCCTGACTTACTCCTACCTTTCCCATGTGGACCTCGTGAA GGACCTGAACAGCGGGCTGATTGGTGCACTTCTCGTGTGCCGCGAAGGTT CGCTCGCTAAGGAAAAGACCCAGACCCTCCATAAGTTCATCCTTTTGTTCG

	CTGTGTTCGATGAAGGAAAGTCATGGCATTCCGAAACTAAGAACTCGCTG
	ATGCAGGACCGGGATGCCGCCTCAGCCCGCGCCTGGCCTAAAATGCATAC
	AGTCAACGGATACGTGAATCGGTCACTGCCCGGGCTCATCGGTTGTCACA
	GAAAGTCCGTGTACTGGCACGTCATCGGCATGGGCACTACGCCTGAAGTG
	CACTCCATCTTCCTGGAAGGGCACACCTTCCTCGTGCGCAACCACCGCCA
	GGCCTCTCTGGAAATCTCCCCGATTACCTTTCTGACCGCCCAGACTCTGCT
	CATGGACCTGGGGCAGTTCCTTCTCTTCTGCCACATCTCCAGCCATCAGCA
	CGACGGAATGGAGGCCTACGTGAAGGTGGACTCATGCCCGGAAGAACCT
	CAGTTGCGGATGAAGAACAACGAGGAGGCCGAGGACTATGACGACGATT
	TGACTGACTCCGAGATGGACGTCGTGCGGTTCGATGACGACAACAGCCCC
	AGCTTCATCCAGATTCGCAGCGTGGCCAAGAAGCACCCCAAAACCTGGGT
	GCACTACATCGCGGCCGAGGAAGAAGATTGGGACTACGCCCCGTTGGTGC
	TGGCACCCGATGACCGGTCGTACAAGTCCCAGTATCTGAACAATGGTCCG
	CAGCGGATTGGCAGAAAGTACAAGAAAGTGCGGTTCATGGCGTACACTG
	ACGAAACGTTTAAGACCCGGGAGGCCATTCAACATGAGAGCGGCATTCTG
	GGACCACTGCTGTACGGAGAGGTCGGCGATACCCTGCTCATCATCTTCAA
	AAACCAGGCCTCCCGGCCTTACAACATCTACCCTCACGGAATCACCGACG
	TGCGGCCACTCTACTCGCGGCGCCTGCCGAAGGGCGTCAAGCACCTGAAA
	GACTTCCCTATCCTGCCGGGCGAAATCTTCAAGTATAAGTGGACCGTCAC
	CGTGGAGGACGGGCCCACCAAGAGCGATCCTAGGTGTCTGACTCGGTACT
	ACTCCAGCTTCGTGAACATGGAACGGGACCTGGCATCGGGACTCATTGGA
	CCGCTGCTGATCTGCTACAAAGAGTCGGTGGATCAACGCGGCAACCAGAT
	CATGTCCGACAAGCGCAACGTGATCCTGTTCTCCGTGTTTGATGAAAACA
	GATCCTGGTACCTCACTGAAAACATCCAGAGGTTCCTCCCAAACCCCGCA
	GGAGTGCAACTGGAGGACCCTGAGTTTCAGGCCTCGAATATCATGCACTC

	GATTAACGGTTACGTGTTCGACTCGCTGCAACTGAGCGTGTGCCTCCATGA
	AGTCGCTTACTGGTACATTCTGTCCATCGGCGCCCAGACTGACTTCCTGAG
	CGTGTTCTTTTCCGGTTACACCTTTAAGCACAAGATGGTGTACGAAGATAC
	CCTGACCCTGTTCCCTTTCTCCGGCGAAACGGTGTTCATGTCGATGGAGAA
	CCCGGGTCTGTGGATTCTGGGATGCCACAACAGCGACTTTCGGAACCGCG
	GAATGACTGCCCTGCTGAAGGTGTCCTCATGCGACAAGAACACCGGAGAC
	TACTACGAGGACTCCTACGAGGATATCTCAGCCTACCTCCTGTCCAAGAA
	CAACGCGATCGAGCCGCGCAGCTTCAGCCAGAACGGCGCGCCAACATCA
	GAGAGCGCCACCCCTGAAAGTGGTCCCGGGAGCGAGCCAGCCACATCTG
	GGTCGGAAACGCCAGGCACAAGTGAGTCTGCAACTCCCGAGTCCGGACCT
	GGCTCCGAGCCTGCCACTAGCGGCTCCGAGACTCCGGGAACTTCCGAGAG
	CGCTACACCAGAAAGCGGACCCGGAACCAGTACCGAACCTAGCGAGGGC
	TCTGCTCCGGGCAGCCCAGCCGGCTCTCCTACATCCACGGAGGAGGGCAC
	TTCCGAATCCGCCACCCCGGAGTCAGGGCCAGGATCTGAACCCGCTACCT
	CAGGCAGTGAGACGCCAGGAACGAGCGAGTCCGCTACACCGGAGAGTGG
	GCCAGGGAGCCCTGCTGGATCTCCTACGTCCACTGAGGAAGGGTCACCAG
	CGGGCTCGCCCACCAGCACTGAAGAAGGTGCCTCGAGCCCGCCTGTGCTG
	AAGAGGCACCAGCGAGAAATTACCCGGACCACCCTCCAATCGGATCAGG
	AGGAAATCGACTACGACGACACCATCTCGGTGGAAATGAAGAAGGAAGA
	TTTCGATATCTACGACGAGGACGAAAATCAGTCCCCTCGCTCATTCCAAA
	AGAAAACTAGACACTACTTTATCGCCGCGGTGGAAAGACTGTGGGACTAT
	GGAATGTCATCCAGCCCTCACGTCCTTCGGAACCGGGCCCAGAGCGGATC
	GGTGCCTCAGTTCAAGAAAGTGGTGTTCCAGGAGTTCACCGACGGCAGCT
	TCACCCAGCCGCTGTACCGGGGAGAACTGAACGAACACCTGGGCCTGCTC
	GGTCCCTACATCCGCGCGGAAGTGGAGGATAACATCATGGTGACCTTCCG

	TAACCAAGCATCCAGACCTTACTCCTTCTATTCCTCCCTGATCTCATACGA
	GGAGGACCAGCGCCAAGGCGCCGAGCCCCGCAAGAACTTCGTCAAGCCC
	AACGAGACTAAGACCTACTTCTGGAAGGTCCAACACCATATGGCCCCGAC
	CAAGGATGAGTTTGACTGCAAGGCCTGGGCCTACTTCTCCGACGTGGACC
	TTGAGAAGGATGTCCATTCCGGCCTGATCGGGCCGCTGCTCGTGTGTCAC
	ACCAACACCCTGAACCCAGCGCATGGACGCCAGGTCACCGTCCAGGAGTT
	TGCTCTGTTCTTCACCATTTTTGACGAAACTAAGTCCTGGTACTTCACCGA
	GAATATGGAGCGAAACTGTAGAGCGCCCTGCAATATCCAGATGGAAGATC
	CGACTTTCAAGGAGAACTATAGATTCCACGCCATCAACGGGTACATCATG
	GATACTCTGCCGGGGCTGGTCATGGCCCAGGATCAGAGGATTCGGTGGTA
	CTTGCTGTCAATGGGATCGAACGAAAACATTCACTCCATTCACTTCTCCGG
	TCACGTGTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCGCTGTACA
	ATCTGTACCCCGGGGTGTTCGAAACTGTGGAGATGCTGCCGTCCAAGGCC
	GGCATCTGGAGAGTGGAGTGCCTGATCGGAGAGCACCTCCACGCGGGGAT
	GTCCACCCTCTTCCTGGTGTACTCGAATAAGTGCCAGACCCCGCTGGGCAT
	GGCCTCGGGCCACATCAGAGACTTCCAGATCACAGCAAGCGGACAATACG
	GCCAATGGGCGCCGAAGCTGGCCCGCTTGCACTACTCCGGATCGATCAAC
	GCATGGTCCACCAAGGAACCGTTCTCGTGGATTAAGGTGGACCTCCTGGC
	CCCTATGATTATCCACGGAATTAAGACCCAGGGCGCCAGGCAGAAGTTCT
	CCTCCCTGTACATCTCGCAATTCATCATCATGTACAGCCTGGACGGGAAG
	AAGTGGCAGACTTACAGGGGAAACTCCACCGGCACCCTGATGGTCTTTTT
	CGGCAACGTGGATTCCTCCGGCATTAAGCACAACATCTTCAACCCACCGA
	TCATAGCCAGATATATTAGGCTCCACCCCACTCACTACTCAATCCGCTCAA
	CTCTTCGGATGGAACTCATGGGGTGCGACCTGAACTCCTGCTCCATGCCGT
	TGGGGATGGAATCAAAGGCTATTAGCGACGCCCAGATCACCGCGAGCTCC

	TACTTCACTAACATGTTCGCCACCTGGAGCCCCTCCAAGGCCAGGCTGCA CTTGCAGGGACGGTCAAATGCCTGGCGGCCGCAAGTGAACAATCCGAAG GAATGGCTTCAAGTGGATTTCCAAAAGACCATGAAAGTGACCGGAGTCAC CACCCAGGGAGTGAAGTCCCTTCTGACCTCGATGTATGTGAAGGAGTTCC TGATTAGCAGCAGCCAGGACGGGCACCAGTGGACCCTGTTCTTCCAAAAC GGAAAGGTCAAGGTGTTCCAGGGGAACCAGGACTCGTTCACACCCGTGGT GAACTCCCTGGACCCCCCACTGCTGACGCGGTACTTGAGGATTCATCCTCA GTCCTGGGTCCATCAGATTGCATTGCGAATGGAAGTCCTGGGCTGCGAGG CCCAGGACCTGTACTGA -- 5562
WPRE（突變的土撥鼠肝炎病毒轉錄後調控元件）（SEQ ID NO: 120）	5582 – TCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCT TAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTG TATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAAT CCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTG GCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTG CCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGC CACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTC GGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCT TTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCT TCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCC TGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGA GTCGGATCTCCCTTTGGGCCGCCTCCCCGCTG -- 6176
bGHpA（牛生長激素多聚腺苷酸化	6185 – CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGC CTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATG

信號）（SEQ ID NO: 122）	AGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTG GGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGC ATGCTGGGGA -- 6395
3' ITR反向末端重複體（SEQ ID NO: 181）	6515 – AAAATTTAAAAGAAGACACCAAATCAGATGCCGCCGGTCGCCGCCGGTA GGCGGGACTTCCGGTACAAGATGGCGGACAATTACGTCATTTCCTGTGAC GTCATTTCCTGTGACGTCACTTCCGGTGGGCGGGACTTCCGGAATTAGGGT TGGCTCTGGGCCAGCGCTTGGGGTTGACGTGCCACTAAGATCAAGCGGCG CGCCGCTTGTCTTAGTGTCAAGGCAACCCCAAGCAAGCTGGCCCAGAG -- 6762

全長序列（SEQ ID NO: 179）

CTCTGGGCCAGCTTGCTTGGGGTTGCCTTGACACTAAGACAAGCGGCGCGCCGCTTGATCTTAG
TGGCACGTCAACCCCAAGCGCTGGCCCAGAGCCAACCCTAATTCCGGAAGTCCCGCCCACCGG
AAGTGACGTCACAGGAAATGACGTCACAGGAAATGACGTAATTGTCCGCCATCTTGTACCGGA
AGTCCCGCCTACCGGCGGCGACCGGCGGCATCTGATTTGGTGTCTTCTTTTAAATTTTGCGGCA
ATTCAGTCGATAACTATAACGGTCCTAAGGTAGCGATTTAAATACGCGCTCTCTTAAGGTAGCC
CCGGGACGCGTCAATTGAGATCTGGATCCGGTACCGAATTCGCGGCCGCCTCGACGACTAGCG
TTTAATTAAACGCGTGTCTGTCTGCACATTTCGTAGAGCGAGTGTTCCGATACTCTAATCTCCCT
AGGCAAGGTTCATATTTGTGTAGGTTACTTATTCTCCTTTTGTTGACTAAGTCAATAATCAGAA
TCAGCAGGTTTGGAGTCAGCTTGGCAGGGATCAGCAGCCTGGGTTGGAAGGAGGGGGTATAA
AAGCCCCTTCACCAGGAGAAGCCGTCACACAGATCCACAAGCTCCTGAGGTAAGTGCCGTGTG
TGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTTGCGTGCCTTGAATTACTGACACT
GACATCCACTTTTTCTTTTTCTCCACAGCTAGCGCCACCATGCAGATTGAGCTGTCCACTTGTTT

CTTCCTGTGCCTCCTGCGCTTCTGTTTCTCCGCCACTCGCCGGTACTACCTTGGAGCCGTGGAGC

TTTCATGGGACTACATGCAGAGCGACCTGGGCGAACTCCCCGTGGATGCCAGATTCCCCCCCC

GCGTGCCAAAGTCCTTCCCCTTTAACACCTCCGTGGTGTACAAGAAAACCCTCTTTGTCGAGTT

CACTGACCACCTGTTCAACATCGCCAAGCCGCGCCCACCTTGGATGGGCCTCCTGGGACCGAC

CATTCAAGCTGAAGTGTACGACACCGTGGTGATCACCCTGAAGAACATGGCGTCCCACCCCGT

GTCCCTGCATGCGGTCGGAGTGTCCTACTGGAAGGCCTCCGAAGGAGCTGAGTACGACGACCA

GACTAGCCAGCGGGAAAAGGAGGACGATAAAGTGTTCCCGGGCGGCTCGCATACTTACGTGTG

GCAAGTCCTGAAGGAAAACGGACCTATGGCATCCGATCCTCTGTGCCTGACTTACTCCTACCTT

TCCCATGTGGACCTCGTGAAGGACCTGAACAGCGGGCTGATTGGTGCACTTCTCGTGTGCCGC

GAAGGTTCGCTCGCTAAGGAAAAGACCCAGACCCTCCATAAGTTCATCCTTTTGTTCGCTGTGT

TCGATGAAGGAAAGTCATGGCATTCCGAAACTAAGAACTCGCTGATGCAGGACCGGGATGCCG

CCTCAGCCCGCGCCTGGCCTAAAATGCATACAGTCAACGGATACGTGAATCGGTCACTGCCCG

GGCTCATCGGTTGTCACAGAAAGTCCGTGTACTGGCACGTCATCGGCATGGGCACTACGCCTG

AAGTGCACTCCATCTTCCTGGAAGGGCACACCTTCCTCGTGCGCAACCACCGCCAGGCCTCTCT

GGAAATCTCCCCGATTACCTTTCTGACCGCCCAGACTCTGCTCATGGACCTGGGGCAGTTCCTT

CTCTTCTGCCACATCTCCAGCCATCAGCACGACGGAATGGAGGCCTACGTGAAGGTGGACTCA

TGCCCGGAAGAACCTCAGTTGCGGATGAAGAACAACGAGGAGGCCGAGGACTATGACGACGA

TTTGACTGACTCCGAGATGGACGTCGTGCGGTTCGATGACGACAACAGCCCCAGCTTCATCCA

GATTCGCAGCGTGGCCAAGAAGCACCCCAAAACCTGGGTGCACTACATCGCGGCCGAGGAAG

AAGATTGGGACTACGCCCCGTTGGTGCTGGCACCCGATGACCGGTCGTACAAGTCCCAGTATC

TGAACAATGGTCCGCAGCGGATTGGCAGAAAGTACAAGAAAGTGCGGTTCATGGCGTACACTG

ACGAAACGTTTAAGACCCGGGAGGCCATTCAACATGAGAGCGGCATTCTGGGACCACTGCTGT

ACGGAGAGGTCGGCGATACCCTGCTCATCATCTTCAAAAACCAGGCCTCCCGGCCTTACAACA

TCTACCCTCACGGAATCACCGACGTGCGGCCACTCTACTCGCGGCGCCTGCCGAAGGGCGTCA

AGCACCTGAAAGACTTCCCTATCCTGCCGGGCGAAATCTTCAAGTATAAGTGGACCGTCACCG

TGGAGGACGGGCCCACCAAGAGCGATCCTAGGTGTCTGACTCGGTACTACTCCAGCTTCGTGA

ACATGGAACGGGACCTGGCATCGGGACTCATTGGACCGCTGCTGATCTGCTACAAAGAGTCGG

TGGATCAACGCGGCAACCAGATCATGTCCGACAAGCGCAACGTGATCCTGTTCTCCGTGTTTG

ATGAAAACAGATCCTGGTACCTCACTGAAAACATCCAGAGGTTCCTCCCAAACCCCGCAGGAG

TGCAACTGGAGGACCCTGAGTTTCAGGCCTCGAATATCATGCACTCGATTAACGGTTACGTGTT

CGACTCGCTGCAACTGAGCGTGTGCCTCCATGAAGTCGCTTACTGGTACATTCTGTCCATCGGC

GCCCAGACTGACTTCCTGAGCGTGTTCTTTTCCGGTTACACCTTTAAGCACAAGATGGTGTACG

AAGATACCCTGACCCTGTTCCCTTTCTCCGGCGAAACGGTGTTCATGTCGATGGAGAACCCGGG

TCTGTGGATTCTGGGATGCCACAACAGCGACTTTCGGAACCGCGGAATGACTGCCCTGCTGAA

GGTGTCCTCATGCGACAAGAACACCGGAGACTACTACGAGGACTCCTACGAGGATATCTCAGC

CTACCTCCTGTCCAAGAACAACGCGATCGAGCCGCGCAGCTTCAGCCAGAACGGCGCGCCAAC

ATCAGAGAGCGCCACCCCTGAAAGTGGTCCCGGGAGCGAGCCAGCCACATCTGGGTCGGAAA

CGCCAGGCACAAGTGAGTCTGCAACTCCCGAGTCCGGACCTGGCTCCGAGCCTGCCACTAGCG

GCTCCGAGACTCCGGGAACTTCCGAGAGCGCTACACCAGAAAGCGGACCCGGAACCAGTACC

GAACCTAGCGAGGGCTCTGCTCCGGGCAGCCCAGCCGGCTCTCCTACATCCACGGAGGAGGGC

ACTTCCGAATCCGCCACCCCGGAGTCAGGGCCAGGATCTGAACCCGCTACCTCAGGCAGTGAG

ACGCCAGGAACGAGCGAGTCCGCTACACCGGAGAGTGGGCCAGGGAGCCCTGCTGGATCTCCT

ACGTCCACTGAGGAAGGGTCACCAGCGGGCTCGCCCACCAGCACTGAAGAAGGTGCCTCGAG

CCCGCCTGTGCTGAAGAGGCACCAGCGAGAAATTACCCGGACCACCCTCCAATCGGATCAGGA

GGAAATCGACTACGACGACACCATCTCGGTGGAAATGAAGAAGGAAGATTTCGATATCTACGA

CGAGGACGAAAATCAGTCCCCTCGCTCATTCCAAAAGAAAACTAGACACTACTTTATCGCCGC

GGTGGAAAGACTGTGGGACTATGGAATGTCATCCAGCCCTCACGTCCTTCGGAACCGGGCCCA

GAGCGGATCGGTGCCTCAGTTCAAGAAAGTGGTGTTCCAGGAGTTCACCGACGGCAGCTTCAC

CCAGCCGCTGTACCGGGGAGAACTGAACGAACACCTGGGCCTGCTCGGTCCCTACATCCGCGC

GGAAGTGGAGGATAACATCATGGTGACCTTCCGTAACCAAGCATCCAGACCTTACTCCTTCTAT

TCCTCCCTGATCTCATACGAGGAGGACCAGCGCCAAGGCGCCGAGCCCCGCAAGAACTTCGTC

AAGCCCAACGAGACTAAGACCTACTTCTGGAAGGTCCAACACCATATGGCCCCGACCAAGGAT

GAGTTTGACTGCAAGGCCTGGGCCTACTTCTCCGACGTGGACCTTGAGAAGGATGTCCATTCCG

GCCTGATCGGGCCGCTGCTCGTGTGTCACACCAACACCCTGAACCCAGCGCATGGACGCCAGG

TCACCGTCCAGGAGTTTGCTCTGTTCTTCACCATTTTTGACGAAACTAAGTCCTGGTACTTCACC

GAGAATATGGAGCGAAACTGTAGAGCGCCCTGCAATATCCAGATGGAAGATCCGACTTTCAAG

GAGAACTATAGATTCCACGCCATCAACGGGTACATCATGGATACTCTGCCGGGGCTGGTCATG

GCCCAGGATCAGAGGATTCGGTGGTACTTGCTGTCAATGGGATCGAACGAAAACATTCACTCC

ATTCACTTCTCCGGTCACGTGTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCGCTGTAC

AATCTGTACCCCGGGGTGTTCGAAACTGTGGAGATGCTGCCGTCCAAGGCCGGCATCTGGAGA

GTGGAGTGCCTGATCGGAGAGCACCTCCACGCGGGGATGTCCACCCTCTTCCTGGTGTACTCG

AATAAGTGCCAGACCCCGCTGGGCATGGCCTCGGGCCACATCAGAGACTTCCAGATCACAGCA

AGCGGACAATACGGCCAATGGGCGCCGAAGCTGGCCCGCTTGCACTACTCCGGATCGATCAAC

GCATGGTCCACCAAGGAACCGTTCTCGTGGATTAAGGTGGACCTCCTGGCCCCTATGATTATCC

ACGGAATTAAGACCCAGGGCGCCAGGCAGAAGTTCTCCTCCCTGTACATCTCGCAATTCATCA

TCATGTACAGCCTGGACGGGAAGAAGTGGCAGACTTACAGGGGAAACTCCACCGGCACCCTG

ATGGTCTTTTTCGGCAACGTGGATTCCTCCGGCATTAAGCACAACATCTTCAACCCACCGATCA

TAGCCAGATATATTAGGCTCCACCCCACTCACTACTCAATCCGCTCAACTCTTCGGATGGAACT

CATGGGGTGCGACCTGAACTCCTGCTCCATGCCGTTGGGGATGGAATCAAAGGCTATTAGCGA

CGCCCAGATCACCGCGAGCTCCTACTTCACTAACATGTTCGCCACCTGGAGCCCCTCCAAGGCC

AGGCTGCACTTGCAGGGACGGTCAAATGCCTGGCGGCCGCAAGTGAACAATCCGAAGGAATG

GCTTCAAGTGGATTTCCAAAAGACCATGAAAGTGACCGGAGTCACCACCCAGGGAGTGAAGTC

CCTTCTGACCTCGATGTATGTGAAGGAGTTCCTGATTAGCAGCAGCCAGGACGGGCACCAGTG

GACCCTGTTCTTCCAAAACGGAAAGGTCAAGGTGTTCCAGGGGAACCAGGACTCGTTCACACC

CGTGGTGAACTCCCTGGACCCCCCACTGCTGACGCGGTACTTGAGGATTCATCCTCAGTCCTGG

GTCCATCAGATTGCATTGCGAATGGAAGTCCTGGGCTGCGAGGCCCAGGACCTGTACTGAATC

AGCCTGAGCTCGCTGATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTAT

TCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTA

TTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGG

AGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCAC

TGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTG

CCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCA

CTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCC

ACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTC

CTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAG

TCGGATCTCCCTTTGGGCCGCCTCCCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCC

ATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTT

CCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGG

GGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCG

GTGGGCTCTATGGCTTCTGAGGCGGAAAGAACGGGCTCGAGAAGCTTCTAGATATCCTCTCTT

AAGGTAGCATCGAGATTTAAATTAGGGATAACAGGGTAATGGCGCGGGCCGCAAAATTTAAA

AGAAGACACCAAATCAGATGCCGCCGGTCGCCGCCGGTAGGCGGGACTTCCGGTACAAGATG

GCGGACAATTACGTCATTTCCTGTGACGTCATTTCCTGTGACGTCACTTCCGGTGGGCGGGACT

TCCGGAATTAGGGTTGGCTCTGGGCCAGCGCTTGGGGTTGACGTGCCACTAAGATCAAGCGGC

GCGCCGCTTGTCTTAGTGTCAAGGCAACCCCAAGCAAGCTGGCCCAGAG

表2C：帶有GPVd162 ITR的例示性GPV-FVIII建構體（核苷酸1-6830；SEQ ID NO: 182）

描述	序列
5' ITR（SEQ ID NO: 183）	1 – CGGTGACGTGTTTCCGGCTGTTAGGTTGACCACGCGCATGCCGCGCGGTC AGCCCAATAGTTAAGCCGGAAACACGTCACCGGAAGTCACATGACCGGA AGTCACGTGACCGGAAACACGTGACAGGAAGCACGTGACCGGAACTAC GTCACCGGATGTGCGTCACCGGAAGCATGTGACCGGAACTTGCGTCACT TCCCCCTCCCCTGATTGGCTGGTTCGAACGAACGAACCCTCCAATGAGAC TCAAGGACAAGAGGATATTTTGCGCGCCAGGAAGTG -- 282
TTPp（肝臟特異性啟動子)(SEQ ID NO: 113）	425 – ACGCGTGTCTGTCTGCACATTTCGTAGAGCGAGTGTTCCGATACTCTAAT CTCCCTAGGCAAGGTTCATATTTGTGTAGGTTACTTATTCTCCTTTTGTTG ACTAAGTCAATAATCAGAATCAGCAGGTTTGGAGTCAGCTTGGCAGGGA TCAGCAGCCTGGGTTGGAAGGAGGGGGTATAAAAGCCCCTTCACCAGGA GAAGCCGTCACACAGATCCACAAGCTCCTG -- 653
合成內含子（SEQ ID NO: 115）	656 – GTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGG CCCTTGCGTGCCTTGAATTACTGACACTGACATCCACTTTTTCTTTTTCTC CACAG -- 761
FVIIIco6XTEN（SEQ ID NO: 117）（含有XTEN144的	773 – ATGCAGATTGAGCTGTCCACTTGTTTCTTCCTGTGCCTCCTGCGCTTCTGT TTCTCCGCCACTCGCCGGTACTACCTTGGAGCCGTGGAGCTTTCATGGGA CTACATGCAGAGCGACCTGGGCGAACTCCCCGTGGATGCCAGATTCCCC CCCCGCGTGCCAAAGTCCTTCCCCTTTAACACCTCCGTGGTGTACAAGAA

密碼子優化的FVIII形式6的開放閱讀框；XTEN序列通過加雙下劃線（SEQ ID NO: 118）標記）	AACCCTCTTTGTCGAGTTCACTGACCACCTGTTCAACATCGCCAAGCCGC GCCCACCTTGGATGGGCCTCCTGGGACCGACCATTCAAGCTGAAGTGTA CGACACCGTGGTGATCACCCTGAAGAACATGGCGTCCCACCCCGTGTCC CTGCATGCGGTCGGAGTGTCCTACTGGAAGGCCTCCGAAGGAGCTGAGT ACGACGACCAGACTAGCCAGCGGGAAAAGGAGGACGATAAAGTGTTCC CGGGCGGCTCGCATACTTACGTGTGGCAAGTCCTGAAGGAAAACGGACC TATGGCATCCGATCCTCTGTGCCTGACTTACTCCTACCTTTCCCATGTGGA CCTCGTGAAGGACCTGAACAGCGGGCTGATTGGTGCACTTCTCGTGTGCC GCGAAGGTTCGCTCGCTAAGGAAAAGACCCAGACCCTCCATAAGTTCAT CCTTTTGTTCGCTGTGTTCGATGAAGGAAAGTCATGGCATTCCGAAACTA AGAACTCGCTGATGCAGGACCGGGATGCCGCCTCAGCCCGCGCCTGGCC TAAAATGCATACAGTCAACGGATACGTGAATCGGTCACTGCCCGGGCTC ATCGGTTGTCACAGAAAGTCCGTGTACTGGCACGTCATCGGCATGGGCA CTACGCCTGAAGTGCACTCCATCTTCCTGGAAGGGCACACCTTCCTCGTG CGCAACCACCGCCAGGCCTCTCTGGAAATCTCCCCGATTACCTTTCTGAC CGCCCAGACTCTGCTCATGGACCTGGGGCAGTTCCTTCTCTTCTGCCACA TCTCCAGCCATCAGCACGACGGAATGGAGGCCTACGTGAAGGTGGACTC ATGCCCGGAAGAACCTCAGTTGCGGATGAAGAACAACGAGGAGGCCGA GGACTATGACGACGATTTGACTGACTCCGAGATGGACGTCGTGCGGTTC GATGACGACAACAGCCCCAGCTTCATCCAGATTCGCAGCGTGGCCAAGA AGCACCCCAAAACCTGGGTGCACTACATCGCGGCCGAGGAAGAAGATTG GGACTACGCCCCGTTGGTGCTGGCACCCGATGACCGGTCGTACAAGTCC CAGTATCTGAACAATGGTCCGCAGCGGATTGGCAGAAAGTACAAGAAAG TGCGGTTCATGGCGTACACTGACGAAACGTTTAAGACCCGGGAGGCCAT TCAACATGAGAGCGGCATTCTGGGACCACTGCTGTACGGAGAGGTCGGC

	GATACCCTGCTCATCATCTTCAAAAACCAGGCCTCCCGGCCTTACAACAT
	CTACCCTCACGGAATCACCGACGTGCGGCCACTCTACTCGCGGCGCCTGC
	CGAAGGGCGTCAAGCACCTGAAAGACTTCCCTATCCTGCCGGGCGAAAT
	CTTCAAGTATAAGTGGACCGTCACCGTGGAGGACGGGCCCACCAAGAGC
	GATCCTAGGTGTCTGACTCGGTACTACTCCAGCTTCGTGAACATGGAACG
	GGACCTGGCATCGGGACTCATTGGACCGCTGCTGATCTGCTACAAAGAG
	TCGGTGGATCAACGCGGCAACCAGATCATGTCCGACAAGCGCAACGTGA
	TCCTGTTCTCCGTGTTTGATGAAAACAGATCCTGGTACCTCACTGAAAAC
	ATCCAGAGGTTCCTCCCAAACCCCGCAGGAGTGCAACTGGAGGACCCTG
	AGTTTCAGGCCTCGAATATCATGCACTCGATTAACGGTTACGTGTTCGAC
	TCGCTGCAACTGAGCGTGTGCCTCCATGAAGTCGCTTACTGGTACATTCT
	GTCCATCGGCGCCCAGACTGACTTCCTGAGCGTGTTCTTTTCCGGTTACA
	CCTTTAAGCACAAGATGGTGTACGAAGATACCCTGACCCTGTTCCCTTTC
	TCCGGCGAAACGGTGTTCATGTCGATGGAGAACCCGGGTCTGTGGATTC
	TGGGATGCCACAACAGCGACTTTCGGAACCGCGGAATGACTGCCCTGCT
	GAAGGTGTCCTCATGCGACAAGAACACCGGAGACTACTACGAGGACTCC
	TACGAGGATATCTCAGCCTACCTCCTGTCCAAGAACAACGCGATCGAGC
	CGCGCAGCTTCAGCCAGAACGGCGCGCCAACATCAGAGAGCGCCACCCC
	TGAAAGTGGTCCCGGGAGCGAGCCAGCCACATCTGGGTCGGAAACGCCA
	GGCACAAGTGAGTCTGCAACTCCCGAGTCCGGACCTGGCTCCGAGCCTG
	CCACTAGCGGCTCCGAGACTCCGGGAACTTCCGAGAGCGCTACACCAGA
	AAGCGGACCCGGAACCAGTACCGAACCTAGCGAGGGCTCTGCTCCGGGC
	AGCCCAGCCGGCTCTCCTACATCCACGGAGGAGGGCACTTCCGAATCCG
	CCACCCCGGAGTCAGGGCCAGGATCTGAACCCGCTACCTCAGGCAGTGA
	GACGCCAGGAACGAGCGAGTCCGCTACACCGGAGAGTGGGCCAGGGAG

	CCCTGCTGGATCTCCTACGTCCACTGAGGAAGGGTCACCAGCGGGCTCG
	CCCACCAGCACTGAAGAAGGTGCCTCGAGCCCGCCTGTGCTGAAGAGGC
	ACCAGCGAGAAATTACCCGGACCACCCTCCAATCGGATCAGGAGGAAAT
	CGACTACGACGACACCATCTCGGTGGAAATGAAGAAGGAAGATTTCGAT
	ATCTACGACGAGGACGAAAATCAGTCCCCTCGCTCATTCCAAAAGAAAA
	CTAGACACTACTTTATCGCCGCGGTGGAAAGACTGTGGGACTATGGAAT
	GTCATCCAGCCCTCACGTCCTTCGGAACCGGGCCCAGAGCGGATCGGTG
	CCTCAGTTCAAGAAAGTGGTGTTCCAGGAGTTCACCGACGGCAGCTTCA
	CCCAGCCGCTGTACCGGGGAGAACTGAACGAACACCTGGGCCTGCTCGG
	TCCCTACATCCGCGCGGAAGTGGAGGATAACATCATGGTGACCTTCCGT
	AACCAAGCATCCAGACCTTACTCCTTCTATTCCTCCCTGATCTCATACGA
	GGAGGACCAGCGCCAAGGCGCCGAGCCCCGCAAGAACTTCGTCAAGCCC
	AACGAGACTAAGACCTACTTCTGGAAGGTCCAACACCATATGGCCCCGA
	CCAAGGATGAGTTTGACTGCAAGGCCTGGGCCTACTTCTCCGACGTGGA
	CCTTGAGAAGGATGTCCATTCCGGCCTGATCGGGCCGCTGCTCGTGTGTC
	ACACCAACACCCTGAACCCAGCGCATGGACGCCAGGTCACCGTCCAGGA
	GTTTGCTCTGTTCTTCACCATTTTTGACGAAACTAAGTCCTGGTACTTCAC
	CGAGAATATGGAGCGAAACTGTAGAGCGCCCTGCAATATCCAGATGGAA
	GATCCGACTTTCAAGGAGAACTATAGATTCCACGCCATCAACGGGTACA
	TCATGGATACTCTGCCGGGGCTGGTCATGGCCCAGGATCAGAGGATTCG
	GTGGTACTTGCTGTCAATGGGATCGAACGAAAACATTCACTCCATTCACT
	TCTCCGGTCACGTGTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGC
	GCTGTACAATCTGTACCCCGGGGTGTTCGAAACTGTGGAGATGCTGCCGT
	CCAAGGCCGGCATCTGGAGAGTGGAGTGCCTGATCGGAGAGCACCTCCA
	CGCGGGGATGTCCACCCTCTTCCTGGTGTACTCGAATAAGTGCCAGACCC

	CGCTGGGCATGGCCTCGGGCCACATCAGAGACTTCCAGATCACAGCAAG CGGACAATACGGCCAATGGGCGCCGAAGCTGGCCCGCTTGCACTACTCC GGATCGATCAACGCATGGTCCACCAAGGAACCGTTCTCGTGGATTAAGG TGGACCTCCTGGCCCCTATGATTATCCACGGAATTAAGACCCAGGGCGC CAGGCAGAAGTTCTCCTCCCTGTACATCTCGCAATTCATCATCATGTACA GCCTGGACGGGAAGAAGTGGCAGACTTACAGGGGAAACTCCACCGGCA CCCTGATGGTCTTTTTCGGCAACGTGGATTCCTCCGGCATTAAGCACAAC ATCTTCAACCCACCGATCATAGCCAGATATATTAGGCTCCACCCCACTCA CTACTCAATCCGCTCAACTCTTCGGATGGAACTCATGGGGTGCGACCTGA ACTCCTGCTCCATGCCGTTGGGGATGGAATCAAAGGCTATTAGCGACGC CCAGATCACCGCGAGCTCCTACTTCACTAACATGTTCGCCACCTGGAGCC CCTCCAAGGCCAGGCTGCACTTGCAGGGACGGTCAAATGCCTGGCGGCC GCAAGTGAACAATCCGAAGGAATGGCTTCAAGTGGATTTCCAAAAGACC ATGAAAGTGACCGGAGTCACCACCCAGGGAGTGAAGTCCCTTCTGACCT CGATGTATGTGAAGGAGTTCCTGATTAGCAGCAGCCAGGACGGGCACCA GTGGACCCTGTTCTTCCAAAACGGAAAGGTCAAGGTGTTCCAGGGGAAC CAGGACTCGTTCACACCCGTGGTGAACTCCCTGGACCCCCCACTGCTGAC GCGGTACTTGAGGATTCATCCTCAGTCCTGGGTCCATCAGATTGCATTGC GAATGGAAGTCCTGGGCTGCGAGGCCCAGGACCTGTACTGA --5596
WPRE(突變的土撥鼠肝炎病毒轉錄後調控元件)(SEQ ID NO: 120)	5616 – TCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTC TTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTT TGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATA AATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAA CGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGG

	CATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCC TATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAG GGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATC ATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGG GACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTT CCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGC CCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCTG -- 6210
bGHpA（牛生長激素多聚腺苷酸化信號）（SEQ ID NO: 122）	6219 – CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGC CTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAAT GAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGG TGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAG GCATGCTGGGGA -- 6429
3' ITR反向末端重複體（SEQ ID NO: 184）	6549 -- CACTTCCTGGCGCGCAAAATATCCTCTTGTCCTTGAGTCTCATTGGAGGG TTCGTTCGTTCGAACCAGCCAATCAGGGGAGGGGGAAGTGACGCAAGTT CCGGTCACATGCTTCCGGTGACGCACATCCGGTGACGTAGTTCCGGTCAC GTGCTTCCTGTCACGTGTTTCCGGTCACGTGACTTCCGGTCATGTGACTTC CGGTGACGTGTTTCCGGCTTAACTATTGGGCTGACCGCGCGGCATGCGCG TGGTCAACCTAACAGCCGGAAACACGTCACCG -- 6830
全長序列（SEQ ID NO: 182）	

CGGTGACGTGTTTCCGGCTGTTAGGTTGACCACGCGCATGCCGCGCGGTCAGCCCAATAGTT

AAGCCGGAAACACGTCACCGGAAGTCACATGACCGGAAGTCACGTGACCGGAAACACGTGA

CAGGAAGCACGTGACCGGAACTACGTCACCGGATGTGCGTCACCGGAAGCATGTGACCGGA

ACTTGCGTCACTTCCCCCTCCCCTGATTGGCTGGTTCGAACGAACGAACCCTCCAATGAGACT

CAAGGACAAGAGGATATTTTGCGCGCCAGGAAGTGGCGGCAATTCAGTCGATAACTATAAC

GGTCCTAAGGTAGCGATTTAAATACGCGCTCTCTTAAGGTAGCCCCGGGACGCGTCAATTGA

GATCTGGATCCGGTACCGAATTCGCGGCCGCCTCGACGACTAGCGTTTAATTAAACGCGTGT

CTGTCTGCACATTTCGTAGAGCGAGTGTTCCGATACTCTAATCTCCCTAGGCAAGGTTCATAT

TTGTGTAGGTTACTTATTCTCCTTTTGTTGACTAAGTCAATAATCAGAATCAGCAGGTTTGGA

GTCAGCTTGGCAGGGATCAGCAGCCTGGGTTGGAAGGAGGGGGTATAAAAGCCCCTTCACC

AGGAGAAGCCGTCACACAGATCCACAAGCTCCTGAGGTAAGTGCCGTGTGTGGTTCCCGCGG

GCCTGGCCTCTTTACGGGTTATGGCCCTTGCGTGCCTTGAATTACTGACACTGACATCCACTT

TTTCTTTTTCTCCACAGCTAGCGCCACCATGCAGATTGAGCTGTCCACTTGTTTCTTCCTGTGC

CTCCTGCGCTTCTGTTTCTCCGCCACTCGCCGGTACTACCTTGGAGCCGTGGAGCTTTCATGG

GACTACATGCAGAGCGACCTGGGCGAACTCCCCGTGGATGCCAGATTCCCCCCCCGCGTGCC

AAAGTCCTTCCCCTTTAACACCTCCGTGGTGTACAAGAAAACCCTCTTTGTCGAGTTCACTGA

CCACCTGTTCAACATCGCCAAGCCGCGCCCACCTTGGATGGGCCTCCTGGGACCGACCATTC

AAGCTGAAGTGTACGACACCGTGGTGATCACCCTGAAGAACATGGCGTCCCACCCCGTGTCC

CTGCATGCGGTCGGAGTGTCCTACTGGAAGGCCTCCGAAGGAGCTGAGTACGACGACCAGA

CTAGCCAGCGGGAAAAGGAGGACGATAAAGTGTTCCCGGGCGGCTCGCATACTTACGTGTG

GCAAGTCCTGAAGGAAAACGGACCTATGGCATCCGATCCTCTGTGCCTGACTTACTCCTACC

TTTCCCATGTGGACCTCGTGAAGGACCTGAACAGCGGGCTGATTGGTGCACTTCTCGTGTGC

CGCGAAGGTTCGCTCGCTAAGGAAAAGACCCAGACCCTCCATAAGTTCATCCTTTTGTTCGC

TGTGTTCGATGAAGGAAAGTCATGGCATTCCGAAACTAAGAACTCGCTGATGCAGGACCGG

GATGCCGCCTCAGCCCGCGCCTGGCCTAAAATGCATACAGTCAACGGATACGTGAATCGGTC

ACTGCCCGGGCTCATCGGTTGTCACAGAAAGTCCGTGTACTGGCACGTCATCGGCATGGGCA

CTACGCCTGAAGTGCACTCCATCTTCCTGGAAGGGCACACCTTCCTCGTGCGCAACCACCGC

CAGGCCTCTCTGGAAATCTCCCCGATTACCTTTCTGACCGCCCAGACTCTGCTCATGGACCTG

GGGCAGTTCCTTCTCTTCTGCCACATCTCCAGCCATCAGCACGACGGAATGGAGGCCTACGT

GAAGGTGGACTCATGCCCGGAAGAACCTCAGTTGCGGATGAAGAACAACGAGGAGGCCGAG

GACTATGACGACGATTTGACTGACTCCGAGATGGACGTCGTGCGGTTCGATGACGACAACAG

CCCCAGCTTCATCCAGATTCGCAGCGTGGCCAAGAAGCACCCCAAAACCTGGGTGCACTACA

TCGCGGCCGAGGAAGAAGATTGGGACTACGCCCCGTTGGTGCTGGCACCCGATGACCGGTC

GTACAAGTCCCAGTATCTGAACAATGGTCCGCAGCGGATTGGCAGAAAGTACAAGAAAGTG

CGGTTCATGGCGTACACTGACGAAACGTTTAAGACCCGGGAGGCCATTCAACATGAGAGCG

GCATTCTGGGACCACTGCTGTACGGAGAGGTCGGCGATACCCTGCTCATCATCTTCAAAAAC

CAGGCCTCCCGGCCTTACAACATCTACCCTCACGGAATCACCGACGTGCGGCCACTCTACTC

GCGGCGCCTGCCGAAGGGCGTCAAGCACCTGAAAGACTTCCCTATCCTGCCGGGCGAAATCT

TCAAGTATAAGTGGACCGTCACCGTGGAGGACGGGCCCACCAAGAGCGATCCTAGGTGTCT

GACTCGGTACTACTCCAGCTTCGTGAACATGGAACGGGACCTGGCATCGGGACTCATTGGAC

CGCTGCTGATCTGCTACAAAGAGTCGGTGGATCAACGCGGCAACCAGATCATGTCCGACAAG

CGCAACGTGATCCTGTTCTCCGTGTTTGATGAAAACAGATCCTGGTACCTCACTGAAAACAT

CCAGAGGTTCCTCCCAAACCCCGCAGGAGTGCAACTGGAGGACCCTGAGTTTCAGGCCTCGA

ATATCATGCACTCGATTAACGGTTACGTGTTCGACTCGCTGCAACTGAGCGTGTGCCTCCATG

AAGTCGCTTACTGGTACATTCTGTCCATCGGCGCCCAGACTGACTTCCTGAGCGTGTTCTTTT

CCGGTTACACCTTTAAGCACAAGATGGTGTACGAAGATACCCTGACCCTGTTCCCTTTCTCCG

GCGAAACGGTGTTCATGTCGATGGAGAACCCGGGTCTGTGGATTCTGGGATGCCACAACAGC

GACTTTCGGAACCGCGGAATGACTGCCCTGCTGAAGGTGTCCTCATGCGACAAGAACACCGG

AGACTACTACGAGGACTCCTACGAGGATATCTCAGCCTACCTCCTGTCCAAGAACAACGCGA

TCGAGCCGCGCAGCTTCAGCCAGAACGGCGCGCCAACATCAGAGAGCGCCACCCCTGAAAG

TGGTCCCGGGAGCGAGCCAGCCACATCTGGGTCGGAAACGCCAGGCACAAGTGAGTCTGCA

ACTCCCGAGTCCGGACCTGGCTCCGAGCCTGCCACTAGCGGCTCCGAGACTCCGGGAACTTC

CGAGAGCGCTACACCAGAAAGCGGACCCGGAACCAGTACCGAACCTAGCGAGGGCTCTGCT

CCGGGCAGCCCAGCCGGCTCTCCTACATCCACGGAGGAGGGCACTTCCGAATCCGCCACCCC

GGAGTCAGGGCCAGGATCTGAACCCGCTACCTCAGGCAGTGAGACGCCAGGAACGAGCGAG

TCCGCTACACCGGAGAGTGGGCCAGGGAGCCCTGCTGGATCTCCTACGTCCACTGAGGAAGG

GTCACCAGCGGGCTCGCCCACCAGCACTGAAGAAGGTGCCTCGAGCCCGCCTGTGCTGAAG

AGGCACCAGCGAGAAATTACCCGGACCACCCTCCAATCGGATCAGGAGGAAATCGACTACG

ACGACACCATCTCGGTGGAAATGAAGAAGGAAGATTTCGATATCTACGACGAGGACGAAAA

TCAGTCCCCTCGCTCATTCCAAAAGAAAACTAGACACTACTTTATCGCCGCGGTGGAAAGAC

TGTGGGACTATGGAATGTCATCCAGCCCTCACGTCCTTCGGAACCGGGCCCAGAGCGGATCG

GTGCCTCAGTTCAAGAAAGTGGTGTTCCAGGAGTTCACCGACGGCAGCTTCACCCAGCCGCT

GTACCGGGGAGAACTGAACGAACACCTGGGCCTGCTCGGTCCCTACATCCGCGCGGAAGTG

GAGGATAACATCATGGTGACCTTCCGTAACCAAGCATCCAGACCTTACTCCTTCTATTCCTCC

CTGATCTCATACGAGGAGGACCAGCGCCAAGGCGCCGAGCCCCGCAAGAACTTCGTCAAGC

CCAACGAGACTAAGACCTACTTCTGGAAGGTCCAACACCATATGGCCCCGACCAAGGATGA

GTTTGACTGCAAGGCCTGGGCCTACTTCTCCGACGTGGACCTTGAGAAGGATGTCCATTCCG

GCCTGATCGGGCCGCTGCTCGTGTGTCACACCAACACCCTGAACCCAGCGCATGGACGCCAG

GTCACCGTCCAGGAGTTTGCTCTGTTCTTCACCATTTTTGACGAAACTAAGTCCTGGTACTTC

ACCGAGAATATGGAGCGAAACTGTAGAGCGCCCTGCAATATCCAGATGGAAGATCCGACTT

TCAAGGAGAACTATAGATTCCACGCCATCAACGGGTACATCATGGATACTCTGCCGGGGCTG

GTCATGGCCCAGGATCAGAGGATTCGGTGGTACTTGCTGTCAATGGGATCGAACGAAAACAT

TCACTCCATTCACTTCTCCGGTCACGTGTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGG

CGCTGTACAATCTGTACCCCGGGGTGTTCGAAACTGTGGAGATGCTGCCGTCCAAGGCCGGC

ATCTGGAGAGTGGAGTGCCTGATCGGAGAGCACCTCCACGCGGGGATGTCCACCCTCTTCCT

GGTGTACTCGAATAAGTGCCAGACCCCGCTGGGCATGGCCTCGGGCCACATCAGAGACTTCC

AGATCACAGCAAGCGGACAATACGGCCAATGGGCGCCGAAGCTGGCCCGCTTGCACTACTC

CGGATCGATCAACGCATGGTCCACCAAGGAACCGTTCTCGTGGATTAAGGTGGACCTCCTGG

CCCCTATGATTATCCACGGAATTAAGACCCAGGGCGCCAGGCAGAAGTTCTCCTCCCTGTAC

ATCTCGCAATTCATCATCATGTACAGCCTGGACGGGAAGAAGTGGCAGACTTACAGGGGAA

ACTCCACCGGCACCCTGATGGTCTTTTTCGGCAACGTGGATTCCTCCGGCATTAAGCACAAC

ATCTTCAACCCACCGATCATAGCCAGATATATTAGGCTCCACCCCACTCACTACTCAATCCGC

TCAACTCTTCGGATGGAACTCATGGGGTGCGACCTGAACTCCTGCTCCATGCCGTTGGGGAT

GGAATCAAAGGCTATTAGCGACGCCCAGATCACCGCGAGCTCCTACTTCACTAACATGTTCG

CCACCTGGAGCCCCTCCAAGGCCAGGCTGCACTTGCAGGGACGGTCAAATGCCTGGCGGCCG

CAAGTGAACAATCCGAAGGAATGGCTTCAAGTGGATTTCCAAAAGACCATGAAAGTGACCG

GAGTCACCACCCAGGGAGTGAAGTCCCTTCTGACCTCGATGTATGTGAAGGAGTTCCTGATT

AGCAGCAGCCAGGACGGGCACCAGTGGACCCTGTTCTTCCAAAACGGAAAGGTCAAGGTGT

TCCAGGGGAACCAGGACTCGTTCACACCCGTGGTGAACTCCCTGGACCCCCCACTGCTGACG

CGGTACTTGAGGATTCATCCTCAGTCCTGGGTCCATCAGATTGCATTGCGAATGGAAGTCCT

GGGCTGCGAGGCCCAGGACCTGTACTGAATCAGCCTGAGCTCGCTGATCATAATCAACCTCT

GGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATG

TGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCC

TCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGT

GGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTG

TCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGC

CTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTC

GGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGAC

GTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCC

GGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGC

CGCCTCCCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCT CCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGG AAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGAC AGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGG CTTCTGAGGCGGAAAGAACGGGCTCGAGAAGCTTCTAGATATCCTCTCTTAAGGTAGCATCG AGATTTAAATTAGGGATAACAGGGTAATGGCGCGGGCCGCCACTTCCTGGCGCGCAAAATAT CCTCTTGTCCTTGAGTCTCATTGGAGGGTTCGTTCGTTCGAACCAGCCAATCAGGGGAGGGG GAAGTGACGCAAGTTCCGGTCACATGCTTCCGGTGACGCACATCCGGTGACGTAGTTCCGGT CACGTGCTTCCTGTCACGTGTTTCCGGTCACGTGACTTCCGGTCATGTGACTTCCGGTGACGT GTTTCCGGCTTAACTATTGGGCTGACCGCGCGGCATGCGCGTGGTCAACCTAACAGCCGGAA ACACGTCACCG

表2D：帶有全長B19 ITR的例示性B19-FVIII建構體（核苷酸1-7032；SEQ ID NO: 189）

描述	序列
5' ITR（SEQ ID NO: 185）	CCAAATCAGATGCCGCCGGTCGCCGCCGGTAGGCGGGACTTCCGGTACAA GATGGCGGACAATTACGTCATTTCCTGTGACGTCATTTCCTGTGACGTCAC TTCCGGTGGGCGGGACTTCCGGAATTAGGGTTGGCTCTGGGCCAGCTTGCT TGGGGTTGCCTTGACACTAAGACAAGCGGCGCGCCGCTTGATCTTAGTGGC ACGTCAACCCCAAGCGCTGGCCCAGAGCCAACCCTAATTCCGGAAGTCCC GCCCACCGGAAGTGACGTCACAGGAAATGACGTCACAGGAAATGACGTAA TTGTCCGCCATCTTGTACCGGAAGTCCCGCCTACCGGCGGCGACCGGCGGC ATCTGATTTGGTGTCTTCTTTTAAATTTT
TTPp（肝臟特異性啟動	ACGCGTGTCTGTCTGCACATTTCGTAGAGCGAGTGTTCCGATACTCTAATC TCCCTAGGCAAGGTTCATATTTGTGTAGGTTACTTATTCTCCTTTTGTTGAC

子)(SEQ ID NO: 113)	TAAGTCAATAATCAGAATCAGCAGGTTTGGAGTCAGCTTGGCAGGGATCA GCAGCCTGGGTTGGAAGGAGGGGGTATAAAAGCCCCTTCACCAGGAGAAG CCGTCACACAGATCCACAAGCTCCTG
合成內含子(SEQ ID NO: 115)	GTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGC CCTTGCGTGCCTTGAATTACTGACACTGACATCCACTTTTTCTTTTTCTCCA CAG
FVIIIco6XTEN(SEQ ID NO: 117)(含有XTEN144的密碼子優化的FVIII形式6的開放閱讀框;XTEN序列通過加雙下劃線(SEQ ID NO: 118)標記)	ATGCAGATTGAGCTGTCCACTTGTTTCTTCCTGTGCCTCCTGCGCTTCTGTT TCTCCGCCACTCGCCGGTACTACCTTGGAGCCGTGGAGCTTTCATGGGACT ACATGCAGAGCGACCTGGGCGAACTCCCCGTGGATGCCAGATTCCCCCCCC GCGTGCCAAAGTCCTTCCCCTTTAACACCTCCGTGGTGTACAAGAAAACCC TCTTTGTCGAGTTCACTGACCACCTGTTCAACATCGCCAAGCCGCGCCCAC CTTGGATGGGCCTCCTGGGACCGACCATTCAAGCTGAAGTGTACGACACCG TGGTGATCACCCTGAAGAACATGGCGTCCCACCCCGTGTCCCTGCATGCGG TCGGAGTGTCCTACTGGAAGGCCTCCGAAGGAGCTGAGTACGACGACCAG ACTAGCCAGCGGGAAAAGGAGGACGATAAAGTGTTCCCGGGCGGCTCGCA TACTTACGTGTGGCAAGTCCTGAAGGAAAACGGACCTATGGCATCCGATCC TCTGTGCCTGACTTACTCCTACCTTTCCCATGTGGACCTCGTGAAGGACCTG AACAGCGGGCTGATTGGTGCACTTCTCGTGTGCCGCGAAGGTTCGCTCGCT AAGGAAAAGACCCAGACCCTCCATAAGTTCATCCTTTTGTTCGCTGTGTTC GATGAAGGAAAGTCATGGCATTCCGAAACTAAGAACTCGCTGATGCAGGA CCGGGATGCCGCCTCAGCCCGCGCCTGGCCTAAAATGCATACAGTCAACG GATACGTGAATCGGTCACTGCCCGGGCTCATCGGTTGTCACAGAAAGTCCG TGTACTGGCACGTCATCGGCATGGGCACTACGCCTGAAGTGCACTCCATCT TCCTGGAAGGGCACACCTTCCTCGTGCGCAACCACCGCCAGGCCTCTCTGG AAATCTCCCCGATTACCTTTCTGACCGCCCAGACTCTGCTCATGGACCTGG

	GGCAGTTCCTTCTCTTCTGCCACATCTCCAGCCATCAGCACGACGGAATGG
	AGGCCTACGTGAAGGTGGACTCATGCCCGGAAGAACCTCAGTTGCGGATG
	AAGAACAACGAGGAGGCCGAGGACTATGACGACGATTTGACTGACTCCGA
	GATGGACGTCGTGCGGTTCGATGACGACAACAGCCCCAGCTTCATCCAGAT
	TCGCAGCGTGGCCAAGAAGCACCCCAAAACCTGGGTGCACTACATCGCGG
	CCGAGGAAGAAGATTGGGACTACGCCCCGTTGGTGCTGGCACCCGATGAC
	CGGTCGTACAAGTCCCAGTATCTGAACAATGGTCCGCAGCGGATTGGCAG
	AAAGTACAAGAAAGTGCGGTTCATGGCGTACACTGACGAAACGTTTAAGA
	CCCGGGAGGCCATTCAACATGAGAGCGGCATTCTGGGACCACTGCTGTAC
	GGAGAGGTCGGCGATACCCTGCTCATCATCTTCAAAAACCAGGCCTCCCGG
	CCTTACAACATCTACCCTCACGGAATCACCGACGTGCGGCCACTCTACTCG
	CGGCGCCTGCCGAAGGGCGTCAAGCACCTGAAAGACTTCCCTATCCTGCCG
	GGCGAAATCTTCAAGTATAAGTGGACCGTCACCGTGGAGGACGGGCCCAC
	CAAGAGCGATCCTAGGTGTCTGACTCGGTACTACTCCAGCTTCGTGAACAT
	GGAACGGGACCTGGCATCGGGACTCATTGGACCGCTGCTGATCTGCTACA
	AAGAGTCGGTGGATCAACGCGGCAACCAGATCATGTCCGACAAGCGCAAC
	GTGATCCTGTTCTCCGTGTTTGATGAAAACAGATCCTGGTACCTCACTGAA
	AACATCCAGAGGTTCCTCCCAAACCCCGCAGGAGTGCAACTGGAGGACCC
	TGAGTTTCAGGCCTCGAATATCATGCACTCGATTAACGGTTACGTGTTCGA
	CTCGCTGCAACTGAGCGTGTGCCTCCATGAAGTCGCTTACTGGTACATTCT
	GTCCATCGGCGCCCAGACTGACTTCCTGAGCGTGTTCTTTTCCGGTTACACC
	TTTAAGCACAAGATGGTGTACGAAGATACCCTGACCCTGTTCCCTTTCTCC
	GGCGAAACGGTGTTCATGTCGATGGAGAACCCGGGTCTGTGGATTCTGGG
	ATGCCACAACAGCGACTTTCGGAACCGCGGAATGACTGCCCTGCTGAAGG
	TGTCCTCATGCGACAAGAACACCGGAGACTACTACGAGGACTCCTACGAG

	GATATCTCAGCCTACCTCCTGTCCAAGAACAACGCGATCGAGCCGCGCAGC
	TTCAGCCAGAACGGCGCGCCAACATCAGAGAGCGCCACCCCTGAAAGTGG
	TCCCGGGAGCGAGCCAGCCACATCTGGGTCGGAAACGCCAGGCACAAGTG
	AGTCTGCAACTCCCGAGTCCGGACCTGGCTCCGAGCCTGCCACTAGCGGCT
	CCGAGACTCCGGGAACTTCCGAGAGCGCTACACCAGAAAGCGGACCCGGA
	ACCAGTACCGAACCTAGCGAGGGCTCTGCTCCGGGCAGCCCAGCCGGCTC
	TCCTACATCCACGGAGGAGGGCACTTCCGAATCCGCCACCCCGGAGTCAG
	GGCCAGGATCTGAACCCGCTACCTCAGGCAGTGAGACGCCAGGAACGAGC
	GAGTCCGCTACACCGGAGAGTGGGCCAGGGAGCCCTGCTGGATCTCCTAC
	GTCCACTGAGGAAGGGTCACCAGCGGGCTCGCCCACCAGCACTGAAGAAG
	GTGCCTCGAGCCCGCCTGTGCTGAAGAGGCACCAGCGAGAAATTACCCGG
	ACCACCCTCCAATCGGATCAGGAGGAAATCGACTACGACGACACCATCTC
	GGTGGAAATGAAGAAGGAAGATTTCGATATCTACGACGAGGACGAAAATC
	AGTCCCCTCGCTCATTCCAAAAGAAAACTAGACACTACTTTATCGCCGCGG
	TGGAAAGACTGTGGGACTATGGAATGTCATCCAGCCCTCACGTCCTTCGGA
	ACCGGGCCCAGAGCGGATCGGTGCCTCAGTTCAAGAAAGTGGTGTTCCAG
	GAGTTCACCGACGGCAGCTTCACCCAGCCGCTGTACCGGGGAGAACTGAA
	CGAACACCTGGGCCTGCTCGGTCCCTACATCCGCGCGGAAGTGGAGGATA
	ACATCATGGTGACCTTCCGTAACCAAGCATCCAGACCTTACTCCTTCTATTC
	CTCCCTGATCTCATACGAGGAGGACCAGCGCCAAGGCGCCGAGCCCCGCA
	AGAACTTCGTCAAGCCCAACGAGACTAAGACCTACTTCTGGAAGGTCCAA
	CACCATATGGCCCCGACCAAGGATGAGTTTGACTGCAAGGCCTGGGCCTA
	CTTCTCCGACGTGGACCTTGAGAAGGATGTCCATTCCGGCCTGATCGGGCC
	GCTGCTCGTGTGTCACACCAACACCCTGAACCCAGCGCATGGACGCCAGGT
	CACCGTCCAGGAGTTTGCTCTGTTCTTCACCATTTTTGACGAAACTAAGTCC

	TGGTACTTCACCGAGAATATGGAGCGAAACTGTAGAGCGCCCTGCAATAT
	CCAGATGGAAGATCCGACTTTCAAGGAGAACTATAGATTCCACGCCATCA
	ACGGGTACATCATGGATACTCTGCCGGGGCTGGTCATGGCCCAGGATCAG
	AGGATTCGGTGGTACTTGCTGTCAATGGGATCGAACGAAAACATTCACTCC
	ATTCACTTCTCCGGTCACGTGTTCACTGTGCGCAAGAAGGAGGAGTACAAG
	ATGGCGCTGTACAATCTGTACCCCGGGGTGTTCGAAACTGTGGAGATGCTG
	CCGTCCAAGGCCGGCATCTGGAGAGTGGAGTGCCTGATCGGAGAGCACCT
	CCACGCGGGGATGTCCACCCTCTTCCTGGTGTACTCGAATAAGTGCCAGAC
	CCCGCTGGGCATGGCCTCGGGCCACATCAGAGACTTCCAGATCACAGCAA
	GCGGACAATACGGCCAATGGGCGCCGAAGCTGGCCCGCTTGCACTACTCC
	GGATCGATCAACGCATGGTCCACCAAGGAACCGTTCTCGTGGATTAAGGT
	GGACCTCCTGGCCCCTATGATTATCCACGGAATTAAGACCCAGGGCGCCAG
	GCAGAAGTTCTCCTCCCTGTACATCTCGCAATTCATCATCATGTACAGCCT
	GGACGGGAAGAAGTGGCAGACTTACAGGGGAAACTCCACCGGCACCCTGA
	TGGTCTTTTTCGGCAACGTGGATTCCTCCGGCATTAAGCACAACATCTTCA
	ACCCACCGATCATAGCCAGATATATTAGGCTCCACCCCACTCACTACTCAA
	TCCGCTCAACTCTTCGGATGGAACTCATGGGGTGCGACCTGAACTCCTGCT
	CCATGCCGTTGGGGATGGAATCAAAGGCTATTAGCGACGCCCAGATCACC
	GCGAGCTCCTACTTCACTAACATGTTCGCCACCTGGAGCCCCTCCAAGGCC
	AGGCTGCACTTGCAGGGACGGTCAAATGCCTGGCGGCCGCAAGTGAACAA
	TCCGAAGGAATGGCTTCAAGTGGATTTCCAAAAGACCATGAAAGTGACCG
	GAGTCACCACCCAGGGAGTGAAGTCCCTTCTGACCTCGATGTATGTGAAGG
	AGTTCCTGATTAGCAGCAGCCAGGACGGGCACCAGTGGACCCTGTTCTTCC
	AAAACGGAAAGGTCAAGGTGTTCCAGGGGAACCAGGACTCGTTCACACCC
	GTGGTGAACTCCCTGGACCCCCCACTGCTGACGCGGTACTTGAGGATTCAT

	CCTCAGTCCTGGGTCCATCAGATTGCATTGCGAATGGAAGTCCTGGGCTGC GAGGCCCAGGACCTGTACTGA
WPRE(突變的土撥鼠肝炎病毒轉錄後調控元件)(SEQ ID NO: 120)	TCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCT TAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTG TATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATC CTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGG CGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGC CACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCC ACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGG CTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTT CCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCT GCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCT GCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCG GATCTCCCTTTGGGCCGCCTCCCCGCTG
bGHpA(牛生長激素多聚腺苷酸化信號)(SEQ ID NO: 122)	CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCC TTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGA GGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGG GGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCAT GCTGGGGA
3' ITR反向末端重複體(SEQ ID NO: 186)	AAAATTTAAAAGAAGACACCAAATCAGATGCCGCCGGTCGCCGCCGGTAG GCGGGACTTCCGGTACAAGATGGCGGACAATTACGTCATTTCCTGTGACGT CATTTCCTGTGACGTCACTTCCGGTGGGCGGGACTTCCGGAATTAGGGTTG GCTCTGGGCCAGCGCTTGGGGTTGACGTGCCACTAAGATCAAGCGGCGCG CCGCTTGTCTTAGTGTCAAGGCAACCCCAAGCAAGCTGGCCCAGAGCCAA CCCTAATTCCGGAAGTCCCGCCCACCGGAAGTGACGTCACAGGAAATGAC

	GTCACAGGAAATGACGTAATTGTCCGCCATCTTGTACCGGAAGTCCCGCCT ACCGGCGGCGACCGGCGGCATCTGATTTGG

全長序列（SEQ ID NO: 189）

CCAAATCAGATGCCGCCGGTCGCCGCCGGTAGGCGGGACTTCCGGTACAAGATGGCGGACAAT

TACGTCATTTCCTGTGACGTCATTTCCTGTGACGTCACTTCCGGTGGGCGGGACTTCCGGAATT

AGGGTTGGCTCTGGGCCAGCTTGCTTGGGGTTGCCTTGACACTAAGACAAGCGGCGCGCCGCTT

GATCTTAGTGGCACGTCAACCCCAAGCGCTGGCCCAGAGCCAACCCTAATTCCGGAAGTCCCG

CCCACCGGAAGTGACGTCACAGGAAATGACGTCACAGGAAATGACGTAATTGTCCGCCATCTT

GTACCGGAAGTCCCGCCTACCGGCGGCGACCGGCGGCATCTGATTTGGTGTCTTCTTTTAAATT

TTGCGGCAATTCAGTCGATAACTATAACGGTCCTAAGGTAGCGATTTAAATACGCGCTCTCTTA

AGGTAGCCCCGGGACGCGTCAATTGAGATCTGGATCCGGTACCGAATTCGCGGCCGCCTCGAC

GACTAGCGTTTAATTAAACGCGTGTCTGTCTGCACATTTCGTAGAGCGAGTGTTCCGATACTCT

AATCTCCCTAGGCAAGGTTCATATTTGTGTAGGTTACTTATTCTCCTTTTGTTGACTAAGTCAAT

AATCAGAATCAGCAGGTTTGGAGTCAGCTTGGCAGGGATCAGCAGCCTGGGTTGGAAGGAGGG

GGTATAAAAGCCCCTTCACCAGGAGAAGCCGTCACACAGATCCACAAGCTCCTGAGGTAAGTG

CCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTTGCGTGCCTTGAATTACT

GACACTGACATCCACTTTTTCTTTTTCTCCACAGCTAGCGCCACCATGCAGATTGAGCTGTCCAC

TTGTTTCTTCCTGTGCCTCCTGCGCTTCTGTTTCTCCGCCACTCGCCGGTACTACCTTGGAGCCG

TGGAGCTTTCATGGGACTACATGCAGAGCGACCTGGGCGAACTCCCCGTGGATGCCAGATTCC

CCCCCCGCGTGCCAAAGTCCTTCCCCTTTAACACCTCCGTGGTGTACAAGAAAACCCTCTTTGT

CGAGTTCACTGACCACCTGTTCAACATCGCCAAGCCGCGCCCACCTTGGATGGGCCTCCTGGGA

CCGACCATTCAAGCTGAAGTGTACGACACCGTGGTGATCACCCTGAAGAACATGGCGTCCCAC

CCCGTGTCCCTGCATGCGGTCGGAGTGTCCTACTGGAAGGCCTCCGAAGGAGCTGAGTACGAC

GACCAGACTAGCCAGCGGGAAAAGGAGGACGATAAAGTGTTCCCGGGCGGCTCGCATACTTAC

GTGTGGCAAGTCCTGAAGGAAAACGGACCTATGGCATCCGATCCTCTGTGCCTGACTTACTCCT

ACCTTTCCCATGTGGACCTCGTGAAGGACCTGAACAGCGGGCTGATTGGTGCACTTCTCGTGTG

CCGCGAAGGTTCGCTCGCTAAGGAAAAGACCCAGACCCTCCATAAGTTCATCCTTTTGTTCGCT

GTGTTCGATGAAGGAAAGTCATGGCATTCCGAAACTAAGAACTCGCTGATGCAGGACCGGGAT

GCCGCCTCAGCCCGCGCCTGGCCTAAAATGCATACAGTCAACGGATACGTGAATCGGTCACTG

CCCGGGCTCATCGGTTGTCACAGAAAGTCCGTGTACTGGCACGTCATCGGCATGGGCACTACG

CCTGAAGTGCACTCCATCTTCCTGGAAGGGCACACCTTCCTCGTGCGCAACCACCGCCAGGCCT

CTCTGGAAATCTCCCCGATTACCTTTCTGACCGCCCAGACTCTGCTCATGGACCTGGGGCAGTT

CCTTCTCTTCTGCCACATCTCCAGCCATCAGCACGACGGAATGGAGGCCTACGTGAAGGTGGAC

TCATGCCCGGAAGAACCTCAGTTGCGGATGAAGAACAACGAGGAGGCCGAGGACTATGACGA

CGATTTGACTGACTCCGAGATGGACGTCGTGCGGTTCGATGACGACAACAGCCCCAGCTTCATC

CAGATTCGCAGCGTGGCCAAGAAGCACCCCAAAACCTGGGTGCACTACATCGCGGCCGAGGAA

GAAGATTGGGACTACGCCCCGTTGGTGCTGGCACCCGATGACCGGTCGTACAAGTCCCAGTAT

CTGAACAATGGTCCGCAGCGGATTGGCAGAAAGTACAAGAAAGTGCGGTTCATGGCGTACACT

GACGAAACGTTTAAGACCCGGGAGGCCATTCAACATGAGAGCGGCATTCTGGGACCACTGCTG

TACGGAGAGGTCGGCGATACCCTGCTCATCATCTTCAAAAACCAGGCCTCCCGGCCTTACAAC

ATCTACCCTCACGGAATCACCGACGTGCGGCCACTCTACTCGCGGCGCCTGCCGAAGGGCGTC

AAGCACCTGAAAGACTTCCCTATCCTGCCGGGCGAAATCTTCAAGTATAAGTGGACCGTCACC

GTGGAGGACGGGCCCACCAAGAGCGATCCTAGGTGTCTGACTCGGTACTACTCCAGCTTCGTG

AACATGGAACGGGACCTGGCATCGGGACTCATTGGACCGCTGCTGATCTGCTACAAAGAGTCG

GTGGATCAACGCGGCAACCAGATCATGTCCGACAAGCGCAACGTGATCCTGTTCTCCGTGTTTG

ATGAAAACAGATCCTGGTACCTCACTGAAAACATCCAGAGGTTCCTCCCAAACCCCGCAGGAG

TGCAACTGGAGGACCCTGAGTTTCAGGCCTCGAATATCATGCACTCGATTAACGGTTACGTGTT

CGACTCGCTGCAACTGAGCGTGTGCCTCCATGAAGTCGCTTACTGGTACATTCTGTCCATCGGC

GCCCAGACTGACTTCCTGAGCGTGTTCTTTTCCGGTTACACCTTTAAGCACAAGATGGTGTACG

AAGATACCCTGACCCTGTTCCCTTTCTCCGGCGAAACGGTGTTCATGTCGATGGAGAACCCGGG

TCTGTGGATTCTGGGATGCCACAACAGCGACTTTCGGAACCGCGGAATGACTGCCCTGCTGAA

GGTGTCCTCATGCGACAAGAACACCGGAGACTACTACGAGGACTCCTACGAGGATATCTCAGC

CTACCTCCTGTCCAAGAACAACGCGATCGAGCCGCGCAGCTTCAGCCAGAACGGCGCGCCAAC

ATCAGAGAGCGCCACCCCTGAAAGTGGTCCCGGGAGCGAGCCAGCCACATCTGGGTCGGAAAC

GCCAGGCACAAGTGAGTCTGCAACTCCCGAGTCCGGACCTGGCTCCGAGCCTGCCACTAGCGG

CTCCGAGACTCCGGGAACTTCCGAGAGCGCTACACCAGAAAGCGGACCCGGAACCAGTACCGA

ACCTAGCGAGGGCTCTGCTCCGGGCAGCCCAGCCGGCTCTCCTACATCCACGGAGGAGGGCAC

TTCCGAATCCGCCACCCCGGAGTCAGGGCCAGGATCTGAACCCGCTACCTCAGGCAGTGAGAC

GCCAGGAACGAGCGAGTCCGCTACACCGGAGAGTGGGCCAGGGAGCCCTGCTGGATCTCCTAC

GTCCACTGAGGAAGGGTCACCAGCGGGCTCGCCCACCAGCACTGAAGAAGGTGCCTCGAGCCC

GCCTGTGCTGAAGAGGCACCAGCGAGAAATTACCCGGACCACCCTCCAATCGGATCAGGAGGA

AATCGACTACGACGACACCATCTCGGTGGAAATGAAGAAGGAAGATTTCGATATCTACGACGA

GGACGAAAATCAGTCCCCTCGCTCATTCCAAAAGAAAACTAGACACTACTTTATCGCCGCGGT

GGAAAGACTGTGGGACTATGGAATGTCATCCAGCCCTCACGTCCTTCGGAACCGGGCCCAGAG

CGGATCGGTGCCTCAGTTCAAGAAAGTGGTGTTCCAGGAGTTCACCGACGGCAGCTTCACCCA

GCCGCTGTACCGGGGAGAACTGAACGAACACCTGGGCCTGCTCGGTCCCTACATCCGCGCGGA

AGTGGAGGATAACATCATGGTGACCTTCCGTAACCAAGCATCCAGACCTTACTCCTTCTATTCC

TCCCTGATCTCATACGAGGAGGACCAGCGCCAAGGCGCCGAGCCCCGCAAGAACTTCGTCAAG

CCCAACGAGACTAAGACCTACTTCTGGAAGGTCCAACACCATATGGCCCCGACCAAGGATGAG

TTTGACTGCAAGGCCTGGGCCTACTTCTCCGACGTGGACCTTGAGAAGGATGTCCATTCCGGCC

TGATCGGGCCGCTGCTCGTGTGTCACACCAACACCCTGAACCCAGCGCATGGACGCCAGGTCA

CCGTCCAGGAGTTTGCTCTGTTCTTCACCATTTTTGACGAAACTAAGTCCTGGTACTTCACCGAG

AATATGGAGCGAAACTGTAGAGCGCCCTGCAATATCCAGATGGAAGATCCGACTTTCAAGGAG

AACTATAGATTCCACGCCATCAACGGGTACATCATGGATACTCTGCCGGGGCTGGTCATGGCCC

AGGATCAGAGGATTCGGTGGTACTTGCTGTCAATGGGATCGAACGAAAACATTCACTCCATTC

ACTTCTCCGGTCACGTGTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCGCTGTACAATC

TGTACCCCGGGGGTGTTCGAAACTGTGGAGATGCTGCCGTCCAAGGCCGGCATCTGGAGAGTGG

AGTGCCTGATCGGAGAGCACCTCCACGCGGGGATGTCCACCCTCTTCCTGGTGTACTCGAATAA

GTGCCAGACCCCGCTGGGCATGGCCTCGGGCCACATCAGAGACTTCCAGATCACAGCAAGCGG

ACAATACGGCCAATGGGCGCCGAAGCTGGCCCGCTTGCACTACTCCGGATCGATCAACGCATG

GTCCACCAAGGAACCGTTCTCGTGGATTAAGGTGGACCTCCTGGCCCCTATGATTATCCACGGA

ATTAAGACCCAGGGCGCCAGGCAGAAGTTCTCCTCCCTGTACATCTCGCAATTCATCATCATGT

ACAGCCTGGACGGGAAGAAGTGGCAGACTTACAGGGGAAACTCCACCGGCACCCTGATGGTCT

TTTTCGGCAACGTGGATTCCTCCGGCATTAAGCACAACATCTTCAACCCACCGATCATAGCCAG

ATATATTAGGCTCCACCCCACTCACTACTCAATCCGCTCAACTCTTCGGATGGAACTCATGGGG

TGCGACCTGAACTCCTGCTCCATGCCGTTGGGGATGGAATCAAAGGCTATTAGCGACGCCCAG

ATCACCGCGAGCTCCTACTTCACTAACATGTTCGCCACCTGGAGCCCCTCCAAGGCCAGGCTGC

ACTTGCAGGGACGGTCAAATGCCTGGCGGCCGCAAGTGAACAATCCGAAGGAATGGCTTCAAG

TGGATTTCCAAAAGACCATGAAAGTGACCGGAGTCACCACCCAGGGAGTGAAGTCCCTTCTGA

CCTCGATGTATGTGAAGGAGTTCCTGATTAGCAGCAGCCAGGACGGGCACCAGTGGACCCTGT

TCTTCCAAAACGGAAAGGTCAAGGTGTTCCAGGGGAACCAGGACTCGTTCACACCCGTGGTGA

ACTCCCTGGACCCCCCACTGCTGACGCGGTACTTGAGGATTCATCCTCAGTCCTGGGTCCATCA

GATTGCATTGCGAATGGAAGTCCTGGGCTGCGAGGCCCAGGACCTGTACTGAATCAGCCTGAG

CTCGCTGATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACT

ATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCC

CGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTG

GCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGG

GGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGC

GGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAAT

TCCGTGGTGTTGTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGAT

TCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCG

GCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCC

CTTTGGGCCGCCTCCCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTT

GCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAAT

GAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAG

GACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTAT

GGCTTCTGAGGCGGAAAGAACGGGCTCGAGAAGCTTCTAGATATCCTCTCTTAAGGTAGCATC

GAGATTTAAATTAGGGATAACAGGGTAATGGCGCGGGCCGCAAAATTTAAAAGAAGACACCA

AATCAGATGCCGCCGGTCGCCGCCGGTAGGCGGGACTTCCGGTACAAGATGGCGGACAATTAC

GTCATTTCCTGTGACGTCATTTCCTGTGACGTCACTTCCGGTGGGCGGGACTTCCGGAATTAGG

GTTGGCTCTGGGCCAGCGCTTGGGGTTGACGTGCCACTAAGATCAAGCGGCGCGCCGCTTGTCT

TAGTGTCAAGGCAACCCCAAGCAAGCTGGCCCAGAGCCAACCCTAATTCCGGAAGTCCCGCCC

ACCGGAAGTGACGTCACAGGAAATGACGTCACAGGAAATGACGTAATTGTCCGCCATCTTGTA

CCGGAAGTCCCGCCTACCGGCGGCGACCGGCGGCATCTGATTTGG

表2E：例示性AAV-FVIII建構體（核苷酸1-6824；SEQ ID NO: 190）

描述	序列
5' ITR（SEQ ID NO: 111）	CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCG GGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGG GAGTGGCCAACTCCATCACTAGGGGTTCCT
CAGp（泛在啟動子）（SEQ ID	TCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCATCTCCCCCCCCTCCCC ACCCCCAATTTTGTATTTATTTATTTTTTAATTATTTTGTGCAGCGATGGGG GCGGGGGGGGGGGGGGGGCGCGCCAGGCGGGGCGGGGCGGGGCGAGG

NO: 191）	GGCGGGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGG CGCGCTCCGAAAGTTTCCTTTTATGGCGAGGCGGCGGCGGCGGCGGCCCTA TAAAAAGCGAAGCGCGCGGCGGGCG
合成內含子（SEQ ID NO: 192）	GTGAGCGGGCGGGACGGCCCTTCTCCTTCGGGCTGTAATTAGCGCTTGGTT TAATGACGGCTTGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTGAGGGGCTC CGGGAGGGCCCTTTGTGCGGGGGGAGCGGCTCGGGGCTGTCCGCGGGGGG ACGGCTGCCTTCGGGGGGGACGGGGCAGGGCGGGGTTCGGCTTCTGGCGT GTGACCGGCGGCTCTAGAGCCTCTGCTAACCATGTTCATGCCTTCTTCTTTT TCCTACAG
FVIIIco6XTEN（SEQ ID NO: 117）(含有XTEN144的密碼子優化的FVIII形式6的開放閱讀框；XTEN序列通過加雙下劃線（SEQ ID NO: 118）標記）	ATGCAGATTGAGCTGTCCACTTGTTTCTTCCTGTGCCTCCTGCGCTTCTGTT TCTCCGCCACTCGCCGGTACTACCTTGGAGCCGTGGAGCTTTCATGGGACT ACATGCAGAGCGACCTGGGCGAACTCCCCGTGGATGCCAGATTCCCCCCCC GCGTGCCAAAGTCCTTCCCCTTTAACACCTCCGTGGTGTACAAGAAAACCC TCTTTGTCGAGTTCACTGACCACCTGTTCAACATCGCCAAGCCGCGCCCAC CTTGGATGGGCCTCCTGGGACCGACCATTCAAGCTGAAGTGTACGACACCG TGGTGATCACCCTGAAGAACATGGCGTCCCACCCCGTGTCCCTGCATGCGG TCGGAGTGTCCTACTGGAAGGCCTCCGAAGGAGCTGAGTACGACGACCAG ACTAGCCAGCGGGAAAAGGAGGACGATAAAGTGTTCCCGGGCGGCTCGCA TACTTACGTGTGGCAAGTCCTGAAGGAAAACGGACCTATGGCATCCGATCC TCTGTGCCTGACTTACTCCTACCTTTCCCATGTGGACCTCGTGAAGGACCTG AACAGCGGGCTGATTGGTGCACTTCTCGTGTGCCGCGAAGGTTCGCTCGCT AAGGAAAAGACCCAGACCCTCCATAAGTTCATCCTTTTGTTCGCTGTGTTC GATGAAGGAAAGTCATGGCATTCCGAAACTAAGAACTCGCTGATGCAGGA CCGGGATGCCGCCTCAGCCCGCGCCTGGCCTAAAATGCATACAGTCAACG GATACGTGAATCGGTCACTGCCCGGGCTCATCGGTTGTCACAGAAAGTCCG

	TGTACTGGCACGTCATCGGCATGGGCACTACGCCTGAAGTGCACTCCATCT
	TCCTGGAAGGGCACACCTTCCTCGTGCGCAACCACCGCCAGGCCTCTCTGG
	AAATCTCCCCGATTACCTTTCTGACCGCCCAGACTCTGCTCATGGACCTGG
	GGCAGTTCCTTCTCTTCTGCCACATCTCCAGCCATCAGCACGACGGAATGG
	AGGCCTACGTGAAGGTGGACTCATGCCCGGAAGAACCTCAGTTGCGGATG
	AAGAACAACGAGGAGGCCGAGGACTATGACGACGATTTGACTGACTCCGA
	GATGGACGTCGTGCGGTTCGATGACGACAACAGCCCCAGCTTCATCCAGAT
	TCGCAGCGTGGCCAAGAAGCACCCCAAAACCTGGGTGCACTACATCGCGG
	CCGAGGAAGAAGATTGGGACTACGCCCCGTTGGTGCTGGCACCCGATGAC
	CGGTCGTACAAGTCCCAGTATCTGAACAATGGTCCGCAGCGGATTGGCAG
	AAAGTACAAGAAAGTGCGGTTCATGGCGTACACTGACGAAACGTTTAAGA
	CCCGGGAGGCCATTCAACATGAGAGCGGCATTCTGGGACCACTGCTGTAC
	GGAGAGGTCGGCGATACCCTGCTCATCATCTTCAAAAACCAGGCCTCCCGG
	CCTTACAACATCTACCCTCACGGAATCACCGACGTGCGGCCACTCTACTCG
	CGGCGCCTGCCGAAGGGCGTCAAGCACCTGAAAGACTTCCCTATCCTGCCG
	GGCGAAATCTTCAAGTATAAGTGGACCGTCACCGTGGAGGACGGGCCCAC
	CAAGAGCGATCCTAGGTGTCTGACTCGGTACTACTCCAGCTTCGTGAACAT
	GGAACGGGACCTGGCATCGGGACTCATTGGACCGCTGCTGATCTGCTACA
	AAGAGTCGGTGGATCAACGCGGCAACCAGATCATGTCCGACAAGCGCAAC
	GTGATCCTGTTCTCCGTGTTTGATGAAAACAGATCCTGGTACCTCACTGAA
	AACATCCAGAGGTTCCTCCCAAACCCCGCAGGAGTGCAACTGGAGGACCC
	TGAGTTTCAGGCCTCGAATATCATGCACTCGATTAACGGTTACGTGTTCGA
	CTCGCTGCAACTGAGCGTGTGCCTCCATGAAGTCGCTTACTGGTACATTCT
	GTCCATCGGCGCCCAGACTGACTTCCTGAGCGTGTTCTTTTCCGGTTACACC
	TTTAAGCACAAGATGGTGTACGAAGATACCCTGACCCTGTTCCCTTTCTCC

	GGCGAAACGGTGTTCATGTCGATGGAGAACCCGGGTCTGTGGATTCTGGG ATGCCACAACAGCGACTTTCGGAACCGCGGAATGACTGCCCTGCTGAAGG TGTCCTCATGCGACAAGAACACCGGAGACTACTACGAGGACTCCTACGAG GATATCTCAGCCTACCTCCTGTCCAAGAACAACGCGATCGAGCCGCGCAGC TTCAGCCAGAACGGCGCGCCAACATCAGAGAGCGCCACCCCTGAAAGTGG TCCCGGGAGCGAGCCAGCCACATCTGGGTCGGAAACGCCAGGCACAAGTG AGTCTGCAACTCCCGAGTCCGGACCTGGCTCCGAGCCTGCCACTAGCGGCT CCGAGACTCCGGGAACTTCCGAGAGCGCTACACCAGAAAGCGGACCCGGA ACCAGTACCGAACCTAGCGAGGGCTCTGCTCCGGGCAGCCCAGCCGGCTC TCCTACATCCACGGAGGAGGGCACTTCCGAATCCGCCACCCCGGAGTCAG GGCCAGGATCTGAACCCGCTACCTCAGGCAGTGAGACGCCAGGAACGAGC GAGTCCGCTACACCGGAGAGTGGGCCAGGGAGCCCTGCTGGATCTCCTAC GTCCACTGAGGAAGGGTCACCAGCGGGCTCGCCCACCAGCACTGAAGAAG GTGCCTCGAGCCCGCCTGTGCTGAAGAGGCACCAGCGAGAAATTACCCGG ACCACCCTCCAATCGGATCAGGAGGAAATCGACTACGACGACACCATCTC GGTGGAAATGAAGAAGGAAGATTTCGATATCTACGACGAGGACGAAAATC AGTCCCCTCGCTCATTCCAAAAGAAAACTAGACACTACTTTATCGCCGCGG TGGAAAGACTGTGGGACTATGGAATGTCATCCAGCCCTCACGTCCTTCGGA ACCGGGCCCAGAGCGGATCGGTGCCTCAGTTCAAGAAAGTGGTGTTCCAG GAGTTCACCGACGGCAGCTTCACCCAGCCGCTGTACCGGGGAGAACTGAA CGAACACCTGGGCCTGCTCGGTCCCTACATCCGCGCGGAAGTGGAGGATA ACATCATGGTGACCTTCCGTAACCAAGCATCCAGACCTTACTCCTTCTATTC CTCCCTGATCTCATACGAGGAGGACCAGCGCCAAGGCGCCGAGCCCCGCA AGAACTTCGTCAAGCCCAACGAGACTAAGACCTACTTCTGGAAGGTCCAA CACCATATGGCCCCGACCAAGGATGAGTTTGACTGCAAGGCCTGGGCCTA

	CTTCTCCGACGTGGACCTTGAGAAGGATGTCCATTCCGGCCTGATCGGGCC
	GCTGCTCGTGTGTCACACCAACACCCTGAACCCAGCGCATGGACGCCAGGT
	CACCGTCCAGGAGTTTGCTCTGTTCTTCACCATTTTTGACGAAACTAAGTCC
	TGGTACTTCACCGAGAATATGGAGCGAAACTGTAGAGCGCCCTGCAATAT
	CCAGATGGAAGATCCGACTTTCAAGGAGAACTATAGATTCCACGCCATCA
	ACGGGTACATCATGGATACTCTGCCGGGGCTGGTCATGGCCCAGGATCAG
	AGGATTCGGTGGTACTTGCTGTCAATGGGATCGAACGAAAACATTCACTCC
	ATTCACTTCTCCGGTCACGTGTTCACTGTGCGCAAGAAGGAGGAGTACAAG
	ATGGCGCTGTACAATCTGTACCCCGGGGTGTTCGAAACTGTGGAGATGCTG
	CCGTCCAAGGCCGGCATCTGGAGAGTGGAGTGCCTGATCGGAGAGCACCT
	CCACGCGGGGATGTCCACCCTCTTCCTGGTGTACTCGAATAAGTGCCAGAC
	CCCGCTGGGCATGGCCTCGGGCCACATCAGAGACTTCCAGATCACAGCAA
	GCGGACAATACGGCCAATGGGCGCCGAAGCTGGCCCGCTTGCACTACTCC
	GGATCGATCAACGCATGGTCCACCAAGGAACCGTTCTCGTGGATTAAGGT
	GGACCTCCTGGCCCCTATGATTATCCACGGAATTAAGACCCAGGGCGCCAG
	GCAGAAGTTCTCCTCCCTGTACATCTCGCAATTCATCATCATGTACAGCCT
	GGACGGGAAGAAGTGGCAGACTTACAGGGGAAACTCCACCGGCACCCTGA
	TGGTCTTTTTCGGCAACGTGGATTCCTCCGGCATTAAGCACAACATCTTCA
	ACCCACCGATCATAGCCAGATATATTAGGCTCCACCCCACTCACTACTCAA
	TCCGCTCAACTCTTCGGATGGAACTCATGGGGTGCGACCTGAACTCCTGCT
	CCATGCCGTTGGGGATGGAATCAAAGGCTATTAGCGACGCCCAGATCACC
	GCGAGCTCCTACTTCACTAACATGTTCGCCACCTGGAGCCCCTCCAAGGCC
	AGGCTGCACTTGCAGGGACGGTCAAATGCCTGGCGGCCGCAAGTGAACAA
	TCCGAAGGAATGGCTTCAAGTGGATTTCCAAAAGACCATGAAAGTGACCG
	GAGTCACCACCCAGGGAGTGAAGTCCCTTCTGACCTCGATGTATGTGAAGG

	AGTTCCTGATTAGCAGCAGCCAGGACGGGCACCAGTGGACCCTGTTCTTCC AAAACGGAAAGGTCAAGGTGTTCCAGGGGAACCAGGACTCGTTCACACCC GTGGTGAACTCCCTGGACCCCCCACTGCTGACGCGGTACTTGAGGATTCAT CCTCAGTCCTGGGTCCATCAGATTGCATTGCGAATGGAAGTCCTGGGCTGC GAGGCCCAGGACCTGTACTGA
WPRE(突變的土撥鼠肝炎病毒轉錄後調控元件)(SEQ ID NO: 120)	TCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCT TAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTG TATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATC CTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGG CGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGC CACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCC ACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGG CTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTT CCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCT GCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCT GCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCG GATCTCCCTTTGGGCCGCCTCCCCGCTG
bGHpA(牛生長激素多聚腺苷酸化信號)(SEQ ID NO: 122)	CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCC TTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGA GGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGG GGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCAT GCTGGGGA
3' ITR反向末端重複體(SEQ ID	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGC TCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGGCTTTGCCCGG GCGGCCTCAGTGAGCGAGCGAGCGCGCAG

NO: 193）	

全長序列（SEQ ID NO: 190）

CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGGT

CGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGG

TTCCTGCGGCAATTCAGTCGATAACTATAACGGTCCTAAGGTAGCGATTTAAATACGCGCTCTC

TTAAGGTAGCCCCGGGACGCGTCAATTGAGATCTGGATCCGGTACCGAATTCGCGGCCGCCTC

GACGACTAGCGTTTAATTAAATCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCATCTCCCC

CCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTATTTTGTGCAGCGATGGGGGCGG

GGGGGGGGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAG

GCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTTATGGCGAG

GCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGCGCGCT

GCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGC

GTTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTTCGGGCTGTAATTAGCGCTTGGTT

TAATGACGGCTTGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTGAGGGGCTCCGGGAGGGCCCTT

TGTGCGGGGGGAGCGGCTCGGGGCTGTCCGCGGGGGGACGGCTGCCTTCGGGGGGGACGGGG

CAGGGCGGGGTTCGGCTTCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTAACCATGTTCAT

GCCTTCTTCTTTTTCCTACAGGCTAGCGCCACCATGCAGATTGAGCTGTCCACTTGTTTCTTCCT

GTGCCTCCTGCGCTTCTGTTTCTCCGCCACTCGCCGGTACTACCTTGGAGCCGTGGAGCTTTCAT

GGGACTACATGCAGAGCGACCTGGGCGAACTCCCCGTGGATGCCAGATTCCCCCCCCGCGTGC

CAAAGTCCTTCCCCTTTAACACCTCCGTGGTGTACAAGAAAACCCTCTTTGTCGAGTTCACTGA

CCACCTGTTCAACATCGCCAAGCCGCGCCCACCTTGGATGGGCCTCCTGGGACCGACCATTCAA

GCTGAAGTGTACGACACCGTGGTGATCACCCTGAAGAACATGGCGTCCCACCCCGTGTCCCTG

CATGCGGTCGGAGTGTCCTACTGGAAGGCCTCCGAAGGAGCTGAGTACGACGACCAGACTAGC

CAGCGGGAAAAGGAGGACGATAAAGTGTTCCCGGGCGGCTCGCATACTTACGTGTGGCAAGTC

CTGAAGGAAAACGGACCTATGGCATCCGATCCTCTGTGCCTGACTTACTCCTACCTTTCCCATG

TGGACCTCGTGAAGGACCTGAACAGCGGGCTGATTGGTGCACTTCTCGTGTGCCGCGAAGGTT

CGCTCGCTAAGGAAAAGACCCAGACCCTCCATAAGTTCATCCTTTTGTTCGCTGTGTTCGATGA

AGGAAAGTCATGGCATTCCGAAACTAAGAACTCGCTGATGCAGGACCGGGATGCCGCCTCAGC

CCGCGCCTGGCCTAAAATGCATACAGTCAACGGATACGTGAATCGGTCACTGCCCGGGCTCAT

CGGTTGTCACAGAAAGTCCGTGTACTGGCACGTCATCGGCATGGGCACTACGCCTGAAGTGCA

CTCCATCTTCCTGGAAGGGCACACCTTCCTCGTGCGCAACCACCGCCAGGCCTCTCTGGAAATC

TCCCCGATTACCTTTCTGACCGCCCAGACTCTGCTCATGGACCTGGGGCAGTTCCTTCTCTTCTG

CCACATCTCCAGCCATCAGCACGACGGAATGGAGGCCTACGTGAAGGTGGACTCATGCCCGGA

AGAACCTCAGTTGCGGATGAAGAACAACGAGGAGGCCGAGGACTATGACGACGATTTGACTG

ACTCCGAGATGGACGTCGTGCGGTTCGATGACGACAACAGCCCCAGCTTCATCCAGATTCGCA

GCGTGGCCAAGAAGCACCCCAAAACCTGGGTGCACTACATCGCGGCCGAGGAAGAAGATTGG

GACTACGCCCCGTTGGTGCTGGCACCCGATGACCGGTCGTACAAGTCCCAGTATCTGAACAAT

GGTCCGCAGCGGATTGGCAGAAAGTACAAGAAAGTGCGGTTCATGGCGTACACTGACGAAAC

GTTTAAGACCCGGGAGGCCATTCAACATGAGAGCGGCATTCTGGGACCACTGCTGTACGGAGA

GGTCGGCGATACCCTGCTCATCATCTTCAAAAACCAGGCCTCCCGGCCTTACAACATCTACCCT

CACGGAATCACCGACGTGCGGCCACTCTACTCGCGGCGCCTGCCGAAGGGCGTCAAGCACCTG

AAAGACTTCCCTATCCTGCCGGGCGAAATCTTCAAGTATAAGTGGACCGTCACCGTGGAGGAC

GGGCCCACCAAGAGCGATCCTAGGTGTCTGACTCGGTACTACTCCAGCTTCGTGAACATGGAA

CGGGACCTGGCATCGGGACTCATTGGACCGCTGCTGATCTGCTACAAAGAGTCGGTGGATCAA

CGCGGCAACCAGATCATGTCCGACAAGCGCAACGTGATCCTGTTCTCCGTGTTTGATGAAAAC

AGATCCTGGTACCTCACTGAAAACATCCAGAGGTTCCTCCCAAACCCCGCAGGAGTGCAACTG

GAGGACCCTGAGTTTCAGGCCTCGAATATCATGCACTCGATTAACGGTTACGTGTTCGACTCGC

TGCAACTGAGCGTGTGCCTCCATGAAGTCGCTTACTGGTACATTCTGTCCATCGGCGCCCAGAC

TGACTTCCTGAGCGTGTTCTTTTCCGGTTACACCTTTAAGCACAAGATGGTGTACGAAGATACC

CTGACCCTGTTCCCTTTCTCCGGCGAAACGGTGTTCATGTCGATGGAGAACCCGGGTCTGTGGA

TTCTGGGATGCCACAACAGCGACTTTCGGAACCGCGGAATGACTGCCCTGCTGAAGGTGTCCTC

ATGCGACAAGAACACCGGAGACTACTACGAGGACTCCTACGAGGATATCTCAGCCTACCTCCT

GTCCAAGAACAACGCGATCGAGCCGCGCAGCTTCAGCCAGAACGGCGCGCCAACATCAGAGA

GCGCCACCCCTGAAAGTGGTCCCGGGAGCGAGCCAGCCACATCTGGGTCGGAAACGCCAGGCA

CAAGTGAGTCTGCAACTCCCGAGTCCGGACCTGGCTCCGAGCCTGCCACTAGCGGCTCCGAGA

CTCCGGGAACTTCCGAGAGCGCTACACCAGAAAGCGGACCCGGAACCAGTACCGAACCTAGCG

AGGGCTCTGCTCCGGGCAGCCCAGCCGGCTCTCCTACATCCACGGAGGAGGGCACTTCCGAAT

CCGCCACCCCGGAGTCAGGGCCAGGATCTGAACCCGCTACCTCAGGCAGTGAGACGCCAGGAA

CGAGCGAGTCCGCTACACCGGAGAGTGGGCCAGGGAGCCCTGCTGGATCTCCTACGTCCACTG

AGGAAGGGTCACCAGCGGGCTCGCCCACCAGCACTGAAGAAGGTGCCTCGAGCCCGCCTGTGC

TGAAGAGGCACCAGCGAGAAATTACCCGGACCACCCTCCAATCGGATCAGGAGGAAATCGACT

ACGACGACACCATCTCGGTGGAAATGAAGAAGGAAGATTTCGATATCTACGACGAGGACGAA

AATCAGTCCCCTCGCTCATTCCAAAAGAAAACTAGACACTACTTTATCGCCGCGGTGGAAAGA

CTGTGGGACTATGGAATGTCATCCAGCCCTCACGTCCTTCGGAACCGGGCCCAGAGCGGATCG

GTGCCTCAGTTCAAGAAAGTGGTGTTCCAGGAGTTCACCGACGGCAGCTTCACCCAGCCGCTGT

ACCGGGGAGAACTGAACGAACACCTGGGCCTGCTCGGTCCCTACATCCGCGCGGAAGTGGAGG

ATAACATCATGGTGACCTTCCGTAACCAAGCATCCAGACCTTACTCCTTCTATTCCTCCCTGATC

TCATACGAGGAGGACCAGCGCCAAGGCGCCGAGCCCCGCAAGAACTTCGTCAAGCCCAACGA

GACTAAGACCTACTTCTGGAAGGTCCAACACCATATGGCCCCGACCAAGGATGAGTTTGACTG

CAAGGCCTGGGCCTACTTCTCCGACGTGGACCTTGAGAAGGATGTCCATTCCGGCCTGATCGGG

CCGCTGCTCGTGTGTCACACCAACACCCTGAACCCAGCGCATGGACGCCAGGTCACCGTCCAG

GAGTTTGCTCTGTTCTTCACCATTTTTGACGAAACTAAGTCCTGGTACTTCACCGAGAATATGG

AGCGAAACTGTAGAGCGCCCTGCAATATCCAGATGGAAGATCCGACTTTCAAGGAGAACTATA

GATTCCACGCCATCAACGGGTACATCATGGATACTCTGCCGGGGCTGGTCATGGCCCAGGATC

AGAGGATTCGGTGGTACTTGCTGTCAATGGGATCGAACGAAAACATTCACTCCATTCACTTCTC

CGGTCACGTGTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCGCTGTACAATCTGTACCC

CGGGGTGTTCGAAACTGTGGAGATGCTGCCGTCCAAGGCCGGCATCTGGAGAGTGGAGTGCCT

GATCGGAGAGCACCTCCACGCGGGGATGTCCACCCTCTTCCTGGTGTACTCGAATAAGTGCCA

GACCCCGCTGGGCATGGCCTCGGGCCACATCAGAGACTTCCAGATCACAGCAAGCGGACAATA

CGGCCAATGGGCGCCGAAGCTGGCCCGCTTGCACTACTCCGGATCGATCAACGCATGGTCCAC

CAAGGAACCGTTCTCGTGGATTAAGGTGGACCTCCTGGCCCCTATGATTATCCACGGAATTAAG

ACCCAGGGCGCCAGGCAGAAGTTCTCCTCCCTGTACATCTCGCAATTCATCATCATGTACAGCC

TGGACGGGAAGAAGTGGCAGACTTACAGGGGAAACTCCACCGGCACCCTGATGGTCTTTTTCG

GCAACGTGGATTCCTCCGGCATTAAGCACAACATCTTCAACCCACCGATCATAGCCAGATATAT

TAGGCTCCACCCCACTCACTACTCAATCCGCTCAACTCTTCGGATGGAACTCATGGGGTGCGAC

CTGAACTCCTGCTCCATGCCGTTGGGGATGGAATCAAAGGCTATTAGCGACGCCCAGATCACC

GCGAGCTCCTACTTCACTAACATGTTCGCCACCTGGAGCCCCTCCAAGGCCAGGCTGCACTTGC

AGGGACGGTCAAATGCCTGGCGGCCGCAAGTGAACAATCCGAAGGAATGGCTTCAAGTGGATT

TCCAAAAGACCATGAAAGTGACCGGAGTCACCACCCAGGGAGTGAAGTCCCTTCTGACCTCGA

TGTATGTGAAGGAGTTCCTGATTAGCAGCAGCCAGGACGGGCACCAGTGGACCCTGTTCTTCC

AAAACGGAAAGGTCAAGGTGTTCCAGGGGAACCAGGACTCGTTCACACCCGTGGTGAACTCCC

TGGACCCCCCACTGCTGACGCGGTACTTGAGGATTCATCCTCAGTCCTGGGTCCATCAGATTGC

ATTGCGAATGGAAGTCCTGGGCTGCGAGGCCCAGGACCTGTACTGAATCAGCCTGAGCTCGCT

GATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTG

CTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATG

GCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTT

GTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTG

CCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTC

ATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGG
TGTTGTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGC
GGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCT
GCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGG
CCGCCTCCCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTC
CCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAA
ATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGC
AAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCT
GAGGCGGAAAGAACGGGCTCGAGAAGCTTCTAGATATCCTCTCTTAAGGTAGCATCGAGATTT
AAATTAGGGATAACAGGGTAATGGCGCGGGCCGCAGGAACCCCTAGTGATGGAGTTGGCCACT
CCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGGC
TTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAG

表2F：帶有全長GPV ITR的例示性GPV-FVIII建構體（核苷酸1-7154；SEQ ID NO: 194）

描述	序列
5' ITR（SEQ ID NO: 187）	CTCATTGGAGGGTTCGTTCGTTCGAACCAGCCAATCAGGGGAGGGGGAAG TGACGCAAGTTCCGGTCACATGCTTCCGGTGACGCACATCCGGTGACGTAG TTCCGGTCACGTGCTTCCTGTCACGTGTTTCCGGTCACGTGACTTCCGGTCA TGTGACTTCCGGTGACGTGTTTCCGGCTGTTAGGTTGACCACGCGCATGCC GCGCGGTCAGCCCAATAGTTAAGCCGGAAACACGTCACCGGAAGTCACAT GACCGGAAGTCACGTGACCGGAAACACGTGACAGGAAGCACGTGACCGG AACTACGTCACCGGATGTGCGTCACCGGAAGCATGTGACCGGAACTTGCG TCACTTCCCCCTCCCCTGATTGGCTGGTTCGAACGAACGAACCCTCCAATG AGACTCAAGGACAAGAGGATATTTTGCGCGCCAGGAAGTG

TTPp（肝臟特異性啟動子）（SEQ ID NO: 113）	ACGCGTGTCTGTCTGCACATTTCGTAGAGCGAGTGTTCCGATACTCTAATC TCCCTAGGCAAGGTTCATATTTGTGTAGGTTACTTATTCTCCTTTTGTTGAC TAAGTCAATAATCAGAATCAGCAGGTTTGGAGTCAGCTTGGCAGGGATCA GCAGCCTGGGTTGGAAGGAGGGGGTATAAAAGCCCCTTCACCAGGAGAAG CCGTCACACAGATCCACAAGCTCCTG
合成內含子（SEQ ID NO: 115）	GTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGC CCTTGCGTGCCTTGAATTACTGACACTGACATCCACTTTTTCTTTTTCTCCA CAG
FVIIIco6XTEN（SEQ ID NO: 117）（含有XTEN144的密碼子優化的FVIII形式6的開放閱讀框；XTEN序列通過加雙下劃線（SEQ ID NO: 118）標記）	ATGCAGATTGAGCTGTCCACTTGTTTCTTCCTGTGCCTCCTGCGCTTCTGTT TCTCCGCCACTCGCCGGTACTACCTTGGAGCCGTGGAGCTTTCATGGGACT ACATGCAGAGCGACCTGGGCGAACTCCCCGTGGATGCCAGATTCCCCCCCC GCGTGCCAAAGTCCTTCCCCTTTAACACCTCCGTGGTGTACAAGAAAACCC TCTTTGTCGAGTTCACTGACCACCTGTTCAACATCGCCAAGCCGCGCCCAC CTTGGATGGGCCTCCTGGGACCGACCATTCAAGCTGAAGTGTACGACACCG TGGTGATCACCCTGAAGAACATGGCGTCCCACCCCGTGTCCCTGCATGCGG TCGGAGTGTCCTACTGGAAGGCCTCCGAAGGAGCTGAGTACGACGACCAG ACTAGCCAGCGGGAAAAGGAGGACGATAAAGTGTTCCCGGGCGGCTCGCA TACTTACGTGTGGCAAGTCCTGAAGGAAAACGGACCTATGGCATCCGATCC TCTGTGCCTGACTTACTCCTACCTTTCCCATGTGGACCTCGTGAAGGACCTG AACAGCGGGCTGATTGGTGCACTTCTCGTGTGCCGCGAAGGTTCGCTCGCT AAGGAAAAGACCCAGACCCTCCATAAGTTCATCCTTTTGTTCGCTGTGTTC GATGAAGGAAAGTCATGGCATTCCGAAACTAAGAACTCGCTGATGCAGGA CCGGGATGCCGCCTCAGCCCGCGCCTGGCCTAAAATGCATACAGTCAACG GATACGTGAATCGGTCACTGCCCGGGCTCATCGGTTGTCACAGAAAGTCCG TGTACTGGCACGTCATCGGCATGGGCACTACGCCTGAAGTGCACTCCATCT

	TCCTGGAAGGGCACACCTTCCTCGTGCGCAACCACCGCCAGGCCTCTCTGG
	AAATCTCCCCGATTACCTTTCTGACCGCCCAGACTCTGCTCATGGACCTGG
	GGCAGTTCCTTCTCTTCTGCCACATCTCCAGCCATCAGCACGACGGAATGG
	AGGCCTACGTGAAGGTGGACTCATGCCCGGAAGAACCTCAGTTGCGGATG
	AAGAACAACGAGGAGGCCGAGGACTATGACGACGATTTGACTGACTCCGA
	GATGGACGTCGTGCGGTTCGATGACGACAACAGCCCCAGCTTCATCCAGAT
	TCGCAGCGTGGCCAAGAAGCACCCCAAAACCTGGGTGCACTACATCGCGG
	CCGAGGAAGAAGATTGGGACTACGCCCCGTTGGTGCTGGCACCCGATGAC
	CGGTCGTACAAGTCCCAGTATCTGAACAATGGTCCGCAGCGGATTGGCAG
	AAAGTACAAGAAAGTGCGGTTCATGGCGTACACTGACGAAACGTTTAAGA
	CCCGGGAGGCCATTCAACATGAGAGCGGCATTCTGGGACCACTGCTGTAC
	GGAGAGGTCGGCGATACCCTGCTCATCATCTTCAAAAACCAGGCCTCCCGG
	CCTTACAACATCTACCCTCACGGAATCACCGACGTGCGGCCACTCTACTCG
	CGGCGCCTGCCGAAGGGCGTCAAGCACCTGAAAGACTTCCCTATCCTGCCG
	GGCGAAATCTTCAAGTATAAGTGGACCGTCACCGTGGAGGACGGGCCCAC
	CAAGAGCGATCCTAGGTGTCTGACTCGGTACTACTCCAGCTTCGTGAACAT
	GGAACGGGACCTGGCATCGGGACTCATTGGACCGCTGCTGATCTGCTACA
	AAGAGTCGGTGGATCAACGCGGCAACCAGATCATGTCCGACAAGCGCAAC
	GTGATCCTGTTCTCCGTGTTTGATGAAAACAGATCCTGGTACCTCACTGAA
	AACATCCAGAGGTTCCTCCCAAACCCCGCAGGAGTGCAACTGGAGGACCC
	TGAGTTTCAGGCCTCGAATATCATGCACTCGATTAACGGTTACGTGTTCGA
	CTCGCTGCAACTGAGCGTGTGCCTCCATGAAGTCGCTTACTGGTACATTCT
	GTCCATCGGCGCCCAGACTGACTTCCTGAGCGTGTTCTTTTCCGGTTACACC
	TTTAAGCACAAGATGGTGTACGAAGATACCCTGACCCTGTTCCCTTTCTCC
	GGCGAAACGGTGTTCATGTCGATGGAGAACCCGGGTCTGTGGATTCTGGG

	ATGCCACAACAGCGACTTTCGGAACCGCGGAATGACTGCCCTGCTGAAGG
	TGTCCTCATGCGACAAGAACACCGGAGACTACTACGAGGACTCCTACGAG
	GATATCTCAGCCTACCTCCTGTCCAAGAACAACGCGATCGAGCCGCGCAGC
	TTCAGCCAGAACGGCGCGCCAACATCAGAGAGCGCCACCCCTGAAAGTGG
	TCCCGGGAGCGAGCCAGCCACATCTGGGTCGGAAACGCCAGGCACAAGTG
	AGTCTGCAACTCCCGAGTCCGGACCTGGCTCCGAGCCTGCCACTAGCGGCT
	CCGAGACTCCGGGAACTTCCGAGAGCGCTACACCAGAAAGCGGACCCGGA
	ACCAGTACCGAACCTAGCGAGGGCTCTGCTCCGGGCAGCCCAGCCGGCTC
	TCCTACATCCACGGAGGAGGGCACTTCCGAATCCGCCACCCCGGAGTCAG
	GGCCAGGATCTGAACCCGCTACCTCAGGCAGTGAGACGCCAGGAACGAGC
	GAGTCCGCTACACCGGAGAGTGGGCCAGGGAGCCCTGCTGGATCTCCTAC
	GTCCACTGAGGAAGGGTCACCAGCGGGCTCGCCCACCAGCACTGAAGAAG
	GTGCCTCGAGCCCGCCTGTGCTGAAGAGGCACCAGCGAGAAATTACCCGG
	ACCACCCTCCAATCGGATCAGGAGGAAATCGACTACGACGACACCATCTC
	GGTGGAAATGAAGAAGGAAGATTTCGATATCTACGACGAGGACGAAAATC
	AGTCCCCTCGCTCATTCCAAAAGAAAACTAGACACTACTTTATCGCCGCGG
	TGGAAAGACTGTGGGACTATGGAATGTCATCCAGCCCTCACGTCCTTCGGA
	ACCGGGCCCAGAGCGGATCGGTGCCTCAGTTCAAGAAAGTGGTGTTCCAG
	GAGTTCACCGACGGCAGCTTCACCCAGCCGCTGTACCGGGGAGAACTGAA
	CGAACACCTGGGCCTGCTCGGTCCCTACATCCGCGCGGAAGTGGAGGATA
	ACATCATGGTGACCTTCCGTAACCAAGCATCCAGACCTTACTCCTTCTATTC
	CTCCCTGATCTCATACGAGGAGGACCAGCGCCAAGGCGCCGAGCCCCGCA
	AGAACTTCGTCAAGCCCAACGAGACTAAGACCTACTTCTGGAAGGTCCAA
	CACCATATGGCCCCGACCAAGGATGAGTTTGACTGCAAGGCCTGGGCCTA
	CTTCTCCGACGTGGACCTTGAGAAGGATGTCCATTCCGGCCTGATCGGGCC

	GCTGCTCGTGTGTCACACCAACACCCTGAACCCAGCGCATGGACGCCAGGT
	CACCGTCCAGGAGTTTGCTCTGTTCTTCACCATTTTTGACGAAACTAAGTCC
	TGGTACTTCACCGAGAATATGGAGCGAAACTGTAGAGCGCCCTGCAATAT
	CCAGATGGAAGATCCGACTTTCAAGGAGAACTATAGATTCCACGCCATCA
	ACGGGTACATCATGGATACTCTGCCGGGGCTGGTCATGGCCCAGGATCAG
	AGGATTCGGTGGTACTTGCTGTCAATGGGATCGAACGAAAACATTCACTCC
	ATTCACTTCTCCGGTCACGTGTTCACTGTGCGCAAGAAGGAGGAGTACAAG
	ATGGCGCTGTACAATCTGTACCCCGGGGTGTTCGAAACTGTGGAGATGCTG
	CCGTCCAAGGCCGGCATCTGGAGAGTGGAGTGCCTGATCGGAGAGCACCT
	CCACGCGGGGATGTCCACCCTCTTCCTGGTGTACTCGAATAAGTGCCAGAC
	CCCGCTGGGCATGGCCTCGGGCCACATCAGAGACTTCCAGATCACAGCAA
	GCGGACAATACGGCCAATGGGCGCCGAAGCTGGCCCGCTTGCACTACTCC
	GGATCGATCAACGCATGGTCCACCAAGGAACCGTTCTCGTGGATTAAGGT
	GGACCTCCTGGCCCCTATGATTATCCACGGAATTAAGACCCAGGGCGCCAG
	GCAGAAGTTCTCCTCCCTGTACATCTCGCAATTCATCATCATGTACAGCCT
	GGACGGGAAGAAGTGGCAGACTTACAGGGGAAACTCCACCGGCACCCTGA
	TGGTCTTTTTCGGCAACGTGGATTCCTCCGGCATTAAGCACAACATCTTCA
	ACCCACCGATCATAGCCAGATATATTAGGCTCCACCCCACTCACTACTCAA
	TCCGCTCAACTCTTCGGATGGAACTCATGGGGTGCGACCTGAACTCCTGCT
	CCATGCCGTTGGGGATGGAATCAAAGGCTATTAGCGACGCCCAGATCACC
	GCGAGCTCCTACTTCACTAACATGTTCGCCACCTGGAGCCCCTCCAAGGCC
	AGGCTGCACTTGCAGGGACGGTCAAATGCCTGGCGGCCGCAAGTGAACAA
	TCCGAAGGAATGGCTTCAAGTGGATTTCCAAAAGACCATGAAAGTGACCG
	GAGTCACCACCCAGGGAGTGAAGTCCCTTCTGACCTCGATGTATGTGAAGG
	AGTTCCTGATTAGCAGCAGCCAGGACGGGCACCAGTGGACCCTGTTCTTCC

	AAAACGGAAAGGTCAAGGTGTTCCAGGGGAACCAGGACTCGTTCACACCC GTGGTGAACTCCCTGGACCCCCCACTGCTGACGCGGTACTTGAGGATTCAT CCTCAGTCCTGGGTCCATCAGATTGCATTGCGAATGGAAGTCCTGGGCTGC GAGGCCCAGGACCTGTACTGA
WPRE(突變的土撥鼠肝炎病毒轉錄後調控元件)(SEQ ID NO: 120)	TCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCT TAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTG TATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATC CTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGG CGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGC CACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCC ACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGG CTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTT CCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCT GCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCT GCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCG GATCTCCCTTTGGGCCGCCTCCCCGCTG
bGHpA(牛生長激素多聚腺苷酸化信號)(SEQ ID NO: 122)	CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCC TTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGA GGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGG GGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCAT GCTGGGGA
3' ITR反向末端重複體(SEQ ID NO: 188)	CACTTCCTGGCGCGCAAAATATCCTCTTGTCCTTGAGTCTCATTGGAGGGTT CGTTCGTTCGAACCAGCCAATCAGGGGAGGGGGAAGTGACGCAAGTTCCG GTCACATGCTTCCGGTGACGCACATCCGGTGACGTAGTTCCGGTCACGTGC TTCCTGTCACGTGTTTCCGGTCACGTGACTTCCGGTCATGTGACTTCCGGTG

	ACGTGTTTCCGGCTTAACTATTGGGCTGACCGCGCGGCATGCGCGTGGTCA ACCTAACAGCCGGAAACACGTCACCGGAAGTCACATGACCGGAAGTCACG TGACCGGAAACACGTGACAGGAAGCACGTGACCGGAACTACGTCACCGGA TGTGCGTCACCGGAAGCATGTGACCGGAACTTGCGTCACTTCCCCCTCCCC TGATTGGCTGGTTCGAACGAACGAACCCTCCAATGAG

全長序列（SEQ ID NO: 194）

CTCATTGGAGGGTTCGTTCGTTCGAACCAGCCAATCAGGGGAGGGGGAAGTGACGCAAGTTCC

GGTCACATGCTTCCGGTGACGCACATCCGGTGACGTAGTTCCGGTCACGTGCTTCCTGTCACGT

GTTTCCGGTCACGTGACTTCCGGTCATGTGACTTCCGGTGACGTGTTTCCGGCTGTTAGGTTGAC

CACGCGCATGCCGCGCGGTCAGCCCAATAGTTAAGCCGGAAACACGTCACCGGAAGTCACATG

ACCGGAAGTCACGTGACCGGAAACACGTGACAGGAAGCACGTGACCGGAACTACGTCACCGG

ATGTGCGTCACCGGAAGCATGTGACCGGAACTTGCGTCACTTCCCCCTCCCCTGATTGGCTGGT

TCGAACGAACGAACCCTCCAATGAGACTCAAGGACAAGAGGATATTTTGCGCGCCAGGAAGTG

GCGGCAATTCAGTCGATAACTATAACGGTCCTAAGGTAGCGATTTAAATACGCGCTCTCTTAAG

GTAGCCCCGGGACGCGTCAATTGAGATCTGGATCCGGTACCGAATTCGCGGCCGCCTCGACGA

CTAGCGTTTAATTAAACGCGTGTCTGTCTGCACATTTCGTAGAGCGAGTGTTCCGATACTCTAA

TCTCCCTAGGCAAGGTTCATATTTGTGTAGGTTACTTATTCTCCTTTTGTTGACTAAGTCAATAA

TCAGAATCAGCAGGTTTGGAGTCAGCTTGGCAGGGATCAGCAGCCTGGGTTGGAAGGAGGGGG

TATAAAAGCCCCTTCACCAGGAGAAGCCGTCACACAGATCCACAAGCTCCTGAGGTAAGTGCC

GTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTTGCGTGCCTTGAATTACTG

ACACTGACATCCACTTTTTCTTTTTCTCCACAGCTAGCGCCACCATGCAGATTGAGCTGTCCACT

TGTTTCTTCCTGTGCCTCCTGCGCTTCTGTTTCTCCGCCACTCGCCGGTACTACCTTGGAGCCGT

GGAGCTTTCATGGGACTACATGCAGAGCGACCTGGGCGAACTCCCCGTGGATGCCAGATTCCC

CCCCCGCGTGCCAAAGTCCTTCCCCTTTAACACCTCCGTGGTGTACAAGAAAACCCTCTTTGTC

GAGTTCACTGACCACCTGTTCAACATCGCCAAGCCGCGCCCACCTTGGATGGGCCTCCTGGGAC

CGACCATTCAAGCTGAAGTGTACGACACCGTGGTGATCACCCTGAAGAACATGGCGTCCCACC

CCGTGTCCCTGCATGCGGTCGGAGTGTCCTACTGGAAGGCCTCCGAAGGAGCTGAGTACGACG

ACCAGACTAGCCAGCGGGAAAAGGAGGACGATAAAGTGTTCCCGGGCGGCTCGCATACTTACG

TGTGGCAAGTCCTGAAGGAAAACGGACCTATGGCATCCGATCCTCTGTGCCTGACTTACTCCTA

CCTTTCCCATGTGGACCTCGTGAAGGACCTGAACAGCGGGCTGATTGGTGCACTTCTCGTGTGC

CGCGAAGGTTCGCTCGCTAAGGAAAAGACCCAGACCCTCCATAAGTTCATCCTTTTGTTCGCTG

TGTTCGATGAAGGAAAGTCATGGCATTCCGAAACTAAGAACTCGCTGATGCAGGACCGGGATG

CCGCCTCAGCCCGCGCCTGGCCTAAAATGCATACAGTCAACGGATACGTGAATCGGTCACTGC

CCGGGCTCATCGGTTGTCACAGAAAGTCCGTGTACTGGCACGTCATCGGCATGGGCACTACGC

CTGAAGTGCACTCCATCTTCCTGGAAGGGCACACCTTCCTCGTGCGCAACCACCGCCAGGCCTC

TCTGGAAATCTCCCCGATTACCTTTCTGACCGCCCAGACTCTGCTCATGGACCTGGGGCAGTTC

CTTCTCTTCTGCCACATCTCCAGCCATCAGCACGACGGAATGGAGGCCTACGTGAAGGTGGACT

CATGCCCGGAAGAACCTCAGTTGCGGATGAAGAACAACGAGGAGGCCGAGGACTATGACGAC

GATTTGACTGACTCCGAGATGGACGTCGTGCGGTTCGATGACGACAACAGCCCCAGCTTCATCC

AGATTCGCAGCGTGGCCAAGAAGCACCCCAAAACCTGGGTGCACTACATCGCGGCCGAGGAA

GAAGATTGGGACTACGCCCCGTTGGTGCTGGCACCCGATGACCGGTCGTACAAGTCCCAGTAT

CTGAACAATGGTCCGCAGCGGATTGGCAGAAAGTACAAGAAAGTGCGGTTCATGGCGTACACT

GACGAAACGTTTAAGACCCGGGAGGCCATTCAACATGAGAGCGGCATTCTGGGACCACTGCTG

TACGGAGAGGTCGGCGATACCCTGCTCATCATCTTCAAAAACCAGGCCTCCCGGCCTTACAAC

ATCTACCCTCACGGAATCACCGACGTGCGGCCACTCTACTCGCGGCGCCTGCCGAAGGGCGTC

AAGCACCTGAAAGACTTCCCTATCCTGCCGGGCGAAATCTTCAAGTATAAGTGGACCGTCACC

GTGGAGGACGGGCCCACCAAGAGCGATCCTAGGTGTCTGACTCGGTACTACTCCAGCTTCGTG

AACATGGAACGGGACCTGGCATCGGGACTCATTGGACCGCTGCTGATCTGCTACAAAGAGTCG

GTGGATCAACGCGGCAACCAGATCATGTCCGACAAGCGCAACGTGATCCTGTTCTCCGTGTTTG

ATGAAAACAGATCCTGGTACCTCACTGAAAACATCCAGAGGTTCCTCCCAAACCCCGCAGGAG

TGCAACTGGAGGACCCTGAGTTTCAGGCCTCGAATATCATGCACTCGATTAACGGTTACGTGTT

CGACTCGCTGCAACTGAGCGTGTGCCTCCATGAAGTCGCTTACTGGTACATTCTGTCCATCGGC

GCCCAGACTGACTTCCTGAGCGTGTTCTTTTCCGGTTACACCTTTAAGCACAAGATGGTGTACG

AAGATACCCTGACCCTGTTCCCTTTCTCCGGCGAAACGGTGTTCATGTCGATGGAGAACCCGGG

TCTGTGGATTCTGGGATGCCACAACAGCGACTTTCGGAACCGCGGAATGACTGCCCTGCTGAA

GGTGTCCTCATGCGACAAGAACACCGGAGACTACTACGAGGACTCCTACGAGGATATCTCAGC

CTACCTCCTGTCCAAGAACAACGCGATCGAGCCGCGCAGCTTCAGCCAGAACGGCGCGCCAAC

ATCAGAGAGCGCCACCCCTGAAAGTGGTCCCGGGAGCGAGCCAGCCACATCTGGGTCGGAAAC

GCCAGGCACAAGTGAGTCTGCAACTCCCGAGTCCGGACCTGGCTCCGAGCCTGCCACTAGCGG

CTCCGAGACTCCGGGAACTTCCGAGAGCGCTACACCAGAAAGCGGACCCGGAACCAGTACCGA

ACCTAGCGAGGGCTCTGCTCCGGGCAGCCCAGCCGGCTCTCCTACATCCACGGAGGAGGGCAC

TTCCGAATCCGCCACCCCGGAGTCAGGGCCAGGATCTGAACCCGCTACCTCAGGCAGTGAGAC

GCCAGGAACGAGCGAGTCCGCTACACCGGAGAGTGGGCCAGGGAGCCCTGCTGGATCTCCTAC

GTCCACTGAGGAAGGGTCACCAGCGGGCTCGCCCACCAGCACTGAAGAAGGTGCCTCGAGCCC

GCCTGTGCTGAAGAGGCACCAGCGAGAAATTACCCGGACCACCCTCCAATCGGATCAGGAGGA

AATCGACTACGACGACACCATCTCGGTGGAAATGAAGAAGGAAGATTTCGATATCTACGACGA

GGACGAAAATCAGTCCCCTCGCTCATTCCAAAAGAAAACTAGACACTACTTTATCGCCGCGGT

GGAAAGACTGTGGGACTATGGAATGTCATCCAGCCCTCACGTCCTTCGGAACCGGGCCCAGAG

CGGATCGGTGCCTCAGTTCAAGAAAGTGGTGTTCCAGGAGTTCACCGACGGCAGCTTCACCCA

GCCGCTGTACCGGGGAGAACTGAACGAACACCTGGGCCTGCTCGGTCCCTACATCCGCGCGGA

AGTGGAGGATAACATCATGGTGACCTTCCGTAACCAAGCATCCAGACCTTACTCCTTCTATTCC

TCCCTGATCTCATACGAGGAGGACCAGCGCCAAGGCGCCGAGCCCCGCAAGAACTTCGTCAAG

CCCAACGAGACTAAGACCTACTTCTGGAAGGTCCAACACCATATGGCCCCGACCAAGGATGAG

TTTGACTGCAAGGCCTGGGCCTACTTCTCCGACGTGGACCTTGAGAAGGATGTCCATTCCGGCC

TGATCGGGCCGCTGCTCGTGTGTCACACCAACACCCTGAACCCAGCGCATGGACGCCAGGTCA

CCGTCCAGGAGTTTGCTCTGTTCTTCACCATTTTTGACGAAACTAAGTCCTGGTACTTCACCGAG

AATATGGAGCGAAACTGTAGAGCGCCCTGCAATATCCAGATGGAAGATCCGACTTTCAAGGAG

AACTATAGATTCCACGCCATCAACGGGTACATCATGGATACTCTGCCGGGGCTGGTCATGGCCC

AGGATCAGAGGATTCGGTGGTACTTGCTGTCAATGGGATCGAACGAAAACATTCACTCCATTC

ACTTCTCCGGTCACGTGTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCGCTGTACAATC

TGTACCCCGGGGTGTTCGAAACTGTGGAGATGCTGCCGTCCAAGGCCGGCATCTGGAGAGTGG

AGTGCCTGATCGGAGAGCACCTCCACGCGGGGATGTCCACCCTCTTCCTGGTGTACTCGAATAA

GTGCCAGACCCCGCTGGGCATGGCCTCGGGCCACATCAGAGACTTCCAGATCACAGCAAGCGG

ACAATACGGCCAATGGGCGCCGAAGCTGGCCCGCTTGCACTACTCCGGATCGATCAACGCATG

GTCCACCAAGGAACCGTTCTCGTGGATTAAGGTGGACCTCCTGGCCCCTATGATTATCCACGGA

ATTAAGACCCAGGGCGCCAGGCAGAAGTTCTCCTCCCTGTACATCTCGCAATTCATCATCATGT

ACAGCCTGGACGGGAAGAAGTGGCAGACTTACAGGGGAAACTCCACCGGCACCCTGATGGTCT

TTTTCGGCAACGTGGATTCCTCCGGCATTAAGCACAACATCTTCAACCCACCGATCATAGCCAG

ATATATTAGGCTCCACCCCACTCACTACTCAATCCGCTCAACTCTTCGGATGGAACTCATGGGG

TGCGACCTGAACTCCTGCTCCATGCCGTTGGGGATGGAATCAAAGGCTATTAGCGACGCCCAG

ATCACCGCGAGCTCCTACTTCACTAACATGTTCGCCACCTGGAGCCCCTCCAAGGCCAGGCTGC

ACTTGCAGGGACGGTCAAATGCCTGGCGGCCGCAAGTGAACAATCCGAAGGAATGGCTTCAAG

TGGATTTCCAAAAGACCATGAAAGTGACCGGAGTCACCACCCAGGGAGTGAAGTCCCTTCTGA

CCTCGATGTATGTGAAGGAGTTCCTGATTAGCAGCAGCCAGGACGGGCACCAGTGGACCCTGT

TCTTCCAAAACGGAAAGGTCAAGGTGTTCCAGGGGAACCAGGACTCGTTCACACCCGTGGTGA

ACTCCCTGGACCCCCCACTGCTGACGCGGTACTTGAGGATTCATCCTCAGTCCTGGGTCCATCA

GATTGCATTGCGAATGGAAGTCCTGGGCTGCGAGGCCCAGGACCTGTACTGAATCAGCCTGAG

CTCGCTGATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACT

ATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCC

CGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTG

GCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGG

GGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGC

GGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAAT

TCCGTGGTGTTGTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGAT

TCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCG

GCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCC

CTTTGGGCCGCCTCCCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTT

GCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAAT

GAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAG

GACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTAT

GGCTTCTGAGGCGGAAAGAACGGGCTCGAGAAGCTTCTAGATATCCTCTCTTAAGGTAGCATC

GAGATTTAAATTAGGGATAACAGGGTAATGGCGCGGGCCGCCACTTCCTGGCGCGCAAAATAT

CCTCTTGTCCTTGAGTCTCATTGGAGGGTTCGTTCGTTCGAACCAGCCAATCAGGGGAGGGGGA

AGTGACGCAAGTTCCGGTCACATGCTTCCGGTGACGCACATCCGGTGACGTAGTTCCGGTCACG

TGCTTCCTGTCACGTGTTTCCGGTCACGTGACTTCCGGTCATGTGACTTCCGGTGACGTGTTTCC

GGCTTAACTATTGGGCTGACCGCGCGGCATGCGCGTGGTCAACCTAACAGCCGGAAACACGTC

ACCGGAAGTCACATGACCGGAAGTCACGTGACCGGAAACACGTGACAGGAAGCACGTGACCG

GAACTACGTCACCGGATGTGCGTCACCGGAAGCATGTGACCGGAACTTGCGTCACTTCCCCCTC

CCCTGATTGGCTGGTTCGAACGAACGAACCCTCCAATGAG

【0283】 在一個實施例中，基因盒包含苯丙胺酸羥化酶（PAH）建構體，其包括如表 10A 和 10B 中所列示的多核苷酸序列。在一個實施例中，基因盒包含 PAH 建構體，其包括表 10A 中所述的多核苷酸序列。在一個實施例中，基因盒包含 PAH 建構體，其包括表 10B 中所述的多核苷酸序列。

【0284】 在某些實施例中，分離的核酸分子包含與 SEQ ID NO: 197 或 198 的核苷酸序列具有至少約 60%、至少約 65%、至少約 70%、至少約

75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或約 100%序列同一性的核苷酸序列。在一些實施例中，分離的核酸分子包含與 SEQ ID NO: 197 的核苷酸序列具有至少約 60%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或約 100%序列同一性的核苷酸序列。在一些實施例中，分離的核酸分子包含與 SEQ ID NO: 198 的核苷酸序列具有至少約 60%、至少約 65%、至少約 70%、至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或約 100%序列同一性的核苷酸序列。在一些實施例中，分離的核酸分子保留表現功能性苯丙胺酸羥化酶的能力。

A. 密碼子適應指數

【0285】 在一個實施例中，基因盒包含編碼 FVIII 多肽的密碼子優化的核苷酸序列，其中密碼子優化的核苷酸序列的人類密碼子適應指數相對於 SEQ ID NO: 16 增加。例如，密碼子優化的核苷酸序列可以具有以下人類密碼子適應指數：至少約 0.75（75%）、至少約 0.76（76%）、至少約 0.77（77%）、至少約 0.78（78%）、至少約 0.79（79%）、至少約 0.80（80%）、至少約 0.81（81%）、至少約 0.82（82%）、至少約 0.83（83%）、至少約 0.84（84%）、至少約 0.85（85%）、至少約 0.86（86%）、至少約 0.87（87%）、至少約 0.88（88%）、至少約 0.89（89%）、至少約 0.90（90%）、至少約 0.91（91%）、至少約 0.92（92%）、至少約 0.93（93%）、至少約 0.94（94%）、至少約 0.95（95%）、至少約 0.96（96%）、至少約 0.97（97%）、至少約 0.98（98%）或至少約 0.99（99%）。在一些實施例中，密碼子優化的核苷酸序列具有至少約.88（88%）的人類密碼子適應指數。在其他實施例中，密碼子優化的核苷酸序列具有至少約.91（91%）的人類密碼子適應指數。在其他實施例中，密碼子優化的核苷酸序列具有至少約.91（97%）的人類密碼子適應指數。

【0286】 在一個特定實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含如下核苷酸序列，其包含編碼 FVIII 多肽的 N 末端部分的第一

核酸序列和編碼 FVIII 多肽的 C 末端部分的第二核酸序列；其中第一核酸序列與 (i) SEQ ID NO: 3 的核苷酸 58-1791；(ii) SEQ ID NO: 3 的核苷酸 1-1791；(iii) SEQ ID NO: 4 的核苷酸 58-1791；或 (iv) SEQ ID NO: 4 的核苷酸 1-1791 具有至少約 80%、至少約 85%、至少約 86%、至少約 87%、至少約 88%、至少約 89%、至少約 90%、至少約 91%、至少約 92%、至少約 93%、至少約 94%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性；其中 N 末端部分和 C 末端部分一起具有 FVIII 多肽活性；並且其中所述核苷酸序列的人類密碼子適應指數相對於 SEQ ID NO: 16 增加。在一些實施例中，核苷酸序列具有以下人類密碼子適應指數：至少約 0.75（75%）、至少約 0.76（76%）、至少約 0.77（77%）、至少約 0.78（78%）、至少約 0.79（79%）、至少約 0.80（80%）、至少約 0.81（81%）、至少約 0.82（82%）、至少約 0.83（83%）、至少約 0.84（84%）、至少約 0.85（85%）、至少約 0.86（86%）、至少約 0.87（87%）、至少約 0.88（88%）、至少約 0.89（89%）、至少約 0.90（90%）或至少約.91（91%）。在一個特定實施例中，核苷酸序列具有至少約.88（88%）的人類密碼子適應指數。在另一個實施例中，核苷酸序列具有至少約.91（91%）的人類密碼子適應指數。

【0287】 在另一個實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含如下核苷酸序列，其包含編碼 FVIII 多肽的 N 末端部分的第一核酸序列和編碼 FVIII 多肽的 C 末端部分的第二核酸序列；其中第二核酸序列與 (i) SEQ ID NO: 5 的核苷酸 1792-2277 和 2320-4374 或 (ii) SEQ ID NO: 6 的核苷酸 1792-2277 和 2320-4374 具有至少約 80%、至少約 85%、至少約 86%、至少約 87%、至少約 88%、至少約 89%、至少約 90%、至少約 91%、至少約 92%、至少約 93%、至少約 94%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性；其中 N 末端部分和 C 末端部分一起具有 FVIII 多肽活性；並且其中所述核苷酸序列的人類密碼子適應指數相對於 SEQ ID NO: 16 增加。在一些實施例中，核苷酸序列具有以下人類密碼子適應指數：至少約 0.75（75%）、至少約 0.76（76%）、至少約 0.77（77%）、至少約 0.78（78%）、至少約 0.79（79%）、至少約 0.80（80%）、

至少約 0.81（81%）、至少約 0.82（82%）、至少約 0.83（83%）、至少約 0.84（84%）、至少約 0.85（85%）、至少約 0.86（86%）、至少約 0.87（87%）或至少約 0.88（88%）。在一個特定實施例中，核苷酸序列具有至少約.83（83%）的人類密碼子適應指數。在另一個實施例中，核苷酸序列具有至少約.88（88%）的人類密碼子適應指數。

【0288】 在其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含編碼具有 FVIII 活性的多肽的核苷酸序列，其中所述核苷酸序列包含與選自 SEQ ID NO: 1、2、3、4、5、6、70 和 71 的胺基酸序列的核苷酸 58-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 1、2、3、4、5、6、70 或 71 的核苷酸 58-4374）具有至少約 80%、至少約 85%、至少約 89%、至少約 90%、至少約 91%、至少約 92%、至少約 93%、至少約 94%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性的核酸序列；並且其中所述核苷酸序列的人類密碼子適應指數相對於 SEQ ID NO: 16 增加。在一些實施例中，核苷酸序列具有以下人類密碼子適應指數：至少約 0.75（75%）、至少約 0.76（76%）、至少約 0.77（77%）、至少約 0.78（78%）、至少約 0.79（79%）、至少約 0.80（80%）、至少約 0.81（81%）、至少約 0.82（82%）、至少約 0.83（83%）、至少約 0.84（84%）、至少約 0.85（85%）、至少約 0.86（86%）、至少約 0.87（87%）或至少約 0.88（88%）。在一個特定實施例中，核苷酸序列具有至少約 0.75（75%）的人類密碼子適應指數。在另一個實施例中，核苷酸序列具有至少約 0.83（83%）的人類密碼子適應指數。在另一個實施例中，核苷酸序列具有至少約 0.88（88%）的人類密碼子適應指數。在另一個實施例中，核苷酸序列具有至少約 0.91（91%）的人類密碼子適應指數。在另一個實施例中，核苷酸序列具有至少約 0.97（97%）的人類密碼子適應指數。

【0289】 在一些實施例中，本公開文本的編碼 FVIII 多肽的密碼子優化的核苷酸序列具有相對於 SEQ ID NO: 16 增加的最優密碼子頻率（FOP）。在某些實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列的 FOP 是至少約 40、至少約 45、至少約 50、至少約 55、至少約 60、至少約 64、至少

約 65、至少約 70、至少約 75、至少約 79、至少約 80、至少約 85 或至少約 90。

【0290】 在其他實施例中，本公開文本的編碼 FVIII 多肽的密碼子優化的核苷酸序列具有相對於 SEQ ID NO: 16 增加的同義密碼子相對使用度（RCSU）。在一些實施例中，分離的核酸分子的 RCSU 大於 1.5。在其他實施例中，分離的核酸分子的 RCSU 大於 2.0。在某些實施例中，分離的核酸分子的 RCSU 是至少約 1.5、至少約 1.6、至少約 1.7、至少約 1.8、至少約 1.9、至少約 2.0、至少約 2.1、至少約 2.2、至少約 2.3、至少約 2.4、至少約 2.5、至少約 2.6 或至少約 2.7。

【0291】 在仍其他實施例中，本公開文本的編碼 FVIII 多肽的密碼子優化的核苷酸序列具有相對於 SEQ ID NO: 16 減少的密碼子有效數。在一些實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列具有以下密碼子有效數：小於約 50、小於約 45、小於約 40、小於約 35、小於約 30 或小於約 25。在一個特定實施例中，分離的核酸分子具有以下密碼子有效數：約 40、約 35、約 30、約 25 或約 20。

B. G/C含量優化

【0292】 在一些實施例中，基因盒包含編碼 FVIII 多肽的密碼子優化的核苷酸序列，其中與 SEQ ID NO: 16 中的 G/C 核苷酸的百分比相比，密碼子優化的核苷酸序列含有更高百分比的 G/C 核苷酸。在其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列具有以下 G/C 含量：至少約 45%、至少約 46%、至少約 47%、至少約 48%、至少約 49%、至少約 50%、至少約 51%、至少約 52%、至少約 53%、至少約 54%、至少約 55%、至少約 56%、至少約 57%、至少約 58%、至少約 59%或至少約 60%。

【0293】 在一個特定實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含編碼 FVIII 多肽的 N 末端部分的第一核酸序列和編碼 FVIII 多肽的 C 末端部分的第二核酸序列；其中第一核酸序列與 (i) SEQ ID NO: 3 的核苷酸 58-1791；(ii) SEQ ID NO: 3 的核苷酸 1-1791；(iii) SEQ ID NO: 4 的核苷酸 58-1791；或 (iv) SEQ ID NO: 4 的核苷酸 1-1791 具有至少約 80%、至少約 85%、至少約 86%、至少約 87%、至少約 88%、至少約 89%、至少約

90%、至少約 91%、至少約 92%、至少約 93%、至少約 94%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性；其中 N 末端部分和 C 末端部分一起具有 FVIII 多肽活性；並且其中與 SEQ ID NO: 16 中的 G/C 核苷酸的百分比相比，所述核苷酸序列含有更高百分比的 G/C 核苷酸。在一些實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含以下 G/C 含量：至少約 45%、至少約 46%、至少約 47%、至少約 48%、至少約 49%、至少約 50%、至少約 51%、至少約 52%、至少約 53%、至少約 54%、至少約 55%、至少約 56%、至少約 57%或至少約 58%。在一個特定實施例中，編碼具有 FVIII 活性的多肽的核苷酸序列具有至少約 58%的 G/C 含量。

【0294】 在另一個實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含編碼 FVIII 多肽的 N 末端部分的第一核酸序列和編碼 FVIII 多肽的 C 末端部分的第二核酸序列；其中第二核酸序列與 (i) SEQ ID NO: 5 的核苷酸 1792-4374；(ii) SEQ ID NO: 6 的核苷酸 1792-4374；(iii) SEQ ID NO: 5 的核苷酸 1792-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 5 的核苷酸 1792-4374）或 (iv) SEQ ID NO: 6 的核苷酸 1792-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 6 的核苷酸 1792-4374）具有至少約 80%、至少約 85%、至少約 86%、至少約 87%、至少約 88%、至少約 89%、至少約 90%、至少約 91%、至少約 92%、至少約 93%、至少約 94%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性；其中 N 末端部分和 C 末端部分一起具有 FVIII 多肽活性；並且其中與 SEQ ID NO: 16 中的 G/C 核苷酸的百分比相比，密碼子優化的核苷酸序列含有更高百分比的 G/C 核苷酸。在其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列具有以下 G/C 含量：至少約 45%、至少約 46%、至少約 47%、至少約 48%、至少約 49%、至少約 50%、至少約 51%、至少約 52%、至少約 53%、至少約 54%、至少約 55%、至少約 56%或至少約 57%。在一個特定實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列具有至少約 52%的 G/C 含量。在另一個實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列具有至少約 55%的 G/C

含量。在另一個實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列具有至少約 57%的 G/C 含量。

【0295】 在其他實施例中，基因盒包含編碼 FVIII 多肽的密碼子優化的核苷酸序列，其中密碼子優化的核苷酸序列包含與 (i) 核苷酸 58-4374 或 (ii) 選自 SEQ ID NO: 1、2、3、4、5、6、70 和 71 的胺基酸序列的核苷酸 58-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 1、2、3、4、5、6、70 或 71 的核苷酸 58-4374）具有至少約 80%、至少約 85%、至少約 89%、至少約 90%、至少約 91%、至少約 92%、至少約 93%、至少約 94%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性的核酸序列；並且其中與 SEQ ID NO: 16 中的 G/C 核苷酸的百分比相比，所述核苷酸序列含有更高百分比的 G/C 核苷酸。在其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列具有至少約 45%的 G/C 含量。在一個特定實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列具有至少約 52%的 G/C 含量。在另一個實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列具有至少約 55%的 G/C 含量。在另一個實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列具有至少約 57%的 G/C 含量。在另一個實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列具有至少約 58%的 G/C 含量。在仍另一個實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列具有至少約 60%的 G/C 含量。

【0296】 「G/C 含量」（或鳥嘌呤-胞嘧啶含量）或「G/C 核苷酸的百分比」是指 DNA 分子中是鳥嘌呤或胞嘧啶的含氮鹼基的百分比。G/C 含量可以使用下式來計算：

$$\frac{G+C}{A+T+G+C} \times 100 \quad \text{(III)}$$

【0297】 人基因的 G/C 含量具有高異質性，其中一些基因的 G/C 含量低至 20%，並且其他基因的 G/C 含量高至 95%。通常，富含 G/C 的基因具有更高表現。事實上，已經證實，增加基因的 G/C 含量可以導致基因表現增加，這主要是由於轉錄增加和較高穩態 mRNA 水準所致。參見 Kudla 等人, PLoS Biol., 4(6): e180 (2006)。

C. 類基質附著區(matrix attachment region-like)序列

【0298】 在一些實施例中，基因盒包含編碼 FVIII 多肽的密碼子優化的核苷酸序列，其中密碼子優化的核苷酸序列相對於 SEQ ID NO: 16 含有更少 MARS/ARS 序列。在其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列含有至多6個、至多5個、至多4個、至多3個或至多2個MARS/ARS序列。在其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列含有至多 1 個 MARS/ARS 序列。在又其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列不含 MARS/ARS 序列。

【0299】 在一個特定實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含編碼 FVIII 多肽的 N 末端部分的第一核酸序列和編碼 FVIII 多肽的 C 末端部分的第二核酸序列；其中第一核酸序列與 (i) SEQ ID NO: 3 的核苷酸 58-1791；(ii) SEQ ID NO: 3 的核苷酸 1-1791；(iii) SEQ ID NO: 4 的核苷酸 58-1791；或 (iv) SEQ ID NO: 4 的核苷酸 1-1791 具有至少約 80%、至少約 85%、至少約 86%、至少約 87%、至少約 88%、至少約 89%、至少約 90%、至少約 91%、至少約 92%、至少約 93%、至少約 94%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性；其中 N 末端部分和 C 末端部分一起具有 FVIII 多肽活性；並且其中密碼子優化的核苷酸序列相對於 SEQ ID NO: 16 含有更少 MARS/ARS 序列。在其他實施例中，編碼具有 FVIII 活性的多肽的核苷酸序列含有至多 6 個、至多 5 個、至多 4 個、至多 3 個或至多 2 個 MARS/ARS 序列。在其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列含有至多 1 個 MARS/ARS 序列。在又其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列不含 MARS/ARS 序列。

【0300】 在另一個實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含編碼 FVIII 多肽的 N 末端部分的第一核酸序列和編碼 FVIII 多肽的 C 末端部分的第二核酸序列；其中第二核酸序列與 (i) SEQ ID NO: 5 的核苷酸 1792-4374；(ii) SEQ ID NO: 6 的核苷酸 1792-4374；(iii) SEQ ID NO: 5 的核苷酸 1792-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 5 的核苷酸 1792-4374）；或 (iv) SEQ ID NO: 6 的核

苷酸 1792-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 6 的核苷酸 1792-4374）具有至少約 80%、至少約 85%、至少約 86%、至少約 87%、至少約 88%、至少約 89%、至少約 90%、至少約 91%、至少約 92%、至少約 93%、至少約 94%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性；其中 N 末端部分和 C 末端部分一起具有 FVIII 多肽活性；並且其中所述核苷酸序列相對於 SEQ ID NO: 16 含有更少 MARS/ARS 序列。在其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列含有至多 6 個、至多 5 個、至多 4 個、至多 3 個或至多 2 個 MARS/ARS 序列。在其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列含有至多 1 個 MARS/ARS 序列。在又其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列不含 MARS/ARS 序列。

【0301】 在其他實施例中，基因盒包含編碼 FVIII 多肽的密碼子優化的核苷酸序列，其中密碼子優化的核苷酸序列包含與 (i) SEQ ID NO: 1、2、3、4、5、6、70 或 71 的核苷酸 58-4374 或 (ii) SEQ ID NO: 1、2、3、4、5、6、70 或 71 的核苷酸 58-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 1、2、3、4、5、6、70 或 71 的核苷酸 58-4374）具有至少約 80%、至少約 85%、至少約 89%、至少約 90%、至少約 91%、至少約 92%、至少約 93%、至少約 94%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性的核酸序列；並且其中密碼子優化的核苷酸序列相對於 SEQ ID NO: 16 含有更少 MARS/ARS 序列。在其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列含有至多 6 個、至多 5 個、至多 4 個、至多 3 個或至多 2 個 MARS/ARS 序列。在其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列含有至多 1 個 MARS/ARS 序列。在又其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列不含 MARS/ARS 序列。

【0302】 已經鑑定出人類 FVIII 核苷酸序列中富含 AT 的元件，其與釀酒酵母（*Saccharomyces cerevisiae*）自主複製序列（ARS）和核基質附著區（MAR）共用序列相似性。（Fallux 等人, *Mol. Cell. Biol.* 16:4264-4272 (1996)）。已經證實這些元件中的一種在體外結合核因子並阻遏氯黴素乙醯

轉移酶（CAT）報告子基因的表現。*同上*。已經假設，這些序列可能有助於人類 FVIII 基因的轉錄阻遏。因此，在一個實施例中，在本公開文本的編碼 FVIII 多肽的密碼子優化的核苷酸序列中消除所有 MAR/ARS 序列。在親代 FVIII 序列（SEQ ID NO: 16）中存在 4 個 MAR/ARS ATATTT 序列（SEQ ID NO: 21）和 3 個 MAR/ARS AAATAT 序列（SEQ ID NO: 22）。所有這些位點都發生突變以破壞優化的 FVIII 序列（SEQ ID NO: 1-6）中的 MAR/ARS 序列。優化的序列中這些元件各自的位置以及相應核苷酸的序列顯示於下表 3 中。

表3：阻遏元件的變化概述

元件位置	起始 BDD FVIII序列（SEQ ID NO: 16）	優化的BDD FVIII序列							
		SEQ ID NO: 1	SEQ ID NO: 2	SEQ ID NO: 3	SEQ ID NO: 4	SEQ ID NO: 5	SEQ ID NO: 6	SEQ ID NO: 70	SEQ ID NO: 71
去穩定序列									
639	ATTTA	GTTTA	GTTCA	GTTCA	GTTCA	GTTCA	GTTCA	GTTCA	GTTCA
1338	ATTTA	GTTTA	GTTCA	CTTCA	GTTCA	GTTCA	GTTCA	CTTCA	GTTCA
1449	ATTTA	CTTTA	CTTCA	CTTCA	CTTCA	CTTCA	CTTCA	CTTCA	CTTCA
1590	TAAAT	TAAAT	CAAGT	CAAGT	TAAGT	CAAGT	CAAGT	CAAGT	TAAGT
1623	TAAAT	CAAAA	GAAGA	CTAAG	CAAGA	CAAGA	CAAGA	TAAGT	CAAGA
2410	ATTTA	ATCTA	ATCTA	ATCTA	ATCTA	ATCTA	ATCTA	ATCTA	ATCTA
2586	ATTTA	GTTTA	GTTCA	GTTCA	GTTCA	GTTCA	GTTCA	GTTCA	GTTCA
2630	TAAAT	TGAAT	TGAAC	TGAAC	TGAAC	TCAAT	TGAAC	TCAAT	TGAAC
3884	ATTTA	ATCTG	ACCTG	ACCTG	ACCTG	ATCTG	ACCTG	ATCTG	ACCTG
3887	TAAAT	TGAAC	TGAAC	TGAAC	TGAAC	TGAAC	TGAAC	TGAAC	TGAAC
潛在啟動子結合位點									

641	TTATA	TTATC	TCATC	TCATT	TCATC	TCATC	TCATC	TCATT	TCATC
1275	TATAA	CTATA	TTACA	CTACA	GTACA	CTACA	CTACA	CTACA	GTACA
1276	TTATA	TATAA	TACAA	TACAA	TACAA	TACAA	TACAA	TACAA	TACAA
1445	TTATA	TCATC	TCATC	TTATC	TCATC	TCATC	TCATC	TTATC	TCATC
1474	TATAA	TATAA	TACAA	TACAA	TACAA	TACAA	TACAA	TACAA	TACAA
1588	TATAA	TATAA	TACAA	TACAA	TATAA	TACAA	TACAA	TACAA	TATAA
2614	TTATA	CTGTA	CTGTA	CTGTA	CTGTA	TTGTA	CTGTA	TTGTA	CTGTA
2661	TATAA	CATCA	CATCA	CATCA	CATCA	CATCA	CATCC	CATCA	CATCC
3286	TATAA	TATAA	TACAA	TACAA	TACAA	TACAA	TACAA	TACAA	TACAA
3840	TTATA	TTATA	TTACT	CTACA	CTACA	CTACA	CTACT	CTACA	CTACT
類基質附著序列（MARS/ARS）									
1287	ATATTT	GTATCT	GTACCT	GTACCT	GTATCT	GTACCT	GTACCT	GTACCT	GTATCT
1447	ATATTT	ATCTTT	ATCTTC	ATCTTC	ATCTTC	ATCTTC	ATCTTC	ATCTTC	ATCTTC
1577	AAATAT	AAATCT	AGATCT	AAATCT	AAATCT	AGATCT	AGATCT	AAATCT	AAATCT
1585	AAATAT	AAGTAT	AAGTAC	AAGTAC	AAGTAT	AAGTAC	AAGTAC	AAGTAC	AAGTAT
2231	ATATTT	ACATCA	ATATCA	ACATCA	ACATCA	ACATCT	ATATCT	ACATCT	ATATCT
3054	AAATAT	AAACAT	GAATAT	GAACAT	GAACAT	GAACAT	GAATAT	GAACAT	GAATAT
3788	ATATTT	ATATCT	ATATCT	ACATCT	ACATCT	ACATCT	ACATCT	ACATCT	ACATCT
富含AU的序列元件（ARE）									
2468	ATTTTATT	ACTTCATC	ACTTCATC	ACTTCATT	ACTTCATT	ACTTTATT	ACTTTATC	ACTTTATT	ACTTTATC
3790	ATTTTTAA	ATCTTTAA	ATCTTCAA	ATCTTCAA	ATCTTCAA	ATCTTCAA	ATCTTCAA	ATCTTCAA	ATCTTCAA

多聚A/多聚T序列									
3273	AAAAAAA	GAAAAAA	GAAGAAG	GAAGAAG	GAAGAAG	GAAGAAG	CAAGAAG	GAAGAAG	CAAGAAG
4195	TTTTTT	TTCTTT	TTCTTC	TTCTTC	TTCTTC	TTCTTC	TTCTTC	TTCTTCC	TTCTTCC
剪接位點									
2203	GGTGAT	GGGGAC	GGCGAC	GGGGAC	GGGGAC	GGAGAC	GGAGAC	GGAGAC	GGAGAC

D. 去穩定序列

【0303】 在一些實施例中，基因盒包含編碼 FVIII 多肽的密碼子優化的核苷酸序列，其中密碼子優化的核苷酸序列相對於 SEQ ID NO: 16 含有更少去穩定元件。在其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列含有至多 9 個、至多 8 個、至多 7 個、至多 6 個或至多 5 個去穩定元件。在其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列含有至多 4 個、至多 3 個、至多 2 個或至多 1 個去穩定元件。在又其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列不含去穩定元件。

【0304】 在一個特定實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含編碼 FVIII 多肽的 N 末端部分的第一核酸序列和編碼 FVIII 多肽的 C 末端部分的第二核酸序列；其中第一核酸序列與 (i) SEQ ID NO: 3 的核苷酸 58-1791；(ii) SEQ ID NO: 3 的核苷酸 1-1791；(iii) SEQ ID NO: 4 的核苷酸 58-1791；或 (iv) SEQ ID NO: 4 的核苷酸 1-1791 具有至少約 80%、至少約 85%、至少約 86%、至少約 87%、至少約 88%、至少約 89%、至少約 90%、至少約 91%、至少約 92%、至少約 93%、至少約 94%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性；其中 N 末端部分和 C 末端部分一起具有 FVIII 多肽活性；並且其中密碼子優化的核苷酸序列相對於 SEQ ID NO: 16 含有更少去穩定元件。在其他實施例中，編碼具有 FVIII 活性的多肽的核苷酸序列含有至多 9 個、至多 8 個、至多 7 個、

至多 6 個或至多 5 個去穩定元件。在其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列含有至多 4 個、至多 3 個、至多 2 個或至多 1 個去穩定元件。在又其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列不含去穩定元件。

【0305】 在另一個實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含編碼 FVIII 多肽的 N 末端部分的第一核酸序列和編碼 FVIII 多肽的 C 末端部分的第二核酸序列；其中第二核酸序列與 (i) SEQ ID NO: 5 的核苷酸 1792-4374；(ii) SEQ ID NO: 6 的核苷酸 1792-4374；(iii) SEQ ID NO: 5 的核苷酸 1792-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 5 的核苷酸 1792-4374）；或 (iv) SEQ ID NO: 6 的核苷酸 1792-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 6 的核苷酸 1792-4374）具有至少約 80%、至少約 85%、至少約 86%、至少約 87%、至少約 88%、至少約 89%、至少約 90%、至少約 91%、至少約 92%、至少約 93%、至少約 94%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性；其中 N 末端部分和 C 末端部分一起具有 FVIII 多肽活性；並且其中密碼子優化的核苷酸序列相對於 SEQ ID NO: 16 含有更少去穩定元件。在其他實施例中，編碼具有 FVIII 活性的多肽的核苷酸序列含有至多 9 個、至多 8 個、至多 7 個、至多 6 個或至多 5 個去穩定元件。在其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列含有至多 4 個、至多 3 個、至多 2 個或至多 1 個去穩定元件。在又其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列不含去穩定元件。

【0306】 在其他實施例中，基因盒包含編碼 FVIII 多肽的密碼子優化的核苷酸序列，其中密碼子優化的核苷酸序列包含與 (i) 選自 SEQ ID NO: 1、2、3、4、5、6、70 和 71 的胺基酸序列的核苷酸 58-4374 或 (ii) 選自 SEQ ID NO: 1、2、3、4、5、6、70 和 71 的胺基酸序列的核苷酸 58-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 1、2、3、4、5、6、70 或 71 的核苷酸 58-4374）具有至少約 80%、至少約 85%、至少約 89%、至少約 90%、至少約 91%、至少約 92%、至少約 93%、至少約

94%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性的核酸序列；並且其中密碼子優化的核苷酸序列相對於 SEQ ID NO: 16 含有更少去穩定元件。在其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列含有至多 9 個、至多 8 個、至多 7 個、至多 6 個或至多 5 個去穩定元件。在其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列含有至多 4 個、至多 3 個、至多 2 個或至多 1 個去穩定元件。在又其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列不含去穩定元件。

【0307】 在親代 FVIII 序列（SEQ ID NO: 16）中存在 10 個去穩定元件：6 個 ATTTA 序列（SEQ ID NO: 23）和 4 個 TAAAT 序列（SEQ ID NO: 24）。在一個實施例中，這些位點的序列發生突變以破壞優化的 FVIII SEQ ID NO: 1-6、70 和 71 中的去穩定元件。優化的序列中這些元件各自的位置以及相應核苷酸的序列顯示於表 3 中。

E. 潛在啟動子結合位點

【0308】 在一些實施例中，基因盒包含編碼 FVIII 多肽的密碼子優化的核苷酸序列，其中所述核苷酸序列相對於 SEQ ID NO: 16 含有更少潛在啟動子結合位點。在其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列含有至多 9 個、至多 8 個、至多 7 個、至多 6 個或至多 5 個潛在啟動子結合位點。在其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列含有至多 4 個、至多 3 個、至多 2 個或至多 1 個潛在啟動子結合位點。在又其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列不含潛在啟動子結合位點。

【0309】 在一個特定實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含編碼 FVIII 多肽的 N 末端部分的第一核酸序列和編碼 FVIII 多肽的 C 末端部分的第二核酸序列；其中第一核酸序列與 (i) SEQ ID NO: 3 的核苷酸 58-1791；(ii) SEQ ID NO: 3 的核苷酸 1-1791；(iii) SEQ ID NO: 4 的核苷酸 58-1791；或 (iv) SEQ ID NO: 4 的核苷酸 1-1791 具有至少約 80%、至少約 85%、至少約 86%、至少約 87%、至少約 88%、至少約 89%、至少約 90%、至少約 91%、至少約 92%、至少約 93%、至少約 94%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性；其中 N

末端部分和C末端部分一起具有FVIII多肽活性；並且其中密碼子優化的核苷酸序列相對於SEQ ID NO: 16含有更少潛在啟動子結合位點。在其他實施例中，編碼FVIII多肽的密碼子優化的核苷酸序列含有至多9個、至多8個、至多7個、至多6個或至多5個潛在啟動子結合位點。在其他實施例中，編碼FVIII多肽的密碼子優化的核苷酸序列含有至多4個、至多3個、至多2個或至多1個潛在啟動子結合位點。在又其他實施例中，編碼FVIII多肽的密碼子優化的核苷酸序列不含潛在啟動子結合位點。

【0310】 在另一個實施例中，編碼FVIII多肽的密碼子優化的核苷酸序列包含編碼FVIII多肽的N末端部分的第一核酸序列和編碼FVIII多肽的C末端部分的第二核酸序列；其中第二核酸序列與(i) SEQ ID NO: 5的核苷酸1792-4374；(ii) SEQ ID NO: 6的核苷酸1792-4374；(iii) SEQ ID NO: 5的核苷酸1792-2277和2320-4374（即，不含編碼B結構域或B結構域片段的核苷酸的SEQ ID NO: 5的核苷酸1792-4374）；或(iv) SEQ ID NO: 6的核苷酸1792-2277和2320-4374（即，不含編碼B結構域或B結構域片段的核苷酸的SEQ ID NO: 6的核苷酸1792-4374）具有至少約80%、至少約85%、至少約86%、至少約87%、至少約88%、至少約89%、至少約90%、至少約91%、至少約92%、至少約93%、至少約94%、至少約95%、至少約96%、至少約97%、至少約98%或至少約99%序列同一性；其中N末端部分和C末端部分一起具有FVIII多肽活性；並且其中密碼子優化的核苷酸序列相對於SEQ ID NO: 16含有更少潛在啟動子結合位點。在其他實施例中，編碼FVIII多肽的密碼子優化的核苷酸序列含有至多9個、至多8個、至多7個、至多6個或至多5個潛在啟動子結合位點。在其他實施例中，編碼FVIII多肽的密碼子優化的核苷酸序列含有至多4個、至多3個、至多2個或至多1個潛在啟動子結合位點。在又其他實施例中，編碼FVIII多肽的密碼子優化的核苷酸序列不含潛在啟動子結合位點。

【0311】 在其他實施例中，基因盒包含編碼FVIII多肽的密碼子優化的核苷酸序列，其中所述核苷酸序列包含與(i)選自SEQ ID NO: 1、2、3、4、5、6、70和71的胺基酸序列的核苷酸58-4374或(ii)選自SEQ ID NO: 1、2、3、4、5、6、70和71的胺基酸序列的核苷酸58-2277和2320-4374（即，

不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 1、2、3、4、5、6、70 或 71 的核苷酸 58-4374）具有至少約 80%、至少約 85%、至少約 89%、至少約 90%、至少約 91%、至少約 92%、至少約 93%、至少約 94%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性的核酸序列；並且其中密碼子優化的核苷酸序列相對於 SEQ ID NO: 16 含有更少潛在啟動子結合位點。在其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列含有至多 9 個、至多 8 個、至多 7 個、至多 6 個或至多 5 個潛在啟動子結合位點。在其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列含有至多 4 個、至多 3 個、至多 2 個或至多 1 個潛在啟動子結合位點。在又其他實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列不含潛在啟動子結合位點。

【0312】 TATA 盒是通常在真核生物的啟動子區中發現的調控序列。其用作通用轉錄因子 TATA 結合蛋白（TBP）的結合位點。TATA 盒通常包含序列 TATAA（SEQ ID NO: 28）或相似變異體。然而，編碼序列內的 TATA 盒可以抑制全長蛋白質的轉譯。在野生型 BDD FVIII 序列（SEQ ID NO: 16）中存在 10 個潛在啟動子結合序列：5 個 TATAA 序列（SEQ ID NO: 28）和 5 個 TTATA 序列（SEQ ID NO: 29）。在一些實施例中，在本公開文本的 FVIII 基因中消除至少 1 個、至少 2 個、至少 3 個或至少 4 個啟動子結合位點。在一些實施例中，在本公開文本的 FVIII 基因中消除至少 5 個啟動子結合位點。在其他實施例中，在本公開文本的 FVIII 基因中消除至少 6 個、至少 7 個或至少 8 個啟動子結合位點。在一個實施例中，在本公開文本的 FVIII 基因中消除至少 9 個啟動子結合位點。在一個特定實施例中，在本公開文本的 FVIII 基因中消除所有啟動子結合位點。優化的序列中每個潛在啟動子結合位點的位置以及相應核苷酸的序列顯示於表 3 中。

F. 其他順式作用負調控元件

【0313】 除了上述的 MAR/ARS 序列、去穩定元件和潛在啟動子位點以外，在野生型 BDD FVIII 序列（SEQ ID NO: 16）中還可以鑑定出若干個其他潛在抑制性序列。在未經優化的 BDD FVIII 序列中可以鑑定出兩個富含 AU 的序列元件（ARE）（ATTTTATT（SEQ ID NO: 30）；和 ATTTTTAA

（SEQ ID NO: 31））以及多聚 A 位點（AAAAAAA；SEQ ID NO: 26）、多聚 T 位點（TTTTTT；SEQ ID NO: 25）和剪接位點（GGTGAT；SEQ ID NO: 27）。這些元件中的一個或多個可以從優化的 FVIII 序列去除。優化的序列中這些位點各自的位置以及相應核苷酸的序列顯示於表 3 中。

【0314】 在某些實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含編碼 FVIII 多肽的 N 末端部分的第一核酸序列和編碼 FVIII 多肽的 C 末端部分的第二核酸序列；其中第一核酸序列與 (i) SEQ ID NO: 3 的核苷酸 58-1791；(ii) SEQ ID NO: 3 的核苷酸 1-1791；(iii) SEQ ID NO: 4 的核苷酸 58-1791；或 (iv) SEQ ID NO: 4 的核苷酸 1-1791 具有至少約 80%、至少約 85%、至少約 86%、至少約 87%、至少約 88%、至少約 89%、至少約 90%、至少約 91%、至少約 92%、至少約 93%、至少約 94%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性；其中 N 末端部分和 C 末端部分一起具有 FVIII 多肽活性；並且其中密碼子優化的核苷酸序列不含一個或多個順式作用負調控元件，例如剪接位點、多聚 T 序列、多聚 A 序列、ARE 序列或其任何組合。

【0315】 在另一個實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含編碼 FVIII 多肽的 N 末端部分的第一核酸序列和編碼 FVIII 多肽的 C 末端部分的第二核酸序列；其中第二核酸序列與 (i) SEQ ID NO: 5 的核苷酸 1792-4374；(ii) SEQ ID NO: 6 的核苷酸 1792-4374；(iii) SEQ ID NO: 5 的核苷酸 1792-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 5 的核苷酸 1792-4374）；或 (iv) SEQ ID NO: 6 的核苷酸 1792-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 6 的核苷酸 1792-4374）具有至少約 80%、至少約 85%、至少約 86%、至少約 87%、至少約 88%、至少約 89%、至少約 90%、至少約 91%、至少約 92%、至少約 93%、至少約 94%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性；其中 N 末端部分和 C 末端部分一起具有 FVIII 多肽活性；並且其中密碼子優化的核苷酸序列不含一個或多個順式作用負調控元件，例如剪接位點、多聚 T 序列、多聚 A 序列、ARE 序列或其任何組合。

【0316】 在其他實施例中，基因盒包含編碼 FVIII 多肽的密碼子優化的核苷酸序列，其中所述核苷酸序列包含與 (i) 選自 SEQ ID NO: 1、2、3、4、5、6、70 和 71 的胺基酸序列的核苷酸 58-4374 或 (ii) 選自 SEQ ID NO: 1、2、3、4、5、6、70 和 71 的胺基酸序列的核苷酸 58-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 1、2、3、4、5、6、70 或 71 的核苷酸 58-4374）具有至少約 80%、至少約 85%、至少約 89%、至少約 90%、至少約 91%、至少約 92%、至少約 93%、至少約 94%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性的核酸序列；並且其中密碼子優化的核苷酸序列不含一個或多個順式作用負調控元件，例如剪接位點、多聚 T 序列、多聚 A 序列、ARE 序列或其任何組合。

【0317】 在一些實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含編碼 FVIII 多肽的 N 末端部分的第一核酸序列和編碼 FVIII 多肽的 C 末端部分的第二核酸序列；其中第一核酸序列與 (i) SEQ ID NO: 3 的核苷酸 58-1791；(ii) SEQ ID NO: 3 的核苷酸 1-1791；(iii) SEQ ID NO: 4 的核苷酸 58-1791；或 (iv) SEQ ID NO: 4 的核苷酸 1-1791 具有至少約 80%、至少約 85%、至少約 86%、至少約 87%、至少約 88%、至少約 89%、至少約 90%、至少約 91%、至少約 92%、至少約 93%、至少約 94%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性；其中 N 末端部分和 C 末端部分一起具有 FVIII 多肽活性；並且其中密碼子優化的核苷酸序列不含剪接位點 GGTGAT（SEQ ID NO: 27）。在一些實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含編碼 FVIII 多肽的 N 末端部分的第一核酸序列和編碼 FVIII 多肽的 C 末端部分的第二核酸序列；其中第一核酸序列與 (i) SEQ ID NO: 3 的核苷酸 58-1791；(ii) SEQ ID NO: 3 的核苷酸 1-1791；(iii) SEQ ID NO: 4 的核苷酸 58-1791；或 (iv) SEQ ID NO: 4 的核苷酸 1-1791 具有至少約 80%、至少約 85%、至少約 86%、至少約 87%、至少約 88%、至少約 89%、至少約 90%、至少約 91%、至少約 92%、至少約 93%、至少約 94%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性；其中 N 末端部分和 C 末端部分一起具有 FVIII 多肽活

性；並且其中密碼子優化的核苷酸序列不含多聚 T 序列（SEQ ID NO: 25）。在一些實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含編碼 FVIII 多肽的 N 末端部分的第一核酸序列和編碼 FVIII 多肽的 C 末端部分的第二核酸序列；其中第一核酸序列與 (i) SEQ ID NO: 3 的核苷酸 58-1791；(ii) SEQ ID NO: 3 的核苷酸 1-1791；(iii) SEQ ID NO: 4 的核苷酸 58-1791；或 (iv) SEQ ID NO: 4 的核苷酸 1-1791 具有至少約 80%、至少約 85%、至少約 86%、至少約 87%、至少約 88%、至少約 89%、至少約 90%、至少約 91%、至少約 92%、至少約 93%、至少約 94%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性；其中 N 末端部分和 C 末端部分一起具有 FVIII 多肽活性；並且其中密碼子優化的核苷酸序列不含多聚 A 序列（SEQ ID NO: 26）。在一些實施例中，編碼 FVIII 多肽的密碼子優化的核苷酸序列包含編碼 FVIII 多肽的 N 末端部分的第一核酸序列和編碼 FVIII 多肽的 C 末端部分的第二核酸序列；其中第一核酸序列與 (i) SEQ ID NO: 3 的核苷酸 58-1791；(ii) SEQ ID NO: 3 的核苷酸 1-1791；(iii) SEQ ID NO: 4 的核苷酸 58-1791；或 (iv) SEQ ID NO: 4 的核苷酸 1-1791 具有至少約 80%、至少約 85%、至少約 86%、至少約 87%、至少約 88%、至少約 89%、至少約 90%、至少約 91%、至少約 92%、至少約 93%、至少約 94%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性；其中 N 末端部分和 C 末端部分一起具有 FVIII 多肽活性；並且其中密碼子優化的核苷酸序列不含 ARE 元件（SEQ ID NO: 30 或 SEQ ID NO: 31）。

【0318】 在一些實施例中，基因盒包含編碼 FVIII 多肽的密碼子優化的核苷酸序列，其中密碼子優化的核苷酸序列包含與 (i) 選自 SEQ ID NO: 1、2、3、4、5、6、70 和 71 的胺基酸序列的核苷酸 58-4374 或 (ii) 選自 SEQ ID NO: 1、2、3、4、5、6、70 和 71 的胺基酸序列的核苷酸 58-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 1、2、3、4、5、6、70 或 71 的核苷酸 58-4374）具有至少約 80%、至少約 85%、至少約 89%、至少約 90%、至少約 91%、至少約 92%、至少約 93%、至少約 94%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性的核酸序列；並且其中密碼子優化的核苷酸序列不含剪接位點

GGTGAT（SEQ ID NO: 27）。在一些實施例中，基因盒包含編碼 FVIII 多肽的密碼子優化的核苷酸序列，其中密碼子優化的核苷酸序列包含與 (i) 選自 SEQ ID NO: 1、2、3、4、5、6、70 和 71 的胺基酸序列的核苷酸 58-4374 或 (ii) 選自 SEQ ID NO: 1、2、3、4、5、6、70 和 71 的胺基酸序列的核苷酸 58-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 1、2、3、4、5、6、70 或 71 的核苷酸 58-4374）具有至少約 80%、至少約 85%、至少約 89%、至少約 90%、至少約 91%、至少約 92%、至少約 93%、至少約 94%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性的核酸序列；並且其中密碼子優化的核苷酸序列不含多聚 T 序列（SEQ ID NO: 25）。在一些實施例中，基因盒包含編碼 FVIII 多肽的密碼子優化的核苷酸序列，其中密碼子優化的核苷酸序列包含與 (i) 選自 SEQ ID NO: 1、2、3、4、5、6、70 和 71 的胺基酸序列的核苷酸 58-4374 或 (ii) 選自 SEQ ID NO: 1、2、3、4、5、6、70 和 71 的胺基酸序列的核苷酸 58-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 1、2、3、4、5、6、70 或 71 的核苷酸 58-4374）具有至少約 80%、至少約 85%、至少約 89%、至少約 90%、至少約 91%、至少約 92%、至少約 93%、至少約 94%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性的核酸序列；並且其中密碼子優化的核苷酸序列不含多聚 A 序列（SEQ ID NO: 26）。在一些實施例中，基因盒包含編碼 FVIII 多肽的密碼子優化的核苷酸序列，其中密碼子優化的核苷酸序列包含與 (i) 選自 SEQ ID NO: 1、2、3、4、5、6、70 和 71 的胺基酸序列的核苷酸 58-4374 或 (ii) 選自 SEQ ID NO: 1、2、3、4、5、6、70 和 71 的胺基酸序列的核苷酸 58-2277 和 2320-4374（即，不含編碼 B 結構域或 B 結構域片段的核苷酸的 SEQ ID NO: 1、2、3、4、5、6、70 或 71 的核苷酸 58-4374）具有至少約 80%、至少約 85%、至少約 89%、至少約 90%、至少約 91%、至少約 92%、至少約 93%、至少約 94%、至少約 95%、至少約 96%、至少約 97%、至少約 98%或至少約 99%序列同一性的核酸序列；並且其中密碼子優化的核苷酸序列不含 ARE 元件（SEQ ID NO: 30 或 SEQ ID NO: 31）。

【0319】 在其他實施例中，本公開文本的優化的 FVIII 序列不含抗病毒基序、莖環結構和重複序列中的一個或多個。

【0320】 在仍其他實施例中，轉錄起始位點周圍的核苷酸變為 kozak 共有序列（GCCGCCACCATGC（SEQ ID NO: 32），其中加下劃線的核苷酸是起始密碼子）。在其他實施例中，可以添加或去除限制位點以有助於選殖過程。

b. FIX和編碼FIX蛋白的多核苷酸序列

【0321】 在一些實施例中，核酸分子包含第一 ITR、第二 ITR 和編碼靶序列的基因盒，其中靶序列編碼治療性蛋白質，其中治療性蛋白質包含 FIX 多肽。在一些實施例中，FIX 多肽包含 FIX 或其變異體或片段，其中 FIX 或其變異體或片段具有 FIX 活性。

【0322】 人類 FIX 是絲胺酸蛋白酶，其是血液凝結級聯的固有路徑的重要組分。如本文所用的「IX 因子」或「FIX」是指凝結因子蛋白和種類及其序列變異體，並且包括但不限於人類 FIX 前體多肽（「prepro」）的 461 單鏈胺基酸序列、成熟人類 FIX（SEQ ID NO: 125）的 415 單鏈胺基酸序列和 R338L FIX（Padua）變異體（SEQ ID NO: 126）。FIX 包括具有血液凝結 FIX 的典型特徵的任何形式的 FIX 分子。如本文所用，「IX 因子」和「FIX」旨在涵蓋包含以下的多肽：結構域 Gla（含有 γ-羧基麩胺酸殘基的區域）、EGF1 和 EGF2（含有與人表皮生長因子同源的序列的區域）、啟動肽（由成熟 FIX 的殘基 R136-R180 形成的「AP」）和 C 末端蛋白酶結構域（「Pro」）或業內已知的這些結構域的同義詞，或者可以是保留天然蛋白的生物學活性的至少一部分的截短片段或序列變異體。已經選殖出 FIX 或序列變異體，如美國專利號 4,770,999 和 7,700,734 中所述，並且編碼人類 FIX 的 cDNA 已經被分離，表徵並選殖至表現載體中（參見例如，Choo 等人, Nature 299:178-180 (1982)；Fair 等人, Blood 64:194-204 (1984)；和 Kurachi 等人, Proc. Natl. Acad. Sci., U.S.A. 79:6461-6464 (1982)）。由 Simioni 等人在 2009 年表徵的 FIX 的一種特定變異體 R338L FIX（Padua）變異體（SEQ ID NO: 2）包含獲得功能突變，所述突變與 Padua 變異體活性相對于天然 FIX 的幾乎 8 倍增加相關（表 4）。FIX 變異體還可以包括具有一個或多個不影響 FIX 多

肽的 FIX 活性的保守胺基酸取代的任何 FIX 多肽。在一些實施例中，FIX 變異體包含通過可切割連接子（例如，IDELVION®）融合的 rFIX-白蛋白。參見 US 7,939,632，通過引用以其整體併入本文。

表4：例示性FIX序列

SEQ ID NO: 125（成熟FIX多肽）
1:YNSGKLEEFV QGNLERECME EKCSFEEARE VFENTERTTE FWKQYVDGDQ CESNPCLNGG 61:SCKDDINSYE CWCPFGFEGK NCELDVTCNI KNGRCEQFCK NSADNKVVCS CTEGYRLAEN 121:QKSCEPAVPF PCGRVSVSQT SKLTRAETVF PDVDYVNSTE AETILDNITQ STQSFNDFTR 181:VVGGEDAKPG QFPWQVVLNG KVDAFCGGSI VNEKWIVTAA HCVETGVKIT VVAGEHNIEE 241:TEHTEQKRNV IRIIPHHNYN AAINKYNHDI ALLELDEPLV LNSYVTPICI ADKEYTNIFL 301:KFGSGYVSGW GRVFHKGRSA LVLQYLRVPL VDRATCLRST KFTIYNNMFC AGFHEGGRDS 361:CQGDSGGPHV TEVEGTSFLT GIISWGEECA MKGKYGIYTK VSRYVNWIKE KTKLT
SEQ ID NO: 126（成熟Padua(R338L)FIX多肽）
1:YNSGKLEEFV QGNLERECME EKCSFEEARE VFENTERTTE FWKQYVDGDQ CESNPCLNGG 61:SCKDDINSYE CWCPFGFEGK NCELDVTCNI KNGRCEQFCK NSADNKVVCS CTEGYRLAEN 121:QKSCEPAVPF PCGRVSVSQT SKLTRAETVF PDVDYVNSTE AETILDNITQ STQSFNDFTR 181:VVGGEDAKPG QFPWQVVLNG KVDAFCGGSI VNEKWIVTAA HCVETGVKIT VVAGEHNIEE 241:TEHTEQKRNV IRIIPHHNYN AAINKYNHDI ALLELDEPLV LNSYVTPICI ADKEYTNIFL 301:KFGSGYVSGW GRVFHKGRSA LVLQYLRVPL VDRATCLLST KFTIYNNMFC AGFHEGGRDS 361:CQGDSGGPHV TEVEGTSFLT GIISWGEECA MKGKYGIYTK VSRYVNWIKE KTKLT
SEQ ID NO: 127（FIX信號多肽和前肽）
1: MQRVNMIMAE SPGLITICLL GYLLSAEC*TV FLDHENANKI LNRPKR*
SEQ ID NO: 160（FIX-連接子-白蛋白）
YNSGKLEEFV QGNLERECME EKCSFEEARE VFENTERTTE FWKQYVDGDQ 50

CESNPCLNGG SCKDDINSYE CWCPFGFEGK NCELDVTCNI KNGRCEQFCK 100

NSADNKVVCS CTEGYRLAEN QKSCEPAVPF PCGRVSVSQT SKLTRAETVF 150

PDVDYVNSTE AETILDNITQ STQSFNDFTR VVGGEDAKPG QFPWQVVLNG 200

KVDAFCGGSI VNEKWIVTAA HCVETGVKIT VVAGEHNIEE TEHTEQKRNV 250

IRIIPHHNYN AAINKYNHDI ALLELDEPLV LNSYVTPICI ADKEYTNIFL 300

KFGSGYVSGW GRVFHKGRSA LVLQYLRVPL VDRATCLRST KFTIYNNMFC 350

AGFHEGGRDS CQGDSGGPHV TEVEGTSFLT GIISWGEECA MKGKYGIYTK 400

VSRYVNWIKE KTKLTPVSQT SKLTRAETVF PDVDAHKSEV AHRFKDLGEE 450

NFKALVLIAF AQYLQQCPFE DHVKLVNEVT EFAKTCVADE SAENCDKSLH 500

TLFGDKLCTV ATLRETYGEM ADCCAKQEPE RNECFLQHKD DNPNLPRLVR 550

PEVDVMCTAF HDNEETFLKK YLYEIARRHP YFYAPELLFF AKRYKAAFTE 600

CCQAADKAAC LLPKLDELRD EGKASSAKQR LKCASLQKFG ERAFKAWAVA 650

RLSQRFPKAE FAEVSKLVTD LTKVHTECCH GDLLECADDR ADLAKYICEN 700

QDSISSKLKE CCEKPLLEKS HCIAEVENDE MPADLPSLAA DFVESKDVCK 750

NYAEAKDVFL GMFLYEYARR HPDYSVVLLL RLAKTYETTL EKCCAAADPH 800

ECYAKVFDEF KPLVEEPQNL IKQNCELFEQ LGEYKFQNAL LVRYTKKVPQ 850

VSTPTLVEVS RNLGKVGSKC CKHPEAKRMP CAEDYLSVVL NQLCVLHEKT 900

PVSDRVTKCC TESLVNRRPC FSALEVDETY VPKEFNAETF TFHADICTLS 950

EKERQIKKQT ALVELVKHKP KATKEQLKAV MDDFAAFVEK CCKADDKETC 1000

FAEEGKKLVA ASQAALGL 1018

SEQ ID NO: 161（FIX）

YNSGKLEEFVQGNLERECMEEKCSFEEAREVFENTERTTEFWKQYVDGDQCESNPCLNGGSCKDDI

NSYECWCPFGFEGKNCELDVTCNIKNGRCEQFCKNSADNKVVCSCTEGYRLAENQKSCEPAVPFPC

GRVSVSQTSKLTRAETVFPDVDYVNSTEAETILDNITQSTQSFNDFTRVVGGEDAKPGQFPWQVVLN
GKVDAFCGGSIVNEKWIVTAAHCVETGVKITVVAGEHNIEETEHTEQKRNVIRIIPHHNYNAAINKY
NHDIALLELDEPLVLNSYVTPICIADKEYTNIFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDR
ATCLRSTKFTIYNNMFCAGFHEGGRDSCQGDSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTK
VSRYVNWIKEKTKLTPVSQTSKLT

SEQ ID NO: 162（連接子）

RAETVFPDV

SEQ ID NO: 163（白蛋白）

DAHKSEVAHRFKDLGEENFKALVLIAFAQYLQQCPFEDHVKLVNEVTEFAKTCVADESAENCDKSL
HTLFGDKLCTVATLRETYGEMADCCAKQEPERNECFLQHKDDNPNLPRLVRPEVDVMCTAFHDNE
ETFLKKYLYEIARRHPYFYAPELLFFAKRYKAAFTECCQAADKAACLLPKLDELRDEGKASSAKQRL
KCASLQKFGERAFKAWAVARLSQRFPKAEFAEVSKLVTDLTKVHTECCHGDLLECADDRADLAKYI
CENQDSISSKLKECCEKPLLEKSHCIAEVENDEMPADLPSLAADFVESKDVCKNYAEAKDVFLGMFL
YEYARRHPDYSVVLLLRLAKTYETTLEKCCAAADPHECYAKVFDEFKPLVEEPQNLIKQNCELFEQL
GEYKFQNALLVRYTKKVPQVSTPTLVEVSRNLGKVGSKCCKHPEAKRMPCAEDYLSVVLNQLCVL
HEKTPVSDRVTKCCTESLVNRRPCFSALEVDETYVPKEFNAETFTFHADICTLSEKERQIKKQTALVE
LVKHKPKATKEQLKAVMDDFAAFVEKCCKADDKETCFAEEGKKLVAASQAALGL

SEQ ID NO: 164（FIX(XTEN)-Fc）*

MQRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRPKRYNSGKLEEFVQGNLERECMEE
KCSFEEAREVFENTERTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYECWCPFGFEGKNCELDVT
CNIKNGRCEQFCKNSADNKVVCSCTEGYRLAENQKSCEPAVPFPCGRVSVSQTSKLTRAETVFPDVD
YVNSTEAETILDGPSPGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSA
PGTSTEPSEGSAPGASSNITQSTQSFNDFTRVVGGEDAKPGQFPWQVVLNGKVDAFCGGSIVNEKWI

VTAAHCVETGVKITVVAGEHNIEETEHTEQKRNVIRIIPHHNYNAAINKYNHDIALLELDEPLVLNSY

VTPICIADKEYTNIFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDRATCLLSTKFTIYNNMFCA

GFHEGGRDSCQGDSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKEKTKLT**DK**

THTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVH

NAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQV

YTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTV

DKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPG

SEQ ID NO: 165（FIX-FXIa-AE288）*

YNSGKLEEFVQGNLERECMEEKCSFEEAREVFENTERTTEFWKQYVDGDQCESNPCLNGGSCKDDI

NSYECWCPFGFEGKNCELDVTCNIKNGRCEQFCKNSADNKVVCSCTEGYRLAENQKSCEPAVPFPC

GRVSVSQTSKLTRAETVFPDVDYVNSTEAETILDNITQSTQSFNDFTRVVGGEDAKPGQFPWQVVLN

GKVDAFCGGSIVNEKWIVTAAHCVETGVKITVVAGEHNIEETEHTEQKRNVIRIIPHHNYNAAINKY

NHDIALLELDEPLVLNSYVTPICIADKEYTNIFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDR

ATCLRSTKFTIYNNMFCAGFHEGGRDSCQGDSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTK

VSRYVNWIKEKTKLTGKLTRAETGGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGS

ETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP

ESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESAT

PESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPAT

SGSETPGTSESATPESGPGTSTEPSEGSAP

SEQ ID NO: 166（FIX-Fc-Fc）**

MQRVNMIMAESPGLITICLLGYLLSAECTVFLDHENANKILNRPKRYNSGKLEEFVQGNLERECMEE

KCSFEEAREVFENTERTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYECWCPFGFEGKNCELDVT

CNIKNGRCEQFCKNSADNKVVCSCTEGYRLAENQKSCEPAVPFPCGRVSVSQTSKLTRAETVFPDVD

YVNSTEAETILDNITQSTQSFNDFTRVVGGEDAKPGQFPWQVVLNGKVDAFCGGSIVNEKWIVTAA

HCVETGVKITVVAGEHNIEETEHTEQKRNVIRIIPHHNYNAAINKYNHDIALLELDEPLVLNSYVTPIC
IADKEYTNIFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDRATCLRSTKFTIYNNMFCAGFHEG
GRDSCQGDSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKEKTKLTDKTHTCPP
CPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQ
YNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKN
QVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSV
MHEALHNHYTQKSLSLSPGKRRRRSGGGGSGGGGSGGGGSGGGGSRRRRDKTHTCPPCPAPELLGG
PSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVV
SVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVK
GFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHY
TQKSLSLSPGK

* 灰色陰影 = 信號肽；下劃線 = XTEN序列；粗體 = Fc。

** 美國專利號9,856,468的SEQ ID NO: 67，將所述專利通過引用以其整體併入本文。

【0323】 FIX 多肽為 55 kDa，合成為由三個區域構成的前原多肽鏈（SEQ ID NO: 125）：28 個胺基酸的信號肽（SEQ ID NO: 127 的胺基酸 1 至 28）、麩胺酸殘基的 γ-羧基化所需的 18 個胺基酸的前肽（胺基酸 29 至 46）和 415 個胺基酸的成熟 IX 因子（SEQ ID NO: 125 或 126）。前肽是 γ-羧基麩胺酸結構域 N 末端的 18 個胺基酸殘基的序列。前肽結合維生素 K 依賴性 γ 羧化酶，然後通過內源蛋白酶從 FIX 的前體多肽切割，所述內源蛋白酶最可能是 PACE（成對鹼性胺基酸切割酶），也稱為弗林蛋白酶或 PCSK3。在不進行 γ 羧基化的情況下，Gla 結構域不能結合鈣以採取將蛋白質錨定至帶負電的磷脂表面所必需的正確構象，從而使 IX 因子無功能。即使將其羧基化，Gla 結構域也依賴於前肽的切割以獲得正常功能，因為保留的前肽干擾與鈣和磷脂的最優結合所必需的 Gla 結構域的構象變化。在人體內，所得成熟 IX 因子由肝細胞作為無活性酶原分泌至血流中，所述酶原

是 415 個胺基酸殘基的單鏈蛋白，其含有以重量計大約 17%的碳水化合物（Schmidt, A. E.等人 (2003) Trends Cardiovasc Med, 13: 39）。

【0324】 成熟 FIX 由若干個結構域構成，所述結構域以 N 末端至 C 末端構型是：GLA 結構域、EGF1 結構域、EGF2 結構域、啟動肽（AP）結構域和蛋白酶（或催化）結構域。短連接子連接 EGF2 結構域與 AP 結構域。FIX 含有兩種啟動肽，分別由 R145-A146 和 R180-V181 形成。在啟動後，單鏈 FIX 變為 2 鏈分子，其中兩條鏈通過二硫鍵連接。凝血因子可以通過替代其啟動肽來工程化，從而得到改變的啟動特異性。在哺乳動物體內，成熟 FIX 必須由啟動的 XI 因子啟動以產生 IX 因子 a。在將 FIX 啟動為 FIXa 之後，蛋白酶結構域提供 FIX 的催化活性。啟動的 VIII 因子（FVIIIa）是完全表現 FIXa 活性的特定輔因子。

【0325】 在某些實施例中，FIX 多肽包含血漿衍生的 FIX 的 Thr148 等位基因形式，並且具有與內源 FIX 類似的結構和功能特徵。

【0326】 許多功能性 FIX 變異體是業內已知的。國際公開號 WO 02/040544 A3 在第 4 頁第 9-30 行和第 15 頁第 6-31 行披露了展現增加的對肝素抑制的抗性的突變體。國際公開號 WO 03/020764 A2 在表 2 和表 3（在第 14-24 頁）和第 12 頁第 1-27 行披露了具有降低的 T 細胞免疫原性的 FIX 突變體。國際公開號 WO 2007/149406 A2 在第 4 頁第 1 行至第 19 頁第 11 行披露了功能性突變體 FIX 分子，其展現增加的蛋白質穩定性、增加的體內和體外半衰期以及增加的蛋白酶抗性。WO 2007/149406 A2 在第 19 頁第 12 行至第 20 行第 9 行還披露了嵌合的和其他變異體 FIX 分子。國際公開號 WO 08/118507 A2 在第 5 頁第 14 行至第 6 頁第 5 行披露了展現增加的凝血活性的 FIX 突變體。國際公開號 WO 09/051717 A2 在第 9 頁第 11 行至第 20 頁第 2 行披露了 FIX 突變體，其具有增加的 N 連接和/或 O 連接糖基化位點數目，這導致增加的半衰期和/或回收率。國際公開號 WO 09/137254 A2 在第 2 頁第[006]段至第 5 頁第[011]段以及第 16 頁第[044]段至第 24 頁第[057]段還披露了具有增加的糖基化位元點數目的 IX 因子突變體。國際公開號 WO 09/130198 A2 在第 4 頁第 26 行至第 12 頁第 6 行披露了功能性突變體 FIX 分子，其具有增加的糖基化位點數目，這導致增加的半衰期。國際公開

號 WO 09/140015 A2 在第 11 頁第[0043]段至第 13 頁第[0053]段披露了功能性 FIX 突變體，其具有增加的可用於聚合物（例如，PEG）綴合的 Cys 殘基數。2011 年 7 月 11 日提交並且作為 WO 2012/006624 於 2012 年 1 月 12 日公開的國際申請號 PCT/US2011/043569 中所述的 FIX 多肽也通過引用以其整體併入本文。在一些實施例中，FIX 多肽包含融合至白蛋白的 FIX 多肽（例如，FIX-白蛋白）。在某些實施例中，FIX 多肽是 IDELVION®或 rIX-FP。

【0327】 另外，已經在血友病個體體內鑑定出 FIX 中的數百個非功能性突變，其中許多突變披露於國際公開號 WO 09/137254 A2 的第 11-14 頁的表 6 中。此類非功能性突變不包括於本發明中，但提供了哪些突變更有可能或更不可能產生功能性 FIX 多肽的額外指導。

【0328】 在一個實施例中，FIX 多肽（或融合多肽的 IX 因子部分）包含與 SEQ ID NO: 1 或 2 中所述的序列（SEQ ID NO: 125 或 126 的胺基酸 1 至 415）或者可替代地與前肽序列或與前肽和信號序列（全長 FIX）至少 70%、至少 80%、至少 85%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%、至少 99%或 100%相同的胺基酸序列。在另一個實施例中，FIX 多肽包含與 SEQ ID NO: 2 中所述的序列至少 70%、至少 80%、至少 85%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%、至少 99%或 100%相同的胺基酸序列。

【0329】 FIX 促凝劑活性表示為一個或多個國際單位（IU）。1 IU 的 FIX 活性大致對應于 1 毫升正常人血漿中 FIX 的量。若干種測定可用於量測 FIX 活性，包括一步式凝血測定（啟動部分促凝血酶原激酶時間；aPTT）、凝血酶產生時間（TGA）和旋轉凝血彈性量測法（ROTEM®）。本發明考慮了與 FIX 序列、天然（如來自人、非人靈長類動物、哺乳動物（包括家畜））序列片段和非天然序列變異體具有同源性的序列，所述序列保留 FIX 的生物活性或生物功能的至少一部分和/或可用於預防、治療、調解或改善凝結因子相關疾病、缺乏、障礙或病症（例如，與創傷、手術或凝結因子缺乏相關的出血發作）。與人類 FIX 具有同源性的序列可以通過標準同源性搜索技術（如 NCBI BLAST）來發現。

【0330】 在某些實施例中，FIX序列經過密碼子優化。密碼子優化的FIX序列的例子包括但不限於國際公開號WO 2016/004113 A1的SEQ ID NO: 1和54-58，將所述專利通過引用以其整體併入本文。

c. FVII和編碼FVII蛋白的多核苷酸序列

【0331】 在一些實施例中，核酸分子包含第一ITR、第二ITR和編碼靶序列的基因盒，其中靶序列編碼治療性蛋白質，其中治療性蛋白質包含VII因子多肽。在一些實施例中，FVII多肽包含FVII或其變異體或片段，其中其變異體或片段具有FVII活性。

【0332】 「VII因子」（「FVII」或「F7」；也稱為因子7、凝結VII因子、血清VII因子、血清凝血酶原轉化加速因子、SPCA、前轉化素（proconvertin）和依他凝血素α（eptacog α））是絲胺酸蛋白酶，其是凝結級聯的一部分。在一個實施例中，本文所述的核酸中的凝血因子是FVII。重組啟動的VII因子（「FVII」）已經廣泛用於治療大出血，如發生在患有以下疾病的患者中的大出血：血友病A或B、缺乏凝結XI因子、FVII、血小板功能缺陷、血小板減少症或血管性血友病。

【0333】 重組啟動的FVII（rFVIIa；Novoseven®）用於治療以下患者的出血發作：(i)具有針對FVIII或FIX的中和抗體（抑制劑）的血友病患者，(ii)具有FVII缺乏的患者，或(iii)正在進行外科手術的患有血友病A或B且具有抑制劑的患者。然而，Novoseven®展示較差功效。通常需要高濃度的FVIIa重複劑量來控制出血，因為FVIIa與啟動的血小板的親和性低，半衰期短，並且在組織因子不存在下酶活性較差。因此，對於具有FVIII和FIX抑制劑和/或具有FVII缺乏的血友病患者，對更好的治療和預防選擇存在未滿足的醫療需求。

【0334】 在一個實施例中，基因盒編碼FVII的成熟形式或其變異體。FVII包括Gla結構域、兩個EGF結構域（EGF-1和EGF-2）以及絲胺酸蛋白酶結構域（或肽酶S1結構域），所述絲胺酸蛋白酶結構域在絲胺酸蛋白酶的肽酶S1家族的所有成員之間高度保守，如例如糜蛋白酶的情況。FVII作為單鏈酶原（即，可啟動的FVII）和完全啟動的雙鏈形式出現。

C. 生長因子

【0335】 在一些實施例中，核酸分子包含第一 ITR、第二 ITR 和編碼靶序列的基因盒，其中靶序列編碼治療性蛋白質，並且其中治療性蛋白質包含生長因子。生長因子可以選自業內已知的任何生長因子。在一些實施例中，生長因子是激素。在其他實施例中，生長因子是細胞激素。在一些實施例中，生長因子是細胞激素。

【0336】 在一些實施例中，生長因子是腎上腺髓質素（AM）。在一些實施例中，生長因子是血管生成素（Ang）。在一些實施例中，生長因子是自分泌運動因子。在一些實施例中，生長因子是骨形態發生蛋白（BMP）。在一些實施例中，BMP 選自 BMP2、BMP4、BMP5 和 BMP7。在一些實施例中，生長因子是睫狀神經營養因子家族成員。在一些實施例中，睫狀神經營養因子家族成員選自睫狀神經營養因子（CNTF）、白血病抑制因子（LIF）、白介素-6（IL-6）。在一些實施例中，生長因子是集落刺激因子。在一些實施例中，集落刺激因子選自巨噬細胞集落刺激因子（m-CSF）、粒細胞集落刺激因子（G-CSF）和粒細胞巨噬細胞集落刺激因子（GM-CSF）。在一些實施例中，生長因子是表皮生長因子（EGF）。在一些實施例中，生長因子是肝配蛋白。在一些實施例中，肝配蛋白選自肝配蛋白 A1、肝配蛋白 A2、肝配蛋白 A3、肝配蛋白 A4、肝配蛋白 A5、肝配蛋白 B1、肝配蛋白 B2 和肝配蛋白 B3。在一些實施例中，生長因子是促紅細胞生成素（EPO）。在一些實施例中，生長因子是成纖維細胞生長因子（FGF）。在一些實施例中，FGF 選自 FGF1、FGF2、FGF3、FGF4、FGF5、FGF6、FGF7、FGF8、FGF9、FGF10、FGF11、FGF12、FGF13、FGF14、FGF15、FGF16、FGF17、FGF18、FGF19、FGF20、FGF21、FGF22 和 FGF23。在一些實施例中，生長因子是胎牛促體素（FBS）。在一些實施例中，生長因子是 GDNF 家族成員。在一些實施例中，GDNF 家族成員選自神經膠質細胞株衍生的神經營養因子（GDNF）、神經秩蛋白、普塞芬蛋白和阿特敏蛋白。在一些實施例中，生長因子是生長分化因子-9（GDF9）。在一些實施例中，生長因子是肝細胞生長因子（HGF）。在一些實施例中，生長因子是肝腫瘤衍生的生長因子（HDGF）。在一些實施例中，生長因子是胰島素。在一些實施例中，生長因子是類胰島素生長因子。在一些實施例中，類胰島素生長

因子是類胰島素生長因子-1（IGF-1）或IGF-2。在一些實施例中，生長因子是白介素（IL）。在一些實施例中，IL選自IL-1、IL-2、IL-3、IL-4、IL-5、IL-6和IL-7。在一些實施例中，生長因子是角質形成細胞生長因子（KGF）。在一些實施例中，生長因子是遷移刺激因子（MSF）。在一些實施例中，生長因子是巨噬細胞刺激蛋白（MSP或類肝細胞生長因子蛋白（HGFLP））。在一些實施例中，生長因子是肌生成抑制蛋白（GDF-8）。在一些實施例中，生長因子是神經調節蛋白。在一些實施例中，神經調節蛋白選自神經調節蛋白1（NRG1）、NRG2、NRG3和NRG4。在一些實施例中，生長因子是神經營養蛋白。在一些實施例中，生長因子是腦衍生的神經營養因子（BDNF）。在一些實施例中，生長因子是神經生長因子（NGF）。在一些實施例中，NGF是神經營養蛋白-3（NT-3）或NT-4。在一些實施例中，生長因子是胎盤生長因子（PGF）。在一些實施例中，生長因子是血小板衍生的生長因子（PDGF）。在一些實施例中，生長因子是腎胺酶（RNLS）。在一些實施例中，生長因子是T細胞生長因子（TCGF）。在一些實施例中，生長因子是血小板生成素（TPO）。在一些實施例中，生長因子是轉化生長因子。在一些實施例中，轉化生長因子是轉化生長因子α（TGF-α）或TGF-β。在一些實施例中，生長因子是腫瘤壞死因子-α（TNF-α）。在一些實施例中，生長因子是血管內皮生長因子（VEGF）。

D. 微小RNA（miRNA）

【0337】 微小RNA（miRNA）是通過抑制轉譯或誘導信使RNA（mRNA）降解來負調控基因表現的小非編碼RNA分子（約18-22個核苷酸）。從其發現以來，miRNA已經參與多種細胞過程，包括細胞凋亡、分化和細胞增生，並且其已顯示在致癌中具有關鍵作用。miRNA調控基因表現的能力使得體內miRNA表現成為基因療法中有價值的工具。

【0338】 本公開文本的某些態樣涉及類質體核酸分子，其包含第一ITR、第二ITR和編碼靶序列的基因盒，其中靶序列編碼miRNA，並且其中第一ITR和/或第二ITR是非腺相關病毒的ITR（例如，第一ITR和/或第二ITR來自非AAV）。miRNA可以是業內已知的任何miRNA。在一些實

施例中，miRNA 下調靶基因的表現。在某些實施例中，靶基因選自 SOD1、HTT、RHO 或其任何組合。

【0339】 在一些實施例中，基因盒編碼一種 miRNA。在一些實施例中，基因盒編碼多於一種 miRNA。在一些實施例中，基因盒編碼兩種或更多種不同的 miRNA。在一些實施例中，基因盒編碼相同 miRNA 的兩個或更多個拷貝。在一些實施例中，基因盒編碼相同治療性蛋白質的兩種或更多種變異體。在某些實施例中，基因盒編碼一種或多種 miRNA 以及一種或多種治療性蛋白質。

【0340】 在一些實施例中，miRNA 是天然存在的 miRNA。在一些實施例中，miRNA 是工程化 miRNA。在一些實施例中，miRNA 是人工 miRNA。在某些實施例中，miRNA 包含 Evers 等人, *Molecular Therapy 26(9)*:1-15（印刷版之前的電子版，2018 年 6 月）披露的 miHTT 工程化的 miRNA。在某些實施例中，miRNA 包含 Dirren 等人, *Annals of Clinical and Translational Neurology 2(2)*:167-84（2015 年 2 月）披露的 miR SOD1 人工 miRNA。在某些實施例中，miRNA 包含 miR-708，其靶向 RHO（參見 Behrman 等人, *JCB 192(6)*:919-27 (2011)）。

【0341】 在一些實施例中，miRNA 通過下調基因抑制劑的表現來上調所述基因的表現。在一些實施例中，抑制劑是天然（例如，野生型）抑制劑。在一些實施例中，抑制劑得自突變的、異源性和/或錯誤表現的基因。

E. 異源部分

【0342】 在一些實施例中，核酸分子包含第一 ITR、第二 ITR 和編碼靶序列的基因盒，其中靶序列編碼治療性蛋白質，並且其中治療性蛋白質包含至少一個異源部分。在一些實施例中，異源部分融合至治療性蛋白質的 N 末端或 C 末端。在其他實施例中，異源部分插入治療性蛋白質內的兩個胺基酸之間。

【0343】 在一些實施例中，治療性蛋白質包含 FVIII 多肽和異源部分，所述異源部分插入 FVIII 多肽內的兩個胺基酸之間。在一些實施例中，異源部分在選自表 5 的一個或多個插入位點插入 FVIII 多肽內。在一些實施例中，異源胺基酸序列可以在以下文獻中披露的任何位點插入本公開文本

的核酸分子編碼的凝血因子多肽內：國際公開號 WO 2013/123457 A1、WO 2015/106052 A1 或美國公開號 2015/0158929 A1，將上述文獻通過引用以其整體併入本文。在一個特定實施例中，治療性蛋白質包含 FVIII 和異源部分，其中所述異源部分緊鄰相對於成熟 FVIII 的胺基酸 745 下游插入 FVIII 內。在一個特定實施例中，治療性蛋白質包含 FVIII 和 XTEN，其中所述 XTEN 緊鄰相對於成熟 FVIII 的胺基酸 745 下游插入 FVIII 內。在一個特定實施例中，FVIII 包含對應於成熟人類 FVIII（SEQ ID NO: 15）的胺基酸 746-1646 的缺失，並且異源部分緊鄰對應於成熟人類 FVIII（SEQ ID NO: 15）的胺基酸 745 下游插入。

表5：FVIII異源部分插入位元點

插入位點	結構域	插入位點	結構域	插入位點	結構域
3	A1	375	A2	1749	A3
18	A1	378	A2	1796	A3
22	A1	399	A2	1802	A3
26	A1	403	A2	1827	A3
40	A1	409	A2	1861	A3
60	A1	416	A2	1896	A3
65	A1	442	A2	1900	A3
81	A1	487	A2	1904	A3
116	A1	490	A2	1905	A3
119	A1	494	A2	1910	A3
130	A1	500	A2	1937	A3
188	A1	518	A2	2019	A3
211	A1	599	A2	2068	C1
216	A1	603	A2	2111	C1

220	A1	713	A2	2120	C1
224	A1	745	B	2171	C2
230	A1	1656	a3區域	2188	C2
333	A1	1711	A3	2227	C2
336	A1	1720	A3	2332	CT
339	A1	1725	A3		

【0344】 在一些實施例中，治療性蛋白質包含 FIX 多肽和異源部分，所述異源部分插入 FIX 多肽內的兩個胺基酸之間。在一些實施例中，異源部分在選自表 5 的一個或多個插入位點插入 FIX 多肽內。在一些實施例中，異源胺基酸序列可以在國際申請號 PCT/US2017/015879 中所披露的任何位點插入本公開文本的核酸分子編碼的凝血因子多肽內，將所述申請通過引用以其整體併入本文。在一個特定實施例中，治療性蛋白質包含 FIX 多肽和異源部分，其中所述異源部分緊鄰相對於成熟 FIX 的胺基酸 166 下游插入 FIX 多肽內。在一個特定實施例中，治療性蛋白質包含 FIX 多肽和 XTEN，其中所述 XTEN 緊鄰相對於成熟 FVIII 的胺基酸 166 下游插入 FIX 內。

表6：FIX異源部分插入位元點

插入位點	**結構域**	**插入位點**	**結構域**	**插入位點**	**結構域**
52	EGF1	149	AP	257	催化
59	EGF1	162	AP	265	催化
66	EGF1	166	AP	277	催化
80	EGF1	174	AP	283	催化
85	EGF2	188	催化	292	催化
89	EGF2	202	催化	316	催化

103	EGF2	224	催化	341	催化
105	EGF2	226	催化	354	催化
113	EGF2	228	催化	392	催化
129	連接子	230	催化	403	催化
142	連接子	240	催化	413	催化

【0345】 在其他實施例中，本公開文本的治療性蛋白質還包含2個、3個、4個、5個、6個、7個或8個異源核苷酸序列。在一些實施例中，所有異源部分是相同的。在一些實施例中，至少一個異源部分與其他異源部分不同。在一些實施例中，本公開文本可以包含2個、3個、4個、5個、6個或多於7個串聯的異源部分。

【0346】 在一些實施例中，異源部分增加治療性蛋白質的半衰期（是「半衰期延長劑」）。

【0347】 在一些實施例中，異源部分是具有非結構化或結構化特徵的肽或多肽，所述特徵與併入本公開文本的蛋白質中時體內半衰期的延長相關。非限制性例子包括白蛋白、白蛋白片段、免疫球蛋白的Fc片段、人絨毛膜促性腺激素的β亞基的C末端肽（CTP）、HAP序列、XTEN序列、轉鐵蛋白或其片段、PAS多肽、聚甘胺酸連接子、聚絲胺酸連接子、白蛋白結合部分或者這些多肽的任何片段、衍生物、變異體或組合。在一個特定實施例中，異源胺基酸序列是免疫球蛋白恒定區或其部分、轉鐵蛋白、白蛋白或PAS序列。在一些態樣中，異源部分包括血管性血友病因子或其片段。在其他相關態樣中，異源部分可以包括非多肽部分的附著位元點（例如，半胱胺酸胺基酸），所述非多肽部分是例如聚乙二醇（PEG）、羥乙基澱粉（HES）、聚唾液酸或者這些元件的任何衍生物、變異體或組合。在一些態樣中，異源部分包含半胱胺酸胺基酸，其用作非多肽部分的附著位元點，所述非多肽部分是例如聚乙二醇（PEG）、羥乙基澱粉（HES）、聚唾液酸或者這些元件的任何衍生物、變異體或組合。

【0348】 在一個具體實施例中，第一異源部分是業內已知的半衰期延長分子，並且第二異源部分是業內已知的半衰期延長分子。在某些實施例中，第一異源部分（例如，第一 Fc 部分）和第二異源部分（例如，第二 Fc 部分）彼此締合形成二聚體。在一個實施例中，第二異源部分是第二 Fc 部分，其中第二 Fc 部分與第一異源部分（例如，第一 Fc 部分）連接或締合。例如，第二異源部分（例如，第二 Fc 部分）可以通過連接子連接至第一異源部分（例如，第一 Fc 部分），或者通過共價或非共價鍵與第一異源部分締合。

【0349】 在一些實施例中，異源部分是如下多肽，其包含至少約 10 個、至少約 100 個、至少約 200 個、至少約 300 個、至少約 400 個、至少約 500 個、至少約 600 個、至少約 700 個、至少約 800 個、至少約 900 個、至少約 1000 個、至少約 1100 個、至少約 1200 個、至少約 1300 個、至少約 1400 個、至少約 1500 個、至少約 1600 個、至少約 1700 個、至少約 1800 個、至少約 1900 個、至少約 2000 個、至少約 2500 個、至少約 3000 個或至少約 4000 個胺基酸，基本上由所述胺基酸組成，或由所述胺基酸組成。在其他實施例中，異源部分是如下多肽，其包含約 100 個至約 200 個胺基酸、約 200 個至約 300 個胺基酸、約 300 個至約 400 個胺基酸、約 400 個至約 500 個胺基酸、約 500 個至約 600 個胺基酸、約 600 個至約 700 個胺基酸、約 700 個至約 800 個胺基酸、約 800 個至約 900 個胺基酸或約 900 個至約 1000 個胺基酸，基本上由所述胺基酸組成，或由所述胺基酸組成。

【0350】 在某些實施例中，異源部分改良治療性蛋白質的一種或多種藥物代謝動力學特性但不顯著影響其生物學活性或功能。

【0351】 在某些實施例中，異源部分增加本公開文本的治療性蛋白質的體內和/或體外半衰期。在其他實施例中，異源部分有助於本公開文本的治療性蛋白質或其片段（例如，在 FVIII 蛋白的蛋白質水解切割後包含異源部分的片段）的可視化或定位。本公開文本的治療性蛋白質或其片段的可視化和/或定位可以是體內的、體外的、離體的或其組合。

【0352】 在其他實施例中，異源部分增加本公開文本的治療性蛋白質或其片段（例如，在治療性蛋白質（例如，凝血因子）的蛋白質水解切割

後包含異源部分的片段）的穩定性。如本文所用，術語「穩定性」是指本領域公認的響應環境條件（例如，升高或降低的溫度）維持治療性蛋白質的一種或多種物理特性的量度。在某些態樣中，物理特性可以是維持治療性蛋白質的共價結構（例如，不存在蛋白質水解切割、不期望的氧化或脫醯胺作用）。在其他態樣中，物理特性還可以是處於正確折疊狀態的治療性蛋白質的存在（例如，不存在可溶性或不溶性聚集體或沈澱物）。在一個態樣中，治療性蛋白質的穩定性是通過測定治療性蛋白質的生物物理特性來量測，所述生物物理特性是例如熱穩定性、pH 解折疊譜、糖基化的穩定去除、溶解度、生化功能（例如，結合至蛋白質、受體或配位體的能力）等和/或其組合。在另一個態樣中，生化功能是通過相互作用的結合親和性來證實。在一個態樣中，蛋白質穩定性的量度是熱穩定性，即對熱激發的抗性。穩定性可以使用業內已知的方法來量測，如 HPLC（高效液相色譜）、SEC（尺寸排阻色譜）、DLS（動態光散射）等。量測熱穩定性的方法包括但不限於差式掃描量熱法（DSC）、差式掃描螢光測定法（DSF）、圓二色性（CD）和熱激發測定。

【0353】 在某些態樣中，由本公開文本的核酸分子編碼的治療性蛋白質包含至少一種半衰期延長劑，即相對於缺少這種異源部分的相應治療性蛋白質的體內半衰期增加治療性蛋白質的體內半衰期的異源部分。治療性蛋白質的體內半衰期可以通過業內熟習此項技術者已知的任何方法來確定，例如活性測定（例如，產色測定或一步式凝血 aPTT 測定，其中治療性蛋白質包含 FVIII 多肽）、ELISA、ROTEM®等。

【0354】 在一些實施例中，一種或多種半衰期延長劑的存在導致治療性蛋白質的半衰期與缺少這一種或多種半衰期延長劑的相應蛋白質的半衰期相比有所增加。包含半衰期延長劑的治療性蛋白質的半衰期是缺少這種半衰期延長劑的相應治療性蛋白質的體內半衰期的至少約 1.5 倍、至少約 2 倍、至少約 2.5 倍、至少約 3 倍、至少約 4 倍、至少約 5 倍、至少約 6 倍、至少約 7 倍、至少約 8 倍、至少約 9 倍、至少約 10 倍、至少約 11 倍或至少約 12 倍長。

【0355】 在一個實施例中，包含半衰期延長劑的治療性蛋白質的半衰期是缺少這種半衰期延長劑的相應蛋白質的體內半衰期的約 1.5 倍至約 20 倍、約 1.5 倍至約 15 倍或約 1.5 倍至約 10 倍長。在另一個實施例中，包含半衰期延長劑的治療性蛋白質的半衰期與缺少這種半衰期延長劑的相應蛋白質的體內半衰期相比延長約 2 倍至約 10 倍、約 2 倍至約 9 倍、約 2 倍至約 8 倍、約 2 倍至約 7 倍、約 2 倍至約 6 倍、約 2 倍至約 5 倍、約 2 倍至約 4 倍、約 2 倍至約 3 倍、約 2.5 倍至約 10 倍、約 2.5 倍至約 9 倍、約 2.5 倍至約 8 倍、約 2.5 倍至約 7 倍、約 2.5 倍至約 6 倍、約 2.5 倍至約 5 倍、約 2.5 倍至約 4 倍、約 2.5 倍至約 3 倍、約 3 倍至約 10 倍、約 3 倍至約 9 倍、約 3 倍至約 8 倍、約 3 倍至約 7 倍、約 3 倍至約 6 倍、約 3 倍至約 5 倍、約 3 倍至約 4 倍、約 4 倍至約 6 倍、約 5 倍至約 7 倍或約 6 倍至約 8 倍。

【0356】 在其他實施例中，包含半衰期延長劑的治療性蛋白質的半衰期是至少約 17 小時、至少約 18 小時、至少約 19 小時、至少約 20 小時、至少約 21 小時、至少約 22 小時、至少約 23 小時、至少約 24 小時、至少約 25 小時、至少約 26 小時、至少約 27 小時、至少約 28 小時、至少約 29 小時、至少約 30 小時、至少約 31 小時、至少約 32 小時、至少約 33 小時、至少約 34 小時、至少約 35 小時、至少約 36 小時、至少約 48 小時、至少約 60 小時、至少約 72 小時、至少約 84 小時、至少約 96 小時或至少約 108 小時。

【0357】 在仍其他實施例中，包含半衰期延長劑的治療性蛋白質的半衰期是約 15 小時至約 2 周、約 16 小時至約 1 周、約 17 小時至約 1 周、約 18 小時至約 1 周、約 19 小時至約 1 周、約 20 小時至約 1 周、約 21 小時至約 1 周、約 22 小時至約 1 周、約 23 小時至約 1 周、約 24 小時至約 1 周、約 36 小時至約 1 周、約 48 小時至約 1 周、約 60 小時至約 1 周、約 24 小時至約 6 天、約 24 小時至約 5 天、約 24 小時至約 4 天、約 24 小時至約 3 天或約 24 小時至約 2 天。

【0358】 在一些實施例中，包含半衰期延長劑的治療性蛋白質的每名個體的平均半衰期是約 15 小時、約 16 小時、約 17 小時、約 18 小時、約 19 小時、約 20 小時、約 21 小時、約 22 小時、約 23 小時、約 24 小時（1

天）、約25小時、約26小時、約27小時、約28小時、約29小時、約30小時、約31小時、約32小時、約33小時、約34小時、約35小時、約36小時、約40小時、約44小時、約48小時（2天）、約54小時、約60小時、約72小時（3天）、約84小時、約96小時（4天）、約108小時、約120小時（5天）、約6天、約7天（1周）、約8天、約9天、約10天、約11天、約12天、約13天或約14天。

【0359】 一種或多種半衰期延長劑可以融合至治療性蛋白質的C末端或N末端或者插入治療性蛋白質內。

1. 免疫球蛋白恒定區或其部分

【0360】 在另一個態樣中，異源部分包含一個或多個免疫球蛋白恒定區或其部分（例如，Fc區）。在一個實施例中，本公開文本的分離的核酸分子還包含編碼免疫球蛋白恒定區或其部分的異源核酸序列。在一些實施例中，免疫球蛋白恒定區或其部分是Fc區。

【0361】 免疫球蛋白恒定區包括表示為CH（恒定重）結構域（CH1、CH2等）的結構域。根據同種型（即IgG、IgM、IgA、IgD或IgE），恒定區可以包括3個或4個CH結構域。一些同種型（例如IgG）恒定區還含有鉸鏈區。參見Janeway等人2001, *Immunobiology*, Garland Publishing, N.Y., N.Y。

【0362】 本公開文本的免疫球蛋白恒定區或其部分可以從多種不同來源獲得。在一個實施例中，免疫球蛋白恒定區或其部分衍生自人類免疫球蛋白。然而，應理解，免疫球蛋白恒定區或其部分可以衍生自另一哺乳動物物種的免疫球蛋白，所述哺乳動物物種包括例如齧齒類動物（例如，小鼠、大鼠、兔、豚鼠）或非人靈長類動物（例如，黑猩猩、獼猴）物種。此外，免疫球蛋白恒定區或其部分可以衍生自任何免疫球蛋白類別（包括IgM、IgG、IgD、IgA和IgE）以及任何免疫球蛋白同種型（包括IgG1、IgG2、IgG3和IgG4）。在一個實施例中，使用人類同種型IgG1。

【0363】 多個免疫球蛋白恒定區基因序列（例如，人恒定區基因序列）能以可公開獲得的存放物的形式獲得。可以選擇具有特定效應子功能（或缺少特定效應子功能）或具有特定修飾的恒定區結構域序列以降低免疫原

性。已經公開了許多抗體和抗體編碼基因的序列並且適宜的 Ig 恒定區序列（例如，鉸鏈、CH2 和/或 CH3 序列或其部分）可以使用本領域公認的技術衍生自這些序列。然後可以改變或合成使用任何前述方法獲得的遺傳物質以獲得本公開文本的多肽。應進一步瞭解，本公開文本的範圍涵蓋恒定區 DNA 序列的等位基因、變異體和突變。

【0364】 免疫球蛋白恒定區或其部分的序列可以例如使用聚合酶鏈式反應和所選用於擴增所關注結構域的引子來選殖。為了從抗體選殖免疫球蛋白恒定區或其部分的序列，可以從雜交瘤、脾臟或淋巴細胞分離 mRNA，逆轉錄成 DNA，並通過 PCR 擴增抗體基因。PCR 擴增方法詳細描述於以下文獻中：美國專利號 4,683,195；4,683,202；4,800,159；4,965,188；以及例如，“PCR Protocols: A Guide to Methods and Applications” Innis 等人編輯, Academic Press, San Diego, CA (1990)；Ho 等人 1989. Gene 77:51；Horton 等人 1993. *Methods Enzymol.* 217:270。PCR 可以基於所公開的重鏈和輕鏈 DNA 和胺基酸序列由共有恒定區引子或由更具特異性的引子來起始。PCR 還可以用於分離編碼抗體輕鏈和重鏈的 DNA 選殖體。在這種情況下，可以通過共有引子或較大同源探針（如小鼠恒定區探針）篩選文庫。適於擴增抗體基因的多個引子組是業內已知的（例如，基於經純化抗體的 N 末端序列的 5’引子（Benhar 和 Pastan. 1994. *Protein Engineering* 7:1509）；cDNA 末端的快速擴增（Ruberti, F.等人 1994. *J. Immunol. Methods* 173:33）；抗體前導序列（Larrick 等人 1989 *Biochem. Biophys. Res. Commun.* 160:1250））。抗體序列的選殖還描述於 Newman 等人於 1995 年 1 月 25 日提交的美國專利號 5,658,570 中，將所述專利通過引用併入本文。

【0365】 本文所用的免疫球蛋白恒定區可以包括所有結構域和鉸鏈區或其部分。在一個實施例中，免疫球蛋白恒定區或其部分包含 CH2 結構域、CH3 結構域和鉸鏈區，即 Fc 區或 FcRn 結合配偶體。

【0366】 如本文所用，術語「Fc 區」定義為多肽中對應于天然 Ig 的 Fc 區的部分，即如通過其兩條重鏈的相應 Fc 結構域的二聚締合所形成。天然 Fc 區與另一 Fc 區形成同二聚體。相比之下，如本文所用的術語「基因融合的 Fc 區」或「單鏈 Fc 區」（scFc 區）是指合成二聚 Fc 區，其包括基因連接於

單條多肽鏈內（即，編碼於單個連續基因序列中）的 Fc 結構域。參見國際公開號 WO 2012/006635，通過引用以其整體併入本文。

【0367】 在一個實施例中，「Fc 區」是指單條 Ig 重鏈的如下部分，其始於緊鄰木瓜蛋白酶切割位點（即，IgG 中的殘基 216，以重鏈恒定區的第一個殘基為 114）上游的鉸鏈區並終於抗體的 C 末端。因此，完整 Fc 區至少包含鉸鏈結構域、CH2 結構域和 CH3 結構域。

【0368】 免疫球蛋白恒定區或其部分可以是 FcRn 結合配偶體。FcRn 在成人上皮組織中有活性並在腸腔、肺氣道、鼻表面、陰道表面、結腸和直腸表面中表現（美國專利號 6,485,726）。FcRn 結合配偶體是免疫球蛋白的結合至 FcRn 的一部分。

【0369】 已經從包括人類在內的若干哺乳動物物種分離出 FcRn 受體。人 FcRn、猴 FcRn、大鼠 FcRn 和小鼠 FcRn 的序列是已知的（Story 等人 1994, J. Exp. Med. 180:2377）。FcRn 受體在相對低 pH 下結合 IgG（但不結合其他免疫球蛋白類別，如 IgA、IgM、IgD 和 IgE），以內腔至漿膜方向跨細胞主動轉運 IgG，然後在間質液中發現的相對高 pH 下釋放 IgG。其在成人上皮組織中表現（美國專利號 6,485,726、6,030,613、6,086,875；WO 03/077834；US2003-0235536A1），包括肺和腸上皮（Israel 等人 1997, Immunology 92:69）、腎近端小管上皮（Kobayashi 等人 2002, Am. J. Physiol. Renal Physiol. 282:F358）以及鼻上皮、陰道表面和膽道系統表面。

【0370】 可用于本公開文本中的 FcRn 結合配偶體涵蓋可由 FcRn 受體特異性結合的分子，包括完整 IgG、IgG 的 Fc 片段和包括 FcRn 受體的完整結合區的其他片段。已經基於 X 射線晶體學描述了 IgG 的 Fc 部分中結合至 FcRn 受體的區域（Burmeister 等人 1994, Nature 372:379）。Fc 與 FcRn 的主要接觸區域靠近 CH2 和 CH3 結構域的接合點。Fc-FcRn 接觸全部位元於單條 Ig 重鏈內。FcRn 結合配偶體包括完整 IgG、IgG 的 Fc 片段和 IgG 的包括 FcRn 的完整結合區的其他片段。主要接觸位點包括 CH2 結構域的胺基酸殘基 248、250-257、272、285、288、290-291、308-311 和 314 以及 CH3 結構域的胺基酸殘基 385-387、428 和 433-436。對免疫球蛋白或免疫球蛋白片段或區域的胺基酸編號的提及都是基於 Kabat 等人 1991, Sequences of

Proteins of Immunological Interest, U.S. Department of Public Health, Bethesda, Md。

【0371】 結合至 FcRn 的 Fc 區或 FcRn 結合配偶體可以通過 FcRn 有效穿梭跨越上皮障壁，由此提供全身投予所需治療分子的非侵入性手段。另外，包含 Fc 區或 FcRn 結合配偶體的融合蛋白被表現 FcRn 的細胞內吞。但是這些融合蛋白再次再循環進循環中，而不是經標記以供降解，由此增加這些蛋白質的體內半衰期。在某些實施例中，免疫球蛋白恒定區的部分是 Fc 區或 FcRn 結合配偶體，其通常通過二硫鍵和其他非特異性相互作用與另一 Fc 區或另一 FcRn 結合配偶體締合，以形成二聚體和更高階多聚體。

【0372】 兩個 FcRn 受體可以結合單個 Fc 分子。晶體學數據表明，每個 FcRn 分子結合 Fc 同二聚體的單一多肽。在一個實施例中，將 FcRn 結合配偶體（例如，IgG 的 Fc 片段）與生物活性分子連接提供口服、經頰、舌下、經直腸、經陰道、作為經鼻投予的氣溶膠或通過肺途徑或通過眼部途徑遞送生物活性分子的手段。在另一個實施例中，凝血因子蛋白能以侵入性方式（例如，皮下、靜脈內）投予。

【0373】 FcRn 結合配偶體區域是可由 FcRn 受體特異性結合，隨後通過 Fc 區的 FcRn 受體主動轉運的分子或其部分。特異性結合是指兩個分子形成在生理條件下相對穩定的複合物。特異性結合的特徵在於區分於通常具有低親和性和中至高容量的非特異性結合的高親和性和低至中容量。通常，在親和常數 KA 高於 10^6 M^{-1} 或高於 10^8 M^{-1} 時，將結合視為特異性的。如果需要，可以通過改變結合條件在不顯著影響特異性結合的情況下減少非特異性結合。熟習此項技術者使用常規技術可以優化適當的結合條件，如分子濃度、溶液的離子強度、溫度、允許結合的時間、阻斷劑（例如，血清白蛋白、乳酪蛋白）濃度等。

【0374】 在某些實施例中，由本公開文本的核酸分子編碼的治療性蛋白質包含一個或多個截短 Fc 區，但是其足以將 Fc 受體（FcR）結合特性賦予 Fc 區。例如，Fc 區中結合至 FcRn 的部分（即，FcRn 結合部分）包含 IgG1 的約胺基酸 282-438，EU 編號（主要接觸位點是 CH2 結構域的胺基酸 248、250-257、272、285、288、290-291、308-311 和 314 以及 CH3 結構域的胺

基酸殘基 385-387、428 和 433-436）。因此，本公開文本的 Fc 區可以包含 FcRn 結合部分或由所述 FcRn 結合部分組成。FcRn 結合部分可以衍生自任何同種型（包括 IgGl、IgG2、IgG3 和 IgG4）的重鏈。在一個實施例中，使用來自人類同種型 IgG1 的抗體的 FcRn 結合部分。在另一個實施例中，使用來自人類同種型 IgG4 的抗體的 FcRn 結合部分。

【0375】 Fc 區可以從多種不同來源獲得。在一個實施例中，多肽的 Fc 區衍生自人類免疫球蛋白。然而，應理解，Fc 部分可以衍生自另一哺乳動物物種的免疫球蛋白，所述哺乳動物物種包括例如齧齒類動物（例如，小鼠、大鼠、兔、豚鼠）或非人類靈長類動物（例如，黑猩猩、獼猴）物種。此外，Fc 結構域或其部分的多肽可以衍生自任何免疫球蛋白類別（包括 IgM、IgG、IgD、IgA 和 IgE）以及任何免疫球蛋白同種型（包括 IgG1、IgG2、IgG3 和 IgG4）。在另一個實施例中，使用人類同種型 IgG1。

【0376】 在某些實施例中，Fc 變異體賦予至少一種由包含所述野生型 Fc 結構域的 Fc 部分賦予的效應子功能的變化（例如，Fc 區結合至 Fc 受體（例如 FcγRI、FcγRII 或 FcγRIII）或補體蛋白（例如，C1q）或觸發抗體依賴性細胞毒性（ADCC）、吞噬作用或補體依賴性細胞毒性（CDCC）的能力的改進或降低）。在其他實施例中，Fc 變異體提供工程化半胱胺酸殘基。

【0377】 本公開文本的 Fc 區可以利用本領域公認的 Fc 變異體，已知所述 Fc 變異體可賦予效應子功能和/或 FcR 或 FcRn 結合的變化（例如，增強或降低）。具體來說，本公開文本的 Fc 區可以包括例如在以下文獻中披露的一個或多個胺基酸位置處的變化（例如，取代）：國際 PCT 公開案 WO 88/07089A1、WO 96/14339A1、WO 98/05787A1、WO 98/23289A1、WO 99/51642A1、WO 99/58572A1、WO 00/09560A2、WO 00/32767A1、WO 00/42072A2、WO 02/44215A2、WO 02/060919A2、WO 03/074569A2、WO 04/016750A2、WO 04/029207A2、WO 04/035752A2、WO 04/063351A2、WO 04/074455A2、WO 04/099249A2、WO 05/040217A2、WO 04/044859、WO 05/070963A1、WO 05/077981A2、WO 05/092925A2、WO 05/123780A2、WO 06/019447A1、WO 06/047350A2 和 WO 06/085967A2；美國專利公開號

US 2007/0231329、US 2007/0231329、US 2007/0237765、US 2007/0237766、US 2007/0237767、US 2007/0243188、US 20070248603、US 20070286859、US 20080057056；或美國專利 5,648,260；5,739,277；5,834,250；5,869,046；6,096,871；6,121,022；6,194,551；6,242,195；6,277,375；6,528,624；6,538,124；6,737,056；6,821,505；6,998,253；7,083,784；7,404,956 和 7,317,091，將每篇文獻通過引用併入本文。在一個實施例中，可以在一個或多個所公開的胺基酸位置進行特定變化（例如，業內披露的一個或多個胺基酸的特定取代）。在另一個實施例中，可以在一個或多個所公開胺基酸位置進行不同變化（例如，業內披露的一個或多個胺基酸位置的不同取代）。

【0378】 IgG 的 Fc 區或 FcRn 結合配偶體可以根據公認程式（如定點突變誘發等）來修飾，以產生將由 FcRn 結合的經修飾 IgG 或其 Fc 片段或部分。此類修飾包括遠離 FcRn 接觸位點的修飾以及在接觸位點內的修飾，所述修飾保留或甚至增強與 FcRn 的結合。例如，可以在不顯著損失 Fc 對 FcRn 的結合親和性的情況下取代人 IgG1 Fc（Fc γ1）中的以下單個胺基酸殘基：P238A、S239A、K246A、K248A、D249A、M252A、T256A、E258A、T260A、D265A、S267A、H268A、E269A、D270A、E272A、L274A、N276A、Y278A、D280A、V282A、E283A、H285A、N286A、T289A、K290A、R292A、E293A、E294A、Q295A、Y296F、N297A、S298A、Y300F、R301A、V303A、V305A、T307A、L309A、Q311A、D312A、N315A、K317A、E318A、K320A、K322A、S324A、K326A、A327Q、P329A、A330Q、P331A、E333A、K334A、T335A、S337A、K338A、K340A、Q342A、R344A、E345A、Q347A、R355A、E356A、M358A、T359A、K360A、N361A、Q362A、Y373A、S375A、D376A、A378Q、E380A、E382A、S383A、N384A、Q386A、E388A、N389A、N390A、Y391F、K392A、L398A、S400A、D401A、D413A、K414A、R416A、Q418A、Q419A、N421A、V422A、S424A、E430A、N434A、T437A、Q438A、K439A、S440A、S444A 和 K447A，其中例如 P238A 代表在位置編號 238 處由丙胺酸取代的野生型脯胺酸。作為例子，一個具體實施例併入 N297A 突變，從而去除高度保守的 N 糖基化位點。除了丙胺酸以外，可以在上文指定的位置用其他胺基酸取代野生型胺基酸。可以將突變單獨引入 Fc 中，從而產生

多於一百個與天然 Fc 不同的 Fc 區。另外，這些單獨突變中的 2 個、3 個或更多個的組合可以一起引入，從而產生另外數百個 Fc 區。

【0379】 某些上述突變可以對 Fc 區或 FcRn 結合配偶體賦予新功能。例如，一個實施例併入 N297A，從而去除高度保守的 N 糖基化位點。這個突變的作用是降低免疫原性，由此增強 Fc 區的循環半衰期，並使 Fc 區不能結合至 FcγRI、FcγRIIA、FcγRIIB 和 FcγRIIIA，並且不損害對 FcRn 的親和性（Routledge 等人 1995, Transplantation 60:847；Friend 等人 1999, Transplantation 68:1632；Shields 等人 1995, J. Biol. Chem. 276:6591）。作為從上述突變產生的新功能的另一例子，對 FcRn 的親和性可以增加，在一些情況下超過野生型的親和性。這種增加的親和性可以反映增加的「締合」速率、降低的「解離」速率或增加的「締合」速率與降低的「解離」速率二者。相信可賦予增加的對 FcRn 的親和性的突變的例子包括但不限於 T256A、T307A、E380A 和 N434A（Shields 等人 2001, J. Biol. Chem. 276:6591）。

【0380】 另外，至少三種人 Fcγ 受體似乎識別 IgG 上下鉸鏈區內的結合位點，通常是胺基酸 234-237。因此，如例如通過將人 IgG1 的胺基酸 233-236「ELLG」（SEQ ID NO: 45）替代為來自 IgG2 的相應序列「PVA」（一個胺基酸缺失），可以從這個區域的突變產生新功能的另一例子和可能降低的免疫原性。已經顯示，在已經引入此類突變時，介導多種效應子功能的 FcγRI、FcγRII 和 FcγRIII 不會結合至 IgG1。Ward 和 Ghetie 1995, Therapeutic Immunology 2:77；和 Armour 等人 1999, Eur. J. Immunol. 29:2613。

【0381】 在另一個實施例中，免疫球蛋白恒定區或其部分包含鉸鏈區或其部分中的與第二免疫球蛋白恒定區或其部分形成一個或多個二硫鍵的胺基酸序列。第二免疫球蛋白恒定區或其部分可以連接至第二多肽，使治療性蛋白質與第二多肽結合在一起。在一些實施例中，第二多肽是強化子部分。如本文所用，術語「強化子部分」是指能夠增強治療性蛋白質的活性的分子、其片段或多肽組分。強化子部分可以是輔因子，如其中治療性蛋白質是凝血因子、可溶性組織因子（sTF）或促凝血肽。因此，在啟動凝血因子後，強化子部分可用於增強凝血因子活性。

【0382】 在某些實施例中，由本公開文本的核酸分子編碼的治療性蛋白質包含對免疫球蛋白恒定區或其部分的胺基酸取代（例如，Fc 變異體），所述胺基酸取代改變 Ig 恒定區的抗原非依賴性效應子功能，特別是蛋白質的循環半衰期。

2. scFc區

【0383】 在另一個態樣中，異源部分包含 scFc（單鏈 Fc）區。在一個實施例中，本公開文本的分離的核酸分子還包含編碼 scFc 區的異源核酸序列。scFc 區在同一線性多肽鏈內包含至少兩個免疫球蛋白恒定區或其部分（例如，Fc 部分或結構域（例如，2、3、4、5、6 個或更多個 Fc 部分或結構域）），所述恒定區或其部分能夠折疊（例如，分子內或分子間折疊）以形成一個由 Fc 肽連接子連接的功能性 scFc 區。例如，在一個實施例中，本公開文本的多肽能夠通過其 scFc 區結合至至少一種 Fc 受體（例如，FcRn、FcγR 受體（例如，FcγRIII）或補體蛋白（例如，C1q）），以改良半衰期或觸發免疫效應子功能（例如，抗體依賴性細胞毒性（ADCC）、吞噬作用或補體依賴性細胞毒性（CDCC）和/或改良可製造性）。

3. CTP

【0384】 在另一個態樣中，異源部分包含人類絨毛膜促性腺激素的 β 亞基的一種 C 末端肽（CTP）或其片段、變異體或衍生物。已知插入重組蛋白中的一種或多種 CTP 肽增加所述蛋白質的體內半衰期。參見例如，美國專利號 5,712,122，通過引用以其整體併入本文。

【0385】 例示性 CTP 肽包括 DPRFQDSSSSKAPPPSLPSPSRLPGPSDTPIL（SEQ ID NO: 33）或 SSSSKAPPPSLPSPSRLPGPSDTPILPQ（SEQ ID NO: 34）。參見例如，美國專利申請公開號 US 2009/0087411 A1，通過引用併入。

4. XTEN序列

【0386】 在一些實施例中，異源部分包含一個或多個 XTEN 序列、其片段、變異體或衍生物。如本文所用，「XTEN 序列」是指延伸長度的多肽，其具有非天然存在的基本上不重複的序列，所述序列主要由小親水胺基酸構成，並且所述序列在生理條件下具有較低程度的或不具有二級或三級結

構。作為異源部分，XTEN 可以用作半衰期延長部分。另外，XTEN 可以提供所需特性，包括但不限於增強的藥物代謝動力學參數和溶解度特徵。

【0387】 將包含 XTEN 序列的異源部分併入本公開文本的蛋白質中可以賦予蛋白質以下有利特性中的一種或多種：構象柔性、增強的水溶解度、高蛋白酶抗性程度、低免疫原性、與哺乳動物受體的低結合或增加的流體力學（或斯托克斯（Stokes））半徑。

【0388】 在某些態樣中，XTEN 序列可以增加藥物代謝動力學特性，如較長體內半衰期或增加的曲線下面積（AUC），使得與相同的但不具有 XTEN 異源部分的蛋白質相比，本公開文本的蛋白質保持在體內並且持續增加的時間段具有促凝血活性。

【0389】 在一些實施例中，可用於本公開文本的 XTEN 序列是具有大於約 20、30、40、50、60、70、80、90、100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、950、1000、1200、1400、1600、1800 或 2000 個胺基酸殘基的肽或多肽。在某些實施例中，XTEN 是具有大於約 20 至約 3000 個胺基酸殘基、大於 30 至約 2500 個殘基、大於 40 至約 2000 個殘基、大於 50 至約 1500 個殘基、大於 60 至約 1000 個殘基、大於 70 至約 900 個殘基、大於 80 至約 800 個殘基、大於 90 至約 700 個殘基、大於 100 至約 600 個殘基、大於 110 至約 500 個殘基或大於 120 至約 400 個殘基的肽或多肽。在一個特定實施例中，XTEN 包含長度長於 42 個胺基酸並且短於 144 個胺基酸的胺基酸序列。

【0390】 本公開文本的 XTEN 序列可以包含一個或多個 5 至 14（例如，9 至 14）個胺基酸殘基的序列基序或與所述序列基序至少 80%、90%、91%、92%、93%、94%、95%、96%、97%、98%或 99%相同的胺基酸序列，其中所述基序包含 4 至 6 種類型的選自甘胺酸（G）、丙胺酸（A）、絲胺酸（S）、蘇胺酸（T）、麩胺酸（E）和脯胺酸（P）的胺基酸（例如，5 種胺基酸），基本上由所述胺基酸組成或由所述胺基酸組成。參見 US 2010-0239554 A1。

【0391】 在一些實施例中，XTEN 包含非重疊序列基序，其中約 80%、或至少約 85%、或至少約 90%、或約 91%、或約 92%、或約 93%、

或約 94%、或約 95%、或約 96%、或約 97%、或約 98%、或約 99%或約 100%的序列由選自從表 7 選擇的單一基序家族的非重疊序列的多個單元組成，從而得到家族序列。如本文所用，“家族”意指，XTEN 具有僅選自表 7 的單一基序類別的基序；即，AD、AE、AF、AG、AM、AQ、BC 或 BD XTEN，並且選擇 XTEN 中並非來自家族基序的任何其他胺基酸以實現所需特性，如允許通過編碼核苷酸併入限制位點、併入切割序列，或實現與治療性蛋白質的更佳連接。在 XTEN 家族的一些實施例中，XTEN 序列包含 AD 基序家族、或 AE 基序家族、或 AF 基序家族、或 AG 基序家族、或 AM 基序家族、或 AQ 基序家族、或 BC 家族、或 BD 家族的非重疊序列基序的多個單元，並且所得 XTEN 展現上述同源性範圍。在其他實施例中，XTEN 包含來自表 7 的兩個或更多個基序家族的基序序列的多個單元。可以選擇這些序列以實現所需的物理/化學特徵，包括諸如淨電荷、親水性、缺少二級結構或缺少重複性等由基序的胺基酸組成所賦予的特性，更全面地描述於下文中。在本段落的上述實施例中，可以使用本文所述方法選擇並組裝併入 XTEN 中的基序，以得到約 36 至約 3000 個胺基酸殘基的 XTEN。

表7. 12個胺基酸的XTEN序列基序和基序家族

基序家族*	**基序序列**	**SEQ ID NO:**
AD	GESPGGSSGSES	73
AD	GSEGSSGPGESS	74
AD	GSSESGSSEGGP	75
AD	GSGGEPSESGSS	76
AE、AM	GSPAGSPTSTEE	77
AE、AM、AQ	GSEPATSGSETP	78
AE、AM、AQ	GTSESATPESGP	79

基序家族*	基序序列	SEQ ID NO:
AE、AM、AQ	GTSTEPSEGSAP	80
AF、AM	GSTSESPSGTAP	81
AF、AM	GTSTPESGSASP	82
AF、AM	GTSPSGESSTAP	83
AF、AM	GSTSSTAESPGP	84
AG、AM	GTPGSGTASSSP	85
AG、AM	GSSTPSGATGSP	86
AG、AM	GSSPSASTGTGP	87
AG、AM	GASPGTSSTGSP	88
AQ	GEPAGSPTSTSE	89
AQ	GTGEPSSTPASE	90
AQ	GSGPSTESAPTE	91
AQ	GSETPSGPSETA	92
AQ	GPSETSTSEPGA	93
AQ	GSPSEPTEGTSA	94
BC	GSGASEPTSTEP	95
BC	GSEPATSGTEPS	96
BC	GTSEPSTSEPGA	97
BC	GTSTEPSEPGSA	98

基序家族*	基序序列	SEQ ID NO:
BD	GSTAGSETSTEA	99
BD	GSETATSGSETA	100
BD	GTSESATSESGA	101
BD	GTSTEASEGSAS	102

* 表示單獨基序序列，其在以各種排列一起使用時產生“家族序列”

【0392】 可以用作本公開文本的治療性蛋白質中的異源部分的XTEN序列的例子披露於例如以下文獻中：美國專利公開號2010/0239554 A1、2010/0323956 A1、2011/0046060 A1、2011/0046061 A1、2011/0077199 A1或2011/0172146 A1，或者國際專利公開號WO 2010091122 A1、WO 2010144502 A2、WO 2010144508 A1、WO 2011028228 A1、WO 2011028229 A1或WO 2011028344 A2，將每篇文獻通過引用以其整體併入本文。

【0393】 XTEN可以具有變化的長度以供插入或連接至治療性蛋白質。在一個實施例中，所述一個或多個XTEN序列的長度是基於要在融合蛋白中實現的特性或功能來選擇。根據預期的特性或功能，XTEN可以是可用作載劑的短或中等長度的序列或較長的序列。在某些實施例中，XTEN包括約6至約99個胺基酸殘基的短區段、約100至約399個胺基酸殘基的中等長度和約400至約1000以及多達約3000個胺基酸殘基的較長長度。因此，插入或連接至治療性蛋白質的XTEN可以具有約6、約12、約36、約40、約42、約72、約96、約144、約288、約400、約500、約576、約600、約700、約800、約864、約900、約1000、約1500、約2000、約2500或多達約3000個胺基酸殘基的長度。在其他實施例中，XTEN序列的長度為約6至約50、約50至約100、約100至150、約150至250、約250至400、約400至約500、約500至約900、約900至1500、約1500至2000或約2000至約3000個胺基酸殘基。在對治療性蛋白質的活性沒有不利影響的情況下，插入或連接至治療性蛋白質的XTEN的精確長度可以變化。在一個實

施例中，本文所用的一個或多個 XTEN 具有 42 個胺基酸、72 個胺基酸、144 個胺基酸、288 個胺基酸、576 個胺基酸或 864 個胺基酸的長度，並且可以選自一個或多個 XTEN 家族序列；即，AD、AE、AF、AG、AM、AQ、BC 或 BD。

【0394】 在一些實施例中，治療性蛋白質包含 FVIII 多肽和 XTEN，其中 XTEN 包含 288 個胺基酸。在一個實施例中，治療性蛋白質包含 FVIII 多肽和 XTEN，其中 XTEN 包含 288 個胺基酸，並且 XTEN 插入 FVIII 多肽的 B 結構域內。在一個特定實施例中，治療性蛋白質包含 FVIII 多肽和包含 SEQ ID NO:109 的 XTEN，並且 XTEN 插入 FVIII 多肽的 B 結構域內。在一個特定實施例中，治療性蛋白質包含 FVIII 多肽和包含 SEQ ID NO:109 的 XTEN，並且 XTEN 緊鄰成熟 FVIII 的胺基酸 745 下游插入 FVIII 多肽內。

【0395】 在一些實施例中，治療性蛋白質包含 FIX 多肽和 XTEN，其中 XTEN 包含 72 個胺基酸。在一個實施例中，治療性蛋白質包含 FIX 多肽和 XTEN，其中 XTEN 包含 72 個胺基酸，並且 XTEN 緊鄰成熟 FIX 的胺基酸 166 下游插入 FIX 多肽內。

【0396】 在一些實施例中，用於本公開文本中的 XTEN 序列與選自由下列組成之群組的序列至少 60%、70%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或 100%相同：AE42、AG42、AE48、AM48、AE72、AG72、AE108、AG108、AE144、AF144、AG144、AE180、AG180、AE216、AG216、AE252、AG252、AE288、AG288、AE324、AG324、AE360、AG360、AE396、AG396、AE432、AG432、AE468、AG468、AE504、AG504、AF504、AE540、AG540、AF540、AD576、AE576、AF576、AG576、AE612、AG612、AE624、AE648、AG648、AG684、AE720、AG720、AE756、AG756、AE792、AG792、AE828、AG828、AD836、AE864、AF864、AG864、AM875、AE912、AM923、AM1318、BC864、BD864、AE948、AE1044、AE1140、AE1236、AE1332、AE1428、AE1524、AE1620、AE1716、AE1812、AE1908、AE2004A、AG948、AG1044、AG1140、AG1236、AG1332、AG1428、AG1524、AG1620、AG1716、AG1812、AG1908、AG2004 及其任何組合。參見 US 2010-0239554 A1。在一個特定實施例中，XTEN 包含

AE42、AE72、AE144、AE288、AE576、AE864、AG 42、AG72、AG144、AG288、AG576、AG864 或其任何組合。

【0397】 可以用作本公開文本的治療性蛋白質中的異源部分的例示性 XTEN 序列包括 XTEN AE42-4(SEQ ID NO: 46，由 SEQ ID NO: 47 編碼)、XTEN AE144-2A(SEQ ID NO: 48，由 SEQ ID NO: 49 編碼)、XTEN AE144-3B（SEQ ID NO: 50，由 SEQ ID NO: 51 編碼）、XTEN AE144-4A（SEQ ID NO: 52，由 SEQ ID NO: 53 編碼）、XTEN AE144-5A（SEQ ID NO: 54，由 SEQ ID NO: 55 編碼）、XTEN AE144-6B（SEQ ID NO: 56，由 SEQ ID NO: 57 編碼）、XTEN AG144-1（SEQ ID NO: 58，由 SEQ ID NO: 59 編碼）、XTEN AG144-A（SEQ ID NO: 60，由 SEQ ID NO: 61 編碼）、XTEN AG144-B（SEQ ID NO: 62，由 SEQ ID NO: 63 編碼）、XTEN AG144-C（SEQ ID NO: 64，由 SEQ ID NO: 65 編碼）和 XTEN AG144-F（SEQ ID NO: 66，由 SEQ ID NO: 67 編碼）。在一個特定實施例中，XTEN 由 SEQ ID NO: 18 編碼。

【0398】 在另一個實施例中，XTEN 序列選自 AE36（SEQ ID NO: 130）、AE42（SEQ ID NO: 131）、AE72（SEQ ID NO: 132）、AE78（SEQ ID NO: 133）、AE144（SEQ ID NO: 134）、AE144_2A（SEQ ID NO: 48）、AE144_3B（SEQ ID NO: 50）、AE144_4A（SEQ ID NO: 52）、AE144_5A（SEQ ID NO: 54）、AE144_6B（SEQ ID NO: 135）、AG144（SEQ ID NO: 136）、AG144_A（SEQ ID NO: 137）、AG144_B（SEQ ID NO: 62）、AG144_C（SEQ ID NO: 64）、AG144_F（SEQ ID NO: 66）、AE288（SEQ ID NO: 138）、AE288_2（SEQ ID NO: 139）、AG288（SEQ ID NO: 140）、AE576（SEQ ID NO: 141）、AG576（SEQ ID NO: 142）、AE864（SEQ ID NO: 143）、AG864（SEQ ID NO: 144）、XTEN_AE72_2A_1（SEQ ID NO: 145）、XTEN_AE72_2A_2（SEQ ID NO: 146）、XTEN_AE72_3B_1（SEQ ID NO: 147）、XTEN_AE72_3B_2（SEQ ID NO: 148）、XTEN_AE72_4A_2（SEQ ID NO: 149）、XTEN_AE72_5A_2（SEQ ID NO: 150）、XTEN_AE72_6B_1（SEQ ID NO: 151）、XTEN_AE72_6B_2（SEQ ID NO: 152）、XTEN_AE72_1A_1（SEQ ID NO: 153）、XTEN_AE72_1A_2（SEQ ID NO: 154）、XTEN_AE144_1A（SEQ ID NO: 155）、AE150（SEQ ID NO: 156）、

AG150（SEQ ID NO: 157）、AE294（SEQ ID NO: 158）、AG294（SEQ ID NO: 159）及其任何組合。在一個具體實施例中，XTEN序列選自AE72、AE144和AE288。本發明的某些XTEN序列的胺基酸序列顯示於表8中。

表8. XTEN序列

XTEN	**胺基酸序列**
AE42-4 （SEQ ID NO: 46）	GAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPASS
AE144-2A （SEQ ID NO: 48）	TSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSE SATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPS EGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPG
A144-3B （SEQ ID NO: 50）	SPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTST EPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPS EGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG
AE144-4A （SEQ ID NO: 52）	TSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGS PTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPG
AE144-5A （SEQ ID NO: 54）	TSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATS GSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEG
AE144-6B （SEQ ID NO: 56）	TSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEP ATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPG
AG144-1	PGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPG

（SEQ ID NO: 58）	SSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTP GSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSS
AG144-A （SEQ ID NO: 60）	GASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGS STPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSST PSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSP
AG144-B （SEQ ID NO: 62）	GTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGS STPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSST PSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSP
AG144-C （SEQ ID NO: 64）	GTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGS SPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASP GTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSP
XTEN AG144-F （SEQ ID NO: 66）	GSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGS STPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPG SGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSP
AE36 （SEQ ID NO: 130）	GSPAGSPTSTEEGTSESATPESGPGSEPATSGSETP
AE42 （SEQ ID NO: 131）	GAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPASS
AE72 （SEQ ID NO: 132）	GAPTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETP GTSESATPESGPGTSTEPSEGSAPGASS
AE78 （SEQ ID NO: 133）	GAPTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETP GTSESATPESGPGTSTEPSEGSAPGASS

AE144 （SEQ ID NO: 134）	GSEPATSGSETPGTSESATPESGPGSEPATSGSETPGSPAGSPTSTEEGT STEPSEGSAPGSEPATSGSETPGSEPATSGSETPGSEPATSGSETPGTSTE PSEGSAPGTSESAPESGPGSEPATSGSETPGTSTEPSEGSAP
AE144_6B （SEQ ID NO: 135）	TSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEP ATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPG
AG144 （SEQ ID NO:136）	GTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGA SPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASP GTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSP
AG144_A （SEQ ID NO: 137）	GASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGS STPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSST PSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSP
AE288 （SEQ ID NO: 138）	GTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGT SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPA TSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSE GSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSE TPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAP GSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP
AE288_2 （SEQ ID NO: 139）	GSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGT STEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTE PSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATP ESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPES

	GPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEE GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAP
AG288 （SEQ ID NO: 140）	PGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPG TPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSS TPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGS GTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTS STGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSST GSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGS
AE576 （SEQ ID NO: 141）	GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT STEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPA TSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSE GSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPES GPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAP GTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGT SESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTE PSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSE GSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP GTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGT SESATPESGPGTSTEPSEGSAP
AG576 （SEQ ID NO: 142）	PGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPG SSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGAS PGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPG

	TSSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTS STGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSST GSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGS PGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPG ASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTP GSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGS GTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSAS TGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTG TGPGSSPSASTGTGPGASPGTSSTGS
AE864 （SEQ ID NO: 143）	GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT STEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPA TSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSE GSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPES GPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAP GTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGT SESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTE PSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSE GSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP GTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGT SESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSP TSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTST

	EEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGP GTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGT STEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTE PSEGSAP
AG864 （SEQ ID NO: 144）	GASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGS STPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSST PSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSG TASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGA TGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSGATG SPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSP GASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSST PSGATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPS GATGSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSS TGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTG SPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSP GTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGS STPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPG SGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGT SSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSS TGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATG SPGASPGTSSTGSP
XTEN_AE72_2A_1	TSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSE

（SEQ ID NO: 145）	SATPESGPGTSTEPSEGSAPG
XTEN_AE72_2A_2 （SEQ ID NO: 146）	TSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSE SATPESGPGTSESATPESGPG
XTEN_AE72_3B_1 （SEQ ID NO:147）	SPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTST EPSEGSAPGTSTEPSEGSAPG
XTEN_AE72_3B_2 （SEQ ID NO: 148）	TSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPG
XTEN_AE72_4A_2 （SEQ ID NO: 149）	TSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSE SATPESGPGTSTEPSEGSAPG
XTEN_AE72_5A_2 （SEQ ID NO: 150）	SPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPA GSPTSTEEGSPAGSPTSTEEG
XTEN_AE72_6B_1 （SEQ ID NO: 151）	TSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEP ATSGSETPGSEPATSGSETPG
XTEN_AE72_6B_2 （SEQ ID NO: 152）	SPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSE SATPESGPGTSTEPSEGSAPG
XTEN_AE72_1A_1 （SEQ ID NO: 153）	SPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTST EPSEGSAPGTSTEPSEGSAPG
XTEN_AE72_1A_2 （SEQ ID NO: 154）	TSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSE SATPESGPGTSTEPSEGSAPG
XTEN_AE144_1A （SEQ ID NO: 155）	SPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTST EPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATS

	GSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPG
AE150 （SEQ ID NO: 156）	GAPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGSPAGSPTSTE EGTSTEPSEGSAPGSEPATSGSETPGSEPATSGSETPGSEPATSGSETPG TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPASS
G150 （SEQ ID NO: 157）	GAPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTG PGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPG ASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPASS
AE294 （SEQ ID NO: 158）	GAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSET PGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGS EPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTE PSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATS GSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGS APGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPASS
AG294 （SEQ ID NO: 159）	GAPPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATG SPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSP GSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGT PGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASP GTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGT SSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSASS

【0399】 在一些實施例中，XTEN的小於100%的胺基酸選自甘胺酸（G）、丙胺酸（A）、絲胺酸（S）、蘇胺酸（T）、麩胺酸（E）和脯胺

酸(P),或者小於100%的序列由來自表7的序列基序或本文所提供的XTEN序列組成。在此類實施例中，XTEN的其餘胺基酸殘基選自其他14種天然L-胺基酸中的任一種，但是可以優先地選自親水胺基酸，使得XTEN序列含有至少約90%、91%、92%、93%、94%、95%、96%、97%、98%或至少約99%的親水胺基酸。綴合建構體中所利用的XTEN中疏水胺基酸的含量可以小於5%、或小於2%、或小於1%疏水胺基酸含量。在XTEN構建中較不有利的疏水殘基包括色胺酸、苯丙胺酸、酪胺酸、白胺酸、異白胺酸、擷胺酸和甲硫胺酸。另外，XTEN序列可以含有小於5%或小於4%或小於3%或小於2%或小於1%或無以下胺基酸：甲硫胺酸（例如，以避免氧化）、或天門冬醯胺酸和麩醯胺酸（以避免脫醯胺）。

【0400】 所述一個或多個XTEN序列可以插入治療性蛋白質的C末端或N末端處，或者插入治療性蛋白質的胺基酸序列中的兩個胺基酸之間。例如，如果治療性蛋白質包含FVIII多肽，則XTEN可以在選自表5的一個或多個插入位點插入兩個胺基酸之間。如果治療性蛋白質包含FIX多肽，則XTEN可以在選自表5的一個或多個插入位點插入兩個胺基酸之間。

【0401】 可以根據本發明使用的XTEN序列的其他例子披露於以下文獻中：美國專利公開號2010/0239554 A1、2010/0323956 A1、2011/0046060 A1、2011/0046061 A1、2011/0077199 A1或2011/0172146 A1，或者國際專利公開號WO 2010091122 A1、WO 2010144502 A2、WO 2010144508 A1、WO 2011028228 A1、WO 2011028229 A1、WO 2011028344 A2、WO 2014/011819 A2或WO 2015/023891。

5. 白蛋白或其片段、衍生物或變異體

【0402】 在一些實施例中，異源部分包含白蛋白或其功能片段。人血清白蛋白（HSA或HA）是其全長形式的609個胺基酸的蛋白質，負責血清滲透壓的大部分，並且還用作內源和外源配位體的載劑。如本文所用的術語“白蛋白”包括全長白蛋白或其功能片段、變異體、衍生物或類似物。白蛋白或其片段或變異體的例子披露於以下文獻中：美國專利公開號2008/0194481A1、2008/0004206 A1、2008/0161243 A1、2008/0261877 A1

或 2008/0153751 A1，或者 PCT 申請公開號 2008/033413 A2、2009/058322 A1 或 2007/021494 A2，將上述文獻通過引用以其整體併入本文。

【0403】 在一個實施例中，本公開文本的治療性蛋白質包含進一步連接至第二異源部分的白蛋白、其片段或變異體，所述第二異源部分選自免疫球蛋白恒定區或其部分（例如，Fc 區）、PAS 序列、HES、PEG 及其任何組合。

6. 白蛋白結合部分

【0404】 在某些實施例中，異源部分是白蛋白結合部分，其包含白蛋白結合肽、細菌白蛋白結合結構域、白蛋白結合抗體片段或其任何組合。

【0405】 例如，白蛋白結合蛋白可以是細菌白蛋白結合蛋白、抗體或抗體片段，包括結構域抗體（參見美國專利號 6,696,245）。例如，白蛋白結合蛋白可以是細菌白蛋白結合結構域，如一種鏈球菌蛋白 G（Konig, T. 和 Skerra, A. (1998) *J. Immunol. Methods* 218, 73-83）。可以用作綴合配偶體的白蛋白結合肽的其他例子是例如具有 Cys-Xaa_1-Xaa_2-Xaa_3-Xaa_4-Cys 共有序列的那些白蛋白結合肽，其中 Xaa_1 是 Asp、Asn、Ser、Thr 或 Trp；Xaa_2 是 Asn、Gln、His、Ile、Leu 或 Lys；Xaa_3 是 Ala、Asp、Phe、Trp 或 Tyr；並且 Xaa_4 是 Asp、Gly、Leu、Phe、Ser 或 Thr，如美國專利申請 2003/0069395 或 Dennis 等人（Dennis 等人 (2002) *J. Biol. Chem.* 277, 35035-35043）中所述。

【0406】 如 Kraulis 等人, FEBS Lett. 378:190-194 (1996)和 Linhult 等人, Protein Sci. 11:206-213 (2002)所披露的來自鏈球菌蛋白 G 的結構域 3 是細菌白蛋白結合結構域的例子。白蛋白結合肽的例子包括具有核心序列 DICLPRWGCLW（SEQ ID NO: 35）的一系列肽。參見例如，Dennis 等人, J. Biol. Chem. 2002, 277: 35035-35043 (2002)。白蛋白結合抗體片段的例子披露於以下文獻中：Muller 和 Kontermann, Curr. Opin. Mol. Ther. 9:319-326 (2007)；Roovers 等人, Cancer Immunol. Immunother. 56:303-317 (2007)；和 Holt 等人, Prot. Eng. Design Sci., 21:283-288 (2008)，將上述文獻通過引用以其整體併入本文。這種白蛋白結合部分的例子是如 Trussel 等人, Bioconjugate Chem. 20:2286-2292 (2009)所披露的 2-(3-馬來醯亞胺基丙醯胺基)-6-(4-(4-碘苯基)丁醯胺基)己酸酯（「Albu」標簽）。

【0407】 脂肪酸、特別是長鏈脂肪酸（LCFA）和長鏈脂肪酸樣白蛋白結合化合物可以用於延長本公開文本的凝血因子蛋白的體內半衰期。LCFA樣白蛋白結合化合物的例子是16-(l-(3-(9-(((2,5-二氧代吡咯啶-l-基氧基)羰氧基)-甲基)-7-磺基-9H-芴-2-基胺基)-3-氧代丙基)-2,5-二氧代吡咯啶-3-基硫基)十六烷酸（參見例如，WO 2010/140148）。

7. PAS序列

【0408】 在其他實施例中，異源部分是PAS序列。如本文所用的PAS序列意指主要包含丙胺酸和絲胺酸殘基或主要包含丙胺酸、絲胺酸和脯胺酸殘基的胺基酸序列，所述胺基酸序列在生理條件下形成隨機螺旋構象。因此，PAS序列是包含丙胺酸、絲胺酸和脯胺酸、基本上由所述胺基酸組成或由所述胺基酸組成的結構單元、胺基酸聚合物或序列盒，其可以用作嵌合蛋白中異源部分的一部分。然而，熟習此項技術者知道，在添加除了丙胺酸、絲胺酸和脯胺酸以外的殘基作為PAS序列中的次要成分時，胺基酸聚合物還可以形成隨機螺旋構象。如本文所用的術語“次要成分”意指，可以在PAS序列中添加除了丙胺酸、絲胺酸和脯胺酸以外的胺基酸至某種程度，例如，高達約12%，即PAS序列的100個胺基酸中的約12個；高達約10%，即PAS序列的100個胺基酸中的約10個；高達約9%，即100個胺基酸中的約9個；高達約8%，即100個胺基酸中的約8個；約6%，即100個胺基酸中的約6個；約5%，即100個胺基酸中的約5個；約4%，即100個胺基酸中的約4個；約3%，即100個胺基酸中的約3個；約2%，即100個胺基酸中的約2個；約1%，即100個胺基酸中的約1個。與丙胺酸、絲胺酸和脯胺酸不同的胺基酸可以選自Arg、Asn、Asp、Cys、Gln、Glu、Gly、His、Ile、Leu、Lys、Met、Phe、Thr、Trp、Tyr和Val。

【0409】 在生理條件下，PAS序列段形成隨機螺旋構象，由此可以介導凝血因子蛋白的增加的體內和/或體外穩定性。由於隨機螺旋結構域自身不具有穩定結構或功能，基本上保留了由凝血因子蛋白介導的生物學活性。在其他實施例中，尤其關於血漿中的蛋白質水解、免疫原性、等電點/靜電行為、與細胞表面受體的結合或內化，形成隨機螺旋結構域的PAS序

列具有生物惰性，但仍然是生物可降解的，這提供了優於諸如 PEG 等合成聚合物的明顯優勢。

【0410】 形成隨機螺旋構象的 PAS 序列的非限制性例子包含選自由下列組成之群組的胺基酸序列：ASPAAPAPASPAAPAPSAPA（SEQ ID NO: 36）、AAPASPAPAAPSAPAPAAPS（SEQ ID NO: 37）、APSSPSPSAPSSPSPASPSS（SEQ ID NO: 38）、APSSPSPSAPSSPSPASPS（SEQ ID NO: 39）、SSPSAPSPSSPASPSPSSPA（SEQ ID NO: 40）、AASPAAPSAPPAAASPAAPSAPPA（SEQ ID NO: 41）、ASAAAPAAASAAASAPSAAA（SEQ ID NO: 42）及其任何組合。PAS 序列的其他例子是從例如以下文獻得知：美國專利公開號 2010/0292130 A1 和 PCT 申請公開號 WO 2008/155134 A1。

8. HAP序列

【0411】 在某些實施例中，異源部分是富含甘胺酸的同胺基酸聚合物（HAP）。HAP 序列可以包含甘胺酸的重複序列，所述重複序列的長度為至少 50 個胺基酸、至少 100 個胺基酸、120 個胺基酸、140 個胺基酸、160 個胺基酸、180 個胺基酸、200 個胺基酸、250 個胺基酸、300 個胺基酸、350 個胺基酸、400 個胺基酸、450 個胺基酸或 500 個胺基酸。在一個實施例中，HAP 序列能夠延長融合至或連接至 HAP 序列的部分的半衰期。HAP 序列的非限制性例子包括但不限於$(Gly)_n$、$(Gly_4Ser)_n$或 $S(Gly_4Ser)_n$，其中 n 是 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19 或 20。在一個實施例中，n 是 20、21、22、23、24、25、26、26、28、29、30、31、32、33、34、35、36、37、38、39 或 40。在另一個實施例中，n 是 50、60、70、80、90、100、110、120、130、140、150、160、170、180、190 或 200。

9. 轉鐵蛋白或其片段

【0412】 在某些實施例中，異源部分是轉鐵蛋白或其片段。任何轉鐵蛋白都可以用於製備本公開文本的凝血因子蛋白。作為例子，野生型人 TF（TF）是 679 個胺基酸的蛋白質，大約 75 KDa（不計糖基化），具有似乎源自基因重複的兩個主要結構域 N（約 330 個胺基酸）和 C（約 340 個胺基

酸）。參見 GenBank 登錄號 NM001063、XM002793、M12530、XM039845、XM 039847 和 S95936（www.ncbi.nlm.nih.gov/），將所有登錄號通過引用以其整體併入本文。轉鐵蛋白包含兩個結構域 N 結構域和 C 結構域。N 結構域包含兩個子結構域 N1 結構域和 N2 結構域，並且 C 結構域包含兩個子結構域 C1 結構域和 C2 結構域。

【0413】 在一個實施例中，轉鐵蛋白異源部分包括轉鐵蛋白剪接變異體。在一個例子中，轉鐵蛋白剪接變異體可以是人轉鐵蛋白的剪接變異體，例如 Genbank 登錄號 AAA61140。在另一個實施例中，嵌合蛋白的轉鐵蛋白部分包括轉鐵蛋白序列的一個或多個結構域，例如 N 結構域、C 結構域、N1 結構域、N2 結構域、C1 結構域、C2 結構域或其任何組合。

10. 清除受體

【0414】 在某些實施例中，異源部分是清除受體、其片段、變異體或衍生物。LRP1 是 600 kDa 膜內在蛋白，其參與受體介導的諸如 X 因子等多種蛋白質的清除。參見例如，Narita 等人, Blood 91:555-560 (1998)。

11. 血管性血友病因子或其片段

【0415】 在某些實施例中，異源部分是血管性血友病因子（VWF）或其一個或多個片段。

【0416】 VWF（也稱為 F8VWF）是存在於血漿中的大多聚糖蛋白，並且在內皮（在懷布爾-帕拉德體（Weibel-Palade body）中）、巨核細胞（血小板的 α 顆粒）和內皮下結締組織中組成性地產生。基本 VWF 單體是 2813 個胺基酸的蛋白質。每個單體含有多個具有特定功能的特定結構域，即 D' 和 D3 結構域（其一起結合至 VIII 因子）、A1 結構域（其結合至血小板 GPIb-受體、肝素和/或可能地膠原）、A3 結構域（其結合至膠原）、C1 結構域（其中 RGD 結構域在被啟動時結合至血小板整聯蛋白 αIIbβ3）和在蛋白質 C 末端的「半胱胺酸結」結構域（其是 VWF 與血小板衍生的生長因子（PDGF）、轉化生長因子-β（TGFβ）和 β-人類絨毛膜促性腺激素（βHCG）共用的）。

【0417】 人類 VWF 的 2813 單體胺基酸序列是作為 Genbank 中的登錄號 NP000543.2 來報道。編碼人類 VWF 的核苷酸序列是作為 Genbank 中

的登錄號 NM000552.3 來報道。SEQ ID NO: 129 是由 SEQ ID NO: 128 編碼的胺基酸序列。D'結構域包括 SEQ ID NO: 129 的胺基酸 764 至 866。D3 結構域包括 SEQ ID NO: 44 的胺基酸 867 至 1240。

【0418】 在血漿中，95%-98%的 FVIII 以與全長 VWF 的緊密非共價複合物來循環。這種複合物的形成對於在體內維持適當的 FVIIII 血漿水準是重要的。Lenting 等人, *Blood.* 92(11): 3983-96 (1998)；Lenting 等人, *J. Thromb. Haemost.* 5(7): 1353-60 (2007)。在 FVIII 由於在重鏈中的位置 372 和 740 以及輕鏈中的位置 1689 的蛋白質水解而被啟動時，從啟動的 FVIII 去除結合至 FVIII 的 VWF。

【0419】 在某些實施例中，異源部分是全長血管性血友病因子。在其他實施例中，異源部分是血管性血友病因子片段。如本文所用，本文所用的術語「VWF 片段（VWF fragment）」或「VWF 片段（VWF fragments）」意指任何 VWF 片段，其與 FVIII 相互作用並保留至少一種或多種通常由全長 VWF 提供給 FVIII 的特性，所述特性是例如防止過早啟動為 FVIIIa、防止過早蛋白質水解、防止可導致過早清除的與磷脂膜的締合、防止與可以結合裸 FVIII 而不是 VWF 結合 FVIII 的 FVIII 清除受體結合和/或穩定 FVIII 重鏈和輕鏈相互作用。在一個具體實施例中，異源部分是包含 VWF 的 D' 結構域和 D3 結構域的（VWF）片段。包含 D'結構域和 D3 結構域的 VWF 片段還可以包含選自由下列組成之群組的 VWF 結構域：A1 結構域、A2 結構域、A3 結構域、D1 結構域、D2 結構域、D4 結構域、B1 結構域、B2 結構域、B3 結構域、C1 結構域、C2 結構域、CK 結構域、其一個或多個片段及其任何組合。融合至 VWF 片段的具有 FVIII 活性的多肽的其他例子披露於以下文獻中：2012 年 7 月 3 日提交的美國臨時專利申請號 61/667,901 和美國公開號 2015/0023959 A1，將這兩篇文獻都通過引用以其整體併入本文。

12. 連接子部分

【0420】 在某些實施例中，異源部分是肽連接子。

【0421】 如本文所用，術語「肽連接子」或「連接子部分」是指連接多肽鏈的線性胺基酸序列中的兩個結構域的肽或多肽序列（例如，合成肽或多肽序列）。

【0422】 在一些實施例中，肽連接子可以插入本公開文本的治療性蛋白質與上述異源部分（如白蛋白）之間。肽連接子可以向嵌合多肽分子提供柔性。連接子通常不被切割，但是可能需要這種切割。在一個實施例中，在加工期間不去除這些連接子。

【0423】 可以存在于本公開文本的嵌合蛋白中的一種類型的連接子是蛋白酶可切割的連接子，其包含切割位點（即，蛋白酶切割位點底物，例如 XI 因子 a、Xa 或凝血酶切割位點）並且可以在切割位點的 N 末端或 C 末端或兩側上包括其他連接子。這些可切割連接子在併入本公開文本的建構體中時產生具有異源切割位點的嵌合分子。

【0424】 在一個實施例中，由本公開文本的核酸分子編碼的治療性蛋白質包含兩個或更多個 Fc 結構域或部分，其通過 cscFc 連接子連接以形成包含於單條多肽鏈中的 Fc 區。cscFc 連接子側接有至少一個細胞內加工位點，即由細胞內酶切割的位點。多肽在所述至少一個細胞內加工位點的切割產生包含至少兩條多肽鏈的多肽。

【0425】 其他肽連接子可以任選地用於本公開文本的建構體中，例如以將凝血因子蛋白連接至 Fc 區。可以結合本公開文本使用的一些例示性連接子包括例如更詳細地描述於下文中的包含 GlySer 胺基酸的多肽。

【0426】 在一個實施例中，肽連接子是合成的，即非天然存在的。在一個實施例中，肽連接子包括包含如下胺基酸序列的肽（或多肽）（其可以是或可以不是天然存在的），所述胺基酸序列將第一線性胺基酸序列與在自然界中並非與其天然地連接或遺傳融合的第二線性胺基酸序列連接或遺傳融合。例如，在一個實施例中，肽連接子可以包含非天然存在的多肽，其是天然存在的多肽的修飾形式（例如，包含突變，如添加、取代或缺失）。在另一個實施例中，肽連接子可以包含非天然存在的胺基酸。在另一個實施例中，肽連接子可以包含存在於在自然界中不存在的線性序列中的天然存在的胺基酸。在仍另一個實施例中，肽連接子可以包含天然存在的多肽序列。

【0427】 例如，在某些實施例中，肽連接子可以用於融合相同的 Fc 部分，由此形成同二聚 scFc 區。在其他實施例中，肽連接子可以用於融合

不同的 Fc 部分（例如野生型 Fc 部分和 Fc 部分變異體），由此形成異二聚 scFc 區。

【0428】 在另一個實施例中，肽連接子包含 gly-ser 連接子或由所述 gly-ser 連接子組成。在一個實施例中，scFc 或 cscFc 連接子包含免疫球蛋白鉸鏈的至少一部分和 gly-ser 連接子。如本文所用，術語「gly-ser 連接子」是指由甘胺酸和絲胺酸殘基組成的肽。在某些實施例中，所述 gly-ser 連接子可以插入肽連接子的兩個其他序列之間。在其他實施例中，gly-ser 連接子連接於肽連接子的另一序列的一個或兩個末端。在又其他實施例中，兩個或更多個 gly-ser 連接子串聯併入肽連接子中。在一個實施例中，本公開文本的肽連接子包含上鉸鏈區的至少一部分（例如，衍生自 IgG1、IgG2、IgG3 或 IgG4 分子）、中鉸鏈區的至少一部分（例如，衍生自 IgG1、IgG2、IgG3 或 IgG4 分子）和一系列 gly/ser 胺基酸殘基。

【0429】 本公開文本的肽連接子的長度為至少一個胺基酸並且可以具有變化的長度。在一個實施例中，本公開文本的肽連接子的長度為約 1 至約 50 個胺基酸。如在此上下文中所用，術語「約」表示+/-兩個胺基酸殘基。由於連接子長度必須是正整數，長度為約 1 至約 50 個胺基酸的長度意指長度為 1-3 至 48-52 個胺基酸的長度。在另一個實施例中，本公開文本的肽連接子的長度為約 10 至約 20 個胺基酸。在另一個實施例中，本公開文本的肽連接子的長度為約 15 至約 50 個胺基酸。在另一個實施例中，本公開文本的肽連接子的長度為約 20 至約 45 個胺基酸。在另一個實施例中，本公開文本的肽連接子的長度為約 15 至約 35 或約 20 至約 30 個胺基酸。在另一個實施例中，本公開文本的肽連接子的長度為約 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、40、50、60、70、80、90、100、500、1000 或 2000 個胺基酸。在一個實施例中，本公開文本的肽連接子的長度為 20 或 30 個胺基酸。

【0430】 在一些實施例中，肽連接子可以包含至少 2、至少 3、至少 4、至少 5、至少 10、至少 20、至少 30、至少 40、至少 50、至少 60、至少 70、至少 80、至少 90 或至少 100 個胺基酸。在其他實施例中，肽連接子可

以包含至少 200、至少 300、至少 400、至少 500、至少 600、至少 700、至少 800、至少 900 或至少 1,000 個胺基酸。在一些實施例中，肽連接子可以包含至少約 10、20、30、40、50、60、70、80、90、100、150、200、300、400、500、600、700、800、900、1000、1100、1200、1300、1400、1500、1600、1700、1800、1900 或 2000 個胺基酸。肽連接子可以包含 1-5 個胺基酸、1-10 個胺基酸、1-20 個胺基酸、10-50 個胺基酸、50-100 個胺基酸、100-200 個胺基酸、200-300 個胺基酸、300-400 個胺基酸、400-500 個胺基酸、500-600 個胺基酸、600-700 個胺基酸、700-800 個胺基酸、800-900 個胺基酸或 900-1000 個胺基酸。

【0431】 可以使用業內已知的技術將肽連接子引入多肽序列中。可以通過 DNA 序列分析來確認修飾。可以使用質體 DNA 轉化宿主細胞用於穩定產生所產生的多肽。

13. 單體-二聚體雜合物

【0432】 在一些實施例中，本公開文本的治療性蛋白質包含含有凝血因子的單體-二聚體雜合物分子。

【0433】 本文所用的術語「單體-二聚體雜合物」是指包含通過二硫鍵相互締合的第一多肽鏈和第二多肽鏈的嵌合蛋白，其中第一鏈包含凝血因子（例如，FVIII）和第一 Fc 區，並且第二鏈包含無凝血因子的第二 Fc 區，基本上由所述第二 Fc 區組成或由所述第二 Fc 區組成。因此，單體-二聚體雜合建構體是包含僅具有一種凝血因子的單體方面和具有兩個 Fc 區的二聚體方面的雜合物。

14. 表現控制序列

【0434】 在一些實施例中，本公開文本的核酸分子還包含至少一個表現控制序列。如本文所用的表現控制序列是有助於與其可操作地連接的編碼核酸的有效轉錄和轉譯的任何調控核苷酸序列，如啟動子序列或啟動子-強化子組合。例如，本公開文本的核酸分子可以可操作地連接至至少一個轉錄控制序列。基因表現控制序列可以是例如哺乳動物或病毒啟動子，如組成型或誘導型啟動子。

【0435】 組成型哺乳動物啟動子包括但不限於以下基因的啟動子：次黃嘌呤磷酸核糖基轉移酶（HPRT）、腺苷脫胺酶、丙酮酸激酶、β-肌動蛋白啟動子和其他組成型啟動子。在真核細胞中組成性地發揮作用的例示性病毒啟動子包括例如來自以下病毒的啟動子：巨細胞病毒（CMV）、猿猴病毒（例如，SV40）、乳頭狀瘤病毒、腺病毒、人免疫缺陷病毒（HIV）、勞斯肉瘤病毒、巨細胞病毒、莫洛尼白血病病毒的長末端重複（LTR）和其他逆轉錄病毒、以及單純皰疹病毒的胸苷激酶啟動子。其他組成型啟動子是業內熟習此項技術者已知的。可用作本公開文本的基因表現序列的啟動子還包括誘導型啟動子。誘導型啟動子是在誘導劑存在下表現。例如，在某些金屬離子存在下誘導金屬硫蛋白啟動子以促進轉錄和轉譯。其他誘導型啟動子是業內熟習此項技術者已知的。

【0436】 在一個實施例中，本公開文本包括轉基因在組織特異性啟動子和/或強化子控制下的表現。在另一個實施例中，啟動子或其他表現控制序列選擇性地增強轉基因在肝細胞中的表現。在某些實施例中，啟動子或其他表現控制序列選擇性地增強轉基因在肝細胞、竇狀細胞和/或內皮細胞中的表現。在一個特定實施例中，啟動子或其他表現控制序列選擇性地增強轉基因在內皮細胞中的表現。在某些實施例中，啟動子或其他表現控制序列選擇性地增強轉基因在以下中的表現：肌肉細胞、中樞神經系統、眼睛、肝臟、心臟或其任何組合。肝臟特異性啟動子的例子包括但不限於小鼠甲狀腺素啟動子（mTTR）、內源人類 VIII 因子啟動子（F8）、人類 α-1-抗胰蛋白酶啟動子（hAAT）、人類白蛋白最小啟動子和小鼠白蛋白啟動子。在一個特定實施例中，啟動子包含 mTTR 啟動子。mTTR 啟動子描述於 R. H. Costa 等人, 1986, Mol. Cell. Biol. 6:4697 中。F8 啟動子描述於 Figueiredo 和 Brownlee, 1995, J. Biol. Chem. 270:11828-11838 中。在一些實施例中，啟動子選自肝臟特異性啟動子（例如，α1-抗胰蛋白酶（AAT））、肌肉特異性啟動子（例如，肌肉肌酸激酶（MCK）、肌球蛋白重鏈 α（αMHC）、肌紅蛋白（MB）和結蛋白（DES））、合成啟動子（例如，SPc5-12、2R5Sc5-12、dMCK 和 tMCK）及其任何組合。

【0437】 在一個實施例中，啟動子選自小鼠甲狀腺素啟動子（mTTR）、內源人類 VIII 因子啟動子（F8）、人類α-1-抗胰蛋白酶啟動子（hAAT）、人類白蛋白最小啟動子、小鼠白蛋白啟動子、TTPp、CASI 啟動子、CAG 啟動子、巨細胞病毒（CMV）啟動子、α1-抗胰蛋白酶（AAT）、肌肉肌酸激酶（MCK）、肌球蛋白重鏈 α（αMHC）、肌紅蛋白（MB）、結蛋白（DES）、SPc5-12、2R5Sc5-12、dMCK 和 tMCK、磷酸甘油酸激酶（PGK）啟動子及其任何組合。

【0438】 可以使用一種或多種強化子進一步增強表現水準以實現治療功效。一種或多種強化子可以單獨提供，或與一種或多種啟動子元件一起提供。通常，表現控制序列包含多個強化子元件和組織特異性啟動子。在一個實施例中，強化子包含一個或多個拷貝的 α-1-微球蛋白/比庫寧蛋白（bikunin）強化子（Rouet 等人, 1992, J. Biol. Chem. 267:20765-20773；Rouet 等人, 1995, Nucleic Acids Res. 23:395-404；Rouet 等人, 1998, Biochem. J. 334:577-584；Ill 等人, 1997, Blood Coagulation Fibrinolysis 8:S23-S30）。在另一個實施例中，強化子衍生自肝臟特異性轉錄因子結合位點，如 EBP、DBP、HNF1、HNF3、HNF4、HNF6 和 Enh1，包含 HNF1、（有義）-HNF3、（有義）-HNF4、（反義）-HNF1、（反義）-HNF6、（有義）-EBP、（反義）-HNF4（反義）。

【0439】 在一個特定例子中，可用於本公開文本的啟動子包含 SEQ ID NO: 69（即，ET 啟動子），其還作為 GenBank 編號 AY661265 而已知。還參見 Vigna 等人, *Molecular Therapy 11(5)*:763 (2005)。其他適宜載體和基因調控元件的例子描述於以下文獻中：WO 02/092134、EP1395293 或美國專利號 6,808,905、7,745,179 或 7,179,903，將上述文獻通過引用以其整體併入本文。

【0440】 在一個實施例中，本公開文本的核酸分子還包含內含子序列。在一些實施例中，內含子序列定位於編碼 FVIII 多肽的核酸序列的 5'。在一些實施例中，內含子序列是天然存在的內含子序列。在一些實施例中，內含子序列是合成序列。在一些實施例中，內含子序列衍生自天然存在的

內含子序列。在某些實施例中，內含子序列包含 SV40 小 T 內含子。在一個實施例中，內含子序列包含 SEQ ID NO: 115。

【0441】 在一些實施例中，核酸分子還包含轉錄後調控元件。在某些實施例中，轉錄後調控元件包含突變的土撥鼠肝炎病毒轉錄後調控元件（WPRE）。在一個特定實施例中，轉錄後調控元件包含 SEQ ID NO: 120。

【0442】 在一些實施例中，核酸分子包含微小 RNA（miRNA）結合位點。在一個實施例中，miRNA 結合位點是 miR-142-3p 的 miRNA 結合位點。在其他實施例中，miRNA 結合位點選自 Rennie 等人, *RNA Biol.* *13(6)*:554-560 (2016)所披露的 miRNA 結合位點以及可在 http://sfold.wadsworth.org/starmirDB.php 獲得的 STarMirDB，將上述文獻通過引用以其整體併入本文。

【0443】 在一些實施例中，核酸分子包含一個或多個 DNA 核靶向序列（DTS）。DTS 促進含有此類序列的 DNA 分子易位至核中。在某些實施例中，DTS 包含 SV40 強化子序列。在某些實施例中，DTS 包含 c-Myc 強化子序列。在一些實施例中，DTS 介於第一 ITR 與第二 ITR 之間。在一些實施例中，DTS 在第一 ITR 的 3'和治療性蛋白質的 5'。在其他實施例中，DTS 在治療性蛋白質的 3'和第二 ITR 的 5'。

【0444】 在一些實施例中，核酸分子還包含 3'UTR 多聚(A)尾序列。在一個實施例中，3'UTR 多聚(A)尾序列包含 bGH 多聚(A)。在一個實施例中，3'UTR 多聚(A)尾包含肌動蛋白多聚(A)位點。在一個實施例中，3'UTR 多聚(A)尾包含血紅蛋白多聚(A)位點。

【0445】 在一個特定實施例中，3'UTR 多聚(A)尾序列包含 SEQ ID NO: 122。

III. 組織特異性表現

【0446】 在某些實施例中，有用的是在載體內包括一個或多個 miRNA 靶序列，其例如與凝血因子轉基因可操作地連接。因此，本公開文本還提供了至少一個與凝血因子核苷酸序列可操作地連接或以其他方式插入載體內的 miRNA 序列靶標。載體中包括的多於一個拷貝的 miRNA 靶序列可以增加系統的有效性。還包括不同 miRNA 靶序列。例如，表現多於一

種轉基因的載體可以具有在多於一個可以相同或不同的miRNA靶序列控制下的轉基因。miRNA靶序列可以是串聯的，但是也包括其他排列。含有miRNA靶序列的轉基因表現盒也能以反義定向插入載體內。反義定向可用於產生病毒顆粒，以避免原本可能對產生細胞有毒的基因產物的表現。在其他實施例中，載體包含1、2、3、4、5、6、7或8個拷貝的相同或不同的miRNA靶序列。然而，在某些其他實施例中，載體將不包括任何miRNA靶序列。是否要包括miRNA靶序列（以及數量）的選擇將通過諸如預期的組織靶標、所需的表現水準等已知參數來指導。

【0447】 在一個實施例中，靶序列是miR-223靶標，已經報道其可在骨髓定向祖細胞中以及至少部分地在更原始的HSPC中最有效地阻斷表現。miR-223靶標可以在分化的髓樣細胞（包括粒細胞、單核細胞、巨噬細胞、髓樣樹突細胞）中阻斷表現。miR-223靶標還可以適於依賴於淋巴或紅細胞譜系中的穩健轉基因表現的基因療法應用。miR-223靶標還可以在人HSC中非常有效地阻斷表現。

【0448】 在另一個實施例中，靶序列是miR142靶標（tccataaagt aggaaacact aca（SEQ ID NO: 43））。在一個實施例中，載體包含4個拷貝的miR-142靶序列。在某些實施例中，將造血特異性微小RNA（如miR-142（142T））的互補序列併入載體（例如，慢病毒載體（LV））的3'非轉譯區中，使得編碼轉基因的轉錄物易發生miRNA介導的下調。通過這種方法，可以在造血譜系抗原呈遞細胞（APC）中防止轉基因表現，同時在非造血細胞中維持所述轉基因表現（Brown等人, Nat Med 2006）。這種策略可以對轉基因表現施加嚴格的轉錄後控制，並且由此使得能夠穩定遞送和長期表現轉基因。在一些實施例中，miR-142調控防止經轉導細胞的免疫介導的清除和/或誘導抗原特異性調控T細胞（Treg），並且介導對轉基因編碼抗原的穩健的免疫耐受。

【0449】 在一些實施例中，靶序列是miR181靶標。Chen C-Z和Lodish H, Seminars in Immunology (2005) 17(2):155-165披露了miR-181，它是在小鼠骨髓內的B細胞中特異性表現的miRNA（Chen和Lodish, 2005）。所述文獻還披露，一些人類miRNA與白血病相關。

【0450】 靶序列可以與 miRNA 完全或部分互補。術語“完全互補”意指，靶序列的核酸序列與識別所述靶序列的 miRNA 的序列 100%互補。術語“部分互補”意指，靶序列與識別所述靶序列的 miRNA 的序列僅部分互補，借此部分互補的序列仍然被 miRNA 識別。換句話說，在本公開文本的上下文中，部分互補的靶序列有效識別相應 miRNA，並且實現在表現所述 miRNA 的細胞中防止或減少轉基因表現。miRNA 靶序列的例子描述於以下文獻中：WO 2007/000668、WO 2004/094642、WO 2010/055413 或 WO 2010/125471，將上述文獻通過引用以其整體併入本文。

【0451】 在一些實施例中，轉基因表現靶向肝臟。在某些實施例中，轉基因表現靶向肝細胞。在其他實施例中，轉基因表現靶向內皮細胞。在一個特定實施例中，轉基因表現靶向天然表現內源 FVIII 的任何組織。

【0452】 在一些實施例中，轉基因表現靶向中樞神經系統。在某些實施例中，轉基因表現靶向神經元。在一些實施例中，轉基因表現靶向傳入神經元。在一些實施例中，轉基因表現靶向傳出神經元。在一些實施例中，轉基因表現靶向中間神經元。在一些實施例中，轉基因表現靶向神經膠質細胞。在一些實施例中，轉基因表現靶向星形膠質細胞。在一些實施例中，轉基因表現靶向少突膠質細胞。在一些實施例中，轉基因表現靶向小神經膠質細胞。在一些實施例中，轉基因表現靶向室管膜細胞。在一些實施例中，轉基因表現靶向許旺(Schwann)細胞。在一些實施例中，轉基因表現靶向衛星細胞。

【0453】 在一些實施例中，轉基因表現靶向肌肉組織。在一些實施例中，轉基因表現靶向平滑肌。在一些實施例中，轉基因表現靶向心肌。在一些實施例中，轉基因表現靶向骨骼肌。

【0454】 在一些實施例中，轉基因表現靶向眼睛。在一些實施例中，轉基因表現靶向感光細胞。在一些實施例中，轉基因表現靶向視網膜神經節細胞。

IV. 宿主細胞

【0455】 本公開文本還提供了包含本公開文本的核酸分子或載體的宿主細胞。如本文所用，術語「轉化」應在廣義上使用，是指將 DNA 引入受體宿主細胞中，這改變了基因型並因此導致受體細胞的變化。

【0456】 「宿主細胞」是指已經用使用重組 DNA 技術構建並且編碼至少一種異源基因的載體轉化的細胞。本公開文本的宿主細胞較佳地具有哺乳動物來源；最佳地具有人或小鼠來源。相信業內熟習此項技術者有能力優先地確定最適合於其目的的特定宿主細胞株。例示性宿主細胞株包括但不限於 CHO、DG44 和 DUXB11（中國倉鼠卵巢系，DHFR-）、HELA（人類子宮頸癌）、CVI（猴腎系）、COS（具有 SV40 T 抗原的 CVI 的衍生物）、R1610（中國倉鼠成纖維細胞）、BALBC/3T3（小鼠成纖維細胞）、HAK（倉鼠腎系）、SP2/O（小鼠骨髓瘤）、P3x63-Ag3.653（小鼠骨髓瘤）、BFA-1c1BPT（牛內皮細胞）、RAJI（人類淋巴細胞）、PER.C6®、NS0、CAP、BHK21 和 HEK 293（人腎）。在一個特定實施例中，宿主細胞選自：CHO 細胞、HEK293 細胞、BHK21 細胞、PER.C6®細胞、NS0 細胞、CAP 細胞及其任何組合。在一些實施例中，本公開文本的宿主細胞具有昆蟲來源。在一個特定實施例中，宿主細胞是 SF9 細胞。宿主細胞株通常可以從商業服務、美國組織培養物保藏中心（American Tissue Culture Collection）或從公開文獻獲得。

【0457】 將本公開文本的核酸分子或載體引入宿主細胞中可以通過業內熟習此項技術者熟知的多種技術來完成。這些技術包括但不限於轉染（包括電泳和電穿孔）、原生質體融合、磷酸鈣沈澱、與包膜 DNA 的細胞融合、顯微注射和完整病毒感染。參見 Ridgway, A. A. G. *"Mammalian Expression Vectors"* 第 24.2 章, 第 470-472 頁 Vectors, Rodriguez 和 Denhardt 編輯(Butterworths, Boston, Mass. 1988)。最佳地，通過電穿孔將質體引入主體中。使轉化的細胞在適於產生輕鏈和重鏈的條件下生長，並測定重鏈和/或輕鏈蛋白質合成。例示性測定技術包括酶聯免疫吸附測定（ELISA）、放射免疫測定（RIA）、或螢光啟動細胞分選儀分析（FACS）、免疫組織化學等。

【0458】 使包含本公開文本的分離的核酸分子或載體的宿主細胞在適當生長培養基中生長。如本文所用，術語「適當生長培養基」意指含有細胞生長所需營養素的培養基。細胞生長所需的營養素可以包括碳源、氮源、必需胺基酸、維生素、礦物質和生長因子。任選地，培養基可以含有一種或多種選擇因子。任選地，培養基可以含有小牛血清或胎牛血清（FCS）。在一個實施例中，培養基基本上不含 IgG。生長培養基通常將例如通過藥物選擇或必需營養素的缺乏來選擇含有 DNA 建構體的細胞，所述營養素通過 DNA 建構體上或與 DNA 建構體共轉染的可選標記來補充。所培養的哺乳動物細胞通常在市售的含血清或無血清培養基（例如，MEM、DMEM、DMEM/F12）中生長。在一個實施例中，培養基是 CDoptiCHO（Invitrogen，加利福尼亞州卡爾斯巴德）。在另一個實施例中，培養基是 CD17（Invitrogen，加利福尼亞州卡爾斯巴德）。適合於所用特定細胞株的培養基的選擇在業內熟習此項技術者的水準內。

V. 多肽的製備

【0459】 本公開文本還提供了由本公開文本的核酸分子編碼的多肽。在其他實施例中，本公開文本的多肽由包含本公開文本的核酸分子的載體編碼。在又其他實施例中，本公開文本的多肽由包含本公開文本的核酸分子的宿主細胞產生。

【0460】 在其他實施例中，本公開文本還提供了產生具有凝血因子（例如，FVIII）活性的多肽的方法，其包括在產生具有凝血因子（例如，FVIII）活性的多肽的條件下培養本公開文本的宿主細胞，以及回收具有凝血因子（例如，FVIII）活性的多肽。在一些實施例中，具有凝血因子（例如，FVIII）活性的多肽的表現相對於在相同條件下培養但包含參考核苷酸序列（例如，SEQ ID NO: 16，親代 FVIII 基因序列）的宿主細胞有所增加。

【0461】 在其他實施例中，本公開文本提供了增加具有凝血因子（例如，FVIII）活性的多肽的表現的方法，其包括在由核酸分子表現具有凝血因子（例如，FVIII）活性的多肽的條件下培養本公開文本的宿主細胞，其中具有凝血因子（例如，FVIII）活性的多肽的表現相對於在相同條件下培

養但包含參考核酸分子（例如，SEQ ID NO: 16，親代 FVIII 基因序列）的宿主細胞有所增加。

【0462】 在其他實施例中，本公開文本提供了改良具有凝血因子（例如，FVIII）活性的多肽的產量的方法，其包括在通過本文所公開的核酸分子產生具有凝血因子（例如，FVIII）活性的多肽的條件下培養宿主細胞，其中具有凝血因子（例如，FVIII）活性的多肽的產量相對於在相同條件下培養但包含參考核酸序列（例如，SEQ ID NO: 16，親代 FVIII 基因序列）的宿主細胞有所提高。

【0463】 本公開文本的治療性蛋白質（例如，凝血因子）可以在轉基因動物（如齧齒類動物、山羊、綿羊、豬或牛）體內合成。術語「轉基因動物」是指已經將外來基因併入基因組中的非人動物。因為這個基因是在種系組織中存在，其從親代傳遞給後代。將外源基因引入單細胞胚胎中（Brinster 等人 1985, Proc. Natl. Acad.Sci. USA 82:4438）。產生轉基因動物的方法是業內已知的，包括產生免疫球蛋白分子的轉基因學（Wagner 等人 1981, Proc. Natl. Acad. Sci. USA 78: 6376；McKnight 等人 1983, Cell 34 : 335；Brinster 等人 1983, Nature 306: 332；Ritchie 等人 1984, Nature 312: 517；Baldassarre 等人 2003, Theriogenology 59 : 831；Robl 等人 2003, Theriogenology 59: 107；Malassagne 等人 2003, Xenotransplantation 10 (3): 267）。

VII. 醫藥組合物

【0464】 含有本公開文本的核酸分子、核酸分子編碼的多肽、載體或宿主細胞的組合物可以含有適宜的醫藥上可接受的載劑。例如，所述組合物可以含有有助於將活性化合物加工成設計用於遞送至作用位點的製劑的賦形劑和/或助劑。

【0465】 在一個實施例中，本公開文本涉及醫藥組合物，其包含 (a) 本文所公開的核酸分子、載體、多肽或宿主細胞；和 (b) 醫藥上可接受的賦形劑。

【0466】 在一些實施例中，醫藥組合物還包含遞送劑。在某些實施例中，遞送劑包含脂質奈米顆粒（LNP）。在其他實施例中，醫藥組合物還包含脂質體、其他聚合分子和外來體。

【0467】 如本文所用，「脂質奈米顆粒」是指包含多個通過分子間力相互物理締合的脂質分子的奈米顆粒。脂質奈米顆粒可以是例如微球體（包括單層和多層囊泡，例如脂質體）、乳液中的分散相、微胞(micelles)或懸浮液中的內相。

【0468】 在一些實施例中，本公開文本提供了包封的核酸分子組合物，其可以包括包封本發明的核酸分子的脂質奈米顆粒主體。脂質奈米顆粒可以包含一種或多種脂質（例如，陽離子脂質、非陽離子脂質和 PEG 修飾的脂質）。在某些實施例中，本公開文本的脂質奈米顆粒配製為將本發明的一種或多種核酸分子遞送至一種或多種靶細胞。適宜脂質的例子包括但不限於磷脂醯基化合物（例如，磷脂醯基乙醇胺、鞘脂、磷脂醯基膽鹼、磷脂醯基絲胺酸、磷脂醯基甘油、神經節苷脂和腦苷脂）。「陽離子脂質」是指在某一 pH（例如，生理 pH）下攜帶淨正電荷的任何脂質種類。

【0469】 在某些實施例中，本公開文本的脂質奈米顆粒具有某一 N/P 比率。如本文所用，「N/P 比率」或「NP 比率」是指可帶正電的聚合物胺基與帶負電的核酸磷酸基的比率。脂質奈米顆粒/核酸分子複合物的 N/P 特徵可以影響諸如表面淨電荷、穩定性和大小等特性。如本文所述的脂質奈米顆粒的 NP 比率可以是約 1、約 2、約 3、約 4、約 5、約 6、約 7、約 8、約 9、約 10、約 15、約 20、約 25、約 30、約 35、約 40、約 45、約 50、約 55、約 60、約 65、約 70、約 75、約 80、約 85、約 90、約 95、約 100 和其間的任何比率。例如，如本文所述的脂質奈米顆粒的 NP 比率可以是約 18、約 36 或約 72。

【0470】 因此，在某些實施例中，醫藥組合物包含包封於脂質奈米顆粒中的本公開文本的核酸分子以及醫藥上可接受的賦形劑。

【0471】 醫藥組合物可以經配製用於通過推注注射腸胃外投予（即靜脈內、皮下或肌內）。注射用配製物能以單位劑型存在，例如在添加有防腐劑的安瓿中或多劑量容器中。組合物可以採取諸如於油性或水性媒劑中的懸浮液、溶液或乳液等形式，並且含有諸如懸浮劑、穩定劑和/或分散劑等配製劑。可替代地，活性成分可以呈粉末形式，用於用適宜媒劑（例如，無熱原水）來構造。

【0472】 適於腸胃外投予的配製物還包括呈水溶性形式（例如，水溶性鹽）的活性化合物的水溶液。另外，可以投予作為適當油性注射懸浮液的活性化合物的懸浮液。適宜親脂溶劑或媒劑包括脂肪油（例如，芝麻油）或合成脂肪酸酯（例如，油酸乙酯或甘油三酯）。水性注射懸浮液可以含有增加懸浮液粘度的物質，包括例如羧甲基纖維素鈉、山梨醇和葡聚糖。任選地，懸浮液還可以含有穩定劑。脂質體還可以用於包封本公開文本的分子，用於遞送至細胞或間質空間中。例示性醫藥上可接受的載劑是生理上相容的溶劑、分散介體、包衣、抗細菌劑和抗真菌劑、等滲劑和吸收延遲劑、水、鹽水、磷酸鹽緩衝鹽水、右旋糖、甘油、乙醇等。在一些實施例中，組合物包含等滲劑，例如糖、多元醇（如甘露醇、山梨醇）或氯化鈉。在其他實施例中，組合物包含增強活性成分的貯存期限或有效性的醫藥上可接受的物質（如潤濕劑）或少量輔助物質（如潤濕劑或乳化劑、防腐劑或緩衝劑）。

【0473】 本公開文本的組合物可以呈多種形式，包括例如液體（例如，可注射和可輸注的溶液）、分散液、懸浮液、半固體和固體劑型。較佳形式取決於投予方式和治療應用。

【0474】 組合物可以配製為溶液、微乳液、分散液、脂質體或其他適於高藥物濃度的有序結構。無菌可注射溶液可以通過以下方式製備：將活性成分以所需量併入視需要具有上文所列舉成分中的一種或組合的適當溶劑中，之後過濾滅菌。通常，分散液是通過以下方式來製備：將活性成分併入無菌媒劑中，所述媒劑含有基礎分散介體和來自上文所列舉的那些成分的所需其他成分。在用於製備無菌可注射溶液的無菌粉末的情況下，較佳製備方法是真空乾燥和冷凍乾燥，其從預先無菌過濾的溶液產生活性成分加上任何其他所需成分的粉末。可以例如通過以下方式維持溶液的適當流動性：通過使用諸如卵磷脂等包衣，在分散液情況下通過維持所需粒徑，以及通過使用表面活性劑。可注射組合物的延長吸收可以通過以下方式來實現：在組合物中包括延遲吸收的試劑，例如單硬脂酸鹽和明膠。

【0475】 活性成分可以用受控釋放配製物或裝置來配製。此類配製物和裝置的例子包括植入物、經皮貼劑和微包封遞送系統。可以使用生物可

降解的生物相容性聚合物，例如乙烯乙酸乙烯酯、聚酐、聚乙醇酸、膠原、聚原酸酯和聚乳酸。製備此類配製物和裝置的方法是本領域已知的。參見例如，Sustained and Controlled Release Drug Delivery Systems, J. R. Robinson 編輯, Marcel Dekker, Inc., New York, 1978。

【0476】 可注射長效配製物可以通過以下方式來製備：形成藥物於生物可降解聚合物（如聚丙交酯-聚乙交酯）中的微包封基質。根據藥物與聚合物的比率以及所採用聚合物的性質，可以控制藥物釋放速率。其他例示性生物可降解聚合物是聚原酸酯和聚酐。可注射長效配製物還可以通過將藥物捕集於脂質體或微乳液中來製備。

【0477】 可以將補充性活性化合物併入組合物中。在一個實施例中，用凝血因子或其變異體、片段、類似物或衍生物配製本公開文本的核酸分子。例如，凝血因子包括但不限於 V 因子、VII 因子、VIII 因子、IX 因子、X 因子、XI 因子、XII 因子、XIII 因子、凝血酶原、纖維蛋白原、血管性血友病因子或重組可溶性組織因子（rsTF）或前述任一種的活化形式。凝血因子或止血劑還可以包括抗纖維蛋白溶藥，例如 ε-胺基己酸、胺甲環酸 (tranexamic acid)。

【0478】 可以調節劑量方案以提供最佳所需反應。例如，可以投予單次推注，可以隨時間投予若干次分開劑量，或者可以如治療情況的緊迫性所示的成比例地減少或增加劑量。有利的是以劑量單位形式配製腸胃外組合物，以便於投予和劑量均勻。參見例如，Remington's Pharmaceutical Sciences (Mack Pub. Co., Easton, Pa. 1980)。

【0479】 除了活性化合物以外，液體劑型還可以含有惰性成分，如水、乙醇、碳酸乙酯、乙酸乙酯、苯甲醇、苯甲酸苄酯、丙二醇、1,3-丁二醇、二甲基甲醯胺、油、甘油、四氫糠醇、聚乙二醇和山梨聚糖的脂肪酸酯。

【0480】 適宜藥物載劑的非限制性例子也描述於 E. W. Martin 的 Remington's Pharmaceutical Sciences 中。賦形劑的一些例子包括澱粉、葡萄糖、乳糖、蔗糖、明膠、麥芽、水稻、麵粉、白堊、矽膠、硬脂酸鈉、單

硬脂酸甘油酯、滑石、氯化鈉、脫脂乳粉、甘油、丙烯、二醇、水、乙醇等。組合物還可以含有 pH 緩衝試劑以及潤濕劑或乳化劑。

【0481】 對於口服投予，醫藥組合物可以採取通過常規手段製備的片劑或膠囊的形式。還可以將組合物製備為液體，例如糖漿或懸浮液。液體可以包括懸浮劑（例如，山梨醇糖漿、纖維素衍生物或氫化食用脂肪）、乳化劑（卵磷脂或阿拉伯膠）、非水性媒劑（例如，扁桃仁油、油性酯、乙醇或分餾植物油）和防腐劑（例如，對羥基苯甲酸甲酯或對羥基苯甲酸丙酯或山梨酸）。製劑還可以包括矯味劑、著色劑和甜味劑。可替代地，組合物可以作為幹產品存在，用於用水或另一種適宜媒劑來組成(constitution)。

【0482】 對於經頰投予，組合物可以採取根據常規方案的片劑或錠劑的形式。

【0483】 對於通過吸入投予，根據本公開文本使用的化合物便捷地以含有或不含賦形劑的霧化氣溶膠的形式或以氣溶膠噴霧劑的形式從任選地具有推進劑的加壓包或霧化器遞送，所述推進劑是例如二氯二氟甲烷、三氯氟甲烷、二氯四氟甲烷、二氧化碳或其他適宜氣體。在加壓氣溶膠的情況下，劑量單位可以通過提供遞送計量量的閥來確定。例如用於吸入器或吹入器中的明膠的膠囊和藥筒可以配製含有化合物與諸如乳糖或澱粉等適宜粉末基質的粉末混合物。

【0484】 醫藥組合物還可以經配製用於作為例如含有常規栓劑基質（如可可脂或其他甘油酯）的栓劑或保留灌腸劑經直腸投予。

【0485】 在一些實施例中，組合物是通過選自由下列組成之群組的途徑來投予：局部投予、眼內投予、腸胃外投予、鞘內投予、硬膜下投予和口服投予。腸胃外投予可以是靜脈內或皮下投予。

VIII. 治療方法

【0486】 在一些態樣中，本公開文本涉及治療有需要的個體的疾病或病症的方法，其包括投予本文所公開的核酸分子、載體、多肽或醫藥組合物。

【0487】 在一些實施例中，核酸分子包含第一 ITR、第二 ITR 和基因盒，其中基因盒編碼靶序列，其中靶序列編碼治療性蛋白質，並且其中核酸分子用於治療有需要的個體的疾病或病症。在一些實施例中，疾病或病症影響選自由下列組成之群組的器官：肌肉、中樞神經系統（CNS）、眼睛、肝臟、心臟、腎臟、胰腺、肺、皮膚、膀胱、泌尿道及其任何組合。在一些實施例中，個體患有選自由下列組成之群組的疾病或病症：DMD（杜氏肌營養不良）、XLMTM（X-關聯性肌小管性肌病）、帕金森病、SMA（脊髓性肌萎縮）、弗裡德賴希共濟失調、GUCY2D-LCA（利伯先天性黑矇）、XLRS（X-關聯性視網膜劈裂症）、AMD（年齡相關性黃斑變性）、ACHM（全色盲）、RPF65 介導的 IRD 及其任何組合。

【0488】 在一些實施例中，核酸分子包含第一 ITR、第二 ITR 和基因盒，其中基因盒編碼靶序列，其中靶序列編碼 miRNA，並且其中核酸分子用於治療有需要的個體的疾病或病症。在一些實施例中，疾病或病症包含肌萎縮側索硬化（ALS）、亨廷頓病和/或體染色體顯性色素性視網膜炎。

【0489】 在一些實施例中，核酸分子包含第一 ITR、第二 ITR 和基因盒，其中基因盒編碼靶序列，其中靶序列編碼凝血因子，並且其中核酸分子用於治療有需要的個體的出血疾病或病症。出血疾病或病症選自出血凝結障礙、關節積血、肌肉出血、口腔出血、出血症、出血至肌肉中、口腔出血症、創傷、頭部創傷（trauma capitis）、胃腸出血、顱內出血、腹內出血、胸腔內出血、骨折、中樞神經系統出血、咽後間隙出血、腹膜後間隙出血、髂腰肌鞘出血及其任何組合。在仍其他實施例中，安排個體進行手術。在又其他實施例中，治療是預防性的或按需的。

【0490】 本公開文本提供了治療出血障礙的方法，其包括將本公開文本的核酸分子、載體或多肽投予有需要的個體。在一些實施例中，出血障礙的特徵在於缺乏凝血因子（例如，FVIII）。在一些實施例中，出血障礙是血友病。在一些實施例中，出血障礙是血友病 A。在治療出血障礙的方法的一些實施例中，在投予後 24 小時，凝血因子（例如，FVIII）的血漿活性相對於投予參考核酸分子（例如，SEQ ID NO: 16，親代 FVIII 基因序列）、包含參考核酸分子的載體或參考核酸分子編碼的多肽的個體有所增加。

【0491】 本公開文本還涉及治療、改善或預防個體的止血障礙的方法，其包括投予治療有效量的本公開文本的分離的核酸分子或由本公開文本的核酸分子編碼的具有凝血因子（例如，FVIII）活性的多肽。通過分離的核酸分子或所編碼多肽進行的治療、改善和預防可以是繞路療法。接受繞路療法的個體可能已經產生針對凝血因子（例如，FVIII）的抑制劑，或者正在產生凝血因子抑制劑。

【0492】 本公開文本的核酸分子、載體或多肽通過促進纖維蛋白凝塊的形成治療或預防止血障礙。由本公開文本的核酸分子編碼的具有凝血因子（例如，FVIII）活性的多肽可以啟動凝結級聯的成員。凝血因子可以是外在路徑、固有路徑或二者的參與者。

【0493】 本公開文本的核酸分子、載體或多肽可以用於治療已知可用凝血因子治療的止血障礙。可以使用本公開文本的方法治療的止血障礙包括但不限於血友病 A、血友病 B、血管性血友病、XI 因子缺乏（PTA 缺乏）、XII 因子缺乏、以及纖維蛋白原、凝血酶原、V 因子、VII 因子、X 因子或 XIII 因子的缺乏或結構異常、關節積血、肌肉出血、口腔出血、出血症、出血至肌肉中、口腔出血症、創傷、頭部創傷、胃腸出血、顱內出血、腹內出血、胸腔內出血、骨折、中樞神經系統出血、咽後間隙出血、腹膜後間隙出血和髂腰肌鞘出血。

【0494】 在一些實施例中，止血障礙是遺傳性障礙。在一個實施例中，個體患有血友病 A。在其他實施例中，止血障礙是缺乏凝血因子的結果。在其他實施例中，止血障礙是缺乏 FVIII 的結果。在其他實施例中，止血障礙可能是缺陷性 FVIII 凝血因子的結果。

【0495】 在另一個實施例中，止血障礙可以是獲得性障礙 (acquired)。獲得性障礙可能源自潛在的繼發性疾病或病症。無關病症可以是（作為例子但不限制）癌症、自身免疫疾病或妊娠。獲得性障礙可能源自衰老或源自治療潛在繼發性障礙的藥物療法（例如，癌症化學療法）。

【0496】 本公開文本還涉及治療未患止血障礙或導致獲得止血障礙的繼發性疾病或病症的個體的方法。本公開文本因此涉及治療需要通用止血劑的個體的方法，其包括投予治療有效量的本公開文本的分離的核酸分

子、載體或多肽。例如，在一個實施例中，需要通用止血劑的個體正在進行或即將進行手術。本公開文本的分離的核酸分子、載體或多肽可以作為預防藥在手術之前或之後投予。本公開文本的分離的核酸分子、載體或多肽可以在手術期間或之後投予以控制急性出血發作。手術可以包括但不限於肝移植、肝切除或幹細胞移植。

【0497】 在另一個實施例中，本公開文本的分離的核酸分子、載體或多肽可以用於治療患有急性出血發作但未患止血障礙的個體。急性出血發作可能源自嚴重創傷（例如，手術）、車禍、傷口、槍擊撕裂（laceration gun shot）或導致不受控制的出血的任何其他創傷性事件。

【0498】 分離的核酸分子、載體或蛋白質可以用於預防性治療患有止血障礙的個體。分離的核酸分子、載體或蛋白質可以用於治療患有止血障礙的個體的急性出血發作。

【0499】 在另一個實施例中，通過投予本公開文本的分離的核酸分子或載體表現凝血因子蛋白不誘導個體的免疫應答。在一些實施例中，免疫應答包含產生針對凝血因子的抗體。在一個實施例中，免疫應答包含產生針對 FVIII 的抗體。在一些實施例中，免疫應答包含細胞激素分泌。在一些實施例中，免疫應答包含啟動 B 細胞、T 細胞或 B 細胞和 T 細胞二者。在一些實施例中，免疫應答是抑制性免疫應答，其中相對於未產生免疫應答的個體的凝血因子的活性，個體的免疫應答降低凝血因子蛋白的活性。在某些實施例中，通過投予本公開文本的分離的核酸分子或載體表現凝血因子蛋白預防針對凝血因子蛋白或從分離的核酸分子或載體表現的凝血因子蛋白的抑制性免疫應答。

【0500】 在一些實施例中，本公開文本的分離的核酸分子、載體或蛋白質組合物與至少一種促進止血的其他藥劑組合投予。促進止血的所述其他藥劑是具有已證實的凝血活性的治療劑。作為例子但不限制，止血劑可以包括 FV、FVII、FIX、FX、FXI、FXII、FXIII、凝血酶原或纖維蛋白原或前述任一種的活化形式。凝血因子或止血劑還可以包括抗纖維蛋白溶藥，例如 ε-胺基己酸、胺甲環酸。

【0501】 在本公開文本的一個實施例中，組合物（例如，分離的核酸分子、載體或多肽）是凝血因子在投予個體時以可啟動的形式存在的組合物。這種可啟動的分子可以在投予個體後在體內於凝血位點被啟動。

【0502】 因此，在一些實施例中，本公開文本提供了治療有需要的個體的出血障礙的方法，其包括將核酸分子投予個體，所述核酸分子包含側接于基因盒的第一反向末端重複體（ITR）和第二 ITR，所述基因盒包含編碼凝血因子的異源多核苷酸序列，其中第一 ITR 和/或第二 ITR 是非腺相關病毒（非 AAV）的 ITR。在一些實施例中，本公開文本提供了治療有需要的個體的出血障礙的方法，其包括將核酸分子投予個體，所述核酸分子包含側接于基因盒的第一反向末端重複體（ITR）和第二 ITR，所述基因盒包含編碼凝血因子的異源多核苷酸序列，其中第一 ITR 和/或第二 ITR 包含 SEQ ID NO: 180、181、183、184、185、186、187 或 188 中所述的核苷酸序列。在一些實施例中，本公開文本提供了治療有需要的個體的出血障礙的方法，其包括將核酸分子投予個體，所述核酸分子包含側接于基因盒的第一反向末端重複體（ITR）和第二 ITR，所述基因盒包含編碼凝血因子的異源多核苷酸序列，其中第一 ITR 和/或第二 ITR 包含與 SEQ ID NO: 180、181、183、184、185、186、187 或 188 中所述的核苷酸序列至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同的核苷酸序列或其功能衍生物。

【0503】 因此，在一些實施例中，本公開文本提供了治療有需要的個體的血友病 A 的方法，其包括將核酸分子投予個體，所述核酸分子包含側接于基因盒的第一反向末端重複體（ITR）和第二 ITR，所述基因盒包含編碼 VIII 因子的異源多核苷酸序列，其中第一 ITR 和/或第二 ITR 是非腺相關病毒（非 AAV）的 ITR。在一些實施例中，本公開文本提供了治療有需要的個體的血友病 A 的方法，其包括將核酸分子投予個體，所述核酸分子包含側接于基因盒的第一反向末端重複體（ITR）和第二 ITR，所述基因盒包含編碼 VIII 因子的異源多核苷酸序列，其中第一 ITR 和/或第二 ITR 包含 SEQ ID NO: 180、181、183、184、185、186、187 或 188 中所述的核苷酸

序列。在一些實施例中，本公開文本提供了治療有需要的個體的血友病 A 的方法，其包括將核酸分子投予個體，所述核酸分子包含側接于基因盒的第一反向末端重複體（ITR）和第二 ITR，所述基因盒包含編碼 VIII 因子的異源多核苷酸序列，其中第一 ITR 和/或第二 ITR 包含與 SEQ ID NO: 180、181、183、184、185、186、187 或 188 中所述的核苷酸序列至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同的核苷酸序列或其功能衍生物。

【0504】 本公開文本還提供了治療肝臟代謝障礙的方法，其包括將本公開文本的核酸分子、載體或多肽投予有需要的個體。在一些實施例中，肝臟代謝障礙選自苯丙酮尿（

【0505】 **圖 7E** 顯示了來自用含有鼠類 PAH 轉基因和 B19d135 或 GPCd165 ITR 的 ssDNA 處理的 PKU 小鼠的肝臟裂解物的蛋白質免疫印跡。在處理後第 81 天收集肝臟，並提取蛋白質裂解物。每個孔代表單隻動物。使用 M2 抗 FLAG 抗體偵檢 FLAG 標記的鼠類 PAH 蛋白，並且包括 GAPDH 上樣對照以供比較。

【0506】 ）、尿素循環疾病（例如，缺乏轉胺甲醯酶（OTC）或精胺琥珀酸合成酶（ASS））、溶酶體貯積障礙（例如，粘多糖貯積症）和糖原貯積病（例如，I、II、III、IV 型糖原貯積病）。其他肝臟代謝障礙包括但不限於威爾森病（Wilson's disease）、α-1 抗胰蛋白酶缺乏、妊娠期同種免疫肝病（GALD）、脂肪酸氧化缺陷、半乳糖血症、脂質貯積病、酪胺酸血症和過氧化物酶體障礙。

【0507】 在一些實施例中，本公開文本提供了治療有需要的個體的肝臟代謝障礙的方法，其包括將核酸分子投予個體，所述核酸分子包含側接于基因盒的第一反向末端重複體（ITR）和第二 ITR，所述基因盒包含編碼在個體體內缺乏的肝臟相關代謝酶的異源多核苷酸序列，其中第一 ITR 和/或第二 ITR 是非腺相關病毒（非 AAV）的 ITR。在一些實施例中，本公開文本提供了治療有需要的個體的肝臟代謝障礙的方法，其包括將核酸分子投予個體，所述核酸分子包含側接于基因盒的第一反向末端重複體（ITR）

和第二 ITR，所述基因盒包含編碼治療性蛋白質（例如，正常的肝臟代謝功能所需的蛋白質）的異源多核苷酸序列，其中第一 ITR 和/或第二 ITR 包含 SEQ ID NO: 180、181、183、184、185、186、187 或 188 中所述的核苷酸序列。在一些實施例中，本公開文本提供了治療有需要的個體的肝臟代謝障礙的方法，其包括將核酸分子投予個體，所述核酸分子包含側接于基因盒的第一反向末端重複體（ITR）和第二 ITR，所述基因盒包含編碼治療性蛋白質（例如，正常的肝臟代謝功能所需的蛋白質）的異源多核苷酸序列，其中第一 ITR 和/或第二 ITR 包含與 SEQ ID NO: 180、181、183、184、185、186、187 或 188 中所述的核苷酸序列至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同的核苷酸序列或其功能衍生物。

【0508】 在一些實施例中，本公開文本提供了治療有需要的個體的苯丙酮尿（PKU）的方法，其包括將核酸分子投予個體，所述核酸分子包含側接于基因盒的第一反向末端重複體（ITR）和第二 ITR，所述基因盒包含編碼苯丙胺酸羥化酶（PAH）的異源多核苷酸序列，其中第一 ITR 和/或第二 ITR 是非腺相關病毒（非 AAV）的 ITR。在一些實施例中，本公開文本提供了治療有需要的個體的苯丙酮尿（PKU）的方法，其包括將核酸分子投予個體，所述核酸分子包含側接于基因盒的第一反向末端重複體（ITR）和第二 ITR，所述基因盒包含編碼苯丙胺酸羥化酶（PAH）的異源多核苷酸序列，其中第一 ITR 和/或第二 ITR 包含 SEQ ID NO: 180、181、183、184、185、186、187 或 188 中所述的核苷酸序列。在一些實施例中，本公開文本提供了治療有需要的個體的苯丙酮尿（PKU）的方法，其包括將核酸分子投予個體，所述核酸分子包含側接于基因盒的第一反向末端重複體（ITR）和第二 ITR，所述基因盒包含編碼苯丙胺酸羥化酶的異源多核苷酸序列，其中第一 ITR 和/或第二 ITR 包含與 SEQ ID NO: 180、181、183、184、185、186、187 或 188 中所述的核苷酸序列至少約 75%、至少約 80%、至少約 85%、至少約 90%、至少約 95%、至少約 96%、至少約 97%、至少約 98%、至少約 99%或 100%相同的核苷酸序列或其功能衍生物。

【0509】 分離的核酸分子、載體或多肽可以用以下方式投予：靜脈內、皮下、肌內或通過任何粘膜表面，例如口服、舌下、經頰、舌下、經鼻、經直腸、經陰道或通過經肺途徑。可以將凝血因子蛋白植入生物聚合物固相支持體內或與所述固相支持體連接，從而允許將嵌合蛋白緩慢釋放至所需位點。

【0510】 對於口服投予，醫藥組合物可以採取通過常規手段製備的片劑或膠囊的形式。還可以將組合物製備為液體，例如糖漿或懸浮液。液體可以包括懸浮劑（例如，山梨醇糖漿、纖維素衍生物或氫化食用脂肪）、乳化劑（卵磷脂或阿拉伯膠）、非水性媒劑（例如，扁桃仁油、油性酯、乙醇或分餾植物油）和防腐劑（例如，對羥基苯甲酸甲酯或對羥基苯甲酸丙酯或山梨酸）。製劑還可以包括矯味劑、著色劑和甜味劑。可替代地，組合物可以作為幹產品存在，用於用水或另一種適宜媒劑來構造。

【0511】 對於經頰和舌下投予，組合物可以採取根據常規方案的片劑、錠劑或即溶薄膜的形式。

【0512】 對於通過吸入投予，根據本公開文本使用的具有凝血因子活性的多肽便捷地以氣溶膠噴霧劑的形式從具有適宜推進劑（例如，二氯二氟甲烷、三氯氟甲烷、二氯四氟甲烷、二氧化碳或其他適宜氣體）的加壓包或霧化器（例如，於 PBS 中）遞送。在加壓氣溶膠的情況下，劑量單位可以通過提供遞送計量量的閥來確定。例如用於吸入器或吹入器中的明膠的膠囊和藥筒(insufflator)可以配製含有化合物與諸如乳糖或澱粉等適宜粉末基質的粉末混合物。

【0513】 在一個實施例中，分離的核酸分子、載體或多肽的投予途徑是腸胃外的。如本文所用的術語腸胃外包括靜脈內、動脈內、腹膜內、肌內、皮下、直腸或陰道投予。腸胃外投予的靜脈內形式是較佳的。雖然所有這些投予形式都明確地考慮在本公開文本的範圍內，但是投予形式可以是注射用溶液，特別是用於靜脈內或動脈內注射或滴注。通常，適於注射的醫藥組合物可以包含緩衝液（例如乙酸鹽、磷酸鹽或檸檬酸鹽緩衝液）、表面活性劑（例如聚山梨醇酯）、任選地穩定劑（例如人類白蛋白）等。然而，在與本文的傳授內容相容的其他方法中，可以將分離的核酸分子、

載體或多肽直接遞送至不良細胞群的位點，由此增加患病組織對治療劑的暴露。

【0514】 用於腸胃外投予的製劑包括無菌的水性或非水性溶液、懸浮液和乳液。非水性溶劑的例子是丙二醇、聚乙二醇、植物油如橄欖油、以及可注射的有機酯如油酸乙酯。水性載劑包括水、醇性/水性溶液、乳液或懸浮液，包括鹽水和緩衝介體。在本公開文本中，醫藥上可接受的載劑包括但不限於 0.01-0.1 M 且較佳地 0.05 M 磷酸鹽緩衝液或 0.8%鹽水。其他常見腸胃外媒劑包括磷酸鈉溶液、林格氏右旋糖、右旋糖和氯化鈉、乳酸林格氏液或固定油。靜脈內媒劑包括流體和營養補充劑、電解質補充劑（如基於林格氏右旋糖的那些）等。也可以存在防腐劑和其他添加劑，如例如抗微生物劑、抗氧化劑、螯合劑和惰性氣體等。

【0515】 更具體地，適於注射使用的醫藥組合物包括無菌的水性溶液（水溶性的）或分散液、以及用於臨時製備無菌可注射溶液或分散液的無菌粉末。在此類情況下，組合物必須是無菌的並且應當是易於注射的程度的流體。其在製造和儲存條件下應當穩定，並且將較佳地抵抗微生物（如細菌和真菌）的污染作用而保存。載劑可以是溶劑或分散介體，其含有例如水、乙醇、多元醇（例如，甘油、丙二醇和液體聚乙二醇等）及其適宜混合物。可以例如通過以下方式維持適當流動性：通過使用諸如卵磷脂等包衣，在分散液情況下通過維持所需粒徑，以及通過使用表面活性劑。

【0516】 醫藥組合物還可以經配製用於作為例如含有常規栓劑基質（如可可脂或其他甘油酯）的栓劑或保留灌腸劑經直腸投予。

【0517】 用於治療病症的本公開文本的組合物的有效劑量根據許多不同因素而變化，所述因素包括投予方式、靶位元點、患者的生理狀態、患者是人還是動物、所投予的其他藥物以及治療是預防性的還是治療性的。通常，患者是人，但是也可以治療非人哺乳動物，包括轉基因哺乳動物。治療劑量可以使用業內熟習此項技術者已知的常規方法逐步增加，以優化安全性和功效。

【0518】 本公開文本的核酸分子、載體或多肽可以任選地與在需要治療（例如，預防性或治療性）的障礙或病症的治療中有效的其他藥劑組合投予。

【0519】 如本文所用，與輔助療法結合或組合的本公開文本的分離的核酸分子、載體或多肽的投予意指依序、同時、同延(coextensive)、並行、伴隨或同期投予或施用所述療法和所公開多肽。業內熟習此項技術者應瞭解，組合治療方案中各種組分的投予或施用可以定時以增強治療的整體有效性。基於所選輔助療法和本說明書的傳授內容，熟習此項技術者（例如，醫師）在不進行過度實驗的情況下便能夠容易地辨別有效的組合治療方案。

【0520】 應進一步瞭解，本公開文本的分離的核酸分子、載體或多肽可以與一種或多種藥劑結合或組合使用（例如，以提供組合治療方案）。本公開文本的多肽或多核苷酸可以組合的例示性藥劑包括代表用於所治療的特定障礙的當前護理標準的藥劑。此類藥劑在本質上可以是化學的或生物的。術語“生物劑（biologic）”或“生物劑（biologic agent）”是指預期用作治療劑的從活的生物體和/或其產物製備的任何藥學活性劑。

【0521】 要與本公開文本的多核苷酸或多肽組合使用的藥劑的量可以隨個體變化，或者可以根據本領域的知識來投予。參見例如，Bruce A Chabner 等人, *Antineoplastic Agents,* 於 Goodman & Gilman's The Pharmacological Basis of Therapeutics 1233-1287 (Joel G. Hardman 等人編輯, 第 9 版 1996).。在另一個實施例中，投予符合護理標準的量的這種藥劑。

【0522】 在一個實施例中，本文還公開了套組，其包含本文所公開的核酸分子以及將核酸分子投予有需要的個體的說明書。在另一個實施例中，本文公開了用於產生本文所提供的核酸分子的桿狀病毒系統。核酸分子是在昆蟲細胞中產生。在另一個實施例中，提供了表現建構體的奈米顆粒遞送系統。表現建構體包含本文所公開的核酸分子。

IX. 基因療法

【0523】 本公開文本的某些態樣提供了在個體體內表現基因建構體的方法，其包括將本公開文本的分離的核酸分子投予有需要的個體。在一些態樣中，本公開文本提供了增加個體體內多肽的表現的方法，其包括將

本公開文本的分離的核酸分子投予有需要的個體。在其他態樣中，本公開文本提供了調節有需要的個體體內多肽的表現的方法，其包括將本公開文本的分離的核酸分子（例如，包含 miRNA 的核酸序列）投予個體。在一些態樣中，本公開文本提供了下調有需要的個體體內靶基因的表現的方法，其包括將本公開文本的分離的核酸分子（例如，包含 miRNA 的核酸序列）投予個體。

【0524】 已經探索出體基因療法作為多種病症的可能的治療，所述病症包括但不限於血友病 A。基因療法是血友病的非常有吸引力的治療，因為其可能在載體的單次投予後通過連續內源性產生凝血因子（例如，FVIII）治癒所述疾病。血友病 A 非常適合於基因替代方法，因為其臨床表現可完全歸因於缺少以微小量（200 ng/ml）在血漿中循環的單一基因產物（例如，FVIII）。

【0525】 已經顯示使用常規的基於病毒的基因遞送誘導人的免疫應答。病毒衣殼蛋白可以觸發人免疫系統的多種組分。基於 AAV 的基因遞送一直是有吸引力的，因為 AAV 是人群中的常見病毒，大多數人已經暴露於 AAV，並且已經顯示 AAV 的免疫原性低於例如腺病毒。因此，大多數人已經產生針對其先前所暴露的特定變異體的免疫應答。這種預先存在的適應性應答可以包括 NAb 和 T 細胞，它們可以降低隨後用 AAV 再感染的臨床功效和/或已經被轉導的細胞的消除，這使具有預先存在的抗 AAV 免疫力的患者沒有資格接受基於 AVV 的基因療法治療。本公開文本的核酸分子可用于基於非病毒的基因療法。由於使用本公開文本的核酸分子的基因遞送不需要病毒衣殼，所以除非之後向個體再次投予（或再次給藥），否則不會產生針對病毒組分的免疫力。因此，本公開文本的核酸分子允許長期基因遞送策略的再次給藥。

【0526】 另外，如本文所述，本公開文本的核酸分子包含非 AAV 細小病毒 ITR，所述 ITR 側接于基因盒以在投予後驅動穩定的轉基因表現。ITR 的存在是穩定的轉基因表現所必需的，如圖 5 中所示，其中不含 ITR 的核酸不能實現穩定的轉基因表現（參見，「無 ITR 的 dsDNA」和「微環」）。

【0527】 本公開文本的凝血因子蛋白可以在哺乳動物（例如，人類患者）體內產生，使用基因治療方法對於選自由下列組成之群組的出血疾病或障礙的治療將是治療上有益的：出血凝結障礙、關節積血、肌肉出血、口腔出血、出血症、出血至肌肉中、口腔出血症、創傷、頭部創傷、胃腸出血、顱內出血、腹內出血、胸腔內出血、骨折、中樞神經系統出血、咽後間隙出血、腹膜後間隙出血和髂腰肌鞘出血。在一個實施例中，出血疾病或障礙是血友病。在另一個實施例中，出血疾病或障礙是血友病 A。

【0528】 其他病症也適於使用本文所公開的核酸分子治療。在某些實施例中，使用本文所述方法治療影響選自由下列組成之群組的靶器官的疾病或病症：肌肉、中樞神經系統（CNS）、眼睛、肝臟、心臟、腎臟、胰腺、肺、皮膚、膀胱、泌尿道或其任何組合。在某些實施例中，使用本文所述方法治療選自由下列組成之群組的疾病或病症：DMD（杜氏肌營養不良）、XLMTM（X-關聯性肌小管性肌病）、帕金森病、SMA（脊髓性肌萎縮）、弗裡德賴希共濟失調、GUCY2D-LCA（利伯先天性黑矇）、XLRS（X-關聯性視網膜劈裂症）、AMD（年齡相關性黃斑變性）、ACHM（全色盲）、RPF65 介導的 IRD（表 9）。

表9.可通過本文所公開的方法治療的疾病和障礙。

疾病	**靶器官**	**缺陷基因**	**基因療法**
DMD（杜氏肌營養不良）	肌肉	X-關聯性肌營養不良蛋白	基因引入
XLMTM（X-關聯性肌小管性肌病）	肌肉	*MTM1*（肌微管素）	基因引入
帕金森病	CNS	酪胺酸羥化酶、AADC、環水解	基因引入

		酶	
SMA （脊髓性肌萎縮）	CNS	SMN1	基因引入
弗裡德賴希共濟失調	CNS	FXN（粒線體型共濟失調蛋白）	基因引入
GUCY2D-LCA 利伯先天性黑矇	眼部	GUCY2D	基因引入
XLRS X-關聯性視網膜劈裂症	眼部	RS1	基因引入
AMD 年齡相關性黃斑變性	眼部	CFH HTRA ARMS CFB/CC2	基因引入
ACHM 全色盲	眼部	CNGA/CNGB	基因引入
RPF65介導的IRD	眼部	Prf65	基因引入
溶酶體貯積障礙	**靶器官**	**缺陷基因**	**基因療法**
MLD 異染性腦白質營養不良	CNS	ARSA PSAP	基因引入

（溶酶體貯積障礙）			
MPS 粘多糖貯積症 （溶酶體貯積障礙）	肝臟	IDUA（MPS I） IDS（MPS II）	基因引入
PKU 苯丙酮尿 （溶酶體貯積障礙）	肝臟	PAH	基因引入
龐貝 糖原貯積病II型	心臟、肝臟、肌肉、CNS	GAA （酸性α-葡糖苷酶）	基因引入
微小RNA療法	**靶器官**	**缺陷基因**	**基因療法**
ALS 肌萎縮側索硬化	CNS	SOD1[1]	miRNA
亨廷頓病	CNS	HTT[2]	miRNA
AdRP 體染色體顯性色素性視網膜炎	眼部	RHO[3] （視紫質）	miRNA

1：SOD1基因的突變占遺傳性ALS病例的20%。野生型SOD1已經在神經培養中顯示抗細胞凋亡特性，同時已經觀察到突變體SOD1在脊髓粒線體中而不是肝臟粒線體中促進細胞凋亡，但是其在這兩種粒線體中同等表現。下調突變的SOD1表現可能在ALS中抑制運動神經元變性。2：HD是由於基因重複部分的長度超過正常範圍而引起的若干種三核苷酸重複障礙中的一

種。*HTT*含有重複多次的三個DNA鹼基（胞嘧啶-腺嘌呤-鳥嘌呤（CAG））的序列（即...CAGCAGCAG...），稱為三核苷酸重複。CAG是胺基酸麩醯胺酸的3字母遺傳密碼（密碼子），因此一系列CAG導致產生稱為聚麩醯胺酸段（或聚Q段）的麩醯胺酸鏈以及基因的重複部分*聚Q區*。通常，人們在聚Q區中具有少於36個重複麩醯胺酸，這導致產生細胞質蛋白亨廷頓蛋白。然而，36個或更多個麩醯胺酸的序列導致產生具有不同特徵的蛋白質。這種改變的形式稱為突變體亨廷頓蛋白（mHTT），增加某些神經元類型的衰變率。通常，CAG重複數與這個過程受影響的程度相關，並且占症狀發作年齡變化的約60%。其餘變化歸因於環境和修改HD機制的其他基因。36-39個重複產生疾病的外顯率降低形式，其症狀具有遲得多的發作和較慢的進展。在一些情況下，發作可能遲至使得從未注意到症狀。在極大重複計數的情況下，HD具有完全外顯率並且可以在20歲以下發生，這時將其稱為幼年型HD、運動不能-僵硬或韋斯特法爾（Westphal）變異體HD。這種情況占HD載劑的約7%。3：大多數導致色素性視網膜炎的*RHO*基因突變改變視紫質蛋白的折疊或轉運。少數突變引起視紫質組成性活化而不是響應光才被活化。研究提示，改變的視紫質形式干擾必需的細胞功能，從而引起視桿(rods)自毀（經歷細胞凋亡）。由於視桿是低光條件下的視力所必需的，因此這些細胞的損失在患有色素性視網膜炎的人中引起進行性夜盲症。

【0529】 在一些實施例中，使用本文所述方法治療溶酶體貯積障礙。在一些實施例中，溶酶體貯積障礙選自 MLD（異染性腦白質營養不良）、MPS（粘多糖貯積症）、PKU（苯丙酮尿）、龐貝糖原貯積病 II 型或其任何組合。

【0530】 在一些實施例中，本文所述方法用於微小 RNA（miRNA）療法中。在一些實施例中，miRNA 治療因基因或蛋白質過表現而引起的病症。在一些實施例中，miRNA 治療因蛋白質積累而引起的病症。在一些實施例中，miRNA 治療因基因或蛋白質錯誤表現而引起的病症。在一些實施例中，miRNA 治療因突變體基因表現而引起的病症。在一些實施例中，miRNA 治療因異源基因表現而引起的病症。在某些實施例中，miRNA 療法治療選自由下列組成之群組的病症：ALS（肌萎縮側索硬化）、亨廷頓病、

AdRP（體染色體顯性色素性視網膜炎）及其任何組合。在某些實施例中，本公開文本的方法包括通過投予本文所公開的核酸分子靶向治療 ALS，其中核酸分子包含編碼 miRNA 的基因盒，其中 miRNA 靶向 SOD1 的表現。在某些實施例中，miRNA 包含 Dirren 等人, *Annals of Clinical and Translational Neurology 2(2)*:167-84（2015 年 2 月）披露的 miR SOD1 人工 miRNA。SOD1 基因的突變占遺傳性 ALS 病例的 20%。野生型 SOD1 已經在神經培養中顯示抗細胞凋亡特性，同時已經觀察到突變體 SOD1 在脊髓粒線體中而不是肝臟粒線體中促進細胞凋亡，但是其在這兩種粒線體中同等表現。下調突變的 SOD1 表現可能在 ALS 中抑制運動神經元變性。

【0531】 在某些實施例中，本公開文本的方法包括通過投予本文所公開的核酸分子靶向治療亨廷頓病，其中核酸分子包含編碼 miRNA 的基因盒，其中 miRNA 靶向 HTT 的表現。在某些實施例中，miRNA 包含 Evers 等人, *Molecular Therapy 26(9)*:1-15（印刷版之前的電子版，2018 年 6 月）披露的 miHTT 工程化的 miRNA。亨廷頓病是由於基因重複部分的長度超過正常範圍而引起的若干種三核苷酸重複障礙中的一種。HTT 含有重複多次的三個 DNA 鹼基（胞嘧啶-腺嘌呤-鳥嘌呤（CAG））的序列（即...CAGCAGCAG...），其被稱為三核苷酸重複。CAG 是胺基酸麩醯胺酸的 3 字母遺傳密碼（密碼子），因此一系列的這些重複體導致產生稱為聚麩醯胺酸段（或聚 Q 段）的麩醯胺酸鏈以及基因的重複部分聚 Q 區。通常，人們在聚 Q 區中具有少於 36 個重複麩醯胺酸，這導致產生細胞質蛋白亨廷頓蛋白。然而，36 個或更多個麩醯胺酸的序列導致產生具有不同特徵的蛋白質。這種改變的形式稱為突變體亨廷頓蛋白（mHTT），增加某些神經元類型的衰變率。通常，CAG 重複體數目與這個過程受影響的程度相關，並且占症狀發作年齡變化的約 60%。其餘變化歸因於環境和修改亨廷頓病機制的其他基因。36-39 個重複體產生疾病的外顯率降低形式，其症狀具有遲得多的發作和較慢的進展。在一些情況下，發作可能遲至使得從未注意到症狀。在極大重複計數的情況下，亨廷頓病具有完全外顯率並且可以在 20 歲以下發生，這時將其稱為幼年型亨廷頓病、運動不能-僵硬或韋斯特法爾變異體亨廷頓病。這種情況占亨廷頓病載劑的約 7%。

【0532】 在某些實施例中，本公開文本的方法包括通過投予本文所公開的核酸分子靶向治療體染色體顯性色素性視網膜炎（AdRP），其中核酸分子包含編碼 miRNA 的基因盒，其中 miRNA 靶向 RHO（視紫質）的表現。在某些實施例中，miRNA 包含 miR-708（參見 Behrman 等人, *JCB 192(6)*:919-27 (2011)）。大多數導致色素性視網膜炎的 RHO 基因突變改變視紫質蛋白的折疊或轉運。少數突變引起視紫質組成性活化而不是響應光才被活化。研究表明，改變的視紫質形式干擾必需的細胞功能，從而引起視桿自毀（經歷細胞凋亡）。由於視桿是低光條件下的視力所必需的，因此這些細胞的損失在患有色素性視網膜炎的人中引起進行性夜盲症。

【0533】 本文所述的所有各個態樣、實施例和選擇都能以任何和所有變化組合。

【0534】 本說明書中所提到的所有出版物、專利和專利申請都通過引用併入本文，併入程度如同指示每個單獨出版物、專利或專利申請明確且單獨地通過引用併入一般。

【0535】 已經一般性地描述了本公開文本，通過參考本文所提供的實例可以獲得進一步的理解。這些實例僅用於說明的目的，而不旨在是限制性的。

實施例

實施例1.帶有AAV和非AAV細小病毒ITR的FVIII表現建構體的產生。

實例1a.將密碼子優化的FVIII基因和來自AAV的反向末端重複體（ITR）區選殖至基因盒中。

【0536】 FVIII 基因盒是基於 AAV 血清型 2 的基因組來產生。然而，源自任何血清型（包括合成的）的 ITR 區可以用於這種方法中（圖 1A）。

【0537】 設計在肝臟特異性啟動子（TTPp）或泛在(ubiquitious)啟動子（CAGp，圖 1A 和 1B）調控下編碼側接有來自 AAV 的反向末端重複體（ITR）區的密碼子優化的 FVIII 編碼序列（AAV-FVIII）的表現質體 AAV2-FVIIIco6XTEN 用於體外和體內表現，如圖 1C 中所示。基因盒還含有 WPRE 和 bGHpA 元件用於轉基因的最佳表現（圖 1A-1C）。將 ITR 側接

的密碼子優化的FVIII序列選殖至包含ColE1複製起點和賦予安比西林抗性的β-內醯胺酶的表現盒的質體骨架中（圖1C）。側接於表現盒的限制性內切核酸酶PvuII的識別位元點經工程化以允許在PvuII消化時準確切除AAV-FVIII建構體（圖1C）。

實例1b.將密碼子優化的FVIII基因和來自非AAV細小病毒的反向末端重複體（ITR）區選殖至基因盒中。

【0538】 基於AAV所屬的病毒科細小病毒科的成員之間的種系發生關係（圖2A），假設依賴病毒屬的其他非AAV成員和紅病毒屬的成員利用類似的細胞機制維持病毒生命週期並建立持久的潛在性感染。因此，可以利用源自這些病毒的基因組的ITR區研發AAV樣（但並非基於AAV）基因表現盒。測試以下細小病毒的ITR區對研發用於基因療法應用的基因建構體的適合性：依賴病毒屬鵝細小病毒（GPV）株B和紅病毒屬B19細小病毒（圖2A）。

【0539】 細小病毒ITR區在質體載體於細菌細胞中繁殖期間的不穩定性對基因建構體的產生和操作造成挑戰。已經成功產生一些含有全長AAV2 ITR（145 nt）的基因建構體，但是這些建構體非常不穩定，並且大多數基於AAV2 ITR的質體含有ITR區的截短的130 nt形式（例示於表1中）。類似地，所產生的帶有B19 ITR和GPV ITR二者的全長序列的質體建構體在細菌主體中展現高度不穩定性（數據未顯示），這顯著限制了這些ITR對於研發用於基因療法應用的基因載體的實用性。

【0540】 先前已經研發出用於拯救帶有ITR的截短形式的重組B19病毒的反求遺傳學系統（Manaresi等人 Virology 508 (2017): 54-62）（表2B，ITR ID：B19d135）。因此，利用B19d135 ITR產生遺傳上穩定的FVIII表現質體B19-FVIIIco6XTEN（圖1D）。為了進一步利用這種方法合成基於GPV ITR的建構體，比較B19 ITR、GPV ITR和AAV2 ITR的全長野生型序列（圖3A）。分別基於與B19 ITR和AVV2 ITR序列中對於ITR功能非必要的前135個和前15個核苷酸的同源性，假設可以去除GPV ITR的前162個核苷酸以合成具有完整功能的ITR（圖3A，加框序列）的穩定基因建構體。因此，與帶有相應ITR的截短形式的建構體AAV2-FVIIIco6XTEN和

B19-FVIIIco6XTEN 類似，使用 GPVd162（表 2C）產生穩定 FVIII 表現質體建構體 GPV-FVIIIco6XTEN（圖 1E）。值得注意的是，全長 B19 ITR 和 GPV ITR 二者都比全長 AAV2 ITR 長得多（表 1），並且不會形成 AAV ITR 的獨特的 T 形髮夾結構（圖 2B）。

【0541】 含有全長 B19 ITR 序列的質體在細菌宿主細胞中展現高度不穩定性，因為可能產生僅含有 3'ITR 的 FVIIIco6XTEN 表現建構體。使用標準分子選殖技術，可能無法獲得含有 5'和 3'全長 B19 ITR 二者的陽性選殖體。為了產生側接有全長 B19 ITR 的 FVIIIco6XTEN 表現建構體 B19wt-FVIIIco6XTEN（圖 1F），使用特定主體大腸桿菌菌株 PMC103。PMC103 含有基因 *sbcC* 的缺失，所述基因編碼識別並消除十字形 DNA 結構的外切核酸酶。不受理論束縛，認為使用缺少 *sbcC* 的菌株 PMC103 可以允許複製長回文序列（即，含有複雜二級結構的序列）並成功選殖 B19wt-FVIIIco6XTEN 以及 GPVwt-FVIIIco6XTEN。所得質體編碼 383 個鹼基對的野生型 B19 5'和 3' ITR 序列（表 2D），並且另一種質體編碼 444 個鹼基對的野生型 GPV 5'和 3' ITR 序列（表 2F）。

【0542】 如實例 1a 中所述產生含有側接有非 AAV 細小病毒 ITR（B19d135、GPVd162 和 B19wt）的 FVIII 表現盒的質體 B19-FVIIIco6XTEN（圖 1D；表 2B）、GPV-FVIIIco6XTEN（圖 1E；表 2C）和 B19wt-FVIIIco6XTEN（圖 1F，表 2D）。使用限制性內切核酸酶 LguI 的識別位點側接於所有 FVIII 表現盒（圖 1D-1F）。

實例1c.含有側接有AAV和非AAV細小病毒ITR的FVIII表現盒的單鏈DNA片段的製備。

【0543】 假設側接於 FVIII 表現盒的 ITR 區內髮夾結構的形成會驅動靶細胞的持久轉導。對於概念驗證研究，分別用 PvuII 和 LguI 消化基於 AAV ITR 的質體 AAV2-FVIIIco6XTEN 以及基於非 AAV ITR 的質體 B19-FVIIIco6XTEN 和 GPV-FVIIIco6XTEN。具有形成的髮夾 ITR 結構的單鏈（ss）AAV-FVIII、B19-FVIII 或 GPV-FVIII 片段是通過以下方式來產生：在 95°C 下使 PvuII 或 LguI 消化的雙鏈 DNA 片段產物（FVIII 表現盒和質體骨架）變性，然後在 4°C 下冷卻以允許回文 ITR 序列折疊（圖 1A-1B）。

在 HemA（血友病 A）小鼠模型中測試所得 ssAAV-FVIII、ssB19-FVIII 或 ssGPV-FVIII 建立肝細胞的持久轉導的能力。

實例1d.使用桿狀病毒表現系統產生FVIII表現建構體。

【0544】 將利用描述於 Li 等人, PLoS ONE 8(8): e69879 (2013)中的桿狀病毒表現系統在昆蟲細胞中產生呈閉端 DNA（ceDNA）分子形式的 AAV-FVIII、B19-FVIII 和 GPV-FVIII 建構體。已經證實 ceDNA 表現盒的全身遞送建立肝細胞的持久轉導，並驅動肝臟中的穩定長期轉基因表現。

實例2.在HemA小鼠體內，包含側接有AAV和非AAV細小病毒ITR的FVIII表現盒的基因建構體的全身注射導致長期FVIII表現。

實例2a. ssAAV-FVIII介導的FVIII表現的體內評價。

【0545】 為了驗證帶有 AAV ITR 區的 ssAAV-FVIII 在體內介導持久轉基因表現的能力，在 5-12 周齡血友病 A（HemA）小鼠（4 隻動物/組）中以 5 μg、10 μg、20 μg ssDNA 基因表現盒（ssAAV-FVIII）通過流體力學注射全身遞送所述基因表現盒（圖 4A）。HDI 導致所注射材料主要遞送至實驗動物的肝臟中。在單次流體力學注射 ssAAV-FVIII 後 18 小時、3 天、2 周、3 周、1 個月、2 個月、3 個月和 4 個月，從實驗動物收集血漿樣本。通過產色 FVIII 活性測定分析血液中的 FVIII 血漿活性。用 5 μg/小鼠的親代表現質體注射的對照動物在投予後不久即顯示高 FVIII 血漿活性水準。然而，循環 FVIII 的水準快速下降，並且截至注射後（p.i.）15 天變得無法偵檢。相比之下，用 5、10 和 20 μg/小鼠的 ssAAV-FVIII 注射的實驗動物產生了轉基因的長期表現，循環 FVIII 的穩定水準分別為正常 FVIII 水準的約 8%、16%和 32%（圖 4A）。觀察到強劑量反應，表明所注射劑量與治療結果之間高度相關。

實例2b. ssB19-FVIII和ssGPV-FVIII介導的FVIII表現的體內評價。

【0546】 為了評價來自分別帶有非 AAV 細小病毒 ITR 區 B19d135 和 GPVd162 的 ssB19-FVIII 和 ssGPV-FVIII 的 FVIII 的體內表現，在 5-12 周齡血友病 A（HemA）小鼠體內通過 HDI 全身遞送 10 或 20 μg/小鼠的 ssB19-FVIII 以及 10 或 50 μg/小鼠的 ssGPV-FVIII 基因表現盒。在注射後 1、

3、7、14、21、28、42、56、84、112、140和168天收集血液樣本，並通過產色FVIII活性測定分析血液中的FVIII活性。如在AAV-FVIII建構體的情況下所觀察到的，用5 μg/小鼠的親代FVIII表現質體注射的對照動物在注射後24小時顯示高FVIII血漿活性水準，其快速下降，並且截至注射後14天變得無法偵檢。用ssB19-FVIII注射的實驗動物在注射後3天顯示峰值FVIII血漿活性，其然後在21天期間逐漸下降，並且在注射後28天左右穩定。（圖4B）另一方面，用ssGPV-FVIII注射的HemA小鼠在第112天左右產生穩定的FVIII血漿活性水準，其在剩餘觀察階段期間維持（圖4C）。值得注意的是，用10 μg/小鼠的ssAAV-FVIII（圖4A）或ssGPV-FVIII（圖4C)注射的動物產生了非常相似的穩定FVIII血漿活性水準，表明AAV2 ITR和GPV ITR區二者都包含有效建立靶細胞的持久轉導所需的遺傳因子。

實例2c. FVIII在hemA小鼠體內的穩定長期表現的ITR和髮夾需求的體內評價。

【0547】 為了比較單鏈DNA盒與替代核酸治療劑的穩定性和長期表現，用PvuII或AflII消化FVIIIco6XTEN質體建構體（圖1A）以產生具有或不具有AAV ITR序列的雙鏈線性DNA。純化不具有ITR的線性雙鏈DNA以產生'無ITR的dsDNA'建構體。最後，通過重疊AflII識別位點連接純化的不具有ITR的dsDNA，導致形成微環DNA。這種小環狀類質體DNA建構體缺乏任何細菌序列和/或ITR序列。通過流體力學注射用等摩爾濃度的DNA建構體注射HemA小鼠，並根據在2-4個月期間收集的血漿確定FVIII活性水準。所有DNA建構體都產生了在30%-60%正常範圍內的初始治療水準的FVIII，然而，只有單鏈DNA在注射後持續4個月顯示32%的轉基因表現的穩定持久性（圖5）。所有雙鏈DNA和微環DNA都在第14-42天達到了6%-10%正常值的穩定表現水準，然而，這些平臺期僅代表所觀察到的初始FVIII活性的10%。因為瞬時升高水準的FVIII表現可以導致形成中和抗藥物抗體，因此在血友病A環境中的免疫耐受性需要穩定表現。

實例2d.野生型與衍生B19 ITR的體內比較。

【0548】 為了比較B19衍生ITR（B19d135，圖1D，表2C）與全長B19 ITR（表2C）的作用，產生側接有248個鹼基對的ITR（圖1F）的

FVIIIco6XTEN表現盒。用30 μg側接有B19d135（圖1D）、GPVd165（圖1E）或野生型B19 ITR（圖1F）的單鏈FVIII-DNA流體力學注射血友病A小鼠。在注射後3、7、14、21、28和35天收集所有群組的血漿，B19d135和GPVd165建構體的其他樣本在第42天、第55天和第84天採集，並通過產色測定分析FVIII活性(圖6)。與衍生B19 ITR相比，全長ITR導致FVIII表現增加約2.5倍。此外，來自野生型ITR的FVIII的表現在開始時是穩定的。

實例2e.體內單鏈裸DNA的再次投予的評價。

【0549】 當前基因治療方式的關鍵限制是由於形成針對基因療法載體的病毒衣殼的抗藥物抗體而無法再次投予治療劑。然而，不存在免疫原性蛋白質的基因療法系統可以再次給藥，以使患者逐步增加至所需治療水準。為了評價我們的側接非AAV ITR的單鏈盒是否可以再次投予，在第0天和第35天用30 μg含有B19d135 ITR和GPVd165 ITR的ssDNA注射hemA小鼠（圖6）。在觀察的第一個月期間，投予GPVd165-FVIII的小鼠達到了大約5%正常值的穩定FVIII水準。在ssDNA的第二劑量後，FVIII水準上升至10%，之後略有下降，顯示FVIII水準的2倍增加。在第一週期間，投予B19d135-FVIII的小鼠達到了8%的穩定FVIII水準，其上升約3.5倍至30%，之後下降至25%。這些數據證實，在血友病A小鼠體內，再次投予具有非AAV ITR的單鏈DNA可以增加FVIII的穩定表現水準。

實例3.帶有B19d135和GPVd162非AAV細小病毒ITR的衍生物的FVIII表現建構體的產生和體內評價。

實例3a. B19 ITR和GPV ITR的最小必需序列的確定。

【0550】 基於依賴病毒屬AAV2和GPV以及紅病毒屬B19的ITR序列之間的比較（Gene Bank登錄號分別為NC_001401.2、U25749.1和KY940273.1），設計用帶有此類ITR的基因建構體持久轉導真核細胞所需的GPV和B19細小病毒ITR的最小序列，其具有或不具有其他序列（間隔子、插入、倒位、添加和/或用其他細小病毒ITR的野生型序列重組）（圖3A和3B）。AAV2 ITR、GPV ITR和B19 ITR的序列比對揭示了所有三種

病毒物種之間的保守區 B19v1 和 GPVv1（呈現於表 2A-2C 中），其作為連續序列而不含可變序列的間隔子區。同樣，基於 B19 ITR 與 GPV ITR 之間的序列比較設計最小必需序列變異體 B19v3 和 GPVv3。由於帶有 GPVd162 ITR 的 FVIII 表現建構體在體內實驗中的表現優於帶有 B19d135 ITR 的基因建構體，假設 B19v3 序列包含在 B19 ITR 與 GPV ITR 序列之間保守的最小 B19 ITR 序列區，並且 GPVv3 序列包含在 GPV ITR 序列中存在並且在 B19 ITR 序列中缺少的最小 GPV ITR 序列區（表 2B 和 2C）。序列 B19v2 和 GPVv2 分別是通過排除 B19 ITR 和 GPV ITR 序列的 ITR 回文序列區中前 135 個和前 162 個核苷酸以及相應互補的 135 個和 162 個核苷酸來產生（表 2B 和 2C）。

實例3b. B19 ITR和GPV ITR及其衍生物的回文區在功能性基因建構體上的定向。

【0551】 細小病毒 ITR 的一部分是由自我互補的回文區組成。先前已證實，對於重組感染性 B19 細小病毒，所拯救的帶有呈正向和反向定向的回文區的病毒展現類似的生長特性（Manaresi 等人 Virology 508 (2017): 54-62）。因此，提出不管此類 ITR 的回文區關於基因表現盒是呈 5'和 3' ITR 組合的正向、反向或任何可能組合，帶有 B19 ITR 和 GPV ITR 及其衍生物的基因表現建構體保持功能性。為了驗證這個假設，將使用相同物種 ITR 的相同以及反向的互補序列，將 B19d135 ITR 和 GPVd162 ITR 以及野生型 B19 ITR 和 GPV ITR 以正向、反向和倒轉定向併入 FVIIIco6XTEN 表現盒中。將產生來自這些質體的單鏈 DNA 並如實例 2a、2b 和 2d 中所述在血友病 A 小鼠體內測試由 TTPp 啟動子驅動的肝臟定向的 FVIII 表現。除了研究相同物種 ITR 的所有定向以外，還將產生 GPV 和 B19 野生型 ITR 及其衍生物的組合並在血友病 A 小鼠體內測試 FVIII 表現。這些表現盒將含有 1 個 B19 來源的 ITR 和 1 個 GPV 來源的 ITR，以確定非同源 ITR 序列是否可以增強所需轉基因的附加型連環化以及長期表現。將通過流體力學注射用 10、20 或 50 μg 含有前述表現盒的 ssDNA 注射血友病 A 小鼠，並且將從在注射後以每週間隔收集的鼠類血漿量測 FVIII。將對投予這些表現盒的小鼠體內的 FVIII 表現和壽命的影響與投予 B19d135、GPVd162 和相應野生型

ITR 表現盒的小鼠體內的 FVIII 表現和壽命進行直接比較（表 2B、2C、2D 和 2F）。

實例3c.在HemA小鼠體內帶有B19d135和GPVd162非AAV細小病毒ITR的衍生物的基因建構體的全身注射。

【0552】 為了評價來自帶有 B19d135 和 GPVd162 非 AAV 細小病毒 ITR 的衍生物的 ssDNA 建構體的 FVIII 體內表現，將在 5-12 周齡 HemA 小鼠體內通過 HDI 全身遞送 5、10、20 或 50 μg/小鼠的每種 ssDNA 基因表現盒。將在注射後 1、3、7、14、21 和 28 天，然後在 4 個月期間每個月一次收集血液樣本。將通過產色 FVIII 活性測定分析血液中的 FVIII 活性。

實例4.帶有B19d135和GPVd162非AAV細小病毒ITR的衍生物的ceDNA表現建構體在昆蟲細胞中的產生和體內評價。

實例4a.使用桿狀病毒表現系統產生帶有B19d135 ITR和GPVd162 ITR的ceDNA表現建構體。

【0553】 與實例 1d 中所述的 AAV-FVIII、B19-FVIII 和 GPV-FVIII 建構體類似，將使用桿狀病毒表現系統在昆蟲細胞中產生呈 ceDNA 形式的帶有 B19d135 和 GPVd162 非 AAV 細小病毒 ITR 的衍生物的基因建構體的 FVIII 表現。

實例4b.在HemA小鼠體內帶有B19d135和GPVd162非AAV細小病毒ITR的衍生物的ceDNA表現建構體的全身注射。

【0554】 為了評價來自帶有 B19d135 和 GPVd162 非 AAV 細小病毒 ITR 的衍生物的 ceDNA 建構體的 FVIII 體內表現，將在 5-12 周齡 HemA 小鼠體內通過 HDI 全身遞送 5、10、20 或 50 μg/小鼠的每種 ceDNA 基因表現盒。將在注射後 1、3、7、14、21 和 28 天，然後在 4 個月期間每個月一次收集血液樣本。將通過產色 FVIII 活性測定分析血液中的 FVIII 活性。

實例5. ssDNA和ceDNA FVIII表現建構體的脂質奈米顆粒配製物的產生。

【0555】 在如實例 1 和 4 中所述產生每種 ssDNA 或 ceDNA 後，將使用適當的脂質組合物通過微流體混合將每種基因建構體配製到脂質奈米顆

粒（LNP）中（LNP-ssDNA 和 LNP-ceDNA）。將調節脂質與 DNA 的比率（N/P）以優化細胞轉導和 FVIII 表現。將使用配製到 LNP 中的編碼側接有 AAV 或非 AAV 細小病毒 ITR 的 FVIII 表現盒的親代質體作為轉導效率的對照。

實例6. LNP-ssDNA和LNP-ceDNA的體外和體內評價。

實例6a.在所培養肝細胞中ssDNA和ceDNA介導的FVIII表現的體外評價。

【0556】 如實例 5 中所述將 ssDNA 或 ceDNA FVIII 表現基因建構體和相應的親代對照質體配製到 LNP 中用於靶向基因遞送。將 Huh7 細胞以 1 x 10^5 個細胞/孔接種至 24 孔組織培養板中並培育過夜。在第二天，將 LNP-ssDNA 或配製物以 1000、500、250、125 和 62.5 ng/孔添加至細胞上。在轉導後 24 小時更換培養基後，在轉導後 48 小時收穫培養基。與人血漿 FACT 標準品相比，通過產色 FVIII 活性測定量測培養基中的 FVIII 活性。將帶有在 CAGp 啟動子下並且側接有 AAV ITR 的 FVIIIco6XTEN 盒的質體以 72、36 和 18 的 N/P 比率包封於脂質奈米顆粒中（圖 8A）。在轉導 Huh7 細胞後，在條件化培養基中量測 FVIII。以 18 的 N/P 比率轉導細胞產生了 36 和 72 的比率的增加的 FVIII 水準，且 1 μg/ml 的峰值劑量產生超過 2 IU/ml。這個數據證實了 LNP 遞送在肝臟靶細胞中的實用性。為了研究通過 LNP 遞送 ssDNA 在肝臟特異性啟動子下的轉導效率，將在 TTPp 啟動子下的 FVIIIco6XTEN 盒以 2 N/p 比率包封並轉導 Huh7 細胞（圖 8B）。與我們的先前數據一致（圖 8A），與 36 的比率相比，18 的 N/P 比率導致增加的 FVIII 活性水準。另外，這個數據證實了 FVIII ssDNA 在肝細胞中的概念驗證 LNP 遞送。在 24 小時後，用 2 μg/ml 單鏈 FVIIIco6XTEN-AAV 轉導的約 2 x 10^5 個 Huh7 細胞產生了 0.33 IU/ml FVIII。

【0557】 另外，文獻中已經顯示，細胞組蛋白沿 rAAV 附加體規則定位，從而產生與細胞染色體 DNA 核小體模式類似的染色質樣結構。因此，還將通過南方(Southern)印跡來評估這些建構體建立持久轉導靶細胞所需的染色質樣核小體結構的能力。

實例6b.在HemA小鼠體內在靜脈內投予後對LNP配製的ssDNA和ceDNA介導

的長期FVIII表現的評價。

【0558】 將以 5、10、20、40、100 ug/小鼠通過 IV 注射將 LNP-ssDNA、LNP-ceDNA 或 LNP-pDNA（質體對照）投予 5-12 周齡 HemA 小鼠，N = 4/組。將在注射後 48 小時開始長達 6 個月中的所選時間點收集血液樣本，並將通過產色 FVIII 活性測定分析血液中的 FVIII 活性。將針對實例 1 和 4 中所述的每種基因建構體比較用 LNP-ssDNA 或 LNP-ceDNA 處理的小鼠體內的 FVIII 表現譜與用 LNP-pDNA 處理的小鼠的 FVIII 表現譜。

實例6c.在加強注射後ssDNA或ceDNA介導的FVIII表現的體內評價。

【0559】 將在初始注射後 2 個月向實例 6b 中用 LNP-ssDNA 或 LNP-ceDNA 處理的小鼠的亞組投予相同劑量的相應 LNP 的額外 IV 注射加強。將在加強注射後 48 小時開始長達 6 個月中的所選時間點收集血液樣本。將通過產色 FVIII 活性測定分析血液中的 FVIII 活性。將用 LNP-ssDNA 或 LNP-ceDNA 處理的小鼠體內的 FVIII 表現譜與用相應 LNP-pDNA 處理的小鼠的 FVIII 表現譜進行比較。

實例7.帶有B19或GPV來源的ITR的基因表現建構體在基因療法中的一般使用的實用性。

實例7a.帶有B19或GPV來源的ITR的報告子基因建構體的產生。

【0560】 為了證實基於非 AAV ITR 的基因表現系統作為平臺在基因療法應用中的一般使用的實用性，基於實例 1b 中所述的建構體用側接有 B19d135 ITR 或 GPVd162 ITR 的綠色螢光蛋白（GFP）或螢光素酶（luc）產生包含表現盒的報告子基因建構體。因此，通過常規分子選殖技術用 GFP 或 luc 的 ORF 替代 B19-FVIIIco6XTEN（圖 1C）和 GPV-FVIIIco6XTEN（圖 1D）中 FVIII 的開放閱讀框（ORF）。

【0561】 還產生含有鼠類苯丙胺酸羥化酶（PAH）轉基因的側接有 B19d135 ITR 或 GPVd162 ITR 的表現盒（圖 7A），使用其在苯丙酮尿的相關小鼠模型中評價 PAH 表現和血液苯丙胺酸濃度的降低。使用這個模型，通過流體力學注射將 200 μg 側接有非 AAV ITR 的 ssDNA 投予 PKU 小鼠（n = 3）用於肝臟表現。在第 3、7、14、28、42、56、70 和 81 天收集血液樣

本，並分離血漿用於苯丙胺酸濃度確定（圖 7B-7C）。接受含有 B19d135 ITR 的表現盒的小鼠在第 3 天展現苯丙胺酸水準從 370 μg/ml 至 210 μg/ml 的下降，所述水準穩定維持至第 81 天（圖 7B）。接受 GPVd162 ITR 盒的小鼠在第 14 天顯示血液苯丙胺酸水準從 350 μg/ml 至 310 μg/ml 的下降，所述水準繼續下降至截至第 42 天的 250 μg/ml 的穩定水準（圖 7C）。血液苯丙胺酸濃度的這些下降代表與注射前濃度相比 45%和 30%的降低（圖 7D）。為了確認鼠類 PAH 蛋白在肝臟中的存在，對在注射後第 81 天取自所處理小鼠的肝臟裂解物進行蛋白質印跡。使用抗 FLAG 標簽抗體偵檢鼠類 PAH 蛋白，圖 7E 顯示在 6 只所處理動物的 5 只中可偵檢的鼠類 PAH 蛋白，並且在用含有 B19d135 ITR 的 ssDNA 處理的小鼠體內觀察到顯著較高的蛋白質水準。這些數據與在圖 7B-7D 中觀察到的血液苯丙胺酸減少一致。總而言之，這些證實示，單鏈 DNA 遞送可以導致功能性肝酶的長期表現。

【0562】 實驗中所用的各種 PAH 建構體的序列陳述於表 10A 和 10B 中。

表10A：帶有B19d135 ITR的B19-PAH建構體（核苷酸1-4146；SEQ ID NO: 197）

描述	序列
5' ITR（SEQ ID NO: 180）	CTCTGGGCCAGCTTGCTTGGGGTTGCCTTGACACTAAGACAAGCGGCGCGCC GCTTGATCTTAGTGGCACGTCAACCCCAAGCGCTGGCCCAGAGCCAACCCTA ATTCCGGAAGTCCCGCCCACCGGAAGTGACGTCACAGGAAATGACGTCACAG GAAATGACGTAATTGTCCGCCATCTTGTACCGGAAGTCCCGCCTACCGGCGG CGACCGGCGGCATCTGATTTGGTGTCTTCTTTTAAATTTT
CAGp 啟動子（SEQ ID NO: 195）	CTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATG GAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCA ACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCC AATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCC

	CACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGT CAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGG GACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGCA TGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCATCTCCCCCCCCTCC CCACCCCCAATTTTGTATTTATTTATTTTTTAATTATTTTGTGCAGCGATGGGG GCGGGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGG GCGGGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGC GCTCCGAAAGTTTCCTTTTATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAA AAGCGAAGCGCGCGGCGGGCG
合成內含子（SEQ ID NO: 192）	GTGAGCGGGCGGGACGGCCCTTCTCCTTCGGGCTGTAATTAGCGCTTGGTTTA ATGACGGCTTGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTGAGGGGCTCCGGG AGGGCCCTTTGTGCGGGGGGAGCGGCTCGGGGCTGTCCGCGGGGGGACGGCT GCCTTCGGGGGGGACGGGGCAGGGCGGGGTTCGGCTTCTGGCGTGTGACCGG CGGCTCTAGAGCCTCTGCTAACCATGTTCATGCCTTCTTCTTTTTCCTACAG
鼠類PAH序列（SEQ ID NO: 196）	ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGATATCGATTACA AGGATGACGATGACAAGGCTGCTGTGGTTCTGGAAAATGGCGTGCTGAGCCG GAAGCTGAGCGACTTCGGACAAGAGACAAGCTACATCGAGGACAACAGCAA CCAGAATGGCGCCGTGTCTCTGATCTTCAGCCTGAAAGAAGAAGTGGGCGCC CTGGCCAAGGTGCTGAGACTGTTCGAGGAAAACGAGATCAATCTGACCCACA TCGAGAGCAGACCCAGCAGACTGAACAAGGACGAGTACGAGTTCTTCACCTA CCTGGACAAGCGGAGCAAGCCTGTGCTGGGCAGCATCATCAAGAGCCTGAGA AACGACATCGGCGCCACCGTGCACGAGCTGAGCAGAGACAAAGAAAAGAAC ACCGTGCCATGGTTCCCCAGGACCATCCAAGAGCTGGACAGATTCGCCAACC AGATCCTGAGCTATGGCGCCGAGCTGGACGCTGATCACCCTGGCTTTAAGGA

	CCCCGTGTACCGGGCCAGAAGAAAGCAGTTTGCCGATATCGCCTACAACTAC CGGCACGGCCAGCCTATTCCTCGGGTCGAGTACACCGAGGAAGAGAGAAAG ACCTGGGGCACCGTGTTCAGAACCCTGAAGGCCCTGTACAAGACCCACGCCT GCTACGAGCACAACCACATCTTCCCACTGCTGGAAAAGTACTGCGGCTTCCG CGAGGACAATATCCCTCAGCTCGAAGACGTGTCCCAGTTCCTGCAGACCTGC ACCGGCTTTAGACTGAGGCCTGTTGCCGGACTGCTGAGCAGCAGAGATTTTCT CGGCGGCCTGGCCTTCAGAGTGTTCCACTGTACCCAGTACATCAGACACGGC AGCAAGCCCATGTACACCCCTGAGCCTGATATCTGCCACGAGCTGCTGGGAC ATGTGCCCCTGTTCAGCGATAGAAGCTTCGCCCAGTTCAGCCAAGAGATCGG ACTGGCTTCTCTGGGAGCCCCTGACGAGTACATTGAGAAGCTGGCCACCATC TACTGGTTCACCGTGGAATTCGGCCTGTGCAAAGAGGGCGACAGCATCAAGG CTTATGGCGCTGGACTGCTGTCTAGCTTCGGCGAGCTGCAGTACTGTCTGAGC GACAAGCCTAAGCTGCTGCCCCTGGAACTGGAAAAGACCGCCTGCCAAGAGT ACACAGTGACCGAGTTCCAGCCTCTGTACTACGTGGCCGAGAGCTTCAACGA CGCCAAAGAAAAAGTGCGGACCTTCGCCGCCACCATTCCTCGGCCTTTTAGC GTCAGATACGACCCCTACACACAGCGCGTGGAAGTGCTGGACAACACACAGC AGCTGAAGATTCTGGCCGACTCCATCAACAGCGAAGTGGGCATTCTGTGTCA CGCCCTGCAGAAGATCAAGAGCTGA
WPRE（突變的土撥鼠肝炎病毒轉錄後調控元件）（SEQ ID NO: 120）	TCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTA ACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATC ATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGT TGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTG TGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTG TCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACT

	CATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTG ACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCC TGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGC CCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTC TTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCC TCCCCGCTG
bGHpA（牛生長激素多聚腺苷酸化信號）(SEQ ID NO: 122)	CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTT CCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAA ATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGG GCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGA
3' ITR反向末端重複體（SEQ ID NO: 181）	AAAATTTAAAAGAAGACACCAAATCAGATGCCGCCGGTCGCCGCCGGTAGGC GGGACTTCCGGTACAAGATGGCGGACAATTACGTCATTTCCTGTGACGTCATT TCCTGTGACGTCACTTCCGGTGGGCGGGACTTCCGGAATTAGGGTTGGCTCTG GGCCAGCGCTTGGGGTTGACGTGCCACTAAGATCAAGCGGCGCGCCGCTTGT CTTAGTGTCAAGGCAACCCCAAGCAAGCTGGCCCAGAG
全長序列（SEQ ID NO: 197）	
CTCTGGGCCAGCTTGCTTGGGGTTGCCTTGACACTAAGACAAGCGGCGCGCCGCTTGATCTTAGT GGCACGTCAACCCCAAGCGCTGGCCCAGAGCCAACCCTAATTCCGGAAGTCCCGCCCACCGGAA GTGACGTCACAGGAAATGACGTCACAGGAAATGACGTAATTGTCCGCCATCTTGTACCGGAAGT CCCGCCTACCGGCGGCGACCGGCGGCATCTGATTTGGTGTCTTCTTTTAAATTTTGCGGCAATTC AGTCGATAACTATAACGGTCCTAAGGTAGCGATTTAAATACGCGCTCTCTTAAGGTAGCCCCGG GACGCGTCAATTGAGATCTGGATCCGGTACCGAATTCGCGGCCGCCTCGACGACTAGCGTTTAG TAATGAGACGCACAAACTAATATCACAAACTGGAAATGTCTATCAATATATAGTTGCTCTAGTT ATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAA	

CTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGA

CGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACG

GTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTC

AATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTG

GCAGTACATCTACGTATTAGTCATCGCTATTACCATGCATGGTCGAGGTGAGCCCCACGTTCTGC

TTCACTCTCCCCATCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTATTTT

GTGCAGCGATGGGGGCGGGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAG

GGGCGGGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAA

GTTTCCTTTTATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGC

GGGAGTCGCTGCGCGCTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCGCGCCGCCCGCCCC

GGCTCTGACTGACCGCGTTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTTCGGGCTG

TAATTAGCGCTTGGTTTAATGACGGCTTGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTGAGGGGC

TCCGGGAGGGCCCTTTGTGCGGGGGGAGCGGCTCGGGGCTGTCCGCGGGGGGACGGCTGCCTTC

GGGGGGGACGGGGCAGGGCGGGGTTCGGCTTCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTG

CTAACCATGTTCATGCCTTCTTCTTTTTCCTACAGCTCCTGGGCAACGTGCTGGTTATTGTGCTGT

CTCATCATTTTGGCAAAGAATTGGATCGCGAAGCCGCCACCATGGACTACAAAGACCATGACGG

TGATTATAAAGATCATGATATCGATTACAAGGATGACGATGACAAGGCTGCTGTGGTTCTGGAA

AATGGCGTGCTGAGCCGGAAGCTGAGCGACTTCGGACAAGAGACAAGCTACATCGAGGACAAC

AGCAACCAGAATGGCGCCGTGTCTCTGATCTTCAGCCTGAAAGAAGAAGTGGGCGCCCTGGCCA

AGGTGCTGAGACTGTTCGAGGAAAACGAGATCAATCTGACCCACATCGAGAGCAGACCCAGCA

GACTGAACAAGGACGAGTACGAGTTCTTCACCTACCTGGACAAGCGGAGCAAGCCTGTGCTGG

GCAGCATCATCAAGAGCCTGAGAAACGACATCGGCGCCACCGTGCACGAGCTGAGCAGAGACA

AAGAAAAGAACACCGTGCCATGGTTCCCCAGGACCATCCAAGAGCTGGACAGATTCGCCAACC

AGATCCTGAGCTATGGCGCCGAGCTGGACGCTGATCACCCTGGCTTTAAGGACCCCGTGTACCG

GGCCAGAAGAAAGCAGTTTGCCGATATCGCCTACAACTACCGGCACGGCCAGCCTATTCCTCGG

GTCGAGTACACCGAGGAAGAGAGAAAGACCTGGGGCACCGTGTTCAGAACCCTGAAGGCCCTG

TACAAGACCCACGCCTGCTACGAGCACAACCACATCTTCCCACTGCTGGAAAAGTACTGCGGCT

TCCGCGAGGACAATATCCCTCAGCTCGAAGACGTGTCCCAGTTCCTGCAGACCTGCACCGGCTT

TAGACTGAGGCCTGTTGCCGGACTGCTGAGCAGCAGAGATTTTCTCGGCGGCCTGGCCTTCAGA

GTGTTCCACTGTACCCAGTACATCAGACACGGCAGCAAGCCCATGTACACCCCTGAGCCTGATA

TCTGCCACGAGCTGCTGGGACATGTGCCCCTGTTCAGCGATAGAAGCTTCGCCCAGTTCAGCCA

AGAGATCGGACTGGCTTCTCTGGGAGCCCCTGACGAGTACATTGAGAAGCTGGCCACCATCTAC

TGGTTCACCGTGGAATTCGGCCTGTGCAAAGAGGGCGACAGCATCAAGGCTTATGGCGCTGGAC

TGCTGTCTAGCTTCGGCGAGCTGCAGTACTGTCTGAGCGACAAGCCTAAGCTGCTGCCCCTGGA

ACTGGAAAAGACCGCCTGCCAAGAGTACACAGTGACCGAGTTCCAGCCTCTGTACTACGTGGCC

GAGAGCTTCAACGACGCCAAAGAAAAAGTGCGGACCTTCGCCGCCACCATTCCTCGGCCTTTTA

GCGTCAGATACGACCCCTACACACAGCGCGTGGAAGTGCTGGACAACACACAGCAGCTGAAGA

TTCTGGCCGACTCCATCAACAGCGAAGTGGGCATTCTGTGTCACGCCCTGCAGAAGATCAAGAG

CTGAGCAAGTAATGAGCGCTGATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGAC

TGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCA

TGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTAT

GAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCC

CCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTA

TTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGG

CACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTG

CCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTT

CCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAG

TCGGATCTCCCTTTGGGCCGCCTCCCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCA

TCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCC

TAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGG

TGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTG GGCTCTATGGCTTCTGAGGCGGAAAGAACGGGCTCGAGAAGCTTCTAGATATCCTCTCTTAAGG TAGCATCGAGATTTAAATTAGGGATAACAGGGTAATGGCGCGGGCCGCAAAATTTAAAAGAAG ACACCAAATCAGATGCCGCCGGTCGCCGCCGGTAGGCGGGACTTCCGGTACAAGATGGCGGAC AATTACGTCATTTCCTGTGACGTCATTTCCTGTGACGTCACTTCCGGTGGGCGGGACTTCCGGAA TTAGGGTTGGCTCTGGGCCAGCGCTTGGGGTTGACGTGCCACTAAGATCAAGCGGCGCGCCGCT TGTCTTAGTGTCAAGGCAACCCCAAGCAAGCTGGCCCAGAG

表10B：帶有GPVd162 ITR的GPV-PAH建構體（核苷酸1-4214；SEQ ID NO: 198）

描述	序列
5' ITR（SEQ ID NO: 183）	CGGTGACGTGTTTCCGGCTGTTAGGTTGACCACGCGCATGCCGCGCGGTCA GCCCAATAGTTAAGCCGGAAACACGTCACCGGAAGTCACATGACCGGAAG TCACGTGACCGGAAACACGTGACAGGAAGCACGTGACCGGAACTACGTCA CCGGATGTGCGTCACCGGAAGCATGTGACCGGAACTTGCGTCACTTCCCCC TCCCCTGATTGGCTGGTTCGAACGAACGAACCCTCCAATGAGACTCAAGGA CAAGAGGATATTTTGCGCGCCAGGAAGTG
CAGp 啟動子（SEQ ID NO: 195）	CTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATA TGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGC CCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAA CGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAA CTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTA TTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGA CCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTAT

	TACCATGCATGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCATCTC CCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTATTTTGTG CAGCGATGGGGGCGGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGCGGG GCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCC AATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTTATGGCGAGGCGGCGGCG GCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCG
合成內含子(SEQ ID NO: 192)	GTGAGCGGGCGGGACGGCCCTTCTCCTTCGGGCTGTAATTAGCGCTTGGTT TAATGACGGCTTGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTGAGGGGCTC CGGGAGGGCCCTTTGTGCGGGGGGAGCGGCTCGGGGCTGTCCGCGGGGGG ACGGCTGCCTTCGGGGGGGACGGGGCAGGGCGGGGTTCGGCTTCTGGCGT GTGACCGGCGGCTCTAGAGCCTCTGCTAACCATGTTCATGCCTTCTTCTTTT TCCTACAG
PAH序列（SEQ ID NO: 196）	ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGATATCGATTA CAAGGATGACGATGACAAGGCTGCTGTGGTTCTGGAAAATGGCGTGCTGA GCCGGAAGCTGAGCGACTTCGGACAAGAGACAAGCTACATCGAGGACAAC AGCAACCAGAATGGCGCCGTGTCTCTGATCTTCAGCCTGAAAGAAGAAGT GGGCGCCCTGGCCAAGGTGCTGAGACTGTTCGAGGAAAACGAGATCAATC TGACCCACATCGAGAGCAGACCCAGCAGACTGAACAAGGACGAGTACGAG TTCTTCACCTACCTGGACAAGCGGAGCAAGCCTGTGCTGGGCAGCATCATC AAGAGCCTGAGAAACGACATCGGCGCCACCGTGCACGAGCTGAGCAGAGA CAAAGAAAAGAACACCGTGCCATGGTTCCCCAGGACCATCCAAGAGCTGG ACAGATTCGCCAACCAGATCCTGAGCTATGGCGCCGAGCTGGACGCTGAT CACCCTGGCTTTAAGGACCCCGTGTACCGGGCCAGAAGAAAGCAGTTTGC CGATATCGCCTACAACTACCGGCACGGCCAGCCTATTCCTCGGGTCGAGTA

	CACCGAGGAAGAGAGAAAGACCTGGGGCACCGTGTTCAGAACCCTGAAGG CCCTGTACAAGACCCACGCCTGCTACGAGCACAACCACATCTTCCCACTGC TGGAAAAGTACTGCGGCTTCCGCGAGGACAATATCCCTCAGCTCGAAGAC GTGTCCCAGTTCCTGCAGACCTGCACCGGCTTTAGACTGAGGCCTGTTGCC GGACTGCTGAGCAGCAGAGATTTTCTCGGCGGCCTGGCCTTCAGAGTGTTC CACTGTACCCAGTACATCAGACACGGCAGCAAGCCCATGTACACCCCTGA GCCTGATATCTGCCACGAGCTGCTGGGACATGTGCCCCTGTTCAGCGATAG AAGCTTCGCCCAGTTCAGCCAAGAGATCGGACTGGCTTCTCTGGGAGCCCC TGACGAGTACATTGAGAAGCTGGCCACCATCTACTGGTTCACCGTGGAATT CGGCCTGTGCAAAGAGGGCGACAGCATCAAGGCTTATGGCGCTGGACTGC TGTCTAGCTTCGGCGAGCTGCAGTACTGTCTGAGCGACAAGCCTAAGCTGC TGCCCCTGGAACTGGAAAAGACCGCCTGCCAAGAGTACACAGTGACCGAG TTCCAGCCTCTGTACTACGTGGCCGAGAGCTTCAACGACGCCAAAGAAAA AGTGCGGACCTTCGCCGCCACCATTCCTCGGCCTTTTAGCGTCAGATACGA CCCCTACACACAGCGCGTGGAAGTGCTGGACAACACACAGCAGCTGAAGA TTCTGGCCGACTCCATCAACAGCGAAGTGGGCATTCTGTGTCACGCCCTGC AGAAGATCAAGAGCTGA
WPRE(突變的土撥鼠肝炎病毒轉錄後調控元件）（SEQ ID NO: 120）	TCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCT TAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTG TATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATC CTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGG CGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGC CACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCC ACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGG

	CTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTT CCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCT GCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCT GCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCG GATCTCCCTTTGGGCCGCCTCCCCGCTG
bGHpA（牛生長激素多聚腺苷酸化信號）(SEQ ID NO: 122)	CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCC TTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGA GGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGG GGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCAT GCTGGGGA
3' ITR反向末端重複體（SEQ ID NO: 184）	CACTTCCTGGCGCGCAAAATATCCTCTTGTCCTTGAGTCTCATTGGAGGGTT CGTTCGTTCGAACCAGCCAATCAGGGGAGGGGGAAGTGACGCAAGTTCCG GTCACATGCTTCCGGTGACGCACATCCGGTGACGTAGTTCCGGTCACGTGC TTCCTGTCACGTGTTTCCGGTCACGTGACTTCCGGTCATGTGACTTCCGGTG ACGTGTTTCCGGCTTAACTATTGGGCTGACCGCGCGGCATGCGCGTGGTCA ACCTAACAGCCGGAAACACGTCACCG
全長序列（SEQ ID NO: 198）	
CGGTGACGTGTTTCCGGCTGTTAGGTTGACCACGCGCATGCCGCGCGGTCAGCCCAATAGTTA AGCCGGAAACACGTCACCGGAAGTCACATGACCGGAAGTCACGTGACCGGAAACACGTGACA GGAAGCACGTGACCGGAACTACGTCACCGGATGTGCGTCACCGGAAGCATGTGACCGGAACTT GCGTCACTTCCCCCTCCCCTGATTGGCTGGTTCGAACGAACGAACCCTCCAATGAGACTCAAGG ACAAGAGGATATTTTGCGCGCCAGGAAGTGGCGGCAATTCAGTCGATAACTATAACGGTCCTA AGGTAGCGATTTAAATACGCGCTCTCTTAAGGTAGCCCCGGGACGCGTCAATTGAGATCTGGA TCCGGTACCGAATTCGCGGCCGCCTCGACGACTAGCGTTTAGTAATGAGACGCACAAACTAAT	

ATCACAAACTGGAAATGTCTATCAATATATAGTTGCTCTAGTTATTAATAGTAATCAATTACGG

GGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCC

TGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAAC

GCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGC

AGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCC

CGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTA

TTAGTCATCGCTATTACCATGCATGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCATCT

CCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTATTTTGTGCAGCGATGGGG

GCGGGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGG

CGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTTATGG

CGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGCG

CGCTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCGCGCCGCCCGCCCCGGCTCTGACTGA

CCGCGTTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTTCGGGCTGTAATTAGCGCTT

GGTTTAATGACGGCTTGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTGAGGGGCTCCGGGAGGGC

CCTTTGTGCGGGGGGAGCGGCTCGGGGCTGTCCGCGGGGGGACGGCTGCCTTCGGGGGGGACG

GGGCAGGGCGGGGTTCGGCTTCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTAACCATGT

TCATGCCTTCTTCTTTTTCCTACAGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATT

TTGGCAAAGAATTGGATCGCGAAGCCGCCACCATGGACTACAAAGACCATGACGGTGATTATA

AAGATCATGATATCGATTACAAGGATGACGATGACAAGGCTGCTGTGGTTCTGGAAAATGGCG

TGCTGAGCCGGAAGCTGAGCGACTTCGGACAAGAGACAAGCTACATCGAGGACAACAGCAAC

CAGAATGGCGCCGTGTCTCTGATCTTCAGCCTGAAAGAAGAAGTGGGCGCCCTGGCCAAGGTG

CTGAGACTGTTCGAGGAAAACGAGATCAATCTGACCCACATCGAGAGCAGACCCAGCAGACT

GAACAAGGACGAGTACGAGTTCTTCACCTACCTGGACAAGCGGAGCAAGCCTGTGCTGGGCAG

CATCATCAAGAGCCTGAGAAACGACATCGGCGCCACCGTGCACGAGCTGAGCAGAGACAAAG

AAAAGAACACCGTGCCATGGTTCCCCAGGACCATCCAAGAGCTGGACAGATTCGCCAACCAGA

TCCTGAGCTATGGCGCCGAGCTGGACGCTGATCACCCTGGCTTTAAGGACCCCGTGTACCGGG

CCAGAAGAAAGCAGTTTGCCGATATCGCCTACAACTACCGGCACGGCCAGCCTATTCCTCGGG

TCGAGTACACCGAGGAAGAGAGAAAGACCTGGGGCACCGTGTTCAGAACCCTGAAGGCCCTG

TACAAGACCCACGCCTGCTACGAGCACAACCACATCTTCCCACTGCTGGAAAAGTACTGCGGC

TTCCGCGAGGACAATATCCCTCAGCTCGAAGACGTGTCCCAGTTCCTGCAGACCTGCACCGGCT

TTAGACTGAGGCCTGTTGCCGGACTGCTGAGCAGCAGAGATTTTCTCGGCGGCCTGGCCTTCAG

AGTGTTCCACTGTACCCAGTACATCAGACACGGCAGCAAGCCCATGTACACCCCTGAGCCTGA

TATCTGCCACGAGCTGCTGGGACATGTGCCCCTGTTCAGCGATAGAAGCTTCGCCCAGTTCAGC

CAAGAGATCGGACTGGCTTCTCTGGGAGCCCCTGACGAGTACATTGAGAAGCTGGCCACCATC

TACTGGTTCACCGTGGAATTCGGCCTGTGCAAAGAGGGCGACAGCATCAAGGCTTATGGCGCT

GGACTGCTGTCTAGCTTCGGCGAGCTGCAGTACTGTCTGAGCGACAAGCCTAAGCTGCTGCCC

CTGGAACTGGAAAAGACCGCCTGCCAAGAGTACACAGTGACCGAGTTCCAGCCTCTGTACTAC

GTGGCCGAGAGCTTCAACGACGCCAAAGAAAAAGTGCGGACCTTCGCCGCCACCATTCCTCGG

CCTTTTAGCGTCAGATACGACCCCTACACACAGCGCGTGGAAGTGCTGGACAACACACAGCAG

CTGAAGATTCTGGCCGACTCCATCAACAGCGAAGTGGGCATTCTGTGTCACGCCCTGCAGAAG

ATCAAGAGCTGAGCAAGTAATGAGCGCTGATCATAATCAACCTCTGGATTACAAAATTTGTGA

AAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGC

CTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGC

TGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGC

TGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCT

TTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGG

CTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTTCCTTGGCT

GCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTC

AATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCC

TTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCTGATCAGCCTCGACTGTGC

CTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCC

ACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATT

CTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGG

CATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACGGGCTCGAGAAGCTTC

TAGATATCCTCTCTTAAGGTAGCATCGAGATTTAAATTAGGGATAACAGGGTAATGGCGCGGG

CCGCCACTTCCTGGCGCGCAAAATATCCTCTTGTCCTTGAGTCTCATTGGAGGGTTCGTTCGTTC

GAACCAGCCAATCAGGGGAGGGGGAAGTGACGCAAGTTCCGGTCACATGCTTCCGGTGACGC

ACATCCGGTGACGTAGTTCCGGTCACGTGCTTCCTGTCACGTGTTTCCGGTCACGTGACTTCCG

GTCATGTGACTTCCGGTGACGTGTTTCCGGCTTAACTATTGGGCTGACCGCGCGGCATGCGCGT

GGTCAACCTAACAGCCGGAAACACGTCACCG

實例7b.帶有B19或GPV來源的ITR的ssDNA報告子基因建構體的製備。

【0563】 將如實例 1c 中所述製備 ssDNA 報告子基因/PAH 建構體。簡單來說，將用 LguI 消化質體。具有形成的髮夾 ITR 結構的 ssDNA 片段將通過以下方式來產生：使 LguI 消化的雙鏈 DNA 片段產物（報告子基因表現盒和質體骨架）在 95°C 下變性，然後在 4°C 下冷卻以允許回文 ITR 序列折疊（圖 1A）。將在小鼠體內測試所得 ssDNA 建構體建立肝臟、肌肉組織、眼睛中的光感受器和中樞神經系統（CNS）的持久轉導的能力。

實例7c. ssDNA介導的報告子基因表現的體內評價。

【0564】 為了驗證實例 7b 中所述的 ssDNA 報告子基因建構體在體內介導持久轉基因表現的能力，將用 5、10 或 20 μg/小鼠的報告子基因 ssDNA 向 5-12 周齡小鼠（4 隻動物/組）全身注射，局部注射至靶肌肉組織和 CNS 細胞，和/或視網膜下注射至靶感光細胞。

【0565】 為了評價來自基於 B19 ITR 和 GPV ITR 的表現建構體的 PAH，將使用相關疾病的小鼠模型。這些基因建構體將通過 HDI 全身遞送以靶向肝臟。

【符號說明】

無

【序列表】

<110> 美商百歐維拉提夫治療公司(BIOVERATIV THERAPEUTICS INC.)

<120> 核酸分子及其用於非病毒基因療法的用途

<130> 615112: SA9-465TW

<140> 108128641
<141> 2019-08-12

<150> US 62/716,826
<151> 2018-08-09

<160> 198

<170> PatentIn version 3.5

<210> 1
<211> 4374
<212> DNA
<213> 人工序列

<220>
<223> coFVIII-5

<400> 1
atgcaaatcg aactgagcac ctgtttcttc ctctgcctgc tgagattctg tttctccgcg 60
acccgccgat actacctggg agcagtggag ctctcctggg attacatgca gagcgacctt 120
ggggagctgc ccgtggatgc caggttccct ccccgggtgc caaagtcgtt tccgttcaac 180
acctccgtgg tgtacaagaa aactctgttc gtggagttca ccgaccacct gttcaatatc 240
gccaagccca gacctccctg gatggggctg ttgggaccta ccatccaagc ggaggtgtac 300
gacactgtgg tcatcactct gaagaacatg gcctcgcatc ccgtgtccct gcacgccgtg 360
ggagtgtctt actggaaagc gtccgagggg gccgaatacg acgaccagac ctcgcagaga 420
gaaaaggaag atgacaaggt gttcccagga ggatcgcaca cctacgtgtg gcaagtgttg 480
aaggagaacg gcccaatggc ctccgacccg ctgtgcctga cctactcgta cctgtcccac 540

gtggacctcg tgaaggacct caactcggga ctgattggag ccctgctggt ctgcagggaa 600

ggctcactgg cgaaagaaaa gactcagacc ttgcacaagt tcattctgct gttcgctgtg 660

ttcgacgagg ggaagtcgtg gcacagcgag actaagaact ccctgatgca agatagagat 720

gccgcctccg cccgggcctg gcctaagatg cacaccgtga acggttacgt gaaccgctcc 780

ctccctggcc tgattggatg ccaccggaag tccgtgtact ggcacgtgat cgggatgggg 840

accacccccg aggtgcacag catcttcctg gaaggtcaca catttctcgt gcgcaaccac 900

cggcaggcct ccctggaaat cagccccatt accttcctca ctgcccagac tctgctgatg 960

gacctgggac agttcctgct gttctgccat atctcctccc accaacatga cggaatggag 1020

gcatacgtga aggtcgattc ctgccctgag gaaccccagc tccgcatgaa gaacaatgag 1080

gaagccgagg actacgacga cgacctgacg gatagcgaga tggatgtggt ccggttcgat 1140

gacgataaca gcccttcctt catccaaatt cgctcggtgg caaagaagca ccccaagacc 1200

tgggtgcatt acattgcggc ggaagaagag gactgggatt atgccccgct tgtcctcgct 1260

cctgacgacc ggagctacaa gagccagtac ctgaacaacg gtccacagag gatcggtaga 1320

aagtacaaga aggtccgctt catggcctat accgacgaaa ccttcaaaac tagagaggcc 1380

atccaacacg aatccggcat cctgggcccg ctcttgtacg gagaagtcgg cgacaccctt 1440

ctcattatct tcaagaacca ggcttcccgg ccgtacaaca tctatccgca tgggatcact 1500

gacgtgcgcc cactgtactc gcggcgcctg cccaagggtg tcaaacacct gaaggatttt 1560

ccgatccttc cgggagaaat cttcaagtac aagtggaccg tgaccgtgga agatggccca 1620

actaagtctg accctagatg cctcacccgc tactactcat ccttcgtcaa catggagcgc 1680

gacctggcca gcggactgat cggcccgctg ctgatttgct acaaggaatc agtggaccaa 1740

cggggaaacc agatcatgtc ggataagagg aacgtcatcc tcttctccgt gtttgacgaa 1800

aaccggtcgt ggtacctgac tgaaaacatc cagcggttcc tccccaaccc cgcgggcgtg 1860

cagctggaag atcctgagtt tcaggcatca aacatcatgc actccattaa cggctacgtg 1920

ttcgattcgc tgcagctgag cgtgtgtctg cacgaagtgg cctactggta catcctgtcc 1980

attggtgccc agactgactt cctgtccgtg tttttctccg gctacacgtt caagcacaag 2040

atggtgtacg aggacaccct gaccctcttc ccttttttccg gcgaaactgt gtttatgagc 2100

atggagaatc ccggcctgtg gatcttgggc tgccacaaca gcgacttccg taacagagga 2160

atgactgcgc tgctcaaggt gtccagctgc gacaagaaca ccggagacta ttatgaggac 2220

tcatacgagg acatctccgc ctacctcctg tccaagaata acgccattga acctcggagc 2280

ttcagccaga acccacccgt gcttaagaga catcaacggg agatcactag gaccaccctg 2340

cagtcagacc aggaggaaat cgactacgat gacaccatct cggtcgagat gaagaaggag 2400

gactttgaca tctacgacga agatgaaaac cagagcccga ggtcgttcca aaagaaaacc 2460

cgccactact ttattgctgc tgtcgagcgg ctgtgggact acggaatgtc gtcctcgccg 2520

cacgtgctcc gcaaccgagc ccagagcggc tcggtgccgc aattcaagaa ggtcgtgttc 2580

caggagttca ctgacgggag cttcactcag cctttgtacc ggggagaact caatgaacat 2640

ctcggcctcc tcggacctta catcagagca gaagtggaag ataacatcat ggtcactttc 2700

cgtaaccaag ccagccgccc gtactcgttc tactcctccc tcatttctta cgaagaggac 2760
cagcggcagg gcgcagaacc gcgcaagaac ttcgtgaagc ccaacgaaac caagacctac 2820
ttctggaaag tgcagcatca tatggccccg actaaggacg agtttgactg caaagcctgg 2880
gcctacttct ccgatgtgga cttggagaag gacgtccact ccggcctcat cggtcccctg 2940
ctcgtgtgcc ataccaatac cctgaacccc gcacacggtc gccaggtcac cgtgcaggag 3000
ttcgctctgt tcttcactat cttcgacgaa actaagtcct ggtacttcac cgagaacatg 3060
gagaggaact gcagagcccc ctgtaacatc cagatggagg acccgacgtt caaggaaaac 3120
taccggttcc acgccattaa cggatacatc atggatacgc tgccgggtct tgtgatggcc 3180
caggatcaac ggatcagatg gtacttattg tcgatgggca gcaacgagaa catccactct 3240
attcacttct ccggtcatgt gttcactgtg cggaagaagg aagagtacaa gatggccctg 3300
tacaaccttt atcccggagt gttcgaaact gtggaaatgc tgccgtcgaa ggccggcatt 3360
tggcgcgtgg agtgtttgat tggagaacat ctccatgcgg ggatgtcaac cctgttcctg 3420
gtgtatagca acaagtgcca gactccgctt gggatggcgt caggacacat tagggatttc 3480
cagatcactg cgtccggcca gtacggccaa tgggccccta agctggcccg cctgcattac 3540
tccggatcca ttaacgcctg gtcaaccaag gagccattct cctggatcaa ggtggacctt 3600
ctggccccca tgattatcca cggaattaag acccaggggg cccggcagaa gttctcctca 3660
ctgtacatca gccagttcat aatcatgtac tccctggacg gaaagaagtg gcaaacctac 3720
aggggaaca gcaccggcac actgatggtc tttttcggaa atgtggactc ctccgggatt 3780
aagcataaca tcttcaaccc tccgattatc gctcggtaca ttagacttca ccctacccac 3840
tacagcattc gctccaccct gcggatggaa ctgatgggct gcgatctgaa ctcgtgcagc 3900
atgccgttgg gaatggagtc caaagcaatt tccgacgcgc agatcaccgc ctcgtcctac 3960
tttaccaaca tgttcgccac gtggtcaccg tccaaggccc ggctgcacct ccagggaaga 4020
tccaacgcat ggcggccaca ggtcaacaac cctaaggagt ggctccaggt ggacttccag 4080
aaaaccatga aggtcaccgg agtcacaacc cagggagtga agtcgctgct gacttctatg 4140
tacgtcaagg agttcctgat ctccagcagc caggacgggc accagtggac cctgttcttc 4200
caaaatggaa aggtcaaggt gtttcagggc aatcaggatt cattcacccc ggtggtgaac 4260
tcccttgatc caccctcct gacccgctac cttcgcatcc acccacagtc ctgggtgcac 4320
cagatcgcgc tgaggatgga ggtcctggga tgcgaagccc aggacctgta ctga 4374

<210> 2
<211> 4374
<212> DNA
<213> 人工序列

<220>

<223> coFVIII-4

<400> 2

atgcagatcg agctgagcac gtgcttcttc ctgtgcctgc tgaggttctg cttcagcgcc 60
accaggaggt actacctggg cgccgtggag ctgagctggg actacatgca gagcgacctg 120
ggcgagctgc ccgtggacgc caggttcccc cccagggtgc ccaagagctt ccccttcaac 180
acgagcgtgg tgtacaagaa gaccctgttc gtggagttca ccgaccatct gttcaatatc 240
gccaagccca ggcccccctg gatggggctg ctggggccca cgatccaggc cgaggtgtac 300
gacaccgtgg tcatcaccct gaagaacatg gccagccacc ccgtgagcct gcacgccgtg 360
ggcgtgagct actggaaggc cagcgagggc gccgagtacg acgaccagac cagccagagg 420
gagaaggagg acgacaaggt gttccccggc ggcagccaca cctacgtgtg gcaggtgctg 480
aaggagaatg ggcccatggc cagcgacccc ctgtgcctga cctactctta cctgagccac 540
gtggatctgg tgaaggacct gaacagcggc ctgatcggcg ccctgctggt gtgcagggag 600
ggcagcctgg ccaaggagaa gacccagacc ctgcacaagt tcatcctgct gttcgccgtg 660
ttcgacgagg gcaagagctg gcacagcgag accaagaaca gcctgatgca ggatagggac 720
gccgccagcg ccagggcctg gcccaagatg cacaccgtga acggctacgt gaacaggtct 780
ctgcccggcc tgatcggctg ccacaggaag agcgtgtact ggcacgtgat cggcatgggg 840
accacccccg aggtgcacag catcttcctg gagggccaca cgttcctggt gaggaatcac 900
aggcaggcca gcctggagat cagcccgatc accttcctga ccgcccagac cctgctgatg 960
gacctggggc agttcctgct gttctgccat atcagctctc accagcacga cggcatggag 1020
gcctacgtga aggtggatag ctgccccgag gagccccagc tgaggatgaa gaacaacgag 1080
gaggccgagg actacgacga cgacctgacc gacagcgaga tggacgtggt gaggttcgac 1140
gacgacaata gcccgagctt catccagatc aggagcgtgg ccaagaagca ccccaagacc 1200
tgggtgcatt acatcgccgc cgaggaggag gattgggact acgccccccct ggtgctggcc 1260
cccgacgaca ggtcttacaa gagccagtac ctgaacaacg ggccccagag gatcggcagg 1320
aagtacaaga aggtgaggtt catggcctac accgacgaga ccttcaagac cagggaggcg 1380
atccagcacg agagcgggat cctggggccc ctgctgtacg gcgaggtggg cgacacgctg 1440
ctgatcatct tcaagaacca ggccagcagg ccgtacaata tctaccccca cgggatcacc 1500
gacgtgaggc ccctgtactc taggaggctg cccaagggcg tgaagcacct gaaggacttc 1560
cccatcctgc ccggcgagat cttcaagtac aagtggaccg tgaccgtgga ggacgggccc 1620
acgaagagcg accccaggtg cctgaccagg tactacagct ctttcgtgaa catggagagg 1680
gacctggcca gcggcctgat cgggcccctg ctgatctgct acaaggagag cgtggatcag 1740
aggggcaacc agatcatgag cgacaagagg aacgtgatcc tgttcagcgt gttcgacgag 1800
aataggtctt ggtacctgac cgagaatatc cagaggttcc tgcccaaccc cgccggcgtg 1860
cagctggagg atcccgagtt ccaggccagc aacatcatgc acagcatcaa cggctacgtg 1920
ttcgacagcc tgcagctgag cgtgtgcctg cacgaggtgg cctactggta catcctgagc 1980

atcggcgccc agaccgactt cctgagcgtg ttcttcagcg gctacacctt caagcacaag 2040

atggtgtacg aggataccct gaccctgttc cccttcagcg gcgagaccgt gttcatgagc 2100

atggagaacc ccggcctgtg gatcctgggc tgccataact ccgacttcag gaataggggc 2160

atgaccgccc tgctgaaggt gagctcttgc gacaagaaca ccggcgacta ctacgaggat 2220

agctacgagg atatcagcgc ctacctgctg agcaagaaca acgccatcga gcccaggtct 2280

ttcagccaga acccccccgt gctgaagagg caccagaggg agatcaccag gacgaccctg 2340

cagagcgacc aggaggagat cgactacgac gacacgatca gcgtggagat gaagaaggag 2400

gatttcgaca tctacgacga ggacgagaat cagagcccca ggtctttcca gaagaagacc 2460

aggcattact tcatcgccgc cgtggagagg ctgtgggact acggcatgag cagctctccc 2520

cacgtgctga ggaatagggc ccagagcggc agcgtgcccc agttcaagaa ggtggtgttc 2580

caggagttca ccgacggcag cttcacccag cccctgtaca ggggcgagct gaacgagcac 2640

ctgggcctgc tggggcccta catcagggcc gaggtggagg ataacatcat ggtgaccttc 2700

aggaatcagg ccagcaggcc ctatagcttc tatagctctc tgatcagcta cgaggaggat 2760

cagaggcagg gcgccgagcc caggaagaac ttcgtgaagc ccaacgagac caagacctac 2820

ttctggaagg tgcagcacca catggccccc acgaaggacg agttcgactg caaggcctgg 2880

gcctacttca gcgacgtgga tctggagaag gacgtgcaca gcggcctgat cgggcccctg 2940

ctggtgtgcc acaccaacac cctgaacccc gcccacggca ggcaggtgac cgtgcaggag 3000

ttcgccctgt tcttcaccat cttcgacgag accaagagct ggtacttcac cgagaatatg 3060

gagaggaatt gcagggcccc ctgcaatatc cagatggagg acccgacctt caaggagaat 3120

tacaggttcc acgccatcaa cggctacatc atggacacgc tgcccggcct ggtcatggcc 3180

caggatcaga ggatcaggtg gtatctgctg agcatgggga gcaacgagaa tatccacagc 3240

atccacttca gcggccacgt gttcaccgtg aggaagaagg aggagtacaa gatggccctg 3300

tacaatctgt accccggcgt gttcgagacc gtggagatgc tgcccagcaa ggccgggatc 3360

tggagggtgg agtgcctgat cggcgagcac ctgcacgccg gcatgagcac gctgttcctg 3420

gtgtactcta acaagtgcca gacccccctg gggatggcca gcggccacat cagggacttc 3480

cagatcaccg ccagcggcca gtacggccag tgggccccca agctggccag gctgcactat 3540

tccggaagca tcaacgcctg gagcacgaag gagcccttca gctggatcaa ggtggatctg 3600

ctggccccca tgatcatcca cgggatcaag acccagggcg ccaggcagaa gttcagctct 3660

ctgtatatca gccagttcat catcatgtac tctctggacg gcaagaagtg gcagacctac 3720

aggggcaaca gcaccggcac gctgatggtg ttcttcggca acgtggactc tagcgggatc 3780

aagcacaata tcttcaaccc ccccatcatc gccaggtaca tcaggctgca ccccacccat 3840

tactctatca ggtctaccct gaggatggag ctgatgggct gcgacctgaa cagctgcagc 3900

atgcccctgg ggatggagag caaggccatc agcgacgccc agatcaccgc cagctcttac 3960

ttcaccaaca tgttcgccac ctggagcccg agcaaggcca ggctgcacct gcagggcagg 4020

tctaacgcct ggaggcccca ggtgaacaac cccaaggagt ggctgcaggt ggatttccag 4080

aagaccatga aggtgaccgg cgtgaccacg cagggcgtga agagcctgct gaccagcatg 4140

tacgtgaagg agttcctgat cagctctagc caggacggcc accagtggac cctgttcttc 4200
cagaacggca aggtgaaggt gttccagggc aaccaggata gcttcacccc cgtggtgaac 4260
agcctggacc ccccccctgct gaccaggtat ctgaggatcc acccccagag ctgggtgcac 4320
cagatcgccc tgaggatgga ggtgctgggc tgcgaggccc aggatctgta ttga 4374

<210> 3
<211> 4374
<212> DNA
<213> 人工序列

<220>
<223> coFVIII-52

<400> 3
atgcaaatcg aactgagcac ctgtttcttc ctctgcctgc tgagattctg tttctccgcg 60
acccgccgat actacctggg agcagtggag ctctcctggg attacatgca gagcgacctt 120
ggggagctgc ccgtggatgc caggttccct ccccgggtgc caaagtcgtt tccgttcaac 180
acctccgtgg tgtacaagaa aactctgttc gtggagttca ccgaccacct gttcaatatc 240
gccaagccca gacctccctg gatggggctg ttgggaccta ccatccaagc ggaggtgtac 300
gacactgtgg tcatcactct gaagaacatg gcctcgcatc ccgtgtccct gcacgccgtg 360
ggagtgtctt actggaaagc gtccgagggg gccgaatacg acgaccagac ctcgcagaga 420
gaaaaggaag atgacaaggt gttcccagga ggatcgcaca cctacgtgtg gcaagtgttg 480
aaggagaacg gcccaatggc ctccgacccg ctgtgcctga cctactcgta cctgtcccac 540
gtggacctcg tgaaggacct caactcggga ctgattggag ccctgctggt ctgcagggaa 600
ggctcactgg cgaaagaaaa gactcagacc ttgcacaagt tcattctgct gttcgctgtg 660
ttcgacgagg ggaagtcgtg gcacagcgag actaagaact ccctgatgca agatagagat 720
gccgcctccg cccgggcctg gcctaagatg cacaccgtga acggttacgt gaaccgctcc 780
ctccctggcc tgattggatg ccaccggaag tccgtgtact ggcacgtgat cgggatgggg 840
accacccccg aggtgcacag catcttcctg gaaggtcaca catttctcgt gcgcaaccac 900
cggcaggcct ccctggaaat cagccccatt accttcctca ctgcccagac tctgctgatg 960
gacctgggac agttcctgct gttctgccat atctcctccc accaacatga cggaatggag 1020
gcatacgtga aggtcgattc ctgccctgag gaaccccagc tccgcatgaa gaacaatgag 1080
gaagccgagg actacgacga cgacctgacg gatagcgaga tggatgtggt ccggttcgat 1140
gacgataaca gcccttcctt catccaaatt cgctcggtgg caaagaagca ccccaagacc 1200
tgggtgcatt acattgcggc ggaagaagag gactgggatt atgccccgct tgtcctcgct 1260

cctgacgacc ggagctacaa gagccagtac ctgaacaacg gtccacagag gatcggtaga 1320

aagtacaaga aggtccgctt catggcctat accgacgaaa ccttcaaaac tagagaggcc 1380

atccaacacg aatccggcat cctgggcccg ctcttgtacg gagaagtcgg cgacaccctt 1440

ctcattatct tcaagaacca ggcttcccgg ccgtacaaca tctatccgca tgggatcact 1500

gacgtgcgcc cactgtactc gcggcgcctg cccaagggtg tcaaacacct gaaggatttt 1560

ccgatccttc cgggagaaat cttcaagtac aagtggaccg tgaccgtgga agatggccca 1620

actaagtctg accctagatg cctcacccgc tactactcat ccttcgtcaa catggagcgc 1680

gacctggcca gcggactgat cggcccgctg ctgatttgct acaaggaatc agtggaccaa 1740

cggggaaacc agatcatgtc ggataagagg aacgtcatcc tcttctccgt gtttgacgaa 1800

aaccggtcgt ggtacctgac cgagaacatc cagaggttcc tgcccaaccc tgctggggtg 1860

cagctggagg accccgagtt ccaggccagc aacatcatgc acagcatcaa tggctacgtg 1920

ttcgacagcc tgcagctgag cgtgtgcctg cacgaggtgg cctactggta catcctgagc 1980

atcggcgccc agaccgactt cctgagcgtg ttcttctctg gctacacctt caagcacaag 2040

atggtgtatg aggacaccct gaccctgttc cccttcagcg gggagactgt cttcatgagc 2100

atggagaacc ctggcctgtg gatcctgggc tgccacaaca gcgacttcag gaacaggggc 2160

atgactgccc tgctgaaagt ctccagctgt gacaagaaca ccggggacta ctacgaggac 2220

agctacgagg acatcagcgc ctacctgctg agcaagaaca atgccatcga gcccaggagc 2280

ttctctcaga acccccccagt gctgaagagg caccagaggg agatcaccag gaccaccctg 2340

cagtctgacc aggaggagat cgactatgat gacaccatca gcgtggagat gaagaaggag 2400

gacttcgaca tctacgacga ggacgagaac cagagcccca ggagcttcca gaagaagacc 2460

aggcactact tcattgctgc tgtggagagg ctgtgggact atggcatgtc cagcagcccc 2520

catgtgctga ggaacagggc ccagtctggc agcgtgcccc agttcaagaa agtcgtgttc 2580

caggagttca ccgacggcag cttcacccag cccctgtaca gaggggagct gaacgagcac 2640

ctgggcctgc tgggccccta catcagggcc gaggtggagg acaacatcat ggtgaccttc 2700

aggaaccagg ccagcaggcc ctacagcttc tacagcagcc tgatcagcta cgaggaggac 2760

cagaggcagg gggctgagcc caggaagaac tttgtgaagc ccaatgaaac caagacctac 2820

ttctggaagg tgcagcacca catggccccc accaaggacg agttcgactg caaggcctgg 2880

gcctacttct ctgacgtgga cctggagaag gacgtgcact ctggcctgat tggcccctg 2940

ctggtgtgcc acaccaacac cctgaaccct gcccatggca ggcaggtgac tgtgcaggag 3000

ttcgccctgt tcttcaccat cttcgatgaa accaagagct ggtacttcac tgagaacatg 3060

gagaggaact gcagggcccc ctgcaacatc cagatggagg accccacctt caaggagaac 3120

tacaggttcc atgccatcaa tggctacatc atggacaccc tgcctggcct ggtcatggcc 3180

caggaccaga ggatcaggtg gtatctgctg agcatgggca gcaacgagaa catccacagc 3240

atccacttct ctggccacgt gttcactgtg aggaagaagg aggagtacaa gatggccctg 3300

tacaacctgt accctggggt gttcgaaacc gtggagatgc tgcccagcaa ggccggcatc 3360

tggagggtgg agtgcctgat tggggagcac ctgcacgccg gcatgagcac cctgttcctg 3420

gtgtacagca acaagtgcca gacccccctg ggcatggcct ctggccacat cagggacttc 3480
cagatcactg cctctggcca gtacggccag tgggccccca agctggccag gctgcactac 3540
tccggaagca tcaatgcctg gagcaccaag gagcccttca gctggatcaa agtggacctg 3600
ctggccccca tgatcatcca cggcatcaag acccaggggg ccaggcagaa gttctccagc 3660
ctgtacatca gccagttcat catcatgtac agcctggacg gcaagaagtg gcagacctac 3720
aggggcaaca gcaccggcac cctgatggtg ttcttcggca acgtggacag cagcggcatc 3780
aagcacaaca tcttcaaccc ccccatcatc gccagataca tcaggctgca ccccacccac 3840
tacagcatca ggagcaccct gaggatggag ctgatgggct gtgacctgaa cagctgcagc 3900
atgcccctgg gcatggagag caaggccatc tctgacgccc agatcactgc ctccagctac 3960
ttcaccaaca tgtttgccac ctggagcccc agcaaggcca ggctgcacct gcagggcagg 4020
agcaatgcct ggaggcccca ggtcaacaac cccaaggagt ggctgcaggt ggacttccag 4080
aagaccatga aggtgactgg ggtgaccacc caggggggtga agagcctgct gaccagcatg 4140
tacgtgaagg agttcctgat ctccagcagc caggacggcc accagtggac cctgttcttc 4200
cagaatggca aggtgaaggt gttccagggc aaccaggaca gcttcacccc tgtggtcaac 4260
agcctggacc ccccccctgct gaccagatac ctgaggatcc acccccagag ctgggtgcac 4320
cagatcgccc tgaggatgga ggtgctgggc tgtgaggccc aggacctgta ctga 4374

<210> 4
<211> 4374
<212> DNA
<213> 人工序列

<220>
<223> coFVIII-62

<400> 4
atgcagattg agctgtccac ttgtttcttc ctgtgcctcc tgcgcttctg tttctccgcc 60
actcgccggt actaccttgg agccgtggag ctttcatggg actacatgca gagcgacctg 120
ggcgaactcc ccgtggatgc cagattcccc ccccgcgtgc caaagtcctt ccccctttaac 180
acctccgtgg tgtacaagaa aaccctcttt gtcgagttca ctgaccacct gttcaacatc 240
gccaagccgc gcccaccttg gatgggcctc ctgggaccga ccattcaagc tgaagtgtac 300
gacaccgtgg tgatcaccct gaagaacatg gcgtcccacc ccgtgtccct gcatgcggtc 360
ggagtgtcct actggaaggc ctccgaagga gctgagtacg acgaccagac tagccagcgg 420
gaaaaggagg acgataaagt gttcccgggc ggctcgcata cttacgtgtg gcaagtcctg 480
aaggaaaacg gacctatggc atccgatcct ctgtgcctga cttactccta cctttcccat 540

gtggacctcg tgaaggacct gaacagcggg ctgattggtg cacttctcgt gtgccgcgaa 600

ggttcgctcg ctaaggaaaa gacccagacc ctccataagt tcatcctttt gttcgctgtg 660

ttcgatgaag gaaagtcatg gcattccgaa actaagaact cgctgatgca ggaccgggat 720

gccgcctcag cccgcgcctg gcctaaaatg catacagtca acggatacgt gaatcggtca 780

ctgcccgggc tcatcggttg tcacagaaag tccgtgtact ggcacgtcat cggcatgggc 840

actacgcctg aagtgcactc catcttcctg gaagggcaca ccttcctcgt gcgcaaccac 900

cgccaggcct ctctggaaat ctccccgatt acctttctga ccgcccagac tctgctcatg 960

gacctggggc agttccttct cttctgccac atctccagcc atcagcacga cggaatggag 1020

gcctacgtga aggtggactc atgcccggaa gaacctcagt tgcggatgaa gaacaacgag 1080

gaggccgagg actatgacga cgatttgact gactccgaga tggacgtcgt gcggttcgat 1140

gacgacaaca gccccagctt catccagatt cgcagcgtgg ccaagaagca ccccaaaacc 1200

tgggtgcact acatcgcggc cgaggaagaa gattgggact acgccccgtt ggtgctggca 1260

cccgatgacc ggtcgtacaa gtcccagtat ctgaacaatg gtccgcagcg gattggcaga 1320

aagtacaaga aagtgcggtt catggcgtac actgacgaaa cgtttaagac ccgggaggcc 1380

attcaacatg agagcggcat tctgggacca ctgctgtacg gagaggtcgg cgataccctg 1440

ctcatcatct tcaaaaacca ggcctcccgg ccttacaaca tctaccctca cggaatcacc 1500

gacgtgcggc cactctactc gcggcgcctg ccgaagggcg tcaagcacct gaaagacttc 1560

cctatcctgc cgggcgaaat cttcaagtat aagtggaccg tcaccgtgga ggacgggccc 1620

accaagagcg atcctaggtg tctgactcgg tactactcca gcttcgtgaa catggaacgg 1680

gacctggcat cgggactcat tggaccgctg ctgatctgct acaaagagtc ggtggatcaa 1740

cgcggcaacc agatcatgtc cgacaagcgc aacgtgatcc tgttctccgt gtttgatgaa 1800

aacagatcct ggtacctgac cgagaacatc cagaggttcc tgcccaaccc tgctggggtg 1860

cagctggagg accccgagtt ccaggccagc aacatcatgc acagcatcaa tggctacgtg 1920

ttcgacagcc tgcagctgag cgtgtgcctg cacgaggtgg cctactggta catcctgagc 1980

atcggcgccc agaccgactt cctgagcgtg ttcttctctg gctacacctt caagcacaag 2040

atggtgtatg aggacaccct gaccctgttc cccttcagcg gggagactgt cttcatgagc 2100

atggagaacc ctggcctgtg gatcctgggc tgccacaaca gcgacttcag gaacaggggc 2160

atgactgccc tgctgaaagt ctccagctgt gacaagaaca ccggggacta ctacgaggac 2220

agctacgagg acatcagcgc ctacctgctg agcaagaaca atgccatcga gcccaggagc 2280

ttctctcaga acccccccagt gctgaagagg caccagaggg agatcaccag gaccaccctg 2340

cagtctgacc aggaggagat cgactatgat gacaccatca gcgtggagat gaagaaggag 2400

gacttcgaca tctacgacga ggacgagaac cagagcccca ggagcttcca gaagaagacc 2460

aggcactact tcattgctgc tgtggagagg ctgtgggact atggcatgtc cagcagcccc 2520

catgtgctga ggaacagggc ccagtctggc agcgtgcccc agttcaagaa agtcgtgttc 2580

caggagttca ccgacggcag cttcacccag cccctgtaca gaggggagct gaacgagcac 2640

ctgggcctgc tgggccccta catcagggcc gaggtggagg acaacatcat ggtgaccttc 2700

aggaaccagg ccagcaggcc ctacagcttc tacagcagcc tgatcagcta cgaggaggac 2760
cagaggcagg gggctgagcc caggaagaac tttgtgaagc ccaatgaaac caagacctac 2820
ttctggaagg tgcagcacca catggccccc accaaggacg agttcgactg caaggcctgg 2880
gcctacttct ctgacgtgga cctggagaag gacgtgcact ctggcctgat tggccccctg 2940
ctggtgtgcc acaccaacac cctgaaccct gcccatggca ggcaggtgac tgtgcaggag 3000
ttcgccctgt tcttcaccat cttcgatgaa accaagagct ggtacttcac tgagaacatg 3060
gagaggaact gcagggcccc ctgcaacatc cagatggagg accccacctt caaggagaac 3120
tacaggttcc atgccatcaa tggctacatc atggacaccc tgcctggcct ggtcatggcc 3180
caggaccaga ggatcaggtg gtatctgctg agcatgggca gcaacgagaa catccacagc 3240
atccacttct ctggccacgt gttcactgtg aggaagaagg aggagtacaa gatggccctg 3300
tacaacctgt accctggggt gttcgaaacc gtggagatgc tgcccagcaa ggccggcatc 3360
tggagggtgg agtgcctgat tggggagcac ctgcacgccg gcatgagcac cctgttcctg 3420
gtgtacagca acaagtgcca gacccccctg ggcatggcct ctggccacat cagggacttc 3480
cagatcactg cctctggcca gtacggccag tgggccccca agctggccag gctgcactac 3540
tccggaagca tcaatgcctg gagcaccaag gagcccttca gctggatcaa agtggacctg 3600
ctggccccca tgatcatcca cggcatcaag acccaggggg ccaggcagaa gttctccagc 3660
ctgtacatca gccagttcat catcatgtac agcctggacg gcaagaagtg gcagacctac 3720
aggggcaaca gcaccggcac cctgatggtg ttcttcggca acgtggacag cagcggcatc 3780
aagcacaaca tcttcaaccc ccccatcatc gccagataca tcaggctgca ccccacccac 3840
tacagcatca ggagcaccct gaggatggag ctgatgggct gtgacctgaa cagctgcagc 3900
atgcccctgg gcatggagag caaggccatc tctgacgccc agatcactgc ctccagctac 3960
ttcaccaaca tgtttgccac ctggagcccc agcaaggcca ggctgcacct gcagggcagg 4020
agcaatgcct ggaggcccca ggtcaacaac cccaaggagt ggctgcaggt ggacttccag 4080
aagaccatga aggtgactgg ggtgaccacc caggggtga agagcctgct gaccagcatg 4140
tacgtgaagg agttcctgat ctccagcagc caggacggcc accagtggac cctgttcttc 4200
cagaatggca aggtgaaggt gttccagggc aaccaggaca gcttcacccc tgtggtcaac 4260
agcctggacc ccccctgct gaccagatac ctgaggatcc acccccagag ctgggtgcac 4320
cagatcgccc tgaggatgga ggtgctgggc tgtgaggccc aggacctgta ctga 4374

<210> 5
<211> 4374
<212> DNA
<213> 人工序列

<220>

<223> coFVIII-25

<400> 5

atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc 60
accaggagat actacctggg cgccgtggag ctgagctggg actacatgca gtctgacctg 120
ggcgagctgc cagtggacgc caggttcccc cccagagtgc ccaagagctt ccccttcaac 180
accagcgtgg tgtacaagaa gaccctgttc gtggagttca ctgaccacct gttcaacatc 240
gccaagccca ggcccccctg gatgggcctg ctgggcccca ccatccaggc cgaggtgtac 300
gacaccgtgg tcatcaccct gaagaacatg gccagccacc ccgtctccct gcacgccgtg 360
ggggtgagct actggaaggc ctctgagggc gccgagtacg acgaccagac cagccagagg 420
gagaaggagg acgacaaggt gttccctggg ggcagccaca cctacgtgtg gcaggtcctg 480
aaggagaacg gccccatggc ctctgacccc ctgtgcctga cctacagcta cctgagccac 540
gtggacctgg tgaaggacct gaactctggc ctgattgggg ccctgctggt gtgcagggag 600
ggcagcctgg ccaaggagaa gacccagacc ctgcacaagt tcatcctgct gttcgccgtg 660
ttcgacgagg gcaagagctg gcactctgaa accaagaaca gcctgatgca ggacagggac 720
gccgcctctg ccagggcctg gcccaagatg cacaccgtca acggctacgt caacaggagc 780
ctgcctggcc tgattggctg ccacaggaag agcgtgtact ggcatgtgat cggcatgggc 840
accacccctg aggtgcacag catcttcctg gagggccaca ccttcctggt caggaaccac 900
aggcaggcca gcctggagat cagccccatc accttcctga ccgcccagac cctgctgatg 960
gacctgggcc agttcctgct gttctgccac atctccagcc accagcacga cggcatggag 1020
gcctacgtga aagtggacag ctgccctgag gagccccagc tgaggatgaa gaacaacgag 1080
gaggccgagg actatgatga cgacctgacc gacagcgaga tggacgtggt caggttcgac 1140
gacgacaaca gccccagctt catccagatc aggagcgtgg ccaagaagca ccccaagacc 1200
tgggtgcact acatcgctgc tgaggaggag gactgggact atgccccccc ggtgctggcc 1260
cctgatgaca ggagctacaa gagccagtac ctgaacaatg gcccccagag gattggcagg 1320
aagtacaaga aagtcaggtt catggcctac actgatgaaa ccttcaagac cagggaggcc 1380
atccagcatg agtctggcat cctgggcccc ctgctgtacg gggaggtggg ggacaccctg 1440
ctgatcatct tcaagaacca ggccagcagg ccctacaaca tctaccccca tggcatcacc 1500
gacgtgaggc ccctgtacag caggaggctg cctaaggggg tgaagcacct gaaagacttc 1560
cccatcctgc ctggggagat cttcaagtac aagtggactg tgactgtgga ggacggcccc 1620
accaagagcg accccaggtg cctgaccaga tactacagca gcttcgtcaa catggagagg 1680
gacctggcct ctggcctgat tggcccctg ctgatctgct acaaggagtc tgtggaccag 1740
aggggcaacc agatcatgag cgacaagagg aacgtgatcc tgttctctgt cttcgacgag 1800
aacaggagct ggtacctgac tgaaaacatc cagcggttcc tccccaaccc cgcgggcgtg 1860
cagctggaag atcctgagtt tcaggcatca aacatcatgc actccattaa cggctacgtg 1920
ttcgattcgc tgcagctgag cgtgtgtctg cacgaagtgg cctactggta catcctgtcc 1980

attggtgccc agactgactt cctgtccgtg tttttctccg gctacacgtt caagcacaag 2040
atggtgtacg aggacaccct gaccctcttc ccttttttccg gcgaaactgt gtttatgagc 2100
atggagaatc ccggcctgtg gatcttgggc tgccacaaca gcgacttccg taacagagga 2160
atgactgcgc tgctcaaggt gtccagctgc gacaagaaca ccggagacta ttatgaggac 2220
tcatacgagg acatctccgc ctacctcctg tccaagaata acgccattga acctcggagc 2280
ttcagccaga acccacccgt gcttaagaga catcaacggg agatcactag gaccaccctg 2340
cagtcagacc aggaggaaat cgactacgat gacaccatct cggtcgagat gaagaaggag 2400
gactttgaca tctacgacga agatgaaaac cagagcccga ggtcgttcca aaagaaaacc 2460
cgccactact ttattgctgc tgtcgagcgg ctgtgggact acggaatgtc gtcctcgccg 2520
cacgtgctcc gcaaccgagc ccagagcggc tcggtgccgc aattcaagaa ggtcgtgttc 2580
caggagttca ctgacgggag cttcactcag cctttgtacc ggggagaact caatgaacat 2640
ctcggcctcc tcggacctta catcagagca gaagtggaag ataacatcat ggtcactttc 2700
cgtaaccaag ccagccgccc gtactcgttc tactcctccc tcatttctta cgaagaggac 2760
cagcggcagg gcgcagaacc gcgcaagaac ttcgtgaagc ccaacgaaac caagacctac 2820
ttctggaaag tgcagcatca tatggccccg actaaggacg agtttgactg caaagcctgg 2880
gcctacttct ccgatgtgga cttggagaag gacgtccact ccggcctcat cggtcccctg 2940
ctcgtgtgcc ataccaatac cctgaacccc gcacacggtc gccaggtcac cgtgcaggag 3000
ttcgctctgt tcttcactat cttcgacgaa actaagtcct ggtacttcac cgagaacatg 3060
gagaggaact gcagagcccc ctgtaacatc cagatggagg acccgacgtt caaggaaaac 3120
taccggttcc acgccattaa cggatacatc atggatacgc tgccgggtct tgtgatggcc 3180
caggatcaac ggatcagatg gtacttattg tcgatgggca gcaacgagaa catccactct 3240
attcacttct ccggtcatgt gttcactgtg cggaagaagg aagagtacaa gatggccctg 3300
tacaaccttt atcccggagt gttcgaaact gtggaaatgc tgccgtcgaa ggccggcatt 3360
tggcgcgtgg agtgtttgat tggagaacat ctccatgcgg ggatgtcaac cctgttcctg 3420
gtgtatagca acaagtgcca gactccgctt gggatggcgt caggacacat tagggatttc 3480
cagatcactg cgtccggcca gtacggccaa tgggccccta agctggcccg cctgcattac 3540
tccggatcca ttaacgcctg gtcaaccaag gagccattct cctggatcaa ggtggacctt 3600
ctggccccca tgattatcca cggaattaag acccaggggg cccggcagaa gttctcctca 3660
ctgtacatca gccagttcat aatcatgtac tccctggacg gaaagaagtg gcaaacctac 3720
aggggaaca gcaccggcac actgatggtc tttttcggaa atgtggactc ctccgggatt 3780
aagcataaca tcttcaaccc tccgattatc gctcggtaca ttagacttca ccctacccac 3840
tacagcattc gctccaccct gcggatggaa ctgatgggct gcgatctgaa ctcgtgcagc 3900
atgccgttgg gaatggagtc caaagcaatt tccgacgcgc agatcaccgc ctcgtcctac 3960
tttaccaaca tgttcgccac gtggtcaccg tccaaggccc ggctgcacct ccagggaaga 4020
tccaacgcat ggcggccaca ggtcaacaac cctaaggagt ggctccaggt ggacttccag 4080
aaaaccatga aggtcaccgg agtcacaacc cagggagtga agtcgctgct gacttctatg 4140

tacgtcaagg agttcctgat ctccagcagc caggacgggc accagtggac cctgttcttc 4200
caaaatggaa aggtcaaggt gtttcagggc aatcaggatt cattcacccc ggtggtgaac 4260
tcccttgatc cacccctcct gacccgctac cttcgcatcc acccacagtc ctgggtgcac 4320
cagatcgcgc tgaggatgga ggtcctggga tgcgaagccc aggacctgta ctga 4374

<210> 6
<211> 4374
<212> DNA
<213> 人工序列

<220>
<223> coFVIII-26

<400> 6
atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc 60
accaggagat actacctggg cgccgtggag ctgagctggg actacatgca gtctgacctg 120
ggcgagctgc cagtggacgc caggttcccc cccagagtgc ccaagagctt ccccttcaac 180
accagcgtgg tgtacaagaa gaccctgttc gtggagttca ctgaccacct gttcaacatc 240
gccaagccca ggccccctg gatgggcctg ctgggcccca ccatccaggc cgaggtgtac 300
gacaccgtgg tcatcaccct gaagaacatg gccagccacc ccgtctccct gcacgccgtg 360
ggggtgagct actggaaggc ctctgagggc gccgagtacg acgaccagac cagccagagg 420
gagaaggagg acgacaaggt gttccctggg ggcagccaca cctacgtgtg gcaggtcctg 480
aaggagaacg gccccatggc ctctgacccc ctgtgcctga cctacagcta cctgagccac 540
gtggacctgg tgaaggacct gaactctggc ctgattgggg ccctgctggt gtgcagggag 600
ggcagcctgg ccaaggagaa gacccagacc ctgcacaagt tcatcctgct gttcgccgtg 660
ttcgacgagg gcaagagctg gcactctgaa accaagaaca gcctgatgca ggacagggac 720
gccgcctctg ccagggcctg gcccaagatg cacaccgtca acggctacgt caacaggagc 780
ctgcctggcc tgattggctg ccacaggaag agcgtgtact ggcatgtgat cggcatgggc 840
accacccctg aggtgcacag catcttcctg gagggccaca ccttcctggt caggaaccac 900
aggcaggcca gcctggagat cagccccatc accttcctga ccgcccagac cctgctgatg 960
gacctgggcc agttcctgct gttctgccac atctccagcc accagcacga cggcatggag 1020
gcctacgtga aagtggacag ctgccctgag gagccccagc tgaggatgaa gaacaacgag 1080
gaggccgagg actatgatga cgacctgacc gacagcgaga tggacgtggt caggttcgac 1140
gacgacaaca gccccagctt catccagatc aggagcgtgg ccaagaagca ccccaagacc 1200
tgggtgcact acatcgctgc tgaggaggag gactgggact atgcccccct ggtgctggcc 1260

cctgatgaca ggagctacaa gagccagtac ctgaacaatg gcccccagag gattggcagg 1320

aagtacaaga aagtcaggtt catggcctac actgatgaaa ccttcaagac cagggaggcc 1380

atccagcatg agtctggcat cctgggcccc ctgctgtacg gggaggtggg ggacaccctg 1440

ctgatcatct tcaagaacca ggccagcagg ccctacaaca tctaccccca tggcatcacc 1500

gacgtgaggc ccctgtacag caggaggctg cctaaggggg tgaagcacct gaaagacttc 1560

cccatcctgc ctggggagat cttcaagtac aagtggactg tgactgtgga ggacggcccc 1620

accaagagcg accccaggtg cctgaccaga tactacagca gcttcgtcaa catggagagg 1680

gacctggcct ctggcctgat tggccccctg ctgatctgct acaaggagtc tgtggaccag 1740

aggggcaacc agatcatgag cgacaagagg aacgtgatcc tgttctctgt cttcgacgag 1800

aacaggagct ggtacctcac tgaaaacatc cagaggttcc tcccaaaccc cgcaggagtg 1860

caactggagg accctgagtt tcaggcctcg aatatcatgc actcgattaa cggttacgtg 1920

ttcgactcgc tgcagctgag cgtgtgcctc catgaagtcg cttactggta cattctgtcc 1980

atcggcgccc agactgactt cctgagcgtg ttcttttccg gttacacctt taagcacaag 2040

atggtgtacg aagataccct gaccctgttc cctttctccg gcgaaacggt gttcatgtcg 2100

atggagaacc cgggtctgtg gattctggga tgccacaaca gcgactttcg gaaccgcgga 2160

atgactgccc tgctgaaggt gtcctcatgc gacaagaaca ccggagacta ctacgaggac 2220

tcctacgagg atatctcagc ctacctcctg tccaagaaca acgcgatcga gccgcgcagc 2280

ttcagccaga acccgcctgt gctgaagagg caccagcgag aaattacccg gaccaccctc 2340

caatcggatc aggaggaaat cgactacgac gacaccatct cggtggaaat gaagaaggaa 2400

gatttcgata tctacgacga ggacgaaaat cagtcccctc gctcattcca aaagaaaact 2460

agacactact ttatcgccgc ggtggaaaga ctgtgggact atggaatgtc atccagccct 2520

cacgtccttc ggaaccgggc ccagagcgga tcggtgcctc agttcaagaa agtggtgttc 2580

caggagttca ccgacggcag cttcacccag ccgctgtacc ggggagaact gaacgaacac 2640

ctgggcctgc tcggtcccta catccgcgcg gaagtggagg ataacatcat ggtgaccttc 2700

cgtaaccaag catccagacc ttactccttc tattcctccc tgatctcata cgaggaggac 2760

cagcgccaag gcgccgagcc ccgcaagaac ttcgtcaagc ccaacgagac taagacctac 2820

ttctggaagg tccaacacca tatggccccg accaaggatg agtttgactg caaggcctgg 2880

gcctacttct ccgacgtgga ccttgagaag gatgtccatt ccggcctgat cgggccgctg 2940

ctcgtgtgtc acaccaacac cctgaaccca gcgcatggac gccaggtcac cgtccaggag 3000

tttgctctgt tcttcaccat ttttgacgaa actaagtcct ggtacttcac cgagaatatg 3060

gagcgaaact gtagagcgcc ctgcaatatc cagatggaag atccgacttt caaggagaac 3120

tatagattcc acgccatcaa cgggtacatc atggatactc tgccggggct ggtcatggcc 3180

caggatcaga ggattcggtg gtacttgctg tcaatgggat cgaacgaaaa cattcactcc 3240

attcacttct ccggtcacgt gttcactgtg cgcaagaagg aggagtacaa gatggcgctg 3300

tacaatctgt accccggggt gttcgaaact gtggagatgc tgccgtccaa ggccggcatc 3360

tggagagtgg agtgcctgat cggagagcac ctccacgcgg ggatgtccac cctcttcctg 3420

gtgtactcga ataagtgcca gaccccgctg ggcatggcct cgggccacat cagagacttc 3480
cagatcacag caagcggaca atacggccaa tgggcgccga agctggcccg cttgcactac 3540
tccggatcga tcaacgcatg gtccaccaag gaaccgttct cgtggattaa ggtggacctc 3600
ctggccccta tgattatcca cggaattaag acccagggcg ccaggcagaa gttctcctcc 3660
ctgtacatct cgcaattcat catcatgtac agcctggacg ggaagaagtg gcagacttac 3720
aggggaaact ccaccggcac cctgatggtc tttttcggca acgtggattc ctccggcatt 3780
aagcacaaca tcttcaaccc accgatcata gccagatata ttaggctcca ccccactcac 3840
tactcaatcc gctcaactct tcggatggaa ctcatggggt gcgacctgaa ctcctgctcc 3900
atgccgttgg ggatggaatc aaaggctatt agcgacgccc agatcaccgc gagctcctac 3960
ttcactaaca tgttcgccac ctggagcccc tccaaggcca ggctgcactt gcagggacgg 4020
tcaaatgcct ggcggccgca agtgaacaat ccgaaggaat ggcttcaagt ggatttccaa 4080
aagaccatga aagtgaccgg agtcaccacc cagggagtga agtcccttct gacctcgatg 4140
tatgtgaagg agttcctgat tagcagcagc caggacgggc accagtggac cctgttcttc 4200
caaaacggaa aggtcaaggt gttccagggg aaccaggact cgttcacacc cgtggtgaac 4260
tccctggacc ccccactgct gacgcggtac ttgaggattc atcctcagtc ctgggtccat 4320
cagattgcat tgcgaatgga agtcctgggc tgcgaggccc aggacctgta ctga 4374

<210> 7
<211> 1734
<212> DNA
<213> 人工序列

<220>
<223> coFVIII-52-NT58

<400> 7
gcgacccgcc gatactacct gggagcagtg gagctctcct gggattacat gcagagcgac 60
cttggggagc tgcccgtgga tgccaggttc cctccccggg tgccaaagtc gtttccgttc 120
aacacctccg tggtgtacaa gaaaactctg ttcgtggagt tcaccgacca cctgttcaat 180
atcgccaagc ccagacctcc ctggatgggg ctgttgggac ctaccatcca agcggaggtg 240
tacgacactg tggtcatcac tctgaagaac atggcctcgc atcccgtgtc cctgcacgcc 300
gtgggagtgt cttactggaa agcgtccgag ggggccgaat acgacgacca gacctcgcag 360
agagaaaagg aagatgacaa ggtgttccca ggaggatcgc acacctacgt gtggcaagtg 420
ttgaaggaga acggcccaat ggcctccgac ccgctgtgcc tgacctactc gtacctgtcc 480
cacgtggacc tcgtgaagga cctcaactcg ggactgattg gagccctgct ggtctgcagg 540

gaaggctcac tggcgaaaga aaagactcag accttgcaca agttcattct gctgttcgct 600
gtgttcgacg aggggaagtc gtggcacagc gagactaaga actccctgat gcaagataga 660
gatgccgcct ccgcccgggc ctggcctaag atgcacaccg tgaacggtta cgtgaaccgc 720
tccctccctg gcctgattgg atgccaccgg aagtccgtgt actggcacgt gatcgggatg 780
gggaccaccc ccgaggtgca cagcatcttc ctggaaggtc acacatttct cgtgcgcaac 840
caccggcagg cctccctgga aatcagcccc attaccttcc tcactgccca gactctgctg 900
atggacctgg gacagttcct gctgttctgc catatctcct cccaccaaca tgacggaatg 960
gaggcatacg tgaaggtcga ttcctgccct gaggaacccc agctccgcat gaagaacaat 1020
gaggaagccg aggactacga cgacgacctg acggatagcg agatggatgt ggtccggttc 1080
gatgacgata acagcccttc cttcatccaa attcgctcgg tggcaaagaa gcaccccaag 1140
acctgggtgc attacattgc ggcggaagaa gaggactggg attatgcccc gcttgtcctc 1200
gctcctgacg accggagcta caagagccag tacctgaaca acggtccaca gaggatcggt 1260
agaaagtaca agaaggtccg cttcatggcc tataccgacg aaaccttcaa aactagagag 1320
gccatccaac acgaatccgg catcctgggc ccgctcttgt acggagaagt cggcgacacc 1380
cttctcatta tcttcaagaa ccaggcttcc cggccgtaca acatctatcc gcatgggatc 1440
actgacgtgc gcccactgta ctcgcggcgc ctgcccaagg gtgtcaaaca cctgaaggat 1500
tttccgatcc ttccgggaga aatcttcaag tacaagtgga ccgtgaccgt ggaagatggc 1560
ccaactaagt ctgaccctag atgcctcacc cgctactact catccttcgt caacatggag 1620
cgcgacctgg ccagcggact gatcggcccg ctgctgattt gctacaagga atcagtggac 1680
caacggggaa accagatcat gtcggataag aggaacgtca tcctcttctc cgtg 1734

<210> 8
<211> 2583
<212> DNA
<213> 人工序列

<220>
<223> coFVIII-52-CT

<400> 8
tttgacgaaa accggtcgtg gtacctgacc gagaacatcc agaggttcct gcccaaccct 60
gctggggtgc agctggagga ccccgagttc caggccagca acatcatgca cagcatcaat 120
ggctacgtgt tcgacagcct gcagctgagc gtgtgcctgc acgaggtggc ctactggtac 180
atcctgagca tcggcgccca gaccgacttc ctgagcgtgt tcttctctgg ctacaccttc 240
aagcacaaga tggtgtatga ggacaccctg accctgttcc ccttcagcgg ggagactgtc 300

ttcatgagca tggagaaccc tggcctgtgg atcctgggct gccacaacag cgacttcagg 360
aacaggggca tgactgccct gctgaaagtc tccagctgtg acaagaacac cggggactac 420
tacgaggaca gctacgagga catcagcgcc tacctgctga gcaagaacaa tgccatcgag 480
cccaggagct tctctcagaa ccccccagtg ctgaagaggc accagaggga gatcaccagg 540
accaccctgc agtctgacca ggaggagatc gactatgatg acaccatcag cgtggagatg 600
aagaaggagg acttcgacat ctacgacgag gacgagaacc agagccccag gagcttccag 660
aagaagacca ggcactactt cattgctgct gtggagaggc tgtgggacta tggcatgtcc 720
agcagccccc atgtgctgag gaacagggcc cagtctggca gcgtgcccca gttcaagaaa 780
gtcgtgttcc aggagttcac cgacggcagc ttcacccagc ccctgtacag aggggagctg 840
aacgagcacc tgggcctgct gggcccctac atcagggccg aggtggagga caacatcatg 900
gtgaccttca ggaaccaggc cagcaggccc tacagcttct acagcagcct gatcagctac 960
gaggaggacc agaggcaggg ggctgagccc aggaagaact ttgtgaagcc caatgaaacc 1020
aagacctact tctggaaggt gcagcaccac atggccccca ccaaggacga gttcgactgc 1080
aaggcctggg cctacttctc tgacgtggac ctggagaagg acgtgcactc tggcctgatt 1140
ggcccccctgc tggtgtgcca caccaacacc ctgaaccctg cccatggcag gcaggtgact 1200
gtgcaggagt tcgccctgtt cttcaccatc ttcgatgaaa ccaagagctg gtacttcact 1260
gagaacatgg agaggaactg cagggccccc tgcaacatcc agatggagga ccccaccttc 1320
aaggagaact acaggttcca tgccatcaat ggctacatca tggacaccct gcctggcctg 1380
gtcatggccc aggaccagag gatcaggtgg tatctgctga gcatgggcag caacgagaac 1440
atccacagca tccacttctc tggccacgtg ttcactgtga ggaagaagga ggagtacaag 1500
atggccctgt acaacctgta ccctggggtg ttcgaaaccg tggagatgct gcccagcaag 1560
gccggcatct ggagggtgga gtgcctgatt ggggagcacc tgcacgccgg catgagcacc 1620
ctgttcctgg tgtacagcaa caagtgccag acccccctgg gcatggcctc tggccacatc 1680
agggacttcc agatcactgc ctctggccag tacggccagt gggcccccaa gctggccagg 1740
ctgcactact ccggaagcat caatgcctgg agcaccaagg agcccttcag ctggatcaaa 1800
gtggacctgc tggcccccat gatcatccac ggcatcaaga cccaggggggc caggcagaag 1860
ttctccagcc tgtacatcag ccagttcatc atcatgtaca gcctggacgg caagaagtgg 1920
cagacctaca ggggcaacag caccggcacc ctgatggtgt tcttcggcaa cgtggacagc 1980
agcggcatca agcacaacat cttcaacccc cccatcatcg ccagatacat caggctgcac 2040
cccacccact acagcatcag gagcaccctg aggatggagc tgatgggctg tgacctgaac 2100
agctgcagca tgcccctggg catggagagc aaggccatct ctgacgccca gatcactgcc 2160
tccagctact tcaccaacat gtttgccacc tggagcccca gcaaggccag gctgcacctg 2220
cagggcagga gcaatgcctg gaggccccag gtcaacaacc ccaaggagtg gctgcaggtg 2280
gacttccaga agaccatgaa ggtgactggg gtgaccaccc agggggtgaa gagcctgctg 2340
accagcatgt acgtgaagga gttcctgatc tccagcagcc aggacggcca ccagtggacc 2400
ctgttcttcc agaatggcaa ggtgaaggtg ttccagggca accaggacag cttcaccccct 2460

gtggtcaaca gcctggaccc ccccctgctg accagatacc tgaggatcca cccccagagc 2520

tgggtgcacc agatcgccct gaggatggag gtgctgggct gtgaggccca ggacctgtac 2580

tga 2583

<210> 9

<211> 1734

<212> DNA

<213> 人工序列

<220>

<223> coFVIII-62-NT58

<400> 9

gccactcgcc ggtactacct tggagccgtg gagctttcat gggactacat gcagagcgac 60

ctgggcgaac tccccgtgga tgccagattc ccccccccgcg tgccaaagtc cttccccttt 120

aacacctccg tggtgtacaa gaaaaccctc tttgtcgagt tcactgacca cctgttcaac 180

atcgccaagc cgcgcccacc ttggatgggc ctcctgggac cgaccattca agctgaagtg 240

tacgacaccg tggtgatcac cctgaagaac atggcgtccc accccgtgtc cctgcatgcg 300

gtcggagtgt cctactggaa ggcctccgaa ggagctgagt acgacgacca gactagccag 360

cgggaaaagg aggacgataa agtgttcccg ggcggctcgc atacttacgt gtggcaagtc 420

ctgaaggaaa acggacctat ggcatccgat cctctgtgcc tgacttactc ctacctttcc 480

catgtggacc tcgtgaagga cctgaacagc gggctgattg gtgcacttct cgtgtgccgc 540

gaaggttcgc tcgctaagga aaagacccag accctccata agttcatcct tttgttcgct 600

gtgttcgatg aaggaaagtc atggcattcc gaaactaaga actcgctgat gcaggaccgg 660

gatgccgcct cagcccgcgc ctggcctaaa atgcatacag tcaacggata cgtgaatcgg 720

tcactgcccg ggctcatcgg ttgtcacaga aagtccgtgt actggcacgt catcggcatg 780

ggcactacgc ctgaagtgca ctccatcttc ctggaagggc acaccttcct cgtgcgcaac 840

caccgccagg cctctctgga aatctccccg attacctttc tgaccgccca gactctgctc 900

atggacctgg ggcagttcct tctcttctgc cacatctcca gccatcagca cgacggaatg 960

gaggcctacg tgaaggtgga ctcatgcccg gaagaacctc agttgcggat gaagaacaac 1020

gaggaggccg aggactatga cgacgatttg actgactccg agatggacgt cgtgcggttc 1080

gatgacgaca acagccccag cttcatccag attcgcagcg tggccaagaa gcaccccaaa 1140

acctgggtgc actacatcgc ggccgaggaa gaagattggg actacgcccc gttggtgctg 1200

gcacccgatg accggtcgta caagtcccag tatctgaaca atggtccgca gcggattggc 1260

agaaagtaca agaaagtgcg gttcatggcg tacactgacg aaacgtttaa gacccgggag 1320

gccattcaac atgagagcgg cattctggga ccactgctgt acggagaggt cggcgatacc 1380

ctgctcatca tcttcaaaaa ccaggcctcc cggccttaca acatctaccc tcacggaatc 1440

accgacgtgc ggccactcta ctcgcggcgc ctgccgaagg gcgtcaagca cctgaaagac 1500

ttccctatcc tgccgggcga aatcttcaag tataagtgga ccgtcaccgt ggaggacggg 1560

cccaccaaga gcgatcctag gtgtctgact cggtactact ccagcttcgt gaacatggaa 1620

cgggacctgg catcgggact cattggaccg ctgctgatct gctacaaaga gtcggtggat 1680

caacgcggca accagatcat gtccgacaag cgcaacgtga tcctgttctc cgtg 1734

<210> 10

<211> 2583

<212> DNA

<213> 人工序列

<220>

<223> coFVIII-62-CT

<400> 10

tttgatgaaa acagatcctg gtacctgacc gagaacatcc agaggttcct gcccaaccct 60

gctggggtgc agctggagga ccccgagttc caggccagca acatcatgca cagcatcaat 120

ggctacgtgt tcgacagcct gcagctgagc gtgtgcctgc acgaggtggc ctactggtac 180

atcctgagca tcggcgccca gaccgacttc ctgagcgtgt tcttctctgg ctacaccttc 240

aagcacaaga tggtgtatga ggacaccctg accctgttcc ccttcagcgg ggagactgtc 300

ttcatgagca tggagaaccc tggcctgtgg atcctgggct gccacaacag cgacttcagg 360

aacaggggca tgactgccct gctgaaagtc tccagctgtg acaagaacac cggggactac 420

tacgaggaca gctacgagga catcagcgcc tacctgctga gcaagaacaa tgccatcgag 480

cccaggagct tctctcagaa ccccccagtg ctgaagaggc accagaggga gatcaccagg 540

accaccctgc agtctgacca ggaggagatc gactatgatg acaccatcag cgtggagatg 600

aagaaggagg acttcgacat ctacgacgag gacgagaacc agagccccag gagcttccag 660

aagaagacca ggcactactt cattgctgct gtggagaggc tgtgggacta tggcatgtcc 720

agcagccccc atgtgctgag gaacagggcc cagtctggca gcgtgcccca gttcaagaaa 780

gtcgtgttcc aggagttcac cgacggcagc ttcacccagc ccctgtacag aggggagctg 840

aacgagcacc tgggcctgct gggcccctac atcagggccg aggtggagga caacatcatg 900

gtgaccttca ggaaccaggc cagcaggccc tacagcttct acagcagcct gatcagctac 960

gaggaggacc agaggcaggg ggctgagccc aggaagaact ttgtgaagcc caatgaaacc 1020

aagacctact tctggaaggt gcagcaccac atggccccca ccaaggacga gttcgactgc 1080

aaggcctggg cctacttctc tgacgtggac ctggagaagg acgtgcactc tggcctgatt 1140

ggcccctgc tggtgtgcca caccaacacc ctgaaccctg cccatggcag gcaggtgact 1200

gtgcaggagt tcgccctgtt cttcaccatc ttcgatgaaa ccaagagctg gtacttcact 1260

gagaacatgg agaggaactg cagggccccc tgcaacatcc agatggagga ccccaccttc 1320

aaggagaact acaggttcca tgccatcaat ggctacatca tggacaccct gcctggcctg 1380

gtcatggccc aggaccagag gatcaggtgg tatctgctga gcatgggcag caacgagaac 1440

atccacagca tccacttctc tggccacgtg ttcactgtga ggaagaagga ggagtacaag 1500

atggccctgt acaacctgta ccctggggtg ttcgaaaccg tggagatgct gcccagcaag 1560

gccggcatct ggagggtgga gtgcctgatt ggggagcacc tgcacgccgg catgagcacc 1620

ctgttcctgg tgtacagcaa caagtgccag acccccctgg gcatggcctc tggccacatc 1680

agggacttcc agatcactgc ctctggccag tacggccagt gggcccccaa gctggccagg 1740

ctgcactact ccggaagcat caatgcctgg agcaccaagg agcccttcag ctggatcaaa 1800

gtggacctgc tggcccccat gatcatccac ggcatcaaga cccaggggc caggcagaag 1860

ttctccagcc tgtacatcag ccagttcatc atcatgtaca gcctggacgg caagaagtgg 1920

cagacctaca ggggcaacag caccggcacc ctgatggtgt tcttcggcaa cgtggacagc 1980

agcggcatca agcacaacat cttcaacccc cccatcatcg ccagatacat caggctgcac 2040

cccacccact acagcatcag gagcaccctg aggatggagc tgatgggctg tgacctgaac 2100

agctgcagca tgcccctggg catggagagc aaggccatct ctgacgccca gatcactgcc 2160

tccagctact tcaccaacat gtttgccacc tggagcccca gcaaggccag gctgcacctg 2220

cagggcagga gcaatgcctg gaggccccag gtcaacaacc ccaaggagtg gctgcaggtg 2280

gacttccaga agaccatgaa ggtgactggg gtgaccaccc agggggtgaa gagcctgctg 2340

accagcatgt acgtgaagga gttcctgatc tccagcagcc aggacggcca ccagtggacc 2400

ctgttcttcc agaatggcaa ggtgaaggtg ttccagggca accaggacag cttcaccct 2460

gtggtcaaca gcctggaccc ccccctgctg accagatacc tgaggatcca cccccagagc 2520

tgggtgcacc agatcgccct gaggatggag gtgctgggct gtgaggccca ggacctgtac 2580

tga 2583

<210> 11

<211> 1734

<212> DNA

<213> 人工序列

<220>

<223> coFVIII-25-NT58

<400> 11

gccaccagga gatactacct gggcgccgtg gagctgagct gggactacat gcagtctgac 60

ctgggcgagc tgccagtgga cgccaggttc ccccccagag tgcccaagag cttccccttc 120

aacaccagcg tggtgtacaa gaagaccctg ttcgtggagt tcactgacca cctgttcaac 180

atcgccaagc ccaggccccc ctggatgggc ctgctgggcc ccaccatcca ggccgaggtg 240

tacgacaccg tggtcatcac cctgaagaac atggccagcc accccgtctc cctgcacgcc 300

gtgggggtga gctactggaa ggcctctgag ggcgccgagt acgacgacca gaccagccag 360

agggagaagg aggacgacaa ggtgttccct gggggcagcc acacctacgt gtggcaggtc 420

ctgaaggaga acggccccat ggcctctgac cccctgtgcc tgacctacag ctacctgagc 480

cacgtggacc tggtgaagga cctgaactct ggcctgattg gggccctgct ggtgtgcagg 540

gagggcagcc tggccaagga gaagacccag accctgcaca agttcatcct gctgttcgcc 600

gtgttcgacg agggcaagag ctggcactct gaaaccaaga acagcctgat gcaggacagg 660

gacgccgcct ctgccagggc ctggcccaag atgcacaccg tcaacggcta cgtcaacagg 720

agcctgcctg gcctgattgg ctgccacagg aagagcgtgt actggcatgt gatcggcatg 780

ggcaccaccc ctgaggtgca cagcatcttc ctggagggcc acaccttcct ggtcaggaac 840

cacaggcagg ccagcctgga gatcagcccc atcaccttcc tgaccgccca gaccctgctg 900

atggacctgg gccagttcct gctgttctgc cacatctcca gccaccagca cgacggcatg 960

gaggcctacg tgaaagtgga cagctgccct gaggagcccc agctgaggat gaagaacaac 1020

gaggaggccg aggactatga tgacgacctg accgacagcg agatggacgt ggtcaggttc 1080

gacgacgaca acagccccag cttcatccag atcaggagcg tggccaagaa gcaccccaag 1140

acctgggtgc actacatcgc tgctgaggag gaggactggg actatgcccc cctggtgctg 1200

gcccctgatg acaggagcta caagagccag tacctgaaca atggccccca gaggattggc 1260

aggaagtaca agaaagtcag gttcatggcc tacactgatg aaaccttcaa gaccagggag 1320

gccatccagc atgagtctgg catcctgggc cccctgctgt acggggaggt gggggacacc 1380

ctgctgatca tcttcaagaa ccaggccagc aggccctaca acatctaccc ccatggcatc 1440

accgacgtga ggcccctgta cagcaggagg ctgcctaagg gggtgaagca cctgaaagac 1500

ttccccatcc tgcctgggga gatcttcaag tacaagtgga ctgtgactgt ggaggacggc 1560

cccaccaaga gcgaccccag gtgcctgacc agatactaca gcagcttcgt caacatggag 1620

agggacctgg cctctggcct gattggcccc ctgctgatct gctacaagga gtctgtggac 1680

cagaggggca accagatcat gagcgacaag aggaacgtga tcctgttctc tgtc 1734

<210> 12

<211> 2583

<212> DNA

<213> 人工序列

<220>

<223> coFVIII-25-CT

<400> 12

ttcgacgaga acaggagctg gtacctgact gaaaacatcc agcggttcct ccccaacccc 60

gcgggcgtgc agctggaaga tcctgagttt caggcatcaa acatcatgca ctccattaac 120

ggctacgtgt tcgattcgct gcagctgagc gtgtgtctgc acgaagtggc ctactggtac 180

atcctgtcca ttggtgccca gactgacttc ctgtccgtgt ttttctccgg ctacacgttc 240

aagcacaaga tggtgtacga ggacaccctg accctcttcc ctttttccgg cgaaactgtg 300

tttatgagca tggagaatcc cggcctgtgg atcttgggct gccacaacag cgacttccgt 360

aacagaggaa tgactgcgct gctcaaggtg tccagctgcg acaagaacac cggagactat 420

tatgaggact catacgagga catctccgcc tacctcctgt ccaagaataa cgccattgaa 480

cctcggagct tcagccagaa cccacccgtg cttaagagac atcaacggga gatcactagg 540

accaccctgc agtcagacca ggaggaaatc gactacgatg acaccatctc ggtcgagatg 600

aagaaggagg actttgacat ctacgacgaa gatgaaaacc agagcccgag gtcgttccaa 660

aagaaaaccc gccactactt tattgctgct gtcgagcggc tgtgggacta cggaatgtcg 720

tcctcgccgc acgtgctccg caaccgagcc cagagcggct cggtgccgca attcaagaag 780

gtcgtgttcc aggagttcac tgacgggagc ttcactcagc ctttgtaccg gggagaactc 840

aatgaacatc tcggcctcct cggaccttac atcagagcag aagtggaaga taacatcatg 900

gtcactttcc gtaaccaagc cagccgcccg tactcgttct actcctccct catttcttac 960

gaagaggacc agcggcaggg cgcagaaccg cgcaagaact tcgtgaagcc caacgaaacc 1020

aagacctact tctggaaagt gcagcatcat atggccccga ctaaggacga gtttgactgc 1080

aaagcctggg cctacttctc cgatgtggac ttggagaagg acgtccactc cggcctcatc 1140

ggtcccctgc tcgtgtgcca taccaatacc ctgaaccccg cacacggtcg ccaggtcacc 1200

gtgcaggagt tcgctctgtt cttcactatc ttcgacgaaa ctaagtcctg gtacttcacc 1260

gagaacatgg agaggaactg cagagccccc tgtaacatcc agatggagga cccgacgttc 1320

aaggaaaact accggttcca cgccattaac ggatacatca tggatacgct gccgggtctt 1380

gtgatggccc aggatcaacg gatcagatgg tacttattgt cgatgggcag caacgagaac 1440

atccactcta ttcacttctc cggtcatgtg ttcactgtgc ggaagaagga agagtacaag 1500

atggccctgt acaaccttta tcccggagtg ttcgaaactg tggaaatgct gccgtcgaag 1560

gccggcattt ggcgcgtgga gtgtttgatt ggagaacatc tccatgcggg gatgtcaacc 1620

ctgttcctgg tgtatagcaa caagtgccag actccgcttg ggatggcgtc aggacacatt 1680

agggatttcc agatcactgc gtccggccag tacggccaat gggcccctaa gctggcccgc 1740

ctgcattact ccggatccat taacgcctgg tcaaccaagg agccattctc ctggatcaag 1800

gtggaccttc tggcccccat gattatccac ggaattaaga cccaggggc ccggcagaag 1860

ttctcctcac tgtacatcag ccagttcata atcatgtact ccctggacgg aaagaagtgg 1920
caaacctaca gggggaacag caccggcaca ctgatggtct ttttcggaaa tgtggactcc 1980
tccgggatta agcataacat cttcaaccct ccgattatcg ctcggtacat tagacttcac 2040
cctacccact acagcattcg ctccaccctg cggatggaac tgatgggctg cgatctgaac 2100
tcgtgcagca tgccgttggg aatggagtcc aaagcaattt ccgacgcgca gatcaccgcc 2160
tcgtcctact ttaccaacat gttcgccacg tggtcaccgt ccaaggcccg gctgcacctc 2220
cagggaagat ccaacgcatg gcggccacag gtcaacaacc ctaaggagtg gctccaggtg 2280
gacttccaga aaaccatgaa ggtcaccgga gtcacaaccc agggagtgaa gtcgctgctg 2340
acttctatgt acgtcaagga gttcctgatc tccagcagcc aggacgggca ccagtggacc 2400
ctgttcttcc aaaatggaaa ggtcaaggtg tttcagggca atcaggattc attcaccccg 2460
gtggtgaact cccttgatcc acccctcctg acccgctacc ttcgcatcca cccacagtcc 2520
tgggtgcacc agatcgcgct gaggatggag gtcctgggat gcgaagccca ggacctgtac 2580
tga 2583

<210> 13
<211> 1734
<212> DNA
<213> 人工序列

<220>
<223> coFVIII-26-NT58

<400> 13
gccaccagga gatactacct gggcgccgtg gagctgagct gggactacat gcagtctgac 60
ctgggcgagc tgccagtgga cgccaggttc ccccccagag tgcccaagag cttccccttc 120
aacaccagcg tggtgtacaa gaagaccctg ttcgtggagt tcactgacca cctgttcaac 180
atcgccaagc ccaggccccc ctggatgggc ctgctgggcc ccaccatcca ggccgaggtg 240
tacgacaccg tggtcatcac cctgaagaac atggccagcc accccgtctc cctgcacgcc 300
gtgggggtga gctactggaa ggcctctgag ggcgccgagt acgacgacca gaccagccag 360
agggagaagg aggacgacaa ggtgttccct gggggcagcc acacctacgt gtggcaggtc 420
ctgaaggaga acggccccat ggcctctgac cccctgtgcc tgacctacag ctacctgagc 480
cacgtggacc tggtgaagga cctgaactct ggcctgattg gggccctgct ggtgtgcagg 540
gagggcagcc tggccaagga gaagacccag accctgcaca agttcatcct gctgttcgcc 600
gtgttcgacg agggcaagag ctggcactct gaaaccaaga acagcctgat gcaggacagg 660
gacgccgcct ctgccagggc ctggcccaag atgcacaccg tcaacggcta cgtcaacagg 720
agcctgcctg gcctgattgg ctgccacagg aagagcgtgt actggcatgt gatcggcatg 780

ggcaccaccc ctgaggtgca cagcatcttc ctggagggcc acaccttcct ggtcaggaac 840
cacaggcagg ccagcctgga gatcagcccc atcaccttcc tgaccgccca gaccctgctg 900
atggacctgg gccagttcct gctgttctgc cacatctcca gccaccagca cgacggcatg 960
gaggcctacg tgaaagtgga cagctgccct gaggagcccc agctgaggat gaagaacaac 1020
gaggaggccg aggactatga tgacgacctg accgacagcg agatggacgt ggtcaggttc 1080
gacgacgaca acagccccag cttcatccag atcaggagcg tggccaagaa gcaccccaag 1140
acctgggtgc actacatcgc tgctgaggag gaggactggg actatgcccc cctggtgctg 1200
gcccctgatg acaggagcta caagagccag tacctgaaca atggccccca gaggattggc 1260
aggaagtaca agaaagtcag gttcatggcc tacactgatg aaaccttcaa gaccagggag 1320
gccatccagc atgagtctgg catcctgggc cccctgctgt acggggaggt gggggacacc 1380
ctgctgatca tcttcaagaa ccaggccagc aggccctaca acatctaccc ccatggcatc 1440
accgacgtga ggcccctgta cagcaggagg ctgcctaagg gggtgaagca cctgaaagac 1500
ttccccatcc tgcctgggga gatcttcaag tacaagtgga ctgtgactgt ggaggacggc 1560
cccaccaaga gcgaccccag gtgcctgacc agatactaca gcagcttcgt caacatggag 1620
agggacctgg cctctggcct gattggcccc ctgctgatct gctacaagga gtctgtggac 1680
cagaggggca accagatcat gagcgacaag aggaacgtga tcctgttctc tgtc 1734

<210> 14
<211> 2583
<212> DNA
<213> 人工序列

<220>
<223> coFVIII-26-CT

<400> 14
ttcgacgaga acaggagctg gtacctcact gaaaacatcc agaggttcct cccaaacccc 60
gcaggagtgc aactggagga ccctgagttt caggcctcga atatcatgca ctcgattaac 120
ggttacgtgt tcgactcgct gcagctgagc gtgtgcctcc atgaagtcgc ttactggtac 180
attctgtcca tcggcgccca gactgacttc ctgagcgtgt tcttttccgg ttacaccttt 240
aagcacaaga tggtgtacga agataccctg accctgttcc ctttctccgg cgaaacggtg 300
ttcatgtcga tggagaaccc gggtctgtgg attctgggat gccacaacag cgactttcgg 360
aaccgcggaa tgactgccct gctgaaggtg tcctcatgcg acaagaacac cggagactac 420
tacgaggact cctacgagga tatctcagcc tacctcctgt ccaagaacaa cgcgatcgag 480
ccgcgcagct tcagccagaa cccgcctgtg ctgaagaggc accagcgaga aattacccgg 540

accaccctcc aatcggatca ggaggaaatc gactacgacg acaccatctc ggtggaaatg 600

aagaaggaag atttcgatat ctacgacgag gacgaaaatc agtcccctcg ctcattccaa 660

aagaaaacta gacactactt tatcgccgcg gtggaaagac tgtgggacta tggaatgtca 720

tccagccctc acgtccttcg gaaccgggcc cagagcggat cggtgcctca gttcaagaaa 780

gtggtgttcc aggagttcac cgacggcagc ttcacccagc cgctgtaccg gggagaactg 840

aacgaacacc tgggcctgct cggtccctac atccgcgcgg aagtggagga taacatcatg 900

gtgaccttcc gtaaccaagc atccagacct tactccttct attcctccct gatctcatac 960

gaggaggacc agcgccaagg cgccgagccc cgcaagaact tcgtcaagcc caacgagact 1020

aagacctact tctggaaggt ccaacaccat atggccccga ccaaggatga gtttgactgc 1080

aaggcctggg cctacttctc cgacgtggac cttgagaagg atgtccattc cggcctgatc 1140

gggccgctgc tcgtgtgtca caccaacacc ctgaacccag cgcatggacg ccaggtcacc 1200

gtccaggagt ttgctctgtt cttcaccatt tttgacgaaa ctaagtcctg gtacttcacc 1260

gagaatatgg agcgaaactg tagagcgccc tgcaatatcc agatggaaga tccgactttc 1320

aaggagaact atagattcca cgccatcaac gggtacatca tggatactct gccggggctg 1380

gtcatggccc aggatcagag gattcggtgg tacttgctgt caatgggatc gaacgaaaac 1440

attcactcca ttcacttctc cggtcacgtg ttcactgtgc gcaagaagga ggagtacaag 1500

atggcgctgt acaatctgta ccccggggtg ttcgaaactg tggagatgct gccgtccaag 1560

gccggcatct ggagagtgga gtgcctgatc ggagagcacc tccacgcggg gatgtccacc 1620

ctcttcctgg tgtactcgaa taagtgccag accccgctgg gcatggcctc gggccacatc 1680

agagacttcc agatcacagc aagcggacaa tacggccaat gggcgccgaa gctggcccgc 1740

ttgcactact ccggatcgat caacgcatgg tccaccaagg aaccgttctc gtggattaag 1800

gtggacctcc tggcccctat gattatccac ggaattaaga cccagggcgc caggcagaag 1860

ttctcctccc tgtacatctc gcaattcatc atcatgtaca gcctggacgg gaagaagtgg 1920

cagacttaca ggggaaactc caccggcacc ctgatggtct ttttcggcaa cgtggattcc 1980

tccggcatta agcacaacat cttcaaccca ccgatcatag ccagatatat taggctccac 2040

cccactcact actcaatccg ctcaactctt cggatggaac tcatggggtg cgacctgaac 2100

tcctgctcca tgccgttggg gatggaatca aaggctatta gcgacgccca gatcaccgcg 2160

agctcctact tcactaacat gttcgccacc tggagccccc ccaaggccag gctgcacttg 2220

cagggacggt caaatgcctg gcggccgcaa gtgaacaatc cgaaggaatg gcttcaagtg 2280

gatttccaaa agaccatgaa agtgaccgga gtcaccaccc agggagtgaa gtcccttctg 2340

acctcgatgt atgtgaagga gttcctgatt agcagcagcc aggacgggca ccagtggacc 2400

ctgttcttcc aaaacggaaa ggtcaaggtg ttccagggga accaggactc gttcacaccc 2460

gtggtgaact ccctggaccc cccactgctg acgcggtact tgaggattca tcctcagtcc 2520

tgggtccatc agattgcatt gcgaatggaa gtcctgggct gcgaggccca ggacctgtac 2580

tga 2583

<210> 15

<211> 2332

<212> PRT

<213> 智人

<400> 15

```
Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser Trp Asp Tyr
1               5                  10                  15

Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg Phe Pro Pro
            20                  25                  30

Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val Tyr Lys Lys
        35                  40                  45

Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile Ala Lys Pro
    50                  55                  60

Arg Pro Pro Trp Met Gly Leu Leu Gly Pro Thr Ile Gln Ala Glu Val
65                  70                  75                  80

Tyr Asp Thr Val Val Ile Thr Leu Lys Asn Met Ala Ser His Pro Val
                85                  90                  95

Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ala Ser Glu Gly Ala
            100                 105                 110

Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp Asp Lys Val
        115                 120                 125

Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu Lys Glu Asn
    130                 135                 140

Gly Pro Met Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser Tyr Leu Ser
145                 150                 155                 160
```

```
His Val Asp Leu Val Lys Asp Leu Asn Ser Gly Leu Ile Gly Ala Leu
                165                 170                 175

Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr Gln Thr Leu
            180                 185                 190

His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly Lys Ser Trp
        195                 200                 205

His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp Ala Ala Ser
    210                 215                 220

Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr Val Asn Arg
225                 230                 235                 240

Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val Tyr Trp His
                245                 250                 255

Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile Phe Leu Glu
            260                 265                 270

Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser Leu Glu Ile
        275                 280                 285

Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met Asp Leu Gly
    290                 295                 300

Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His Asp Gly Met
305                 310                 315                 320

Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro Gln Leu Arg
                325                 330                 335

Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp Leu Thr Asp
            340                 345                 350
```

```
Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser Pro Ser Phe
        355                 360                 365

Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr Trp Val His
    370                 375                 380

Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro Leu Val Leu
385                 390                 395                 400

Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn Asn Gly Pro
                405                 410                 415

Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met Ala Tyr Thr
            420                 425                 430

Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu Ser Gly Ile
        435                 440                 445

Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu Leu Ile Ile
    450                 455                 460

Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro His Gly Ile
465                 470                 475                 480

Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys Gly Val Lys
                485                 490                 495

His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe Lys Tyr Lys
            500                 505                 510

Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp Pro Arg Cys
        515                 520                 525

Leu Thr Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg Asp Leu Ala
    530                 535                 540
```

Ser Gly Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu Ser Val Asp
545 550 555 560

Gln Arg Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val Ile Leu Phe
565 570 575

Ser Val Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu Asn Ile Gln
580 585 590

Arg Phe Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp Pro Glu Phe
595 600 605

Gln Ala Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val Phe Asp Ser
610 615 620

Leu Gln Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp Tyr Ile Leu
625 630 635 640

Ser Ile Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe Ser Gly Tyr
645 650 655

Thr Phe Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr Leu Phe Pro
660 665 670

Phe Ser Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro Gly Leu Trp
675 680 685

Ile Leu Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly Met Thr Ala
690 695 700

Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp Tyr Tyr Glu
705 710 715 720

Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys Asn Asn Ala
725 730 735

```
Ile Glu Pro Arg Ser Phe Ser Gln Asn Ser Arg His Pro Ser Thr Arg
            740                 745                 750

Gln Lys Gln Phe Asn Ala Thr Thr Ile Pro Glu Asn Asp Ile Glu Lys
        755                 760                 765

Thr Asp Pro Trp Phe Ala His Arg Thr Pro Met Pro Lys Ile Gln Asn
    770                 775                 780

Val Ser Ser Ser Asp Leu Leu Met Leu Leu Arg Gln Ser Pro Thr Pro
785                 790                 795                 800

His Gly Leu Ser Leu Ser Asp Leu Gln Glu Ala Lys Tyr Glu Thr Phe
                805                 810                 815

Ser Asp Asp Pro Ser Pro Gly Ala Ile Asp Ser Asn Asn Ser Leu Ser
            820                 825                 830

Glu Met Thr His Phe Arg Pro Gln Leu His His Ser Gly Asp Met Val
        835                 840                 845

Phe Thr Pro Glu Ser Gly Leu Gln Leu Arg Leu Asn Glu Lys Leu Gly
    850                 855                 860

Thr Thr Ala Ala Thr Glu Leu Lys Lys Leu Asp Phe Lys Val Ser Ser
865                 870                 875                 880

Thr Ser Asn Asn Leu Ile Ser Thr Ile Pro Ser Asp Asn Leu Ala Ala
                885                 890                 895

Gly Thr Asp Asn Thr Ser Ser Leu Gly Pro Pro Ser Met Pro Val His
            900                 905                 910

Tyr Asp Ser Gln Leu Asp Thr Thr Leu Phe Gly Lys Lys Ser Ser Pro
        915                 920                 925
```

```
Leu Thr Glu Ser Gly Gly Pro Leu Ser Leu Ser Glu Glu Asn Asn Asp
    930                 935                 940

Ser Lys Leu Leu Glu Ser Gly Leu Met Asn Ser Gln Glu Ser Ser Trp
945                 950                 955                 960

Gly Lys Asn Val Ser Ser Thr Glu Ser Gly Arg Leu Phe Lys Gly Lys
                965                 970                 975

Arg Ala His Gly Pro Ala Leu Leu Thr Lys Asp Asn Ala Leu Phe Lys
            980                 985                 990

Val Ser Ile Ser Leu Leu Lys Thr  Asn Lys Thr Ser Asn  Asn Ser Ala
        995                 1000                1005

Thr Asn  Arg Lys Thr His Ile  Asp Gly Pro Ser Leu  Leu Ile Glu
    1010                1015                1020

Asn Ser  Pro Ser Val Trp Gln  Asn Ile Leu Glu Ser  Asp Thr Glu
    1025                1030                1035

Phe Lys  Lys Val Thr Pro Leu  Ile His Asp Arg Met  Leu Met Asp
    1040                1045                1050

Lys Asn  Ala Thr Ala Leu Arg  Leu Asn His Met Ser  Asn Lys Thr
    1055                1060                1065

Thr Ser  Ser Lys Asn Met Glu  Met Val Gln Gln Lys  Lys Glu Gly
    1070                1075                1080

Pro Ile  Pro Pro Asp Ala Gln  Asn Pro Asp Met Ser  Phe Phe Lys
    1085                1090                1095

Met Leu  Phe Leu Pro Glu Ser  Ala Arg Trp Ile Gln  Arg Thr His
    1100                1105                1110
```

Gly Lys Asn Ser Leu Asn Ser Gly Gln Gly Pro Ser Pro Lys Gln
1115 1120 1125

Leu Val Ser Leu Gly Pro Glu Lys Ser Val Glu Gly Gln Asn Phe
1130 1135 1140

Leu Ser Glu Lys Asn Lys Val Val Val Gly Lys Gly Glu Phe Thr
1145 1150 1155

Lys Asp Val Gly Leu Lys Glu Met Val Phe Pro Ser Ser Arg Asn
1160 1165 1170

Leu Phe Leu Thr Asn Leu Asp Asn Leu His Glu Asn Asn Thr His
1175 1180 1185

Asn Gln Glu Lys Lys Ile Gln Glu Glu Ile Glu Lys Lys Glu Thr
1190 1195 1200

Leu Ile Gln Glu Asn Val Val Leu Pro Gln Ile His Thr Val Thr
1205 1210 1215

Gly Thr Lys Asn Phe Met Lys Asn Leu Phe Leu Leu Ser Thr Arg
1220 1225 1230

Gln Asn Val Glu Gly Ser Tyr Asp Gly Ala Tyr Ala Pro Val Leu
1235 1240 1245

Gln Asp Phe Arg Ser Leu Asn Asp Ser Thr Asn Arg Thr Lys Lys
1250 1255 1260

His Thr Ala His Phe Ser Lys Lys Gly Glu Glu Glu Asn Leu Glu
1265 1270 1275

Gly Leu Gly Asn Gln Thr Lys Gln Ile Val Glu Lys Tyr Ala Cys
1280 1285 1290

```
Thr Thr  Arg Ile Ser Pro Asn  Thr Ser Gln Gln Asn  Phe Val Thr
    1295                 1300                 1305

Gln Arg  Ser Lys Arg Ala Leu  Lys Gln Phe Arg Leu  Pro Leu Glu
    1310                 1315                 1320

Glu Thr  Glu Leu Glu Lys Arg  Ile Ile Val Asp Asp  Thr Ser Thr
    1325                 1330                 1335

Gln Trp  Ser Lys Asn Met Lys  His Leu Thr Pro Ser  Thr Leu Thr
    1340                 1345                 1350

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

Pro Leu  Ser Asp Cys Leu Thr  Arg Ser His Ser Ile  Pro Gln Ala
    1370                 1375                 1380

Asn Arg  Ser Pro Leu Pro Ile  Ala Lys Val Ser Ser  Phe Pro Ser
    1385                 1390                 1395

Ile Arg  Pro Ile Tyr Leu Thr  Arg Val Leu Phe Gln  Asp Asn Ser
    1400                 1405                 1410

Ser His  Leu Pro Ala Ala Ser  Tyr Arg Lys Lys Asp  Ser Gly Val
    1415                 1420                 1425

Gln Glu  Ser Ser His Phe Leu  Gln Gly Ala Lys Lys  Asn Asn Leu
    1430                 1435                 1440

Ser Leu  Ala Ile Leu Thr Leu  Glu Met Thr Gly Asp  Gln Arg Glu
    1445                 1450                 1455

Val Gly  Ser Leu Gly Thr Ser  Ala Thr Asn Ser Val  Thr Tyr Lys
    1460                 1465                 1470
```

Lys Val Glu Asn Thr Val Leu Pro Lys Pro Asp Leu Pro Lys Thr
1475 1480 1485

Ser Gly Lys Val Glu Leu Leu Pro Lys Val His Ile Tyr Gln Lys
1490 1495 1500

Asp Leu Phe Pro Thr Glu Thr Ser Asn Gly Ser Pro Gly His Leu
1505 1510 1515

Asp Leu Val Glu Gly Ser Leu Leu Gln Gly Thr Glu Gly Ala Ile
1520 1525 1530

Lys Trp Asn Glu Ala Asn Arg Pro Gly Lys Val Pro Phe Leu Arg
1535 1540 1545

Val Ala Thr Glu Ser Ser Ala Lys Thr Pro Ser Lys Leu Leu Asp
1550 1555 1560

Pro Leu Ala Trp Asp Asn His Tyr Gly Thr Gln Ile Pro Lys Glu
1565 1570 1575

Glu Trp Lys Ser Gln Glu Lys Ser Pro Glu Lys Thr Ala Phe Lys
1580 1585 1590

Lys Lys Asp Thr Ile Leu Ser Leu Asn Ala Cys Glu Ser Asn His
1595 1600 1605

Ala Ile Ala Ala Ile Asn Glu Gly Gln Asn Lys Pro Glu Ile Glu
1610 1615 1620

Val Thr Trp Ala Lys Gln Gly Arg Thr Glu Arg Leu Cys Ser Gln
1625 1630 1635

Asn Pro Pro Val Leu Lys Arg His Gln Arg Glu Ile Thr Arg Thr
1640 1645 1650

Thr Leu Gln Ser Asp Gln Glu Glu Ile Asp Tyr Asp Asp Thr Ile
1655 1660 1665

Ser Val Glu Met Lys Lys Glu Asp Phe Asp Ile Tyr Asp Glu Asp
1670 1675 1680

Glu Asn Gln Ser Pro Arg Ser Phe Gln Lys Lys Thr Arg His Tyr
1685 1690 1695

Phe Ile Ala Ala Val Glu Arg Leu Trp Asp Tyr Gly Met Ser Ser
1700 1705 1710

Ser Pro His Val Leu Arg Asn Arg Ala Gln Ser Gly Ser Val Pro
1715 1720 1725

Gln Phe Lys Lys Val Val Phe Gln Glu Phe Thr Asp Gly Ser Phe
1730 1735 1740

Thr Gln Pro Leu Tyr Arg Gly Glu Leu Asn Glu His Leu Gly Leu
1745 1750 1755

Leu Gly Pro Tyr Ile Arg Ala Glu Val Glu Asp Asn Ile Met Val
1760 1765 1770

Thr Phe Arg Asn Gln Ala Ser Arg Pro Tyr Ser Phe Tyr Ser Ser
1775 1780 1785

Leu Ile Ser Tyr Glu Glu Asp Gln Arg Gln Gly Ala Glu Pro Arg
1790 1795 1800

Lys Asn Phe Val Lys Pro Asn Glu Thr Lys Thr Tyr Phe Trp Lys
1805 1810 1815

Val Gln His His Met Ala Pro Thr Lys Asp Glu Phe Asp Cys Lys
1820 1825 1830

```
Ala Trp  Ala Tyr Phe Ser Asp  Val Asp Leu Glu Lys  Asp Val His
    1835                 1840                 1845

Ser Gly  Leu Ile Gly Pro Leu  Leu Val Cys His Thr  Asn Thr Leu
    1850                 1855                 1860

Asn Pro  Ala His Gly Arg Gln  Val Thr Val Gln Glu  Phe Ala Leu
    1865                 1870                 1875

Phe Phe  Thr Ile Phe Asp Glu  Thr Lys Ser Trp Tyr  Phe Thr Glu
    1880                 1885                 1890

Asn Met  Glu Arg Asn Cys Arg  Ala Pro Cys Asn Ile  Gln Met Glu
    1895                 1900                 1905

Asp Pro  Thr Phe Lys Glu Asn  Tyr Arg Phe His Ala  Ile Asn Gly
    1910                 1915                 1920

Tyr Ile  Met Asp Thr Leu Pro  Gly Leu Val Met Ala  Gln Asp Gln
    1925                 1930                 1935

Arg Ile  Arg Trp Tyr Leu Leu  Ser Met Gly Ser Asn  Glu Asn Ile
    1940                 1945                 1950

His Ser  Ile His Phe Ser Gly  His Val Phe Thr Val  Arg Lys Lys
    1955                 1960                 1965

Glu Glu  Tyr Lys Met Ala Leu  Tyr Asn Leu Tyr Pro  Gly Val Phe
    1970                 1975                 1980

Glu Thr  Val Glu Met Leu Pro  Ser Lys Ala Gly Ile  Trp Arg Val
    1985                 1990                 1995

Glu Cys  Leu Ile Gly Glu His  Leu His Ala Gly Met  Ser Thr Leu
    2000                 2005                 2010
```

Phe Leu Val Tyr Ser Asn Lys Cys Gln Thr Pro Leu Gly Met Ala
2015 2020 2025

Ser Gly His Ile Arg Asp Phe Gln Ile Thr Ala Ser Gly Gln Tyr
2030 2035 2040

Gly Gln Trp Ala Pro Lys Leu Ala Arg Leu His Tyr Ser Gly Ser
2045 2050 2055

Ile Asn Ala Trp Ser Thr Lys Glu Pro Phe Ser Trp Ile Lys Val
2060 2065 2070

Asp Leu Leu Ala Pro Met Ile Ile His Gly Ile Lys Thr Gln Gly
2075 2080 2085

Ala Arg Gln Lys Phe Ser Ser Leu Tyr Ile Ser Gln Phe Ile Ile
2090 2095 2100

Met Tyr Ser Leu Asp Gly Lys Lys Trp Gln Thr Tyr Arg Gly Asn
2105 2110 2115

Ser Thr Gly Thr Leu Met Val Phe Phe Gly Asn Val Asp Ser Ser
2120 2125 2130

Gly Ile Lys His Asn Ile Phe Asn Pro Pro Ile Ile Ala Arg Tyr
2135 2140 2145

Ile Arg Leu His Pro Thr His Tyr Ser Ile Arg Ser Thr Leu Arg
2150 2155 2160

Met Glu Leu Met Gly Cys Asp Leu Asn Ser Cys Ser Met Pro Leu
2165 2170 2175

Gly Met Glu Ser Lys Ala Ile Ser Asp Ala Gln Ile Thr Ala Ser
2180 2185 2190

Ser Tyr Phe Thr Asn Met Phe Ala Thr Trp Ser Pro Ser Lys Ala
2195 2200 2205

Arg Leu His Leu Gln Gly Arg Ser Asn Ala Trp Arg Pro Gln Val
2210 2215 2220

Asn Asn Pro Lys Glu Trp Leu Gln Val Asp Phe Gln Lys Thr Met
2225 2230 2235

Lys Val Thr Gly Val Thr Thr Gln Gly Val Lys Ser Leu Leu Thr
2240 2245 2250

Ser Met Tyr Val Lys Glu Phe Leu Ile Ser Ser Ser Gln Asp Gly
2255 2260 2265

His Gln Trp Thr Leu Phe Phe Gln Asn Gly Lys Val Lys Val Phe
2270 2275 2280

Gln Gly Asn Gln Asp Ser Phe Thr Pro Val Val Asn Ser Leu Asp
2285 2290 2295

Pro Pro Leu Leu Thr Arg Tyr Leu Arg Ile His Pro Gln Ser Trp
2300 2305 2310

Val His Gln Ile Ala Leu Arg Met Glu Val Leu Gly Cys Glu Ala
2315 2320 2325

Gln Asp Leu Tyr
2330

<210> 16
<211> 4371
<212> DNA
<213> 人工序列

<220>

<223> BDD-FVIII（未優化；"親代"），核苷酸序列

<400> 16

atgcaaatag agctctccac ctgcttcttt ctgtgccttt tgcgattctg ctttagtgcc 60

accagaagat actacctggg tgcagtggaa ctgtcatggg actatatgca aagtgatctc 120

ggtgagctgc ctgtggacgc aagatttcct cctagagtgc caaaatcttt tccattcaac 180

acctcagtcg tgtacaaaaa gactctgttt gtagaattca cggatcacct tttcaacatc 240

gctaagccaa ggccaccctg gatgggtctg ctaggtccta ccatccaggc tgaggtttat 300

gatacagtgg tcattacact taagaacatg gcttcccatc ctgtcagtct tcatgctgtt 360

ggtgtatcct actggaaagc ttctgaggga gctgaatatg atgatcagac cagtcaaagg 420

gagaaagaag atgataaagt cttccctggt ggaagccata catatgtctg gcaggtcctg 480

aaagagaatg gtccaatggc ctctgaccca ctgtgcctta cctactcata tctttctcat 540

gtggacctgg taaaagactt gaattcaggc ctcattggag ccctactagt atgtagagaa 600

gggagtctgg ccaaggaaaa gacacagacc ttgcacaaat ttatactact ttttgctgta 660

tttgatgaag ggaaaagttg gcactcagaa acaaagaact ccttgatgca ggatagggat 720

gctgcatctg ctcgggcctg gcctaaaatg cacacagtca atggttatgt aaacaggtct 780

ctgccaggtc tgattggatg ccacaggaaa tcagtctatt ggcatgtgat tggaatgggc 840

accactcctg aagtgcactc aatattcctc gaaggtcaca catttcttgt gaggaaccat 900

cgccaggcgt ccttggaaat ctcgccaata actttcctta ctgctcaaac actcttgatg 960

gaccttggac agtttctact gttttgtcat atctcttccc accaacatga tggcatggaa 1020

gcttatgtca aagtagacag ctgtccagag gaaccccaac tacgaatgaa aaataatgaa 1080

gaagcggaag actatgatga tgatcttact gattctgaaa tggatgtggt caggtttgat 1140

gatgacaact ctccttcctt tatccaaatt cgctcagttg ccaagaagca tcctaaaact 1200

tgggtacatt acattgctgc tgaagaggag gactgggact atgctccctt agtcctcgcc 1260

cccgatgaca gaagttataa aagtcaatat ttgaacaatg gccctcagcg gattggtagg 1320

aagtacaaaa aagtccgatt tatggcatac acagatgaaa cctttaagac tcgtgaagct 1380

attcagcatg aatcaggaat cttgggacct ttactttatg gggaagttgg agacacactg 1440

ttgattatat ttaagaatca agcaagcaga ccatataaca tctaccctca cggaatcact 1500

gatgtccgtc ctttgtattc aaggagatta ccaaaaggtg taaaacattt gaaggatttt 1560

ccaattctgc caggagaaat attcaaatat aaatggacag tgactgtaga agatgggcca 1620

actaaatcag atcctcggtg cctgacccgc tattactcta gtttcgttaa tatggagaga 1680

gatctagctt caggactcat tggccctctc ctcatctgct acaaagaatc tgtagatcaa 1740

agaggaaacc agataatgtc agacaagagg aatgtcatcc tgttttctgt atttgatgag 1800

aaccgaagct ggtacctcac agagaatata caacgctttc tccccaatcc agctggagtg 1860

cagcttgagg atccagagtt ccaagcctcc aacatcatgc acagcatcaa tggctatgtt 1920

tttgatagtt tgcagttgtc agtttgtttg catgaggtgg catactggta cattctaagc 1980

attggagcac agactgactt cctttctgtc ttcttctctg gatatacctt caaacacaaa 2040

atggtctatg aagacacact caccctattc ccattctcag gagaaactgt cttcatgtcg 2100

atggaaaacc caggtctatg gattctgggg tgccacaact cagactttcg gaacagaggc 2160

atgaccgcct tactgaaggt ttctagttgt gacaagaaca ctggtgatta ttacgaggac 2220

agttatgaag atatttcagc atacttgctg agtaaaaaca atgccattga accaagaagc 2280

ttctctcaaa acccaccagt cttgaaacgc catcaacggg aaataactcg tactactctt 2340

cagtcagatc aagaggaaat tgactatgat gataccatat cagttgaaat gaagaaggaa 2400

gattttgaca tttatgatga ggatgaaaat cagagccccc gcagctttca aaagaaaaca 2460

cgacactatt ttattgctgc agtggagagg ctctgggatt atgggatgag tagctcccca 2520

catgttctaa gaaacagggc tcagagtggc agtgtccctc agttcaagaa agttgttttc 2580

caggaattta ctgatggctc ctttactcag cccttatacc gtggagaact aaatgaacat 2640

ttgggactcc tggggccata tataagagca gaagttgaag ataatatcat ggtaactttc 2700

agaaatcagg cctctcgtcc ctattccttc tattctagcc ttatttctta tgaggaagat 2760

cagaggcaag gagcagaacc tagaaaaaac tttgtcaagc ctaatgaaac caaaacttac 2820

ttttggaaag tgcaacatca tatggcaccc actaaagatg agtttgactg caaagcctgg 2880

gcttatttct ctgatgttga cctggaaaaa gatgtgcact caggcctgat tggacccctt 2940

ctggtctgcc acactaacac actgaaccct gctcatggga gacaagtgac agtacaggaa 3000

tttgctctgt ttttcaccat ctttgatgag accaaaagct ggtacttcac tgaaaatatg 3060

gaaagaaact gcagggctcc ctgcaatatc cagatggaag atcccacttt taaagagaat 3120

tatcgcttcc atgcaatcaa tggctacata atggatacac tacctggctt agtaatggct 3180

caggatcaaa ggattcgatg gtatctgctc agcatgggca gcaatgaaaa catccattct 3240

attcatttca gtggacatgt gttcactgta cgaaaaaaag aggagtataa aatggcactg 3300

tacaatctct atccaggtgt ttttgagaca gtggaaatgt taccatccaa agctggaatt 3360

tggcgggtgg aatgccttat tggcgagcat ctacatgctg ggatgagcac acttttctg 3420

gtgtacagca ataagtgtca gactcccctg ggaatggctt ctggacacat tagagatttt 3480

cagattacag cttcaggaca atatggacag tgggccccaa agctggccag acttcattat 3540

tccggatcaa tcaatgcctg gagcaccaag gagccctttt cttggatcaa ggtggatctg 3600

ttggcaccaa tgattattca cggcatcaag acccagggtg cccgtcagaa gttctccagc 3660

ctctacatct ctcagtttat catcatgtat agtcttgatg ggaagaagtg gcagacttat 3720

cgaggaaatt ccactggaac cttaatggtc ttctttggca atgtggattc atctgggata 3780

aaacacaata tttttaaccc tccaattatt gctcgataca tccgtttgca cccaactcat 3840

tatagcattc gcagcactct tcgcatggag ttgatgggct gtgatttaaa tagttgcagc 3900

atgccattgg gaatggagag taaagcaata tcagatgcac agattactgc ttcatcctac 3960

tttaccaata tgtttgccac ctggtctcct tcaaaagctc gacttcacct ccaagggagg 4020

agtaatgcct ggagacctca ggtgaataat ccaaaagagt ggctgcaagt ggacttccag 4080
aagacaatga aagtcacagg agtaactact cagggagtaa aatctctgct taccagcatg 4140
tatgtgaagg agttcctcat ctccagcagt caagatggcc atcagtggac tctcttttt 4200
cagaatggca aagtaaaggt ttttcaggga aatcaagact ccttcacacc tgtggtgaac 4260
tctctagacc caccgttact gactcgctac cttcgaattc acccccagag ttgggtgcac 4320
cagattgccc tgaggatgga ggttctgggc tgcgaggcac aggacctcta c 4371

<210> 17
<211> 1438
<212> PRT
<213> 人工序列

<220>
<223> BDD-FVIII（未優化；"親代"），胺基酸序列

<400> 17
Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser Trp Asp Tyr
1 5 10 15

Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg Phe Pro Pro
20 25 30

Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val Tyr Lys Lys
35 40 45

Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile Ala Lys Pro
50 55 60

Arg Pro Pro Trp Met Gly Leu Leu Gly Pro Thr Ile Gln Ala Glu Val
65 70 75 80

Tyr Asp Thr Val Val Ile Thr Leu Lys Asn Met Ala Ser His Pro Val
85 90 95

Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ala Ser Glu Gly Ala

```
            100                 105                 110

Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp Asp Lys Val
        115                 120                 125

Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu Lys Glu Asn
    130                 135                 140

Gly Pro Met Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser Tyr Leu Ser
145                 150                 155                 160

His Val Asp Leu Val Lys Asp Leu Asn Ser Gly Leu Ile Gly Ala Leu
                165                 170                 175

Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr Gln Thr Leu
            180                 185                 190

His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly Lys Ser Trp
        195                 200                 205

His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp Ala Ala Ser
    210                 215                 220

Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr Val Asn Arg
225                 230                 235                 240

Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val Tyr Trp His
                245                 250                 255

Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile Phe Leu Glu
            260                 265                 270

Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser Leu Glu Ile
        275                 280                 285

Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met Asp Leu Gly
```

```
    290                 295                 300

Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His Asp Gly Met
305                 310                 315                 320

Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro Gln Leu Arg
                325                 330                 335

Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp Leu Thr Asp
            340                 345                 350

Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser Pro Ser Phe
        355                 360                 365

Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr Trp Val His
    370                 375                 380

Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro Leu Val Leu
385                 390                 395                 400

Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn Asn Gly Pro
                405                 410                 415

Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met Ala Tyr Thr
            420                 425                 430

Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu Ser Gly Ile
        435                 440                 445

Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu Leu Ile Ile
    450                 455                 460

Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro His Gly Ile
465                 470                 475                 480

Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys Gly Val Lys
```

485 490 495

His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe Lys Tyr Lys
500 505 510

Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp Pro Arg Cys
515 520 525

Leu Thr Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg Asp Leu Ala
530 535 540

Ser Gly Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu Ser Val Asp
545 550 555 560

Gln Arg Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val Ile Leu Phe
565 570 575

Ser Val Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu Asn Ile Gln
580 585 590

Arg Phe Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp Pro Glu Phe
595 600 605

Gln Ala Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val Phe Asp Ser
610 615 620

Leu Gln Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp Tyr Ile Leu
625 630 635 640

Ser Ile Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe Ser Gly Tyr
645 650 655

Thr Phe Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr Leu Phe Pro
660 665 670

Phe Ser Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro Gly Leu Trp

675 680 685

Ile Leu Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly Met Thr Ala
690 695 700

Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp Tyr Tyr Glu
705 710 715 720

Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys Asn Asn Ala
725 730 735

Ile Glu Pro Arg Ser Phe Ser Gln Asn Pro Pro Val Leu Lys Arg His
740 745 750

Gln Arg Glu Ile Thr Arg Thr Thr Leu Gln Ser Asp Gln Glu Glu Ile
755 760 765

Asp Tyr Asp Asp Thr Ile Ser Val Glu Met Lys Lys Glu Asp Phe Asp
770 775 780

Ile Tyr Asp Glu Asp Glu Asn Gln Ser Pro Arg Ser Phe Gln Lys Lys
785 790 795 800

Thr Arg His Tyr Phe Ile Ala Ala Val Glu Arg Leu Trp Asp Tyr Gly
805 810 815

Met Ser Ser Ser Pro His Val Leu Arg Asn Arg Ala Gln Ser Gly Ser
820 825 830

Val Pro Gln Phe Lys Lys Val Val Phe Gln Glu Phe Thr Asp Gly Ser
835 840 845

Phe Thr Gln Pro Leu Tyr Arg Gly Glu Leu Asn Glu His Leu Gly Leu
850 855 860

Leu Gly Pro Tyr Ile Arg Ala Glu Val Glu Asp Asn Ile Met Val Thr

865 870 875 880

Phe Arg Asn Gln Ala Ser Arg Pro Tyr Ser Phe Tyr Ser Ser Leu Ile
885 890 895

Ser Tyr Glu Glu Asp Gln Arg Gln Gly Ala Glu Pro Arg Lys Asn Phe
900 905 910

Val Lys Pro Asn Glu Thr Lys Thr Tyr Phe Trp Lys Val Gln His His
915 920 925

Met Ala Pro Thr Lys Asp Glu Phe Asp Cys Lys Ala Trp Ala Tyr Phe
930 935 940

Ser Asp Val Asp Leu Glu Lys Asp Val His Ser Gly Leu Ile Gly Pro
945 950 955 960

Leu Leu Val Cys His Thr Asn Thr Leu Asn Pro Ala His Gly Arg Gln
965 970 975

Val Thr Val Gln Glu Phe Ala Leu Phe Phe Thr Ile Phe Asp Glu Thr
980 985 990

Lys Ser Trp Tyr Phe Thr Glu Asn Met Glu Arg Asn Cys Arg Ala Pro
995 1000 1005

Cys Asn Ile Gln Met Glu Asp Pro Thr Phe Lys Glu Asn Tyr Arg
1010 1015 1020

Phe His Ala Ile Asn Gly Tyr Ile Met Asp Thr Leu Pro Gly Leu
1025 1030 1035

Val Met Ala Gln Asp Gln Arg Ile Arg Trp Tyr Leu Leu Ser Met
1040 1045 1050

Gly Ser Asn Glu Asn Ile His Ser Ile His Phe Ser Gly His Val

1055 1060 1065

Phe Thr Val Arg Lys Lys Glu Glu Tyr Lys Met Ala Leu Tyr Asn
1070 1075 1080

Leu Tyr Pro Gly Val Phe Glu Thr Val Glu Met Leu Pro Ser Lys
1085 1090 1095

Ala Gly Ile Trp Arg Val Glu Cys Leu Ile Gly Glu His Leu His
1100 1105 1110

Ala Gly Met Ser Thr Leu Phe Leu Val Tyr Ser Asn Lys Cys Gln
1115 1120 1125

Thr Pro Leu Gly Met Ala Ser Gly His Ile Arg Asp Phe Gln Ile
1130 1135 1140

Thr Ala Ser Gly Gln Tyr Gly Gln Trp Ala Pro Lys Leu Ala Arg
1145 1150 1155

Leu His Tyr Ser Gly Ser Ile Asn Ala Trp Ser Thr Lys Glu Pro
1160 1165 1170

Phe Ser Trp Ile Lys Val Asp Leu Leu Ala Pro Met Ile Ile His
1175 1180 1185

Gly Ile Lys Thr Gln Gly Ala Arg Gln Lys Phe Ser Ser Leu Tyr
1190 1195 1200

Ile Ser Gln Phe Ile Ile Met Tyr Ser Leu Asp Gly Lys Lys Trp
1205 1210 1215

Gln Thr Tyr Arg Gly Asn Ser Thr Gly Thr Leu Met Val Phe Phe
1220 1225 1230

Gly Asn Val Asp Ser Ser Gly Ile Lys His Asn Ile Phe Asn Pro

1235 1240 1245

Pro Ile Ile Ala Arg Tyr Ile Arg Leu His Pro Thr His Tyr Ser
1250 1255 1260

Ile Arg Ser Thr Leu Arg Met Glu Leu Met Gly Cys Asp Leu Asn
1265 1270 1275

Ser Cys Ser Met Pro Leu Gly Met Glu Ser Lys Ala Ile Ser Asp
1280 1285 1290

Ala Gln Ile Thr Ala Ser Ser Tyr Phe Thr Asn Met Phe Ala Thr
1295 1300 1305

Trp Ser Pro Ser Lys Ala Arg Leu His Leu Gln Gly Arg Ser Asn
1310 1315 1320

Ala Trp Arg Pro Gln Val Asn Asn Pro Lys Glu Trp Leu Gln Val
1325 1330 1335

Asp Phe Gln Lys Thr Met Lys Val Thr Gly Val Thr Thr Gln Gly
1340 1345 1350

Val Lys Ser Leu Leu Thr Ser Met Tyr Val Lys Glu Phe Leu Ile
1355 1360 1365

Ser Ser Ser Gln Asp Gly His Gln Trp Thr Leu Phe Phe Gln Asn
1370 1375 1380

Gly Lys Val Lys Val Phe Gln Gly Asn Gln Asp Ser Phe Thr Pro
1385 1390 1395

Val Val Asn Ser Leu Asp Pro Pro Leu Leu Thr Arg Tyr Leu Arg
1400 1405 1410

Ile His Pro Gln Ser Trp Val His Gln Ile Ala Leu Arg Met Glu

1415 1420 1425

Val Leu Gly Cys Glu Ala Gln Asp Leu Tyr
1430 1435

<210> 18
<211> 450
<212> DNA
<213> 人工序列

<220>
<223> XTEN核苷酸序列

<400> 18
ggcgcgccaa catcagagag cgccaccct gaaagtggtc ccgggagcga gccagccaca 60
tctgggtcgg aaacgccagg cacaagtgag tctgcaactc ccgagtccgg acctggctcc 120
gagcctgcca ctagcggctc cgagactccg ggaacttccg agagcgctac accagaaagc 180
ggacccggaa ccagtaccga acctagcgag ggctctgctc cgggcagccc agccggctct 240
cctacatcca cggaggaggg cacttccgaa tccgccaccc cggagtcagg gccaggatct 300
gaacccgcta cctcaggcag tgagacgcca ggaacgagcg agtccgctac accggagagt 360
gggccaggga gccctgctgg atctcctacg tccactgagg aagggtcacc agcgggctcg 420
cccaccagca ctgaagaagg tgcctcgagc 450

<210> 19
<211> 4824
<212> DNA
<213> 人工序列

<220>
<223> coFVIII-52-XTEN

<400> 19
atgcaaatcg aactgagcac ctgtttcttc ctctgcctgc tgagattctg tttctccgcg 60
acccgccgat actacctggg agcagtggag ctctcctggg attacatgca gagcgacctt 120

ggggagctgc ccgtggatgc caggttccct ccccgggtgc caaagtcgtt tccgttcaac 180
acctccgtgg tgtacaagaa aactctgttc gtggagttca ccgaccacct gttcaatatc 240
gccaagccca gacctccctg gatggggctg ttgggaccta ccatccaagc ggaggtgtac 300
gacactgtgg tcatcactct gaagaacatg gcctcgcatc ccgtgtccct gcacgccgtg 360
ggagtgtctt actggaaagc gtccgagggg gccgaatacg acgaccagac ctcgcagaga 420
gaaaaggaag atgacaaggt gttcccagga ggatcgcaca cctacgtgtg gcaagtgttg 480
aaggagaacg gcccaatggc ctccgacccg ctgtgcctga cctactcgta cctgtcccac 540
gtggacctcg tgaaggacct caactcggga ctgattggag ccctgctggt ctgcagggaa 600
ggctcactgg cgaaagaaaa gactcagacc ttgcacaagt tcattctgct gttcgctgtg 660
ttcgacgagg ggaagtcgtg gcacagcgag actaagaact ccctgatgca agatagagat 720
gccgcctccg cccgggcctg gcctaagatg cacaccgtga acggttacgt gaaccgctcc 780
ctccctggcc tgattggatg ccaccggaag tccgtgtact ggcacgtgat cgggatgggg 840
accacccccg aggtgcacag catcttcctg gaaggtcaca catttctcgt gcgcaaccac 900
cggcaggcct ccctggaaat cagccccatt accttcctca ctgcccagac tctgctgatg 960
gacctgggac agttcctgct gttctgccat atctcctccc accaacatga cggaatggag 1020
gcatacgtga aggtcgattc ctgccctgag gaaccccagc tccgcatgaa gaacaatgag 1080
gaagccgagg actacgacga cgacctgacg gatagcgaga tggatgtggt ccggttcgat 1140
gacgataaca gcccttcctt catccaaatt cgctcggtgg caaagaagca ccccaagacc 1200
tgggtgcatt acattgcggc ggaagaagag gactgggatt atgccccgct tgtcctcgct 1260
cctgacgacc ggagctacaa gagccagtac ctgaacaacg gtccacagag gatcggtaga 1320
aagtacaaga aggtccgctt catggcctat accgacgaaa ccttcaaaac tagagaggcc 1380
atccaacacg aatccggcat cctgggcccg ctcttgtacg gagaagtcgg cgacaccctt 1440
ctcattatct tcaagaacca ggcttcccgg ccgtacaaca tctatccgca tgggatcact 1500
gacgtgcgcc cactgtactc gcggcgcctg cccaagggtg tcaaacacct gaaggatttt 1560
ccgatccttc cgggagaaat cttcaagtac aagtggaccg tgaccgtgga agatggccca 1620
actaagtctg accctagatg cctcacccgc tactactcat ccttcgtcaa catggagcgc 1680
gacctggcca gcggactgat cggcccgctg ctgatttgct acaaggaatc agtggaccaa 1740
cggggaaacc agatcatgtc ggataagagg aacgtcatcc tcttctccgt gtttgacgaa 1800
aaccggtcgt ggtacctgac cgagaacatc cagaggttcc tgcccaaccc tgctggggtg 1860
cagctggagg accccgagtt ccaggccagc aacatcatgc acagcatcaa tggctacgtg 1920
ttcgacagcc tgcagctgag cgtgtgcctg cacgaggtgg cctactggta catcctgagc 1980
atcggcgccc agaccgactt cctgagcgtg ttcttctctg gctacacctt caagcacaag 2040
atggtgtatg aggacaccct gaccctgttc cccttcagcg gggagactgt cttcatgagc 2100
atggagaacc ctggcctgtg gatcctgggc tgccacaaca gcgacttcag gaacaggggc 2160
atgactgccc tgctgaaagt ctccagctgt gacaagaaca ccggggacta ctacgaggac 2220
agctacgagg acatcagcgc ctacctgctg agcaagaaca atgccatcga gcccaggagc 2280

ttctctcaga acggcgcgcc aacatcagag agcgccaccc ctgaaagtgg tcccgggagc 2340

gagccagcca catctgggtc ggaaacgcca ggcacaagtg agtctgcaac tcccgagtcc 2400

ggacctggct ccgagcctgc cactagcggc tccgagactc cgggaacttc cgagagcgct 2460

acaccagaaa gcggacccgg aaccagtacc gaacctagcg agggctctgc tccgggcagc 2520

ccagccggct ctcctacatc cacggaggag ggcacttccg aatccgccac cccggagtca 2580

gggccaggat ctgaacccgc tacctcaggc agtgagacgc caggaacgag cgagtccgct 2640

acaccggaga gtgggccagg gagccctgct ggatctccta cgtccactga ggaagggtca 2700

ccagcgggct cgcccaccag cactgaagaa ggtgcctcga gccccccagt gctgaagagg 2760

caccagaggg agatcaccag gaccaccctg cagtctgacc aggaggagat cgactatgat 2820

gacaccatca gcgtggagat gaagaaggag gacttcgaca tctacgacga ggacgagaac 2880

cagagcccca ggagcttcca gaagaagacc aggcactact tcattgctgc tgtggagagg 2940

ctgtgggact atggcatgtc cagcagcccc catgtgctga ggaacagggc ccagtctggc 3000

agcgtgcccc agttcaagaa agtcgtgttc caggagttca ccgacggcag cttcacccag 3060

cccctgtaca gaggggagct gaacgagcac ctgggcctgc tgggccccta catcagggcc 3120

gaggtggagg acaacatcat ggtgaccttc aggaaccagg ccagcaggcc ctacagcttc 3180

tacagcagcc tgatcagcta cgaggaggac cagaggcagg gggctgagcc caggaagaac 3240

tttgtgaagc ccaatgaaac caagacctac ttctggaagg tgcagcacca catggccccc 3300

accaaggacg agttcgactg caaggcctgg gcctacttct ctgacgtgga cctggagaag 3360

gacgtgcact ctggcctgat tggccccctg ctggtgtgcc acaccaacac cctgaaccct 3420

gcccatggca ggcaggtgac tgtgcaggag ttcgccctgt tcttcaccat cttcgatgaa 3480

accaagagct ggtacttcac tgagaacatg gagaggaact gcagggcccc ctgcaacatc 3540

cagatggagg accccacctt caaggagaac tacaggttcc atgccatcaa tggctacatc 3600

atggacaccc tgcctggcct ggtcatggcc caggaccaga ggatcaggtg gtatctgctg 3660

agcatgggca gcaacgagaa catccacagc atccacttct ctggccacgt gttcactgtg 3720

aggaagaagg aggagtacaa gatggccctg tacaacctgt accctggggt gttcgaaacc 3780

gtggagatgc tgcccagcaa ggccggcatc tggagggtgg agtgcctgat tggggagcac 3840

ctgcacgccg gcatgagcac cctgttcctg gtgtacagca acaagtgcca gacccccctg 3900

ggcatggcct ctggccacat cagggacttc cagatcactg cctctggcca gtacggccag 3960

tgggccccca agctggccag gctgcactac tccggaagca tcaatgcctg gagcaccaag 4020

gagcccttca gctggatcaa agtggacctg ctggccccca tgatcatcca cggcatcaag 4080

acccaggggg ccaggcagaa gttctccagc ctgtacatca gccagttcat catcatgtac 4140

agcctggacg gcaagaagtg gcagacctac aggggcaaca gcaccggcac cctgatggtg 4200

ttcttcggca acgtggacag cagcggcatc aagcacaaca tcttcaaccc ccccatcatc 4260

gccagataca tcaggctgca ccccacccac tacagcatca ggagcaccct gaggatggag 4320

ctgatgggct gtgacctgaa cagctgcagc atgcccctgg gcatggagag caaggccatc 4380

tctgacgccc agatcactgc ctccagctac ttcaccaaca tgtttgccac ctggagcccc 4440

agcaaggcca ggctgcacct gcagggcagg agcaatgcct ggaggcccca ggtcaacaac 4500
cccaaggagt ggctgcaggt ggacttccag aagaccatga aggtgactgg ggtgaccacc 4560
caggggggtga agagcctgct gaccagcatg tacgtgaagg agttcctgat ctccagcagc 4620
caggacggcc accagtggac cctgttcttc cagaatggca aggtgaaggt gttccagggc 4680
aaccaggaca gcttcacccc tgtggtcaac agcctggacc ccccccctgct gaccagatac 4740
ctgaggatcc acccccagag ctgggtgcac cagatcgccc tgaggatgga ggtgctgggc 4800
tgtgaggccc aggacctgta ctga 4824

<210> 20
<211> 4824
<212> DNA
<213> 人工序列

<220>
<223> coFVIII-1-XTEN

<400> 20
atgcagattg agctgtctac ttgctttttc ctgtgcctgc tgaggttttg cttttccgct 60
acacgaaggt attatctggg ggctgtggaa ctgtcttggg attacatgca gagtgacctg 120
ggagagctgc cagtggacgc aaggtttccc cctagagtcc ctaagtcatt ccccttcaac 180
actagcgtgg tctacaagaa aacactgttc gtggagttta ctgatcacct gttcaacatc 240
gcaaagccta ggccaccctg gatgggactg ctggggccaa caatccaggc cgaggtgtac 300
gacaccgtgg tcattacact taagaacatg gcctcacacc ccgtgagcct gcatgctgtg 360
ggcgtcagct actggaaggc ttccgaagga gcagagtatg acgatcagac ttcccagaga 420
gaaaaagagg acgataaggt gtttcctggc ggatctcata cctacgtgtg gcaggtcctg 480
aaagagaatg gccctatggc ctccgaccct ctgtgcctga cctactctta tctgagtcac 540
gtggacctgg tcaaggatct gaacagcggc ctgatcggag ccctgctggt gtgcagggaa 600
ggaagcctgg ctaaggagaa aacccagaca ctgcataagt tcattctgct gttcgccgtg 660
tttgacgaag ggaaatcatg gcacagcgag acaaagaata gtctgatgca ggacagggat 720
gccgcttcag ccagagcttg gcccaaaatg cacactgtga acggctacgt caatcgctca 780
ctgcctgggc tgatcggctg ccaccgaaag agcgtgtatt ggcatgtcat cgggatgggc 840
accacacctg aagtgcactc cattttcctg gagggacata cctttctggt ccgcaaccac 900
cgacaggctt ccctggagat ctctccaatt accttcctga cagcacagac tctgctgatg 960
gacctggggc agttcctgct gttttgccac atcagctccc accagcatga tggcatggag 1020
gcttacgtga aagtggactc ttgtcccgag gaacctcagc tgcggatgaa gaacaatgag 1080

gaagcagaag actatgacga tgacctgacc gactccgaga tggatgtggt ccgattcgat 1140
gacgataaca gcccctcctt tatccagatt agatctgtgg ccaagaaaca ccctaagaca 1200
tgggtccatt acatcgcagc cgaggaagag gactgggatt atgcaccact ggtgctggca 1260
ccagacgatc gctcctacaa atctcagtat ctgaacaatg ggccacagag gattggcaga 1320
aagtacaaga aagtgcggtt catggcatat accgatgaga ccttcaagac tcgcgaagcc 1380
atccagcacg agagcggcat cctgggacca ctgctgtacg gagaagtggg agacaccctg 1440
ctgatcattt tcaagaacca ggccagccgg ccttacaata tctatccaca tgggattaca 1500
gatgtgcgcc ctctgtacag caggagactg ccaaagggcg tcaaacacct gaaggacttc 1560
ccaatcctgc ccggagaaat cttcaagtac aagtggactg tcaccgtcga ggatggcccc 1620
actaagagcg accctcggtg cctgacccgc tactattcta gtttcgtgaa tatggaaaga 1680
gatctggcaa gcggactgat cggaccactg ctgatttgtt acaaagagag cgtggatcag 1740
agaggcaacc agatcatgtc cgacaagcgg aatgtgattc tgttcagtgt ctttgacgaa 1800
aacaggtcat ggtacctgac cgagaacatc cagagattcc tgcctaatcc agctggggtg 1860
cagctggaag atcctgagtt tcaggcatct aacatcatgc atagtattaa tggctacgtg 1920
ttcgacagtt tgcagctgag cgtgtgcctg cacgaggtcg cttactggta tatcctgagc 1980
attggggcac agacagattt cctgagcgtg ttcttttccg gctacacttt taagcataaa 2040
atggtctatg aggacacact gactctgttc cccttcagcg gcgaaaccgt gtttatgagc 2100
atggagaatc ccggactgtg gattctgggg tgccacaaca gcgatttcag aaatcgcgga 2160
atgactgccc tgctgaaagt gtcaagctgt gacaagaaca ccggggacta ctatgaagat 2220
tcatacgagg acatcagcgc atatctgctg tccaaaaaca atgccattga accccggtct 2280
tttagtcaga atggcgcgcc aacatcagag agcgccaccc ctgaaagtgg tcccgggagc 2340
gagccagcca catctgggtc ggaaacgcca ggcacaagtg agtctgcaac tcccgagtcc 2400
ggacctggct ccgagcctgc cactagcggc tccgagactc cgggaacttc cgagagcgct 2460
acaccagaaa gcggacccgg aaccagtacc gaacctagcg agggctctgc tccgggcagc 2520
ccagccggct ctcctacatc cacggaggag ggcacttccg aatccgccac cccggagtca 2580
gggccaggat ctgaacccgc tacctcaggc agtgagacgc caggaacgag cgagtccgct 2640
acaccggaga gtgggccagg gagccctgct ggatctccta cgtccactga ggaagggtca 2700
ccagcgggct cgcccaccag cactgaagaa ggtgcctcga gccctccagt gctgaagcgg 2760
caccagcgcg agatcacccg cactaccctg cagagtgatc aggaagagat cgactacgac 2820
gatacaattt ctgtggaaat gaagaaagag gacttcgata tctatgacga agatgagaac 2880
cagagtcctc gatcattcca gaagaaaacc aggcattact ttattgccgc agtggagcgg 2940
ctgtgggatt atggcatgtc ctctagtcct cacgtgctgc gaaatagggc ccagtcagga 3000
agcgtcccac agttcaagaa agtggtcttc caggagttta cagacgggtc ctttactcag 3060
ccactgtaca ggggcgaact gaacgagcac ctgggactgc tggggcccta tatcagagca 3120
gaagtggagg ataacattat ggtcaccttc agaaatcagg cctctcggcc ttacagtttt 3180
tattcaagcc tgatctctta cgaagaggac cagcgacagg gagctgaacc acgaaaaaac 3240

ttcgtgaagc ctaatgagac caaaacatac ttttggaagg tgcagcacca tatggcccca 3300
acaaaagacg agttcgattg caaggcatgg gcctattttt ctgacgtgga tctggagaag 3360
gacgtgcaca gtggcctgat tggcccactg ctggtgtgcc atactaacac cctgaatcca 3420
gcccacggcc ggcaggtcac tgtccaggag ttcgctctgt tctttaccat ctttgatgag 3480
acaaagagct ggtacttcac cgaaaacatg gagcgaaatt gcagggctcc atgtaacatt 3540
cagatggaag accccacatt caaggagaac taccgctttc atgctatcaa tggatacatc 3600
atggatactc tgcccgggct ggtcatggca caggaccaga gaatccggtg gtatctgctg 3660
agcatgggca gcaacgagaa tatccactca attcatttca gcgggcacgt gtttactgtc 3720
aggaagaaag aagagtacaa gatggccctg tacaacctgt atcccggcgt gttcgaaacc 3780
gtcgagatgc tgcctagcaa ggccggaatc tggagagtgg aatgcctgat tggagagcac 3840
ctgcatgctg ggatgtctac cctgtttctg gtgtacagta ataagtgtca gacacccctg 3900
ggaatggcat ccgggcatat cagggatttc cagattaccg catctggaca gtacggacag 3960
tgggcaccta agctggctag actgcactat tccggatcta tcaacgcttg gtccacaaaa 4020
gagcctttct cttggattaa ggtggacctg ctggccccaa tgatcattca tggcatcaaa 4080
actcagggag ctcggcagaa gttctcctct ctgtacatct cacagtttat catcatgtac 4140
agcctggatg ggaagaaatg gcagacatac cgcggcaata gcacaggaac tctgatggtg 4200
ttctttggca acgtggacag cagcggaatc aagcacaaca ttttcaatcc ccctatcatt 4260
gctagataca tccggctgca cccaacccat tattctattc gaagtacact gaggatggaa 4320
ctgatgggat gcgatctgaa cagttgttca atgcccctgg ggatggagtc caaggcaatc 4380
tctgacgccc agattaccgc cagctcctac ttcactaata tgtttgctac ctggagccct 4440
tccaaagcaa gactgcacct gcaaggccgc agcaacgcat ggcgaccaca ggtgaacaat 4500
cccaaggagt ggttgcaggt cgattttcag aaaactatga aggtgaccgg ggtcacaact 4560
cagggcgtga aaagtctgct gacctcaatg tacgtcaagg agttcctgat ctctagttca 4620
caggacggac atcagtggac actgttcttt cagaacggga aggtgaaagt cttccagggc 4680
aatcaggatt cctttacacc tgtggtcaac agtctagacc ctccactgct gaccagatac 4740
ctgagaatcc accctcagtc ctgggtgcac cagattgccc tgagaatgga agtgctggga 4800
tgcgaggccc aggatctgta ctga 4824

<210> 21
<211> 6
<212> DNA
<213> 人工序列

<220>
<223> MAR/ARS

<400> 21
atattt 6

<210> 22
<211> 6
<212> DNA
<213> 人工序列

<220>
<223> MAR/ARS

<400> 22
aaatat 6

<210> 23
<211> 5
<212> DNA
<213> 人工序列

<220>
<223> 去穩定元件

<400> 23
attta 5

<210> 24
<211> 5
<212> DNA
<213> 人工序列

<220>
<223> 去穩定元件

<400> 24

taaat 5

<210> 25

<211> 6

<212> DNA

<213> 人工序列

<220>

<223> 多聚T位點

<400> 25

tttttt 6

<210> 26

<211> 7

<212> DNA

<213> 人工序列

<220>

<223> 多聚A位點

<400> 26

aaaaaaa 7

<210> 27

<211> 6

<212> DNA

<213> 人工序列

<220>

<223> 剪接位點

<400> 27
ggtgat 6

<210> 28
<211> 5
<212> DNA
<213> 人工序列

<220>
<223> TATA盒

<400> 28
tataa 5

<210> 29
<211> 5
<212> DNA
<213> 人工序列

<220>
<223> TATA盒

<400> 29
ttata 5

<210> 30
<211> 8
<212> DNA
<213> 人工序列

<220>
<223> 富含AU的序列元件

<400> 30
attttatt 8

<210> 31
<211> 8
<212> DNA
<213> 人工序列

<220>
<223> 富含AU的序列元件

<400> 31
atttttaa 8

<210> 32
<211> 13
<212> DNA
<213> 人工序列

<220>
<223> kozak共有序列

<400> 32
gccgccacca tgc 13

<210> 33
<211> 32
<212> PRT
<213> 人工序列

<220>
<223> CTP肽

<400> 33

Asp Pro Arg Phe Gln Asp Ser Ser Ser Ser Lys Ala Pro Pro Pro Ser
1 5 10 15

Leu Pro Ser Pro Ser Arg Leu Pro Gly Pro Ser Asp Thr Pro Ile Leu
20 25 30

<210> 34

<211> 28

<212> PRT

<213> 人工序列

<220>

<223> CTP肽

<400> 34

Ser Ser Ser Ser Lys Ala Pro Pro Pro Ser Leu Pro Ser Pro Ser Arg
1 5 10 15

Leu Pro Gly Pro Ser Asp Thr Pro Ile Leu Pro Gln
20 25

<210> 35

<211> 11

<212> PRT

<213> 人工序列

<220>

<223> 白蛋白結合肽核心序列

<400> 35

Asp Ile Cys Leu Pro Arg Trp Gly Cys Leu Trp
1 5 10

<210> 36
<211> 20
<212> PRT
<213> 人工序列

<220>
<223> PAS序列

<400> 36
Ala Ser Pro Ala Ala Pro Ala Pro Ala Ser Pro Ala Ala Pro Ala Pro
1 5 10 15

Ser Ala Pro Ala
20

<210> 37
<211> 20
<212> PRT
<213> 人工序列

<220>
<223> PAS序列

<400> 37
Ala Ala Pro Ala Ser Pro Ala Pro Ala Ala Pro Ser Ala Pro Ala Pro
1 5 10 15

Ala Ala Pro Ser
20

<210> 38
<211> 20

<212> PRT
<213> 人工序列

<220>
<223> PAS序列

<400> 38
Ala Pro Ser Ser Pro Ser Pro Ser Ala Pro Ser Ser Pro Ser Pro Ala
1 5 10 15

Ser Pro Ser Ser
20

<210> 39
<211> 19
<212> PRT
<213> 人工序列

<220>
<223> PAS序列

<400> 39
Ala Pro Ser Ser Pro Ser Pro Ser Ala Pro Ser Ser Pro Ser Pro Ala
1 5 10 15

Ser Pro Ser

<210> 40
<211> 20
<212> PRT
<213> 人工序列

<220>

<223> PAS序列

<400> 40

Ser Ser Pro Ser Ala Pro Ser Pro Ser Ser Pro Ala Ser Pro Ser Pro
1 5 10 15

Ser Ser Pro Ala
20

<210> 41
<211> 24
<212> PRT
<213> 人工序列

<220>
<223> PAS序列

<400> 41

Ala Ala Ser Pro Ala Ala Pro Ser Ala Pro Pro Ala Ala Ala Ser Pro
1 5 10 15

Ala Ala Pro Ser Ala Pro Pro Ala
20

<210> 42
<211> 20
<212> PRT
<213> 人工序列

<220>
<223> PAS序列

<400> 42

Ala Ser Ala Ala Ala Pro Ala Ala Ala Ser Ala Ala Ala Ser Ala Pro

1 5 10 15

Ser Ala Ala Ala
20

<210> 43
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> miR142靶標

<400> 43
tccataaagt aggaaacact aca 23

<210> 44
<211> 2332
<212> PRT
<213> 智人

<400> 44
Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser Trp Asp Tyr
1 5 10 15

Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg Phe Pro Pro
20 25 30

Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val Tyr Lys Lys
35 40 45

Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile Ala Lys Pro
50 55 60

```
Arg Pro Pro Trp Met Gly Leu Leu Gly Pro Thr Ile Gln Ala Glu Val
65                  70                  75                  80

Tyr Asp Thr Val Val Ile Thr Leu Lys Asn Met Ala Ser His Pro Val
                85                  90                  95

Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ala Ser Glu Gly Ala
            100                 105                 110

Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp Asp Lys Val
        115                 120                 125

Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu Lys Glu Asn
    130                 135                 140

Gly Pro Met Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser Tyr Leu Ser
145                 150                 155                 160

His Val Asp Leu Val Lys Asp Leu Asn Ser Gly Leu Ile Gly Ala Leu
                165                 170                 175

Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr Gln Thr Leu
            180                 185                 190

His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly Lys Ser Trp
        195                 200                 205

His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp Ala Ala Ser
    210                 215                 220

Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr Val Asn Arg
225                 230                 235                 240

Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val Tyr Trp His
                245                 250                 255
```

```
Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile Phe Leu Glu
            260                 265                 270

Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser Leu Glu Ile
        275                 280                 285

Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met Asp Leu Gly
    290                 295                 300

Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His Asp Gly Met
305                 310                 315                 320

Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro Gln Leu Arg
                325                 330                 335

Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp Leu Thr Asp
            340                 345                 350

Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser Pro Ser Phe
        355                 360                 365

Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr Trp Val His
    370                 375                 380

Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro Leu Val Leu
385                 390                 395                 400

Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn Asn Gly Pro
                405                 410                 415

Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met Ala Tyr Thr
            420                 425                 430

Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu Ser Gly Ile
        435                 440                 445
```

```
Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu Leu Ile Ile
    450                 455                 460

Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro His Gly Ile
465                 470                 475                 480

Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys Gly Val Lys
                485                 490                 495

His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe Lys Tyr Lys
            500                 505                 510

Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp Pro Arg Cys
        515                 520                 525

Leu Thr Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg Asp Leu Ala
    530                 535                 540

Ser Gly Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu Ser Val Asp
545                 550                 555                 560

Gln Arg Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val Ile Leu Phe
                565                 570                 575

Ser Val Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu Asn Ile Gln
            580                 585                 590

Arg Phe Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp Pro Glu Phe
        595                 600                 605

Gln Ala Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val Phe Asp Ser
    610                 615                 620

Leu Gln Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp Tyr Ile Leu
625                 630                 635                 640
```

Ser Ile Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe Ser Gly Tyr
645 650 655

Thr Phe Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr Leu Phe Pro
660 665 670

Phe Ser Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro Gly Leu Trp
675 680 685

Ile Leu Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly Met Thr Ala
690 695 700

Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp Tyr Tyr Glu
705 710 715 720

Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys Asn Asn Ala
725 730 735

Ile Glu Pro Arg Ser Phe Ser Gln Asn Ser Arg His Pro Ser Thr Arg
740 745 750

Gln Lys Gln Phe Asn Ala Thr Thr Ile Pro Glu Asn Asp Ile Glu Lys
755 760 765

Thr Asp Pro Trp Phe Ala His Arg Thr Pro Met Pro Lys Ile Gln Asn
770 775 780

Val Ser Ser Ser Asp Leu Leu Met Leu Leu Arg Gln Ser Pro Thr Pro
785 790 795 800

His Gly Leu Ser Leu Ser Asp Leu Gln Glu Ala Lys Tyr Glu Thr Phe
805 810 815

Ser Asp Asp Pro Ser Pro Gly Ala Ile Asp Ser Asn Asn Ser Leu Ser
820 825 830

```
Glu Met Thr His Phe Arg Pro Gln Leu His His Ser Gly Asp Met Val
        835                 840                 845

Phe Thr Pro Glu Ser Gly Leu Gln Leu Arg Leu Asn Glu Lys Leu Gly
    850                 855                 860

Thr Thr Ala Ala Thr Glu Leu Lys Lys Leu Asp Phe Lys Val Ser Ser
865                 870                 875                 880

Thr Ser Asn Asn Leu Ile Ser Thr Ile Pro Ser Asp Asn Leu Ala Ala
                885                 890                 895

Gly Thr Asp Asn Thr Ser Ser Leu Gly Pro Pro Ser Met Pro Val His
            900                 905                 910

Tyr Asp Ser Gln Leu Asp Thr Thr Leu Phe Gly Lys Lys Ser Ser Pro
        915                 920                 925

Leu Thr Glu Ser Gly Gly Pro Leu Ser Leu Ser Glu Glu Asn Asn Asp
    930                 935                 940

Ser Lys Leu Leu Glu Ser Gly Leu Met Asn Ser Gln Glu Ser Ser Trp
945                 950                 955                 960

Gly Lys Asn Val Ser Ser Thr Glu Ser Gly Arg Leu Phe Lys Gly Lys
                965                 970                 975

Arg Ala His Gly Pro Ala Leu Leu Thr Lys Asp Asn Ala Leu Phe Lys
            980                 985                 990

Val Ser Ile Ser Leu Leu Lys Thr  Asn Lys Thr Ser Asn  Asn Ser Ala
        995                 1000                1005

Thr Asn  Arg Lys Thr His Ile  Asp Gly Pro Ser Leu  Leu Ile Glu
    1010                1015                1020
```

```
Asn Ser  Pro Ser Val Trp Gln  Asn Ile Leu Glu Ser  Asp Thr Glu
    1025                1030                1035

Phe Lys  Lys Val Thr Pro Leu  Ile His Asp Arg Met  Leu Met Asp
    1040                1045                1050

Lys Asn  Ala Thr Ala Leu Arg  Leu Asn His Met Ser  Asn Lys Thr
    1055                1060                1065

Thr Ser  Ser Lys Asn Met Glu  Met Val Gln Gln Lys  Lys Glu Gly
    1070                1075                1080

Pro Ile  Pro Pro Asp Ala Gln  Asn Pro Asp Met Ser  Phe Phe Lys
    1085                1090                1095

Met Leu  Phe Leu Pro Glu Ser  Ala Arg Trp Ile Gln  Arg Thr His
    1100                1105                1110

Gly Lys  Asn Ser Leu Asn Ser  Gly Gln Gly Pro Ser  Pro Lys Gln
    1115                1120                1125

Leu Val  Ser Leu Gly Pro Glu  Lys Ser Val Glu Gly  Gln Asn Phe
    1130                1135                1140

Leu Ser  Glu Lys Asn Lys Val  Val Val Gly Lys Gly  Glu Phe Thr
    1145                1150                1155

Lys Asp  Val Gly Leu Lys Glu  Met Val Phe Pro Ser  Ser Arg Asn
    1160                1165                1170

Leu Phe  Leu Thr Asn Leu Asp  Asn Leu His Glu Asn  Asn Thr His
    1175                1180                1185

Asn Gln  Glu Lys Lys Ile Gln  Glu Glu Ile Glu Lys  Lys Glu Thr
    1190                1195                1200
```

```
Leu Ile  Gln Glu Asn Val Val  Leu Pro Gln Ile His  Thr Val Thr
    1205                 1210                 1215

Gly Thr  Lys Asn Phe Met Lys  Asn Leu Phe Leu Leu  Ser Thr Arg
    1220                 1225                 1230

Gln Asn  Val Glu Gly Ser Tyr  Asp Gly Ala Tyr Ala  Pro Val Leu
    1235                 1240                 1245

Gln Asp  Phe Arg Ser Leu Asn  Asp Ser Thr Asn Arg  Thr Lys Lys
    1250                 1255                 1260

His Thr  Ala His Phe Ser Lys  Lys Gly Glu Glu Glu  Asn Leu Glu
    1265                 1270                 1275

Gly Leu  Gly Asn Gln Thr Lys  Gln Ile Val Glu Lys  Tyr Ala Cys
    1280                 1285                 1290

Thr Thr  Arg Ile Ser Pro Asn  Thr Ser Gln Gln Asn  Phe Val Thr
    1295                 1300                 1305

Gln Arg  Ser Lys Arg Ala Leu  Lys Gln Phe Arg Leu  Pro Leu Glu
    1310                 1315                 1320

Glu Thr  Glu Leu Glu Lys Arg  Ile Ile Val Asp Asp  Thr Ser Thr
    1325                 1330                 1335

Gln Trp  Ser Lys Asn Met Lys  His Leu Thr Pro Ser  Thr Leu Thr
    1340                 1345                 1350

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

Pro Leu  Ser Asp Cys Leu Thr  Arg Ser His Ser Ile  Pro Gln Ala
    1370                 1375                 1380
```

```
Asn Arg  Ser Pro Leu Pro Ile  Ala Lys Val Ser Ser  Phe Pro Ser
    1385                 1390                 1395

Ile Arg  Pro Ile Tyr Leu Thr  Arg Val Leu Phe Gln  Asp Asn Ser
    1400                 1405                 1410

Ser His  Leu Pro Ala Ala Ser  Tyr Arg Lys Lys Asp  Ser Gly Val
    1415                 1420                 1425

Gln Glu  Ser Ser His Phe Leu  Gln Gly Ala Lys Lys  Asn Asn Leu
    1430                 1435                 1440

Ser Leu  Ala Ile Leu Thr Leu  Glu Met Thr Gly Asp  Gln Arg Glu
    1445                 1450                 1455

Val Gly  Ser Leu Gly Thr Ser  Ala Thr Asn Ser Val  Thr Tyr Lys
    1460                 1465                 1470

Lys Val  Glu Asn Thr Val Leu  Pro Lys Pro Asp Leu  Pro Lys Thr
    1475                 1480                 1485

Ser Gly  Lys Val Glu Leu Leu  Pro Lys Val His Ile  Tyr Gln Lys
    1490                 1495                 1500

Asp Leu  Phe Pro Thr Glu Thr  Ser Asn Gly Ser Pro  Gly His Leu
    1505                 1510                 1515

Asp Leu  Val Glu Gly Ser Leu  Leu Gln Gly Thr Glu  Gly Ala Ile
    1520                 1525                 1530

Lys Trp  Asn Glu Ala Asn Arg  Pro Gly Lys Val Pro  Phe Leu Arg
    1535                 1540                 1545

Val Ala  Thr Glu Ser Ser Ala  Lys Thr Pro Ser Lys  Leu Leu Asp
    1550                 1555                 1560
```

```
Pro Leu  Ala Trp Asp Asn His  Tyr Gly Thr Gln Ile  Pro Lys Glu
    1565                 1570                 1575

Glu Trp  Lys Ser Gln Glu Lys  Ser Pro Glu Lys Thr  Ala Phe Lys
    1580                 1585                 1590

Lys Lys  Asp Thr Ile Leu Ser  Leu Asn Ala Cys Glu  Ser Asn His
    1595                 1600                 1605

Ala Ile  Ala Ala Ile Asn Glu  Gly Gln Asn Lys Pro  Glu Ile Glu
    1610                 1615                 1620

Val Thr  Trp Ala Lys Gln Gly  Arg Thr Glu Arg Leu  Cys Ser Gln
    1625                 1630                 1635

Asn Pro  Pro Val Leu Lys Arg  His Gln Arg Glu Ile  Thr Arg Thr
    1640                 1645                 1650

Thr Leu  Gln Ser Asp Gln Glu  Glu Ile Asp Tyr Asp  Asp Thr Ile
    1655                 1660                 1665

Ser Val  Glu Met Lys Lys Glu  Asp Phe Asp Ile Tyr  Asp Glu Asp
    1670                 1675                 1680

Glu Asn  Gln Ser Pro Arg Ser  Phe Gln Lys Lys Thr  Arg His Tyr
    1685                 1690                 1695

Phe Ile  Ala Ala Val Glu Arg  Leu Trp Asp Tyr Gly  Met Ser Ser
    1700                 1705                 1710

Ser Pro  His Val Leu Arg Asn  Arg Ala Gln Ser Gly  Ser Val Pro
    1715                 1720                 1725

Gln Phe  Lys Lys Val Val Phe  Gln Glu Phe Thr Asp  Gly Ser Phe
    1730                 1735                 1740
```

Thr Gln Pro Leu Tyr Arg Gly Glu Leu Asn Glu His Leu Gly Leu
1745 1750 1755

Leu Gly Pro Tyr Ile Arg Ala Glu Val Glu Asp Asn Ile Met Val
1760 1765 1770

Thr Phe Arg Asn Gln Ala Ser Arg Pro Tyr Ser Phe Tyr Ser Ser
1775 1780 1785

Leu Ile Ser Tyr Glu Glu Asp Gln Arg Gln Gly Ala Glu Pro Arg
1790 1795 1800

Lys Asn Phe Val Lys Pro Asn Glu Thr Lys Thr Tyr Phe Trp Lys
1805 1810 1815

Val Gln His His Met Ala Pro Thr Lys Asp Glu Phe Asp Cys Lys
1820 1825 1830

Ala Trp Ala Tyr Phe Ser Asp Val Asp Leu Glu Lys Asp Val His
1835 1840 1845

Ser Gly Leu Ile Gly Pro Leu Leu Val Cys His Thr Asn Thr Leu
1850 1855 1860

Asn Pro Ala His Gly Arg Gln Val Thr Val Gln Glu Phe Ala Leu
1865 1870 1875

Phe Phe Thr Ile Phe Asp Glu Thr Lys Ser Trp Tyr Phe Thr Glu
1880 1885 1890

Asn Met Glu Arg Asn Cys Arg Ala Pro Cys Asn Ile Gln Met Glu
1895 1900 1905

Asp Pro Thr Phe Lys Glu Asn Tyr Arg Phe His Ala Ile Asn Gly
1910 1915 1920

```
Tyr Ile  Met Asp Thr Leu Pro  Gly Leu Val Met Ala  Gln Asp Gln
    1925                 1930                 1935

Arg Ile  Arg Trp Tyr Leu Leu  Ser Met Gly Ser Asn  Glu Asn Ile
    1940                 1945                 1950

His Ser  Ile His Phe Ser Gly  His Val Phe Thr Val  Arg Lys Lys
    1955                 1960                 1965

Glu Glu  Tyr Lys Met Ala Leu  Tyr Asn Leu Tyr Pro  Gly Val Phe
    1970                 1975                 1980

Glu Thr  Val Glu Met Leu Pro  Ser Lys Ala Gly Ile  Trp Arg Val
    1985                 1990                 1995

Glu Cys  Leu Ile Gly Glu His  Leu His Ala Gly Met  Ser Thr Leu
    2000                 2005                 2010

Phe Leu  Val Tyr Ser Asn Lys  Cys Gln Thr Pro Leu  Gly Met Ala
    2015                 2020                 2025

Ser Gly  His Ile Arg Asp Phe  Gln Ile Thr Ala Ser  Gly Gln Tyr
    2030                 2035                 2040

Gly Gln  Trp Ala Pro Lys Leu  Ala Arg Leu His Tyr  Ser Gly Ser
    2045                 2050                 2055

Ile Asn  Ala Trp Ser Thr Lys  Glu Pro Phe Ser Trp  Ile Lys Val
    2060                 2065                 2070

Asp Leu  Leu Ala Pro Met Ile  Ile His Gly Ile Lys  Thr Gln Gly
    2075                 2080                 2085

Ala Arg  Gln Lys Phe Ser Ser  Leu Tyr Ile Ser Gln  Phe Ile Ile
    2090                 2095                 2100
```

```
Met Tyr  Ser Leu Asp Gly Lys  Lys Trp Gln Thr Tyr  Arg Gly Asn
    2105                 2110                 2115

Ser Thr  Gly Thr Leu Met Val  Phe Phe Gly Asn Val  Asp Ser Ser
    2120                 2125                 2130

Gly Ile  Lys His Asn Ile Phe  Asn Pro Pro Ile Ile  Ala Arg Tyr
    2135                 2140                 2145

Ile Arg  Leu His Pro Thr His  Tyr Ser Ile Arg Ser  Thr Leu Arg
    2150                 2155                 2160

Met Glu  Leu Met Gly Cys Asp  Leu Asn Ser Cys Ser  Met Pro Leu
    2165                 2170                 2175

Gly Met  Glu Ser Lys Ala Ile  Ser Asp Ala Gln Ile  Thr Ala Ser
    2180                 2185                 2190

Ser Tyr  Phe Thr Asn Met Phe  Ala Thr Trp Ser Pro  Ser Lys Ala
    2195                 2200                 2205

Arg Leu  His Leu Gln Gly Arg  Ser Asn Ala Trp Arg  Pro Gln Val
    2210                 2215                 2220

Asn Asn  Pro Lys Glu Trp Leu  Gln Val Asp Phe Gln  Lys Thr Met
    2225                 2230                 2235

Lys Val  Thr Gly Val Thr Thr  Gln Gly Val Lys Ser  Leu Leu Thr
    2240                 2245                 2250

Ser Met  Tyr Val Lys Glu Phe  Leu Ile Ser Ser Ser  Gln Asp Gly
    2255                 2260                 2265

His Gln  Trp Thr Leu Phe Phe  Gln Asn Gly Lys Val  Lys Val Phe
    2270                 2275                 2280
```

Gln Gly Asn Gln Asp Ser Phe Thr Pro Val Val Asn Ser Leu Asp
2285 2290 2295

Pro Pro Leu Leu Thr Arg Tyr Leu Arg Ile His Pro Gln Ser Trp
2300 2305 2310

Val His Gln Ile Ala Leu Arg Met Glu Val Leu Gly Cys Glu Ala
2315 2320 2325

Gln Asp Leu Tyr
2330

<210> 45
<211> 4
<212> PRT
<213> 人工序列

<220>
<223> 人IgG1的胺基酸233-236

<400> 45
Glu Leu Leu Gly
1

<210> 46
<211> 42
<212> PRT
<213> 人工序列

<220>
<223> XTEN AE42-4，蛋白質序列

<400> 46
Gly Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly

1 5 10 15

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala
20 25 30

Thr Ser Gly Ser Glu Thr Pro Ala Ser Ser
35 40

<210> 47
<211> 126
<212> DNA
<213> 人工序列

<220>
<223> XTEN AE42-4，DNA序列

<400> 47
ggcgcgccag gttctcctgc tggctccccc acctcaacag aagaggggac aagcgaaagc 60
gctacgcctg agagtggccc tggctctgag ccagccacct ccggctctga aacccctgcc 120
tcgagc 126

<210> 48
<211> 144
<212> PRT
<213> 人工序列

<220>
<223> XTEN AE144-2A，蛋白質序列

<400> 48
Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly
1 5 10 15

Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly

20 25 30

Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
35 40 45

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr Glu
50 55 60

Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
65 70 75 80

Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly
85 90 95

Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu
100 105 110

Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
115 120 125

Ser Gly Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly
130 135 140

<210> 49
<211> 450
<212> DNA
<213> 人工序列

<220>
<223> XTEN AE144-2A，DNA序列

<400> 49
ggcgcgccaa ccagtacgga gccgtccgag gggagcgcac caggaagccc ggctgggagc 60
ccgacttcta ccgaagaggg tacatctacc gaaccaagtg aaggttcagc accaggcacc 120
tcaacagaac cctctgaggg ctcggcgcct ggtacaagtg agtccgccac cccagaatcc 180

gggcctggga caagcacaga accttcggaa gggagtgccc ctggaacatc cgaatcggca 240

accccagaat cagggccagg atctgagccc gcgacttcgg gctccgagac gcctgggaca 300

tccaccgagc cctccgaagg atcagcccca ggcaccagca cggagccctc tgagggaagc 360

gcacctggta ccagcgaaag cgcaactccc gaatcaggtc ccggtacgag cgagtcggcg 420

accccggaga gcgggccagg tgcctcgagc 450

<210> 50

<211> 144

<212> PRT

<213> 人工序列

<220>

<223> XTEN AE144-3B，蛋白質序列

<400> 50

Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Glu Ser
1 5 10 15

Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser
20 25 30

Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly
35 40 45

Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu
50 55 60

Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly
65 70 75 80

Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
85 90 95

Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu
100 105 110

Pro Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser
115 120 125

Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
130 135 140

<210> 51
<211> 450
<212> DNA
<213> 人工序列

<220>
<223> XTEN AE144-3B，DNA序列

<400> 51
ggcgcgccaa gtcccgctgg aagcccaact agcaccgaag aggggacctc agagtccgcc 60
acccccgagt ccggccctgg ctctgagcct gccactagcg gctccgagac tcctggcaca 120
tccgaaagcg ctacacccga gagtggaccc ggcacctcta ccgagcccag tgagggctcc 180
gcccctggaa caagcaccga gcccagcgaa ggcagcgccc cagggacctc cacagagccc 240
agtgaaggca gtgctcctgg caccagcacc gaaccaagcg agggctctgc acccgggacc 300
tccaccgagc caagcgaagg ctctgccccc ggcacttcca ccgagcccag cgaaggcagc 360
gcccctggga gccccgctgg ctctcccacc agcactgagg agggcacatc taccgaacca 420
agtgaaggct ctgcaccagg tgcctcgagc 450

<210> 52
<211> 144
<212> PRT
<213> 人工序列

<220>
<223> XTEN AE144-4A，蛋白質序列

<400> 52

```
Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala
1               5                  10                  15

Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
            20                  25                  30

Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly
        35                  40                  45

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr Glu
    50                  55                  60

Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
65                  70                  75                  80

Ser Gly Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly
                85                  90                  95

Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Ser Pro Ala Gly
            100                 105                 110

Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Glu Ser Ala Thr Pro Glu
        115                 120                 125

Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
    130                 135                 140
```

<210> 53
<211> 450
<212> DNA
<213> 人工序列

<220>
<223> XTEN AE144-4A，DNA序列

<400> 53

ggcgcgccaa cgtccgaaag tgctacccct gagtcaggcc ctggtagtga gcctgccaca 60

agcggaagcg aaactccggg gacctcagag tctgccactc ccgaatcggg gccaggctct 120

gaaccggcca cttcagggag cgaaacacca ggaacatcgg agagcgctac cccggagagc 180

gggccaggaa ctagtactga gcctagcgag ggaagtgcac ctggtacaag cgagtccgcc 240

acacccgagt ctggccctgg ctctccagcg ggctcaccca cgagcactga agagggctct 300

cccgctggca gcccaacgtc gacagaagaa ggatcaccag caggctcccc cacatcaaca 360

gaggagggta catcagaatc tgctactccc gagagtggac ccggtacctc cactgagccc 420

agcgagggga gtgcaccagg tgcctcgagc 450

<210> 54

<211> 144

<212> PRT

<213> 人工序列

<220>

<223> XTEN AE144-5A，蛋白質序列

<400> 54

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala
1 5 10 15

Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
20 25 30

Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly
35 40 45

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr Glu
50 55 60

Pro Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser
65 70 75 80

Thr Glu Glu Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly

85 90 95

Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
100 105 110

Ala Thr Pro Glu Ser Gly Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser
115 120 125

Thr Glu Glu Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly
130 135 140

<210> 55
<211> 450
<212> DNA
<213> 人工序列

<220>
<223> XTEN AE144-5A，DNA序列

<400> 55
ggcgcgccaa catcagagag cgccaccccct gaaagtggtc ccgggagcga gccagccaca 60
tctgggtcgg aaacgccagg cacaagtgag tctgcaactc ccgagtccgg acctggctcc 120
gagcctgcca ctagcggctc cgagactccg ggaacttccg agagcgctac accagaaagc 180
ggacccggaa ccagtaccga acctagcgag ggctctgctc cgggcagccc agccggctct 240
cctacatcca cggaggaggg cacttccgaa tccgccaccc cggagtcagg gccaggatct 300
gaacccgcta cctcaggcag tgagacgcca ggaacgagcg agtccgctac accggagagt 360
gggccaggga gccctgctgg atctcctacg tccactgagg aagggtcacc agcgggctcg 420
cccaccagca ctgaagaagg tgcctcgagc 450

<210> 56
<211> 144
<212> PRT
<213> 人工序列

<220>

<223> XTEN AE144-6B，蛋白質序列

<400> 56

Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser
1 5 10 15

Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
20 25 30

Ser Gly Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly
35 40 45

Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Ser Glu Pro Ala
50 55 60

Thr Ser Gly Ser Glu Thr Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser
65 70 75 80

Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
85 90 95

Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Ser Glu Pro Ala
100 105 110

Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
115 120 125

Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
130 135 140

<210> 57

<211> 450

<212> DNA

<213> 人工序列

<220>

<223> XTEN AE144-6B，DNA序列

<400> 57

ggcgcgccaa catctaccga gccttccgaa ggctctgccc ctgggacctc agaatctgca 60

acccctgaaa gcggccctgg aacctccgaa agtgccactc ccgagagcgg cccagggaca 120

agcgagtcag caacccctga gtctggaccc ggcagcgagc ctgcaacctc tggctcagag 180

actcccggct cagaacccgc tacctcaggc tccgagacac ccggctctcc tgctgggagt 240

cccacttcca ccgaggaagg aacatccact gagcctagtg agggctctgc ccctggaacc 300

agcacagagc caagtgaggg cagtgcacca ggatccgagc cagcaaccag cgggtccgag 360

actcccggga cctctgagtc tgccacccca gagagcggac ccggcacttc aaccgagccc 420

tccgaaggat cagcaccagg tgcctcgagc 450

<210> 58

<211> 144

<212> PRT

<213> 人工序列

<220>

<223> XTEN AG144-1，蛋白質序列

<400> 58

Pro Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser
1 5 10 15

Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Thr Pro Gly Ser Gly Thr
20 25 30

Ala Ser Ser Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser
35 40 45

Pro Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Ala Ser
50 55 60

Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr
65 70 75 80

Ala Ser Ser Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser
85 90 95

Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ala Ser
100 105 110

Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser
115 120 125

Ser Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser
130 135 140

<210> 59
<211> 450
<212> DNA
<213> 人工序列

<220>
<223> XTEN AG144-1，DNA序列

<400> 59
ggcgcgccac ccgggtcgtc cccgtcggcg tccaccggaa cagggccagg gtcatccccg 60
tcagcgtcga ctgggacggg acccgggaca cccggttcgg ggactgcatc ctcctcgcct 120
ggttcgtcca ccccgtcagg agccacgggt tcgccgggaa gcagcccaag cgcatccact 180
ggtacagggc ctggggcttc accgggtact tcatccacgg ggtcaccggg aacgcccgga 240
tcggggacgg cttcctcatc accaggatcg tcaacaccct cgggcgcaac gggcagcccc 300
ggaacccctg gttcgggtac ggcgtcgtcg agccccggtg cgagcccggg aacaagctcg 360
acaggatcgc ctggggcgtc acccggcacg tcgagcacag gcagccccgg aaccccctgga 420
tcgggaaccg cgtcgtcaag cgcctcgagc 450

<210> 60

<211> 144

<212> PRT

<213> 人工序列

<220>

<223> XTEN AG144-A，蛋白質序列

<400> 60

```
Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ser Ser Pro
1               5                  10                  15

Ser Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser Pro Ser Ala Ser Thr
            20                  25                  30

Gly Thr Gly Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro
        35                  40                  45

Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Pro
    50                  55                  60

Ser Ala Ser Thr Gly Thr Gly Pro Gly Ala Ser Pro Gly Thr Ser Ser
65                  70                  75                  80

Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro
                85                  90                  95

Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Thr Pro Gly
            100                 105                 110

Ser Gly Thr Ala Ser Ser Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
        115                 120                 125

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
    130                 135                 140
```

<210> 61
<211> 450
<212> DNA
<213> 人工序列

<220>
<223> XTEN AG144-A，DNA序列

<400> 61
ggcgcgccag gtgcctcgcc gggaacatca tcaactggtt cacccgggtc atccccctcg 60
gcctcaaccg ggacgggtcc cggctcatcc cccagcgcca gcactggaac aggtcctggc 120
actcctggtt ccggtacggc atcgtcatcc ccgggaagct caacaccgtc cggagcgaca 180
ggatcacctg gctcgtcacc ttcggcgtca actggaacgg ggccaggggc ctcacccgga 240
acgtcctcga ctgggtcgcc tggtacgccg ggatcaggaa cggcctcatc ctcgcctggg 300
tcctcaacgc cctcgggtgc gactggttcg ccgggaactc ctggctcggg gacggcctcg 360
tcgtcgcctg gggcatcacc ggggacgagc tccacggggt cccctggagc gtcaccgggg 420
acctcctcga caggtagccc ggcctcgagc 450

<210> 62
<211> 144
<212> PRT
<213> 人工序列

<220>
<223> XTEN AG144-B，蛋白質序列

<400> 62
Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ser Ser Thr
1 5 10 15

Pro Ser Gly Ala Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
20 25 30

Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro
35 40 45

Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Pro
50 55 60

Ser Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser Pro Ser Ala Ser Thr
65 70 75 80

Gly Thr Gly Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro
85 90 95

Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ala Ser Pro
100 105 110

Gly Thr Ser Ser Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
115 120 125

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
130 135 140

<210> 63
<211> 450
<212> DNA
<213> 人工序列

<220>
<223> XTEN AG144-B，DNA序列

<400> 63
ggcgcgccag gtacaccggg cagcggcacg gcttcgtcgt cacccggctc gtccacaccg 60
tcgggagcta cgggaagccc aggagcgtca ccgggaacgt cgtcaacggg gtcaccgggt 120
acgccaggta gcggcacggc cagcagctcg ccaggttcat cgaccccgtc gggagcgact 180
gggtcgcccg gatcaagccc gtcagcttcc actggaacag gacccgggtc gtcgccgtca 240
gcctcaacgg ggacaggacc tggttcatcg acgccgtcag gggcgacagg ctcgcccgga 300
tcgtcaacac cctcgggggc aacggggagc cctggtgcgt cgcctggaac ctcatccacc 360
ggaagcccgg gggcctcgcc gggtacgagc tccacgggat cgcccggagc gtcccccgga 420

acttcaagca cagggagccc tgcctcgagc 450

<210> 64
<211> 144
<212> PRT
<213> 人工序列

<220>
<223> XTEN AG144-C，蛋白質序列

<400> 64

Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ala Ser Pro
1 5 10 15

Gly Thr Ser Ser Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
20 25 30

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
35 40 45

Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Thr Pro Gly
50 55 60

Ser Gly Thr Ala Ser Ser Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
65 70 75 80

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
85 90 95

Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ser Ser Thr
100 105 110

Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala
115 120 125

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
130 135 140

<210> 65
<211> 450
<212> DNA
<213> 人工序列

<220>
<223> XTEN AG144-C，DNA序列

<400> 65
ggcgcgccag gtacacccgg atcgggtaca gcgtcatcga gccccggtgc gtcacctggt 60
acgtcgagca cggggtcgcc aggggcgtcc cctgggacgt cctcaacagg ctcgcccggt 120
gcgtcacccg gcacgtcgtc cacgggttca cctggtagct ccccttccgc gtccactggc 180
accgggcctg gaactccggg gagcggcaca gcgagctcgt cgccgggagc atcgcctggg 240
acatcgagca ccgggtcgcc aggagcatcg cccggaacat ccagcacagg aagccccggc 300
gcgtcgcccg ggacatcaag cacaggttcc ccgggatcga gcacgccgtc cggagccact 360
ggatcaccag ggagctcgac accttccggc gcaacgggat cgcccggagc cagcccgggt 420
acgtcaagca ctggctcccc tgcctcgagc 450

<210> 66
<211> 144
<212> PRT
<213> 人工序列

<220>
<223> XTEN AG144-F，蛋白質序列

<400> 66
Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser Pro
1 5 10 15

Ser Ala Ser Thr Gly Thr Gly Pro Gly Ala Ser Pro Gly Thr Ser Ser

20 25 30

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
35 40 45

Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Pro
50 55 60

Ser Ala Ser Thr Gly Thr Gly Pro Gly Ala Ser Pro Gly Thr Ser Ser
65 70 75 80

Thr Gly Ser Pro Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro
85 90 95

Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ser Ser Thr
100 105 110

Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala
115 120 125

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
130 135 140

<210> 67
<211> 450
<212> DNA
<213> 人工序列

<220>
<223> XTEN AG144-F，DNA序列

<400> 67
ggcgcgccag gctccagccc ctccgcgagc acgggaaccg gaccaggttc gtcaccctca 60
gcatcaacgg ggacgggacc gggggcgtca ccaggaacgt cctccaccgg ctcgccgggt 120
gcatcacccg gaacgtcatc gaccggatcg ccagggagct cgacgccatc aggcgcaaca 180

ggatcacctg gctcaagccc tagcgcgtca accggcacgg gtccgggtgc ctcccctggc 240
acgtccagca ccggatcacc cggatcgagc ccatccgcct caaccggaac cggacccggt 300
acaccagggt cgggaacagc ctcctcgtca ccaggctcct caacccccctc gggagccacg 360
ggttcgcccg gttcgtcaac gccttccgga gcaactggta gccccggagc atcgccagga 420
acttcgagca cggggtcgcc cgcctcgagc 450

<210> 68
<211> 4374
<212> DNA
<213> 人工序列

<220>
<223> coFVIII-1 DNA序列

<400> 68
atgcagattg agctgtctac ttgcttttc ctgtgcctgc tgaggttttg cttttccgct 60
acacgaaggt attatctggg ggctgtggaa ctgtcttggg attacatgca gagtgacctg 120
ggagagctgc cagtggacgc aaggtttccc cctagagtcc ctaagtcatt ccccttcaac 180
actagcgtgg tctacaagaa aacactgttc gtggagttta ctgatcacct gttcaacatc 240
gcaaagccta ggccaccctg gatgggactg ctggggccaa caatccaggc cgaggtgtac 300
gacaccgtgg tcattacact taagaacatg gcctcacacc ccgtgagcct gcatgctgtg 360
ggcgtcagct actggaaggc ttccgaagga gcagagtatg acgatcagac ttcccagaga 420
gaaaaagagg acgataaggt gtttcctggc ggatctcata cctacgtgtg gcaggtcctg 480
aaagagaatg gccctatggc ctccgaccct ctgtgcctga cctactctta tctgagtcac 540
gtggacctgg tcaaggatct gaacagcggc ctgatcggag ccctgctggt gtgcagggaa 600
ggaagcctgg ctaaggagaa aacccagaca ctgcataagt tcattctgct gttcgccgtg 660
tttgacgaag ggaaatcatg gcacagcgag acaaagaata gtctgatgca ggacagggat 720
gccgcttcag ccagagcttg gcccaaaatg cacactgtga acggctacgt caatcgctca 780
ctgcctgggc tgatcggctg ccaccgaaag agcgtgtatt ggcatgtcat cgggatgggc 840
accacacctg aagtgcactc cattttcctg gagggacata cctttctggt ccgcaaccac 900
cgacaggctt ccctggagat ctctccaatt accttcctga cagcacagac tctgctgatg 960
gacctggggc agttcctgct gttttgccac atcagctccc accagcatga tggcatggag 1020
gcttacgtga aagtggactc ttgtcccgag gaacctcagc tgcggatgaa gaacaatgag 1080
gaagcagaag actatgacga tgacctgacc gactccgaga tggatgtggt ccgattcgat 1140
gacgataaca gcccctcctt tatccagatt agatctgtgg ccaagaaaca ccctaagaca 1200

tgggtccatt acatcgcagc cgaggaagag gactgggatt atgcaccact ggtgctggca 1260
ccagacgatc gctcctacaa atctcagtat ctgaacaatg ggccacagag gattggcaga 1320
aagtacaaga aagtgcggtt catggcatat accgatgaga ccttcaagac tcgcgaagcc 1380
atccagcacg agagcggcat cctgggacca ctgctgtacg gagaagtggg agacaccctg 1440
ctgatcattt tcaagaacca ggccagccgg ccttacaata tctatccaca tgggattaca 1500
gatgtgcgcc ctctgtacag caggagactg ccaaagggcg tcaaacacct gaaggacttc 1560
ccaatcctgc ccggagaaat cttcaagtac aagtggactg tcaccgtcga ggatggcccc 1620
actaagagcg accctcggtg cctgacccgc tactattcta gtttcgtgaa tatggaaaga 1680
gatctggcaa gcggactgat cggaccactg ctgatttgtt acaaagagag cgtggatcag 1740
agaggcaacc agatcatgtc cgacaagcgg aatgtgattc tgttcagtgt ctttgacgaa 1800
aacaggtcat ggtacctgac cgagaacatc cagagattcc tgcctaatcc agctggggtg 1860
cagctggaag atcctgagtt tcaggcatct aacatcatgc atagtattaa tggctacgtg 1920
ttcgacagtt tgcagctgag cgtgtgcctg cacgaggtcg cttactggta tatcctgagc 1980
attggggcac agacagattt cctgagcgtg ttcttttccg gctacacttt taagcataaa 2040
atggtctatg aggacacact gactctgttc cccttcagcg gcgaaaccgt gtttatgagc 2100
atggagaatc ccggactgtg gattctgggg tgccacaaca gcgatttcag aaatcgcgga 2160
atgactgccc tgctgaaagt gtcaagctgt gacaagaaca ccggggacta ctatgaagat 2220
tcatacgagg acatcagcgc atatctgctg tccaaaaaca atgccattga accccggtct 2280
tttagtcaga atcctccagt gctgaagagg caccagaggg agatcacccg cactaccctg 2340
cagagtgatc aggaagagat cgactacgac gatacaattt ctgtggaaat gaagaaagag 2400
gacttcgata tctatgacga agatgagaac cagagtcctc gatcattcca gaagaaaacc 2460
aggcattact ttattgccgc agtggagcgg ctgtgggatt atggcatgtc ctctagtcct 2520
cacgtgctgc gaaatagggc ccagtcagga agcgtcccac agttcaagaa agtggtcttc 2580
caggagttta cagacgggtc ctttactcag ccactgtaca ggggcgaact gaacgagcac 2640
ctgggactgc tggggcccta tatcagagca gaagtggagg ataacattat ggtcaccttc 2700
agaaatcagg cctctcggcc ttacagtttt tattcaagcc tgatctctta cgaagaggac 2760
cagcgacagg gagctgaacc acgaaaaaac ttcgtgaagc ctaatgagac caaaacatac 2820
ttttggaagg tgcagcacca tatggcccca acaaaagacg agttcgattg caaggcatgg 2880
gcctattttt ctgacgtgga tctggagaag gacgtgcaca gtggcctgat tggcccactg 2940
ctggtgtgcc atactaacac cctgaatcca gcccacggcc ggcaggtcac tgtccaggag 3000
ttcgctctgt tctttaccat ctttgatgag acaaagagct ggtacttcac cgaaaacatg 3060
gagcgaaatt gcagggctcc atgtaacatt cagatggaag accccacatt caaggagaac 3120
taccgctttc atgctatcaa tggatacatc atggatactc tgcccgggct ggtcatggca 3180
caggaccaga gaatccggtg gtatctgctg agcatgggca gcaacgagaa tatccactca 3240
attcatttca gcgggcacgt gtttactgtc aggaagaaag aagagtacaa gatggccctg 3300
tacaacctgt atcccggcgt gttcgaaacc gtcgagatgc tgcctagcaa ggccggaatc 3360

tggagagtgg aatgcctgat tggagagcac ctgcatgctg ggatgtctac cctgtttctg 3420
gtgtacagta ataagtgtca gacacccctg ggaatggcat ccgggcatat cagggatttc 3480
cagattaccg catctggaca gtacggacag tgggcaccta agctggctag actgcactat 3540
tccggatcta tcaacgcttg gtccacaaaa gagcctttct cttggattaa ggtggacctg 3600
ctggccccaa tgatcattca tggcatcaaa actcagggag ctcggcagaa gttctcctct 3660
ctgtacatct cacagtttat catcatgtac agcctggatg ggaagaaatg gcagacatac 3720
cgcggcaata gcacaggaac tctgatggtg ttctttggca acgtggacag cagcggaatc 3780
aagcacaaca ttttcaatcc ccctatcatt gctagataca tccggctgca cccaacccat 3840
tattctattc gaagtacact gaggatggaa ctgatgggat gcgatctgaa cagttgttca 3900
atgcccctgg ggatggagtc caaggcaatc tctgacgccc agattaccgc cagctcctac 3960
ttcactaata tgtttgctac ctggagccct tccaaagcaa gactgcacct gcaaggccgc 4020
agcaacgcat ggcgaccaca ggtgaacaat cccaaggagt ggttgcaggt cgattttcag 4080
aaaactatga aggtgaccgg ggtcacaact cagggcgtga aaagtctgct gacctcaatg 4140
tacgtcaagg agttcctgat ctctagttca caggacggac atcagtggac actgttcttt 4200
cagaacggga aggtgaaagt cttccagggc aatcaggatt cctttacacc tgtggtcaac 4260
agtctagacc ctccactgct gaccagatac ctgagaatcc accctcagtc ctgggtgcac 4320
cagattgccc tgagaatgga agtgctggga tgcgaggccc aggatctgta ctga 4374

<210> 69
<211> 577
<212> DNA
<213> 人工序列

<220>
<223> ET啟動子，DNA序列

<400> 69
ctcgaggtca attcacgcga gttaataatt accagcgcgg gccaaataaa taatccgcga 60
ggggcaggtg acgtttgccc agcgcgcgct ggtaattatt aacctcgcga atattgattc 120
gaggccgcga ttgccgcaat cgcgaggggc aggtgacctt tgcccagcgc gcgttcgccc 180
cgccccggac ggtatcgata agcttaggag cttgggctgc aggtcgaggg cactgggagg 240
atgttgagta agatggaaaa ctactgatga cccttgcaga gacagagtat taggacatgt 300
ttgaacaggg gccgggcgat cagcaggtag ctctagagga tccccgtctg tctgcacatt 360
tcgtagagcg agtgttccga tactctaatc tccctaggca aggttcatat ttgtgtaggt 420
tacttattct ccttttgttg actaagtcaa taatcagaat cagcaggttt ggagtcagct 480

tggcagggat cagcagcctg ggttggaagg aggggtata aaagcccctt caccaggaga 540

agccgtcaca cagatccaca agctcctgcc accatgg 577

<210> 70

<211> 4374

<212> DNA

<213> 人工序列

<220>

<223> coFVIII-5

<400> 70

atgcaaatcg aactgagcac ctgtttcttc ctctgcctgc tgagattctg tttctccgcg 60

acccgccgat actacctggg agcagtggag ctctcctggg attacatgca gagcgacctt 120

ggggagctgc ccgtggatgc caggttccct ccccgggtgc caaagtcgtt tccgttcaac 180

acctccgtgg tgtacaagaa aactctgttc gtggagttca ccgaccacct gttcaatatc 240

gccaagccca gacctccctg gatggggctg ttgggaccta ccatccaagc ggaggtgtac 300

gacactgtgg tcatcactct gaagaacatg gcctcgcatc ccgtgtccct gcacgccgtg 360

ggagtgtctt actggaaagc gtccgagggg gccgaatacg acgaccagac ctcgcagaga 420

gaaaaggaag atgacaaggt gttcccagga ggatcgcaca cctacgtgtg gcaagtgttg 480

aaggagaacg gcccaatggc ctccgacccg ctgtgcctga cctactcgta cctgtcccac 540

gtggacctcg tgaaggacct caactcggga ctgattggag ccctgctggt ctgcagggaa 600

ggctcactgg cgaaagaaaa gactcagacc ttgcacaagt tcattctgct gttcgctgtg 660

ttcgacgagg ggaagtcgtg gcacagcgag actaagaact ccctgatgca agatagagat 720

gccgcctccg cccgggcctg gcctaagatg cacaccgtga acggttacgt gaaccgctcc 780

ctccctggcc tgattggatg ccaccggaag tccgtgtact ggcacgtgat cgggatgggg 840

accacccccg aggtgcacag catcttcctg gaaggtcaca catttctcgt gcgcaaccac 900

cggcaggcct ccctggaaat cagccccatt accttcctca ctgcccagac tctgctgatg 960

gacctgggac agttcctgct gttctgccat atctcctccc accaacatga cggaatggag 1020

gcatacgtga aggtcgattc ctgccctgag gaaccccagc tccgcatgaa gaacaatgag 1080

gaagccgagg actacgacga cgacctgacg gatagcgaga tggatgtggt ccggttcgat 1140

gacgataaca gcccttcctt catccaaatt cgctcggtgg caaagaagca ccccaagacc 1200

tgggtgcatt acattgcggc ggaagaagag gactgggatt atgccccgct tgtcctcgct 1260

cctgacgacc ggagctacaa gagccagtac ctgaacaacg gtccacagag gatcggtaga 1320

aagtacaaga aggtccgctt catggcctat accgacgaaa ccttcaaaac tagagaggcc 1380

atccaacacg aatccggcat cctgggcccg ctcttgtacg gagaagtcgg cgacaccctt 1440

ctcattatct tcaagaacca ggcttcccgg ccgtacaaca tctatccgca tgggatcact 1500

gacgtgcgcc cactgtactc gcggcgcctg cccaagggtg tcaaacacct gaaggatttt 1560

ccgatccttc cgggagaaat cttcaagtac aagtggaccg tgaccgtgga agatggccca 1620

actaagtctg accctagatg cctcacccgc tactactcat ccttcgtcaa catggagcgc 1680

gacctggcca gcggactgat cggcccgctg ctgatttgct acaaggaatc agtggaccaa 1740

cggggaaacc agatcatgtc ggataagagg aacgtcatcc tcttctccgt gtttgacgaa 1800

aaccggtcgt ggtacctgac tgaaaacatc cagcggttcc tccccaaccc cgcgggcgtg 1860

cagctggaag atcctgagtt tcaggcatca aacatcatgc actccattaa cggctacgtg 1920

ttcgattcgc tgcagctgag cgtgtgtctg cacgaagtgg cctactggta catcctgtcc 1980

attggtgccc agactgactt cctgtccgtg tttttctccg gctacacgtt caagcacaag 2040

atggtgtacg aggacaccct gaccctcttc ccttttttccg gcgaaactgt gtttatgagc 2100

atggagaatc ccggcctgtg gatcttgggc tgccacaaca gcgacttccg taacagagga 2160

atgactgcgc tgctcaaggt gtccagctgc gacaagaaca ccggagacta ttatgaggac 2220

tcatacgagg acatctccgc ctacctcctg tccaagaata acgccattga acctcggagc 2280

ttcagccaga acccacccgt gcttaagaga catcaacggg agatcactag gaccaccctg 2340

cagtcagacc aggaggaaat cgactacgat gacaccatct cggtcgagat gaagaaggag 2400

gactttgaca tctacgacga agatgaaaac cagagcccga ggtcgttcca aaagaaaacc 2460

cgccactact ttattgctgc tgtcgagcgg ctgtgggact acggaatgtc gtcctcgccg 2520

cacgtgctcc gcaaccgagc ccagagcggc tcggtgccgc aattcaagaa ggtcgtgttc 2580

caggagttca ctgacgggag cttcactcag cctttgtacc ggggagaact caatgaacat 2640

ctcggcctcc tcggacctta catcagagca gaagtggaag ataacatcat ggtcactttc 2700

cgtaaccaag ccagccgccc gtactcgttc tactcctccc tcatttctta cgaagaggac 2760

cagcggcagg gcgcagaacc gcgcaagaac ttcgtgaagc ccaacgaaac caagacctac 2820

ttctggaaag tgcagcatca tatggccccg actaaggacg agtttgactg caaagcctgg 2880

gcctacttct ccgatgtgga cttggagaag gacgtccact ccggcctcat cggtcccctg 2940

ctcgtgtgcc ataccaatac cctgaacccc gcacacggtc gccaggtcac cgtgcaggag 3000

ttcgctctgt tcttcactat cttcgacgaa actaagtcct ggtacttcac cgagaacatg 3060

gagaggaact gcagagcccc ctgtaacatc cagatggagg acccgacgtt caaggaaaac 3120

taccggttcc acgccattaa cggatacatc atggatacgc tgccgggtct tgtgatggcc 3180

caggatcaac ggatcagatg gtacttattg tcgatgggca gcaacgagaa catccactct 3240

attcacttct ccggtcatgt gttcactgtg cggaagaagg aagagtacaa gatggccctg 3300

tacaaccttt atcccggagt gttcgaaact gtggaaatgc tgccgtcgaa ggccggcatt 3360

tggcgcgtgg agtgtttgat tggagaacat ctccatgcgg ggatgtcaac cctgttcctg 3420

gtgtatagca acaagtgcca gactccgctt gggatggcgt caggacacat tagggatttc 3480

cagatcactg cgtccggcca gtacggccaa tgggccccta agctggcccg cctgcattac 3540

tccggatcca ttaacgcctg gtcaaccaag gagccattct cctggatcaa ggtggacctt 3600
ctggccccca tgattatcca cggaattaag acccaggggg cccggcagaa gttctcctca 3660
ctgtacatca gccagttcat aatcatgtac tccctggacg gaaagaagtg gcaaacctac 3720
aggggggaaca gcaccggcac actgatggtc tttttcggaa atgtggactc ctccgggatt 3780
aagcataaca tcttcaaccc tccgattatc gctcggtaca ttagacttca ccctacccac 3840
tacagcattc gctccaccct gcggatggaa ctgatgggct gcgatctgaa ctcgtgcagc 3900
atgccgttgg gaatggagtc caaagcaatt tccgacgcgc agatcaccgc ctcgtcctac 3960
tttaccaaca tgttcgccac gtggtcaccg tccaaggccc ggctgcacct ccagggaaga 4020
tccaacgcat ggcggccaca ggtcaacaac cctaaggagt ggctccaggt ggacttccag 4080
aaaaccatga aggtcaccgg agtcacaacc cagggagtga agtcgctgct gacttctatg 4140
tacgtcaagg agttcctgat ctccagcagc caggacgggc accagtggac cctgttcttc 4200
caaaatggaa aggtcaaggt gtttcagggc aatcaggatt cattcacccc ggtggtgaac 4260
tcccttgatc cacccctcct gacccgctac cttcgcatcc acccacagtc ctgggtgcac 4320
cagatcgcgc tgaggatgga ggtcctggga tgcgaagccc aggacctgta ctga 4374

<210> 71
<211> 4374
<212> DNA
<213> 人工序列

<220>
<223> coFVIII-6

<400> 71
atgcagattg agctgtccac ttgtttcttc ctgtgcctcc tgcgcttctg tttctccgcc 60
actcgccggt actaccttgg agccgtggag ctttcatggg actacatgca gagcgacctg 120
ggcgaactcc ccgtggatgc cagattcccc ccccgcgtgc caaagtcctt ccccttaac 180
acctccgtgg tgtacaagaa aaccctcttt gtcgagttca ctgaccacct gttcaacatc 240
gccaagccgc gcccaccttg gatgggcctc ctgggaccga ccattcaagc tgaagtgtac 300
gacaccgtgg tgatcaccct gaagaacatg gcgtcccacc ccgtgtccct gcatgcggtc 360
ggagtgtcct actggaaggc ctccgaagga gctgagtacg acgaccagac tagccagcgg 420
gaaaaggagg acgataaagt gttcccgggc ggctcgcata cttacgtgtg gcaagtcctg 480
aaggaaaacg gacctatggc atccgatcct ctgtgcctga cttactccta cctttcccat 540
gtggacctcg tgaaggacct gaacagcggg ctgattggtg cacttctcgt gtgccgcgaa 600
ggttcgctcg ctaaggaaaa gacccagacc ctccataagt tcatcctttt gttcgctgtg 660

ttcgatgaag gaaagtcatg gcattccgaa actaagaact cgctgatgca ggaccgggat 720
gccgcctcag cccgcgcctg gcctaaaatg catacagtca acggatacgt gaatcggtca 780
ctgcccgggc tcatcggttg tcacagaaag tccgtgtact ggcacgtcat cggcatgggc 840
actacgcctg aagtgcactc catcttcctg gaagggcaca ccttcctcgt gcgcaaccac 900
cgccaggcct ctctggaaat ctccccgatt acctttctga ccgcccagac tctgctcatg 960
gacctggggc agttccttct cttctgccac atctccagcc atcagcacga cggaatggag 1020
gcctacgtga aggtggactc atgcccggaa gaacctcagt tgcggatgaa gaacaacgag 1080
gaggccgagg actatgacga cgatttgact gactccgaga tggacgtcgt gcggttcgat 1140
gacgacaaca gccccagctt catccagatt cgcagcgtgg ccaagaagca ccccaaaacc 1200
tgggtgcact acatcgcggc cgaggaagaa gattgggact acgccccgtt ggtgctggca 1260
cccgatgacc ggtcgtacaa gtcccagtat ctgaacaatg gtccgcagcg gattggcaga 1320
aagtacaaga aagtgcggtt catggcgtac actgacgaaa cgtttaagac ccgggaggcc 1380
attcaacatg agagcggcat tctgggacca ctgctgtacg gagaggtcgg cgataccctg 1440
ctcatcatct tcaaaaacca ggcctcccgg ccttacaaca tctaccctca cggaatcacc 1500
gacgtgcggc cactctactc gcggcgcctg ccgaagggcg tcaagcacct gaaagacttc 1560
cctatcctgc cgggcgaaat cttcaagtat aagtggaccg tcaccgtgga ggacgggccc 1620
accaagagcg atcctaggtg tctgactcgg tactactcca gcttcgtgaa catggaacgg 1680
gacctggcat cgggactcat tggaccgctg ctgatctgct acaaagagtc ggtggatcaa 1740
cgcggcaacc agatcatgtc cgacaagcgc aacgtgatcc tgttctccgt gtttgatgaa 1800
aacagatcct ggtacctcac tgaaaacatc cagaggttcc tcccaaaccc cgcaggagtg 1860
caactggagg accctgagtt tcaggcctcg aatatcatgc actcgattaa cggttacgtg 1920
ttcgactcgc tgcagctgag cgtgtgcctc catgaagtcg cttactggta cattctgtcc 1980
atcggcgccc agactgactt cctgagcgtg ttcttttccg gttacacctt taagcacaag 2040
atggtgtacg aagataccct gaccctgttc cctttctccg gcgaaacggt gttcatgtcg 2100
atggagaacc cgggtctgtg gattctggga tgccacaaca gcgactttcg gaaccgcgga 2160
atgactgccc tgctgaaggt gtcctcatgc gacaagaaca ccggagacta ctacgaggac 2220
tcctacgagg atatctcagc ctacctcctg tccaagaaca acgcgatcga gccgcgcagc 2280
ttcagccaga acccgcctgt gctgaagagg caccagcgag aaattacccg gaccaccctc 2340
caatcggatc aggaggaaat cgactacgac gacaccatct cggtggaaat gaagaaggaa 2400
gatttcgata tctacgacga ggacgaaaat cagtcccctc gctcattcca aaagaaaact 2460
agacactact ttatcgccgc ggtggaaaga ctgtgggact atggaatgtc atccagccct 2520
cacgtccttc ggaaccgggc ccagagcgga tcggtgcctc agttcaagaa agtggtgttc 2580
caggagttca ccgacggcag cttcacccag ccgctgtacc ggggagaact gaacgaacac 2640
ctgggcctgc tcggtcccta catccgcgcg gaagtggagg ataacatcat ggtgaccttc 2700
cgtaaccaag catccagacc ttactccttc tattcctccc tgatctcata cgaggaggac 2760
cagcgccaag gcgccgagcc ccgcaagaac ttcgtcaagc ccaacgagac taagacctac 2820

ttctggaagg tccaacacca tatggccccg accaaggatg agtttgactg caaggcctgg 2880
gcctacttct ccgacgtgga ccttgagaag gatgtccatt ccggcctgat cgggccgctg 2940
ctcgtgtgtc acaccaacac cctgaaccca gcgcatggac gccaggtcac cgtccaggag 3000
tttgctctgt tcttcaccat ttttgacgaa actaagtcct ggtacttcac cgagaatatg 3060
gagcgaaact gtagagcgcc ctgcaatatc cagatggaag atccgacttt caaggagaac 3120
tatagattcc acgccatcaa cgggtacatc atggatactc tgccggggct ggtcatggcc 3180
caggatcaga ggattcggtg gtacttgctg tcaatgggat cgaacgaaaa cattcactcc 3240
attcacttct ccggtcacgt gttcactgtg cgcaagaagg aggagtacaa gatggcgctg 3300
tacaatctgt accccggggt gttcgaaact gtggagatgc tgccgtccaa ggccggcatc 3360
tggagagtgg agtgcctgat cggagagcac ctccacgcgg ggatgtccac cctcttcctg 3420
gtgtactcga ataagtgcca gaccccgctg ggcatggcct cgggccacat cagagacttc 3480
cagatcacag caagcggaca atacggccaa tgggcgccga agctggcccg cttgcactac 3540
tccggatcga tcaacgcatg gtccaccaag gaaccgttct cgtggattaa ggtggacctc 3600
ctggccccta tgattatcca cggaattaag acccagggcg ccaggcagaa gttctcctcc 3660
ctgtacatct cgcaattcat catcatgtac agcctggacg ggaagaagtg gcagacttac 3720
aggggaaact ccaccggcac cctgatggtc tttttcggca acgtggattc ctccggcatt 3780
aagcacaaca tcttcaaccc accgatcata gccagatata ttaggctcca ccccactcac 3840
tactcaatcc gctcaactct tcggatggaa ctcatggggt gcgacctgaa ctcctgctcc 3900
atgccgttgg ggatggaatc aaaggctatt agcgacgccc agatcaccgc gagctcctac 3960
ttcactaaca tgttcgccac ctggagcccc tccaaggcca ggctgcactt gcagggacgg 4020
tcaaatgcct ggcggccgca agtgaacaat ccgaaggaat ggcttcaagt ggatttccaa 4080
aagaccatga aagtgaccgg agtcaccacc cagggagtga agtcccttct gacctcgatg 4140
tatgtgaagg agttcctgat tagcagcagc caggacgggc accagtggac cctgttcttc 4200
caaaacggaa aggtcaaggt gttccagggg aaccaggact cgttcacacc cgtggtgaac 4260
tccctggacc ccccactgct gacgcggtac ttgaggattc atcctcagtc ctgggtccat 4320
cagattgcat tgcgaatgga agtcctgggc tgcgaggccc aggacctgta ctga 4374

<210> 72
<211> 4824
<212> DNA
<213> 人工序列

<220>
<223> coFVIII-6-XTEN

<400> 72

atgcagattg agctgtccac ttgtttcttc ctgtgcctcc tgcgcttctg tttctccgcc 60

actcgccggt actaccttgg agccgtggag ctttcatggg actacatgca gagcgacctg 120

ggcgaactcc ccgtggatgc cagattcccc ccccgcgtgc caaagtcctt ccccttttaac 180

acctccgtgg tgtacaagaa aaccctcttt gtcgagttca ctgaccacct gttcaacatc 240

gccaagccgc gcccaccttg gatgggcctc ctgggaccga ccattcaagc tgaagtgtac 300

gacaccgtgg tgatcaccct gaagaacatg gcgtcccacc ccgtgtccct gcatgcggtc 360

ggagtgtcct actggaaggc ctccgaagga gctgagtacg acgaccagac tagccagcgg 420

gaaaaggagg acgataaagt gttcccgggc ggctcgcata cttacgtgtg gcaagtcctg 480

aaggaaaacg gacctatggc atccgatcct ctgtgcctga cttactccta cctttcccat 540

gtggacctcg tgaaggacct gaacagcggg ctgattggtg cacttctcgt gtgccgcgaa 600

ggttcgctcg ctaaggaaaa gacccagacc ctccataagt tcatcctttt gttcgctgtg 660

ttcgatgaag gaaagtcatg gcattccgaa actaagaact cgctgatgca ggaccgggat 720

gccgcctcag cccgcgcctg gcctaaaatg catacagtca acggatacgt gaatcggtca 780

ctgcccgggc tcatcggttg tcacagaaag tccgtgtact ggcacgtcat cggcatgggc 840

actacgcctg aagtgcactc catcttcctg gaagggcaca ccttcctcgt gcgcaaccac 900

cgccaggcct ctctggaaat ctccccgatt acctttctga ccgcccagac tctgctcatg 960

gacctggggc agttccttct cttctgccac atctccagcc atcagcacga cggaatggag 1020

gcctacgtga aggtggactc atgcccggaa gaacctcagt tgcggatgaa gaacaacgag 1080

gaggccgagg actatgacga cgatttgact gactccgaga tggacgtcgt gcggttcgat 1140

gacgacaaca gccccagctt catccagatt cgcagcgtgg ccaagaagca ccccaaaacc 1200

tgggtgcact acatcgcggc cgaggaagaa gattgggact acgccccgtt ggtgctggca 1260

cccgatgacc ggtcgtacaa gtcccagtat ctgaacaatg gtccgcagcg gattggcaga 1320

aagtacaaga aagtgcggtt catggcgtac actgacgaaa cgtttaagac ccgggaggcc 1380

attcaacatg agagcggcat tctgggacca ctgctgtacg gagaggtcgg cgataccctg 1440

ctcatcatct tcaaaaacca ggcctcccgg ccttacaaca tctaccctca cggaatcacc 1500

gacgtgcggc cactctactc gcggcgcctg ccgaagggcg tcaagcacct gaaagacttc 1560

cctatcctgc cgggcgaaat cttcaagtat aagtggaccg tcaccgtgga ggacgggccc 1620

accaagagcg atcctaggtg tctgactcgg tactactcca gcttcgtgaa catggaacgg 1680

gacctggcat cgggactcat tggaccgctg ctgatctgct acaaagagtc ggtggatcaa 1740

cgcggcaacc agatcatgtc cgacaagcgc aacgtgatcc tgttctccgt gtttgatgaa 1800

aacagatcct ggtacctcac tgaaaacatc cagaggttcc tcccaaaccc cgcaggagtg 1860

caactggagg accctgagtt tcaggcctcg aatatcatgc actcgattaa cggttacgtg 1920

ttcgactcgc tgcagctgag cgtgtgcctc catgaagtcg cttactggta cattctgtcc 1980

atcggcgccc agactgactt cctgagcgtg ttcttttccg gttacacctt taagcacaag 2040

atggtgtacg aagataccct gaccctgttc cctttctccg gcgaaacggt gttcatgtcg 2100

atggagaacc cgggtctgtg gattctggga tgccacaaca gcgactttcg gaaccgcgga 2160
atgactgccc tgctgaaggt gtcctcatgc gacaagaaca ccggagacta ctacgaggac 2220
tcctacgagg atatctcagc ctacctcctg tccaagaaca acgcgatcga gccgcgcagc 2280
ttcagccaga acggcgcgcc aacatcagag agcgccaccc ctgaaagtgg tcccgggagc 2340
gagccagcca catctgggtc ggaaacgcca ggcacaagtg agtctgcaac tcccgagtcc 2400
ggacctggct ccgagcctgc cactagcggc tccgagactc cgggaacttc cgagagcgct 2460
acaccagaaa gcggacccgg aaccagtacc gaacctagcg agggctctgc tccgggcagc 2520
ccagccggct ctcctacatc cacggaggag ggcacttccg aatccgccac cccggagtca 2580
gggccaggat ctgaacccgc tacctcaggc agtgagacgc caggaacgag cgagtccgct 2640
acaccggaga gtgggccagg gagccctgct ggatctccta cgtccactga ggaagggtca 2700
ccagcgggct cgcccaccag cactgaagaa ggtgcctcga gcccgcctgt gctgaagagg 2760
caccagcgag aaattacccg gaccaccctc caatcggatc aggaggaaat cgactacgac 2820
gacaccatct cggtggaaat gaagaaggaa gatttcgata tctacgacga ggacgaaaat 2880
cagtcccctc gctcattcca aaagaaaact agacactact ttatcgccgc ggtggaaaga 2940
ctgtgggact atggaatgtc atccagccct cacgtccttc ggaaccgggc ccagagcgga 3000
tcggtgcctc agttcaagaa agtggtgttc caggagttca ccgacggcag cttcacccag 3060
ccgctgtacc ggggagaact gaacgaacac ctgggcctgc tcggtcccta catccgcgcg 3120
gaagtggagg ataacatcat ggtgaccttc cgtaaccaag catccagacc ttactccttc 3180
tattcctccc tgatctcata cgaggaggac cagcgccaag gcgccgagcc ccgcaagaac 3240
ttcgtcaagc ccaacgagac taagacctac ttctggaagg tccaacacca tatggccccg 3300
accaaggatg agtttgactg caaggcctgg gcctacttct ccgacgtgga ccttgagaag 3360
gatgtccatt ccggcctgat cgggccgctg ctcgtgtgtc acaccaacac cctgaaccca 3420
gcgcatggac gccaggtcac cgtccaggag tttgctctgt tcttcaccat ttttgacgaa 3480
actaagtcct ggtacttcac cgagaatatg gagcgaaact gtagagcgcc ctgcaatatc 3540
cagatggaag atccgacttt caaggagaac tatagattcc acgccatcaa cgggtacatc 3600
atggatactc tgccggggct ggtcatggcc caggatcaga ggattcggtg gtacttgctg 3660
tcaatgggat cgaacgaaaa cattcactcc attcacttct ccggtcacgt gttcactgtg 3720
cgcaagaagg aggagtacaa gatggcgctg tacaatctgt accccggggt gttcgaaact 3780
gtggagatgc tgccgtccaa ggccggcatc tggagagtgg agtgcctgat cggagagcac 3840
ctccacgcgg ggatgtccac cctcttcctg gtgtactcga ataagtgcca gaccccgctg 3900
ggcatggcct cgggccacat cagagacttc cagatcacag caagcggaca atacggccaa 3960
tgggcgccga agctggcccg cttgcactac tccggatcga tcaacgcatg gtccaccaag 4020
gaaccgttct cgtggattaa ggtggacctc ctggccccta tgattatcca cggaattaag 4080
acccagggcg ccaggcagaa gttctcctcc ctgtacatct cgcaattcat catcatgtac 4140
agcctggacg ggaagaagtg gcagacttac aggggaaact ccaccggcac cctgatggtc 4200
tttttcggca acgtggattc ctccggcatt aagcacaaca tcttcaaccc accgatcata 4260

gccagatata ttaggctcca ccccactcac tactcaatcc gctcaactct tcggatggaa 4320

ctcatggggt gcgacctgaa ctcctgctcc atgccgttgg ggatggaatc aaaggctatt 4380

agcgacgccc agatcaccgc gagctcctac ttcactaaca tgttcgccac ctggagcccc 4440

tccaaggcca ggctgcactt gcagggacgg tcaaatgcct ggcggccgca agtgaacaat 4500

ccgaaggaat ggcttcaagt ggatttccaa aagaccatga aagtgaccgg agtcaccacc 4560

cagggagtga agtcccttct gacctcgatg tatgtgaagg agttcctgat tagcagcagc 4620

caggacgggc accagtggac cctgttcttc caaaacggaa aggtcaaggt gttccagggg 4680

aaccaggact cgttcacacc cgtggtgaac tccctggacc ccccactgct gacgcggtac 4740

ttgaggattc atcctcagtc ctgggtccat cagattgcat tgcgaatgga agtcctgggc 4800

tgcgaggccc aggacctgta ctga 4824

<210> 73

<211> 12

<212> PRT

<213> 人工序列

<220>

<223> AD基序

<400> 73

Gly Glu Ser Pro Gly Gly Ser Ser Gly Ser Glu Ser

1 5 10

<210> 74

<211> 12

<212> PRT

<213> 人工序列

<220>

<223> AD基序

<400> 74

Gly Ser Glu Gly Ser Ser Gly Pro Gly Glu Ser Ser

1 5 10

<210> 75
<211> 12
<212> PRT
<213> 人工序列

<220>
<223> AD基序

<400> 75
Gly Ser Ser Glu Ser Gly Ser Ser Glu Gly Gly Pro
1 5 10

<210> 76
<211> 12
<212> PRT
<213> 人工序列

<220>
<223> AD基序

<400> 76
Gly Ser Gly Gly Glu Pro Ser Glu Ser Gly Ser Ser
1 5 10

<210> 77
<211> 12
<212> PRT
<213> 人工序列

<220>
<223> AE、AM基序

<400> 77

Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu

1 5 10

<210> 78

<211> 12

<212> PRT

<213> 人工序列

<220>

<223> AE、AM、AQ基序

<400> 78

Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro

1 5 10

<210> 79

<211> 12

<212> PRT

<213> 人工序列

<220>

<223> AE、AM、AQ基序

<400> 79

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro

1 5 10

<210> 80

<211> 12

<212> PRT

<213> 人工序列

<220>

<223> AE、AM、AQ基序

<400> 80

Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro

1 5 10

<210> 81

<211> 12

<212> PRT

<213> 人工序列

<220>

<223> AF、AM基序

<400> 81

Gly Ser Thr Ser Glu Ser Pro Ser Gly Thr Ala Pro

1 5 10

<210> 82

<211> 12

<212> PRT

<213> 人工序列

<220>

<223> AF、AM基序

<400> 82

Gly Thr Ser Thr Pro Glu Ser Gly Ser Ala Ser Pro

1 5 10

<210> 83

<211> 12

<212> PRT

<213> 人工序列

<220>

<223> AF、AM基序

<400> 83

Gly Thr Ser Pro Ser Gly Glu Ser Ser Thr Ala Pro

1 5 10

<210> 84

<211> 12

<212> PRT

<213> 人工序列

<220>

<223> AF、AM基序

<400> 84

Gly Ser Thr Ser Ser Thr Ala Glu Ser Pro Gly Pro

1 5 10

<210> 85

<211> 12

<212> PRT

<213> 人工序列

<220>

<223> AG、AM基序

<400> 85

Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro

1 5 10

<210> 86
<211> 12
<212> PRT
<213> 人工序列

<220>
<223> AG、AM基序

<400> 86
Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro
1 5 10

<210> 87
<211> 12
<212> PRT
<213> 人工序列

<220>
<223> AG、AM基序

<400> 87
Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro
1 5 10

<210> 88
<211> 12
<212> PRT
<213> 人工序列

<220>
<223> AG、AM基序

<400> 88

Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
1 5 10

<210> 89
<211> 12
<212> PRT
<213> 人工序列

<220>
<223> AQ基序

<400> 89
Gly Glu Pro Ala Gly Ser Pro Thr Ser Thr Ser Glu
1 5 10

<210> 90
<211> 12
<212> PRT
<213> 人工序列

<220>
<223> AQ基序

<400> 90
Gly Thr Gly Glu Pro Ser Ser Thr Pro Ala Ser Glu
1 5 10

<210> 91
<211> 12
<212> PRT
<213> 人工序列

<220>

<223> AQ基序

<400> 91

Gly Ser Gly Pro Ser Thr Glu Ser Ala Pro Thr Glu

1 5 10

<210> 92

<211> 12

<212> PRT

<213> 人工序列

<220>

<223> AQ基序

<400> 92

Gly Ser Glu Thr Pro Ser Gly Pro Ser Glu Thr Ala

1 5 10

<210> 93

<211> 12

<212> PRT

<213> 人工序列

<220>

<223> AQ基序

<400> 93

Gly Pro Ser Glu Thr Ser Thr Ser Glu Pro Gly Ala

1 5 10

<210> 94

<211> 12

<212> PRT

<213> 人工序列

<220>

<223> AQ基序

<400> 94

Gly Ser Pro Ser Glu Pro Thr Glu Gly Thr Ser Ala

1 5 10

<210> 95

<211> 12

<212> PRT

<213> 人工序列

<220>

<223> BC基序

<400> 95

Gly Ser Gly Ala Ser Glu Pro Thr Ser Thr Glu Pro

1 5 10

<210> 96

<211> 12

<212> PRT

<213> 人工序列

<220>

<223> BC基序

<400> 96

Gly Ser Glu Pro Ala Thr Ser Gly Thr Glu Pro Ser

1 5 10

<210> 97

<211> 12

<212> PRT

<213> 人工序列

<220>
<223> BC基序

<400> 97
Gly Thr Ser Glu Pro Ser Thr Ser Glu Pro Gly Ala
1 5 10

<210> 98
<211> 12
<212> PRT
<213> 人工序列

<220>
<223> BC基序

<400> 98
Gly Thr Ser Thr Glu Pro Ser Glu Pro Gly Ser Ala
1 5 10

<210> 99
<211> 12
<212> PRT
<213> 人工序列

<220>
<223> BD基序

<400> 99
Gly Ser Thr Ala Gly Ser Glu Thr Ser Thr Glu Ala
1 5 10

<210> 100
<211> 12
<212> PRT
<213> 人工序列

<220>
<223> BD基序

<400> 100
Gly Ser Glu Thr Ala Thr Ser Gly Ser Glu Thr Ala
1 5 10

<210> 101
<211> 12
<212> PRT
<213> 人工序列

<220>
<223> BD基序

<400> 101
Gly Thr Ser Glu Ser Ala Thr Ser Glu Ser Gly Ala
1 5 10

<210> 102
<211> 12
<212> PRT
<213> 人工序列

<220>
<223> BD基序

<400> 102
Gly Thr Ser Thr Glu Ala Ser Glu Gly Ser Ala Ser

1 5 10

<210> 103
<211> 1607
<212> PRT
<213> 人工序列

<220>
<223> coFVIII-6-XTEN蛋白質序列

<400> 103

Met Gln Ile Glu Leu Ser Thr Cys Phe Phe Leu Cys Leu Leu Arg Phe
1 5 10 15

Cys Phe Ser Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser
20 25 30

Trp Asp Tyr Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg
35 40 45

Phe Pro Pro Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val
50 55 60

Tyr Lys Lys Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile
65 70 75 80

Ala Lys Pro Arg Pro Pro Trp Met Gly Leu Leu Gly Pro Thr Ile Gln
85 90 95

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

His Pro Val Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ala Ser
115 120 125

```
Glu Gly Ala Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp
    130                 135                 140

Asp Lys Val Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu
145                 150                 155                 160

Lys Glu Asn Gly Pro Met Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser
                165                 170                 175

Tyr Leu Ser His Val Asp Leu Val Lys Asp Leu Asn Ser Gly Leu Ile
            180                 185                 190

Gly Ala Leu Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr
        195                 200                 205

Gln Thr Leu His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly
    210                 215                 220

Lys Ser Trp His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp
225                 230                 235                 240

Ala Ala Ser Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr
                245                 250                 255

Val Asn Arg Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val
            260                 265                 270

Tyr Trp His Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile
        275                 280                 285

Phe Leu Glu Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser
    290                 295                 300

Leu Glu Ile Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met
305                 310                 315                 320
```

```
Asp Leu Gly Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His
                325                 330                 335

Asp Gly Met Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro
            340                 345                 350

Gln Leu Arg Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp
        355                 360                 365

Leu Thr Asp Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser
    370                 375                 380

Pro Ser Phe Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr
385                 390                 395                 400

Trp Val His Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro
                405                 410                 415

Leu Val Leu Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn
            420                 425                 430

Asn Gly Pro Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met
        435                 440                 445

Ala Tyr Thr Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu
    450                 455                 460

Ser Gly Ile Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu
465                 470                 475                 480

Leu Ile Ile Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro
                485                 490                 495

His Gly Ile Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys
            500                 505                 510
```

Gly Val Lys His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe
515 520 525

Lys Tyr Lys Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp
530 535 540

Pro Arg Cys Leu Thr Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg
545 550 555 560

Asp Leu Ala Ser Gly Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu
565 570 575

Ser Val Asp Gln Arg Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val
580 585 590

Ile Leu Phe Ser Val Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu
595 600 605

Asn Ile Gln Arg Phe Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp
610 615 620

Pro Glu Phe Gln Ala Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val
625 630 635 640

Phe Asp Ser Leu Gln Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp
645 650 655

Tyr Ile Leu Ser Ile Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe
660 665 670

Ser Gly Tyr Thr Phe Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr
675 680 685

Leu Phe Pro Phe Ser Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro
690 695 700

Gly Leu Trp Ile Leu Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly
705 710 715 720

Met Thr Ala Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp
725 730 735

Tyr Tyr Glu Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys
740 745 750

Asn Asn Ala Ile Glu Pro Arg Ser Phe Ser Gln Asn Gly Ala Pro Thr
755 760 765

Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr
770 775 780

Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser
785 790 795 800

Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Thr
805 810 815

Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr Glu Pro
820 825 830

Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr
835 840 845

Glu Glu Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser
850 855 860

Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala
865 870 875 880

Thr Pro Glu Ser Gly Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr
885 890 895

```
Glu Glu Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Ala
            900                 905                 910

Ser Ser Pro Pro Val Leu Lys Arg His Gln Arg Glu Ile Thr Arg Thr
        915                 920                 925

Thr Leu Gln Ser Asp Gln Glu Glu Ile Asp Tyr Asp Asp Thr Ile Ser
    930                 935                 940

Val Glu Met Lys Lys Glu Asp Phe Asp Ile Tyr Asp Glu Asp Glu Asn
945                 950                 955                 960

Gln Ser Pro Arg Ser Phe Gln Lys Lys Thr Arg His Tyr Phe Ile Ala
                965                 970                 975

Ala Val Glu Arg Leu Trp Asp Tyr Gly Met Ser Ser Ser Pro His Val
            980                 985                 990

Leu Arg Asn Arg Ala Gln Ser Gly  Ser Val Pro Gln Phe  Lys Lys Val
        995                 1000                1005

Val Phe  Gln Glu Phe Thr Asp  Gly Ser Phe Thr Gln  Pro Leu Tyr
    1010                1015                1020

Arg Gly  Glu Leu Asn Glu His  Leu Gly Leu Leu Gly  Pro Tyr Ile
    1025                1030                1035

Arg Ala  Glu Val Glu Asp Asn  Ile Met Val Thr Phe  Arg Asn Gln
    1040                1045                1050

Ala Ser  Arg Pro Tyr Ser Phe  Tyr Ser Ser Leu Ile  Ser Tyr Glu
    1055                1060                1065

Glu Asp  Gln Arg Gln Gly Ala  Glu Pro Arg Lys Asn  Phe Val Lys
    1070                1075                1080
```

```
Pro Asn  Glu Thr Lys Thr Tyr  Phe Trp Lys Val Gln  His His Met
    1085                1090                1095

Ala Pro  Thr Lys Asp Glu Phe  Asp Cys Lys Ala Trp  Ala Tyr Phe
    1100                1105                1110

Ser Asp  Val Asp Leu Glu Lys  Asp Val His Ser Gly  Leu Ile Gly
    1115                1120                1125

Pro Leu  Leu Val Cys His Thr  Asn Thr Leu Asn Pro  Ala His Gly
    1130                1135                1140

Arg Gln  Val Thr Val Gln Glu  Phe Ala Leu Phe Phe  Thr Ile Phe
    1145                1150                1155

Asp Glu  Thr Lys Ser Trp Tyr  Phe Thr Glu Asn Met  Glu Arg Asn
    1160                1165                1170

Cys Arg  Ala Pro Cys Asn Ile  Gln Met Glu Asp Pro  Thr Phe Lys
    1175                1180                1185

Glu Asn  Tyr Arg Phe His Ala  Ile Asn Gly Tyr Ile  Met Asp Thr
    1190                1195                1200

Leu Pro  Gly Leu Val Met Ala  Gln Asp Gln Arg Ile  Arg Trp Tyr
    1205                1210                1215

Leu Leu  Ser Met Gly Ser Asn  Glu Asn Ile His Ser  Ile His Phe
    1220                1225                1230

Ser Gly  His Val Phe Thr Val  Arg Lys Lys Glu Glu  Tyr Lys Met
    1235                1240                1245

Ala Leu  Tyr Asn Leu Tyr Pro  Gly Val Phe Glu Thr  Val Glu Met
    1250                1255                1260
```

```
Leu Pro  Ser Lys Ala Gly Ile  Trp Arg Val Glu Cys  Leu Ile Gly
    1265                 1270                 1275

Glu His  Leu His Ala Gly Met  Ser Thr Leu Phe Leu  Val Tyr Ser
    1280                 1285                 1290

Asn Lys  Cys Gln Thr Pro Leu  Gly Met Ala Ser Gly  His Ile Arg
    1295                 1300                 1305

Asp Phe  Gln Ile Thr Ala Ser  Gly Gln Tyr Gly Gln  Trp Ala Pro
    1310                 1315                 1320

Lys Leu  Ala Arg Leu His Tyr  Ser Gly Ser Ile Asn  Ala Trp Ser
    1325                 1330                 1335

Thr Lys  Glu Pro Phe Ser Trp  Ile Lys Val Asp Leu  Leu Ala Pro
    1340                 1345                 1350

Met Ile  Ile His Gly Ile Lys  Thr Gln Gly Ala Arg  Gln Lys Phe
    1355                 1360                 1365

Ser Ser  Leu Tyr Ile Ser Gln  Phe Ile Ile Met Tyr  Ser Leu Asp
    1370                 1375                 1380

Gly Lys  Lys Trp Gln Thr Tyr  Arg Gly Asn Ser Thr  Gly Thr Leu
    1385                 1390                 1395

Met Val  Phe Phe Gly Asn Val  Asp Ser Ser Gly Ile  Lys His Asn
    1400                 1405                 1410

Ile Phe  Asn Pro Pro Ile Ile  Ala Arg Tyr Ile Arg  Leu His Pro
    1415                 1420                 1425

Thr His  Tyr Ser Ile Arg Ser  Thr Leu Arg Met Glu  Leu Met Gly
    1430                 1435                 1440
```

```
Cys Asp  Leu Asn Ser Cys Ser  Met Pro Leu Gly Met  Glu Ser Lys
    1445                 1450                 1455

Ala Ile  Ser Asp Ala Gln Ile  Thr Ala Ser Ser Tyr  Phe Thr Asn
    1460                 1465                 1470

Met Phe  Ala Thr Trp Ser Pro  Ser Lys Ala Arg Leu  His Leu Gln
    1475                 1480                 1485

Gly Arg  Ser Asn Ala Trp Arg  Pro Gln Val Asn Asn  Pro Lys Glu
    1490                 1495                 1500

Trp Leu  Gln Val Asp Phe Gln  Lys Thr Met Lys Val  Thr Gly Val
    1505                 1510                 1515

Thr Thr  Gln Gly Val Lys Ser  Leu Leu Thr Ser Met  Tyr Val Lys
    1520                 1525                 1530

Glu Phe  Leu Ile Ser Ser Ser  Gln Asp Gly His Gln  Trp Thr Leu
    1535                 1540                 1545

Phe Phe  Gln Asn Gly Lys Val  Lys Val Phe Gln Gly  Asn Gln Asp
    1550                 1555                 1560

Ser Phe  Thr Pro Val Val Asn  Ser Leu Asp Pro Pro  Leu Leu Thr
    1565                 1570                 1575

Arg Tyr  Leu Arg Ile His Pro  Gln Ser Trp Val His  Gln Ile Ala
    1580                 1585                 1590

Leu Arg  Met Glu Val Leu Gly  Cys Glu Ala Gln Asp  Leu Tyr
    1595                 1600                 1605
```

<210> 104

<211> 16

<212> DNA
<213> 人工序列

<220>
<223> AAV2的Rep結合位點（RBS）

<400> 104
gcgcgctcgc tcgctc 16

<210> 105
<211> 6
<212> DNA
<213> 人工序列

<220>
<223> AAV2的末端解鏈位點（TRS）

<400> 105
agttgg 6

<210> 106
<211> 2328
<212> PRT
<213> 人工序列

<220>
<223> 成熟全長FVIII

<400> 106
Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser Trp Asp Tyr
1 5 10 15

Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg Phe Pro Pro
20 25 30

```
Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val Tyr Lys Lys
        35                  40                  45

Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile Ala Lys Pro
    50                  55                  60

Arg Pro Pro Trp Met Gly Leu Leu Gly Pro Thr Ile Gln Ala Glu Val
65                  70                  75                  80

Tyr Asp Thr Val Val Ile Thr Leu Lys Asn Met Ala Ser His Pro Val
                85                  90                  95

Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ala Ser Glu Gly Ala
            100                 105                 110

Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp Asp Lys Val
        115                 120                 125

Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu Lys Glu Asn
    130                 135                 140

Gly Pro Met Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser Tyr Leu Ser
145                 150                 155                 160

His Val Asp Leu Val Lys Asp Leu Asn Ser Gly Leu Ile Gly Ala Leu
                165                 170                 175

Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr Gln Thr Leu
            180                 185                 190

His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly Lys Ser Trp
        195                 200                 205

His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp Ala Ala Ser
    210                 215                 220
```

```
Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr Val Asn Arg
225                 230                 235                 240

Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val Tyr Trp His
                245                 250                 255

Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile Phe Leu Glu
            260                 265                 270

Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser Leu Glu Ile
        275                 280                 285

Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met Asp Leu Gly
    290                 295                 300

Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His Asp Gly Met
305                 310                 315                 320

Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro Gln Leu Arg
                325                 330                 335

Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp Leu Thr Asp
            340                 345                 350

Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser Pro Ser Phe
        355                 360                 365

Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr Trp Val His
    370                 375                 380

Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro Leu Val Leu
385                 390                 395                 400

Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn Asn Gly Pro
                405                 410                 415
```

Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met Ala Tyr Thr
420 425 430

Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu Ser Gly Ile
435 440 445

Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu Leu Ile Ile
450 455 460

Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro His Gly Ile
465 470 475 480

Thr Asp Val Arg Pro Lys Arg Arg Leu Pro Lys Gly Val Lys His Leu
485 490 495

Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe Lys Tyr Lys Trp Thr
500 505 510

Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp Pro Arg Cys Leu Thr
515 520 525

Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg Asp Leu Ala Ser Gly
530 535 540

Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu Ser Val Asp Gln Arg
545 550 555 560

Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val Ile Leu Phe Ser Val
565 570 575

Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu Asn Ile Gln Arg Phe
580 585 590

Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp Pro Glu Phe Gln Ala
595 600 605

```
Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val Phe Asp Ser Leu Gln
    610                 615                 620

Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp Tyr Ile Leu Ser Ile
625                 630                 635                 640

Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe Ser Gly Tyr Thr Phe
                645                 650                 655

Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr Leu Phe Pro Phe Ser
            660                 665                 670

Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro Gly Leu Trp Ile Leu
        675                 680                 685

Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly Met Thr Ala Leu Leu
    690                 695                 700

Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp Tyr Tyr Glu Asp Ser
705                 710                 715                 720

Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys Asn Asn Ala Ile Glu
                725                 730                 735

Pro Arg Ser Phe Ser Gln Asn Ser Arg His Pro Ser Thr Arg Gln Lys
            740                 745                 750

Gln Phe Asn Ala Thr Thr Ile Pro Glu Asn Asp Ile Glu Lys Thr Asp
        755                 760                 765

Pro Trp Phe Ala His Arg Thr Pro Met Pro Lys Ile Gln Asn Val Ser
    770                 775                 780

Ser Ser Asp Leu Leu Met Leu Leu Arg Gln Ser Pro Thr Pro His Gly
785                 790                 795                 800
```

Leu Ser Leu Ser Asp Leu Gln Glu Ala Lys Tyr Glu Thr Phe Ser Asp
805 810 815

Asp Pro Ser Pro Gly Ala Ile Asp Ser Asn Asn Ser Leu Ser Glu Met
820 825 830

Thr His Phe Arg Pro Gln Leu His His Ser Gly Asp Met Val Phe Thr
835 840 845

Pro Glu Ser Gly Leu Gln Leu Arg Leu Asn Glu Lys Leu Gly Thr Thr
850 855 860

Ala Ala Thr Glu Leu Lys Lys Leu Asp Phe Lys Val Ser Ser Thr Ser
865 870 875 880

Asn Asn Leu Ile Ser Thr Ile Pro Ser Asp Asn Leu Ala Ala Gly Thr
885 890 895

Asp Asn Thr Ser Ser Leu Gly Pro Pro Ser Met Pro Val His Tyr Asp
900 905 910

Ser Gln Leu Asp Thr Thr Leu Phe Gly Lys Lys Ser Ser Pro Leu Thr
915 920 925

Glu Ser Gly Gly Pro Leu Ser Leu Ser Glu Glu Asn Asn Asp Ser Lys
930 935 940

Leu Leu Glu Ser Gly Leu Met Asn Ser Gln Glu Ser Ser Trp Gly Lys
945 950 955 960

Asn Val Ser Ser Thr Glu Ser Gly Arg Leu Phe Lys Gly Lys Arg Ala
965 970 975

His Gly Pro Ala Leu Leu Thr Lys Asp Asn Ala Leu Phe Lys Val Ser
980 985 990

```
Ile Ser Leu Leu Lys Thr Asn Lys  Thr Ser Asn Asn Ser  Ala Thr Asn
        995                 1000                1005

Arg Lys  Thr His Ile Asp Gly  Pro Ser Leu Leu Ile  Glu Asn Ser
    1010                1015                1020

Pro Ser  Val Trp Gln Asn Ile  Ser Asp Thr Glu Phe  Lys Lys Val
    1025                1030                1035

Thr Pro  Leu Ile His Asp Arg  Met Leu Met Asp Lys  Asn Ala Thr
    1040                1045                1050

Ala Leu  Arg Leu Asn His Met  Ser Asn Lys Thr Thr  Ser Ser Lys
    1055                1060                1065

Asn Met  Glu Met Val Gln Gln  Lys Lys Glu Gly Pro  Ile Pro Pro
    1070                1075                1080

Asp Ala  Gln Asn Pro Asp Met  Ser Phe Phe Lys Met  Leu Phe Leu
    1085                1090                1095

Pro Glu  Ser Ala Arg Trp Ile  Gln Arg Thr His Gly  Lys Asn Ser
    1100                1105                1110

Leu Asn  Ser Gly Gln Gly Pro  Ser Pro Lys Gln Leu  Val Ser Leu
    1115                1120                1125

Gly Pro  Glu Lys Ser Val Glu  Gly Gln Asn Phe Leu  Ser Glu Lys
    1130                1135                1140

Asn Lys  Val Val Val Gly Lys  Gly Glu Phe Thr Lys  Asp Val Gly
    1145                1150                1155

Leu Lys  Glu Met Val Phe Pro  Ser Ser Arg Asn Leu  Phe Leu Thr
    1160                1165                1170
```

```
Asn Leu  Asp Asn Leu His Glu  Asn Asn Thr His Asn  Gln Glu Lys
    1175                 1180                 1185

Lys Ile  Gln Glu Glu Ile Glu  Lys Lys Glu Thr Leu  Ile Gln Glu
    1190                 1195                 1200

Asn Val  Val Leu Pro Gln Ile  His Thr Val Thr Gly  Thr Lys Asn
    1205                 1210                 1215

Phe Met  Lys Asn Leu Phe Leu  Leu Ser Thr Arg Gln  Asn Val Glu
    1220                 1225                 1230

Gly Ser  Tyr Asp Gly Ala Tyr  Ala Pro Val Leu Gln  Asp Phe Arg
    1235                 1240                 1245

Ser Leu  Asn Asp Ser Thr Asn  Arg Thr Lys Lys His  Thr Ala His
    1250                 1255                 1260

Phe Ser  Lys Lys Gly Glu Glu  Glu Asn Leu Glu Gly  Leu Gly Asn
    1265                 1270                 1275

Gln Thr  Lys Gln Ile Val Glu  Lys Tyr Ala Cys Thr  Thr Arg Ile
    1280                 1285                 1290

Ser Pro  Asn Thr Ser Gln Gln  Asn Phe Val Thr Gln  Arg Ser Lys
    1295                 1300                 1305

Arg Ala  Leu Lys Gln Phe Arg  Leu Pro Leu Glu Glu  Thr Glu Leu
    1310                 1315                 1320

Glu Lys  Arg Ile Ile Val Asp  Asp Thr Ser Thr Gln  Trp Ser Lys
    1325                 1330                 1335

Asn Met  Lys His Leu Thr Pro  Ser Thr Leu Thr Gln  Ile Asp Tyr
    1340                 1345                 1350
```

```
Asn Glu  Lys Glu Lys Gly Ala  Ile Thr Gln Ser Pro  Leu Ser Asp
    1355                 1360                 1365

Cys Leu  Thr Arg Ser His Ser  Ile Pro Gln Ala Asn  Arg Ser Pro
    1370                 1375                 1380

Leu Pro  Ile Ala Lys Val Ser  Ser Phe Pro Ser Ile  Arg Pro Ile
    1385                 1390                 1395

Tyr Leu  Thr Arg Val Leu Phe  Gln Asp Asn Ser Ser  His Leu Pro
    1400                 1405                 1410

Ala Ala  Ser Tyr Arg Lys Lys  Asp Ser Gly Val Gln  Glu Ser Ser
    1415                 1420                 1425

His Phe  Leu Gln Gly Ala Lys  Lys Asn Asn Leu Ser  Leu Ala Ile
    1430                 1435                 1440

Leu Thr  Leu Glu Met Thr Gly  Asp Gln Arg Glu Val  Gly Ser Leu
    1445                 1450                 1455

Gly Thr  Ser Ala Thr Asn Ser  Val Thr Tyr Lys Lys  Val Glu Asn
    1460                 1465                 1470

Thr Val  Leu Pro Lys Pro Asp  Leu Pro Lys Thr Ser  Gly Lys Val
    1475                 1480                 1485

Glu Leu  Leu Pro Lys Val His  Ile Tyr Gln Lys Asp  Leu Phe Pro
    1490                 1495                 1500

Thr Glu  Thr Ser Asn Gly Ser  Pro Gly His Leu Asp  Leu Val Glu
    1505                 1510                 1515

Gly Ser  Leu Leu Gln Gly Thr  Glu Gly Ala Ile Lys  Trp Asn Glu
    1520                 1525                 1530
```

```
Ala Asn  Arg Pro Gly Lys Val  Pro Phe Leu Arg Val  Ala Thr Glu
    1535                 1540                 1545

Ser Ser  Ala Lys Thr Pro Ser  Lys Leu Leu Asp Pro  Leu Ala Trp
    1550                 1555                 1560

Asp Asn  His Tyr Gly Thr Gln  Ile Pro Lys Glu Glu  Trp Lys Ser
    1565                 1570                 1575

Gln Glu  Lys Ser Pro Glu Lys  Thr Ala Phe Lys Lys  Lys Asp Thr
    1580                 1585                 1590

Ile Leu  Ser Leu Asn Ala Cys  Glu Ser Asn His Ala  Ile Ala Ala
    1595                 1600                 1605

Ile Asn  Glu Gly Gln Asn Lys  Pro Glu Ile Glu Val  Thr Trp Ala
    1610                 1615                 1620

Lys Gln  Gly Arg Thr Glu Arg  Leu Cys Ser Gln Asn  Pro Pro Val
    1625                 1630                 1635

Leu Lys  Arg His Gln Arg Glu  Ile Thr Arg Thr Thr  Leu Gln Ser
    1640                 1645                 1650

Asp Gln  Glu Glu Ile Asp Tyr  Asp Asp Thr Ile Ser  Val Glu Met
    1655                 1660                 1665

Lys Lys  Glu Asp Phe Asp Ile  Tyr Asp Glu Asp Glu  Asn Gln Ser
    1670                 1675                 1680

Pro Arg  Ser Phe Gln Lys Lys  Thr Arg His Tyr Phe  Ile Ala Ala
    1685                 1690                 1695

Val Glu  Arg Leu Trp Asp Tyr  Gly Met Ser Ser Ser  Pro His Val
    1700                 1705                 1710
```

Leu Arg Asn Arg Ala Gln Ser Gly Ser Val Pro Gln Phe Lys Lys
1715 1720 1725

Val Val Phe Gln Glu Phe Thr Asp Gly Ser Phe Thr Gln Pro Leu
1730 1735 1740

Tyr Arg Gly Glu Leu Asn Glu His Leu Gly Leu Leu Gly Pro Tyr
1745 1750 1755

Ile Arg Ala Glu Val Glu Asp Asn Ile Met Val Thr Phe Arg Asn
1760 1765 1770

Gln Ala Ser Arg Pro Tyr Ser Phe Tyr Ser Ser Leu Ile Ser Tyr
1775 1780 1785

Glu Glu Asp Gln Arg Gln Gly Ala Glu Pro Arg Lys Asn Phe Val
1790 1795 1800

Lys Pro Asn Glu Thr Lys Thr Tyr Phe Trp Lys Val Gln His His
1805 1810 1815

Met Ala Pro Thr Lys Asp Glu Phe Asp Cys Lys Ala Trp Ala Tyr
1820 1825 1830

Phe Ser Asp Val Asp Leu Glu Lys Asp Val His Ser Gly Leu Ile
1835 1840 1845

Gly Pro Leu Leu Val Cys His Thr Asn Thr Leu Asn Pro Ala His
1850 1855 1860

Gly Arg Gln Val Thr Val Gln Glu Phe Ala Leu Phe Phe Thr Ile
1865 1870 1875

Phe Asp Glu Thr Lys Ser Trp Tyr Phe Thr Glu Asn Met Glu Arg
1880 1885 1890

```
Asn Cys  Arg Ala Pro Cys Asn  Ile Gln Met Glu Asp  Pro Thr Phe
    1895                 1900                 1905

Lys Glu  Asn Tyr Arg Phe His  Ala Ile Asn Gly Tyr  Ile Met Asp
    1910                 1915                 1920

Thr Leu  Pro Gly Leu Val Met  Ala Gln Asp Gln Arg  Ile Arg Trp
    1925                 1930                 1935

Tyr Leu  Leu Ser Met Gly Ser  Asn Glu Asn Ile His  Ser Ile His
    1940                 1945                 1950

Phe Ser  Gly His Val Phe Thr  Val Arg Lys Lys Glu  Glu Tyr Lys
    1955                 1960                 1965

Met Ala  Leu Tyr Asn Leu Tyr  Pro Gly Val Phe Glu  Thr Val Glu
    1970                 1975                 1980

Met Leu  Pro Ser Lys Ala Gly  Ile Trp Arg Val Glu  Cys Leu Ile
    1985                 1990                 1995

Gly Glu  His Leu His Ala Gly  Met Ser Thr Leu Phe  Leu Val Tyr
    2000                 2005                 2010

Ser Asn  Lys Cys Gln Thr Pro  Leu Gly Met Ala Ser  Gly His Ile
    2015                 2020                 2025

Arg Asp  Phe Gln Ile Thr Ala  Ser Gly Gln Tyr Gly  Gln Trp Ala
    2030                 2035                 2040

Pro Lys  Leu Ala Arg Leu His  Tyr Ser Gly Ser Ile  Asn Ala Trp
    2045                 2050                 2055

Ser Thr  Lys Glu Pro Phe Ser  Trp Ile Lys Val Asp  Leu Leu Ala
    2060                 2065                 2070
```

Pro Met Ile Ile His Gly Ile Lys Thr Gln Gly Ala Arg Gln Lys
2075 2080 2085

Phe Ser Ser Leu Tyr Ile Ser Gln Phe Ile Ile Met Tyr Ser Leu
2090 2095 2100

Asp Gly Lys Lys Trp Gln Thr Tyr Arg Gly Asn Ser Thr Gly Thr
2105 2110 2115

Leu Met Val Phe Phe Gly Asn Val Asp Ser Ser Gly Ile Lys His
2120 2125 2130

Asn Ile Phe Asn Pro Pro Ile Ile Ala Arg Tyr Ile Arg Leu His
2135 2140 2145

Pro Thr His Tyr Ser Ile Arg Ser Thr Leu Arg Met Glu Leu Met
2150 2155 2160

Gly Cys Asp Leu Asn Ser Cys Ser Met Pro Leu Gly Met Glu Ser
2165 2170 2175

Lys Ala Ile Ser Asp Ala Gln Ile Thr Ala Ser Ser Tyr Phe Thr
2180 2185 2190

Asn Met Phe Ala Thr Trp Ser Pro Ser Lys Ala Arg Leu His Leu
2195 2200 2205

Gln Gly Arg Ser Asn Ala Trp Arg Pro Gln Val Asn Asn Pro Lys
2210 2215 2220

Glu Trp Leu Gln Val Asp Phe Gln Lys Thr Met Lys Val Thr Gly
2225 2230 2235

Val Thr Thr Gln Gly Val Lys Ser Leu Leu Thr Ser Met Tyr Val
2240 2245 2250

Lys Glu Phe Leu Ile Ser Ser Ser Gln Asp Gly His Gln Trp Thr
2255 2260 2265

Leu Phe Phe Gln Asn Gly Lys Val Lys Val Phe Gln Gly Asn Gln
2270 2275 2280

Asp Ser Phe Thr Pro Val Val Asn Ser Leu Asp Pro Pro Leu Leu
2285 2290 2295

Thr Arg Tyr Leu Arg Ile His Pro Gln Ser Trp Val His Gln Ile
2300 2305 2310

Ala Leu Arg Met Glu Val Leu Gly Cys Glu Ala Gln Asp Leu Tyr
2315 2320 2325

<210> 107
<211> 7053
<212> DNA
<213> 人工序列

<220>
<223> 成熟全長FVIII

<400> 107
atgcaaatag agctctccac ctgcttcttt ctgtgccttt tgcgattctg ctttagtgcc 60
accagaagat actacctggg tgcagtggaa ctgtcatggg actatatgca aagtgatctc 120
ggtgagctgc ctgtggacgc aagatttcct cctagagtgc caaaatcttt tccattcaac 180
acctcagtcg tgtacaaaaa gactctgttt gtagaattca cggatcacct tttcaacatc 240
gctaagccaa ggccaccctg gatgggtctg ctaggtccta ccatccaggc tgaggtttat 300
gatacagtgg tcattacact taagaacatg gcttcccatc ctgtcagtct tcatgctgtt 360
ggtgtatcct actggaaagc ttctgaggga gctgaatatg atgatcagac cagtcaaagg 420
gagaaagaag atgataaagt cttccctggt ggaagccata catatgtctg gcaggtcctg 480
aaagagaatg gtccaatggc ctctgaccca ctgtgcctta cctactcata tctttctcat 540
gtggacctgg taaaagactt gaattcaggc ctcattggag ccctactagt atgtagagaa 600

gggagtctgg ccaaggaaaa gacacagacc ttgcacaaat ttatactact ttttgctgta 660

tttgatgaag ggaaaagttg gcactcagaa acaaagaact ccttgatgca ggatagggat 720

gctgcatctg ctcgggcctg gcctaaaatg cacacagtca atggttatgt aaacaggtct 780

ctgccaggtc tgattggatg ccacaggaaa tcagtctatt ggcatgtgat tggaatgggc 840

accactcctg aagtgcactc aatattcctc gaaggtcaca catttcttgt gaggaaccat 900

cgccaggcgt ccttggaaat ctcgccaata actttcctta ctgctcaaac actcttgatg 960

gaccttggac agtttctact gttttgtcat atctcttccc accaacatga tggcatggaa 1020

gcttatgtca aagtagacag ctgtccagag gaaccccaac tacgaatgaa aaataatgaa 1080

gaagcggaag actatgatga tgatcttact gattctgaaa tggatgtggt caggtttgat 1140

gatgacaact ctccttcctt tatccaaatt cgctcagttg ccaagaagca tcctaaaact 1200

tgggtacatt acattgctgc tgaagaggag gactgggact atgctccctt agtcctcgcc 1260

cccgatgaca gaagttataa aagtcaatat ttgaacaatg gccctcagcg gattggtagg 1320

aagtacaaaa aagtccgatt tatggcatac acagatgaaa cctttaagac tcgtgaagct 1380

attcagcatg aatcaggaat cttgggacct ttactttatg gggaagttgg agacacactg 1440

ttgattatat ttaagaatca agcaagcaga ccatataaca tctaccctca cggaatcact 1500

gatgtccgtc ctttgtattc aaggagatta ccaaaaggtg taaaacattt gaaggatttt 1560

ccaattctgc caggagaaat attcaaatat aaatggacag tgactgtaga agatgggcca 1620

actaaatcag atcctcggtg cctgacccgc tattactcta gtttcgttaa tatggagaga 1680

gatctagctt caggactcat tggccctctc ctcatctgct acaaagaatc tgtagatcaa 1740

agaggaaacc agataatgtc agacaagagg aatgtcatcc tgttttctgt atttgatgag 1800

aaccgaagct ggtacctcac agagaatata caacgctttc tccccaatcc agctggagtg 1860

cagcttgagg atccagagtt ccaagcctcc aacatcatgc acagcatcaa tggctatgtt 1920

tttgatagtt tgcagttgtc agtttgtttg catgaggtgg catactggta cattctaagc 1980

attggagcac agactgactt cctttctgtc ttcttctctg gatatacctt caaacacaaa 2040

atggtctatg aagacacact caccctattc ccattctcag gagaaactgt cttcatgtcg 2100

atggaaaacc caggtctatg gattctgggg tgccacaact cagactttcg gaacagaggc 2160

atgaccgcct tactgaaggt ttctagttgt gacaagaaca ctggtgatta ttacgaggac 2220

agttatgaag atatttcagc atacttgctg agtaaaaaca atgccattga accaagaagc 2280

ttctcccaga attcaagaca ccctagcact aggcaaaagc aatttaatgc caccacaatt 2340

ccagaaaatg acatagagaa gactgaccct tggtttgcac acagaacacc tatgcctaaa 2400

atacaaaatg tctcctctag tgatttgttg atgctcttgc gacagagtcc tactccacat 2460

gggctatcct tatctgatct ccaagaagcc aaatatgaga ctttttctga tgatccatca 2520

cctggagcaa tagacagtaa taacagcctg tctgaaatga cacacttcag gccacagctc 2580

catcacagtg gggacatggt atttacccct gagtcaggcc tccaattaag attaaatgag 2640

aaactgggga caactgcagc aacagagttg aagaaacttg atttcaaagt ttctagtaca 2700

tcaaataatc tgatttcaac aattccatca gacaatttgg cagcaggtac tgataataca 2760

agttccttag gacccccaag tatgccagtt cattatgata gtcaattaga taccactcta 2820
tttggcaaaa agtcatctcc ccttactgag tctggtggac ctctgagctt gagtgaagaa 2880
aataatgatt caaagttgtt agaatcaggt ttaatgaata gccaagaaag ttcatgggga 2940
aaaaatgtat cgtcaacaga gagtggtagg ttatttaaag ggaaaagagc tcatggacct 3000
gctttgttga ctaaagataa tgccttattc aaagttagca tctctttgtt aaagacaaac 3060
aaaacttcca ataattcagc aactaataga aagactcaca ttgatggccc atcattatta 3120
attgagaata gtccatcagt ctggcaaaat atattagaaa gtgacactga gtttaaaaaa 3180
gtgacacctt tgattcatga cagaatgctt atggacaaaa atgctacagc tttgaggcta 3240
aatcatatgt caaataaaac tacttcatca aaaaacatgg aaatggtcca acagaaaaaa 3300
gagggcccca ttccaccaga tgcacaaaat ccagatatgt cgttctttaa gatgctattc 3360
ttgccagaat cagcaaggtg gatacaaagg actcatggaa agaactctct gaactctggg 3420
caaggcccca gtccaaagca attagtatcc ttaggaccag aaaaatctgt ggaaggtcag 3480
aatttcttgt ctgagaaaaa caaagtggta gtaggaaagg gtgaatttac aaaggacgta 3540
ggactcaaag agatggtttt tccaagcagc agaaacctat ttcttactaa cttggataat 3600
ttacatgaaa ataatacaca caatcaagaa aaaaaaattc aggaagaaat agaaaagaag 3660
gaaacattaa tccaagagaa tgtagttttg cctcagatac atacagtgac tggcactaag 3720
aatttcatga agaacctttt cttactgagc actaggcaaa atgtagaagg ttcatatgac 3780
ggggcatatg ctccagtact tcaagatttt aggtcattaa atgattcaac aaatagaaca 3840
aagaaacaca cagctcattt ctcaaaaaaa ggggaggaag aaaacttgga aggcttggga 3900
aatcaaacca agcaaattgt agagaaatat gcatgcacca caaggatatc tcctaataca 3960
agccagcaga attttgtcac gcaacgtagt aagagagctt tgaaacaatt cagactccca 4020
ctagaagaaa cagaacttga aaaaaggata attgtggatg acacctcaac ccagtggtcc 4080
aaaaacatga aacatttgac cccgagcacc ctcacacaga tagactacaa tgagaaggag 4140
aaaggggcca ttactcagtc tcccttatca gattgcctta cgaggagtca tagcatccct 4200
caagcaaata gatctccatt acccattgca aaggtatcat catttccatc tattagacct 4260
atatatctga ccagggtcct attccaagac aactcttctc atcttccagc agcatcttat 4320
agaaagaaag attctggggt ccaagaaagc agtcatttct tacaaggagc caaaaaaaat 4380
aacctttctt tagccattct aaccttggag atgactggtg atcaaagaga ggttggctcc 4440
ctggggacaa gtgccacaaa ttcagtcaca tacaagaaag ttgagaacac tgttctcccg 4500
aaaccagact tgcccaaaac atctggcaaa gttgaattgc ttccaaaagt tcacatttat 4560
cagaaggacc tattccctac ggaaactagc aatgggtctc ctggccatct ggatctcgtg 4620
gaagggagcc ttcttcaggg aacagaggga gcgattaagt ggaatgaagc aaacagacct 4680
ggaaaagttc cctttctgag agtagcaaca gaaagctctg caaagactcc ctccaagcta 4740
ttggatcctc ttgcttggga taaccactat ggtactcaga taccaaaaga agagtggaaa 4800
tcccaagaga agtcaccaga aaaaacagct tttaagaaaa aggataccat tttgtccctg 4860
aacgcttgtg aaagcaatca tgcaatagca gcaataaatg agggacaaaa taagcccgaa 4920

atagaagtca cctgggcaaa gcaaggtagg actgaaaggc tgtgctctca aaacccacca 4980
gtcttgaaac gccatcaacg ggaaataact cgtactactc ttcagtcaga tcaagaggaa 5040
attgactatg atgataccat atcagttgaa atgaagaagg aagattttga catttatgat 5100
gaggatgaaa atcagagccc ccgcagcttt caaaagaaaa cacgacacta ttttattgct 5160
gcagtggaga ggctctggga ttatgggatg agtagctccc cacatgttct aagaaacagg 5220
gctcagagtg gcagtgtccc tcagttcaag aaagttgttt tccaggaatt tactgatggc 5280
tcctttactc agcccttata ccgtggagaa ctaaatgaac atttgggact cctggggcca 5340
tatataagag cagaagttga agataatatc atggtaactt tcagaaatca ggcctctcgt 5400
ccctattcct tctattctag ccttatttct tatgaggaag atcagaggca aggagcagaa 5460
cctagaaaaa actttgtcaa gcctaatgaa accaaaactt acttttggaa agtgcaacat 5520
catatggcac ccactaaaga tgagtttgac tgcaaagcct gggcttattt ctctgatgtt 5580
gacctggaaa aagatgtgca ctcaggcctg attggacccc ttctggtctg ccacactaac 5640
acactgaacc ctgctcatgg gagacaagtg acagtacagg aatttgctct gtttttcacc 5700
atctttgatg agaccaaaag ctggtacttc actgaaaata tggaaagaaa ctgcagggct 5760
ccctgcaata tccagatgga agatcccact tttaaagaga attatcgctt ccatgcaatc 5820
aatggctaca taatggatac actacctggc ttagtaatgg ctcaggatca aaggattcga 5880
tggtatctgc tcagcatggg cagcaatgaa aacatccatt ctattcattt cagtggacat 5940
gtgttcactg tacgaaaaaa agaggagtat aaaatggcac tgtacaatct ctatccaggt 6000
gtttttgaga cagtggaaat gttaccatcc aaagctggaa tttggcgggt ggaatgcctt 6060
attggcgagc atctacatgc tgggatgagc acacttttc tggtgtacag caataagtgt 6120
cagactcccc tgggaatggc ttctggacac attagagatt ttcagattac agcttcagga 6180
caatatggac agtgggcccc aaagctggcc agacttcatt attccggatc aatcaatgcc 6240
tggagcacca aggagccctt ttcttggatc aaggtggatc tgttggcacc aatgattatt 6300
cacggcatca agacccaggg tgcccgtcag aagttctcca gcctctacat ctctcagttt 6360
atcatcatgt atagtcttga tgggaagaag tggcagactt atcgaggaaa ttccactgga 6420
accttaatgg tcttctttgg caatgtggat tcatctggga taaaacacaa tatttttaac 6480
cctccaatta ttgctcgata catccgtttg cacccaactc attatagcat tcgcagcact 6540
cttcgcatgg agttgatggg ctgtgattta aatagttgca gcatgccatt gggaatggag 6600
agtaaagcaa tatcagatgc acagattact gcttcatcct actttaccaa tatgtttgcc 6660
acctggtctc cttcaaaagc tcgacttcac ctccaaggga ggagtaatgc ctggagacct 6720
caggtgaata atccaaaaga gtggctgcaa gtggacttcc agaagacaat gaaagtcaca 6780
ggagtaacta ctcagggagt aaaatctctg cttaccagca tgtatgtgaa ggagttcctc 6840
atctccagca gtcaagatgg ccatcagtgg actctctttt ttcagaatgg caaagtaaag 6900
gtttttcagg gaaatcaaga ctccttcaca cctgtggtga actctctaga cccaccgtta 6960
ctgactcgct accttcgaat tcacccccag agttgggtgc accagattgc cctgaggatg 7020
gaggttctgg gctgcgaggc acaggacctc tac 7053

<210> 108

<400> 108
000

<210> 109
<211> 1718
<212> PRT
<213> 人工序列

<220>
<223> BIVV001 BDD-FVIII(XTEN)序列（無Fc）

<400> 109

```
Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser Trp Asp Tyr
1               5                  10                  15

Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg Phe Pro Pro
            20                  25                  30

Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val Tyr Lys Lys
        35                  40                  45

Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile Ala Lys Pro
    50                  55                  60

Arg Pro Pro Trp Met Gly Leu Leu Gly Pro Thr Ile Gln Ala Glu Val
65                  70                  75                  80

Tyr Asp Thr Val Val Ile Thr Leu Lys Asn Met Ala Ser His Pro Val
                85                  90                  95

Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ala Ser Glu Gly Ala
```

```
            100                 105                 110

Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp Asp Lys Val
        115                 120                 125

Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu Lys Glu Asn
    130                 135                 140

Gly Pro Met Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser Tyr Leu Ser
145                 150                 155                 160

His Val Asp Leu Val Lys Asp Leu Asn Ser Gly Leu Ile Gly Ala Leu
                165                 170                 175

Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr Gln Thr Leu
            180                 185                 190

His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly Lys Ser Trp
        195                 200                 205

His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp Ala Ala Ser
    210                 215                 220

Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr Val Asn Arg
225                 230                 235                 240

Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val Tyr Trp His
                245                 250                 255

Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile Phe Leu Glu
            260                 265                 270

Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser Leu Glu Ile
        275                 280                 285

Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met Asp Leu Gly
```

```
    290                 295                 300

Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His Asp Gly Met
305                 310                 315                 320

Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro Gln Leu Arg
                325                 330                 335

Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp Leu Thr Asp
            340                 345                 350

Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser Pro Ser Phe
        355                 360                 365

Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr Trp Val His
    370                 375                 380

Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro Leu Val Leu
385                 390                 395                 400

Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn Asn Gly Pro
                405                 410                 415

Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met Ala Tyr Thr
            420                 425                 430

Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu Ser Gly Ile
        435                 440                 445

Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu Leu Ile Ile
    450                 455                 460

Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro His Gly Ile
465                 470                 475                 480

Thr Asp Val Arg Pro Lys Arg Arg Leu Pro Lys Gly Val Lys His Leu
```

485 490 495

Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe Lys Tyr Lys Trp Thr
500 505 510

Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp Pro Arg Cys Leu Thr
515 520 525

Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg Asp Leu Ala Ser Gly
530 535 540

Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu Ser Val Asp Gln Arg
545 550 555 560

Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val Ile Leu Phe Ser Val
565 570 575

Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu Asn Ile Gln Arg Phe
580 585 590

Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp Pro Glu Phe Gln Ala
595 600 605

Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val Phe Asp Ser Leu Gln
610 615 620

Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp Tyr Ile Leu Ser Ile
625 630 635 640

Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe Ser Gly Tyr Thr Phe
645 650 655

Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr Leu Phe Pro Phe Ser
660 665 670

Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro Gly Leu Trp Ile Leu

675 680 685

Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly Met Thr Ala Leu Leu
690 695 700

Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp Tyr Tyr Glu Asp Ser
705 710 715 720

Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys Asn Asn Ala Ile Glu
725 730 735

Pro Arg Ser Phe Ser Gln Asn Gly Thr Ser Glu Ser Ala Thr Pro Glu
740 745 750

Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly
755 760 765

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala
770 775 780

Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
785 790 795 800

Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
805 810 815

Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Glu Ser
820 825 830

Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser
835 840 845

Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly
850 855 860

Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Ser Pro Ala Gly

865 870 875 880

Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly
885 890 895

Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly
900 905 910

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Glu Ser
915 920 925

Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser
930 935 940

Glu Thr Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly
945 950 955 960

Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Thr Glu
965 970 975

Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly
980 985 990

Ser Ala Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly
995 1000 1005

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr
1010 1015 1020

Glu Pro Ser Glu Gly Ser Ala Pro Ala Ser Ser Glu Ile Thr Arg
1025 1030 1035

Thr Thr Leu Gln Ser Asp Gln Glu Glu Ile Asp Tyr Asp Asp Thr
1040 1045 1050

Ile Ser Val Glu Met Lys Lys Glu Asp Phe Asp Ile Tyr Asp Glu

1055 1060 1065

Asp Glu Asn Gln Ser Pro Arg Ser Phe Gln Lys Lys Thr Arg His
1070 1075 1080

Tyr Phe Ile Ala Ala Val Glu Arg Leu Trp Asp Tyr Gly Met Ser
1085 1090 1095

Ser Ser Pro His Val Leu Arg Asn Arg Ala Gln Ser Gly Ser Val
1100 1105 1110

Pro Gln Phe Lys Lys Val Val Phe Gln Glu Phe Thr Asp Gly Ser
1115 1120 1125

Phe Thr Gln Pro Leu Tyr Arg Gly Glu Leu Asn Glu His Leu Gly
1130 1135 1140

Leu Leu Gly Pro Tyr Ile Arg Ala Glu Val Glu Asp Asn Ile Met
1145 1150 1155

Val Thr Phe Arg Asn Gln Ala Ser Arg Pro Tyr Ser Phe Tyr Ser
1160 1165 1170

Ser Leu Ile Ser Tyr Glu Glu Asp Gln Arg Gln Gly Ala Glu Pro
1175 1180 1185

Arg Lys Asn Phe Val Lys Pro Asn Glu Thr Lys Thr Tyr Phe Trp
1190 1195 1200

Lys Val Gln His His Met Ala Pro Thr Lys Asp Glu Phe Asp Cys
1205 1210 1215

Lys Ala Trp Ala Tyr Phe Ser Asp Val Asp Leu Glu Lys Asp Val
1220 1225 1230

His Ser Gly Leu Ile Gly Pro Leu Leu Val Cys His Thr Asn Thr

1235 1240 1245

Leu Asn Pro Ala His Gly Arg Gln Val Thr Val Gln Glu Phe Ala
1250 1255 1260

Leu Phe Phe Thr Ile Phe Asp Glu Thr Lys Ser Trp Tyr Phe Thr
1265 1270 1275

Glu Asn Met Glu Arg Asn Cys Arg Ala Pro Cys Asn Ile Gln Met
1280 1285 1290

Glu Asp Pro Thr Phe Lys Glu Asn Tyr Arg Phe His Ala Ile Asn
1295 1300 1305

Gly Tyr Ile Met Asp Thr Leu Pro Gly Leu Val Met Ala Gln Asp
1310 1315 1320

Gln Arg Ile Arg Trp Tyr Leu Leu Ser Met Gly Ser Asn Glu Asn
1325 1330 1335

Ile His Ser Ile His Phe Ser Gly His Val Phe Thr Val Arg Lys
1340 1345 1350

Lys Glu Glu Tyr Lys Met Ala Leu Tyr Asn Leu Tyr Pro Gly Val
1355 1360 1365

Phe Glu Thr Val Glu Met Leu Pro Ser Lys Ala Gly Ile Trp Arg
1370 1375 1380

Val Glu Cys Leu Ile Gly Glu His Leu His Ala Gly Met Ser Thr
1385 1390 1395

Leu Phe Leu Val Tyr Ser Asn Lys Cys Gln Thr Pro Leu Gly Met
1400 1405 1410

Ala Ser Gly His Ile Arg Asp Phe Gln Ile Thr Ala Ser Gly Gln

1415 1420 1425

Tyr Gly Gln Trp Ala Pro Lys Leu Ala Arg Leu His Tyr Ser Gly
1430 1435 1440

Ser Ile Asn Ala Trp Ser Thr Lys Glu Pro Phe Ser Trp Ile Lys
1445 1450 1455

Val Asp Leu Leu Ala Pro Met Ile Ile His Gly Ile Lys Thr Gln
1460 1465 1470

Gly Ala Arg Gln Lys Phe Ser Ser Leu Tyr Ile Ser Gln Phe Ile
1475 1480 1485

Ile Met Tyr Ser Leu Asp Gly Lys Lys Trp Gln Thr Tyr Arg Gly
1490 1495 1500

Asn Ser Thr Gly Thr Leu Met Val Phe Phe Gly Asn Val Asp Ser
1505 1510 1515

Ser Gly Ile Lys His Asn Ile Phe Asn Pro Pro Ile Ile Ala Arg
1520 1525 1530

Tyr Ile Arg Leu His Pro Thr His Tyr Ser Ile Arg Ser Thr Leu
1535 1540 1545

Arg Met Glu Leu Met Gly Cys Asp Leu Asn Ser Cys Ser Met Pro
1550 1555 1560

Leu Gly Met Glu Ser Lys Ala Ile Ser Asp Ala Gln Ile Thr Ala
1565 1570 1575

Ser Ser Tyr Phe Thr Asn Met Phe Ala Thr Trp Ser Pro Ser Lys
1580 1585 1590

Ala Arg Leu His Leu Gln Gly Arg Ser Asn Ala Trp Arg Pro Gln

1595 1600 1605

Val Asn Asn Pro Lys Glu Trp Leu Gln Val Asp Phe Gln Lys Thr
1610 1615 1620

Met Lys Val Thr Gly Val Thr Thr Gln Gly Val Lys Ser Leu Leu
1625 1630 1635

Thr Ser Met Tyr Val Lys Glu Phe Leu Ile Ser Ser Ser Gln Asp
1640 1645 1650

Gly His Gln Trp Thr Leu Phe Phe Gln Asn Gly Lys Val Lys Val
1655 1660 1665

Phe Gln Gly Asn Gln Asp Ser Phe Thr Pro Val Val Asn Ser Leu
1670 1675 1680

Asp Pro Pro Leu Leu Thr Arg Tyr Leu Arg Ile His Pro Gln Ser
1685 1690 1695

Trp Val His Gln Ile Ala Leu Arg Met Glu Val Leu Gly Cys Glu
1700 1705 1710

Ala Gln Asp Leu Tyr
1715

<210> 110
<211> 6526
<212> DNA
<213> 人工序列

<220>
<223> NB15.1v32建構體（AAV-FVIIIco6XTEN-ssDNA）

<400> 110

ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgaccttt 60

ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggccaa ctccatcact 120

aggggttcct gcggcaattc agtcgataac tataacggtc ctaaggtagc gatttaaata 180

cgcgctctct taaggtagcc ccgggacgcg tcaattgaga tctggatccg gtaccgaatt 240

cgcggccgcc tcgacgacta gcgtttaatt aaacgcgtgt ctgtctgcac atttcgtaga 300

gcgagtgttc cgatactcta atctccctag gcaaggttca tatttgtgta ggttacttat 360

tctccttttg ttgactaagt caataatcag aatcagcagg tttggagtca gcttggcagg 420

gatcagcagc ctgggttgga aggagggggt ataaaagccc cttcaccagg agaagccgtc 480

acacagatcc acaagctcct gaggtaagtg ccgtgtgtgg ttcccgcggg cctggcctct 540

ttacgggtta tggcccttgc gtgccttgaa ttactgacac tgacatccac tttttctttt 600

tctccacagc tagcgccacc atgcagattg agctgtccac ttgtttcttc ctgtgcctcc 660

tgcgcttctg tttctccgcc actcgccggt actaccttgg agccgtggag ctttcatggg 720

actacatgca gagcgacctg ggcgaactcc ccgtggatgc cagattcccc ccccgcgtgc 780

caaagtcctt ccccttaac acctccgtgg tgtacaagaa aaccctcttt gtcgagttca 840

ctgaccacct gttcaacatc gccaagccgc gcccaccttg gatgggcctc ctgggaccga 900

ccattcaagc tgaagtgtac gacaccgtgg tgatcaccct gaagaacatg gcgtcccacc 960

ccgtgtccct gcatgcggtc ggagtgtcct actggaaggc ctccgaagga gctgagtacg 1020

acgaccagac tagccagcgg gaaaaggagg acgataaagt gttcccgggc ggctcgcata 1080

cttacgtgtg gcaagtcctg aaggaaaacg gacctatggc atccgatcct ctgtgcctga 1140

cttactccta cctttcccat gtggacctcg tgaaggacct gaacagcggg ctgattggtg 1200

cacttctcgt gtgccgcgaa ggttcgctcg ctaaggaaaa gacccagacc ctccataagt 1260

tcatcctttt gttcgctgtg ttcgatgaag gaaagtcatg gcattccgaa actaagaact 1320

cgctgatgca ggaccgggat gccgcctcag cccgcgcctg gcctaaaatg catacagtca 1380

acggatacgt gaatcggtca ctgcccgggc tcatcggttg tcacagaaag tccgtgtact 1440

ggcacgtcat cggcatgggc actacgcctg aagtgcactc catcttcctg gaagggcaca 1500

ccttcctcgt gcgcaaccac cgccaggcct ctctggaaat ctccccgatt acctttctga 1560

ccgcccagac tctgctcatg gacctggggc agttccttct cttctgccac atctccagcc 1620

atcagcacga cggaatggag gcctacgtga aggtggactc atgcccggaa gaacctcagt 1680

tgcggatgaa gaacaacgag gaggccgagg actatgacga cgatttgact gactccgaga 1740

tggacgtcgt gcggttcgat gacgacaaca gccccagctt catccagatt cgcagcgtgg 1800

ccaagaagca ccccaaaacc tgggtgcact acatcgcggc cgaggaagaa gattgggact 1860

acgccccgtt ggtgctggca cccgatgacc ggtcgtacaa gtcccagtat ctgaacaatg 1920

gtccgcagcg gattggcaga aagtacaaga aagtgcggtt catggcgtac actgacgaaa 1980

cgtttaagac ccgggaggcc attcaacatg agagcggcat tctgggacca ctgctgtacg 2040

gagaggtcgg cgataccctg ctcatcatct tcaaaaacca ggcctcccgg ccttacaaca 2100

tctaccctca cggaatcacc gacgtgcggc cactctactc gcggcgcctg ccgaagggcg 2160

tcaagcacct gaaagacttc cctatcctgc cgggcgaaat cttcaagtat aagtggaccg 2220
tcaccgtgga ggacgggccc accaagagcg atcctaggtg tctgactcgg tactactcca 2280
gcttcgtgaa catggaacgg gacctggcat cgggactcat tggaccgctg ctgatctgct 2340
acaaagagtc ggtggatcaa cgcggcaacc agatcatgtc cgacaagcgc aacgtgatcc 2400
tgttctccgt gtttgatgaa aacagatcct ggtacctcac tgaaaacatc cagaggttcc 2460
tcccaaaccc cgcaggagtg caactggagg accctgagtt tcaggcctcg aatatcatgc 2520
actcgattaa cggttacgtg ttcgactcgc tgcaactgag cgtgtgcctc catgaagtcg 2580
cttactggta cattctgtcc atcggcgccc agactgactt cctgagcgtg ttcttttccg 2640
gttacacctt taagcacaag atggtgtacg aagataccct gaccctgttc cctttctccg 2700
gcgaaacggt gttcatgtcg atggagaacc cgggtctgtg gattctggga tgccacaaca 2760
gcgactttcg gaaccgcgga atgactgccc tgctgaaggt gtcctcatgc gacaagaaca 2820
ccggagacta ctacgaggac tcctacgagg atatctcagc ctacctcctg tccaagaaca 2880
acgcgatcga gccgcgcagc ttcagccaga acggcgcgcc aacatcagag agcgccaccc 2940
ctgaaagtgg tcccgggagc gagccagcca catctgggtc ggaaacgcca ggcacaagtg 3000
agtctgcaac tcccgagtcc ggacctggct ccgagcctgc cactagcggc tccgagactc 3060
cgggaacttc cgagagcgct acaccagaaa gcggacccgg aaccagtacc gaacctagcg 3120
agggctctgc tccgggcagc ccagccggct ctcctacatc cacggaggag ggcacttccg 3180
aatccgccac cccggagtca gggccaggat ctgaacccgc tacctcaggc agtgagacgc 3240
caggaacgag cgagtccgct acaccggaga gtgggccagg gagccctgct ggatctccta 3300
cgtccactga ggaagggtca ccagcgggct cgcccaccag cactgaagaa ggtgcctcga 3360
gcccgcctgt gctgaagagg caccagcgag aaattacccg gaccaccctc caatcggatc 3420
aggaggaaat cgactacgac gacaccatct cggtggaaat gaagaaggaa gatttcgata 3480
tctacgacga ggacgaaaat cagtcccctc gctcattcca aaagaaaact agacactact 3540
ttatcgccgc ggtggaaaga ctgtgggact atggaatgtc atccagccct cacgtccttc 3600
ggaaccgggc ccagagcgga tcggtgcctc agttcaagaa agtggtgttc caggagttca 3660
ccgacggcag cttcacccag ccgctgtacc ggggagaact gaacgaacac ctgggcctgc 3720
tcggtcccta catccgcgcg gaagtggagg ataacatcat ggtgaccttc cgtaaccaag 3780
catccagacc ttactccttc tattcctccc tgatctcata cgaggaggac cagcgccaag 3840
gcgccgagcc ccgcaagaac ttcgtcaagc ccaacgagac taagacctac ttctggaagg 3900
tccaacacca tatggccccg accaaggatg agtttgactg caaggcctgg gcctacttct 3960
ccgacgtgga ccttgagaag gatgtccatt ccggcctgat cgggccgctg ctcgtgtgtc 4020
acaccaacac cctgaaccca gcgcatggac gccaggtcac cgtccaggag tttgctctgt 4080
tcttcaccat ttttgacgaa actaagtcct ggtacttcac cgagaatatg gagcgaaact 4140
gtagagcgcc ctgcaatatc cagatggaag atccgacttt caaggagaac tatagattcc 4200
acgccatcaa cgggtacatc atggatactc tgccggggct ggtcatggcc caggatcaga 4260
ggattcggtg gtacttgctg tcaatgggat cgaacgaaaa cattcactcc attcacttct 4320

ccggtcacgt gttcactgtg cgcaagaagg aggagtacaa gatggcgctg tacaatctgt 4380
accccggggt gttcgaaact gtggagatgc tgccgtccaa ggccggcatc tggagagtgg 4440
agtgcctgat cggagagcac ctccacgcgg ggatgtccac cctcttcctg gtgtactcga 4500
ataagtgcca gaccccgctg ggcatggcct cgggccacat cagagacttc cagatcacag 4560
caagcggaca atacggccaa tgggcgccga agctggcccg cttgcactac tccggatcga 4620
tcaacgcatg gtccaccaag gaaccgttct cgtggattaa ggtggacctc ctggccccta 4680
tgattatcca cggaattaag acccagggcg ccaggcagaa gttctcctcc ctgtacatct 4740
cgcaattcat catcatgtac agcctggacg ggaagaagtg gcagacttac aggggaaact 4800
ccaccggcac cctgatggtc tttttcggca acgtggattc ctccggcatt aagcacaaca 4860
tcttcaaccc accgatcata gccagatata ttaggctcca ccccactcac tactcaatcc 4920
gctcaactct tcggatggaa ctcatggggt gcgacctgaa ctcctgctcc atgccgttgg 4980
ggatggaatc aaaggctatt agcgacgccc agatcaccgc gagctcctac ttcactaaca 5040
tgttcgccac ctggagcccc tccaaggcca ggctgcactt gcagggacgg tcaaatgcct 5100
ggcggccgca agtgaacaat ccgaaggaat ggcttcaagt ggatttccaa aagaccatga 5160
aagtgaccgg agtcaccacc cagggagtga agtcccttct gacctcgatg tatgtgaagg 5220
agttcctgat tagcagcagc caggacgggc accagtggac cctgttcttc caaaacggaa 5280
aggtcaaggt gttccagggg aaccaggact cgttcacacc cgtggtgaac tccctggacc 5340
ccccactgct gacgcggtac ttgaggattc atcctcagtc ctgggtccat cagattgcat 5400
tgcgaatgga agtcctgggc tgcgaggccc aggacctgta ctgaatcagc ctgagctcgc 5460
tgatcataat caacctctgg attacaaaat ttgtgaaaga ttgactggta ttcttaacta 5520
tgttgctcct tttacgctat gtggatacgc tgctttaatg cctttgtatc atgctattgc 5580
ttcccgtatg gctttcattt tctcctcctt gtataaatcc tggttgctgt ctctttatga 5640
ggagttgtgg cccgttgtca ggcaacgtgg cgtggtgtgc actgtgtttg ctgacgcaac 5700
ccccactggt tggggcattg ccaccacctg tcagctcctt tccgggactt tcgctttccc 5760
cctccctatt gccacggcgg aactcatcgc cgcctgcctt gcccgctgct ggacaggggc 5820
tcggctgttg ggcactgaca attccgtggt gttgtcgggg aaatcatcgt cctttccttg 5880
gctgctcgcc tgtgttgcca cctggattct gcgcgggacg tccttctgct acgtcccttc 5940
ggccctcaat ccagcggacc ttccttcccg cggcctgctg ccggctctgc ggcctcttcc 6000
gcgtcttcgc cttcgccctc agacgagtcg gatctccctt tgggccgcct ccccgctgat 6060
cagcctcgac tgtgccttct agttgccagc catctgttgt ttgcccctcc cccgtgcctt 6120
ccttgaccct ggaaggtgcc actcccactg tcctttccta ataaaatgag gaaattgcat 6180
cgcattgtct gagtaggtgt cattctattc tgggggggtgg ggtggggcag gacagcaagg 6240
gggaggattg ggaagacaat agcaggcatg ctggggatgc ggtgggctct atggcttctg 6300
aggcggaaag aacgggctcg agaagcttct agatatcctc tcttaaggta gcatcgagat 6360
ttaaattagg gataacaggg taatggcgcg ggccgcagga acccctagtg atggagttgg 6420
ccactccctc tctgcgcgct cgctcgctca ctgaggccgg gcgaccaaag gtcgcccgac 6480

gcccgggctt tgcccgggcg gcctcagtga gcgagcgagc gcgcag 6526

<210> 111
<211> 130
<212> DNA
<213> 人工序列

<220>
<223> 5'ITR（5'端AAV2反向末端重複體）

<400> 111
ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgaccttt 60
ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggccaa ctccatcact 120
aggggttcct 130

<210> 112
<211> 142
<212> DNA
<213> 人工序列

<220>
<223> 質體骨架序列（PBS）-1

<400> 112
gcggcaattc agtcgataac tataacggtc ctaaggtagc gatttaaata cgcgctctct 60
taaggtagcc ccgggacgcg tcaattgaga tctggatccg gtaccgaatt cgcggccgcc 120
tcgacgacta gcgtttaatt aa 142

<210> 113
<211> 229
<212> DNA
<213> 人工序列

<220>

<223> TTPp（肝臟特異性啟動子）

<400> 113

acgcgtgtct gtctgcacat ttcgtagagc gagtgttccg atactctaat ctccctaggc 60

aaggttcata tttgtgtagg ttacttattc tccttttgtt gactaagtca ataatcagaa 120

tcagcaggtt tggagtcagc ttggcaggga tcagcagcct gggttggaag gaggggtat 180

aaaagcccct tcaccaggag aagccgtcac acagatccac aagctcctg 229

<210> 114

<211> 2

<212> DNA

<213> 人工序列

<220>

<223> 質體骨架序列（PBS）-2

<400> 114

ag 2

<210> 115

<211> 106

<212> DNA

<213> 人工序列

<220>

<223> 合成內含子

<400> 115

gtaagtgccg tgtgtggttc ccgcgggcct ggcctcttta cgggttatgg cccttgcgtg 60

ccttgaatta ctgacactga catccacttt ttctttttct ccacag 106

<210> 116

<211> 11

<212> DNA

<213> 人工序列

<220>

<223> 質體骨架序列（PBS）-3

<400> 116

ctagcgccac c 11

<210> 117

<211> 4824

<212> DNA

<213> 人工序列

<220>

<223> FVIIIco6XTEN（含有XTEN144的密碼子優化的FVIII形式6的開放閱讀框）

<400> 117

atgcagattg agctgtccac ttgtttcttc ctgtgcctcc tgcgcttctg tttctccgcc 60

actcgccggt actaccttgg agccgtggag ctttcatggg actacatgca gagcgacctg 120

ggcgaactcc ccgtggatgc cagattcccc ccccgcgtgc caaagtcctt ccccttaac 180

acctccgtgg tgtacaagaa aaccctcttt gtcgagttca ctgaccacct gttcaacatc 240

gccaagccgc gcccaccttg gatgggcctc ctgggaccga ccattcaagc tgaagtgtac 300

gacaccgtgg tgatcaccct gaagaacatg gcgtcccacc ccgtgtccct gcatgcggtc 360

ggagtgtcct actggaaggc ctccgaagga gctgagtacg acgaccagac tagccagcgg 420

gaaaaggagg acgataaagt gttcccgggc ggctcgcata cttacgtgtg gcaagtcctg 480

aaggaaaacg gacctatggc atccgatcct ctgtgcctga cttactccta cctttcccat 540

gtggacctcg tgaaggacct gaacagcggg ctgattggtg cacttctcgt gtgccgcgaa 600

ggttcgctcg ctaaggaaaa gacccagacc ctccataagt tcatcctttt gttcgctgtg 660

ttcgatgaag gaaagtcatg gcattccgaa actaagaact cgctgatgca ggaccgggat 720

gccgcctcag cccgcgcctg gcctaaaatg catacagtca acggatacgt gaatcggtca 780

ctgcccgggc tcatcggttg tcacagaaag tccgtgtact ggcacgtcat cggcatgggc 840

actacgcctg aagtgcactc catcttcctg gaagggcaca ccttcctcgt gcgcaaccac 900

cgccaggcct ctctggaaat ctccccgatt acctttctga ccgcccagac tctgctcatg 960

gacctggggc agttccttct cttctgccac atctccagcc atcagcacga cggaatggag 1020
gcctacgtga aggtggactc atgcccggaa gaacctcagt tgcggatgaa gaacaacgag 1080
gaggccgagg actatgacga cgatttgact gactccgaga tggacgtcgt gcggttcgat 1140
gacgacaaca gccccagctt catccagatt cgcagcgtgg ccaagaagca ccccaaaacc 1200
tgggtgcact acatcgcggc cgaggaagaa gattgggact acgccccgtt ggtgctggca 1260
cccgatgacc ggtcgtacaa gtcccagtat ctgaacaatg gtccgcagcg gattggcaga 1320
aagtacaaga aagtgcggtt catggcgtac actgacgaaa cgtttaagac ccgggaggcc 1380
attcaacatg agagcggcat tctgggacca ctgctgtacg gagaggtcgg cgataccctg 1440
ctcatcatct tcaaaaacca ggcctcccgg ccttacaaca tctaccctca cggaatcacc 1500
gacgtgcggc cactctactc gcggcgcctg ccgaagggcg tcaagcacct gaaagacttc 1560
cctatcctgc cgggcgaaat cttcaagtat aagtggaccg tcaccgtgga ggacgggccc 1620
accaagagcg atcctaggtg tctgactcgg tactactcca gcttcgtgaa catggaacgg 1680
gacctggcat cgggactcat tggaccgctg ctgatctgct acaaagagtc ggtggatcaa 1740
cgcggcaacc agatcatgtc cgacaagcgc aacgtgatcc tgttctccgt gtttgatgaa 1800
aacagatcct ggtacctcac tgaaaacatc cagaggttcc tcccaaaccc cgcaggagtg 1860
caactggagg accctgagtt tcaggcctcg aatatcatgc actcgattaa cggttacgtg 1920
ttcgactcgc tgcaactgag cgtgtgcctc catgaagtcg cttactggta cattctgtcc 1980
atcggcgccc agactgactt cctgagcgtg ttcttttccg gttacacctt taagcacaag 2040
atggtgtacg aagataccct gaccctgttc cctttctccg gcgaaacggt gttcatgtcg 2100
atggagaacc cgggtctgtg gattctggga tgccacaaca gcgactttcg gaaccgcgga 2160
atgactgccc tgctgaaggt gtcctcatgc gacaagaaca ccggagacta ctacgaggac 2220
tcctacgagg atatctcagc ctacctcctg tccaagaaca acgcgatcga gccgcgcagc 2280
ttcagccaga acggcgcgcc aacatcagag agcgccaccc ctgaaagtgg tcccgggagc 2340
gagccagcca catctgggtc ggaaacgcca ggcacaagtg agtctgcaac tcccgagtcc 2400
ggacctggct ccgagcctgc cactagcggc tccgagactc cgggaacttc cgagagcgct 2460
acaccagaaa gcggacccgg aaccagtacc gaacctagcg agggctctgc tccgggcagc 2520
ccagccggct ctcctacatc cacggaggag ggcacttccg aatccgccac cccggagtca 2580
gggccaggat ctgaacccgc tacctcaggc agtgagacgc caggaacgag cgagtccgct 2640
acaccggaga gtgggccagg gagccctgct ggatctccta cgtccactga ggaagggtca 2700
ccagcgggct cgcccaccag cactgaagaa ggtgcctcga gcccgcctgt gctgaagagg 2760
caccagcgag aaattacccg gaccaccctc caatcggatc aggaggaaat cgactacgac 2820
gacaccatct cggtggaaat gaagaaggaa gatttcgata tctacgacga ggacgaaaat 2880
cagtcccctc gctcattcca aaagaaaact agacactact ttatcgccgc ggtggaaaga 2940
ctgtgggact atggaatgtc atccagccct cacgtccttc ggaaccgggc ccagagcgga 3000
tcggtgcctc agttcaagaa agtggtgttc caggagttca ccgacggcag cttcacccag 3060
ccgctgtacc ggggagaact gaacgaacac ctgggcctgc tcggtcccta catccgcgcg 3120

gaagtggagg ataacatcat ggtgaccttc cgtaaccaag catccagacc ttactccttc 3180

tattcctccc tgatctcata cgaggaggac cagcgccaag gcgccgagcc ccgcaagaac 3240

ttcgtcaagc ccaacgagac taagacctac ttctggaagg tccaacacca tatggccccg 3300

accaaggatg agtttgactg caaggcctgg gcctacttct ccgacgtgga ccttgagaag 3360

gatgtccatt ccggcctgat cgggccgctg ctcgtgtgtc acaccaacac cctgaaccca 3420

gcgcatggac gccaggtcac cgtccaggag tttgctctgt tcttcaccat ttttgacgaa 3480

actaagtcct ggtacttcac cgagaatatg gagcgaaact gtagagcgcc ctgcaatatc 3540

cagatggaag atccgacttt caaggagaac tatagattcc acgccatcaa cgggtacatc 3600

atggatactc tgccggggct ggtcatggcc caggatcaga ggattcggtg gtacttgctg 3660

tcaatgggat cgaacgaaaa cattcactcc attcacttct ccggtcacgt gttcactgtg 3720

cgcaagaagg aggagtacaa gatggcgctg tacaatctgt accccggggt gttcgaaact 3780

gtggagatgc tgccgtccaa ggccggcatc tggagagtgg agtgcctgat cggagagcac 3840

ctccacgcgg ggatgtccac cctcttcctg gtgtactcga ataagtgcca gaccccgctg 3900

ggcatggcct cgggccacat cagagacttc cagatcacag caagcggaca atacggccaa 3960

tgggcgccga agctggcccg cttgcactac tccggatcga tcaacgcatg gtccaccaag 4020

gaaccgttct cgtggattaa ggtggacctc ctggccccta tgattatcca cggaattaag 4080

acccagggcg ccaggcagaa gttctcctcc ctgtacatct cgcaattcat catcatgtac 4140

agcctggacg ggaagaagtg gcagacttac aggggaaact ccaccggcac cctgatggtc 4200

tttttcggca acgtggattc ctccggcatt aagcacaaca tcttcaaccc accgatcata 4260

gccagatata ttaggctcca ccccactcac tactcaatcc gctcaactct tcggatggaa 4320

ctcatggggt gcgacctgaa ctcctgctcc atgccgttgg ggatggaatc aaaggctatt 4380

agcgacgccc agatcaccgc gagctcctac ttcactaaca tgttcgccac ctggagcccc 4440

tccaaggcca ggctgcactt gcagggacgg tcaaatgcct ggcggccgca agtgaacaat 4500

ccgaaggaat ggcttcaagt ggatttccaa aagaccatga aagtgaccgg agtcaccacc 4560

cagggagtga agtcccttct gacctcgatg tatgtgaagg agttcctgat tagcagcagc 4620

caggacgggc accagtggac cctgttcttc caaaacggaa aggtcaaggt gttccagggg 4680

aaccaggact cgttcacacc cgtggtgaac tccctggacc ccccactgct gacgcggtac 4740

ttgaggattc atcctcagtc ctgggtccat cagattgcat tgcgaatgga agtcctgggc 4800

tgcgaggccc aggacctgta ctga 4824

<210> 118

<211> 450

<212> DNA

<213> 人工序列

<220>

<223> XTEN144

<400> 118

ggcgcgccaa catcagagag cgccaccccct gaaagtggtc ccgggagcga gccagccaca 60

tctgggtcgg aaacgccagg cacaagtgag tctgcaactc ccgagtccgg acctggctcc 120

gagcctgcca ctagcggctc cgagactccg ggaacttccg agagcgctac accagaaagc 180

ggacccggaa ccagtaccga acctagcgag ggctctgctc cgggcagccc agccggctct 240

cctacatcca cggaggaggg cacttccgaa tccgccaccc cggagtcagg gccaggatct 300

gaacccgcta cctcaggcag tgagacgcca ggaacgagcg agtccgctac accggagagt 360

gggccaggga gccctgctgg atctcctacg tccactgagg aagggtcacc agcgggctcg 420

cccaccagca ctgaagaagg tgcctcgagc 450

<210> 119

<211> 19

<212> DNA

<213> 人工序列

<220>

<223> 質體骨架序列（PBS）-4

<400> 119

atcagcctga gctcgctga 19

<210> 120

<211> 595

<212> DNA

<213> 人工序列

<220>

<223> WPRE（突變的土撥鼠肝炎病毒轉錄後調控元件）

<400> 120

tcataatcaa cctctggatt acaaaatttg tgaaagattg actggtattc ttaactatgt 60

tgctcctttt acgctatgtg gatacgctgc tttaatgcct ttgtatcatg ctattgcttc 120
ccgtatggct ttcattttct cctccttgta taaatcctgg ttgctgtctc tttatgagga 180
gttgtggccc gttgtcaggc aacgtggcgt ggtgtgcact gtgtttgctg acgcaacccc 240
cactggttgg ggcattgcca ccacctgtca gctcctttcc gggactttcg ctttcccct 300
ccctattgcc acggcggaac tcatcgccgc ctgccttgcc cgctgctgga caggggctcg 360
gctgttgggc actgacaatt ccgtggtgtt gtcggggaaa tcatcgtcct ttccttggct 420
gctcgcctgt gttgccacct ggattctgcg cgggacgtcc ttctgctacg tcccttcggc 480
cctcaatcca gcggaccttc cttcccgcgg cctgctgccg gctctgcggc ctcttccgcg 540
tcttcgcctt cgccctcaga cgagtcggat ctccctttgg gccgcctccc cgctg 595

<210> 121
<211> 8
<212> DNA
<213> 人工序列

<220>
<223> 質體骨架序列（PBS）-5

<400> 121
atcagcct 8

<210> 122
<211> 211
<212> DNA
<213> 人工序列

<220>
<223> bGHpA（牛生長激素多聚腺苷酸化信號）

<400> 122
cgactgtgcc ttctagttgc cagccatctg ttgtttgccc ctcccccgtg ccttccttga 60
ccctggaagg tgccactccc actgtccttt cctaataaaa tgaggaaatt gcatcgcatt 120
gtctgagtag gtgtcattct attctggggg gtggggtggg gcaggacagc aaggggagg 180
attgggaaga caatagcagg catgctgggg a 211

<210> 123
<211> 119
<212> DNA
<213> 人工序列

<220>
<223> 質體骨架序列（PBS）-6

<400> 123
tgcggtgggc tctatggctt ctgaggcgga aagaacgggc tcgagaagct tctagatatc 60
ctctcttaag gtagcatcga gatttaaatt agggataaca gggtaatggc gcgggccgc 119

<210> 124
<211> 130
<212> DNA
<213> 人工序列

<220>
<223> 3'ITR（3'端AAV2反向末端重複體）

<400> 124
aggaacccct agtgatggag ttggccactc cctctctgcg cgctcgctcg ctcactgagg 60
ccgggcgacc aaaggtcgcc cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc 120
gagcgcgcag 130

<210> 125
<211> 415
<212> PRT
<213> 人工序列

<220>
<223> 成熟FIX多肽

<400> 125

Tyr Asn Ser Gly Lys Leu Glu Glu Phe Val Gln Gly Asn Leu Glu Arg
1 5 10 15

Glu Cys Met Glu Glu Lys Cys Ser Phe Glu Glu Ala Arg Glu Val Phe
20 25 30

Glu Asn Thr Glu Arg Thr Thr Glu Phe Trp Lys Gln Tyr Val Asp Gly
35 40 45

Asp Gln Cys Glu Ser Asn Pro Cys Leu Asn Gly Gly Ser Cys Lys Asp
50 55 60

Asp Ile Asn Ser Tyr Glu Cys Trp Cys Pro Phe Gly Phe Glu Gly Lys
65 70 75 80

Asn Cys Glu Leu Asp Val Thr Cys Asn Ile Lys Asn Gly Arg Cys Glu
85 90 95

Gln Phe Cys Lys Asn Ser Ala Asp Asn Lys Val Val Cys Ser Cys Thr
100 105 110

Glu Gly Tyr Arg Leu Ala Glu Asn Gln Lys Ser Cys Glu Pro Ala Val
115 120 125

Pro Phe Pro Cys Gly Arg Val Ser Val Ser Gln Thr Ser Lys Leu Thr
130 135 140

Arg Ala Glu Thr Val Phe Pro Asp Val Asp Tyr Val Asn Ser Thr Glu
145 150 155 160

Ala Glu Thr Ile Leu Asp Asn Ile Thr Gln Ser Thr Gln Ser Phe Asn
165 170 175

Asp Phe Thr Arg Val Val Gly Gly Glu Asp Ala Lys Pro Gly Gln Phe

180 185 190

Pro Trp Gln Val Val Leu Asn Gly Lys Val Asp Ala Phe Cys Gly Gly
195 200 205

Ser Ile Val Asn Glu Lys Trp Ile Val Thr Ala Ala His Cys Val Glu
210 215 220

Thr Gly Val Lys Ile Thr Val Val Ala Gly Glu His Asn Ile Glu Glu
225 230 235 240

Thr Glu His Thr Glu Gln Lys Arg Asn Val Ile Arg Ile Ile Pro His
245 250 255

His Asn Tyr Asn Ala Ala Ile Asn Lys Tyr Asn His Asp Ile Ala Leu
260 265 270

Leu Glu Leu Asp Glu Pro Leu Val Leu Asn Ser Tyr Val Thr Pro Ile
275 280 285

Cys Ile Ala Asp Lys Glu Tyr Thr Asn Ile Phe Leu Lys Phe Gly Ser
290 295 300

Gly Tyr Val Ser Gly Trp Gly Arg Val Phe His Lys Gly Arg Ser Ala
305 310 315 320

Leu Val Leu Gln Tyr Leu Arg Val Pro Leu Val Asp Arg Ala Thr Cys
325 330 335

Leu Arg Ser Thr Lys Phe Thr Ile Tyr Asn Asn Met Phe Cys Ala Gly
340 345 350

Phe His Glu Gly Gly Arg Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro
355 360 365

His Val Thr Glu Val Glu Gly Thr Ser Phe Leu Thr Gly Ile Ile Ser

370 375 380

Trp Gly Glu Glu Cys Ala Met Lys Gly Lys Tyr Gly Ile Tyr Thr Lys
385 390 395 400

Val Ser Arg Tyr Val Asn Trp Ile Lys Glu Lys Thr Lys Leu Thr
405 410 415

<210> 126
<211> 415
<212> PRT
<213> 人工序列

<220>
<223> 成熟Padua(R338L)FIX多肽

<400> 126
Tyr Asn Ser Gly Lys Leu Glu Glu Phe Val Gln Gly Asn Leu Glu Arg
1 5 10 15

Glu Cys Met Glu Glu Lys Cys Ser Phe Glu Glu Ala Arg Glu Val Phe
20 25 30

Glu Asn Thr Glu Arg Thr Thr Glu Phe Trp Lys Gln Tyr Val Asp Gly
35 40 45

Asp Gln Cys Glu Ser Asn Pro Cys Leu Asn Gly Gly Ser Cys Lys Asp
50 55 60

Asp Ile Asn Ser Tyr Glu Cys Trp Cys Pro Phe Gly Phe Glu Gly Lys
65 70 75 80

Asn Cys Glu Leu Asp Val Thr Cys Asn Ile Lys Asn Gly Arg Cys Glu
85 90 95

```
Gln Phe Cys Lys Asn Ser Ala Asp Asn Lys Val Val Cys Ser Cys Thr
            100                 105                 110

Glu Gly Tyr Arg Leu Ala Glu Asn Gln Lys Ser Cys Glu Pro Ala Val
        115                 120                 125

Pro Phe Pro Cys Gly Arg Val Ser Val Ser Gln Thr Ser Lys Leu Thr
    130                 135                 140

Arg Ala Glu Thr Val Phe Pro Asp Val Asp Tyr Val Asn Ser Thr Glu
145                 150                 155                 160

Ala Glu Thr Ile Leu Asp Asn Ile Thr Gln Ser Thr Gln Ser Phe Asn
                165                 170                 175

Asp Phe Thr Arg Val Val Gly Gly Glu Asp Ala Lys Pro Gly Gln Phe
            180                 185                 190

Pro Trp Gln Val Val Leu Asn Gly Lys Val Asp Ala Phe Cys Gly Gly
        195                 200                 205

Ser Ile Val Asn Glu Lys Trp Ile Val Thr Ala Ala His Cys Val Glu
    210                 215                 220

Thr Gly Val Lys Ile Thr Val Val Ala Gly Glu His Asn Ile Glu Glu
225                 230                 235                 240

Thr Glu His Thr Glu Gln Lys Arg Asn Val Ile Arg Ile Ile Pro His
                245                 250                 255

His Asn Tyr Asn Ala Ala Ile Asn Lys Tyr Asn His Asp Ile Ala Leu
            260                 265                 270

Leu Glu Leu Asp Glu Pro Leu Val Leu Asn Ser Tyr Val Thr Pro Ile
        275                 280                 285
```

Cys Ile Ala Asp Lys Glu Tyr Thr Asn Ile Phe Leu Lys Phe Gly Ser
290 295 300

Gly Tyr Val Ser Gly Trp Gly Arg Val Phe His Lys Gly Arg Ser Ala
305 310 315 320

Leu Val Leu Gln Tyr Leu Arg Val Pro Leu Val Asp Arg Ala Thr Cys
325 330 335

Leu Leu Ser Thr Lys Phe Thr Ile Tyr Asn Asn Met Phe Cys Ala Gly
340 345 350

Phe His Glu Gly Gly Arg Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro
355 360 365

His Val Thr Glu Val Glu Gly Thr Ser Phe Leu Thr Gly Ile Ile Ser
370 375 380

Trp Gly Glu Glu Cys Ala Met Lys Gly Lys Tyr Gly Ile Tyr Thr Lys
385 390 395 400

Val Ser Arg Tyr Val Asn Trp Ile Lys Glu Lys Thr Lys Leu Thr
405 410 415

<210> 127
<211> 46
<212> PRT
<213> 人工序列

<220>
<223> FIX信號多肽和前肽

<400> 127
Met Gln Arg Val Asn Met Ile Met Ala Glu Ser Pro Gly Leu Ile Thr
1 5 10 15

Ile Cys Leu Leu Gly Tyr Leu Leu Ser Ala Glu Cys Thr Val Phe Leu
20 25 30

Asp His Glu Asn Ala Asn Lys Ile Leu Asn Arg Pro Lys Arg
35 40 45

<210> 128
<211> 8442
<212> DNA
<213> 人工序列

<220>
<223> 人類VWF

<400> 128
atgattcctg ccagatttgc cggggtgctg cttgctctgg ccctcatttt gccagggacc 60
ctttgtgcag aaggaactcg cggcaggtca tccacggccc gatgcagcct tttcggaagt 120
gacttcgtca acacctttga tgggagcatg tacagctttg cgggatactg cagttacctc 180
ctggcagggg gctgccagaa acgctccttc tcgattattg gggacttcca gaatggcaag 240
agagtgagcc tctccgtgta tcttggggaa ttttttgaca tccatttgtt tgtcaatggt 300
accgtgacac agggggacca aagagtctcc atgccctatg cctccaaagg gctgtatcta 360
gaaactgagg ctgggtacta caagctgtcc ggtgaggcct atggctttgt ggccaggatc 420
gatggcagcg gcaactttca agtcctgctg tcagacagat acttcaacaa gacctgcggg 480
ctgtgtggca actttaacat ctttgctgaa gatgacttta tgacccaaga agggaccttg 540
acctcggacc cttatgactt tgccaactca tgggctctga gcagtggaga acagtggtgt 600
gaacgggcat ctcctcccag cagctcatgc aacatctcct ctggggaaat gcagaagggc 660
ctgtgggagc agtgccagct tctgaagagc acctcggtgt ttgcccgctg ccaccctctg 720
gtggaccccg agccttttgt ggccctgtgt gagaagactt tgtgtgagtg tgctgggggg 780
ctggagtgcg cctgccctgc cctcctggag tacgcccgga cctgtgccca ggagggaatg 840
gtgctgtacg gctggaccga ccacagcgcg tgcagcccag tgtgccctgc tggtatggag 900
tataggcagt gtgtgtcccc ttgcgccagg acctgccaga gcctgcacat caatgaaatg 960
tgtcaggagc gatgcgtgga tggctgcagc tgccctgagg gacagctcct ggatgaaggc 1020
ctctgcgtgg agagcaccga gtgtccctgc gtgcattccg gaaagcgcta ccctcccggc 1080
acctccctct ctcgagactg caacacctgc atttgccgaa acagccagtg gatctgcagc 1140

aatgaagaat gtccagggga gtgccttgtc actggtcaat cccacttcaa gagctttgac 1200

aacagatact tcaccttcag tgggatctgc cagtacctgc tggcccggga ttgccaggac 1260

cactccttct ccattgtcat tgagactgtc cagtgtgctg atgaccgcga cgctgtgtgc 1320

acccgctccg tcaccgtccg gctgcctggc ctgcacaaca gccttgtgaa actgaagcat 1380

ggggcaggag ttgccatgga tggccaggac atccagctcc ccctcctgaa aggtgacctc 1440

cgcatccagc atacagtgac ggcctccgtg cgcctcagct acggggagga cctgcagatg 1500

gactgggatg gccgcgggag gctgctggtg aagctgtccc ccgtctatgc cgggaagacc 1560

tgcggcctgt gtgggaatta caatggcaac cagggcgacg acttccttac cccctctggg 1620

ctggcrgagc cccgggtgga ggacttcggg aacgcctgga agctgcacgg ggactgccag 1680

gacctgcaga agcagcacag cgatccctgc gccctcaacc cgcgcatgac caggttctcc 1740

gaggaggcgt gcgcggtcct gacgtccccc acattcgagg cctgccatcg tgccgtcagc 1800

ccgctgccct acctgcggaa ctgccgctac gacgtgtgct cctgctcgga cggccgcgag 1860

tgcctgtgcg gcgccctggc cagctatgcc gcggcctgcg cggggagagg cgtgcgcgtc 1920

gcgtggcgcg agccaggccg ctgtgagctg aactgcccga aaggccaggt gtacctgcag 1980

tgcgggaccc cctgcaacct gacctgccgc tctctctctt acccggatga ggaatgcaat 2040

gaggcctgcc tggagggctg cttctgcccc ccagggctct acatggatga gaggggggac 2100

tgcgtgccca aggcccagtg cccctgttac tatgacggtg agatcttcca gccagaagac 2160

atcttctcag accatcacac catgtgctac tgtgaggatg gcttcatgca ctgtaccatg 2220

agtggagtcc ccggaagctt gctgcctgac gctgtcctca gcagtcccct gtctcatcgc 2280

agcaaaagga gcctatcctg tcggcccccc atggtcaagc tggtgtgtcc cgctgacaac 2340

ctgcgggctg aagggctcga gtgtaccaaa acgtgccaga actatgacct ggagtgcatg 2400

agcatgggct gtgtctctgg ctgcctctgc cccccgggca tggtccggca tgagaacaga 2460

tgtgtggccc tggaaaggtg tccctgcttc catcagggca aggagtatgc ccctggagaa 2520

acagtgaaga ttggctgcaa cacttgtgtc tgtcgggacc ggaagtggaa ctgcacagac 2580

catgtgtgtg atgccacgtg ctccacgatc ggcatggccc actacctcac cttcgacggg 2640

ctcaaatacc tgttccccgg ggagtgccag tacgttctgg tgcaggatta ctgcggcagt 2700

aaccctggga cctttcggat cctagtgggg aataagggat gcagccaccc ctcagtgaaa 2760

tgcaagaaac gggtcaccat cctggtggag ggaggagaga ttgagctgtt tgacggggag 2820

gtgaatgtga agaggcccat gaaggatgag actcactttg aggtggtgga gtctggccgg 2880

tacatcattc tgctgctggg caaagccctc tccgtggtct gggaccgcca cctgagcatc 2940

tccgtggtcc tgaagcagac ataccaggag aaagtgtgtg gcctgtgtgg gaattttgat 3000

ggcatccaga acaatgacct caccagcagc aacctccaag tggaggaaga ccctgtggac 3060

tttgggaact cctggaaagt gagctcgcag tgtgctgaca ccagaaaagt gcctctggac 3120

tcatcccctg ccacctgcca taacaacatc atgaagcaga cgatggtgga ttcctcctgt 3180

agaatcctta ccagtgacgt cttccaggac tgcaacaagc tggtggaccc cgagccatat 3240

ctggatgtct gcatttacga cacctgctcc tgtgagtcca ttggggactg cgcctgcttc 3300

tgcgacacca ttgctgccta tgcccacgtg tgtgcccagc atggcaaggt ggtgacctgg 3360

aggacggcca cattgtgccc ccagagctgc gaggagagga atctccggga gaacgggtat 3420

gagtgtgagt ggcgctataa cagctgtgca cctgcctgtc aagtcacgtg tcagcaccct 3480

gagccactgg cctgccctgt gcagtgtgtg gagggctgcc atgcccactg ccctccaggg 3540

aaaatcctgg atgagctttt gcagacctgc gttgaccctg aagactgtcc agtgtgtgag 3600

gtggctggcc ggcgttttgc ctcaggaaag aaagtcacct tgaatcccag tgaccctgag 3660

cactgccaga tttgccactg tgatgttgtc aacctcacct gtgaagcctg ccaggagccg 3720

ggaggcctgg tggtgcctcc cacagatgcc ccggtgagcc ccaccactct gtatgtggag 3780

gacatctcgg aaccgccgtt gcacgatttc tactgcagca ggctactgga cctggtcttc 3840

ctgctggatg gctcctccag gctgtccgag gctgagtttg aagtgctgaa ggcctttgtg 3900

gtggacatga tggagcggct gcgcatctcc cagaagtggg tccgcgtggc cgtggtggag 3960

taccacgacg gctcccacgc ctacatcggg ctcaaggacc ggaagcgacc gtcagagctg 4020

cggcgcattg ccagccaggt gaagtatgcg ggcagccagg tggcctccac cagcgaggtc 4080

ttgaaataca cactgttcca aatcttcagc aagatcgacc gccctgaagc ctcccgcatc 4140

gccctgctcc tgatggccag ccaggagccc caacggatgt cccggaactt tgtccgctac 4200

gtccagggcc tgaagaagaa gaaggtcatt gtgatcccgg tgggcattgg gccccatgcc 4260

aacctcaagc agatccgcct catcgagaag caggcccctg agaacaaggc cttcgtgctg 4320

agcagtgtgg atgagctgga gcagcaaagg gacgagatcg ttagctacct ctgtgacctt 4380

gcccctgaag ccctcctctc tactctgccc cccgacatgg cacaagtcac tgtgggcccg 4440

gggctcttgg gggtttcgac cctggggccc aagaggaact ccatggttct ggatgtggcg 4500

ttcgtcctgg aaggatcgga caaaattggt gaagccgact tcaacaggag caaggagttc 4560

atggaggagg tgattcagcg gatggatgtg ggccaggaca gcatccacgt cacggtgctg 4620

cagtactcct acatggtgac cgtggagtac cccttcagcg aggcacagtc caaaggggac 4680

atcctgcagc gggtgcgaga gatccgctac cagggcggca acaggaccaa cactgggctg 4740

gccctgcggt acctctctga ccacagcttc ttggtcagcc agggtgaccg ggagcaggcg 4800

cccaacctgg tctacatggt caccggaaat cctgcctctg atgagatcaa gaggctgcct 4860

ggagacatcc aggtggtgcc cattggagtg ggccctaatg ccaacgtgca ggagctggag 4920

aggattggct ggcccaatgc ccctatcctc atccaggact ttgagacgct cccccgagag 4980

gctcctgacc tggtgctgca gaggtgctgc tccggagagg ggctgcagat ccccaccctc 5040

tcccctgcac ctgactgcag ccagcccctg gacgtgatcc ttctcctgga tggctcctcc 5100

agtttcccag cttcttattt tgatgaaatg aagagtttcg ccaaggcttt catttcaaaa 5160

gccaatatag ggcctcgtct cactcaggtg tcagtgctgc agtatggaag catcaccacc 5220

attgacgtgc catggaacgt ggtcccggag aaagcccatt tgctgagcct tgtggacgtc 5280

atgcagcggg agggaggccc cagccaaatc ggggatgcct tgggctttgc tgtgcgatac 5340

ttgacttcag aaatgcatgg tgccaggccg ggagcctcaa aggcggtggt catcctggtc 5400

acggacgtct ctgtggattc agtggatgca gcagctgatg ccgccaggtc caacagagtg 5460

acagtgttcc ctattggaat tggagatcgc tacgatgcag cccagctacg gatcttggca 5520

ggcccagcag gcgactccaa cgtggtgaag ctccagcgaa tcgaagacct ccctaccatg 5580

gtcaccttgg gcaattcctt cctccacaaa ctgtgctctg gatttgttag gatttgcatg 5640

gatgaggatg ggaatgagaa gaggcccggg gacgtctgga ccttgccaga ccagtgccac 5700

accgtgactt gccagccaga tggccagacc ttgctgaaga gtcatcgggt caactgtgac 5760

cgggggctga ggccttcgtg ccctaacagc cagtcccctg ttaaagtgga agagacctgt 5820

ggctgccgct ggacctgccc ctgygtgtgc acaggcagct ccactcggca catcgtgacc 5880

tttgatgggc agaatttcaa gctgactggc agctgttctt atgtcctatt tcaaaacaag 5940

gagcaggacc tggaggtgat tctccataat ggtgcctgca gccctggagc aaggcagggc 6000

tgcatgaaat ccatcgaggt gaagcacagt gccctctccg tcgagstgca cagtgacatg 6060

gaggtgacgg tgaatgggag actggtctct gttccttacg tgggtgggaa catggaagtc 6120

aacgtttatg gtgccatcat gcatgaggtc agattcaatc accttggtca catcttcaca 6180

ttcactccac aaaacaatga gttccaactg cagctcagcc ccaagacttt tgcttcaaag 6240

acgtatggtc tgtgtgggat ctgtgatgag aacggagcca atgacttcat gctgagggat 6300

ggcacagtca ccacagactg gaaaacactt gttcaggaat ggactgtgca gcggccaggg 6360

cagacgtgcc agcccatcct ggaggagcag tgtcttgtcc ccgacagctc ccactgccag 6420

gtcctcctct taccactgtt tgctgaatgc cacaaggtcc tggctccagc cacattctat 6480

gccatctgcc agcaggacag ttgccaccag gagcaagtgt gtgaggtgat cgcctcttat 6540

gcccacctct gtcggaccaa cggggtctgc gttgactgga ggacacctga tttctgtgct 6600

atgtcatgcc caccatctct ggtctacaac cactgtgagc atggctgtcc ccggcactgt 6660

gatggcaacg tgagctcctg tggggaccat ccctccgaag gctgtttctg ccctccagat 6720

aaagtcatgt tggaaggcag ctgtgtccct gaagaggcct gcactcagtg cattggtgag 6780

gatggagtcc agcaccagtt cctggaagcc tgggtcccgg accaccagcc ctgtcagatc 6840

tgcacatgcc tcagcgggcg gaaggtcaac tgcacaacgc agccctgccc cacggccaaa 6900

gctcccacgt gtggcctgtg tgaagtagcc cgcctccgcc agaatgcaga ccagtgctgc 6960

cccgagtatg agtgtgtgtg tgacccagtg agctgtgacc tgcccccagt gcctcactgt 7020

gaacgtggcc tccagcccac actgaccaac cctggcgagt gcagacccaa cttcacctgc 7080

gcctgcagga aggaggagtg caaaagagtg tccccaccct cctgcccccc gcaccgtttg 7140

cccacccttc ggaagaccca gtgctgtgat gagtatgagt gtgcctgcaa ctgtgtcaac 7200

tccacagtga gctgtcccct tgggtacttg gcctcaaccg ccaccaatga ctgtggctgt 7260

accacaacca cctgccttcc cgacaaggtg tgtgtccacc gaagcaccat ctaccctgtg 7320

ggccagttct gggaggaggg ctgcgatgtg tgcacctgca ccgacatgga ggatgccgtg 7380

atgggcctcc gcgtggccca gtgctcccag aagccctgtg aggacagctg tcggtcgggc 7440

ttcacttacg ttctgcatga aggcgagtgc tgtggaaggt gcctgccatc tgcctgtgag 7500

gtggtgactg gctcaccgcg gggggactcc cagtcttcct ggaagagtgt cggctcccag 7560

tgggcctccc cggagaaccc ctgcctcatc aatgagtgtg tccgagtgaa ggaggaggtc 7620

tttatacaac aaaggaacgt ctcctgcccc cagctggagg tccctgtctg cccctcgggc 7680
tttcagctga gctgtaagac ctcagcgtgc tgcccaagct gtcgctgtga gcgcatggag 7740
gcctgcatgc tcaatggcac tgtcattggg cccgggaaga ctgtgatgat cgatgtgtgc 7800
acgacctgcc gctgcatggt gcaggtgggg gtcatctctg gattcaagct ggagtgcagg 7860
aagaccacct gcaacccctg ccccctgggt tacaaggaag aaaataacac aggtgaatgt 7920
tgtgggagat gtttgcctac ggcttgcacc attcagctaa gaggaggaca gatcatgaca 7980
ctgaagcgtg atgagacgct ccaggatggc tgtgatactc acttctgcaa ggtcaatgag 8040
agaggagagt acttctggga gaagagggtc acaggctgcc cacccтttga tgaacacaag 8100
tgtcttgctg agggaggtaa aattatgaaa attccaggca cctgctgtga cacatgtgag 8160
gagcctgagt gcaacgacat cactgccagg ctgcagtatg tcaaggtggg aagctgtaag 8220
tctgaagtag aggtggatat ccactactgc cagggcaaat gtgccagcaa agccatgtac 8280
tccattgaca tcaacgatgt gcaggaccag tgctcctgct gctctccgac acggacggag 8340
cccatgcagg tggccctgca ctgcaccaat ggctctgttg tgtaccatga ggttctcaat 8400
gccatggagt gcaaatgctc ccccaggaag tgcagcaagt ga 8442

<210> 129
<211> 2809
<212> PRT
<213> 人工序列

<220>
<223> 人類VWF
<220>
<221> 尚未歸類的特徵
<222> (540)..(540)
<223> Xaa可以是任何天然存在的胺基酸
<220>
<221> 尚未歸類的特徵
<222> (1944)..(1944)
<223> Xaa可以是任何天然存在的胺基酸
<220>
<221> 尚未歸類的特徵
<222> (2012)..(2012)
<223> Xaa可以是任何天然存在的胺基酸

<400> 129

```
Met Ile Pro Ala Arg Phe Ala Gly Val Leu Leu Ala Leu Ala Leu Ile
1               5                  10                  15

Leu Pro Gly Thr Leu Cys Ala Glu Gly Thr Arg Gly Arg Ser Ser Thr
            20                  25                  30

Ala Arg Cys Ser Leu Phe Gly Ser Asp Phe Val Asn Thr Phe Asp Gly
        35                  40                  45

Ser Met Tyr Ser Phe Ala Gly Tyr Cys Ser Tyr Leu Leu Ala Gly Gly
    50                  55                  60

Cys Gln Lys Arg Ser Phe Ser Ile Ile Gly Asp Phe Gln Asn Gly Lys
65                  70                  75                  80

Arg Val Ser Leu Ser Val Tyr Leu Gly Glu Phe Phe Asp Ile His Leu
                85                  90                  95

Phe Val Asn Gly Thr Val Thr Gln Gly Asp Gln Arg Val Ser Met Pro
            100                 105                 110

Tyr Ala Ser Lys Gly Leu Tyr Leu Glu Thr Glu Ala Gly Tyr Tyr Lys
        115                 120                 125

Leu Ser Gly Glu Ala Tyr Gly Phe Val Ala Arg Ile Asp Gly Ser Gly
    130                 135                 140

Asn Phe Gln Val Leu Leu Ser Asp Arg Tyr Phe Asn Lys Thr Cys Gly
145                 150                 155                 160

Leu Cys Gly Asn Phe Asn Ile Phe Ala Glu Asp Asp Phe Met Thr Gln
                165                 170                 175

Glu Gly Thr Leu Thr Ser Asp Pro Tyr Asp Phe Ala Asn Ser Trp Ala
            180                 185                 190
```

Leu Ser Ser Gly Glu Gln Trp Cys Glu Arg Ala Ser Pro Pro Ser Ser
195 200 205

Ser Cys Asn Ile Ser Ser Gly Glu Met Gln Lys Gly Leu Trp Glu Gln
210 215 220

Cys Gln Leu Leu Lys Ser Thr Ser Val Phe Ala Arg Cys His Pro Leu
225 230 235 240

Val Asp Pro Glu Pro Phe Val Cys Glu Lys Thr Leu Cys Glu Cys Ala
245 250 255

Gly Gly Leu Glu Cys Ala Cys Pro Ala Leu Leu Glu Tyr Ala Arg Thr
260 265 270

Cys Ala Gln Glu Gly Met Val Leu Tyr Gly Trp Thr Asp His Ser Ala
275 280 285

Cys Ser Pro Val Cys Pro Ala Gly Met Glu Tyr Arg Gln Cys Val Ser
290 295 300

Pro Cys Ala Arg Thr Cys Gln Ser Leu His Ile Asn Glu Met Cys Gln
305 310 315 320

Glu Arg Cys Val Asp Gly Cys Ser Cys Pro Glu Gly Gln Leu Leu Asp
325 330 335

Glu Gly Leu Cys Val Glu Ser Thr Glu Cys Pro Cys Val His Ser Gly
340 345 350

Lys Arg Tyr Pro Pro Gly Thr Ser Leu Ser Arg Asp Cys Asn Thr Cys
355 360 365

Ile Cys Arg Asn Ser Gln Trp Ile Cys Ser Asn Glu Glu Cys Pro Gly
370 375 380

```
Glu Cys Leu Val Thr Gly Gln Ser His Phe Lys Ser Phe Asp Asn Arg
385                 390                 395                 400

Tyr Phe Thr Phe Ser Gly Ile Cys Gln Tyr Leu Leu Ala Arg Asp Cys
                405                 410                 415

Gln Asp His Ser Phe Ser Ile Val Ile Glu Thr Val Gln Cys Ala Asp
            420                 425                 430

Asp Arg Asp Ala Val Cys Thr Arg Ser Val Thr Val Arg Leu Pro Gly
        435                 440                 445

Leu His Asn Ser Leu Val Lys Leu Lys His Gly Ala Gly Val Ala Met
    450                 455                 460

Asp Gly Gln Asp Ile Gln Leu Pro Leu Leu Lys Gly Asp Leu Arg Ile
465                 470                 475                 480

Gln His Thr Val Thr Ala Ser Val Arg Leu Ser Tyr Gly Glu Asp Leu
                485                 490                 495

Gln Met Asp Trp Asp Gly Arg Gly Arg Leu Leu Val Lys Leu Ser Pro
            500                 505                 510

Val Tyr Ala Gly Lys Thr Cys Gly Leu Cys Gly Asn Tyr Asn Gly Asn
        515                 520                 525

Gln Gly Asp Asp Phe Leu Thr Pro Ser Gly Leu Xaa Glu Pro Arg Val
    530                 535                 540

Glu Asp Phe Gly Asn Ala Trp Lys Leu His Gly Asp Cys Gln Asp Leu
545                 550                 555                 560

Gln Lys Gln His Ser Asp Pro Cys Ala Leu Asn Pro Arg Met Thr Arg
                565                 570                 575
```

Phe Ser Glu Glu Ala Cys Ala Val Leu Thr Ser Pro Thr Phe Glu Ala
580 585 590

Cys His Arg Ala Val Ser Pro Leu Pro Tyr Leu Arg Asn Cys Arg Tyr
595 600 605

Asp Val Cys Ser Cys Ser Asp Gly Arg Glu Cys Leu Cys Gly Ala Ser
610 615 620

Tyr Ala Ala Ala Cys Ala Gly Arg Gly Val Arg Val Ala Trp Arg Glu
625 630 635 640

Pro Gly Arg Cys Glu Leu Asn Cys Pro Lys Gly Gln Val Tyr Leu Gln
645 650 655

Cys Gly Thr Pro Cys Asn Leu Thr Cys Arg Ser Leu Ser Tyr Pro Asp
660 665 670

Glu Glu Cys Asn Glu Ala Cys Leu Glu Gly Cys Phe Cys Pro Pro Gly
675 680 685

Leu Tyr Met Asp Glu Arg Gly Asp Cys Val Pro Lys Ala Gln Cys Pro
690 695 700

Cys Tyr Tyr Asp Gly Glu Ile Phe Gln Pro Glu Asp Ile Phe Ser Asp
705 710 715 720

His His Thr Met Cys Tyr Cys Glu Asp Gly Phe Met His Cys Thr Met
725 730 735

Ser Gly Val Pro Gly Ser Leu Leu Pro Asp Ala Val Leu Ser Ser Pro
740 745 750

Leu Ser His Arg Ser Lys Arg Ser Leu Ser Cys Arg Pro Pro Met Val
755 760 765

Lys Leu Val Cys Pro Ala Asp Asn Leu Arg Ala Glu Gly Leu Glu Cys
770 775 780

Thr Lys Thr Cys Gln Asn Tyr Asp Leu Glu Cys Met Ser Met Gly Cys
785 790 795 800

Val Ser Gly Cys Leu Cys Pro Pro Gly Met Val Arg His Glu Asn Arg
805 810 815

Cys Val Ala Leu Glu Arg Cys Pro Cys Phe His Gln Gly Lys Glu Tyr
820 825 830

Ala Pro Gly Glu Thr Val Lys Ile Gly Cys Asn Thr Cys Val Cys Arg
835 840 845

Asp Arg Lys Trp Asn Cys Thr Asp His Val Cys Asp Ala Thr Cys Ser
850 855 860

Thr Ile Gly Met Ala His Tyr Leu Thr Phe Asp Gly Leu Lys Tyr Leu
865 870 875 880

Phe Pro Gly Glu Cys Gln Tyr Val Leu Val Gln Asp Tyr Cys Gly Ser
885 890 895

Asn Pro Gly Thr Phe Arg Ile Leu Val Gly Asn Lys Gly Cys Ser His
900 905 910

Pro Ser Val Lys Cys Lys Lys Arg Val Thr Ile Leu Val Glu Gly Gly
915 920 925

Glu Ile Glu Leu Phe Asp Gly Glu Val Asn Val Lys Arg Pro Met Lys
930 935 940

Asp Glu Thr His Phe Glu Val Val Glu Ser Gly Arg Tyr Ile Ile Leu
945 950 955 960

Leu Leu Gly Lys Ala Leu Ser Val Val Trp Asp Arg His Leu Ser Ile
965 970 975

Ser Val Val Leu Lys Gln Thr Tyr Gln Glu Lys Val Cys Gly Leu Cys
980 985 990

Gly Asn Phe Asp Gly Ile Gln Asn Asn Asp Leu Thr Ser Ser Asn Leu
995 1000 1005

Gln Val Glu Glu Asp Pro Val Asp Phe Gly Asn Ser Trp Lys Val
1010 1015 1020

Ser Ser Gln Cys Ala Asp Thr Arg Lys Val Pro Leu Asp Ser Ser
1025 1030 1035

Pro Ala Thr Cys His Asn Asn Ile Met Lys Gln Thr Met Val Asp
1040 1045 1050

Ser Ser Cys Arg Ile Leu Thr Ser Asp Val Phe Gln Asp Cys Asn
1055 1060 1065

Lys Leu Val Asp Pro Glu Pro Tyr Leu Asp Val Cys Ile Tyr Asp
1070 1075 1080

Thr Cys Ser Cys Glu Ser Ile Gly Asp Cys Ala Cys Phe Cys Asp
1085 1090 1095

Thr Ile Ala Ala Tyr Ala His Val Cys Ala Gln His Gly Lys Val
1100 1105 1110

Val Thr Trp Arg Thr Ala Thr Leu Cys Pro Gln Ser Cys Glu Glu
1115 1120 1125

Arg Asn Leu Arg Glu Asn Gly Tyr Glu Cys Glu Trp Arg Tyr Asn
1130 1135 1140

Ser Cys Ala Pro Ala Cys Gln Val Thr Cys Gln His Pro Glu Pro
1145 1150 1155

Leu Ala Cys Pro Val Gln Cys Val Glu Gly Cys His Ala His Cys
1160 1165 1170

Pro Pro Gly Lys Ile Leu Asp Glu Leu Leu Gln Thr Cys Val Asp
1175 1180 1185

Pro Glu Asp Cys Pro Val Cys Glu Val Ala Gly Arg Arg Phe Ala
1190 1195 1200

Ser Gly Lys Lys Val Thr Leu Asn Pro Ser Asp Pro Glu His Cys
1205 1210 1215

Gln Ile Cys His Cys Asp Val Val Asn Leu Thr Cys Glu Ala Cys
1220 1225 1230

Gln Glu Pro Gly Gly Leu Val Val Pro Pro Thr Asp Ala Pro Val
1235 1240 1245

Ser Pro Thr Thr Leu Tyr Val Glu Asp Ile Ser Glu Pro Pro Leu
1250 1255 1260

His Asp Phe Tyr Cys Ser Arg Leu Leu Asp Leu Val Phe Leu Leu
1265 1270 1275

Asp Gly Ser Ser Arg Leu Ser Glu Ala Glu Phe Glu Val Leu Lys
1280 1285 1290

Ala Phe Val Val Asp Met Met Glu Arg Leu Arg Ile Ser Gln Lys
1295 1300 1305

Trp Val Arg Val Ala Val Val Glu Tyr His Asp Gly Ser His Ala
1310 1315 1320

```
Tyr Ile  Gly Leu Lys Asp Arg  Lys Arg Pro Ser Glu  Leu Arg Arg
    1325                 1330                 1335

Ile Ala  Ser Gln Val Lys Tyr  Ala Gly Ser Gln Val  Ala Ser Thr
    1340                 1345                 1350

Ser Glu  Val Leu Lys Tyr Thr  Leu Phe Gln Ile Phe  Ser Lys Ile
    1355                 1360                 1365

Asp Arg  Pro Glu Ala Ser Arg  Ile Ala Leu Leu Leu  Met Ala Ser
    1370                 1375                 1380

Gln Glu  Pro Gln Arg Met Ser  Arg Asn Phe Val Arg  Tyr Val Gln
    1385                 1390                 1395

Gly Leu  Lys Lys Lys Lys Val  Ile Val Ile Pro Val  Gly Ile Gly
    1400                 1405                 1410

Pro His  Ala Asn Leu Lys Gln  Ile Arg Leu Ile Glu  Lys Gln Ala
    1415                 1420                 1425

Pro Glu  Asn Lys Ala Phe Val  Leu Ser Ser Val Asp  Glu Leu Glu
    1430                 1435                 1440

Gln Gln  Arg Asp Glu Ile Val  Ser Tyr Leu Cys Asp  Leu Ala Pro
    1445                 1450                 1455

Glu Ala  Pro Pro Pro Thr Leu  Pro Pro Asp Met Ala  Gln Val Thr
    1460                 1465                 1470

Val Gly  Pro Gly Leu Leu Gly  Val Ser Thr Leu Gly  Pro Lys Arg
    1475                 1480                 1485

Asn Ser  Met Val Leu Asp Val  Ala Phe Val Leu Glu  Gly Ser Asp
    1490                 1495                 1500
```

Lys Ile Gly Glu Ala Asp Phe Asn Arg Ser Lys Glu Phe Met Glu
1505 1510 1515

Glu Val Ile Gln Arg Met Asp Val Gly Gln Asp Ser Ile His Val
1520 1525 1530

Thr Val Leu Gln Tyr Ser Tyr Met Val Thr Val Glu Tyr Pro Phe
1535 1540 1545

Ser Glu Ala Gln Ser Lys Gly Asp Ile Leu Gln Arg Val Arg Glu
1550 1555 1560

Ile Arg Tyr Gln Gly Gly Asn Arg Thr Asn Thr Gly Leu Ala Leu
1565 1570 1575

Arg Tyr Leu Ser Asp His Ser Phe Leu Val Ser Gln Gly Asp Arg
1580 1585 1590

Glu Gln Ala Pro Asn Leu Val Tyr Met Val Thr Gly Asn Pro Ala
1595 1600 1605

Ser Asp Glu Ile Lys Arg Leu Pro Gly Asp Ile Gln Val Val Pro
1610 1615 1620

Ile Gly Val Gly Pro Asn Ala Asn Val Gln Glu Leu Glu Arg Ile
1625 1630 1635

Gly Trp Pro Asn Ala Pro Ile Leu Ile Gln Asp Phe Glu Thr Leu
1640 1645 1650

Pro Arg Glu Ala Pro Asp Leu Val Leu Gln Arg Cys Cys Ser Gly
1655 1660 1665

Glu Gly Leu Gln Ile Pro Thr Leu Ser Pro Ala Pro Asp Cys Ser
1670 1675 1680

```
Gln Pro  Leu Asp Val Ile Leu  Leu Leu Asp Gly Ser  Ser Ser Phe
    1685                 1690                 1695

Pro Ala  Ser Tyr Phe Asp Glu  Met Lys Ser Phe Ala  Lys Ala Phe
    1700                 1705                 1710

Ile Ser  Lys Ala Asn Ile Gly  Pro Arg Leu Thr Gln  Val Ser Val
    1715                 1720                 1725

Leu Gln  Tyr Gly Ser Ile Thr  Thr Ile Asp Val Pro  Trp Asn Val
    1730                 1735                 1740

Val Pro  Glu Lys Ala His Leu  Leu Ser Leu Val Asp  Val Met Gln
    1745                 1750                 1755

Arg Glu  Gly Gly Pro Ser Gln  Ile Gly Asp Ala Leu  Gly Phe Ala
    1760                 1765                 1770

Val Arg  Tyr Leu Thr Ser Glu  Met His Gly Ala Arg  Pro Gly Ala
    1775                 1780                 1785

Ser Lys  Ala Val Val Ile Leu  Val Thr Asp Val Ser  Val Asp Ser
    1790                 1795                 1800

Val Asp  Ala Ala Ala Asp Ala  Ala Arg Ser Asn Arg  Val Thr Val
    1805                 1810                 1815

Phe Pro  Ile Gly Ile Gly Asp  Arg Tyr Asp Ala Ala  Gln Leu Arg
    1820                 1825                 1830

Ile Leu  Ala Gly Pro Ala Gly  Asp Ser Asn Val Val  Lys Leu Gln
    1835                 1840                 1845

Arg Ile  Glu Asp Leu Pro Thr  Met Val Thr Leu Gly  Asn Ser Phe
    1850                 1855                 1860
```

```
Leu His  Lys Leu Cys Ser Gly  Phe Val Arg Ile Cys  Met Asp Glu
    1865                 1870                 1875

Asp Gly  Asn Glu Lys Arg Pro  Gly Asp Val Trp Thr  Leu Pro Asp
    1880                 1885                 1890

Gln Cys  His Thr Val Thr Cys  Gln Pro Asp Gly Gln  Thr Leu Leu
    1895                 1900                 1905

Lys Ser  His Arg Val Asn Cys  Asp Arg Gly Leu Arg  Pro Ser Cys
    1910                 1915                 1920

Pro Asn  Ser Gln Ser Pro Val  Lys Val Glu Glu Thr  Cys Gly Cys
    1925                 1930                 1935

Arg Trp  Thr Cys Pro Xaa Val  Cys Thr Gly Ser Ser  Thr Arg His
    1940                 1945                 1950

Ile Val  Thr Phe Asp Gly Gln  Asn Phe Lys Leu Thr  Gly Ser Cys
    1955                 1960                 1965

Ser Tyr  Val Leu Phe Gln Asn  Lys Glu Gln Asp Leu  Glu Val Ile
    1970                 1975                 1980

Leu His  Asn Gly Ala Cys Ser  Pro Gly Ala Arg Gln  Gly Cys Met
    1985                 1990                 1995

Lys Ser  Ile Glu Val Lys His  Ser Ala Leu Ser Val  Glu Xaa His
    2000                 2005                 2010

Ser Asp  Met Glu Val Thr Val  Asn Gly Arg Leu Val  Ser Val Pro
    2015                 2020                 2025

Tyr Val  Gly Gly Asn Met Glu  Val Asn Val Tyr Gly  Ala Ile Met
    2030                 2035                 2040
```

His Glu Val Arg Phe Asn His Leu Gly His Ile Phe Thr Phe Thr
2045 2050 2055

Pro Gln Asn Asn Glu Phe Gln Leu Gln Leu Ser Pro Lys Thr Phe
2060 2065 2070

Ala Ser Lys Thr Tyr Gly Leu Cys Gly Ile Cys Asp Glu Asn Gly
2075 2080 2085

Ala Asn Asp Phe Met Leu Arg Asp Gly Thr Val Thr Thr Asp Trp
2090 2095 2100

Lys Thr Leu Val Gln Glu Trp Thr Val Gln Arg Pro Gly Gln Thr
2105 2110 2115

Cys Gln Pro Ile Leu Glu Glu Gln Cys Leu Val Pro Asp Ser Ser
2120 2125 2130

His Cys Gln Val Leu Leu Leu Pro Leu Phe Ala Glu Cys His Lys
2135 2140 2145

Val Leu Ala Pro Ala Thr Phe Tyr Ala Ile Cys Gln Gln Asp Ser
2150 2155 2160

Cys His Gln Glu Gln Val Cys Glu Val Ile Ala Ser Tyr Ala His
2165 2170 2175

Leu Cys Arg Thr Asn Gly Val Cys Val Asp Trp Arg Thr Pro Asp
2180 2185 2190

Phe Cys Ala Met Ser Cys Pro Pro Ser Leu Val Tyr Asn His Cys
2195 2200 2205

Glu His Gly Cys Pro Arg His Cys Asp Gly Asn Val Ser Ser Cys
2210 2215 2220

Gly Asp His Pro Ser Glu Gly Cys Phe Cys Pro Pro Asp Lys Val
2225 2230 2235

Met Leu Glu Gly Ser Cys Val Pro Glu Glu Ala Cys Thr Gln Cys
2240 2245 2250

Ile Gly Glu Asp Gly Val Gln His Gln Phe Leu Glu Ala Trp Val
2255 2260 2265

Pro Asp His Gln Pro Cys Gln Ile Cys Thr Cys Leu Ser Gly Arg
2270 2275 2280

Lys Val Asn Cys Thr Thr Gln Pro Cys Pro Thr Ala Lys Ala Pro
2285 2290 2295

Thr Cys Gly Leu Cys Glu Val Ala Arg Leu Arg Gln Asn Ala Asp
2300 2305 2310

Gln Cys Cys Pro Glu Tyr Glu Cys Val Cys Asp Pro Val Ser Cys
2315 2320 2325

Asp Leu Pro Pro Val Pro His Cys Glu Arg Gly Leu Gln Pro Thr
2330 2335 2340

Leu Thr Asn Pro Gly Glu Cys Arg Pro Asn Phe Thr Cys Ala Cys
2345 2350 2355

Arg Lys Glu Glu Cys Lys Arg Val Ser Pro Pro Ser Cys Pro Pro
2360 2365 2370

His Arg Leu Pro Thr Leu Arg Lys Thr Gln Cys Cys Asp Glu Tyr
2375 2380 2385

Glu Cys Ala Cys Asn Cys Val Asn Ser Thr Val Ser Cys Pro Leu
2390 2395 2400

Gly Tyr Leu Ala Ser Thr Ala Thr Asn Asp Cys Gly Cys Thr Thr
2405 2410 2415

Thr Thr Cys Leu Pro Asp Lys Val Cys Val His Arg Ser Thr Ile
2420 2425 2430

Tyr Pro Val Gly Gln Phe Trp Glu Glu Gly Cys Asp Val Cys Thr
2435 2440 2445

Cys Thr Asp Met Glu Asp Ala Val Met Gly Leu Arg Val Ala Gln
2450 2455 2460

Cys Ser Gln Lys Pro Cys Glu Asp Ser Cys Arg Ser Gly Phe Thr
2465 2470 2475

Tyr Val Leu His Glu Gly Glu Cys Cys Gly Arg Cys Leu Pro Ser
2480 2485 2490

Ala Cys Glu Val Val Thr Gly Ser Pro Arg Gly Asp Ser Gln Ser
2495 2500 2505

Ser Trp Lys Ser Val Gly Ser Gln Trp Ala Ser Pro Glu Asn Pro
2510 2515 2520

Cys Leu Ile Asn Glu Cys Val Arg Val Lys Glu Glu Val Phe Ile
2525 2530 2535

Gln Gln Arg Asn Val Ser Cys Pro Gln Leu Glu Val Pro Val Cys
2540 2545 2550

Pro Ser Gly Phe Gln Leu Ser Cys Lys Thr Ser Ala Cys Cys Pro
2555 2560 2565

Ser Cys Arg Cys Glu Arg Met Glu Ala Cys Met Leu Asn Gly Thr
2570 2575 2580

```
Val Ile  Gly Pro Gly Lys Thr  Val Met Ile Asp Val  Cys Thr Thr
    2585                 2590                 2595

Cys Arg  Cys Met Val Gln Val  Gly Val Ile Ser Gly  Phe Lys Leu
    2600                 2605                 2610

Glu Cys  Arg Lys Thr Thr Cys  Asn Pro Cys Pro Leu  Gly Tyr Lys
    2615                 2620                 2625

Glu Glu  Asn Asn Thr Gly Glu  Cys Cys Gly Arg Cys  Leu Pro Thr
    2630                 2635                 2640

Ala Cys  Thr Ile Gln Leu Arg  Gly Gly Gln Ile Met  Thr Leu Lys
    2645                 2650                 2655

Arg Asp  Glu Thr Leu Gln Asp  Gly Cys Asp Thr His  Phe Cys Lys
    2660                 2665                 2670

Val Asn  Glu Arg Gly Glu Tyr  Phe Trp Glu Lys Arg  Val Thr Gly
    2675                 2680                 2685

Cys Pro  Pro Phe Asp Glu His  Lys Cys Leu Ala Glu  Gly Gly Lys
    2690                 2695                 2700

Ile Met  Lys Ile Pro Gly Thr  Cys Cys Asp Thr Cys  Glu Glu Pro
    2705                 2710                 2715

Glu Cys  Asn Asp Ile Thr Ala  Arg Leu Gln Tyr Val  Lys Val Gly
    2720                 2725                 2730

Ser Cys  Lys Ser Glu Val Glu  Val Asp Ile His Tyr  Cys Gln Gly
    2735                 2740                 2745

Lys Cys  Ala Ser Lys Ala Met  Tyr Ser Ile Asp Ile  Asn Asp Val
    2750                 2755                 2760
```

Gln Asp Gln Cys Ser Cys Cys Ser Pro Thr Arg Thr Glu Pro Met
2765 2770 2775

Gln Val Ala Leu His Cys Thr Asn Gly Ser Val Val Tyr His Glu
2780 2785 2790

Val Leu Asn Ala Met Glu Cys Lys Cys Ser Pro Arg Lys Cys Ser
2795 2800 2805

Lys

<210> 130
<211> 36
<212> PRT
<213> 人工序列

<220>
<223> AE36

<400> 130
Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Glu
1 5 10 15

Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly
20 25 30

Ser Glu Thr Pro
35

<210> 131
<211> 42
<212> PRT

<213> 人工序列

<220>
<223> AE42

<400> 131
Gly Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly
1 5 10 15

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala
20 25 30

Thr Ser Gly Ser Glu Thr Pro Ala Ser Ser
35 40

<210> 132
<211> 78
<212> PRT
<213> 人工序列

<220>
<223> AE72

<400> 132
Gly Ala Pro Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser
1 5 10 15

Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala
20 25 30

Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu
35 40 45

Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr
50 55 60

Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Ala Ser Ser
65 70 75

<210> 133
<211> 78
<212> PRT
<213> 人工序列

<220>
<223> AE78

<400> 133
Gly Ala Pro Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser
1 5 10 15

Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala
20 25 30

Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu
35 40 45

Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr
50 55 60

Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Ala Ser Ser
65 70 75

<210> 134
<211> 143
<212> PRT
<213> 人工序列

<220>

<223> AE144

<400> 134

Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu
1 5 10 15

Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly
20 25 30

Ser Glu Thr Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu
35 40 45

Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Ser Glu Pro
50 55 60

Ala Thr Ser Gly Ser Glu Thr Pro Gly Ser Glu Pro Ala Thr Ser Gly
65 70 75 80

Ser Glu Thr Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro
85 90 95

Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu
100 105 110

Ser Ala Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser
115 120 125

Glu Thr Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro
130 135 140

<210> 135

<211> 144

<212> PRT

<213> 人工序列

<220>

<223> AE144_6B

<400> 135

Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser
1 5 10 15

Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
20 25 30

Ser Gly Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly
35 40 45

Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Ser Glu Pro Ala
50 55 60

Thr Ser Gly Ser Glu Thr Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser
65 70 75 80

Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
85 90 95

Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Ser Glu Pro Ala
100 105 110

Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
115 120 125

Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
130 135 140

<210> 136

<211> 144

<212> PRT

<213> 人工序列

<220>

<223> AG144

<400> 136

Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ser Ser Thr
1 5 10 15

Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Pro Ser Ala Ser Thr
20 25 30

Gly Thr Gly Pro Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro
35 40 45

Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ala Ser Pro
50 55 60

Gly Thr Ser Ser Thr Gly Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala
65 70 75 80

Thr Gly Ser Pro Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro
85 90 95

Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ser Ser Pro
100 105 110

Ser Ala Ser Thr Gly Thr Gly Pro Gly Thr Pro Gly Ser Gly Thr Ala
115 120 125

Ser Ser Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro
130 135 140

<210> 137

<211> 144

<212> PRT

<213> 人工序列

<220>

<223> AG144_A

<400> 137

Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ser Ser Pro
1 5 10 15

Ser Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser Pro Ser Ala Ser Thr
20 25 30

Gly Thr Gly Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro
35 40 45

Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Pro
50 55 60

Ser Ala Ser Thr Gly Thr Gly Pro Gly Ala Ser Pro Gly Thr Ser Ser
65 70 75 80

Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro
85 90 95

Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Thr Pro Gly
100 105 110

Ser Gly Thr Ala Ser Ser Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
115 120 125

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
130 135 140

<210> 138

<211> 288

<212> PRT

<213> 人工序列

<220>

<223> AE288

<400> 138

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro
1 5 10 15

Ala Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
20 25 30

Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro
35 40 45

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr
50 55 60

Glu Pro Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr
65 70 75 80

Ser Thr Glu Glu Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro
85 90 95

Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu
100 105 110

Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Pro Ala Gly Ser Pro Thr
115 120 125

Ser Thr Glu Glu Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu
130 135 140

Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu
145 150 155 160

Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Glu Ser Ala Thr Pro
165 170 175

Glu Ser Gly Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro
180 185 190

Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Ser Glu Pro
195 200 205

Ala Thr Ser Gly Ser Glu Thr Pro Gly Ser Pro Ala Gly Ser Pro Thr
210 215 220

Ser Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro
225 230 235 240

Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Ser Glu Pro
245 250 255

Ala Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
260 265 270

Glu Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro
275 280 285

<210> 139
<211> 288
<212> PRT
<213> 人工序列

<220>
<223> AE288_2

<400> 139
Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Glu

```
1               5                   10                  15

Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly
            20                  25                  30

Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro
        35                  40                  45

Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr
    50                  55                  60

Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu
65                  70                  75                  80

Gly Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro
                85                  90                  95

Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr
            100                 105                 110

Glu Pro Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr
        115                 120                 125

Ser Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro
    130                 135                 140

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro
145                 150                 155                 160

Ala Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
                165                 170                 175

Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro
            180                 185                 190

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr
```

195 200 205

Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr Pro
210 215 220

Glu Ser Gly Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu
225 230 235 240

Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Ser Pro Ala
245 250 255

Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Glu Ser Ala Thr Pro
260 265 270

Glu Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro
275 280 285

<210> 140
<211> 288
<212> PRT
<213> 人工序列

<220>
<223> AG288

<400> 140
Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ala Ser
1 5 10 15

Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr
20 25 30

Ala Ser Ser Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser
35 40 45

```
Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ser Ser
    50                  55                  60

Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr
65                  70                  75                  80

Ala Ser Ser Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser
                85                  90                  95

Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser
            100                 105                 110

Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser Pro Ser Ala Ser
        115                 120                 125

Thr Gly Thr Gly Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser
    130                 135                 140

Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ser Ser
145                 150                 155                 160

Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Pro Ser Ala Ser
                165                 170                 175

Thr Gly Thr Gly Pro Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly
            180                 185                 190

Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ala Ser
        195                 200                 205

Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ser Ser Thr Pro Ser Gly
    210                 215                 220

Ala Thr Gly Ser Pro Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly
225                 230                 235                 240
```

Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ser Ser
245 250 255

Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Thr Pro Gly Ser Gly Thr
260 265 270

Ala Ser Ser Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser
275 280 285

<210> 141
<211> 576
<212> PRT
<213> 人工序列

<220>
<223> AE576

<400> 141
Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Glu
1 5 10 15

Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser Glu
20 25 30

Gly Ser Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu
35 40 45

Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr
50 55 60

Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr Pro
65 70 75 80

Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro
85 90 95

Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Ser Pro Ala
100 105 110

Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Glu Ser Ala Thr Pro
115 120 125

Glu Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro
130 135 140

Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala
145 150 155 160

Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu
165 170 175

Gly Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro
180 185 190

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr
195 200 205

Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr Pro
210 215 220

Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro
225 230 235 240

Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr
245 250 255

Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr Pro
260 265 270

Glu Ser Gly Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro
275 280 285

Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Glu
290 295 300

Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly
305 310 315 320

Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro
325 330 335

Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr
340 345 350

Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu
355 360 365

Gly Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro
370 375 380

Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr
385 390 395 400

Glu Pro Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr
405 410 415

Ser Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro
420 425 430

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro
435 440 445

Ala Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
450 455 460

Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro
465 470 475 480

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr
485 490 495

Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr Pro
500 505 510

Glu Ser Gly Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu
515 520 525

Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Ser Pro Ala
530 535 540

Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Glu Ser Ala Thr Pro
545 550 555 560

Glu Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro
565 570 575

<210> 142
<211> 576
<212> PRT
<213> 人工序列

<220>
<223> AG576

<400> 142
Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ser Ser
1 5 10 15

Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Pro Ser Ala Ser
20 25 30

Thr Gly Thr Gly Pro Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly

```
        35                  40                  45

Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser
    50                  55                  60

Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser
65                  70                  75                  80

Ser Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser
                85                  90                  95

Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Thr Pro
            100                 105                 110

Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ala Ser Pro Gly Thr Ser
        115                 120                 125

Ser Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser
    130                 135                 140

Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ser Ser
145                 150                 155                 160

Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Thr Pro Gly Ser Gly Thr
                165                 170                 175

Ala Ser Ser Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser
            180                 185                 190

Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ala Ser
        195                 200                 205

Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ser Ser Thr Pro Ser Gly
    210                 215                 220

Ala Thr Gly Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser
```

225 230 235 240

Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Thr Pro
245 250 255

Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ser Ser Thr Pro Ser Gly
260 265 270

Ala Thr Gly Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser
275 280 285

Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser
290 295 300

Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Ala Ser Pro Gly Thr Ser
305 310 315 320

Ser Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser
325 330 335

Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ala Ser
340 345 350

Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser
355 360 365

Ser Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser
370 375 380

Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Thr Pro
385 390 395 400

Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ser Ser Thr Pro Ser Gly
405 410 415

Ala Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser

420 425 430

Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Thr Pro
435 440 445

Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ser Ser Thr Pro Ser Gly
450 455 460

Ala Thr Gly Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser
465 470 475 480

Pro Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser
485 490 495

Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Ala Ser Pro Gly Thr Ser
500 505 510

Ser Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser
515 520 525

Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser
530 535 540

Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser Pro Ser Ala Ser
545 550 555 560

Thr Gly Thr Gly Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser
565 570 575

<210> 143
<211> 864
<212> PRT
<213> 人工序列

<220>

<223> AE864

<400> 143

Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Glu
1 5 10 15

Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser Glu
20 25 30

Gly Ser Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu
35 40 45

Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr
50 55 60

Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr Pro
65 70 75 80

Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro
85 90 95

Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Ser Pro Ala
100 105 110

Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Glu Ser Ala Thr Pro
115 120 125

Glu Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro
130 135 140

Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala
145 150 155 160

Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu
165 170 175

Gly Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro
180 185 190

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr
195 200 205

Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr Pro
210 215 220

Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro
225 230 235 240

Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr
245 250 255

Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr Pro
260 265 270

Glu Ser Gly Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro
275 280 285

Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Glu
290 295 300

Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly
305 310 315 320

Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro
325 330 335

Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr
340 345 350

Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu
355 360 365

```
Gly Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro
    370                 375                 380

Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr
385                 390                 395                 400

Glu Pro Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr
                405                 410                 415

Ser Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro
            420                 425                 430

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro
        435                 440                 445

Ala Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
    450                 455                 460

Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro
465                 470                 475                 480

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr
                485                 490                 495

Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr Pro
            500                 505                 510

Glu Ser Gly Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu
        515                 520                 525

Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Ser Pro Ala
    530                 535                 540

Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Glu Ser Ala Thr Pro
545                 550                 555                 560
```

```
Glu Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro
                565                 570                 575

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro
            580                 585                 590

Ala Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
        595                 600                 605

Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro
    610                 615                 620

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr
625                 630                 635                 640

Glu Pro Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr
                645                 650                 655

Ser Thr Glu Glu Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro
            660                 665                 670

Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu
        675                 680                 685

Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Pro Ala Gly Ser Pro Thr
    690                 695                 700

Ser Thr Glu Glu Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu
705                 710                 715                 720

Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu
                725                 730                 735

Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Glu Ser Ala Thr Pro
            740                 745                 750
```

Glu Ser Gly Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro
755 760 765

Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Ser Glu Pro
770 775 780

Ala Thr Ser Gly Ser Glu Thr Pro Gly Ser Pro Ala Gly Ser Pro Thr
785 790 795 800

Ser Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro
805 810 815

Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Ser Glu Pro
820 825 830

Ala Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
835 840 845

Glu Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro
850 855 860

<210> 144
<211> 864
<212> PRT
<213> 人工序列

<220>
<223> AG864

<400> 144
Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ser Ser Pro
1 5 10 15

Ser Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser Pro Ser Ala Ser Thr
20 25 30

Gly Thr Gly Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro
35 40 45

Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Pro
50 55 60

Ser Ala Ser Thr Gly Thr Gly Pro Gly Ala Ser Pro Gly Thr Ser Ser
65 70 75 80

Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro
85 90 95

Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Thr Pro Gly
100 105 110

Ser Gly Thr Ala Ser Ser Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
115 120 125

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
130 135 140

Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ser Ser Thr
145 150 155 160

Pro Ser Gly Ala Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
165 170 175

Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro
180 185 190

Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Pro
195 200 205

Ser Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser Pro Ser Ala Ser Thr
210 215 220

Gly Thr Gly Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro
225 230 235 240

Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ala Ser Pro
245 250 255

Gly Thr Ser Ser Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
260 265 270

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
275 280 285

Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ala Ser Pro
290 295 300

Gly Thr Ser Ser Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
305 310 315 320

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
325 330 335

Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Thr Pro Gly
340 345 350

Ser Gly Thr Ala Ser Ser Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
355 360 365

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
370 375 380

Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ser Ser Thr
385 390 395 400

Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala
405 410 415

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
420 425 430

Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ser Ser Thr
435 440 445

Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala
450 455 460

Thr Gly Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro
465 470 475 480

Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Ala Ser Pro
485 490 495

Gly Thr Ser Ser Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
500 505 510

Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro
515 520 525

Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ala Ser Pro
530 535 540

Gly Thr Ser Ser Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
545 550 555 560

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
565 570 575

Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ser Ser Thr
580 585 590

Pro Ser Gly Ala Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr Ala
595 600 605

Ser Ser Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro
610 615 620

Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ser Ser Thr
625 630 635 640

Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala
645 650 655

Thr Gly Ser Pro Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro
660 665 670

Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Ala Ser Pro
675 680 685

Gly Thr Ser Ser Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr Ala
690 695 700

Ser Ser Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro
705 710 715 720

Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser Pro
725 730 735

Ser Ala Ser Thr Gly Thr Gly Pro Gly Ala Ser Pro Gly Thr Ser Ser
740 745 750

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
755 760 765

Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Pro
770 775 780

Ser Ala Ser Thr Gly Thr Gly Pro Gly Ala Ser Pro Gly Thr Ser Ser
785 790 795 800

Thr Gly Ser Pro Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro
805 810 815

Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ser Ser Thr
820 825 830

Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala
835 840 845

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
850 855 860

<210> 145
<211> 72
<212> PRT
<213> 人工序列

<220>
<223> XTEN_AE72_2A_1

<400> 145
Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly
1 5 10 15

Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly
20 25 30

Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
35 40 45

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr Glu
50 55 60

Pro Ser Glu Gly Ser Ala Pro Gly

65 70

<210> 146
<211> 72
<212> PRT
<213> 人工序列

<220>
<223> XTEN_AE72_2A_2

<400> 146
Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala
1 5 10 15

Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly
20 25 30

Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
35 40 45

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Glu Ser
50 55 60

Ala Thr Pro Glu Ser Gly Pro Gly
65 70

<210> 147
<211> 72
<212> PRT
<213> 人工序列

<220>
<223> XTEN_AE72_3B_1

<400> 147

Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Glu Ser
1 5 10 15

Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser
20 25 30

Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly
35 40 45

Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu
50 55 60

Pro Ser Glu Gly Ser Ala Pro Gly
65 70

<210> 148

<211> 72

<212> PRT

<213> 人工序列

<220>

<223> XTEN_AE72_3B_2

<400> 148

Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu
1 5 10 15

Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly
20 25 30

Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
35 40 45

Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Thr Glu

50 55 60

Pro Ser Glu Gly Ser Ala Pro Gly
65 70

<210> 149
<211> 72
<212> PRT
<213> 人工序列

<220>
<223> XTEN_AE72_4A_2

<400> 149
Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Pro Ala Gly
1 5 10 15

Ser Pro Thr Ser Thr Glu Glu Gly Ser Pro Ala Gly Ser Pro Thr Ser
20 25 30

Thr Glu Glu Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly
35 40 45

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr Glu
50 55 60

Pro Ser Glu Gly Ser Ala Pro Gly
65 70

<210> 150
<211> 72
<212> PRT
<213> 人工序列

<220>
<223> XTEN_AE72_5A_2

<400> 150
Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Glu Ser
1 5 10 15

Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser
20 25 30

Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly
35 40 45

Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Ser Pro Ala Gly
50 55 60

Ser Pro Thr Ser Thr Glu Glu Gly
65 70

<210> 151
<211> 72
<212> PRT
<213> 人工序列

<220>
<223> XTEN_AE72_6B_1

<400> 151
Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser
1 5 10 15

Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
20 25 30

Ser Gly Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly

35 40 45

Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Ser Glu Pro Ala
50 55 60

Thr Ser Gly Ser Glu Thr Pro Gly
65 70

<210> 152
<211> 72
<212> PRT
<213> 人工序列

<220>
<223> XTEN_AE72_6B_2

<400> 152
Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Thr Glu
1 5 10 15

Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly
20 25 30

Ser Ala Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly
35 40 45

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr Glu
50 55 60

Pro Ser Glu Gly Ser Ala Pro Gly
65 70

<210> 153
<211> 72

<212> PRT
<213> 人工序列

<220>
<223> XTEN_AE72_1A_1

<400> 153
Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Glu Ser
1 5 10 15

Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly
20 25 30

Ser Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly
35 40 45

Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu
50 55 60

Pro Ser Glu Gly Ser Ala Pro Gly
65 70

<210> 154
<211> 72
<212> PRT
<213> 人工序列

<220>
<223> XTEN_AE72_1A_2

<400> 154
Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala
1 5 10 15

Thr Ser Gly Ser Glu Thr Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser

```
            20                  25                  30

Glu Thr Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly
        35                  40                  45

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr Glu
    50                  55                  60

Pro Ser Glu Gly Ser Ala Pro Gly
65                  70
```

<210> 155
<211> 144
<212> PRT
<213> 人工序列

<220>
<223> XTEN_AE144_1A

<400> 155

```
Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Glu Ser
1               5                   10                  15

Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly
            20                  25                  30

Ser Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly
        35                  40                  45

Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu
    50                  55                  60

Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
65                  70                  75                  80
```

Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly
85 90 95

Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Ser Pro Ala Gly
100 105 110

Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Glu Ser Ala Thr Pro Glu
115 120 125

Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
130 135 140

<210> 156
<211> 150
<212> PRT
<213> 人工序列

<220>
<223> AE150

<400> 156
Gly Ala Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly
1 5 10 15

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala
20 25 30

Thr Ser Gly Ser Glu Thr Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser
35 40 45

Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
50 55 60

Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Ser Glu Pro Ala
65 70 75 80

Thr Ser Gly Ser Glu Thr Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser
85 90 95

Glu Thr Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
100 105 110

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala
115 120 125

Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly
130 135 140

Ser Ala Pro Ala Ser Ser
145 150

<210> 157
<211> 150
<212> PRT
<213> 人工序列

<220>
<223> G150

<400> 157
Gly Ala Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly
1 5 10 15

Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Pro Ser
20 25 30

Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser Pro Ser Ala Ser Thr Gly
35 40 45

Thr Gly Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly

50 55 60

Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ser Ser Thr Pro
65 70 75 80

Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Pro Ser Ala Ser Thr Gly
85 90 95

Thr Gly Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly
100 105 110

Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Thr Pro Gly Ser
115 120 125

Gly Thr Ala Ser Ser Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr
130 135 140

Gly Ser Pro Ala Ser Ser
145 150

<210> 158
<211> 294
<212> PRT
<213> 人工序列

<220>
<223> AE294

<400> 158
Gly Ala Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly
1 5 10 15

Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
20 25 30

```
Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser
        35                  40                  45

Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly
    50                  55                  60

Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly
65                  70                  75                  80

Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Glu Ser Ala Thr Pro Glu
                85                  90                  95

Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly
            100                 105                 110

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Pro Ala Gly
        115                 120                 125

Ser Pro Thr Ser Thr Glu Glu Gly Ser Pro Ala Gly Ser Pro Thr Ser
    130                 135                 140

Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
145                 150                 155                 160

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Glu Ser
                165                 170                 175

Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
            180                 185                 190

Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly
        195                 200                 205

Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Ser Pro Ala Gly
    210                 215                 220
```

Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly
225 230 235 240

Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
245 250 255

Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
260 265 270

Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly
275 280 285

Ser Ala Pro Ala Ser Ser
290

<210> 159
<211> 294
<212> PRT
<213> 人工序列

<220>
<223> AG294

<400> 159
Gly Ala Pro Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
1 5 10 15

Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Thr Pro Gly
20 25 30

Ser Gly Thr Ala Ser Ser Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala
35 40 45

Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro
50 55 60

Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Thr Pro Gly
65 70 75 80

Ser Gly Thr Ala Ser Ser Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala
85 90 95

Thr Gly Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro
100 105 110

Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser Pro
115 120 125

Ser Ala Ser Thr Gly Thr Gly Pro Gly Ala Ser Pro Gly Thr Ser Ser
130 135 140

Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro
145 150 155 160

Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Pro
165 170 175

Ser Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser Pro Ser Ala Ser Thr
180 185 190

Gly Thr Gly Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
195 200 205

Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ser Ser Thr
210 215 220

Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Pro Ser Ala Ser Thr
225 230 235 240

Gly Thr Gly Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
245 250 255

Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Thr Pro Gly
260 265 270

Ser Gly Thr Ala Ser Ser Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala
275 280 285

Thr Gly Ser Ala Ser Ser
290

<210> 160
<211> 1016
<212> PRT
<213> 人工序列

<220>
<223> FIX-連接子-白蛋白

<400> 160
Tyr Asn Ser Gly Lys Leu Glu Glu Phe Val Gln Gly Asn Leu Glu Arg
1 5 10 15

Glu Cys Met Glu Glu Lys Cys Ser Phe Glu Glu Ala Arg Glu Val Phe
20 25 30

Glu Asn Thr Glu Arg Thr Thr Glu Phe Trp Lys Gln Tyr Val Asp Gly
35 40 45

Asp Gln Cys Glu Ser Asn Pro Cys Leu Asn Gly Gly Ser Cys Lys Asp
50 55 60

Asp Ile Asn Ser Tyr Glu Cys Trp Cys Pro Phe Gly Phe Glu Gly Lys
65 70 75 80

Asn Cys Glu Leu Asp Val Thr Cys Asn Ile Lys Asn Gly Arg Cys Glu

```
                85                  90                  95

Gln Phe Cys Lys Asn Ser Ala Asp Asn Lys Val Val Cys Ser Cys Thr
            100                 105                 110

Glu Gly Tyr Arg Leu Ala Glu Asn Gln Lys Ser Cys Glu Pro Ala Val
        115                 120                 125

Pro Phe Pro Cys Gly Arg Val Ser Val Ser Gln Thr Ser Lys Leu Thr
    130                 135                 140

Arg Ala Glu Thr Val Phe Pro Asp Val Asp Tyr Val Asn Ser Thr Glu
145                 150                 155                 160

Ala Glu Thr Ile Leu Asp Asn Ile Thr Gln Ser Thr Gln Ser Phe Asn
                165                 170                 175

Asp Phe Thr Arg Val Val Gly Gly Glu Asp Ala Lys Pro Gly Gln Phe
            180                 185                 190

Pro Trp Gln Val Val Leu Asn Gly Lys Val Asp Ala Phe Cys Gly Gly
        195                 200                 205

Ser Ile Val Asn Glu Lys Trp Ile Val Thr Ala Ala His Cys Val Glu
    210                 215                 220

Thr Gly Val Lys Ile Thr Val Val Ala Gly Glu His Asn Ile Glu Glu
225                 230                 235                 240

Thr Glu His Thr Glu Gln Lys Arg Asn Val Ile Arg Ile Ile Pro His
                245                 250                 255

His Asn Tyr Asn Ala Ala Ile Asn Lys Tyr Asn His Asp Ile Ala Leu
            260                 265                 270

Leu Glu Leu Asp Glu Pro Leu Val Leu Asn Ser Tyr Val Thr Pro Ile
```

```
        275                 280                 285

Cys Ile Ala Asp Lys Glu Tyr Thr Asn Ile Phe Leu Lys Phe Gly Ser
    290                 295                 300

Gly Tyr Val Ser Gly Trp Gly Arg Val Phe His Lys Gly Arg Ser Ala
305                 310                 315                 320

Leu Val Leu Gln Tyr Leu Arg Val Pro Leu Val Asp Arg Ala Thr Cys
                325                 330                 335

Leu Arg Ser Thr Lys Phe Thr Ile Tyr Asn Asn Met Phe Cys Ala Gly
            340                 345                 350

Phe His Glu Gly Gly Arg Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro
        355                 360                 365

His Val Thr Glu Val Glu Gly Thr Ser Phe Leu Thr Gly Ile Ile Ser
    370                 375                 380

Trp Gly Glu Glu Cys Ala Met Lys Gly Lys Tyr Gly Ile Tyr Thr Lys
385                 390                 395                 400

Val Ser Arg Tyr Val Asn Trp Ile Lys Glu Lys Thr Lys Leu Thr Pro
                405                 410                 415

Val Ser Gln Thr Ser Lys Leu Thr Arg Ala Glu Thr Val Phe Pro Asp
            420                 425                 430

Val Asp Ala His Lys Ser Glu Val Ala His Arg Phe Lys Asp Leu Gly
        435                 440                 445

Glu Glu Asn Phe Lys Ala Leu Val Leu Ile Ala Phe Ala Gln Tyr Leu
    450                 455                 460

Gln Gln Cys Pro Phe Glu Asp His Val Lys Leu Val Asn Glu Val Thr
```

```
465                 470                 475                 480

Glu Phe Ala Lys Thr Cys Val Ala Asp Glu Ser Ala Glu Asn Cys Asp
                485                 490                 495

Lys Ser Leu His Thr Leu Phe Gly Asp Lys Leu Cys Thr Val Ala Thr
            500                 505                 510

Leu Arg Glu Thr Tyr Gly Glu Met Ala Asp Cys Cys Ala Lys Gln Glu
        515                 520                 525

Pro Glu Arg Asn Glu Cys Phe Leu Gln His Lys Asp Asp Asn Pro Asn
    530                 535                 540

Leu Pro Arg Leu Val Arg Pro Glu Val Asp Val Met Cys Thr Ala Phe
545                 550                 555                 560

His Asp Asn Glu Glu Thr Phe Leu Lys Lys Tyr Leu Tyr Glu Ile Ala
                565                 570                 575

Arg Arg His Pro Tyr Phe Tyr Ala Pro Glu Leu Leu Phe Phe Ala Lys
            580                 585                 590

Arg Tyr Lys Ala Ala Phe Thr Glu Cys Cys Gln Ala Ala Asp Lys Ala
        595                 600                 605

Ala Cys Leu Leu Pro Lys Leu Asp Glu Leu Arg Asp Glu Gly Lys Ala
    610                 615                 620

Ser Ser Ala Lys Gln Arg Leu Lys Cys Ala Ser Leu Gln Lys Phe Gly
625                 630                 635                 640

Glu Arg Ala Phe Lys Ala Trp Ala Val Ala Arg Leu Ser Gln Arg Phe
                645                 650                 655

Pro Lys Ala Glu Phe Ala Glu Val Ser Lys Leu Val Thr Asp Leu Thr
```

660 665 670

Lys Val His Thr Glu Cys Cys His Gly Asp Leu Leu Glu Cys Ala Asp
675 680 685

Asp Arg Ala Asp Leu Ala Lys Tyr Ile Cys Glu Asn Gln Asp Ser Ile
690 695 700

Ser Ser Lys Leu Lys Glu Cys Cys Glu Lys Pro Leu Leu Glu Lys Ser
705 710 715 720

His Cys Ile Ala Glu Val Glu Asn Asp Glu Met Pro Ala Asp Leu Pro
725 730 735

Ser Leu Ala Ala Asp Phe Val Glu Ser Lys Asp Val Cys Lys Asn Tyr
740 745 750

Ala Glu Ala Lys Asp Val Phe Leu Gly Met Phe Leu Tyr Glu Tyr Ala
755 760 765

Arg Arg His Pro Asp Tyr Ser Val Val Leu Leu Leu Arg Leu Ala Lys
770 775 780

Thr Tyr Glu Thr Thr Leu Glu Lys Cys Cys Ala Ala Ala Asp Phe Cys
785 790 795 800

Tyr Ala Lys Val Phe Asp Glu Phe Lys Pro Leu Val Glu Glu Pro Gln
805 810 815

Asn Leu Ile Lys Gln Asn Cys Glu Leu Phe Glu Gln Leu Gly Glu Tyr
820 825 830

Lys Phe Gln Asn Ala Leu Leu Val Arg Tyr Thr Lys Lys Val Pro Gln
835 840 845

Val Ser Thr Pro Thr Leu Val Glu Val Ser Arg Asn Leu Gly Lys Val

```
    850                 855                 860

Gly Ser Lys Cys Cys Lys His Pro Glu Ala Lys Arg Met Pro Cys Ala
865                 870                 875                 880

Glu Asp Tyr Leu Ser Val Val Leu Asn Gln Leu Cys Val Leu His Glu
                885                 890                 895

Lys Thr Pro Val Ser Asp Arg Val Thr Lys Cys Cys Thr Glu Ser Leu
            900                 905                 910

Val Asn Arg Arg Pro Cys Phe Ser Ala Leu Glu Val Asp Glu Thr Tyr
        915                 920                 925

Val Pro Lys Glu Phe Asn Ala Glu Thr Phe Thr Phe His Ala Asp Ile
    930                 935                 940

Cys Thr Leu Ser Glu Lys Glu Arg Gln Ile Lys Lys Gln Thr Ala Leu
945                 950                 955                 960

Val Glu Leu Val Lys His Lys Pro Lys Ala Thr Lys Glu Gln Leu Lys
                965                 970                 975

Ala Val Met Asp Asp Phe Ala Ala Phe Val Glu Lys Cys Cys Lys Ala
            980                 985                 990

Asp Asp Lys Glu Thr Cys Phe Ala  Glu Glu Gly Lys Lys  Leu Val Ala
        995                 1000                1005

Ala Ser  Gln Ala Ala Leu Gly  Leu
    1010                1015
```

<210> 161

<211> 424

<212> PRT

<213> 人工序列

<220>

<223> FIX

<400> 161

Tyr Asn Ser Gly Lys Leu Glu Glu Phe Val Gln Gly Asn Leu Glu Arg
1 5 10 15

Glu Cys Met Glu Glu Lys Cys Ser Phe Glu Glu Ala Arg Glu Val Phe
20 25 30

Glu Asn Thr Glu Arg Thr Thr Glu Phe Trp Lys Gln Tyr Val Asp Gly
35 40 45

Asp Gln Cys Glu Ser Asn Pro Cys Leu Asn Gly Gly Ser Cys Lys Asp
50 55 60

Asp Ile Asn Ser Tyr Glu Cys Trp Cys Pro Phe Gly Phe Glu Gly Lys
65 70 75 80

Asn Cys Glu Leu Asp Val Thr Cys Asn Ile Lys Asn Gly Arg Cys Glu
85 90 95

Gln Phe Cys Lys Asn Ser Ala Asp Asn Lys Val Val Cys Ser Cys Thr
100 105 110

Glu Gly Tyr Arg Leu Ala Glu Asn Gln Lys Ser Cys Glu Pro Ala Val
115 120 125

Pro Phe Pro Cys Gly Arg Val Ser Val Ser Gln Thr Ser Lys Leu Thr
130 135 140

Arg Ala Glu Thr Val Phe Pro Asp Val Asp Tyr Val Asn Ser Thr Glu
145 150 155 160

```
Ala Glu Thr Ile Leu Asp Asn Ile Thr Gln Ser Thr Gln Ser Phe Asn
                165                 170                 175

Asp Phe Thr Arg Val Val Gly Gly Glu Asp Ala Lys Pro Gly Gln Phe
            180                 185                 190

Pro Trp Gln Val Val Leu Asn Gly Lys Val Asp Ala Phe Cys Gly Gly
        195                 200                 205

Ser Ile Val Asn Glu Lys Trp Ile Val Thr Ala Ala His Cys Val Glu
    210                 215                 220

Thr Gly Val Lys Ile Thr Val Val Ala Gly Glu His Asn Ile Glu Glu
225                 230                 235                 240

Thr Glu His Thr Glu Gln Lys Arg Asn Val Ile Arg Ile Ile Pro His
                245                 250                 255

His Asn Tyr Asn Ala Ala Ile Asn Lys Tyr Asn His Asp Ile Ala Leu
            260                 265                 270

Leu Glu Leu Asp Glu Pro Leu Val Leu Asn Ser Tyr Val Thr Pro Ile
        275                 280                 285

Cys Ile Ala Asp Lys Glu Tyr Thr Asn Ile Phe Leu Lys Phe Gly Ser
    290                 295                 300

Gly Tyr Val Ser Gly Trp Gly Arg Val Phe His Lys Gly Arg Ser Ala
305                 310                 315                 320

Leu Val Leu Gln Tyr Leu Arg Val Pro Leu Val Asp Arg Ala Thr Cys
                325                 330                 335

Leu Arg Ser Thr Lys Phe Thr Ile Tyr Asn Asn Met Phe Cys Ala Gly
            340                 345                 350
```

Phe His Glu Gly Gly Arg Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro
355 360 365

His Val Thr Glu Val Glu Gly Thr Ser Phe Leu Thr Gly Ile Ile Ser
370 375 380

Trp Gly Glu Glu Cys Ala Met Lys Gly Lys Tyr Gly Ile Tyr Thr Lys
385 390 395 400

Val Ser Arg Tyr Val Asn Trp Ile Lys Glu Lys Thr Lys Leu Thr Pro
405 410 415

Val Ser Gln Thr Ser Lys Leu Thr
420

<210> 162
<211> 9
<212> PRT
<213> 人工序列

<220>
<223> 連接子

<400> 162
Arg Ala Glu Thr Val Phe Pro Asp Val
1 5

<210> 163
<211> 585
<212> PRT
<213> 人工序列

<220>
<223> 白蛋白

<400> 163

Asp Ala His Lys Ser Glu Val Ala His Arg Phe Lys Asp Leu Gly Glu
1 5 10 15

Glu Asn Phe Lys Ala Leu Val Leu Ile Ala Phe Ala Gln Tyr Leu Gln
20 25 30

Gln Cys Pro Phe Glu Asp His Val Lys Leu Val Asn Glu Val Thr Glu
35 40 45

Phe Ala Lys Thr Cys Val Ala Asp Glu Ser Ala Glu Asn Cys Asp Lys
50 55 60

Ser Leu His Thr Leu Phe Gly Asp Lys Leu Cys Thr Val Ala Thr Leu
65 70 75 80

Arg Glu Thr Tyr Gly Glu Met Ala Asp Cys Cys Ala Lys Gln Glu Pro
85 90 95

Glu Arg Asn Glu Cys Phe Leu Gln His Lys Asp Asp Asn Pro Asn Leu
100 105 110

Pro Arg Leu Val Arg Pro Glu Val Asp Val Met Cys Thr Ala Phe His
115 120 125

Asp Asn Glu Glu Thr Phe Leu Lys Lys Tyr Leu Tyr Glu Ile Ala Arg
130 135 140

Arg His Pro Tyr Phe Tyr Ala Pro Glu Leu Leu Phe Phe Ala Lys Arg
145 150 155 160

Tyr Lys Ala Ala Phe Thr Glu Cys Cys Gln Ala Ala Asp Lys Ala Ala
165 170 175

Cys Leu Leu Pro Lys Leu Asp Glu Leu Arg Asp Glu Gly Lys Ala Ser

180 185 190

Ser Ala Lys Gln Arg Leu Lys Cys Ala Ser Leu Gln Lys Phe Gly Glu
195 200 205

Arg Ala Phe Lys Ala Trp Ala Val Ala Arg Leu Ser Gln Arg Phe Pro
210 215 220

Lys Ala Glu Phe Ala Glu Val Ser Lys Leu Val Thr Asp Leu Thr Lys
225 230 235 240

Val His Thr Glu Cys Cys His Gly Asp Leu Leu Glu Cys Ala Asp Asp
245 250 255

Arg Ala Asp Leu Ala Lys Tyr Ile Cys Glu Asn Gln Asp Ser Ile Ser
260 265 270

Ser Lys Leu Lys Glu Cys Cys Glu Lys Pro Leu Leu Glu Lys Ser His
275 280 285

Cys Ile Ala Glu Val Glu Asn Asp Glu Met Pro Ala Asp Leu Pro Ser
290 295 300

Leu Ala Ala Asp Phe Val Glu Ser Lys Asp Val Cys Lys Asn Tyr Ala
305 310 315 320

Glu Ala Lys Asp Val Phe Leu Gly Met Phe Leu Tyr Glu Tyr Ala Arg
325 330 335

Arg His Pro Asp Tyr Ser Val Val Leu Leu Leu Arg Leu Ala Lys Thr
340 345 350

Tyr Glu Thr Thr Leu Glu Lys Cys Cys Ala Ala Ala Asp Pro His Glu
355 360 365

Cys Tyr Ala Lys Val Phe Asp Glu Phe Lys Pro Leu Val Glu Glu Pro

370 375 380

Gln Asn Leu Ile Lys Gln Asn Cys Glu Leu Phe Glu Gln Leu Gly Glu
385 390 395 400

Tyr Lys Phe Gln Asn Ala Leu Leu Val Arg Tyr Thr Lys Lys Val Pro
405 410 415

Gln Val Ser Thr Pro Thr Leu Val Glu Val Ser Arg Asn Leu Gly Lys
420 425 430

Val Gly Ser Lys Cys Cys Lys His Pro Glu Ala Lys Arg Met Pro Cys
435 440 445

Ala Glu Asp Tyr Leu Ser Val Val Leu Asn Gln Leu Cys Val Leu His
450 455 460

Glu Lys Thr Pro Val Ser Asp Arg Val Thr Lys Cys Cys Thr Glu Ser
465 470 475 480

Leu Val Asn Arg Arg Pro Cys Phe Ser Ala Leu Glu Val Asp Glu Thr
485 490 495

Tyr Val Pro Lys Glu Phe Asn Ala Glu Thr Phe Thr Phe His Ala Asp
500 505 510

Ile Cys Thr Leu Ser Glu Lys Glu Arg Gln Ile Lys Lys Gln Thr Ala
515 520 525

Leu Val Glu Leu Val Lys His Lys Pro Lys Ala Thr Lys Glu Gln Leu
530 535 540

Lys Ala Val Met Asp Asp Phe Ala Ala Phe Val Glu Lys Cys Cys Lys
545 550 555 560

Ala Asp Asp Lys Glu Thr Cys Phe Ala Glu Glu Gly Lys Lys Leu Val

565 570 575

Ala Ala Ser Gln Ala Ala Leu Gly Leu
580 585

<210> 164
<211> 763
<212> PRT
<213> 人工序列

<220>
<223> FIX(XTEN)-Fc

<400> 164
Met Gln Arg Val Asn Met Ile Met Ala Glu Ser Pro Gly Leu Ile Thr
1 5 10 15

Ile Cys Leu Leu Gly Tyr Leu Leu Ser Ala Glu Cys Thr Val Phe Leu
20 25 30

Asp His Glu Asn Ala Asn Lys Ile Leu Asn Arg Pro Lys Arg Tyr Asn
35 40 45

Ser Gly Lys Leu Glu Glu Phe Val Gln Gly Asn Leu Glu Arg Glu Cys
50 55 60

Met Glu Glu Lys Cys Ser Phe Glu Glu Ala Arg Glu Val Phe Glu Asn
65 70 75 80

Thr Glu Arg Thr Thr Glu Phe Trp Lys Gln Tyr Val Asp Gly Asp Gln
85 90 95

Cys Glu Ser Asn Pro Cys Leu Asn Gly Gly Ser Cys Lys Asp Asp Ile
100 105 110

```
Asn Ser Tyr Glu Cys Trp Cys Pro Phe Gly Phe Glu Gly Lys Asn Cys
        115                 120                 125

Glu Leu Asp Val Thr Cys Asn Ile Lys Asn Gly Arg Cys Glu Gln Phe
    130                 135                 140

Cys Lys Asn Ser Ala Asp Asn Lys Val Val Cys Ser Cys Thr Glu Gly
145                 150                 155                 160

Tyr Arg Leu Ala Glu Asn Gln Lys Ser Cys Glu Pro Ala Val Pro Phe
                165                 170                 175

Pro Cys Gly Arg Val Ser Val Ser Gln Thr Ser Lys Leu Thr Arg Ala
            180                 185                 190

Glu Thr Val Phe Pro Asp Val Asp Tyr Val Asn Ser Thr Glu Ala Glu
        195                 200                 205

Thr Ile Leu Asp Gly Pro Ser Pro Gly Ser Pro Thr Ser Thr Glu Glu
    210                 215                 220

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro
225                 230                 235                 240

Ala Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
                245                 250                 255

Glu Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro
            260                 265                 270

Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Ala Ser Ser
        275                 280                 285

Asn Ile Thr Gln Ser Thr Gln Ser Phe Asn Asp Phe Thr Arg Val Val
    290                 295                 300
```

```
Gly Gly Glu Asp Ala Lys Pro Gly Gln Phe Pro Trp Gln Val Val Leu
305                 310                 315                 320

Asn Gly Lys Val Asp Ala Phe Cys Gly Gly Ser Ile Val Asn Glu Lys
                325                 330                 335

Trp Ile Val Thr Ala Ala His Cys Val Glu Thr Gly Val Lys Ile Thr
            340                 345                 350

Val Val Ala Gly Glu His Asn Ile Glu Glu Thr Glu His Thr Glu Gln
        355                 360                 365

Lys Arg Asn Val Ile Arg Ile Ile Pro His His Asn Tyr Asn Ala Ala
    370                 375                 380

Ile Asn Lys Tyr Asn His Asp Ile Ala Leu Leu Glu Leu Asp Glu Pro
385                 390                 395                 400

Leu Val Leu Asn Ser Tyr Val Thr Pro Ile Cys Ile Ala Asp Lys Glu
                405                 410                 415

Tyr Thr Asn Ile Phe Leu Lys Phe Gly Ser Gly Tyr Val Ser Gly Trp
            420                 425                 430

Gly Arg Val Phe His Lys Gly Arg Ser Ala Leu Val Leu Gln Tyr Leu
        435                 440                 445

Arg Val Pro Leu Val Asp Arg Ala Thr Cys Leu Leu Ser Thr Lys Phe
    450                 455                 460

Thr Ile Tyr Asn Asn Met Phe Cys Ala Gly Phe His Glu Gly Gly Arg
465                 470                 475                 480

Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro His Val Thr Glu Val Glu
                485                 490                 495
```

```
Gly Thr Ser Phe Leu Thr Gly Ile Ile Ser Trp Gly Glu Glu Cys Ala
            500                 505                 510

Met Lys Gly Lys Tyr Gly Ile Tyr Thr Lys Val Ser Arg Tyr Val Asn
        515                 520                 525

Trp Ile Lys Glu Lys Thr Lys Leu Thr Asp Lys Thr His Thr Cys Pro
    530                 535                 540

Pro Cys Pro Ala Pro Glu Leu Leu Gly Gly Pro Ser Val Phe Leu Phe
545                 550                 555                 560

Pro Pro Lys Pro Lys Asp Thr Leu Met Ile Ser Arg Thr Pro Glu Val
                565                 570                 575

Thr Cys Val Val Val Asp Val Ser His Glu Asp Pro Glu Val Lys Phe
            580                 585                 590

Asn Trp Tyr Val Asp Gly Val Glu Val His Asn Ala Lys Thr Lys Pro
        595                 600                 605

Arg Glu Glu Gln Tyr Asn Ser Thr Tyr Arg Val Val Ser Val Leu Thr
    610                 615                 620

Val Leu His Gln Asp Trp Leu Asn Gly Lys Glu Tyr Lys Cys Lys Val
625                 630                 635                 640

Ser Asn Lys Ala Leu Pro Ala Pro Ile Glu Lys Thr Ile Ser Lys Ala
                645                 650                 655

Lys Gly Gln Pro Arg Glu Pro Gln Val Tyr Thr Leu Pro Pro Ser Arg
            660                 665                 670

Asp Glu Leu Thr Lys Asn Gln Val Ser Leu Thr Cys Leu Val Lys Gly
        675                 680                 685
```

Phe Tyr Pro Ser Asp Ile Ala Val Glu Trp Glu Ser Asn Gly Gln Pro
690 695 700

Glu Asn Asn Tyr Lys Thr Thr Pro Pro Val Leu Asp Ser Asp Gly Ser
705 710 715 720

Phe Phe Leu Tyr Ser Lys Leu Thr Val Asp Lys Ser Arg Trp Gln Gln
725 730 735

Gly Asn Val Phe Ser Cys Ser Val Met His Glu Ala Leu His Asn His
740 745 750

Tyr Thr Gln Lys Ser Leu Ser Leu Ser Pro Gly
755 760

<210> 165
<211> 712
<212> PRT
<213> 人工序列

<220>
<223> FIX-FXIa-AE288

<400> 165
Tyr Asn Ser Gly Lys Leu Glu Glu Phe Val Gln Gly Asn Leu Glu Arg
1 5 10 15

Glu Cys Met Glu Glu Lys Cys Ser Phe Glu Glu Ala Arg Glu Val Phe
20 25 30

Glu Asn Thr Glu Arg Thr Thr Glu Phe Trp Lys Gln Tyr Val Asp Gly
35 40 45

Asp Gln Cys Glu Ser Asn Pro Cys Leu Asn Gly Gly Ser Cys Lys Asp
50 55 60

```
Asp Ile Asn Ser Tyr Glu Cys Trp Cys Pro Phe Gly Phe Glu Gly Lys
65                  70                  75                  80

Asn Cys Glu Leu Asp Val Thr Cys Asn Ile Lys Asn Gly Arg Cys Glu
                85                  90                  95

Gln Phe Cys Lys Asn Ser Ala Asp Asn Lys Val Val Cys Ser Cys Thr
            100                 105                 110

Glu Gly Tyr Arg Leu Ala Glu Asn Gln Lys Ser Cys Glu Pro Ala Val
        115                 120                 125

Pro Phe Pro Cys Gly Arg Val Ser Val Ser Gln Thr Ser Lys Leu Thr
    130                 135                 140

Arg Ala Glu Thr Val Phe Pro Asp Val Asp Tyr Val Asn Ser Thr Glu
145                 150                 155                 160

Ala Glu Thr Ile Leu Asp Asn Ile Thr Gln Ser Thr Gln Ser Phe Asn
                165                 170                 175

Asp Phe Thr Arg Val Val Gly Gly Glu Asp Ala Lys Pro Gly Gln Phe
            180                 185                 190

Pro Trp Gln Val Val Leu Asn Gly Lys Val Asp Ala Phe Cys Gly Gly
        195                 200                 205

Ser Ile Val Asn Glu Lys Trp Ile Val Thr Ala Ala His Cys Val Glu
    210                 215                 220

Thr Gly Val Lys Ile Thr Val Val Ala Gly Glu His Asn Ile Glu Glu
225                 230                 235                 240

Thr Glu His Thr Glu Gln Lys Arg Asn Val Ile Arg Ile Ile Pro His
                245                 250                 255
```

```
His Asn Tyr Asn Ala Ala Ile Asn Lys Tyr Asn His Asp Ile Ala Leu
            260                 265                 270

Leu Glu Leu Asp Glu Pro Leu Val Leu Asn Ser Tyr Val Thr Pro Ile
        275                 280                 285

Cys Ile Ala Asp Lys Glu Tyr Thr Asn Ile Phe Leu Lys Phe Gly Ser
    290                 295                 300

Gly Tyr Val Ser Gly Trp Gly Arg Val Phe His Lys Gly Arg Ser Ala
305                 310                 315                 320

Leu Val Leu Gln Tyr Leu Arg Val Pro Leu Val Asp Arg Ala Thr Cys
                325                 330                 335

Leu Arg Ser Thr Lys Phe Thr Ile Tyr Asn Asn Met Phe Cys Ala Gly
            340                 345                 350

Phe His Glu Gly Gly Arg Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro
        355                 360                 365

His Val Thr Glu Val Glu Gly Thr Ser Phe Leu Thr Gly Ile Ile Ser
    370                 375                 380

Trp Gly Glu Glu Cys Ala Met Lys Gly Lys Tyr Gly Ile Tyr Thr Lys
385                 390                 395                 400

Val Ser Arg Tyr Val Asn Trp Ile Lys Glu Lys Thr Lys Leu Thr Gly
                405                 410                 415

Lys Leu Thr Arg Ala Glu Thr Gly Gly Thr Ser Glu Ser Ala Thr Pro
            420                 425                 430

Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro
        435                 440                 445
```

```
Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro
    450                 455                 460

Ala Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
465                 470                 475                 480

Glu Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro
                485                 490                 495

Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Glu
            500                 505                 510

Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly
        515                 520                 525

Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro
    530                 535                 540

Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Ser Pro Ala
545                 550                 555                 560

Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu
                565                 570                 575

Gly Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro
            580                 585                 590

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Glu
        595                 600                 605

Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly
    610                 615                 620

Ser Glu Thr Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro
625                 630                 635                 640
```

Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Thr
645 650 655

Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu
660 665 670

Gly Ser Ala Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro
675 680 685

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr
690 695 700

Glu Pro Ser Glu Gly Ser Ala Pro
705 710

<210> 166
<211> 944
<212> PRT
<213> 人工序列

<220>
<223> FIX-Fc-Fc

<400> 166
Met Gln Arg Val Asn Met Ile Met Ala Glu Ser Pro Gly Leu Ile Thr
1 5 10 15

Ile Cys Leu Leu Gly Tyr Leu Leu Ser Ala Glu Cys Thr Val Phe Leu
20 25 30

Asp His Glu Asn Ala Asn Lys Ile Leu Asn Arg Pro Lys Arg Tyr Asn
35 40 45

Ser Gly Lys Leu Glu Glu Phe Val Gln Gly Asn Leu Glu Arg Glu Cys

```
    50                  55                  60

Met Glu Glu Lys Cys Ser Phe Glu Glu Ala Arg Glu Val Phe Glu Asn
65                  70                  75                  80

Thr Glu Arg Thr Thr Glu Phe Trp Lys Gln Tyr Val Asp Gly Asp Gln
                85                  90                  95

Cys Glu Ser Asn Pro Cys Leu Asn Gly Gly Ser Cys Lys Asp Asp Ile
            100                 105                 110

Asn Ser Tyr Glu Cys Trp Cys Pro Phe Gly Phe Glu Gly Lys Asn Cys
        115                 120                 125

Glu Leu Asp Val Thr Cys Asn Ile Lys Asn Gly Arg Cys Glu Gln Phe
    130                 135                 140

Cys Lys Asn Ser Ala Asp Asn Lys Val Val Cys Ser Cys Thr Glu Gly
145                 150                 155                 160

Tyr Arg Leu Ala Glu Asn Gln Lys Ser Cys Glu Pro Ala Val Pro Phe
                165                 170                 175

Pro Cys Gly Arg Val Ser Val Ser Gln Thr Ser Lys Leu Thr Arg Ala
            180                 185                 190

Glu Thr Val Phe Pro Asp Val Asp Tyr Val Asn Ser Thr Glu Ala Glu
        195                 200                 205

Thr Ile Leu Asp Asn Ile Thr Gln Ser Thr Gln Ser Phe Asn Asp Phe
    210                 215                 220

Thr Arg Val Val Gly Gly Glu Asp Ala Lys Pro Gly Gln Phe Pro Trp
225                 230                 235                 240

Gln Val Val Leu Asn Gly Lys Val Asp Ala Phe Cys Gly Gly Ser Ile
```

245 250 255

Val Asn Glu Lys Trp Ile Val Thr Ala Ala His Cys Val Glu Thr Gly
260 265 270

Val Lys Ile Thr Val Val Ala Gly Glu His Asn Ile Glu Glu Thr Glu
275 280 285

His Thr Glu Gln Lys Arg Asn Val Ile Arg Ile Ile Pro His His Asn
290 295 300

Tyr Asn Ala Ala Ile Asn Lys Tyr Asn His Asp Ile Ala Leu Leu Glu
305 310 315 320

Leu Asp Glu Pro Leu Val Leu Asn Ser Tyr Val Thr Pro Ile Cys Ile
325 330 335

Ala Asp Lys Glu Tyr Thr Asn Ile Phe Leu Lys Phe Gly Ser Gly Tyr
340 345 350

Val Ser Gly Trp Gly Arg Val Phe His Lys Gly Arg Ser Ala Leu Val
355 360 365

Leu Gln Tyr Leu Arg Val Pro Leu Val Asp Arg Ala Thr Cys Leu Arg
370 375 380

Ser Thr Lys Phe Thr Ile Tyr Asn Asn Met Phe Cys Ala Gly Phe His
385 390 395 400

Glu Gly Gly Arg Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro His Val
405 410 415

Thr Glu Val Glu Gly Thr Ser Phe Leu Thr Gly Ile Ile Ser Trp Gly
420 425 430

Glu Glu Cys Ala Met Lys Gly Lys Tyr Gly Ile Tyr Thr Lys Val Ser

435 440 445

Arg Tyr Val Asn Trp Ile Lys Glu Lys Thr Lys Leu Thr Asp Lys Thr
450 455 460

His Thr Cys Pro Pro Cys Pro Ala Pro Glu Leu Leu Gly Gly Pro Ser
465 470 475 480

Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu Met Ile Ser Arg
485 490 495

Thr Pro Glu Val Thr Cys Val Val Val Asp Val Ser His Glu Asp Pro
500 505 510

Glu Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu Val His Asn Ala
515 520 525

Lys Thr Lys Pro Arg Glu Glu Gln Tyr Asn Ser Thr Tyr Arg Val Val
530 535 540

Ser Val Leu Thr Val Leu His Gln Asp Trp Leu Asn Gly Lys Glu Tyr
545 550 555 560

Lys Cys Lys Val Ser Asn Lys Ala Leu Pro Ala Pro Ile Glu Lys Thr
565 570 575

Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln Val Tyr Thr Leu
580 585 590

Pro Pro Ser Arg Asp Glu Leu Thr Lys Asn Gln Val Ser Leu Thr Cys
595 600 605

Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val Glu Trp Glu Ser
610 615 620

Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro Pro Val Leu Asp

625 630 635 640

Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr Val Asp Lys Ser
645 650 655

Arg Trp Gln Gln Gly Asn Val Phe Ser Cys Ser Val Met His Glu Ala
660 665 670

Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser Pro Gly Lys
675 680 685

Arg Arg Arg Arg Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
690 695 700

Gly Gly Gly Ser Gly Gly Gly Gly Ser Arg Arg Arg Arg Asp Lys Thr
705 710 715 720

His Thr Cys Pro Pro Cys Pro Ala Pro Glu Leu Leu Gly Gly Pro Ser
725 730 735

Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu Met Ile Ser Arg
740 745 750

Thr Pro Glu Val Thr Cys Val Val Val Asp Val Ser His Glu Asp Pro
755 760 765

Glu Val Lys Phe Asn Trp Tyr Val Asp Gly Val Glu Val His Asn Ala
770 775 780

Lys Thr Lys Pro Arg Glu Glu Gln Tyr Asn Ser Thr Tyr Arg Val Val
785 790 795 800

Ser Val Leu Thr Val Leu His Gln Asp Trp Leu Asn Gly Lys Glu Tyr
805 810 815

Lys Cys Lys Val Ser Asn Lys Ala Leu Pro Ala Pro Ile Glu Lys Thr

820 825 830

Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln Val Tyr Thr Leu
835 840 845

Pro Pro Ser Arg Asp Glu Leu Thr Lys Asn Gln Val Ser Leu Thr Cys
850 855 860

Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val Glu Trp Glu Ser
865 870 875 880

Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro Pro Val Leu Asp
885 890 895

Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr Val Asp Lys Ser
900 905 910

Arg Trp Gln Gln Gly Asn Val Phe Ser Cys Ser Val Met His Glu Ala
915 920 925

Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser Pro Gly Lys
930 935 940

<210> 167
<211> 383
<212> DNA
<213> 人工序列

<220>
<223> 細小病毒B19 KY940273.1

<400> 167
ccaaatcaga tgccgccggt cgccgccggt aggcgggact tccggtacaa gatggcggac 60
aattacgtca tttcctgtga cgtcatttcc tgtgacgtca cttccggtgg gcgggacttc 120
cggaattagg gttggctctg ggccagcttg cttggggttg ccttgacact aagacaagcg 180

gcgcgccgct tgatcttagt ggcacgtcaa ccccaagcgc tggcccagag ccaaccctaa 240
ttccggaagt cccgcccacc ggaagtgacg tcacaggaaa tgacgtcaca ggaaatgacg 300
taattgtccg ccatcttgta ccggaagtcc cgcctaccgg cggcgaccgg cggcatctga 360
tttggtgtct tcttttaaat ttt 383

<210> 168
<211> 248
<212> DNA
<213> 人工序列

<220>
<223> 細小病毒B19 d135

<400> 168
ctctgggcca gcttgcttgg ggttgccttg acactaagac aagcggcgcg ccgcttgatc 60
ttagtggcac gtcaacccca agcgctggcc cagagccaac cctaattccg gaagtcccgc 120
ccaccggaag tgacgtcaca ggaaatgacg tcacaggaaa tgacgtaatt gtccgccatc 180
ttgtaccgga agtcccgcct accggcggcg accggcggca tctgatttgg tgtcttcttt 240
taaatttt 248

<210> 169
<211> 129
<212> DNA
<213> 人工序列

<220>
<223> 細小病毒v1 AAV2

<400> 169
cggcgcgccg cttgatctta gtggcacgtc aaccagcgct ggcccagagc caaccctaat 60
tccggaagtc ctcagtccgc catcttgccc gcctaccggc ggcgaccggc ggcatcattt 120
ggtgttctt 129

<210> 170
<211> 113
<212> DNA
<213> 人工序列

<220>
<223> 細小病毒V2

<400> 170
ctctgggcca gcttgcttgg ggttgccttg acactaagac aagcggcgcg ccgcttgatc 60
ttagtggcac gtcaacccca agcgctggcc cagagtgtct tcttttaaat ttt 113

<210> 171
<211> 340
<212> DNA
<213> 人工序列

<220>
<223> 細小病毒V3

<400> 171
caaatcagat gccgccggtc gccgccggta ggcgggactt ccggtacaag atggcggaca 60
attacgtcat ttcctgtgac gtatttcctg tgacgtactt ccggtggcgg gacttccgga 120
attttggctc tgggccagct tgcttggggt tgccttgacc aagcgcgcgc cgcttgatca 180
ccccaagcgc tggcccagag ccacctaacc ggaagtcccc ccaccggaag tgacgtcaca 240
ggaaagacgt cacaggaagt aattgtccgc catcttgtac cggaagtccc gcaccggcgg 300
cgaccggcgg catctgattt ggtgtcttct tttaaatttt 340

<210> 172
<211> 444
<212> DNA
<213> 人工序列

<220>

<223> 細小病毒U25749.1

<400> 172

ctcattggag ggttcgttcg ttcgaaccag ccaatcaggg gaggggggaag tgacgcaagt 60

tccggtcaca tgcttccggt gacgcacatc cggtgacgta gttccggtca cgtgcttcct 120

gtcacgtgtt tccggtcacg tgacttccgg tcatgtgact tccggtgacg tgtttccggc 180

tgttaggttg accacgcgca tgccgcgcgg tcagcccaat agttaagccg gaaacacgtc 240

accggaagtc acatgaccgg aagtcacgtg accggaaaca cgtgacagga agcacgtgac 300

cggaactacg tcaccggatg tgcgtcaccg gaagcatgtg accggaactt gcgtcacttc 360

cccctcccct gattggctgg ttcgaacgaa cgaaccctcc aatgagactc aaggacaaga 420

ggatattttg cgcgccagga agtg 444

<210> 173

<211> 282

<212> DNA

<213> 人工序列

<220>

<223> 細小病毒GPV AAV2 V1

<400> 173

cggtgacgtg tttccggctg ttaggttgac cacgcgcatg ccgcgcggtc agcccaatag 60

ttaagccgga aacacgtcac cggaagtcac atgaccggaa gtcacgtgac cggaaacacg 120

tgacaggaag cacgtgaccg gaactacgtc accggatgtg cgtcaccgga agcatgtgac 180

cggaacttgc gtcacttccc cctcccctga ttggctggtt cgaacgaacg aaccctccaa 240

tgagactcaa ggacaagagg atattttgcg cgccaggaag tg 282

<210> 174

<211> 145

<212> DNA

<213> 人工序列

<220>

<223> 細小病毒GPV AAV2 V1

<400> 174
ttgaccacgc gcatgccgcg cggtcagccc aatagttaag ccgggtgacc acacgtgaca 60
ggaagcacgg gatgtgcgtc accggaagca gtgaccgggc tggttcgaac gaacgaaccc 120
tccaactcaa ggacaagagg atatt 145

<210> 175
<211> 120
<212> DNA
<213> 人工序列

<220>
<223> 細小病毒GPV V2

<400> 175
cggtgacgtg tttccggctg ttaggttgac cacgcgcatg ccgcgcggtc agcccaatag 60
ttaagccgga aacacgtcac cgactcaagg acaagaggat attttgcgcg ccaggaagtg 120

<210> 176
<211> 102
<212> DNA
<213> 人工序列

<220>
<223> 細小病毒GPV V3

<400> 176
gggaacaatc aggggaagtg accggtgacg tcatgtaact tgcgtcactt cccgttcgaa 60
cgaacgaacg agactcaagg acaagaggcg cgccaggaag tg 102

<210> 177
<211> 145
<212> DNA

<213> 人工序列

<220>

<223> 細小病毒AAV2 NC_001401.2

<400> 177

ttggccactc cctctctgcg cgctcgctcg ctcactgagg ccgggcgacc aaaggtcgcc 60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgcag agagggagtg 120

gccaactcca tcactagggg ttcct 145

<210> 178

<211> 130

<212> DNA

<213> 人工序列

<220>

<223> 細小病毒Gtx

<400> 178

ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgaccttt 60

ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggccaa ctccatcact 120

aggggttcct 130

<210> 179

<211> 6762

<212> DNA

<213> 人工序列

<220>

<223> B19-FVIII建構體

<400> 179

ctctgggcca gcttgcttgg ggttgccttg acactaagac aagcggcgcg ccgcttgatc 60

ttagtggcac gtcaacccca agcgctggcc cagagccaac cctaattccg gaagtcccgc 120

ccaccggaag tgacgtcaca ggaaatgacg tcacaggaaa tgacgtaatt gtccgccatc 180
ttgtaccgga agtcccgcct accggcggcg accggcggca tctgatttgg tgtcttcttt 240
taaattttgc ggcaattcag tcgataacta taacggtcct aaggtagcga tttaaatacg 300
cgctctctta aggtagcccc gggacgcgtc aattgagatc tggatccggt accgaattcg 360
cggccgcctc gacgactagc gtttaattaa acgcgtgtct gtctgcacat ttcgtagagc 420
gagtgttccg atactctaat ctccctaggc aaggttcata tttgtgtagg ttacttattc 480
tccttttgtt gactaagtca ataatcagaa tcagcaggtt tggagtcagc ttggcaggga 540
tcagcagcct gggttggaag gaggggtat aaaagcccct tcaccaggag aagccgtcac 600
acagatccac aagctcctga ggtaagtgcc gtgtgtggtt cccgcgggcc tggcctcttt 660
acgggttatg gcccttgcgt gccttgaatt actgacactg acatccactt tttcttttc 720
tccacagcta gcgccaccat gcagattgag ctgtccactt gtttcttcct gtgcctcctg 780
cgcttctgtt tctccgccac tcgccggtac taccttggag ccgtggagct ttcatgggac 840
tacatgcaga gcgacctggg cgaactcccc gtggatgcca gattcccccc ccgcgtgcca 900
aagtccttcc cctttaacac ctccgtggtg tacaagaaaa ccctctttgt cgagttcact 960
gaccacctgt tcaacatcgc caagccgcgc ccaccttgga tgggcctcct gggaccgacc 1020
attcaagctg aagtgtacga caccgtggtg atcaccctga agaacatggc gtcccacccc 1080
gtgtccctgc atgcggtcgg agtgtcctac tggaaggcct ccgaaggagc tgagtacgac 1140
gaccagacta gccagcggga aaaggaggac gataaagtgt tcccgggcgg ctcgcatact 1200
tacgtgtggc aagtcctgaa ggaaaacgga cctatggcat ccgatcctct gtgcctgact 1260
tactcctacc tttcccatgt ggacctcgtg aaggacctga acagcgggct gattggtgca 1320
cttctcgtgt gccgcgaagg ttcgctcgct aaggaaaaga cccagaccct ccataagttc 1380
atccttttgt tcgctgtgtt cgatgaagga aagtcatggc attccgaaac taagaactcg 1440
ctgatgcagg accgggatgc cgcctcagcc cgcgcctggc ctaaaatgca tacagtcaac 1500
ggatacgtga atcggtcact gcccgggctc atcggttgtc acagaaagtc cgtgtactgg 1560
cacgtcatcg gcatgggcac tacgcctgaa gtgcactcca tcttcctgga agggcacacc 1620
ttcctcgtgc gcaaccaccg ccaggcctct ctggaaatct ccccgattac ctttctgacc 1680
gcccagactc tgctcatgga cctggggcag ttccttctct tctgccacat ctccagccat 1740
cagcacgacg gaatggaggc ctacgtgaag gtggactcat gcccggaaga acctcagttg 1800
cggatgaaga acaacgagga ggccgaggac tatgacgacg atttgactga ctccgagatg 1860
gacgtcgtgc ggttcgatga cgacaacagc cccagcttca tccagattcg cagcgtggcc 1920
aagaagcacc ccaaaacctg ggtgcactac atcgcggccg aggaagaaga ttgggactac 1980
gccccgttgg tgctggcacc cgatgaccgg tcgtacaagt cccagtatct gaacaatggt 2040
ccgcagcgga ttggcagaaa gtacaagaaa gtgcggttca tggcgtacac tgacgaaacg 2100
tttaagaccc gggaggccat tcaacatgag agcggcattc tgggaccact gctgtacgga 2160
gaggtcggcg ataccctgct catcatcttc aaaaaccagg cctcccggcc ttacaacatc 2220
taccctcacg gaatcaccga cgtgcggcca ctctactcgc ggcgcctgcc gaagggcgtc 2280

aagcacctga aagacttccc tatcctgccg ggcgaaatct tcaagtataa gtggaccgtc 2340

accgtggagg acgggcccac caagagcgat cctaggtgtc tgactcggta ctactccagc 2400

ttcgtgaaca tggaacggga cctggcatcg ggactcattg gaccgctgct gatctgctac 2460

aaagagtcgg tggatcaacg cggcaaccag atcatgtccg acaagcgcaa cgtgatcctg 2520

ttctccgtgt ttgatgaaaa cagatcctgg tacctcactg aaaacatcca gaggttcctc 2580

ccaaaccccg caggagtgca actggaggac cctgagtttc aggcctcgaa tatcatgcac 2640

tcgattaacg gttacgtgtt cgactcgctg caactgagcg tgtgcctcca tgaagtcgct 2700

tactggtaca ttctgtccat cggcgcccag actgacttcc tgagcgtgtt cttttccggt 2760

tacaccttta agcacaagat ggtgtacgaa gataccctga ccctgttccc tttctccggc 2820

gaaacggtgt tcatgtcgat ggagaacccg ggtctgtgga ttctgggatg ccacaacagc 2880

gactttcgga accgcggaat gactgccctg ctgaaggtgt cctcatgcga caagaacacc 2940

ggagactact acgaggactc ctacgaggat atctcagcct acctcctgtc caagaacaac 3000

gcgatcgagc cgcgcagctt cagccagaac ggcgcgccaa catcagagag cgccaccct 3060

gaaagtggtc ccgggagcga gccagccaca tctgggtcgg aaacgccagg cacaagtgag 3120

tctgcaactc ccgagtccgg acctggctcc gagcctgcca ctagcggctc cgagactccg 3180

ggaacttccg agagcgctac accagaaagc ggacccggaa ccagtaccga acctagcgag 3240

ggctctgctc cgggcagccc agccggctct cctacatcca cggaggaggg cacttccgaa 3300

tccgccaccc cggagtcagg gccaggatct gaacccgcta cctcaggcag tgagacgcca 3360

ggaacgagcg agtccgctac accggagagt gggccaggga gccctgctgg atctcctacg 3420

tccactgagg aagggtcacc agcgggctcg cccaccagca ctgaagaagg tgcctcgagc 3480

ccgcctgtgc tgaagaggca ccagcgagaa attacccgga ccaccctcca atcggatcag 3540

gaggaaatcg actacgacga caccatctcg gtggaaatga agaaggaaga tttcgatatc 3600

tacgacgagg acgaaaatca gtcccctcgc tcattccaaa agaaaactag acactacttt 3660

atcgccgcgg tggaaagact gtgggactat ggaatgtcat ccagccctca cgtccttcgg 3720

aaccgggccc agagcggatc ggtgcctcag ttcaagaaag tggtgttcca ggagttcacc 3780

gacggcagct tcacccagcc gctgtaccgg ggagaactga acgaacacct gggcctgctc 3840

ggtccctaca tccgcgcgga agtggaggat aacatcatgg tgaccttccg taaccaagca 3900

tccagacctt actccttcta ttcctccctg atctcatacg aggaggacca gcgccaaggc 3960

gccgagcccc gcaagaactt cgtcaagccc aacgagacta agacctactt ctggaaggtc 4020

caacaccata tggccccgac caaggatgag tttgactgca aggcctgggc ctacttctcc 4080

gacgtggacc ttgagaagga tgtccattcc ggcctgatcg ggccgctgct cgtgtgtcac 4140

accaacaccc tgaacccagc gcatggacgc caggtcaccg tccaggagtt tgctctgttc 4200

ttcaccattt ttgacgaaac taagtcctgg tacttcaccg agaatatgga gcgaaactgt 4260

agagcgccct gcaatatcca gatggaagat ccgactttca aggagaacta tagattccac 4320

gccatcaacg ggtacatcat ggatactctg ccggggctgg tcatggccca ggatcagagg 4380

attcggtggt acttgctgtc aatgggatcg aacgaaaaca ttcactccat tcacttctcc 4440

ggtcacgtgt tcactgtgcg caagaaggag gagtacaaga tggcgctgta caatctgtac 4500
cccggggtgt tcgaaactgt ggagatgctg ccgtccaagg ccggcatctg gagagtggag 4560
tgcctgatcg gagagcacct ccacgcgggg atgtccaccc tcttcctggt gtactcgaat 4620
aagtgccaga ccccgctggg catggcctcg ggccacatca gagacttcca gatcacagca 4680
agcggacaat acggccaatg ggcgccgaag ctggcccgct tgcactactc cggatcgatc 4740
aacgcatggt ccaccaagga accgttctcg tggattaagg tggacctcct ggcccctatg 4800
attatccacg gaattaagac ccagggcgcc aggcagaagt tctcctccct gtacatctcg 4860
caattcatca tcatgtacag cctggacggg aagaagtggc agacttacag gggaaactcc 4920
accggcaccc tgatggtctt tttcggcaac gtggattcct ccggcattaa gcacaacatc 4980
ttcaacccac cgatcatagc cagatatatt aggctccacc ccactcacta ctcaatccgc 5040
tcaactcttc ggatggaact catggggtgc gacctgaact cctgctccat gccgttgggg 5100
atggaatcaa aggctattag cgacgcccag atcaccgcga gctcctactt cactaacatg 5160
ttcgccacct ggagcccctc caaggccagg ctgcacttgc agggacggtc aaatgcctgg 5220
cggccgcaag tgaacaatcc gaaggaatgg cttcaagtgg atttccaaaa gaccatgaaa 5280
gtgaccggag tcaccaccca gggagtgaag tcccttctga cctcgatgta tgtgaaggag 5340
ttcctgatta gcagcagcca ggacgggcac cagtggaccc tgttcttcca aaacggaaag 5400
gtcaaggtgt tccaggggaa ccaggactcg ttcacacccg tggtgaactc cctggacccc 5460
ccactgctga cgcggtactt gaggattcat cctcagtcct gggtccatca gattgcattg 5520
cgaatggaag tcctgggctg cgaggcccag gacctgtact gaatcagcct gagctcgctg 5580
atcataatca acctctggat tacaaaattt gtgaaagatt gactggtatt cttaactatg 5640
ttgctccttt tacgctatgt ggatacgctg ctttaatgcc tttgtatcat gctattgctt 5700
cccgtatggc tttcattttc tcctccttgt ataaatcctg gttgctgtct ctttatgagg 5760
agttgtggcc cgttgtcagg caacgtggcg tggtgtgcac tgtgtttgct gacgcaaccc 5820
ccactggttg gggcattgcc accacctgtc agctcctttc cgggactttc gctttccccc 5880
tccctattgc cacggcggaa ctcatcgccg cctgccttgc ccgctgctgg acaggggctc 5940
ggctgttggg cactgacaat tccgtggtgt tgtcggggaa atcatcgtcc tttccttggc 6000
tgctcgcctg tgttgccacc tggattctgc gcgggacgtc cttctgctac gtcccttcgg 6060
ccctcaatcc agcggacctt ccttcccgcg gcctgctgcc ggctctgcgg cctcttccgc 6120
gtcttcgcct tcgccctcag acgagtcgga tctccctttg ggccgcctcc ccgctgatca 6180
gcctcgactg tgccttctag ttgccagcca tctgttgttt gcccctcccc cgtgccttcc 6240
ttgaccctgg aaggtgccac tcccactgtc ctttcctaat aaaatgagga aattgcatcg 6300
cattgtctga gtaggtgtca ttctattctg gggggtgggg tggggcagga cagcaagggg 6360
gaggattggg aagacaatag caggcatgct ggggatgcgg tgggctctat ggcttctgag 6420
gcggaaagaa cgggctcgag aagcttctag atatcctctc ttaaggtagc atcgagattt 6480
aaattaggga taacagggta atggcgcggg ccgcaaaatt taaaagaaga caccaaatca 6540
gatgccgccg gtcgccgccg gtaggcggga cttccggtac aagatggcgg acaattacgt 6600

catttcctgt gacgtcattt cctgtgacgt cacttccggt gggcgggact tccggaatta 6660
gggttggctc tgggccagcg cttggggttg acgtgccact aagatcaagc ggcgcgccgc 6720
ttgtcttagt gtcaaggcaa ccccaagcaa gctggcccag ag 6762

<210> 180
<211> 248
<212> DNA
<213> 人工序列

<220>
<223> 5' ITR

<400> 180
ctctgggcca gcttgcttgg ggttgccttg acactaagac aagcggcgcg ccgcttgatc 60
ttagtggcac gtcaacccca agcgctggcc cagagccaac cctaattccg gaagtcccgc 120
ccaccggaag tgacgtcaca ggaaatgacg tcacaggaaa tgacgtaatt gtccgccatc 180
ttgtaccgga agtcccgcct accggcggcg accggcggca tctgatttgg tgtcttcttt 240
taaatttt 248

<210> 181
<211> 248
<212> DNA
<213> 人工序列

<220>
<223> 3' ITR

<400> 181
aaaatttaaa agaagacacc aaatcagatg ccgccggtcg ccgccggtag gcgggacttc 60
cggtacaaga tggcggacaa ttacgtcatt tcctgtgacg tcatttcctg tgacgtcact 120
tccggtgggc gggacttccg gaattagggt tggctctggg ccagcgcttg gggttgacgt 180
gccactaaga tcaagcggcg cgccgcttgt cttagtgtca aggcaacccc aagcaagctg 240
gcccagag 248

<210> 182

<211> 6830

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸序列

<400> 182

cggtgacgtg tttccggctg ttaggttgac cacgcgcatg ccgcgcggtc agcccaatag 60

ttaagccgga aacacgtcac cggaagtcac atgaccggaa gtcacgtgac cggaaacacg 120

tgacaggaag cacgtgaccg gaactacgtc accggatgtg cgtcaccgga agcatgtgac 180

cggaacttgc gtcacttccc cctcccctga ttggctggtt cgaacgaacg aaccctccaa 240

tgagactcaa ggacaagagg atattttgcg cgccaggaag tggcggcaat tcagtcgata 300

actataacgg tcctaaggta gcgatttaaa tacgcgctct cttaaggtag ccccgggacg 360

cgtcaattga gatctggatc cggtaccgaa ttcgcggccg cctcgacgac tagcgtttaa 420

ttaaacgcgt gtctgtctgc acatttcgta gagcgagtgt tccgatactc taatctccct 480

aggcaaggtt catatttgtg taggttactt attctccttt tgttgactaa gtcaataatc 540

agaatcagca ggtttggagt cagcttggca gggatcagca gcctgggttg gaaggagggg 600

gtataaaagc cccttcacca ggagaagccg tcacacagat ccacaagctc ctgaggtaag 660

tgccgtgtgt ggttcccgcg ggcctggcct ctttacgggt tatggccctt gcgtgccttg 720

aattactgac actgacatcc actttttctt tttctccaca gctagcgcca ccatgcagat 780

tgagctgtcc acttgtttct tcctgtgcct cctgcgcttc tgtttctccg ccactcgccg 840

gtactacctt ggagccgtgg agctttcatg ggactacatg cagagcgacc tgggcgaact 900

ccccgtggat gccagattcc ccccccgcgt gccaaagtcc ttccccttta acacctccgt 960

ggtgtacaag aaaaccctct ttgtcgagtt cactgaccac ctgttcaaca tcgccaagcc 1020

gcgcccacct tggatgggcc tcctgggacc gaccattcaa gctgaagtgt acgacaccgt 1080

ggtgatcacc ctgaagaaca tggcgtccca ccccgtgtcc ctgcatgcgg tcggagtgtc 1140

ctactggaag gcctccgaag gagctgagta cgacgaccag actagccagc gggaaaagga 1200

ggacgataaa gtgttcccgg gcggctcgca tacttacgtg tggcaagtcc tgaaggaaaa 1260

cggacctatg gcatccgatc ctctgtgcct gacttactcc tacctttccc atgtggacct 1320

cgtgaaggac ctgaacagcg ggctgattgg tgcacttctc gtgtgccgcg aaggttcgct 1380

cgctaaggaa aagacccaga ccctccataa gttcatcctt ttgttcgctg tgttcgatga 1440

aggaaagtca tggcattccg aaactaagaa ctcgctgatg caggaccggg atgccgcctc 1500

agcccgcgcc tggcctaaaa tgcatacagt caacggatac gtgaatcggt cactgcccgg 1560

gctcatcggt tgtcacagaa agtccgtgta ctggcacgtc atcggcatgg gcactacgcc 1620
tgaagtgcac tccatcttcc tggaagggca caccttcctc gtgcgcaacc accgccaggc 1680
ctctctggaa atctccccga ttacctttct gaccgcccag actctgctca tggacctggg 1740
gcagttcctt ctcttctgcc acatctccag ccatcagcac gacggaatgg aggcctacgt 1800
gaaggtggac tcatgcccgg aagaacctca gttgcggatg aagaacaacg aggaggccga 1860
ggactatgac gacgatttga ctgactccga gatggacgtc gtgcggttcg atgacgacaa 1920
cagccccagc ttcatccaga ttcgcagcgt ggccaagaag caccccaaaa cctgggtgca 1980
ctacatcgcg gccgaggaag aagattggga ctacgccccg ttggtgctgg cacccgatga 2040
ccggtcgtac aagtcccagt atctgaacaa tggtccgcag cggattggca gaaagtacaa 2100
gaaagtgcgg ttcatggcgt acactgacga aacgtttaag acccgggagg ccattcaaca 2160
tgagagcggc attctgggac cactgctgta cggagaggtc ggcgataccc tgctcatcat 2220
cttcaaaaac caggcctccc ggccttacaa catctaccct cacggaatca ccgacgtgcg 2280
gccactctac tcgcggcgcc tgccgaaggg cgtcaagcac ctgaaagact tccctatcct 2340
gccgggcgaa atcttcaagt ataagtggac cgtcaccgtg gaggacgggc ccaccaagag 2400
cgatcctagg tgtctgactc ggtactactc cagcttcgtg aacatggaac gggacctggc 2460
atcgggactc attggaccgc tgctgatctg ctacaaagag tcggtggatc aacgcggcaa 2520
ccagatcatg tccgacaagc gcaacgtgat cctgttctcc gtgtttgatg aaaacagatc 2580
ctggtacctc actgaaaaca tccagaggtt cctcccaaac cccgcaggag tgcaactgga 2640
ggaccctgag tttcaggcct cgaatatcat gcactcgatt aacggttacg tgttcgactc 2700
gctgcaactg agcgtgtgcc tccatgaagt cgcttactgg tacattctgt ccatcggcgc 2760
ccagactgac ttcctgagcg tgttcttttc cggttacacc tttaagcaca agatggtgta 2820
cgaagatacc ctgaccctgt tccctttctc cggcgaaacg gtgttcatgt cgatggagaa 2880
cccgggtctg tggattctgg gatgccacaa cagcgacttt cggaaccgcg gaatgactgc 2940
cctgctgaag gtgtcctcat gcgacaagaa caccggagac tactacgagg actcctacga 3000
ggatatctca gcctacctcc tgtccaagaa caacgcgatc gagccgcgca gcttcagcca 3060
gaacggcgcg ccaacatcag agagcgccac ccctgaaagt ggtcccggga gcgagccagc 3120
cacatctggg tcggaaacgc caggcacaag tgagtctgca actcccgagt ccggacctgg 3180
ctccgagcct gccactagcg gctccgagac tccgggaact tccgagagcg ctacaccaga 3240
aagcggaccc ggaaccagta ccgaacctag cgagggctct gctccgggca gcccagccgg 3300
ctctcctaca tccacggagg agggcacttc cgaatccgcc accccggagt cagggccagg 3360
atctgaaccc gctacctcag gcagtgagac gccaggaacg agcgagtccg ctacaccgga 3420
gagtgggcca gggagccctg ctggatctcc tacgtccact gaggaagggt caccagcggg 3480
ctcgcccacc agcactgaag aaggtgcctc gagcccgcct gtgctgaaga ggcaccagcg 3540
agaaattacc cggaccaccc tccaatcgga tcaggaggaa atcgactacg acgacaccat 3600
ctcggtggaa atgaagaagg aagatttcga tatctacgac gaggacgaaa atcagtcccc 3660
tcgctcattc caaaagaaaa ctagacacta ctttatcgcc gcggtggaaa gactgtggga 3720

ctatggaatg tcatccagcc ctcacgtcct tcggaaccgg gcccagagcg gatcggtgcc 3780
tcagttcaag aaagtggtgt tccaggagtt caccgacggc agcttcaccc agccgctgta 3840
ccggggagaa ctgaacgaac acctgggcct gctcggtccc tacatccgcg cggaagtgga 3900
ggataacatc atggtgacct tccgtaacca agcatccaga ccttactcct tctattcctc 3960
cctgatctca tacgaggagg accagcgcca aggcgccgag ccccgcaaga acttcgtcaa 4020
gcccaacgag actaagacct acttctggaa ggtccaacac catatggccc cgaccaagga 4080
tgagtttgac tgcaaggcct gggcctactt ctccgacgtg gaccttgaga aggatgtcca 4140
ttccggcctg atcgggccgc tgctcgtgtg tcacaccaac accctgaacc cagcgcatgg 4200
acgccaggtc accgtccagg agtttgctct gttcttcacc atttttgacg aaactaagtc 4260
ctggtacttc accgagaata tggagcgaaa ctgtagagcg ccctgcaata tccagatgga 4320
agatccgact ttcaaggaga actatagatt ccacgccatc aacgggtaca tcatggatac 4380
tctgccgggg ctggtcatgg cccaggatca gaggattcgg tggtacttgc tgtcaatggg 4440
atcgaacgaa aacattcact ccattcactt ctccggtcac gtgttcactg tgcgcaagaa 4500
ggaggagtac aagatggcgc tgtacaatct gtaccccggg gtgttcgaaa ctgtggagat 4560
gctgccgtcc aaggccggca tctggagagt ggagtgcctg atcggagagc acctccacgc 4620
ggggatgtcc accctcttcc tggtgtactc gaataagtgc cagaccccgc tgggcatggc 4680
ctcgggccac atcagagact tccagatcac agcaagcgga caatacggcc aatgggcgcc 4740
gaagctggcc cgcttgcact actccggatc gatcaacgca tggtccacca aggaaccgtt 4800
ctcgtggatt aaggtggacc tcctggcccc tatgattatc cacggaatta agacccaggg 4860
cgccaggcag aagttctcct ccctgtacat ctcgcaattc atcatcatgt acagcctgga 4920
cgggaagaag tggcagactt acaggggaaa ctccaccggc accctgatgg tctttttcgg 4980
caacgtggat tcctccggca ttaagcacaa catcttcaac ccaccgatca tagccagata 5040
tattaggctc caccccactc actactcaat ccgctcaact cttcggatgg aactcatggg 5100
gtgcgacctg aactcctgct ccatgccgtt ggggatggaa tcaaaggcta ttagcgacgc 5160
ccagatcacc gcgagctcct acttcactaa catgttcgcc acctggagcc cctccaaggc 5220
caggctgcac ttgcagggac ggtcaaatgc ctggcggccg caagtgaaca atccgaagga 5280
atggcttcaa gtggatttcc aaaagaccat gaaagtgacc ggagtcacca cccagggagt 5340
gaagtccctt ctgacctcga tgtatgtgaa ggagttcctg attagcagca gccaggacgg 5400
gcaccagtgg accctgttct tccaaaacgg aaaggtcaag gtgttccagg ggaaccagga 5460
ctcgttcaca cccgtggtga actccctgga ccccccactg ctgacgcggt acttgaggat 5520
tcatcctcag tcctgggtcc atcagattgc attgcgaatg gaagtcctgg gctgcgaggc 5580
ccaggacctg tactgaatca gcctgagctc gctgatcata atcaacctct ggattacaaa 5640
atttgtgaaa gattgactgg tattcttaac tatgttgctc cttttacgct atgtggatac 5700
gctgctttaa tgcctttgta tcatgctatt gcttcccgta tggctttcat tttctcctcc 5760
ttgtataaat cctggttgct gtctctttat gaggagttgt ggcccgttgt caggcaacgt 5820
ggcgtggtgt gcactgtgtt tgctgacgca acccccactg gttggggcat tgccaccacc 5880

tgtcagctcc tttccgggac tttcgctttc cccctcccta ttgccacggc ggaactcatc 5940
gccgcctgcc ttgcccgctg ctggacaggg gctcggctgt tgggcactga caattccgtg 6000
gtgttgtcgg ggaaatcatc gtcctttcct tggctgctcg cctgtgttgc cacctggatt 6060
ctgcgcggga cgtccttctg ctacgtccct tcggccctca atccagcgga ccttccttcc 6120
cgcggcctgc tgccggctct gcggcctctt ccgcgtcttc gccttcgccc tcagacgagt 6180
cggatctccc tttgggccgc ctccccgctg atcagcctcg actgtgcctt ctagttgcca 6240
gccatctgtt gtttgcccct cccccgtgcc ttccttgacc ctggaaggtg ccactcccac 6300
tgtcctttcc taataaaatg aggaaattgc atcgcattgt ctgagtaggt gtcattctat 6360
tctgggggggt ggggtggggc aggacagcaa gggggaggat tgggaagaca atagcaggca 6420
tgctggggat gcggtgggct ctatggcttc tgaggcggaa agaacgggct cgagaagctt 6480
ctagatatcc tctcttaagg tagcatcgag atttaaatta gggataacag ggtaatggcg 6540
cgggccgcca cttcctggcg cgcaaaatat cctcttgtcc ttgagtctca ttggagggtt 6600
cgttcgttcg aaccagccaa tcaggggagg gggaagtgac gcaagttccg gtcacatgct 6660
tccggtgacg cacatccggt gacgtagttc cggtcacgtg cttcctgtca cgtgtttccg 6720
gtcacgtgac ttccggtcat gtgacttccg gtgacgtgtt tccggcttaa ctattgggct 6780
gaccgcgcgg catgcgcgtg gtcaacctaa cagccggaaa cacgtcaccg 6830

<210> 183
<211> 282
<212> DNA
<213> 人工序列

<220>
<223> 5' ITR

<400> 183
cggtgacgtg tttccggctg ttaggttgac cacgcgcatg ccgcgcggtc agcccaatag 60
ttaagccgga aacacgtcac cggaagtcac atgaccggaa gtcacgtgac cggaaacacg 120
tgacaggaag cacgtgaccg gaactacgtc accggatgtg cgtcaccgga agcatgtgac 180
cggaacttgc gtcacttccc cctcccctga ttggctggtt cgaacgaacg aaccctccaa 240
tgagactcaa ggacaagagg atattttgcg cgccaggaag tg 282

<210> 184
<211> 282

<212> DNA
<213> 人工序列

<220>
<223> 3' ITR

<400> 184
cacttcctgg cgcgcaaaat atcctcttgt ccttgagtct cattggaggg ttcgttcgtt 60
cgaaccagcc aatcagggga gggggaagtg acgcaagttc cggtcacatg cttccggtga 120
cgcacatccg gtgacgtagt tccggtcacg tgcttcctgt cacgtgtttc cggtcacgtg 180
acttccggtc atgtgacttc cggtgacgtg tttccggctt aactattggg ctgaccgcgc 240
ggcatgcgcg tggtcaacct aacagccgga aacacgtcac cg 282

<210> 185
<211> 383
<212> DNA
<213> 人工序列

<220>
<223> 5' ITR

<400> 185
ccaaatcaga tgccgccggt cgccgccggt aggcgggact tccggtacaa gatggcggac 60
aattacgtca tttcctgtga cgtcatttcc tgtgacgtca cttccggtgg gcgggacttc 120
cggaattagg gttggctctg ggccagcttg cttggggttg ccttgacact aagacaagcg 180
gcgcgccgct tgatcttagt ggcacgtcaa ccccaagcgc tggcccagag ccaaccctaa 240
ttccggaagt cccgcccacc ggaagtgacg tcacaggaaa tgacgtcaca ggaaatgacg 300
taattgtccg ccatcttgta ccggaagtcc cgcctaccgg cggcgaccgg cggcatctga 360
tttggtgtct tcttttaaat ttt 383

<210> 186
<211> 383
<212> DNA
<213> 人工序列

<220>

<223> 3' ITR

<400> 186

aaaatttaaa agaagacacc aaatcagatg ccgccggtcg ccgccggtag gcgggacttc 60

cggtacaaga tggcggacaa ttacgtcatt tcctgtgacg tcatttcctg tgacgtcact 120

tccggtgggc gggacttccg gaattagggt tggctctggg ccagcgcttg gggttgacgt 180

gccactaaga tcaagcggcg cgccgcttgt cttagtgtca aggcaacccc aagcaagctg 240

gcccagagcc aaccctaatt ccggaagtcc cgcccaccgg aagtgacgtc acaggaaatg 300

acgtcacagg aaatgacgta attgtccgcc atcttgtacc ggaagtcccg cctaccggcg 360

gcgaccggcg gcatctgatt tgg 383

<210> 187

<211> 444

<212> DNA

<213> 人工序列

<220>

<223> 5' ITR

<400> 187

ctcattggag ggttcgttcg ttcgaaccag ccaatcaggg gaggggggaag tgacgcaagt 60

tccggtcaca tgcttccggt gacgcacatc cggtgacgta gttccggtca cgtgcttcct 120

gtcacgtgtt tccggtcacg tgacttccgg tcatgtgact tccggtgacg tgtttccggc 180

tgttaggttg accacgcgca tgccgcgcgg tcagcccaat agttaagccg gaaacacgtc 240

accggaagtc acatgaccgg aagtcacgtg accggaaaca cgtgacagga agcacgtgac 300

cggaactacg tcaccggatg tgcgtcaccg gaagcatgtg accggaactt gcgtcacttc 360

cccctcccct gattggctgg ttcgaacgaa cgaaccctcc aatgagactc aaggacaaga 420

ggatattttg cgcgccagga agtg 444

<210> 188

<211> 444

<212> DNA

<213> 人工序列

<220>

<223> 3' ITR

<400> 188

cacttcctgg cgcgcaaaat atcctcttgt ccttgagtct cattggaggg ttcgttcgtt 60

cgaaccagcc aatcagggga gggggaagtg acgcaagttc cggtcacatg cttccggtga 120

cgcacatccg gtgacgtagt tccggtcacg tgcttcctgt cacgtgtttc cggtcacgtg 180

acttccggtc atgtgacttc cggtgacgtg tttccggctt aactattggg ctgaccgcgc 240

ggcatgcgcg tggtcaacct aacagccgga aacacgtcac cggaagtcac atgaccggaa 300

gtcacgtgac cggaaacacg tgacaggaag cacgtgaccg gaactacgtc accggatgtg 360

cgtcaccgga agcatgtgac cggaacttgc gtcacttccc cctcccctga ttggctggtt 420

cgaacgaacg aaccctccaa tgag 444

<210> 189

<211> 7032

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸序列

<400> 189

ccaaatcaga tgccgccggt cgccgccggt aggcgggact tccggtacaa gatggcggac 60

aattacgtca tttcctgtga cgtcatttcc tgtgacgtca cttccggtgg gcgggacttc 120

cggaattagg gttggctctg ggccagcttg cttggggttg ccttgacact aagacaagcg 180

gcgcgccgct tgatcttagt ggcacgtcaa ccccaagcgc tggcccagag ccaaccctaa 240

ttccggaagt cccgcccacc ggaagtgacg tcacaggaaa tgacgtcaca ggaaatgacg 300

taattgtccg ccatcttgta ccggaagtcc cgcctaccgg cggcgaccgg cggcatctga 360

tttggtgtct tcttttaaat tttgcggcaa ttcagtcgat aactataacg gtcctaaggt 420

agcgatttaa atacgcgctc tcttaaggta gccccgggac gcgtcaattg agatctggat 480

ccggtaccga attcgcggcc gcctcgacga ctagcgttta attaaacgcg tgtctgtctg 540

cacatttcgt agagcgagtg ttccgatact ctaatctccc taggcaaggt tcatatttgt 600

gtaggttact tattctcctt ttgttgacta agtcaataat cagaatcagc aggtttggag 660

tcagcttggc agggatcagc agcctgggtt ggaaggaggg ggtataaaag ccccttcacc 720

aggagaagcc gtcacacaga tccacaagct cctgaggtaa gtgccgtgtg tggttcccgc 780

gggcctggcc tctttacggg ttatggccct tgcgtgcctt gaattactga cactgacatc 840

cacttttct ttttctccac agctagcgcc accatgcaga ttgagctgtc cacttgtttc 900

ttcctgtgcc tcctgcgctt ctgtttctcc gccactcgcc ggtactacct tggagccgtg 960

gagctttcat gggactacat gcagagcgac ctgggcgaac tccccgtgga tgccagattc 1020

ccccccgcg tgccaaagtc cttccccttt aacacctccg tggtgtacaa gaaaaccctc 1080

tttgtcgagt tcactgacca cctgttcaac atcgccaagc cgcgcccacc ttggatgggc 1140

ctcctgggac cgaccattca agctgaagtg tacgacaccg tggtgatcac cctgaagaac 1200

atggcgtccc accccgtgtc cctgcatgcg gtcggagtgt cctactggaa ggcctccgaa 1260

ggagctgagt acgacgacca gactagccag cgggaaaagg aggacgataa agtgttcccg 1320

ggcggctcgc atacttacgt gtggcaagtc ctgaaggaaa acggacctat ggcatccgat 1380

cctctgtgcc tgacttactc ctacctttcc catgtggacc tcgtgaagga cctgaacagc 1440

gggctgattg gtgcacttct cgtgtgccgc gaaggttcgc tcgctaagga aaagacccag 1500

accctccata agttcatcct tttgttcgct gtgttcgatg aaggaaagtc atggcattcc 1560

gaaactaaga actcgctgat gcaggaccgg gatgccgcct cagcccgcgc ctggcctaaa 1620

atgcatacag tcaacggata cgtgaatcgg tcactgcccg ggctcatcgg ttgtcacaga 1680

aagtccgtgt actggcacgt catcggcatg ggcactacgc ctgaagtgca ctccatcttc 1740

ctggaagggc acaccttcct cgtgcgcaac caccgccagg cctctctgga aatctccccg 1800

attacctttc tgaccgccca gactctgctc atggacctgg ggcagttcct tctcttctgc 1860

cacatctcca gccatcagca cgacggaatg gaggcctacg tgaaggtgga ctcatgcccg 1920

gaagaacctc agttgcggat gaagaacaac gaggaggccg aggactatga cgacgatttg 1980

actgactccg agatggacgt cgtgcggttc gatgacgaca acagccccag cttcatccag 2040

attcgcagcg tggccaagaa gcaccccaaa acctgggtgc actacatcgc ggccgaggaa 2100

gaagattggg actacgcccc gttggtgctg gcacccgatg accggtcgta caagtcccag 2160

tatctgaaca atggtccgca gcggattggc agaaagtaca agaaagtgcg gttcatggcg 2220

tacactgacg aaacgtttaa gacccgggag gccattcaac atgagagcgg cattctggga 2280

ccactgctgt acggagaggt cggcgatacc ctgctcatca tcttcaaaaa ccaggcctcc 2340

cggccttaca acatctaccc tcacggaatc accgacgtgc ggccactcta ctcgcggcgc 2400

ctgccgaagg gcgtcaagca cctgaaagac ttccctatcc tgccgggcga aatcttcaag 2460

tataagtgga ccgtcaccgt ggaggacggg cccaccaaga gcgatcctag gtgtctgact 2520

cggtactact ccagcttcgt gaacatggaa cgggacctgg catcgggact cattggaccg 2580

ctgctgatct gctacaaaga gtcggtggat caacgcggca accagatcat gtccgacaag 2640

cgcaacgtga tcctgttctc cgtgtttgat gaaaacagat cctggtacct cactgaaaac 2700

atccagaggt tcctcccaaa ccccgcagga gtgcaactgg aggaccctga gtttcaggcc 2760

tcgaatatca tgcactcgat taacggttac gtgttcgact cgctgcaact gagcgtgtgc 2820

ctccatgaag tcgcttactg gtacattctg tccatcggcg cccagactga cttcctgagc 2880
gtgttctttt ccggttacac ctttaagcac aagatggtgt acgaagatac cctgaccctg 2940
ttccctttct ccggcgaaac ggtgttcatg tcgatggaga acccgggtct gtggattctg 3000
ggatgccaca acagcgactt tcggaaccgc ggaatgactg ccctgctgaa ggtgtcctca 3060
tgcgacaaga acaccggaga ctactacgag gactcctacg aggatatctc agcctacctc 3120
ctgtccaaga acaacgcgat cgagccgcgc agcttcagcc agaacggcgc gccaacatca 3180
gagagcgcca cccctgaaag tggtcccggg agcgagccag ccacatctgg gtcggaaacg 3240
ccaggcacaa gtgagtctgc aactcccgag tccggacctg gctccgagcc tgccactagc 3300
ggctccgaga ctccgggaac ttccgagagc gctacaccag aaagcggacc cggaaccagt 3360
accgaaccta gcgagggctc tgctccgggc agcccagccg gctctcctac atccacggag 3420
gagggcactt ccgaatccgc caccccggag tcagggccag gatctgaacc cgctacctca 3480
ggcagtgaga cgccaggaac gagcgagtcc gctacaccgg agagtgggcc agggagccct 3540
gctggatctc ctacgtccac tgaggaaggg tcaccagcgg gctcgcccac cagcactgaa 3600
gaaggtgcct cgagcccgcc tgtgctgaag aggcaccagc gagaaattac ccggaccacc 3660
ctccaatcgg atcaggagga aatcgactac gacgacacca tctcggtgga aatgaagaag 3720
gaagatttcg atatctacga cgaggacgaa aatcagtccc ctcgctcatt ccaaaagaaa 3780
actagacact actttatcgc cgcggtggaa agactgtggg actatggaat gtcatccagc 3840
cctcacgtcc ttcggaaccg ggcccagagc ggatcggtgc ctcagttcaa gaaagtggtg 3900
ttccaggagt tcaccgacgg cagcttcacc cagccgctgt accggggaga actgaacgaa 3960
cacctgggcc tgctcggtcc ctacatccgc gcggaagtgg aggataacat catggtgacc 4020
ttccgtaacc aagcatccag accttactcc ttctattcct ccctgatctc atacgaggag 4080
gaccagcgcc aaggcgccga gccccgcaag aacttcgtca agcccaacga gactaagacc 4140
tacttctgga aggtccaaca ccatatggcc ccgaccaagg atgagtttga ctgcaaggcc 4200
tgggcctact tctccgacgt ggaccttgag aaggatgtcc attccggcct gatcgggccg 4260
ctgctcgtgt gtcacaccaa caccctgaac ccagcgcatg gacgccaggt caccgtccag 4320
gagtttgctc tgttcttcac cattttgac gaaactaagt cctggtactt caccgagaat 4380
atggagcgaa actgtagagc gccctgcaat atccagatgg aagatccgac tttcaaggag 4440
aactatagat tccacgccat caacgggtac atcatggata ctctgccggg gctggtcatg 4500
gcccaggatc agaggattcg gtggtacttg ctgtcaatgg gatcgaacga aaacattcac 4560
tccattcact tctccggtca cgtgttcact gtgcgcaaga aggaggagta caagatggcg 4620
ctgtacaatc tgtaccccgg ggtgttcgaa actgtggaga tgctgccgtc caaggccggc 4680
atctggagag tggagtgcct gatcggagag cacctccacg cggggatgtc caccctcttc 4740
ctggtgtact cgaataagtg ccagaccccg ctgggcatgg cctcgggcca catcagagac 4800
ttccagatca cagcaagcgg acaatacggc caatgggcgc cgaagctggc ccgcttgcac 4860
tactccggat cgatcaacgc atggtccacc aaggaaccgt tctcgtggat taaggtggac 4920
ctcctggccc ctatgattat ccacggaatt aagacccagg gcgccaggca gaagttctcc 4980

tccctgtaca tctcgcaatt catcatcatg tacagcctgg acgggaagaa gtggcagact 5040
tacaggggaa actccaccgg caccctgatg gtctttttcg gcaacgtgga ttcctccggc 5100
attaagcaca acatcttcaa cccaccgatc atagccagat atattaggct ccaccccact 5160
cactactcaa tccgctcaac tcttcggatg gaactcatgg ggtgcgacct gaactcctgc 5220
tccatgccgt tggggatgga atcaaaggct attagcgacg cccagatcac cgcgagctcc 5280
tacttcacta acatgttcgc cacctggagc ccctccaagg ccaggctgca cttgcaggga 5340
cggtcaaatg cctggcggcc gcaagtgaac aatccgaagg aatggcttca agtggatttc 5400
caaaagacca tgaaagtgac cggagtcacc acccagggag tgaagtccct tctgacctcg 5460
atgtatgtga aggagttcct gattagcagc agccaggacg ggcaccagtg gaccctgttc 5520
ttccaaaacg gaaaggtcaa ggtgttccag gggaaccagg actcgttcac acccgtggtg 5580
aactccctgg acccccccact gctgacgcgg tacttgagga ttcatcctca gtcctgggtc 5640
catcagattg cattgcgaat ggaagtcctg ggctgcgagg cccaggacct gtactgaatc 5700
agcctgagct cgctgatcat aatcaacctc tggattacaa aatttgtgaa agattgactg 5760
gtattcttaa ctatgttgct ccttttacgc tatgtggata cgctgcttta atgcctttgt 5820
atcatgctat tgcttcccgt atggctttca ttttctcctc cttgtataaa tcctggttgc 5880
tgtctcttta tgaggagttg tggcccgttg tcaggcaacg tggcgtggtg tgcactgtgt 5940
ttgctgacgc aacccccact ggttggggca ttgccaccac ctgtcagctc ctttccggga 6000
ctttcgcttt ccccctccct attgccacgg cggaactcat cgccgcctgc cttgcccgct 6060
gctggacagg ggctcggctg ttgggcactg acaattccgt ggtgttgtcg gggaaatcat 6120
cgtcctttcc ttggctgctc gcctgtgttg ccacctggat tctgcgcggg acgtccttct 6180
gctacgtccc ttcggccctc aatccagcgg accttccttc ccgcggcctg ctgccggctc 6240
tgcggcctct tccgcgtctt cgccttcgcc ctcagacgag tcggatctcc ctttgggccg 6300
cctccccgct gatcagcctc gactgtgcct tctagttgcc agccatctgt tgtttgcccc 6360
tcccccgtgc cttccttgac cctggaaggt gccactccca ctgtcctttc ctaataaaat 6420
gaggaaattg catcgcattg tctgagtagg tgtcattcta ttctgggggg tggggtgggg 6480
caggacagca agggggagga ttgggaagac aatagcaggc atgctgggga tgcggtgggc 6540
tctatggctt ctgaggcgga aagaacgggc tcgagaagct tctagatatc ctctcttaag 6600
gtagcatcga gatttaaatt agggataaca gggtaatggc gcgggccgca aaatttaaaa 6660
gaagacacca aatcagatgc cgccggtcgc cgccggtagg cgggacttcc ggtacaagat 6720
ggcggacaat tacgtcattt cctgtgacgt catttcctgt gacgtcactt ccggtgggcg 6780
ggacttccgg aattagggtt ggctctgggc cagcgcttgg ggttgacgtg ccactaagat 6840
caagcggcgc gccgcttgtc ttagtgtcaa ggcaacccca agcaagctgg cccagagcca 6900
accctaattc cggaagtccc gcccaccgga agtgacgtca caggaaatga cgtcacagga 6960
aatgacgtaa ttgtccgcca tcttgtaccg gaagtcccgc ctaccggcgg cgaccggcgg 7020
catctgattt gg 7032

<210> 190

<211> 6824

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸序列

<400> 190

ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgaccttt 60

ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggccaa ctccatcact 120

aggggttcct gcggcaattc agtcgataac tataacggtc ctaaggtagc gatttaaata 180

cgcgctctct taaggtagcc ccgggacgcg tcaattgaga tctggatccg gtaccgaatt 240

cgcggccgcc tcgacgacta gcgtttaatt aaatcgaggt gagccccacg ttctgcttca 300

ctctccccat ctccccccccc tccccacccc caattttgta tttatttatt ttttaattat 360

tttgtgcagc gatgggggcg gggggggggg gggggcgcgc gccaggcggg gcggggcggg 420

gcgaggggcg gggcggggcg aggcggagag gtgcggcggc agccaatcag agcggcgcgc 480

tccgaaagtt tccttttatg gcgaggcggc ggcggcggcg gccctataaa aagcgaagcg 540

cgcggcgggc gggagtcgct gcgcgctgcc ttcgccccgt gccccgctcc gccgccgcct 600

cgcgccgccc gccccggctc tgactgaccg cgttactccc acaggtgagc gggcgggacg 660

gcccttctcc ttcgggctgt aattagcgct tggtttaatg acggcttgtt tcttttctgt 720

ggctgcgtga aagccttgag gggctccggg agggcccttt gtgcgggggg agcggctcgg 780

ggctgtccgc gggggggacgg ctgccttcgg gggggacggg gcagggcggg gttcggcttc 840

tggcgtgtga ccggcggctc tagagcctct gctaaccatg ttcatgcctt cttcttttttc 900

ctacaggcta gcgccaccat gcagattgag ctgtccactt gtttcttcct gtgcctcctg 960

cgcttctgtt tctccgccac tcgccggtac taccttggag ccgtggagct ttcatgggac 1020

tacatgcaga gcgacctggg cgaactcccc gtggatgcca gattcccccc ccgcgtgcca 1080

aagtccttcc cctttaacac ctccgtggtg tacaagaaaa ccctctttgt cgagttcact 1140

gaccacctgt tcaacatcgc caagccgcgc ccaccttgga tgggcctcct gggaccgacc 1200

attcaagctg aagtgtacga caccgtggtg atcaccctga agaacatggc gtcccacccc 1260

gtgtccctgc atgcggtcgg agtgtcctac tggaaggcct ccgaaggagc tgagtacgac 1320

gaccagacta gccagcggga aaaggaggac gataaagtgt tcccgggcgg ctcgcatact 1380

tacgtgtggc aagtcctgaa ggaaaacgga cctatggcat ccgatcctct gtgcctgact 1440

tactcctacc tttcccatgt ggacctcgtg aaggacctga acagcgggct gattggtgca 1500

cttctcgtgt gccgcgaagg ttcgctcgct aaggaaaaga cccagaccct ccataagttc 1560

atccttttgt tcgctgtgtt cgatgaagga aagtcatggc attccgaaac taagaactcg 1620
ctgatgcagg accgggatgc cgcctcagcc cgcgcctggc ctaaaatgca tacagtcaac 1680
ggatacgtga atcggtcact gcccgggctc atcggttgtc acagaaagtc cgtgtactgg 1740
cacgtcatcg gcatgggcac tacgcctgaa gtgcactcca tcttcctgga agggcacacc 1800
ttcctcgtgc gcaaccaccg ccaggcctct ctggaaatct ccccgattac ctttctgacc 1860
gcccagactc tgctcatgga cctggggcag ttccttctct tctgccacat ctccagccat 1920
cagcacgacg gaatggaggc ctacgtgaag gtggactcat gcccggaaga acctcagttg 1980
cggatgaaga acaacgagga ggccgaggac tatgacgacg atttgactga ctccgagatg 2040
gacgtcgtgc ggttcgatga cgacaacagc cccagcttca tccagattcg cagcgtggcc 2100
aagaagcacc ccaaaacctg ggtgcactac atcgcggccg aggaagaaga ttgggactac 2160
gccccgttgg tgctggcacc cgatgaccgg tcgtacaagt cccagtatct gaacaatggt 2220
ccgcagcgga ttggcagaaa gtacaagaaa gtgcggttca tggcgtacac tgacgaaacg 2280
tttaagaccc gggaggccat tcaacatgag agcggcattc tgggaccact gctgtacgga 2340
gaggtcggcg ataccctgct catcatcttc aaaaaccagg cctcccggcc ttacaacatc 2400
taccctcacg gaatcaccga cgtgcggcca ctctactcgc ggcgcctgcc gaagggcgtc 2460
aagcacctga aagacttccc tatcctgccg ggcgaaatct tcaagtataa gtggaccgtc 2520
accgtggagg acgggcccac caagagcgat cctaggtgtc tgactcggta ctactccagc 2580
ttcgtgaaca tggaacggga cctggcatcg ggactcattg gaccgctgct gatctgctac 2640
aaagagtcgg tggatcaacg cggcaaccag atcatgtccg acaagcgcaa cgtgatcctg 2700
ttctccgtgt ttgatgaaaa cagatcctgg tacctcactg aaaacatcca gaggttcctc 2760
ccaaaccccg caggagtgca actggaggac cctgagtttc aggcctcgaa tatcatgcac 2820
tcgattaacg gttacgtgtt cgactcgctg caactgagcg tgtgcctcca tgaagtcgct 2880
tactggtaca ttctgtccat cggcgcccag actgacttcc tgagcgtgtt cttttccggt 2940
tacaccttta agcacaagat ggtgtacgaa gataccctga ccctgttccc tttctccggc 3000
gaaacggtgt tcatgtcgat ggagaacccg ggtctgtgga ttctgggatg ccacaacagc 3060
gactttcgga accgcggaat gactgccctg ctgaaggtgt cctcatgcga caagaacacc 3120
ggagactact acgaggactc ctacgaggat atctcagcct acctcctgtc caagaacaac 3180
gcgatcgagc cgcgcagctt cagccagaac ggcgcgccaa catcagagag cgccaccct 3240
gaaagtggtc ccgggagcga gccagccaca tctgggtcgg aaacgccagg cacaagtgag 3300
tctgcaactc ccgagtccgg acctggctcc gagcctgcca ctagcggctc cgagactccg 3360
ggaacttccg agagcgctac accagaaagc ggacccggaa ccagtaccga acctagcgag 3420
ggctctgctc cgggcagccc agccggctct cctacatcca cggaggaggg cacttccgaa 3480
tccgccaccc cggagtcagg gccaggatct gaacccgcta cctcaggcag tgagacgcca 3540
ggaacgagcg agtccgctac accggagagt gggccaggga gccctgctgg atctcctacg 3600
tccactgagg aagggtcacc agcgggctcg cccaccagca ctgaagaagg tgcctcgagc 3660
ccgcctgtgc tgaagaggca ccagcgagaa attacccgga ccaccctcca atcggatcag 3720

gaggaaatcg actacgacga caccatctcg gtggaaatga agaaggaaga tttcgatatc 3780
tacgacgagg acgaaaatca gtcccctcgc tcattccaaa agaaaactag acactacttt 3840
atcgccgcgg tggaaagact gtgggactat ggaatgtcat ccagccctca cgtccttcgg 3900
aaccgggccc agagcggatc ggtgcctcag ttcaagaaag tggtgttcca ggagttcacc 3960
gacggcagct tcacccagcc gctgtaccgg ggagaactga acgaacacct gggcctgctc 4020
ggtccctaca tccgcgcgga agtggaggat aacatcatgg tgaccttccg taaccaagca 4080
tccagacctt actccttcta ttcctccctg atctcatacg aggaggacca gcgccaaggc 4140
gccgagcccc gcaagaactt cgtcaagccc aacgagacta agacctactt ctggaaggtc 4200
caacaccata tggccccgac caaggatgag tttgactgca aggcctgggc ctacttctcc 4260
gacgtggacc ttgagaagga tgtccattcc ggcctgatcg ggccgctgct cgtgtgtcac 4320
accaacaccc tgaacccagc gcatggacgc caggtcaccg tccaggagtt tgctctgttc 4380
ttcaccattt ttgacgaaac taagtcctgg tacttcaccg agaatatgga gcgaaactgt 4440
agagcgccct gcaatatcca gatggaagat ccgactttca aggagaacta tagattccac 4500
gccatcaacg ggtacatcat ggatactctg ccggggctgg tcatggccca ggatcagagg 4560
attcggtggt acttgctgtc aatgggatcg aacgaaaaca ttcactccat tcacttctcc 4620
ggtcacgtgt tcactgtgcg caagaaggag gagtacaaga tggcgctgta caatctgtac 4680
cccggggtgt tcgaaactgt ggagatgctg ccgtccaagg ccggcatctg gagagtggag 4740
tgcctgatcg gagagcacct ccacgcgggg atgtccaccc tcttcctggt gtactcgaat 4800
aagtgccaga ccccgctggg catggcctcg ggccacatca gagacttcca gatcacagca 4860
agcggacaat acggccaatg ggcgccgaag ctggcccgct tgcactactc cggatcgatc 4920
aacgcatggt ccaccaagga accgttctcg tggattaagg tggacctcct ggcccctatg 4980
attatccacg gaattaagac ccagggcgcc aggcagaagt tctcctccct gtacatctcg 5040
caattcatca tcatgtacag cctggacggg aagaagtggc agacttacag gggaaactcc 5100
accggcaccc tgatggtctt tttcggcaac gtggattcct ccggcattaa gcacaacatc 5160
ttcaacccac cgatcatagc cagatatatt aggctccacc ccactcacta ctcaatccgc 5220
tcaactcttc ggatggaact catggggtgc gacctgaact cctgctccat gccgttgggg 5280
atggaatcaa aggctattag cgacgcccag atcaccgcga gctcctactt cactaacatg 5340
ttcgccacct ggagcccctc caaggccagg ctgcacttgc agggacggtc aaatgcctgg 5400
cggccgcaag tgaacaatcc gaaggaatgg cttcaagtgg atttccaaaa gaccatgaaa 5460
gtgaccggag tcaccaccca gggagtgaag tcccttctga cctcgatgta tgtgaaggag 5520
ttcctgatta gcagcagcca ggacgggcac cagtggaccc tgttcttcca aaacggaaag 5580
gtcaaggtgt tccaggggaa ccaggactcg ttcacacccg tggtgaactc cctggacccc 5640
ccactgctga cgcggtactt gaggattcat cctcagtcct gggtccatca gattgcattg 5700
cgaatggaag tcctgggctg cgaggcccag gacctgtact gaatcagcct gagctcgctg 5760
atcataatca acctctggat tacaaaattt gtgaaagatt gactggtatt cttaactatg 5820
ttgctccttt tacgctatgt ggatacgctg ctttaatgcc tttgtatcat gctattgctt 5880

cccgtatggc tttcattttc tcctccttgt ataaatcctg gttgctgtct ctttatgagg 5940
agttgtggcc cgttgtcagg caacgtggcg tggtgtgcac tgtgtttgct gacgcaaccc 6000
ccactggttg gggcattgcc accacctgtc agctcctttc cgggactttc gctttccccc 6060
tccctattgc cacggcggaa ctcatcgccg cctgccttgc ccgctgctgg acaggggctc 6120
ggctgttggg cactgacaat tccgtggtgt tgtcggggaa atcatcgtcc tttccttggc 6180
tgctcgcctg tgttgccacc tggattctgc gcgggacgtc cttctgctac gtcccttcgg 6240
ccctcaatcc agcggacctt ccttcccgcg gcctgctgcc ggctctgcgg cctcttccgc 6300
gtcttcgcct tcgccctcag acgagtcgga tctccctttg ggccgcctcc ccgctgatca 6360
gcctcgactg tgccttctag ttgccagcca tctgttgttt gcccctcccc cgtgccttcc 6420
ttgaccctgg aaggtgccac tcccactgtc ctttcctaat aaaatgagga aattgcatcg 6480
cattgtctga gtaggtgtca ttctattctg gggggtgggg tggggcagga cagcaagggg 6540
gaggattggg aagacaatag caggcatgct ggggatgcgg tgggctctat ggcttctgag 6600
gcggaaagaa cgggctcgag aagcttctag atatcctctc ttaaggtagc atcgagattt 6660
aaattaggga taacagggta atggcgcggg ccgcaggaac ccctagtgat ggagttggcc 6720
actccctctc tgcgcgctcg ctcgctcact gaggccgggc gaccaaaggt cgcccgacgc 6780
ccgggctttg cccgggcggc ctcagtgagc gagcgagcgc gcag 6824

<210> 191
<211> 278
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸序列

<400> 191
tcgaggtgag ccccacgttc tgcttcactc tccccatctc ccccccctcc ccacccccaa 60
ttttgtattt atttatttt taattatttt gtgcagcgat gggggcgggg gggggggggg 120
ggcgcgcgcc aggcggggcg gggcggggcg aggggcgggg cggggcgagg cggagaggtg 180
cggcggcagc caatcagagc ggcgcgctcc gaaagtttcc ttttatggcg aggcggcggc 240
ggcggcggcc ctataaaaag cgaagcgcgc ggcgggcg 278

<210> 192
<211> 262

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸序列

<400> 192

gtgagcgggc gggacggccc ttctccttcg ggctgtaatt agcgcttggt ttaatgacgg 60

cttgtttctt ttctgtggct gcgtgaaagc cttgaggggc tccgggaggg ccctttgtgc 120

ggggggagcg gctcggggct gtccgcgggg ggacggctgc cttcgggggg gacggggcag 180

ggcggggttc ggcttctggc gtgtgaccgg cggctctaga gcctctgcta accatgttca 240

tgccttcttc tttttcctac ag 262

<210> 193

<211> 130

<212> DNA

<213> 人工序列

<220>

<223> 3' ITR

<400> 193

aggaacccct agtgatggag ttggccactc cctctctgcg cgctcgctcg ctcactgagg 60

ccgggcgacc aaaggtcgcc cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc 120

gagcgcgcag 130

<210> 194

<211> 7154

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸序列

<400> 194

ctcattggag ggttcgttcg ttcgaaccag ccaatcaggg gaggggggaag tgacgcaagt 60

tccggtcaca tgcttccggt gacgcacatc cggtgacgta gttccggtca cgtgcttcct 120

gtcacgtgtt tccggtcacg tgacttccgg tcatgtgact tccggtgacg tgtttccggc 180

tgttaggttg accacgcgca tgccgcgcgg tcagcccaat agttaagccg gaaacacgtc 240

accggaagtc acatgaccgg aagtcacgtg accggaaaca cgtgacagga agcacgtgac 300

cggaactacg tcaccggatg tgcgtcaccg gaagcatgtg accggaactt gcgtcacttc 360

cccctcccct gattggctgg ttcgaacgaa cgaaccctcc aatgagactc aaggacaaga 420

ggatattttg cgcgccagga agtggcggca attcagtcga taactataac ggtcctaagg 480

tagcgattta aatacgcgct ctcttaaggt agccccggga cgcgtcaatt gagatctgga 540

tccggtaccg aattcgcggc cgcctcgacg actagcgttt aattaaacgc gtgtctgtct 600

gcacatttcg tagagcgagt gttccgatac tctaatctcc ctaggcaagg ttcatatttg 660

tgtaggttac ttattctcct tttgttgact aagtcaataa tcagaatcag caggtttgga 720

gtcagcttgg cagggatcag cagcctgggt tggaaggagg gggtataaaa gccccttcac 780

caggagaagc cgtcacacag atccacaagc tcctgaggta agtgccgtgt gtggttcccg 840

cgggcctggc ctctttacgg gttatggccc ttgcgtgcct tgaattactg acactgacat 900

ccacttttc ttttctccca cagctagcgc caccatgcag attgagctgt ccacttgttt 960

cttcctgtgc ctcctgcgct tctgtttctc cgccactcgc cggtactacc ttggagccgt 1020

ggagctttca tgggactaca tgcagagcga cctgggcgaa ctccccgtgg atgccagatt 1080

ccccccccgc gtgccaaagt ccttcccctt taacacctcc gtggtgtaca agaaaaccct 1140

ctttgtcgag ttcactgacc acctgttcaa catcgccaag ccgcgcccac cttggatggg 1200

cctcctggga ccgaccattc aagctgaagt gtacgacacc gtggtgatca ccctgaagaa 1260

catggcgtcc caccccgtgt ccctgcatgc ggtcggagtg tcctactgga aggcctccga 1320

aggagctgag tacgacgacc agactagcca gcgggaaaag gaggacgata aagtgttccc 1380

gggcggctcg catacttacg tgtggcaagt cctgaaggaa aacggaccta tggcatccga 1440

tcctctgtgc ctgacttact cctacctttc ccatgtggac ctcgtgaagg acctgaacag 1500

cgggctgatt ggtgcacttc tcgtgtgccg cgaaggttcg ctcgctaagg aaaagaccca 1560

gaccctccat aagttcatcc ttttgttcgc tgtgttcgat gaaggaaagt catggcattc 1620

cgaaactaag aactcgctga tgcaggaccg ggatgccgcc tcagcccgcg cctggcctaa 1680

aatgcataca gtcaacggat acgtgaatcg gtcactgccc gggctcatcg gttgtcacag 1740

aaagtccgtg tactggcacg tcatcggcat gggcactacg cctgaagtgc actccatctt 1800

cctggaaggg cacaccttcc tcgtgcgcaa ccaccgccag gcctctctgg aaatctcccc 1860

gattaccttt ctgaccgccc agactctgct catggacctg gggcagttcc ttctcttctg 1920

ccacatctcc agccatcagc acgacggaat ggaggcctac gtgaaggtgg actcatgccc 1980

ggaagaacct cagttgcgga tgaagaacaa cgaggaggcc gaggactatg acgacgattt 2040

gactgactcc gagatggacg tcgtgcggtt cgatgacgac aacagcccca gcttcatcca 2100

gattcgcagc gtggccaaga agcaccccaa aacctgggtg cactacatcg cggccgagga 2160
agaagattgg gactacgccc cgttggtgct ggcacccgat gaccggtcgt acaagtccca 2220
gtatctgaac aatggtccgc agcggattgg cagaaagtac aagaaagtgc ggttcatggc 2280
gtacactgac gaaacgttta agacccggga ggccattcaa catgagagcg gcattctggg 2340
accactgctg tacggagagg tcggcgatac cctgctcatc atcttcaaaa accaggcctc 2400
ccggccttac aacatctacc ctcacggaat caccgacgtg cggccactct actcgcggcg 2460
cctgccgaag ggcgtcaagc acctgaaaga cttccctatc ctgccgggcg aaatcttcaa 2520
gtataagtgg accgtcaccg tggaggacgg gcccaccaag agcgatccta ggtgtctgac 2580
tcggtactac tccagcttcg tgaacatgga acgggacctg gcatcgggac tcattggacc 2640
gctgctgatc tgctacaaag agtcggtgga tcaacgcggc aaccagatca tgtccgacaa 2700
gcgcaacgtg atcctgttct ccgtgtttga tgaaaacaga tcctggtacc tcactgaaaa 2760
catccagagg ttcctcccaa accccgcagg agtgcaactg gaggaccctg agtttcaggc 2820
ctcgaatatc atgcactcga ttaacggtta cgtgttcgac tcgctgcaac tgagcgtgtg 2880
cctccatgaa gtcgcttact ggtacattct gtccatcggc gcccagactg acttcctgag 2940
cgtgttcttt tccggttaca cctttaagca caagatggtg tacgaagata ccctgaccct 3000
gttccctttc tccggcgaaa cggtgttcat gtcgatggag aacccgggtc tgtggattct 3060
gggatgccac aacagcgact ttcggaaccg cggaatgact gccctgctga aggtgtcctc 3120
atgcgacaag aacaccggag actactacga ggactcctac gaggatatct cagcctacct 3180
cctgtccaag aacaacgcga tcgagccgcg cagcttcagc cagaacggcg cgccaacatc 3240
agagagcgcc acccctgaaa gtggtcccgg gagcgagcca gccacatctg ggtcggaaac 3300
gccaggcaca agtgagtctg caactcccga gtccggacct ggctccgagc ctgccactag 3360
cggctccgag actccgggaa cttccgagag cgctacacca gaaagcggac ccggaaccag 3420
taccgaacct agcgagggct ctgctccggg cagcccagcc ggctctccta catccacgga 3480
ggagggcact tccgaatccg ccaccccgga gtcagggcca ggatctgaac ccgctacctc 3540
aggcagtgag acgccaggaa cgagcgagtc cgctacaccg gagagtgggc cagggagccc 3600
tgctggatct cctacgtcca ctgaggaagg gtcaccagcg ggctcgccca ccagcactga 3660
agaaggtgcc tcgagcccgc ctgtgctgaa gaggcaccag cgagaaatta cccggaccac 3720
cctccaatcg gatcaggagg aaatcgacta cgacgacacc atctcggtgg aaatgaagaa 3780
ggaagatttc gatatctacg acgaggacga aaatcagtcc cctcgctcat tccaaaagaa 3840
aactagacac tactttatcg ccgcggtgga aagactgtgg gactatggaa tgtcatccag 3900
ccctcacgtc cttcggaacc gggcccagag cggatcggtg cctcagttca agaaagtggt 3960
gttccaggag ttcaccgacg gcagcttcac ccagccgctg taccggggag aactgaacga 4020
acacctgggc ctgctcggtc cctacatccg cgcggaagtg gaggataaca tcatggtgac 4080
cttccgtaac caagcatcca gaccttactc cttctattcc tccctgatct catacgagga 4140
ggaccagcgc caaggcgccg agccccgcaa gaacttcgtc aagcccaacg agactaagac 4200
ctacttctgg aaggtccaac accatatggc cccgaccaag gatgagtttg actgcaaggc 4260

ctgggcctac ttctccgacg tggaccttga gaaggatgtc cattccggcc tgatcgggcc 4320
gctgctcgtg tgtcacacca acaccctgaa cccagcgcat ggacgccagg tcaccgtcca 4380
ggagtttgct ctgttcttca ccatttttga cgaaactaag tcctggtact tcaccgagaa 4440
tatggagcga aactgtagag cgccctgcaa tatccagatg gaagatccga ctttcaagga 4500
gaactataga ttccacgcca tcaacgggta catcatggat actctgccgg ggctggtcat 4560
ggcccaggat cagaggattc ggtggtactt gctgtcaatg ggatcgaacg aaaacattca 4620
ctccattcac ttctccggtc acgtgttcac tgtgcgcaag aaggaggagt acaagatggc 4680
gctgtacaat ctgtaccccg gggtgttcga aactgtggag atgctgccgt ccaaggccgg 4740
catctggaga gtggagtgcc tgatcggaga gcacctccac gcggggatgt ccaccctctt 4800
cctggtgtac tcgaataagt gccagacccc gctgggcatg gcctcgggcc acatcagaga 4860
cttccagatc acagcaagcg gacaatacgg ccaatgggcg ccgaagctgg cccgcttgca 4920
ctactccgga tcgatcaacg catggtccac caaggaaccg ttctcgtgga ttaaggtgga 4980
cctcctggcc cctatgatta tccacggaat taagacccag ggcgccaggc agaagttctc 5040
ctccctgtac atctcgcaat tcatcatcat gtacagcctg gacgggaaga agtggcagac 5100
ttacagggga aactccaccg gcaccctgat ggtctttttc ggcaacgtgg attcctccgg 5160
cattaagcac aacatcttca acccaccgat catagccaga tatattaggc tccaccccac 5220
tcactactca atccgctcaa ctcttcggat ggaactcatg gggtgcgacc tgaactcctg 5280
ctccatgccg ttggggatgg aatcaaaggc tattagcgac gcccagatca ccgcgagctc 5340
ctacttcact aacatgttcg ccacctggag cccctccaag gccaggctgc acttgcaggg 5400
acggtcaaat gcctggcggc cgcaagtgaa caatccgaag gaatggcttc aagtggattt 5460
ccaaaagacc atgaaagtga ccggagtcac cacccaggga gtgaagtccc ttctgacctc 5520
gatgtatgtg aaggagttcc tgattagcag cagccaggac gggcaccagt ggaccctgtt 5580
cttccaaaac ggaaaggtca aggtgttcca ggggaaccag gactcgttca cacccgtggt 5640
gaactccctg gacccccccac tgctgacgcg gtacttgagg attcatcctc agtcctgggt 5700
ccatcagatt gcattgcgaa tggaagtcct gggctgcgag gcccaggacc tgtactgaat 5760
cagcctgagc tcgctgatca taatcaacct ctggattaca aaatttgtga aagattgact 5820
ggtattctta actatgttgc tccttttacg ctatgtggat acgctgcttt aatgcctttg 5880
tatcatgcta ttgcttcccg tatggctttc attttctcct ccttgtataa atcctggttg 5940
ctgtctcttt atgaggagtt gtggcccgtt gtcaggcaac gtggcgtggt gtgcactgtg 6000
tttgctgacg caacccccac tggttggggc attgccacca cctgtcagct cctttccggg 6060
actttcgctt tccccctccc tattgccacg gcggaactca tcgccgcctg ccttgcccgc 6120
tgctggacag gggctcggct gttgggcact gacaattccg tggtgttgtc ggggaaatca 6180
tcgtcctttc cttggctgct cgcctgtgtt gccacctgga ttctgcgcgg gacgtccttc 6240
tgctacgtcc cttcggccct caatccagcg gaccttcctt cccgcggcct gctgccggct 6300
ctgcggcctc ttccgcgtct tcgccttcgc cctcagacga gtcggatctc cctttgggcc 6360
gcctccccgc tgatcagcct cgactgtgcc ttctagttgc agccatctg ttgtttgccc 6420

ctcccccgtg ccttccttga ccctggaagg tgccactccc actgtccttt cctaataaaa 6480
tgaggaaatt gcatcgcatt gtctgagtag gtgtcattct attctggggg gtggggtggg 6540
gcaggacagc aaggggggagg attgggaaga caatagcagg catgctgggg atgcggtggg 6600
ctctatggct tctgaggcgg aaagaacggg ctcgagaagc ttctagatat cctctcttaa 6660
ggtagcatcg agatttaaat tagggataac agggtaatgg cgcgggccgc cacttcctgg 6720
cgcgcaaaat atcctcttgt ccttgagtct cattggaggg ttcgttcgtt cgaaccagcc 6780
aatcagggga gggggaagtg acgcaagttc cggtcacatg cttccggtga cgcacatccg 6840
gtgacgtagt tccggtcacg tgcttcctgt cacgtgtttc cggtcacgtg acttccggtc 6900
atgtgacttc cggtgacgtg tttccggctt aactattggg ctgaccgcgc ggcatgcgcg 6960
tggtcaacct aacagccgga aacacgtcac cggaagtcac atgaccggaa gtcacgtgac 7020
cggaaacacg tgacaggaag cacgtgaccg gaactacgtc accggatgtg cgtcaccgga 7080
agcatgtgac cggaacttgc gtcacttccc cctcccctga ttggctggtt cgaacgaacg 7140
aaccctccaa tgag 7154

<210> 195
<211> 648
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸序列

<400> 195
ctagttatta atagtaatca attacggggt cattagttca tagcccatat atggagttcc 60
gcgttacata acttacggta aatggcccgc ctggctgacc gcccaacgac ccccgcccat 120
tgacgtcaat aatgacgtat gttcccatag taacgccaat agggactttc cattgacgtc 180
aatgggtgga gtatttacgg taaactgccc acttggcagt acatcaagtg tatcatatgc 240
caagtacgcc ccctattgac gtcaatgacg gtaaatggcc cgcctggcat tatgcccagt 300
acatgacctt atgggacttt cctacttggc agtacatcta cgtattagtc atcgctatta 360
ccatgcatgg tcgaggtgag ccccacgttc tgcttcactc tccccatctc ccccccctcc 420
ccacccccaa ttttgtattt atttatttttt taattatttt gtgcagcgat gggggcgggg 480
ggggggggggg ggcgcgcgcc aggcggggcg gggcggggcg aggggcgggg cggggcgagg 540
cggagaggtg cggcggcagc caatcagagc ggcgcgctcc gaaagtttcc ttttatggcg 600
aggcggcggc ggcggcggcc ctataaaaag cgaagcgcgc ggcgggcg 648

<210> 196

<211> 1428

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸序列

<400> 196

atggactaca aagaccatga cggtgattat aaagatcatg atatcgatta caaggatgac 60

gatgacaagg ctgctgtggt tctggaaaat ggcgtgctga gccggaagct gagcgacttc 120

ggacaagaga caagctacat cgaggacaac agcaaccaga atggcgccgt gtctctgatc 180

ttcagcctga aagaagaagt gggcgccctg gccaaggtgc tgagactgtt cgaggaaaac 240

gagatcaatc tgacccacat cgagagcaga cccagcagac tgaacaagga cgagtacgag 300

ttcttcacct acctggacaa gcggagcaag cctgtgctgg gcagcatcat caagagcctg 360

agaaacgaca tcggcgccac cgtgcacgag ctgagcagag acaaagaaaa gaacaccgtg 420

ccatggttcc ccaggaccat ccaagagctg gacagattcg ccaaccagat cctgagctat 480

ggcgccgagc tggacgctga tcaccctggc tttaaggacc ccgtgtaccg ggccagaaga 540

aagcagtttg ccgatatcgc ctacaactac cggcacggcc agcctattcc tcgggtcgag 600

tacaccgagg aagagagaaa gacctggggc accgtgttca gaaccctgaa ggccctgtac 660

aagacccacg cctgctacga gcacaaccac atcttcccac tgctggaaaa gtactgcggc 720

ttccgcgagg acaatatccc tcagctcgaa gacgtgtccc agttcctgca gacctgcacc 780

ggctttagac tgaggcctgt tgccggactg ctgagcagca gagattttct cggcggcctg 840

gccttcagag tgttccactg tacccagtac atcagacacg gcagcaagcc catgtacacc 900

cctgagcctg atatctgcca cgagctgctg ggacatgtgc ccctgttcag cgatagaagc 960

ttcgcccagt tcagccaaga gatcggactg gcttctctgg gagcccctga cgagtacatt 1020

gagaagctgg ccaccatcta ctggttcacc gtggaattcg gcctgtgcaa agagggcgac 1080

agcatcaagg cttatggcgc tggactgctg tctagcttcg gcgagctgca gtactgtctg 1140

agcgacaagc ctaagctgct gcccctggaa ctggaaaaga ccgcctgcca agagtacaca 1200

gtgaccgagt tccagcctct gtactacgtg gccgagagct tcaacgacgc caaagaaaaa 1260

gtgcggacct tcgccgccac cattcctcgg ccttttagcg tcagatacga cccctacaca 1320

cagcgcgtgg aagtgctgga caacacacag cagctgaaga ttctggccga ctccatcaac 1380

agcgaagtgg gcattctgtg tcacgccctg cagaagatca agagctga 1428

<210> 197

<211> 4146

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸序列

<400> 197

ctctgggcca gcttgcttgg ggttgccttg acactaagac aagcggcgcg ccgcttgatc 60

ttagtggcac gtcaacccca agcgctggcc cagagccaac cctaattccg gaagtcccgc 120

ccaccggaag tgacgtcaca ggaaatgacg tcacaggaaa tgacgtaatt gtccgccatc 180

ttgtaccgga agtcccgcct accggcggcg accggcggca tctgatttgg tgtcttcttt 240

taaattttgc ggcaattcag tcgataacta taacggtcct aaggtagcga tttaaatacg 300

cgctctctta aggtagcccc gggacgcgtc aattgagatc tggatccggt accgaattcg 360

cggccgcctc gacgactagc gtttagtaat gagacgcaca aactaatatc acaaactgga 420

aatgtctatc aatatatagt tgctctagtt attaatagta atcaattacg gggtcattag 480

ttcatagccc atatatggag ttccgcgtta cataacttac ggtaaatggc ccgcctggct 540

gaccgcccaa cgacccccgc ccattgacgt caataatgac gtatgttccc atagtaacgc 600

caatagggac tttccattga cgtcaatggg tggagtattt acggtaaact gcccacttgg 660

cagtacatca agtgtatcat atgccaagta cgcccccctat tgacgtcaat gacggtaaat 720

ggcccgcctg gcattatgcc cagtacatga ccttatggga ctttcctact tggcagtaca 780

tctacgtatt agtcatcgct attaccatgc atggtcgagg tgagccccac gttctgcttc 840

actctcccca tctccccccc ctccccaccc ccaattttgt atttatttat tttttaatta 900

ttttgtgcag cgatgggggc gggggggggg gggggcgcgc cgccaggcgg ggcggggcgg 960

ggcgaggggc ggggcggggc gaggcggaga ggtgcggcgg cagccaatca gagcggcgcg 1020

ctccgaaagt ttccttttat ggcgaggcgg cggcggcggc ggccctataa aaagcgaagc 1080

gcgcggcggg cgggagtcgc tgcgcgctgc cttcgccccg tgccccgctc cgccgccgcc 1140

tcgcgccgcc cgccccggct ctgactgacc gcgttactcc cacaggtgag cgggcgggac 1200

ggcccttctc cttcgggctg taattagcgc ttggtttaat gacggcttgt ttcttttctg 1260

tggctgcgtg aaagccttga ggggctccgg gagggccctt tgtgcggggg gagcggctcg 1320

gggctgtccg cgggggacg gctgccttcg gggggacgg ggcagggcgg ggttcggctt 1380

ctggcgtgtg accggcggct ctagagcctc tgctaaccat gttcatgcct tcttcttttt 1440

cctacagctc ctgggcaacg tgctggttat tgtgctgtct catcattttg gcaaagaatt 1500

ggatcgcgaa gccgccacca tggactacaa agaccatgac ggtgattata aagatcatga 1560

tatcgattac aaggatgacg atgacaaggc tgctgtggtt ctggaaaatg gcgtgctgag 1620

ccggaagctg agcgacttcg gacaagagac aagctacatc gaggacaaca gcaaccagaa 1680
tggcgccgtg tctctgatct tcagcctgaa agaagaagtg ggcgccctgg ccaaggtgct 1740
gagactgttc gaggaaaacg agatcaatct gacccacatc gagagcagac ccagcagact 1800
gaacaaggac gagtacgagt tcttcaccta cctggacaag cggagcaagc ctgtgctggg 1860
cagcatcatc aagagcctga gaaacgacat cggcgccacc gtgcacgagc tgagcagaga 1920
caaagaaaag aacaccgtgc catggttccc caggaccatc caagagctgg acagattcgc 1980
caaccagatc ctgagctatg gcgccgagct ggacgctgat caccctggct ttaaggaccc 2040
cgtgtaccgg gccagaagaa agcagtttgc cgatatcgcc tacaactacc ggcacggcca 2100
gcctattcct cgggtcgagt acaccgagga agagagaaag acctggggca ccgtgttcag 2160
aaccctgaag gccctgtaca agacccacgc ctgctacgag cacaaccaca tcttcccact 2220
gctggaaaag tactgcggct tccgcgagga caatatccct cagctcgaag acgtgtccca 2280
gttcctgcag acctgcaccg gctttagact gaggcctgtt gccggactgc tgagcagcag 2340
agattttctc ggcggcctgg ccttcagagt gttccactgt acccagtaca tcagacacgg 2400
cagcaagccc atgtacaccc ctgagcctga tatctgccac gagctgctgg gacatgtgcc 2460
cctgttcagc gatagaagct tcgcccagtt cagccaagag atcggactgg cttctctggg 2520
agcccctgac gagtacattg agaagctggc caccatctac tggttcaccg tggaattcgg 2580
cctgtgcaaa gagggcgaca gcatcaaggc ttatggcgct ggactgctgt ctagcttcgg 2640
cgagctgcag tactgtctga gcgacaagcc taagctgctg cccctggaac tggaaaagac 2700
cgcctgccaa gagtacacag tgaccgagtt ccagcctctg tactacgtgg ccgagagctt 2760
caacgacgcc aaagaaaaag tgcggacctt cgccgccacc attcctcggc cttttagcgt 2820
cagatacgac ccctacacac agcgcgtgga agtgctggac aacacacagc agctgaagat 2880
tctggccgac tccatcaaca gcgaagtggg cattctgtgt cacgccctgc agaagatcaa 2940
gagctgagca agtaatgagc gctgatcata atcaacctct ggattacaaa atttgtgaaa 3000
gattgactgg tattcttaac tatgttgctc cttttacgct atgtggatac gctgctttaa 3060
tgcctttgta tcatgctatt gcttcccgta tggctttcat tttctcctcc ttgtataaat 3120
cctggttgct gtctctttat gaggagttgt ggcccgttgt caggcaacgt ggcgtggtgt 3180
gcactgtgtt tgctgacgca acccccactg gttggggcat tgccaccacc tgtcagctcc 3240
tttccgggac tttcgctttc cccctcccta ttgccacggc ggaactcatc gccgcctgcc 3300
ttgcccgctg ctggacaggg gctcggctgt tgggcactga caattccgtg gtgttgtcgg 3360
ggaaatcatc gtcctttcct tggctgctcg cctgtgttgc cacctggatt ctgcgcggga 3420
cgtccttctg ctacgtccct tcggccctca atccagcgga ccttccttcc cgcggcctgc 3480
tgccggctct gcggcctctt ccgcgtcttc gccttcgccc tcagacgagt cggatctccc 3540
tttgggccgc ctccccgctg atcagcctcg actgtgcctt ctagttgcca gccatctgtt 3600
gtttgccccc cccccgtgcc ttccttgacc ctggaaggtg ccactcccac tgtcctttcc 3660
taataaaatg aggaaattgc atcgcattgt ctgagtaggt gtcattctat tctgggggggt 3720
ggggtggggc aggacagcaa gggggaggat tgggaagaca atagcaggca tgctggggat 3780

gcggtgggct ctatggcttc tgaggcggaa agaacgggct cgagaagctt ctagatatcc 3840

tctcttaagg tagcatcgag atttaaatta gggataacag ggtaatggcg cgggccgcaa 3900

aatttaaaag aagacaccaa atcagatgcc gccggtcgcc gccggtaggc gggacttccg 3960

gtacaagatg gcggacaatt acgtcatttc ctgtgacgtc atttcctgtg acgtcacttc 4020

cggtgggcgg gacttccgga attagggttg gctctgggcc agcgcttggg gttgacgtgc 4080

cactaagatc aagcggcgcg ccgcttgtct tagtgtcaag gcaaccccaa gcaagctggc 4140

ccagag 4146

<210> 198

<211> 4214

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸序列

<400> 198

cggtgacgtg tttccggctg ttaggttgac cacgcgcatg ccgcgcggtc agcccaatag 60

ttaagccgga aacacgtcac cggaagtcac atgaccggaa gtcacgtgac cggaaacacg 120

tgacaggaag cacgtgaccg gaactacgtc accggatgtg cgtcaccgga agcatgtgac 180

cggaacttgc gtcacttccc cctcccctga ttggctggtt cgaacgaacg aaccctccaa 240

tgagactcaa ggacaagagg atattttgcg cgccaggaag tggcggcaat tcagtcgata 300

actataacgg tcctaaggta gcgatttaaa tacgcgctct cttaaggtag ccccgggacg 360

cgtcaattga gatctggatc cggtaccgaa ttcgcggccg cctcgacgac tagcgtttag 420

taatgagacg cacaaactaa tatcacaaac tggaaatgtc tatcaatata tagttgctct 480

agttattaat agtaatcaat tacggggtca ttagttcata gcccatatat ggagttccgc 540

gttacataac ttacggtaaa tggcccgcct ggctgaccgc ccaacgaccc ccgcccattg 600

acgtcaataa tgacgtatgt tcccatagta acgccaatag ggactttcca ttgacgtcaa 660

tgggtggagt atttacggta aactgcccac ttggcagtac atcaagtgta tcatatgcca 720

agtacgcccc ctattgacgt caatgacggt aaatggcccg cctggcatta tgcccagtac 780

atgaccttat gggactttcc tacttggcag tacatctacg tattagtcat cgctattacc 840

atgcatggtc gaggtgagcc ccacgttctg cttcactctc cccatctccc ccccctcccc 900

acccccaatt ttgtatttat ttatttttta attattttgt gcagcgatgg gggcgggggg 960

gggggggggg cgcgcgccag gcggggcggg gcggggcgag gggcggggcg gggcgaggcg 1020

gagaggtgcg gcggcagcca atcagagcgg cgcgctccga aagtttcctt ttatggcgag 1080

gcggcggcgg cggcggccct ataaaaagcg aagcgcgcgg cgggcgggag tcgctgcgcg 1140
ctgccttcgc cccgtgcccc gctccgccgc cgcctcgcgc cgcccgcccc ggctctgact 1200
gaccgcgtta ctcccacagg tgagcgggcg ggacggccct tctccttcgg gctgtaatta 1260
gcgcttggtt taatgacggc ttgtttcttt tctgtggctg cgtgaaagcc ttgaggggct 1320
ccgggagggc cctttgtgcg gggggagcgg ctcggggctg tccgcggggg gacggctgcc 1380
ttcgggggggg acggggcagg gcggggttcg gcttctggcg tgtgaccggc ggctctagag 1440
cctctgctaa ccatgttcat gccttcttct ttttcctaca gctcctgggc aacgtgctgg 1500
ttattgtgct gtctcatcat tttggcaaag aattggatcg cgaagccgcc accatggact 1560
acaaagacca tgacggtgat tataaagatc atgatatcga ttacaaggat gacgatgaca 1620
aggctgctgt ggttctggaa aatggcgtgc tgagccggaa gctgagcgac ttcggacaag 1680
agacaagcta catcgaggac aacagcaacc agaatggcgc cgtgtctctg atcttcagcc 1740
tgaaagaaga agtgggcgcc ctggccaagg tgctgagact gttcgaggaa aacgagatca 1800
atctgaccca catcgagagc agacccagca gactgaacaa ggacgagtac gagttcttca 1860
cctacctgga caagcggagc aagcctgtgc tgggcagcat catcaagagc ctgagaaacg 1920
acatcggcgc caccgtgcac gagctgagca gagacaaaga aaagaacacc gtgccatggt 1980
tccccaggac catccaagag ctggacagat tcgccaacca gatcctgagc tatggcgccg 2040
agctggacgc tgatcaccct ggctttaagg accccgtgta ccgggccaga agaaagcagt 2100
ttgccgatat cgcctacaac taccggcacg gccagcctat tcctcgggtc gagtacaccg 2160
aggaagagag aaagacctgg ggcaccgtgt tcagaaccct gaaggccctg tacaagaccc 2220
acgcctgcta cgagcacaac cacatcttcc cactgctgga aaagtactgc ggcttccgcg 2280
aggacaatat ccctcagctc gaagacgtgt cccagttcct gcagacctgc accggcttta 2340
gactgaggcc tgttgccgga ctgctgagca gcagagattt tctcggcggc ctggccttca 2400
gagtgttcca ctgtacccag tacatcagac acggcagcaa gcccatgtac acccctgagc 2460
ctgatatctg ccacgagctg ctgggacatg tgcccctgtt cagcgataga agcttcgccc 2520
agttcagcca agagatcgga ctggcttctc tgggagcccc tgacgagtac attgagaagc 2580
tggccaccat ctactggttc accgtggaat tcggcctgtg caaagagggc gacagcatca 2640
aggcttatgg cgctggactg ctgtctagct tcggcgagct gcagtactgt ctgagcgaca 2700
agcctaagct gctgcccctg gaactggaaa agaccgcctg ccaagagtac acagtgaccg 2760
agttccagcc tctgtactac gtggccgaga gcttcaacga cgccaaagaa aaagtgcgga 2820
ccttcgccgc caccattcct cggcctttta gcgtcagata cgacccctac acacagcgcg 2880
tggaagtgct ggacaacaca cagcagctga agattctggc cgactccatc aacagcgaag 2940
tgggcattct gtgtcacgcc ctgcagaaga tcaagagctg agcaagtaat gagcgctgat 3000
cataatcaac ctctggatta caaaatttgt gaaagattga ctggtattct taactatgtt 3060
gctcctttta cgctatgtgg atacgctgct ttaatgcctt tgtatcatgc tattgcttcc 3120
cgtatggctt tcattttctc ctccttgtat aaatcctggt tgctgtctct ttatgaggag 3180
ttgtggcccg ttgtcaggca acgtggcgtg gtgtgcactg tgtttgctga cgcaaccccc 3240

actggttggg gcattgccac cacctgtcag ctcctttccg ggactttcgc tttccccctc 3300

cctattgcca cggcggaact catcgccgcc tgccttgccc gctgctggac aggggctcgg 3360

ctgttgggca ctgacaattc cgtggtgttg tcggggaaat catcgtcctt tccttggctg 3420

ctcgcctgtg ttgccacctg gattctgcgc gggacgtcct tctgctacgt cccttcggcc 3480

ctcaatccag cggaccttcc ttcccgcggc ctgctgccgg ctctgcggcc tcttccgcgt 3540

cttcgccttc gccctcagac gagtcggatc tccctttggg ccgcctcccc gctgatcagc 3600

ctcgactgtg ccttctagtt gccagccatc tgttgtttgc ccctcccccg tgccttcctt 3660

gaccctggaa ggtgccactc ccactgtcct ttcctaataa aatgaggaaa ttgcatcgca 3720

ttgtctgagt aggtgtcatt ctattctggg gggtggggtg gggcaggaca gcaaggggga 3780

ggattgggaa gacaatagca ggcatgctgg ggatgcggtg ggctctatgg cttctgaggc 3840

ggaaagaacg ggctcgagaa gcttctagat atcctctctt aaggtagcat cgagatttaa 3900

attagggata acagggtaat ggcgcgggcc gccacttcct ggcgcgcaaa atatcctctt 3960

gtccttgagt ctcattggag ggttcgttcg ttcgaaccag ccaatcaggg gaggggggaag 4020

tgacgcaagt tccggtcaca tgcttccggt gacgcacatc cggtgacgta gttccggtca 4080

cgtgcttcct gtcacgtgtt tccggtcacg tgacttccgg tcatgtgact tccggtgacg 4140

tgtttccggc ttaactattg ggctgaccgc gcggcatgcg cgtggtcaac ctaacagccg 4200

gaaacacgtc accg 4214

【發明申請專利範圍】

【第1項】一種核酸分子，其包含：

側接於基因盒的第一反向末端重複體（ITR）和第二 ITR，所述基因盒包含異源多核苷酸序列，其中所述第一 ITR 由 SEQ ID NO: 180 所示的核苷酸序列組成以及所述第二 ITR 由 SEQ ID NO: 181 所示的核苷酸序列組成；或所述第一 ITR 由 SEQ ID NO: 183 所示的核苷酸序列組成以及所述第二 ITR 由 SEQ ID NO: 184 所示的核苷酸序列組成。

【第2項】如請求項 1 的核酸分子，其還包含啟動子，其中：

所述啟動子定位於所述異源多核苷酸序列的 5'；

所述啟動子是組織特異性啟動子；

所述啟動子驅動所述異源多核苷酸序列在選自由下列組成之群組的器官中的表現：肌肉、中樞神經系統（CNS）、眼睛、肝臟、心臟、腎臟、胰腺、肺、皮膚、膀胱、泌尿道或其任何組合；

所述啟動子驅動所述異源多核苷酸序列在以下細胞中的表現：肝細胞、內皮細胞、心肌細胞、骨骼肌細胞、竇狀細胞、傳入神經元、傳出神經元、中間神經元、神經膠質細胞、星形膠質細胞、少突膠質細胞、小神經膠質細胞、室管膜細胞、肺上皮細胞、許旺(Schwann)細胞、衛星細胞、感光細胞、視網膜神經節細胞或其任何組合；或

所述啟動子選自由下列組成之群組：小鼠甲狀腺素啟動子（mTTR）、內源人類 VIII 因子啟動子（F8）、人類 α-1-抗胰蛋白酶啟動子（hAAT）、人類白蛋白最小啟動子、小鼠白蛋白啟動子、三重四脯胺酸（TTP）啟動子、CASI 啟動子、CAG 啟動子、巨細胞病毒（CMV）啟動子、α1-抗胰蛋白酶（AAT）、肌肉肌酸激酶（MCK）、肌球蛋白重鏈 α（αMHC）、肌紅蛋白（MB）、結蛋白（DES）、SPc5-12、2R5Sc5-12、dMCK、tMCK 和磷酸甘油酸激酶（PGK）啟動子。

【第3項】如請求項 2 的核酸分子，其中：

所述異源多核苷酸序列還包含內含子序列，其中：

所述內含子序列定位於所述異源多核苷酸序列的 5'；

所述內含子序列定位於所述啟動子的 3'；

所述內含子序列包含合成內含子序列；或

所述內含子序列包含 SEQ ID NO: 115 或 192；

所述異源多核苷酸序列編碼凝血因子、生長因子、激素、細胞激素、抗體、或其片段。

【第4項】如請求項 1 的核酸分子，其中：

所述基因盒還包含轉錄後調控元件，其中：

所述轉錄後調控元件定位於所述異源多核苷酸序列的 3'；

所述轉錄後調控元件包含突變的土撥鼠肝炎病毒轉錄後調控元件（WPRE）、微小 RNA 結合位點、DNA 核靶向序列或其任何組合，或

所述轉錄後調控元件包含微小 RNA 結合位點，其包含與 miR142-3p 的結合位點；或

所述基因盒還包含 3'UTR 多聚(A)尾序列，其中所述 3'UTR 多聚(A)尾序列選自由下列組成之群組：bGH 多聚(A)、肌動蛋白多聚(A)、血紅蛋白多聚(A)、包含 bGH 多聚(A)的 3'UTR 多聚(A)尾序列及其任何組合；

其中所述基因盒還包含強化子序列，其中所述強化子序列定位於所述第一 ITR 與所述第二 ITR 之間；

所述基因盒包含單鏈核酸；或

所述基因盒包含雙鏈核酸。

【第5項】如請求項 1 的核酸分子，其中所述核酸分子從 5'至 3'包含：所述第一 ITR、所述基因盒和所述第二 ITR；其中所述基因盒包含組織特異性啟動子序列、內含子序列、異源多核苷酸序列、轉錄後調控元件和 3'UTR 多聚(A)尾序列；或

所述基因盒從 5'至 3'包含：組織特異性啟動子序列、內含子序列、所述異源多核苷酸序列、轉錄後調控元件和 3'UTR 多聚(A)尾序列，其中：

(a) 所述組織特異性啟動子序列包含 TTP 啟動子；

(b) 所述內含子是合成內含子；

(c) 所述轉錄後調控元件包含 WPRE；並且

(d) 所述 3'UTR 多聚(A)尾序列包含 bGHp 多聚(A)。

【第6項】如請求項 3 的核酸分子，其中所述凝血因子選自由下列組成之群組：I 因子（FI）、II 因子（FII）、III 因子（FIII）、IV 因子（FVI）、V 因子（FV）、VI 因子（FVI）、VII 因子（FVII）、VIII 因子（FVIII）、IX 因子（FIX）、X 因子（FX）、XI 因子（FXI）、XII 因子（FXII）、XIII 因子（FVIII）、血管性血友病(Von Willebrand)因子（VWF）、前微血管增滲素、高分子量激肽原、纖連蛋白、抗凝血酶 III、肝素輔因子 II、蛋白質 C、蛋白質 S、蛋白質 Z、蛋白質 Z 相關蛋白酶抑制劑（ZPI）、纖維蛋白溶酶原、α2-抗纖維蛋白溶酶、組織纖維蛋白溶酶原活化劑（tPA）、尿激酶、纖維蛋白溶酶原活化物抑制劑-1（PAI-1）、纖維蛋白溶酶原活化物抑制劑-2（PAI2）及其任何組合；所述凝血因子是全長成熟 FVIII、包含 SEQ ID NO: 106 的胺基酸序列之 FVIII，包含 A1 結構域、A2 結構域、A3 結構域、C1 結構域、C2 結構域、部分 B 結構域或無 B 結構域之 FVIII，或包含 SEQ ID NO:109 的胺基酸序列之 FVIII；或

所述凝血因子包含異源部分。

【第7項】如請求項 1 的核酸分子，其中：

所述核酸分子是用遞送劑配製，其中所述遞送劑包含脂質奈米顆粒，或所述遞送劑選自由下列組成之群組：脂質體、非脂質聚合分子和內體及其任何組合；或

所述核酸分子經配製用於靜脈內、經皮、真皮內、皮下、經肺或口服遞送或其任何組合，或其中所述核酸分子經配製用於靜脈內遞送。

【第8項】一種載體，其包含如請求項 1 的核酸分子。

【第9項】一種宿主細胞，其包含如請求項 1 的核酸分子。

【第10項】一種醫藥組合物，其包含如請求項 1 的核酸分子和醫藥上可接受的賦形劑。

【第11項】一種醫藥組合物，其包含如請求項 9 的宿主細胞和醫藥上可接受的賦形劑。

【第12項】一種套組，其包含如請求項 1 的核酸分子和用於將所述核酸分子投予有需要的個體的說明書。

【第13項】 如請求項 1 的核酸分子，其中所述核酸分子是在昆蟲細胞中產生。

【第14項】 一種奈米顆粒遞送系統，其包含如請求項 1 的核酸分子。

【第15項】 一種產生多肽的方法，其包括在適宜條件下培養如請求項 9 的宿主細胞和回收所述多肽，其中所述多肽係具有凝血活性的多肽且該方法包含回收具有凝血活性的所述多肽。

【第16項】 一種如請求項 1 的核酸分子用於製備在有需要的個體體內表現如請求項 1 所記載之異源多核苷酸序列的藥物的用途。

【第17項】 一種如請求項 3 的核酸分子用於製備在有需要的個體體內表現凝血因子的藥物的用途。

【第18項】 一種如請求項 6 的核酸分子用於製備治療有需要的個體的凝血因子缺乏的藥物的用途。

【第19項】 如請求項 18 的用途，：

其中所述核酸分子是靜脈內、經皮、真皮內、皮下、口服、經肺或其任何組合地投予，

其中所述核酸分子是靜脈內投予；

其中所述核酸分子與第二藥劑被投予所述個體；

其中所述個體是哺乳動物；

其中所述個體是人類；

其中相對於所述投予前所述個體體內的 FVIII 活性，所述核酸分子導致增加的 FVIII 活性；

其中所述個體患有出血障礙，

其中所述個體具有血友病，或

其中所述個體具有血友病 A。

【第20項】 一種核酸分子用於製備治療有需要的個體的出血障礙的藥物的用途，其中所述出血障礙是血友病 A，其中所述核酸分子包含側接於基因盒的第一反向末端重複體（ITR）和第二 ITR，所述基因盒包含編碼凝血因子的異源多核苷酸序列，其中所述第一 ITR 由 SEQ ID NO: 180 所示的核苷酸序列組成以及所述第二 ITR 由 SEQ ID NO: 181 所示的核苷酸序列組

成；或所述第一 ITR 由 SEQ ID NO: 183 所示的核苷酸序列組成以及所述第二 ITR 由 SEQ ID NO: 184 所示的核苷酸序列組成。

【第21項】 一種核酸分子用於製備治療有需要的個體的肝臟代謝障礙的藥物的用途，其中所述核酸分子包含側接於基因盒的第一反向末端重複體（ITR）和第二 ITR，所述基因盒包含編碼在所述個體體內缺乏的肝臟相關代謝酶的異源多核苷酸序列，其中所述第一 ITR 由 SEQ ID NO: 180 所示的核苷酸序列組成以及所述第二 ITR 由 SEQ ID NO: 181 所示的核苷酸序列組成；或所述第一 ITR 由 SEQ ID NO: 183 所示的核苷酸序列組成以及所述第二 ITR 由 SEQ ID NO: 184 所示的核苷酸序列組成。

【第22項】 如請求項 21 的用途，

其中所述肝臟代謝障礙選自苯丙酮尿（PKU）、尿素循環疾病、溶酶體貯積障礙和糖原貯積病，其中所述肝臟代謝障礙是苯丙酮尿（PKU）；

其中所述核酸分子是靜脈內、經皮、真皮內、皮下、口服、經肺或其任何組合地投予，

其中所述核酸分子是靜脈內投予；

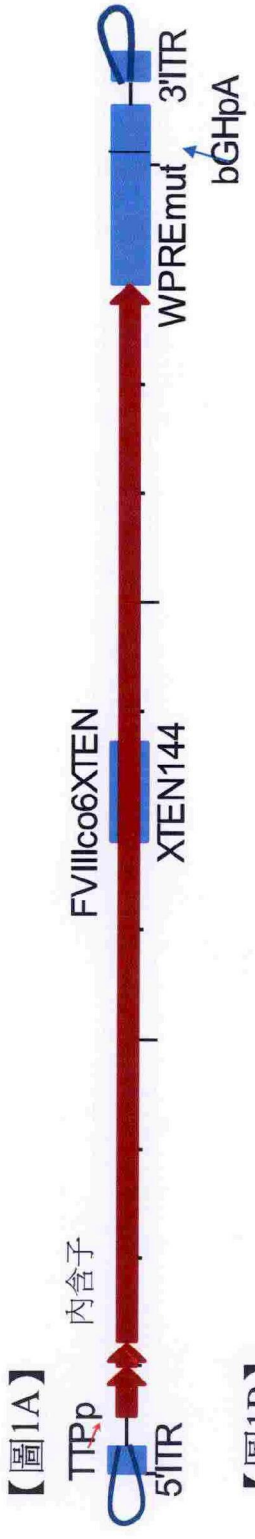
其中所述核酸分子與第二藥劑被投予所述個體；

其中所述個體是哺乳動物；或

其中所述個體是人類。

【發明圖式】

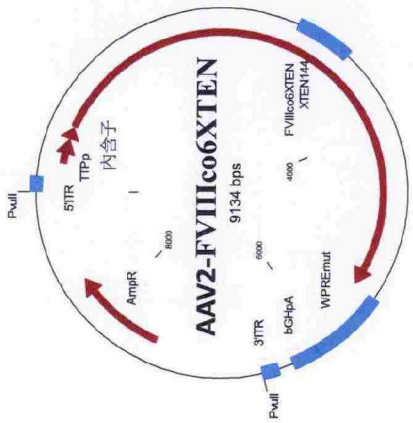
【圖1A】



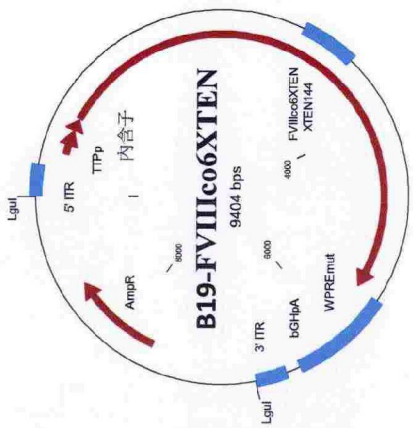
【圖1B】



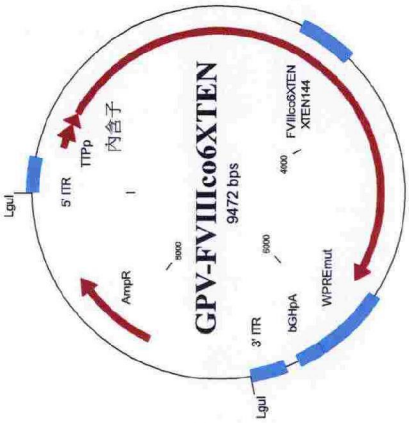
【圖1C】



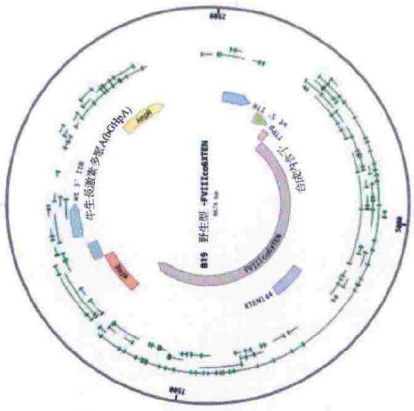
【圖1D】



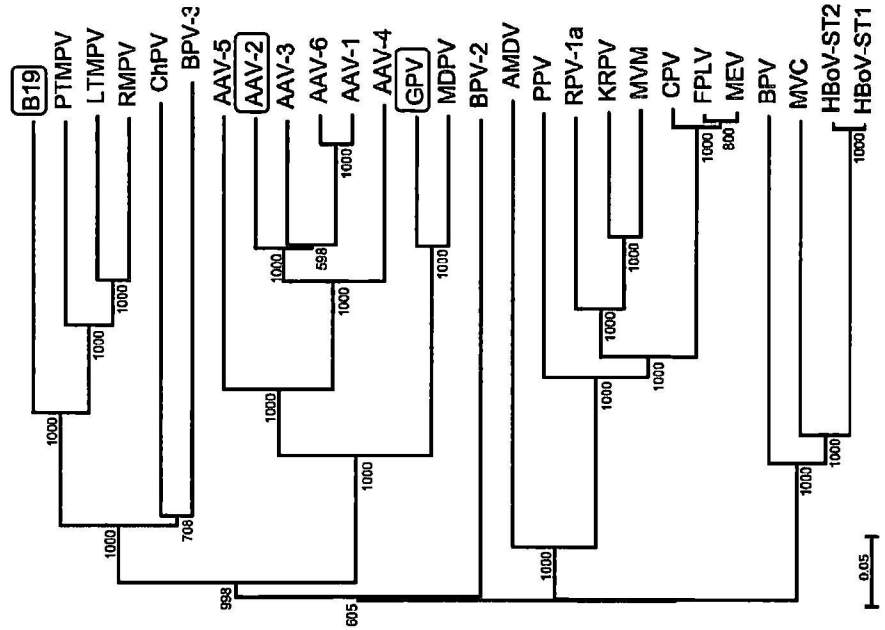
【圖1E】



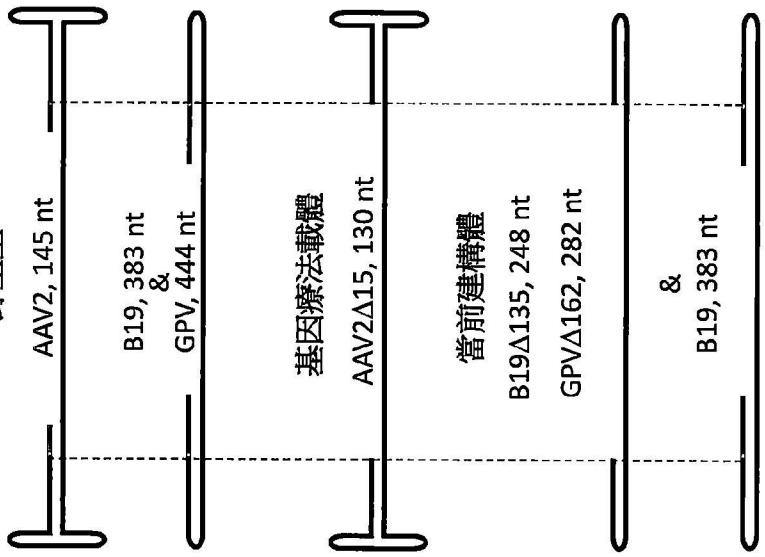
【圖1F】



【圖2A】


B19
PTMPV
LTMPV
RMPV
ChPV
BPV-3
AAV-5
AAV-2
AAV-3
AAV-6
AAV-1
AAV-4
GPV
MDPV
BPV-2
AMDV
PPV
RPV-1a
KRPV
MVM
CPV
FPLV
MEV
BPV
MVC
HBoV-ST2
HBoV-ST1
1000
998
605
708
598
800
0.05

【圖2B】


野生型
AAV2, 145 nt
B19, 383 nt
&
GPV, 444 nt
基因療法載體
AAV2Δ15, 130 nt
當前建構體
B19Δ135, 248 nt
GPVΔ162, 282 nt
&
B19, 383 nt

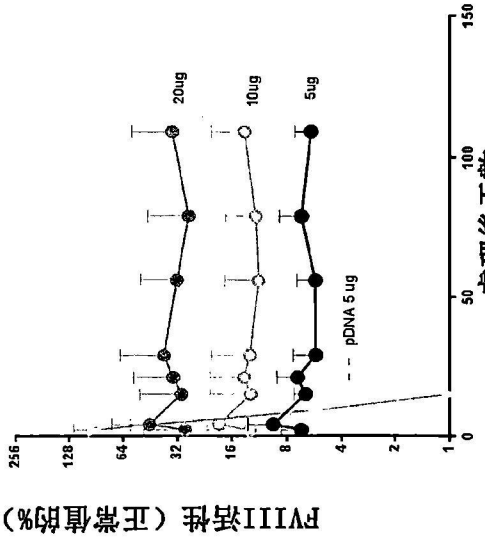
【圖3A】


B19 ⟶ 1 ccaaatcagatgccgc---cggtcg---ccgcc------ggtaggcgggact---------tccggt-acaagatggcggacaattacgtc---------atttcctgtgacgtcatttcctgtgacgtcacttccggtgggcgggacttccggaattagggttggct--
GPV ⟶ 1 -ctcattggagggttcgttcgttcgaaccagccaatcaggggaggggggaagtgacgcaagttccggtcacatgcttccggtgacgcacatccggtgacgtagttccggtcacgt-gcttcctgtcacgt-gtttccggtc-acgtgacttccggtcat----gtgacttc
AAV2 ⟶ 1 ----------ttggc----------cactcc----------ct--
138 ctgggccagcttgcttggggttgccttgacactaagacaagcggcgcgc-cgcttgatcttagtggcacgtcaaccccaagcgctggcccagagccaaccctaatttccggaagtcccgcccaccggaagtgacgtcacaggaaatgacgtcacaggaaatgacgtcacaggaaatgacgtaattg
163 cggtgacgtgtttccggctgttaggttgac-------cacgc-gcatgc-cgcgcggtc--------------agcccaatagttaagccggaaaca--cgtca--ccggaagtcac-atgaccggaagtcacgtgaccggaaa-cacgtgacaggaa-----gcacgtg
14 ct----------ctgcgcgctcgctcgctcactgaggccgggcgacc---aaaggtcgcccgacgcccgggctttgcccgggcggcc----------tca----------g
307 tccgccatcttgt-accggaagtcccgcctaccggcggc----gaccgg------cggcat----------ctgatttggtgtcttctttaa----------atttt----------
299 accggaactacgtcaccggatgtgcgtc--accggaagcatgtgaccggaacttgcgtcacttcccctccctgattggctggttcgaacgaacgaaccctccaatgagactcaaggacaagaggatattttgcgcgccaggaagtg
92 tgagcgagcgag----------cgcgcagagagggagt-----ggccaa------ctccat----------c--actagggg--ttcct----------

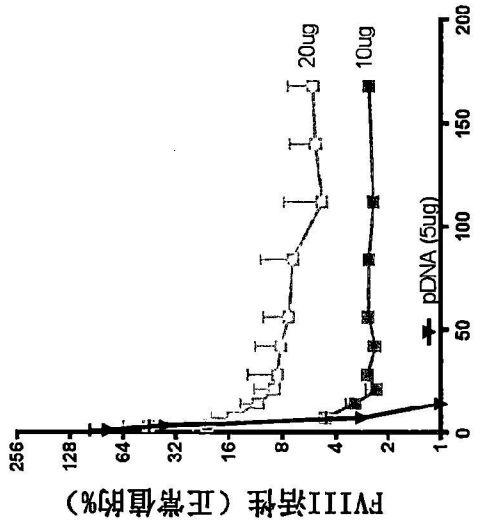
【圖3B】


B19 ⟶ 1 ccaaatcagatgccgc---cggtcg---ccgcc------ggtaggcgggact---------tccggt-acaagatggcggacaattacgtc---------atttcctgtgacgtcatttcctgtgacgtcacttccggtgggcgggacttccggaattagggttggct--
GPV ⟶ 1 -ctcattggagggttcgttcgttcgaaccagccaatcaggggaggggggaagtgacgcaagttccggtcacatgcttccggtgacgcacatccggtgacgtagttccggtcacgt-gcttcctgtcacgt-gtttccggtc-acgtgacttccggtcat----gtgacttc
138 ctgggccagcttgcttggggttgccttgacactaagacaagcggcgcgccgcttgatcttagtggcacgtcaaccccaagcgctggcccagagccaaccctaatttccggaagtcccgcccaccggaagtgacgtcacaggaaatgacgtcacaggaaatgacgtaattgt
163 cggtgacgtgtttccggctgttaggttgac-------cacgc-gcatgccgcgcggtc--------------agcccaatagttaagccggaaaca--cgtca--ccggaagtcac-atgaccggaagtcacgtgaccggaaa-cacgtgacaggaa-----gcacgtga
308 ccgccatcttgt-accggaagtcccgcctaccggcggc----gaccgg------cggcat----------ctgatttggtgtcttctttaa----------atttt----------
300 ccggaactacgtcaccggatgtgcgtc--accggaagcatgtgaccggaacttgcgtcacttcccctccctgattggctggttcgaacgaacgaaccctccaatgagactcaaggacaagaggatattttgcgcgccaggaagtg

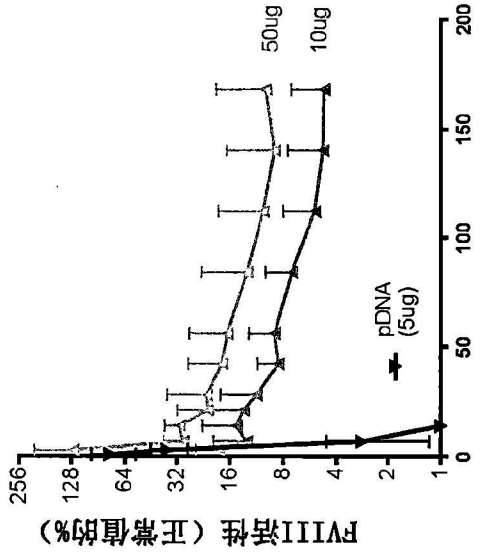
【圖4A】


FVIII活性（正常值的%）
256
128
64
32
16
8
4
2
1
0
50
100
150
處理後天數
20ug
10ug
5ug
pDNA 5 ug
ssAAV-FVIII

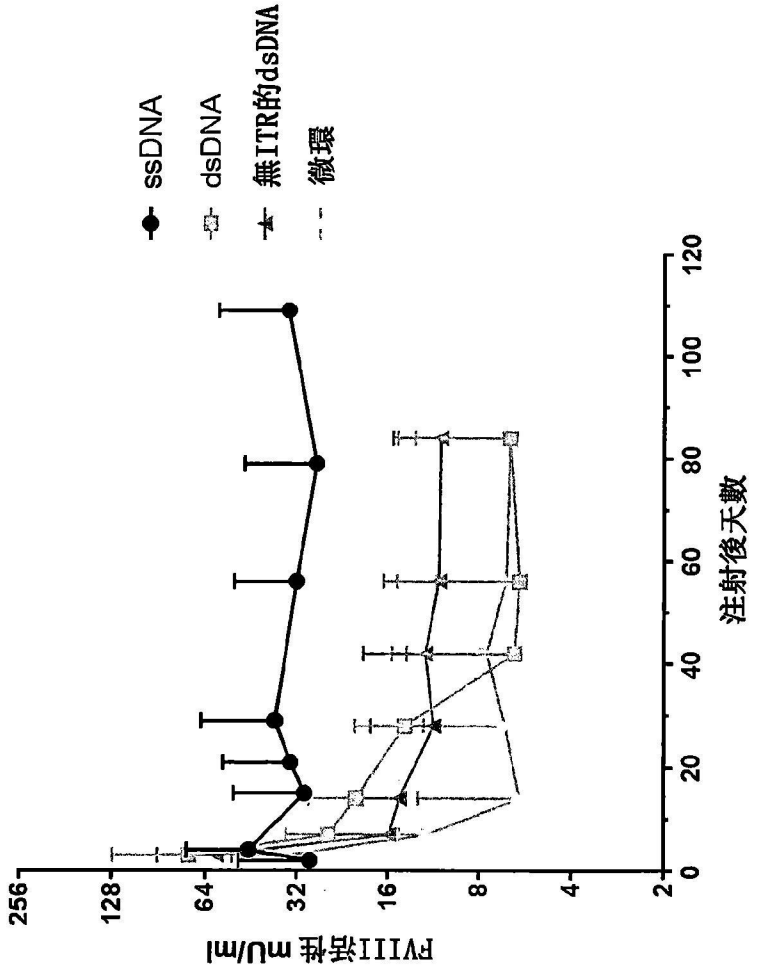
【圖4B】


FVIII活性（正常值的%）
256
128
64
32
16
8
4
2
1
0
50
100
150
200
處理後天數
20ug
10ug
pDNA (5ug)
ssB19-FVIII

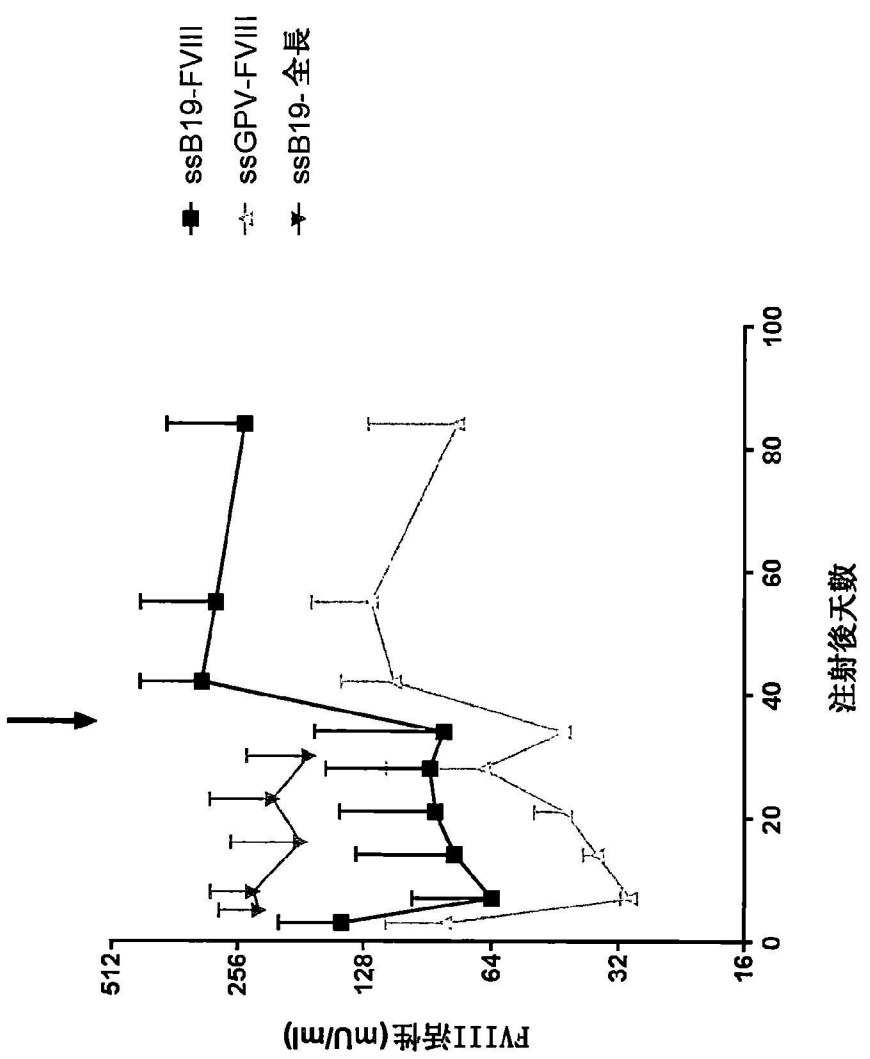
【圖4C】


FVIII活性（正常值的%）
256
128
64
32
16
8
4
2
1
0
50
100
150
200
處理後天數
50ug
10ug
pDNA
(5ug)
ssGPV-FVIII

【圖5】


ssDNA
dsDNA
無ITR的dsDNA
微環
FVIII活性 mU/ml
256
128
64
32
16
8
4
2
0
20
40
60
80
100
120
注射後天數

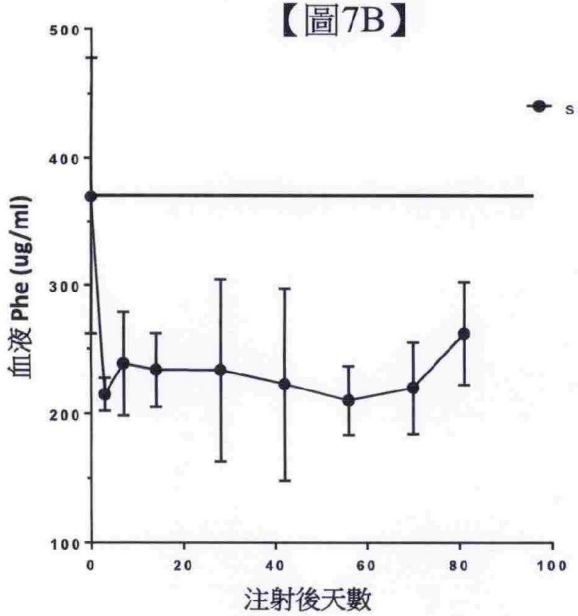
【圖6】


ssB19-FVIII
ssGPV-FVIII
ssB19-全長
FVIII活性(mU/ml)
512
256
128
64
32
16
0
20
40
60
80
100
注射後天數

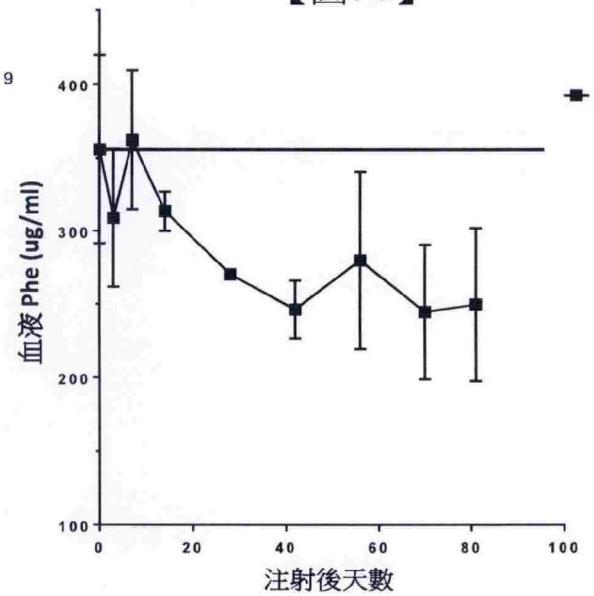
【圖7A】



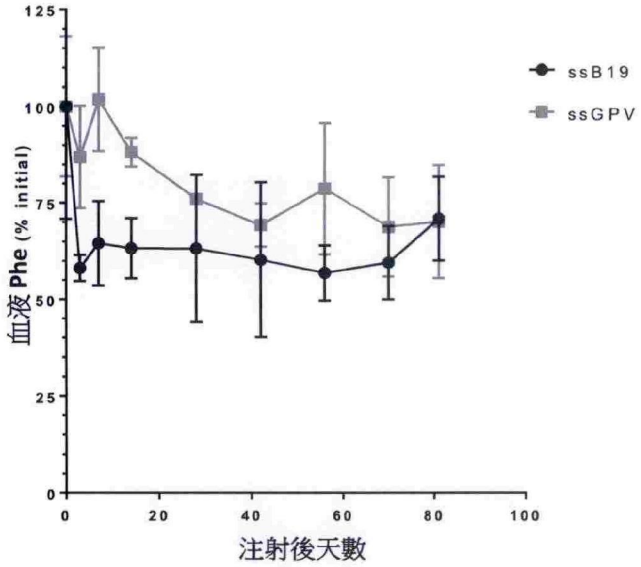
【圖7B】



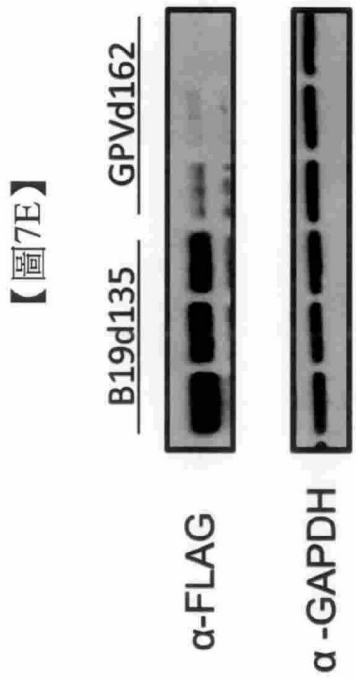
【圖7C】



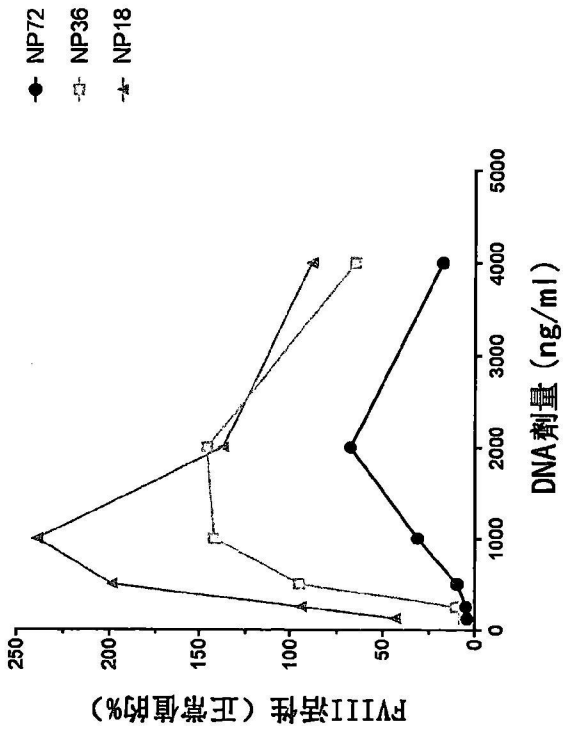
【圖7D】



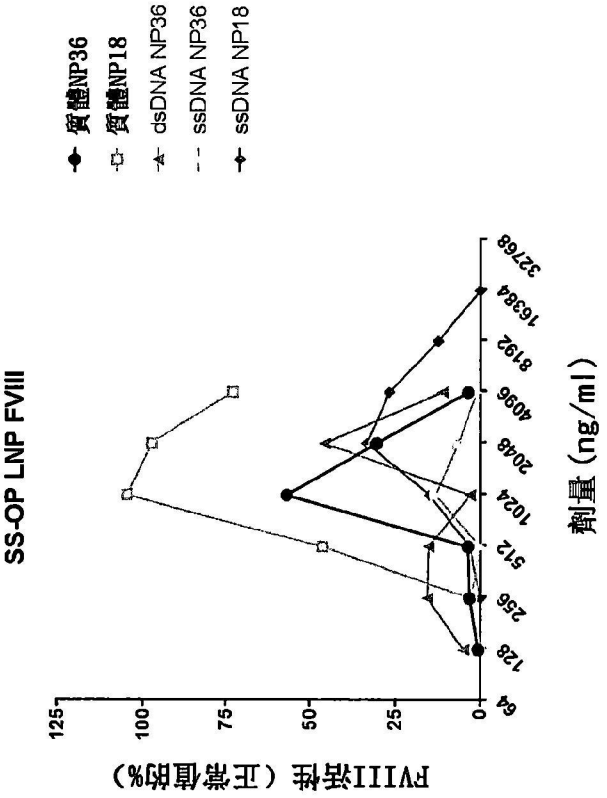
【圖7E】



【圖8A】


NP72
NP36
NP18
FVIII活性（正常值的%）
DNA劑量（ng/ml）
0
50
100
150
200
250
0
1000
2000
3000
4000
5000

【圖8B】


SS-OP LNP FVIII
質體NP36
質體NP18
dsDNA NP36
ssDNA NP36
ssDNA NP18
FVIII活性（正常值的%）
劑量（ng/ml）
0
25
50
75
100
125
64
128
256
512
1024
2048
4096
8192
16384
32768