



(12) 发明专利



(10) 授权公告号 CN 109072214 B

(45) 授权公告日 2023. 02. 17

(21) 申请号 201780020889.1	(73) 专利权人 比奥贝拉蒂治疗公司
(22) 申请日 2017.01.31	地址 美国马萨诸塞州
(65) 同一申请的已公布的文献号 申请公布号 CN 109072214 A	(72) 发明人 谭思元 刘彤瑶
(43) 申请公布日 2018.12.21	(74) 专利代理机构 北京坤瑞律师事务所 11494 专利代理师 封新琴
(30) 优先权数据 62/289,696 2016.02.01 US 62/409,739 2016.10.18 US	(51) Int.Cl. C12N 9/64 (2006.01)
(85) PCT国际申请进入国家阶段日 2018.09.28	(56) 对比文件 W0 2016004113 A1,2016.01.07 CN 104427995 A,2015.03.18 W0 2014127215 A1,2014.08.21
(86) PCT国际申请的申请数据 PCT/US2017/015879 2017.01.31	审查员 段珊
(87) PCT国际申请的公布数据 W02017/136358 EN 2017.08.10	

权利要求书1页 说明书85页
序列表81页 附图46页

(54) 发明名称
优化的因子VIII基因

(57) 摘要
本公开提供密码子优化的因子VIII序列、包含密码子优化的因子VIII序列的载体和宿主细胞、由密码子优化的因子VIII序列编码的多肽和产生此类多肽的方法。本公开还提供治疗如血友病等出血病症的方法,所述方法包括向受试者施用密码子优化的因子VIII核酸序列或由其编码的多肽。

1. 一种分离的核酸分子,其核苷酸序列编码具有FVIII活性的多肽且如SEQ ID NO:71的核苷酸58-2277和2320-4374或SEQ ID NO:71的核苷酸58-4374所示。

2. 如权利要求1所述的分离的核酸分子,其中所述核苷酸序列连接至编码信号肽的核酸序列,其中所述编码信号肽的核酸序列包含:

- (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。

3. 如权利要求1或2所述的分离的核酸分子,其中所述核苷酸序列连接至编码异源氨基酸序列的异源核苷酸序列,其中所述异源氨基酸序列是免疫球蛋白恒定区或其一部分、XTEN、转铁蛋白、白蛋白或PAS序列。

4. 如权利要求3所述的分离的核酸分子,其中所述异源氨基酸序列连接至由所述核苷酸序列编码的氨基酸序列的N端或C端或插入由所述核苷酸序列编码的氨基酸序列中的在选自如SEQ ID NO:15所示的成熟人FVIII的位置3,18,22,26,40,60,65,81,116,119,130,188,211,216,220,224,230,333,336,339,375,378,399,403,409,416,442,487,490,494,500,518,588,603,713,745,1656,1711,1720,1725,1749,1796,1802,1827,1861,1896,1900,1904,1905,1910,1937,2019,2068,2111,2120,2171,2188,2227和2332的一个或多个插入位点处的两个氨基酸之间。

5. 一种分离的核酸分子,其核苷酸序列编码具有FVIII活性的FVIII多肽且如SEQ ID NO:72所示。

6. 一种载体,其包含如权利要求1至5中任一项所述的核酸分子。

7. 如权利要求6所述的载体,其中所述载体是慢病毒载体。

8. 一种宿主细胞,其包含如权利要求1至5中任一项所述的核酸分子或如权利要求6或7所述的载体。

9. 一种产生具有FVIII活性的多肽的方法,所述方法包括:在一定条件下培养如权利要求8所述的宿主细胞,由此产生具有FVIII活性的多肽;和回收所述具有FVIII活性的多肽。

10. 如权利要求1至5中任一项所述的核酸分子或如权利要求6或7所述的载体用于制备治疗出血病症的药物的用途。

优化的因子VIII基因

[0001] 以电子方式提交的序列表的引用

[0002] 呈ASCII文本文件形式的以电子方式提交的序列表(名称:2159_469PC02_SequenceListing_ST25.txt;大小:204,138字节;和创建日期:2017年1月31日)的内容以引用的方式整体并入本文中。

背景技术

[0003] 凝血途径部分地涉及血小板的表面上因子VIIIa(FVIIIa)和因子IXa(FIXa)的酶复合物(Xase复合物)的形成。FIXa是一种丝氨酸蛋白酶,在其辅因子FVIIIa不存在下具有相对微弱催化活性。所述Xase复合物将因子X(FX)裂解成因子Xa(FXa),所述因子Xa(FXa)又与因子Va(FVa)相互作用以裂解凝血酶原并且产生凝血酶。血友病A是由FVIII(FVIII)基因中的突变和/或缺失,从而导致FVIII活性的缺陷而引起的出血病症(Peyvandi等人2006)。在一些情况下,患者由于如抗FVIII抗体等FVIII抑制剂的存在而具有降低的FVIII水平。

[0004] 血友病A的特征在于自发出血和过多出血。通常在儿童早期开始的重复出血至肌肉和关节中随时间导致血友病性关节病和不可逆关节损害。这种损害是进行性的并且可导致严重受限的关节活动性、肌肉萎缩和慢性疼痛(Rodriguez-Merchan, E.C., Semin. Thromb. Hemost. 29:87-96 (2003), 其以引用的方式整体并入本文中)。

[0005] 所述疾病可通过旨在使FVIII活性恢复至正常水平的1至5%以预防自发出血的替代疗法来治疗(参见例如Mannucci, P.M.等人, N. Engl. J. Med. 344:1773-9 (2001), 以引用的方式整体并入本文中)。存在可用于按需要治疗出血事件或用于通过预防性治疗来预防出血事件发生的来源于血浆的重组FVIII产品。基于这些产品的半衰期(10-12小时)(White G.C.等人, Thromb. Haemost. 77:660-7 (1997); Morfini, M., Haemophilia 9(增刊1):94-99; discussion 100 (2003)), 治疗方案需要频繁静脉内施用, 通常每周两次至三次用于预防并且每天一次至三次用于按需要治疗(Manco-Johnson, M.J.等人, N. Engl. J. Med. 357:535-544 (2007)), 其中每一者均以引用的方式整体并入本文中。所述频繁施用是不便的并且昂贵的。

[0006] 向患者提供低成本重组FVIII蛋白的主要障碍是商业制造的高成本。FVIII蛋白在异源表达系统中表达不良, 比相似大小的蛋白质低两个至三个数量级。(Lynch等人, Hum. Gene. Ther.; 4:259-72 (1993)。FVIII的不良表达部分地归因于FVIII编码序列中存在抑制FVIII表达的顺式作用元件, 如转录沉默子元件(Hoeben等人, Blood 85:2447-2454 (1995))、基质附着样序列(MAR)(Fallux等人, Mol. Cell. Biol. 16:4264-4272 (1996))和转录延伸抑制元件(Koeberl等人, Hum. Gene. Ther.; 6:469-479 (1995))。

[0007] 我们对FVIII表达的生物学的理解的进展已经引起了更有效FVIII变体的开发。例如, 生物化学研究证明了FVIII B结构域对于FVIII辅因子活性并非必需的。B结构域的缺失导致了mRNA水平相对全长野生型FVIII的17倍增加和分泌蛋白的30%增加。(Toole等人, Proc Natl Acad Sci USA 83:5939-42 (1986))。这引起了B结构域缺失(BDD)的FVIII蛋白浓缩物的开发, 所述B结构域缺失(BDD)的FVIII蛋白浓缩物现在广泛用于临床中。然而, 最

近的研究指示了ER腔中的全长和BDD hFVIII错误折叠,导致未折叠蛋白反应(UPR)的活化和鼠科动物肝细胞的细胞凋亡。

[0008] 因此,所属领域中需要在异源系统中有效地表达的FVIII序列。

发明内容

[0009] 公开了编码具有FVIII活性的多肽的密码子优化的核酸分子。一方面,本公开提供包含如下核苷酸序列的分离的核酸分子,所述核苷酸序列包含编码因子VIII (FVIII) 多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列;其中所述第一核酸序列与(i) SEQ ID NO:3的核苷酸58-2277和2320-1791或(ii) SEQ ID NO:4的核苷酸58-2277和2320-1791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;并且其中所述N端部分和所述C端部分合起来具有FVIII多肽活性。

[0010] 本公开还提供包含如下核苷酸序列的分离的核酸分子,所述核苷酸序列包含编码FVIII多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列;其中所述第二核酸序列与(i) SEQ ID NO:5的核苷酸1792-2277和2320-4374或(ii) SEQ ID NO:6的1792-2277和2320-4374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;并且其中所述N端部分和所述C端部分合起来具有FVIII多肽活性。

[0011] 在一些实施方案中,本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与SEQ ID NO:1的核苷酸58-2277和2320-4374具有至少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。在其他实施方案中,本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与SEQ ID NO:2的核苷酸58-2277和2320-4374具有至少约94%、至少约95%、至少96%、至少97%、至少98%或至少99%序列同一性的核酸序列。在一个实施方案中,本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与SEQ ID NO:70的核苷酸58-2277和2320-4374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。在另一实施方案中,本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与SEQ ID NO:71的核苷酸58-2277和2320-4374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。

[0012] 在一些实施方案中,本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与SEQ ID NO:3的核苷酸58-2277和2320-4374具有至少约92%、至少约93%、至少约94%、至少约95%、至少96%、至少97%、至少98%或至少99%序列同一性的核酸序列。

[0013] 在一些实施方案中,本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与SEQ ID NO:4的核苷酸58-2277和2320-4374具有至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少96%、至少97%、至少98%或至少99%序列同一性的核酸序列。

[0014] 在一些实施方案中,本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的

分离的核酸分子,其中所述核苷酸序列包含与SEQ ID NO:5的核苷酸58-2277和2320-4374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。

[0015] 在一些实施方案中,本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与SEQ ID NO:6的核苷酸58-2277和2320-4374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。

[0016] 在一些实施方案中,本公开提供一种增加具有FVIII活性的多肽在受试者中的表达的方法,所述方法包括向有需要的受试者施用本文所公开的分离的核酸分子或载体,其中所述多肽的表达相对于包含SEQ ID NO:16的参考核酸分子或包含所述参考核酸分子的载体增加。

[0017] 在一些实施方案中,本公开提供一种治疗出血病症的方法,所述方法包括:向有需要的受试者施用本文所公开的核酸分子、载体或多肽。

[0018] 实施方案

[0019] E1.一种包含如下核苷酸序列的分离的核酸分子,所述核苷酸序列包含编码因子VIII (FVIII) 多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列;其中所述第一核酸序列与(i) SEQ ID NO:3的核苷酸58-2277和2320-1791或(ii) SEQ ID NO:4的核苷酸58-2277和2320-1791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;并且其中所述N端部分和所述C端部分合起来具有FVIII多肽活性。

[0020] E2.E1的分离的核酸分子,其中所述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。

[0021] E3.E1或E2的分离的核酸分子,其中所述第二核苷酸序列与SEQ ID NO:3的核苷酸1792-2277和2320-4374或SEQ ID NO:4的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。

[0022] E4.E1或E2的分离的核酸分子,其中所述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。

[0023] E5.一种包含如下核苷酸序列的分离的核酸分子,所述核苷酸序列包含编码FVIII多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列;其中所述第二核酸序列与(i) SEQ ID NO:5的核苷酸1792-2277和2320-4374或(ii) SEQ ID NO:6的1792-2277和2320-4374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;并且其中所述N端部分和所述C端部分合起来具有FVIII多肽活性。

[0024] E6.E5的分离的核酸分子,其中所述第二核苷酸序列包含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。

[0025] E7.E5或E6的分离的核酸分子,其中所述第一核苷酸序列与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。

[0026] E8.E5或E6的分离的核酸分子,其中所述第一核苷酸序列包含SEQ ID NO:5的核苷酸58-1791或SEQ ID NO:6的核苷酸58-1791。

[0027] E9.一种包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与SEQ ID NO:1的核苷酸58-2277和2320-4374具有至少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。

[0028] E10.E9的分离的核酸分子,其中所述核苷酸序列包含SEQ ID NO:1的核苷酸58-2277和2320-4374。

[0029] E11.一种包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与SEQ ID NO:2的核苷酸58-2277和2320-4374具有至少约94%、至少约95%、至少96%、至少97%、至少98%或至少99%序列同一性的核酸序列。

[0030] E12.E10的分离的核酸分子,其中所述核苷酸序列包含SEQ ID NO:2的核苷酸58-2277和2320-4374。

[0031] E13.一种包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与SEQ ID NO:70的核苷酸58-2277和2320-4374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。

[0032] E14.E13的分离的核酸分子,其中所述核苷酸序列包含SEQ ID NO:70的核苷酸58-2277和2320-4374。

[0033] E15.一种包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与SEQ ID NO:71的核苷酸58-2277和2320-4374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。

[0034] E16.E15的分离的核酸分子,其中所述核苷酸序列包含SEQ ID NO:71的核苷酸58-2277和2320-4374。

[0035] E17.一种包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与SEQ ID NO:3的核苷酸58-2277和2320-4374具有至少约92%、至少约93%、至少约94%、至少约95%、至少96%、至少97%、至少98%或至少99%序列同一性的核酸序列。

[0036] E18.E17的分离的核酸分子,其中所述核苷酸序列包含SEQ ID NO:3的核苷酸58-2277和2320-4374。

[0037] E19.一种包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与SEQ ID NO:4的核苷酸58-2277和2320-4374具有至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少96%、至少97%、至少98%或至少99%序列同一性的核酸序列。

[0038] E20.E19的分离的核酸分子,其中所述核苷酸序列包含SEQ ID NO:4的核苷酸58-2277和2320-4374。

[0039] E21.一种包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与SEQ ID NO:5的核苷酸58-2277和2320-4374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。

[0040] E22.E21的分离的核酸分子,其中所述核苷酸序列包含SEQ ID NO:5的核苷酸58-

2277和2320-4374。

[0041] E23.一种包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与SEQ ID NO:6的核苷酸58-2277和2320-4374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。

[0042] E24.E23的分离的核酸分子,其中所述核苷酸序列包含SEQ ID NO:6的核苷酸58-2277和2320-4374。

[0043] E25.E1至E24中任一者的分离的核酸分子,其中所述核苷酸序列进一步包含编码信号肽的核酸序列。

[0044] E26.E25的分离的核酸分子,其中所述信号肽是FVIII信号肽。

[0045] E27.E25或E26的分离的核酸分子,其中所述编码信号肽的核酸序列经过密码子优化。

[0046] E28.E25至E27中任一者的分离的核酸分子,其中所述编码信号肽的核酸序列与(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。

[0047] E29.E1至E28中任一者的分离的核酸分子,其中所述核酸分子包含选自由以下组成的组的一种或多种特性:

[0048] (a) 所述核酸分子或其一部分的人密码子适应指数相对于SEQ ID NO:16增加;

[0049] (b) 所述核苷酸序列或其一部分的最优密码子的频率相对于SEQ ID NO:16增加;

[0050] (c) 所述核苷酸序列或其一部分含有与SEQ ID NO:16中的G/C核苷酸的百分比相比较较高百分比的G/C核苷酸;

[0051] (d) 所述核苷酸序列或其一部分的相对同义密码子使用相对于SEQ ID NO:16增加;

[0052] (e) 所述核苷酸序列或其一部分的密码子的有效数目相对于SEQ ID NO:16降低;

[0053] (f) 所述核苷酸序列含有相对于SEQ ID NO:16较少的MARS/ARS序列(SEQ ID NO:21和22);

[0054] (g) 所述核苷酸序列含有相对于SEQ ID NO:16较少的去稳定元件(SEQ ID NO:23和24);和

[0055] (h) 其任何组合。

[0056] E30.E1至E29中任一者的分离的核酸分子,其进一步包含异源核苷酸序列。

[0057] E31.E30的分离的核酸分子,其中所述异源核苷酸序列编码异源氨基酸序列,所述异源氨基酸序列是半衰期延长剂。

[0058] E32.E30或E31的分离的核酸分子,其中所述异源氨基酸序列是免疫球蛋白恒定区或其一部分、转铁蛋白、白蛋白或PAS序列。

[0059] E33.E30或E31的分离的核酸分子,其中所述异源氨基酸序列是Fc区。

[0060] E34.E30至E33中任一者的分离的核酸分子,其中所述异源氨基酸序列连接至由所

述核苷酸序列编码的氨基酸序列的N端或C端或插入由所述核苷酸序列编码的氨基酸序列中的两个氨基酸之间。

[0061] E35.E34的分离的核酸分子,其中所述异源氨基酸序列插入选自表3的一个或多个插入位点处的两个氨基酸之间。

[0062] E36.E30至E35中任一者的分离的核酸分子,其编码包含FVIII的单体-二聚体杂合分子。

[0063] E37.E1至E36中任一者的分离的核酸分子,其中所述FVIII多肽是全长FVIII或B结构域缺失的FVIII。

[0064] E38.E1至E37中任一者的分离的核酸分子,其可操作性连接至至少一个转录控制序列。

[0065] E39.一种载体,其包含E1至E38中任一者的核酸分子。

[0066] E40.E39的载体,其中所述载体是病毒载体。

[0067] E41.一种宿主细胞,其包含E1至E32中任一者的核酸分子或E39或E40的载体。

[0068] E42.E41的宿主细胞,其中所述宿主细胞是选自由以下组成的组:CHO细胞、HEK293细胞、BHK21细胞、PER.C6细胞、NS0细胞和CAP细胞。

[0069] E43.一种多肽,其由E1至E37中任一者的核酸分子或E39或40的载体编码或由E41或E42的宿主细胞产生。

[0070] E44.一种产生具有FVIII活性的多肽的方法,所述方法包括:在一定条件下培养E41或E42的宿主细胞,由此产生具有FVIII活性的多肽;和回收所述具有FVIII活性的多肽。

[0071] E45.E44的方法,其中所述具有FVIII活性的多肽的表达相对于在包括包含SEQ ID NO:16的参考核苷酸序列的相同条件下培养的宿主细胞增加。

[0072] E46.一种增加受试者中具有FVIII活性的多肽的表达的方法,所述方法包括向有需要的受试者施用E1至E38中任一者的分离的核酸分子或E39或E40的载体,其中所述多肽的表达相对于包含SEQ ID NO:16的参考核酸分子或包含所述参考核酸分子的载体增加。

[0073] E47.一种增加具有FVIII活性的多肽的表达的方法,所述方法包括在一定条件下培养E41或E42的宿主细胞,由此由所述核酸分子表达具有FVIII活性的多肽,其中所述具有FVIII活性的多肽的表达相对于在包括包含SEQ ID NO:16的参考核酸序列的相同条件下培养的宿主细胞增加。

[0074] E48.E44至E47中任一者的方法,其中所述FVIII多肽的表达增加至少约2倍、至少约3倍、至少约4倍、至少约5倍、至少约6倍、至少约7倍、至少约8倍、至少约9倍、至少约10倍、至少约11倍、至少约12倍、至少约13倍、至少约14倍、至少约15倍、至少约20倍、至少约25倍、至少约30倍、至少约35倍、至少约40倍、至少约50倍、至少约60倍、至少约70倍、至少约80倍、至少约90倍或至少约100倍。

[0075] E49.一种提高具有FVIII活性的多肽的产率的方法,所述方法包括在一定条件下培养E41或E42的宿主细胞,由此由所述核酸分子产生具有FVIII活性的多肽,其中具有FVIII活性的多肽的产率相对于在包括包含SEQ ID NO:16的参考核酸序列的相同条件下培养的宿主细胞增加。

[0076] E50.一种治疗出血病症的方法,所述方法包括:向有需要的受试者施用E1至E38中任一者的核酸分子、E39或E40的载体或E43的多肽。

- [0077] E51.E50的方法,其中所述出血病症的特征在于FVIII缺陷。
- [0078] E52.E50的方法,其中所述出血病症是血友病。
- [0079] E53.E50的方法,其中所述出血病症是血友病A。
- [0080] E54.E50至E53中任一者的方法,其中在施用后24小时的血浆FVIII活性相对于施用包含SEQ ID NO:16的参考核酸分子、包含所述参考核酸分子的载体或由所述参考核酸分子编码的多肽的受试者增加。
- [0081] E55.E54的方法,其中所述血浆FVIII活性增加至少约2倍、至少约3倍、至少约4倍、至少约5倍、至少约6倍、至少约7倍、至少约8倍、至少约9倍、至少约10倍、至少约11倍、至少约12倍、至少约13倍、至少约14倍、至少约15倍、至少约20倍、至少约25倍、至少约30倍、至少约35倍、至少约40倍、至少约50倍、至少约60倍、至少约70倍、至少约80倍、至少约90倍或至少约100倍。
- [0082] E56.E39或40的载体,其中所述载体是慢病毒载体。
- [0083] E57.E39、E40和E56中任一者的载体,其进一步包含组织特异性启动子、组织特异性增强子或组织特异性启动子和组织特异性增强子两者。
- [0084] E58.E57的载体,其中所述组织特异性启动子和/或所述组织特异性增强子选择性地增强肝细胞中所述转基因的表达。
- [0085] E59.E57或E58的载体,其中所述组织特异性启动子包括选自由以下组成的组的启动子序列:由小鼠甲状腺素启动子(mTTR)、内源性人因子VIII启动子(F8)、人 α -1-抗胰蛋白酶启动子(hAAT)、人白蛋白最小启动子和小鼠白蛋白启动子。
- [0086] E60.E57至E59中任一者的载体,其中所述组织特异性启动子包括mTTR启动子。
- [0087] E61.一种治疗出血病症的方法,所述方法包括:向有需要的受试者施用E56至E60中任一者的载体。

附图说明

[0088] 图1A至图1J提供了编码B结构域缺失的因子VIII的密码子优化的核苷酸序列(SEQ ID NO:17)。图1A示出了coFVIII-3的核苷酸序列(SEQ ID NO:1)。图1B示出了coFVIII-4的核苷酸序列(SEQ ID NO:2)。图1C示出了coFVIII-5的核苷酸序列(SEQ ID NO:70)。图1D示出了coFVIII-6的核苷酸序列(SEQ ID NO:71)。图1E示出了coFVIII-52的核苷酸序列(SEQ ID NO:3)。图1F示出了coFVIII-62的核苷酸序列(SEQ ID NO:4)。图1G示出了coFVIII-25的核苷酸序列(SEQ ID NO:5)。图1H示出了coFVIII-26的核苷酸序列(SEQ ID NO:6)。图1I和图1J分别示出了B结构域缺失的(BDD-FVIII)的核苷酸序列和氨基酸序列(分别为SEQ ID NO:16和17)。

[0089] 图2A至图2J示出了编码BDD-FVIII的密码子优化的核苷酸序列中的密码子使用偏好调节。图2A示出了编码BDD-FVIII(例如,非优化BDD-FVIII)的野生型核苷酸序列(在密码子优化之前)中的密码子的相对频率。所述非优化BDD-FVIII序列的人密码子适应指数(CAI)是74%。图2B示出了coFVIII-1变体序列中的密码子的相对频率,所述coFVIII-1变体序列具有88%的人CAI。图2C示出了coFVIII-3变体序列中的密码子的相对频率,所述coFVIII-3变体序列具有91%的人CAI。图2D示出了coFVIII-4变体序列中的密码子的相对频率,所述coFVIII-4变体序列具有97%的人CAI。图2E示出了coFVIII-5变体序列中的密码

子的相对频率,所述coFVIII-5变体序列具有83%的人CAI。图2F示出了coFVIII-6变体序列中的密码子的相对频率,所述coFVIII-6变体序列具有83%的人CAI。图2G示出了coFVIII-52变体序列中的密码子的相对频率,所述coFVIII-52变体序列具有91%的人CAI。图2H示出了coFVIII-62变体序列中的密码子的相对频率,所述coFVIII-62变体序列具有91%的人CAI。图2I示出了coFVIII-25变体序列中的密码子的相对频率,所述coFVIII-25变体序列具有88%的人CAI。图2J示出了coFVIII-26变体序列中的密码子的相对频率,所述coFVIII-26变体序列具有88%的人CAI。

[0090] 图3提供FVIII-303的质粒图,所述FVIII-303在pcDNA3骨架中包含在ET增强的转甲状腺素蛋白启动子的控制下的coFVIII-1,所述coFVIII-1定位于coFVIII-1翻译起始位点的上游并且包含合成增强子、mTIR增强子和mTIR启动子。

[0091] 图4示出了在流体动力学注射5 μ g FVIII-303 (coFVIII-1;圆形)或5 μ g FVIII-311 (BDD-FVIII;正方形)之后HemA小鼠中的FVIII血浆活性的图形表示。FVIII血浆活性通过在注射后24、48和72小时进行FVIII特异性发色测定来测定。示出了针对FVIII-311的表达水平标准化的在72小时时的相对活性水平。

[0092] 图5示出了pLV-coFVIII-52的质粒图,所述pLV-coFVIII-52在慢病毒质粒中包含在ET启动子的控制下的coFVIII-52,所述coFVIII-52定位于coFVIII-52翻译起始位点的上游并且包含合成增强子、mTTR增强子和mTTR启动子。

[0093] 图6A至图6C示出了在流体动力学注射多种FVIII编码核苷酸之后HemA小鼠中的FVIII血浆活性的图形表示。FVIII血浆活性通过在注射后24、48和72小时进行FVIII特异性发色测定来测定。图6A示出了在流体动力学注射5 μ g LV-coFVIII-1 (实心圆形)、5 μ g LV-coFVIII-3 (三角形)、5 μ g LV-coFVIII-4 (倒三角形)、5 μ g LV-coFVIII-5 (菱形)或5 μ g LV-coFVIII-6 (空心圆形)之后HemA小鼠中的FVIII血浆活性。图6B示出了在流体动力学注射5 μ g LV-coFVIII-1 (圆形)、5 μ g LV-coFVIII-25 (三角形)或5 μ g LV-coFVIII-26 (倒三角形)之后HemA小鼠中的FVIII血浆活性。图6C示出了在流体动力学注射20 μ g LV-2116 (非密码子优化的(WT) BDD-FVIII核苷酸序列;空心圆形)、20 μ g LV-coFVIII-1 (三角形)、20 μ g LV-coFVIII-52 (正方形)或20 μ g LV-coFVIII-62 (实心圆形)之后HemA小鼠中的FVIII血浆活性。示出各质粒在72小时时的相对活性水平,如所指示的,针对LV-coFVIII-1 (图6A、图6B和图6C)和/或LV-2116 (图6C)的表达水平标准化。

[0094] 图7示出了如与LV-2116 (BDD-FVIII)对照相比,并且如通过FVIII特异性发色测定所测量,在注射每只小鼠1E8TU的包含coFVIII-1、coFVIII-5、coFVIII-52、coFVIII-6或coFVIII-62的慢病毒载体之后24小时HemA小鼠中的血浆FVIII活性。误差线指示标准偏差。

[0095] 图8A至图8C提供编码融合至XTEN的BDD-FVIII的多种密码子优化的核苷酸序列。图8A示出了coFVIII-52-XTEN的核苷酸序列 (SEQ ID NO:19),其中将编码具有144个氨基酸的XTEN的核苷酸序列 (“XTEN₁₄₄”;SEQ ID NO:18;加下划线)插入coFVIII-52核苷酸序列内。图8B示出了coFVIII-1-XTEN的核苷酸序列 (SEQ ID NO:20),其中将编码具有144个氨基酸的XTEN的核苷酸序列 (“XTEN₁₄₄”;SEQ ID NO:18;加下划线)插入coFVIII-1核苷酸序列内。图8C示出了coFVIII-6-XTEN的核苷酸序列 (SEQ ID NO:72),其中将编码具有144个氨基酸的XTEN的核苷酸序列 (“XTEN₁₄₄”;SEQ ID NO:18;加下划线)插入coFVIII-6核苷酸序列 (例如,对应于成熟FVIII序列的氨基酸残基745)内。

[0096] 图9提供pLV-coFVIII-52-XTEN的质粒图,所述pLV-coFVIII-52-XTEN在慢病毒载体中包含在ET启动子的控制下的coFVIII-52-XTEN。以与pLV-coFVIII-52-XTEN相同的方式构建包含如本文所述的编码具有FVIII活性的多肽的剩余密码子优化的核酸分子中的每一者的慢病毒载体,其中插入相同XTEN序列以替换FVIII的B结构域。

[0097] 图10A和图10B示出了在注射包含编码BDD-FVIII的多种密码子优化的核苷酸序列的质粒DNA(图10A)或慢病毒载体(图10B)之后HemA小鼠中的FVIII活性。图10A示出了在流体动力学注射5 μ g FVIII-311(非密码子优化,BDD-FVIII编码核苷酸序列;正方形)、5 μ g FVIII-303(coFVIII-1;小圆形)或FVIII-306(coFVIII-1-XTEN₁₄₄;大圆形)之后HemA小鼠中的FVIII血浆活性的图形表示。示出每种质粒在72小时针对FVIII-311标准化的相对活性。图10B示出了如与LV-2116(BDD-FVIII)对照相比,并且如通过FVIII特异性发色测定所测量,在注射每只小鼠1E8TU的包含coFVIII-52或coFVIII-52-XTEN的慢病毒载体之后21天HemA小鼠中的血浆FVIII活性。误差线指示标准偏差。

[0098] 图11A示出了全长成人因子VIII的氨基酸序列。图11B示出了全长人血管性血友病因子的氨基酸序列(SEQ ID NO:44)。图11C和图11D分别示出了具有42个氨基酸的XTEN多肽的氨基酸序列和核苷酸序列(XTEN AE42-4;分别为SEQ ID NO:46和47)。具有144个氨基酸的多种XTEN多肽的氨基酸序列在图11E、图11G、图11I、图11K、图11M、图11O、图11Q、图11S、图11U和图11W中示出(分别为SEQ ID NO:48、50、52、54、56、58、60、62、64和66),并且相应核苷酸序列在图11F、图11H、图11J、图11L、图11N、图11P、图11R、图11T、图11V和图11X中示出(分别为SEQ ID NO:49、51、53、55、57、59、61、63、65和67)。图11Y示出了ET启动子的核苷酸序列(SEQ ID NO:69)。图11Z示出了coFVIII-1的核苷酸序列(SEQ ID NO:68)(参见国际公布号WO 2014/127215,SEQ ID NO:1)。

[0099] 图12A是在IV施用约1.5E10TU/kg的LV-wtBDD-FVIII(圆形)、LV-coFVIII-6(正方形)或LV-coFVIII-6XTEN(三角形)之后14天龄的HemA小鼠中的FVIII血浆活性(IU/mL)的图形表示。图12B是在通过IV施用约1.5E10TU/kg的表达wtBDD-FVIII、coFVIII-1、coFVIII-3、coFVIII-4、coFVIII-5、coFVIII-6、coFVIII-52、coFVIII-62、coFVIII-25或coFVIII-26的慢病毒载体治疗14天龄的HemA小鼠之后150天的载体拷贝数(VCN)的图形表示。图12C是在通过IV施用约1.5E10TU/kg的表达wtBDD-FVIII、coFVIII-1、coFVIII-3、coFVIII-4、coFVIII-5、coFVIII-6、coFVIII-52、coFVIII-62、coFVIII-25或coFVIII-26的慢病毒载体治疗14天龄的HemA小鼠之后21天的FVIII血浆活性(IU/mL)的图形表示。

[0100] 图13A和图13B是说明用表达coFVIII-5变体的慢病毒治疗的五只HemA小鼠中的FVIII血浆活性水平(图13A)和抗FVIII抗体水平(图13B)的图形表示。通过静脉内注射向十四天龄的HemA同窝出生小鼠施用大约1.5E10TU/kg的表达coFVIII-5变体的慢病毒。各小鼠由数字指明(即,1、2、3、4和5;图13A和图13B)。

[0101] 图14是在如通过慢病毒治疗后21天的FVIII血浆活性所证明的LV-FVIII表达水平与抗FVIII抗体的存在之间的关系的图形表示。各数据点对应于单一HemA小鼠。各小鼠通过静脉内注射接受1.5E10TU/kg的剂量的表达本文所公开的coFVIII变体之一的慢病毒。水平线指示平均FVIII血浆活性。

[0102] 图15是在慢病毒治疗后150天时的每个细胞的载体拷贝数(VCN)与抗FVIII抗体的存在之间的关系的图形表示。各数据点对应于单一HemA小鼠。各小鼠通过静脉内注射接受

1.5E10TU/kg的剂量的表达本文所公开的coFVIII变体之一的慢病毒。水平线指示平均VCN。

[0103] 图16A和图16B是说明用表达coFVIII-52变体的慢病毒治疗的两只HemA小鼠(coFVIII-52-A和coFVIII-52-B)中的FVIII血浆活性水平(图16A)和抗FVIII抗体水平(图16B)的图形表示。通过静脉内注射向十四天龄的HemA同窝出生小鼠施用大约1.5E10TU/kg的表达coFVIII-52变体的慢病毒。图16C和图16D是示出从图16A和图16B的coFVIII-52-A(图16C)和coFVIII-52-B(图16D)小鼠收集的肝组织中的FVIII表达的RNA原位杂交染色(暗染色)的图像。

[0104] 图17是示出用表达野生型B结构域缺的失FVIII(wtBDD-FVIII;三角形)、coFVIII-52XTEN(圆形)或coFVIII-6XTEN(倒三角形)变体的慢病毒治疗的HemA新生小鼠中的长期FVIII表达的图形表示。通过静脉内注射向新生HemA小鼠施用大约1.5A10TU/kg的表达wtBDD-FVIII、coFVIII-52XTEN或coFVIII-6XTEN的慢病毒。在大约16周内测量FVIII血浆活性。

[0105] 图18是在施用 1.3×10^9 转导单位/kg的包含编码融合至XTEN的因子VIII的核苷酸的慢病毒载体(SEQ ID NO:72;LV-coFVIII-6-XTEN)之后HemA新生狗(S3或K4)中的循环FVIII水平的图形表示。由实线连接的正方形表示aPTT-S3样品,并且由虚线连接的三角形表示aPTT-K4样品。y轴示出了以正常值的百分比表示的血浆FVIII活性,其中正常人FVIII活性为100%。x轴示出了慢病毒治疗后的天数,其中慢病毒治疗在第0天施用。

[0106] 图19A至图19C是全血止血的图形表示,如通过针对原生HemA狗(图19A)、在慢病毒治疗后2周的狗S3(图19B)和在慢病毒治疗后2周的狗K4(图19C)的旋转式血栓弹力描记(ROTEM)测定所监测。关于图19A至图19C中的每一者,凝固时间(CT)以秒(s)示出,凝块形成时间(CFT)以秒(s)示出, α 角(α)以度($^\circ$)示出,CT后5分钟振幅(A5)以毫米(mm)示出,CT后20分钟振幅(A20)以毫米(mm)示出,并且最大凝块坚固度(MCF)以毫米(mm)示出。图19D是表格,概述了针对图19A至图19C中展示的参数CT、CFT、 α 、A5、A2和MCF中的每一者的正常范围。

具体实施方式

[0107] 本公开描述了编码具有因子VIII(FVIII)活性的多肽的密码子优化的基因。本公开针对编码具有因子VIII活性的多肽的密码子优化的核酸分子、包含优化的核酸分子的载体和宿主细胞、由优化的核酸分子编码的多肽和产生所述多肽的方法。本公开还针对治疗如血友病等出血病症的方法,所述方法包括向受试者施用优化的因子VIII核酸序列、包含所述优化的核酸序列的载体或由此编码的多肽。本公开通过提供优化的因子VIII序列满足了所属领域中的重要需要,所述优化的因子VIII序列在宿主细胞中展示了增加的表达,改进了产生重组因子VIII的方法中因子VIII蛋白的产率,并且当用于基因疗法方法中时潜在地引起较大的治疗功效。在某些实施方案中,本公开描述了一种包含如下核苷酸序列的分离的核酸分子,所述核苷酸序列与选自SEQ ID NO:1-14、70和71的核苷酸序列具有序列同源性。

[0108] 本公开的示例性构建体说明于附图和序列表中。为了提供对本说明书和权利要求的清楚理解,下文提供以下定义。

[0109] I. 定义

[0110] 应注意,术语“一个(a)”或“一个(an)”实体是指一个或多个那一实体:例如,“一个

核苷酸序列”应理解为表示一个或多个核苷酸序列。因而，术语“一个(a)”(或“一个(an)”)、“一个或多个”和“至少一个”在本文中可互换使用。

[0111] 术语“约”在本文中用于意指大概、约略地、大约或在区域中。当术语“约”结合数值范围使用时，其通过在其所陈述的数值上下延伸边界来修饰那一范围。通常，术语“约”在本文中用于修饰在所陈述的值上下的数值达10%的变化，向上或向下(更高或更低)。

[0112] 出于本公开的目的，术语“分离的”指明生物材料(细胞、多肽、多核苷酸或其片段、变体或衍生物)已经从其原始环境(其中天然地存在所述生物材料的环境)中去除。例如，以天然状态存在于植物或动物中的多核苷酸并非分离的，然而，从其中天然地存在所述多核苷酸的相邻核酸分离的所述多核苷酸被视为“分离的”。不需要特定纯化程度。出于本公开的目的，在宿主细胞中表达的重组产生的多肽和蛋白质被视为分离的，已经通过任何合适技术分离、分馏或部分地或大体上纯化的原生或重组多肽也是如此。

[0113] “核酸”、“核酸分子”、“寡核苷酸”和“多核苷酸”可互换使用并且是指核糖核苷(腺苷、鸟苷、尿苷或胞嘧啶核苷；“RNA分子”)或脱氧核糖核苷(脱氧腺苷、脱氧鸟苷、脱氧胸苷或脱氧胞嘧啶核苷；“DNA分子”)的磷酸酯聚合物形式，或呈单链形式或双链螺旋的其任何磷脂类似物，如硫代磷酸酯和硫酯。双链DNA-DNA、DNA-RNA和RNA-RNA螺旋是可能的。术语核酸分子并且特别地DNA或RNA分子仅指所述分子的一级和二级结构，并且并未使其限于任何特定三级形式。因此，这一术语包括尤其发现于线性或环状DNA分子(例如，限制片段)、质粒、超螺旋DNA和染色体中的双链DNA。在论述特定双链DNA分子的结构时，序列可在本文中根据正常惯例描述，所述正常惯例仅沿DNA的非转录链(即，具有与mRNA同源的序列的链)依5'至3'方向给出所述序列。“重组DNA分子”是已经历分子生物操纵的DNA分子。DNA包括但不限于cDNA、基因组DNA、质粒DNA、合成DNA和半合成DNA。本公开的“核酸组合物”包含如本文所述的一种或多种核酸。

[0114] 如本文所用，“编码区”或“编码序列”是多核苷酸的一部分，其由可翻译成氨基酸的密码子组成。尽管“终止密码子”(TAG、TGA或TAA)通常不翻译成氨基酸，其可被视为编码区的一部分，但任何侧接序列，例如启动子、核糖体结合位点、转录终止子、内含子等不是编码区的一部分。编码区的边界典型地由在5'端的编码所得多肽的氨基末端的起始密码子和在3'端的编码所得多肽的羧基末端的翻译终止密码子确定。两个或更多个编码区可存在于单一多核苷酸构建体中(例如，在单一载体上)，或存在于单独多核苷酸构建体中(例如，在单独(不同)载体上)。接着，由此可见单一载体可含有仅单一编码区，或包含两个或更多个编码区。

[0115] 由哺乳动物细胞分泌的某些蛋白质与分泌信号肽相关，一旦生长的蛋白质链跨糙面内质网的输出已经开始，所述分泌信号肽即从成熟蛋白质裂解。本领域技术人员知道，信号肽一般融合至所述多肽的N端，并且从完整或“全长”多肽裂解以产生所述多肽的分泌或“成熟”形式。在某些实施方案中，原生信号肽或保持指导与其可操作性缔合的多肽的分泌的能力的那一序列的功能衍生物。或者，可使用异源哺乳动物信号肽，例如人组织血纤维蛋白溶酶原活化剂(TPA)或小鼠 β -葡糖苷酸酶信号肽，或其功能衍生物。

[0116] 术语“下游”是指位于参考核苷酸序列的3'处的核苷酸序列。在某些实施方案中，下游核苷酸序列涉及转录起始点之后的序列。例如，基因的翻译起始密码子位于转录起始位点的下游。

[0117] 术语“上游”是指位于参考核苷酸序列的5'处的核苷酸序列。在某些实施方案中，上游核苷酸序列涉及位于编码区或转录起始点的5'侧的序列。例如，大多数启动子位于转录起始位点的上游。

[0118] 如本文所用，术语“基因调控区”或“调控区”是指位于编码区的上游(5'非编码序列)、内部或下游(3'非编码序列)，并且影响缔合的编码区的转录、RNA加工、稳定性或翻译的核苷酸序列。调控区可包括启动子、翻译前导序列、内含子、多腺苷酸化识别序列、RNA加工位点、效应子结合位点和茎-环结构。如果编码区意图用于在真核细胞中表达，那么多腺苷酸化信号和转录终止序列将通常位于所述编码区的3'处。

[0119] 编码例如多肽等基因产物的多核苷酸可包括与一个或多个编码区可操作性缔合的启动子和/或其他表达(例如，转录或翻译)控制元件。在可操作性缔合中，例如多肽等基因产物的编码区与一个或多个调控区缔合，其缔合方式使得将所述基因产物的表达置于所述调控区的影响或控制下。例如，如果启动子功能的诱导导致编码由所述编码区编码的基因产物的mRNA的转录，并且如果所述启动子与所述编码区之间的连接的性质不会干扰所述启动子指导所述基因产物的表达的能力或干扰DNA模板转录的能力，那么所述编码区与所述启动子“可操作性缔合”。除了启动子以外的其他表达控制元件也可与编码区可操作性缔合以指导基因产物表达，所述其他表达控制元件例如增强子、操纵子、阻遏子和转录终止信号。

[0120] “转录控制序列”是指提供宿主细胞中编码序列的表达的DNA调控序列，如启动子、增强子、终止子等。本领域技术人员已知多种转录控制区。这些转录控制区包括但不限于在脊椎动物细胞中发挥作用的转录控制区，如但不限于来自巨细胞病毒(即刻早期启动子，结合内含子-A)、猿猴病毒40(早期启动子)和逆转录病毒(如劳斯肉瘤病毒)的启动子和增强子区段。其他转录控制区包括源于脊椎动物基因的那些，如肌动蛋白、热休克蛋白、牛生长激素和兔 β -球蛋白，以及能够控制真核细胞中的基因表达的其他序列。另外的合适转录控制区包括组织特异性启动子和增强子以及淋巴因子诱导性启动子(例如，可由干扰素或白细胞介素诱导的启动子)。

[0121] 同样，本领域技术人员已知多种翻译控制元件。这些翻译控制元件包括但不限于核糖体结合位点、翻译起始和终止密码子和源于小核糖核酸病毒的元件(特别是内部核糖体进入位点或IRES，还称作CITE序列)。

[0122] 如本文所用，术语“表达”是指使多核苷酸产生基因产物(例如，RNA或多肽)的过程。其包括但不限于将所述多核苷酸转录成信使RNA(mRNA)、转运RNA(tRNA)、小发夹RNA(shRNA)、小干扰RNA(siRNA)或任何其他RNA产物，和将mRNA翻译成多肽。表达产生“基因产物”。如本文所用，基因产物可以是核酸，例如通过基因的转录产生的信使RNA，或自转录物翻译的多肽。本文所述的基因产物进一步包括具有转录后修饰(例如，多腺苷酸化或剪接)的核酸，或具有翻译后修饰(例如，甲基化、糖基化、添加脂质、与其他蛋白质亚单位缔合或蛋白水解裂解)的多肽。如本文所用，术语“产率”是指通过基因的表达产生的多肽的量。

[0123] “载体”是指用于将核酸克隆和/或转移至宿主细胞中的任何媒介。载体可以是可附接另一核酸区段以便引起所附接的区段的复制的复制子。“复制子”是指用作体内自主复制单位(即，能够在其本身的控制下复制)的任何遗传元件(例如，质粒、噬菌体、粘粒、染色体、病毒)。术语“载体”包括用于体外、离体或体内将核酸引入细胞中的病毒和非病毒媒介。

本领域中已知并使用大量载体,包括例如质粒、经过修饰的真核病毒或经过修饰的细菌病毒。将多核苷酸插入至合适载体中可通过将适当的多核苷酸片段连接至具有互补粘性末端的所选载体中而实现。

[0124] 载体可经过工程改造以编码提供已经并入所述载体的细胞的选择或鉴别的可选择标记或报告基因。可选择标记或报告基因的表达允许并入和表达含于所述载体上的其他编码区的宿主细胞的鉴别和/或选择。本领域中已知并且使用的可选择标记基因的实例包括:提供对氨苄青霉素、链霉素、庆大霉素、卡那霉素、潮霉素、双丙氨膦除草剂、磺酰胺等的抗性的基因;和用作表型标记的基因,即花青素调控基因、异戊烷基转移酶基因等。本领域中已知并且使用的报告基因的实例包括:荧光素酶 (Luc)、绿色荧光蛋白 (GFP)、氯霉素乙酰转移酶 (CAT)、 β -半乳糖苷酶 (LacZ)、 β -葡萄糖苷酸酶 (Gus) 等。可选择标记也可以被视为报告基因。

[0125] 术语“可选择标记”是指能够基于所述标记基因的效应被加以选择的鉴别因子,通常是抗生素或化学抗性基因,即对抗生素的抗性、对除草剂的抗性、比色标记、酶、荧光标记等,其中所述效应用于追踪所关注的核酸的遗传性和/或鉴别已经遗传所关注的核酸的细胞或生物体。本领域中已知并且使用的可选择标记基因的实例包括:提供对氨苄青霉素、链霉素、庆大霉素、卡那霉素、潮霉素、双丙氨膦除草剂、磺酰胺等的抗性的基因;和用作表型标记的基因,即花青素调控基因、异戊烷基转移酶基因等。

[0126] 术语“报告基因”是指编码能够基于所述报告基因的效应加以鉴别的鉴别因子的核酸,其中所述效应用于追踪所关注的核酸的遗传性、鉴别已经遗传所关注的核酸的细胞或生物体和/或测量基因表达诱导或转录。本领域中已知并且使用的报告基因的实例包括:荧光素酶 (Luc)、绿色荧光蛋白 (GFP)、氯霉素乙酰转移酶 (CAT)、 β -半乳糖苷酶 (LacZ)、 β -葡萄糖苷酸酶 (Gus) 等。可选择标记基因也可以被视为报告基因。

[0127] “启动子”和“启动子序列”可互换使用并且是指能够控制编码序列或功能RNA的表达的DNA序列。通常,编码序列位于启动子序列的3'处。启动子可整体源于原生基因,或由源于在自然界中发现的不同启动子的不同元件构成,或甚至包含合成DNA区段。本领域技术人员应理解,不同启动子可指导不同组织或细胞类型中或在不同发育阶段或响应于不同环境或生理条件的基因的表达。使基因在大多数时间在大多数细胞类型中表达的启动子通常被称作“组成性启动子”。使基因在特定细胞类型中表达的启动子通常被称作“细胞特异性启动子”或“组织特异性启动子”。使基因在特定发育或细胞分化阶段表达的启动子通常被称作“发育特异性启动子”或“细胞分化特异性启动子”。经过诱导并且在使细胞暴露于诱导所述启动子的试剂、生物分子、化学品、配体、光等或用所述试剂、生物分子、化学品、配体、光等处理之后使基因表达的启动子通常被称作“诱导性启动子”或“可控启动子”。进一步认识到,由于在大多数情况下尚未完全地界定调控序列的精确边界,不同长度的DNA片段可以具有相同的启动子活性。

[0128] 启动子序列通常在其3'端处以转录起始位点为界并且向上游延伸(5'方向)以包括开始在背景以上可检测的水平转录所必需的最少数目的碱基或元件。在所述启动子序列内将发现转录起始位点(便利地例如通过用核酸酶S1定位来界定),以及负责RNA聚合酶的结合的蛋白质结合结构域(共有序列)。

[0129] 术语“限制核酸内切酶”和“限制酶”可互换使用并且是指在双链DNA内的特定核苷

酸序列内结合并且切割的酶。

[0130] 术语“质粒”是指通常携带并非细胞的中心代谢的一部分的基因并且通常呈环状双链DNA分子的形式的染色体外元件。所述元件可以是自主复制的序列、基因组整合序列、噬菌体或核苷酸序列、源于任何来源的单链或双链DNA或RNA的线性、环状或超螺旋,其中多个核苷酸序列已经接合或重组成独特的构造,所述独特构造能够将所选择的基因产物的启动子片段和DNA序列连同适当的3'未翻译序列引入至细胞中。

[0131] 可使用的真核病毒载体包括但不限于腺病毒载体、逆转录病毒载体、腺相关病毒载体、痘病毒(例如,牛痘病毒载体)、杆状病毒载体或疱疹病毒载体。非病毒载体包括质粒、脂质体、带电脂质(细胞转染剂(cytofectin))、DNA-蛋白质复合物和生物聚合物。

[0132] “克隆载体”是指“复制子”,所述复制子是单位长度的核酸,所述核酸顺序复制并且包含复制起点,如质粒、噬菌体或粘粒,另一核酸区段可附接至所述复制子以便引起所附接的区段的复制。某些克隆载体能够在一种细胞类型(例如,细菌)中复制并且在另一细胞类型(例如,真核细胞)中表达。克隆载体典型地包含一个或多个可用于包含所述载体的细胞的选择的序列和/或一个或多个用于所关注的核酸序列的插入的克隆位点。

[0133] 术语“表达载体”是指经过设计以使所插入的核酸序列能够在插入至宿主细胞中之后表达的媒介。使所插入的核酸序列放置成与如上文所述的调控区的可操作缔合。

[0134] 载体通过本领域中众所周知的方法引入至宿主细胞中,例如转染、电穿孔、微注射、转导、细胞融合、DEAE葡聚糖、磷酸钙沉淀、脂质转染(溶酶体融合)、基因枪使用或DNA载体转运体(transporter)。

[0135] 如本文所用,“培养(Culture)”、“培养(to culture)”和“培养(culturing)”意指在允许细胞生长或分裂的体外条件下孵育细胞或使细胞维持生存状态。如本文所用,“培养的细胞”意指体外繁殖的细胞。

[0136] 如本文所用,术语“多肽”意图涵盖单数“多肽”以及复数“多肽”并且是指由通过酰胺键(还称作肽键)线性连接的单体(氨基酸)构成的分子。术语“多肽”是指两个或更多个氨基酸的任何一个或多个链,并且不指特定长度的产物。因此,肽、二肽、三肽、寡肽、“蛋白质”、“氨基酸链”或用于指两个或更多个氨基酸的一个或多个链的任何其他术语均包括于“多肽”的定义内,并且术语“多肽”可替代这些术语中的任一者或可互换使用。术语“多肽”还意图指多肽的表达后修饰的产物,所述表达后修饰包括但不限于糖基化、乙酰化、磷酸化、酰胺化、通过已知的保护/阻断基团衍生化、蛋白水解裂解或通过非天然存在的氨基酸修饰。多肽可源于天然生物来源或以重组技术产生,但不必自指定的核酸序列翻译。其可以任何方式,包括通过化学合成产生。

[0137] 术语“氨基酸”包括丙氨酸(Ala或A);精氨酸(Arg或R);天冬酰胺(Asn或N);天冬氨酸(Asp或D);半胱氨酸(Cys或C);谷酰胺(Gln或Q);谷氨酸(Glu或E);甘氨酸(Gly或G);组氨酸(His或H);异亮氨酸(Ile或I);亮氨酸(Leu或L);赖氨酸(Lys或K);甲硫氨酸(Met或M);苯丙氨酸(Phe或F);脯氨酸(Pro或P);丝氨酸(Ser或S);苏氨酸(Thr或T);色氨酸(Trp或W);酪氨酸(Tyr或Y);和缬氨酸(Val或V)。非传统氨基酸也在本公开的范围并且包括正亮氨酸、鸟氨酸、正缬氨酸、高丝氨酸和其他氨基酸残基类似物,如Ellman等人Meth.Enzym.202:301-336(1991)中所述的那些。为了产生所述非天然存在的氨基酸残基,可使用Noren等人Science 244:182(1989)和Ellman等人(同上)的程序。简单地说,这些程序涉及用非天然存

在的氨基酸残基以化学方式活化抑制因子tRNA,随后为所述RNA的体外转录和翻译。所述非传统氨基酸的引入也可使用本领域中已知的肽化学来实现。如本文所用,术语“极性氨基酸”包括具有净零电荷,但在其侧链的不同部分中具有非零部分电荷的氨基酸(例如,M、F、W、S、Y、N、Q、C)。这些氨基酸可参与疏水性相互作用和静电相互作用。如本文所用,术语“带电氨基酸”包括可在其侧链上具有非零净电荷的氨基酸(例如,R、K、H、E、D)。这些氨基酸可参与疏水性相互作用和静电相互作用。

[0138] 本公开还包括多肽的片段或变体,和其任何组合。术语“片段”或“变体”在提及本公开的多肽结合结构域或结合分子时包括保持参考多肽的至少一些特性(例如,对FcRn结合结构域或Fc变体的FcRn结合亲和力、针对FVIII变体的凝固活性或针对VWF片段的FVIII结合活性)的任何多肽。多肽的片段包括蛋白水解片段,以及缺失片段,加上本文中别处论述的特定抗体片段,但不包括天然存在的全长多肽(或成熟多肽)。本公开的多肽结合结构域或结合分子的变体包括如上文所述的片段,以及由于氨基酸取代、缺失或插入而具有改变的氨基酸序列的多肽。变体可为天然或非天然存在的。非天然存在的变体可使用本领域已知的诱变技术产生。变体多肽可包含保守或非保守氨基酸取代、缺失或添加。

[0139] “保守氨基酸取代”是如下取代,其中氨基酸残基用具有相似侧链的氨基酸残基置换。具有相似侧链的氨基酸残基的家族已经在本领域中定义,包括碱性侧链(例如,赖氨酸、精氨酸、组氨酸)、酸性侧链(例如,天冬氨酸、谷氨酸)、不带电极性侧链(例如,甘氨酸、天冬酰胺、谷酰胺、丝氨酸、苏氨酸、酪氨酸、半胱氨酸)、非极性侧链(例如,丙氨酸、缬氨酸、亮氨酸、异亮氨酸、脯氨酸、苯丙氨酸、甲硫氨酸、色氨酸)、 β -分支链侧链(例如,苏氨酸、缬氨酸、异亮氨酸)和芳香族侧链(例如,酪氨酸、苯丙氨酸、色氨酸、组氨酸)。因此,如果多肽中的氨基酸用来自相同侧链家族的另一氨基酸置换,那么所述取代被视为保守的。在另一实施方案中,氨基酸串可以用侧链家族成员的顺序和/或组成不同的结构上相似的串保守性置换。

[0140] 如本领域中已知的术语“同一性百分比”是如通过比较所述序列所确定的两个或更多个多肽序列或两个或更多个多核苷酸序列之间的关系。在本领域中,“同一性”还意指如通过所述序列的串之间的匹配所确定的多肽或多核苷酸序列(视情况而定)之间的序列相关的程度。“同一性”可容易地通过已知方法计算,所述方法包括但不限于以下描述的那些:Computational Molecular Biology (Lesk,A.M.编)Oxford University Press,New York(1988);Biocomputing:Informatics and Genome Projects (Smith,D.W.编)Academic Press,New York(1993);Computer Analysis of Sequence Data,第I部分 (Griffin,A.M.和Griffin,H.G.编)Humana Press,New Jersey(1994);Sequence Analysis in Molecular Biology (von Heinje,G.编)Academic Press(1987);和Sequence Analysis Primer (Gribskov,M.和Devereux,J.编)Stockton Press,New York(1991)。测定同一性的优选方法经过设计以给出所测试的序列之间的最佳匹配。测定同一性的方法编纂于公开可得的计算机程序中。序列比对和同一性百分比计算可使用序列分析软件执行,所述序列分析软件如LASERGENE生物信息计算套件的Megalign程序(DNASTAR Inc.,Madison,WI)、GCG程序套件(Wisconsin Package 9.0版,Genetics Computer Group(GCG),Madison,WI)、BLASTP、BLASTN、BLASTX(Altschul等人,J.Mol.Biol.215:403(1990))和DNASTAR(DNASTAR, Inc.1228S.Park St.Madison,WI53715USA)。在本申请的上下文内,应理解在使用序列分析软件进行分析的情况下,除非另外规定,否则所述分析的结果将基于所参考的程序的“默认

值”。如本文所用,“默认值”将意指当首次初始化时最初装载所述软件的值或参数的任何集合。出于测定本公开的优化的BDD FVIII序列与参考序列之间的同一性百分比的目的,仅使用所述参考序列中对应于本公开的优化的BDD FVIII序列中的核苷酸的核苷酸来计算同一性百分比。例如,当比较含有B结构域的全长FVIII核苷酸序列与本公开的优化的B结构域缺失(BDD)的FVIII核苷酸序列时,将使用所述比对中包括A1、A2、A3、C1和C2结构域的部分来计算同一性百分比。全长FVIII序列中编码B结构域(其将导致所述比对中的大“空位”)的部分的核苷酸将作为错误匹配未进行计算。另外,在测定本公开的优化的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)与参考序列之间的同一性百分比时,同一性百分比将通过将匹配的核苷酸的数目除以如本文所陈述的优化的BDD-FVIII序列或其指定部分的完整序列中的核苷酸的总数来计算。

[0141] 如本文所用,“对应于本公开的优化的BDD FVIII序列中的核苷酸的核苷酸”是通过本公开的优化的BDD FVIII序列的比对来鉴别以最大化与参考FVIII序列的同一性。用于鉴别参考FVIII序列中的等同氨基酸的编号是基于用于鉴别本公开的优化的BDD FVIII序列中的相应氨基酸的编号。

[0142] “融合”或“嵌合”蛋白包含第一氨基酸序列连接至在自然界中未与其天然地连接的第二氨基酸序列。可使通常存在于单独蛋白质中的氨基酸序列集合于融合多肽中,或可将通常存在于同一蛋白质中的氨基酸序列以新的布置放置于融合多肽中,例如本公开的因子VIII结构域与Ig Fc结构域的融合。融合蛋白例如通过化学合成,或通过产生并且翻译其中肽区以所需关系进行编码的多核苷酸而产生。嵌合蛋白可进一步包含通过共价键、非肽键或非共价键与所述第一氨基酸序列缔合的第二氨基酸序列。

[0143] 如本文所用,术语“插入位点”是指FVIII多肽或其片段、变体或衍生物中的位置,所述位置紧邻可插入异源部分的位置的上游。“插入位点”以编号指定,所述编号是成熟原生FVIII(SEQ ID NO:15;图11A)中对应于所述插入位点的氨基酸的编号,所述氨基酸紧邻插入的位置的N端。例如,措辞“a3包含在对应于SEQ ID NO:15的氨基酸1656的插入位点处的异源部分”指示所述异源部分位于对应于SEQ ID NO:15的氨基酸1656和氨基酸1657的两个氨基酸之间。

[0144] 如本文所用,措辞“紧邻氨基酸的下游”是指紧挨着所述氨基酸的末端羧基的位置。同样,措辞“紧邻氨基酸的上游”是指紧挨着所述氨基酸的末端胺基的位置。

[0145] 如本文所用,术语“插入(inserted)”、“插入(is inserted)”、“插入至.....中(inserted into)”或语法相关术语是指相对于原生成熟FVIII中的类似位置,重组FVIII多肽中的异源部分的位置。如本文所用,所述术语是指所述重组FVIII多肽相对于原生成熟FVIII的特征,并且不指示、暗示或推断制得所述重组FVIII多肽的任何方法或过程。

[0146] 如本文所用,术语“半衰期”是指特定多肽的体内生物半衰期。半衰期可由施用于受试者的量的一半从动物中的循环和/或其他组织清除所需的时间表示。当给定多肽的清除曲线被绘制为随时间而变时,所述曲线通常是具有快速 α -相和较长 β -相的双相。所述 α -相通常表示所施用的Fc多肽在血管内间隙与血管外间隙之间的平衡并且部分地由所述多肽的大小决定。所述 β -相通常表示所述多肽在血管内间隙中的分解代谢。在一些实施方案中,FVIII和包含FVIII的嵌合蛋白是单相的,并且因此不具有 α 相,而仅具有单一 β 相。因此,在某些实施方案中,如本文所用的术语半衰期是指所述多肽在 β -相中的半衰期。

[0147] 如本文所用,术语“连接(linked)”分别是指第一氨基酸序列或核苷酸序列共价或非共价接合至第二氨基酸序列或核苷酸序列。所述第一氨基酸序列或核苷酸序列可直接地接合至所述第二氨基酸序列或核苷酸序列或与所述第二氨基酸序列或核苷酸序列并置,或替代地,间插序列可使所述第一序列共价地接合至所述第二序列。术语“连接”不仅意指第一氨基酸序列与第二氨基酸序列在C端或N端处的融合,而且包括整个第一氨基酸序列(或所述第二氨基酸序列)插入至所述第二氨基酸序列(或分别地,所述第一氨基酸序列)中的任何两个氨基酸中。在一个实施方案中,所述第一氨基酸序列可通过肽键或接头连接至第二氨基酸序列。所述第一核苷酸序列可通过磷酸二酯键或接头连接至第二核苷酸序列。所述接头可以是肽或多肽(针对多肽链)或核苷酸或核苷酸链(针对核苷酸链)或任何化学部分(针对多肽链和多核苷酸链两者)。术语“连接”还由连字符(-)指示。

[0148] 如本文所用,术语“与.....缔合”是指在第一氨基酸链与第二氨基酸链之间形成的共价或非共价键。在一个实施方案中,术语“与.....缔合”意指共价、非肽键或非共价键。这种缔合可由冒号,即(:)指示。在另一实施方案中,其意指除肽键以外的共价键。例如,氨基酸半胱氨酸包含巯醇基,其可与第二半胱氨酸残基上的巯醇基形成二硫键或桥。在大多数天然存在的IgG分子中,CH1和CL区由二硫键缔合并且两条重链由使用Kabat编号系统时对应于239和242的位置处的两个二硫键(位置226或229,EU编号系统)缔合。共价键的实例包括但不限于肽键、金属键、氢键、二硫键、 σ 键、 π 键、 δ 键、糖苷键、不确定键、弯键、偶极键、 π 骨架、双键、三键、四重键、五重键、六重键、共轭、超共轭、芳香性、哈普托数或反键。非共价键的非限制性实例包括离子键(例如,阳离子- π 键或盐键)、金属键、氢键(例如,二氢键、二氢复合物、低屏障氢键或对称氢键)、范德华力、伦敦色散力(London dispersion force)、机械键、卤素键、亲金作用、嵌入、堆叠、熵力或化学极性。

[0149] 本文所用的术语“单体-二聚体杂合物”是指包含第一多肽链和第二多肽链的嵌合蛋白,所述多肽链通过二硫键彼此缔合,其中所述第一链包含凝固因子(例如,因子VIII)和第一Fc区并且所述第二链包含第二Fc区,基本上由第二Fc区组成,或由第二Fc区组成,而无所述凝固因子。因此,所述单体-二聚体杂合物构建体是包含仅具有一种凝固因子的单体方面和具有两个Fc区的二聚体方面的杂合物。

[0150] 如本文所用,止血意指出血的停止或减慢;或通过血管或身体部分的血流的停止或减慢。

[0151] 如本文所用,止血病症意指基因遗传或获得性病况,其特征在于自发地或由于创伤而出血的倾向,归因于形成纤维蛋白凝块的能力受损或不能形成纤维蛋白凝块。所述病症的实例包括血友病。三种主要形式是血友病A(因子VIII缺陷)、血友病B(因子IX缺陷或“克雷司马斯病”)和血友病C(因子XI缺陷,轻度出血倾向)。其他止血病症包括例如血管假性血友病、因子XI缺陷(PTA缺陷)、因子XII缺陷、纤维蛋白原、凝血酶原、因子V、因子VII、因子X或因子XIII的缺陷或结构异常、Bernard-Soulier综合征,所述Bernard-Soulier综合征是GPIIb的缺乏或缺陷。vWF的受体GPIIb可以是有缺陷的并且导致初期凝块形成(初期止血)的缺乏和增加的出血倾向),和Glanzmann和Naegeli的血小板机能不全(血小板无力症(Glanzmann thrombasthenia))。在肝功能衰竭(急性和慢性形式)中,存在由肝脏引起的凝固因子的产生不足;这可增加出血风险。

[0152] 本公开的分离的核酸分子、分离的多肽或包含所述分离的核酸分子的载体可以预

防性地加以使用。如本文所用,术语“预防性治疗”是指在出血事件之前施用分子。在一个实施方案中,需要一般止血剂的受试者正在经历或将要经历手术。本公开的多核苷酸、多肽或载体可在作为预防剂在手术之前或之后施用。本公开的多核苷酸、多肽或载体可在手术期间或之后施用以控制急性出血事件。所述手术可包括但不限于肝移植、肝切除、牙科手术或干细胞移植。

[0153] 本公开的分离的核酸分子、分离的多肽或载体也用于按需要治疗。术语“按需要治疗”是指响应于出血事件的症状或在可引起出血的行为之前施用分离的核酸分子、分离的多肽或载体。一方面,当出血开始时(如在损伤之后),或当预期出血时(如在手术之前),可对受试者给予按需要治疗。另一方面,在增加出血风险的活动(如接触运动)之前可给予按需要治疗。

[0154] 如本文所用,术语“急性出血”是指出血事件,而与潜在原因无关。例如,受试者可具有创伤、尿毒症、遗传性出血病症(例如,因子VII缺陷)、血小板病症或由于凝固因子的抗体的发展而产生的抗性。

[0155] 如本文所用,治疗(Treat)、治疗(treatment)、治疗(treating)是指例如疾病或病况的严重性的降低;疾病过程的持续时间的减少;与疾病或病况相关的一种或多种症状的改善;向患有疾病或病况的受试者提供有益作用,而未必治愈所述疾病或病况,或与疾病或病况相关的一种或多种症状的预防。在一个实施方案中,术语“治疗(treating)”或“治疗(treatment)”意指通过施用本公开的分离的核酸分子、分离的多肽或载体在受试者中维持至少约1IU/dL、2IU/dL、3IU/dL、4IU/dL、5IU/dL、6IU/dL、7IU/dL、8IU/dL、9IU/dL、10IU/dL、11IU/dL、12IU/dL、13IU/dL、14IU/dL、15IU/dL、16IU/dL、17IU/dL、18IU/dL、19IU/dL或20IU/dL的FVIII谷水平。在另一实施方案中,治疗意指维持介于约1与约20IU/dL之间、介于约2与约20IU/dL之间、介于约3与约20IU/dL之间、介于约4与约20IU/dL之间、介于约5与约20IU/dL之间、介于约6与约20IU/dL之间、介于约7与约20IU/dL之间、介于约8与约20IU/dL之间、介于约9与约20IU/dL之间或介于约10与约20IU/dL之间的FVIII谷水平。疾病或病况的治疗也可包括在受试者中维持可与非血友病受试者中的FVIII活性的至少约1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%或20%相当的水平的FVIII活性。治疗所需的最低谷水平可通过一种或多种已知的方法来测量并且可针对每个人加以调节(增加或减少)。

[0156] 如本文所用,“施用”意指经由药学上可接受的途径对受试者给予药学上可接受的因子VIII编码核酸分子、因子VIII多肽或包含本公开的因子VIII编码核酸分子的载体。施用途径可以是静脉内,例如静脉内注射和静脉内输注。另外的施用途径包括例如皮下、肌肉内、口服、鼻和肺施用。所述核酸分子、多肽和载体可作为包含至少一种赋形剂的药物组合物的一部分施用。

[0157] 如本文所用,措辞“有需要的受试者”包括将受益于本公开的核酸分子、多肽或载体的施用的受试者,如哺乳动物受试者,例如以改善止血。在一个实施方案中,所述受试者包括但不限于患有血友病的个体。在另一实施方案中,所述受试者包括但不限于已经发展FVIII抑制因子并且因此需要旁路疗法的个体。所述受试者可以是成体或未成年体(例如,不足12岁)。

[0158] 如本文所用,术语“凝固因子”是指天然存在或重组产生的分子或其类似物,其预

防受试者中的出血事件或减少受试者中的出血事件的持续时间。换句话说,其意指具有促凝固活性的分子,即,负责将纤维蛋白原转化为不溶性纤维蛋白网,从而使血液凝结或凝固。“可活化凝固因子”是呈无活性形式(例如,呈其酶原形式)的凝固因子,其能够被转化为活性形式。

[0159] 如本文所用,凝固活性意指参与生物化学反应的级联的能力,所述级联导致形成纤维蛋白凝块和/或降低出血或出血事件的严重性、持续时间或频率。

[0160] 如本文所用,术语“异源”或“外源”是指在给定情况中,例如在细胞中或在多肽中通常未发现的所述分子。例如,外源或异源分子可被引入至细胞中并且仅在例如通过转染或遗传工程改造的其他形式操纵细胞之后存在,或异源氨基酸序列可存在于其中未天然地发现所述序列的蛋白质中。

[0161] 如本文所用,术语“异源核苷酸序列”是指未天然地与给定多核苷酸序列一起存在的核苷酸序列。在一个实施方案中,所述异源核苷酸序列编码能够延长FVIII的半衰期的多肽。在另一实施方案中,所述异源核苷酸序列编码增加FVIII的流体动力学半径的多肽。在其他实施方案中,所述异源核苷酸序列编码改善FVIII的一种或多种药物动力学特性而不会显著地影响其生物活性或功能(例如,其促凝血活性)的多肽。在一些实施方案中,FVIII通过接头连接至(linked or connected to)由所述异源核苷酸序列编码的多肽。由异源核苷酸序列编码的多肽部分的非限制性实例包括免疫球蛋白恒定区或其一部分、白蛋白或其片段、白蛋白结合部分、转铁蛋白、美国专利申请号20100292130的PAS多肽、HAP序列、转铁蛋白或其片段、人绒毛膜促性腺激素的β亚单位的C端肽(CTP)、白蛋白结合小分子、XTEN序列、FcRn结合部分(例如,完全Fc区或其结合FcRn的部分)、单链Fc区(ScFc区,例如,如US 2008/0260738、WO 2008/012543或WO 2008/1439545中所述)、聚甘氨酸接头、聚丝氨酸接头、选自甘氨酸(G)、丙氨酸(A)、丝氨酸(S)、苏氨酸(T)、谷氨酸(E)和脯氨酸(P)的两种类型的氨基酸的6-40个氨基酸的肽和短多肽(其中具有小于50%至大于50%的变化程度的二级结构)或其两种或更多种组合。在一些实施方案中,由所述异源核苷酸序列编码的多肽连接至非多肽部分。所述非多肽部分的非限制性实例包括聚乙二醇(PEG)、白蛋白结合小分子、聚唾液酸、羟基乙基淀粉(HES)、其衍生物或其任何组合。

[0162] 如本文所用,术语“Fc区”是定义为多肽中对应于原生Ig的Fc区的部分,即,如通过其两条重链的相应Fc域的二聚缔合所形成。原生Fc区与另一Fc区形成同二聚体。相比之下,如本文所用,术语“基因融合的Fc区”或“单链Fc区”(scFc区)是指包含基因连接于单一多肽链内(即,编码于单一邻接基因序列中)的Fc结构域的合成二聚Fc区。

[0163] 在一个实施方案中,“Fc区”是指单一Ig重链的部分,在铰链区中紧邻木瓜蛋白酶裂解位点的上游(即,IgG中的残基216,考虑重链恒定区的第一残基是114)开始并且在所述抗体的C端处终止。相应地,完整Fc结构域至少包含铰链结构域、CH2结构域和CH3结构域。

[0164] 视Ig同种型而定,Ig恒定区的Fc区可包括CH2、CH3和CH4结构域以及铰链区。包含Ig的Fc区的嵌合蛋白赋予嵌合蛋白数种可需要的特性,包括增加的稳定性、增加的血清半衰期(参见Capon等人,1989,Nature 337:525)以及结合于Fc受体,如新生儿Fc受体(FcRn)(美国专利号6,086,875、6,485,726、6,030,613;WO 03/077834;US2003-0235536A1),所述专利以引用的方式整体并入本文中。

[0165] “参考核苷酸序列”当作为与本公开的核苷酸序列的比较用于本文中时是与本公

开的核苷酸序列基本上同一的多核苷酸序列,除了对应于FVIII序列的部分未优化。例如,针对由SEQ ID NO:1的密码子优化的BDD FVIII和编码在3'端连接至SEQ ID NO:1的单链Fc区的异源核苷酸序列组成的核酸分子的参考核苷酸序列是由SEQ ID NO:16的原始(或“亲本”)BDD FVIII(图1I)和编码在3'端连接至SEQ ID NO:16的单链Fc区的同一异源核苷酸序列组成的核酸分子。

[0166] 如本文所用,“密码子适应指数”是指密码子使用偏好的量度。密码子适应指数(CAI)量度了给定的蛋白质编码基因序列相对于参考基因集合的偏差(Sharp PM和Li WH, Nucleic Acids Res.15(3):1281-95(1987))。CAI是通过测定在所述基因序列的长度内与各密码子相关的权重的几何平均值(在密码子中测量)来计算:

$$[0167] \quad CAI = \exp(1/L \sum_{l=1}^L \ln(w_l(l))), \quad (I)$$

[0168] 关于各氨基酸,在CAI中其密码子中的每一者的权重计算为针对所述氨基酸所观察到的密码子(f_i)的频率与同义密码子(f_j)的频率之间的比率:

[0169] 式2:

$$[0170] \quad w_i = \frac{f_i}{\max(f_j)} \quad ij \in [\text{氨基酸的同义密码子}] \quad (II)$$

[0171] 如本文所用,关于核苷酸序列的术语“优化”是指编码多肽的多核苷酸序列,其中所述多核苷酸序列已经发生突变以增强所述多核苷酸序列的特性。在一些实施方案中,进行优化以增加转录水平,增加翻译水平,增加稳态mRNA水平,增加或减少调控蛋白(如通用转录因子)的结合,增加或减少剪接,或增加由所述多核苷酸序列产生的多肽的产率。可针对多核苷酸序列进行的以优化所述多核苷酸序列的改变的实例包括密码子优化、G/C含量优化、重复序列的去除、富AT元件的去除、隐蔽剪接位点的去除、抑制转录或翻译的顺式作用元件的去除、添加或去除poly-T或poly-A序列、在转录起始位点附近添加增强转录的序列(如Kozak共有序列)、可能形成茎环结构的序列的去除、去稳定序列的去除和其两种或更多种组合。

[0172] 编码FVIII蛋白的多核苷酸序列

[0173] 在一些实施方案中,本公开针对编码具有FVIII活性的多肽的密码子优化的核酸分子。在一些实施方案中,所述多核苷酸编码全长FVIII多肽。在其他实施方案中,所述核酸分子编码B结构域缺失(BDD)的FVIII多肽,其中FVIII的B结构域的全部或一部分缺失。在一个特定实施方案中,所述核酸分子编码包含如下氨基酸序列的多肽,所述氨基酸序列与SEQ ID NO:17(图1J)或其片段具有至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性。在一个实施方案中,所述核酸分子编码具有氨基酸序列SEQ ID NO:17的多肽或其片段。

[0174] 在一些实施方案中,本公开的核酸分子编码包含信号肽或其片段的FVIII多肽。在其他实施方案中,所述核酸分子编码缺乏信号肽的FVIII多肽。在一些实施方案中,所述信号肽包含SEQ ID NO:17的氨基酸1-19。

[0175] 在一些实施方案中,本公开提供包含如下核苷酸序列的分离的核酸分子,所述核

核苷酸序列包含编码FVIII多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列;其中所述第一核酸序列与(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;并且其中所述N端部分和所述C端部分合起来具有FVIII多肽活性。在一个特定实施方案中,所述第一核酸序列与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。在另一实施方案中,所述第一核酸序列与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。在其他实施方案中,所述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。

[0176] 在其他实施方案中,本公开提供包含如下核苷酸序列的分离的核酸分子,所述核苷酸序列包含编码FVIII多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列;其中所述第一核酸序列与(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;并且其中所述N端部分和所述C端部分合起来具有FVIII多肽活性。在一个实施方案中,所述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。在另一实施方案中,所述第二核苷酸序列与SEQ ID NO:3的核苷酸1792-4374或SEQ ID NO:4的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。在一个特定实施方案中,所述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1792-4374。在另一实施方案中,所述第二核苷酸序列与SEQ ID NO:3的核苷酸1792-2277和2320-4374或SEQ ID NO:4的1792-2277和2320-4374(即,SEQ ID NO:3的核苷酸1792-4374或SEQ ID NO:4的1792-4374,而无编码B结构域或B结构域片段的核苷酸)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。在一个特定实施方案中,所述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1792-2277和2320-4374(即,SEQ ID NO:3的核苷酸1792-4374或SEQ ID NO:4的1792-4374,而无编码B结构域或B结构域片段的核苷酸)。

[0177] 在一些实施方案中,本公开提供包含如下核苷酸序列的分离的核酸分子,所述核苷酸序列包含编码FVIII多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列;其中所述第二核酸序列与(i) SEQ ID NO:5的核苷酸1792-4374或(ii) SEQ ID NO:6的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;并且其中所述N端部分和所述C端部分合起来具有FVIII多肽活性。在某些实施方案中,所述第二核酸序列与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。在其他实施方案中,所述第二核酸序列与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。在一个特定实施方案中,所述第二核酸序列包含SEQ ID NO:5的核苷酸1792-4374或SEQ ID NO:6的1792-4374。在一些实施方案中,连接至上文列出的第二核酸序列的第一核酸序列与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。在其他实施方案中,连接至上

文列出的第二核酸序列的第一核酸序列与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。

[0178] 在其他实施方案中,本公开提供包含如下核苷酸序列的分离的核酸分子,所述核苷酸序列包含编码FVIII多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列;其中所述第二核酸序列与(i) SEQ ID NO:5的核苷酸1792-2277和2320-4374(即,SEQ ID NO:5的核苷酸1792-4374,而无编码B结构域或B结构域片段的核苷酸)或(ii) SEQ ID NO:6的1792-2277和2320-4374(即,SEQ ID NO:6的核苷酸1792-4374,而无编码B结构域或B结构域片段的核苷酸)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;并且其中所述N端部分和所述C端部分合起来具有FVIII多肽活性。在某些实施方案中,所述第二核酸序列与SEQ ID NO:5的核苷酸1792-2277和2320-4374(即,SEQ ID NO:5的核苷酸1792-4374,而无编码B结构域或B结构域片段的核苷酸)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。在其他实施方案中,所述第二核酸序列与SEQ ID NO:6的核苷酸1792-2277和2320-4374(即,SEQ ID NO:6的核苷酸1792-4374,而无编码B结构域或B结构域片段的核苷酸)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。在一个特定实施方案中,所述第二核酸序列包含SEQ ID NO:5的核苷酸1792-2277和2320-4374或SEQ ID NO:6的1792-2277和2320-4374(即,SEQ ID NO:5的核苷酸1792-4374或SEQ ID NO:6的1792-4374,而无编码B结构域或B结构域片段的核苷酸)。在一些实施方案中,连接至上文列出的第二核酸序列的第一核酸序列与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。在其他实施方案中,连接至上文列出的第二核酸序列的第一核酸序列与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。

[0179] 在一些实施方案中,本公开提供包含如下核苷酸序列的分离的核酸分子,所述核苷酸序列包含编码FVIII多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列;其中所述第一核酸序列与(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;并且其中所述N端部分和所述C端部分合起来具有FVIII多肽活性。在其他实施方案中,所述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。

[0180] 在其他实施方案中,本公开提供包含如下核苷酸序列的分离的核酸分子,所述核苷酸序列包含编码FVIII多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列;其中所述第一核酸序列与(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;并且其中所述N端部分和所述C端部分合起来具有FVIII多肽活性。在一个实施方案中,所述第一

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。在另一实施方案中,连接至所述第一核苷酸序列的第二核苷酸序列与SEQ ID NO:1的核苷酸1792-4374、SEQ ID NO:2的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。在一个特定实施方案中,连接至所述第一核苷酸序列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。在其他实施方案中,连接至所述第一核苷酸序列的第二核苷酸序列与(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。在一个实施方案中,所述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。

[0181] 在另一实施方案中,本公开提供包含如下核苷酸序列的分离的核酸分子,所述核苷酸序列包含编码FVIII多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列;其中所述第二核酸序列与(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;并且其中所述N端部分和所述C端部分合起来具有FVIII多肽活性。在一个特定实施方案中,所述第二核酸序列包含(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。在一些实施方案中,本公开提供包含如下核苷酸序列的分离的核酸分子,所述核苷酸序列包含编码FVIII多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列;其中所述第二核酸序列与(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,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,而无编码B结构域或B结构域片段的核苷酸)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;并且其中所述N端部分和所述C端部分合起来具有FVIII多肽活性。在一个实施方案中,所述第二核酸序列包含(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,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,而无编码B结构域或B结构域片段的核苷酸)。

[0182] 在一些实施方案中,本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的

分离的核酸分子,其中所述核苷酸序列包含与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。在其他实施方案中,所述核苷酸序列包含与SEQ ID NO:1的核苷酸58-2277和2320-4374(即,SEQ ID NO:1的核苷酸58-4374,而无编码B结构域或B结构域片段的核苷酸)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。在其他实施方案中,所述核酸序列与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。在其他实施方案中,所述核苷酸序列包含SEQ ID NO:1的核苷酸58-2277和2320-4374(即,SEQ ID NO:1的核苷酸58-4374,而无编码B结构域或B结构域片段的核苷酸)或SEQ ID NO:1的核苷酸58-4374。在其他实施方案中,所述核苷酸序列包含SEQ ID NO:1的核苷酸1-2277和2320-4374(即,SEQ ID NO:1的核苷酸1-4374,而无编码B结构域或B结构域片段的核苷酸)或SEQ ID NO:1的核苷酸1-4374。

[0183] 在一些实施方案中,本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与SEQ ID NO:2的核苷酸58-4374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。在其他实施方案中,所述核苷酸序列包含与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。在其他实施方案中,所述核酸序列与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。在其他实施方案中,所述核苷酸序列包含SEQ ID NO:2的核苷酸58-2277和2320-4374(即,SEQ ID NO:2的核苷酸58-4374,而无编码B结构域或B结构域片段的核苷酸)或SEQ ID NO:2的核苷酸58-4374。在其他实施方案中,所述核苷酸序列包含SEQ ID NO:2的核苷酸1-2277和2320-4374(即,SEQ ID NO:2的核苷酸1-4374,而无编码B结构域或B结构域片段的核苷酸)或SEQ ID NO:2的核苷酸1-4374。

[0184] 在一些实施方案中,本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与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。在其他实施方案中,所述核苷酸序列包含与SEQ ID NO:70的核苷酸58-2277和2320-4374(即,SEQ ID NO:70的核苷酸58-4374,而无编码B结构域或B结构域片段的核苷酸)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。在其他实施方案中,所述核酸序列与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。在其他实施方案中,所述核苷酸序列包含SEQ ID NO:70的核苷酸58-2277和2320-4374(即,SEQ ID NO:70的核苷酸58-4374,而无编码B结构域或B结构域片段的核苷酸)或SEQ ID NO:70的核苷酸58-4374。在其他实施方案中,所述核苷酸序列包含SEQ ID NO:70的核苷酸1-2277和2320-4374(即,SEQ ID NO:70的核苷酸1-4374,而无编码B结构域或B结构域片段的核苷酸)或SEQ ID NO:70的核苷酸1-4374。

[0185] 在一些实施方案中,本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的

分离的核酸分子,其中所述核苷酸序列包含与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。在其他实施方案中,所述核苷酸序列包含与SEQ ID NO:71的核苷酸58-2277和2320-4374(即,SEQ ID NO:71的核苷酸58-4374,而无编码B结构域或B结构域片段的核苷酸)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。在其他实施方案中,所述核苷酸序列与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。在其他实施方案中,所述核苷酸序列包含SEQ ID NO:71的核苷酸58-2277和2320-4374(即,SEQ ID NO:71的核苷酸58-4374,而无编码B结构域或B结构域片段的核苷酸)或SEQ ID NO:71的核苷酸58-4374。在其他实施方案中,所述核苷酸序列包含SEQ ID NO:71的核苷酸1-2277和2320-4374(即,SEQ ID NO:71的核苷酸1-4374,而无编码B结构域或B结构域片段的核苷酸)或SEQ ID NO:71的核苷酸1-4374。

[0186] 在一些实施方案中,本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与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。在其他实施方案中,所述核苷酸序列包含与SEQ ID NO:3的核苷酸58-2277和2320-4374(即,SEQ ID NO:3的核苷酸58-4374,而无编码B结构域或B结构域片段的核苷酸)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。在某些实施方案中,所述核苷酸序列与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。在一些实施方案中,所述核苷酸序列包含SEQ ID NO:3的核苷酸58-2277和2320-4374(即,SEQ ID NO:3的核苷酸58-4374,而无编码B结构域或B结构域片段的核苷酸)或SEQ ID NO:3的核苷酸58-4374。在其他实施方案中,所述核苷酸序列包含SEQ ID NO:3的核苷酸58-2277和2320-4374(即,SEQ ID NO:3的核苷酸1-4374,而无编码B结构域或B结构域片段的核苷酸)或SEQ ID NO:3的核苷酸1-4374。

[0187] 在一些实施方案中,本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与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。在其他实施方案中,所述核苷酸序列包含与SEQ ID NO:4的核苷酸58-2277和2320-4374(即,SEQ ID NO:4的核苷酸58-4374,而无编码B结构域或B结构域片段的核苷酸)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。在其他实施方案中,所述核苷酸序列与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。在其他实施方案中,所述核苷酸序列包含SEQ ID NO:4的核苷酸58-2277和2320-4374(即,SEQ ID NO:4的核苷酸58-4374,而无编码B结构域或B结构域片段的核苷酸)或SEQ ID NO:4的核苷酸58-4374。在其他实施方案中,所述核苷酸序列包含SEQ ID NO:4的核苷酸1-2277和2320-4374(即,SEQ ID NO:4的核苷酸1-4374,而无编码B结构域或B结构域片段的核苷酸)或SEQ ID NO:4的核苷酸1-4374。

[0188] 在一些实施方案中,本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与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。在其他实施方案中,所述核苷酸序列包含与SEQ ID NO:5的核苷酸58-2277和2320-4374(即,SEQ ID NO:5的核苷酸58-4374,而无编码B结构域或B结构域片段的核苷酸)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。在某些实施方案中,所述核酸序列与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。在一些实施方案中,所述核苷酸序列包含SEQ ID NO:5的核苷酸58-2277和2320-4374(即,SEQ ID NO:5的核苷酸58-4374,而无编码B结构域或B结构域片段的核苷酸)或SEQ ID NO:5的核苷酸58-4374。在其他实施方案中,所述核苷酸序列包含SEQ ID NO:5的核苷酸1-2277和2320-4374(即,SEQ ID NO:5的核苷酸1-4374,而无编码B结构域或B结构域片段的核苷酸)或SEQ ID NO:5的核苷酸1-4374。

[0189] 在一些实施方案中,本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与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。在其他实施方案中,所述核苷酸序列包含与SEQ ID NO:6的核苷酸58-2277和2320-4374(即,SEQ ID NO:6的核苷酸58-4374,而无编码B结构域或B结构域片段的核苷酸)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。在某些实施方案中,所述核酸序列与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。在一些实施方案中,所述核苷酸序列包含SEQ ID NO:6的核苷酸58-2277和2320-4374(即,SEQ ID NO:6的核苷酸58-4374,而无编码B结构域或B结构域片段的核苷酸)或SEQ ID NO:6的核苷酸58-4374。在其他实施方案中,所述核苷酸序列包含SEQ ID NO:6的核苷酸1-2277和2320-4374(即,SEQ ID NO:6的核苷酸1-4374,而无编码B结构域或B结构域片段的核苷酸)或SEQ ID NO:6的核苷酸1-4374。

[0190] 在一些实施方案中,所述核苷酸序列包含编码信号肽的核酸序列。在某些实施方案中,所述信号肽是FVIII信号肽。在一些实施方案中,所述编码信号肽的核酸序列经过密码子优化。在一个特定实施方案中,所述编码信号肽的核酸序列与(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。

[0191] SEQ ID NO:1-6、70和71是开始或“亲本”或“野生型”FVIII核苷酸序列SEQ ID NO:16的优化形式。SEQ ID NO:16编码B结构域缺失的人FVIII。虽然SEQ ID NO:1-6、70和71是源于FVIII的特定B结构域缺失形式(SEQ ID NO:16),但应理解,本公开还针对编码FVIII的

其他形式的核酸的优化形式。例如，FVIII的其他形式可包括全长FVIII、FVIII的其他B结构域缺失(下文所述)或保持FVIII活性的FVIII的其他片段。

[0192] 除非另外规定,否则如本文所用,“具有FVIII活性的多肽”意指在凝固方面具有其正常作用的功能性FVIII多肽。术语具有FVIII活性的多肽包括保持全长野生型因子VIII在凝固途径中的功能的其功能性片段、变体、类似物或衍生物。“具有FVIII活性的多肽”可与FVIII蛋白、FVIII多肽或FVIII互换使用。FVIII功能的实例包括但不限于活化凝固的能力、用作因子IX的辅因子的能力或在Ca²⁺和磷脂存在下与因子IX形成因子X酶复合物(tenase complex)的能力,所述因子X酶复合物接着将因子X转化为活化形式Xa。在一个实施方案中,具有FVIII活性的多肽包含两条多肽链,具有FVIII重链的第一链和具有FVIII轻链的第二链。在另一实施方案中,具有FVIII活性的多肽是单链FVIII。单链FVIII可在对应于成熟FVIII序列的氨基酸残基1645和/或1648处含有一种或多种突变或取代。参见国际申请号PCT/US2012/045784,其以引用的方式整体并入本文中。所述FVIII蛋白可以是人、猪、犬、大鼠或鼠科动物FVIII蛋白。另外,来自人和其他物种的FVIII之间的比较已经鉴别出对功能可能所需的保守残基(Cameron等人,Thromb.Haemost.79:317-22(1998);US 6,251,632)。

[0193] 如本文所用,FVIII的“B结构域”与本领域中已知的B结构域相同,本领域中已知的B结构域由内部氨基酸序列同一性和通过凝血酶进行蛋白水解裂解的位点限定(例如,全长人FVIII的残基Ser741-Arg1648)。其他人FVIII域由以下氨基酸残基限定:A1,残基Ala1-Arg372;A2,残基Ser373-Arg740;A3,残基Ser1690-Ile2032;C1,残基Arg2033-Asn2172;C2,残基Ser2173-Tyr2332。A3-C1-C2序列包括残基Ser1690-Tyr2332。剩余序列(残基Glu1649-Arg1689)通常称作FVIII轻链活化肽。猪、小鼠和犬FVIII的所有结构域(包括所述B结构域)的边界的位置也是本领域中已知的。BDD FVIII的一个实例是REFACTO®重组BDD FVIII(Wyeth Pharmaceuticals,Inc.)。

[0194] “B结构域缺失的FVIII”可以具有美国专利号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中所公开的全部或部分缺失,所述专利中的每一者均以引用的方式整体并入本文中。在一些实施方案中,本公开的B结构域缺失的FVIII序列包含美国专利号6,316,226的第4栏第4行至第5栏第28行和实施例1-5中所公开的缺失中的任一者(也在US 6,346,513中)。在一些实施方案中,本公开的B结构域缺失的FVIII具有美国专利号5,789,203的第2栏第26-51行和实施例5-8中所公开的缺失(另外,US 6,060,447、US 5,595,886和US 6,228,620)。在一些实施方案中,B结构域缺失的FVIII具有美国专利号5,972,885的第1栏第25行至第2栏第40行;美国专利号6,048,720的第6栏第1-22行和实施例1;美国专利号5,543,502的第2栏第17-46行;美国专利号5,171,844的第4栏第22行至第5栏第36行;美国专利号5,112,950的第2栏第55-68行、图2和实施例1;美国专利号4,868,112的第2栏第2行至第19栏第21行和表2;美国专利号7,041,635的第2栏第1行至第3栏第19行、第3栏第40行至第4栏第67行、第7栏第43行至第8栏第26行和第11栏第5行至第13栏第39行;或美国专利号6,458,563的第4栏第25-53行中所述的缺失。在一些实施方案中,B结构域缺失的FVIII具有B结构域中的大部分的缺失,但仍含有B结构域中将初级翻译产物体内蛋白水解加工为两条多肽链所必需的氨基端序列,如WO 91/09122中所公开,WO 91/09122以引用的方式整体并入本文中。在一些实施方案中,B结构域缺失的FVIII用氨基

酸747-1638的缺失,即实际上所述B结构域的完全缺失构建。Hoeben R.C.等人J.Biol.Chem.265(13):7318-7323(1990),其以引用的方式整体并入本文中。B结构域缺失的FVIII也可以含有FVIII的氨基酸771-1666或氨基酸868-1562的缺失。Meulien P.等人Protein Eng.2(4):301-6(1988),其以引用的方式整体并入本文中。作为本公开的一部分的另外的B结构域缺失包括例如:氨基酸982-1562或760-1639(Toole等人,Proc.Natl.Acad.Sci.U.S.A.(1986)83,5939-5942))、797-1562(Eaton等人Biochemistry(1986)25:8343-8347))、741-1646(Kaufman(PCT公布申请号W0 87/04187))、747-1560(Sarver等人,DNA(1987)6:553-564))、741-1648(Pasek(PCT申请号88/00831))、816-1598或741-1689(Lagner(Behring Inst.Mitt.(1988)No 82:16-25,EP 295597))的缺失,其中每一者均以引用的方式整体并入本文中。前述缺失中的每一者均可在任何FVIII序列中产生。

[0195] 多种包括B结构域缺失的功能性FVIII分子公开于以下专利中:受让给Baxter的US 6,316,226和US 6,346,513;受让给In2Gen的US 7,041,635;受让给Chiron的US 5,789,203、US 6,060,447、US 5,595,886和US 6,228,620;受让给Biovitrum的US 5,972,885和US 6,048,720;受让给Novo Nordisk的US 5,543,502和US 5,610,278;受让给Immuno Ag的US 5,171,844;受让给Transgene S.A.的US 5,112,950;受让给Genetics Institute的US 4,868,112,其中每一者均以引用的方式整体并入本文中。

[0196] 密码子优化

[0197] 在一个实施方案中,本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核酸序列已经过密码子优化。在另一实施方案中,编码具有FVIII活性的多肽并且经受密码子优化的起始核酸序列是SEQ ID NO:16。在一些实施方案中,编码具有FVIII活性的多肽的序列针对人表达经过密码子优化。在其他实施方案中,编码具有FVIII活性的多肽的序列针对鼠科动物表达经过密码子优化。SEQ ID NO:1-6、70和71是SEQ ID NO:16的密码子优化形式,针对人表达经过优化。

[0198] 当术语“密码子优化”是指用于多种宿主的转化的核酸分子的基因或编码区时,所述术语是指所述核酸分子的基因或编码区中的密码子的改变以反映宿主生物体的典型密码子使用而不改变由DNA编码的多肽。所述优化包括用更频繁用于所述生物体的基因中的一种或多种密码子置换至少一种或超过一种或大部分密码子。

[0199] 包含编码任何多肽链的氨基酸的密码子的核苷酸序列中的偏差允许编码所述基因的序列的变异。由于各密码子由三个核苷酸组成,并且构成DNA的核苷酸限于四种特定碱基,存在核苷酸的64种可能组合,其中61种编码氨基酸(剩余三种密码子编码终止翻译的信号)。显示何种密码子编码何种氨基酸的“遗传密码”以表1复制于本文中。结果,多种氨基酸由超过一种密码子指定。例如,氨基酸丙氨酸和脯氨酸由四个三联体编码,丝氨酸和精氨酸由六个三联体编码,而色氨酸和甲硫氨酸仅由一个三联体编码。这种简并性允许DNA碱基组成在广泛范围内改变而不改变由所述DNA编码的蛋白质的氨基酸序列。

[0200] 表1:标准遗传密码

[0201]

	T	C	A	G
T	TTT Phe (F) TTC " TTA Leu (L) TTG "	TCT Ser (S) TCC " TCA " TCG "	TAT Tyr (Y) TAC " TAA 终止 TAG 终止	TGT Cys (C) TGC TGA 终止 TGG Trp (W)
C	CTT Leu (L) CTC " CTA " CTG "	CCT Pro (P) CCC " CCA " CCG "	CAT His (H) CAC " CAA Gln (Q) CAG "	CGT Arg (R) CGC " CGA " CGG "
A	ATT Ile (I) ATC " ATA " ATG Met (M)	ACT Thr (T) ACC " ACA " ACG "	AAT Asn (N) AAC " AAA Lys (K) AAG "	AGT Ser (S) AGC " AGA Arg (R) AGG "
G	GTT Val (V) GTC " GTA " GTG "	GCT Ala (A) GCC " GCA " GCG "	GAT Asp (D) GAC " GAA Glu (E) GAG "	GGT Gly (G) GGC " GGA " GGG "

[0202] 多种生物体展示使用特定密码子来编码的偏好以将特定氨基酸插入生长的肽链中。生物体之间密码子使用的密码子偏爱或密码子偏好差异由遗传密码的简并性提供,并且充分记录于多种生物体中。密码子偏好通常与信使RNA (mRNA) 的翻译效率有关,相信信使RNA (mRNA) 的翻译效率又尤其依赖于正在翻译的密码子的特性和特定转运RNA (tRNA) 分子的可用性。细胞中所选择的tRNA的优势一般是最频繁用于肽合成的密码子的反映。因此,基因可基于密码子优化经过调适用于给定生物体中的最佳基因表达。

[0203] 鉴于多种动物、植物和微生物物种的大量基因序列可用,已经计算出密码子使用的相对频率。密码子使用表可例如在www.kazusa.or.jp/codon/ (2012年6月18日访问) 中可获得的“密码子使用数据库”中获得。参见Nakamura, Y. 等人Nucl. Acids Res. 28:292 (2000)。

[0204] 以优化的频率随机地分配密码子以编码给定多肽序列可以通过计算各氨基酸的密码子频率,并且接着将所述密码子随机地分配至所述多肽序列而手动地进行。另外,多种算法和计算机软件程序可用于计算最佳序列。

[0205] 在一些实施方案中,所述核酸分子包含一种或多种特性:(a) 所述核酸分子或其一部分相对于SEQ ID NO:16具有增加的人密码子适应指数;(b) 所述核苷酸序列或其一部分相对于SEQ ID NO:16具有增加的最优密码子频率;(c) 与SEQ ID NO:16中的G/C核苷酸百分比相比,所述核苷酸序列或其一部分含有较高百分比的G/C核苷酸;(d) 所述核苷酸序列或其一部分相对于SEQ ID NO:16具有增加的相对同义密码子使用;(e) 所述核苷酸序列或其一部分相对于SEQ ID NO:16是降低的有效数目的密码子;(f) 所述核苷酸序列相对于SEQ

ID NO:16含有较少MARS/ARS序列(SEQ ID NO:21和22);(g)所述核苷酸序列相对于SEQ ID NO:16含有较少去稳定元件(SEQ ID NO:23和24);(i)所述核苷酸序列不含poly-T序列,(j)所述核苷酸序列不含poly-A序列;或(k)其任何组合。在一些实施方案中,所述核酸分子含有(a)-(j)中的至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、至少八种、至少九种或十种特征。

[0206] 密码子适应指数

[0207] 在一个实施方案中,本公开提供包含编码具有FVIII活性的多肽的本文所述的核苷酸序列的分离的核酸分子,其中所述人密码子适应指数相对于SEQ ID NO:16增加。例如,所述核苷酸序列可具有至少约0.75 (75%)、至少约0.76 (76%)、至少约0.77 (77%)、至少约0.78 (78%)、至少约0.79 (79%)、至少约0.80 (80%)、至少约0.81 (81%)、至少约0.82 (82%)、至少约0.83 (83%)、至少约0.84 (84%)、至少约0.85 (85%)、至少约0.86 (86%)、至少约0.87 (87%)、至少约0.88 (88%)、至少约0.89 (89%)、至少约0.90 (90%)、至少约0.91 (91%)、至少约0.92 (92%)、至少约0.93 (93%)、至少约0.94 (94%)、至少约0.95 (95%)、至少约0.96 (96%)、至少约0.97 (97%)、至少约0.98 (98%)或至少约0.99 (99%)的人密码子适应指数。在一些实施方案中,所述核苷酸序列具有至少约.88 (88%)的人密码子适应指数。在其他实施方案中,所述核苷酸序列具有至少约.91 (91%)的人密码子适应指数。在其他实施方案中,所述核苷酸序列具有至少约.91 (97%)的人密码子适应指数。

[0208] 在一个特定实施方案中,本公开提供包含如下核苷酸序列的分离的核酸分子,所述核苷酸序列包含编码FVIII多肽的N端部分的第一核苷酸序列和编码FVIII多肽的C端部分的第二核苷酸序列;其中所述第一核苷酸序列与(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具有至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性;其中所述N端部分和所述C端部分合起来具有FVIII多肽活性;并且其中所述核苷酸序列的人密码子适应指数相对于SEQ ID NO:16增加。在一些实施方案中,所述核苷酸序列具有至少约0.75 (75%)、至少约0.76 (76%)、至少约0.77 (77%)、至少约0.78 (78%)、至少约0.79 (79%)、至少约0.80 (80%)、至少约0.81 (81%)、至少约0.82 (82%)、至少约0.83 (83%)、至少约0.84 (84%)、至少约0.85 (85%)、至少约0.86 (86%)、至少约0.87 (87%)、至少约0.88 (88%)、至少约0.89 (89%)、至少约0.90 (90%)或至少约.91 (91%)的人密码子适应指数。在一个特定中,所述核苷酸序列具有至少约.88 (88%)的人密码子适应指数。在另一实施方案中,所述核苷酸序列具有至少约.91 (91%)的人密码子适应指数。

[0209] 在另一实施方案中,本公开提供包含如下核苷酸序列的分离的核酸分子,所述核苷酸序列包含编码FVIII多肽的N端部分的第一核苷酸序列和编码FVIII多肽的C端部分的第二核苷酸序列;其中所述第二核苷酸序列与(i) SEQ ID NO:5的核苷酸1792-2277和2320-4374或(ii) SEQ ID NO:6的1792-2277和2320-4374具有至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性;其中所述N端部分和所述C端部分合起来具有FVIII多肽活性;并且其中所述核苷酸序列的人

密码子适应指数相对于SEQ ID NO:16增加。在一些实施方案中,所述核苷酸序列具有至少约0.75 (75%)、至少约0.76 (76%)、至少约0.77 (77%)、至少约0.78 (78%)、至少约0.79 (79%)、至少约0.80 (80%)、至少约0.81 (81%)、至少约0.82 (82%)、至少约0.83 (83%)、至少约0.84 (84%)、至少约0.85 (85%)、至少约0.86 (86%)、至少约0.87 (87%)或至少约0.88 (88%)的人密码子适应指数。在一个特定中,所述核苷酸序列具有至少约.83 (83%)的人密码子适应指数。在另一实施方案中,所述核苷酸序列具有至少约.88 (88%)的人密码子适应指数。

[0210] 在其他实施方案中,本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与选自SEQ ID NO:1、2、3、4、5、6、70和71的氨基酸序列的核苷酸58-2277和2320-4374(即,SEQ ID NO:1、2、3、4、5、6、70和71的核苷酸58-4374,而无编码B结构域或B结构域片段的核苷酸)具有至少约80%、至少约85%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列;并且其中所述核苷酸序列的人密码子适应指数相对于SEQ ID NO:16增加。在一些实施方案中,所述核苷酸序列具有至少约0.75 (75%)、至少约0.76 (76%)、至少约0.77 (77%)、至少约0.78 (78%)、至少约0.79 (79%)、至少约0.80 (80%)、至少约0.81 (81%)、至少约0.82 (82%)、至少约0.83 (83%)、至少约0.84 (84%)、至少约0.85 (85%)、至少约0.86 (86%)、至少约0.87 (87%)或至少约0.88 (88%)的人密码子适应指数。在一个特定中,所述核苷酸序列具有至少约.75 (75%)的人密码子适应指数。在另一实施方案中,所述核苷酸序列具有至少约.83 (83%)的人密码子适应指数。在另一实施方案中,所述核苷酸序列具有至少约.88 (88%)的人密码子适应指数。在另一实施方案中,所述核苷酸序列具有至少约.91 (91%)的人密码子适应指数。在另一实施方案中,所述核苷酸序列具有至少约.97 (97%)的人密码子适应指数。

[0211] 在一些实施方案中,本公开的分离的核酸分子相对于SEQ ID NO:16具有增加的最优密码子频率(FOP)。在某些实施方案中,所述分离的核酸分子的FOP是至少约40、至少约45、至少约50、至少约55、至少约60、至少约64、至少约65、至少约70、至少约75、至少约79、至少约80、至少约85或至少约90。

[0212] 在其他实施方案中,本公开的分离的核酸分子相对于SEQ ID NO:16具有增加的相对同义密码子使用(RCSU)。在一些实施方案中,所述分离的核酸分子的RCSU大于1.5。在其他实施方案中,所述分离的核酸分子的RCSU大于2.0。在某些实施方案中,所述分离的核酸分子的RCSU是至少约1.5、至少约1.6、至少约1.7、至少约1.8、至少约1.9、至少约2.0、至少约2.1、至少约2.2、至少约2.3、至少约2.4、至少约2.5、至少约2.6或至少约2.7。

[0213] 在其他实施方案中,本公开的分离的核酸分子相对于SEQ ID NO:16具有减少的有效数目的密码子。在一些实施方案中,所述分离的核酸分子具有少于约50、少于约45、少于约40、少于约35、少于约30或少于约25的有效数目的密码子。在一个特定实施方案中,所述分离的核酸分子具有约40、约35、约30、约25或约20的有效数目的密码子。

[0214] G/C含量优化

[0215] 在一些实施方案中,本公开提供包含编码具有FVIII活性的多肽的本文所述的核苷酸序列的分离的核酸分子,其中与SEQ ID NO:16中的G/C核苷酸百分比相比,所述核苷酸序列含有较高百分比的G/C核苷酸。在其他实施方案中,所述编码具有FVIII活性的多肽的

核苷酸序列具有至少约45%、至少约46%、至少约47%、至少约48%、至少约49%、至少约50%、至少约51%、至少约52%、至少约53%、至少约54%、至少约55%、至少约56%、至少约57%、至少约58%、至少约59%或至少约60%的G/C含量。

[0216] 在一个特定实施方案中,本公开提供包含如下核苷酸序列的分离的核酸分子,所述核苷酸序列包含编码FVIII多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列;其中所述第一核酸序列与(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具有至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性;其中所述N端部分和所述C端部分合起来具有FVIII多肽活性;并且其中与SEQ ID NO:16中的G/C核苷酸百分比相比,所述核苷酸序列含有较高百分比的G/C核苷酸。在一些实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列具有至少约45%、至少约46%、至少约47%、至少约48%、至少约49%、至少约50%、至少约51%、至少约52%、至少约53%、至少约54%、至少约55%、至少约56%、至少约57%或至少约58%的G/C含量。在一个特定实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列具有至少约58%的G/C含量。

[0217] 在另一实施方案中,本公开提供包含如下核苷酸序列的分离的核酸分子,所述核苷酸序列包含编码FVIII多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列;其中所述第二核酸序列与(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,SEQ ID NO:5的核苷酸1792-4374,而无编码B结构域或B结构域片段的核苷酸)或(iv) SEQ ID NO:6的核苷酸1792-2277和2320-4374(即,SEQ ID NO:6的核苷酸1792-4374,而无编码B结构域或B结构域片段的核苷酸)具有至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性;其中所述N端部分和所述C端部分合起来具有FVIII多肽活性;并且其中与SEQ ID NO:16中的G/C核苷酸百分比相比,所述核苷酸序列含有较高百分比的G/C核苷酸。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列具有至少约45%、至少约46%、至少约47%、至少约48%、至少约49%、至少约50%、至少约51%、至少约52%、至少约53%、至少约54%、至少约55%、至少约56%或至少约57%的G/C含量。在一个特定实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列具有至少约52%的G/C含量。在另一实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列具有至少约55%的G/C含量。在另一实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列具有至少约57%的G/C含量。

[0218] 在其他实施方案中,本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与选自SEQ ID NO:1、2、3、4、5、6、70和71的氨基酸序列的(i)核苷酸58-4374或(ii)核苷酸58-2277和2320-4374(即,SEQ ID NO:1、2、3、4、5、6、70和71的核苷酸58-4374,而无编码B结构域或B结构域片段的核苷酸)具有至少约80%、至少约85%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸

序列;并且其中与SEQ ID NO:16中的G/C核苷酸百分比相比,所述核苷酸序列含有较高百分比的G/C核苷酸。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列具有至少约45%的G/C含量。在一个特定实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列具有至少约52%的G/C含量。在另一实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列具有至少约55%的G/C含量。在另一实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列具有至少约57%的G/C含量。在另一实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列具有至少约58%的G/C含量。在另一实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列具有至少约60%的G/C含量。

[0219] “G/C含量”(或鸟嘌呤-胞嘧啶含量)或“G/C核苷酸百分比”是指DNA分子中含氮碱基(为鸟嘌呤或胞嘧啶)的百分比。G/C含量可以使用下式计算:

$$[0220] \quad \frac{G+C}{A+T+G+C} \times 100 \quad (III)$$

[0221] 人基因在其G/C含量方面是高度不同的,其中一些基因具有低达20%的G/C含量,并且其他基因具有高达95%的G/C含量。一般说来,富G/C基因更高地表达。事实上,已经证明了增加基因的G/C含量可以引起所述基因的增加的表达,这主要是归因于转录的增加和较高的稳态mRNA水平。参见Kudla等人,PLoS Biol.,4(6):e180(2006)。

[0222] 基质附着区样序列

[0223] 在一些实施方案中,本公开提供包含编码具有FVIII活性的多肽的本文所述的核苷酸序列的分离的核酸分子,其中所述核苷酸序列相对于SEQ ID NO:16含有较少MARS/ARS序列。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列含有至多6个、至多5个、至多4个、至多3个或至多2个MARS/ARS序列。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列含有至多1个MARS/ARS序列。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列不含MARS/ARS序列。

[0224] 在一个特定实施方案中,本公开提供包含如下核苷酸序列的分离的核酸分子,所述核苷酸序列包含编码FVIII多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列;其中所述第一核酸序列与(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具有至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性;其中所述N端部分和所述C端部分合起来具有FVIII多肽活性;并且其中所述核苷酸序列相对于SEQ ID NO:16含有较少MARS/ARS序列。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列含有至多6个、至多5个、至多4个、至多3个或至多2个MARS/ARS序列。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列含有至多1个MARS/ARS序列。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列不含MARS/ARS序列。

[0225] 在另一实施方案中,本公开提供包含如下核苷酸序列的分离的核酸分子,所述核苷酸序列包含编码FVIII多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列;其中所述第二核酸序列与(i) SEQ ID NO:5的核苷酸1792-4374;(ii) SEQ ID

N0:6的核苷酸1792-4374; (iii) SEQ ID N0:5的核苷酸1792-2277和2320-4374 (即, SEQ ID N0:5的核苷酸1792-4374, 而无编码B结构域或B结构域片段的核苷酸) 或 (iv) SEQ ID N0:6的核苷酸1792-2277和2320-4374 (即, SEQ ID N0:6的核苷酸1792-4374, 而无编码B结构域或B结构域片段的核苷酸) 具有至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性; 其中所述N端部分和所述C端部分合起来具有FVIII多肽活性; 并且其中所述核苷酸序列相对于SEQ ID N0:16含有较少MARS/ARS序列。在其他实施方案中, 所述编码具有FVIII活性的多肽的核苷酸序列含有至多6个、至多5个、至多4个、至多3个或至多2个MARS/ARS序列。在其他实施方案中, 所述编码具有FVIII活性的多肽的核苷酸序列含有至多1个MARS/ARS序列。在其他实施方案中, 所述编码具有FVIII活性的多肽的核苷酸序列不含MARS/ARS序列。

[0226] 在其他实施方案中, 本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子, 其中所述核苷酸序列包含与 (i) SEQ ID N0:1、2、3、4、5、6、70或71的核苷酸58-4374或 (ii) SEQ ID N0:1、2、3、4、5、6、70或71的核苷酸58-2277和2320-4374 (即, SEQ ID N0:1、2、3、4、5、6、70或71的核苷酸58-4374, 而无编码B结构域或B结构域片段的核苷酸) 具有至少约80%、至少约85%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列; 并且其中所述核苷酸序列相对于SEQ ID N0:16含有较少MARS/ARS序列。在其他实施方案中, 所述编码具有FVIII活性的多肽的核苷酸序列含有至多6个、至多5个、至多4个、至多3个或至多2个MARS/ARS序列。在其他实施方案中, 所述编码具有FVIII活性的多肽的核苷酸序列含有至多1个MARS/ARS序列。在其他实施方案中, 所述编码具有FVIII活性的多肽的核苷酸序列不含MARS/ARS序列。

[0227] 已经鉴别出人FVIII核苷酸序列中与酿酒酵母自主复制序列 (ARS) 和核基质附着区 (MAR) 共享序列相似性的富AT元件。(Fallux等人, Mol. Cell. Biol. 16:4264-4272 (1996)。已经证明这些元件之一体外结合核因子并且抑制氯霉素乙酰转移酶 (CAT) 报告基因的表达。同上。已经假设这些序列可以促进人FVIII基因的转录抑制。因此, 在一个实施方案中, 所有MAR/ARS序列均在本公开的FVIII基因中消除。在亲本FVIII序列 (SEQ ID N0:16) 中存在四个MAR/ARS ATATT序列 (SEQ ID N0:21) 和三个MAR/ARS AAATAT序列 (SEQ ID N0:22)。使所有这些位点均发生突变以破坏优化的FVIII序列 (SEQ ID N0:1-6) 中的MAR/ARS序列。在所述优化序列中这些元件中的每一者的位置和相应核苷酸的序列显示于下表2中。

[0228] 表2: 抑制性元件的改变的概述

[0229]

元件的位置	起始 BDD FVIII 序列 (SEQ ID NO:16)	优化的 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
去稳定序列									
639	ATTTA	GTT TA	GTTC A	GTT CA	GTT CA	GTTC A	GTTC A	GTTC A	GTTC A
1338	ATTTA	GTT TA	GTTC A	CTT CA	GTT CA	GTTC A	GTTC A	CTTC A	GTTC A
1449	ATTTA	CTT TA	CTTC A	CTT CA	CTT CA	CTTC A	CTTC A	CTTC A	CTTC A
1590	TAAAT	TAA AT	CAA GT	CA AGT	TAA GT	CAAG T	CAAG T	CAAG T	TAAG T
1623	TAAAT	CAA AA	GAA GA	CTA AG	CAA GA	CAAG A	CAAG A	TAAG T	CAAG A
2410	ATTTA	ATC TA	ATCT A	ATC TA	ATC TA	ATCT A	ATCT A	ATCT A	ATCT A
2586	ATTTA	GTT TA	GTTC A	GTT CA	GTT CA	GTTC A	GTTC A	GTTC A	GTTC A
2630	TAAAT	TGA AT	TGA AC	TGA AC	TGA AC	TCAA T	TGAA C	TCAA T	TGAA C
3884	ATTTA	ATC TG	ACCT G	ACC TG	ACC TG	ATCT G	ACCT G	ATCT G	ACCT G
3887	TAAAT	TGA AC	TGA AC	TGA AC	TGA AC	TGAA C	TGAA C	TGAA C	TGAA C
潜在启动子结合位点									
641	TTATA	TTA TC	TCAT C	TCA TT	TCA TC	TCAT C	TCAT C	TCAT T	TCAT C
1275	TATAA	CTA TA	TTAC A	CTA CA	GTA CA	CTAC A	CTAC A	CTAC A	GTAC A
1276	TTATA	TAT AA	TAC AA	TAC AA	TAC AA	TACA A	TACA A	TACA A	TACA A

[0230]

1445	TTATA	TCA TC	TCAT C	TTA TC	TCA TC	TCAT C	TCAT C	TTAT C	TCAT C
1474	TATAA	TAT AA	TAC AA	TAC AA	TAC AA	TACA A	TACA A	TACA A	TACA A
1588	TATAA	TAT AA	TAC AA	TAC AA	TAT AA	TACA A	TACA A	TACA A	TATA A
2614	TTATA	CTG TA	CTGT A	CTG TA	CTG TA	TTGT A	CTGT A	TTGT A	CTGT A
2661	TATAA	CAT CA	CATC A	CAT CA	CAT CA	CATC A	CATC C	CATC A	CATC C
3286	TATAA	TAT AA	TAC AA	TAC AA	TAC AA	TACA A	TACA A	TACA A	TACA A
3840	TTATA	TTA TA	TTAC T	CTA CA	CTA CA	CTAC A	CTAC T	CTAC A	CTAC T
基质附着样序列(MARS/ARS)									
1287	ATATTT	GTA TCT	GTA CCT	GTA CCT	GTA TCT	GTAC CT	GTAC CT	GTAC CT	GTAT CT
1447	ATATTT	ATC TTT	ATCT TC	ATC TTC	ATC TTC	ATCT TC	ATCT TC	ATCT TC	ATCT TC
1577	AAATAT	AAA TCT	AGA TCT	AA ATC T	AAA TCT	AGAT CT	AGAT CT	AAAT CT	AAAT CT
1585	AAATAT	AAG TAT	AAG TAC	AA GTA C	AAG TAT	AAGT AC	AAGT AC	AAGT AC	AAGT AT
2231	ATATTT	ACA TCA	ATAT CA	AC ATC A	ACA TCA	ACAT CT	ATAT CT	ACAT CT	ATAT CT
3054	AAATAT	AAA CAT	GAA TAT	GA AC AT	GAA CAT	GAAC AT	GAAT AT	GAAC AT	GAAT AT
3788	ATATTT	ATA TCT	ATAT CT	AC ATC T	ACA TCT	ACAT CT	ACAT CT	ACAT CT	ACAT CT
富 AU 序列元件(ARE)									
2468	1. TTTTATT	ACT TCA TC	ACTT CATC	ACT TCA TT	ACT TCA TT	ACTT TATT	ACTT TATC	ACTT TATT	ACTT TATC
3790	2. TTTTTAA	ATC TTT AA	ATCT TCA A	ATC TTC AA	ATC TTC AA	ATCT TCAA	ATCT TCAA	ATCT TCAA	ATCT TCAA
Poly A/Poly T 序列									
3273	AAAAAA A	GAA AAA A	GAA GAA G	GA AG AA G	GAA GAA G	GAA GAA G	CAAG AAG	GAA GAA G	CAAG AAG
4195	TTTTTT	TTC TTT	TTCT TC	TTC TTC	TTC TTC	TTCT TC	TTCT TC	TTCT TCC	TTCT TCC
剪接位点									
2203	GGTGAT	GGG GAC	GGC GAC	GG GG AC	GGG GAC	GGA GAC	GGA GAC	GGA GAC	GGA GAC

[0231] 去稳定序列

[0232] 在一些实施方案中,本公开提供包含编码具有FVIII活性的多肽的本文所述的核苷酸序列的分离的核酸分子,其中所述核苷酸序列相对于SEQ ID NO:16含有较少去稳定元件。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列含有至多9个、至多8个、至多7个、至多6个或至多5个去稳定元件。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列含有至多4个、至多3个、至多2个或至多1个去稳定元件。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列不含去稳定元件。

[0233] 在一个特定实施方案中,本公开提供包含如下核苷酸序列的分离的核酸分子,所述核苷酸序列包含编码FVIII多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列;其中所述第一核酸序列与(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具有至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性;其中所述N端部分和所述C端部分合起来具有FVIII多肽活性;并且其中所述核苷酸序列相对于SEQ ID NO:16含有较少去稳定元件。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列含有至多9个、至多8个、至多7个、至多6个或至多5个去稳定元件。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列含有至多4个、至多3个、至多2个或至多1个去稳定元件。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列不含去稳定元件。

[0234] 在另一实施方案中,本公开提供包含如下核苷酸序列的分离的核酸分子,所述核苷酸序列包含编码FVIII多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列;其中所述第二核酸序列与(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,SEQ ID NO:5的核苷酸1792-4374,而无编码B结构域或B结构域片段的核苷酸);或(iv) SEQ ID NO:6的核苷酸1792-2277和2320-4374(即,SEQ ID NO:6的核苷酸1792-4374,而无编码B结构域或B结构域片段的核苷酸)具有至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性;其中所述N端部分和所述C端部分合起来具有FVIII多肽活性;并且其中所述核苷酸序列相对于SEQ ID NO:16含有较少去稳定元件。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列含有至多9个、至多8个、至多7个、至多6个或至多5个去稳定元件。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列含有至多4个、至多3个、至多2个或至多1个去稳定元件。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列不含去稳定元件。

[0235] 在其他实施方案中,本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与(i)选自SEQ ID NO:1、2、3、4、5、6、70和71的氨基酸序列的核苷酸58-4374或(ii)选自SEQ ID NO:1、2、3、4、5、6、70和71的氨基酸序列的核苷酸58-2277和2320-4374(即,SEQ ID NO:1、2、3、4、5、6、70或71的核苷酸58-4374,而无编码B结构域或B结构域片段的核苷酸)具有至少约80%、至少约85%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列;并且其中所述核苷酸序列相对于

SEQ ID NO:16含有较少去稳定元件。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列含有至多9个、至多8个、至多7个、至多6个或至多5个去稳定元件。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列含有至多4个、至多3个、至多2个或至多1个去稳定元件。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列不含去稳定元件。

[0236] 在亲本FVIII序列(SEQ ID NO:16)中存在十个去稳定元件:六个ATTTA序列(SEQ ID NO:23)和四个TAAAT序列(SEQ ID NO:24)。在一个实施方案中,使这些位点的序列发生突变以破坏优化的FVIII SEQ ID NO:1-6、70和71中的去稳定元件。在所述优化序列中这些元件中的每一者的位置和相应核苷酸的序列显示于表2中。

[0237] 潜在启动子结合位点

[0238] 在一些实施方案中,本公开提供包含编码具有FVIII活性的多肽的本文所述的核苷酸序列的分离的核酸分子,其中所述核苷酸序列相对于SEQ ID NO:16含有较少潜在启动子结合位点。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列含有至多9个、至多8个、至多7个、至多6个或至多5个潜在启动子结合位点。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列含有至多4个、至多3个、至多2个或至多1个潜在启动子结合位点。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列不含潜在启动子结合位点。

[0239] 在一个特定实施方案中,本公开提供包含如下核苷酸序列的分离的核酸分子,所述核苷酸序列包含编码FVIII多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列;其中所述第一核酸序列与(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具有至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性;其中所述N端部分和所述C端部分合起来具有FVIII多肽活性;并且其中所述核苷酸序列相对于SEQ ID NO:16含有较少潜在启动子结合位点。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列含有至多9个、至多8个、至多7个、至多6个或至多5个潜在启动子结合位点。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列含有至多4个、至多3个、至多2个或至多1个潜在启动子结合位点。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列不含潜在启动子结合位点。

[0240] 在另一实施方案中,本公开提供包含如下核苷酸序列的分离的核酸分子,所述核苷酸序列包含编码FVIII多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列;其中所述第二核酸序列与(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,SEQ ID NO:5的核苷酸1792-4374,而无编码B结构域或B结构域片段的核苷酸);或(iv) SEQ ID NO:6的核苷酸1792-2277和2320-4374(即,SEQ ID NO:6的核苷酸1792-4374,而无编码B结构域或B结构域片段的核苷酸)具有至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性;其中所述N端部分和

所述C端部分合起来具有FVIII多肽活性;并且其中所述核苷酸序列相对于SEQ ID NO:16含有较少潜在启动子结合位点。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列含有至多9个、至多8个、至多7个、至多6个或至多5个潜在启动子结合位点。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列含有至多4个、至多3个、至多2个或至多1个潜在启动子结合位点。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列不含潜在启动子结合位点。

[0241] 在其他实施方案中,本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与(i)选自SEQ ID NO:1、2、3、4、5、6、70和71的氨基酸序列的核苷酸58-4374或(ii)选自SEQ ID NO:1、2、3、4、5、6、70和71的氨基酸序列的核苷酸58-2277和2320-4374(即,SEQ ID NO:1、2、3、4、5、6、70或71的核苷酸58-4374,而无编码B结构域或B结构域片段的核苷酸)具有至少约80%、至少约85%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列;并且其中所述核苷酸序列相对于SEQ ID NO:16含有较少潜在启动子结合位点。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列含有至多9个、至多8个、至多7个、至多6个或至多5个潜在启动子结合位点。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列含有至多4个、至多3个、至多2个或至多1个潜在启动子结合位点。在其他实施方案中,所述编码具有FVIII活性的多肽的核苷酸序列不含潜在启动子结合位点。

[0242] TATA盒是通常在真核生物的启动子区中发现的调控序列。其充当通用转录因子TATA结合蛋白(TBP)的结合位点。TATA盒通常包含序列TATAA(SEQ ID NO:28)或闭合变体。然而,编码序列内的TATA盒可以抑制全长蛋白质的翻译。在野生型BDD FVIII序列(SEQ ID NO:16)中存在十个潜在启动子结合序列:五个TATAA序列(SEQ ID NO:28)和五个TTATA序列(SEQ ID NO:29)。在一些实施方案中,至少1个、至少2个、至少3个或至少4个启动子结合位点在本公开的FVIII基因中消除。在一些实施方案中,至少5个启动子结合位点在本公开的FVIII基因中消除。在其他实施方案中,至少6个、至少7个或至少8个启动子结合位点在本公开的FVIII基因中消除。在一个实施方案中,至少9个启动子结合位点在本公开的FVIII基因中消除。在一个特定实施方案中,所有启动子结合位点均在本公开的FVIII基因中消除。在所述优化序列中每一个潜在启动子结合位点的位置和相应核苷酸的序列显示于表2中。

[0243] 其他顺式作用负调控元件

[0244] 除了上文所述的MAR/ARS序列、去稳定元件和潜在启动子位点以外,还可在野生型BDD FVIII序列(SEQ ID NO:16)中鉴别出数个另外的潜在抑制序列。在所述非优化BDD FVIII序列中可鉴别出两个富AU序列元件(ARE)(ATTTTATT(SEQ ID NO:30);和ATTTTTAA(SEQ ID NO:31),连同poly-A位点(AAAAAA;SEQ ID NO:26)、poly-T位点(TTTTTT;SEQ ID NO:25)和剪接位点(GGTGAT;SEQ ID NO:27)。这些元件中的一者或多者可从所述优化FVIII序列中去除。在所述优化序列中这些位点中的每一者的位置和相应核苷酸的序列显示于表2中。

[0245] 在某些实施方案中,本公开提供包含如下核苷酸序列的分离的核酸分子,所述核苷酸序列包含编码FVIII多肽的N端部分的第一核苷酸序列和编码FVIII多肽的C端部分的第二核苷酸序列;其中所述第一核苷酸序列与(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具有至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性; 其中所述N端部分和所述C端部分合起来具有FVIII多肽活性; 并且其中所述核苷酸序列不含一种或多种顺式作用负调控元件, 例如剪接位点、poly-T序列、poly-A序列、ARE序列或其任何组合。

[0246] 在另一实施方案中, 本公开提供包含如下核苷酸序列的分离的核酸分子, 所述核苷酸序列包含编码FVIII多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列; 其中所述第二核酸序列与(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, SEQ ID NO:5的核苷酸1792-4374, 而无编码B结构域或B结构域片段的核苷酸); 或(iv) SEQ ID NO:6的核苷酸1792-2277和2320-4374(即, SEQ ID NO:6的核苷酸1792-4374, 而无编码B结构域或B结构域片段的核苷酸) 具有至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性; 其中所述N端部分和所述C端部分合起来具有FVIII多肽活性; 并且其中所述核苷酸序列不含一种或多种顺式作用负调控元件, 例如剪接位点、poly-T序列、poly-A序列、ARE序列或其任何组合。

[0247] 在其他实施方案中, 本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子, 其中所述核苷酸序列包含与(i) 选自SEQ ID NO:1、2、3、4、5、6、70和71的氨基酸序列的核苷酸58-4374或(ii) 选自SEQ ID NO:1、2、3、4、5、6、70和71的氨基酸序列的核苷酸58-2277和2320-4374(即, SEQ ID NO:1、2、3、4、5、6、70或71的核苷酸58-4374, 而无编码B结构域或B结构域片段的核苷酸) 具有至少约80%、至少约85%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列; 并且其中所述核苷酸序列不含一种或多种顺式作用负调控元件, 例如剪接位点、poly-T序列、poly-A序列、ARE序列或其任何组合。

[0248] 在一些实施方案中, 本公开提供包含如下核苷酸序列的分离的核酸分子, 所述核苷酸序列包含编码FVIII多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列; 其中所述第一核酸序列与(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具有至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性; 其中所述N端部分和所述C端部分合起来具有FVIII多肽活性; 并且其中所述核苷酸序列不含剪接位点GGTGAT (SEQ ID NO:27)。在一些实施方案中, 本公开提供包含如下核苷酸序列的分离的核酸分子, 所述核苷酸序列包含编码FVIII多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列; 其中所述第一核酸序列与(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具有至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少

约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性;其中所述N端部分和所述C端部分合起来具有FVIII多肽活性;并且其中所述核苷酸序列不含poly-T序列(SEQ ID NO:25)。在一些实施方案中,本公开提供包含如下核苷酸序列的分离的核酸分子,所述核苷酸序列包含编码FVIII多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列;其中所述第一核酸序列与(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具有至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性;其中所述N端部分和所述C端部分合起来具有FVIII多肽活性;并且其中所述核苷酸序列不含poly-A序列(SEQ ID NO:26)。在一些实施方案中,本公开提供包含如下核苷酸序列的分离的核酸分子,所述核苷酸序列包含编码FVIII多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列;其中所述第一核酸序列与(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具有至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性;其中所述N端部分和所述C端部分合起来具有FVIII多肽活性;并且其中所述核苷酸序列不含ARE元件(SEQ ID NO:30或SEQ ID NO:31)。

[0249] 在一些实施方案中,本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与(i)选自SEQ ID NO:1、2、3、4、5、6、70和71的氨基酸序列的核苷酸58-4374或(ii)选自SEQ ID NO:1、2、3、4、5、6、70和71的氨基酸序列的核苷酸58-2277和2320-4374(即,SEQ ID NO:1、2、3、4、5、6、70或71的核苷酸58-4374,而无编码B结构域或B结构域片段的核苷酸)具有至少约80%、至少约85%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列;并且其中所述核苷酸序列不含剪接位点GGTGAT(SEQ ID NO:27)。在一些实施方案中,本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与(i)选自SEQ ID NO:1、2、3、4、5、6、70和71的氨基酸序列的核苷酸58-4374或(ii)选自SEQ ID NO:1、2、3、4、5、6、70和71的氨基酸序列的核苷酸58-2277和2320-4374(即,SEQ ID NO:1、2、3、4、5、6、70或71的核苷酸58-4374,而无编码B结构域或B结构域片段的核苷酸)具有至少约80%、至少约85%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列;并且其中所述核苷酸序列不含poly-T序列(SEQ ID NO:25)。在一些实施方案中,本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与(i)选自SEQ ID NO:1、2、3、4、5、6、70和71的氨基酸序列的核苷酸58-4374或(ii)选自SEQ ID NO:1、2、3、4、5、6、70和71的氨基酸序列的核苷酸58-2277和2320-4374(即,SEQ ID NO:1、2、3、4、5、6、70或71的核苷酸58-4374,而无编码B结构域或B结构域片段的核苷酸)具有至少约80%、至少约85%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约

94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列;并且其中所述核苷酸序列不含poly-A序列(SEQ ID NO:26)。在一些实施方案中,本公开提供包含编码具有FVIII活性的多肽的核苷酸序列的分离的核酸分子,其中所述核苷酸序列包含与(i)选自SEQ ID NO:1、2、3、4、5、6、70和71的氨基酸序列的核苷酸58-4374或(ii)选自SEQ ID NO:1、2、3、4、5、6、70和71的氨基酸序列的核苷酸58-2277和2320-4374(即,SEQ ID NO:1、2、3、4、5、6、70或71的核苷酸58-4374,而无编码B结构域或B结构域片段的核苷酸)具有至少约80%、至少约85%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列;并且其中所述核苷酸序列不含ARE元件(SEQ ID NO:30或SEQ ID NO:31)。

[0250] 在其他实施方案中,本公开的优化的FVIII序列不包含抗病毒基序、茎-环结构和重复序列中的一者或多者。

[0251] 在其他实施方案中,将围绕转录起始位点的核苷酸改变为kozak共有序列(GCCGCCACCATGC(SEQ ID NO:32),其中加下划线的核苷酸是起始密码子)。在其他实施方案中,可添加或去除限制位点以促进克隆过程。

[0252] 异源核苷酸序列

[0253] 在一些实施方案中,本公开的分离的核酸分子进一步包含异源核苷酸序列。在一些实施方案中,本公开的分离的核酸分子进一步包含至少一个异源核苷酸序列。所述异源核苷酸序列可与本公开的优化的BDD-FVIII核苷酸序列在5'端、3'端连接,或插入至所述优化的BDD-FVIII核苷酸序列的中间。因此,在一些实施方案中,由所述异源核苷酸序列编码的异源氨基酸序列连接至由所述核苷酸序列编码的FVIII氨基酸序列的N端或C端或插入所述FVIII氨基酸序列中的两个氨基酸之间。在一些实施方案中,所述异源氨基酸序列可插入选自表3的一个或多个插入位点处的两个氨基酸之间。在一些实施方案中,所述异源氨基酸序列可插入由本公开的核酸分子编码的FVIII多肽内的在国际公布号WO 2013/123457 A1、WO 2015/106052 A1或美国公布号2015/0158929 A1中公开的任何位点处,所述公布以引用的方式整体并入本文中。

[0254] 在一些实施方案中,由所述异源核苷酸序列编码的异源氨基酸序列插入所述B结构域或其片段内。在一些实施方案中,所述异源氨基酸序列插入FVIII内紧邻对应于成熟人FVIII(SEQ ID NO:15)的氨基酸745的氨基酸的下游。在一个特定实施方案中,所述FVIII包含对应于成熟人FVIII(SEQ ID NO:15)的氨基酸746-1646的缺失,并且由所述异源核苷酸序列编码的异源氨基酸序列紧邻对应于成熟人FVIII(SEQ ID NO:15)的氨基酸745的下游插入。

[0255] 表3:异源部分插入位点

[0256]	插入位点	结构域	插入位点	结构域	插入位点	结构域
	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
[0257]	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 区	2188	C2
	333	A1	1711	A3	2227	C2
	336	A1	1720	A3	2332	CT
	339	A1	1725	A3		

[0258] 注：插入位点指示对应于成熟人FVIII (SEQ ID NO:15) 的氨基酸位置的氨基酸位置。

[0259] 在其他实施方案中，本公开的分离的核酸分子进一步包含两个、三个、四个、五个、六个、七个或八个异源核苷酸序列。在一些实施方案中，所有所述异源核苷酸序列是相同的。在一些实施方案中，至少一个异源核苷酸序列不同于其它异源核苷酸序列。在一些实施方案中，本公开可包含串联的两个、三个、四个、五个、六个或超过七个异源核苷酸序列。

[0260] 在一些实施方案中，所述异源核苷酸序列编码氨基酸序列。在一些实施方案中，由所述异源核苷酸序列编码的氨基酸序列是可增加FVIII分子的半衰期的异源部分（“半衰期延长剂”）。

[0261] 在一些实施方案中，所述异源部分是具有未结构化或结构化特征的肽或多肽，所述未结构化或结构化特征与并入本公开的蛋白质中时体内半衰期的延长有关。非限制性实例包括白蛋白、白蛋白片段、免疫球蛋白的Fc片段、人绒毛膜促性腺激素的β亚单位的C端肽(CTP)、HAP序列、XTEN序列、转铁蛋白或其片段、PAS多肽、聚甘氨酸接头、聚丝氨酸接头、白蛋白结合部分或这些多肽的任何片段、衍生物、变体或组合。在一个特定实施方案中，所述异源氨基酸序列是免疫球蛋白恒定区或其一部分、转铁蛋白、白蛋白或PAS序列。在一些方面，异源部分包括血管性血友病因子或其片段。在其他相关方面，异源部分可包括针对非多肽部分的附着位点（例如，半胱氨酸氨基酸），所述非多肽部分如聚乙二醇(PEG)、羟基乙基

淀粉 (HES)、聚唾液酸或这些元件的任何衍生物、变体或组合。在一些方面,异源部分包含充当针对非多肽部分的附着位点的半胱氨酸氨基酸,所述非多肽部分如聚乙二醇 (PEG)、羟基乙基淀粉 (HES)、聚唾液酸或这些元件的任何衍生物、变体或组合。

[0262] 在一个具体实施方案中,第一异源核苷酸序列编码作为本领域中已知的半衰期延长分子的第一异源部分,并且第二异源核苷酸序列编码也可以是本领域中已知的半衰期延长分子的第二异源部分。在某些实施方案中,所述第一异源部分(例如,第一Fc部分)和所述第二异源部分(例如,第二Fc部分)彼此缔合以形成二聚体。在一个实施方案中,所述第二异源部分是第二Fc部分,其中所述第二Fc部分连接至所述第一异源部分(例如,第一Fc部分)或与所述第一异源部分(例如,第一Fc部分)缔合。例如,所述第二异源部分(例如,第二Fc部分)可通过接头连接至所述第一异源部分(例如,第一Fc部分)或通过共价或非共价键与所述第一异源部分缔合。

[0263] 在一些实施方案中,所述异源部分是多肽,所述多肽包含至少约10个、至少约100个、至少约200个、至少约300个、至少约400个、至少约500个、至少约600个、至少约700个、至少约800个、至少约900个、至少约1000个、至少约1100个、至少约1200个、至少约1300个、至少约1400个、至少约1500个、至少约1600个、至少约1700个、至少约1800个、至少约1900个、至少约2000个、至少约2500个、至少约3000个或至少约4000个氨基酸,基本上由所述氨基酸组成,或由所述氨基酸组成。在其他实施方案中,所述异源部分是多肽,所述多肽包含约100个至约200个氨基酸、约200个至约300个氨基酸、约300个至约400个氨基酸、约400个至约500个氨基酸、约500个至约600个氨基酸、约600个至约700个氨基酸、约700个至约800个氨基酸、约800个至约900个氨基酸或约900个至约1000个氨基酸,基本上由所述氨基酸组成,或由所述氨基酸组成。

[0264] 在某些实施方案中,异源部分会改善所述FVIII蛋白的一种或多种药物动力学特性,而不会显著影响其生物活性或功能。

[0265] 在某些实施方案中,异源部分会增加本公开的FVIII蛋白的体内和/或体外半衰期。在其他实施方案中,异源部分会促进本公开的FVIII蛋白或其片段(例如,在所述FVIII蛋白的蛋白水解裂解之后包含异源部分的片段)的显现或定位。本公开的FVIII蛋白或其片段的显现和/或定位可以是体内、体外、离体或其组合。

[0266] 在其他实施方案中,异源部分会增加本公开的FVIII蛋白或其片段(例如,在所述FVIII蛋白的蛋白水解裂解之后包含异源部分的片段)的稳定性。如本文所用,术语“稳定性”是指所述FVIII蛋白的一种或多种物理特性响应于环境条件(例如,升高或降低的温度)的维持的公认量度。在某些方面,所述物理特性可以是所述FVIII蛋白的共价结构的维持(例如,蛋白水解裂解、不想要的氧化或去酰胺化的不存在)。在其他方面,所述物理特性也可以是呈正确折叠状态的FVIII蛋白的存在(例如,可溶性或不溶性聚集体或沉淀物的不存在)。一方面,所述FVIII蛋白的稳定性是通过测定所述FVIII蛋白的生物物理特性,例如热稳定性、pH去折叠概况、糖基化的稳定去除、溶解度、生物化学功能(例如,结合于蛋白质、受体或配体的能力)等和/或其组合来测量。另一方面,通过所述相互作用的结合亲和力来证明生物化学功能。一方面,蛋白质稳定性的量度是热稳定性,即对热激发(thermal challenge)的抗性。稳定性可以使用本领域中已知的方法来测量,如HPLC(高效液相色谱法)、SEC(尺寸排阻色谱法)、DLS(动态光散射)等。测量热稳定性的方法包括但不限于差示

扫描量热法 (DSC)、差示扫描荧光测定法 (DSF)、圆二色性 (CD) 和热激发测定。

[0267] 在某些方面,由本公开的核酸分子编码的FVIII蛋白包含至少一种半衰期延长剂,即异源部分,所述异源部分相对于缺乏所述异源部分的相应FVIII蛋白的体内半衰期增加所述FVIII蛋白的体内半衰期。FVIII蛋白的体内半衰期可以通过本领域技术人员已知的任何方法测量,例如活性测定(发色测定或一期凝固aPTT测定)、ELISA、ROTEM™等。

[0268] 在一些实施方案中,一种或多种半衰期延长剂的存在导致与缺乏所述一种或多种半衰期延长剂的相应蛋白质的半衰期相比所述FVIII蛋白的半衰期增加。包含半衰期延长剂的FVIII蛋白的半衰期比缺乏所述半衰期延长剂的相应FVIII蛋白的体内半衰期长至少约1.5倍、至少约2倍、至少约2.5倍、至少约3倍、至少约4倍、至少约5倍、至少约6倍、至少约7倍、至少约8倍、至少约9倍、至少约10倍、至少约11倍或至少约12倍。

[0269] 在一个实施方案中,包含半衰期延长剂的FVIII蛋白的半衰期比缺乏所述半衰期延长剂的相应蛋白的体内半衰期长约1.5倍至约20倍、约1.5倍至约15倍或约1.5倍至约10倍。在另一实施方案中,如与缺乏所述半衰期延长剂的相应蛋白的体内半衰期相比,包含半衰期延长剂的FVIII蛋白的半衰期延长约2倍至约10倍、约2倍至约9倍、约2倍至约8倍、约2倍至约7倍、约2倍至约6倍、约2倍至约5倍、约2倍至约4倍、约2倍至约3倍、约2.5倍至约10倍、约2.5倍至约9倍、约2.5倍至约8倍、约2.5倍至约7倍、约2.5倍至约6倍、约2.5倍至约5倍、约2.5倍至约4倍、约2.5倍至约3倍、约3倍至约10倍、约3倍至约9倍、约3倍至约8倍、约3倍至约7倍、约3倍至约6倍、约3倍至约5倍、约3倍至约4倍、约4倍至约6倍、约5倍至约7倍或约6倍至约8倍。

[0270] 在其他实施方案中,包含半衰期延长剂的FVIII蛋白的半衰期是至少约17小时、至少约18小时、至少约19小时、至少约20小时、至少约21小时、至少约22小时、至少约23小时、至少约24小时、至少约25小时、至少约26小时、至少约27小时、至少约28小时、至少约29小时、至少约30小时、至少约31小时、至少约32小时、至少约33小时、至少约34小时、至少约35小时、至少约36小时、至少约48小时、至少约60小时、至少约72小时、至少约84小时、至少约96小时或至少约108小时。

[0271] 在其他实施方案中,包含半衰期延长剂的FVIII蛋白的半衰期是约15小时至约两周、约16小时至约一周、约17小时至约一周、约18小时至约一周、约19小时至约一周、约20小时至约一周、约21小时至约一周、约22小时至约一周、约23小时至约一周、约24小时至约一周、约36小时至约一周、约48小时至约一周、约60小时至约一周、约24小时至约六天、约24小时至约五天、约24小时至约四天、约24小时至约三天或约24小时至约两天。

[0272] 在一些实施方案中,每名受试者中包含半衰期延长剂的FVIII蛋白的平均半衰期是约15小时、约16小时、约17小时、约18小时、约19小时、约20小时、约21小时、约22小时、约23小时、约24小时(1天)、约25小时、约26小时、约27小时、约28小时、约29小时、约30小时、约31小时、约32小时、约33小时、约34小时、约35小时、约36小时、约40小时、约44小时、约48小时(2天)、约54小时、约60小时、约72小时(3天)、约84小时、约96小时(4天)、约108小时、约120小时(5天)、约六天、约七天(一周)、约八天、约九天、约10天、约11天、约12天、约13天或约14天。

[0273] 一种或多种半衰期延长剂可以融合至FVIII的C端或N端或插入FVIII内。

[0274] 1. 免疫球蛋白恒定区或其部分

[0275] 另一方面,异源部分包括一个或多个免疫球蛋白恒定区或其部分(例如,Fc区)。在一个实施方案中,本公开的分离的核酸分子进一步包含编码免疫球蛋白恒定区或其部分的异源核酸序列。在一些实施方案中,所述免疫球蛋白恒定区或其部分是Fc区。

[0276] 免疫球蛋白恒定区包含指示CH(恒定重链)结构域(CH1、CH2等)的结构域。视同种型(即,IgG、IgM、IgA、IgD或IgE)而定,所述恒定区可以包含三个或四个CH结构域。一些同种型(例如,IgG)恒定区还含有铰链区。参见Janeway等人2001,Immunobiology, Garland Publishing, N.Y., N.Y.

[0277] 用于产生本公开的FVIII蛋白的免疫球蛋白恒定区或其部分可以从多种不同来源获得。在一个实施方案中,免疫球蛋白恒定区或其部分是源于人免疫球蛋白。然而,应理解免疫球蛋白恒定区或其部分可源于另一哺乳动物物种的免疫球蛋白,所述另一哺乳动物物种包括例如啮齿动物(例如,小鼠、大鼠、兔、豚鼠)或非人灵长类动物(例如,黑猩猩、猕猴)物种。此外,免疫球蛋白恒定区或其部分可源于任何免疫球蛋白类别(包括IgM、IgG、IgD、IgA和IgE)和任何免疫球蛋白同种型(包括IgG1、IgG2、IgG3和IgG4)。在一个实施方案中,使用人同种型IgG1。

[0278] 多种免疫球蛋白恒定区基因序列(例如,人恒定区基因序列)可以可公开存取的寄存物形式获得。可选择具有特定效应功能(或缺乏特定效应功能)或具有降低免疫原性的特定修饰的恒定区结构域序列。已经公布抗体和抗体编码基因的多个序列并且合适的Ig恒定区序列(例如,铰链、CH2和/或CH3序列,或其部分)可使用公认技术源于这些序列。使用任何前述方法获得的遗传材料可接着发生改变或进行合成以获得本公开的多肽。应进一步理解,本公开的范围涵盖恒定区DNA序列的等位基因、变体和突变。

[0279] 免疫球蛋白恒定区或其部分的序列可例如使用聚合酶链反应和选择用于扩增所关注的结构域的引物来克隆。为了克隆来自抗体的免疫球蛋白恒定区或其部分的序列,可从杂交瘤、脾或淋巴细胞分离mRNA,反转录至DNA中,并且通过PCR扩增抗体基因。PCR扩增方法详细地描述于美国专利号4,683,195;4,683,202;4,800,159;4,965,188;和例如“PCR Protocols: A Guide to Methods and Applications”Innis等人编, Academic Press, San Diego, CA (1990); Ho等人1989. Gene 77:51; Horton等人1993. Methods Enzymol. 217:270)中。PCR可基于公布的重链和轻链DNA和氨基酸序列通过共有恒定区引物或通过更特定引物来开始。PCR也可用于分离编码抗体轻链和重链的DNA克隆。在这种情况下,可通过共有引物或较大的同源探针,如小鼠恒定区探针来筛选文库。适用于抗体基因的扩增的多种引物集合是本领域中已知的(例如,基于纯化的抗体的N端序列(Benhar和Pastan. 1994. Protein Engineering 7:1509); cDNA末端的快速扩增(Ruberti, F.等人1994. J. Immunol. Methods 173:33); 抗体前导序列(Larrick等人1989 Biochem. Biophys. Res. Commun. 160:1250)的5'引物。抗体序列的克隆进一步描述于1995年1月25日提交的Newman等人的美国专利号5,658,570中,所述美国专利以引用的方式并入本文中。

[0280] 本文所用的免疫球蛋白恒定区可以包括所有结构域和铰链区或其部分。在一个实施方案中,免疫球蛋白恒定区或其部分包含CH2结构域、CH3结构域和铰链区(即,Fc区或FcRn结合配偶体)。

[0281] 如本文所用,术语“Fc区”是定义为多肽中对应于原生Ig的Fc区的部分,即,如通过其两条重链的相应Fc结构域的二聚缔合所形成。原生Fc区与另一Fc区形成同二聚体。相比

之下,如本文所用,术语“遗传融合的Fc区”或“单链Fc区”(scFc区)是指包含遗传连接于单一多肽链内(即,编码于单一邻接遗传序列中)的Fc结构域的合成二聚Fc区。参见国际公布号WO 2012/006635,其以引用的方式整体并入本文中。

[0282] 在一个实施方案中,“Fc区”是指单一Ig重链的部分,在铰链区中木瓜蛋白酶裂解位点的正好上游(即,IgG中的残基216,考虑重链恒定区的第一残基是114)开始并且在所述抗体的C端处终止。相应地,完全Fc区至少包含铰链域、CH2结构域和CH3结构域。

[0283] 免疫球蛋白恒定区或其部分可以是FcRn结合配偶体。FcRn在成体上皮组织中具活性并且表达于肠腔、肺气道、鼻表面、阴道表面、结肠和直肠表面中(美国专利号6,485,726)。FcRn结合配偶体是免疫球蛋白中结合于FcRn的部分。

[0284] FcRn受体已经从包括人在内的数种哺乳动物物种中分离。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受体以相对低pH结合IgG(而非其他免疫球蛋白类别,如IgA、IgM、IgD和IgE),沿腔至浆膜方向主动地跨细胞转运IgG,并且接着以细胞间隙液中发现的相对较高pH释放所述IgG。其表达于成体上皮组织中(美国专利号6,485,726、6,030,613、6,086,875;WO 03/077834;US2003-0235536A1),所述成体上皮组织包括肺和肠上皮(Israel等人1997,Immunology 92:69)肾近端肾小管上皮(Kobayashi等人2002,Am.J.Physiol.Renal Physiol.282:F358)以及鼻上皮、阴道表面和胆系表面。

[0285] 适用于本公开的FcRn结合配偶体涵盖可由FcRn受体特异性结合的分子,包括完整IgG、IgG的Fc片段和包括所述FcRn受体的完全结合区的其他片段。IgG的Fc部分中结合于FcRn受体的区已经基于X射线结晶学加以描述(Burmeister等人1994,Nature 372:379)。Fc与FcRn的主要接触区域是在CH2和CH3结构域的接合部附近。Fc-FcRn接触均在单一Ig重链内。FcRn结合配偶体包括完整IgG、IgG的Fc片段和包括FcRn的完全结合区的IgG的其他片段。主要接触位点包括CH2结构域的氨基酸残基248、250-257、272、285、288、290-291、308-311和314和CH3结构域的氨基酸残基385-387、428和433-436。对免疫球蛋白或免疫球蛋白片段或区的氨基酸编号的提及均是基于Kabat等人1991,Sequences of Proteins of Immunological Interest,U.S.Department of Public Health,Bethesda,Md。

[0286] 结合于FcRn的Fc区或FcRn结合配偶体可由FcRn有效地跨上皮屏障穿梭,因此提供非侵入手段来全身性地施用所需的治疗分子。另外,包含Fc区或FcRn结合配偶体的融合蛋白由表达FcRn的细胞内吞。并未经过标记用于降解,这些融合蛋白再次再循环进入循环中,因此增加这些蛋白质的体内半衰期。在某些实施方案中,免疫球蛋白恒定区的部分是Fc区或FcRn结合配偶体,所述Fc区或FcRn结合配偶体典型地经由二硫键和其他非特异性相互作用与另一Fc区或另一FcRn结合配偶体缔合以形成二聚体和高级多聚体。

[0287] 两种FcRn受体可以结合单一Fc分子。结晶学数据表明,每种FcRn分子结合Fc同二聚体的单一多肽。在一个实施方案中,使FcRn结合配偶体(例如,IgG的Fc片段)连接至生物活性分子会提供口服、经颊、舌下、经直肠、经阴道、以经鼻或经由肺途径或经由眼途径施用的气雾剂形式递送所述生物活性分子的手段。在另一实施方案中,FVIII蛋白可侵入性,例如皮下、静脉内施用。

[0288] FcRn结合配偶体区是可由FcRn受体特异性结合的分子或其部分,随后由所述Fc区的FcRn受体主动转运。特异性结合是指两种分子形成在生理学条件下相对稳定的复合物。

特异性结合的特征在于高亲和力和低至中等能力,如与通常具有低亲和力和中等至高能力的非特异性结合相区分。典型地,当亲和力常数 KA 高于 $10^6 M^{-1}$ 或高于 $10^8 M^{-1}$ 时,结合被视为特异性的。必要时,非特异性结合可以降低,而不会通过改变结合条件显著影响特异性结合。熟练技术人员可以使用常规技术优化适当结合条件,如分子的浓度、溶液的离子强度、温度、允许用于结合的时间、阻断剂(例如,血清白蛋白、乳酪蛋白)的浓度等。

[0289] 在某些实施方案中,由本公开的核酸分子编码的FVIII蛋白包含一个或多个截短的Fc区,所述截短的Fc区仍足以向所述Fc区赋予Fc受体(FcR)结合特性。例如,Fc区中结合于FcRn的部分(即,FcRn结合部分)包含IgG1EU编号的约氨基酸282-438(其中主要接触位点是CH2结构域的氨基酸248、250-257、272、285、288、290-291、308-311和314和CH3结构域的氨基酸残基385-387、428和433-436)。因此,本公开的Fc区可包含FcRn结合部分或由FcRn结合部分组成。FcRn结合部分可源于任何同种型(包括IgG1、IgG2、IgG3和IgG4)的重链。在一个实施方案中,使用来自人同种型IgG1的抗体的FcRn结合部分。在另一实施方案中,使用来自人同种型IgG4的抗体的FcRn结合部分。

[0290] 所述Fc区可获自多种不同来源。在一个实施方案中,所述多肽的Fc区是源于人免疫球蛋白。然而,应理解Fc部分可源于另一哺乳动物物种的免疫球蛋白,所述另一哺乳动物物种包括例如啮齿动物(例如,小鼠、大鼠、兔、豚鼠)或非人灵长类动物(例如,黑猩猩、猕猴)物种。此外,所述Fc结构域的多肽或其部分可源于任何免疫球蛋白类别(包括IgM、IgG、IgD、IgA和IgE)和任何免疫球蛋白同种型(包括IgG1、IgG2、IgG3和IgG4)。在另一实施方案中,使用人同种型IgG1。

[0291] 在某些实施方案中,所述Fc变体赋予由包含所述野生型Fc结构域的Fc部分赋予的至少一种效应功能的改变(例如,所述Fc区结合于Fc受体(例如,Fc γ RI、Fc γ RII或Fc γ RIII)或补体蛋白(例如,C1q)或触发抗体依赖性细胞毒性(ADCC)、吞噬或补体依赖性细胞毒性(CDCC)的能力的改进或降低)。在其他实施方案中,所述Fc变体提供经过工程改造的半胱氨酸残基。

[0292] 本公开的Fc区可使用公认Fc变体,已知所述公认Fc变体赋予效应功能和/或FcR或FcRn结合的改变(例如,增强或降低)。特别地,本公开的Fc区可包括例如在国际PCT公布W088/07089A1、W096/14339A1、W098/05787A1、W098/23289A1、W099/51642A1、W099/58572A1、W000/09560A2、W000/32767A1、W000/42072A2、W002/44215A2、W002/060919A2、W003/074569A2、W004/016750A2、W004/029207A2、W004/035752A2、W004/063351A2、W004/074455A2、W004/099249A2、W005/040217A2、W004/044859、W005/070963A1、W005/077981A2、W005/092925A2、W005/123780A2、W006/019447A1、W006/047350A2和W006/085967A2;美国专利公布号US2007/0231329、US2007/0231329、US2007/0237765、US2007/0237766、US2007/0237767、US2007/0243188、US20070248603、US20070286859、US20080057056;或美国专利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中公开的氨基酸位置中的一处或多处的改变(例如,取代),其中每一者均以引用的方式并入本文中。在一个实施方案中,特定改变(例如,所属领域中公开的一种或多种氨基酸的特定取代)可在所公开的氨基酸位置中的一处或多处进行。在另一实施方案中,可在所公开的氨基酸位置中的一处或多处进行不同改变(例如,所属领域中公开的

一个或多个氨基酸位置的不同取代)。

[0293] IgG的Fc区或FcRn结合配偶体可根据公认的程序(如定点诱变等)进行修饰以产生将由FcRn结合的经过修饰的IgG或Fc片段或其部分。所述修饰包括远离FcRn接触位点的修饰以及在所述接触位点内的修饰,所述修饰保持或甚至增强与FcRn的结合。例如,人IgG1Fc(Fc γ 1)中的以下单一氨基酸残基可经过取代而无Fc对FcRn的结合亲和力的显著损失: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表示在位置编号238处的野生型脯氨酸由丙氨酸取代。例如,一个特定实施方案并入N297A突变,从而去除了高度保守的N-糖基化位点。除了丙氨酸以外,其他氨基酸也可以在上文规定的位置处取代野生型氨基酸。突变可单独引入至Fc中,从而产生超过一百个不同于原生Fc的Fc区。另外,可合起来引入这些个别突变中的两者、三者或更多者的组合,从而产生数百个Fc区。

[0294] 上述突变中的某些可对Fc区或FcRn结合配偶体赋予新的功能性。例如,一个实施方案并入N297A,从而去除了高度保守的N-糖基化位点。这种突变的作用是降低免疫原性,由此增强Fc区的循环半衰期,和使Fc区不能结合于Fc γ RI、Fc γ RIIA、Fc γ RIIB和Fc γ RIIIA,而不会损害对FcRn的亲和力(Routledge等人1995,Transplantation 60:847; Friend等人1999,Transplantation 68:1632; Shields等人1995, J. Biol. Chem. 276:6591)。作为由上述突变产生的新的功能性的另一实例,对FcRn的亲和力可在一些情况下增加超出野生型的所述亲和力。这种增加的亲和力可反映增加的“缔合”速率、减少的“解离”速率或增加的“缔合”速率和减少的“解离”速率两者。相信赋予增加的对FcRn的亲和力的突变的实例包括但不限于T256A、T307A、E380A和N434A(Shields等人2001, J. Biol. Chem. 276:6591)。

[0295] 另外,至少三种人Fc γ 受体看来识别IgG中在下部铰链区内的结合位点(一般是氨基酸234-237)。因此,新的功能性和潜在减少的免疫原性的另一实例可由这一区的突变引起,如例如通过将人IgG1“ELLG”(SEQ ID NO:45)的氨基酸233-236置换成来自IgG2“PVA”的相应序列(具有一个氨基酸缺失)。已经显示,当已经引入所述突变时,介导多种效应功能的Fc γ RI、Fc γ RII和Fc γ RIII将不会结合于IgG1。Ward和Ghetie 1995, Therapeutic Immunology 2:77和Armour等人1999, Eur. J. Immunol. 29:2613。

[0296] 在另一实施方案中,免疫球蛋白恒定区或其部分包含在铰链区中的氨基酸序列或其部分,所述氨基酸序列或其部分与第二免疫球蛋白恒定区或其部分形成一个或多个二硫键。所述第二免疫球蛋白恒定区或其部分可连接至第二多肽,从而使FVIII蛋白和所述第二多肽合在一起。在一些实施方案中,所述第二多肽是增强子部分。如本文所用,术语“增强子部分”是指能够增强FVIII的促凝血活性的分子、其片段或多肽的组分。所述增强子部分可以是辅因子,如可溶性组织因子(sTF)或促凝血肽。因此,在活化FVIII时,所述增强子部分

可用于增强FVIII活性。

[0297] 在某些实施方案中,由本公开的核酸分子编码的FVIII蛋白包含免疫球蛋白恒定区或其部分的氨基酸取代(例如,Fc变体),所述氨基酸取代会改变Ig恒定区的抗原独立效应功能,尤其是所述蛋白质的循环半衰期。

[0298] 2.scFc区

[0299] 另一方面,异源部分包括scFc(单链Fc)区。在一个实施方案中,本公开的分离的核酸分子进一步包含编码scFc区的异源核酸序列。所述scFc区包含在同一直链多肽链内的至少两个免疫球蛋白恒定区或其部分(例如,Fc部分或结构域(例如,2、3、4、5、6个或更多个Fc部分或结构域)),所述免疫球蛋白恒定区或其部分能够折叠(例如,分子内或分子间折叠)以形成由Fc肽接头连接的一个功能性scFc区。例如,在一个实施方案中,本公开的多肽能够经由其scFc区结合至至少一种Fc受体(例如,FcRn、Fc γ R受体(例如,Fc γ RIII)或补体蛋白(例如,C1q))以便改进半衰期或触发免疫效应功能(例如,抗体依赖性细胞毒性(ADCC)、吞噬或补体依赖性细胞毒性(CDCC)和/或改进可制造性)。

[0300] 3.CTP

[0301] 另一方面,异源部分包括人绒毛膜促性腺激素的 β 亚单位的一种C端肽(CTP)或其片段、变体或衍生物。已知插入至重组蛋白中的一种或多种CTP肽增加所述蛋白质的体内半衰期。参见例如美国专利号5,712,122,以引用的方式整体并入本文中。

[0302] 示例性CTP肽包括DPRFQDSSSSKAPPSLPSPSRLPGPSDTPIL(SEQ ID NO:33)或SSSSKAPPSLPSPSRLPGPSDTPILPQ(SEQ ID NO:34)。参见例如美国专利申请公布号US 2009/0087411A1,以引用的方式并入。

[0303] 4.XTEN序列

[0304] 在一些实施方案中,异源部分包括一个或多个XTEN序列、其片段、变体或衍生物。如此处所用,“XTEN序列”是指具有非天然存在、大体上非重复性序列的延长长度的多肽,所述序列主要由小的亲水性氨基酸构成,其中所述序列在生理学条件下具有低度或不具有二级或三级结构。作为异源部分,XTEN可充当半衰期延长部分。另外,XTEN可提供所需特性,包括但不限于增强的药物动力学参数和溶解度特征。

[0305] 将包含XTEN序列的异源部分并入至本公开的蛋白质中可以向所述蛋白质赋予以下有利特性中的一者或多者:构象柔性、增强的水溶性、高度蛋白酶抗性、低免疫原性、与哺乳动物受体的低结合或增加的流体动力学(或斯托克斯)半径。

[0306] 在某些方面,XTEN序列可以增加药物动力学特性,如较长的体内半衰期或增加的曲线下面积(AUC),使得与不具有XTEN异源部分的相同蛋白质相比,本公开的蛋白质在一段增加的时期内停留于体内并且具有促凝血活性。

[0307] 在一些实施方案中,适用于本公开的XTEN序列是具有超过约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个氨基酸残基的肽或多肽。在某些实施方案中,XTEN是具有超过约20个至约3000个氨基酸残基、超过30个至约2500个残基、超过40个至约2000个残基、超过50个至约1500个残基、超过60个至约1000个残基、超过70个至约900个残基、超过80个至约800个残基、超过90个至约700个残基、超过100个至约600个残基、超过110个至约500个残基或超过120个至约400个残基的肽或多肽。在一个特定实施方案中,所述

XTEN包含长于42个氨基酸并且短于144个氨基酸的长度的氨基酸序列。

[0308] 本公开的XTEN序列可以包含5至14个(例如,9至14个)氨基酸残基的一个或多个序列基序或与所述序列基序具有至少80%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性的氨基酸序列,其中所述基序包含选自由甘氨酸(G)、丙氨酸(A)、丝氨酸(S)、苏氨酸(T)、谷氨酸(E)和脯氨酸(P)组成的组的4至6种类型的氨基酸(例如,5种氨基酸),基本上由所述氨基酸组成,或由所述氨基酸组成。参见US 2010-0239554 A1。

[0309] 在一些实施方案中,所述XTEN包含非重叠序列基序,其中所述序列的约80%或至少约85%或至少约90%或约91%或约92%或约93%或约94%或约95%或约96%或约97%或约98%或约99%或约100%由选自从表4选择的单一基序家族的非重叠序列的多个单位组成,从而产生家族序列。如本文所用,“家族”意指所述XTEN具有仅选自来自表4的单一基序类别的基序;即,AD、AE、AF、AG、AM、AQ、BC或BD XTEN,并且选择所述XTEN中不是来自家族基序的任何其他氨基酸以实现所需的特性,如允许通过编码核苷酸并入限制位点、并入裂解序列或实现与FVIII的更佳连接。在XTEN家族的一些实施方案中,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家族的非重叠序列基序的多个单位,其中所得XTEN展现上述同源性的范围。在其他实施方案中,所述XTEN包含来自表2A的两个或更多个基序家族的基序序列的多个单位。可选择这些序列以实现所需物理/化学特征,包括如下文更充分描述的净电荷、疏水性、缺乏二级结构或缺乏由所述基序的氨基酸组成赋予的重复性等特性。在上文中描述于这一段落中的实施方案中,并入至XTEN中的基序可使用本文所述的方法来选择并且组装以实现约36个至约3000个氨基酸残基的XTEN。

[0310] 表4. 12个氨基酸的XTEN序列基序和基序家族

[0311]

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
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
BD	GSTAGSETSTEA	99
BD	GSETATSGSETA	100
BD	GTSESATSESGA	101
BD	GTSTEASEGSAS	102

[0312] *表示当合起来用于多种排列中时产生“家族序列”的个别基序序列

[0313] 可用作本公开的嵌合蛋白中的异源部分的XTEN序列的实例公开于例如美国专利公布号2010/0239554 A1、2010/0323956 A1、2011/0046060 A1、2011/0046061 A1、2011/0077199 A1或2011/0172146 A1或国际专利公布号WO 2010091122 A1、WO 2010144502 A2、WO 2010144508 A1、WO 2011028228 A1、WO 2011028229 A1或WO 2011028344 A2中,其中每一者均以引用的方式整体并入本文中。

[0314] XTEN可具有变化的长度以用于插入至或连接至FVIII。在一个实施方案中,XTEN序列的长度是基于所述融合蛋白中欲实现的特性或功能来选择。视预期的特性或功能而定,XTEN可以是短或中等长度序列或较长序列,所述序列可以用作载体。在某些实施方案中,所述XTEN包括约6个至约99个氨基酸残基的短区段、约100个至约399个氨基酸残基的中等长度和约400个至约1000个并且长达约3000个氨基酸残基的较长长度。因此,插入至或连接至FVIII的XTEN可以具有约6个、约12个、约36个、约40个、约42个、约72个、约96个、约144个、约288个、约400个、约500个、约576个、约600个、约700个、约800个、约864个、约900个、约1000个、约1500个、约2000个、约2500个或长达约3000个氨基酸残基长度的长度。在其他实施方案中,所述XTEN序列是约6个至约50个、约50个至约100个、约100个至150个、约150个至250个、约250个至400个、约400个至约500个、约500个至约900个、约900个至1500个、约1500个至2000或约2000个至约3000个氨基酸残基长度。插入至或连接至FVIII的XTEN的精确长度可以变化,而不会不利地影响所述FVIII的活性。在一个实施方案中,一种或多种本文所用的XTEN具有42个氨基酸、72个氨基酸、144个氨基酸、288个氨基酸、576个氨基酸或864个氨基酸长度并且可以选自一个或多个XTENT家族序列;即AD、AE、AF、AG、AM、AQ、BC或BD。

[0315] 在一些实施方案中,本公开中所用的XTEN序列与选自由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和其任何组合组成的组的序列具有至少60%、70%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性。参见US 2010-0239554 A1。在一个特定实施方案中,所述XTEN包含AE42、AE72、AE144、AE288、AE576、AE864、AG 42、AG72、AG144、AG288、AG576、AG864或其任何组合。

[0316] 可用作本公开的嵌合蛋白中的异源部分的示例性XTEN序列包括XTEN AE42-4 (SEQ ID NO:46,由SEQ ID NO:47编码;分别是图11C和图11D)、XTEN 144-2A (SEQ ID NO:48,由SEQ ID NO:49编码;分别是图11E和图11F)、XTEN A144-3B (SEQ ID NO:50,由SEQ ID NO:51编码;分别是图11G和图11H)、XTEN AE144-4A (SEQ ID NO:52,由SEQ ID NO:53编码;分别是图11I和图11J)、XTEN AE144-5A (SEQ ID NO:54,由SEQ ID NO:55编码;分别是图11K和图11L)、XTEN AE144-6B (SEQ ID NO:56,由SEQ ID NO:57编码;分别是图11M和图11N)、XTEN AG144-1 (SEQ ID NO:58,由SEQ ID NO:59编码;分别是图11O和图11P)、XTEN AG144-A (SEQ ID NO:60,由SEQ ID NO:61编码;分别是图11Q和图11R)、XTEN AG144-B (SEQ ID NO:62,由SEQ ID NO:63编码;分别是图11S和图11T)、XTEN AG144-C (SEQ ID NO:64,由SEQ ID NO:65编码;分别是图11U和图11V)和XTEN AG144-F (SEQ ID NO:66,由SEQ ID NO:67编码;分别是图11W和图11X)。在一个特定实施方案中,所述XTEN是由SEQ ID NO:18编码。

[0317] 在一些实施方案中,XTEN的氨基酸的少于100%选自甘氨酸(G)、丙氨酸(A)、丝氨酸(S)、苏氨酸(T)、谷氨酸(E)和脯氨酸(P),或所述序列的少于100%由来自表2A的序列基序或本文所提供的XTEN序列组成。在所述实施方案中,所述XTEN的剩余氨基酸残基选自任何其他14种天然L-氨基酸,但可以优先地选自亲水性氨基酸,使得所述XTEN序列含有至少约90%、91%、92%、93%、94%、95%、96%、97%、98%或至少约99%亲水性氨基酸。用于缀合构建体中的XTEN的疏水性氨基酸的含量可以是少于5%或少于2%或少于1%疏水性氨基酸含量。在XTEN的构建中不受青睐的疏水性残基包括色氨酸、苯丙氨酸、酪氨酸、亮氨酸、异亮氨酸、缬氨酸和甲硫氨酸。另外,XTEN序列可以含有少于5%或少于4%或少于3%或少于2%或少于1%或不含有以下氨基酸:甲硫氨酸(例如,以避免氧化)或天冬酰胺和谷氨酰胺(以避免去酰胺化)。

[0318] 所述一个或多个XTEN序列可插入由所述核苷酸序列编码的氨基酸序列的C端或N端或插入由所述核苷酸序列编码的氨基酸序列中的两个氨基酸之间。例如,所述XTEN可插入选自表3的一个或多个插入位点处的两个氨基酸之间。FVIII内可允许XTEN插入的位点的实例可发现于例如国际公布号W0 2013/123457 A1或美国公布号2015/0158929 A1中,所述公布以引用的方式整体并入本文中。

[0319] 5. 白蛋白或其片段、衍生物或变体

[0320] 在一些实施方案中,异源部分包括白蛋白或其功能片段。人血清白蛋白(HSA或HA)是其全长形式具有609个氨基酸的蛋白质,负责血清的大部分的渗透压并且还充当内源和外源配体的载体。如本文所用,术语“白蛋白”包括全长白蛋白或其功能片段、变体、衍生物或类似物。白蛋白或其片段或变体的实例公开于美国专利公布号2008/0194481A1、2008/0004206 A1、2008/0161243 A1、2008/0261877 A1或2008/0153751 A1或PCT申请公布号2008/033413 A2、2009/058322 A1或2007/021494 A2中,所述公布以引用的方式整体并入

本文中。

[0321] 在一个实施方案中,由本公开的核酸分子编码的FVIII蛋白包含白蛋白、其片段或变体,所述白蛋白、其片段或变体进一步连接至选自由免疫球蛋白恒定区或其部分(例如,Fc区)、PAS序列、HES和PEG组成的组的第二异源部分。

[0322] 6.白蛋白结合部分

[0323] 在某些实施方案中,所述异源部分是白蛋白结合部分,其包含白蛋白结合肽、细菌白蛋白质结合结构域、白蛋白结合抗体片段或其任何组合。

[0324] 例如,白蛋白结合蛋白可以是细菌白蛋白结合蛋白、抗体或包括结构域抗体的抗体片段(参见美国专利号6,696,245)。例如,白蛋白结合蛋白可以是细菌白蛋白质结合结构域,如链球菌蛋白G的细菌白蛋白质结合结构域(Konig, T. 和Skerra, A. (1998) J. Immunol. Methods 218, 73-83)。可用作缀合配偶体的白蛋白结合肽的其他实例是例如具有Cys-Xaa₁-Xaa₂-Xaa₃-Xaa₄-Cys共有序列的那些,其中Xaa₁是Asp、Asn、Ser、Thr或Trp;Xaa₂是Asn、Gln、His、Ile、Leu或Lys;Xaa₃是Ala、Asp、Phe、Trp或Tyr;并且Xaa₄是Asp、Gly、Leu、Phe、Ser或Thr,如美国专利申请2003/0069395或Dennis等人(Dennis等人(2002) J. Biol. Chem. 277, 35035-35043)所述。

[0325] 如由Kraulis等人,FEBS Lett. 378:190-194(1996)和Linhult等人,Protein Sci. 11:206-213(2002)公开的来自链球菌蛋白G的结构域3是细菌白蛋白质结合结构域的一个实例。白蛋白结合肽的实例包括具有核心序列DICLPRWGCLW(SEQ ID NO:35)的肽的系列。参见例如Dennis等人,J. Biol. Chem. 2002, 277:35035-35043(2002)。白蛋白结合抗体片段的实例公开于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)中,所述文献以引用的方式整体并入本文中。所述白蛋白结合部分的一个实例是如由Trussel等人, Bioconjugate Chem. 20:2286-2292(2009)公开的2-(3-顺丁烯二酰亚胺基丙酰胺基)-6-(4-(4-碘苯基)丁酰胺基)己酸酯(“Albu”标签)。

[0326] 脂肪酸、尤其长链脂肪酸(LCFA)和长链脂肪酸样白蛋白结合化合物可用于延长本公开的FVIII蛋白的体内半衰期。LCFA样白蛋白结合化合物的一个实例是16-(1-(3-(9-((2,5-二氧化吡咯烷-1-基氧基)羰基氧基)-甲基)-7-磺基-9H-芴-2-基氨基)-3-氧代丙基)-2,5-二氧化吡咯烷-3-基硫基)十六烷酸(参见例如WO 2010/140148)。

[0327] 7. PAS序列

[0328] 在其他实施方案中,所述异源部分是PAS序列。如本文所用,PAS序列意指主要包含丙氨酸和丝氨酸残基或主要包含丙氨酸、丝氨酸和脯氨酸残基的氨基酸序列,所述氨基酸序列在生理学条件下形成无规卷曲构象。因此,所述PAS序列是包含丙氨酸、丝氨酸和脯氨酸,基本上由丙氨酸、丝氨酸和脯氨酸组成,或由丙氨酸、丝氨酸和脯氨酸组成的结构单元、氨基酸聚合物或序列盒,其可用作所述嵌合蛋白中的异源部分的一部分。然而,熟练人员应知晓,当除了丙氨酸、丝氨酸和脯氨酸以外的残基作为所述PAS序列中的次要成分添加时,氨基酸聚合物也可以形成无规卷曲构象。如本文所用,术语“次要成分”意指除了丙氨酸、丝氨酸和脯氨酸以外的氨基酸可添加于所述PAS序列中至某一程度,例如多达约12%(即,所述PAS序列的100个氨基酸中的约12个)、多达约10%(即,所述PAS序列的100个氨基酸中的约10个)、多达约9%(即,100个氨基酸中的约9个)、多达约8%(即,100个氨基酸中的约8

个)、约6% (即,100个氨基酸中的约6个)、约5% (即,100个氨基酸中的约5个)、约4% (即,100个氨基酸中的约4个)、约3% (即,100个氨基酸中的约3个)、约2% (即,100个氨基酸中的约2个)、约1% (即,100个氨基酸中的约1个)。不同于丙氨酸、丝氨酸和脯氨酸的氨基酸可选自由Arg、Asn、Asp、Cys、Gln、Glu、Gly、His、Ile、Leu、Lys、Met、Phe、Thr、Trp、Tyr和Val组成的组。

[0329] 在生理学条件下,所述PAS序列区段形成无规卷曲构象并且由此可以介导所述FVIII蛋白的增加的体内和/或体外稳定性。由于所述无规卷曲域本身不采取稳定结构或功能,由所述FVIII蛋白介导的生物活性基本上得以保持。在其他实施方案中,形成无规卷曲结构域的PAS序列是生物学惰性的,尤其相对于血浆中的蛋白水解、免疫原性、等电点/静电行为、结合于细胞表面受体或内化,但仍然是生物可降解的,其提供了优于如PEG等合成聚合物的明显优势。

[0330] 形成无规卷曲构象的PAS序列的非限制性实例包括选自由以下组成的组的氨基酸序列:ASPAAPAPASPAAPSAPA (SEQ ID NO:36)、AAPASPAPAAPSAPAPAAPS (SEQ ID NO:37)、APSSPSPS APSSPSPASPSS (SEQ ID NO:38)、APSSPSPSAPSSPSPASPS (SEQ ID NO:39)、SSPSAPSPSPASPSPSSPA (SEQ ID NO:40)、AA SPAAPSAPPAAASPAAPSAPPA (SEQ ID NO:41)和ASAAAPAAA SAAASAPSAAA (SEQ ID NO:42)或其任何组合。PAS序列的另外的实例由例如美国专利公布号2010/0292130A1和PCT申请公布号WO 2008/155134 A1已知。

[0331] 8.HAP序列

[0332] 在某些实施方案中,所述异源部分富甘氨酸的同型氨基酸聚合物(HAP)。所述HAP序列可包括甘氨酸的重复性序列,所述重复性序列具有至少50个氨基酸、至少100个氨基酸、120个氨基酸、140个氨基酸、160个氨基酸、180个氨基酸、200个氨基酸、250个氨基酸、300个氨基酸、350个氨基酸、400个氨基酸、450个氨基酸或500个氨基酸的长度。在一个实施方案中,所述HAP序列能够延长融合至或连接至所述HAP序列的部分的半衰期。所述HAP序列的非限制性实例包括但不限于 $(\text{Gly})_n$ 、 $(\text{Gly}_4\text{Ser})_n$ 或 $\text{S}(\text{Gly}_4\text{Ser})_n$,其中n为1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20。在一个实施方案中,n为20、21、22、23、24、25、26、26、28、29、30、31、32、33、34、35、36、37、38、39或40。在另一实施方案中,n为50、60、70、80、90、100、110、120、130、140、150、160、170、180、190或200。

[0333] 9.转铁蛋白或其片段

[0334] 在某些实施方案中,所述异源部分是转铁蛋白或其片段。任何转铁蛋白均可用于制造本公开的FVIII蛋白。例如,野生型人TF (TF) 是大约75kDa (未考虑糖基化) 的679个氨基酸的蛋白质,具有两个主要的结构域N (约330个氨基酸) 和C (约340个氨基酸),所述结构域看来源于基因复制。参见GenBank登录号NM001063、XM002793、M12530、XM039845、XM 039847和S95936 (www.ncbi.nlm.nih.gov/),其均以引用的方式整体并入本文中。转铁蛋白包含两个结构域,N结构域和C结构域。N结构域包含两个子结构域N1结构域和N2结构域,并且C结构域包含两个子结构域C1结构域和C2结构域。

[0335] 在一个实施方案中,转铁蛋白异源部分包括转铁蛋白剪接变体。在一个实施例中,转铁蛋白剪接变体可以是人转铁蛋白的剪接变体,例如Genbank登陆AAA61140。在另一实施方案中,所述嵌合蛋白的转铁蛋白部分包括转铁蛋白序列的一个或多个结构域,例如N结构域、C结构域、N1结构域、N2结构域、C1结构域、C2结构域或其任何组合。

[0336] 10. 清除受体

[0337] 在某些实施方案中,所述异源部分是清除受体、其片段、变体或衍生物。LRP1是600kDa完整膜蛋白,其牵涉于多种蛋白质(如因子X)的受体介导清除中。参见例如Narita等人,Blood 91:555-560(1998)。

[0338] 11. 血管性血友病因子或其片段

[0339] 在某些实施方案中,所述异源部分是血管性血友病因子(VWF)或其一个或多个片段。

[0340] VWF(还称作F8VWF)是存在于血浆中并且在内皮(在Weibel-Palade小体中)、巨核细胞(血小板的 α -颗粒)和内皮下结缔组织中组成性产生的大的多聚体糖蛋白。基础VWF单体是2813个氨基酸的蛋白质。每一种单体均含有多个具有特定功能的特定结构域,即D'和D3结构域(其合起来结合于因子VIII)、A1结构域(其结合于血小板GPIb-受体、肝素和/或可能结合于胶原蛋白)、A3结构域(其结合于胶原蛋白)、C1结构域(其中当这个结构域被活化时,RGD结构域结合于血小板整合素 α IIb β 3)和在所述蛋白质的C末端处的“半胱氨酸结”结构域(VWF与血小板源性生长因子(PDGF)、转化生长因子- β (TGF β)和 β -人绒毛膜促性腺激素(BHCG)共享所述结构域)。

[0341] 人VWF的2813单体氨基酸序列在Genbank中以登录号NP000543.2报告。编码人VWF的核苷酸序列在Genbank中以登录号NM000552.3报告。SEQ ID NO:44(图11B)是由SEQ ID NO:43编码的氨基酸序列。所述D'结构域包括SEQ ID NO:44的氨基酸764至866。所述D3结构域包括SEQ ID NO:44的氨基酸867至1240。

[0342] 在血浆中,95-98%的FVIII以与全长VWF的紧密非共价复合物形式循环。这种复合物的形成对于体内FVIII的适当血浆水平的维持至关重要。Lenting等人,Blood.92(11):3983-96(1998);Lenting等人,J.Thromb.Haemost.5(7):1353-60(2007)。当FVIII由于在重链中的位置372和740处和在轻链中的位置1689处的蛋白水解而活化时,结合于FVIII的VWF从活化的FVIII去除。

[0343] 在某些实施方案中,所述异源部分是全长血管性血友病因子。在其他实施方案中,所述异源部分是血管性血友病因子片段。如本文所用,本文所用的术语“VWF片段”意指与FVIII相互作用并且保持通常由全长VWF提供给FVIII的至少一种或多种特性的任何VWF片段,所述至少一种或多种特性例如防止过早活化为FVIIIa、防止过早蛋白水解、防止可能导致过早清除的与磷脂膜的缔合、防止结合于可结合裸FVIII而非VWF结合的FVIII的FVIII清除受体和/或使FVIII重链和轻链相互作用稳定化。在一个特定实施方案中,所述异源部分是包含VWF的D'结构域和D3结构域的(VWF)片段。包含D'结构域和D3结构域的VWF片段可进一步包含选自A1结构域、A2结构域、A3结构域、D1结构域、D2结构域、D4结构域、B1结构域、B2结构域、B3结构域、C1结构域、C2结构域、CK结构域、其一个或多个片段和其任何组合组成的组的VWF结构域。融合至VWF片段的具有FVIII活性的多肽的另外的实例公开于2012年7月3日提交的美国临时专利申请号61/667,901和美国公布号2015/0023959A1中,所述文献均以引用的方式整体并入本文中。

[0344] 12. 接头部分

[0345] 在某些实施方案中,所述异源部分是肽接头。

[0346] 如本文所用,术语“肽接头”或“接头部分”是指在多肽链的线性氨基酸序列中连接

两个结构域的肽或多肽序列(例如,合成肽或多肽序列)。

[0347] 在一些实施方案中,编码肽接头的异源核苷酸序列可插入本公开的优化的FVIII多核苷酸序列与编码例如上述异源部分之一(如白蛋白)的异源核苷酸序列之间。肽接头可向嵌合多肽分子提供柔性。接头并非典型地裂解,然而所述裂解可以是所需的。在一个实施方案中,这些接头在加工期间未去除。

[0348] 可存在于本公开的嵌合蛋白中的一种类型的接头是蛋白酶可裂解接头,其包含裂解位点(即蛋白酶裂解位点底物,例如因子XIa、Xa或凝血酶裂解位点)并且其可在所述裂解位点的N端或C端或两侧包括另外的接头。这些可裂解接头当并入至本公开的构建体中时产生具有异源裂解位点的嵌合分子。

[0349] 在一个实施方案中,由本公开的核酸分子编码的FVIII多肽包含经由cscFc接头连接的两个或更多个Fc结构域或部分以形成包含于单一多肽链中的Fc区。所述cscFc接头侧接至少一个细胞内加工位点,即通过细胞内酶裂解的位点。所述多肽在所述至少一个细胞内加工位点处的裂解产生包含至少两个多肽链的多肽。

[0350] 其他肽接头可任选地用于本公开的构建体中,例如以连接FVIII蛋白至Fc区。可结合本公开使用的一些示例性连接体包括例如下文更详细描述地包含GlySer氨基酸的多肽。

[0351] 在一个实施方案中,所述肽接头是合成的,即非天然存在的。在一个实施方案中,肽接头包括包含如下氨基酸序列的肽(或多肽)(其可以是或可以不是天然存在的),所述氨基酸序列将氨基酸的第一线性序列连接或基因融合至在自然界中并未与其天然地连接或基因融合的氨基酸的第二线性序列。例如,在一个实施方案中,所述肽接头可包含非天然存在的多肽,所述非天然存在的多肽是天然存在的多肽的经过修饰的形式(例如,包含如添加、取代或缺失等突变)。在另一实施方案中,所述肽接头可包含非天然存在的氨基酸。在另一实施方案中,所述肽接头可包含存在于在自然界中不存在的线性序列中的天然存在的氨基酸。在另一实施方案中,所述肽接头可包含天然存在的多肽序列。

[0352] 例如,在某些实施方案中,肽接头可用于融合相同的Fc部分,由此形成同二聚体scFc区。在其他实施方案中,肽接头可用于融合不同的Fc部分(例如,野生型Fc部分和Fc部分变体),由此形成异二聚体scFc区。

[0353] 在另一实施方案中,肽接头包含gly-ser接头或由gly-ser接头组成。在一个实施方案中,scFc或cscFc接头包含免疫球蛋白铰链的至少一部分和gly-ser接头。如本文所用,术语“gly-ser接头”是指由甘氨酸和丝氨酸残基组成的肽。在某些实施方案中,所述gly-ser接头可插入所述肽接头的两个其他序列之间。在其他实施方案中,gly-ser接头附接于所述肽接头的另一序列的一个或两个末端。在其他实施方案中,两个或更多个gly-ser接头连续并入肽接头中。在一个实施方案中,本公开的肽接头包含上部铰链区的至少一部分(例如,源于IgG1、IgG2、IgG3或IgG4分子)、中间铰链区的至少一部分(例如,源于IgG1、IgG2、IgG3或IgG4分子)和一系列gly/ser氨基酸残基。

[0354] 本公开的肽接头是至少一个氨基酸长度并且可具有变化的长度。在一个实施方案中,本公开的肽接头是约1个至约50个氨基酸长度。如本文所用,术语“约”指示+/-两个氨基酸残基。由于接头长度必须是正整数,约1个至约50个氨基酸长度的长度意指1-3至48-52个氨基酸长度的长度。在另一实施方案中,本公开的肽接头是约10个至约20个氨基酸长度。在另一实施方案中,本公开的肽接头是约15个至约50个氨基酸长度。在另一实施方案中,本公

开的肽接头是约20个至约45个氨基酸长度。在另一实施方案中,本公开的肽接头是约15个至约35个或约20个至约30个氨基酸长度。在另一实施方案中,本公开的肽接头是约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个氨基酸长度。在一个实施方案中,本公开的肽接头是20个或30个氨基酸长度。

[0355] 在一些实施方案中,所述肽接头可包含至少两个、至少三个、至少四个、至少五个、至少10个、至少20个、至少30个、至少40个、至少50个、至少60个、至少70个、至少80个、至少90个或至少100个氨基酸。在其他实施方案中,所述肽接头可包含至少200个、至少300个、至少400个、至少500个、至少600个、至少700个、至少800个、至少900个或至少1,000个氨基酸。在一些实施方案中,所述肽接头可包含至少约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个氨基酸。所述肽接头可包含1-5个氨基酸、1-10个氨基酸、1-20个氨基酸、10-50个氨基酸、50-100个氨基酸、100-200个氨基酸、200-300个氨基酸、300-400个氨基酸、400-500个氨基酸、500-600个氨基酸、600-700个氨基酸、700-800个氨基酸、800-900个氨基酸或900-1000个氨基酸。

[0356] 肽接头可使用本领域中已知的技术引入至多肽序列中。修饰可通过DNA序列分析来确认。质粒DNA可用于转化宿主细胞以供所产生的多肽的稳定产生。

[0357] 单体-二聚体杂合物

[0358] 在一些实施方案中,进一步包含异源核苷酸序列的本公开的分离的核酸分子编码包含FVIII的单体-二聚体杂合分子。

[0359] 本文所用的术语“单体-二聚体杂合物”是指包含第一多肽链和第二多肽链的嵌合蛋白,所述多肽链通过二硫键彼此缔合,其中所述第一链包含因子VIII和第一Fc区并且所述第二链包含第二Fc区,基本上由第二Fc区组成,或由第二Fc区组成,而无所述FVIII。因此,所述单体-二聚体杂合物构建体是包含仅具有一种凝固因子的单体方面和具有两个Fc区的二聚体方面的杂合物。

[0360] 表达控制元件

[0361] 在一些实施方案中,本公开的核酸分子或载体进一步包含至少一个表达控制序列。如本文所用的表达控制序列是任何调控核苷酸序列(如启动子序列或启动子-增强子组合),其促进与其可操作性连接的编码核酸的有效转录和翻译。例如,本公开的分离的核酸分子可操作性连接至至少一个转录控制序列。所述基因表达控制序列可例如是哺乳动物或病毒启动子,如组成性或诱导性启动子。组成性哺乳动物启动子包括但不限于用于以下基因的启动子:次黄嘌呤磷酸核糖基转移酶(HPRT)、腺苷脱氨酶、丙酮酸激酶、 β -肌动蛋白启动子和其他组成性启动子。在真核细胞中组成性发挥作用的示例性病毒启动子包括例如来自巨细胞病毒(CMV)、猿猴病毒(例如,SV40)、乳头状瘤病毒、腺病毒、人免疫缺陷病毒(HIV)、劳斯肉瘤病毒、巨细胞病毒、莫洛尼白血病病毒的长末端重复序列(LTR)和其他逆转录病毒的启动子,和单纯疱疹病毒的胸苷激酶启动子。本领域技术人员已知其他组成性启动子。适用作本公开的基因表达序列的启动子还包括诱导性启动子。诱导性启动子在诱导剂存在下表达。例如,金属硫蛋白启动子在某些金属离子存在下经过诱导以促进转录和翻译。本领域普通技术人员已知其他诱导性启动子。

[0362] 在一个实施方案中,本公开包括转基因在组织特异性启动子和/或增强子的控制下的表达。在另一实施方案中,所述启动子或其他表达控制序列选择性地增强肝细胞中所述转基因的表达。肝特异性启动子的实例包括但不限于小鼠甲状腺素启动子(mTTR)、内源人因子VIII启动子(F8)、人 α -1-抗胰蛋白酶启动子(hAAT)、人白蛋白最小启动子和小鼠白蛋白启动子。在一个特定实施方案中,所述启动子包括mTTR启动子。所述mTTR启动子描述于R.H.Costa等人,1986,Mol.Cell.Biol.6:4697中。所述F8启动子描述于Figueiredo和Brownlee,1995,J.Biol.Chem.270:11828-11838中。

[0363] 表达水平可使用一种或多种增强子进一步增强以实现治疗功效。一种或多种增强子可单独或与一种或多种启动子元件一起提供。典型地,所述表达控制序列包含多种增强子元件,和组织特异性启动子。在一个实施方案中,增强子包含 α -1-微球蛋白/尿抑胰酶素(bikunin)增强子的一个或多个拷贝(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)。在另一实施方案中,增强子源于具有Enh1的肝特异性转录因子结合位点,如EBP、DBP、HNF1、HNF3、HNF4、HNF6,包括HNF1、(有义)-HNF3、(有义)-HNF4、(反义)-HNF1、(反义)-HNF6、(有义)-EBP、(反义)-HNF4(反义)。

[0364] 在一个特定实施例,适用于本公开的启动子包括SEQ ID NO:69(即,ET启动子;图11Y),其还称作GenBank号AY661265。还参见Vigna等人,Molecular Therapy 11(5):763(2005)。其他合适载体和基因调控元件的实例描述于WO 02/092134、EP1395293或美国专利号6,808,905、7,745,179或7,179,903中,所述文献以引用的方式整体并入本文中。

[0365] 一般来说,所述表达控制序列应在必要时包括分别牵涉于转录和翻译的起始中的5'非转录和5'非翻译序列,如TATA盒、加帽序列、CAAT序列等。尤其,所述5'非转录序列应包括启动子区,所述启动子区包括用于可操作性接合的编码核酸的转录控制的启动子序列。所述基因表达序列在需要时任选地包括增强子序列或上游活化子序列。

[0366] 载体系统

[0367] 本公开的一些实施方案针对包含本文所述的一种或多种编码具有FVIII活性的多肽的密码子优化的核酸分子的载体,包含所述载体的宿主细胞,和使用所述载体治疗出血病症的方法。本公开通过提供包含优化的FVIII序列的载体满足了本领域中的重要需要,所述优化的FVIII序列展示了受试者中增加的表达并且当用于基因疗法方法中时潜在地引起较大的治疗功效。

[0368] 在一些实施方案中,本公开提供包含分离的核酸分子的载体,所述分离的核酸分子包含如下核苷酸序列,所述核苷酸序列包含编码FVII多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列;其中所述第一核酸序列与(i) SEQ ID NO:3的核苷酸58-1791或(ii) SEQ ID NO:4的核苷酸58-1791具有至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性;并且其中所述N端部分和所述C端部分合起来具有FVIII多肽活性。在其他实施方案中,所述核酸分子包含如下核苷酸序列,所述核苷酸序列包含编码FVIII多肽的N端部分的第一核酸序列和编码FVIII多肽的C端部分的第二核酸序列;其中所述第二核酸序列与(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; 或 (iv) SEQ ID NO:6的核苷酸1792-2277和2320-4374具有至少约80%、至少约85%、至少约86%、至少约87%、至少约88%、至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性; 并且其中所述N端部分和所述C端部分合起来具有FVIII多肽活性。

[0369] 在一些实施方案中, 本公开提供包含分离的核酸分子的载体, 所述分离的核酸分子包含编码具有FVIII活性的多肽的核苷酸序列, 其中所述核苷酸序列包含与 (i) SEQ ID NO:1的核苷酸58-4374或 (ii) SEQ ID NO:1的核苷酸58-2277和2320-4374具有至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列并且可操作性连接至启动子、靶标序列或两者。在其他实施方案中, 所述核酸序列包含 (i) SEQ ID NO:1的核苷酸58-4374或 (ii) SEQ ID NO:1的核苷酸58-2277和2320-4374。

[0370] 在一些实施方案中, 本公开提供包含分离的核酸分子的载体, 所述分离的核酸分子包含编码具有FVIII活性的多肽的核苷酸序列, 其中所述核苷酸序列包含与 (i) SEQ ID NO:2的核苷酸58-4374或 (ii) SEQ ID NO:2的核苷酸58-2277和2320-4374具有至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列并且可操作性连接至启动子、靶标序列或两者。在其他实施方案中, 所述核酸序列包含 (i) SEQ ID NO:2的核苷酸58-4374或 (ii) SEQ ID NO:2的核苷酸58-2277和2320-4374。

[0371] 在一些实施方案中, 本公开提供包含分离的核酸分子的载体, 所述分离的核酸分子包含编码具有FVIII活性的多肽的核苷酸序列, 其中所述核苷酸序列包含与 (i) SEQ ID NO:70的核苷酸58-4374或 (ii) SEQ ID NO:70的核苷酸58-2277和2320-4374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并且可操作性连接至启动子、靶标序列或两者。在其他实施方案中, 所述核酸序列包含 (i) SEQ ID NO:70的核苷酸58-4374或 (ii) SEQ ID NO:70的核苷酸58-2277和2320-4374。

[0372] 在一些实施方案中, 本公开提供包含分离的核酸分子的载体, 所述分离的核酸分子包含编码具有FVIII活性的多肽的核苷酸序列, 其中所述核苷酸序列包含与 (i) SEQ ID NO:71的核苷酸58-4374或 (ii) SEQ ID NO:71的核苷酸58-2277和2320-4374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并且可操作性连接至启动子、靶标序列或两者。在其他实施方案中, 所述核酸序列包含 (i) SEQ ID NO:71的核苷酸58-4374或 (ii) SEQ ID NO:71的核苷酸58-2277和2320-4374。

[0373] 在一些实施方案中, 本公开提供包含分离的核酸分子的载体, 所述分离的核酸分子包含编码具有FVIII活性的多肽的核苷酸序列, 其中所述核苷酸序列包含与 (i) SEQ ID NO:3的核苷酸58-4374或 (ii) SEQ ID NO:3的核苷酸58-2277和2320-4374具有至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列并且可操作性连接至启动子、靶标序列或两者。在其他实施方案中, 所述核酸序列包含 (i) SEQ ID NO:3的核苷酸58-4374或 (ii) SEQ ID NO:3的核苷酸58-2277

和2320-4374。

[0374] 在一些实施方案中,本公开提供包含分离的核酸分子的载体,所述分离的核酸分子包含编码具有FVIII活性的多肽的核苷酸序列,其中所述核苷酸序列包含与(i) SEQ ID NO:4的核苷酸58-4374或(ii) SEQ ID NO:4的核苷酸58-2277和2320-4374具有至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列并且可操作性连接至启动子、靶标序列或两者。在其他实施方案中,所述核苷酸序列包含(i) SEQ ID NO:4的核苷酸58-4374或(ii) SEQ ID NO:4的核苷酸58-2277和2320-4374。

[0375] 在一些实施方案中,本公开提供包含分离的核酸分子的载体,所述分离的核酸分子包含编码具有FVIII活性的多肽的核苷酸序列,其中所述核苷酸序列包含与(i) SEQ ID NO:5的核苷酸58-4374或(ii) SEQ ID NO:5的核苷酸58-2277和2320-4374具有至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列并且可操作性连接至启动子、靶标序列或两者。在其他实施方案中,所述核苷酸序列包含(i) SEQ ID NO:5的核苷酸58-4374或(ii) SEQ ID NO:5的核苷酸58-2277和2320-4374。

[0376] 在一些实施方案中,本公开提供包含分离的核酸分子的载体,所述分离的核酸分子包含编码具有FVIII活性的多肽的核苷酸序列,其中所述核苷酸序列包含与(i) SEQ ID NO:6的核苷酸58-4374或(ii) SEQ ID NO:6的核苷酸58-2277和2320-4374具有至少约89%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%序列同一性的核酸序列并且可操作性连接至启动子、靶标序列或两者。在其他实施方案中,所述核苷酸序列包含(i) SEQ ID NO:6的核苷酸58-4374或(ii) SEQ ID NO:6的核苷酸58-2277和2320-4374。

[0377] 用于本公开的合适载体包括表达载体、病毒载体和质粒载体。在一个实施方案中,所述载体是病毒载体。

[0378] 如本文所用,表达载体是指当引入至适当宿主细胞中时,含有用于所插入的编码序列的转录和翻译的必需元件或在RNA病毒载体的情况下用于复制和翻译的必需元件的任何核酸构建体。表达载体可包括质粒、噬菌粒、病毒和其衍生物。

[0379] 本公开的表达载体将包含编码本文所述的BDD FVIII蛋白的优化的多核苷酸。在一个实施方案中,针对BDD FVIII蛋白的优化的编码序列可操作性连接至表达控制序列。如本文所用,当两个核酸序列以允许各组分核酸序列保持其功能性的方式共价连接时,所述两个核酸序列是可操作性连接的。当编码序列和基因表达控制序列以将所述编码序列的表达或转录和/或翻译置于基因表达控制序列的影响或控制下的方式共价连接时,据说所述编码序列和所述基因表达控制序列是可操作性连接的。如果5'基因表达序列中的启动子的诱导导致编码序列的转录并且如果两个DNA序列之间的连接的性质不会(1)导致移码突变的引入,(2)干扰所述启动子区指导所述编码序列的转录的能力,或(3)干扰相应的RNA转录物翻译成蛋白质的能力,那么据说所述两个DNA序列是可操作性连接的。因此,如果基因表达序列能够实现编码核酸序列的转录,使得所得转录物被翻译成所需的蛋白质或多肽,那么所述基因表达序列将可操作性连接至所述编码核酸序列。

[0380] 病毒载体包括但不限于来自以下病毒的核酸序列:逆转录病毒,如莫洛尼鼠科动

物白血病毒、哈维鼠科动物肉瘤病毒、鼠科动物乳腺肿瘤病毒和劳氏肉瘤病毒；慢病毒；腺病毒；腺相关病毒；SV40类型病毒；多瘤病毒；Epstein-Barr病毒；乳头状瘤病毒；疱疹病毒；牛痘病毒；脊髓灰质炎病毒；和RNA病毒，如逆转录病毒。可容易地使用本领域中众所周知的其他载体。某些病毒载体是基于非细胞病变真核病毒，其中非必需基因已经用所关注的基因置换。非细胞病变病毒包括逆转录病毒，其生命周期涉及将基因组病毒RNA反转录成DNA，所述DNA后续前病毒整合至宿主细胞DNA中。逆转录病毒已经被批准用于人基因疗法试验。最适用的是那些复制缺陷型逆转录病毒（即，能够指导所需的蛋白质的合成，但不能制造传染性粒子）。此类基因改变的逆转录病毒表达载体具有用于体内基因的高效转导的一般效用。用于产生复制缺陷型逆转录病毒的标准方案（包括以下步骤：将外源遗传材料并入至质粒中，用质粒转染包装细胞系，由所述包装细胞系产生重组逆转录病毒，从组织培养基收集病毒粒子，和用病毒粒子感染靶标细胞）提供于Kriegler, M., Gene Transfer and Expression, A Laboratory Manual, W.H. Freeman Co., New York (1990) 和Murry, E. J., Methods in Molecular Biology, 第7卷, Humana Press, Inc., Clifton, N.J. (1991) 中。

[0381] 在一个实施方案中，所述病毒是腺相关病毒（双链DNA病毒）。所述腺相关病毒可工程改造成复制缺陷型并且能够感染广泛范围的细胞类型和物种。其进一步具有以下优势，如热和脂质溶剂稳定性；在不同谱系的细胞中的高转导频率，包括造血细胞；和缺乏重复感染抑制，因此允许多个系列的转导。据报告，所述腺相关病毒可以位点特异性方式整合至人细胞DNA中，由此最小化插入诱变的可能性和逆转录病毒感染所特有的插入基因表达的变异性。另外，野生型腺相关病毒感染已经在选择压力不存在下在组织培养物中持续超过100代，暗示腺相关病毒基因组整合是相对稳定的事件。所述腺相关病毒也可以染色体外形式起作用。

[0382] 在另一实施方案中，所述病毒载体是已经操纵以携带编码如本文所公开的FVIII蛋白的多核苷酸的腺相关病毒（AAV）。已经公开了用于获得重组AAV（rAAV）的一般方法。参见例如USP 8,734,809、2013/0195801以及其中所引用的参考文献。在一些实施方案中，rAAV载体包含一个或多个AAV反向末端重复序列（ITR）和所关注的转基因（例如，优化的FVIII多核苷酸序列）。在某些实施方案中，制造rAAV的方法涉及培养所需的宿主细胞，所述宿主细胞含有编码AAV衣壳蛋白或其片段的核酸序列；功能性rep基因；由AAV反向末端重复序列（ITR）和所关注的转基因构成的rAAV载体；和充分的辅助功能以允许将重组AAV载体包装至所述AAV衣壳蛋白中。用于执行这些和相关程序的材料和方法已经公开于例如USP 8,734,809、2013/0195801、PCT/US1997/015692、PCT/US2002/033692、PCT/US2002/033630、W02007/148971、W000/20561、W003/042361和W02007/04670中。

[0383] 源于几乎任何血清型的不同AAV载体序列中的一者或多者可根据本公开使用。特定AAV载体序列的选择将由如所关注的向性、所需的载体产率等已知参数指导。一般来说，AAV血清型具有在氨基酸和核酸层面上具有显著同源性的基因组序列，提供基因功能的相关集合，产生相关的病毒体，并且相似地复制和组装。关于多种AAV血清型的基因组序列和基因组相似性的概述，参见例如GenBank登录号U89790；GenBank登录号J01901；GenBank登录号AF043303；GenBank登录号AF085716；Chlorini等人（1997, J. Vir. 71:6823-33）；Srivastava等人（1983, J. Vir. 45:555-64）；Chlorini等人（1999, J. Vir. 73:1309-1319）；Rutledge等人（1998, J. Vir. 72:309-319）；和Wu等人（2000, J. Vir. 74:8635-47）。AAV血清型

1、2、3、4和5是用于本公开的上下文的AAV核苷酸序列的说明性来源。通过例如衣壳穿梭技术和AAV衣壳文库或由新近设计、开发或发展的ITR获得的AAV6、AAV7、AAV8或AAV9或新近开发的AAV样粒子也适用于某些公开应用。参见Dalkara,D等人(2013),*Sci.Transl.Med.* 5 (189):189ra76;Kotterman,MA *Nat.Rev.Genet.* (2014) 15 (7):455。

[0384] 然而,在某些实施方案中,对肝和相关组织具有显著向性的AAV载体将在表达本文所公开的FVIII蛋白方面受到关注。非限制性实例包括AAV血清型1、2、6和8。参见例如Torres-Torranteras等人(2014) 22:901和其中引用的参考文献。

[0385] 在其他实施方案中,所述载体源于慢病毒。在某些实施方案中,所述载体是能够感染非分裂细胞的重组慢病毒的载体。

[0386] 慢病毒基因组和前病毒DNA典型地具有在逆转录病毒中发现的三种基因:gag、pol和env,所述基因侧接两个长末端重复(LTR)序列。gag基因编码内部结构(基质、衣壳和核衣壳)蛋白;pol基因编码RNA指导的DNA聚合酶(反转录酶)、蛋白酶和整合酶;并且env基因编码病毒包膜糖蛋白。5'和3'LTR用于促进病毒体RNA的转录和多腺苷酸化。所述LTR含有病毒复制所必需的所有其他顺式作用序列。慢病毒具有另外的基因,包括vif、vpr、tat、rev、vpu、nef和vpx(在HIV-1、HIV-2和/或SIV中)。

[0387] 与5'LTR相邻的是基因组的反转录(tRNA引物结合位点)和将病毒RNA有效衣壳化至粒子(Psi位点)中所必需的序列。如果衣壳化(或将逆转录病毒RNA包装至传染性病毒体中)所必需的序列从病毒基因组中缺失,那么所述顺式缺陷会防止基因组RNA的衣壳化。

[0388] 然而,所得突变体仍能够指导所有病毒体蛋白的合成。本公开提供一种产生能够感染非分裂细胞的重组慢病毒的方法,所述方法包括用携带包装功能的两种或更多种载体(即,gag、pol和env,以及rev和tat)转染合适的宿主细胞。如下文中将公开,缺乏功能性tat基因的载体是某些应用所需要的。因此,例如,第一载体可提供编码病毒gag和病毒pol的核酸并且另一载体可提供编码病毒env的核酸以产生包装细胞。将提供异源基因的载体(本文中鉴别为转移载体)引入至所述包装细胞中会产生生产细胞,所述生产细胞释放携带所关注的外来基因的传染性病毒粒子。

[0389] 根据上文所指示的载体和外来基因的配置,所述第二载体可提供编码病毒包膜(env)基因的核酸。所述env基因可源于几乎任何合适的病毒,包括逆转录病毒。在一些实施方案中,所述env蛋白是允许人和其他物种的细胞的转导的两性包膜蛋白。

[0390] 逆转录病毒源性env基因的实例包括但不限于:莫洛尼鼠科动物白血病病毒(MoMuLV或MMLV)、哈维鼠科动物肉瘤病毒(HaMuSV或HSV)、鼠科动物乳腺肿瘤病毒(MuMTV或MMTV)、长臂猿ape白血病病毒(GaLV或GALV)、人免疫缺陷病毒(HIV)和劳氏肉瘤病毒(RSV)。也可使用其他env基因,如水疱性口炎病毒(VSV)蛋白G(VSV G)、肝炎病毒和流感的env基因。

[0391] 提供病毒env核酸序列的载体与本文中别处所述的调控序列可操作性缔合。

[0392] 在某些实施方案中,所述载体包括慢病毒载体,其中HIV毒力基因env、vif、vpr、vpu和nef缺失,而未损害所述载体转导非分裂细胞的能力。

[0393] 在一些实施方案中,所述载体包括慢病毒载体,所述慢病毒载体包含3'LTR的U3区的缺失。U3区的缺失可以是完全缺失或部分缺失。

[0394] 在一些实施方案中,包含本文所述的FVIII核苷酸序列的本公开的慢病毒载体可

在细胞中用 (a) 包含 gag、pol 或 gag 和 pol 基因的第一核苷酸序列和 (b) 包含异源 env 基因的第二核苷酸序列转染; 其中所述慢病毒载体缺乏功能性 tat 基因。在其他实施方案中, 所述细胞进一步用包含 rev 基因的第四核苷酸序列转染。在某些实施方案中, 所述慢病毒载体缺乏选自 vif、vpr、vpu、vpx 和 nef 或其组合的功能性基因。

[0395] 在某些实施方案中, 慢病毒载体包含编码 gag 蛋白、Rev 应答元件、中央聚嘌呤序列 (central polypurine track, cPPT) 或其任何组合的一个或多个核苷酸序列。

[0396] 所述慢病毒载体的实例公开于 W09931251、W09712622、W09817815、W09817816 和 W09818934 中, 所述文献以引用的方式整体并入本文中。

[0397] 其他载体包括质粒载体。质粒载体已经广泛地描述于本领域中并且是本领域技术人员众所周知的。参见例如 Sambrook 等人, *Molecular Cloning: A Laboratory Manual*, 第 2 版, Cold Spring Harbor Laboratory Press, 1989。最近数年来, 已经发现质粒载体由于其无法在宿主基因组中复制和整合至宿主基因组中而特别有利于体内将基因递送至细胞。然而, 这些具有与宿主细胞可相容的启动子的质粒可自在所述质粒内可操作性编码的基因表达肽。可获自商业供应商的一些常用的质粒包括 pBR322、pUC18、pUC19、多种 pcDNA 质粒、pRC/CMV、多种 pCMV 质粒、pSV40 和 pBlueScript。特定质粒的另外实例包括 pcDNA3.1, 目录号 V79020; pcDNA3.1/hygro, 目录号 V87020; pcDNA4/myc-His, 目录号 V86320; 和 pBudCE4.1, 目录号 V53220, 均来自 Invitrogen (Carlsbad, CA.)。其他质粒是本领域普通技术人员众所周知的。另外, 质粒可使用标准分子生物学技术定制设计以去除和/或添加 DNA 的特定片段。

[0398] 组织特异性表达

[0399] 在某些实施方案中, 在所述载体内包括一个或多个 miRNA 靶标序列将是适用的, 所述 miRNA 靶标序列例如可操作性连接至优化的 FVIII 转基因。因此, 本公开还提供可操作性连接至优化的 FVIII 核苷酸序列或以其他方式插入载体内的至少一个 miRNA 序列靶标。包括于所述载体中的 miRNA 靶标序列的超过一个拷贝可增加系统的有效性。还包括不同的 miRNA 靶标序列。例如, 表达超过一种转基因的载体可具有在超过一个 miRNA 靶标序列的控制下的转基因, 所述超过一个 miRNA 靶标序列可以是相同或不同的。所述 miRNA 靶标序列可以是串联的, 但也包括其他布置。含有 miRNA 靶标序列的转基因表达盒也可以反义取向插入所述载体内。反义取向可适用于病毒粒子的产生以避免可在其他方面对所述生产细胞有毒的基因产物的表达。在其他实施方案中, 所述载体包含相同或不同的 miRNA 靶标序列的 1、2、3、4、5、6、7 或 8 个拷贝。然而, 在某些其他实施方案中, 所述载体将不包括任何 miRNA 靶标序列。是否包括 miRNA 靶标序列 (和多少个) 的选择将由如预期的组织靶标、所需的表达水平等已知参数指导。

[0400] 在一个实施方案中, 所述靶标序列是 miR-223 靶标, 已经报告所述 miR-223 靶标最有效地阻断骨髓定向祖细胞中的表达并且至少部分地阻断更原始的 HSPC 中的表达。miR-223 靶标可阻断分化的骨髓细胞 (包括粒细胞、单核细胞、巨噬细胞、骨髓树突状细胞) 中的表达。miR-223 靶标也可适用于基因疗法应用, 所述基因疗法应用依赖于淋巴或红细胞谱系中的稳固转基因表达。miR-223 靶标也可极有效地阻断人 HSC 中的表达。

[0401] 在另一实施方案中, 所述靶标序列是 miR142 靶标 (tccataaagt aggaacact aca (SEQ ID NO:43))。在一个实施方案中, 所述载体包含 miR-142 靶标序列的 4 个拷贝。在某些实施方案中, 将造血特异性微 RNA (如 miR-142 (142T)) 的互补序列并入至例如慢病毒载体

(LV) 等载体的3'非翻译区中,使得转基因编码转录物易受到miRNA介导的下调。通过这种方法,可防止造血谱系抗原递呈细胞(APC)中的转基因表达,同时在非造血细胞中维持转基因表达(Brown等人,Nat Med 2006)。这种策略可对转基因表达施加严苛的转录后控制并且因此使得转基因能够稳定递送并且长期表达。在一些实施方案中,miR-142调控会防止免疫介导的转导细胞清除和/或诱导抗原特异性调控T细胞(T_{reg})并且介导对转基因编码抗原的稳固免疫耐受性。

[0402] 在一些实施方案中,所述靶标序列是miR181靶标。Chen C-Z和Lodish H,Seminars in Immunology (2005) 17 (2):155-165公开了miR-181,所述miR-181是在小鼠骨髓内的B细胞中特异性表达的miRNA(Chen和Lodish,2005)。其还公开了一些人miRNA与白血病有关。

[0403] 所述靶标序列可与所述miRNA完全或部分互补。术语“完全互补”意指所述靶标序列具有与识别所述靶标序列的miRNA的序列100%互补的核酸序列。术语“部分互补”意指所述靶标序列仅与识别所述靶标序列的miRNA的序列部分互补,由此所述部分互补序列仍由所述miRNA识别。换句话说,在本公开的上下文中的部分互补靶标序列有效识别相应的miRNA并且实现表达所述miRNA的细胞中的转基因表达的防止或降低。所述miRNA靶标序列的实例描述于W02007/000668、W02004/094642、W02010/055413或W02010/125471中,所述文献以引用的方式整体并入本文中。

[0404] 宿主细胞

[0405] 本公开还提供包含本公开的核酸分子或载体的宿主细胞。如本文所用,术语“转化”应以广泛含义使用以指将DNA引入至受体宿主细胞中,所述引入会改变基因型并且因而导致受体细胞中的改变。

[0406] “宿主细胞”是指已经用使用重组DNA技术构建并且编码至少一种异源基因的载体转化的细胞。本公开的宿主细胞优选地具有哺乳动物起源;最优选地具有人或小鼠起源。相信本领域技术人员具有优先确定最适用于其目的的特定宿主细胞系的能力。示例性宿主细胞系包括但不限于CHO、DG44和DUXB11(中国仓鼠卵巢细胞系,DHFR⁻)、HELA(人子宫颈癌)、CVI(猴肾细胞系)、COS(具有SV40T抗原的CVI的衍生物)、R1610(中国仓鼠成纤维细胞)BALBC/3T3(小鼠成纤维细胞)、HAK(仓鼠肾细胞系)、SP2/0(小鼠骨髓瘤)、P3.times.63-Ag3.653(小鼠骨髓瘤)、BFA-1c1BPT(牛内皮细胞)、RAJI(人淋巴细胞)、PER.C6[®]、NS0、CAP、BHK21和HEK 293(人肾)。在一个特定实施方案中,所述宿主细胞选自由以下组成的组:CHO细胞、HEK293细胞、BHK21细胞、PER.C6[®]细胞、NS0细胞和CAP细胞。宿主细胞系典型地可获自商业服务美国培养物保藏中心(American Tissue Culture Collection),或获自公布的文献。

[0407] 将本公开的分离的核酸分子或载体引入至宿主细胞中可通过本领域技术人员众所周知的多种技术实现。这些技术包括但不限于转染(包括电泳和电穿孔)、原生质体融合、磷酸钙沉淀、与包膜DNA的细胞融合、微注射和用完整病毒感染。参见Ridgway,A.A.G.“Mammalian Expression Vectors”第24.2章,第470-472页Vectors,Rodriguez和Denhardt编(Butterworths,Boston,Mass.1988)。最优选地,质粒引入至宿主中是经由电穿孔。转化细胞在适合轻链和重链的产生的条件下生长,并且针对重链和/或轻链蛋白合成进行测定。示例性测定技术包括酶联免疫吸附测定(ELISA)、放射免疫测定(RIA)或荧光活化的细胞分

选仪分析 (FACS)、免疫组织化学等。

[0408] 包含本公开的分离的核酸分子或载体的宿主细胞在适当的生长培养基中生长。如本文所用,术语“适当的生长培养基”意指含有细胞生长所需的营养物的培养基。细胞生长所需的营养物可包括碳源、氮源、必需氨基酸、维生素、矿物质和生长因子。任选地,所述培养基可含有一种或多种选择因子。任选地,所述培养基可含有牛犊血清或胎牛血清 (FCS)。在一个实施方案中,所述培养基大体上不含IgG。所述生长培养基一般将通过例如药物选择或必需营养物的缺乏而选择含有所述DNA构建体的细胞,所述必需营养物通过所述DNA构建体上的可选择标记物来补充或用所述DNA构建体共转染。培养的哺乳动物细胞一般在市面上有售的含血清或无血清培养基(例如, MEM、DMEM、DMEM/F12)中生长。在一个实施方案中,所述培养基是CDoptiCHO (Invitrogen, Carlsbad, CA.)。在另一实施方案中,所述培养基是CD17 (Invitrogen, Carlsbad, CA.)。适合用于所用的特定细胞系的培养基的选择在本领域普通技术人员的水平内。

[0409] 多肽的制备

[0410] 本公开还提供由本公开的核酸分子编码的多肽。在其他实施方案中,本公开的多肽由包含本公开的分离的核酸分子的载体编码。在其他实施方案中,本公开的多肽由包含本公开的分离的核酸分子的宿主细胞产生。

[0411] 在其他实施方案中,本公开还提供一种产生具有FVIII活性的多肽的方法,所述方法包括在一定条件下培养本公开的宿主细胞,由此产生具有FVIII活性的多肽;和回收所述具有FVIII活性的多肽。在一些实施方案中,所述具有FVIII活性的多肽的表达相对于在包含包含SEQ ID NO:16(亲本FVIII基因序列)的参考核苷酸序列的相同条件下培养的宿主细胞增加。

[0412] 在其他实施方案中,本公开提供一种增加具有FVIII活性的多肽的表达的方法,所述方法包括在一定条件下培养本公开的宿主细胞,由此由所述核酸分子表达具有FVIII活性的多肽,其中所述具有FVIII活性的多肽的表达相对于在包含包含SEQ ID NO:16的参考核酸分子的相同条件下培养的宿主细胞增加。

[0413] 在其他实施方案中,本公开提供一种改进具有FVIII活性的多肽的产率的方法,所述方法包括在一定条件下培养宿主细胞,由此由所述核酸分子产生具有FVIII活性的多肽,其中具有FVIII活性的多肽的产率相对于在包含包含SEQ ID NO:16的参考核酸序列的相同条件下培养的宿主细胞增加。

[0414] 在其他实施方案中,本公开提供一种改进具有FVIII活性的多肽的产率的方法,所述方法包括培养包含编码所述多肽的核苷酸序列的宿主细胞,其中所述核苷酸序列的3'部分的密码子适应指数相对于所述核苷酸序列的5'部分增加;其中具有FVIII活性的多肽的产率相对于在包含含有SEQ ID NO:16的参考核酸序列的相同条件下培养的宿主细胞增加。在一些实施方案中,所述核苷酸序列的5'部分的密码子适应指数相对于SEQ ID NO:16的密码子适应指数增加,减少,或未改变。

[0415] 在其他实施方案中,本公开提供一种改进具有FVIII活性的多肽的产率的方法,所述方法包括培养包含编码所述多肽的核苷酸序列的宿主细胞,其中所述核苷酸序列的5'部分的密码子适应指数相对于所述核苷酸序列的3'部分增加;其中具有FVIII活性的多肽的产率相对于在包含含有SEQ ID NO:16的参考核酸序列的相同条件下培养的宿主细胞增加。

在一些实施方案中,所述核苷酸序列的3'部分的密码子适应指数相对于SEQ ID NO:16的密码子适应指数增加,减少,或未改变。

[0416] 在其他实施方案中,本公开提供一种改进具有FVIII活性的多肽的产率的方法,所述方法包括培养包含编码所述多肽的核苷酸序列的宿主细胞,其中所述核苷酸中编码所述多肽的C端部分的部分的密码子适应指数相对于所述核苷酸中编码所述多肽的N端部分的部分增加;其中具有FVIII活性的多肽的产率相对于在包含含有SEQ ID NO:16的参考核酸序列的相同条件下培养的宿主细胞增加。在一些实施方案中,所述核苷酸中编码所述多肽的N端部分的部分的密码子适应指数相对于SEQ ID NO:16的密码子适应指数增加,减少,或未改变。

[0417] 在其他实施方案中,本公开提供一种改进具有FVIII活性的多肽的产率的方法,所述方法包括培养包含编码所述多肽的核苷酸序列的宿主细胞,其中所述核苷酸中编码所述多肽的N端部分的部分的密码子适应指数相对于所述核苷酸中编码所述多肽的C端部分的部分增加;其中具有FVIII活性的多肽的产率相对于在包含含有SEQ ID NO:16的参考核酸序列的相同条件下培养的宿主细胞增加。在一些实施方案中,所述核苷酸中编码所述多肽的C端部分的部分的密码子适应指数相对于SEQ ID NO:16的密码子适应指数增加,减少,或未改变。

[0418] 在改进具有FVIII活性的多肽的产率的某些实施方案中,所述核苷酸序列的5'部分适当地比时与SEQ ID NO:1的约核苷酸1-1791、SEQ ID NO:1的核苷酸58-1791或其片段一致。在其他实施方案中,由所述核苷酸的5'部分编码的多肽适当地比时与SEQ ID NO:17的约氨基酸1-497、SEQ ID NO:17的氨基酸20-497或其片段一致。在某些实施方案中,所述核苷酸序列中编码所述多肽的N端部分的部分适当地比时与SEQ ID NO:1的约核苷酸1-1791、SEQ ID NO:1的核苷酸58-1791或其片段一致。在一些实施方案中,所述核苷酸序列的3'部分适当地比时与SEQ ID NO:1的约核苷酸1792-4374或其片段一致。在其他实施方案中,由所述核苷酸的3'部分编码的多肽适当地比时与SEQ ID NO:17的约氨基酸498-1458或其片段一致。在某些实施方案中,所述核苷酸序列中编码所述多肽的C端部分的部分适当地比时与SEQ ID NO:1的约核苷酸1792-4374或其片段一致。

[0419] 在一些实施方案中,所述FVIII多肽的表达相对于在包含含有SEQ ID NO:16的参考核酸序列的相同条件下培养的宿主细胞增加至少约1.5倍、至少约2倍、至少约3倍、至少约4倍、至少约5倍、至少约6倍、至少约7倍、至少约8倍、至少约9倍、至少约10倍、至少约11倍、至少约12倍、至少约13倍、至少约14倍、至少约15倍、至少约20倍、至少约25倍、至少约30倍、至少约35倍、至少约40倍、至少约50倍、至少约60倍、至少约70倍、至少约80倍、至少约90倍、至少约100倍、至少约150倍或至少约200倍。

[0420] 多种方法可用于由本公开的优化的核酸分子重组产生FVIII蛋白。所需序列的多核苷酸可通过从头固相DNA合成或通过早先制备的多核苷酸的PCR诱变产生。寡核苷酸介导的诱变是一种用于制备核苷酸序列中的取代、插入、缺失或改变(例如,改变的密码子)的方法。例如,起始DNA通过使编码所需突变的寡核苷酸与单链DNA模板杂交而改变。在杂交之后,使用DNA聚合酶来合成所述模板的完整第二互补链,所述完整第二互补链并入寡核苷酸引物。在一个实施方案中,基因工程改造(例如,基于引物的PCR诱变)足以并入如本文所定义的改变,以产生本公开的多核苷酸。

[0421] 关于重组蛋白产生,将编码FVIII蛋白的本公开的优化的多核苷酸序列插入至适当的表达媒介物中,所述表达媒介物即含有用于所插入的编码序列的转录和翻译的必需元件或在RNA病毒载体的情况下含有用于复制和翻译的必需元件的载体。

[0422] 本公开的多核苷酸序列在适当阅读框中插入至所述载体中。所述表达载体接着转染至将表达所述多肽的合适靶细胞中。本领域中已知的转染技术包括但不限于磷酸钙沉淀(Wigler等人1978,Cell 14:725)和电穿孔(Neumann等人1982,EMBO,J.1:841)。多种宿主-表达载体系统可用于在真核细胞中表达本文所述的FVIII蛋白。在一个实施方案中,所述真核细胞是动物细胞,包括哺乳动物细胞(例如,HEK293细胞、PER.C6[®]、CHO、BHK、Cos、HeLa细胞)。本公开的多核苷酸序列也可编码将允许分泌所述FVIII蛋白的信号序列。本领域技术人员应理解,虽然所述FVIII蛋白经过翻译,但所述信号序列由所述细胞裂解以形成成熟蛋白。多种信号序列是本领域中已知的,例如原生因子V11信号序列、原生因子IX信号序列和小鼠IgK轻链信号序列。或者,在未包括信号序列的情况下,所述FVIII蛋白可通过溶解所述细胞来回收。

[0423] 本公开的FVIII蛋白可在转基因动物,如啮齿动物、山羊、绵羊、猪或牛中合成。术语“转基因动物”是指已经将外来基因并入至其基因组中的非人动物。因为这种基因存在于生殖系组织中,所以其从母体传代至后代。外源基因被引入至单细胞胚胎中(Brinster等人1985,Proc.Natl.Acad.Sci.USA 82:4438)。产生转基因动物的方法是本领域中已知的,包括产生免疫球蛋白分子的转基因技术(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)。

[0424] 所述表达载体可编码允许重组产生的蛋白质的容易纯化或鉴别的标签。实例包括但不限于载体pUR278(Ruther等人1983,EMBO J.2:1791),其中本文所述的FVIII蛋白的编码序列可与lac Z编码区同框连接至所述载体中,使得产生杂合蛋白;pGEX载体可用于表达具有谷胱甘肽S-转移酶(GST)标签的蛋白质。这些蛋白质通常是可溶的并且可容易地通过吸附至基质谷胱甘肽-琼脂糖珠粒、随后在游离谷胱甘肽存在下洗脱而从细胞纯化。所述载体包括用于纯化之后所述标签的容易去除的裂解位点(例如,PreCission Protease(Pharmacia,Peapack,N.J.))。

[0425] 出于本公开的目的,可使用多种表达载体系统。这些表达载体通常在宿主生物体中可复制,呈游离体形式或作为宿主染色体DNA的组成部分。表达载体可包括表达控制序列,包括但不限于启动子(例如,天然缔合或异源启动子)、增强子、信号序列、剪接信号、增强子元件和转录终止序列。优选地,所述表达控制序列是能够转化或转染真核宿主细胞的载体中的真核启动子系统。表达载体还可使用源于动物病毒的DNA元件,所述动物病毒如牛乳头状瘤病毒、多瘤病毒、腺病毒、牛痘病毒、杆状病毒、逆转录病毒(RSV、MMTV或MOMLV)、巨细胞病毒(CMV)或SV40病毒。其他涉及使用具有内部核糖体结合位点的多顺反子系统。

[0426] 通常,表达载体含有选择标记(例如,氨苄青霉素抗性、潮霉素抗性、四环素抗性 or 新霉素抗性)以允许检测用所需的DNA序列转化的那些细胞(参见例如Itakura等人,美国专利4,704,362)。已经将所述DNA整合至其染色体中的细胞可通过引入一种或多种允许转化的宿主细胞的选择的标记来选择。所述标记可提供针对营养缺陷型宿主的原养型、生物杀

灭剂抗性(例如,抗生素)或对如铜等重金属的抗性。所述可选择标记基因可直接地连接至欲表达的DNA序列,或通过共转化引入至同一细胞中。

[0427] 适用于表达优化的FVIII序列的载体的实例是NEOSPLA(美国专利号6,159,730)。这种载体含有巨细胞病毒启动子/增强子、小鼠 β 球蛋白主要启动子、SV40复制起点、牛生长激素多腺苷酸化序列、新霉素磷酸转移酶外显子1和外显子2、二氢叶酸还原酶基因和前导序列。已经发现这种载体在可变区和恒定区基因的并入、细胞中的转染、随后含G418培养基中的选择和甲氨蝶呤扩增后引起极高水平的抗体表达。载体系统也教导于美国专利号5,736,137和5,658,570中,所述美国专利中的每一者均以引用的方式整体并入本文中。这种系统提供高表达水平,例如>30pg/细胞/天。其他示例性载体系统公开于美国专利号6,413,777中。

[0428] 在其他实施方案中,本公开的公开的多肽可使用多顺反子构建体表达。在这些表达系统中,可从单一多顺反子构建体产生多种所关注的基因产物,如多聚体结合蛋白的多种多肽。这些系统有利地使用内部核糖体进入位点(IRES)以在真核宿主细胞中提供相对高水平的多肽。可相容的IRES序列公开于美国专利号6,193,980中,所述美国专利也并入本文中。

[0429] 更一般地,一旦已经制备编码多肽的载体或DNA序列,即可将所述表达载体引入至适当宿主细胞中。即,所述宿主细胞可被转化。将质粒引入至宿主细胞中可通过本领域技术人员众所周知的多种技术实现,如上文所论述。转化细胞在适合FVIII多肽的产生的条件下生长,并且针对FVIII多肽合成进行测定。示例性测定技术包括酶联免疫吸附测定(ELISA)、放射免疫测定(RIA)或荧光活化的细胞分选仪分析(FACS)、免疫组织化学等。

[0430] 在用于从重组宿主分离多肽的方法的描述中,除非另外清楚地规定,否则术语“细胞”和“细胞培养物”可互换使用以表示多肽的来源。换句话说,从“细胞”回收多肽可意指从短暂离心的全细胞,或从含有培养基和悬浮的细胞两者的细胞培养物回收。

[0431] 用于蛋白质表达的宿主细胞系优选地具有哺乳动物起源;最优选地具有人或小鼠起源,因为本公开的分离的核酸已经针对人细胞中的表达经过优化。上文已经描述了示例性宿主细胞系。在产生具有FVIII活性的多肽的方法的一个实施方案中,所述宿主细胞是HEK293细胞。在产生具有FVIII活性的多肽的方法的另一实施方案中,所述宿主细胞是CHO细胞。

[0432] 编码本公开的多肽的基因也可在如细菌或酵母或植物细胞等非哺乳动物细胞中表达。在这方面,应理解如细菌等多种单细胞非哺乳动物微生物也可被转化;即,能够在培养物中或在发酵中生长的那些。易经受转化的细菌包括以下的成员:肠杆菌科,如大肠杆菌或沙门氏菌属菌株;杆菌科,如枯草芽胞杆菌;肺炎球菌;链球菌,和流感嗜血杆菌。应进一步理解,当在细菌中表达时,所述多肽典型地成为包涵体的一部分。所述多肽必须经过分离,经过纯化并且接着组装至功能分子中。

[0433] 或者,本公开的优化的核苷酸序列可并入转基因中以引入至转基因动物的基因组中并且后续在所述转基因动物的乳汁中表达(参见例如Deboer等人,US 5,741,957,Rosen,US 5,304,489,和Meade等人,US5,849,992)。合适的转基因包括用于与来自乳腺特异性基因(如酪蛋白或 β 微球蛋白)的启动子和增强子可操作性连接的多肽的编码序列。

[0434] 体外产生允许按比例增加以生成大量的所需多肽。用于组织培养条件下的哺乳动

物细胞培养的技术是本领域中已知的并且包括均质悬浮培养(例如,在气升式反应器中或在连续搅拌反应器中)或固定或截留细胞培养(例如,在中空纤维、微胶囊中,在琼脂糖微珠或陶瓷筒上)。必需和/或必要时,多肽的溶液可通过惯用色谱方法,例如凝胶过滤、离子交换色谱法、DEAE-纤维素色谱法或(免疫)亲和色谱法来纯化,例如在合成铰链区多肽的优先生物合成之后或在本文所述的HIC色谱法步骤之前或之后。亲和标签序列(例如,His(6)标签)可任选地附接或包括于所述多肽序列内以促进下游纯化。

[0435] 一旦经过表达,所述FVIII蛋白可根据本领域的标准程序来纯化,包括硫酸铵沉淀、亲和柱色谱法、HPLC纯化、凝胶电泳等(一般参见Scopes,Protein Purification (Springer-Verlag,N.Y.,(1982))。关于药物用途,至少约90-95%均质性的的大体上纯蛋白质是优选的,并且98-99%或更高均质性是最优选的。

[0436] 药物组合物

[0437] 含有本公开的分离的核酸分子、由所述核酸分子编码的具有FVIII活性的多肽、载体或宿主细胞的组合物可含有合适的药学上可接受的载体。例如,其可含有促进将活性化合物加工成设计用于递送至作用位点的制剂的赋形剂和/或助剂。

[0438] 所述药物组合物可经过配制以通过团注进行肠胃外施用(即,静脉内、皮下或肌肉内)。用于注射的制剂可以单位剂型,例如在安瓿中或在具有添加的防腐剂的多剂量容器中呈递。所述组合物可采用如油性或水性媒剂中的悬浮液、溶液或乳液等形式,并且含有如悬浮剂、稳定剂和/或分散剂等配制剂。或者,所述活性成分可呈用于由合适媒剂(例如,无热原水)复原的粉末形式。

[0439] 用于肠胃外施用的合适制剂还包括呈水溶性形式的活性化合物(例如,水溶性盐)的水溶液。另外,可施用呈适当油性注射悬浮液形式的所述活性化合物的悬浮液。合适的亲脂性溶剂或媒剂包括脂肪油(例如,芝麻油)或合成脂肪酸酯(例如,油酸乙酯或三酸甘油酯)。水性注射悬浮液可含有增加所述悬浮液的粘度的物质,包括例如羧基甲基纤维素钠、山梨醇和葡聚糖。任选地,所述悬浮液还可含有稳定剂。脂质体也可用于封装本公开的分子以递送至细胞或胞间隙中。示例性药学上可接受的载剂是生理学上可相容的溶剂、分散介质、涂层、抗菌和抗真菌剂、等渗和吸收延迟剂、水、生理盐水、磷酸盐缓冲生理盐水、右旋糖、甘油、乙醇等。在一些实施方案中,所述组合物包含等渗剂,例如糖、多元醇(如甘露醇、山梨醇)或氯化钠。在其他实施方案中,所述组合物包含药学上可接受的物质,如湿润剂或微量的辅助物质(如湿润或乳化剂、防腐剂或缓冲剂),所述药学上可接受的物质增强活性成分的保质期或有效性。

[0440] 本公开的组合物可呈多种形式,包括例如液体(例如,可注射和可输注溶液)、分散液、悬浮液、半固体和固体剂型。优选的形式取决于施用模式和治疗应用。

[0441] 所述组合物可配制成溶液、微乳液、分散液、脂质体或适合高药物浓度的其他有序结构。无菌可注射溶液可通过在有需要时具有一种上文所列举的成分或成分的组合的适当溶剂中并入所需量的活性成分、随后过滤灭菌来制备。一般地,分散液通过将活性成分并入无菌媒剂中来制备,所述无菌媒剂含有基础分散介质和来自上文所列举的那些的所需其他成分。在用于制备无菌可注射溶液的无菌粉末的情况下,优选制备方法是真空干燥和冷冻干燥,其会生成活性成分加上来自先前无菌过滤的溶液的任何另外所需成分的粉末。溶液的适当流动性可例如通过使用涂层(如卵磷脂),在分散液的情况下通过维持所需粒径,和

通过使用表面活性剂来维持。可注射组合物的延长吸收可通过在所述组合物中包括延迟吸收的剂(例如单硬脂酸盐和明胶)来引起。

[0442] 活性成分可用控制释放制剂或器件来配制。所述制剂和器件的实例包括植入物、透皮贴剂和微封装的递送系统。可使用生物可降解、生物可相容聚合物,例如乙烯乙酸乙烯酯、聚酐、聚乙醇酸、胶原蛋白、聚原酸酯和聚乳酸。本领域技术人员已知用于制备所述制剂和器件的方法。参见例如Sustained and Controlled Release Drug Delivery Systems, J.R.Robinson编,Marcel Dekker,Inc.,New York,1978。

[0443] 可注射储库制剂可通过在如聚丙交酯-聚乙交酯等生物可降解聚合物中形成所述药物的微封装基质而制得。视药物与聚合物的比率和所用聚合物的性质而定,可控制药物释放速率。其他示例性生物可降解聚合物是聚原酸酯和聚酐。储库可注射制剂还可通过将药物截留于脂质体或微乳液中来制备。

[0444] 补充活性化合物可并入至所述组合物中。在一个实施方案中,本公开的嵌合蛋白与另一凝固因子或其变体、片段、类似物或衍生物配制。例如,所述凝固因子包括但不限于因子V、因子VII、因子VIII、因子IX、因子X、因子XI、因子XII、因子XIII、凝血酶原、纤维蛋白原、血管性血友病因子或重组可溶性组织因子(rsTF)或前述任一者的活化形式。止血剂的凝固因子也可包括抗纤维蛋白溶解药物,例如 ϵ -氨基-己酸、氨甲环酸。

[0445] 给药方案可经过调节以提供最佳所需响应。例如,可施用单一大丸药,可随时间施用数个分次剂量,或剂量可如治疗情形的紧急性所指示按比例降低或增加。为了便于剂量的施用和均一性,以剂量单位形式配制肠胃外组合物是有利的。参见例如Remington's Pharmaceutical Sciences (Mack Pub.Co.,Easton,Pa.1980)。

[0446] 除了活性化合物以外,液体剂型还可含有惰性成分,如水、乙醇、碳酸乙酯、乙酸乙酯、苯甲醇、苯甲酸苯甲酯、丙二醇、1,3-丁二醇、二甲基甲酰胺、油、甘油、四氢糠醇、聚乙二醇和山梨聚糖的脂肪酸酯。

[0447] 合适的药物载剂的非限制性实例也由E.W.Martin描述于Remington's Pharmaceutical Sciences中。赋形剂的一些实例包括淀粉、葡萄糖、乳糖、蔗糖、明胶、麦芽、稻、面粉、白垩、硅胶、硬脂酸钠、甘油单硬脂酸酯、滑石、氯化钠、脱脂奶粉、甘油、丙烯、二醇、水、乙醇等。所述组合物还可含有pH缓冲剂和湿润或乳化剂。

[0448] 对于口服施用,所述药物组合物可采用通过常规手段制备的片剂或胶囊的形式。所述组合物还可制备为液体,例如糖浆或悬浮液。所述液体可包括悬浮剂(例如,山梨醇糖浆、纤维素衍生物或氢化食用脂肪)、乳化剂(卵磷脂或阿拉伯树胶)、非水性媒剂(例如,杏仁油、油性酯、乙醇或分馏植物油)和防腐剂(例如,对羟基苯甲酸甲酯或对羟基苯甲酸丙酯或山梨酸)。所述制剂还可包括调味剂、着色剂和甜味剂。或者,所述组合物可呈用于以水或另一合适的媒剂复原的干燥产品形式提供。

[0449] 对于经颊施用,所述组合物可根据常规方案采用片剂或口含锭的形式。

[0450] 对于通过吸入施用,根据本公开使用的化合物便利地以具有或不具有赋形剂的雾化气雾剂的形式或以来自加压包装或喷雾器的气雾剂喷雾的形式递送,任选地具有推进剂,例如二氯二氟甲烷、三氯氟甲烷、二氯四氟甲烷、二氧化钛或其他合适气体。在加压气雾剂的情况下,剂量单位可通过提供阀以递送计量的量来测定。例如用于吸入器或吹入器中的明胶的胶囊和筒可配制成含有所述化合物和合适的粉末基质(如乳糖或淀粉)的粉末混

合物。

[0451] 所述药物组合物还可配制用于以栓剂或保留灌肠形式直肠施用,例如含有常规栓剂基质,如可可脂或其他甘油酯。

[0452] 在一个实施方案中,药物组合物包含具有因子VIII活性的多肽、编码所述具有因子VIII活性的多肽的优化的核酸分子、包含所述核酸分子的载体或包含所述载体的宿主细胞,和医药学上可接受的载剂。在一些实施方案中,所述组合物通过选自以下组成的组的途径施用:表面施用、眼内施用、肠胃外施用、鞘内施用、硬膜下施用和口服施用。所述肠胃外施用可以是静脉内或皮下施用。

[0453] 在其他实施方案中,所述组合物用于治疗有需要的受试者中的出血疾病或病况。所述出血疾病或病况选自以下组成的组:出血性凝固病症、关节积血、肌肉出血、口腔出血(oral bleed)、出血、出血至肌肉中、口腔出血(oral hemorrhage)、创伤、创伤性头皮炎、胃肠出血、颅内出血、腹腔内出血、胸腔内出血、骨折、中枢神经系统出血、咽后间隙出血、腹膜后间隙出血、髂腰肌鞘出血和其任何组合。在其他实施方案中,预定所述受试者经历手术。在其他实施方案中,所述治疗是预防性的或按需要的。

[0454] 治疗方法

[0455] 本公开提供一种治疗出血病症的方法,所述方法包括向有需要的受试者施用本公开的核酸分子、载体或多肽。在一些实施方案中,所述出血病症的特征在于FVIII缺陷。在一些实施方案中,所述出血病症是血友病。在一些实施方案中,所述出血病症是血友病A。在治疗出血病症的方法的一些实施方案中,在施用后24小时的血浆FVIII活性相对于施用包含SEQ ID NO:16的参考核酸分子、包含所述参考核酸分子的载体或由所述参考核酸分子编码的多肽的受试者增加。

[0456] 在一些实施方案中,血浆FVIII活性在施用后约6小时、约12小时、约18小时、约24小时、约36小时、约48小时、约3天、约4天、约5天、约6天、约7天、约8天、约9天、约10天、约11天、约12天、约13天、约14天、约15天、约16天、约17天、约18天、约19天、约20天、约21天、约22天、约23天、约24天、约25天、约26天、约27天或约28天相对于施用包含SEQ ID NO:16的参考核酸分子、包含所述参考核酸分子的载体或由所述参考核酸分子编码的多肽的受试者增加。在某些实施方案中,血浆FVIII活性在施用后约24小时相对于施用包含SEQ ID NO:16的参考核酸分子、包含所述参考核酸分子的病毒载体或由所述参考核酸分子编码的多肽的受试者增加。在另一实施方案中,血浆FVIII活性在施用后约21天相对于施用包含SEQ ID NO:16的参考核酸分子、包含所述参考核酸分子的病毒载体或由所述参考核酸分子编码的多肽的受试者增加。

[0457] 在一些实施方案中,施用后的血浆FVIII活性相对于施用包含SEQ ID NO:16的参考核酸分子、包含所述参考核酸分子的病毒载体或由所述参考核酸分子编码的多肽的受试者增加至少约1.5倍、至少约2倍、至少约3倍、至少约4倍、至少约5倍、至少约6倍、至少约7倍、至少约8倍、至少约9倍、至少约10倍、至少约11倍、至少约12倍、至少约13倍、至少约14倍、至少约15倍、至少约20倍、至少约25倍、至少约30倍、至少约35倍、至少约40倍、至少约50倍、至少约60倍、至少约70倍、至少约80倍、至少约90倍、至少约100倍、至少约150倍、至少约200倍、至少约250倍、至少约300倍、至少约350倍、至少约400倍、至少约450倍或至少约500倍。在一些实施方案中,施用后的血浆FVIII活性相对于生理学正常循环FVIII水平增加至

少约150%、至少约200%、至少约250%、至少约300%、至少约350%、至少约400%、至少约450%、至少约500%、至少约550%、至少约600%、至少约650%、至少约700%、至少约750%、至少约800%、至少约850%、至少约900%、至少约950%、至少约1000%、至少约1500%、至少约2000%、至少约2500%、至少约3000%、至少约3500%、至少约4000%、至少约4500%、至少约5000%、至少约5500%、至少约6000%、至少约7000%、至少约8000%、至少约9000%、至少约10,000%。在一个实施方案中,施用后的血浆FVIII活性相对于生理学正常循环FVIII水平增加至少约3000至约5000%。在一些实施方案中,在施用包含编码本文所述的具有因子VIII (FVIII) 活性的多肽的密码子优化的基因的质粒后24小时,或在施用包含编码本文所述的具有因子VIII (FVIII) 活性的多肽的密码子优化的基因的慢病毒或AAV载体后21天,血浆FVIII活性相对于施用包含SEQ ID NO:16的参考核酸分子、包含所述参考核酸分子的载体或由所述参考核酸分子编码的多肽的受试者增加至少约6倍。在一些实施方案中,在施用包含编码本文所述的具有因子VIII (FVIII) 活性的多肽的密码子优化的基因的质粒后24小时,或在施用包含编码本文所述的具有因子VIII (FVIII) 活性的多肽的密码子优化的基因的慢病毒或AAV载体后21天,血浆FVIII活性相对于施用包含SEQ ID NO:16的参考核酸分子、包含所述参考核酸分子的载体或由所述参考核酸分子编码的多肽的受试者增加至少约10倍。在一些实施方案中,在施用包含编码本文所述的具有因子VIII (FVIII) 活性的多肽的密码子优化的基因的质粒后24小时,或在施用包含编码本文所述的具有因子VIII (FVIII) 活性的多肽的密码子优化的基因的慢病毒或AAV载体后21天,血浆FVIII活性相对于施用包含SEQ ID NO:16的参考核酸分子、包含所述参考核酸分子的载体或由所述参考核酸分子编码的多肽的受试者增加至少约23倍。在一些实施方案中,在施用包含编码本文所述的具有因子VIII (FVIII) 活性的多肽的密码子优化的基因的质粒后24小时,或在施用包含编码本文所述的具有因子VIII (FVIII) 活性的多肽的密码子优化的基因的慢病毒或AAV载体后21天,血浆FVIII活性相对于施用包含SEQ ID NO:16的参考核酸分子、包含所述参考核酸分子的载体或由所述参考核酸分子编码的多肽的受试者增加至少约18倍。在一些实施方案中,在施用包含编码本文所述的具有因子VIII (FVIII) 活性的多肽的密码子优化的基因的质粒后24小时,或在施用包含编码本文所述的具有因子VIII (FVIII) 活性的多肽的密码子优化的基因的慢病毒或AAV载体后21天,血浆FVIII活性相对于施用包含SEQ ID NO:16的参考核酸分子、包含所述参考核酸分子的载体或由所述参考核酸分子编码的多肽的受试者增加至少约30倍。在一些实施方案中,在施用包含编码本文所述的具有因子VIII (FVIII) 活性的多肽的密码子优化的基因的质粒后24小时,或在施用包含编码本文所述的具有因子VIII (FVIII) 活性的多肽的密码子优化的基因的慢病毒或AAV载体后21天,血浆FVIII活性相对于施用包含SEQ ID NO:16的参考核酸分子、包含所述参考核酸分子的载体或由所述参考核酸分子编码的多肽的受试者增加至少约50倍。在一些实施方案中,在施用包含编码本文所述的具有因子VIII (FVIII) 活性的多肽的密码子优化的基因的质粒后24小时,或在施用包含编码本文所述的具有因子VIII (FVIII) 活性的多肽的密码子优化的基因的慢病毒或AAV载体后21天,血浆FVIII活性相对于施用包含SEQ ID NO:16的参考核酸分子、包含所述参考核酸分子的载体或由所述参考核酸分子编码的多肽的受试者增加至少约100倍。

[0458] 本公开还涉及一种治疗、改善或预防受试者的止血病症的方法,所述方法包括施

用治疗有效量的本公开的分离的核酸分子或由本公开的核酸分子编码的具有FVIII活性的多肽。由所述分离的核酸分子或所述编码的多肽实现的治疗、改善和预防可以是旁路疗法。接受旁路疗法的受试者可已经发展凝固因子(例如,FVIII)的抑制剂,或是正在发展凝固因子抑制剂的受试者。

[0459] 本公开的核酸分子、载体或多肽通过促进纤维蛋白凝块的形成来治疗或预防止血病症。由本公开的核酸分子编码的具有FVIII活性的多肽可活化凝固级联的成员。所述凝固因子可以是外在途径、内在途径或两者的参与者。

[0460] 本公开的核酸分子、载体或多肽可用于治疗已知可用FVIII治疗的止血病症。可使用本公开的方法治疗的止血病症包括但不限于血友病A、血友病B、血管假性血友病、因子XI缺陷(PTA缺陷)、因子XII缺陷以及纤维蛋白原、凝血酶原、因子V、因子VII、因子X或因子XIII的缺陷或结构异常、关节积血、肌肉出血、口腔出血(oral bleed)、出血、出血至肌肉中、口腔出血(oral hemorrhage)、创伤、创伤性头皮炎、胃肠出血、颅内出血、腹腔内出血、胸腔内出血、骨折、中枢神经系统出血、咽后间隙出血、腹膜后间隙出血和髂腰肌鞘出血。用于施用于受试者的组合物包括包含编码FVIII凝固因子(用于基因疗法应用)以及FVIII多肽分子的本公开的优化的核苷酸序列的核酸分子。

[0461] 在一些实施方案中,所述止血病症是遗传病症。在一个实施方案中,所述受试者具有血友病A。在其他实施方案中,所述止血病症是FVIII缺陷的结果。在其他实施方案中,所述止血病症可以是缺陷型FVIII凝固因子的结果。

[0462] 在另一实施方案中,所述止血病症可以是获得性病症。所述获得性病症可由于潜在的继发疾病或病况而引起。所述无关病况可以是例如但不限于癌症、自身免疫疾病或怀孕。所述获得性病症可由于老年或由于治疗潜在的继发病症的药物治疗(例如,癌症化学疗法)而引起。

[0463] 本公开还涉及一种治疗未患有止血病症或导致获得止血病症的继发疾病或病况的受试者的方法。本公开因此涉及一种治疗需要一般止血剂的受试者的方法,所述方法包括施用治疗有效量的本公开的分离的核酸分子、载体或FVIII多肽。例如,在一个实施方案中,需要一般止血剂的受试者正在经历或将要经历手术。本公开的分离的核酸分子、载体或FVIII多肽可在手术之前或之后作为预防药施用。本公开的分离的核酸分子、载体或FVIII多肽可在手术期间或之后施用以控制急性出血事件。所述手术可包括但不限于肝移植、肝切除或干细胞移植。

[0464] 在另一实施方案中,本公开的分离的核酸分子、载体或FVIII多肽可用于治疗患有有止血病症但具有急性出血事件的受试者。所述急性出血事件可由于严重创伤,例如手术、车祸、伤口、撕裂性枪击或导致不受控出血的任何其他创伤性事件而引起。

[0465] 所述分离的核酸分子、载体或FVIII蛋白可用于预防性治疗具有止血病症的受试者。所述分离的核酸分子、载体或FVIII蛋白可用于治疗具有止血病症的受试者中的急性出血事件。

[0466] 在另一实施方案中,通过施用本公开的分离的核酸分子或载体实现的FVIII蛋白的表达不会在受试者中诱导免疫反应。在一些实施方案中,所述免疫反应包含针对FVIII的抗体的发展。在一些实施方案中,所述免疫反应包含细胞因子分泌。在一些实施方案中,所述免疫反应包含B细胞、T细胞或B细胞和T细胞两者的活化。在一些实施方案中,所述免疫反

应是抑制免疫反应,其中相对于尚未发展免疫反应的受试者中的FVIII活性,所述受试者中的免疫反应会降低FVIII蛋白的活性。在某些实施方案中,通过施用本公开的分离的核酸分子或载体实现的FVIII蛋白的表达会防止针对由所述分离的核酸分子或所述载体表达的FVIII蛋白或FVIII蛋白的抑制免疫反应。

[0467] 在一些实施方案中,本公开的分离的核酸分子、载体或FVIII蛋白组合物与至少一种促进止血的其他药剂组合施用。所述促进止血的其他药剂是具有经过证明的凝固活性的治疗剂。所述止血剂可包括例如但不限于因子V、因子VII、因子IX、因子X、因子XI、因子XII、因子XIII、凝血酶原或纤维蛋白原或前述任一者的活化形式。所述凝固因子或止血剂也可包括抗纤维蛋白溶解药物,例如 ϵ -氨基-己酸、氨甲环酸。

[0468] 在本公开的一个实施方案中,所述组合物(例如,所述分离的核酸分子、载体或FVIII多肽)是如下组合物,其中当施用于受试者时,FVIII以可活化形式存在。所述可活化分子可在施用于受试者之后在凝固位点处体内活化。

[0469] 所述分离的核酸分子、载体或FVIII多肽可静脉内、皮下、肌肉内或经由任何粘膜表面,例如口服、舌下、经颊、舌下、经鼻、经直肠、经阴道或经由肺途径施用。所述FVIII蛋白可植入生物聚合物固体支撑物内或连接至生物聚合物固体支撑物,所述生物聚合物固体支撑物允许所述嵌合蛋白缓慢释放至所需位点。

[0470] 关于口服施用,所述药物组合物可采用通过常规手段制备的片剂或胶囊的形式。所述组合物还可制备为液体,例如糖浆或悬浮液。所述液体可包括悬浮剂(例如,山梨醇糖浆、纤维素衍生物或氢化食用脂肪)、乳化剂(卵磷脂或阿拉伯树胶)、非水性媒剂(例如,杏仁油、油性酯、乙醇或分馏植物油)和防腐剂(例如,对羟基苯甲酸甲酯或对羟基苯甲酸丙酯或山梨酸)。所述制剂还可包括调味剂、着色剂和甜味剂。或者,所述组合物可呈用于以水或另一合适媒剂复原的干燥产品形式提供。

[0471] 关于经颊和舌下施用,所述组合物可根据常规方案采用片剂、口含锭或快速溶解膜的形式。

[0472] 关于通过吸入施用,根据本公开使用的具有FVIII活性的多肽便利地以来自加压包装或喷雾器的气雾剂喷雾(例如,在PBS中)的形式递送,利用合适的推进剂,例如二氯二氟甲烷、三氯氟甲烷、二氯四氟甲烷、二氧化碳或其他合适气体。在加压气雾剂的情况下,剂量单位可通过提供阀以递送计量的量来测定。例如用于吸入器或吹入器中的明胶的胶囊和筒可配制成含有所述化合物和合适的粉末基质(如乳糖或淀粉)的粉末混合物。

[0473] 在一个实施方案中,所述分离的核酸分子、载体或FVIII多肽的施用途径是肠胃外。如本文所用,术语肠胃外包括静脉内、动脉内、腹膜内、肌肉内、皮下、直肠或阴道施用。肠胃外施用的静脉内形式是优选的。虽然清楚地预期所有这些施用形式均在本公开的范围内,但一种施用形式将是用于注射、特别是用于静脉内或动脉内注射或滴注的溶液。通常,用于注射的合适药物组合物可包含缓冲剂(例如,乙酸盐、磷酸盐或柠檬酸盐缓冲液)、表面活性剂(例如,聚山梨醇酯)、任选地稳定剂(例如,人白蛋白)等。然而,在与本文教导可相容的其他方法中,所述分离的核酸分子、载体或FVIII多肽可直接地递送至不良细胞群体的位点,由此增加患病组织的治疗剂暴露。

[0474] 用于肠胃外施用的制剂包括无菌水溶液或非水溶液、悬浮液和乳液。非水性溶剂的实例是丙二醇、聚乙二醇、植物油(如橄榄油)和可注射有机酯(如油酸乙酯)。水性载体包

括水、醇/水溶液、乳液或悬浮液,包括生理盐水和缓冲介质。在本公开中,药学上可接受的载剂包括但不限于0.01-0.1M和优选地0.05M磷酸盐缓冲液或0.8%生理盐水。其他常见肠胃外媒剂包括磷酸钠溶液、林格氏右旋糖、右旋糖和氯化钠、乳酸林格氏或固定油。静脉内媒剂包括流体和营养补充剂、电解质补充剂(如基于林格氏右旋糖的那些)等。防腐剂和其他添加剂也可存在,例如抗微生物剂、抗氧化剂、螯合剂和惰性气体等。

[0475] 更特定地,适用于可注射用途的药物组合物包括无菌水溶液(其中可溶于水)或分散液和用于无菌可注射溶液或分散液的临时制备的无菌粉末。在所述情况下,所述组合物必须是无菌的并且应具流动性至存在可容易注射性的程度。其应在制造和存储条件下稳定且将优选地针对如细菌和真菌等微生物的污染作用进行防腐处理。载体可以是含有例如水、乙醇、多元醇(例如,甘油、丙二醇和液体聚乙二醇等)和其合适混合物的溶剂或分散介质。适当流动性可例如通过使用涂层(如卵磷脂),在分散液的情况下通过维持所需粒径和通过使用表面活性剂来维持。

[0476] 微生物作用的预防可通过多种抗菌剂和抗真菌剂,例如对羟基苯甲酸酯、氯丁醇、苯酚、抗坏血酸、硫柳汞等实现。在多种情况下,在所述组合物中包括例如糖、多元醇(如甘露醇、山梨糖醇)或氯化钠的等渗剂将是优选的。可注射组合物的延长吸收可通过在所述组合物中包括延迟吸收的剂(例如单硬脂酸铝和明胶)来引起。

[0477] 在任何情况下,无菌可注射溶液可通过在有需要时在适当溶剂中并入所需量的活性化合物(例如,单独或与其他活性剂组合的多肽)与一种本文所列举的成分或成分的组合、随后过滤灭菌来制备。一般地,分散液通过将活性化合物并入无菌媒剂中来制备,所述无菌媒剂含有基础分散介质和来自上文所列举的那些的所需其他成分。在用于制备无菌可注射溶液的无菌粉末的情况下,优选制备方法是真空干燥和冷冻干燥,其会生成活性成分加上来自其先前无菌过滤的溶液的任何另外所需成分的粉末。根据本领域中已知的方法,用于注射的制剂经过加工,填充至如安瓿、袋、瓶、注射器或小瓶等容器中,并且在无菌条件下密封。此外,所述制剂可以试剂盒的形式包装并且出售。所述制品将优选地具有标签或包装插页,指示出相关组合物适用于治疗患有或倾向于患有凝固病症的受试者。

[0478] 所述药物组合物还可配制用于以栓剂或保留灌肠形式直肠施用,例如含有常规栓剂基质,如可可脂或其他甘油酯。

[0479] 用于病况的治疗的本公开组合物的有效剂量视多种不同因素而变化,所述因素包括施用方式、靶标位点、患者的生理学状态、患者是人或动物、所施用的其他药物和治疗是预防性的或治疗性的。通常,患者是人,但包括转基因动物在内的非人哺乳动物也可进行治疗。治疗剂量可使用本领域技术人员已知的常规方法来滴定以优化安全性和功效。

[0480] 关于施用包含编码如本文所述的具有因子VIII (FVIII) 活性的多肽的密码子优化的基因的质粒或由所述密码子优化的基因编码的多肽的剂量可介于1000ug/kg至0.1ng/kg体重范围内。在一个实施方案中,给药范围是1ug/kg至100ug/kg。关于施用包含编码如本文所述的具有因子VIII (FVIII) 活性的多肽的密码子优化的基因的慢病毒载体的剂量可介于 10^3 至 10^{15} TU/kg范围内。关于施用包含编码如本文所述的具有因子VIII (FVIII) 活性的多肽的密码子优化的基因的AAV载体的剂量可介于 10^5 至 10^{18} VG/kg范围内。

[0481] 所述分离的核酸分子、质粒、载体或FVIII多肽可作为单一剂量或作为多个剂量施用,其中所述多个剂量可连续地或以特定定时的时间间隔施用。体外测定可用于测定关于

施用的最佳剂量范围和/或方案。本领域中已知测量凝固因子活性的体外测定。另外,有效剂量可从获自动物模型,例如血友病狗(Mount等人2002,Blood 99(8):2670)的剂量-反应曲线推断。

[0482] 在上述范围中间的剂量也预期在本公开的范围。受试者可每天、隔天、每周或根据通过经验分析确定的任何其他方案施用所述剂量。示例性治疗需要在例如至少六个月的一段延长时期内以多个剂量施用。在一些方法中,可同时施用两种或更多种多肽,在所述情况下所施用的各多肽的剂量属于所指示的范围内。

[0483] 本公开的分离的核酸分子、载体或FVIII多肽可在多个场合施用。单一剂量之间的时间间隔可为每天、每周、每个月或每年。时间间隔也可以是不规则的,如通过测量患者中经过修饰的多肽或抗原的血液水平所指示。或者,多肽可以持续释放制剂形式施用,在所述情况下需要不太频繁施用。剂量和频率视所述多肽或多核苷酸在患者中的半衰期而变化。

[0484] 施用的剂量和频率可视治疗是预防性的或治疗性的而变化。在预防性应用中,含有所述分离的核酸分子、载体或FVIII多肽或其混合物的组合物施用于尚未呈疾病状态的患者以增强所述患者的抗性 or 最小化疾病的效应。所述量被定义为“预防有效剂量”。在很长一段时期内,相对低剂量以相对不频繁时间间隔施用。一些患者在其余生中继续接受治疗。

[0485] 本公开的分离的核酸分子、载体或FVIII多肽可任选地与有效治疗需要治疗(例如,预防性或治疗性)的病症或病况的其他剂组合施用。

[0486] 如本文所用,本公开的分离的核酸分子、载体或FVIII多肽与辅助疗法结合或组合施用意指所述疗法和所公开的多肽的连续、同时、同延、并发、伴随或同时期施用或应用。本领域技术人员应理解,所述组合治疗方案的多种组分的施用或应用可定时以增强所述治疗的总体有效性。熟练技术人员(例如,医师)将基于所选择的辅助疗法和本说明书的教导容易地能够辨别有效的组合治疗方案而无不当实验。

[0487] 应进一步理解,本公开的分离的核酸分子、载体或FVIII多肽可与一种或多种剂结合或组合使用(例如,以提供组合治疗方案)。可与本公开的多肽或多核苷酸组合的示例性剂包括代表针对所治疗的特定病症的现行护理标准的剂。所述剂本质上可以是化学的或生物的。术语“生物”或“生物剂”是指由活的生物体和/或其产品制得的任何药物活性剂,所述药物活性剂预期用作治疗剂。

[0488] 欲与本公开的多核苷酸或多肽组合使用的剂的量可随受试者而变化或可根据本领域中已知来施用。参见例如Bruce A Chabner等人,Antineoplastic Agents,GOODMAN&GILMAN'S THE PHARMACOLOGICAL BASIS OF THERAPEUTICS 1233-1287((Joel G.Hardman等人编,第9版1996)。在另一实施方案中,施用与所述护理标准一致的量的所述剂。

[0489] 如先前所论述,本公开的多核苷酸和多肽可以针对凝固病症的体内治疗的药物有效量施用。在这方面,应理解本公开的多肽或多核苷酸可经过配制以促进活性剂的施用并且促进活性剂的稳定性。优选地,根据本公开的药物组合物包含药学上可接受、无毒、无菌载体,如生理盐水、无毒缓冲液、防腐剂等。当然,本公开的药物组合物可以单一或多个剂量施用以提供药物有效量的所述多肽。

[0490] 多种测试可用于评估凝固系统的功能:活化部分凝血酶原时间(aPTT)测试、发色测定、ROTEM[®]测定、凝血酶原时间(PT)测试(还用于测定INR)、纤维蛋白原测试(通常通

过Clauss方法)、血小板计数、血小板功能测试(通常通过PFA-100)、TCT、出血时间、混合测试(异常性纠正患者的血浆是否与正常血浆混合)、凝固因子测定、抗磷脂抗体、D-二聚体、基因测试(例如,因子V Leiden、凝血酶原突变G20210A)、稀释印度蝰蛇毒液时间(dRVVT)、杂项血小板功能测试、血栓弹力描记术(TEG或Sonoclot)、血栓弹力图试验(TEM[®],例如ROTEM[®])或优球蛋白溶解时间(ELT)。

[0491] 所述aPTT测试是量度“内在”(还称作接触活化途径)和共同凝固途径两者的功效的性能指示剂。这种测试通常用于测量市面上有售的重组凝固因子(例如,FVIII或FIX)的凝固活性。其与凝血酶原时间(PT)结合使用,所述凝血酶原时间(PT)测量外在途径。

[0492] ROTEM[®]分析提供关于止血法的整体动力学的信息:凝固时间、凝块形成、凝块稳定性和溶解。血栓弹力图试验中的不同参数取决于血浆凝固系统的活性、血小板功能、纤维蛋白溶解或影响这些相互作用的多种因素。这种测定可提供二期止血的完整视图。

[0493] 基因疗法

[0494] 本公开提供一种增加具有FVIII活性的多肽在受试者中的表达的方法,所述方法包括向有需要的受试者施用本公开的分离的核酸分子,其中所述多肽的表达相对于包含SEQ ID NO:16的参考核酸分子增加。本公开还提供一种增加具有FVIII活性的多肽在受试者中的表达的方法,所述方法包括向有需要的受试者施用本公开的载体,其中所述多肽的表达相对于包含参考核酸分子的载体增加。

[0495] 已经探索体细胞基因疗法作为血友病A的可能治疗。基因疗法是血友病的特别引人注意的治疗,因为其可能通过在载体的单一施用之后连续内源产生FVIII来治愈所述疾病。血友病A充分适用于基因置换方法,因为其临床显现完全可归因于以微量(200ng/ml)在血浆中循环的单一基因产物(FVIII)的缺乏。

[0496] 本公开的FVIII蛋白可在哺乳动物(例如,人患者)中体内产生,使用基因疗法方法来治疗选自以下组成的组的出血疾病或病症将是治疗上有益的:出血性凝固病症、关节积血、肌肉出血、口腔出血(oral bleed)、出血、出血至肌肉中、口腔出血(oral hemorrhage)、创伤、创伤性头皮炎、胃肠出血、颅内出血、腹腔内出血、胸腔内出血、骨折、中枢神经系统出血、咽后间隙出血、腹膜后间隙出血和髂腰肌鞘出血。在一个实施方案中,所述出血疾病或病症是血友病。在另一实施方案中,所述出血疾病或病症是血友病A。这涉及施用可操作性连接至合适的表达控制序列的优化的FVIII编码核酸。在某些实施方案中,这些序列并入至病毒载体中。用于所述基因疗法的合适病毒载体包括腺病毒载体、慢病毒载体、杆状病毒载体、Epstein Barr病毒载体、乳多空病毒载体、牛痘病毒载体、单纯疱疹病毒载体和腺相关病毒(AAV)载体。所述病毒载体可以是复制缺陷型病毒载体。在其他实施方案中,腺病毒载体在其E1基因或E3基因中具有缺失。在其他实施方案中,所述序列并入至本领域技术人员已知的非病毒载体中。

[0497] 本文所述的多个方面、实施方案和选项均可以任何和所有变化组合。

[0498] 本说明书中提及的所有公布、专利和专利申请均以引用的方式并入,其并入程度就如同各个别公布、专利或专利申请特定地并且个别地经指示以引用的方式并入一般。

[0499] 已经一般地描述了本公开,可通过参考本文所提供的实施例获得进一步理解。这些实施例仅出于说明目的并且不意图具限制性。

[0500] 实施例

[0501] 实施例1.密码子优化策略

[0502] 通过控制密码子使用偏好产生了八种密码子优化的BDD FVIII变体,包括coFVIII-3 (SEQ ID NO:1;图1A)、coFVIII-4 (SEQ ID NO:2;图1B)、coFVIII-5 (SEQ ID NO:70;图1C)、coFVIII-6 (SEQ ID NO:71;图1D)、coFVIII-52 (SEQ ID NO:3;图1E)、coFVIII-62 (SEQ ID NO:4;图1F)、coFVIII-25 (SEQ ID NO:5;图1G)和coFVIII-26 (SEQ ID NO:6;图1H)。使用在线工具Eugene来促进密码子优化,如先前所述(参见Gaspar等人,“EuGene: maximizing synthetic gene design for heterologous expression,” Bioinformatics28:2683-84 (2012)),并且监测数种密码子使用参数,如密码子适应指数(CAI)和相对同义密码子使用(RSCU)(表5)。所有变体均调整至CAI $\geq$ 83%并且RSCU $\geq$ 1.63,而亲本B结构域缺失的FVIII序列在优化之前的CAI为74%并且RSCU为1.12(表5)。

[0503] 表5:密码子优化参数

	亲本 BDD FVIII	coF VIII -3	coF VIII- 4	coFVI II-5	coF VIII- 6	coF VIII- 52	coF VIII- 62	coF VIII- 25	coF VIII- 26
[0504] 密码子适应 指数(CAI; %)	74	91	97	83	83	91	91	88	88
最优密码子 的 频 率 (FOP)	39	65	92	64	64	79	79	74	75
GC 含量(%)	44.10	52.1 0	60.8 0	55.7	55.9 %	58.3 0	58.3 0	57.3 0	57.6 0
相对同义密 码 子 使 用 (RSCU)	1.12	2.32	2.72	1.63	1.63	2.22	2.19	2.04	2.58
密码子对偏 好	0.19	0.43	0.04	0.11	0.11	0.27	0.27	0.23	0.48
[0505] 密码子的有 效数目	54.2	25.6	22.8	39.7	39.1	30.9	31.4	34.1	26.7

[0506] 相对于非优化的BDD FVIII序列(图2A),除了所述CAI的总体增加以外,所述八种变体还基于如图2中所说明的CAI在编码区中的分布被设计成三个类别。第一个类别包含在整个编码区中具有高CAI的平均分布的BDD FVIII变体(参见图2C至图2F)。第一个类别包括coFVIII-3(图2C)、coFVIII-4(图2D)、coFVIII-5(图2E)、coFVIII-6(图2F),以及先前描述的coFVIII-1(参见国际公布号WO 2014/127215 (SEQ ID NO:1))(图2B)。第二个类别包含在所述编码序列的N端半段处具有较低CAI并且在所述编码序列的C端半段处具有较高CAI的BDD-FVIII变体(参见图2G和图2H)。第二个类别包括coFVIII-52(图2G)和coFVIII-62(图2H)。第三个类别包含在所述编码序列的N端半段处具有较高CAI并且在所述编码序列的C端半段处具有较低CAI的BDD FVIII变体(参见图2I和图2J)。第三个类别包括coFVIII-25(图2I)和coFVIII-26(图2J)。

[0507] 不受任何理论束缚,推测较高CAI可能与较快蛋白质翻译有关,并且这三种类别可能代表从开始至结束的不同蛋白质合成速率。例如,具有较低CAI的区的翻译可能相对于具

有较高CAI的区的翻译缓慢地进行。假如这样的话,例如具有较低CAI的coFVIII-52或coFVIII-62的N端半段的翻译可能最初缓慢地进行,随后具有较高CAI的C端半段更快翻译。这可能优选用于翻译期间的蛋白质折叠和翻译后修饰,而不会减慢总体蛋白质合成。关于coFVIII-25和coFVIII-26变体可见相反效应,所述变体在N端半段处具有较高CAI并且在C端半段处具有较低CAI。

[0508] 为了确保mRNA的稳定性,所有FVIII密码子优化的变体均经过调整以避免包括隐蔽剪接位点、过早polyA位点、RNA不稳定基序(ARE)和重复序列在内的多个位点,和调整GC含量(参见表2)。

[0509] 实施例2. 从pcDNA3质粒克隆和表达coFVIII变体

[0510] 针对体内表达设计含有多种FVIII变体的表达质粒。将非优化BDD FVIII(图1I; SEQ ID NO:16)和coFVIII-1(图11Z; SEQ ID NO:68)多核苷酸克隆至pcDNA3骨架(Invitrogen)中,其中CMV启动子由ET启动子置换(参见图3)。所得质粒FVIII-311(BDD FVIII)和FVIII-303(coFVIII-1)分别地驱动非优化BDD FVIII和coFVIII-1的表达。

[0511] 在Hem A小鼠中通过FVIII-303或FVIII-311的流体动力学注射(5 μ g DNA/小鼠)来评估FVIII-311和FVIII-303的体内表达。在注射后24、48和72小时收集血浆样品,并且通过FVIII特异性发色测定来测定FVIII活性。

[0512] 如图4所示,用FVIII-311(BDD FVIII; 正方形)治疗的小鼠的血浆FVIII活性在注射后72小时为74 $\pm$ 43mU/mL,而用FVIII-303(coFVIII-1; 圆形)治疗的小鼠的血浆FVIII活性在注射后72小时为452 $\pm$ 170mU/mL(图4)。这代表coFVIII-1的表达相对于非优化BDD FVIII增加大概六倍。

[0513] 实施例3. 使用慢病毒载体系统克隆和表达coFVIII变体

[0514] 为了进一步评估密码子优化的BDD FVIII变体的表达水平,在ET启动子的控制下将编码序列克隆至慢病毒质粒中(参见Amendola等人,“Coordinate dual-gene transgenesis by lentiviral vectors carrying synthetic bidirectional promoters,”*Nature Biol.*23:108-16(2005);国际公布号W0 2000/066759 A1)。pLV-coFVIII-52的质粒图显示于图5中;并且含有非优化BDD FVIII(LV-2116)、coFVIII-1(LV-coFVIII-1)、coFVIII-3(LV-coFVIII-3)、coFVIII-4(LV-coFVIII-4)、coFVIII-5(LV-coFVIII-5)和coFVIII-6(LV-coFVIII-6)、coFVIII-62(LV-coFVIII-62)、coFVIII-25(LV-coFVIII-25)和coFVIII-26(LV-coFVIII-26)的质粒以相同方式构建,除了coFVIII-52片段由每一个所指示的编码序列使用NheI和SalI位点置换(表6)。

[0515] 表6: 编码FVIII变体的表达质粒

质粒 ID	描述
FVIII-303 (coFVIII-1)	在 pcDNA3 中的 ET 启动子下的 coFVIII-1
FVIII-311 (BDD FVIII)	在 pcDNA3 中的 ET 启动子下的亲本 BDD-FVIII
LV-2116 (BDD FVIII)	在慢病毒质粒中的 ET 启动子下的亲本 BDD-FVIII
LV-coFVIII-1	在慢病毒质粒中的 ET 启动子下的 coFVIII-1
[0516] LV-coFVIII-3	在慢病毒质粒中的 ET 启动子下的 coFVIII-3
LV-coFVIII-4	在慢病毒质粒中的 ET 启动子下的 coFVIII-4
LV-coFVIII-5	在慢病毒质粒中的 ET 启动子下的 coFVIII-5
LV-coFVIII-6	在慢病毒质粒中的 ET 启动子下的 coFVIII-6
LV-coFVIII-52	在慢病毒质粒中的 ET 启动子下的 coFVIII-52
LV-coFVIII-62	在慢病毒质粒中的 ET 启动子下的 coFVIII-62
LV-coFVIII-25	在慢病毒质粒中的 ET 启动子下的 coFVIII-25
LV-coFVIII-26	在慢病毒质粒中的 ET 启动子下的 coFVIII-26

[0517] 在HemA小鼠中通过以5 μ g DNA/小鼠(图6A、图6B)或20 μ g DNA/小鼠(图6C)的剂量进行流体动力学注射来评估慢病毒密码子优化的FVIII变体。如图6所示,coFVIII-3(图6A;三角形)、coFVIII-4(图6A;倒三角形)、coFVIII-5(图6A;菱形)、coFVIII-6(图6A;空心圆形)、coFVIII-25(图6B;三角形)、coFVIII-26(图6B;倒三角形)、coFVIII-52(图6C;正方形)和coFVIII-62(图6C;实心圆形)中的每一者均展现高于coFVIII-1(图6A,圆形;图6B,圆形;和图6C,三角形)的FVIII活性。特别地,coFVIII-25和coFVIII-26在注射后72小时展现相似的表达水平,达到比coFVIII-1的活性高约3倍的活性(图6B),与非优化、亲本BDD FVIII相比,所述活性翻译成高24倍的FVIII活性(参见图4)。coFVIII-52(正方形)和coFVIII-62(实心圆形)均在注射后72小时实现甚至更高的表达,分别展现比coFVIII-1(三角形)高6倍和4倍的表达,并非分别展现比非优化、亲本BDD FVIII(空心圆形)高50倍和30倍的表达(图6C)。这些数据指示,在编码序列的N端半段处的较低CAI和在编码序列的C端半段处的较高CAI的组合如与CAI的反向分布相比可能更有益于FVIII表达。

[0518] 实施例4:HemA小鼠中密码子优化的FVIII变体的长期慢病毒表达

[0519] 针对通过慢病毒载体介导的基因转移实现的长期FVIII表达,评估经过鉴别驱动在流体动力学注射后72小时HemA小鼠中FVIII的高表达的变体。慢病毒载体通过短暂转染在293T细胞中产生并且通过超离心浓缩至约5E9TU/ml。所述慢病毒载体接着通过眶后注射以1E8TU/小鼠的剂量施用于12-14日龄的HemA小鼠。在慢病毒注射后21天,注射LV-2116(BDD FVIII;图7)的小鼠的平均血浆FVIII活性是约0.04IU/ml。在注射后21天,相对于LV-2116(非优化B结构域缺失的FVIII)对照,coFVIII-1、coFVIII-5、coFVIII-52、coFVIII-6和coFVIII-62中的每一者均导致较高循环FVIII水平。特别地,在注射后21天,coFVIII-1和coFVIII-5注射会产生约1.8IU/mL的FVIII血浆活性水平,coFVIII-52会产生约4.9IU/mL的FVIII血浆活性水平,coFVIII-6会产生约4.6IU/mL的FVIII血浆活性水平,并且coFVIII-62

会产生约2.5IU/mL的FVIII血浆活性水平(图7)。在注射LV-coFVIII-6和LV-coFVIII-52(分别为4.6IU/ml和4.9IU/ml)的小鼠中观察到的FVIII血浆水平比注射LV-2116(非优化BDD-FVIII)对照的小鼠中观察到的血浆水平高超过100倍。

[0520] 实施例5.coFVIII-XTEN融合构建体

[0521] 测试XTEN改进稳态FVIII表达的能力。首先,将144个氨基酸的XTEN的编码序列(“XTEN₁₄₄”;SEQ ID NO:18)分别插入coFVIII-52和coFVIII-1的核苷酸1193处(或在编码多肽的最初764个氨基酸之后)以产生coFVIII-52-XTEN(图8A;SEQ ID NO:19)和coFVIII-1-XTEN(图8B;SEQ ID NO:20)。所述coFVIII-1-XTEN序列接着在如上文所述的ET启动子的控制下克隆至pcDNA3骨架(Invitrogen)中,以产生FVIII-306表达质粒;并且所述coFVIII-52-XTEN序列在如上文所公开的ET启动子的控制下克隆至慢病毒质粒中,以产生pLV-coFVIII-52-XTEN(图9)。FVIII-306(coFVIII-1-XTEN)通过流体动力学注射以5μg DNA/小鼠施用于HemA小鼠。如与FVIII-303(coFVIII-1;图10A,小圆形)和FVIII-311(BDD FVIII;图10A,正方形)相比,XTEN₁₄₄融合至coFVIII-1(FVIII-306;图10A,大圆形)会在注射后72小时时在HemA小鼠中分别引起约5倍和33倍高的FVIII表达。XTEN插入对FVIII表达的影响也在HemA小鼠中使用慢病毒载体来评估(图10B)。LV-coFVIII-52-XTEN通过眶后注射以1E8TU/小鼠施用于12-14日龄的HemA小鼠。如与LV-coFVIII52和LV-2116(BDD-FVIII)相比,XTEN₁₄₄融合至coFVIII-52(图10B)会在注射后21天时HemA小鼠中分别引起约4倍和450倍高的FVIII表达。

[0522] 包含融合至XTEN₁₄₄并且融合至ET启动子的coFVIII-3、co-FVIII-4、coFVIII-5、coFVIII-6、coFVIII-62、coFVIII-25和coFVIII-26中的每一者的慢病毒载体将如上文所述进行制备。所述载体将针对其FVIII蛋白表达进行测试。

[0523] 实施例6.coFVIII构建体的表达

[0524] 通过标准分子克隆技术将密码子优化的FVIII变体克隆至如图9所说明的慢病毒质粒中。慢病毒载体接着通过短暂转染在HEK293细胞中产生并且通过超离心分离。

[0525] 将FVIII慢病毒载体通过静脉内注射以1.5E10TU/kg LV-FVIII变体的剂量施用于14日龄的HemA小鼠幼崽。在LV-FVIII治疗后第21天测量FVIII血浆活性,并且在LV-FVIII治疗后第150天在从LV-FVIII治疗的动物收集的肝尸检样品中测量每个细胞的载体拷贝数(VCN)。虽然VCN值在所有动物中均相似而与所施用的LV-FVIII变体无关(图12B),但用coFVIII变体治疗的动物中的FVIII活性水平比用wtBDD-FVIII治疗的动物中高30至100倍(图12A和图12C;表7)。这些数据指示,FVIII密码子优化会改进慢病毒载体设置中的FVIII表达。

[0526] 表7:密码子优化的FVIII构建体的相对表达

[0527]

LV-FVIII 变体	CoFVII I-1	CoFVIII -3	CoFVIII- 4	CoFVIII- 5	CoFVIII- 6
FVIII 表达相对于 LV-wtBDD-FVII I 的倍数改进	57	34	33	74	107
LV-FVIII 变体	CoFVII I-52	CoFVIII -62	CoFVIII- 25	CoFVIII- 26	
FVIII 表达相对于 LV-wtBDD-FVII I 的倍数改进	96	59	87	83	

[0528] 实施例7. 在慢病毒治疗后HemA小鼠中的FVIII转基因表达介导的免疫反应

[0529] 针对长期FVIII表达和抗FVIII抗体形成, 评估实施例6的LV-FVIII治疗的小鼠。如通过FVIII血浆活性证明的FVIII表达在同一治疗组内的动物之间变化(图13A)。例如, 用表达coFVIII-5变体的慢病毒载体治疗的三只小鼠(指定1、2和3)在大约16周内显示一致的FVIII表达, 而用相同慢病毒载体治疗的三只同窝出生小鼠(指定4、5和6)在治疗后约10周显示FVIII血浆活性水平的急剧下降(图13A)。在小鼠1、2和3中观察到的一致的FVIII血浆活性与不可检测或极低水平的抗FVIII抗体相关(图13B; 小鼠1、2和3)。相反, 展现FVIII血浆活性的急剧下降的小鼠还展现抗FVIII抗体的增加的水平(图13B; 小鼠4、5和6)。这些数据表明, FVIII转基因表达会在一小组动物中诱导抗FVIII抗体形成, 并且所得的抗FVIII抗体从循环中消除转基因FVIII蛋白。

[0530] 评估FVIII表达与抗FVIII抗体形成之间的关系。将实施例6的LV-FVIII治疗的小鼠分成两组: 呈抗FVIII抗体阴性的小鼠和呈抗FVIII抗体阳性的小鼠。如图14所示, 转基因FVIII在生理学水平下的表达不会诱导对转基因FVIII的免疫反应(图14, 圆形)然而, 超生理水平的FVIII表达看来诱导抗FVIII抗体形成, 以致FVIII表达水平越高, 抗FVIII抗体形成的机会越高。这些数据表明, 在经受FVIII基因疗法治疗的患者中维持FVIII表达的生理水平看来是有益的。

[0531] 为了确定FVIII表达诱导的免疫反应是否导致转基因表达肝细胞的损失, 在来自抗FVIII抗体阳性和阴性小鼠的肝尸检样品中评估载体拷贝数(图15)和FVIII RNA转录水平(图16)。如图15所示, 抗FVIII抗体阳性和阴性小鼠中的载体拷贝数的分布是相同的, 指示尽管发展抗FVIII抗体, 仍维持具有LV-FVIII整合的细胞。这表明LV-FVIII介导的FVIII转基因表达不会诱导针对FVIII表达肝细胞的细胞毒性T淋巴细胞(CTL)反应。为了进一步确认这些结果, 通过RNA原位杂交评估FVIII RNA转录(图16C和图16D)。在肝收集时, 小鼠coFVIII-52-B不具有可检测的循环FVIII和高水平的抗FVIII抗体(图16A和图16B)。然而, 来自coFVIII-52-B小鼠的肝组织中的RNA转录信号和FVIII-RNA阳性细胞的数目可与FVIII-52-A小鼠相当, 所述FVIII-52-A小鼠在尸检时具有约4IU/ml的循环FVIII。因此, FVIII表达不会在实验HemA小鼠中诱导CTL反应。

[0532] 实施例8. LV-FVIII治疗的HemA新生小鼠中的FVIII长期表达

[0533] 为了评估使用慢病毒系统通过靶向肝来治疗儿科HemA患者的功效, 通过太阳穴静

脉注射向2日龄的HemA小鼠施用约 1.5×10^{10} TU/kg的LV-coFVIII-52XTEN、LV-coFVIII6-XTEN或表达wtBDD-FVIII的慢病毒载体。对于变体和对照两者观察到一致的长期FVIII表达,证明了在经过治疗的小鼠的分裂的肝细胞中维持了整合的FVIII表达级联(图17)。这些数据表明,LV-FVIII可能潜在地用于治疗儿科和成年HemA患者。

[0534] 实施例9.HemA新生狗中的LV-FVIII的评估

[0535] 为了进一步评估LV-FVIII在较大动物模型中的功效,通过静脉内注射向两只一周龄的HemA新生狗(指定S3和K4)施用 1.3×10^9 TU/kg LV-coFVIII-6-XTEN。这一剂量比先前用于HemA小鼠模型中的剂量低超过10倍。在施用所述慢病毒载体之后,通过一期凝固测定(aPTT)监测血浆FVIII活性(图18)并且通过旋转式血栓弹力描记(ROTEM)测定监测全血止血(图19A至图19D)。在LV-FVIII治疗之前,针对S3的FVIII水平是正常值的0.7%(图18)。在慢病毒载体治疗后,S3的FVIII水平在第7天和第14天分别增加至79%和103%(图18)。关于K4,施用前FVIII水平是正常值的1.4%(图18)。在慢病毒载体治疗后,FVIII水平在第6天和第14天分别增加至22%和25%(图18)。

[0536] 与FVIII水平相关,在治疗后2周对于两只动物均观察到标准化ROTEM(图19A至图19C),证明由LV-FVIII介导的治疗益处。HemA狗中由LV-FVIII实现的治疗有益的FVIII表达水平确认了LV-FVIII用于血友病A的治疗的潜在用途。

[0537] 实施例10.HemA新生小鼠中的LV-FVIII的评估

[0538] 用于基因置换的使用慢病毒载体(LV)的离体基因疗法已经证明了针对多种适应症的临床功效并且在经过治疗的患者中的多年随访未显示肿瘤发生的迹象。LV-FIX的全身性递送会介导持续的FIX表达并且在血友病动物模型中经充分耐受。大的包装能力、经由基因整合维持长期转基因表达的能力、人群中已存在的抗LV抗体(ab)的缺乏和在临床前和临床情形中证明的鼓舞人心的体内概况使LV成为用于体内基因递送、尤其用于具有大cDNA尺寸的基因候选物(如FVIII)的有希望的媒介。

[0539] 为了评估LV-FVIII用于治疗血友病A(HemA)的潜在用途,使置于肝细胞特异性启动子下的密码子优化的人FVIII(hFVIII)变体成为LV系统的一部分,所述LV系统含有微RNA-142靶标序列的多个拷贝以最小化抗原递呈细胞中的FVIII表达并且降低诱导性抗FVIII抗体的概率。通过293T细胞的短暂转染产生LV-hFVIII载体,随后通过超离心进行1000倍浓缩并且在HemA小鼠模型中进行评估。在LV-hFVIII的静脉内施用后,通过FVIII活性和抗原测定监测循环hFVIII水平,通过经由定量PCR测量LV DNA拷贝并且经由原位杂交测量转基因RNA来评估肝中的LV转导效率,通过总抗hFVIII抗体ELISA测量抗hFVIII抗体。

[0540] 针对在新生阶段经过治疗的HemA小鼠中的所有LV-hFVIII变体,观察到持续的FVIII表达。在 1.5×10^{10} 转导单位/kg的剂量下,LV编码密码子优化的hFVIII(LV-cohFVIII)引起比LV编码野生型hFVIII高30倍至100倍的循环FVIII(图12C),而肝细胞中的载体拷贝数和FVIII RNA阳性细胞的百分比在所有测试的组中可相当(图12B)。密码子优化与XTEN的组合(LV-cohFVIII-XTEN)是一种非结构化亲水性多肽,据推测其通过增加有效负荷的流体动力学尺寸来改进循环半衰期,所述组合在血浆中产生30-50IU/mL FVIII活性,代表正常循环FVIII水平的3,000-5,000%(图12A、图17)。此外,仅在具有超生理水平的hFVIII的小鼠中检测到抗hFVIII ab(图14),而在抗hFVIII抗体阳性小鼠中未观察到针对LV诱导细胞的细胞毒性T淋巴细胞反应(图15和图16A至图16D)。我们的结果支持用于血友病A的体内基因

疗法的LV-FVIII的进一步发展。

[0541] 特定实施方案的前述描述将充分地揭露本公开的一般性质,使得其他人可通过应用在本领域的技术内的知识容易地针对多种应用来修改和/或改编所述特定实施方案,而无不当实验,不偏离本公开的一般概念。因此,基于本文所提供的教导和指导,所述改变和修改意图在所公开的实施方案的相等物的含义和范围内。应理解,本文中的措辞或术语是出于描述而非限制的目的,以致本说明书的术语或措辞欲由熟练技术人员根据所述教导和指导来解释。

[0542] 本公开的其他实施方案将由本领域技术人员通过考虑本文所公开的公开的说明书和实践显而易见。预期所述说明书和实施例仅被视为示例性的,本公开的真实范围和精神由以下权利要求指示。

[0543] 本文所引用的所有专利和公布均以引用的方式整体并入本文中。

[0544] 本申请要求2016年2月1日提交的美国临时申请号62/289,696和2016年10月18日提交的美国临时申请号62/409,739的优先权权益,所述美国临时申请以引用的方式整体并入本文中。

[0001]	序列表
[0002]	<110> 比奥贝拉蒂治疗公司(BIOVERATIV THERAPEUTICS INC.)
[0003]	<120> 优化的因子VIII基因
[0004]	<130> 2159.469PC02
[0005]	<150> 62/409,739
[0006]	<151> 2016-10-18
[0007]	<150> 62/289,696
[0008]	<151> 2016-02-01
[0009]	<160> 103
[0010]	<170> PatentIn 3.5版
[0011]	<210> 1
[0012]	<211> 4374
[0013]	<212> DNA
[0014]	<213> 人工序列
[0015]	<220>
[0016]	<223> coFVIII-3
[0017]	<400> 1
[0018]	atgcagattg agctgtccac ttgtttcttc ctgtgcctcc tgcgcttctg tttctccgcc 60
[0019]	actcgccggt actaccttg agccgtggag ctttcatggg actacatgca gagcgacctg 120
[0020]	ggcgaactcc ccgtggatgc cagattcccc ccccgctgc caaagtcctt cccctttaac 180
[0021]	acctccgtgg tgtacaagaa aacctcttt gtgcagttca ctgaccacct gtccaacatc 240
[0022]	gccaagccgc gccaccttg gatgggcctc ctgggaccga ccattcaagc tgaagtgtac 300
[0023]	gacaccgtgg tgatcacct gaagaacatg gcgtcccacc ccgtgtccct gcatgcggtc 360
[0024]	ggagtgtcct actggaaggc ctccgaagga gctgagtacg acgaccagac tagccagcgg 420
[0025]	gaaaaggagg acgataaagt gttcccgggc ggctcgcata cttacgtgtg gcaagtctg 480
[0026]	aaggaaaacg gacctatggc atccgatcct ctgtgcctga cttactccta cttttcccat 540
[0027]	gtggacctcg tgaaggacct gaacagcggg ctgattggtg cacttctcgt gtgccgcgaa 600
[0028]	ggttcgctcg ctaaggaaaa gaccagacc ctccataagt tcatcctttt gttcgctgtg 660
[0029]	ttcgatgaag gaaagtcatg gcattccgaa actaagaact cgctgatgca ggaccgggat 720
[0030]	gccgcctcag cccgcgctg gcctaaaatg catacagtca acgatacgt gaatcggtea 780
[0031]	ctgcccgggc tcatcggttg tcacagaaag tccgtgtact ggcacgtcat cggcatgggc 840
[0032]	actacgcctg aagtgcactc catcttctg gaagggcaca ctttctcgt gcgcaaccac 900
[0033]	cgccaggcct ctctggaaat ctccccgatt acctttctga ccgccagac tctgctcatg 960
[0034]	gacctggggc agttccttct cttctgccac atctccagcc atcagcacga cggaatggag 1020
[0035]	gcctacgtga aggtggactc atgcccggaa gaacctcagt tgcggatgaa gaacaacgag 1080
[0036]	gagccgagg actatgacga cgatttgact gactccgaga tggacgtcgt gcggttcgat 1140
[0037]	gacgacaaca gcccagctt catccagatt cgcagcgtgg ccaagaagca ccccaaac 1200
[0038]	tgggtgcaact acatcgcggc cgaggaagaa gattgggact acgccccgtt ggtgctggca 1260
[0039]	ccgatgacc ggtcgtacaa gtcccagtat ctgaacaatg gtccgcagcg gattggcaga 1320
[0040]	aagtacaaga aagtgcggtt catggcgtag actgacgaaa cgtttaagac ccgggaggcc 1380
[0041]	attcaacatg agagcggcat tctgggacca ctgctgtacg gagaggtcgg cgataccctg 1440

[0042]	ctcatcatct tcaaaaacca ggcctcccg cttacaaca tctaccctca cggaatcacc	1500
[0043]	gacgtgcggc cactctactc gcggcgcctg ccgaaggcg tcaagcacct gaaagacttc	1560
[0044]	cctatcctgc cgggcgaaat cttcaagtat aagtggaccg tcaccgtgga ggacgggccc	1620
[0045]	accaagagcg atcctaggtg tctgactcgg tactactcca gcttcgtgaa catggaacgg	1680
[0046]	gacctggcat cgggactcat tggaccgctg ctgatctgct acaaagagtc ggtggatcaa	1740
[0047]	cgcggcaacc agatcatgtc cgacaagcgc aacgtgatcc tgttctccgt gtttgatgaa	1800
[0048]	aacagatcct ggtacctcac tgaaaacatc cagaggttcc tcccaaacc cgcaggagt	1860
[0049]	caactggagg accctgagtt tcaggcctcg aatatcatgc actcgattaa cggttacgtg	1920
[0050]	ttcgactcgc tgcagctgag cgtgtgcctc catgaagtcg cttactggtg cattctgtcc	1980
[0051]	atcggcggcc agactgactt cctgagcgtg ttcttttccg gttacacct taagcacaag	2040
[0052]	atggtgtacg aagataccct gacctgttc cttttctccg gcgaaacggt gttcatgtcg	2100
[0053]	atggagaacc cgggtctgtg gattctggga tgccacaaca gcgactttcg gaaccgcgga	2160
[0054]	atgactgccc tgctgaaggt gtctcatgc gacaagaaca ccggagacta ctacgaggac	2220
[0055]	tcctacgagg atatctcagc ctacctctg tccaagaaca acgcgatcga gccgcgcagc	2280
[0056]	ttcagccaga accgcctgt gctgaagagg caccagcgag aaattaccg gaccaccctc	2340
[0057]	caatcggatc aggaggaaat cgactacgac gacaccatct cgggtgaaat gaagaaggaa	2400
[0058]	gatttcgata tctacgacga ggacgaaaat cagtcccctc gctcattcca aaagaaaact	2460
[0059]	agacactact ttatcgccgc ggtggaaaga ctgtgggact atggaatgc atccagccct	2520
[0060]	cacgtccttc ggaaccgggc ccagagcgga tcggtgcctc agttcaagaa agtgggtgtc	2580
[0061]	caggagtcca ccgacggcag cttcaccag ccgctgtacc ggggagaact gaacgaacac	2640
[0062]	ctgggcctgc tcggtcccta catccgcgcg gaagtggagg ataacatcat ggtgaccttc	2700
[0063]	cgtaaccaag catccagacc ttactccttc tattctccc tgatctcata cgaggaggac	2760
[0064]	cagcgccaag gcgccgagcc ccgcaagaac ttcgtcaagc ccaacgagac taagacctac	2820
[0065]	ttctggaagg tccaacacca tatggccccg accaaggatg agtttgactg caaggcctgg	2880
[0066]	gcctacttct ccgacgtgga ccttgagaag gatgtccatt ccggcctgat cgggccgctg	2940
[0067]	ctcgtgtgtc acaccaacac cctgaacca gcgcattgac gccaggtcac cgtccaggag	3000
[0068]	tttgctctgt tcttcaccat ttttgacgaa actaagtcct ggtacttcac cgagaatatg	3060
[0069]	gagcgaact gtagagcgcc ctgcaatc cagatggaag atccgacttt caaggagaac	3120
[0070]	tatagattcc acgccatcaa cgggtacatc atggatactc tgccggggct ggtcatggcc	3180
[0071]	caggatcaga ggattcgggt gtacttctgt tcaatgggat cgaacgaaaa cattcactcc	3240
[0072]	attcacttct ccggtcacgt gttcactgtg cgcaagaagg aggagtacaa gatggcgctg	3300
[0073]	tacaatctgt accccggggt gttcgaaact gtggagatgc tgccgtccaa ggccggcatc	3360
[0074]	tggagagtgg agtgctgat cggagagcac ctccacgcgg ggatgtccac cctcttctg	3420
[0075]	gtgtactcga ataagtcca gaccccgctg ggcattggcct cgggccacat cagagacttc	3480
[0076]	cagatcacag caagcggaca atacggccaa tgggcgccga agctggccc cttgcactac	3540
[0077]	tccggatcga tcaacgcatg gtccaccaag gaaccgttct cgtggattaa ggtggacctc	3600
[0078]	ctggccccta tgattatcca cggaattaag acccaggcg ccaggcagaa gttctctcc	3660
[0079]	ctgtacatct cgcaattcat catcatgtac agcctggacg ggaagaagtg gcagacttac	3720
[0080]	aggggaaact ccaccggcac cctgatggtc tttttcgca acgtggattc ctccggcatt	3780
[0081]	aagcacaaca tcttcaacc accgatcata gccagatata ttaggtcca cccactcac	3840
[0082]	tactcaatcc gctcaactct tcggatggaa ctcatgggt gcgacctgaa ctctgtctcc	3900
[0083]	atgccgttgg ggatggaatc aaaggctatt agcgacgcc agatcaccgc gagctctac	3960

[0084]	ttcactaaca tgttgccac ctggagcccc tccaaggcca ggctgcactt gcaggacgg	4020
[0085]	tcaaatgcct ggcgccgca agtgaacaat ccgaaggaat ggcttcaagt ggatttccaa	4080
[0086]	aagaccatga aagtgaccgg agtcaccacc caggagtgga agtcccttct gacctcgatg	4140
[0087]	tatgtgaagg agttcctgat tagcagcagc caggacgggc accagtggac cctgtttctt	4200
[0088]	caaaacggaa aggtcaaggt gttccagggg aaccaggact cgttcacacc cgtggtgaac	4260
[0089]	tccttgacc cccactgct gacgcggtac ttgaggattc atcctcagtc ctgggtccat	4320
[0090]	cagattgcat tgcgaatgga agtcctgggc tgcgagggcc aggacctgta ctga	4374
[0091]	<210> 2	
[0092]	<211> 4374	
[0093]	<212> DNA	
[0094]	<213> 人工序列	
[0095]	<220>	
[0096]	<223> coFVIII-4	
[0097]	<400> 2	
[0098]	atgcagatcg agctgagcac gtgcttcttc ctgtgcctgc tgaggttctg cttcagcgcc	60
[0099]	accaggaggt actacctggg cgccgtggag ctgagctggg actacatgca gagegacctg	120
[0100]	ggcgagctgc ccgtggacgc caggttcccc ccagggtgc ccaagagctt ccccttcaac	180
[0101]	acgagcgtgg tgtacaagaa gacctgttc gtggagttca ccgacctct gttcaatatc	240
[0102]	gccaagccca ggccccctg gatggggctg ctggggccca cgatccaggc cgaggtgtac	300
[0103]	gacaccgtgg tcatcacct gaagaacatg gccagccacc ccgtgagcct gcacgccgtg	360
[0104]	ggcgtgagct actggaaggc cagcgagggc gccgagtacg acgaccagac cagccagagg	420
[0105]	gagaaggagg acgacaaggt gttccccggc ggcagccaca cctacgtgtg gcaggtgctg	480
[0106]	aaggagaatg ggccccatggc cagcgacccc ctgtgcctga cctactctta cctgagccac	540
[0107]	gtggatctgg tgaaggacct gaacagcggc ctgatcggcg ccctgctggt gtgcaggagg	600
[0108]	ggcagcctgg ccaaggagaa gaccagacc ctgcacaagt tcatcctgct gttegccgtg	660
[0109]	ttcgacgagg gcaagagctg gcacagcgag accaagaaca gcctgatgca ggataggac	720
[0110]	gccgccagcg ccagggcctg gccaagatg cacaccgtga acggctacgt gaacaggtct	780
[0111]	ctgcccggcc tgatcggtg ccacaggaag agcgtgtact ggcacgtgat cggcatgggg	840
[0112]	accaccccg aggtgcacag catcttctg gagggccaca cgttctggt gaggaatcac	900
[0113]	aggcaggcca gcctggagat cagcccgatc accttctga ccgccagac cctgctgatg	960
[0114]	gacctggggc agttctgct gttctgcat atcagctctc accagcacga cggcatggag	1020
[0115]	gcctacgtga aggtggatag ctgccccgag gagccccagc tgaggatgaa gaacaacgag	1080
[0116]	gaggccgagg actacacga cgacctgacc gacagcgaga tggacgtggt gaggttcgac	1140
[0117]	gacgacaata gcccagctt catccagatc aggagcgtgg ccaagaagca cccaagacc	1200
[0118]	tgggtgcatt acatcgccgc cgaggaggag gattgggact acgccccct ggtgctggcc	1260
[0119]	cccagcaca ggtcttaca gagccagtac ctgaacaacg ggccccagag gatcggcagg	1320
[0120]	aagtacaaga aggtgaggtt catggcctac accgacgaga cttcaagac caggaggcg	1380
[0121]	atccagcacg agagcgggat cctggggccc ctgctgtacg gcgaggtggg cgacacgtg	1440
[0122]	ctgatcatct tcaagaacca ggccagcagg ccgtacaata tctacccca cgggatcacc	1500
[0123]	gacgtgaggc ccctgtactc taggaggctg cccaaggcg tgaagcacct gaaggacttc	1560
[0124]	cccatcctgc ccggcgagat cttcaagtac aagtggaccg tgaccgtgga ggacggggcc	1620
[0125]	acgaagagcg accccaggtg cctgaccagg tactacagct ctttctgtaa catggagagg	1680

[0126]	gacctggcca gcggcctgat cgggccccctg ctgatctgct acaaggagag cgtggatcag	1740
[0127]	aggggcaacc agatcatgag cgacaagagg aacgtgatcc tgttcagcgt gttcgacgag	1800
[0128]	aataggtctt ggtacctgac cgagaatata cagaggttcc tgcccaaccc cgccggcgtg	1860
[0129]	cagctggagg atccccagtt ccaggccagc aacatcatgc acagcatcaa cggctacgtg	1920
[0130]	ttcgacagcc tgcagctgag cgtgtgcctg cagcaggtgg cctactggta catcctgagc	1980
[0131]	atcggcgccc agaccgactt cctgagcgtg ttcttcagcg gctacacctt caagcacaag	2040
[0132]	atggtgtacg aggataccct gacctgttc ccttcagcg gcgagaccgt gttcatgagc	2100
[0133]	atggagaacc ccggcctgtg gatcctgggc tgccataact ccgacttcag gaataggggc	2160
[0134]	atgaccgccc tgctgaagggt gagctcttgc gacaagaaca ccggcgacta ctacgaggat	2220
[0135]	agctacgagg atatcagcgc ctacctgtg agcaagaaca acgccatcga gcccaggtct	2280
[0136]	ttcagccaga accccccctg gctgaagagg caccagagg agatcaccag gacgacctg	2340
[0137]	cagagcgacc aggaggagat cgactacgac gacacgatca gcgtggagat gaagaaggag	2400
[0138]	gatttcgaca tctacgacga ggacgagaat cagagcccca ggtctttcca gaagaagacc	2460
[0139]	aggcattact tcatcgccgc cgtggagagg ctgtgggact acggcatgag cagctctccc	2520
[0140]	cacgtgctga ggaatagggc ccagagcggc agcgtgcccc agttcaagaa ggtggtgttc	2580
[0141]	caggagttca ccgacggcag cttcaccag ccctgtaca ggggcgagct gaacgagcac	2640
[0142]	ctgggcctgc tggggcccta catcagggcc gaggtggagg ataacatcat ggtgaccttc	2700
[0143]	aggaatcagg ccagcaggcc ctatagcttc tatagctctc tgatcagcta cgaggaggat	2760
[0144]	cagaggcagg gcgccgagcc caggaagaac ttcgtgaagc ccaacgagac caagacctac	2820
[0145]	ttctggaagg tgcagacca catggcccc acgaaggacg agttcgactg caaggcctgg	2880
[0146]	gcctacttca gcgacgtgga tctggagaag gacgtgcaca gcggcctgat cgggccccctg	2940
[0147]	ctggtgtgcc acaccaacac cctgaacccc gccacggca ggcaggtgac cgtgcaggag	3000
[0148]	ttcgccctgt tcttcacat cttcgacgag accaagagct ggtacttcac cgagaatatg	3060
[0149]	gagaggaatt gcaggcccc ctgcaatata cagatggagg acccgacctt caaggagaat	3120
[0150]	tacaggttcc acgccatcaa cggctacata atggacacgc tgcccggcct ggtcatggcc	3180
[0151]	caggatcaga ggatcaggtg gtatctgtg agcatgggga gcaacgagaa tatccacagc	3240
[0152]	atccacttca gcggccacgt gttcacctg aggaagaagg aggagtacaa gatggccctg	3300
[0153]	tacaatctgt accccggcgt gttcgagacc gtggagatgc tgcccagcaa ggccgggac	3360
[0154]	tggagggtgg agtgctgat cggcgagcac ctgcacgcc gcattgagcac gctgttctg	3420
[0155]	gtgtactcta acaagtcca gacccccctg gggatggcca gcggccacat cagggacttc	3480
[0156]	cagatcacccg ccagcggcca gtacggccag tgggccccca agctggccag gctgcactat	3540
[0157]	tccggaagca tcaacgcctg gagcacgaag gagcccttca gctggatcaa ggtggatctg	3600
[0158]	ctggccccca tgatcatcca cgggatcaag acccaggcgc ccaggcagaa gttcagctct	3660
[0159]	ctgtatatca gccagttcat catcatgtac tctctggacg gcaagaagtg gcagacctac	3720
[0160]	aggggcaaca gcaccggcac gctgatggtg ttcttcggca acgtggactc tagcgggatc	3780
[0161]	aagcacaata tcttcaaccc ccccatcata gccaggtaca tcaggctgca cccacccat	3840
[0162]	tactctatca ggtctaccct gaggatggag ctgatgggct gcgacctgaa cagctgcagc	3900
[0163]	atgccccctg ggatggagag caaggccatc agcgacgcc agatcacgc cagctcttac	3960
[0164]	ttaccaaca tgttcgccac ctggagcccc agcaaggcca ggctgcacct gcagggcagg	4020
[0165]	tctaacgcct ggaggcccc ggtgaacaac cccaaggagt ggctgcaggt ggatttcag	4080
[0166]	aagaccatga aggtgaccgg cgtgaccacg cagggcgtga agagcctgct gaccagcatg	4140
[0167]	tacgtgaagg agttctgat cagctctagc caggacggcc accagtggac cctgttcttc	4200

[0168] cagaacggca aggtgaaggt gttccagggc aaccaggata gcttcacccc cgtggtgaac 4260
 [0169] agcctggacc cccccctgct gaccaggtat ctgaggatcc acccccagag ctgggtgcac 4320
 [0170] cagatcgccc tgaggatgga ggtgctgggc tgcgaggccc aggatctgta ttga 4374
 [0171] <210> 3
 [0172] <211> 4374
 [0173] <212> DNA
 [0174] <213> 人工序列
 [0175] <220>
 [0176] <223> coFVIII-52
 [0177] <400> 3
 [0178] atgcaaatcg aactgagcac ctgtttcttc ctctgectgc tgagattctg tttctccgcg 60
 [0179] acccgccgat actacctggg agcagtggag ctctcctggg attacatgca gagegacctt 120
 [0180] ggggagctgc ccgtggatgc caggttcctt ccccggtgc caaagtcgtt tccgttcaac 180
 [0181] acctccgtgg tgtacaagaa aactctgttc gtggagtcca ccgaccacct gttcaatatc 240
 [0182] gccaaagcca gacctccctg gatggggctg ttgggacctt ccatccaagc ggaggtgtac 300
 [0183] gacactgtgg tcatcactct gaagaacatg gctcgcac ccgtgtccct gcacgccgtg 360
 [0184] ggagtgtctt actggaaagc gtccgagggg gccgaatacg acgaccagac ctgcagaga 420
 [0185] gaaaaggaag atgacaaggt gttcccagga ggatcgcaca cctacgtgtg gcaagtgttg 480
 [0186] aaggagaacg gcccaatggc ctccgacccg ctgtgcctga cctactcgta cctgtccac 540
 [0187] gtggacctcg tgaaggacct caactcgga ctgattggag ccctgctggt ctgcagggaa 600
 [0188] ggctcactgg cgaaagaaa gactcagacc ttgcacaagt tcattctgct gttcgtgtg 660
 [0189] ttgcacgagg ggaagtcgtg gcacagcgag actaagaact ccctgatgca agatagagat 720
 [0190] gccgcctccg cccgggcttg gcctaagatg cacaccgtga acggttacgt gaaccgctcc 780
 [0191] ctccctggcc tgattggatg ccaccggaag tccgtgtact ggcacgtgat cgggatgggg 840
 [0192] accacccccg aggtgcacag catcttcctg gaaggtcaca catttctcgt gcgcaaccac 900
 [0193] cggcaggcct ccctggaaat cagccccatt accttctca ctgcccagac tctgtgatg 960
 [0194] gacctgggac agttcctgct gttctgcat atctcctccc accaacaatga cggaatggag 1020
 [0195] gcatacgtga aggtcgatc ctgcccagag gaacccagc tccgcatgaa gaacaatgag 1080
 [0196] gaagccgagg actacgacga cgacctgacg gatagcgaga tggatgtggt ccggttcgat 1140
 [0197] gacgataaca gcccttcctt catccaaatt cgctcgttg caaagaagca cccaagacc 1200
 [0198] tgggtgcatt acattcggc ggaagaagag gactgggatt atgccccgct tgtcctcgt 1260
 [0199] cctgacgacc ggagctacaa gagccagtac ctgaacaacg gtccacagag gatcggtaga 1320
 [0200] aagtacaaga aggtccgctt catggcctat accgacgaaa ctttcaaac tagagaggcc 1380
 [0201] atccaacacg aatccggcat cctgggcccg ctcttgtagc gagaagtcgg cgacaccctt 1440
 [0202] ctattatct tcaagaacca ggcttcccgg ccgtacaaca tctatccga tgggatcact 1500
 [0203] gactgcgcc cactgtactc gcggcgctg cccaagggtg tcaaacacct gaaggatttt 1560
 [0204] ccgatccttc cgggagaaat cttcaagtac aagtggaccg tgaccgtgga agatggccca 1620
 [0205] actaagtctg acctagatg cctcaccgc tactactcat ccttcgtcaa catggagcgc 1680
 [0206] gacctggcca gcgactgat cggcccgtg ctgatttctg acaaggaatc agtggaccaa 1740
 [0207] cggggaaacc agatcatgtc ggataagagg aacgtcatcc tcttctccgt gtttgacgaa 1800
 [0208] aaccggtcgt ggtacgtgac cgagaacatc cagaggttcc tgcccaacc tgctggggtg 1860
 [0209] cagctggagg accccgagtt ccaggccagc aacatcatgc acagcatcaa tggctacgtg 1920

[0210]	ttcgacagcc tgcagctgag cgtgtgcctg caccaggtgg cctactggta catcctgagc	1980
[0211]	atcggcgccc agaccgactt cctgagcgtg ttctttctctg gctacacctt caagcacaag	2040
[0212]	atgggtgatg aggacacctt gacctgttc ccttcagcg gggagactgt cttcatgagc	2100
[0213]	atggagaacc ctggcctgtg gatcctgggc tgccacaaca gcgacttcag gaacaggggc	2160
[0214]	atgactgccc tgctgaaagt ctccagctgt gacaagaaca ccggggacta ctacgaggac	2220
[0215]	agctacgagg acatcagcgc ctacctgtg agcaagaaca atgccatcga gcccaggagc	2280
[0216]	ttctctcaga acccccagt gctgaagagg caccagagg agatcaccag gaccacctg	2340
[0217]	cagtctgacc aggaggagat cgactatgat gacaccatca gcgtggagat gaagaaggag	2400
[0218]	gacttcgaca tctacgacga ggacgagaac cagagcccca ggagcttcca gaagaagacc	2460
[0219]	aggcactact tcattgctgc tgtggagagg ctgtgggact atggcatgtc cagcagcccc	2520
[0220]	catgtgctga ggaacagggc ccagtctggc agcgtgcccc agttcaagaa agtcgtgttc	2580
[0221]	caggagtcca ccgacggcag cttcaccag cccctgtaca gaggggagct gaacgagcac	2640
[0222]	ctgggcctgc tgggccccta catcagggcc gaggtggagg acaacatcat ggtgaccttc	2700
[0223]	aggaaccagg ccagcaggcc ctacagcttc tacagcagcc tgatcagcta cgaggaggac	2760
[0224]	cagaggcagg gggctgagcc caggaagaac tttgtgaagc ccaatgaaac caagacctac	2820
[0225]	ttctggaagg tgcagacca catggcccc accaaggacg agttcgaact caaggcctgg	2880
[0226]	gcctacttct ctgacgtgga cctggagaag gacgtgcact ctggcctgat tggccccctg	2940
[0227]	ctggtgtgcc acaccaacac cctgaaccct gcccatggca ggcagggtgac tgtgcaggag	3000
[0228]	ttcgccctgt tcttcacat cttcgatgaa accaagagct ggtacttcac tgagaacatg	3060
[0229]	gagaggaact gcagggcccc ctgcaacatc cagatggagg accccacctt caaggagaac	3120
[0230]	tacaggttcc atgccatcaa tggctacatc atggacaccc tgcttgccct ggtcatggcc	3180
[0231]	caggaccaga ggatcagggt gtatctgtct agcatgggca gcaacgagaa catccacagc	3240
[0232]	atccacttct ctggccacgt gttcactgtg aggaagaagg aggagtacaa gatggccctg	3300
[0233]	tacaacctgt accctggggt gttcgaaacc gtggagatgc tgcccagcaa ggccggcatc	3360
[0234]	tggagggtgg agtgctgat tggggagcac ctgcacgccg gcatgagcac cctgttctctg	3420
[0235]	gtgtacagca acaagtgcc gacccccctg ggcattggcct ctggccacat cagggacttc	3480
[0236]	cagatcactg cctctggcca gtacggccag tgggccccca agctggccag gctgcactac	3540
[0237]	tccggaagca tcaatgcctg gagcaccaag gagcccttca gctggatcaa agtggacctg	3600
[0238]	ctggccccca tgatcatcca cggcatcaag acccaggggg ccaggcagaa gttctccagc	3660
[0239]	ctgtacatca gccagttcat catcatgtac agcctggacg gcaagaagtg gcagacctac	3720
[0240]	aggggcaaca gcaccggcac cctgatggtg ttcttcggca acgtggacag cagcggcatc	3780
[0241]	aagcacaaca tcttcaaccc ccccatcatc gccagataca tcaggctgca cccacccac	3840
[0242]	tacagcatca ggagaccctt gaggatggag ctgatgggct gtgacctgaa cagctgcagc	3900
[0243]	atggccctgg gcatggagag caaggccatc tctgacgcc agatcactgc ctccagctac	3960
[0244]	ttaccaaca tgtttgccac ctggagcccc agcaaggcca ggctgcacct gcagggcagg	4020
[0245]	agcaatgcct ggaggcccc ggtcaacaac cccaaggagt ggctgcagggt ggacttcag	4080
[0246]	aagaccatga aggtgactgg ggtgaccacc cagggggtga agagcctgct gaccagcatg	4140
[0247]	tacgtgaagg agttctgat ctccagcagc caggacggcc accagtggac cctgttcttc	4200
[0248]	cagaatggca aggtgaaggt gttccagggc aaccaggaca gcttcacccc tgtgtgtaac	4260
[0249]	agcctggacc cccccctgct gaccagatac ctgaggatcc accccagag ctgggtgcac	4320
[0250]	cagatcgccc tgaggatgga ggtgctgggc tgtgaggccc aggacctgta ctga	4374
[0251]	<210> 4	

[0252]	<211> 4374
[0253]	<212> DNA
[0254]	<213> 人工序列
[0255]	<220>
[0256]	<223> coFVIII-62
[0257]	<400> 4
[0258]	atgcagattg agctgtccac ttgtttcttc ctgtgcctcc tgcgcttctg tttctccgcc 60
[0259]	actcgcgggt actaccttgg agccgtggag ctttcatggg actacatgca gagcgacctg 120
[0260]	ggcgaactcc ccgtggatgc cagattcccc ccccgctgc caaagtcctt cccctttaac 180
[0261]	acctccgtgg tgtacaagaa aaccctcttt gtgcagttca ctgaccacct gttcaacatc 240
[0262]	gccaaagccgc gccaccttg gatgggcctc ctgggaccga ccattcaagc tgaagtgtac 300
[0263]	gacaccgtgg tgatcacctt gaagaacatg gcgtcccacc ccgtgtccct gcattgcggtc 360
[0264]	ggagtgtcct actggaaggc ctccgaagga gctgagtacg acgaccagac tagccagcgg 420
[0265]	gaaaaggagg acgataaagt gttcccgggc ggctcgcata cttacgtgtg gcaagtcctg 480
[0266]	aaggaaaacg gacctatggc atccgatcct ctgtgcctga cttactccta cttttcccat 540
[0267]	gtggacctcg tgaaggacct gaacagcggg ctgattggtg cacttctcgt gtgccgcgaa 600
[0268]	ggttcgctcg ctaaggaaaa gaccagacc ctccataagt tcactctttt gttcgctgtg 660
[0269]	ttcgatgaag gaaagtcatt gcattccgaa actaagaact cgctgatgca ggaccgggat 720
[0270]	gccgcctcag cccgcgcctg gcctaaaatg catacagtca acgatacgt gaatcggatc 780
[0271]	ctgcccgggc tcactcggtt tcacagaaag tccgtgtact ggcacgtcat cggcatgggc 840
[0272]	actacgcctg aagtgcactc catcttctct gaagggcaca ctttctcgt gcgcaaccac 900
[0273]	cgccaggcct ctctggaaat ctccccgatt acctttctga ccgccagac tctgtctatg 960
[0274]	gacctggggc agttccttct cttctgccac atctccagcc atcagcacga cggaatggag 1020
[0275]	gcctacgtga aggtggactc atgcccggaa gaacctcagt tgcggatgaa gaacaacgag 1080
[0276]	gaggccgagg actatgacga cgatttgact gactccgaga tggacgtcgt gcggttcgat 1140
[0277]	gacgacaaca gcccagctt catccagatt cgcagcgtgg ccaagaagca ccccaaaacc 1200
[0278]	tgggtgcact acatgcggc cgaggaagaa gattgggact acgccccgtt ggtgctggca 1260
[0279]	ccgatgacc ggtcgtacaa gtcccagtat ctgaacaatg gtccgcagcg gattggcaga 1320
[0280]	aagtacaaga aagtgcggtt catggcgtag actgacgaaa cgtttaagac ccgggaggcc 1380
[0281]	attcaacatg agagcggcat tctgggacca ctgctgtacg gagaggtcgg cgataccctg 1440
[0282]	ctcatcatct taaaaacca ggcctcccg cttacaaca tctaccctca cggaatcacc 1500
[0283]	gacgtgcggc cactctactc gcggcgctg ccgaaggcg tcaagcacct gaaagacttc 1560
[0284]	cctatcctgc cgggcgaaat cttcaagtat aagtggaccg tcaccgtgga ggacgggccc 1620
[0285]	accaagagcg atcctaggtg tctgactcgg tactactcca gtttcgtgaa catggaacgg 1680
[0286]	gacctggcat cgggactcat tggaccgtg ctgatctgct acaaagagtc ggtggatcaa 1740
[0287]	cgcggcaacc agatcatgtc cgacaagcgc aacgtgatcc tgttctccgt gtttgatgaa 1800
[0288]	aacagatcct ggtacctgac cgagaacatc cagaggttcc tgcccaacc tgctggggtg 1860
[0289]	cagctggagg accccgagtt ccaggccagc aacatcatgc acagcatcaa tggctacgtg 1920
[0290]	ttcgacagcc tgcagctgag cgtgtgcctg cagcaggtgg cctactggtg catcctgagc 1980
[0291]	atcggcggcc agaccgactt cctgagcgtg ttcttctctg gctacacctt caagcacaag 2040
[0292]	atggtgtatg aggacacctt gaccctgttc cccttcagcg gggagactgt cttcatgagc 2100
[0293]	atggagaacc ctggcctgtg gatcctgggc tgccacaaca gcgacttcag gaacaggggc 2160

[0294]	atgactgccc tgctgaaagt ctccagctgt gacaagaaca ccggggacta ctacaggagc	2220
[0295]	agctacgagg acatcagcgc ctacctgtg agcaagaaca atgccatcga gccaggagc	2280
[0296]	ttctctcaga accccccagt gctgaagagg caccagaggg agatcaccag gaccacctg	2340
[0297]	cagtctgacc aggaggagat cgactatgat gacacatca gcgtggagat gaagaaggag	2400
[0298]	gacttcgaca tctacgacga ggacgagaac cagagcccca ggagcttcca gaagaagacc	2460
[0299]	aggcactact tcattgctgc tgtggagagg ctgtgggact atggcatgtc cagcagcccc	2520
[0300]	catgtgctga ggaacagggc ccagtctggc agcgtgcccc agttcaagaa agtcgtgttc	2580
[0301]	caggagttca ccgacggcag cttcaccag cccctgtaca gaggggagct gaacgagcac	2640
[0302]	ctgggcctgc tgggccccta catcagggcc gaggtggagg acaacatcat ggtgaccttc	2700
[0303]	aggaaccagg ccagcaggcc ctacagcttc tacagcagcc tgatcagcta cgaggaggac	2760
[0304]	cagaggcagg gggctgagcc caggaagaac tttgtgaagc ccaatgaaac caagacctac	2820
[0305]	ttctggaagg tgcagacca catggcccc accaaggacg agttcgactg caaggcctgg	2880
[0306]	gcctacttct ctgacgtgga cctggagaag gacgtgcact ctggcctgat tggccccctg	2940
[0307]	ctgggtgtgcc acaccaacac cctgaaccct gcccatggca ggcagggtgac tgtgcaggag	3000
[0308]	ttcgccctgt tcttcacat cttcgatgaa accaagagct ggtacttcac tgagaacatg	3060
[0309]	gagaggaact gcagggcccc ctgcaacatc cagatggagg accccacctt caaggagaac	3120
[0310]	tacaggttcc atgccatcaa tggctacatc atggacaccc tgccctggcct ggtcatggcc	3180
[0311]	caggaccaga ggatcagggt gtatctgtg agcatgggca gcaacgagaa catccacagc	3240
[0312]	atccacttct ctggccacgt gttcactgtg aggaagaagg aggagtacaa gatggccctg	3300
[0313]	tacaacctgt acctgggggt gttcgaaacc gtggagatgc tgcccagcaa ggccggcatc	3360
[0314]	tggagggtgg agtgccctgat tggggagcac ctgcacgccg gcatgagcac cctgttccctg	3420
[0315]	gtgtacagca acaagtgcc gacccccctg ggcatggcct ctggccacat cagggacttc	3480
[0316]	cagatcactg cctctggcca gtacggccag tgggccccca agctggccag gctgcactac	3540
[0317]	tccggaagca tcaatgcctg gagcaccaag gagcccttca gctggatcaa agtggacctg	3600
[0318]	ctggccccca tgatcatcca cggcatcaag acccaggggg ccaggcagaa gttctccagc	3660
[0319]	ctgtacatca gccagttcat catcatgtac agcctggacg gcaagaagtg gcagacctac	3720
[0320]	aggggcaaca gcaccggcac cctgatggtg ttcttcggca acgtggacag cagcggcatc	3780
[0321]	aagcacaaca tttcaaccc ccccatcatc gccagatata tcaggctgca cccacccac	3840
[0322]	tacagcatca ggagaccct gaggatggag ctgatgggct gtgacctgaa cagctgcagc	3900
[0323]	atgcccctgg gcatggagag caaggccatc tctgacgcc agatcactgc ctccagctac	3960
[0324]	ttaccaaca tgtttgccac ctggagcccc agcaaggcca ggctgcacct gcagggcagg	4020
[0325]	agcaatgcct ggaggcccc ggtcaacaac cccaaggagt ggctgcaggt ggacttccag	4080
[0326]	aagaccatga aggtgactgg ggtgaccacc cagggggtga agagcctgct gaccagcatg	4140
[0327]	tacgtgaagg agttcctgat ctccagcagc caggacggcc accagtggac cctgttcttc	4200
[0328]	cagaatggca aggtgaaggt gttccagggc aaccaggaca gcttcacccc tgttgtcaac	4260
[0329]	agcctggacc cccccctgct gaccagatac ctgaggatcc acccccagag ctgggtgcac	4320
[0330]	cagatcgccc tgaggatgga ggtgctgggc tgtgaggccc aggacctgta ctga	4374
[0331]	<210> 5	
[0332]	<211> 4374	
[0333]	<212> DNA	
[0334]	<213> 人工序列	
[0335]	<220>	

[0336]	<223> coFVIII-25	
[0337]	<400> 5	
[0338]	atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc	60
[0339]	accaggagat actacctggg cgccgtggag ctgagctggg actacatgca gtctgacctg	120
[0340]	ggcgagctgc cagtggacgc caggttcccc cccagagtgc ccaagagctt ccccttcaac	180
[0341]	accagcgtgg tgtacaagaa gacctgttc gtggagtcca ctgaccacct gttcaacatc	240
[0342]	gccaaagcca ggccccctg gatgggcctg ctgggcccc ccatccaggc cgaggtgtac	300
[0343]	gacaccgtgg tcatcacct gaagaacatg gccagccacc ccgtctccct gcacgccgtg	360
[0344]	ggggtgagct actggaaggc ctctgagggc gccgagtacg acgaccagac cagccagagg	420
[0345]	gagaaggagg acgacaaggt gttccctggg ggcagccaca cctacgtgtg gcaggtcctg	480
[0346]	aaggagaacg gccccatggc ctctgacccc ctgtgcctga cctacagcta cctgagccac	540
[0347]	gtggacctgg tgaaggacct gaactctggc ctgattgggg ccctgctggt gtgcaggag	600
[0348]	ggcagcctgg ccaaggagaa gaccagacc ctgcacaagt tcatcctgct gttcgccgtg	660
[0349]	ttcgacgagg gcaagagctg gcactctgaa accaagaaca gcctgatgca ggacaggac	720
[0350]	gccgcctctg ccagggcctg gcccaagatg cacaccgtca acggctacgt caacaggagc	780
[0351]	ctgcctggcc tgattggctg ccacaggaag agcgtgtact ggcatgtgat cggcatgggc	840
[0352]	accacccctg aggtgcacag catcttctctg gagggccaca ctttctggt caggaaccac	900
[0353]	aggcaggcca gcctggagat cagccccatc accttctga ccgccagac cctgctgatg	960
[0354]	gacctgggcc agttcctgct gttctgccac atctccagcc accagcacga cggcatggag	1020
[0355]	gcctacgtga aagtggacag ctgcccctgag gagccccagc tgaggatgaa gaacaacgag	1080
[0356]	gaggccgagg actatgatga cgacctgacc gacagcgaga tggacgtggt caggttcgac	1140
[0357]	gacgacaaca gcccagctt catccagatc aggagcgtgg ccaagaagca cccaagacc	1200
[0358]	tgggtgcaact acatcgctgc tgaggaggag gactgggact atgccccct ggtgctggcc	1260
[0359]	cctgatgaca ggagctacaa gagccagtac ctgaacaatg gccccagag gattggcagg	1320
[0360]	aagtacaaga aagtcagggt catggcctac actgatgaaa cttcaagac cagggaggcc	1380
[0361]	atccagcatg agtctggcat cctgggcccc ctgctgtacg gggaggtggg ggacaccctg	1440
[0362]	ctgatcatct tcaagaacca ggccagcagg ccctacaaca tctacccca tggcatcacc	1500
[0363]	gacgtgaggc ccctgtacag caggaggtg cctaagggg tgaagcacct gaaagacttc	1560
[0364]	cccatcctgc ctggggagat cttcaagtac aagtggactg tgactgtgga ggacggcccc	1620
[0365]	accaagagcg accccagggt cctgaccaga tactacagca gcttcgtcaa catggagagg	1680
[0366]	gacctggcct ctggcctgat tggccccctg ctgatctgct acaaggagtc tgtggaccag	1740
[0367]	aggggcaacc agatcatgag cgacaagagg aacgtgatcc tgttctctgt cttcgacgag	1800
[0368]	aacaggagct ggtacctgac tgaaaacatc cagcggttcc tccccaaacc cgcgggcgtg	1860
[0369]	cagctggaag atcctgagtt tcaggcatca aacatcatgc actccattaa cggctacgtg	1920
[0370]	ttcgattcgc tgcagctgag cgtgtgtctg cacgaagtgg cctactggtg catcctgtcc	1980
[0371]	attggtgccc agactgactt cctgtccgtg tttttctccg gctacagtt caagcacaag	2040
[0372]	atggtgtacg aggacacct gacctcttc ctttttccg gcgaaactgt gtttatgagc	2100
[0373]	atggagaatc ccggcctgtg gatcttgggc tgccacaaca gcgacttccg taacagagga	2160
[0374]	atgactgcgc tgctcaaggt gtccagctgc gacaagaaca ccggagacta ttatgaggac	2220
[0375]	tcatacgagg acatctccgc ctacctctg tccaagaata acgccattga acctcgagc	2280
[0376]	ttcagccaga acccaccctg gcttaagaga catcaacggg agatcactag gaccaccctg	2340
[0377]	cagtcagacc aggaggaaat cgactacgat gacaccatct cggtcgagat gaagaaggag	2400

[0378]	gactttgaca tctacgacga agatgaaaac cagagcccca ggtcgttcca aaagaaaacc	2460
[0379]	cgccactact ttattgctgc tgtcgagcgg ctgtgggact acggaatgtc gtcctcgccg	2520
[0380]	cacgtgctcc gcaaccgagc ccagagcggc tgggtgccgc aattcaagaa ggtcgtgttc	2580
[0381]	caggagtcca ctgacgggag cttcactcag cttttgtacc ggggagaact caatgaacat	2640
[0382]	ctcggcctcc tcggacctta catcagagca gaagtggag ataacatcat ggtcactttc	2700
[0383]	cgtaaccaag ccagccgccc gtactcgttc tactcctccc tcattttcta cgaagaggac	2760
[0384]	cagcggcagg gcgcagaacc gcgcaagaac ttcgtgaagc ccaacgaaac caagacctac	2820
[0385]	ttctggaaag tgcagcatca tatggccccg actaaggacg agtttgactg caaagcctgg	2880
[0386]	gcctacttct ccgatgtgga cttggagaag gacgtccact ccggcctcat cggccccctg	2940
[0387]	ctcgtgtgcc ataccaatac cctgaacccc gcacacggtc gccaggtcac cgtgcaggag	3000
[0388]	ttcgtctgtg tcttactat cttcgacgaa actaagtcct ggtacttcac cgagaacatg	3060
[0389]	gagaggaact gcagagcccc ctgtaacatc cagatggagg acccgacgtt caagggaaac	3120
[0390]	taccggttcc acgccattaa cggatacatc atggatacgc tgccgggtct tgtgatggcc	3180
[0391]	caggatcaac ggatcagatg gtacttattg tcgatgggca gcaacgagaa catccactct	3240
[0392]	attcacttct ccggtcatgt gttcactgtg cggaagaagg aagagtacaa gatggccctg	3300
[0393]	tacaaccttt atcccggagt gttcgaaact gtggaaatgc tgccgtcgaa ggccggcatt	3360
[0394]	tggcgcgtgg agtgtttgat tggagaacat ctccatgcgg ggatgtcaac cctgttcctg	3420
[0395]	gtgtatagca acaagtgcc aactccgctt gggatggcgt caggacacat tagggatttc	3480
[0396]	cagatcactg cgtccggcca gtacggccaa tgggccccta agctggcccc cctgcattac	3540
[0397]	tccggatcca ttaacgcctg gtcaaccaag gagccattct cctggatcaa ggtggacctt	3600
[0398]	ctggccccca tgattatcca cggaattaag acccaggggg cccggcagaa gttctctca	3660
[0399]	ctgtacatca gccagttcat aatcatgtac tccctggacg gaaagaagtg gcaaacctac	3720
[0400]	agggggaaca gcaccggcac actgatggtc tttttcgaa atgtggactc ctccgggatt	3780
[0401]	aagcataaca tcttcaaccc tccgattatc gctcgttaca ttagacttca ccctaccac	3840
[0402]	tacagcattc gctccaccct gcggatggaa ctgatgggct gcgatctgaa ctcgtgcagc	3900
[0403]	atgccgttgg gaatggagtc caaagcaatt tccgacgcgc agatcaccgc ctctctctac	3960
[0404]	tttaccaaca tgttcgccac gtggtcaccg tccaaggccc ggctgcacct ccagggaaga	4020
[0405]	tccaacgcat ggcgccacaa ggtcaacaac cctaaggagt ggctccagggt ggacttccag	4080
[0406]	aaaaccatga aggtcaccgg agtcacaacc caggagtgta agtcgtctgt gacttctatg	4140
[0407]	tacgtcaagg agttcctgat ctccagcagc caggacgggc accagtggac cctgtttctc	4200
[0408]	caaaatggaa aggtcaaggt gtttcagggc aatcaggatt cattcacccc ggtggtgaac	4260
[0409]	tcccttgatc caccctcct gaccgcctac cttcgcatcc acccacagtc ctgggtgcac	4320
[0410]	cagatcgccg tgaggatgga ggtcctggga tgcgaagccc aggacctgta ctga	4374
[0411]	<210>	6
[0412]	<211>	4374
[0413]	<212>	DNA
[0414]	<213>	人工序列
[0415]	<220>	
[0416]	<223>	coFVIII-26
[0417]	<400>	6
[0418]	atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc	60
[0419]	accaggagat actacctggg cgccgtggag ctgagctggg actacatgca gtctgacctg	120

[0420]	ggcgagctgc cagtggacgc caggttcccc cccagagtgc ccaagagctt ccccttcaac	180
[0421]	accagcgtgg tgtacaagaa gacctgttc gtggagttca ctgaccacct gttcaacatc	240
[0422]	gccaaagccca ggccccctg gatgggcctg ctggggccca ccatccaggc cgaggtgtac	300
[0423]	gacaccgtgg tcatcacct gaagaacatg gccagccacc cgtctccct gcacgccgtg	360
[0424]	ggggtgagct actggaaggc ctctgagggc gccgagtacg acgaccagac cagccagagg	420
[0425]	gagaaggagg acgacaaggt gttccctggg ggcagccaca cctacgtgtg gcaggtcctg	480
[0426]	aaggagaacg gccccatggc ctctgacccc ctgtgcctga cctacagcta cctgagccac	540
[0427]	gtggacctgg tgaaggacct gaactctggc ctgattgggg ccctgctggt gtgcagggag	600
[0428]	ggcagcctgg ccaaggagaa gaccagacc ctgcacaagt tcatcctgct gttcgccgtg	660
[0429]	ttcgacgagg gcaagagctg gcactctgaa accaagaaca gcctgatgca ggacagggac	720
[0430]	gccgcctctg ccagggcctg gcccaagatg cacaccgtca acggctacgt caacaggagc	780
[0431]	ctgcctggcc tgattggctg ccacaggaag agcgtgtact ggcatgtgat cggcatgggc	840
[0432]	accaccctg aggtgcacag catcttctg gagggccaca ctttctggt caggaaccac	900
[0433]	aggcaggcca gcctggagat cagccccatc accttctga cgcgccagac cctgctgatg	960
[0434]	gacctgggcc agttcttgct gttctgccac atctccagcc accagcacga cggcatggag	1020
[0435]	gcctacgtga aagtggacag ctgccctgag gagccccagc tgaggatgaa gaacaacgag	1080
[0436]	gaggccgagg actatgatga cgacctgacc gacagcgaga tggacgtggt caggttcgac	1140
[0437]	gacgacaaca gcccagctt catccagatc aggagcgtgg ccaagaagca cccaagacc	1200
[0438]	tgggtgcaact acatcgctgc tgaggaggag gactgggact atgccccct ggtgctggcc	1260
[0439]	cctgatgaca ggagctacaa gagccagtac ctgaacaatg gccccagag gattggcagg	1320
[0440]	aagtacaaga aagtcaggtt catggcctac actgatgaaa cttcaagac caggagggcc	1380
[0441]	atccagcatg agtctggcat cctgggcccc ctgctgtacg gggaggtggg ggacaccctg	1440
[0442]	ctgatcatct tcaagaacca ggccagcagg ccctacaaca tctacccca tggcatcacc	1500
[0443]	gacgtgaggc ccctgtacag caggagctg cctaaggggg tgaagcacct gaaagacttc	1560
[0444]	cccatcctgc ctggggagat cttcaagtac aagtggactg tgactgtgga ggacggcccc	1620
[0445]	accaagagcg accccagtg cctgaccaga tactacagca gcttcgtcaa catggagagg	1680
[0446]	gacctggcct ctggcctgat tggccccctg ctgatctgct acaaggagtc tgtggaccag	1740
[0447]	aggggaacc agatcatgag cgacaagagg aacgtgatcc tgttctctgt cttcgacgag	1800
[0448]	aacaggagct ggtacctcac tgaaaacatc cagaggttcc tccaaaccc cgcaggagtg	1860
[0449]	caactggagg accctgagtt tcaggcctcg aatatcatgc actcgattaa cggttacgtg	1920
[0450]	ttcgactcgc tgcagctgag cgtgtgcctc catgaagtgc cttactgta cattctgtcc	1980
[0451]	atcggcgccc agactgactt cctgagcgtg ttcttttccg gttacacctt taagcacaag	2040
[0452]	atggtgtacg aagataccct gacctgttc cttttctccg gcgaaacggt gttcatgtcg	2100
[0453]	atggagaacc cgggtctgtg gattctggga tgccacaaca gcgactttcg gaaccgcgga	2160
[0454]	atgactgccc tgctgaaggt gtcctcatgc gacaagaaca ccggagacta ctacgaggac	2220
[0455]	tcctacgagg atatctcagc ctacctctg tccaagaaca acgcgatcga gccgcgcagc	2280
[0456]	ttcagccaga accgcctgt gctgaagagg caccagcgag aaattacccg gaccaccctc	2340
[0457]	caatcggatc aggaggaaat cgactacgac gacaccatct cgggtgaaat gaagaaggaa	2400
[0458]	gatttcgata tctacgacga ggacgaaaat cagtccccctc gctcattcca aaagaaaact	2460
[0459]	agacactact ttatcgccgc ggtggaaaga ctgtgggact atggaatgtc atccagccct	2520
[0460]	cacgtccttc ggaaccgggc ccagagcgga tcggtgcctc agttcaagaa agtgggtgtc	2580
[0461]	caggagtcca ccgacggcag cttcaccag ccgctgtacc ggggagaact gaacgaacac	2640

[0462]	ctgggcctgc tcggtcccta catccgcgcg gaagtggagg ataacatcat ggtgaccttc	2700
[0463]	cgtaaccaag catccagacc ttactccttc tattcctccc tgatctcata cgaggaggac	2760
[0464]	cagcgccaag gcgccgagcc ccgcaagaac ttcgtcaagc ccaacgagac taagacctac	2820
[0465]	ttctggaagg tccaacacca tatggccccg accaaggatg agtttgactg caaggcctgg	2880
[0466]	gcctacttct ccgacgtgga ccttgagaag gatgtccatt ccggcctgat cgggccgctg	2940
[0467]	ctcgtgtgtc acaccaacac cctgaaccca gcgcattggac gccaggtcac cgtccaggag	3000
[0468]	tttgcctgtg tcttcacccat ttttgacgaa actaagtcct ggtacttcac cgagaatatg	3060
[0469]	gagcgaaact gtagagcgcc ctgcaatata cagatggaag atccgacttt caaggagaac	3120
[0470]	tatagattcc acgccatcaa cgggtacatc atggatactc tgccggggct ggtcatggcc	3180
[0471]	caggatcaga ggattcggtg gtacttgctg tcaatgggat cgaacgaaaa cattcactcc	3240
[0472]	attcacttct ccggtcacgt gttcactgtg cgcaagaagg aggagtacaa gatggcgctg	3300
[0473]	tacaatctgt accccggggg gttcgaaact gtggagatgc tgccgtccaa ggccggcatc	3360
[0474]	tggagagtgg agtgctgat cggagagcac ctccacgcgg ggatgtccac cctcttcctg	3420
[0475]	gtgtactcga ataagtcca gaccccgctg ggcatggcct cgggccacat cagagacttc	3480
[0476]	cagatcacag caagcggaca atacggccaa tgggcgccga agctggcccg cttgcactac	3540
[0477]	tccggatcga tcaacgcatt gtccaccaag gaaccgttct cgtggattaa ggtggacctc	3600
[0478]	ctggccccta tgattatcca cggaattaag acccaggcgc ccaggcagaa gttctcctcc	3660
[0479]	ctgtacatct cgcaattcat catcatgtac agcctggacg ggaagaagtg gcagacttac	3720
[0480]	aggggaaact ccaccggcac cctgatggtc tttttcggca acgtggattc ctccggcatt	3780
[0481]	aagcacaaca tcttcaaccc accgatcata gccagatata ttaggtcca cccactcac	3840
[0482]	tactcaatcc gctcaactct tcggatggaa ctcatggggg gcgacctgaa ctctgtctcc	3900
[0483]	atgccgttgg ggatggaatc aaaggctatt agcgacgccc agatcaccgc gagctcctac	3960
[0484]	ttcactaaca tgttgccac ctggagcccc tccaaggcca ggctgcactt gcagggacgg	4020
[0485]	tcaaatgcct ggcgcccgca agtgaacaat ccgaaggaaat ggcttcaagt ggatttccaa	4080
[0486]	aagaccatga aagtgaccgg agtcaccacc caggagtgga agtcccttct gacctcgatg	4140
[0487]	tatgtgaagg agttcctgat tagcagcagc caggacgggc accagtggac cctgtttctc	4200
[0488]	caaacggaa aggtcaaggt gttccagggg aaccaggact cgttcacacc cgtggtgaac	4260
[0489]	tccctggacc cccactgct gacgcggtac ttgaggattc atcctcagtc ctgggtccat	4320
[0490]	cagattgcat tgcgaatgga agtcctgggc tgcgaggccc aggacctgta ctga	4374
[0491]	<210> 7	
[0492]	<211> 1734	
[0493]	<212> DNA	
[0494]	<213> 人工序列	
[0495]	<220>	
[0496]	<223> coFVIII-52-NT58	
[0497]	<400> 7	
[0498]	gcgacccgcc gatactacct gggagcagtg gagctctcct gggattacat gcagagcgac	60
[0499]	cttggggagc tgcccgtgga tgccaggttc cctccccggg tgccaaagtc gtttccgttc	120
[0500]	aacacctccg tgggtgtacaa gaaaactctg ttcgtggagt tcaccgacca cctgttcaat	180
[0501]	atcgccaagc ccagacctcc ctggatgggg ctgttgggac ctaccatcca agcggagggtg	240
[0502]	tacgacactg tggatcatcac tctgaagaac atggcctcgc atcccgtgtc cctgcacgcc	300
[0503]	gtgggagtggt cttactggaa agcgtccgag ggggccgaat acgacgacca gacctcgag	360

[0504]	agagaaaagg aagatgacaa ggtgttccca ggaggatcgc acacctacgt gtggcaagtg	420
[0505]	ttgaaggaga acggcccaat ggcctccgac ccgctgtgcc tgacctactc gtacctgtcc	480
[0506]	cacgtggacc tcgtgaagga cctcaactcg ggactgattg gagecctgct ggtctgcagg	540
[0507]	gaaggctcac tggcgaaaga aaagactcag accttgacaca agttcattct gctgttcgct	600
[0508]	gtgttcgacg aggggaagtc gtggcacagc gagactaaga actccctgat gcaagataga	660
[0509]	gatgccgcct ccgcccgggc ctggcctaag atgcacaccg tgaacggta cgtgaaccgc	720
[0510]	tcctccctg gcctgattgg atgccaccgg aagtccgtgt actggcacgt gatcgggatg	780
[0511]	gggaccaccc ccgaggtgca cagcatcttc ctggaagtc acacatttct cgtgcgcaac	840
[0512]	caccggcagg cctccctgga aatcagcccc attaccttcc tctactgcca gactctgctg	900
[0513]	atggacctgg gacagtccct gctgttctgc catatctcct cccaccaaca tgacggaatg	960
[0514]	gaggcatacg tgaaggtcga ttctgtccct gaggaacccc agctccgcat gaagaacaat	1020
[0515]	gaggaagccg aggactacga cgacgacctg acggatagcg agatggatgt ggtccggttc	1080
[0516]	gatgacgata acagcccttc cttcatccaa attcgctcgg tggcaaagaa gcaccccaag	1140
[0517]	acctgggtgc attacattgc ggcggaagaa gaggactggg attatgcccc gcttgtcctc	1200
[0518]	gctcctgacg accggagcta caagagccag tacctgaaca acggtccaca gaggatcggg	1260
[0519]	agaaagtaca agaaggtccg cttcatggcc tataccgacg aaaccttcaa aactagagag	1320
[0520]	gccatccaac acgaatccgg catcctgggc ccgctcttgt acggagaagt cggcgacacc	1380
[0521]	cttctcatta tcttcaagaa ccaggcttcc cggccgtaca acatctatcc gcatgggatc	1440
[0522]	actgacgtgc gccactgta ctgcggcgc ctgcccagg gtgtcaaaca cctgaaggat	1500
[0523]	tttccgatcc ttccgggaga aatcttcaag tacaagtga ccgtgaccgt ggaagatggc	1560
[0524]	ccaactaagt ctgaccctag atgcctcacc cgctactact catccttcgt caacatggag	1620
[0525]	cgcgacctgg ccagcggact gatcggcccc ctgctgattt gctacaagga atcagtggac	1680
[0526]	caacggggaa accagatcat gtcggataag aggaacgtca tcctcttctc cgtg	1734
[0527]	<210>	8
[0528]	<211>	2583
[0529]	<212>	DNA
[0530]	<213>	人工序列
[0531]	<220>	
[0532]	<223>	coFVIII-52-CT
[0533]	<400>	8
[0534]	tttgacgaaa accggtcgtg gtacctgacc gagaacatcc agaggttcct gccaaccct	60
[0535]	gctggggtgc agctggagga ccccgagttc caggccagca acatcatgca cagcatcaat	120
[0536]	ggctacgtgt tcgacagcct gcagctgagc gtgtgcctgc acgaggtggc ctactggtac	180
[0537]	atcctgagca tcggcgccca gaccgacttc ctgagcgtgt tcttctctgg ctacaccttc	240
[0538]	aagcacaaga tgggtgatga ggacaccctg accctgttcc ctttcagcgg ggagactgtc	300
[0539]	ttcatgagca tggagaaccc tggcctgtgg atcctgggct gccacaacag cgacttcagg	360
[0540]	aacaggggca tgactgccct gctgaaagtc tccagctgtg acaagaacac cggggactac	420
[0541]	tacgaggaca gctacgagga catcagcgcc tacctgctga gcaagaacaa tgccatcgag	480
[0542]	cccaggagct tctctcagaa cccccagtg ctgaagaggc accagaggga gatcaccagg	540
[0543]	accacctgc agtctgacca ggaggagatc gactatgatg acaccatcag cgtggagatg	600
[0544]	aagaaggagg acttcgacat ctacgacgag gacgagaacc agagccccag gagcttcagg	660
[0545]	aagaagacca ggcactactt cattgtctgt gtggagaggc tgtgggacta tggcatgtcc	720

[0546]	agcagccccc atgtgctgag gaacagggcc cagtctggca gcgtgcccc gttcaagaaa	780
[0547]	gtcgtgttcc aggagttcac cgacggcagc ttcacccagc ccctgtacag aggggagctg	840
[0548]	aacgagcacc tgggcctgct gggcccttac atcagggccg aggtggagga caacatcatg	900
[0549]	gtgaccttca ggaaccaggc cagcaggccc tacagcttct acagcagcct gatcagctac	960
[0550]	gaggaggacc agaggcaggg ggctgagccc aggaagaact ttgtgaagcc caatgaaacc	1020
[0551]	aagacctact tctggaaggt gcagcaccac atggccccc ccaaggacga gttcgactgc	1080
[0552]	aaggcctggg cctacttctc tgacgtggac ctggagaagg acgtgcactc tggcctgatt	1140
[0553]	ggcccccctgc tgggtgtgcca caccaacacc ctgaaccctg cccatggcag gcaggtgact	1200
[0554]	gtgcaggagt tcgccctgtt cttcaccatc ttcgatgaaa ccaagagctg gtacttcact	1260
[0555]	gagaacatgg agaggaactg cagggccccc tgcaacatcc agatggagga cccaccttc	1320
[0556]	aaggagaact acaggttcca tgccatcaat ggctacatca tggacaccct gcctggcctg	1380
[0557]	gtcatggccc aggaccagag gatcagggtg tatctgctga gcatgggcag caacgagaac	1440
[0558]	atccacagca tccacttctc tggccacgtg ttcactgtga ggaagaagga ggagtacaag	1500
[0559]	atggccctgt acaacctgta ccctgggggtg ttcgaaaccg tggagatgct gccagcaag	1560
[0560]	gccggcatct ggagggtgga gtgcctgatt ggggagcacc tgcacgccg catgagcacc	1620
[0561]	ctgttcttgg tgtacagaa caagtgccag accccctgg gcatggcctc tggccacatc	1680
[0562]	agggacttcc agatcactgc ctctggccag tacggccagt gggccccc aa gctggccagg	1740
[0563]	ctgcactact ccggaagcat caatgcctgg agcaccaagg agcccttcag ctggatcaaa	1800
[0564]	gtggacctgc tggccccc at gatcatccac ggcatcaaga cccagggggc caggcagaag	1860
[0565]	ttctccagcc tgtacatcag ccagttcatc atcatgtaca gcctggacgg caagaagtgg	1920
[0566]	cagacctaca ggggcaacag caccggcacc ctgatggtgt tcttcggcaa cgtggacagc	1980
[0567]	agcggcatca agcacaacat cttcaacccc cccatcatcg ccagatacat caggctgcac	2040
[0568]	cccaccact acagcatcag gagcacctg aggatggagc tgatgggctg tgacctgaac	2100
[0569]	agctgcagca tgccctggg catggagagc aaggccatct ctgacgccc gatcactgcc	2160
[0570]	tccagctact tcaccaacat gtttgccacc tggagcccca gcaaggccag gctgcacctg	2220
[0571]	cagggcagga gcaatgcctg gagggcccag gtcaacaacc ccaaggagtg gctgcagggtg	2280
[0572]	gacttcaga agaccatgaa ggtgactggg gtgaccacc aggggggtgaa gagcctgctg	2340
[0573]	accagcatgt acgtgaagga gttcctgac tccagcagcc aggacggcca ccagtggacc	2400
[0574]	ctgtttctcc agaattgcaa ggtgaagggtg ttccagggca accaggacag cttcaccct	2460
[0575]	gtggtcaaca gcctggaccc cccctgctg accagatacc tgaggatcca ccccagagc	2520
[0576]	tgggtgcacc agatcgccct gaggatggag gtgctgggct gtgaggccca ggacctgtac	2580
[0577]	tga	2583
[0578]	<210>	9
[0579]	<211>	1734
[0580]	<212>	DNA
[0581]	<213>	人工序列
[0582]	<220>	
[0583]	<223>	coFVIII-62-NT58
[0584]	<400>	9
[0585]	gccactcgcc ggtactacct tggagccgtg gagctttcat gggactacat gcagagcgac	60
[0586]	ctgggcgaac tccccgtgga tgccagattc cccccccg tgccaaagtc cttccccctt	120
[0587]	aacacctccg tgggtgtacaa gaaaaccctc tttgtcgagt tcaactgacca cctgttcaac	180

[0588]	atcgccaagc cgcgcccacc ttggatgggc ctctggggac cgaccattca agctgaagtg	240
[0589]	tacgacaccg tggatgacac cctgaagaac atggcgctccc acccgtgtc cctgcatgcg	300
[0590]	gtcggagtgt cctactggaa ggctccgaa ggagctgagt acgacgacca gactagccag	360
[0591]	cgggaaaagg aggacgataa agtgttcccg ggcggtcgc atacttacgt gtggcaagtc	420
[0592]	ctgaaggaaa acggacctat ggcacccgat cctctgtgcc tgacttactc ctacctttcc	480
[0593]	catgtggacc tcgtgaagga cctgaacagc gggctgattg gtgcacttct cgtgtgccgc	540
[0594]	gaaggttcgc tcgctaagga aaagaccag accctccata agttcatcct tttgttcgct	600
[0595]	gtgttcgatg aaggaaagtc atggcattcc gaaactaaga actcgctgat gcaggaccgg	660
[0596]	gatgccgcct cagcccgcgc ctggcctaaa atgcatacag tcaacggata cgtgaatcgg	720
[0597]	tcactgcccg ggctcatcgg ttgtcacaga aagtccgtgt actggcacgt catcgcatg	780
[0598]	ggcactacgc ctgaagtga ctccatcttc ctggaagggc acaccttcct cgtgcgcaac	840
[0599]	caccgccagg cctctctgga aatctccccg attaccttcc tgaccgccc gactctgctc	900
[0600]	atggacctgg ggcagttcct tctcttctgc cacatctcca gccatcagca cgacggaatg	960
[0601]	gaggcctacg tgaaggtgga ctcatgcccg gaagaacctc agttgcggat gaagaacaac	1020
[0602]	gaggaggccg aggactatga cgacgatttg actgactccg agatggacgt cgtgcggttc	1080
[0603]	gatgacgaca acagccccag ctccatccag attcgacgc tggccaagaa gcaccccaaa	1140
[0604]	acctgggtgc actacatcgc ggccgaggaa gaagattggg actacgccc gttggtgctg	1200
[0605]	gcacccgatg accggtcgta caagtcacg tatctgaaca atgggtccgca gcggattggc	1260
[0606]	agaaagtaca agaaagtgcg gttcatggcg tactactgacg aaacgtttaa gaccgggag	1320
[0607]	gccattcaac atgagagcgg cattctggga cactgctgt acggagaggt cggcgatacc	1380
[0608]	ctgctcatca tcttcaaaaa ccaggcctcc cgcccttaca acatctaccc tcacggaatc	1440
[0609]	accgacgtgc ggccactcta ctgcggcgc ctgccgaagg gcgtcaagca cctgaaagac	1500
[0610]	ttccctatcc tgccgggcga aatcttcaag tataagtga ccgtcacctg ggaggacggg	1560
[0611]	cccaccaaga gcatcctag gtgtctgact cggctactact ccagcttcgt gaacatggaa	1620
[0612]	cgggacctgg catcgggact cattggaccg ctgctgatct gctacaaaga gtcggtggat	1680
[0613]	caacgcggca accagatcat gtccgacaag cgcaacgtga tcctgttctc cgtg	1734
[0614]	<210> 10	
[0615]	<211> 2583	
[0616]	<212> DNA	
[0617]	<213> 人工序列	
[0618]	<220>	
[0619]	<223> coFVIII-62-CT	
[0620]	<400> 10	
[0621]	tttgatgaaa acagatcctg gtacctgacc gagaacatcc agaggttcct gcccaccct	60
[0622]	gctggggtgc agctggagga ccccgagttc caggccagca acatcatgca cagcatcaat	120
[0623]	ggctacgtgt tcgacagcct gcagctgagc gtgtgcctgc acgaggtggc ctactgttac	180
[0624]	atcctgagca tcggcgccca gaccgacttc ctgagcgtgt tcttctctgg ctacaccttc	240
[0625]	aagcacaaga tgggtgatga ggacaccctg accctgttcc ccttcagcgg ggagactgtc	300
[0626]	ttcatgagca tggagaacct tggcctgtgg atcctgggct gccacaacag cgacttcagg	360
[0627]	aacaggggca tgactgccct gctgaaagtc tccagctgtg acaagaacac cggggactac	420
[0628]	tacgaggaca gctacgagga catcagcgcc tacctgctga gcaagaacaa tgccatcgag	480
[0629]	cccaggagct tctctcagaa cccccagtg ctgaagaggc accagaggga gatcaccagg	540

[0630]	accacacctgc agtctgacca ggaggagatc gactatgatg acaccatcag cgtggagatg	600
[0631]	aagaaggagg acttcgacat ctacgacgag gacgagaacc agagccccag gagcttccag	660
[0632]	aagaagacca ggcactactt cattgtctgt gtggagaggc tgtgggacta tggcatgtcc	720
[0633]	agcagcccc atgtgtctgag gaacagggcc cagtctggca gcgtgcccc gttcaagaaa	780
[0634]	gtcgtgttcc aggagttcac cgacggcagc ttcaccagc ccctgtacag aggggagctg	840
[0635]	aacgagcacc tgggcctgct gggcccctac atcagggccg aggtggagga caacatcatg	900
[0636]	gtgaccttca ggaaccagge cagcaggccc tacagcttct acagcagcct gatcagctac	960
[0637]	gaggaggacc agaggcaggg ggctgagccc aggaagaact ttgtgaagcc caatgaaacc	1020
[0638]	aagacctact tctggaaggt gcagcaccac atggccccca ccaaggacga gttcgactgc	1080
[0639]	aaggcctggg cctacttctc tgacgtggac ctggagaagg acgtgcactc tggcctgatt	1140
[0640]	ggccccctgc tgggtgtgcca caccaacacc ctgaacctg cccatggcag gcaggtgact	1200
[0641]	gtgcaggagt tcgccctgtt cttcaccatc ttcatgaaa ccaagagctg gtacttcaact	1260
[0642]	gagaacatgg agaggaactg cagggcccc tgcaacatcc agatggagga cccaccttc	1320
[0643]	aaggagaact acaggttcca tgccatcaat ggctacatca tggacaccct gcctggcctg	1380
[0644]	gtcatggccc aggaccagag gatcagggtg tatctgtga gcatgggcag caacgagaac	1440
[0645]	atccacagca tccacttctc tggccacgtg ttcactgtga ggaagaagga ggagtacaag	1500
[0646]	atggccctgt acaacctgta ccctgggggtg ttcgaaaccg tggagatgct gccagcaag	1560
[0647]	gccggcatct ggagggtgga gtgcctgatt ggggagcacc tgcacgccgg catgagcacc	1620
[0648]	ctgttcctgg tgtacagcaa caagtgccag acccccctgg gcatggcctc tggccacatc	1680
[0649]	agggacttcc agatcactgc ctctggccag tacggccagt gggccccaa gctggccagg	1740
[0650]	ctgcactact ccggaagcat caatgcctgg agcaccaagg agcccttcag ctggatcaaa	1800
[0651]	gtggacctgc tggcccccat gatcatccac ggcatcaaga cccagggggc caggcagaag	1860
[0652]	ttctccagcc tgtacatcag ccagttcatc atcatgtaca gcctggacgg caagaagtgg	1920
[0653]	cagacctaca ggggcaacag caccggcacc ctgatgggtg tcttcggcaa cgtggacagc	1980
[0654]	agcggcatca agcacaacat cttcaacccc cccatcatcg ccagatacat caggctgcac	2040
[0655]	cccaccact acagcatcag gagcaccctg aggatggagc tgatgggctg tgacctgaac	2100
[0656]	agctgcagca tgcccctggg catggagagc aaggccatct ctgacgcca gatcactgcc	2160
[0657]	tccagctact tcaccaacat gtttgccacc tggagcccca gcaaggccag gctgcacctg	2220
[0658]	cagggcagga gcaatgctg gaggccccag gtcaacaacc ccaaggagtg gctgcagggtg	2280
[0659]	gacttccaga agaccatgaa ggtgactggg gtgaccaccc agggggtgaa gagcctgctg	2340
[0660]	accagcatgt acgtgaagga gttcctgatc tccagcagcc aggacggcca ccagtggacc	2400
[0661]	ctgtttcttc agaatggcaa ggtgaagggtg ttccagggca accaggacag cttcaccctc	2460
[0662]	gtgggtcaaca gcctggaccc cccctgctg accagatacc tgaggatcca ccccagagc	2520
[0663]	tgggtgcacc agatcgccct gaggatggag gtgctgggct gtgaggccca ggacctgtac	2580
[0664]	tga	2583
[0665]	<210>	11
[0666]	<211>	1734
[0667]	<212>	DNA
[0668]	<213>	人工序列
[0669]	<220>	
[0670]	<223>	coFVIII-25-NT58
[0671]	<400>	11

[0672]	gccaccagga gatactacct gggcgccgtg gagctgagct gggactacat gcagtctgac	60
[0673]	ctgggcgagc tgccagtgga cgccagggtc ccccccagag tgcccaagag cttccccttc	120
[0674]	aacaccagcg tgggtgtacaa gaagaccctg ttctgtggagt tcaactgacca cctgttcaac	180
[0675]	atcgccaagc ccaggccccc ctggatgggc ctgctgggcc ccaccatcca ggccgagggtg	240
[0676]	tacgacaccg tggatcatcac cctgaagaac atggccagcc acccgtctc cctgcacgcc	300
[0677]	gtgggggtga gctactggaa ggcctctgag ggcgccgagt acgacgacca gaccagccag	360
[0678]	agggagaagg aggacgacaa ggtgttccct gggggcagcc acacctacgt gtggcaggtc	420
[0679]	ctgaaggaga acggccccc atggcctctgac cccctgtgcc tgacctacag ctacctgagc	480
[0680]	cacgtggacc tgggtgaagga cctgaactct ggcctgattg gggccctgct ggtgtgcagg	540
[0681]	gagggcagcc tggccaagga gaagaccag accctgcaca agttcatcct gctgttcgcc	600
[0682]	gtgttcgacg agggcaagag ctggcactct gaaaccaaga acagcctgat gcaggacagg	660
[0683]	gacgccgct ctgccagggc ctggcccaag atgcacaccg tcaacggcta cgtcaacagg	720
[0684]	agcctgcctg gcctgattgg ctgccacagg aagagcgtgt actggcatgt gatcgcatg	780
[0685]	ggcaccaccc ctgagggtgca cagcatcttc ctggagggcc acaccttctt ggtcaggaac	840
[0686]	cacaggcagg ccagcctgga gatcagcccc atcaccttc tgaccgcccc gaccctgctg	900
[0687]	atggacctgg gccagttcct gctgttctgc cacatctcca gccaccagca cgacggcatg	960
[0688]	gaggcctacg tgaaagtgga cagctgccct gaggagcccc agctgaggat gaagaacaac	1020
[0689]	gaggaggccg aggactatga tgacgacctg accgacagcg agatggacgt ggtcaggttc	1080
[0690]	gacgacgaca acagccccag cttcatccag atcaggagcg tggccaagaa gcacccaag	1140
[0691]	acctgggtgc actacatcgc tgctgaggag gaggactggg actatgcccc cctggtgctg	1200
[0692]	gcccctgatg acaggagcta caagagccag tacctgaaca atggccccc gaggattggc	1260
[0693]	aggaagtaca agaaagtcag gttcatggcc taaactgatg aaaccttcaa gaccaggag	1320
[0694]	gccatccagc atgagctctg catcctgggc cccctgctgt acggggagggt gggggacacc	1380
[0695]	ctgctgatca tcttcaagaa ccaggccagc aggcctaca acatctaccc ccatggcatc	1440
[0696]	accgacgtga ggcccctgta cagcaggagg ctgcctaagg ggggtgaagca cctgaaagac	1500
[0697]	ttccccatcc tgccctggga gatcttcaag tacaagtgga ctgtgactgt ggaggacggc	1560
[0698]	cccaccaaga gcgaccccag gtgcctgacc agatactaca gcagcttcgt caacatggag	1620
[0699]	agggacctgg cctctggcct gattggcccc ctgctgatct gctacaagga gtctgtggac	1680
[0700]	cagaggggca accagatcat gagcgacaag aggaacgtga tctgtttctc tgctc	1734
[0701]	<210> 12	
[0702]	<211> 2583	
[0703]	<212> DNA	
[0704]	<213> 人工序列	
[0705]	<220>	
[0706]	<223> coFVIII-25-CT	
[0707]	<400> 12	
[0708]	ttcgacgaga acaggagctg gtacctgact gaaaacatcc agcggttctt cccaacccc	60
[0709]	gcgggcgtgc agctggaaga tcttgagttt caggcatcaa acatcatgca ctccattaac	120
[0710]	ggctacgtgt tcgattcgtc gcagctgagc gtgtgtctgc acgaagtggc ctactgttac	180
[0711]	atcctgtcca ttgggtgccc gactgacttc ctgtccgtgt ttttctccgg ctacacgttc	240
[0712]	aagcacaaga tgggtgtacga ggacaccctg accctcttcc ctttttccgg cgaaactgtg	300
[0713]	tttatgagca tggagaatcc cggcctgtgg atcttgggct gccacaacag cgacttccgt	360

[0714]	aacagaggaa tgactgcgct gctcaagggtg tccagctgcg acaagaacac cggagactat	420
[0715]	tatgaggact catacagga catctccgcc tacctcctgt ccaagaataa cgccattgaa	480
[0716]	cctcgagct tcagccagaa cccaccgtg cttaagagac atcaacggga gatcactagg	540
[0717]	accaccctgc agtcagacca ggaggaaatc gactacgatg acaccatctc ggtcgagatg	600
[0718]	aagaaggagg actttgacat ctacgacgaa gatgaaaacc agagcccag gtcgttccaa	660
[0719]	aagaaaaccc gccactactt tattgtctgt gtcgagcggc tgtgggacta cggaatgtcg	720
[0720]	tcctcgccgc acgtgtccg caaccgagcc cagagcggct cggtgccgca attcaagaag	780
[0721]	gtcgtgttcc aggagtac tgacgggagc ttactcagc ctttgtaccg gggagaactc	840
[0722]	aatgaacatc tcggcctcct cggaccttac atcagagcag aagtggaaga taacatcatg	900
[0723]	gtcactttcc gtaaccaagc cagccgccc gactcgttct actcctccct catttcttac	960
[0724]	gaagaggacc agcggcagg cgacgaacc cgcaagaact tcgtgaagcc caacgaaacc	1020
[0725]	aagacctact tctggaaagt gcagcatcat atggccccga ctaaggacga gtttgactgc	1080
[0726]	aaagcctggg cctactctc cgatgtggac ttggagaagg acgtccactc cggcctcatc	1140
[0727]	ggtccctgc tcgtgtgcca taccaatacc ctgaacccg cacacggtcg ccaggtcacc	1200
[0728]	gtgcaggagt tcgctctgtt ttactatc ttgcagaaa ctaagtctg gtacttcacc	1260
[0729]	gagaacatgg agaggaactg cagagcccc tgtaacatcc agatggagga cccgacgttc	1320
[0730]	aaggaaaact accggttcca cgccattaac ggatacatca tggatacgt gccgggtctt	1380
[0731]	gtgatggccc aggatcaacg gatcagatgg tacttattgt cgatgggcag caacgagaac	1440
[0732]	atccactcta ttacttctc cggatcatgt ttactgtgc ggaagaagga agagtacaag	1500
[0733]	atggccctgt acaaccttta tcccgagtg ttcgaaactg tggaatgct gccgtcgaag	1560
[0734]	gccggcattt ggcgcgtgga gtgtttgatt ggagaacatc tccatgcggg gatgtcaacc	1620
[0735]	ctgttcttg tgtatagcaa caagtgccag actccgctg ggatggcgtc aggacacatt	1680
[0736]	agggatttcc agatcactgc gtccggccag tacggccaat gggcccctaa gctggcccgc	1740
[0737]	ctgcattact ccgatccat taacgcctg tcaaccaagg agccattctc ctgatcaag	1800
[0738]	gtggaccttc tggcccccat gattatccac ggaattaaga cccagggggc ccggcagaag	1860
[0739]	ttctctcac tgtacatcag ccagttcata atcatgtact ccctggacgg aaagaagtgg	1920
[0740]	caaacctaca gggggaacag caccggcaca ctgatgtct ttttcggaaa tgttgactcc	1980
[0741]	tccgggatta agcataacat cttcaaccct ccgattatcg ctcggtacat tagacttcac	2040
[0742]	cctaccact acagattcg ctccacctg cggatggaac tgatgggtg cgatctgaac	2100
[0743]	tcgtgcagca tgccgttggg aatggagtcc aaagcaattt ccgacgcga gatcaccgcc	2160
[0744]	tcgtctact ttaccaacat gtccgccag tggtcaccgt ccaaggcccc gctgcacctc	2220
[0745]	cagggaagat ccaacgcatg gcggccacag gtcaacaacc ctaaggagt gctccagggtg	2280
[0746]	gacttcaga aaacctgaa ggtcaccgga gtcacaacc agggagtga gtcgtgtg	2340
[0747]	acttctatgt acgtcaagga gttcctgac tccagagcc aggcgggca ccagtggacc	2400
[0748]	ctgttcttcc aaaatgaaa ggtcaagggtg ttccaggga atcaggattc attcaccgcc	2460
[0749]	gtggtgaact cccttgatcc acccctctg acccgtacc ttgcaccca cccacagtcc	2520
[0750]	tgggtgcacc agatcgcgct gaggatggag gtcctgggat gcgaagccca ggacctgtac	2580
[0751]	tga	2583
[0752]	<210>	13
[0753]	<211>	1734
[0754]	<212>	DNA
[0755]	<213>	人工序列

[0756]	<220>	
[0757]	<223>	coFVIII-26-NT58
[0758]	<400>	13
[0759]	gccaccagga gatactacct gggcgccgtg gagctgagct gggactacat gcagtctgac	60
[0760]	ctgggcgagc tgccagtgga cgccaggttc cccccagag tgcccaagag cttccccttc	120
[0761]	aacaccagcg tgggtgtacaa gaagaccctg ttcgtggagt tctactgacca cctgttcaac	180
[0762]	atcgccaagc ccaggccccc ctggatgggc ctgctgggcc ccaccatcca ggccgaggtg	240
[0763]	tacgacaccg tggatcatcac cctgaagaac atggccagcc acccgtctc cctgcacgcc	300
[0764]	gtgggggtga gctactggaa ggcctctgag ggcgccgagt acgacgacca gaccagccag	360
[0765]	agggagaagg aggacgacaa ggtgttccct gggggcagcc acacctacgt gtggcaggtc	420
[0766]	ctgaaggaga acggccccc atggcctctgac cccctgtgcc tgacctacag ctacctgagc	480
[0767]	cacgtggacc tgggtgaagga cctgaactct ggcctgattg gggccctgct ggtgtgcagg	540
[0768]	gagggcagcc tggccaagga gaagaccag accctgcaca agttcatcct gctgttcgcc	600
[0769]	gtgttcgacg agggcaagag ctggcactct gaaaccaaga acagcctgat gcaggacagg	660
[0770]	gacgccgcct ctgccagggc ctggcccaag atgcacaccg tcaacggcta cgtcaacagg	720
[0771]	agcctgcctg gcctgattgg ctgccacagg aagagcgtgt actggcatgt gatcgcatg	780
[0772]	ggcaccaccc ctgaggtgca cagcatcttc ctggagggcc acaccttcct ggtcaggaac	840
[0773]	cacaggcagg ccagcctgga gatcagcccc atcaccttc tgaccgcca gaccctgctg	900
[0774]	atggacctgg gccagttcct gctgttctgc cacatctcca gccaccagca cgacggcatg	960
[0775]	gaggcctacg tgaaagtgga cagctgcctt gaggagcccc agctgaggat gaagaacaac	1020
[0776]	gaggaggccg aggactatga tgacgacctg accgacagcg agatggacgt ggtcaggttc	1080
[0777]	gacgacgaca acagccccag cttcatccag atcaggagcg tggccaagaa gcacccaag	1140
[0778]	acctgggtgc actacatgc tgctgaggag gaggactggg actatgcccc cctggtgctg	1200
[0779]	gcccctgatg acaggagcta caagagccag tacctgaaca atggccccca gaggattggc	1260
[0780]	aggaagtaca agaaagtcag gttcatggcc tacactgatg aaaccttcaa gaccaggag	1320
[0781]	gccatccagc atgagtctgg catcctgggc cccctgctgt acggggaggt gggggacacc	1380
[0782]	ctgctgatca tcttcaagaa ccaggccagc aggcctaca acatctaccc ccatggcatc	1440
[0783]	accgacgtga ggccccgtga cagcaggagg ctgcctaagg ggggtgaagca cctgaaagac	1500
[0784]	ttccccatcc tgcttgggga gatcttcaag tacaagtgga ctgtgactgt ggaggacggc	1560
[0785]	cccaccaaga gcgaccccag gtgcctgacc agatactaca gcagcttcgt caacatggag	1620
[0786]	agggacctgg cctctggcct gattggcccc ctgctgatct gctacaagga gtctgtggac	1680
[0787]	cagaggggca accagatcat gagcgacaag aggaacgtga tcctgttctc tgctc	1734
[0788]	<210>	14
[0789]	<211>	2583
[0790]	<212>	DNA
[0791]	<213>	人工序列
[0792]	<220>	
[0793]	<223>	coFVIII-26-CT
[0794]	<400>	14
[0795]	ttcgacgaga acaggagctg gtacctcact gaaaacatcc agaggttcct cccaaacccc	60
[0796]	gcaggagtgc aactggagga ccctgagttt caggcctcga atatcatgca ctcgattaac	120
[0797]	ggttacgtgt tcgactcgct gcagctgagc gtgtgcctcc atgaagtcgc ttactgttac	180

[0798]	attctgtcca tcggcgccca gactgacttc ctgagcgtgt tcttttccgg ttacaccttt	240
[0799]	aagcacaaga tgggtgtacga agataccctg accctgttcc ctttctccgg cgaaacgggtg	300
[0800]	ttcatgtcga tggagaaccc ggggtctgtgg attctgggat gccacaacag cgactttccgg	360
[0801]	aaccgcggaa tgactgcctt gctgaagggtg tcctcatgcg acaagaacac cggagactac	420
[0802]	tacgaggact cctacgagga tatctcagcc tacctcctgt ccaagaacaa cgcgatcgag	480
[0803]	ccgcgcagct tcagccagaa cccgcctgtg ctgaagaggc accagcgaga aattaccggg	540
[0804]	accaccctcc aatcgatca ggaggaaatc gactacgacg acaccatctc ggtggaaatg	600
[0805]	aagaaggaag atttcgatat ctacgacgag gacgaaaatc agtcccctcg ctcatctcaa	660
[0806]	aagaaaacta gacactactt tatcgccgcg gtggaaagac tgtgggacta tggaatgtca	720
[0807]	tccagccctc acgtccttcg gaaccgggccc cagagcggat cgggtgcctca gttcaagaaa	780
[0808]	gtgggtgttc aggagttcac cgacggcagc ttcaccagc cgctgtaccg gggagaactg	840
[0809]	aacgaacacc tgggcctgct cggtccttac atccgcgagg aagtggagga taacatcatg	900
[0810]	gtgaccttcc gtaaccaagc atccagacct tactccttct attcctccct gatctcatac	960
[0811]	gaggaggacc agcgccaagg cgccgagccc cgcaagaact tcgtcaagcc caacgagact	1020
[0812]	aagacctact tctggaaggt ccaacaccat atggccccga ccaaggatga gtttgactgc	1080
[0813]	aaggcctggg cctacttctc cgacgtggac cttgagaagg atgtccattc cggcctgac	1140
[0814]	gggccgctgc tcgtgtgtca caccaacacc ctgaaccag cgcatggacg ccaggtcacc	1200
[0815]	gtccaggagt ttgctctgtt cttcaccatt tttgacgaaa ctaagtcttg gtacttcacc	1260
[0816]	gagaatatgg agcgaaactg tagagcggcc tgcaatatcc agatggaaga tccgactttc	1320
[0817]	aaggagaact atagattcca cgccatcaac gggatcatca tggatactct gccggggctg	1380
[0818]	gtcatggccc aggatcagag gattcgggtg tacttgtgt caatgggac gaacgaaaac	1440
[0819]	attcactcca ttacttctc cggtcacgtg ttactgtgc gcaagaagga ggagtacaag	1500
[0820]	atggcgtgtg acaatctgta ccccggggtg ttcgaaactg tggagatgct gccgtccaag	1560
[0821]	gccgcatctt ggagagtgga gtgcctgacg ggagagcacc tccacgcggg gatgtccacc	1620
[0822]	ctcttcttgg tgtactcgaa taagtccag acccgcctgg gcatggcctc gggccacatc	1680
[0823]	agagacttcc agatcacagc aagcggacaa tacggccaat gggcgccgaa gctggccccg	1740
[0824]	ttgcactact ccgatcgat caacgcatgg tccaccaagg aaccgttctc gtggattaag	1800
[0825]	gtggacctcc tggcccttat gattatccac ggaattaaga cccagggcgc caggcagaag	1860
[0826]	ttctctctcc tgtacatctc gcaattcatc atcatgtaca gcctggacgg gaagaagtgg	1920
[0827]	cagacttaca ggggaaactc caccggcacc ctgatggtct ttttcggcaa cgtggattcc	1980
[0828]	tccggcatta agcacaacat cttcaacca ccgatcatag ccagatatat taggctccac	2040
[0829]	cccactcact actcaatccg ctcaactctt cggatggaac tcatgggggtg cgacctgaac	2100
[0830]	tcctgtctcca tgccgttggg gatggaatca aaggctatta gcgacgccc gatcaccgcg	2160
[0831]	agctcctact tactaacaat gttcgccacc tggagcccct ccaaggccag gctgcacttg	2220
[0832]	cagggacggg caaatgcctg gcggccgcaa gtgaacaatc cgaaggaatg gcttcaagt	2280
[0833]	gatttcaaaa agaccatgaa agtgaccgga gtcaccaccc agggagtga gtccttctg	2340
[0834]	acctgatgtg atgtgaagga gttcctgatt agcagcagcc aggacgggca ccagtggacc	2400
[0835]	ctgttcttcc aaaacggaaa ggtcaagggtg ttccagggga accaggactc gttcacaccc	2460
[0836]	gtggatgaact ccctggaccc cccactgtg acgcggtact tgaggattca tcctcagtcc	2520
[0837]	tgggtccatc agattgcatt gcgaatggaa gtcctgggct gcgaggccca ggacctgtac	2580
[0838]	tga	2583
[0839]	<210>	15

[0840]	<211>	2332
[0841]	<212>	PRT
[0842]	<213>	智人
[0843]	<400>	15
[0844]	Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser Trp Asp Tyr	
[0845]	1	5 10 15
[0846]	Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg Phe Pro Pro	
[0847]		20 25 30
[0848]	Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val Tyr Lys Lys	
[0849]		35 40 45
[0850]	Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile Ala Lys Pro	
[0851]		50 55 60
[0852]	Arg Pro Pro Trp Met Gly Leu Leu Gly Pro Thr Ile Gln Ala Glu Val	
[0853]	65	70 75 80
[0854]	Tyr Asp Thr Val Val Ile Thr Leu Lys Asn Met Ala Ser His Pro Val	
[0855]		85 90 95
[0856]	Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ala Ser Glu Gly Ala	
[0857]		100 105 110
[0858]	Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp Asp Lys Val	
[0859]		115 120 125
[0860]	Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu Lys Glu Asn	
[0861]		130 135 140
[0862]	Gly Pro Met Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser Tyr Leu Ser	
[0863]	145	150 155 160
[0864]	His Val Asp Leu Val Lys Asp Leu Asn Ser Gly Leu Ile Gly Ala Leu	
[0865]		165 170 175
[0866]	Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr Gln Thr Leu	
[0867]		180 185 190
[0868]	His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly Lys Ser Trp	
[0869]		195 200 205
[0870]	His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp Ala Ala Ser	
[0871]		210 215 220
[0872]	Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr Val Asn Arg	
[0873]	225	230 235 240
[0874]	Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val Tyr Trp His	
[0875]		245 250 255
[0876]	Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile Phe Leu Glu	
[0877]		260 265 270
[0878]	Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser Leu Glu Ile	
[0879]		275 280 285
[0880]	Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met Asp Leu Gly	
[0881]		290 295 300

[0882]	Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His Asp Gly Met
[0883]	305 310 315 320
[0884]	Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro Gln Leu Arg
[0885]	325 330 335
[0886]	Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp Leu Thr Asp
[0887]	340 345 350
[0888]	Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser Pro Ser Phe
[0889]	355 360 365
[0890]	Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr Trp Val His
[0891]	370 375 380
[0892]	Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro Leu Val Leu
[0893]	385 390 395 400
[0894]	Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn Asn Gly Pro
[0895]	405 410 415
[0896]	Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met Ala Tyr Thr
[0897]	420 425 430
[0898]	Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu Ser Gly Ile
[0899]	435 440 445
[0900]	Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu Leu Ile Ile
[0901]	450 455 460
[0902]	Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro His Gly Ile
[0903]	465 470 475 480
[0904]	Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys Gly Val Lys
[0905]	485 490 495
[0906]	His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe Lys Tyr Lys
[0907]	500 505 510
[0908]	Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp Pro Arg Cys
[0909]	515 520 525
[0910]	Leu Thr Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg Asp Leu Ala
[0911]	530 535 540
[0912]	Ser Gly Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu Ser Val Asp
[0913]	545 550 555 560
[0914]	Gln Arg Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val Ile Leu Phe
[0915]	565 570 575
[0916]	Ser Val Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu Asn Ile Gln
[0917]	580 585 590
[0918]	Arg Phe Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp Pro Glu Phe
[0919]	595 600 605
[0920]	Gln Ala Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val Phe Asp Ser
[0921]	610 615 620
[0922]	Leu Gln Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp Tyr Ile Leu
[0923]	625 630 635 640

[0924]	Ser	Ile	Gly	Ala	Gln	Thr	Asp	Phe	Leu	Ser	Val	Phe	Phe	Ser	Gly	Tyr
[0925]					645					650					655	
[0926]	Thr	Phe	Lys	His	Lys	Met	Val	Tyr	Glu	Asp	Thr	Leu	Thr	Leu	Phe	Pro
[0927]					660					665					670	
[0928]	Phe	Ser	Gly	Glu	Thr	Val	Phe	Met	Ser	Met	Glu	Asn	Pro	Gly	Leu	Trp
[0929]					675					680					685	
[0930]	Ile	Leu	Gly	Cys	His	Asn	Ser	Asp	Phe	Arg	Asn	Arg	Gly	Met	Thr	Ala
[0931]					690					695					700	
[0932]	Leu	Leu	Lys	Val	Ser	Ser	Cys	Asp	Lys	Asn	Thr	Gly	Asp	Tyr	Tyr	Glu
[0933]					705					710					715	
[0934]	Asp	Ser	Tyr	Glu	Asp	Ile	Ser	Ala	Tyr	Leu	Leu	Ser	Lys	Asn	Asn	Ala
[0935]					725					730					735	
[0936]	Ile	Glu	Pro	Arg	Ser	Phe	Ser	Gln	Asn	Ser	Arg	His	Pro	Ser	Thr	Arg
[0937]					740					745					750	
[0938]	Gln	Lys	Gln	Phe	Asn	Ala	Thr	Thr	Ile	Pro	Glu	Asn	Asp	Ile	Glu	Lys
[0939]					755					760					765	
[0940]	Thr	Asp	Pro	Trp	Phe	Ala	His	Arg	Thr	Pro	Met	Pro	Lys	Ile	Gln	Asn
[0941]					770					775					780	
[0942]	Val	Ser	Ser	Ser	Asp	Leu	Leu	Met	Leu	Leu	Arg	Gln	Ser	Pro	Thr	Pro
[0943]					785					790					795	
[0944]	His	Gly	Leu	Ser	Leu	Ser	Asp	Leu	Gln	Glu	Ala	Lys	Tyr	Glu	Thr	Phe
[0945]					805					810					815	
[0946]	Ser	Asp	Asp	Pro	Ser	Pro	Gly	Ala	Ile	Asp	Ser	Asn	Asn	Ser	Leu	Ser
[0947]					820					825					830	
[0948]	Glu	Met	Thr	His	Phe	Arg	Pro	Gln	Leu	His	His	Ser	Gly	Asp	Met	Val
[0949]					835					840					845	
[0950]	Phe	Thr	Pro	Glu	Ser	Gly	Leu	Gln	Leu	Arg	Leu	Asn	Glu	Lys	Leu	Gly
[0951]					850					855					860	
[0952]	Thr	Thr	Ala	Ala	Thr	Glu	Leu	Lys	Lys	Leu	Asp	Phe	Lys	Val	Ser	Ser
[0953]					865					870					875	
[0954]	Thr	Ser	Asn	Asn	Leu	Ile	Ser	Thr	Ile	Pro	Ser	Asp	Asn	Leu	Ala	Ala
[0955]					885					890					895	
[0956]	Gly	Thr	Asp	Asn	Thr	Ser	Ser	Leu	Gly	Pro	Pro	Ser	Met	Pro	Val	His
[0957]					900					905					910	
[0958]	Tyr	Asp	Ser	Gln	Leu	Asp	Thr	Thr	Leu	Phe	Gly	Lys	Lys	Ser	Ser	Pro
[0959]					915					920					925	
[0960]	Leu	Thr	Glu	Ser	Gly	Gly	Pro	Leu	Ser	Leu	Ser	Glu	Glu	Asn	Asn	Asp
[0961]					930					935					940	
[0962]	Ser	Lys	Leu	Leu	Glu	Ser	Gly	Leu	Met	Asn	Ser	Gln	Glu	Ser	Ser	Trp
[0963]					945					950					955	
[0964]	Gly	Lys	Asn	Val	Ser	Ser	Thr	Glu	Ser	Gly	Arg	Leu	Phe	Lys	Gly	Lys
[0965]					965					970					975	

[0966]	Arg Ala His Gly Pro Ala Leu Leu Thr Lys Asp Asn Ala Leu Phe Lys
[0967]	980 985 990
[0968]	Val Ser Ile Ser Leu Leu Lys Thr Asn Lys Thr Ser Asn Asn Ser Ala
[0969]	995 1000 1005
[0970]	Thr Asn Arg Lys Thr His Ile Asp Gly Pro Ser Leu Leu Ile Glu
[0971]	1010 1015 1020
[0972]	Asn Ser Pro Ser Val Trp Gln Asn Ile Leu Glu Ser Asp Thr Glu
[0973]	1025 1030 1035
[0974]	Phe Lys Lys Val Thr Pro Leu Ile His Asp Arg Met Leu Met Asp
[0975]	1040 1045 1050
[0976]	Lys Asn Ala Thr Ala Leu Arg Leu Asn His Met Ser Asn Lys Thr
[0977]	1055 1060 1065
[0978]	Thr Ser Ser Lys Asn Met Glu Met Val Gln Gln Lys Lys Glu Gly
[0979]	1070 1075 1080
[0980]	Pro Ile Pro Pro Asp Ala Gln Asn Pro Asp Met Ser Phe Phe Lys
[0981]	1085 1090 1095
[0982]	Met Leu Phe Leu Pro Glu Ser Ala Arg Trp Ile Gln Arg Thr His
[0983]	1100 1105 1110
[0984]	Gly Lys Asn Ser Leu Asn Ser Gly Gln Gly Pro Ser Pro Lys Gln
[0985]	1115 1120 1125
[0986]	Leu Val Ser Leu Gly Pro Glu Lys Ser Val Glu Gly Gln Asn Phe
[0987]	1130 1135 1140
[0988]	Leu Ser Glu Lys Asn Lys Val Val Val Gly Lys Gly Glu Phe Thr
[0989]	1145 1150 1155
[0990]	Lys Asp Val Gly Leu Lys Glu Met Val Phe Pro Ser Ser Arg Asn
[0991]	1160 1165 1170
[0992]	Leu Phe Leu Thr Asn Leu Asp Asn Leu His Glu Asn Asn Thr His
[0993]	1175 1180 1185
[0994]	Asn Gln Glu Lys Lys Ile Gln Glu Glu Ile Glu Lys Lys Glu Thr
[0995]	1190 1195 1200
[0996]	Leu Ile Gln Glu Asn Val Val Leu Pro Gln Ile His Thr Val Thr
[0997]	1205 1210 1215
[0998]	Gly Thr Lys Asn Phe Met Lys Asn Leu Phe Leu Leu Ser Thr Arg
[0999]	1220 1225 1230
[1000]	Gln Asn Val Glu Gly Ser Tyr Asp Gly Ala Tyr Ala Pro Val Leu
[1001]	1235 1240 1245
[1002]	Gln Asp Phe Arg Ser Leu Asn Asp Ser Thr Asn Arg Thr Lys Lys
[1003]	1250 1255 1260
[1004]	His Thr Ala His Phe Ser Lys Lys Gly Glu Glu Glu Asn Leu Glu
[1005]	1265 1270 1275
[1006]	Gly Leu Gly Asn Gln Thr Lys Gln Ile Val Glu Lys Tyr Ala Cys
[1007]	1280 1285 1290

[1008]	Thr Thr Arg Ile Ser Pro Asn Thr Ser Gln Gln Asn Phe Val Thr		
[1009]	1295	1300	1305
[1010]	Gln Arg Ser Lys Arg Ala Leu Lys Gln Phe Arg Leu Pro Leu Glu		
[1011]	1310	1315	1320
[1012]	Glu Thr Glu Leu Glu Lys Arg Ile Ile Val Asp Asp Thr Ser Thr		
[1013]	1325	1330	1335
[1014]	Gln Trp Ser Lys Asn Met Lys His Leu Thr Pro Ser Thr Leu Thr		
[1015]	1340	1345	1350
[1016]	Gln Ile Asp Tyr Asn Glu Lys Glu Lys Gly Ala Ile Thr Gln Ser		
[1017]	1355	1360	1365
[1018]	Pro Leu Ser Asp Cys Leu Thr Arg Ser His Ser Ile Pro Gln Ala		
[1019]	1370	1375	1380
[1020]	Asn Arg Ser Pro Leu Pro Ile Ala Lys Val Ser Ser Phe Pro Ser		
[1021]	1385	1390	1395
[1022]	Ile Arg Pro Ile Tyr Leu Thr Arg Val Leu Phe Gln Asp Asn Ser		
[1023]	1400	1405	1410
[1024]	Ser His Leu Pro Ala Ala Ser Tyr Arg Lys Lys Asp Ser Gly Val		
[1025]	1415	1420	1425
[1026]	Gln Glu Ser Ser His Phe Leu Gln Gly Ala Lys Lys Asn Asn Leu		
[1027]	1430	1435	1440
[1028]	Ser Leu Ala Ile Leu Thr Leu Glu Met Thr Gly Asp Gln Arg Glu		
[1029]	1445	1450	1455
[1030]	Val Gly Ser Leu Gly Thr Ser Ala Thr Asn Ser Val Thr Tyr Lys		
[1031]	1460	1465	1470
[1032]	Lys Val Glu Asn Thr Val Leu Pro Lys Pro Asp Leu Pro Lys Thr		
[1033]	1475	1480	1485
[1034]	Ser Gly Lys Val Glu Leu Leu Pro Lys Val His Ile Tyr Gln Lys		
[1035]	1490	1495	1500
[1036]	Asp Leu Phe Pro Thr Glu Thr Ser Asn Gly Ser Pro Gly His Leu		
[1037]	1505	1510	1515
[1038]	Asp Leu Val Glu Gly Ser Leu Leu Gln Gly Thr Glu Gly Ala Ile		
[1039]	1520	1525	1530
[1040]	Lys Trp Asn Glu Ala Asn Arg Pro Gly Lys Val Pro Phe Leu Arg		
[1041]	1535	1540	1545
[1042]	Val Ala Thr Glu Ser Ser Ala Lys Thr Pro Ser Lys Leu Leu Asp		
[1043]	1550	1555	1560
[1044]	Pro Leu Ala Trp Asp Asn His Tyr Gly Thr Gln Ile Pro Lys Glu		
[1045]	1565	1570	1575
[1046]	Glu Trp Lys Ser Gln Glu Lys Ser Pro Glu Lys Thr Ala Phe Lys		
[1047]	1580	1585	1590
[1048]	Lys Lys Asp Thr Ile Leu Ser Leu Asn Ala Cys Glu Ser Asn His		
[1049]	1595	1600	1605

[1050]	Ala Ile Ala Ala Ile Asn Glu Gly Gln Asn Lys Pro Glu Ile Glu
[1051]	1610 1615 1620
[1052]	Val Thr Trp Ala Lys Gln Gly Arg Thr Glu Arg Leu Cys Ser Gln
[1053]	1625 1630 1635
[1054]	Asn Pro Pro Val Leu Lys Arg His Gln Arg Glu Ile Thr Arg Thr
[1055]	1640 1645 1650
[1056]	Thr Leu Gln Ser Asp Gln Glu Glu Ile Asp Tyr Asp Asp Thr Ile
[1057]	1655 1660 1665
[1058]	Ser Val Glu Met Lys Lys Glu Asp Phe Asp Ile Tyr Asp Glu Asp
[1059]	1670 1675 1680
[1060]	Glu Asn Gln Ser Pro Arg Ser Phe Gln Lys Lys Thr Arg His Tyr
[1061]	1685 1690 1695
[1062]	Phe Ile Ala Ala Val Glu Arg Leu Trp Asp Tyr Gly Met Ser Ser
[1063]	1700 1705 1710
[1064]	Ser Pro His Val Leu Arg Asn Arg Ala Gln Ser Gly Ser Val Pro
[1065]	1715 1720 1725
[1066]	Gln Phe Lys Lys Val Val Phe Gln Glu Phe Thr Asp Gly Ser Phe
[1067]	1730 1735 1740
[1068]	Thr Gln Pro Leu Tyr Arg Gly Glu Leu Asn Glu His Leu Gly Leu
[1069]	1745 1750 1755
[1070]	Leu Gly Pro Tyr Ile Arg Ala Glu Val Glu Asp Asn Ile Met Val
[1071]	1760 1765 1770
[1072]	Thr Phe Arg Asn Gln Ala Ser Arg Pro Tyr Ser Phe Tyr Ser Ser
[1073]	1775 1780 1785
[1074]	Leu Ile Ser Tyr Glu Glu Asp Gln Arg Gln Gly Ala Glu Pro Arg
[1075]	1790 1795 1800
[1076]	Lys Asn Phe Val Lys Pro Asn Glu Thr Lys Thr Tyr Phe Trp Lys
[1077]	1805 1810 1815
[1078]	Val Gln His His Met Ala Pro Thr Lys Asp Glu Phe Asp Cys Lys
[1079]	1820 1825 1830
[1080]	Ala Trp Ala Tyr Phe Ser Asp Val Asp Leu Glu Lys Asp Val His
[1081]	1835 1840 1845
[1082]	Ser Gly Leu Ile Gly Pro Leu Leu Val Cys His Thr Asn Thr Leu
[1083]	1850 1855 1860
[1084]	Asn Pro Ala His Gly Arg Gln Val Thr Val Gln Glu Phe Ala Leu
[1085]	1865 1870 1875
[1086]	Phe Phe Thr Ile Phe Asp Glu Thr Lys Ser Trp Tyr Phe Thr Glu
[1087]	1880 1885 1890
[1088]	Asn Met Glu Arg Asn Cys Arg Ala Pro Cys Asn Ile Gln Met Glu
[1089]	1895 1900 1905
[1090]	Asp Pro Thr Phe Lys Glu Asn Tyr Arg Phe His Ala Ile Asn Gly
[1091]	1910 1915 1920

[1092]	Tyr Ile Met Asp Thr Leu Pro Gly Leu Val Met Ala Gln Asp Gln
[1093]	1925 1930 1935
[1094]	Arg Ile Arg Trp Tyr Leu Leu Ser Met Gly Ser Asn Glu Asn Ile
[1095]	1940 1945 1950
[1096]	His Ser Ile His Phe Ser Gly His Val Phe Thr Val Arg Lys Lys
[1097]	1955 1960 1965
[1098]	Glu Glu Tyr Lys Met Ala Leu Tyr Asn Leu Tyr Pro Gly Val Phe
[1099]	1970 1975 1980
[1100]	Glu Thr Val Glu Met Leu Pro Ser Lys Ala Gly Ile Trp Arg Val
[1101]	1985 1990 1995
[1102]	Glu Cys Leu Ile Gly Glu His Leu His Ala Gly Met Ser Thr Leu
[1103]	2000 2005 2010
[1104]	Phe Leu Val Tyr Ser Asn Lys Cys Gln Thr Pro Leu Gly Met Ala
[1105]	2015 2020 2025
[1106]	Ser Gly His Ile Arg Asp Phe Gln Ile Thr Ala Ser Gly Gln Tyr
[1107]	2030 2035 2040
[1108]	Gly Gln Trp Ala Pro Lys Leu Ala Arg Leu His Tyr Ser Gly Ser
[1109]	2045 2050 2055
[1110]	Ile Asn Ala Trp Ser Thr Lys Glu Pro Phe Ser Trp Ile Lys Val
[1111]	2060 2065 2070
[1112]	Asp Leu Leu Ala Pro Met Ile Ile His Gly Ile Lys Thr Gln Gly
[1113]	2075 2080 2085
[1114]	Ala Arg Gln Lys Phe Ser Ser Leu Tyr Ile Ser Gln Phe Ile Ile
[1115]	2090 2095 2100
[1116]	Met Tyr Ser Leu Asp Gly Lys Lys Trp Gln Thr Tyr Arg Gly Asn
[1117]	2105 2110 2115
[1118]	Ser Thr Gly Thr Leu Met Val Phe Phe Gly Asn Val Asp Ser Ser
[1119]	2120 2125 2130
[1120]	Gly Ile Lys His Asn Ile Phe Asn Pro Pro Ile Ile Ala Arg Tyr
[1121]	2135 2140 2145
[1122]	Ile Arg Leu His Pro Thr His Tyr Ser Ile Arg Ser Thr Leu Arg
[1123]	2150 2155 2160
[1124]	Met Glu Leu Met Gly Cys Asp Leu Asn Ser Cys Ser Met Pro Leu
[1125]	2165 2170 2175
[1126]	Gly Met Glu Ser Lys Ala Ile Ser Asp Ala Gln Ile Thr Ala Ser
[1127]	2180 2185 2190
[1128]	Ser Tyr Phe Thr Asn Met Phe Ala Thr Trp Ser Pro Ser Lys Ala
[1129]	2195 2200 2205
[1130]	Arg Leu His Leu Gln Gly Arg Ser Asn Ala Trp Arg Pro Gln Val
[1131]	2210 2215 2220
[1132]	Asn Asn Pro Lys Glu Trp Leu Gln Val Asp Phe Gln Lys Thr Met
[1133]	2225 2230 2235

[1134]	Lys Val Thr Gly Val Thr Thr Gln Gly Val Lys Ser Leu Leu Thr	
[1135]	2240	2245 2250
[1136]	Ser Met Tyr Val Lys Glu Phe Leu Ile Ser Ser Ser Gln Asp Gly	
[1137]	2255	2260 2265
[1138]	His Gln Trp Thr Leu Phe Phe Gln Asn Gly Lys Val Lys Val Phe	
[1139]	2270	2275 2280
[1140]	Gln Gly Asn Gln Asp Ser Phe Thr Pro Val Val Asn Ser Leu Asp	
[1141]	2285	2290 2295
[1142]	Pro Pro Leu Leu Thr Arg Tyr Leu Arg Ile His Pro Gln Ser Trp	
[1143]	2300	2305 2310
[1144]	Val His Gln Ile Ala Leu Arg Met Glu Val Leu Gly Cys Glu Ala	
[1145]	2315	2320 2325
[1146]	Gln Asp Leu Tyr	
[1147]	2330	
[1148]	<210> 16	
[1149]	<211> 4371	
[1150]	<212> DNA	
[1151]	<213> 人工序列	
[1152]	<220>	
[1153]	<223> BDD-FVIII (非优化;“亲本”),核苷酸序列	
[1154]	<400> 16	
[1155]	atgcaaataag agctctccac ctgcttcttt ctgtgccitt tgcgattctg ctttagtgcc	60
[1156]	accagaagat actacctggg tgcaagtggaa ctgtcatggg actatatgca aagtgatctc	120
[1157]	ggtgagctgc ctgtggacgc aagatttcct cctagagtgc caaaatcttt tccattcaac	180
[1158]	acctcagtcg tgtacaaaaa gactctgttt gtagaattca cggatcacct tttcaacatc	240
[1159]	gctaagccaa ggccaccctg gatgggtctg ctaggtccta ccatccaggc tgaggtttat	300
[1160]	gatacagtggt tcattacact taagaacatg gcttcccatc ctgtcagtct tcatgctgtt	360
[1161]	ggtgtatcct actggaaagc ttctgaggga gctgaatatg atgatcagac cagtcaaagg	420
[1162]	gagaaagaag atgataaagt cttccctggt ggaagccata catatgtctg gcaggctctg	480
[1163]	aaagagaatg gtccaatggc ctctgaccca ctgtgcctta cctactcata tctttctcat	540
[1164]	gtggacctgg taaaagactt gaattcaggc ctcatggag ccctactagt atgtagagaa	600
[1165]	gggagtctgg ccaaggaaaa gacacagacc ttgcacaaat ttatactact ttttgctgta	660
[1166]	tttgatgaag ggaaaagttg gcactcagaa acaaagaact ctttgatgca ggatagggat	720
[1167]	gctgcatctg ctcgggcctg gcctaaaatg cacacagtca atggttatgt aaacaggtct	780
[1168]	ctgccaggtc tgattggatg ccacaggaaa tcagtctatt ggcatgtgat tggaatgggc	840
[1169]	accactcctg aagtgcactc aatattcctc gaaggtcaca catttcttgt gaggaacct	900
[1170]	cgccaggcgt ctttggaat ctcgccaata actttcctta ctgctcaaac actcttgatg	960
[1171]	gaccttgac agtttctact gttttgtcat atctcttccc accaactga tggcatggaa	1020
[1172]	gcttatgtca aagtagacag ctgtccagag gaacccaac tacgaatgaa aaataatgaa	1080
[1173]	gaagcggaag actatgatga tgatcttact gattctgaaa tggatgtggt caggtttgat	1140
[1174]	gatgacaact ctcttccctt tatccaaatt cgctcagttg ccaagaagca tcctaaaact	1200
[1175]	tggttacatt acattgctgc tgaagaggag gactgggact atgctcctt agtctctgcc	1260

[1176]	cccgatgaca gaagttataa aagtcaatat ttgaacaatg gccctcagcg gatttgtagg	1320
[1177]	aagtacaaaa aagtccgatt tatggcatat acagatgaaa cttttaagac tcgtgaagct	1380
[1178]	attcagcatg aatcaggaat cttgggacct ttactttatg gggaagttgg agacacactg	1440
[1179]	ttgattatat ttaagaatca agcaagcaga ccatataaca tctacctca cggaatcact	1500
[1180]	gatgtccgtc ctttgtattc aaggagatta caaaaagggtg taaaacattt gaaggatttt	1560
[1181]	ccaattctgc caggagaaat attcaaatat aaatggacag tgactgtaga agatgggcca	1620
[1182]	actaaatcag atcctcgggtg cctgacccgc tattactcta gtttcgttaa tatggagaga	1680
[1183]	gatctagctt caggactcat tggccctctc ctcatctgct acaaagaatc tgtagatcaa	1740
[1184]	agaggaaacc agataatgtc agacaagagg aatgtcatcc tgttttctgt atttgatgag	1800
[1185]	aaccgaagct ggtacctcac agagaatata caacgcttc tcccaatcc agctggagt	1860
[1186]	cagcttgagg atccagagtt ccaagcctcc aacatcatgc acagcatcaa tggctatgtt	1920
[1187]	tttgatagtt tgcagttgtc agtttgtttg catgaggtgg catactggta cattctaagc	1980
[1188]	attggagcac agactgactt cttttctgtc ttcttctctg gatataacct caaacacaaa	2040
[1189]	atggtctatg aagacacact caccctattc ccattctcag gagaaactgt cttcatgtcg	2100
[1190]	atggaaaacc caggtctatg gattctgggg tgccacaact cagactttcg gaacagaggc	2160
[1191]	atgaccgctt tactgaaggt ttctagtgtg gacaagaaca ctggtgatta ttacgaggac	2220
[1192]	agttatgaag atatttcagc atacttctgt agtaaaaaa atgccattga accaagaagc	2280
[1193]	ttctctcaaa acccaccagt cttgaaacgc catcaacggg aaataactcg tactactctt	2340
[1194]	cagtcagatc aagaggaaat tgactatgat gataccatat cagttgaaat gaagaaggaa	2400
[1195]	gattttgaca tttatgatga ggatgaaaat cagagccccc gcagctttca aaagaaaaca	2460
[1196]	cgacactatt ttattgtctc agtggagagg ctctgggatt atgggatgag tagctcccca	2520
[1197]	catgttctaa gaaacagggc tcagagtggc agtgtccctc agttcaagaa agttgttttc	2580
[1198]	caggaattta ctgatgctc ctttactcag cccttatacc gtggagaact aaatgaacat	2640
[1199]	ttgggactcc tggggccata tataagagca gaagttgaag ataatatcat ggtaactttc	2700
[1200]	agaaatcagg cctctcgctc ctattccttc tattctagcc ttattttcta tgaggaagat	2760
[1201]	cagaggcaag gagcagaacc tagaaaaaac tttgtcaagc ctaatgaaac caaaacttac	2820
[1202]	ttttgaaag tgcaacatca tatggcacc ctaaaagatg agtttgactg caaagcctgg	2880
[1203]	gcttatttct ctgatgttga cctggaaaaa gatgtgcact caggcctgat tggacccctt	2940
[1204]	ctggtctgcc aactaacac actgaaccct gctcatggga gacaagtac agtacaggaa	3000
[1205]	tttgcctgtt ttttcaccat ctttgatgag accaaaagct ggtacttcac tgaaaatatg	3060
[1206]	gaaagaaact gcagggtcc ctgcaatac cagatggaag atcccacttt taaagagaat	3120
[1207]	tatcgcttcc atgcaatcaa tggctacata atggatacac tacctggctt agtaatggct	3180
[1208]	caggatcaaa ggattcgatg gtatctgtc agcatgggca gcaatgaaa catccattct	3240
[1209]	attcatttca gtggacatgt gttcactgta cgaaaaaaag aggagtataa aatggcactg	3300
[1210]	tacaatctct atccaggtgt ttttgagaca gtggaaatgt taccatcaa agctggaatt	3360
[1211]	tggcggttgg aatgccttat tggcgagcat ctacatgtcg ggatgagcac actttttctg	3420
[1212]	gtgtacagca ataagtgtca gactcccctg ggaatggctt ctggacacat tagagatttt	3480
[1213]	cagattacag cttcaggaca atatggacag tgggccccaa agctggccag acttcattat	3540
[1214]	tccggatcaa tcaatgcctg gagcaccaag gagccctttt cttggatcaa ggtggatctg	3600
[1215]	ttggcaccaa tgattattca cggcatcaag acccagggtg cccgtcagaa gttctccagc	3660
[1216]	ctctacatct ctcatgttat catcatgtat agtcttgatg ggaagaagtg gcagacttat	3720
[1217]	cgaggaaatt ccaactggaac cttaatggtc ttctttggca atgtggattc atctgggata	3780

[1218]	aaacacaata tttttaaccc tccaattatt gctcgataca tccgtttgca cccaactcat	3840
[1219]	tatagcattc gcagcactct tcgcatggag ttgatgggct gtgatttaaa tagttgcagc	3900
[1220]	atgccattgg gaatggagag taaagcaata tcagatgcac agattactgc ttcatectac	3960
[1221]	tttaccaata tgtttgccac ctggtctcct tcaaaagctc gacttcacct ccaagggagg	4020
[1222]	agtaatgcct ggagacctca ggtgaataat ccaaaagagt ggctgcaagt ggacttccag	4080
[1223]	aagacaatga aagtcacagg agtaactact cagggagtaa aatctctgct taccagcatg	4140
[1224]	tatgtgaagg agttcctcat ctccagcagt caagatggcc atcagtggac tctctttttt	4200
[1225]	cagaatggca aagtaaagg ttttcagga aatcaagact ctttcacacc tgtggtgaac	4260
[1226]	tctctagacc caccgttact gactcgctac cttcgaattc acccccagag ttgggtgcac	4320
[1227]	cagattgccc tgaggatgga ggttctgggc tgcgaggcac aggacctcta c	4371
[1228]	<210> 17	
[1229]	<211> 1438	
[1230]	<212> PRT	
[1231]	<213> 人工序列	
[1232]	<220>	
[1233]	<223> BDD-FVIII (非优化;“亲本”),氨基酸序列	
[1234]	<400> 17	
[1235]	Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser Trp Asp Tyr	
[1236]	1 5 10 15	
[1237]	Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg Phe Pro Pro	
[1238]	20 25 30	
[1239]	Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val Tyr Lys Lys	
[1240]	35 40 45	
[1241]	Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile Ala Lys Pro	
[1242]	50 55 60	
[1243]	Arg Pro Pro Trp Met Gly Leu Leu Gly Pro Thr Ile Gln Ala Glu Val	
[1244]	65 70 75 80	
[1245]	Tyr Asp Thr Val Val Ile Thr Leu Lys Asn Met Ala Ser His Pro Val	
[1246]	85 90 95	
[1247]	Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ala Ser Glu Gly Ala	
[1248]	100 105 110	
[1249]	Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp Asp Lys Val	
[1250]	115 120 125	
[1251]	Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu Lys Glu Asn	
[1252]	130 135 140	
[1253]	Gly Pro Met Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser Tyr Leu Ser	
[1254]	145 150 155 160	
[1255]	His Val Asp Leu Val Lys Asp Leu Asn Ser Gly Leu Ile Gly Ala Leu	
[1256]	165 170 175	
[1257]	Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr Gln Thr Leu	
[1258]	180 185 190	
[1259]	His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly Lys Ser Trp	

[1260]	195	200	205
[1261]	His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp Ala Ala Ser		
[1262]	210	215	220
[1263]	Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr Val Asn Arg		
[1264]	225	230	235
[1265]	Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val Tyr Trp His		
[1266]	245	250	255
[1267]	Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile Phe Leu Glu		
[1268]	260	265	270
[1269]	Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser Leu Glu Ile		
[1270]	275	280	285
[1271]	Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met Asp Leu Gly		
[1272]	290	295	300
[1273]	Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His Asp Gly Met		
[1274]	305	310	315
[1275]	Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro Gln Leu Arg		
[1276]	325	330	335
[1277]	Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp Leu Thr Asp		
[1278]	340	345	350
[1279]	Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser Pro Ser Phe		
[1280]	355	360	365
[1281]	Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr Trp Val His		
[1282]	370	375	380
[1283]	Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro Leu Val Leu		
[1284]	385	390	395
[1285]	Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn Asn Gly Pro		
[1286]	405	410	415
[1287]	Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met Ala Tyr Thr		
[1288]	420	425	430
[1289]	Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu Ser Gly Ile		
[1290]	435	440	445
[1291]	Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu Leu Ile Ile		
[1292]	450	455	460
[1293]	Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro His Gly Ile		
[1294]	465	470	475
[1295]	Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys Gly Val Lys		
[1296]	485	490	495
[1297]	His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe Lys Tyr Lys		
[1298]	500	505	510
[1299]	Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp Pro Arg Cys		
[1300]	515	520	525
[1301]	Leu Thr Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg Asp Leu Ala		

[1302]	530	535	540
[1303]	Ser Gly Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu Ser Val Asp		
[1304]	545	550	555 560
[1305]	Gln Arg Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val Ile Leu Phe		
[1306]	565	570	575
[1307]	Ser Val Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu Asn Ile Gln		
[1308]	580	585	590
[1309]	Arg Phe Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp Pro Glu Phe		
[1310]	595	600	605
[1311]	Gln Ala Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val Phe Asp Ser		
[1312]	610	615	620
[1313]	Leu Gln Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp Tyr Ile Leu		
[1314]	625	630	635 640
[1315]	Ser Ile Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe Ser Gly Tyr		
[1316]	645	650	655
[1317]	Thr Phe Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr Leu Phe Pro		
[1318]	660	665	670
[1319]	Phe Ser Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro Gly Leu Trp		
[1320]	675	680	685
[1321]	Ile Leu Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly Met Thr Ala		
[1322]	690	695	700
[1323]	Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp Tyr Tyr Glu		
[1324]	705	710	715 720
[1325]	Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys Asn Asn Ala		
[1326]	725	730	735
[1327]	Ile Glu Pro Arg Ser Phe Ser Gln Asn Pro Pro Val Leu Lys Arg His		
[1328]	740	745	750
[1329]	Gln Arg Glu Ile Thr Arg Thr Thr Leu Gln Ser Asp Gln Glu Glu Ile		
[1330]	755	760	765
[1331]	Asp Tyr Asp Asp Thr Ile Ser Val Glu Met Lys Lys Glu Asp Phe Asp		
[1332]	770	775	780
[1333]	Ile Tyr Asp Glu Asp Glu Asn Gln Ser Pro Arg Ser Phe Gln Lys Lys		
[1334]	785	790	795 800
[1335]	Thr Arg His Tyr Phe Ile Ala Ala Val Glu Arg Leu Trp Asp Tyr Gly		
[1336]	805	810	815
[1337]	Met Ser Ser Ser Pro His Val Leu Arg Asn Arg Ala Gln Ser Gly Ser		
[1338]	820	825	830
[1339]	Val Pro Gln Phe Lys Lys Val Val Phe Gln Glu Phe Thr Asp Gly Ser		
[1340]	835	840	845
[1341]	Phe Thr Gln Pro Leu Tyr Arg Gly Glu Leu Asn Glu His Leu Gly Leu		
[1342]	850	855	860
[1343]	Leu Gly Pro Tyr Ile Arg Ala Glu Val Glu Asp Asn Ile Met Val Thr		

[1344]	865	870	875	880
[1345]	Phe Arg Asn Gln Ala Ser Arg Pro Tyr Ser Phe Tyr Ser Ser Leu Ile			
[1346]		885	890	895
[1347]	Ser Tyr Glu Glu Asp Gln Arg Gln Gly Ala Glu Pro Arg Lys Asn Phe			
[1348]		900	905	910
[1349]	Val Lys Pro Asn Glu Thr Lys Thr Tyr Phe Trp Lys Val Gln His His			
[1350]		915	920	925
[1351]	Met Ala Pro Thr Lys Asp Glu Phe Asp Cys Lys Ala Trp Ala Tyr Phe			
[1352]		930	935	940
[1353]	Ser Asp Val Asp Leu Glu Lys Asp Val His Ser Gly Leu Ile Gly Pro			
[1354]	945	950	955	960
[1355]	Leu Leu Val Cys His Thr Asn Thr Leu Asn Pro Ala His Gly Arg Gln			
[1356]		965	970	975
[1357]	Val Thr Val Gln Glu Phe Ala Leu Phe Phe Thr Ile Phe Asp Glu Thr			
[1358]		980	985	990
[1359]	Lys Ser Trp Tyr Phe Thr Glu Asn Met Glu Arg Asn Cys Arg Ala Pro			
[1360]		995	1000	1005
[1361]	Cys Asn Ile Gln Met Glu Asp Pro Thr Phe Lys Glu Asn Tyr Arg			
[1362]		1010	1015	1020
[1363]	Phe His Ala Ile Asn Gly Tyr Ile Met Asp Thr Leu Pro Gly Leu			
[1364]		1025	1030	1035
[1365]	Val Met Ala Gln Asp Gln Arg Ile Arg Trp Tyr Leu Leu Ser Met			
[1366]		1040	1045	1050
[1367]	Gly Ser Asn Glu Asn Ile His Ser Ile His Phe Ser Gly His Val			
[1368]		1055	1060	1065
[1369]	Phe Thr Val Arg Lys Lys Glu Glu Tyr Lys Met Ala Leu Tyr Asn			
[1370]		1070	1075	1080
[1371]	Leu Tyr Pro Gly Val Phe Glu Thr Val Glu Met Leu Pro Ser Lys			
[1372]		1085	1090	1095
[1373]	Ala Gly Ile Trp Arg Val Glu Cys Leu Ile Gly Glu His Leu His			
[1374]		1100	1105	1110
[1375]	Ala Gly Met Ser Thr Leu Phe Leu Val Tyr Ser Asn Lys Cys Gln			
[1376]		1115	1120	1125
[1377]	Thr Pro Leu Gly Met Ala Ser Gly His Ile Arg Asp Phe Gln Ile			
[1378]		1130	1135	1140
[1379]	Thr Ala Ser Gly Gln Tyr Gly Gln Trp Ala Pro Lys Leu Ala Arg			
[1380]		1145	1150	1155
[1381]	Leu His Tyr Ser Gly Ser Ile Asn Ala Trp Ser Thr Lys Glu Pro			
[1382]		1160	1165	1170
[1383]	Phe Ser Trp Ile Lys Val Asp Leu Leu Ala Pro Met Ile Ile His			
[1384]		1175	1180	1185
[1385]	Gly Ile Lys Thr Gln Gly Ala Arg Gln Lys Phe Ser Ser Leu Tyr			

[1386]	1190	1195	1200
[1387]	Ile Ser Gln Phe Ile Ile Met Tyr Ser Leu Asp Gly Lys Lys Trp		
[1388]	1205	1210	1215
[1389]	Gln Thr Tyr Arg Gly Asn Ser Thr Gly Thr Leu Met Val Phe Phe		
[1390]	1220	1225	1230
[1391]	Gly Asn Val Asp Ser Ser Gly Ile Lys His Asn Ile Phe Asn Pro		
[1392]	1235	1240	1245
[1393]	Pro Ile Ile Ala Arg Tyr Ile Arg Leu His Pro Thr His Tyr Ser		
[1394]	1250	1255	1260
[1395]	Ile Arg Ser Thr Leu Arg Met Glu Leu Met Gly Cys Asp Leu Asn		
[1396]	1265	1270	1275
[1397]	Ser Cys Ser Met Pro Leu Gly Met Glu Ser Lys Ala Ile Ser Asp		
[1398]	1280	1285	1290
[1399]	Ala Gln Ile Thr Ala Ser Ser Tyr Phe Thr Asn Met Phe Ala Thr		
[1400]	1295	1300	1305
[1401]	Trp Ser Pro Ser Lys Ala Arg Leu His Leu Gln Gly Arg Ser Asn		
[1402]	1310	1315	1320
[1403]	Ala Trp Arg Pro Gln Val Asn Asn Pro Lys Glu Trp Leu Gln Val		
[1404]	1325	1330	1335
[1405]	Asp Phe Gln Lys Thr Met Lys Val Thr Gly Val Thr Thr Gln Gly		
[1406]	1340	1345	1350
[1407]	Val Lys Ser Leu Leu Thr Ser Met Tyr Val Lys Glu Phe Leu Ile		
[1408]	1355	1360	1365
[1409]	Ser Ser Ser Gln Asp Gly His Gln Trp Thr Leu Phe Phe Gln Asn		
[1410]	1370	1375	1380
[1411]	Gly Lys Val Lys Val Phe Gln Gly Asn Gln Asp Ser Phe Thr Pro		
[1412]	1385	1390	1395
[1413]	Val Val Asn Ser Leu Asp Pro Pro Leu Leu Thr Arg Tyr Leu Arg		
[1414]	1400	1405	1410
[1415]	Ile His Pro Gln Ser Trp Val His Gln Ile Ala Leu Arg Met Glu		
[1416]	1415	1420	1425
[1417]	Val Leu Gly Cys Glu Ala Gln Asp Leu Tyr		
[1418]	1430	1435	
[1419]	<210> 18		
[1420]	<211> 450		
[1421]	<212> DNA		
[1422]	<213> 人工序列		
[1423]	<220>		
[1424]	<223> XTEN核苷酸序列		
[1425]	<400> 18		
[1426]	ggcgcgccaa catcagagag cgccaccct gaaagtggtc ccgggagcga gccagccaca 60		
[1427]	tctgggtcgg aaacgccagg cacaagttag tctgcaactc ccgagtccgg acctggctcc 120		

[1428]	gagcctgcc	ctagcggctc	cgagactccg	ggaacttccg	agagcgctac	accagaaagc	180
[1429]	ggacccggaa	ccagtaccga	acctagcgag	ggctctgctc	cgggcagccc	agccggctct	240
[1430]	cctacatcca	cggaggaggg	cacttccgaa	tccgccaccc	cggagtcagg	gccaggatct	300
[1431]	gaaccgcgta	cctcaggcag	tgagacgcca	ggaacgagcg	agtccgctac	accggagagt	360
[1432]	gggccaggga	gccctgctgg	atctcctacg	tccactgagg	aagggtcacc	agcgggctcg	420
[1433]	cccaccagca	ctgaagaagg	tgccctcgagc				450
[1434]	<210>	19					
[1435]	<211>	4824					
[1436]	<212>	DNA					
[1437]	<213>	人工序列					
[1438]	<220>						
[1439]	<223>	coFVIII-52-XTEN					
[1440]	<400>	19					
[1441]	atgcaaateg	aactgagcac	ctgtttcttc	ctctgcctgc	tgagattctg	tttctccgcg	60
[1442]	acccgccgat	actacctggg	agcagtggag	ctctcctggg	attacatgca	gagcgacctt	120
[1443]	ggggagctgc	ccgtggatgc	caggttccct	ccccgggtgc	caaagtcgtt	tccgttcaac	180
[1444]	acctccgtgg	tgtacaagaa	aactctgttc	gtggagtcca	ccgaccacct	gttcaatatc	240
[1445]	gccaaagcca	gacctccctg	gatggggctg	ttgggacctc	ccatccaagc	ggaggtgtac	300
[1446]	gacactgtgg	tcatcactct	gaagaacatg	gcctcgcac	ccgtgtccct	gcacgccgtg	360
[1447]	ggagtgtctt	actggaaagc	gtccgagggg	gccgaatacg	acgaccagac	ctcgcagaga	420
[1448]	gaaaaggaag	atgacaaggt	gttcccagga	ggatcgcaca	cctacgtgtg	gcaagtgttg	480
[1449]	aaggagaacg	gccaatggc	ctccgaccgc	ctgtgcctga	cctactcgta	cctgtcccac	540
[1450]	gtggacctcg	tgaaggacct	caactcggga	ctgattggag	ccctgctggt	ctgcagggaa	600
[1451]	ggctcactgg	cgaagaaaaa	gactcagacc	ttgcacaagt	tcattctgct	gttcgctgtg	660
[1452]	ttcgacgagg	ggaagtcgtg	gcacagcgag	actaagaact	ccctgatgca	agatagagat	720
[1453]	gccgcctccg	cccgggcctg	gcctaagatg	cacaccgtga	acggttacgt	gaaccgctcc	780
[1454]	ctccctggcc	tgattggatg	ccaccggaag	tccgtgtact	ggcacgtgat	cgggatgggg	840
[1455]	accacccccg	aggtgcacag	catcttcctg	gaaggtcaca	catttctcgt	gcgcaaccac	900
[1456]	cggcaggcct	ccctggaaat	cagccccatt	accttctcca	ctgcccagac	tctgctgatg	960
[1457]	gacctgggac	agttcctgct	gttctgcca	atctcctccc	accaacatga	cggaatggag	1020
[1458]	gcatacgtga	aggtcgattc	ctgccctgag	gaacccagc	tccgcatgaa	gaacaatgag	1080
[1459]	gaagccgagg	actacgacga	cgacctgacg	gatagcgaga	tgatgtggt	ccggttcgat	1140
[1460]	gacgataaca	gcccttcctt	catccaaatt	cgctcgggtg	caaagaagca	ccccaaagacc	1200
[1461]	tggtgctatt	acattgcggc	ggaagaagag	gactgggatt	atgccccgct	tgctctcgct	1260
[1462]	cctgacgacc	ggagctacaa	gagccagtac	ctgaacaacg	gtccacagag	gatcggtaga	1320
[1463]	aagtacaaga	aggtccgctt	catggcctat	accgacgaaa	ccttcaaaac	tagagaggcc	1380
[1464]	atccaacacg	aatccggcat	cctgggcccg	ctcttgtagc	gagaagtcgg	cgacaccctt	1440
[1465]	ctcattatct	tcaagaacca	ggcttcccgg	ccgtacaaca	tctatccgca	tgggatcact	1500
[1466]	gacgtgcgcc	cactgtactc	gcggcgccctg	cccaagggtg	tcaaacacct	gaaggatttt	1560
[1467]	ccgatccttc	cgggagaaat	cttcaagtac	aagtggaccg	tgaccgtgga	agatggccca	1620
[1468]	actaagtctg	accctagatg	cctcacccgc	tactactcat	ccttcgtcaa	catggagcgc	1680
[1469]	gacctggcca	gcggactgat	cggcccgcgtg	ctgatttgct	acaaggaatc	agtggaacaa	1740

[1470]	cggggaaacc agatcatgtc ggataagagg aacgtcatcc tcttctccgt gtttgacgaa	1800
[1471]	aaccggtcgt ggtacctgac cgagaacatc cagaggttcc tgcccaaccc tgctgggggtg	1860
[1472]	cagctggagg accccgagtt ccaggccagc aacatcatgc acagcatcaa tggctacgtg	1920
[1473]	ttcgacagcc tgcagctgag cgtgtgcctg cagcaggtgg cctactggta catcctgagc	1980
[1474]	atcggcgccc agaccgactt cctgagcgtg ttcttctctg gctacacctt caagcacaag	2040
[1475]	atggtgtatg aggacaccct gaccctgttc cccttcagcg gggagactgt cttcatgagc	2100
[1476]	atggagaacc ctggcctgtg gatcctgggc tgccacaaca ggcacttcag gaacaggggc	2160
[1477]	atgactgccc tgctgaaagt ctccagctgt gacaagaaca ccggggacta ctacgaggac	2220
[1478]	agctacgagg acatcagcgc ctacctgtg agcaagaaca atgccatcga gcccgaggagc	2280
[1479]	ttctctcaga acggcgcgcc aacatcagag agcgccaccc ctgaaagtgg tcccgggagc	2340
[1480]	gagccagcca catctgggtc ggaaacgcca ggcacaagtg agtctgcaac tcccgagtcc	2400
[1481]	ggacctggct ccgagcctgc cactagcggc tccgagactc cgggaacttc cgagagcgt	2460
[1482]	acaccagaaa gcggaccggg aaccagtacc gaacctagcg agggctctgc tccgggcagc	2520
[1483]	ccagccggct ctctacatc cacggaggag ggcacttcg aatccgccac cccggagtca	2580
[1484]	gggccaggat ctgaacccgc tacctcaggc agtgagacgc caggaacgag cgagtccgct	2640
[1485]	acaccggaga gtgggccagg gagccctgct ggatctccta cgtccactga ggaagggtca	2700
[1486]	ccagcgggct cggccaccag cactgaagaa ggtgcctcga gccccccagt gctgaagagg	2760
[1487]	caccagaggg agatcaccag gaccaccctg cagtctgacc aggaggagat cgactatgat	2820
[1488]	gacaccatca gcgtggagat gaagaaggag gacttcgaca tctacgacga ggacgagaac	2880
[1489]	cagagcccca ggagcttcca gaagaagacc aggcaactt catttctgctg tgtggagagg	2940
[1490]	ctgtgggact atggcatgtc cagcagcccc catgtgtgta ggaacagggc ccagtctggc	3000
[1491]	agcgtgcccc agttcaagaa agtcgtgttc caggagttca ccgacggcag cttcaccag	3060
[1492]	ccccgtaca gaggggagct gaacgagcac ctgggcctgc tgggccccta catcagggcc	3120
[1493]	gaggtggagg acaacatcat ggtgacctc aggaaccagg ccagcaggcc ctacagcttc	3180
[1494]	tacagcagcc tgatcagcta cgaggaggac cagaggcagg gggctgagcc caggaagaac	3240
[1495]	tttgtgaagc ccaatgaaac caagacctac ttctggaagg tgcagacca catggcccc	3300
[1496]	accaaggacg agttcgactg caaggcctgg gcctacttct ctgacgtgga cctggagaag	3360
[1497]	gacgtgcaact ctggcctgat tggccccctg ctggtgtgcc acaccaacac cctgaaccct	3420
[1498]	gcccattgca ggcagtgac tgtgcaggag ttgcctgt tcttcaccat cttegatgaa	3480
[1499]	accaagagct ggtacttcac tgagaacatg gagaggaaact gcagggcccc ctgcaacatc	3540
[1500]	cagatggagg accccacctt caaggagaac tacaggttcc atgccatcaa tggctacatc	3600
[1501]	atggacaccc tgcctggcct ggtcatggcc caggaccaga ggatcaggtg gtatctgctg	3660
[1502]	agcatgggca gcaacgagaa catccacagc atccacttct ctggccacgt gttcactgtg	3720
[1503]	aggaagaagg aggagtacaa gatggccctg tacaacctgt accctggggg gttcgaaacc	3780
[1504]	gtggagatgc tgcccagcaa ggccggcatc tggagggtgg agtgcctgat tggggagcac	3840
[1505]	ctgcacgccg gcatgagcac cctgttctctg gtgtacagca acaagtgcc gacccccctg	3900
[1506]	ggcatggcct ctggccacat cagggaactc cagatcactg cctctggcca gtacggccag	3960
[1507]	tgggccccca agctggccag gctgcactac tccggaagca tcaatgcctg gagcaccaag	4020
[1508]	gagcccttca gctggatcaa agtggacctg ctggccccca tgatcatcca cggcatcaag	4080
[1509]	accagggggg ccaggcagaa gttctccagc ctgtacatca gccagttcat catcatgtac	4140
[1510]	agcctggacg gcaagaagtg gcagacctac aggggcaaca gcaccggcac cctgatgggtg	4200
[1511]	ttcttcggca acgtggacag cagcggcatc aagcacaaca tcttcaaccc ccccatcatc	4260

[1512]	gccagataca tcaggctgca cccacccac tacagcatca ggagcaccct gaggatggag	4320
[1513]	ctgatgggct gtgacctgaa cagctgcagc atgcccctgg gcatggagag caaggccatc	4380
[1514]	tctgacgcc agatcactgc ctccagctac ttaccaaca tgtttgccac ctggagcccc	4440
[1515]	agcaaggcca ggctgcacct gcagggcagg agcaatgcct ggaggcccca ggtcaacaac	4500
[1516]	cccaaggagt ggctgcaggt ggacttccag aagacatga aggtgactgg ggtgaccacc	4560
[1517]	cagggggtga agagcctgct gaccagcatg tacgtgaagg agttcctgat ctccagcagc	4620
[1518]	caggacggcc accagtggac cctgttcttc cagaatggca aggtgaaggt gttccagggc	4680
[1519]	aaccaggaca gttcacccc tgtgttcaac agcctggacc cccccctgct gaccagatac	4740
[1520]	ctgaggatcc acccccagag ctgggtgcac cagatcgccc tgaggatgga ggtgctgggc	4800
[1521]	tgtgaggccc aggacctgta ctga	4824
[1522]	<210>	20
[1523]	<211>	4824
[1524]	<212>	DNA
[1525]	<213>	人工序列
[1526]	<220>	
[1527]	<223>	coFVIII-1-XTEN
[1528]	<400>	20
[1529]	atgcagattg agctgtctac ttgcttttct ctgtgcctgc tgaggttttg cttttccgct	60
[1530]	acacgaaggt attatctggg ggctgtggaa ctgtcttggg attacatgca gagtgcctg	120
[1531]	ggagagctgc cagtggacgc aaggtttccc cctagagtcc ctaagtcatt ccccttcaac	180
[1532]	actagcgtgg tctacaagaa aacactgttc gtggagttta ctgatcacct gttcaacatc	240
[1533]	gcaaagccta ggccaccctg gatgggactg ctggggccaa caatccaggc cgaggtgtac	300
[1534]	gacaccgtgg tcattacact taagaacatg gcctcacacc ccgtgagcct gcatgctgtg	360
[1535]	ggcgtcagct actggaaggc ttccgaagga gcagagtatg acgatcagac ttcccagaga	420
[1536]	gaaaaagagg acgataaggt gtttcttggc ggatctcata cctacgtgtg gcaggtcctg	480
[1537]	aaagagaatg gccctatggc ctccgaccct ctgtgcctga cctactctta tctgagtcac	540
[1538]	gtggacctgg tcaaggatct gaacagcggc ctgatcggag ccctgctggt gtgcagggaa	600
[1539]	ggaagcctgg ctaaggagaa aaccagaca ctgcataagt tcattctgct gttcgccgtg	660
[1540]	tttgacgaag ggaaatcatg gcacagcgag acaaagaata gtctgatgca ggacagggat	720
[1541]	gccgttcag ccagagcttg gcccaaaatg cactgtga acggctacgt caatcgtca	780
[1542]	ctgcctgggc tgatcgctg ccaccgaaag agcgtgtatt ggcatgtcat cgggatgggc	840
[1543]	accacacctg aagtgcactc cattttcctg gagggacata cttttctggt ccgcaaccac	900
[1544]	cgacaggctt ccctggagat ctctccaatt accttctga cagcacagac tctgctgatg	960
[1545]	gacctggggc agttcctgct gttttgccac atcagctccc accagcatga tggcatggag	1020
[1546]	gcttacgtga aagtggactc ttgtcccag gaacctcagc tgcggatgaa gaacaatgag	1080
[1547]	gaagcagaag actatgacga tgacctgacc gactccgaga tggatgtggt ccgattcgat	1140
[1548]	gacgataaca gccctcctt tatccagatt agatctgtgg ccaagaaaca ccctaagaca	1200
[1549]	tgggtccatt acatgcagc cgaggaagag gactgggatt atgcaccact ggtgctggca	1260
[1550]	ccagacgatc gtcctacaa atctcagtat ctgaacaatg ggccacagag gattggcaga	1320
[1551]	aagtacaaga aagtgcggtt catggcatat accgatgaga cttcaagac tcgcaagcc	1380
[1552]	atccagcacg agagcgcat cctgggacca ctgctgtacg gagaagtggg agacaccctg	1440
[1553]	ctgatcattt tcaagaacca ggccagccgg cttacaata tctatccaca tgggattaca	1500

[1554]	gatgtgcgcc ctctgtacag caggagactg ccaaagggcg tcaaacacct gaaggacttc	1560
[1555]	ccaatcctgc cgggagaaat cttcaagtac aagtggactg tcaccgtcga ggatggcccc	1620
[1556]	actaagagcg accctcgggtg cctgaccgcg tactattcta gtttcgtgaa tatggaaaga	1680
[1557]	gatctggcaa gcggactgat cggaccactg ctgatttgtt acaaagagag cgtggatcag	1740
[1558]	agaggcaacc agatcatgtc cgacaagcgg aatgtgattc tgttcagtgt ctttgacgaa	1800
[1559]	aacaggtcat ggtacctgac cgagaacatc cagagattcc tgcctaattc agctgggggtg	1860
[1560]	cagctggaag atcctgagtt tcaggcatct aacatcatgc atagtattaa tggctacgtg	1920
[1561]	ttcgacagtt tgcagctgag cgtgtgcctg cagcaggtcg cttactggta taccctgagc	1980
[1562]	attggggcac agacagattt cctgagcgtg ttcttttccg gctacacttt taagcataaa	2040
[1563]	atggtctatg aggacacact gactctgttc cccttcagcg gcgaaaccgt gtttatgagc	2100
[1564]	atggagaatc cgggactgtg gattctgggg tgccacaaca gcgatttcag aaatcgcgga	2160
[1565]	atgactgcc tgcctgaaagt gtcaagctgt gacaagaaca ccggggacta ctatgaagat	2220
[1566]	tcatacagg acatcagcgc atatctgtg tccaaaaaca atgccattga accccggtct	2280
[1567]	tttagtcaga atggcgcgcc aacatcagag agcggcacc ctgaaagtgg tcccgggagc	2340
[1568]	gagccagcca catctgggtc ggaaacgcca ggcacaagt agtctgcaac tcccaggtcc	2400
[1569]	ggacctggct ccgagcctgc cactagcggc tccgagactc cgggaacttc cgagagcgt	2460
[1570]	acaccagaaa gcggaccgg aaccagtacc gaacctagcg agggctctgc tccgggcagc	2520
[1571]	ccagccggt ctcctacatc cagggaggag ggcacttcg aatccgccac cccggagtca	2580
[1572]	gggccaggat ctgaacccgc tacctcaggc agtgagacgc caggaacgag cgagtccgt	2640
[1573]	acaccggaga gtgggccagg gacccctgct ggatctccta cgtccactga ggaagggtca	2700
[1574]	ccagcgggct cgcccaccag cactgaagaa ggtgcctcga gccctccagt gctgaagcgg	2760
[1575]	caccagcgcg agatcacccg cactaccctg cagagtgatc aggaagagat cgactacgac	2820
[1576]	gatacaattt ctgtggaaat gaagaaagag gacttcgata tctatgacga agatgagaac	2880
[1577]	cagagtccct gatcattcca gaagaaaacc aggcttact ttattgccgc agtggagcgg	2940
[1578]	ctgtgggatt atggcatgtc ctctagtcct cagtgctgc gaaataggc ccagtcagga	3000
[1579]	agcgtcccac agttcaagaa agtggctctc caggagtta cagacgggtc ctttactcag	3060
[1580]	ccactgtaca ggggcgaact gaacgagcac ctgggactgc tggggcccta tatcagagca	3120
[1581]	gaagtggagg ataacattat ggtcaccttc agaaatcagg cctctcggcc ttacagtttt	3180
[1582]	tattcaagcc tgatctctta cgaagaggac cagcgacagg gagctgaacc acgaaaaaac	3240
[1583]	ttcgtgaagc ctaatgagac caaacatac ttttggaagg tgcagacca tatggcccca	3300
[1584]	acaaaagacg agttcgattg caaggcatgg gcctattttt ctgacgtgga tctggagaag	3360
[1585]	gacgtgcaca gtggcctgat tggccactg ctggtgtgcc atactaacac cctgaatcca	3420
[1586]	gcccacggcc ggcaggtcac tgtccaggag ttcgctctgt tctttaccat ctttgatgag	3480
[1587]	acaaagagct ggtacttcac cgaaaacatg gagcgaaatt gcagggtcc atgtaacatt	3540
[1588]	cagatggaag accccacatt caaggagaac taccgcttc atgctatcaa tggatacatc	3600
[1589]	atggatactc tgcccgggt ggtcatggca caggaccaga gaatccgggt gtatctgctg	3660
[1590]	agcatgggca gcaacgagaa tatccactca attcatttca gcgggcacgt gtttactgtc	3720
[1591]	aggaagaaag aagagtacaa gatggccctg tacaacctgt atcccggcgt gttcgaaacc	3780
[1592]	gtcgagatgc tgcctagcaa ggccggaatc tggagagtgg aatgcctgat tggagagcac	3840
[1593]	ctgcatgctg ggatgtctac cctgtttctg gtgtacagta ataagtgtca gacaccctg	3900
[1594]	ggaatggcat ccgggcatat cagggtttc cagattaccg catctggaca gtacggacag	3960
[1595]	tgggcaccta agctggctag actgcactat tccgatcta tcaacgctt gtccacaaaa	4020

[1596] gagcctttct cttggattaa ggtggacctg ctggcccaaa tgatcattca tggcatcaaa 4080
[1597] actcagggag ctcggcagaa gttctcctct ctgtacatct cacagtttat catcatgtac 4140
[1598] agcctggatg ggaagaaatg gcagacatac cgcggcaata gcacaggaaac tctgatgggtg 4200
[1599] ttctttggca acgtggacag cagcggaatc aagcacaaca ttttcaatcc ccctatcatt 4260
[1600] gctagatata tccggctgca cccaacccat tattctattc gaagtacact gaggatggaa 4320
[1601] ctgatgggat gcgatctgaa cagttgttca atgccctgg ggatggagtc caaggcaatc 4380
[1602] tctgacgccc agattaccgc cagctcctac ttactaata tgtttgctac ctggagccct 4440
[1603] tccaaagcaa gactgcacct gcaaggccgc agcaacgcat ggcgaccaca ggtgaacaat 4500
[1604] cccaaggagt ggttgacagt cgattttcag aaaactatga aggtgaccgg ggtcacaact 4560
[1605] cagggcgtga aaagtctgct gacctcaatg tacgtcaagg agttcctgat ctctagttca 4620
[1606] caggacggac atcagtggac actgttcttt cagaacggga aggtgaaagt cttccagggc 4680
[1607] aatcaggatt cttttacacc tgttgtcaac agtctagacc ctccactgct gaccagatac 4740
[1608] ctgagaatcc accctcagtc ctgggtgcac cagattgccc tgagaatgga agtgctggga 4800
[1609] tgcgaggccc aggatctgta ctga 4824
[1610] <210> 21
[1611] <211> 6
[1612] <212> DNA
[1613] <213> 人工序列
[1614] <220>
[1615] <223> MAR/ARS
[1616] <400> 21
[1617] atatatt 6
[1618] <210> 22
[1619] <211> 6
[1620] <212> DNA
[1621] <213> 人工序列
[1622] <220>
[1623] <223> MAR/ARS
[1624] <400> 22
[1625] aaatat 6
[1626] <210> 23
[1627] <211> 5
[1628] <212> DNA
[1629] <213> 人工序列
[1630] <220>
[1631] <223> 去稳定元件
[1632] <400> 23
[1633] attta 5
[1634] <210> 24
[1635] <211> 5
[1636] <212> DNA
[1637] <213> 人工序列

[1638]	<220>
[1639]	<223> 去稳定元件
[1640]	<400> 24
[1641]	taaat 5
[1642]	<210> 25
[1643]	<211> 6
[1644]	<212> DNA
[1645]	<213> 人工序列
[1646]	<220>
[1647]	<223> poly-T位点
[1648]	<400> 25
[1649]	tttttt 6
[1650]	<210> 26
[1651]	<211> 7
[1652]	<212> DNA
[1653]	<213> 人工序列
[1654]	<220>
[1655]	<223> poly-A位点
[1656]	<400> 26
[1657]	aaaaaaa 7
[1658]	<210> 27
[1659]	<211> 6
[1660]	<212> DNA
[1661]	<213> 人工序列
[1662]	<220>
[1663]	<223> 剪接位点
[1664]	<400> 27
[1665]	ggtgat 6
[1666]	<210> 28
[1667]	<211> 5
[1668]	<212> DNA
[1669]	<213> 人工序列
[1670]	<220>
[1671]	<223> TATA盒
[1672]	<400> 28
[1673]	tataa 5
[1674]	<210> 29
[1675]	<211> 5
[1676]	<212> DNA
[1677]	<213> 人工序列
[1678]	<220>
[1679]	<223> TATA盒

[1680]	<400> 29
[1681]	ttata 5
[1682]	<210> 30
[1683]	<211> 8
[1684]	<212> DNA
[1685]	<213> 人工序列
[1686]	<220>
[1687]	<223> 富AU序列元件
[1688]	<400> 30
[1689]	atatttatt 8
[1690]	<210> 31
[1691]	<211> 8
[1692]	<212> DNA
[1693]	<213> 人工序列
[1694]	<220>
[1695]	<223> 富AU序列元件
[1696]	<400> 31
[1697]	atatttta 8
[1698]	<210> 32
[1699]	<211> 13
[1700]	<212> DNA
[1701]	<213> 人工序列
[1702]	<220>
[1703]	<223> kozak共有序列
[1704]	<400> 32
[1705]	gccgccacca tgc 13
[1706]	<210> 33
[1707]	<211> 32
[1708]	<212> PRT
[1709]	<213> 人工序列
[1710]	<220>
[1711]	<223> CTP肽
[1712]	<400> 33
[1713]	Asp Pro Arg Phe Gln Asp Ser Ser Ser Ser Lys Ala Pro Pro Pro Ser
[1714]	1 5 10 15
[1715]	Leu Pro Ser Pro Ser Arg Leu Pro Gly Pro Ser Asp Thr Pro Ile Leu
[1716]	20 25 30
[1717]	<210> 34
[1718]	<211> 28
[1719]	<212> PRT
[1720]	<213> 人工序列
[1721]	<220>

[1722]	<223> CTP肽
[1723]	<400> 34
[1724]	Ser Ser Ser Ser Lys Ala Pro Pro Pro Ser Leu Pro Ser Pro Ser Arg
[1725]	1 5 10 15
[1726]	Leu Pro Gly Pro Ser Asp Thr Pro Ile Leu Pro Gln
[1727]	20 25
[1728]	<210> 35
[1729]	<211> 11
[1730]	<212> PRT
[1731]	<213> 人工序列
[1732]	<220>
[1733]	<223> 白蛋白结合肽核心序列
[1734]	<400> 35
[1735]	Asp Ile Cys Leu Pro Arg Trp Gly Cys Leu Trp
[1736]	1 5 10
[1737]	<210> 36
[1738]	<211> 20
[1739]	<212> PRT
[1740]	<213> 人工序列
[1741]	<220>
[1742]	<223> PAS序列
[1743]	<400> 36
[1744]	Ala Ser Pro Ala Ala Pro Ala Pro Ala Ser Pro Ala Ala Pro Ala Pro
[1745]	1 5 10 15
[1746]	Ser Ala Pro Ala
[1747]	20
[1748]	<210> 37
[1749]	<211> 20
[1750]	<212> PRT
[1751]	<213> 人工序列
[1752]	<220>
[1753]	<223> PAS序列
[1754]	<400> 37
[1755]	Ala Ala Pro Ala Ser Pro Ala Pro Ala Ala Pro Ser Ala Pro Ala Pro
[1756]	1 5 10 15
[1757]	Ala Ala Pro Ser
[1758]	20
[1759]	<210> 38
[1760]	<211> 20
[1761]	<212> PRT
[1762]	<213> 人工序列
[1763]	<220>

[1764]	<223> PAS序列
[1765]	<400> 38
[1766]	Ala Pro Ser Ser Pro Ser Pro Ser Ala Pro Ser Ser Pro Ser Pro Ala
[1767]	1 5 10 15
[1768]	Ser Pro Ser Ser
[1769]	20
[1770]	<210> 39
[1771]	<211> 19
[1772]	<212> PRT
[1773]	<213> 人工序列
[1774]	<220>
[1775]	<223> PAS序列
[1776]	<400> 39
[1777]	Ala Pro Ser Ser Pro Ser Pro Ser Ala Pro Ser Ser Pro Ser Pro Ala
[1778]	1 5 10 15
[1779]	Ser Pro Ser
[1780]	<210> 40
[1781]	<211> 20
[1782]	<212> PRT
[1783]	<213> 人工序列
[1784]	<220>
[1785]	<223> PAS序列
[1786]	<400> 40
[1787]	Ser Ser Pro Ser Ala Pro Ser Pro Ser Ser Pro Ala Ser Pro Ser Pro
[1788]	1 5 10 15
[1789]	Ser Ser Pro Ala
[1790]	20
[1791]	<210> 41
[1792]	<211> 24
[1793]	<212> PRT
[1794]	<213> 人工序列
[1795]	<220>
[1796]	<223> PAS序列
[1797]	<400> 41
[1798]	Ala Ala Ser Pro Ala Ala Pro Ser Ala Pro Pro Ala Ala Ala Ser Pro
[1799]	1 5 10 15
[1800]	Ala Ala Pro Ser Ala Pro Pro Ala
[1801]	20
[1802]	<210> 42
[1803]	<211> 20
[1804]	<212> PRT
[1805]	<213> 人工序列

[1806]	<220>
[1807]	<223> PAS序列
[1808]	<400> 42
[1809]	Ala Ser Ala Ala Ala Pro Ala Ala Ala Ser Ala Ala Ala Ser Ala Pro
[1810]	1 5 10 15
[1811]	Ser Ala Ala Ala
[1812]	20
[1813]	<210> 43
[1814]	<211> 23
[1815]	<212> DNA
[1816]	<213> 人工序列
[1817]	<220>
[1818]	<223> miR142标靶
[1819]	<400> 43
[1820]	tccataaagt aggaacact aca 23
[1821]	<210> 44
[1822]	<211> 2332
[1823]	<212> PRT
[1824]	<213> 智人
[1825]	<400> 44
[1826]	Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser Trp Asp Tyr
[1827]	1 5 10 15
[1828]	Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg Phe Pro Pro
[1829]	20 25 30
[1830]	Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val Tyr Lys Lys
[1831]	35 40 45
[1832]	Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile Ala Lys Pro
[1833]	50 55 60
[1834]	Arg Pro Pro Trp Met Gly Leu Leu Gly Pro Thr Ile Gln Ala Glu Val
[1835]	65 70 75 80
[1836]	Tyr Asp Thr Val Val Ile Thr Leu Lys Asn Met Ala Ser His Pro Val
[1837]	85 90 95
[1838]	Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ala Ser Glu Gly Ala
[1839]	100 105 110
[1840]	Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp Asp Lys Val
[1841]	115 120 125
[1842]	Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu Lys Glu Asn
[1843]	130 135 140
[1844]	Gly Pro Met Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser Tyr Leu Ser
[1845]	145 150 155 160
[1846]	His Val Asp Leu Val Lys Asp Leu Asn Ser Gly Leu Ile Gly Ala Leu
[1847]	165 170 175

[1848]	Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr Gln Thr Leu		
[1849]		180	185
[1850]	His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly Lys Ser Trp		190
[1851]		195	200
[1852]	His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp Ala Ala Ser		205
[1853]		210	215
[1854]	Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr Val Asn Arg		220
[1855]		225	230
[1856]	Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val Tyr Trp His		235
[1857]		245	250
[1858]	Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile Phe Leu Glu		255
[1859]		260	265
[1860]	Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser Leu Glu Ile		270
[1861]		275	280
[1862]	Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met Asp Leu Gly		285
[1863]		290	295
[1864]	Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His Asp Gly Met		300
[1865]		305	310
[1866]	Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro Gln Leu Arg		315
[1867]		325	330
[1868]	Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp Leu Thr Asp		335
[1869]		340	345
[1870]	Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser Pro Ser Phe		350
[1871]		355	360
[1872]	Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr Trp Val His		365
[1873]		370	375
[1874]	Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro Leu Val Leu		380
[1875]		385	390
[1876]	Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn Asn Gly Pro		395
[1877]		405	410
[1878]	Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met Ala Tyr Thr		415
[1879]		420	425
[1880]	Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu Ser Gly Ile		430
[1881]		435	440
[1882]	Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu Leu Ile Ile		445
[1883]		450	455
[1884]	Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro His Gly Ile		460
[1885]		465	470
[1886]	Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys Gly Val Lys		475
[1887]		485	490
[1888]	His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe Lys Tyr Lys		495
[1889]		500	505
			510

[1890]	Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp Pro Arg Cys
[1891]	515 520 525
[1892]	Leu Thr Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg Asp Leu Ala
[1893]	530 535 540
[1894]	Ser Gly Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu Ser Val Asp
[1895]	545 550 555 560
[1896]	Gln Arg Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val Ile Leu Phe
[1897]	565 570 575
[1898]	Ser Val Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu Asn Ile Gln
[1899]	580 585 590
[1900]	Arg Phe Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp Pro Glu Phe
[1901]	595 600 605
[1902]	Gln Ala Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val Phe Asp Ser
[1903]	610 615 620
[1904]	Leu Gln Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp Tyr Ile Leu
[1905]	625 630 635 640
[1906]	Ser Ile Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe Ser Gly Tyr
[1907]	645 650 655
[1908]	Thr Phe Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr Leu Phe Pro
[1909]	660 665 670
[1910]	Phe Ser Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro Gly Leu Trp
[1911]	675 680 685
[1912]	Ile Leu Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly Met Thr Ala
[1913]	690 695 700
[1914]	Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp Tyr Tyr Glu
[1915]	705 710 715 720
[1916]	Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys Asn Asn Ala
[1917]	725 730 735
[1918]	Ile Glu Pro Arg Ser Phe Ser Gln Asn Ser Arg His Pro Ser Thr Arg
[1919]	740 745 750
[1920]	Gln Lys Gln Phe Asn Ala Thr Thr Ile Pro Glu Asn Asp Ile Glu Lys
[1921]	755 760 765
[1922]	Thr Asp Pro Trp Phe Ala His Arg Thr Pro Met Pro Lys Ile Gln Asn
[1923]	770 775 780
[1924]	Val Ser Ser Ser Asp Leu Leu Met Leu Leu Arg Gln Ser Pro Thr Pro
[1925]	785 790 795 800
[1926]	His Gly Leu Ser Leu Ser Asp Leu Gln Glu Ala Lys Tyr Glu Thr Phe
[1927]	805 810 815
[1928]	Ser Asp Asp Pro Ser Pro Gly Ala Ile Asp Ser Asn Asn Ser Leu Ser
[1929]	820 825 830
[1930]	Glu Met Thr His Phe Arg Pro Gln Leu His His Ser Gly Asp Met Val
[1931]	835 840 845

[1932]	Phe Thr Pro Glu Ser Gly Leu Gln Leu Arg Leu Asn Glu Lys Leu Gly
[1933]	850 855 860
[1934]	Thr Thr Ala Ala Thr Glu Leu Lys Lys Leu Asp Phe Lys Val Ser Ser
[1935]	865 870 875 880
[1936]	Thr Ser Asn Asn Leu Ile Ser Thr Ile Pro Ser Asp Asn Leu Ala Ala
[1937]	885 890 895
[1938]	Gly Thr Asp Asn Thr Ser Ser Leu Gly Pro Pro Ser Met Pro Val His
[1939]	900 905 910
[1940]	Tyr Asp Ser Gln Leu Asp Thr Thr Leu Phe Gly Lys Lys Ser Ser Pro
[1941]	915 920 925
[1942]	Leu Thr Glu Ser Gly Gly Pro Leu Ser Leu Ser Glu Glu Asn Asn Asp
[1943]	930 935 940
[1944]	Ser Lys Leu Leu Glu Ser Gly Leu Met Asn Ser Gln Glu Ser Ser Trp
[1945]	945 950 955 960
[1946]	Gly Lys Asn Val Ser Ser Thr Glu Ser Gly Arg Leu Phe Lys Gly Lys
[1947]	965 970 975
[1948]	Arg Ala His Gly Pro Ala Leu Leu Thr Lys Asp Asn Ala Leu Phe Lys
[1949]	980 985 990
[1950]	Val Ser Ile Ser Leu Leu Lys Thr Asn Lys Thr Ser Asn Asn Ser Ala
[1951]	995 1000 1005
[1952]	Thr Asn Arg Lys Thr His Ile Asp Gly Pro Ser Leu Leu Ile Glu
[1953]	1010 1015 1020
[1954]	Asn Ser Pro Ser Val Trp Gln Asn Ile Leu Glu Ser Asp Thr Glu
[1955]	1025 1030 1035
[1956]	Phe Lys Lys Val Thr Pro Leu Ile His Asp Arg Met Leu Met Asp
[1957]	1040 1045 1050
[1958]	Lys Asn Ala Thr Ala Leu Arg Leu Asn His Met Ser Asn Lys Thr
[1959]	1055 1060 1065
[1960]	Thr Ser Ser Lys Asn Met Glu Met Val Gln Gln Lys Lys Glu Gly
[1961]	1070 1075 1080
[1962]	Pro Ile Pro Pro Asp Ala Gln Asn Pro Asp Met Ser Phe Phe Lys
[1963]	1085 1090 1095
[1964]	Met Leu Phe Leu Pro Glu Ser Ala Arg Trp Ile Gln Arg Thr His
[1965]	1100 1105 1110
[1966]	Gly Lys Asn Ser Leu Asn Ser Gly Gln Gly Pro Ser Pro Lys Gln
[1967]	1115 1120 1125
[1968]	Leu Val Ser Leu Gly Pro Glu Lys Ser Val Glu Gly Gln Asn Phe
[1969]	1130 1135 1140
[1970]	Leu Ser Glu Lys Asn Lys Val Val Val Gly Lys Gly Glu Phe Thr
[1971]	1145 1150 1155
[1972]	Lys Asp Val Gly Leu Lys Glu Met Val Phe Pro Ser Ser Arg Asn
[1973]	1160 1165 1170

[1974]	Leu Phe Leu Thr Asn Leu Asp Asn Leu His Glu Asn Asn Thr His
[1975]	1175 1180 1185
[1976]	Asn Gln Glu Lys Lys Ile Gln Glu Glu Ile Glu Lys Lys Glu Thr
[1977]	1190 1195 1200
[1978]	Leu Ile Gln Glu Asn Val Val Leu Pro Gln Ile His Thr Val Thr
[1979]	1205 1210 1215
[1980]	Gly Thr Lys Asn Phe Met Lys Asn Leu Phe Leu Leu Ser Thr Arg
[1981]	1220 1225 1230
[1982]	Gln Asn Val Glu Gly Ser Tyr Asp Gly Ala Tyr Ala Pro Val Leu
[1983]	1235 1240 1245
[1984]	Gln Asp Phe Arg Ser Leu Asn Asp Ser Thr Asn Arg Thr Lys Lys
[1985]	1250 1255 1260
[1986]	His Thr Ala His Phe Ser Lys Lys Gly Glu Glu Glu Asn Leu Glu
[1987]	1265 1270 1275
[1988]	Gly Leu Gly Asn Gln Thr Lys Gln Ile Val Glu Lys Tyr Ala Cys
[1989]	1280 1285 1290
[1990]	Thr Thr Arg Ile Ser Pro Asn Thr Ser Gln Gln Asn Phe Val Thr
[1991]	1295 1300 1305
[1992]	Gln Arg Ser Lys Arg Ala Leu Lys Gln Phe Arg Leu Pro Leu Glu
[1993]	1310 1315 1320
[1994]	Glu Thr Glu Leu Glu Lys Arg Ile Ile Val Asp Asp Thr Ser Thr
[1995]	1325 1330 1335
[1996]	Gln Trp Ser Lys Asn Met Lys His Leu Thr Pro Ser Thr Leu Thr
[1997]	1340 1345 1350
[1998]	Gln Ile Asp Tyr Asn Glu Lys Glu Lys Gly Ala Ile Thr Gln Ser
[1999]	1355 1360 1365
[2000]	Pro Leu Ser Asp Cys Leu Thr Arg Ser His Ser Ile Pro Gln Ala
[2001]	1370 1375 1380
[2002]	Asn Arg Ser Pro Leu Pro Ile Ala Lys Val Ser Ser Phe Pro Ser
[2003]	1385 1390 1395
[2004]	Ile Arg Pro Ile Tyr Leu Thr Arg Val Leu Phe Gln Asp Asn Ser
[2005]	1400 1405 1410
[2006]	Ser His Leu Pro Ala Ala Ser Tyr Arg Lys Lys Asp Ser Gly Val
[2007]	1415 1420 1425
[2008]	Gln Glu Ser Ser His Phe Leu Gln Gly Ala Lys Lys Asn Asn Leu
[2009]	1430 1435 1440
[2010]	Ser Leu Ala Ile Leu Thr Leu Glu Met Thr Gly Asp Gln Arg Glu
[2011]	1445 1450 1455
[2012]	Val Gly Ser Leu Gly Thr Ser Ala Thr Asn Ser Val Thr Tyr Lys
[2013]	1460 1465 1470
[2014]	Lys Val Glu Asn Thr Val Leu Pro Lys Pro Asp Leu Pro Lys Thr
[2015]	1475 1480 1485

[2016]	Ser Gly Lys Val Glu Leu Leu Pro Lys Val His Ile Tyr Gln Lys		
[2017]	1490	1495	1500
[2018]	Asp Leu Phe Pro Thr Glu Thr Ser Asn Gly Ser Pro Gly His Leu		
[2019]	1505	1510	1515
[2020]	Asp Leu Val Glu Gly Ser Leu Leu Gln Gly Thr Glu Gly Ala Ile		
[2021]	1520	1525	1530
[2022]	Lys Trp Asn Glu Ala Asn Arg Pro Gly Lys Val Pro Phe Leu Arg		
[2023]	1535	1540	1545
[2024]	Val Ala Thr Glu Ser Ser Ala Lys Thr Pro Ser Lys Leu Leu Asp		
[2025]	1550	1555	1560
[2026]	Pro Leu Ala Trp Asp Asn His Tyr Gly Thr Gln Ile Pro Lys Glu		
[2027]	1565	1570	1575
[2028]	Glu Trp Lys Ser Gln Glu Lys Ser Pro Glu Lys Thr Ala Phe Lys		
[2029]	1580	1585	1590
[2030]	Lys Lys Asp Thr Ile Leu Ser Leu Asn Ala Cys Glu Ser Asn His		
[2031]	1595	1600	1605
[2032]	Ala Ile Ala Ala Ile Asn Glu Gly Gln Asn Lys Pro Glu Ile Glu		
[2033]	1610	1615	1620
[2034]	Val Thr Trp Ala Lys Gln Gly Arg Thr Glu Arg Leu Cys Ser Gln		
[2035]	1625	1630	1635
[2036]	Asn Pro Pro Val Leu Lys Arg His Gln Arg Glu Ile Thr Arg Thr		
[2037]	1640	1645	1650
[2038]	Thr Leu Gln Ser Asp Gln Glu Glu Ile Asp Tyr Asp Asp Thr Ile		
[2039]	1655	1660	1665
[2040]	Ser Val Glu Met Lys Lys Glu Asp Phe Asp Ile Tyr Asp Glu Asp		
[2041]	1670	1675	1680
[2042]	Glu Asn Gln Ser Pro Arg Ser Phe Gln Lys Lys Thr Arg His Tyr		
[2043]	1685	1690	1695
[2044]	Phe Ile Ala Ala Val Glu Arg Leu Trp Asp Tyr Gly Met Ser Ser		
[2045]	1700	1705	1710
[2046]	Ser Pro His Val Leu Arg Asn Arg Ala Gln Ser Gly Ser Val Pro		
[2047]	1715	1720	1725
[2048]	Gln Phe Lys Lys Val Val Phe Gln Glu Phe Thr Asp Gly Ser Phe		
[2049]	1730	1735	1740
[2050]	Thr Gln Pro Leu Tyr Arg Gly Glu Leu Asn Glu His Leu Gly Leu		
[2051]	1745	1750	1755
[2052]	Leu Gly Pro Tyr Ile Arg Ala Glu Val Glu Asp Asn Ile Met Val		
[2053]	1760	1765	1770
[2054]	Thr Phe Arg Asn Gln Ala Ser Arg Pro Tyr Ser Phe Tyr Ser Ser		
[2055]	1775	1780	1785
[2056]	Leu Ile Ser Tyr Glu Glu Asp Gln Arg Gln Gly Ala Glu Pro Arg		
[2057]	1790	1795	1800

[2058]	Lys Asn Phe Val Lys Pro Asn Glu Thr Lys Thr Tyr Phe Trp Lys
[2059]	1805 1810 1815
[2060]	Val Gln His His Met Ala Pro Thr Lys Asp Glu Phe Asp Cys Lys
[2061]	1820 1825 1830
[2062]	Ala Trp Ala Tyr Phe Ser Asp Val Asp Leu Glu Lys Asp Val His
[2063]	1835 1840 1845
[2064]	Ser Gly Leu Ile Gly Pro Leu Leu Val Cys His Thr Asn Thr Leu
[2065]	1850 1855 1860
[2066]	Asn Pro Ala His Gly Arg Gln Val Thr Val Gln Glu Phe Ala Leu
[2067]	1865 1870 1875
[2068]	Phe Phe Thr Ile Phe Asp Glu Thr Lys Ser Trp Tyr Phe Thr Glu
[2069]	1880 1885 1890
[2070]	Asn Met Glu Arg Asn Cys Arg Ala Pro Cys Asn Ile Gln Met Glu
[2071]	1895 1900 1905
[2072]	Asp Pro Thr Phe Lys Glu Asn Tyr Arg Phe His Ala Ile Asn Gly
[2073]	1910 1915 1920
[2074]	Tyr Ile Met Asp Thr Leu Pro Gly Leu Val Met Ala Gln Asp Gln
[2075]	1925 1930 1935
[2076]	Arg Ile Arg Trp Tyr Leu Leu Ser Met Gly Ser Asn Glu Asn Ile
[2077]	1940 1945 1950
[2078]	His Ser Ile His Phe Ser Gly His Val Phe Thr Val Arg Lys Lys
[2079]	1955 1960 1965
[2080]	Glu Glu Tyr Lys Met Ala Leu Tyr Asn Leu Tyr Pro Gly Val Phe
[2081]	1970 1975 1980
[2082]	Glu Thr Val Glu Met Leu Pro Ser Lys Ala Gly Ile Trp Arg Val
[2083]	1985 1990 1995
[2084]	Glu Cys Leu Ile Gly Glu His Leu His Ala Gly Met Ser Thr Leu
[2085]	2000 2005 2010
[2086]	Phe Leu Val Tyr Ser Asn Lys Cys Gln Thr Pro Leu Gly Met Ala
[2087]	2015 2020 2025
[2088]	Ser Gly His Ile Arg Asp Phe Gln Ile Thr Ala Ser Gly Gln Tyr
[2089]	2030 2035 2040
[2090]	Gly Gln Trp Ala Pro Lys Leu Ala Arg Leu His Tyr Ser Gly Ser
[2091]	2045 2050 2055
[2092]	Ile Asn Ala Trp Ser Thr Lys Glu Pro Phe Ser Trp Ile Lys Val
[2093]	2060 2065 2070
[2094]	Asp Leu Leu Ala Pro Met Ile Ile His Gly Ile Lys Thr Gln Gly
[2095]	2075 2080 2085
[2096]	Ala Arg Gln Lys Phe Ser Ser Leu Tyr Ile Ser Gln Phe Ile Ile
[2097]	2090 2095 2100
[2098]	Met Tyr Ser Leu Asp Gly Lys Lys Trp Gln Thr Tyr Arg Gly Asn
[2099]	2105 2110 2115

[2100]	Ser Thr Gly Thr Leu Met Val Phe Phe Gly Asn Val Asp Ser Ser
[2101]	2120 2125 2130
[2102]	Gly Ile Lys His Asn Ile Phe Asn Pro Pro Ile Ile Ala Arg Tyr
[2103]	2135 2140 2145
[2104]	Ile Arg Leu His Pro Thr His Tyr Ser Ile Arg Ser Thr Leu Arg
[2105]	2150 2155 2160
[2106]	Met Glu Leu Met Gly Cys Asp Leu Asn Ser Cys Ser Met Pro Leu
[2107]	2165 2170 2175
[2108]	Gly Met Glu Ser Lys Ala Ile Ser Asp Ala Gln Ile Thr Ala Ser
[2109]	2180 2185 2190
[2110]	Ser Tyr Phe Thr Asn Met Phe Ala Thr Trp Ser Pro Ser Lys Ala
[2111]	2195 2200 2205
[2112]	Arg Leu His Leu Gln Gly Arg Ser Asn Ala Trp Arg Pro Gln Val
[2113]	2210 2215 2220
[2114]	Asn Asn Pro Lys Glu Trp Leu Gln Val Asp Phe Gln Lys Thr Met
[2115]	2225 2230 2235
[2116]	Lys Val Thr Gly Val Thr Thr Gln Gly Val Lys Ser Leu Leu Thr
[2117]	2240 2245 2250
[2118]	Ser Met Tyr Val Lys Glu Phe Leu Ile Ser Ser Ser Gln Asp Gly
[2119]	2255 2260 2265
[2120]	His Gln Trp Thr Leu Phe Phe Gln Asn Gly Lys Val Lys Val Phe
[2121]	2270 2275 2280
[2122]	Gln Gly Asn Gln Asp Ser Phe Thr Pro Val Val Asn Ser Leu Asp
[2123]	2285 2290 2295
[2124]	Pro Pro Leu Leu Thr Arg Tyr Leu Arg Ile His Pro Gln Ser Trp
[2125]	2300 2305 2310
[2126]	Val His Gln Ile Ala Leu Arg Met Glu Val Leu Gly Cys Glu Ala
[2127]	2315 2320 2325
[2128]	Gln Asp Leu Tyr
[2129]	2330
[2130]	<210> 45
[2131]	<211> 4
[2132]	<212> PRT
[2133]	<213> 人工序列
[2134]	<220>
[2135]	<223> 人IgG1的氨基酸233-236
[2136]	<400> 45
[2137]	Glu Leu Leu Gly
[2138]	1
[2139]	<210> 46
[2140]	<211> 42
[2141]	<212> PRT

[2142]	<213> 人工序列
[2143]	<220>
[2144]	<223> XTEN AE42-4,蛋白质序列
[2145]	<400> 46
[2146]	Gly Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly
[2147]	1 5 10 15
[2148]	Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala
[2149]	20 25 30
[2150]	Thr Ser Gly Ser Glu Thr Pro Ala Ser Ser
[2151]	35 40
[2152]	<210> 47
[2153]	<211> 126
[2154]	<212> DNA
[2155]	<213> 人工序列
[2156]	<220>
[2157]	<223> XTEN AE42-4,DNA序列
[2158]	<400> 47
[2159]	ggcgcgccag gttctcctgc tggctcccc acctcaacag aagaggggac aagcgaaagc 60
[2160]	gctacgcctg agagtggccc tggctctgag ccagccacct ccggctctga aaccctgcc 120
[2161]	tcgagc 126
[2162]	<210> 48
[2163]	<211> 144
[2164]	<212> PRT
[2165]	<213> 人工序列
[2166]	<220>
[2167]	<223> XTEN AE144-2A,蛋白质序列
[2168]	<400> 48
[2169]	Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly
[2170]	1 5 10 15
[2171]	Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly
[2172]	20 25 30
[2173]	Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
[2174]	35 40 45
[2175]	Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr Glu
[2176]	50 55 60
[2177]	Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
[2178]	65 70 75 80
[2179]	Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly
[2180]	85 90 95
[2181]	Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu
[2182]	100 105 110
[2183]	Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu

[2184]	115	120	125
[2185]	Ser Gly Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly		
[2186]	130	135	140
[2187]	<210> 49		
[2188]	<211> 450		
[2189]	<212> DNA		
[2190]	<213> 人工序列		
[2191]	<220>		
[2192]	<223> XTEN AE144-2A,DNA序列		
[2193]	<400> 49		
[2194]	ggcgcgccaa ccagtacgga gccgtccgag gggagcgcac caggaagccc ggctggggagc	60	
[2195]	ccgacttcta ccgaagaggg tacatctacc gaaccaagtg aaggttcagc accaggcacc	120	
[2196]	tcaacagaac cctctgaggg ctcggcgcct ggtacaagtg agtccgccac cccagaatcc	180	
[2197]	gggcctggga caagcacaga accttcgaa gggagtgtccc ctggaacatc cgaatcggca	240	
[2198]	accccagaat cagggccagg atctgagccc gcgacttcgg gctccgagac gcctgggaca	300	
[2199]	tccaccgagc cctccgaagg atcagcccca ggcaccagca cggagccctc tgagggaagc	360	
[2200]	gcacctggta ccagcgaaag cgcaactccc gaatcaggtc ccggtacgag cgagtcggcg	420	
[2201]	accccggaga gcgggccagg tgcctcgagc	450	
[2202]	<210> 50		
[2203]	<211> 144		
[2204]	<212> PRT		
[2205]	<213> 人工序列		
[2206]	<220>		
[2207]	<223> XTEN AE144-3B,蛋白质序列		
[2208]	<400> 50		
[2209]	Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Glu Ser		
[2210]	1 5 10 15		
[2211]	Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser		
[2212]	20 25 30		
[2213]	Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly		
[2214]	35 40 45		
[2215]	Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu		
[2216]	50 55 60		
[2217]	Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly		
[2218]	65 70 75 80		
[2219]	Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly		
[2220]	85 90 95		
[2221]	Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu		
[2222]	100 105 110		
[2223]	Pro Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser		
[2224]	115 120 125		
[2225]	Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly		

[2226]	130	135	140
[2227]	<210> 51		
[2228]	<211> 450		
[2229]	<212> DNA		
[2230]	<213> 人工序列		
[2231]	<220>		
[2232]	<223> XTEN AE144-3B,DNA序列		
[2233]	<400> 51		
[2234]	ggcgcgccaa gtcccgtctgg aagcccaact agcaccgaag aggggacctc agagtccgcc 60		
[2235]	acccccgagt ccggccctgg ctctgagcct gccactagcg gctccgagac tcctggcaca 120		
[2236]	tccgaaagcg ctacaccga gaggtagccc ggcacctcta ccgagcccag tgagggtcc 180		
[2237]	gcccttgaa caagaccga gccagcgaa ggcagcgccc caggacctc cacagagccc 240		
[2238]	agtgaaggca gtgctcctgg caccagcacc gaaccaagcg agggctctgc acccgggacc 300		
[2239]	tccaccgagc caagcgaagg ctctgccctt ggcacttcca ccgagcccag cgaaggcagc 360		
[2240]	gcccttgga gcccgtctgg ctctccacc agcactgagg agggcacatc taccgaacca 420		
[2241]	agtgaaggct ctgcaccagg tgcctcgagc 450		
[2242]	<210> 52		
[2243]	<211> 144		
[2244]	<212> PRT		
[2245]	<213> 人工序列		
[2246]	<220>		
[2247]	<223> XTEN AE144-4A,蛋白质序列		
[2248]	<400> 52		
[2249]	Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala		
[2250]	1 5 10 15		
[2251]	Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu		
[2252]	20 25 30		
[2253]	Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly		
[2254]	35 40 45		
[2255]	Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr Glu		
[2256]	50 55 60		
[2257]	Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu		
[2258]	65 70 75 80		
[2259]	Ser Gly Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly		
[2260]	85 90 95		
[2261]	Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Ser Pro Ala Gly		
[2262]	100 105 110		
[2263]	Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Glu Ser Ala Thr Pro Glu		
[2264]	115 120 125		
[2265]	Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly		
[2266]	130 135 140		
[2267]	<210> 53		

[2268]	<211> 450
[2269]	<212> DNA
[2270]	<213> 人工序列
[2271]	<220>
[2272]	<223> XTEN AE144-4A,DNA序列
[2273]	<400> 53
[2274]	ggcgcgccaa cgtccgaaag tgctaccctt gagtcaggcc ctggtagtga gcctgccaca 60
[2275]	agcgggaagcg aaactccggg gacctcagag tctgccactc ccgaatcggg gccaggctct 120
[2276]	gaaccggcca cttcaggagg cgaaacacca ggaacatcgg agagcgctac cccggagagc 180
[2277]	gggccaggaa ctagtactga gcctagcgag ggaagtgcac ctggtacaag cgagtccgcc 240
[2278]	acacccgagt ctggccctgg ctctccagcg ggctcaccca cgagcactga agagggtctt 300
[2279]	cccgtggca gcccaacgtc gacagaagaa ggatcaccag caggctcccc cacatcaaca 360
[2280]	gaggagggtg catcagaatc tgctactccc gagagtggac ccggtacctc cactgagccc 420
[2281]	agcgagggga gtgcaccagg tgcctcgagc 450
[2282]	<210> 54
[2283]	<211> 144
[2284]	<212> PRT
[2285]	<213> 人工序列
[2286]	<220>
[2287]	<223> XTEN AE144-5A,蛋白质序列
[2288]	<400> 54
[2289]	Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala
[2290]	1 5 10 15
[2291]	Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
[2292]	20 25 30
[2293]	Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly
[2294]	35 40 45
[2295]	Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr Glu
[2296]	50 55 60
[2297]	Pro Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser
[2298]	65 70 75 80
[2299]	Thr Glu Glu Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly
[2300]	85 90 95
[2301]	Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
[2302]	100 105 110
[2303]	Ala Thr Pro Glu Ser Gly Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser
[2304]	115 120 125
[2305]	Thr Glu Glu Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly
[2306]	130 135 140
[2307]	<210> 55
[2308]	<211> 450
[2309]	<212> DNA

- [2310] <213> 人工序列
- [2311] <220>
- [2312] <223> XTEN AE144-5A,DNA序列
- [2313] <400> 55
- [2314] ggcgcgccaa catcagagag cgccaccctt gaaagtggc cgggagcga gccagccaca 60
- [2315] tctgggtcgg aaacgccagg cacaagttag tctgcaactc ccgagtcagg acctggctcc 120
- [2316] gagcctgcca ctagcggtc cgagactccg ggaacttccg agagcgtac accagaaagc 180
- [2317] ggaccgggaa ccagtaccga acctagcgag ggctctgctc cgggcagccc agccggctct 240
- [2318] cctacatcca cggaggaggg cacttccgaa tccgccacc cggagtcagg gccaggatct 300
- [2319] gaaccgcta cctcaggcag tgagacgcca ggaacgagcg agtccgctac accggagagt 360
- [2320] gggccaggga gccctgctgg atctctacg tccactgagg aagggtcacc agcgggctcg 420
- [2321] cccaccagca ctgaagaagg tgctctgagc 450
- [2322] <210> 56
- [2323] <211> 144
- [2324] <212> PRT
- [2325] <213> 人工序列
- [2326] <220>
- [2327] <223> XTEN AE144-6B,蛋白质序列
- [2328] <400> 56
- [2329] Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser
- [2330] 1 5 10 15
- [2331] Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
- [2332] 20 25 30
- [2333] Ser Gly Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly
- [2334] 35 40 45
- [2335] Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Ser Glu Pro Ala
- [2336] 50 55 60
- [2337] Thr Ser Gly Ser Glu Thr Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser
- [2338] 65 70 75 80
- [2339] Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
- [2340] 85 90 95
- [2341] Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Ser Glu Pro Ala
- [2342] 100 105 110
- [2343] Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
- [2344] 115 120 125
- [2345] Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
- [2346] 130 135 140
- [2347] <210> 57
- [2348] <211> 450
- [2349] <212> DNA
- [2350] <213> 人工序列
- [2351] <220>

- [2352] <223> XTEN AE144-6B,DNA序列
- [2353] <400> 57
- [2354] ggcgcgccaa catctaccga gccttccgaa ggctctgccc ctgggacctc agaatctgca 60
- [2355] acccctgaaa gcggccctgg aacctccgaa agtgccactc ccgagagcgg cccagggaca 120
- [2356] agcgagtcag caaccctga gtctggaccc ggcagcgagc ctgcaacctc tggctcagag 180
- [2357] actcccggtc cagaaccgc tacctcaggc tccgagacac ccggctctcc tgctgggagt 240
- [2358] cccacttcca ccgaggaagg aacatccact gaggctagtg agggctctgc ccctggaacc 300
- [2359] agcacagagc caagtgaggg cagtgcacca ggatccgagc cagcaaccag cgggtccgag 360
- [2360] actcccgga cctctgagtc tgccaccca gagagcggac ccggcacttc aaccgagccc 420
- [2361] tccgaaggat cagcaccagg tgcctcgagc 450
- [2362] <210> 58
- [2363] <211> 144
- [2364] <212> PRT
- [2365] <213> 人工序列
- [2366] <220>
- [2367] <223> XTEN AG144-1,蛋白质序列
- [2368] <400> 58
- [2369] Pro Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser
- [2370] 1 5 10 15
- [2371] Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Thr Pro Gly Ser Gly Thr
- [2372] 20 25 30
- [2373] Ala Ser Ser Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser
- [2374] 35 40 45
- [2375] Pro Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Ala Ser
- [2376] 50 55 60
- [2377] Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr
- [2378] 65 70 75 80
- [2379] Ala Ser Ser Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser
- [2380] 85 90 95
- [2381] Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ala Ser
- [2382] 100 105 110
- [2383] Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser
- [2384] 115 120 125
- [2385] Ser Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser
- [2386] 130 135 140
- [2387] <210> 59
- [2388] <211> 450
- [2389] <212> DNA
- [2390] <213> 人工序列
- [2391] <220>
- [2392] <223> XTEN AG144-1,DNA序列
- [2393] <400> 59

[2394]	ggcgcgccac ccgggtcgtc cccgtcggcg tccaccggaa cagggccagg gtcatccccg	60
[2395]	tcagcgtcga ctgggacggg acccgggaca cccggttcgg ggactgcac ctcctcgcct	120
[2396]	ggttcgtcca ccccgtcagg agccacgggt tcgccgggaa gcagcccaag cgcateccact	180
[2397]	ggtacagggc ctggggcttc accgggtact tcatccacgg ggtcaccggg aacgcccgga	240
[2398]	tcggggacgg cttcctcatc accaggatcg tcaacaccct cgggcgcaac gggcagcccc	300
[2399]	ggaacccctg gttcgggtac ggcgtcgtcg agccccggcg cgagccccgg aacaagctcg	360
[2400]	acaggatcgc ctggggcgtc acccggcacg tcgagcacag gcagccccgg aacccttgga	420
[2401]	tcgggaaccg cgtcgtcaag cgcctcgagc	450
[2402]	<210>	60
[2403]	<211>	144
[2404]	<212>	PRT
[2405]	<213>	人工序列
[2406]	<220>	
[2407]	<223>	XTEN AG144-A,蛋白质序列
[2408]	<400>	60
[2409]	Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ser Ser Pro	
[2410]	1 5 10 15	
[2411]	Ser Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser Pro Ser Ala Ser Thr	
[2412]	20 25 30	
[2413]	Gly Thr Gly Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro	
[2414]	35 40 45	
[2415]	Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Pro	
[2416]	50 55 60	
[2417]	Ser Ala Ser Thr Gly Thr Gly Pro Gly Ala Ser Pro Gly Thr Ser Ser	
[2418]	65 70 75 80	
[2419]	Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro	
[2420]	85 90 95	
[2421]	Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Thr Pro Gly	
[2422]	100 105 110	
[2423]	Ser Gly Thr Ala Ser Ser Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser	
[2424]	115 120 125	
[2425]	Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro	
[2426]	130 135 140	
[2427]	<210>	61
[2428]	<211>	450
[2429]	<212>	DNA
[2430]	<213>	人工序列
[2431]	<220>	
[2432]	<223>	XTEN AG144-A,DNA序列
[2433]	<400>	61
[2434]	ggcgcgccag gtgcctcgcc gggaacatca tcaactggtt caccggggtc atccccctcg	60
[2435]	gcctcaaccg ggacgggtcc cggtcatcc ccagcgcca gcactggaac aggtctctggc	120

[2436] actcctggtt ccggtacggc atcgatcatcc ccgggaagct caacaccgtc cggagcgcaca 180
 [2437] ggatcacctg gctcgtcacc ttggcgctca actggaacgg ggccaggggc ctcaccgcga 240
 [2438] acgtcctega ctgggtcgcc tggtagccg ggatcaggaa cggcctcatc ctcgcttggg 300
 [2439] tcctcaacgc cctcgggtgc gactggttcg ccgggaactc ctggctcggg gacggcctcg 360
 [2440] tcgtcgcttg gggcatcacc ggggacgagc tccacggggt cccctggagc gtcaccgggg 420
 [2441] acctcctega caggtagccc ggcctcgagc 450
 [2442] <210> 62
 [2443] <211> 144
 [2444] <212> PRT
 [2445] <213> 人工序列
 [2446] <220>
 [2447] <223> XTEN AG144-B,蛋白质序列
 [2448] <400> 62
 [2449] Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ser Ser Thr
 [2450] 1 5 10 15
 [2451] Pro Ser Gly Ala Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
 [2452] 20 25 30
 [2453] Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro
 [2454] 35 40 45
 [2455] Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Pro
 [2456] 50 55 60
 [2457] Ser Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser Pro Ser Ala Ser Thr
 [2458] 65 70 75 80
 [2459] Gly Thr Gly Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro
 [2460] 85 90 95
 [2461] Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ala Ser Pro
 [2462] 100 105 110
 [2463] Gly Thr Ser Ser Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
 [2464] 115 120 125
 [2465] Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
 [2466] 130 135 140
 [2467] <210> 63
 [2468] <211> 450
 [2469] <212> DNA
 [2470] <213> 人工序列
 [2471] <220>
 [2472] <223> XTEN AG144-B,DNA序列
 [2473] <400> 63
 [2474] ggcgccag gtacaccggg cagcggcacg gcttcgtcgt caccggctc gtccacaccg 60
 [2475] tcgggagcta cggaagccc aggagcgtca ccgggaacgt cgtcaacggg gtcaccgggt 120
 [2476] acgccaggta gcggcacggc cagcagctcg ccaggttcat cgaccccgtc gggagcgact 180
 [2477] gggtcgccc gatcaagccc gtcagcttcc actggaacag gaccgggtc gtcgcgtca 240

[2478] gcctcaacgg ggacaggacc tggttcatcg acgccgtcag gggcgacagg ctgccccgga 300
 [2479] tcgtcaacac cctcgggggc aacgggggagc cctggtgcgt cgcctggaac ctcatccacc 360
 [2480] ggaagcccg gggcctcgcc gggtacgagc tccacgggat cgcccgagc gtcccccgga 420
 [2481] acttcaagca caggagccc tgcctcgagc 450
 [2482] <210> 64
 [2483] <211> 144
 [2484] <212> PRT
 [2485] <213> 人工序列
 [2486] <220>
 [2487] <223> XTEN AG144-C,蛋白质序列
 [2488] <400> 64
 [2489] Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ala Ser Pro
 [2490] 1 5 10 15
 [2491] Gly Thr Ser Ser Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
 [2492] 20 25 30
 [2493] Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
 [2494] 35 40 45
 [2495] Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Thr Pro Gly
 [2496] 50 55 60
 [2497] Ser Gly Thr Ala Ser Ser Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
 [2498] 65 70 75 80
 [2499] Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
 [2500] 85 90 95
 [2501] Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ser Ser Thr
 [2502] 100 105 110
 [2503] Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala
 [2504] 115 120 125
 [2505] Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
 [2506] 130 135 140
 [2507] <210> 65
 [2508] <211> 450
 [2509] <212> DNA
 [2510] <213> 人工序列
 [2511] <220>
 [2512] <223> XTEN AG144-C,DNA序列
 [2513] <400> 65
 [2514] ggcgcgccag gtacaccgg atcgggtaca gcgtcatcga gccccgtgc gtcacctggt 60
 [2515] acgtcgagca cggggtcgcc aggggcgtcc cctgggacgt cctcaacagg ctgccccggt 120
 [2516] gcgtcaccgg gcacgtcgtc cacgggttca cctggtagct ccccttcgc gtccactggc 180
 [2517] accgggcctg gaactccggg gagcggcaca gcgagctcgt cgccgggagc atgcctggg 240
 [2518] acatcgagca ccgggtcgcc aggagcatcg cccggaacat ccagcacagg aagccccggc 300
 [2519] gcgtcgccgg ggacatcaag cacaggttcc ccgggatcga gcacgccgtc cggagccact 360

[2520]	ggatcaccag ggagctcgac accttcggc gcaacgggat cgcccggagc cagcccgggt 420
[2521]	acgtcaagca ctggctcccc tgcctcgagc 450
[2522]	<210> 66
[2523]	<211> 144
[2524]	<212> PRT
[2525]	<213> 人工序列
[2526]	<220>
[2527]	<223> XTEN AG144-F,蛋白质序列
[2528]	<400> 66
[2529]	Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser Pro
[2530]	1 5 10 15
[2531]	Ser Ala Ser Thr Gly Thr Gly Pro Gly Ala Ser Pro Gly Thr Ser Ser
[2532]	20 25 30
[2533]	Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
[2534]	35 40 45
[2535]	Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Pro
[2536]	50 55 60
[2537]	Ser Ala Ser Thr Gly Thr Gly Pro Gly Ala Ser Pro Gly Thr Ser Ser
[2538]	65 70 75 80
[2539]	Thr Gly Ser Pro Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro
[2540]	85 90 95
[2541]	Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ser Ser Thr
[2542]	100 105 110
[2543]	Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala
[2544]	115 120 125
[2545]	Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
[2546]	130 135 140
[2547]	<210> 67
[2548]	<211> 450
[2549]	<212> DNA
[2550]	<213> 人工序列
[2551]	<220>
[2552]	<223> XTEN AG144-F,DNA序列
[2553]	<400> 67
[2554]	ggcgccag gctccagccc ctccgcgagc acgggaaccg gaccaggttc gtcaccctca 60
[2555]	gcatcaacgg ggacgggacc gggggcgctca ccaggaacgt cctccaccgg ctgcgccgggt 120
[2556]	gcatcaccgg gaacgtcatc gaccggatcg ccaggagct cgacgccatc aggcgcaaca 180
[2557]	ggatcacctg gctcaagccc tagcgctca accggcacgg gtccgggtgc ctcccctggc 240
[2558]	acgtccagca ccggtacacc cggatcgagc ccatccgcct caaccggaac cggaccgggt 300
[2559]	acaccagggt cggaacagc ctctcgtca ccaggctcct caacccctc gggagccacg 360
[2560]	ggttcgcccg gttcgtcaac gccttcggga gcaactggta gcccggagc atcgccagga 420
[2561]	acttcgagca cggggtcgcc cgctcgagc 450

[2562]	<210> 68
[2563]	<211> 4374
[2564]	<212> DNA
[2565]	<213> 人工序列
[2566]	<220>
[2567]	<223> coFVIII-1 DNA序列
[2568]	<400> 68
[2569]	atgcagattg agctgtctac ttgctttttc ctgtgcctgc tgaggttttg cttttccgct 60
[2570]	acacgaaggt attatctggg ggctgtggaa ctgtcttggg attacatgca gaggacctg 120
[2571]	ggagagctgc cagtgacgc aagggtttcc ctagagtcc ctaagtcatt ccccttcaac 180
[2572]	actagcgtgg tctacaagaa aacctgttc gtggagtta ctgacacac gttcaacatc 240
[2573]	gcaaagccta ggccaccctg gatgggactg ctggggccaa caatccagc cgagggtgtac 300
[2574]	gacaccgtgg tcattacact taagaacatg gcctcacacc ccgtgagcct gcatgctgtg 360
[2575]	ggcgtcagct actggaagc ttccgaagga gcagagtatg acgatcagac ttcccagaga 420
[2576]	gaaaaagagg acgataaggt gtttctctgc ggatctcata cctacgtgtg gcaggctcctg 480
[2577]	aaagagaatg gccctatggc ctccgaccct ctgtgcctga cctactctta tctgagtcac 540
[2578]	gtggacctgg tcaaggatct gaacagcggc ctgacggag ccctgctggt gtgcagggaa 600
[2579]	ggaagcctgg ctaaggagaa aaccagaca ctgcataagt tcattctgct gttcgccgtg 660
[2580]	tttgacgaag ggaaatcatg gcacagcag acaaagaata gtctgatgca ggacagggat 720
[2581]	gccgttcag ccagagcttg gcccataatg cactgtga acggctacgt caatcgtca 780
[2582]	ctgcctgggc tgatcggtg ccaccgaaag agcgtgtatt ggcatgtcat cgggatgggc 840
[2583]	accacacctg aagtgcactc cttttcctg gagggacata ctttctggt ccgcaaccac 900
[2584]	cgacaggctt ccctggagat ctctccaatt accttctga cagcacagac tctgtgatg 960
[2585]	gacctgggac agttctgtg gttttgccac atcagctccc accagcatga tggcatggag 1020
[2586]	gcttacgtga aagtggactc ttgtcccgag gaacctcagc tgcggatgaa gaacaatgag 1080
[2587]	gaagcagaag actatgacga tgacctgacc gactccgaga tggatgtggt ccgattcgat 1140
[2588]	gacgataaca gcccctcctt tatccagatt agatctgtgg ccaagaaaca ccctaagaca 1200
[2589]	tgggtccatt acatcgagc cgaggaagag gactgggatt atgcaccact ggtgctggca 1260
[2590]	ccagacgac gctcctacaa atctcagtat ctgaacaatg ggccacagag gattggcaga 1320
[2591]	aagtacaaga aagtgcggtt catggcatat accgatgaga ctttcaagac tcgcgaagcc 1380
[2592]	atccagcacg agagcggcat cctgggacca ctgctgtacg gagaagtggg agacaccctg 1440
[2593]	ctgatcattt tcaagaacca ggccagccg cttacaata tctatccaca tgggattaca 1500
[2594]	gatgtgcgcc ctctgtacag caggagactg ccaaaggcgc tcaaacacct gaaggacttc 1560
[2595]	ccaatcctgc ccggagaaat cttcaagtac aagtggactg tcaccgtcga ggatggcccc 1620
[2596]	actaagagcg accctcgtg cctgaccgc tactattcta gtttcgtgaa tatggaaaga 1680
[2597]	gatctggcaa gcggactgat cggaccactg ctgattgtt acaaagagag cgtggatcag 1740
[2598]	agaggcaacc agatcatgtc cgacaagcgg aatgtgattc tgttcagtgt ctttgacgaa 1800
[2599]	aacaggtcat ggtacctgac cgagaacatc cagagattcc tgcctaatcc agctggggtg 1860
[2600]	cagctggaag atcctgagtt tcaggcatct aacatcatgc atagtattaa tggctacgtg 1920
[2601]	ttcgacagtt tgcagctgag cgtgtgcctg cagaggtcg cttactggta taccctgagc 1980
[2602]	attggggcac agacagattt cctgagcgtg ttcttttccg gctacattt taagcataaa 2040
[2603]	atggtctatg aggacacact gactctgttc ccttcagcg gcgaaaccgt gtttatgagc 2100

[2604]	atggagaatc ccggactgtg gattctgggg tgccacaaca gcgatttcag aaatcgcgga	2160
[2605]	atgactgccc tgctgaaagt gtcaagctgt gacaagaaca ccggggacta ctatgaagat	2220
[2606]	tcatacgagg acatcagcgc atatctgctg tccaaaaaca atgccattga accccggtct	2280
[2607]	tttagtcaga atcctccagt gctgaagagg caccagaggg agatcacccg cactaccctg	2340
[2608]	cagagtgatc aggaagagat cgactacgac gatacaattt ctgtggaaat gaagaaagag	2400
[2609]	gacttcgata tctatgacga agatgagaac cagagtccct gatcattcca gaagaaaacc	2460
[2610]	aggcattact ttattgccgc agtggagcgg ctgtgggatt atggcatgtc ctctagtcct	2520
[2611]	cacgtgctgc gaaatagggc ccagtcagga agcgtccac agttcaagaa agtgggtctc	2580
[2612]	caggagttaa cagacgggtc ctttactcag ccactgtaca ggggcgaact gaacgagcac	2640
[2613]	ctgggactgc tggggcccta tatcagagca gaagtggagg ataacattat ggtcaccttc	2700
[2614]	agaaatcagg cctctcgcc ttacagtttt tattcaagcc tgatctctta cgaagaggac	2760
[2615]	cagcgacagg gagctgaacc acgaaaaaac ttctgtaagc ctaatgagac caaacatac	2820
[2616]	ttttggaagg tgcagacca tatggcccca acaaaagacg agttcgattg caaggcatgg	2880
[2617]	gcctatTTTT ctgacgtgga tctggagaag gacgtgcaca gtggcctgat tggccactg	2940
[2618]	ctggtgtgcc atactaacac cctgaatcca gccacggcc ggcaggtcac tgtccaggag	3000
[2619]	ttcgtctgt tctttaccat ctttgatgag acaagagct ggtacttcac cgaaaacatg	3060
[2620]	gagcgaaatt gcagggtcc atgtaacatt cagatggaag accccacatt caaggagaac	3120
[2621]	taccgtttc atgctatcaa tggatacacc atggatactc tgcccgggct ggtcatggca	3180
[2622]	caggaccaga gaatccggtg gtatctgctg agcatgggca gcaacgagaa tatccactca	3240
[2623]	attcatttca gcgggcacgt gtttactgtc aggaagaaag aagagtacaa gatggccctg	3300
[2624]	tacaacctgt atcccggcgt gttcgaaacc gtcgagatgc tgcctagcaa ggccggaatc	3360
[2625]	tggagagtgg aatgcctgat tggagagcac ctgcatgctg ggatgtctac cctgtttctg	3420
[2626]	gtgtacagta ataagtgtca gacacccctg ggaatggcat ccgggcatac cagggatttc	3480
[2627]	cagattaccg catctggaca gtacggacag tgggcaccta agctggctag actgcactat	3540
[2628]	tccggatcta tcaacgcttg gtccacaaaa gagectttct cttggattaa ggtggacctg	3600
[2629]	ctggcccaa tgatcattca tggcatcaaa actcaggag ctcggcagaa gttctctct	3660
[2630]	ctgtacatct cacagtttat catcatgtac agcctggatg ggaagaaatg gcagacatac	3720
[2631]	cgcggcaata gcacaggaac tctgatggtg ttctttggca acgtggacag cagcggaatc	3780
[2632]	aagcacaaca ttttcaatcc ccctatcatt gctagataca tccggtgca cccaacccat	3840
[2633]	tattctattc gaagtacact gaggatggaa ctgatgggat gcgatctgaa cagttgttca	3900
[2634]	atgcccttgg ggatggagtc caaggcaatc tctgacgcc agattaccgc cagctcctac	3960
[2635]	ttcactaata tgtttgctac ctggagccct tccaaagcaa gactgcacct gcaaggccgc	4020
[2636]	agcaacgcat ggcgaccaca ggtgaacaat cccaaggagt ggttgacagt cgattttcag	4080
[2637]	aaaactatga aggtgaccgg ggtcacaact caggcgctga aaagtctgct gacctcaatg	4140
[2638]	tacgtcaagg agttcctgat ctctagtcca caggacggac atcagtggac actgtttctt	4200
[2639]	cagaacggga aggtgaaagt cttccagggc aatcaggatt cttttacacc tgtgtgtcaac	4260
[2640]	agtctagacc ctccactgct gaccagatac ctgagaatcc accctcagtc ctgggtgcac	4320
[2641]	cagattgccc tgagaatgga agtgcctggga tgcgaggccc aggatctgta ctga	4374
[2642]	<210> 69	
[2643]	<211> 577	
[2644]	<212> DNA	
[2645]	<213> 人工序列	

[2646]	<220>	
[2647]	<223>	ET启动子,DNA序列
[2648]	<400>	69
[2649]	ctcgagggtca attcacgcga gttaataatt accagcgcgg gccaaataaa taatccgcga	60
[2650]	ggggcagggtg acgtttgccc agcgcgcgct ggtaattatt aacctcgcga atattgattc	120
[2651]	gaggccgcga ttgccgcaat cgcgaggggc aggtgacctt tgcccagcgc gcgttcgccc	180
[2652]	cgccccggac ggtatcgata agcttaggag cttgggctgc aggtcgaggg cactgggagg	240
[2653]	atgttgagta agatggaaaa ctactgatga cccttgacaga gacagagtat taggacatgt	300
[2654]	ttgaacaggg gccgggcgat cagcaggtag ctctagagga tccccgtctg tctgcacatt	360
[2655]	tcgtagagcg agtgttccga tactctaatac tccctaggca aggttcatat ttgtgtaggt	420
[2656]	tacttattct cttttgttg actaagtcaa taatcagaat cagcaggttt ggagtcagct	480
[2657]	tggcagggat cagcagcctg ggttggaaag aggggggtata aaagccccct caccaggaga	540
[2658]	agccgtcaca cagatccaca agctcctgcc accatgg	577
[2659]	<210>	70
[2660]	<211>	4374
[2661]	<212>	DNA
[2662]	<213>	人工序列
[2663]	<220>	
[2664]	<223>	coFVIII-5
[2665]	<400>	70
[2666]	atgcaaatcg aactgagcac ctgtttcttc ctctgcctgc tgagattctg tttctccgcg	60
[2667]	acccgccgat actacctggg agcagtgagg ctctcctggg attacatgca gagcgacctt	120
[2668]	ggggagctgc ccgtggatgc caggttcctt ccccggtgc caaagtcgtt tccgttcaac	180
[2669]	acctccgtgg tgtacaagaa aactctgttc gtggagttca ccgaccacct gttcaatata	240
[2670]	gccaaagcca gacctccctg gatggggctg ttgggacctt ccatccaagc ggaggtgtac	300
[2671]	gacactgtgg tcatcactct gaagaacatg gcctcgcac ccgtgtccct gcacgccgtg	360
[2672]	ggagtgtctt actggaaagc gtccgagggg gccgaatacg acgaccagac ctgcagaga	420
[2673]	gaaaaggaag atgacaaggt gttcccagga ggatcgacac cctacgtgtg gcaagtgttg	480
[2674]	aaggagaacg gcccaatggc ctccgacctg ctgtgcctga cctactcgta cctgtccac	540
[2675]	gtggacctcg tgaaggacct caactcgga ctgattggag ccctgctggt ctgcaggaa	600
[2676]	ggctcactgg cgaaagaaaa gactcagacc ttgcacaagt tcattctgct gttcgtctgt	660
[2677]	ttcgacgagg ggaagtcgtg gcacagcgag actaagaact ccctgatgca agatagagat	720
[2678]	gccgcctccg cccgggcctg gcctaagatg cacaccgtga acggttacgt gaaccgctcc	780
[2679]	ctccctggcc tgattggatg ccaccggaag tccgtgtact ggcacgtgat cgggatggg	840
[2680]	accacccccg aggtgcacag catcttcctg gaaggtcaca catttctcgt gcgcaaccac	900
[2681]	cggcaggcct ccctggaaat cagccccatt accttctca ctgcccagac tctgtgatg	960
[2682]	gacctgggac agttctgtct gttctgccat atctcctccc accaacaatga cggaatggag	1020
[2683]	gcatacgtga aggtcgattc ctgcctgag gaacccagc tccgcatgaa gaacaatgag	1080
[2684]	gaagccgagg actacgacga cgacctgac gatagcgaga tggatgtggt ccggttcgat	1140
[2685]	gacgataaca gcccttcctt catccaaatt cgctcgttg caaagaagca cccaagacc	1200
[2686]	tgggtgcatt acattgcggc ggaagaagag gactgggatt atgccccgt tgcctcgt	1260
[2687]	cctgacgacc ggagctacaa gagccagtac ctgaacaacg gtccacagag gatcggtaga	1320

[2688]	aagtacaaga aggtccgctt catggcctat accgacgaaa ccttcaaaac tagagaggcc	1380
[2689]	atccaacacg aatccggcat cctgggcccg ctcttgtacg gagaagtcgg cgacaccctt	1440
[2690]	ctcattatct tcaagaacca ggcttcccgg cegtacaaca tctatccgca tgggatcact	1500
[2691]	gacgtgcgcc cactgtactc gcggcgccgtg cccaagggtg tcaaacacct gaaggatttt	1560
[2692]	ccgatccttc cgggagaaat cttcaagtac aagtggaccg tgaccgtgga agatggccca	1620
[2693]	actaagtctg accctagatg cctcacccgc tactactcat ccttcgtcaa catggagcgc	1680
[2694]	gacctggcca gcggactgat cggcccgtg ctgatttgc tacaaggaatc agtggacca	1740
[2695]	cggggaaacc agatcatgtc ggataagagg aacgtcatcc tcttctccgt gtttgacgaa	1800
[2696]	aaccggtcgt ggtacctgac tgaaaacatc cagcggttcc tcccaaccc cgcgggcgtg	1860
[2697]	cagctggaag atcctgagtt tcaggcatca aacatcatgc actccattaa cggctacgtg	1920
[2698]	ttcgattcgc tgcagctgag cgtgtgtctg cacgaagtgg cctactggta catcctgtcc	1980
[2699]	attggtgccc agactgactt cctgtccgtg tttttctccg gctacacgtt caagcacaag	2040
[2700]	atggtgtacg aggacaccct gaccctcttc ctttttccg gcgaaactgt gtttatgagc	2100
[2701]	atggagaatc ccggcctgtg gatcttgggc tgccacaaca gcgacttccg taacagagga	2160
[2702]	atgactgcgc tgctcaaggt gtccagctgc gacaagaaca ccggagacta ttatgaggac	2220
[2703]	tcatacgagg acatctccgc ctacctctg tccaagaata acgccattga acctcggagc	2280
[2704]	ttcagccaga acccaccgt gcttaagaga catcaacggg agatcactag gaccaccctg	2340
[2705]	cagtcagacc aggaggaaat cgactacgat gacaccatct cggtcgagat gaagaaggag	2400
[2706]	gactttgaca tctacgacga agatgaaaac cagagcccga ggctgttcca aaagaaaacc	2460
[2707]	cgccactact ttattgctgc tgcgagcgg ctgtgggact acggaatgtc gtccctcgccg	2520
[2708]	cacgtgtcc gcaaccgagc ccagagcggc tcgggtgccg aattcaagaa ggctgtgttc	2580
[2709]	caggagttca ctgacgggag cttcactcag cttttgtacc ggggagaact caatgaacat	2640
[2710]	ctcgccctcc tcggacctta catcagagca gaagtggaag ataacatcat ggtcactttc	2700
[2711]	cgtaaccaag ccagccgccc gtactcgttc tactcctccc tcatttctta cgaagaggac	2760
[2712]	cagcggcagg gcgcagaacc gcgcaagaac ttcgtgaagc ccaacgaaac caagacctac	2820
[2713]	ttctggaaag tgcagcatca tatggccccg actaaggacg agtttgactg caaagcctgg	2880
[2714]	gcctacttct ccgatgtgga cttggagaag gacgtccact ccggcctcat cggctcccctg	2940
[2715]	ctcgtgtgcc ataccaatac cctgaacccc gcacacggtc gccaggtcac cgtgcaggag	3000
[2716]	ttcgtctgt tcttactat cttcgacgaa actaagtcct ggtacttcac cgagaacatg	3060
[2717]	gagaggaact gcagagcccc ctgtaacatc cagatggagg acccgacgtt caagggaaaac	3120
[2718]	taccggttcc acgccattaa cggatacatc atggatacgc tgccgggtct tgtgatggcc	3180
[2719]	caggatcaac ggatcagatg gtacttattg tcgatgggca gcaacgagaa catccactct	3240
[2720]	attcacttct ccggtcatgt gttcactgtg cggaagaagg aagagtacaa gatggccctg	3300
[2721]	tacaaccttt atcccggagt gttcgaaact gtggaaatgc tgccgtcgaa ggccggcatt	3360
[2722]	tggcgcgtgg agtgtttgat tggagaacat ctccatgcgg ggatgtcaac cctgttcctg	3420
[2723]	gtgtatagca acaagtgcc aactccgctt gggatggcgt caggacacat tagggatttc	3480
[2724]	cagatcactg cgtccggcca gtacggccaa tgggccccta agctggcccc cctgcattac	3540
[2725]	tccggatcca ttaacgctg gtcaaccaag gagccattct cctggatcaa ggtggacctt	3600
[2726]	ctggccccca tgattatcca cggaattaag acccaggggg cccggcagaa gttctctca	3660
[2727]	ctgtacatca gccagttcat aatcatgtac tccctggacg gaaagaagt gcaaacctac	3720
[2728]	agggggaaca gcaccggcac actgatggtc tttttcgaa atgtggactc ctccgggatt	3780
[2729]	aagcataaca tcttcaaccc tccgattatc gctcgttaca ttagacttca ccctaccac	3840

[2730]	tacagcattc gctccaccct gcggatggaa ctgatgggct gcgatctgaa ctctgtgcagc	3900
[2731]	atgccgttgg gaatggagtc caaagcaatt tccgacgcgc agatcaccgc ctctgtcctac	3960
[2732]	tttaccaaca tggtcgccac gtggtcaccg tccaaggccc ggctgcacct ccagggaaga	4020
[2733]	tccaacgcat ggcggccaca ggtcaacaac cctaaggagt ggctccaggt ggacttccag	4080
[2734]	aaaaccatga aggtcaccgg agtcacaacc cagggagtga agtcgctgct gacttctatg	4140
[2735]	tacgtcaagg agttcctgat ctccagcagc caggacgggc accagtggac cctgttcttc	4200
[2736]	caaaatggaa aggtcaagggt gtttcagggc aatcaggatt cattcacccc ggtgggtgaac	4260
[2737]	tcccttgatc caccctcct gacccgctac ctctgcatcc acccacagtc ctgggtgcac	4320
[2738]	cagatcgcgc tgaggatgga ggtcctggga tgcgaagccc aggacctgta ctga	4374
[2739]	<210> 71	
[2740]	<211> 4374	
[2741]	<212> DNA	
[2742]	<213> 人工序列	
[2743]	<220>	
[2744]	<223> coFVIII-6	
[2745]	<400> 71	
[2746]	atgcagattg agctgtccac ttgtttcttc ctgtgcctcc tgcgttctg tttctccgcc	60
[2747]	actgcceggt actaccttgg agccgtggag ctttcatggg actacatgca gacgcacctg	120
[2748]	ggcgaactcc ccgtggatgc cagattcccc ccccgctgc caaagtcctt cccctttaac	180
[2749]	acctccgtgg tgtacaagaa aacctcttt gtcgagtcca ctgaccacct gttcaacatc	240
[2750]	gccaaagcgc gccaccttg gatgggcctc ctgggaccga ccattcaagc tgaagtgtac	300
[2751]	gacaccgtgg tgatcacct gaagaacatg gcgtcccacc ccgtgtccct gcatgcggtc	360
[2752]	ggagtgtcct actggaaggc ctccgaagga gctgagtacg acgaccagac tagccagcgg	420
[2753]	gaaaaggagg acgataaagt gttcccgggc ggctcgcata cttacgtgtg gcaagtctg	480
[2754]	aaggaaaacg gacctatggc atccgatcct ctgtgcctga cttactccta ctttcccat	540
[2755]	gtggacctcg tgaaggacct gaacagcggg ctgattggtg cacttctcgt gtgccgcgaa	600
[2756]	ggttcgctcg ctaaggaaaa gaccagacc ctccataagt tcattctttt gttcgctgtg	660
[2757]	ttcgatgaag gaaagtcatg gcattccgaa actaagaact cgctgatgca ggaccgggat	720
[2758]	gccgcctcag cccgcgctg gcctaaaatg catacagtca acggatacgt gaatcggtca	780
[2759]	ctgcccgggc tcacggttg tcacagaaag tccgtgtact ggcacgtcat cggcatgggc	840
[2760]	actacgcctg aagtgcactc catcttctg gaagggcaca ctttctcgt gcgcaaccac	900
[2761]	cgccaggcct ctctggaaat ctccccgatt acctttctga ccgccagac tctgtcatg	960
[2762]	gacctggggc agttccttct cttctgccac atctccagcc atcagcacga cggaatggag	1020
[2763]	gcctacgtga aggtggactc atgcccggaa gaacctcagt tgcggatgaa gaacaacgag	1080
[2764]	gaggccgagg actatgacga cgatttgact gactccgaga tggacgtcgt gcggttcgat	1140
[2765]	gacgacaaca gcccagctt catccagatt cgcagcgtgg ccaagaagca cccaaaacc	1200
[2766]	tggtgtcact acatcgcggc cgaggaagaa gattgggact acgccccgtt ggtgttgga	1260
[2767]	ccgatgacc ggtcgtacaa gtcccagtat ctgaacaatg gtccgcagcg gattggcaga	1320
[2768]	aagtacaaga aagtgcggtt catggcgtag actgacgaaa cgtttaagac ccgggaggcc	1380
[2769]	attcaacatg agagcgcat tctgggacca ctgctgtacg gagaggtcgg cgataccctg	1440
[2770]	ctcatcatct tcaaaaacca ggcctcccg cttacaaca tctaccctca cggaatcacc	1500
[2771]	gacgtgcggc cactctactc gcggcgctg ccgaaggcgc tcaagcacct gaaagacttc	1560

[2772]	cctatcctgc cgggcgaaat cttcaagtat aagtggaccg tcaccgtgga ggacgggccc	1620
[2773]	accaagagcg atcctaggtg tctgactcgg tactactcca gcttcgtgaa catggaacgg	1680
[2774]	gacctggcat cgggactcat tggaccgtg ctgatctgct acaaagagtc ggtggatcaa	1740
[2775]	cgcggaacc agatcatgtc cgacaagcgc aacgtgatcc tgttctccgt gtttgatgaa	1800
[2776]	aacagatcct ggtacctcac tgaaaacatc cagaggttcc tcccaaacc cgcaggagt	1860
[2777]	caactggagg accctgagtt tcaggcctcg aatatcatgc actcgattaa cggttacgtg	1920
[2778]	ttcgactcgc tgcagctgag cgtgtgcctc catgaagtcg cttactggta cattctgtcc	1980
[2779]	atcggcgcgc agactgactt cctgagcgtg ttcttttccg gttacacctt taagcacaag	2040
[2780]	atggtgtacg aagataccct gaccctgttc cttttctccg gcgaaacggg gttcatgtcg	2100
[2781]	atggagaacc cgggtctgtg gattctggga tgccacaaca gcgactttcg gaaccgcgga	2160
[2782]	atgactgcc tgcgaaggt gtccctcatgc gacaagaaca ccggagacta ctacaggac	2220
[2783]	tcctacgagg atatctcagc ctacctcctg tccaagaaca acgcgatcga gccgcgcagc	2280
[2784]	ttcagccaga accgcctgt gctgaagagg caccagcgag aaattaccg gaccaccctc	2340
[2785]	caatcggatc aggaggaaat cgactacgac gacaccatct cgggtgaaat gaagaaggaa	2400
[2786]	gatttcgata tctacgacga ggacgaaaat cagtccccctc gctcattcca aaagaaaact	2460
[2787]	agacactact ttatcgccgc ggtggaaaga ctgtgggact atggaatgtc atccagccct	2520
[2788]	cacgtccttc ggaaccgggc ccagagcgga tcggtgcctc agttcaagaa agtgggtgttc	2580
[2789]	caggagttca ccgacggcag cttcaccag ccgctgtacc ggggagaact gaacgaacac	2640
[2790]	ctgggcctgc tcggtcccta catccgcgcg gaagtggagg ataacatcat ggtgaccttc	2700
[2791]	cgtaaccaag catccagacc ttactccttc tattctctcc tgatctcata cgaggaggac	2760
[2792]	cagcgccaag gcgccgagcc ccgcaagaac ttctgcaagc ccaacgagac taagacctac	2820
[2793]	ttctggaagg tccaacacca tatggccccg accaaggatg agtttgactg caaggcctgg	2880
[2794]	gcctacttct ccgactgga ccttgagaag gatgtccatt ccggcctgat cgggccgctg	2940
[2795]	ctcgtgtgtc acaccaacac cctgaaccca gcgcatggac gccaggtcac cgtccaggag	3000
[2796]	tttgcctgt tcttcacat ttttgacgaa actaagtcct ggtacttcac cgagaatatg	3060
[2797]	gagcgaact gtagagcgc ctgcaatc cagatggaag atccgacttt caaggagaac	3120
[2798]	tatagattcc acgccatcaa cgggtacatc atggatactc tgccggggct ggtcatggcc	3180
[2799]	caggatcaga ggattcgtg gtacttctg tcaatgggat cgaacgaaaa cattactcc	3240
[2800]	attcacttct ccggtcacgt gttcactgtg cgcaagaagg aggagtacaa gatggcgtg	3300
[2801]	tacaatctgt accccggggt gttcgaaact gtggagatgc tgccgtccaa ggccggcatc	3360
[2802]	tggagagtgg agtgctgat cggagagcac ctccacgcgg ggatgtccac cctcttctg	3420
[2803]	gtgtactcga ataagtcca gaccccgctg ggcattggcct cgggccacat cagagacttc	3480
[2804]	cagatcacag caagcgga atacggccaa tgggcgccga agctggcccc cttgcactac	3540
[2805]	tccggtcga tcaacgatg gtccaccaag gaaccgttct cgtggattaa ggtggacctc	3600
[2806]	ctggcccccta tgattatcca cggaattaag acccaggcg ccaggcagaa gttctctcc	3660
[2807]	ctgtacatct cgcaattcat catcatgtac agcctggacg ggaagaagtg gcagacttac	3720
[2808]	aggggaaact ccaccggcac cctgatggtc tttttcgca acgtggattc ctccggcatt	3780
[2809]	aagcacaaca tcttcaaccc accgatcata gccagatata ttaggtcca cccactcac	3840
[2810]	tactcaatcc gctcaactct tcggatggaa ctcatggggg gcgacctgaa ctctgtctcc	3900
[2811]	atgccgttgg ggatggaatc aaaggctatt agcgacgcc agatcaccgc gagctcttac	3960
[2812]	ttcactaaca tgttgccac ctggagcccc tccaaggcca ggctgcactt gcagggacgg	4020
[2813]	tcaaatgcct ggccggccga agtgaacaat ccgaaggaaat ggcttcaagt ggatttccaa	4080

[2814]	aagaccatga aagtgaccgg agtcaccacc caggagtgga agtcccttct gacctcgatg	4140
[2815]	tatgtgaagg agttcctgat tagcagcagc caggacgggc accagtggac cctgttcttc	4200
[2816]	caaaacggaa aggtcaagggt gttccagggg aaccaggact cgttcacacc cgtggtgaac	4260
[2817]	tccctggacc cccactgct gacgcggtac ttgaggattc atcctcagtc ctgggtccat	4320
[2818]	cagattgcat tgcgaatgga agtcctgggc tgcgaggccc aggacctgta ctga	4374
[2819]	<210> 72	
[2820]	<211> 4824	
[2821]	<212> DNA	
[2822]	<213> 人工序列	
[2823]	<220>	
[2824]	<223> coFVIII-6-XTEN	
[2825]	<400> 72	
[2826]	atgcagattg agctgtccac ttgtttcttc ctgtgcctcc tgcgcttctg tttctccgcc	60
[2827]	actcgccggt actaccttg agccgtggag ctttcatggg actacatgca gagcgacctg	120
[2828]	ggcgaactcc ccgtggatgc cagattcccc cccgcgtgc caaagtcctt cccctttaac	180
[2829]	acctccgtgg tgtacaagaa aacctcttt gtgcagttca ctgaccacct gttcaacatc	240
[2830]	gccaaagccgc gccaccttg gatgggcctc ctgggaccga ccattcaagc tgaagtgtac	300
[2831]	gacaccgtgg tgatcacct gaagaacatg gcgtcccacc ccgtgtccct gcatgcggtc	360
[2832]	ggagtgtcct actggaaggc ctccgaagga gctgagtacg acgaccagac tagccagcgg	420
[2833]	gaaaaggagg acgataaagt gttcccgggc ggctcgata cttacgttg gcaagtctg	480
[2834]	aaggaaaacg gacctatggc atccgatcct ctgtgcctga cttactccta cttttcccat	540
[2835]	gtggacctcg tgaaggacct gaacagcggg ctgattggtg cacttctcgt gtgccgcgaa	600
[2836]	ggttcgctcg ctaaggaaaa gaccagacc ctccataagt tcatcctttt gttcgctgtg	660
[2837]	ttcgatgaag gaaagtcatg gcattccgaa actaagaact cgctgatgca ggaccgggat	720
[2838]	gccgcctcag cccgcgctg gcctaaaatg catacagtca acgatacgt gaatcggatc	780
[2839]	ctgcccgggc tcatcggttg tcacagaaag tccgtgtact ggcacgtcat cggcatgggc	840
[2840]	actacgcctg aagtgcactc catcttctg gaagggcaca ctttctcgt gcgcaaccac	900
[2841]	cgccaggcct ctctggaaat ctccccgatt acctttctga ccgccagac tctgtcatg	960
[2842]	gacctggggc agttcttct cttctgccac atctccagcc atcagcacga cggaatggag	1020
[2843]	gcctacgtga aggtggactc atgcccggaa gaacctcagt tgcggatgaa gaacaacgag	1080
[2844]	gaggccgagg actatgacga cgatttgact gactccgaga tggacgtcgt gcggttcgat	1140
[2845]	gacgacaaca gcccagctt catccagatt cgcagcgtgg ccaagaagca cccaaaacc	1200
[2846]	tgggtgcact acatcgcggc cgaggaagaa gattgggact acgccccgtt ggtgctggca	1260
[2847]	cccgatgacc ggtcgtacaa gtcccagtat ctgaacaatg gtccgcagcg gattggcaga	1320
[2848]	aagtacaaga aagtgcggtt catggcgtag actgacgaaa cgtttaagac ccgggaggcc	1380
[2849]	attcaacatg agagcggcat tctgggacca ctgctgtacg gagaggtcgg cgataccctg	1440
[2850]	ctcatcatct taaaaacca ggctcccg cttacaaca tctacctca cggaatcacc	1500
[2851]	gacgtcggc cactctactc gcggcgctg ccgaaggcg tcaagcacct gaaagacttc	1560
[2852]	cctatcctgc cgggcgaaat cttcaagtat aagtggaccg tcaccgtgga ggacgggccc	1620
[2853]	accaagagcg atcctaggtg tctgactcgg tactactcca gttcgtgaa catggaacgg	1680
[2854]	gacctggcat cgggactcat tggaccgtg ctgatctgct acaaagagtc ggtggatcaa	1740
[2855]	cgcggcaacc agatcatgtc cgacaagcgc aacgtgatcc tgttctccgt gtttgatgaa	1800

[2856]	aacagatcct ggtacctcac tgaaaacatc cagaggttcc tcccaaacc cgcaggagtg	1860
[2857]	caactggagg accctgagtt tcaggcctcg aatatcatgc actcgattaa cggttacgtg	1920
[2858]	ttcgactgcg tgcagctgag cgtgtgcctc catgaagtcg cttactggta cattctgtcc	1980
[2859]	atcggcgccc agactgactt cctgagcgtg ttcttttccg gttacacctt taagcacaag	2040
[2860]	atggtgtacg aagataccct gaccctgttc cttttctccg gcgaaacggt gttcatgtcg	2100
[2861]	atggagaacc cgggtctgtg gattctggga tgccacaaca gcgactttcg gaaccgcgga	2160
[2862]	atgactgccc tgctgaaggt gtctcatgc gacaagaaca cgggagacta ctacgaggac	2220
[2863]	tcctacgagg atatctcagc ctacctctg tccaagaaca acgcgatcga gccgcgcagc	2280
[2864]	ttcagccaga acggcgcgcc aacatcagag agcgccaccc ctgaaagtgg tcccgggagc	2340
[2865]	gagccagcca catctgggtc ggaaacgcca ggcacaagtg agtctgcaac tcccaggtcc	2400
[2866]	ggacctggct ccgagcctgc cactagcggc tccgagactc cgggaacttc cgagagcgct	2460
[2867]	acaccagaaa gcggaccgga aaccagtacc gaacctagcg agggctctgc tccgggcagc	2520
[2868]	ccagccggct ctctacatc cacggaggag ggcacttccg aatccgccac cccgagtcga	2580
[2869]	gggccaggat ctgaaccgcg tacttcaggc agtgagacgc caggaacgag cgagtccgct	2640
[2870]	acaccggaga gtgggcccagg gagccctgct ggatctccta cgtccactga ggaagggtca	2700
[2871]	ccagcgggct cgcaccag cactgaagaa ggtgcctcga gccgcctgt gctgaagagg	2760
[2872]	caccagcgag aaattaccg gaccacctc caatcggatc aggaggaaat cgactacgac	2820
[2873]	gacaccatct cgggtgaaat gaagaaggaa gatttcgata tctacgacga ggacgaaaat	2880
[2874]	cagtccctc gctcattcca aaagaaaact agacactact ttatcgccgc ggtggaaaga	2940
[2875]	ctgtgggact atggaatgc atccagccct cacgtccttc ggaaccgggc ccagagcgga	3000
[2876]	tcggtgcctc agttcaagaa agtgggtgtc caggagtcca ccgacggcag cttcaccag	3060
[2877]	ccgctgtacc ggggagaact gaacgaacac ctgggcctgc tcggtcccta catccgcgcg	3120
[2878]	gaagtggagg ataacatcat ggtgaccttc cgtaaccaag catccagacc ttactccttc	3180
[2879]	tattctctcc tgatctcata cgaggaggac cagcgccaag gcgccgagcc ccgcaagaac	3240
[2880]	ttcgtcaagc ccaacgagac taagacctac ttctggaagg tccaacacca tatggccccg	3300
[2881]	accaaggatg agtttgactg caaggcctgg gcctacttct ccgacgtgga ccttgagaag	3360
[2882]	gatgtccatt ccggcctgat cgggccgctg ctctgtgtgc acaccaacac cctgaacca	3420
[2883]	gcgcatggac gccaggtcac cgtccaggag tttgctctgt tcttcacct ttttgacgaa	3480
[2884]	actaagtcct ggtacttcac cgagaatatg gagcgaaact gtagagcgcc ctgcaatatc	3540
[2885]	cagatggaag atccgacttt caaggagaac tatagattcc acgccatcaa cgggtacatc	3600
[2886]	atggatactc tgccggggct ggtcatggcc caggatcaga ggattcgggt gtacttgctg	3660
[2887]	tcaatgggat cgaacgaaaa cattcactcc attcacttct ccggtcacgt gttcactgtg	3720
[2888]	cgcaagaagg aggagtacaa gatggcgctg tacaatctgt accccggggt gttcgaaact	3780
[2889]	gtggagatgc tgccgtccaa ggccggcatc tggagagtgg agtgcctgat cggagagcac	3840
[2890]	ctccacgcgg ggatgtccac cctcttctg gtgtactcga ataagtcca gaccccgctg	3900
[2891]	ggcatggcct cgggccacat cagagacttc cagatcacag caagcggaca atacggcaa	3960
[2892]	tgggcgccga agctggcccg cttgcactac tccggtcga tcaacgatg gtccaccaag	4020
[2893]	gaaccgttct cgtggattaa ggtggacctc ctggccccta tgattatcca cggaattaag	4080
[2894]	accaggcg ccaggcagaa gttctctctc ctgtacatct cgcaattcat catcatgtac	4140
[2895]	agcctggacg ggaagaagtg gcagacttac aggggaaact ccaccggcac cctgatggtc	4200
[2896]	tttttcggca acgtggattc ctccggcatt aagcacaaca tcttcaacc accgatcata	4260
[2897]	gccagatata ttaggtcca cccactcac tactcaatcc gctcaactct tcggatggaa	4320

[2898]	ctcatggggt gcgacctgaa ctcctgctcc atgccgttgg ggatggaatc aaaggctatt 4380
[2899]	agcgacgccc agatcacccg gagctcctac ttcactaaca tgttcgccac ctggagcccc 4440
[2900]	tccaaggcca ggctgcactt gcagggacgg tcaaatgcct ggcgggccgca agtgaacaat 4500
[2901]	ccgaaggaat ggcttcaagt ggatttccaa aagaccatga aagtgaccgg agtcaccacc 4560
[2902]	cagggagtga agtcccttct gacctcgatg tatgtgaagg agttcctgat tagcagcagc 4620
[2903]	caggacgggc accagtggac cctgttcttc caaaacggaa aggtcaaggt gttccagggg 4680
[2904]	aaccaggact cgttcacacc cgtggtgaac tccctggacc cccactgct gacgcggtac 4740
[2905]	ttgaggattc atcctcagtc ctgggtccat cagattgcat tgcgaatgga agtcctgggc 4800
[2906]	tgcgaggccc aggacctgta ctga 4824
[2907]	<210> 73
[2908]	<211> 12
[2909]	<212> PRT
[2910]	<213> 人工序列
[2911]	<220>
[2912]	<223> AD基序
[2913]	<400> 73
[2914]	Gly Glu Ser Pro Gly Gly Ser Ser Gly Ser Glu Ser
[2915]	1 5 10
[2916]	<210> 74
[2917]	<211> 12
[2918]	<212> PRT
[2919]	<213> 人工序列
[2920]	<220>
[2921]	<223> AD基序
[2922]	<400> 74
[2923]	Gly Ser Glu Gly Ser Ser Gly Pro Gly Glu Ser Ser
[2924]	1 5 10
[2925]	<210> 75
[2926]	<211> 12
[2927]	<212> PRT
[2928]	<213> 人工序列
[2929]	<220>
[2930]	<223> AD基序
[2931]	<400> 75
[2932]	Gly Ser Ser Glu Ser Gly Ser Ser Glu Gly Gly Pro
[2933]	1 5 10
[2934]	<210> 76
[2935]	<211> 12
[2936]	<212> PRT
[2937]	<213> 人工序列
[2938]	<220>
[2939]	<223> AD基序

[2940]	<400> 76
[2941]	Gly Ser Gly Gly Glu Pro Ser Glu Ser Gly Ser Ser
[2942]	1 5 10
[2943]	<210> 77
[2944]	<211> 12
[2945]	<212> PRT
[2946]	<213> 人工序列
[2947]	<220>
[2948]	<223> AE、AM基序
[2949]	<400> 77
[2950]	Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu
[2951]	1 5 10
[2952]	<210> 78
[2953]	<211> 12
[2954]	<212> PRT
[2955]	<213> 人工序列
[2956]	<220>
[2957]	<223> AE、AM、AQ基序
[2958]	<400> 78
[2959]	Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro
[2960]	1 5 10
[2961]	<210> 79
[2962]	<211> 12
[2963]	<212> PRT
[2964]	<213> 人工序列
[2965]	<220>
[2966]	<223> AE、AM、AQ基序
[2967]	<400> 79
[2968]	Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro
[2969]	1 5 10
[2970]	<210> 80
[2971]	<211> 12
[2972]	<212> PRT
[2973]	<213> 人工序列
[2974]	<220>
[2975]	<223> AE、AM、AQ基序
[2976]	<400> 80
[2977]	Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro
[2978]	1 5 10
[2979]	<210> 81
[2980]	<211> 12
[2981]	<212> PRT

[2982]	<213>	人工序列
[2983]	<220>	
[2984]	<223>	AF、AM基序
[2985]	<400>	81
[2986]	Gly Ser Thr Ser Glu Ser Pro Ser Gly Thr Ala Pro	
[2987]	1	5 10
[2988]	<210>	82
[2989]	<211>	12
[2990]	<212>	PRT
[2991]	<213>	人工序列
[2992]	<220>	
[2993]	<223>	AF、AM基序
[2994]	<400>	82
[2995]	Gly Thr Ser Thr Pro Glu Ser Gly Ser Ala Ser Pro	
[2996]	1	5 10
[2997]	<210>	83
[2998]	<211>	12
[2999]	<212>	PRT
[3000]	<213>	人工序列
[3001]	<220>	
[3002]	<223>	AF、AM基序
[3003]	<400>	83
[3004]	Gly Thr Ser Pro Ser Gly Glu Ser Ser Thr Ala Pro	
[3005]	1	5 10
[3006]	<210>	84
[3007]	<211>	12
[3008]	<212>	PRT
[3009]	<213>	人工序列
[3010]	<220>	
[3011]	<223>	AF、AM基序
[3012]	<400>	84
[3013]	Gly Ser Thr Ser Ser Thr Ala Glu Ser Pro Gly Pro	
[3014]	1	5 10
[3015]	<210>	85
[3016]	<211>	12
[3017]	<212>	PRT
[3018]	<213>	人工序列
[3019]	<220>	
[3020]	<223>	AG、AM基序
[3021]	<400>	85
[3022]	Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro	
[3023]	1	5 10

[3024]	<210> 86
[3025]	<211> 12
[3026]	<212> PRT
[3027]	<213> 人工序列
[3028]	<220>
[3029]	<223> AG、AM基序
[3030]	<400> 86
[3031]	Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro
[3032]	1 5 10
[3033]	<210> 87
[3034]	<211> 12
[3035]	<212> PRT
[3036]	<213> 人工序列
[3037]	<220>
[3038]	<223> AG、AM基序
[3039]	<400> 87
[3040]	Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro
[3041]	1 5 10
[3042]	<210> 88
[3043]	<211> 12
[3044]	<212> PRT
[3045]	<213> 人工序列
[3046]	<220>
[3047]	<223> AG、AM基序
[3048]	<400> 88
[3049]	Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
[3050]	1 5 10
[3051]	<210> 89
[3052]	<211> 12
[3053]	<212> PRT
[3054]	<213> 人工序列
[3055]	<220>
[3056]	<223> AQ基序
[3057]	<400> 89
[3058]	Gly Glu Pro Ala Gly Ser Pro Thr Ser Thr Ser Glu
[3059]	1 5 10
[3060]	<210> 90
[3061]	<211> 12
[3062]	<212> PRT
[3063]	<213> 人工序列
[3064]	<220>
[3065]	<223> AQ基序

[3066]	<400> 90
[3067]	Gly Thr Gly Glu Pro Ser Ser Thr Pro Ala Ser Glu
[3068]	1 5 10
[3069]	<210> 91
[3070]	<211> 12
[3071]	<212> PRT
[3072]	<213> 人工序列
[3073]	<220>
[3074]	<223> AQ基序
[3075]	<400> 91
[3076]	Gly Ser Gly Pro Ser Thr Glu Ser Ala Pro Thr Glu
[3077]	1 5 10
[3078]	<210> 92
[3079]	<211> 12
[3080]	<212> PRT
[3081]	<213> 人工序列
[3082]	<220>
[3083]	<223> AQ基序
[3084]	<400> 92
[3085]	Gly Ser Glu Thr Pro Ser Gly Pro Ser Glu Thr Ala
[3086]	1 5 10
[3087]	<210> 93
[3088]	<211> 12
[3089]	<212> PRT
[3090]	<213> 人工序列
[3091]	<220>
[3092]	<223> AQ基序
[3093]	<400> 93
[3094]	Gly Pro Ser Glu Thr Ser Thr Ser Glu Pro Gly Ala
[3095]	1 5 10
[3096]	<210> 94
[3097]	<211> 12
[3098]	<212> PRT
[3099]	<213> 人工序列
[3100]	<220>
[3101]	<223> AQ基序
[3102]	<400> 94
[3103]	Gly Ser Pro Ser Glu Pro Thr Glu Gly Thr Ser Ala
[3104]	1 5 10
[3105]	<210> 95
[3106]	<211> 12
[3107]	<212> PRT

[3108]	<213>	人工序列
[3109]	<220>	
[3110]	<223>	BC基序
[3111]	<400>	95
[3112]	Gly Ser Gly Ala Ser Glu Pro Thr Ser Thr Glu Pro	
[3113]	1	5 10
[3114]	<210>	96
[3115]	<211>	12
[3116]	<212>	PRT
[3117]	<213>	人工序列
[3118]	<220>	
[3119]	<223>	BC基序
[3120]	<400>	96
[3121]	Gly Ser Glu Pro Ala Thr Ser Gly Thr Glu Pro Ser	
[3122]	1	5 10
[3123]	<210>	97
[3124]	<211>	12
[3125]	<212>	PRT
[3126]	<213>	人工序列
[3127]	<220>	
[3128]	<223>	BC基序
[3129]	<400>	97
[3130]	Gly Thr Ser Glu Pro Ser Thr Ser Glu Pro Gly Ala	
[3131]	1	5 10
[3132]	<210>	98
[3133]	<211>	12
[3134]	<212>	PRT
[3135]	<213>	人工序列
[3136]	<220>	
[3137]	<223>	BC基序
[3138]	<400>	98
[3139]	Gly Thr Ser Thr Glu Pro Ser Glu Pro Gly Ser Ala	
[3140]	1	5 10
[3141]	<210>	99
[3142]	<211>	12
[3143]	<212>	PRT
[3144]	<213>	人工序列
[3145]	<220>	
[3146]	<223>	BD基序
[3147]	<400>	99
[3148]	Gly Ser Thr Ala Gly Ser Glu Thr Ser Thr Glu Ala	
[3149]	1	5 10

[3150]	<210>	100
[3151]	<211>	12
[3152]	<212>	PRT
[3153]	<213>	人工序列
[3154]	<220>	
[3155]	<223>	BD基序
[3156]	<400>	100
[3157]	Gly Ser Glu Thr Ala Thr Ser Gly Ser Glu Thr Ala	
[3158]	1	5 10
[3159]	<210>	101
[3160]	<211>	12
[3161]	<212>	PRT
[3162]	<213>	人工序列
[3163]	<220>	
[3164]	<223>	BD基序
[3165]	<400>	101
[3166]	Gly Thr Ser Glu Ser Ala Thr Ser Glu Ser Gly Ala	
[3167]	1	5 10
[3168]	<210>	102
[3169]	<211>	12
[3170]	<212>	PRT
[3171]	<213>	人工序列
[3172]	<220>	
[3173]	<223>	BD基序
[3174]	<400>	102
[3175]	Gly Thr Ser Thr Glu Ala Ser Glu Gly Ser Ala Ser	
[3176]	1	5 10
[3177]	<210>	103
[3178]	<211>	1607
[3179]	<212>	PRT
[3180]	<213>	人工序列
[3181]	<220>	
[3182]	<223>	coFVIII-6-XTEN蛋白质序列
[3183]	<400>	103
[3184]	Met Gln Ile Glu Leu Ser Thr Cys Phe Phe Leu Cys Leu Leu Arg Phe	
[3185]	1	5 10 15
[3186]	Cys Phe Ser Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser	
[3187]		20 25 30
[3188]	Trp Asp Tyr Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg	
[3189]		35 40 45
[3190]	PhePro Pro Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val	
[3191]		50 55 60

[3192]	Tyr	Lys	Lys	Thr	Leu	Phe	Val	Glu	Phe	Thr	Asp	His	Leu	Phe	Asn	Ile
[3193]	65					70					75					80
[3194]	Ala	Lys	Pro	Arg	Pro	Pro	Trp	Met	Gly	Leu	Leu	Gly	Pro	Thr	Ile	Gln
[3195]					85					90						95
[3196]	Ala	Glu	Val	Tyr	Asp	Thr	Val	Val	Ile	Thr	Leu	Lys	Asn	Met	Ala	Ser
[3197]				100					105					110		
[3198]	His	Pro	Val	Ser	Leu	His	Ala	Val	Gly	Val	Ser	Tyr	Trp	Lys	Ala	Ser
[3199]				115					120					125		
[3200]	Glu	Gly	Ala	Glu	Tyr	Asp	Asp	Gln	Thr	Ser	Gln	Arg	Glu	Lys	Glu	Asp
[3201]		130						135					140			
[3202]	Asp	Lys	Val	Phe	Pro	Gly	Gly	Ser	His	Thr	Tyr	Val	Trp	Gln	Val	Leu
[3203]	145					150					155					160
[3204]	Lys	Glu	Asn	Gly	Pro	Met	Ala	Ser	Asp	Pro	Leu	Cys	Leu	Thr	Tyr	Ser
[3205]					165					170						175
[3206]	Tyr	Leu	Ser	His	Val	Asp	Leu	Val	Lys	Asp	Leu	Asn	Ser	Gly	Leu	Ile
[3207]				180						185						190
[3208]	Gly	Ala	Leu	Leu	Val	Cys	Arg	Glu	Gly	Ser	Leu	Ala	Lys	Glu	Lys	Thr
[3209]				195						200				205		
[3210]	Gln	Thr	Leu	His	Lys	Phe	Ile	Leu	Leu	Phe	Ala	Val	Phe	Asp	Glu	Gly
[3211]		210						215						220		
[3212]	Lys	Ser	Trp	His	Ser	Glu	Thr	Lys	Asn	Ser	Leu	Met	Gln	Asp	Arg	Asp
[3213]	225					230					235					240
[3214]	Ala	Ala	Ser	Ala	Arg	Ala	Trp	Pro	Lys	Met	His	Thr	Val	Asn	Gly	Tyr
[3215]					245						250					255
[3216]	Val	Asn	Arg	Ser	Leu	Pro	Gly	Leu	Ile	Gly	Cys	His	Arg	Lys	Ser	Val
[3217]				260						265				270		
[3218]	Tyr	Trp	His	Val	Ile	Gly	Met	Gly	Thr	Thr	Pro	Glu	Val	His	Ser	Ile
[3219]				275							280				285	
[3220]	Phe	Leu	Glu	Gly	His	Thr	Phe	Leu	Val	Arg	Asn	His	Arg	Gln	Ala	Ser
[3221]		290						295						300		
[3222]	Leu	Glu	Ile	Ser	Pro	Ile	Thr	Phe	Leu	Thr	Ala	Gln	Thr	Leu	Leu	Met
[3223]	305					310					315					320
[3224]	Asp	Leu	Gly	Gln	Phe	Leu	Leu	Phe	Cys	His	Ile	Ser	Ser	His	Gln	His
[3225]					325						330				335	
[3226]	Asp	Gly	Met	Glu	Ala	Tyr	Val	Lys	Val	Asp	Ser	Cys	Pro	Glu	Glu	Pro
[3227]					340						345				350	
[3228]	Gln	Leu	Arg	Met	Lys	Asn	Asn	Glu	Glu	Ala	Glu	Asp	Tyr	Asp	Asp	Asp
[3229]				355							360				365	
[3230]	Leu	Thr	Asp	Ser	Glu	Met	Asp	Val	Val	Arg	Phe	Asp	Asp	Asp	Asn	Ser
[3231]				370							375				380	
[3232]	Pro	Ser	Phe	Ile	Gln	Ile	Arg	Ser	Val	Ala	Lys	Lys	His	Pro	Lys	Thr
[3233]	385					390						395				400

[3234]	Trp Val His Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro		
[3235]		405	410 415
[3236]	Leu Val Leu Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn		
[3237]		420	425 430
[3238]	Asn Gly Pro Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met		
[3239]		435	440 445
[3240]	Ala Tyr Thr Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu		
[3241]		450	455 460
[3242]	Ser Gly Ile Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu		
[3243]		465	470 475 480
[3244]	Leu Ile Ile Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro		
[3245]		485	490 495
[3246]	His Gly Ile Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys		
[3247]		500	505 510
[3248]	Gly Val Lys His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe		
[3249]		515	520 525
[3250]	Lys Tyr Lys Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp		
[3251]		530	535 540
[3252]	Pro Arg Cys Leu Thr Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg		
[3253]		545	550 555 560
[3254]	Asp Leu Ala Ser Gly Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu		
[3255]		565	570 575
[3256]	Ser Val Asp Gln Arg Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val		
[3257]		580	585 590
[3258]	Ile Leu Phe Ser Val Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu		
[3259]		595	600 605
[3260]	Asn Ile Gln Arg Phe Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp		
[3261]		610	615 620
[3262]	Pro Glu Phe Gln Ala Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val		
[3263]		625	630 635 640
[3264]	Phe Asp Ser Leu Gln Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp		
[3265]		645	650 655
[3266]	Tyr Ile Leu Ser Ile Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe		
[3267]		660	665 670
[3268]	Ser Gly Tyr Thr Phe Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr		
[3269]		675	680 685
[3270]	Leu Phe Pro Phe Ser Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro		
[3271]		690	695 700
[3272]	Gly Leu Trp Ile Leu Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly		
[3273]		705	710 715 720
[3274]	Met Thr Ala Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp		
[3275]		725	730 735

[3276]	Tyr Tyr Glu Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys
[3277]	740 745 750
[3278]	Asn Asn Ala Ile Glu Pro Arg Ser Phe Ser Gln Asn Gly Ala Pro Thr
[3279]	755 760 765
[3280]	Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr
[3281]	770 775 780
[3282]	Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser
[3283]	785 790 795 800
[3284]	Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Thr
[3285]	805 810 815
[3286]	Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr Glu Pro
[3287]	820 825 830
[3288]	Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr
[3289]	835 840 845
[3290]	Glu Glu Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser
[3291]	850 855 860
[3292]	Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala
[3293]	865 870 875 880
[3294]	Thr Pro Glu Ser Gly Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr
[3295]	885 890 895
[3296]	Glu Glu Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Ala
[3297]	900 905 910
[3298]	Ser Ser Pro Pro Val Leu Lys Arg His Gln Arg Glu Ile Thr Arg Thr
[3299]	915 920 925
[3300]	Thr Leu Gln Ser Asp Gln Glu Glu Ile Asp Tyr Asp Asp Thr Ile Ser
[3301]	930 935 940
[3302]	Val Glu Met Lys Lys Glu Asp Phe Asp Ile Tyr Asp Glu Asp Glu Asn
[3303]	945 950 955 960
[3304]	Gln Ser Pro Arg Ser Phe Gln Lys Lys Thr Arg His Tyr Phe Ile Ala
[3305]	965 970 975
[3306]	Ala Val Glu Arg Leu Trp Asp Tyr Gly Met Ser Ser Ser Pro His Val
[3307]	980 985 990
[3308]	Leu Arg Asn Arg Ala Gln Ser Gly Ser Val Pro Gln Phe Lys Lys Val
[3309]	995 1000 1005
[3310]	Val Phe Gln Glu Phe Thr Asp Gly Ser Phe Thr Gln Pro Leu Tyr
[3311]	1010 1015 1020
[3312]	Arg Gly Glu Leu Asn Glu His Leu Gly Leu Leu Gly Pro Tyr Ile
[3313]	1025 1030 1035
[3314]	Arg Ala Glu Val Glu Asp Asn Ile Met Val Thr Phe Arg Asn Gln
[3315]	1040 1045 1050
[3316]	Ala Ser Arg Pro Tyr Ser Phe Tyr Ser Ser Leu Ile Ser Tyr Glu
[3317]	1055 1060 1065

[3318]	Glu Asp Gln Arg Gln Gly Ala Glu Pro Arg Lys Asn Phe Val Lys		
[3319]	1070	1075	1080
[3320]	Pro Asn Glu Thr Lys Thr Tyr Phe Trp Lys Val Gln His His Met		
[3321]	1085	1090	1095
[3322]	Ala Pro Thr Lys Asp Glu Phe Asp Cys Lys Ala Trp Ala Tyr Phe		
[3323]	1100	1105	1110
[3324]	Ser Asp Val Asp Leu Glu Lys Asp Val His Ser Gly Leu Ile Gly		
[3325]	1115	1120	1125
[3326]	Pro Leu Leu Val Cys His Thr Asn Thr Leu Asn Pro Ala His Gly		
[3327]	1130	1135	1140
[3328]	Arg Gln Val Thr Val Gln Glu Phe Ala Leu Phe Phe Thr Ile Phe		
[3329]	1145	1150	1155
[3330]	Asp Glu Thr Lys Ser Trp Tyr Phe Thr Glu Asn Met Glu Arg Asn		
[3331]	1160	1165	1170
[3332]	Cys Arg Ala Pro Cys Asn Ile Gln Met Glu Asp Pro Thr Phe Lys		
[3333]	1175	1180	1185
[3334]	Glu Asn Tyr Arg Phe His Ala Ile Asn Gly Tyr Ile Met Asp Thr		
[3335]	1190	1195	1200
[3336]	Leu Pro Gly Leu Val Met Ala Gln Asp Gln Arg Ile Arg Trp Tyr		
[3337]	1205	1210	1215
[3338]	Leu Leu Ser Met Gly Ser Asn Glu Asn Ile His Ser Ile His Phe		
[3339]	1220	1225	1230
[3340]	Ser Gly His Val Phe Thr Val Arg Lys Lys Glu Glu Tyr Lys Met		
[3341]	1235	1240	1245
[3342]	Ala Leu Tyr Asn Leu Tyr Pro Gly Val Phe Glu Thr Val Glu Met		
[3343]	1250	1255	1260
[3344]	Leu Pro Ser Lys Ala Gly Ile Trp Arg Val Glu Cys Leu Ile Gly		
[3345]	1265	1270	1275
[3346]	Glu His Leu His Ala Gly Met Ser Thr Leu Phe Leu Val Tyr Ser		
[3347]	1280	1285	1290
[3348]	Asn Lys Cys Gln Thr Pro Leu Gly Met Ala Ser Gly His Ile Arg		
[3349]	1295	1300	1305
[3350]	Asp Phe Gln Ile Thr Ala Ser Gly Gln Tyr Gly Gln Trp Ala Pro		
[3351]	1310	1315	1320
[3352]	Lys Leu Ala Arg Leu His Tyr Ser Gly Ser Ile Asn Ala Trp Ser		
[3353]	1325	1330	1335
[3354]	Thr Lys Glu Pro Phe Ser Trp Ile Lys Val Asp Leu Leu Ala Pro		
[3355]	1340	1345	1350
[3356]	Met Ile Ile His Gly Ile Lys Thr Gln Gly Ala Arg Gln Lys Phe		
[3357]	1355	1360	1365
[3358]	Ser Ser Leu Tyr Ile Ser Gln Phe Ile Ile Met Tyr Ser Leu Asp		
[3359]	1370	1375	1380

[3360]	Gly Lys Lys Trp Gln Thr Tyr Arg Gly Asn Ser Thr Gly Thr Leu
[3361]	1385 1390 1395
[3362]	Met Val Phe Phe Gly Asn Val Asp Ser Ser Gly Ile Lys His Asn
[3363]	1400 1405 1410
[3364]	Ile Phe Asn Pro Pro Ile Ile Ala Arg Tyr Ile Arg Leu His Pro
[3365]	1415 1420 1425
[3366]	Thr His Tyr Ser Ile Arg Ser Thr Leu Arg Met Glu Leu Met Gly
[3367]	1430 1435 1440
[3368]	Cys Asp Leu Asn Ser Cys Ser Met Pro Leu Gly Met Glu Ser Lys
[3369]	1445 1450 1455
[3370]	Ala Ile Ser Asp Ala Gln Ile Thr Ala Ser Ser Tyr Phe Thr Asn
[3371]	1460 1465 1470
[3372]	Met Phe Ala Thr Trp Ser Pro Ser Lys Ala Arg Leu His Leu Gln
[3373]	1475 1480 1485
[3374]	Gly Arg Ser Asn Ala Trp Arg Pro Gln Val Asn Asn Pro Lys Glu
[3375]	1490 1495 1500
[3376]	Trp Leu Gln Val Asp Phe Gln Lys Thr Met Lys Val Thr Gly Val
[3377]	1505 1510 1515
[3378]	Thr Thr Gln Gly Val Lys Ser Leu Leu Thr Ser Met Tyr Val Lys
[3379]	1520 1525 1530
[3380]	Glu Phe Leu Ile Ser Ser Ser Gln Asp Gly His Gln Trp Thr Leu
[3381]	1535 1540 1545
[3382]	Phe Phe Gln Asn Gly Lys Val Lys Val Phe Gln Gly Asn Gln Asp
[3383]	1550 1555 1560
[3384]	Ser Phe Thr Pro Val Val Asn Ser Leu Asp Pro Pro Leu Leu Thr
[3385]	1565 1570 1575
[3386]	Arg Tyr Leu Arg Ile His Pro Gln Ser Trp Val His Gln Ile Ala
[3387]	1580 1585 1590
[3388]	Leu Arg Met Glu Val Leu Gly Cys Glu Ala Gln Asp Leu Tyr
[3389]	1595 1600 1605

ATGCAGATCGAACTGAGCACCTGCTTCTTTCTGTGCCTGCTGAGGTTTGTCTTTAGCGCCACCAGGAGATACTATCTGGGCGC
CGTGGAAGTGAAGTGGGACTATATGAGTCTGATCTGGGCGAACTGCCAGTGGATGCCAGGTTTCCCCCAGAGTGCCCAAAA
GCTTTCCCTTTAATACCAGCGTGGTGTATAAGAAAACCTGTTTGTGGAATTCACTGATCATCTGTTTAAATATCGCCAAGCCC
AGGCCCCCTGGATGGGCGTGTGGGCCCCACCATCCAGGCCGAAGTGATGATACCGTGGTCAACCCCTGAAAAACATGGC
CAGCCATCCAGTGAAGCTGCATGCTGTGGGCGTCTCTATTGGAAGCCTCTGAAGGCGCCGAGTATGATGATCAGACCAGCC
AGAGGAAAAAGAGATGATAAAGTCTTTCTGGGGCAGCCATACCTATGTCTGGCAGGTCTGAAAGAAAAATGGCCCCATG
GCCAGCGATCCCCTGTGCCTGACCTATAGCTATCTGAGCCATGTGGACCTGGTGAAGGATCTGAACAGCGGCGCTGATTGGGGC
CCTGCTGGTGTGCAGGGAAGGCGAGCCTGGCCAAAGAAAAACCCAGACCCTGCATAAGTTTATCCTGCTGTTTGGCGTGTG
ATGAAGGCAAAAGCTGGCATTCTGAAACCAAAAACAGCCTGATGCAGGACAGGGATGCCGCGCTCTGCCAGGGCCTGGCCCAAA
ATGCATACCGTGAATGGCTATGTGAATAGGAGCCTGCCTGGCCTGATTGGCTGCCACAGGAAAAAGCGTGTATTGGCATGTGAT
CGGCATGGGCACCAACCCCGAAGTGCATAGCATCTTTCTGGAAGGCCATACCTTCTGGTCAGGAACCAAGGCGAGCCAGCC
TGGAATCAGCCCCATCACCTTCTGACCGCCAGACCCTGCTGATGGATCTGGGCCAGTTTCTGCTGTTTGGCACATCTCC
AGCCATCAGCATGATGGCATGGAAGCCTATGTGAAAGTGCATAGCTGCCCGAAGAACCCAGCTGAGGATGAAAAACAATGA
AGAAGCCGAAGACTATGATGATGATCTGACTGATTCTGAAATGGATGTGGTCAGGTTTATGATGATAATAGCCCCAGCTTTA
TCCAGATCAGGAGCGTGGCCAAAAACATCCCAAGACCTGGGTGCATTATATCGCTGCTGAGGAAGAAGATTGGGACTATGCC
CCCTGGTGTGGCCCCCTGATGATAGGAGCTATAAAGCCAGTATCTGAACAATGGCCCCAGAGGATTGGCAGGAAGTATPA
AAAAAGTCAGGTTTATGGCCTACACTGATGAAACCTTCAAGACCAGGGAAGCCATCCAGCATGAGTCTGGCATCCTGGGCCCCC
TGCTGTATGGCGAAGTGGGGACACCCCTGCTGATCATCTTTAAAAATCAGGCCAGCAGGCCCTATAATATCTATCCCATGGC
ATCACTGATGTGAGGCCCCCTGTACAGCAGGAGGCTGCCCAAAGGCGTGAAACATCTGAAAGATTTCCTATCCTGCTGGCGA
AATCTTTAAGTATAAATGGACTGTGACTGTGGAAGATGGCCCCACCAAAAGCGATCCAGGTGCCTGACCAGGTATTATTCCA
GCTTTGTGAATATGGAACGCGATCTGGCCTCTGGCCTGATTGGCCCCCTGCTGATCTGCTATAAAGAGTCTGTGGACAGAGG
GGCAATCAGATCATGAGCGATAAAAGGAATGTCTCTGTTCTGTCTTTGATGAGAATAGGAGCTGGTACCTGACCGAAAA
CATCCAGAGGTTTCTGCCCAATCCCGCCGGCGTGCAGCTGGAAGATCCCGAGTTTCAGGCCAGCAATATCATGCATAGCATCA
ATGGCTATGTCTTTGATAGCCTGCAGCTGAGCGTGTGCCTGCATGAGGTGGCCTATTGGTATATCCTGAGCATCGGCGCCAG
ACCGATTCTTCTGAGCGTGTCTTTCTCTGGCTATACCTTTAAACATAAAATGGTGTATGAGGACACCCCTGACCCTGTTTCCCTT
CTCTGGCGAAACCGTGTATGAGCATGGAATCCCGCCCTGTGGATCTGGGCTGCCACAACAGCGATTTCAGGAACAGGG
GCATGACTGCTGCTGAAAGTCTCCAGCTGCGATAAAACACTGGGGACTATTATGAGGACAGCTATGAGGACATCAGCGCC
TATCTGCTGAGCAAGAACAATGCCATCGAACCAGGAGCTTTAGCCAGAATCCCCAGTGTGAAAAGGCATCAGAGGGAAAT
CACCAGGACCACCTGCAGTCTGATCAGGAAGAAATCGACTATGATGATACCATCAGCGTGGAATGAAGAAAGAGATTG
ATATCTATGATGAAGATGAAATCAGAGCCCCAGGAGCTTTAGCCAGAATCCCCAGTGTGAAAAGGCATCAGAGGGAAAT
CTGTGGGACTATGGCATGTCCAGCAGCCCCCATGTGCTGAGGAACAGGGCCAGTCTGGCAGCGTGCCCCAGTTTAAAAAGT
CGTGTTCAGGAGTTTACCGATGGCAGCTTTACCCAGCCCCCTGTATAGGGGCGAACTGAATGAACATCTGGGCTGCTGGGCG
CCTACATCAGGGCCGAAGTGAAGATAATATCATGGTGACCTTCAGGAACCAGGCCAGCAGGCCCTACAGCTTTTATTCCAGC
CTGATCAGCTATGAGGAAGATCAGAGGCGAGGGGCTGAGCCAGGAAAACTTTGTGAAACCAATGAAACCAAGACCTACTT
TTGGAAAGTCCAGCATCATATGGCCCCACCAAGGATGAATTTGATTGCAAAGCCTGGGCTACTTCTCTGATGTGGACCTGG
AAAAAGATGTGCATAGCGGCTGATTGGCCCCCTGCTGGTGTGCCACCAATACCTTGAACCTTGCCCATGGCAGGCAGGTG
ACTGTGCAGGAGTTTGGCTGTTTCTTTACCATCTTTGATGAAACCAAAAGCTGGTACTTCACCGAAAAACATGGAAGGAAGT
CAGGGCCCCCTGCAACATCCAGATGGAAGATCCACCTTTAAAGAAATATAGGTTCCATGCCATCAATGGCTATATCATGG
ATACCCTGCTGGCTGGTGTATGGCCAGGACAGAGGATCAGGTGGTATCTGCTGAGCATGGGCGCAATGAAACATCCAT
AGCATCCATTTCTCTGGCCATGTCTTTACCGTCAAGAAAAAGAGATATAAATGGCCCTGTATAATCTGTACCTGGGGT
GTTTGAACCGTGGAAATGCTGCCAGCAAGCCGGCATCTGGAGGGTGAATGCCTGATTGGCGAACATCTGCATGCTGGCA
TGAGCACCTGTTTCTGGTGTATAGCAATAAGTGCCAGACCCCCCTGGGCATGGCCTCTGGCCATATCAGGGATTTTCAGATC
ACTGCCTCTGGCCAGTATGGCCAGTGGGCCCCCAACTGGCCAGGCTGCATTATTCGGAAGCATCAATGCCTGGAGCACCAA
AGAACCCCTTAGCTGGATCAAAGTGCATCTGCTGGCCCCCATGATCATCCATGGCATCAAGACCCAGGGGGCCAGGCAGAAAT
TTTCAGGCTGTACATCAGCCAGTTTATCATCATGTATAGCCTGGATGGCAAAAAATGGCAGACCTATAGGGGCAATAGCACC
GGCACCTGATGGTGTCTTTGGCAATGTGGACAGCAGCGGCATCAACATAATATCTTTAATCCCCCATCATCGCCAGGTA
TATCAGGCTGCATCCCAACCATTATAGCATCAGGAGCACCTGAGGATGGAATGATGGGCTGCGATCTGAACAGCTGCAGCA
TGCCCCCTGGGCATGGAAGCAAGCCATCAGCGATGCCAGATCACTGCCTCCAGCTACTTCACTAATATGTTTGGCACCTGG
AGCCCCAGCAAGCCAGGCTGCATCTGCAGGGCAGGAGCAATGCCTGGAGGGCCAGGTCAACAATCCCAAGAAATGGCTGCA
GGTCGATTTTCAGAAAACCATGAAAGTCACTGGGGTGACCACCCAGGGGGTGAAGGCTGCTGACCAGCATGTATGTGAAG
AGTTTCTGATCTCCAGCAGCCAGGATGGCCATCAGTGGACCTGTCTTTTCAAGATGGCAAGTCAAAGTCTTTCAGGGCAAT
CAGGACAGCTTTACCCCTGTGGTGAATAGCCTGGATCCCCCTGCTGACCAGGTATCTGAGGATCCATCCCCAGAGCTGGGT
GCATCAGATCGCCCTGAGGATGGAAGTGTGGGCTGCGAAGCCAGGACCTGTACTGA

图1A:coFVIII-3-SEQ ID NO:1

ATGCAGATCGAGCTGAGCACGTGCTTCTTCTGTGCCTGCTGAGGTTCTGCTTCAGCGCCACCAGGAGGTACTACCTGGGCGC
CGTGGAGCTGAGCTGGGACTACATGCAGAGCGACCTGGGCGAGCTGCCCCTGGACGCCAGGTTCCCCCCCAGGGTGGCCAAAG
GCTTCCCCCTTCAACACGAGCGTGGTGTACAAGAAGACCCTGTTCTGTGGAGTTACCGACCATCTGTTCAATATCGCCAAAGCCC
AGGCCCCCTGGATGGGGCTGCTGGGGCCACGATCCAGGCGAGGTGTACGACACCGTGGTTCATCACCTGAAGAATATGGC
CAGCCACCCCTGAGCCTGCACGCCGTGGGCTGAGCTACTGGAAGGCCAGCGAGGGCGCCGAGTACGACGACCAGACCAGCC
AGAGGGAGAGGAGGACGACAAGGTGTTCCCCGGCGGACGCCACACCTACGTGTGGCAGGTGCTGAAGGAGAATGGGCCCATG
GCCAGCGACCCCTGTGCTGACCTACTCTTACCTGAGCCACGTGGATCTGGTGAAGGACCTGAACAGCGGCCCTGATCGGCGC
CCTGCTGGTGTGACGGGAGGGCAGCCTGGCCAAAGGAGAAGACCAGACCCTGCACAAGTTTCATCCTGCTGTTCCGCCGTGTTCTG
ACGAGGGCAAGAGCTGGCACAGCGAGACCAAGAACAGCCTGATGCAGGATAGGGACGCCGCCAGCGCCAGGGCCCTGGCCCAAG
ATGCACACCGTGAACGGCTACGTGAACAGGTCTCTGCCGGCCTGATCGGCTGCCACAGGAAGAGCGTGTACTGGCACGTGAT
CGGCATGGGGACCCACCCCGAGGTGCACAGCATCTTCTGGAGGGCCACACGTTCTGCTGAGGAATCACAGGACGGCCAGCC
TGGAGATCAGCCCGATCACCTTCTGACCGCCAGACCCTGCTGATGGACCTGGGGCAGTTCTGCTGTTCTGCCATATCAGC
TCTCACCAGCACGCGCATGGAGGCCTACGTGAAGGTGGATAGCTGCCCGAGGAGCCCGAGCTGAGGATGAAGAACAACGA
GGAGGCGGAGGACTACGACGACGACCTGACCGACAGCGAGATGGACGTGGTGAAGTTTCGACGACGACAATAGCCCGAGCTTCA
TCCAGATCAGGAGCGTGGCCAAAGACCCCAAGACCTGGGTGCATTACATCGCCGCGGAGGAGGATTGGGACTACGCC
CCCCCTGGTGTGGCCCCCGACGACAGGTCTTACAAGAGCCAGTACCTGAACAACGGGGCCCGAGAGGATCGGCAGGAAGTACAA
GAAGGTGAGGTTTCATGGCCTACACCGACGAGACCTTCAAGACCAGGGAGGGGATCCAGCACGAGAGCGGGATCTTGGGGCCCC
TGCTGTACGGCGAGGTGGGGCAGACGCTGCTGATCATCTTCAAGAACCAGGCCAGCAGGCCCTACAATATCTACCCCCACGGG
ATCACCGACGTGAGGCCCTGTACTCTAGGAGGCTGCCCAAGGGCGTGAAGCACCTGAAGGACTTCCCCATCTGCCCGGCGA
GATCTTCAAGTACAAGTGGACCTGACCGTGGAGGACGGGGCCACGAAGAGCGACCCAGGTGCCTGACCAGGTACTACAGCT
CTTTCGTGAACATGGAGAGGGACCTGGCCAGCGGCTGATCGGGCCCTGCTGATCTGCTACAGGAGAGCGTGGATCAGAGG
GGCAACCAGATCATGAGCGACAAGAGGAACGTGATCCTGTTGAGCGTGTTCGACGAGAATAGGTCTTGGTACCTGACCGAGAA
TATCCAGAGGTTCTGCCCAACCCCGCGCGTGCAGCTGGAGGATCCCGAGTTCAGGCGCAGCAACATCATGCACAGCATCA
ACGGCTACGTGTTTCGACAGCCTGCAGCTGAGCGTGTGCTGTCACGAGGTGGCCTACTGGTACATCTGAGCATCGCGCCAG
ACCGACTTCTGAGCGTGTCTTTCAGCGGCTACACCTTCAAGCAAGATGGTGTACGAGGATACCTGACCCTGTTCCCTT
CAGGGCCGAGACCTGTTTCATGAGCATGGAGAACCCTGGCTGTGGATCCTGGGCTGCCATAACTCCGACTTCAAGAAATAGG
GCATGACCGCCCTGCTGAAGGTGAGCTCTTTCGACAAGAACCCTGGCGACTACTACGAGGATAGCTACGAGGATATCAGCGCC
TACCTGCTGAGCAAGAACAACGCCATCGAGCCAGGTCTTTTCAGCCAGAACCCCGCTGCTGAAGAGGCACCGAGGGAGAT
CACCAGGACGACCTGACAGCGCACCAGGAGGAGATCGACTACGACGACAGATCAGCGTGGAGATGAAGAAGGAGGATTTG
ACATCTACGACGAGGACGAGAATCAGAGCCCAAGTCTTTCCAGAAGAAGACCAGGCATTACTTCATCGCCGCGCTGGAGAG
CTGTGGGACTACGGCATAGCAGCTCTCCCCACGTGCTGAGGAATAGGGCCAGAGCGGACGCGTGCACGCTTCAAGAGGT
GGTGTTCAGGAGTTTACCAGCGGACGCTTCAACCAGCCCTGTACAGGGCGAGCTGAACGAGCACCTGGGCCCTGCTGGGGC
CCTACATCAGGGCCGAGGTGGAGGATAACATCATGTTGACCTTCAGGAATCAGGCCAGCAGGCCCTATAGCTTCTATAGCTCT
CTGATCAGCTACGAGGAGGATCAGAGGCGAGGGCGCCGAGCCAGGAAGAATTCGTGAAGCCCAACGAGACCAAGACCTACTT
CTGGAAGGTGCAGCACCATGCCCCCACGAAGGACGAGTTCGACTGCAAGGCTGGGCTACTTACGCGACGTGGATCTGG
AGAAGGACGTGCACAGCGGCTGATCGGGCCCTGCTGTTGTCGCCACACCAACACCTGAACCCCGCCACGGCAGGCAAGGTG
ACCGTGCAGGAGTTGCGCCTGTTCTTCAACCATCTTCGACGAGACCAAGAGCTGGTACTTCACCGAGAATATGGAGAGGAATTG
CAGGGCCCTGCAATATCCAGATGGAGGACCCGACCTTCAAGGAGAATTACAGGTTCCACGCCATCAACGGCTACATCATGG
ACACGCTGCCCGGCTGGTCATGGCCAGGATCAGAGGATCAGGTGGTATCTGCTGAGCATGGGGAGCAACGAGAATATCCAC
AGCATCCTTTCAGCGCCACCTGTTTACCGTGAGGAAGAAGGAGGAGTACAAGATGGCCCTGTACAATCTGTACCCCGGCT
GTTTCGAGACCGTGGAGATGCTGCCAGCAAGGCGGGATCTGGAGGGTGGAGTGCTGATCGGCGAGCACCTGCACGCCGGCA
TGAGCACGCTGTTCTGCTGACTCTAACAAGTGCCAGACCCCTGGGGATGGCCAGCGGCCACATCAGGGACTTCCAGATC
ACCGCAGCGGCCAGTACGGCCAGTGGGCCCCAAGCTGGCCAGGCTGCACCTATTCGGAAGCATCAACGCCTGGAGCACGAA
GGAGCCCTTCAGCTGGATCAAGGTGGATCTGCTGGCCCCATGATCATCCACGGGATCAAGACCCAGGGCGCCAGGCAGAGT
TCAGCTCTCTGTATATCAGCCAGTTCATCATGATGACTCTCTGGACGGCAAGAAGTGGCAGACCTACAGGGGCAACAGCACC
GGCAGCTGATGGTGTCTTCGGCAACGTGGACTCTAGCGGGATCAAGCAATATCTTCAACCCCCCATCATCGCCAGGTA
CATCAGGCTGCACCCACCCATTACTCTATCAGGTCTACCTGAGGATGGAGCTGATGGGCTGCGACCTGAACAGCTGCAGCA
TGCCCCCTGGGGATGGAGAGCAAGGCCATCAGCGACGCCAGATCACCGCCAGCTCTTACTTCAACAATGTTTCGCCACCTGG
AGCCCCAGCAAGGCCAGGTGCACCTGCAGGGCAGGTCTAACGCCTGGAGGCCCGAGGTGAACAACCCCAAGGAGTGGCTGCA
GGTGGATTTCAGAAGACCATGAAGGTGACCGCGTGACCACGAGGGCGTGAAGAGCTGCTGACCAGCATGTACGTGAAGG
AGTTCCTGATCAGCTCTAGCCAGGACGGCCACCAAGTGGACCGTGTCTTTCAGAACGGCAAGGTGAAGGTGTTCCAGGGCAAC
CAGGATAGCTTACCCCCGTGGTGAACAGCCTGGACCCCCCTGCTGACCAGGTATCTGAGGATCCACCCCGAGGCTGGGT
GCACCAGATCGCCCTGAGGATGGAGGTGCTGGGCTGCGAGGCCAGGATCTGTATTGA

图1B:coFVIII-4-SEQ ID NO:2

ATGCAAACTCGAACTGAGCACCTGTTTCTTCTCTGCCTGCTGAGATTCTGTTTCTCCGCGACCCGCCGATACCTGGGAGCAGTGG
AGCTCTCTGGGATTACATGCAGAGCGACCTTGGGGAGCTGCCCGTGGATGCCAGGTTCCCTCCCGGGTGCCAAAGTCGTTTCCGTT
CAACACCTCCGTGGGTACAGAACTCTGTTCTGTGGAGTTCACCGACCACTGTTCAATATCGCCAGGCCAGACCTCCCTGGATG
GGGCTGTTGGGACCTACCATCCAAGCGGAGGTGTACGACACTGTGGTCATCACTCTGAAGAACATGGCCCTCGCATCCCGTGTCCCTGC
ACGCCGTGGGAGTGTCTTACTGGAAGCGTCCGAGGGGGCCGAATACGACGACCACTCGCAGAGAGAAAAGGAAGATGACAAGT
GTTCCAGGAGGATCGCACACCTACGTGTGGCAAGTGTGAAGGAGAACGGCCCAATGGCCCTCCGACCCGCTGTGCCCTGACCTACTCG
TACCTGTCCACGTGGACCTCGTGAAGGACCTCAACTCGGGACTGATTGGAGCCCTGCTGGTCTGCAGGGAAGGCTCACTGGCGAAAG
AAAAGACTCAGACCTTGCAACAAGTTCATTCTGCTGTTGCTGTGTCGACGAGGGGAAAGTCGTGGCACAGCGAGACTAAGAAGTCCCT
GATGCAAGATAGAGATGCCGCTCCGCCCGGGCTGGCCTAAGATGCACACCGTGAACGGTTACGTGAACCGCTCCCTCCCTGGCCCTG
ATTGGATGCCACCGGAAGTCCGTGTAAGTGGCAGTGTGCGGATGGGGACACCCCGAGGTGCACAGCATCTTCTGGGAAGGTACACA
CATTTCTCGTGCAGAACACCGGCGAGGCTCCCTGGAATCAGCCCCATTACCTTCTCTACTGCCAGACTCTGCTGATGGACCTGGG
ACAGTTCTGCTGTTCTGCCATATCTCCTCCACCAAGATGACGGAATGGAGGCATACGTGAAGGTGATTCCTGCCCTGAGGAACCC
CAGCTCCGATGAAGAACAATGAGGAAGCCGAGGACTACGACGACGACCTGACGGATAGCGAGATGGATGTGGTCCGCTTCGATGACG
ATAACAGCCCTTCTTCTCATCAAAATTCTGCTCGGTGGCAAGAGACACCCCAAGACCTGGGTGCATTACATTGCGGGGGAAGAGAGGA
CTGGGATTATGCCCGCTTGTCTGCTCCTGACGACCGGAGCTACAAGAGCCAGTACCTGAACAACGGTCCACAGAGGATCGGTAGA
AAGTACAAGAGGTCCGCTTCATGGCCTATACCGACGAAACCTTCAAACTAGAGAGGCCATCCAACAGGAATCCGGCATCTGGGCC
CGCTCTGTACGGAGAAGTCGGCGACACCTTCTCATTATCTTCAAGAACCAGGCTTCCCGGCCGTACAACATCTATCCGATGGGAT
CACTGACGTGCGCCACTGTAAGTCTGCGGGCGCTGGCCAGGGTGTCAAACACCTGAAGGATTTTCCGATCTTCCGGGAGAAATCTTC
AAGTACAAGTGGACCGTGACCGTGAAGATGGCCCAACTAAGTCTGACCCTAGATGCCTCACCCTACTACTCATCTTCTGTCACA
TGGAGCGCGACCTGGCCAGCGGACTGATCGGCCCGCTGCTGATTTGCTACAGGAATCAGTGGACCAACGGGGAAACAGATCATGTC
GGATAAGAGGAACGTCACTCTTCTCCGTGTTTACGAAACCGGTCGTGGTACCTGACTGAAACATCCAGCGGTTCTCCCAAC
CCCGCGGGCGTGCAGCTGGAAGATCCTGAGTTTCAAGCATCACTGACTCCATTAACGGTACGTGTTCCGATTCGCTGCGAGC
TGAGCGTGTGCTGCACGAAGTGGCCTACTGGTACATCTGTCCATTGGTGGCCAGACTGACTTCTGTCGCTGTTTTCTCCGGTA
CAGGTTCAAGCACAAGATGGTGTACGAGGACACCTGACCCTCTTCCCTTTTCCGGCGAACTGTGTTTATGAGCATGGAGAATCCC
GGCCTGTGGATCTTGGGCTGCCACAACAGCGACTTCCGTAACAGAGGAATGACTGCGCTGCTCAAGGTGTCCAGCTGCGACAGAACA
CCGGAGACTATTATGAGGACTCATACGAGGACATCTCCGCTACCTCCTGTCGAAGAATAACGCCATTGAACCTCGGAGCTTCAGCCA
GAACCCACCCGTGCTTAAGAGACATCAACGGGAGATCACTAGGACCACTTGCAGTCAAGCCAGGAGGAAATCGACTACGATGACACC
ATCTCGGTGAGATGAAGAAGGAGGACTTTGACATCTACGACGAAGATGAAACCCAGAGCCCGAGGTCGTTCCAAAAGAAAACCCGCC
ACTACTTTATTTGCTGCTGTGAGCGGCTGTGGGACTACGGAATGTGCTCCTCGCCGACGCTGCTCCGCAACCGAGCCAGAGCGGCTC
GGTGGCCGAATTCAAGAAGGTGCTGTTCCAGGAGTTCCTGACGGGAGCTTCACTCAGCCTTTGTACCGGGGAGAACTCAATGAACAT
CTCGGCCTCCTCGGACCTTACATCAGAGCAGAAGTGGAGATCAACATCATGCTGCTCACTTTCCGTAACCAAGCCAGCCGCGCTACTCGT
TCTACTCTCCCTCATTTCTTACGAAGAGGACCGCGGCGAGGCGCAGAACCAGCAAGAACTTCGTGAAGCCCAACGAAACCAAGAC
CTACTTCTGGAAGTGCAGCATCATATGGCCCGGACTAAGGACGAGTTTACTGCAAGGCTGGGCTACTTCTCCGATGTGGACTTG
GAGAAGGACGTCCACTCCGGCCTCATCGGTCCCTGCTGCTGTGCCATACCAATACCCTGAACCCCGCACAGGTGCGCAGGTACCG
TGAGGAGTTTCGCTCTGTTCTTCACTATCTTCAAGCAAACTAAGTCTGGTACTTCAAGGAGAACATGAGAGGAACTGCAGAGCCCC
CTGTAACATCCAGATGGAGGACCCGACGTTCAAGGAAAACCTACCGGTTCCAGCCATTAAACGGATACATCATGGATACGCTGCCGGGT
CTTGTGATGGCCAGGATCAACGGATCAGATGGTACTTATTGTGATGGGACGAAACGAGAACATCCACTCTATTCACTTCTCCGGTC
ATGTGTTCACTGTGCGGAAGAAGGAGATACAGATGGCCCTGTACAACCTTTATCCCGGAGTGTTCGAAACTGTGGAAATGCTGCC
GTGGAAGGCCGCAATTGGCGCGTGGAGTGTGTTGATTGGAGAATCTCCATGCGGGGATGTCAACCTGTTCTGTTGATATAGCAAC
AAGTGCCAGACTCCGCTTGGGATGGCGTCAGGACACATTAGGAGATTCCAGATCACTGCGTCCGGCCAGTACGGCCATGGGCCCCTA
AGCTGGCCCGCTGCATTACTCCGATCCATTAAAGCCTGGTCAACCAAGGAGCCATTCTCCTGGATCAAGGTGGACCTTCTGGCCCC
CATGATTATCCACGGAATTAAAGACCCAGGGGGCCCGGAGAAAGTTCTCCTCACTGTACATCAGCCAGTTCATAATCATGTACTCCCTG
GACGGAAGAAGTGGCAAACCTACAGGGGGAACGACACCGGCACACTGATGGTCTTTTTCGGAATGTGGACTCCTCCGGGATTAAGC
ATAACATCTTCAACCTCCGATTATCGCTCGGTACATTAGACTTCAACCTACCCACTACAGCATTCGCTCCACCTGCGGATGGAAT
GATGGGCTGCGATCTGAACCTGTCAGCATGCCGTTGGGAATGGAGTCCAAAGCAATTTCCGACGCGCAGATCACCGCTCGTCTAC
TTTACCAACATGTTGCCACGTGGTCACCGTCCAAGGCCCGGCTGCACCTCCAGGGAAGATCCAACGATGGCGGCCACAGGTCAACA
ACCCTAAGGAGTGGCTCCAGGTGGACTTCCAGAAAACATGAAGGTACCGGAGTCAACCCAGGGAGTGAAGTCTGCTGCTGACTTC
TATGTACGTCAAGGAGTTCTGATCTCCAGCAGCCAGGACGGGACCAAGTGGACCTGTTCTTCCAAATGGAAGGTCAAGGTGTTT
CAGGGCAATCAGGATTCATTACCCCGGTGGTGAACCTCCTTGATCCACCCCTCCTGACCCGCTACCTTCGCATCCACCCACAGTCTC
GGGTGCACAGATCGCGCTGAGGATGGAGGTCTGGGATGCAAGCCAGGACCTGTACTGA

图1C:coFVIII-5-SEQ ID NO:70

ATGCAGATTGAGCTGTCCACTTGTTCCTTCTGTGCCTCCTGCGCTTCTGTTTCTCCGCCACTCGCCGGTACTACCTTGGAGCCGTGG
AGCTTTTCATGGGACTACATGCAGAGCGACCTGGGCGAACTCCCCGTGGATGCCAGATTCCCCCCCCGCGTCCCAAAGTCTTCCCCCTT
TAACACCTCCGTGGTGTACAGAGAAACCCCTCTTTGTCTGAGTTCACTGACCACCTGTTCAACATCGCCAGCCGCGCCACCTTGGATG
GGCCTCCTGGGACCGACCATCAAGCTGAAGTGTACGACACCGTGGTGATCACCCTGAAGAACATGGCGTCCCACCCCGTGTCCCTGC
ATGCGGTGGAGTGTCTACTGGAAGGCCCTCCGAAGGAGCTGAGTACGACGACCACTAGCCAGCGGGAAAAGGAGGACGATAAAGT
GTTCCCGGGCGGCTCGCATACTTACGTGTGGCAAGTCTGAAGGAAAACGGACCTATGGCATCCGATCCTCTGTGCCTGACTTACTCC
TACCTTTCCCATGTGGACCTCGTGAAGGACCTGAACAGCGGGCTGATTGGTGCACTTCTCGTGTGCCGCGAAGGTTTCGCTCGCTAAGG
AAAAGACCCAGACCCCTCCATAAGTTCATCCTTTTGTTCGCTGTGTTTCGATGAAGGAAAGTCATGGCATTCCGAAACTAAGAACTCGCT
GATGCAGGACCGGATGCCGCTCAGCCCGCGCCTGGCCTAAAATGCATACAGTCAACGGATACGTGAATCGGTCACTGCCCGGGCTC
ATCGGTTGTACAGAAAGTCCGTGTACTGGCACGTATCGGCATGGGCACTACGCCCTGAAGTGCACTCCATCTTCTGGAAGGGCACA
CCTTCTCGTGCGCAACCACCGCCAGGCCCTCTGGAATCTCCCGATTACCTTCTGACCGCCAGACTCTGCTCATGGACCTGGG
GCAGTTCTCTTCTCTGCCACATCTCCAGCCATCAGCACGACGGAATGGAGGCCCTACGTGAAGGTGGACTCATGCCCGGAAGAACCT
CAGTTGCCGATGAAGAACACGAGGAGGCGGAGGACTATGACGACGATTTGACTGACTCCGAGATGGACGTGCTGCGTTTCGATGACG
ACAACAGCCCCAGCTTCATCCAGATTGCGAGCGTGGCCAAAGAAGCACCCCAAACCTGGGTGCACTACATCGCGCCGAGGAAGAAGA
TTGGGACTACGCCCGTTGGTGCTGGCACCCGATGACCGGTCTACAAGTCCAGTATCTGAACAATGGTCCGACGCGGATTGGCAGA
AAGTACAGAAAGTGGCGTTTCATGGCGTACACTGACGAAACGTTTAAAGACCCGGGAGGCCATTCAACATGAGAGCGGCATTCTGGGAC
CACTGCTGTACGGAGAGGTGGCGATACCTGCTCATCTTCAAAAACAGGCCCTCCCGGCCCTACAACATCTACCCTCAGGAAAT
CACCAGCTGCGGCCACTCTACTCGCGCGCCTGCCGAAGGGCTCAAGCACCTGAAAGACTTCCCTATCTCTCGCGGCGAAATCTTC
AAGTATAAGTGGACCGTCAACGCTGGAGGACGGGCCCAAGAGCGATCCTAGGTGTCTGACTCGGTACTACTCCAGCTTCGTGAACA
TGGAACGGGACCTGGCATCGGGAATCATTGGACCGCTGCTGATCTGCTACAAAGAGTCCGTGGATCAACGCGGCAACCAGATCATGTG
CGACAAGCGCAACGTGATCCTGTTCTCCGTGTTGATGAAAACAGATCCTGGTACCTCACTGAAAACATCCAGAGGTTCTCCCAAAC
CCCGCAGAGTCAACTGAGGAGCCCTGAGTTTCAGGCCCTCGAATATCATGCACTCGATTACGGTTACGTTTCGCTGCTCGAGTGCAGC
TGAGCGTGTGCTCCATGAAGTGCCTTACTGGTACATCTGTCCATCGCGCGCCAGACTGACTTCTTGAGCGTGTCTTTTCCGGTTA
CACCTTTAAGCACAAGATGGTGTACGAAGATACCTGACCCTGTTCCCTTCTCCGGCGAAACGGTGTTCATGTGATGGAGAACC
GGTCTGTGGATTCTGGGATGCCACAACAGCGACTTTCGGAACCGCGGAATGACTGCCCTGCTGAAGGTGTCTCATGCGACAAGAACA
CCGCGAGACTACTACGAGGACTCCTACGAGGATATCTCAGCCTACCTCTGTCACAGAACACCGCATGAGCGCGCGGACTTCAGCCA
GAACCCGCTGTGCTGAAGAGGCCACGCGAGAAATTACCCGAGCCACCTCCAATCGGATCAGGAGGAAATCGACTACGACGACACC
ATCTCGGTGGAATGAAGAAGGAAGATTTGATATCTACGACGAGGACGAAAATCAGTCCCTCGCTCATCCAAAAGAAAAGTAGAC
ACTACTTTATCGCCGCGGTGGAAAGACTGTGGGACTATGGAATGTATCCAGCCCTCAGTCTCTCGGAACCGGGCCAGAGCGGATC
GGTGCTCAGTTCAAGAAAGTGGTGTCCAGGAGTTCCCGGAGCGGAGCTTCAACCGCGCGCTGTACCGGGGAGAACTGAACGAACAC
CTGGGCTGCTCGGTCCCTACATCGCGCGGAAGTGGAGGATAACATCATGGTGACCTTCCGTAAACCAAGCATCCAGACCTTACTCCT
TCTATTCTCCCTGATCTCATACGAGGAGGACGCGCAAGGCGCGGAGCCCGCAAGAACTTCGTCAAGCCCAACGAGACTAAGAC
CTACTTCTGGAAGGTCCAACACCATATGGCCCCGACCAAGGATGAGTTTGAAGTGAAGGCTGGGCTACTTCTCCGACGTGGACCTT
GAGAAGGATGTCATTCCGGCTGATCGGGCGCTGCTGCTGTGTCACACCAACACCTGAACCCAGCGCATGGACCGCAGGTCAACG
TCCAGGAGTTTGTCTGTTCTTCAACATTTTGAAGAACTAAGTCTGTTACTTCAACGAGAAATATGGAGCGAAACTGTAGAGCGCC
CTGCAATATCCAGATGGAAGATCCGACTTCAAGGAGAACTATAGATTCCACGCCATCAACGGGTACATCATGGTACTCTGCCGGG
CTGGTCAATGGCCAGGATCAGAGGATTCGGTGGTACTTGCTGTCAATGGGATCGAACGAAAACATTCACTCCATTCACTTCTCGGT
ACGTGTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCGCTGTACAATCTGTACCCCGGGGTGTTGAAACTGTGGAGATGTGCC
GTCCAAGGCCGCGCATCTGGAGAGTGGAGTGCCTGATCGGAGAGCACCTCCACGCGGGGATGTCCACCTCTTCTGTTGTACTCGAAT
AAGTGCCAGACCCGCTGGGCATGGCCTCGGGCCACATCAGAGACTTCCAGATCAGCAAGCGGACAATACGGCCATGGGCGCCGA
AGCTGGCCCGCTTGCACTACTCCGGATCGATCAACGCATGGTCCACCAAGGAACCGTTCTCGTGGATTAAAGTGGACCTCCTGGCCCC
TATGATTATCCACGGAATTAAGACCCAGGGCGCCAGGCGAAGTTCTCCTCCCTGTACATCTCGCAATTATCATCATGTACAGCCTG
GACGGGAAGAAAGTGGCAGACTTACAGGGGAACTCCACCGGCACCTGATGGTCTTTTTCGGCAACGTGGATTCTCCGGCATTAAGC
ACAACATCTTCAACCCACCGATCATAGCCAGATATATTAGGCTCCACCCCACTCACTACTCAATCCGCTCAACTCTTCGGATGGAAT
CATGGGTGCGACCTGAACCTCTGCTCCATGCCGTGGGGATGGAATCAAAGGCTATTAGCGACGCCAGATCACCAGGAGCTCCTAC
TTCACTAATCATGTTCCGCCACCTGGAGCCCTCCAAGGCCAGGCTGCACTTGACGGGACGGTCAAAATGCTGGCGGCGCAAGTGAACA
ATCCGAAGGAATGGCTTCAAGTGGATTTCCAAAAGACCATGAAAGTGACCGGAGTCAACCCAGGGAGTGAAGTCCCTTCTGACCTC
GATGTATGTGAAGGAGTCTGATTAGCAGCAGCCAGGACGGCACCAAGTGGACCTGTTCTTCCAAAACGGAAAGGTCAAGGTGTTT
CAGGGGAACAGGACTCGTTCAACCCGTGGTGAACCTCCTGGACCCCACTGCTGACGCGGTACTTGAGGATTCATCCTCAGTCCT
GGGTCCATCAGATTGCATTGCGAATGGAAGTCTGGGCTGCGAGGCGCCAGGACCTGTACTGA

图1D:coFVIII-6-SEQ ID NO:71

ATGCAAACTCGAAGTACGACCTGTTCTCTCTGCTGCTGAGATTCTGTTTCTCCGCGACCCGCGGATACTACCTGGGAGC
AGTGGAGCTCTCTGGGATTACATGCAGAGCGACCTTGGGGAGCTGCCGTTGGATGCCAGGTTCCCTCCCCGGGTGCCAAAGT
CGTTTCCGTTCAACACCTCCGTGGTGTACAAGAACTCTGTTCTGGTGGATTACCGACCACTGTTCAATATCGCAAGCCC
AGACCTCCCTGGATGGGGCTGTTGGGACCTACCATCCAAGCGGAGGTGTACGACACTGTGGTCACTCTGAAGAATATGGC
CTCGCATCCGTGTCCCTGCACGCCGTGGGAGTGTCTTACTGAAAGCGTCCGAGGGGGCCGAATACGACAGCCAGACCTCGC
AGAGAGAAAGGAAGATGACAAGGTGTTCCAGGAGGATCGCACACCTACGTGTGGCAAGTGTGAAGGAGAACGGCCCAATG
GCCTCCGACCCGCTGTGCCTGACCTACTCGTACCTGTCCACGTGGACCTCGTGAAGGACCTCAACTCGGGACTGATTGGAGC
CCTGCTGGTCTGCAGGGAAGGCTCACTGGCGAAAGAAAGACTCAGACCTTGACAAAGTTCATTCTGCTGTTCCGTGTGTTTCG
ACGAGGGGAAGTCTGGCACAGCGACACTAAGAACTCCCTGATGCAAGATAGAGATGCCGCTCCGCCCGGGCCCTGGCCTAAG
ATGCACACCGTGAACGGTTACGTGAACCGCTCCCTCCCTGGCTGATTGGATGCCACCGGAAGTCCGTGTACTGGCAGCTGAT
CGGGATGGGGACCAACCCCGAGGTGCACAGCATCTTCCCTGGAAGGTCAACATTTCTCGTGGCAACCAACCGGCGAGCCCTCCC
TGGAAATCAGCCCCATTACCTTCTCACTGCCCAGACTCTGCTGATGGACCTGGGACAGTTCTGCTGTTCTGCCATATCTCC
TCCACCAACATGACGGAATGGAGGCATACGTGAAGTCTGATTCTGCCCTGAGGAACCCAGCTCCGATGAAGAAATGA
GGAAGCCGAGGACTACGACGACGACCTGACGGATAGCGAGATGGATGTGGTCCGGTTCGATGACGATAACAGCCCTTCTTCA
TCCAAATTCTCGTCCGTGGCAAGAAAGCACCCCAAGACCTGGGTGCATTACATTGCGGCGGAAGAAGAGGACTGGGATTATGCC
CCGCTTGTCTCGCTCTGACGACCGGAGCTACAAGAGCCGATACCTGAACAACGGTCCACAGAGGATCGGTAGAAGTACAA
GAAGTCCGCTTCATGGCTATACCGACGAAACCTTCAAACTAGAGAGGCCATCCAACACGAATCCGGCATCTTGGGCCCCG
TCTGTACGGAGAAGTGGCGACACCTTCTCATTATCTTCAAGAACAGGCTTCCGCGCCGTACAACATCTATCCGCATGGG
ATCACTGACGTGCGCCACTGTACTCGCGGCGCCTGCCAAGGGTGTCAAACACCTGAAGGATTTCCGATCTTCCGGGAGA
AATCTTCAAGTACAAGTGGACCGTGACCGTGAAGATGGCCCACTAAGTCTGACCTAGATGCCTCACCCTACTACTCAT
CCTTCGTCACATGGAGCGCGACCTGGCCAGCGGACTGATCGGCGCCGCTGCTGATTGCTACAAGGAATCAGTGGACCAACGG
GGAACACGATCATGTCCGATAAGAGGAACGTCATCTTCTCCGTCCTTGACGAAAACCGGTCGTGGTACCTGACCGAGAA
CATCCAGAGGTTCTTGCCCAACCTGCTGGGGTGCAGCTGGAGGACCCGAGTTCAGGCCAGCAACATCATGCACAGCATCA
ATGGCTACGTGTTTCGACAGCTGCAGCTGAGCGTGTGCCTGCACGAGGTGGCTTACTGGTACATCTTGAGCATCGGCGCCAG
ACCGACTTCTTGAGCGTGTCTTCTCTGGCTACACCTTCAAGCACAAGATGGTGTATGAGGACACCTGACCCCTGTTCCCTT
CAGCGGGGAGACTGTCTTCATGAGCATGGAGAACCCTGGCCCTGTGGATCCTGGGCTGCCACAACGCGACTTCAGGAACAGGG
GCATGACTGCCCTGTGAAAGTCTCCAGCTGTGACAGAACCAGGGGACTACTACGAGGACAGCTACGAGGACATCAGCGCC
TACCTGCTGAGCAAGAACAATGCCATCGAGCCAGGAGCTTCTCTCAGAACCCCCAGTGTGAAGAGGACCAGAGGGAGAT
CACCAGGACCAACCTGCAGTCTGACGAGGAGGATCGACTATGATGACACCATCAGCGTGGAGATGAAGAAGGAGGACTTCG
ACATCTACGACGAGGACGAGAACAGAGCCCAAGGAGCTTCCAGCAAGAAGACCAGGCACTACTTCATTGCTGCTGTGGAGAGG
CTGTGGGACTATGGCATGTCCAGCAGCCCCATGTGCTGAGGACAGGCGCCAGTCTGGCAGCGTGCCTTCAGAAAGT
CGTGTTCAGGAGTTCACCGACGGGAGCTTACCCAGCCCTGTACAGAGGGAGCTGAACGAGCACCCTGGGCTGCTGGGCGC
CCTACATCAGGGCCGAGGTGGAGGACAACATCATGTGACCTTCAGGAACAGGCCAGCAGGCTTACAGCTTCTACAGCAGC
CTGATCAGCTACGAGGAGGACAGAGGCGAGGGGCTGAGCCAGGAAGAATTTGTGAAGCCCAATGAACCAAGACCTACTT
CTGGAAGGTGCAGCACCACATGGCCCCACCAAGGACGAGTTCGACTGCAAGGCTGGGCTACTTCTCTGACGTGGACCTGG
AGAAGGACGTGCACCTCTGCTGATGGCCCCCTGCTGGTGTGCCACACCAACACCTGAACCTGCCATGGCAGGCAAGTG
ACTGTGACGAGGTTCCGCTGTTCTTCAACCATCTTCGATGAACCAAGAGCTGGTACTTCACTGAGAACATGGAGAGGAATG
CAGGGCCCCCTGCAACATCCAGATGGAGGACCCACCTTCAAGGAGAACTACAGGTTCCATGCCATCAATGGCTACATCATGG
ACACCTTGCTTGGCTGGTTCATGGCCAGGACAGGATCAGGTGGTATCTGCTGAGCATGGGACGCAACGAGAACATCCAC
AGCATCCACTTCTCTGGCCACGTGTTCACTGTGAGGAAGAAGGAGGAGTACAAGATGGCCCTGTACAACCTGTACCTGGGGT
GTTGGAACCCGTGGAGATGCTGCCAGCAAGGCCGGCATCTGGAGGGTGGAGTGCCTGATTGGGGAGCACCTGCACGCCGCA
TGAGCACCTGTCTCTGGTGTACAGCAACAAGTGCACACCCCTGGGCGATGGCCTCTGGCCACATCAGGGACTTCCAGATC
ACTGCTCTGGCCAGTACGGCCAGTGGGCCCCCAAGCTGGCCAGGCTGCACTACTCCGGAAGCATCAATGCTTGGAGCACC
GGAGCCCTTCAGCTGGATCAAGTGGACCTGCTGGCCCCCATGATCATCCACGGCATCAAGACCCAGGGGGCCAGGCGAAGT
TCTCCAGCTGTACATCAGCCAGTTCATCATGATGATCAGGCTGGACGGCAAGAAGTGGCAGACCTACAGGGGCAACAGCACC
GGCACCTGATGGTGTCTTCCGCAACGTGGACAGCAGCGGCATCAAGCACAACATCTTCAACCCCCCATCATCGCCAGATA
CATCAGGCTGCACCCCAACCTACAGCATCAGGAGCACCCTGAGGATGGAGCTGATGGGCTGTGACCTGAACAGCTGCAGCA
TGCCCTTGGGATGGAGAGCAAGGCCATCTCTGACGCCAGATCACTGCCTCCAGCTACTTACCAACATGTTTGGCACCTGG
AGCCCCAGCAAGGCCAGGCTGCACCTGCAGGGCAGGAGCAATGCCTGGAGGGCCAGGTCACAACCCCAAGGAGTGGCTGCA
GGTGGACTTCCAGAGACCATGAAGGTGACTGGGGTGACACCCAGGGGGTGAAGAGCTGCTGACAGCATGTACGTGAAGG
AGTTCCTGATCTCCAGCAGCCAGGACGGCCACCAAGTGGACCTGTTCTTCCAGAATGGCAAGGTGAAGGTGTTCCAGGGCAAC
CAGGACAGCTTACCCCTGTGGTCAACAGCTGGACCCCCCTGCTGACAGATACCTGAGGATCCACCCCAAGAGCTGGGT
GCACAGATCGCCCTGAGGATGGAGGTGCTGGGCTGTGAGGCCAGGACCTGTACTGA

实线下划线 = coFVIII-52-NT58 (SEQ ID NO: 7)

虚线下划线 = coFVIII-52-CT (SEQ ID NO: 8)

图1E:coFVIII-52-SEQ ID NO:3

ATGCAGATTGAGCTGTCCACTTGTCTTCTCTGTGCTCTCGGCTTCTGTTTCTCCGCCACTCGCCGGTACTACCTTGGAGC
 CGTGGAGCTTTCATGGGACTACATGCAGAGCGACCTGGGCGAACTCCCCGTGGATGCCAGATTCCCCCCCCCGCTGCCAAAGT
 CCTTCCCCCTTAACACCTCCGTGGTGTACAAGAAAACCTCTTTGTGCGATTCACTGACCACCTGTTCAACATCGCCAAGCCG
 CGCCACCTTGGATGGGCTCTGGGACCGACCATTCAAGCTGAAGTGTACGACACCGTGGTGATCACCTGAAGAATATGGC
 GTCCACCCCGTGTCCCTGCATGCGGTGCGAGTGTCTTACTGGAAGGCTCCGAAGGAGCTGAGTACGACGACGACACTAGCC
 AGCGGGAAAAGGAGGACGATAAAGTGTTCGGGGCGGCTCGCATACTTACGTGTGGCAAGTCTTGAAGGAAAACGACCTATG
 GCATCCGATCCTCTGTGCTGACTTACTCCTACCTTTCCCATGTGGACCTCGTGAAGGACCTGAACAGCGGGCTGATTGGTGC
 ACTTCTCGTGTGCGCGAAGGTTCCGTGCTAAGGAAAAGACCCAGACCTCCATAAGTTCATCTTTTGTTCGGTGTGTTTCG
 ATGAAGGAAAGTCAATGGCATTCCGAACTAAGAACTCGCTGATGCAGGACCGGGATGCCGCTCAGCCCGCGCTGGCTAAA
 ATGCATACAGTCAACGGATACGTGAATCGGTCACTGCCGGGCTCATCGGTTGTACAGAAAGTCCGTGACTGGCAGCTCAT
 CGGCATGGGCACTACGCTGAAGTGCATCCATCTTCTGGAGGGCACACCTTCCTCGTGGCAACCCAGCCGACGCTCTC
 TGGAAATCTCCCGATTACCTTTCTGACCGCCAGACTCTGCTCATGGACCTGGGGCAGTTCTTCTCTTCTGCCACATCTCC
 AGCCATCAGCAGCAGGAATGGAGGCTACGTGAAGTGGACTCATGCCGGAAGAACCTCAGTTCGGGATGAAGAACACGA
 GGAGGCGGAGGACTATGACGACGATTTGACTGACTCCGAGATGGACGTCGTGCGGTCGATGACGACAACAGCCCGAGCTTCA
 TCCAGATTCCGACGCTGGCCAAAGACCCCAAACCTGGGTGCACTACATCGCGGCGAGGAAGAAGATTGGGACTACGCG
 CGTGTGGTGTGGCACCCGATGACCGGTGCTACAAGTCCAGTATCTGAACAATGGTCCGACGCGGATTGGCAGAAAGTACAA
 GAAAGTGGGTTTCATGGCTACACTGACGAAACGTTTAAGACCCGGGAGGCCATTCAACATGAGAGCGGCATTCTGGGACAC
 TGCTGTACGAGAGGTCGGCGATACCTGCTCATCATCTTAAAAACAGGCTCCCGGCTTACAACATCTACCTCAGCGA
 ATCACCAGCTGCGGCCACTTACTCGCGGCGCTGCCGAAGGGCTCAAGCACCTGAAAGACTTCCCTATCTTCCCGGCGCA
 AATCTTCAAGTATGACCGCTCACCGTGGAGGACGGGCCACCAAGAGCGATCTAGGTGTCTGACTCGGTACTACTCCA
 GCTTCGTGAACATGGAACGGGACCTGGCATCGGACTCATTGGACCGCTGCTGATCTGCTACAAAGAGTCGGTGGATCAACGC
 GGCAACCAGATCATCTCCGACAAGCGCAACGTGATCTGTTCTCCGCTTTTGATGAAAACAGATCCTGGTACCTGACCGAGAA
 CATCCAGAGGTTCTGCCCAACCTGCTGGGGTGCAGCTGGAGGACCCGAGTTCCAGGCCAGCAACATCATGCACAGCATCA
 ATGGCTACGTGTTCCGACAGCCTGCAGCTGACGCTGTGCTGCACGAGGTGGCCTACTGGTACATCTTGAACATCGGCGCCAG
 ACCGACTTCTCGAGCGTGTCTTCTCTGGCTACACCTTCAAGCACAAGATGGTGTATGAGGACACCTGACCTGTTCCTTCCCTT
 CAGCGGGGAGACTGTCTTATGAGCATGGAGAACCTGGGCTGTGGATCCTGGGCTGCCACAACAGCGACTTCAAGAACAGGG
 GCATGACTGCCCTGTGAAAGTCTCCAGCTGTGACAAGAACCCGGGGACTACTACGAGGACAGCTACGAGGACATCAGCGCC
 TACCTGTGTAGCAAGAACATGCCATCGAGGCCAGGAGCTTCTCTCAGAACCCCGAGTGTGTAAGAGGCCAGAGGGAGAT
 CACCGAGGACACCTGCACTGTGACGAGGAGGATCGACTATGATGACCATCAGCGTGGAGATGAAGAAGGAGGACTTCG
 ACATCTACGACGAGGACGAGAACCAGAGCCCGAGGAGCTTCCAGAAGAGACAGGCACTACTTCTGCTGTGGAGAGG
 CTGTGGGACTATGGCATGTCCAGCAGCCCCATGTGCTGAGGACAGGGCCAGTCTGGCAGCGTGGCCGAGTTCAAGAAAGT
 CGTGTTCAGGAGTTTACCGACGGCAGCTTCAACCCAGCCCCGTGACAGAGGGGAGCTGAACGAGCACCTGGGCTGTCTGGGCC
 CCTACATCAGGGCCGAGGTGGAGGACAACATCATGTGTGACCTTCAGGAACAGGCCAGCAGGCCCTACAGCTTCTACAGAGC
 CTGATCAGCTACGAGGAGGACAGAGGAGGGGCTGAGCCAGGAAGAACTTGTGAAGCCCAATGAACCAAGACCTACTT
 CTGGAAGGTGCAGCACCATGAGCCCAAGGACGAGTTTCGACTGCAAGGCTGGGCTACTTCTCTGACGTGGACCTGG
 AGAAGGACGTGCACTCTGGCTGATTGGCCCCCTGTGCTGTGCCACCAACACCTGAACCTGGCCATGGCAGGCGAGTG
 ACTGTGACGAGGTTTCCCTGTCTTCTCACCATCTTCGATGAACCAAGAGCTGGTACTTCACTGAGAACATGGAGAGGAATG
 CAGGGCCCCCTGCAACATCCAGATGGAGGACCCACCTTCAAGGAGAACTACAGGTTCCATGCCATCAATGGCTACATCATGG
 ACACCTGCTTGGCTGGTTCATGGCCAGGACAGAGGATCAGGTGGTATCTGCTGAGCATGGGACGCAACGAGAACATCCAC
 AGCATCCACTTCTCTGGCCAGTGTCTACTGTGAGGAAGAAGGAGGAGTACAAGATGGCCCTGTACAACCTGTACCCTGGGGT
 GTTCCGAAACCGTGGAGATGTGCCCCAGCAAGCGCGCATCTGGAGGGTGGAGTGCCTGATTGGGGAGCACTGACGCGCGCA
 TGAGCACCTGTCTCTGGTGTACAGCAACAAGTGCCAGACCCCCCTGGGCATGGCCTCTGGCCACATCAGGGACTTCCAGATC
 ACTGCTCTGGCCAGTACGGCCAGTGGGCCCCAAGCTGGCCAGGCTGCACTACTCCGGAAGCATCAATGGCTGGAGCACCAA
 GGAGCCCTTCACTGATGATCAAGTGGACCTGCTGGCCCCATGATCATCCACGGCATCAAGACCCAGGGGGCCAGGCAGAAAGT
 TCTCCAGCCTGTATCATCAGCCAGTTTCATCATATGTACAGCCTGGACGGCAAGAAGTGGCAGACCTACAGGGGCAACAGCACC
 GGCACCTGTATGGTGTCTTCCGCAACGTGGACAGCAGCGGCATCAAGCACAACATCTTCAACCCCCCATCATCGCCAGATA
 CATCAGGCTGCACCCACCCACTACAGCATCAGGAGCACCTGAGGATGGAGCTGATGGGCTGTGACCTGAACAGCTGCAGCA
 TGCCCTTGGGCATGGAGAGCAAGGCCATCTCTGACGCCCAGATCACTGCCTCCAGCTACTTACCAACATGTTTGGCACCTGG
 AGCCCCAGCAAGGCCAGGCTGCACCTGCAGGGCAGGAGCAATGCCCTGGAGGCCAGGTCACCAACCCCAAGGAGTGGCTGCA
 GGTGGACTTCCAGAAGACCATGAAGTGAAGTGGGGTGACACCCAGGGGGTGAAGAGCCTGCTGACAGCATGTACGTGAAGG
 AGTTCTGATCTCCAGCAGCCAGGACGGCCACAGTGGACCTGTCTTCCAGAATGGCAAGGTGAAGTGTTCAGGGCAAC
 CAGGACAGCTTACCCCTGTGGTCAACAGCCTGGACCCCCCTGTGACAGATACCTGAGGATCCACCCCCAGAGCTGGGT
 GCACGAGATCGCCCTGAGGATGGAGGTGCTGGGCTGTGAGGCCAGGACCTGTACTGA

实线下划线 = coFVIII-62-NT58 (SEQ ID NO: 9)

虚线下划线 = coFVIII-62-CT (SEQ ID NO: 10)

图1F:coFVIII-62-SEQ ID NO:4

ATGCAGATTGAGCTGAGCACCTGCTTCTTCTGCTGCTGCTGAGGTTCTGCTTCTCTGCCACCAGGAGATACTACCTGGGGCG
 CGTGGAGCTGAGCTGGGACTACATGCAGTCTGACCTGGGCGAGCTGCCAGTGGACGCCAGGTTCCCCCCCAGAGTGCCCAAGA
 GCTTCCCCCTCAACACCAGCGTGGTGTACAAGAAGACCTGTTCGTGGAGTTCACTGACCACCTGTTCAACATCGCCAAAGCCC
 AGGCCCCCTGGATGGGCTGCTGGGCCCCACCATCCAGGCCGAGGTGTACGACACCGTGGTCACTACCTGAAGAATATGGC
 CAGCCACCCGCTCCTCCGTGCACGCCGTGGGGGTGAGCTACTGGAAGGCTCTGAGGGCGCCGAGTACGAGCAGCAGCAGCCAGCC
 AGAGGGAGAAGGAGGACACAAAGGTGTTCCCTGGGGGACGCCACCTACCTGTGGCAGGTCTGAAGGAGAAGCGGCCCATG
 GCCTCTGACCCCTGTGCCTGACCTACAGTACCTGAGCCACGTGGACCTGGTGAAGGACCTGAACTCTGGCCTGATTGGGGC
 CCTGCTGGTGTGACGGGAGGCGAGCCTGGCCAAAGAGAAGACCCAGACCTGCACAAGTTCATCTGCTGTTTCGCCGTGTTCG
 ACGAGGGCAAGAGCTGGCACTCTGAAACCAAGAACAGCCTGATGACGAGCAGGACGCCGCTCTGCCAGGGCCTGGCCCAAG
 ATGCACACCCGTCAACGGCTACGTCAACAGGAGCCTGCTTGCCTGATTGGCTGCCACAGGAAGAGCGTCTACTGGCATCTGAT
 CGGCATGGGACACCCTGAGGTGCACAGCATCTTCTGGAGGGCCACCTTCTGCTCAGGAACCAAGGACAGCCAGCCAGCC
 TGGAGATCAGCCCCATCACCCTTCTGACCCGCCAGACCTGCTGATGGACCTGGGCCAGTTCCTGCTGTTCTGCCACATCTCC
 AGCCACCAGCAGCAGCGCATGGAGGCTACGTGAAAGTGGACAGCTGCCCTGAGGAGCCCCAGCTGAGGATGAAGAACAACGA
 GGAGGCCGAGGACTATGATGACGACCTGACCGACAGCGAGATGGACGTGGTCAAGGTTGACGACGACAACAGCCCCAGCTTCA
 TCCAGATCAGGAGCGTGGCCAAAGAAGCACCCCAAGACCTGGGTGCACCTACATCGCTGCTGAGGAGGAGGACTGGGACTATGCC
 CCCCTGGTGTGGCCCTGATGACAGGAGCTACAAGAGCCAGTACCTGAACAATGGCCCCCAGAGGATTGGCAGGAAGTACAA
 GAAAGTCAGGTTTATGGCTTACCTGATGAAACCTTCAAGACAGGGAGGCCATCCAGCATGAGTCTGGCATCTTGGGCCCCC
 TGCTGTACGGGGAGGTGGGGGACACCTGCTGATCATCTTCAAGAACCAGGCCAGCAGGCCCTACAACATCTACCCCCATGGC
 ATCACCAGCTGAGGGCCCTGTACAGCAGGAGGCTGCCTAAGGGGGTGAAGCACCTGAAAGACTTCCCCATCTGCCTGGGGA
 GATCTTCAAGTCAAGTGGACTGTGACTGTGGAGGACGCCCCACCAAGAGCGACCCAGGTGCTGACCATGATCTACAGCA
 GCTTCGTCAACATGGAGAGGACCTGGCCTCTGGCCTGATTGGCCCCCTGCTGATCTGCTACAGGAGTCTGTGGACAGAGG
 GGCACACAGATCATGAGCGACAAGAGGAACCTGATCTCTCTCTCTTCTGCTTCTGACGAGAACAGGAGCTGGTACTGACTGAAAA
 CATCCAGCGGTTCTCCCCAACCCCGCGGGCTGCAGCTGGAAGATCCTGAGTTTCAGGCATCAAAATCATGCACTCCATTA
 ACGGCTACGCTTTCGATTTCGCTGCAGCTGAGCGTGTGCTGCAACGAAGTGGGCTACTGGTACATCCTGTCCATTGGTGCCAG
 ACTGACTTCCCTGTCTCGTGTCTTCTCCGGCTACAGTTCAAGCACAAGATGGTGTACGAGGACACCTGACCTCTTCCCTTT
 TTCCGGCGAAACTGTGTTTATGAGCATGGAGAAATCCCGGCTGTGGATCTTGGGCTGCCACAACAGCGACTTCCGTAACAGAG
 GAATGACTGCGCTGCTCAAGGTGTCCAGCTGCGACAAGAACCAGGAGACTATTATGAGGACTCATACGAGGACATCTCCGCC
 TACCTCTCTGCTCAAGATAACGCCATTGAACCTCGGAGCTTCAGCCAGAACCACCCGTGCTTAAGAGACATCAACGGGAGAT
 CACTAGGACACCTTCAGCTCAGACAGGAGGAAATCGACTACGATGACACCATCTCGCTCAGATGAAGAAGGAGGACTTTG
 ACATCTACGAGCAAGATGAAAACAGAGCCCGAGGTCTTCCAAAAGAAAACCGCCACTACTTTATGCTGCTCTCGAGCGG
 CTGTGGGACTACGGAATGTCTCTCCCGCAGCTGCTCCGCAACCGAGCCAGAGCGGCTCGGTGCGGCAATTCAAGAAGGT
 CGTGTTCAGGAGTTCACTGACGGGAGCTTCACTCAGCCTTTGTACCGGGGAGAACTCAATGAACATCTCGGCCCTCTCGGAC
 CTTAGATCAGAGCAGAGTGGAAAGATAACATCATGGTCACTTTCCGTAACCAAGCCAGCCGCCCCGCTACTCGTTCTACTCTCC
 CTCATTTCTTACGAAGAGGACAGCGGCGAGGCGCAGAACCGCGAAGAACTTCGTGAAGCCCAACGAACCAAGACCTACTT
 CTGGAAAGTGCAGCATATATGGCCCGGACTAAGGACGAGTTTGACTGCAAAAGCTGGGCTACTTCTCTCGATGTGGACTTGG
 AGAAGGACGTCACCTCCGGCTCATCGGTCCCTGCTGCTGTGCTGCTACCAATACCTGAAACCCGACAGGCTGCGCAGGTC
 ACCGTGCAGGAGTTCGCTCTGTTCTTCACTATCTTCGACGAAACTAAGTCTCGGTACTTCACCGAGAACATGGAGAGGAATG
 CAGAGCCCCCTGTAAATCCAGATGGAGGACCCGACGTTCAAGGAAAATACCGGTTCCACGCCATTAACGGATACATCATGG
 ATACGCTGCCGGGCTTTGTGATGGCCAGGATCAACGGATCAGATGGTACTTATTGTGATGGGAGCAACGAGAATCCAC
 TCTATTCACTTCTCCGGTCATGTGTTCACTGTGCGGAAGAAGGAGTACAAGATGGCCCTGTACAACCTTTATCCCGGAGT
 GTTTCGAAACTGTGGAAATGCTGCGCTCGAAGGCGCGCATTTGCGCGCTGGAGTGTGTTGATTGGAGAATCTCCATGCGGGGA
 TGTCAAACCTGTTCTGGTGTATAGCAACAAGTGCCAGACTCCGCTTGGGATGGCGTCAGGACACATTAGGGATTTCAGATC
 ACTGCGTCCGGCCAGTACGGCCAATGGGCCCCTAAGCTGGGCCGCTGCATTACTCCGGATCCATTAACGCCCTGGTCAACCAA
 GGAGCCATTCTCTTGGATCAAGGTGGACCTTCTGGCCCCATGATTATCCACGGAATTAAGACCCAGGGGGGCGGCGAGAAGT
 TCTCTCACTGTACATCAGCCAGTTTCAATCATGTACTCCCTGGACGGAAGAAGTGGCAAACTTACAGGGGGAAACAGCACC
 GGCACACTGATGGTCTTTTTCGGAATGTGGACTCTCCGGGATTAAGCATAACATCTTCAACCTCCGATTATCGCTCGGTA
 CATTAGACTTCACCTACCCACTACAGCATTCGCTCCACCCTGCGGATGGAACCTGATGGGCTGCGATCTGAACTCGTGCAGCA
 TGCCGTTGGGAATGGAGTCCAAAGCAATTTCCGACGCGCAGATCACCGCCTCGTCTACTTTACCAACATGTTTCGCCACGTGG
 TCACCGTCCAAAGGCCGCTGCACCTCCAGGGAAGATCCAACGCATGGCGGCCAGGTTCAACAACCTAAGGAGTGGCTCCA
 GGTGGACTTCCAGAAAACCATGAAGTCAACGGAGTCAACCCAGGGAGTGAAGTCTGCTGCTGACTTCTATGTACGCTCAAGG
 AGTTCCTGATCTCAGCAGCCAGGACGGGACCAAGTGGACCTGTTCTTCCAAAATGGAAAGGTCAAGGTGTTTCAGGGCAAT
 CAGGATTCATTCACCCCGGTGGTGAACCTCCCTGATCCACCCCTCTGACCCGCTACCTTCGCATCCACCACAGTCTCTGGGT
 GCACCAGATCGCGCTGAGGATGGAGGTCTTGGGATGCGAAGCCAGGACCTGTACTGA

实线下划线 = coFVIII-25-NT58 (SEQ ID NO: 11)

虚线下划线 = coFVIII-25-CT (SEQ ID NO: 12)

图1G:coFVIII-25-SEQ ID NO:5

ATGCAGATTGAGCTGAGCACCTGCTTCTTCTGTGCTGCTGAGGTTCTGCTTCTCTGCCACGAGGAGATACTACCTGGGCGC
CGTGGAGCTGAGCTGGGACTACATGCAGTCTGACCTGGGCGAGCTGCCAGTGGACGCCAGGTTCCCCCCAGAGTGCCCAAGA
GCTTCCCCTTCAACACCAGCGTGGTGTACAAGAAGACCTGTTCGTGGAGTTCACTGACCACCTGTTCAACATCGCCAAAGCCC
AGGCCCCCTGGATGGGCTGCTGGGCCCCACCATCCAGGCCGAGGTGTACGACACCGTGGTCACTACCTGAAGAATATGGC
CAGCCACCCGCTCTCCCTGCACGCCGTGGGGGTGAGTACTGGAAGGCCTCTGAGGGGCGCGAGTACGACGACACGACACGCC
AGAGGGAGAAGGAGGACGACAAGGTGTTCCCTGGGGGCGAGCCACACCTACGTGTGGCAGGTCTGAAGGAGAAGCGCCCCATG
GCCTCTGACCCCTGTGCTGACCTACAGCTACCTGAGCCACGTGGACCTGGTGAAGGACCTGAACTCTGGCCGTATTGGGGC
CCTGCTGGTGTGACGGGAGGGCAGCCTGGCCAAAGGAGAAGACCCAGACCTGCACAAGTTCACTCTGCTGTTCCCGCTGTTCG
ACGAGGGCAAGAGCTGGCACTCTGAAACCAAGAAGACGCTGATGCAGGACAGGGACGCCCTCTGCCAGGGCTGGCCCAAG
ATGCACACCGTCAACGGCTACGTCAACAGGAGCCTGCTTCCCTGGAGGCTGATTGGCTGCCACAGGAAGAGCGTACTGGCATGTGAT
CGGATGGGACCCCTGAGGTGCACAGCATCTTCTGGAGGCGCCACACCTTCTGCTGAGGAACACAGGACGGCCAGCC
TGGAGATCAGCCCCATCACCTTCTGACCGCCAGACCTGCTGATGGACCTGGGCGAGTTCTGCTGTTCTGCCACATCTCC
AGCCACGACGACGAGCGCATGGAGGCTACGTGAAAGTGGACAGCTGCCCTGAGGAGCCCGAGCTGAGGATGAAGAACAACGA
GGAGGCGGAGGACTATGATGACGACCTGACCGACAGCGAGATGGACGTGGTCAAGGTTGACGACGACAACAGCCCCAGCTTCA
TCCAGATCAGGAGCGTGGCCAAAGAAGCACCCCAAGACCTGGGTGCACTACATCGTGTGAGGAGGAGGACTGGGACTATGGCC
CCCCTGGTGTGGCCCTGATGACAGGAGCTACAAGAGCCAGTACCTGAACAATGGCCCCCAGAGGATTGGCAGGAAGTACAA
GAAAGTCAAGTTCATGGCTACACTGATGAACTTCAAGACCAGGGAGGCCATCCAGCATGAGTCTGGCATCTTGGGCCCCC
TGCTGTACGGGGAGGTGGGGGACACCTGCTGATCATCTTCAAGAACCAGGCCAGCAGGCCCTACAACATCTACCCCCATGGC
ATCACCGAGCTGAGGCCCCGTACAGCAGGAGGCTGCCTAAGGGGGTGAAGCACCTGAAAGACTTCCCCATCTGCCTGGGGA
GATCTTCAAGTACAAGTGGACTGTGACTGTGGAGGAGCGGCCACCAAGAGCGACCCAGGTGCTGACCATGACTACAGCA
GCTTCGTCAACATGGAGAGGACCTGGCCTCTGGCCTGATTGGCCCCCTGCTGATCTGCTACAGGAGTCTGTGGACGAGG
GGCAACCAGATCATGAGCGACAAGAGGAACGTGATCTGTTCTCTGCTTTCGACGAGAACAGGAGCTGGTACCTCACTGAAAA
CATCCAGAGGTTCTTCCAAACCCCGCAGGAGTGCAACTGGAGGACCTGAGTTTCAAGGCTCGAATATCATGCACTCGATTA
ACGTTTACGTGTTTCGACTCGCTGCAGCTGAGCGTGTGCTTCCATGAAGTCCCTTACTGGTACATCTGTCCTATCGCCGCGCCAG
ACTGACTTCCGTGAGCGTGTCTTTTCCGGTTACACCTTTAAGCACAAGATGGTGTACGAAGATACCTGACCTGTTCCTTT
CTCCGGCGAAACGGTGTTCATGTGATGGAGAACCCTGGGTCTGTGGATTCTGGGATGCCACAACAGCGACTTTCGGAACCGCG
GAATGACTGCCCTGTGGAAGGTGTCTCATGCGACAAGAACCCTGGAGACTACTACGAGGACTCTTACGAGGATATCTACGCC
TACCTCTCTGTCGAAGAACAACGCGATCGAGGCGCGCAGCTTACGCCAGAACCCTGCTGTGAAGAGGACCCAGCGAGAAAT
TACCCGGACCCCTTCAATCGGATCAGGAGGAAATCGACTACGACGACCATCTCGGTGGAAATGAAGAAGGAGATTTCG
ATATCTACGACGAGGACGAAATCAGTCCCTCGCTCATTTCCAAAAGAACTAGACACTACTTTATCGCCGCGGTGGAAAGA
CTGTGGGACTATGGAATGTATCCAGCCCTCACGCTCTTGGAAACCGGGCCAGAGCGGATCGGTGCCCTCAGTTCAAGAAAGT
GGTGTTCAGGAGTTTACCGACGGCAGCTTACCCAGCCGCTGTACCGGGGAGAACTGAACGAACACCTGGGGCTGTCTCGGT
CCTACATCCGCGCGGAAGTGGAGGATAACATCATGTTGACCTTCCGTAACCAAGCATCCAGACCTTACTCTCTTATTCCTCC
CTGATCTCATACGAGGAGGACAGCGCCAAAGCGCGCAGCCCGCAAGAACTTCGTCAAGCCCAACGAGACTAAGACCTACTT
CTGGAAGGTCCAACACCATATGGCCCCGACCAAGGATGAGTTTGTACTGCAAGGCTTGGGCTACTTCTCCGACGTGGACCTTG
AGAAGGATGTCCATTCCGGCTGATCGGGCGCTGCTGCTGTGTACACCAACACCTGAACCCAGCGCATGGACGCCAGGTC
ACCGTCCAGGAGTTTCTCTGTTCTTACCATTTTGTACGAACTAAGTCTGCTGCTTACCGGAGATATGGAGCGAACTG
TAGAGCGCCTGCAATATCCAGATGGAAGATCCGACTTTCAAGGAGAACTATAGATTCCACGCCATCAACGGGTACATCATGG
ATACTCTGCGGGGCTGGTTCATGGCCAGGATCAGAGGATTCGGTGGTACTTGTCTGTCATGGGATCGAACGAAACATTAC
TCCATTCACTTCTCCGGTCACGTGTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCGCTGTACAATCTGTACCCCGGGGT
GTTGAAACTGTGGAGATGTGCGCTCCAAGGCGCGCATCTGGAGAGTGGAGTGCCTGATCGGAGAGCCTCCACGCGGGGA
TGTCCACCCTTCTCTGGTGTACTCGAATAAGTGCCAGACCCGCTGGGCATGGCCTCGGGCCACATCAGAGACTTCCAGATC
ACAGCAAGCGGACAAATACGGCCAATGGGCGCGGAAGCTGGCCGCTTGCACCTACTCCGATCGATCAACGCATGTTCCACCAA
GGAACCGTTCTGCTGGATTAAGGTGGACCTCTGGCCCTATGATTATCCACGGAATTAAGACCCAGGGCGCCAGGCGAAGT
TCTCTCCCTGTACATCTCGCAATTATCATCATGTACAGCCTGGACGGGAAGAAGTGGCAGACTTACAGGGGAACTCCACC
GGCACCTGATGGTCTTTTTTCGGCAACGTGGATTCTCCGGCATTAAGCACAACATCTTCAACCCACCGATCATAGCCAGATA
TATTAGGCTCCACCCCACTCACTACTCAATCCGCTCAACTCTTCGGATGGAACCTCATGGGGTGCAGCTGAACCTCTGTCCA
TGCCGTTGGGGATGGAATCAAAGGCTATTAGCGACGCCAGATCACCAGCGAGCTCTACTTCACTAACATGTTTCGCCACCTGG
AGCCCTCCAAAGGCTAGGCTGCACTTGCAGGACGGTCAAAATGCCTGGCGGCCGCAAGTGAACAATCCGAAGGAATGGCTTCA
AGTGGATTTCCAAAAGACCATGAAAGTGACCGGAGTACCACCCAGGGAGTGAAGTCCCTTCTGACCTCGATGTATGTGAAGG
AGTTCCTGATTAGCAGCAGCCAGGACGGGACCAAGTGGACCTGTCTTCCAAAACGAAAGGTCAAGGTGTTCCAGGGGAAC
CAGGACTCGTTCACACCCGTGGTGAATCCCTGGACCCCCACTGCTGACGCGGTACTTGAGGATTATCCTCAGTCTGGGT
CCATCAGATTGCATTGCGAATGGAAGTCTTGGGCTGCGAGGCCAGGACCTGTACTGA

实线下划线 = coFVIII-26-NT58 (SEQ ID NO: 13)

虚线下划线 = coFVIII-26-CT (SEQ ID NO: 14)

图1H:coFVIII-26-SEQ ID NO:6

ATGCAATAGAGCTCTCCACCTGCTCTTTCTGTGCTTTTGTGCGATTCTGCTTTAGTGCCACCAGAAGATACTACCTGGGTGCAGTG
GAACTGTCATGGGACTATATGCAAGTGATCTCGGTGAGCTGCTGTGGACGCAAGATTTCTCTAGAGTGCCAAATCTTTCC
ATTCAACACCTCAGTCGTGTACAAAAAGACTCTGTTGTAGAATTCACGGATCACCTTTTCAACATCGCTAAGCCAAGGCCACCCT
GGATGGGTCTGCTAGGTCTACCATCCAGGCTGAGGTTTATGATACAGTGGTCACTTACACTAAGAACATGGCTTCCCATCTGTG
AGTCTTCATGCTGTTGGTGTATCCTACTGGAAAGCTTCTGAGGGAGCTGAATATGATGATCAGACCAGTCAAAGGGAGAAAGAA
GATGATAAAGTCTTCCCTGGTGGAGCCATACATATGTCTGGCAGGTCCTGAAAGAGAATGGTCAATGGCCTCTGACCCACTGT
GCCTTACCTACTCATATCTTTCTCATGTGGACCTGGTAAAGACTTGAATTCAGGCCTCATTGGAGCCCTACTAGTATGTAGAGAA
GGGAGTCTGGCCAAGGAAAAGACACAGACCTTGACAAAATTTATACTACTTTTTGCTGTATTTGATGAAGGGAAAAGTTGGCACT
CAGAAACAAAGAACTCCTTGATGCAGGATAGGGATGCTGCATCTGCTCGGGCCTGGCCTAAAATGCACACAGTCAATGGTTATGT
AAACAGGTCTCTGCCAGGTCTGATTGGATGCCACAGGAAATCAGTCTATTGGCATGTGATTGGAATGGGCACCACCTCTGAAGTG
CACTCAATATTCCTCGAAGGTCACACATTTCTGTGAGGAACCATCGCCAGGCGTCTTGGAATCTCGCCAATAACTTTCTTACT
GCTCAACACTCTTGATGGACCTTGGACAGTTTCTACTGTTTGTATATCTTCTCCCAACATGATGGCATGGAAGCTTATGTC
AAAGTAGACAGCTGTCCAGAGGAACCCCACTACGAATGAAAAATAATGAAGAAGCGGAAGACTATGATGATGATCTTACTGAT
TCTGAAATGGATGTGGTCAGGTTTGTGATGACAACTCTCTTCTTATCCAAATTCGCTCAGTTGCCAAGAAGCATCTAAAAT
TGGGTACATTACATTGCTGCTGAAGAGGAGGACTGGGACTATGCTCCCTAGTCTCGCCCCGATGACAGAAGTTATAAAAGTC
AATATTTGAACATGGCCCTCAGCGGATTGGTAGGAAGTACAAAAAGTCCGATTTATGGCATAACAGATGAAACCTTTAAGAC
TCGTGAAGCTATTAGCATGAATCAGGAATCTGGGACCTTACTTTATGGGGAAGTTGGAGACACACTGTTGATTATATTAAGA
ATCAAGCAAGCAGACCATATAACATCTACCTCACGGAATCACTGATGTCCTGCTTTGATTCAAGGAGATTACCAAAAGGTGTA
AAACATTTGAAGGATTTTCAATTTCTGCCAGGAGAAATATCAAATATAATGGACAGTGACTGTAGAAGATGGGCCAACTAAAT
CAGATCCTCGGTGCTGACCCGCTATTACTAGTTTCTGTTAATATGAGAGAGATCTAGCTTCAGGACTCATTGGCCCTCTCTCA
TCTGCTACAAAGAACTGTAGATCAAAGAGGAAACCAGATAATGTCAGACAAGAGGAATGTCATCTGTTTTCTGTATTTGATGA
GAACCGAAGCTGGTACCTCAGAGAAATATACAACGCTTCTCCCAATCCAGCTGGAGTGCAGCTTGAGGATCCAGAGTTCCAA
GCCTCAACATCATGCACAGCATCAATGGCTATGTTTTGATAGTTTGCAGTTGTCAGTTGTTTGCATGAGGTGGCATACTGGTA
CATTCTAAGCATTGGAGCACAGACTGACTTCTTCTGTCTTCTCTGATATACCTCAAACACAAAATGGTCTATGAAGACAC
ACTCACCTATTCCATTCTCAGGAGAACTGTCTTCATGTCGATGGAAAACCCAGGTCTATGGATTCTGGGGTGCCCACTCAG
ACTTTCGGAACAGAGCATGACCCGCTTACTGAAGTTTTCTAGTTGTGACAAGAACACTGGTGATTATTACGAGGACAGTTATGA
AGATATTTAGCATACTTGTGAGTAAAAACAATGCCATTGAACCAAGAAGCTTCTCTCAAAACCCACAGCTTGAAACGCCATC
AACGGGAAATAACTCGTACTACTCTCAGTCAGATCAAGAGGAAATGACTATGATGATACCATATCAGTTGAAATGAAGAAGGA
AGATTTTGACATTTATGATGAGGATGAAAATCAGAGCCCCGAGCTTTCAAAGAAAACACGACACTATTTTATTGCTGCAGTGG
AGAGGCTCTGGGATTATGGGATGAGTAGTCCCCACATGTTCTAAGAAACAGGGCTCAGAGTGGCAGTGTCCCTCAGTTCAAGA
AAGTTGTTTTCCAGGAATTTACTGATGGCTCTTTACTCAGCCCTTATACCGTGGAGAACTAAATGAACATTTGGGACTCCTGGGG
CCATATATAAGAGCAGAAGTTGAAGATAATATCATGGTAACCTTTCAGAAATCAGGCCTCTCGTCCCTATTCTTCTATTCTAGCCTT
ATTTCTTATGAGGAAGATCAGAGGCAAGGAGCAGAACCTAGAAAAAATTTGTCAAGCCTAATGAACCAAAACTTACTTTTGA
AAGTGCAACATCATATGGCACCCACTAAAGATGAGTTTGACTGCAAGCCTGGGCTTATTTCTCTGATGTTGACCTGGAAAAAGAT
GTGCACTCAGGCCTGATTGGACCCCTCTGCTGTGCCACACTAACACACTGAACCCTGCTCATGGGAGACAAGTGACAGTACAGG
AATTTGCTCTGTTTTTACCATCTTTGATGAGACAAAAGCTGGTACTTCACTGAAAATATGGAAGAACTGCAGGGCTCCCTGC
AATATCCAGATGGAAGATCCCACTTTTAAAGAGAATTATCGCTTCCATGCAATCAATGGCTACATAATGGATACACTACCTGGCT
AGTAATGGCTCAGGATCAAAGGATTGATGGTATCTGCTCAGCATGGGCAGCAATGAAAACATCCATTCTATTCAATTCAGTGGA
CATGTGTTCACTGTACGAAAAAAGAGGAGTATAAAATGGCACTGTACAATCTCTATCCAGGTGTTTTGAGACAGTGGAATGT
TACCATCCAAAGCTGGAATTTGGCGGTGGAATGCCTTATTGGCGAGCATCTACATGCTGGGATGAGCACACTTTTTCTGGTGA
CAGCAATAAGTGTGAGACTCCCTGGGAATGGCTTCTGGACACATTAGAGATTTTCAAGATTACAGCTTCAGGACAATATGGACAG
TGGGCCCCAAAGCTGGCCAGACTTCAATTATCCGGATCAATCAATGCCTGGAGCACCAGGAGCCCTTTCTTGGATCAAGGTGG
ATCTGTTGGCACCAATGATTATTCACGGCATCAAGACCCAGGGTGGCCGTGAGAAGTTCTCCAGCTCTACATCTCTCAGTTTATCA
TCATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATCCACTGGAACCTTAATGGTCTTCTTTGGCAATGTGGAT
TCATCTGGGATAAAACAATATTTTTAACCTTCAATTAATGCTCGATACATCCGTTTGCACCCAATCATTATAGCATTGCGAGC
ACTCTTCGATGGAGTTGATGGGCTGTGATTTAAATAGTTGCAGCATGCCATTGGGAATGGAGAGTAAAGCAATATCAGATGCAC
AGATTACTGCTTCATCTACTTTACCAATATGTTGCCACCTGGTCTCTTCAAAAGCTCGACTTCACCTCCAAGGGAGGAGTAATG
CCTGGAGACCTCAGGTGAATAATCCAAAAGAGTGGCTGCAAGTGGACTCCAGAAGACAATGAAAGTCACAGGAGTAATACTC
AGGGAGTAAATCTCTGCTTACCAGCATGTATGTGAAGGAGTTCCTCATCTCCAGCAGTCAAGATGGCCATCAGTGGACTCTTT
TTTCAGAAATGGCAAAGTAAAGGTTTTTCAGGGAAATCAAGACTCTTACACCTGTGGTGAACCTCTAGACCCACCGTTACTGAC
TCGCTACCTTGAATTCACCCAGAGTTGGGTGCCAGATTGCCCTGAGGATGGAGGTTCTGGGCTGCGAGGCACAGGACCTC
TAC

***BDD-FVIII =B结构域缺失的FVIII**

图1I:BDD-FVIII (非优化;“亲本”),核苷酸序列(SEQ ID NO:16)

ATRRYYLGAVELSDWYMQSDLGELPVDARFPVRPKSFPFNTSVVYKTLFVEFTDHLFNIAPRPPWMGLLGPTIQAEVYDVTVITLK
 NMASHPVSLHAVGVSYWKASEGAEYDDQTSQREKEDDKVFPGGSHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGA
 LLVCREGSLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNRSLPGLIGCHRKSYYWHVIGMG
 TPEVHSIFLEGHTFLVRNHRQASLEISPIFLTAQTLMDLGQFLFCHISSHQHDGMEAYVKVDSCPEEPQLRMKNNEEAEDYDDDLT
 DSEMDVVRFDDDNSPSFIQIRSAKKHPKTWVHYIAEEEDWDYAPLVLPAPDDRSYKSQYLNNGPQRIGRKYKVRFMAYTDETFKTR
 EAIQHESGILGPLLYGEVGDLLIIFKNQASRPYNIYPHGIDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRY
 SSFVNMRDLASGLIGPLLYCYKESVDQRGNQIMSDKRNVLFSVFDENRSWYLTENIQRFLPNPAGVQLEDPFQASNIMHSINGYVF
 DSLQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKHKMVYEDTLTLFPFSGETVFMSENPGLWILGCHNSDFRNRGMTALLKVSSC
 DKNTGDYEDSYEDISAYLLSKNNAIEPRSFQNPVLRKHQREITRTTLQSDQEEIDYDDTISVEMKKEDFDIYDEDENQSPRSFQKTR
 HYFIAAVERLWDYGMSSSPHVLNRNAQSGSVPPQFKKVVFEFTDGSFTQPLRYGELNEHLGLLGPYIRAEVEDNIMVTFRNQASRPYSF
 YSSLISYEEDQRQGAEPKRNFKVKNETKTYFWKVQHMHMAPTKDEFDCAWAYFSDVDLEKDVHSGLIGPLLVCHTNTLNPAGHRQVT
 VQEFALFFTIFDETKSWYFTENMERNCRAPCNQMEDPTFKENYRFHAIINGYIMDTLPGLVMAQDQIRWYLLSMGSNENIHSIFSG
 HVFTVRKKEEYKMALYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPLGMASGHIRDFQITASGQYQWA
 PKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIIHGKTQGARQKFSSLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKH
 NIFNPPIAIRYLHPHTHYSIRSTLRMELMGCDLNSCSMPLGMESKAISDAQITASSYFTNMFATWSPSKARLHLQGRSNAWRPQVNNP
 KEWLQVDFQKTMKVTVTTQGVKSLTSMYVKEFLISSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLTRYLRIHPQSW
 VHQIALRMEVLGCEAQDLY

*BDD-FVIII = B结构域缺失的FVIII

图1J:BDD-FVIII (非优化;“亲本”),氨基酸序列 (SEQ ID NO:17)

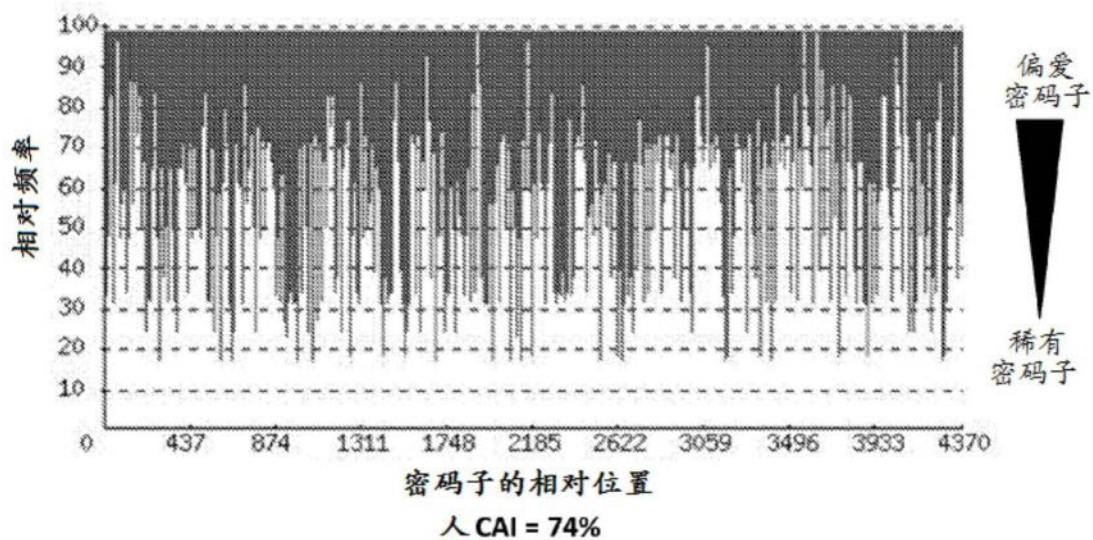
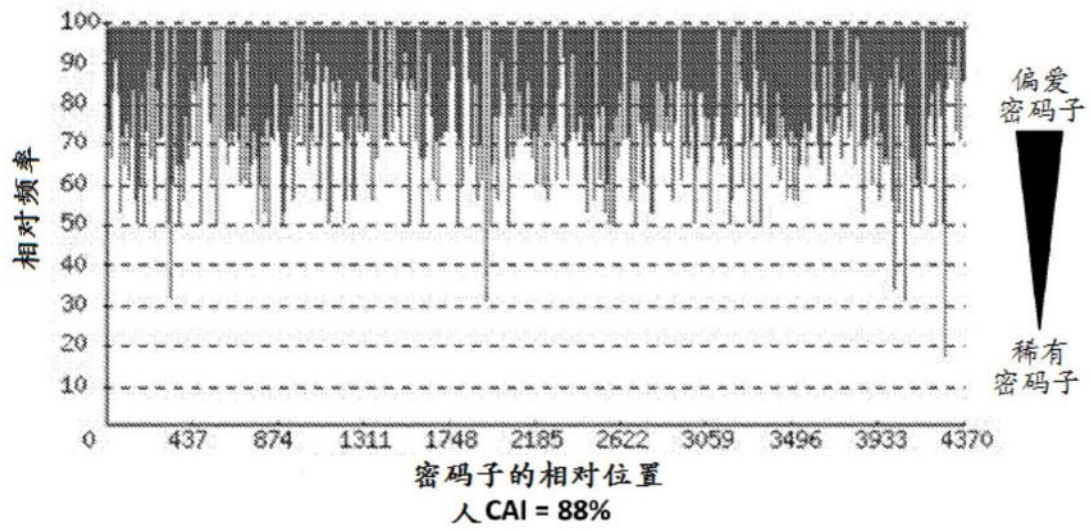


图2A-非优化BDD FVIII



BDD FVIII = B结构域缺失的FVIII

图2B-coFVIII-1

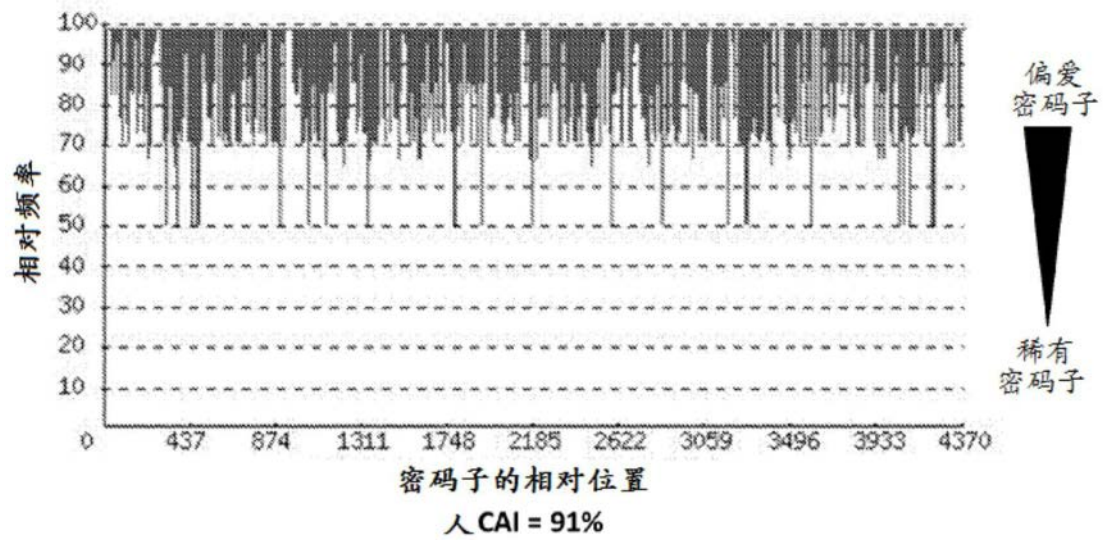


图2C-coFVIII-3

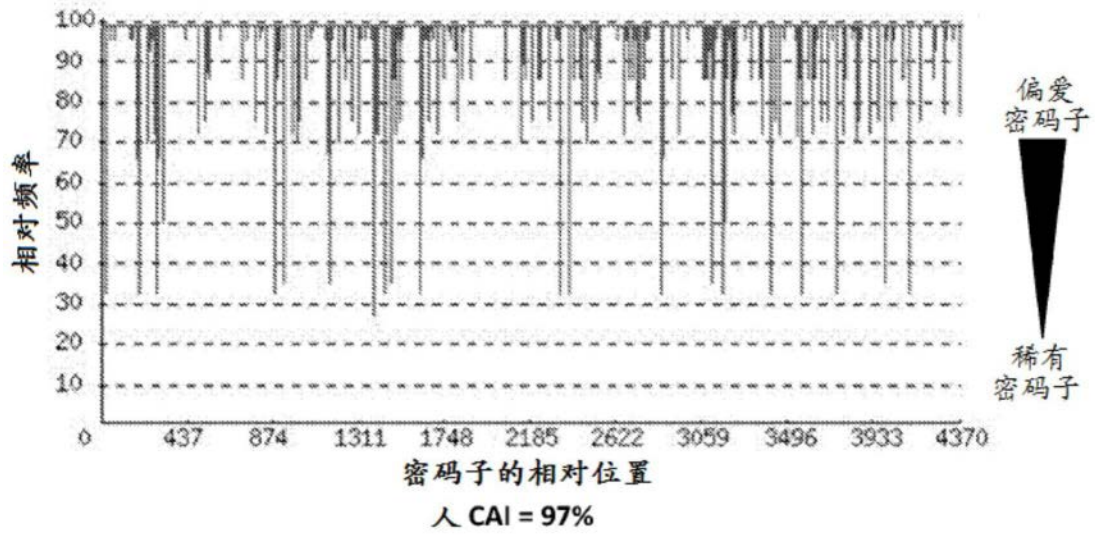


图2D-coFVIII-4

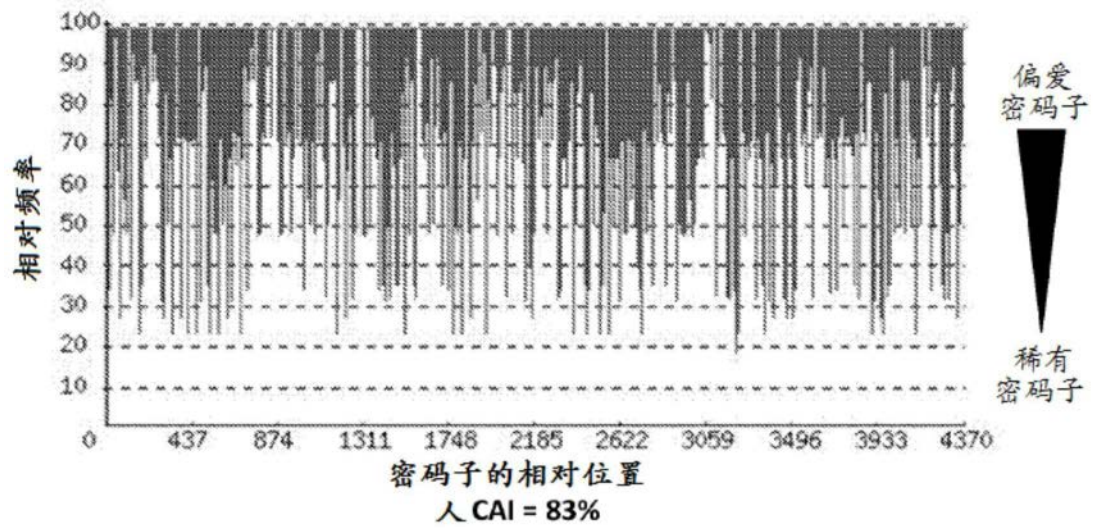


图2E-coFVIII-5

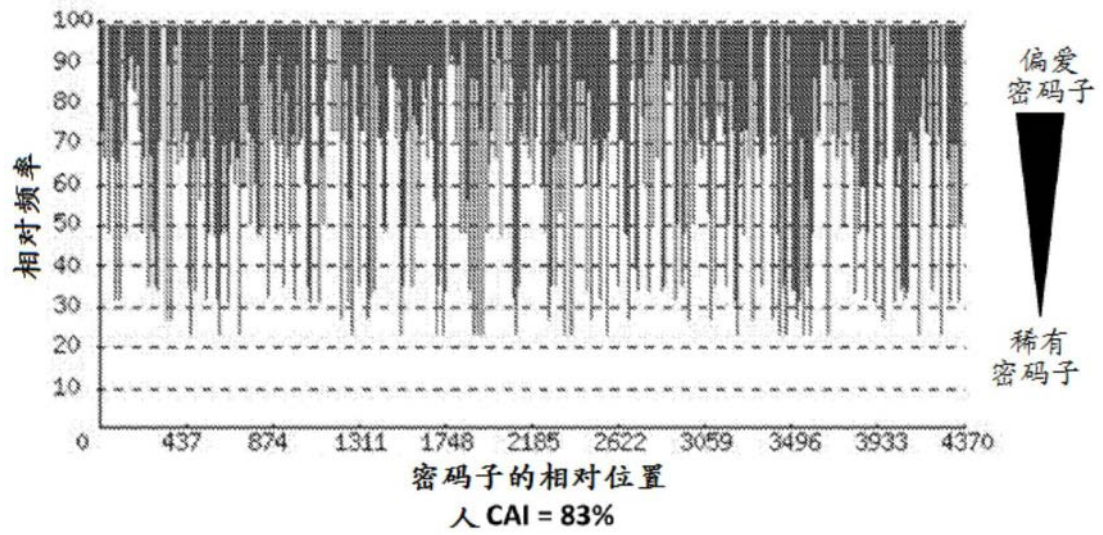


图2F-coFVIII-6

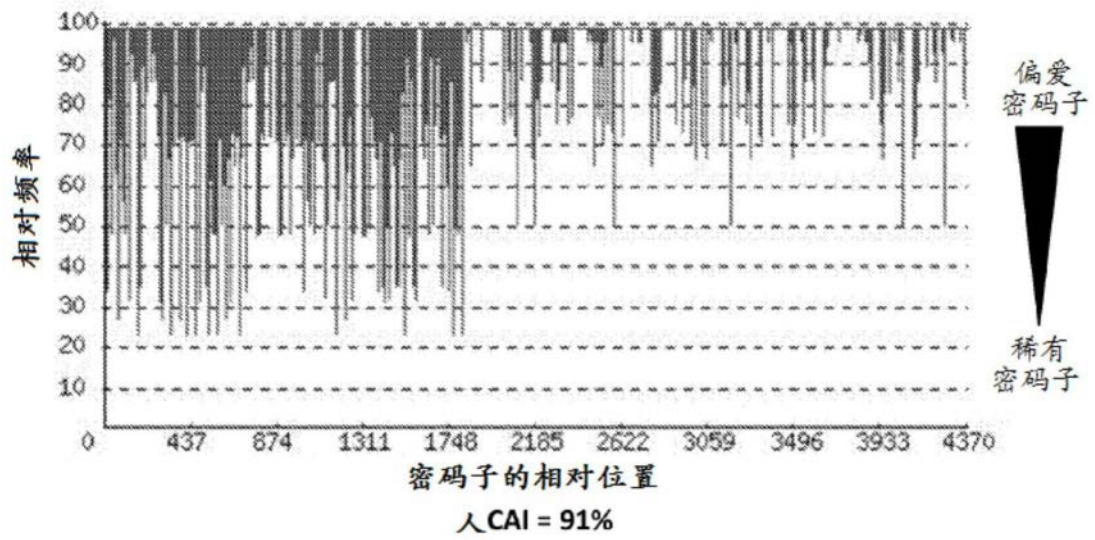


图2G-coFVIII-52

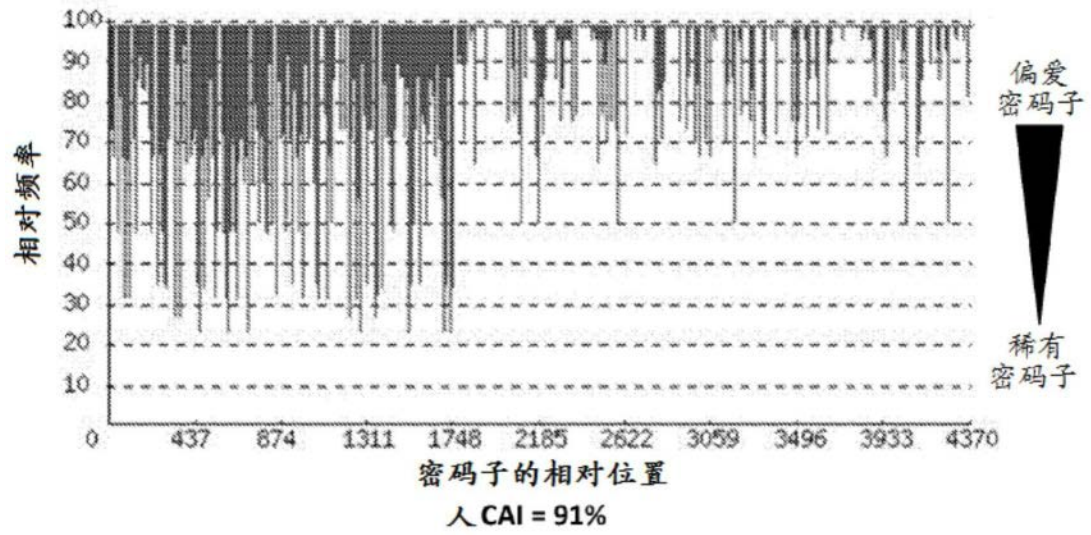


图2H-coFVIII-62

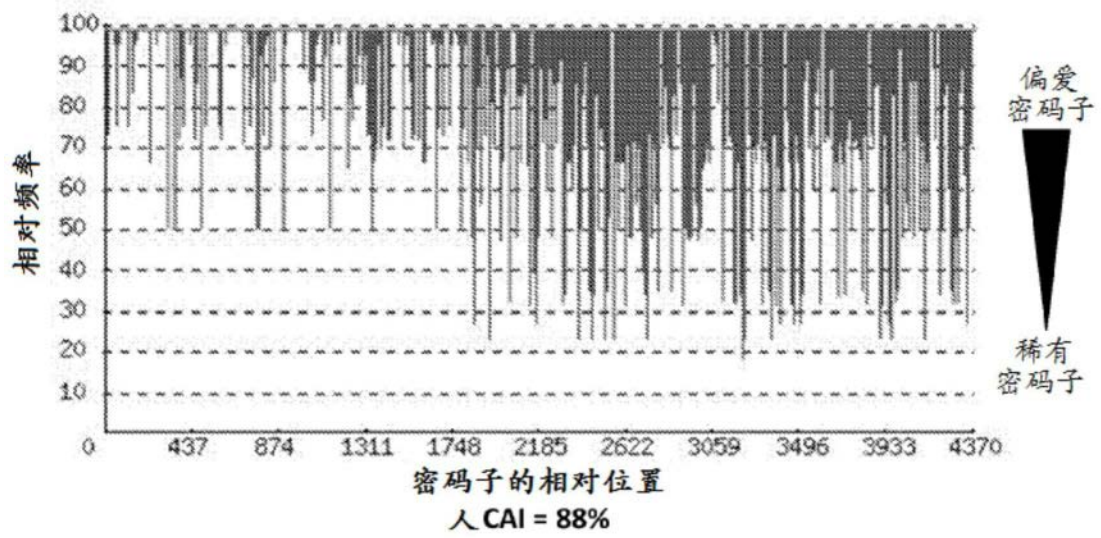


图2I-coFVIII-25

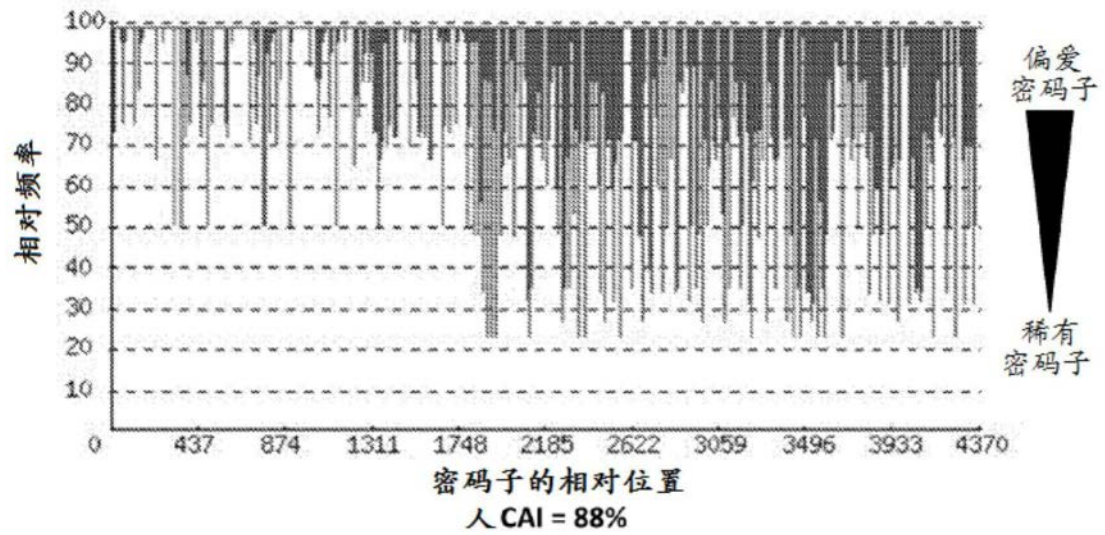


图2J-coFVIII-26

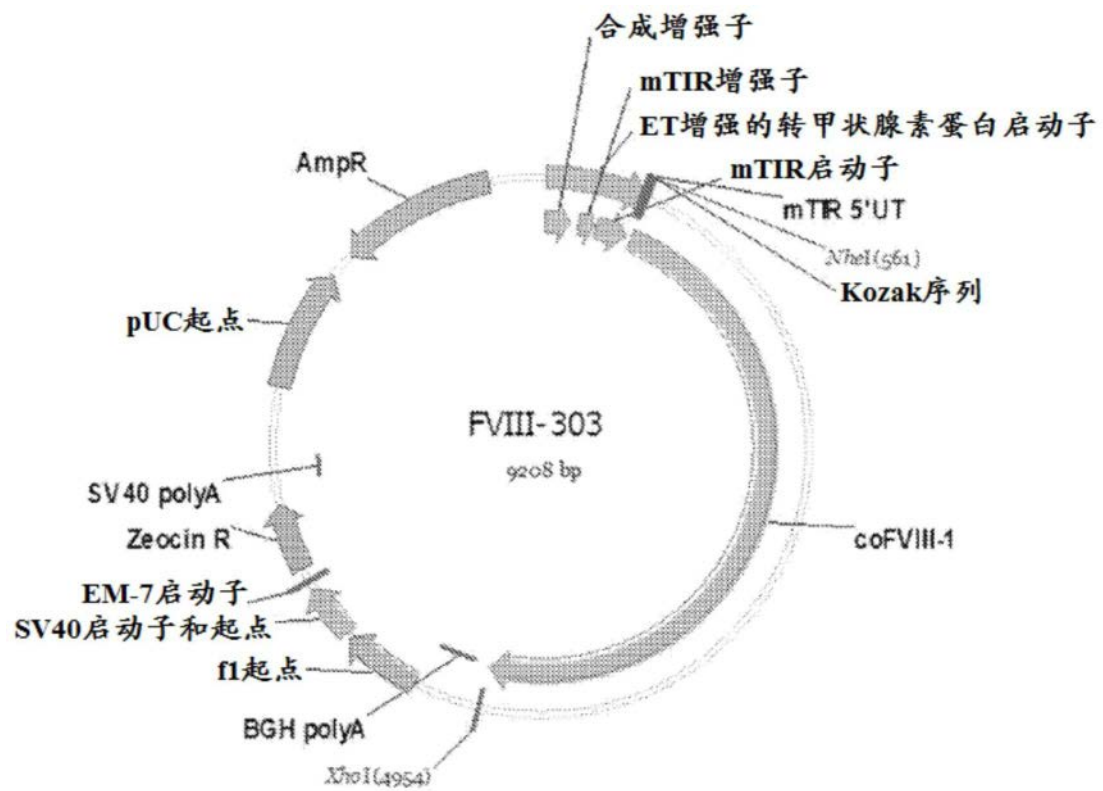


图3:密码子优化的FVIII表达质粒的质粒图

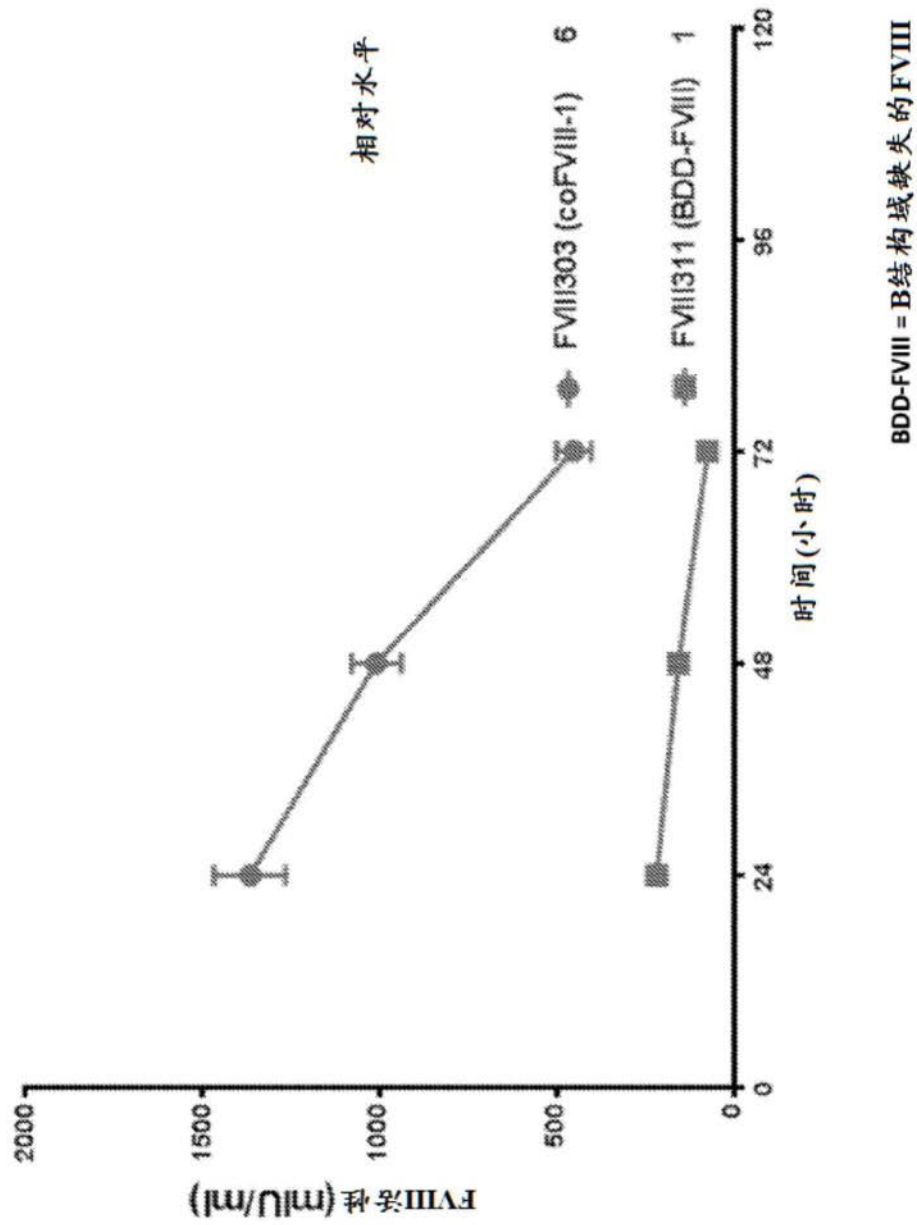


图4:HemA小鼠中的FVIII活性

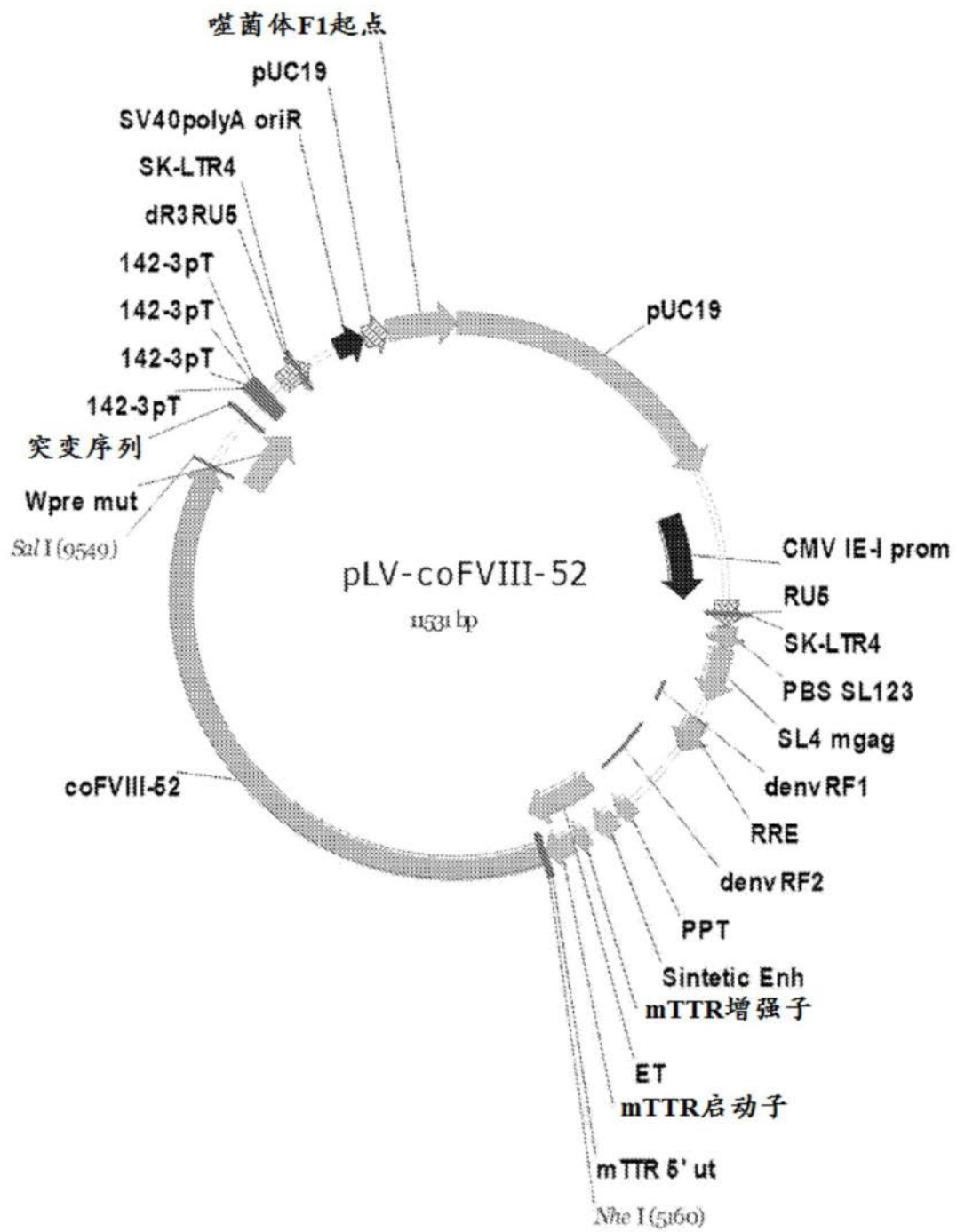


图5:密码子优化的FVIII表达质粒的质粒图

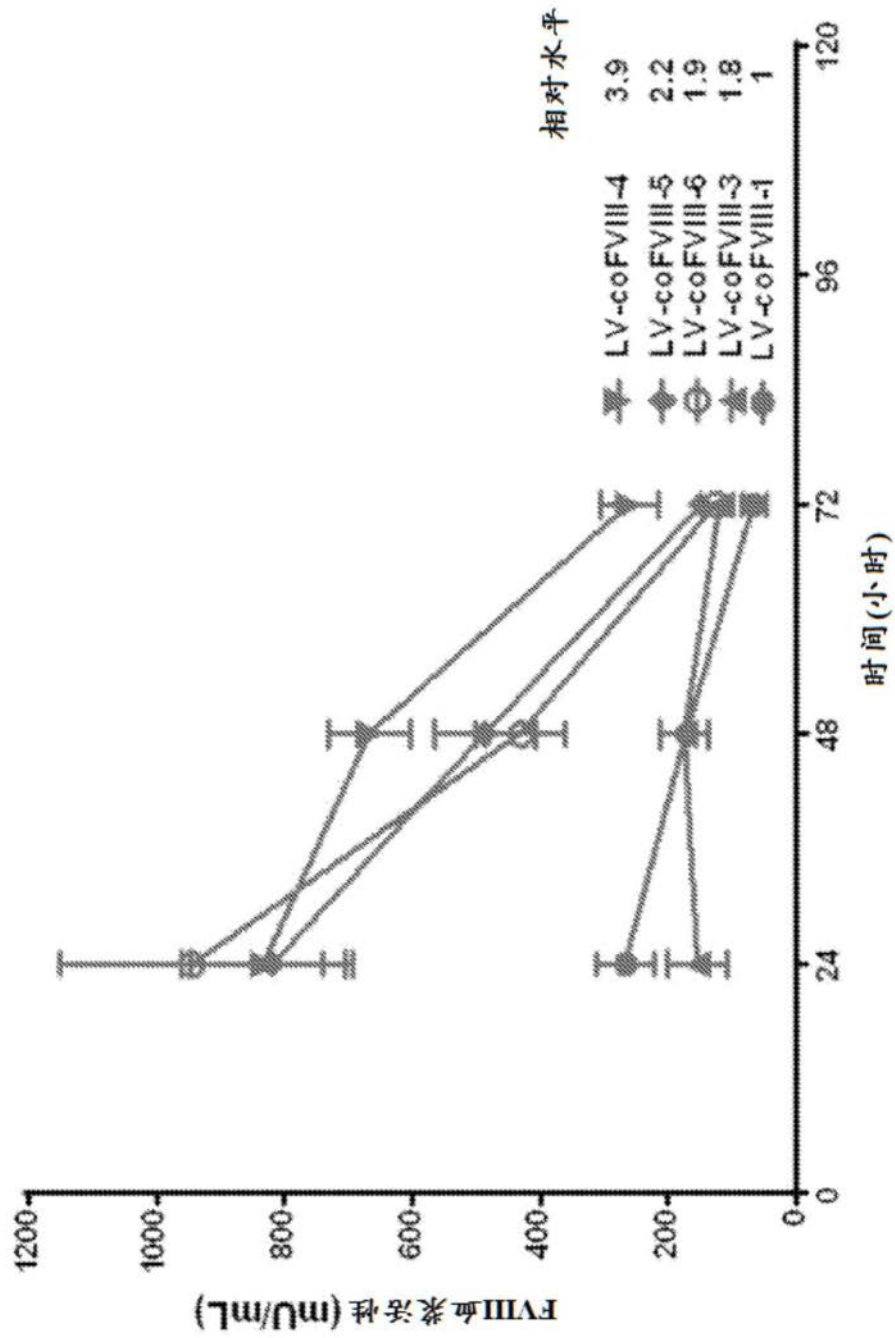


图6A: 相对于coFVIII-1的coFVIII-3、coFVIII-4、coFVIII-5和coFVIII-6

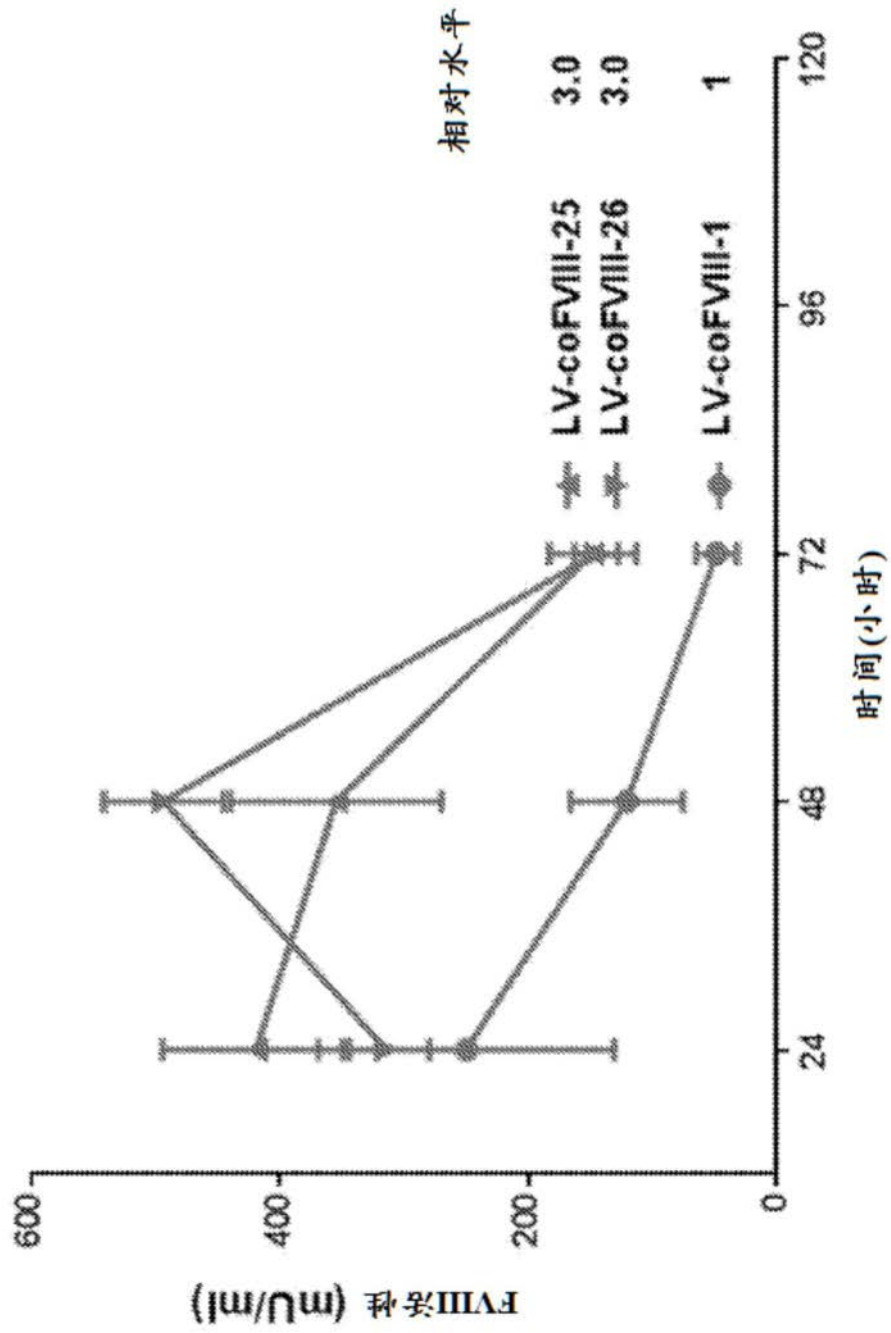


图6B: 相对于coFVIII-1的coFVIII-25和coFVIII-26

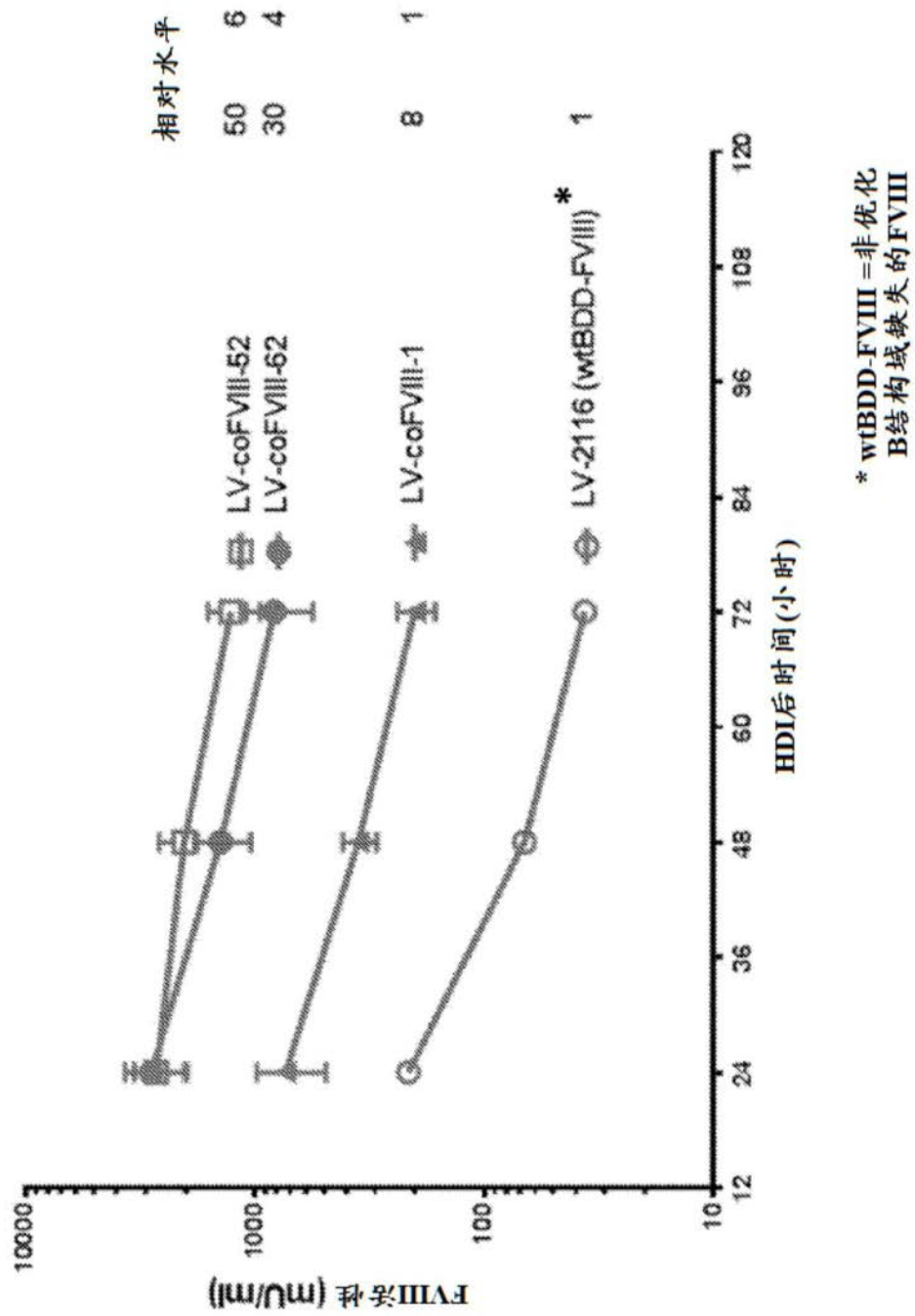


图6C: 相对于BDD-FVIII和coFVIII-1的coFVIII-52和coFVIII-62

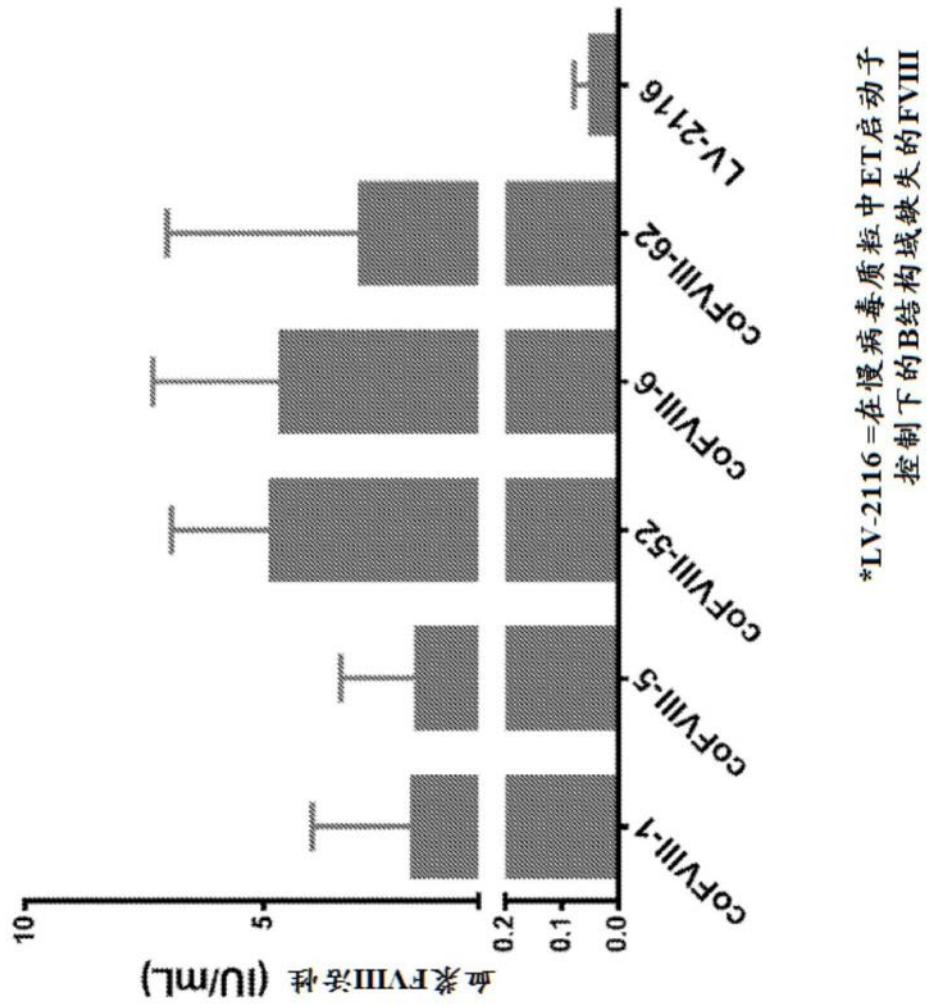


图7:HemA小鼠中的慢病毒载体介导的FVIII活性

ATGCAAACTCGAACTGAGCACCTGTTTCTTCTCTGCCTGCTGAGATTCTGTTTCTCCGCGACCCGCCGATACTACCTGGGAGCAGTGG
AGCTCTCTCTGGGATTACATGCAGAGCGACCTTGGGGAGCTGCCCCTGGATGCCAGGTTCCCTCCCGGGTGCCAAAGTCGTTTCCGTT
CAACACCTCCCGTGGTGTACAGAACTCTGTTCTGTTGAGGTTTACCGACCACTGTTCAATATCGCCAGCCAGACCTCCCTGGATG
GGGCTGTTGGGACCTACCATCCAAGCGGAGGTGTACGACACTGTGGTCATCACTCTGAAGAACATGGCCTCGCATCCCGTGTCCCTGC
ACGCCGTGGGAGTGTCTTACTGGAAAGCGTCCGAGGGGGCCGAATACGACACCGACCGACCTCGCAGAGAGAAAAGGAAGATGACAAAGT
GTTCCAGGAGGATCGCACACCTACGTGTGGCAAGTGTGAAGGAGAACGGCCCAATGGCCTCCGACCCGCTGTGCCTGACCTACTCG
TACCTGTCCACGTGGACCTCGTGAAGGACCTCAACTCGGGACTGATTGGAGCCCTGCTGGTCTGCAAGGAAAGGCTCACTGGCGAAAG
AAAAGACTCAGACCTTGCACAAGTTCATTCTGCTGTTCGCTGTGTTGACAGGGGGAAGTCGTGGCACAGCGAGACTAAGAATCCCT
GATGCAAGATAGAGATGCCGCTCCGCCCGGGCCTGGCCTAAGATGACACCGTGAACGGTTACGTGAACCGCTCCCTCCCTGGCCTG
ATTGGATGCCACCGGAAGTCCGTGTACTGGCACGTGATCGGGATGGGGACCAACCCCGAGGTGCACAGCATCTTCTGGAAGGTGACA
CATTTCTCGTGCAGAACCCACCGGACGGCCTCCCTGGAAATCAGCCCCATTACCTTCTCTACTGCCAGACTCTGCTGATGGACCTGGG
ACAGTTCTCTGCTGTTCTGCCATATCTCCTCCCAACACATGACGGAATGGAGGCATACGTGAAGGTCGATTCTTCCCTGAGGAAACCC
CAGCTCCGCATGAAGAACAATGAGGAAGCCGAGGACTACGACGACGACCTGACGGATAGCGAGATGGATGTGTCGGTTCCGATGACG
ATAACAGCCCTTCTCTCATCCAAATTCGCTCGGTGGCAAGGAGTCAACACCTGGGTGCATTACATTGCGGGGAGAAAGAGGA
CTGGGATTATGCCCGCTTGTCTCGCTCTGACGACCGGAGCTACAAGAGCCAGTACCTGAACAACGGTCCACAGAGGATCGGTAGA
AAGTACAGAAGGTCCGCTTCATGGCCTATACCGACGAAACCTTCAAACTAGAGAGGCCATCCAACAGGAATCCGGCATCTGTGGCC
CGCTCTTGTACGGAAGTCCGGGACACCTTCTCATTATCTTCAAGAACCAGGCTTCCGGCCGTACAACATCTATCCGCATGGGAT
CACTGACGTTGCCCGACTGTACTCGCGGCGCCTGCCCAAGGTTGTCAAACACCTGAAGGATTTCCGATCTCTCCGGGAGAAATCTTC
AAGTACAAGTGGACCGTGACCGTGAAGATGGCCCAACTAAGTCTGACCCTAGATGCCTCACCCTACTACTCATCTTCTGTCAACA
TGGAGCCGACCTGGCCAGCGGACTGATCGGCCCGCTGCTGATTGCTACAGGAATCAGTGGACCAACGGGGAAACAGATCATGTC
GGATAAGAGGAACGTCTCTCTCTCGTGTGTTGACGAAACCGGTCGTGGTACCTGACCGAGAATCCAGAGGTTCTTCCCAAC
CCTGTCGGGTTCAGCTGAGGACCCGAGTTCAGGCCAGGACATCAGTCTGCTGAGCAAGAACAATGCCATCGAGCCAGGAGCTTCTCTCA
TGAGCGTGTGCTGCACGAGGTGGCTACTGGTACATCTGAGCATCGGCGCCAGACCGACTTCTGAGCGTGTCTTCTCTGGCTA
CACCTTCAAGCACAAGATGGTGTATGAGGACACCTGACCCCTGTTCCTTTCAGCGGGGAGACTGTCTTCATGAGCATGGAGAACCCT
GGCCTGTGGATCTTGGGCTGCCACAACAGCGACTTCAGGAACAGGGGCATGACTGCCCTGCTGAAAGTCTCCAGCTGTGACAAGAACA
CCGGGGACTACTACGAGACAGCTACGAGGACATCAGCGCCTACCTGCTGAGCAAGAACAATGCCATCGAGCCAGGAGCTTCTCTCA
GAACGGCGCGCCCAACATCAGAGAGCGCCACCCCTGAAAGTGGTCCCGGGAGCGAGCCAGCCACATCTGGGTCCGGAACCCAGGCA
AGTGAGTCTGCAACTCCCGAGTCCGGACCTGGCTCCGAGCCTGCCACTAGCGGCTCCGAGACTCCGGGAACCTCCGAGAGCGCTACAC
CAGAAAGCGGACCCGGAACCACTAGCGAAGCTAGCGAGGGCTCTGCTCCGGGACGCCAGCCGGCTCTCTACATCCACGGAGGAGGG
CACTTCCGAATCCCGCACCCCGAGTTCAGGGCCAGGATCTGTAACCCCTACTCTCAGGCACTGAGACGCCAGGACGAGCTCCGCT
ACACCGGAGTGGGGCCAGGAGCCCTGCTGGATCTCTACGTCCACTGAGGAAGGGTCACCAAGCGGCTCGCCACCAAGCACTGAAG
AAGGTGCTCGAGCCCCCAGTGTGTAAGAGGACCAAGAGGAGATCACCAGGACCACTGCGAGTCTGACAGGAGGAGATCGACTA
TGATGACACCATCAGCGTGGAGATGAAGAAGGAGGACTTCAGCATCTACGACGAGGACGAGAACCAGAGCCCCAGGAGCTTCCAGAG
AAGACAGGCACTACTTCATTGCTGCTGTGGAGGGCTGTGGGACTATGGCATGTCCAGCAGCCCCATGTGCTGAGGAACAGGGCCC
AGTCTGGCCACGTGCCCACTTCAAGAAAGTCTGTTCAGGAGTTCACGACGGCAGCTTCAACCAAGCCCTGTACAGAGGGGAGCT
GAACGAGCACCTGGGCTGCTGGGCCCTTACATCAGGGCCGAGGTGGAGGACAACATCATGGTGACCTTCAGGAACAGGCCAGCAGG
CCCTACAGCTTCTACAGCAGCTGATCAGCTACGAGGAGGACAGAGGCGAGGGGCTGAGCCAGGAAGAATTTGTGAAGCCCAATG
AAACCAAGACCTACTTCTGGAAGGTGCAGCACCATGGCCCCACCAAGGACGAGTTCGACTGCAAGGCCTGGGCTACTTCTCTGA
CGTGGACCTGGAGAAGGACGTGCACTCTGGCCTGATTGGCCCCCTGCTGCTGTGTCACACCAACACCTGAAACCTGCCATGGCAGG
CAGGTGACTGTGCAGGAGTTCGCCCTGTTCTTCAACCATCTTCGATGAAACCAAGAGCTGGTACTTCACTGAGAACATGGAGGAACT
GCAGGGCCCCCTGCAACATCCAGATGGAGGACCCACCTTCAAGGAGAACTACAGGTTCATGCCATCAATGGCTACATCATGGACAC
CCTGCTGGCCTGGTCTGATGGCCAGGACAGAGGATCAGGTGGTATCTGCTGAGCATGGGCGCAACGAGAACATCCACAGCATCCAC
TTCTCTGGCCACGTGTTCACTGTGAGGAAGAAGGAGGATGACAGATGGCCCTGTACAACCTGTACCTGGGGTGTTCGAACCGTGG
AGATGCTGCCAGCAAGGCCGGCATCTGGAGGGTGGAGTGCCTGATTGGGGAGCACCTGCACGCCGGCATGAGCACCTGTTCTCTGGT
GTACAGCAACAAGTGCCAGACCCCTGGGCATGGCCTCTGGCCACATCAGGACTTCCAGATCACTGCTCTGGCCAGTACGGCCAG
TGGGCCCCCAAGCTGGCCAGGCTGCACTACTCCGGAAGCATCAATGCTGGAGCACAAGGAGCCCTTCAGCTGGATCAAAGTGGACC
TGCTGGCCCCATGATCATCCAGGCATCAAGACCCAGGGGGCCAGGCAAGTTCCTCCAGCTGTACATCAGCCAGTTTATCATCAT
GTACAGCTGGACGGCAAGAAGTGGCAGACCTACAGGGGCAACAGCACCCGACCCCTGATGGTGTCTTCGGCAACGTGGACAGCAGC
GGCATCAAGCACAACATCTTCAACCCCCCATCATCGCCAGATACATCAGGCTGCACCCACCCACTACAGCATCAGGAGCACCTGA
GGATGGAGCTGATGGCTGTGACCTGAACAGCTGCAGCATGCCCTGGGCATGGAGAGCAAGGCCATCTTGACGCCAGATCACTGC
CTCCAGTACTTCAACAACATGTTTGCACCTGGAGCCCCAGCAAGGCCAGGCTGCACCTGCAGGGCAGGAGCAATGCCTGGAGGGCC
CAGGTCAACAACCCCAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATGAAGGTGACTGGGGTGACCAACCCAGGGGGTGAAGAGCC
TGCTGACAGCATGTACGTGAAGGAGTTCCTGATCTCCAGCAGCCAGGACGGCCACCAGTGGACCTGTTCTTCCAGAATGGCAAGGT
GAAGGTGTTCCAGGGCAACAGGACAGCTTACCCCTGTGGTCAACAGCCTGGACCCCCCTGCTGACAGATACCTGAGGATCCAC
CCCCAGAGCTGGGTGCACAGATCGCCCTGAGGATGGAGGTGCTGGGCTGTGAGGCCAGGACCTGTACTGA

加下划线= XTEN核苷酸序列 (SEQ ID NO: 18)

图8A:coFVIII-52-XTEN-SEQ ID NO:19

ATGCAGATTGAGCTGTCTACTTGCTTTTTCTGTGCCTGCTGAGGTTTTGCTTTCCGCTACACGAAGTATTATCTGGGGCTGTGG
AACTGTCTTGGGATTACATGCAGAGTGACCTGGGAGAGCTGCCAGTGGACGCAAGGTTTCCCCCTAGAGTCCCTAAGTCATTCCCCCTT
CAACACTAGCGTGGGTACAGAGAAACACTGTTCTGTGGAGTTTACTGATCACCTGTTCAACATCGCAAGCCCTAGGCCACCGTGGATG
GGACTGCTGGGGCCAAACATCCAGGCCGAGGTGTACGACACCGTGGTCATTACACTTAAGAACATGGCCCTCACACCCCGTGAGCCCTGC
ATGCTTGGGGCTCAGCTACTGGGAAGGCTTCCGAAGGAGCAGAGTATGACGATCAGACTTCCAGAGAGAAAAGAGGACGATAAGGT
GTTTCTGTGGCGGATCTCATACCTACGTGTGGCAGGTCTGAAAGAGAATGGCCCTATGGCCCTCCGACCCCTCTGTGCCCTGACCTACTCT
TATCTGAGTCAGTGGACCTGGTCAAGGATCTGAACAGCGGCCCTGATCGGAGCCCTGCTGGTGTGCAGGGAAGGAAGCCTGGCTAAGG
AGAAAACCCAGACACTGCATAAGTTTCTGCTGTTGCGCGTGTGACGAAGGGAAATCATGGCACAGCGAGACAAAGAATAGTCT
GATGCAGGACAGGGATGCCGCTTCAGCCAGAGCTTGGCCCAAATGCACACTGTGAACGGCTACGTCAATCGCTCACTGCCCTGGGCTG
ATCGGCTGCCACCGAAAGAGCGTGTATTGGCATGTCTCGGGATGGGCACACACCTGAAGTGCCTCCATTTTCTGGAGGGACATA
CCTTTCTGGTCCGCAACCACCGACAGGCTTCCCTGGAGATCTCTCCAATTACCTTCTGACAGCACAGACTCTGCTGATGGACCTGGG
GCAGTTCTGCTGTTTTGCCACATCAGCTCCCACCGAGATGATGGCATGGAGGCTTACGTGAAAGTGGACTCTGTCCCGAGGAACCT
CAGCTCGGGATGAAGAACAATGAGGAAGCAGAACTATGACGATGACCTGACCGACTCCGAGATGGATGTGGTCCGATTCGATGACG
ATAACAGCCCCCTCCTTTATCCAGATTAGATCTGTGGCCAAAGAAACACCTTAAGACATGGGTCCATTACATCCGAGCCGAGGAAGAGA
CTGGGATTATGCACCACTGGTGTGGCACCAGACGATCGCTCTACAATCTCAGTATCTGAACAATGGGCCACAGAGGATTGGCAGA
AAGTACAGAAAGTGGGTTTATGGCATATACCGATGAGACCTTCAAGACTCGCGAAGCCATCCAGCAGGAGAGCGGCATCTGGGAC
CACTGCTGTACGGAGAAGTGGGAGACACCTGCTGATCATTTTCAAGAACCAGGCCAGCCGGCCTTACAATATCTATCCACATGGGAT
TACAGATGTGCCCTCTGTACAGCAGGAGACTGCCAAAGGGCTCAAAACACCTGAAGGACTTCCCAATCTCGAGCCGAGGAATCTTC
AAGTACAAGTGGACTGTCAACGTCGAGGATGGCCCACTAAGAGCGACCCCTCGGTGCCTGACCCGCTACTATTCTAGTTTCGTGAATA
TGGAAAGAGATCTGGCAAGCGGACTGATCGGACCACTGCTGATTTGTACAAAGAGAGCGTGGATCAGAGAGGCAACCAGATCATGTC
CGACAAGCGGAATGTGATTCTGTTCAGTGTCTTTGACGAAACAGGTCATGGTACCTGACCGAGAATCCAGAGATTCTGCCTAAT
CCAGCTGGGCTGCCCTCTGTACAGCAGGAGACTTCCAGGATCTAACATCATGATGATTAATGGCTAGCTGTCCGAGAGAAATCTTC
TGAGCGTGTGCCCTGCACGAGGTGCGTTACTGGTATATCTGAGCATTGGGGCACAGACAGATTTCCTGAGCGTGTCTTTTCCGGCTA
CAGTTTAAAGCATAAAATGGTCTATGAGGACACACTGACTCTGTTCCCTTCAGCGGCGAAACCGTGTATGAGCATGGAGAATCCC
GGACTGTGGATTCTGGGGTGCCAACAACAGCGATTTCAGAAATCGCGGAATGACTGCCCTGCTGAAAGTGTCAAGCTGTGACAAGACA
CCGGGGTACTATGAAGATTATACGAGGACATCAGCGCATCTAAGCATGCTGCTGCCAAACAAATGCCATTGAACCCGGTCTTTTAGTCA
GAATGGCGCGCAACATCAGAGAGCGCCACCCCTGAAAGTGTTCGGGGAGCGAGCCAGCCACATCTGGGTCCGAAACGCCAGGCACA
AGTGAGTCTGCAACTCCCGAGTCCGGACCTGGCTCCGAGCCTGCCACTAGCGGCTCCGAGACTCCGGGAATCCGAGAGCGCTACAC
CAGAAAGCGGACCCGGAACAGTACCGAACCTAGCGAGGGCTCTGCTCCGGGCGAGCCAGCCGGCTCTCTACATCCACGGAGGAGGG
CACTTCCGAATCCGCCACCCCGGAGTCAGGGCCAGGATCTGAACCCGCTACCTCAGGCAGTGAGACGCCAGGAACGAGCGAGTCCGGT
ACACCGGAGAGTGGGCGAGGAGCCCTGCTGGATCTCCTACCTGCTGAGGAGGGTCAACGCGGGCTCGCCACCGAGCACTGAAG
AAGGTGCCTCGAGCCCTCCAGTGTGAAGCGGCACGCGGAGATCACCCGCACTACCTGCAGAGTGATCAGGAAGAGATCGACTA
CGACGATACAAATTTCTGTGGAATGAAGAAAGAGGACTTCGATATCTATGACGAAGATGAGAACCAGAGTCTCTGATCATTCCAGAA
AAAACAGGCATTACTTTATGCGCAGTGGAGCGGCTGTGGGATTATGGCATGTCTCTAGTCTCAGTGTGCGAAATAGGGCCC
AGTCAGGAAGCTCCCAAGTTCAAGAAAGTGGTCTTCCAGGAGTTTACAGACGGGTCTTTACTCAGCCAGTGTACAGGGCGAACT
GAACGAGCACCTGGGACTGCTGGGCGCTATATCAGAGCAGAAGTGGAGGATAACATTATGGTCACCTTCAGAAATCAGGCCTCTCGG
CCTTACAGTTTTTATTCAAGCCTGATCTCTTACGAAGAGGACAGCGACAGGGAGCTGAACCAAGAAACCTTCGTGAAGCCTAATG
AGACAAAACATACTTTTGAAGGTGCAGCACCATATGGCCCAACAAAGACGAGTTCGATTGCAAGGCATGGGCCATTTTTCTGA
CGTGATCTGGAGAAGGACGTGCACAGTGGCTGATTTGGCCCACTGCTGGTGTGCCATACTAACACCTGAATCCAGGCCACGGCCGG
CAGGTCACTGTCCAGGAGTTCGCTCTGTTCTTTACCATCTTTGATGAGACAAAGAGCTGGTACTTCACCGAAAACATGGAGCGAAAT
GCAGGGCTCCATGTAACTTCAGATGGAAGACCCACATTCAAGGAGAACTACCGCTTTCATGCTATCAATGGATACATCATGGATAC
TCTGCCCGGGCTGGTTCATGGCACAGGACAGAGAATCCGGTGGTATCTGCTGAGCATGGGCAGCAACGAGAATATCCACTCAATTCAT
TTCAGCGGGCAGCTGTTTACTGTGAGGAAGAAAGAGTACAAGATGGCCCTGTACAACCTGTATCCCGCGTGTTCGAAACCGTCG
AGATGCTGCTAGCAAGGCCGGAATCTGGAGAGTGGAAATGCCCTGATTTGGAGAGCAGCTGCATGCTGGGATGTACCCCTGTTTCTGGT
GTACAGTAATAAGTGTGACACACCCCTGGGAATGGCATCCGGGCATATCAGGATTTCCAGATTACCGCATCTGGACAGTACGGACAG
TGGGCACCTAAGCTGGCTAGACTGCACTATTCCGGATCTATCAACGCTTGGTCCACAAAAGAGCCTTTCTCTTGGATTAAAGTGGACC
TGCTGGCCCAATGATCATTCATGGCATCAAACTCAGGGAGCTCGGCAGAGTTCTCTCTCTGTACATCTCACAGTTTATCATCAT
GTACAGCCTGGATGGGAAGAAATGGCAGACATACCGCGCAATAGCACAGGAATCTGATGGTGTCTTTGGCAACGTGGACAGCAGC
GGAATCAAGCACAACATTTTCAATCCCCCTATCATTGCTAGATACATCCGGCTGCACCAACCCATTATTCTATTGAAAGTACACTGA
GGATGGAATGATGGGATGCGATCTGAACAGTTGTTCAATGCCCTGGGGATGGAGTCCAAGGCAATCTGTACGCCCAGATTACCGC
CAGCTCCTACTTCACTAATATGTTTGCTACCTGGAGCCCTTCCAAAGCAAGACTGCACCTGCAAGGCCGAGCAACGATGGCGACCA
CAGGTGAACAATCCCAAGGAGTGGTTGCAGGTGATTTTCAAGAAACTATGAAGGTGACCGGGGTCACTCAGGGCGTGAAGATGTC
TGCTGACCTCAATGTACGTCAAGGAGTTCCTGATCTCTAGTTTACAGGACGACATCAGTGGACACTGTTCTTTCAGAACGGGAAGGT
GAAAGTCTTCCAGGGCAATCAGGATTCTTTACACCTGTGGTCAACAGTCTAGACCTCCACTGCTGACAGATACCTGAGAATCCAC
CCTCAGTCTGGGTGCACAGATTGCCCTGAGAATGGAAGTGTGGGATGCGAGGCCAGGATCTGTACTGA

加下划线= XTEN核苷酸序列 (SEQ ID NO: 18)

图8B:coFVIII-1-XTEN-SEQ ID NO:20

ATGCAGATTGAGCTGTCCACTTGTTCCTTCTGTGCCTCCTGCGCTTCTGTTTCTCCGCCACTCGCCGGTACTACCTTGGAGCCGTGG
 AGCTTTTACATGGGACTACATGCAGAGCGACCTGGGCGAACTCCCGTGGATGCCAGATTCCCCCCCCGCTGCCAAAGTCTTCCCTT
 TAACACCTCCGTGGGTACAGAAACCCCTCTTGTGCGAGTTCACTGACCACCTGTTCAACATCGCCAGCCGCGCCACCTTGGATG
 GGCCTCCTGGGACCGACCATCAAGCTGAAGTGTACGACACCGTGGTGATCACCCTGAAGAACATGGCGTCCCACCCCGTGTCCCTGC
 ATGCGGTGCGAGTGTCTACTGGAAGGCCCTCCGAAGGAGCTGAGTACGACGACGACTAGCCAGCGGGGAAAGGAGGACGATAAAGT
 GTTCCCGGGCGGCTCGCATACTTACGTGTGGCAAGTCTGAAGGAAAACGGACCTATGGCATCCGATCCTGTGCGCTGACTTACTCC
 TACCTTTCCCATGTGGACCTCGTGAAGGACCTGAACAGCGGGCTGATGGTGCACTTCTCGTGTGCCGCGAAGGTTCGCTCGCTAAGG
 AAAAGACCCAGACCTCCATAAGTTCATCTTTTGTTCGCTGTGTTTCGATGAAGGAAAGTATGGCATTCGAAAACTAAGAACTCGCT
 GATGCGAGGACCGGATGCCGCTCAGCCGCGCCTGGCCTAAATGCATACAGTCAACGGATACGTGAATCGGTCACTGCCCGGGCTC
 ATCGGTTGTACAGAAAGTCCGTGTAAGTGGCAGCTATCGGCATGGGCACTACGCTGAAGTGCATCCATCTTCTGGAAAGGGCACA
 CTTCTCTCGTGGCAACACCGCCAGGCTCTCTGGAATCTCCCGATTACCTTCTGACCGCCAGACTCTGCTCATGGACCTGGG
 GCAGTTCCTTCTCTGCGACATCTCCAGCCATCAGCAGCGGAATGGAGGCTACGTGAAGTGGACTCATGCCCGGAAGAACCT
 CAGTTGCGGATGAAGAACACGAGGAGGCCGAGGACTATGACGACGATTTGACTGACTCCGAGATGGACGTCGTGCGGTTCGATGACG
 ACAACAGCCCGAGCTTCATCCAGATTCCGAGCCTGGCGAGGCTGGCCAAAGAGCACCCGAAACCTGGGTGCACTACATCGCGGCGGAGGAAGA
 TTGGGACTACGCCCGTGGTGCTGGCACCCGATGACCGGTCTGACAAGTCCAGTATCTGAACAATGGTCCGCGAGCGGATTGGCAGA
 AAGTACAGAAAGTGGGTTTCATGGCGTACACTGACGAAACGTTTAAAGACCCGGGAGGCCATTCAACATGAGAGCGGCATTCTGGGAC
 CACTGCTGTACGGAGAGGTGGCGATACCTGCTCATCTTCAAAAACAGGCCCTCCCGGCTTACAACATCTACCTCAGCGAAT
 CACCGAGGTGGCCACTTACTCGCGCGCCTTCCGGAAGGCGCTCAAGCACCTGAAAGACTTCCCTATCTGCGGCGAAATCTTTC
 AAGTATAAGTGGACCGTCAACGTTGGAGGACGGGCCCCAAGAGCGATCCTAGGTGTGACTCGGTACTACTCCAGCTTCGTGAACA
 TGGAAACGGGACCTGGCATCGGACTCATTGGACCGCTGCTGATCTGCTACAAAGAGTCCGTGGATCAACGCGGCAACAGATCATGTC
 CGACAAGCGCAACGTGATCCTGTTCTCCGTGTTTGTGAAAACAGATCCTGGTACCTCACTGAAAACATCCAGAGGTTCTCCCAAAC
 GAGCGGCGCGCAACATCAGAGAGCGCCACCTGAAAGTGGTCCGCGGAGCGAGCCAGCCACATCTGGGTCCGAAACCGCCAGGCCA
 AGTGAGTCTGCAACTCCCGAGTCCGGACCTGGCTCCGAGCCTGCCACTAGCGGCTCCGAGACTCCGGGAATCCGAGAGCGCTACAC
 CAGAAAGCGGACCCGGAACAGTACCGAACCTAGCGAGGCTCTGCTCCGGGCGAGCCAGCCGGCTCTCTACATCCAGGAGGAGGG
 CACTTCCGAATCCGCCACCCCGAGTCAAGGCGAGGATCTGAACCCGCTACCTCAGGCAGTGAGACGCCAGGAACGAGCGAGTCCGCT
 ACACCGGAGAGTGGGCCAGGAGCCCTGCTGGATCTCTTACGTTCCAGTGGAGGAGGTCAACGCGGGCTCCGCCACCGACTGAAG
 AAGGTGCTCGAGCCCGCTGTGCTGAAGAGGACCCAGCGAGAAATTACCCGACCCCTCCAAATCGGATCAGGAGGAAATCGACTA
 CGACGACACCATCTCGGTGGAAATGAAGAAGGAAGATTTCATATCTACGACGAGGACGAAAATCAGTCCCTCGCTATTCCAAAG
 AAAACTAGACACTACTTTATCGCCGCGGTGGAAAGACTGTGGGACTATGGAATGTATCCAGCCCTCAGTCTTTCGGAACCGGGCCC
 AGACGGGATCGGTGCCCTCAGTTCAAGAAAGTGGTGTCCAGGAGTTCACCGACGCGAGTTCAACCGAGCCGCTGATCCGGGAGAACT
 GAACGAACACCTGGGCTGCTCGGTCCCTACATCCGCGCGGAAGTGGAGGATAACATCATGGTGACCTTCCGTAACCAAGCATCCAGA
 CTTTACTCTTCTATTCTCCCTGATCTCATACGAGGAGGACGAGCGCAAGGCGCCGAGCCCGCAAGAACTTCGTCAGCCCAAGC
 AGACTAAGACCTACTTCTGGAAGTCCAAACACCATATGGCCCGACCAAGGATGAGTTTGAAGTCAAGGCTGGGCTACTTCTCCGA
 CGTGGACCTTGAGAAGGATGTCCATTCCGGCTGATCGGGCGCTGCTCGTGTGTACACCAACCCCTGAACCCAGCGCATGGACGC
 CAGGTCAACCGTCCAGGAGTTGCTGTTCTTCAACATTTTTGACGAACTAAGTCTGGTACTTCAACGAGAATATGGAGCGAACT
 GTAGAGCGCCCTGCAATATCCAGATGGAAGATCCGACTTTCAAGGAGAACTATAGATTCCAGCCATCAACGGGTACATCATGGATAC
 TCTGCCGGGGCTGGTCATGGCCAGGATCAGAGGATCGGTGGTACTTGTGTCAATGGGATCGAACGAAAACATTCACTCCATTAC
 TTCTCCGGTCACTGTGCTGCGCAAGAGGAGGAGTACAAGATGGCGCTGTACAATCTGTACCCCGGGGTGTTCCAAACTGTGG
 AGATGCTCGCGTCCAGGCGCGCATCTGGAGAGTGGAGTGGCTGATCGGAGAGCACTCCACGCGGGGATGTCCACCTCTTCTGTGT
 GTACTCGAATAAGTGCCAGACCCCGCTGGGCATGGCCTCGGGCCACATCAGAGACTTCAGATCAGCAGCGGACAAATACGGCCAA
 TGGGCGCGGAAGCTGGCCCGCTTGCATCTCCGATCGATCAACGCATGGTCCACCAAGGAACCGTTCTCGTGGATTAAAGTGGACC
 TCCTGGCCCTATGATTATCCACGGAATTAAGACCCAGGGCGCCAGGCGAGGTTCTCTCTCTGTACATCTCGCAATTATCATCAT
 GTACAGCTGGACGGGAGAAAGTGGCAGACTTACAGGGGAACTCCACCGCACCTGATGGTCTTTTTCGGCAACGTGGATTCTCTCC
 GGCATTAAAGCAACAATCTTCAACCCACCGATCATAGCCAGATATATTAGGCTCCACCCCACTCACTACTCAATCCGCTCAACTCTTC
 GGATGGAATCATGGGTGCGACCTGAATCTCTGCTCATGCCGTTGGGGATGGAATCAAAGGCTATTAGCGAGCCGAGATCACCGC
 GAGCTCCTACTTCACTAATCATGTTCCGCACTGGAGCCCTCAAGGCCAGGCTGCACTTGCAGGGACGGTCAATGCCTGGCGGCGG
 CAAGTGAACAATCCGAAGGAATGGCTTCAAGTGGATTTCAAAAGACCATGAAAGTGACCGGAGTCAACCCAGGGAGTGAAGTCCC
 TTCTGACCTCGATGTATGTGAAGGAGTCTCTAGTAGCAGCCAGGACGGGACCAAGTGGACCTGTTCTTCCAAACCGGAAAGGT
 CAAGGTGTTCAGGGGAACAGGACTCGTTCAACCCGTTGGTGAATCCCTGGACCCCACTGCTGACGCGGTACTTGGAGTTTCAT
 CCTCAGTCTGGGTCCATCAGATTGCAATTGCGAATGGAAGTCTGGGCTGGGAGGCCAGGACCTGTACTGA

加下划线= XTEEN核苷酸序列 (SEQ ID NO: 18)

图8C:coFVIII-6-XTEEN-SEQ ID NO:72

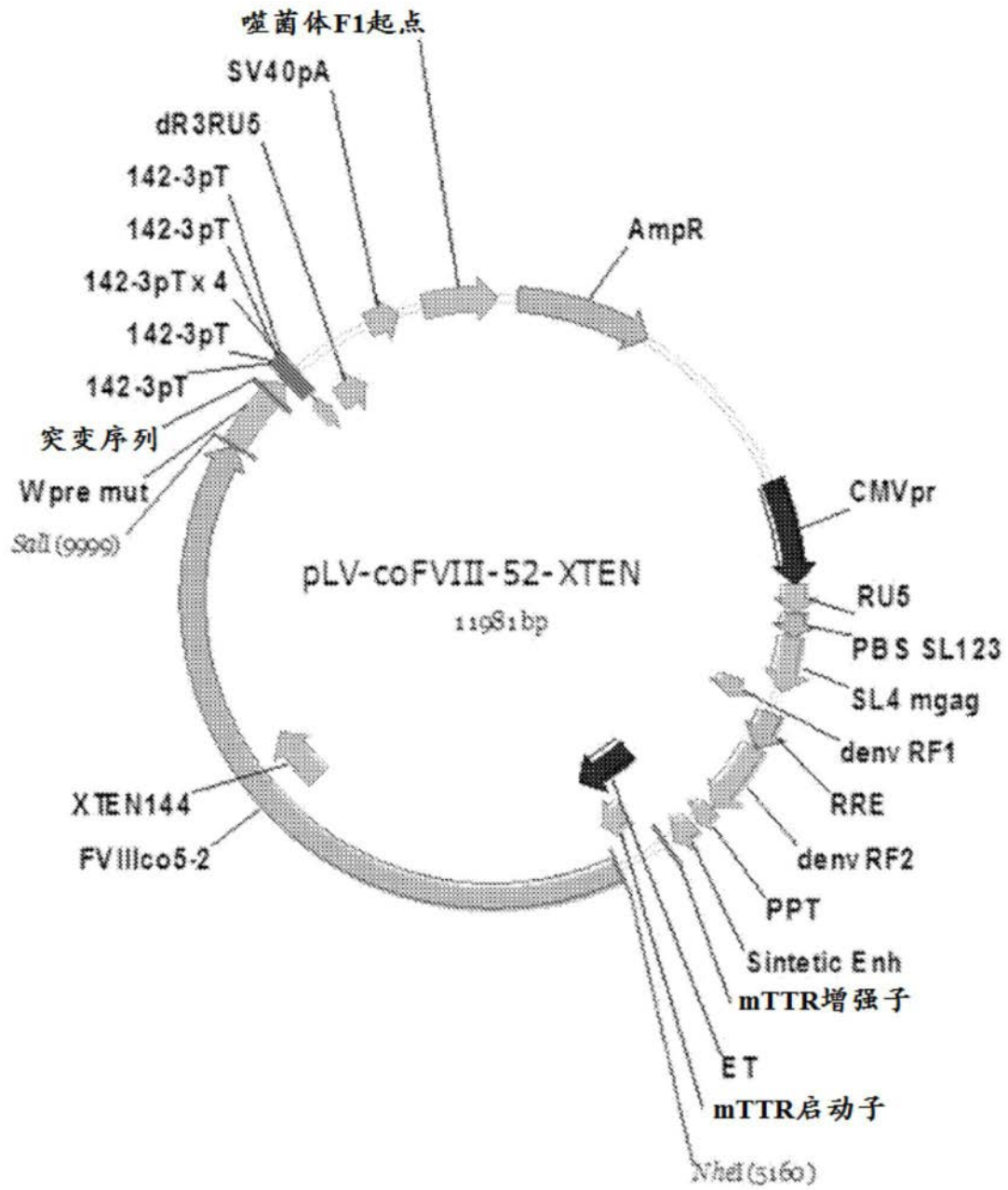


图9:pLV-coFVIII-52-XTEN的质粒图

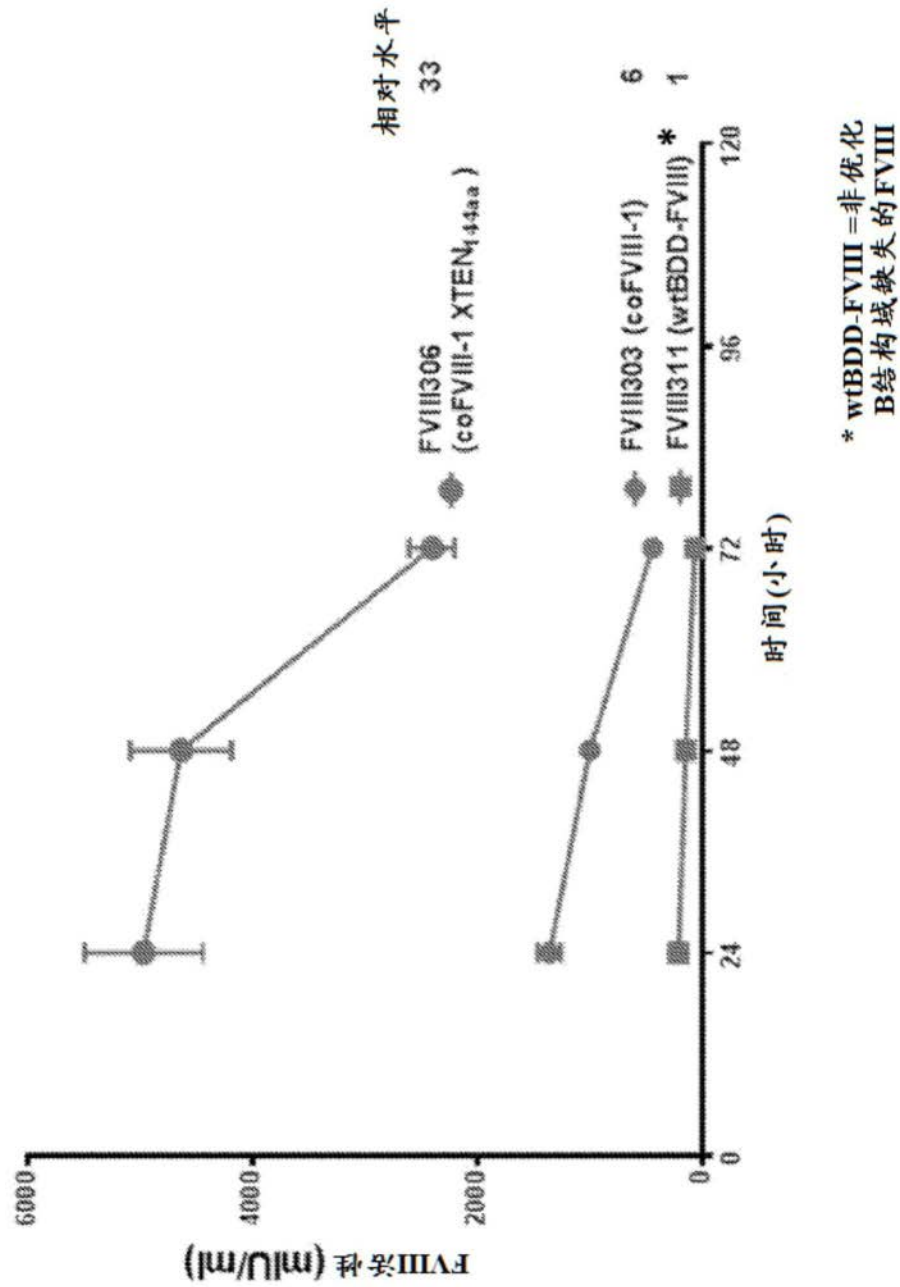
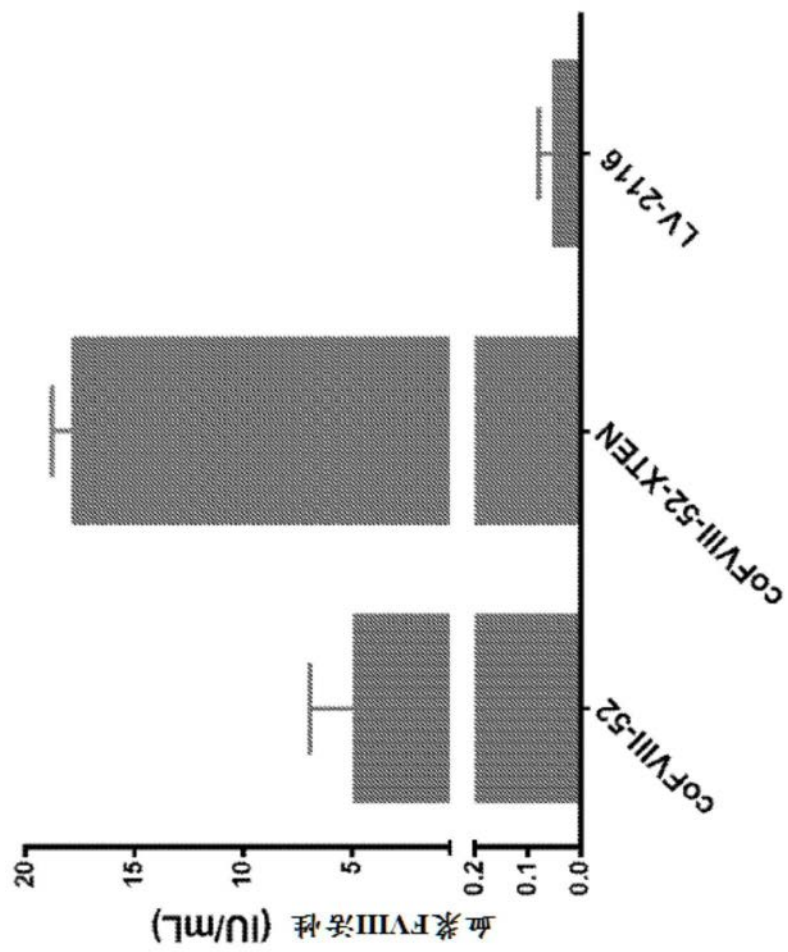


图10A:流体动力学注射质粒DNA后HemA小鼠中的FVIII活性



*LV-2116 = 在慢病毒质粒中的ET启动子
控制下的B结构域缺失的FVIII

图10B:慢病毒注射后21天HemA小鼠中的FVIII活性

ATRRYYLGAVELSWDYMQSDLGELPVDARFPPRPVPSFNFNTSVVYKTLFVEFTDHLFNIAPRPPWMLLGPITQAEVYDVTITLK
 NMASHPVSLHAGVSVYWKASEGAEDDQTSQREKEDDKVFPGGSHYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGA
 LLVCREGLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGVYNRSLPGLIGCHRKSVYWHVIGMG
 TPEVHSIFLEGHTFLVRNHRQASLEISPTFLTAQTLLMDLGQFLFCISSHQHDGMEAYVKVDSCEEPQLRMKNNEEAEDYDDDLT
 DSEMDVVRFDNNSPSFIQIRSVAKKHPKTWVHYIAEEEDWDYAPLVLPAPDRSRYKSQYLNNGPQRIGRKYKVRFMAYTDETFKTR
 EAIQHESGILGPLLYGEVGDITLLIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRY
 SSFVNMERDLASGLIGPLLYCYKESVDQQRNQIMSDKRNVLFSVFDENRSWYLTENIQRFLPNPAGVQLEDPEFQASNIMHSINGYVF
 DSLQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKHKMVYEDTLTFPFSGETVFMMSMENPGLWILGCHNSDFNRNMGMTALLKVSSC
 DKNTGDYEDSYEDISAYLLSKNNAIEPRFSQNSRHPSTRQKQFNATTIPENDIEKTDPWFAHRTMPKIQNVSSDMLLRQSPTPH
 GLSLDLQEAKEYTFSDDPSPGAIDSNNSLSEMTHFRPQLHHSGLDMVFTPEGLQLRLNEKLGTTAATELKKLDFKVSSTSNLISTIPSD
 NLAAGTDNTSSLGPPSMPVHYDSQLDITLFGKKSSPLTESGGPLSLSEENNDKLLSGLMNSQESSWGKNVSSTESGRLFKGRKRAHGP
 ALLTKDNALFKVISILLKTNKTSNNSATNRKTHIDGPSLLIENSPSVWQNILESDETFKKTPLIHDRMLMDKNATALRLNHSNKTSS
 KNMEMVQKKEGPIPPDAQNPDMSEFFKMLFLPESARWIQRTHGKNSLSNGQGPSKQLVSLGPEKSVGEQNFLEKNKVVGKGEF
 TKDVGKEMVFPSSRNFLTLNLDNLHENNTHNQEKIQEEIEKKEKLIQENVVLPQIHTVTGTKNFMKNLFLSTRQNVESYDGAAYP
 VLQDFRSLNDSTNRKTHAHFSKKGEEENLEGLNQTKQIVEKYACTTRISPNTSQNFVTQSRKALKQFRPLEETELEKRIIVDDTS
 TQWSKNMKHLTPSTLTQIDYNEKEGAITQSPSLDCLTRSHSIPQANRSLPIAKVSSFPSIRPIYLTRVLFDQDNSSHLPAASYRKKDSGVQ
 ESSHFLQGAKKNNLSAILTLEMTGDQREVGLSGTSATNSVTYKKVENTVLPKPDLPKTSKGVELLPKVHIYQKDLFPTETSNNGSPGHLDL
 VEGSLQGTGAIKWNEANRPGKVPFLRVATESAKTPSKLLDPLAWDNHYGTQIPKEEWKSQEKSPKTAFFKKDITLSLNACESNHAI
 AAINEGQNKPEIEVTWAKQGRTERLCSQNPVLRHQRERITRTLQSDQEEIDYDDTISVEMKKEDFDIYDEDENQSPRSFQKTRHYFI
 AAVERLWDYGMSSSPHVLNRNAQSGSVPPQFKKVVFEFTDGSFTQPLRGELNEHLGLLGPYIRAEVEDNIMVTFRNQASRPYSFYSSL
 ISYEEDQRQGAEPKRNFKVNETKTYFWKVQHMAPTKDEFDCAWAYFSDVLEKDVHSLGIGPLLVCHTNTLPAHGRQYTVQE
 FALFFTIFDETKSWYFTENMERNCRAPCNQMEDPTFKNYRFHAINGYIMDTLPGLVMAQDQIRIRWYLLSMGSNENIHSIHFSGHVF
 TVRKKEEYKMALYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPLGMASGHIRDQITASGQYQGWAPKLA
 RLHYSGSINAWSTKEPFSWIKVDLLAPMIIHGKIQGARQKFSLSYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKNIFN
 PPIIARYIRLHPHTHSIRSLRMELMGCDLNSCSMPLGMESKAISDAQITASSYFTNMFATWSPSKARLHLQGRSNAWRPQVNNPKEW
 LQVDFQKTMKVTGVTQGVKSLTSMYVKEFLISSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLTRYLRHPQSWVHQ
 IALRMEVLGCEAQDLY

图11A:全长成熟人因子VIII氨基酸序列 (SEQ ID NO:15)

ATRRYYLGAVELSWDYMQSDLGELPVDARFPPRPVPSFNFNTSVVYKTLFVEFTDHLFNIAPRPPWMLLGPITQAEVYDVTITLK
 NMASHPVSLHAGVSVYWKASEGAEDDQTSQREKEDDKVFPGGSHYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGA
 LLVCREGLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGVYNRSLPGLIGCHRKSVYWHVIGMG
 TPEVHSIFLEGHTFLVRNHRQASLEISPTFLTAQTLLMDLGQFLFCISSHQHDGMEAYVKVDSCEEPQLRMKNNEEAEDYDDDLT
 DSEMDVVRFDNNSPSFIQIRSVAKKHPKTWVHYIAEEEDWDYAPLVLPAPDRSRYKSQYLNNGPQRIGRKYKVRFMAYTDETFKTR
 EAIQHESGILGPLLYGEVGDITLLIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRY
 SSFVNMERDLASGLIGPLLYCYKESVDQQRNQIMSDKRNVLFSVFDENRSWYLTENIQRFLPNPAGVQLEDPEFQASNIMHSINGYVF
 DSLQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKHKMVYEDTLTFPFSGETVFMMSMENPGLWILGCHNSDFNRNMGMTALLKVSSC
 DKNTGDYEDSYEDISAYLLSKNNAIEPRFSQNSRHPSTRQKQFNATTIPENDIEKTDPWFAHRTMPKIQNVSSDMLLRQSPTPH
 GLSLDLQEAKEYTFSDDPSPGAIDSNNSLSEMTHFRPQLHHSGLDMVFTPEGLQLRLNEKLGTTAATELKKLDFKVSSTSNLISTIPSD
 NLAAGTDNTSSLGPPSMPVHYDSQLDITLFGKKSSPLTESGGPLSLSEENNDKLLSGLMNSQESSWGKNVSSTESGRLFKGRKRAHGP
 ALLTKDNALFKVISILLKTNKTSNNSATNRKTHIDGPSLLIENSPSVWQNILESDETFKKTPLIHDRMLMDKNATALRLNHSNKTSS
 KNMEMVQKKEGPIPPDAQNPDMSEFFKMLFLPESARWIQRTHGKNSLSNGQGPSKQLVSLGPEKSVGEQNFLEKNKVVGKGEF
 TKDVGKEMVFPSSRNFLTLNLDNLHENNTHNQEKIQEEIEKKEKLIQENVVLPQIHTVTGTKNFMKNLFLSTRQNVESYDGAAYP
 VLQDFRSLNDSTNRKTHAHFSKKGEEENLEGLNQTKQIVEKYACTTRISPNTSQNFVTQSRKALKQFRPLEETELEKRIIVDDTS
 TQWSKNMKHLTPSTLTQIDYNEKEGAITQSPSLDCLTRSHSIPQANRSLPIAKVSSFPSIRPIYLTRVLFDQDNSSHLPAASYRKKDSGVQ
 ESSHFLQGAKKNNLSAILTLEMTGDQREVGLSGTSATNSVTYKKVENTVLPKPDLPKTSKGVELLPKVHIYQKDLFPTETSNNGSPGHLDL
 VEGSLQGTGAIKWNEANRPGKVPFLRVATESAKTPSKLLDPLAWDNHYGTQIPKEEWKSQEKSPKTAFFKKDITLSLNACESNHAI
 AAINEGQNKPEIEVTWAKQGRTERLCSQNPVLRHQRERITRTLQSDQEEIDYDDTISVEMKKEDFDIYDEDENQSPRSFQKTRHYFI
 AAVERLWDYGMSSSPHVLNRNAQSGSVPPQFKKVVFEFTDGSFTQPLRGELNEHLGLLGPYIRAEVEDNIMVTFRNQASRPYSFYSSL
 ISYEEDQRQGAEPKRNFKVNETKTYFWKVQHMAPTKDEFDCAWAYFSDVLEKDVHSLGIGPLLVCHTNTLPAHGRQYTVQE
 FALFFTIFDETKSWYFTENMERNCRAPCNQMEDPTFKNYRFHAINGYIMDTLPGLVMAQDQIRIRWYLLSMGSNENIHSIHFSGHVF
 TVRKKEEYKMALYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPLGMASGHIRDQITASGQYQGWAPKLA
 RLHYSGSINAWSTKEPFSWIKVDLLAPMIIHGKIQGARQKFSLSYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKNIFN
 PPIIARYIRLHPHTHSIRSLRMELMGCDLNSCSMPLGMESKAISDAQITASSYFTNMFATWSPSKARLHLQGRSNAWRPQVNNPKEW
 LQVDFQKTMKVTGVTQGVKSLTSMYVKEFLISSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLTRYLRHPQSWVHQ
 IALRMEVLGCEAQDLY

(X是任何天然氨基酸)

图11B:全长血管性血友病因子,氨基酸序列 (SEQ ID NO:44)

GAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPASS

图11C:XTEN AE42-4,蛋白质序列(SEQ ID NO:46)

GGCGCGCCAGGTTCTCTGCTGGCTCCCCACCTCAACAGAAGAGGGGACAAGCGAAAGCGCTACGCCTGAGAGTGGCCCTGGCT
CTGAGCCAGCCACCTCCGGCTCTGAAACCCCTGCCTCGAGC

图11D:XTEN AE42-4,DNA序列(SEQ ID NO:47)

TSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSE
TPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPG

图11E:XTEN AE144-2A,蛋白质序列(SEQ ID NO:48)

GGCGCGCCAACCAAGTACGGAGCCGTCCGAGGGGAGCGCACCAGGAAGCCCGGCTGGGAGCCCGACTTCTACCGAAGAGGGTACA
TCTACCGAACCAAGTGAAGGTTACGACACAGGCACCTCAACAGAACCTCTGAGGGCTCGGCGCCTGGTACAAGTGAGTCCGCCAC
CCCAGAATCCGGGCTGGGACAAGCACAGAACCTTCGGAAGGGAGTGCCCTGGAACATCCGAATCGGCAACCCAGAACATCAGGG
CCAGGATCTGAGCCCGGACTTCGGGCTCCGAGACGCTGGGACATCCACCGAGCCCTCCGAAGGATCAGCCCCAGGCACACAGCA
CGGAGCCCTCTGAGGGAAGCGCACCTGGTACCAGCGAAAGCGCAACTCCCGAATCAGGTCCCGGTACGAGCGAGTCGGCGACCCC
GGAGAGCGGGCCAGGTGCCTCGAGC

图11F:XTEN AE144-2A,DNA序列(SEQ ID NO:49)

SPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGS
APGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG

图11G:XTEN AE144-3B,蛋白质序列(SEQ ID NO:50)

GGCGCGCCAAGTCCCGCTGGAAGCCCACTAGCACCGAAGAGGGGACCTCAGAGTCCGCCACCCCGAGTCCGGCCCTGGCTCTG
AGCCTGCCACTAGCGGCTCCGAGACTCTGGCACATCCGAAAGCGCTACACCCGAGAGTGGAACCCGGCACCTCTACCGAGCCCAAGT
GAGGGCTCCGCCCTGGAACAAGCACCGAGCCAGCGAAGGCAGCGCCCGAGGGACCTCCACAGAGCCCAAGTGAAGGCAGTGCT
CCTGGCACCAAGCAACCAAGCGAGGGCTCTGCACCCGGGACCTCCACCGAGCCAAGCGAAGGCTCTGCCCTGGCACTTCCA
CCGAGCCCAAGGCAAGGCAGCGCCCTGGGAGCCCGCTGGCTCTCCACCAAGCACTGAGGAGGGCACATCTACCGAACCAAGTGA
AGGCTCTGCACCAAGTGCCTCGAGC

图11H:XTEN AE144-3B,DNA序列(SEQ ID NO:51)

TSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTST
EEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPG

图11I:XTEN AE144-4A,蛋白质序列(SEQ ID NO:52)

GGCGCGCCAACGTCCGAAAGTGCTACCCCTGAGTCAGGCCCTGGTAGTGAGCCTGCCACAAGCGGAAGCGAAACTCCGGGGACCT
CAGAGTCTGCCACTCCCGAATCGGGGCCAGGCTCTGAACCGGCCACTTCAGGGAGCGAAACACCAGGAACATCGGAGAGCGCTAC
CCCGGAGAGCGGGCCAGGAAGTACTGAGCCTAGCGAGGGAAGTGCACCTGGTACAAGCGAGTCCGCCACACCCGAGTCTGG
CCCTGGCTCTCCAGCGGGCTACCCACGAGCACTGAAGAGGGCTCTCCCGCTGGCAGCCCAACGTCGACAGAAGAAGGATACCA
GCAGGCTCCCCACATCAACAGAGGAGGGTACATCAGAATCTGCTACTCCCGAGAGTGGACCCGGTACCTCCACTGAGCCAGCG
AGGGGAGTGCACCAAGTGCCTCGAGC

图11J:XTEN AE144-4A,DNA序列 (SEQ ID NO:53)

TSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGP
GSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEG

图11K:XTEN AE144-5A,蛋白质序列 (SEQ ID NO:54)

GGCGCGCCAACATCAGAGAGCGCCACCCCTGAAAGTGGTCCGGGAGCGAGCCAGCCACATCTGGGTCCGAAACGCCAGGCACAA
GTGAGTCTGCACTCCCGAGTCCGGACCTGGCTCCGAGCCTGCCACTAGCGGCTCCGAGACTCCGGGAACTTCCGAGAGCGCTACAC
CAGAAAGCGGACCCGGAACCACTAGCGAGGGCTCTGCTCCGGGCAGCCAGCCGGCTCTCTACATCCACGGAGGAG
GGCACTTCCGAATCCGCCACCCGGAGTCAGGGCCAGGATCTGAACCCGCTACCTCAGGCAGTGAGACGCCAGGAACGAGCGAGTC
CGCTACACCGGAGAGTGGGCCAGGGAGCCCTGCTGGATCTCTACGTCCACTGAGGAAGGGTCACCAGCGGGCTCGCCACACAGCA
CTGAAGAAGGTGCCTCGAGC

图11L:XTEN AE144-5A,DNA序列 (SEQ ID NO:55)

TSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAP
GTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPG

图11M:XTEN AE144-6B,蛋白质序列 (SEQ ID NO:56)

GGCGCGCCAACATCTACCGAGCCTCCGAAGGCTCTGCCCTGGGACCTCAGAATCTGCAACCCCTGAAAGCGGCCCTGGAACCTCC
GAAAGTGCCACTCCCGAGAGCGGCCAGGGACAAGCGAGTCAGCAACCCCTGAGTCTGGACCCGGCAGCGAGCCTGCAACCTCTGG
CTCAGAGACTCCCGGCTCAGAACCCGCTACCTCAGGCTCCGAGACCCGGCTCTCTGCTGGGAGTCCCACTTCCACCGAGGAAGG
AACATCCACTGAGCCTAGTGAGGGCTCTGCCCTGGAACCAAGCAGACAGAGCCAAGTGAGGGCAGTGACACAGGATCCGAGCCAGCAA
CCAGCGGGTCCGAGACTCCCGGACCTCTGAGTCTGCCACCCAGAGAGCGGACCCGGCACTTCAACCGAGCCCTCCGAAGGATCA
GCACCAAGTGCCTCGAGC

图11N:XTEN AE144-6B,DNA序列 (SEQ ID NO:57)

PGSSPSASTGTGPGSSPSASTGTGPGTSGTASSSPGSSTPSGATSGPSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGA
TGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSS

图11O:XTEN AG144-1,蛋白质序列 (SEQ ID NO:58)

GGCGCGCCACCCGGGTCGTCCTCCCGTCGGCGTCCACCGGAACAGGGCCAGGGTCATCCCGTCAGCGTCGACTGGGACGGGACCCGG
GACACCCGGTTCGGGGACTGCATCCTCCTCGCCTGGTTCGTCCACCCCGTCAGGAGCCACGGGTTCCCGGGAAGCAGCCCAAGCGC
ATCCACTGGTACAGGGCCTGGGGCTTACCGGGTACTTCATCCACGGGGTCACCGGGAACGCCGGATCGGGGACGGCTTCTCATC
ACCAGGATCGTCAACACCTCGGGCGAACGGGCAGCCCCGAACCCCTGGTTCGGGTACGGCGTCGTCGAGCCCCGGTGCGAGCC
CGGGAACAAGCTCGACAGGATCGCCTGGGGCGTCACCCGGCACGTGAGCACAGGCAGCCCCGAACCCCTGGATCGGGAACCGC
GTCGTCAAGCGCCTCGAGC

图11P:XTEN AG144-1,DNA序列 (SEQ ID NO:59)

GASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTAS
SSPGSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSP

图11Q:XTEN AG144-A,蛋白质序列 (SEQ ID NO:60)

GGCGCGCCAGGTGCCTCGCCGGGAACATCATCAACTGGTTCACCCGGGTATCCCTCGGCCTCAACCGGGACGGGTCCCGGCTCA
TCCCCAGCGCCAGCACTGGAACAGGTCTGGCACTCCTGGTTCGGTACGGCATCGTCATCCCGGGAAGCTCAACACCGTCCGGA
GCGACAGGATCACTGGCTCGTCACCTTCGGCGTCAACTGGAACGGGGCCAGGGGCTCACCCGGAACGTCTCGACTGGGTGCGCT
GGTACGCCGGGATCAGGAACGGCCTATCCTCGCCTGGGTCTCAACGCCCTCGGGTGCAGTGGTTCGCCGGGAACCTCTGGCTCG
GGGACGGCCTCGTCGCTGCCTGGGGCATACCGGGGACGAGCTCCACGGGGTCCCTGGAGCGTCACCGGGGACCTCTCGACAGG
TAGCCCGGCCTCGAGC

图11R:XTEN AG144-A,DNA序列 (SEQ ID NO:61)

GTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGAT
GSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSP

图11S:XTEN AG144-B,蛋白质序列 (SEQ ID NO:62)

GGCGCGCCAGGTACACCGGGCAGCGGCACGGCTTCGTGTCACCCGGCTCGTCCACACCGTCGGGAGCTACGGGAAGCCAGGAGC
GTCACCGGGAACGTGTCACCGGGTACCGGGTACGCCAGGTAGCGGCACGGCCAGCAGCTCGCCAGGTTTCATCGACCCCGTCGG
GAGCGACTGGGTGCGCCGGATCAAGCCCGTCAGCTTCACTGGAACAGGACCCGGGTCGTGCGCGTCAGCCTCAACGGGGACAGGA
CCTGGTTCATCGACGCCGTAGGGGCGACAGGCTCGCCCGGATCGTCAACACCTCGGGGGCAACGGGGAGCCCTGGTGCGTCGCC
TGGAACCTCATCCACCGGAAGCCGGGGGCTCGCCGGGTACGAGCTCCACGGGATCGCCCGGAGCGTCCCCGGAACCTCAAGCA
CAGGGAGCCCTGCCTCGAGC

图11T:XTEN AG144-B,DNA序列 (SEQ ID NO:63)

GTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTSGTASSSPGASPGTSSTGSPGASPGTSST
GSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSP

图11U:XTEN AG144-C,蛋白质序列 (SEQ ID NO:64)

GGCGGCCAGGTACACCGGATCGGGTACAGCGTCATCGAGCCCCGGTGCCTCACCTGGTACGTCGAGCACGGGGTCGCCAGGGG
GTCCCTGGGACGTCCTCAACAGGCTCGCCGGTGCCTACCCGGCACGTCGTCACGGGTTACCTGGTAGCTCCCTTCCGCGTCC
ACTGGCACCGGGCTGGAATCCGGGGAGCGGCACAGCGAGCTCGTCGCCGGGAGCATCGCTGGGACATCGAGCACCGGGTCGC
CAGGAGCATCGCCCGAATCCAGCACAGGAAGCCCCGGCGCTCGCCGGGACATCAAGCACAGTTCCCGGGATCGAGCACG
CCGTCCGGAGCCACTGGATCACCAGGGAGCTCGACACCTTCGGCGCAACGGGATCGCCCGAGCCAGCCGGGTACGTCAAGCAC
TGGTCCCTGCCTCGAGC

图11V:XTEN AG144-C,DNA序列 (SEQ ID NO:65)

GSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTG
TGP TPGSGTASSPSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSP

图11W:XTEN AG144-F,蛋白质序列 (SEQ ID NO:66)

GGCGCGCCAGGCTCCAGCCCCTCGCGAGCACGGGAACCGGACCAGGTTTCGTACCCCTCAGCATCAACGGGGACGGGACCGGGGG
CGTCACCAGGAACGTCTCCACCGGCTCGCCGGTGCATACCCGGAACGTCATCGACCGGATCGCCAGGGAGCTCGACGCCATCAG
GCGCAACAGGATCACCTGGCTCAAGCCCTAGCGCGTCAACCGGCACGGTCCGGGTGCCTCCCTGGCACGTCCAGCACCGGATCAG
CCGGATCGAGCCATCCGCTCAACCGGAACCGGACCGGTACACCAGGTCGGGAACAGCCTCCTCGTCACCAGGCTCCTCAACCC
CTCGGGAGCCACGGGTTCCCGGTTCTGTCACGCTTCGGAGCAACTGGTAGCCCCGGAGCATCGCCAGGAATTCGAGCACG
GGTCGCCCCGCTCGAGC

图11X:XTEN AG144-F,DNA序列 (SEQ ID NO:67)

CTCGAGGTCAATTCACGCGAGTTAATAATTACCAGCGCGGGCCAAATAAATAATCCGCGAGGGGCAGGTGACGTTTGCCAGCGCGC
GCTGGTAATTATTAACCTCGCAATATTGATTCGAGGCCGCGATTGCCGCAATCGCGAGGGGCAGGTGACCTTTGCCAGCGCGCGT
TCGCCCCGCCCCGACGGTATCGATAAGCTTAGGAGCTTGGGCTGCAGGTCGAGGGCACTGGGAGGATGTTGAGTAAGATGAAAA
CTACTGATGACCTTGCAGAGACAGAGTATTAGGACATGTTGAACAGGGGCGGGCGATCAGCAGGTAGCTCTAGAGGATCCCCGT
CTGTCTGCACATTTCTAGAGCGAGTGTTCCGATACTCTAATCTCCCTAGGCAAGGTTCAATTTGTGTAGGTTACTTATTCCTTTTG
TTGACTAAGTCAATAATCAGAATCAGCAGGTTTGAGTCAGCTTGGCAGGGATCAGCAGCCTGGGTTGGAAGGAGGGGGTATAAAA
GCCCCCTCACCAGGAGAAGCCGTCACACAGATCCACAAGCTCCTGCCACCATGG

图11Y:ET启动子,DNA序列 (SEQ ID NO:69)

ATGCAGATTGAGCTGCTACTTGCTTTTCTGTGCTGCTGAGGTTTGTCTTTCCGCTACACGAAGGTATTATCTGGGGGCTGTGGA
ACTGTCTTGGGATTACATGCAGAGTGACCTGGGAGAGCTGCCAGTGGACGCAAGGTTTCCCCCTAGAGTCCCTAAGTCATTCCTTC
AACACTAGCGTGTCTACAAGAAACACTGTTCTGTGGAGTTTACTGATCACCTGTTCAACATCGCAAAGCCTAGGCCACCTGGATGG
GACTGCTGGGGCCAAACATCCAGGCCGAGGTGTACGACCCGTGGTCATTACACTTAAGAACATGGCCTCACACCCCGTGAGCCTGC
ATGCTGTGGGCGTCAGTACTGGAAGGCTCCGAAGGAGCAGAGTATGACGATCAGACTTCCAGAGAGAAAAAGAGGACGATAAG
GTGTTTCTGGCGGATCTCATACCTACGTGTGGCAGGTCTGTAAAGAGAAATGGCCCTATGGCCTCCGACCTCTGTGCTGACCTACT
CTTATCTGAGTCACGTGGACCTGGTCAAGGATCTGAACAGCGGCCTGATCGGAGCCCTGCTGGTGTGCAGGGAAGGAAGCCTGGCTA
AGGAGAAAACCCAGACACTGCATAAGTTCATTCTGCTGTTGCGCGTGTGGACGAAGGAAATCATGGCACAGCGAGACAAAGAATA
GTCTGATGCAGGACAGGGATGCCGCTTCAGCCAGAGCTTGGCCCAAAATGCACACTGTGAACGGCTACGTCAATCGCTCACTGCCTG
GGCTGATCGGCTGCCACCGAAAGAGCGTGTATTGGCATGTCATCGGGATGGGCCACACACTGAAGTGAAGTCCATTTCTGGAGG
GACATACCTTTCTGGTCCGCAACCACCGACAGGCTTCCCTGGAGATCTCTCCAATTACCTTCTGACAGCACAGACTCTGCTGATGGAC
CTGGGGCAGTTCTGCTGTTTTGCCACATCAGCTCCACCCAGCATGATGGCATGGAGGCTTACGTGAAAGTGGACTCTTGTCCCGAGG
AACCTCAGCTGCGGATGAAGAACAATGAGGAAGCAGAAGACTATGACGATGACCTGACCGACTCCGAGATGGATGTGGTCCGATT
GATGACGATAACAGCCCTCTTTATCCAGATTAGATCTGTGGCAAGAAACACCCTAAGACATGGGTCCATTACATCGCAGCCGAGG
AAGAGGACTGGGATTATGCACCACTGGTGTGGCACCAGACGATCGCTCTACAATCTCAGTATCTGAACAATGGGCCACAGAGGA
TTGGCAGAAAGTACAAGAAAGTGGGTTTCATGGCATATACCGATGAGACCTTCAAGACTCGCGAAGCCATCCAGCAGGAGAGCGGCA
TCTGGGACCACTGCTGTACGGAGAAAGTGGGAGACCCCTGTGATCATTTTCAAGAACCAGGCCAGCGCCTTACAATATCTATCC
ACATGGGATTACAGATGTGCGCCCTGTACAGCAGGAGACTGCCAAAGGGCGTCAAACACCTGAAGGACTTCCCAATCTGCCCGG
AGAAATCTTCAAGTACAAGTGGACTGTACCGCTCAGGATGGCCCCACTAAGAGCGACCCCTCGGTGCCTGACCCGCTACTATTCTAGT
TTCGTGAATATGGAAGAGATCTGGCAAGCGGACTGATCGGACCACTGCTGATTTGTTACAAAGAGAGCGTGGATCAGAGAGGCAAC
CAGATCATGTCCGACAAGCGGAATGTGATTCTGTTCAGTGTCTTTGACGAAAACAGGTGATGGTACCTGACCGAGAACATCCAGAGAT
TCTGCCTAATCCAGCTGGGGTGACGCTGGAAGATCCTGAGTTTCAAGCATTAACATCATGCATAGTATTAATGGCTACGTGTTGCA
CAGTTTGACGCTGAGCGTGTGCTGCACGAGGTGCTTACTGGTATATCTGAGCATTGGGGCACAGACAGATTTCTGAGCGTGTTC
TTTTCCGGCTACACTTTTAAGCATAAAATGGTCTATGAGGACACACTGACTCTGTTCCCTTCAGCGGCGAAACCGTGTATTATGAGCAT
GGAGAATCCCGGACTGTGGATTCTGGGGTGCCACACAGCGATTTAGAAATCGCGGAATGACTGCCCTGCTGAAAGTGTCAAGCTG
TGACAAGAACACCGGGGACTACTATGAAGATTATACGAGGACATCAGCGCATATCTGCTGTCCAAAAACAATGCCATTGAACCCG
GTCTTTTATGTCAGAACTCTCCAGTGTGAAGAGGCACAGAGGAGATCACCGCACTACCTGACAGAGTATCAGGAAGAGATCGA
CTACGAGATACAATTTCTGTGGAATGAAGAAAGAGGACTTCGATATCTATGACGAAGATGAGAACCAGAGTCCCTGATCATTTCCAG
AAGAAAACAGGCATTACTTTATTGCCGAGTGGAGCGGCTGTGGGATTATGGCATGTCTCTAGTCTCAGTGCTGCGAAATAGG
GCCAGTCAGGAAGCGTCCACAGTTCAAGAAAGTGGTCTTCAGGAGTTTACAGACGGGTCTTTACTCAGCCATGTACAGGGGC
GAACTGAACGAGCACCTGGGACTGCTGGGGCCCTATATCAGAGCAGAAGTGGAGGATAACATTATGGTCACCTTCAGAAATCAGGCC
TCTCGGCTTACAGTTTTTATTCAAGCCTGATCTTACGAAGAGGACACGACAGGGAGCTGAACACGAAAAAACTTCGTGAAGC
CTAATGAGACCAAAACATACTTTTGAAGGTGCAGCACCATATGGCCCCAACAAAAGACGAGTTCGATTGCAAGGCATGGGCTATTT
TTCTGACGTGGATCTGGAGAAGGACGTGCACAGTGGCCTGATTGGCCCACTGCTGGTGTGCCATACTAACCCCTGAATCCAGCCAC
GGCCGGCAGGTCACTGTCCAGGAGTTCGCTCTGTTCTTACCATCTTTGATGAGACAAAGAGCTGGTACTTCACCGAAAACATGGAGC
GAAATTGCAGGGCTCCATGTAACATTAGATGGAAGACCCACATTCAAGGAGAATACCGCTTTCATGCTATCAATGGATACATCAT
GGATACTCTGCCCCGGGCTGGTATGGCACAGGACAGAGAATCCGGTGGTATCTGCTGAGCATGGGCAGCAACGAGAATATCCACTC
AATTCAATTCAGCGGCGACGTGTTTACTGTACGAAGAAAGAAGAGTACAAGATGGCCCTGTACAACCTGTATCCCGCGTGTTCGAA
ACCGTCGAGATGCTGCCTAGCAAGGCCGGAATCTGGAGAGTGAATGCCTGATTGGAGAGCACCTGCATGCTGGGATGTCTACCTG
TTTCTGGTGTACAGTAATAAGTGTACAGACCCCTGGGAATGGCATCCGGGCATATCAGGGATTTCAGATTACCGCATCTGGACAGT
ACGGACAGTGGGCACCTAAGCTGGCTAGACTGCACTATTCGGATCTATCAACGCTTGGTCCACAAAAGAGCCTTTCTCTTGGATTAA
GGTGGACCTGCTGGCCCAATGATCATTATGGCATCAAACTCAGGGAGCTCGGCAGAAATTCTCTCTCTGTACATCTCAGATTTA
TCATCATGTACAGCCTGGATGGGAAGAAATGGCAGACATACCGCGGCAATAGCACAGGAATCTGATGGTGTCTTTGGCAACGTGG
ACAGCAGCGGAATCAAGCACAACTTTCAATCCCCCTATCATTGCTAGATACATCCGGCTGCACCAACCCATTATCTATTGAAAGT
ACACTGAGGATGGAATGATGGGATGCGATCTGAACAGTTGTTCAATGCCCTGGGGATGGAGTCCAAGGCAATCTGTACGCCCAG
ATTACCGCCAGCTCCTACTTCACTAATATGTTGCTACCTGGAGCCCTCCAAAGCAAGACTGCACCTGCAAGGCCGCGACGACGATG
GCGACCACAGGTGAACAATCCCAAGGAGTGGTTCAGGTCGATTTTCAGAAAATATGAAGGTGACCGGGGTCAAACTCAGGGCG
TGAAAAGTCTGCTGACCTCAATGTACGTCAAGGAGTTCCTGATCTAGTTTCACAGGACGGACATCAGTGGACACTGTTCTTTAGAA
CGGGAAGGTGAAAGTCTTCCAGGGCAATCAGGATTCCTTACACCTGTGGTCAACAGTCTAGACCCTCACTGCTGACCAAGATACCTG
AGAATCCACCCTCAGTCTGGGTGCACCAAGATTGCCCTGAGAATGGAAGTGTGGGATGCGAGGCCAGGATCTGTACTGA

图11Z:coFVIII-1-DNA序列 (SEQ ID NO:68)

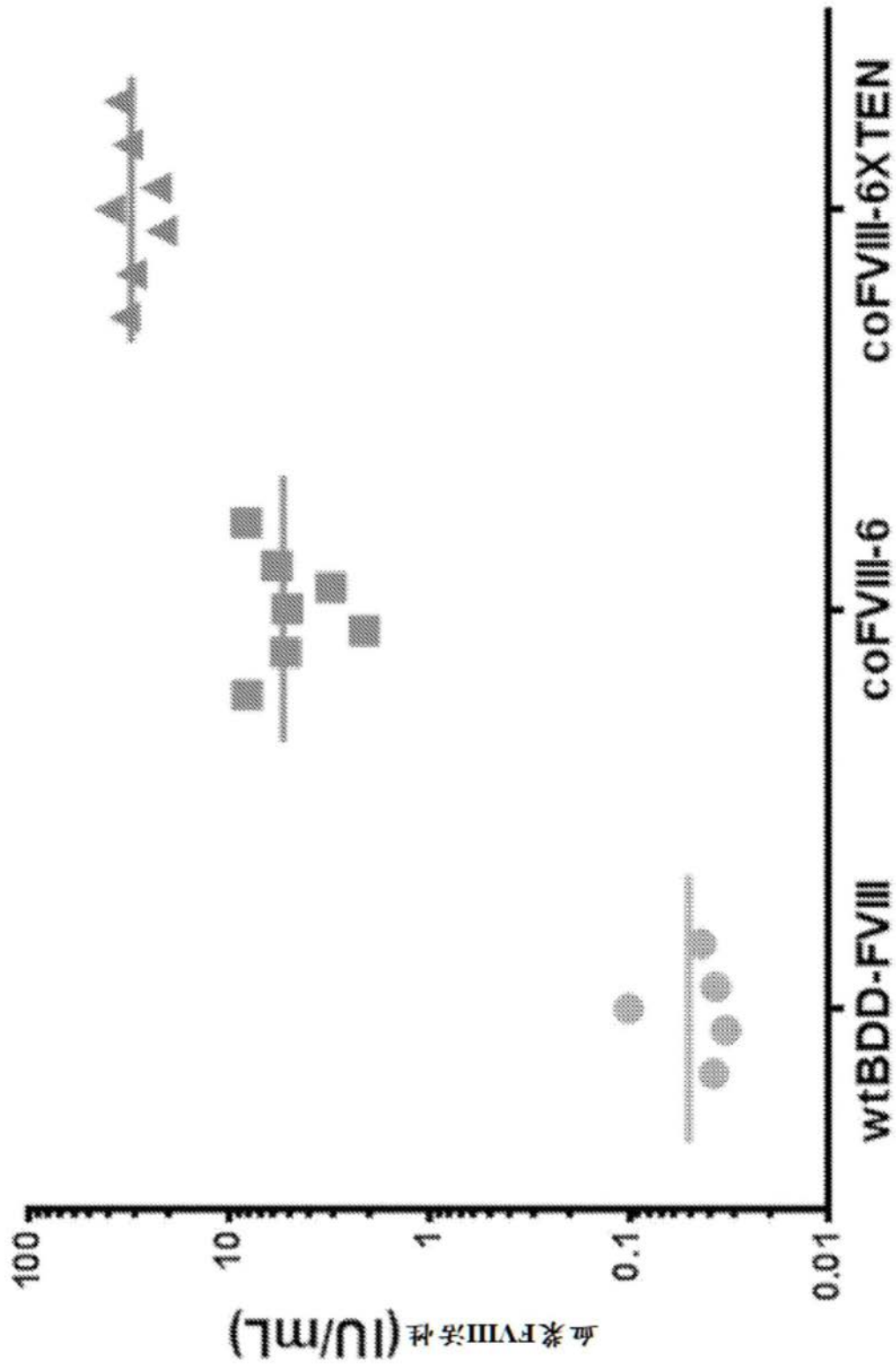


图12A: 用LV-FVIII对HemA小鼠进行IV注射

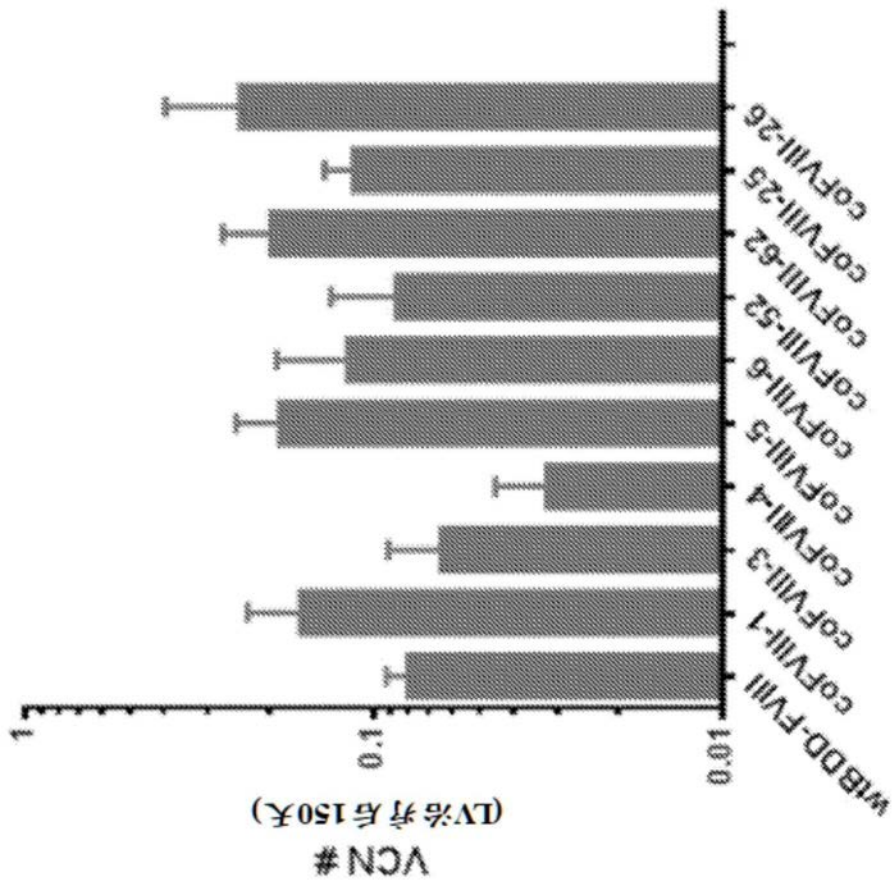


图12B

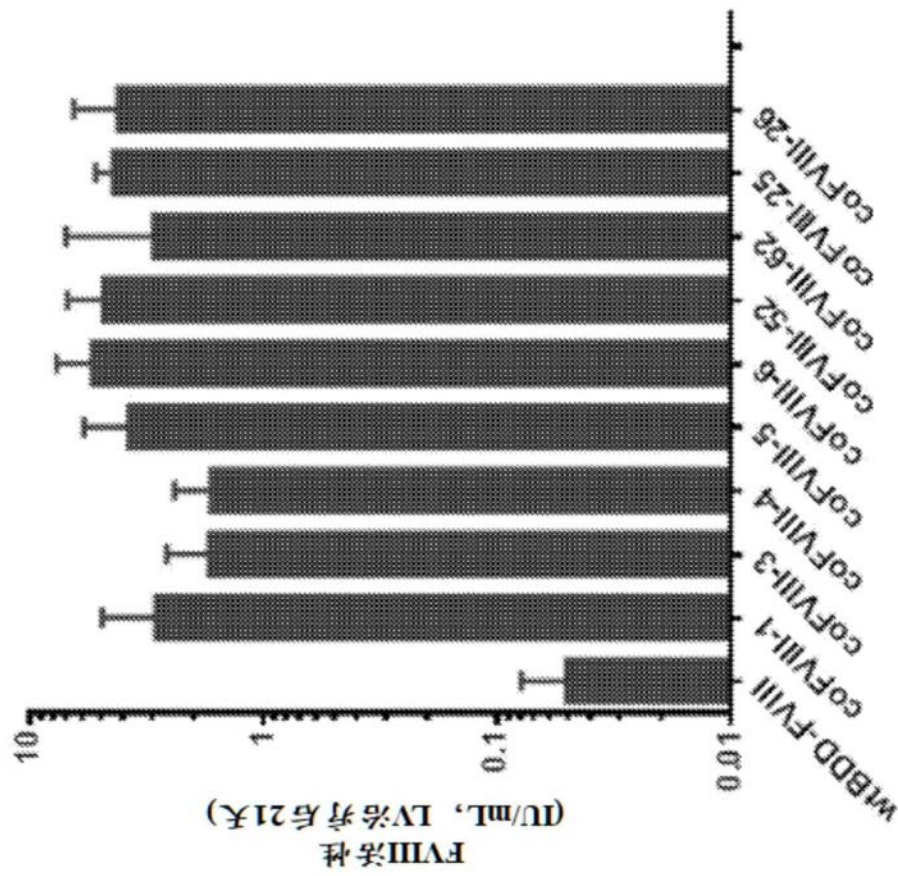


图12C

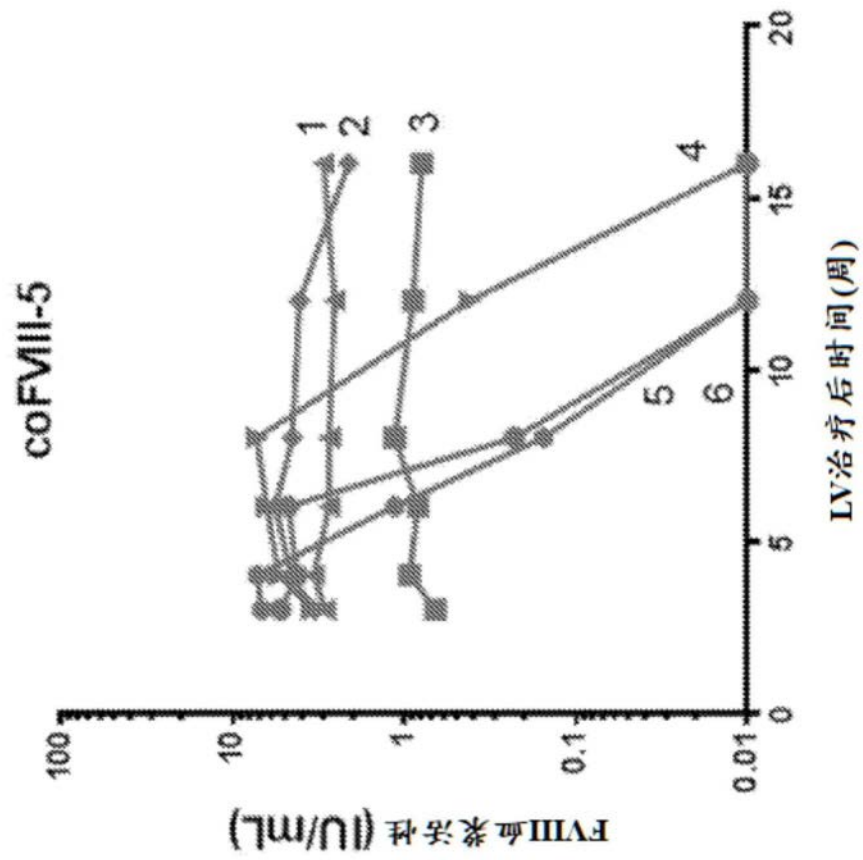


图13A

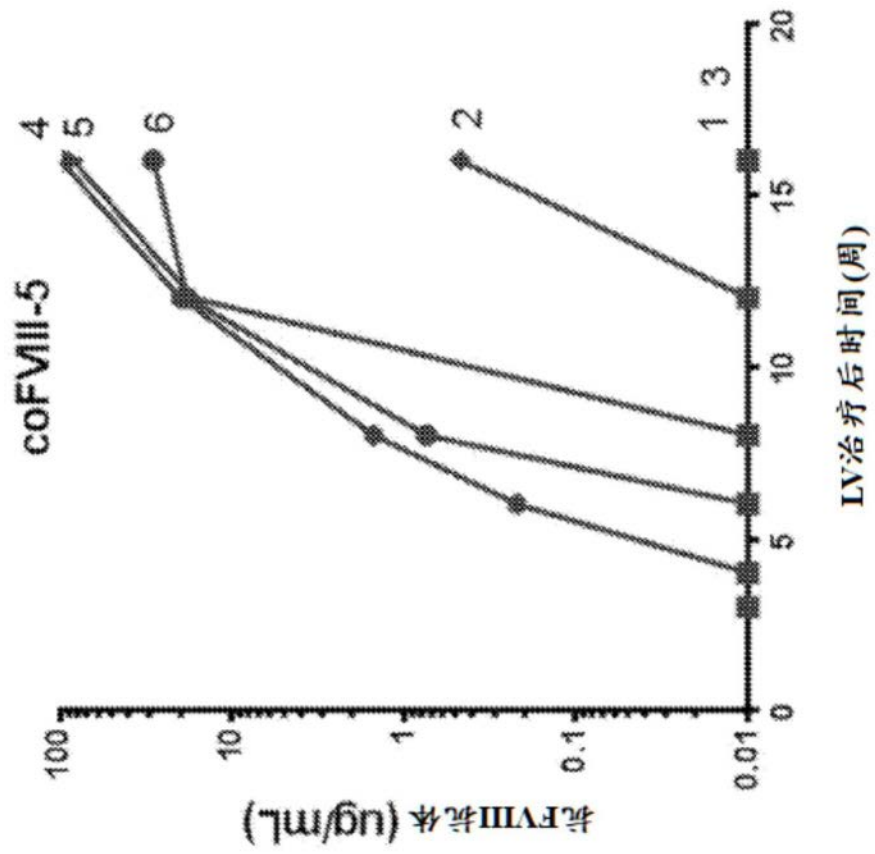


图13B

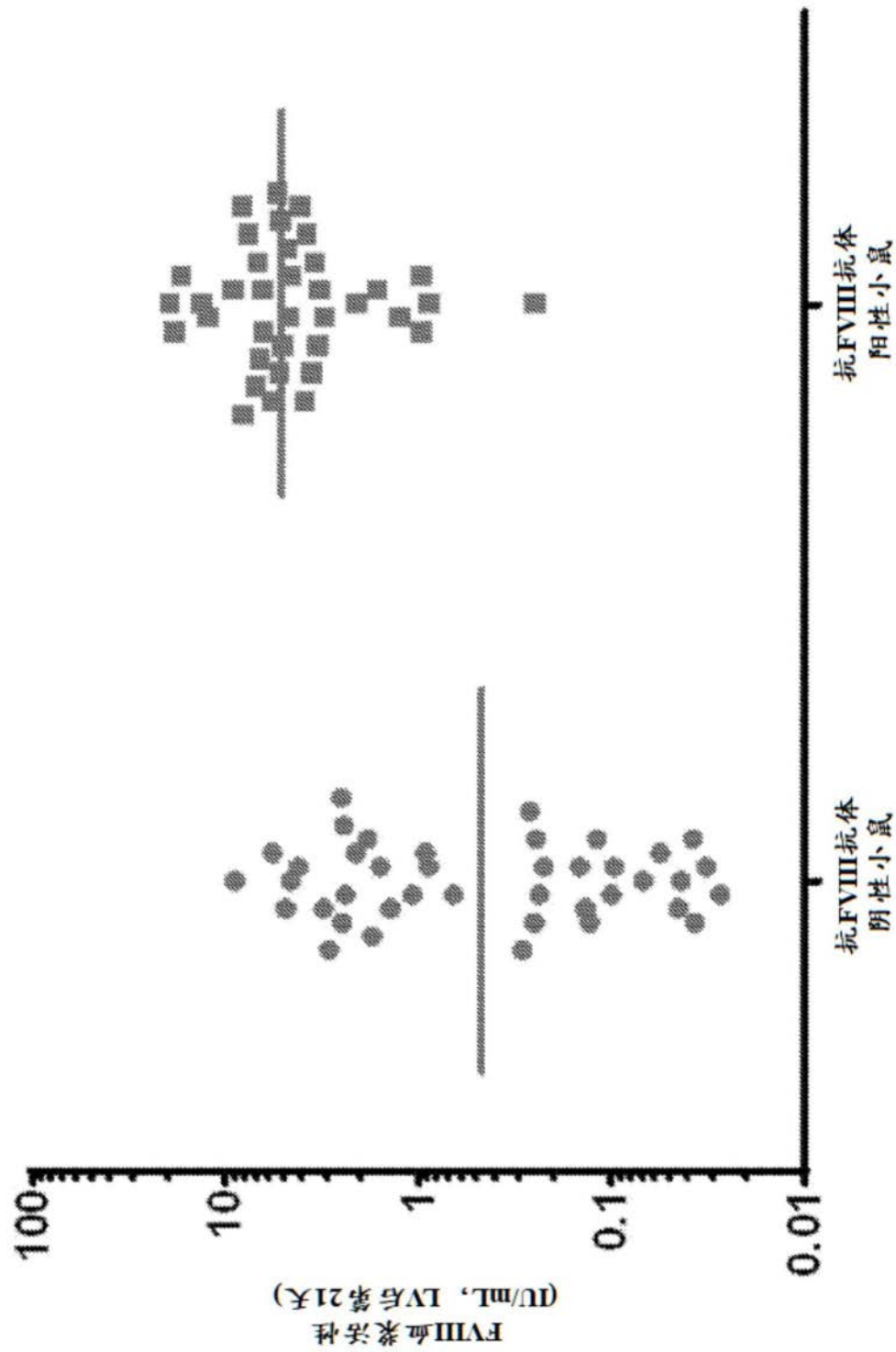


图14

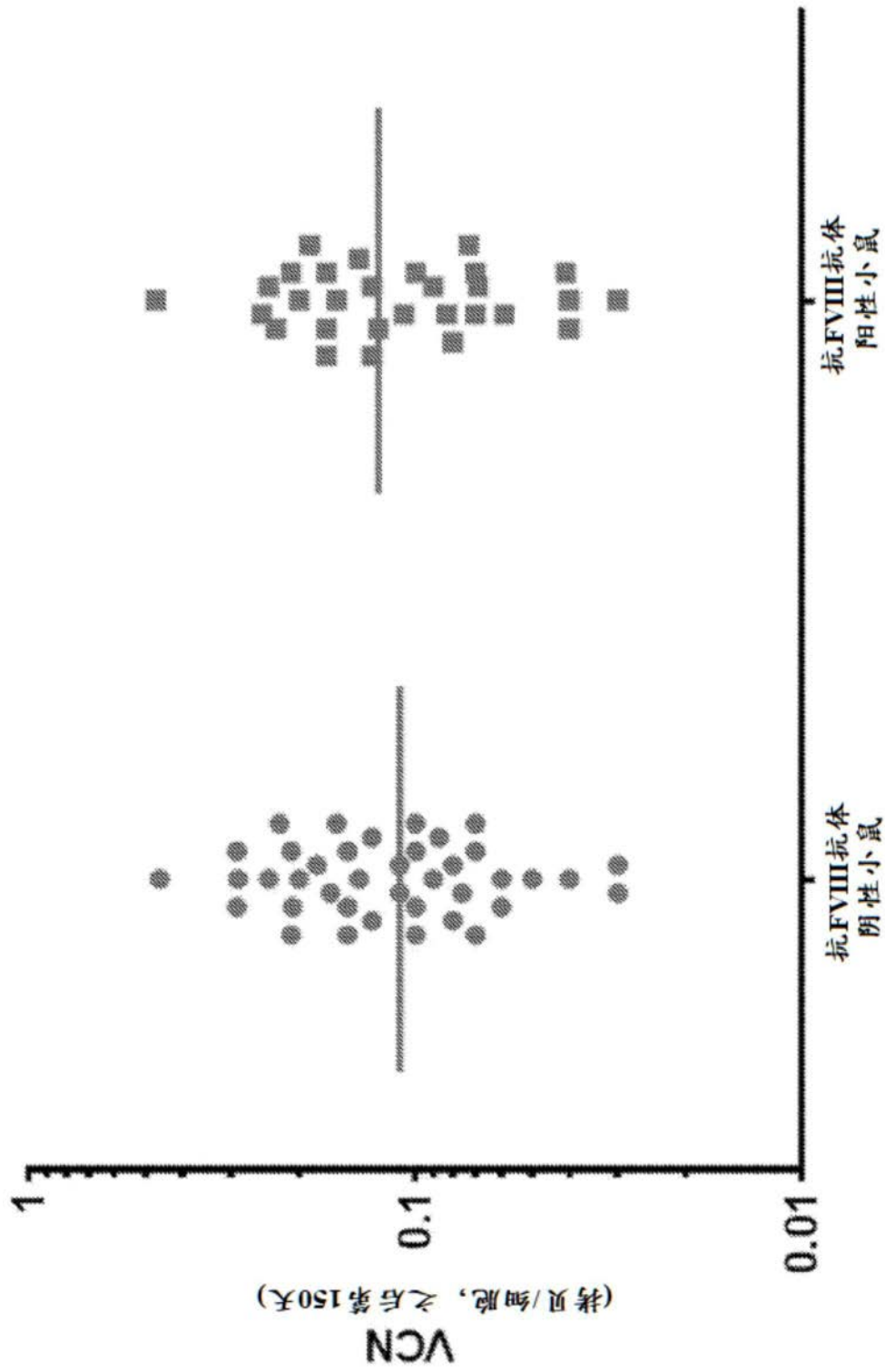


图15

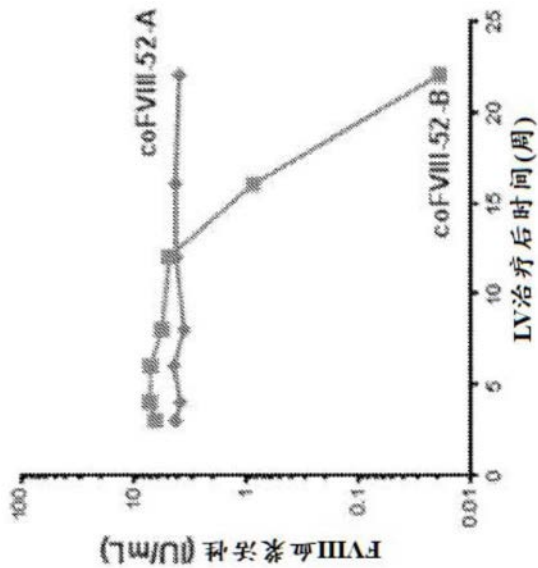


图16A

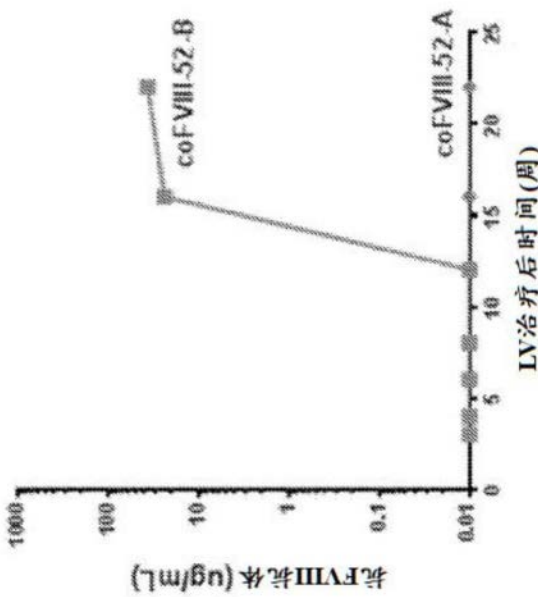


图16B

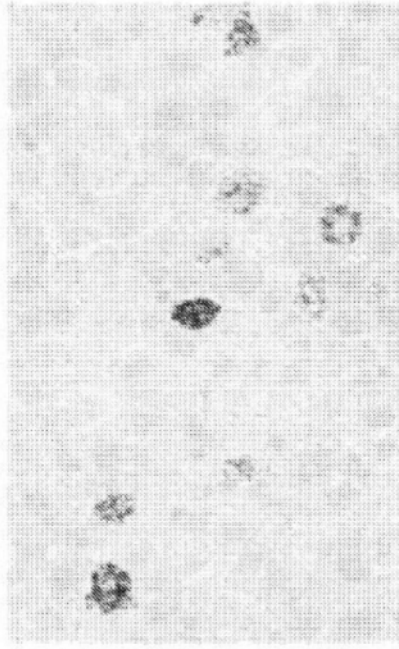


图16C

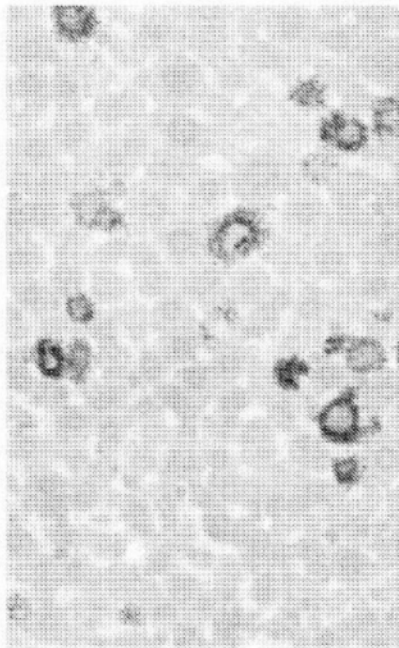


图16D

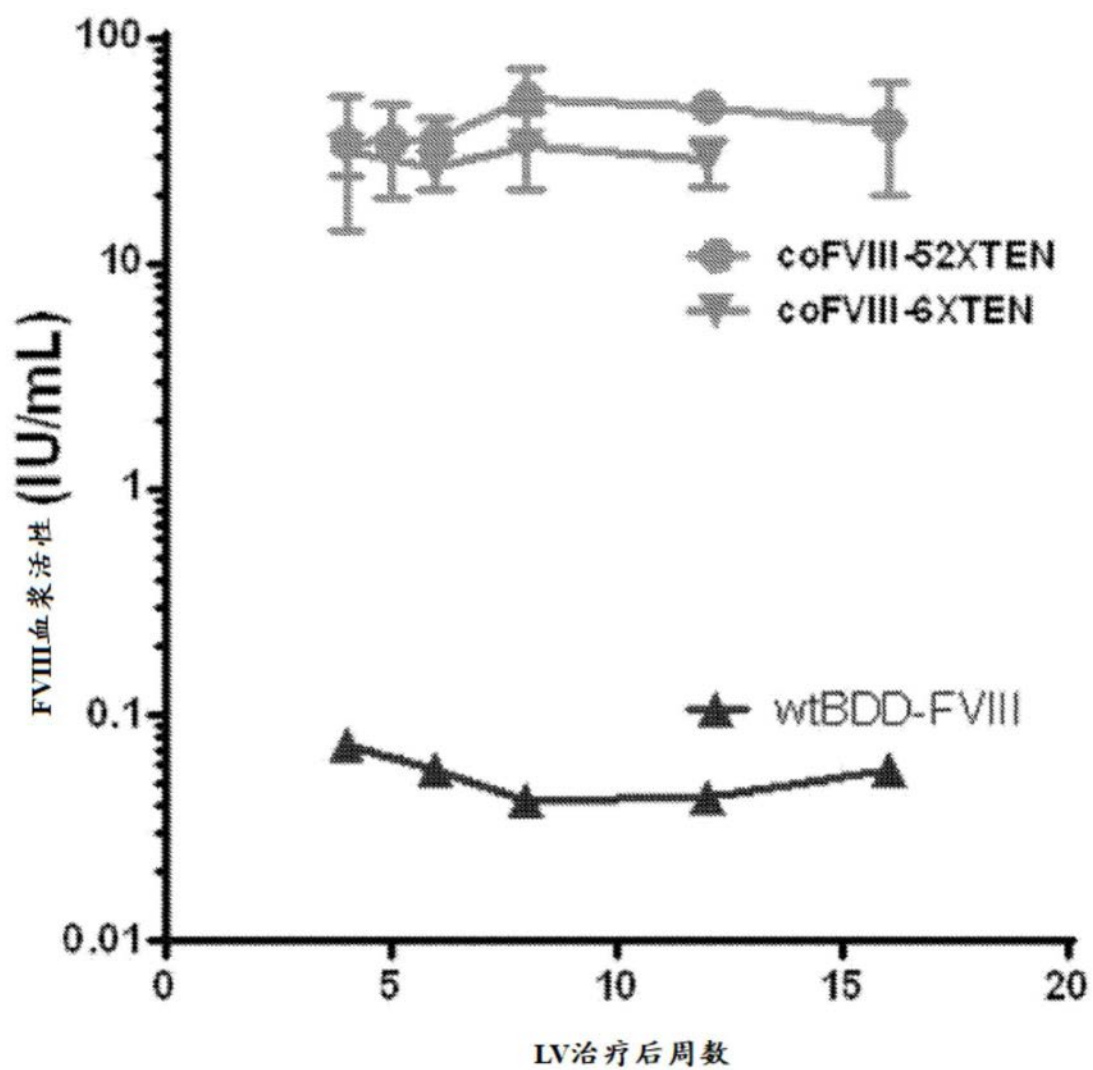


图17

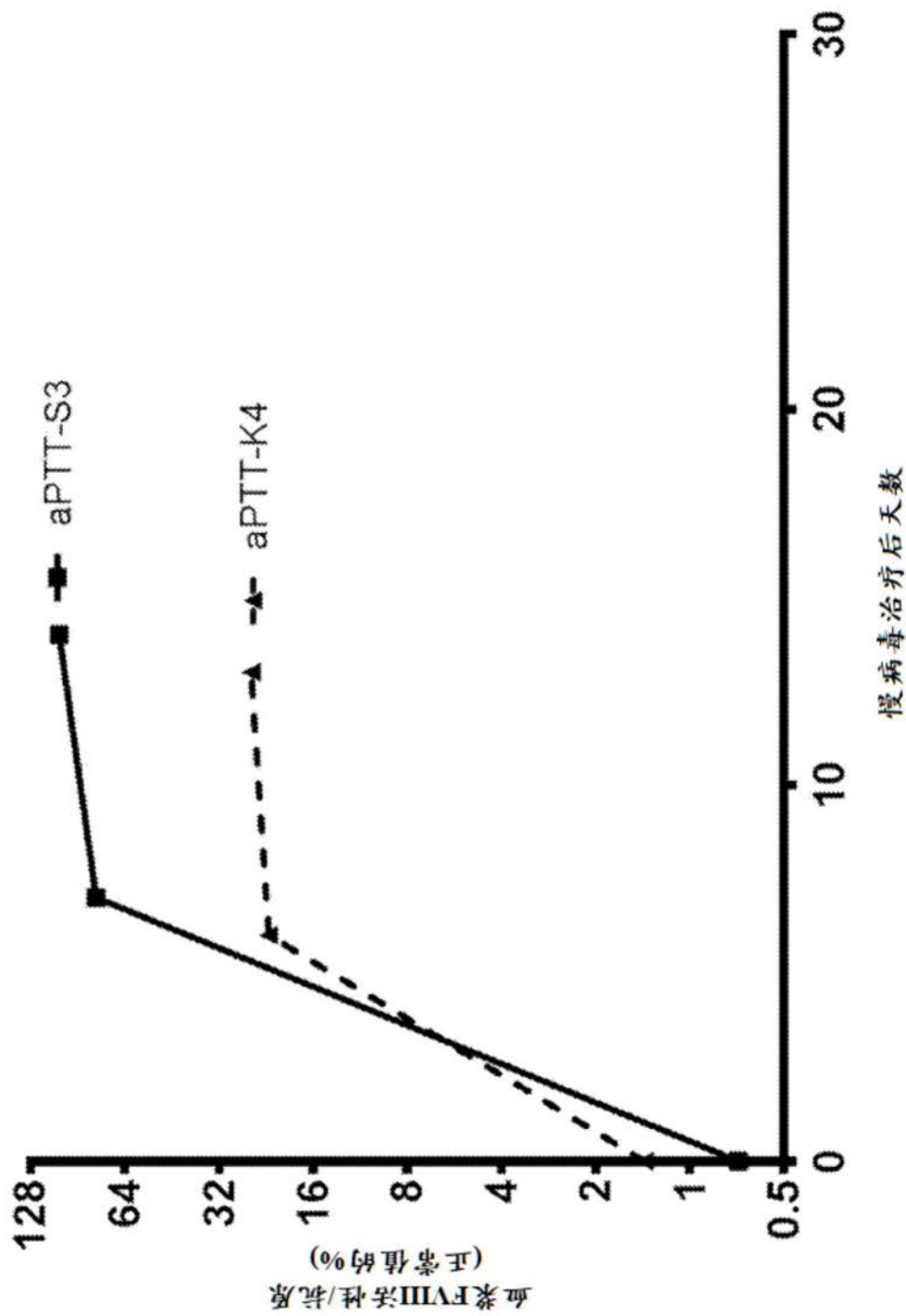


图18

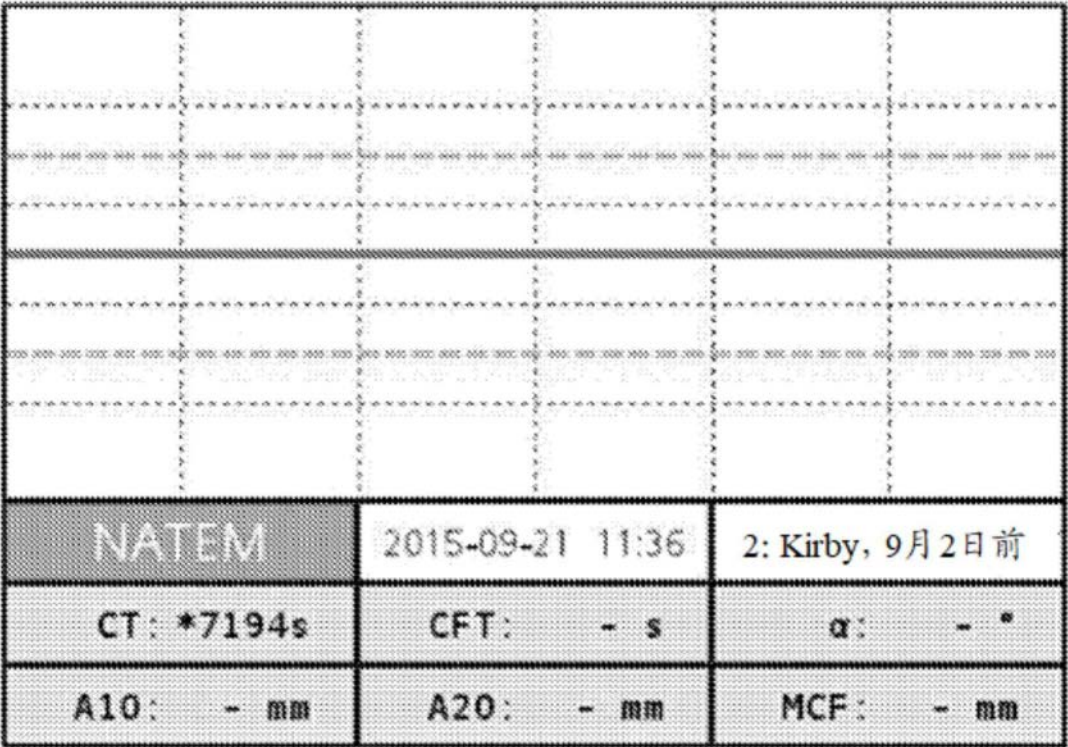


图19A

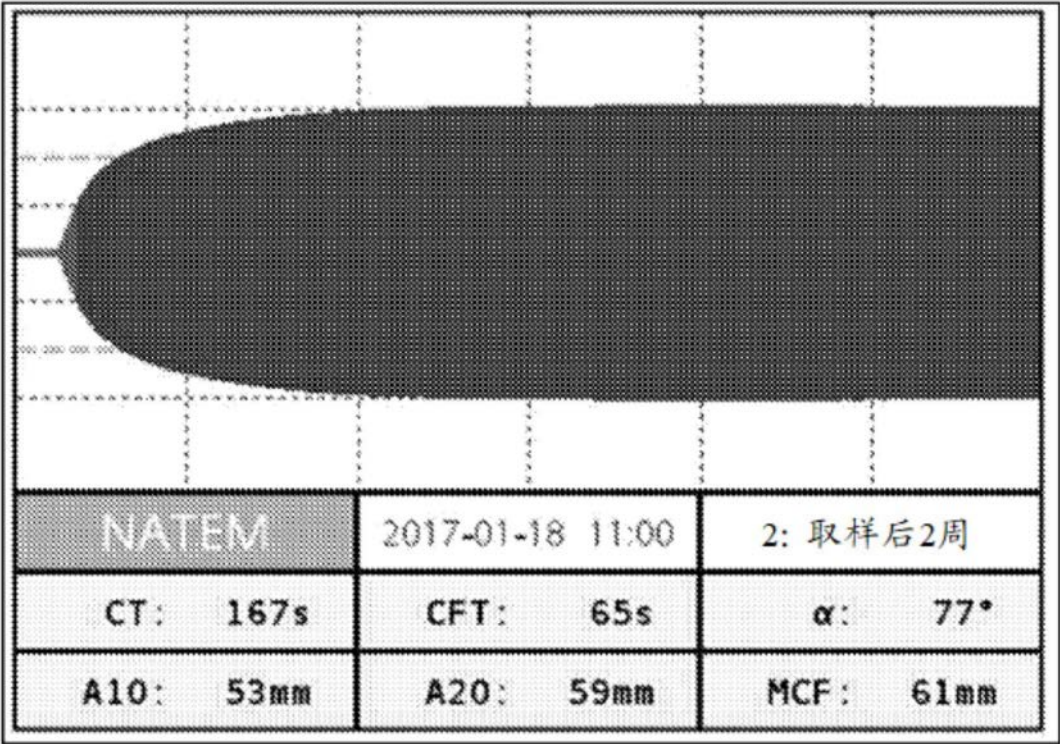


图19B

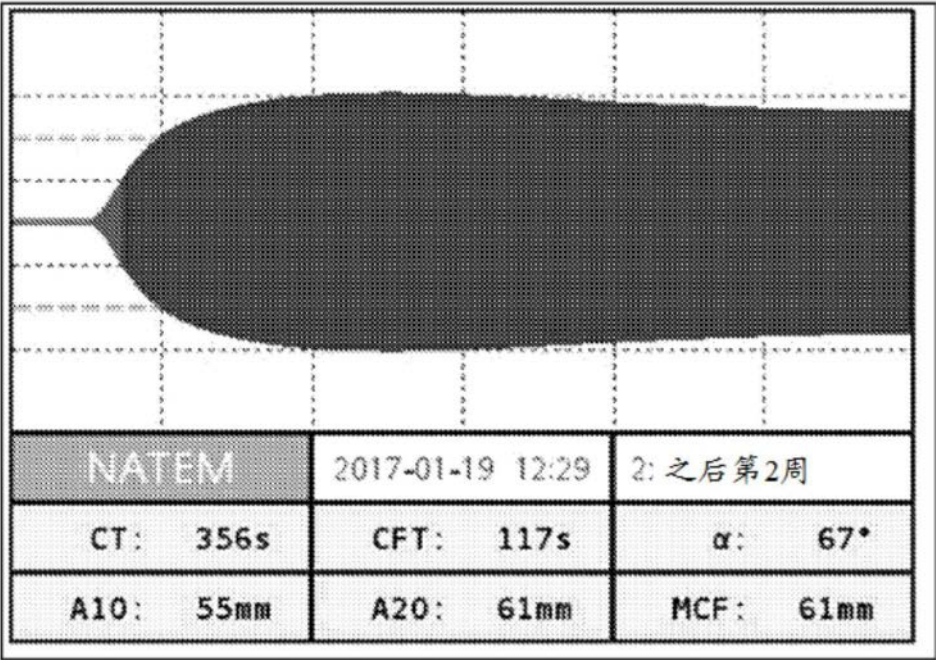


图19C

参数	参数描述	正常范围
CT (秒)	凝固时间	120-480
CFT (秒)	凝块形成时间	60-240
$\alpha(^{\circ})$	α 角	27-60
A5 (mm)	CT后5分钟的振幅	na
A20 (mm)	CT后20分钟的振幅	na
MCF (mm)	最大凝块坚度	40-60

图19D