



# (12) 发明专利申请

(10) 申请公布号 CN 116731201 A

(43) 申请公布日 2023. 09. 12

(21) 申请号 202310334224.5

C12N 5/10 (2006.01)

(22) 申请日 2015.01.09

A61K 38/37 (2006.01)

(30) 优先权数据

A61K 47/68 (2017.01)

61/926,226 2014.01.10 US

A61P 7/04 (2006.01)

61/988,104 2014.05.02 US

(62) 分案原申请数据

201580009942.9 2015.01.09

(71) 申请人 比奥贝拉蒂治疗公司

地址 美国马萨诸塞州

(72) 发明人 埃克塔·赛斯·恰布拉 刘同耀

(74) 专利代理机构 北京坤瑞律师事务所 11494

专利代理师 岑晓东

(51) Int. Cl.

C07K 19/00 (2006.01)

权利要求书2页 说明书165页

C12N 15/62 (2006.01)

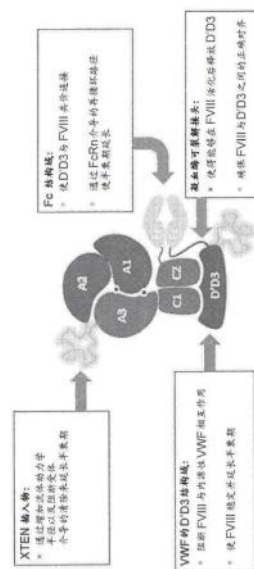
序列表(电子公布) 附图14页

(54) 发明名称

因子VIII嵌合蛋白及其用途

(57) 摘要

本发明涉及因子VIII嵌合蛋白及其用途。本发明提供一种嵌合蛋白,所述嵌合蛋白包含第一多肽,所述第一多肽包含FVIII蛋白和第一Ig恒定区或其一部分;和第二多肽,所述第二多肽包含含有VWF的D'结构域和D3结构域的VWF蛋白、在长度方面具有小于288个氨基酸的XTEN序列以及第二Ig恒定区或其一部分,其中所述第一多肽和所述第二多肽彼此缔合。本发明也包括核苷酸、载体、宿主细胞、使用嵌合蛋白的方法。



1. 一种嵌合蛋白,其包含下式(a) - (hh)中的每一个:

- (a) FVIII-F1:F2-L2-X-L1-V;
  - (b) FVIII-F1:V-L1-X-L2-F2;
  - (c) F1-FVIII:F2-L2-X-L1-V;
  - (d) F1-FVIII:V-L1-X-L2-F2;
  - (e) FVIII-X2-F1:F2-L2-X1-L1-V;
  - (f) FVIII-X2-F1:V-L1-X1-L2-F2;
  - (g) FVIII (X2) -F1:F2-L2-X1-L1-V;
  - (h) FVIII (X2) -F1:V-L1-X1-L2-F2;
  - (i) F1-X2-F1:F2-L2-X1-L1-V;
  - (j) F1-X2-F1:V-L1-X1-L2-F2;
  - (k) V-L1-X-L2-F2-L3-FVIII-L4-F1;
  - (l) V-L1-X-L2-F2-L3-F1-L4-FVIII;
  - (m) F1-L4-FVIII-L3-F2-L2-X-L1-V;
  - (n) FVIII-L4-F1-L3-F2-L2-X-L1-V;
  - (o) FVIII-L4-F1-L3-V-L1-X-L2-F2;
  - (p) FVIII-L4-F1-L3-F2-L2-X-L1-V;
  - (q) F2-L2-X-L1-V-L3-F1-L4-FVIII;
  - (r) F2-L2-X-L1-V-L3-FVIII-L4-F1;
  - (s) V-L1-X1-L2-F2-L3-FVIII (X2) -L4-F1;
  - (t) V-L1-X1-L2-F2-L3-F1-L4-FVIII (X2) ;
  - (u) F1-L4-FVIII (X2) -L3-F2-L2-X1-L1-V;
  - (v) F-L4-FVIII (X2) -L3-V-L1-X1-L2-F2;
  - (w) FVIII (X2) -L4-F1-L3-V-L1-X1-L2-F2;
  - (x) FVIII (X2) -L4-F1-L3-F2-L2-X1-L1-V;
  - (y) F2-L2-X1-L1-V-L3-F1-L4-FVIII (X2) ;
  - (z) F2-L2-X1-L1-V-L3-FVIII (X2) -L4-F1;
  - (aa) V-L1-X2-L2-F2-L3-FVIII-L4-X2-L5-F1;
  - (bb) V-L1-X2-L2-F2-L3-F1-L5-X2-L4-FVIII;
  - (cc) F1-L5-X2-L4-FVIII-L3-F2-L2-X2-L1-V;
  - (dd) F1-L5-X2-L4-FVIII-L3-V-L1-X2-L2-F2;
  - (ee) FVIII-L5-X2-L4-F2-L3-V-L1-X1-L2-F1;
  - (ff) FVIII-L5-X2-L4-F2-L3-F1-L2-X1-L1-V;
  - (gg) F1-L2-X1-L1-V-L3-F2-L4-X2-L5-FVIII;或
  - (hh) F1-L2-X1-L1-V-L3-FVIII-L5-X2-L4-F2;
- 其中V是VWF蛋白,其包含D'结构域和D3结构域,  
X或X1是含有小于288个氨基酸的第一XTEN序列,  
X2是第二XTEN序列,  
FVIII包括FVIII蛋白,

FVIII (X2) 包括具有插入在FVIII蛋白内一个或多个插入位点中的第二XTEN序列的所述FVIII蛋白，

F1是第一Ig恒定区或其一部分，

F2是第二Ig恒定区或其一部分，

L1、L2、L3、L4或L5是任选接头，

(-) 是肽键；并且

(:) 是共价键或非共价键。

2. 一种多核苷酸或一组多核苷酸，其编码如权利要求1所述的嵌合蛋白。

3. 一种载体，其包含如权利要求2所述的多核苷酸和一种或多种可操作地连接于所述多核苷酸或所述一组多核苷酸的启动子。

4. 一种宿主细胞，其包含如权利要求2所述的多核苷酸或如权利要求3所述的载体。

5. 一种药物组合物，其包含如权利要求1所述的嵌合蛋白、如权利要求2所述的多核苷酸、如权利要求3所述的载体或如权利要求4所述的宿主细胞以及药学上可接受的载体。

6. 一种延长或增加嵌合蛋白的半衰期的方法，其中所述方法包括向有需要的受试者添加有效量的如权利要求1所述的嵌合蛋白、如权利要求2所述的多核苷酸、如权利要求3所述的载体、如权利要求4所述的宿主细胞或如权利要求5所述的组合物，其中所述VWF蛋白、所述XTEN序列、所述第一Ig恒定区或其一部分和所述第二Ig恒定区或其一部分增加所述嵌合蛋白的所述半衰期。

7. 一种治疗有需要的受试者的流血疾病或病症的方法，其包括施用有效量的如权利要求1所述的嵌合蛋白、如权利要求2所述的多核苷酸、如权利要求3所述的载体、如权利要求4所述的宿主细胞或如权利要求5所述的组合物，其中所述流血疾病或病症选自由以下组成的组：流血凝血病症、关节积血、肌肉流血、口腔流血、出血、向肌肉中出血、口腔出血、创伤、头创伤、胃肠流血、颅内出血、腹内出血、胸内出血、骨折、中枢神经系统流血、咽后隙中流血、腹膜后隙中流血和髂腰肌鞘中流血。

8. 一种制备嵌合蛋白的方法，其包括用如权利要求2所述的多核苷酸或如权利要求3所述的载体转染一种或多种宿主细胞，以及在所述宿主细胞中表达所述嵌合蛋白。

## 因子VIII嵌合蛋白及其用途

[0001] 本申请是申请日为2015年01月09日、中国申请号为201580009942.9、发明名称为“因子VIII嵌合蛋白及其用途”的发明申请的分案申请。

[0002] 对电子提交的序列表的引用

[0003] 以ASCII文本文件(名称:2159\_441PC02\_SequenceListing\_ST25.txt;大小:823,500字节;以及创建日期:2015年1月9日)形式电子提交的序列表的内容以引用的方式整体并入本文。

[0004] 发明背景

[0005] 甲型血友病(Haemophilia A)是一种由编码凝血因子VIII(FVIII)的基因的缺陷引起的流血病症,并且影响万分之一至万分之二的男性出生者。Graw等,Nat.Rev.Genet.6(6):488-501(2005)。受甲型血友病影响的患者可以输注纯化或重组产生的FVIII加以治疗。然而,已知所有可商购获得的FVIII产品的半衰期都是约8-12小时,从而要求频繁向患者进行静脉内施用。参见Weiner M.A.和Cairo,M.S.,Pediatric Hematology Secrets, Lee,M.T.,12.Disorders of Coagulation,Elsevier Health Sciences,2001;Lillicrap, D.Thromb.Res.122增刊4:S2-8(2008)。此外,已尝试许多方法来延长FVIII半衰期。举例来说,用以延长凝结因子的半衰期的处于开发中的方法包括聚乙二醇化、糖聚乙二醇化和与白蛋白缀合。参见Dumont等,Blood.119(13):3024-3030(2012年1月13日在线发表)。然而,无论所用蛋白质工程改造如何,当前在开发中的长效FVIII产品据报道具有有限半衰期-在临床前动物模型中仅约1.5至2小时。参见同上。已在人中证明一致结果,例如据报道相较于ADVATE®,rFVIIIIFc在甲型血友病患者中的半衰期改进多达约1.7倍。参见同上。因此,尽管存在微小改进,但半衰期增加可指示存在其它T<sub>1/2</sub>限制因素。参见Liu,T.等,2007ISTH会议,摘要#P-M-035;Henrik,A.等,2011ISTH会议,摘要#P=MO-181;Liu,T.等,2011ISTH会议摘要#P-WE-131。

[0006] 血浆范威尔邦德因子(von Willebrand Factor,VWF)的半衰期是约16小时(在13至18小时的范围内)。Goudemand J等J Thromb Haemost 2005;3:2219-27。VWF半衰期可受以下许多因素影响:糖基化样式、ADAMTS-13(具有血栓反应素(thrombospondin)基序-13的解整合素(disintegrin)和金属蛋白酶)和VWF中的各种突变。

[0007] 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。这个复合物的形成对于维持体内FVIII的适当血浆水平是重要的。Lenting等,Blood.92(11):3983-96(1998);Lenting等,J.Thromb.Haemost.5(7):1353-60(2007)。全长野生型FVIII主要以具有重链(MW 200kD)和轻链(MW 73kD)的异二聚体形式存在。当FVIII由于在重链中的位置372和740处以及在轻链中的位置1689处的蛋白水解而活化时,结合于FVIII的VWF被从活化FVIII移除。活化FVIII连同活化因子IX、钙和磷脂一起(“因子X活化酶复合物(tenase complex)”)诱导因子X活化,从而产生大量凝血酶。凝血酶转而接着裂解纤维蛋白原以形成可溶性纤维蛋白单体,所述单体接着自发聚合以形成可溶性纤维蛋白聚合物。凝血酶也活化因子XIII,因子XIII连同钙一起用于交联和稳定化可溶性纤维蛋白聚合物,从而形成交联(不溶性)纤维蛋白。活化FVIII通过蛋白水解从循环快速清除。

[0008] 归因于频繁给药和由给药时程引起的不便,所以仍然需要开发要求较小施用频率的FVIII产品,即半衰期长于半衰期限制1.5至2倍的FVIII产品。

[0009] 发明简述

[0010] 本发明提供一种嵌合蛋白,其包含(i)第一多肽,所述第一多肽包含融合于第一免疫球蛋白(“Ig”)恒定区或其一部分的因子VIII(“FVIII”)蛋白,和(ii)第二多肽,所述第二多肽包含含有VWF的D’结构域和D3结构域的范威尔邦德因子(“VWF”)蛋白,所述VWF蛋白通过在之间的XTEN序列融合于第二Ig恒定区或其一部分,其中所述XTEN序列含有小于288个氨基酸残基,并且其中所述第一多肽与所述第二多肽连接或缔合。某些实施方案包括如本文所述的嵌合蛋白,其中第二多肽中的XTEN序列由长度在12个氨基酸与287个氨基酸之间的氨基酸序列组成。

[0011] 也公开如本文所述的嵌合蛋白,其中相较于包含第一多肽和第二多肽的相应融合蛋白,其中所述融合蛋白的第二多肽包含含有至少288个氨基酸的XTEN序列,所述嵌合蛋白展现较长半衰期。一些实施方案包括含有至少288个氨基酸的XTEN序列AE288。在一些实施方案中,AE288是SEQ ID NO:8。

[0012] 也公开如本文所述的嵌合蛋白,其中第二多肽的XTEN序列含有约36、约42、约72或约144个氨基酸。在一些实施方案中,第二多肽的XTEN序列选自AE42、AE72、AE144、AG42、AG72或AG144。

[0013] 一些实施方案包括如本文所述的嵌合蛋白,其中第二多肽的XTEN序列选自SEQ ID NO:9;SEQ ID NO:10;SEQ ID NO:11;SEQ ID NO:55;SEQ ID NO:56;SEQ ID NO:57;SEQ ID NO:58;SEQ ID NO:58;SEQ ID NO:59;SEQ ID NO:14;SEQ ID NO:60;SEQ ID NO:61;SEQ ID NO:62;或SEQ ID NO:63。

[0014] 在某些实施方案中,第一多肽进一步包含使FVIII蛋白与第一Ig恒定区或其一部分连接的第二XTEN序列。也公开如本文所述的嵌合蛋白,其中第一多肽包含插入在FVIII蛋白内一个或多个插入位点处的第三XTEN序列。在一些实施方案中,第一多肽进一步包含插入在FVIII蛋白内一个或多个插入位点处的第二XTEN序列。在某些实施方案中,第一多肽包含使FVIII蛋白与第一Ig恒定区或其一部分连接的第三XTEN序列。

[0015] 也公开如本文所述的嵌合蛋白,其中第二XTEN序列、第三XTEN序列、或第二XTEN序列和第三XTEN序列各自独立地选自AE42、AE72、AE864、AE576、AE288、AE144、AG864、AG576、AG288和AG144。在一些实施方案中,第二XTEN序列、第三XTEN序列、或第二XTEN序列和第三XTEN序列各自独立地选自SEQ ID NO:8;SEQ ID NO:9;SEQ ID NO:10;SEQ ID NO:11;SEQ ID NO:17;SEQ ID NO:54;SEQ ID NO:19;SEQ ID NO:16;SEQ ID NO:18;SEQ ID NO:15;SEQ ID NO:55;SEQ ID NO:56;SEQ ID NO:57;SEQ ID NO:58;SEQ ID NO:59;SEQ ID NO:14;SEQ ID NO:60;SEQ ID NO:61;SEQ ID NO:62;或SEQ ID NO:63。在某些实施方案中,第二XTEN序列、第三XTEN序列、或第二XTEN序列与第三XTEN序列两者各自独立地是AE288或AG288。在一些实施方案中,通过接头使第二多肽中的XTEN序列融合于第二Ig恒定区或其一部分。在某些实施方案中,接头是可裂解接头。

[0016] 一些实施方案包括如本文所述的嵌合蛋白,其中接头可由选自以下的蛋白酶裂解:因子XIa、因子XIIa、胰舒血管素(kallikrein)、因子VIIa、因子IXa、因子Xa、因子IIa(凝血酶)、弹性蛋白酶-2(Elastase-2)、粒酶-B(Granzyme-B)、TEV、肠激酶(Enterokinase)、蛋

白酶3C、分选酶A(Sortase A)、MMP-12、MMP-13、MMP-17和MMP-20。在一些实施方案中,接头可由因子IIa(凝血酶)裂解。

[0017] 也公开如本文所述的嵌合蛋白,其中接头包含一个或多个包含选自以下的氨基酸序列的裂解位点:RRRR(SEQ ID NO:102)、RKRRKR(SEQ ID NO:103)、RRRRS(SEQ ID NO:104)、TQSFNDFTR(SEQ ID NO:1)、SVSQTSLKTR(SEQ ID NO:3)、DFLAEGGGVR(SEQ ID NO:4)、TTKIKPR(SEQ ID NO:5)、LVPRG(SEQ ID NO:6)、ALRPR(SEQ ID NO:7)、KLTRAET(SEQ ID NO:121)、DFTRVVG(SEQ ID NO:122)、TMTRIVGG(SEQ ID NO:123)、SPFRSTGG(SEQ ID NO:124)、LQVRIVGG(SEQ ID NO:125)、PLGRIVGG(SEQ ID NO:126)、IEGRTVGG(SEQ ID NO:127)、LTPRSLLV(SEQ ID NO:128)、LGPVSGVP(SEQ ID NO:129)、VAGDSLEE(SEQ ID NO:130)、GPAGLGGA(SEQ ID NO:131)、GPAGLRGA(SEQ ID NO:132)、APLGLRLR(SEQ ID NO:133)、PALPLVAQ(SEQ ID NO:134)、ENLYFQG(SEQ ID NO:135)、DDDKIVGG(SEQ ID NO:136)、LEVLFGQP(SEQ ID NO:137)、LPKTGSES(SEQ ID NO:138)、DKNTGDYEDSYEDISAYLLSKNNAIEPRRSFS(SEQ ID NO:88)和IEPRSFS(SEQ ID NO:194)。在一些实施方案中,接头包含TLDPRSFLLRNPNDKYEPFWEDEEK(SEQ ID NO:146)。在某些实施方案中,裂解位点包含氨基酸序列LVPRG(SEQ ID NO:6)。在其它实施方案中,裂解位点包含氨基酸序列IEPRSFS(SEQ ID NO:194)。在其它实施方案中,裂解位点包含氨基酸序列IEPRSFS(SEQ ID NO:194),其中所述裂解位点不是FVIII的全长a2区。在一些实施方案中,裂解位点包括FVIII的a2区的包含至少序列IEPR(SEQ ID NO:200)的片段。在其它实施方案中,裂解位点包括FVIII的a2区的包含至少序列IEPR(SEQ ID NO:200)的片段,其中所述裂解位点不是全长a2区。在某些实施方案中,在如本文提供或如本领域中已知的凝血酶裂解测定中,裂解位点是可裂解的。

[0018] 一些实施方案包括如本文所述的嵌合蛋白,其中所述第一Ig恒定区或其一部分包含第一Fc区和/或第二Ig恒定区或其一部分包含第二Fc区。在一些实施方案中,第一Ig恒定区或其一部分和第二Ig恒定区或其一部分使嵌合蛋白的半衰期延长。在一些实施方案中,第一多肽和第二多肽通过接头融合。在某些实施方案中,第一多肽和第二多肽通过可加工接头融合。在一些实施方案中,第一Ig恒定区或其一部分与第二Ig恒定区或其一部分缔合。在某些实施方案中,第一Ig恒定区或其一部分通过共价键与第二Ig恒定区或其一部分缔合。在一些实施方案中,共价键是二硫键。

[0019] 也公开包含下式(a) - (hh)中的每一个的嵌合蛋白:

[0020] (a) FVIII-F1:F2-L2-X-L1-V;

[0021] (b) FVIII-F1:V-L1-X-L2-F2;

[0022] (c) F1-FVIII:F2-L2-X-L1-V;

[0023] (d) F1-FVIII:V-L1-X-L2-F2;

[0024] (e) FVIII-X2-F1:F2-L2-X1-L1-V;

[0025] (f) FVIII-X2-F1:V-L1-X1-L2-F2;

[0026] (g) FVIII(X2)-F1:F2-L2-X1-L1-V;

[0027] (h) FVIII(X2)-F1:V-L1-X1-L2-F2;

[0028] (i) F1-X2-F1:F2-L2-X1-L1-V;

[0029] (j) F1-X2-F1:V-L1-X1-L2-F2;

[0030] (k) V-L1-X-L2-F2-L3-FVIII-L4-F1;

- [0031] (l)V-L1-X-L2-F2-L3-F1-L4-FVIII;
- [0032] (m)F1-L4-FVIII-L3-F2-L2-X-L1-V;
- [0033] (n)FVIII-L4-F1-L3-F2-L2-X-L1-V;
- [0034] (o)FVIII-L4-F1-L3-V-L1-X-L2-F2;
- [0035] (p)FVIII-L4-F1-L3-F2-L2-X-L1-V;
- [0036] (q)F2-L2-X-L1-V-L3-F1-L4-FVIII;
- [0037] (r)F2-L2-X-L1-V-L3-FVIII-L4-F1;
- [0038] (s)V-L1-X1-L2-F2-L3-FVIII(X2)-L4-F1;
- [0039] (t)V-L1-X1-L2-F2-L3-F1-L4-FVIII(X2);
- [0040] (u)F1-L4-FVIII(X2)-L3-F2-L2-X1-L1-V;
- [0041] (v)F-L4-FVIII(X2)-L3-V-L1-X1-L2-F2;
- [0042] (w)FVIII(X2)-L4-F1-L3-V-L1-X1-L2-F2;
- [0043] (x)FVIII(X2)-L4-F1-L3-F2-L2-X1-L1-V;
- [0044] (y)F2-L2-X1-L1-V-L3-F1-L4-FVIII(X2);
- [0045] (z)F2-L2-X1-L1-V-L3-FVIII(X2)-L4-F1;
- [0046] (aa)V-L1-X2-L2-F2-L3-FVIII-L4-X2-L5-F1;
- [0047] (bb)V-L1-X2-L2-F2-L3-F1-L5-X2-L4-FVIII;
- [0048] (cc)F1-L5-X2-L4-FVIII-L3-F2-L2-X2-L1-V;
- [0049] (dd)F1-L5-X2-L4-FVIII-L3-V-L1-X2-L2-F2;
- [0050] (ee)FVIII-L5-X2-L4-F2-L3-V-L1-X1-L2-F1;
- [0051] (ff)FVIII-L5-X2-L4-F2-L3-F1-L2-X1-L1-V;
- [0052] (gg)F1-L2-X1-L1-V-L3-F2-L4-X2-L5-FVIII;或
- [0053] (hh)F1-L2-X1-L1-V-L3-FVIII-L5-X2-L4-F2;

[0054] 其中V是VWF蛋白,其包含D'结构域和D3结构域,X或X1是含有小于288个氨基酸的第一XTEN序列,X2是第二XTEN序列,FVIII包括FVIII蛋白,FVIII(X2)包括具有插入在FVIII蛋白内一个或多个插入位点中的第二XTEN序列的FVIII蛋白,F1是第一Ig恒定区或其一部分,F2是第二Ig恒定区或其一部分,L1、L2、L3、L4或L5是任选接头,(-)是肽键;并且(:)是共价键或非共价键。

[0055] 一些实施方案包括如本文所述的嵌合蛋白,其中X或X1由长度在12个氨基酸与287个氨基酸之间的氨基酸序列组成。

[0056] 在某些实施方案中,相较于包含所述式,例外之处是X或X1是AE288的相应嵌合蛋白,如本文所述的嵌合蛋白展现较长半衰期。在一些实施方案中,AE288是SEQ ID NO:8。

[0057] 一些实施方案包括如本文所述的嵌合蛋白,其中式X或X1含有约36、约42、约72或约144个氨基酸。在某些实施方案中,式X或X1选自AE42、AE72、AE144、AG42、AG72或AG144。在一些实施方案中,式X或X1选自SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:55;SEQ ID NO:56;SEQ ID NO:57;SEQ ID NO:58;SEQ ID NO:59;SEQ ID NO:14;SEQ ID NO:60;SEQ ID NO:61;SEQ ID NO:62;或SEQ ID NO:63。在某些实施方案中,X2包含长度是至少约36个氨基酸、至少约42个氨基酸、至少约144个氨基酸、至少约288个氨基酸、至少约576个氨基酸、至少约864个氨基酸的氨基酸序列。在某些实施方案中,X2选自AE42、AE72、

AE864、AE576、AE288、AE144、AG864、AG576、AG288和AG144。在一些实施方案中，X2选自SEQ ID NO:8;SEQ ID NO:9;SEQ ID NO:10;SEQ ID NO:11;SEQ ID NO:17;SEQ ID NO:54;SEQ ID NO:19;SEQ ID NO:16;SEQ ID NO:18;SEQ ID NO:15;SEQ ID NO:55;SEQ ID NO:56;SEQ ID NO:57;SEQ ID NO:58;SEQ ID NO:59;SEQ ID NO:14;SEQ ID NO:60;SEQ ID NO:61;SEQ ID NO:62;或SEQ ID NO:63。在某些实施方案中，X2是AE288或AG288。

[0058] 也公开如本文所述的包含X或X1和/或X2的嵌合蛋白，其相较于不包含X或X1和/或X2的嵌合蛋白展现较长半衰期。在一些实施方案中，L1和/或L2是可裂解接头。在某些实施方案中，L4和/或L5是可裂解接头。在某些实施方案中，接头可由选自以下的蛋白酶裂解：因子XIa、因子XIIa、胰舒血管素、因子VIIa、因子IXa、因子Xa、因子IIa(凝血酶)、弹性蛋白酶-2、粒酶-B、TEV、肠激酶、蛋白酶3C、分选酶A、MMP-12、MMP-13、MMP-17和MMP-20。在一些实施方案中，接头可由因子IIa(凝血酶)裂解。

[0059] 一些实施方案包括如本文所述的嵌合蛋白，其中接头包含一个或多个包含选自以下的氨基酸序列的裂解位点：RRRR(SEQ ID NO:102)、RKRRKR(SEQ ID NO:103)、RRRRS(SEQ ID NO:104)、TQSFNDFTR(SEQ ID NO:2)、SVSQTSLKTR(SEQ ID NO:3)、DFLAEGGGVR(SEQ ID NO:4)、TTKIKPR(SEQ ID NO:5)、LVPRG(SEQ ID NO:6)、ALRPR(SEQ ID NO:7)、KLTRAET(SEQ ID NO:121)、DFTRVVG(SEQ ID NO:122)、TMTRIVGG(SEQ ID NO:123)、SPFRSTGG(SEQ ID NO:124)、LQVRIVGG(SEQ ID NO:125)、PLGRIVGG(SEQ ID NO:126)、IEGRVGG(SEQ ID NO:127)、LTPRSLLV(SEQ ID NO:128)、LGPVSGVP(SEQ ID NO:129)、VAGDSLEE(SEQ ID NO:130)、GPAGLGGA(SEQ ID NO:131)、GPAGLRGA(SEQ ID NO:132)、APLGLRLR(SEQ ID NO:133)、PALPLVAQ(SEQ ID NO:134)、ENLYFQG(SEQ ID NO:135)、DDDKIVGG(SEQ ID NO:136)、LEVLFFQGP(SEQ ID NO:137)和LPKTGSES(SEQ ID NO:138)。在一些实施方案中，接头包含TLDPRSFLLRNPNDKYEPFWEDEEK(SEQ ID NO:146)。在某些实施方案中，接头包含氨基酸序列LVPRG(SEQ ID NO:6)。在一些实施方案中，接头包括FVIII的a1区、FVIII的a2区、FVIII的a3区或其任何组合。在某些实施方案中，接头包括FVIII的a2区的片段。在一些情况下，a2区的片段可包含序列DKNTGDYEDSYEDISAYLLSKNNAIEPRSFS(SEQ ID NO:88)。在其它实施方案中，可使用FVIII的a2区的较小片段，包括具有序列IEPRSFS(SEQ ID NO:194)的片段。在一个特定实施方案中，接头包含氨基酸序列IEPRSFS(SEQ ID NO:194)。在另一实施方案中，接头包含氨基酸序列IEPRSFS(SEQ ID NO:194)，其中接头不是FVIII的全长a2区。

[0060] 也公开如本文所述的嵌合蛋白，其中FVIII的a2区包含与ISDKNTGDYEDSYEDISAYLLSKNNAIEPRSFS(SEQ ID NO:106)或DKNTGDYEDSYEDISAYLLSKNNAIEPRSFS(SEQ ID NO:88)至少约80%、约85%、约90%、约95%或100%同一的氨基酸序列。在一些实施方案中，a1区包含与ISMKNNEEAEDYDDDLTDSEMDVVRFDNNSPSFIQIRSV(SEQ ID NO:107)至少约80%、约85%、约90%、约95%或100%同一的氨基酸序列。在某些实施方案中，a3区包含与ISEITRTLQSDQEEIDYDDTISVEMKKEDFDIYDEDENQSPRSFQ(SEQ ID NO:108)至少约80%、约85%、约90%、约95%或100%同一的氨基酸序列。在一些实施方案中，F1包含第一Fc区和/或F2包含第二Fc区。

[0061] 一些实施方案包括如本文所述的嵌合蛋白，其中相较于不包含F1和F2的嵌合蛋白，包含F1和F2的嵌合蛋白展现较长半衰期。在某些实施方案中，L3是可加工接头。在一些实施方案中，VWF蛋白通过非共价键与FVIII蛋白缔合。在一些实施方案中，相较于无VWF蛋



白和/或XTEN序列的FVIII蛋白或相较于野生型FVIII,嵌合蛋白的半衰期被延长。在某些实施方案中,嵌合蛋白的半衰期比无VWF蛋白或XTEN序列的FVIII蛋白或比野生型FVIII延长至少约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。

[0062] 也公开如本文所述的嵌合蛋白,其中嵌合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。在一些实施方案中,嵌合蛋白在HemA小鼠中的半衰期是约40小时。在某些实施方案中,VWF蛋白不实质上结合VWF清除受体。在一些实施方案中,VWF蛋白能够保护FVIII蛋白免遭一种或多种蛋白酶裂解,保护FVIII蛋白免遭活化,使FVIII蛋白的重链和/或轻链稳定,或防止FVIII蛋白被一种或多种清道夫受体清除。

[0063] 一些实施方案包括如本文所述的嵌合蛋白,其中VWF蛋白通过遮蔽或阻断FVIII蛋白上的VWF结合位点来抑制或防止内源性VWF结合FVIII蛋白。在某些实施方案中,VWF结合位点位于FVIII蛋白的A3结构域或C2结构域、或A3结构域与C2结构域两者中。在一些实施方案中,VWF结合位点包含对应于SEQ ID NO:65的氨基酸1669至1689以及2303至2332的氨基酸序列。在一些实施方案中,第一Ig恒定区或其一部分和第二Ig恒定区或其一部分是相同的或不同的。在某些实施方案中,FVIII蛋白连接于和/或插入有至少两个XTEN序列、至少三个XTEN序列、至少四个XTEN序列、至少五个XTEN序列或至少六个XTEN序列。

[0064] 也公开如本文所述的嵌合蛋白,其中FVIII蛋白包含FVIII的一个或多个选自以下的结构域:A1结构域、a1酸性区、A2结构域、a2酸性区、B结构域、A3结构域、a3酸性区、C1结构域、C2结构域、其一个或多个片段及其任何组合。

[0065] 也公开如本文所述的嵌合蛋白,其中FVIII蛋白中的一个或多个插入位点位于FVIII蛋白的一个或多个选自以下的结构域内:A1结构域、a1酸性区、A2结构域、a2酸性区、A3结构域、B结构域、C1结构域、C2结构域及其任何组合,或位于FVIII蛋白的一个或多个选自由以下组成的组的结构域之间:A1结构域和a1酸性区、a1酸性区和A2结构域、A2结构域和a2酸性区、a2酸性区和B结构域、B结构域和A3结构域、A3结构域和C1结构域、C1结构域和C2结构域及其任何组合,或位于FVIII蛋白的选自以下的两个结构域之间:A1结构域和a1酸性区、a1酸性区和A2结构域、A2结构域和a2酸性区、a2酸性区和B结构域、B结构域和A3结构域、A3结构域和C1结构域、C1结构域和C2结构域及其任何组合。在一些实施方案中,FVIII蛋白中的一个或多个插入位点是一个或多个选自由表7、表8、表9和表10中的氨基酸残基组成的组的氨基酸。在某些实施方案中,FVIII蛋白中的插入位点位于紧靠对应于成熟FVIII蛋白(SEQ ID NO:65)的氨基酸745的下游。在一些实施方案中,FVIII蛋白中的插入位点位于紧靠对应于成熟FVIII蛋白(SEQ ID NO:65)的残基1656和残基1900的下游。在一些实施方案中,FVIII蛋白中的插入位点紧靠对应于成熟FVIII蛋白(SEQ ID NO:65)的残基26、1656和1900的下游。在某些实施方案中,FVIII蛋白中的插入位点紧靠对应于成熟FVIII蛋白(SEQ ID NO:65)的残基403和745的下游。在一些实施方案中,FVIII蛋白中的插入位点紧靠对应于成熟FVIII蛋白(SEQ ID NO:65)的残基745和1900的下游。在某些实施方案中,FVIII蛋白

中的插入位点紧靠对应于成熟FVIII蛋白 (SEQ ID NO:65) 的残基18和745的下游。在一些实施方案中, FVIII蛋白是双链FVIII亚型。在一些实施方案中, FVIII蛋白是单链FVIII亚型。在某些实施方案中, FVIII蛋白包含B结构域或其一部分。在一些实施方案中, FVIII蛋白是SQ B结构域缺失的FVIII。

[0066] 一些实施方案包括如本文所述的嵌合蛋白, 其中单链FVIII亚型在对应于以下的残基处含有至少一个氨基酸取代: 对应于全长成熟因子VIII多肽 (SEQ ID NO:65) 的残基1648、残基1645或两个残基或SQ BDD因子VIII (SEQ ID NO:67) 的残基754、残基751或两个残基。在某些实施方案中, 氨基酸取代是除精氨酸以外的氨基酸。在一些实施方案中, 双链FVIII亚型包含含有FVIII的重链的第一链和含有FVIII的轻链的第二链, 其中所述重链和所述轻链通过金属键彼此缔合。在某些实施方案中, D' 结构域包含与SEQ ID NO:21的氨基酸764至866至少90%、95%、96%、97%、98%、99%或100%同一的氨基酸序列。在一些实施方案中, D3结构域包含与SEQ ID NO:21的氨基酸867至1240至少90%、95%、96%、97%、98%、99%或100%同一的氨基酸序列。在某些实施方案中, VWF蛋白是单体。

[0067] 也公开如本文所述的嵌合蛋白, 其包含至少两个VWF蛋白、至少三个VWF蛋白、至少四个VWF蛋白、至少五个VWF蛋白或至少六个VWF蛋白。在某些实施方案中, VWF蛋白包含与SEQ ID NO:21的氨基酸764至1240至少90%、95%、96%、97%、98%、99%或100%同一的氨基酸序列。在一些实施方案中, VWF蛋白基本上由以下组成或由以下组成: SEQ ID NO:21的氨基酸764至1240。在某些实施方案中, VWF蛋白在对应于SEQ ID NO:21的残基1099、残基1142、或残基1099与1142两者的残基处含有至少一个氨基酸取代。在一些实施方案中, VWF蛋白含有取代对应于SEQ ID NO:21的残基1099、残基1142、或残基1099与1142两者的残基的除半胱氨酸以外的氨基酸。在某些实施方案中, VWF蛋白进一步包含VWF的D1结构域、D2结构域、或D1和D2结构域。

[0068] 一些实施方案包括如本文所述的嵌合蛋白, 其中VWF蛋白进一步包含选自以下的VWF结构域: A1结构域、A2结构域、A3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、其一个或多个片段及其任何组合。

[0069] 也公开如本文所述的嵌合蛋白, 其中VWF蛋白基本上由以下组成或由以下组成: (1) VWF的D' 和D3结构域或其片段; (2) VWF的D1、D' 和D3结构域或其片段; (3) VWF的D2、D' 和D3结构域或其片段; (4) VWF的D1、D2、D' 和D3结构域或其片段; 或(5) VWF的D1、D2、D'、D3和A1结构域或其片段。

[0070] 一些实施方案包括如本文所述的嵌合蛋白, 其中VWF蛋白进一步包含VWF或FVIII的可操作地连接于VWF蛋白的信号肽。

[0071] 也公开如本文所述的嵌合蛋白, 其中一个或多个接头的长度是至少约10、20、30、40、50、60、70、80、90、100、110、120、130、140、150、160、170、180、190、200、210、220、230、240、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。在一些实施方案中, 一个或多个接头的长度是约1至约2000个氨基酸残基。在某些实施方案中, 一个或多个接头包括gly/ser肽。在一些实施方案中, gly/ser肽具有式  $(\text{Gly}_4\text{Ser})_n$  (SEQ ID NO:94) 或  $\text{S}(\text{Gly}_4\text{Ser})_n$  (SEQ ID NO:164), 其中n是选自自由1、2、3、4、5、6、7、8、9和10组成的组的正整数。在某些实施方案中,  $(\text{Gly}_4\text{Ser})_n$  接头是  $(\text{Gly}_4\text{Ser})_3$  (SEQ ID NO:100) 或  $(\text{Gly}_4\text{Ser})_4$  (SEQ ID NO:165)。在一些实施方案中, 接头包

含20个氨基酸、35个氨基酸、48个氨基酸、73个氨基酸或95个氨基酸。在某些实施方案中,可裂解接头是SGGGSGGGSGGGSGGGSGGGSLVPRGSGG (SEQ ID NO:166)。

[0072] 在一些实施方案中,如本文所述的嵌合蛋白是聚唾液酸化、聚乙二醇化或羟乙基淀粉化的。

[0073] 也公开如本文所述的嵌合蛋白,其中第一多肽包括与FVIII161 (SEQ ID NO:69)、FVIII169 (SEQ ID NO:70)、FVIII173 (SEQ ID NO:72)、FVIII195 (SEQ ID NO:73)、FVIII196 (SEQ ID NO:74)、FVIII199 (SEQ ID NO:75)、FVIII201 (SEQ ID NO:76)、FVIII203 (SEQ ID NO:77)、FVIII204 (SEQ ID NO:78)、FVIII205 (SEQ ID NO:79)、FVIII266 (SEQ ID NO:80)、FVIII267 (SEQ ID NO:81)、FVIII268 (SEQ ID NO:82)、FVIII269 (SEQ ID NO:83)、FVIII271 (SEQ ID NO:84)、FVIII272 (SEQ ID NO:85)或FVIII282 (SEQ ID NO:159)至少约80%、90%、95%、99%或100%同一,并且第二多肽包括与VWF057 (SEQ ID NO:152)或VWF059 (SEQ ID NO:197)至少约80%、90%、95%、99%或100%同一。在一些实施方案中,第一多肽包含FVIII169 (SEQ ID NO:70),并且第二多肽包含VWF057 (SEQ ID NO:152)。在其它实施方案中,第一多肽包含FVIII169 (SEQ ID NO:70),并且第二多肽包含VWF059 (SEQ ID NO:197)。在另一实施方案中,第一多肽包含FVIII169 (SEQ ID NO:70),并且第二多肽包含VWF062 (SEQ ID NO:199)。在一些实施方案中,嵌合蛋白有效预防和/或阻止从有需要的受试者流血。

[0074] 也公开编码如本文所述的嵌合蛋白的一种多核苷酸或一组多核苷酸。在一些实施方案中,如本文所述的多核苷酸进一步包含编码PC5或PC7的多核苷酸链。

[0075] 一些实施方案包括一种载体,其包含如本文所述的多核苷酸和一种或多种可操作地连接于所述多核苷酸或所述一组多核苷酸的启动子。

[0076] 在一些实施方案中,如本文所述的载体进一步包括包含编码PC5或PC7的多核苷酸链的另外的载体。

[0077] 也公开一种包含如本文所述的多核苷酸或载体的宿主细胞。在一些实施方案中,宿主细胞是哺乳动物细胞。在某些实施方案中,哺乳动物细胞选自HEK293细胞、CHO细胞和BHK细胞。

[0078] 也公开一种药物组合物,其包含如本文所述的嵌合蛋白、多核苷酸、载体或宿主细胞以及药学上可接受的载体。在一些实施方案中,相较于野生型FVIII蛋白,嵌合蛋白具有延长的半衰期。在某些实施方案中,嵌合蛋白的半衰期比野生型FVIII延长至少约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。

[0079] 一些实施方案包括如本文所述的组合物,其中嵌合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。在某些实施方案中,嵌合蛋白在HemA小鼠中的半衰期是约40小时。在一些实施方案中,通过选自由以下组成的组的途径施用如本文所述的组合物:局部施用、眼内施用、胃肠外施用、鞘内施用、硬膜下施用和口服施用。在某些实施方案中,胃肠

外施用是静脉内或皮下施用。

[0080] 在一些实施方案中,如本文所述的组合物用于治疗有需要的受试者的流血疾病或病状。在某些实施方案中,流血疾病或病状选自由以下组成的组:流血凝血病症、关节积血、肌肉流血、口腔流血、出血、向肌肉中出血、口腔出血、创伤、头创伤、胃肠流血、颅内出血、腹内出血、胸内出血、骨折、中枢神经系统流血、咽后隙中流血、腹膜后隙中流血、髂腰肌鞘中流血及其任何组合。在一些实施方案中,受试者预定经受手术。在某些实施方案中,治疗是防治性的或按需进行。

[0081] 也公开一种延长或增加嵌合蛋白的半衰期的方法,其中所述方法包括向有需要的受试者添加有效量的如本文所述的嵌合蛋白、多核苷酸、载体、宿主细胞或组合物,其中VWF蛋白、XTEN序列、第一Ig恒定区或其一部分和第二Ig恒定区或其一部分使嵌合蛋白的半衰期增加。

[0082] 一些实施方案包括一种治疗有需要的受试者的流血疾病或病症的方法,其包括施用有效量的如本文所述的嵌合蛋白、多核苷酸、载体、宿主细胞或组合物,其中所述流血疾病或病症选自由以下组成的组:流血凝血病症、关节积血、肌肉流血、口腔流血、出血、向肌肉中出血、口腔出血、创伤、头创伤、胃肠流血、颅内出血、腹内出血、胸内出血、骨折、中枢神经系统流血、咽后隙中流血、腹膜后隙中流血和髂腰肌鞘中流血。在一些实施方案中,受试者是动物。在某些实施方案中,动物是人。在一些实施方案中,受试者正罹患甲型血友病。在某些实施方案中,治疗是防治性的或按需进行。在一些实施方案中,有效量是0.1 $\mu$ g/kg至500mg/kg。

[0083] 也公开一种如本文所述的方法,其中通过选自由以下组成的组的途径施用如本文所述的嵌合蛋白、多核苷酸、载体、宿主细胞或组合物:局部施用、眼内施用、胃肠外施用、鞘内施用、硬膜下施用和口服施用。在某些实施方案中,胃肠外施用选自由以下组成的组:静脉内施用、皮下施用、肌肉内施用和皮内施用。

[0084] 一些实施方案包括一种制备嵌合蛋白的方法,其包括用如本文所述的多核苷酸或载体转染一种或多种宿主细胞,以及在宿主细胞中表达所述嵌合蛋白。在一些实施方案中,如本文所述的方法进一步包括分离嵌合蛋白。在某些实施方案中,嵌合蛋白有效阻止和/或预防受试者的流血。

[0085] 附图简述

[0086] 图1显示嵌合蛋白的示意图,所述嵌合蛋白包含第一多肽,所述第一多肽包含融合于Fc区的FVIII蛋白(A1-A2-部分或完全B-A3-C1-C2),其中XTEN插入在所述FVIII蛋白内的插入位点处;和第二多肽,所述第二多肽包含含有D' D3结构域的VWF蛋白、具有小于288个氨基酸的XTEN、凝血酶可裂解接头和第二Fc区。XTEN插入在FVIII蛋白中和/或融合于VWF蛋白通过增加流体动力学半径以及通过阻断受体介导的清除来延长嵌合蛋白的半衰期。VWF的D' D3结构域阻断FVIII与内源性VWF相互作用,使FVIII蛋白稳定,并且延长嵌合蛋白的半衰期。Fc结构域可使D' D3结构域与FVIII蛋白共价连接,并且通过FcRn介导的再循环路径来延长嵌合蛋白的半衰期。凝血酶可裂解接头使得能够在FVIII活化后释放D' D3结构域,并且确保FVIII与VWF的D' D3结构域之间的正确对齐。

[0087] 图2显示FVIII-XTEN-Fc:D' D3-XTEN-Fc异二聚体的三质粒表达系统:第一质粒包含编码单链FVIII-XTEN-Fc的核苷酸序列,其中XTEN插入B结构域中;第二质粒包含编码

D1D2D' D3-XTEN-Fc的核苷酸序列,其中XTEN序列包含小于288个氨基酸;并且第三质粒包含编码PACE(一种原肽加工酶)的核苷酸序列。当三种多肽从三种质粒表达时,VWF的D1D2原肽结构域可通过细胞内加工来从D' D3结构域加以加工。所得复合物含有三种产物,第一分子是FVIII-XTEN/D' D3异二聚体,第二分子是副产物,即D' D3-XTEN-Fc的同二聚体,并且第三分子是另一副产物,即FVIII(XTEN)-Fc。

[0088] 图3显示XTEN插入物对异二聚体的半衰期延长的累加作用。FVIII169包含融合于Fc区的B结构域缺失的FVIII蛋白,其中XTEN序列(例如AE288)插入在对应于成熟全长FVIII的氨基酸745处。FVIII205包含融合于Fc区的B结构域缺失的FVIII蛋白,其中XTEN序列(例如AE144)插入在对应于成熟全长FVIII的氨基酸18处,并且另一XTEN序列(例如AE288)插入在对应于成熟全长FVIII的氨基酸745处。VWF031包含通过凝血酶可裂解接头融合于Fc区的VWF的D' 结构域和D3结构域(无XTEN)。VWF034包含融合于AE288和Fc区的VWF的D' 结构域和D3结构域。FVIII169/VWF031(倒三角)在HemA小鼠中的半衰期是16.7小时;FVIII205/VWF031(圆圈)在HemA小鼠中的半衰期是29.4小时;并且FVIII169/VWF034(方块)在HemA小鼠中的半衰期是31.1小时。

[0089] 图4显示当插入在VWF的D' D3结构域与Fc结构域之间时,AE144 XTEN比AE288 XTEN赋予更佳半衰期延长。举例来说,尽管VWF169/VWF034(方块)在HemA小鼠中的半衰期是31.1.小时,但FVIII169/VWF057(圆圈)在HemA小鼠中的半衰期是42小时。VWF057包含融合于AE144和Fc区的VWF的D' D3结构域。

[0090] 图5显示Fc结构域为嵌合蛋白异二聚体的半衰期延长所需。当将FVIII205/VWF031(圆圈)在HemA小鼠中的半衰期与在FcRn结合位点处含有突变(IHH三重突变Fc)并且因此不能通过FcRn路径再循环的FVIII263/VWF050(方块)的半衰期进行比较时,FVIII263/VWF050的半衰期(23小时)短于VWF205/VWF031的半衰期(29.4小时)。这指示Fc区为半衰期延长所必需。

[0091] 图6小图A显示在HemA小鼠尾部钳夹模型中,FVIII-XTEN-Fc/D' D3-XTEN-Fc异二聚体相较于B结构域缺失的FVIII(SQ BDD FVIII)的类似急性功效。在75IU/kg下对小鼠给药,并且通过aPTT测定来测量活性。SQ BDD FVIII显示为圆圈,而FVIII169/VWF034显示为方块,FVIII169/VWF057显示为菱形,并且媒介物显示为倒三角。FVIII169、VWF034和VWF057的构建体细节显示在本文其它地方。图6小图B显示在37.5IU/kg剂量下,FVIII169/VWF034与B结构域缺失的FVIII(SQ BDD FVIII)在HemA小鼠中的急性功效的比较,并且通过aPTT测定来测量活性。各治疗组中小鼠的中值失血(uL)由水平线指示,C57/BL6小鼠中的失血(uL)显示为中空三角;在37.5IU/kg的rBDD-FVIII给药之后的失血(uL)显示为中空圆圈;在37.5IU/kg的FVIII169/VWF034给药之后的失血(uL)显示为中空方块,并且在媒介物给药之后的失血(uL)显示为倒三角。

[0092] 图7小图A-B显示rFVIII169/VWF057异二聚体在尾部静脉横切流血模型中对HemA小鼠提供较长久保护。图7小图A显示在尾部损伤之前72小时接受rFVIII169/VWF057(方块)、在尾部损伤之前48小时接受SQ BDD-FVIII(菱形)、在尾部损伤之前24小时接受SQBDD FVIII(倒三角)以及接受媒介物(圆圈)的小鼠中的再流血数据。通过aPTT测定来测量活性。X轴显示以小时计的时间,并且Y轴显示非流血者的百分比。图7小图B显示图7小图A中所示的四个种类的小鼠中的相应存活数据。相较于在尾部损伤之前24小时接受SQBDD FVIII治

疗的小鼠,在尾部损伤之前72小时接受12IU/kg的FVIII169/VWF057的小鼠显示关于再流血和存活的类似保护。

[0093] 图8小图A显示在损伤之前96小时接受rFVIII-XTEN-Fc/D' D3-XTEN-Fc异二聚体的小鼠中相对于在损伤之前24小时接受rBDD-FVIII的小鼠中的类似再流血数据。实心方块显示在损伤之前24小时接受FVIII169/VWF034的小鼠中的再流血数据;中空方块显示在损伤之前96小时接受FVIII169/VWF034的小鼠中的再流血数据;实心菱形显示在损伤之前24小时接受FVIII169/VWF057的小鼠中的再流血数据;中空菱形显示在损伤之前96小时接受FVIII169/VWF057的小鼠中的再流血数据;实心圆圈显示在损伤之前24小时接受rBDD-FVIII的小鼠中的再流血数据;中空圆圈显示在损伤之前48小时接受rBDD-FVIII的小鼠中的再流血数据;并且实心三角显示接受媒介物的小鼠中的再流血数据。X轴显示以小时计的时间,并且y轴显示非流血者的百分比。

[0094] 图8小图B显示在损伤之前96小时接受rFVIII-XTEN-Fc/D' D3-XTEN-Fc异二聚体的小鼠中相对于在损伤之前24小时接受rBDD-FVIII的小鼠中的存活曲线。X轴显示以小时计的时间,并且y轴显示存活百分比。符号与图8小图A相同。

[0095] 图9显示代表性FVIII-VWF异二聚体以及FVIII169、FVIII286、VWF057、VWF059和VWF062构建体的图解。举例来说,FVIII169构建体包含融合于Fc区的具有R1648A取代的B结构域缺失的FVIII蛋白,其中XTEN序列(例如AE288)插入在对应于成熟全长FVIII的氨基酸745处(A1-a1-A2-a2-288XTEN-a3-A3-C1-C2-Fc)。FVIII286构建体包含融合于Fc区的具有R1648取代的B结构域缺失的FVIII蛋白,其中XTEN序列(例如AE288)插入在对应于成熟全长FVIII的氨基酸745处,在FVIII与Fc之间具有另外的a2区(A1-a1-A2-a2-288XTEN-a3-A3-C1-C2-a2-Fc)。VWF057是一种VWF-Fc融合构建体,其包含通过包含LVPRG凝血酶位点(“LVPRG”;SEQ ID NO:6)和GS接头(“GS”)的VWF接头连接于Fc区的VWF蛋白的D' D3结构域(在D' D3结构域中具有两个氨基酸取代,即C336A和C379A),其中XTEN序列(即AE144)插入在D' D3结构域与VWF接头之间(D' D3-144XTEN-GS+LVPRG-Fc)。VWF059是一种VWF-Fc融合构建体,其包含通过FVIII的酸性区2(a2)作为VWF接头连接于Fc区的VWF蛋白的D' D3结构域(在D' D3结构域中具有两个氨基酸取代,即C336A和C379A),其中XTEN序列(即AE144)插入在D' D3结构域与VWF接头之间。VWF062是一种VWF-Fc融合构建体,其包含连接于Fc区的VWF蛋白的D' D3结构域(在D' D3结构域中具有两个氨基酸取代,即C336A和C379A),其中XTEN序列(即AE144)插入在D' D3结构域与Fc区之间(D' D3-144XTEN-Fc)。

[0096] 图10显示表示FVIII/VWF异二聚体构建体,例如FVIII169/VWF057、FVIII169/VWF059、FVIII169/VWF059A和FVIII169/VWF073的示意图。箭头显示添加任选接头以引入凝血酶裂解位点所处的位点。FVIII169/VWF057具有包含LVPRG(SEQ ID NO:6)的接头。FVIII169/VWF059具有包含FVIII a2区(即 ISDKNTGDYYEDSYEDISAYLLSKNNAIEPRSFSDKTH (SEQ ID NO:106))的接头。FVIII169/VWF059A具有包含截短FVIII a2区(即 DKNTGDYYEDSYEDISAYLLSKNNAIEPRSFSDKTH (SEQ ID NO:88))的接头。FVIII169/VWF073在VWF073构建体(SEQ ID NO:175)内具有包含FVIII a2区的由IEPRSF(SAQ ID NO:194)组成的片段的接头。

[0097] 图11小图A-C显示在凝血酶消化FVIII169/VWF057和FVIII-Fc对照之后的SDS-

PAGE图像。图11小图A显示用抗D3抗体(AB 96340)对SDS-PAGE凝胶的染色。箭头突出“LCFc:D'D3-XTEN-Fc”,其是未裂解的全长FVIII169/VWF057;和“D'D3-144 XTEN”,其是在由凝血酶裂解之后的所得片段。图11小图B显示用抗HC抗体(GMA012)对SDS-PAGE凝胶的染色。箭头突出FVIII重链(“HC”)和FVIII A2结构域。图11小图C显示小图A和B的重叠图。在各图的顶部指示的时间点收集样品。箭头指向相关蛋白质。

[0098] 图12小图A-C显示在凝血酶消化FVIII169/VWF059之后的SDS-PAGE图像。图12小图A显示用抗D3抗体(AB 96340)对SDS-PAGE凝胶的染色。箭头突出“LCFc:D'D3-XTEN-Fc,”其是未裂解的全长FVIII169/VWF059;和“D'D3-144 XTEN”,其是在由凝血酶裂解之后的所得片段。图12小图B显示用抗HC抗体(GMA012)对SDS-PAGE凝胶的染色。箭头突出未裂解的全长FVIII169/VWF059;D'D3-144XTEN-a3,其是在由凝血酶裂解之后的所得片段;和“A2”,其是FVIII的A2结构域。图12小图C显示小图A和B的重叠图。在各图的顶部指示的时间点收集样品。

[0099] 图13显示相较于用BDD-FVIII对照(方块)治疗的HemA小鼠,用FVIII169/VWF059(圆圈)治疗的HemA小鼠的急性功效数据。在尾部钳夹之后测量失血值。 $p=0.9883$ 。

[0100] 发明详述

[0101] 本发明涉及一种包含两个多肽的嵌合蛋白,第一多肽包含融合于第一Ig恒定区的FVIII蛋白,并且第二多肽包含通过XTEN序列融合于第二Ig恒定区或其一部分的VWF蛋白,其中所述XTEN序列含有小于288个氨基酸。

[0102] I. 定义

[0103] 应注意术语“一”实体是指一个或多个那个实体;例如“一核苷酸序列”应理解为表示一个或多个核苷酸序列。因此,术语“一”、“一个(种)或多个(种)”和“至少一个(种)”在本文中可互换使用。

[0104] 此外,当在本文中使用时,“和/或”应视为特定公开两个指定特征或组分中的每一个,伴有或不伴有另一个。因此,如本文在诸如“A和/或B”的短语中所用的术语“和/或”意图包括“A和B”、“A或B”、“A”(单独)和“B”(单独)。同样,如诸如“A、B和/或C”的短语中所用的术语“和/或”意图涵盖以下方面中的每一个:A、B和C;A、B或C;A或C;A或B;B或C;A和C;A和B;B和C;A(单独);B(单独);和C(单独)。

[0105] 应了解每当在本文中措辞“包含”描述各个方面的情况下,也提供以“由…组成”和/或“基本上由…组成”描述的另外类似方面。

[0106] 除非另外定义,否则本文所用的所有技术和科学术语都具有与由本公开所相关领域中的普通技术人员通常理解相同的含义。举例来说,Concise Dictionary of Biomedicine and Molecular Biology, Juo, Pei-Show, 第2版, 2002, CRC Press; The Dictionary of Cell and Molecular Biology, 第3版, 1999, Academic Press; 以及 Oxford Dictionary Of Biochemistry And Molecular Biology, 修订版, 2000, Oxford University Press 提供技术人员以本公开中使用的许多术语的综合词典。

[0107] 单位、前缀和符号以它们的Système International de Unites (SI) 接受形式表示。数值范围包括界定范围的数值。除非另外指示, 否则氨基酸序列以氨基至羧基定向从左至右书写。本文提供的标题并非是对本公开的各个方面的限制, 所述各个方面可通过参考说明书整体而得到。因此, 通过参考说明书整体, 以下紧接着定义的术语被更充分阐释。

[0108] 术语“约”在本文中用于意指近似、大致、大约或在...左右。当术语“约”与数值范围联合使用时，它通过使边界延伸高于和低于阐述的数值来修饰那个范围。一般来说，术语“约”可通过向上或向下(增高或降低)变化例如10%来修饰数值高于和低于陈述的值。

[0109] 术语“多核苷酸”或“核苷酸”意图涵盖单个核酸以及多个核酸，并且是指分离的核酸分子或构建体，例如信使RNA (mRNA) 或质粒DNA (pDNA)。在某些实施方案中，多核苷酸包含常规磷酸二酯键或非常规键(例如酰胺键，诸如见于肽核酸(PNA)中)。术语“核酸”是指存在于多核苷酸中的任何一个或多个核酸区段，例如DNA或RNA片段。就“分离的”核酸或多核苷酸来说，其意指已从它的天然环境移除的核酸分子DNA或RNA。举例来说，出于本发明的目的，载体中含有的编码因子VIII多肽的重组多核苷酸被视为是分离的。分离的多核苷酸的其它实例包括维持在异源性宿主细胞中的重组多核苷酸，或呈溶液形式的从其它多核苷酸纯化(部分或实质上纯化)的重组多核苷酸。分离的RNA分子包括本发明的多核苷酸的体内或体外RNA转录物。本发明的分离的多核苷酸或核酸进一步包括合成产生的所述分子。此外，多核苷酸或核酸可包括调控元件，诸如启动子、增强子、核糖体结合位点或转录终止信号。

[0110] 如本文所用，“编码区”或“编码序列”是多核苷酸的由可翻译成氨基酸的密码子组成的一部分。尽管“终止密码子”(TAG、TGA或TAA)通常不翻译成氨基酸，但它可被视为编码区的一部分，然而任何侧接序列(例如启动子、核糖体结合位点、转录终止子、内含子等)都不是编码区的一部分。编码区的边界通常由在5'末端的起始密码子(编码所得多肽的氨基末端)和在3'末端的翻译终止密码子(编码所得多肽的羧基末端)来确定。本发明的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。那么由此可见单一载体可仅含有单一编码区，或包含两个或更多个编码区，例如单一载体可单独编码如下所述的结合结构域A和结合结构域B。此外，本发明的载体、多核苷酸或核酸可编码融合或未融合于编码本发明的结合结构域的核酸的异源性编码区。异源性编码区包括不限于专门化元件或基序，诸如分泌信号肽或异源性功能性结构域。

[0111] 由哺乳动物细胞分泌的某些蛋白质与分泌信号肽缔合，一旦逐渐增长的蛋白质链跨越粗糙内质网的排出已被引发，所述信号肽即从成熟蛋白质裂解。本领域普通技术人员应了解信号肽通常融合于多肽的N末端，并且从完全或“全长”多肽裂解以产生多肽的分泌或“成熟”形式。在某些实施方案中，使用天然信号肽，或那个序列的保留引导可操作地与它关联的多肽分泌的能力的功能性衍生物。或者，可使用异源性哺乳动物信号肽，例如人组织纤维蛋白溶酶原活化因子(tissue plasminogen activator, TPA)或小鼠 $\beta$ -葡萄糖醛酸酶( $\beta$ -glucuronidase)信号肽或其功能性衍生物。

[0112] 术语“下游”在涉及核苷酸序列时意指核酸或核苷酸序列位于参照核苷酸序列的3'。在某些实施方案中，下游核苷酸序列涉及在转录起始点之后的序列。举例来说，基因的翻译起始密码子位于转录起始位点的下游。术语“下游”在涉及多肽序列时意指氨基酸或氨基酸插入位点位于参照氨基酸的C末端。举例来说，紧靠对应于成熟野生型FVIII蛋白的氨基酸745的下游的插入位点意指插入位点在对应于成熟野生型FVIII蛋白的氨基酸745与氨基酸746之间。

[0113] 术语“上游”是指核苷酸序列位于参照核苷酸序列的5'。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。举例来说,大多数启动子位于转录起始位点的上游。

[0114] 如本文所用,术语“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),并且影响关联的编码区的转录、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、聚腺苷酸化识别序列、RNA加工位点、效应物结合位点和茎-环结构。如果编码区意图在真核细胞中表达,那么聚腺苷酸化信号和转录终止序列将通常位于编码序列的3'。

[0115] 编码基因产物(例如多肽)的多核苷酸可包括可操作地与一个或多个编码区关联的启动子和/或其它转录或翻译控制元件。在一可操作关联中,基因产物(例如多肽)的编码区以将所述基因产物的表达置于调控区的影响或控制下的方式与一个或多个调控区关联。举例来说,如果诱导启动子功能导致编码由编码区编码的基因产物的mRNA转录,并且如果启动子与编码区之间的连接的性质不干扰启动子引导基因产物表达的能力或不干扰DNA模板被转录的能力,那么编码区和启动子被“可操作地关联”。除启动子之外,其它转录控制元件(例如增强子、操纵子、阻遏子和转录终止信号)也可可操作地与编码区关联以引导基因产物表达。

[0116] 多种转录控制区为本领域技术人员所知。这些转录控制区包括不限于在脊椎动物细胞中起作用的转录控制区,诸如但不限于来自巨细胞病毒(立即早期启动子,与内含子-A联合)、猿猴病毒40(早期启动子)和逆转录病毒(诸如劳斯肉瘤病毒(Rous sarcoma virus))的启动子和增强子区段。其它转录控制区包括源于脊椎动物基因(诸如肌动蛋白(actin)、热休克蛋白、牛生长激素和兔 $\beta$ -球蛋白)的那些转录控制区,以及能够控制真核细胞中的基因表达的其它序列。另外适合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(lymphokine)诱导性启动子(例如可由干扰素或白介素诱导的启动子)。

[0117] 类似地,多种翻译控制元件为本领域普通技术人员所知。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、以及源于微小RNA病毒的元件(特别是内部核糖体进入位点或IRES,也被称为CITE序列)。

[0118] 如本文所用的术语“表达”是指多核苷酸产生基因产物(例如RNA或多肽)所采用的过程。它包括不限于多核苷酸转录成信使RNA(mRNA)、转移RNA(tRNA)、小发夹RNA(shRNA)、小干扰RNA(siRNA)或任何其它RNA产物,以及mRNA翻译成多肽。表达产生“基因产物”。如本文所用,基因产物可为核酸(例如通过基因转录产生的信使RNA)或从转录物翻译的多肽。本文所述的基因产物进一步包括具有转录后修饰(例如聚腺苷酸化或剪接)的核酸、或具有翻译后修饰(例如甲基化、糖基化、添加脂质、与其它蛋白质亚单位缔合、或蛋白水解裂解)的多肽。

[0119] “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物。载体可为另一核酸区段可与其连接以便致使所连接区段复制的复制子。“复制子”是指在体内充当自主复制单元,即能够在它自身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”包括用于在体外、离体或在体内将核酸引入细胞中的病毒媒介物与非病毒媒介物两者。本领域中已知并使用许多载体,包括例如质粒、修饰的真核病毒或修饰的细菌病毒。将多核苷酸插入适合载体中可通过将适当多核苷酸片段连接至具有互补粘性末端的所选载体中来实现。

[0120] 载体可被工程改造以编码提供对已并有载体的细胞的选择或鉴定的可选择标记或报道体。可选择标记或报道体的表达允许鉴定和/或选择并有且表达载体上含有的其它编码区的宿主细胞。本领域中已知并使用的可选择标记基因的实例包括：提供对氨苄青霉素(ampicillin)、链霉素(streptomycin)、庆大霉素(gentamycin)、卡那霉素(kanamycin)、潮霉素(hygromycin)、双丙胺膦(bialaphos)除草剂、磺酰胺等的抗性的基因；以及用作表型标记的基因，即花青素(anthocyanin)调控基因、异戊基转移酶基因等。本领域中已知并使用的报道体的实例包括：荧光素酶(Luc)、绿色荧光蛋白(GFP)、氯霉素乙酰转移酶(CAT)、-半乳糖苷酶(LacZ)、-葡萄糖醛酸酶(Gus)等。可选择标记也可被视为报道体。

[0121] 术语“质粒”是指染色体外元件，其常携带不是细胞的中心代谢的一部分的基因，并且通常呈环状双链DNA分子形式。所述元件可为源于任何来源的具有单链或双链DNA或RNA的线性、环状或超螺旋自主复制序列、基因组整合序列、噬菌体或核苷酸序列，其中许多核苷酸序列已被接合或重组成能够将启动子片段和所选基因产物的DNA序列连同适当3'非翻译序列一起引入细胞中的独特构造。

[0122] 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、腺相关病毒载体和痘病毒(例如痘苗病毒)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、带电荷脂质(细胞转染剂)、DNA-蛋白质复合物和生物聚合物。

[0123] “克隆载体”是指是依序复制的具有一定单元长度的核酸，并且包含复制起点的“复制子”(诸如质粒、噬菌体或粘粒)，另一核酸区段可与其连接以便致使所连接区段复制。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。克隆载体通常包含一种或多种可用于选择包含载体的细胞的序列和/或一个或多个用于插入目标核酸序列的多克隆位点。

[0124] 术语“表达载体”是指被设计以使得能够在插入宿主细胞中之后表达插入的核酸序列的媒介物。插入的核酸序列以与如上所述的调控区可操作关联的方式放置。

[0125] 通过本领域中熟知的方法将载体引入宿主细胞中，所述方法例如转染、电穿孔、显微注射、转导、细胞融合、DEAE右旋糖苷、磷酸钙沉淀、脂质体转染(溶酶体融合)、使用基因枪、或DNA载体转运体。

[0126] 如本文所用的“培养”意指在允许细胞生长或分裂的体外条件下孵育细胞或以存活状态维持细胞。如本文所用的“培养的细胞”意指在体外繁殖的细胞。

[0127] 如本文所用，术语“多肽”意图涵盖单个“多肽”以及多个“多肽”，并且是指由通过酰胺键(也称为肽键)线性连接的单体(氨基酸)组成的分子。术语“多肽”是指具有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”或用于指代具有两个或更多个氨基酸的一个或多个链的任何其它术语包括在“多肽”的定义内，并且术语“多肽”可替代任何这些术语加以使用或可与任何这些术语互换使用。术语“多肽”也意指具有多肽的表达后修饰的产物，所述修饰包括但不限于糖基化、乙酰化、磷酸化、酰胺化、由已知保护/封闭基团达成的衍生化、蛋白水解裂解、或由非天然存在的氨基酸达成的修饰。多肽可源于天然生物来源或通过重组技术产生，但未必从指定核酸序列翻译。它可以任何方式，包括通过化学合成产生。

[0128] “分离的”多肽或其片段、变体或衍生物是指不在它的天然环境中的多肽。不要求特定纯化水平。举例来说，分离的多肽可仅从它的天然或自然环境移除。出于本发明的目

的,在宿主细胞中表达的重组产生的多肽和蛋白质被视为是分离的,已通过任何适合技术分离、分级分离、或部分或实质上纯化的天然或重组多肽也被视为是分离的。

[0129] 本发明中也包括多肽的片段或变体及其任何组合。术语“片段”或“变体”在涉及本发明的多肽结合结构域或结合分子时包括保留参照多肽的至少一些性质(例如FcRn结合结构域或Fc变体的FcRn结合亲和力、FVIII变体的凝血活性或VWF片段的FVIII结合活性)的任何多肽。除在本文其它地方讨论的特定抗体片段之外,多肽的片段也包括蛋白水解片段以及缺失片段,但不包括天然存在的全长多肽(或成熟多肽)。本发明的多肽结合结构域或结合分子的变体包括如上所述的片段以及氨基酸序列由于氨基酸取代、缺失或插入而改变的多肽。变体可为天然存在的或非天然存在的。非天然存在的变体可使用本领域已知的诱变技术产生。变异多肽可包含保守性或非保守性氨基酸取代、缺失或添加。

[0130] 本文所用的术语“VWF蛋白”或“VWF蛋白”意指与FVIII相互作用,并且保留至少一种或多种通常由全长VWF对FVIII提供的性质的任何VWF片段,所述性质例如防止过早活化成FVIIIa,防止过早蛋白水解,防止与磷脂膜进行可导致过早清除的缔合,防止结合可结合裸露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。

[0131] “保守性氨基酸取代”是其中氨基酸残基被具有类似侧链的氨基酸残基置换的氨基酸取代。具有类似侧链的氨基酸残基的家族已在本领域中加以定义,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)、不带电荷极性侧链(例如甘氨酸、天冬酰胺、谷氨酰胺、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)、 $\beta$ -分支侧链(例如苏氨酸、缬氨酸、异亮氨酸)和芳族侧链(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被来自同一侧链家族的另一氨基酸置换,那么取代被视为是保守性的。在另一实施方案中,一串氨基酸可被在侧链家族成员的顺序和/或组成方面不同的在结构上类似的一串保守置换。

[0132] 如本领域中所知,通过比较一个多肽的氨基酸序列与第二多肽的序列来确定两个多肽之间的“序列同一性”。当在本文中讨论时,任何特定多肽是否与另一多肽至少约50%、60%、70%、75%、80%、85%、90%、95%、99%或100%同一可使用本领域中已知的方法和计算机程序/软体来确定,诸如但不限于BESTFIT程序(用于Unix的Wisconsin序列分析包,第8版,Genetics Computer Group,University Research Park,575Science Drive, Madison,WI 53711)。BESTFIT使用Smith和Waterman,Advances in Applied Mathematics 2:482-489(1981)的局部同源性算法以得到两个序列之间具有同源性的最佳区段。当使用BESTFIT或任何其它序列比对程序来确定特定序列是否与本发明的参照序列例如95%同一时,当然应设置参数以使历经参照多肽序列的全长计算同一性百分比,并且允许同源性空位占参照序列中氨基酸总数的多达5%。

[0133] 如本文所用,通过比对以使第一VWF或FVIII序列与第二VWF或FVIII序列之间的同一性或类似性最大来鉴定VWF序列或FVIII蛋白质序列中“对应于…的氨基酸”或“等效氨基酸”。用于鉴定第二VWF或FVIII序列中等效氨基酸的编号基于用于鉴定第一VWF或FVIII序列中相应氨基酸的编号。

[0134] 如本文所用,术语“插入位点”是指FVIII多肽或其片段、变体或衍生物中紧靠可插

入异源性部分所处的位置的上游的位置。以某一编号指定“插入位点”，所述编号是成熟天然FVIII (SEQ ID NO:65) 中由插入位点所对应的氨基酸的编号，所述氨基酸紧靠插入位置的N末端。举例来说，短语“a3在对应于SEQ ID NO:65的氨基酸1656的插入位点处包含XTEN”指示异源性部分位于对应于SEQ ID NO:65的氨基酸1656和氨基酸1657的两个氨基酸之间。

[0135] 如本文所用的短语“紧靠氨基酸的下游”是指紧接于所述氨基酸的末端羧基的位置。类似地，短语“紧靠氨基酸的上游”是指紧接于所述氨基酸的末端胺基团的位置。因此，如本文所用的短语“在插入位点的两个氨基酸之间”是指XTEN或任何其它多肽插入在两个邻近氨基酸之间所处的位置。因此，短语“紧靠氨基酸的下游插入”和“插入在插入位点的两个氨基酸之间”与“在插入位点处插入”同义使用。

[0136] 如本文所用的术语“插入”、“被插入”、“插入…中”或语法相关术语是指相对于天然成熟人FVIII中的类似位置，XTEN在嵌合多肽中的位置。如本文所用，所述术语是指重组FVIII多肽相对于天然成熟人FVIII的特征，并且不指示、暗示或意指制备嵌合多肽所采用的任何方法或过程。举例来说，关于本文提供的嵌合多肽，短语“将XTEN紧靠FVIII多肽的残基745的下游插入”意指所述嵌合多肽在紧靠对应于天然成熟人FVIII中的氨基酸745的氨基酸的下游包含XTEN，其例如由对应于天然成熟人FVIII的氨基酸745和746的氨基酸界定。

[0137] “融合”或“嵌合”蛋白包含连接于在自然界中它不天然连接的第二氨基酸序列的第一氨基酸序列。通常存在于单独蛋白质中的氨基酸序列可以融合多肽形式集合在一起，或通常存在于同一蛋白质中的氨基酸序列可以新排列放置在融合多肽中，例如本发明的因子VIII结构域与Ig Fc结构域的融合。例如通过化学合成，或通过产生并翻译其中肽区域以所需关系被编码的多核苷酸来产生融合蛋白。嵌合蛋白可进一步包含通过共价非肽键或非共价键与第一氨基酸序列缔合的第二氨基酸序列。

[0138] 如本文所用，术语“半衰期”是指特定多肽在体内的生物半衰期。半衰期可由向受试者施用的一半数量从动物中的循环和/或其它组织清除所需的时间表示。当构建给定多肽随时间变化的清除曲线时，曲线通常是双相的，即快速 $\alpha$ 相和较长 $\beta$ 相。 $\alpha$ 相通常表示施用的Fc多肽在血管内间隙与血管外间隙之间达到平衡，并且部分地由多肽的大小决定。 $\beta$ 相通常表示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 $\alpha$ 相，而仅具有单一 $\beta$ 相。因此，在某些实施方案中，如本文所用的术语半衰期是指处于 $\beta$ 相的多肽的半衰期。人抗体在人体中的典型 $\beta$ 相半衰期是21天。

[0139] 如本文所用的术语“连接”是指第一氨基酸序列或核苷酸序列分别共价或非共价接合于第二氨基酸序列或核苷酸序列。第一氨基酸或核苷酸序列可直接接合或毗连于第二氨基酸或核苷酸序列，或者间插序列可使第一序列共价接合于第二序列。术语“连接”不仅意指第一氨基酸序列在C末端或N末端融合于第二氨基酸序列，而且也包括将整个第一氨基酸序列(或第二氨基酸序列)插入第二氨基酸序列(或相应地，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，第一氨基酸序列可通过肽键或接头连接于第二氨基酸序列。第一核苷酸序列可通过磷酸二酯键或接头连接于第二核苷酸序列。接头可为肽或多肽(用于多肽链)或核苷酸或核苷酸链(用于核苷酸链)或任何化学部分(用于多肽链与多核苷酸链两者)。术语“连接”也用连字符(-)指示。

[0140] 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。这个缔合可

用冒号,即(:)指示。在另一实施方案中,它意指除肽键之外的共价键。举例来说,氨基酸半胱氨酸包含可与第二半胱氨酸残基上的硫醇基团形成二硫键或二硫桥的硫醇基团。在大多数天然存在的IgG分子中,CH1区和CL区通过二硫键缔合,并且两个重链通过两个二硫键在对应于239和242的位置处缔合,所述位置使用Kabat编号系统(位置226或229,EU编号系统)。共价键的实例包括但不限于肽键、金属键、氢键、二硫键、 $\sigma$ 键、 $\pi$ 键、 $\delta$ 键、糖苷键、抓氢键(agnostic bond)、弯曲键、偶极键、 $\pi$ 反向键、双键、三键、四键、五键、六键、共轭、超共轭、芳香性、哈普托数(hapticity)或反键合。非共价键的非限制性实例包括离子键(例如阳离子- $\pi$ 键或盐键)、金属键、氢键(例如二氢键、二氢复合物、低能障氢键或对称氢键)、范德华力(van der Waals force)、伦敦分散力(London dispersion force)、机械键、卤素键、亲金性、插入、堆积、熵力或化学极性。

[0141] 本文所用的术语“单体-二聚体杂交物”是指包含通过二硫键彼此缔合的第一多肽链和第二多肽链的嵌合蛋白,其中所述第一链包含凝结因子(例如因子VIII)和第一Fc区,并且所述第二链包含以下、基本上由以下组成或由以下组成:无凝结因子的第二Fc区。因此,单体-二聚体杂交物构建体是一种杂交物,其包含仅具有一个凝结因子的单体方面和具有两个Fc区的二聚体方面。

[0142] 如本文所用,术语“裂解位点”或“酶促裂解位点”是指由酶识别的位点。某些酶促裂解位点包括细胞内加工位点。在一个实施方案中,多肽具有由在凝结级联期间活化的酶裂解的酶促裂解位点,以致所述位点的裂解发生在凝块形成的部位处。示例性所述位点包括例如由凝血酶、因子XIa或因子Xa识别的那些。示例性FXIa裂解位点包括例如TQSFNDFTR (SEQ ID NO:1)和SVSQTSKLTR (SEQ ID NO:3)。示例性凝血酶裂解位点包括例如DFLAEGGGVR (SEQ ID NO:4)、TTKIKPR (SEQ ID NO:5)、LVPRG (SEQ ID NO:6)、ALRPR (SEQ ID NO:7)、ISDKNTGDYYEDSYEDISAYLLSKNNAIEPRSF (SEQ ID NO:106)、DKNTGDYYEDSYEDISAYLLSKNNAIEPRSF (SEQ ID NO:88)和IEPRSF (SEQ ID NO:194)。其它酶促裂解位点在本领域中是已知的,并且描述于本文其它地方。

[0143] 如本文所用,术语“加工位点”或“细胞内加工位点”是指多肽中的一种类型的酶促裂解位点,其是在所述多肽翻译之后起作用的酶的靶标。在一个实施方案中,所述酶在从高尔基体腔(Golgi lumen)转运至反面高尔基体(trans-Golgi)区室期间起作用。细胞内加工酶在蛋白质从细胞分泌之前裂解多肽。所述加工位点的实例包括例如由内肽酶的PACE/弗林蛋白酶(furin)(其中PACE是成对碱性氨基酸裂解酶(Paired basic Amino acid Cleaving Enzyme)的首字母缩略词)家族靶向的那些。这些酶定位于高尔基体膜,并且在序列基序Arg-[任何残基]-(Lys或Arg)-Arg的羧基末端侧上裂解蛋白质。如本文所用,“弗林蛋白酶”家族的酶包括例如PCSK1(也称为PC1/Pc3)、PCSK2(也称为PC2)、PCSK3(也称为弗林蛋白酶或PACE)、PCSK4(也称为PC4)、PCSK5(也称为PC5或PC6)、PCSK6(也称为PACE4)或PCSK7(也称为PC7/LPC、PC8或SPC7)。其它加工位点在本领域中是已知的。

[0144] 在包括超过一个加工或裂解位点的构建体中,应了解所述位点可相同或不同。

[0145] 术语“弗林蛋白酶”是指对应于EC编号3.4.21.75的酶。弗林蛋白酶是枯草杆菌蛋白酶样原蛋白转化酶,其也称为PACE(成对碱性氨基酸裂解酶)。弗林蛋白酶使非活性前体蛋白质的各段缺失以使它们转化成生物活性蛋白质。在它的细胞内转运期间,VWF的原肽可由弗林蛋白酶从成熟VWF分子裂解。在一些实施方案中,弗林蛋白酶从VWF的D'D3裂解D1D2。

在其它实施方案中,编码弗林蛋白酶的核苷酸序列可连同编码VWF片段的核苷酸序列一起表达以使D1D2结构域可由弗林蛋白酶在细胞内裂解去除。

[0146] 在包括超过一个加工或裂解位点的构建体中,应了解所述位点可相同或不同。

[0147] 如本文所用的“可加工接头”是指包含至少一个在本文其它地方描述的细胞内加工位点的接头。

[0148] 如本文所用的止血病症意指特征在于由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而有自发出血或由于创伤而出血的倾向的基因遗传性或获得性病状。所述病症的实例包括血友病。三种主要形式是甲型血友病(因子VIII缺乏症)、乙型血友病(因子IX缺乏症或“克雷司马斯病(Christmas disease)”)和丙型血友病(因子XI缺乏症,轻微流血倾向)。其它止血病症包括例如范威尔邦德病;因子XI缺乏症(PTA缺乏症);因子XII缺乏症;纤维蛋白原、凝血酶原、因子V、因子VII、因子X或因子XIII缺乏症或结构异常;伯纳德-苏里尔综合征(Bernard-Soulier syndrome),其是一种GPIb缺陷症或缺乏症。VWF的受体GPIb可为缺陷性的,并且导致缺乏初级凝块形成(初级止血)和流血倾向增加、以及格兰茨曼和内格利血小板无力症(thrombasthenia of Glanzman and Naegeli)(格兰茨曼血小板无力症)。在肝衰竭(急性和慢性形式)中,由肝产生的凝血因子存在不足;此可增加流血风险。

[0149] 本发明的嵌合分子可防治性加以使用。如本文所用,术语“防治性治疗”是指在流血发作之前施用分子。在一个实施方案中,需要一般性止血剂的受试者正经受或即将经受手术。本发明的嵌合蛋白可在手术之前或之后作为防治剂加以施用。本发明的嵌合蛋白可在手术期间或之后施用以控制急性流血发作。手术可包括但不限于肝移植、肝切除、牙科程序或干细胞移植。

[0150] 本发明的嵌合蛋白也用于按需治疗。术语“按需治疗”是指回应于流血发作的症状或在可能导致流血的活动之前施用嵌合分子。在一个方面,按需治疗可在流血开始时(诸如在损伤之后)或在预期会流血时(诸如在手术之前)给予受试者。在另一方面,按需治疗可在会增加流血风险的活动(诸如接触性运动)之前给予。

[0151] 如本文所用,术语“急性流血”是指无论潜伏原因如何的流血发作。举例来说,受试者可具有创伤、尿毒症、遗传性流血病症(例如因子VII缺乏症)、血小板病症,或由于产生针对凝结因子的抗体而具有抗性。

[0152] 如本文所用的治疗是指例如减轻疾病或病状的严重性;降低疾病病程的持续时间;改善一种或多种与疾病或病状相关的症状;向患有疾病或病状的受试者提供有益作用,而未必治愈所述疾病或病状;或防治一种或多种与疾病或病状相关的症状。在一个实施方案中,术语“治疗”意指通过施用本发明的嵌合蛋白或VWF片段来使受试者中的FVIII谷底水平维持在至少约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。在另一实施方案中,治疗意指维持FVIII谷底水平在以下之间:约1与约20IU/dL、约2与约20IU/dL、约3与约20IU/dL、约4与约20IU/dL、约5与约20IU/dL、约6与约20IU/dL、约7与约20IU/dL、约8与约20IU/dL、约9与约20IU/dL或约10与约20IU/dL。治疗疾病或病状也可包括维持受试者中的FVIII活性在类似于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%的水平下。为治疗所需的最小谷底水平可通过一种或多种已知方法测量，并且可针对每个人加以调整（增加或减少）。

#### [0153] II. 嵌合蛋白

[0154] 本发明涉及使用融合于XTEN序列的VWF蛋白，通过防止或抑制FVIII半衰期限制因素（即内源性VWF）与FVIII蛋白缔合来延长嵌合蛋白的半衰期。内源性VWF与约95%至约98%的FVIII以非共价复合物形式缔合。尽管内源性VWF是FVIII半衰期限制因素，但也已知内源性VWF结合于FVIII蛋白质会以各种方式保护FVIII。举例来说，全长VWF（呈具有约250kDa的多聚体形式）可保护FVIII免遭蛋白酶裂解和FVIII活化，使FVIII重链和/或轻链稳定，并且防止FVIII被清道夫受体清除。但同时，内源性VWF通过防止胞饮作用以及通过VWF清除路径从系统清除FVIII-VWF复合物而限制FVIII半衰期。在不受理论束缚下，据信内源性VWF是阻止融合于半衰期延长物的嵌合蛋白的半衰期比野生型FVIII的半衰期长约两倍的半衰期限制因素。因此，本发明涉及使用包含D'结构域和D3结构域的VWF蛋白（例如VWF片段）来防止或抑制内源性VWF与FVIII蛋白之间的相互作用，以及同时通过使用XTEN序列与Ig恒定区或其一部分组合来增加所得FVIII蛋白的半衰期。特定来说，本发明显示较短XTEN序列（即在长度方面含有小于288个氨基酸的XTEN，即短于288个氨基酸的XTEN）在延长嵌合蛋白的半衰期方面更佳。

[0155] 在一个实施方案中，本发明涉及一种嵌合蛋白，其包含(i)第一多肽，所述第一多肽包含融合于第一Ig恒定区或其一部分的FVIII蛋白，和(ii)第二多肽，所述第二多肽包含含有VWF的D'结构域和D3结构域的VWF蛋白，所述VWF蛋白通过在之间的XTEN序列融合于第二Ig恒定区或其一部分，其中所述XTEN序列含有小于288个氨基酸残基，并且其中所述第一多肽与所述第二多肽连接或缔合。在另一实施方案中，第二多肽中的XTEN序列由长度在12个氨基酸与287个氨基酸之间的氨基酸序列组成。在其它实施方案中，相较于包含第一多肽和第二多肽的相应融合蛋白，其中第二多肽包含含有至少288个氨基酸的XTEN序列（例如AE288，例如SEQ ID NO:8），嵌合蛋白展现较长半衰期。在其它实施方案中，第二多肽中的XTEN序列含有至少约36、至少约42、至少约72或至少约144个氨基酸，但小于288个氨基酸，例如AE42、AE72、AE144（AE144、AE144\_2A、AE144\_3B、AE144\_4A、AE144\_5A、AE144\_6B）、AG42、AG72或AG144（AG144、AG144\_A、AG144\_B、AG144\_C、AG144\_F），例如SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:55；SEQ ID NO:56；SEQ ID NO:57；SEQ ID NO:58；SEQ ID NO:59；SEQ ID NO:14；SEQ ID NO:60；SEQ ID NO:61；SEQ ID NO:62；或SEQ ID NO:63。

[0156] 本发明的嵌合蛋白可进一步包含使FVIII蛋白与第一Ig恒定区或其一部分连接的第二XTEN序列。

[0157] 在某些实施方案中，本发明涉及一种嵌合蛋白，其包含(i)第一多肽，所述第一多肽包含融合于第一Ig恒定区或其一部分的FVIII蛋白，和(ii)第二多肽，所述第二多肽包含含有VWF的D'结构域和D3结构域的VWF蛋白，所述VWF蛋白通过在之间的第一XTEN序列融合于第二Ig恒定区或其一部分，其中所述XTEN序列含有小于288个氨基酸残基，并且其中所述第一多肽与所述第二多肽连接或缔合，并且其中所述第一多肽进一步包含插入在所述FVIII蛋白内一个或多个插入位点处或融合于所述FVIII蛋白和/或所述第一Ig恒定区或其一部分的第二XTEN序列。因此，在一个实施方案中，第二XTEN序列插入在FVIII蛋白内一个或多个插入位点处。在另一实施方案中，第二XTEN序列融合于FVIII蛋白和/或第一Ig恒定

区或其一部分。在其它实施方案中,第二XTEN序列插入在FVIII蛋白内一个或多个插入位点处,并且第三XTEN序列融合于FVIII蛋白和/或第一Ig恒定区或其一部分。

[0158] 第二XTEN序列和/或第三XTEN序列可为任何长度的XTEN氨基酸。举例来说,在本文其它地方公开第二XTEN序列和/或第三XTEN序列,例如AE42、AE72、AE864、AE576、AE288、AE144、AG864、AG576、AG288和AG144,例如SEQ ID NO:8;SEQ ID NO:9;SEQ ID NO:10;SEQ ID NO:11;SEQ ID NO:17;SEQ ID NO:54;SEQ ID NO:19;SEQ ID NO:16;SEQ ID NO:18;SEQ ID NO:15;SEQ ID NO:55;SEQ ID NO:56;SEQ ID NO:57;SEQ ID NO:58;SEQ ID NO:59;SEQ ID NO:14;SEQ ID NO:60;SEQ ID NO:61;SEQ ID NO:62;或SEQ ID NO:63。在特定实施方案中,第二XTEN序列和/或第三XTEN序列是AE288或AG288,例如SEQ ID NO:8或19。

[0159] 在某些实施方案中,本发明涉及一种嵌合蛋白,其包含(i)第一多肽,所述第一多肽包含通过任选接头融合于第一Ig恒定区或其一部分的FVIII蛋白,其中任选XTEN序列(X2)插入在所述FVIII蛋白内一个或多个插入位点处或融合于所述FVIII蛋白或所述第一Ig恒定区或其一部分,和(ii)第二多肽,所述第二多肽包含含有VWF的D'结构域和D3结构域的VWF蛋白,所述VWF蛋白通过在所述VWF蛋白与第二Ig恒定区或其一部分之间的XTEN序列(X1)融合于所述第二Ig恒定区或其一部分,其中所述XTEN序列(X1)含有小于288个氨基酸残基,并且通过接头融合于所述VWF蛋白,并且其中所述第一多肽和所述第二多肽相缔合。在一些实施方案中,本发明涉及一种嵌合蛋白,其包含(i)第一多肽,所述第一多肽包含通过任选接头融合于第一Ig恒定区或其一部分的FVIII蛋白,其中任选XTEN序列(X2)插入在所述FVIII蛋白内一个或多个插入位点处或融合于所述FVIII蛋白或所述第一Ig恒定区或其一部分,和(ii)第二多肽,所述第二多肽包含含有VWF的D'结构域和D3结构域的VWF蛋白,所述VWF蛋白通过在所述VWF蛋白与第二Ig恒定区或其一部分之间的XTEN序列(X1)融合于所述第二Ig恒定区或其一部分,其中所述XTEN序列(X1)含有小于288个氨基酸残基,并且通过接头融合于所述第二Ig恒定区或其一部分,并且其中所述第一多肽和所述第二多肽相缔合。在其它实施方案中,使XTEN序列(X1)与VWF蛋白或第二Ig恒定区或其一部分融合的接头是可裂解接头。在本文其它地方显示可裂解接头的非限制性实例。在特定实施方案中,接头是凝血酶可裂解接头。

[0160] 在一些实施方案中,嵌合蛋白是两个多肽链,第一链包含上述第一多肽,并且第二链包含上述第二多肽。举例来说,两个多肽链包含(i)第一链,其包含单链FVIII蛋白、第一Ig恒定区或其一部分以及插入在所述FVIII蛋白内一个或多个插入位点处或融合于所述FVIII蛋白或所述第一Ig恒定区或其一部分的任选XTEN序列,和(ii)第二链,其包含通过在之间的XTEN序列(X1)融合于第二Ig恒定区或其一部分的VWF蛋白,其中所述XTEN序列(X1)含有小于288个氨基酸。

[0161] 在某些实施方案中,嵌合蛋白是两个多肽链,第一链包含FVIII蛋白的重链,并且第二链从N末端至C末端包含FVIII蛋白的轻链、插入在所述FVIII蛋白内一个或多个插入位点处或融合于所述FVIII蛋白或第一Ig恒定区或其一部分的任选XTEN序列以及第一Ig恒定区或其一部分、任选接头(例如可加工接头)、VWF蛋白、XTEN序列(X1)、第二任选接头(例如可裂解接头)和第二Ig恒定区或其一部分。

[0162] 在其它实施方案中,嵌合蛋白是三个多肽链,(i)第一链包含FVIII蛋白的重链,(ii)第二链包含FVIII蛋白的轻链、第一Ig恒定区或其一部分以及插入在所述FVIII蛋白的



所述重链或所述轻链内一个或多个插入位点处或融合于所述FV1111蛋白或所述第一Ig恒定区或其一部分的任选XTEN序列,以及(iii)第三链包含通过在之间的XTEN序列(X1)融合于第二Ig恒定区或其一部分的VWF蛋白,其中所述第一链和所述第二链通过例如金属键的非共价键缔合,并且所述第二链和所述第三链通过例如二硫键的共价键缔合。

[0163] 在其它实施方案中,嵌合蛋白是单链,其从N末端至C末端包含单链FV1111蛋白、插入在所述FV1111蛋白内一个或多个插入位点处或融合于所述FV1111蛋白或第一Ig恒定区或其一部分的任选XTEN序列以及第一Ig恒定区或其一部分、任选接头(例如可加工接头)、VWF蛋白、XTEN序列(X1)、第二任选接头(例如可裂解接头)和第二Ig恒定区或其一部分。

[0164] 在某些实施方案中,嵌合蛋白包含下式(a) - (hh)中的一个:

- [0165] (a) FV1111-F1:F2-L2-X-L1-V;
- [0166] (b) FV1111-F1:V-L1-X-L2-F2;
- [0167] (c) F1-FV1111:F2-L2-X-L1-V;
- [0168] (d) F1-FV1111:V-L1-X-L2-F2;
- [0169] (e) FV1111-X2-F1:F2-L2-X1-L1-V;
- [0170] (f) FV1111-X2-F1:V-L1-X1-L2-F2;
- [0171] (g) FV1111(X2)-F1:F2-L2-X1-L1-V;
- [0172] (h) FV1111(X2)-F1:V-L1-X1-L2-F2;
- [0173] (i) F1-X2-F1:F2-L2-X1-L1-V;
- [0174] (j) F1-X2-F1:V-L1-X1-L2-F2;
- [0175] (k) V-L1-X-L2-F2-L3-FV1111-L4-F1;
- [0176] (l) V-L1-X-L2-F2-L3-F1-L4-FV1111;
- [0177] (m) F1-L4-FV1111-L3-F2-L2-X-L1-V;
- [0178] (n) FV1111-L4-F1-L3-F2-L2-X-L1-V;
- [0179] (o) FV1111-L4-F1-L3-V-L1-X-L2-F2;
- [0180] (p) FV1111-L4-F1-L3-F2-L2-X-L1-V;
- [0181] (q) F2-L2-X-L1-V-L3-F1-L4-FV1111;
- [0182] (r) F2-L2-X-L1-V-L3-FV1111-L4-F1;
- [0183] (s) V-L1-X1-L2-F2-L3-FV1111(X2)-L4-F1;
- [0184] (t) V-L1-X1-L2-F2-L3-F1-L4-FV1111(X2);
- [0185] (u) F1-L4-FV1111(X2)-L3-F2-L2-X1-L1-V;
- [0186] (v) F-L4-FV1111(X2)-L3-V-L1-X1-L2-F2;
- [0187] (w) FV1111(X2)-L4-F1-L3-V-L1-X1-L2-F2;
- [0188] (x) FV1111(X2)-L4-F1-L3-F2-L2-X1-L1-V;
- [0189] (y) F2-L2-X1-L1-V-L3-F1-L4-FV1111(X2);
- [0190] (z) F2-L2-X1-L1-V-L3-FV1111(X2)-L4-F1;
- [0191] (aa) V-L1-X2-L2-F2-L3-FV1111-L4-X2-L5-F1;
- [0192] (bb) V-L1-X2-L2-F2-L3-F1-L5-X2-L4-FV1111;
- [0193] (cc) F1-L5-X2-L4-FV1111-L3-F2-L2-X2-L1-V;
- [0194] (dd) F1-L5-X2-L4-FV1111-L3-V-L1-X2-L2-F2;

- [0195] (ee)FVIII-L5-X2-L4-F2-L3-V-L1-X1-L2-F1;
- [0196] (ff)FVIII-L5-X2-L4-F2-L3-F1-L2-X1-L1-V;
- [0197] (gg)F1-L2-X1-L1-V-L3-F2-L4-X2-L5-FVIII;或
- [0198] (hh)F1-L2-X1-L1-V-L3-FVIII-L5-X2-L4-F2;
- [0199] 其中V是VWF蛋白,其包含D'结构域和D3结构域,
- [0200] X或X1是含有小于288个氨基酸的第一XTEN序列,
- [0201] X2是第二XTEN序列,
- [0202] FVIII包括FVIII蛋白,
- [0203] FVIII (X2) 包括具有插入在FVIII蛋白内一个或多个插入位点中的第二XTEN序列的FVIII蛋白,
- [0204] F1是第一Ig恒定区或其一部分,
- [0205] F2是第二Ig恒定区或其一部分,
- [0206] L1、L2、L3、L4或L5是任选接头,
- [0207] (-)是肽键;并且
- [0208] (:)是共价键或非共价键。
- [0209] 在一个实施方案中,X或X1由长度在12个氨基酸与287个氨基酸之间的氨基酸序列组成。在另一实施方案中,相较于包含其中X或X1是AE288(例如SEQ ID NO:8)的式的相应融合蛋白,嵌合蛋白展现较长半衰期。
- [0210] 在其它实施方案中,式中X或X1含有至少约36、至少约42、至少约72或至少约144个氨基酸,但小于288个氨基酸,例如AE42、AE72、AE144(AE144、AE144\_2A、AE144\_3B、AE144\_4A、AE144\_5A、AE144\_6B)、AG42、AG72或AG144(AG144、AG144\_A、AG144\_B、AG144\_C、AG144\_F),例如SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:55;SEQ ID NO:56;SEQ ID NO:57;SEQ ID NO:58;SEQ ID NO:59;SEQ ID NO:14;SEQ ID NO:60;SEQ ID NO:61;SEQ ID NO:62;或SEQ ID NO:63。
- [0211] 在其它实施方案中,X2包含长度是至少约36个氨基酸、至少42个氨基酸、至少144个氨基酸、至少288个氨基酸、至少576个氨基酸或至少864个氨基酸的氨基酸序列,例如AE42、AE72、AE864、AE576、AE288、AE144、AG864、AG576、AG288或AG144,例如SEQ ID NO:9;SEQ ID NO:10;SEQ ID NO:15;SEQ ID NO:16;SEQ ID NO:8;SEQ ID NO:11;SEQ ID NO:17;SEQ ID NO:18、SEQ ID NO:19或SEQ ID NO:14。在特定实施方案中,X2是AE288或AG288,例如SEQ ID NO:8或19。
- [0212] 在某些实施方案中,相较于无X或X1和/或X2的嵌合蛋白,包含X或X1和/或X2的嵌合蛋白具有延长的半衰期。在其它实施方案中,L1和/或L2是可裂解接头。在其它实施方案中,L4和/或L5是可裂解接头。
- [0213] II.A. 范威尔邦德因子(VWF)蛋白
- [0214] VWF(也称为F8VWF)是一种存在于血浆中,并且在内皮(在怀布尔-帕拉德体(Weibel-Palade body)中)、巨核细胞(血小板的 $\alpha$ -颗粒)和内皮下结缔组织中组成性产生的大型多聚糖蛋白。基本VWF单体是具有2813个氨基酸的蛋白质。每个单体都含有许多具有特定功能的特定结构域,即D'/D3结构域(其结合因子VIII)、A1结构域(其结合血小板GPIIb受体、肝素和/或可能胶原蛋白)、A3结构域(其结合胶原蛋白)、C1结构域(其中RGD结构域在

这个结构域被活化时结合血小板整合素(integrin)αIIbβ3)和在蛋白质的C末端的“半胱氨酸结(cysteine knot)”结构域(VWF与血小板源性生长因子(PDGF)、转化生长因子-β(TGFβ)和β-人绒毛膜促性腺激素(β-human chorionic gonadotropin,βHCG)共有所述结构域)。

[0215] 在一个实施方案中,VWF蛋白是VWF片段。如本文所用的术语“VWF片段”包括但不限于能够抑制内源性VWF结合FVIII的包含D’结构域和D3结构域的功能性VWF片段。在一个实施方案中,VWF片段结合FVIII蛋白。在另一实施方案中,VWF片段阻断FVIII蛋白上的VWF结合位点,由此抑制FVIII蛋白与内源性VWF的相互作用。VWF片段包括保留VWF的这些活性的衍生物、变体、突变体或类似物。

[0216] 人VWF的2813单体氨基酸序列在Genbank中以登记号NP\_000543.2加以报道。编码人VWF的核苷酸序列在Genbank中以登记号NM\_000552.3加以报道。人VWF的核苷酸序列指定为SEQ ID NO:20。SEQ ID NO:21是全长VWF的氨基酸序列。VWF的各结构域列于表1中。

[0217] 表1.VWF序列

VWF 结构域	氨基酸序列
VWF 信号肽 (SEQ ID NO: 21 的 氨基酸 1 至 22)	1 <u>MIPARFAGVL LALALILPGT LC</u> 22
VWF DID2 区 (SEQ ID NO: 21 的 氨基酸 23 至 763)	23 AEGTRGRS STARCSLFGS DFVNTFDGSM 51 YSFAGYCSYL LAGGCQKRSF SIIGDFQNGK RVSLSVYLGE FFDIHLFVNG 101 TVTQGDQRVS MPYASKGLYL ETEAGYYKLS GEAYGFVARI DGSGNFQVLL 151 SDRYFNKTCG LCGNFNIFAE DDFMTQEGTL TSDPYDFANS WALSSGEQWC 201 ERASPPSSC NISSGEMQKG LWEQCQLLKS TSVFARCHPL VDPEPFVALC 251 EKTLCECAGG LECACPALLE YARTCAQEGM VLYGWDHSA CSPVCPAGME 301 YRQCVSPCAR TCQSLHINEM CQERCVDGCS CPEGQLLEDEG LCVESTECPC 351 VHSGKRYPPG TSLSRDCNTC ICRNSQWICS NEECPGECLV TGQSHFKSFD 401 NRYFTFSGIC QYLLARDCQD HSFIVIVETV QCADDRDAVC TRSVTVRLPG 451 LHNSLVKLVK GAGVAMDGQD IQLPLLKGD LRIQHTVTASV RLSYGEDLQM 501 DWDGRGRLLV KLSPVYAGKT CGLCGNYNGN QGDDFLTPSG

[0218]

[0219]

	551	LAEP RVEDFG NAWK LHGDCQ DLQK QHSDPC ALNPRMTRFS EEACAVLTSP TFEACHRAVS		
	601	PLPYLRNCRY DVCS CSDGRE CLCGALASYA AACAGRGVRV AWREPGRCEL		
	651	NCPKGQVYLQ CGTPCNLTCR SLSYPDEECN EACLEGCFCP PGLYMDERGD		
	701	CVPKAQCPCY YDGEIFQPED IFSDHHTMCY CEDGFMHCTM SGVPGSLLPD		
	751	AVLSSPLSHR SKR		763
VWF D' 结构域	764		SLSCRPP MVKLVCPADN LRAEGLECTK	
	801	<u>TCQNYDLECM</u> <u>SMGCVSGCLC</u> PPGMVRHENR CVALERCPCF HQGKEYAPGE <u>TVKIGCNTCV</u>		
	851	<u>CRDRKWNCTD</u> HVC DAT		866
VWF D3 结构域	867		CSTI GMAHYLTFDG LKYLFPGEQ	
	901	<u>YVLVQDYCGS</u> <u>NPGTFRILVG</u> NKGCSHPSVK CKKRVTILVE GGEIELFDGE <u>VNVKRPMKDE</u>		
	951	<u>THFEVVEGR</u> YIILLGKAL SVVWDRHLSI SVVLKQTYQE <u>KVCGLCGNFD</u>		
	1001	<u>GIQNNDLTSS</u> NLQVEEDPVD FGNSWKVSSQ CADTRKVP LD <u>SSPATCHNNI</u>		
	1051	<u>MKQTMVDSSC</u> RILTSDFVD CNKLVDPPEY LDVCIYDTCS <u>CESIGDCACF</u>		
	1101	<u>CDTIAAYAHV</u> CAOHGKVVTW RTATLCPQSC EERNLRENGY <u>ECEWRYN SCA</u>		
	1151	<u>PACQVTCQHP</u> EPLACPVCV EGCHAHCPPG KILDELLOTC <u>VDPEDCPVCE</u>		
	1201	<u>VAGRRFASGK</u> KVTLNPSDPE HCQICHCDVV NLTCEACQEP		
	1240			
VWF A1 结构域	1241	GGLVVPPTDA		
	1251	PVSPTTLYVE DISEPPLHDF YCSRLLDLVF LLDGSSRLSE AEFEVLKAFV		
	1301	VDMMERLRIS QKWVRVAVVE YHDGSHAYIG LKDRKRPSSEL RRIASQVKYA		
	1351	GSQVASTSEV LKYTLFQIFS KIDRPEASRI ALLMASQEP QRMSRNFVRY		
	1401	VQGLKKKKVI VIPVGIGPHA NLKQIRLIEK QAPENKAFVL SSVDELEQQR		
	1451	DEIVSYLCDL APEAPPPTLP PDMAQVTVG		1479
	1480		P GLLGVSTLGP KRNSMVL DVA	
	1501	FVLEGS DKIG EADFNRSKEF MEEVIQRMDV GQDSIHVTVL QYSYMTVEY		
	1551	PFSEAQSKGD ILQRVREIRY QGGNRTNTGL ALRYLSDHSF LVSQGDREQA		1600
	1601	PNLVYMTGN PASDEIKRLP GDIQVVP IGV GPANANVQELE RIGWPNAPIL		
	1651	IQDFETLPRE APDLVLQRCC SGEGLQIPTL SPAPDCSQPL DVILLLDGSS		
	1701	SFPASYFDEM KSF AKAFISK ANIGPRLTQV SVLQYGSITT IDVPWNVVPE		
	1751	KAHLLSLVDV MQREGGPSQI GDALGFAVRY LTSEM HGARP GASKAVVILV		
	1801	TDVSVDSVDA AADAARSNRV TVFP IIGDR YDAAQLRILA GPAGDSNVVK		

[0220]

1851	LQRIEDLPTM VTLGNSFLHK LCSGFVTRICM DEDGNEKRPG DWTLPDQCH
1901	TVTCQPDGQT LLKSHRVNCD RGLRPPSPNS QSPVKVEETC GCRWTCPCVC
1951	TGSSTRHIVT FDGQNFKLTG SCSYVLFQNK EQDLEVILHN GACSPGARQG
2001	CMKSIEVKHS ALSVEXHSDM EVTVNGLRVS VPYVGGNMEV NVYGAIMHEV
2051	RFNHLGHIFT FTPQNNFQL QLSPKTFASK TYGLCGICDE NGANDFMLRD
2101	GTVTTDWKTLL VQEWTVQRPQ QTCQPILEEQ CLVPDSSHCQ VLLLPLFAEC
2151	HKVLAPATFY AICQQDSCHQ EQVCEVIASY AHLCRTNGVC VDWRTPDFCA
2201	MSCPPSLVYN HCEHGCPRHC DGNVSSCGDH PSEGCFPCPD KVMLEGSCVP
2251	EEACTQCIGE DGVQHQFLEA WVPDHQPCQI CTCLSGRKVN CTTQPCPTAK
2301	APTCGLCEVA RLRQNADQCC PEYECVCDPV SCDLPPVPHC ERGLQPTLTN
2351	PGECRPNFTC ACRKEECKRV SPPSCPPHRL PTLRKTQCCD EYECACNCVN
2401	STVSCPLGYL ASTATNDCGC TTTTCLPDKV CVHRSTIYPV GQFWEEGCDV
2451	CTCTDMEDAV MGLRVAQCSQ KPCEDSCRSG FTYVLHEGEC CGRCLPSACE
2501	VVTGSPRGDS QSSWKSVMGQ WASPENPCLI NECVRVKEEV FIQQRNVSCP
2551	QLEVPVPCPSG FQLSCKTSAC CPSCRCERME ACMLNGTVIG PGKTVMIDVC
2601	TTCRCMVQVG VISGFKLECR KTTCNPCPLG YKEENNTGEC CGRCLPTACT
2651	IQLRGGQIMT LKRDETLQDG CDTHFCKVNE RGEYFWEKRV TGCPPFDEHK
2701	CLAEGGKIMK IPGTCCDTCE EPECNDITAR LQYVKVGSCK SEVEVDIHYC
2751	QGKCASKAMY SIDINDVQDQ CSCCSPTRTE PMQVALHCTN GSVVYHEVLN
2801	AMECKCSPRK CSK
<b>核苷酸序列(SEQ ID NO: 20)</b>	
全长 VWF	1 ATGATTCTCG CCAGATTGCG CGGGGTGCTG CTGCTCTGG CCCTCATTTC 51 GCCAGGGACC CTTTGTGCAG AAGGAACTCG CGGCAGGTCA TCCACGGCCC 101 GATGCAGCCT TTTCGGAAGT GACTTCGTCA ACACCTTTGA TGGGAGCATG 151 TACAGCTTTG CGGGATACTG CAGTTACCTC CTGGCAGGGG GCTGCCAGAA 201 ACGCTCCTTC TCGATTATTG GGGACTTCCA GAATGGCAAG AGAGTGAGCC 251 TCTCCGTGTA TCTTGGGGAA TTTTTTGACA TCCATTTGTT TGTCATGGT 301 ACCGTGACAC AGGGGGACCA AAGAGTCTCC ATGCCCTATG CCTCCAAAGG 351 GCTGTATCTA GAAACTGAGG CTGGGTACTA CAAGCTGTCC GGTGAGGCCT 401 ATGGCTTTGT GGCCAGGATC GATGGCAGCG GCAACTTTCA AGTCCTGCTG 451 TCAGACAGAT ACTTCAACAA GACCTGCGGG CTGTGTGGCA ACTTTAACAT

[0221]

501	CTTTGCTGAA	GATGACTTTA	TGACCCAAGA	AGGGACCTTG
	ACCTCGGACC			
551	CTTATGACTT	TGCCAACTCA	TGGGCTCTGA	GCAGTGGAGA
	ACAGTGGTGT			
601	GAACGGGCAT	CTCCTCCCAG	CAGCTCATGC	AACATCTCCT
	CTGGGGAAAT			
651	GCAGAAGGGC	CTGTGGGAGC	AGTGCCAGCT	TCTGAAGAGC
	ACCTCGGTGT			
701	TTGCCCGCTG	CCACCCTCTG	GTGGACCCCG	AGCCTTTTGT
	GGCCCTGTGT			
751	GAGAAGACTT	TGTGTGAGTG	TGCTGGGGGG	CTGGAGTGCG
	CCTGCCCTGC			
801	CCTCCTGGAG	TACGCCCGGA	CCTGTGCCCA	GGAGGGAATG
	GTGCTGTACG			
851	GCTGGACCGA	CCACAGCGCG	TGCAGCCCG	TGTGCCCTGC
	TGGTATGGAG			
901	TATAGGCAGT	GTGTGTCCCC	TTGCGCCAGG	ACCTGCCAGA
	GCCTGCACAT			
951	CAATGAAATG	TGTCAGGAGC	GATGCGTGGA	TGGCTGCAGC
	TGCCCTGAGG			
1001	GACAGCTCCT	GGATGAAGGC	CTCTGCGTGG	AGAGCACCGA
	GTGTCCCTGC			
1051	GTGCATTCCG	GAAAGCGCTA	CCCTCCCGGC	ACCTCCCTCT
	CTCGAGACTG			
1101	CAACACCTGC	ATTTGCCGAA	ACAGCCAGTG	GATCTGCAGC
	AATGAAGAAT			
1151	GTCCAGGGGA	GTGCCTTGTC	ACTGGTCAAT	CCCACCTCAA
	GAGCTTTGAC			
1201	AACAGATACT	TCACCTTCAG	TGGGATCTGC	CAGTACCTGC
	TGGCCCGGGA			
1251	TTGCCAGGAC	CACTCCTTCT	CCATTGTCAT	TGAGACTGTC
	CAGTGTGCTG			
1301	ATGACCGCGA	CGCTGTGTGC	ACCCGCTCCG	TCACCGTCCG
	GCTGCCTGGC			
1351	CTGCACAACA	GCCTTGTA	ACTGAAGCAT	GGGGCAGGAG
	TTGCCATGGA			
1401	TGGCCAGGAC	ATCCAGCTCC	CCCTCTGAA	AGGTGACCTC
	CGCATCCAGC			
1451	ATACAGTGAC	GGCCTCCGTG	CGCCTCAGCT	ACGGGGAGGA
	CCTGCAGATG			
1501	GACTGGGATG	GCCGCGGGAG	GCTGCTGGTG	AAGCTGTCCC
	CCGTCTATGC			
1551	CGGGAAGACC	TGCGGCCTGT	GTGGGAATTA	CAATGGCAAC
	CAGGGCGACG			
1601	ACTTCCTTAC	CCCCTCTGGG	CTGGCRGAGC	CCCGGTGGA
	GGACTTCGGG			
1651	AACGCCTGGA	AGCTGCACGG	GGACTGCCAG	GACCTGCAGA
	AGCAGCACAG			
1701	CGATCCCTGC	GCCCTCAACC	CGCGCATGAC	CAGGTCTCC
	GAGGAGGCGT			
1751	GCGCGGTCT	GACGTCCCC	ACATTGAGG	CCTGCCATCG
	TGCCGTCAGC			
1801	CCGCTGCCCT	ACCTGCGGAA	CTGCCGCTAC	GACGTGTGCT
	CCTGCTCGGA			
1851	CGGCCGCGAG	TGCCTGTGCG	GCGCCCTGGC	CAGCTATGCC
	GCGGCCTGCG			
1901	CGGGGAGAGG	CGTGCGCGTC	GCGTGGCGCG	AGCCAGGCCG
	CTGTGAGCTG			
1951	AACTGCCCGA	AAGGCCAGGT	GTACCTGCAG	TGCGGGACCC
	CCTGCAACCT			
2001	GACCTGCCCG	TCTCTCTCTT	ACCCGGATGA	GGAATGCAAT
	GAGGCCTGCC			
2051	TGGAGGGCTG	CTTCTGCCCC	CCAGGGCTCT	ACATGGATGA
	GAGGGGGGAC			
2101	TGCGTGCCCA	AGGCCAGTG	CCCCTGTTAC	TATGACGGTG
	AGATCTTCCA			

[0222]

2151	GCCAGAAGAC	ATCTTCTCAG	ACCATCACAC	CATGTGCTAC
	TGTGAGGATG			
2201	GCTTCATGCA	CTGTACCATG	AGTGGAGTCC	CCGGAAGCTT
	GCTGCCTGAC			
2251	GCTGTCCTCA	GCAGTCCCCT	GTCTCATCGC	AGCAAAAGGA
	GCCTATCCTG			
2301	TCGGCCCCCC	ATGGTCAAGC	TGGTGTGTCC	CGCTGACAAC
	CTGCGGGCTG			
2351	AAGGGCTCGA	GTGTACCAAA	ACGTGCCAGA	ACTATGACCT
	GGAGTGCATG			
2401	AGCATGGGCT	GTGTCTCTGG	CTGCCTCTGC	CCCCGGGCA
	TGGTCCGGCA			
2451	TGAGAACAGA	TGTGTGGCCC	TGAAAAGGTG	TCCCTGCTTC
	CATCAGGGCA			
2501	AGGAGTATGC	CCCTGGAGAA	ACAGTGAAGA	TTGGCTGCAA
	CACTGTGTGC			
2551	TGTCGGGACC	GGAAGTGAA	CTGCACAGAC	CATGTGTGTG
	ATGCCACGTG			
2601	CTCCACGATC	GGCATGGCCC	ACTACCTCAC	CTTCGACGGG
	CTCAAATACC			
2651	TGTTCCCCGG	GGAGTGCCAG	TACGTTCTGG	TGCAGGATTA
	CTGCGGCAGT			
2701	AACCCTGGGA	CCTTTCGGAT	CCTAGTGGGG	AATAAGGGAT
	GCAGCCACCC			
2751	CTCAGTGAAA	TGCAAGAAAC	GGGTCACCAT	CCTGGTGGAG
	GGAGGAGAGA			
2801	TTGAGCTGTT	TGACGGGGAG	GTGAATGTGA	AGAGGCCCAT
	GAAGGATGAG			
2851	ACTCACTTTG	AGGTGGTGGG	GTCTGGCCGG	TACATCATTC
	TGCTGCTGGG			
2901	CAAAGCCCTC	TCCGTGGTCT	GGGACCGCCA	CCTGAGCATC
	TCCGTGGTCC			
2951	TGAAGCAGAC	ATACCAGGAG	AAAGTGTGTG	GCCTGTGTGG
	GAATTTTGAT			
3001	GGCATCCAGA	ACAATGACCT	CACCAGCAGC	AACCTCCAAG
	TGGAGGAAGA			
3051	CCCTGTGGAC	TTTGGGAACT	CCTGGAAAGT	GAGCTCGCAG
	TGTGCTGACA			
3101	CCAGAAAAGT	GCCTCTGGAC	TCATCCCCTG	CCACCTGCCA
	TAACAACATC			
3151	ATGAAGCAGA	CGATGGTGGG	TTCCTCCTGT	AGAATCCTTA
	CCAGTGACGT			
3201	CTTCCAGGAC	TGCAACAAGC	TGGTGGACCC	CGAGCCATAT
	CTGGATGTCT			
3251	GCATTACGA	CACCTGCTCC	TGTGAGTCCA	TTGGGGACTG
	CGCCTGCTTC			
3301	TGCGACACCA	TTGCTGCCTA	TGCCACGTG	TGTGCCCAGC
	ATGGCAAGGT			
3351	GGTGACCTGG	AGGACGGCCA	CATTGTGCC	CCAGAGCTGC
	GAGGAGAGGA			
3401	ATCTCCGGGA	GAACGGGTAT	GAGTGTGAGT	GGCGCTATAA
	CAGCTGTGCA			
3451	CCTGCCTGTC	AAGTCACGTG	TCAGCACCTT	GAGCCACTGG
	CCTGCCCTGT			
3501	GCAGTGTGTG	GAGGGCTGCC	ATGCCCACTG	CCCTCCAGGG
	AAAATCCTGG			
3551	ATGAGCTTTT	GCAGACCTGC	GTTGACCCTG	AAGACTGTCC
	AGTGTGTGAG			
3601	GTGGCTGGCC	GGCGTTTTCG	CTCAGGAAAG	AAAGTCACCT
	TGAATCCCAG			
3651	TGACCCTGAG	CACTGCCAGA	TTTGCCACTG	TGATGTTGTG
	AACCTCACCT			
3701	GTGAAGCCTG	CCAGGAGCCG	GGAGGCCTGG	TGGTGCCTCC
	CACAGATGCC			
3751	CCGGTGAGCC	CCACCACTCT	GTATGTGGAG	GACATCTCGG
	AACCGCCGTT			

[0223]

3801	GCACGATTTCTACTGCAGCAGGCTACTGGA	CCTGGTCTTCCTGCTGGATG
3851	GCTCCTCCAGGCTGTCCGAGGCTGAGTTG	AAGTGCTGAA
3901	GTGGACATGTGGAGCGGCTGCGCATCTCC	CAGAAGTGGG
3951	CGTGGTGGAGTACCACGACGCTCCCACGC	CTACATCGGG
4001	GGAAGCGACGTCAGAGCTGCGGCGATTG	CCAGCCAGGT
4051	GGCAGCCAGGTGGCTCCACAGCGAGGTC	TTGAAATACA
4101	AATCTTCAGCAAGATCGACGCCCTGAAGC	CTCCCGCATC
4151	TGATGGCCAGCCAGGAGCCC	CAACGGATGT
4201	GTCCAGGGCC	TGAAGAAGAA
4251	GCCCCATGCCAACCTCAAGC	AGATCCGCCT
4301	AGAACAAGGCTTTCGTGCTG	AGCAGTGTGG
4351	GACGAGATCGTTAGCTACCT	CTGTGACCTT
4401	TACTCTGCCCCCGACATGG	CACAAGTCAC
4451	GGGTTTCGACCCTGGGGCCC	AAGAGGAACT
4501	TTCGTCCTGGAAGGATCGGA	CAAAATTGGT
4551	CAAGGAGTTCATGGAGGAG	TGATTCAGCG
4601	GCATCCACGTCACGGTGCTG	CAGTACTCCT
4651	CCCTCAGCGAGGCACAGTC	CAAAGGGGAC
4701	GATCCGCTAC	CAGGGCGGCA
4751	ACCTCTCTGACCACAGCTTC	TTGGTCAGCC
4801	CCCAACCTGTTCTACATGGT	CACCGGAAAT
4851	GAGGCTGCCTGGAGACATCC	AGGTGGTGCC
4901	CCAACGTGCA	GGAGCTGGAG
4951	ATCCAGGACTTTGAGACGCT	CCCCGAGAG
5001	GAGGTGCTGCTCCGAGAGG	GGCTGCAGAT
5051	CTGACTGCAGCCAGCCCCTG	GACGTGATCC
5101	AGTTTCCCAGCTTCTTATTT	TGATGAAATG
5151	CATTTCAAAA	GCCAATATAG
5201	AGTATGGAAGCATCACCACC	ATTGACGTGC
5251	AAAGCCCATTTGCTGAGCCT	TGTGGACGTC
5301	CAGCCAAATCGGGATGCCT	TGGGCTTTGC
5351	AAATGCATGGTGCCAGGCCG	GGAGCCTCAA
5401	ACGGACGTCTCTGTGGATTC	AGTGGATGCA



[0224]

5451	CAACAGAGTG	ACAGTGTTC	CTATTGGAAT	TGGAGATCGC
	TACGATGCAG			
5501	CCCAGCTACG	GATCTTGGCA	GGCCCAGCAG	GCGACTCCAA
	CGTGGTGAAG			
5551	CTCCAGCGAA	TCGAAGACCT	CCCTACCATG	GTCACCTTGG
	GCAATTCCTT			
5601	CCTCCACAAA	CTGTGCTCTG	GATTTGTTAG	GATTTGCATG
	GATGAGGATG			
5651	GGAATGAGAA	GAGGCCCGGG	GACGTCTGGA	CCTTGCCAGA
	CCAGTGCCAC			
5701	ACCGTGACTT	GCCAGCCAGA	TGGCCAGACC	TTGCTGAAGA
	GTCATCGGGT			
5751	CAACTGTGAC	CGGGGGCTGA	GGCCTTCGTG	CCCTAACAGC
	CAGTCCCCTG			
5801	TTAAAGTGGG	AGAGACCTGT	GGCTGCCGCT	GGACCTGCC
	CTGYGTGTGC			
5851	ACAGGCAGCT	CCACTCGGCA	CATCGTGACC	TTTGATGGGC
	AGAAATTCAA			
5901	GCTGACTGGC	AGCTGTTCTT	ATGTCCTATT	TCAAAAACAAG
	GAGCAGGACC			
5951	TGGAGGTGAT	TCTCCATAAT	GGTGCCTGCA	GCCCTGGAGC
	AAGGCAGGGC			
6001	TGCATGAAAT	CCATCGAGGT	GAAGCACAGT	GCCCTCTCCG
	TCGAGSTGCA			
6051	CAGTGACATG	GAGGTGACGG	TGAATGGGAG	ACTGGTCTCT
	GTTCTTACG			
6101	TGGGTGGGAA	CATGGAAAGT	AACGTTTATG	GTGCCATCAT
	GCATGAGGTC			
6151	AGATTCAAATC	ACCTTGGTCA	CATCTTCACA	TTCACTCCAC
	AAAACAATGA			
6201	GTTCCAATG	CAGCTCAGCC	CCAAGACTTT	TGCTTCAAAG
	ACGTATGGTC			
6251	TGTGTGGGAT	CTGTGATGAG	AACGGAGCCA	ATGACTTCAT
	GCTGAGGGAT			
6301	GGCACAGTCA	CCACAGACTG	GAAAACACTT	GTTCAGGAAT
	GGACTGTGCA			
6351	GCGGCCAGGG	CAGACGTGCC	AGCCCATCCT	GGAGGAGCAG
	TGTCCTGTCC			
6401	CCGACAGCTC	CCACTGCCAG	GTCCCTCTCT	TACCACTGTT
	TGCTGAATGC			
6451	CACAAGGTCC	TGGCTCCAGC	CACATTCTAT	GCCATCTGCC
	AGCAGGACAG			
6501	TTGCCACCAG	GAGCAAGTGT	GTGAGGTGAT	CGCCTCTTAT
	GCCCACCTCT			
6551	GTCGGACCAA	CGGGGTCTGC	GTTGACTGGA	GGACACCTGA
	TTTCTGTGCT			
6601	ATGTCAATGCC	CACCATCTCT	GGTCTACAAC	CACTGTGAGC
	ATGGCTGTCC			
6651	CCGGCACTGT	GATGGCAACG	TGAGCTCCTG	TGGGGACCAT
	CCCTCCGAAG			
6701	GCTGTTTCTG	CCCTCCAGAT	AAAGTCATGT	TGGAAGGCAG
	CTGTGTCCCT			
6751	GAAGAGGCCT	GCACTCAGTG	CATTGGTGAG	GATGGAGTCC
	AGCACCAGTT			
6801	CCTGGAAGCC	TGGGTCCCGG	ACCACCAGCC	CTGTCAGATC
	TGCACATGCC			
6851	TCAGCGGGCG	GAAGGTCAAC	TGCACAACGC	AGCCCTGCC
	CACGGCCAAA			
6901	GCTCCCACGT	GTGGCCTGTG	TGAAGTAGCC	CGCCTCCGCC
	AGAATGCAGA			
6951	CCAGTGCTGC	CCCGAGTATG	AGTGTGTGTG	TGACCCAGTG
	AGCTGTGACC			
7001	TGCCCCAGT	GCCTCACTGT	GAACGTGGCC	TCCAGCCCAC
	ACTGACCAAAC			
7051	CCTGGCGAGT	GCAGACCCAA	CTTCACCTGC	GCCTGCAGGA
	AGGAGGAGTG			

[0225]

7101	CAAAAGAGTG	TCCCCACCCT	CCTGCCCCCC	GCACCGTTTG
	CCCACCCTTC			
7151	GGAAGACCCA	GTGCTGTGAT	GAGTATGAGT	GTGCCGTGCAA
	CTGTGTCAAC			
7201	TCCACAGTGA	GCTGTCCCCT	TGGGTACTION	GCCTCAACCG
	CCACCAATGA			
7251	CTGTGGCTGT	ACCACAACCA	CCTGCCCTCC	CGACAAGGTG
	TGTGTCCACC			
7301	GAAGACCCAT	CTACCCTGTG	GGCCAGTTCT	GGGAGGAGGG
	CTGCGATGTG			
7351	TGCACCTGCA	CCGACATGGA	GGATGCCGTG	ATGGGCCCTCC
	GCGTGGCCCA			
7401	GTGCTCCCAG	AAGCCCTGTG	AGGACAGCTG	TCGGTCGGGC
	TTCACTTACG			
7451	TTCTGCATGA	AGGCGAGTGC	TGTGGAAGGT	GCCTGCCATC
	TGCCGTGAG			
7501	GTGGTGACTG	GCTCACCGCG	GGGGGACTCC	CAGTCTTCCT
	GGAAGAGTGT			
7551	CGGCTCCCAG	TGGGCCTCCC	CGGAGAACCC	CTGCCCTCATC
	AATGAGTGTG			
7601	TCCGAGTGAA	GGAGGAGGTC	TTTATACAAC	AAAGGAACGT
	CTCCTGCCCC			
7651	CAGCTGGAGG	TCCCTGTCTG	CCCCTCGGGC	TTTCAGCTGA
	GCTGTAAGAC			
7701	CTCAGCGTGC	TGCCCAAGCT	GTCGCTGTGA	GCGCATGGAG
	GCCTGCATGC			
7751	TCAATGGCAC	TGTCATTGGG	CCCGGAAGA	CTGTGATGAT
	CGATGTGTGC			
7801	ACGACCTGCC	GCTGCATGGT	GCAGGTGGGG	GTCATCTCTG
	GATTCAAGCT			
7851	GGAGTGCAGG	AAGACCACCT	GCAACCCCTG	CCCCCTGGGT
	TACAAGGAAG			
7901	AAAATAACAC	AGGTGAATGT	TGTGGGAGAT	GTTTGCCTAC
	GGCTTGCACC			
7951	ATTCAGCTAA	GAGGAGGACA	GATCATGACA	CTGAAGCGTG
	ATGAGACGCT			
8001	CCAGGATGGC	TGTGATACTC	ACTTCTGCAA	GGTCAATGAG
	AGAGGAGAGT			
8051	ACTTCTGGGA	GAAGAGGGTC	ACAGGCTGCC	CACCCTTTGA
	TGAACACAAG			
8101	TGTCTTGCTG	AGGGAGGTAA	AATTATGAAA	ATTCCAGGCA
	CCTGCTGTGA			
8151	CACATGTGAG	GAGCCTGAGT	GCAACGACAT	CACTGCCAGG
	CTGCAGTATG			
8201	TCAAGGTGGG	AAGCTGTAAG	TCTGAAGTAG	AGGTGGATAT
	CCACTACTGC			
8251	CAGGGCAAAT	GTGCCAGCAA	AGCCATGTAC	TCCATTGACA
	TCAACGATGT			
8301	GCAGGACCAG	TGCTCCTGCT	GCTCTCCGAC	ACGACCGGAG
	CCCATGCAGG			
8351	TGGCCCTGCA	CTGCACCAAT	GGCTCTGTTG	TGTACCATGA
	GGTTCTCAAT			
8401	GCCATGGAGT	GCAAATGCTC	CCCAGGAAG	TGCAGCAAGT GA

[0226] 如本文所用的VWF蛋白可为包含VWF的D'结构域和D3结构域的VWF片段,其中所述VWF片段结合因子VIII(FVIII),并且抑制内源性VWF(全长VWF)结合FVIII。包含D'结构域和D3结构域的VWF片段可进一步包含选自由以下组成的组的VWF结构域: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、其一个或多个片段及其任何组合。在一个实施方案中,VWF片段包含以下、基本上由以下组成或由以下组成:(1)VWF的D'和D3结构域或其片段;(2)VWF的D1、D'和D3结构域或其片段;(3)VWF的D2、D'和D3结构域或其片段;(4)VWF的D1、D2、D'和D3结构域或其片段;或(5)VWF的D1、D2、D'、D3和A1结构域或其片段。本文所述的VWF片段不含有结合VWF清除受体的位点。在另一实施方案中,本文所述的VWF片段不是SEQ ID NO:21的氨基酸764至1274。本发明的VWF片段可包含连接于或融合于VWF片段的任何其它序列。举例来说,本文所

述的VWF片段可进一步包含信号肽。

[0227] 在一个实施方案中,包含D'结构域和D3结构域的VWF片段与FVIII蛋白结合或缔合。通过与FVIII蛋白结合或缔合,本发明的VWF片段保护FVIII免遭蛋白酶裂解和FVIII活化,使FVIII的重链和轻链稳定,并且防止FVIII被清道夫受体清除。在另一实施方案中,VWF片段与FVIII蛋白结合或缔合,并且阻断或防止FVIII蛋白结合磷脂和活化的蛋白质C。通过防止或抑制FVIII蛋白与内源性全长VWF结合,本发明的VWF片段降低FVIII被VWF清除受体清除,并且因此延长嵌合蛋白的半衰期。因此,嵌合蛋白的半衰期延长是由于缺乏VWF清除受体结合位点的VWF片段与FVIII蛋白结合或缔合以及由VWF片段遮蔽或保护FVIII蛋白免遭含有VWF清除受体结合位点的内源性VWF。结合于VWF片段或由VWF片段保护的FVIII蛋白也可允许FVIII蛋白再循环。通过消除全长VWF分子中含有的VWF清除路径受体结合位点,本发明的FVIII/VWF异二聚体被遮蔽免遭VWF清除路径,从而进一步延长FVIII半衰期。

[0228] 在一个实施方案中,适用于本发明的VWF蛋白包含VWF的D'结构域和D3结构域,其中所述D'结构域与SEQ ID NO:21的氨基酸764至866至少60%、70%、80%、85%、90%、95%、96%、97%、98%、99%或100%同一,其中所述VWF蛋白防止或抑制内源性VWF结合FVIII。在另一实施方案中,VWF蛋白包含VWF的D'结构域和D3结构域,其中所述D3结构域与SEQ ID NO:21的氨基酸867至1240至少60%、70%、80%、85%、90%、95%、96%、97%、98%、99%或100%同一,其中所述VWF蛋白防止或抑制内源性VWF结合FVIII。在一些实施方案中,本文所述的VWF蛋白包含以下、基本上由以下组成或由以下组成:VWF的D'结构域和D3结构域,所述结构域与SEQ ID NO:21的氨基酸764至1240至少60%、70%、80%、85%、90%、95%、96%、97%、98%、99%或100%同一,其中所述VWF蛋白防止或抑制内源性VWF结合FVIII。在其它实施方案中,VWF蛋白包含以下、基本上由以下组成或由以下组成:D1、D2、D'和D3结构域,所述结构域与SEQ ID NO:21的氨基酸23至1240至少60%、70%、80%、85%、90%、95%、96%、97%、98%、99%或100%同一,其中所述VWF蛋白防止或抑制内源性VWF结合FVIII。在其它实施方案中,VWF蛋白进一步包含可操作地与其连接的信号肽。

[0229] 在一些实施方案中,适用于本发明的VWF蛋白基本上由以下组成或由以下组成:(1) D'D3结构域、D1D'D3结构域、D2D'D3结构域或D1D2D'D3结构域和(2) 多达约10个氨基酸(例如SEQ ID NO:21的氨基酸764至1240至SEQ ID NO:21的氨基酸764至1250的任何序列)、多达约15个氨基酸(例如SEQ ID NO:21的氨基酸764至1240至SEQ ID NO:21的氨基酸764至1255的任何序列)、多达约20个氨基酸(例如SEQ ID NO:21的氨基酸764至1240至SEQ ID NO:21的氨基酸764至1260的任何序列)、多达约25个氨基酸(例如SEQ ID NO:21的氨基酸764至1240至SEQ ID NO:21的氨基酸764至1265的任何序列)或多达约30个氨基酸(例如SEQ ID NO:21的氨基酸764至1240至SEQ ID NO:21的氨基酸764至1260的任何序列)的另外的VWF序列。在特定实施方案中,包含D'结构域和D3结构域或基本上由D'结构域和D3结构域组成的VWF蛋白既不是SEQ ID NO:21的氨基酸764至1274也不是全长成熟VWF。在一些实施方案中,D1D2结构域以反式与D'D3结构域一起表达。在一些实施方案中,D1D2结构域以顺式与D'D3结构域一起表达。

[0230] 在其它实施方案中,包含连接于D1D2结构域的D'D3结构域的VWF蛋白进一步包含细胞内裂解位点(例如PACE(弗林蛋白酶)或PC5的裂解位点),从而允许在表达后,从D'D3结构域裂解D1D2结构域。细胞内裂解位点的非限制性实例在本文其它地方加以公开。

[0231] 在其它实施方案中,VWF蛋白包含D'结构域和D3结构域,但不包含选自以下组成的组的氨基酸序列:(1)对应于SEQ ID NO:21的氨基酸1241至2813,(2)对应于SEQ ID NO:21的氨基酸1270至氨基酸2813,(3)对应于SEQ ID NO:21的氨基酸1271至氨基酸2813,(4)对应于SEQ ID NO:21的氨基酸1272至氨基酸2813,(5)对应于SEQ ID NO:21的氨基酸1273至氨基酸2813,(6)对应于SEQ ID NO:21的氨基酸1274至氨基酸2813,及其任何组合。

[0232] 在其它实施方案中,本发明的VWF蛋白包含以下、基本上由以下组成或由以下组成:对应于D'结构域、D3结构域和A1结构域的氨基酸序列,其中所述氨基酸序列与SEQ ID NO:21的氨基酸764至1479至少60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或100%同一,其中所述VWF蛋白防止内源性VWF结合FVIII。在特定实施方案中,VWF蛋白不是SEQ ID NO:21的氨基酸764至1274。

[0233] 在一些实施方案中,本发明的VWF蛋白包含D'结构域和D3结构域,但不包含至少一个选自以下组成的组的VWF结构域:(1)A1结构域,(2)A2结构域,(3)A3结构域,(4)D4结构域,(5)B1结构域,(6)B2结构域,(7)B3结构域,(8)C1结构域,(9)C2结构域,(10)CK结构域,(11)CK结构域和C2结构域,(12)CK结构域、C2结构域和C1结构域,(13)CK结构域、C2结构域、C1结构域、B3结构域,(14)CK结构域、C2结构域、C1结构域、B3结构域、B2结构域,(15)CK结构域、C2结构域、C1结构域、B3结构域、B2结构域和B1结构域,(16)CK结构域、C2结构域、C1结构域、B3结构域、B2结构域、B1结构域和D4结构域,(17)CK结构域、C2结构域、C1结构域、B3结构域、B2结构域、B1结构域、D4结构域和A3结构域,(18)CK结构域、C2结构域、C1结构域、B3结构域、B2结构域、B1结构域、D4结构域、A3结构域和A2结构域,(19)CK结构域、C2结构域、C1结构域、B3结构域、B2结构域、B1结构域、D4结构域、A3结构域、A2结构域和A1结构域,和(20)其任何组合。

[0234] 在其它实施方案中,VWF蛋白包含D'D3结构域和一个或多个结构域或模块。所述结构域或模块的实例包括但不限于Zhour等,Blood2012年4月6日在线发表:DOI 10.1182/blood-2012-01-405134中公开的结构域和模块,所述文献以引用的方式整体并入本文。举例来说,VWF蛋白可包含D'D3结构域和一个或多个选自以下组成的组的结构域或模块:A1结构域、A2结构域、A3结构域、D4N模块、VWD4模块、C8-4模块、TIL-4模块、C1模块、C2模块、C3模块、C4模块、C5模块、C5模块、C6模块及其任何组合。

[0235] 在其它实施方案中,VWF蛋白连接于异源性部分,其中所述异源性部分连接于所述VWF蛋白的N末端或C末端或紧靠所述VWF蛋白中一个或多个氨基酸(例如一个或多个XTEN插入位点)的下游插入。举例来说,异源性部分在VWF蛋白中的插入位点可在D'结构域、D3结构域或两者中。异源性部分可为半衰期延长物。

[0236] 在某些实施方案中,适用于本发明的VWF蛋白形成多聚体,例如二聚体、三聚体、四聚体、五聚体、六聚体、七聚体或更高级多聚体。在其它实施方案中,VWF蛋白是仅具有一个VWF蛋白的单体。在一些实施方案中,本发明的VWF蛋白可具有一个或多个氨基酸取代、缺失、添加或修饰。在一个实施方案中,VWF蛋白可包括氨基酸取代、缺失、添加或修饰以使VWF蛋白不能形成二硫键或不能形成二聚体或多聚体。在另一实施方案中,氨基酸取代在D'结构域和D3结构域内。在特定实施方案中,适用于本发明的VWF蛋白在对应于以下的残基处含有至少一个氨基酸取代:对应于SEQ ID NO:21的残基1099、残基1142、或残基1099与1142两者。至少一个氨基酸取代可为不天然存在于野生型VWF中的任何氨基酸。举例来说,氨基酸

取代可为除半胱氨酸以外的任何氨基酸,例如异亮氨酸、丙氨酸、亮氨酸、天冬酰胺、赖氨酸、天冬氨酸、甲硫氨酸、苯丙氨酸、谷氨酸、苏氨酸、谷氨酰胺、色氨酸、甘氨酸、缬氨酸、脯氨酸、丝氨酸、酪氨酸、精氨酸或组氨酸。在另一实例中,氨基酸取代具有一个或多个防止或抑制VWF蛋白形成多聚体的氨基酸。

[0237] 在某些实施方案中,适用于本文中的VWF蛋白可进一步被修饰以改进它与FVIII的相互作用,例如改进对FVIII的结合亲和力。作为非限制性实例,VWF蛋白包含在对应于SEQ ID NO:21的氨基酸764的残基处的丝氨酸残基和在对应于SEQ ID NO:21的氨基酸773的残基处的赖氨酸残基。残基764和/或773可有助于VWF蛋白对FVIII的结合亲和力。在其它实施方案中,适用于本发明的VWF蛋白可具有其它修饰,例如蛋白质可被聚乙二醇化、糖基化、羟乙基淀粉化或聚唾液酸化。

#### [0238] II. B. XTEN序列

[0239] 如本文所用,“XTEN序列”是指具有主要由小亲水性氨基酸组成的非天然存在的大致上非重复序列的延长长度多肽,其中所述序列在生理条件下具有低度或不具有二级或三级结构。作为嵌合蛋白配偶体,XTEN可充当载体,从而例如当连接于本发明的VWF蛋白或FVIII序列以产生嵌合蛋白时,赋予某些合乎需要的药物动力学、物理化学和药物性质。所述合乎需要的性质包括但不限于药物动力学参数和溶解性特征增强。如本文所用,“XTEN”明确排除抗体或抗体片段,诸如单链抗体或轻链或重链的Fc片段。

[0240] 本发明提供一种较短XTEN序列,当所述XTEN序列融合于VWF蛋白和/或第二Ig恒定区或其一部分时,相较于较长XTEN序列,所述较短XTEN序列提供改进的半衰期延长性质。因此,融合于VWF蛋白和/或第二Ig恒定区或其一部分的XTEN序列在长度方面含有小于288个氨基酸,即短于288个氨基酸。在一个实施方案中,融合于VWF蛋白和/或第二Ig恒定区或其一部分的XTEN序列由长度在12个氨基酸与287个氨基酸之间的氨基酸序列组成。在另一实施方案中,融合于VWF蛋白和/或第二Ig恒定区或其一部分的XTEN序列包含至少约36个氨基酸、至少约42个氨基酸、至少约72个氨基酸或至少约144个氨基酸,但小于288个氨基酸。在其它实施方案中,融合于VWF蛋白和/或第二Ig恒定区或其一部分的XTEN序列选自AE36、AG36、AE42、AG42、AE72、AG72、AE144或AG144。在一个实施方案中,融合于VWF蛋白和/或第二Ig恒定区或其一部分的XTEN序列是与SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11或SEQ ID NO:14至少80%、85%、90%、95%、96%、97%、98%、99%或100%同一的氨基酸序列,其中相较于无XTEN序列的嵌合蛋白,嵌合蛋白展现改进的半衰期。

[0241] 本发明的嵌合蛋白可进一步包含另外(第二、第三或更多)的XTEN序列。另外的XTEN序列可进一步融合于FVIII蛋白或第一Ig恒定区或其一部分。另外的XTEN序列可为任何长度。举例来说,融合于FVIII蛋白或第一Ig恒定区或其一部分的另外的XTEN序列是具有大于约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,另外的XTEN序列是具有大于约20至约3000个氨基酸残基、大于约30至约2500个残基、大于约40至约2000个残基、大于约50至约1500个残基、大于约60至约1000个残基、大于约70至约900个残基、大于约80至约800个残基、大于约90至约700个残基、大于约100至约600个残基、大于约110至约500个残基、或大于约120至约400个残基的肽或多肽。

[0242] XTEN序列(即融合于VWF蛋白和/或第二Ig恒定区或其一部分的XTEN序列或融合于FVIII蛋白和/或第一Ig恒定区或其一部分或插入在FVIII蛋白内一个或多个插入位点处的XTEN序列)可包含一个或多个具有9至14个氨基酸残基的序列基序或与所述序列基序至少80%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一的氨基酸序列,其中所述基序包含以下、基本上由以下组成或由以下组成:4至6种类型的选自由以下组成的组的氨基酸: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)。参见US 2010-0239554 A1。

[0243] 在一些实施方案中,XTEN序列包含非重叠序列基序,其中序列的至少约80%、或至少约85%、或至少约90%、或至少约91%、或至少约92%、或至少约93%、或至少约94%、或至少约95%、或至少约96%、或至少约97%、或至少约98%、或至少约99%或约100%由多个选自从表2A选择的单一基序家族的非重叠序列单元组成,从而产生家族序列。如本文所用,“家族”意指XTEN具有仅选自来自表2A的单一基序种类的基序;即AD、AE、AF、AG、AM、AQ、BC或BD XTEN,并且选择XTEN中不来自家族基序的任何其它氨基酸以实现所需性质,诸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或VWF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家族的多个非重叠序列基序单元,从而所得XTEN展现上述同源性范围。在其它实施方案中,XTEN包含来自表2A的两个或更多个基序家族的多个基序序列单元。可选择这些序列以实现以下更充分描述的由基序的氨基酸组成赋予的所需物理/化学特征,包括诸如净电荷、亲水性、缺乏二级结构或缺乏重复性的性质。在上文于这个段落中所述的实施方案中,可使用本文所述的方法选择并装配并入XTEN中的基序以实现XTEN具有约36至约3000个氨基酸残基。

[0244] 表2A. 具有12个氨基酸的XTEN序列基序和基序家族

[0245]

基序家族*	基序序列
AD	GESPGGSSGSES (SEQ ID NO: 24)
AD	GSEGSSGPGESS (SEQ ID NO: 25)
AD	GSSESGSSEGGP (SEQ ID NO: 26)
AD	GSGGEPSESGSS (SEQ ID NO: 27)
AE、AM	GSPAGSPTSTEE (SEQ ID NO: 28)
AE、AM、AQ	GSEPATSGSETP (SEQ ID NO: 29)
AE、AM、AQ	GTSESATPESGP (SEQ ID NO: 30)
AE、AM、AQ	GTSTEPSEGSAP (SEQ ID NO: 31)
AF、AM	GSTSESPSGTAP (SEQ ID NO: 32)

基序家族*	基序序列
AF, AM	GTSTPESGSASP (SEQ ID NO: 33)
AF, AM	GTSPSGESSTAP (SEQ ID NO: 34)
AF, AM	GSTSSTAESPGP (SEQ ID NO: 35)
AG, AM	GTPGSGTASSSP (SEQ ID NO: 36)
AG, AM	GSSTPSGATGSP (SEQ ID NO: 37)
AG, AM	GSSPSASTGTGP (SEQ ID NO: 38)
AG, AM	GASPGTSSTGSP (SEQ ID NO: 39)
AQ	GEPAGSPTSTSE (SEQ ID NO: 40)
AQ	GTGEPSSTPASE (SEQ ID NO: 41)
AQ	GSGPSTESAPTE (SEQ ID NO: 42)
AQ	GSETPSGPSETA (SEQ ID NO: 43)
AQ	GPSETSTSEPGA (SEQ ID NO: 44)
AQ	GSPSEPTEGTSA (SEQ ID NO: 45)
BC	GSGASEPTSTEP (SEQ ID NO: 46)
BC	GSEPATSGTEPS (SEQ ID NO: 47)
BC	GTSEPSTSEPGA (SEQ ID NO: 48)
BC	GTSTEPSEPGSA (SEQ ID NO: 49)
BD	GSTAGSETSTEA (SEQ ID NO: 50)
BD	GSETATSGSETA (SEQ ID NO: 51)
BD	GTSESATSESGA (SEQ ID NO: 52)
BD	GTSTEASEGSAS (SEQ ID NO: 53)

[0246]

[0247] • 表示当以各种排列一起使用时产生“家族序列”的个别基序序列

[0248] 在一些实施方案中,本发明中使用的XTEN序列与选自由以下组成的组的序列至少60%、70%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一: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和AG2004。参见US 2010-0239554 A1。

[0249] 在一个实施方案中,XTEN序列与选自由以下组成的组的氨基酸序列至少60%、70%、80%、90%、95%、96%、97%、98%、99%或100%同一:AE42 (SEQ ID NO:9)、AE72 (SEQ ID NO:10)、AE144\_2A (SEQ ID NO:55)、AE144\_3B (SEQ ID NO:56)、AE144\_4A (SEQ ID NO:57)、AE144\_5A (SEQ ID NO:58)、AE144\_6B (SEQ ID NO:59)、AG144\_A (SEQ ID NO:60)、AG144\_B (SEQ ID NO:61)、AG144\_C (SEQ ID NO:62)、AG144\_F (SEQ ID NO:63)、AE864 (SEQ ID NO:15)、AE576 (SEQ ID NO:16)、AE288 (SEQ ID NO:8)、AE288\_2 (SEQ ID NO:54)、AE144 (SEQ ID NO:11)、AG864 (SEQ ID NO:17)、AG576 (SEQ ID NO:18)、AG288 (SEQ ID NO:19)、AG144 (SEQ ID NO:14) 及其任何组合。在另一实施方案中,XTEN序列选自由以下组成的组:

AE42 (SEQ ID NO:9)、AE72 (SEQ ID NO:10)、AE144\_2A (SEQ IDNO:55)、AE144\_3B (SEQ ID NO:56)、AE144\_4A (SEQ ID NO:57)、AE144\_5A (SEQ ID NO:58)、AE144\_6B (SEQ ID NO:59)、AG144\_A (SEQ ID NO:60)、AG144\_B (SEQ ID NO:61)、AG144\_C (SEQ ID NO:62)、AG144\_F (SEQ ID NO:63)、AE864 (SEQ ID NO:15)、AE576 (SEQ ID NO:16)、AE288 (SEQ ID NO:8)、AE288\_2 (SEQ ID NO:54)、AE144 (SEQ ID NO:11)、AG864 (SEQ ID NO:17)、AG576 (SEQ ID NO:18)、AG288 (SEQ ID NO:19)、AG144 (SEQ ID NO:14) 及其任何组合。在特定实施方案中,XTEN序列是AE288。本发明的某些XTEN序列的氨基酸序列显示于表2B中。

[0250] 表2B.XTEN序列

XTEN	氨基酸序列
AE42 SEQ ID NO: 9	GAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPASS
AE72 SEQ ID NO: 10	GAP TSESATPESG PGSEPATSGS ETPGTSESAT PESGPGSEPA TSGSETPGTS ESATPESGPG TSTEPSEGS A PGASS
AE144 SEQ ID NO:11	GSEPATSGSETPGTSESATPESGPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEG SAPGSEPATSGSETPGSEPATSGSETPGSEPATSGSETPGTSTEPSEGSAPGTSESA PESGPGSEPATSGSETPGTSTEPSEGSAP
AE144_2A (SEQ ID NO: 55)	TSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTS TEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSES ATPESGPGTSESATPESGPG
AE144_3B (SEQ ID NO: 56)	SPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTS TEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAG SPTSTEEGTSTEPSEGSAPG
AE144_4A (SEQ ID NO: 57)	TSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTS TEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSES ATPESGPGTSTEPSEGSAPG
AE144_5A (SEQ ID NO: 58)	TSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTS TEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAG SPTSTEEGSPAGSPTSTEEG
AE144_6B (SEQ ID NO: 59)	TSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSE PATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSES ATPESGPGTSTEPSEGSAPG

[0251]



[0252]

AG144 SEQ ID NO:14	GTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSST GSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSA STGTGPGTTPGSGTASSSPGSSTPSGATGSP
AG144_A (SEQ ID NO: 60)	GASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTTPGSGTASSSPGSSTPSGATGSPGS SPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGASP GTSSTGSPGASPGTSSTGSP
AG144_B (SEQ ID NO: 61)	GTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGS SPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASP GTSSTGSPGASPGTSSTGSP
AG144_C (SEQ ID NO: 62)	GTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSS PSGATGSPGASPGTSSTGSP
AG144_F (SEQ ID NO: 63)	GSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGS SPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTTPGSGTASSSPGSSTPSGATGSPGSS PSGATGSPGASPGTSSTGSP
AE288 SEQ ID NO: 8	GTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESG PGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPES GPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSSESATPESGPGTSESATPE SGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSE GSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP
AE288_2 (SEQ ID NO: 54)	GSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGT STEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPAT SGSETPGTSESATPESGPGTSTEPSEGSAPGTSSESATPESGPGSPAGSPTSTEEGSPAGSPT STEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAP
AG288 SEQ ID NO:19	PGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASS SPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTG TGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSAST GTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSAS TGTGPGASPGTSSTGSPGSSPSASTGTGPGTTPGSGTASSSPGSSTPSGATGSP
AE576 SEQ ID NO:16	GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGS PGTSTEPSEGSAPGTSSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTST EEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSE SAPGTSTEPSEGSAPGTSSESATPESGPGTSTEPSEGSAPGTSSESATPESGPGSEPATSG SETPGTSTEPSEGSAPGTSTEPSEGSAPGTSSESATPESGPGTSESATPESGPGSEPATSG TSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEP SEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAG SPTSTEEGTSTEPSEGSAPGTSSESATPESGPGSEPATSGSETPGTSESATPESGPGSE ATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSSESATPESGPGSPAGSPTSTEEGSP AGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAP
AG576 SEQ ID NO:18	PGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATG SPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTAS SSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTTPGSGTA SSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSG ATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPS GATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPG

[0253]

	TSSTGSPGASPGTSSSTGSPGASPGTSSSTGSPGASPGTSSSTGSPGTPGSGTASSSPGSST PSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSS TPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSSTGSPGTPGSGTASSSPGS STPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSSTGS
AE864 SEQ ID NO:15	GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGS PGTSTEPSEGSAPGTSSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTST EEGTSESATPESGPGTSTEPSEGSAPGTSSTEPSEGSAPGSPAGSPTSTEEGTSTEPSE SAPGTSSTEPSEGSAPGTSSESATPESGPGTSTEPSEGSAPGTSSESATPESGPGSEPATSG SETPGTSTEPSEGSAPGTSSTEPSEGSAPGTSSESATPESGPGTSESATPESGPGSPAGSP TSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSSTEP SEGSAPGTSSTEPSEGSAPGTSSTEPSEGSAPGTSSTEPSEGSAPGTSSTEPSEGSAPGSPAG SPTSTEEGTSTEPSEGSAPGTSSESATPESGPGSEPATSGSETPGTSESATPESGPGSE ATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSSESATPESGPGSPAGSPTSTEEGSP AGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSSESATPESGPGS EPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPG SPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEE GSPAGSPTSTEEGTSTEPSEGSAPGTSSESATPESGPGTSESATPESGPGTSESATPESG PGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSSTEPSEGS APGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP
AG864 SEQ ID NO:17	GASPGTSSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTSGTASSSPGSSTPSGATGS PGSSPSASTGTGPGASPGTSSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASS SPGASPGTSSSTGSPGASPGTSSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSST GSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGA TGSPGSSTPSGATGSPGASPGTSSSTGSPGASPGTSSSTGSPGASPGTSSSTGSPGTPGSGT ASSSPGASPGTSSSTGSPGASPGTSSSTGSPGASPGTSSSTGSPGSSPSASTGTGPGTSGG TASSSPGASPGTSSSTGSPGASPGTSSSTGSPGASPGTSSSTGSPGSSTPSGATGSPGSSTP SGATGSPGASPGTSSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTP PSGATGSPGSSPSASTGTGPGASPGTSSSTGSPGASPGTSSSTGSPGTPGSGTASSSPGAS PGTSSSTGSPGASPGTSSSTGSPGASPGTSSSTGSPGASPGTSSSTGSPGTPGSGTASSSPGS STPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPG SSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSSTGSPGTPGSGTASSSP GSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSSTGSPGASPGTSSSTGS PGSSTPSGATGSPGSSPSASTGTGPGASPGTSSSTGSPGSSPSASTGTGPGTPGSGTASS SPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSSTGSP

[0254] 在其中XTEN组分具有小于100%的它的氨基酸由4、5或6种类型的选自甘氨酸(G)、丙氨酸(A)、丝氨酸(S)、苏氨酸(T)、谷氨酸(E)和脯氨酸(P)的氨基酸组成,或小于100%的序列由来自表3的序列基序或表4和13-17的XTEN序列组成的那些实施方案中,XTEN的其它氨基酸残基选自其它14种天然L-氨基酸中的任一个,但优先选自亲水性氨基酸,以使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。不是甘氨酸(G)、丙氨酸(A)、丝氨酸(S)、苏氨酸(T)、谷氨酸(E)和脯氨酸(P)的XTEN氨基酸分散在整个XTEN序列中,位于序列基序内部或之间,或集中在XTEN序列的一个或多个短链段中,例如以在XTEN组分与FVIII或VWF组分之间产生接头。在XTEN组分包含除甘氨酸(G)、丙氨酸(A)、丝氨酸(S)、苏氨酸(T)、谷氨酸(E)和脯氨酸(P)以外的氨基酸的所述情况下,优选的是小于约2%或小于约1%氨基酸是疏水性残基以使所得序列通常缺乏二级结构,例如不具有超过2%α螺旋或2%β折叠,如通过本文公开的方法所确定。在构建XTEN时有利性较小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,可设计XTEN序列以含有小于5%或小于4%或小于3%或小于2%或小于1%或不含有以下氨基酸:半胱氨酸(以避免形成二硫化物和氧化)、甲硫氨酸(以避免氧化)、天冬酰胺和谷氨酰胺(以避免脱酰胺)。因此,在一些实施方案中,除甘氨酸(G)、丙氨酸(A)、丝氨酸(S)、苏氨酸(T)、谷氨酸(E)和脯氨酸(P)之外也包含其它氨基酸的XTEN组分具有促成α螺旋和β折叠的残基小于5%(如通过Chou-Fasman算法所测量)的序列,并且具有至少90%或至少约95%或更多无规卷曲形成,如通过GOR算法所测量。

[0255] 在其它实施方案中,本发明中使用的XTEN序列影响本发明的嵌合蛋白的物理或化

学性质,例如药物动力学。本发明中使用的XTEN序列可展现一种或多种以下有利性质: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性或增加的流体动力学(或斯托克斯(Stokes))半径。在特定实施方案中,连接于本发明中的FVIII蛋白的XTEN序列增加药物动力学性质,诸如较长的终末半衰期或增加的曲线下面积(AUC),以使本文所述的嵌合蛋白相较于野生型FVIII在体内停留时期增加。在其它实施方案中,本发明中使用的XTEN序列增加药物动力学性质,诸如较长的终末半衰期或增加的曲线下面积(AUC),以使FVIII蛋白相较于野生型FVIII在体内停留时期增加。

[0256] 本发明的一个实施方案是一种FVIII/VWF融合蛋白,其包含融合于Fc区的FVIII部分和融合于Fc区的VWF部分,其中XTEN序列(例如AE288)插入在所述FVIII部分内,并且其中具有小于288个氨基酸的XTEN序列(例如AE144)插入在所述VWF部分与Fc部分之间。如实施例中所述,在VWF部分与Fc部分之间插入具有小于288个氨基酸的XTEN对嵌合蛋白的药物动力学的影响大于在VWF部分与Fc部分之间插入具有288个氨基酸的XTEN。举例来说,相较于具有288个氨基酸的XTEN,在FVIII/VWF融合蛋白中的VWF部分与Fc部分之间插入具有小于288个氨基酸的XTEN序列可增加嵌合蛋白的终末半衰期。在一些实施方案中,相对于插入具有288个氨基酸的XTEN序列,终末半衰期增加至少约5%、至少约10%、至少约15%、至少约20%、至少约25%或至少约30%。在一个特定实施方案中,相对于插入具有288个氨基酸的XTEN,终末半衰期增加至少约35%。插入具有小于288个氨基酸的XTEN序列也可增加嵌合蛋白的AUC值。在一些实施方案中,相对于插入具有288个氨基酸的XTEN,AUC增加至少约50%、至少约100%或至少约200%。在一个特定实施方案中,AUC增加约两倍。插入具有小于288个氨基酸的XTEN序列也可降低嵌合蛋白的清除率。举例来说,相对于插入具有288个氨基酸的XTEN序列,清除率可降低至少约5%、至少约10%、至少约15%、至少约20%、至少约25%或至少约30%。相对于插入具有288个氨基酸的XTEN,插入具有小于288个氨基酸的XTEN序列可增加平均滞留时间(MRT)和/或降低稳态表观分布体积( $V_{ss}$ )。

[0257] 多种方法和测定可用于测定包含XTEN序列的蛋白质的物理/化学性质。所述方法包括但不限于分析性离心、EPR、HPLC-离子交换、HPLC-尺寸排阻、反相HPLC、光散射、毛细管电泳、圆二色性、差示扫描量热法、荧光、HPLC-离子交换、HPLC-尺寸排阻、IR、NMR、拉曼(Raman)光谱法、折射法和UV/可见光谱法。其它方法公开于Amau等,Prot Expr and Purif 48,1-13(2006)中。

[0258] 可根据本发明使用的XTEN序列的其它实例公开于美国专利公布号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、WO 2011028344 A2或WO 20130122617 A1中。

[0259] II.C. 因子VIII(FVIII)蛋白

[0260] 除非另外规定,否则如本文所用的“FVIII蛋白”意指在凝血中发挥它的正常作用的功能性FVIII多肽。术语FVIII蛋白包括其保留全长野生型因子VIII在凝血路径中的功能的功能性片段、变体、类似物或衍生物。“FVIII蛋白”可与FVIII多肽(或蛋白质)或FVIII互换使用。FVIII功能的实例包括但不限于能够活化凝血、能够充当因子IX的辅因子、或能够在 $Ca^{2+}$ 和磷脂存在下与因子IX形成因子X活化酶复合物,其接着使因子X转化成活化形式Xa。FVIII蛋白可为人、猪、犬、大鼠或鼠类FVIII蛋白。此外,来自人与其它物种的FVIII之间的

比较已鉴定可能为功能所需的保守残基(Cameron等,Thromb.Haemost.79:317-22(1998); US 6,251,632)。

[0261] 许多测试可用于评估凝血系统的功能:活化部分凝血激素时间(aPTT)测试、显色测定、ROTEM测定、凝血酶原时间(PT)测试(也用于确定INR)、纤维蛋白原测试(常通过克劳斯(Clauss)方法)、血小板计数、血小板功能测试(常通过PFA-100)、TCT、流血时间、混合测试(如果患者的血浆与正常血浆混合,那么是否进行异常矫正)、凝血因子测定、抗磷脂抗体、D-二聚体、遗传测试(例如因子V Leiden、凝血酶原突变G20210A)、稀释鲁塞尔氏蝰蛇毒液时间(Russell's viper venom time,dRVVT)、杂项血小板功能测试、凝血弹性描记法(TEG或Sonoclot)、凝血弹性测量术(TEM<sup>®</sup>,例如ROTEM<sup>®</sup>)或优球蛋白(euglobulin)溶解时间(ELT)。

[0262] aPTT测试是衡量“内在”凝血路径(也称为接触活化路径)与共同凝血路径两者的功效的性能指标。这个测试通常用于测量可商购获得的重组凝血因子(例如FVIII或FIX)的凝血活性。它与衡量外在路径的凝血酶原时间(PT)联合使用。

[0263] ROTEM分析提供关于以下完整止血动力学的信息:凝结时间、凝块形成、凝块稳定性和溶解。凝血弹性测量术中的不同参数取决于血浆凝血系统的活性、血小板功能、纤维蛋白溶解或影响这些相互作用的许多因素。这个测定可提供对次级止血的完全综览。

[0264] 已知FVIII多肽和多核苷酸序列,也已知许多功能性片段、突变体和修饰形式。以下显示人FVIII序列(全长)的实例。

[0265] 表3. 全长因子VIII的氨基酸序列

[0266] (全长FVIII(FVIII信号肽加下划线;FVIII重链加双下划线;B结构域用斜体表示;并且FVIII轻链呈普通文本形式)

信号肽:(SEQ ID NO: 64)  
MQIELSTCFFLCLLRFCFS

成熟因子 VIII (SEQ ID NO: 65)\*

ATTRYYLGAVELSDYMQSDLGELPVDARFPPRVPKSFPPNTSVVYKKTLEFVEFDHLEFNIAKPRPPWMLL  
GPTIQAEVYDVTVITLKNMASHPVSLHAVGVSYWKASEGAEYDDQTSOREKEDDKVFPGGSHTYVWQVLKEN  
GPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVFDGKSWHSETKNSL  
MQRDAASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPEVHSIFLEGHTFLVRNHRQASLEI

SPITFLTAQTLMLDLGQFLLEFCHISSHQHDGMEAYVKVSDSCPEEPQLRMKNNEEAEDYDDDLTDSEMDVVR  
DDDNPSFSFIQIRSVAKKHPKTWVHYIAAEEEDWDYAPLVLAPDDRYSYKSYQLNNGEPQIRGRKYKQVRFMAYT  
DETFFKTREAIQHESGILGPELLYGEVGDITLLIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPIL  
PGEIFKYKWTVTVEDEGPTKSDPRCLTRYSSVFNMERDLASGLIGPLLYCYKESVDQRGNQIMSDKRNVILE  
SVFDENRSWYLTENIQRFLEPNPAGVQLEDPEFOASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGAOTDF  
LSVFFSGYTFKHKMVEEDTLTLFPFSGETVFMSEMPGLWILGCHNSDFRNRGMTALLKVSSCDKNTGDYDE  
DSYEDISAYLLSKNNAIEPRSFSSQNSRHPSTRQKQFNATFI PENDIEKTDPPFAHRTMPKIQNVSSSDLLM  
LLRQSPTPHGLSLSLQEAKEYETFSDDPSPGAIDSNNSLSEMTFRPQLHHSQDMVFTPESSLQRLRNEKLG  
TTAATELKKLDFKVSSTSNLI STIPSDNLAAGTONTSSLGPPSMFVHYDSQLDITLFGKKS SPLTESGGFL  
SLSEENND SKLLESGLMNSQESSWGKNVSSTESGRLFRKRAHGPAALLTKDNALFKVSI SLLKTNKTSNNSA  
TNRKTHIDGPSLLIENS PSVWQNI LESDTEFKKVTPLIHDRMLMDKNATALRLNHMSNKTTSSKNMEMVQOK  
KEGPIPPDAQNPDMSEFFKMLFLPE SARWIQRTHGKNSLNSGQGPSKQLVSLGPEKSEVQGNFLSEKNKVVV  
GKGEFTKDVGLKEMVFPSSRNLF LTNLDNLHENNTHNQEKKI QEEIEKKETLIQENVVLPQIHTVTGTKNFM  
KNLFLSTRQNVESYDGAYAPVLQDFRSLNDSNRTKHTAHFSKKGEEENLEGLNQTKQIVEKYACTTR  
ISPNTSQQNFVTQRSKRALKQFRLEETELEKRI IVDDTSTQWSKNMKHLTPSTLTQIDYNEKEKGAITQS  
PLSDCLTRSHSIPOANRSLPLIAKVSSFP SIRPIYLTRVLPQDNSSHLPAASYRKKDSGVQESSHFLQGAKK  
NNLSLAILTLEMTGDQREVGSLGTSATNSVTYKVENTVLPKPDLPKTSKVELLPKVHIYQKDLFPPTETSN  
GSPGHLDLVEGSLQGTGAIKWNEANRPGKVPFLRVATESSAKTPSKLLDPLANDNHYGTQIPKEEWSQE  
KSPKTAFAKKKDTI LSLNACESNHAI AAINEGQNKPSIEVFWAKQGRTERLCSQNPVVKRHHQREITRTTLQ  
SDQEEIDYDDTI SVEMKKEDFDIYDEDENQSPRSFQKKTRHYFIAAVERLWDYGMSSSPHVLNRNRAQSGSVP  
QFKKVVFOEFTDGSFTQPLYRGELNEHLGLLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLSIYEEDQRQGA  
EPRKNFVKPNETKYFWKVQHMMAPT KDEFCKAWAYFSDVDLEKDVHSGLIGPLLVCHTNTLNPAHGRQVT  
VQEFALFFTI FDETKSWYFTENMERNCRAPCNIQMEDPTFKENYRFHAINGYIMDTLPGLVMAQDQIRIWYL  
LSMGSNENIHSIHFSGHVFTVRKKEEYKMALYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLV  
YSNKCQTP LGMASGHIRD FQITASGQYQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMI IHGIKTQG  
ARQKFSSLYISQFI IMYSLDGKKWQTYRGNSTGLMVFFGNVDSSGIKHNIFNPPIIARYIRLHPHYSIRS  
TLRMELMGCDLNSCSMPLGMESKAI SDAQITASSYFTNM FATWSPSKARLHLQGRSNAWRPQVNNPKEWLQV  
DFQKTMKVTGVTQGVKSLTSMYVKEFLISSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLLTR  
YLRIHPQSWVHQIALRMEVLGCEAQDLY

[0267]

[0268] 表4. 编码全长FVIII的核苷酸序列 (SEQ ID NO:66)\*

[0269]

```

661                                     ATG CAAATAGAGC TCTCCACCTG
721 CTTCTTTCTG TGCCTTTTGC GATTCTGCTT TAGTGCCACC AGAAGATACT ACCTGGGTGC
781 AGTGGAACCTG TCATGGGACT ATATGCAAAG TGATCTCGGT GAGCTGCCTG TGGACGCAAG
841 ATTCCTCCTCCT AGAGTGCCAA AATCTTTTCC ATTCAACACC TCAGTCGTGT ACAAAAAGAC
901 TCTGTTTGTGTA GAATTCACGG ATCACCTTTT CAACATCGCT AAGCCAAGGC CACCCTGGAT
961 GGGTCTGCTA GGTCCCTACCA TCCAGGCTGA GGTTTATGAT ACAGTGGTCA TTACTACTTAA
1021 GAACATGGCT TCCCATCCTG TCAGTCTTCA TGCTGTTGGT GTATCCTACT GGAAAGCTTC
1081 TGAGGGAGCT GAATATGATG ATCAGACCAG TCAAAGGGAG AAAGAAGATG ATAAAGTCTT
1141 CCCTGGTGGA AGCCATACAT ATGTCTGGCA GGTCTGAAA GAGAATGGTC CAATGGCCTC
1201 TGACCCACTG TGCCTTACCT ACTCATATCT TTCTCATGTG GACCTGGTAA AAGACTTGAA
1261 TTCAGGCCTC ATTGGAGCCC TACTAGTATG TAGAGAAGGG AGTCTGGCCA AGGAAAAGAC
1321 ACAGACCTTG CACAAATTTA TACTACTTTT TGCTGTATTT GATGAAGGGA AAAGTTGGCA
1381 CTCAGAAACA AAGAACTCCT TGATGCAGGA TAGGGATGCT GCATCTGCTC GGGCCTGGCC
1441 TAAAATGCAC ACAGTCAATG GTTATGTAAA CAGGTCTCTG CCAGGTCTGA TTGGATGCCA
1501 CAGGAAATCA GTCTATTGGC ATGTGATTGG AATGGGCACC ACTCCTGAAG TGCACCTCAAT
1561 ATTCTCGAA GGTCACACAT TTCTTGAGAG GAACCATCGC CAGGCGTCTT TGGAAATCTC
1621 GCCAATAACT TTCCTTACTG CTCAAACACT CTTGATGGAC CTTGGACAGT TTCTACTGTT
1681 TTGTCATATC TCTTCCCACC AACATGATGG CATGGAAGCT TATGTCAAAG TAGACAGCTG
1741 TCCAGAGGAA CCCCAACTAC GAATGAAAAA TAATGAAGAA GCGGAAGACT ATGATGATGA
1801 TCTTACTGAT TCTGAAATGG ATGTGGTCAG GTTTGATGAT GACAACTCTC CTTCCCTTAT
1861 CCAAATTCGC TCAGTTGCCA AGAAGCATCC TAAAACCTGG GTACATTACA TTGCTGCTGA

```

[0270]

1921 AGAGGAGGAC TGGGACTATG CTCCTTAGT CCTCGCCCC GATGACAGAA GTTATAAAAG  
1981 TCAATATTTG AACAAATGGCC CTCAGCGGAT TGGTAGGAAG TACAAAAAAG TCCGATTTAT  
2041 GGCATACACA GATGAAACCT TTAAGACTCG TGAAGCTATT CAGCATGAAT CAGGAATCTT  
2101 GGGACCTTTA CTTTATGGGG AAGTTGGAGA CACACTGTTG ATTATATTTA AGAATCAAGC  
2161 AAGCAGACCA TATAACATCT ACCCTCACGG AATCACTGAT GTCCGTCCTT TGTATTCAAG  
2221 GAGATTACCA AAAGGTGTAA AACATTTGAA GGATTTTCCA ATTCTGCCAG GAGAAATATT  
2281 CAAATATAAA TGGACAGTGA CTGTAGAAGA TGGGCCAACT AAATCAGATC CTCGGTGCCT  
2341 GACCCGCTAT TACTCTAGTT TCGTTAATAT GGAGAGAGAT CTAGCTTCAG GACTCATTGG  
2401 CCTCTCCTC ATCTGCTACA AAGAATCTGT AGATCAAAGA GGAAACCAGA TAATGTCAGA  
2461 CAAGAGGAAT GTCATCCTGT TTTCTGTATT TGATGAGAAC CGAAGCTGGT ACCTCACAGA  
2521 GAATATACAA CGCTTTCTCC CCAATCCAGC TGGAGTGCAG CTTGAGGATC CAGAGTTCCA  
2581 AGCTCCAAC ATCATGCACA GCATCAATGG CTATGTTTTT GATAGTTTGC AGTTGTCAGT  
2641 TTGTTTGCAT GAGGTGGCAT ACTGGTACAT TCTAAGCATT GGAGCACAGA CTGACTTCCT  
2701 TTCTGTCTTC TTCTCTGGAT ATACCTTCAA ACACAAAATG GTCTATGAAG ACACACTCAC  
2761 CCTATTCCTA TTCTCAGGAG AAACTGTCTT CATGTCGATG GAAAACCCAG GTCTATGGAT  
2821 TCTGGGGTGC CACAACCTAG ACTTTGCGAA CAGAGGCATG ACCGCCTTAC TGAAGGTTTC  
2881 TAGTTGTGAC AAGAACACTG GTGATTATTA CGAGGACAGT TATGAAGATA TTTCAGCATA  
2941 CTTGCTGAGT AAAACAATG CCATTGAACC AAGAAGCTTC TCCCAGAATT CAAGACACCC  
3001 TAGCACTAGG CAAAAGCAAT TTAATGCCAC CACAATTCCA GAAAATGACA TAGAGAAGAC  
3061 TGACCCTTGG TTTGCACACA GAACACCTAT GCCTAAAATA CAAAATGTCT CCTCTAGTGA  
3121 TTGTTGATG CTCTTGCGAC AGAGTCTTAC TCCACATGGG CTATCCTTAT CTGATCTCCA  
3181 AGAAGCCAAA TATGAGACTT TTTCTGATGA TCCATCACCT GGAGCAATAG ACAGTAATAA  
3241 CAGCCTGTCT GAAATGACAC ACTTCAGGCC ACAGCTCCAT CACAGTGGGG ACATGGTATT  
3301 TACCCCTGAG TCAGGCCCTCC AATTAAGATT AAATGAGAAA CTGGGGACAA CTGCAGCAAC  
3361 AGAGTTGAAG AAACCTGAT TCAAAGTTTC TCAAAGTACA TAGTACATCA AATAATCTGA TTTCACAAT  
3421 TCCATCAGAC AATTTGGCAG CAGGTACTGA TAATACAAGT TCCTTAGGAC CCCCAGTAT  
3481 GCCAGTTCAT TATGATAGTC AATTAGATAC CACTCTATTT GGCAAAAAGT CATCTCCCCT  
3541 TACTGAGTCT GGTGGACCTC TGAGCTTGAG TGAAGAAAAT AATGATTCAA AGTTGTTAGA  
3601 ATCAGGTTTA ATGAATAGCC AAGAAAGTTC ATGGGGAAAA AATGTATCGT CAACAGAGAG  
3661 TGGTAGGTTA TTTAAAGGGA AAAGAGCTCA TGGACCTGCT TTGTTGACTA AAGATAATGC  
3721 CTTATTCAAA GTTAGCATCT CTTTGTAAA GACAAAACAA ACTTCCAATA ATTCAGCAAC  
3781 TAATAGAAAAG ACTCACATTG ATGGCCCATC ATTATTAATT GAGAAATGCA CATCAGTCTG  
3841 GCAAAAATA TTAGAAAAGT ACCTAGATT TAAAAAGTG ACACCTTTGA TTCATGACAG  
3901 AATGCTTATG GACAAAAATG CTACAGCTTT GAGGCTAAAT CATATGTCAA ATAAAACTAC  
3961 TTCATCAAAA AACATGGAAA TGGTCCAACA GAAAAAAGAG GGCCCCATTC CACCAGATGC  
4021 ACAAATCCA GATATGTCGT TCTTTAAGAT GCTATCTTGG CCAGAATCAG CAAGGTGGAT  
4081 ACAAAGGACT CATGGAAAGA ACTCTCTGAA CTC TGGGCAA GGCCCCAGTC CAAAGCAATT  
4141 AGTATCCTTA GGACCAGAAA AATCTGTGGA AGGTCAGAA TTTCTGTCTG AGAAAAACAA  
4201 AGTGGTAGTA GGAAAGGGTG AATTTACAAA GGACGTAGGA CTCAAAGAGA TGGTTTTTCC  
4261 AAGCAGCAGA AACCTATTTT TTAATAACTT GGATAATTTA CATGAAAATA ATACACACAA  
4321 TCAAGAAAAA AAAATTCAGG AAGAAATAGA AAAGAAGGAA ACATTAATCC AAGAGAATGT  
4381 AGTTTTGCCT CAGATACATA CAGTGACTGG CACTAAGAAT TTCATGAAGA ACCTTTTCTT  
4441 ACTGAGCACT AGGCAAAAATG TAGAAGGTTT ATATGACGGG GCATATGCTC CAGTACTTCA  
4501 AGATTTTAGG TCATTAATG ATTCAACAAA TAGAACAAAG AAACACACAG CTCATTTCTC  
4561 AAAAAAGGG GAGGAAGAAA ACTTGGGAAG CTTGGGAAAT CAAACCAAGC AAATTGTAGA  
4621 GAAATATGCA TGCACCACAA GGATATCTCC TAATACAAGC CAGCAGAATT TTGTCACGCA  
4681 ACGTAGTAAG AGAGCTTTGA AACAATTCAG ACTCCCCTA GAAGAAACAG AACTTGAAAA  
4741 AAGGATAATT GTGGATGACA CCTCAACCCA GTGGTCCAAA AACATGAAAC ATTTGACCCC  
4801 GAGCACCTC ACACAGATAG ACTACAATGA GAAGGAGAAA GGGGCCATTA CTCGACTCTC  
4861 CTTATCAGAT TGCCTTACGA GGAGTCATAG CATCCCTCAA CCAAATAGAT CTCCATTACC  
4921 CATTGCAAAG GTATCATCAT TTCCATCTAT TAGACCTATA TATCTGACCA GGGTCTTATT  
4981 CCAAGACAAC TCTTCTCATC TTCCAGCAGC ATCTTATAGA AAGAAAGATT CTGGGGTCCA  
5041 AGAAAGCAGT CATTCTTTAC AAGGAGCCAA AAAAAATAAC CTTTCTTTAG CCATTCTAAC  
5101 CTTGGAGATG ACTGGTGATC AAAGAGAGGT TGGCTCCCTG GGGACAAGTG CCACAAATTC  
5161 AGTCACATAC AAGAAAGTTG AGAACACTGT TCTCCCGAAA CCAGACTTGC CCAAAACATC  
5221 TGGCAAAGTT GAATTGCTTC CAAAAGTTCA CATTATTCAG AAGGACCTAT TCCCTACGGA  
5281 AACTAGCAAT GGGTCTCCTG GCCATCTGGA TCTCGTGGA GGGAGCTTTC TTCAGGGAAC  
5341 AAGGGAGCG ATTAAGTGA ATGAAGCAA CAGACCTGGA AAAGTCCCT TTCTGAGAGT  
5401 AGCAACAGAA AGCTCTGCAA AGACTCCCTC CAAGCTATTG GATCCTCTTG CTTGGGATAA

5461	CCACTATGGT	ACTCAGATAC	CAAAAAGAAGA	GTGGAAATCC	CAAGAGAAGT	CACCAGAAAA	
5521	AACAGCTTTT	AAGAAAAAGG	ATACCATTTT	GTCCTGAAAC	GCTTGTGAAA	GCAATCATGC	
5581	AATAGCAGCA	ATAAATGAGG	GACAAAATAA	GCCCGAAATA	GAAGTCACCT	GGGCAAAGCA	
5641	AGGTAGGACT	GAAAGGCTGT	GCTCTCAAAA	CCCACCAGTC	TTGAAACGCC	ATCAACGGGA	
5701	AATAACTCGT	ACTACTCTTC	AGTCAGATCA	AGAGGAAATT	GACTATGATG	ATACCATATC	
5761	AGTTGAAATG	AAGAAGGAAG	ATTTTGACAT	TTATGATGAG	GATGAAAATC	AGAGCCCCCG	
5821	CAGCTTTCAA	AAGAAAACAC	GACACTATTT	TATTGCTGCA	GTGGAGAGGC	TCTGGGATTA	
5881	TGGGATGAGT	AGCTCCCCAC	ATGTTCTAAG	AAACAGGGCT	CAGAGTGGCA	GTGTCCCTCA	
5941	GTTCAAGAAA	GTTGTTTTCC	AGGAATTTAC	TGATGGCTCC	TTTACTCAGC	CCTTATACCG	
6001	TGGAGAACTA	AATGAACATT	TGGGACTCCT	GGGGCCATAT	ATAAGAGCAG	AAGTTGAAGA	
6061	TAATATCATG	GTAACTTTCA	GAAATCAGGC	CTCTCGTCCC	TATTCCTTCT	ATTCTAGCCT	
6121	TATTTCTTAT	GAGGAAGATC	AGAGGCAAGG	AGCAGAACCT	AGAAAAAACT	TTGTCAAGCC	
6181	TAATGAAACC	AAAACTTACT	TTTGGAAAGT	GCAACATCAT	ATGGCACCCA	CTAAAGATGA	
6241	GTTTGACTGC	AAAGCCTGGG	CTTATTTCTC	TGATGTTGAC	CTGGAAAAAG	ATGTGCACTC	
6301	AGGCCTGATT	GGACCCCTTC	TGGTCTGCCA	CACTAACACA	CTGAACCCCTG	CTCATGGGAG	
6361	ACAAGTGACA	GTACAGGAAT	TTGTCTGTGT	TTTCACCATC	TTTGATGAGA	CCAAAAGCTG	
6421	GTACTTCACT	GAAAAATATG	AAAGAAACTG	CAGGGCTCCC	TGCAATATCC	AGATGGAAGA	
6481	TCCCACCTTT	AAAGAGAATT	ATCGCTTCCA	TGCCAATCAAT	GGCTACATAA	TGGATACACT	
6541	ACCTGGCTTA	GTAATGGCTC	AGGATCAAAG	GATTTCGATGG	TATCTGCTCA	GCATGGGCAG	
[0271]	6601	CAATGAAAAC	ATCCATTCTA	TTCATTTTCAG	TGGACATGTG	TTCACTGTAC	GAAAAAAAGA
6661	GGAGTATAAA	ATGGCACGTG	ACAATCTCTA	TCCAGGTGTT	TTTGAGACAG	TGGAAATGTT	
6721	ACCATCCAAA	GCTGGAATTT	GGCGGGTGGG	ATGCCCTTAT	GGCGAGCATC	TACATGCTGG	
6781	GATGAGCACA	CTTTTTCTGG	TGTACAGCAA	TAAGTGTGAG	ACTCCCTG	GAATGGCTTC	
6841	TGGACACATT	AGAGATTTTC	AGATTACAGC	TTCAGGACAA	TATGGCAGT	GGGCCCAAA	
6901	GCTGGCCAGA	CTTCATTAT	CCGGATCAAT	CAATGCCTGG	AGCACCAAGG	AGCCCTTTTC	
6961	TTGGATCAAG	GTGGATCTGT	TGGCACCAAT	GATTATTCAC	GGCATCAAGA	CCCAGGGTGC	
7021	CCGTCAGAAG	TTCTCCAGCC	TCTACATCTC	TCAGTTTATC	ATCATGTATA	GTCTTGATGG	
7081	GAAGAAGTGG	CAGACTTATC	GAGGAAATTC	CACTGGAACC	TTAATGGTCT	TCTTTGGCAA	
7141	TGTGGATTCA	TCTGGGATAA	AACACAATAT	TTTAAACCTT	CCAATTATTG	CTCGATACAT	
7201	CCGTTTGCAC	CCAACCTCATT	ATAGCATTTCG	CAGCACTCTT	CGCATGGAGT	TGATGGGCTG	
7261	TGATTTAAAT	AGTTGCAGCA	TGCCATTGGG	AATGGAGAGT	AAAGCAATAT	CAGATGCACA	
7321	GATTACTGCT	TCATCCTACT	TTACCAATAT	GTTTGCCACC	TGTTCTCCTT	CAAAAAGCTG	
7381	ACTTCACCTC	CAAGGGAGGA	GTAATGCCTG	GAGACCTCAG	GTGAATAATC	CAAAAAGAGTG	
7441	GCTGCAAGTG	GACTTCCAGA	AGACAATGAA	AGTCACAGGA	GTAACTACTC	AGGGAGTAAA	
7501	ATCTCTGCTT	ACCAGCATGT	ATGTGAAGGA	GTTCTCATC	TCCAGCAGTC	AAGATGGCCA	
7561	TCAGTGGACT	CTCTTTTTTC	AGAATGGCAA	AGTAAAGGTT	TTTCAGGGAA	ATCAAGACTC	
7621	CTTCACACCT	GTGGTGAAC	CTCTAGACCC	ACCGTTACTG	ACTCGCTACC	TTCGAATTCA	
7681	CCCCCAGAGT	TGGGTGCACC	AGATTGCCCT	GAGGATGGAG	GTTCTGGGCT	GCGAGGCACA	
7741	GGACCTCTAC						

[0272] \*加下划线的核酸编码信号肽。

[0273] FVIII多肽包括全长FVIII、全长FVIII减去N末端Met、成熟FVIII(减去信号序列)、在N末端具有另外的Met的成熟FVIII和/或B结构域完全或部分缺失的FVIII。在某些实施方案中,FVIII变体包括B结构域缺失,无论部分或完全缺失。

[0274] 天然成熟人FVIII的序列呈现为SEQ ID NO:65。天然FVIII蛋白具有下式:A1-a1-A2-a2-B-a3-A3-C1-C2,其中A1、A2和A3是结构相关的“A结构域”,B是“B结构域”,C1和C2是结构相关的“C结构域”,并且a1、a2和a3是酸性间隔体区。关于SEQ ID NO:65中的一级氨基酸序列位置,人FVIII的A1结构域从Ala1延伸至约Arg336,a1间隔体区从约Met337延伸至约Val374,A2结构域从约Ala375延伸至约Tyr719,a2间隔体区从约Glu720延伸至约Arg740,B结构域从约Ser741延伸至约Arg1648,a3间隔体区从约Glu1649延伸至约Arg1689,A3结构域从约Ser1690延伸至约Leu2025,C1结构域从约Gly2026延伸至约Asn2072,并且C2结构域从约Ser2073延伸至Tyr2332。除特定蛋白水解裂解位点以外,对FVIII的结构域和区域之间的边界的位置的指定在不同参考文献中可有所不同。因此,本文中指示的边界通过使用术语“约”来指定为近似边界。

[0275] 分离人FVIII基因,并且在哺乳动物细胞中表达(Toole,J.J.等,Nature 312:342-



347(1984);Gitschier,J.等,Nature 312:326-330(1984);Wood,W.I.等,Nature 312:330-337(1984);Vehar,G.A.等,Nature 312:337-342(1984);WO 87/04187;WO 88/08035;WO 88/03558;以及美国专利号4,757,006)。从如美国专利号4,965,199中所示的cDNA推断FVIII氨基酸序列。此外,B结构域部分或完全缺失的FVIII显示于美国专利号4,994,371和4,868,112中。在一些实施方案中,人FVIII B结构域被如美国专利号5,004,803中所示的人因子V B结构域置换。编码人因子VIII的cDNA序列和氨基酸序列分别以美国申请公布号2005/0100990的SEQ ID NO:1和2显示。

[0276] 猪FVIII序列发表在Toole,J.J.等,Proc.Natl.Acad.Sci.USA 83:5939-5942(1986)中。此外,由从猪脾cDNA文库PCR扩增FVIII序列获得的完全猪cDNA序列已报道于Healey,J.F.等,Blood 88:4209-4214(1996)中。具有所有结构域、所有亚单位和特定氨基酸序列的取代的杂交人/猪FVIII公开于Lollar和Runge的美国专利号5,364,771中和WO 93/20093中。更新近地,猪FVIII的A1和A2结构域和用猪A1和/或A2结构域取代相应人结构域的嵌合FVIII的核苷酸和相应氨基酸序列报道于WO 94/11503中。Lollar,J.S.的美国专利号5,859,204也公开猪cDNA和推断的氨基酸序列。美国专利号6,458,563公开一种B结构域缺失的猪FVIII。

[0277] Lollar,J.S.的美国专利号5,859,204报道具有降低的抗原性和降低的免疫反应性的FVIII功能性突变体。Lollar,J.S.的美国专利号6,376,463也报道具有降低的免疫反应性的FVIII突变体。Saenko等的美国申请公布号2005/0100990报道FVIII的A2结构域中的功能性突变。

[0278] 在一个实施方案中,FVIII(或嵌合蛋白的FVIII部分)可与具有SEQ ID NO:67的氨基酸1至1438或SEQ ID NO:65的氨基酸1至2332的FVIII氨基酸序列(无信号序列)或具有SEQ ID NO:64的氨基酸1至19和SEQ ID NO:67的氨基酸1至1438或SEQ ID NO:64的氨基酸1至19和SEQ ID NO:65的氨基酸1至2332的FVIII氨基酸序列(具有信号序列)至少50%、60%、70%、80%、90%、95%、96%、97%、98%、99%或100%同一,其中所述FVIII具有凝结活性,例如作为辅因子活化因子IX以使因子X转化成活化因子X。FVIII(或嵌合蛋白的FVIII部分)可与具有SEQ ID NO:67的氨基酸1至1438或SEQ ID NO:65的氨基酸1至2332的FVIII氨基酸序列(无信号序列)同一。FVIII可进一步包含信号序列。

[0279] 如本文所用的FVIII的“B结构域”与本领域中已知的通过内部氨基酸序列同一性和蛋白水解裂解位点定义的B结构域相同,例如是全长人FVIII的残基Ser741-Arg1648。其它人FVIII结构域由以下氨基酸残基界定:A1,残基Ala1-Arg372;A2,残基Ser373-Arg740;A3,残基Ser1690-Asn2019;C1,残基Lys2020-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通常被称为a3酸性区。猪、小鼠和犬FVIII的包括B结构域的所有结构域的边界位置在本领域中也是已知的。在一个实施方案中,缺失FVIII的B结构域(“B结构域缺失的因子VIII”或“BDD FVIII”)。BDD FVIII的一实例是REFACTO<sup>®</sup>(重组BDD FVIII),其与具有表5中的序列的因子VIII部分具有相同序列。(BDD FVIII重链加双下划线;B结构域用斜体表示;并且BDD FVIII轻链呈普通文本形式)。编码表6的核苷酸序列(SEQ ID NO:68)显示于表6中。

[0280] 表5.B结构域缺失的因子VIII(BDD FVIII)的氨基酸序列

BDD FVIII (SEQ ID NO: 67)

ATRRYYLGAVELSDYMQSDLGELPVDARFPPRPVKSPFENTSVVYKKTLEFVEFTDHLFNIAKPRPPWMGLL  
GPTIQAEVYDTVVITLKNMASHPVSLHAVGVSYWKASEGAEYDDQTSQREKEDDKVFPGGSHTYVVOVLKEN  
GPMASDPLCLTYSYLSHVLDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVFDGKSWHSETKNSL  
MODRDAASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPEVHSIFLEGHTFLVRNHRQASLEI  
SPITFLTAQTLMLMDLQGLLFCHESSHQHDGMEAYVKVDSCPEEPQLRMKNNEEAEDYDDDLTDSEMDVVR  
DDNSPSFIQIRSVAKKHKPTWVHYIAEEEDWDYAPLVLAPDDRSYKSOYLNGPQRIGRKYKQVRFMAYT  
DETFKTREAIQHESGILGPLYGEVGDILLIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPIL  
PGEIFKYKWTVTVEDGPTKSDPRCLTRYSSFVNMERDLASGLIGPLLYCYKESVDQRGNOIMSDKRNVILF  
SVFDENRSWYLTENIQRFLEPNPAGVQLEDPEFQASNIMHSINGYVFDLSQLSVCLEHVAYWYILSIGAOTDF  
LSVFFSGYTFKHKMVYEDTLTLFFPSGETVFMSENPGLWILGCHNSDFRNRGMTALLKVS SCDKNTGDYIE  
DSYEDISAYLLSKNNAIEPRSFSONPPVLRHQRFEITRRTLQSDQEEIDYDDTISVEMKKEDFDIYDEDENQ  
SPRSFQKKTRHYFIAAVERLWDYGMSSSPHVLNRNRAQSGSVPQFKKVVFOEFTDGSFTQPLYRGELNEHLGL  
LGPIRAEVEDNIMVTFRNQASRPYSFYSSLSIYEEDQRQGAEPKRFVKNPNETKTYFWKVQHMAPTKDEF  
DCKAWAYFSDVDLEKDVHSGLIGPLLVCHTNTLNPAHGRQVTVQEFALFFTIFDETKSWYFTENMERNCRAP  
CNIQMEDPTFKENYRFHAINGYIMDTLPGLVMAQDQIRWYLLSMGSNENIHSIHFSGHVFTVRKKEEYKMA  
LYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPGLMASGHIRDFQITASGQYGQW  
APKLARLHYSGSINAWSTKEPFSWIKVDLLAPMI IHGIKTQGARQKFS SLYISQFIIMYSLDGKKWQTYRGN  
STGTLMVFFGNVDSSGIKHNI FNPP IARYIRLHPHYSIRSTLRMELMGCDLNSCSMPLGMESKAISDAQI

[0281]

TASSYFTNMFATWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTTQGVKSLLTSMYVKEFLI  
SSSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPELLTRYLRIHQPQSWVHQIALRMEVLGCEAQDLY

[0282] 表6. 编码BDD FVIII的核苷酸序列 (SEQ ID NO:68)\*

[0283]

```

661                A TGCAAATAGA GCTCTCCACC TGCTTCTTTC
721 TGTGCCTTTT GCGATTCTGC TTTAGTGCCA CCAGAAGATA CTACCTGGGT GCAGTGGAAC
781 TGTTCATGGGA CTATATGCAA AGTGATCTCG GTGAGCTGCC TGTGGACGCA AGATTTCCCTC
841 CTAGAGTGCC AAAATCTTTT CCATTCAACA CCTCAGTCGT GTACAAAAAG ACTCTGTTTG
901 TAGAATTCAC GGATCACCTT TTCAACATCG CTAAGCCAAG GCCACCCTGG ATGGGTCTGC
961 TAGGTCCTAC CATCCAGGCT GAGGTTTATG ATACAGTGGT CATTACACTT AAGAACATGG
1021 CTTCOCATCC TGTCAGTCTT CATGCTGTTG GTGTATCCTA CTGGAAAGCT TCTGAGGGAG
1081 CTGAATATGA TGATCAGACC AGTCAAAGGG AGAAAAGAAGA TGATAAAGTC TTCCCTGGTG
1141 GAAGCCATAC ATATGTCTGG CAGGTCCTGA AAGAGAATGG TCCAATGGCC TCTGACCCAC
1201 TGTGCCTTAC CTACTCATAT CTTTCTCATG TGGACCTGGT AAAAGACTTG AATTCAGGCC
1261 TCATTGGAGC CCTACTAGTA TGTAGAGAAG GGAGTCTGGC CAAGGAAAAG ACACAGACCT
1321 TGCACAAATT TATACTACTT TTTGCTGTAT TTGATGAAGG GAAAAGTTGG CACTCAGAAA
1381 CAAAGAACTC CTTGATGCAG GATAGGGATG CTGCATCTGC TCGGGCCTGG CCTAAAATGC
1441 ACACAGTCAA TGGTTATGTA AACAGGTCTC TGCCAGGTCT GATTGGATGC CACAGGAAAT
1501 CAGTCTATTG GCATGTGATT GGAATGGGCA CCACTCCTGA AGTGCCTCA ATATTCCCTCG
1561 AAGGTCACAC ATTTCTTGTG AGGAACCATC GCCAGGCGTC CTTGGAAATC TCGCCAATAA
1621 CTTTCCTTAC TGCTCAAACA CTCTTGATGG ACCTTGACA GTTTCCTACTG TTTTGTCTATA
1681 TCTCTTCCCA CCAACATGAT GGCATGGAAG CTTATGTCAA AGTAGACAGC TGTCCAGAGG
1741 AACCCCAACT ACGAATGAAA AATAATGAAG AAGCGGAAGA CTATGATGAT GATCTTACTG
1801 ATTCTGAAAT GGATGTGGTC AGGTTTGATG ATGACAACCTC TCCTTCCTTT ATCCAAATTC
1861 GCTCAGTTGC CAAGAAGCAT CCTAAAACCTT GGGTACATTA CATTGCTGCT GAAGAGGAGG
1921 ACTGGGACTA TGCTCCCTTA GTCCTCGCCC CCGATGACAG AAGTTATAAAA AGTCAATATT
1981 TGAACAATGG CCCTCAGCGG ATTGGTAGGA AGTACAAAAA AGTCCGATTT ATGGCATAACA
2041 CAGATGAAAC CTTTAAGACT CGTGAAGCTA TTCAGCATGA ATCAGGAATC TTGGGACCTT
2101 TACTTTATGG GGAAGTTGGA GACACACTGT TGATTATATT TAAGAATCAA GCAAGCAGAC
2161 CATATAACAT CTACCCTCAC GGAATCACTG ATGTCCGTCC TTTGTATTCA AGGAGATTAC
2221 CAAAAGGTGT AAAACATTTG AAGGATTTTC CAATTCTGCC AGGAGAAATA TTCAAATATA
2281 AATGGACAGT GACTGTAGAA GATGGGCCAA CTAATCAGA TCCTCGGTGC CTGACCCGCT
2341 ATTACTCTAG TTTCGTTAAT ATGGAGAGAG ATCTAGCTTC AGGACTCATT GGCCCTCTCC
2401 TCATCTGCTA CAAAGAATCT GTAGATCAA GAGGAAACCA GATAATGTCA GACAAGAGGA
2461 ATGTCATCCT GTTTCTGTA TTTGATGAGA ACCGAAGCTG GTACCTCACA GAGAATATAC
2521 AACGCTTTCT CCCCATCCA GCTGGAGTGC AGCTTGAGGA TCCAGAGTTC CAAGCCTCCA
2581 ACATCATGCA CAGCATCAAT GGCTATGTTT TTGATAGTTT GCAGTTGTCA GTTTGTTTGC
2641 ATGAGGTGGC ATACTGGTAC ATTCTAAGCA TTGGAGCACA GACTGACTTC CTTTCTGTCT
2701 TCTTCTCTGG ATATACCCTC AAACACAAAA TGGTCTATGA AGACACACTC ACCCTATTCC
2761 CATTCTCAGG AGAAACTGTC TTCATGTCGA TGGAAAACCC AGGTCTATGG ATTCTGGGGT
2821 GCCACAACCTC AGACTTTCGG AACAGAGGCA TGACCGCCTT ACTGAAGGTT TCTAGTTGTG
2881 ACAAGAACAC TGGTGATTAT TACGAGGACA GTTATGAAGA TATTTAGCA TACTTGCTGA
2941 GTAAAAACAA TGCCATTGAA CCAAGAAGCT TCTCTCAAAA CCCACCAGTC TTGAAACGCC
3001 ATCAACGGGA AATAACTCGT ACTACTCTTC AGTCAGATCA AGAGGAAATT GACTATGATG
3061 ATACCATATC AGTTGAAATG AAGAAGGAAG ATTTTGACAT TTATGATGAG GATGAAAATC
3121 AGAGCCCCCG CAGCTTCAA AAGAAAACAC GACTATTTT TATTGCTGCA GTGGAGAGGC
3181 TCTGGGATTA TGGGATGAGT AGCTCCCCAC ATGTTCTAAG AAACAGGGCT CAGAGTGGCA
3241 GTGTCCCTCA GTTCAAGAAA GTTGTTTTCC AGGAATTTAC TGATGGCTCC TTTACTCAGC
3301 CCTTATACCG TGGAGAACTA AATGAACATT TGGGACTCCT GGGGCCATAT ATAAGAGCAG
3361 AAGTTGAAGA TAATATCATG GTAACCTTCA GAAATCAGGC CTCTCGTCCC TATTCCTTCT
3421 ATTCTAGCCT TATTTCTTAT GAGGAAGATC AGAGGCAAGG AGCAGAACCT AGAAAAACT
3481 TTGTCAAGCC TAATGAAACC AAAACTTACT TTTGGAAAGT GCAACATCAT ATGGCACCCA
3541 CTAAAGATGA GTTTGACTGC AAAGCCTGGG CTTATTTCTC TGATGTTGAC CTGGAAAAG
3601 ATGTGCACTC AGGCCTGATT GGACCCCTTC TGGTCTGCCA CACTAACACA CTGAACCCTG

```

3661	CTCATGGGAG	ACAAGTGACA	GTACAGGAAT	TTGCTCTGTT	TTTCACCATC	TTTGATGAGA	
3721	CCAAAAGCTG	GTACTTCACT	GAAAATATGG	AAAGAAACTG	CAGGGCTCCC	TGCAATATCC	
3781	AGATGGAAGA	TCCCACITTT	AAAGAGAATT	ATCGCTTCCA	TGCAATCAAT	GGCTACATAA	
3841	TGGATACACT	ACCTGGCTTA	GTAATGGCTC	AGGATCAAAG	GATTTCGATGG	TATCTGCTCA	
3901	GCATGGGCAG	CAATGAAAAC	ATCCATTCTA	TTCATTTTCAG	TGGACATGTG	TTCACTGTAC	
3961	GAAAAAAGA	GGAGTATAAA	ATGGCACTGT	ACAATCTCTA	TCCAGGTGTT	TTTGAGACAG	
4021	TGGAAATGTT	ACCATCCAAA	GCTGGAATTT	GGCGGGTGG	ATGCCTTATT	GGCGAGCATC	
4081	TACATGCTGG	GATGAGCACA	CTTTTCTGG	TGTACAGCAA	TAAGTGTGAG	ACTCCCCTGG	
4141	GAATGGCTTC	TGGACACATT	AGAGATTTTC	AGATTACAGC	TTCAGGACAA	TATGGACAGT	
4201	GGGCCCCAAA	GCTGGCCAGA	CTTCATTATT	CCGGATCAAT	CAATGCCTGG	AGCACCAAGG	
4261	AGCCCTTTTC	TTGGATCAAG	GTGGATCTGT	TGGCACCAAT	GATTATTTCAC	GGCATCAAGA	
4321	CCCAGGGTGC	CCGTCAGAAG	TTCTCCAGCC	TCTACATCTC	TCAGTTTATC	ATCATGTATA	
[0284]	4381	GTCTTGATGG	GAAGAAGTGG	CAGACTTATC	GAGGAAATTC	CACTGGAACC	TTAATGGTCT
4441	TCTTTGGCAA	TGTGGATTCA	TCTGGGATAA	AACACAATAT	TTTTAACCCCT	CCAATTATTG	
4501	CTCGATACAT	CCGTTTGCAC	CCAACTCATT	ATAGCATTCG	CAGCACTCTT	CGCATGGAGT	
4561	TGATGGGCTG	TGATTTAAAT	AGTTGCAGCA	TGCCATTGGG	AATGGAGAGT	AAAGCAATAT	
4621	CAGATGCACA	GATTACTGCT	TCATCCTACT	TTACCAATAT	GTTTGCCACC	TGGTCTCCTT	
4681	CAAAAGCTCG	ACTTCACCTC	CAAGGGAGGA	GTAATGCCTG	GAGACCTCAG	GTGAATAATC	
4741	CAAAAGAGTG	GCTGCAAGTG	GACTTCCAGA	AGACAATGAA	AGTCACAGGA	GTAACTACTC	
4801	AGGGAGTAAA	ATCTCTGCTT	ACCAGCATGT	ATGTGAAGGA	GTTCCCTCATC	TCCAGCAGTC	
4861	AAGATGGCCA	TCAGTGGACT	CTCTTTTTTC	AGAATGGCAA	AGTAAAGGTT	TTTCAGGGAA	
4921	ATCAAGACTC	CTTCACACCT	GTGGTGAACT	CTCTAGACCC	ACCGTTACTG	ACTCGCTACC	
4981	TTCGAATTCA	CCCCCAGAGT	TGGGTGCACC	AGATTGCCCT	GAGGATGGAG	GTTCTGGGCT	
5041	GCGAGGCACA	GGACCTCTAC					

[0285] \*加下划线的核酸编码信号肽。

[0286] “B结构域缺失的FVIII”可具有公开于美国专利号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中的完全或部分缺失。在一些实施方案中,本发明的B结构域缺失的FVIII序列包含在美国专利号6,316,226(也在US 6,346,513中)的第4栏第4行至第5栏第28行和实施例1-5处公开的任一缺失。在另一实施方案中,B结构域缺失的因子VIII是S743/Q1638 B结构域缺失的因子VIII(SQ BDD FVIII)(例如因子VIII具有从氨基酸744至氨基酸1637的缺失,例如因子VIII具有SEQ ID NO:65的氨基酸1-743和氨基酸1638-2332,即SEQ ID NO:67)。在一些实施方案中,本发明的B结构域缺失的FVIII具有在美国专利号5,789,203(以及US 6,060,447、US 5,595,886和US 6,228,620)的第2栏第26-51行和实施例5-8处公开的缺失。在一些实施方案中,B结构域缺失的因子VIII具有以下中描述的缺失: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。在一些实施方案中,B结构域缺失的FVIII缺失大部分B结构域,但仍然含有为在体内将初级翻译产物蛋白水解加工成两个多肽链所必需的B结构域氨基末端序列,如WO 91/09122中所公开。在一些实施方案中,在缺失氨基酸747-1638,即实际上完全缺失B结构域下构建B结构域缺失的FVIII。Hoeben R.C.等J.Biol.Chem.265(13):7318-7323(1990)。B结构域缺失的因子VIII也可含有FVIII的氨基酸771-1666或氨基酸868-1562的缺失。Meulien P.等Protein Eng.2(4):301-6(1988)。作为本发明的一部分的另外的B结构域缺失包括以下:缺失氨基酸982至1562或760至1639(Toole等,Proc.Natl.Acad.Sci.U.S.A.(1986)83,5939-5942))、

797至1562(Eaton等Biochemistry(1986)25:8343-8347)、741至1646(Kaufman(PCT公布申请号WO 87/04187))、747-1560(Sarver等,DNA(1987)6:553-564)、741至1648(Pasek(PCT申请号88/00831))、或816至1598或741至1648(Lagner(Behring Inst.Mitt.(1988)第82期:16-25);EP 295597)。在其它实施方案中,BDD FVIII包括含有B结构域的保留一个或多个N连接的糖基化位点的片段的FVIII多肽,所述位点例如是对应于全长FVIII序列的氨基酸序列的残基757、784、828、900、963或任选943。B结构域片段的实例包括如Miao,H.Z.等,Blood 103(a):3412-3419(2004);Kasuda,A等,J.Thromb.Haemost.6:1352-1359(2008);以及Pipe,S.W.等,J.Thromb.Haemost.9:2235-2242(2011)中公开的B结构域的226个氨基酸或163个氨基酸(即保留B结构域的前226个氨基酸或163个氨基酸)。在其它实施方案中,BDD FVIII进一步包含在残基309处的点突变(从Phe成为Ser)以改进BDD FVIII蛋白的表达。参见Miao,H.Z.等,Blood 103(a):3412-3419(2004)。在其它实施方案中,BDD FVIII包括含有B结构域的一部分,但不含有一个或多个弗林蛋白酶裂解位点(例如Arg1313和Arg1648)的FVIII多肽。参见Pipe,S.W.等,J.Thromb.Haemost.9:2235-2242(2011)。可在任何FVIII序列中进行各前述缺失。

[0287] 在一些实施方案中,FVIII具有部分B结构域。在一些实施方案中,具有部分B结构域的FVIII蛋白是FVIII198。FVIII198是一种含有部分B结构域的单链FVIII Fc分子-226N6。数字226表示FVIII B结构域的N末端226个氨基酸,并且N6表示B结构域中的六个N-糖基化位点。

[0288] 在一个实施方案中,就在氨基酸1648(在全长因子VIII或SEQ ID NO:65中)、氨基酸754(在S743/Q1638 B结构域缺失的因子VIII或SEQ ID NO:67中)处的精氨酸或相应精氨酸残基(在其它变体中)之后裂解FVIII,由此产生重链和轻链。在另一实施方案中,FVIII包含通过金属离子介导的非共价键连接或缔合的重链和轻链。

[0289] 在其它实施方案中,FVIII是单链FVIII,其尚未就在氨基酸1648(在全长FVIII或SEQ ID NO:65中)、氨基酸754(在S743/Q1638 B结构域缺失的FVIII或SEQ ID NO:67中)处的精氨酸或相应精氨酸残基(在其它变体中)之后被裂解。单链FVIII可包含一个或多个氨基酸取代。在一个实施方案中,氨基酸取代在对应于全长成熟因子VIII多肽(SEQ ID NO:65)的残基1648、残基1645或两者或SQ BDD因子VIII(SEQ ID NO:67)的残基754、残基751或两者的残基处。氨基酸取代可为除精氨酸以外的任何氨基酸,例如异亮氨酸、亮氨酸、赖氨酸、甲硫氨酸、苯丙氨酸、苏氨酸、色氨酸、缬氨酸、丙氨酸、天冬酰胺、天冬氨酸、半胱氨酸、谷氨酸、谷氨酰胺、甘氨酸、脯氨酸、硒代半胱氨酸、丝氨酸、酪氨酸、组氨酸、鸟氨酸、吡咯赖氨酸或牛磺酸。

[0290] FVIII可进一步由凝血酶裂解,接着活化成FVIIIa,从而充当活化因子IX(FIXa)的辅因子。活化FIX接着连同活化FVIII一起形成X酶复合物,并且使因子X转化成活化因子X(FXa)。对于活化,FVIII由凝血酶在氨基酸372、740和1689(对应于B结构域缺失的FVIII序列中的氨基酸372、740和795)处的三个精氨酸残基之后裂解,所述裂解产生具有50kDa A1、43kDa A2和73kDa A3-C1-C2链的FVIIIa。在一个实施方案中,适用于本发明的FVIII蛋白是非活性FVIII。在另一实施方案中,FVIII蛋白是活化FVIII。

[0291] 具有与VWF蛋白连接或缔合的FVIII多肽的蛋白质可包含与SEQ ID NO:65或67至少50%、60%、70%、80%、90%、95%、96%、97%、98%、99%或100%同一的序列,其中所述

序列具有FVIII凝结活性,例如活化因子IX作为辅因子以使因子X转化成活化因子X(FXa)。

[0292] 如本文所用的“杂交”或“嵌合”多肽和蛋白质包括第一多肽链(例如融合于具有小于288个氨基酸的XTEN序列和第一Ig恒定区或其一部分的VWF蛋白)与第二多肽链(例如融合于第二Ig恒定区或其一部分的FVIII蛋白)的组合,由此形成异二聚体。在一个实施方案中,杂交物中的第一多肽与第二多肽通过蛋白质-蛋白质相互作用(诸如电荷-电荷或疏水性相互作用)彼此缔合。在另一实施方案中,第一多肽包含VWF蛋白-XTEN-Fc融合蛋白,并且第二多肽包含FVIII-Fc融合蛋白,从而使得杂交物是异二聚体,其中XTEN含有小于288个氨基酸。在其它实施方案中,第一多肽包含VWF蛋白-XTEN-Fc融合蛋白,并且第二多肽包含FVIII(X)-Fc融合蛋白,从而使得杂交物是异二聚体,其中XTEN含有小于288个氨基酸。第一多肽与第二多肽可通过第一Fc区与第二Fc区之间的共价键(例如二硫键)缔合。第一多肽与第二多肽可进一步通过VWF片段与FVIII蛋白之间的结合而彼此缔合。

[0293] 适用于本发明中的FVIII蛋白可包括具有一个或多个不影响FVIII凝血活性的另外的XTEN序列的FVIII。所述XTEN序列可融合于FVIII蛋白的C末端或N末端或插入在FVIII蛋白中一组或多组两个氨基酸残基之间,同时插入不影响FVIII凝血活性或FVIII功能。在一个实施方案中,插入改进FVIII蛋白的药物动力学性质(例如半衰期)。在另一实施方案中,插入可为多个插入,例如大于两个、三个、四个、五个、六个、七个、八个、九个或十个插入。插入位点的实例包括但不限于表7、8、9、10、11、12、13、14、15中所列的位点或其任何组合。

[0294] 连接于一个或多个XTEN序列的FVIII蛋白可表示为FVIII(X<sub>2</sub>)或FVIII<sub>(a→b)</sub>-X-FVIII<sub>(c→d)</sub>,其中FVIII<sub>(a→b)</sub>包含以下、基本上由以下组成或由以下组成:FVIII蛋白的从氨基酸残基“a”至氨基酸残基“b”的第一部分;X<sub>2</sub>包含以下、基本上由以下组成或由以下组成:一个或多个XTEN序列,FVIII<sub>(c→d)</sub>包含以下、基本上由以下组成或由以下组成:FVIII蛋白的从氨基酸残基“c”至氨基酸残基“d”的第二部分;

[0295] a是FVIII蛋白的第一部分的N末端氨基酸残基,

[0296] b是FVIII蛋白的第一部分的C末端氨基酸残基,但也是其中插入XTEN序列的插入位点的两个氨基酸的N末端氨基酸残基,

[0297] c是FVIII蛋白的第二部分的N末端氨基酸残基,但也是其中插入XTEN序列的插入位点的两个氨基酸的C末端氨基酸残基,并且

[0298] d是FVIII蛋白的C末端氨基酸残基,并且

[0299] 其中FVIII蛋白的第一部分与FVIII蛋白的第二部分彼此不相同,并且共同具有足以使FVIII蛋白具有FVIII凝血活性的长度。

[0300] 在一个实施方案中,FVIII蛋白的第一部分和FVIII蛋白的第二部分是SEQ ID NO: 65[全长成熟FVIII序列]或SEQ ID NO:67[B结构域缺失的FVIII]的片段,例如分别是N末端部分和C末端部分。在某些实施方案中,FVIII蛋白的第一部分包含FVIII蛋白的A1结构域和A2结构域。FVIII蛋白的第二部分包含A3结构域、C1结构域和任选C2结构域。在其它实施方案中,FVIII蛋白的第一部分包含A1结构域和A2结构域,并且FVIII蛋白的第二部分包含B结构域的一部分、A3结构域、C1结构域和任选C2结构域。在其它实施方案中,FVIII蛋白的第一部分包含FVIII蛋白的A1结构域、A2结构域和B结构域的一部分,并且FVIII蛋白的第二部分包含A3结构域、C1结构域和任选C2结构域。在其它实施方案中,FVIII蛋白的第一部分包含

FVIII蛋白的A1结构域、A2结构域和B结构域的第一部分。FVIII蛋白的第二部分包含B结构域的第二部分、A3结构域、C1结构域和任选C2结构域。在一些实施方案中，两个氨基酸(“b”和“c”)可为表7、8、9、10、11、12、13、14和15中所示的氨基酸残基插入位点中的任何一个或多个。举例来说，“b”可为紧靠其中插入或连接一个或多个XTEN序列的位点的上游的氨基酸残基，并且“c”可为紧靠其中插入或连接一个或多个XTEN序列的位点的下游的氨基酸残基。在一些实施方案中，“a”是FVIII蛋白的首个成熟氨基酸序列，并且“d”是FVIII蛋白的末个氨基酸序列。举例来说，FVIII<sub>(a→b)</sub>可为与SEQ ID NO:67[B结构域缺失的FVIII氨基酸序列]或SEQ ID NO:65[全长FVIII]的氨基酸1至745至少70%、80%、90%、95%、96%、97%、98%、99%或100%同一的氨基酸序列，并且FVIII<sub>(c→d)</sub>可为分别SEQ ID NO:67的氨基酸746至1438或SEQ ID NO:65的氨基酸1641至2332。

[0301] 在一些方面，FVIII蛋白中的插入位点位于FVIII蛋白的一个或多个结构域(其是N末端、A1结构域、A2结构域、A3结构域、B结构域、C1结构域、C2结构域、C末端或其两个或更多个组合)中或位于FVIII蛋白的两个结构域(其是A1结构域和a1酸性区、以及a1酸性区和A2结构域、A2结构域和a2酸性区、a2酸性区和B结构域、B结构域和A3结构域、以及A3结构域和C1结构域、C1结构域和C2结构域或其任何组合)之间。举例来说，其中可插入XTEN序列的插入位点选自由以下组成的组：N末端和A1结构域、N末端和A2结构域、N末端和A3结构域、N末端和B结构域、N末端和C1结构域、N末端和C2结构域、N末端和C末端、A1结构域和A2结构域、A1结构域和A3结构域、A1结构域和B结构域、A1结构域和C1结构域、A1结构域和C2结构域、A1结构域和C末端、A2结构域和A3结构域、A2结构域和B结构域、A2结构域和C1结构域、A2结构域和C2结构域、A2结构域和C末端、A3结构域和B结构域、A3结构域和C1结构域、A3结构域和C2结构域、A3结构域和C末端、B结构域和C1结构域、B结构域和C2结构域、B结构域和C末端、C1结构域和C2结构域、C1结构域和C末端、C2结构域和C末端及其两个或更多个组合。插入位点的非限制性实例列于表7、8、9、10、11、12、13、14和15中。

[0302] 其中紧靠FVIII蛋白中的一个或多个氨基酸(例如一个或多个XTEN插入位点)的下游插入或在C末端或N末端连接XTEN序列的FVIII蛋白在连接于XTEN序列或由XTEN序列插入之后保留FVIII活性。XTEN序列可插入FVIII蛋白中一次或大于一次、两次、三次、四次、五次或六次以使插入不影响FVIII活性(即FVIII蛋白仍然保留凝血性质)。

[0303] 适用于本发明中的FVIII蛋白可通过任选接头在FVIII蛋白的N末端或C末端连接于一个或多个XTEN多肽或通过一个或多个任选接头紧靠FVIII蛋白中的一个或多个氨基酸(例如一个或多个XTEN插入位点)的下游插入。在一个实施方案中，其中插入XTEN序列的两个氨基酸残基或XTEN序列所连接的氨基酸残基对应于SEQ ID NO:65[全长成熟FVIII]的两个或一个选自由表7、表8、表9和表10中的残基及其任何组合组成的组的氨基酸残基。

[0304] 在其它实施方案中，至少一个XTEN序列插入本文公开的任何一个或多个XTEN插入位点或其任何组合中。在一个方面，至少一个XTEN序列插入表7中公开的一个或多个氨基酸中公开的一个或多个XTEN插入位点中。

[0305] 表7：示例性XTEN插入位点

[0306]

编号	XTEN 插入点 *	插入残基	FVIII BDD 下游序列	FVIII 结 构域
1	0	(N 末端)	ATR	A1
2	3	R	RYY	A1
3	17	M	QSD	A1
4	18	Q	SDL	A1
5	22	G	ELP	A1
6	24	L	PVD	A1
7	26	V	DAR	A1
8	28	A	RFP	A1
9	32	P	RVP	A1
10	38	F	PFN	A1
11	40	F	NTS	A1
12	41	N	TSV	A1
13	60	N	IAK	A1
14	61	I	AKP	A1
15	65	R	PPW	A1
16	81	Y	DTV	A1
17	111	G	AEY	A1
18	116	D	QTS	A1
19	119	S	QRE	A1
20	120	Q	REK	A1
21	128	V	FPG	A1
22	129	F	PGG	A1
23	130	P	GGG	A1
24	182	G	SLA	A1
25	185	A	KEK	A1
26	188	K	TQT	A1
27	205	G	KSW	A1
28	210	S	ETK	A1
29	211	E	TKN	A1
30	216	L	MQD	A1
31	220	R	DAA	A1
32	222	A	ASA	A1
33	223	A	SAR	A1
34	224	S	ARA	A1
35	230	K	MHT	A1
36	243	P	GLI	A1
37	244	G	LIG	A1



[0307]

编号	XTEN 插入点 *	插入残基	FVIII BDD 下游序列	FVIII 结 构域
38	250	R	KSV	A1
39	318	D	GME	A1
40	333	P	QLR	A1
42	334	Q	LRM	A1
43	336	R	MKN	a1
44	339	N	NEE	a1
45	345	D	YDD	a1
46	357	V	VRF	a1
47	367	S	FIQ	a1
48	370	S	RPY	a1
49	375	A	KKH	A2
50	376	K	KHP	A2
51	378	H	PKT	A2
52	399	V	LAP	A2
53	403	D	DRS	A2
54	405	R	SYK	A2
55	409	S	QYL	A2
56	416	P	QRI	A2
57	434	E	TFK	A2
58	438	T	REA	A2
59	441	A	IQH	A2
60	442	I	QHE	A2
61	463	I	IFK	A2
62	487	Y	SRR	A2
63	490	R	LPK	A2
64	492	P	KGV	A2
65	493	K	GVK	A2
66	494	G	VKH	A2
67	500	D	FPI	A2
68	506	G	EIF	A2
69	518	E	DGP	A2
70	556	K	ESV	A2
71	565	Q	IMS	A2
72	566	I	MSD	A2
73	598	P	AGV	A2
74	599	A	GVQ	A2
75	603	L	EDP	A2
76	616	S	ING	A2
77	686	G	LWI	A2

[0308]

编号	XTEN 插入点 *	插入残基	FVIII BDD 下游序列	FVIII 结 构域
78	713	K	NTG	A2
79	719	Y	EDS	A2
80	730	L	LSK	A2
81	733	K	NNA	A2
82	745	N	PPV**	B
83	1640	P	PVL	B
84	1652	R	TTL	B
85	1656	Q	SDQ	A3
86	1685	N	QSP	A3
87	1711	M	SSS	A3
88	1713	S	SPH	A3
89	1720	N	RAQ	A3
90	1724	S	GSV	A3
91	1725	G	SVP	A3
92	1726	S	VPQ	A3
93	1741	G	SFT	A3
94	1744	T	QPL	A3
95	1749	R	GEL	A3
96	1773	V	TFR	A3
97	1792	Y	EED	A3
98	1793	E	EDQ	A3
99	1796	Q	RQG	A3
100	1798	Q	GAE	A3
101	1799	G	AEP	A3
102	1802	P	RKN	A3
103	1803	R	KNF	A3
104	1807	V	KPN	A3
105	1808	K	PNE	A3
106	1827	K	DEF	A3
107	1844	E	KDV	A3
108	1861	N	TLN	A3
109	1863	L	NPA	A3
110	1896	E	RNC	A3
111	1900	R	APC	A3
112	1904	N	IQM	A3
113	1905	I	QME	A3
114	1910	P	TFK	A3
115	1920	A	ING	A3
116	1937	D	QRI	A3

编号	XTEN 插入点 *	插入残基	FVIII BDD 下游序列	FVIII 结 构域
117	1981	G	VFE	A3
118	2019	N	KCQ	A3
119	2020	K	CQT	C1
120	2044	G	QWA	C1
121	2068	F	SWI	C1
122	2073	V	DLL	C1
123	2090	R	QKF	C1
124	2092	K	FSS	C1
125	2093	F	SSL	C1
126	2111	K	WQT	C1
127	2115	Y	RGN	C1
[0309] 128	2120	T	GTL	C1
129	2125	V	FFG	C1
130	2171	L	NSC	C1
131	2173	S	CSM	C2
132	2188	A	QIT	C2
133	2223	V	NNP	C2
134	2224	N	NPK	C2
135	2227	K	EWL	C2
136	2268	G	HQW	C2
137	2277	N	GKV	C2
138	2278	G	KVK	C2
139	2290	F	TPV	C2
140	2332	Y	FVIII 的 C 末端	CT

[0310] \*指示XTEN的基于成熟全长人FVIII的氨基酸编号的插入点,其中插入可在指示的氨基酸的N末端侧或C末端侧上。

[0311] 在一些实施方案中,一个或多个XTEN序列插入在对应于SEQ ID NO:65的氨基酸32、220、224、336、339、399、416、603、1656、1711、1725、1905或1910或其任何组合的上游或下游约六个氨基酸内。

[0312] 表8. 示例性XTEN插入范围

编号	XTEN 插入点	插入残基	FVIII BDD 下游序列	FVIII 结构域	离插入残基的距离*
9	32	P	RVP	A1	-3, +6
31	220	R	DAA	A1	-
34	224	S	ARA	A1	+5
43	336	R	MKN	a1	-1, +6
44	339	N	NEE	a1	-4, +5
[0313] 52	399	V	LAP	A2	-6, +3
56	416	P	QRI	A2	+6
75	603	L	EDP	A2	6, +6
85	1656	Q	SDQ	B	-3, +6
87	1711	M	SSS	A3	-6, +1
91	1725	G	SVP	A3	+6
113	1905	I	QME	A3	+6
114	1910	P	TFK	A3	-5, +6

[0314] \*离插入残基的距离是指离可进行插入所处的指定插入残基(残基“0”)的N末端(负数)或C末端(正数)的相对氨基酸数目。标号“-x”是指插入位点离指定插入残基的N末端侧x个氨基酸。类似地,标号“+x”是指插入位点离指定插入残基的C末端侧x个氨基酸。

[0315] 举例来说,“-1,+2”指示在表示为-1、0、+1或+2的氨基酸残基的N末端或C末端进行插入。

[0316] 在其它实施方案中,一个或多个XTEN序列紧靠对应于全长成熟人FVIII的一个或多个选自表9中的一个或多个插入位点组成的组的氨基酸的下游插入。

[0317] 表9. 示例性XTEN插入位点或范围

编号	XTEN 插入点范围*	第一插入残基	FVIII 结构域
3	18-32	Q	A1
8	40	F	A1
[0318] 18	211-224	E	A1
27	336-403	R	A1, A2
43	599	A	A2
47	745-1640	N	B
50	1656-1728	Q	B, a3, A3

编号	XTEN 插入点范围*	第一插入残基	FVIII 结构域
[0319] 57	1796-1804	R	A3
65	1900-1912	R	A3
81	2171-2332	L	C1, C2

[0320] \*指示相对于成熟人FVIII的氨基酸编号加以编号的插入位点的范围

[0321] 在其它实施方案中,一个或多个XTEN插入FVIII的B结构域中。在一个实例中,XTEN插入在对应于SEQ ID NO:65的氨基酸740与1640之间,其中氨基酸740与1640之间的FVIII

序列任选不存在。在另一实例中,XTEN插入在对应于SEQ ID NO:65的氨基酸741与1690之间,其中氨基酸740与1690之间的FVIII序列任选不存在。在其它实例中,XTEN插入在对应于SEQ ID NO:65的氨基酸741与1648之间,其中氨基酸741与1648之间的FVIII序列任选不存在。在其它实例中,XTEN插入在对应于SEQ ID NO:65的氨基酸743与1638之间,其中氨基酸743与1638之间的FVIII序列任选不存在。在其它实例中,XTEN插入在对应于SEQ ID NO:65的氨基酸745与1656之间,其中氨基酸745与1656之间的FVIII序列任选不存在。在一些实例中,XTEN插入在对应于SEQ ID NO:65的氨基酸745与1657之间,其中氨基酸745与1657之间的FVIII序列任选不存在。在某些实例中,XTEN插入在对应于SEQ ID NO:65的氨基酸745与1667之间,其中氨基酸745与1667之间的FVIII序列任选不存在。在其它实例中,XTEN插入在对应于SEQ ID NO:65的氨基酸745与1686之间,其中氨基酸745与1686之间的FVIII序列任选不存在。在一些其它实例中,XTEN插入在对应于SEQ ID NO:65的氨基酸747与1642之间,其中氨基酸747与1642之间的FVIII序列任选不存在。在其它实例中,XTEN插入在对应于SEQ ID NO:65的氨基酸751与1667之间,其中氨基酸751与1667之间的FVIII序列任选不存在。

[0322] 在一些实施方案中,一个或多个XTEN插入紧靠选自表10中的氨基酸残基组成的组的插入位点的氨基酸的下游的一个或多个氨基酸中。

[0323] 表10:FVIII XTEN插入位点和构建体标号

[0324]

构建体编号	结构域	上游残基编号*	下游残基编号*	上游序列	下游序列
F8X-1	A1	3	4	ATR	RYY
F8X-2	A1	18	19	YMQ	SDL
F8X-3	A1	22	23	DLG	ELP
F8X-4	A1	26	27	LPV	DAR
F8X-5	A1	40	41	FPF	NTS
F8X-6	A1	60	61	LFN	IAK
F8X-7	A1	116	117	YDD	QTS
F8X-8	A1	130	131	VFP	GGG
F8X-9	A1	188	189	KEK	TQT
F8X-10	A1	216	217	NSL	MQD
F8X-11	A1	230	231	WPK	MHT
F8X-12	A1	333	334	EED	QLR
F8X-13	A2	375	376	SVA	KKH
F8X-14	A2	403	404	APD	DRS
F8X-15	A2	442	443	EAI	QHE
F8X-16	A2	490	491	RRL	PKG
F8X-17	A2	518	519	TVE	DGP
F8X-18	A2	599	600	NPA	GVQ
F8X-19	A2	713	714	CDK	NTG
F8X-20	BD	745	746	SQN	PPV
F8X-21	BD	745	746	SQN	PPV
F8X-22	BD**	745	746	SQN	PPV
F8X-23	A3	1720	1721	APT	KDE
F8X-24	A3	1796	1797	EDQ	RQG
F8X-25	A3	1802	1803	AEP	RKN
F8X-26	A3	1827	1828	PTK	DEF
F8X-27	A3	1861	1862	HTN	TLN
F8X-28	A3	1896	1897	NME	RNC
F8X-29	A3	1900	1901	NCR	APC
F8X-30	A3	1904	1905	PCN	IQM
F8X-31	A3	1937	1938	AQD	QRI
F8X-32	C1	2019	2020	YSN	KCQ
F8X-33	C1	2068	2069	EPF	SWI
F8X-34	C1	2111	2112	GKK	WQT
F8X-35	C1	2120	2121	NST	GTL

[0325]

F8X-36	C2	2171	2172	CDL	NSC
F8X-37	C2	2188	2189	SDA	QIT
F8X-38	C2	2227	2228	NPK	EWL
F8X-39	C2	2277	2278	FQN	GKV
F8X-40	CT	2332	NA	DLY	NA
F8X-41	CT	2332	NA	DLY	NA
F8X-42	A1	3	4	ATR	ATR
pSD0001	A2	403	404		
pSD0002	A2	599	600		
pSD0021	N 未 端	0	1		
pSD0022	A1	32	33		
pSD0023	A1	65	66		
pSD0024	A1	81	82		
pSD0025	A1	119	120		
pSD0026	A1	211	212		
pSD0027	A1	220	221		
pSD0028	A1	224	225		
pSD0029	A1	336	337		
pSD0030	A1	339	340		
pSD0031	A2	378	379		
pSD0032	A2	399	400		
pSD0033	A2	409	410		
pSD0034	A2	416	417		
pSD0035	A2	487	488		
pSD0036	A2	494	495		
pSD0037	A2	500	501		
pSD0038	A2	603	604		
pSD0039	A3	1656	1657		
pSD0040	A3	1711	1712		
pSD0041	A3	1725	1726		
pSD0042	A3	1749	1750		
pSD0043	A3	1905	1906		
pSD0044	A3	1910	1911		
pDS0062	A3	1900	1901		

[0326] \*指示成熟FVIII蛋白的氨基酸编号

[0327] 在一个实施方案中,一个或多个XTEN插入位点位于FVIII蛋白的一个或多个表面暴露的柔性环结构(例如容许环)内。举例来说,至少一个XTEN序列可插入包含其中可插入至少一个XTEN多肽而不消除重组蛋白的促凝血活性或重组蛋白质在体内表达或在体外于宿主细胞中表达的能力的至少两个“容许环”的各FVIII“A”结构域中。容许环是允许插入至

少一个XTEN序列的除了其它属性之外也具有高表面或溶剂暴露和高构象柔性的区域。A1结构域包含容许环-1(A1-1)区域和容许环-2(A1-2)区域，A2结构域包含容许环-1(A2-1)区域和容许环-2(A2-2)区域，A3结构域包含容许环-1(A3-1)区域和容许环-2(A3-2)区域。

[0328] 在一个方面，FVIII A1结构域中的第一容许环(A1-1)位于 $\beta$ 链1与 $\beta$ 链2之间，并且FVIII A2结构域中的第二容许环(A1-2)位于 $\beta$ 链11与 $\beta$ 链12之间。FVIII A2结构域中的第一容许环(A2-1)位于 $\beta$ 链22与 $\beta$ 链23之间，并且FVIII A2结构域中的第二容许环(A2-2)位于 $\beta$ 链32与 $\beta$ 链33之间。FVIII A3结构域中的第一容许环(A3-1)位于 $\beta$ 链38与 $\beta$ 链39之间，并且FVIII A3中的第二容许环(A3-2)位于 $\beta$ 链45与 $\beta$ 链46之间。在某些方面，包含A1-1的表面暴露的柔性环结构对应于天然成熟人FVIII中从SEQ ID NO:65的约氨基酸15至约氨基酸45，例如从SEQ ID NO:65的约氨基酸18至约氨基酸41的区域。在其它方面，包含A1-2的表面暴露的柔性环结构对应于天然成熟人FVIII中从SEQ ID NO:65的约氨基酸201至约氨基酸232，例如从SEQ ID NO:65的约氨基酸218至约氨基酸229的区域。在其它方面，包含A2-1的表面暴露的柔性环结构对应于天然成熟人FVIII中从SEQ ID NO:65的约氨基酸395至约氨基酸421，例如从SEQ ID NO:65的约氨基酸397至约氨基酸418的区域。在其它实施方案中，包含A2-2的表面暴露的柔性环结构对应于天然成熟人FVIII中从SEQ ID NO:65的约氨基酸577至约氨基酸635，例如从SEQ ID NO:65的约氨基酸595至约氨基酸607的区域。在某些方面，包含A3-1的表面暴露的柔性环结构对应于天然成熟人FVIII中从SEQ ID NO:65的约氨基酸1705至约氨基酸1732，例如从SEQ ID NO:65的约氨基酸1711至约氨基酸1725的区域。在其它方面，包含A3-2的表面暴露的柔性环结构对应于天然成熟人FVIII中从SEQ ID NO:65的约氨基酸1884至约氨基酸1917，例如从SEQ ID NO:65的约氨基酸1899至约氨基酸1911的区域。

[0329] 在另一实施方案中，其中插入至少一个XTEN序列的一个或多个氨基酸位于a3结构域，例如对应于全长成熟FVIII多肽的氨基酸1649至1689内。在特定实施方案中，XTEN序列插入在SEQ ID NO:65(全长成熟FVIII)的氨基酸1656与1657之间。在特定实施方案中，包含紧靠对应于SEQ ID NO:65的氨基酸1656的下游插入的XTEN序列的FVIII蛋白进一步包含从对应于SEQ ID NO:65的氨基酸745至氨基酸1656的缺失。

[0330] 在一些实施方案中，用于一个或多个XTEN插入的一个或多个插入位点紧靠一个或多个选自由以下组成的组的对应于成熟全长FVIII的氨基酸的下游：

- [0331] (1)氨基酸3, (2)氨基酸18, (3)氨基酸22,
- [0332] (4)氨基酸26, (5)氨基酸32, (6)氨基酸40,
- [0333] (7)氨基酸60, (8)氨基酸65, (9)氨基酸81,
- [0334] (10)氨基酸116, (11)氨基酸119, (12)氨基酸130,
- [0335] (13)氨基酸188, (14)氨基酸211, (15)氨基酸216,
- [0336] (16)氨基酸220, (17)氨基酸224, (18)氨基酸230,
- [0337] (19)氨基酸333, (20)氨基酸336, (21)氨基酸339,
- [0338] (22)氨基酸375, (23)氨基酸399, (24)氨基酸403,
- [0339] (25)氨基酸409, (26)氨基酸416, (26)氨基酸442,
- [0340] (28)氨基酸487, (29)氨基酸490, (30)氨基酸494,
- [0341] (31)氨基酸500, (32)氨基酸518, (33)氨基酸599,



- [0342] (34)氨基酸603, (35)氨基酸713, (36)氨基酸745,  
 [0343] (37)氨基酸1656, (38)氨基酸1711, (39)氨基酸1720,  
 [0344] (40)氨基酸1725, (41)氨基酸1749, (42)氨基酸1796,  
 [0345] (43)氨基酸1802, (44)氨基酸1827, (45)氨基酸1861,  
 [0346] (46)氨基酸1896, (47)氨基酸1900, (48)氨基酸1904,  
 [0347] (49)氨基酸1905, (50)氨基酸1910, (51)氨基酸1937,  
 [0348] (52)氨基酸2019, (53)氨基酸2068, (54)氨基酸2111,  
 [0349] (55)氨基酸2120, (56)氨基酸2171, (57)氨基酸2188,  
 [0350] (58)氨基酸2227, (59)氨基酸2277, 和  
 [0351] (60)其两个或更多个组合。

[0352] 在一个实施方案中,适用于本发明的FVIII蛋白包含两个XTEN序列,插入第一XTEN插入位点中的第一XTEN序列和插入第二XTEN插入位点中的第二XTEN。第一XTEN插入位点和第二XTEN插入位点的非限制性实例列于表11中。

[0353] 表11. 两个XTEN的示例性插入位点

插入 1		插入 2	
插入位点	结构域	插入位点	结构域
745	B	2332	CT
26	A1	403	A2
40	A1	403	A2
18	A1	403	A2
26	A1	599	A2
40	A1	599	A2

[0354]

[0355]

插入 1		插入 2	
插入位点	结构域	插入位点	结构域
18	A1	599	A2
1720	A3	1900	A3
1725	A3	1900	A3
1711	A3	1905	A3
1720	A3	1905	A3
1725	A3	1905	A3
1656	A3	26	A1
1656	A3	18	A1
1656	A3	40	A1
1656	A3	399	A2
1656	A3	403	A2
1656	A3	1725	A3
1656	A3	1720	A3
1900	A3	18	A1
1900	A3	26	A1
1900	A3	40	A1
1905	A3	18	A1
1905	A3	40	A1
1905	A3	26	A1
1910	A3	26	A1
18	A1	399	A2
26	A1	399	A2
40	A1	399	A2
18	A1	403	A2
1656	A3	1900	A3
1656	A3	1905	A3
1711	A3	40	A1
1711	A3	26	A1
1720	A3	26	A1
1720	A3	40	A1
1720	A3	18	A1
1725	A3	26	A1
1725	A3	40	A1
1725	A3	18	A1
1720	A3	403	A2
1720	A3	399	A2
1711	A3	403	A2
1720	A3	403	A2
1725	A3	403	A2

[0356]	插入 1		插入 2	
	插入位点	结构域	插入位点	结构域
	1725	A3	399	A2
	1711	A3	403	A2
	1900	A3	399	A2
	1900	A3	403	A2
	1905	A3	403	A2
	1905	A3	399	A2
	1910	A3	403	A2

[0357] 插入或连接于FVIII蛋白的两个XTEN可相同或不同。在一些实施方案中,适用于本发明的FVIII蛋白包含两个插入FVIII蛋白中的XTEN序列,紧靠对应于SEQ ID NO:65的氨基酸745的下游插入的第一XTEN序列,以及紧靠对应于SEQ ID NO:65的氨基酸2332(C末端)的下游插入的第二XTEN序列。在其它实施方案中,第一XTEN序列紧靠对应于SEQ ID NO:65的氨基酸18、26、40、1656或1720的下游插入,并且第二XTEN序列紧靠对应于SEQ ID NO:65的氨基酸403的下游插入。在其它实施方案中,第一XTEN序列紧靠对应于SEQ ID NO:65的氨基酸18、26或40的下游插入,并且第二XTEN序列紧靠对应于SEQ ID NO:65的氨基酸599的下游插入。在其它实施方案中,第一XTEN序列紧靠对应于SEQ ID NO:65的氨基酸1656的下游插入,并且第二XTEN序列紧靠对应于SEQ ID NO:65的氨基酸18、26、40、399、403、1725、1720、1900、1905或2332的下游插入。在某些实施方案中,第一XTEN序列紧靠对应于SEQ ID NO:65的氨基酸1900的下游插入,并且第二XTEN序列紧靠对应于SEQ ID NO:65的氨基酸18、26或40的下游插入。在一些实施方案中,第一XTEN序列紧靠对应于SEQ ID NO:65的氨基酸18、26或40的下游插入,并且第二XTEN序列紧靠对应于SEQ ID NO:65的氨基酸399的下游插入。在其它实施方案中,第一XTEN序列紧靠对应于SEQ ID NO:65的氨基酸1720的下游插入,并且第二XTEN序列紧靠对应于SEQ ID NO:65的氨基酸18、26或40的下游插入。在其它实施方案中,第一XTEN序列紧靠对应于SEQ ID NO:65的氨基酸1720的下游插入,并且第二XTEN序列紧靠对应于SEQ ID NO:65的氨基酸18的下游插入。在特定实施方案中,FVIII蛋白包含两个XTEN序列,紧靠对应于SEQ ID NO:65的氨基酸745的下游插入的第一XTEN序列,以及紧靠对应于SEQ ID NO:65的氨基酸2332的下游插入的第二XTEN序列,其中所述FVIII蛋白进一步具有从对应于SEQ ID NO:65的氨基酸745至对应于SEQ ID NO:65的氨基酸1685的缺失、在对应于SEQ ID NO:65的氨基酸1680处的突变或取代(例如Y1680F)、在对应于SEQ ID NO:65的氨基酸1648处的突变或取代(例如R1648A)、或在对应于SEQ ID NO:65的氨基酸1648(例如R1648A)和对应于SEQ ID NO:65的氨基酸1680(例如Y1680F)处的至少两个突变或取代。在特定实施方案中,FVIII蛋白包含两个XTEN序列,紧靠对应于SEQ ID NO:65的氨基酸1656的下游插入的第一XTEN,以及紧靠SEQ ID NO:65的氨基酸2332的下游插入的第二XTEN序列,其中FVIII蛋白进一步具有从对应于SEQ ID NO:65的氨基酸745至氨基酸1656的缺失。

[0358] 在某些实施方案中,FVIII蛋白包含三个XTEN序列,插入第一XTEN插入位点中的第一XTEN序列、插入第二XTEN序列中的第二XTEN序列以及插入第三XTEN插入位点中的第三XTEN序列。第一、第二或第三XTEN序列可相同或不同。第一、第二和第三插入位点可选自本文公开的任一插入位点的组。在一些实施方案中,包含三个XTEN序列的FVIII蛋白可进一步

包含突变或取代,例如对应于SEQ ID NO:65的氨基酸1648,例如R1648A。举例来说,第一、第二和第三XTEN插入位点的非限制性实例列于表12中。

[0359] 表12. 三个XTEN的示例性插入位点

	插入 1		插入 2		插入 3	
	插入位点	结构域	插入位点	结构域	插入位点	结构域
[0360]	26	A1	403	A2	1656	A3
	26	A1	403	A2	1720	A3
	26	A1	403	A2	1900	A3
	26	A1	1656	A3	1720	A3
	26	A1	1656	A3	1900	A3

	插入 1		插入 2		插入 3	
	插入位点	结构域	插入位点	结构域	插入位点	结构域
[0361]	26	A1	1720	A3	1900	A3
	403	A2	1656	A3	1720	A3
	403	A2	1656	A3	1900	A3
	403	A2	1720	A3	1900	A3
	1656	A3	1720	A3	1900	A3
	745	B	1900		2332	
	18	A1	745	B	2332	CT
	26	A1	745	B	2332	CT
	40	A1	745	B	2332	CT
	18	A1	745	B	2332	CT
	40	A1	745	B	2332	CT
	403	A2	745	B	2332	CT
	399	A2	745	B	2332	CT
	1725	A3	745	B	2332	CT
	1720	A3	745	B	2332	CT
	1711	A3	745	B	2332	CT
	1900	A3	745	B	2332	CT
	1905	A3	745	B	2332	CT
	1910	A3	745	B	2332	CT

[0362] 在一些实施方案中,FVIII蛋白包含三个XTEN序列,紧靠对应于SEQ ID NO:65的氨基酸26的下游插入的第一XTEN序列,插入在对应于SEQ ID NO:65的氨基酸403的下游的第二XTEN序列,以及插入在对应于SEQ ID NO:65的氨基酸1656、1720或1900的下游的第三XTEN序列。在其它实施方案中,第一XTEN序列紧靠对应于SEQ ID NO:65的氨基酸26的下游插入,第二XTEN序列插入在对应于SEQ ID NO:65的氨基酸1656的下游,并且第三XTEN序列插入在对应于SEQ ID NO:65的氨基酸1720或1900的下游。在其它实施方案中,第一XTEN序列紧靠对应于SEQ ID NO:65的氨基酸26的下游插入,第二XTEN序列插入在对应于SEQ ID NO:65的氨基酸1720的下游,并且第三XTEN序列插入在对应于SEQ ID NO:65的氨基酸1900

的下游。在其它实施方案中,第一XTEN序列紧靠对应于SEQ ID NO:65的氨基酸403的下游插入,第二XTEN序列插入在对应于SEQ ID NO:65的氨基酸1656的下游,并且第三XTEN序列插入在对应于SEQ ID NO:65的氨基酸1720或1900的下游。在其它实施方案中,第一XTEN序列紧靠对应于SEQ ID NO:65的氨基酸403或1656的下游插入,第二XTEN序列插入在对应于SEQ ID NO:65的氨基酸1720的下游,并且第三XTEN序列插入在对应于SEQ ID NO:65的氨基酸1900的下游。在其它实施方案中,第一XTEN序列紧靠对应于SEQ ID NO:65的氨基酸18、26、40、399、403、1711、1720、1725、1900、1905或1910的下游插入,第二XTEN序列插入在对应于SEQ ID NO:65的氨基酸745的下游,并且第三XTEN序列插入在对应于SEQ ID NO:65的氨基酸2332的下游。

[0363] 在其它实施方案中,本发明中的FVIII蛋白包含四个XTEN序列,插入第一插入位点中的第一XTEN序列、插入第二插入位点中的第二XTEN序列、插入第三插入位点中的第三XTEN序列以及插入第四插入位点中的第四XTEN序列。第一、第二、第三和第四XTEN序列可相同、不同或呈其组合形式。在一些实施方案中,包含四个XTEN序列的FVIII蛋白可进一步包含突变或取代,例如对应于SEQ ID NO:65的氨基酸1648,例如R1648A。第一、第二、第三和第四XTEN插入位点的非限制性实例列于表13中。

[0364] 表13. 四个XTEN的示例性插入位点

[0365]

插入 1		插入 2		插入 3		插入 4	
插入位点	结构域	插入位点	结构域	插入位点	结构域	插入位点	结构域
26	A1	403	A2	1656	a3	1720	A3
26	A1	403	A2	1656	a3	1900	A3
26	A1	403	A2	1720	A3	1900	A3
26	A1	1656	a3	1720	A3	1900	A3
403	A2	1656	a3	1720	A3	1900	A3
0040	A1	0403	A2	745	B	2332	CT
0040	A1	0403	A2	745	B	2332	CT
0018	A1	0409	A2	745	B	2332	CT
0040	A1	0409	A2	745	B	2332	CT
0040	A1	0409	A2	745	B	2332	CT
0018	A1	0409	A2	745	B	2332	CT

[0366]

插入 1		插入 2		插入 3		插入 4	
插入位点	结构域	插入位点	结构域	插入位点	结构域	插入位点	结构域
0040	A1	1720	A3	745	B	2332	CT
0026	A1	1720	A3	745	B	2332	CT
0018	A1	1720	A3	745	B	2332	CT
0018	A1	1720	A3	745	B	2332	CT
0018	A1	1720	A3	745	B	2332	CT
0026	A1	1720	A3	745	B	2332	CT
0018	A1	1720	A3	745	B	2332	CT
0018	A1	1900	A3	745	B	2332	CT
0018	A1	1900	A3	745	B	2332	CT
0026	A1	1900	A3	745	B	2332	CT
0040	A1	1900	A3	745	B	2332	CT
0040	A1	1905	A3	745	B	2332	CT
0018	A1	1905	A3	745	B	2332	CT
0040	A1	1905	A3	745	B	2332	CT
0026	A1	1905	A3	745	B	2332	CT
0018	A1	1905	A3	745	B	2332	CT
0018	A1	1905	A3	745	B	2332	CT
0018	A1	1910	A3	745	B	2332	CT
0018	A1	1910	A3	745	B	2332	CT
0040	A1	1910	A3	745	B	2332	CT
0026	A1	1910	A3	745	B	2332	CT
0018	A1	1910	A3	745	B	2332	CT
0026	A1	1910	A3	745	B	2332	CT
0040	A1	1910	A3	745	B	2332	CT
0018	A1	1910	A3	745	B	2332	CT
0409	A2	1720	A3	745	B	2332	CT
0403	A2	1720	A3	745	B	2332	CT
0409	A2	1720	A3	745	B	2332	CT
0403	A2	1720	A3	745	B	2332	CT
0403	A2	1720	A3	745	B	2332	CT
0403	A2	1900	A3	745	B	2332	CT
0403	A2	1900	A3	745	B	2332	CT
0409	A2	1900	A3	745	B	2332	CT
0403	A2	1900	A3	745	B	2332	CT
0403	A2	1900	A3	745	B	2332	CT
0409	A2	1900	A3	745	B	2332	CT
0409	A2	1905	A3	745	B	2332	CT
0403	A2	1905	A3	745	B	2332	CT

插入 1		插入 2		插入 3		插入 4	
插入位点	结构域	插入位点	结构域	插入位点	结构域	插入位点	结构域
0403	A2	1905	A3	745	B	2332	CT
0403	A2	1905	A3	745	B	2332	CT
0409	A2	1905	A3	745	B	2332	CT
0403	A2	1905	A3	745	B	2332	CT
0409	A2	1910	A3	745	B	2332	CT
0403	A2	1910	A3	745	B	2332	CT
0403	A2	1910	A3	745	B	2332	CT
0403	A2	1910	A3	745	B	2332	CT
0403	A2	1910	A3	745	B	2332	CT
1720	A3	1900	A3	745	B	2332	CT
1720	A3	1905	A3	745	B	2332	CT
1720	A3	1910	A3	745	B	2332	CT
1720	A3	1910	A3	745	B	2332	CT
0403	A2	1656	a3	1720	A3	2332	CT
0403	A2	1656	a3	1900	A3	2332	CT
0403	A2	1720	A3	1900	A3	2332	CT
1656	a3	1720	A3	1900	A3	2332	CT
0018	A1	0403	A2	1656	a3	2332	CT
0018	A1	0403	A2	1720	A3	2332	CT
0018	A1	0403	A2	1900	A3	2332	CT
0018	A1	1656	a3	1720	A3	2332	CT
0018	A1	1656	a3	1900	A3	2332	CT
0018	A1	1720	A3	1900	A3	2332	CT
0018	A1	0403	A2	0745	B	2332	CT
0018	A1	0745	B	1720	A3	2332	CT
0018	A1	0745	B	1900	A3	2332	CT
0403	A2	0745	B	1720	A3	2332	CT
0403	A2	0745	B	1900	A3	2332	CT
0745	B	1720	A3	1900	A3	2332	CT
0188	A1	1900	A3	0745	B	2332	CT
0599		1900	A3	0745	B	2332	CT
2068		1900	A3	0745	B	2332	CT
2171		1900	A3	0745	B	2332	CT
2227		1900	A3	0745	B	2332	CT
2277		1900	A3	0745	B	2332	CT

[0367]

[0368] 在一些实施方案中, FVIII蛋白包含五个XTEN序列, 插入第一插入位点中的第一XTEN序列、插入第二插入位点中的第二XTEN序列、插入第三XTEN插入位点中的第三XTEN序列、插入第四XTEN插入位点中的第四XTEN序列以及插入第五XTEN插入位点中的第五XTEN序

列。第一、第二、第三、第四或第五XTEN序列可相同、不同或呈其组合形式。第一、第二、第三、第四和第五插入位点的非限制性实例列于表14中。

[0369] 表14. 五个XTEN的示例性插入位点

	XTEN 插入 1	XTEN 插入 2	XTEN 插入 3	XTEN 插入 4	XTEN 插入 5
	0403	1656	1720	1900	2332
	0018	0403	1656	1720	2332
	0018	0403	1656	1900	2332
[0370]	0018	0403	1720	1900	2332
	0018	1656	1720	1900	2332
	0018	0403	0745	1720	2332
	0018	0403	0745	1900	2332
	0018	0745	1720	1900	2332
	0403	0745	1720	1900	2332

[0371] 在某些实施方案中，FVIII蛋白包含六个XTEN序列，插入第一XTEN插入位点中的第一XTEN序列、插入第二XTEN插入位点中的第二XTEN序列、插入第三XTEN插入位点中的第三XTEN序列、插入第四XTEN插入位点中的第四XTEN序列、插入第五XTEN插入位点中的第五XTEN序列以及插入第六XTEN插入位点中的第六XTEN序列。第一、第二、第三、第四、第五或第六XTEN序列可相同、不同或呈其组合形式。六个XTEN插入位点的实例包括但不限于表15中所列的插入位点。

[0372] 表15. 六个XTEN的示例性XTEN插入位点

	XTEN 插入 1	XTEN 插入 2	XTEN 插入 3	XTEN 插入 4	XTEN 插入 5	XTEN 插入 5
	0018	0403	1656	1720	1900	2332
[0373]	0018	0403	0745	1720	1900	2332

[0374] 在一特定实例中，第一XTEN插入在对应于SEQ ID NO:65的氨基酸26与27之间，并且第二XTEN插入在对应于SEQ ID NO:65(全长成熟FVIII)的氨基酸1720与1721之间。在另一实例中，第一XTEN插入在对应于SEQ ID NO:65的氨基酸403与404之间，并且第二XTEN插入在对应于SEQ ID NO:65的氨基酸1720与1721之间。在一些实例中，第一XTEN插入在对应于SEQ ID NO:65的氨基酸1656与1657之间，并且第二XTEN插入在对应于SEQ ID NO:65的氨基酸1720与1721之间。在其它实例中，第一XTEN插入在对应于SEQ ID NO:65的氨基酸26与27之间，第二XTEN插入在对应于SEQ ID NO:65的氨基酸1656与1657之间，并且第三XTEN插入在对应于SEQ ID NO:65的氨基酸1720与1721之间。在其它实施方案中，第一XTEN插入在对应于SEQ ID NO:65的氨基酸403与404之间，第二XTEN插入在对应于SEQ ID NO:65的氨基酸1656与1657之间，并且第三XTEN插入在对应于SEQ ID NO:65的氨基酸1720与1721之间。在其它实施方案中，第一XTEN插入在对应于SEQ ID NO:65的氨基酸403与404之间，第二XTEN插入在对应于SEQ ID NO:65的氨基酸1656与1657之间，并且第三XTEN插入在对应于SEQ ID NO:65的氨基酸1720与1721之间。在某些实施方案中，第一XTEN插入在对应于SEQ



ID NO:65的氨基酸26与27之间,第二XTEN插入在对应于SEQ ID NO:65的氨基酸1720与1721之间,并且第三XTEN插入在对应于SEQ ID NO:65的氨基酸1900与1901之间。在一些实施方案中,第一XTEN插入在对应于SEQ ID NO:65的氨基酸26与27之间,第二XTEN插入在对应于SEQ ID NO:65的氨基酸1656与1657之间,第三XTEN插入在对应于SEQ ID NO:65的氨基酸1720与1721之间,并且第四XTEN插入在对应于SEQ ID NO:65的1900与1901之间。

[0375] 在特定实施方案中,XTEN序列插入在全长因子VIII的氨基酸745与746之间或插入B结构域缺失的因子VIII的相应插入位点。

[0376] 在一些实施方案中,本发明的嵌合蛋白包含两个多肽序列,第一多肽序列包含与选自FVIII-161 (SEQ ID NO:69)、FVIII-169 (SEQ ID NO:70)、FVIII-170 (SEQ ID NO:71)、FVIII-173 (SEQ ID NO:72)、FVIII-195 (SEQ ID NO:73)、FVIII-196 (SEQ ID NO:74)、FVIII-199 (SEQ ID NO:75)、FVIII-201 (SEQ ID NO:76)、FVIII-203 (SEQ ID NO:77)、FVIII-204 (SEQ ID NO:78)、FVIII-205 (SEQ ID NO:79)、FVIII-266 (SEQ ID NO:80)、FVIII-267 (SEQ ID NO:81)、FVIII-268 (SEQ ID NO:82)、FVIII-269 (SEQ ID NO:83)、FVIII-271 (SEQ ID NO:84)或FVIII-272 (SEQ ID NO:85)的序列至少约80%、90%、95%或100%同一的氨基酸序列,并且第二多肽序列包含与选自VWF031 (SEQ ID NO:86)、VWF034 (SEQ ID NO:87)或VWF-036的序列至少约80%、90%、95%或100%同一的氨基酸序列。

[0377] II.D. Ig恒定区或其一部分

[0378] 本发明的嵌合蛋白也包括两个Ig恒定区或其一部分,通过任选接头融合于FVIII蛋白的第一Ig恒定区或其一部分,以及通过具有小于288个氨基酸的XTEN序列融合于VWF蛋白的第二Ig恒定区或其一部分。Ig恒定区或其一部分可与XTEN序列和VWF蛋白组合改进嵌合蛋白的药物动力学或药理学性质。在某些实施方案中,Ig恒定区或其一部分延长融合于Ig恒定区或其一部分的分子的半衰期。

[0379] Ig恒定区包含表示为CH(恒定重)结构域(CH1、CH2等)的结构域。视同种型(即IgG、IgM、IgA、IgD或IgE)而定,恒定区可包含三个或四个CH结构域。一些同种型(例如IgG)恒定区也含有铰链区。参见Janeway等2001, Immunobiology, Garland Publishing, N.Y., N.Y.

[0380] 用于产生本发明的嵌合蛋白的Ig恒定区或其一部分可从许多不同来源获得。在一些实施方案中,Ig恒定区或其一部分源于人Ig。然而,应了解Ig恒定区或其一部分可源于另一哺乳动物物种的Ig,所述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,Ig恒定区或其一部分可源于任何Ig类别(包括IgM、IgG、IgD、IgA和IgE)和任何Ig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。

[0381] 多种Ig恒定区基因序列(例如人恒定区基因序列)可以公开可得的寄存物形式获得。可选择具有特定效应物功能(或缺乏特定效应物功能)或具有用以降低免疫原性的特定修饰的恒定区结构域序列。许多抗体和抗体编码基因序列已被公开,并且可使用本领域认可的技术从这些序列获得适合Ig恒定区序列(例如铰链、CH2和/或CH3序列或其部分)。使用任何前述方法获得的遗传物质可接着加以改变或合成以获得本发明的多肽。应进一步了解本发明的范围涵盖恒定区DNA序列的等位基因、变体和突变。

[0382] Ig恒定区或其一部分的序列可例如使用被选择用以扩增目标结构域的聚合酶链反应和引物来克隆。为从抗体克隆Ig恒定区或其一部分的序列,可从杂交瘤、脾或淋巴细胞

分离mRNA,逆转录成DNA,并且通过PCR扩增抗体基因。PCR扩增方法详细描述于美国专利号4,683,195;4,683,202;4,800,159;4,965,188中;以及例如“PCR Protocols:A Guide to Methods and Applications”Innis等编,Academic Press,San Diego,CA(1990);Ho等1989.Gene 77:51;Horton等1993.Methods Enzymol.217:270)中。PCR可通过共有恒定区引物或通过基于公开的重链和轻链DNA和氨基酸序列的更特异性引物来引发。如上所讨论,PCR也可用于分离编码抗体轻链和重链的DNA克隆。在这个情况下,可通过共有引物或较大同源性探针(诸如小鼠恒定区探针)来筛选文库。适于扩增抗体基因的众多引物组在本领域中是已知的(例如基于纯化抗体的N末端序列的5'引物(Benhar和Pastan.1994.Protein Engineering 7:1509);cDNA末端的快速扩增物(Ruberti,F.等1994.J.Immunol.Methods 173:33);抗体前导序列(Larrick等1989Biochem.Biophys.Res.Commun.160:1250)。对抗体序列的克隆进一步描述于Newman等,1995年1月25日提交的美国专利号5,658,570中,所述专利以引用的方式并入本文。

[0383] 本文所用的Ig恒定区可包括所有结构域和铰链区或其部分。在一个实施方案中,Ig恒定区或其一部分包含CH2结构域、CH3结构域和铰链区,即Fc区或FcRn结合配偶体。

[0384] 如本文所用,术语“Fc区”定义为多肽的对应于天然Ig的Fc区的部分,即如通过它的两个重链的相应Fc结构域的二聚缔合所形成。天然Fc区与另一Fc区形成同二聚体。相反,如本文所用的术语“遗传融合Fc区”或“单链Fc区”(scFc区)是指合成二聚Fc区,其包含在单一多肽链内遗传连接的Fc结构域(即在单一连续遗传序列中编码)。

[0385] 在一个实施方案中,“Fc区”是指单一Ig重链的一部分,其在恰好在木瓜蛋白酶(papain)裂解位点(即IgG中的残基216,将重链恒定区的第一残基视为114)上游的铰链区中开始,并且在抗体的C末端结束。因此,完全Fc结构域包含至少铰链结构域、CH2结构域和CH3结构域。

[0386] 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区的嵌合蛋白对嵌合蛋白赋予若干合乎需要的性质,包括稳定性增加、血清半衰期增加(参见Capon等,1989,Nature 337:525)以及结合Fc受体,诸如新生儿Fc受体(FcRn)(美国专利号6,086,875、6,485,726、6,030,613;WO 03/077834;US2003-0235536A1),所述文献和专利以引用的方式整体并入本文。

[0387] Ig恒定区或其一部分可为FcRn结合配偶体。FcRn在成人上皮组织中具有活性,并且在肠腔、肺气道、鼻表面、阴道表面、结肠和直肠表面中表达(美国专利号6,485,726)。FcRn结合配偶体是Ig的结合FcRn的部分。

[0388] 已从包括人的若干哺乳动物物种分离FcRn受体。已知人FcRn、猴FcRn、大鼠FcRn和小鼠FcRn的序列(Story等1994,J.Exp.Med.180:2377)。FcRn受体在相对较低pH下结合IgG(但不结合其它Ig类别,诸如IgA、IgM、IgD和IgE),以细胞腔至浆膜方向跨细胞主动转运IgG,接着在见于间隙液中的相对较高pH下释放IgG。它在成人上皮组织(美国专利号6,485,726、6,030,613、6,086,875;WO 03/077834;US2003-0235536A1),包括肺和肠上皮(Israel等1997,Immunology 92:69)、肾近端管状上皮(Kobayashi等2002,Am.J.Physiol.Renal Physiol.282:F358)以及鼻上皮;阴道表面;和胆系表面中表达。

[0389] 适用于本发明中的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FcRn受体的完全结合区的其它片段。已基于X射线结晶学描述

IgG的Fc部分的结合FcRn受体的区域(Burmeister等1994,Nature 372:379)。Fc与FcRn的主要接触区域接近CH2结构域与CH3结构域的接合点。Fc-FcRn接触都在单一Ig重链内。FcRn结合配偶体包括完整IgG、IgG的Fc片段和IgG的包括FcRn的完全结合区域的其它片段。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以及CH3结构域的氨基酸残基385-387、428和433-436。对Ig或Ig片段的氨基酸编号所作的提及都基于Kabat等1991,Sequences of Proteins of Immunological Interest,U.S.Department of Public Health,Bethesda,Md。

[0390] 结合于FcRn的Fc区或FcRn结合配偶体可通过FcRn有效地跨越上皮屏障穿梭,由此提供一种全身性施用所需治疗分子的非侵入性手段。另外,包含Fc区或FcRn结合配偶体的融合蛋白由表达FcRn的细胞胞饮。但替代被标记以进行降解,这些融合蛋白再次再循环进入循环中,由此增加这些蛋白质的体内半衰期。在某些实施方案中,Ig恒定区的部分是Fc区或FcRn结合配偶体,其通常通过二硫键和其它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更高级多聚体。

[0391] 两个FcRn受体可结合单一Fc分子。结晶学数据表明各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)连接于生物活性分子会提供一种口服、经颊、舌下、经直肠、经阴道、以经鼻或通过肺途径施用的气雾剂形式、或通过眼途径递送所述生物活性分子的手段。在另一实施方案中,可侵入性施用嵌合蛋白,例如皮下、静脉内。

[0392] FcRn结合配偶体区域是可由FcRn受体特异性结合,随后由Fc区的FcRn受体主动转运的分子或其一部分。特异性结合是指两个分子形成在生理条件下相对稳定的复合物。特异性结合的特征在于亲和力较高而能力较低至中等,如与通常亲和力较低而能力中等至较高的非特异性结合相区分。通常,当亲和常数 $K_A$ 高于 $10^6 M^{-1}$ 或高于 $10^8 M^{-1}$ 时,结合被视为是特异性的。必要时,可通过改变结合条件来降低非特异性结合而不实质上影响特异性结合。诸如分子浓度、溶液的离子强度、温度、允许结合时间、阻断剂(例如血清白蛋白、乳酪蛋白)浓度等的适当结合条件可由熟练技术人员使用常规技术加以优化。

[0393] 在某些实施方案中,本发明的嵌合蛋白包含一个或多个截短Fc区,尽管如此,所述Fc区仍然足以对Fc区赋予Fc受体(FcR)结合性质。举例来说,Fc区的结合FcRn的部分(即FcRn结合部分)包含IgG1的约氨基酸282-438(EU编号),(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以及CH3结构域的氨基酸残基385-387、428和433-436。因此,本发明的Fc区可包含FcRn结合部分或由FcRn结合部分组成。FcRn结合部分可源于包括IgG1、IgG2、IgG3和IgG4的任何同种型的重链。在一个实施方案中,使用来自具有人同种型IgG1的抗体的FcRn结合部分。在另一实施方案中,使用来自具有人同种型IgG4的抗体的FcRn结合部分。

[0394] 在另一实施方案中,“Fc区”包括Fc结构域的氨基酸序列或源于Fc结构域的氨基酸序列。在某些实施方案中,Fc区包含以下中的至少一个:铰链(例如上、中和/或下铰链区)结构域(抗体Fc区的约氨基酸216-230,根据EU编号)、CH2结构域(抗体Fc区的约氨基酸231-340,根据EU编号)、CH3结构域(抗体Fc区的约氨基酸341-438,根据EU编号)、CH4结构域、或其变体、部分或片段。在其它实施方案中,Fc区包含完全Fc结构域(即铰链结构域、CH2结构域和CH3结构域)。在一些实施方案中,Fc区包含以下、基本上由以下组成或由以下组成:融

合于CH3结构域(或其一部分)的铰链结构域(或其一部分)、融合于CH2结构域(或其一部分)的铰链结构域(或其一部分)、融合于CH3结构域(或其一部分)的CH2结构域(或其一部分)、融合于铰链结构域(或其一部分)与CH3结构域(或其一部分)两者的CH2结构域(或其一部分)。在其它实施方案中,Fc区缺乏CH2结构域的至少一部分(例如CH2结构域的全部或一部分)。在特定实施方案中,Fc区包含以下或由以下组成:对应于EU编号221至447的氨基酸。

[0395] 在本文中表示为F、F1或F2的Fc区可从许多不同来源获得。在一个实施方案中,多肽的Fc区源于人Ig。然而,应了解Fc区可源于另一哺乳动物物种的Ig,所述物种包括例如啮齿动物(例如小鼠、大鼠、兔或豚鼠)或非人灵长类动物(例如黑猩猩、猕猴)物种。此外,Fc结构域或其部分的多肽可源于任何Ig类别(包括IgM、IgG、IgD、IgA和IgE)和任何Ig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。

[0396] 在某些实施方案中,Fc变体赋予由包含所述野生型Fc结构域的Fc区赋予的至少一种效应物功能的变化(例如Fc区结合Fc受体(例如Fc $\gamma$ RI、Fc $\gamma$ RII或Fc $\gamma$ RIII)或补体蛋白(例如C1q)、或触发抗体依赖性细胞毒性(ADCC)、吞噬作用或补体依赖性细胞毒性(CDCC)的能力改进或降低)。在其它实施方案中,Fc变体提供工程改造的半胱氨酸残基。

[0397] 本发明的Fc区可采用本领域认可的已知会赋予效应物功能和/或FcR或FcRn结合变化(例如增强或降低)的Fc变体。具体来说,本发明的结合分子可包括例如在以下中公开的一个或多个氨基酸位置处的变化(例如取代):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,其各自以引用的方式并入本文。在一个实施方案中,可在一个或多个公开的氨基酸位置处进行特定变化(例如特定取代本领域中公开的一个或多个氨基酸)。在另一实施方案中,可在一个或多个公开的氨基酸位置处进行不同变化(例如对本领域中公开的一个或多个氨基酸位置的不同取代)。

[0398] IgG的Fc区或FcRn结合配偶体可根据充分认可的程序(诸如定点诱变等)加以修饰以产生将由FcRn结合的修饰的IgG或其Fc片段或部分。所述修饰包括保持或甚至增强与FcRn的结合的远离FcRn接触位点的修饰以及在接触位点内的修饰。举例来说,人IgG1 Fc(Fc $\gamma$ 1)中的以下单一氨基酸残基可被取代而不显著损失Fc对FcRn的结合亲和力: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表示野生型脯氨酸在位置编号238处被丙氨酸取代。举例来说,一特定实施方案并有N297A突变,从而移除高度保守的N-糖基化位点。除丙氨酸之外,其它氨基酸也可在以上指定的位置处取代野生型氨基酸。突变可单一引入Fc中,从而产生超过一百个不同于天然Fc的Fc区。另外,两个、三个或更多个这些个别突变的组合可一起引入,从而再产生数百个Fc区。此外,本发明的构建体的一个Fc区可被突变,并且所述构建体的另一Fc区完全不突变,或它们两者均可被突变,但突变不同。

[0399] 某些以上突变可对Fc区或FcRn结合配偶体赋予新功能性。举例来说,一个实施例并有N297A,从而移除高度保守的N-糖基化位点。这个突变的作用在于降低免疫原性,由此增强Fc区的循环半衰期,以及在不损害对FcRn的亲合力下致使Fc区不能结合Fc $\gamma$ RI、Fc $\gamma$ RIIA、Fc $\gamma$ RIIB和Fc $\gamma$ RIIIA (Routledge等1995, Transplantation60:847; Friend等1999, Transplantation 68:1632; Shields等1995, J. Biol. Chem. 276:6591)。作为由上述突变而产生的新功能性的另一实例,对FcRn的亲合力在一些情况下可增加超过野生型对FcRn的亲合力。这个亲合力增加可反映“缔合”速率增加、“解离”速率降低、或“缔合”速率增加与“解离”速率降低两者。据信会赋予对FcRn的亲合力增加的突变的实例包括但不限于T256A、T307A、E380A和N434A (Shields等2001, J. Biol. Chem. 276:6591)。

[0400] 另外,至少三种人Fc $\gamma$ 受体似乎识别IgG上在下铰链区内的结合位点,通常是氨基酸234-237。因此,新功能性和潜在免疫原性降低的另一实例可由这个区域的突变而产生,如例如通过将人IgG1的氨基酸233-236“ELLG”置换成来自IgG2的相应序列“PVA”(有一个氨基酸缺失)。已显示当已引入所述突变时,介导各种效应物功能的Fc $\gamma$ RI、Fc $\gamma$ RII和Fc $\gamma$ RIII将不结合IgG1。Ward和Ghetie 1995, Therapeutic Immunology 2:77以及Armour等1999, Eur. J. Immunol. 29:2613。

[0401] 在一个实施方案中,Ig恒定区或其一部分(例如Fc区)是包括序列PKNSSMISNTP (SEQ ID NO:89或美国专利号5,739,277的SEQ ID NO:3),并且任选进一步包括选自HQSLGTQ (SEQ ID NO:90)、HQNLSDGK (SEQ ID NO:91)、HQNISDGK (SEQ ID NO:92)或VISSHLGQ (SEQ ID NO:93)(或美国专利号5,739,277的分别SEQ ID NO:11、1、2和31)的序列的多肽。

[0402] 在另一实施方案中,免疫球蛋白恒定区或其一部分包含在铰链区或其一部分中的与另一免疫球蛋白恒定区或其一部分形成一个或多个二硫键的氨基酸序列。由免疫球蛋白恒定区或其一部分达成的二硫键将包含FVIII的第一多肽和包含VWF片段的第二多肽放置在一起以使内源性VWF不置换VWF片段,并且不结合FVIII。因此,第一免疫球蛋白恒定区或其一部分与第二免疫球蛋白恒定区或其一部分之间的二硫键会防止内源性VWF与FVIII蛋白之间的相互作用。对VWF与FVIII蛋白之间的相互作用的这个抑制允许嵌合蛋白的半衰期超过两倍限制。铰链区或其一部分可进一步连接于以下一个或多个结构域:CH1、CH2、CH3、其片段及其任何组合。在特定实施方案中,免疫球蛋白恒定区或其一部分是铰链区和CH2。

[0403] 在某些实施方案中,Ig恒定区或其一部分被半糖基化。举例来说,包含两个Fc区或FcRn结合配偶体的嵌合蛋白可含有第一糖基化Fc区(例如糖基化CH2区)或FcRn结合配偶体和第二无糖基化Fc区(例如无糖基化CH2区)或FcRn结合配偶体。在一个实施方案中,接头可

插入在糖基化Fc区与无糖基化Fc区之间。在另一实施方案中，Fc区或FcRn结合配偶体被完全糖基化，即所有Fc区都被糖基化。在其它实施方案中，Fc区可为无糖基化的，即无Fc部分被糖基化。

[0404] 在某些实施方案中，本发明的嵌合蛋白包含对Ig恒定区或其一部分的氨基酸取代（例如Fc变体），其改变Ig恒定区的抗原非依赖性效应物功能，特别是蛋白质的循环半衰期。

[0405] 当相较于缺乏这些取代的蛋白质时，所述蛋白质展现与FcRn的结合增加或降低，并且因此在血清中的半衰期分别增加或降低。预期对FcRn的亲合力改进的Fc变体会具有较长血清半衰期，并且所述分子在治疗哺乳动物的方法（其中施用的多肽的长久半衰期例如为治疗慢性疾病或病症所需）中具有适用应用（参见例如美国专利7,348,004、7,404,956和7,862,820）。相反，预期FcRn结合亲和力降低的Fc变体会具有较短半衰期，并且所述分子也适用于例如在循环时间缩短可为有利时向哺乳动物施用，例如对于体内诊断成像或在起始多肽在持续延长时期存在于循环中时具有毒性副作用的情况下。FcRn结合亲和力降低的Fc变体跨越胎盘的可能性也较小，并且因此也适用于治疗妊娠妇女的疾病或病症。此外，其中可能需要FcRn结合亲和力降低的其它应用包括其中需要定位于脑、肾和/或肝的那些应用。在一个示范性实施方案中，本发明的嵌合蛋白展现从血管结构跨越肾小球的上皮的转运降低。在另一实施方案中，本发明的嵌合蛋白展现从脑跨越血脑屏障（BBB）进入血管间隙中的转运降低。在一个实施方案中，FcRn结合改变的蛋白质包含至少一个在Ig恒定区的“FcRn结合环”内具有一个或多个氨基酸取代的Fc区或FcRn结合配偶体（例如一或两个Fc区或FcRn结合配偶体）。FcRn结合环包含野生型全长Fc区的氨基酸残基280-299（根据EU编号）。在其它实施方案中，本发明的嵌合蛋白中具有改变的FcRn结合亲和力的Ig恒定区或其一部分包括至少一个在15 Å FcRn“接触区”内具有一个或多个氨基酸取代的Fc区或FcRn结合配偶体。如本文所用，术语15 Å FcRn“接触区”包括在野生型全长Fc部分的以下位置处的残基：243-261、275-280、282-293、302-319、336-348、367、369、372-389、391、393、408、424、425-440（EU编号）。在其它实施方案中，本发明的具有改变的FcRn结合亲和力的Ig恒定区或其一部分包括至少一个在对应于任一以下EU位置的氨基酸位置处具有一个或多个氨基酸取代的Fc区或FcRn结合配偶体：256、277-281、283-288、303-309、313、338、342、376、381、384、385、387、434（例如N434A或N434K）和438。改变FcRn结合活性的示范性氨基酸取代公开于国际PCT公布号W005/047327中，所述公布以引用的方式并入本文。

[0406] 本发明中使用的Fc区或FcRn结合配偶体也可包含本领域认可的改变嵌合蛋白的糖基化的氨基酸取代。举例来说，嵌合蛋白的连接于VWF片段或FVIII蛋白的Fc区或FcRn结合配偶体可包括具有导致糖基化（例如N连接或O连接的糖基化）降低的突变的Fc区，或可包括糖形式改变（例如低海藻糖（fucose）或无海藻糖聚糖）的野生型Fc部分。

[0407] 在一个实施方案中，本发明的未加工嵌合蛋白可包含遗传融合Fc区（即scFc区），所述Fc区的组成Ig恒定区或其一部分中的两个或更多个独立地选自本文所述的Ig恒定区或其一部分。在一个实施方案中，二聚Fc区的Fc区是相同的。在另一实施方案中，至少两个Fc区是不同的。举例来说，本发明的蛋白质的Fc区或FcRn结合配偶体包含相同数目的氨基酸残基，或它们可在长度方面相差一个或多个氨基酸残基（例如约5个氨基酸残基（例如1、2、3、4或5个氨基酸残基）、约10个残基、约15个残基、约20个残基、约30个残基、约40个残基

或约50个残基)。在其它实施方案中,本发明的蛋白质的Fc区或FcRn结合配偶体可在序列方面在一个或多个氨基酸位置处不同。举例来说,至少两个Fc区或FcRn结合配偶体可在约5个氨基酸位置(例如1、2、3、4或5个氨基酸位置)、约10个位置、约15个位置、约20个位置、约30个位置、约40个位置、或约50个位置处不同。

#### [0408] II.E. 接头

[0409] 本发明的嵌合蛋白进一步包含一个或多个接头。一种接头类型是可裂解接头,其在体内向受试者施用,例如在凝血部位处施用时可由各种蛋白酶裂解。在一个实施方案中,可裂解接头允许在凝血级联部位处从XTEN序列,由此从嵌合蛋白裂解部分(例如VWF蛋白),由此使活化FVIII(FVIIIa)具有它的FVIIIa活性。另一接头类型是可加工接头,其含有细胞内裂解位点,并且因此可由宿主细胞中的细胞内加工酶裂解,从而允许适当表达多肽和形成嵌合蛋白。

[0410] 一个或多个接头可存在于嵌合蛋白中的任何两个蛋白质之间。在一个实施方案中,嵌合蛋白包含第一多肽,所述第一多肽包含(i)FVIII蛋白和(ii)第一Ig恒定区或其一部分;和第二多肽,所述第二多肽包含(iii)VWF蛋白,(iv)接头(例如可裂解接头),(v)XTEN序列,和(vi)第二Ig恒定区或其一部分。在另一实施方案中,嵌合蛋白包含第一多肽,所述第一多肽包含(i)FVIII蛋白和(ii)第一Ig恒定区或其一部分;和第二多肽,所述第二多肽包含(iii)VWF蛋白,(iv)XTEN序列,(v)接头(例如可裂解接头),和(vi)第二Ig恒定区或其一部分。在其它实施方案中,嵌合蛋白包含第一多肽,所述第一多肽包含(i)FVIII蛋白和(ii)第一Ig恒定区或其一部分;和第二多肽,所述第二多肽包含(iii)VWF蛋白,(iv)第一接头(例如可裂解接头),(v)XTEN序列,(vi)第二接头(例如可裂解接头),和(vii)第二Ig恒定区或其一部分。在一些实施方案中,第一多肽进一步在FVIII蛋白与第一Ig恒定区之间包含接头,例如可裂解接头。

[0411] 在某些实施方案中,嵌合蛋白包含单链,其包含(i)FVIII蛋白,(ii)第一Ig恒定区或其一部分,(iii)接头(例如可加工接头),(iv)VWF蛋白,(v)XTEN序列,和(vi)第二Ig恒定区或其一部分。在其它实施方案中,嵌合蛋白包含单链,其包含(i)FVIII蛋白,(ii)第一Ig恒定区或其一部分,(iii)第一接头(例如可加工接头),(iv)VWF蛋白,(v)第二接头(例如可裂解接头),(vi)XTEN序列,和(vii)第二Ig恒定区或其一部分。在嵌合蛋白在宿主细胞中表达之后,可加工接头可被加工;因此在宿主细胞中产生的嵌合蛋白可呈包含两个或三个多肽链的最终形式。

[0412] 适用于本发明中的接头可包括任何有机分子。在一个实施方案中,接头包括聚合物,例如聚乙二醇(PEG)或羟乙基淀粉(HES)。在另一实施方案中,接头包括氨基酸序列。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。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。在一个实施方案中,接头包括XTEN序列。XTEN的其它实例可根据本发明加以使用,并且公开于美国专利公布号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。在另一实施方案中,接头是PAS序列。

[0413] 在一个实施方案中,接头是聚合物,例如聚乙二醇(PEG)或羟乙基淀粉(HES)。在另一实施方案中,接头是氨基酸序列。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。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。

[0414] 接头的实例在本领域中是熟知的。在一个实施方案中,接头包含序列 $G_n$ 。接头可包含序列 $(GA)_n$ 。接头可包含序列 $(GGS)_n$ 。在其它实施方案中,接头包含 $(GGGS)_n$ (SEQ ID NO:101)。在其它实施方案中,接头包含序列 $(GGS)_n$   $(GGGS)_n$ (SEQ ID NO:95)。在这些情况下,n可为整数1-100。在其它情况下,n可为整数1-20,即1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20。接头的实例包括但不限于GGG、SGGSGGS(SEQ ID NO:96)、GGSGGSGGSGGSGGG(SEQ ID NO:97)、GGSGGSGGGGSGGGGS(SEQ ID NO:98)、GGSGGSGGSGGSGGSGGS(SEQ ID NO:99)或GGGGSGGGGSGGGGS(SEQ ID NO:100)。接头不消除或减弱VWF蛋白活性或因子VIII的凝结活性。任选地,接头增强VWF蛋白活性或因子VIII蛋白的凝结活性,例如通过进一步减弱空间位阻效应,以及使VWF蛋白或因子VIII部分更易于到达它的靶标结合位点。

[0415] 在一个实施方案中,适用于嵌合蛋白的接头的长度是15-25个氨基酸。在另一实施方案中,适用于嵌合蛋白的接头的长度是15-20个氨基酸。在一些实施方案中,用于嵌合蛋白的接头的长度是10-25个氨基酸。在其它实施方案中,用于嵌合蛋白的接头的长度是15个氨基酸。在其它实施方案中,用于嵌合蛋白的接头是 $(GGGS)_n$ (SEQ ID NO:94),其中G表示甘氨酸,S表示丝氨酸,并且n是整数1-20。

#### [0416] II.F. 裂解位点

[0417] 可裂解接头可并有能够以化学方式(例如水解酯键)、以酶促方式(即并有蛋白酶裂解序列)或以光解方式(例如发色团,诸如3-氨基-3-(2-硝基苯基)丙酸(ANP))裂解的部分以使一种分子从另一分子释放。

[0418] 在一个实施方案中,可裂解接头在N末端或C末端或两者处包含一个或多个裂解位点。在另一实施方案中,可裂解接头基本上由一个或多个可裂解位点组成或由一个或多个可裂解位点组成。在其它实施方案中,可裂解接头包含本文所述的异源性氨基酸接头序列或聚合物和一个或多个可裂解位点。

[0419] 在某些实施方案中,可裂解接头包含一个或多个可在宿主细胞中裂解的裂解位点(即细胞内加工位点)。裂解位点的非限制性实例包括RRRR(SEQ ID NO:102)、RKRRKR(SEQ ID NO:103)和RRRRS(SEQ ID NO:104)。

[0420] 在一些实施方案中,可裂解接头包括来自FVIII的a1区、来自FVIII的a2区、来自FVIII的a3区、包含X-V-P-R(SEQ ID NO:105)和PAR1外结合位点相互作用基序的凝血酶可裂解位点(其中X是脂族氨基酸)或其任何组合。包括a2区,其包含与对应于全长FVIII的Glu720至Arg740至少约80%、约85%、约90%、约95%或100%同一的氨基酸序列,其中所述



a2区能够由凝血酶裂解。在特定实施方案中,适用于本发明的可裂解接头包括a2区,其包含ISDKNTGDYEDSYEDISAYLLSKNNAIEPRFS(SEQ ID NO:106)。在其它实施方案中,用于本发明的可裂解接头包括a1区,其包含与对应于全长FVIII的Met337至Arg372至少约80%、约85%、约90%、约95%或100%同一的氨基酸序列,其中所述a1区能够由凝血酶裂解。在特定实施方案中,a1区包含ISMKNNEEAEDYDDDLTDSEMDVVRFDNNSPSFIQIRSV(SEQ ID NO:107)。在一些实施方案中,本发明的可裂解接头包括a3区,其包含与对应于全长FVIII的Glu1649至Arg1689至少约80%、约85%、约90%、约95%或100%同一的氨基酸序列,其中所述a3区能够由凝血酶裂解。在特定实施方案中,用于本发明的可裂解接头包括a3区,其包含ISEITRTLQSDQEEIDYDDTISVEMKKEDFDIYDEDENQSPRSFQ(SEQ ID NO:108)。

[0421] 在其它实施方案中,可裂解接头包括凝血酶裂解位点,其包含X-V-P-R(SEQ ID NO:105)和PAR1外结合位点相互作用基序,并且其中所述PAR1外结合位点相互作用基序包含S-F-L-L-R-N(SEQ ID NO:109)。PAR1外结合位点相互作用基序可进一步包含选自以下的氨基酸序列:P、P-N、P-N-D、P-N-D-K(SEQ ID NO:110)、P-N-D-K-Y(SEQ ID NO:111)、P-N-D-K-Y-E(SEQ ID NO:112)、P-N-D-K-Y-E-P(SEQ ID NO:113)、P-N-D-K-Y-E-P-F(SEQ ID NO:114)、P-N-D-K-Y-E-P-F-W(SEQ ID NO:115)、P-N-D-K-Y-E-P-F-W-E(SEQ ID NO:116)、P-N-D-K-Y-E-P-F-W-E-D(SEQ ID NO:117)、P-N-D-K-Y-E-P-F-W-E-D-E(SEQ ID NO:118)、P-N-D-K-Y-E-P-F-W-E-D-E-E(SEQ ID NO:119)、P-N-D-K-Y-E-P-F-W-E-D-E-E-S(SEQ ID NO:120)或其任何组合。在一些实施方案中,脂族氨基酸选自甘氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸。

[0422] 在其它实施方案中,可裂解接头包含一个或多个在向受试者施用包含所述可裂解接头的嵌合蛋白之后由蛋白酶裂解的裂解位点。在一个实施方案中,裂解位点由选自由以下组成的组的蛋白酶裂解:因子XIa、因子XIIa、胰舒血管素、因子VIIa、因子IXa、因子Xa、因子IIa(凝血酶)、弹性蛋白酶-2、MMP-12、MMP-13、MMP-17和MMP-20。在另一实施方案中,裂解位点选自由以下组成的组:FXIa裂解位点(例如KLTR↓AET(SEQ ID NO:121))、FXIa裂解位点(例如DFTR↓VVG(SEQ ID NO:122))、FXIIa裂解位点(例如TMTR↓IVGG(SEQ ID NO:123))、胰舒血管素裂解位点(例如SPFR↓STGG(SEQ ID NO:124))、FVIIa裂解位点(例如LQVR↓IVGG(SEQ ID NO:125))、FIXa裂解位点(例如PLGR↓IVGG(SEQ ID NO:126))、FXa裂解位点(例如IEGR↓TVGG(SEQ ID NO:127))、FIIa(凝血酶)裂解位点(例如LTPR↓SLLV(SEQ ID NO:128))、弹性蛋白酶-2裂解位点(例如LGPV↓SGVP(SEQ ID NO:129))、粒酶-B裂解(例如VAGD↓SLEE(SEQ ID NO:130))、MMP-12裂解位点(例如GPAG↓LGGA(SEQ ID NO:131))、MMP-13裂解位点(例如GPAG↓LRGA(SEQ ID NO:132))、MMP-17裂解位点(例如APLG↓LRLR(SEQ ID NO:133))、MMP-20裂解位点(例如PALP↓LVAQ(SEQ ID NO:134))、TEV裂解位点(例如ENLYFQ↓G(SEQ ID NO:135))、肠激酶裂解位点(例如DDDK↓IVGG(SEQ ID NO:136))、蛋白酶3C(PRESCISSION<sup>TM</sup>)裂解位点(例如LEVLVQ↓GP(SEQ ID NO:137))和分选酶A裂解位点(例如LPKT↓GSES)(SEQ ID NO:138)。在某些实施方案中,FXIa裂解位点包括但不限于例如TQSFNDFTR(SEQ ID NO:1)和SVSQTSLKTR(SEQ ID NO:3)。非限制性示例性凝血酶裂解位点包括例如DFLAEGGGVR(SEQ ID NO:4)、TTKIKPR(SEQ ID NO:5)、LVPRG(SEQ ID NO:6)、DKNTGDYEDSYEDISAYLLSKNNAIEPRFS(SEQ ID NO:88)或IEPRFS(SEQ ID NO:194)以及包含ALRPR(SEQ ID NO:7)、基本上由ALRPR(SEQ ID NO:7)组成或由ALRPR(SEQ ID NO:7)组成的序列(例如ALRPRVVGGA(SEQ ID

NO:145)。

[0423] 在特定实施方案中,裂解位点是TLDPRSFLLRNPNDKYEPFWEDEEK (SEQ ID NO:146)。在另一实施方案中,裂解位点包含DKNTGDYEDSYEDISAYLLSKNNAIEPRSFS (SEQ ID NO:88) 或其片段。在一个特定实施方案中,裂解位点包含IEPRSFS (SEQ ID NO:194)。在另一实施方案中,裂解位点包含EPRSFS (SEQ ID NO:195),其中所述裂解位点不是FVIII的全长a2区。在另一实施方案中,裂解位点包含IEPR (SEQ ID NO:200)。在另一实施方案中,裂解位点包含IEPR (SEQ ID NO:200),其中所述裂解位点不是FVIII的全长a2区,或不包含FVIII的全长a2区。在其它实施方案中,裂解位点包含DKNTGDYEDSYEDISAYLLSKNNAIEPRSFS (SEQ ID NO:88)、KNTGDYEDSYEDISAYLLSKNNAIEPRSFS (SEQ ID NO:139)、NTGDYEDSYEDISAYLLSKNNAIEPRSFS (SEQ ID NO:140)、TGDYEDSYEDISAYLLSKNNAIEPRSFS (SEQ ID NO:141)、GDYEDSYEDISAYLLSKNNAIEPRSFS (SEQ ID NO:142)、DYEDSYEDISAYLLSKNNAIEPRSFS (SEQ ID NO:143)、YEDSYEDISAYLLSKNNAIEPRSFS (SEQ ID NO:144)、YEDSYEDISAYLLSKNNAIEPRSFS (SEQ ID NO:176)、EDSYEDISAYLLSKNNAIEPRSFS (SEQ ID NO:177)、DSYEDISAYLLSKNNAIEPRSFS (SEQ ID NO:178)、SYEDISAYLLSKNNAIEPRSFS (SEQ ID NO:179)、YEDISAYLLSKNNAIEPRSFS (SEQ ID NO:180)、EDISAYLLSKNNAIEPRSFS (SEQ ID NO:181)、DISAYLLSKNNAIEPRSFS (SEQ ID NO:182)、ISAYLLSKNNAIEPRSFS (SEQ ID NO:183)、SAYLLSKNNAIEPRSFS (SEQ ID NO:184)、AYLLSKNNAIEPRSFS (SEQ ID NO:185)、YLLSKNNAIEPRSFS (SEQ ID NO:186)、LLSKNNAIEPRSFS (SEQ ID NO:187)、LSKNNAIEPRSFS (SEQ ID NO:188)、SKNNAIEPRSFS (SEQ ID NO:189)、KNNAIEPRSFS (SEQ ID NO:190)、NNAIEPRSFS (SEQ ID NO:191)、NAIEPRSFS (SEQ ID NO:192)、AIEPRSFS (SEQ ID NO:193) 或 IEPRSFS (SEQ ID NO:194)。在其它实施方案中,裂解位点包含DKNTGDYEDSYEDISAYLLSKNNAIEPRSFS (SEQ ID NO:88)、KNTGDYEDSYEDISAYLLSKNNAIEPRSFS (SEQ ID NO:139)、NTGDYEDSYEDISAYLLSKNNAIEPRSFS (SEQ ID NO:140)、TGDYEDSYEDISAYLLSKNNAIEPRSFS (SEQ ID NO:141)、GDYEDSYEDISAYLLSKNNAIEPRSFS (SEQ ID NO:142)、DYEDSYEDISAYLLSKNNAIEPRSFS (SEQ ID NO:143)、YEDSYEDISAYLLSKNNAIEPRSFS (SEQ ID NO:144)、YEDSYEDISAYLLSKNNAIEPRSFS (SEQ ID NO:176)、EDSYEDISAYLLSKNNAIEPRSFS (SEQ ID NO:177)、DSYEDISAYLLSKNNAIEPRSFS (SEQ ID NO:178)、SYEDISAYLLSKNNAIEPRSFS (SEQ ID NO:179)、YEDISAYLLSKNNAIEPRSFS (SEQ ID NO:180)、EDISAYLLSKNNAIEPRSFS (SEQ ID NO:181)、DISAYLLSKNNAIEPRSFS (SEQ ID NO:182)、ISAYLLSKNNAIEPRSFS (SEQ ID NO:183)、SAYLLSKNNAIEPRSFS (SEQ ID NO:184)、AYLLSKNNAIEPRSFS (SEQ ID NO:185)、YLLSKNNAIEPRSFS (SEQ ID NO:186)、LLSKNNAIEPRSFS (SEQ ID NO:187)、LSKNNAIEPRSFS (SEQ ID NO:188)、SKNNAIEPRSFS (SEQ ID NO:189)、KNNAIEPRSFS (SEQ ID NO:190)、NNAIEPRSFS (SEQ ID NO:191)、NAIEPRSFS (SEQ ID NO:192)、AIEPRSFS (SEQ ID NO:193) 或 IEPRSFS (SEQ ID NO:194),其中所述裂解位点不是全长FVIII a2区。在某些实施方案中,在如本文提供或如本领域中已知的凝血酶裂解测定中,可裂解接头是可裂解的。

[0424] III. 多核苷酸、载体和宿主细胞

[0425] 本发明中也提供一种编码本发明的嵌合蛋白的多核苷酸。在一个实施方案中,第一多肽链和第二多肽链可由单一多核苷酸链编码。在另一实施方案中,第一多肽链和第二多肽链由两个不同多核苷酸(即第一核苷酸序列和第二核苷酸序列)编码。在另一实施方案

中,第一核苷酸序列和第二核苷酸序列在两个不同多核苷酸(例如不同载体)上。

[0426] 本发明包括一种编码单一多肽链(例如FVIII(X2)-F1-L3-F2-L2-X1-L1-V)的多核苷酸,其中FVIII(X2)包括其中在一个或多个插入位点处插入XTEN序列的FVIII蛋白,F1包括第一Ig恒定区或其一部分,例如第一Fc区,L1包括第一接头,V包括VWF蛋白,X1包括在长度方面具有小于288个氨基酸的XTEN序列,L2包括第二接头,L3包括第三接头,并且F2包括第二Ig恒定区或其一部分,例如第二Fc区。本发明也包括两个多核苷酸,第一多核苷酸序列编码包含融合于第一Ig恒定区或其一部分的FVIII蛋白的第一多肽,并且第二多核苷酸序列编码包含VWF蛋白、在长度方面具有小于288个氨基酸的XTEN序列以及第二Ig恒定区或其一部分的第二多肽。在一些实施方案中,包含两个多肽链或三个多肽链的嵌合蛋白可由单一多核苷酸链编码,接着加工成两个或三个(或更多个)多肽链。在其它实施方案中,包含这些多肽链的嵌合蛋白可由两个或三个多核苷酸链编码。

[0427] 在其它实施方案中,多核苷酸组进一步包含编码蛋白质转化酶的另外的核苷酸链(例如当嵌合多肽由单一多核苷酸链编码时的第二核苷酸链或当嵌合蛋白由两个多核苷酸链编码时的第三核苷酸链)。蛋白质转化酶可选自由以下组成的组:原蛋白转化酶枯草杆菌蛋白酶/kexin 5型(PCSK5或PC5)、原蛋白转化酶枯草杆菌蛋白酶/kexin 7型(PCSK7或PC5)、酵母Kex 2、原蛋白转化酶枯草杆菌蛋白酶/kexin 3型(PACE或PCSK3)及其两个或更多个组合。在一些实施方案中,蛋白质转化酶是PACE、PC5或PC7。在特定实施方案中,蛋白质转化酶是PC5或PC7。参见国际申请号PCT/US2011/043568。

[0428] 如本文所用,表达载体是指含有在引入适当宿主细胞中时,为转录和翻译插入编码序列所必需的元件,或在RNA病毒载体的情况下,为复制和翻译所必需的元件的任何核酸构建体。表达载体可包括质粒、噬菌粒、病毒及其衍生物。

[0429] 本发明的表达载体将包括编码本文所述的嵌合蛋白的多核苷酸。在一个实施方案中,包含FVIII蛋白和第一Ig恒定区的第一多肽;包含VWF蛋白、具有小于288个氨基酸的XTEN序列以及第二Ig恒定区或其一部分的第二多肽;或两者的编码序列中的一个或多个被可操作地连接于表达控制序列。如本文所用,当两个核酸序列以允许各组成核酸序列保留它的功能性的方式共价连接时,它们被可操作地连接。当编码序列和基因表达控制序列以将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,它们被称为被可操作地连接。如果诱导5'基因表达序列中的启动子导致编码序列转录,并且如果两个DNA序列之间的键联的性质不(1)导致引入框移突变,(2)干扰启动子区域引导编码序列转录的能力,或(3)干扰相应RNA转录物被翻译成蛋白质的能力,那么两个DNA序列被称为被可操作地连接。因此,如果基因表达序列能够实现编码核酸序列的转录以使所得转录物翻译成所需蛋白质或多肽,那么所述基因表达序列将被可操作地连接于那个编码核酸序列。

[0430] 如本文所用的基因表达控制序列是有助于它所可操作地连接的编码核酸的高效转录和翻译的任何调控核苷酸序列,诸如启动子序列或启动子-增强子组合。基因表达控制序列可例如是哺乳动物或病毒启动子,诸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、 $\beta$ -肌动蛋白启动子和其它组成性启动子。在真核细胞中组成性起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(papilloma

virus)、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(Moloney leukemia virus)的长末端重复序列(LTR)和其它逆转录病毒的启动子以及单纯疱疹病毒的胸苷激酶启动子。其它组成性启动子为本领域普通技术人员所知。适用作本发明的基因表达序列的启动子也包括诱导性启动子。诱导性启动子在诱导剂存在下表达。举例来说,诱导金属硫蛋白启动子以促进在某些金属离子存在下的转录和翻译。其它诱导性启动子为本领域普通技术人员所知。

[0431] 一般来说,基因表达控制序列在必要时将包括分别涉及转录和翻译起始的5'非转录和5'非翻译序列,诸如TATA盒、加帽序列、CAAT序列等。尤其地,所述5'非转录序列将包括启动子区域,所述区域包括用于对可操作地接合的编码核酸进行转录控制的启动子序列。基因表达序列任选包括如所需的增强子序列或上游活化子序列。

[0432] 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,诸如莫洛尼鼠类白血病病毒、哈维(Harvey)鼠类肉瘤病毒、鼠类乳腺肿瘤病毒和劳斯肉瘤病毒;腺病毒、腺相关病毒;SV40型病毒;多型瘤病毒;艾伯斯坦-巴尔病毒(Epstein-Barr virus);乳头状瘤病毒;疱疹病毒;痘苗病毒;脊髓灰质炎病毒;和RNA病毒,诸如逆转录病毒。可易于采用本领域中熟知的其它载体。某些病毒载体基于其中非必需基因已被目标基因置换的非细胞病变真核病毒。非细胞病变病毒包括逆转录病毒,其生命周期涉及基因组病毒RNA逆转录成DNA,随后前病毒整合至宿主细胞DNA中。逆转录病毒已被核准用于人基因疗法试验。最适用的是具有复制缺陷性的那些逆转录病毒(即能够引导合成所需蛋白质,但不能制造感染性粒子)。所述遗传改变的逆转录病毒表达载体具有用于在体内高效率转导基因的一般效用。用于产生复制缺陷性逆转录病毒的标准方案(包括以下步骤:将外源性遗传物质并入质粒中,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,以及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) 以及 Murry, E.J., *Methods in Molecular Biology*, 第7卷, Humana Press, Inc., Clifton, N.J. (1991) 中。

[0433] 在一个实施方案中,病毒是腺相关病毒,其是一种双链DNA病毒。腺相关病毒可被工程改造以具有复制缺陷性,并且能够感染广泛范围的细胞类型和物种。它进一步具有诸如以下的优势:热和脂质溶剂稳定性;在不同谱系的细胞(包括造血细胞)中具有高转导频率;以及缺乏重复感染抑制,因此允许进行多级转导。据报道,腺相关病毒可以位点特异性方式整合至人细胞DNA中,由此使逆转录病毒感染的插入诱变的可能性和插入基因表达特征的可变性最小。此外,野生型腺相关病毒感染已在不存在选择压力下在组织培养中进行大于100继代,从而暗示腺相关病毒基因组整合是相对稳定事件。腺相关病毒也可以染色体外方式起作用。

[0434] 其它载体包括质粒载体。质粒载体已在本领域中广泛描述,并且为本领域技术人员所熟知。参见例如 Sambrook 等, *Molecular Cloning: A Laboratory Manual*, 第2版, Cold Spring Harbor Laboratory Press, 1989。在过去数年中,已发现质粒载体由于不能在宿主基因组内复制以及整合至宿主基因组中而特别有利于在体内将基因递送至细胞中。然而,具有可与宿主细胞相容的启动子的这些质粒可从可操作地编码于质粒内的基因表达肽。一些可从商业供应商获得的通常使用的质粒包括 pBR322、pUC18、pUC19、各种 pcDNA 质粒、pRC/CMV、各种 pCMV 质粒、pSV40 和 pBlueScript。特定质粒的其它实例包括 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,全部来自Invitrogen(Carlsbad,CA.)。其它质粒为本领域普通技术人员所熟知。另外,可使用标准分子生物学技术定制设计质粒以移除和/或添加特定DNA片段。

[0435] 在一种可用于产生本发明的蛋白质的昆虫表达系统中,苜蓿银纹夜蛾核多角体病毒(*Autographa californica nuclear polyhidrosis virus*,AcNPV)用作表达外来基因的载体。病毒生长在草地贪夜蛾(*Spodoptera frugiperda*)细胞中。编码序列可被克隆至病毒的非必需区域(例如多角体基因)中,并且置于ACNPV启动子(例如多角体启动子)的控制下。成功插入编码序列将导致多角体基因失活,以及产生非封闭重组病毒(即缺乏由多角体基因编码的蛋白质外壳的病毒)。这些重组病毒接着用于感染其中表达插入基因的草地贪夜蛾细胞。(参见例如Smith等(1983)*J Virol* 46:584;美国专利号4,215,051)。这个表达系统的其它实例可见于Ausubel等编(1989)*Current Protocols in Molecular Biology*,第2卷,Greene Publish.Assoc.&Wiley Interscience中。

[0436] 可用于表达本发明的蛋白质的另一系统是谷氨酰胺合成酶基因表达系统,也被称为“GS表达系统”(Lonza Biologics PLC,Berkshire UK)。这个表达系统详述于美国专利号5,981,216中。

[0437] 在哺乳动物宿主细胞中,可利用许多基于病毒的表达系统。在其中腺病毒用作表达载体的情况下,编码序列可连接于腺病毒转录/翻译控制复合物,例如晚期启动子和三联前导序列。这个嵌合基因可接着通过体外或体内重组插入腺病毒基因组中。插入在病毒基因组的非必需区域(例如区域E1或E3)中将产生在受感染宿主中具有活力,并且能够表达肽的重组病毒。参见例如Logan和Shenk(1984)*Proc Natl Acad Sci USA* 81:3655)。或者,可使用痘苗7.5K启动子。参见例如Mackett等(1982)*Proc Natl Acad Sci USA* 79:7415;Mackett等(1984)*J Virol* 49:857;Panicali等(1982)*Proc Natl Acad Sci USA* 79:4927。

[0438] 为增加产生效率,多核苷酸可被设计以编码由酶裂解位点分开的本发明的蛋白质的多个单元。所得多肽可被裂解(例如通过用适当酶处理)以回收多肽单元。这可增加由单一启动子驱动的多肽的产量。当用于适当病毒表达系统中时,由mRNA编码的各多肽的翻译在转录物内部受引导;例如由内部核糖体进入位点IRES引导。因此,多顺反子构建体引导单一大型多顺反子mRNA的转录,此转而引导多个个别多肽的翻译。这个方法消除多聚蛋白的产生和酶促加工,并且可显著增加由单一启动子驱动的多肽的产量。

[0439] 转化中使用的载体将通常含有用于鉴定转化体的可选择标记。在细菌系统中,此可包括抗生素抗性基因,诸如氨苄青霉素或卡那霉素。用于培养的哺乳动物细胞中的可选择标记包括赋予对诸如新霉素(neomycin)、潮霉素和甲氨蝶呤(methotrexate)的药物的抗性的基因。可选择标记可为可扩增可选择标记。一种可扩增可选择标记是二氢叶酸还原酶(dihydrofolate reductase,DHFR)基因。Simonsen C C等(1983)*Proc Natl Acad Sci USA* 80:2495-9。可选择标记由Thilly(1986)*Mammalian Cell Technology*,Butterworth Publishers,Stoneham,Mass.综述,并且对可选择标记的选择完全属于本领域中的普通技能水平。

[0440] 可选择标记可于单独质粒上与目标基因同时引入细胞中,或它们可于同一质粒上引入。如果于同一质粒上,那么可选择标记和目标基因可在不同启动子或相同启动子控制下,后述排列会产生双顺反子信息。这个类型的构建体在本领域中是已知的(例如美国专利

号4,713,339)。

[0441] 表达载体可编码允许对重组产生的蛋白质进行简易纯化的标签。实例包括但不限于载体pUR278(Ruther等(1983)EMBO J 2:1791),其中待表达的蛋白质的编码序列可与lac z编码区同框连接至载体中以使标签化融合蛋白得以产生;pGEX载体可用于表达具有谷胱甘肽(glutathione)S转移酶(GST)标签的本发明的蛋白质。这些蛋白质通常是可溶性的,并且可易于通过吸附于谷胱甘肽-琼脂糖珠粒,随后在游离谷胱甘肽存在下洗脱来从细胞纯化。载体包括用于在纯化之后简易移除标签的裂解位点(凝血酶或因子Xa蛋白酶或PRESCISSIIONPROTEASE™(Pharmacia,Peapack,N.J.))。

[0442] 一种或多种表达载体接着转染或共转染至将表达多肽的适合靶标细胞中。本领域中已知的转染技术包括但不限于磷酸钙沉淀(Wigler等(1978)Cell 14:725)、电穿孔(Neumann等(1982)EMBO J 1:841)和基于脂质体的试剂。多种宿主表达载体系统可用于表达本文所述的蛋白质,所述系统包括原核细胞与真核细胞两者。这些系统包括但不限于微生物,诸如用含有适当编码序列的重组噬菌体DNA或质粒DNA表达载体转化的细菌(例如大肠杆菌(E.coli));用含有适当编码序列的重组酵母或真菌表达载体转化的酵母或丝状真菌;用含有适当编码序列的重组病毒表达载体(例如杆状病毒)感染的昆虫细胞系统;用重组病毒表达载体(例如花椰菜花叶病毒(cauliflower mosaic virus)或烟草花叶病毒(tobacco mosaic virus))感染或用含有适当编码序列的重组质粒表达载体(例如Ti质粒)转化的植物细胞系统;或动物细胞系统,包括哺乳动物细胞(例如HEK 293、CHO、Cos、HeLa、HKB11和BHK细胞)。

[0443] 在一个实施方案中,宿主细胞是真核细胞。如本文所用,真核细胞是指具有定形细胞核的任何动物或植物细胞。动物的真核细胞包括脊椎动物(例如哺乳动物)的细胞和无脊椎动物(例如昆虫)的细胞。植物的真核细胞可明确包括不限于酵母细胞。真核细胞不同于例如细菌的原核细胞。

[0444] 在某些实施方案中,真核细胞是哺乳动物细胞。哺乳动物细胞是源于哺乳动物的任何细胞。哺乳动物细胞明确包括但不限于哺乳动物细胞系。在一个实施方案中,哺乳动物细胞是人细胞。在另一实施方案中,哺乳动物细胞是HEK 293细胞,其是一种人胚肾细胞系。HEK 293细胞可以CRL-1533从美国典型培养物保藏中心(American Type Culture Collection,Manassas,VA)以及以293-H细胞(目录号11631-017)或293-F细胞(目录号11625-019)从Invitrogen(Carlsbad,Calif.)获得。在一些实施方案中,哺乳动物细胞是PER.C6®细胞,其是一种源于视网膜的人细胞系。PER.C6®细胞可从Crucell(Leiden,The Netherlands)获得。在其它实施方案中,哺乳动物细胞是中国仓鼠卵巢(CHO)细胞。CHO细胞可从美国典型培养物保藏中心(Manassas,VA.)获得(例如CHO-K1;CCL-61)。在其它实施方案中,哺乳动物细胞是幼小仓鼠肾(BHK)细胞。BHK细胞可从美国典型培养物保藏中心(Manassas,VA.)获得(例如CRL-1632)。在一些实施方案中,哺乳动物细胞是HKB11细胞,其是HEK293细胞与人B细胞系的杂交细胞系。Mei等,Mol.Biotechnol.34(2):165-78(2006)。

[0445] 在一个实施方案中,将包括FVIII(X2)-Fc融合物编码序列、VWF蛋白-L1-X1-L2-Fc编码序列或两者和可选择标记(例如博莱霉素(zeocin)抗性)的质粒转染至HEK 293细胞中以产生嵌合蛋白。

[0446] 在另一实施方案中,将包括FVIII-Fc融合物编码序列、VWF蛋白-L1-X-L2-Fc编码

序列或两者和可选择标记(例如博莱霉素(zeocin)抗性)的质粒转染至HEK 293细胞中以产生嵌合蛋白。

[0447] 在一些实施方案中,将包括FVIII(X2)-Fc融合物编码序列和第一可选择标记(例如博莱霉素抗性基因)的第一质粒以及包括VWF蛋白-L1-X1-L2-Fc编码序列和第二可选择标记(例如新霉素抗性基因)的第二质粒和包括蛋白质转化酶编码序列和第三可选择标记(例如潮霉素抗性基因)的第三质粒共转染至HEK 293细胞中以产生嵌合蛋白。第一和第二质粒可以相等量(即1:1摩尔比)引入,或它们可以不等量引入。

[0448] 在其它实施方案中,将包括FVIII-Fc融合物编码序列和第一可选择标记(例如博莱霉素抗性基因)的第一质粒以及包括VWF蛋白-L1-X-L2-Fc编码序列和第二可选择标记(例如新霉素抗性基因)的第二质粒和包括蛋白质转化酶编码序列和第三可选择标记(例如潮霉素抗性基因)的第三质粒共转染至HEK 293细胞中以产生嵌合蛋白。第一和第二质粒可以相等量(即1:1摩尔比)引入,或它们可以不等量引入。

[0449] 在其它实施方案中,将包括FVIII(X2)-Fc融合物编码序列和第一可选择标记(例如博莱霉素抗性基因)的第一质粒以及包括VWF蛋白-L1-X1-L2-Fc融合物编码序列和第二可选择标记(例如新霉素抗性基因)的第二质粒和包括蛋白质转化酶编码序列和第三可选择标记(例如潮霉素抗性基因)的第三质粒共转染至HEK 293细胞中以产生嵌合蛋白。第一和第二质粒可以相等量(即1:1摩尔比)引入,或它们可以不等量引入。

[0450] 在某些实施方案中,将包括嵌合蛋白编码FVIII(有或无XTEN)-F1-L3-F2-L2-X-L1-V编码序列和第一可选择标记(例如博莱霉素抗性基因)的第一质粒和包括蛋白质转化酶编码序列和第二可选择标记(例如潮霉素抗性基因)的第二质粒共转染至HEK 293细胞中以产生嵌合蛋白。FVIII(X)-F1编码序列的启动子和V-L2-X-L1-F2编码序列的启动子可不同,或它们可相同。

[0451] 在其它实施方案中,转染的细胞是稳定转染的。使用本领域技术人员已知的常规技术,这些细胞可被选择并维持为稳定细胞系。

[0452] 使含有蛋白质的DNA构建体的宿主细胞在适当生长培养基中生长。如本文所用,术语“适当生长培养基”意指含有为细胞生长所需的营养物的培养基。为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,培养基可含有一种或多种选择因子。任选地,培养基可含有小牛血清或胎牛血清(FCS)。在一个实施方案中,培养基大致上不含有IgG。生长培养基将通常通过例如药物选择或缺乏必需营养物来选择含有DNA构建体的细胞,所述必需营养物由DNA构建体上或与DNA构建体共转染的可选择标记补充。通常使培养的哺乳动物细胞生长在可商购获得的含血清或无血清培养基(例如MEM、DMEM、DMEM/F12)中。在一个实施方案中,培养基是CD293(Invitrogen,Carlsbad,CA.)。在另一实施方案中,培养基是CD17(Invitrogen,Carlsbad,CA.)。对适于所用特定细胞系的培养基的选择属于本领域普通技术人员的水平。

[0453] 为共同表达嵌合蛋白的两个多肽链,在允许表达两个链的条件下培养宿主细胞。如本文所用,培养是指在体外维持活细胞至少一定时间。维持可但无需包括活细胞群体增加。举例来说,培养中维持的细胞可在群体方面是静态的,但仍然具有活力,并且能够产生所需产物,例如重组蛋白或重组融合蛋白。适用于培养真核细胞的条件在本领域中是熟知的,并且包括适当选择培养基、培养基补充剂、温度、pH、氧饱和度等。出于商业目的,培养可

包括使用各种类型的按比例扩大系统中的任一个,所述系统包括振荡器烧瓶、滚瓶、中空纤维生物反应器、搅拌槽生物反应器、气升生物反应器、摇袋式生物反应器(Wave bioreactor)和其它系统。

[0454] 也选择细胞培养条件以允许VWF片段与FVIII蛋白缔合。允许表达VWF片段和/或FVIII蛋白的条件可包括存在维生素K来源。举例来说,在一个实施方案中,在补充有4mM谷氨酰胺的CD293培养基(Invitrogen, Carlsbad, CA)或OptiCHO培养基(Invitrogen, Carlsbad, CA)中培养稳定转染的HEK 293细胞。

[0455] 在一个方面,本发明涉及一种表达、制备或产生本发明的嵌合蛋白的方法,其包括a) 转染包含编码所述嵌合蛋白的多核苷酸的宿主细胞以及b) 在适于表达所述嵌合蛋白的条件下在培养基中培养所述宿主细胞,其中所述嵌合蛋白得以表达。

[0456] 在其它实施方案中,含有连接于第一Ig恒定区或其一部分的FVIII蛋白和/或通过XTEN序列融合于第二Ig恒定区或其一部分的VWF蛋白的蛋白质产物被分泌至培养基中。使培养基与细胞分离,浓缩,过滤,接着穿过两个或三个亲和柱,例如蛋白质A柱和一个或两个阴离子交换柱。

[0457] 在某些方面,本发明涉及通过本文所述的方法产生的嵌合蛋白。

[0458] 体外生产允许按比例扩大以产生本发明的大量所需改变多肽。用于在组织培养条件下进行哺乳动物细胞培养的技术在本领域中是已知的,并且包括例如在气升反应器中或在连续搅拌器反应器中进行均质悬浮培养,或例如在中空纤维、微囊中、在琼脂糖微珠粒或陶瓷柱筒上进行固定或包埋细胞培养。必要和/或需要时,多肽的溶液可通过惯用色谱方法纯化,所述方法例如凝胶过滤、离子交换色谱法、疏水性相互作用色谱法(HIC、经DEAE-纤维素的色谱法或亲和色谱法)。

#### [0459] IV. 药物组合物

[0460] 含有本发明的嵌合蛋白的组合物可含有适合药学上可接受的载体。举例来说,它们可含有有助于将活性化合物加工成被设计用于递送至作用部位的制剂的赋形剂和/或助剂。

[0461] 药物组合物可被配制以用于通过快速浓注进行胃肠外施用(即静脉内、皮下或肌肉内)。注射用制剂可以例如于安瓿中或于多剂量容器中的添加有防腐剂的单位剂型提供。组合物可采用诸如于油性或水性媒介物中的混悬液、溶液或乳液的形式,并且可含有配制剂,诸如混悬剂、稳定剂和/或分散剂。或者,活性成分可呈用于用适合媒介物(例如无热原水)复原的粉末形式。

[0462] 适于胃肠外施用的制剂也包括呈水溶性形式(例如水溶性盐)的活性化合物的水溶液。此外,可施用呈适当油性注射混悬液形式的活性化合物混悬液。适合亲脂性溶剂或媒介物包括脂肪油(例如芝麻油)或合成脂肪酸酯(例如油酸乙酯或甘油三酯)。水性注射混悬液可含有增加混悬液的粘度的物质,包括例如羧甲基纤维素钠、山梨糖醇和右旋糖苷。任选地,混悬液也可含有稳定剂。脂质体也可用于囊封本发明的分子以递送至细胞或间质间隙中。示例性药学上可接受的载体是生理可相容溶剂、分散介质、包衣剂、抗菌剂和抗真菌剂、等张剂和吸收延迟剂、水、盐水、磷酸盐缓冲盐水、右旋糖、甘油、乙醇等。在一些实施方案中,组合物包含等张剂,例如糖、多元醇(诸如甘露糖醇、山梨糖醇)或氯化钠。在其它实施方案中,组合物包含增强活性成分的储存期限或有效性的药学上可接受的物质,诸如湿润



剂或少量辅助物质,诸如湿润剂或乳化剂、防腐剂或缓冲剂。

[0463] 本发明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、混悬液、半固体和固体剂型。优选形式取决于施用模式和治疗应用。

[0464] 组合物可被配制成溶液、微乳液、分散液、脂质体或适于高药物浓度的其它有序结构。可通过将活性成分以所需量根据需要与以上列举的一种成分或成分的组合一起并入适当溶剂中,随后进行过滤灭菌来制备无菌可注射溶液。通常,通过将活性成分并入含有基本分散介质和来自以上列举的成分的所需其它成分的无菌媒介物中来制备分散液。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,其产生活性成分外加来自先前无菌过滤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包衣剂(诸如卵磷脂(lecithin)),在分散液的情况下通过维持所需粒度以及通过使用表面活性剂加以维持。可注射组合物的延长吸收可通过在组合物中包括延迟吸收的试剂(例如单硬脂酸盐和明胶)来达成。

[0465] 活性成分可与控制释放制剂或装置一起配制。所述制剂和装置的实例包括植入物、经皮贴片和微囊封递送系统。可使用可生物降解的生物相容性聚合物,例如乙烯乙酸乙酯、聚酸酐、聚乙醇酸、胶原蛋白、聚原酸酯和聚乳酸。用于制备所述制剂和装置的方法在本领域中是已知的。参见例如Sustained and Controlled Release Drug Delivery Systems, J.R. Robinson编, Marcel Dekker, Inc., New York, 1978。

[0466] 可通过形成药物于诸如聚丙交酯-聚乙交酯的可生物降解的聚合物中的微囊封基质来制备可注射储槽制剂。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。其它示例性可生物降解的聚合物是聚原酸酯和聚酸酐。储槽可注射制剂也可通过将药物包埋在脂质体或微乳液中来制备。

[0467] 补充性活性化合物可并入组合物中。在一个实施方案中,本发明的嵌合蛋白与另一凝结因子或其变体、片段、类似物或衍生物一起配制。举例来说,凝结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、范威尔邦德因子或重组可溶性组织因子(rsTF)或任何前述各物的活化形式。止血剂的凝结因子也可包括抗纤维蛋白溶解药物,例如 $\epsilon$ -氨基-己酸、氨甲环酸。

[0468] 可调整剂量方案以提供最优所需响应。举例来说,可施用单次大丸剂,可随时间施用若干分次剂量,或可如通过治疗情况的紧急性所指示,按比例降低或增加剂量。有利的是以剂量单位形式配制胃肠外组合物以易于施用和达成剂量均一性。参见例如Remington's Pharmaceutical Sciences (Mack Pub. Co., Easton, Pa. 1980)。

[0469] 除活性化合物之外,液体剂型也可含有惰性成分,诸如水、乙醇、碳酸乙酯、乙酸乙酯、苯甲醇、苯甲酸苄酯、丙二醇、1,3-丁二醇、二甲基甲酰胺、油、甘油、四氢呋喃甲醇、聚乙二醇和脱水山梨醇的脂肪酸酯。

[0470] 适合药物载体的非限制性实例也描述于由E.W. Martin所著的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、稻米、面粉、白垩、硅胶、硬脂酸钠、甘油单硬脂酸酯、滑石、氯化钠、脱脂乳粉、甘油、丙二醇、水、乙醇等。组合物也可含有pH缓冲试剂和湿润剂或乳化剂。

[0471] 对于口服施用,药物组合物可采用通过常规手段制备的片剂或胶囊形式。组合物也可制备成液体,例如糖浆或混悬液。液体可包括混悬剂(例如山梨糖醇糖浆、纤维素衍生

物或氢化食用脂肪)、乳化剂(卵磷脂或阿拉伯胶(acacia))、非水性媒介物(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)。制剂也可包括调味剂、着色剂和甜味剂。或者,组合物可以用于用水或另一适合媒介物复原的干燥产品形式提供。

[0472] 对于经颊施用,组合物可采用根据常规方案的片剂或糖锭形式。

[0473] 对于通过吸入施用,供根据本发明使用的化合物宜以有或无赋形剂的雾化气雾剂形式或以来自任选具有推进剂的加压包装或雾化器的气雾喷雾剂形式递送,所述推进剂例如是二氯二氟甲烷、三氯氟甲烷、二氯四氟甲烷、二氧化碳或其它适合气体。在加压气雾剂的情况下,剂量单位可通过提供用以递送计量数量的阀来确定。可配制用于吸入器或吹入器中的例如明胶的胶囊和药筒,其含有化合物与诸如乳糖或淀粉的适合粉末基质的粉末混合物。

[0474] 药物组合物也可以例如含有常规栓剂基质(诸如可可脂或其它甘油酯)的栓剂或滞留灌肠剂形式配制用于经直肠施用。

[0475] 在一个实施方案中,药物组合物包含嵌合蛋白、编码所述嵌合蛋白的多核苷酸、包含所述多核苷酸的载体或包含所述载体的宿主细胞以及药学上可接受的载体。相较于野生型FVIII蛋白或无VWF片段的相应FVIII蛋白,嵌合蛋白中的FVIII蛋白具有延长的半衰期。在一个实施方案中,其中嵌合蛋白的半衰期比野生型FVIII延长至少约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。在另一实施方案中,因子VIII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。

[0476] 在一些实施方案中,通过选自由以下组成的组的途径施用组合物:局部施用、眼内施用、胃肠外施用、鞘内施用、硬膜下施用和口服施用。胃肠外施用可为静脉内或皮下施用。

[0477] 在其它实施方案中,组合物用于治疗有需要的受试者的流血疾病或病状。流血疾病或病状选自由以下组成的组:流血凝血病症、关节积血、肌肉流血、口腔流血、出血、向肌肉中出血、口腔出血、创伤、头创伤、胃肠流血、颅内出血、腹内出血、胸内出血、骨折、中枢神经系统流血、咽后隙中流血、腹膜后隙中流血、髂腰肌鞘中流血及其任何组合。在其它实施方案中,受试者预定经受手术。在其它实施方案中,治疗是防治性的或按需进行。

[0478] V. 基因疗法

[0479] 本发明的其嵌合蛋白可在例如人患者的哺乳动物中体内产生,使用用以治疗选自由以下组成的组的流血疾病或病症的基因疗法方法将是治疗有益的:流血凝血病症、关节积血、肌肉流血、口腔流血、出血、向肌肉中出血、口腔出血、创伤、头创伤、胃肠流血、颅内出血、腹内出血、胸内出血、骨折、中枢神经系统流血、咽后隙中流血、腹膜后隙中流血和髂腰肌鞘中流血。在一个实施方案中,流血疾病或病症是血友病。在另一实施方案中,流血疾病或病症是甲型血友病。这涉及施用可操作地连接于适合表达控制序列的适合嵌合蛋白编码核酸。在某一实施方案中,这些序列并入病毒载体中。适用于所述基因疗法的病毒载体包括

腺病毒载体、慢病毒载体、杆状病毒载体、艾伯斯坦巴尔病毒载体、乳多泡病毒载体 (papovaviral vector)、痘苗病毒载体、单纯性疱疹病毒载体和腺相关病毒 (AAV) 载体。病毒载体可为复制缺陷性病毒载体。在其它实施方案中,腺病毒载体在它的E1基因或E3基因中具有缺失。当使用腺病毒载体时,哺乳动物可不暴露于编码可选择标记基因的核酸。在其它实施方案中,将序列并入本领域技术人员已知的非病毒载体中。

[0480] VI. 使用嵌合蛋白的方法

[0481] 本发明涉及一种使用本文所述的嵌合蛋白来防止或抑制内源性VWF结合FVIII蛋白的方法。本发明也涉及一种使用具有连接于XTEN和Ig恒定区或其一部分的FVIII蛋白的嵌合蛋白的方法。

[0482] 本发明的一个方面涉及通过阻断或遮蔽FVIII上的VWF结合位点免遭内源性VWF来防止或抑制FVIII与内源性VWF相互作用,以及同时使用与也可为半衰期延长物的Ig恒定区或其一部分组合的XTEN序列来延长嵌合蛋白的半衰期。在一个实施方案中,本发明涉及一种构建半衰期长于野生型FVIII的FVIII蛋白的方法。适用于方法中的嵌合蛋白包括本文所述的任何一种或多种嵌合蛋白。

[0483] 本发明的另一方面包括一种向有需要的受试者施用包含半衰期长于野生型FVIII的FVIII蛋白的嵌合蛋白的方法,其中所述方法包括向所述受试者施用本文所述的嵌合蛋白。

[0484] 在一个实施方案中,本发明涉及一种使用XTEN序列和Ig恒定区或其一部分来改进包含FVIII蛋白和VWF蛋白的嵌合蛋白的半衰期的方法,其防止或抑制内源性VWF与FVIII蛋白相互作用。连接于XTEN序列(例如FVIII(X)),接着与融合于XTEN和Ig恒定区或其一部分的VWF蛋白结合或缔合的FVIII蛋白被遮蔽或保护免遭VWF的清除路径,并且因此相较于未结合于VWF蛋白的FVIII蛋白,清除率降低。因此,相较于未与XTEN序列和VWF蛋白结合或缔合的FVIII蛋白,遮蔽的FVIII蛋白的半衰期得以最大延长。在某些实施方案中,与VWF蛋白缔合或由VWF蛋白保护,并且连接于XTEN序列的FVIII蛋白不由VWF清除受体清除。在其它实施方案中,与VWF蛋白缔合或由VWF蛋白保护,并且连接于XTEN序列的FVIII蛋白从系统被清除慢于未与VWF蛋白缔合或由VWF蛋白保护,并且连接于XTEN序列的FVIII蛋白。

[0485] 在一个方面,包含连接于XTEN序列的FVIII蛋白或与连接于XTEN的VWF蛋白结合或缔合的FVIII蛋白的嵌合蛋白从循环的清除率降低,因为VWF蛋白不含有VWF清除受体结合位点。VWF蛋白防止或抑制与VWF蛋白结合或缔合的FVIII通过VWF清除路径从系统清除。适用于本发明的VWF蛋白也可提供至少一种或多种由内源性VWF提供的VWF样FVIII保护性质。在某些实施方案中,VWF蛋白或XTEN序列也可掩蔽一个或多个FVIII清除受体结合位点,由此防止FVIII通过它自身的清除路径清除。

[0486] 在一些实施方案中,由VWF蛋白或XTEN序列防止或抑制FVIII蛋白结合内源性VWF可发生在体外或体内。

[0487] 也提供一种增加嵌合蛋白的半衰期的方法,其包括向有需要的受试者施用本文所述的嵌合蛋白。与全长VWF结合或缔合的非活化FVIII在血浆中的半衰期是约12至14小时。其中在循环中几乎不存在VWF的3型VWD中,FVIII的半衰期仅为约六小时,从而由于FVIII浓度降低而在所述患者中产生轻度至中度甲型血友病的症状。本发明的与VWF片段或XTEN序列连接或缔合的嵌合蛋白的半衰期可比与全长VWF结合或缔合的非活化FVIII的半衰期增

加高至少约1.5倍、1.6倍、1.7倍、1.8倍、1.9倍、2.0倍、2.1倍、2.2倍、2.3倍、2.4倍、2.6倍、2.7倍、2.8倍、2.9倍、3.0倍、3.1倍、3.2倍、3.3倍、3.4倍、3.5倍、3.6倍、3.7倍、3.8倍、3.9倍或4.0倍。

[0488] 在一个实施方案中,包含第一多肽(其包含FVIII蛋白和第一Ig恒定区或其一部分)和第二多肽(其包含VWF蛋白、具有小于288个氨基酸的XTEN以及Ig恒定区或其一部分)的嵌合蛋白展现的半衰期比包含相同第一多肽和无所述XTEN序列的第二多肽的相应嵌合蛋白或野生型FVIII高至少约2倍、2.5倍、3.0倍、3.5倍、4.0倍、4.5倍、5.0倍、5.5倍、6.0倍、7倍、8倍、9倍或10倍。在另一实施方案中,包含第一多肽(其包含FVIII蛋白和第一Ig恒定区或其一部分)和第二多肽(其包含VWF蛋白、具有小于288个氨基酸的XTEN以及Ig恒定区或其一部分)的嵌合蛋白展现的半衰期比包含相同第一多肽和无所述XTEN序列的第二多肽的相应嵌合蛋白或野生型FVIII高约2至约5倍、约3至约10倍、约5至约15倍、约10至约20倍、约15至约25倍、约20至约30倍、约25至约35倍、约30至约40倍、约35至约45倍。在特定实施方案中,在FVIII和VWF双重敲除小鼠中,本发明的嵌合蛋白的半衰期比野生型FVIII的半衰期增加高至少约30、31、32、33、34、35、36、37、38、39或40倍。

[0489] 在某些实施方案中,嵌合蛋白在小鼠中展现约40小时的半衰期。

[0490] 在一些实施方案中,嵌合蛋白的半衰期长于与内源性VWF缔合的FVIII的半衰期。在其它实施方案中,嵌合蛋白的半衰期是野生型FVIII或与内源性VWF缔合的FVIII蛋白的半衰期的至少约1.5倍、2倍、2.5倍、3.5倍、3.6倍、3.7倍、3.8倍、3.9倍、4.0倍、4.5倍或5.0倍。

[0491] 在一些实施方案中,作为本发明的结果,相较于无VWF蛋白的FVIII蛋白或野生型FVIII,嵌合蛋白的半衰期得以延长。本发明的嵌合蛋白的半衰期比无VWF蛋白的嵌合蛋白或野生型FVIII的半衰期长至少约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。在一个实施方案中,FVIII的半衰期比野生型FVIII的半衰期长约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,相较于野生型FVIII或无VWF蛋白的FVIII蛋白,FVIII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。在其它实施方案中,本发明的嵌合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40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。在其它实施方案中,本发明的嵌合蛋白的半衰期是约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。

[0492] 在一些实施方案中,本发明的嵌合蛋白根据受试者的平均半衰期是约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。

[0493] 此外,本发明提供一种治疗或预防流血疾病或病症的方法,其包括施用有效量的嵌合蛋白。在一个实施方案中,流血疾病或病症选自由以下组成的组:流血凝血病症、关节积血、肌肉流血、口腔流血、出血、向肌肉中出血、口腔出血、创伤、头创伤、胃肠流血、颅内出血、腹内出血、胸内出血、骨折、中枢神经系统流血、咽后隙中流血、腹膜后隙中流血和髂腰肌鞘中流血。在特定实施方案中,流血疾病或病症是甲型血友病。

[0494] 通过本发明制备的包含与本文所述的VWF蛋白(其防止或抑制FVIII蛋白与内源性VWF相互作用)组合的XTEN序列和Ig恒定区或其一部分的嵌合蛋白具有如将由本领域技术人员认识到的许多用途,包括但不限于治疗患有止血病症的受试者的方法和治疗需要一般性止血剂的受试者的方法。在一个实施方案中,本发明涉及一种治疗患有止血病症的受试者的方法,其包括施用治疗有效量的嵌合蛋白。

[0495] 嵌合蛋白中的FVIII蛋白部分通过在带负电荷磷脂表面上充当因子IX的辅因子,由此形成X酶复合物来治疗或预防止血病症。活化凝血因子结合磷脂表面使这个过程定位于血管损害部位。在磷脂表面上,因子VIIIa使由因子IXa达成的因子X活化的最大速度增加约200,000倍,从而导致凝血酶产生的大规模瞬间爆发。

[0496] 本发明的嵌合蛋白可用于治疗任何止血病症。可通过施用本发明的嵌合蛋白治疗的止血病症包括但不限于甲型血友病以及与因子VIII相关的缺乏症或结构异常。在一个实施方案中,止血病症是甲型血友病。

[0497] 本发明的嵌合蛋白可防治性用于治疗患有止血病症的受试者。本发明的嵌合蛋白可用于治疗患有止血病症的受试者的急性流血发作。在另一实施方案中,止血病症可为缺陷性凝结因子,例如范威尔邦德氏因子的结果。在一个实施方案中,止血病症是遗传病症。在另一实施方案中,止血病症是获得性病症。获得性病症可由潜伏继发性疾病或病状所致。无关病状可为例如但不限于癌症、自体免疫疾病或妊娠。获得性病症可由老龄或用以治疗潜伏继发性病症的药物疗法(例如癌症化学疗法)所致。

[0498] 本发明也涉及治疗不患有先天性止血病症,但患有导致获得止血病症(例如由于产生抗FVIII抗体或手术)的继发性疾病或病状的受试者的方法。因此,本发明涉及一种治疗需要一般性止血剂的受试者的方法,其包括施用治疗有效量的通过本发明方法制备的嵌合蛋白。

[0499] 本发明也涉及降低FVIII的免疫原性或诱导针对FVIII的较小免疫原性的方法,其包括施用有效量的本文所述的嵌合蛋白或其编码多核苷酸。

[0500] 在一个实施方案中,需要一般性止血剂的受试者正经受或即将经受手术。本发明的嵌合蛋白可在手术之前、期间或之后作为防治性方案加以施用。本发明的嵌合蛋白可在手术之前、期间或之后施用以控制急性流血发作。

[0501] 本发明的嵌合蛋白可用于治疗具有急性流血发作,未患有止血病症的受试者。急性流血发作可由严重创伤,例如手术、车祸、伤口、枪射击裂伤或导致流血不受控制的任何其它创伤事件所致。流血发作的非限制性实例包括流血凝血病症、关节积血、肌肉流血、口腔流血、出血、向肌肉中出血、口腔出血、创伤、头创伤、胃肠流血、颅内出血、腹内出血、胸内出血、骨折、中枢神经系统流血、咽后隙中流血、腹膜后隙中流血、髂腰肌鞘中流血及其任何组合。

[0502] 在防治性应用中,向尚未处于疾病状态中的患者施用一种或多种含有本发明的嵌合蛋白或其混合物的组合物以增强所述患者的抗性或减轻与疾病或病症相关的症状。这种量定义为“防治有效剂量”。在治疗性应用中,有时需要在相对较短间隔下的相对较高剂量(例如每剂约1至400mg/kg的多肽,其中5至25mg的剂量更通常用于放射免疫缀合物,并且更高剂量用于细胞毒素药物修饰的多肽)直至疾病进展降低或终止,以及直至患者显示部分或完全疾病症状改善。此后,可向患者施用防治性方案。

[0503] 在一些实施方案中,本发明的嵌合蛋白或组合物用于按需治疗,其包括治疗流血发作、关节积血、肌肉流血、口腔流血、出血、向肌肉中出血、口腔出血、创伤、头创伤(头部创伤)、胃肠流血、颅内出血、腹内出血、胸内出血、骨折、中枢神经系统流血、咽后隙中流血、腹膜后隙中流血或髂腰肌鞘中流血。受试者可需要手术防治、围手术期管理或手术治疗。所述手术包括例如小手术、大手术、拔牙、扁桃体切除术、腹链沟疝切开术、滑膜切除术、全部膝盖置换、开颅术、骨缝合术、创伤手术、颅内手术、腹内手术、胸内手术或关节置换手术。

[0504] 在一个实施方案中,本发明的嵌合蛋白是静脉内、皮下、肌肉内施用或通过任何粘膜表面施用,例如口服、舌下、经颊、经鼻、经直肠、经阴道或通过肺途径施用。本发明的包含VWF片段和FVIII蛋白的嵌合蛋白可植入在允许嵌合蛋白缓慢释放至流血部位的生物聚合物固体载体内或连接于所述生物聚合物固体载体,或植入绷带/敷料中。嵌合蛋白的剂量将视受试者和所用特定施用途径而变化。剂量可在0.1至100,000 $\mu$ g/kg体重的范围内。在一个实施方案中,剂量范围是0.1-1,000 $\mu$ g/kg。在另一实施方案中,剂量范围是0.1-500 $\mu$ g/kg。可连续或在特定时间间隔下施用蛋白质。可采用体外测定来确定最优剂量范围和/或施用时程。测量凝结因子活性的体外测定在本领域中是已知的,例如STA-CLOT VIIa-rTF凝结测定或ROTEM凝结测定。另外,可从由动物模型(例如血友病狗)获得的剂量-响应曲线外推有效剂量(Mount等2002,Blood 99(8):2670)。

[0505] 现已详细描述本发明,通过参照以下实施例将更明确了解本发明,所述实施例仅出于说明目的包括于此,并且不意图限制本发明。本文提及的所有专利、公布和文章都以引用的方式明确和特定并入本文。

## 实施例

[0506] 除非另外陈述,否则在整篇实施例中,使用以下材料和方法。

[0507] 材料和方法

[0508] 一般来说,除非另外指示,否则本发明的实施采用常规化学技术、生物物理学技术、分子生物学技术、重组DNA技术、免疫学技术(尤其例如抗体技术)和标准电泳技术。参见例如Sambrook,Fritsch和Maniatis,Molecular Cloning: Cold Spring Harbor Laboratory Press(1989);Antibody Engineering Protocols(Methods in Molecular

Biology), 510, Paul, S., Humana Pr (1996); Antibody Engineering: A Practical Approach (Practical Approach Series, 169), McCafferty 编, Irl Pr (1996); Antibodies: A Laboratory Manual, Harlow 等, CS. H. L. Press, Pub. (1999); 以及 Current Protocols in Molecular Biology, Ausubel 等编, John Wiley & Sons (1992)。

[0509] 实施例1: FVIII-XTEN-Fc/D' D3-XTEN-Fc 异二聚体

[0510] 本发明涉及产生一种通过IgG的Fc结构域偶联于范威尔邦德因子(VWF)蛋白的D' D3结构域的嵌合FVIII分子。连接的D' D3结构域防止FVIII与内源性VWF多聚体相互作用。这个分子充当合并其它半衰期延长技术以改进嵌合蛋白的药物动力学的平台。XTEN序列被并入FVIII B结构域中以及D' D3与Fc区之间以增加FVIII/VWF异二聚体的半衰期。

[0511] D' D3与Fc之间的凝血酶裂解位点允许在由凝血酶活化FVIII分子后释放D' D3结构域。

[0512] 实施例2: FVIII-XTEN-Fc/D' D3-Fc 异二聚体的质粒构建

[0513] 克隆VWF050-含IHH三重突变的VWF031

[0514] Fc中的IHH三重突变防止与FcRn相互作用,因此不存在通过FcRn路径使含有Fc的分子再循环。Fc中的3个突变是I253A、H310A、H435A。

[0515] 通过在RsRII限制位点与Not I限制位点之间用含有IHH三重突变的Fc片段交换VWF031质粒的Fc区来产生VWF050。

[0516] 克隆VWF057-克隆具有144 AE XTEN+35 aa凝血酶可裂解接头的VWF-Fc。

[0517] 寡聚物

[0518] ESC 155-针对VWF034中的144 AE XTEN的寡聚物-反向

[0519] CCCC GCC ACCG GAT CCCCC GCC ACCG GAT CCCCC GCC ACCG GAT CCCCC GCC ACCG GAACCTCCACCGC  
CGCTCGAGGCACCTTCTTCAGTGCTGGTGGGCGAGCCCGCTGGTGACCCCTTCCTC

[0520] ESC 156-针对VWF034中的144 AE XTEN-GS接头的寡聚物-反向

[0521] GGGGAAGAGGAAGACTGACGGTCCGCCAGGAGTTCTGGAGCTGGGCACGGTGGGCATGTGTGAGTTTT  
GTCGCCTCCGCTGCCCCGGGGACCAGGGATCCCCGCCACCGGATCCCCGCCACCGGATCCCCGCCACCGGATC  
CCCCGCC

[0522] ESC 157-针对VWF031中的144 AE XTEN的寡聚物-正向

[0523] GTGAAGCCTGCCAGGAGCCGATATCGGGCGCGCCAACATCAGAGAGCGCCACCCCTGAAAGTGGTCCCG  
GGAGCGAGCCAGC

[0524] 进行两次PCR以获得144 AE-XTEN+具有凝血酶裂解位点的35aa GS接头。

[0525] 使用作为模板的144-AE XTEN编码DNA以及ESC 157/ESC 155引物对进行第一PCR反应。由这个反应获得的约550bp长的PCR产物用作第二PCR反应的模板,并且使用ESC 157/156引物对加以扩增。这个反应产生约700bp长的产物。接着用EcoRV-HF和RsRII消化这个700bp PCR产物和VWF034质粒。消化质粒骨架。

[0526] 接着使用VWF034连接700bp PCR产物。

[0527] 克隆VWF058-VWF034中的IHH三重突变

[0528] Fc中的IHH三重突变防止与FcRn相互作用,因此不存在通过FcRn路径使含有Fc的分子再循环。Fc中的3个突变是I253A、H310A、H435A。

[0529] 通过在RsRII限制位点与Not I限制位点之间用含有IHH三重突变的Fc片段交换

VWF034质粒的Fc区来产生VWF058。

[0530] 克隆FVIII-263-具有IHH三重突变的FVIII 205

[0531] Fc中的IHH三重突变防止与FcRn相互作用,因此不存在通过FcRn路径使含有Fc的分子再循环。Fc中的3个突变是I253A、H310A、H435A。

[0532] 通过在RsRII限制位点与Not 1限制位点之间用含有IHH三重突变的Fc片段交换FVIII 205质粒的Fc区来产生FVIII-263。

[0533] 克隆FVIII-282-在B结构域中具有144 AE XTEN的FVIII-Fc

[0534] ESC 158-针对B结构域中的144 AE XTEN的寡聚物-正向

[0535] AAGAAGCTTCTCTCAAAACGGCGCGCAACATCAGAGAGCGCCACCCCTGAAAGTGGTCCCGGGAGCGA  
GCCAGCCACATCTGGGTCGAAACGCCAGGC

[0536] ESC 159-针对B结构域中的144 AE XTEN的寡聚物-反向

[0537] GGTATCATCATAATCGATTTCTCTTGATCTGACTGAAGAGTAGTACGAGTTATTTTCAGCTTGATGGCG  
TTTCAAGACTGGTGGGCTCGAGGCACCTTCTTCAGTGCTGGTGGGCGAGCCCGCTGGTGACCCTTCTCAGTGGACG  
TAGG

[0538] 使用作为模板的144-AE XTEN编码DNA以及ESC 158/ESC 159引物对进行第一PCR反应。接着用AscI和C1a1消化由这个反应获得的约550bp长的PCR产物和FVIII 169质粒。由消化的FVIII 169获得的质粒骨架接着用于连接550bp PCR产物以获得FVIII 282。

[0539] 克隆FVIII-283-具有IHH三重突变的FVIII 169

[0540] Fc中的IHH三重突变防止与FcRn相互作用,因此不存在通过FcRn路径使含有Fc的分子再循环。Fc中的3个突变是I253A、H310A、H435A。

[0541] 通过在RsRII限制位点与Not 1限制位点之间用含有IHH三重突变的Fc片段交换FVIII 169质粒的Fc区来产生FVIII-283。

[0542] 实施例3:在HEK293细胞中产生FVIII-XTEN-Fc/D' D3-XTEN-Fc

[0543] 图2.显示FVIII-XTEN-Fc/D' D3-XTEN-Fc构建体的表达的示意图。使用聚乙烯亚胺(PEI)在HEK293细胞中进行三质粒共转染。第一质粒获得FVIII-XTEN-Fc的表达,第二质粒表达D1D2D' D3-XTEN-Fc,并且第三质粒表达PACE/弗林蛋白酶,其为从D1D2D' D3-XTEN-Fc酶促移除原肽(即D1D2结构域)所需。这个三质粒表达系统的产物包括FVIII-XTEN-Fc/D' D3-XTEN-Fc异二聚体、D' D3-XTEN-Fc同二聚体和微量FVIII-XTEN-Fc半合子样貌物质。

[0544] 实施例4:纯化FVIII-XTEN-Fc/D' D3-XTEN-Fc异二聚体

[0545] 为纯化FVIII-XTEN-Fc/D' D3-XTEN-Fc异二聚体,切向流过滤(TFF)步骤用于使条件培养基首先浓缩10倍。接着依次使用亲和色谱法和脱盐柱来进一步纯化滤液中的产物。根据HPLC-SEC,分子的纯度是可接受的,并且通过蛋白质印迹加以进一步确认。分子的比活性与B结构域缺失的FVIII类似,如通过FVIII活性测定(实施例5)和OD280测量所测量。

[0546] 实施例5:FVIII-XTEN-Fc/D' D3-XTEN-Fc异二聚体的比活性

[0547] 通过FVIII显色测定和活化部分凝血激素时间(aPTT)测定来测量FVIII-XTEN-Fc/D' D3-XTEN-Fc异二聚体的活性。SQ BDD-FVIII、rFVIII169/VWF034和rFVIII169/VWF057的比显色活性和比aPTT活性列于表16中。相较于SQ BDD-FVIII,对于rFVIII169/VWF034和rFVIII169/VWF057,我们已观察到类似比显色活性以及比aPTT活性降低60%。

[0548] 表16:异二聚体变体的比活性



	FVIII	SQ BDD-FVIII	rFVIII169/V WF034	rFVIII160/V WF057
[0549]	比显色活性 (IU/pmol)	0.9 – 2.0	1.1-1.2	0.8-1.6
	比 aPTT 活性 (IU/pmol)	0.75-1.7	0.4	0.3-0.6

[0550] FVIII显色测定

[0551] 使用来自DiaPharma的COATEST SP FVIII试剂盒(产品号:K824086)测量FVIII活性,并且所有孵育都在振荡下在37℃板加热器上进行。

[0552] 凝血因子VIII:C的WHO第8国际标准物(浓缩物,代号07/350)用作测定标准物,所述标准物的范围是100mIU/mL至0.78mIU/mL。一式两份添加汇合的正常人血浆测定对照和测试样品(用1X Coatest缓冲液稀释)至Immulon 2HB 96孔板中(25μL/孔)。依序添加新鲜制备的IXa/FX/磷脂混合物(50μL)、25μL 25mM CaCl<sub>2</sub>和50μL FXa底物至各孔中,其中在各次添加之间孵育5分钟。在与底物一起孵育之后,添加25μL 20%乙酸以终止显色反应,并且用SpectraMAX plus (Molecular Devices) 仪器测量OD405的吸光度。用SoftMax Pro软件(5.2版)分析数据。最低定量水平(LLoQ)是7.8mIU/mL。

[0553] FVIII aPTT测定

[0554] 如下在Sysmex CA-1500凝血分析器上进行FVIII aPTT测定:首先,通过仪器添加50uL含手动稀释的样品、标准物和对照的aPTT缓冲液(50mM Tris、100mM NaCl、1% HSA,pH 7.4)至反应杯中,随后添加50uL缺乏FVIII的血浆(George King Bio-Medical,产品号:0800)。在37℃下孵育1分钟之后,添加50uL aPTT试剂(Actin®FSL活化脑透明质试剂-Dade Behring,参考号B4219-2)至反应混合物中,并且在37℃下孵育4分钟。随后,添加50uL 20mM CaCl<sub>2</sub>(Dade Behring,参考号ORF037),并且将反应杯立刻转移至四个分光光度计通道位置中的一个中以测量混合物中折射光的量,其通过仪器的软件算法被转换成凝结的发生。报道的凝结时间是从添加CaCl<sub>2</sub>开始直至发生凝块形成的时长。通过以100mIU/ml至0.78mIU/ml的范围将WHO第8国际FVIII标准物稀释于aPTT缓冲液中来产生测定标准物。在MS Excel中将标准曲线绘制为作为Y轴的凝结时间(以秒计)相对于作为X轴的FVIII活性对数(以10为底)(mIU/mL),并且使用针对这个标准曲线的线性回归线的公式计算个别样品的活性。基于测定性能,量化下限(LLoQ)是7.8mIU/mL。

[0555] 实施例6:XTEN插入物对异二聚体的半衰期延长的累加作用

[0556] 将XTEN插入物并入异二聚体中以达成半衰期延长。在FVIII B结构域处插入单一288氨基酸(aa)AE-XTEN导致异二聚体在HemA小鼠中的半衰期是16.7小时,如图3中通过rFVIII169/VWF031所证明。为进一步改进异二聚体的半衰期,将长度是144 aa或288 aa的第二XTEN插入物分别在FVIII A1结构域中或紧靠D' D3片段的下游并入FVIII169/VWF031中,异二聚体变体被命名为FVIII205/VWF031和FVIII169/VWF034。

[0557] 通过在200IU/kg剂量下单次静脉内施用测试分子来在缺乏FVIII的(HemA)小鼠中评估rFVIII169/VWF031、rFVIII205/VWF031和rFVIII169/VWF034的半衰期。在如图3中指示的指定时间点收集血浆样品,通过FVIII显色测定来测定样品的FVIII活性,使用

WinNonlin-Phoenix程序计算PK参数并列于表17中。

[0558] 如图3和表17中所示,在FVIII的A1结构域处或在D' D3的下游添加第二XTEN插入物分别进一步改进异二聚体的半衰期至29.45或31.10。此外,也由于两种XTEN插入物而观察到清除率和AUC改进大于2倍。

[0559] 表17:异二聚体在HemA小鼠中的PK参数

FVIII	XTEN 插入物		$T_{1/2}$ (h)	MRT (h)	Cl (mL/h/kg)	Vss (mL/kg)	AUC <sub>D</sub> (kg*h/mL)
	插入物 1	插入物 2					
[0560] rFVIII169/VWF031	B*-AE2 88		16.6 5	18.44	3.57	85.72	0.28
rFVIII205/VWF031	B*-AE2 88	A1-AE 144	29.4 5	36.02	1.76	63.56	0.57
rFVIII169/VWF034	B*-AE2 88	D'D3- AE288	31.1 0	34.57	1.73	59.77	0.58

[0561] 实施例7:当插入在D' D3与Fc结构域之间时,144 aa AE-XTEN比288 aa AE-XTEN赋予更佳半衰期益处。

[0562] 构建另一异二聚体-FVIII169/VWF057以试图鉴定XTEN插入物在D' D3-XTEN-Fc链内的最优长度,其中使XTEN插入物的长度从288 aa降低至144 aa。如图4中所示,相较于rFVIII169/VWF034,rFVIII169/VWF057的半衰期从31小时增加至42小时。也观察到rFVIII169/VWF057的清除率和AUC改进,数据列于表18中。因此,当插入在FVIII-XTEN-Fc/D' D3-XTEN-Fc异二聚体的D' D3与Fc结构域之间时,144 aa AE-XTEN插入物比AE-288 aa XTEN更优。

[0563] 表18:rFVIII169/VWF034和rFVIII169/VWF057在HemA小鼠中的PK参数

FVIII	$T_{1/2}$ (h)	MRT (h)	Cl (mL/h/kg)	Vss (mL/kg)	AUC <sub>D</sub> (kg*h/mL)
[0564] rFVIII169/VWF034	31.10	34.57	1.73	59.77	0.58
[0565] rFVIII169/VWF057	42.23	53.24	0.97	51.44	1.03

[0566] 实施例8:Fc结构域延长异二聚体的半衰期

[0567] Fc结构域通过FcRn介导的再循环路径来延长它的融合蛋白的半衰期。为确认Fc结构域对异二聚体的半衰期延长的必要性,用三重突变体(I253A/H310A/H435A;IHH)置换rFVIII205/VWF031中的野生型Fc结构域以形成rFVIII263/VWF050,并且通过表面等离子体共振(Biacore)测定确认rFVIII263/VWF050的FcRn结合被完全消除。相较于rFVIII205/VWF031在HemA小鼠中评估FVIII263/VWF050的半衰期。观察到rFVIII263/VWF050的清除率增加以及半衰期和AUC降低,如图5和表19中所示。这个结果证明除确保FVIII和D' D3共价结合之外,Fc结构域也为异二聚体的半衰期改进所必需。

[0568] 表19:rFVIII205/VWF031和rFVIII263/VWF040在HemA小鼠中的PK参数

	<b>FVIII</b>	<b>Fc 结构域中的突变</b>	<b>T<sub>1/2</sub> (h)</b>	<b>MRT (h)</b>	<b>Cl (mL/h/kg)</b>	<b>Vss (mL/kg)</b>	<b>AUC_D (kg*h/mL)</b>
[0569]	rFVIII205/VWF031	无	29.4 5	36.02	1.76	63.56	0.57
	rFVIII263/VWF050	IHH	22.9 6	26.15	2.36	61.69	0.42

[0570] 实施例9:FVIII-XTEN-Fc/D' D3-XTEN-Fc异二聚体在HemA小鼠尾部钳夹流血模型中的急性功效

[0571] 使用HemA小鼠尾部钳夹流血模型评估先导异二聚体候选物的急性功效。

[0572] 将8-12周龄雄性HemA小鼠随机分入4个治疗组中,并且分别以单次静脉内施用SQ BDD-FVIII、rFVIII169/VWF034、rFVIII169/VWF057或媒介物溶液治疗。为模拟FVIII的阵发性治疗(以恢复正常FVIII血浆水平的50-100%),所选FVIII治疗剂量是如通过FVIII aPTT活性衡量的75IU/kg。在这个剂量水平下,所有测试FVIII变体都将在给药后5分钟恢复正常鼠类血浆FVIII活性的约70%。

[0573] 将研究中各个别动物的失血体积绘制于图6中。相较于媒介物治疗的动物,观察到所有FVIII治疗组的失血体积都显著降低。在三个FVIII治疗组内,未发现失血降低有统计显著差异,从而表明异二聚体分子关于按需治疗可潜在地与SQ BDD-FVIII同样有效。

[0574] 将研究中各个别动物的失血体积绘制于图6中。相较于媒介物治疗的动物,观察到所有FVIII治疗组的失血体积都显著降低。在三个FVIII治疗组内,未发现失血降低有统计显著差异,从而表明异二聚体分子关于按需治疗可潜在地与SQ BDD-FVIII同样有效。

[0575] 此外,用较低剂量(37.5IU/kg)的rBDD-FVIII或rFVIII169/VWF034治疗HemA小鼠,并且结果显示于图6小图B中。与75IU/kg剂量相同,在尾部钳夹损伤后,rFVIII169/VWF034对HemA小鼠提供与BDD-FVIII类似的保护,从而指示所述分子在约35%的正常鼠类循环FVIII水平下仍然有效治疗HemA小鼠的严重流血发作。

[0576] 如下进行尾部钳夹程序。简要来说,在尾部损伤之前用50mg/kg氯胺酮(Ketamine)/0.5mg/kg右旋美托咪啉(Dexmedetomidine)混合物使小鼠麻醉,并且放置在37°C加热板上以帮助维持体温。接着使小鼠的尾部在37°C盐水中浸渍10分钟以扩张侧静脉。在静脉扩张之后,通过尾部静脉注射FVIII变体或媒介物溶液,并且接着在给药后5分钟使用直边11号解剖刀切断远端5mm尾部。持续30分钟将流出的血液收集至13ml 37°C盐水中,并且通过血液收集管的重量变化来测定失血体积:失血体积=(收集管结束重量-开始重量+0.10)ml。使用t检验(曼惠特尼检验(Mann Whitney test))和单因素ANOVA(克鲁斯凯-沃利斯检验(KRUSKAL-Wallis test),事后检验:邓氏多重比较检验(Dunns multiple comparison test))进行统计分析。

[0577] 实施例10:FVIII-XTEN-Fc/D' D3-XTEN-Fc异二聚体在HemA小鼠尾部静脉横切流血模型中的防治性功效

[0578] 在HemA小鼠尾部静脉横切(TVT)模型中测试FVIII169/VWF057的防治性功效。TVT模型通过将损伤引入小鼠尾部的侧静脉中来诱发流血,此模拟患有血友病流血病症的患者

的自发性流血发作。

[0579] 将8-10周龄雄性HemA小鼠随机分入四个治疗组中,并且在尾部静脉损伤之前72小时用FVIII169/VWF057治疗,或在损伤之前24小时或48小时用SQ BDD-FVIII治疗。媒介物治疗的动物用作阴性对照。将损伤后24小时内的再流血或归因于过度失血的安乐死的事件绘制于图7中。

[0580] 如图7中所示,不同于在TVT之前48小时用SQ BDD-FVIII治疗的小鼠(其中在损伤后观察到仅仅有限保护),相较于在TVT之前24小时接受SQ BDD-FVIII治疗的小鼠,在尾部损伤之前72小时接受rFVIII169/VWF057的小鼠具有关于再流血和存活的类似保护,从而指示在TVT模型中,rFVIII169/VWF057可对HemA小鼠提供长久至少3倍或更多倍(例如4倍)的保护。因此,rFVIII169/VWF057可能显著降低当前FVIII防治的治疗频率。

[0581] 类似地,在尾部静脉损伤之前24或96小时用FVIII-XTEN-Fc/D' D3-XTEN-Fc异二聚体:rFVIII169/VWF034和rFVIII169/VWF057治疗HemA小鼠。将治疗的再流血和存活数据与通过rBDD-FVIII(在损伤之前24或48小时)和媒介物获得的数据进行比较。尽管在尾部静脉损伤之前24小时用rBDD-FVIII治疗的小鼠的再流血类似于用媒介物治疗的小鼠,但在损伤之前24小时用异二聚体治疗的小鼠的再流血数据显著好于媒介物治疗组。此外,在损伤之前96小时用异二聚体治疗的小鼠的再流血数据与在损伤之前24小时接受rBDD-FVIII的小鼠类似。就TVT损伤后24小时的存活率而言,与用rBDD-FVIII治疗的小鼠的存活率小于50%相对比,当在损伤之前24小时施用FVIII分子时,在FVIII-XTEN-Fc/D' D3-XTEN-Fc异二聚体治疗下,大于90%的小鼠存活于TVT损伤。此外,当相较于在损伤之前24小时接受rBDD-FVIII治疗的小鼠时,在尾部静脉损伤之前96小时用FVIII-XTEN-Fc/D' D3-XTEN-Fc异二聚体治疗的小鼠的存活更佳(在rFVIII169/VWF034的情况下)或类似(在rFVIII169/VWF057的情况下)。再流血数据与存活数据两者均已指示相对于rBDD-FVIII治疗,FVIII-XTEN-Fc/D' D3-XTEN-Fc异二聚体治疗的功效延长4倍。

[0582] HemA小鼠尾部静脉横切模型

[0583] 如下进行尾部静脉横切程序。用含有50mg/kg氯胺酮、0.125mg/kg右旋美托咪啶和0.1mg/kg布诺啡(Buprenex)的混合物使小鼠麻醉。在足够麻醉深度处,用直边11号手术刀片在尾部的直径是约2.7mm所处的区域处横切小鼠的侧向尾部静脉。用温热盐水洗去流血以确保清晰观察伤口。接着持续下一24小时将治疗的小鼠单独舍饲在具有白色纸垫料的清洁笼中。每小时观察并记录尾部再流血和小鼠的身体活动直至尾部损伤后12小时。立刻使濒死小鼠安乐死,并且在尾部损伤后24小时进行最终观察。为模拟血友病患者的流血情况以及为确保动物从麻醉完全恢复,给予1mg/kg阿替美唑(Atipamezole)溶液以逆转在尾部静脉横切开始时的右旋美托咪啶作用。在12小时观察时期结束时另外施用0.1mg/kg剂量的布诺啡以用于过夜疼痛管理。产生时间相对于再流血以及时间相对于安乐死的存活曲线以进行数据分析,并且对数秩(曼特尔-考克斯(Mantel-COX))检验用于统计评估。

[0584] 实施例11:制备FVIII169/VWF059和其它构建体

[0585] pSYN FVIII 310克隆:

[0586] 以商业手段制备侧接有在N末端的BamH1位点和在C末端的Cla1位点的合成DNA片段。这个合成DNA用于置换pSYN FVIII 169构建体(SEQ ID NO:155)中的BamH1至Cla1区域。合成DNA与pSYN FVIII 169DNA两者均用BamH1和Cla1双重消化,将消化的合成DNA插入消化

的pSYN FVIII 169中以产生pSYN FVIII 310(SEQ ID NO:168;表20)。

[0587] 克隆pSYN FVIII 312:

[0588] 以商业手段制备侧接有在N末端的BamH1位点和在C末端的Afe 1位点的合成DNA片段。这个合成DNA用于置换pSYN FVIII 169构建体(SEQ ID NO:155)中的BamH1至Afe1区域。合成DNA与pSYN FVIII 169DNA两者均用BamH1和Afe1双重消化,将消化的合成DNA插入消化的pSYN FVIII 169中以产生pSYN FVIII 312(SEQ ID NO:169;表20)。从pSYN FVIII312产生pSYN FVIII 312A(SEQ ID NO:2;表20)以移除在FVIII和XTEN的接合点处的编码氨基酸残基GAP的AscI位点。

[0589] 表20:合成FVIII构建体。

构建体	蛋白质序列
[0590] pSYN FVIII 169	<p><u>PRSF</u><u>SQNG</u><u>APGT</u><u>SESAT</u><u>PESG</u><u>PGSE</u><u>PATSG</u><u>SETPGT</u>  <u>SESAT</u><u>PESG</u><u>PGSE</u><u>PATSG</u><u>SETPGT</u><u>SESAT</u><u>PESG</u><u>PGTS</u>  <u>TEP</u><u>SEGS</u><u>APG</u><u>SPAG</u><u>SPTST</u><u>EEGT</u><u>SESAT</u><u>PESG</u><u>PGSE</u>  <u>PATSG</u><u>SETPGT</u><u>SESAT</u><u>PESG</u><u>PGSP</u><u>AGSPT</u><u>STEEG</u><u>SPA</u>  <u>GSPT</u><u>STEEGT</u><u>STEP</u><u>SEGS</u><u>APGT</u><u>SESAT</u><u>PESG</u><u>PGTSE</u>  <u>SAT</u><u>PESG</u><u>PGTSE</u><u>SESAT</u><u>PESG</u><u>PGSE</u><u>PATSG</u><u>SETPG</u><u>SEPA</u>  <u>TSG</u><u>SETPG</u><u>SPAG</u><u>SPTST</u><u>EEGT</u><u>STEP</u><u>SEGS</u><u>APGT</u><u>STE</u>  <u>P</u><u>SEGS</u><u>APG</u><u>SEP</u><u>PATSG</u><u>SETPGT</u><u>SESAT</u><u>PESG</u><u>PGTSE</u>  <u>P</u><u>SEGS</u><u>APASS</u><u>PPVL</u><u>KRHQAE</u><u>ITR</u> (SEQ ID NO: 167)            (加下划线=XTEN 残基; 未加下划线=FVIII 残基)</p>
pSYN FVIII 310	<p><u>PRSF</u><u>G</u><u>APGT</u><u>SESAT</u><u>PESG</u><u>PGSE</u><u>PATSG</u><u>SETPGT</u><u>SES</u>  <u>AT</u><u>PESG</u><u>PGSE</u><u>PATSG</u><u>SETPGT</u><u>SESAT</u><u>PESG</u><u>PGTSTE</u>  <u>P</u><u>SEGS</u><u>APG</u><u>SPAG</u><u>SPTST</u><u>EEGT</u><u>SESAT</u><u>PESG</u><u>PGSEPA</u>  <u>TSG</u><u>SETPGT</u><u>SESAT</u><u>PESG</u><u>PGSP</u><u>AGSPT</u><u>STEEG</u><u>SPAG</u>  <u>SPT</u><u>STEEGT</u><u>STEP</u><u>SEGS</u><u>APGT</u><u>SESAT</u><u>PESG</u><u>PGTSES</u>  <u>AT</u><u>PESG</u><u>PGTSES</u><u>AT</u><u>PESG</u><u>PGSE</u><u>PATSG</u><u>SETPG</u><u>SEPAT</u></p>

	<p><u>SGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEP</u>  <u>SEGSAPGSEPATSGSETPGTSESATPESGPGTSTEP</u>  <u>SEGSAPASSEITR (SEQ ID NO: 168)</u>                  (加下划线=XTEN 残基; 未加下划线=FVIII 残基)</p>
<p>[0591]</p>	<p>pSYN FVIII 312</p> <p><u>PRFSQNGAPGTSESATPESGPGSEPATSGSETPGT</u>  <u>SESATPESGPGSEPATSGSETPGTSESATPESGPGTS</u>  <u>TEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSE</u>  <u>PATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPA</u>  <u>GSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSE</u>  <u>SATPESGPGTSESATPESGPGSEPATSGSETPGSEPA</u>  <u>TSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTE</u>  <u>PSEGSAPGSEPATSGSETPGTSESATPESGPGTSTE</u>  <u>PSEGSAPASSEITR (SEQ ID NO: 169)</u>                  (加下划线=XTEN 残基; 未加下划线=FVIII 残基)</p>
	<p>pSYN FVIII 312A</p> <p><u>PRFSQNGTSESATPESGPGSEPATSGSETPGTSES</u>  <u>ATPESGPGSEPATSGSETPGTSESATPESGPGTSTE</u>  <u>PSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPA</u>  <u>TSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAG</u>  <u>SPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSES</u>  <u>ATPESGPGTSESATPESGPGSEPATSGSETPGSEPAT</u>  <u>SGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEP</u>  <u>SEGSAPGSEPATSGSETPGTSESATPESGPGTSTEP</u>  <u>SEGSAPASSEITR (SEQ ID NO: 2)</u>                  (序列与 pSYN FVIII312 类似, 只是对应于 AscI 位点的残基(即 GAP)被移除)(加下划线=XTEN 残基; 未加下划线=FVIII 残基)</p>

[0592] 克隆pSYN VWF059和VWF073:

[0593] 制备编码在D' D3-XTEN与Fc之间的不同接头区域的各种合成DNA片段。这些合成DNA片段侧接有在N末端的Asc1位点和在C末端的Not 1位点。这些合成DNA用于置换pSYN VWF057构建体(SEQ ID NO:152)中的Asc1至Not1区域。pSYN VWF059构建体(表21)包含接头区域(SEQ ID NO:13),其包括整个FVIII酸性区2(a2)。这个位点被报道由凝血酶裂解,并且在FVIII活化后,D' D3XTEN被释放。pSYN VWF073构建体(表21)仅含有FVIII酸性区2(a2)的凝血酶裂解位点(即IEPRFS) (SEQ ID NO:23)。合成DNA与pSYNVWF057 DNA两者均用Asc1和Not1双重消化。将消化的合成DNA插入消化的pSYN VWF057中以产生pSYN VWF059和pSYN VWF073。通过移除EcoRV限制位点来从pSYN VWF059产生pSYN VWF59A构建体(表21)。通过在HEK293细胞中共同表达FVIII169和VWF057或VWF059来产生FVIII169/VWF057和FVIII169/VWF059异二聚体蛋白。

[0594] 表21:合成VWF构建体-可裂解接头区域

构建体	蛋白质序列
pSYN VWF057	<u>TSTEEGASSGGGGSGGGGSGGGGSGGGGSGGGG</u> <u>SGGGGSGGGGSLVPRGSGGDKTH</u> (SEQ ID NO: 12)  斜体和加下划线序列显示 GS 接头和 LVPR 凝血酶裂解位点(也被阴影化)。
pSYN VWF059	<u>TSTEEGASISDKNTGDYYEDSYEDISAYLLSKNNAIE</u> <u>PRSFSDKTH</u> (SEQ ID NO: 13)  斜体和加下划线序列显示来自 FVIII 酸性区 2 (a2) 的 32 aa。阴影化序列显示 pSYN VWF059A 中使用的凝血酶裂解位点。
[0595] pSYN VWF059A	<u>TSTEEGASSDKNTGDYYEDSYEDISAYLLSKNNAIE</u> <u>PRSFSDKTH</u> (SEQ ID NO: 22)  斜体和加下划线序列显示来自 FVIII 酸性区 2 (a2) 的 32 aa。这个序列是与 VWF059 类似的序列，例外之处是对应于 EcoRV 位点的残基(即 IS)被移除。
pSYN VWF073	<u>TSTEEGASSGGGGSGGGGSGGGGSGGGGSGGGG</u> <u>SGGGGSI EPRFSGSGGDKTH</u> (SEQ ID NO: 23)  斜体和加下划线序列显示具有来自 FVIII 酸性区 2 的截短凝血酶裂解位点(阴影化 7 个氨基酸 -IEPRFS)的 GS 接头。

[0596] 实施例12:凝血酶消化FVIII异二聚体以分析D'D3从Fc释放

[0597] 在凝血酶消化实验中测试两种FVIII异二聚体蛋白,并且考查由凝血酶对它们的裂解速率。这个实验中使用的两种异二聚体构建体是FVIII169/VWF057异二聚体和FVIII169/VWF059异二聚体,连同FVIIIIFc一起。以上描述FVIII169/VWF057和FVIII169/VWF059异二聚体。进行三个消化反应:i) FVIIIIFc ii) FVIII169/VWF057(图11),和iii) FVIII 169/VWF059(图12)。以FVIII:凝血酶的摩尔比约22:1,用人 $\alpha$ -凝血酶处理测试样品。在37°C水浴中孵育各反应。在各指示的时间点( $t=5, 15, 30, 45, 60$ 分钟),取出22.5 $\mu$ L样品,用22.5 $\mu$ L非还原性2x SDS装载染料终止,并且加热3分钟。接着在SDS-PAGE凝胶上操作消化的蛋白质。利用LICOR系统,使用抗FVIII重链(GMA012)和抗VWF-D3(Ab96340)抗体进行蛋白质印迹。

[0598] 如图11中所示,FVIII169/VWF057暴露于凝血酶导致D'D3-XTEN-Fc的检测水平逐渐降低,同时关联有裂解产物D'D3-144 XTEN的水平增加。在15分钟之后剩余有未裂解的

FVIII169/VWF057。相反,图12显示FVIII 169/VWF059由凝血酶更快速裂解,如通过在5分钟之后有少许乃至无可检测的未裂解FVIII 169/VWF059所证实。因此,FVIII 169/VWF059显示相较于FVIII169/VWF057,在凝血酶活化后,D' D3从Fc更佳释放。

[0599] 进行平行实验以使用质谱分析(MS)探究凝血酶裂解。根据MS,FVIII 169/VWF059再次显示相较于VWF057,D' D3从Fc更佳释放。

[0600] 实施例13:FVIII169/VWF059在HemA小鼠中的体内评估

[0601] 为进一步评估FVIII169/VWF059的药物动力学概况和体内效价,通过在150IU/kg剂量下静脉内施用来用FVIII169/VWF059治疗HemA小鼠。在注射后5分钟、24、48、72、96和120小时通过腔静脉血液收集来收集血浆样品。通过FVIII显色测定来测量血浆样品中的FVIII活性,并且使用Phoenix程序计算PK参数。如表22中所示,相较于FVIII169/VWF057,观察到FVIII169/VWF059的类似PK概况,从而指示a2凝血酶裂解接头对异二聚体的PK概况不具有负面影响。

[0602] 表22:FVIII169/VWF057和FVIII169/VWF059在HemA小鼠中的PK概况

异二聚体	T <sub>1/2</sub> (h)	AUC/D (h*kg*mIU /mL/mIU)	Cl (mL/h/kg)	MRT (h)	Vss (mL/kg)
FVIII169/VWF 057	38.53	0.80	1.26	44.92	56.38
FVIII169/VWF 059	40.51	0.74	1.35	49.22	66.26

[0604] 在HemA小鼠尾部钳夹模型(描述于实施例9中)中相较于野生型BDD-FVIII评估FVIII169/VWF059的急性功效。用75IU/kg的FVIII169/VWF059或BDD-FVIII治疗HemA小鼠,并且将各实验小鼠的失血体积绘制于图13中。相较于BDD-FVIII,FVIII169/VWF059对HemA小鼠提供相同保护程度(p=0.9883),从而指示FVIII169/VWF059在体内具有完全功能性。

[0605] FVIII-XTEN-Fc/D' D3-Fc异二聚体的质粒构建

[0606] VWF031核苷酸序列(SEQ ID NO:147)



[0607]

```

1   ATGAT TCCTG CCAGA TTTGC CGGGG TGCTG CTTGC TCTGG CCCTC ATTTT
51  GCCAG GGACC CTTTG TGCAG AAGGA ACTCG CGGCA GGTC A TCCAC GGCCC
101 GATGC AGCCT TTTGC GAAGT GACTT CGTCA ACACC TTTGA TGGGA GCATG
151 TACAG CTTTG CGGGA TACTG CAGTT ACCTC CTGGC AGGGG GCTGC CAGAA
201 ACGCT CCTTC TCGAT TATTG GGGAC TTCCA GAATG GCAAG AGAGT GAGCC
251 TCTCC GTGTA TCTTG GGGAA TTTTT TGACA TCCAT TTGTT TGTCA ATGGT
301 ACCGT GACAC AGGGG GACCA AAGAG TCTCC ATGCC CTATG CCTCC AAAGG
351 GCTGT ATCTA GAAAC TGAGG CTGGG TACTA CAAGC TGTCC GGTGA GGCCT
401 ATGGC TTTGT GGCCA GGATC GATGG CAGCG GCAAC TTTCA AGTCC TGCTG
451 TCAGA CAGAT ACTTC AACAA GACCT GCGGG CTGTG TGGCA ACTTT AACAT
501 CTTTG CTGAA GATGA CTTTA TGACC CAAGA AGGGA CCTTG ACCTC GGACC
551 CTTAT GACTT TGCCA ACTCA TGGGC TCTGA GCAGT GGAGA ACAGT GGTGT
601 GAACG GGCAT CTCCT CCCAG CAGCT CATGC AACAT CTCCT CTGGG GAAAT
651 GCAGA AGGGC CTGTG GGAGC AGTGC CAGCT TCTGA AGAGC ACCTC GGTGT
701 TTGCC CGCTG CCACC CTCTG GTGGA CCCCC AGCCT TTTGT GGCCC TGTGT
751 GAGAA GACTT TGTGT GAGTG TGCTG GGGGG CTGGA GTGCG CCTGC CCTGC
801 CCTCC TGGAG TACGC CCGGA CCTGT GCCCA GGAGG GAATG GTGCT GTACG
851 GCTGG ACCGA CCACA GCGCG TGCAG CCCAG TGTGC CCTGC TGGTA TGGAG
901 TATAG GCAGT GTGTG TCCCC TTGCG CCAGG ACCTG CCAGA GCCTG CACAT
951 CAATG AAATG TGTCA GGAGC GATGC GTGGA TGGCT GCAGC TGCCC TGAGG
1001 GACAG CTCCT GGATG AAGGC CTCTG CGTGG AGAGC ACCGA GTGTC CCTGC
1051 GTGCA TTCCG GAAAG CGCTA CCCTC CCGGC ACCTC CCTCT CTCGA GACTG
1101 CAACA CCTGC ATTTG CCGAA ACAGC CAGTG GATCT GCAGC AATGA AGAAT
1151 GTCCA GGGGA GTGCC TTGTC ACTGG TCAAT CCCAC TTCAA GAGCT TTGAC
1201 AACAG ATACT TCACC TTCAG TGGGA TCTGC CAGTA CCTGC TGGCC CGGGA

```

1251 TTGCC AGGAC CACTC CTTCT CCATT GTCAT TGAGA CTGTC CAGTG TGCTG  
1301 ATGAC CGCGA CGCTG TGTGC ACCCG CTCCG TCACC GTCCG GCTGC CTGGC  
1351 CTGCA CAACA GCCTT GTGAA ACTGA AGCAT GGGGC AGGAG TTGCC ATGGA  
1401 TGGCC AGGAC ATCCA GCTCC CCCTC CTGAA AGGTG ACCTC CGCAT CCAGC  
1451 ATACA GTGAC GGCCT CCGTG CGCCT CAGCT ACGGG GAGGA CCTGC AGATG  
1501 GACTG GGATG GCCGC GGGAG GCTGC TGGTG AAGCT GTCCC CCGTC TATGC  
1551 CGGGA AGACC TGC GG CCTGT GTGGG AATTA CAATG GCAAC CAGGG CGACC  
1601 ACTTC CTTAC CCCCT CTGGG CTGGC GGAGC CCCGG GTGGA GGACT TCGGG  
1651 AACGC CTGGA AGCTG CACGG GGACT GCCAG GACCT GCAGA AGCAG CACAG  
1701 CGATC CCTGC GCCCT CAACC CGCGC ATGAC CAGGT TCTCC GAGGA GGCGT  
1751 GCGCG GTCCCT GACGT CCCC ACATT CGAGG CCTGC CATCG TGCCG TCAGC  
1801 CCGCT GCCCT ACCTG CCGAA TGTCC GCTAC GACGT GTGCT CCTGC TCGGA  
1851 CGGCC GCGAG TGCCT GTGCG GCGCC CTGGC CAGCT ATGCC GCGGC CTGCG  
1901 CGGGG AGAGG CGTGC GCGTC GCGTG GCGCG AGCCA GGCCG CTGTG AGCTG  
1951 AACTG CCCGA AAGGC CAGGT GTACC TGCAG TGC GG GACCC CCTGC AACCT  
2001 GACCT GCCGC TCTCT CTCTT ACCCG GATGA GGAAT GCAAT GAGGC CTGCC  
2051 TGGAG GGCTG CTTCT GCCCC CCAGG GCTCT ACATG GATGA GAGGG GGGAC  
2101 TGCGT GCCCA AGGCC CAGTG CCCCT GTTAC TATGA CGGTG AGATC TTCCA  
2151 GCCAG AAGAC ATCTT CTCAG ACCAT CACAC CATGT GCTAC TGTGA GGATG  
2201 GCTTC ATGCA CTGTA CCATG AGTGG AGTCC CCGGA AGCTT GCTGC CTGAC  
2251 GCTGT CCTCA GCAGT CCCCT GTCTC ATCGC AGCAA AAGGA GCCTA TCCTG  
2301 TCGGC CCCCC ATGGT CAAGC TGGTG TGTCC CGCTG ACAAC CTGCG GGCTG  
2351 AAGGG CTCGA GTGTA CAAA ACGTG CCAGA ACTAT GACCT GGAGT GCATG  
2401 AGCAT GGGCT GTGTC TCTGG CTGCC TCTGC CCCCC GGGCA TGGT CCGCA  
2451 TGAGA ACAGA TGTGT GGCC TGGAA AGGTG TCCCT GCTTC CATCA GGGCA  
2501 AGGAG TATGC CCCTG GAGAA ACAGT GAAGA TTGGC TGCAA CACTT GTGTC  
2551 TGTCG GGACC GGAAG TGGAA CTGCA CAGAC CATGT GTGTG ATGCC ACGTG  
2601 CTCCA CGATC GGCAT GGCCC ACTAC CTCAC CTTCG ACGGG CTCAA ATACC  
2651 TGTTC CCCGG GGAGT GCCAG TACGT TCTGG TGCAG GATTA CTGCG GCAGT  
[0608] 2701 AACCC TGGGA CTTTT CGGAT CCTAG TGGGG AATAA GGGAT GCAGC CACCC  
2751 CTCAG TGAAA TGCAA GAAAC GGGTC ACCAT CCTGG TGGAG GGAGG AGAGA  
2801 TTGAG CTGTT TGACG GGGAG GTGAA TGTGA AGAGG CCCAT GAAGG ATGAG  
2851 ACTCA CTTTG AGGTG GTGGA GTCTG GCCGG TACAT CATTG TGCTG CTGGG  
2901 CAAAG CCCTC TCCGT GGTCT GGGAC CGCCA CCTGA GCATC TCCGT GGTCC  
2951 TGAAG CAGAC ATACC AGGAG AAAGT GTGTG GCCTG TGTGG GAATT TTGAT  
3001 GGCAT CCAGA ACAAT GACCT CACCA GCAGC AACCT CCAAG TGGAG GAAGA  
3051 CCCTG TGGAC TTTGG GAACT CCTGA AAAGT GAGCT CGCAG TGTGC TGACA  
3101 CCAGA AAAGT GCCTC TGGAC TCATC CCTG CCACC TGCCA TAACA ACATC  
3151 ATGAA GCAGA CGATG GTGGA TTCCCT CCTGT AGAAT CCTTA CCAGT GACGT  
3201 CTTCC AGGAC TGCAA CAAGC TGGTG GACCC CGAGC CATAT CTGGA TGTCT  
3251 GCATT TACGA CACCT GCTCC TGTGA GTCCA TTGGG GACTG CGCCG CATTG  
3301 TGCGA CACCA TTGCT GCCTA TGCCC ACGTG TGTGC CCAGC ATGGC AAGGT  
3351 GGTGA CCTGG AGGAC GGCCA CATTG TGCCC CCAGA GCTGC GAGGA GAGGA  
3401 ATCTC CGGGA GAACG GGTAT GAGGC TGAGT GGCGC TATAA CAGCT GTGCA  
3451 CCTGC CTGTC AAGTC ACGTG TCAGC ACCCT GAGCC ACTGG CCTGC CCTGT  
3501 GCAGT GTGTG GAGGG CTGCC ATGCC CACTG CCCTC CAGGG AAAAT CCTGG  
3551 ATGAG CTTTT GCAGA CCTGC GTTGA CCCTG AAGAC TGTC AGTGT GTGAG  
3601 GTGGC TGGCC GGCGT TTTGC CTCAG GAAAG AAAGT CACCT TGAAT CCCAG  
3651 TGACC CTGAG CACTG CCAGA TTTGC CACTG TGATG TTGTC AACCT CACCT  
3701 GTGAA GCCTG CCAGG AGCCG ATATC TGGCG GTGGA GGTTC CGGTG GCGGG  
3751 GGATC CGGCG GTGGA GGTTC CGGCG GTGGA GGTTC CGGTG GCGGG GGATC  
3801 CGGTG GCGGG GGATC CCTGG TCCCC CGGGG CAGCG GCGGT GGAGG TTCCG  
3851 GTGGC GGGGG ATCCG ACAA ACTCA CACAT GCCCA CCGTG CCCAG CTCCA  
3901 GAACT CCTGG GCGGA CCGTC AGTCT TCCTC TTCCC CCCAA AACCC AAGGA  
3951 CACCC TCATG ATCTC CCGGA CCCCT GAGGT CACAT GCGTG GTGGT GGACG  
4001 TGAGC CACGA AGACC CTGAG GTCAA GTTCA ACTGG TACGT GGACG GCGTG  
4051 GAGGT GCATA ATGCC AAGAC AAAGC CGCGG GAGGA GCAGT ACAAC AGCAC  
4101 GTACC GTGTG GTCAG CGTCC TCACC GTCC TGCACC AGGAC TGGCT GAATG  
4151 GCAAG GAGTA CAAGT GCAAG GTCTC CAACA AAGCC CTCCC AGCCC CCATC

[0609] 4201 GAGAA AACCA TCTCC AAAGC CAAAG GGCAG CCCC GAGAAC CACAG GTGTA  
 4251 CACCC TGCC CCGC ATGAG CTGAC CAAGA ACCAG GTCAG CCTGA  
 4301 CCTGC CTGGT CAAAG GCTTC TATCC CAGCG ACATC GCCGT GGAGT GGGAG  
 4351 AGCAA TGGGC AGCCG GAGAA CAACT ACAAG ACCAC GCCTC CCGTG TTGGA  
 4401 CTCCG ACGGC TCCTT CTTCC TCTAC AGCAA GCTCA CCGTG GACAA GAGCA  
 4451 GGTGG CAGCA GGGGA ACGTC TTCTC ATGCT CCGTG ATGCA TGAGG CTCTG  
 4501 CACAA CCACT ACACG CAGAA GAGCC TCTCC CTGTC TCCGG GTAAA TGA

[0610] VWF031蛋白质序列(SEQ ID NO: 86)

1 MIPARFAGVL LALALILPGT LCAEGTRGRS STARCSLFGS DFNVTFDGSM  
 51 YSFAGYCSYL LAGGCQKRSF SIIGDFQNGK RVLSLVYLGE FFDIHLFVNG  
 101 TVTQGDQRVS MPYASKGLYL ETEAGYYKLS GEAYGFVARI DGSGNFQVLL  
 151 SDRYFNKTCG LCGNFNIFAE DDFMTQEGTL TSDPYDFANS WALSSGEQWC  
 201 ERASPPSSC NISSGEMQKG LWEQCQLLKS TSVFARCHPL VDPEPFVALC  
 251 EKTLECECAGG LECACPALLE YARTCAQEGM VLYGWDHSA CSPVCPAGME  
 301 YRQCVSPCAR TCQSLHINEM CQERCVDGCS CPEGQLLEDEG LCVESTTECPC  
 351 VHSBKRYPPG TSLSRDCNTC ICRNSQWICS NEECPGECV TGQSHFKSFD  
 401 NRYFTFSGIC QYLLARDCQD HSFSIVIVTV QCADDRDAVC TRSVTVRLPG  
 451 LHNSLVKLVK GAGVAMDQD IQLPLLKGD LRIQHTVTASV RLSYGEDLQM  
 501 DWDGRGRLLV KLSVPYAGKT CGLCGNYNGN QGDDFLTPSG LAEPRVEDFG  
 551 NAWKLHGDCQ DLQKQHS DPC ALNPRMTRFS EEACAVLTSP TFEACHRAVS  
 601 PLPYLRNCRY DVCSCSDGRE CLCGALASYA AACAGRGRV AWREPGRCEL  
 651 NCPKGQVYLQ CGTPCNLTCR SLSYPDEECN EACLEGCFCP PGLYMDERGD  
 701 CVPKAQCPCY YDGEIFQPED IFSDHHTMCY CEDGFMHCTM SGVPGSLLPD  
 [0611] 751 AVLSSPLSHR SKRSLSCRPP MVKLVCPADN LRAEGLECTK TCQNYDLECM  
 801 SMGCVSGCLC PPGMVRHENR CVALERCPCF HQGKEYAPGE TVKIGCNTCV  
 851 CRDRKWNCTD HVCDATCSTI GMAHYLTFDG LKYLFPGECQ YVLVQDYCGS  
 901 NPGTFRILVG NKGCSHPSVK CKKRVTLVE GGEIELFDGE VNVKRPMDKE  
 951 THFEVVEGR YIILLGKAL SVVWDRHLSI SVVLKQTYQE KVCGLCGNFD  
 1001 GIQNDLTSS NLQVEEDPVD FGNSWKVSSQ CADTRKVPD SSPATCHNNI  
 1051 MKQTMVDSSC RILTSDFVQD CNKLVDPPEY LDVCIYDTCES CESIGDCAAF  
 1101 CDTIAAYAHV CAQHGVVVTW RTATLCPQSC EERNLRENGY EAEWRYNSCA  
 1151 PACQVTCQHP EPLACPVQCV EGCHAHCPPG KILDELLQTC VDPEDCPVCE  
 1201 VAGRRFASGK KVTLNPSDPE HCQICHCDVV NLTCEACQEP ISGGGGSGGG  
 1251 GSGGGSGGG GSGGGSGGG GSLVPRGSGG GSGGGSGGSDK THTCPPCPAP  
 1301 ELLGGPSVFL FPPKPKDTLM ISRTPEVTCV VVDVSHEDPE VKNFWYVDGV  
 1351 EVHNAKTKPR EEQYNSTYRV VSVLTVLHQD WLNKEYKCK VSNKALPAPI  
 1401 EKTISKAKGQ PREPQVYTLPSRDELTKNQ VSLTCLVKGF YPSDIAVEWE  
 1451 SNGQPENNYK TTPVLDSDG SFFLYSKLTV DKSRWQQGNV FSCSVMHEAL  
 1501 HNHYTQKSL SPSGK\*

[0612] VWF034核苷酸序列(SEQ ID NO:148)

1 ATGAT TCCTG CCAGA TTTGC CGGGG TGCTG CTTGC TCTGG CCCTC ATTTT  
 51 GCCAG GGACC CTTTG TGCAG AAGGA ACTCG CGGCA GGTC TCCAC GGCCC  
 101 GATGC AGCCT TTTGC GAAGT GACTT CGTCA ACACC TTTGA TGGGA GCATG  
 151 TACAG CTTTG CGGGA TACTG CAGTT ACCTC CTGGC AGGGG GCTGC CAGAA  
 201 ACGCT CCTTC TCGAT TATTG GGGAC TTCCA GAATG GCAAG AGAGT GAGCC  
 251 TCTCC GTGTA TCTTG GGGAA TTTT TGACA TCCAT TTGTT TGTC ATGGT  
 [0613] 301 ACCGT GACAC AGGGG GACCA AAGAG TCTCC ATGCC CTATG CCTCC AAAGG  
 351 GCTGT ATCTA GAAAC TGAGG CTGGG TACTA CAAGC TGTC GGTGA GGCCT  
 401 ATGGC TTTGT GGCCA GGATC GATGG CAGCG GCAAC TTTCA AGTCC TGCTG  
 451 TCAGA CAGAT ACTTC AACAA GACCT GCGGG CTGTG TGGCA ACTTT AACAT  
 501 CTTTG CTGAA GATGA CTTTA TGACC CAAGA AGGGA CCTG ACCTC GGACC  
 551 CTTAT GACTT TGCCA ACTCA TGGGC TCTGA GCAGT GGAGA ACAGT GGTGT  
 601 GAACG GGCAT CTCCT CCCAG CAGCT CATGC AACAT CTCCT CTGGG GAAAT

651 GCAGA AGGGC CTGTG GGAGC AGTGC CAGCT TCTGA AGAGC ACCTC GGTGT  
701 TTGCC CGCTG CCACC CTCTG GTGGA CCCCC AGCCT TTTGT GGCCC TGTGT  
751 GAGAA GACTT TGTGT GAGTG TGCTG GGGGG CTGGA GTGCG CCTGC CCTGC  
801 CCTCC TGGAG TACGC CCGGA CCTGT GCCCA GGAGG GAATG GTGCT GTACG  
851 GCTGG ACCGA CCACA GCGCG TGCAG CCCAG TGTGC CCTGC TGGTA TGGAG  
901 TATAG GCAGT GTGTG TCCCC TTGCG CCAGG ACCTG CCAGA GCCTG CACAT  
951 CAATG AAATG TGTCA GGAGC GATGC GTGGA TGGCT GCAGC TGCCC TGAGG  
1001 GACAG CTCCT GGATG AAGGC CTCTG CGTGG AGAGC ACCGA GTGTC CCTGC  
1051 GTGCA TTCCG GAAAG CGCTA CCCTC CCGGC ACCTC CCTCT CTCGA GACTG  
1101 CAACA CCTGC ATTTG CCGAA ACAGC CAGTG GATCT GCAGC AATGA AGAAT  
1151 GTCCA GGGGA GTGCC TTGTC ACTGG TCAAT CCCAC TTCAA GAGCT TTGAC  
1201 AACAG ATACT TCACC TTCAG TGGGA TCTGC CAGTA CCTGC TGGCC CGGGA  
1251 TTGCC AGGAC CACTC CTTCT CCATT GTCAT TGAGA CTGTC CAGTG TGCTG  
1301 ATGAC CGCGA CGCTG TGTGC ACCCG CTCCG TCACC GTCCG GCTGC CTGGC  
1351 CTGCA CAACA GCCTT GTGAA ACTGA AGCAT GGGGC AGGAG TTGCC ATGGA  
1401 TGGCC AGGAC ATCCA GCTCC CCCTC CTGAA AGGTG ACCTC CGCAT CCAGC  
1451 ATACA GTGAC GGCCT CCGTG CGCCT CAGCT ACGGG GAGGA CCTGC AGATG  
1501 GACTG GGATG GCCGC GGGAG GCTGC TGGTG AAGCT GTCCC CCGTC TATGC  
1551 CGGGA AGACC TGC GG CCTGT GTGGG AATTA CAATG GCAAC CAGGG CGACG  
1601 ACTTC CTTAC CCCCT CTGGG CTGGC GGAGC CCCGG GTGGA GGACT TCGGG  
1651 AACGC CTGGA AGCTG CACGG GGACT GCCAG GACCT GCAGA AGCAG CACAG  
1701 CGATC CCTGC GCCCT CAACC CGCGC ATGAC CAGGT TCTCC GAGGA GGCGT  
1751 GCGCG GTCCT GACGT CCCCC ACATT CGAGG CCTGC CATCG TGCCG TCAGC  
1801 CCGCT GCCCT ACCTG CGGAA CTGCC GCTAC GACGT GTGCT CCTGC TCGGA  
1851 CGGCC GCGAG TGCCT GTGCG GTGCC CAGCT ATGCC GCGGC CTGGC  
1901 CGGGG AGAGG CGTGC GCGTC GCGTG GCGCG AGCCA GGCCG CTGTG AGCTG  
1951 AACTG CCCGA AAGGC CAGGT GTACC TGCAG TCGGG GACCC CCTGC AACCT  
2001 GACCT GCCGC TCTCT CTCTT ACCCG GATGA GGAAT GCAAT GAGGC CTGCC  
2051 TGGAG GGCTG CTTCT GCCCC CCAGG GCTCT ACATG GATGA GAGGG GGGAC  
[0614] 2101 TGCGT GCCCA AGGCC CAGTG CCCCT GTTAC TATGA CGGTG AGATC TTCCA  
2151 GCCAG AAGAC ATCTT CTCAG ACCAT CACAC CATGT GCTAC TGTGA GGATG  
2201 GCTTC ATGCA CTGTA CCATG AGTGG AGTCC CCGGA AGCTT GCTGC CTGAC  
2251 GCTGT CCTCA GCAGT CCCCT GTCTC ATCGC AGCAA AAGGA GCCTA TCCTG  
2301 TCGGC CCCCC ATGGT CAAGC TGGTG TGTCC CGCTG ACAAC CTGCG GGCTG  
2351 AAGGG CTCGA GTGTA CAAA ACGTG CCAGA ACTAT GACCT GGAGT GCATG  
2401 AGCAT GGGCT GTGTC TCTGG CTGCC TCTGC CCCCC GGGCA TGGTC CGGCA  
2451 TGAGA ACAGA TGTGT GGCCC TGGAA AGGTG TCCCT GCTTC CATCA GGGCA  
2501 AGGAG TATGC CCCTG GAGAA ACAGT GAAGA TTGGC TGCAA CACTT GTGTC  
2551 TGTGCG GGACC GGAAG TGGAA CTGCA CAGAC CATGT GTGTG ATGCC AGCTG  
2601 CTCCA CGATC GGCAT GGCCC ACTAC CTCAC CTTCCG ACGGG CTCAA ATACC  
2651 TGTTT CCCGG GGAGT GCCAG TACGT TCTGG TGCAG GATTA CTGCG GCAGT  
2701 AACCC TGGGA CCTTT CGGAT CCTAG TGGGG AATAA GGGAT GCAGC CACCC  
2751 CTCAG TGAAA TGCAA GAAAC GGGTC ACCAT CCTGG TGGAG GGAGG AGAGA  
2801 TTGAG CTGTT TGACG GGGAG GTGAA TGTGA AGAGG CCCAT GAAGG ATGAG  
2851 ACTCA CTTTG AGGTG GTGGA GTCTG GCCGG TACAT CATTG TGCTG CTGGG  
2901 CAAAG CCCTC TCCGT GGTCT GGGAC CGCCA CCTGA GCATC TCCGT GTTCC  
2951 TGAAG CAGAC ATACC AGGAG AAAGT GTGTG GCCTG TGTGG GAATT TTGAT  
3001 GGCAT CCAGA ACAAT GACCT CACCA GCAGC AACCT CCAAG TGGAG GAAGA  
3051 CCCTG TGGAC TTTGG GAACT CCTGG AAAGT GAGCT CGCAG TGTGC TGACA  
3101 CCAGA AAAGT GCCTC TGGAG TCATC CCCTG CCACC TGCCA TAACA ACATC  
3151 ATGAA GCAGA CGATG GTGGA TTCTT CCTGT AGAAT CTTA CCAGT GACGT  
3201 CTTCC AGGAC TGCAA CAAGC TGGTG GACCC CGAGC CATAT CTGGA TGTCT  
3251 GCATT TACGA CACCT GCTCC TGTGA GTCCA TTGGG GACTG CGCCG CATTG  
3301 TGCGA CACCA TTGCT GCCTA TGCCC ACGTG TGTGC CCAGC ATGGC AAGGT  
3351 GGTGA CCTGG AGGAC GGCCA CATTG TGCCC CCAGA GCTGC GAGGA GAGGA  
3401 ATCTC CGGGA GAACG GGTAT GAGGC TGAGT GGCGC TATAA CAGCT GTGCA  
3451 CCTGC CTGTC AAGTC ACGTG TCAGC ACCCT GAGCC ACTGG CCTGC CCTGT  
3501 GCAGT GTGTG GAGGG CTGCC ATGCC CACTG CCCTC CAGGG AAAAT CCTGG  
3551 ATGAG CTTTT GCAGA CCTGC GTTGA CCCTG AAGAC TGTCC AGTGT GTGAG

3601 GTGGC TGGCC GGCCT TTTGC CTCAG GAAAG AAAGT CACCT TGAAT CCCAG  
 3651 TGACC CTGAG CACTG CCAGA TTTGC CACTG TGATG TTGTC AACCT CACCT  
 3701 GTGAA GCCTG CCAGG AGCCG ATATC GGGTA CCTCA GAGTC TGCTA CCCCC  
 3751 GAGTC AGGGC CAGGA TCAGA GCCAG CCACC TCCGG GTCTG AGACA CCCGG  
 3801 GACTT CCGAG AGTGC CACCC CTGAG TCCGG ACCCG GGTCC GAGCC CGCCA  
 3851 CTTCC GGCTC CGAAA CTCCC GGCAC AAGCG AGAGC GCTAC CCCAG AGTCA  
 3901 GGACC AGGAA CATCT ACAGA GCCCT CTGAA GGCTC CGCTC CAGGG TCCCC  
 3951 AGCCG GCAGT CCCAC TAGCA CCGAG GAGGG AACCT CTGAA AGCGC CACAC  
 4001 CCGAA TCAGG GCCAG GGTCT GAGCC TGCTA CCAGC GGCAG CGAGA CACCA  
 4051 GGCAC CTCTG AGTCC GCCAC ACCAG AGTCC GGACC CGGAT CTCCC GCTGG  
 4101 GAGCC CCACC TCCAC TGAGG AGGGA TCTCC TGCTG GCTCT CCAAC ATCTA  
 4151 CTGAG GAAGG TACCT CAACC GAGCC ATCCG AGGGA TCAGC TCCCG GCACC  
 4201 TCAGA GTCCG CAACC CCGGA GTCTG GACCC GGAAC TTCCG AAAGT GCCAC  
 4251 ACCAG AGTCC GGTCC CGGGA CTTCA GAATC AGCAA CACCC GAGTC CGGCC  
 4301 CTGGG TCTGA ACCCG CCACA AGTGG TAGTG AGACA CCAGG ATCAG AACCT  
 4351 GCTAC CTCAG GGTC A GAGAC ACCCG GATCT CCGGC AGGCT CACCA ACCTC  
 4401 CACTG AGGAG GGCAC CAGCA CAGAA CCAAG CGAGG GCTCC GCACC CGGAA  
 [0615] 4451 CAAGC ACTGA ACCCA GTGAG GGTTC AGCAC CCGGC TCTGA GCCGG CCACA  
 4501 AGTGG CAGTG AGACA CCCGG CACTT CAGAG AGTGC CACCC CCGAG AGTGG  
 4551 CCCAG GCACT AGTAC CGAGC CCTCT GAAGG CAGTG CGCCA GATTC TGGCG  
 4601 GTGGA GGTTC CGGTG GCGGG GGATC CGGTG GCGGG GGATC CGGTG GCGGG  
 4651 GGATC CGGTG GCGGG GGATC CCTGG TCCCC CCGGG CAGCG GAGGC GACAA  
 4701 AACTC ACACA TGCCC ACCGT GCCCA GCTCC AGAAC TCCTG GCGGG ACCGT  
 4751 CAGTC TTCCCT CTTCC CCCCC AAACC CAAGG ACACC CTCAT GATCT CCCGG  
 4801 ACCCC TGAGG TCACA TGCGT GGTGG TGGAC GTGAG CCACG AAGAC CCTGA  
 4851 GGTCA AGTTC AACTG GTACG TGGAC GGCGT GGAGG TGCAT AATGC CAAGA  
 4901 CAAAG CCGCG GGAGG AGCAG TACAA CAGCA CGTAC CGTGT GGTCA GCGTC  
 4951 CTCAC CGTCC TGCAC CAGGA CTGGC TGAAT GGCAA GGAGT ACAAG TGCAA  
 5001 GGTCT CCAAC AAAGC CCTCC CAGCC CCCAT CGAGA AAACC ATCTC CAAAG  
 5051 CCAAA GGGCA GCCCC GAGAA CCACA GGTGT ACACC CTGCC CCCAT CCCGG  
 5101 GATGA GCTGA CCAAG AACCA GGTCA GCCTG ACCTG CCTGG TCAA GGCTT  
 5151 CTATC CCAGC GACAT CGCCG TGGAG TGGGA GAGCA ATGGG CAGCC GGAGA  
 5201 ACAAC TACAA GACCA CGCCT CCCGT GTTGG ACTCC GACGG CTCCT TCTTC  
 5251 CTCTA CAGCA AGCTC ACCGT GGACA AGAGC AGGTG GCAGC AGGGG AACGT  
 5301 CTTCT CATGC TCCGT GATGC ATGAG GCTCT GCACA ACCAC TACAC GCAGA  
 5351 AGAGC CTCTC CCTGT CTCCG GGTAA ATGA

[0616] VWF034蛋白质序列(SEQ ID NO:87)

```

1     MIPARFAGVL LALALILPGT LCAEGTRGRS STARCSLFGS DFNNTFDGSM
51    YSFAGYCSYL LAGGCQKRSF SIIGDFQNGK RVSLSVYLGE FFDIHLFVNG
101   TVTQGDQRVS MPYASKGLYL ETEAGYYKLS GEAYGFVARI DGSGNFQVLL
151   SDRYFNKTCG LCGNFNIFAE DDFMTQEGTL TSDPYDFANS WALSSGEQWC
201   ERASPPSSSC NISSGEMQKG LWEOCQLLKS TSVFARCHPL VDPEPFVALC
251   EKTLCCECAGG LECACPALLE YARTCAQEGM VLYGWDHSA CSPVCPAGME
301   YRQCVSPCAR TCQSLHINEM CQERCVDGCS CPEGQLLDEG LCVESTTEPC
351   VHSGKRYPPG TSLSRDCNTC ICRNSQWICS NEECPGECLV TGQSHFKSFD
401   NRYFTFSGIC QYLLARDCQD HSFSIVIETV QCADDRDAVC TRSVTVRLPG
451   LHNSLVKCLKH GAGVAMDGQD IQLPLLKGD LRIQHTVTASV RLSYGEDLQM
501   DWDGRGRLLV KLSPVYAGKT CGLCGNYNGN QGDDFLTPSG LAEPRVEDFG
551   NAWKLHGDCQ DLQKQHSDFC ALNPRMTRFS EEACAVLTSP TFEACHRAVS
601   PLPYLRNCRY DVCSCSDGRE CLCGALASYA AACAGRGVRV AWREPGRCCL
651   NCPKGQVYLQ CGTPCNLTCT SLSYPDEECN EACLEGCFCP PGLYMDERGD
701   CVPKAQCPCY YDGEIFQPED IFSDHHTMCY CEDGFMHCTM SGVPGSLLPD
751   AVLSSPLSHR SKRSLSCRPP MVKLVCPADN LRAEGLECTK TCQNYDLECM
801   SMGCVSGCLC PPGMVRHENR CVALERCPCF HQGKEYAPGE TVKIGCNTCV
851   CRDRKWNCTD HVCDATCSTI GMAHYLTFDG LKYLFPGEQY YVLVQDYCGS

```

[0617]

```

901   NPGTFRILVG NKGCSHPSVK CKKRVTILVE GGEIELFDGE VNVKRPMKDE
951   THFEVVEGR YIILLGKAL SVVWDRHLSI SVVLKQTYQE KVCGLCGNFD
1001  GIQNNDLTSS NLQVEEDPVD FGNSWKVSSQ CADTRKVPLD SSPATCHNNI
1051  MKQTMVDSSC RILTSDFVQD CNKLVDPPEY LDVCIYDTCS CESIGDCAAF
1101  CDTIAAYAHV CAQHGVVVTW RTATLCPQSC EERNLRENGY EAEWRYNSCA
1151  PACQVTCQHP EPLACPVCQV EGCHAHCPPG KILDELLQTC VDPEDCPVCE
1201  VAGRRFASGK KVTLNPSDPE HCQICHCDVV NLTCEACQEP ISGTSESATP
1251  ESGPGSEPAT SGSETPGTSE SATPESGPGS EPATSGSETP GTSESATPES
1301  GPGTSTEPSE GSAPGSPAGS PTSTEEGTSE SATPESGPGS EPATSGSETP
1351  GTSESATPES GPGSPAGSPT STEEGSPAGS PTSTEEGTST EPSEGSAPGT
1401  SESATPESGP GTSESATPES GPGTSESATP ESGPGSEPAT SGSETPGSEP
1451  ATSGSETPGS PAGESPTSTEE GTSTEPSEGS APGTSTEPSE GSAPGSEPAT
1501  SGSETPGTSE SATPESGPGT STEPSEGSAP DIGGGGGSGG GGSVLPVPRSG
1551  GDKTHTCPPC PAPELLGGPS VFLFPPKPKD TLMISRTPEV TCVVVDVSHE
1601  DPEVKFNWYV DGVEVHNAKT KPREEQYNST YRVVSVLTVL HQDWLNGKEY
1651  KCKVSNKALP APIEKTISKA KGQPREPQVY TLPPSRDEL TKNQVSLTCLV
1701  KGFYPSDIAV EWESNGQPEN NYKTTTPVLD SDGSFFLYSK LTVDKSRWQQ
1751  GNVFSCSVMH EALHNHYTQK SLSLSPGK*

```

[0618] VWF050核苷酸序列(IHH三重突变体) (SEQ ID NO:149)

1 ATGAT TCCTG CCAGA TTTGC CGGGG TGCTG CTTGC TCTGG CCCTC ATTTT  
 51 GCCAG GGACC CTTTG TGCAG AAGGA ACTCG CGGCA GGTCA TCCAC GGCC  
 101 GATGC AGCCT TTTCG GAAGT GACTT CGTCA ACACC TTTGA TGGGA GCATG  
 151 TACAG CTTTG CGGGA TACTG CAGTT ACCTC CTGGC AGGGG GCTGC CAGAA  
 201 ACGCT CCTTC TCGAT TATTG GGGAC TTCCA GAATG GCAAG AGAGT GAGCC  
 251 TCTCC GTGTA TCTTG GGGAA TTTTT TGACA TCCAT TTGTT TGTCA ATGGT  
 301 ACCGT GACAC AGGGG GACCA AAGAG TCTCC ATGCC CTATG CCTCC AAAGG  
 351 GCTGT ATCTA GAAAC TGAGG CTGGG TACTA CAAGC TGTCC GGTGA GGCCT  
 401 ATGGC TTTGT GGCCA GGATC GATGG CAGCG GCAAC TTTCA AGTCC TGCTG  
 451 TCAGA CAGAT ACTTC AACAA GACCT GCGGG CTGTG TGGCA ACTTT AACAT  
 501 CTTTG CTGAA GATGA CTTTA TGACC CAAGA AGGGA CCTTG ACCTC GGACC  
 551 CTTAT GACTT TGCCA ACTCA TGGGC TCTGA GCAGT GGAGA ACAGT GGTGT  
 601 GAACG GGCAT CTCCT CCCAG CAGCT CATGC AACAT CTCCT CTGGG GAAAT  
 651 GCAGA AGGGC CTGTG GGAGC AGTGC CAGCT TCTGA AGAGC ACCTC GGTGT  
 701 TTGCC CGCTG CCACC CTCCT GTGGA CCCCC AGCCT TTTGT GGCC TGTGT  
 751 GAGAA GACTT TGTGT GAGTG TGCTG GGGGG CTGGA GTGCG CCTGC CCTGC  
 801 CCTCC TGGAG TACGC CCGGA CCTGT GCCCA GGAGG GAATG GTGCT GTACG  
 851 GCTGG ACCGA CCACA GCGCG TGCAG CCCAG TGTGC CCTGC TGGTA TGGAG  
 [0619] 901 TATAG GCAGT GTGTG TCCCC TTGCG CCAGG ACCTG CCAGA GCCTG CACAT  
 951 CAATG AAATG TGTC A GGAGC GATGC GTGGA TGGCT GCAGC TGCCC TAGAG  
 1001 GACAG CTCCT GGATG AAGGC CTCCT CGTGG AGAGC ACCGA GTGTC CCTGC  
 1051 GTGCA TTCCG GAAAG CGCTA CCCTC CCGGC ACCTC CCTCT CTCGA GACTG  
 1101 CAACA CCTGC ATTTG CCGAA ACAGC CAGTG GATCT GCAGC AATGA AGAAT  
 1151 GTCCA GGGGA GTGCC TTGTC ACTGG TCAAT CCCAC TTCAA GAGCT TTGAC  
 1201 AACAG ATACT TCACC TTCAG TGGGA TCTGC CAGTA CCTGC TGGCC CGGGA  
 1251 TTGCC AGGAC CACTC CTTCT CCATT GTCAT TGAGA CTGTC CAGTG TGCTG  
 1301 ATGAC CGCGA CGCTG TGTGC ACCCG CTCCG TCACC GTCCG GCTGC CTGGC  
 1351 CTGCA CAACA GCCTT GTGAA ACTGA AGCAT GGGGC AGGAG TTGCC ATGGA  
 1401 TGGCC AGGAC ATCCA GCTCC CCCTC CTGAA AGGTG ACCTC CGCAT CCAGC  
 1451 ATACA GTGAC GGCCT CCGTG CGCCT CAGCT ACGGG GAGGA CCTGC AGATG  
 1501 GACTG GGATG GCCGC GGGAG GCTGC TGGTG AAGCT GTCCC CCGTC TATGC  
 1551 CGGGA AGACC TGCAG CTTGT GTGGG AATTA CAATG GCAAC CAGGG CGACG  
 1601 ACTTC CTTAC CCCCT CTGGG CTGGC GGAGC CCCGG GTGGA GGACT TCGGG  
 1651 AACGC CTGGA AGCTG CACGG GGACT GCCAG GACCT GCAGA AGCAG CACAG  
 1701 CGATC CCTGC GCCCT CAACC CGCGC ATGAC CAGGT TCTCC GAGGA GGCGT  
 1751 GCGCG GTCCCT GACGT CCCCC ACATT CGAGG CCTGC CATCG TGCCG TCAGC  
 1801 CCGCT GCCCT ACCTG CCGAA CTGCC GCTAC GACGT GTGCT CCTGC TCGGA

1851 CGGCC GCGAG TGCCT GTGCG GCGCC CTGGC CAGCT ATGCC GCGGC CTGCC  
 1901 CGGGG AGAGG CGTGC GCGTC GCGTG GCGCG AGCCA GGCCG CTGTG AGCTG  
 1951 AACTG CCCGA AAGGC CAGGT GTACC TGCAG TCGCG GACCC CCTGC AACCT  
 2001 GACCT GCCGC TCTCT CTCTT ACCCG GATGA GGAAT GCAAT GAGGC CTGCC  
 2051 TGGAG GGCTG CTTCT GCCCC CCAGG GCTCT ACATG GATGA GAGGG GGGAC  
 2101 TGGCT GCCCA AGGCC CAGTG CCCCT GTTAC TATGA CGGTG AGATC TTCCA  
 2151 GCCAG AAGAC ATCTT CTCAG ACCAT CACAC CATGT GCTAC TGTGA GGATG  
 2201 GCTTC ATGCA CTGTA CCATG AGTGG AGTCC CCGGA AGCTT GCTGC CTGAC  
 2251 GCTGT CCTCA GCAGT CCCCT GTCTC ATCGC AGCAA AAGGA GCCTA TCCTG  
 2301 TCGGC CCCCC ATGGT CAAGC TGGTG TGTCC CGCTG ACAAC CTGCG GGCTG  
 2351 AAGGG CTCGA GTGTA CCAA ACGTG CCAGA ACTAT GACCT GGAGT GCATG  
 2401 AGCAT GGGCT GTGTC TCTGG CTGCC TCTGC CCCCC GGGCA TGGTC CGGCA  
 2451 TGAGA ACAGA TGTGT GGCCC TGGAA AGGTG TCCCT GCTTC CATCA GGGCA  
 2501 AGGAG TATGC CCCTG GAGAA ACAGT GAAGA TTGGC TGCAA CACTT GTGTC  
 2551 TGTGC GGACC GGAAG TGGAA CTGCA CAGAC CATGT GTGTG ATGCC ACGTG  
 2601 CTCCA CGATC GGCAT GGCCC ACTAC CTCAC CTTCC ACGGG CTCAA ATACC  
 2651 TGTTC CCCGG GGAGT GCCAG TACGT TCTGG TGCAG GATTA CTGCG GCAGT  
 2701 AACCC TGGGA CCTTT CGGAT CCTAG TGGGG AATAA GGGAT GCAGC CACCC  
 2751 CTCAG TGAAA TGCAA GAAAC GGGTC ACCAT CCTGG TGGAG GGAGG AGAGA  
 2801 TTGAG CTGTT TGACG GGGAG GTGAA TGTGA AGAGG CCCAT GAGG ATGAG  
 2851 ACTCA CTTTG AGGTG GTGGA GTCTG GCCGG TACAT CATTG TGCTG CTGGG  
 2901 CAAAG CCTTC TCCGT GGTCT GGGAC CGCCA CCTGA GCATC TCCGT GGTCC  
 2951 TGAAG CAGAC ATACC AGGAG AAAGT GTGTG GCCTG TGTGG GAATT TTGAT  
 3001 GGCAT CCAGA ACAAT GACCT CACCA GCAGC AACCT CCAAG TGGAG GAAGA  
 3051 CCCTG TGGAC TTTGG GAACT CCTGG AAAGT GAGCT CGCAG TGTGC TGACA  
 3101 CCAGA AAAGT GCCTC TGGAC TCATC CCCTG CCACC TGCCA TAACA ACATC  
 [0620] 3151 ATGAA GCAGA CGATG GTGGA TTCCT CCTGT AGAAT CCTTA CCAGT GACGT  
 3201 CTTCC AGGAC TGCAA CAAGC TGGTG GACCC CGAGC CATAT CTGGA TGTCT  
 3251 GCATT TACGA CACCT GCTCC TGTGA GTCCA TTGGG GACTG CGCCG CATTG  
 3301 TGCGA CACCA TTGCT GCCTA TGCCC ACGTG TGTGC CCAGC ATGGC AAGGT  
 3351 GGTGA CCTGG AGGAC GGCCA CATTG TGCCC CCAGA GCTGC GAGGA GAGGA  
 3401 ATCTC CGGGA GAACG GGTAT GAGGC TGAGT GGCGC TATAA CAGCT GTGCA  
 3451 CCTGC CTGTC AAGTC ACGTG TCAGC ACCCT GAGCC ACTGG CCTGC CCTGT  
 3501 GCAGT GTGTG GAGGG CTGCC ATGCC CACTG CCCTC CAGGG AAAAT CCTGG  
 3551 ATGAG CTTTT GCAGA CCTGC GTTGA CCCTG AAGAC TGTCC AGTGT GTGAG  
 3601 GTGGC TGGCC GGCCT TTTGC CTCAG GAAAG AAAGT CACCT TGAAT CCCAG  
 3651 TGACC CTGAG CACTG CCAGA TTTGC CACTG TGATG TTGTC AACCT CACCT  
 3701 GTGAA GCCTG CCAGG AGCCG ATATC TGGCG GTGGA GGTTC CGGTG GCGGG  
 3751 GGATC CGGCG GTGGA GGTTC CGGCG GTGGA GGTTC CGGTG GCGGG GGATC  
 3801 CGGTG GCGGG GGATC CCTGG TCCCC CGGGG CAGCG GCGGT GGAGG TTCCG  
 3851 GTGGC GGGGG ATCCG ACAA ACTCA CACAT GCCCA CCGTG CCCAG CTCCA  
 3901 GAACT CCTGG GCGGA CCGTC AGTCT TCCTC TTCCC CCCAA AACCC AAGGA  
 3951 CACCC TCATG GCCTC CCGGA CCCCT GAGGT CACAT GCGTG GTGGT GGACG  
 4001 TGAGC CACGA AGACC CTGAG GTCAA GTTCA ACTGG TACGT GGACG GCGTG  
 4051 GAGGT GCATA ATGCC AAGAC AAAGC CGCGG GAGGA GCAGT ACAAC AGCAC  
 4101 GTACC GTGTG GTCAG CGTCC TCACC GTCCT GGCCC AGGAC TGGCT GAATG  
 4151 GCAAG GAGTA CAAGT GCAAG GTCTC CAACA AAGCC CTCCC AGCCC CCATC  
 4201 GAGAA AACCA TCTCC AAAGC CAAAG GGCAG CCCCC AGAAC CACAG GTGTA  
 4251 CACCC TGCCC CCATC CCGCG ATGAG CTGAC CAAGA ACCAG GTCAG CCTGA  
 4301 CCTGC CTGGT CAAAG GCTTC TATCC CAGCG ACATC GCCGT GGAGT GGGAG  
 4351 AGCAA TGGGC AGCCG GAGAA CAACT ACAAG ACCAC GCCTC CCGTG TTGGA  
 4401 CTCCG ACGGC TCCTT CTTCC TCTAC AGCAA GCTCA CCGTG GACAA GAGCA  
 4451 GGTGG CAGCA GGGGA ACGTC TTCTC ATGCT CCGTG ATGCA TGAGG CTCTG  
 4501 CACAA CGCCT ACACG CAGAA GAGCC TCTCC CTGTC TCCGG GTAAA TGA

[0621] VWF050蛋白质序列(IHH三重突变体)(SEQ ID NO:150)

[0622] 1 MIPARFAGVL LALALILPGT LCAEGTRGRS STARCSLFGS DFNFTFDGSM



[0623]

```

51   YSFAGYCSYL LAGGCQKRSF SIIGDFQNGK RVLSLVYLGE FFDIHLFVNG
101  TVTQGDQQRVS MPYASKGLYL ETEAGYYKLS GEAYGFVARI DGSGNFQVLL
151  SDRYFNKTCG LCGNFNIFAE DDFMTQEGTL TSDPYDFANS WALSSGEQWC
201  ERASPPSSSC NISSGEMQKG LWEQCQLLKS TSVFARCHPL VDPEPFVALC
251  EKTLCECAGG LECACPALLE YARTCAQEGM VLYGWTDHSA CSPVCPAGME
301  YRQCVSPCAR TCQSLHINEM CQERCVDGCS CPEGQLLDEG LCVESTTECPC
351  VHS GKRYPPG TSLSRDCNTC ICRNSQWICS NEECPGECLV TGQSHFKSFD
401  NRYFTFSGIC QYLLARDCQD HSFSIVIETV QCADDRDAVC TRSVTVRLPG
451  LHNSLVKCLKH GAGVAMDGDQ IQLPLLKGD LRIQHTVTASV RLSYGEDLQM
501  DWDGRGRLLV KLSPVYAGKT CGLCGNYNGN QGDDFLTPSG LAEPRVEDFG
551  NAWKLGDCQ DLQKQHSDFC ALNPRMTRFS EEACAVLTSP TFEACHRAVS
601  PLPYLRNCRY DVCSCSDGRE CLCGALASYA AACAGRGVRV AWREPGRCEL
651  NCPKGQVYLQ CGTPCNLTCR SLSYPDEECN EACLEGCFCP PGLYMDERGD
701  CVPKAQCPCY YDGEIFQPED IFSDHHTMCY CEDGFMHCTM SGVPGSLLPD
751  AVLSSPLSHR SKRSLSCRPP MVKLVCPADN LRAEGLECTK TCQNYDLECM
801  SMGCVSGCLC PPGMVRHENR CVALERCPCF HQGKEYAPGE TVKIGCNTCV
851  CRDRKWNCTD HVCDATCSTI GMAHYLTFDG LKYLFPGECQ YVLVQDYCGS
901  NPGTFRILVG NKGCSHPSVK CKKRVTILVE GGEIELFDGE VNVKRPMKDE
951  THFEVVESGR YII LLLGKAL SVVWDRHLSI SVVLKQTYQE KVCGLCGNFD
1001 GIQNNDLTSS NLQVEEDPVD FGNSWKVSSQ CADTRKVPLD SSPATCHNNI
1051 MKQTMVDSSC RILTSDFVQD CNKLVDPPEY LDVCIYDTCS CESIGDCAAF
1101 CDTIAAYAHV CAQHGVVVTW RTATLCPQSC EERNLRENGY EAEWRYNSCA
1151 PACQVTCQHP EPLACPVCQV EGCHAHCPPG KILDELLQTC VDPEDCPVCE
1201 VAGRRFASGK KVTLNPSDPE HCQICHCDVV NLTCEACQEP ISGGGGSGGG
1251 GSGGGGSGGG GSGGGGSGGG GSLVPRGSGG GSGGGGSDK THTCPPCPAP
1301 ELLGGPSVFL FPPKPKDTLM ASRTPEVTCV VVDVSHEDPE VKFNWYVDGV
1351 EVHNAKTKPR EEQYNSTYRV VSVLTVLAQD WLNKKEYKCK VSNKALPAPI
1401 EKTISKAKGQ PREPQVYTL PPSRDELTKNQ VSLTCLVKGF YPSDIAVEWE
1451 SNGQPENNYK TTPPVLDSDG SFFLYSKLTV DKSRWQQGNV FSCSVMHEAL
1501 HNAYTQKSL S LSPGK*
    
```

[0624] VWF057核昔酸序列 (SEQ ID NO:151)

[0625]

```

1   ATGAT TCCTG CCAGA TTTGC CGGGG TGCTG CTTGC TCTGG CCCTC ATTTT
51  GCCAG GGACC CTTTG TGCAG AAGGA ACTCG CGGCA GGTC A TCCAC GGCCC
101 GATGC AGCCT TTTGC GAAGT GACTT CGTCA ACACC TTTGA TGGGA GCATG
151 TACAG CTTTG CGGGA TACTG CAGTT ACCTC CTGGC AGGGG GCTGC CAGAA
201 ACGCT CCTTC TCGAT TATTG GGGAC TTCCA GAATG GCAAG AGAGT GAGCC
251 TCTCC GTGTA TCTTG GGGAA TTTTT TGACA TCCAT TTGTT TGTC A ATGGT
301 ACCGT GACAC AGGGG GACCA AAGAG TCTCC ATGCC CTATG CCTCC AAAGG
351 GCTGT ATCTA GAAAC TGAGG CTGGG TACTA CAAGC TGTC C GGTGA GGCCT
401 ATGGC TTTGT GGCCA GGATC GATGG CAGCG GCAAC TTTCA AGTCC TGCTG
451 TCAGA CAGAT ACTTC AACAA GACCT GCGGG CTGTG TGGCA ACTTT AACAT
501 CTTTG CTGAA GATGA CTTTA TGACC CAAGA AGGGA CCTTG ACCTC GGACC
551 CTTAT GACTT TGCCA ACTCA TGGGC TCTGA GCAGT GGAGA ACAGT GGTGT
601 GAACG GGCAT CTCCT CCCAG CAGCT CATGC AACAT CTCCT CTGGG GAAAT
651 GCAGA AGGGC CTGTG GGAGC AGTGC CAGCT TCTGA AGAGC ACCTC GGTGT
701 TTGCC CGCTG CCACC CTCTG GTGGA CCCCC AGCCT TTTGT GGCCC TGTGT
751 GAGAA GACTT TGTGT GAGTG TGCTG GGGGG CTGGA GTGCG CCTGC CCTGC
801 CCTCC TGGAG TACGC CCGGA CCTGT GCCCA GGAGG GAATG GTGCT GTACG
851 GCTGG ACCGA CCACA GCGCG TGCAG CCCAG TGTGC CCTGC TGGTA TGGAG
901 TATAG GCAGT GTGTG TCCCC TTGCG CCAGG ACCTG CCAGA GCCTG CACAT
951 CAATG AAATG TGTC A GGAGC GATCG GTGGA TGGCT GCAGC TGCCC TGAGG
1001 GACAG CTCCT GGATG AAGC CTCTG CTTGG AGAGC ACCGA GTGCC CCTGC
1051 GTGCA TTCCG GAAAG CGCTA CCTTC CCGGC ACCTC CCTCT CTCGA CACTG
1101 CAACA CCTGC ATTTG CCGAA ACAGC CAGTG GATCT GCAGC AATGA AGAAT
    
```

1151 GTCCA GGGGA GTGCC TTGTC ACTGG TCAAT CCCAC TTCAA GAGCT TTGAC  
1201 AACAG ATACT TCACC TTCAG TGGGA TCTGC CAGTA CCTGC TGGCC CGGGA  
1251 TTGCC AGGAC CACTC CTTCT CCATT GTCAT TGAGA CTGTC CAGTG TGCTG  
1301 ATGAC CGCGA CGCTG TGTGC ACCCG CTCCG TCACC GTCCG GCTGC CTGGC  
1351 CTGCA CAACA GCCTT GTGAA ACTGA AGCAT GGGGC AGGAG TTGCC ATGGA  
1401 TGGCC AGGAC ATCCA GCTCC CCCTC CTGAA AGGTG ACCTC CGCAT CCAGC  
1451 ATACA GTGAC GGCCT CCGTG CGCCT CAGCT ACGGG GAGGA CCTGC AGATG  
1501 GACTG GGATG GCCGC GGGAG GCTGC TGGTG AAGCT GTCCC CCGTC TATGC  
1551 CGGGA AGACC TGC GG CTTGT GTGGG AATTA CAATG GCAAC CAGGG CGACG  
1601 ACTTC CTTAC CCCCT CTGGG CTGGC GGAGC CCCGG GTGGA GGA CT TCGGG  
1651 AACGC CTGGA AGCTG CACGG GGACT GCCAG GACCT GCAGA AGCAG CACAG  
1701 CGATC CCTGC GCCCT CAACC CGCGC ATGAC CAGGT TCTCC GAGGA GCGCT  
1751 GCGCG GTCCT GACGT CCCCC ACATT CGAGG CCTGC CATCG TGCCG TCAGC  
1801 CCGCT GCCCT ACCTG CGGAA CTGCC GCTAC GACGT GTGCT CCTGC TCGGA  
1851 CGGCC GCGAG TGCCT GTGCG GCGCC CTGGC CAGCT ATGCC GCGGC CTGCG  
1901 CGGGG AGAGG CGTGC GCGTC GCGTG GCGCG AGCCA GGCCG CTGTG AGCTG  
1951 AACTG CCCGA AAGGC CAGGT GTACC TGCAG TCGGG GACCC CCTGC AACCT  
2001 GACCT GCCGC TCTCT CTCTT ACCCG GATGA GGAAT GCAAT GAGGC CTGCC  
2051 TGGAG GGCTG CTTCT GCCCC CCAGG GCTCT ACATG GATGA GAGGG GGGAC  
2101 TGCGT GCCCA AGGCC CAGTG CCCCT GTTAC TATGA CGGTG AGATC TTCCA  
2151 GCCAG AAGAC ATCTT CTCAG ACCAT CACAC CATGT GCTAC TGTGA GGATG  
2201 GCTTC ATGCA CTGTA CCATG AGTGG AGTCC CCGGA AGCTT GCTGC CTGAC  
2251 GCTGT CCTCA GCAGT CCCCT GTCTC ATCGC AGCAA AAGGA GCCTA TCCTG  
2301 TCGGC CCCCC ATGGT CAAGC TGGTG TGTC CGCTG ACAAC CTGCG GGCTG  
2351 AAGGG CTCGA GTGTA CAAAA ACGTG CCAGA ACTAT GACCT GGAGT GCATG  
2401 AGCAT GGGCT GTGTC TCTGG CTGCC TCTGC CCCCC GGGCA TGGTC GGGCA  
2451 TGAGA ACAGA TGTGT GGCCC TGGAA AGGTG TCCCT GCTTC CATCA GGGCA  
2501 AGGAG TATGC CCCTG GAGAA ACAGT GAAGA TTGGC TGCAA CACTT GTGTC  
2551 TGTCG GGACC GGAAG TGGAA CTGCA CAGAC CATGT GTGTG ATGCC ACGTG  
[0626] 2601 CTCCA CGATC GGCAT GGCCC ACTAC CTCAC CTTCG ACGGG CTCAA ATACC  
2651 TGTTT CCCGG GGAGT GCCAG TACGT TCTGG TGCAG GATTA CTGCG GCAGT  
2701 AACCC TGGGA CCTTT CGGAT CCTAG TGGGG AATAA GGGAT GCAGC CACCC  
2751 CTCAG TGAAA TGCAA GAAAC GGGTC ACCAT CCTGG TGGAG GGAGG AGAGA  
2801 TTGAG CTGTT TGACG GGGAG GTGAA TGTGA AGAGG CCCAT GAAGG ATGAG  
2851 ACTCA CTTTG AGGTG GTGGA GTCTG GCCGG TACAT CATTC TGCTG CTGGG  
2901 CAAAG CCCTC TCCGT GGTCT GGGAG CGCCA CCTGA GCATC TCCGT GGTCC  
2951 TGAAG CAGAC ATACC AGGAG AAAGT GTGTG GCCTG TGTG GAATT TTGAT  
3001 GGCAAT CCAGA ACAAT GACCT CACCA GCAGC AACCT CCAAG TGGAG GAAGA  
3051 CCCTG TGGAC TTTGG GAACT CCTGG AAAGT GAGCT CGCAG TGTGC TGACA  
3101 CCAGA AAAGT GCCTC TGGAC TCATC CCCTG CCACC TGCCA TAACA ACATC  
3151 ATGAA GCAGA CGATG GTGGA TTCCT CCTGT AGAAT CCTTA CCAGT GACGT  
3201 CTTCC AGGAC TGCAA CAAGC TGGTG GACCC CGAGC CATAT CTGGA TGTCT  
3251 GCATT TACGA CACCT GCTCC TGTGA GTCCA TTGGG GACTG CGCCG CATTC  
3301 TGCGA CACCA TTGCT GCCTA TGCCC ACGTG TGTGC CCAGC ATGGC AAGGT  
3351 GGTGA CCTGG AGGAC GGCCA CATTG TGCCC CCAGA GCTGC GAGGA GAGGA  
3401 ATCTC CGGGA GAACG GGTAT GAGGC TGAGT GGCGC TATAA CAGCT GTGCA  
3451 CCTGC CTGTC AAGTC ACGTG TCAGC ACCCT GAGCC ACTGG CCTGC CCTGT  
3501 GCAGT GTGTG GAGGG CTGCC ATGCC CACTG CCCTC CAGGG AAAAT CCTGG  
3551 ATGAG CTTTT GCAGA CCTGC GTTGA CCCTG AAGAC TGTC AGTGT GTGAG  
3601 GTGGC TGGCC GGCGT TTTGC CTCAG GAAAG AAAGT CACCT TGAAT CCCAG  
3651 TGACC CTGAG CACTG CCAGA TTTGC CACTG TGATG TTGTC AACCT CACCT  
3701 GTGAA GCCTG CCAGG AGCCG ATATC GGGCG CGCCA ACATC AGAGA GCGCC  
3751 ACCCC TGAAA GTGGT CCCGG GAGCG AGCCA GCCAC ATCTG GGTCG GAAAC  
3801 GCCAG GCACA AGTGA GTCTG CAACT CCCGA GTCCG GACCT GGCTC CGAGC  
3851 CTGCC ACTAG CGGCT CCGAG ACTCC GGGAA CTTC GAGAG CGCTA CACCA  
3901 GAAAG CGGAC CCGGA ACCAG TACCG AACCT AGCGA GGGCT CTGCT CCGGG  
3951 CAGCC CAGCC GGCTC TCCTA CATCC ACGGA GGAGG GCACT TCCGA ATCCG  
4001 CCACC CCGGA GTCAG GGCCA GGATC TGAAC CCGCT ACCTC AGGCA GTGAG  
4051 ACGCC AGGAA CGAGC GAGTC CGCTA CACCG GAGAG TGGGC CAGGG AGCCC

4101 TGCTG GATCT CCTAC GTCCA CTGAG GAAGG GTCAC CAGCG GGCTC GCCCA  
 4151 CCAGC ACTGA AGAAG GTGCC TCGAG CGGCG GTGGA GGTTT CCGTG GCGGG  
 4201 GGATC CGGTG GCGGG GGATC CGGTG GCGGG GGATC CGGTG GCGGG GGATC  
 4251 CCTGG TCCCC CCGGG CAGCG GAGGC GACAA AACTC ACACA TGCCC ACCGT  
 4301 GCCCA GCTCC AGAAC TCCTG GGCGG ACCGT CAGTC TTCCT CTCC CCCC  
 4351 AAACC CAAGG ACACC CTCAT GATCT CCCGG ACCCC TGAGG TCACA TGCCT  
 4401 GGTGG TGGAC GTGAG CCACG AAGAC CCTGA GGTCA AGTTC AACTG GTACG  
 4451 TGGAC GGCCT GGAGG TGCAT AATGC CAAGA CAAAG CCGCG GGAGG AGCAG  
 [0627] 4501 TACAA CAGCA CGTAC CGTGT GGTCA GCGTC CTCAC CGTCC TGCAC CAGGA  
 4551 CTGGC TGAAT GGCAA GGAGT ACAAG TGCAA GGTCT CCAAC AAAGC CCTCC  
 4601 CAGCC CCCAT CGAGA AAACC ATCTC CAAAG CAAA GGGCA GCCCC GAGAA  
 4651 CCACA GGTGT ACACC CTGCC CCCAT CCCGG GATGA GCTGA CCAAG AACCA  
 4701 GGTCA GCCTG ACCTG CCTGG TCAAA GGCTT CTATC CCAGC GACAT CGCCG  
 4751 TGGAG TGGGA GAGCA ATGGG CAGCC GGAGA ACAAC TACAA GACCA CGCCT  
 4801 CCCGT GTTGG ACTCC GACGG CTCCT TCTTC CTCTA CAGCA AGCTC ACCGT  
 4851 GGACA AGAGC AGGTG GCAGC AGGGG AACGT CTTCT CATGC TCCGT GATGC  
 4901 ATGAG GCTCT GCACA ACCAC TACAC GCAGA AGAGC CTCTC CCTGT CTCCG  
 4951 GGTAA ATGA

[0628] VWF057蛋白质序列 (SEQ ID NO:152)

1 MIPARFAGVL LALALILPGT LCAEGTRGRS STARCSLFGS DFNFTFDGSM  
 51 YSFAGYCSYL LAGGCQKRSF SIIIGDFQNGK RVSLSVYLGE FFDIHLFVNG  
 101 TVTQGDQRVV MPYASKGLYL ETEAGYYKLS GEAYGFVARI DGSGNFQVLL  
 151 SDRYFNKTCG LCGNFNIFAE DDFMTQEGTL TSDPYDFANS WALSSGEQWC  
 201 ERASPPSSC NISSGEMQKG LWEQCQLLKS TSVFARCHPL VDPEPFVALC  
 251 EKTLCECAGG LECACPALLE YARTCAQEGM VLYGWTDHSA CSPVCPAGME  
 301 YRQCVSPCAR TCQSLHINEM CQERCVDGCS CPEGQLLDEG LCVESTCEPC  
 351 VHSBKRYPPG TSLSRDCNTC ICRNSQWICS NEECPGECLV TGQSHFKSFD  
 401 NRYFTFSGIC QYLLARDCQD HSFSIVIVTV QCADDRDAVC TRSVTVRLPG  
 451 LHNSLVKLVK GAGVAMDGQD IQLPLLKGD LRIQHTVTASV RLSYGEDLQM  
 501 DWDGRGRLLV KLSVPYAGKT CGLCGNYNGN QGDDFLTPSG LAEPRVEDFG  
 551 NAWKLHGDCQ DLQKQHS DPC ALNPRMTRFS EEACAVLTSP TFEACHRAVS  
 601 PLPYLRNCRY DVCSCSDGRE CLCGALASYA AACAGRGVRV AWREPRCEL  
 651 NCPKGQVYLQ CGTPCNLTCR SLSYPDEECN EACLEGCFCP PGLYMDERGD  
 701 CVPKAQCPCY YDGEIFQPED IFSDHHTMCY CEDGFMHCTM SGVPGSLLPD  
 751 AVLSSPLSHR SKRSLSCRPP MVKLVCPADN LRAEGLECTK TCQNYDLECM  
 [0629] 801 SMGCVSGCLC PPGMVRHENR CVALERCPCF HQGKEYAPGE TVKIGCNTCV  
 851 CRDRKWNCTD HVCDATCSTI GMAHYLTFDG LKYLFPGECQ YVLVQDYCGS  
 901 NPGTFRILVG NKGCSHPSVK CKKRVTVLVE GGEIELFDGE VNVKRP MKDE  
 951 THFEVVEGR YIIILLGKAL SVVWDRHLSI SVVLKQTYQE KVCGLCGNFD  
 1001 GIQNDLTSS NLQVEEDPVD FGNSWKVSSQ CADTRKVPLD SSPATCHNNI  
 1051 MKQTMVDSSC RILTSDFVQD CNKLVDPPEY LDVCIYDTCS CESIGDCAAF  
 1101 CDTIAAYAHV CAQHGVVVTW RTATLCPQSC EERNLRENGY EAEWRYNSCA  
 1151 PACQVTCQHP EPLACPVCV EGCHAHCPPG KILDELLQTC VDPEDCPVCE  
 1201 VAGRRFASGK KVTLNPSDPE HCQICHCDVV NLTCEACQEP ISGAPTSESA  
 1251 TPESGPGSEP ATSGSETPGT SESATPESGP GSEPATSGSE TPGTSESATP  
 1301 ESGPGTSTEP SEGSA PGSPA GSPTSTEEGT SESATPESGP GSEPATSGSE  
 1351 TPGTSESATP ESGPGSPAGS PTSTEEGSPA GSPTSTEEGA SSGGGGSGGG  
 1401 GSGGGGSGGG GSGGGGSLVP RSGGGDKTHT CPPCPAPELL GGPSVFLFPP  
 1451 KPKDTLMISR TPEVTCVVVD VSHEDPEVKF NWYVDGVEVH NAKTKPREEQ  
 1501 YNSTYRVVSV LTVLHQDWLN GKEYKCKVSN KALPAPIEKT ISKAKQPRE  
 1551 PQVYTLPPSR DELTKNQVSL TCLVKGFYPS DIAVEWESNG QPENNYK TTP  
 1601 PVLDS DGSFF LYSKLTVDKS RWQQGNV FSC SVMHEALHNH YTQKSLSLSP  
 1651 GK\*

[0630] VWF058核苷酸序列 (具有IHH突变的VWF034) (SEQ ID NO:153)

1 ATGAT TCCTG CCAGA TTTGC CGGGG TGCTG CTTGC TCTGG CCCTC ATTTT

51 GCCAG GGACC CTTTG TGCAG AAGGA ACTCG CGGCA GGTCA TCCAC GGCCC

101 GATGC AGCCT TTTTC GAAGT GACTT CGTCA ACACC TTTGA TGGGA GCATG

151 TACAG CTTTG CGGGA TACTG CAGTT ACCTC CTGGC AGGGG GCTGC CAGAA

201 ACGCT CCTTC TCGAT TATTG GGGAC TTCCA GAATG GCAAG AGAGT GAGCC

251 TCTCC GTGTA TCTTG GGGAA TTTTT TGACA TCCAT TTGTT TGTCA ATGGT

301 ACCGT GACAC AGGGG GACCA AAGAG TCTCC ATGCC CTATG CCTCC AAAGG

351 GCTGT ATCTA GAAAC TGAGG CTGGG TACTA CAAGC TGTCC GGTGA GGCCT

401 ATGGC TTTGT GGCCA GGATC GATGG CAGCG GCAAC TTTCA AGTCC TGCTG

451 TCAGA CAGAT ACTTC AACAA GACCT GCGGG CTGTG TGGCA ACTTT AACAT

501 CTTTG CTGAA GATGA CTTTA TGACC CAAGA AGGGA CCTTG ACCTC GGACC

551 CTTAT GACTT TGCCA ACTCA TGGGC TCTGA GCAGT GGAGA ACAGT GGTGT

601 GAACG GGCAT CTCCT CCCAG CAGCT CATGC AACAT CTCCT CTGGG GAAAT

651 GCAGA AGGGC CTGTG GGAGC AGTGC CAGCT TCTGA AGAGC ACCTC GGTGT

701 TTGCC CGCTG CCACC CTCTG GTGGA CCCC GAGCCT TTTGT GGCCC TGTGT

751 GAGAA GACTT TGTGT GAGTG TGCTG GGGGG CTGGA GTGCG CCTGC CCTGC

801 CCTCC TGGAG TACGC CCGGA CCTGT GCCCA GGAGG GAATG GTGCT GTACG

851 GCTGG ACCGA CCACA GCGCG TGCAG CCCAG TGTGC CCTGC TGGA TGGAG

901 TATAG GCAGT GTGTG TCCCC TTGCG CCAGG ACCTG CCAGA GCCTG CACAT

951 CAATG AAATG TGTCA GGAGC GATGC GTGGA TGGCT GCAGC TGCCC TGAGG

1001 GACAG CTCCT GGATG AAGGC CTCTG CGTGG AGAGC ACCGA GTGTC CCTGC

1051 GTGCA TTCCG GAAAG CGCTA CCCTC CCGGC ACCTC CCTCT CTCGA GACTG

1101 CAACA CCTGC ATTTG CCGAA ACAGC CAGTG GATCT GCAGC AATGA AGAAT

1151 GTCCA GGGGA GTGCC TTGTC ACTGG TCAAT CCCAC TTCAA GAGCT TTGAC

1201 AACAG ATACT TCACC TTCAG TGGGA TCTGC CAGTA CCTGC TGGCC CGGGA

1251 TTGCC AGGAC CACTC CTTCT CCATT GTCAT TGAGA CTGTC CAGTG TGCTG

1301 ATGAC CGCGA CGCTG TGTGC ACCCG CTCGG TCACC GTCCG GCTGC CTGGC

1351 CTGCA CAACA GCCTT GTGAA ACTGA AGCAT GGGGC AGGAG TTGCC ATGGA

1401 TGGCC AGGAC ATCCA GCTCC CCCTC CTGAA AGGTG ACCTC CGCAT CCAGC

[0631] 1451 ATACA GTGAC GGCCT CCGTG CCGCT CAGCT ACGGG GAGGA CCTGC AGATG

1501 GACTG GGATG GCCGC GGGAG GCTGC TGGTG AAGCT GTCCC CCGTC TATGC

1551 CGGGA AGACC TGCGG CCTGT GTGGG AATTA CAATG GCAAC CAGGG CGACG

1601 ACTTC CTTAC CCCCT CTGGG CTGGC GGAGC CCCGG GTGGA GGAAT TCGGG

1651 AACGC CTGGA AGCTG CACGG GGAAT GCCAG GACCT GCAGA AGCAG CACAG

1701 CGATC CCTGC GCCCT CAACC CGCGC ATGAC CAGGT TCTCC GAGGA GGCGT

1751 GCGCG GTCCT GACGT CCCCC ACATT CGAGG CCTGC CATCG TGCCG TCAGC

1801 CCGCT GCCCT ACCTG CCGAA CTGCC GCTAC GACGT GTGCT CCTGC TCGGA

1851 CGGCC GCGAG TGCCT GTGCG GCGCC CTGGC CAGCT ATGCC GCGGC CTGCG

1901 CGGGG AGAGG CGTGC GCGTC GCGTG GCGCG AGCCA GGCCG CTGTG AGCTG

1951 AACTG CCCGA AAGGC CAGGT GTACC TGCAAG TGCGG GACCC CCTGC AACCT

2001 GACCT GCCGC TCTCT CTCTT ACCCG GATGA GGAAT GCAAT GAGGC CTGCC

2051 TGGAG GGCTG CTTCT GCGCC CCAGG GCTCT ACATG GATGA GAGGG GGGAC

2101 TGCGT GCCCA AGGCC CAGTG CCGCT CCGCT GTTAC TATGA CGGTG AGATC TTCCA

2151 GCCAG AAGAC ATCTT CTCAG ACCAT CACAC CATGT GCTAC TGTGA GGAAT

2201 GCTTC ATGCA CTGTA CCATG AGTGG AGTCC CCGGA AGCTT GCTGC CTGAC

2251 GCTGT CCTCA GCAGT CCCCT GTCTC ATGCG AGCAA AAGGA GCCTA TCCTG

2301 TCGGC CCCC ATGGT CAAGC TGGTG TGCC CGCTG ACAAC CTGCG GGCTG

2351 AAGGG CTCGA GTGTA CCAA ACGTG CCAGA ACTAT GACCT GGAGT GCATG

2401 AGCAT GGGCT GTGTC TCTGG CTGCC TCTGC CCCCC GGGCA TGGT CCGCA

2451 TGAGA ACAGA TGTGT GGCCC TGGA AGGTG TCCCT GCTTC CATCA GGGCA

2501 AGGAG TATGC CCCTG GAGAA ACAGT GAAGA TTGGC TGCAA CACTT GTGTC

2551 TGTCG GGACC GGAAG TGGAA CTGCA CAGAC CATGT GTGTG ATGCC ACGTG

2601 CTCCA CGATC GGCAT GGCCC ACTAC CTCAC CTTCG ACGGG CTCAA ATACC

2651 TGTTT CCCGG GGAGT GCCAG TACGT TCTGG TGCAG GATTA CTGCG GCAGT

2701 AACCC TGGGA CCTTT CGGAT CCTAG TGGGG AATAA GGGAT GCAGC CACCC

2751 CTCAG TGAAA TGCAA GAAAC GGGTC ACCAT CCTGG TGGAG GGAGG AGAGA

2801 TTGAG CTGTT TGACG GGGAG GTGAA TGTGA AGAGG CCCAT GAAGG ATGAG

2851 ACTCA CTTTG AGGTG GTGGA GTCTG GCCGG TACAT CATTC TGCTG CTGGG

2901 CAAAG CCCTC TCCGT GGTCT GGGAC CGCCA CCTGA GCATC TCCGT GGTCC

2951 TGAAG CAGAC ATACC AGGAG AAAGT GTGTG GCCTG TGTGG GAATT TTGAT

3001 GGCAT CCAGA ACAAT GACCT CACCA GCAGC AACCT CCAAG TGGAG GAAGA  
 3051 CCCTG TGGAC TTTGG GAACT CCTGG AAAGT GAGCT CGCAG TGTGC TGACA  
 3101 CCAGA AAAGT GCCTC TGGAC TCATC CCCTG CCACC TGCCA TAACA ACATC  
 3151 ATGAA GCAGA CGATG GTGGA TTCCT CCTGT AGAAT CCTTA CCAGT GACGT  
 3201 CTTCC AGGAC TGCAA CAAGC TGGTG GACCC CGAGC CATAT CTGGA TGTCT  
 3251 GCATT TACGA CACCT GCTCC TGTGA GTCCA TTGGG GACTG CGCCG CATTC  
 3301 TGCGA CACCA TTGCT GCCTA TGCCC ACGTG TGTGC CCAGC ATGGC AAGGT  
 3351 GGTGA CCTGG AGGAC GGCCA CATTG TGCCC CCAGA GCTGC GAGGA GAGGA  
 3401 ATCTC CGGGA GAACG GGTAT GAGGC TGAGT GGCGC TATAA CAGCT GTGCA  
 3451 CCTGC CTGTC AAGTC ACGTG TCAGC ACCCT GAGCC ACTGG CCTGC CCTGT  
 3501 GCAGT GTGTG GAGGG CTGCC ATGCC CACTG CCCTC CAGGG AAAAT CCTGG  
 3551 ATGAG CTTTT GCAGA CCTGC GTTGA CCCTG AAGAC TGTCC AGTGT GTGAG  
 3601 GTGGC TGGCC GGCCT TTTGC CTCAG GAAAG AAAGT CACCT TGAAT CCCAG  
 3651 TGACC CTGAG CACTG CCAGA TTTGC CACTG TGATG TTGTC AACCT CACCT  
 3701 GTGAA GCCTG CCAGG AGCCG ATATC GGGTA CCTCA GAGTC TGCTA CCCCC  
 3751 GAGTC AGGGC CAGGA TCAGA GCCAG CCACC TCCGG GTCTG AGACA CCCGG  
 3801 GACTT CCGAG AGTGC CACCC CTGAG TCCGG ACCCG GGTCC GAGCC CGCCA  
 3851 CTTCC GGCTC CGAAA CTCCC GGCAC AAGCG AGAGC GCTAC CCCAG AGTCA  
 3901 GGACC AGGAA CATCT ACAGA GCCCT CTGAA GGCTC CGCTC CAGGG TCCCC  
 3951 AGCCG GCAGT CCCAC TAGCA CCGAG GAGGG AACCT CTGAA AGCGC CACAC  
 4001 CCGAA TCAGG GCCAG GGTCT GAGCC TGCTA CCAGC GGCAG GCAGA CACCA  
 4051 GGCAC CTCTG AGTCC GCCAC ACCAG AGTCC GGACC CGGAT CTCCC GCTGG  
 4101 GAGCC CCACC TCCAC TGAGG AGGGA TCTCC TGCTG GCTCT CCAAC ATCTA  
 [0632] 4151 CTGAG GAAAG TACCT CAACC GAGCC ATCCG AGGGA TCAGC TCCCG GCACC  
 4201 TCAGA GTCGG CAACC CCGGA GTCTG GACCC GGAAC TTCCG AAAGT GCCAC  
 4251 ACCAG AGTCC GGTCC CGGGA CTTCA GAATC AGCAA CACCC GAGTC CGGCC  
 4301 CTGGG TCTGA ACCCG CCACA AGTGG TAGTG AGACA CCAGG ATCAG AACCT  
 4351 GCTAC CTCAG GGTCA GAGAC ACCCG GATCT CCGGC AGGCT CACCA ACCTC  
 4401 CACTG AGGAG GGCAC CAGCA CAGAA CCAAG CGAGG GCTCC GCACC CGGAA  
 4451 CAAGC ACTGA ACCCA GTGAG GGTTC AGCAC CCGGC TCTGA GCCGG CCACA  
 4501 AGTGG CAGTG AGACA CCCGG CACTT CAGAG AGTGC CACCC CCGAG AGTGG  
 4551 CCCAG GCACT AGTAC CGAGC CCTCT GAAGG CAGTG CGCCA GATTC TGCCG  
 4601 GTGGA GGTTC CGGTG GCGGG GGATC CGGTG GCGGG GGATC CGGTG GCGGG  
 4651 GGATC CGGTG GCGGG GGATC CCTGG TCCCC CCGGG CAGCG GAGGC GACAA  
 4701 AACTC ACACA TGCCC ACCGT GCCCA GCTCC AGAAC TCCTG GGCGG ACCGT  
 4751 CAGTC TTCCT CTTCC CCCCC AAACC CAAGG ACACC CTCAT GGCTT CCCGG  
 4801 ACCCC TGAGG TCACA TGCCT GGTGG TGGAC GTGAG CCACG AAGAC CCTGA  
 4851 GGTCA AGTTC AACTG GTACG TGGAC GGCCT GGAGG TGCAT AATGC CAAGA  
 4901 CAAAG CCGCG GGAGG AGCAG TACAA CAGCA CGTAC CGTGT GGTCA GCGTC  
 4951 CTCAC CGTCC TGGCC CAGGA CTGGC TGAAT GGCAA GGAGT ACAAG TGCAA  
 5001 GGTCT CCAAC AAAGC CCTCC CAGCC CCCAT CGAGA AAACC ATCTC CAAAG  
 5051 CCAAA GGGCA GCCCC GAGAA CCACA GGTGT ACACC CTGCC CCCAT CCGC  
 5101 GATGA GCTGA CCAAG AACCA GGTCA GCCTG ACCTG CCTGG TCAAA GGCTT  
 5151 CTATC CCAGC GACAT CGCCG TGGAG TGGGA GAGCA ATGGG CAGCC GGAGA  
 5201 ACAAC TACAA GACCA CGCCT CCCGT GTTGG ACTCC GACGG CTCCT TCTTC  
 5251 CTCTA CAGCA AGCTC ACCGT GGACA AGAGC AGGTG GCAGC AGGGG AACGT  
 5301 CTTCT CATGC TCCGT GATGC ATGAG GCTCT GCACA ACGCC TACAC GCAGA  
 5351 AGAGC CTCTC CCTGT CTCCG GGTAA ATGA

[0633] VWF058蛋白质序列(具有IHH突变的VWF034) (SEQ ID NO:154)

```

1     MIPARFAGVL LALALILPGT LCAEGTRGRS STARCSLFGS DFNNTFDGSM
51    YSFAGYCSYL LAGGCQKRSF SIIGDFQNGK RVSLSVYLGE FFDIHLFVNG
101   TVTQGDQRVS MPYASKGLYL ETEAGYYKLS GEAYGFVARI DGSGNFQVLL
151   SDRYFNKTCG LCGNFNIFAE DDFMTQEGTL TSDPYDFANS WALSSGEQWC
201   ERASPPSSSC NISSGEMQKG LWEQCQLLKS TSVFARCHPL VDPEPFVALC
251   EKTLCCECAGG LECACPALLE YARTCAQEGM VLYGWTDHSA CSPVCPAGME
301   YRQCVSPCAR TCQSLHINEM CQERCVDGCS CPEGQLLDEG LCVESTTECPC

351   VHSGKRYPPG TSLSRDCNTC ICRNSQWICS NEECPGECLV TGQSHFKSFD
401   NRYFTFSGIC QYLLARDCQD HSFSIVIETV QCADDRDAVC TRSVTVRLPG
451   LHNSLVKCLKH GAGVAMDGQD IQLPLLKGD LRIQHTVTASV RLSYGEDLQM
501   DWDGRGRLLV KLSVPYAGKT CGLCGNYNGN QGDDFLTPSG LAEPRVEDFG
551   NAWKLGDCQ DLQKQHSDFC ALNPRMTRFS EEACAVLTSP TFEACHRAVS
601   PLPYLRNCRY DVCSCSDGRE CLCGALASYA AACAGRGVRV AWREPGRCEL
651   NCPKGQVYLQ CGTPCNLTCR SLSYPDEECN EACLEGCFCP PGLYMDERGD
701   CVPKAQCPCY YDGEIFQPED IFSDHHTMCY CEDGFMHCTM SGVPGSLLPD
751   AVLSSPLSHR SKRSLSCRPP MVKLVCPADN LRAEGLECTK TCQNYDLECM
801   SMGCVSGCLC PPGMVRHENR CVALERCPCF HQGKEYAPGE TVKIGCNTCV
[0634] 851   CRDRKWNCTD HVCDATCSTI GMAHYLTFDG LKYLFPGEQ YVLVQDYCGS
901   NPGTFRILVG NKGCSHPSVK CKKRVTILVE GGEIELFDGE VNVKRPMKDE
951   THFEVVEGR YIILLGKAL SVVWDRHLSI SVVLKQTYQE KVCGLCGNFD
1001  GIQNNDLTSS NLQVEEDPVD FGNSWKVSSQ CADTRKVPLD SSPATCHNNI
1051  MKQTMVDSSC RILTSDFVQD CNKLVDPPEY LDVCIYDTCS CESIGDCAAF
1101  CDTIAAYAHV CAQHGVVVTW RTATLCPQSC EERNLRENGY EAEWRYNCSA
1151  PACQVTCQHP EPLACPVQCV EGCHAHCPPG KILDELLQTC VDPEDCPVCE
1201  VAGRRFASGK KVTLNPSDPE HCQICHCDVV NLTCEACQEP ISGTSESATP
1251  ESGPGSEPAT SGSETPGTSE SATPESGPGS EPATSGSETP GTSESATPES
1301  GPGTSTEPSE GSAPGSPAGS PTSTEEGTSE SATPESGPGS EPATSGSETP
1351  GTSESATPES GPGSPAGSPT STEEGSPAGS PTSTEEGTST EPSEGSAPGT
1401  SESATPESGP GTSESATPES GPGTSESATP ESGPGSEPAT SGSETPGSEP
1451  ATSGSETPGS PAGSPTSTEE GTSTEPSEGS APGTSTEPSE GSAPGSEPAT
1501  SGSETPGTSE SATPESGPGT STEPSEGSAP DSGGGGSGGG GSGGGGSGGG
1551  GSGGGGSLVP RGSAGDKTHT CPPCPAPELL GGPSVFLFPP KPKDTLMASR
1601  TPEVTCVVVD VSHEDPEVKF NWYVDGVEVH NAKTKPREEQ YNSTYRVVSV
1651  LTVLAQDWLN GKEYKCKVSN KALPAPIEKT ISKAKQPRE PQVYTLPPSR
1701  DELTKNQVSL TCLVKGFYPS DIAVEWESNG QPENNYKTP PVLDSDGSFF
1751  LYSKLTVDKS RWQQGNVFC SVMHEALHNA YTQKSLSLSP GK*

[0635] FVIII 169核苷酸序列(SEQ ID NO:155)

```

[0636]

```

1   ATGCA AATAG AGCTC TCCAC CTGCT TCTTT CTGTG CCTTT TGC GA TTCTG
51  CTTTA GTGCC ACCAG AAGAT ACTAC CTGGG TGCAG TGGAA CTGTC ATGGG
101 ACTAT ATGCA AAGTG ATCTC GGTGA GCTGC CTGTG GACGC AAGAT TTCCT
151 CCTAG AGTGC CAAAA TCTTT TCCAT TCAAC ACCTC AGTCG TGTAC AAAAA
201 GACTC TGTTT GTAGA ATTCA CGGAT CACCT TTTCA ACATC GCTAA GCCAA
251 GGCCA CCCTG GATGG GTCTG CTAGG TCCTA CCATC CAGGC TGAGG TTTAT
301 GATAC AGTGG TCATT ACACT TAAGA ACATG GCTTC CCATC CTGTC AGTCT
351 TCATG CTGTT GGTGT ATCCT ACTGG AAAGC TTCTG AGGGA GCTGA ATATG
401 ATGAT CAGAC CAGTC AAAGG GAGAA AGAAG ATGAT AAAGT CTTCC CTGGT
451 GGAAG CCATA CATAT GTCTG GCAGG TCCTG AAAGA GAATG GTCCA ATGGC
501 CTCTG ACCCA CTGTG CCTTA CCTAC TCATA TCTTT CTCAT GTGGA CCTGG
551 TAAAA GACTT GAATT CAGGC CTCAT TGGAG CCCTA CTAGT ATGTA GAGAA
601 GGGAG TCTGG CCAAG GAAAA GACAC AGACC TTGCA CAAAT TTATA CTACT
651 TTTTG CTGTA TTTGA TGAAG GGAAA AGTTG GCACT CAGAA ACAA GAACT
701 CCTTG ATGCA GGATA GGGAT GCTGC ATCTG CTCGG GCCTG GCCTA AAATG
751 CACAC AGTCA ATGGT TATGT AAACA GGTCT CTGCC AGGTC TGATT GGATG
801 CCACA GGAAA TCAGT CTATT GGCAT GTGAT TGGAA TGGGC ACCAC TCCTG
851 AAGTG CACTC AATAT TCCTC GAAGG TCACA CATTT CTTGT GAGGA ACCAT
901 CGCCA GGCTA GCTTG GAAAT CTCGC CAATA ACTTT CCTTA CTGCT CAAAC
951 ACTCT TGATG GACCT TGGAC AGTTT CTACT GTTTT GTCAT ATCTC TTCCC
1001 ACCAA CATGA TGGCA TGGAA GCTTA TGTCA AAGTA GACAG CTGTC CAGAG
1051 GAACC CCAAC TACGA ATGAA AAATA ATGAA GAAGC GGAAG ACTAT GATGA
1101 TGATC TTA CT GATTG TGA AAA TGGAT GTGGT CAGGT TTGAT GATGA CAACT
1151 TCCTT TCCTT TATCC AAATT CGCTC AGTTG CCAAG AAGCA TCCTA AA ACT
1201 TGGGT ACATT ACATT GCTGC TGAAG AGGAG GACTG GGACT ATGCT CCCTT
1251 AGTCC TCGCC CCCGA TGACA GAAGT TATAA AAGTC AATAT TTGAA CAATG

```

1301 GCCCT CAGCG GATTG GTAGG AAGTA CAAAA AAGTC CGATT TATGG CATACT  
1351 ACAGA TGAAA CCTTT AAGAC TCGTG AAGCT ATTCA GCATG AATCA GGAAT  
1401 CTTGG GACCT TTACT TTATG GGGAA GTTGG AGACA CACTG TTGAT TATAT  
1451 TTAAG AATCA AGCAA GCAGA CCATA TAACA TCTAC CCTCA CGGAA TCACT  
1501 GATGT CCGTC CTTTG TATTC AAGGA GATTA CAAA AGGTG TAAA CATT  
1551 GAAGG ATTTT CCAAT TCTGC CAGGA GAAAT ATTCA AATAT AAATG GACAG  
1601 TGACT GTAGA AGATG GGCCA ACTAA ATCAG ATCCT CGGTG CCTGA CCCGC  
1651 TATTA CTCTA GTTTC GTTAA TATGG AGAGA GATCT AGCTT CAGGA CTCAT  
1701 TGGCC CTCTC CTCAT CTGCT ACAA GAATC TGTAG ATCAA AGAGG AAACC  
1751 AGATA ATGTC AGACA AGAGG AATGT CATCC TGTTT TCTGT ATTTG ATGAG  
1801 AACCG AAGCT GGTAC CTCAC AGAGA ATATA CAACG CTTTC TCCCC AATCC  
1851 AGCTG GAGTG CAGCT TGAGG ATCCA GAGTT CCAAG CCTCC AACAT CATGC  
1901 ACAGC ATCAA TGGCT ATGTT TTTGA TAGTT TGCAG TTGTC AGTTT GTTTG  
1951 CATGA GGTGG CATACT TGGTA CATTG TAAGC ATTGG AGCAC AGACT GACTT  
2001 CCTTT CTGTC TTCTT CTCTG GATAT ACCTT CAAAC ACAA ATGGT CTATG  
2051 AAGAC AACT CACCC TATTC CCATT CTCAG GAGAA ACTGT CTTCA TGTCG  
2101 ATGGA AAACC CAGGT CTATG GATTC TGGGG TGCCA CAACT CAGAC TTTCG  
2151 GAACA GAGGC ATGAC CGCCT TACTG AAGGT TTCTA GTTGT GACAA GAACA  
2201 CTGGT GATTA TTACG AGGAC AGTTA TGAAG ATATT TCAGC ATACT TGCTG  
2251 AGTAA AAACA ATGCC ATTGA ACCAA GAAGC TTCTC TCAA ACCGC GCGCC  
2301 AGGTA CCTCA GAGTC TGCTA CCCCC GAGTC AGGGC CAGGA TCAGA GCCAG  
2351 CCACC TCCGG GTCTG AGACA CCGG GACTT CCGAG AGTGC CACCC CTGAG  
2401 TCCGG ACCCG GGTCC GAGCC CGCCA CTTC GGCTC CGAAA CTCCC GGCAC  
2451 AAGCG AGAGC GCTAC CCCAG AGTCA GGACC AGGAA CATCT ACAGA GCCCT  
2501 CTGAA GGCTC CGCTC CAGGG TCCCC AGCCG GCAGT CCCAC TAGCA CCGAG  
2551 GAGGG AACCT CTGAA AGCGC CACAC CCGAA TCAGG GCCAG GGCTT GAGCC  
2601 TGCTA CCAGC GGCAG CGAGA CACCA GGCAC CTCTG AGTCC GCCAC ACCAG  
2651 AGTCC GGACC CGGAT CTCCC GCTGG GAGCC CCACC TCCAC TGAGG AGGGA  
2701 TCTCC TGCTG GCTCT CCAAC ATCTA CTGAG GAAGG TACCT CAACC GAGCC  
[0637] 2751 ATCCG AGGGA TCAGC TCCCG GCACC TCAGA GTCGG CAACC CCGGA GTCTG  
2801 GACCC GGAAC TTCCG AAAGT GCCAC ACCAG AGTCC GGTCC CGGGA CTTCA  
2851 GAATC AGCAA CACCC GAGTC CGGCC CTGGG TCTGA ACCCG CCACA AGTGG  
2901 TAGTG AGACA CCAGG ATCAG AACCT GCTAC CTCAG GGTC A GAGAC ACCCG  
2951 GATCT CCGGC AGGCT CACCA ACCTC CACTG AGGAG GGCAC CAGCA CAGAA  
3001 CCAAG CGAGG GCTCC GCACC CGGAA CAAGC ACTGA ACCCA GTGAG GGTT  
3051 AGCAC CCGGC TCTGA GCCGG CCACA AGTGG CAGTG AGACA CCGGC CACTT  
3101 CAGAG AGTGC CACCC CCGAG AGTGG CCCAG GCACT AGTAC CGAGC CCTCT  
3151 GAAGG CAGTG GCCCA GCCTC GAGCC CACCA GTCTT GAAAC GCCAT CAAAG  
3201 TGAAA TAACT CGTAC TACTC TTCAG TCAGA TCAAG AGGAA ATCGA TTATG  
3251 ATGAT ACCAT ATCAG TTGAA ATGAA GAAGG AAGAT TTTGA CATT ATGAT  
3301 GAGGA TGAAA ATCAG AGCCC CCGCA GCTTT CAAA GAAA CACGA CACTA  
3351 TTTTA TTGCT GCAGT GGAGA GGCTC TGGGA TTATG GGATG AGTAG CTCCC  
3401 CACAT GTTCT AAGAA ACAGG GCTCA GAGTG GCAGT GTCCC TCAGT TCAAG  
3451 AAAGT TGTTT TCCAG GAATT TACTG ATGGC TCCTT TACTC AGCCC TTATA  
3501 CCGTG GAGAA CTAAG TGAAC ATTTG GGACT CCTGG GGCCA TATAT AAGAG  
3551 CAGAA GTTGA AGATA ATATC ATGGT AACTT TCAGA AATCA GGCC CTCTG  
3601 CCCTA TTCCT TCTAT TCTAG CCTTA TTTCT TATGA GGAAG ATCAG AGGCA  
3651 AGGAG CAGAA CCTAG AAAAA ACTTT GTCAA GCCTA ATGAA ACCAA AACTT  
3701 ACTTT TGGA AGTGC AACAT CATAT GGCAC CCACT AAAGA TGAGT TTGAC  
3751 TGCAA AGCCT GGGCT TATT CTCTG ATGTT GACCT GGAAA AAGAT GTGCA  
3801 CTCAG GCCTG ATTGG ACCCC TTCTG GTCTG CCACA CTAAC ACAC GAACC  
3851 CTGCT CATGG GAGAC AAGTG ACAGT ACAGG AATTT GCTCT GTTT TCACC  
3901 ATCTT TGATG AGACC AAAAG CTGGT ACTTC ACTGA AAATA TGGAA AGAAA  
3951 CTGCA GGGCT CCCTG CAATA TCCAG ATGGA AGATC CCACT TTAA AGAGA  
4001 ATTAT CGCTT CCATG CAATC AATGG CTACA TAATG GATAC ACTAC CTGGC  
4051 TTAGT AATGG CTCAG GATCA AAGGA TTCGA TGGTA TCTGC TCAGC ATGGG  
4101 CAGCA ATGAA AACAT CCATT CTATT CATT CAGTG GACAT GTGTT CACTG  
4151 TACGA AAAAA AGAGG AGTAT AAAAT GGCAC TGTAC AATCT CTATC CAGGT  
4201 GTTTT TGAGA CAGTG GAAAT GTTAC CATCC AAAGC TGGA TTTGG CGGGT



4251 GGAAT GCCTT ATTGG CGAGC ATCTA CATGC TGGGA TGAGC ACACT TTTTC  
 4301 TGGTG TACAG CAATA AGTGT CAGAC TCCCC TGGGA ATGGC TTCTG GACAC  
 4351 ATTAG AGATT TTCAG ATTAC AGCTT CAGGA CAATA TGGAC AGTGG GCCCC  
 4401 AAAGC TGGCC AGACT TCATT ATTCC GGATC AATCA ATGCC TGGAG CACCA  
 4451 AGGAG CCCTT TTCTT GGATC AAGGT GGATC TGTTG GCACC AATGA TTATT  
 4501 CACGG CATCA AGACC CAGGG TGCCC GTCAG AAGTT CTCCA GCCTC TACAT  
 4551 CTCTC AGTTT ATCAT CATGT ATAGT CTTGA TGGGA AGAAG TGGCA GACTT  
 4601 ATCGA GGAAA TTCCA CTGGA ACCTT AATGG TCTTC TTTGG CAATG TGGAT  
 4651 TCATC TGGGA TAAAA CACAA TATTT TTAAC CCTCC AATTA TTGCT CGATA  
 4701 CATCC GTTTG CACCC AACTC ATTAT AGCAT TCGCA GCACT CTTCG CATGG  
 4751 AGTTG ATGGG CTGTG ATTTA AATAG TTGCA GCATG CCATT GGGAA TGGAG  
 4801 AGTAA AGCAA TATCA GATGC ACAGA TTACT GCTTC ATCCT ACTTT ACCAA  
 4851 TATGT TTGCC ACCTG GTCTC CTTCA AAAGC TCGAC TTCAC CTCCA AGGGA  
 4901 GGAGT AATGC CTGGA GACCT CAGGT GAATA ATCCA AAAGA GTGGC TGCAA  
 4951 GTGGA CTTCC AGAAG ACAAT GAAAG TCACA GGAGT AACTA CTCAG GGAGT  
 5001 AAAAT CTCTG CTTAC CAGCA TGTAT GTGAA GGAGT TCCTC ATCTC CAGCA  
 [0638] 5051 GTCAA GATGG CCATC AGTGG ACTCT CTTTT TTCAG AATGG CAAAG TAAAG  
 5101 GTTTT TCAGG GAAAT CAAGA CTCCT TCACA CCTGT GGTGA ACTCT CTAGA  
 5151 CCCAC CGTTA CTGAC TCGCT ACCTT CGAAT TCACC CCCAG AGTTG GGTGC  
 5201 ACCAG ATTGC CCTGA GGATG GAGGT TCTGG GCTGC GAGGC ACAGG ACCTC  
 5251 TACGA CAAA CTCAC ACATG CCCAC CGTGC CCAGC TCCAG AACTC CTGGG  
 5301 CGGAC CGTCA GTCTT CCTCT TCCCC CAAA ACCCA AGGAC ACCCT CATGA  
 5351 TCTCC CGGAC CCCTG AGGTC ACATG CGTGG TGGTG GACGT GAGCC ACGAA  
 5401 GACCC TGAGG TCAAG TTCAA CTGGT ACGTG GACGG CGTGG AGGTG CATAA  
 5451 TGCCA AGACA AAGCC GCGGG AGGAG CAGTA CAACA GCACG TACCG TGTGG  
 5501 TCAGC GTCTT CACCG TCCTG CACCA GGACT GGCTG AATGG CAAGG AGTAC  
 5551 AAGTG CAAGG TCTCC AACAA AGCCC TCCCA GCCCC CATCG AGAAA ACCAT  
 5601 CTCCA AAGCC AAAGG GCAGC CCCGA GAACC ACAGG TGTAC ACCCT GCCCC  
 5651 CATCC CGGGA TGAGC TGACC AAGAA CCAGG TCAGC CTGAC CTGCC TGGTC  
 5701 AAAGG CTTCT ATCCC AGCGA CATCG CCGTG GAGTG GGAGA GCAAT GGGCA  
 5751 GCCGG AGAAC AACTA CAAGA CCACG CCTCC CGTGT TGGAC TCCGA CGGCT  
 5801 CCTTC TTCCT CTACA GCAAG CTCAC CGTGG ACAAG AGCAG GTGGC AGCAG  
 5851 GGGAA CGTCT TCTCA TGCTC CGTGA TGCAT GAGGC TCTGC ACAAC CACTA  
 5901 CACGC AGAAG AGCCT CTCCC TGTCT CCGGG TAAAT GA

[0639] FVIII 169蛋白质序列 (SEQ ID NO:70)

1	MQIELSTCFF	LCLLRFCFSA	TRRYYLGAVE	LSWDYMQSDL	GELPVDARFP
51	PRVPKSFPFN	TSVVYKKTFL	VEFTDHLFNI	AKPRPPWMGL	LGPTIQAEVY
101	DTVVITLKNM	ASHPVSLHAV	GVSYWKASEG	AEYDDQTSQR	EKEDDKVFPG
151	GSHTYVWQVL	KENGPMA SDP	LCLTYSYLSH	VDLVKDLNSG	LIGALLVCRE
201	GSLAKEKTQT	LHKFILLFAV	FDEGKSWHSE	TKNSLMQDRD	AASARAWPKM
251	HTVNGYVNRS	LPGLIGCHRK	SVYWHVIGMG	TTPEVHSIFL	EGHTFLVRNH
301	RQASLEISPI	TFLTAQTLLM	DLGQFLLFCH	ISSHQHDGME	AYVKVDSCPE
351	EPQLRMKNNE	EAEDYDDDLT	DSEMDVVRFD	DDNSPSFIQI	RSVAKKHPKT
401	WVHYIAAEEE	DWDYAPLVLA	PDRSYKSQY	LNNGPQRIGR	KYKKVRFMAY
451	TDETFKTREA	IQHESGILGP	LLYGEVGDTL	LIIFKNQASR	PYNIYPHGIT
501	DVRPLYSRRL	PKGVKHLKDF	PILPGEIFKY	KWTVTVEDGP	TKSDPRCLTR
551	YYSSFVNMER	DLASGLIGPL	LICYKESVDQ	RGNQIMSDKR	NVILFSVFDE
601	NRSWYL TENI	QRFLPNPAGV	QLEDPEFQAS	NIMHSINGYV	FDSLQLSVCL
651	HEVAYWYILS	IGAQTDFLSV	FFSGYTFKHK	MVYEDTLTLF	PFSGETVFMS
701	MENPGLWILG	CHNSDFRNRG	MTALLKVSSC	DKNTGDY YED	SYEDISAYLL
751	SKNNAIEPRS	FSQNGAPGTS	ESATPESGPG	SEPATSGSET	PGTSESATPE
801	SGPGSEPATS	GSETPGTSES	ATPESGPGTS	TEPSEGSAPG	SPAGSPTSTE
851	EGTSESATPE	SGPGSEPATS	GSETPGTSES	ATPESGPGSP	AGSPTSTEEG
901	SPAGSPTSTE	EGTSTEPSEG	SAPGTSESAT	PESGPGTSES	ATPESGPGTS
951	ESATPESGPG	SEPATSGSET	PGSEPATSGS	ETPGSPAGSP	TSTEEGTSTE
[0640] 1001	PSEGSAPGTS	TEPSEGSAPG	SEPATSGSET	PGTSESATPE	SGPGTSTEPS
1051	EGSAPASSPP	VLKRHQAEIT	RTTLQSDQEE	IDYDDTISVE	MKKEDFDIYD
1101	EDENQSPRSF	QKKTRHYFIA	AVERLWDYGM	SSSPHVLNRN	AQSGSVPQFK
1151	KVVVFQFTDG	SFTQPLYRGE	LNEHLGLLGP	YIRAEVEDNI	MVTFRNQASR
1201	PYSFYSSLIS	YEEDQRQGAE	PRKNFVKPNE	TKTYFWKVQH	HMAPTKDEFD
1251	CKAWAYFSDV	DLEKDVHSGL	IGPLLVCHTN	TLNPAHGRQV	TVQEFALFFT
1301	IFDETKSWYF	TENMERN CRA	PCNIQMEDPT	FKENYRFHAI	NGYIMDTLPG
1351	LVMAQDQRIR	WYLLSMGSNE	NIHSIHFSGH	VFTVRKKEEY	KMALYNLYPG
1401	VFETVEMLPS	KAGIWRVECL	IGEHLHAGMS	TLFLVYSNKC	QTPLGMASGH
1451	IRDFQITASG	QYGQWAPKLA	RLHYSGSINA	WSTKEPFSWI	KVDLLAPMII
1501	HGIKTQGARQ	KFSSLYISQF	IIMYSLDGKK	WQTYRGNSTG	TLMVFFGNVD
1551	SSGIKHNIFN	PPIIARYIRL	HPTHYSIRST	LRMELMGCDL	NSCSMPLGME
1601	SKAISDAQIT	ASSYFTNMFA	TWSPSKARLH	LQGRSNAWRP	QVNNPKEWLQ
1651	VDFQKTMKVT	GVTTQGVKSL	LTSMYVKEFL	ISSSQDGHQW	TLFFQNGKVK
1701	VFQGNQDSFT	PVVNSLDPPL	LTRYLRIHPQ	SWVHQIALRM	EVLGCEAQDL
1751	YDKTHTCPPC	PAPELLGGPS	VFLFPPKPKD	TLMISRTPEV	TCVVVDVSHE
1801	DPEVKFNWYV	DGVEVHNAKT	KPREEQYNST	YRVVSVLTVL	HQDWLNGKEY
1851	KCKVSNKALP	APIEKTISKA	KGQPREPQVY	TLPPSRDEL T	KNQVSLTCLV
1901	KGFYPSDIAV	EWESNGQPEN	NYKTPPVLD	SDGSFFLYSK	LTVDKSRWQQ
1951	GNVFSCSVMH	EALHNHYTQK	SLSLSPGK*		

[0641] FVIII 263核苷酸序列(IHH三重突变体)(SEQ ID NO:156)

[0642]

```

1      ATGCA AATAG AGCTC TCCAC CTGCT TCTTT CTGTG CCTTT TGC GA TTCTG
51     CTTTA GTGCC ACCAG AAGAT ACTAC CTGGG TGCAG TGGAA CTGTC ATGGG
101    ACTAT ATGCA AGGCG CGCCA ACATC AGAGA GCGCC ACCCC TGAAA GTGGT
151    CCCGG GAGCG AGCCA GCCAC ATCTG GGTCG GAAAC GCCAG GCACA AGTGA
201    GTCTG CAACT CCCGA GTCCG GACCT GGCTC CGAGC CTGCC ACTAG CGGCT
251    CCGAG ACTCC GGGAA CTTCC GAGAG CGCTA CACCA GAAAG CGGAC CCGGA
301    ACCAG TACCG AACCT AGCGA GGGCT CTGCT CCGGG CAGCC CAGCC GGCTC
351    TCCTA CATCC ACGGA GGAGG GCACT TCCGA ATCCG CCACC CCGGA GTCAG
401    GGCCA GGATC TGAAC CCGCT ACCTC AGGCA GTGAG ACGCC AGGAA CGAGC
451    GAGTC CGCTA CACCG GAGAG TGGGC CAGGG AGCCC TGCTG GATCT CCTAC
501    GTCCA CTGAG GAAGG GTCAC CAGCG GGCTC GCCCA CCAGC ACTGA AGAAG
551    GTGCC TCGAG CAGTG ATCTC GGTGA GCTGC CTGTG GACGC AAGAT TTCCT
601    CCTAG AGTGC CAAAA TCTTT TCCAT TCAAC ACCTC AGTCG TGTAC AAAAA
651    GACTC TGTTT GTAGA ATTCA CGGAT CACCT TTTCA ACATC GCTAA GCCAA
701    GGCCA CCCTG GATGG GTCTG CTAGG TCCTA CCATC CAGGC TGAGG TTTAT
751    GATAC AGTGG TCATT ACACT TAAGA ACATG GCTTC CCATC CTGTC AGTCT
801    TCATG CTGTT GGTGT ATCCT ACTGG AAAGC TTCTG AGGGA GCTGA ATATG
851    ATGAT CAGAC CAGTC AAAGG GAGAA AGAAG ATGAT AAAGT CTTCCT CTGGT
901    GGAAG CCATA CATAT GTCTG GCAGG TCCTG AAAGA GAATG GTCCA ATGGC
951    CTCTG ACCCA CTGTG CCTTA TCATA TCTTT CTCAT CTGTA CTTGG
1001   TAAAA GACTT GAATT CAGGC CTCAT TGGAG CCCTA CTAGT ATGTA GAGAA
1051   GGGAG TCTGG CCAAG GAAAA GACAC AGACC TTGCA CAAAT TTATA CTACT
1101   TTTTG CTGTA TTTGA TGAAG GGAAA AGTTG GCACT CAGAA ACAA GAACT
1151   CCTTG ATGCA GGATA GGGAT GCTGC ATCTG CTCGG GCCTG GCCTA AAATG
1201   CACAC AGTCA ATGGT TATGT AAACA GGTCT CTGCC AGGTC TGATT GGATG
1251   CCACA GGAAA TCAGT CTATT GGCAT GTGAT TGGAA TGGGC ACCAC TCCTG
1301   AAGTG CACTC AATAT TCCTC GAAGG TCACA CATTT CTTGT GAGGA ACCAT
1351   CGCCA GGCTA GCTTG GAAAT CTCGC CAATA ACTTT CCTTA CTGCT CAAAC
1401   ACTCT TGATG GACCT TGGAC AGTTT CTACT GTTTT GTCAT ATCTC TTCCC
1451   ACCAA CATGA TGGCA TGGAA GCTTA TGTCA AAGTA GACAG CTGTC CAGAG
1501   GAACC CCAAC TACGA ATGAA AAATA ATGAA GAAGC GGAAG ACTAT GATGA
1551   TGATC TTA CT GATTG TGAAG TGGAT GTGGT CAGGT TTGAT GATGA CAACT
1601   CTCTT TCCTT TATCC AAATT CGCTC AGTTG CCAAG AAGCA TCCTA AA ACT
1651   TGGGT ACATT ACATT GCTGC TGAAG AGGAG GACTG GGACT ATGCT CCCTT
1701   AGTCC TCGCC CCCGA TGACA GAAGT TATAA AAGTC AATAT TTGAA CAATG
1751   GCCCT CAGCG GATTG GTAGG AAGTA CAAAA AAGTC CGATT TATGG CATA C
1801   ACAGA TGAAA CCTTT AAGAC TCGTG AAGCT ATTCA GCATG AATCA GGAAT
1851   CTTGG GACCT TTA CT TTATG GGGAA GTTGG AGACA CACTG TTGAT TATAT

```

1901 TTAAG AATCA AGCAA GCAGA CCATA TAACA TCTAC CCTCA CGGAA TCACT  
1951 GATGT CCGTC CTTTG TATTC AAGGA GATTA CCAA AGGTG TAAAA CATT  
2001 GAAGG ATTTT CCAAT TCTGC CAGGA GAAAT ATTCA AATAT AAATG GACAG  
2051 TGACT GTAGA AGATG GGCCA ACTAA ATCAG ATCCT CGGTG CCTGA CCCGC  
2101 TATTA CTCTA GTTTC GTTAA TATGG AGAGA GATCT AGCTT CAGGA CTCAT  
2151 TGGCC CTCTC CTCAT CTGCT ACAAA GAATC TGTAG ATCAA AGAGG AAACC  
2201 AGATA ATGTC AGACA AGAGG AATGT CATCC TGTTT TCTGT ATTTG ATGAG  
2251 AACCG AAGCT GGTAC CTCAC AGAGA ATATA CAACG CTTTC TCCCC AATCC  
2301 AGCTG GAGTG CAGCT TGAGG ATCCA GAGTT CCAAG CCTCC AACAT CATGC  
2351 ACAGC ATCAA TGGCT ATGTT TTTGA TAGTT TGCAG TTGTC AGTTT GTTTG  
2401 CATGA GGTGG CATA TGGTA CATT TAAGC ATTGG AGCAG AGACT GACTT  
2451 CCTTT CTGTC TTCTT CTCTG GATAT ACCTT CAAAC ACAA ATGGT CTATG  
2501 AAGAC ACACT CACCC TATTC CCATT CTCAG GAGAA ACTGT CTTC A TGTCG  
2551 ATGGA AAACC CAGGT CTATG GATTC TGGGG TGCCA CAACT CAGAC TTTCG  
2601 GAACA GAGGC ATGAC CGCCT TACTG AAGGT TTCTA GTTGT GACAA GAACA  
2651 CTGGT GATTA TTACG AGGAC AGTTA TGAAG ATATT TCAGC ATACT TGCTG  
2701 AGTAA AAACA ATGCC ATTGA ACCAA GAAGC TTCTC TCAA ACGGC GCGCC  
2751 AGGTA CCTCA GAGTC TGCTA CCCCC GAGTC AGGGC CAGGA TCAGA GCCAG  
2801 CCACC TCCGG GTCTG AGACA CCCGG GACTT CCGAG AGTGC CACCC CTGAG  
2851 TCCGG ACCCG GGTCC GAGCC CGCCA CTTCC GGCTC CGAAA CTCCC GGCAC  
2901 AAGCG AGAGC GCTAC CCCAG AGTCA GGACC AGGAA CATCT ACAGA GCCCT  
2951 CTGAA GGCTC CGCTC CAGGG TCCCC AGCCG GCAGT CCCAC TAGCA CCGAG  
3001 GAGGG AACCT CTGAA AGCGC CACAC CCGAA TCAGG GCCAG GGTCT GAGCC  
3051 TGCTA CCAGC GGCAG CGAGA CACCA GGCAC CTCTG AGTCC GCCAC ACCAG  
3101 AGTCC GGACC CGGAT CTCCC GCTGG GAGCC CCACC TCCAC TGAGG AGGGA  
3151 TCTCC TGCTG GCTCT CCAAC ATCTA CTGAG GAAGG TACCT CAACC GAGCC  
3201 ATCCG AGGGA TCAGC TCCCG GCACC TCAGA GTCGG CAACC CCGGA GTCTG  
3251 GACCC GGAAC TTCCG AAAGT GCCAC ACCAG AGTCC GGTCC CGGGA CTTC A  
3301 GAATC AGCAA CACCC GAGTC CGGCC CTGGG TCTGA ACCCG CCACA AGTGG  
[0643] 3351 TAGTG AGACA CCAGG ATCAG AACCT GCTAC CTCAG GGTCA GAGAC ACCCG  
3401 GATCT CCGGC AGGCT CACCA ACCTC CACTG AGGAG GGCAC CAGCA CAGAA  
3451 CCAAG CGAGG GCTCC GCACC CGGAA CAAGC ACTGA ACCCA GTGAG GGTT  
3501 AGCAC CCGGC TCTGA GCCGG CCACA AGTGG CAGTG AGACA CCGGG CACTT  
3551 CAGAG AGTGC CACCC CCGAG AGTGG CCCAG GCACT AGTAC CGAGC CCTCT  
3601 GAAGG CAGTG CGCCA GCCTC GAGCC CACCA GTCTT GAAA GCCAT CAAGC  
3651 TGAAA TAACT CGTAC TACTC TTCAG TCAGA TCAAG AGGAA ATCGA TTATG  
3701 ATGAT ACCAT ATCAG TTGAA ATGAA GAAGG AAGAT TTTGA CATT ATGAT  
3751 GAGGA TGAAA ATCAG AGCCC CCGCA GCTTT CAAA GAAA CACGA CACTA  
3801 TTTTA TTGCT GCAGT GGAGA GGCTC TGGGA TTATG GGATG AGTAG CTCCC  
3851 CACAT GTTCT AAGAA ACAGG GCTCA GAGTG GCAGT GTCCC TCAGT TCAAG  
3901 AAAGT TGTTT TCCAG GAATT TACTG ATGGC TCCTT TACTC AGCCC TTATA  
3951 CCGTG GAGAA CTAAG TGAAC ATTTG GGACT CCTGG GGCCA TATAT AAGAG  
4001 CAGAA GTTGA AGATA ATATC ATGGT AACTT TCAGA AATCA GGCTT CTCGT  
4051 CCCTA TTCCT TCTAT TCTAG CCTTA TTTCT TATGA GGAAG ATCAG AGGCA  
4101 AGGAG CAGAA CCTAG AAAAA ACTTT GTCAA GCCTA ATGAA ACCAA AACTT  
4151 ACTTT TGGAA AGTGC AACAT CATAT GGCAC CCACT AAAGA TGAGT TTGAC  
4201 TGCAA AGCCT GGGCT TATTT CTCTG ATGTT GACCT GGAAA AAGAT GTGCA  
4251 CTCAG GCCTG ATTGG ACCCC TTCTG GTCTG CCACA CTAAC ACACT GAACC  
4301 CTGCT CATGG GAGAC AAGTG ACAGT ACAGG AATTT GCTCT GTTTT TCACC  
4351 ATCTT TGATG AGACC AAAAG CTGGT ACTTC ACTGA AAATA TGGAA AGAAA  
4401 CTGCA GGGCT CCCTG CAATA TCCAG ATGGA AGATC CCACT TTTAA AGAGA  
4451 ATTAT CGCTT CCATG CAATC AATGG CTACA TAATG GATAC ACTAC CTGGC  
4501 TTAGT AATGG CTCAG GATCA AAGGA TTCGA TGGTA TCTGC TCAGC ATGGG  
4551 CAGCA ATGAA AACAT CCATT CTATT CATTT CAGTG GACAT GTGTT CACTG  
4601 TACGA AAAAA AGAGG AGTAT AAAAT GGCAC TGTAC AATCT CTATC CAGGT  
4651 GTTTT TGAGA CAGTG GAAAT GTTAC CATCC AAAGC TGGAA TTTGG CGGGT  
4701 GGAAT GCCTT ATTGG CGAGC ATCTA CATGC TGGGA TGAGC ACACT TTTTC  
4751 TGGTG TACAG CAATA AGTGT CAGAC TCCCC TGGGA ATGGC TTCTG GACAC  
4801 ATTAG AGATT TTCAG ATTAC AGCTT CAGGA CAATA TGGAC AGTGG GCCCC

4851 AAAGC TGGCC AGACT TCATT ATTCC GGATC AATCA ATGCC TGGAG CACCA  
 4901 AGGAG CCCTT TTCTT GGATC AAGGT GGATC TGTTG GCACC AATGA TTATT  
 4951 CACGG CATCA AGACC CAGGG TGCCC GTCAG AAGTT CTCCA GCCTC TACAT  
 5001 CTCTC AGTTT ATCAT CATGT ATAGT CTTGA TGGGA AGAAG TGGCA GACTT  
 5051 ATCGA GGAAA TTCCA CTGGA ACCTT AATGG TCTTC TTTGG CAATG TGGAT  
 5101 TCATC TGGGA TAAAA CACAA TATTT TTAAC CCTCC AATTA TTGCT CGATA  
 5151 CATCC GTTTG CACCC AACTC ATTAT AGCAT TCGCA GCACT CTTCG CATGG  
 5201 AGTTG ATGGG CTGTG ATTTA AATAG TTGCA GCATG CCATT GGGAA TGGAG  
 5251 AGTAA AGCAA TATCA GATGC ACAGA TTAAT GCTTC ATCCT ACTTT ACCAA  
 5301 TATGT TTGCC ACCTG GTCTC CTTCA AAAGC TCGAC TTCAC CTCCA AGGGA  
 5351 GGAGT AATGC CTGGA GACCT CAGGT GAATA ATCCA AAAGA GTGGC TGCAA  
 5401 GTGGA CTTCC AGAAG ACAAT GAAAG TCACA GGAGT AACTA CTCAG GGAGT  
 5451 AAAAT CTCTG CTTAC CAGCA TGTAT GTGAA GGAGT TCCTC ATCTC CAGCA  
 5501 GTCAA GATGG CCATC AGTGG ACTCT CTTTT TTCAG AATGG CAAAAG TAAAG  
 5551 GTTTT TCAGG GAAAT CAAGA CTCCT TCACA CCTGT GGTGA ACTCT CTAGA  
 [0644] 5601 CCCAC CGTTA CTGAC TCGCT ACCTT CGAAT TCACC CCCAG AGTTG GGTGC  
 5651 ACCAG ATTGC CCTGA GGATG GAGGT TCTGG GCTGC GAGGC ACAGG ACCTC  
 5701 TACGA CAAAA CTCAC ACATG CCCAC CGTGC CCAGC TCCAG AACTC CTGGG  
 5751 CGGAC CGTCA GTCTT CCTCT TCCCC CAAA ACCCA AGGAC ACCCT CATGG  
 5801 CCTCC CGGAC CCCTG AGGTC ACATG CGTGG TGGTG GACGT GAGCC ACGAA  
 5851 GACCC TGAGG TCAAG TTCAA CTGGT ACGTG GACGG CGTGG AGGTG CATAA  
 5901 TGCCA AGACA AAGCC GCGGG AGGAG CAGTA CAACA GCACG TACCG TGTGG  
 5951 TCAGC GTCCT CACCG TCCTG GCCCA GGACT GGCTG AATGG CAAGG AGTAC  
 6001 AAGTG CAAGG TCTCC AACAA AGCCC TCCCA GCCCC CATCG AGAAA ACCAT  
 6051 CTCCA AAGCC AAAGG GCAGC CCCGA GAACC ACAGG TGTAC ACCCT GCCCC  
 6101 CATCC CGCGA TGAGC TGACC AAGAA CCAGG TCAGC CTGAC CTGCC TGGTC  
 6151 AAAGG CTTCT ATCCC AGCGA CATCG CCGTG GAGTG GGAGA GCAAT GGGCA  
 6201 GCCGG AGAAC AACTA CAAGA CCACG CCTCC CGTGT TGGAC TCCGA CGGCT  
 6251 CCTTC TTCCT CTACA GCAAG CTCAC CGTGG ACAAG AGCAG GTGGC AGCAG  
 6301 GGGAA CGTCT TCTCA TGCTC CGTGA TGCAT GAGGC TCTGC ACAAC GCCTA  
 6351 CACGC AGAAG AGCCT CTCCC TGTCT CCGGG TAAAT GA  
 [0645] FVIII 263蛋白质序列(IHH三重突变体) (SEQ ID NO:157)

1	MQIELSTCFF	LCLLRFCFSA	TRRYYLGAVE	LSWDYMQGAP	TSESATPESG
51	PGSEPATSGS	ETPGTSESAT	PESGPGSEPA	TSGSETPGTS	ESATPESGPG
101	TSTEPSEGS	PGSPAGSPTS	TEEGTSESAT	PESGPGSEPA	TSGSETPGTS
151	ESATPESGPG	SPAGSPTSTE	EGSPAGSPTS	TEEGASSSDL	GELPVDARFP
201	PRVPKSFPFN	TSVVYKKTFL	VEFTDHLFNI	AKPRPPWMGL	LGPTIQAEVY
251	DTVVITLKNM	ASHPVSLHAV	GVSYWKASEG	AEYDDQTSQR	EKEDDKVFPG
301	GSHTYVWQVL	KENGPASDP	LCLTYSYLSH	VDLVKDLNSG	LIGALLVCRE
351	GSLAKEKTQT	LHKFILLFAV	FDEGKSWHSE	TKNSLMQDRD	AASARAWPKM
401	HTVNGYVNR	LPGLIGCHRK	SVYWHVIGMG	TTPEVHSIFL	EGHTFLVRNH
451	RQASLEISPI	TFLTAQTLLM	DLGQFLLFCH	ISSHQHDGME	AYVKVDSCPE
501	EPQLRMKNNE	EAEDYDDDLT	DSEMDVVRFD	DDNSPSFIQI	RSVAKKHPKT
551	WVHYIAAEEE	DWDYAPLVLA	PDDRSYKSQY	LNNGPQRIGR	KYKKVRFMAY
601	TDETFKTREA	IQHESGILGP	LLYGEVGDTL	LIIFKNQASR	PYNIYPHGIT
651	DVRPLYSRRL	PKGVKHLKDF	PILPGEIFKY	KWTVTVEDGP	TKSDPRCLTR
701	YYSSFVNMER	DLASGLIGPL	LICYKESVDQ	RGNQIMSDKR	NVILFSVFDE
751	NRSWYLTEI	QRFLPNPAGV	QLEDPEFQAS	NIMHSINGYV	FDSLQLSVCL
801	HEVAYWYILS	IGAQTDFLSV	FFSGYTFKHK	MVYEDTLTLF	PFSGETVFMS
851	MENPGLWILG	CHNSDFRNRG	MTALLKVSSC	DKNTGDYED	SYEDISAYLL
901	SKNNAIEPRS	FSQNGAPGTS	ESATPESGPG	SEPATSGSET	PGTSESATPE
951	SGPGSEPATS	GSETPGTSES	ATPESGPGTS	TEPSEGSAPG	SPAGSPTSTE
1001	EGTSESATPE	SGPGSEPATS	GSETPGTSES	ATPESGPGSP	AGSPTSTEEG
[0646] 1051	SPAGSPTSTE	EGTSTEPSEG	SAPGTSESAT	PESGPGTSES	ATPESGPGTS
1101	ESATPESGPG	SEPATSGSET	PGSEPATSGS	ETPGSPAGSP	TSTEEGTSTE
1151	PSEGSAPGTS	TEPSEGSAPG	SEPATSGSET	PGTSESATPE	SGPGTSTEPS
1201	EGSAPASSPP	VLKRHQAEIT	RTTLQSDQEE	IDYDDTISVE	MKKEDFDIYD
1251	EDENQSPRSF	QKKTRHYFIA	AVERLWDYGM	SSSPHVLNR	AQSGSVPOFK
1301	KVVVFQFTDG	SFTQPLYRGE	LNEHLGLLGP	YIRAEVEDNI	MVTFRNQASR
1351	PYSFYSSLIS	YEEDQRQGAE	PRKNFVKPNE	TKTYFWKVQH	HMAPTKDEFD
1401	CKAWAYFSDV	DLEKDVHSGL	IGPLLVCHTN	TLNPAHGRQV	TVQEFALFFT
1451	IFDETKSWYF	TENMERN CRA	PCNIQMEDPT	FKENYRFHAI	NGYIMDTLPG
1501	LVMAQDQIR	WYLLSMGSNE	NIHSIHFSGH	VFTVRKKEEY	KMALYNLYPG
1551	VFETVEMLPS	KAGIWRVECL	IGEHLHAGMS	TLFLVYSNKC	QTPLGMASGH
1601	IRDFQITASG	QYGQWAPKLA	RLHYSGSINA	WSTKEPFSWI	KVDLLAPMI I
1651	HGIKTQGARQ	KFSSLYISQF	IIMYSLDGKK	WQTYRGNSTG	TLMVFFGNVD
1701	SSGIKHNIFN	PPIIARYIRL	HPTHYSIRST	LRMELMGCDL	NSCSMPLGME
1751	SKAISDAQIT	ASSYFTNMFA	TWSPSKARLH	LQGRSNAWRP	QVNNPKEWLQ
1801	VDFQKTMKVT	GVTTQGVKSL	LTSMYVKEFL	ISSSQDGHQW	TLFFQNGKVK
1851	VFQGNQDSFT	PVNSLDPPL	LTRYLRIHPQ	SWVHQIALRM	EVLGCEAQDL
1901	YDKTHTCPPC	PAPELLGGPS	VFLFPPKPKD	TLMASRTPEV	TCVVVDVSHE
1951	DPEVKFNWYV	DGVEVHNAKT	KPREEQYNST	YRVVSVLTVL	AQDWLNGKEY
2001	KCKVSNKALP	APIEKTISKA	KGQPREPQVY	TLPPSRDEL T	KNQVSLTCLV
2051	KGFYPSDIAV	EWESNGQPEN	NYKTTPPVLD	SDGSFFLYSK	LTVDKSRWQQ
2101	GNVFSCSVMH	EALHNAYTQK	SLSLSPGK*		

[0647] FVIII 282核苷酸序列 (SEQ ID NO:158)

[0648]

```

1   ATGCA AATAG AGCTC TCCAC CTGCT TCTTT CTGTG CCTTT TGCGA TTCTG
51  CTTTA GTGCC ACCAG AAGAT ACTAC CTGGG TGCAG TGGAA CTGTC ATGGG
101 ACTAT ATGCA AAGTG ATCTC GGTGA GCTGC CTGTG GACGC AAGAT TTCCT
151 CCTAG AGTGC CAAAA TCTTT TCCAT TCAAC ACCTC AGTCG TGTAC AAAAA
201 GACTC TGTTT GTAGA ATTCA CGGAT CACCT TTTCA ACATC GCTAA GCCAA
251 GGCCA CCCTG GATGG GTCTG CTAGG TCCTA CCATC CAGGC TGAGG TTTAT
301 GATAC AGTGG TCATT ACACT TAAGA ACATG GCTTC CCATC CTGTC AGTCT
351 TCATG CTGTT GGTGT ATCCT ACTGG AAAGC TTCTG AGGGA GCTGA ATATG
401 ATGAT CAGAC CAGTC AAAGG GAGAA AGAAG ATGAT AAAGT CTTCC CTGGT
451 GGAAG CCATA CATAT GTCTG GCAGG TCCTG AAAGA GAATG GTCCA ATGGC
501 CTCTG ACCCA CTGTG CCTTA CCTAC TCATA TCTTT CTCAT GTGGA CCTGG
551 TAAAA GACTT GAATT CAGGC CTCAT TGGAG CCCTA CTAGT ATGTA GAGAA
601 GGGAG TCTGG CCAAG GAAAA GACAC AGACC TTGCA CAAAT TTATA CTA CTACT
651 TTTTG CTGTA TTTGA TGAAG GGAAA AGTTG GCACT CAGAA ACAA GAACT
701 CCTTG ATGCA GGATA GGGAT GCTGC ATCTG CTCGG GCCTG GCCTA AAATG
751 CACAC AGTCA ATGGT TATGT AAACA GGTCT CTGCC AGGTC TGATT GGATG
801 CCACA GGAAA TCAGT CTATT GGCAT GTGAT TGGAA TGGGC ACCAC TCCTG
851 AAGTG CACTC AATAT TCCTC GAAGG TCACA CATTT CTTGT GAGGA ACCAT
901  GCCA GGCTA GCTTG GAAAT CTCGC CAATA ACTTT CCTTA CTGCT CAAAC
951  ACTCT TGATG GACCT TGGAC AGTTT CTA CTACT GTTTT GTCAT ATCTC TTCCC
1001 ACCAA CATGA TGGCA TGGAA GCTTA TGTCA AAGTA GACAG CTGTC CAGAG
1051 GAACC CCAAC TACGA ATGAA AAATA ATGAA GAAGC GGAAG ACTAT GATGA
1101 TGATC TTA CTACT GATTC TGAAA TGGAT GTGGT CAGGT TTGAT GATGA CAACT
1151 CTCCT TCCTT TATCC AAATT CGCTC AGTTG CCAAG AAGCA TCCTA AA ACT
1201 TGGGT ACATT ACATT GCTGC TGAAG AGGAG GACTG GGACT ATGCT CCCTT
1251 AGTCC TCGCC CCCGA TGACA GAAGT TATAA AAGTC AATAT TTGAA CAATG
1301 GCCCT CAGCG GATTG GTAGG AAGTA CAAAA AAGTC CGATT TATGG CATA C
1351 ACAGA TGAAA CCTTT AAGAC TCGTG AAGCT ATTCA GCATG AATCA GGAAT
1401 CTTGG GACCT TTA CTACT TTATG GGGAA GTTGG AGACA CACTG TTGAT TATA T
1451 TTAAG AATCA AGCAA GCAGA CCATA TAACA TCTAC CCTCA CGGAA TCACT
1501 GATGT CCGTC CTTTG TATTC AAGGA GATTA CCAA AGGTG TAAAA CATT T
1551 GAAGG ATTTT CCAAT TCTGC CAGGA GAAAT ATTCA AATAT AAATG GACAG
1601 TGACT GTAGA AGATG GGCCA ACTAA ATCAG ATCCT CGGTG CCTGA CCCGC
1651 TATTA CTCTA GTTTC GTTAA TATGG AGAGA GATCT AGCTT CAGGA CTCAT
1701 TGGCC CTCTC CTCAT CTGCT ACAA GAATC TGTAG ATCAA AGAGG AAACC
1751 AGATA ATGTC AGACA AGAGG AATGT CATCC TGTTT TCTGT ATTTG ATGAG
1801 AACCG AAGCT GGTAC CTCAC AGAGA ATATA CAACG CTTTC TCCC AATCC
1851 AGCTG GAGTG CAGCT TGAGG ATCCA GAGTT CCAAG CCTCC AACAT CATGC

```

1901 ACAGC ATCAA TGGCT ATGTT TTTGA TAGTT TGCAG TTGTC AGTTT GTTTG  
1951 CATGA GGTGG CATA C TGGTA CATT C TAAGC ATTGG AGCAC AGACT GACTT  
2001 CCTTT CTGTC TTCTT CTCTG GATAT ACCTT CAAAC ACAA ATGGT CTATG  
2051 AAGAC AACT CACCC TATTC CCATT CTCAG GAGAA ACTGT CTTCA TGTCG  
2101 ATGGA AAACC CAGGT CTATG GATTC TGGGG TGCCA CAACT CAGAC TTTCCG  
2151 GAACA GAGGC ATGAC CGCCT TACTG AAGGT TTCTA GTTGT GACAA GAACA  
2201 CTGGT GATTA TTACG AGGAC AGTTA TGAAG ATATT TCAGC ATACT TGCTG  
2251 AGTAA AAACA ATGCC ATTGA ACCAA GAAGC TTCTC TCAA ACGGC GCGCC  
2301 AACAT CAGAG AGCGC CACCC CTGAA AGTGG TCCCG GGAGC GAGCC AGCCA  
2351 CATCT GGGTC GGAAA CGCCA GGCAC AAGTG AGTCT GCAAC TCCCG AGTCC  
2401 GGAGC TGGCT CCGAG CCTGC CACTA GCGGC TCCGA GACTC CGGGA ACTTC  
2451 CGAGA GCGCT ACACC AGAAA GCGGA CCCGG AACCA GTACC GAACC TAGCG  
2501 AGGGC TCTGC TCCGG GCAGC CCAGC CGGCT CTCCT ACATC CACGG TAGGAG  
2551 GGCAC TTCCG AATCC GCCAC CCCGG AGTCA GGGCC AGGAT CTGAA CCCGC  
2601 TACCT CAGGC AGTGA GACGC CAGGA ACGAG CGAGT CCGCT ACACC GGAGA  
2651 GTGGG CCAGG GAGCC CTGCT GGATC TCCTA CGTCC ACTGA GGAAG GGTC A  
2701 CCAGC GGGCT CGCCC ACCAG CACTG AAGAA GGTGC CTCGA GCCCA CCAGT  
2751 CTTGA AACGC CATCA AGCTG AAATA ACTCG TACTA CTCTT CAGTC AGATC  
2801 AAGAG GAAAT CGATT ATGAT GATAC CATAT CAGTT GAAAT GAAGA AGGAA  
2851 GATTT TGACA TTTAT GATGA GGATG AAAAT CAGAG CCCCC GCAGC TTTCA  
2901 AAAGA AAACA CGACA CTATT TTATT GCTGC AGTGG AGAGG CTCTG GGATT  
2951 ATGGG ATGAG TAGCT CCCCC CATGT TCTAA GAAAC AGGGC TCAGA GTGGC  
3001 AGTGT CCCTC AGTTC AAGAA AGTTG TTTTC CAGGA ATTTA CTGAT GGCTC  
3051 CTTTA CTCAG CCCTT ATACC GTGGA GAACT AAATG AACAT TTGGG ACTCC  
3101 TGGGG CCATA TATAA GAGCA GAAGT TGAAG ATAAT ATCAT GGTAA CTTTC  
3151 AGAAA TCAGG CCTCT CGTCC CTATT CCTTC TATTC TAGCC TTATT CTTTA  
3201 TGAGG AAGAT CAGAG GCAAG GAGCA GAACC TAGAA AAAAC TTTGT CAAGC  
3251 CTAAT GAAAC CAAAA CTTAC TTTTG GAAAG TGCAA CATCA TATGG CACCC  
3301 ACTAA AGATG AGTTT GACTG CAAAG CCTGG GCTTA TTTCT CTGAT GTTGA  
[0649] 3351 CCTGG AAAAA GATGT GCACT CAGGC CTGAT TGGAC CCCTT CTGGT CTGCC  
3401 AACT AACAC ACTGA ACCCT GCTCA TGGGA GACAA GTGAC AGTAC AGGAA  
3451 TTTGC TCTGT TTTTC ACCAT CTTTG ATGAG ACCAA AAGCT GGTAC TTCAC  
3501 TGAAA ATATG GAAAG AAAC TGCAGG GCTCC CTGCA ATATC CAGAT GGAAG  
3551 ATCCC ACTTT TAAAG AGAAT TATCG CTTCC ATGCA ATCAA TGGCT ACATA  
3601 ATGGA TACAC TACCT GGCTT AGTAA TGGCT CAGGA TCAA GGATT CGATG  
3651 GTATC TGCTC AGCAT GGGCA GCAAT GAAAA CATCC ATTCT ATTCA TTTCA  
3701 GTGGA CATGT GTTCA CTGTA CGAAA AAAAG AGGAG TATAA AATGG CACTG  
3751 TACAA TCTCT ATCCA GGTGT TTTTG AGACA GTGGA AATGT TACCA TCCAA  
3801 AGCTG GAATT TGGCG GGTGG AATGC CTTAT TGGCG AGCAT CTACA TGCTG  
3851 GGATG AGCAC ACTTT TTCTG GTGTA CAGCA ATAAG TGTC A GACTC CCCTG  
3901 GGAAT GGCTT CTGGA CACAT TAGAG ATTTT CAGAT TACAG CTTC A GGACA  
3951 ATATG GACAG TGGGC CCCAA AGCTG GCCAG ACTTC ATTAT TCCGG ATCAA  
4001 TCAAT GCCTG GAGCA CCAAG GAGCC CTTTT CTTGG ATCAA GGIGG ATCTG  
4051 TTGGC ACCAA TGATT ATTCA CGGCA TCAAG ACCCA GGGTG CCCGT CAGAA  
4101 GTTCT CCAGC CTCTA CATCT CTCAG TTTAT CATCA TGTAT AGTCT TGATG  
4151 GGAAG AAGTG GCAGA CTTAT CGAGG AAATT CCACT GGAAC CTTAA TGGTC  
4201 TTCTT TGGCA ATGTG GATTC ATCTG GGATA AAACA CAATA TTTTT AACCC  
4251 TCCAA TTATT GCTCG ATACA TCCGT TTGCA CCCAA CTCAT TATAG CATTC  
4301 GCAGC ACTCT TCGCA TGGAG TTGAT GGGCT GTGAT TTA A TAGTT GCAGC  
4351 ATGCC ATTGG GAATG GAGAG TAAAG CAATA TCAGA TGCAC AGATT ACTGC  
4401 TTCAT CCTAC TTTAC CAATA TGTTT GCCAC CTGGT CTCCT TCAA AGCTC  
4451 GACTT CACCT CCAAG GGAGG AGTAA TGCC T GGAGA CCTCA GTGA ATAAT  
4501 CAAA AGAGT GGCTG CAAGT GGA TCCAG AAGAC AATGA AAGTC ACAGG  
4551 AGTAA CTACT CAGGG AGTAA AATCT CTGCT TACCA GCATG TATGT GAAGG  
4601 AGTTC CTCAT CTCCA GCAGT CAAGA TGGCC ATCAG TGGAC TCTCT TTTTT  
4651 CAGAA TGGCA AAGTA AAGGT TTTTC AGGGA AATCA AGACT CCTTC ACACC  
4701 TGTGG TGAAC TCTCT AGACC CACCG TTACT GACTC GCTAC CTTCG AATTC  
4751 ACCCC CAGAG TTGGG TGCAC CAGAT TGCCC TGAGG ATGGA GTTC TGGGC  
4801 TGCGA GGCAC AGGAC CTCTA CGACA AAAC T CACAC ATGCC CACCG TGCCC



4851 AGCTC CAGAA CTCCT GGGCG GACCG TCAGT CTTC CTTCC CCCCC AAAAC  
 4901 CCAAG GACAC CCTCA TGATC TCCCG GACCC CTGAG GTCAC ATGCG TGGTG  
 4951 GTGGA CGTGA GCCAC GAAGA CCCTG AGGTC AAGTT CAACT GGTAC GTGGA  
 5001 CGGCG TGGAG GTGCA TAATG CCAAG ACAA GCCGC GGGAG GAGCA GTACA  
 5051 ACAGC ACGTA CCGTG TGGTC AGCGT CCTCA CCGTC CTGCA CCAGG ACTGG  
 5101 CTGAA TGGCA AGGAG TACAA GTGCA AGGTC TCCAA CAAAG CCCTC CCAGC  
 [0650] 5151 CCCC TCGAG AAAAC CATCT CCAA GCCAA AGGGC AGCCC CGAGA ACCAC  
 5201 AGGTG TACAC CCTGC CCCC TCCCG GGATG AGCTG ACCAA GAACC AGGTC  
 5251 AGCCT GACCT GCCTG GTCAA AGGCT TCTAT CCCAG CGACA TCGCC GTGGA  
 5301 GTGGG AGAGC AATGG GCAGC CGGAG AACAA CTACA AGACC ACGCC TCCCG  
 5351 TGTGT GACTC CGACG GCTCC TTCTT CCTCT ACAGC AAGCT CACCG TGGAC  
 5401 AAGAG CAGGT GGCAG CAGGG GAACG TCTTC TCATG CTCCG TGATG CATGA  
 5451 GGCTC TGCAC AACCA CTACA CGCAG AAGAG CCTCT CCCTG TCTCC GGGTA  
 5501 AATGA

[0651] FVIII 282蛋白质序列 (SEQ ID NO:159)

1 MQIELSTCFF LCLLRFCFSA TRRYYLGA VE LSWDYMQSDL GELPVDARFP  
 51 PRVPKSFPFN TSVVYKKT LF VEFTDHLFNI AKPRPPWMGL LGPTIQAEVY  
 101 DTVVITLKNM ASHPVSLHAV GVS YWKASEG AEYDDQTSQR EKEDDKVFPG  
 151 GSHTYVWQVL KENGP MASDP LCLTYSYLSH VDLVKDLNSG LIGALLVCRE  
 201 GSLAKEKTQT LHKFILLFAV FDEGKSWHSE TKNSLMQDRD AASARAWPKM  
 251 HTVNGYVNRSLPGLIGCHRK SVYWHVIGMG TTPEVHSIFL EGHTFLVRNH  
 301 RQASLEISPI TFLTAQTLLM DLGQFLLFCH ISSHQHDGME AYVKVDSCPE  
 351 EPQLRMKNNE EAEDYDDDLT DSEMDVVRFD DDNSPSFIQI RSVAKKHPKT  
 401 WVHYIAAEEE DWDYAPLVLA PDDRSYKSQY LNNGPQRIGR KYKKVRFMAY  
 451 TDETFKTREA IQHESGILGP LLYGEVGD TL LIIFKNQASR PNYIYPHGIT  
 501 DVRPLYSRRL PKGVKHLKDF PILPGEIFKY KWTVTVEDGP TKSDPRCLTR  
 551 YYSSFVNMER DLASGLIGPL LICYKESVDQ RGNQIMSDKR NVILFSVFDE  
 601 NRSWYL TENI QRFLPNPAGV QLEDPEFQAS NIMHSINGYV FDSLQLSVCL  
 651 HEVAYWYILS IGAQTDFLSV FFSGYTFKHK MVEYEDTLTLF PFSGETVFMMS  
 701 MENPGLWILG CHNSDFRNRG MTALLKVSSC DKNTGDYIED SYEDISAYLL  
 751 SKNNAIEPRS FSQNGAPTSE SATPESGPGS EPATSGSETP GTSESATPES  
 801 GPGSEPATSG SETPGTSESA TPESGPGTST EPSEGSAPGS PAGSPTSTEE  
 851 GTSESATPES GPGSEPATSG SETPGTSESA TPESGPGSPA GSPTSTEEGS  
 [0652] 901 PAGSPTSTEE GASSPPVLKR HQAEITR TTL QSDQEEIDYD DTISVEMKKE  
 951 DFDIYDEDEN QSPRSFQKKT RHYFIAAVER LW DYGMSSSP HVLRNRAQSG  
 1001 SV PQFKKVV FQEFTDGSFTQ PLYRGELNEH LGLLGPYIRA EVEDNIMVTF  
 1051 RNQASRPYSF YSSLISYEED QRQGAEP RKN FVKPNETKTY FWKVQH HMAP  
 1101 TKDEFDCKAW AYFSDVDLEK DVHSGLIGPL LVCHTNTLNP AHGRQVTVQE  
 1151 FALFFTIFDE TKS WYFTENM ERNCRAPCNI QMEDPTFKEN YRFHAINGYI  
 1201 MDTLPGLVMA QDQRIRWYLL SMGSNENIHS IHFSGHVFTV RKKEEYKMAL  
 1251 YNLYPGVFET VEMLP SKAGI WRVECLIGEH LHAGMSTLFL VYSNKCQTPL  
 1301 GMASGHIRDF QITASGQY GQ WAPKLARLHY SGSINAWSTK EPFSWIKVDL  
 1351 LAPMI IHG IKTQGARQKFSS LYISQFIIMY SLDGKKWQTY RGNSTGTLMV  
 1401 FFGNVDSSGI KHNIFNPPII ARYIRLHPTH YSIRSTLRME LMGCDLNSCS  
 1451 MPLGMESKAI SDAQITASSY FTNMFATWSP SKARLHLQGR SNAWRPQVNN  
 1501 PKEWLQVDFQ KTMKVTG VTT QGVKSLTSM YVKEFLISS QDGHQWTLFF  
 1551 QNGKVKV FQG NQDSFTP VVN SLDPPLLTRY LRIHPQSWVH QIALRMEVLG  
 1601 CEAQDLYDKT HTCPCPAPE LLGGPSVFLF PPKPKDTLMI SRTPEVTCVV  
 1651 VDVSHEDPEV KFNWYVDGVE VHNAKTKPRE EQYNSTYRVV SVLTVLHQDW  
 1701 LNGKEYKCKV SNKALPAPIE KTISKAKGQP REPQVYTLPP SRDELTKNQV  
 1751 SLTCLVKGFY PSDIAVEWES NGQPENNYKT TPPVLDS DGS FFLYSKLTV D  
 1801 KSRWQQGNVF SCSVMHEALH NHYTQKSLSL SPGK\*

[0653] FVIII 283核苷酸序列 (具有IHH三重突变的FVIII 169) (SEQ ID NO:160)

[0654]

1 ATGCA AATAG AGCTC TCCAC CTGCT TCTTT CTGTG CCTTT TGC GA TTCTG  
51 CTTTA GTGCC ACCAG AAGAT ACTAC CTGGG TGCAG TGGAA CTGTC ATGGG  
101 ACTAT ATGCA AAGTG ATCTC GGTGA GCTGC CTGTG GACGC AAGAT TTCCT  
151 CCTAG AGTGC CAAAA TCTTT TCCAT TCAAC ACCTC AGTCG TGTAC AAAAA  
201 GACTC TGTTT GTAGA ATTCA CGGAT CACCT TTTCA ACATC GCTAA GCCAA  
251 GGCCA CCCTG GATGG GTCTG CTAGG TCCTA CCATC CAGGC TGAGG TTTAT  
301 GATAC AGTGG TCATT ACACT TAAGA ACATG GCTTC CCATC CTGTC AGTCT  
351 TCATG CTGTT GGTGT ATCCT ACTGG AAAGC TTCTG AGGGA GCTGA ATATG  
401 ATGAT CAGAC CAGTC AAAGG GAGAA AGAAG ATGAT AAAGT CTTCC CTGGT  
451 GGAAG CCATA CATAT GTCTG GCAGG TCCTG AAAGA GAATG GTCCA ATGGC  
501 CTCTG ACCCA CTGTG CCTTA CCTAC TCATA TCTTT CTCAT GTGGA CCTGG  
551 TAAAA GACTT GAATT CAGGC CTCAT TGGAG CCCTA CTAGT ATGTA GAGAA  
601 GGGAG TCTGG CCAAG GAAAA GACAC AGACC TTGCA CAAAT TTATA CTA CT  
651 TTTTG CTGTA TTTGA TGAAG GGAAA AGTTG GCACT CAGAA ACAAA GAACT  
701 CCTTG ATGCA GGATA GGGAT GCTGC ATCTG CTCGG GCCTG GCCTA AAATG  
751 CACAC AGTCA ATGGT TATGT AAACA GGTCT CTGCC AGGTC TGATT GGATG  
801 CCACA GGAAA TCAGT CTATT GGCAT GTGAT TGGAA TGGGC ACCAC TCCTG  
851 AAGTG CACTC AATAT TCCTC GAAGG TCACA CATTT CTTGT GAGGA ACCAT  
901 CGCCA GGCTA GCTTG GAAAT CTCGC CAATA ACTTT CCTTA CTGCT CAAAC  
951 ACTCT TGATG GACCT TGGAC AGTTT CTA CTACT GTTTT GTCAT ATCTC TTCCC  
1001 ACCAA CATGA TGGCA TGGAA GCTTA TGTCA AAGTA GACAG CTGTC CAGAG  
1051 GAACC CCAAC TACGA ATGAA AAATA ATGAA GAAGC GGAAG ACTAT GATGA  
1101 TGATC TTA CT GATTC TGAAA TGGAT GTGGT CAGGT TTGAT GATGA CAACT  
1151 CTCCT TCCTT TATCC AAATT CGCTC AGTTG CCAAG AAGCA TCCTA AAACT  
1201 TGGGT ACATT ACATT GCTGC TGAAG AGGAG GACTG GGACT ATGCT CCCTT  
1251 AGTCC TCGCC CCCG TGACA GAAGT TATAA AAGTC AATAT TTGAA CAATG  
1301 GCCCT CAGCG GATTG GTAGG AAGTA CAAAA AAGTC CGATT TATGG CATA  
1351 ACAGA TGAAA CCTTT AAGAC TCGTG AAGCT ATTCA GCATG AATCA GGAAT  
1401 CTTGG GACCT TTA CT TTATG TGGAA GTTGG AGACA CACTG TTGAT TATAT  
1451 TTAAG AATCA AGCAA GCAGA CCATA TAACA TCTAC CCTCA CGGAA TCA CT  
1501 GATGT CCGTC CTTTG TATTC AAGGA GATTA CCAA AGGTG TAAAA CATTT  
1551 GAAGG ATTTT CCAAT TCTGC CAGGA GAAAT ATTCA AATAT AAATG GACAG  
1601 TGACT GTAGA AGATG GGCCA ACTAA ATCAG ATCCT CGGTG CCTGA CCCGC  
1651 TATTA CTCTA GTTTC GTTAA TATGG AGAGA GATCT AGTT CAGGA CTCAT  
1701 TGGCC CTCTC CTCAT CTGCT ACAA GAATC TGTAG ATCAA AGAGG AAACC  
1751 AGATA ATGTC AGACA AGAGG AATGT CATCC TGTTT TCTGT ATTTG ATGAG  
1801 AACCG AAGCT GGTAC CTCAC AGAGA ATATA CAACG CTTTC TCCCC AATCC  
1851 AGCTG GAGTG CAGCT TGAGG ATCCA GAGTT CCAAG CCTCC AACAT CATGC  
1901 ACAGC ATCAA TGGCT ATGTT TTTGA TAGTT TGCAG TTGTC AGTTT GTTTG  
1951 CATGA GGTGG CATA TGGA CATTC TAAGC ATTGG AGCAC AGACT GACTT  
2001 CCTTT CTGTC TTCTT CTCTG GATAT ACCTT CAAAC ACAA ATGGT CTATG  
2051 AAGAC ACACT CACCC TATTC CCATT CTCAG GAGAA ACTGT CTCA TGTCG  
2101 ATGGA AAACC CAGGT CTATG GATTC TGGGG TGCCA CAACT CAGAC TTTCC  
2151 GAACA GAGGC ATGAC CGCCT TACTG AAGGT TTCTA GTTGT GACAA GAACA  
2201 CTGGT GATTA TTACG AGGAC AGTTA TGAAG ATATT TCAGC ATACT TGCTG  
2251 AGTAA AAACA ATGCC ATTGA ACCAA GAAGC TTCTC TCAA ACGGC GCGCC  
2301 AGGTA CCTCA GAGTC TGCTA CCCCC GAGTC AGGGC CAGGA TCAGA GCCAG  
2351 CCACC TCCGG GTCTG AGACA CCCC GACTT CCGAG AGTGC CACCC CTGAG  
2401 TCCGG ACCCG GGTCC GAGCC CGCCA CTTCC GGCTC CGAAA CTCCC GGCAC  
2451 AAGCG AGAGC GCTAC CCCAG AGTCA GGACC AGGAA CATCT ACAGA GCCCT  
2501 CTGAA GGCTC CGCTC CAGGG TCCCC AGCCG GCAGT CCCAC TAGCA CCGAG  
2551 GAGGG AACCT CTGAA AGCGC CACAC CCGAA TCAGG GCCAG GGTCT GAGCC  
2601 TGCTA CCAGC GGCAG CGAGA CACCA GGCAC CTCTG AGTCC GCCAC ACCAG  
2651 AGTCC GGACC CGGAT CTCCC GCTGG GAGCC CCACC TCCAC TGAGG AGGGA  
2701 TCTCC TGCTG GCTCT CCAAC ATCTA CTGAG GAAGG TACCT CAACC GACC  
2751 ATCCG AGGGA TCAGC TCCCG GCACC TCAGA GTCGG CAACC CCGGA GTCTG  
2801 GACCC GGAAC TTCCG AAAGT GCCAC ACCAG AGTCC GGTCC CGGGA CTTCA  
2851 GAATC AGCAA CACCC GAGTC CGGCC CTGGG TCTGA ACCCG CCACA AGTGG  
2901 TAGTG AGACA CCAGG ATCAG AACCT GCTAC CTCAG GGTCA GAGAC ACCCG

2951 GATCT CCGGC AGGCT CACCA ACCTC CACTG AGGAG GGCAC CAGCA CAGAA  
 3001 CCAAG CGAGG GCTCC GCACC CGGAA CAAGC ACTGA ACCCA GTGAG GGTTT  
 3051 AGCAC CCGGC TCTGA GCCGG CCACA AGTGG CAGTG AGACA CCCGG CACTT  
 3101 CAGAG AGTGC CACCC CCGAG AGTGG CCCAG GCACT AGTAC CGAGC CCTCT  
 3151 GAAGG CAGTG CGCCA GCCTC GAGCC CACCA GTCTT GAAAC GCCAT CAAGC  
 3201 TGAAA TAACT CGTAC TACTC TTCAG TCAGA TCAAG AGGAA ATCGA TTATG  
 3251 ATGAT ACCAT ATCAG TTGAA ATGAA GAAGG AAGAT TTTGA CATTT ATGAT  
 3301 GAGGA TGAAA ATCAG AGCCC CCGCA GCTTT CAAA GAAA CACGA CACTA  
 3351 TTTTA TTGCT GCAGT GGAGA GGCTC TGGGA TTATG GGATG AGTAG CTCCC  
 3401 CACAT GTTCT AAGAA ACAGG GCTCA GAGTG GCAGT GTCCC TCAGT TCAAG  
 3451 AAAGT TGTTT TCCAG GAATT TACTG ATGGC TCCTT TACTC AGCCC TTATA  
 3501 CCGTG GAGAA CTAAA TGAAC ATTTG GGACT CCTGG GGCCA TATAT AAGAG  
 3551 CAGAA GTTGA AGATA ATATC ATGGT AACTT TCAGA AATCA GGCCCT CTCGT  
 3601 CCCTA TTCCT TCTAT TCTAG CCTTA TTTCT TATGA GGAAG ATCAG AGGCA  
 3651 AGGAG CAGAA CCTAG AAAAA ACTTT GTCAA GCCTA ATGAA ACCAA AACTT  
 3701 ACTTT TGGAA AGTGC AACAT CATAT GGCAC CCACT AAAGA TGAGT TTGAC  
 3751 TGCAA AGCCT GGGCT TATTT CTCTG ATGTT GACCT GGAAA AAGAT GTGCA  
 3801 CTCAG GCCTG ATTGG ACCCC TTCTG GTCTG CCACA CTAAC ACACT GAACC  
 3851 CTGCT CATGG GAGAC AAGTG ACAGT ACAGG AATTT GCTCT GTTTT TCACC  
 3901 ATCTT TGATG AGACC AAAAG CTGGT ACTTC ACTGA AAATA TGGAA AGAAA  
 3951 CTGCA GGGCT CCCTG CAATA TCCAG ATGGA AGATC CCACT TTTAA AGAGA  
 4001 ATTAT CGCTT CCATG CAATC AATGG CTACA TAATG GATAC ACTAC CTGGC  
 4051 TTAGT AATGG CTCAG GATCA AAGGA TTCGA TGGTA TCTGC TCAGC ATGGG  
 4101 CAGCA ATGAA AACAT CCATT CTATT CATTT CAGTG GACAT GTGTT CACTG  
 4151 TACGA AAAAA AGAGG AGTAT AAAAT GGCAC TGTAC AATCT CTATC CAGGT  
 4201 GTTTT TGAGA CAGTG GAAAT GTTAC CATCC AAAGC TGGAA TTTGG CGGGT  
 4251 GGAAT GCCTT ATTGG CGAGC ATCTA CATGC TGGGA TGAGC ACACT TTTTC  
 4301 TGGTG TACAG CAATA AGTGT CAGAC TCCCC TGGGA ATGGC TTCTG GACAC  
 4351 ATTAG AGATT TTCAG ATTAC AGCTT CAGGA CAATA TGGAC AGTGG GCCCC  
 [0655] 4401 AAAGC TGGCC AGACT TCATT ATTCC GGATC AATCA ATGCC TGGAG CACCA  
 4451 AGGAG CCCTT TTCTT GGATC AAGGT GGATC TGTTG GCACC AATGA TTATT  
 4501 CACGG CATCA AGACC CAGGG TGCCC GTCAG AAGTT CTCCA GCCTC TACAT  
 4551 CTCTC AGTTT ATCAT CATGT ATAGT CTTGA TGGGA AGAAG TGGCA GACTT  
 4601 ATCGA GGAAA TTCCA CTGGA ACCTT AATGG TCTTC TTTGG CAATG TGGAT  
 4651 TCATC TGGGA TAAAA CACAA TATTT TTAAC CCTCC AATTA TTGCT CGATA  
 4701 CATCC GTTTG CACCC AACTC ATTAT AGCAT TCGCA GCACT CTTCG CATGG  
 4751 AGTTG ATGGG CTGTG ATTTA AATAG TTGCA GCATG CCATT GGGAA TGGAG  
 4801 AGTAA AGCAA TATCA GATGC ACAGA TTACT GCTTC ATCTT ACTTA ACCAA  
 4851 TATGT TTGCC ACCTG GTCTC CTTCA AAAGC TCGAC TTCAC TCCA ACCGA  
 4901 GGAGT AATGC CTGGA GACCT CAGGT GAATA ATCCA AAAGA GTGGC TGCAA  
 4951 GTGGA CTTCC AGAAG ACAAT GAAAG TCACA GGAGT AACTA CTCAG GGAGT  
 5001 AAAAT CTCTG CTTAC CAGCA TGTAT GTGAA GGAGT TCCTC ATCTC CAGCA  
 5051 GTCAA GATGG CCATC AGTGG ACTCT CTTTT TTCAG AATGG CAAAG TAAAG  
 5101 GTTTT TCAGG GAAAT CAAGA CTCCT TCACA CCTGT GGTGA ACTCT CTAGA  
 5151 CCCAC CGTTA CTGAC TCGCT ACCTT CGAAT TCACC CCCAG AGTTG GGTGC  
 5201 ACCAG ATTGC CCTGA GGATG GAGGT TCTGG GCTGC GAGGC ACAGG ACCTC  
 5251 TACGA CAAA CTCAC ACATG CCCAC CGTGC CCAGC TCCAG AACTC CTGGG  
 5301 CGGAC CGTCA GTCTT CCTCT TCCCC CAAA ACCCA AGGAC ACCCT CATGG  
 5351 CCTCC CGGAC CCCTG AGGTC ACATG CGTGG TGGTG GACGT GAGCC ACGAA  
 5401 GACCC TGAGG TCAAG TTCAA CTGGT ACGTG GACGG CGTGG AGGTG CATAA  
 5451 TGCCA AGACA AAGCC GCGGG AGGAG CAGTA CAACA GCACG TACCG TGTGG  
 5501 TCAGC GTCCT CACCG TCC TG CCGCA GGACT GGCTG AATGG CAAGG AGTAC  
 5551 AAGTG CAAGG TCTCC AACAA AGCCC TCCCA GCCCC CATCG AGAAA ACCAT  
 5601 CTCCA AAGCC AAAGG GCAGC CCCGA GAACC ACAGG TGTAC ACCCT GCCCC  
 5651 CATCC CGGGA TGAGC TGACC AAGAA CCAGG TCAGC CTGAC CTGCC TGGTC  
 5701 AAAGG CTTCT ATCCC AGCGA CATCG CCGTG GAGTG GGAGA GCAAT GGGCA  
 5751 GCCGG AGAAC AACTA CAAGA CCACG CCTCC CGTGT TGGAC TCCGA CGGCT  
 5801 CCTTC TTCCT CTACA GCAAG CTCAC CGTGG ACAAG AGCAG GTGGC AGCAG  
 5851 GGGAA CGTCT TCTCA TGCTC CGTGA TGCAT GAGGC TCTGC ACAAC GCCTA  
 [0656] 5901 CACGC AGAAG AGCCT CTCCC TGTCT CCGGG TAAAT GA  
 [0657] FVIII 283蛋白质序列(具有IHH三重突变的FVIII 169) (SEQ ID NO:161)

1 MQIELSTCFF LCLLRFCFSA TRRYYLGAWE LSWDYMQSDL GELPVDARFP  
 51 PRVPKSFPFN TSVVYKKTFL VEFTDHLFNI AKPRPPWMGL LGPTIQAEVY  
 101 DTVVITLKNM ASHPVSLHAV GVSYWKASEG AEYDDQTSQR EKEDDKVFPG  
 151 GSHTYVWQVL KENGPMA SDP LCLTYSYLSH VDLVKDLNSG LIGALLVCRE  
 201 GSLAKEKTQT LHKFILLFAV FDEGKSWHSE TKNSLMQDRD AASARAWPKM  
 251 HTVNGYVNRS LPGLIGCHRK SVYWHVIGMG TTPEVHSIFL EGHTFLVRNH  
 301 RQASLEISPI TFLTAQTLLM DLGQFLLFCH ISSHQHDGME AYVKVDSCPE  
 351 EPQLRMKNNE EAEDYDDDLT DSEMDVVRFD DDNSPSFIQI RSVAKKHPKT  
 401 WVHYIAAEEE DWDYAPLVLA PDDRSYKSQY LNNGPQRIGR KYKKVRFMAY  
 451 TDETFKTREA IQHESGILGP LLYGEVGD TL LIIFKNQASR PNYIYPHGIT  
 501 DVRPLYSRRL PKGVKHLKDF PILPGEIFKY KWTVTVEDGP TKSDPRCLTR  
 551 YYSSFVNMER DLASGLIGPL LICYKESVDQ RGNQIMSDKR NVILFSVFDE  
 601 NRSWYL TENI QRFLPNPAGV QLEDPEFQAS NIMHSINGYV FDSLQLSVCL  
 651 HEVAYWYILS IGAQTDFLSV FFSGYTFKHK MVYEDTLTLF PFSGETVFMMS  
 701 MENPGLWILG CHNSDFRNRG MTALLKVSSC DKNTGDYED SYEDISAYLL  
 751 SKNNAIEPRS FSQNGAPGTS ESATPESGPG SEPATSGSET PGTSESATPE  
 801 SGP GSEPATS GSETPGTSES ATPESGPGTS TEPSEGSAPG SPAGSPTSTE  
 851 EGTSESATPE SGP GSEPATS GSETPGTSES ATPESGPGSP AGSPTSTEEG  
 901 SPAGSPTSTE EGTSTEPSEG SAPGTSESAT PESGPGTSES ATPESGPGTS  
 951 ESATPESGPG SEPATSGSET PGSEPATSGS ETPGSPAGSP TSTEEGTSTE  
 [0658] 1001 PSEGSAPGTS TEPSEGSAPG SEPATSGSET PGTSESATPE SGP GTSTEPS  
 1051 EGSAPASSPP VLKRHQAEIT RTTLQSDQEE IDYDDTISVE MKKEDFDIYD  
 1101 EDENQSPRSF QKKTRHYFIA AVERLWDYGM SSSPHVLRNR AQSGSVPQFK  
 1151 KVVVFQFTDG SFTQPLYRGE LNEHLGLLGP YIRAEVEDNI MVTFRNQASR  
 1201 PYSFYSSLIS YEEDQRQGAE PRKNFVKPNE TKTYFWKVQH HMAPTKDEFD  
 1251 CKAWAYFSDV DLEKDVHSGL IGPLLVCHTN TLNPAHGRQV TVQEFALFFT  
 1301 IFDETKSWYF TENMERN CRA PCNIQMEDPT FKENYRFHAI NGYIMDTLPG  
 1351 LVMAQDQRIR WYLLSMGSNE NIHSIHFSGH VFTVRKKEY KMALYNLYPG  
 1401 VFETVEMLPS KAGIWRVECL IGEHLHAGMS TLFLVYSNKC QTPLGMASGH  
 1451 IRDFQITASG QYGQWAPKLA RLHYSGSINA WSTKEPFSWI KVDLLAPMII  
 1501 HGIKTQGARQ KFSSLYISQF IIMYSLDGKK WQTYRGNSTG TLMVFFGNVD  
 1551 SSGIKHNIFN PPIIARYIRL HPTHYSIRST LRMELMGCDL NSCSMPLGME  
 1601 SKAISDAQIT ASSYFTNMFA TWSPSKARLH LQGRSNAWRP QVNNPKEWLQ  
 1651 VDFQKTMKVT GVTTQGVKSL LTSMYVKEFL ISSSQDGHQW TLFFQNGKVK  
 1701 VFQGNQDSFT PVVNSLDPPL LTRYLRIHPQ SWVHQIALRM EVLGCEAQL  
 1751 YDKTHTCPPC PAPELLGGPS VFLFPPKPKD TLMASRTPEV TCVVVDVSHE  
 1801 DPEVKFNWYV DGVEVHNAKT KPREEQYNST YRVVSVLTVL AQDWLNGKEY  
 1851 KCKVSNKALP APIEKTISKA KGQPREPQVY TLPPSRDEL T KNQVSLTCLV  
 1901 KGFYPSDIAV EWESNGQPEN NYKTTPPVLD SDGSFFLYSK LTVDKSRWQQ  
 1951 GNVFSCSVMH EALHNAYTQK SLSLSPGK\*

[0659] pSYNFVIII 010核苷酸序列- (双链FVIIIIFc) (SEQ ID NO:162)

1 ATGCAAATAG AGCTCTCCAC CTGCTTCTTT CTGTGCCTTT TGCGATTCTG  
 51 CTTTAGTGCC ACCAGAAGAT ACTACCTGGG TGCAGTGGAA CTGTCATGGG  
 101 ACTATATGCA AAGTGATCTC GGTGAGCTGC CTGTGGACGC AAGATTTCTT  
 [0660] 151 CCTAGAGTGC CAAAATCTTT TCCATTCAAC ACCTCAGTCG TGTACAAAAA  
 201 GACTCTGTTT GTAGAATTCA CGGATCACCT TTTCAACATC GCTAAGCCAA  
 251 GGCCACCCTG GATGGGTCTG CTAGGTCCTA CCATCCAGGC TGAGGTTTAT  
 301 GATACAGTGG TCATTACACT TAAGAACATG GCTTCCCATC CTGTCAGTCT  
 351 TCATGCTGTT GGTGTATCCT ACTGGAAAGC TTCTGAGGGA GCTGAATATG

[0661]

```

401   ATGATCAGAC CAGTCAAAGG GAGAAAAGAAG ATGATAAAGT CTTCCCTGGT
451   GGAAGCCATA CATATGTCTG GCAGGTCCCTG AAAGAGAATG GTCCAATGGC
501   CTCTGACCCA CTGTGCCTTA CCTACTCATA TCTTTCTCAT GTGGACCTGG
551   TAAAAGACTT GAATTCAGGC CTCATTGGAG CCCTACTAGT ATGTAGAGAA
601   GGGAGTCTGG CCAAGGAAAA GACACAGACC TTGCACAAAT TTATACTACT
651   TTTTGCTGTA TTTGATGAAG GGAAAAGTTG GCACTCAGAA ACAAAGAACT
701   CTTTGATGCA GGATAGGGAT GCTGCATCTG CTCGGGCCTG GCCTAAAATG
751   CACACAGTCA ATGGTTATGT AAACAGGTCT CTGCCAGGTC TGATTGGATG
801   CCACAGGAAA TCAGTCTATT GGCATGTGAT TGGAATGGGC ACCACTCCTG
851   AAGTGCACTC AATATTCCTC GAAGGTCACA CATTTCCTGT GAGGAACCAT
901   CGCCAGGCGT CCTTGGAAT CTGCGCAATA ACTTTCCTTA CTGACTCAAAC
951   ACTCTTGATG GACCTTGGAC AGTTTCTACT GTTTTGTCAT ATCTCTTCCC
1001  ACCAACATGA TGGCATGGAA GCTTATGTCA AAGTAGACAG CTGTCCAGAG
1051  GAACCCCAAC TACGAATGAA AAATAATGAA GAAGCGGAAG ACTATGATGA
1101  TGATCTTACT GATTCTGAAA TGGATGTGGT CAGGTTTGAT GATGACAAC
1151  CTCCTTCCCT TATCCAAAT CGCTCAGTTG CCAAGAAGCA TCCTAAAAC
1201  TGGGTACATT ACATTGCTGC TGAAGAGGAG GACTGGGACT ATGCTCCCTT
1251  AGTCCTCGCC CCCGATGACA GAAGTTATAA AAGTCAATAT TTGAACAATG
1301  GCCCTCAGCG GATTGGTAGG AAGTACAAAA AAGTCCGATT TATGGCATA
1351  ACAGATGAAA CCTTTAAGAC TCGTGAAGCT ATTCAGCATG AATCAGGAAT
1401  CTTGGGACCT TTACTTTATG GGGAAAGTTG AGACACACTG TTGATTATAT
1451  TTAAGAATCA AGCAAGCAGA CCATATAACA TCTACCCTCA CGGAATCACT
1501  GATGTCCGTC CTTTGTATTC AAGGAGATTA CCAAAGGTG TAAAACATTT
1551  GAAGGATTTT CCAATTCTGC CAGGAGAAAT ATTCAAATAT AAATGGACAG
1601  TGACTGTAGA AGATGGGCCA ACTAAATCAG ATCCTCGGTG CCTGACCCGC
1651  TATTACTCTA GTTTCGTTAA TATGGAGAGA GATCTAGCTT CAGGACTCAT
1701  TGGCCCTCTC CTCATCTGCT ACAAAGAATC TGTAGATCAA AGAGGAAACC
1751  AGATAATGTC AGACAAGAGG AATGTCAATC TGTTTTCTGT ATTTGATGAG
1801  AACCGAAGCT GGTACCTCAC AGAGAATATA CAACGCTTTC TCCCAATCC
1851  AGCTGGAGTG CAGCTTGAGG ATCCAGAGTT CCAAGCCTCC AACATCATGC
1901  ACAGCATCAA TGGCTATGTT TTTGATAGTT TGCAGTTGTC AGTTTGTGTTG
1951  CATGAGGTGG CATACTGGTA CATCTAAGC ATTGAGCAC AGACTGACTT
2001  CCTTCTGTCT TTCTTCTCTG GATATACCTT CAAACACAAA ATGGTCTATG
2051  AAGACACT CACCTATTC CCATTCTCAG GAGAACTGT CTTGATCTCG
2101  ATGGAAAACC CAGGTCTATG GATCTGGGG TGCCACAAC CTGACTTTCG
2151  GAACAGAGGC ATGACCGCCT TACTGAAGGT TTCTAGTTGT GACAAGAACA
2201  CTGGTGATTA TTACGAGGAC AGTTATGAAG ATATTTAGC ATACTTGCTG
2251  AGTAAAAACA ATGCCATTGA ACCAAGAAGC TTCTCTCAA ACCCACCAGT
2301  CTTGAAACGC CATCAACGGG AAATAACTCG TACTACTCTT CAGTCAGATC
2351  AAGAGGAAAT TGACTATGAT GATACCATAT CAGTTGAAAT GAAGAAGGAA
2401  GATTTTGACA TTTATGATGA GGATGAAAAT CAGAGCCCC GCAGCTTTC
2451  AAAGAAAACA CGACACTATT TTATTGCTGC AGTGGAGAGG CTCTGGGATT
2501  ATGGGATGAG TAGCTCCCCA CATGTTCTAA GAAACAGGGC TCAGAGTGGC
2551  AGTGTCCCTC AGTTCAGAA AGTTGTTTTC CAGGAATTTA CTGATGGCTC
2601  CTTTACTCAG CCCTTATACC GTGGAGAACT AAATGAACAT TTGGGACTCC
2651  TGGGGCCATA TATAAGAGCA GAAGTTGAAG ATAATATCAT GGTAACCTTTC
2701  AGAAATCAGG CCTCTCGTCC CTATTCCCTC TATTCTAGCC TTATTTCTTA
2751  TGAGGAAGAT CAGAGGCAAG GAGCAGAACC TAGAAAAAAC TTTGTCAAGC
2801  CTAATGAAAC CAAACTTAC TTTTGGAAAG TGCAACATCA TATGGCACCC
2851  ACTAAAGATG AGTTTGACTG CAAAGCCTGG GCTTATTTCT CTGATGTTGA
2901  CCTGGAAAAA GATGTGCACT CAGGCCTGAT TGGACCCCTT CTGGTCTGCC
2951  AACTAACAC ACTGAACCTT GCTCATGGGA GACAAGTGAC AGTACAGGAA
3001  TTTGCTCTGT TTTTCACCAT CTTTGATGAG ACCAAAAGCT GGTACTTCAC
3051  TGAAAATATG GAAAGAACT GCAGGGCTCC CTGCAATATC CAGATGGAAG
3101  ATCCCACTTT TAAAGAGAAT TATCGCTTCC ATGCAATCAA TGGCTACATA
3151  ATGGATACAC TACCTGGCTT AGTAATGGCT CAGGATCAA GGATTCGATG
3201  GTATCTGCTC AGCATGGGCA GCAATGAAAA CATCCATTCT ATTCATTTCA
3251  GTGGACATGT GTTCACTGTA CGAAAAAAG AGGAGTATA AATGGCACTG
3301  TACAATCTCT ATCCAGGTGT TTTTGAGACA GTGGAAATGT TACCATCCAA

```

3351 AGCTGGAATT TGGCGGGTGG AATGCCTTAT TGGCGAGCAT CTACATGCTG  
3401 GGATGAGCAC ACTTTTTCTG GTGTACAGCA ATAAGTGTCA GACTCCCCTG  
3451 GGAATGGCTT CTGGACACAT TAGAGATTTT CAGATTACAG CTTCAGGACA  
3501 ATATGGACAG TGGGCCCAA AGCTGGCCAG ACTTCATTAT TCCGGATCAA  
3551 TCAATGCCTG GAGCACC AAG GAGCCCTTTT CTTGGATCAA GGTGGATCTG  
3601 TTGGCACCAA TGATTATTCA CGGCATCAAG ACCCAGGGTG CCCGTCAGAA  
3651 GTTCTCCAGC CTCTACATCT CTCAGTTTAT CATCATGTAT AGTCTTGATG  
3701 GGAAGAAGTG GCAGACTTAT CGAGGAAATT CCACTGGAAC CTTAATGGTC  
3751 TTCTTTGGCA ATGTGGATTC ATCTGGGATA AAACACAATA TTTTAAACCC  
3801 TCCAATTATT GCTCGATACA TCCGTTTGCA CCCAACTCAT TATAGCATTC  
3851 GCAGCACTCT TCGCATGGAG TTGATGGGCT GTGATTTAAA TAGTTGCAGC  
3901 ATGCCATTGG GAATGGAGAG TAAAGCAATA TCAGATGCAC AGATTACTGC  
3951 TTCATCCTAC TTTACCAATA TGTTTGCCAC CTGGTCTCCT TCAAAAGCTC  
4001 GACTTCACCT CCAAGGGAGG AGTAATGCCT GGAGACCTCA GGTGAATAAT  
4051 CCAAAAAGAGT GGCTGCAAGT GGACTTCCAG AAGACAATGA AAGTCACAGG  
4101 AGTAACTACT CAGGGAGTAA AATCTCTGCT TACCAGCATG TATGTGAAGG  
4151 AGTTCCTCAT CTCCAGCAGT CAAGATGGCC ATCAGTGGAC TCTCTTTTTT  
[0662] 4201 CAGAATGGCA AAGTAAAGGT TTTTCAGGGA AATCAAGACT CTTTCACACC  
4251 TGTGGTGAAC TCTCTAGACC CACCGTTACT GACTCGCTAC CTTCGAATTC  
4301 ACCCCCAGAG TTGGGTGCAC CAGATTGCC TGAGGATGGA GGTTCTGGGC  
4351 TGCAGGCAC AGGACCTCTA CGACAAA ACT CACACATGCC CACCGTGCCC  
4401 AGCTCCAGAA CTCCTGGGCG GACCGTCAGT CTTCCTCTTC CCCCCAAAC  
4451 CCAAGGACAC CCTCATGATC TCCCGGACCC CTGAGGTCAC ATGCGTGGTG  
4501 GTGGACGTGA GCCACGAAGA CCCTGAGGTC AAGTTCAACT GGTACGTGGA  
4551 CGGCGTGGAG GTGCATAATG CCAAGACAAA GCCCGGGAG GAGCAGTACA  
4601 ACAGCACGTA CCGTGTGGTC AGCGTCCTCA CCGTCCTGCA CCAGGACTGG  
4651 CTGAATGGCA AGGAGTACAA GTGCAAGGTC TCCAACAAAG CCCTCCCAGC  
4701 CCCCATCGAG AAAACCATCT CCAAAGCCAA AGGGCAGCCC CGAGAACCAC  
4751 AGGTGTACAC CCTGCCCCCA TCCCGGGATG AGCTGACCAA GAACCAGGTC  
4801 AGCCTGACCT GCCTGGTCAA AGGCTTCTAT CCCAGCGACA TCGCCGTGGA  
4851 GTGGGAGAGC AATGGGCAGC CGGAGAACAA CTACAAGACC ACGCCTCCCG  
4901 TGTTGGACTC CGACGGCTCC TTCTTCCTCT ACAGCAAGCT CACCGTGGAC  
4951 AAGAGCAGGT GGCAGCAGGG GAACGTCTTC TCATGCTCCG TGATGCATGA  
5001 GGCTCTGCAC AACCCTACA CGCAGAAGAG CCTCTCCCTG TCTCCGGGTA  
5051 AATGA

[0663] pSYNFVIII 010蛋白质序列-(双链FVIII Fc) (SEQ ID NO:163)

1	MQIELSTCF	LCLLRFCFSA	TRRYYLGA	LSWDYMQSD	GELPVDARFP
51	PRVPKSF	TSVVYKTL	VEFTDHLF	AKPRPPWM	LGPTIQAEV
101	DTVVITL	ASHPVSLH	GVSYWKAS	AEYDDQTS	EKEDDKVFP
151	GSHTYVW	KENGPMA	LCLTYSYL	VDLVKDLN	LIGALLVCR
201	GSLAKEKT	LHKFILLF	FDEGKSWH	TKNSLMQD	AASARAWPK
251	HTVNGYV	LPGLIGCH	SVYWHVIG	TTPEVHSI	EGHTFLVRN
301	RQASLEIS	TFLTAQTLL	DLGQFLLF	ISSHQHDG	AYVKVDSC
351	EPQLRMKN	EAEDYDDDL	DSEMDVVR	DDNSPSFI	RSVAKKHPK
401	WVHYIAA	DWDYAPLV	PDDRSYKS	LNNGPQR	KYKKVRFM
451	TDEFKTR	IQHESGIL	LLYGEVGD	LIIFKNQAS	PYNIYPHG
501	DVRPLYSR	PKGVKHLK	PILPGEIF	KWTVTVED	TKSDPRCL
551	YYSSFVN	DLASGLIG	LICYKESV	RGNQIMSD	NVILFSVF
601	NRSWYL	QRFLPNPA	QLEDPEFQ	NIMHSING	FDSLQLSV
651	HEVAYWY	IGAQTDFL	FFSGYTFK	MVYEDTL	PFSGETVFM
701	MENPGLW	CHNSDFRN	MTALLKV	DKNTGDY	SYEDISAY
751	SKNNAIE	FSQNPPVL	HQREITRT	QSDQEEID	DTISVEMK
801	DFDIYDE	QSPRSFQK	RHYFIAAV	LWDYGMSS	HVLRNRAQ
[0664] 851	SVPQFKK	QEFTDGSF	PLYRGELN	LGLLGPYI	EVEDNIMV
901	RNQASRP	YSSLISYE	QRQGAEP	FVKPNETK	FWKVQHMA
951	TKDEFDCK	AYFSDVDL	DVHSGLI	LVCHTNTL	AHGRQVTV
1001	FALFFTIF	TKSWYFTEN	ERNCRAPCN	QMEDPTFK	YRFHAINGY
1051	MDTLPLG	VDQIRWYL	SMGSNENI	IHFSGHVFT	RKKEEYKMA
1101	YNLYPGV	FETVEMLP	SKAGIWRV	ECLIGELH	LHAGMSTLF
1151	GMASGH	IRDFQITAS	GQYQGWAP	KLARLHYS	SGSINAWST
1201	LAPMIIH	GIKTQGAR	QKFSSLYI	SQFIIMYSL	DGKKWQTYR
1251	FFGNVD	SSGKHNI	FNPIIARY	IRLHPHYS	SIRSTLRME
1301	MPLGMES	KAISDAQIT	ASSYFTN	MFATWSPSK	ARLHLQGRS
1351	PKEWLQV	DFQKTMKV	TGVTTQGV	KSLTSMYV	KEFLISSQD
1401	QNGKVKV	FQGNQDS	FTPVVNSL	DPPLLTRYL	RIHPQSWVH
1451	CEAQDLY	DKTHTCPP	CPAPELLG	GPSVFLFPP	KPKDTLMSR
1501	VDVSHED	PEVKFNWY	VDGVEVH	NAKTKPRE	EQYNSTYRV
1551	LNGKEYK	CKVSNKAL	PAPIEKTI	SKAKGQPRE	QVYTLPPSR
1601	SLTCLVK	GFYPSDIA	VEWESNGQ	PENNYKTPP	VLDSDGSFF
1651	KSRWQQG	NVFS	SCSVMHEALH	NHYTQKSLSL	SPGK*

[0665] FVIII 195蛋白质序列 (在氨基酸1656和1900处具有两个144 AE XTEN的双链 FVIII Fc) (SEQ ID NO:73)

```

1      MQIELSTCFF LCLLRFCFSA TRRYYLGAVE LSWDYMQSDL GELPVDARFP
51     PRVPKSFPFN TSVVYKKTFL VEFTDHLFNI AKPRPPWMGL LGPTIQAEVY
101    DTVVITLKNM ASHPVSLHAV GVSYWKASEG AEYDDQTSQR EKEDDKVFPF
151    GSHTYVWQVL KENGPASDP LCLTYSYLSH VDLVKDLNSG LIGALLVCRE
201    GSLAKEKTQT LHKFILLFAV FDEGKSWHSE TKNSLMQDRD AASARAWPKM
251    HTVNGYVNRS LPGLIGCHRK SVYWHVIGMG TTPEVHSIFL EGHTFLVRNH
301    RQASLEISPI TFLTAQTLLM DLGQFLLFCH ISSHQHDGME AYVKVDSCPE
351    EPQLRMKNNE EAEDYDDDLT DSEMDVVRFD DDNSPSFIQI RSVAKKHPKT
401    WVHYIAAEEE DWDYAPLVLA PDDRSYKSQY LNNGPQRIGR KYKKVRFMAY
451    TDETFKTREA IQHESGILGP LLYGEVGDTL LIIFKNQASR PYNIPHGIT
501    DVRPLYSRRL PKGVKHLKDF PILPGEIFKY KWTVTVEDGP TKSDPRCLTR
551    YYSSFVNMER DLASGLIGPL LICYKESVDQ RGNQIMSDKR NVILFSVFDE
601    NRSWYLTEI QRFLPNPAGV QLEDPEFQAS NIMHSINGYV FDSLQLSVCL
651    HEVAYWYILS IGAQTDFLSV FFSGYTFKHK MVYEDTLTLF PFSGETVFMS
701    MENPGLWILG CHNSDFRNRG MTALLKVSSC DKNTGDYED SYEDISAYLL
751    SKNNAIEPRS FSQNPPVLKR HQREITRRTL QGAPGTPGSG TASSSPGASP
801    GTSSTGSPGA SPGTSSTGSP GASPGTSSTG SPGSSPSAST GTGPGTPGSG
851    TASSSPGASP GTSSTGSPGA SPGTSSTGSP GASPGTSSTG SPGSSTPSGA
901    TGSPGSSTPS GATGSPGASP GTSSTGSPAS SSDQEEIDYD DTISVEMKKE
[0666] 951    DFDIYDEDEN QSPRSFQKKT RHYFIAAVER LWDYGMSSSP HVLNRNAQSG
1001   SVPQFKKVFV QEFTDGSFTQ PLYRGELNEH LGLLGPYIRA EVEDNIMVTF
1051   RNQASRPYSF YSSLISYEED QRQGAEPKRN FVKPNETKTY FWKVQHMAP
1101   TKDEFDCKAW AYFSDVDLEK DVHSGLIGPL LVCHTNTLNP AHGRQVTVQE
1151   FALFFTIFDE TKSIFYTENM ERNCRGAPTS ESATPESGPG SEPATSGSET
1201   PGTSESATPE SGPGSEPATS GSETPGTSES ATPESGPGTS TEPSEGSAPG
1251   TSESATPESG PGSPAGSPTS TEEGSPAGSP TSTEEGSPAG SPTSTEEGTS
1301   ESATPESGPG TSTEPSEGS SA PGASSAPCNI QMEDPTFKEN YRFHAINGYI
1351   MDTLPGLVMA QDQRIRWYLL SMGSNENIHS IHFSGHVFTV RKKEEYKMAL
1401   YNLYPGVFET VEMLPKAGI WRVECLIGEH LHAGMSTLFL VYSNKCQTP
1451   GMASGHIRDF QITASGQYQW WAPKLARLHY SGSINAWSTK EPFSWIKVDL
1501   LAPMI IHGK TQGARQKFSS LYISQFIIMY SLDGKKWQTY RGNSTGTLMV
1551   FFGNVDSSGI KHNI FNPII ARYIRLHPTH YSIRSTLRME LMGCDLNSCS
1601   MPLGMESKAI SDAQITASSY FTNMFATWSP SKARLHLQGR SNAWRPQVNN
1651   PKEWLQVDFQ KTMKVTGVT QGVKSLTSM YVKEFLISS QDGHQWTLFF
1701   QNGKVKVFQG NQDSFTPVVN SLDPPLLTRY LRIHPQSWVH QIALRMEVLG
1751   CEAQDLYDKT HTCPCPAPE LLGGPSVFLF PPKPKDTLMI SRTPEVTCVV
1801   VDVSHEDPEV KFNWYVDGVE VHNAKTKPRE EQYNSTYRVV SVLTVLHQDW
1851   LNGKEYKCKV SNKALPAPIE KTISKAKGQP REPQVYTLPP SRDELTKNQV
1901   SLTCLVKGFY PSDIAVEWES NGQPENNYKT TPPVLDSGDS FFLYSKLTVD
1951   KSRWQQGNVF SCSVMHEALH NHYTQKSLSL SPGK*

```

[0667] pSYN-FVIII-173成熟蛋白质序列(SEQ ID NO:72):



```

1   ATRRYYLGA V ELSWDYMQSD LGELPVDARF PPRVPKSFPF NTSVVKKTL
51  FVEFTDHLFN IAKPRPPWGM LLGPTIQAEV YDTVVITLKN MASHPVSLHA
101 VGVSYWKASE GA EYDDQTSQ REKEDDKVFP GGSHTYVWQV LKENGPMASD
151 PLCLTYSYLS HVDLVKDLNS GLIGALLVCR EGSLAKEKTQ TLHKFILLFA
201 VFDEGKSWHS ETKNSLMQDR DAASARAWPK MHTVNGYVNR SLPGLIGCHR
251 KSVYWHVIGM GTTPEVHSIF LEGHTFLVRN HRQASLEISP ITFLTAQTLL
301 MDLGQFLLFC HISSHQHDGM EAYVKVDSCP EEPQLRMKNN EEAEDYDDDL
351 TDSEMDVVRF DDDNSPSFIQ IRSVAKKHPK TWVHYIAAEE EDWDYAPLVL
401 APDDRSYKSQ YLNNGPQRIG RKYKKVRFMA YTDEFKTR E AIQHESGILG
451 PLLYGEVGD T LLIIFKNQAS RPYNIYPHGI TDVRPLYSRR LPKGVKHLKD
501 FPILPGEIFK YKWTVTVEDG PTKSDPRCLT RYSSSFVNME RDLASGLIGP
551 LLICYKESVD QRGNQIMSDK RNVILFSVFD ENRSWYL TEN IQRFLPNPAG
601 VQLEDPEFQA SNIMHSINGY VFDSLQLSVC LHEVAYWYIL SIGAQTDFLS
651 VFFSGYTFKH KMVYEDTLTL FPFSGETVFM SMENPGLWIL GCHNSDFRNR
701 GMTALLKVSS CDKNTGDYEE DSYEDISAYL LSKNNAIEPR SFSQNGAPGT
751 SESATPESGP GSEPATSGSE TPGTSESATP ESGPGSEPAT SGSETPGTSE
801 SATPESGPGT STEPSEGSAP GSPAGSPTST EEGTSESATP ESGPGSEPAT
851 SGSETPGTSE SATPESGPGS PAGSPTSTEE GSPAGSPTST EEGTSTEPSE
901 GSAPGTSESA TPESGPGTSE SATPESGPGT SESATPESGP GSEPATSGSE
951 TPGSEPATSG SETPGSPAGS PTSTEEGTST EPSEGSAPGT STEPSEGSAP
[0668] 1001 GSEPATSGSE TPGTSESATP ESGPGTSTEP SEGSAPASSP PVLKRHQREI
1051 TRTTLQSDQE EIDYDDTISV EMKKEDFDIY DEDENQSPRS FQKKTRHYFI
1101 AAVERLWDYG MSSSPHVLRN RAQSGSVPQF KKVVVFQFTD GSFTQPLYRG
1151 ELNEHLGLLG PYIRAEVEDN IMVTFRNQAS RPYSFYSSLI SYEEDQRQGA
1201 EPRKNFVKPN ETKTYFWKVQ HHMAPTKDEF DCKAWAYFSD VDLEKDVHSG
1251 LIGPLLVCHT NTLNPAHGRQ VTVQEFALFF TIFDETKSWY FTENMERNCR
1301 APCNIQMEDP TFKENYRFHA INGYIMDTLP GLVMAQDQRI RWYLLSMGSN
1351 ENIHSIHFSG HVFTVRKKEE YKMALYNLYP GVFETVEMLP SKAGIWRVEC
1401 LIGEHLHAGM STLFLVYSNK CQTPLGMASG HIRDFQITAS GOYGQWAPKL
1451 ARLHYSGSIN AWSTKEPFSW IKVDLLAPMI IHGIKTQGAR QKFSSLYISQ
1501 FIIMYSLDGK KWQTYRGNST GTLMVFFGNV DSSGIKHNI F NPPIIARYIR
1551 LHPTHYSIRS TLRMELMGCD LNCSMPLGM ESKAISDAQI TASSYFTNMF
1601 ATWSPSKARL HLQGRSNAWR PQVNNPKEWL QVDFQKTMKV TGVTTQGVKS
1651 LLTSMYVKEF LISSSQDGHQ WTLFFQNGKV KVFQGNQDSF TPVVNSLDPP
1701 LLTRYLRIHP QSWVHQIALR MEVLGCEAQD LYDKTHTCPP CPAPELLGGP
1751 SVFLFPPKPK DTLMISRTPE VTCVVVDVSH EDPEVKFNWY VDGVEVHNAK
1801 TKPREEQYNS TYRVVSVLTV LHQDWLNGKE YKCKVSNKAL PAPIEKTISK
1851 AKGQPREPQV YTLPPSRDEL TKNQVSLTCL VKGFYPSDIA VEWESNGQPE
1901 NNYKTTTPVL DSDGSFFLYS KLTVDKSRWQ QGNVFSCSVM HEALHNHYTQ
1951 KSLSLSPGK

```

[0669] FVIII 196蛋白质序列(在氨基酸26、1656和1900处具有三个144AE XTEN的双链 FVIIIIFc) (SEQ ID NO:74)

1	MQIELSTCFF	LCLLRFCFSA	TRRYYLGAVE	LSWDYMQSDL	GELPVGAPGS
51	SPSASTGTGP	GSSPSASTGT	GPGASPGTSS	TGSPGASPGT	SSTGSPGSST
101	PSGATGSPGS	SPSASTGTGP	GASPGTSSSTG	SPGSSPSAST	GTGPGTPGSG
151	TASSSPGSST	PSGATGSPGS	STPSGATGSP	GASPGTSSSTG	SPASSDARFP
201	PRVPKSFPPN	TSVVYKKTFL	VEFTDHLFNI	AKPRPPWMGL	LGPTIQAEVY
251	DTVVITLKNM	ASHPVSLHAV	GVSYWKASEG	AEYDDQTSQR	EKEDDKVFPG
301	GSHTYVWQVL	KENGPMSADP	LCLTYSYLSH	VDLVKDLNSG	LIGALLVCRE
351	GSLAKEKTQT	LHKFILLFAV	FDEGKSWHSE	TKNSLMQDRD	AASARAWPKM
401	HTVNGYVNRS	LPGLIGCHRK	SVYWHVIGMG	TTPEVHSIFL	EGHTFLVRNH
451	RQASLEISPI	TFLTAQTLLM	DLGQFLFCH	ISSHQHDGME	AYVKVDSCPE
501	EPQLRMKNNE	EAEDYDDDLT	DSEMDVVRFD	DDNSPSFIQI	RSVAKKHPKT
551	WVHYIAAEEE	DWDYAPLVLA	PDDRSYKSQY	LNNGPQRIGR	KYKKVRFMAY
601	TDETFKTREA	IQHESGILGP	LLYGEVGDTL	LIIFKNQASR	PYNIYPHGIT
651	DVRPLYSRRL	PKGVKHLKDF	PILPGEIFKY	KWTVTVEDGP	TKSDPRCLTR
701	YYSSFVNMER	DLASGLIGPL	LICYKESVDQ	RGNQIMSDKR	NVILFSVFDE
751	NRSWYLTEI	QRFLPNPAGV	QLEDPEFQAS	NIMHSINGYV	FDSLQLSVCL
801	HEVAYWYILS	IGAQTDFLSV	FFSGYTFKHK	MVYEDTLTLF	PFSGETVFMMS
851	MENPGLWILG	CHNSDFRNRG	MTALLKVSSC	DKNTGDYED	SYEDISAYLL
901	SKNNAIEPRS	FSQNPPVLKR	HQREITRRTL	QGAPGTPGSG	TASSSPGASP
951	GTSSSTGSPGA	SPGTSSSTGSP	GASPGTSSSTG	SPGSSPSAST	GTGPGTPGSG
[0670] 1001	TASSSPGASP	GTSSSTGSPGA	SPGTSSSTGSP	GASPGTSSSTG	SPGSSTPSGA
1051	TGSPGSSTPS	GATGSPGASP	GTSSSTGSPAS	SSDQEEIDYD	DTISVEMKKE
1101	DFDIYDEDEN	QSPRSFQKKT	RHYFIAAVER	LWDYGMSSSP	HVLRNRAQSG
1151	SVPQFKKVVF	QEFTDGSFTQ	PLYRGELNEH	LGLLGPYIRA	EVEDNIMVTF
1201	RNQASRPYSF	YSSLISYEED	QRQGAEPKRN	FVKPNETKTY	FWKVQHMAP
1251	TKDEFDCKAW	AYFSDVDLEK	DVHSGLIGPL	LVCHTNTLNP	AHGRQVTVOE
1301	FALFFTIFDE	TKSWYFTENM	ERNCRGAPTS	ESATPESGPG	SEPATSGSET
1351	PGTSESATPE	SGPGSEPATS	GSETPGTSES	ATPESGPGTS	TEPSEGSAPG
1401	TSESATPESG	PGSPAGSPTS	TEEGSPAGSP	TSTEEGSPAG	SPTSTEEGTS
1451	ESATPESGPG	TSTEPSEGS	PGASSAPCNI	QMEDPTFKEN	YRFHAINGYI
1501	MDTLPGLVMA	QDQRIRWYLL	SMGSNENIHS	IHFSGHVFTV	RKKEEYKMAL
1551	YNLYPGVFET	VEMLPSKAGI	WRVECLIGEH	LHAGMSTLFL	VYSNKCQTPL
1601	GMASGHIRDF	QITASGQYGO	WAPKLARLHY	SGSINAWSTK	EPFSWIKVDL
1651	LAPMIIHGIK	TQGARQKFSS	LYISQFIIMY	SLDGKKWQTY	RGNSTGTLMV
1701	FFGNVDSSGI	KHNIFNPPII	ARYIRLHPTH	YSIRSTLRME	LMGCDLNSSC
1751	MPLGMESKAI	SDAQITASSY	FTNMFATWSP	SKARLHLQGR	SNAWRPQVNN
1801	PKEWLQVDFQ	KTMKVTGVT	QGVKSLTSM	YVKEFLISS	QDGHQWTLFF
1851	QNGKVVFQ	NQDSFTPVVN	SLDPPLLTRY	LRIHPQSWVH	QIALRMEVLG
1901	CEAQDLYDKT	HTCPPCPAPE	LLGGPSVFLF	PPKPKDTLMI	SRTPEVTCVV
1951	VDVSHEDPEV	KFNWYVDGVE	VHNAKTKPRE	EQYNSTYRVV	SVLTVLHQDW
2001	LNGKEYKCKV	SNKALPAPIE	KTISKAKGQP	REPQVYTLPP	SRDELTKNQV
2051	SLTCLVKGFY	PSDIAVEWES	NGQPENNYKT	TPPVLDSDGS	FFLYSKLTVD
2101	KSRWQQGNVF	SCSVMHEALH	NHYTQKSLSL	SPGK*	

[0671] FVIII 199蛋白质序列 (在氨基酸1656和1900处具有三个144 AE XTEN的单链 FVIIIIFc) (SEQ ID NO:75)

1	MQIELSTCFF	LCLLRFCFSA	TRRYYLGAVE	LSWDYMQSDL	GELPVDARFP
51	PRVPKSFPFN	TSVVYKKTFL	VEFTDHLFNI	AKPRPPWMGL	LGPTIQAEVY
101	DTVVITLKNM	ASHPVSLHAV	GVSYWKASEG	AEYDDQTSQR	EKEDDKVFPG
151	GSHTYVWQVL	KENGPMSADP	LCLTYSYLSH	VDLVKDLNSG	LIGALLVCRE
201	GSLAKEKTQT	LHKFILLFAV	FDEGKSWHSE	TKNSLMQDRD	AASARAWPKM
251	HTVNGYVNRS	LPGLIGCHRK	SVYWHVIGMG	TTPEVHSIFL	EGHTFLVRNH
301	RQASLEISPI	TFLTAQTLLM	DLGQFLLFCH	ISSHQHDGME	AYVKVDSCPE
351	EPQLRMKNNE	EAEDYDDDLT	DSEMDVVRFD	DDNSPSFIQI	RSVAKKHPKT
401	WVHYIAAEEE	DWDYAPLVLA	PDDRSYKSQY	LNNGPQRIGR	KYKKVRFMAY
451	TDETFKTREA	IQHESGILGP	LLYGEVGDTL	LIIFKNQASR	PYNIYPHGIT
501	DVRPLYSRRL	PKGVKHLKDF	PILPGEIFKY	KWTVTVEDGP	TKSDPRCLTR
551	YYSSFVNMER	DLASGLIGPL	LICYKESVDQ	RGNQIMSDKR	NVILFSVFDE
601	NRSWYL TENI	QRFLPNPAGV	QLEDPEFQAS	NIMHSINGYV	FDSLQLSVCL
651	HEVAYWYILS	IGAQTDFLSV	FFSGYTFKHK	MVYEDTLTLF	PFSGETV FMS
701	MENPGLWILG	CHNSDFRNRG	MTALLKVSSC	DKNTGDYIED	SYEDISAYLL
751	SKNNAIEPRS	FSQNPPVLKR	HQAEITR TTL	QGAPGTPGSG	TASSSPGASP
801	GTSSSTGSPGA	SPGTSSTGSP	GASPGTSSTG	SPGSSPSAST	GTGPGTPGSG
851	TASSSPGASP	GTSSSTGSPGA	SPGTSSTGSP	GASPGTSSTG	SPGSSTPSGA
901	TGSPGSSTPS	GATGSPGASP	GTSSSTGSPAS	SSDQEEIDYD	DTISVEMKKE
951	DFDIYDEDEN	QSPRSFQKKT	RHYFIAAVER	LWDYGMSSSP	HVLRNRAQSG
[0672] 1001	SVPQFKKVVV	QEFTDGSFTQ	PLYRGELNEH	LGLLGPYIRA	EVEDNIMVTF
1051	RNQASRPYSF	YSSLISYEED	QRQGAEPKRN	FVKPNETKTY	FWKVQH HMAP
1101	TKDEFDCKAW	AYFSDVDLEK	DVHSGLIGPL	LVCHTNTLNP	AHGRQVTVQE
1151	FALFFTIFDE	TKSWYFTENM	ERNCRGAPTS	ESATPESGPG	SEPATSGSET
1201	PGTSESATPE	SGPGSEPATS	GSETPGTSES	ATPESGPGTS	TEPSEGSAPG
1251	TSESATPESG	PGSPAGSPTS	TEEGSPAGSP	TSTEEGSPAG	SPTSTEEGTS
1301	ESATPESGPG	TSTEPSEGS A	PGASSAPCNI	QMEDPTFKEN	YRFHAINGYI
1351	MDTLPGLVMA	QDQRIRWYLL	SMGSNENIHS	IHFSGHVFTV	RKKEEYKMAL
1401	YNLYPGVFET	VEMLPSKAGI	WRVECLIGEH	LHAGMSTLFL	VYSNKCQ TPL
1451	GMASGHIRDF	QITASGQYGO	WAPKLARLHY	SGSINAWSTK	EPFSWIKVDL
1501	LAPMI IHGIK	TQGARQKFSS	LYISQFIIMY	SLDGKKWQTY	RGNSTGTLMV
1551	FFGNVDSSGI	KHNIFNPPII	ARYIRLHPTH	YSIRSTLRME	LMGCDLNSCS
1601	MPLGMESKAI	SDAQITASSY	FTNMFATWSP	SKARLHLQGR	SNAWRPQVNN
1651	PKEWLQVDFQ	KTMKVTGVT T	QGVKSLLTSM	YVKEFLISSS	QDGHQWTLFF
1701	QNGKVKVFQG	NQDSFTPVVN	SLDPPLLTRY	LRIHPQSWVH	QIALRMEVLG
1751	CEAQDLYDKT	HTCPPCPAPE	LLGGPSVFLF	PPKPKDTLMI	SRTPEVTCVV
1801	VDVSHEDPEV	KFNWYVDGVE	VHNAKTKPRE	EQYNSTYRVV	SVLTVLHQDW
1851	LNGKEYKCKV	SNKALPAPIE	KTISKAKGQP	REPQVYTLPP	SRDELTKNQV
1901	SLTCLVKGFY	PSDIAVEWES	NGQPENNYKT	TPPVLDS DGS	FFLYSKLTV D
1951	KSRWQQGNVF	SCSVMHEALH	NHYTQKSLSL	SPGK*	

[0673] FVIII 201蛋白质序列(在氨基酸26、1656和1900处具有三个144 AE XTEN的单链 FVIIIFc) (SEQ ID NO:76)

1	MQIELSTCFF	LCLLRFCFSA	TRRYYLGAVE	LSWDYMQSDL	GELPVGAPGS
51	SPSASTGTGP	GSSPSASTGT	GPGASPGTSS	TGSPGASPGT	SSTGSPGSST
101	PSGATGSPGS	SPSASTGTGP	GASPGTSSTG	SPGSSPSAST	GTGPGTPGSG
151	TASSSPGSST	PSGATGSPGS	STPSGATGSP	GASPGTSSTG	SPASSDARFP
201	PRVPKSFPFN	TSVVYKKTFL	VEFTDHLFNI	AKPRPPWMGL	LGPTIQAEVY
251	DTVVITLKNM	ASHPVSLHAV	GVSYWKASEG	AEYDDQTSQR	EKEDDKVFPG
301	GSHTYVWQVL	KENGPASDP	LCLTYSYLSH	VDLVKDLNSG	LIGALLVCRE
351	GSLAKEKTQT	LHKFILLFAV	FDEGKSWHSE	TKNSLMQDRD	AASARAWPKM
401	HTVNGYVNRS	LPGLIGCHRK	SVYWHVIGMG	TTPEVHSIFL	EGHTFLVRNH
451	RQASLEISPI	TFLTAQTLLM	DLGQFLLFCH	ISSHQHDGME	AYVKVDSCPE
501	EPQLRMKNNE	EAEDYDDDLT	DSEMDVVRFD	DDNSPSFIQI	RSVAKKHPKT
551	WVHYIAAEEE	DWDYAPLVLA	PDDRSYKSQY	LNNGPQRIGR	KYKKVRFMAY
601	TDETFKTREA	IQHESGILGP	LLYGEVGDTL	LIIFKNQASR	PYNIYPHGIT
651	DVRPLYSRRL	PKGVKHLKDF	PILPGEIFKY	KWTVTVEDGP	TKSDPRCLTR
701	YYSSFVNMER	DLASGLIGPL	LICYKESVDQ	RGNQIMSDKR	NVILFSVFDE
751	NRSWYL TENI	QRFLPNPAGV	QLEDPEFQAS	NIMHSINGYV	FDSLQLSVCL
801	HEVAYWYILS	IGAQTDFLSV	FFSGYTFKHK	MVYEDTLTLF	PFSGETVFMS
851	MENPGLWILG	CHNSDFRNRG	MTALLKVSSC	DKNTGDYDYE	SYEDISAYLL
901	SKNNAIEPRS	FSQNPPVLKR	HQAEITRRTL	QGAPGTPGSG	TASSSPGASP
951	GTSSSTGSPGA	SPGTSSTGSP	GASPGTSSTG	SPGSSPSAST	GTGPGTPGSG
1001	TASSSPGASP	GTSSSTGSPGA	SPGTSSTGSP	GASPGTSSTG	SPGSSTPSGA
[0674] 1051	TGSPGSSTPS	GATGSPGASP	GTSSSTGSPAS	SSDQEEIDYD	DTISVEMKKE
1101	DFDIYDEDEN	QSPRSFQKKT	RHYFIAAVER	LWDYGMSSSP	HVLRNRAQSG
1151	SVPQFKKVVV	QEFTDGSFTQ	PLYRGELNEH	LGLLGPYIRA	EVEDNIMVTF
1201	RNQASRPYSF	YSSLISYEED	QRQGAEPKRN	FVKPNETKTY	FWKVQHMAP
1251	TKDEFDCKAW	AYFSDVDLEK	DVHSGLIGPL	LVCHTNTLNP	AHGRQVTVQE
1301	FALFFTIFDE	TKSWYFTENM	ERNCRGAPTS	ESATPESGPG	SEPATSGSET
1351	PGTSESATPE	SGPGSEPATS	GSETPGTSES	ATPESGPGTS	TEPSEGSAPG
1401	TSESATPESG	PGSPAGSPTS	TEEGSPAGSP	TSTEEGSPAG	SPTSTEEGTS
1451	ESATPESGPG	TSTEPSEGS	PGASSAPCNI	QMEDPTFKEN	YRFHAINGYI
1501	MDTLPGLVMA	QDQIRWYLL	SMGSNENIHS	IHFSGHVFTV	RKKEEYKMAL
1551	YNLYPGVFET	VEMLPSKAGI	WRVECLIGEH	LHAGMSTLFL	VYSNKCQTPL
1601	GMASGHIRDF	QITASGQYGQ	WAPKLARLHY	SGSINAWSTK	EPFSWIKVDL
1651	LAPMIIHGK	TQGARQKFSS	LYISQFIIMY	SLDGKKWQTY	RGNSTGTLMV
1701	FFGNVDSSGI	KHNIFNPPII	ARYIRLHPTH	YSIRSTLRME	LMGCDLNSCS
1751	MPLGMESKAI	SDAQITASSY	FTNMFATWSP	SKARLHLQGR	SNAWRPQVNN
1801	PKEWLQVDFQ	KTMKVTGVTT	QGVKSLTSM	YVKEFLISS	QDGHQWTLFF
1851	QNGKVKVFGQ	NQDSFTPVVN	SLDPPLLTRY	LRIHPQSWVH	QIALRMEVLG
1901	CEAQDLYDKT	HTCPCPAPE	LLGGPSVFLF	PPKPKDTLMI	SRTPEVTCVV
1951	VDVSHEDPEV	KFNWYVDGVE	VHNAKTKPRE	EQYNSTYRVV	SVLTVLHQDW
2001	LNGKEYKCKV	SNKALPAPIE	KTISKAKGQP	REPQVYTLPP	SRDELTKNQV
2051	SLTCLVKGFY	PSDIAVEWES	NGQPENNYKT	TPPVLDSDGS	FFLYSKLTVD
2101	KSRWQQGNVF	SCSVMEALH	NHYTQKLSL	SPGK*	

[0675] FVIII 203蛋白质序列(具有两个AE XTEN的单链FVIIIIFc;一个288AE XTEN在B结构域中,并且一个144 AE XTEN在氨基酸1900处) (SEQ ID NO:77)

```

1      MQIELSTCFF LCLLRFCSA TRRYYLGA VE LSWDYMQSDL GELPVDARFP
51     PRVPKSFPFN TSVVYKKTFL VEFTDHLFNI AKPRPPWMGL LGPTIQAEVY
101    DTVVITLKNM ASHPVSLHAV GVSYWKA SEG AEYDDQTSQR EKEDDKVFPG
151    GSHTYVWQVL KENGPMSADP LCLTYSYLSH VDLVKDLNSG LIGALLVCRE
201    GSLAKEKTQT LHKFILLFAV FDEGKSWHSE TKNSLMQDRD AASARAWPKM
251    HTVNGYVNRS LPGLIGCHRK SVYWHVIGMG TTPEVHSIFL EGHTFLVRNH
301    RQASLEISPI TFLTAQTL LM DLGQFLLFCH ISSHQHDGME AYVKVDSCPE
351    EPQLRMKNNE EAEDYDDDLT DSEMDVVRFD DDNSPSFIQI RSVAKKHPKT
401    WVHYIAAEEE DWDYAPLVLA PDDRSYKSQY LNNGPQRIGR KYKKVRFMAY
451    TDETFKTREA IQHESGILGP LLYGEVGD TL LIIFKNQASR PYNIIYPHGIT
501    DVRPLYSRRL PKGVKHLKDF PILPGEIFKY KWTVTVEDGP TKSDPRCLTR
551    YYSSFVNMER DLASGLIGPL LICYKESVDQ RGNQIMSDKR NVILFSVFDE
601    NRSWYL TENI QRFLPNPAGV QLEDPEFQAS NIMHSINGYV FDSLQLSVCL
651    HEVAYWYILS IGAQTDFLSV FFSGYTFKHK MVYEDTLTLF PFSGETVFMS
701    MENPGLWILG CHNSDFRNRG MTALLKVSSC DKNTGDY YED SYEDISAYLL
751    SKNNAIEPRS FSQNGAPGTS ESATPESGPG SEPATSGSET PGTSESATPE
801    SGP GSEPATS GSETPGTSES ATPESGPGTS TEPSEGSAPG SPAGSPTSTE
851    EGTSESATPE SGP GSEPATS GSETPGTSES ATPESGPGSP AGSPTSTEEG
901    SPAGSPTSTE EGTSTEPSEG SAPGTSESAT PESGPGTSES ATPESGPGTS
951    ESATPESGPG SEPATSGSET PGSEPATSGS ETPGSPAGSP TSTEEGTSTE
1001   PSEGSAPGTS TEPSEGSAPG SEPATSGSET PGTSESATPE SGP GTSTEPS
[0676] 1051   EGSAPASSPP VLKRHQAEIT RTTLQSDQEE IDYDDTISVE MKKEDFDIYD
1101   EDENQSPRSF QKKTRHYFIA AVERLWDYGM SSSPHVLRNR AQSGSVPQFK
1151   KVVVFQFTDG SFTQPLYRGE LNEHLGLLGP YIRAEVEDNI MVTFRNQASR
1201   PYSFYSSLIS YEEDQRQGA E PRKNFVKPNE TKTYFWKVQH HMAPTKDEFD
1251   CKAWAYFSDV DLEKDVHSGL IGPLL VCHTN TLNPAHGRQV TVQEFALFFT
1301   IFDETKSWYF TENMERNCRG APTSESATPE SGP GSEPATS GSETPGTSES
1351   ATPESGPGSE PATSGSETPG TSESATPESG PGTSTEPSEG SAPGTSESAT
1401   PESGPGSPAG SPTSTEEGSP AGSPTSTEEG SPAGSPTSTE EGTSESATPE
1451   SGP GTSTEPS EGSAPGASSA PCNIQMEDPT FKENYRFHAI NGYIMDTLPG
1501   LVMAQDQRIR WYLLSMGSNE NIHSIHFSGH VFTVRKKEEY KMALYNLYPG
1551   VFETVEMLPS KAGIWRVECL IGEHLHAGMS TLFLVYSNKC QTPLGMASGH
1601   IRDFQITASG QYGQWAPKLA RLHYSGSINA WSTKEPFSWI KVDLLAPMII
1651   HGIKTQGARQ KFSSLYISQF IIMYSLDGKK WQTYRGNSTG TLMVFFGNVD
1701   SSGIKHNIFN PPIIARYIRL HPTHYSIRST LRMELMGCDL NSCSMPLGME
1751   SKAISDAQIT ASSYFTNMFA TWSPSKARLH LQGRSNAWRP QVNNPKEWLQ
1801   VDFQKTMKVT GVTTQGVKSL LTSMYVKEFL ISSSQDGHQW TLFFQNGKVK
1851   VFQGNQDSFT PVVNSLDPPL LTRYLRIH PQ SWVHQIALRM EVLGCEA QDL
1901   YDKTHTCPPC PAPELLGGPS VFLFPPKPKD TLMISRTPEV TCVVVDVSHE
1951   DPEVKFNWYV DGVEVHNAKT KPREEQYNST YRVVSVLTVL HQDWLNGKEY
2001   KCKVSNKALP APIEKTISKA KGQPREPQVY TLPPSRDELT KNQVSLTCLV
2051   KGFYPSDIAV EWESNGQPEN NYKTPPVLD SDGSFFLYSK LTVDKSRWQQ
2101   GNVFSCSVMH EALHNHYTQK SLSLSPGK*

```

[0677] FVIII 204蛋白质序列(具有两个AE XTEN的单链FVIIIIFc;一个288AE XTEN在B结构域中,并且一个144 AE XTEN在氨基酸403处)(SEQ ID NO:78)

[0678] 1 MQIELSTCFF LCLLRFCSA TRRYYLGA VE LSWDYMQSDL GELPVDARFP

51 PRVPKSFPFN TSVVYKKTFL VEFTDHLFNI AKPRPPWMGL LGPTIQAEVY  
101 DTVVITLKNM ASHPVSLHAV GVSYWKASEG AEYDDQTSQR EKEDDKVFPF  
151 GSHTYVWQVL KENGPMSADP LCLTYSYLSH VDLVKDLNSG LIGALLVCRE  
201 GSLAKEKTQT LHKFILLFAV FDEGKSWHSE TKNSLMQDRD AASARAWPKM  
251 HTVNGYVNRSLPGLIGCHRK SVYWHVIGMG TTPEVHSIFL EIGHTFLVRNH  
301 RQASLEISPI TFLTAQTLLM DLGQFLLFCH ISSHQHDGME AYVKVDSCPE  
351 EPQLRMKNNE EAEDYDDDLT DSEMDVVRFD DDNSPSFIQI RSVAKKHPKT  
401 WVHYIAAEEE DWDYAPLVLA PDGAPTSTEP SEGSAPGSPA GSPTSTEEGT  
451 STEPSEGSAP GTSTEPSEGS APGTSESATP ESGPGTSTEP SEGSAPGTSE  
501 SATPESGPGS EPATSGSETP GTSTEPSEGS APGTSTEPSE GSAPGTSESA  
551 TPESGPGTSE SATPESGPGA SDRSYKSQY LNNGPQRIGR KYKKVRFMAY  
601 TDEFKTRREA IQHESGILGP LLYGEVGDTL LIIFKNQASR PYNIYPHGIT  
651 DVRPLYSRRL PKGVKHLKDF PILPGEIFKY KWTVTVEDGP TKSDPRCLTR  
701 YYSSFVNMER DLASGLIGPL LICYKESVDQ RGNQIMSDKR NVILFSVFDE  
751 NRSWYL TENI QRFLPNPAGV QLEDPEFQAS NIMHSINGYV FDSLQLSVCL  
801 HEVAYWYILS IGAQTDFLSV FFSGYTFKHK MVYEDTLTLF PFSGETVFMMS  
851 MENPGLWILG CHNSDFRNRG MTALLKVSSC DKNTGDYED SYEDISAYLL  
901 SKNNAIEPRS FSQNGAPGTS ESATPESGPG SEPATSGSET PGTSESATPE  
951 SGPGSEPATS GSETPGTSES ATPESGPGTS TEPSEGSAPG SPAGSPTSTE  
1001 EGTSESATPE SGPGSEPATS GSETPGTSES ATPESGPGSP AGSPTSTEEG  
[0679] 1051 SPAGSPTSTE EGTSTEPSEG SAPGTSESAT PESGPGTSES ATPESGPGTS  
1101 ESATPESGPG SEPATSGSET PGSEPATSGS ETPGSPAGSP TSTEEGTSTE  
1151 PSEGSAPGTS TEPSEGSAPG SEPATSGSET PGTSESATPE SGPGTSTEPS  
1201 EGSAPASSPP VLKRHQAEIT RTTLQSDQEE IDYDDTISVE MKKEDFDIYD  
1251 EDENQSPRSF QKKTRHYFIA AVERLWDYGM SSSPHVLRNR AQSGSVPQFK  
1301 KVVVFQFTDG SFTQPLYRGE LNEHLGLLGP YIRAEVEDNI MVTFRNQASR  
1351 PYSFYSSLIS YEEDQRQGA PRKNFVKPNE TKTYFWKVQH HMAPTKDEFD  
1401 CKAWAYFSDV DLEKDVHSGL IGPLLVCHTN TLNPAHGRQV TVQEFALFFT  
1451 IFDETKSWYF TENMERN CRA PCNIQMEDPT FKENYRFHAI NGYIMDTLPG  
1501 LVMAQDQRIR WYLLSMGSNE NIHSIHFSGH VFTVRKKEEY KMALYNLYPG  
1551 VFETVEMLPS KAGIWRVECL IGEHLHAGMS TFLFLVYSNKC QTPLGMASGH  
1601 IRDFQITASG QYGQWAPKLA RLHYSGSINA WSTKEPFSWI KVDLLAPMI I  
1651 HGIKTQGARQ KFSSLYISQF IIMYSLDGKK WQTYRGNSTG TLMVFFGNVD  
1701 SSGIKHNIFN PPIIARYIRL HPTHYSIRST LRMELMGCDL NSCSMPLGME  
1751 SKAISDAQIT ASSYFTNMFA TWSPSKARLH LQGRSNAWRP QVNNPKEWLQ  
1801 VDFQKTMKVT GVTTQGVKSL LTSMYVKEFL ISSSQDGHQW TLFFQNGKVK  
1851 VFQGNQDSFT PVVNSLDPPL LTRYLRIHPQ SWVHQIALRM EVLGCEAQDL  
1901 YDKHTHCPPC PAPELLGGPS VFLFPPKPKD TLMISRTPEV TCVVVDVSHE  
1951 DPEVKFNWYV DGVEVHNAKT KPREEQYNST YRVVSVLTVL HQDWLNGKEY  
2001 KCKVSNKALP APIEKTISKA KGQPREPQVY TLPPSRDEL TKNQVSLTCLV  
2051 KGFYPSDIAV EWESNGQPEN NYKTTPPVLD SDGSFFLYSK LTVDKSRWQQ  
2101 GNVFSCSVMH EALHNHYTQK SLSLSPGK\*

[0680] FVIII 205蛋白质序列(具有两个AE XTEN的单链FVIIIIfc;一个288AE XTEN在B结构域中,并且一个144 AE XTEN在氨基酸18处)(SEQ ID NO:79)

1	MQIELSTCFF	LCLLRFCFSA	TRRYYLGAVE	LSWDYMQGAP	TSESATPESG
51	PGSEPATSGS	ETPGTSESAT	PESGPGSEPA	TSGSETPGTS	ESATPESGPG
101	TSTEPSEGS	PGSPAGSPTS	TEEGTSESAT	PESGPGSEPA	TSGSETPGTS
151	ESATPESGPG	SPAGSPTSTE	EGSPAGSPTS	TEEGASSSDL	GELPVDARFP
201	PRVPKSFPFN	TSVVYKKTFL	VEFTDHLFNI	AKPRPPWMGL	LGPTIQAEVY
251	DTVVITLKNM	ASHPVSLHAV	GVSYWKASEG	AEYDDQTSQR	EKEDDKVFPF
301	GSHTYVWQVL	KENGPMSADP	LCLTYSYLSH	VDLVKDLNSG	LIGALLVCRE
351	GSLAKEKTQT	LHKFILLFAV	FDEGKSWHSE	TKNSLMQDRD	AASARAWPKM
401	HTVNGYVNRS	LPGLIGCHRK	SVYWHVIGMG	TTPEVHSIFL	EGHTFLVRNH
451	RQASLEISPI	TFLTAQTLML	DLGQFLLFCH	ISSHQHDGME	AYVKVDSCPE
501	EPQLRMKNNE	EAEDYDDDLT	DSEMDVVRFD	DDNSPSFIQI	RSVAKKHPKT
551	WVHYIAAEEE	DWDYAPLVLA	PDDRSYKSQY	LNNGPQRIGR	KYKKVRFMAY
601	TDETFKTREA	IQHESGILGP	LLYGEVGDTL	LIIFKNQASR	PYNIYPHGIT
651	DVRPLYSRRL	PKGVKHLKDF	PILPGEIFKY	KWTVTVEDGP	TKSDPRCLTR
701	YYSSFVNMER	DLASGLIGPL	LICYKESVDQ	RGNQIMSDKR	NVILFSVFDE
751	NRSWYLTEI	QRFLPNPAGV	QLEDPEFQAS	NIMHSINGYV	FDSLQLSVCL
801	HEVAYWYILS	IGAQTDFLSV	FFSGYTFKHK	MVYEDTLTLF	PFSGETVFMMS
851	MENPGLWILG	CHNSDFRNRG	MTALLKVSSC	DKNTGDYDYE	SYEDISAYLL
901	SKNNAIEPRS	FSQNGAPGTS	ESATPESGPG	SEPATSGSET	PGTSESATPE
951	SGPGSEPATS	GSETPGTSES	ATPESGPGTS	TEPSEGSAPG	SPAGSPTSTE
[0681] 1001	EGTSESATPE	SGPGSEPATS	GSETPGTSES	ATPESGPGSP	AGSPTSTEEG
1051	SPAGSPTSTE	EGTSTEPSEG	SAPGTSESAT	PESGPGTSES	ATPESGPGTS
1101	ESATPESGPG	SEPATSGSET	PGSEPATSGS	ETPGSPAGSP	TSTEEGTSTE
1151	PSEGSAPGTS	TEPSEGSAPG	SEPATSGSET	PGTSESATPE	SGPGTSTEPS
1201	EGSAPASSPP	VLKRHQAEIT	RTTLQSDQEE	IDYDDTISVE	MKKEDFDIYD
1251	EDENQSPRSF	QKKTRHYFIA	AVERLWDYGM	SSSPHVLNRN	AQSGSVPOFK
1301	KVVFQEFDTG	SFTQPLYRGE	LNEHLGLLGP	YIRAEVEDNI	MVTFRNQASR
1351	PYSFYSSLIS	YEEDQRQGAE	PRKNFVKPNE	TKTYFWKVQH	HMAPTKDEFD
1401	CKAWAYFSDV	DLEKDVHSGL	IGPLLVCHTN	TLNPAHGRQV	TVQEFALFFT
1451	IFDETKSWYF	TENMERN CRA	PCNIQMEDPT	FKENYRFHAI	NGYIMDTLPG
1501	LVMAQDQRIR	WYLLSMGSNE	NIHSIHFSGH	VFTVRKKEEY	KMALYNLYPG
1551	VFETVEMLPS	KAGIWRVECL	IGEHLHAGMS	TLFLVYSNKC	QTPLGMASGH
1601	IRDFQITASG	QYGQWAPKLA	RLHYSGSINA	WSTKEPFSWI	KVDLLAPMI I
1651	HGIKTQGARQ	KFSSLYISQF	IIMYSLDGKK	WQTYRGNSTG	TLMVFFGNVD
1701	SSGIKHNIFN	PPIIARYIRL	HPTHYSIRST	LRMELMGCDL	NSCSMPLGME
1751	SKAISDAQIT	ASSYFTNMFA	TWSPSKARLH	LQGRSNAWRP	QVNNPKEWLQ
1801	VDFQKTMKVT	GVTTQGVKSL	LTSMYVKEFL	ISSSQDGHQW	TLFFQNGKVK
1851	VFQGNQDSFT	PVVNSLDPPL	LTRYLRIHPQ	SWVHQIALRM	EVLGCEAQDL
1901	YDKTHTCPPC	PAPELLGGPS	VFLFPPKPKD	TLMISRTPEV	TCVVVDVSHE
1951	DPEVKFNWYV	DGVEVHNAKT	KPREEQYNST	YRVVSVLTVL	HQDWLNGKEY
2001	KCKVSNKALP	APIEKTISKA	KGQPREPQVY	TLPPSRDEL T	KNQVSLTCLV
2051	KGFYPSDIAV	EWESNGQPEN	NYKTTPPVLD	SDGSFFLYSK	LTVDKSRWQQ
2101	GNVFSCSVMH	EALHNHYTQK	SLSLSPGK*		

[0682] pSYN FVIII 266蛋白质序列(具有在氨基酸18处的42 AE-XTEN和在B结构域中的288 AE XTEN的FVIII Fc) (SEQ ID NO:80)

1	MQIELSTCFF	LCLLRFCFSA	TRRYYLGAWE	LSWDYMQGAP	GSPAGSPTST
51	EEGTSESATP	ESGPGSEPAT	SGSETPASSS	DLGELPVDAR	FPPRVPKSFP
101	FNTSVVYKKT	LFVEFTDHLF	NIAKPRPPWM	GLLGPTIQAE	VYDTVVITLK
151	NMASHPVSLH	AVGVSYWKAS	EGAEDDQTS	QREKEDDKVF	PGGSHTYVWQ
201	VLKENGPMAS	DPLCLTYSYL	SHVDLVKDLN	SGLIGALLVC	REGSLAKEKT
251	QTLHKFILLF	AVFDEGKSWH	SETKNSLMQD	RDAASARAWP	KMHTVNGYVN
301	RSLPGLIGCH	RKSVYWHVIG	MGTTPPEVHSI	FLEGHTFLVR	NHRQASLEIS
351	PITFLTAQTL	LMDLGQFLLF	CHISSHQHDG	MEAYVKVDSC	PEEPQLRMKN
401	NEEAEDYDDD	LTDSEMDVVR	FDDDNSPSFI	QIRSVAKKHP	KTWVHYIAAE
451	EEDWDYAPLV	LAPDDRSYKS	QYLNNGPQRI	GRKYKKVRFM	AYTDETFKTR
501	EAIQHESGIL	GPLLYGEVGD	TLLIIFKNQA	SRPYNIPYHG	ITDVRPLYSR
551	RLPKGVKHLK	DFPILPGEIF	KYKWTVTVED	GPTKSDPRCL	TRYSSSFVNM
601	ERDLASGLIG	PLLCYKESV	DQRGNQIMSD	KRNVILFSVF	DENRSWYLTE
651	NIQRFLPNPA	GVQLEDPEFQ	ASNIMHSING	YVFDSLQLSV	CLHEVAYWYI
701	LSIGAQTDFL	SVFFSGYTFK	HKMVYEDTLT	LFPFSGETVF	MSMENPGLWI
751	LGCHNSDFRN	RGMTALLKVS	SCDKNTGDYD	EDSYEDISAY	LLSKNNAIEP
801	RSFSQNGAPG	TSESATPESG	PGSEPATSGS	ETPGTSESAT	PESGPGSEPA
851	TSGSETPGTS	ESATPESGPG	TSTEPSEGSA	PGSPAGSPTS	TEEGTSESAT
901	PESGPGSEPA	TSGSETPGTS	ESATPESGPG	SPAGSPTSTE	EGSPAGSPTS
951	TEEGTSTEPS	EGSAPGTSES	ATPESGPGTS	ESATPESGPG	TSESATPESG
[0683] 1001	PGSEPATSGS	ETPGSEPAT	GSETPGSPAG	SPTSTEEGTS	TEPSEGSAPG
1051	TSTEPSEGSA	PGSEPATSGS	ETPGTSESAT	PESGPGTSTE	PSEGSAPASS
1101	PPVLKRHQAE	ITRRTLQSDQ	EEIDYDDTIS	VEMKKEDFDI	YDEDENQSPR
1151	SFQKKTRHYF	IAAVERLWDY	GMSSSPHVLR	NRAQSGSVPQ	FKKVVFQEFT
1201	DGSFTQPLYR	GELNEHLGLL	GPYIRAEVED	NIMVTFRNQA	SRPYSFYSSL
1251	ISYEEDQRQG	AEPRKNFVKP	NETKTYFWKV	QHHMPTKDE	FDCKAWAYFS
1301	DVDLEKDVHS	GLIGPLLCH	TNTLNPAHGR	QVTVQEFALF	FTIFDETKSW
1351	YFTENMERN	RAPCNIQMED	PTFKENYRFH	AINGYIMDTL	PGLVMAQDQR
1401	IRWYLLSMGS	NENIHSIHFS	GHVFTVRKKE	EYKMALYNLY	PGVFETVEML
1451	PSKAGIWRVE	CLIGEHLHAG	MSTLFLVYSN	KCQTPMGAS	GHIRDFQITA
1501	SGQYGQWAPK	LARLHYSGSI	NAWSTKEPFS	WIKVDLLAPM	IIHGKIQGA
1551	RQKFSSLYIS	QFIIMYSLDG	KKWQTYRGN	TGTLMVFFGN	VDSSGIKHNI
1601	FNPPIIARYI	RLHPHYSIR	STLRMELMGC	DLNSCSMPLG	MESKAISDAQ
1651	ITASSYFTNM	FATWSPSKAR	LHLQGRSNAW	RPQVNNPKEW	LQVDFQKTMK
1701	VTGVTTQGVK	SLTSMYVKE	FLISSQDGH	QWTLFFQNGK	VKVFQGNQDS
1751	FTPVVNSLDP	PLLTRYLRH	PQSWVHQIAL	RMEVLGCEAQ	DLYDKTHTCP
1801	PCPAPELLGG	PSVFLFPPKP	KDTLMISRTP	EVTVCVVVDVS	HEDPEVKFNW
1851	YVDGVEVHNA	KTKPREEQYN	STYRVVSVLT	VLHQDWLNGK	EYKCKVSNKA
1901	LPAPIEKTIS	KAKGQPREPQ	VYTLPPSRDE	LTKNQVSLTC	LVKGFYPSDI
1951	AVEWESNGQP	ENNYKTTPPV	LDSDGSFFLY	SKLTVDKSRW	QQGNVFCSCV
2001	MHEALHNHYT	QKSLSLSPGK	*		

[0684] pSYN FVIII 267蛋白质序列(具有在氨基酸18处的72 AE-XTEN和在B结构域中的288 AE XTEN的FVIII Fc) (SEQ ID NO:81)



1	MQIELSTCFF	LCLLRFCFSA	TRRYYLGAWE	LSWDYMQGAP	TSESATPESG
51	PGSEPATSGS	ETPGTSESAT	PESGPGSEPA	TSGSETPGTS	ESATPESGPG
101	TSTEPSEGSA	PGASSDLGE	LPVDARFPPR	VPKSFPPNTS	VVYKKTLFVE
151	FTDHLFNIK	PRPPWMGLLG	PTIQAEVYDT	VVITLKNMAS	HPVSLHAVGV
201	SYWKASEGAE	YDDQTSQREK	EDDKVFPGGS	HTYVWQVLKE	NGPMASDPLC
251	LTYSYLSHVD	LVKDLNSGLI	GALLVCREGS	LAKEKTQTLH	KFILLFAVFD
301	EGKSWHSETK	NSLMQDRDAA	SARAWPKMHT	VNGYVNRSLP	GLIGCHRKSV
351	YWHVIGMGTT	PEVHSIFLEG	HTFLVRNHRQ	ASLEISPITF	LTAQTLLMDL
401	GQFLLFCHIS	SHQHGDMEAY	VKVDSCPEEP	QLRMKNNEEA	EDYDDDLTDS
451	EMDVVRFDDD	NSPSFIQIRS	VAKKHPKTWV	HYIAAEEEDW	DYAPLVLAPD
501	DRSYKSQYLN	NGPQRIGRKY	KKVRFMAYTD	ETFKTREAIO	HESGILGPLL
551	YGEVGDTHLI	IFKNQASRPY	NIYPHGITDV	RPLYSRRLPK	GVKHLKDFPI
601	LPGEIFKYKW	TVTVEDGPTK	SDPRCLTRYI	SSFVNMERDL	ASGLIGPLLI
651	CYKESVDQRG	NQIMSDKRN	ILFSVFDENR	SWYLTENIQR	FLPNPAGVQL
701	EDPEFQASNI	MHSINGYVFD	SLQLSVCLHE	VAYWYILSIG	AQTDFLSVFF
751	SGYTFKHKMV	YEDTLTLFPF	SGETVFMMSME	NPGLWILGCH	NSDFRNRGMT
801	ALLKVSSCDK	NTGDYEDSY	EDISAYLLSK	NNAIEPRSF	QNGAPGTSES
851	ATPESGPGSE	PATSGSETPG	TSESATPESG	PGSEPATSGS	ETPGTSESAT
901	PESGPGTSTE	PSEGSAPGSP	AGSPTSTEEG	TSESATPESG	PGSEPATSGS
951	ETPGTSESAT	PESGPGSPAG	SPTSTEEGSP	AGSPTSTEEG	TSTEPSEGSA
1001	PGTSESATPE	SGPGTSESAT	PESGPGTSES	ATPESGPGSE	PATSGSETPG
[0685] 1051	SEPATSGSET	PGSPAGSPTS	TEEGTSTEPS	EGSAPGTSTE	PSEGSAPGSE
1101	PATSGSETPG	TSESATPESG	PGTSTEPSEG	SAPASSPPVL	KRHQAEITRT
1151	TLQSDQEEID	YDDTISVEMK	KEDFDIYDED	ENQSPRSFQK	KTRHYFIAAV
1201	ERLWDYGMSS	SPHVLNRNAQ	SGSVPQFKKV	VFQEFTDGSF	TQPLYRGELN
1251	EHLGLLGPYI	RAEVEDNIMV	TFRNQASRPY	SFYSSLISYE	EDQRQGAEP
1301	KNFVKPNETK	TYFWKVQHMM	APTKDEFDCK	AWAYFSDVDL	EKDVHSGSLIG
1351	PLLVCHTNTL	NPAHGRQVTV	QEFALFFTI	DETKSWYFTE	NMERNCRAPC
1401	NIQMEDPTFK	ENYRFHAING	YIMDTLPLGLV	MAQDQIRIRWY	LLSMGSNENI
1451	HSIHFSGHVF	TVRKKEEYKM	ALYNLYPGVF	ETVEMLPKA	GIWRVECLIG
1501	EHLHAGMSTL	FLVYSNKCQT	PLGMASGHIR	DFQITASGQY	GQWAPKLARL
1551	HYSGSINAWS	TKEPFSWIKV	DLAPMIHIG	IKTQGARQKF	SSLYISQFII
1601	MYSLDGKKWQ	TYRGNSTGTL	MVFFGNVDSS	GIKHNI FNPP	IIARYIRLHP
1651	THYSIRSTLR	MELMGCDLNS	CSMPLGMESK	AISDAQITAS	SYFTNM FATW
1701	SPSKARLHLQ	GRSNAWRPQV	NNPKEWLQVD	FQKTMKVTGV	TTQGVKSLLT
1751	SMYVKEFLIS	SSQDGHQWTL	FFQNGKV KVF	QGNQDSFTPV	VNSLDPPLLT
1801	RYLRIHPQSW	VHQIALRMEV	LGCEAQDLYD	KTHTCPPCPA	PELLGGPSVF
1851	LFPPKPKDTL	MISRTPEVTC	VVDVSHEDP	EVKFNWYVDG	VEVHNAKTKP
1901	REEQYNSTYR	VVSVLTVLHQ	DWLNKEYKC	KVSNKALPAP	IEKTISKAKG
1951	QPREPQVYTL	PPSRDELTKN	QVSLTCLVKG	FYPSDIAVEW	ESNGQPENNY
2001	KTPPVLDSD	GSFFLYSKLT	VDKSRWQQGN	VFSCSVMHEA	LHNHYTQKSL
2051	SLSPGK*				

[0686] pSYN FVIII 268蛋白质序列(在氨基酸18处具有144 AE-XTEN的FVIII Fc) (SEQ ID NO:82)

```

1    MQIELSTCFF LCLLRFCFSA TRRYYLGA VE LSWDYMQGAP TSESATPESG
51   PGSEPATSGS ETPGTSESAT PESGPGSEPA TSGSETPGTS ESATPESGPG
101  TSTEPSEGS SA PGSPAGSPTS TEEGTSESAT PESGPGSEPA TSGSETPGTS
151  ESATPESGPG SPAGSPTSTE EGSPAGSPTS TEEGASSSDL GELPVDARFP
201  PRVPKSF PFN TSVVYKKT LF VEFTDHLFNI AKPRPPWMGL LGPTIQAEVY
251  DTVVITLKNM ASHPVSLHAV GVSYWKASEG AEYDDQTSQR EKEDDKVFPG
301  GSHTYVWQVL KENGPMSADP LCLTYSYLSH VDLVKDLNSG LIGALLVCRE
351  GSLAKEKTQT LHKFILLFAV FDEGKSWHSE TKNSLMQDRD AASARAWPKM
401  HTVNGYVNRS LPGLIGCHRK SVYWHVIGMG TTPEVHSIFL EGHTFLVRNH
451  RQASLEISPI TFLTAQTLLM DLGQFLLFCH ISSHQHDGME AYVKVDSCPE
501  EPQLRMKNNE EAEDYDDDLT DSEMDVVRFD DDNSPSFIQI RSVAKKHPKT
551  WVHYIAAEEE DWDYAPLVLA PDDRSYKSQY LNNGPQRIGR KYKKVRFMAY
601  TDETFKTREA IQHESGILGP LLYGEVGD TL LIIFKNQASR PYNIIYPHGIT
651  DVRPLYSRRL PKGVKHLKDF PILPGEIFKY KWTVTVEDGP TKSDPRCLTR
701  YYSSFVNMER DLASGLIGPL LICYKESVDQ RGNQIMSDKR NVILFSVFDE
751  NRSWYL TENI QRFLPNPAGV QLEDPEFQAS NIMHSINGYV FDSLQLSVCL
801  HEVAYWYILS IGAQTDFLSV FFSGYTFKHK MVEYEDTLTF PFSGETVFMMS
851  MENPGLWILG CHNSDFRNRG MTALLKVSSC DKNTGDY YED SYEDISAYLL
[0687] 901  SKNNAIEPRS FSQNPPVLKR HQAEITR TTL QSDQEEIDYD DTISVEMKKE
951  DFDIYDEDEN QSPRSFQKKT RHYFIAAVER LW DYGMSSSP HVLRNRAQSG
1001 SVPQFKK VVF QEFTDGSFTQ PLYRGELNEH LGLLGPYIRA EVEDNIMVTF
1051 RNQASRPYSF YSSLISYEED QRQGAEP RKN FVKPNETKTY FWKVQH HMAP
1101 TKDEFDCKAW AYFSDVDLEK DVHSGLIGPL LVCHTNTLNP AHGRQVTVQE
1151 FALFFTIFDE TKS WYFTENM ERNCRAPCNI QMEDPTFKEN YRFHAINGYI
1201 MDTLPGLVMA QDQIRWYLL SMGSNENIHS IHFSGHVFTV RKKEEYKMAL
1251 YNLYPGVFET VEMLPSKAGI WRVECLIGEH LHAGMSTLFL VYSNKCQTPL
1301 GMASGHIRDF QITASGQY GQ WAPKLARLHY SGSINAWSTK EPPSWIKVDL
1351 LAPMI IHG I K TQGARQKFSS LYISQFIIMY SLDGKKWQTY RGNSTGTLMV
1401 FFGNVDSSGI KHNIFNPPII ARYIRLHPTH YSIRSTLRME LMGCDLNSCS
1451 MPLGMESKAI SDAQITASSY FTNMFATWSP SKARLHLQGR SNAWRPQVNN
1501 PKEWLQVDFQ KTMKVTG VTT QGVKSLLTSM YVKEFLISS QDGHQWTLFF
1551 QNGKVKVFQ G NQDSFTP VVN SLDPPLLTRY LRIHPQSWVH QIALRMEVLG
1601 CEAQDLYDKT HTPPCPAPE LLGGPSVFLF PPKPKDTLMI SRTPEVTCVV
1651 VDVSHEDPEV KFNWYVDGVE VHNATKPRE EQYNSTYRVV SVLTVLHQDW
1701 LNGKEYKCKV SNKALPAPIE KTISKAKGQP REPQVYTLPP SRDELTKNQV
1751 SLTCLVKGFY PSDIAVEWES NGQPENNYKT TPPVLDS DGS FFLYSKLTVD
1801 KSRWQQGNVF SCSVMHEALH NHYTQKSLSL SPGK*

```

[0688] pSYN FVIII 269蛋白质序列(在氨基酸18处具有72 AE-XTEN的FVIII Fc) (SEQ ID NO:83)

```

1    MQIELSTCFF LCLLRFCFSA TRRYYLGA VE LSWDYMQGAP TSESATPESG
51   PGSEPATSGS ETPGTSESAT PESGPGSEPA TSGSETPGTS ESATPESGPG
[0689] 101  TSTEPSEGS SA PGASSSDLGE LPVDARFPPR VPKSF PFNTS VVYKKT LFVE
151  FTDHLFNI AK PRPPWMGL LG PTIQAEVYDT VVITLKNMAS HPVSLHAVGV
201  SYWKASEGAE YDDQTSQREK EDDKVFPGGS HTYVWQVLKE NGPMASDPLC
251  LTYSYLSHVD LVKDLNSGLI GALLVCREGS LAKEKTQTLH KFILLFAVFD

```

301 EGKSWHSETK NSLMQDRDAA SARAWPKMHT VNGYVNRSLP GLIGCHRKSV  
 351 YWHVIGMGTT PEVHSIFLEG HTFLVRNHRQ ASLEISPITF LTAQTLMLDL  
 401 GQFLLFCHIS SHQHDGMEAY VKVDSCPEEP QLRMKNNEEA EDYDDDLTDS  
 451 EMDVVRFDLDD NSPSFIQIRS VAKKHPKTWV HYIAAEEEDW DYAPLVLAPD  
 501 DRSYKSQYLN NGPQRIGRKY KKVRFMAYTD ETFKTREAIQ HESGILGPLL  
 551 YGEVGDTHLI IFKNQASRPY NIYPHGITDV RPLYSRRLPK GVKHLKDFPI  
 601 LPGEIFKYKW TVTVEDGPTK SDPRCLTRY Y SSFVNMERDL ASGLIGPLLI  
 651 CYKESVDQRG NQIMSDKRN V ILFSVFDENR SWYL TENIQR FL PNPAGVQL  
 701 EDPEFQASNI MHSINGYVFD SLQLSVCLHE VAYWYILSIG AQTDFLSVFF  
 751 SGYTFKHK MV YEDTLTLFPF SGETV FMSME NPGLWILGCH NSDFRNRGMT  
 801 ALLKVSSCDK NTGDYYEDSY EDISAYLLSK NNAIEPRSF S QNPPVLKRHQ  
 851 AEITRRTLQS DQEEIDYDDT ISVEMKKEDF DIYDEDENQS PRSFQKKTRH  
 901 YFIAAVERLW DYGMSSSPHV LRNRAQSGSV PQFKKVV FQE FTDGSFTQPL  
 951 YRGELNEHLG LLGPYIRAEV EDNIMVTFRN QASRPYSFYS SLISYEEDQR  
 [0690] 1001 QGAEPRKNFV KPNETKTYFW KVQHHM APTK DEFDC KAWAY FSDVDLEKDV  
 1051 HSLIGPLLV CHTNTLNPAH GRQVTVQEFA LFFTIFDETK SWYFTENMER  
 1101 NCRAPCNIQM EDPTFKENYR FHAINGYIMD TLPGLVMAQD QRIRWYLLSM  
 1151 GSNENIHSIH FSGHVFTVRK KEEYKMALYN LYPGVFETVE MLPSKAGIWR  
 1201 VECLIGEHLH AGMSTLFLVY SNKCQTPLGM ASGHIRDFQI TASGQYQWA  
 1251 PKLARLHYSG SINAWSTKEP FSWIKVDLLA PMIIHG IKTQ GARQKFSSLY  
 1301 ISQFIIMYSL DGKKWQTYRG NSTGTL MVFF GNV DSSGIKH NIFNPPIAR  
 1351 YIRLHPHYS IRSTLRMELM GCDLN SC SMP LGMESKAISD AQITASSYFT  
 1401 NMFATWSPSK ARLHLQGRSN AWRPQVNNPK EWLQVDFQKT MKVTGVTTQG  
 1451 VKSLTSMYV KEFLISSSQD GHQWTLFFQN GKVKVFQGNQ DSFTPVVNSL  
 1501 DPPLLTRYLR IHPQSWVHQI ALRMEVLGCE AQDLYDKTHT CPPCPAPELL  
 1551 GGPSVFLFPP KPKDTLMISR TPEVTCVVVD VSHEDPEVKF N WYVDGVEVH  
 1601 NAKTKPREEQ YNSTYRVVSV LTVLHQDWLN GKEYKCKVSN KALPAPIEKT  
 1651 ISKAKQPRE PQVYTLPPSR DELTKNQVSL TCLVKGFYPS DIAVEWESNG  
 1701 QPENNYKTP PVLDSDGSFF LYSKLTVDKS RWQQGNV FSC SVMHEALHNH  
 1751 YTQKSLSLSP GK\*

[0691] pSYNFVIII 271蛋白质序列(在氨基酸18处具有42 AE-XTEN的FVIII Fc) (SEQ ID NO:84)

```

1      MQIELSTCFF LCLLRFCFSA TRRYYLGAWE LSWDYMQGAP GSPAGSPTST
51     EEGTSESATP ESGPGSEPAT SGSETPASSS DLGELPVDAR FPPRPVKSFP
101    FNTSVVYKKT LFVEFTDHLF NIAKPRPPWM GLLGPTIQAE VYDTVVITLK
151    NMASHPVSLH AVGVSYWKAS EGAEYDDQTS QREKEDDKVF PGGSHTYVWQ
201    VLKENGPMAS DPLCLTYSYL SHVDLVKDLN SGLIGALLVC REGSLAKEKT
251    QTLHKFILLF AVFDEGKSWH SETKNLSMQD RDAASARAWP KMHTVNGYVN
301    RSLPGLIGCH RKSVMYWHVIG MGTTPQVHSI FLEGHTFLVR NHRQASLEIS
351    PITFLTAQTL LMDLGQFLLF CHISSHQHDG MEAYVKVDSC PEEPQLRMKN
401    NEEAEDYDDD LTDSEMDVVR FDDDNSPSFI QIRSVAKKHP KTWVHYIAAE
451    EEDWDYAPLV LAPDDRSYKS QYLNNGPQRI GRKYKKVRFM AYTDETFKTR
501    EAIQHESGIL GPLLYGEVGD TLLIIFKNQA SRPYNIPYHG ITDVRPLYSR
551    RLPKGVKHLK DFPILPGEIF KYKWTVTVED GPTKSDPRCL TRYSSSFVNM
601    ERDLASGLIG PLLICYKESV DQRGNQIMSD KRNVIILFSVF DENRSWYLTE
651    NIQRFLPNPA GVQLEDPEFQ ASNIMHSING YVFDSLQLSV CLHEVAYWYI
701    LSIGAQTDFL SVFFSGYTFK HKMVEYEDTLT LFPFSGETVF MSMENPGLWI
751    LGCHNSDFRN RGMTALLKVS SCDKNTGDYD EDSYEDISAY LLSKNNAIEP
801    RSFSQNPPVL KRHQAEITRT TLQSDQEEID YDDTISVEMK KEDFDIYDED
[0692] 851    ENQSPRSFQK KTRHYFIAAV ERLWDYGMSS SPHVLRNRAQ SGSVPQFKKV
901    VFQEFDTGSF TQPLYRGELN EHLGLLGPYI RAEVEDNIMV TFRNQASRPY
951    SFYSSLISYE EDQRQGAEPK KNFVKPNETK TYFWKVQHHM APTKDEFDCK
1001   AWAYFSDVDL EKDVHSGLIG PLLVCHTNTL NPAHGRQVTV QEFALFFTIF
1051   DETKSWYFTE NMERNCRAPC NIQMEDPTFK ENYRFHAING YIMDTLPLGLV
1101   MAQDQIRIRWY LLSMGSNENI HSIHFSGHVF TVRKKEEYKM ALYNLYPGVF
1151   ETVEMLPSKA GIWRVECLIG EHLHAGMSTL FLVYSNKCQT PLGMASGHIR

1201   DFQITASGQY GQWAPKLARL HYSGSINAW S TKEPFSWIKV DLLAPMI IHG
1251   IKTOGARQKF SSLYISQFII MYSLDGKKWQ TYRGNSTGTL MVFFGNVDSS
1301   GIKHNIFFNP IIARYIRLHP THYSIRSTLR MELMGCDLNS CSMPPLGMESK
1351   AISDAQITAS SYFTNMFATW SPSKARLHLQ GRSNAWRPQV NNPKEWLQVD
1401   FQKTMKVTGV TTQGVKSLLT SMYVKEFLIS SSQDGHQWTL FFQNGKVKVF
1451   QGNQDSFTPV VNSLDPPLLT RYLRIHPQSW VHQIALRMEV LGCEAQDLYD
1501   KTHTCPPCPA PELLGGPSVF LFPPKPKDTL MISRTPEVTC VVVDVSHEDP
1551   EVKFNWYVDG VEVHNAKTKP REEQYNSTYR VVSVLTVLHQ DWLNGKEYKC
1601   KVSNKALPAP IEKTISKAKG QPREPQVYTL PPSRDELTKN QVSLTCLVKG
1651   FYPSDIAVEW ESNGQPENNY KTTPPVLDSG GSFFLYSKLT VDKSRWQOQN
1701   VFSCSVMHEA LHNHYTQKSL SLSPGK*

```

[0693] pSYN FVIII蛋白质序列272(具有在氨基酸18处的144 AE XTEN和在B结构域中的244 AE XTEN的FVIII-无Fc) (SEQ ID NO:85)

```

1      MQIELSTCFE LCLLRFCFSA TRRYYLGAWE LSWDYMQGAP TSESATPESG
51     PGSEPATSGS ETPGTSESAT PESGPGSEPA TSGSETPGTS ESATPESGPG
101    TSTEPSEGS SA PGSPAGSPTS TEEGTSESAT PESGPGSEPA TSGSETPGTS
151    ESATPESGPG SPAGSPTSTE EGSPAGSPTS TEEGASSSDL GELPVDARFP
201    PRVPKSFPFN TSVVYKKTLE VEFTDHLFNI AKPRPPWMGL LGPTIQAEVY
251    DTVVITLKNM ASHPVSLHAV GVSYWKASEG AEYDDQTSQR EKEDDKVFPF
301    GSHTYVWQVL KENGPMSADP LCLTYSYLSH VDLVKDLNSG LIGALLVCRE
351    GSLAKEKTQT LHKFILLFAV FDEGKSWHSE TKNSLMQDRD AASARAWPKM
401    HTVNGYVNRS LPGLIGCHRK SVYWHVIGMG TPPEVHSIFL EGHTFLVRNH
451    RQASLEISPI TFLTAQTLLM DLGQFLLFCH ISSHQHDGME AYVKVDSCPE
501    EPQLRMKNNE EAEDYDDDLT DSEMDVVRFD DDNSPSFIQI RSVAKKHPKT
551    WVHYIAAEEE DWDYAPLVLA PDDRSYKSQY LNNGPQRIGR KYKKVRFMAY
601    TDETFKTREA IQHESGILGP LLYGEVGDLE LIIFKNQASR PYNIIYPHGIT
651    DVRPLYSRRL PKGVKHLKDF PILPGEIFKY KWTVTVEDGP TKSDPRCLTR
701    YYSSFVNMER DLASGLIGPL LICYKESVDQ RGNQIMSDKR NVILFSVFDE
751    NRSWYLTEI QRFLPNPAGV QLEDPEFQAS NIMHSINGYV FDSLQLSVCL
801    HEVAYWYILS IGAQTDFLSV FFSGYTFKHK MVEYEDTLTF PFSGETVFMF
851    MENPGLWILG CHNSDFRNRG MTALLKVSSC DKNTGDYED SYEDISAYLL
901    SKNNAIEPRS FSQNGAPGTS ESATPESGPG SEPATSGSET PGTSESATPE
[0694] 951    SGPGESEPAT S GSETPGTSES ATPESGPGTS TEPSEGSAPG SPAGSPTSTE
1001   EGTSESATPE SGPGESEPAT S GSETPGTSES ATPESGPGSP AGSPTSTEEG
1051   SPAGSPTSTE EGTSTEPSEG SAPGTSESAT PESGPGTSES ATPESGPGTS
1101   ESATPESGPG SEPATSGSET PGSEPATSGS ETPGSPAGSP TSTEEGTSTE
1151   PSEGSAPGTS TEPSEGSAPG SEPATSGSET PGTSESATPE SGPSTSTEPS
1201   EGSAPASSPP VLKRHQAEIT RTTLQSDQEE IDYDDTISVE MKKEDFDIYD
1251   EDENQSPRSF QKKTRHYFIA AVERLWDYGM SSSPHVLRNR AQSGSVPOFK
1301   KVVVFQFTDG SFTQPLYRGE LNEHLGLLGP YIRAEVEDNI MVTFRNQASR
1351   PYSFYSSLIS YEEDQRQGAE PRKNFVKPNE TKTYFWKVQH HMAPTKDEFD
1401   CKAWAYFSDV DLEKDVHSGE IGPLLVCHTN TLNPAHGRQV TVQEFALFFT
1451   IFDETKSWYF TENMERNCR A PCNIQMEDPT FKENYRFHAI NGYIMDTLPG
1501   LVMAQDQRIR WYLLSMGSNE NIHSIHFSGH VFTVRKKEEY KMALYNLYPG
1551   VFETVEMLPS KAGIWRVECL IGEHLHAGMS TLFLVYSNKC QTPLGMASGH
1601   IRDFQITASG QYGQWAPKLA RLHYSGSINA WSTKEPFSWI KVDLLAPMII
1651   HGIKTQGARQ KFSSLYISQF IIMYSLDGKK WQTYRGNSTG TLMVFFGNVD
1701   SSGIKHNIFN PPIIARYIRL HPTHYSIRST LRMELMGCDL NSCSMPLGME
1751   SKAISDAQIT ASSYFTNMFA TWSPSKARLH LQGRSNAWRP QVNNPKEWLQ
1801   VDFQKTMKVT GVTTQGVKSL LTSMYVKEFL ISSSQDGHQW TLFFQNGKVK
1851   VFQGNQDSFT PVVNSLDPL LTRYLRIHPQ SWVHQIALRM EVLGCEAQDL
1901   Y*

```

[0695] pSYN-FVIII-161蛋白质序列(SEQ ID NO:69) (FVIII序列氨基酸位置1-1457;加下划线区域表示Fc区;弯曲下划线表示在第一Fc与VWF片段之间的可裂解接头;加双下划线区域表示VWF片段;粗体区域表示在VWF片段与Fc之间的可裂解接头)。

1 MQIELSTCFF LCLLRFCFSA TRRYYLGAWE LSWDYMQSDL GELPVDARFP  
51 PRVPKSFPFN TSVVYKKTFL VEFTDHLFNI AKPRPPWMGL LGPTIQAEVY  
101 DTVVITLKNM ASHPVSLHAV GVSYWKASEG AEYDDQTSQR EKEDDKVFPG  
151 GSHTYVWQVL KENGPASDP LCLTYSYLSH VDLVKDLNSG LIGALLVCRE  
201 GSLAKEKTQT LHKFILLFAV FDEGKSWHSE TKNSLMQDRD AASARAWPKM  
251 HTVNGYVNRS LPGLIGCHRK SVYWHVIGMG TTPEVHSIFL EIGHTFLVRNH  
301 RQASLEISPI TFLTAQTLLM DLGQFLLFCH ISSHQHDGME AYVKVDSCPE  
351 EPQLRMKNNE EAEDYDDDLT DSEMDVVRFD DDNSPSFIQI RSVAKKHPKT  
401 WVHYIAAEEE DWDYAPLVLA PDDRSYKSQY LNNGPQRIGR KYKKVRFMAY  
451 TDETFKTREA IQHESGILGP LLYGEVGDTL LIIFKNQASR PYNIIYPHGIT  
501 DVRPLYSRRL PKGVKHLKDF PILPGEIFKY KWTVTVEDGP TKSDPRCLTR  
551 YYSSFVNMER DLASGLIGPL LICYKESVDQ RGNQIMSDKR NVILFSVFDE  
601 NRSWYL TENI QRFLPNPAGV QLEDPEFQAS NIMHSINGYV FDSLQLSVCL  
651 HEVAYWYILS IGAQTDFLSV FFSGYTFKHK MVEDTLTLF PFSGETVFMS  
701 MENPGLWILG CHNSDFRNRG MTALLKVSSC DKNTGDYED SYEDISAYLL  
751 SKNNAIEPRS FSQNPVVKR HQREITR TTL QSDQEEIDYD DTISVEMKKE  
801 DFDIYDEDEN QSPRSFQKKT RHYFIAAVER LWDYGMSSSP HVLRNRAQSG  
851 SVPQFKKVV FQEFTDGSFTQ PLYRGELNEH LGLLGPYIRA EVEDNIMVTF  
901 RNQASRPYSF YSSLISYEED QRQGAEPKRN FVKPNETKTY FWKVQHMAP  
951 TKDEFDCKAW AYFSDVDLEK DVHSGLIGPL LVCHTNTLNP AHGRQVTVQE  
1001 FALFFTIFDE TKSIFYTENM ERNCRAPCNI QMEDPTFKEN YRFHAINGYI  
1051 MDTLPGLVMA QDQIRWYLL SMGSNENIHS IHFSGHVFTV RKKEEYKMAL  
1101 YNLYPGVFET VEMLPSKAGI WRVECLIGEH LHAGMSTLFL VYSNKCQTPL  
1151 GMASGHIRDF QITASGQYGO WAPKLARLHY SGSINAWSTK EPFSWIKVDL  
1201 LAPMIIHGK TQGARQKFSS LYISQFIIMY SLDGKKWQTY RGNSTGTLMV  
[0696] 1251 FFGNVDSSGI KHNI FNPII ARYIRLHPTH YSIRSTLRME LMGCDLNSCS  
1301 MPLGMESKAI SDAQITASSY FTNMFATWSP SKARLHLQGR SNAWRPQVNN  
1351 PKEWLQVDFQ KTMKVTGVT QGVKSLTSM YVKEFLISS QDGHQWTLFF  
1401 QNGKVKVFQO NQDSFTPVVN SLDPPLLTRY LRIHPQSWVH QIALRMEVLG  
1451 CEAQDLYDKT HTCPCPAPE LLGGPSVFLF PPKPKDTLMI SRTPEVTCVV  
1501 VDVSHEDPEV KFNWYVDGVE VHNATKPRE EQYNSTYRVV SVLTVLHQDW  
1551 LNGKEYKCKV SNKALPAPIE KTISKAKQOP REPQVYTLPP SRDELTKNQV  
1601 SLTCLVKGFY PSDIAVEWES NGQPENNYKT TPPVLDS DGS FFLYKLTVD  
1651 KSRWQQGNVF SCSVMHEALH NHYTQKLSL SPGKRRRRSG GGGSGGGSG  
1701 GGGSGGGSG GGGSGGGSR KRRKRSLSR PPMVKLVCPA DNLRAEGLEC  
1751 TKTCQNYDLE CMSMGCVSGC LCPPGMVRHE NRCVALERCP CFHQKEYAP  
1801 GETVKIGCNT CVCRDRKWC TDHVCDATCS TIGMAHYLTF DGLKYLFPGE  
1851 COYVLVQDYC GSNPGTFRIL VGNKGC SHPS VKCKKRV TIL VEGGEI ELD  
1901 GEVNVKRP MK DETHFEVVE GRYIILLGK ALSVVDRHL SISVVLKQTY  
1951 QEKVCGLCGN FDGIQNDLT SSNLQVEEDP VDFGNSWKVS SQCADTRKVP  
2001 LDSSPATCHN NIMKQTMVDS SCRILTSDFV QDCNKLVDPE PYLDVCIYDT  
2051 CSCESIGDCA AFCDTIAAYA HVCAOHGKVV TWRTATLCPQ SCEERNLREN  
2101 GYEAEWRYS CAPACQVTCQ HPEPLACPVO CVEGCHAHCP PGKILDELLO  
2151 TCVDPEDCPV CEVAGRRFAS GKKVTLNPSD PEHCQICHCD VVNLTCEACQ  
2201 EPISGTSESA TPESGPGSEP ATSGSETPGT SESATPESGP GSEPATSGSE  
2251 TPGTSESATP ESGPGTSTEP SEGSAPGSPA GSPTSTEEGT SESATPESGP  
2301 GSEPATSGSE TPGTSESATP ESGPGSPAGS PTSTEEGSPA GSPTSTEEGT  
2351 STEPSEGSAP GTSESATPES GPGTSESATP ESGPGTSESA TPESGPGSEP  
2401 ATSGSETPGS EPATSGSETP GSPAGSPTST EEGTSTEPSE GSAPGTSTEP  
2451 SEGSAPGSEP ATSGSETPGT SESATPESGP GTSTEPSEGS APDSGGGGSG  
2501 GGGSGGGSG GGGSGGGSL VPRGSGDKT HTCPCPAPE LLGGPSVFLF

2551 PPKPKDTLMI SRTPEVTCVV VDVSHEDPEV KFNWYVDGVE VHNAKTKPRE  
 2601 EQYNSTYRVV SVLTVLHQDW LNGKEYKCKV SNKALPAPIE KTISKAKGQP  
 [0697] 2651 REPOVYTLPP SRDELTKNOV SLTCLVKGFY PSDIAVEWES NGQOPENNYKT  
 2701 TPPVLDSDGS FFLYSKLTVD KSRWQQGNVF SCSVMHEALH NHYTQKSLSL  
 2751 SPGK

[0698] pSYN-FVIII-170蛋白质序列(SEQ ID NO:71)

1 SLSCRPPMVK LVCPADNLRA EGLECTKTCQ NYDLECMSMG CVSGCLCPPG  
 51 MVRHENRCVA LERCPCFHQG KEYAPGETVK IGCNTCVCRD RKWNCTDHVC  
 101 DATCSTIGMA HYLTFDGLKY LFPGECQYVL VQDYCGSNPG TFRILVGNKG  
 151 CSHPSVKCKK RVTILVEGGE IELFDGEVNV KRPMKDETHF EVVESGRYII  
 201 LLLGKALSVV WDRHLSISVV LKQTYQEKV GLCGNFDGIQ NNDLTSSNLQ  
 251 VEEDPVDFGN SWKVSSQCAD TRKVPLDSSP ATCHNNIMKQ TMVDSSCRIL  
 301 TSDVFQDCNK LVDPEPYLDV CIYDTCSCES IGDCAAFCDT IAAYAHVCAQ  
 351 HGKVVWRTA TLCPOSCEER NLRENGYAE WRYNSCAPAC QVTCQHPEPL  
 401 ACPVQCVEGC HAHCPPGKIL DELLOTCVDP EDCPVCEVAG RRFASGKKVT  
 451 LNPSDPEHCQ ICHCDVNL TCEACQEPISG TSESATPESG PGSEPATSGS  
 501 ETPGTSESAT PESGPGSEPA TSGSETPGTS ESATPESGPG TSTEPSEGSA  
 551 PGSPAGSPTS TEEGTSESAT PESGPGSEPA TSGSETPGTS ESATPESGPG  
 601 SPAGSPTSTE EGSPAGSPTS TEEGTSTEPS EGSAPGTSES ATPESGPGTS  
 651 ESATPESGPG TSESATPESG PGSEPATSGS ETPGSEPATS GSETPGSPAG  
 701 SPTSTEEGTS TEPSEGSAPG TSTEPSEGSA PGSEPATSGS ETPGTSESAT  
 751 PESGPGTSTE PSEGSAPDSG GGGSGGGGSG GGGSGGGGSG GGGSLVPRGS  
 801 GGASATRRYY LGAVELSWDY MQSDLGELPV DARFPPRVPK SFPFNSTSVVY  
 851 KKTLFVEFTD HLFNIAKPRP PWMGLLGPTI QAEVYDTVVI TLKNMASHPV  
 901 SLHAVGVSYW KASEGAEYDD QTSQREKEDD KVFPGGSHY VWQVLKENG  
 951 MASDPLCLTY SYLSHVDLVK DLNSGLIGAL LVCREGSLAK EKTQTLHKFI  
 1001 LLFAVFDEGK SWHSETKNSL MQDRDAASAR AWPKMHTVNG YVNRSLPGLI  
 [0699] 1051 GCHRKSVYWH VIGMGTPEV HSIFLEGHTF LVRNHRQASL EISPITFLTA  
 1101 QTLMLDLGQF LLFCHISSHQ HDGMEAYVKV DSCPEEPQLR MKNNEEAEDY  
 1151 DDDLDTSEMD VVRFDDNSP SFIQIRSVAK KHPKTWVHYI AAEEDWDYA  
 1201 PLVLAPDDRS YKSQYLNNP QRIGRKYKKV RFMAYTDEF KTREAIQES  
 1251 GILGPLLYGE VGDTLIIIFK NQASRPYNIY PHGITDVRPL YSRRLPKGVK  
 1301 HLKDFPILPG EIFKYKWTVT VEDGPTKSDP RCLTRYSSV VNMERDLAS  
 1351 LIGPLLYCYK ESDVQRGNQI MSDKRNILF SVFDENRSWY LTNERDLFLP  
 1401 NPAGVQLEDP EFQASNIMHS INGYVFDLQ LSVCLHEVAY WYILSIGAQT  
 1451 DFLSVFFSGY TFKHKMVYED TLTLFPFSGE TVFMSMENPG LWILGCHNSD  
 1501 FRNRGMTALL KVSSCDKNTG DYYEDSYEDI SAYLLSKNNA IEPRSFQNP  
 1551 PVLKRHQREI TRTTLQSDQE EIDYDDTISV EMKKEDFDIY DEDENQSPRS  
 1601 FQKKTRHYFI AAVERLWDYG MSSSPHVLN RAQSGSVPQF KKVVQEFQTD  
 1651 GSFTQPLYRG ELNEHLGLL PYIRAEVEDN IMVTFRNQAS RPYSFYSSLI  
 1701 SYEEDQRQGA EPRKNFVKPN ETKTYFWKVQ HHMAPTKDEF DCKAWAYFSD  
 1751 VDLEKDVHSG LIGPLLVCHT NTLNPAHGRQ VTVQEFALFF TIFDETKSWY  
 1801 FTE NMERNCR APCNIQMEDP TFKENYRFHA INGYIMDTLP GLVMAQDQRI  
 1851 RWYLLSMGSN ENIHSIHFSG HVFTVRKKEE YKMALYNLYP GVFETVEMLP  
 1901 SKAGIWRVEC LIGELHAGM STLFLVYSNK CQTPLGMASG HIRDFQITAS  
 1951 GQYQWAPKL ARLHYSIN AWSTKEPFSW IKVDLLAPMI IHGKTQGAR  
 2001 QKFSSLYISQ FIIMYSLDGK KWQTYRGNST GTLMVFFGNV DSSGIKHNI  
 2051 NPPIIARYIR LHPTHYSIRS TLRMELMGCD LNCSMPLGM ESKAISDAQI  
 2101 TASSYFTNMF ATWSPSKARL HLQGRSNAWR PQVNNPKEWL QVDFQKTMKV  
 2151 TGVTTQGVKS LLTSMYVKEF LISSSQDGHQ WTLFFQNGKV KVFQGNQDSF  
 2201 TPVVNSLDPP LLTRYLRHQP QSVVHQIALR MEVLGCEAQD LY

[0700] pSYN FVIII 310核苷酸序列(编码除2个氨基酸残基之外,B结构域完全缺失,并且在aa 742之后插入288 AE-XTEN的FVIII)(SEQ ID NO:170)

1 ATGCAAATAG AGCTCTCCAC CTGCTTCTTT CTGTGCCTTT TGCGATTCTG  
 51 CTTTAGTGCC ACCAGAAGAT ACTACCTGGG TGCAGTGGAA CTGTCATGGG  
 101 ACTATATGCA AAGTGATCTC GGTGAGCTGC CTGTGGACGC AAGATTTCTT  
 151 CCTAGAGTGC CAAAATCTTT TCCATTCAAC ACCTCAGTCG TGTACAAAAA  
 201 GACTCTGTTT GTAGAATTCA CGGATCACCT TTTCAACATC GCTAAGCCAA  
 251 GGCCACCCTG GATGGGTCTG CTAGGTCCTA CCATCCAGGC TGAGGTTTAT  
 301 GATACAGTGG TCATTACACT TAAGAACATG GCTTCCCATC CTGTCACTCT  
 351 TCATGCTGTT GGTGTATCCT ACTGGAAAAGC TTCTGAGGGA GCTGAATATG  
 401 ATGATCAGAC CAGTCAAAGG GAGAAAAGAAG ATGATAAAGT CTTCCCTGGT  
 451 GGAAGCCATA CATATGTCTG GCAGGTCCTG AAAGAGAATG GTCCAATGGC  
 501 CTCTGACCCA CTGTGCCTTA CCTACTCATA TCTTTCATCAT GTGGACCTGG  
 551 TAAAAGACTT GAATTCAGGC CTCATTGGAG CCCTACTAGT ATGTAGAGAA  
 601 GGGAGTCTGG CCAAGGAAAA GACACAGACC TTGCACAAAT TTATACTACT  
 651 TTTTGCTGTA TTTGATGAAG GGAAAAGTTG GCACTCAGAA ACAAAGAACT  
 701 CCTTGATGCA GGATAGGGAT GCTGCATCTG CTCGGGCTG GCCTAATATG  
 751 CACACAGTCA ATGGTTATGT AAACAGGTCT CTGCCAGGTC TGATTGGATG  
 801 CCACAGGAAA TCAGTCTATT GGCATGTGAT TGGAAATGGC ACCACTCCTG  
 851 AAGTGCACCT AATATTCTC GAAGGTCACA CATTTCCTGT GAGGAACCAT  
 901 CGCCAGGCGT CCTTGGAAAT CTCGCCAATA ACTTTCCTTA CTGCTCAAAC  
 951 ACTCTTGATG GACCTTGGAC AGTTTCTACT GTTTTGTCAT ATCTCTTCCC  
 1001 ACCAACATGA TGGCATGGAA GCTTATGTCA AAGTAGACAG CTGTCCAGAG  
 1051 GAACCCCAAC TACGAATGAA AAATAATGAA GAAGCGGAAG ACTATGATGA  
 1101 TGATCTTACT GATTCTGAAA TGGATGTGGT CAGGTTTGAT GATGACAAC  
 1151 CTCCTTCCTT TATCCAAATT CGCTCAGTTG CCAAGAAGCA TCCTAAAAC  
 1201 TGGGTACATT ACATTGCTGC TGAAGAGGAG GACTGGGACT ATGCTCCCTT  
 1251 AGTCCTCGCC CCGATGACA GAAGTTATAA AAGTCAATAT TTGAACAATG  
 1301 GCCCTCAGCG GATTGGTAGG AAGTACAAAA AAGTCCGATT TATGGCATA  
 1351 ACAGATGAAA CCTTAAAGAC TCGTGAAGCT ATTCAGCATG AATCAGGAAT  
 1401 CTTGGGACCT TTACTTTATG GGGAAAGTTG AGACACACTG TTGATTATAT  
 1451 TTAAGAATCA AGCAAGCAGA CCATATAACA TCTACCCTCA CGGAATCACT  
 1501 GATGTCCGTC CTTTGTATTC AAGGAGATTA CCAAAAAGGTG TAAAACATTT  
 1551 GAAGGATTTT CCAATTCTGC CAGGAGAAAT ATTCAAATAT AAATGGACAG  
 1601 TGACTGTAGA AGATGGGCCA ACTAAATCAG ATCCTCGGTG CCTGACCCGC  
 1651 TATTACTCTA GTTTCGTTAA TATGGAGAGA GATCTAGCTT CAGGACTCAT  
 1701 TGGCCCTCTC CTCATCTGCT ACAAAGAATC TGTAGATCAA AGAGGAAACC  
 1751 AGATAATGTC AGACAAGAGG AATGTATCC TGTTTTCTGT ATTTGATGAG  
 1801 AACCAGAGCT GGTACCTCAC AGAGAATATA CAACGCTTTC TCCCAATCC  
 1851 AGCTGGAGTG CAGCTTGAGG ATCCAGAGTT CCAAGCCTCC AACATCATGC  
 1901 ACAGCATCAA TGGCTATGTT TTTGATAGTT TGCAGTTGTC AGTTTGTGTTG  
 1951 CATGAGGTGG CATACTGGTA CATTCTAAGC ATTTGGAGCAC AGACTGACTT  
 2001 CCTTCTGTC TTCTTCTCTG GATATACCTT CAAACACAAA ATGGTCTATG  
 2051 AAGACACACT CACCCTATTC CCATTCTCAG GAGAAACTGT CTTTATGTCG  
 2101 ATGAAAACC CAGGTCATG GATTCTGGGG TGCCACAACCT CAGACTTTCG  
 2151 GAACAGAGGC ATGACCGCCT TACTGAAGGT TTCTAGTTGT GACAAGAACA  
 2201 CTGGTGATTA TTACGAGGAC AGTTATGAAG ATATTTACAG ATACTTGCTG  
 2251 AGTAAAAACA ATGCCATTGA ACCAAGAAGC TTCGGTACCT CAGAGTCTGC  
 2301 TACCCCGGAG TCAGGGCCAG GATCAGAGCC AGCCACCTCC GGGTCTGAGA  
 2351 CACCCGGGAC TTCCGAGAGT GCCACCCCTG AGTCCGGACC CGGGTCCGAG  
 2401 CCCGCCACTT CCGGCTCCGA AACTCCCGGC ACAAGCGAGA GCGTACCCC  
 2451 AGAGTCAGGA CCAGGAACAT CTACAGAGCC CTCTGAAGGC TCCGCTCCAG  
 2501 GGTCCCCAGC CGGCAGTCCC ACTAGCACCG AGGAGGGAAC CTCTGAAAGC  
 2551 GCCACACCCG AATCAGGGCC AGGGTCTGAG CCTGCTACCA GCGGCAGCGA  
 2601 GACACCAGGC ACCTCTGAGT CCGCCACACC AGAGTCCGGA CCCGGATCTC  
 2651 CCGCTGGGAG CCCACCTCC ACTGAGGAGG GATCTCTGTC TGGCTCTCCA  
 2701 ACATCTACTG AGGAAGGTAC CTCACCGAG CCATCCGAGG GATCAGCTCC  
 2751 CGGCACCTCA GAGTCGGCAA CCCCAGGATC TGGACCCGGA ACTTCCGAAA  
 2801 GTGCCACACC AGAGTCCGGT CCCGGGACTT CAGAATCAGC AACACCCGAG  
 2851 TCCGGCCCTG GGTCTGAACC CGCCACAAGT GGTAGTGAGA CACCAGGATC  
 2901 AGAACCTGCT ACCTCAGGGT CAGAGACACC CGGATCTCCG GCAGGCTCAC

[0701]



2951 CAACCTCCAC TGAGGAGGGC ACCAGCACAG AACCAAGCGA GGGCTCCGCA  
3001 CCCGGAACAA GCACTGAACC CAGTGAGGGT TCAGCACCCG GCTCTGAGCC  
3051 GGCCACAAGT GGCAGTGAGA CACCCGGCAC TTCAGAGAGT GCCACCCCCG  
3101 AGAGTGGCCC AGGCACTAGT ACCGAGCCCT CTGAAGGCAG TGCGCCAGCC  
3151 TCGAGCGAAA TAACTCGTAC TACTCTTCAG TCAGATCAAG AGGAAATCGA  
3201 TTATGATGAT ACCATATCAG TTGAAATGAA GAAGGAAGAT TTTGACATTT  
3251 ATGATGAGGA TGAAAATCAG AGCCCCCGCA GC'TTTCAAAA GAAAACACGA  
3301 CACTATTTTA TTGCTGCAGT GGAGAGGCTC TGGGATTATG GGATGAGTAG  
3351 CTCCCCACAT GTTCTAAGAA ACAGGGCTCA GAGTGGCAGT GTCCCTCAGT  
3401 TCAAGAAAGT TGTTTTCCAG GAATTTACTG ATGGCTCCTT TACTCAGCCC  
3451 TTATACCGTG GAGAATAAA TGAACATTTG GGACTCCTGG GGCCATATAT  
3501 AAGAGCAGAA GTTGAAGATA ATATCATGGT AACTTTCAGA AATCAGGCCT  
3551 CTCGTCCCTA TTCCTTCTAT TCTAGCCTTA TTTCTTATGA GGAGATCAG  
3601 AGGCAAGGAG CAGAACCCTAG AAAAAACTTT GTCAAGCCTA ATGAAACCAA  
3651 AACTTACTTT TGGAAGTGC AACATCATAT GGCACCCACT AAAGATGAGT  
3701 TTGACTGCAA AGCCTGGGCT TATTTCTCTG ATGTTGACCT GGAAAAAGAT  
3751 GTGCACTCAG GCCTGATTGG ACCCCTTCTG GTCTGCCACA CTAACACACT  
3801 GAACCCTGCT CATGGGAGAC AAGTGACAGT ACAGGAATTT GCTCTGTTTT  
3851 TCACCATCTT TGATGAGACC AAAAGCTGGT ACTTCACTGA AAATATGGAA  
3901 AGAAACTGCA GGGCTCCCTG CAATATCCAG ATGGAAGATC CCACTTTTAA  
3951 AGAGAATTAT CGCTTCCATG CAATCAATGG CTACATAATG GATACACTAC  
4001 CTGGCTTAGT AATGGCTCAG GATCAAAGGA TTCGATGGTA TCTGCTCAGC  
4051 ATGGGCAGCA ATGAAAACAT CCATTCATAT CATTTCAGTG GACATGTGTT  
4101 CACTGTACGA AAAAAAGAG AGTATAAAAT GGCACGTAC AATCTCTATC  
4151 CAGGTGTTTT TGAGACAGT GAAATGTTAC CATCCAAAGC TGGAAATTTGG  
4201 CGGGTGGAAT GCCTTATTGG CGAGCATCTA CATGCTGGGA TGAGCACACT  
4251 TTTTCTGGTG TACAGCAATA AGTGTGAGAC TCCCCTGGGA ATGGCTTCTG  
4301 GACACATTAG AGATTTTCAG ATTACAGCTT CAGGACAATA TGGACAGTGG  
4351 GCCCCAAAGC TGGCCAGACT TCATTATTC GGATCAATCA ATGCCTGGAG  
[0702] 4401 CACCAAGGAG CCCTTTTCTT GGATCAAGGT GGATCTGTTG GCACCAATGA  
4451 TTATTCACGG CATCAAGACC CAGGGTGCCG GTCAGAAGTT CTCCAGCCTC  
4501 TACATCTCTC AGTTTATCAT CATGTATAGT CTTGATGGGA AGAAGTGCCA  
4551 GACTTATCGA GGAAATTCCA CTGGAACCTT AATGGTCTTC TTTGGCAATG  
4601 TGGATTATC TGGGATAAAA CACAATATTT TTAACCCCTC AATTATTGCT  
4651 CGATACATCC GTTTGCACCC AACTCATTAT AGCATTCGCA GCACTCTTCG  
4701 CATGGAGTTG ATGGGCTGTG ATTTAAATAG TTGCAGCATG CCATTGGGAA  
4751 TGGAGAGTAA AGCAATATCA GATGCACAGA TTA CTGCTTC ATCCTACTTT  
4801 ACCAATATGT TTGCCACCTG GTCTCCTTCA AAAGCTCGAC TTCACCTCCA  
4851 AGGGAGGAGT AATGCCTGGA GACCTCAGGT GAATAATCCA AAAGAGTGGC  
4901 TGCAAGTGGA CTTCCAGAAG ACAATGAAAG TCACAGGAGT AACTACTCAG  
4951 GGAGTAAAT CTCTGCTTAC CAGCATGTAT GTGAAGGAGT TCCTCATCTC  
5001 CAGCAGTCAA GATGGCCATC AGTGGACTCT CTTTTTTCAG AATGGCAAAG  
5051 TAAAGGTTTT TCAGGGAAAT CAAGACTCCT TCACACCTGT GGTGAACTCT  
5101 CTAGACCCAC CGTTACTGAC TCGCTACCTT CGAATTCACC CCCAGAGTTG  
5151 GGTGCACCAG ATTGCCCTGA GGATGGAGGT TCTGGGCTGC GAGGCACAGG  
5201 ACCTCTACGA CAAAACCTAC ACATGCCAC CGTGCCAGC TCCAGAATC  
5251 CTGGGCGGAC CGTCAGTCTT CCTCTTCCCC CAAAACCCA AGGACACCCCT  
5301 CATGATCTCC CGGACCCCTG AGGTACATG CGTGGTGGTG GACGTGAGCC  
5351 ACGAAGACCC TGAGGTCAAG TTCAACTGGT ACGTGGACGG CGTGGAGGTG  
5401 CATAATGCCA AGACAAAGC GCGGGAGGAG CAGTACAACA GCACGTACCG  
5451 TGTGGTCAGC GTCCTCACCG TCCTGCACCA GGACTGGCTG AATGGCAAGG  
5501 AGTACAAGTG CAAGGTCTCC AACAAAGCCC TCCCAGCCCC CATCGAGAAA  
5551 ACCATCTCCA AAGCCAAAGG GCAGCCCCGA GAACCACAGG TGTACACCCCT  
5601 GCCCCATCC CGGGATGAGC TGACCAAGAA CCAGGTCAGC CTGACCTGCC  
5651 TGGTCAAAGG CTTCTATCCC AGCGACATCG CCGTGGAGTG GGAGAGCAAT  
5701 GGGCAGCCGG AGAACAATA CAAGACCACG CCTCCCCTGT TGGACTCCGA  
5751 CGGCTCCTTC TTCCTCTACA GCAAGCTCAC CGTGGACAAG AGCAGGTGGC  
5801 AGCAGGGGAA CGTCTTCTCA TGCTCCGTGA TGCATGAGGC TCTGCACAAC  
5851 CACTACACGC AGAAGAGCCT CTCCCTGTCT CCGGGTAAAT GA

[0703] pSYN FVIII 310蛋白质序列(除2个氨基酸残基之外,B结构域完全缺失,并且在aa 742之后插入288 AE-XTEN的FVIII) (SEQ ID NO:171)

```

1   ATRRYYLGA V ELSWDYMQSD LGELPVDARF PPRVPKSFPP NTSVVYKCTL
51  FVEFTDHLFN IAKPRPPWVG LLGPTIQAEV YDTVVITLKN MASHPVSLHA
101 VGVSYWKASE GA EYDDQTSQ REKEDDKVFP GGSHTYVWQV LKENGPMASD
151 PLCLTYSYLS HVDLVKDLNS GLIGALLVCR EGSLAKEKTQ TLHKFILLFA
201 VFDEGKSWHS ETKNSLMQDR DAASARAWPK MHTVNGYVNR SLPGLIGCHR
251 KSVYWHVIGM GTTPEVHSIF LEGHTFLVRN HRQASLEISP ITFLTAQTLL
301 MDLGQFLLFC HISSHQHDGM EAYVKVDSCP EEPQLRMKNN EEAEDYDDDL
351 TDSEMDVVRF DDDNSPSFIQ IRSVAKKHPK TWVHYIAAEE EDWDYAPLVL
401 APDDRSYKSQ YLNNGPQRIG RKYKKVRFMA YDDEFKTR EAIQHEGILG
451 PLLYGEVGD T LLIIFKNQAS RPYNIYPHGI TDVRPLYSR RLPKGVKHLKD
501 FPILPGEIFK YKWTVTVEDG PTKSDPRCLT RYYSSFVNME RDLASGLIGP
551 LLICYKESVD QRGNQIMSDK RNVILFSVFD ENRSWYLTE N IQRFLPNPAG
601 VQLEDPEFQA SNIMHSINGY VFDSLQLSVC LHEVAYWYIL SIGAQTDFLS
651 VFFSGYTFKH KMYEDTLTL FPFSGETVFM SMENPGLWIL GCHNSDFRNR
701 GMTALLKVSS CDKNTGDYEE DSYEDISAYL LSKNNAIEPR SFGTSESATP
751 ESGPGSEPAT SGSETPGTSE SATPESGPGS EPATSGSETP GTSESATPES
801 GPGTSTEPSE GSAPGSPAGS PTSTEEGTSE SATPESGPGS EPATSGSETP
851 GTSESATPES GPGSPAGSPT STEEGSPAGS PTSTEEGTST EPSEGSAPGT
901 SESATPESGP GTSESATPES GPGTSESATP ESGPGSEPAT SGSETPGSEP
[0704] 951 ATSGSETPGS PAGESPTSTEE GTSTEPSEGS APGTSTEPSE GSAPGSEPAT
1001 SGSETPGTSE SATPESGPGT STEPSEGSAP ASSEITR TTL QSDQEEIDYD
1051 DTISVEMKKE DFDIYDEDEN QSPRSFQKKT RHYFIAAVER LWDYGMSSSP
1101 HVLRNRAQSG SVPQFKKVV F QEFTDGSFTQ PLYRGELNEH LGLLGPYIRA
1151 EVEDNIMVTF RNQASRPYSF YSSLISYEED QRQGAEP RKN FVKPNETKTY
1201 FWKVQHMAP TKDEFDCKAW AYFSDVDLEK DVHSLGIGPL LVCHTNTLNP
1251 AHGRQVTVQE FALFFTIFDE TKSWYFTENM ERNCRAPCNI QMEDPTFKEN
1301 YRFHAINGYI MDTLPGLVMA QDQRIRWYLL SMGSNENIHS IHFSGHVFTV
1351 RKKEEYKMAL YNLYPGVFET VEMLPKAGI WRVECLIGEH LHAGMSTLFL
1401 VYSNKCQTPL GMASGHIRDF QITASGQYGO WAPKLARLHY SGSINAWSTK
1451 EPPSWIKVDL LAPMIIHG I K TQGARQKFSS LYISQFIIMY SLDGKKWQTY
1501 RGNSTGTLMV FFGNV DSSGI KHNIFNPPII ARYIRLHP TH YSIRSTLRME
1551 LMGCDLNSCS MPLGMESKAI SDAQITASSY FTNMFATWSP SKARLHLQGR
1601 SNAWRPQVNN PKEWLQVDFQ KTMKVTGVT T QGVKSLTSM YVKEFLISSS
1651 QDGHQWTLFF QNGKVKVFGQ NQDSFTP VVN SLDPPLLTRY LRIHPQSWVH
1701 QIALRMEVLG CEAQDLYDKT HTCPCPAP E LLGGPSVFLF PPKPKDTLMI
1751 SRTPEVTCV V VDVSHEDPEV KFNWYVDGVE VHNATKPRE EQYNSTYRVV
1801 SVLTVLHQDW LNGKEYKCKV SNKALPAP IE KTISKAKGQP REPQVYTLPP
1851 SRDELTKNQV SLTCLVKGFY PSDIAVEWES NGQPENNYKT TPPVLDS DGS
1901 FFLYSKLTV D KSRWQQGNV F SCSVMHEALH NHYTQKSLSL SPGK*

```

[0705] pSYN FVIII 312核苷酸序列(编码除5个氨基酸残基之外,B结构域完全缺失,并且在aa 745之后插入288 AE-XTEN的FVIII-B5形式)(SEQ ID NO:172)

```

1   ATGCAAATAG AGCTCTCCAC CTGCTTCTTT CTGTGCCTTT TGCGATTCTG
51  CTTTAGTGCC ACCAGAAGAT ACTACCTGGG TGCAGTGGAA CTGTCATGGG
[0706] 101 ACTATATGCA AAGTGATCTC GGTGAGCTGC CTGTGGACGC AAGATTT CCT
151 CCTAGAGTGC CAAAATCTTT TCCATTCAAC ACCTCAGTCG TGTACAAAAA
201 GACTCTGTTT GTAGAATTCA CGGATCACCT TTTCAACATC GCTAAGCCAA
251 GGCCACCCTG GATGGGTCTG CTAGGTCCTA CCATCCAGGC TGAGGTTTAT

```

301 GATACAGTGG TCATTACACT TAAGAACATG GCTTCCCATC CTGTCAGTCT  
 351 TCATGCTGTT GGTGTATCCT ACTGGAAAGC TTCTGAGGGA GCTGAATATG  
 401 ATGATCAGAC CAGTCAAAGG GAGAAAGAAG ATGATAAAGT CTTCCCTGGT  
 451 GGAAGCCATA CATATGTCTG GCAGGTCCCTG AAAGAGAATG GTCCAATGGC  
 501 CTCTGACCCA CTGTGCCTTA CCTACTCATA TCTTTCTCAT GTGGACCTGG  
 551 TAAAAGACTT GAATTCAGGC CTCATTGGAG CCCTACTAGT ATGTAGAGAA  
 601 GGGAGTCTGG CCAAGGAAAA GACACAGACC TTGCACAAAT TTATACTACT  
 651 TTTTGCTGTA TTTGATGAAG GGAAAAGTTG GCACTCAGAA ACAAAGAACT  
 701 CCTTGATGCA GGATAGGGAT GCTGCATCTG CTCGGGCCCTG GCCTAAAATG  
 751 CACACAGTCA ATGGTTATGT AAACAGGTCT CTGCCAGGTC TGATTGGATG  
 801 CCACAGGAAA TCAGTCTATT GGCATGTGAT TGGAAATGGGC ACCACTCCTG  
 851 AAGTGCACCTC AATATTCTCTC GAAGGTCACA CATTTCCTGT GAGGAACCAT  
 901 CGCCAGGCGT CTTGGAAAT CTCGCCAATA ACTTTCCTTA CTGCTCAAAC  
 951 ACTCTTGATG GACCTTGGAC AGTTTCTACT GTTTTGTCTAT ATCTCTTCCC  
 1001 ACCAACATGA TGGCATGGAA GCTTATGTCA AAGTAGACAG CTGTCCAGAG  
 1051 GAACCCCAAC TACGAATGAA AAATAATGAA GAAGCGGAAG ACTATGATGA  
 1101 TGATCTTACT GATTCTGAAA TGGATGTGGT CAGGTTTGAT GATGACAACCT  
 1151 CTCCTTCCTT TATCCAAATT CGCTCAGTTG CCAAGAAGCA TCCTAAAACCT  
 1201 TGGGTACATT ACATTGCTGC TGAAGAGGAG GACTGGGACT ATGCTCCCTT  
 1251 AGTCCTCGCC CCCGATGACA GAAGTTATAA AAGTCAATAT TTGAACAATG  
 1301 GCCCTCAGCG GATTGGTAGG AAGTACAAAA AAGTCCGATT TATGGCATACT  
 1351 ACAGATGAAA CCTTTAAGAC TCGTGAAGCT ATTCAGCATG AATCAGGAAT  
 1401 CTTGGGACCT TTACTTTATG GGGAAGTTGG AGACACACTG TTGATTATAT  
 1451 TTAAGAATCA AGCAAGCAGA CCATATAACA TCTACCCTCA CGGAATCACT  
 1501 GATGTCGTC CTTTGTATTC AAGGAGATTA CAAAAGGTG TAAAACATTT  
 1551 GAAGGATTTT CCAATTCTGC CAGGAGAAAT ATTCAAATAT AAATGGACAG  
 1601 TGACTGTAGA AGATGGGCCA ACTAAATCAG ATCCTCGGTG CCTGACCCGC  
 1651 TATTACTCTA GTTTCGTTAA TATGGAGAGA GATCTAGCTT CAGGACTCAT  
 1701 TGGCCCTCTC CTCATCTGCT ACAAAGAATC TGTAGATCAA AGAGGAAACC  
 [0707] 1751 AGATAATGTC AGACAAGAGG AATGTCATCC TGTTTTCTGT ATTTGATGAG  
 1801 AACCGAAGCT GGTACCTCAC AGAGAATATA CAACGCTTTC TCCCAATCC  
 1851 AGCTGGAGTG CAGCTTGAGG ATCCAGAGTT CCAAGCCTCC AACATCATGC  
 1901 ACAGCATCAA TGGCTATGTT TTTGATAGTT TGCAGTTGTC AGTTTGTGTTG  
 1951 CATGAGGTGG CATACTGGTA CATTCTAAGC ATTGGAGCAC AGACTGACTT  
 2001 CCTTTCTGTC TTCTTCTCTG GATATACCTT CAAACACAAA ATGGTCTATG  
 2051 AAGACACACT CACCCTATTC CCATTCTCAG GAGAACTGT CTTTATGTCG  
 2101 ATGGAAAACC CAGGTCTATG GATTCTGGGG TGCCACAACCT CAGACTTTTCG  
 2151 GAACAGAGGC ATGACCCGCT TACTGAAGGT TTCTAGTTGT GACAAGAACA  
 2201 CTGGTGATTA TTACGAGGAC AGTTATGAAG ATATTTTCAGC ATACTTGCTG  
 2251 AGTAAAAACA ATGCCATTGA ACCAAGAAGC TTCTCTCAA ACGGTACCTC  
 2301 AGAGTCTGCT ACCCCCGAGT CAGGGCCAGG ATCAGAGCCA GCCACCTCCG  
 2351 GGTCTGAGAC ACCCGGGACT TCCGAGAGTG CCACCCCTGA GTCCGGACCC  
 2401 GGGTCCGAGC CCGCCACTTC CGGCTCCGAA ACTCCCGGCA CAAGCGAGAG  
 2451 CGTACCCCA GAGTCAGGAC CAGGAACATC TACAGAGCCC TCTGAAGGCT  
 2501 CCGCTCCAGG GTCCCCAGCC GGCAGTCCCA CTAGCACCGA GGAGGGAACC  
 2551 TCTGAAAGCG CCACACCCGA ATCAGGGCCA GGGTCTGAGC CTGCTACCAG  
 2601 CGGCAGCGAG ACACCAGGCA CCTCTGAGTC CGCCACACCA GAGTCCGGAC  
 2651 CCGGATCTCC CGCTGGGAGC CCCACCTCCA CTGAGGAGGG ATCTCCTGCT  
 2701 GGCTCTCCAA CATCTACTGA GGAAGGTACC TCAACCGAGC CATCCGAGGG  
 2751 ATCAGTCCC GGCACCTCAG AGTCGGCAAC CCCGGAGTCT GGACCCGGAA  
 2801 CTTCCGAAAAG TGCCACACCA GAGTCCGGTC CCGGGACTTC AGAATCAGCA  
 2851 ACACCCGAGT CCGGCCCTGG GTCTGAACCC GCCACAAGTG GTAGTGAGAC  
 2901 ACCAGGATCA GAACCTGCTA CCTCAGGGTC AGAGACACCC GGATCTCCGG  
 2951 CAGGCTCACC AACCTCCACT GAGGAGGGCA CCAGCACAGA ACCAAGCGAG  
 3001 GGCTCCGCAC CCGGAACAAG CACTGAACCC AGTGAGGGTT CAGCACCCGG  
 3051 CTCTGAGCCG GCCACAAGTG GCAGTGAGAC ACCCGGCACT TCAGAGAGTG  
 3101 CCACCCCCGA GAGTGGCCCA GGCCTAGTA CCGAGCCCTC TGAAGGCAGT  
 3151 GCGCCAGCCT CGAGCGAAAT AACTCGTACT ACTCTTCAGT CAGATCAAGA  
 3201 GGAAATCGAT TATGATGATA CCATATCAGT TGAAATGAAG AAGGAAGATT

```

3251   TTGACATTTA TGATGAGGAT GAAAATCAGA GCCCCCGCAG CTTTCAAAG
3301   AAAACACGAC ACTATTTTAT TGCTGCAGTG GAGAGGCTCT GGGATTATGG
3351   GATGAGTAGC TCCCCACATG TTCTAAGAAA CAGGGGCTCAG AGTGGCAGTG
3401   TCCCTCAGTT CAAGAAAGTT GTTTCCAGG AATTTACTGA TGGCTCCTTT
3451   ACTCAGCCCT TATACCGTGG AGAACTAAAT GAACATTTGG GACTCCTGGG
3501   GCCATATATA AGAGCAGAAG TTGAAGATAA TATCATGGTA ACTTTCAGAA
3551   ATCAGGCCCT TCGTCCCTAT TCCTTCTATT CTAGCCTTAT TTCTTATGAG
3601   GAAGATCAGA GGCAAGGAGC AGAACCTAGA AAAAAGTTTG TCAAGCCTAA
3651   TGAAACCAA ACTTACTTTT GGAAAGTGCA ACATCATATG GCACCCACTA
3701   AAGATGAGTT TGACTGCAA GCCTGGGCTT ATTTCTCTGA TGTTGACCTG
3751   GAAAAAGATG TGCCTCAGG CCTGATTGGA CCCCTTCTGG TCTGCCACAC
3801   TAACACACTG AACCTGCTC ATGGGAGACA AGTGACAGTA CAGGAATTTG
3851   CTCTGTTTTT CACCATCTTT GATGAGACCA AAAGCTGGTA CTTCACTGAA
3901   AATATGGAAA GAAACTGCAG GGCTCCCTGC AATATCCAGA TGGAAGATCC
3951   CACTTTTAAA GAGAATTATC GCTTCCATGC AATCAATGGC TACATAATGG
4001   ATACACTACC TGGCTTAGTA ATGGCTCAGG ATCAAAGGAT TCGATGGTAT
4051   CTGCTCAGCA TGGGCAGCAA TGAAAACATC CATTCTATTC ATTTCAGTGG
4101   ACATGTGTTT ACTGTACGAA AAAAAGAGGA GTATAAATG GCACTGTACA
4151   ATCTCTATCC AGGTGTTTTT GAGACAGTGG AAATGTTACC ATCCAAAGCT
4201   GGAATTTGGC GGGTGGAAAT CCTTATTGGC GAGCATCTAC ATGCTGGGAT
4251   GAGCACACTT TTTCTGGTGT ACAGCAATAA GTGTCAGACT CCCCTGGGAA
4301   TGGCTTCTGG ACACATTAGA GATTTTCAGA TTACAGCTTC AGGACAATAT
4351   GGACAGTGGG CCCCAAAGCT GGCCAGACTT CATTATTCCG GATCAATCAA
4401   TGCCTGGAGC ACCAAGGAGC CCTTTTCTTG GATCAAGGTG GATCTGTTGG
4451   CACCAATGAT TATTCACGGC ATCAAGACCC AGGGTGCCCG TCAGAAGTTC
4501   TCCAGCCTCT ACATCTCTCA GTTTATCATC ATGTATAGTC TTGATGGGAA
[0708] 4551   GAAGTGGCAG ACTTATCGAG GAAATTCCAC TGGAACCTTA ATGGTCTTCT
4601   TTGGCAATGT GGATTCATCT GGGATAAAAC ACAATATTTT TAACCCTCCA
4651   ATTATTGCTC GATACATCCG TTTGCACCCA ACTCATTATA GCATTGCGAG
4701   CACTCTTCGC ATGGAGTTGA TGGGCTGTGA TTTAAATAGT TGCAGCATGC
4751   CATTGGGAAT GGAGAGTAAA GCAATATCAG ATGCACAGAT TACTGCTTCA
4801   TCCTACTTTA CCAATATGTT TGCCACCTGG TCTCCTTCAA AAGCTCGACT
4851   TCACCTCCAA GGGAGGAGTA ATGCCTGGAG ACCTCAGGTG AATAATCCAA
4901   AAGAGTGGCT GCAAGTGGAC TTCCAGAAGA CAATGAAAGT CACAGGAGTA
4951   ACTACTCAGG GAGTAAAATC TCTGCTTACC AGCATGTATG TGAAGGAGTT
5001   CCTCATCTCC AGCAGTCAAG ATGGCCATCA GTGGACTCTC TTTTTTCAGA
5051   ATGGCAAAGT AAAGGTTTTT CAGGGAAATC AAGACTCCTT CACACCTGTG
5101   GTGAACTCTC TAGACCCACC GTTACTGACT CGCTACCTTC GAATTCACCC
5151   CCAGAGTTGG GTGCACCAGA TTGCCCTGAG GATGGAGGTT CTGGGCTGCG
5201   AGGCACAGGA CCTCTACGAC AAAACTCACA CATGCCACC GTGCCAGCT
5251   CCAGAACTCC TGGGCGGACC GTCAGTCTTC CTCTTCCCC CAAAACCCAA
5301   GGACACCCTC ATGATCTCCC GGACCCTGA GGTACATGC GTGGTGGTGG
5351   ACGTGAGCCA CGAAGACCCT GAGGTCAAGT TCAACTGGTA CGTGGACGGC
5401   GTGGAGGTGC ATAATGCCAA GACAAAGCCG CGGGAGGAGC AGTACAACAG
5451   CACGTACCGT GTGGTCAGCG TCCTCACCGT CCTGCACCAG GACTGGCTGA
5501   ATGGCAAGGA GTACAAGTGC AAGGTCTCCA ACAAAGCCCT CCCAGCCCCC
5551   ATCGAGAAA CCATCTCCAA AGCCAAAGGG CAGCCCCGAG AACCACAGGT
5601   GTACACCCTG CCCCCATCCC GGGATGAGCT GACCAAGAAC CAGGTCAGCC
5651   TGACCTGCCT GGTCAAAGGC TTCTATCCCA GCGACATCGC CGTGGAGTGG
5701   GAGAGCAATG GGCAGCCGGA GAACAACACT AAGACCACGC CTCCCCTGTT
5751   GGACTCCGAC GGCTCCTTCT TCCTCTACAG CAAGCTCACC GTGGACAAGA
5801   GCAGGTGGCA GCAGGGGAAC GTCTTCTCAT GCTCCGTGAT GCATGAGGCT
5851   CTGCACAACC ACTACACGCA GAAGAGCCTC TCCCTGTCTC CGGGTAATG

```

[0709] pSYN FVIII 312蛋白质序列(除5个氨基酸残基之外,B结构域完全缺失,并且在aa 745之后插入288 AE-XTEN的FVIII-B5形式)(SEQ ID NO:173)

```

1   ATRRYYLGAV ELSWDYMQSD LGELPVDARF PPRVPKSFPF NTSVVKKTL

51   FVEFTDHLFN IAKPRPPWVG LIGPTIQAEV YDTVVITLKN MASHPVSLHA
101  VGVSYWKASE GAHYDDQTSQ REKEDDKVFP GGSHTYVWQV LKENGPMASD
151  PLCLTYSYLS HVDLVKDLNS GLIGALLVCR EGSLAKEKTQ TLHKFILLFA
201  VFDEGKSWHS ETKNSLMQDR DAASARAWPK MHTVNGYVNR SLPGLIGCHR
251  KSVYWHVIGM GTTPEVHSIF LEGHTFLVRN HRQASLEISP ITFLTAQTL
301  MDLGQFLLFC HISSHQHDGM EAYVKVDSCP EEPQLRMKNN EEAEDYDDDL
351  TDSEMDVVRV DDDNSPSFIQ IRSVAKKHPK TWVHYIAAEE EDWDYAPLVL
401  APDDRSYKSQ YLNNGPQRIG RKYKKVRFMA YTDETFKTR AIQHESGILG
451  PLYGGEVGDV LLIIFKNQAS RPYNIYPHGI TDVRPLYSRR LPKGVKHLKD
501  FPILPGEIFK YKWTVTVEDG PTKSDPRCLT RYSSFFVNM RDLASGLIGP
551  LLICYKESVD QRGNQIMSDK RNVILFSVFD ENRSWYLTEN IQRFLPNPAG
601  VQLEDPEFQA SNIMHSINGY VFDSLQLSVC LHEVAYWYIL SIGAQDFLS
651  VFFSGYTFKH KMYEDTLTL FPFSGETVFM SMENPGLWIL GCHNSDFRNR
701  GMTALLKVSS CDKNTGDYEE DSYEDISAYL LSKNNAIEPR SFSQNGTSES
751  ATPESGPGSE PATSGSETPG TSESATPESG PGSEPATSGS ETPGTSESAT
801  PESGPGTSTE PSEGSAPGSP AGSPTSTEEG TSESATPESG PGSEPATSGS
851  ETPGTSESAT PESGPGSPAG SPTSTEEGSP AGSPTSTEEG TSTEPSEGSA
[0710] 901  PGTSESATPE SGPGTSESAT PESGPGTSES ATPESGPGSE PATSGSETPG
951  SEPATSGSET PGSPAGSPTS TEEGTSTEPS EGSAPGTSTE PSEGSAPGSE
1001 PATSGSETPG TSESATPESG PGTSTEPSEG SAPASSEITR TTLQSDQEEI
1051 DYDDTISVEM KKEDFDIYDE DENQSPRSFQ KKTRHYFIAA VERLWDYGMS
1101 SSPHVLNRRA QSGSVPQFKK VVFQEFTDGS FTQPLYRGEL NEHLGLLGPY
1151 IRAEVEDNIM VTFRNQASRP YSFYSSLISY EEDQRQGAEP RKNFVKPNET
1201 KTYFWKVQHH MAPTKDEFDC KAWAYFSDVD LEKDVHSGLI GPLLVCHTNT
1251 LNPAHGRQVT VQEFALFFT I FDETKSWYFT ENMERNCRAP CNIQMEDPTF
1301 KENYRFHAIN GYIMDTLPGL VMAQDQRIRW YLLSMGSNEN IHSIHFSGHV
1351 FTVRKKEEYK MALYNLYPGV FETVEMLPSK AGIWRVECLI GEHLHAGMST
1401 LFLVYSNKCQ TPLGMASGHI RDFQITASGQ YGQWAPKLAR LHYSGSINAW
1451 STKEPFSWIK VDLLAPMI IH GIKTQGARQK FSSLYISQFI IMYSLDGKKW
1501 QTYRGNSTGT LMVFFGNVDS SGIKHNIFNP PIIARYIRLH PTHYSIRSTL
1551 RMELMGCDLN SCSMPLGMES KAISDAQITA SSYFTNMFAT WSPSKARLHL
1601 QGRSNAWRPQ VNNPKEWLQV DFQKTMKVTG VTTQGVKSL TSMYVKEFLI
1651 SSSQDGHQWT LFFQNGKVKV FQGNQDSFTP VVNSLDPPLL TRYLRHPQS
1701 WVHQIALRME VLGCEAQDLY DKTHTCPPCP APELLGGPSV FLFPPKPKDT
1751 LMISRTPEVT CVVVDVSHED PEVKFNWYVD GVEVHNAKTK PREEQYNSTY
1801 RVVSVLTVLH QDWLNGKEYK CKVSNKALPA PIEKTISKAK GQPREPQVYT
1851 LPPSRDELTK NQVSLTCLVK GFYPSDIAVE WESNGQPENN YKTTTPVLDL
1901 DGSFFLYSKL TVDKSRWQOG NVFSCSVME ALHNHYTQKS LSLSPGK*

```

[0711] pSYN VWF059核苷酸序列(编码在接头中具有酸性区2(a2)凝血酶位点的VWF D' D3-Fc) (SEQ ID NO:196)

[0712]

```
1   ATGATTCCTG CCAGATTTGC CGGGGTGCTG CTGCTCTGG CCCTCATTTT
51  GCCAGGGACC CTTTGTGCAG AAGGAACTCG CGGCAGGTCA TCCACGGCCC
101 GATGCAGCCT TTTCGGAAGT GACTTCGTCA ACACCTTTGA TGGGAGCATG
151 TACAGCTTTG CGGGATACTG CAGTTACCTC CTGGCAGGGG GCTGCCAGAA
201 ACGCTCCTTC TCGATTATTG GGGACTTCCA GAATGGCAAG AGAGTGAGCC
251 TCTCCGTGTA TCTTGGGGAA TTTTTTGACA TCCATTTGTT TGTCAATGGT
301 ACCGTGACAC AGGGGGACCA AAGAGTCTCC ATGCCCTATG CCTCCAAAGG
351 GCTGTATCTA GAAACTGAGG CTGGGTACTA CAAGCTGTCC GGTGAGGCCT
401 ATGGCTTTGT GGCCAGGATC GATGGCAGCG GCAACTTTCA AGTCCTGCTG
451 TCAGACAGAT ACTTCAACAA GACCTGCGGG CTGTGTGGCA ACTTTAACAT
501 CTTTGCTGAA GATGACTTTA TGACCCAAGA AGGGACCTTG ACCTCGGACC
551 CTTATGACTT TGCCAACTCA TGGGCTCTGA GCAGTGGAGA ACAGTGGTGT
601 GAACGGGCAT CTCCTCCCAG CAGCTCATGC AACATCTCCT CTGGGGAAAT
651 GCAGAAGGGC CTGTGGGAGC AGTGCCAGCT TCTGAAGAGC ACCTCGGTGT
701 TTGCCCGCTG CCACCCTCTG GTGGACCCCG AGCCTTTTGT GGCCCTGTGT
751 GAGAAGACTT TGTGTGAGTG TGCTGGGGGG CTGGAGTGCG CCTGCCCTGC
```

801 CCTCTGGAG TACGCCCGGA CCTGTGCCCA GGAGGGAATG GTGCTGTACG  
 851 GCTGGACCGA CCACAGCGCG TGCAGCCCAG TGTGCCCTGC TGGTATGGAG  
 901 TATAGGCAGT GTGTGTCCCC TTGCGCCAGG ACCTGCCAGA GCCTGCACAT  
 951 CAATGAAATG TGTCAGGAGC GATGCGTGGG TGGCTGCAGC TGCCCTGAGG  
 1001 GACAGCTCCT GGATGAAGGC CTCTGCGTGG AGAGCACCAG GTGTCCCTGC  
 1051 GTGCATTCCG GAAAGCGCTA CCCTCCCAGG ACCTCCCTCT CTCGAGACTG  
 1101 CAACACCTGC ATTTGCCGAA ACAGCCAGTG GATCTGCAGC AATGAAGAAT  
 1151 GTCCAGGGGA GTGCCTTGTC ACTGGTCAAT CCCACTTCAA GAGCTTTGAC  
 1201 AACAGATACT TCACCTTCAG TGGGATCTGC CAGTACCTGC TGGCCCCGGA  
 1251 TTGCCAGGAC CACTCCTTCT CCATTGTTCAT TGAGACTGTC CAGTGTGCTG  
 1301 ATGACCGCGA CGCTGTGTGC ACCCGCTCCG TCACCGTCCG GCTGCTGCG  
 1351 CTGCACAACA GCCTTGTGAA ACTGAAGCAT GGGGCAGGAG TTGCCATGGA  
 1401 TGGCCAGGAC ATCCAGCTCC CCCTCCTGAA AGGTGACCTC CGCATCCAGC  
 1451 ATACAGTGAC GGCCTCCGTG CGCCTCAGCT ACGGGGAGGA CCTGCAGATG  
 1501 GACTGGGATG GCCGCGGGAG GCTGCTGGTG AAGCTGTCCC CCGTCTATGC  
 1551 CGGGAAGACC TCGGCGCTGT GTGGGAATTA CAATGGCAAC CAGGGCGACG  
 1601 ACTTCCTTAC CCCCTCTGGG CTGGCGGAGC CCCGGGTGGA GGACTTCGGG  
 1651 AACGCCTGGA AGCTGCACGG GGACTGCCAG GACCTGCAGA AGCAGCACAG  
 1701 CGATCCCTGC GCCCTCAACC CGCGCATGAC CAGGTTCTCC GAGGAGGCGT  
 1751 GCGCGGTCTT GACGTCCCC ACATTCGAGG CCTGCCATCG TGCCGTACAG  
 1801 CCGTGCCTT ACCTGCGGAA CTGCCGCTAC GACGTGTGCT CCTGCTCGGA  
 1851 CGGCCGCGAG TGCCTGTGCG GCGCCCTGGC CAGCTATGCC GCGGCTGCG  
 1901 CGGGGAGAGG CGTGCGCGTC GCGTGGCGCG AGCCAGGCCG CTGTGAGCTG  
 1951 AACTGCCCGA AAGGCCAGT GTACCTGCAG TCGGGGACCC CCTGCAACCT  
 2001 GACCTGCCGC TCTCTCTTT ACCCGGATGA GGAATGCAAT GAGGCTGCC  
 2051 TGGAGGGCTG CTTCTGCCCC CCAGGGCTCT ACATGGATGA GAGGGGGGAC  
 2101 TGCGTGCCCA AGGCCAGTG CCCCTGTTAC TATGACGGTG AGATCTTCCA  
 2151 GCCAGAAGAC ATCTTCTCAG ACCATCACAC CATGTGCTAC TGTGAGGATG  
 2201 GCTTCATGCA CTGTACCATG AGTGGAGTCC CCGGAAGCTT GCTGCCCTGAC  
 [0713] 2251 GCTGTCTCTA GCAGTCCCCT GTCTCATCGC AGCAAAGGA GCCTATCCTG  
 2301 TCGGCCCCCC ATGGTCAAGC TGGTGTGTCC CGCTGACAAC CTGCGGGCTG  
 2351 AAGGGCTCGA GTGTACCAA ACGTGCCAGA ACTATGACCT GGAGTGATG  
 2401 AGCATGGGCT GTGTCTCTGG CTGCCTCTGC CCCCCGGGCA TGGTCCGGCA  
 2451 TGAGAACAGA TGTGTGGCCC TGGAAAGGTG TCCCTGCTTC CATCAGGGCA  
 2501 AGGAGTATGC CCCTGGAGAA ACAGTGAAGA TTGGCTGCAA CACTTGTGTC  
 2551 TGTCGGGACC GGAAGTGGAA CTGCACAGAC CATGTGTGTG ATGCCACGTG  
 2601 CTCCACGATC GGCATGGCCC ACTACCTCAC CTTGACGGG CTCAAATACC  
 2651 TGTCCCCGGG GGAGTGCCAG TACGTTCTGG TGCAGGATTA CTGCGCAGT  
 2701 AACCTGGGA CCTTTCGGAT CCTAGTGGGG AATAAGGGAT GCAGCCACCC  
 2751 CTCAGTGAAA TGCAAGAAAC GGGTCACCAT CCTGGTGGAG GGAGGAGAGA  
 2801 TTGAGCTGTT TGACGGGGAG GTGAATGTGA AGAGGCCCAT GAAGGATGAG  
 2851 ACTCACTTTG AGGTGGTGGG GTCTGGCCGG TACATCATTC TGCTGCTGGG  
 2901 CAAAGCCCTC TCCGTGGTCT GGGACCGCCA CCTGAGCATC TCCGTGGTCC  
 2951 TGAAGCAGAC ATACCAGGAG AAAGTGTGTG GCCTGTGTGG GAATTTTGTG  
 3001 GGCATCCAGA ACAATGACCT CACCAGCAGC AACCTCCAAG TGGAGGAAGA  
 3051 CCCTGTGGAC TTTGGGAACT CCTGGAAAGT GAGCTCGCAG TGTGCTGACA  
 3101 CCAGAAAAGT GCCTCTGGAC TCATCCCCTG CCACCTGCCA TAACAACATC  
 3151 ATGAAGCAGA CGATGGTGGG TTCTCCTGTG AGAATCCTTA CCAGTGACGT  
 3201 CTTCCAGGAC TGCAACAAGC TGGTGGACCC CGAGCCATAT CTGGATGTCT  
 3251 GCATTTACGA CACCTGCTCC TGTGAGTCCA TTGGGGACTG CGCCGCATTC  
 3301 TGCACACCA TTGCTGCCTA TGCCACGTG TGTGCCAGC ATGGCAAGGT  
 3351 GGTGACCTGG AGGACGGCCA CATTGTGCC CCAGAGCTGC GAGGAGAGGA  
 3401 ATCTCCGGGA GAACGGGTAT GAGGCTGAGT GCGCTATAA CAGCTGTGCA  
 3451 CCTGCCTGTC AAGTCACGTG TCAGCACCTT GAGCCACTGG CCTGCCCTGT  
 3501 GCAGTGTGTG GAGGGCTGCC ATGCCACTG CCTCCAGGG AAAATCCTGG  
 3551 ATGAGCTTTT GCAGACCTGC GTTGACCCTG AAGACTGTCC AGTGTGTGAG  
 3601 GTGGCTGGCC GCGTTTTTCG CTCAGGAAAG AAAGTCACCT TGAATCCCAG  
 3651 TGACCCTGAG CACTGCCAGA TTTGCCACTG TGATGTTGTC AACCTCACCT  
 3701 GTGAAGCCTG CCAGGAGCCG ATATCGGGCG CGCCAACATC AGAGAGCGCC

```

3751  ACCCCTGAAA GTGGTCCCGG GAGCGAGCCA GCCACATCTG GGTCGGAAAC
3801  GCCAGGCACA AGTGAGTCTG CAACTCCCGA GTCCGGACCT GGCTCCGAGC
3851  CTGCCACTAG CGGCTCCGAG ACTCCGGGAA CTTCGAGAG CGCTACACCA
3901  GAAAGCGGAC CCGGAACCAG TACCGAACCT AGCGAGGGCT CTGCTCCGGG
3951  CAGCCCAGCC GGCTCTCCTA CATCCACGGA GGAGGGCACT TCCGAATCCG
4001  CCACCCCGGA GTCAGGGCCA GGATCTGAAC CCGCTACCTC AGGCAGTGAG
4051  ACGCCAGGAA CGAGCGAGTC CGCTACACCG GAGAGTGGGC CAGGGAGCCC
4101  TGCTGGATCT CCTACGTCCA CTGAGGAAGG GTCACCAGCG GGCTCGCCCA
4151  CCAGCACTGA AGAAGGTGCC TCGATATCTG ACAAGAACAC TGGTGATTAT
4201  TACGAGGACA GTTATGAAGA TATTTAGCA TACTTGCTGA GTAAAAACAA
4251  TGCCATTGAA CCAAGAAGCT TCTCTGACAA AACTCACACA TGCCCACCGT
[0714] 4301  GCCCAGCTCC AGAACTCCTG GGCGGACCGT CAGTCTTCTT CTCCCCCA
4351  AAACCAAGG ACACCCTCAT GATCTCCCGG ACCCCTGAGG TCACATGCGT
4401  GGTGGTGGAC GTGAGCCACG AAGACCCTGA GGTCAGTTC AACTGGTACG
4451  TGGACGGCGT GGAGGTGCAT AATGCCAAGA CAAAGCCGCG GGAGGAGCAG
4501  TACAACAGCA CGTACCGTGT GGTCAGCGTC CTCACCGTCC TGCACCAGGA
4551  CTGGCTGAAT GGCAAGGAGT ACAAGTGCAA GGTCTCCAAC AAAGCCCTCC
4601  CAGCCCCCAT CGAGAAAACC ATCTCCAAAG CCAAAGGGCA GCCCCGAGAA
4651  CCACAGGTGT ACACCCTGCC CCCATCCCGG GATGAGCTGA CCAAGAACCA
4701  GGTCAGCCTG ACCTGCCTGG TCAAAGGCTT CTATCCCAGC GACATCGCCG
4751  TGGAGTGGGA GAGCAATGGG CAGCCGGAGA ACAACTACAA GACCACGCCT
4801  CCCGTGTTGG ACTCCGACGG CTCCTTCTTC CTCTACAGCA AGCTCACCGT
4851  GGACAAGAGC AGGTGGCAGC AGGGGAACGT CTTCTCATGC TCCGTGATGC
4901  ATGAGGCTCT GCACAACCAC TACACGCAGA AGAGCCTCTC CCTGTCTCCG
4951  GGTAATGA

```

[0715] pSYN VWF059蛋白质序列(在接头中具有a2区FVIII凝血酶位点的VWF D' D3-Fc) - 粗体加下划线区域显示a2区 (SEQ ID NO:197)



1 MI PARFAGVL LALALILPGT LCAEGTRGRS STARCSLFGS DFVNTFDGSM  
 51 YSFAGYCSYL LAGGCQKRSF SIIGDFQNGK RVSLSVYLGE FFDIHLFVNG  
 101 TVTQGDQRVS MPYASKGLYL ETEAGYYKLS GEAYGFVARI DGSGNFQVLL  
 151 SDRYFNKTCG LCGNFNIFAE DDFMTQEGTL TSDPYDFANS WALSSGEQWC  
 201 ERASPPSSSC NISSGEMQKG LWEQCQLLKS TSVFARCHPL VDPEPFVALC  
 251 EKTLCCECAGG LECACPALLE YARTCAQEGM VLYGWTDHSA CSPVCPAGME  
 301 YRQCVSPCAR TCQSLHINEM CQERCVDGCS CPEGQLLDEG LCVESTECPC  
 351 VHS GKRYPPG TSLSRDCNTC ICRNSQWICS NEECPGECIV TGQSHFKSFD  
 401 NRYFTFSGIC QYLLARDCQD HSFSIVIETV QCADDRDAVC TRSVTVRLPG  
 451 LHNSLVKCLKH GAGVAMDGQD IQLPLLKGD L RIQHTVTASV RLSYGEDLQM  
 501 DWDGRGRLLV KLS PVYAGKT CGLCGNYNGN QGDDFLTPSG LAEPRVEDFG  
 551 NAWKLGDCQ DLQKQHS DPC ALNPRMTRFS EEACAVLTSP TFEACHRAVS  
 601 PLPYLRNCRY DVCSCSDGRE CLCGALASYA AACAGRGVRV AWREPGRCEL  
 651 NCPKGQVYLQ CGTPCNLTCR SLSYPDEECN EACLEGCFCP PGLYMDERGD  
 701 CVPKAQCPCY YDGEIFQPED IFSDHHTMCY CEDGFMHCTM SGVPGSLLPD  
 751 AVLSSPLSHR SKRSLSCRPP MVKLVCPADN LRAEGLECTK TCQNYDLECM  
 [0716] 801 SMGCVSGCLC PPGMVRHENR CVALERCPCF HQGKEYAPGE TVKIGCNTCV  
 851 CRDRKWNCTD HVCDATCSTI GMAHYLTFDG LKYLFPGECQ YVLVQDYCGS  
 901 NPGTFRILVG NKGCSHPSVK CKKRVTILVE GGEIELFDGE VNVKRPMKDE  
 951 THFEVVESGR YIILLGKAL SVVWDRHLSI SVVLKQTYQE KVCGLCGNFD  
 1001 GIQNNDLTSS NLQVEEDPVD FGNSWKVSSQ CADTRKVPLD SSPATCHNNI  
 1051 MKQTMVDSSC RILTSDFVQD CNKLVDP EPY LDVCIYDTCS CESIGDCAAF  
 1101 CDTIAAYAHV CAQH GKVV TW RTATLCPQSC EERNLRENGY EAEWRYNSCA  
 1151 PACQVTCQHP EPLACPVQCV EGCHAHCPPG KILDELLQTC VDPEDCPVCE  
 1201 VAGRRFASGK KVTLNPSDPE HCQICHCDVV NLTCEACQEP ISGAPTESA  
 1251 TPESGPGSEP ATSGSETPGT SESATPESGP GSEPATSGSE TPGTSESATP  
 1301 ESGPGTSTEP SEGSAPGSPA GSPTSTEEGT SESATPESGP GSEPATSGSE  
 1351 TPGTSESATP ESGPGSPAGS PTSTEEGSPA GSPTSTEEGA **SISDKNTGDY**  
 1401 **YEDSYEDISA YLLSKNNAIE PRSFS**DKTHT CPPCPAPELL GGPSVFLFPP  
 1451 KPKDTLMISR TPEVTCVVVD VSHEDPEVKF NWYVDGVEVH NAKTKPREEQ  
  
 1501 YNSTYRVVSV LTVLHQDWLN GKEYKCKVSN KALPAPIEKT ISKAKGQPRE  
 1551 PQVYTLPPSR DELTKNQVSL TCLVKGFYPS DIAVEWESNG QPENNYKTTP  
 1601 PVLDS DGSFF LYSKLTVDKS RWQQGNV FSC SVMHEALHNH YTQKSLSLSP  
 1651 GK\*

[0717] pSYN VWF062核苷酸序列(编码在接头中不具有凝血酶位点的VWF D' D3-Fc) (SEQ ID NO:198)

1 ATGATTCCTG CCAGATTTGC CGGGGTGCTG CTTGCTCTGG CCCTCATTTT  
 51 GCCAGGGACC CTTTGTGCAG AAGGAACCTG CGGCAGGTCA TCCACGGCCC  
 101 GATGCAGCCT TTTCGGAAGT GACTTCGTCA ACACCTTTGA TGGGAGCATG  
 151 TACAGCTTTG CGGGATACTG CAGTTACCTC CTGGCAGGGG GCTGCCAGAA  
 201 ACGCTCCTTC TCGATTATTG GGGACTTCCA GAATGGCAAG AGAGTGAGCC  
 251 TCTCCGTGTA TCTTGGGGAA TTTTTTGACA TCCATTTGTT TGTCAATGGT  
 301 ACCGTGACAC AGGGGGACCA AAGAGTCTCC ATGCCCTATG CCTCCAAAGG  
 351 GCTGTATCTA GAAACTGAGG CTGGGTACTA CAAGCTGTCC GGTGAGGCCCT  
 401 ATGGCTTTGT GGCCAGGATC GATGGCAGCG GCAACTTTCA AGTCCTGCTG  
 451 TCAGACAGAT ACTTCAACAA GACCTGCGGG CTGTGTGGCA ACTTTAACAT  
 501 CTTTGTCTGAA GATGACTTTA TGACCCAAGA AGGGACCTTG ACCTCGGACC  
 551 CTTATGACTT TGCCAACTCA TGGGCTCTGA GCAGTGGAGA ACAGTGGTGT  
 601 GAACGGGCAT CTCTCCCGAG CAGCTCATGC AACATCTCCT CTGGGGAAAT  
 651 GCAGAAGGGC CTGTGGGAGC AGTGCCAGCT TCTGAAGAGC ACCTCGGTGT  
 701 TTGCCCCTG CCACCCTCTG GTGGACCCCG AGCCTTTTGT GGCCCTGTGT  
 751 GAGAAGACTT TGTGTGAGTG TGCTGGGGGG CTGGAGTGC GCTGCCCTGC  
 801 CCTCCTGGAG TACGCCCGGA CCTGTGCCCA GGAGGGAAATG GTGCTGTACG  
 851 GCTGGACCGA CCACAGCGCG TGCAGCCAG TGTGCCCTGC TGGTATGGAG  
 901 TATAGGCAGT GTGTGTCCCC TTGCGCCAGG ACCTGCCAGA GCCTGCACAT  
 951 CAATGAAATG TGTCAGGAGC GATGCGTGGG TGGCTGCAGC TGCCCTGAGG  
 1001 GACAGCTCCT GGATGAAGGC CTCTGCGTGG AGAGCACCGA GTGTCCCTGC  
 1051 GTGCATTCCG GAAAGCGCTA CCCTCCCGGC ACCTCCCTCT CTCGAGACTG  
 1101 CAACACCTGC ATTTGCCGAA ACAGCCAGTG GATCTGCAGC AATGAAGAAT  
 1151 GTCCAGGGGA GTGCCTTGTC ACTGGTCAAT CCCACTTCAA GAGCTTTGAC  
 1201 AACAGATACT TCACCTTCAG TGGGATCTGC CAGTACCTGC TGGCCCGGGA  
 [0718] 1251 TTGCCAGGAC CACTCCTTCT CCATTGTCAT TGAGACTGTC CAGTGTGCTG  
 1301 ATGACCGCGA CGCTGTGTGC ACCCGCTCCG TCACCGTCCG GCTGCCTGGC  
 1351 CTGCACAACA GCCTTGTTGAA ACTGAAGCAT GGGGCAGGAG TTGCCATGGA  
 1401 TGGCCAGGAC ATCCAGCTCC CCCTCCTGAA AGGTGACCTC CGCATCCAGC  
 1451 ATACAGTGAC GGCTCCCGTG CGCCTCAGCT ACGGGGAGGA CCTGCAGATG  
 1501 GACTGGGATG GCCGCGGGAG GCTGCTGGTG AAGCTGTCCC CCGTCTATGC  
 1551 CGGGAAGACC TCGGCCTGT GTGGGAATTA CAATGGCAAC CAGGGCGACG  
 1601 ACTTCCTTAC CCCCTCTGGG CTGGCGGAGC CCCGGGTGGA GGACTTCGGG  
 1651 AACGCCTGGA AGCTGCACGG GGACTGCCAG GACCTGCAGA AGCAGCACAG  
 1701 CGATCCCTGC GCCCTCAACC CGCGCATGAC CAGGTTCTCC GAGGAGCGCT  
 1751 GCGCGTCCCT GACGTCCCC ACATTCGAGG CCTGCCATCG TGCCGTGAGC  
 1801 CCGCTGCCCT ACCTGCGGAA CTGCCGCTAC GACGTGTGCT CCTGCTCGGA  
 1851 CGGCCGCGAG TGCTGTGCG GCGCCCTGGC CAGCTATGCC GCGGCCTGCG  
 1901 CGGGGAGAGG CGTGCGCGTC GCGTGGCGCG AGCCAGGCCG CTGTGAGCTG  
 1951 AACTGCCC GA AAGGCCAGGT GTACCTGCAG TGCGGGACCC CTGCAACCT  
 2001 GACCTGCCGC TCTCTCTCT ACCCGGATGA GGAATGCAAT GAGGCCTGCC  
 2051 TGGAGGGCTG CTCTGCCCC CCAGGGCTCT ACATGGATGA GAGGGGGGAC  
 2101 TGCGTGCCCA AGGCCAGTG CCCCTGTTAC TATGACGGTG AGATCTTCCA  
 2151 GCCAGAAGAC ATCTTCTCAG ACCATCACAC CATGTGCTAC TGTGAGGATG  
 2201 GCTTCATGCA CTGTACCATG AGTGGAGTCC CCGGAAGCTT GCTGCCTGAC  
 2251 GCTGTCCCTCA GCAGTCCCCT GTCTCATCGC AGCAAAAGGA GCCTATCCTG  
 2301 TCGGCCCCCC ATGGTCAAGC TGGTGTGTCC CGCTGACAAC CTGCGGGCTG  
 2351 AAGGGCTCGA GTGTACCAA ACGTGCCAGA ACTATGACCT GGAGTGCATG  
 2401 AGCATGGGCT GTGTCTCTGG CTGCCTCTGC CCCCCGGGCA TGGTCCGGCA  
 2451 TGAGAACAGA TGTGTGGCC TGGAAAGGTG TCCCTGCTTC CATCAGGGCA  
 2501 AGGAGTATGC CCCTGGAGAA ACAGTGAAGA TTGGCTGCAA CACTTGTGTC

2551 TGTCGGGACC GGAAGTGGA CTGCACAGAC CATGTGTGTG ATGCCACGTG  
 2601 CTCCACGATC GGCATGGCCC ACTACCTCAC CTTGACGGG CTCAAATACC  
 2651 TGTTCCCCGG GGAGTGCCAG TACGTTCTGG TGCAGGATTA CTGCGGCAGT  
 2701 AACCCCTGGGA CCTTTCGGAT CCTAGTGGGG AATAAGGGAT GCAGCCACCC  
 2751 CTCAGTGAAA TGCAAGAAAC GGGTCACCAT CCTGGTGGAG GGAGGAGAGA  
 2801 TTGAGCTGTT TGACGGGGAG GTGAATGTGA AGAGGCCCAT GAAGGATGAG  
 2851 ACTCACTTTG AGGTGGTGGG GTCTGGCCGG TACATCATTC TGCTGCTGGG  
 2901 CAAAGCCCTC TCCGTGGTCT GGGACCGCCA CCTGAGCATC TCCGTGGTCC  
 2951 TGAAGCAGAC ATACCAGGAG AAAGTGTGTG GCCTGTGTGG GAATTTTGAT  
 3001 GGCATCCAGA ACAATGACCT CACCAGCAGC AACCTCCAAG TGGAGGAAGA  
 3051 CCCTGTGGAC TTTGGGAACT CCTGGAAAGT GAGCTCGCAG TGTGCTGACA  
 3101 CCAGAAAAGT GCCTCTGGAC TCATCCCCTG CCACCTGCCA TAACAACATC  
 3151 ATGAAGCAGA CGATGGTGGG TTCCTCCTGT AGAATCCTTA CCAGTGACGT  
 3201 CTTCCAGGAC TGCAACAAGC TGGTGGACCC CGAGCCATAT CTGGATGTCT  
 3251 GCATTTACGA CACCTGCTCC TGTGAGTCCA TTGGGGACTG CGCCGCATTC  
 3301 TGCAGACCA TTGCTGCCTA TGCCCACGTG TGTGCCCAGC ATGGCAAGGT  
 3351 GGTGACCTGG AGGACGGCCA CATTGTGCC CCAGAGCTGC GAGGAGAGGA  
 3401 ATCTCCGGGA GAACGGGTAT GAGGCTGAGT GGCCTATAA CAGCTGTGCA  
 3451 CCTGCCCTGTC AAGTCACGTG TCAGCACCCCT GAGCCACTGG CCTGCCCTGT  
 3501 GCAGTGTGTG GAGGGCTGCC ATGCCCACTG CCCTCCAGGG AAAATCCTGG  
 3551 ATGAGCTTTT GCAGACCTGC GTTGACCCTG AAGACTGTCC AGTGTGTGAG  
 3601 GTGGCTGGCC GCGTTTTGTC CTCAGGAAAG AAAGTCACCT TGAATCCAG  
 3651 TGACCCTGAG CACTGCCAGA TTTGCCACTG TGATGTTGTC AACCTCACCT  
 [0719] 3701 GTGAAGCCTG CCAGGAGCCG ATATCGGGCG CGCCAACATC AGAGAGCGCC  
 3751 ACCCCTGAAA GTGGTCCCGG GAGCGAGCCA GCCACATCTG GGTCGGAAAC  
 3801 GCCAGGCACA AGTGAGTCTG CAACTCCCGA GTCCGGACCT GGCTCCGAGC  
 3851 CTGCCACTAG CGGCTCCGAG ACTCCGGGAA CTTCCGAGAG CGCTACACCA  
 3901 GAAAGCGGAC CCGGAACCAG TACCGAACCT AGCGAGGGCT CTGCTCCGGG  
 3951 CAGCCCAGCC GGCTCTCCTA CATCCACGGA GGAGGGCACT TCCGAATCCG  
 4001 CCACCCCGGA GTCAGGGCCA GGATCTGAAC CCGCTACCTC AGGCAGTGAG  
 4051 ACGCCAGGAA CGAGCGAGTC CGCTACACCG GAGAGTGGC CAGGGAGCCC  
 4101 TGCTGATCT CCTACGTCCA CTGAGGAAGG GTCACCAGCG GGCTCGCCCA  
 4151 CCAGCACTGA AGAAGGTGCC TCGAGCGACA AAACCTCACAC ATGCCACCG  
 4201 TGCCCAGCTC CAGAATCCT GGGCGGACCG TCAGTCTTCC TCTTCCCCC  
 4251 AAAACCCAAG GACACCCTCA TGATCTCCCG GACCCCTGAG GTCACATGCG  
 4301 TGGTGGTGGG CGTGAGCCAC GAAGACCCTG AGGTCAAGTT CAACTGGTAC  
 4351 GTGGACGGCG TGGAGGTGCA TAATGCCAAG ACAAAGCCGC GGGAGGAGCA  
 4401 GTACAACAGC ACGTACCCTG TGGTCAGCGT CCTCACCGTC CTGCACCAGG  
 4451 ACTGGCTGAA TGGCAAGGAG TACAAGTGCA AGGTCTCCAA CAAAGCCCTC  
 4501 CCAGCCCCCA TCGAGAAAAC CATCTCCAA GCCAAAGGGC AGCCCCGAGA  
 4551 ACCACAGGTG TACACCCTGC CCCCATCCCG GGATGAGCTG ACCAAGAACC  
 4601 AGGTCAGCCT GACCTGCCTG GTCAAAGGCT TCTATCCCAG CGACATCGCC  
 4651 GTGGAGTGGG AGAGCAATGG GCAGCCGGAG ACAAACACTACA AGACCACGCC  
 4701 TCCCGTGTG GACTCCGACG GCTCCTTCTT CCTCTACAGC AAGCTCACCG  
 4751 TGGACAAGAG CAGGTGGCAG CAGGGGAACG TCTTCTCATG CTCCGTGATG  
 4801 CATGAGGCTC TGCACAACCA CTACACGCAG AAGAGCCTCT CCCTGTCTCC  
 4851 GGGTAAATGA

[0720] pSYN VWF062蛋白质序列(在接头中不具有凝血酶位点的VWF D' D3-Fc) (SEQ ID NO:199)

1 MIPARFAGVL LALALILPGT LCAEGTRGRS STARCSLFGS DFNNTFDGSM  
 51 YSFAGYCSYL LAGGCQKRSF SIIIGDFQNGK RVSLSVYLGE FFDIHLFVNG  
 101 TVTQGDQRVS MPYASKGLYL ETEAGYYKLS GEAYGFVARI DGSGNFQVLL  
 151 SDRYFNKTCG LCGNFNIFAE DDFMTQEGTL TSDPYDFANS WALSSGEQWC  
 201 ERASPPSSSC NISSGEMQKG LWEQCQLLKS TSVFARCHPL VDPEPFVALC  
 251 EKTLCCECAGG LECACPALLE YARTCAQEGM VLYGWTDHSA CSPVCPAGME  
 301 YRQCVSPCAR TCQSLHINEM CQERCVDGCS CPEGQLLDEG LCVESTTECPC  
  
 351 VHSGKRYPPG TSLSRDCNTC ICRNSQWICS NEECPGECV TGQSHFKSFD  
 401 NRYFTFSGIC QYLLARDCQD HSFSIVIVTV QCADDRDAVC TRSVTVRLPG  
 451 LHNSLVKCLKH GAGVAMDGQD IQLPLLKGD LRIQHTVTASV RLSYGEDLQM  
 501 DWDGRGRLLV KLSPVYAGKT CGLCGNYNGN QGDDFLTPSG LAEPRVEDFG  
 551 NAWKLGDCQ DLQKQHS DPC ALNPRMTRFS EEACAVLTSP TFEACHRAVS  
 601 PLPYLRNCRY DVCSCSDGRE CLCGALASYA AACAGRGVRV AWREPGRCEL  
 651 NCPKGQVYLQ CGTPCNLTCR SLSYPDEECN EACLEGCFCP PGLYMDERGD  
 701 CVPKAQCPCY YDGEIFQPED IFSDHHTMCY CEDGFMHCTM SGVPGSLLPD  
 751 AVLSSPLSHR SKRSLSCRPP MVKLVCPADN LRAEGLECTK TCQNYDLECM  
 [0721] 801 SMGCVSGCLC PPGMVRHENR CVALERCPCF HQGKEYAPGE TVKIGCNTCV  
 851 CRDRKWNCTD HVCDATCSTI GMAHYLTFDG LKYLFPGECQ YVLVQDYCGS  
 901 NPGTFRILVG NKGCSHPSVK CKKRVTILVE GGEIELFDGE VNVKRPMKDE  
 951 THFEVVESEGR YIILLGKAL SVVWDRHLSI SVVLKQTYQE KVCGLCGNFD  
 1001 GIQNNDLTSS NLQVEEDPVD FGNSWKVSSQ CADTRKVPLD SSPATCHNNI  
 1051 MKQTMVDSSC RILTSDFVQD CNKLVDPPEY LDVCIYDTCES SIGDCAAF  
 1101 CDTIAAYAHV CAQHGVVVTW RTATLCPQSC EERNLRENGY EAEWRYNSCA  
 1151 PACQVTCQHP EPLACPVQCV EGCHAHCPPG KILDELLQTC VDPEDCPVCE  
 1201 VAGRRFASGK KVTLNPSDPE HCQICHCDVV NLTCEACQEP ISGAPTESA  
 1251 TPESGPGSEP ATSGSETPGT SESATPESGP GSEPATSGSE TPGTSESATP  
 1301 ESGPGTSTEP SEGSAPGSPA GSPTSTEEGT SESATPESGP GSEPATSGSE  
 1351 TPGTSESATP ESGPGSPAGS PTSTEEGSPA GSPTSTEEGA SSDKHTCPP  
 1401 CPAPELLGGP SVFLFPPKPK DTLMISRTPE VTCVVVDVSH EDPEVKFNWY  
 1451 VDGVEVHNAK TKPREEQYNS TYRVVSVLTV LHQDWLNGKE YKCKVSNKAL  
 1501 PAPIEKTISK AKGQPREPQV YTLPPSRDEL TKNQVSLTCL VKGFYPSDIA  
 1551 VEWESNGQPE NNYKTTTPVL DSDGSFFLYS KLTVDKSRWQ QGNVFSVSM  
 1601 HEALHNHYTQ KSLSLSPGK\*

[0722] pSYN VWF073核苷酸序列- (编码VWFD1D2D' D3-144 AE XTEN-FVIII截短a2凝血酶位点-Fc) (SEQ ID NO:174)

1 ATGATTCCTG CCAGATTTGC CGGGGTGCTG CTTGCTCTGG CCCTCATTTT  
51 GCCAGGGACC CTTTGTGCAG AAGGAACTCG CGGCAGGTCA TCCACGGCCC  
101 GATGCAGCCT TTTCGGAAGT GACTTCGTCA ACACCTTTGA TGGGAGCATG  
151 TACAGCTTTG CGGGATACTG CAGTTACCTC CTGGCAGGGG GCTGCCAGAA  
201 ACGCTCCTTC TCGATTATTG GGGACTTCCA GAATGGCAAG AGAGTGAGCC  
251 TCTCCGTGTA TCTTGGGGAA TTTTTTGACA TCCATTTGTT TGTCAATGGT  
301 ACCGTGACAC AGGGGGACCA AAGAGTCTCC ATGCCCTATG CCTCCAAAGG  
351 GCTGTATCTA GAAACTGAGG CTGGGTACTA CAAGCTGTCC GGTGAGGCCCT  
401 ATGGCTTTGT GGCCAGGATC GATGGCAGCG GCAACTTTCA AGTCCTGCTG  
451 TCAGACAGAT ACTTCAACAA GACCTGCGGG CTGTGTGGCA ACTTTAACAT  
501 CTTTGTCTGAA GATGACTTTA TGACCCAAGA AGGGACCTTG ACCTCGGACC  
551 CTTATGACTT TGCCAACTCA TGGGCTCTGA GCAGTGGAGA ACAGTGGTGT  
601 GAACGGGCAT CTCTCCCGAG CAGCTCATGC AACATCTCCT CTGGGGAAAT  
[0723] 651 GCAGAAGGGC CTGTGGGAGC AGTGCCAGCT TCTGAAGAGC ACCTCGGTGT  
701 TTGCCCCTG CCACCCTCTG GTGGACCCCG AGCCTTTTGT GGCCCTGTGT  
751 GAGAAGACTT TGTGTGAGTG TGCTGGGGGG CTGGAGTGC GCTGCCCTGC  
801 CCTCCTGGAG TACGCCCGGA CCTGTGCCCA GGAGGGAAATG GTGCTGTACG  
851 GCTGGACCGA CCACAGCGCG TGCAGCCAG TGTGCCCTGC TGGTATGGAG  
901 TATAGGCAGT GTGTGTCCCC TTGCGCCAGG ACCTGCCAGA GCCTGCACAT  
951 CAATGAAATG TGTCAGGAGC GATGCGTGGA TGGCTGCAGC TGCCCTGAGG  
1001 GACAGCTCCT GGATGAAGGC CTCTGCGTGG AGAGCACCGA GTGTCCCTGC  
1051 GTGCATTCCG GAAAGCGCTA CCCTCCCGG ACCTCCCTCT CTCGAGACTG  
1101 CAACACCTGC ATTTGCCGAA ACAGCCAGTG GATCTGCAGC AATGAAGAAT  
1151 GTCCAGGGGA GTGCCTTGTC ACTGGTCAAT CCCACTTCAA GAGCTTTGAC  
1201 AACAGATACT TCACCTTCAG TGGGATCTGC CAGTACCTGC TGGCCCGGGA  
1251 TTGCCAGGAC CACTCCTTCT CCATTGTCAT TGAGACTGTC CAGTGTGCTG  
1301 ATGACC CGA CGCTGTGTGC ACCCGCTCCG TCACCGTCCG GCTGCCTGGC

1351 CTGCACAACA GCCTTGTGAA ACTGAAGCAT GGGGCAGGAG TTGCCATGGA  
1401 TGGCCAGGAC ATCCAGCTCC CCCTCCTGAA AGGTGACCTC CGCATCCAGC  
1451 ATACAGTGAC GGCCTCCGTG CGCCTCAGCT ACGGGGAGGA CCTGCAGATG  
1501 GACTGGGATG GCCGCGGGAG GCTGCTGGTG AAGCTGTCCC CCGTCTATGC  
1551 CGGGAAGACC TCGGCGCTGT GTGGGAATTA CAATGGCAAC CAGGGCGACG  
1601 ACTTCCTTAC CCCCTCTGGG CTGGCGGAGC CCCGGGTGGA GGACTTCGGG  
1651 AACGCCTGGA AGCTGCACGG GGACTGCCAG GACCTGCAGA AGCAGCACAG  
1701 CGATCCCTGC GCCCTCAACC CGCGCATGAC CAGGTTCTCC GAGGAGGCGT  
1751 GCGCGGTCTT GACGTCCCC ACATTCGAGG CCTGCCATCG TGCCGTCAGC  
1801 CCGCTGCCCT ACCTGCGGAA CTGCCGCTAC GACGTGTGCT CCTGCTCGGA  
1851 CGGCGCGGAG TGCCTGTGCG GCGCCCTGGC CAGCTATGCC GCGGCCCTGCG  
1901 CGGGGAGAGG CGTGCAGTGC GCGTGGCGCG AGCCAGGCCG CTGTGAGCTG  
1951 AACTGCCCGA AAGGCCAGGT GTACCTGCAG TGCGGGACCC CCTGCAACCT  
2001 GACCTGCCGC TCTCTCTCTT ACCCGGATGA GGAATGCAAT GAGGCCCTGCC  
2051 TGGAGGGCTG CTTCTGCCCC CCAGGGCTCT ACATGGATGA GAGGGGGGAC  
2101 TGCGTGCCCA AGGCCAGTG CCCCTGTTAC TATGACGGTG AGATCTTCCA  
2151 GCCAGAAGAC ATCTTCTCAG ACCATCACAC CATGTGCTAC TGTGAGGATG  
2201 GCTTCATGCA CTGTACCATG AGTGGAGTCC CCGGAAGCTT GCTGCCCTGAC  
2251 GCTGTCTCTA GCAGTCCCC GTCTCATCGC AGCAAAGGA GCCTATCCTG  
2301 TCGGCCCCCC ATGGTCAAGC TGGTGTGTCC CGCTGACAAC CTGCGGCTG  
2351 AAGGGCTCGA GTGTACAAA ACGTGCCAGA ACTATGACCT GGAGTGCATG  
2401 AGCATGGGCT GTGTCTCTGG CTGCCTCTGC CCCCCGGGCA TGGTCCGGCA  
2451 TGAGAACAGA TGTGTGGCCC TGGAAAGGTG TCCCTGCTTC CATCAGGGCA  
2501 AGGAGTATGC CCCTGGAGAA ACAGTGAAGA TTGGCTGCAA CACTTGTGTC  
2551 TGTGCGGACC GGAAGTGGAA CTGCACAGAC CATGTGTGTG ATGCCACGTG  
2601 CTCCACGATC GGCATGGCCC ACTACCTCAC CTTTCGACGGG CTCAAATACC  
2651 TGTTCCCCGG GGAGTGCCAG TACGTTCTGG TGCAGGATTA CTGCGGCAGT  
2701 AACCCTGGGA CCTTTCGGAT CCTAGTGGGG AATAAGGGAT GCAGCCACCC  
2751 CTCAGTGAAA TGCAAGAAAC GGGTCACCAT CCTGGTGGAG GGAGGAGAGA  
[0724] 2801 TTGAGCTGTT TGACGGGGAG GTGAATGTGA AGAGGCCCAT GAAGGATGAG  
2851 ACTCACTTTG AGGTGGTGGG GTCTGGCCGG TACATCATTG TGCTGCTGGG  
2901 CAAAGCCCTC TCCGTGGTCT GGGACCGCCA CCTGAGCATC TCCGTGGTCC  
2951 TGAAGCAGAC ATACCAGGAG AAAGTGTGTG GCCTGTGTGG GAATTTTGAT  
3001 GGCATCCAGA ACAATGACCT CACCAGCAGC AACCTCCAAG TGGAGGAAGA  
3051 CCCTGTGGAC TTTGGGAACT CCTGGAAAGT GAGCTCGCAG TGTGCTGACA  
3101 CCAGAAAAGT GCCTCTGGAC TCATCCCCTG CCACCTGCCA TAACAACATC  
3151 ATGAAGCAGA CGATGGTGGG TTCTCCTGT AGAATCCTTA CCAGTGACGT  
3201 CTTCCAGGAC TGCAACAAGC TGGTGGACCC CGAGCCATAT CTGGATGTCT  
3251 GCATTTACGA CACCTGCTCC TGTGAGTCCA TTGGGGACTG CGCCGCATTC  
3301 TGCGACACCA TTGCTGCCTA TGCCCACGTG TGTGCCAGC ATGGCAAGGT  
3351 GGTGACCTGG AGGACGGCCA CATTGTGCCC CCAGAGCTGC GAGGAGAGGA  
3401 ATCTCCGGGA GAACGGGTAT GAGGCTGAGT GGCCTATAA CAGCTGTGCA  
3451 CCTGCCTGTC AAGTCACGTG TCAGCACCTT GAGCCACTGG CCTGCCCTGT  
3501 GCAGTGTGTG GAGGGCTGCC ATGCCACTG CCCTCCAGGG AAAATCCTGG  
3551 ATGAGCTTTT GCAGACCTGC GTTGACCTTG AAGACTGTCC AGTGTGTGAG  
3601 GTGGCTGGCC GCGTTTTTGC CTCAGGAAAG AAAGTCACCT TGAATCCCAG  
3651 TGACCCTGAG CACTGCCAGA TTTGCCACTG TGATGTTGTC AACCTCACCT  
3701 GTGAAGCCTG CCAGGAGCCG GGCAGCCAA CATCAGAGAG CGCCACCCCT  
3751 GAAAGTGGTC CCGGGAGCGA GCCAGCCACA TCTGGGTGCG AAACGCCAGG  
3801 CACAAGTGAG TCTGCAACTC CCGAGTCCGG ACCTGGCTCC GAGCCTGCCA  
3851 CTAGCGGCTC CGAGACTCCG GGAACCTCCG AGAGCGCTAC ACCAGAAAGC  
3901 GGACCCGGAA CCAGTACCGA ACCTAGCGAG GGCTCTGCTC CGGGCAGCCC  
3951 AGCCGGCTCT CCTACATCCA CGGAGGAGGG CACTTCCGAA TCCGCCACCC  
4001 CGGAGTCAGG GCCAGGATCT GAACCCGCTA CCTCAGGCAG TGAGACGCCA  
4051 GGAACGAGCG AGTCCGCTAC ACCGGAGAGT GGGCCAGGGA GCCCTGCTGG  
4101 ATCTCCTACG TCCACTGAGG AAGGGTCACC AGCGGGCTCG CCCACCAGCA  
4151 CTGAAGAAGG TGCCTCGAGC GGCGGTGGAG GATCCGGTGG CGGGGGATCC  
4201 GGTGGCGGGG GATCCGGTGG CGGGGGATCC GGTGGCGGGG GATCCGGTGG  
4251 CGGGGGATCC ATTGAACCAA GAAGCTTCTC TGGCAGCGGA GGCGACAAA

4301 CTCACACATG CCCACCGTGC CCAGCTCCAG AACTCCTGGG CGGACCGTCA  
 4351 GTCTTCCTCT TCCCCCAA ACCCAAGGAC ACCCTCATGA TCTCCCGGAC  
 4401 CCCTGAGGTC ACATGCGTGG TGGTGGACGT GAGCCACGAA GACCCTGAGG  
 4451 TCAAGTTCAA CTGGTACGTG GACGGCGTGG AGGTGCATAA TGCCAAGACA  
 4501 AAGCCGCGGG AGGAGCAGTA CAACAGCACG TACCGTGTGG TCAGCGTCCT  
 4551 CACCGTCCCTG CACCAGGACT GGCTGAATGG CAAGGAGTAC AAGTGCAAGG  
 [0725] 4601 TCTCCAACAA AGCCCTCCCA GCCCCCATCG AGAAAACCAT CTCCAAAGCC  
 4651 AAAGGGCAGC CCCGAGAACC ACAGGTGTAC ACCCTGCCCC CATCCCGGGA  
 4701 TGAGCTGACC AAGAACCAGG TCAGCCTGAC CTGCCTGGTC AAAGGCTTCT  
 4751 ATCCAGCGA CATCGCCGTG GAGTGGGAGA GCAATGGGCA GCCGGAGAAC  
 4801 AACTACAAGA CCACGCCTCC CGTGTTGGAC TCCGACGGCT CTTTCTTCCT  
 4851 CTACAGCAAG CTCACCGTGG ACAAGAGCAG GTGGCAGCAG GGAACGTCT  
 4901 TCTCATGCTC CGTGATGCAT GAGGCTCTGC ACAACCACTA CACGCAGAAG  
 4951 AGCCTCTCCC TGTCTCCGGG TAAATGA

[0726] pSYN VWF073蛋白质序列 - (VWFD1D2D' D3-144 AE XTEN-截短a2凝血酶位点-Fc)  
 (SEQ ID NO:175)

1 MIPARFAGVL LALALILPGT LCAEGTRGRS STARCSLFGS DFNVTFDGSM  
 51 YSFAGYCSYL LAGGCQKRSF SIIGDFQNGK RVSLSVYLGE FFDIHLFVNG  
 101 TVTQGDQRV MPYASKGLYL ETEAGYYKLS GEAYGFVARI DGSGNFQVLL  
 151 SDRYFNKTCG LCGNFNIFAE DDFMTQEGTL TSDPYDFANS WALSSGEQWC  
 201 ERASPPSSC NISSGEMQKG LWEQCQLLKS TSVFARCHPL VDPEPFVALC  
 251 EKTLCCECAGG LECACPALLE YARTCAQEGM VLYGWDHSA CSPVCPAGME  
 301 YRQCVSPCAR TCQSLHINEM CQERCVDGCS CPEGQLLDEG LCVESTTEPC  
 351 VHSGKRYPPG TSLSRDCNTC ICRNSQWICS NEECPGECIV TGQSHFKSFD  
 401 NRYFTFSGIC QYLLARDCQD HSFSIVIVTV QCADDRDAVC TRSVTVRLPG  
 451 LHNSLVKCLKH GAGVAMDQD IQLPLLKGD LRIQHTVTASV RLSYGEDLQM  
 501 DWDGRGRLLV KLSVPYAGKT CGLCGNYNGN QGDDFLTPSG LAEPRVEDFG  
 551 NAWKLGDCQ DLQKQHS DPC ALNPRMTRFS EEACAVLTSP TFEACHRAVS  
 601 PLPYLRNCRY DVCSCSDGRE CLCGALASYA AACAGRGVRV AWREPGRCEL  
 651 NCPKGQVYLQ CGTPCNLTCR SLSYPDEECN EACLEGCFCP PGLYMDERGD  
 701 CVPKAQCPCY YDGEIFQPED IFSDHHTMCY CEDGFMHCTM SGVPGSLLPD  
 751 AVLSSPLSHR SKRSLSCRPP MVKLVCPADN LRAEGLECTK TCQNYDLECM  
 [0727] 801 SMGCVSGCLC PPGMVRHENR CVALERCPCF HQGKEYAPGE TVKIGCNTCV  
 851 CRDRKWNCTD HVCDATCSTI GMAHYLTFDG LKYLFPGECQ YVLVQDYCGS  
 901 NPGTFRILVG NKGCSHPSVK CKKRVTLVVE GGEIELFDGE VNVKRPMKDE  
 951 THFEVVESGR YIIILLGKAL SVVWDRHLSI SVVLKQTYQE KVCGLGNFND  
 1001 GIQNNDLTSS NLQVEEDPVD FGNSWKVSSQ CADTRKVPLD SSPATCHNNI  
 1051 MKQTMVDSSC RILTSDFVQD CNKLVDPPEY LDVCIYDTCS CESIGDCAAF  
 1101 CDTIAAYAHV CAQHGVVVTW RTATLCPQSC EERNLRENGY EAEWRYNSCA  
 1151 PACQVTCQHP EPLACPVQCV EGCHAHCPPG KILDELLQTC VDPEDCPVCE  
 1201 VAGRRFASGK KVTLNPSDPE HCQICHCDVV NLTCEACQEP GAPTSESATP  
 1251 ESGPGSEPAT SGSETPGTSE SATPESGPGS EPATSGSETP GTSESATPES  
 1301 GPGTSTEPSE GSAPGSPAGS PTSTEEGTSE SATPESGPGS EPATSGSETP  
 1351 GTSESATPES GPGSPAGSPT STEEGSPAGS PTSTEEGASS GGGGSGGGGS  
 1401 GGGGSGGGGS GGGGSGGGGS IEPRFSFGSG GDKTHTCPPC PAPELLGGPS  
 1451 VFLFPPKPKD TLMISRTPEV TCVVVDVSHE DPEVKFNWYV DGVEVHNAKT  
 1501 KPREEQYNST YRVSVLTVL HQDWLNGKEY KCKVSNKALP APIEKTISKA  
 1551 KGQPREPQVY TLPPSRDEL TKNQVSLTCLV KGFYPSDIAV EWESNGQPEN  
 1601 NYKTTPPVLD SDGSFFLYSK LTVDKSRWQQ GNVFSCSVMH EALHNHYTQK  
 1651 SLSLSPGK\*

[0728] 对特定实施方案的上文描述将如此充分揭示本发明的一般性特性以致他人可在不脱离本发明的一般性概念下,在不进行过度实验下,通过应用本领域的技能内的知识而易于针对各种应用来修改和/或改适所述特定实施方案。因此,基于本文呈现的教导和指

导,所述改适和修改意图在公开的实施方案的等效物的含义和范围内。应了解本文的措辞或术语是出于描述而非限制目的,因此本说明书的术语或措辞应由熟练技术人员根据教导和指导加以解释。

[0729] 本发明的其它实施方案将由考虑本文公开的本发明的说明书和实施而为本领域技术人员显而易见。意图说明书和实施例仅被视为具有示例性,其中本发明的真正范围和精神由以下权利要求指示。

[0730] 本文引用的所有专利和出版物都以引用的方式整体并入本文。



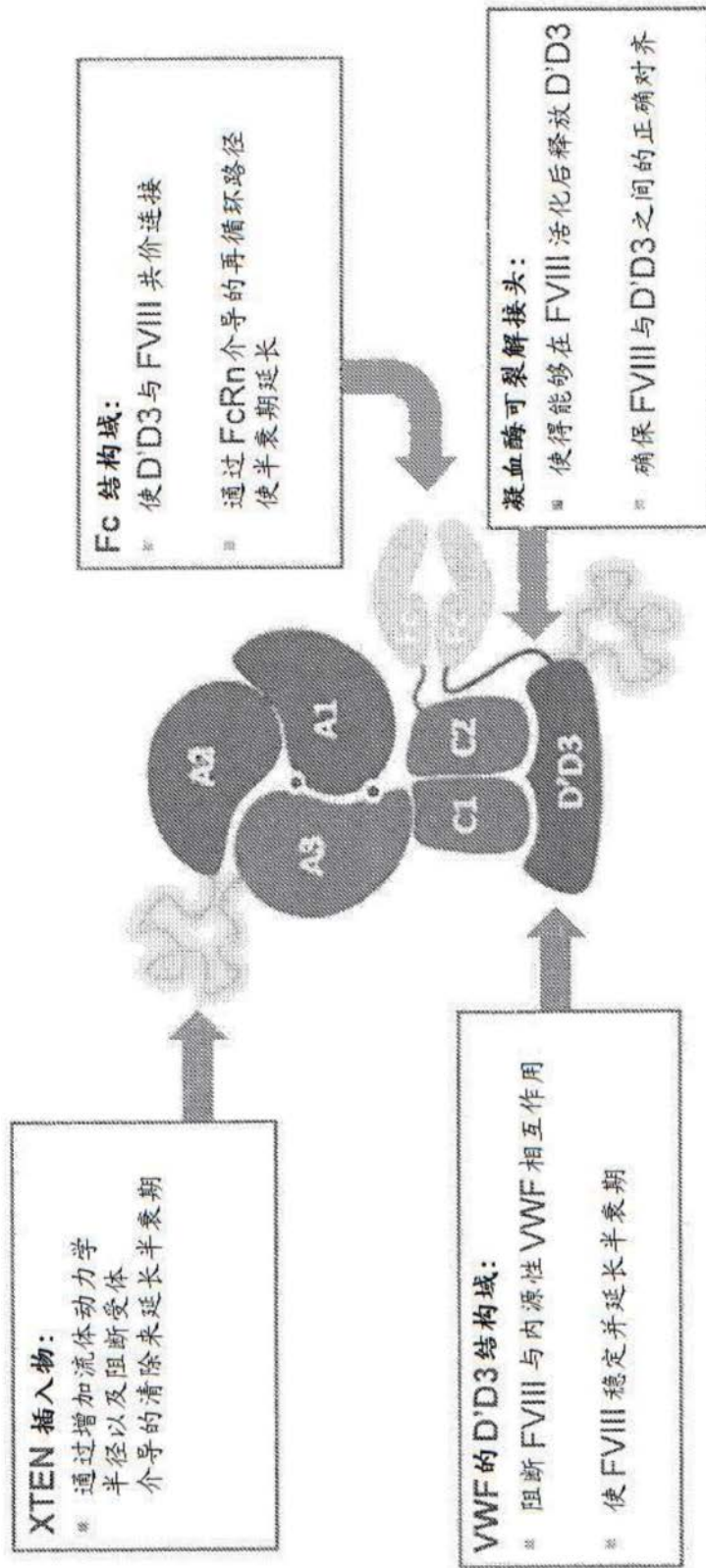


图1

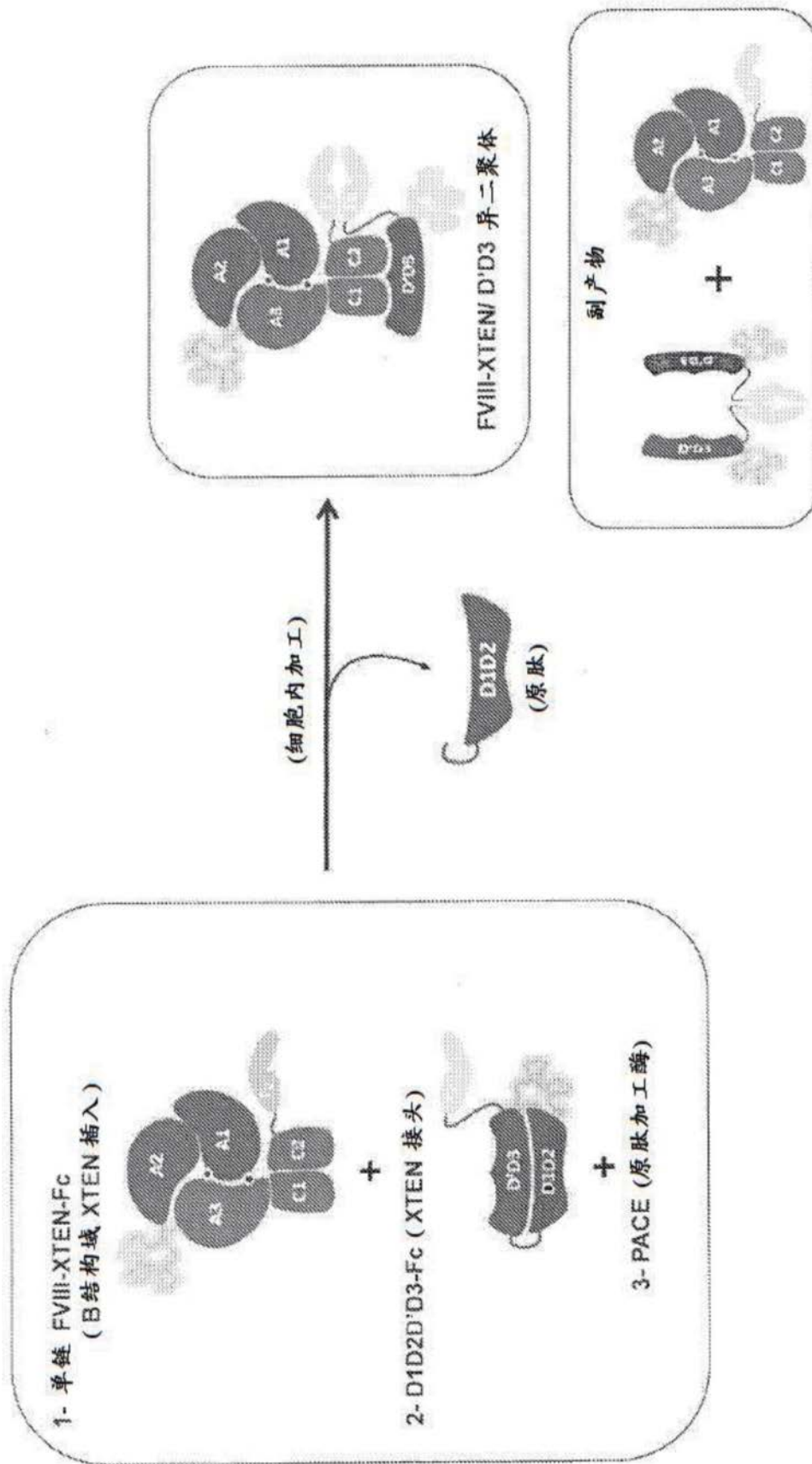


图2

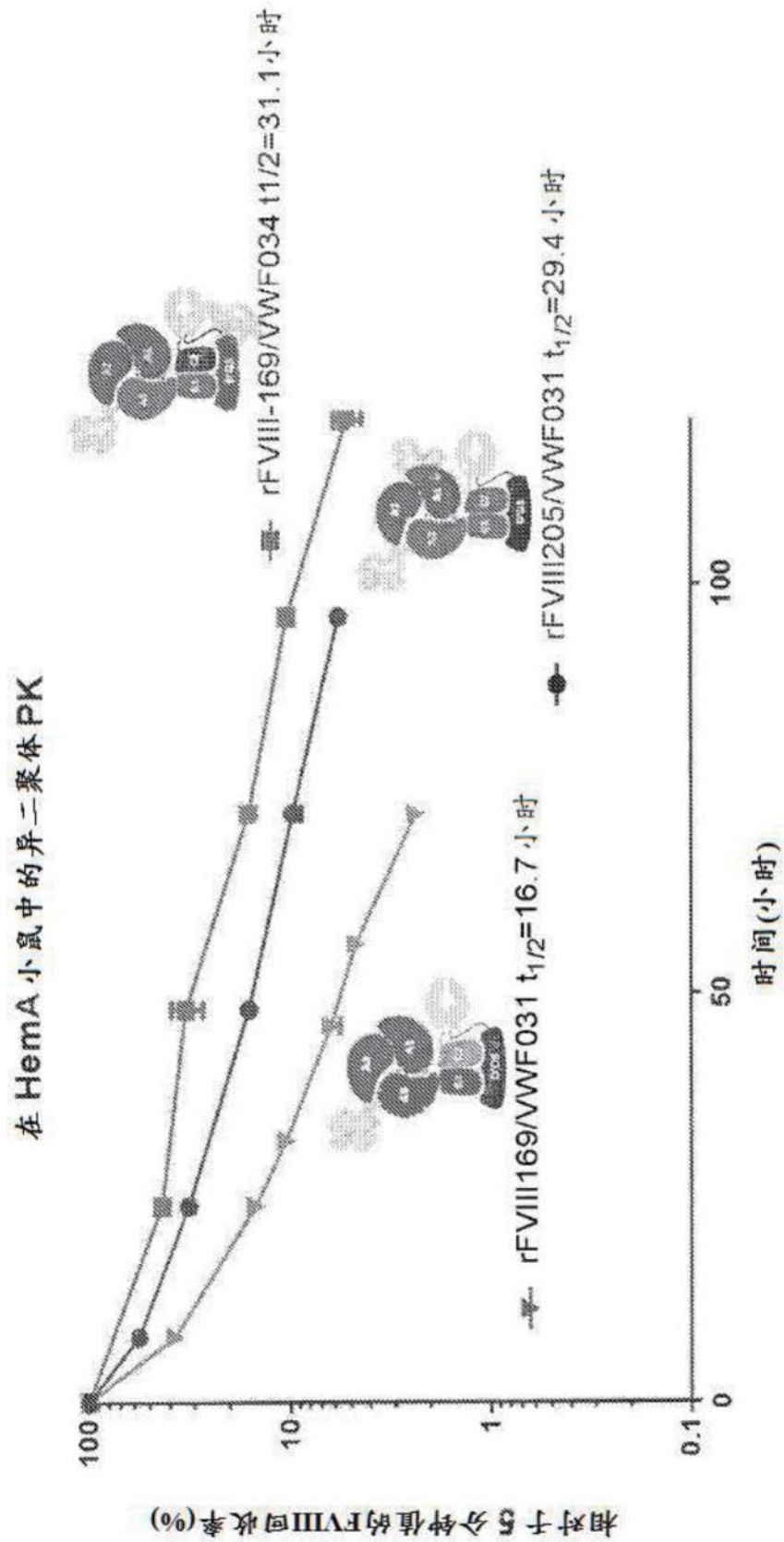


图3

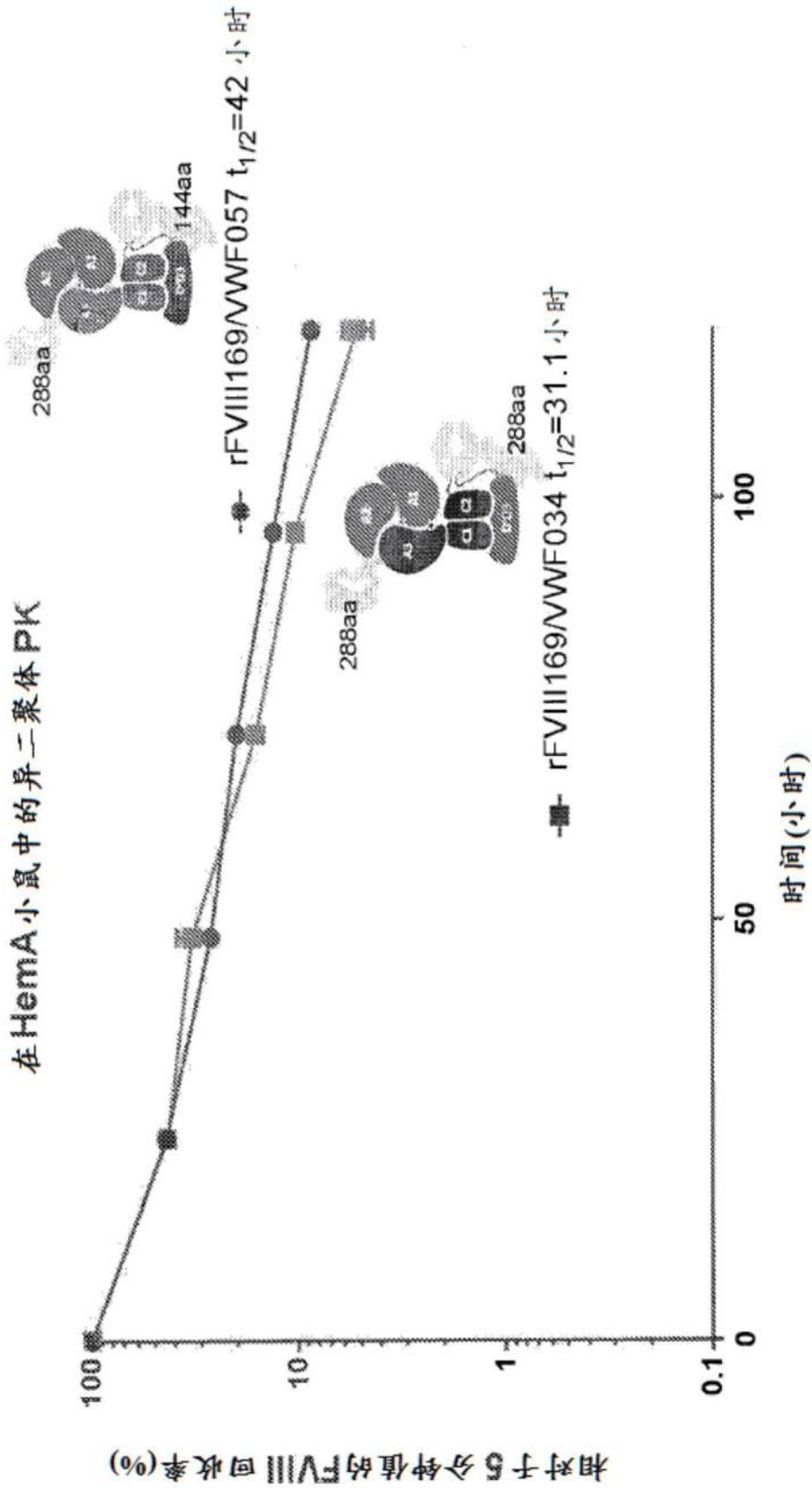


图4

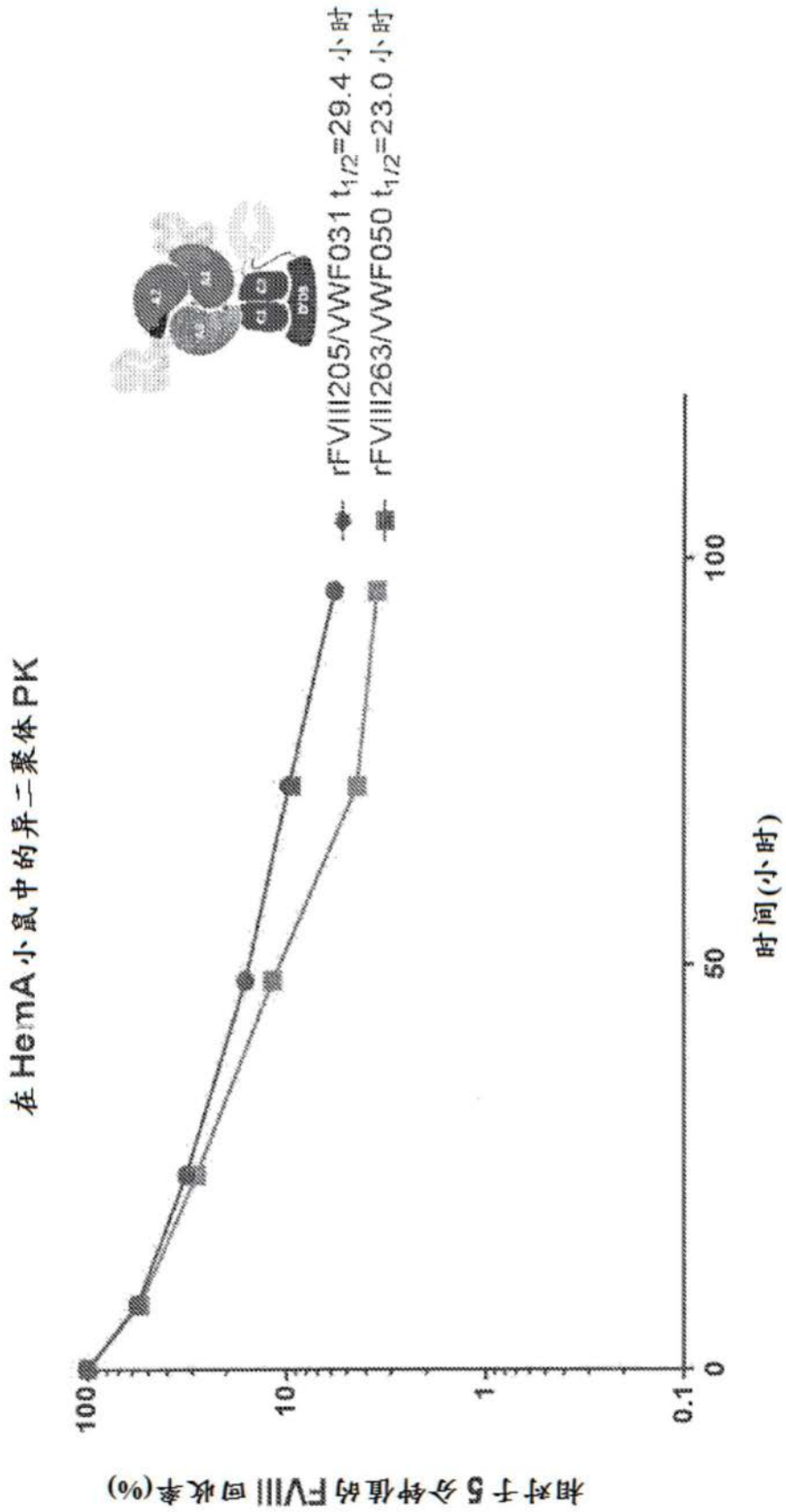


图5

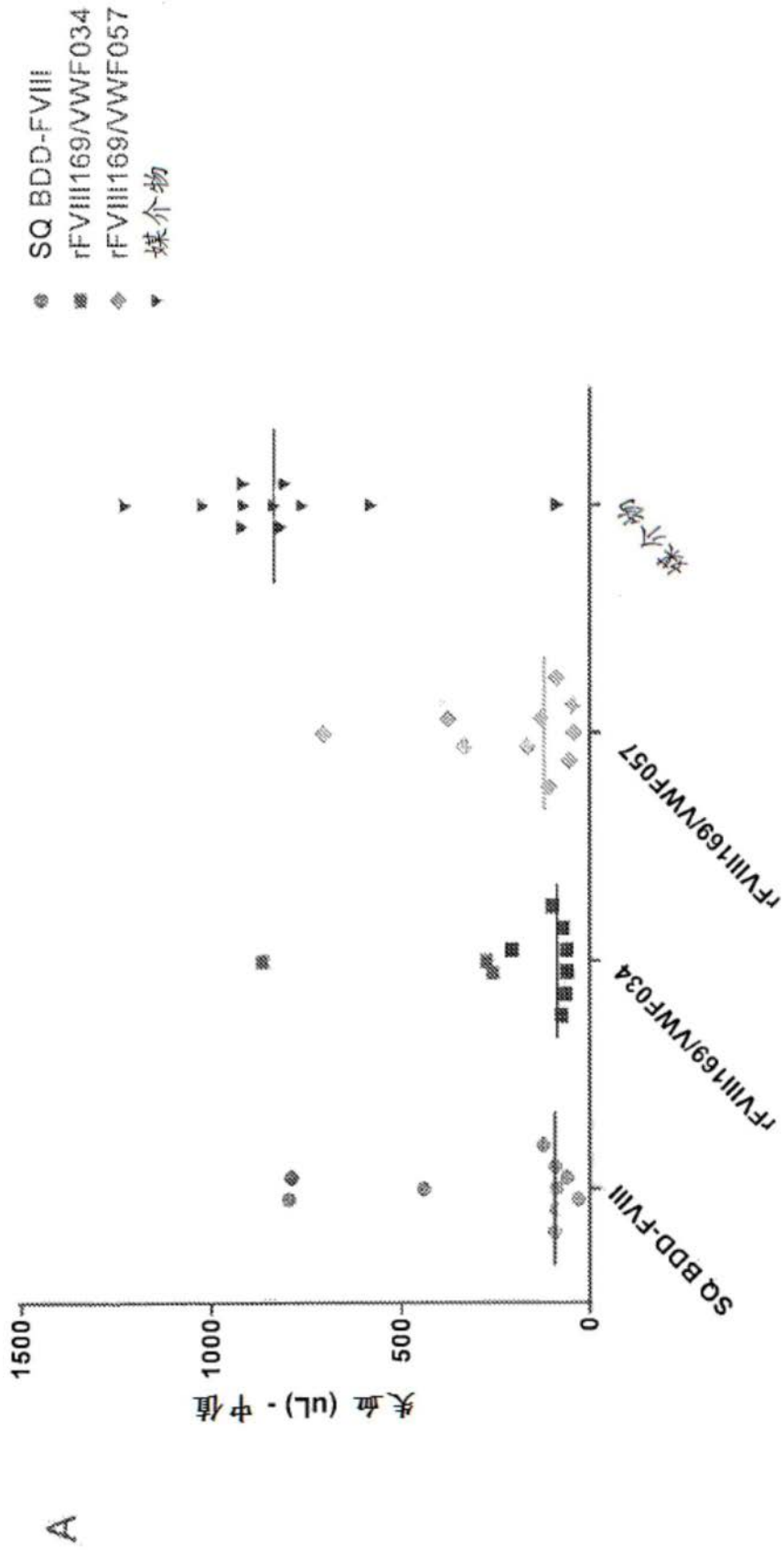


图6

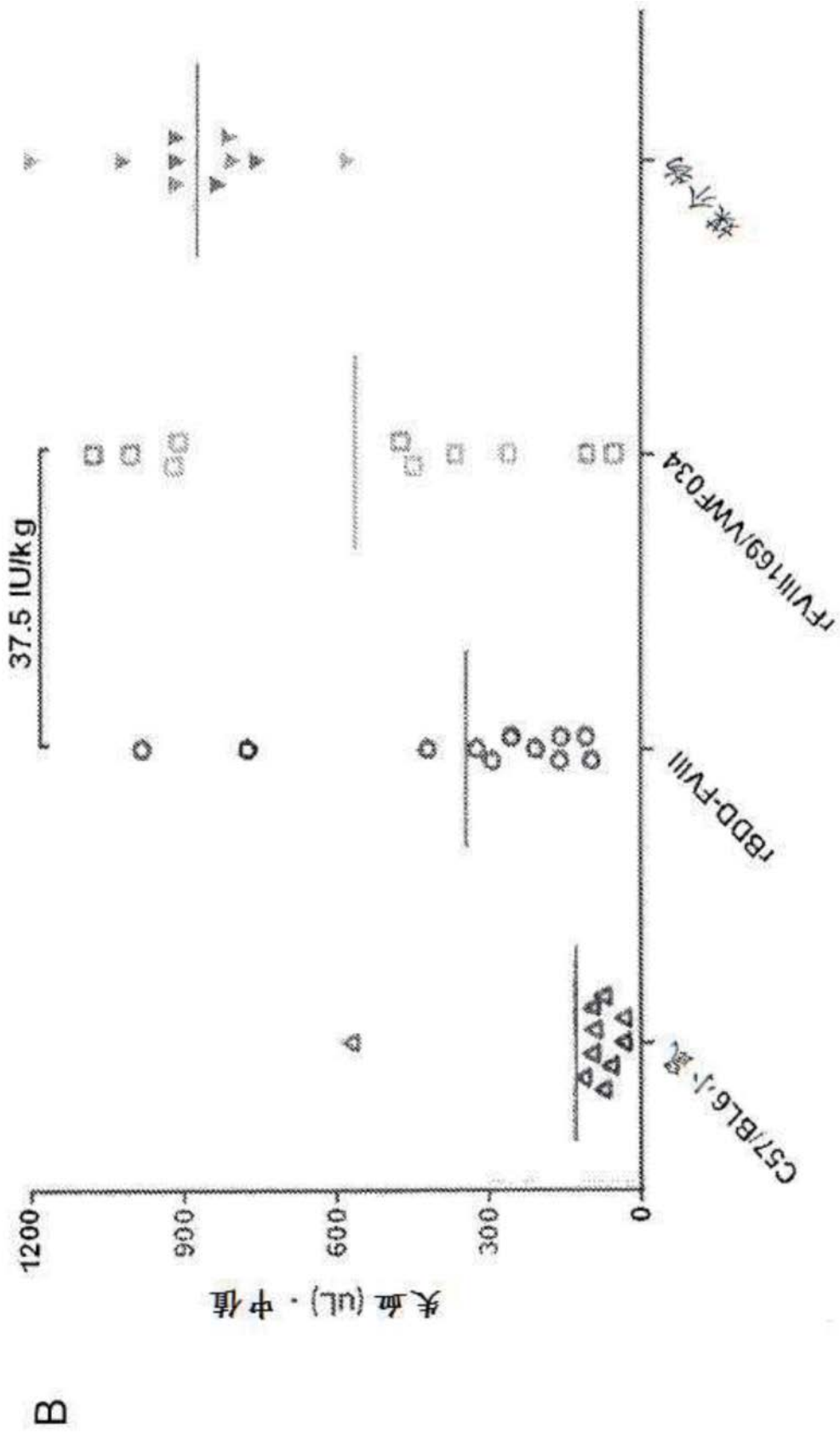


图6(续)

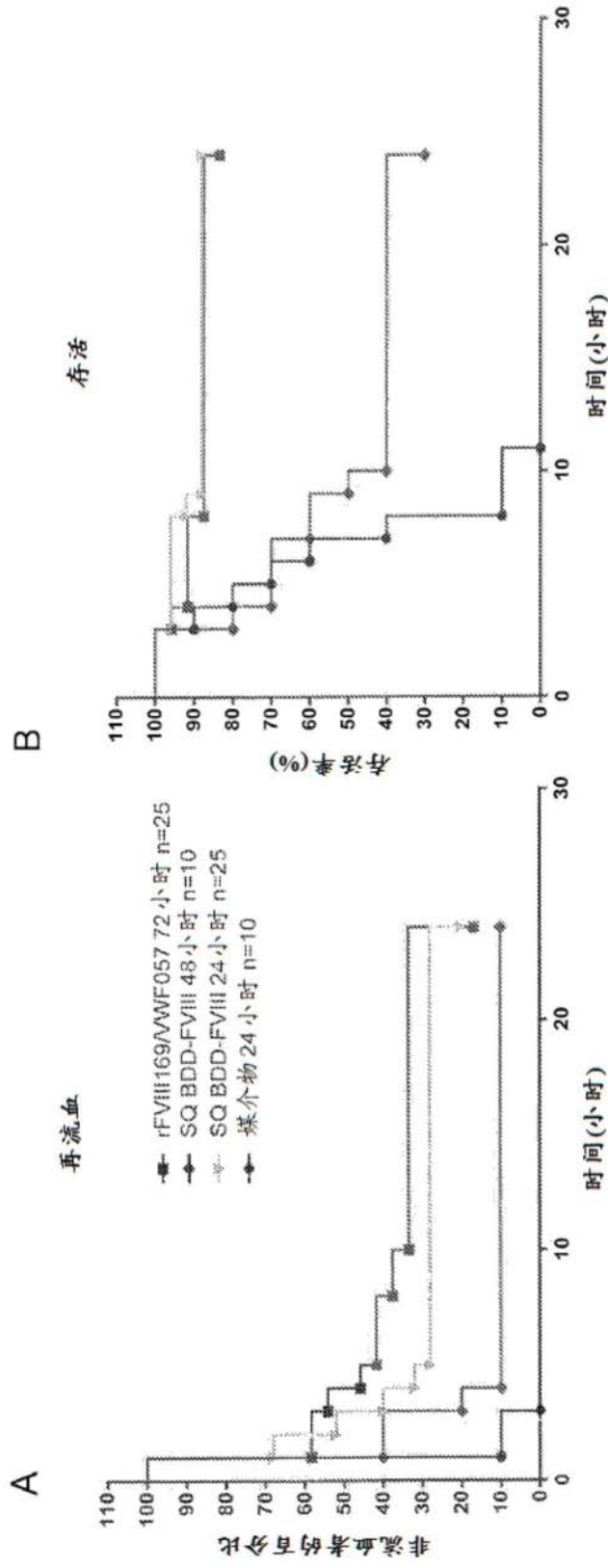


图7



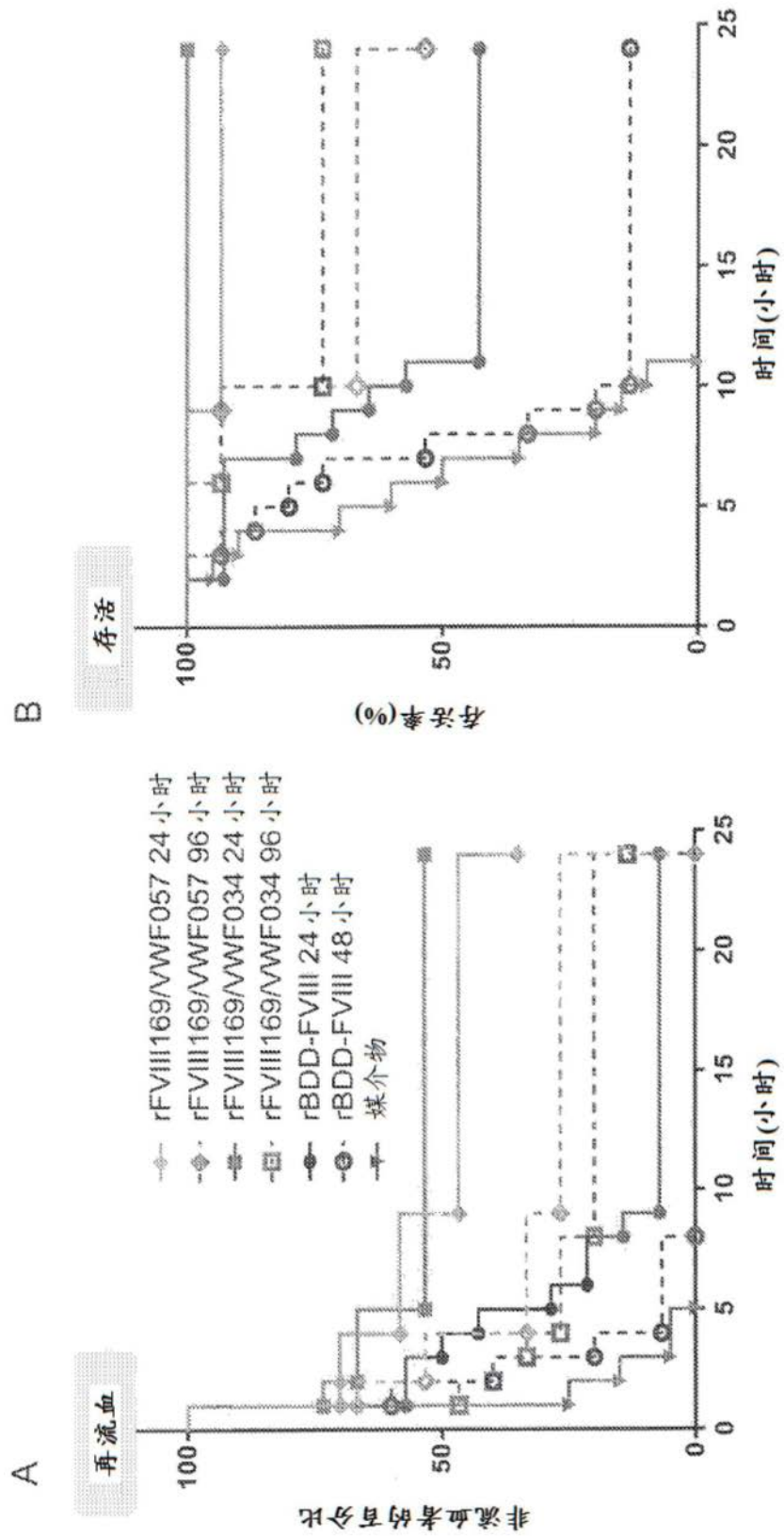


图8

代表性 FVIII-VWF 异二聚体的条块图

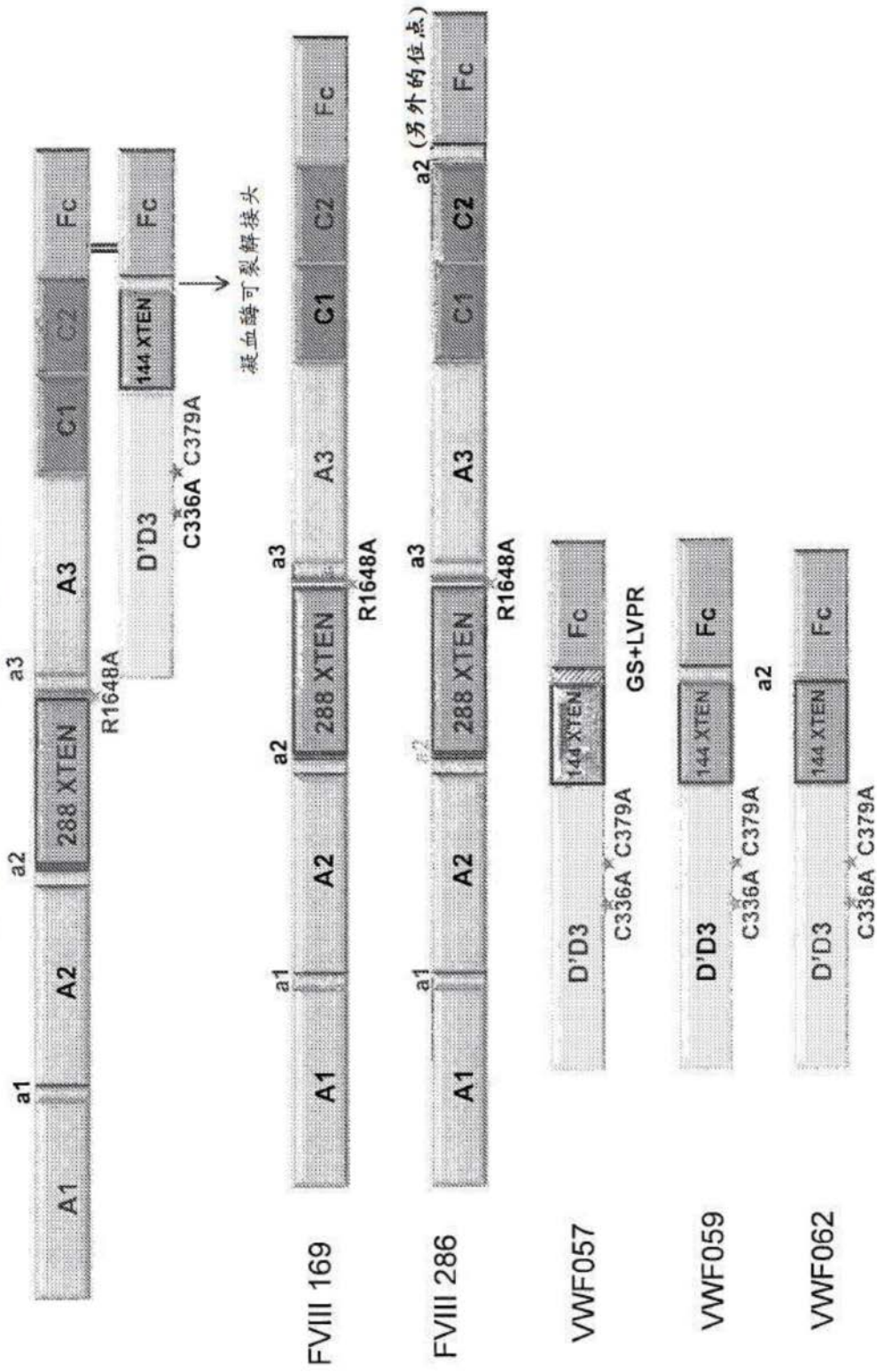


图9

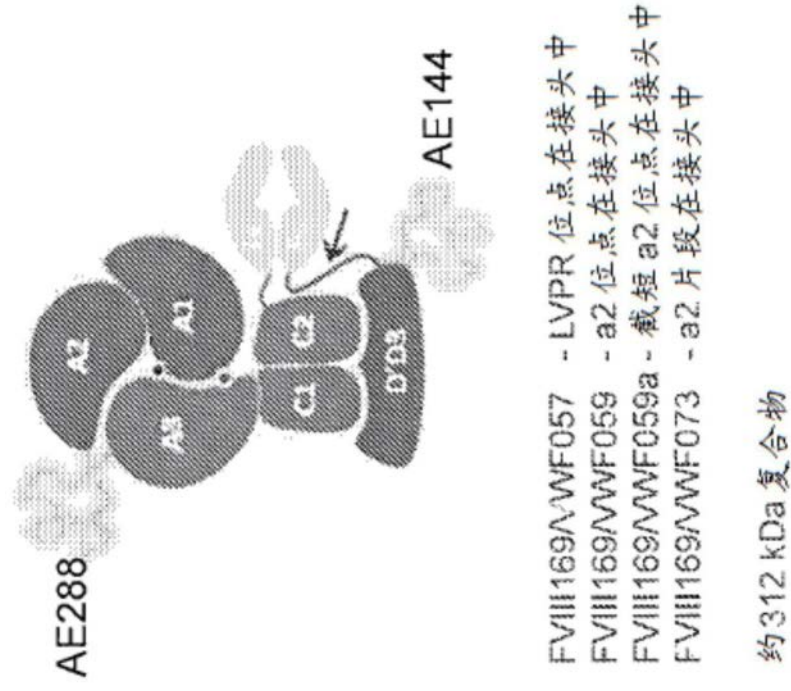
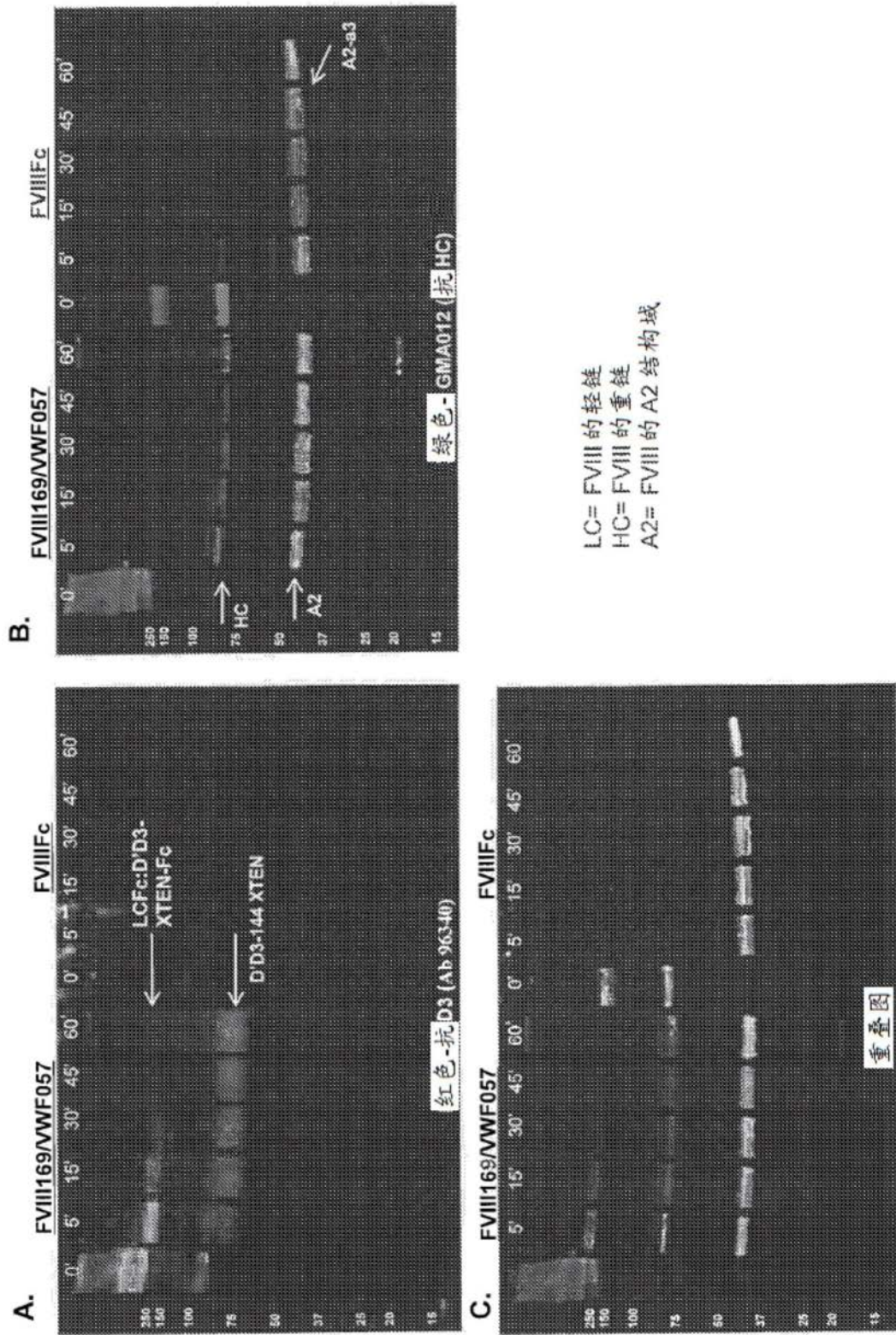


图10



LC= FVIII 的轻链  
 HC= FVIII 的重链  
 A2= FVIII 的 A2 结构域

图11

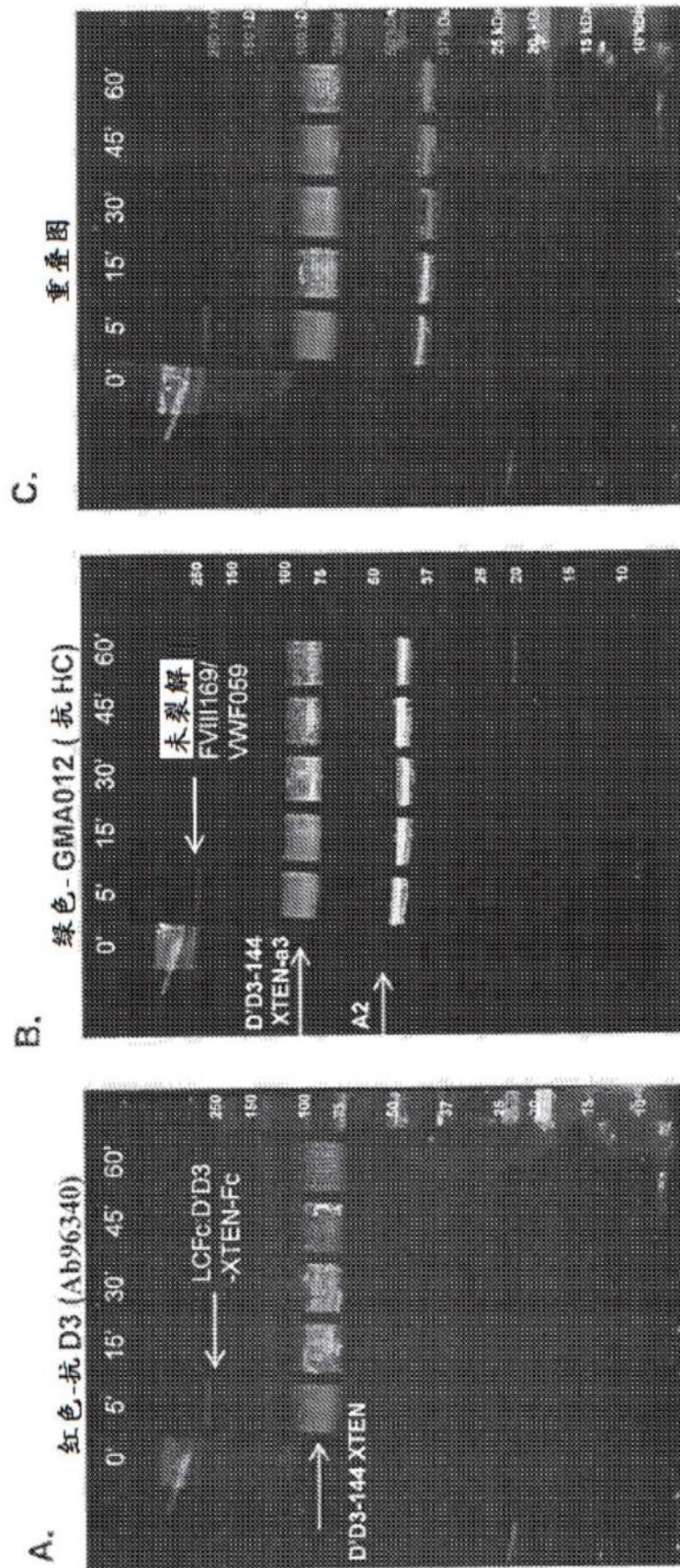


图12

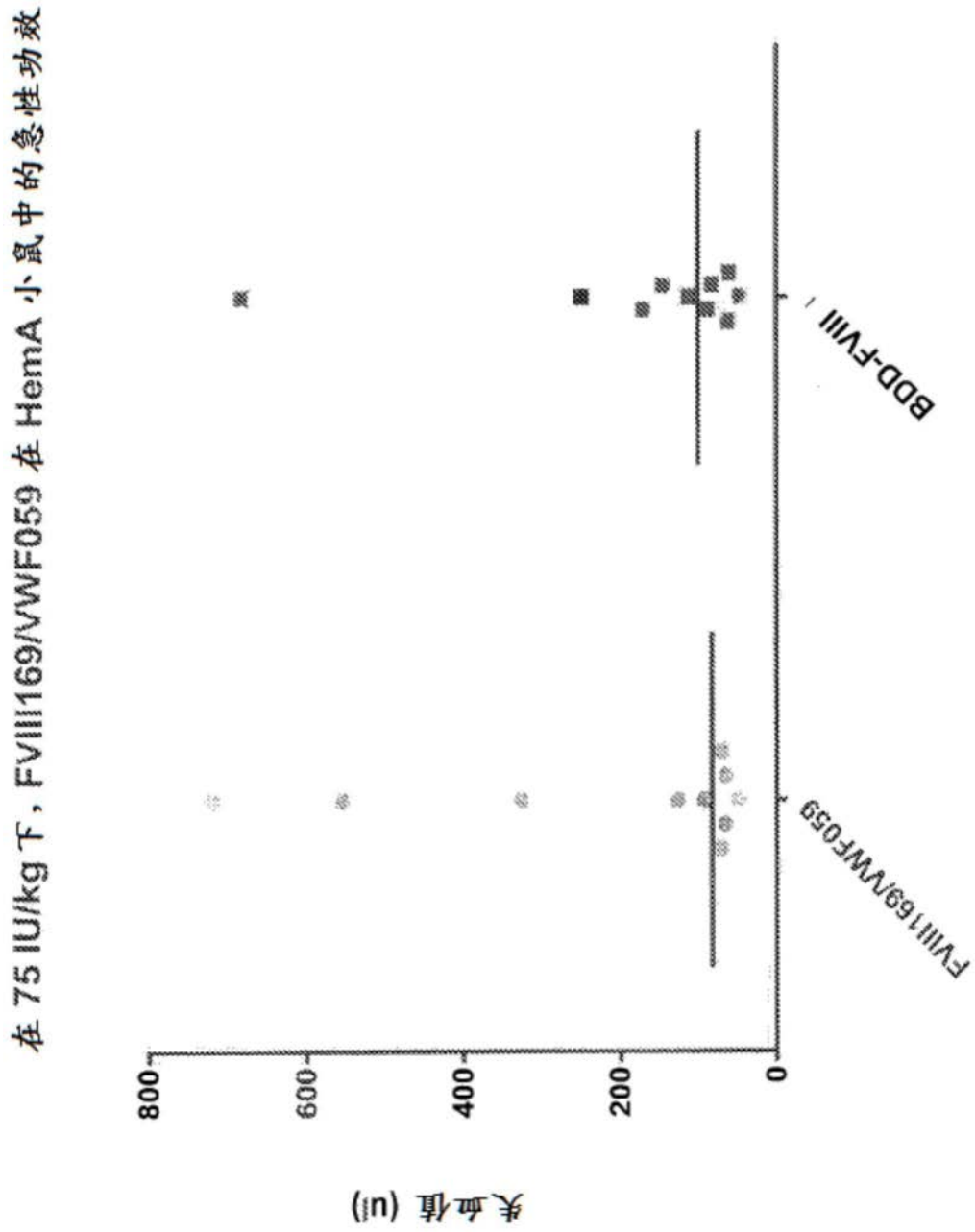


图13