

(45) 授权公告日 2024.07.12

约翰内斯·莱格勒

审查员 华菁菁

序列表33页 附图18页

本公开提供了编码用于在哺乳动物细胞中表达的因子IX变体的密码子改变的多核苷酸以及其他方面。在一些实施方案中,本公开还提供了哺乳动物基因疗法载体和治疗血友病B的方法。



1. 一种核酸组合物,所述组合物包含编码因子IX蛋白的因子IX多核苷酸,所述编码因子IX蛋白的因子IX多核苷酸的核酸序列为SEQ ID NO:17。

2. 如权利要求1所述的核酸组合物,其中因子IX多核苷酸具有不超过10个CpG二核苷酸。

3. 如权利要求1所述的核酸组合物,其中因子IX多核苷酸具有不超过3个CpG二核苷酸。

4. 根据权利要求1-3中任一项所述的核酸组合物,其中因子IX多核苷酸编码相对于SEQ ID NO:2在氨基酸残基384处具有亮氨酸突变的因子IX多肽。

5. 根据权利要求1-3中任一项所述的核酸组合物,其中由因子IX多核苷酸编码的因子IX蛋白具有SEQ ID NO:12的氨基酸序列。

6. 根据权利要求1-3中任一项所述的核酸组合物,其中因子IX多核苷酸进一步包含编码前原前导肽的前原前导多核苷酸,所述前原前导肽包含SEQ ID NO:36的氨基酸序列。

7. 如权利要求6所述的核酸组合物,其中所述前原前导多核苷酸具有SEQ ID NO:23的核酸序列。

8. 根据权利要求6所述的核酸组合物,其中因子IX多核苷酸具有与SEQ ID NO:9的核酸序列具有至少99%同一性的核酸序列。

9. 根据权利要求1-3中任一项所述的核酸组合物,其中所述编码因子IX蛋白的因子IX多核苷酸的核酸序列为SEQ ID NO:9。

10. 一种核酸组合物,所述核酸组合物包含编码因子IX蛋白的因子IX多核苷酸,所述编码因子IX蛋白的因子IX多核苷酸的核酸序列为SEQ ID NO:13。

11. 一种核酸组合物,所述核酸组合物包含编码因子IX蛋白的因子IX多核苷酸,所述编码因子IX蛋白的因子IX多核苷酸的核酸序列为SEQ ID NO:14。

12. 一种核酸组合物,所述核酸组合物包含编码因子IX蛋白的因子IX多核苷酸,所述编码因子IX蛋白的因子IX多核苷酸的核酸序列为SEQ ID NO:15。

13. 一种核酸组合物,所述核酸组合物包含编码因子IX蛋白的因子IX多核苷酸,所述编码因子IX蛋白的因子IX多核苷酸的核酸序列为SEQ ID NO:16。

14. 根据权利要求10-13中任一项所述的核酸组合物,其中因子IX多核苷酸具有不超过10个CpG二核苷酸。

15. 根据权利要求10-13中任一项所述的核酸组合物,其中因子IX多核苷酸具有不超过3个CpG二核苷酸。

16. 根据权利要求10-13中任一项所述的核酸组合物,其中因子IX多核苷酸编码相对于SEQ ID NO:2在氨基酸残基384处具有亮氨酸突变的因子IX多肽。

17. 根据权利要求10-13中任一项所述的核酸组合物,其中由因子IX多核苷酸编码的因子IX蛋白具有SEQ ID NO:12的氨基酸序列。

18. 根据权利要求10-13中任一项所述的核酸组合物,其中因子IX多核苷酸进一步包含编码前原前导肽的前原前导多核苷酸,所述前原前导肽包含SEQ ID NO:36的氨基酸序列。

19. 根据权利要求1-3和10-13中任一项所述的核酸组合物,所述核酸组合物进一步包含可操作地连接至因子IX多核苷酸的肝特异性启动子元件。

20. 如权利要求19所述的核酸组合物,其中所述肝特异性启动子元件包含启动子多核苷酸的一个拷贝,所述启动子多核苷酸包含与SEQ ID NO:39具有至少95%同一性的核酸序

列。

21. 如权利要求19所述的核酸组合物,其中所述肝特异性启动子元件包含启动子多核苷酸的三个拷贝,所述启动子多核苷酸包含与SEQ ID NO:39具有至少95%同一性的核酸序列。

22. 如权利要求20所述的核酸组合物,其中所述启动子多核苷酸包含SEQ ID NO:39的核酸序列。

23. 根据权利要求1-3和10-13中任一项所述的核酸组合物,所述核酸组合物进一步包含可操作地连接至因子IX多核苷酸的内含子。

24. 如权利要求23所述的核酸组合物,其中所述内含子包含含有与SEQ ID NO:53具有至少95%同一性的核酸序列的MVM内含子多核苷酸。

25. 如权利要求24所述的核酸组合物,其中所述MVM内含子多核苷酸包含SEQ ID NO:53的核酸序列。

26. 根据权利要求23所述的核酸组合物,其中所述内含子位于启动子元件与编码因子IX多肽的核苷酸序列的翻译起始位点之间。

27. 如权利要求1所述的核酸组合物,所述核酸组合物包含SEQ ID NO:40的核酸序列。

28. 根据权利要求1-3和10-13中任一项所述的核酸组合物,所述核酸组合物包含哺乳动物基因疗法载体。

29. 如权利要求28所述的核酸组合物,其中所述哺乳动物基因疗法载体为腺相关病毒AAV载体。

30. 如权利要求29所述的核酸组合物,其中所述腺相关病毒载体为血清型8腺相关病毒AAV-8载体。

31. 根据权利要求29所述的核酸组合物,其中所述哺乳动物基因疗法载体包含编码所述因子IX蛋白的单链多核苷酸。

32. 根据权利要求1-3和10-13中任一项所述的核酸组合物,所述核酸组合物用于治疗血友病B。

33. 一种根据权利要求1-3和10-13中任一项所述的核酸组合物的用途,所述用途用于制造用以治疗血友病B的药剂。

34. 一种产生腺相关病毒AAV颗粒的方法,所述方法包括将根据权利要求1-3和10-13中任一项所述的核酸组合物引入至哺乳动物宿主细胞中,其中所述核酸组合物有能力在所述哺乳动物宿主细胞中进行复制。

用于血友病B的基因疗法的编码具有增加的表达的重组FIX的病毒载体

[0001] 相关申请的交叉参考

[0002] 本申请要求于2017年5月22日提交的美国临时申请No.62/509,616的权益,所述美国临时申请出于所有目的以全文引用的方式明确地并入本文中。

[0003] 对以光盘形式提交的“序列表”、表格或计算机程序列表附件的引用

[0004] 本公开通过引用合并了随同提交的于2018年5月21日创建的、名为008073_5117_WO_Sequence_Listing.txt并且大小为73KB的序列表文本副本。

[0005] 发明背景

[0006] 血液凝固是通过相互依赖的生化反应的复杂又动态的生物途径(称为凝血级联反应)来进行。凝血因子VIII(FVIII)和因子IX(FIX)是所述级联反应中的关键组分。因子VIII被募集至出血部位,并且与活化因子IX和因子X(FX)形成X酶复合物。所述X酶复合物使FX活化,这进而使凝血酶原活化为凝血酶,所述凝血酶接着使凝血级联反应中的其他组分活化,从而产生稳定的凝块(综述于Saenko等人,Trends Cardiovasc.Med.,9:185-192(1999); Lenting等人,Blood,92:3983-3996(1998)中)。

[0007] 血友病B是以因子IX活性缺乏为特征的先天性X连锁出血病症。通常,减低的因子VIII/因子IX活性抑制凝血级联反应中的正反馈回路。这导致不完全凝固,表现为出血事件持续时间增加、大范围瘀伤、自发性口腔和鼻腔出血、关节僵硬和慢性疼痛,并且在严重情况下可能有内出血和贫血现象。(Zhang等人,Clinic.Rev.Allerg.Immunol.,37:114-124(2009))。

[0008] 按照惯例,利用因子IX替代疗法来治疗血友病B,所述因子IX替代疗法由向患有血友病B的个体施用因子IX蛋白(例如血浆源性的或以重组方式产生的因子IX)组成。预防性地施用因子IX来预防出血事件或响应于急性出血事件而降低出血事件的频率,和/或在围手术期施用因子IX来管理手术期间的出血。然而,因子IX替代疗法存在有几个不合需要的特征。

[0009] 首先,施用因子IX替代疗法来治疗或管理血友病B,但不能治愈潜在的因子IX缺乏症。由于此原因,患有血友病B的个体在其寿命持续时间内需要因子IX替代疗法。连续治疗是昂贵的并且需要个体维持严格依从性,因为对于患有重度血友病B的个体来说仅遗漏几个预防性剂量就可能会具有严重后果。

[0010] 其次,因为常规因子IX产品在体内具有相对较短的半衰期,约为24小时,所以预防性因子IX替代疗法需要每周施用两次或三次。此给个体增加了在其整个生命中维持依从性的负担。虽然第三代“长效”因子IX药物可以降低施用频率,但使用这些药物的预防性因子FIX替代疗法仍然需要每月、每周或更频繁地永久施用。举例来说,使用Nonacog β pegol[聚乙二醇化重组因子IX](诺和诺德公司(Novo Nordisk),待美国和欧洲监管机构批准)的预防性治疗仍然需要每周施用(Collins P.W.等人,Blood,124(26):3880-86(2014))。此外,对化学修饰的生物制剂(例如聚乙二醇化多肽)的长期效果尚未完全了解。

[0011] 再次,多达5%的重度血友病B患者因子IX替代疗法形成抗因子IX抑制剂抗体,从

而致使疗法效率低 (Osooli和Berntorp, J. Intern. Med., 277(1):1-15(2015))。不同于可以用于治疗已产生抗因子VIII抑制抗体的血友病A患者的因子VIII旁路疗法, 针对治疗血友病B不存在因子IX旁路疗法。

[0012] 最后, 因子IX替代疗法是昂贵的, 视患者体重而定每一剂量在约\$1,000至约\$3,000范围内 (美国血友病联合会 (Hemophilia Federation of America) 在线材料)。因此, 在每周给药两次的情况下, 因子IX替代疗法每年的成本可能高达\$300,000。

[0013] 基因疗法对于治疗血友病B来说前景很好, 因为它将会补救功能因子IX活性潜在的低表达现象 (例如因错义或无义突变所致), 而不是向个体提供一次性剂量的因子IX活性。由于提供因子IX的机制的差异, 与因子IX替代疗法相比, 单次施用因子IX基因疗法载体可以为个体提供足够水平的因子IX达数年, 甚至更久。此降低了治疗成本并且消除了对持续患者依从性的需要。

[0014] 已显示了血友病B的因子IX基因疗法治疗的概念证明。参见例如Manno C.S.等人, Nat Med., 12(3):342-47(2006)。然而, 关于是否能够在足够长的时间内表达治疗有效量的因子IX依然存在疑问。参见例如Giangrande, Semin Thromb Hemost. 42(5):513-17(2016)。

[0015] 已进行了几个尝试来构建密码子优化的因子IX。举例来说, WO 2006/036502公开了一种具有ApoE HCR-1增强子和 α -1抗胰蛋白酶 (AAT) 启动子的密码子优化的因子IX AAV基因疗法载体。类似地, WO 2014/064277和WO 2016/146757公开了包括肝特异性SERPIN调控元件的一个或多个拷贝的密码子优化的因子VIII和因子IX AAV基因疗法载体。最后, WO 2016/210170公开了具有ApoE HCR-1增强子和 α -1抗胰蛋白酶 (AAT) 启动子的密码子优化的因子IX AAV基因疗法载体。

发明内容

[0016] 因此, 对改进的因子IX基因疗法构建体存在需要。举例来说, 对编码被更有效地包装至基因疗法载体中并且经由基因疗法载体递送的因子IX的合成的密码子改变的核酸存在需要。对更有效地表达因子IX的合成的密码子改变的核酸也存在需要。对编码与野生型因子IX相比具有改进的折叠特性、来自表达细胞的分泌得到改善和/或具有增加的活性的因子IX多肽的密码子改变的核酸也存在需要。此类编码因子IX的密码子改变的核酸允许改善对因子IX缺乏症 (例如血友病B) 的治疗。利用所公开的编码因子IX蛋白的密码子改变的核酸来减轻或清除以上缺乏症和与因子IX缺乏症 (例如血友病B) 治疗相关的其他问题。

[0017] 在一个方面中, 描述了编码因子IX和因子IX变体的核酸组合物 (例如密码子改变的多核苷酸)。如本文所描述, 所述核酸组合物包括与编码因子IX的CS02、CS03、CS04、CS05和CS06序列具有高度序列同一性的多核苷酸。本文所描述的核酸组合物相对于野生型因子IX编码序列提供增加的因子IX表达。所述核酸组合物还允许增加基于AAV的基因疗法病毒颗粒的产生。在一些实施方案中, 与编码因子IX的野生型序列相比, 本文所描述的核酸组合物具有减低的GC含量并且或包括较少的CpG二核苷酸。

[0018] 在一些实施方案中, 核酸组合物包括编码因子IX的多核苷酸, 所述多核苷酸的核苷酸序列与所公开的选自以下的序列具有至少95%的序列同一性 (例如至少95%、96%、97%、98%、99%或100%的序列同一性): CS02-FL-NA (SEQ ID NO:5)、CS02-MP-NA (SEQ ID NO:13)、CS03-FL-NA (SEQ ID NO:6)、CS03-MP-NA (SEQ ID NO:14)、CS04-FL-NA (SEQ ID NO:

7)、CS04-MP-NA (SEQ ID NO:15)、CS05-FL-NA (SEQ ID NO:8)、CS05-MP-NA (SEQ ID NO:16)、CS06-FL-NA (SEQ ID NO:9) 和 CS06-MP-NA (SEQ ID NO:17)。

[0019] 在一些实施方案中,核酸组合物包括编码因子IX的多核苷酸,所述多核苷酸的核苷酸序列与所公开的编码因子IX轻链的序列(例如CS02-LC-NA (SEQ ID NO:42)、CS03-LC-NA (SEQ ID NO:44)、CS04-LC-NA (SEQ ID NO:46)、CS05-LC-NA (SEQ ID NO:48) 或 CS06-LC-NA (SEQ ID NO:50)) 和所公开的编码因子IX重链的序列(例如CS02-HC-NA (SEQ ID NO:41)、CS03-HC-NA (SEQ ID NO:43)、CS04-HC-NA (SEQ ID NO:45)、CS05-HC-NA (SEQ ID NO:47) 或 CS06-HC-NA (SEQ ID NO:49)) 具有至少95%的序列同一性(例如至少95%、96%、97%、98%、99%或100%的序列同一性)。

[0020] 在一些实施方案中,核酸组合物包括编码具有轻链、重链和将轻链的C端连接至重链的N端的多肽接头(例如活化肽)的因子IX多肽的多核苷酸。所述因子IX多肽的轻链由与以下一者具有高度序列同一性的第一核苷酸序列编码:CS02-LC-NA (SEQ ID NO:42)、CS03-LC-NA (SEQ ID NO:44)、CS04-LC-NA (SEQ ID NO:46)、CS05-LC-NA (SEQ ID NO:48) 和 CS06-LC-NA (SEQ ID NO:50)。所述因子IX多肽的重链由与以下一者具有高度序列同一性的第二核苷酸序列编码:CS02-HC-NA (SEQ ID NO:41)、CS03-HC-NA (SEQ ID NO:43)、CS04-HC-NA (SEQ ID NO:45)、CS05-HC-NA (SEQ ID NO:47) 和 CS06-HC-NA (SEQ ID NO:49)。所述多肽接头包含蛋白酶切割位点(例如两个因子XIa切割位点)。

[0021] 在上文所描述的多核苷酸的一些实施方案中,所述多肽接头具有与野生型因子IX活化肽FIX-AP-AA (SEQ ID NO:56;FIX-FL-AA (SEQ ID NO:2) 的氨基酸192-226) 具有高度序列同一性的氨基酸序列。在一些实施方案中,所述多肽接头由与以下一者具有高度序列同一性的第三核苷酸序列编码:CS02-AP-NA (SEQ ID NO:57)、CS03-AP-NA (SEQ ID NO:58)、CS04-AP-NA (SEQ ID NO:59)、CS05-AP-NA (SEQ ID NO:60) 和 CS06-AP-NA (SEQ ID NO:61)。

[0022] 在一些实施方案中,本文所描述的密码子改变的多核苷酸编码前原因子IX多肽,例如其中所编码的因子IX蛋白包括信号肽和原肽。在一些实施方案中,信号肽、原肽或信号肽与原肽两者由密码子改变的序列编码。在一些实施方案中,信号肽、原肽或信号肽与原肽两者由野生型序列编码,而编码成熟因子IX单链多肽(例如FIX-MP-AA (SEQ ID NO:10); FIX-FL-AA (SEQ ID NO:2) 的氨基酸47-461) 的核酸的部分为密码子改变的。

[0023] 在一些实施方案中,本文所描述的密码子改变的多核苷酸编码因子IX变体多肽,所述因子IX变体多肽例如相对于野生型因子IX氨基酸序列(例如FIX-FL-AA (SEQ ID NO:2) 或 FIX-MP-AA (SEQ ID NO:10)) 具有一个或多个氨基酸取代。在一些实施方案中,因子IX变体为与野生型因子IX相比具有增加的因子IX活性的超高活性因子IX变体。在一个具体实施方案中,所编码的因子IX多肽具有‘Padua’ R384L氨基酸取代(相对于因子IX前原多肽序列 FIX-FL-AA (SEQ ID NO:2), 相对于成熟因子IX单链序列 FIX-MP-AA (SEQ ID NO:10) 为 R338L 氨基酸取代)。

[0024] 在一个方面中,描述了治疗血友病B的方法。所述方法包括向有需要的患者施用本文所描述的核酸组合物(例如密码子改变的因子IX多核苷酸构建体)(例如与CS02、CS03、CS04、CS05或CS06因子IX编码序列具有高度序列同一性的多核苷酸)。在一些实施方案中,如本文所描述,因子IX多核苷酸构建体为哺乳动物基因疗法载体。在一个具体实施方案中,因子IX多核苷酸构建体为腺相关病毒(AAV)载体。在一些实施方案中,基因疗法载体包括肝

特异性调控控制元件的一个或多个拷贝(例如CRM8调控控制元件的1个至3个拷贝)。

[0025] 在一个方面中,描述了产生腺相关病毒(AAV)颗粒的方法。所述方法包括将本文所描述的核酸组合物(例如密码子改变的因子IX多核苷酸构建体,例如与CS02、CS03、CS04、CS05或CS06因子IX编码序列具有高度序列同一性的多核苷酸)引入至哺乳动物宿主细胞中,其中所述多核苷酸有能力在所述哺乳动物宿主细胞中进行复制。

附图说明

[0026] 图1说明了根据一些实现方式的示例性因子IX基因疗法构建体。图上显示自身互补型(A、B)和单链(C、D)载体的序列元件不具有(A、C)和具有(B、D)肝特异性顺式调控模块(CRM8)。

[0027] 图2显示登录号CCDS14666.1的野生型因子IX编码序列(SEQ ID NO:1) (“FIX-FL-NA”)。

[0028] 图3显示在人体中表达的两种野生型因子IX前原多肽同种型的氨基酸序列。图3A显示对应于UniProt登录号P00740和NCBI登录号NP_000124.1的第一较长因子IX前原多肽同种型的野生型氨基酸序列(SEQ ID NO:2) (“FIX-FL-AA”)。图3B显示对应于NCBI登录号NP_001300842.1的第二较短因子IX前原多肽同种型的野生型氨基酸序列(SEQ ID NO:3) (“FIX2-FL-AA”)。

[0029] 图4显示Padua(R384L)因子IX氨基酸序列(SEQ ID NO:4; “FIXp-FL-AA”)。

[0030] 图5显示根据一些实现方式编码具有R384L氨基酸取代的因子IX变体的CS02密码子改变的核苷酸序列(SEQ ID NO:5) (CS02-FL-NA)。

[0031] 图6显示根据一些实现方式编码具有R384L氨基酸取代的因子IX变体的CS03密码子改变的核苷酸序列(SEQ ID NO:6) (CS03-FL-NA)。

[0032] 图7显示根据一些实现方式编码具有R384L氨基酸取代的因子IX变体的CS04密码子改变的核苷酸序列(SEQ ID NO:7) (CS04-FL-NA)。

[0033] 图8显示根据一些实现方式编码具有R384L氨基酸取代的因子IX变体的CS05密码子改变的核苷酸序列(SEQ ID NO:8) (CS05-FL-NA)。

[0034] 图9显示根据一些实现方式编码具有R384L氨基酸取代的因子IX变体的CS06密码子改变的核苷酸序列(SEQ ID NO:9) (CS06-FL-NA)。

[0035] 图10说明以每千克体重 2×10^{11} vg的剂量注射了具有CRM8肝特异性顺式调控控制元件的0个、1个、2个或3个拷贝的CS02基因疗法构建体的野生型小鼠体内的FIX抗原水平。

[0036] 图11显示在人体中表达的两种单链野生型因子IX蛋白同种型(例如缺乏信号肽和原肽)的氨基酸序列。图11A显示对应于UniProt登录号P00740和NCBI登录号NP_000124.1的第一较长因子IX前原多肽同种型的野生型氨基酸序列(SEQ ID NO:10) (“FIX-MA-AA”)。图11B显示对应于NCBI登录号NP_001300842.1的第二较短因子IX前原多肽同种型的野生型氨基酸序列(SEQ ID NO:11) (“FIX2-MA-AA”)。

[0037] 图12显示单链因子IX(R338L) “Padua”氨基酸序列(SEQ ID NO:12; “FIXp-MP-AA”)。

[0038] 图13显示根据一些实现方式编码具有R338L氨基酸取代的单链因子IX变体的CS02

密码子改变的核苷酸序列 (SEQ ID NO:13) (CS02-MP-NA)。

[0039] 图14显示根据一些实现方式编码具有R338L氨基酸取代的单链因子IX变体的CS03密码子改变的核苷酸序列 (SEQ ID NO:14) (CS03-MP-NA)。

[0040] 图15显示根据一些实现方式编码具有R338L氨基酸取代的单链因子IX变体的CS04密码子改变的核苷酸序列 (SEQ ID NO:15) (CS04-MP-NA)。

[0041] 图16显示根据一些实现方式编码具有R338L氨基酸取代的单链因子IX变体的CS05密码子改变的核苷酸序列 (SEQ ID NO:16) (CS05-MP-NA)。

[0042] 图17显示根据一些实现方式编码具有R338L氨基酸取代的单链因子IX变体的CS06密码子改变的核苷酸序列 (SEQ ID NO:17) (CS06-MP-NA)。

[0043] 图18显示根据一些实现方式编码本文所描述的一些构建体的前原肽 (PPP) 的核酸序列 (NA)。

[0044] 图19显示根据一些实现方式编码本文所描述的一些构建体的信号肽 (SP) 的核酸序列 (NA)。

[0045] 图20显示根据一些实现方式编码本文所描述的一些构建体的原肽 (PP) 的核酸序列 (NA)。

[0046] 图21显示FIX前原肽 (PPP) 的氨基酸序列 (AA)。

[0047] 图22显示FIX信号肽 (SP) 的氨基酸序列 (AA)。

[0048] 图23显示FIX原肽 (PP) 的氨基酸序列 (AA)。

[0049] 图24显示CRM8序列的核酸序列 (SEQ ID NO:39)。

[0050] 图25A和图25B显示CS06-CRM8.3-ssV构建体的核酸序列 (SEQ ID NO:40)。

具体实施方式

[0051] I. 引言

[0052] 基于AAV的基因疗法对于血友病治疗来说前景很好。对于血友病B, 第一临床数据是令人鼓舞的, 因为在至少一些患者中能够维持约10%的FIX水平超过1年。举例来说, 在最初的人体试验中, 证实AAV-FIX构建体的肝动脉插管术使得因子IX在体内有短暂表达。Kay M. 等人, Nat Genet. 24 (3) : 257-61 (2000)。然而, 转导使得免疫系统针对AAV源性衣壳抗原发生适度活化。Manno C.S. 等人, Nat Med. 12 (3) : 342-47 (2006) 和Mingozzi F. 等人, Nat Med. 13 (4) : 419-22 (2007)。

[0053] 非病毒载体可能具有较低免疫原性, 因为它们是基于裸DNA或与非抗原性载体 (例如惰性聚合物、脂质或纳米粒子) 相关的DNA的递送。然而, 非病毒载体的细胞转染率低于病毒递送载体。另外, 来自非病毒载体的长期表达因用于大规模产生构建体的细菌序列的存在而受阻。

[0054] 然而, 这些挑战不能简单地通过施用较高剂量的基因疗法构建体来解决。根据目前的知识, 基于AAV的基因疗法载体的载体剂量不应增至超过每千克体重 2×10^{12} vg。这是因为在此类高剂量下T细胞免疫反应被触发, 此会破坏所转导的细胞, 并且因此, 降低或甚至消除转基因表达。因此, 需要改进FIX表达的策略来使FIX基因疗法成为对于血友病B患者来说为可行的治疗选择。

[0055] 因此, 支持改进因子IX表达和活性的改进型因子IX多肽构建体将会改善两种治疗

方法。举例来说,将会通过降低构建体的初始剂量,从而降低对受试者免疫系统的刺激来改善病毒递送方法。将会通过使用较少的治疗性多核苷酸拷贝支持较大的因子IX活性来改善依赖于施用裸DNA的方法。

[0056] 本公开涉及发现解决这些和其他与因子IX基因疗法相关的问题的密码子改变的因子IX变体编码序列。举例来说,本文所公开的多核苷酸在哺乳动物宿主中提供有明显改进的因子IX表达和活性。在一些实现方式中,通过使用与密码子改变的CS02、CS03、CS04、CS05和CS06构建体具有高度序列同一性的编码因子IX的多核苷酸来实现这些优点。在一些实施方案中,如下文更全面地描述,与野生型构建体相比,这些序列包括显著较少的CpG二核苷酸。

[0057] 有利的是,本文所描述的CS02、CS03、CS04、CS05和CS06密码子改变的因子IX序列与同等的野生型序列相比在体内提供优良的因子IX表达。举例来说,实例1显示相对于携带野生型因子IX编码序列的自身互补型AAV载体,携带CS02、CS03、CS04、CS05或CS06密码子改变的因子IX(R384L)编码序列的自身互补型AAV载体使得体内因子IX活性增加20倍至40倍。类似地,观察到相对于携带野生型因子IX(R384L)编码序列的自身互补型AAV载体,因子IX表达增加了2倍至4倍(表2)。

[0058] 有利的是,可以通过在因子IX编码序列上游引入肝特异性调控元件的一个或多个拷贝来使由CS02、CS03、CS04、CS05和CS06密码子改变的序列产生的改进的因子IX活性进一步增强。举例来说,如实例2和实例3中所证实,在自身互补型AAV因子IX载体中包括一个或多个肝特异性CRM8调控控制元件使小鼠模型中的因子IX表达进一步增加2倍至3倍并且在人肝细胞中则为2倍至13倍(分别见表3和表4)。同样地,在单链AAV因子IX载体中包括肝特异性CRM8调控控制元件的一个或多个拷贝使体内因子IX活性增加2倍(小鼠模型;表5)并且在人肝细胞中则高达26倍(表6)。

[0059] 令人惊讶的是,虽然缺乏肝特异性CRM8调控控制元件的编码密码子改变的因子IX多肽的自身互补型AAV载体与类似单链AAV载体相比提供更大的因子IX表达增加(将表6中由CS06-CRM.0-scV提供的6.2倍的FIX活性增加与由CS06-CRM.0-ssV(SEQ ID NO:40)提供的3.9倍的因子IX活性增加相比较),但含有肝特异性CRM8调控控制元件的多个拷贝的单链AAV因子IX载体显著优于类似自身互补型AAV载体(将表4中由CS02-CRM8.3-scV相对于CS02-CRM8.0-scV提供的12.8倍的因子IX活性增加与表6中由CS06-CRM8.3-ssV(SEQ ID NO:40)相对于CS06-CRM8.0-scV提供的16.8倍的因子IX活性增加相比较)。

[0060] II. 定义

[0061] 除非另有说明,否则如本文所用的以下术语具有归属于它们的含义。

[0062] 如本文所用的术语“因子IX”和“FIX”(其中“IX”是指意指“九”的罗马数字)是可互换使用的,并且是指具有因子IX活性的任何蛋白质(例如活性FIX,常常称为“FIXa”)或具有因子IX活性的蛋白质的蛋白质前体(例如原蛋白或前原蛋白,常常称为pFIX和ppFIX),所述因子IX活性特别是在因子VIII存在下的例如如使用欧洲药典9.0章节2.7.11中所描述的一期因子IX凝集测定所测量的因子X切割活性,所述参考文献的内容通过引用并入本文中。

[0063] 因子IX被翻译为非活性单链多肽,所述多肽包括信号肽、原肽、轻链、活化肽和重链,常常被称为因子IX前原多肽。因子IX前原肽经历翻译后加工以形成活性因子IX蛋白(例如FIXa)。此加工包括移除(例如通过切割)信号肽,随后移除(例如通过切割)原肽,从而形

成仍为非活性的含有因子IX轻链和因子IX重链的单链成熟因子IX多肽。进一步切割成熟因子IX多肽以切除因子IX轻链与因子IX重链之间的活化肽,从而形成活性因子IX蛋白(例如FIXa)。因子IX轻链和因子IX重链通过二硫键保持缔合。

[0064] 举例来说,首先切割野生型人因子IX前原蛋白以释放所编码的信号肽(FIX-FL-AA (SEQ ID NO:2)的氨基酸1-28),从而形成第一单链原蛋白。接着切割此单链原蛋白以释放原肽(FIX-FL-AA (SEQ ID NO:2)的氨基酸29-46),从而形成第二单链原蛋白(例如FIX-MP-AA (SEQ ID NO:10),其中“MP”称号代表“成熟蛋白”)。接着通过因子XIa在FIX轻链与FIX重链之间将第二单链原蛋白切割两次,以释放活化肽(FIX-FL-AA (SEQ ID NO:2)的氨基酸192-226)。此形成由独立的轻链和重链通过二硫键缔合组成的活性因子IXa蛋白。关于因子IX的结构、功能和活化的其他信息,参见例如Brandstetter H.等人P.N.A.S.USA,92(21):9796-800(1995);Hopfner KP等人,Structure,7(8):989-96(1999);和Gailani D.等人,Thromb Res.,133增刊1:S48-51(2014),这些参考文献的内容出于所有目的以全文引用的方式并入本文中。

[0065] 如本文所描述,此活性因子IXa蛋白可以包括一种或多种变体,并且R338L变体在一些实施方案中有特定用途。此被称为具有在本文中称为“FIXp-MP-NA”的核酸序列的“FIXp-MP-AA”(SEQ ID NO:12);“FIXp”代表在最终蛋白质中包括Padua R338L变体。应注意,本文例示的密码子优化的序列CS02至CS06编码FIXp蛋白,包括R338L变体。因此,在FIX的定义中明确涵盖FIXp。

[0066] 如本文所用,术语“因子IX多肽”和“FIX多肽”是指在特定条件下例如如使用欧洲药典9.0章节2.7.11中所描述的一期因子IX凝集测定所测量的具有因子IX丝氨酸蛋白酶活性的多肽。因子IX多肽包括单链前体多肽(包括因子IX前原多肽、因子IX原肽和成熟单链因子IX多肽),所述单链前体多肽在通过上文所描述的翻译后加工被活化时变成具有因子IX丝氨酸蛋白酶活性的活性因子IX蛋白,以及活性因子IX蛋白本身。在因子IX多肽的定义中明确涵盖包括R338L变体的因子IX多肽。在一个示例性实施方案中,人因子IX多肽是指包括与野生型人因子IX多肽中包括轻链和重链的部分(FIX-MP-AA,SEQ ID NO:10,显示于图11A中)或与padua人因子IX多肽中包括轻链和重链的部分(FIXp-MP-AA,SEQ ID NO:12,显示于图12中)具有高度序列同一性(例如至少85%、90%、95%、99%或更高百分比)的氨基酸序列的多肽。

[0067] 如本文所用,术语“因子IX轻链”或简化的“轻链”是指活化因子IXa蛋白中源自因子IX单链多肽的N端部分的多肽,其含有因子IX的Gla模块、EGF样1和EGF样2结构域。在一个示例性实施方案中,人前原因子IX多肽(FIX-FL-AA (SEQ ID NO:2))的氨基酸47-191构成因子IX轻链。如本文所用,野生型因子IX轻链的氨基酸序列被称为FIX-LC-AA (SEQ ID NO:62)。

[0068] 如本文所用,术语“因子IX重链”或简化的“重链”是指活化因子IXa蛋白中源自因子IX单链多肽的C端部分的多肽,其含有因子IX的肽酶S1结构域。在一个示例性实施方案中,人前原因子IX多肽(FIX-FL-AA (SEQ ID NO:2))的氨基酸227-461构成因子IX重链。如本文所用,野生型因子IX重链的氨基酸序列被称为FIX-HC-AA (SEQ ID NO:63)并且当包括R338L变体时被称为FIXp-HC-AA (SEQ ID NO:64)。

[0069] 通常,因子IX轻链和重链被表达为单一多肽链,例如与活化肽一起。然而,在一些

实施方案中,因子IX轻链和因子VIII重链被表达为独立的多肽链(例如共表达),并且加以重构以形成因子IX蛋白(例如体内或体外)。一般来说,出于本发明的目的,即使两条链分开表达,它们也通常处于同一表达载体(例如病毒基因组)上,而不是处于不同表达载体上。

[0070] 如本文所用,术语“因子IX活化肽”或简化的“活化肽”是指在使因子IXa蛋白活化后从因子IX单链多肽切除的肽。在一个示例性实施方案中,人前原因子IX多肽(FIX-FL-AA (SEQ ID NO:2))的氨基酸192-226构成因子IX活化肽。如本文所用,野生型因子IX活化肽的氨基酸序列被称为FIX-AP-AA (SEQ ID NO:56)。

[0071] 如本文所用,术语“因子IX信号肽”或简化的“信号肽”是指由信号肽酶从因子IX前原多肽的N端切除的肽。所述信号肽将新翻译的因子IX前原蛋白引导至内质网。在一个示例性实施方案中,人前原因子IX多肽(FIX-FL-AA (SEQ ID NO:2))的氨基酸1-28构成因子IX信号肽。如本文所用,野生型因子IX信号肽的氨基酸序列被称为FIX-SP-AA (SEQ ID NO:37)。本发明的一些信号肽显示于图19和图22中。

[0072] 如本文所用,术语“因子IX原肽”或简化的“原肽”是指由弗林蛋白酶(Furin)从因子IX原多肽的N端切除(例如在切割信号肽之后)的肽。所述原肽包括 γ -羧化识别位点,所述羧化识别位点将羧化酶募集至相邻Gla模块,从而促进谷氨酰胺残基的羧化。在一个示例性实施方案中,人前原因子IX多肽(FIX-FL-AA (SEQ ID NO:2))的氨基酸29-46构成因子IX原肽。如本文所用,野生型因子IX原肽的氨基酸序列被称为FIX-PP-AA (SEQ ID NO:38)。

[0073] 如本文所用,术语“因子IX前原肽”或简化的“前原肽”是指因子IX信号肽和原多肽的聚集体。在一个示例性实施方案中,人前原因子IX多肽(FIX-FL-AA (SEQ ID NO:2))的氨基酸1-46构成因子IX前原肽。如本文所用,野生型因子IX前原肽的氨基酸序列被称为FIX-PPP-AA (SEQ ID NO:36),其具有图18中所示的称为FIX-PPP-NA (SEQ ID NO:18)的核酸序列(当使用R338L变体时具有相应的FIXp-PPP-AA和FIXp-PPP-NA)。

[0074] 除非本文中另有说明,否则因子IX氨基酸的编号是指图3A中以SEQ ID NO:2呈现的全长野生型人因子IX前原多肽序列(FIX-FL-AA)中的相应氨基酸。因而,当提及本文所公开的因子IX多肽中的氨基酸取代时,所叙述的氨基酸编号是指全长野生型因子IX前原多肽序列中的类似(例如在结构上或功能上等效)和/或同源(例如在进化上在一级氨基酸序列中为保守的)氨基酸。举例来说,R384L氨基酸取代是指在全长野生型人因子IX前原肽序列(FIX-FL-AA (SEQ ID NO:2))的位置384处R至L的取代、在成熟野生型因子IX单链多肽(FIX-MP-AA (SEQ ID NO:10))的位置338处R至L的取代、在全长野生型人因子IX前原肽同种型2序列(FIX2-FL-AA (SEQ ID NO:3))的位置346处R至L的取代、在成熟野生型人因子IX前原肽同种型2序列(FIX2-FL-AA (SEQ ID NO:3))的位置300处R至L的取代以及在野生型人因子IX重链序列(FIX-HC-AA (SEQ ID NO:63))的位置158处R至L的取代。因此,所有这些命名法描述了不同因子IX构建体中的相同“Padua”氨基酸取代。

[0075] 如本文所描述,因子IX氨基酸编号系统视是否包括因子IX前原肽(例如全长野生型人因子IX序列的氨基酸1-46,包括信号肽和原肽)而定。在包括前原肽的情况下,所述编号被称为“包括前原肽”或“PPI”。在不包括前原肽的情况下,所述编号被称为“不包括前原肽”或“PPE”。举例来说,R384L为对与PPE编号中的R338L相同的氨基酸取代的PPI编号。类似地,因子IX氨基酸编号也视因子IX同种型而定。举例来说,R384L为对与同种型2编号中的R346L相同的氨基酸取代的同种型1编号。除非另外指出,否则所有氨基酸编号是指图3A中

以SEQ ID NO:2呈现的全长野生型人因子IX同种型1序列(FIX-FL-AA)中的相应氨基酸。此编号对于除R384L“Padua”突变以外具有相同氨基酸序列的FIXp-FL-AA(SEQ ID NO:4)来说为相同的。

[0076] 野生型因子IX多肽的非限制性实例包括人前原因子IX(例如基因库登录号NP_000124.1(FIX-FL-AA(SEQ ID NO:2)))和NP_001300842.1(FIX2-FL-AA(SEQ ID NO:3))、缺乏信号肽(前原蛋白的氨基酸1-28)和/或原肽(前原蛋白的氨基酸29-46)的相应单链因子IX,和其天然变体;猪前原因子IX(例如UniProt登录号P00741)、缺乏信号肽的相应单链因子IX,和其天然变体;鼠前原因子IX(例如UniProt登录号P16294)、缺乏信号肽的相应单链因子IX,和其天然变体;大鼠前原因子IX(例如UniProt登录号P16296)、缺乏信号肽的相应单链因子IX,和其天然变体;和其他哺乳动物因子VIII同系物(例如黑猩猩、猿、仓鼠、豚鼠等)。

[0077] 如本文所用,因子IX多肽包括在因子VIII存在下具有因子X切割活性的天然变体和人工构建体。如本公开中所用,因子IX涵盖任何天然变体、替代序列、同种型或突变体蛋白质,其保留一定基础因子IX切割活性(例如如在根据欧洲药典9.0章节2.7.11的一期凝聚测定中所测定,相应野生型活性的至少5%、10%、25%、50%、75%或更高百分比,所述参考文献关于其在章节2.7.11中的人凝血因子IX测定的教导内容通过引用明确地并入本文中。在人类群体中所发现的因子IX氨基酸变异(相对于FIX-FL-AA(SEQ ID NO:2))的实例包括但不限于I17N、L20S、C28R、C28Y、V30I、R43L、R43Q、R43W、K45N、R46S、R46T、N48I、S49P、L52S、E53A、E54D、E54G、F55C、G58A、G58E、G58R、E66V、E67K、F71S、E73K、E73V、R75Q、E79D、T84R、Y91C、D93G、Q96P、C97S、P101R、C102R、C102R、G106D、G106S、C108S、D110N、I112S、N113K、Y115C、C119F、C119R、E124K、G125E、G125R、G125V、C134Y、I136T、N138H、G139D、G139S、C155F、G160E、Q167H、S169C、C170F、C178R、C178W、R191C、R191H、R226G、R226Q、R226W、V227D、V227F、V228F、V228L、Q241H、Q241K、C252S、C252Y、G253E、G253R、A265T、C268W、A279T、N283D、E291V、R294G、R294Q、V296M、H302R、N306S、I316F、L318R、L321Q、N328K、N328Y、P333H、P333T、T342K、T342M、I344L、G351D、W356C、G357E、G357R、K362E、G363W、A366D、R379G、R379Q、C382Y、L392F、L383I、R384L、K387E、I390F、M394K、F395I、F395L、C396F、C396S、A397P、R404T、C407R、C407S、D410H、S411G、S411I、G412E、G413R、P414T、V419E、F424V、T426P、S430T、W431G、W431R、G432S、E433A、G433K、C435Y、A436V、G442E、G442R、I443T、R449Q、R449W、Y450C、W453R和I454T。如下文更全面地论述,此编号是相对于野生型人FIX来说的。在人类群体中识别出的其他氨基酸变异为已知的并且可以例如使用国家生物技术信息中心(National Center for Biotechnology Information,“NCBI”)的变异查看器发现,登录号GCF_000001405.25。因子VIII蛋白还包括含有翻译后修饰的多肽。

[0078] 在本公开中特别有用的是FIX蛋白,所述FIX蛋白包括所谓的“Padua”突变,所述突变为在成熟单链因子IX蛋白的位置338(R338L)、因子IX前原多肽的位置384(R384L)处精氨酸变为亮氨酸。此突变赋予FIX蛋白高功能活性。举例来说,已证实“Padua”蛋白(例如含有R338L突变的因子IX)在体内的活性为野生型因子IX的5倍至10倍。美国专利No.6,531,298;Simioni P.等人,N Engl J Med.361(17):1671-75(2009),其以全文引用的方式并入本文中。因此,本公开提供编码Padua-FIX蛋白(有时在本文中称为“FIXp”或“pFIX”)的氨基酸和

核酸构建体。

[0079] 如本文所用,术语“因子IX多核苷酸”和“FIX多核苷酸”是指编码在特定条件下例如如使用欧洲药典9.0章节2.7.11中所描述的一期因子IX凝聚测定所测量具有因子IX丝氨酸蛋白酶活性的因子IX多肽的多核苷酸。因子IX多核苷酸包括编码因子IX单链前体多肽(包括因子IX前原多肽、因子IX原肽和成熟单链因子IX多肽)的多核苷酸,这些因子IX单链前体多肽当通过上文所描述的翻译后加工被活化时变成具有因子IX丝氨酸蛋白酶活性的活性因子IX蛋白。在因子IX多核苷酸的定义中明确涵盖编码包括R338L变体的因子IX多肽的多核苷酸。在一个示例性实施方案中,人因子IX多核苷酸是指编码包括与野生型人因子IX多肽中包括轻链和重链的部分(FIX-MP-AA,SEQ ID NO:10,显示于图11A中)或与padua人因子IX多核苷酸中包括轻链和重链的部分(FIXp-MP-AA,SEQ ID NO:12,显示于图12中)具有高度序列同一性(例如至少85%、90%、95%、99%或更高百分比)的氨基酸序列的多肽的多核苷酸。

[0080] 如本文所描述,因子IX多核苷酸可以包括调控元件,诸如启动子、增强子、终止子、多聚腺苷酸化序列和内含子;以及病毒包装元件,诸如反向末端重复序列(“ITR”);和/或支持多核苷酸在非病毒宿主细胞中的复制的其他元件,例如支持多核苷酸例如在细菌、酵母或哺乳动物宿主细胞中的增殖的复制子。

[0081] 在本公开中特别有用的是密码子改变的因子IX多核苷酸。如本文所描述,与由天然编码的因子IX构建体(例如使用野生型人密码子编码相同因子IX氨基酸序列的多核苷酸)提供的因子IX表达水平相比,密码子改变的FIX多核苷酸提供转基因因子IX在体内增加的表达。如本文所用,术语“增加的表达”是指与施用了天然编码的因子IX构建体的动物的血液中的转基因因子IX蛋白水平相比,施用了编码因子IX的密码子改变的多核苷酸的动物的血液中的转基因因子IX蛋白水平发生增加。增加的蛋白质表达使得因子IX活性增加;因此,增加的表达使得活性增加。

[0082] 在一些实施方案中,增加的表达是指与施用了天然编码的因子IX多核苷酸的动物的血液中的转基因因子IX多肽水平相比,施用了密码子改变的因子IX多核苷酸的动物的血液中的转基因因子IX多肽多至少25%。出于本公开的目的,增加的表达是指由改变密码子序列而产生的效果,而不是由潜在氨基酸取代(例如“Padua”突变)引起的高活性。也就是说,相对于由天然编码的“Padua”蛋白获得的表达水平比较由编码“Padua”因子IX多核苷酸的密码子优化的序列获得的表达水平。在一些实施方案中,增加的表达是指与施用了天然编码的因子IX多核苷酸的动物的血液中的转基因因子IX多肽水平相比,施用了密码子改变的因子IX多核苷酸的动物的血液中的转基因因子IX多肽多至少50%、多至少75%、多至少100%、多至少3倍、多至少4倍、多至少5倍、多至少6倍、多至少7倍、多至少8倍、多至少9倍、多至少10倍、多至少15倍、多至少20倍、多至少25倍、多至少30倍、多至少40倍、多至少50倍、多至少60倍、多至少70倍、多至少80倍、多至少90倍、多至少100倍、多至少125倍、多至少150倍、多至少175倍、多至少200倍、多至少225倍或多至少250倍。动物血液中的因子IX多肽水平可以例如使用对因子IX多肽具有特异性的ELISA测定来测量。

[0083] “因子IX活性”或“因子IX丝氨酸蛋白酶活性”在本文中意指在存在因子VIIIa辅因子的情况下例如经由野生型因子IX中的Arg194-Ile195肽键的水解切割因子X多肽,因此将因子X活化为因子Xa的能力。活性水平可以使用本领域中已知的任何因子IX活性来测量;合

适的测定概述于本文中；用于测定因子IX活性的一种示例性测定为本文所提供的实例中所用的欧洲药典9.0章节2.7.11中所描述的一期因子IX凝聚测定。在一些实施方案中，在一期凝聚测定中使用缺乏FIX活性的人血浆作为对照来测定因子IX特异性。

[0084] 因为某些因子IX变体与野生型因子IX相比在体内具有增强的特定活性，例如人“Padua”变体的因子IX丝氨酸蛋白酶活性为天然编码类型的人因子IX的5倍至10倍，所以在一些实施方案中，例如替代增加的因子IX表达或除增加的因子IX表达以外，通过施用了因子IX多核苷酸的动物的血液中的因子IX活性增加来评估因子IX多核苷酸组合物的治疗潜能。在一些实施方案中，如本文所用，增加的因子IX活性是指与施用了天然编码的因子IX多核苷酸的动物的血液中的因子IX活性相对于在施用天然编码的因子IX多核苷酸之前动物血液中的基线因子IX活性的增加相比，施用了密码子改变的因子IX多核苷酸的动物的血液中的因子IX活性相对于在施用密码子改变的因子IX多核苷酸之前动物血液中的基线因子IX活性的增加更大。在一些实施方案中，增加的因子IX活性是指与施用了天然编码的因子IX多核苷酸的动物的血液中的因子IX活性水平相对于在施用天然编码的因子IX多核苷酸之前动物中的因子IX活性基线水平的增加相比，施用了密码子改变的因子IX多核苷酸的动物的血液中的因子IX活性相对于在施用密码子改变的因子IX多核苷酸之前动物血液中的因子IX活性基线水平的增加至少大25%。在一些实施方案中，增加的因子IX活性是指与施用了天然编码的因子IX多核苷酸的动物的血液中的因子IX活性水平相对于在施用天然编码的因子IX多核苷酸之前动物中的因子IX活性基线水平的增加相比，施用了密码子改变的因子IX多核苷酸的动物的血液中的因子IX活性相对于在施用密码子改变的因子IX多核苷酸之前动物血液中的因子IX活性基线水平的增加至少大50%、大至少75%、大至少100%、大至少3倍、大至少4倍、大至少5倍、大至少6倍、大至少7倍、大至少8倍、大至少9倍、大至少10倍、大至少15倍、大至少20倍、大至少25倍、大至少30倍、大至少40倍、大至少50倍、大至少60倍、大至少70倍、大至少80倍、大至少90倍、大至少100倍、大至少125倍、大至少150倍、大至少175倍、大至少200倍、大至少225倍或大至少250倍。如本文所描述，使用欧洲药典9.0章节2.7.11中所描述的一期因子IX凝聚测定来测量活性。

[0085] 如本文所用，术语“血友病”是指一组广泛地以血液凝结或凝固减少为特征的疾病状态。血友病可以指A型、B型或C型血友病，或指所有三种疾病类型的复合疾病。A型血友病（血友病A）是由因子VIII（FVIII）活性降低或损失引起并且是最突出的血友病亚型。B型血友病（血友病B）由因子IX（FIX）凝结功能的损失或降低引起。C型血友病（血友病C）为因子XI（FXI）凝结活性损失或降低的结果。血友病A和血友病B为X连锁疾病，而血友病C为常染色体疾病。血友病的常规治疗包括预防性地和按需地施用凝血因子，诸如FVIII、FIX（包括Bebulin®-VH和FXI）；以及FEIBA-VH、去氨加压素和血浆输注。

[0086] 如本文所用，术语“因子IX基因疗法”或“FIX基因疗法”包括为患者提供编码因子IX的核酸来缓解、减少或预防与因子IX缺乏症（例如血友病B）相关的一个或多个症状（例如临床因素）的复发的任何治疗方法。所述术语涵盖施用包含编码因子IX分子，包括任何修饰形式的因子IX（例如因子VIII R384L变体）的核酸的任何化合物、药物、程序或方案来维持或改善患有因子IX缺乏症（例如血友病B）的个体的健康状况。本领域技术人员将了解到，FIX基因疗法的过程或FIX基因疗法治疗剂的剂量可以例如基于根据本公开所获得的结果而有所变化。

[0087] 术语“治疗有效量或剂量”或“治疗充足量或剂量”或“有效或充足量或剂量”是指对施用对象产生治疗效果的剂量。举例来说,适用于治疗血友病的药物的治疗有效量可以是能够预防或缓解与血友病相关的一个或多个症状的量。

[0088] 在一些实施方案中,治疗有效性治疗使得受试者的出血事件的频率和/或严重程度降低。

[0089] 如本文所用,术语“基因”是指编码多肽链的DNA分子区段(例如编码区)。在一些实施方案中,基因按区域紧接在产生多肽链时所涉及的编码区之前、之后和/或插入其中而定位(例如调控元件,诸如启动子、增强子、多聚腺苷酸化序列、5'-非翻译区、3'-非翻译区或内含子)。

[0090] 如本文所用,术语“调控元件”是指提供编码序列在细胞中的表达的核苷酸序列,诸如启动子、增强子、终止子、多聚腺苷酸化序列、内含子等。

[0091] 如本文所用,术语“启动子元件”是指帮助控制编码序列表达的核苷酸序列。通常,启动子元件位于基因翻译起始位点的5'处。然而,在某些实施方案中,启动子元件可以位于内含子序列内或编码序列的3'处。在一些实施方案中,适用于基因疗法载体的启动子是源自靶蛋白的天然基因(例如因子VIII启动子)。在一些实施方案中,适用于基因疗法载体的启动子针对在靶生物体的特定细胞或组织中的表达具特异性(例如肝特异性启动子)。在其他实施方案中,将多个已经充分表征的启动子元件之一用于本文所描述的基因疗法载体中。已经充分表征的启动子元件的非限制性实例包括CMV早期启动子、 β -肌动蛋白启动子和甲基CpG结合蛋白2(MeCP2)启动子。在一些实施方案中,启动子为组成性启动子,其基本上驱动靶蛋白的恒定表达。在其他实施方案中,启动子为诱导型启动子,其响应于特定刺激(例如暴露于特定治疗或药剂)而驱动靶蛋白的表达。关于设计用于AAV介导的基因疗法的启动子的综述,参见Gray等人(Human Gene Therapy22:1143-53(2011)),其内容出于所有目的以全文引用的方式明确并入本文中。

[0092] 如本文所用,“CRM8”元件是指源自SERPINA1基因(NCBI登录号NM_000295.4)的顺式作用调控模块,其以肝特异性方式增强可操作连接的基因(例如编码因子IX多肽的序列)的表达,与SEQ ID NO:39具有高度序列同一性。如本文所用,CRM8元件是指其调控元件的单个拷贝,在一些实施方案中,包括在因子IX多核苷酸内的一个或多个拷贝(例如1个、2个、3个或更多个拷贝)中。关于CRM元件(诸如CRM8)的其他信息,参见Chuah MK等人,Mol Ther., 22(9):1605-13(2014),其通过引用并入本文中。

[0093] 如本文所用,“MVM内含子”是指源自与SEQ ID NO:53具有高度序列同一性的小鼠微小病毒的内含子序列。关于MVM内含子本身的其他信息,参见Haut和Pintel,J Virol.72(3):1834-43(1998),并且关于MVM内含子在AAV基因疗法载体中的用途,参见Wu Z等人,Mol Ther., 16(2):280-9(2008),该两个参考文献通过引用并入本文中。

[0094] 如本文所用,术语“可操作地连接”是指第一参考核苷酸序列(例如基因)与第二核苷酸序列(例如调控控制元件)之间允许第二核苷酸序列影响与第一参考核苷酸序列相关的一种或多种特性(例如转录速率)的关系。在本公开的上下文中,当调控元件位于基因疗法载体内使得其发挥对因子IX转基因转录的影响(例如促进性或组织选择性影响)时,调控控制元件可操作地连接至因子IX转基因。

[0095] 如本文所用,术语“载体”是指用于将因子IX核酸转移至宿主细胞中的任何核酸构

建体。在一些实施方案中,载体包括复制子,其用以复制核酸构建体。适用于基因疗法的载体的非限制性实例包括质粒、噬菌体、粘粒、人工染色体和病毒,其充当体内复制的自主性单元。在一些实施方案中,载体为用于将因子IX核酸引入宿主细胞中的病毒载体。许多适用于基因疗法的修饰的真核病毒为本领域中已知的。举例来说,腺相关病毒(AAV)特别适合于在人类基因疗法中使用,因为人类为该病毒的天然宿主,已知天然病毒不会促成任何疾病,并且这些病毒会引起轻度免疫反应。

[0096] 如本文所用,术语“因子IX病毒载体”是指包含编码因子IX多肽的因子IX多核苷酸的重组病毒,其足以使因子IX多肽在引入合适的动物宿主(例如人类)中时进行表达。在因子IX病毒载体的定义内明确包括其中编码因子IX多肽的密码子改变的因子IX多核苷酸已被插入病毒基因组中的重组病毒。在因子IX病毒载体的定义内也明确包括其中病毒的天然基因组已被编码因子IX多肽的因子IX多核苷酸置换的重组病毒。在因子IX病毒载体的定义内包括包含编码因子IX的“Padua”变体的因子IX多核苷酸的重组病毒。

[0097] 如本文所用,术语“因子IX病毒颗粒”是指封装编码因子IX多肽的因子IX多核苷酸的病毒颗粒,其对因子IX多肽在引入合适的动物宿主(例如人类)中时的表达具特异性。在因子IX病毒颗粒的定义内明确包括封装其中已插入编码因子IX多肽的密码子改变的因子IX多核苷酸的基因组的重组病毒颗粒。在因子IX病毒颗粒的定义内也明确包括封装置换病毒天然基因组的编码因子IX多肽的因子IX多核苷酸的重组病毒颗粒。在因子IX病毒颗粒的定义内包括封装编码因子IX的“Padua”变体的因子IX多核苷酸的重组病毒颗粒。

[0098] “AAV”或“腺相关病毒”在本文中意指细小病毒科病毒属内的依赖细小病毒。如本文所用,AAV可以指源自其中已插入因子IX多核苷酸的天然存在的“野生型”AAV基因组的病毒、源自使用由天然存在的AAV cap基因编码的衣壳蛋白包装至衣壳中的重组因子IX多核苷酸的重组病毒或源自使用由非天然衣壳cap基因编码的衣壳蛋白包装至衣壳中的重组因子IX多核苷酸的重组病毒。在AAV的定义内包括封装因子IX多核苷酸的AAV 1型(AAV1)、AAV 2型(AAV2)、AAV 3型(AAV3)、AAV 4型(AAV4)、AAV 5型(AAV5)、AAV 6型(AAV6)、AAV 7型(AAV7)、AAV 8型(AAV8)和AAV 9型(AAV9)病毒以及由一种或多种封装因子IX多核苷酸的变体AAV衣壳蛋白形成的病毒。

[0099] “AAV8”、“AAV-8”或“AAV血清型8”在本文中意指由封装因子IX多核苷酸的AAV8衣壳病毒蛋白形成的病毒。

[0100] 如本文所用,术语“CpG”是指沿循单链DNA的胞嘧啶-鸟嘌呤二核苷酸,其中“p”表示两者之间的磷酸酯键。

[0101] 如本文所用,术语“CpG岛”是指多核苷酸内具有统计学上升高的CpG二核苷酸密度的区域。如本文所用,若在200个碱基对窗内:(i)多核苷酸(例如编码密码子改变的因子IX蛋白的多核苷酸)的区域的GC含量大于50%,并且(ii)如通过以下关系所定义,每个预期CpG二核苷酸观测到的CpG二核苷酸的比率为至少0.6:

$$[0102] \quad \frac{N[CpG]*N[\text{窗长度}]}{N[C]*N[G]} \geq 0.6,$$

[0103] 那么所述区域为CpG岛。关于识别CpG岛的方法的其他信息,参见Gardiner-Garden M.等人,J Mol Biol.,196(2):261-82(1987),所述参考文献的内容出于所有目的以全文引用的方式明确并入本文中。

[0104] 如本文所用,术语“核酸”是指脱氧核糖核苷酸或核糖核苷酸和其呈单链或双链形式的聚合物,以及其互补序列。所述术语涵盖含有已知核苷酸类似物或修饰的主链残基或键联的核酸,其为合成的、天然存在的和非天然存在的,并且其具有与参考核酸类似的结合特性,并且其以类似于参考核苷酸的方式被代谢。此类类似物的实例包括但不限于硫代磷酸酯、氨基磷酸酯、磷酸甲酯、手性磷酸甲酯、2-0-甲基核糖核苷酸和肽-核酸(PNA)。然而,本文中特别适合在患者中用于基因疗法中的实施方案使用磷酸二酯键。

[0105] “核酸组合物”在本文中意指包括编码因子IX多核苷酸的因子IX多核苷酸的任何分子或分子制剂。在核酸组合物的定义内包括因子IX多核苷酸、因子IX多核苷酸的水溶液、封装因子IX多核苷酸的病毒颗粒和封装因子IX多核苷酸的病毒颗粒的水性制剂。如本文所公开的核酸组合物包括编码FIX多肽的密码子改变的FIX基因。

[0106] 术语“氨基酸”是指天然存在和非天然的氨基酸,包括以类似于天然存在的氨基酸的方式起作用的氨基酸类似物和氨基酸模拟物。天然存在的氨基酸包括由遗传密码编码的那些氨基酸,以及随后经过修饰的那些氨基酸,例如羟基脯氨酸、 γ -羧基谷氨酸和O-磷酸丝氨酸。天然存在的氨基酸可以包括例如D-氨基酸和L-氨基酸。关于氨基酸序列,本领域技术人员将认识到针对核酸或肽序列的使所编码序列中的单个氨基酸或小百分比的氨基酸改变、增加或缺失的个别取代、缺失或添加为“保守修饰的变体”,其中所述改变使得氨基酸被化学上类似的氨基酸取代。能够提供功能类似的氨基酸的保守取代在本领域中为人熟知。此类保守修饰的变体为除本公开的多态变体、种间同源体和等位基因以外的并且并不排除本公开的多态变体、种间同源体和等位基因。

[0107] 提供功能类似的氨基酸的保守氨基酸取代在本领域中为人熟知。视特定氨基酸(例如催化性、结构性或空间上重要的氨基酸)的功能性而定,可将不同氨基酸分组视为彼此的保守取代。表1提供基于氨基酸的电荷和极性、氨基酸的疏水性、氨基酸的表面暴露/结构性质以及氨基酸的二级结构倾向性而视为保守取代的氨基酸的分组。

[0108] 表1. 基于蛋白质中的残基功能性的保守氨基酸取代分组。

[0109]

重要特征	保守分组
电荷/极性	1. H、R 和 K 2. D 和 E 3. C、T、S、G、N、Q 和 Y 4. A、P、M、L、I、V、F 和 W
疏水性	1. D、E、N、Q、R 和 K 2. C、S、T、P、G、H 和 Y 3. A、M、I、L、V、F 和 W
结构/表面暴露	1. D、E、N、Q、H、R 和 K 2. C、S、T、P、A、G、W 和 Y 3. M、I、L、V 和 F
二级结构倾向性	1. A、E、Q、H、K、M、L 和 R 2. C、T、I、V、F、Y 和 W 3. S、G、P、D 和 N
进化保守性	1. D 和 E 2. H、K 和 R 3. N 和 Q 4. S 和 T 5. L、I 和 V 6. F、Y 和 W 7. A 和 G 8. M 和 C

[0110] 在两个或更多个核酸或肽序列的背景下,术语“同一”或“同一性”百分比是指如使用BLAST或BLAST 2.0序列比较算法用下文所描述的默认参数或通过手动比对和目视检查所测量,两个或更多个序列或子序列为相同的或具有指定百分比的相同氨基酸残基或核苷酸(即:当在比较窗或指定区域内比较和比对最大对应性时,在指定区域内具有约60%的同一性,优选为65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高同一性)。

[0111] 如本领域中已知,许多不同程序可以用于识别蛋白质(或如下文所论述的核酸)是否与已知序列具有序列同一性或相似性。序列同一性和/或相似性是使用本领域中已知的标准技术来测定,所述标准技术包括但不限于:Smith和Waterman,Adv. Appl. Math., 2:482 (1981)的局部序列同一性算法、Needleman和Wunsch, J. Mol. Biol., 48:443 (1970)的序列同一性比对算法、Pearson和Lipman, Proc. Natl. Acad. Sci. U.S.A., 85:2444 (1988)的相似性搜索法、这些算法的计算机化实现方式(Wisconsin Genetics软件包, Genetics Computer Group, 575 Science Drive, Madison, WI中的GAP、BESTFIT、FASTA和TFASTA)、由Devereux等人, Nucl. Acid Res., 12:387-395 (1984)描述的Best Fit序列程序, 优选使用默认设置或通过检查来进行。优选地,通过FastDB基于以下参数来计算同一性百分比:错配罚分为1;空位罚分为1;空位尺寸罚分为0.33;并且连接罚分为30, “Current Methods in Sequence Comparison and Analysis”, Macromolecule Sequencing and Synthesis, Selected Methods and Applications, 第127-149页 (1988), Alan R. Liss, Inc, 所述所有文献通过引用并入本文中。

[0112] 适用算法的实例为PILEUP。PILEUP使用渐进性成对比对由一组相关序列创建多重序列比对。它也可绘制显示用于创建比对的聚类关系的树形图。PILEUP使用Feng和Doolittle, *J. Mol. Evol.* 35:351-360 (1987) 的渐进性比对方法的简化型式;所述方法类似于由Higgins和Sharp *CABIOS* 5:151-153 (1989) 描述的方法,这两个参考文献均通过引用并入本文中。适用PILEUP参数包括默认空位权重3.00、默认空位长度权重0.10和加权末端空位。

[0113] 适用算法的另一实例为以下文献中所描述的BLAST算法:Altschul等人, *J. Mol. Biol.* 215,403-410, (1990);Altschul等人, *Nucleic Acids Res.* 25:3389-3402 (1997);和Karlin等人, *Proc. Natl. Acad. Sci. U.S.A.* 90:5873-5787 (1993), 这些文献均通过引用并入本文中。特别适用的BLAST程序为WU-BLAST-2程序,其获自Altschul等人, *Methods in Enzymology*, 266:460-480 (1996); <http://blast.wustl.edu/blast/README.html>。WU-BLAST-2使用几个搜索参数,其中大多数被设置为默认值。将可调整参数设置为以下值:重叠跨度=1,重叠分数=0.125,词阈值(T)=11。HSP S和HSP S2参数为动态值并且由程序本身根据特定序列的组成和搜索目标序列所针对的特定数据库的组成来建立;然而,可以调整这些值来增加灵敏度。

[0114] 另一适用算法为如由Altschul等人, *Nucl. Acids Res.*, 25:3389-3402报道的带空位的BLAST,该参考文献通过引用并入本文中。带空位的BLAST使用BLOSUM-62取代得分;阈值T参数设置为9;引发无空位延伸的双击法;使空位长度k承担10+k的代价;Xu设置为16,并且对于数据库搜索阶段,Xg设置为40,而对于算法的输出阶段设置为67。通过对应于约22位的得分来引发带空位的比对。

[0115] 氨基酸序列同一性百分比是通过用匹配的同残基数目除以比对区域中“较长”序列的残基总数目来确定。“较长”序列为具有比对区域中的大多数实际残基的序列(忽略通过WU-Blast-2使比对得分最大化而引入的空位)。以类似方式,将关于所识别多肽的编码序列的“核酸序列同一性百分比(%)”定义为候选序列中与细胞周期蛋白的编码序列中的核苷酸残基具有同一性的核苷酸残基的百分比。优选方法利用设置为默认参数的WU-BLAST-2的BLASTN模块,其中重叠跨度和重叠分数分别设置为1和0.125。

[0116] 比对可以包括在所要比对的序列中引入空位。此外,对于含有比由图3A的野生型因子IX序列(SEQ ID NO:2)编码的蛋白质多或少的氨基酸的序列,应了解,在一个实施方案中,序列同一性百分比将基于相对于氨基酸或核苷酸总数目的同一氨基酸或核苷酸数目来确定。因此,举例来说,如下文所论述,在一个实施方案中,比图3A中所示的序列(SEQ ID NO:2)短的序列的序列同一性将是使用较短序列中的核苷酸数目来确定。在同一性百分比计算中,未针对序列变异的诸如插入、缺失、取代等各种表现形式分配相对权重。

[0117] 在一个实施方案中,仅同一性为正评分(+1)并且向所有形式的序列变异(包括空位)分配值“0”,此排除了对如下文关于序列相似性计算所描述的加权标度或参数的需要。序列同一性百分比可以例如通过用匹配的同残基的数目除以比对区域内“较短”序列的残基总数目并且乘以100来计算。“较长”序列为具有比对区域中的大多数实际残基的序列。

[0118] 术语“等位基因变体”是指特定基因位点处基因的多态形式以及源自基因的mRNA转录物的cDNA,和由它们编码的多肽。术语“优选哺乳动物密码子”是指来自如选自以下列表的在哺乳动物细胞中表达的蛋白质中最常用的编码氨基酸的密码子集合中的密码子

集:Gly (GGC,GGG);Glu (GAG);Asp (GAC);Val (GTG,GTC);Ala (GCC,GCT);Ser (AGC,TCC);Lys (AAG);Asn (AAC);Met (ATG);Ile (ATC);Thr (ACC);Trp (TGG);Cys (TGC);Tyr (TAT,TAC);Leu (CTG);Phe (TTC);Arg (CGC,AGG,AGA);Gln (CAG);His (CAC);和Pro (CCC)。

[0119] 如本文所用,术语密码子改变的是指编码多肽(例如因子IX蛋白)的多核苷酸序列,其中编码多肽的天然多核苷酸的至少一个密码子已经改变以改善多核苷酸序列的特性。在一些实施方案中,改进的特性促使编码多肽的mRNA的转录增加、mRNA稳定性增加(例如改善mRNA半衰期)、多肽翻译增加和/或多核苷酸在载体内的包装增加。可以用于实现改进的性质的变化的非限制性实例包括改变针对特定氨基酸对密码子的使用和/或分布、调整总体和/或局部GC含量、移除富含AT的序列、移除重复的序列元件、调整总体和/或局部CpG二核苷酸含量、移除隐蔽调控元件(例如TATA盒和CCAAT盒元件)、移除内含子/外显子剪接位点、改善调控序列(例如引入Kozak共有序列),以及移除能够在所转录的mRNA中形成二级结构(例如茎环)的序列元件。

[0120] 如本文所论述,存在各种命名法来指本文中的本公开组分。“CS-编号”(例如“CS02”、“CS03”、“CS04”、“CS05”、“CS06”等)是指编码FIX多肽的密码子改变的多核苷酸和/或所编码的多肽,包括变体。举例来说,CS02-FL是指全长(Full Length)密码子改变的CS02多核苷酸序列或由CS02多核苷酸序列编码的氨基酸序列(对于氨基酸(Amino Acid)序列在本文中有时称为“CS02-FL-AA”,而对于核酸(Nucleic Acid)序列则为“CS02-FL-NA”(SEQ ID NO:5))。类似地,“CS02-LC”是指编码FIX多肽的轻链的密码子改变的核酸序列(“CS02-LC-NA”(SEQ ID NO:42))或由CS02多核苷酸序列编码的FIX轻链的氨基酸序列(在本文中有时也称为“CS02-LC-AA”)。同样地,CS02-HC、CS02-HC-AA和CS02-HC-NA(SEQ ID NO:41)对于FIX重链来说同样如此。如本领域技术人员将了解到,对于仅密码子改变(例如与Padua因子IX变体相比不含其他氨基酸取代)的诸如CS02、CS03、CS04、CS05、CS06等构建体,氨基酸序列将为具有同一性的,因为氨基酸序列不因密码子优化而改变。因此,本公开的序列构建体包括但不限于CS02-FL-NA(SEQ ID NO:5)、CS02-FL-AA、CS02-LC-NA(SEQ ID NO:42)、CS02-LC-AA、CS02-HC-AA、CS02-HC-NA(SEQ ID NO:41)、CS03-FL-NA(SEQ ID NO:6)、CS03-FL-AA、CS03-LC-NA(SEQ ID NO:44)、CS03-LC-AA、CS03-HC-AA、CS03-HC-NA(SEQ ID NO:43)、CS04-FL-NA(SEQ ID NO:7)、CS04-FL-AA、CS04-LC-NA(SEQ ID NO:46)、CS04-LC-AA、CS04-HC-AA、CS04-HC-NA、CS05-FL-NA(SEQ ID NO:8)、CS05-FL-AA、CS05-LC-NA(SEQ ID NO:48)、CS05-LC-AA、CS05-HC-AA、CS05-HC-NA(SEQ ID NO:47)、CS06-FL-NA(SEQ ID NO:9)、CS06-FL-AA、CS06-LC-NA(SEQ ID NO:50)、CS06-LC-AA、CS06-HC-AA和CS06-HC-NA(SEQ ID NO:49)。应注意,本文中的所有“CS”构建体编码或含有FIXp氨基酸序列,不过那些编码或含有人野生型FIX氨基酸序列的CS构建体也包括在CS构建体的定义内。

[0121] 如本文所用,术语“肝特异性表达”是指与在其他组织中相比,特定基因(例如密码子改变的转基因因子IX基因)在体内在肝组织中的优先或主导表达。在一些实施方案中,肝特异性表达意指特定基因的所有表达中至少50%发生在受试者的肝组织内。在其他实施方案中,肝特异性表达意指特定基因的所有表达中至少55%、60%、65%、70%、75%、80%、85%、90%、95%、99%或100%发生在受试者的肝组织内。因此,肝特异性调控元件为驱动基因在肝组织中的肝特异性表达的调控元件。

[0122] 如本文所用,术语“小于”X和“小于”X%是指0至X的范围,不包括值X,例如0%至

X%, 不包括X%。如本文所用, 这些术语与以0或0%开始直至, 但不包括X或X%的范围可互换使用。

[0123] 如本文所用, 术语“不超过”X或“不超过”X%是指0至X的范围, 包括值X, 例如0%至X%, 包括X%。如本文所用, 这些术语与以0或0%开始直至, 并且包括X或X%的范围可互换使用。

[0124] 如本文所用, 术语“大于”X或“大于”X%是指X至上限的范围, 不包括值X, 例如X%至100%, 不包括X%。如本文所用, 这些术语在百分比的情形中与以X或X%开始 (但不包括X或X%) 直至100%的上限的范围可互换使用。

[0125] 如本文所用, 术语“至少”X或“至少”X%是指X至上限的范围, 包括值X, 例如X%至100%, 包括X%。如本文所用, 这些术语在百分比的情形中与以X或X%开始 (并且包括X或X%) 至100%的上限的范围可互换使用。

[0126] 如本文所用, 术语“在‘X’与‘Y’之间”、“在‘X’%与‘Y’%之间”、“‘X’至‘Y’”和“‘X’%至‘Y’%”是指X至Y的范围, 包括值X和Y, 例如X%至Y%, 包括X%和Y%。如本文所用, 这些术语与以X或X%开始直至, 并且包括Y或Y%的范围可互换使用。

[0127] III. 密码子改变的因子IX多核苷酸

[0128] 在一些实施方案中, 本公开提供编码因子IX或因子IX变体 (其中FIXp在具体实施方案中获得应用) 的密码子改变的核酸组合物。这些密码子改变的多核苷酸当于基于AAV的基因疗法构建体中施用时提供明显改进的因子IX表达。与常规密码子优化的构建体相比, 这些密码子改变的多核苷酸也展示改进的AAV-病毒颗粒包装。如实例1中所证实, 申请人已通过发现编码具有超高活性R338L氨基酸取代 (基于成熟单链因子IX多肽序列; 基于因子IX前原蛋白序列为R384L) 的因子IXp多肽的几个密码子改变的多核苷酸 (例如CS02-FL-NA、CS03-FL-NA、CS04-FL-NA、CS05-FL-NA和CS06-FL-NA (分别为SEQ ID NO: 5-9)) 而实现这些优点。如实例2和实例3中所证实, 一个或多个肝特异性调控控制元件 (例如CRM8) 并入编码因子IX分子的基因疗法载体中进一步增加了因子IX的体内和体外表达和因子IX活性。

[0129] 野生型因子IX被编码而具有28个氨基酸的信号肽 (FIX-SP-AA (SEQ ID NO: 37)) 和18个氨基酸的原肽 (FIX-PP-AA (SEQ ID NO: 38)), 它们是在因子IXa活化之前从所编码的多肽切割出来。如本领域技术人员所了解, 信号肽和/或原肽可为突变的, 被来自其他基因或其他生物体的信号肽和/或原肽置换, 或被完全移除, 而不会影响在通过细胞加工移除信号肽和原肽之后所留下的成熟多肽的序列。

[0130] 因此, 在一些实施方案中, 本文所提供的密码子改变的多核苷酸 (例如核酸组合物) 具有与编码成熟因子IX单链多肽的CS02-FL-NA、CS03-FL-NA、CS04-FL-NA、CS05-FL-NA或CS06-FL-NA (分别为SEQ ID NO: 5-9) 具有高度序列同一性的核苷酸序列, 所述成熟因子IX单链多肽即为因子IX轻链、活化肽和重链 (例如由野生型因子IX基因编码的全长多肽的氨基酸47-461; FIX-FL-AA (SEQ ID NO: 2))。

[0131] 另外, 如本领域中所知, 人野生型因子IX具有位于因子IX轻链与重链之间的34个氨基酸的活化肽, 其在蛋白质活化后从单链因子IX多肽切除。因为从活性因子IX多肽移除所述活化肽, 该肽本身对于最终因子IX活性来说为非必需的。因此, 并不要求由本文所公开的密码子改变的多核苷酸编码的因子IX多肽与人野生型活化肽序列 (FIX-AP-AA (SEQ ID NO: 56)) 具有高度序列同一性。然而, 所编码的活化肽应在因子IX多肽活化后可切除。举例

来说,在一些实施方案中,所编码的活化肽应在其N端和C端包括因子XI切割位点,这些因子XI切割位点可由人因子IX在体内所识别和切割。

[0132] 因此,在一些实施方案中,本文所提供的密码子改变的多核苷酸(例如核酸组合物)编码与人野生型FIX轻链序列(FIX-LC-AA(SEQ ID NO:62))和人野生型FIX重链序列(FIX-HC-AA(SEQ ID NO:63))具有高度序列同一性的单链因子IX多肽,并且另外编码具有两个因子XI切割位点的将轻链C端连接至重链N端的多肽接头(例如活化肽)。

[0133] 在一些实施方案中,由密码子改变的多核苷酸编码的因子IX轻链和重链分别为人因子IX轻链和重链,包括FIXp重链。在其他实施方案中,由密码子改变的多核苷酸编码的因子IX轻链和重链为来自另一哺乳动物(例如猪因子IX)的重链和轻链序列。在其他实施方案中,因子IX轻链和重链为嵌合轻链和重链(例如人和第二哺乳动物序列的组合)。在其他实施方案中,因子IX轻链和重链为来自另一哺乳动物的轻链和重链的人源化型式,例如来自另一哺乳动物的其中人残基在所选位置被取代以降低所得肽当向人类施用时的免疫原性的轻链和重链序列。

[0134] 人类基因的GC含量变化幅度很大,从小于25%至大于90%。然而,一般来说,具有较高GC含量的人类基因以较高水平表达。举例来说,Kudla等人(PLoS Biol.,4(6):80(2006))证实增加基因的GC含量主要通过增加转录和实现较高的mRNA转录物稳态水平来增加所编码多肽的表达。通常,密码子优化的基因构建体的所需GC含量被认为等于或大于60%。举例来说,使用GeneOptimizer软件(Geneart)对scAAV8.FIXR338L基因疗法载体中的因子IX基因特定地进行密码子改变,以使野生型编码序列的GC含量从41%GC增加至61%GC。参见Wu Z.等人,Mol Ther 16:280-89(2008),和Monahan PE等人,Hum Gene Ther.,26(2):69-81(2015)。然而,天然AAV基因组具有约56%的GC含量。

[0135] 因此,在一些实施方案中,本文所提供的密码子改变的多核苷酸(例如核酸组合物)的CG含量更接近地匹配天然AAV病毒颗粒的GC含量(例如约56%GC),此含量低于常规针对哺乳动物细胞中的表达进行了密码子优化的多核苷酸的优选CG含量(例如等于或高于60%GC)。举例来说,CS02-FL-NA(SEQ ID NO:5)具有约54%的GC含量,CS03-FL-NA(SEQ ID NO:6)具有约55%的GC含量,CS04-FL-NA(SEQ ID NO:7)具有约54.5%的GC含量,CS05-FL-NA(SEQ ID NO:8)具有约56.6%的GC含量,并且CS06-FL-NA(SEQ ID NO:9)具有约55%的GC含量。这些构建体与具有较高GC含量的类似密码子改变的序列相比将提供改进的病毒颗粒包装。

[0136] 因此,在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸(例如与CS02-CS06因子IX编码序列之一具有高度序列同一性的多核苷酸)的总GC含量小于60%。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量小于59%。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量小于58%。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量小于57%。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量不超过56%。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量不超过55%。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量不超过54%。

[0137] 在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为53%至59%。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为54%

至59%。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为55%至59%。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为56%至59%。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为53%至58%。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为54%至58%。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为55%至58%。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为56%至58%。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为53%至57%。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为54%至57%。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为55%至57%。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为56%至57%。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为53%至56%。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为54%至56%。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为55%至56%。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为53%至55%。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为54%至55%。

[0138] 在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为 $54 \pm 0.5\%$ 。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为 $54 \pm 0.4\%$ 。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为 $54 \pm 0.3\%$ 。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为 $54 \pm 0.2\%$ 。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为 $54 \pm 0.1\%$ 。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为54%。

[0139] 在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为 $55 \pm 0.5\%$ 。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为 $55 \pm 0.4\%$ 。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为 $55 \pm 0.3\%$ 。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为 $55 \pm 0.2\%$ 。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为 $55 \pm 0.1\%$ 。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为55%。

[0140] 在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为 $56 \pm 0.5\%$ 。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为 $56 \pm 0.4\%$ 。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为 $56 \pm 0.3\%$ 。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为 $56 \pm 0.2\%$ 。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为 $56 \pm 0.1\%$ 。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的总GC含量为56%。

[0141] 从理论上讲,这些CpG二核苷酸(即:胞嘧啶核苷酸,随后鸟嘌呤核苷酸)在体内经由toll样受体诱导免疫反应。一些证据表明在某些情况下,CpG耗尽的AAV载体避开小鼠中的免疫检测(Faust等人,J.Clin.Invest.2013;123,2994-3001)。野生型因子IX编码序列(FIX-FL-NA(SEQ ID NO:1))含有20个CpG二核苷酸。

[0142] 因此,在一些实施方案中,对本文所提供的核酸组合物(例如密码子改变的多核苷

酸)进行密码子改变以减少因子IX编码序列中的CpG二核苷酸数目。举例来说,CS02-FL-NA (SEQ ID NO:5)不具有CpG二核苷酸,CS03-FL-NA (SEQ ID NO:6)不具有CpG二核苷酸,CS04-FL-NA (SEQ ID NO:7)不具有CpG二核苷酸,CS05-FL-NA (SEQ ID NO:8)具有11个CpG二核苷酸,并且CS06-FL-NA (SEQ ID NO:9)具有3个CpG二核苷酸。这些构建体与野生型因子IX编码序列和具有较多CpG二核苷酸数目的类似密码子改变的序列相比应导致较低免疫原性反应。

[0143] 因此,在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸(例如与CS02-CS06因子IX编码序列之一具有高度序列同一性的多核苷酸)的序列具有少于20个CpG二核苷酸。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的序列具有少于15个CpG二核苷酸。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的序列具有少于12个CpG二核苷酸。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的序列具有少于10个CpG二核苷酸。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的序列具有少于5个CpG二核苷酸。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的序列具有少于3个CpG二核苷酸。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的序列不具有CpG二核苷酸。

[0144] 在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的序列具有超过15个CpG二核苷酸。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的序列具有不超过12个CpG二核苷酸。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的序列具有不超过10个CpG二核苷酸。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的序列具有不超过5个CpG二核苷酸。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的序列具有不超过3个CpG二核苷酸。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的序列不具有CpG二核苷酸。在一些实施方案中,编码因子IX多肽的密码子改变的多核苷酸的序列具有不超过18、17、16、15、14、13、12、11、10、9、8、7、6、5、4、3、2、1个CpG二核苷酸或没有CpG二核苷酸。

[0145] A. 因子IX氨基酸取代

[0146] 为了进一步增加本文所描述的因子IX构建体的基于AAV载体的表达效率,根据一些实现方式,进一步并入已知改善分泌、增加特定活性和/或增强因子IX稳定性的氨基酸取代。在本领域中已知许多潜在因子IX变体增加血浆FIX活性水平。这些变体包括增加因子IX催化活性(例如超高活性突变体)、增加对抗凝血酶III和/或肝素的抗性、增加血清半衰期和导致翻译后修饰模式发生改变的氨基酸取代。

[0147] 举例来说,残基R338 (PPE) 的突变可以增加因子IX的凝结活性。关于综述,参见美国专利No.6,531,298,其内容出于所有目的以全文引用的方式并入本文中。如美国专利No.6,531,298中所公开,在此位置精氨酸变为亮氨酸的氨基酸取代增加因子IX的活性。此随后在体内得到证实,其中R338L (PPE) 突变在体内使因子IX活性增加至5倍至10倍。关于综述,参见Simioni P.等人,N Engl J Med.361(17):1671-75(2009),其以全文引用的方式并入本文中。因此,在一些实施方案中,本文所描述的密码子改变的多核苷酸编码在精氨酸384 (PPI;残基338 (PPE) 处具有氨基酸取代的因子IX多肽。在一个具体实施方案中,氨基酸取代为R384L (PPI)。在其他实施方案中,在残基384 (PPI) /338 (PPE) 处的氨基酸取代为变成并非亮氨酸的残基。举例来说,据报道R384A (PPI) 氨基酸取代在小鼠中提供2倍至6倍高的

活性。Schuettrumpf J等人, Blood, 105(6):2316-23(2005), 其内容出于所有目的以全文引用的方式明确并入本文中。

[0148] 类似地, 残基Y305、K311、S365和Y391的突变导致合成底物上的因子IX活性增加。特别是, K311M和K311T单一突变导致合成切割底物上的活性增加2.8倍和6.7倍。Sichler K.等人, J Biol Chem. 278(6):4121-26(2003) (使用不同残基编号)。另外, Y305F/K311T/Y391T三重突变体导致合成底物上的活性增加7000倍。同上。因此, 在一些实施方案中, 本文所描述的密码子改变的多核苷酸编码在酪氨酸305 (PPI)、赖氨酸311 (PPI) 和酪氨酸391 (PPI) 处具有氨基酸取代的因子IX多肽。在一个具体实施方案中, 氨基酸取代为K311M (PPI)。在一个具体实施方案中, 氨基酸取代为K311T (PPI)。在另一个具体实施方案中, 氨基酸取代为Y305F/K311T/Y391T (PPI)。

[0149] 提供改进的特性的其他氨基酸取代为本领域中已知的且可以并入所描述的密码子改变的因子IX多核苷酸中。举例来说, 参见美国专利No. 8,778,870, 其内容出于所有目的以全文引用的方式明确并入本文中。

[0150] B. 编码因子IX蛋白的密码子改变的多核苷酸

[0151] CS02密码子改变的多核苷酸

[0152] 在一个实施方案中, 本文所提供的核酸组合物包括编码单链因子IX多肽的因子IX多核苷酸(例如密码子优化的多核苷酸), 其中因子IX多核苷酸包括与CS02-FL-NA (SEQ ID NO:5) 具有高度序列同一性的核苷酸序列。在一些实施方案中, 与CS02-FL-NA (SEQ ID NO:5) 具有高度序列同一性的因子IX多核苷酸的核苷酸序列与野生型因子IX编码序列 (FIX-FL-NA (SEQ ID NO:1)) 相比具有降低的GC含量。在一些实施方案中, 与CS02-FL-NA (SEQ ID NO:5) 具有高度序列同一性的因子IX多核苷酸的核苷酸序列与野生型因子IX编码序列 (FIX-FL-NA (SEQ ID NO:1)) 相比具有减少的CpG二核苷酸数目。

[0153] 在一个具体实施方案中, 密码子改变的多核苷酸的序列与CS02-FL-NA (SEQ ID NO:5) 具有至少95%的同一性。在一个具体实施方案中, 密码子改变的多核苷酸的序列与CS02-FL-NA (SEQ ID NO:5) 具有至少96%的同一性。在一个具体实施方案中, 密码子改变的多核苷酸的序列与CS02-FL-NA (SEQ ID NO:5) 具有至少97%的同一性。在一个具体实施方案中, 密码子改变的多核苷酸的序列与CS02-FL-NA (SEQ ID NO:5) 具有至少98%的同一性。在一个具体实施方案中, 密码子改变的多核苷酸的序列与CS02-FL-NA (SEQ ID NO:5) 具有至少99%的同一性。在一个具体实施方案中, 密码子改变的多核苷酸的序列与CS02-FL-NA (SEQ ID NO:5) 具有至少99.5%的同一性。在一个具体实施方案中, 密码子改变的多核苷酸的序列与CS02-FL-NA (SEQ ID NO:5) 具有至少99.9%的同一性。在另一个具体实施方案中, 密码子改变的多核苷酸的序列为CS02-FL-NA (SEQ ID NO:5)。

[0154] 在一些实施方案中, 与CS02-FL-NA (SEQ ID NO:5) 具有高度序列同一性的密码子改变的多核苷酸的序列具有小于60%的GC含量。在一些实施方案中, 与CS02-FL-NA (SEQ ID NO:5) 具有高度序列同一性的密码子改变的多核苷酸的序列具有小于59%的GC含量。在一些实施方案中, 与CS02-FL-NA (SEQ ID NO:5) 具有高度序列同一性的密码子改变的多核苷酸的序列具有小于58%的GC含量。在一些实施方案中, 与CS02-FL-NA (SEQ ID NO:5) 具有高度序列同一性的密码子改变的多核苷酸的序列具有小于57%的GC含量。在一些实施方案中, 与CS02-FL-NA (SEQ ID NO:5) 具有高度序列同一性的密码子改变的多核苷酸的序列具

多核苷酸的序列具有不超过6个CpG二核苷酸。在一些实施方案中,与CS02-FL-NA (SEQ ID NO:5) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过5个CpG二核苷酸。在一些实施方案中,与CS02-FL-NA (SEQ ID NO:5) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过4个CpG二核苷酸。在一些实施方案中,与CS02-FL-NA (SEQ ID NO:5) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过3个CpG二核苷酸。在一些实施方案中,与CS02-FL-NA (SEQ ID NO:5) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过2个CpG二核苷酸。在一些实施方案中,与CS02-FL-NA (SEQ ID NO:5) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过1个CpG二核苷酸。在一些实施方案中,与CS02-FL-NA (SEQ ID NO:5) 具有高度序列同一性的密码子改变的多核苷酸的序列不具有CpG二核苷酸。

[0158] 在一些实施方案中,所编码的因子IX多肽(例如由与CS02-FL-NA (SEQ ID NO:5) 具有高度序列同源性的多核苷酸编码的多肽)与野生型因子IX前原蛋白序列FIX-FL-AA (SEQ ID NO:2) 和/或Padua (hFIX (R384L))前原蛋白序列FIXp-FL-AA (SEQ ID NO:4) 具有高度序列同一性。所编码的因子IX多肽应保留被活化为功能因子IXa蛋白(例如通过移除信号肽和原肽,和通过切除活化多肽)的能力。

[0159] 在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少85%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少90%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少95%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少96%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少97%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少98%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少99%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少99.5%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少99.9%的同一性。在一个实施方案中,所编码的因子IX多肽的序列为FIX-FL-AA (SEQ ID NO:2)。

[0160] 在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少85%的同一性并且在前原多肽的位置384(例如成熟因子IX单链多肽FIXp-MP-AA (SEQ ID NO:12) 的位置338)处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少90%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少95%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少96%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少97%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少98%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少99%的同一性并且在前原多肽的位置384处包括亮氨

酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA(SEQ ID NO:4)具有至少99.5%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA(SEQ ID NO:4)具有至少99.9%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列为FIXp-FL-AA(SEQ ID NO:4)。

[0161] 在一个实施方案中,本文所提供的核酸组合物包括编码单链因子IX多肽(例如具有丝氨酸蛋白酶活性)的因子IX多核苷酸(例如密码子改变的多核苷酸),其中因子IX多核苷酸包括与CS02-MP-NA(SEQ ID NO:13)具有高度序列同一性的核苷酸序列。在一些实施方案中,与CS02-MP-NA(SEQ ID NO:13)具有高度序列同一性的因子IX多核苷酸的核苷酸序列与野生型因子IX编码序列(FIX-FL-NA(SEQ ID NO:1))相比具有降低的GC含量。在一些实施方案中,与CS02-MP-NA(SEQ ID NO:13)具有高度序列同一性的因子IX多核苷酸的核苷酸序列与野生型因子IX编码序列(FIX-FL-NA(SEQ ID NO:1))相比具有减少的CpG二核苷酸数目。

[0162] 在一个具体实施方案中,密码子改变的多核苷酸的序列与CS02-MP-NA(SEQ ID NO:13)具有至少95%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS02-MP-NA(SEQ ID NO:13)具有至少96%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS02-MP-NA(SEQ ID NO:13)具有至少97%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS02-MP-NA(SEQ ID NO:13)具有至少98%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS02-MP-NA(SEQ ID NO:13)具有至少99%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS02-MP-NA(SEQ ID NO:13)具有至少99.5%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS02-MP-NA(SEQ ID NO:13)具有至少99.9%的同一性。在另一个具体实施方案中,密码子改变的多核苷酸的序列为CS02-MP-NA(SEQ ID NO:13)。

[0163] 在一些实施方案中,与CS02-MP-NA(SEQ ID NO:13)具有高度序列同一性的密码子改变的多核苷酸的序列具有低于60%的GC含量。在一些实施方案中,与CS02-MP-NA(SEQ ID NO:13)具有高度序列同一性的密码子改变的多核苷酸的序列具有低于59%的GC含量。在一些实施方案中,与CS02-MP-NA(SEQ ID NO:13)具有高度序列同一性的密码子改变的多核苷酸的序列具有低于58%的GC含量。在一些实施方案中,与CS02-MP-NA(SEQ ID NO:13)具有高度序列同一性的密码子改变的多核苷酸的序列具有低于57%的GC含量。在一些实施方案中,与CS02-MP-NA(SEQ ID NO:13)具有高度序列同一性的密码子改变的多核苷酸的序列具有低于56%的GC含量。在一些实施方案中,与CS02-MP-NA(SEQ ID NO:13)具有高度序列同一性的密码子改变的多核苷酸的序列具有低于55%的GC含量。在一些实施方案中,与CS02-MP-NA(SEQ ID NO:13)具有高度序列同一性的密码子改变的多核苷酸的序列具有低于54%的GC含量。

[0164] 在一些实施方案中,与CS02-MP-NA(SEQ ID NO:13)具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至60%的GC含量。在一些实施方案中,与CS02-MP-NA(SEQ ID NO:13)具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至59%的GC含量。在一些实施方案中,与CS02-MP-NA(SEQ ID NO:13)具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至58%的GC含量。在一些实施方案中,与CS02-MP-NA(SEQ ID

子改变的多核苷酸的序列不具有CpG二核苷酸。

[0167] 在一些实施方案中,与CS02-MP-NA (SEQ ID NO:13) 具有高度序列同一性的因子IX多核苷酸进一步包括编码具有FIX-SP-AA (SEQ ID NO:37) 的氨基酸序列的因子IX信号肽的因子IX信号多核苷酸。在一些实施方案中,因子IX信号多核苷酸与CS02-SP-NA (SEQ ID NO:25) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX信号多核苷酸与CS03-SP-NA (SEQ ID NO:26) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX信号多核苷酸与CS04-SP-NA (SEQ ID NO:27) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX信号多核苷酸与CS05-SP-NA (SEQ ID NO:28) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX信号多核苷酸与CS06-SP-NA (SEQ ID NO:29) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。

[0168] 在一些实施方案中,与CS02-MP-NA (SEQ ID NO:13) 具有高度序列同一性的因子IX多核苷酸进一步包括编码具有FIX-PP-AA (SEQ ID NO:38) 的氨基酸序列的因子IX原肽的因子IX原肽多核苷酸。在一些实施方案中,因子IX原肽多核苷酸与CS02-PP-NA (SEQ ID NO:31) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX原肽多核苷酸与CS03-PP-NA (SEQ ID NO:32) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX原肽多核苷酸与CS04-PP-NA (SEQ ID NO:33) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX原肽多核苷酸与CS05-PP-NA (SEQ ID NO:34) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX原肽多核苷酸与CS06-PP-NA (SEQ ID NO:35) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。

[0169] 在一些实施方案中,与CS02-MP-NA (SEQ ID NO:13) 具有高度序列同一性的因子IX多核苷酸进一步包括编码具有FIX-PPP-AA (SEQ ID NO:36) 的氨基酸序列的因子IX前原肽的因子IX前原肽多核苷酸。在一些实施方案中,因子IX前原肽多核苷酸与CS02-PPP-NA (SEQ ID NO:19) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX前原肽多核苷酸与CS03-PPP-NA (SEQ ID NO:20) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX前原肽多核苷酸与CS04-PPP-NA (SEQ ID NO:21) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX前原肽多核苷酸与CS05-PPP-NA (SEQ ID NO:22) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX前原肽多核苷酸与CS06-PPP-NA (SEQ ID NO:23) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。

[0170] 在一些实施方案中,所编码的因子IX多肽(例如由与CS02-FL-NA (SEQ ID NO:5) 具有高度序列同源性的多核苷酸编码的多肽)与野生型成熟因子IX单链多肽序列FIX-MP-AA (SEQ ID NO:10) 和/或成熟Padua (hFIX (R384L)) 单链序列FIXp-MP-AA (SEQ ID NO:12) 具有高度序列同一性。所编码的因子IX多肽应保留被活化为功能因子IXa蛋白(例如通过移除任何信号肽和原肽,和通过切除活化多肽)的能力。

[0171] 在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少85%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少90%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少95%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少96%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少97%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少98%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少99%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少99.5%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少99.9%的同一性。在一个实施方案中,所编码的因子IX多肽的序列为FIX-MP-AA (SEQ ID NO:10)。

[0172] 在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少85%的同一性并且在前原多肽的位置384 (例如成熟因子IX单链多肽FIXp-MP-AA (SEQ ID NO:12) 的位置338) 处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少90%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少95%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少96%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少97%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少98%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少99%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少99.5%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少99.9%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列为FIXp-MP-AA (SEQ ID NO:12)。

[0173] 在一个实施方案中,本文所提供的密码子改变的多核苷酸编码单链因子IX多肽,所述单链因子IX多肽包括轻链、重链和将轻链C端连接至重链N端的多肽接头。因子IX多肽的轻链由第一核苷酸序列编码,所述第一核苷酸序列与作为CS02-FL-NA (SEQ ID NO:5) 的编码因子IX轻链的部分的CS02-LC-NA (SEQ ID NO:42) 具有高度序列同一性。因子IX多肽的重链由第二核苷酸序列编码,所述第二核苷酸序列与作为CS02-FL-NA (SEQ ID NO:5) 的编码因子IX重链的部分的CS02-HC-NA (SEQ ID NO:41) 具有高度序列同一性。多肽接头包括因子XI切割位点,其允许体内成熟 (例如在前体单链因子IX多肽表达之后)。

[0174] 在一些实施方案中,第一和第二核苷酸序列分别与CS02-LC-NA和CS02-HC-NA (SEQ ID NO:42和41) 具有至少95%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS02-LC-NA和CS02-HC-NA (SEQ ID NO:42和41) 具有至少96%的序列同一性。在一些实

施方案中,第一和第二核苷酸序列分别与CS02-LC-NA和CS02-HC-NA(SEQ ID NO:42和41)具有至少97%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS02-LC-NA和CS02-HC-NA(SEQ ID NO:42和41)具有至少98%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS02-LC-NA和CS02-HC-NA(SEQ ID NO:42和41)具有至少99%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS02-LC-NA和CS02-HC-NA(SEQ ID NO:42和41)具有至少99.5%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS02-LC-NA和CS02-HC-NA(SEQ ID NO:42和41)具有至少99.9%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别为CS02-LC-NA和CS02-HC-NA(SEQ ID NO:42和41)。

[0175] 在一些实施方案中,因子IX构建体的多肽接头由第三核苷酸序列编码,所述第三核苷酸序列与作为编码野生型因子IX活化多肽(例如FIX-FL-AA(SEQ ID NO:2)的氨基酸192-226)的密码子改变的序列的CS02-AP-NA(SEQ ID NO:57)具有高度序列同一性。在一些实施方案中,第三核苷酸序列与CS02-AP-NA(SEQ ID NO:57)具有至少80%的同一性。在一些实施方案中,第三核苷酸序列与CS02-AP-NA(SEQ ID NO:57)具有至少90%的同一性。在一些实施方案中,第三核苷酸序列与CS02-AP-NA(SEQ ID NO:57)具有至少95%的同一性。在一些实施方案中,第三核苷酸序列与CS02-AP-NA(SEQ ID NO:57)具有至少96%的同一性。在一些实施方案中,第三核苷酸序列与CS02-AP-NA(SEQ ID NO:57)具有至少97%的同一性。在一些实施方案中,第三核苷酸序列与CS02-AP-NA(SEQ ID NO:57)具有至少98%的同一性。在一些实施方案中,第三核苷酸序列与CS02-AP-NA(SEQ ID NO:57)具有至少99%的同一性。在一些实施方案中,第三核苷酸序列为CS02-AP-NA(SEQ ID NO:57)。

[0176] 在一些实施方案中,所编码的因子IX多肽还包括信号肽(例如因子IX信号肽)和/或原肽(例如因子IX原肽)。在一些实施方案中,信号肽为野生型因子IX信号肽(FIX-SP-AA(SEQ ID NO:37))。在一些实施方案中,信号肽由与CS02-SP-NA(SEQ ID NO:25)具有高度序列同一性(例如至少95%、96%、97%、98%或99%)的密码子改变的多核苷酸序列编码。在一些实施方案中,原肽为野生型因子IX原肽(FIX-PP-AA(SEQ ID NO:38))。在一些实施方案中,原肽由与CS02-PP-NA(SEQ ID NO:31)具有高度序列同一性(例如至少95%、96%、97%、98%或99%)的密码子改变的多核苷酸序列编码。

[0177] 在一些实施方案中,所编码的因子IX多肽(例如由与CS02-LC-NA(SEQ ID NO:42)和CS02-HC-NA(SEQ ID NO:41)具有高度序列同源性的多核苷酸编码的多肽)与野生型成熟因子IX单链多肽序列FIX-MP-AA(SEQ ID NO:10)和/或成熟Padua(hFIX(R384L))单链序列FIXp-MP-AA(SEQ ID NO:12)具有高度序列同一性。所编码的因子IX多肽应保留被活化为功能因子IXa蛋白(例如通过移除任何信号肽和原肽,和通过切除活化多肽)的能力。

[0178] 在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少85%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少90%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少95%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少96%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少97%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少98%的同一性。在一个

实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少99%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少99.5%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少99.9%的同一性。在一个实施方案中,所编码的因子IX多肽的序列为FIX-MP-AA(SEQ ID NO:10)。

[0179] 在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA(SEQ ID NO:12)具有至少85%的同一性并且在前原多肽的位置384(例如成熟因子IX单链多肽FIXp-MP-AA(SEQ ID NO:12)的位置338)处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA(SEQ ID NO:12)具有至少90%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA(SEQ ID NO:12)具有至少95%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA(SEQ ID NO:12)具有至少96%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA(SEQ ID NO:12)具有至少97%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA(SEQ ID NO:12)具有至少98%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA(SEQ ID NO:12)具有至少99%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA(SEQ ID NO:12)具有至少99.5%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA(SEQ ID NO:12)具有至少99.9%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列为FIXp-MP-AA(SEQ ID NO:12)。

[0180] 在一些实施方案中,参考图1,提供包括结构A的自身互补型多核苷酸的核酸组合,其中所述多核苷酸的FIX编码序列部分包括编码成熟因子IX多肽的核酸序列,所述核酸序列与CS02-MP-NA(SEQ ID NO:13)具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX信号肽的核酸序列,所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-SP-NA(SEQ ID NO:24)、CS02-SP-NA(SEQ ID NO:25)、CS03-SP-NA(SEQ ID NO:26)、CS04-SP-NA(SEQ ID NO:27)、CS05-SP-NA(SEQ ID NO:28)和CS06-SP-NA(SEQ ID NO:29)。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX原肽的核酸序列(任选地与如上文所描述的因子IX信号肽的核酸序列组合),所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-PP-NA(SEQ ID NO:30)、CS02-PP-NA(SEQ ID NO:31)、CS03-PP-NA(SEQ ID NO:32)、CS04-PP-NA(SEQ ID NO:33)、CS05-PP-NA(SEQ ID NO:34)和CS06-PP-NA(SEQ ID NO:35)。在一些实施方案中,多核苷酸的FIX编码序列部分包括编码前原因子IX多肽的核酸序列,所述核酸序列与CS02-FL-NA(SEQ ID NO:5)具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。

[0181] 在一些实施方案中,参考图1,提供包括结构B的自身互补型多核苷酸的核酸组合,其中多核苷酸的FIX编码序列部分包括编码成熟因子IX多肽的核酸序列,所述核酸序列与CS02-MP-NA(SEQ ID NO:13)具有至少95%、96%、97%、98%、99%、99.5%、99.9%或

100%的同一性。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX信号肽的核酸序列,所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-SP-NA(SEQ ID NO:24)、CS02-SP-NA(SEQ ID NO:25)、CS03-SP-NA(SEQ ID NO:26)、CS04-SP-NA(SEQ ID NO:27)、CS05-SP-NA(SEQ ID NO:28)和CS06-SP-NA(SEQ ID NO:29)。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX原肽的核酸序列(任选地与如上文所描述的因子IX信号肽的核酸序列组合),所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-PP-NA(SEQ ID NO:30)、CS02-PP-NA(SEQ ID NO:31)、CS03-PP-NA(SEQ ID NO:32)、CS04-PP-NA(SEQ ID NO:33)、CS05-PP-NA(SEQ ID NO:34)和CS06-PP-NA(SEQ ID NO:35)。在一些实施方案中,多核苷酸的FIX编码序列部分包括编码前原因子IX多肽的核酸序列,所述核酸序列与CS02-FL-NA(SEQ ID NO:5)具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。

[0182] 在一些实施方案中,参考图1,提供包括结构C的多核苷酸(例如单链多核苷酸)的核酸组合物,其中多核苷酸的FIX编码序列部分包括编码成熟因子IX多肽的核酸序列,所述核酸序列与CS02-MP-NA(SEQ ID NO:13)具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX信号肽的核酸序列,所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-SP-NA(SEQ ID NO:24)、CS02-SP-NA(SEQ ID NO:25)、CS03-SP-NA(SEQ ID NO:26)、CS04-SP-NA(SEQ ID NO:27)、CS05-SP-NA(SEQ ID NO:28)和CS06-SP-NA(SEQ ID NO:29)。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX原肽的核酸序列(任选地与如上文所描述的因子IX信号肽的核酸序列组合),所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-PP-NA(SEQ ID NO:30)、CS02-PP-NA(SEQ ID NO:31)、CS03-PP-NA(SEQ ID NO:32)、CS04-PP-NA(SEQ ID NO:33)、CS05-PP-NA(SEQ ID NO:34)和CS06-PP-NA(SEQ ID NO:35)。在一些实施方案中,多核苷酸的FIX编码序列部分包括编码前原因子IX多肽的核酸序列,所述核酸序列与CS02-FL-NA(SEQ ID NO:5)具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。

[0183] 在一些实施方案中,参考图1,提供包括结构D的多核苷酸(例如单链多核苷酸)的核酸组合物,其中多核苷酸的FIX编码序列部分包括编码成熟因子IX多肽的核酸序列,所述核酸序列与CS02-MP-NA(SEQ ID NO:13)具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX信号肽的核酸序列,所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-SP-NA(SEQ ID NO:24)、CS02-SP-NA(SEQ ID NO:25)、CS03-SP-NA(SEQ ID NO:26)、CS04-SP-NA(SEQ ID NO:27)、CS05-SP-NA(SEQ ID NO:28)和CS06-SP-NA(SEQ ID NO:29)。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX原肽的核酸序列(任选地与如上文所描述的因子IX信号肽的核酸序列组合),所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-PP-NA(SEQ ID NO:30)、CS02-PP-NA(SEQ ID NO:31)、CS03-PP-NA(SEQ ID NO:32)、CS04-PP-NA(SEQ ID NO:33)、CS05-PP-NA(SEQ ID NO:34)和CS06-PP-NA(SEQ ID NO:35)。在一些实施方案中,多核苷酸的FIX编码序列部分包括编码前原因子IX多肽的核酸序列,所述核酸序列与CS02-

FL-NA (SEQ ID NO:5) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。

[0184] CS03密码子改变的多核苷酸

[0185] 在一个实施方案中,本文所提供的核酸组合物包括编码单链因子IX多肽的因子IX多核苷酸(例如密码子改变的多核苷酸),其中因子IX多核苷酸包括与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的核苷酸序列。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的因子IX多核苷酸的核苷酸序列与野生型因子IX编码序列 (FIX-FL-NA (SEQ ID NO:1)) 相比具有降低的GC含量。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的因子IX多核苷酸的核苷酸序列与野生型因子IX编码序列 (FIX-FL-NA (SEQ ID NO:1)) 相比具有减少的CpG二核苷酸数目。

[0186] 在一个具体实施方案中,密码子改变的多核苷酸的序列与CS03-FL-NA (SEQ ID NO:6) 具有至少95%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS03-FL-NA (SEQ ID NO:6) 具有至少96%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS03-FL-NA (SEQ ID NO:6) 具有至少97%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS03-FL-NA (SEQ ID NO:6) 具有至少98%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS03-FL-NA (SEQ ID NO:6) 具有至少99%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS03-FL-NA (SEQ ID NO:6) 具有至少99.5%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS03-FL-NA (SEQ ID NO:6) 具有至少99.9%的同一性。在另一个具体实施方案中,密码子改变的多核苷酸的序列为CS03-FL-NA (SEQ ID NO:6) 。

[0187] 在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于60%的GC含量。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于59%的GC含量。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于58%的GC含量。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于57%的GC含量。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于56%的GC含量。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于55%的GC含量。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于54%的GC含量。

[0188] 在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至60%的GC含量。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至59%的GC含量。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至58%的GC含量。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至57%的GC含量。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至56%的GC含量。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度

序列同一性的密码子改变的多核苷酸的序列具有50%至55%的GC含量。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至54%的GC含量。

[0189] 在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 1.0$ 的GC含量。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.8$ 的GC含量。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.6$ 的GC含量。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.5$ 的GC含量。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.4$ 的GC含量。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.3$ 的GC含量。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.2$ 的GC含量。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.1$ 的GC含量。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有53.8%的GC含量。

[0190] 在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过15个CpG二核苷酸。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过12个CpG二核苷酸。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过10个CpG二核苷酸。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过9个CpG二核苷酸。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过8个CpG二核苷酸。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过7个CpG二核苷酸。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过6个CpG二核苷酸。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过5个CpG二核苷酸。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过4个CpG二核苷酸。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过3个CpG二核苷酸。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过2个CpG二核苷酸。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过1个CpG二核苷酸。在一些实施方案中,与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同一性的密码子改变的多核苷酸的序列不具有CpG二核苷酸。

[0191] 在一些实施方案中,所编码的因子IX多肽(例如由与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同源性的多核苷酸编码的多肽) 与野生型因子IX前蛋白序列FIX-FL-AA (SEQ

ID NO:2) 和/或Padua (hFIX (R384L)) 前原蛋白序列FIXp-FL-AA (SEQ ID NO:4) 具有高度序列同一性。所编码的因子IX多肽应保留被活化为功能因子IXa蛋白 (例如通过移除信号肽和原肽, 和通过切除活化多肽) 的能力。

[0192] 在一个实施方案中, 所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少85%的同一性。在一个实施方案中, 所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少90%的同一性。在一个实施方案中, 所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少95%的同一性。在一个实施方案中, 所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少96%的同一性。在一个实施方案中, 所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少97%的同一性。在一个实施方案中, 所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少98%的同一性。在一个实施方案中, 所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少99%的同一性。在一个实施方案中, 所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少99.5%的同一性。在一个实施方案中, 所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少99.9%的同一性。在一个实施方案中, 所编码的因子IX多肽的序列为FIX-FL-AA (SEQ ID NO:2)。

[0193] 在一个实施方案中, 所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少85%的同一性并且在前原多肽的位置384 (例如成熟因子IX单链多肽FIXp-MP-AA (SEQ ID NO:12) 的位置338) 处包括亮氨酸。在一个实施方案中, 所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少90%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中, 所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少95%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中, 所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少96%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中, 所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少97%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中, 所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少98%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中, 所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少99%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中, 所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少99.5%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中, 所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少99.9%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中, 所编码的因子IX多肽的序列为FIXp-FL-AA (SEQ ID NO:4)。

[0194] 在一个实施方案中, 本文所提供的核酸组合物包括编码单链因子IX多肽 (例如具有丝氨酸蛋白酶活性) 的因子IX多核苷酸 (例如密码子改变的多核苷酸), 其中因子IX多核苷酸包括与CS03-MP-NA (SEQ ID NO:14) 具有高度序列同一性的核苷酸序列。在一些实施方案中, 与CS03-MP-NA (SEQ ID NO:14) 具有高度序列同一性的因子IX多核苷酸的核苷酸序列与野生型因子IX编码序列 (FIX-FL-NA (SEQ ID NO:1)) 相比具有降低的GC含量。在一些实施方案中, 与CS03-MP-NA (SEQ ID NO:14) 具有高度序列同一性的因子IX多核苷酸的核苷酸序列与野生型因子IX编码序列 (FIX-FL-NA (SEQ ID NO:1)) 相比具有减少的CpG二核苷酸数

目。

[0195] 在一个具体实施方案中,密码子改变的多核苷酸的序列与CS03-MP-NA (SEQ ID NO:14) 具有至少95%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS03-MP-NA (SEQ ID NO:14) 具有至少96%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS03-MP-NA (SEQ ID NO:14) 具有至少97%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS03-MP-NA (SEQ ID NO:14) 具有至少98%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS03-MP-NA (SEQ ID NO:14) 具有至少99%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS03-MP-NA (SEQ ID NO:14) 具有至少99.5%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS03-MP-NA (SEQ ID NO:14) 具有至少99.9%的同一性。在另一个具体实施方案中,密码子改变的多核苷酸的序列为CS03-MP-NA (SEQ ID NO:14)。

[0196] 在一些实施方案中,与CS03-MP-NA (SEQ ID NO:14) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于60%的GC含量。在一些实施方案中,与CS03-MP-NA (SEQ ID NO:14) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于59%的GC含量。在一些实施方案中,与CS03-MP-NA (SEQ ID NO:14) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于58%的GC含量。在一些实施方案中,与CS03-MP-NA (SEQ ID NO:14) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于57%的GC含量。在一些实施方案中,与CS03-MP-NA (SEQ ID NO:14) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于56%的GC含量。在一些实施方案中,与CS03-MP-NA (SEQ ID NO:14) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于55%的GC含量。在一些实施方案中,与CS03-MP-NA (SEQ ID NO:14) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于54%的GC含量。

[0197] 在一些实施方案中,与CS03-MP-NA (SEQ ID NO:14) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至60%的GC含量。在一些实施方案中,与CS03-MP-NA (SEQ ID NO:14) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至59%的GC含量。在一些实施方案中,与CS03-MP-NA (SEQ ID NO:14) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至58%的GC含量。在一些实施方案中,与CS03-MP-NA (SEQ ID NO:14) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至57%的GC含量。在一些实施方案中,与CS03-MP-NA (SEQ ID NO:14) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至56%的GC含量。在一些实施方案中,与CS03-MP-NA (SEQ ID NO:14) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至55%的GC含量。在一些实施方案中,与CS03-MP-NA (SEQ ID NO:14) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至54%的GC含量。

[0198] 在一些实施方案中,与CS03-MP-NA (SEQ ID NO:14) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 1.0$ 的GC含量。在一些实施方案中,与CS03-MP-NA (SEQ ID NO:14) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.8$ 的GC含量。在一些实施方案中,与CS03-MP-NA (SEQ ID NO:14) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.6$ 的GC含量。在一些实施方案中,与CS03-MP-NA (SEQ ID NO:14) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.5$ 的GC含量。

在一些实施方案中,与CS03-MP-NA(SEQ ID NO:14)具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.4$ 的GC含量。在一些实施方案中,与CS03-MP-NA(SEQ ID NO:14)具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.3$ 的GC含量。在一些实施方案中,与CS03-MP-NA(SEQ ID NO:14)具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.2$ 的GC含量。在一些实施方案中,与CS03-MP-NA(SEQ ID NO:14)具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.1$ 的GC含量。在一些实施方案中,与CS03-MP-NA(SEQ ID NO:14)具有高度序列同一性的密码子改变的多核苷酸的序列具有53.8%的GC含量。

[0199] 在一些实施方案中,与CS03-MP-NA(SEQ ID NO:14)具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过15个CpG二核苷酸。在一些实施方案中,与CS03-MP-NA(SEQ ID NO:14)具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过12个CpG二核苷酸。在一些实施方案中,与CS03-MP-NA(SEQ ID NO:14)具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过10个CpG二核苷酸。在一些实施方案中,与CS03-MP-NA(SEQ ID NO:14)具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过9个CpG二核苷酸。在一些实施方案中,与CS03-MP-NA(SEQ ID NO:14)具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过8个CpG二核苷酸。在一些实施方案中,与CS03-MP-NA(SEQ ID NO:14)具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过7个CpG二核苷酸。在一些实施方案中,与CS03-MP-NA(SEQ ID NO:14)具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过6个CpG二核苷酸。在一些实施方案中,与CS03-MP-NA(SEQ ID NO:14)具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过5个CpG二核苷酸。在一些实施方案中,与CS03-MP-NA(SEQ ID NO:14)具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过4个CpG二核苷酸。在一些实施方案中,与CS03-MP-NA(SEQ ID NO:14)具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过3个CpG二核苷酸。在一些实施方案中,与CS03-MP-NA(SEQ ID NO:14)具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过2个CpG二核苷酸。在一些实施方案中,与CS03-MP-NA(SEQ ID NO:14)具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过1个CpG二核苷酸。在一些实施方案中,与CS03-MP-NA(SEQ ID NO:14)具有高度序列同一性的密码子改变的多核苷酸的序列不具有CpG二核苷酸。

[0200] 在一些实施方案中,与CS03-MP-NA(SEQ ID NO:14)具有高度序列同一性的因子IX多核苷酸进一步包括编码具有FIX-SP-AA(SEQ ID NO:37)的氨基酸序列的因子IX信号肽的因子IX信号多核苷酸。在一些实施方案中,因子IX信号多核苷酸与CS02-SP-NA(SEQ ID NO:25)具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX信号多核苷酸与CS03-SP-NA(SEQ ID NO:26)具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX信号多核苷酸与CS04-SP-NA(SEQ ID NO:27)具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX信号多核苷酸与CS05-SP-NA(SEQ ID NO:28)具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX信号多核苷酸与CS06-SP-NA(SEQ ID NO:29)具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。

[0201] 在一些实施方案中,与CS03-MP-NA (SEQ ID NO:14) 具有高度序列同一性的因子IX多核苷酸进一步包括编码具有FIX-PP-AA (SEQ ID NO:38) 的氨基酸序列的因子IX原肽的因子IX原肽多核苷酸。在一些实施方案中,因子IX原肽多核苷酸与CS02-PP-NA (SEQ ID NO:31) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX原肽多核苷酸与CS03-PP-NA (SEQ ID NO:32) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX原肽多核苷酸与CS04-PP-NA (SEQ ID NO:33) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX原肽多核苷酸与CS05-PP-NA (SEQ ID NO:34) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX原肽多核苷酸与CS06-PP-NA (SEQ ID NO:35) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。

[0202] 在一些实施方案中,与CS03-MP-NA (SEQ ID NO:14) 具有高度序列同一性的因子IX多核苷酸进一步包括编码具有FIX-PPP-AA (SEQ ID NO:36) 的氨基酸序列的因子IX前原肽的因子IX前原肽多核苷酸。在一些实施方案中,因子IX前原肽多核苷酸与CS02-PPP-NA (SEQ ID NO:19) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX前原肽多核苷酸与CS03-PPP-NA (SEQ ID NO:20) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX前原肽多核苷酸与CS04-PPP-NA (SEQ ID NO:21) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX前原肽多核苷酸与CS05-PPP-NA (SEQ ID NO:22) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX前原肽多核苷酸与CS06-PPP-NA (SEQ ID NO:23) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。

[0203] 在一些实施方案中,所编码的因子IX多肽(例如由与CS03-FL-NA (SEQ ID NO:6) 具有高度序列同源性的多核苷酸编码的多肽)与野生型成熟因子IX单链多肽序列FIX-MP-AA (SEQ ID NO:10) 和/或成熟Padua (hFIX (R384L)) 单链序列FIXp-MP-AA (SEQ ID NO:12) 具有高度序列同一性。所编码的因子IX多肽应保留被活化为功能因子IXa蛋白(例如通过移除任何信号肽和原肽,和通过切除活化多肽)的能力。

[0204] 在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少85%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少90%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少95%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少96%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少97%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少98%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少99%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少99.5%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少99.9%的同一性。在一个实施方案中,所编码的因子IX多肽的序列为FIX-MP-AA (SEQ ID NO:10)。

[0205] 在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少85%的同一性并且在前原多肽的位置384 (例如成熟因子IX单链多肽FIXp-MP-AA (SEQ ID NO:12) 的位置338) 处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少90%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少95%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少96%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少97%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少98%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少99%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少99.5%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少99.9%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列为FIXp-MP-AA (SEQ ID NO:12)。

[0206] 在一个实施方案中,本文所提供的密码子改变的多核苷酸编码单链因子IX多肽,所述单链因子IX多肽包括轻链、重链和将轻链C端连接至重链N端的多肽接头。因子IX多肽的轻链由第一核苷酸序列编码,所述第一核苷酸序列与作为CS03-FL-NA (SEQ ID NO:6) 的编码因子IX轻链的部分的CS03-LC-NA (SEQ ID NO:44) 具有高度序列同一性。因子IX多肽的重链由第二核苷酸序列编码,所述第二核苷酸序列与作为CS03-FL-NA (SEQ ID NO:6) 的编码因子IX重链的部分的CS03-HC-NA (SEQ ID NO:43) 具有高度序列同一性。多肽接头包括因子XI切割位点,其允许体内成熟 (例如在前体单链因子IX多肽表达之后)。

[0207] 在一些实施方案中,第一和第二核苷酸序列分别与CS03-LC-NA (SEQ ID NO:44) 和CS03-HC-NA (SEQ ID NO:43) 具有至少95%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS03-LC-NA (SEQ ID NO:44) 和CS03-HC-NA (SEQ ID NO:43) 具有至少96%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS03-LC-NA (SEQ ID NO:44) 和CS03-HC-NA (SEQ ID NO:43) 具有至少97%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS03-LC-NA (SEQ ID NO:44) 和CS03-HC-NA (SEQ ID NO:43) 具有至少98%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS03-LC-NA (SEQ ID NO:44) 和CS03-HC-NA (SEQ ID NO:43) 具有至少99%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS03-LC-NA (SEQ ID NO:44) 和CS03-HC-NA (SEQ ID NO:43) 具有至少99.5%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS03-LC-NA (SEQ ID NO:44) 和CS03-HC-NA (SEQ ID NO:43) 具有至少99.9%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别为CS03-LC-NA (SEQ ID NO:44) 和CS03-HC-NA (SEQ ID NO:43)。

[0208] 在一些实施方案中,因子IX构建体的多肽接头由第三核苷酸序列编码,所述第三核苷酸序列与作为编码野生型因子IX活化多肽 (例如FIX-FL-AA (SEQ ID NO:2) 的氨基酸

192-226)的密码子改变的序列的CS03-AP-NA (SEQ ID NO:58) 具有高度序列同一性。在一些实施方案中,第三核苷酸序列与CS03-AP-NA (SEQ ID NO:58) 具有至少80%的同一性。在一些实施方案中,第三核苷酸序列与CS03-AP-NA (SEQ ID NO:58) 具有至少90%的同一性。在一些实施方案中,第三核苷酸序列与CS03-AP-NA (SEQ ID NO:58) 具有至少95%的同一性。在一些实施方案中,第三核苷酸序列与CS03-AP-NA (SEQ ID NO:58) 具有至少96%的同一性。在一些实施方案中,第三核苷酸序列与CS03-AP-NA (SEQ ID NO:58) 具有至少97%的同一性。在一些实施方案中,第三核苷酸序列与CS03-AP-NA (SEQ ID NO:58) 具有至少98%的同一性。在一些实施方案中,第三核苷酸序列与CS03-AP-NA (SEQ ID NO:58) 具有至少99%的同一性。在一些实施方案中,第三核苷酸序列为CS03-AP-NA (SEQ ID NO:58)。

[0209] 在一些实施方案中,所编码的因子IX多肽还包括信号肽(例如因子IX信号肽)和/或原肽(例如因子IX原肽)。在一些实施方案中,信号肽为野生型因子IX信号肽(FIX-SP-AA (SEQ ID NO:37))。在一些实施方案中,信号肽由与CS03-SP-NA (SEQ ID NO:26) 具有高度序列同一性(例如至少95%、96%、97%、98%或99%)的密码子改变的多核苷酸序列编码。在一些实施方案中,原肽为野生型因子IX原肽(FIX-PP-AA (SEQ ID NO:38))。在一些实施方案中,原肽由与CS03-PP-NA (SEQ ID NO:32) 具有高度序列同一性(例如至少95%、96%、97%、98%或99%)的密码子改变的多核苷酸序列编码。

[0210] 在一些实施方案中,所编码的因子IX多肽(例如由与CS03-LC-NA (SEQ ID NO:44) 和CS03-HC-NA (SEQ ID NO:43) 具有高度序列同源性的多核苷酸编码的多肽)与野生型成熟因子IX单链多肽序列FIX-MP-AA (SEQ ID NO:10) 和/或成熟Padua (hFIX (R384L)) 单链序列FIXp-MP-AA (SEQ ID NO:12) 具有高度序列同一性。所编码的因子IX多肽应保留被活化为功能因子IXa蛋白(例如通过移除任何信号肽和原肽,和通过切除活化多肽)的能力。

[0211] 在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少85%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少90%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少95%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少96%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少97%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少98%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少99%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少99.5%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少99.9%的同一性。在一个实施方案中,所编码的因子IX多肽的序列为FIX-MP-AA (SEQ ID NO:10)。

[0212] 在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少85%的同一性并且在前原多肽的位置384(例如成熟因子IX单链多肽FIXp-MP-AA (SEQ ID NO:12) 的位置338)处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少90%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少95%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码

的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少96%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少97%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少98%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少99%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少99.5%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少99.9%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列为FIXp-MP-AA (SEQ ID NO:12)。

[0213] 在一些实施方案中,参考图1,提供包括结构A的自身互补型多核苷酸的核酸组合物,其中多核苷酸的FIX编码序列部分包括编码成熟因子IX多肽的核酸序列,所述核酸序列与CS03-MP-NA (SEQ ID NO:14) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX信号肽的核酸序列,所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-SP-NA (SEQ ID NO:24)、CS02-SP-NA (SEQ ID NO:25)、CS03-SP-NA (SEQ ID NO:26)、CS04-SP-NA (SEQ ID NO:27)、CS05-SP-NA (SEQ ID NO:28) 和CS06-SP-NA (SEQ ID NO:29)。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX原肽的核酸序列(任选地与如上文所描述的因子IX信号肽的核酸序列组合),所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-PP-NA (SEQ ID NO:30)、CS02-PP-NA (SEQ ID NO:31)、CS03-PP-NA (SEQ ID NO:32)、CS04-PP-NA (SEQ ID NO:33)、CS05-PP-NA (SEQ ID NO:34) 和CS06-PP-NA (SEQ ID NO:35)。在一些实施方案中,多核苷酸的FIX编码序列部分包括编码前原因子IX多肽的核酸序列,所述核酸序列与CS03-FL-NA (SEQ ID NO:6) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。

[0214] 在一些实施方案中,参考图1,提供包括结构B的自身互补型多核苷酸的核酸组合物,其中多核苷酸的FIX编码序列部分包括编码成熟因子IX多肽的核酸序列,所述核酸序列与CS03-MP-NA (SEQ ID NO:14) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX信号肽的核酸序列,所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-SP-NA (SEQ ID NO:24)、CS02-SP-NA (SEQ ID NO:25)、CS03-SP-NA (SEQ ID NO:26)、CS04-SP-NA (SEQ ID NO:27)、CS05-SP-NA (SEQ ID NO:28) 和CS06-SP-NA (SEQ ID NO:29)。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX原肽的核酸序列(任选地与如上文所描述的因子IX信号肽的核酸序列组合),所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-PP-NA (SEQ ID NO:30)、CS02-PP-NA (SEQ ID NO:31)、CS03-PP-NA (SEQ ID NO:32)、CS04-PP-NA (SEQ ID NO:33)、CS05-PP-NA (SEQ ID NO:34) 和CS06-PP-NA (SEQ ID NO:35)。在一些实施方案中,多核苷酸的FIX编码序列部分包括编码前原因子IX多肽的核酸序列,所述核酸序列与CS03-FL-NA (SEQ ID NO:6) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。

[0215] 在一些实施方案中,参考图1,提供包括结构C的多核苷酸(例如单链多核苷酸)的核酸组合物,其中多核苷酸的FIX编码序列部分包括编码成熟因子IX多肽的核酸序列,所述核酸序列与CS03-MP-NA(SEQ ID NO:14)具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX信号肽的核酸序列,所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-SP-NA(SEQ ID NO:24)、CS02-SP-NA(SEQ ID NO:25)、CS03-SP-NA(SEQ ID NO:26)、CS04-SP-NA(SEQ ID NO:27)、CS05-SP-NA(SEQ ID NO:28)和CS06-SP-NA(SEQ ID NO:29)。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX原肽的核酸序列(任选地与如上文所描述的因子IX信号肽的核酸序列组合),所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-PP-NA(SEQ ID NO:30)、CS02-PP-NA(SEQ ID NO:31)、CS03-PP-NA(SEQ ID NO:32)、CS04-PP-NA(SEQ ID NO:33)、CS05-PP-NA(SEQ ID NO:34)和CS06-PP-NA(SEQ ID NO:35)。在一些实施方案中,多核苷酸的FIX编码序列部分包括编码前原因子IX多肽的核酸序列,所述核酸序列与CS03-FL-NA(SEQ ID NO:6)具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。

[0216] 在一些实施方案中,参考图1,提供包括结构D的多核苷酸(例如单链多核苷酸)的核酸组合物,其中多核苷酸的FIX编码序列部分包括编码成熟因子IX多肽的核酸序列,所述核酸序列与CS03-MP-NA(SEQ ID NO:14)具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX信号肽的核酸序列,所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-SP-NA(SEQ ID NO:24)、CS02-SP-NA(SEQ ID NO:25)、CS03-SP-NA(SEQ ID NO:26)、CS04-SP-NA(SEQ ID NO:27)、CS05-SP-NA(SEQ ID NO:28)和CS06-SP-NA(SEQ ID NO:29)。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX原肽的核酸序列(任选地与如上文所描述的因子IX信号肽的核酸序列组合),所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-PP-NA(SEQ ID NO:30)、CS02-PP-NA(SEQ ID NO:31)、CS03-PP-NA(SEQ ID NO:32)、CS04-PP-NA(SEQ ID NO:33)、CS05-PP-NA(SEQ ID NO:34)和CS06-PP-NA(SEQ ID NO:35)。在一些实施方案中,多核苷酸的FIX编码序列部分包括编码前原因子IX多肽的核酸序列,所述核酸序列与CS03-FL-NA(SEQ ID NO:6)具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。

[0217] CS04密码子改变的多核苷酸

[0218] 在一个实施方案中,本文所提供的核酸组合物包括编码单链因子IX多肽的因子IX多核苷酸(例如密码子改变的多核苷酸),其中因子IX多核苷酸包括与CS04-FL-NA(SEQ ID NO:7)具有高度序列同一性的核苷酸序列。在一些实施方案中,与CS04-FL-NA(SEQ ID NO:7)具有高度序列同一性的因子IX多核苷酸的核苷酸序列与野生型因子IX编码序列(FIX-FL-NA(SEQ ID NO:1))相比具有降低的GC含量。在一些实施方案中,与CS04-FL-NA(SEQ ID NO:7)具有高度序列同一性的因子IX多核苷酸的核苷酸序列与野生型因子IX编码序列(FIX-FL-NA(SEQ ID NO:1))相比具有减少的CpG二核苷酸数目。

[0219] 在一个具体实施方案中,密码子改变的多核苷酸的序列与CS04-FL-NA(SEQ ID

N0:7) 具有至少95%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS04-FL-NA(SEQ ID N0:7) 具有至少96%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS04-FL-NA(SEQ ID N0:7) 具有至少97%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS04-FL-NA(SEQ ID N0:7) 具有至少98%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS04-FL-NA(SEQ ID N0:7) 具有至少99%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS04-FL-NA(SEQ ID N0:7) 具有至少99.5%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS04-FL-NA(SEQ ID N0:7) 具有至少99.9%的同一性。在另一个具体实施方案中,密码子改变的多核苷酸的序列为CS04-FL-NA(SEQ ID N0:7)。

[0220] 在一些实施方案中,与CS04-FL-NA(SEQ ID N0:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于60%的GC含量。在一些实施方案中,与CS04-FL-NA(SEQ ID N0:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于59%的GC含量。在一些实施方案中,与CS04-FL-NA(SEQ ID N0:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于58%的GC含量。在一些实施方案中,与CS04-FL-NA(SEQ ID N0:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于57%的GC含量。在一些实施方案中,与CS04-FL-NA(SEQ ID N0:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于56%的GC含量。在一些实施方案中,与CS04-FL-NA(SEQ ID N0:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于55%的GC含量。在一些实施方案中,与CS04-FL-NA(SEQ ID N0:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于54%的GC含量。

[0221] 在一些实施方案中,与CS04-FL-NA(SEQ ID N0:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至60%的GC含量。在一些实施方案中,与CS04-FL-NA(SEQ ID N0:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至59%的GC含量。在一些实施方案中,与CS04-FL-NA(SEQ ID N0:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至58%的GC含量。在一些实施方案中,与CS04-FL-NA(SEQ ID N0:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至57%的GC含量。在一些实施方案中,与CS04-FL-NA(SEQ ID N0:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至56%的GC含量。在一些实施方案中,与CS04-FL-NA(SEQ ID N0:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至55%的GC含量。在一些实施方案中,与CS04-FL-NA(SEQ ID N0:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至54%的GC含量。

[0222] 在一些实施方案中,与CS04-FL-NA(SEQ ID N0:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 1.0$ 的GC含量。在一些实施方案中,与CS04-FL-NA(SEQ ID N0:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.8$ 的GC含量。在一些实施方案中,与CS04-FL-NA(SEQ ID N0:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.6$ 的GC含量。在一些实施方案中,与CS04-FL-NA(SEQ ID N0:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.5$ 的GC含量。在一些实施方案中,与CS04-FL-NA(SEQ ID N0:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.4$ 的GC含量。在一些实施方案中,与CS04-FL-NA(SEQ ID N0:7)

具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.3$ 的GC含量。在一些实施方案中,与CS04-FL-NA (SEQ ID NO:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.2$ 的GC含量。在一些实施方案中,与CS04-FL-NA (SEQ ID NO:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.1$ 的GC含量。在一些实施方案中,与CS04-FL-NA (SEQ ID NO:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有53.8%的GC含量。

[0223] 在一些实施方案中,与CS04-FL-NA (SEQ ID NO:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过15个CpG二核苷酸。在一些实施方案中,与CS04-FL-NA (SEQ ID NO:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过12个CpG二核苷酸。在一些实施方案中,与CS04-FL-NA (SEQ ID NO:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过10个CpG二核苷酸。在一些实施方案中,与CS04-FL-NA (SEQ ID NO:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过9个CpG二核苷酸。在一些实施方案中,与CS04-FL-NA (SEQ ID NO:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过8个CpG二核苷酸。在一些实施方案中,与CS04-FL-NA (SEQ ID NO:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过7个CpG二核苷酸。在一些实施方案中,与CS04-FL-NA (SEQ ID NO:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过6个CpG二核苷酸。在一些实施方案中,与CS04-FL-NA (SEQ ID NO:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过5个CpG二核苷酸。在一些实施方案中,与CS04-FL-NA (SEQ ID NO:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过4个CpG二核苷酸。在一些实施方案中,与CS04-FL-NA (SEQ ID NO:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过3个CpG二核苷酸。在一些实施方案中,与CS04-FL-NA (SEQ ID NO:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过2个CpG二核苷酸。在一些实施方案中,与CS04-FL-NA (SEQ ID NO:7) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过1个CpG二核苷酸。在一些实施方案中,与CS04-FL-NA (SEQ ID NO:7) 具有高度序列同一性的密码子改变的多核苷酸的序列不具有CpG二核苷酸。

[0224] 在一些实施方案中,所编码的因子IX多肽(例如由与CS04-FL-NA (SEQ ID NO:7) 具有高度序列同源性的多核苷酸编码的多肽)与野生型因子IX前原蛋白序列FIX-FL-AA (SEQ ID NO:2) 和/或Padua (hFIX (R384L)) 前原蛋白序列FIXp-FL-AA (SEQ ID NO:4) 具有高度序列同一性。所编码的因子IX多肽应保留被活化为功能因子IXa蛋白(例如通过移除信号肽和原肽,和通过切除活化多肽)的能力。

[0225] 在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少85%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少90%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少95%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少96%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少97%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少98%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少99%的同一性。在

一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少99.5% 的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少99.9%的同一性。在一个实施方案中,所编码的因子IX多肽的序列为FIX-FL-AA (SEQ ID NO:2)。

[0226] 在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少85%的同一性并且在前原多肽的位置384 (例如成熟因子IX单链多肽FIXp-MP-AA (SEQ ID NO:12) 的位置338) 处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少90%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少95%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少96%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少97%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少98%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少99%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少99.5%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少99.9%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列为FIXp-FL-AA (SEQ ID NO:4)。

[0227] 在一个实施方案中,本文所提供的核酸组合物包括编码单链因子IX多肽 (例如具有丝氨酸蛋白酶活性) 的因子IX多核苷酸 (例如密码子改变的多核苷酸), 其中因子IX多核苷酸包括与CS04-MP-NA (SEQ ID NO:15) 具有高度序列同一性的核苷酸序列。在一些实施方案中,与CS04-MP-NA (SEQ ID NO:15) 具有高度序列同一性的因子IX多核苷酸的核苷酸序列与野生型因子IX编码序列 (FIX-FL-NA (SEQ ID NO:1)) 相比具有降低的GC含量。在一些实施方案中,与CS04-MP-NA (SEQ ID NO:15) 具有高度序列同一性的因子IX多核苷酸的核苷酸序列与野生型因子IX编码序列 (FIX-FL-NA (SEQ ID NO:1)) 相比具有减少的CpG二核苷酸数目。

[0228] 在一个具体实施方案中,密码子改变的多核苷酸的序列与CS04-MP-NA (SEQ ID NO:15) 具有至少95%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS04-MP-NA (SEQ ID NO:15) 具有至少96%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS04-MP-NA (SEQ ID NO:15) 具有至少97%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS04-MP-NA (SEQ ID NO:15) 具有至少98%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS04-MP-NA (SEQ ID NO:15) 具有至少99%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS04-MP-NA (SEQ ID NO:15) 具有至少99.5%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS04-MP-NA (SEQ ID NO:15) 具有至少99.9%的同一性。在另一个具体实施方案中,密码子改变的多核苷酸的序列为CS04-MP-NA (SEQ ID NO:15)。

二核苷酸。在一些实施方案中,与CS04-MP-NA (SEQ ID NO:15) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过10个CpG二核苷酸。在一些实施方案中,与CS04-MP-NA (SEQ ID NO:15) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过9个CpG二核苷酸。在一些实施方案中,与CS04-MP-NA (SEQ ID NO:15) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过8个CpG二核苷酸。在一些实施方案中,与CS04-MP-NA (SEQ ID NO:15) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过7个CpG二核苷酸。在一些实施方案中,与CS04-MP-NA (SEQ ID NO:15) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过6个CpG二核苷酸。在一些实施方案中,与CS04-MP-NA (SEQ ID NO:15) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过5个CpG二核苷酸。在一些实施方案中,与CS04-MP-NA (SEQ ID NO:15) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过4个CpG二核苷酸。在一些实施方案中,与CS04-MP-NA (SEQ ID NO:15) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过3个CpG二核苷酸。在一些实施方案中,与CS04-MP-NA (SEQ ID NO:15) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过2个CpG二核苷酸。在一些实施方案中,与CS04-MP-NA (SEQ ID NO:15) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过1个CpG二核苷酸。在一些实施方案中,与CS04-MP-NA (SEQ ID NO:15) 具有高度序列同一性的密码子改变的多核苷酸的序列不具有CpG二核苷酸。

[0233] 在一些实施方案中,与CS04-MP-NA (SEQ ID NO:15) 具有高度序列同一性的因子IX多核苷酸进一步包括编码具有FIX-SP-AA (SEQ ID NO:37) 的氨基酸序列的因子IX信号肽的因子IX信号多核苷酸。在一些实施方案中,因子IX信号多核苷酸与CS02-SP-NA (SEQ ID NO:25) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX信号多核苷酸与CS03-SP-NA (SEQ ID NO:26) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX信号多核苷酸与CS04-SP-NA (SEQ ID NO:27) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX信号多核苷酸与CS05-SP-NA (SEQ ID NO:28) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX信号多核苷酸与CS06-SP-NA (SEQ ID NO:29) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。

[0234] 在一些实施方案中,与CS04-MP-NA (SEQ ID NO:15) 具有高度序列同一性的因子IX多核苷酸进一步包括编码具有FIX-PP-AA (SEQ ID NO:38) 的氨基酸序列的因子IX原肽的因子IX原肽多核苷酸。在一些实施方案中,因子IX原肽多核苷酸与CS02-PP-NA (SEQ ID NO:31) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX原肽多核苷酸与CS03-PP-NA (SEQ ID NO:32) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX原肽多核苷酸与CS04-PP-NA (SEQ ID NO:33) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX原肽多核苷酸与CS05-PP-NA (SEQ ID NO:34) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX原肽多核苷酸与CS06-PP-NA (SEQ ID NO:35) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。

[0235] 在一些实施方案中,与CS04-MP-NA(SEQ ID NO:15)具有高度序列同一性的因子IX多核苷酸进一步包括编码具有FIX-PPP-AA(SEQ ID NO:36)的氨基酸序列的因子IX前原肽的因子IX前原肽多核苷酸。在一些实施方案中,因子IX前原肽多核苷酸与CS02-PPP-NA(SEQ ID NO:19)具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX前原肽多核苷酸与CS03-PPP-NA(SEQ ID NO:20)具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX前原肽多核苷酸与CS04-PPP-NA(SEQ ID NO:21)具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX前原肽多核苷酸与CS05-PPP-NA(SEQ ID NO:22)具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX前原肽多核苷酸与CS06-PPP-NA(SEQ ID NO:23)具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。

[0236] 在一些实施方案中,所编码的因子IX多肽(例如由与CS04-FL-NA(SEQ ID NO:7)具有高度序列同源性的多核苷酸编码的多肽)与野生型成熟因子IX单链多肽序列FIX-MP-AA(SEQ ID NO:10)和/或成熟Padua(hFIX(R384L))单链序列FIXp-MP-AA(SEQ ID NO:12)具有高度序列同一性。所编码的因子IX多肽应保留被活化为功能因子IXa蛋白(例如通过移除任何信号肽和原肽,和通过切除活化多肽)的能力。

[0237] 在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少85%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少90%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少95%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少96%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少97%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少98%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少99%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少99.5%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少99.9%的同一性。在一个实施方案中,所编码的因子IX多肽的序列为FIX-MP-AA(SEQ ID NO:10)。

[0238] 在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA(SEQ ID NO:12)具有至少85%的同一性并且在前原多肽的位置384(例如成熟因子IX单链多肽FIXp-MP-AA(SEQ ID NO:12)的位置338)处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA(SEQ ID NO:12)具有至少90%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA(SEQ ID NO:12)具有至少95%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA(SEQ ID NO:12)具有至少96%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA(SEQ ID NO:12)具有至少97%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA(SEQ ID NO:12)具有至少98%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的

序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少99%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少99.5%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少99.9%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列为FIXp-MP-AA (SEQ ID NO:12)。

[0239] 在一个实施方案中,本文所提供的密码子改变的多核苷酸编码单链因子IX多肽,所述单链因子IX多肽包括轻链、重链和将轻链C端连接至重链N端的多肽接头。因子IX多肽的轻链由第一核苷酸序列编码,所述第一核苷酸序列与作为CS04-FL-NA (SEQ ID NO:7) 的编码因子IX轻链的部分的CS04-LC-NA (SEQ ID NO:46) 具有高度序列同一性。因子IX多肽的重链由第二核苷酸序列编码,所述第二核苷酸序列与作为CS04-FL-NA (SEQ ID NO:7) 的编码因子IX重链的部分的CS04-HC-NA (SEQ ID NO:45) 具有高度序列同一性。多肽接头包括因子XI切割位点,其允许体内成熟(例如在前体单链因子IX多肽表达之后)。

[0240] 在一些实施方案中,第一和第二核苷酸序列分别与CS04-LC-NA和CS04-HC-NA (SEQ ID NO:46和45) 具有至少95%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS04-LC-NA和CS04-HC-NA (SEQ ID NO:46和45) 具有至少96%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS04-LC-NA和CS04-HC-NA (SEQ ID NO:46和45) 具有至少97%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS04-LC-NA和CS04-HC-NA (SEQ ID NO:46和45) 具有至少98%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS04-LC-NA和CS04-HC-NA (SEQ ID NO:46和45) 具有至少99%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS04-LC-NA和CS04-HC-NA (SEQ ID NO:46和45) 具有至少99.5%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS04-LC-NA和CS04-HC-NA (SEQ ID NO:46和45) 具有至少99.9%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别为CS04-LC-NA和CS04-HC-NA (SEQ ID NO:46和45)。

[0241] 在一些实施方案中,因子IX构建体的多肽接头由第三核苷酸序列编码,所述第三核苷酸序列与作为编码野生型因子IX活化多肽(例如FIX-FL-AA (SEQ ID NO:2) 的氨基酸192-226) 的密码子改变的序列的CS04-AP-NA (SEQ ID NO:59) 具有高度序列同一性。在一些实施方案中,第三核苷酸序列与CS04-AP-NA (SEQ ID NO:59) 具有至少80%的同一性。在一些实施方案中,第三核苷酸序列与CS04-AP-NA (SEQ ID NO:59) 具有至少90%的同一性。在一些实施方案中,第三核苷酸序列与CS04-AP-NA (SEQ ID NO:59) 具有至少95%的同一性。在一些实施方案中,第三核苷酸序列与CS04-AP-NA (SEQ ID NO:59) 具有至少96%的同一性。在一些实施方案中,第三核苷酸序列与CS04-AP-NA (SEQ ID NO:59) 具有至少97%的同一性。在一些实施方案中,第三核苷酸序列与CS04-AP-NA (SEQ ID NO:59) 具有至少98%的同一性。在一些实施方案中,第三核苷酸序列与CS04-AP-NA (SEQ ID NO:59) 具有至少99%的同一性。在一些实施方案中,第三核苷酸序列为CS04-AP-NA (SEQ ID NO:59)。

[0242] 在一些实施方案中,所编码的因子IX多肽还包括信号肽(例如因子IX信号肽)和/或原肽(例如因子IX原肽)。在一些实施方案中,信号肽为野生型因子IX信号肽(FIX-SP-AA (SEQ ID NO:37))。在一些实施方案中,信号肽由与CS04-SP-NA (SEQ ID NO:27) 具有高度序

列同一性(例如至少95%、96%、97%、98%或99%)的密码子改变的多核苷酸序列编码。在一些实施方案中,原肽为野生型因子IX原肽(FIX-PP-AA(SEQ ID NO:38))。在一些实施方案中,原肽由与CS04-PP-NA(SEQ ID NO:33)具有高度序列同一性(例如至少95%、96%、97%、98%或99%)的密码子改变的多核苷酸序列编码。

[0243] 在一些实施方案中,所编码的因子IX多肽(例如由与CS04-LC-NA(SEQ ID NO:46)和CS04-HC-NA(SEQ ID NO:45)具有高度序列同源性的多核苷酸编码的多肽)与野生型成熟因子IX单链多肽序列FIX-MP-AA(SEQ ID NO:10)和/或成熟Padua(hFIX(R384L))单链序列FIXp-MP-AA(SEQ ID NO:12)具有高度序列同一性。所编码的因子IX多肽应保留被活化为功能因子IXa蛋白(例如通过移除任何信号肽和原肽,和通过切除活化多肽)的能力。

[0244] 在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少85%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少90%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少95%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少96%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少97%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少98%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少99%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少99.5%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA(SEQ ID NO:10)具有至少99.9%的同一性。在一个实施方案中,所编码的因子IX多肽的序列为FIX-MP-AA(SEQ ID NO:10)。

[0245] 在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA(SEQ ID NO:12)具有至少85%的同一性并且在前原多肽的位置384(例如成熟因子IX单链多肽FIXp-MP-AA(SEQ ID NO:12)的位置338)处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA(SEQ ID NO:12)具有至少90%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA(SEQ ID NO:12)具有至少95%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA(SEQ ID NO:12)具有至少96%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA(SEQ ID NO:12)具有至少97%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA(SEQ ID NO:12)具有至少98%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA(SEQ ID NO:12)具有至少99%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA(SEQ ID NO:12)具有至少99.5%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA(SEQ ID NO:12)具有至少99.9%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列为FIXp-MP-AA(SEQ ID NO:12)。

[0246] 在一些实施方案中,参考图1,提供包括结构A的自身互补型多核苷酸的核酸组合

物,其中多核苷酸的FIX编码序列部分包括编码成熟因子IX多肽的核酸序列,其与CS04-MP-NA (SEQ ID NO:15) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX信号肽的核酸序列,所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-SP-NA (SEQ ID NO:24)、CS02-SP-NA (SEQ ID NO:25)、CS03-SP-NA (SEQ ID NO:26)、CS04-SP-NA (SEQ ID NO:27)、CS05-SP-NA (SEQ ID NO:28) 和CS06-SP-NA (SEQ ID NO:29)。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX原肽的核酸序列(任选地与如上文所描述的因子IX信号肽的核酸序列组合),所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-PP-NA (SEQ ID NO:30)、CS02-PP-NA (SEQ ID NO:31)、CS03-PP-NA (SEQ ID NO:32)、CS04-PP-NA (SEQ ID NO:33)、CS05-PP-NA (SEQ ID NO:34) 和CS06-PP-NA (SEQ ID NO:35)。在一些实施方案中,多核苷酸的FIX编码序列部分包括编码前原因子IX多肽的核酸序列,所述核酸序列与CS04-FL-NA (SEQ ID NO:7) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。

[0247] 在一些实施方案中,参考图1,提供包括结构B的自身互补型多核苷酸的核酸组合物,其中多核苷酸的FIX编码序列部分包括编码成熟因子IX多肽的核酸序列,所述核酸序列与CS04-MP-NA (SEQ ID NO:15) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX信号肽的核酸序列,所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-SP-NA (SEQ ID NO:24)、CS02-SP-NA (SEQ ID NO:25)、CS03-SP-NA (SEQ ID NO:26)、CS04-SP-NA (SEQ ID NO:27)、CS05-SP-NA (SEQ ID NO:28) 和CS06-SP-NA (SEQ ID NO:29)。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX原肽的核酸序列(任选地与如上文所描述的因子IX信号肽的核酸序列组合),所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-PP-NA (SEQ ID NO:30)、CS02-PP-NA (SEQ ID NO:31)、CS03-PP-NA (SEQ ID NO:32)、CS04-PP-NA (SEQ ID NO:33)、CS05-PP-NA (SEQ ID NO:34) 和CS06-PP-NA (SEQ ID NO:35)。在一些实施方案中,多核苷酸的FIX编码序列部分包括编码前原因子IX多肽的核酸序列,所述核酸序列与CS04-FL-NA (SEQ ID NO:7) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。

[0248] 在一些实施方案中,参考图1,提供包括结构C的多核苷酸(例如单链多核苷酸)的核酸组合物,其中多核苷酸的FIX编码序列部分包括编码成熟因子IX多肽的核酸序列,所述核酸序列与CS04-MP-NA (SEQ ID NO:15) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX信号肽的核酸序列,所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-SP-NA (SEQ ID NO:24)、CS02-SP-NA (SEQ ID NO:25)、CS03-SP-NA (SEQ ID NO:26)、CS04-SP-NA (SEQ ID NO:27)、CS05-SP-NA (SEQ ID NO:28) 和CS06-SP-NA (SEQ ID NO:29)。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX原肽的核酸序列(任选地与如上文所描述的因子IX信号肽的核酸序列组合),所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-PP-NA (SEQ ID NO:30)、CS02-PP-NA (SEQ ID NO:31)、CS03-PP-NA (SEQ ID NO:32)、CS04-PP-NA (SEQ ID NO:33)、CS05-PP-NA (SEQ ID NO:34) 和CS06-PP-NA (SEQ ID NO:35)。在一些实施方案中,多

核苷酸的FIX编码序列部分包括编码前原因子IX多肽的核酸序列,所述核酸序列与CS04-FL-NA(SEQ ID NO:7)具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。

[0249] 在一些实施方案中,参考图1,提供包括结构D的多核苷酸(例如单链多核苷酸)的核酸组合物,其中多核苷酸的FIX编码序列部分包括编码成熟因子IX多肽的核酸序列,所述核酸序列与CS04-MP-NA(SEQ ID NO:15)具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX信号肽的核酸序列,所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-SP-NA(SEQ ID NO:24)、CS02-SP-NA(SEQ ID NO:25)、CS03-SP-NA(SEQ ID NO:26)、CS04-SP-NA(SEQ ID NO:27)、CS05-SP-NA(SEQ ID NO:28)和CS06-SP-NA(SEQ ID NO:29)。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX原肽的核酸序列(任选地与如上文所描述的因子IX信号肽的核酸序列组合),所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-PP-NA(SEQ ID NO:30)、CS02-PP-NA(SEQ ID NO:31)、CS03-PP-NA(SEQ ID NO:32)、CS04-PP-NA(SEQ ID NO:33)、CS05-PP-NA(SEQ ID NO:34)和CS06-PP-NA(SEQ ID NO:35)。在一些实施方案中,多核苷酸的FIX编码序列部分包括编码前原因子IX多肽的核酸序列,所述核酸序列与CS04-FL-NA(SEQ ID NO:7)具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。

[0250] CS05密码子改变的多核苷酸

[0251] 在一个实施方案中,本文所提供的核酸组合物包括编码单链因子IX多肽的因子IX多核苷酸(例如密码子改变的多核苷酸),其中因子IX多核苷酸包括与CS05-FL-NA(SEQ ID NO:8)具有高度序列同一性的核苷酸序列。在一些实施方案中,与CS05-FL-NA(SEQ ID NO:8)具有高度序列同一性的因子IX多核苷酸的核苷酸序列与野生型因子IX编码序列(FIX-FL-NA(SEQ ID NO:1))相比具有降低的GC含量。在一些实施方案中,与CS05-FL-NA(SEQ ID NO:8)具有高度序列同一性的因子IX多核苷酸的核苷酸序列与野生型因子IX编码序列(FIX-FL-NA(SEQ ID NO:1))相比具有减少的CpG二核苷酸数目。

[0252] 在一个具体实施方案中,密码子改变的多核苷酸的序列与CS05-FL-NA(SEQ ID NO:8)具有至少95%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS05-FL-NA(SEQ ID NO:8)具有至少96%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS05-FL-NA(SEQ ID NO:8)具有至少97%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS05-FL-NA(SEQ ID NO:8)具有至少98%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS05-FL-NA(SEQ ID NO:8)具有至少99%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS05-FL-NA(SEQ ID NO:8)具有至少99.5%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS05-FL-NA(SEQ ID NO:8)具有至少99.9%的同一性。在另一个具体实施方案中,密码子改变的多核苷酸的序列为CS05-FL-NA(SEQ ID NO:8)。

[0253] 在一些实施方案中,与CS05-FL-NA(SEQ ID NO:8)具有高度序列同一性的密码子改变的多核苷酸的序列具有低于60%的GC含量。在一些实施方案中,与CS05-FL-NA(SEQ ID NO:8)具有高度序列同一性的密码子改变的多核苷酸的序列具有低于59%的GC含量。在一

核苷酸。在一些实施方案中,与CS05-FL-NA (SEQ ID NO:8) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过8个CpG二核苷酸。在一些实施方案中,与CS05-FL-NA (SEQ ID NO:8) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过7个CpG二核苷酸。在一些实施方案中,与CS05-FL-NA (SEQ ID NO:8) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过6个CpG二核苷酸。在一些实施方案中,与CS05-FL-NA (SEQ ID NO:8) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过5个CpG二核苷酸。在一些实施方案中,与CS05-FL-NA (SEQ ID NO:8) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过4个CpG二核苷酸。在一些实施方案中,与CS05-FL-NA (SEQ ID NO:8) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过3个CpG二核苷酸。在一些实施方案中,与CS05-FL-NA (SEQ ID NO:8) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过2个CpG二核苷酸。在一些实施方案中,与CS05-FL-NA (SEQ ID NO:8) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过1个CpG二核苷酸。在一些实施方案中,与CS05-FL-NA (SEQ ID NO:8) 具有高度序列同一性的密码子改变的多核苷酸的序列不具有CpG二核苷酸。

[0257] 在一些实施方案中,所编码的因子IX多肽(例如由与CS05-FL-NA (SEQ ID NO:8) 具有高度序列同源性的多核苷酸编码的多肽)与野生型因子IX前原蛋白序列FIX-FL-AA (SEQ ID NO:2) 和/或Padua (hFIX (R384L)) 前原蛋白序列FIXp-FL-AA (SEQ ID NO:4) 具有高度序列同一性。所编码的因子IX多肽应保留被活化为功能因子IXa蛋白(例如通过移除信号肽和原肽,和通过切除活化多肽)的能力。

[0258] 在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少85%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少90%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少95%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少96%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少97%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少98%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少99%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少99.5%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA (SEQ ID NO:2) 具有至少99.9%的同一性。在一个实施方案中,所编码的因子IX多肽的序列为FIX-FL-AA (SEQ ID NO:2)。

[0259] 在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少85%的同一性并且在前原多肽的位置384(例如成熟因子IX单链多肽FIXp-MP-AA (SEQ ID NO:12) 的位置338)处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少90%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少95%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少96%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ

ID NO:4) 具有至少97%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少98%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少99%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少99.5%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA (SEQ ID NO:4) 具有至少99.9%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列为FIXp-FL-AA (SEQ ID NO:4)。

[0260] 在一个实施方案中,本文所提供的核酸组合物包括编码单链因子IX多肽(例如具有丝氨酸蛋白酶活性)的因子IX多核苷酸(例如密码子改变的多核苷酸),其中因子IX多核苷酸包括与CS05-MP-NA (SEQ ID NO:16) 具有高度序列同一性的核苷酸序列。在一些实施方案中,与CS05-MP-NA (SEQ ID NO:16) 具有高度序列同一性的因子IX多核苷酸的核苷酸序列与野生型因子IX编码序列 (FIX-FL-NA (SEQ ID NO:1)) 相比具有降低的GC含量。在一些实施方案中,与CS05-MP-NA (SEQ ID NO:16) 具有高度序列同一性的因子IX多核苷酸的核苷酸序列与野生型因子IX编码序列 (FIX-FL-NA (SEQ ID NO:1)) 相比具有减少的CpG二核苷酸数目。

[0261] 在一个具体实施方案中,密码子改变的多核苷酸的序列与CS05-MP-NA (SEQ ID NO:16) 具有至少95%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS05-MP-NA (SEQ ID NO:16) 具有至少96%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS05-MP-NA (SEQ ID NO:16) 具有至少97%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS05-MP-NA (SEQ ID NO:16) 具有至少98%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS05-MP-NA (SEQ ID NO:16) 具有至少99%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS05-MP-NA (SEQ ID NO:16) 具有至少99.5%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS05-MP-NA (SEQ ID NO:16) 具有至少99.9%的同一性。在另一个具体实施方案中,密码子改变的多核苷酸的序列为CS05-MP-NA (SEQ ID NO:16)。

[0262] 在一些实施方案中,与CS05-MP-NA (SEQ ID NO:16) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于60%的GC含量。在一些实施方案中,与CS05-MP-NA (SEQ ID NO:16) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于59%的GC含量。在一些实施方案中,与CS05-MP-NA (SEQ ID NO:16) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于58%的GC含量。在一些实施方案中,与CS05-MP-NA (SEQ ID NO:16) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于57%的GC含量。在一些实施方案中,与CS05-MP-NA (SEQ ID NO:16) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于56%的GC含量。在一些实施方案中,与CS05-MP-NA (SEQ ID NO:16) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于55%的GC含量。在一些实施方案中,与CS05-MP-NA (SEQ ID NO:16) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于54%的GC含量。

[0263] 在一些实施方案中,与CS05-MP-NA (SEQ ID NO:16) 具有高度序列同一性的密码子

二核苷酸。在一些实施方案中,与CS05-MP-NA (SEQ ID NO:16) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过2个CpG二核苷酸。在一些实施方案中,与CS05-MP-NA (SEQ ID NO:16) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过1个CpG二核苷酸。在一些实施方案中,与CS05-MP-NA (SEQ ID NO:16) 具有高度序列同一性的密码子改变的多核苷酸的序列不具有CpG二核苷酸。

[0266] 在一些实施方案中,与CS05-MP-NA (SEQ ID NO:16) 具有高度序列同一性的因子IX多核苷酸进一步包括编码具有FIX-SP-AA (SEQ ID NO:37) 的氨基酸序列的因子IX信号肽的因子IX信号多核苷酸。在一些实施方案中,因子IX信号多核苷酸与CS02-SP-NA (SEQ ID NO:25) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX信号多核苷酸与CS03-SP-NA (SEQ ID NO:26) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX信号多核苷酸与CS04-SP-NA (SEQ ID NO:27) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX信号多核苷酸与CS05-SP-NA (SEQ ID NO:28) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX信号多核苷酸与CS06-SP-NA (SEQ ID NO:29) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。

[0267] 在一些实施方案中,与CS05-MP-NA (SEQ ID NO:16) 具有高度序列同一性的因子IX多核苷酸进一步包括编码具有FIX-PP-AA (SEQ ID NO:38) 的氨基酸序列的因子IX原肽的因子IX原肽多核苷酸。在一些实施方案中,因子IX原肽多核苷酸与CS02-PP-NA (SEQ ID NO:31) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX原肽多核苷酸与CS03-PP-NA (SEQ ID NO:32) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX原肽多核苷酸与CS04-PP-NA (SEQ ID NO:33) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX原肽多核苷酸与CS05-PP-NA (SEQ ID NO:34) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX原肽多核苷酸与CS06-PP-NA (SEQ ID NO:35) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。

[0268] 在一些实施方案中,与CS05-MP-NA (SEQ ID NO:16) 具有高度序列同一性的因子IX多核苷酸进一步包括编码具有FIX-PPP-AA (SEQ ID NO:36) 的氨基酸序列的因子IX前原肽的因子IX前原肽多核苷酸。在一些实施方案中,因子IX前原肽多核苷酸与CS02-PPP-NA (SEQ ID NO:19) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX前原肽多核苷酸与CS03-PPP-NA (SEQ ID NO:20) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX前原肽多核苷酸与CS04-PPP-NA (SEQ ID NO:21) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX前原肽多核苷酸与CS05-PPP-NA (SEQ ID NO:22) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX前原肽多核苷酸与CS06-PPP-NA (SEQ ID NO:23) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。

[0269] 在一些实施方案中,所编码的因子IX多肽(例如由与CS05-FL-NA (SEQ ID NO:8) 具

有高度序列同源性的多核苷酸编码的多肽)与野生型成熟因子IX单链多肽序列FIX-MP-AA (SEQ ID NO:10)和/或成熟Padua (hFIX (R384L))单链序列FIXp-MP-AA (SEQ ID NO:12)具有高度序列同一性。所编码的因子IX多肽应保留被活化为功能因子IXa蛋白(例如通过移除任何信号肽和原肽,和通过切除活化多肽)的能力。

[0270] 在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10)具有至少85%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10)具有至少90%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10)具有至少95%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10)具有至少96%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10)具有至少97%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10)具有至少98%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10)具有至少99%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10)具有至少99.5%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10)具有至少99.9%的同一性。在一个实施方案中,所编码的因子IX多肽的序列为FIX-MP-AA (SEQ ID NO:10)。

[0271] 在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12)具有至少85%的同一性并且在前原多肽的位置384(例如成熟因子IX单链多肽FIXp-MP-AA (SEQ ID NO:12)的位置338)处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12)具有至少90%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12)具有至少95%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12)具有至少96%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12)具有至少97%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12)具有至少98%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12)具有至少99%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12)具有至少99.5%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12)具有至少99.9%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列为FIXp-MP-AA (SEQ ID NO:12)。

[0272] 在一个实施方案中,本文所提供的密码子改变的多核苷酸编码单链因子IX多肽,所述单链因子IX多肽包括轻链、重链和将轻链C端连接至重链N端的多肽接头。因子IX多肽的轻链由第一核苷酸序列编码,所述第一核苷酸序列与作为CS05-FL-NA (SEQ ID NO:8)的编码因子IX轻链的部分的CS05-LC-NA (SEQ ID NO:48)具有高度序列同一性。因子IX多肽的重链由第二核苷酸序列编码,所述第二核苷酸序列与作为CS05-FL-NA (SEQ ID NO:8)的编码因子IX重链的部分的CS05-HC-NA (SEQ ID NO:47)具有高度序列同一性。多肽接头包括因

子XI切割位点,其允许体内成熟(例如在前体单链因子IX多肽表达之后)。

[0273] 在一些实施方案中,第一和第二核苷酸序列分别与CS05-LC-NA和CS05-HC-NA (SEQ ID NO:48和47) 具有至少95%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS05-LC-NA和CS05-HC-NA (SEQ ID NO:48和47) 具有至少96%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS05-LC-NA和CS05-HC-NA (SEQ ID NO:48和47) 具有至少97%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS05-LC-NA和CS05-HC-NA (SEQ ID NO:48和47) 具有至少98%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS05-LC-NA和CS05-HC-NA (SEQ ID NO:48和47) 具有至少99%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS05-LC-NA和CS05-HC-NA (SEQ ID NO:48和47) 具有至少99.5%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS05-LC-NA和CS05-HC-NA (SEQ ID NO:48和47) 具有至少99.9%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别为CS05-LC-NA和CS05-HC-NA (SEQ ID NO:48和47)。

[0274] 在一些实施方案中,因子IX构建体的多肽接头由第三核苷酸序列编码,所述第三核苷酸序列与作为编码野生型因子IX活化多肽(例如FIX-FL-AA (SEQ ID NO:2) 的氨基酸192-226) 的密码子改变的序列的CS05-AP-NA (SEQ ID NO:60) 具有高度序列同一性。在一些实施方案中,第三核苷酸序列与CS05-AP-NA (SEQ ID NO:60) 具有至少80%的同一性。在一些实施方案中,第三核苷酸序列与CS05-AP-NA (SEQ ID NO:60) 具有至少90%的同一性。在一些实施方案中,第三核苷酸序列与CS05-AP-NA (SEQ ID NO:60) 具有至少95%的同一性。在一些实施方案中,第三核苷酸序列与CS05-AP-NA (SEQ ID NO:60) 具有至少96%的同一性。在一些实施方案中,第三核苷酸序列与CS05-AP-NA (SEQ ID NO:60) 具有至少97%的同一性。在一些实施方案中,第三核苷酸序列与CS05-AP-NA (SEQ ID NO:60) 具有至少98%的同一性。在一些实施方案中,第三核苷酸序列与CS05-AP-NA (SEQ ID NO:60) 具有至少99%的同一性。在一些实施方案中,第三核苷酸序列为CS05-AP-NA (SEQ ID NO:60)。

[0275] 在一些实施方案中,所编码的因子IX多肽还包括信号肽(例如因子IX信号肽)和/或原肽(例如因子IX原肽)。在一些实施方案中,信号肽为野生型因子IX信号肽(FIX-SP-AA (SEQ ID NO:37))。在一些实施方案中,信号肽由与CS05-SP-NA (SEQ ID NO:28) 具有高度序列同一性(例如至少95%、96%、97%、98%或99%) 的密码子改变的多核苷酸序列编码。在一些实施方案中,原肽为野生型因子IX原肽(FIX-PP-AA (SEQ ID NO:38))。在一些实施方案中,原肽由与CS05-PP-NA (SEQ ID NO:34) 具有高度序列同一性(例如至少95%、96%、97%、98%或99%) 的密码子改变的多核苷酸序列编码。

[0276] 在一些实施方案中,所编码的因子IX多肽(例如由与CS05-LC-NA (SEQ ID NO:48) 和CS05-HC-NA (SEQ ID NO:47) 具有高度序列同源性的多核苷酸编码的多肽) 与野生型成熟因子IX单链多肽序列FIX-MP-AA (SEQ ID NO:10) 和/或成熟Padua (hFIX (R384L)) 单链序列FIXp-MP-AA (SEQ ID NO:12) 具有高度序列同一性。所编码的因子IX多肽应保留被活化为功能因子IXa蛋白(例如通过移除任何信号肽和原肽,和通过切除活化多肽) 的能力。

[0277] 在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少85%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少90%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-

MP-AA (SEQ ID NO:10) 具有至少95%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少96%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少97%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少98%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少99%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少99.5%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少99.9%的同一性。在一个实施方案中,所编码的因子IX多肽的序列为FIX-MP-AA (SEQ ID NO:10)。

[0278] 在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少85%的同一性并且在前原多肽的位置384 (例如成熟因子IX单链多肽FIXp-MP-AA (SEQ ID NO:12) 的位置338) 处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少90%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少95%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少96%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少97%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少98%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少99%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少99.5%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少99.9%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列为FIXp-MP-AA (SEQ ID NO:12)。

[0279] 在一些实施方案中,参考图1,提供包括结构A的自身互补型多核苷酸的核酸组合,其中多核苷酸的FIX编码序列部分包括编码成熟因子IX多肽的核酸序列,所述核酸序列与CS03-MP-NA (SEQ ID NO:14) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX信号肽的核酸序列,所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-SP-NA (SEQ ID NO:24)、CS02-SP-NA (SEQ ID NO:25)、CS03-SP-NA (SEQ ID NO:26)、CS04-SP-NA (SEQ ID NO:27)、CS05-SP-NA (SEQ ID NO:28) 和CS06-SP-NA (SEQ ID NO:29)。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX原肽的核酸序列 (任选地与如上文所描述的因子IX信号肽的核酸序列组合),所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-PP-NA (SEQ ID NO:30)、CS02-PP-NA (SEQ ID NO:31)、CS03-PP-NA (SEQ ID NO:32)、CS04-PP-NA (SEQ ID NO:33)、CS05-PP-NA (SEQ ID NO:34) 和CS06-PP-NA (SEQ ID NO:35)。在一些实施方案中,多核苷酸的FIX编码序列部分包括编码前原因子IX多肽的核酸序列,所述核酸序列与CS05-FL-

NA (SEQ ID NO:8) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。

[0280] 在一些实施方案中,参考图1,提供包括结构B的自身互补型多核苷酸的核酸组合,其中多核苷酸的FIX编码序列部分包括编码成熟因子IX多肽的核酸序列,所述核酸序列与CS05-MP-NA (SEQ ID NO:16) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX信号肽的核酸序列,所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-SP-NA (SEQ ID NO:24)、CS02-SP-NA (SEQ ID NO:25)、CS03-SP-NA (SEQ ID NO:26)、CS04-SP-NA (SEQ ID NO:27)、CS05-SP-NA (SEQ ID NO:28) 和CS06-SP-NA (SEQ ID NO:29)。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX原肽的核酸序列(任选地与如上文所描述的因子IX信号肽的核酸序列组合),所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-PP-NA (SEQ ID NO:30)、CS02-PP-NA (SEQ ID NO:31)、CS03-PP-NA (SEQ ID NO:32)、CS04-PP-NA (SEQ ID NO:33)、CS05-PP-NA (SEQ ID NO:34) 和CS06-PP-NA (SEQ ID NO:35)。在一些实施方案中,多核苷酸的FIX编码序列部分包括编码前原因子IX多肽的核酸序列,所述核酸序列与CS05-FL-NA (SEQ ID NO:8) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。

[0281] 在一些实施方案中,参考图1,提供包括结构C的多核苷酸(例如单链多核苷酸)的核酸组合,其中多核苷酸的FIX编码序列部分包括编码成熟因子IX多肽的核酸序列,所述核酸序列与CS05-MP-NA (SEQ ID NO:16) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX信号肽的核酸序列,所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-SP-NA (SEQ ID NO:24)、CS02-SP-NA (SEQ ID NO:25)、CS03-SP-NA (SEQ ID NO:26)、CS04-SP-NA (SEQ ID NO:27)、CS05-SP-NA (SEQ ID NO:28) 和CS06-SP-NA (SEQ ID NO:29)。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX原肽的核酸序列(任选地与如上文所描述的因子IX信号肽的核酸序列组合),所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-PP-NA (SEQ ID NO:30)、CS02-PP-NA (SEQ ID NO:31)、CS03-PP-NA (SEQ ID NO:32)、CS04-PP-NA (SEQ ID NO:33)、CS05-PP-NA (SEQ ID NO:34) 和CS06-PP-NA (SEQ ID NO:35)。在一些实施方案中,多核苷酸的FIX编码序列部分包括编码前原因子IX多肽的核酸序列,所述核酸序列与CS05-FL-NA (SEQ ID NO:8) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。

[0282] 在一些实施方案中,参考图1,提供包括结构D的多核苷酸(例如单链多核苷酸)的核酸组合,其中多核苷酸的FIX编码序列部分包括编码成熟因子IX多肽的核酸序列,所述核酸序列与CS05-MP-NA (SEQ ID NO:16) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX信号肽的核酸序列,所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-SP-NA (SEQ ID NO:24)、CS02-SP-NA (SEQ ID NO:25)、CS03-SP-NA (SEQ ID NO:26)、CS04-SP-NA (SEQ ID NO:27)、CS05-SP-NA (SEQ ID NO:28) 和CS06-SP-NA (SEQ ID NO:29)。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX原肽的核酸序列(任选地与如上文所描述的因子IX信号肽的核酸序列组合),所述核酸序列与

以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-PP-NA (SEQ ID NO:30)、CS02-PP-NA (SEQ ID NO:31)、CS03-PP-NA (SEQ ID NO:32)、CS04-PP-NA (SEQ ID NO:33)、CS05-PP-NA (SEQ ID NO:34) 和CS06-PP-NA (SEQ ID NO:35)。在一些实施方案中,多核苷酸的FIX编码序列部分包括编码前因子IX多肽的核酸序列,所述核酸序列与CS05-FL-NA (SEQ ID NO:8) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。

[0283] CS06密码子改变的多核苷酸

[0284] 在一个实施方案中,本文所提供的核酸组合物包括编码单链因子IX多肽的因子IX多核苷酸(例如密码子改变的多核苷酸),其中因子IX多核苷酸包括与CS06-FL-NA (SEQ ID NO:9) 具有高度序列同一性的核苷酸序列。在一些实施方案中,与CS06-FL-NA (SEQ ID NO:9) 具有高度序列同一性的因子IX多核苷酸的核苷酸序列与野生型因子IX编码序列 (FIX-FL-NA (SEQ ID NO:1)) 相比具有降低的GC含量。在一些实施方案中,与CS06-FL-NA (SEQ ID NO:9) 具有高度序列同一性的因子IX多核苷酸的核苷酸序列与野生型因子IX编码序列 (FIX-FL-NA (SEQ ID NO:1)) 相比具有减少的CpG二核苷酸数目。

[0285] 在一个具体实施方案中,密码子改变的多核苷酸的序列与CS06-FL-NA (SEQ ID NO:9) 具有至少95%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS06-FL-NA (SEQ ID NO:9) 具有至少96%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS06-FL-NA (SEQ ID NO:9) 具有至少97%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS06-FL-NA (SEQ ID NO:9) 具有至少98%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS06-FL-NA (SEQ ID NO:9) 具有至少99%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS06-FL-NA (SEQ ID NO:9) 具有至少99.5%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS06-FL-NA (SEQ ID NO:9) 具有至少99.9%的同一性。在另一个具体实施方案中,密码子改变的多核苷酸的序列为CS06-FL-NA (SEQ ID NO:9)。

[0286] 在一些实施方案中,与CS06-FL-NA (SEQ ID NO:9) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于60%的GC含量。在一些实施方案中,与CS06-FL-NA (SEQ ID NO:9) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于59%的GC含量。在一些实施方案中,与CS06-FL-NA (SEQ ID NO:9) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于58%的GC含量。在一些实施方案中,与CS06-FL-NA (SEQ ID NO:9) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于57%的GC含量。在一些实施方案中,与CS06-FL-NA (SEQ ID NO:9) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于56%的GC含量。在一些实施方案中,与CS06-FL-NA (SEQ ID NO:9) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于55%的GC含量。在一些实施方案中,与CS06-FL-NA (SEQ ID NO:9) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于54%的GC含量。

[0287] 在一些实施方案中,与CS06-FL-NA (SEQ ID NO:9) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至60%的GC含量。在一些实施方案中,与CS06-FL-NA (SEQ ID NO:9) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至59%的GC含量。在一些实施方案中,与CS06-FL-NA (SEQ ID NO:9) 具有高度序列同一性的密码子改变的多

施方案中,与CS06-FL-NA(SEQ ID NO:9)具有高度序列同一性的密码子改变的多核苷酸的序列不具有CpG二核苷酸。

[0290] 在一些实施方案中,所编码的因子IX多肽(例如由与CS06-FL-NA(SEQ ID NO:9)具有高度序列同源性的多核苷酸编码的多肽)与野生型因子IX前原蛋白序列FIX-FL-AA(SEQ ID NO:2)和/或Padua(hFIX(R384L))前原蛋白序列FIXp-FL-AA(SEQ ID NO:4)具有高度序列同一性。所编码的因子IX多肽应保留被活化为功能因子IXa蛋白(例如通过移除信号肽和原肽,和通过切除活化多肽)的能力。

[0291] 在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA(SEQ ID NO:2)具有至少85%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA(SEQ ID NO:2)具有至少90%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA(SEQ ID NO:2)具有至少95%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA(SEQ ID NO:2)具有至少96%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA(SEQ ID NO:2)具有至少97%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA(SEQ ID NO:2)具有至少98%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA(SEQ ID NO:2)具有至少99%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA(SEQ ID NO:2)具有至少99.5%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-FL-AA(SEQ ID NO:2)具有至少99.9%的同一性。在一个实施方案中,所编码的因子IX多肽的序列为FIX-FL-AA(SEQ ID NO:2)。

[0292] 在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA(SEQ ID NO:4)具有至少85%的同一性并且在前原多肽的位置384(例如成熟因子IX单链多肽FIXp-MP-AA(SEQ ID NO:12)的位置338)处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA(SEQ ID NO:4)具有至少90%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA(SEQ ID NO:4)具有至少95%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA(SEQ ID NO:4)具有至少96%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA(SEQ ID NO:4)具有至少97%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA(SEQ ID NO:4)具有至少98%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA(SEQ ID NO:4)具有至少99%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA(SEQ ID NO:4)具有至少99.5%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-FL-AA(SEQ ID NO:4)具有至少99.9%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列为FIXp-FL-AA(SEQ ID NO:4)。

[0293] 在一个实施方案中,本文所提供的核酸组合物包括编码单链因子IX多肽(例如具有丝氨酸蛋白酶活性)的因子IX多核苷酸(例如密码子改变的多核苷酸),其中因子IX多核苷酸包括与CS06-MP-NA(SEQ ID NO:17)具有高度序列同一性的核苷酸序列。在一些实施方

案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的因子IX多核苷酸的核苷酸序列与野生型因子IX编码序列 (FIX-FL-NA (SEQ ID NO:1)) 相比具有降低的GC含量。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的因子IX多核苷酸的核苷酸序列与野生型因子IX编码序列 (FIX-FL-NA (SEQ ID NO:1)) 相比具有减少的CpG二核苷酸数目。

[0294] 在一个具体实施方案中,密码子改变的多核苷酸的序列与CS06-MP-NA (SEQ ID NO:17) 具有至少95%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS06-MP-NA (SEQ ID NO:17) 具有至少96%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS06-MP-NA (SEQ ID NO:17) 具有至少97%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS06-MP-NA (SEQ ID NO:17) 具有至少98%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS06-MP-NA (SEQ ID NO:17) 具有至少99%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS06-MP-NA (SEQ ID NO:17) 具有至少99.5%的同一性。在一个具体实施方案中,密码子改变的多核苷酸的序列与CS06-MP-NA (SEQ ID NO:17) 具有至少99.9%的同一性。在另一个具体实施方案中,密码子改变的多核苷酸的序列为CS06-MP-NA (SEQ ID NO:17)。

[0295] 在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于60%的GC含量。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于59%的GC含量。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于58%的GC含量。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于57%的GC含量。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于56%的GC含量。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于55%的GC含量。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有低于54%的GC含量。

[0296] 在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至60%的GC含量。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至59%的GC含量。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至58%的GC含量。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至57%的GC含量。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至56%的GC含量。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至55%的GC含量。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有50%至54%的GC含量。

[0297] 在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 1.0$ 的GC含量。在一些实施方案中,与CS06-MP-NA (SEQ

ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.8$ 的GC含量。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.6$ 的GC含量。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.5$ 的GC含量。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.4$ 的GC含量。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.3$ 的GC含量。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.2$ 的GC含量。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有 $53.8\% \pm 0.1$ 的GC含量。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有53.8%的GC含量。

[0298] 在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过15个CpG二核苷酸。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过12个CpG二核苷酸。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过10个CpG二核苷酸。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过9个CpG二核苷酸。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过8个CpG二核苷酸。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过7个CpG二核苷酸。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过6个CpG二核苷酸。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过5个CpG二核苷酸。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过4个CpG二核苷酸。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过3个CpG二核苷酸。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过2个CpG二核苷酸。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列具有不超过1个CpG二核苷酸。在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的密码子改变的多核苷酸的序列不具有CpG二核苷酸。

[0299] 在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的因子IX多核苷酸进一步包括编码具有FIX-SP-AA (SEQ ID NO:37) 的氨基酸序列的因子IX信号肽的因子IX信号多核苷酸。在一些实施方案中,因子IX信号多核苷酸与CS02-SP-NA (SEQ ID NO:25) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX信号多核苷酸与CS03-SP-NA (SEQ ID NO:26) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX信号多核苷酸与CS04-SP-NA (SEQ ID NO:27) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序

列。在一些实施方案中,因子IX信号多核苷酸与CS05-SP-NA (SEQ ID NO:28) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX信号多核苷酸与CS06-SP-NA (SEQ ID NO:29) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。

[0300] 在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的因子IX多核苷酸进一步包括编码具有FIX-PP-AA (SEQ ID NO:38) 的氨基酸序列的因子IX原肽的因子IX原肽多核苷酸。在一些实施方案中,因子IX原肽多核苷酸与CS02-PP-NA (SEQ ID NO:31) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX原肽多核苷酸与CS03-PP-NA (SEQ ID NO:32) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX原肽多核苷酸与CS04-PP-NA (SEQ ID NO:33) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX原肽多核苷酸与CS05-PP-NA (SEQ ID NO:34) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX原肽多核苷酸与CS06-PP-NA (SEQ ID NO:35) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。

[0301] 在一些实施方案中,与CS06-MP-NA (SEQ ID NO:17) 具有高度序列同一性的因子IX多核苷酸进一步包括编码具有FIX-PPP-AA (SEQ ID NO:36) 的氨基酸序列的因子IX前原肽的因子IX前原肽多核苷酸。在一些实施方案中,因子IX前原肽多核苷酸与CS02-PPP-NA (SEQ ID NO:19) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX前原肽多核苷酸与CS03-PPP-NA (SEQ ID NO:20) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX前原肽多核苷酸与CS04-PPP-NA (SEQ ID NO:21) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX前原肽多核苷酸与CS05-PPP-NA (SEQ ID NO:22) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。在一些实施方案中,因子IX前原肽多核苷酸与CS06-PPP-NA (SEQ ID NO:23) 具有至少90%、95%、96%、97%、98%、99%或100%同一性的核酸序列。

[0302] 在一些实施方案中,所编码的因子IX多肽(例如由与CS06-FL-NA (SEQ ID NO:9) 具有高度序列同源性的多核苷酸编码的多肽)与野生型成熟因子IX单链多肽序列FIX-MP-AA (SEQ ID NO:10) 和/或成熟Padua (hFIX (R384L)) 单链序列FIXp-MP-AA (SEQ ID NO:12) 具有高度序列同一性。所编码的因子IX多肽应保留被活化为功能因子IXa蛋白(例如通过移除任何信号肽和原肽,和通过切除活化多肽)的能力。

[0303] 在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少85%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少90%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少95%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少96%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少97%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少98%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少99%的同

一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少99.5%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10) 具有至少99.9%的同一性。在一个实施方案中,所编码的因子IX多肽的序列为FIX-MP-AA (SEQ ID NO:10)。

[0304] 在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少85%的同一性并且在前原多肽的位置384 (例如成熟因子IX单链多肽FIXp-MP-AA (SEQ ID NO:12) 的位置338) 处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少90%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少95%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少96%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少97%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少98%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少99%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少99.5%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少99.9%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列为FIXp-MP-AA (SEQ ID NO:12)。

[0305] 在一个实施方案中,本文所提供的密码子改变的多核苷酸编码单链因子IX多肽,所述单链因子IX多肽包括轻链、重链和将轻链C端连接至重链N端的多肽接头。因子IX多肽的轻链由第一核苷酸序列编码,所述第一核苷酸序列与作为CS06-FL-NA (SEQ ID NO:9) 的编码因子IX轻链的部分的CS06-LC-NA (SEQ ID NO:50) 具有高度序列同一性。因子IX多肽的重链由第二核苷酸序列编码,所述第二核苷酸序列与作为CS06-FL-NA (SEQ ID NO:9) 的编码因子IX重链的部分的CS06-HC-NA (SEQ ID NO:49) 具有高度序列同一性。多肽接头包括因子XI切割位点,其允许体内成熟 (例如在前体单链因子IX多肽表达之后)。

[0306] 在一些实施方案中,第一和第二核苷酸序列分别与CS06-LC-NA和CS06-HC-NA (SEQ ID NO:50和49) 具有至少95%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS06-LC-NA和CS06-HC-NA (SEQ ID NO:50和49) 具有至少96%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS06-LC-NA和CS06-HC-NA (SEQ ID NO:50和49) 具有至少97%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS06-LC-NA和CS06-HC-NA (SEQ ID NO:50和49) 具有至少98%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS06-LC-NA和CS06-HC-NA (SEQ ID NO:50和49) 具有至少99%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS06-LC-NA和CS06-HC-NA (SEQ ID NO:50和49) 具有至少99.5%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别与CS06-LC-NA和CS06-HC-NA (SEQ ID NO:50和49) 具有至少99.9%的序列同一性。在一些实施方案中,第一和第二核苷酸序列分别为CS06-LC-NA和CS06-HC-NA (SEQ ID

NO:50和49)。

[0307] 在一些实施方案中,因子IX构建体的多肽接头由第三核苷酸序列编码,所述第三核苷酸序列与作为编码野生型因子IX活化多肽(例如FIX-FL-AA (SEQ ID NO:2)的氨基酸192-226)的密码子改变的序列的CS06-AP-NA (SEQ ID NO:61)具有高度序列同一性。在一些实施方案中,第三核苷酸序列与CS06-AP-NA (SEQ ID NO:61)具有至少80%的同一性。在一些实施方案中,第三核苷酸序列与CS06-AP-NA (SEQ ID NO:61)具有至少90%的同一性。在一些实施方案中,第三核苷酸序列与CS06-AP-NA (SEQ ID NO:61)具有至少95%的同一性。在一些实施方案中,第三核苷酸序列与CS06-AP-NA (SEQ ID NO:61)具有至少96%的同一性。在一些实施方案中,第三核苷酸序列与CS06-AP-NA (SEQ ID NO:61)具有至少97%的同一性。在一些实施方案中,第三核苷酸序列与CS06-AP-NA (SEQ ID NO:61)具有至少98%的同一性。在一些实施方案中,第三核苷酸序列与CS06-AP-NA (SEQ ID NO:61)具有至少99%的同一性。在一些实施方案中,第三核苷酸序列为CS06-AP-NA (SEQ ID NO:61)。

[0308] 在一些实施方案中,所编码的因子IX多肽还包括信号肽(例如因子IX信号肽)和/或原肽(例如因子IX原肽)。在一些实施方案中,信号肽为野生型因子IX信号肽(FIX-SP-AA (SEQ ID NO:37))。在一些实施方案中,信号肽由与CS06-SP-NA (SEQ ID NO:29)具有高度序列同一性(例如至少95%、96%、97%、98%或99%)的密码子改变的多核苷酸序列编码。在一些实施方案中,原肽为野生型因子IX原肽(FIX-PP-AA (SEQ ID NO:38))。在一些实施方案中,原肽由与CS06-PP-NA (SEQ ID NO:35)具有高度序列同一性(例如至少95%、96%、97%、98%或99%)的密码子改变的多核苷酸序列编码。

[0309] 在一些实施方案中,所编码的因子IX多肽(例如由与CS06-LC-NA (SEQ ID NO:50)和CS06-HC-NA (SEQ ID NO:49)具有高度序列同源性的多核苷酸编码的多肽)与野生型成熟因子IX单链多肽序列FIX-MP-AA (SEQ ID NO:10)和/或成熟Padua (hFIX (R384L))单链序列FIXp-MP-AA (SEQ ID NO:12)具有高度序列同一性。所编码的因子IX多肽应保留被活化为功能因子IXa蛋白(例如通过移除任何信号肽和原肽,和通过切除活化多肽)的能力。

[0310] 在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10)具有至少85%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10)具有至少90%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10)具有至少95%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10)具有至少96%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10)具有至少97%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10)具有至少98%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10)具有至少99%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10)具有至少99.5%的同一性。在一个实施方案中,所编码的因子IX多肽的序列与FIX-MP-AA (SEQ ID NO:10)具有至少99.9%的同一性。在一个实施方案中,所编码的因子IX多肽的序列为FIX-MP-AA (SEQ ID NO:10)。

[0311] 在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12)具有至少85%的同一性并且在前原多肽的位置384(例如成熟因子IX单链多肽FIXp-MP-AA (SEQ ID NO:12)的位置338)处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序

列与FIXp-MP-AA (SEQ ID NO:12) 具有至少90%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少95%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少96%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少97%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少98%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少99%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少99.5%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列与FIXp-MP-AA (SEQ ID NO:12) 具有至少99.9%的同一性并且在前原多肽的位置384处包括亮氨酸。在一个实施方案中,所编码的因子IX多肽的序列为FIXp-MP-AA (SEQ ID NO:12)。

[0312] 在一些实施方案中,参考图1,提供包括结构A的自身互补型多核苷酸的核酸组合物,其中多核苷酸的FIX编码序列部分包括编码成熟因子IX多肽的核酸序列,其与CS06-MP-NA (SEQ ID NO:17) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX信号肽的核酸序列,所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-SP-NA (SEQ ID NO:24)、CS02-SP-NA (SEQ ID NO:25)、CS03-SP-NA (SEQ ID NO:26)、CS04-SP-NA (SEQ ID NO:27)、CS05-SP-NA (SEQ ID NO:28) 和CS06-SP-NA (SEQ ID NO:29)。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX原肽的核酸序列(任选地与如上文所描述的因子IX信号肽的核酸序列组合),所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-PP-NA (SEQ ID NO:30)、CS02-PP-NA (SEQ ID NO:31)、CS03-PP-NA (SEQ ID NO:32)、CS04-PP-NA (SEQ ID NO:33)、CS05-PP-NA (SEQ ID NO:34) 和CS06-PP-NA (SEQ ID NO:35)。在一些实施方案中,多核苷酸的FIX编码序列部分包括编码前原因子IX多肽的核酸序列,所述核酸序列与CS06-FL-NA (SEQ ID NO:9) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。

[0313] 在一些实施方案中,参考图1,提供包括结构B的自身互补型多核苷酸的核酸组合物,其中多核苷酸的FIX编码序列部分包括编码成熟因子IX多肽的核酸序列,所述核酸序列与CS06-MP-NA (SEQ ID NO:17) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX信号肽的核酸序列,所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-SP-NA (SEQ ID NO:24)、CS02-SP-NA (SEQ ID NO:25)、CS03-SP-NA (SEQ ID NO:26)、CS04-SP-NA (SEQ ID NO:27)、CS05-SP-NA (SEQ ID NO:28) 和CS06-SP-NA (SEQ ID NO:29)。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX原肽的核酸序列(任选地与如上文所描述的因子IX信号肽的核酸序列组合),所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-PP-NA (SEQ ID NO:30)、CS02-PP-NA (SEQ ID NO:31)、CS03-PP-NA (SEQ ID NO:32)、CS04-PP-NA (SEQ ID NO:

33)、CS05-PP-NA (SEQ ID NO:34) 和CS06-PP-NA (SEQ ID NO:35)。在一些实施方案中,多核苷酸的FIX编码序列部分包括编码前原因子IX多肽的核酸序列,所述核酸序列与CS06-FL-NA (SEQ ID NO:9) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。

[0314] 在一些实施方案中,参考图1,提供包括结构C的多核苷酸(例如单链多核苷酸)的核酸组合物,其中多核苷酸的FIX编码序列部分包括编码成熟因子IX多肽的核酸序列,所述核酸序列与CS06-MP-NA (SEQ ID NO:17) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX信号肽的核酸序列,所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-SP-NA (SEQ ID NO:24)、CS02-SP-NA (SEQ ID NO:25)、CS03-SP-NA (SEQ ID NO:26)、CS04-SP-NA (SEQ ID NO:27)、CS05-SP-NA (SEQ ID NO:28) 和CS06-SP-NA (SEQ ID NO:29)。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX原肽的核酸序列(任选地与如上文所描述的因子IX信号肽的核酸序列组合),所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-PP-NA (SEQ ID NO:30)、CS02-PP-NA (SEQ ID NO:31)、CS03-PP-NA (SEQ ID NO:32)、CS04-PP-NA (SEQ ID NO:33)、CS05-PP-NA (SEQ ID NO:34) 和CS06-PP-NA (SEQ ID NO:35)。在一些实施方案中,多核苷酸的FIX编码序列部分包括编码前原因子IX多肽的核酸序列,所述核酸序列与CS06-FL-NA (SEQ ID NO:9) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。

[0315] 在一些实施方案中,参考图1,提供包括结构D的多核苷酸(例如单链多核苷酸)的核酸组合物,其中多核苷酸的FIX编码序列部分包括编码成熟因子IX多肽的核酸序列,所述核酸序列与CS06-MP-NA (SEQ ID NO:17) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX信号肽的核酸序列,所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-SP-NA (SEQ ID NO:24)、CS02-SP-NA (SEQ ID NO:25)、CS03-SP-NA (SEQ ID NO:26)、CS04-SP-NA (SEQ ID NO:27)、CS05-SP-NA (SEQ ID NO:28) 和CS06-SP-NA (SEQ ID NO:29)。在一些实施方案中,多核苷酸的FIX编码序列部分还包括编码因子IX原肽的核酸序列(任选地与如上文所描述的因子IX信号肽的核酸序列组合),所述核酸序列与以下一者具有至少90%、95%、96%、97%、98%、99%或100%的同一性:FIX-PP-NA (SEQ ID NO:30)、CS02-PP-NA (SEQ ID NO:31)、CS03-PP-NA (SEQ ID NO:32)、CS04-PP-NA (SEQ ID NO:33)、CS05-PP-NA (SEQ ID NO:34) 和CS06-PP-NA (SEQ ID NO:35)。在一些实施方案中,多核苷酸的FIX编码序列部分包括编码前原因子IX多肽的核酸序列,所述核酸序列与CS06-FL-NA (SEQ ID NO:9) 具有至少95%、96%、97%、98%、99%、99.5%、99.9%或100%的同一性。

[0316] C. 密码子改变的因子IX信号肽和原肽

[0317] 在一个方面中,本公开提供编码因子IX信号肽、因子IX原肽和两者(例如因子IX前原肽)的密码子改变的多核苷酸。这些密码子改变的多核苷酸改善因子IX表达并且可以被置于例如密码子改变的或以其他方式编码因子IX单链多肽(例如因子IX轻链、活化肽和重链)的多核苷酸的上游。通常,所编码的肽为野生型因子IX信号肽(例如FIX-SP-AA (SEQ ID NO:37))、原肽(例如FIX-PP-AA (SEQ ID NO:38) 和前原肽 (FIX-PPP-AA (SEQ ID NO:36)))。

[0318] 在某些实施方案中,编码因子IX信号肽、原肽和前原肽的密码子改变的多核苷酸与以下一者具有高度同一性(例如至少90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)的序列:CS02-SP-NA(SEQ ID NO:25)、CS03-SP-NA(SEQ ID NO:26)、CS04-SP-NA(SEQ ID NO:27)、CS05-SP-NA(SEQ ID NO:28)、CS06-SP-NA(SEQ ID NO:29)、CS02-PP-NA(SEQ ID NO:31)、CS03-PP-NA(SEQ ID NO:32)、CS04-PP-NA(SEQ ID NO:33)、CS05-PP-NA(SEQ ID NO:34)、CS06-PP-NA(SEQ ID NO:35)、CS02-PPP-NA(SEQ ID NO:19)、CS03-PPP-NA(SEQ ID NO:20)、CS04-PPP-NA(SEQ ID NO:21)、CS05-PPP-NA(SEQ ID NO:22)和CS06-PPP-NA(SEQ ID NO:23)。

[0319] CS02信号肽和原肽

[0320] 在一个实施方案中,编码因子IX信号肽的密码子改变的多核苷酸与CS02-SP-NA(SEQ ID NO:25)具有至少95%的序列同一性。在其他实施方案中,编码因子IX信号肽的密码子改变的多核苷酸与CS02-SP-NA(SEQ ID NO:25)具有至少96%、97%、98%、99%或100%的同一性。

[0321] 在一个实施方案中,编码因子IX原肽的密码子改变的多核苷酸与CS02-PP-NA(SEQ ID NO:31)具有至少95%的序列同一性。在其他实施方案中,编码因子IX原肽的密码子改变的多核苷酸与CS02-PP-NA(SEQ ID NO:31)具有至少96%、97%、98%、99%或100%的同一性。

[0322] 在一个实施方案中,编码因子IX前原肽的密码子改变的多核苷酸与CS02-PPP-NA(SEQ ID NO:19)具有至少95%的序列同一性。在其他实施方案中,编码因子IX前原肽的密码子改变的多核苷酸与CS02-PPP-NA(SEQ ID NO:19)具有至少96%、97%、98%、99%或100%的同一性。

[0323] CS03信号肽和原肽

[0324] 在一个实施方案中,编码因子IX信号肽的密码子改变的多核苷酸与CS03-SP-NA(SEQ ID NO:26)具有至少95%的序列同一性。在其他实施方案中,编码因子IX信号肽的密码子改变的多核苷酸与CS03-SP-NA(SEQ ID NO:26)具有至少96%、97%、98%、99%或100%的同一性。

[0325] 在一个实施方案中,编码因子IX原肽的密码子改变的多核苷酸与CS03-PP-NA(SEQ ID NO:32)具有至少95%的序列同一性。在其他实施方案中,编码因子IX原肽的密码子改变的多核苷酸与CS03-PP-NA(SEQ ID NO:32)具有至少96%、97%、98%、99%或100%的同一性。

[0326] 在一个实施方案中,编码因子IX前原肽的密码子改变的多核苷酸与CS03-PPP-NA(SEQ ID NO:20)具有至少95%的序列同一性。在其他实施方案中,编码因子IX前原肽的密码子改变的多核苷酸与CS03-PPP-NA(SEQ ID NO:20)具有至少96%、97%、98%、99%或100%的同一性。

[0327] CS04信号肽和原肽

[0328] 在一个实施方案中,编码因子IX信号肽的密码子改变的多核苷酸与CS04-SP-NA(SEQ ID NO:27)具有至少95%的序列同一性。在其他实施方案中,编码因子IX信号肽的密码子改变的多核苷酸与CS04-SP-NA(SEQ ID NO:27)具有至少96%、97%、98%、99%或100%的同一性。

[0329] 在一个实施方案中,编码因子IX原肽的密码子改变的多核苷酸与CS04-PP-NA (SEQ ID NO:33) 具有至少95%的序列同一性。在其他实施方案中,编码因子IX原肽的密码子改变的多核苷酸与CS04-PP-NA (SEQ ID NO:33) 具有至少96%、97%、98%、99%或100%的同一性。

[0330] 在一个实施方案中,编码因子IX前原肽的密码子改变的多核苷酸与CS04-PPP-NA (SEQ ID NO:21) 具有至少95%的序列同一性。在其他实施方案中,编码因子IX前原肽的密码子改变的多核苷酸与CS04-PPP-NA (SEQ ID NO:21) 具有至少96%、97%、98%、99%或100%的同一性。

[0331] CS05信号肽和原肽

[0332] 在一个实施方案中,编码因子IX信号肽的密码子改变的多核苷酸与CS05-SP-NA (SEQ ID NO:28) 具有至少95%的序列同一性。在其他实施方案中,编码因子IX信号肽的密码子改变的多核苷酸与CS05-SP-NA (SEQ ID NO:28) 具有至少96%、97%、98%、99%或100%的同一性。

[0333] 在一个实施方案中,编码因子IX原肽的密码子改变的多核苷酸与CS05-PP-NA (SEQ ID NO:34) 具有至少95%的序列同一性。在其他实施方案中,编码因子IX原肽的密码子改变的多核苷酸与CS05-PP-NA (SEQ ID NO:34) 具有至少96%、97%、98%、99%或100%的同一性。

[0334] 在一个实施方案中,编码因子IX前原肽的密码子改变的多核苷酸与CS05-PPP-NA (SEQ ID NO:22) 具有至少95%的序列同一性。在其他实施方案中,编码因子IX前原肽的密码子改变的多核苷酸与CS05-PPP-NA (SEQ ID NO:22) 具有至少96%、97%、98%、99%或100%的同一性。

[0335] CS06信号肽和原肽

[0336] 在一个实施方案中,编码因子IX信号肽的密码子改变的多核苷酸与CS06-SP-NA (SEQ ID NO:29) 具有至少95%的序列同一性。在其他实施方案中,编码因子IX信号肽的密码子改变的多核苷酸与CS06-SP-NA (SEQ ID NO:29) 具有至少96%、97%、98%、99%或100%的同一性。

[0337] 在一个实施方案中,编码因子IX原肽的密码子改变的多核苷酸与CS06-PP-NA (SEQ ID NO:35) 具有至少95%的序列同一性。在其他实施方案中,编码因子IX原肽的密码子改变的多核苷酸与CS06-PP-NA (SEQ ID NO:35) 具有至少96%、97%、98%、99%或100%的同一性。

[0338] 在一个实施方案中,编码因子IX前原肽的密码子改变的多核苷酸与CS06-PPP-NA (SEQ ID NO:23) 具有至少95%的序列同一性。在其他实施方案中,编码因子IX前原肽的密码子改变的多核苷酸与CS06-PPP-NA (SEQ ID NO:23) 具有至少96%、97%、98%、99%或100%的同一性。

[0339] IV. 因子IX表达载体

[0340] 在一些实施方案中,本文所描述的密码子改变的多核苷酸被整合至表达载体中。如本领域技术人员将了解到,许多形式的载体可以用于使用本文所公开的密码子改变的因子IX多核苷酸序列实现因子IX基因疗法。表达载体的非限制性实例包括病毒载体(例如适用于基因疗法的载体)、质粒载体、噬菌体载体、粘粒、噬菌粒、人工染色体等等。

[0341] 在一些实施方案中,本文所描述的密码子改变的多核苷酸被整合至病毒基因疗法载体中。病毒载体的非限制性实例包括:逆转录病毒,例如莫罗尼鼠白血病病毒(Moloney murine leukemia virus, MMLV)、哈维鼠肉瘤病毒(Harvey murine sarcoma virus)、鼠乳腺肿瘤病毒(murine mammary tumor virus)和劳斯氏肉瘤病毒(Rous sarcoma virus);腺病毒、腺相关病毒;SV40型病毒;多瘤病毒;爱泼斯坦-巴尔病毒(Epstein-Barr virus);乳头瘤病毒;疱疹病毒;牛痘病毒;和脊髓灰质炎病毒。

[0342] 在体内,因子IX主要在肝脏中合成。因而,已靶向肝细胞作为适合于因子IX基因疗法构建体的宿主细胞。几类病毒载体已显示有能力进行基因疗法构建体的肝靶向递送,所述基因疗法构建体包括逆转录病毒载体(参见例如Axelrod等人,1990;Kay等人,1992;Van den Driessche等人,1999;和Xu等人,2003,2005,其公开内容出于所有目的以全文引用的方式明确地并入本文中)、慢病毒(参见例如Ward等人,2011;Brown等人,2007;和Matrai等人,2011,其公开内容出于所有目的以全文引用的方式明确地并入本文中)、腺相关病毒(AAV)(参见例如Herzog等人,1999,其公开内容出于所有目的以全文引用的方式明确地并入本文中)和腺病毒载体(参见例如Brown等人,2004;和Ehrhardt和Kay,2002,其公开内容出于所有目的以全文引用的方式明确地并入本文中)。

[0343] 在一些实施方案中,基因疗法载体为逆转录病毒,并且特别是复制缺陷型逆转录病毒。用于产生复制缺陷型逆转录病毒的方案为本领域中已知的。关于综述,参见Kriegler, M., Gene Transfer and Expression, 实验室手册(A Laboratory Manual), W.H. Freeman Co., New York (1990); 和Murry, E.J., Methods in Molecular Biology, 第7卷, Humana Press, Inc., Clifton, N.J. (1991)。

[0344] 在一个实施方案中,基因疗法载体为基于腺相关病毒(AAV)的基因疗法载体。AAV系统先前已有描述并且为本领域中普遍熟知的(Kelleher和Vos, Biotechniques, 17(6): 1110-17 (1994); Cotten等人, P.N.A.S.U.S.A., 89(13): 6094-98 (1992); Curiel, Nat Immun, 13(2-3): 141-64 (1994); Muzyczka, Curr Top Microbiol Immunol, 158: 97-129 (1992); 和Asokan A等人, Mol. Ther., 20(4): 699-708 (2012), 各自出于所有目的以全文引用的方式并入本文中)。关于rAAV载体的生成和使用的细节描述于例如美国专利No. 5,139,941和4,797,368中,这些专利各自出于所有目的以全文引用的方式并入本文中。在一个具体实施方案中,AAV载体为AAV-8载体。

[0345] 用于肝特异性因子IX表达的示例性AAV递送载体描述于W02009/130208中,所述专利的内容出于所有目的以全文引用的方式明确并入本文中。所述载体为编码人因子IX的单链AAV载体,并且包括驱动因子cDNA的TTR Serp调控序列。所述载体还包括人因子IX基因的内含子I和多聚腺苷酸化信号。

[0346] 在一些实施方案中,本文所描述的密码子改变的多核苷酸被整合至逆转录病毒表达载体中。这些系统先前已有描述,并且为本领域中普遍熟知的(Mann等人, Cell, 33: 153-159, 1983; Nicolas和Rubinstein, Vectors: A survey of molecular cloning vectors and their uses, Rodriguez和Denhardt编, Stoneham: Butterworth, 第494-513页, 1988; Temin, Gene Transfer, Kucherlapati (编), New York: Plenum Press, 第149-188页, 1986)。在一个具体实施方案中,逆转录病毒载体为慢病毒载体(参见例如Naldini等人, Science, 272(5259): 263-267, 1996; Zufferey等人, Nat Biotechnol, 15(9): 871-875, 1997; Blomer

等人, *J Virol.*, 71(9):6641-6649, 1997; 美国专利No. 6,013,516和5,994,136)。

[0347] 在一些实施方案中,本文所描述的密码子改变的多核苷酸可以通过非病毒方法向受试者施用。举例来说,裸DNA可以通过电穿孔、声孔效应、粒子轰击或水力递送施用至细胞中。也可以将DNA用聚合物(例如脂质体、多核糖体、多聚复合物(polypleses)、树状大分子)封装或使DNA与之偶合,并且作为复合物施用给受试者。同样地,DNA可以与无机纳米粒子,例如金、二氧化硅、氧化铁或磷酸钙粒子偶合,或附接至细胞穿透肽以在体内递送至细胞。

[0348] 密码子改变的因子IX编码多核苷酸也可以并入人工染色体中,诸如人工染色体表达(ACE)(参见例如Lindenbaum等人, *Nucleic Acids Res.*, 32(21):e172(2004))和哺乳动物人工染色体(MAC)。关于综述,参见例如Pérez-Luz和Díaz-Nido, *J Biomed Biotechnol.* 2010:Article ID 642804(2010)。

[0349] 多种载体可以用于在细胞培养物中由密码子改变的多肽表达因子IX多肽,包括真核和原核表达载体。在某些实施方案中,涵盖用于在细胞培养物中表达因子IX多肽的质粒载体。一般来说,结合这些宿主使用含有源自与宿主细胞相容的物种的复制子和控制序列的质粒载体。所述载体可以携带复制位点,以及能够在转化细胞中提供表型选择的标志序列。质粒将包括编码因子IX多肽的密码子改变的多核苷酸,其可操作地连接至一个或多个控制序列,例如启动子。

[0350] 用于原核表达的载体的非限制性实例包括质粒,诸如pRSET、pET、pBAD等,其中用于原核表达载体中的启动子包括lac、trc、trp、recA、araBAD等。用于真核表达的载体的实例包括:(i)对于在酵母中表达,诸如pA0、pPIC、pYES、pMET的载体,使用诸如AOX1、GAP、GAL1、AUG1等的启动子;(ii)对于在昆虫细胞中表达,诸如pMT、pAc5、pIB、pMIB、pBAC等的载体,使用诸如PH、p10、MT、Ac5、OpIE2、gp64、polh等的启动子;以及(iii)对于在哺乳动物细胞中表达,诸如pSVL、pCMV、pRc/RSV、pcDNA3、pBPV等的载体,以及源自诸如牛痘病毒、腺相关病毒、疱疹病毒、逆转录病毒等病毒系统的载体,使用诸如CMV、SV40、EF-1、UbC、RSV、ADV、BPV和 β -肌动蛋白的启动子。

[0351] 在一些实施方案中,本公开提供AAV基因疗法载体,所述基因疗法载体包括如本文所描述的密码子改变的因子IX多核苷酸、载体5'端和3'端的内部末端重复(ITR)序列、可操作地连接至因子IX多核苷酸的一个或多个启动子和/或增强子序列以及在因子IX多核苷酸序列3'端之后的多聚腺苷酸化信号。在一些实施方案中,一个或多个启动子和/或增强子序列包括肝特异性调控控制元件的一个或多个拷贝。

[0352] 图1说明根据一些实现方式的因子IX基因疗法载体的几个示例性架构。图1A说明具有突变的5' ITR、截短的TTR增强子/启动子序列、MVM病毒内含子序列、密码子改变的因子IX编码序列、多聚腺苷酸化序列和3'-ITR的自身互补型AAV载体。图1B说明类似于图1A,但进一步包括一个或多个(例如一个、两个、三个或更多个)肝特异性调控控制元件的编码因子IX多肽的自身互补型AAV载体。图1C说明具有与图1A相同的元件,但存在以下不同之处的单链载体:5'-ITR未突变,从而防止自身互补性。图1D说明类似于图1A,但进一步包括一个或多个(例如一个、两个、三个或更多个)肝特异性调控控制元件的编码因子IX多肽的单链AAV载体。虽然参考图1中包括R384L 'Padua' 氨基酸取代的因子IX蛋白进行了说明,但在一些实施方案中,具有如图1中所示的一般结构(例如结构A、B、C或D)的因子IX核苷酸构建体编码并不包括R384L 'Padua' 氨基酸取代的因子IX蛋白。

[0353] 图25显示AAV因子IX基因疗法载体CS06-CRM8.3-ssV(SEQ ID NO:40)的核苷酸序列,其例示了图1D中所说明的基因疗法载体架构。CS06-CRM8.3-ssV(SEQ ID NO:40)的核苷酸1-145为AAV2 5'-ITR序列(SEQ ID NO:51)。在5'-ITR序列之后为位于核苷酸165-236、238-309和311-382处的肝特异性CRM8调控控制元件CRM8(SEQ ID NO:39)的三个拷贝。在CRM8序列之后为位于核苷酸383-712处的截短的TTR增强子/启动子序列(SEQ ID NO:52)。接着,所述载体包括位于核苷酸724-800处的小鼠微小病毒(MVM)内含子(SEQ ID NO:53)。所述载体的核苷酸814-2199为CS06密码子改变的因子IX(R384L)编码序列(CS06-FL-NA(SEQ ID NO:9))。在因子IX多核苷酸序列之后为位于核苷酸2208-2441处的BGH多聚腺苷酸化信号,并且最后为位于核苷酸2458-2602处的AAV2 3'-ITR序列(SEQ ID NO:55)。

[0354] 在一些实施方案中,本公开提供包含与SEQ ID NO:40的核苷酸1-2602具有至少95%同一性的序列的因子IX多核苷酸。在一些实施方案中,本公开提供包含与SEQ ID NO:40的核苷酸1-2602具有至少99%同一性的序列的因子IX多核苷酸。在一些实施方案中,本公开提供包含与SEQ ID NO:40的核苷酸1-2602具有至少99.5%同一性的序列的因子IX多核苷酸。在一些实施方案中,本公开提供包含SEQ ID NO:40的核苷酸1-2602的序列的因子IX多核苷酸。

[0355] 已经表征了几种AAV血清型,包括AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8和AAV9。通常,任何AAV血清型均可以用于本文所描述的因子IX基因疗法构建体。然而,这些血清型具有不同趋向性,例如它们优先感染不同组织。在一个实施方案中,因为因子IX主要在肝脏中产生,所以基于至少在血清型AAV7、AAV8和AAV9中发现的肝趋向性来选择用于所公开的基因疗法构建体的AAV血清型。因此,在一个实施方案中,因子IX基因疗法构建体为AAV7血清型载体。在另一个实施方案中,因子IX基因疗法构建体为AAV8血清型载体。在又一个实施方案中,因子IX基因疗法构建体为AAV9血清型载体。

[0356] 本文所描述的因子IX基因疗法构建体可以为单链的(例如ssAAV载体,如在图1C和图1D中所说明)或自身互补的(例如scAAV载体,如在图1A和图1B中所说明)。虽然研究和理论已表明自身互补型AAV载体应有助于更好的转基因表达,但通过避开在翻译之前对第二条链合成的需要,如实施例5中所报道,识别出单链AAV载体促成比可比的自身互补型载体更好的因子IX表达。

[0357] 启动子和增强子

[0358] 本文所描述的因子IX基因疗法构建体通常包括驱动体内基因表达的一个或多个启动子和/或增强子元件,例如调控元件。在一些实施方案中,启动子或增强子元件以组织依赖性方式,例如主要在特定组织中驱动表达。因为因子IX主要在肝脏中合成,所以在一些实施方案中,本文所描述的基因疗法载体包括肝特异性调控元件,其基本上使基因疗法载体的表达限于肝细胞。

[0359] 通常,肝特异性调控元件可以源自已知只在肝脏中表达的任何基因。WO 2009/130208识别出几种以肝特异性方式表达的基因,包括serpin肽酶抑制剂分支A成员1(也称为 α -抗胰蛋白酶)(SERPINA1;GeneID 5265)、载脂蛋白C-I(APOC1;GeneID 341)、载脂蛋白C-IV(APOC4;GeneID 346)、载脂蛋白H(APOH;GeneID 350);甲状腺素运载蛋白(TTR;GeneID 7276)、白蛋白(ALB;GeneID 213)、醛缩酶B(ALDOB;GeneID 229)、细胞色素P450家族2亚家族E多肽1(CYP2E1;GeneID 1571)、纤维蛋白原 α 链(FGA;GeneID 2243)、转铁蛋白(TF;

GeneID 7018)、结合珠蛋白相关蛋白(HPR;GeneID 3250)。在一些实施方案中,本文所描述的因子IX基因疗法构建体包括源自这些蛋白质中的一个或多个的基因组基因座的肝特异性调控元件。此类元件的几个实例描述于WO 2009/130208中,所述专利的内容出于所有目的以全文引用的方式明确并入本文中。

[0360] 肝特异性调控元件的一个实例是来自于甲状腺素运载蛋白(TTR)基因,通常称为“TTR_e”或“TTR_{Enh}”。Hsieh J.L.等人,Cancer Sci.,100(3):537-45(2009),其内容出于所有目的以全文引用的方式明确并入本文中。在一些实施方案中,本文所描述的因子IX基因疗法构建体包括截短的TTR增强子和启动子元件。这些元件的实例提供于如图25所提供的CS06-CRM8.3-ssV(SEQ ID NO:40)的核苷酸383-712处。在一些实施方案中,截短的TTR增强子和启动子元件与CS06-CRM8.3-ssV(SEQ ID NO:40)的核苷酸383-712具有至少85%的序列同一性。在其他实施方案中,截短的TTR增强子和启动子元件与CS06-CRM8.3-ssV(SEQ ID NO:40)的核苷酸383-712具有至少90%、95%、96%、97%、98%、99%、99.5%或100%的序列同一性。

[0361] 肝特异性调控元件的另一个实例是来自于如PCT公布号W02016/146757中所描述的SERPINA1基因,所述公布的内容出于所有目的以全文引用的方式明确并入本文中。此类元件的一个实例为提供于CS06-CRM8.3-ssV(SEQ ID NO:40)的核苷酸165-236处的CRM8调控控制元件(SEQ ID NO:39)。在一些实施方案中,SERPINA1源性调控控制元件与CRM8(SEQ ID NO:39)具有至少85%的序列同一性。在其他实施方案中,截短的SERPINA1源性调控控制元件与CRM8(SEQ ID NO:39)具有至少90%、95%、96%、97%、98%、99%或100%的序列同一性。

[0362] 在一些实施方案中,如由图1B和图1D中说明的构建体所例示,因子IX基因疗法构建体包括一个或多个SERPINA1源性调控控制元件。在一个实施方案中,构建体包括一个SERPINA1源性调控控制元件(例如CRM8)。在另一个实施方案中,构建体包括两个SERPINA1源性调控控制元件(例如CRM8)。在另一个实施方案中,构建体包括三个SERPINA1源性调控控制元件(例如CRM8)。在其他实施方案中,构建体包括4个、5个、6个或更多个SERPINA1源性调控控制元件(例如CRM8)。

[0363] 在一个实施方案中,如图1B、图1D和图25中所例示,因子IX基因疗法构建体包括一个或多个SERPINA1源性调控控制元件(例如CRM8)以及截短的TTR增强子和启动子元件。

[0364] 内含子

[0365] 在一些实施方案中,本文所描述的因子IX基因疗法构建体包括内含子,例如病毒源性内含子,以增加因子IX基因的表达。适用于基因疗法构建体的表达的内含子为本领域中已知的。通常,如图1和图25中所示的因子IX构建体中所例示,内含子位于转基因编码序列的5'处。然而,在一些实施方案中,内含子可以例如在天然因子IX内含子接点处位于转基因编码序列内,或另外地,位于转基因编码序列的3'处。可以用于本文所描述的因子IX基因疗法构建体中的内含子的非限制性实例包括源自以下各项的内含子:小鼠微小病毒(MVM)内含子、 β -球蛋白内含子(β 1VS-11)、因子IX(FIX)内含子A、猿猴病毒40(SV40)小T内含子和 β -肌动蛋白内含子。

[0366] 在一个实施方案中,本文所描述的因子IX基因疗法构建体包括例如如在图1中所说明和由图25中的CS06-CRM8.3-ssV(SEQ ID NO:40)的核苷酸724-800处的MVM内含子(SEQ

ID NO:53)所例示的MVM源性内含子。在一些实施方案中,用于本文所描述的基因疗法构建体中的内含子与MVM(SEQ ID NO:53)具有至少85%的序列同一性。在其他实施方案中,用于本文所描述的基因疗法构建体中的内含子与MVM(SEQ ID NO:53)具有至少90%、95%、96%、97%、98%、99%或100%的序列同一性。

[0367] 多聚腺苷酸化信号

[0368] 在一些实施方案中,例如如在图1中所说明,本文所描述的因子IX基因疗法构建体包括多聚腺苷酸化信号。多聚腺苷酸化信号在由因子IX转基因生成的mRNA转录物的3'端引导poly-A尾的合成。因此,多聚腺苷酸化信号位于因子IX编码序列的3'处。可以用于本文所描述的因子IX基因疗法构建体中的多聚腺苷酸化信号的非限制性实例包括源自以下各项的多聚腺苷酸化信号:猿猴病毒40(SV40)晚期基因、牛生长激素(BGH)多聚腺苷酸化信号和最小兔 β -球蛋白(mRBG)基因。

[0369] 在一个实施方案中,本文所描述的因子IX基因疗法构建体包括例如如在图1中所说明和由图25中的CS06-CRM8.3-ssV(SEQ ID NO:40)的核苷酸2208-2441处的BGHpA信号(SEQ ID NO:54)所例示的源自牛生长激素(BGH)多聚腺苷酸化信号的多聚腺苷酸化信号。在一些实施方案中,用于本文所描述的基因疗法构建体中的多聚腺苷酸化信号与BGHpA信号(SEQ ID NO:54)具有至少85%的序列同一性。在其他实施方案中,用于本文所描述的基因疗法构建体中的多聚腺苷酸化信号与BGHpA信号(SEQ ID NO:54)具有至少90%、95%、96%、97%、98%、99%或100%的序列同一性。

[0370] V. 方法

[0371] 产生

[0372] 本文所描述的密码子改变的因子IX多核苷酸和病毒载体(例如核酸组合物)是根据用于核酸扩增和载体产生的常规方法来产生。已经开发出两种主要平台用于大规模产生重组AAV载体。第一平台是基于在哺乳动物细胞中的复制,而第二平台是基于在无脊椎动物细胞中的复制。关于综述,参见Kotin R.M., Hum.Mol.Genet., 20(R1):R2-6(2011),其内容出于所有目的以全文引用的方式明确并入本文中。

[0373] 因此,本公开提供用于产生腺相关病毒(AAV)颗粒的方法。在一些实施方案中,这些方法包括将如本文所描述与CS02、CS03、CS04、CS05或CS06序列之一具有高度核苷酸序列同一性(例如至少95%、96%、97%、98%、99%、99.5%、99.9%或100%)的密码子改变的因子IX多核苷酸构建体引入宿主细胞中,其中多核苷酸构建体有能力在宿主细胞中进行复制。

[0374] 在一些实施方案中,宿主细胞为哺乳动物宿主细胞,例如HEK、CHO或BHK细胞。在一个具体实施方案中,宿主细胞为HEK 293细胞。在一些实施方案中,宿主细胞为无脊椎动物细胞,例如昆虫细胞。在一个具体实施方案中,宿主细胞为SF9细胞。

[0375] 制剂

[0376] 本文中提供用于治疗出血病症的组合物。此类组合物含有治疗有效量的密码子改变的因子IX多核苷酸,例如如本文所描述包括编码因子IX的密码子改变的多核苷酸的AAV基因疗法载体。治疗有效量的密码子改变的FIX多核苷酸(例如包括密码子改变的因子IX编码序列的AAV基因疗法载体)与适用于全身、表面或局部施用的药物载体或媒介物混合。本文所公开的密码子改变的因子IX多核苷酸的最终制剂将属于本领域技术人员的能力范围

内。

[0377] 剂量

[0378] 将本发明的核酸组合物施用给有需要的患者。所施用的治疗性基因疗法药剂的量或剂量取决于诸如以下的因素：特定密码子改变的FIX多核苷酸构建体、所用的递送载体、疾病严重程度和个体的一般特征。确切剂量将视治疗目的而定，并且将可由本领域技术人员使用已知技术确定（参见例如Lieberman, Pharmaceutical Dosage Forms (第1-3卷, 1992); Lloyd, The Art, Science and Technology of Pharmaceutical Compounding (1999); Pickar, Dosage Calculations (1999); 和Remington: The Science and Practice of Pharmacy, 第20版, 2003, Gennaro编, Lippincott, Williams & Wilkins)。确定用于治疗特定受试者的特定剂量和给药方案在熟练的医生的能力范围内。

[0379] 在一些实施方案中，以治疗有效剂量向有需要的受试者（例如患有轻度、中度或重度血友病B的受试者）静脉内施用具有密码子改变的因子IX多核苷酸的基因疗法载体（例如AAV基因疗法载体颗粒）。在一些实施方案中，治疗有效剂量在每千克受试者体重约 2×10^{11} 与 2×10^{14} 个载体基因组之间。在一个具体实施方案中，治疗有效剂量在每千克受试者体重约 2×10^{12} 与 2×10^{13} 个载体基因组之间。在一些实施方案中，向受试者施用每千克受试者体重约 2×10^{11} 、 3×10^{11} 、 4×10^{11} 、 5×10^{11} 、 6×10^{11} 、 7×10^{11} 、 8×10^{11} 、 9×10^{11} 、 1×10^{12} 、 2×10^{12} 、 3×10^{12} 、 4×10^{12} 、 5×10^{12} 、 6×10^{12} 、 7×10^{12} 、 8×10^{12} 、 9×10^{12} 、 1×10^{13} 、 2×10^{13} 、 3×10^{13} 、 4×10^{13} 、 5×10^{13} 、 6×10^{13} 、 7×10^{13} 、 8×10^{13} 、 9×10^{13} 、 1×10^{14} 或 2×10^{14} 个载体基因组。

[0380] 因此，本公开提供用于治疗因子IX缺乏症（例如血友病B）的方法。在一些实施方案中，这些方法包括向有需要的患者施用如本文所描述与CS02、CS03、CS04、CS05或CS06序列之一具有高度核苷酸序列同一性（例如至少95%、96%、97%、98%、99%、99.5%、99.9%或100%）的密码子改变的因子IX多核苷酸构建体。在一些实施方案中，密码子改变的因子多核苷酸与密码子改变的因子IX前原多肽编码序列具有高度序列同一性，例如与CS02-FL-NA (SEQ ID NO:5)、CS03-FL-NA (SEQ ID NO:6)、CS04-FL-NA (SEQ ID NO:7)、CS05-FL-NA (SEQ ID NO:8) 或CS06-FL-NA (SEQ ID NO:9) 之一具有高度序列同一性。在一些实施方案中，密码子改变的因子多核苷酸与密码子改变的成熟因子IX单链多肽编码序列具有高度序列同一性，例如与CS02-MP-NA (SEQ ID NO:13)、CS03-MP-NA (SEQ ID NO:14)、CS04-MP-NA (SEQ ID NO:15)、CS05-MP-NA (SEQ ID NO:16) 或CS06-MP-NA (SEQ ID NO:17) 之一具有高度序列同一性。

[0381] 在一些实施方案中，治疗包括向有需要的患者施用包括如本文所描述与CS02、CS03、CS04、CS05或CS06序列之一具有高度核苷酸序列同一性（例如至少95%、96%、97%、98%、99%、99.5%、99.9%或100%）的密码子改变的因子IX多核苷酸构建体的基因疗法载体。在一个实施方案中，基因疗法载体为哺乳动物基因疗法载体。在一个具体实施方案中，哺乳动物基因疗法载体为病毒载体，例如慢病毒、逆转录病毒、腺病毒或腺相关病毒载体。

[0382] 在一个实施方案中，基因疗法载体为具有编码密码子改变的因子IX编码序列的病毒载体的腺相关病毒 (AAV) 颗粒。通常，病毒载体包括各末端的反向末端重复序列 (ITR)、一个或多个表达调控元件、密码子改变的因子IX编码序列和poly-A信号序列。在一个具体实施方案中，基因疗法载体包括肝特异性调控控制元件（例如CRM8元件的一个或多个拷贝）。

[0383] 产生

[0384] 本文所描述的密码子改变的因子IX多核苷酸和病毒载体(例如核酸组合物)是根据用于核酸扩增和载体产生的常规方法来产生。已经开发出两种主要平台用于大规模产生重组AAV载体。第一平台是基于在哺乳动物细胞中的复制,而第二平台是基于在无脊椎动物细胞中的复制。关于综述,参见Kotin R.M., Hum.Mol.Genet., 20(R1):R2-6(2011),其内容出于所有目的以全文引用的方式明确并入本文中。

[0385] 因此,本公开提供用于产生腺相关病毒(AAV)颗粒的方法。在一些实施方案中,这些方法包括将如本文所描述与CS02、CS03、CS04、CS05或CS06序列之一具有高度核苷酸序列同一性(例如至少95%、96%、97%、98%、99%、99.5%、99.9%或100%)的密码子改变的因子IX多核苷酸构建体引入宿主细胞中,其中多核苷酸构建体有能力在宿主细胞中进行复制。

[0386] 在一些实施方案中,宿主细胞为哺乳动物宿主细胞,例如HEK、CHO或BHK细胞。在一个具体实施方案中,宿主细胞为HEK 293细胞。在一些实施方案中,宿主细胞为无脊椎动物细胞,例如昆虫细胞。在一个具体实施方案中,宿主细胞为SF9细胞。

[0387] 治疗

[0388] 在一些实施方案中,根据已知施用方法向有需要的受试者施用本文所描述的核酸组合物(例如密码子改变的多核苷酸)。用于施用基因疗法载体的方法在本领域中为人熟知。这些方法包括但不限于静脉内施用、肌肉内注射、间质内注射和肝内施用(例如肝内动脉或静脉)。举例来说,参见Chuah MK等人, Hum Gene Ther., 23(6):557-65(2012); Chuah MK等人, J Thromb Haemost., 10(8):1566-69(2012); Chuah MK等人, J Thromb Haemost. 11增刊1:99-110(2013); VandenDriessche等人, Hum Gene Ther. 23(1):4-6(2012); High KA, Blood, 120(23):4482-87(2012); Matrai等人, Mol Ther., 18(3):477-90(2010); 和Matrai等人, Curr Opin Hematol., 17(5):387-92(2010),其各自关于综述通过引用并入本文中。

[0389] 评估治疗功效

[0390] 血友病B治疗的治疗功效可以例如通过测量来自所治疗受试者的血液的因子IX依赖性凝结潜能来评估。用于评估凝结潜能的计量法包括但不限于体外活化部分凝血活酶时间测定(APPT)、因子IX生色活性测定、血液凝结时间和因子IX抗原水平(例如使用因子IX特异性ELISA)。应注意,治疗剂量并不需要在患者中产生野生型水平的FIX;确切地说,足以以有意义或可测量的方式减轻症状的表达出于本发明的目的被视为治疗性的。

[0391] 根据美国血友病基金会(National Hemophilia Foundation),当受试者的血浆含有正常人血浆因子IX活性的6%至49%时,将该受试者归类为患有轻度血友病B。患有轻度血友病B的受试者通常仅在重伤、创伤或手术之后发生出血。在许多情况下,直至受伤、手术或拔牙导致长时间出血才诊断为轻度血友病。可能直至成年才发生第一事件。患有轻度血友病的女性常常会出现月经过多、月经期反应大,且可能会产后出血。

[0392] 根据美国血友病基金会,当受试者的血浆含有正常人血浆的因子IX活性的1%至5%时,将该受试者归类为患有中度血友病B。患有中度血友病B的受试者倾向于在受伤后发生出血事件。没有明显原因发生的出血被称为自发出血事件。

[0393] 根据美国血友病基金会,当受试者的血浆含有不到正常人血浆因子IX活性的1%时,将该受试者归类为患有重度血友病B。患有重度血友病B的受试者在受伤后会流血并且

可能会频繁发生自发出血事件,常常进入其关节和肌肉中。

[0394] 在一些实施方案中,正常人血浆被定义为每毫升含有1IU的因子IX活性。因此,在一些实施方案中,归类为患有轻度血友病B的受试者的血浆含有每毫升0.06与0.49IU之间的因子IX活性。在一些实施方案中,归类为患有中度血友病B的受试者的血浆含有每毫升0.01与0.05IU之间的因子IX活性。在一些实施方案中,归类为患有重度血友病B的受试者的血浆含有每毫升0.01与0.05IU之间的因子IX活性。

[0395] 因此,在一些实施方案中,当血友病B疗法使受试者的血液/血浆中的平均因子IX活性水平升高时,所述血友病B疗法为治疗有效的。在一些实施方案中,治疗有效的治疗使受试者的血液/血浆中的平均因子IX活性水平升高至少1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、50%或更高百分比。在一个具体实施方案中,治疗有效的血友病疗法使受试者血液/血浆中的平均因子IX活性增加至少5%。在另一个具体实施方案中,治疗有效的血友病疗法使受试者血液/血浆中的平均因子IX活性增加至少10%。在另一个具体实施方案中,治疗有效的血友病疗法使受试者血液/血浆中的平均因子IX活性增加至少15%。在另一个具体实施方案中,治疗有效的血友病疗法使受试者血液/血浆中的平均因子IX活性增加至少20%。在另一个具体实施方案中,治疗有效的血友病疗法使受试者血液/血浆中的平均因子IX活性增加至少25%。在另一个具体实施方案中,治疗有效的血友病疗法使受试者血液/血浆中的平均因子IX活性增加至少30%。

[0396] 在一些实施方案中,治疗有效的治疗使受试者血液中的平均因子IX活性水平升高,使得受试者归类为患有不太严重的血友病B形式。举例来说,在一个实施方案中,最初归类为患有重度血友病B的受试者在经历治疗有效的治疗之后重新归类为患有中度血友病B或轻度血友病B。在另一个实施方案中,最初归类为患有中度血友病B的受试者在经历治疗有效的治疗之后重新归类为患有轻度血友病B。

[0397] VI. 实施例

[0398] 实施例1-密码子改变的因子IX表达序列提高FIX表达水平

[0399] 为了产生提供改进的异源因子IX体内表达的基因疗法构建体,构建一组编码具有R384L氨基酸取代的全长因子IX前原蛋白(FIXp-FL-AA SEQ ID NO:4)的自身互补型基于AAV8的载体。通过几个步骤改变各构建体的因子IX编码序列来改善在人体中的表达。根据被设计用来说明优选/不利序列基序并且使密码子使用偏向于优选人密码子的算法对各因子IX编码序列进行修饰。在此第一步骤中使用如表2中所报道的几个算法。接着对由应用表2中所报道的算法产生的中间密码子改变的序列进一步作出手工修饰以减少或消除CpG二核苷酸,调整最终GC含量,为了允许优选密码子对而作出调整,为了避免不利密码子对而作出调整,并且调整最终密码子使用。关于这些考虑因素的其他信息,参见例如Fath S.等人,PLoS.One.,6,e17596(2011);Haas J.等人,Curr.Biol.,6,315-324(1996);Tats A.,BMC Genomics.9:463(2008);Grote A.等人,Nucleic Acids Research,33(网站服务器问题(Web Server issue)),W526-W531(2005);Mirsafian H.等人,Scientific World Journal.,639682(2014);和Pechmann S.等人,Nat Struct Mol Biol.20(2):237-43(2013),这些文献的内容出于所有目的,尤其关于其对于密码子改变考虑因素的教导内容以全文引用的方式明确并入本文中。

[0400] 各个所产生的密码子改变的编码序列(例如分别显示于图5至图9中的CS02、CS03、

CS04、CS05和CS06) 编码同一FIX (R384L) 蛋白 (FIXp-FL-AA (SEQ ID NO:4))。CS02、CS03和CS04构建体不含CpG基序,而CS05和CS06分别含有11个和3个CpG。

[0401] 为了用作对照,还产生了在有和没有R384L Padua氨基酸取代的情况下合并野生型FIX编码序列的载体构建体。WH01构建体编码不具有R384L Padua突变的野生型FIX前原蛋白,并且包括20个CpG二核苷酸。WH02构建体编码具有R384L Padua突变的野生型FIX前原蛋白,并且包括19个CpG二核苷酸。

[0402] WH01和WH02构建体分别在其编码序列中包括20个和19个CpG。相比之下,CS02、CS03和CS04构建体不含CpG基序,而CS05和CS06构建体分别含有11个和3个CpG。

[0403] 如图1A中所示,将密码子改变的因子IX编码序列插入腺相关病毒 (“AAV”) 转基因盒中,所述腺相关病毒转基因盒含有小鼠甲状腺素运载蛋白增强子/启动子 (SEQ ID NO:52)、小鼠微小病毒 (“MVM”) 内含子 (SEQ ID NO:53)、包括R384L “Padua” 氨基酸取代的密码子改变的FIX构建体 (美国专利No.6,531,298;Simione等人,NEJM361:1671-75(2009);参考野生型精氨酸在缺乏信号肽和原肽的人单链FIX蛋白中的位置,通常将R384L突变报道为R338L突变),随后为牛生长激素polyA元件 (SEQ ID NO:54)。所述基因盒侧接AAV2反向末端重复序列 (“ITR”) (SEQ ID NO:51和55)。左侧ITR重复序列在末端多联体拆分位点中包括突变,从而产生载体的自身互补型 (sc) 表型。基本载体设计详细描述于Wu等人,Mol. Ther. 16:280-89(2008) 和PCT公布号W0 2014/064277A1中,这些参考文献的内容出于所有目的以全文引用的方式并入本文中。

[0404] 向B6/129P2-F9tm1Dws FIX敲除小鼠 (描述于Lin等人,Blood,90:3962-66(1997) 中,所述参考文献的内容出于所有目的以全文引用的方式并入本文中) 施用CS和WH因子IX AAV构建体。基于个别动物体重 (4x10E11个载体基因组/千克 (vg/kg) 体重) 经由外侧尾部静脉将AAV载体稀释液注入动物 (每组4-8只动物) 体内。在给药之后根据已知程序使用玻璃毛细管通过眶后穿刺按规定的间隔收集血液样品。接着将血液转移至预先填充有柠檬酸钠抗凝剂的管子中并且通过标准程序获得血浆并在-20℃下冷冻。

[0405] 测定各种因子IX构建体的表达并且如表2中所报道,使用FIX敲除小鼠中第14天的血浆FIX水平来判断在将载体经尾部静脉注射至小鼠体内之后的构建体效能。到第14天,基因敲除小鼠模型中的表达水平几乎达至最大FIX表达。如表2中所示,在施用4x10E11个载体基因组/千克 (vg/kg) 体重之后第14天WH02 FIX (R384L) 对照构建体以1.03个单位的FIX进行表达。使用此表达水平作为基线来测定密码子改变的因子IX构建体的富集倍数。如表2中所报道,与由野生型多核苷酸序列编码的WH02对照构建体相比,CS密码子改变的构建体提供增加约2倍与4倍之间的表达。最值得注意的是,CS06密码子改变的构建体提供为WH02对照构建体4.2倍大的因子IX活性和为WH01 (野生型因子IX) 对照构建体21.6倍大的因子IX活性。

[0406] 表2. 来自具有野生型密码子序列 (WH01-wtFIX;WH02-FIX (R384L) 和密码子改变的序列 (CS02-CS06) 的构建体的因子IX表达。

[0407]

AAV 构建体	载体基因组的 修饰	CpG 的 数目	第 14 天表 达水平 [人 FIX%]	与 WH01 相比的表 达倍数	与 WH02 相比的表 达倍数
WH01	不具有 R338L (Padua)突变的人 FIX 野生型序列(GeneBank NM000133.3)	20	0.20	1.0	0.19
WH02	具有 R338L (Padua)突变的人 FIX 野生型序列	19	1.03	5.2	1.0
CS02	具有 R338L 突变的人 FIX 序列; 针对人类血清白蛋白密码子进一步优化的 Geneart 基本算法。	0	2.12	10.6	2.1
CS03	具有 R338L 突变的人 FIX 序列; 针对最常用的人密码子进一步优化的 Geneart 基本算法(Haas 等人, 1996. Curr Biol. 6, 315-324)。	0	1.98	9.9	1.9
CS04	具有 R338L 突变的人 FIX 序列; 如 Uhlen 等	0	2.77	13.9	2.7

[0408]		人, 2015 Science 347, 6220 中所描述针对肝密码子使用进一步优化的 Geneart 基本算法。				
	CS05	具有 R338L 突变的人 FIX 序列; 被修改以减少 CpG 的 JCAT 算法; (Grote 等人, 2005. Nucleic Acids Res 1,33)。	11	3.93	19.7	3.8
	CS06	具有 R338L 突变的人 FIX 序列; 针对最常用的人密码子进一步优化的 Geneart 基本算法(Haas 等人, 1996. Curr Biol. 6, 315-324)。	3	4.32	21.6	4.2

[0409] 实施例2-肝特异性CRM8元件提高FIX在小鼠中的表达

[0410] 为了进一步增加来自密码子改变的构建体的因子IX表达和活性, 将如Nair等人, Blood 123:3195-99 (2014) 中所报道的肝特异性顺式调控控制元件 (CRM8 (SEQ ID NO:39)) 的一个至三个拷贝并入基因盒中, 形成图示于图1B中的构建体。通过尾部静脉途径将具有CS02密码子改变的FIX编码序列加上零个 (CS02-CRM8.0-V)、一个 (CS02-CRM8.1-V)、两个 (CS02-CRM8.2-V) 或三个 (CS02-CRM8.3-V) CRM8对照元件的AAV载体注入野生型小鼠体内。接着使用人FIX特异性ELISA测定随时间推移测量小鼠血浆中的人FIX抗原。

[0411] 如表3中所报道, 在感染后21天, 与来自缺乏CRM8元件的对照构建体的表达相比, CRM8调控元件的使用使因子IX体内表达增加至约2倍和4倍。举例来说, 含有单个CRM8元件的CS02-CRM8.1-V载体提供两倍于CS02-CRM8.0-V对照载体的FIX表达。包括CRM8元件的多个拷贝进一步改善此表达。举例来说, 相对于对照载体, 含有CRM8元件的2个拷贝的载体提供三倍的表达, 并且含有CRM8元件的3个拷贝的载体提供3.4倍的表达。

[0412] 表3. 注射了具有CRM8调控控制元件的0-3个拷贝的密码子改变的AAV载体的野生型小鼠的血浆中的因子IX表达水平。

[0413]	#	AAV 构建体	CRM8 元件的数目	FIX (ng/ml) 第 4 天	FIX (ng/ml) 第 11 天	FIX (ng/ml) 第 21 天	倍数增加 第 21 天
	1	CS02-CRM8.0-scV	0	65.8	133.4	239.2	1.0
	2	CS02-CRM8.1-scV	1	120.7	250.7	442.8	1.9
	3	CS02-CRM8.2-scV	2	152.9	417.3	713.8	3.0
	4	CS02-CRM8.3-scV	3	130.9	432.6	800.9	3.4

[0414] 实施例3-肝特异性CRM8元件提高FIX在人肝细胞中的表达

[0415] 通过用人肝细胞系HepG2进行的体外生物效能测定对如实施例2中所描述含有CRM8肝特异性调控控制元件的0-3个拷贝的CS02因子IX基因疗法构建体进行进一步测试。简单来说,如实施例2中所描述使HepG2细胞感染CS02-CRM8-V AAV载体之一,并且在感染之后三天,通过生色底物测定来测量FIX活性。与实施例2中所报道的结果一致,如表4中所报道,含有CRM8调控控制元件的所有载体提供较高的FIX表达。明显地,与在小鼠模型中相比,在人HepG2细胞中使用多个CRM8元件的效果甚至更显著。举例来说,相对于对照载体,含有CRM8元件的2个拷贝的载体提供6.7倍的表达并且含有CRM8元件的3个拷贝的载体提供12.8倍的表达。此证实了在这些载体中CRM8调控控制元件对FIX表达具有的积极作用。

[0416] 表4.注射了具有CRM8调控控制元件的0-3个拷贝的密码子改变的AAV载体的人肝HepG2细胞中的因子IX表达水平。

[0417]	AAV 构建体	CRM8 元件的	FIX 活性[生物效	倍数增加
		数目	能单位]	
	CS02-CRM8.0-scV	0	0.35	1
	CS02-CRM8.1-scV	1	0.82	2.3
	CS02-CRM8.2-scV	2	2.36	6.7
[0418]	CS02-CRM8.3-scV	3	4.48	12.8

[0419] 实施例4-单链FIX AAV8载体提供与可比的自身互补型载体类似的体内表达

[0420] 在一些情况下,自身互补型(sc)AAV载体比类似单链(ss)AAV载体更有效地表达转基因盒。此大概是因为自身互补型载体基因组在细胞核中脱壳之后更快速地形成双链所致。关于综述,参见McCarty D., Mol. Ther., (16):1648-56(2008),其内容出于所有目的以全文引用的方式并入本文中。

[0421] 最近的一项研究使用EGFP载体证实了此效果。Bell等人, Hum. Gene Ther. Methods, (27):228-37(2016)。然而,所述研究还表明此效果对转基因和剂量具依赖性。举例来说,与相应单链载体相比,在自身互补型AAV8载体中的人鸟氨酸氨甲酰基转移酶(h0TC)基因盒显示在小鼠肝中在低剂量下有更好的表达。然而,在高剂量下未能展示出此效果,表明至少在所研究的非分泌性基因中该效果对转基因和剂量具依赖性。同上。

[0422] 为了探索所公开的密码子改变的FIX基因构建体在单链和自身互补型设计的情形中的特性,分别如图1D和图1C中所图示,在有和无CRM8调控控制元件的情况下构建具有CS06密码子改变的FIX(R338L)基因和两个完整ITR的单链构建体。在HEK293细胞系统中产生单链(ss)载体,并且将因子IX表达与实施例1-3中所报道的自身互补型构建体的表达相

比较。

[0423] 首先,在如上文所描述将自身互补型(sc)和单链(ss)CS06-CRM8.0-V构建体注入B6/129P2-F9tm1Dws FIX敲除小鼠体内之后,对所述构建体进行体内测试。令人惊讶的是,如表5中所报道,自身互补型(sc)和单链(ss)CS06载体构建体显示非常类似的血浆FIX活性水平,从而表明对于本文所描述的密码子改变的因子IX构建体,所报道的sc载体与ss载体相比的优点并不成立。表达取决于许多参数,包括转基因构建体、转录物稳定性、构建体中所用的启动子、时间和剂量。如表5中所示,在被选择用来校正出血并且在FIX ko小鼠中获得长期表达的条件下,相应的sc和ss载体提供基本上类似的表达水平。

[0424] 还在单链载体背景中研究了肝特异性CRM8调控控制元件对FIX表达的影响。如表5中所报道,在单链载体中包括一个CRM8元件改善了B6/129P2-F9tm1Dws FIX敲除小鼠中的FIX表达。包括三个CRM8元件进一步改善了来自单链CS06构建体的FIX表达,直至略为缺乏CRM8元件的自身互补型CS06对照物的2倍的水平。与野生型WH02构建体相比,当与三个CRM8调控控制元件配对时,单链CS06载体提供高达7倍大的表达。

[0425] 表5.注射了各个单链(ss)和自身互补型(sc)AAV因子IX载体的FIX敲除小鼠中的因子IX表达水平。

[0426]	AAV 构建体	CRM 8 元件的数目	表达水平 ¹ (第 7 天)	表达水平 (第 14 天)	表达水平 (第 28 天)	相较于 CS06 的倍数增加 (第 7 天/第 14 天/第 28 天)	相较于 WH02 的倍数增加 (第 7 天/第 14 天/第 28 天)
	CS06-CRM 8.0-ssV	0	1.38	2.73	2.99	0.7 / 0.9 / 0.9	1.5 / 2.7 / 2.7
	CS06-CRM 8.1-ssV	1	1.92	3.57	3.47	1.0 / 1.1 / 1.0	2.1 / 3.5 / 3.2
	CS06-CRM 8.3-ssV (SEQ ID NO:40)	3	4.43	6.65	7.78	2.3 / 2.1 / 2.2	4.9 / 6.5 / 7.1
	CS06-CRM 8.0-scV	0	1.89	3.17	3.50	1.0 / 1.0 / 1.0	2.1 / 3.1 / 3.2
	WH02-CRM 8.0-scV	0	0.90	1.03	1.10	0.5 / 0.3 / 0.3	1.0
¹ 以国际单位计的 FIX 活性(7-8 只小鼠的平均值); d: 天;							

[0427] 实施例5-单链FIX AAV8载体在人肝细胞中提供比可比的自身互补型载体更好的FIX表达

[0428] 接着在人HepG2细胞中研究来自实施例4中所描述的单链CS06载体的因子IX表达并且与类似自身互补型载体构建体相比较。与实施例4中所报道的体内结果一致,在HepG2细胞中不具有CRM8元件的单链CS06载体以比较性自身互补型载体略微较低的水平提供FIX表达。然而,如表6中所报道,包括单个CRM8元件使来自单链CS06载体的FIX表达增加至为来自自身互补型CS06载体的表达的2.6倍大的水平。

[0429] 然而,最令人惊讶的是,在单链CS06载体中包括三个CRM8元件使FIX表达增加至为

来自自身互补型CS06载体的表达的16.8倍大的水平。增加的FIX表达大于来自WH02对照载体的FIX表达超过100倍。概括地说,在体内和体外生物效能测定中含有三个CRM8元件的单链CS06载体提供最高的表达水平。

[0430] 表6. 在人肝细胞中来自单链(ss)和自身互补型载体的因子IX表达水平。

[0431]	AAV 构建体	CRM8 元件的数目	FIX 活性[生物效能单位]	相较于CS06-CRM8.0-scV 的倍数增加	相较于WH02-CRM8.0-scV 的倍数增加
	CS06-CRM8.0-ssV	0	0.24	0.6	3.9
	CS06-CRM8.1-ssV	1	0.95	2.6	15.9
	CS06-CRM8.3-ssV (SEQ ID NO: 40)	3	6.20	16.8	103.3
	CS06-CRM8.0-scV	0	0.37	1.0	6.2
	WH02-CRM8.0-scV	0	0.06	0.2	1.0

[0432] 实施例1-5的材料及方法

[0433] 动物实验。对于FIX敲除模型中的实验,使用FIX ko小鼠品系B6/129P2-F9tm1Dws (由Lin等人,1997.Blood 90:3962-6开发)。在野生型小鼠模型中,使用4-5周龄的雄性C57BL6-J B16小鼠。两种品系均获自商业育种者。基于个别动物体重经由外侧尾部静脉将AAV载体稀释液注入动物(每组4-8只动物)体内。在给药之后根据已知程序使用玻璃毛细管通过眶后穿刺按所规定的时间间隔收集血样。接着将血液转移至预先填充有柠檬酸钠抗凝剂的管子中并且通过标准程序获得血浆并在-20℃下冷冻。

[0434] 在HepG2细胞中进行的包括FIX生色底物测定的体外生物效能测定。在人肝细胞系HepG2(ATCC HB-8065)中进行对基因疗法载体制剂的体外生物效能测定。在用羟基脲处理之后,使细胞感染AAV8FIX载体并且孵育约96小时。在孵育时间期间,使FIX表达并且释放至细胞上清液中,并且通过生色端点测量(Rossix AB,Sweden)测定FIX活性。各测定操作包括所纯化的AAV-FIX载体材料的标准曲线,使用在700与7000之间的范围内的MOI。将标准物在MOI 3270下的FIX活性设置为生物效能单位(BPU)1。

[0435] 在小鼠血浆中的人FIX定量。为了定量基因敲除小鼠血浆中的人FIX,使用标准FIX凝血分析方法进行FIX凝聚测定。为了定量野生型小鼠血浆中的人FIX抗原,使用专门检测人FIX的市售ELISA试剂盒(ASSERACHROM IX:AG(产品目录号00943Stago BNL))。

[0436] 实施例6-因并入CRM8元件而改善转录功效

[0437] 为了解释含有CRM8的载体的改进的生物效能是否由增加的转录功效引起,将人肝细胞系(HepG2)和小鼠肝细胞(FIX敲除小鼠)用含有0个、1个或3个CRM8元件的单链CS06载体转导。测定FIX mRNA和DNA水平并且呈现为归一化的FIX mRNA与DNA水平之间的比率。

[0438] 在体外模型中,包括一个CRM8元件(CS06-CRM8.1-ssV)或三个CRM8元件(CS06-CRM8.3-ssV(SEQ ID NO:40))在所转导的人肝细胞中分别产生为在用缺乏CRM8元件的载体(CS06-CRM8.0-ssV)转导的细胞中的5倍和23倍高的人FIX mRNA水平(表6)。类似地,在体内模型中,在鼠肝中来自含有一个或三个CRM8元件的载体的FIX表达分别为来自缺乏CRM8元件的载体的FIX表达的2.0倍和2.8倍高(表6)。两种模型均支持CRM8元件在改善FIX构建体

的转录活性方面提供有益效果。

[0439] 表7. 人肝细胞系或小鼠肝在AAV8-FIX转导之后的FIX mRNA水平。

AAV 构建体	体外肝细胞系(HepG2)		体内小鼠肝组织	
	归一化的 FIX 比率: mRNA/DNA	相较于 CS06-CRM8.0-ssV 的倍数增加	归一化的 FIX 比率: mRNA/DNA	相较于 CS06-CRM8.0-ssV 的倍数增加
CS06-CRM8.0-ssV	0.025	1	0.38	1
CS06-CRM8.1-ssV	0.14	5.4	0.78	2.0
CS06-CRM8.3-ssV	0.58	23.4	1.09	2.8

[0441] 实施例6的方法

[0442] 包括RNA和DNA提取的定量实时聚合酶链反应。通过标准程序从冷冻肝(参见动物实验)或HepG2细胞(参见在HepG2细胞中的体外生物效能测定)提取基因组DNA和总RNA。对于体内实验的分析,每个处理组选择在第14天接近于各自组的平均FIX活性(在平均值±SD内)的三只动物的子集。使用寡核苷酸(dT20)引物、SuperScript III逆转录酶(RT)和用DNA酶处理的总RNA根据手册(DNeasy血液和组织试剂盒,Qiagen,Germany;RNeasy微型试剂盒,Qiagen)合成cDNA。

[0443] 通过基于荧光的定量实时聚合酶链反应(qPCR)扩增FIX外显子6的96bp序列来测定gDNA和cDNA样品两者中的FIX转基因拷贝数。鼠β-肌动蛋白充当内源性对照物并且使用市售TaqMan测定来定量。使用特定装置的软件进行qPCR数据分析,基于标准曲线的线性回归参数计算每一反应的FIX或β-肌动蛋白拷贝数。另外,针对1μg的RNA或DNA对结果进行归一化并且计算mRNA:DNA比率。

[0444] 应了解,本文所描述的实施例和实施方案仅用于说明性目的,并且本领域技术人员将会想到有鉴于其的各种修改或变化,而且这些修改或变化应包括在本申请的精神和范围内和包括在随附权利要求书的范畴内。本文引用的所有公布、专利和专利申请出于所有目的以全文引用的方式并入本文中。

- [0001] 序列表
- [0002] <110> 百深公司(Baxalta Incorporated)
- [0003] 百深有限责任公司(Baxalta GmbH)
- [0004] <120> 用于血友病B的基因疗法的编码具有增加的表达的重组FIX的病毒载体
- [0005] <130> 008073-5117-W0
- [0006] <150> US 62/509,616
- [0007] <151> 2017-05-22
- [0008] <160> 64
- [0009] <170> PatentIn version 3.5
- [0010] <210> 1
- [0011] <211> 1386
- [0012] <212> DNA
- [0013] <213> 人工序列(Artificial Sequence)
- [0014] <220>
- [0015] <223> FIX-FL-NA
- [0016] <400> 1
- [0017] atgcagcgcg tgaacatgat catggcagaa tcaccaggcc tcatcaccat ctgcctttta 60
- [0018] ggatatctac tcagtgtga atgtacagtt tttcttgatc atgaaaacgc caacaaaatt 120
- [0019] ctgaatcggc caaagaggta taattcaggt aaattggaag agtttgttca agggaacctt 180
- [0020] gagagagaat gtatggaaga aaagtgtagt ttggaagaag cacgagaagt ttttgaaaac 240
- [0021] actgaaagaa caactgaatt ttggaagcag tatgttgatg gagatcagtg tgagtccaat 300
- [0022] ccatgtttaa atggcggcag ttgcaaggat gacattaatt cctatgaatg ttggtgtccc 360
- [0023] tttggatttg aaggaaagaa ctgtgaatta gatgtaacat gtaacattaa gaatggcaga 420
- [0024] tgcgagcagt tttgtaaaa tagtgctgat aacaagggtg tttgctcctg tactgaggga 480
- [0025] tatcgacttg cagaaaacca gaagtcctgt gaaccagcag tgccatttcc atgtggaaga 540
- [0026] gtttctgttt cacaaacttc taagctcacc cgtgctgaga ctgtttttcc tgatgtggac 600
- [0027] tatgtaaatt ctactgaagc tgaaccatt ttggataaca tcaactaaag cacccaatca 660
- [0028] tttaatgact tcaactcgggt tggttggtga gaagatgcca aaccaggta attcccttgg 720
- [0029] caggttggtt tgaatggtaa agttgatgca ttctgtggag gctctatcgt taatgaaaaa 780
- [0030] tggattgtaa ctgctgccc ctgtgttgaa actggtgtta aaattacagt tgtegcaggt 840
- [0031] gaacataata ttgaggagac agaacataca gagcaaaagc gaaatgtgat tcgaattatt 900
- [0032] cctcaccaca actacaatgc agctattaat aagtacaacc atgacattgc ctttctggaa 960
- [0033] ctggacgaac ctttagtgct aaacagctac gttacaccta tttgcattgc tgacaaggaa 1020
- [0034] tacacgaaca tcttcctcaa atttggatct ggctatgtaa gtggctgggg aagagtcttc 1080
- [0035] cacaaaggga gatcagcttt agttcttcag taccttagag ttccacttgt tgaccgagcc 1140
- [0036] acatgtcttc gatctacaaa gtccaccatc tataacaaca tgttctgtgc tggcttccat 1200
- [0037] gaaggaggta gagattcatg tcaaggagat agtgggggac cccatgttac tgaagtggaa 1260
- [0038] gggaccagtt tcttaactgg aattattagc tggggtgaag agtgtgcaat gaaaggcaaa 1320
- [0039] tatggaatat ataccaaggt atcccgttat gtcaactgga ttaaggaaaa aacaaagctc 1380
- [0040] acttaa 1386
- [0041] <210> 2

[0042]	<211>	461	
[0043]	<212>	PRT	
[0044]	<213>	人工序列(Artificial Sequence)	
[0045]	<220>		
[0046]	<223>	FIX-FL-AA	
[0047]	<400>	2	
[0048]	Met	Gln	Arg Val Asn Met Ile Met Ala Glu Ser Pro Gly Leu Ile Thr
[0049]	1	5	10 15
[0050]	Ile	Cys	Leu Leu Gly Tyr Leu Leu Ser Ala Glu Cys Thr Val Phe Leu
[0051]		20	25 30
[0052]	Asp	His	Glu Asn Ala Asn Lys Ile Leu Asn Arg Pro Lys Arg Tyr Asn
[0053]		35	40 45
[0054]	Ser	Gly	Lys Leu Glu Glu Phe Val Gln Gly Asn Leu Glu Arg Glu Cys
[0055]		50	55 60
[0056]	Met	Glu	Glu Lys Cys Ser Phe Glu Glu Ala Arg Glu Val Phe Glu Asn
[0057]	65	70	75 80
[0058]	Thr	Glu	Arg Thr Thr Glu Phe Trp Lys Gln Tyr Val Asp Gly Asp Gln
[0059]		85	90 95
[0060]	Cys	Glu	Ser Asn Pro Cys Leu Asn Gly Gly Ser Cys Lys Asp Asp Ile
[0061]		100	105 110
[0062]	Asn	Ser	Tyr Glu Cys Trp Cys Pro Phe Gly Phe Glu Gly Lys Asn Cys
[0063]		115	120 125
[0064]	Glu	Leu	Asp Val Thr Cys Asn Ile Lys Asn Gly Arg Cys Glu Gln Phe
[0065]		130	135 140
[0066]	Cys	Lys	Asn Ser Ala Asp Asn Lys Val Val Cys Ser Cys Thr Glu Gly
[0067]	145	150	155 160
[0068]	Tyr	Arg	Leu Ala Glu Asn Gln Lys Ser Cys Glu Pro Ala Val Pro Phe
[0069]		165	170 175
[0070]	Pro	Cys	Gly Arg Val Ser Val Ser Gln Thr Ser Lys Leu Thr Arg Ala
[0071]		180	185 190
[0072]	Glu	Thr	Val Phe Pro Asp Val Asp Tyr Val Asn Ser Thr Glu Ala Glu
[0073]		195	200 205
[0074]	Thr	Ile	Leu Asp Asn Ile Thr Gln Ser Thr Gln Ser Phe Asn Asp Phe
[0075]		210	215 220
[0076]	Thr	Arg	Val Val Gly Gly Glu Asp Ala Lys Pro Gly Gln Phe Pro Trp
[0077]	225	230	235 240
[0078]	Gln	Val	Val Leu Asn Gly Lys Val Asp Ala Phe Cys Gly Gly Ser Ile
[0079]		245	250 255
[0080]	Val	Asn	Glu Lys Trp Ile Val Thr Ala Ala His Cys Val Glu Thr Gly
[0081]		260	265 270
[0082]	Val	Lys	Ile Thr Val Val Ala Gly Glu His Asn Ile Glu Glu Thr Glu
[0083]		275	280 285

[0084]	His Thr Glu Gln Lys Arg Asn Val Ile Arg Ile Ile Pro His His Asn
[0085]	290 295 300
[0086]	Tyr Asn Ala Ala Ile Asn Lys Tyr Asn His Asp Ile Ala Leu Leu Glu
[0087]	305 310 315 320
[0088]	Leu Asp Glu Pro Leu Val Leu Asn Ser Tyr Val Thr Pro Ile Cys Ile
[0089]	325 330 335
[0090]	Ala Asp Lys Glu Tyr Thr Asn Ile Phe Leu Lys Phe Gly Ser Gly Tyr
[0091]	340 345 350
[0092]	Val Ser Gly Trp Gly Arg Val Phe His Lys Gly Arg Ser Ala Leu Val
[0093]	355 360 365
[0094]	Leu Gln Tyr Leu Arg Val Pro Leu Val Asp Arg Ala Thr Cys Leu Arg
[0095]	370 375 380
[0096]	Ser Thr Lys Phe Thr Ile Tyr Asn Asn Met Phe Cys Ala Gly Phe His
[0097]	385 390 395 400
[0098]	Glu Gly Gly Arg Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro His Val
[0099]	405 410 415
[0100]	Thr Glu Val Glu Gly Thr Ser Phe Leu Thr Gly Ile Ile Ser Trp Gly
[0101]	420 425 430
[0102]	Glu Glu Cys Ala Met Lys Gly Lys Tyr Gly Ile Tyr Thr Lys Val Ser
[0103]	435 440 445
[0104]	Arg Tyr Val Asn Trp Ile Lys Glu Lys Thr Lys Leu Thr
[0105]	450 455 460
[0106]	<210> 3
[0107]	<211> 423
[0108]	<212> PRT
[0109]	<213> 人工序列(Artificial Sequence)
[0110]	<220>
[0111]	<223> FIX2-FL-AA
[0112]	<400> 3
[0113]	Met Gln Arg Val Asn Met Ile Met Ala Glu Ser Pro Gly Leu Ile Thr
[0114]	1 5 10 15
[0115]	Ile Cys Leu Leu Gly Tyr Leu Leu Ser Ala Glu Cys Thr Val Phe Leu
[0116]	20 25 30
[0117]	Asp His Glu Asn Ala Asn Lys Ile Leu Asn Arg Pro Lys Arg Tyr Asn
[0118]	35 40 45
[0119]	Ser Gly Lys Leu Glu Glu Phe Val Gln Gly Asn Leu Glu Arg Glu Cys
[0120]	50 55 60
[0121]	Met Glu Glu Lys Cys Ser Phe Glu Glu Ala Arg Glu Val Phe Glu Asn
[0122]	65 70 75 80
[0123]	Thr Glu Arg Thr Thr Glu Phe Trp Lys Gln Tyr Val Asp Val Thr Cys
[0124]	85 90 95
[0125]	Asn Ile Lys Asn Gly Arg Cys Glu Gln Phe Cys Lys Asn Ser Ala Asp

[0126]	100	105	110
[0127]	Asn Lys Val Val Cys Ser Cys Thr Glu Gly Tyr Arg Leu Ala Glu Asn		
[0128]	115	120	125
[0129]	Gln Lys Ser Cys Glu Pro Ala Val Pro Phe Pro Cys Gly Arg Val Ser		
[0130]	130	135	140
[0131]	Val Ser Gln Thr Ser Lys Leu Thr Arg Ala Glu Thr Val Phe Pro Asp		
[0132]	145	150	155
[0133]	Val Asp Tyr Val Asn Ser Thr Glu Ala Glu Thr Ile Leu Asp Asn Ile		
[0134]	165	170	175
[0135]	Thr Gln Ser Thr Gln Ser Phe Asn Asp Phe Thr Arg Val Val Gly Gly		
[0136]	180	185	190
[0137]	Glu Asp Ala Lys Pro Gly Gln Phe Pro Trp Gln Val Val Leu Asn Gly		
[0138]	195	200	205
[0139]	Lys Val Asp Ala Phe Cys Gly Gly Ser Ile Val Asn Glu Lys Trp Ile		
[0140]	210	215	220
[0141]	Val Thr Ala Ala His Cys Val Glu Thr Gly Val Lys Ile Thr Val Val		
[0142]	225	230	235
[0143]	Ala Gly Glu His Asn Ile Glu Glu Thr Glu His Thr Glu Gln Lys Arg		
[0144]	245	250	255
[0145]	Asn Val Ile Arg Ile Ile Pro His His Asn Tyr Asn Ala Ala Ile Asn		
[0146]	260	265	270
[0147]	Lys Tyr Asn His Asp Ile Ala Leu Leu Glu Leu Asp Glu Pro Leu Val		
[0148]	275	280	285
[0149]	Leu Asn Ser Tyr Val Thr Pro Ile Cys Ile Ala Asp Lys Glu Tyr Thr		
[0150]	290	295	300
[0151]	Asn Ile Phe Leu Lys Phe Gly Ser Gly Tyr Val Ser Gly Trp Gly Arg		
[0152]	305	310	315
[0153]	Val Phe His Lys Gly Arg Ser Ala Leu Val Leu Gln Tyr Leu Arg Val		
[0154]	325	330	335
[0155]	Pro Leu Val Asp Arg Ala Thr Cys Leu Arg Ser Thr Lys Phe Thr Ile		
[0156]	340	345	350
[0157]	Tyr Asn Asn Met Phe Cys Ala Gly Phe His Glu Gly Gly Arg Asp Ser		
[0158]	355	360	365
[0159]	Cys Gln Gly Asp Ser Gly Gly Pro His Val Thr Glu Val Glu Gly Thr		
[0160]	370	375	380
[0161]	Ser Phe Leu Thr Gly Ile Ile Ser Trp Gly Glu Glu Cys Ala Met Lys		
[0162]	385	390	395
[0163]	Gly Lys Tyr Gly Ile Tyr Thr Lys Val Ser Arg Tyr Val Asn Trp Ile		
[0164]	405	410	415
[0165]	Lys Glu Lys Thr Lys Leu Thr		
[0166]	420		
[0167]	<210> 4		

95

[0210]	His Thr Glu Gln Lys Arg Asn Val Ile Arg Ile Ile Pro His His Asn		
[0211]	290	295	300
[0212]	Tyr Asn Ala Ala Ile Asn Lys Tyr Asn His Asp Ile Ala Leu Leu Glu		
[0213]	305	310	315 320
[0214]	Leu Asp Glu Pro Leu Val Leu Asn Ser Tyr Val Thr Pro Ile Cys Ile		
[0215]	325	330	335
[0216]	Ala Asp Lys Glu Tyr Thr Asn Ile Phe Leu Lys Phe Gly Ser Gly Tyr		
[0217]	340	345	350
[0218]	Val Ser Gly Trp Gly Arg Val Phe His Lys Gly Arg Ser Ala Leu Val		
[0219]	355	360	365
[0220]	Leu Gln Tyr Leu Arg Val Pro Leu Val Asp Arg Ala Thr Cys Leu Leu		
[0221]	370	375	380
[0222]	Ser Thr Lys Phe Thr Ile Tyr Asn Asn Met Phe Cys Ala Gly Phe His		
[0223]	385	390	395 400
[0224]	Glu Gly Gly Arg Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro His Val		
[0225]	405	410	415
[0226]	Thr Glu Val Glu Gly Thr Ser Phe Leu Thr Gly Ile Ile Ser Trp Gly		
[0227]	420	425	430
[0228]	Glu Glu Cys Ala Met Lys Gly Lys Tyr Gly Ile Tyr Thr Lys Val Ser		
[0229]	435	440	445
[0230]	Arg Tyr Val Asn Trp Ile Lys Glu Lys Thr Lys Leu Thr		
[0231]	450	455	460
[0232]	<210> 5		
[0233]	<211> 1385		
[0234]	<212> DNA		
[0235]	<213> 人工序列(Artificial Sequence)		
[0236]	<220>		
[0237]	<223> CS02-FL-NA		
[0238]	<400> 5		
[0239]	atgcagaggg tgaacatgat catggctgag agccctggcc tgatcaccat ctgcctgctg	60	
[0240]	ggctacctgc tgtcagcaga gtgcacagtg ttcttgacc atgagaatgc caacaagatc	120	
[0241]	ctgaacaggc ccaagagata caactcaggc aagctggagg agtttgtgca gggcaacctg	180	
[0242]	gagagggagt gcatggagga gaagtgcagc tttagaggagg ccagagaggt gtttgagaac	240	
[0243]	acagagagga ccacagagtt ctggaagcag tatgtggatg gagaccagtg tgagagcaac	300	
[0244]	ccttgctga atggaggcag ctgcaaggat gacatcaaca gctatgagtg ctggtgcct	360	
[0245]	tttgctttg agggcaagaa ctgtgagctg gatgtgacct gcaacatcaa gaatggcagg	420	
[0246]	tgtgagcagt tctgcaagaa ctgagctgac aacaaagtgg tgtgtagctg cacagagggc	480	
[0247]	tacagactgg ctgagaacca gaagagctgt gaggctgctg tgcccttccc ctgtggcaga	540	
[0248]	gtgtcagtgt cccagaccag caagctgacc agagctgaga cagtgttccc tgatgtggac	600	
[0249]	tatgtgaata gcacagaggc tgagaccatc ctggacaaca tcaccagag caccagctcc	660	
[0250]	ttcaatgact tcaccagagt tgtgggagga gaggatgcca agcctggcca gttcccctgg	720	
[0251]	caggtggtgc tgaatggcaa agtggatgcc ttctgtggag gcagcattgt gaatgagaag	780	

[0252]	tggattgtga cagctgcccc ctgtgtggag acaggagtga agatcacagt ggtggctgga	840
[0253]	gaacacaata ttgaggagac agagcacaca gagcagaaga ggaatgtcat caggattatc	900
[0254]	ccccaccaca actacaatgc tgccatcaac aagtacaacc atgacattgc cctgctggag	960
[0255]	ctggatgagc ctctggtgct gaatagctat gtgaccccc tctgcattgc tgacaaggag	1020
[0256]	tacaccaaca tcttcctgaa gtttggctca ggctatgtgt caggctgggg cagagtgttc	1080
[0257]	cacaagggca gatcagccct ggtgctgcag tacctgagag tgcccctggt ggacagagcc	1140
[0258]	acctgcctgt tgagcaccaa gttcaccatc tacaacaaca tgttctgtgc tggcttccat	1200
[0259]	gagggaggca gagacagctg ccaggagagac tcaggaggac cccatgtgac agaagtggag	1260
[0260]	ggcaccagct tcctgacagg catcatcagc tggggagagg agtgtgcat gaagggaag	1320
[0261]	tatggcatct acaccaaagt gagcagatat gtgaactgga tcaaggagaa aaccaagctg	1380
[0262]	acctg	1385
[0263]	<210>	6
[0264]	<211>	1386
[0265]	<212>	DNA
[0266]	<213>	人工序列(Artificial Sequence)
[0267]	<220>	
[0268]	<223>	CS03-FL-NA
[0269]	<400>	6
[0270]	atgcagaggg tgaacatgat catggctgag agccctggcc tgatcaccat ctgcctgctg	60
[0271]	ggctacctgc tgtctgctga gtgactgtg ttcttgacc atgagaatgc caacaagatc	120
[0272]	ctgaacaggc ccaagagata caactctggc aagctggagg agtttgtgca gggcaacctg	180
[0273]	gagagggagt gcatggagga gaagtgcagc tttgaggagg ccagggaagt gtttgagaac	240
[0274]	actgagagga ccactgagtt ctggaagcag tatgtggatg gggaccagtg tgagagcaac	300
[0275]	ccttgctga atgggggcag ctgcaaggat gacatcaaca gctatgagt ctggtgccct	360
[0276]	tttgctttg agggcaagaa ctgtgactg gatgtgacct gcaacatcaa gaatggcagg	420
[0277]	tgtgagcagt tctgcaagaa ctctgctgac aacaaagtgg tgtgtagctg cactgagggc	480
[0278]	tacagactgg ctgagaacca gaagagctgt gagcctgctg tgcccttccc ctgtggcaga	540
[0279]	gtgtctgtgt cccagaccag caagctgacc agagctgaga ctgtgttccc tgatgtggac	600
[0280]	tatgtgaata gcaactgaggc tgagaccatc ctggacaaca taccagag caccagtc	660
[0281]	ttcaatgact tcaccagagt ggtggggggg gaggatgcca agcctggcca gttcccttg	720
[0282]	caggtggtgc tgaatggcaa agtggatgcc ttctgtgggg gcagcattgt gaatgagaag	780
[0283]	tggattgtga ctgctgcccc ctgtgtggag actgggtga agatcactgt ggtggctggg	840
[0284]	gaacacaata ttgaggagac tgagcacact gagcagaaga ggaatgtcat caggattatc	900
[0285]	ccccaccaca actacaatgc tgccatcaac aagtacaacc atgacattgc cctgctggag	960
[0286]	ctggatgagc ctctggtgct gaatagctat gtgaccccc tctgcattgc tgacaaggag	1020
[0287]	tacaccaaca tcttcctgaa gtttggctct ggctatgtgt ctggctgggg cagagtgttc	1080
[0288]	cacaagggca ggtctgccct ggtgctgcag tacctgagag tgcccctggt ggacagagcc	1140
[0289]	acctgcctgc tgagcaccaa gttcaccatc tacaacaaca tgttctgtgc tggcttccat	1200
[0290]	gaggggggca gagacagctg ccagggggac tctgggggcc cccatgtgac tgaagtggag	1260
[0291]	ggcaccagct tcctgactgg catcatcagc tggggggagg agtgtgcat gaagggaag	1320
[0292]	tatggcatct acaccaaagt gagcaggtat gtgaactgga tcaaggagaa aaccaagctg	1380
[0293]	acctga	1386

[0294]	<210> 7
[0295]	<211> 1386
[0296]	<212> DNA
[0297]	<213> 人工序列(Artificial Sequence)
[0298]	<220>
[0299]	<223> CS04-FL-NA
[0300]	<400> 7
[0301]	atgcagaggg tgaacatgat tatggctgag agccctggcc tgatcaccat ctgcctgctg 60
[0302]	ggctacctgc tgtctgctga gtgcacagtgt ttcttgacc atgagaatgc caacaagatc 120
[0303]	ctgaacaggc ccaagagata caactctggc aagctggagg agtttgtgca gggcaacctg 180
[0304]	gagagggagt gcatggagga gaagtgcagc tttagaggagg ccaggagggt gtttgagaac 240
[0305]	acagagagga ccacagagtt ctggaagcag tatgtggatg gtgaccagtg tgagagcaac 300
[0306]	ccttgctga atggaggcag ctgcaaggat gacatcaaca gctatgagt ctggtgccct 360
[0307]	tttgctttg agggcaagaa ctgtgagctg gatgtgacct gcaacatcaa gaatggcagg 420
[0308]	tgtgagcagt tctgcaagaa ctctgtgac aacaagggtg tgtgtagctg cacagagggc 480
[0309]	tacagactgg ctgagaacca gaagagctgt gacgtgctg tgcccttccc ctgtggcaga 540
[0310]	gtgtctgtgt ccagaccag caagctgacc agagctgaga cagtgttccc tgatgtggac 600
[0311]	tatgtgaaca gcacagaggc tgagaccatc ctggacaaca taccagag caccagtc 660
[0312]	ttcaatgact taccagagt ggtgggagga gaggatgcca agcctggcca gttccctgg 720
[0313]	caggtgggtgc tgaatggcaa ggtggatgcc ttctgtggag gcagcattgt gaatgagaag 780
[0314]	tggattgtga cagctgcca ctgtgtggag acaggagtga agatcacagt ggtggctgga 840
[0315]	gagcacaaca ttgaggagac agagcacaca gagcagaaga ggaatgtgat caggatcatc 900
[0316]	cctcaccaca actacaatgc tgccatcaac aagtacaacc atgacattgc cctgctggag 960
[0317]	ctggatgagc ctctggtgct gaacagctat gtgacccta tctgcattgc tgacaaggag 1020
[0318]	tacaccaaca tcttctgaa gtttggtct ggctatgtgt ctggctgggg cagagtgttc 1080
[0319]	cacaagggca ggtctgccct ggtgctgcag tacctgagag tgccctggt ggacagagcc 1140
[0320]	acctgcctgt tgagaccaa gttcaccatc tacaacaaca tgttctgtgc tggcttccat 1200
[0321]	gagggaggca gagacagctg ccagggtgac tctggaggac cccatgtgac agaggtggag 1260
[0322]	ggcaccagct tctgacagg catcatcagc tggggagagg agtgtgcat gaagggaag 1320
[0323]	tatggcatct acacaaagt gagcagatat gtgaactgga tcaaggagaa gaccaagctg 1380
[0324]	acctga 1386
[0325]	<210> 8
[0326]	<211> 1386
[0327]	<212> DNA
[0328]	<213> 人工序列(Artificial Sequence)
[0329]	<220>
[0330]	<223> CS05-FL-NA
[0331]	<400> 8
[0332]	atgcagaggg tgaacatgat tatggctgag agccctggcc tgatcaccat ctgcctgctg 60
[0333]	ggctacctgc tgtctgctga gtgcactgtg ttcttgacc atgagaatgc caacaagatc 120
[0334]	ctgaaccgcc ccaagcgcta caactctggc aagctggagg agtttgtgca gggcaacctg 180
[0335]	gagagggagt gcatggagga gaagtgcagc tttagaggagg ccaggagggt gtttgagaac 240

[0336]	actgagcgca ccactgagtt ctggaagcag tatgtggatg gggaccagtg tgagagcaac	300
[0337]	ccctgcctga atggggggag ctgcaaggat gacatcaaca gctatgagtg ctggtgcccc	360
[0338]	tttgctttg agggcaagaa ctgtgagctg gatgtgacct gcaacatcaa gaatggccgc	420
[0339]	tgtgagcagt tctgcaagaa ctctgctgac aacaaggtgg tgtgctcttg cactgagggc	480
[0340]	taccgcctgg ctgagaacca gaagagctgt gaggctgctg tgcccttccc ctgtggcagg	540
[0341]	gtgtctgtga gccagaccag caagctgacc agggctgaga ctgtgttccc tgacgtggac	600
[0342]	tatgtgaaca gactgagge tgagaccatc ctggacaaca tcaccagag caccagagc	660
[0343]	ttcaatgact tcaccaggt ggtgggagga gaggatgcca agcctggcca gtccccctgg	720
[0344]	caggtggtgc tgaatggcaa ggtggatgcc ttctgtggag gcagcattgt gaatgagaag	780
[0345]	tggattgtga ccgctccca ctgtgtggag actggagtga agatcactgt ggtggctggg	840
[0346]	gagcacaaca ttgaggagac agagcacaca gacagaagc gcaatgtgat caggatcatc	900
[0347]	ccccaccaca actacaatgc tgccatcaac aagtacaacc atgacattgc cctgctggag	960
[0348]	ctggatgagc ccctggtgct gaacagctac gtgacccca tctgcattgc agacaaggag	1020
[0349]	tacaccaaca tcttctgaa gtttggctct ggctatgtgt ctggctgggg cagggtgttc	1080
[0350]	cacaagggca ggtctgccct ggtgctgcag tacctgaggg tgcccctggt ggacagggcc	1140
[0351]	acctgcctgc tgagaccaa gtccaccatc tacaacaaca tgttctgcgc tggcttccat	1200
[0352]	gagggaggaa gggacagctg ccaggagac tctggaggcc cccatgtgac agaggtggag	1260
[0353]	ggcaccagct tcctgacagg catcatcagc tggggggagg agtgtgccat gaagggcaag	1320
[0354]	tatggcatct acaccaaagt gtcccgtat gtgaactgga tcaaggagaa gaccaagctg	1380
[0355]	acctga	1386
[0356]	<210>	9
[0357]	<211>	1386
[0358]	<212>	DNA
[0359]	<213>	人工序列(Artificial Sequence)
[0360]	<220>	
[0361]	<223>	CS06-FL-NA
[0362]	<400>	9
[0363]	atgcagaggg tcaacatgat catggctgag tcccctggcc tcatcaccat ctgcctgctg	60
[0364]	ggctacctgc tgtctgctga gtgcactgtc ttcttgacc atgagaatgc caacaagatc	120
[0365]	ctcaacaggc ccaagagata caactctggc aaactggagg agtttgtcca gggcaacctg	180
[0366]	gagagggagt gcatggagga gaagtgtcc tttgaggagg ccagggaggt ctttgagaac	240
[0367]	actgagcgca ccactgagtt ctggaacag tatgtggatg gggaccagtg tgagtccaac	300
[0368]	ccctgcctga atgggggag ctgcaaggat gacatcaaca gctatgagtg ctggtgcccc	360
[0369]	tttgctttg agggcaagaa ctgtgagctg gatgtgacct gcaacatcaa gaatggcaga	420
[0370]	tgtgagcagt tctgcaagaa ctctgctgac aacaaggtgg tgtgctcctg cactgagggc	480
[0371]	taccgcctgg ctgagaacca gaagagctgt gaggctgctg tgccattccc atgtggcaga	540
[0372]	gtctctgtga gccagaccag caagctcacc agggctgaga ctgtgttccc tgatgtggac	600
[0373]	tatgtgaaca gactgagge tgaaaccatc ctggacaaca tcaccagag caccagagc	660
[0374]	ttcaatgact tcaccaggt ggtgggagga gaggatgcca agcctggcca gtccccctgg	720
[0375]	caagtgggtgc tcaatggcaa ggtggatgcc ttctgtgggg gctccattgt gaatgagaag	780
[0376]	tggattgtca ctgctccca ctgtgtggag actgggtca agatcactgt ggtggctggg	840
[0377]	gagcacaaca ttgaggagac tgagcacact gacagaagc gcaatgtgat caggatcatc	900

[0378] cccaccaca actacaatgc tgccatcaac aagtacaacc atgacattgc cctgctggag 960
 [0379] ctggatgagc ccctggtcct caacagctat gtgaccccca tctgcattgc tgacaaggag 1020
 [0380] tacaccaaca tcttcctcaa gtttggctct ggctatgtct ctggctgggg cagagtgttc 1080
 [0381] cacaaaggca ggtctgccct ggtgctccag tacctgagag tgcccctggt ggacagggcc 1140
 [0382] acctgcctct tgagacacaa gtccaccatc tacaacaaca tgttctgtgc tggcttccat 1200
 [0383] gagggaggaa gagacagctg ccagggggac tctggaggac cccatgtcac tgaggtggag 1260
 [0384] ggcacctcct tctcactgg catcatctcc tggggagagg agtgtgcat gaaaggcaaa 1320
 [0385] tatggcatct acaccaaagt ctccagatat gtcaactgga tcaaggagaa gaccaagctg 1380
 [0386] acctga 1386
 [0387] <210> 10
 [0388] <211> 415
 [0389] <212> PRT
 [0390] <213> 人工序列(Artificial Sequence)
 [0391] <220>
 [0392] <223> FIX-MP-AA
 [0393] <400> 10
 [0394] Tyr Asn Ser Gly Lys Leu Glu Glu Phe Val Gln Gly Asn Leu Glu Arg
 [0395] 1 5 10 15
 [0396] Glu Cys Met Glu Glu Lys Cys Ser Phe Glu Glu Ala Arg Glu Val Phe
 [0397] 20 25 30
 [0398] Glu Asn Thr Glu Arg Thr Thr Glu Phe Trp Lys Gln Tyr Val Asp Gly
 [0399] 35 40 45
 [0400] Asp Gln Cys Glu Ser Asn Pro Cys Leu Asn Gly Gly Ser Cys Lys Asp
 [0401] 50 55 60
 [0402] Asp Ile Asn Ser Tyr Glu Cys Trp Cys Pro Phe Gly Phe Glu Gly Lys
 [0403] 65 70 75 80
 [0404] Asn Cys Glu Leu Asp Val Thr Cys Asn Ile Lys Asn Gly Arg Cys Glu
 [0405] 85 90 95
 [0406] Gln Phe Cys Lys Asn Ser Ala Asp Asn Lys Val Val Cys Ser Cys Thr
 [0407] 100 105 110
 [0408] Glu Gly Tyr Arg Leu Ala Glu Asn Gln Lys Ser Cys Glu Pro Ala Val
 [0409] 115 120 125
 [0410] Pro Phe Pro Cys Gly Arg Val Ser Val Ser Gln Thr Ser Lys Leu Thr
 [0411] 130 135 140
 [0412] Arg Ala Glu Thr Val Phe Pro Asp Val Asp Tyr Val Asn Ser Thr Glu
 [0413] 145 150 155 160
 [0414] Ala Glu Thr Ile Leu Asp Asn Ile Thr Gln Ser Thr Gln Ser Phe Asn
 [0415] 165 170 175
 [0416] Asp Phe Thr Arg Val Val Gly Gly Glu Asp Ala Lys Pro Gly Gln Phe
 [0417] 180 185 190
 [0418] Pro Trp Gln Val Val Leu Asn Gly Lys Val Asp Ala Phe Cys Gly Gly
 [0419] 195 200 205

[0420]	Ser Ile Val Asn Glu Lys Trp Ile Val Thr Ala Ala His Cys Val Glu		
[0421]	210	215	220
[0422]	Thr Gly Val Lys Ile Thr Val Val Ala Gly Glu His Asn Ile Glu Glu		
[0423]	225	230	235 240
[0424]	Thr Glu His Thr Glu Gln Lys Arg Asn Val Ile Arg Ile Ile Pro His		
[0425]	245	250	255
[0426]	His Asn Tyr Asn Ala Ala Ile Asn Lys Tyr Asn His Asp Ile Ala Leu		
[0427]	260	265	270
[0428]	Leu Glu Leu Asp Glu Pro Leu Val Leu Asn Ser Tyr Val Thr Pro Ile		
[0429]	275	280	285
[0430]	Cys Ile Ala Asp Lys Glu Tyr Thr Asn Ile Phe Leu Lys Phe Gly Ser		
[0431]	290	295	300
[0432]	Gly Tyr Val Ser Gly Trp Gly Arg Val Phe His Lys Gly Arg Ser Ala		
[0433]	305	310	315 320
[0434]	Leu Val Leu Gln Tyr Leu Arg Val Pro Leu Val Asp Arg Ala Thr Cys		
[0435]	325	330	335
[0436]	Leu Arg Ser Thr Lys Phe Thr Ile Tyr Asn Asn Met Phe Cys Ala Gly		
[0437]	340	345	350
[0438]	Phe His Glu Gly Gly Arg Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro		
[0439]	355	360	365
[0440]	His Val Thr Glu Val Glu Gly Thr Ser Phe Leu Thr Gly Ile Ile Ser		
[0441]	370	375	380
[0442]	Trp Gly Glu Glu Cys Ala Met Lys Gly Lys Tyr Gly Ile Tyr Thr Lys		
[0443]	385	390	395 400
[0444]	Val Ser Arg Tyr Val Asn Trp Ile Lys Glu Lys Thr Lys Leu Thr		
[0445]	405	410	415
[0446]	<210> 11		
[0447]	<211> 377		
[0448]	<212> PRT		
[0449]	<213> 人工序列(Artificial Sequence)		
[0450]	<220>		
[0451]	<223> FIX2-MP-AA		
[0452]	<400> 11		
[0453]	Tyr Asn Ser Gly Lys Leu Glu Glu Phe Val Gln Gly Asn Leu Glu Arg		
[0454]	1	5	10 15
[0455]	Glu Cys Met Glu Glu Lys Cys Ser Phe Glu Glu Ala Arg Glu Val Phe		
[0456]	20	25	30
[0457]	Glu Asn Thr Glu Arg Thr Thr Glu Phe Trp Lys Gln Tyr Val Asp Val		
[0458]	35	40	45
[0459]	Thr Cys Asn Ile Lys Asn Gly Arg Cys Glu Gln Phe Cys Lys Asn Ser		
[0460]	50	55	60
[0461]	Ala Asp Asn Lys Val Val Cys Ser Cys Thr Glu Gly Tyr Arg Leu Ala		

[0462]	65	70	75	80
[0463]	Glu Asn Gln Lys Ser Cys Glu Pro Ala Val Pro Phe Pro Cys Gly Arg			
[0464]		85	90	95
[0465]	Val Ser Val Ser Gln Thr Ser Lys Leu Thr Arg Ala Glu Thr Val Phe			
[0466]		100	105	110
[0467]	Pro Asp Val Asp Tyr Val Asn Ser Thr Glu Ala Glu Thr Ile Leu Asp			
[0468]		115	120	125
[0469]	Asn Ile Thr Gln Ser Thr Gln Ser Phe Asn Asp Phe Thr Arg Val Val			
[0470]		130	135	140
[0471]	Gly Gly Glu Asp Ala Lys Pro Gly Gln Phe Pro Trp Gln Val Val Leu			
[0472]	145	150	155	160
[0473]	Asn Gly Lys Val Asp Ala Phe Cys Gly Gly Ser Ile Val Asn Glu Lys			
[0474]		165	170	175
[0475]	Trp Ile Val Thr Ala Ala His Cys Val Glu Thr Gly Val Lys Ile Thr			
[0476]		180	185	190
[0477]	Val Val Ala Gly Glu His Asn Ile Glu Glu Thr Glu His Thr Glu Gln			
[0478]		195	200	205
[0479]	Lys Arg Asn Val Ile Arg Ile Ile Pro His His Asn Tyr Asn Ala Ala			
[0480]		210	215	220
[0481]	Ile Asn Lys Tyr Asn His Asp Ile Ala Leu Leu Glu Leu Asp Glu Pro			
[0482]	225	230	235	240
[0483]	Leu Val Leu Asn Ser Tyr Val Thr Pro Ile Cys Ile Ala Asp Lys Glu			
[0484]		245	250	255
[0485]	Tyr Thr Asn Ile Phe Leu Lys Phe Gly Ser Gly Tyr Val Ser Gly Trp			
[0486]		260	265	270
[0487]	Gly Arg Val Phe His Lys Gly Arg Ser Ala Leu Val Leu Gln Tyr Leu			
[0488]		275	280	285
[0489]	Arg Val Pro Leu Val Asp Arg Ala Thr Cys Leu Arg Ser Thr Lys Phe			
[0490]		290	295	300
[0491]	Thr Ile Tyr Asn Asn Met Phe Cys Ala Gly Phe His Glu Gly Gly Arg			
[0492]	305	310	315	320
[0493]	Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro His Val Thr Glu Val Glu			
[0494]		325	330	335
[0495]	Gly Thr Ser Phe Leu Thr Gly Ile Ile Ser Trp Gly Glu Glu Cys Ala			
[0496]		340	345	350
[0497]	Met Lys Gly Lys Tyr Gly Ile Tyr Thr Lys Val Ser Arg Tyr Val Asn			
[0498]		355	360	365
[0499]	Trp Ile Lys Glu Lys Thr Lys Leu Thr			
[0500]		370	375	
[0501]	<210> 12			
[0502]	<211> 415			
[0503]	<212> PRT			

[0504]	<213> 人工序列(Artificial Sequence)																		
[0505]	<220>																		
[0506]	<223> FIXp-MP-AA																		
[0507]	<400> 12																		
[0508]	Tyr	Asn	Ser	Gly	Lys	Leu	Glu	Glu	Phe	Val	Gln	Gly	Asn	Leu	Glu	Arg			
[0509]	1				5					10					15				
[0510]	Glu	Cys	Met	Glu	Glu	Lys	Cys	Ser	Phe	Glu	Glu	Ala	Arg	Glu	Val	Phe			
[0511]				20					25					30					
[0512]	Glu	Asn	Thr	Glu	Arg	Thr	Thr	Glu	Phe	Trp	Lys	Gln	Tyr	Val	Asp	Gly			
[0513]			35					40					45						
[0514]	Asp	Gln	Cys	Glu	Ser	Asn	Pro	Cys	Leu	Asn	Gly	Gly	Ser	Cys	Lys	Asp			
[0515]	50						55					60							
[0516]	Asp	Ile	Asn	Ser	Tyr	Glu	Cys	Trp	Cys	Pro	Phe	Gly	Phe	Glu	Gly	Lys			
[0517]	65				70					75				80					
[0518]	Asn	Cys	Glu	Leu	Asp	Val	Thr	Cys	Asn	Ile	Lys	Asn	Gly	Arg	Cys	Glu			
[0519]				85					90				95						
[0520]	Gln	Phe	Cys	Lys	Asn	Ser	Ala	Asp	Asn	Lys	Val	Val	Cys	Ser	Cys	Thr			
[0521]				100					105				110						
[0522]	Glu	Gly	Tyr	Arg	Leu	Ala	Glu	Asn	Gln	Lys	Ser	Cys	Glu	Pro	Ala	Val			
[0523]			115					120					125						
[0524]	Pro	Phe	Pro	Cys	Gly	Arg	Val	Ser	Val	Ser	Gln	Thr	Ser	Lys	Leu	Thr			
[0525]	130						135					140							
[0526]	Arg	Ala	Glu	Thr	Val	Phe	Pro	Asp	Val	Asp	Tyr	Val	Asn	Ser	Thr	Glu			
[0527]	145				150					155				160					
[0528]	Ala	Glu	Thr	Ile	Leu	Asp	Asn	Ile	Thr	Gln	Ser	Thr	Gln	Ser	Phe	Asn			
[0529]				165					170				175						
[0530]	Asp	Phe	Thr	Arg	Val	Val	Gly	Gly	Glu	Asp	Ala	Lys	Pro	Gly	Gln	Phe			
[0531]			180						185				190						
[0532]	Pro	Trp	Gln	Val	Val	Leu	Asn	Gly	Lys	Val	Asp	Ala	Phe	Cys	Gly	Gly			
[0533]			195					200					205						
[0534]	Ser	Ile	Val	Asn	Glu	Lys	Trp	Ile	Val	Thr	Ala	Ala	His	Cys	Val	Glu			
[0535]	210					215						220							
[0536]	Thr	Gly	Val	Lys	Ile	Thr	Val	Val	Ala	Gly	Glu	His	Asn	Ile	Glu	Glu			
[0537]	225				230					235				240					
[0538]	Thr	Glu	His	Thr	Glu	Gln	Lys	Arg	Asn	Val	Ile	Arg	Ile	Ile	Pro	His			
[0539]				245					250				255						
[0540]	His	Asn	Tyr	Asn	Ala	Ala	Ile	Asn	Lys	Tyr	Asn	His	Asp	Ile	Ala	Leu			
[0541]			260						265				270						
[0542]	Leu	Glu	Leu	Asp	Glu	Pro	Leu	Val	Leu	Asn	Ser	Tyr	Val	Thr	Pro	Ile			
[0543]			275						280				285						
[0544]	Cys	Ile	Ala	Asp	Lys	Glu	Tyr	Thr	Asn	Ile	Phe	Leu	Lys	Phe	Gly	Ser			
[0545]	290					295						300							

[0546]	Gly Tyr Val Ser Gly Trp Gly Arg Val Phe His Lys Gly Arg Ser Ala	
[0547]	305	310 315 320
[0548]	Leu Val Leu Gln Tyr Leu Arg Val Pro Leu Val Asp Arg Ala Thr Cys	
[0549]	325	330 335
[0550]	Leu Leu Ser Thr Lys Phe Thr Ile Tyr Asn Asn Met Phe Cys Ala Gly	
[0551]	340	345 350
[0552]	Phe His Glu Gly Gly Arg Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro	
[0553]	355	360 365
[0554]	His Val Thr Glu Val Glu Gly Thr Ser Phe Leu Thr Gly Ile Ile Ser	
[0555]	370	375 380
[0556]	Trp Gly Glu Glu Cys Ala Met Lys Gly Lys Tyr Gly Ile Tyr Thr Lys	
[0557]	385	390 395 400
[0558]	Val Ser Arg Tyr Val Asn Trp Ile Lys Glu Lys Thr Lys Leu Thr	
[0559]	405	410 415
[0560]	<210> 13	
[0561]	<211> 1248	
[0562]	<212> DNA	
[0563]	<213> 人工序列(Artificial Sequence)	
[0564]	<220>	
[0565]	<223> CS02-MP-NA	
[0566]	<400> 13	
[0567]	tacaactcag gcaagctgga ggagtttgtg cagggcaacc tggagaggga gtgcatggag	60
[0568]	gagaagtgca gctttgagga ggccagagag gtgtttgaga acacagagag gaccacagag	120
[0569]	ttctggaagc agtatgtgga tggagaccag tgtgagagca acccttgccct gaatggaggc	180
[0570]	agctgcaagg atgacatcaa cagctatgag tgctggtgcc cttttggctt tgagggcaag	240
[0571]	aactgtgagc tggatgtgac ctgcaacatc aagaatggca ggtgtgagca gttctgcaag	300
[0572]	aactcagctg acaacaaagt ggtgtgtagc tgcacagagg gctacagact ggctgagaac	360
[0573]	cagaagagct gtgagcctgc tgtgcccttc ccctgtggca gagtgtcagt gtcccagacc	420
[0574]	agcaagctga ccagagctga gacagtgttc cctgatgtgg actatgtgaa tagcacagag	480
[0575]	gctgagacca tcttgacaa catcaccagc agcaccagc ctttcaatga cttcaccaga	540
[0576]	gtttgtgggag gagaggatgc caagcctggc cagttcccct ggcagggtgt gctgaatggc	600
[0577]	aaagtggatg cttctgtgtg aggcagcatt gtgaatgaga agtggattgt gacagctgcc	660
[0578]	cactgtgtgg agacaggagt gaagatcaca gtgggtggctg gagaacacaa tattgaggag	720
[0579]	acagagcaca cagagcagaa gaggaatgtc atcaggatta tccccacca caactacaat	780
[0580]	gctgccatca acaagtacaa ccatgacatt gccctgctgg agctggatga gcctctggtg	840
[0581]	ctgaatagct atgtgacccc catctgcatt gctgacaagg agtacacaa catcttcctg	900
[0582]	aagtttggtc caggctatgt gtcaggctgg ggcagagtgt tccacaaggg cagatcagcc	960
[0583]	ctggtgctgc agtacctgag agtgcccctg gtggacagag ccacctgcct gttgagcacc	1020
[0584]	aagttcacca tctacaacaa catgtttctgt gctggtctcc atgaggaggc cagagacagc	1080
[0585]	tgccaggagg actcaggagg acccatgtg acagaagtgg agggcaccag cttcctgaca	1140
[0586]	ggcatcatca gctggggaga ggagtgtgcc atgaagggca agtatggcat ctacacaaaa	1200
[0587]	gtgagcagat atgtgaactg gatcaaggag aaaaccaagc tgacctga	1248

[0588]	<210> 14
[0589]	<211> 1248
[0590]	<212> DNA
[0591]	<213> 人工序列(Artificial Sequence)
[0592]	<220>
[0593]	<223> CS03-MP-NA
[0594]	<400> 14
[0595]	tacaactctg gcaagctgga ggagtttgtg cagggcaacc tggagaggga gtgcatggag 60
[0596]	gagaagtgca gctttgagga ggccaggga gtgtttgaga aacttgagag gaccactgag 120
[0597]	ttctggaagc agtatgtgga tggggaccag tgtgagagca acccttgcct gaatgggggc 180
[0598]	agctgcaagg atgacatcaa cagctatgag tgctgggtgcc cttttggctt tgagggcaag 240
[0599]	aactgtgagc tggatgtgac ctgcaacatc aagaatggca ggtgtgagca gttctgcaag 300
[0600]	aactctgctg acaacaaagt ggtgtgtagc tgcactgagg gctacagact ggctgagaac 360
[0601]	cagaagagct gtgagcctgc tgtgcccttc ccctgtggca gagtgtctgt gtcccagacc 420
[0602]	agcaagctga ccagagctga gactgtgttc cctgatgtgg actatgtgaa tagcactgag 480
[0603]	gctgagacca tcctggacaa catcaccagc agcaccagc cttcaatga cttaccaga 540
[0604]	gtggtggggg gggaggatgc caagcctggc cagttcccct ggcagggtgt gctgaatggc 600
[0605]	aaagtggatg cttctgtgtg gggcagcatt gtgaatgaga agtggattgt gactgctgcc 660
[0606]	cactgtgtgg agactggggt gaagatcact gtggtggctg gggaacacaa tattgaggag 720
[0607]	actgagcaca ctgagcagaa gaggaatgtc atcaggatta tccccacca caactacaat 780
[0608]	gctgccatca acaagtacaa ccatgacatt gccctgctgg agctggatga gcctctggtg 840
[0609]	ctgaatagct atgtgacccc catctgcatt gctgacaagg agtacaccaa catcttcctg 900
[0610]	aagtttggct ctggctatgt gtctggctgg ggcagagtgt tccacaaggg caggtctgcc 960
[0611]	ctggtgctgc agtacctgag agtgcccctg gtggacagag ccacctgcct gctgagcacc 1020
[0612]	aagttcacca tctacaacaa catgttctgt gctggcttcc atgagggggg cagagacagc 1080
[0613]	tgccaggggg actctggggg ccccatgtg actgaagtgg agggcaccag cttcctgact 1140
[0614]	ggcatcatca gctgggggga ggagtgtgcc atgaaggga agtatggcat ctacacaaa 1200
[0615]	gtgagcaggt atgtgaactg gatcaaggag aaaaccaagc tgacctga 1248
[0616]	<210> 15
[0617]	<211> 1250
[0618]	<212> DNA
[0619]	<213> 人工序列(Artificial Sequence)
[0620]	<220>
[0621]	<223> CS04-MP-NA
[0622]	<400> 15
[0623]	cstacaactc tggcaagctg gaggagtgtg tgcagggcaa cctggagagg gagtgcattg 60
[0624]	aggagaagtg cagctttgag gaggccaggg aggtgtttga gaacacagag aggaccacag 120
[0625]	agttctggaa gcagtatgtg gatggtgacc agtgtgagag caacccttgc ctgaatggag 180
[0626]	gcagctgcaa ggatgacatc aacagctatg agtgctggtg cccttttggc tttagaggca 240
[0627]	agaactgtga gctggatgtg acctgcaaca tcaagaatgg caggtgtgag cagttctgca 300
[0628]	agaactctgc tgacaacaag gtggtgtgta gctgcacaga gggctacaga ctggctgaga 360
[0629]	accagaagag ctgtgagcct gctgtgccct tcccctgtgg cagagtgtct gtgtcccaga 420

[0630]	ccagcaagct gaccagagct gagacagtgt tccctgatgt ggactatgtg aacagcacag	480
[0631]	aggetgagac catcctggac aacatcaccc agagcaccca gtccttcaat gacttcacca	540
[0632]	gagtgggtggg aggagaggat gccaaagcctg gccagttccc ctggcagggtg gtgctgaatg	600
[0633]	gcaaggtgga tgccttctgt ggaggcagca ttgtgaatga gaagtggatt gtgacagctg	660
[0634]	cccactgtgt ggagacagga gtgaagatca cagtgggtggc tggagagcac aacattgagg	720
[0635]	agacagagca cacagagcag aagaggaatg tgatcaggat catccctcac cacaactaca	780
[0636]	atgtgccat caacaagtac aacctgaca ttgccctgct ggagctggat gagectctgg	840
[0637]	tgctgaacag ctatgtgacc cctatctgca ttgctgacaa ggagtacacc aacatcttcc	900
[0638]	tgaagtttgg ctctggctat gtgtctggct ggggcagagt gttccacaag ggcaggtctg	960
[0639]	ccctggtgct gcagtacctg agagtgtccc tgggtggacag agccacctgc ctgttgagca	1020
[0640]	ccaagttcac catctacaac aacatgttct gtgctggctt ccatgaggga ggcagagaca	1080
[0641]	gctgccaggg tgactctgga ggaccccatg tgacagaggt ggagggcacc agcttctctga	1140
[0642]	caggcatcat cagctgggga gaggagtgtg ccatgaaggg caagtatggc atctacacca	1200
[0643]	aagtgagcag atatgtgaac tggatcaagg agaagaccaa gctgacctga	1250
[0644]	<210> 16	
[0645]	<211> 1248	
[0646]	<212> DNA	
[0647]	<213> 人工序列(Artificial Sequence)	
[0648]	<220>	
[0649]	<223> CS05-MP-NA	
[0650]	<400> 16	
[0651]	tacaactctg gcaagctgga ggagtttgtg cagggaacc tggagaggga gtgcatggag	60
[0652]	gagaagtgca gctttgagga ggccaggag gtgtttgaga acactgagcg caccactgag	120
[0653]	ttctggaagc agtatgtgga tggggaccag tgtgagagca acccctgcct gaatgggggg	180
[0654]	agctgcaagg atgacatcaa cagctatgag tgctgggtgcc cctttggctt tgagggcaag	240
[0655]	aactgtgagc tggatgtgac ctgcaacatc aagaatggcc gctgtgagca gttctgcaag	300
[0656]	aactctgctg acaacaaggt ggtgtgctct tgcactgagg gctaccgcct ggctgagaac	360
[0657]	cagaagagct gtgagcctgc tgtgcccttc ccctgtggca ggggtgtctgt gagccagacc	420
[0658]	agcaagctga ccagggtga gactgtgttc cctgacgtgg actatgtgaa cagcactgag	480
[0659]	gctgagacca tcttgacaa catcacccag agcacccaga gcttcaatga cttcaccagg	540
[0660]	gtggtgggag gagaggatgc caagcctggc cagttcccct ggcaggtggt gctgaatggc	600
[0661]	aaggtggatg cttctgtgtg aggcagcatt gtgaatgaga agtggattgt gaccgctgcc	660
[0662]	cactgtgtgg agactggagt gaagatcact gtggtggctg gggagcaca cattgaggag	720
[0663]	acagagcaca cagagcagaa gcgcaatgtg atcaggatca tccccacca caactacaat	780
[0664]	gctgccatca acaagtacaa ccatgacatt gccctgctgg agctggatga gccctgggtg	840
[0665]	ctgaacagct acgtgacccc catctgcatt gcagacaagg agtacaccaa catcttcctg	900
[0666]	aagtttggct ctggctatgt gtctggctgg ggcagggtgt tccacaaggg caggtctgcc	960
[0667]	ctggtgctgc agtacctgag ggtgccctg gtggacagg ccacctgcct gctgagcacc	1020
[0668]	aagttcacca tctacaacaa catgttctgc gctggcttcc atgaggagg aaggacagc	1080
[0669]	tgccaggag actctggagg ccccatgtg acagaggtgg agggcaccag cttcctgaca	1140
[0670]	ggcatcatca gctgggggga ggagtgtgcc atgaaggga agtatggcat ctacaccaa	1200
[0671]	gtgtcccgt atgtgaactg gatcaaggag aagaccaagc tgacctga	1248

[0672]	<210> 17
[0673]	<211> 1248
[0674]	<212> DNA
[0675]	<213> 人工序列(Artificial Sequence)
[0676]	<220>
[0677]	<223> CS06-MP-NA
[0678]	<400> 17
[0679]	tacaactctg gcaaactgga ggagtttgtc cagggcaacc tggagaggga gtgcatggag 60
[0680]	gagaagtgct cctttgagga ggccaggag gtctttgaga aacttgagcg caccactgag 120
[0681]	ttctggaaac agtatgtgga tggggaccag tgtgagtcca acccctgcct gaatgggggc 180
[0682]	agctgcaagg atgacatcaa cagctatgag tgctgggtgcc cctttggctt tgagggcaag 240
[0683]	aactgtgagc tggatgtgac ctgcaacatc aagaatggca gatgtgagca gttctgcaag 300
[0684]	aactctgctg acaacaaggt ggtgtgtctc tgcactgagg gctaccgcct ggctgagaac 360
[0685]	cagaagagct gtgagcctgc tgtgccattc ccatgtggca gagtctctgt gagccagacc 420
[0686]	agcaagctca ccagggtga gactgtgttc cctgatgtgg actatgtgaa cagcactgag 480
[0687]	gctgaaacca tcctggacaa catcaccag agcaccaga gcttcaatga cttaccaga 540
[0688]	gtggtgggag gagaggatgc caagcctggc cagttcccct ggcaagtgtt gctcaatggc 600
[0689]	aaggtggatg cttctgtgg gggctccatt gtgaatgaga agtggattgt cactgtgtcc 660
[0690]	cactgtgtgg agactgggt caagatcact gtggtggctg gggagcaca cattgaggag 720
[0691]	actgagcaca ctgagcagaa gcgcaatgtg atcaggatca tccccacca caactacaat 780
[0692]	gctgccatca acaagtacaa ccatgacatt gccctgctgg agctggatga gccctggtc 840
[0693]	ctcaacagct atgtgacccc catctgcatt gctgacaagg agtacaccaa catcttcctc 900
[0694]	aagtttgct ctggctatgt ctctggctgg ggcagagtgt tccacaaagg caggtctgcc 960
[0695]	ctggtgtctc agtacctgag agtgccctg gtggacagg ccacctgcct cttgagcacc 1020
[0696]	aagttcacca tctacaacaa catgttctgt gctggcttc atgaggagg aagagacagc 1080
[0697]	tgccagggg actctggagg acccatgtc actgaggtgg agggcacctc cttctcact 1140
[0698]	ggcatcatct cctggggaga ggagtgtgcc atgaaaggca aatatggcat ctacaccaa 1200
[0699]	gtctccagat atgtcaactg gatcaaggag aagaccaagc tgacctga 1248
[0700]	<210> 18
[0701]	<211> 138
[0702]	<212> DNA
[0703]	<213> 人工序列(Artificial Sequence)
[0704]	<220>
[0705]	<223> FIX-PPP-NA
[0706]	<400> 18
[0707]	atgcagcgcg tgaacatgat catggcagaa tcaccaggcc tcatcaccat ctgcctttta 60
[0708]	ggatatctac tcaagtctga atgtacagtt tttcttgatc atgaaaacgc caacaaaatt 120
[0709]	ctgaatcggc caagagg 138
[0710]	<210> 19
[0711]	<211> 138
[0712]	<212> DNA
[0713]	<213> 人工序列(Artificial Sequence)

[0714] <220>
[0715] <223> CS02-PPP-NA
[0716] <400> 19
[0717] atgcagaggg tgaacatgat catggctgag agccctggcc tgatcaccat ctgcctgctg 60
[0718] ggctacctgc tgtcagcaga gtgcacagtg ttcttgacc atgagaatgc caacaagatc 120
[0719] ctgaacaggc ccaagaga 138
[0720] <210> 20
[0721] <211> 138
[0722] <212> DNA
[0723] <213> 人工序列(Artificial Sequence)
[0724] <220>
[0725] <223> CS03-PPP-NA
[0726] <400> 20
[0727] atgcagaggg tgaacatgat catggctgag agccctggcc tgatcaccat ctgcctgctg 60
[0728] ggctacctgc tgtctgctga gtgcactgtg ttcttgacc atgagaatgc caacaagatc 120
[0729] ctgaacaggc ccaagaga 138
[0730] <210> 21
[0731] <211> 138
[0732] <212> DNA
[0733] <213> 人工序列(Artificial Sequence)
[0734] <220>
[0735] <223> CS04-PPP-NA
[0736] <400> 21
[0737] atgcagaggg tgaacatgat tatggctgag agccctggcc tgatcaccat ctgcctgctg 60
[0738] ggctacctgc tgtctgctga gtgcacagtg ttcttgacc atgagaatgc caacaagatc 120
[0739] ctgaacaggc ccaagaga 138
[0740] <210> 22
[0741] <211> 138
[0742] <212> DNA
[0743] <213> 人工序列(Artificial Sequence)
[0744] <220>
[0745] <223> CS05-PPP-NA
[0746] <400> 22
[0747] atgcagaggg tgaacatgat tatggctgag agccctggcc tgatcaccat ctgcctgctg 60
[0748] ggctacctgc tgtctgctga gtgcactgtg ttcttgacc atgagaatgc caacaagatc 120
[0749] ctgaaccgcc ccaagcgc 138
[0750] <210> 23
[0751] <211> 138
[0752] <212> DNA
[0753] <213> 人工序列(Artificial Sequence)
[0754] <220>
[0755] <223> CS06-PPP-NA

[0756] <400> 23
[0757] atgcagaggg tcaacatgat catggctgag tcccctggcc tcatcaccat ctgcctgctg 60
[0758] ggctacctgc tgtctgctga gtgcactgtc ttcttgacc atgagaatgc caacaagatc 120
[0759] ctcaacaggc ccaagaga 138
[0760] <210> 24
[0761] <211> 84
[0762] <212> DNA
[0763] <213> 人工序列(Artificial Sequence)
[0764] <220>
[0765] <223> FIX-SP-NA
[0766] <400> 24
[0767] atgcagcgcg tgaacatgat catggcagaa tcaccaggcc tcatcaccat ctgcctttta 60
[0768] ggatatctac tcagtctga atgt 84
[0769] <210> 25
[0770] <211> 84
[0771] <212> DNA
[0772] <213> 人工序列(Artificial Sequence)
[0773] <220>
[0774] <223> CS02-SP-NA
[0775] <400> 25
[0776] atgcagaggg tgaacatgat catggctgag agccctggcc tgatcaccat ctgcctgctg 60
[0777] ggctacctgc tgcagcaga gtgc 84
[0778] <210> 26
[0779] <211> 84
[0780] <212> DNA
[0781] <213> 人工序列(Artificial Sequence)
[0782] <220>
[0783] <223> CS03-SP-NA
[0784] <400> 26
[0785] atgcagaggg tgaacatgat catggctgag agccctggcc tgatcaccat ctgcctgctg 60
[0786] ggctacctgc tgtctgctga gtgc 84
[0787] <210> 27
[0788] <211> 84
[0789] <212> DNA
[0790] <213> 人工序列(Artificial Sequence)
[0791] <220>
[0792] <223> CS04-SP-NA
[0793] <400> 27
[0794] atgcagaggg tgaacatgat tatggctgag agccctggcc tgatcaccat ctgcctgctg 60
[0795] ggctacctgc tgtctgctga gtgc 84
[0796] <210> 28
[0797] <211> 84

- [0798] <212> DNA
[0799] <213> 人工序列(Artificial Sequence)
[0800] <220>
[0801] <223> CS05-SP-NA
[0802] <400> 28
[0803] atgcagaggg tgaacatgat tatggctgag agccctggcc tgatcaccat ctgcctgctg 60
[0804] ggctacctgc tgtctgctga gtgc 84
[0805] <210> 29
[0806] <211> 84
[0807] <212> DNA
[0808] <213> 人工序列(Artificial Sequence)
[0809] <220>
[0810] <223> CS06--SP-NA
[0811] <400> 29
[0812] atgcagaggg tcaacatgat catggctgag tccctggcc tcatcaccat ctgcctgctg 60
[0813] ggctacctgc tgtctgctga gtgc 84
[0814] <210> 30
[0815] <211> 54
[0816] <212> DNA
[0817] <213> 人工序列(Artificial Sequence)
[0818] <220>
[0819] <223> FIX-PP-NA
[0820] <400> 30
[0821] acagtttttc ttgatcatga aaacgccaac aaaattctga atcgccaaa gagg 54
[0822] <210> 31
[0823] <211> 54
[0824] <212> DNA
[0825] <213> 人工序列(Artificial Sequence)
[0826] <220>
[0827] <223> CS02-PP-NA
[0828] <400> 31
[0829] acagtgttcc tggaccatga gaatgccaac aagatcctga acaggcccaa gaga 54
[0830] <210> 32
[0831] <211> 54
[0832] <212> DNA
[0833] <213> 人工序列(Artificial Sequence)
[0834] <220>
[0835] <223> CS03-PP-NA
[0836] <400> 32
[0837] actgtgttcc tggaccatga gaatgccaac aagatcctga acaggcccaa gaga 54
[0838] <210> 33
[0839] <211> 54

[0840]	<212> DNA
[0841]	<213> 人工序列(Artificial Sequence)
[0842]	<220>
[0843]	<223> CS04-PP-NA
[0844]	<400> 33
[0845]	acagtgttcc tggaccatga gaatgccaac aagatcctga acaggcccaa gaga 54
[0846]	<210> 34
[0847]	<211> 54
[0848]	<212> DNA
[0849]	<213> 人工序列(Artificial Sequence)
[0850]	<220>
[0851]	<223> CS05-PP-NA
[0852]	<400> 34
[0853]	actgtgttcc tggaccatga gaatgccaac aagatcctga accgccccaa gcgc 54
[0854]	<210> 35
[0855]	<211> 54
[0856]	<212> DNA
[0857]	<213> 人工序列(Artificial Sequence)
[0858]	<220>
[0859]	<223> CS06-PP-NA
[0860]	<400> 35
[0861]	actgtcttcc tggaccatga gaatgccaac aagatcctca acaggcccaa gaga 54
[0862]	<210> 36
[0863]	<211> 46
[0864]	<212> PRT
[0865]	<213> 人工序列(Artificial Sequence)
[0866]	<220>
[0867]	<223> FIX-PPP-AA
[0868]	<400> 36
[0869]	Met Gln Arg Val Asn Met Ile Met Ala Glu Ser Pro Gly Leu Ile Thr
[0870]	1 5 10 15
[0871]	Ile Cys Leu Leu Gly Tyr Leu Leu Ser Ala Glu Cys Thr Val Phe Leu
[0872]	20 25 30
[0873]	Asp His Glu Asn Ala Asn Lys Ile Leu Asn Arg Pro Lys Arg
[0874]	35 40 45
[0875]	<210> 37
[0876]	<211> 28
[0877]	<212> PRT
[0878]	<213> 人工序列(Artificial Sequence)
[0879]	<220>
[0880]	<223> FIX-SP-AA
[0881]	<400> 37

[0882]	Met Gln Arg Val Asn Met Ile Met Ala Glu Ser Pro Gly Leu Ile Thr
[0883]	1 5 10 15
[0884]	Ile Cys Leu Leu Gly Tyr Leu Leu Ser Ala Glu Cys
[0885]	20 25
[0886]	<210> 38
[0887]	<211> 18
[0888]	<212> PRT
[0889]	<213> 人工序列(Artificial Sequence)
[0890]	<220>
[0891]	<223> FIX-PP-AA
[0892]	<400> 38
[0893]	Thr Val Phe Leu Asp His Glu Asn Ala Asn Lys Ile Leu Asn Arg Pro
[0894]	1 5 10 15
[0895]	Lys Arg
[0896]	<210> 39
[0897]	<211> 72
[0898]	<212> DNA
[0899]	<213> 人工序列(Artificial Sequence)
[0900]	<220>
[0901]	<223> CRM8
[0902]	<400> 39
[0903]	gggggaggct gctggtgaat attaaccaag gtcacccag ttatcggagg agcaaacagg 60
[0904]	ggctaagtcc ac 72
[0905]	<210> 40
[0906]	<211> 5276
[0907]	<212> DNA
[0908]	<213> 人工序列(Artificial Sequence)
[0909]	<220>
[0910]	<223> CS06-CRM8.3-ssV
[0911]	<400> 40
[0912]	ttggccactc cctctctgcg cgctcgctcg ctactgagg ccgggcgacc aaaggtcgcc 60
[0913]	cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgag agagggagtg 120
[0914]	gccaaactcca tactagggg ttcttgagtt taaacttcgt cgacggggga ggctgctggt 180
[0915]	gaatattaac caaggtcacc ccagttatcg gaggagcaaa caggggctaa gtccaccggg 240
[0916]	ggaggctgct ggtgaatatt aaccaaggtc accccagtta tcggaggagc aaacaggggc 300
[0917]	taagtccacc gggggaggct gctggtgaat attaaccaag gtcacccag ttatcggagg 360
[0918]	agcaaacagg ggctaagtcc accgagggca ctgggaggat gttgagtaag atggaaaact 420
[0919]	actgatgacc cttgcagaga cagagtatta ggacatgttt gaacaggggc cgggcgatca 480
[0920]	gcaggtagct ctagaggatc cccgtctgtc tgcacatttc gtagagcgag tgttccgata 540
[0921]	ctctaatactc cctaggcaag gttcatatth gtgtaggtta cttattctcc ttttgttgac 600
[0922]	taagtcaata atcagaatca gcaggtttgg agtcagcttg gcagggatca gcagcctggg 660
[0923]	ttggaaggag ggggtataaa agccccctca ccaggagaag ccgtcacaca gactaggcgc 720

[0924]	gccctaaggt aagttggcgc cgtttaaggg atggttggtt ggtggggtat taatgtttaa	780
[0925]	ttaccttttt tacaggcctg aagatctgcc accatgcaga gggatcaacat gatcatggct	840
[0926]	gagtcacctg gcctcatcac catctgcctg ctgggctacc tgcctgtctgc tgagtgcact	900
[0927]	gtcttccttg accatgagaa tgccaacaag atcctcaaca ggcccaagag atacaactct	960
[0928]	ggcaaaactgg aggagtttgt ccagggaac ctggagaggg agtgcatgga ggagaagtgc	1020
[0929]	tcctttgagg aggccaggga ggtctttgag aacactgagc gcaccactga gttctggaaa	1080
[0930]	cagtatgtgg atggggacca gtgtgagtc aacctctgcc tgaatggggg cagctgcaag	1140
[0931]	gatgacatca acagctatga gtgctggtgc ccttttggtt ttgagggcaa gaactgtgag	1200
[0932]	ctggatgtga cctgcaacat caagaatggc agatgtgagc agttctgcaa gaactctgct	1260
[0933]	gacaacaagg tgggtgtctc ctgcatgag ggctaccgcc tggctgagaa ccagaagagc	1320
[0934]	tgtgagcctg ctgtgccatt cccatgtggc agagtctctg tgagccagac cagcaagctc	1380
[0935]	accagggtg agactgtgtt cctgtatgtg gactatgtga acagcactga ggctgaaacc	1440
[0936]	atcctggaca acatcaccca gagcaccag agcttcaatg acttcaccag agtggtggga	1500
[0937]	ggagaggatg ccaagcctgg ccagttcccc tggcaagtgg tgctcaatgg caaggtggat	1560
[0938]	gccttctgtg ggggtcccat tgtgaatgag aagtggattg tcatgctgc cactgtgtg	1620
[0939]	gagactggg tcaagatcac tgtgtgtggt ggggagcaca acattgagga gactgagcac	1680
[0940]	actgagcaga agcgcaatgt gatcaggatc atccccacc acaactacaa tgctgccatc	1740
[0941]	aacaagtaca accatgacat tgccctgctg gagctggatg agcccctggt cctcaacagc	1800
[0942]	tatgtgaccc ccatctgcat tgctgacaag gactacacca acatcttctt caagtttggc	1860
[0943]	tctggctatg tctctggctg gggcagagtg ttccacaaag gcaggtctgc cctggtgctc	1920
[0944]	cagtacctga gactgcccct ggtggacagg gccacctgcc tcttgagcac caagttcacc	1980
[0945]	atctacaaca acatgttctg tgctggcttc catgaggag gaagagacag ctgccagggg	2040
[0946]	gactctggag gaccccatgt cactgaggtg gagggcacct ctttctctac tggcatcatc	2100
[0947]	tcctggggag aggagtgtgc catgaaaggc aaatatggca tctacaccaa agtctccaga	2160
[0948]	tatgtcaact ggatcaagga gaagaccaag ctgacctgat gagcatgcct agagctcgtc	2220
[0949]	gatcagctc gactgtgcct tctagttgcc agccatctgt tgtttgcccc tccccctgc	2280
[0950]	cttccttgac cctggaaggt gccactccca ctgtccttc ctaataaaat gaggaaattg	2340
[0951]	catgcattg tctgagtagg tgtcattcta ttctgggggg tggggtgggg caggacagca	2400
[0952]	agggggagga ttgggaagac aatagcaggc atgctgggga attaatgaag ctgcgaagg	2460
[0953]	aacctctagt gatggagttg gccactccct ctctgcgcgc tcgctcgtc actgaggccg	2520
[0954]	ggcgacaaa ggtcgcccga cgccgggct ttgcccgggc ggcctcagtg agcgagcgag	2580
[0955]	cgcgagaga gggagtggcc aagacgattt aaatgacaag cttggcgtaa tcatggtcat	2640
[0956]	agctgtttcc tgtgtgaaat tgttatccgc tcacaattcc acacaacata cgagccggaa	2700
[0957]	gcataaagtg taaagcctgg ggtgcctaata gactgagcta actcacatta attgcgttgc	2760
[0958]	gctcactgcc cgctttccag tcgggaaacc tgctgtgcca gctgcattaa tgaatcgcc	2820
[0959]	aacgcgcggg gagaggcggg ttgcgtattg ggcgtcttc cgcttcctcg ctactgact	2880
[0960]	cgctgcgtc ggtcgttcgg ctgcggcgag cggtatcagc tactcaaag gcgtaatac	2940
[0961]	ggttatccac agaatcaggg gataacgag gaaagaacat gtgagcaaaa ggccagcaaa	3000
[0962]	aggccaggaa ccgtaaaaag gccgcgttgc tggcgttttt ccataggctc cgccccctg	3060
[0963]	acgagcatca caaaaatcga cgctcaagtc agaggtggcg aaacccgaca ggactataaa	3120
[0964]	gataccaggc gtttccccct ggaagctccc tcgtgcgtc tcctgttccg accctgccgc	3180
[0965]	ttaccggata cctgtccgcc tttctccctt cgggaagcgt ggcgtttct catagctcac	3240

[0966]	gctgtaggta tctcagttcg gtgtaggtcg ttcgctccaa gctgggctgt gtgcacgaac	3300
[0967]	cccccgttca gcccgaaccgc tgcgcccttat ccggttaacta tcgtcttgag tccaacccgg	3360
[0968]	taagacacga cttatcgcca ctggcagcag ccactggtaa caggattagc agagcgaggt	3420
[0969]	atgtaggcgg tgctacagag ttcttgaagt ggtggcctaa ctacggctac actagaagaa	3480
[0970]	cagtatttgg tatctgcgct ctgctgaagc cagttacctt cggaaaaaga gttggtagct	3540
[0971]	cttgatccgg caaacaacc accgctggta gcggtggtt ttttgtttgc aagcagcaga	3600
[0972]	ttacgcgcag aaaaaagga tctcaagaag atcctttgat cttttctacg gggcttgacg	3660
[0973]	ctcagtggaa cgaaaactca cgtaagga ttttggcat gagattatca aaaaggatct	3720
[0974]	tcacctagat ctttttaaat taaaaatgaa gttttaaatc aatctaaagt atatatgagt	3780
[0975]	aaacttggtc tgacagttac caatgcttaa tcagtgggc acctatctca gcgatctgtc	3840
[0976]	tatttcgttc atccatagtt gcctgactcc ccgtcgtgta gataactacg atacgggagg	3900
[0977]	gcttaccatc tggccccagt gctgcaatga taccgcgaga cccacgctca ccggctccag	3960
[0978]	atttatcagc aataaaccag ccagccgga gggccgagc cagaagtggc cctgcaactt	4020
[0979]	tatccgcctc catccagtct attaatgtt gccgggaagc tagagtaagt agttcgccag	4080
[0980]	ttaatagttt gcgcaacgtt gttgccattg ctacaggcat cgtgggtgca cgctcgtcgt	4140
[0981]	ttggtagtgc ttcatcagc tccggttccc aacgatcaag gcgagttaca tgatccccc	4200
[0982]	tggtgtgcaa aaaagcgggt agctccttcg gtctccgat cgttgtcaga agtaagtgg	4260
[0983]	ccgcagtgtt atcactcatg gttatggcag cactgcataa ttctcttact gtcatgcat	4320
[0984]	ccgtaagatg cttttctgtg actggtgagt actcaaccaa gtcattctga gaatagtga	4380
[0985]	tgccgcgacc gagttgctct tgcccggcgt caatacggga taataccgcg ccacatagca	4440
[0986]	gaactttaa agtgctcatc attggaac gttcttcggg gcgaaaactc tcaaggatct	4500
[0987]	taccgctgtt gagatccagt tcgatgtaac ccactcgtgc acccaactga tcttcagcat	4560
[0988]	ctttacttt caccagcgtt tctgggtgag caaaaacagg aaggcaaat gccgcaaaa	4620
[0989]	agggataaag ggcgacacgg aaatgttgaa tactcatact cttcctttt caatattatt	4680
[0990]	gaagcattta tcagggttat tgtctcatga gcgatacat atttgaatgt atttagaaaa	4740
[0991]	ataaacaaat aggggttccg cgcacattc cccgaaaagt gccacctgac gtctaagaaa	4800
[0992]	ccattattat catgacatta acctataaaa ataggcgtat caccaggccc tttcgtctcg	4860
[0993]	cgcgtttcgg tgatgacggt gaaaacctct gacacatgca gctcccggag acggtcacag	4920
[0994]	cttgctgta agcggatgcc gggagcagac aagcccgta gggcgcgtca gcgggtgttg	4980
[0995]	gcgggtgtcg gggctggtt aactatgcgg catcagagca gattgtactg agagtgcacc	5040
[0996]	atatgcggtg tgaaataccg cacagatgcg taaggagaaa ataccgcatc aggcgccatt	5100
[0997]	cgccattcag gctgcgcaac tgttggaag ggcgatcgtt gcgggcctct tcgctattac	5160
[0998]	gccagctggc gaaaggggga tgtgctgcaa ggcgattaag ttgggtaacg ccagggtttt	5220
[0999]	cccagtcagc acgttgtaaa acgacggcca gtgaattcct cgagatttaa atgacg	5276
[1000]	<210> 41	
[1001]	<211> 702	
[1002]	<212> DNA	
[1003]	<213> 人工序列(Artificial Sequence)	
[1004]	<220>	
[1005]	<223> CS02-HC-NA	
[1006]	<400> 41	
[1007]	gttgtgggag gagaggatgc caagcctggc cagttcccct ggcaggtggt gctgaatggc	60

[1008]	aaagtggatg ctttctgtgg aggcagcatt gtgaatgaga agtggattgt gacagctgcc	120
[1009]	cactgtgtgg agacaggagt gaagatcaca gtggtggctg gagaacacaa tattgaggag	180
[1010]	acagagcaca cagagcagaa gaggaatgtc atcaggatta tccccacca caactacaat	240
[1011]	gctgccatca acaagtacaa ccatgacatt gccctgctgg agctggatga gcctctggtg	300
[1012]	ctgaatagct atgtgacccc catctgcatt gctgacaagg agtacaccaa catcttcctg	360
[1013]	aagtttgct caggctatgt gtcaggctgg ggcagagtgt tccacaaggg cagatcagcc	420
[1014]	ctggtgctgc agtacctgag agtgcccctg gtggacagag ccacctgcct gttgagcacc	480
[1015]	aagttcacca tctacaacaa catgtttctgt gctggcttcc atgagggagg cagagacagc	540
[1016]	tgccaggagg actcaggagg accccatgtg acagaagtgg agggcaccag cttcctgaca	600
[1017]	ggcatcatca gctggggaga ggagtgtgcc atgaaggga agtatggcat ctacaccaa	660
[1018]	gtgagcagat atgtgaactg gatcaaggag aaaaccaagc tg	702
[1019]	<210>	42
[1020]	<211>	435
[1021]	<212>	DNA
[1022]	<213>	人工序列(Artificial Sequence)
[1023]	<220>	
[1024]	<223>	CS02-LC-NA
[1025]	<400>	42
[1026]	tacaactcag gcaagctgga ggagtttgtg cagggaacc tggagaggga gtgcatggag	60
[1027]	gagaagtgc gctttgagga ggccagagag gtgtttgaga acacagagag gaccacagag	120
[1028]	ttctggaagc agtatgtgga tggagaccag tgtgagagca acccttgctt gaattggaggc	180
[1029]	agctgcaagg atgacatcaa cagctatgag tgctggtgcc cttttggctt tgagggaag	240
[1030]	aactgtgagc tggatgtgac ctgcaacatc aagaatggca ggtgtgagca gttctgcaag	300
[1031]	aactcagctg acaacaaagt ggtgtgtagc tgcacagagg gctacagact ggctgagaac	360
[1032]	cagaagagct gtgagcctgc tgtgcccttc cctgtggca gagtgtcagt gtcccagacc	420
[1033]	agcaagctga ccaga	435
[1034]	<210>	43
[1035]	<211>	705
[1036]	<212>	DNA
[1037]	<213>	人工序列(Artificial Sequence)
[1038]	<220>	
[1039]	<223>	CS03-HC-NA
[1040]	<400>	43
[1041]	gtggtggggg gggaggatgc caagcctggc cagttcccct ggcaggtggt gctgaatggc	60
[1042]	aaagtggatg ctttctgtgg ggcagcatt gtgaatgaga agtggattgt gactgctgcc	120
[1043]	cactgtgtgg agactgggt gaagatcact gtggtggctg gggaacacaa tattgaggag	180
[1044]	actgagcaca ctgagcagaa gaggaatgtc atcaggatta tccccacca caactacaat	240
[1045]	gctgccatca acaagtacaa ccatgacatt gccctgctgg agctggatga gcctctggtg	300
[1046]	ctgaatagct atgtgacccc catctgcatt gctgacaagg agtacaccaa catcttcctg	360
[1047]	aagtttgct ctggctatgt gtctggctgg ggcagagtgt tccacaaggg caggtctgcc	420
[1048]	ctggtgctgc agtacctgag agtgcccctg gtggacagag ccacctgcct gctgagcacc	480
[1049]	aagttcacca tctacaacaa catgtttctgt gctggcttcc atgagggggg cagagacagc	540

- [1050] tgccaggggg actctggggg ccccatgtg actgaagtgg agggcaccag cttctgact 600
- [1051] ggcatcatca gctgggggga ggagtgtgcc atgaaggga agtatggcat ctacacaaa 660
- [1052] gtgagcaggt atgtgaactg gatcaaggag aaaaccaagc tgacc 705
- [1053] <210> 44
- [1054] <211> 435
- [1055] <212> DNA
- [1056] <213> 人工序列(Artificial Sequence)
- [1057] <220>
- [1058] <223> CS03-LC-NA
- [1059] <400> 44
- [1060] tacaactctg gcaagctgga ggagtttgtg cagggcaacc tggagaggga gtgcatggag 60
- [1061] gagaagtgca gctttgagga ggccaggga gtgtttgaga aactgagag gaccactgag 120
- [1062] ttctggaagc agtatgtgga tggggaccag tgtgagagca acccttgctt gaatgggggc 180
- [1063] agctgcaagg atgacatcaa cagctatgag tgctggtgcc cttttggctt tgagggaag 240
- [1064] aactgtgagc tggatgtgac ctgcaacatc aagaatggca ggtgtgagca gttctgcaag 300
- [1065] aactctgtg acaacaaagt ggtgtgtagc tgcactgagg gctacagact ggctgagaac 360
- [1066] cagaagagct gtgagcctgc tgtgcccttc ccctgtggca gagtgtctgt gtcccagacc 420
- [1067] agcaagctga ccaga 435
- [1068] <210> 45
- [1069] <211> 705
- [1070] <212> DNA
- [1071] <213> 人工序列(Artificial Sequence)
- [1072] <220>
- [1073] <223> CS04-HC-NA
- [1074] <400> 45
- [1075] gtggtgggag gagaggatgc caagcctggc cagttcccct ggcaggtggt gctgaatggc 60
- [1076] aaggtggatg cttctgtgg aggcagcatt gtgaatgaga agtggattgt gacagctgcc 120
- [1077] cactgtgtgg agacaggagt gaagatcaca gtggtggctg gagagcaca cattgaggag 180
- [1078] acagagcaca cagagcagaa gaggaatgtg atcaggatca tccctacca caactacaat 240
- [1079] gctgccatca acaagtacaa ccatgacatt gccctgctgg agctggatga gcctctggtg 300
- [1080] ctgaacagct atgtgacccc tatctgcatt gctgacaagg agtacacaa catcttcctg 360
- [1081] aagtttgct ctggctatgt gtctggctgg ggcagagtgt tccacaagg caggtctgcc 420
- [1082] ctggtgctgc agtacctgag agtgcccctg gtggacagag ccacctgctt gttgagcacc 480
- [1083] aagttacca tctacaacaa catgttctgt gctggcttcc atgaggagg cagagacagc 540
- [1084] tgccagggtg actctggagg acccatgtg acagaggtgg agggcaccag cttctgaca 600
- [1085] ggcatcatca gctggggaga ggagtgtgcc atgaaggga agtatggcat ctacacaaa 660
- [1086] gtgagcagat atgtgaactg gatcaaggag aagaccaagc tgacc 705
- [1087] <210> 46
- [1088] <211> 435
- [1089] <212> DNA
- [1090] <213> 人工序列(Artificial Sequence)
- [1091] <220>

- [1092] <223> CS04-LC-NA
- [1093] <400> 46
- [1094] tacaactctg gcaagctgga ggagtttgtg cagggcaacc tggagaggga gtgcatggag 60
- [1095] gagaagtgca gctttgagga ggccaggag gtgtttgaga acacagagag gaccacagag 120
- [1096] ttctggaagc agtatgtgga tggtagaccag tgtgagagca acccttgccct gaatggaggc 180
- [1097] agctgcaagg atgacatcaa cagctatgag tgctggtgcc cttttggctt tgagggaag 240
- [1098] aactgtgagc tggatgtgac ctgcaacatc aagaatggca ggtgtgagca gttctgcaag 300
- [1099] aactctgctg acaacaaggt ggtgtgtagc tgcacagagg gctacagact ggctgagaac 360
- [1100] cagaagagct gtgagcctgc tgtgcccttc ccctgtggca gagtgtctgt gtcccagacc 420
- [1101] agcaagctga ccaga 435
- [1102] <210> 47
- [1103] <211> 705
- [1104] <212> DNA
- [1105] <213> 人工序列(Artificial Sequence)
- [1106] <220>
- [1107] <223> CS05-HC-NA
- [1108] <400> 47
- [1109] gtggtgggag gagaggatgc caagcctggc cagttcccct ggcagggtgt gctgaatggc 60
- [1110] aaggtggatg cttctgttg aggagcatt gtgaatgaga agtggattgt gaccgctgcc 120
- [1111] cactgtgtgg agactggagt gaagatcact gtggtggtg gggagcaca cattgaggag 180
- [1112] acagagcaca cagagcagaa gcgcaatgtg atcaggatca tccccacca caactacaat 240
- [1113] gctgccatca acaagtacaa ccatgacatt gccctgctgg agctggatga gccctggtg 300
- [1114] ctgaacagct acgtgacccc catctgcatt gcagacaagg agtacacaa catcttcctg 360
- [1115] aagtttgct ctggctatgt gtctggctgg ggcagggtgt tccacaagg caggtctgcc 420
- [1116] ctggtgctgc agtacctgag ggtgccctg gtggacagg ccacctgcct gctgagcacc 480
- [1117] aagttcacca tctacaacaa catgttctgc gctggcttcc atgaggagg aaggacagc 540
- [1118] tgccaggag actctggagg ccccatgtg acagaggtg agggcaccag ctctctgaca 600
- [1119] ggcatcatca gctgggggga ggagtgtgcc atgaaggca agtatggcat ctacacaaa 660
- [1120] gtgtcccgt atgtgaactg gatcaaggag aagaccaagc tgacc 705
- [1121] <210> 48
- [1122] <211> 435
- [1123] <212> DNA
- [1124] <213> 人工序列(Artificial Sequence)
- [1125] <220>
- [1126] <223> CS05-LC-NA
- [1127] <400> 48
- [1128] tacaactctg gcaagctgga ggagtttgtg cagggcaacc tggagaggga gtgcatggag 60
- [1129] gagaagtgca gctttgagga ggccaggag gtgtttgaga aactgagcg caccactgag 120
- [1130] ttctggaagc agtatgtgga tggggaccag tgtgagagca accctgcct gaatggggg 180
- [1131] agctgcaagg atgacatcaa cagctatgag tgctggtgcc cttttggctt tgagggaag 240
- [1132] aactgtgagc tggatgtgac ctgcaacatc aagaatggcc gctgtgagca gttctgcaag 300
- [1133] aactctgctg acaacaaggt ggtgtgtct tgcactgagg gctaccgcct ggctgagaac 360

[1134]	cagaagagct gtgagcctgc tgtgcccttc ccctgtggca ggggtgtctgt gagccagacc 420
[1135]	agcaagctga ccagg 435
[1136]	<210> 49
[1137]	<211> 705
[1138]	<212> DNA
[1139]	<213> 人工序列(Artificial Sequence)
[1140]	<220>
[1141]	<223> CS06-HC-NA
[1142]	<400> 49
[1143]	gtgggtgggag gagaggatgc caagcctggc cagttcccct ggcaagtggc gctcaatggc 60
[1144]	aagggtgatg ctttctgtgg gggctccatt gtgaatgaga agtggattgt cactgctgcc 120
[1145]	cactgtgtgg agactggggc caagatcact gtgggtggctg gggagcacia cattgaggag 180
[1146]	actgagcaca ctgagcagaa gcgcaatgtg atcaggatca tccccacca caactacaat 240
[1147]	gctgccatca acaagtacaa ccatgacatt gccctgctgg agctggatga gccctggtc 300
[1148]	ctcaacagct atgtgacccc catctgcatt gctgacaagg agtacaccaa catcttcctc 360
[1149]	aagtttgct ctggctatgt ctctggctgg ggcagagtgt tccacaaagg caggtctgcc 420
[1150]	ctggtgctcc agtacctgag agtgccctg gtggacaggg ccacctgcct cttgagcacc 480
[1151]	aagttcacca tctacaacaa catgttctgt gctggcttcc atgagggagg aagagacagc 540
[1152]	tgccaggggg actctggagg accccatgtc actgaggtgg agggcacctc cttcctcact 600
[1153]	ggcatcatct cctggggaga ggagtgtgcc atgaaaggca aatatggcat ctacacaaaa 660
[1154]	gtctccagat atgtcaactg gatcaaggag aagaccaagc tgacc 705
[1155]	<210> 50
[1156]	<211> 435
[1157]	<212> DNA
[1158]	<213> 人工序列(Artificial Sequence)
[1159]	<220>
[1160]	<223> CS06-LC-NA
[1161]	<400> 50
[1162]	tacaactctg gcaaaactga ggagtttgtc cagggaacc ttgagaggga gtgcatggag 60
[1163]	gagaagtgtc cttttgagga ggccaggag gtctttgaga aactgagcg caccactgag 120
[1164]	ttctggaaac agtatgtgga tggggaccag tgtgagtcca acccctgcct gaatgggggc 180
[1165]	agctgcaagg atgacatcaa cagctatgag tgctgggtgcc cttttggctt tgagggcaag 240
[1166]	aactgtgagc tggatgtgac ctgcaacatc aagaatggca gatgtgagca gttctgcaag 300
[1167]	aactctgtg acaacaagg ggtgtgctcc tgcactgagg gctaccgcct ggctgagaac 360
[1168]	cagaagagct gtgagcctgc tgtgccattc ccatgtggca gagtctctgt gagccagacc 420
[1169]	agcaagctca ccagg 435
[1170]	<210> 51
[1171]	<211> 145
[1172]	<212> DNA
[1173]	<213> 人工序列(Artificial Sequence)
[1174]	<220>
[1175]	<223> AAV2_5'-ITR

- [1176] <400> 51
- [1177] ttggccactc cctctctgcg cgctcgctcg ctactgagg ccgggcgacc aaaggtcgcc 60
- [1178] cgacgcccgg gctttgcccg ggccggcctca gtgagcgagc gagcgcgag agagggagtg 120
- [1179] gccaaactcca tcactagggg ttcct 145
- [1180] <210> 52
- [1181] <211> 330
- [1182] <212> DNA
- [1183] <213> 人工序列(Artificial Sequence)
- [1184] <220>
- [1185] <223> 截短的TTR增强子/启动子
- [1186] <400> 52
- [1187] cgagggcact gggaggatgt tgagtaagat ggaaaactac tgatgaccct tgcagagaca 60
- [1188] gagtattagg acatgtttga acaggggccc ggcatcagc aggtagctct agaggatccc 120
- [1189] cgtctgtctg cacatttcgt agagcgagtg ttccgatact ctaatctccc taggcaaggt 180
- [1190] tcatatttgt gtaggttact tattctcctt ttgttgacta agtcaataat cagaatcagc 240
- [1191] aggtttggag tcagcttggc agggatcagc agcctgggtt ggaaggaggg ggtataaaag 300
- [1192] ccccttcacc aggagaagcc gtcacacaga 330
- [1193] <210> 53
- [1194] <211> 77
- [1195] <212> DNA
- [1196] <213> 人工序列(Artificial Sequence)
- [1197] <220>
- [1198] <223> MVM内含子
- [1199] <400> 53
- [1200] ctaaggtaa ttggcgccgt ttaagggatg gttggttggg ggggtattaa tgtttaatta 60
- [1201] ccttttttac aggcctg 77
- [1202] <210> 54
- [1203] <211> 234
- [1204] <212> DNA
- [1205] <213> 人工序列(Artificial Sequence)
- [1206] <220>
- [1207] <223> BGH_多聚腺苷酸化信号
- [1208] <400> 54
- [1209] cctagagctc gctgatcagc ctgactgtg ctttctagtt gccagccatc tgttgtttgc 60
- [1210] ccctcccccg tgccttcctt gaccctggaa ggtgccactc cactgtcct ttcctaataa 120
- [1211] aatgaggaaa ttgcatcgca ttgtctgagt aggtgtcatt ctattctggg ggggtggggtg 180
- [1212] gggcaggaca gcaaggggga ggattgggaa gacaatagca ggcattgctgg ggaa 234
- [1213] <210> 55
- [1214] <211> 145
- [1215] <212> DNA
- [1216] <213> 人工序列(Artificial Sequence)
- [1217] <220>

- [1218] <223> AAV2 3'-ITR
[1219] <400> 55
[1220] aggaaccct agtgatggag ttggcactc cctctctgcg cgtcgtcgcg ctcactgagg 60
[1221] ccgggcgacc aaaggtcgcc cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc 120
[1222] gagcgcgcag agagggagtg gccaa 145
[1223] <210> 56
[1224] <211> 35
[1225] <212> PRT
[1226] <213> 人工序列(Artificial Sequence)
[1227] <220>
[1228] <223> FIX-AP-AA
[1229] <400> 56
[1230] Ala Glu Thr Val Phe Pro Asp Val Asp Tyr Val Asn Ser Thr Glu Ala
[1231] 1 5 10 15
[1232] Glu Thr Ile Leu Asp Asn Ile Thr Gln Ser Thr Gln Ser Phe Asn Asp
[1233] 20 25 30
[1234] Phe Thr Arg
[1235] 35
[1236] <210> 57
[1237] <211> 105
[1238] <212> DNA
[1239] <213> 人工序列(Artificial Sequence)
[1240] <220>
[1241] <223> CS02-AP-NA
[1242] <400> 57
[1243] gctgagacag tgttccctga tgtggactat gtgaatagca cagaggctga gaccatcctg 60
[1244] gacaacatca ccagagcac ccagtccttc aatgacttca ccaga 105
[1245] <210> 58
[1246] <211> 105
[1247] <212> DNA
[1248] <213> 人工序列(Artificial Sequence)
[1249] <220>
[1250] <223> CS03-AP-NA
[1251] <400> 58
[1252] gctgagactg tgttccctga tgtggactat gtgaatagca ctgaggctga gaccatcctg 60
[1253] gacaacatca ccagagcac ccagtccttc aatgacttca ccaga 105
[1254] <210> 59
[1255] <211> 105
[1256] <212> DNA
[1257] <213> 人工序列(Artificial Sequence)
[1258] <220>
[1259] <223> CS04-AP-NA

[1260] <400> 59
 [1261] gctgagacag tggtccctga tgtggactat gtgaacagca cagaggetga gaccatcctg 60
 [1262] gacaacatca cccagagcac ccagtccttc aatgacttca ccaga 105
 [1263] <210> 60
 [1264] <211> 105
 [1265] <212> DNA
 [1266] <213> 人工序列(Artificial Sequence)
 [1267] <220>
 [1268] <223> CS05-AP-NA
 [1269] <400> 60
 [1270] gctgagactg tggtccctga cgtggactat gtgaacagca ctgaggetga gaccatcctg 60
 [1271] gacaacatca cccagagcac ccagagcttc aatgacttca ccagg 105
 [1272] <210> 61
 [1273] <211> 105
 [1274] <212> DNA
 [1275] <213> 人工序列(Artificial Sequence)
 [1276] <220>
 [1277] <223> CS06-AP-NA
 [1278] <400> 61
 [1279] gctgagactg tggtccctga tgtggactat gtgaacagca ctgaggetga aaccatcctg 60
 [1280] gacaacatca cccagagcac ccagagcttc aatgacttca ccaga 105
 [1281] <210> 62
 [1282] <211> 145
 [1283] <212> PRT
 [1284] <213> 人工序列(Artificial Sequence)
 [1285] <220>
 [1286] <223> FIX-LC-AA
 [1287] <400> 62
 [1288] Tyr Asn Ser Gly Lys Leu Glu Glu Phe Val Gln Gly Asn Leu Glu Arg
 [1289] 1 5 10 15
 [1290] Glu Cys Met Glu Glu Lys Cys Ser Phe Glu Glu Ala Arg Glu Val Phe
 [1291] 20 25 30
 [1292] Glu Asn Thr Glu Arg Thr Thr Glu Phe Trp Lys Gln Tyr Val Asp Gly
 [1293] 35 40 45
 [1294] Asp Gln Cys Glu Ser Asn Pro Cys Leu Asn Gly Gly Ser Cys Lys Asp
 [1295] 50 55 60
 [1296] Asp Ile Asn Ser Tyr Glu Cys Trp Cys Pro Phe Gly Phe Glu Gly Lys
 [1297] 65 70 75 80
 [1298] Asn Cys Glu Leu Asp Val Thr Cys Asn Ile Lys Asn Gly Arg Cys Glu
 [1299] 85 90 95
 [1300] Gln Phe Cys Lys Asn Ser Ala Asp Asn Lys Val Val Cys Ser Cys Thr
 [1301] 100 105 110

[1302]	Glu Gly Tyr Arg Leu Ala Glu Asn Gln Lys Ser Cys Glu Pro Ala Val
[1303]	115 120 125
[1304]	Pro Phe Pro Cys Gly Arg Val Ser Val Ser Gln Thr Ser Lys Leu Thr
[1305]	130 135 140
[1306]	Arg
[1307]	145
[1308]	<210> 63
[1309]	<211> 235
[1310]	<212> PRT
[1311]	<213> 人工序列(Artificial Sequence)
[1312]	<220>
[1313]	<223> FIX-HC-AA
[1314]	<400> 63
[1315]	Val Val Gly Gly Glu Asp Ala Lys Pro Gly Gln Phe Pro Trp Gln Val
[1316]	1 5 10 15
[1317]	Val Leu Asn Gly Lys Val Asp Ala Phe Cys Gly Gly Ser Ile Val Asn
[1318]	20 25 30
[1319]	Glu Lys Trp Ile Val Thr Ala Ala His Cys Val Glu Thr Gly Val Lys
[1320]	35 40 45
[1321]	Ile Thr Val Val Ala Gly Glu His Asn Ile Glu Glu Thr Glu His Thr
[1322]	50 55 60
[1323]	Glu Gln Lys Arg Asn Val Ile Arg Ile Ile Pro His His Asn Tyr Asn
[1324]	65 70 75 80
[1325]	Ala Ala Ile Asn Lys Tyr Asn His Asp Ile Ala Leu Leu Glu Leu Asp
[1326]	85 90 95
[1327]	Glu Pro Leu Val Leu Asn Ser Tyr Val Thr Pro Ile Cys Ile Ala Asp
[1328]	100 105 110
[1329]	Lys Glu Tyr Thr Asn Ile Phe Leu Lys Phe Gly Ser Gly Tyr Val Ser
[1330]	115 120 125
[1331]	Gly Trp Gly Arg Val Phe His Lys Gly Arg Ser Ala Leu Val Leu Gln
[1332]	130 135 140
[1333]	Tyr Leu Arg Val Pro Leu Val Asp Arg Ala Thr Cys Leu Arg Ser Thr
[1334]	145 150 155 160
[1335]	Lys Phe Thr Ile Tyr Asn Asn Met Phe Cys Ala Gly Phe His Glu Gly
[1336]	165 170 175
[1337]	Gly Arg Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro His Val Thr Glu
[1338]	180 185 190
[1339]	Val Glu Gly Thr Ser Phe Leu Thr Gly Ile Ile Ser Trp Gly Glu Glu
[1340]	195 200 205
[1341]	Cys Ala Met Lys Gly Lys Tyr Gly Ile Tyr Thr Lys Val Ser Arg Tyr
[1342]	210 215 220
[1343]	Val Asn Trp Ile Lys Glu Lys Thr Lys Leu Thr

[1344]	225	230	235
[1345]	<210> 64		
[1346]	<211> 235		
[1347]	<212> PRT		
[1348]	<213> 人工序列(Artificial Sequence)		
[1349]	<220>		
[1350]	<223> FIXp-HC-AA		
[1351]	<400> 64		
[1352]	Val Val Gly Gly Glu Asp Ala Lys Pro Gly Gln Phe Pro Trp Gln Val		
[1353]	1 5 10 15		
[1354]	Val Leu Asn Gly Lys Val Asp Ala Phe Cys Gly Gly Ser Ile Val Asn		
[1355]	20 25 30		
[1356]	Glu Lys Trp Ile Val Thr Ala Ala His Cys Val Glu Thr Gly Val Lys		
[1357]	35 40 45		
[1358]	Ile Thr Val Val Ala Gly Glu His Asn Ile Glu Glu Thr Glu His Thr		
[1359]	50 55 60		
[1360]	Glu Gln Lys Arg Asn Val Ile Arg Ile Ile Pro His His Asn Tyr Asn		
[1361]	65 70 75 80		
[1362]	Ala Ala Ile Asn Lys Tyr Asn His Asp Ile Ala Leu Leu Glu Leu Asp		
[1363]	85 90 95		
[1364]	Glu Pro Leu Val Leu Asn Ser Tyr Val Thr Pro Ile Cys Ile Ala Asp		
[1365]	100 105 110		
[1366]	Lys Glu Tyr Thr Asn Ile Phe Leu Lys Phe Gly Ser Gly Tyr Val Ser		
[1367]	115 120 125		
[1368]	Gly Trp Gly Arg Val Phe His Lys Gly Arg Ser Ala Leu Val Leu Gln		
[1369]	130 135 140		
[1370]	Tyr Leu Arg Val Pro Leu Val Asp Arg Ala Thr Cys Leu Leu Ser Thr		
[1371]	145 150 155 160		
[1372]	Lys Phe Thr Ile Tyr Asn Asn Met Phe Cys Ala Gly Phe His Glu Gly		
[1373]	165 170 175		
[1374]	Gly Arg Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro His Val Thr Glu		
[1375]	180 185 190		
[1376]	Val Glu Gly Thr Ser Phe Leu Thr Gly Ile Ile Ser Trp Gly Glu Glu		
[1377]	195 200 205		
[1378]	Cys Ala Met Lys Gly Lys Tyr Gly Ile Tyr Thr Lys Val Ser Arg Tyr		
[1379]	210 215 220		
[1380]	Val Asn Trp Ile Lys Glu Lys Thr Lys Leu Thr		
[1381]	225 230 235		

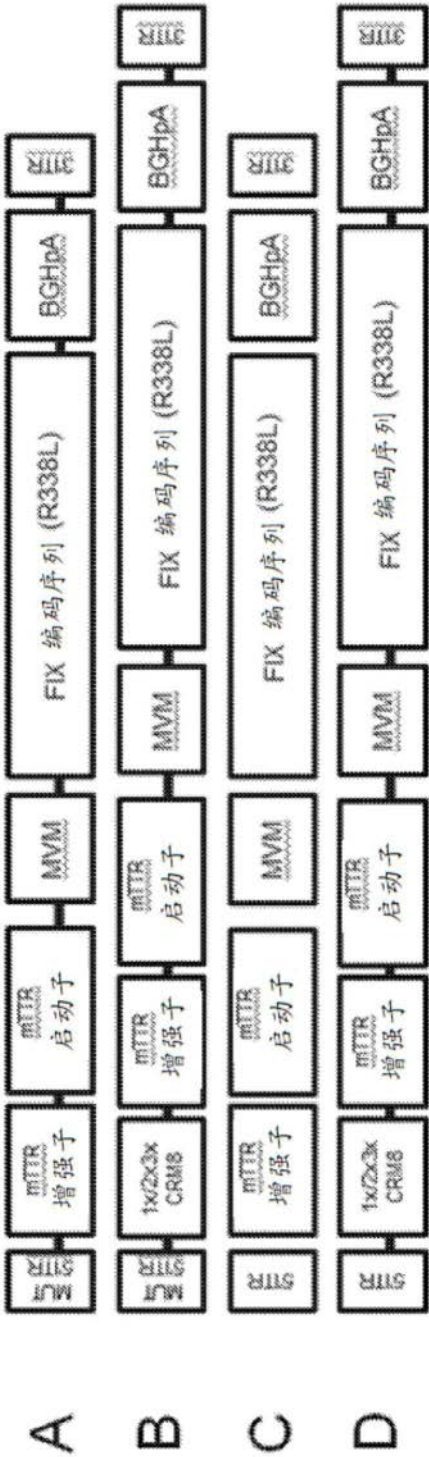


图1

FIX-FL-NA

ATGCAGCGCGTGAACATGATCATGGCAGAATCACCAGGCCTCATCACCATCTGCCTTTTAGGATATC
TACTCAGTGCTGAATGTACAGTTTTTCTTGATCATGAAAACGCCAACAAAATTCTGAATCGGCCAAA
GAGGTATAATTCAGGTAAATTGGAAGAGTTTGTTCAGGGAACTTGAGAGAGAATGTATGGAAGAA
AAGTGTAGTTTTGAAGAAGCACGAGAAGTTTTTGAAAACACTGAAAGAACAACCTGAATTTTGGAAGC
AGTATGTTGATGGAGATCAGTGTGAGTCCAATCCATGTTTAAATGGCGGCAGTTGCAAGGATGACAT
TAATTCCTATGAATGTTGGTGTCCCTTTGGATTTGAAGGAAAGAACTGTGAATTAGATGTAACATGT
AACATTAAGAATGGCAGATGCGAGCAGTTTTGTAAAAATAGTGCTGATAACAAGGTGGTTTGCTCCT
GTACTGAGGGATATCGACTTGCAGAAAACCAGAACTCCTGTGAACCAGCAGTGCCATTTCCATGTGG
AAGAGTTTCTGTTTACAAACTTCTAAGCTCACCCGTGCTGAGACTGTTTTTCTGATGTGGACTAT
GTAAATTCTACTGAAGCTGAAACCATTTTGGATAACATCACTCAAAGCACCCAATCATTTAATGACT
TCACTCGGGTTGTTGGTGGAGAAGATGCCAAACCAGGTCAATTCCCTTGGCAGGTTGTTTTGAATGG
TAAAGTTGATGCATTCTGTGGAGGCTCTATCGTTAATGAAAAATGGATTGTAACCTGCTGCCACTGT
GTTGAACTGGTGTAAAATTACAGTTGTGCGAGGTGAACATAATATTGAGGAGACAGAACATACAG
AGCAAAAGCGAAATGTGATTGCAATTATTCCTCACCACAACTACAATGCAGCTATTAATAAGTACAA
CCATGACATTGCCCTTCTGGAAGTGGACGAACCTTAGTGCTAAACAGCTACGTTACACCTATTTGC
ATTGCTGACAAGGAATACACGAACATCTTCCTCAAATTTGGATCTGGCTATGTAAGTGGCTGGGGAA
GAGTCTTCCACAAAGGGAGATCAGCTTTAGTTCCTCAGTACCTTAGAGTTCACCTTGTGACCGAGC
CACATGTCTTCGATCTACAAAGTTCACCATCTATAACAACATGTTCTGTGCTGGCTTCCATGAAGGA
GGTAGAGATTGATGTCAAGGAGATAGTGGGGGACCCCATGTTACTGAAGTGGGAAGGGACCGATTTCT
TAACTGGAATTATTAGCTGGGGTGAAGAGTGTGCAATGAAAGGCAAATATGGAATATATACCAAGGT
ATCCCGGTATGTCAACTGGATTAAGGAAAAACAAAGCTCACTTAA (SEQ ID NO:1)

图2

FIX-FL-AA

MQRVNMIMAE SPGLITICLL GYLLSAECTV FLDHENANKI LNRPKRYNSG
KLEEFVQGNL ERECMEEKCS FEEAREVFEN TERTTEFWKQ YVDGDQCESN
PCLNGGSKD DINSYECWCP FGFEGKNCEL DVTCNIKNGR CEQFCKNSAD
NKVVCSCTEG YRLAENQKSC EPAVPFPCGR VSVSQTSKLT RAETVFPDVD
YVNSTEAEI LDNITQSTQS FNDFTRVVGG EDAKPGQFPW QVVLNGKVDA
FCGGSIVNEK WIVTAAHCVE TGVKITVVAG EHNIEETEHT EQKRNVIIRI
PHHYNAAIN KYNHDIALLE LDEPLVLNSY VTPICIIDKE YTNIFLKFGS
GYVSGWGRVF HKGRSALVLQ YLRVPLVDRA TCLRSTKFTI YNNMFCAGFH
EGGRDSCQGD SGGPHVTEVE GTSFLTGIIS WGEECAMKGK YGIYTKVSRY
VNWIKEKTKL T (SEQ ID NO:2)

图3A

FIX2-FL-AA

mqrvmimae spgliticll gyllsaectv fldhenanki lnrpkrynsg kleefvqgnl
 erecmeecks feearevfen terttefwkq yvdtcnikn grceqfckns adnkvvcsct
 egyrlaenqk scepavpfpc grvsvsqtsk ltraetvfpd vdyvnsteae tildnitqst
 qsfndftrvv ggedakpggf pwqvvlngkv dafcggsivn ekwivtaahc vetgvkitvv
 agehnieete hteqkrnvir iiphhnynaa inkynhdial leldeplvln syvtpiciad
 keytniflkf gsgyvsgwgr vfhkgrsalv lqylrvplvd ratclrstkf tiynnmfcag
 fhegrdscq gdsggphvte vegtsfltgi iswgeecamk gkygiytkvs ryvnwikekt
 klt (SEQ ID NO:3)

图3B

FIXp-FL-AA

MQRVNMIMAE SPGLITICLL GYLLSAECTV FLDHENANKI LNRPKRYNSG
 KLEEFVQGNL ERECMEEKCS FEEAREVFEN TERTTEFWKQ YVDGDQCESN
 PCLNGGSCKD DINSYECWCP FGFEGKNCEL DVTCNIKNGR CEQFCKNSAD
 NKVVCSCTEG YRLAENQKSC EPAVPFPCGR VSVSQTSKLT RAETVFPDVD
 YVNSTEAEI LDNITQSTQS FNDFTRVVG GEDAKPGQFPW QVVLNGKVDA
 FCGGSIVNEK WIVTAAHCVE TGVKITVVAG EHNIEETEHT EQKRN VIRII
 PHHYNAAIN KYNHDIALLE LDEPLVLNSY VTPICIADKE YTNIFLKFGS
 GYVSGWGRVF HKGRSALVLQ YLRVPLVDRA TCLLSTKFTI YNNMFCAGFH
 EGGRDSCQGD SGGPHVTEVE GTSFLTGIIS WGEECAMKGK YGIYTKVSRY
 VNWIKEKTKL T (SEQ ID NO:4)

图4

CS02-FL-NA

atgcagaggggtgaacatgatcatggctgagagccctggcctgatcaccatctgcctgctgggctac
ctgctgtcagcagagtgcacagtgttcctggaccatgagaatgccacaagatcctgaacaggccc
aagagatacaactcaggcaagctggaggagtttgtgcagggcaacctggagagggagtgcatggag
gagaagtgcagctttgaggaggccagagaggtgtttgagaacacagagaggaccacagagttctgg
aagcagtatgtggatggagaccagtgtgagagcaacccttgccctgaatggaggcagctgcaaggat
gacatcaacagctatgagtgtggtgcccttttggctttgagggcaagaactgtgagctggatgtg
acctgcaacatcaagaatggcaggtgtgagcagttctgcaagaactcagctgacaacaaagtgggtg
tgtagctgcacagagggctacagactggctgagaaccagaagagctgtgagcctgctgtgcccttc
ccctgtggcagagtgtcagtgtcccagaccagcaagctgaccagagctgagacagtgttcctgat
gtggactatgtgaatagcacagaggctgagaccatcctggacaacatcacccagagcacccagtc
ttcaatgacttcaccagagttgtgggaggagaggatgccaaagcctggccagttcccctggcaggtg
gtgctgaatggcaagtggatgccttctgtggaggcagcattgtgaatgagaagtggattgtgaca
gctgccactgtgtggagacaggagtgaagatcacagtgggtggctggagaacacaatattgaggag
acagagcacacagagcagaagaggaatgtcatcaggattatccccaccacaactacaatgctgcc
atcaacaagtacaacatgacattgccctgctggagctggatgagcctctggtgctgaatagctat
gtgacccccatctgcattgctgacaaggagtacaccaacatcttcctgaagtttggctcaggctat
gtgtcaggctggggcagagtgttcacaagggcagatcagccctggtgctgcagtacctgagagt
cccctggtggacagagccacctgctgttgagcaccaagttcaccatctacaacaacatgttctgt
gctggcttccatgagggaggcagagacagctgccaggagagactcaggaggacccccatgtgacagaa
gtggagggcaccagcttctgacaggcatcatcagctgggggagaggagtgtgcatgaagggaag
tatggcatctacaccaaagtgagcagatatgtgaactggatcaaggagaaaaccaagctgacctga
(SEQ ID NO:5)

图5

CS03-FL-NA

atgcagaggggtgaacatgatcatggctgagagccctggcctgatcaccatctgcctgctgggctac
ctgctgtcgtgagtgcactgtgttcctggaccatgagaatgccacaagatcctgaacaggccc
aagagatacaactctggcaagctggaggagtttgtgcagggcaacctggagagggagtgcatggag
gagaagtgcagctttgaggaggccagggaagtgtttgagaacactgagaggaccactgagttctgg
aagcagtatgtggatggggaccagtgtgagagcaacccttgccctgaatgggggagctgcaaggat
gacatcaacagctatgagtgtggtgcccttttggctttgagggcaagaactgtgagctggatgtg
acctgcaacatcaagaatggcaggtgtgagcagttctgcaagaactctgctgacaacaaagtgggtg
tgtagctgcactgagggctacagactggctgagaaccagaagagctgtgagcctgctgtgcccttc
ccctgtggcagagtgtctgtgtcccagaccagcaagctgaccagagctgagactgtgttcctgat
gtggactatgtgaatagcactgaggctgagaccatcctggacaacatcacccagagcacccagtc
ttcaatgacttcaccagagttgtgggggggaggatgccaaagcctggccagttcccctggcaggtg
gtgctgaatggcaagtggatgccttctgtggggggcagcattgtgaatgagaagtggattgtgact
gctgccactgtgtggagactggggtgaagatcactgtggtggctggggaacacaatattgaggag
actgagcacactgagcagaagaggaatgtcatcaggattatccccaccacaactacaatgctgcc
atcaacaagtacaacatgacattgccctgctggagctggatgagcctctggtgctgaatagctat
gtgacccccatctgcattgctgacaaggagtacaccaacatcttcctgaagtttggctctggctat
gtgtctggctggggcagagtgttcacaagggcaggtctgcctggtgctgcagtacctgagagt
cccctggtggacagagccacctgctgtgagcaccaagttcaccatctacaacaacatgttctgt
gctggcttccatgagggggcagagacagctgccagggggactctgggggcccccatgtgactgaa
gtggaGggcaccagcttctgactggcatcatcagctgggggggaggagtgtgcatgaagggaag
tatggcatctacaccaaagtgagcaggtatgtgaactggatcaaggagaaaaccaagctgacctga
(SEQ ID NO:6)

图6

CS04-FL-NA

atgcagaggggtgaacatgattatggctgagagccctggcctgatcaccatctgcctgctgggctac
ctgctgtctgctgagtgacagtgttcctggaccatgagaatgccacaagatcctgaacaggccc
aagagatacaactctggcaagctggaggagtttgtgcagggcaacctggagagggagtgcatggag
gagaagtgcagctttgaggaggccaggaggtgtttgagaacacagagaggaccacagagtcttg
aagcagtatgtggatggtgaccagtgtgagagcaaccttgccctgaatggaggcagctgcaagga
gacatcaacagctatgagtgtggtgcccttttggctttgagggcaagaactgtgagctggatgtg
acctgcaacatcaagaatggcaggtgtgagcagttctgcaagaactctgctgacaacaagggtgtg
tgtagctgcacagagggctacagactggctgagaaccagaagagctgtgagcctgctgtgcccttc
ccctgtggcagagtgtctgtgtcccagaccagcaagctgaccagagctgagacagtgttccctgat
gtggactatgtgaacagcacagaggctgagaccatcctggacaacatcaccagagcaccagtc
ttcaatgacttcaccagagtgggtgggaggagaggatgccaaagcctggccagttcccctggcaggtg
gtgctgaatggcaaggtggatgccttctgtggaggcagcattgtgaatgagaagtggattgtgaca
gctgccactgtgtggagacaggagtgaagatcacagtgggtggctggagagcacaacattgaggag
acagagcacacagagcagaagagggaatgtgatcaggatcatccctcaccacaactacaatgctgcc
atcaacaagtacaacatgacattgccctgctggagctggatgagcctctggtgctgaacagctat
gtgaccctatctgcattgtgacaaaggagtacaccaacatcttccctgaagtttggctctggctat
gtgtctggctggggcagagtgttccacaagggcaggtctgccctgggtgctgcagtacctgagagt
cccctggtggacagagccacctgctgttgagcaccaagttcaccatctacaacaacatgttctgt
gctggcttccatgagggaggcagagacagctgccaggggtgactctggaggacccccatgtgacagag
gtggaGggcaccagcttccctgacaggcatcatcagctgggggagaggagtgtgccatgaagggcaag
tatggcatctacaccaaagtgagcagatatgtgaactggatcaaggagaagaccaagctgacctga
(SEQ ID NO:7)

图7

CS05-FL-NA

atgcagaggggtgaacatgattatggctgagagccctggcctgatcaccatctgcctgctgggctac
ctgctgtctgctgagtgacagtgttcctggaccatgagaatgccacaagatcctgaacggcccc
aagcgctacaactctggcaagctggaggagtttgtgcagggcaacctggagagggagtgcatggag
gagaagtgcagctttgaggaggccaggaggtgtttgagaacactgagcgcaccactgagttctgg
aagcagtatgtggatggggaccagtgtgagagcaaccttgccctgaatgggggggagctgcaaggat
gacatcaacagctatgagtgtggtgcccttttggctttgagggcaagaactgtgagctggaatgtg
acctgcaacatcaagaatggccgctgtgagcagttctgcaagaactctgctgacaacaagggtgtg
tgctcttgactgagggctaccgctggctgagaaccagaagagctgtgagcctgctgtgcccttc
ccctgtggcaggggtgtctgtgagccagaccagcaagctgaccagggctgagactgtgttccctgac
gtggactatgtgaacagcactgaggctgagaccatcctggacaacatcaccagagcaccagagc
ttcaatgacttcaccaggggtgggtgggaggagaggatgccaaagcctggccagttcccctggcaggtg
gtgctgaatggcaaggtggatgccttctgtggaggcagcattgtgaatgagaagtggattgtgacc
gctgccactgtgtggagactggagtgaagatcactgtggtggctggggagcacaacattgaggag
acagagcacacagagcagaagcgcaatgtgatcaggatcatccccaccacaactacaatgctgcc
atcaacaagtacaacatgacattgccctgctggagctggatgagccctgggtgctgaacagctac
gtgacccccatctgcattgcagacaaggagtacaccaacatcttccctgaagtttggctctggctat
gtgtctggctggggcaggggtgttccacaagggcaggtctgccctgggtgctgcagtacctgaggggtg
cccctggtggacagggccacctgctgtgagcaccaagttcaccatctacaacaacatgttctgtg
gctggcttccatgagggagggaagggaacagctgccagggagactctggaggcccccatgtgacagag
gtggagggcaccagcttccctgacaggcatcatcagctggggggagaggagtgtgccatgaagggcaag
tatggcatctacaccaaagtgtcccgtatgtgaactggatcaaggagaagaccaagctgacctga
(SEQ ID NO:8)

图8

CS06-FL-NA

atgcagaggggtcaacatgatcatggctgagtcacctggcctcatcaccatctgcctgctgggctac
ctgctgtctgctgagtgcaactgtcttcttgacctgagaaatgccacaagatcctcaacaggccc
aagagatacaactctggcaactggaggagtttgtccagggcaacctggagagggagtgcatggag
gagaagtgtcctttgaggaggccaggaggtctttgagaacactgagcgccacctgagttctgg
aaacagtatgtggatggggaccagtgtgagtcacaacctgcctgaatgggggagctgcaaggat
gacatcaacagctatgagtgctggtgcccctttggctttgagggaagaactgtgagctggatgtg
acctgcaacatcaagaatggcagatgtgagcagttctgcaagaactctgctgacaacaagggtgtg
tgctcctgcaactgagggctaccgcctggctgagaaccagaagagctgtgagcctgctgtgccattc
ccatgtggcagagtctctgtgagccagaccagcaagctcaccagggctgagactgtgttccctgat
gtggactatgtgaacagcaactgaggctgaaacctcctggacaacatcaccagagcaccagagc
ttcaatgacttcaccagagtgggtgggaggagaggatgccaaacctggccagtctccctggcaagt
gtgctcaatggcaagggtggatgccttctgtgggggctccattgtgaatgagaagtggattgtcact
gctgccactgtgtggagactgggtcaagatcactgtggtggctggggagcacaacttgaggag
actgagcacactgagcagaagcgcaatgtgatcaggatcatccccaccacaactacaatgctgcc
atcaacaagtacaacctgacattgccctgctggagctggatgagccctgggtcctcaacagctat
gtgaccccatctgcattgctgacaaggagtacaccaacatcttctcaagtttggctctggctat
gtctctggctggggcagagtgttcacaaaaggcaggtctgcctgggtgctccagtaacctgagagt
ccctgggtggacagggccacctgcctcttgagcaccagttcaccatctacaacaacatgttctgt
gctggcttccatgagggaggaagagacagctgccagggggactctggaggaccccatgtcactgag
gtggagggcacctccttctcactggcatcatctcctggggagaggagtggtgccatgaaaggcaaa
tatggcatctacaccaaagtctccagatatgtcaactggatcaaggagaagaccaagctgacctga
(SEQ ID NO:9)

图9

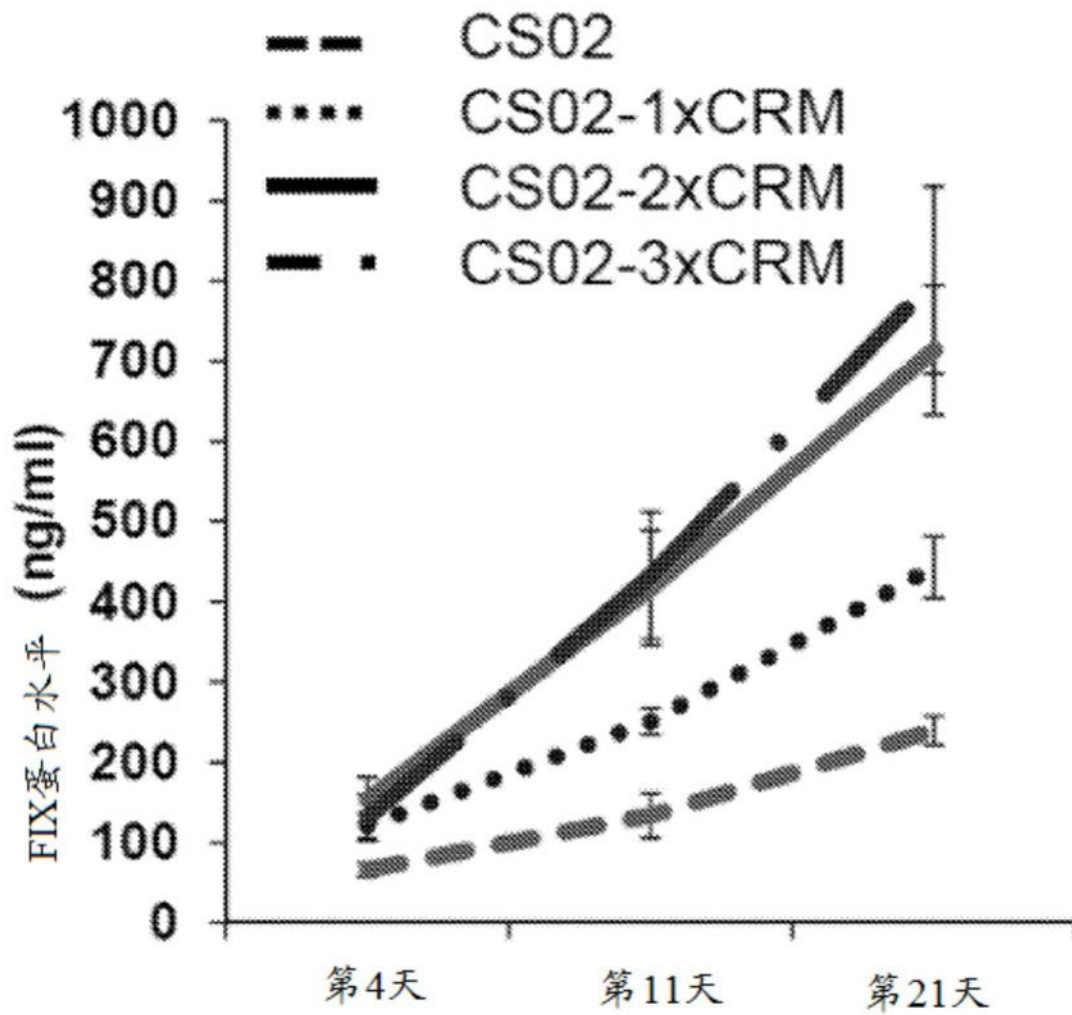


图10

FIX-MP-AA

YNSG KLEEFVQGNL ERECMECKCS FEEAREVFEN TERTTEFWKQ YVDGDQCESN
 PCLNGGSKCD DINSYECWCP FGFEGKNCEL DVTCNIKNGR CEQFCKNSAD
 NKVVCSCTEG YRLAENQKSC EPAVPFPCGR VSVSQTSLT RAETVFPDWD
 YVNSTEAETI LDNITQSTQS FNDFTRVVGG EDAKPGQFPW QVVLNGKVDA
 FCGGSIVNEK WIVTAAHCVE TGVKITVVAG EHNIEETEHT EQKRNVIIRI
 PHHNYNAAIN KYNHDIALLE LDEPLVLNSY VTPICIADKE YTNIFLKFGS
 GYVSGWGRVF HKGRSALVLQ YLRVPLVDRA TCLRSTKFTI YNNMFCAGFH
 EGGRDSCQGD SGGPHVTEVE GTSFLTGIIS WGEECAMKGK YGIYTKVSRV
 VNWIKEKTKL T (SEQ ID NO:10)

图11A

FIX2-MP-AA

ynsg kleefvqgnl erecmeecks feearevfen terttefwkq yvdtvcnikn
grceqfckns adnkvvcsct egyrlaenqk scepavpfpc grvsvsqtsk ltraetvfpd
vdyvnsteae tildnitqst qsfndftrvv ggedakpggf pwqvvlngkv dafcggsivn
ekwivtaahc vetgvkitvv agehnieete hteqkrnvir iiphhnynaa inkynhdial
leldeplvln syvtpiciad keytniflkf gsgyvsgwgr vfhkgrsalv lqylrvplvd
ratclrstkf tiynnmfcag fheggrdscq gdsggphvte vegtsfltgi iswgeecamk
gkygiytkvs ryvnwikekt klt (SEQ ID NO:11)

图11B

FIXp-MP-AA

YNSG KLEEFVQGNL ERECMEEKCS FEEAREVFEN TERTTEFWKQ YVDGDQCESN
PCLNGGSCKD DINSYECWCP FGFEKNCCEL DVTCNIKNGR CEQFCKNSAD
NKVVCSTEG YRLAENQKSC EPAVPFPCGR VSVSQTSLT RAETVFPDVS
YVNSTEAEI LDNITQSTQS FNDFTRVVGG EDAKPGQFPW QVVLNGKVDA
FCGGSIVNEK WIVTAAHCVE TGVKITVVAG EHNIEETEHT EQKRNVIIRI
PHHNYNAAIN KYNHDIALLE LDEPLVLNSY VTPICIADKE YTNIFLKFGS
GYVSGWGRVF HKGRSALVLQ YLRVPLVDRA TLLSTKFTI YNNMFCAGFH
EGGRDSCQGD SGGPHVTEVE GTSFLTGIIS WGEECAMKGK YGIYTKVSRY
VNWIKEKTKL T (SEQ ID NO:12)

图12

CS02-MP-NA

```
tacaactcag gcaagctgga ggagtttgtg cagggcaacc tggagagggga gtgcatggag
gagaagtgca gctttgagga ggccagagag gtgtttgaga acacagagag gaccacagag
ttctggaagc agtatgtgga tggagaccag tgtgagagca acccttgccct gaatggaggc
agctgcaagg atgacatcaa cagctatgag tgctgggtgcc cttttggctt tgaggggcaag
aactgtgagc tggatgtgac ctgcaacatc aagaatggca ggtgtgagca gttctgcaag
aactcagctg acaacaaagt ggtgtgtagc tgcacagagg gctacagact ggctgagaac
cagaagagct gtgagcctgc tgtgcccttc ccctgtggca gagtgtcagt gtcccagacc
agcaagctga ccagagctga gacagtgttc cctgatgtgg actatgtgaa tagcacagag
gctgagacca tcctggacaa catcacccag agcacccagt ctttcaatga cttcaccaga
gttggtgggag gagaggatgc caagcctggc cagttcccct ggcaggtggg gctgaatggc
aaagtggatg ctttctgtgg aggcagcatt gtgaatgaga agtggattgt gacagctgcc
cactgtgtgg agacaggagt gaagatcaca gtgggtggctg gagaacacaa tattgaggag
acagagcaca cagagcagaa gaggaatgtc atcaggatta tccccacca caactacaat
gctgccatca acaagtacaa ccatgacatt gccctgctgg agctggatga gcctctggtg
ctgaatagct atgtgacccc catctgcatt gctgacaagg agtacaccaa catcttcctg
aagtttggct caggctatgt gtcaggctgg ggcagagtgt tccacaaggg cagatcagcc
ctggtgctgc agtacctgag agtgcccctg gtggacagag ccacctgcct gttgagcacc
aagttcacca tctacaacaa catgttctgt gctggcttcc atgagggagg cagagacagc
tgccagggag actcaggagg accccatgtg acagaagtgg agggcaccag cttcctgaca
ggcatcatca gctggggaga ggagtgtgcc atgaagggca agtatggcat ctacaccaa
gtgagcagat atgtgaactg gatcaaggag aaaaccaagc tgacctga
(SEQ ID NO:13)
```

图13

CS03-MP-NA

```
tacaactctg gcaagctgga ggagtttgtg cagggcaacc tggagagggg gtgcatggag
gagaagtgca gctttgagga ggccagggaa gtgtttgaga aactgagag gaccactgag
ttctggaagc agtatgtgga tggggaccag tgtgagagca acccttgctt gaatgggggc
agctgcaagg atgacatcaa cagctatgag tgctgggtgcc cttttggctt tgagggcaag
aactgtgagc tggatgtgac ctgcaacatc aagaatggca ggtgtgagca gttctgcaag
aactctgctg acaacaaagt ggtgtgtagc tgcactgagg gctacagact ggctgagaac
cagaagagct gtgagcctgc tgtgcccttc cctgtggca gagtgtctgt gtcccagacc
agcaagctga ccagagctga gactgtgttc cctgatgtgg actatgtgaa tagcactgag
gctgagacca tcctggacaa catcacccag agcacccagt ctttcaatga cttcaccaga
gtggtggggg gggaggatgc caagcctggc cagttcccct ggcaggtggt gctgaatggc
aaagtggatg ctttctgtgg gggcagcatt gtgaatgaga agtggattgt gactgctgcc
cactgtgtgg agactggggg gaagatcact gtggtggctg gggaacacaa tattgaggag
actgagcaca ctgagcagaa gaggaatgtc atcaggatta tccccacca caactacaat
gctgccatca acaagtacaa ccatgacatt gccctgctgg agctggatga gcctctggtg
ctgaatagct atgtgacccc catctgcatt gctgacaagg agtacaccaa catcttcctg
aagtttggct ctggctatgt gtctggctgg ggcagagtgt tccacaaggg caggtctgcc
ctggtgctgc agtacctgag agtgcccctg gtggacagag ccacctgcct gctgagcacc
aagttcacca tctacaacaa catgttctgt gctggcttcc atgagggggg cagagacagc
tgccaggggg actctggggg ccccatgtg actgaagtgg agggcaccag cttcctgact
ggcatcatca gctgggggga ggagtgtgcc atgaaggga agtatggcat ctacaccaa
gtgagcaggt atgtgaactg gatcaaggag aaaaccaagc tgacctga
(SEQ ID NO:14)
```

图14

CS04-MP-NA

```
tacaactctg gcaagctgga ggagtttgtg cagggcaacc tggagagggga gtgcatggag
gagaagtgca gctttgagga ggccagggag gtgtttgaga acacagagag gaccacagag
ttctggaagc agtatgtgga tgggtgaccag tgtgagagca acccttgccct gaatggaggc
agctgcaagg atgacatcaa cagctatgag tgctgggtgcc cttttggctt tgagggcaag
aactgtgagc tggatgtgac ctgcaacatc aagaatggca ggtgtgagca gttctgcaag
aactctgctg acaacaaggt ggtgtgtagc tgcacagagg gctacagact ggctgagaac
cagaagagct gtgagcctgc tgtgcccttc ccctgtggca gagtgtctgt gtcccagacc
agcaagctga ccagagctga gacagtgttc cctgatgtgg actatgtgaa cagcacagag
gctgagacca tcctggacaa catcacccag agcacccagt ccttcaatga cttcaccaga
gtggtgggag gagaggatgc caagcctggc cagttcccct ggcaggtggt gctgaatggc
aaggtggatg ccttctgtgg aggcagcatt gtgaatgaga agtggattgt gacagctgcc
cactgtgtgg agacaggagt gaagatcaca gtggtggctg gagagcaca cttgaggag
acagagcaca cagagcagaa gaggaatgtg atcaggatca tccctacca caactacaat
gctgccatca acaagtacaa ccatgacatt gccctgctgg agctggatga gcctctggtg
ctgaacagct atgtgacccc tatctgcatt gctgacaagg agtacacca catcttcctg
aagtttggct ctggctatgt gtctggctgg ggcagagtgt tccacaaggg caggtctgcc
ctggtgctgc agtacctgag agtgcccctg gtggacagag ccacctgcct gttgagcacc
aagttcacca tctacaacaa catgttctgt gctggcttcc atgagggagg cagagacagc
tgccaggggtg actctggagg accccatgtg acagaggtgg agggcaccag cttcctgaca
ggcatcatca gctggggaga ggagtgtgcc atgaagggca agtatggcat ctacacaaa
gtgagcagat atgtgaactg gatcaaggag aagaccaagc tgacctga
(SEQ ID NO:15)
```

图15

CS05-MP-NA

```
tacaactctg gcaagctgga ggagtttgtg cagggcaacc tggagagggga gtgcatggag
gagaagtgca gctttgagga ggccagggag gtgtttgaga acactgagcg caccactgag
ttctggaagc agtatgtgga tggggaccag tgtgagagca accctgcct gaatggggg
agctgcaagg atgacatcaa cagctatgag tgctgggtgcc cttttggctt tgagggcaag
aactgtgagc tggatgtgac ctgcaacatc aagaatggcc gctgtgagca gttctgcaag
aactctgctg acaacaaggt ggtgtgctct tgcactgagg gctaccgcct ggctgagaac
cagaagagct gtgagcctgc tgtgcccttc ccctgtggca ggggtgtctgt gagccagacc
agcaagctga ccagggctga gactgtgttc cctgacgtgg actatgtgaa cagcactgag
gctgagacca tcctggacaa catcacccag agcacccaga gcttcaatga cttcaccagg
gtggtgggag gagaggatgc caagcctggc cagttcccct ggcaggtggt gctgaatggc
aaggtggatg ccttctgtgg aggcagcatt gtgaatgaga agtggattgt gaccgctgcc
cactgtgtgg agactggagt gaagatcact gtggtggctg gggagcaca cttgaggag
acagagcaca cagagcagaa ggcgaatgtg atcaggatca tccccacca caactacaat
gctgccatca acaagtacaa ccatgacatt gccctgctgg agctggatga gccctggtg
ctgaacagct acgtgacccc catctgcatt gcagacaagg agtacacca catcttcctg
aagtttggct ctggctatgt gtctggctgg ggcaggtgt tccacaaggg caggtctgcc
ctggtgctgc agtacctgag ggtgccctg gtggacaggg ccacctgcct gctgagcacc
aagttcacca tctacaacaa catgttctgc gctggcttcc atgagggagg aaggacagc
tgccaggggag actctggagg ccccatgtg acagaggtgg agggcaccag cttcctgaca
ggcatcatca gctgggggga ggagtgtgcc atgaagggca agtatggcat ctacacaaa
gtgtcccgt atgtgaactg gatcaaggag aagaccaagc tgacctga
(SEQ ID NO:16)
```

图16

CS06-MP-NA

```
tacaactctg gcaaactgga ggagtttgtc cagggcaacc tggagagggga gtgcatggag
gagaagtgtc cctttgagga ggccagggag gtctttgaga aactgagcg caccactgag
ttctggaaac agtatgtgga tggggaccag tgtgagtcca acccctgcct gaatgggggc
agctgcaagg atgacatcaa cagctatgag tgctgggtgcc cctttggctt tgagggcaag
aactgtgagc tggatgtgac ctgcaacatc aagaatggca gatgtgagca gttctgcaag
aactctgtctg acaacaaggt ggtgtgctcc tgcactgagg gctaccgcct ggctgagaac
cagaagagct gtgagcctgc tgtgccattc ccatgtggca gagtctctgt gagccagacc
agcaagctca ccagggctga gactgtgttc cctgatgtgg actatgtgaa cagcactgag
gctgaaacca tcctggacaa catcacccag agcacccaga gcttcaatga cttcaccaga
gtggtgggag gagaggatgc caagcctggc cagttcccct ggcaagtggg gctcaatggc
aaggtggatg ccttctgtgg gggctccatt gtgaatgaga agtggattgt cactgctgcc
cactgtgtgg agactggggg caagatcact gtggtggctg gggagcacaa cattgaggag
actgagcaca ctgagcagaa ggcgaatgtg atcaggatca tccccacca caactacaat
gctgccatca acaagtacaa ccatgacatt gccctgctgg agctggatga gccctggtc
ctcaacagct atgtgacccc catctgcatt gctgacaagg agtacacca catcttcctc
aagtttggct ctggctatgt ctctggctgg ggcagagtgt tccacaaagg caggtctgcc
ctggtgctcc agtacctgag agtgcccctg gtggacaggg ccacctgcct cttgagcacc
aagttcacca tctacaacaa catgttctgt gctggcttcc atgagggagg aagagacagc
tgccaggggg actctggagg accccatgtc actgaggtgg agggcacctc cttcctcact
ggcatcatct cctggggaga ggagtgtgcc atgaaaggca aatatggcat ctacaccaa
gtctccagat atgtcaactg gatcaaggag aagaccaagc tgacctga
(SEQ ID NO:17)
```

图17

FIX-PPP-NA

atgcagcgcg tgaacatgat catggcagaa tcaccaggcc tcatcaccat ctgcctttta
ggatatctac tcagtgtga atgtacagtt tttcttgatc atgaaaacgc caacaaaatt
ctgaatcggc caaagagg (SEQ ID NO:18)

CS02-PPP-NA

atgcagaggg tgaacatgat catggctgag agccctggcc tgatcaccat ctgcctgctg
ggctacctgc tgtcagcaga gtgcacagtgt ttcctggacc atgagaatgc caacaagatc
ctgaacaggc ccaagaga (SEQ ID NO:19)

CS03-PPP-NA

atgcagaggg tgaacatgat catggctgag agccctggcc tgatcaccat ctgcctgctg
ggctacctgc tgtctgctga gtgcactgtg ttcctggacc atgagaatgc caacaagatc
ctgaacaggc ccaagaga (SEQ ID NO:20)

CS04-PPP-NA

atgcagaggg tgaacatgat tatggctgag agccctggcc tgatcaccat ctgcctgctg
ggctacctgc tgtctgctga gtgcacagtgt ttcctggacc atgagaatgc caacaagatc
ctgaacaggc ccaagaga (SEQ ID NO:21)

CS05-PPP-NA

atgcagaggg tgaacatgat tatggctgag agccctggcc tgatcaccat ctgcctgctg
ggctacctgc tgtctgctga gtgcactgtg ttcctggacc atgagaatgc caacaagatc
ctgaaccgcc ccaagcgc (SEQ ID NO:22)

CS06-PPP-NA

atgcagaggg tcaacatgat catggctgag tcccctggcc tcatcaccat ctgcctgctg
ggctacctgc tgtctgctga gtgcactgtc ttcctggacc atgagaatgc caacaagatc
ctcaacaggc ccaagaga (SEQ ID NO:23)

图18

FIX-SP-NA

atgcagcgcg tgaacatgat catggcagaa tcaccaggcc tcatcaccat ctgcctttta
ggatatctac tcagtgtga atgt (SEQ ID NO:24)

CS02-SP-NA

atgcagaggg tgaacatgat catggctgag agccctggcc tgatcaccat ctgcctgtg
ggctacctgc tgtcagcaga gtgc (SEQ ID NO:25)

CS03-SP-NA

atgcagaggg tgaacatgat catggctgag agccctggcc tgatcaccat ctgcctgtg
ggctacctgc tgtctgtga gtgc (SEQ ID NO:26)

CS04-SP-NA

atgcagaggg tgaacatgat tatggctgag agccctggcc tgatcaccat ctgcctgtg
ggctacctgc tgtctgtga gtgc (SEQ ID NO:27)

CS05-SP-NA

atgcagaggg tgaacatgat tatggctgag agccctggcc tgatcaccat ctgcctgtg
ggctacctgc tgtctgtga gtgc (SEQ ID NO:28)

CS06-SP-NA

atgcagaggg tcaacatgat catggctgag tcccctggcc tcatcaccat ctgcctgtg
ggctacctgc tgtctgtga gtgc (SEQ ID NO:29)

图19

FIX-PP-NA

acagtttttc ttgatcatga aaacgccaac aaaattctga atcggccaaa gagg
(SEQ ID NO:30)

CS02-PP-NA

acagtgttcc tggaccatga gaatgccaac aagatcctga acaggcccaa gaga
(SEQ ID NO:31)

CS03-PP-NA

actgtgttcc tggaccatga gaatgccaac aagatcctga acaggcccaa gaga
(SEQ ID NO:32)

CS04-PP-NA

acagtgttcc tggaccatga gaatgccaac aagatcctga acaggcccaa gaga
(SEQ ID NO:33)

CS05-PP-NA

actgtgttcc tggaccatga gaatgccaac aagatcctga accgccccaa gcgc
(SEQ ID NO:34)

CS06-PP-NA

actgtcttcc tggaccatga gaatgccaac aagatcctca acaggcccaa gaga
(SEQ ID NO:35)

图20

FIX-PPP-AA

MQRVNMIMAE SPGLITICLL GYLLSAECTV FLDHENANKI LNRPKR (SEQ ID NO:36)

图21

FIX-SP-AA

MQRVNMIMAE SPGLITICLL GYLLSAEC (SEQ ID NO:37)

图22

FIX-PP-AA

TV FLDHENANKI LNRPKR (SEQ ID NO:38)

图23

CRM8gggggaggctgctggtgaatattaaccaaggtcacccagttatcggaggagcaaacaggggctaag
tccac (SEQ ID NO:39)

图24

CS06-CRM8.3-ssV

ttggccactccctctctgcgcgctcgtcgtcactgagggccggcgaccaaaggctcgcccgacgcc
cgggctttgccccggcgccctcagtgagcgagcgagcgcgagagagggagtgcccaactccatcac
taggggttccctgagtttaaacttcgtcgacgggggaggctgctggtgaatattaaccaaggtcaccc
cagttatcggaggagcaaacaggggctaagtccaccgggggaggctgctggtgaatattaaccaagg
tcaccccagttatcggaggagcaaacaggggctaagtccaccgggggaggctgctggtgaatatta
ccaaggtcaccccagttatcggaggagcaaacaggggctaagtccaccgagggcactgggaggatgt
tgagtaagatggaaaactactgatgacccttgagagacagagtattaggacatgtttgaacagggg
ccgggagatcagcaggttagctctagaggatccccgtctgtctgcacatttcgtagagcgagtgttcc
gatactctaatactccctaggcaagggttcataatttggttaggttacttattctcctttgttgactaa
gtcaataatcagaatcagcaggtttggagtcagcttggcagggatcagcagcctgggttggaaggag
ggggtataaaaagcccccttcaccaggagaagccgtcacacagactaggcgcgccctaaggtaagttgg
cgccgtttaagggatggttgggtgggtggttattaatgtttaattacctttttacaggcctgaaga
tctgccaccatgcagaggggtcaacatgatcatggctgagtcacctggcctcatcaccatctgcctgc
tgggctacctgctgtctgctgagtgactgtcttcttgaccatgagaatgccaacaagatcctcaa
cagggccaagagatacaactctggcaaactggaggagtttgccaggggcaacctggagagggagtg
atggaggagaagtgtcctttgaggaggccaggagggtctttgagaacactgagcgcaccactgagt
tctggaaacagtatgtggatggggaccagtggtgagtcacaacctgcctgaatgggggcagctgcaa
ggatgacatcaacagctatgagtgctgggtgccccctttggctttgagggaagaactgtgagctggat
gtgacctgcaacatcaagaatggcagatgtgagcagttctgcaagaactctgctgacaacaagggtg
tgtgctcctgcactgagggctaccgcctggctgagaaccagaagagctgtgagcctgctgtgccatt
cccatgtggcagagtctctgtgagccagaccagcaagctcaccagggtgagactgtgttccctgat
gtggactatgtgaacagcactgaggtgaaacctcctggacaacatcacccagagcacccagagct
tcaatgacttcaccagagtgggtgggaggagaggatgccaagcctggccagttccccctggcaagtgg
gctcaatggcaagggtggatgccttctgtgggggtccattgtgaatgagaagtggaattgtcactgct
gcccactgtgtggagactggggtcaagatcactgtggtggctggggagcacaacattgaggagactg
agcacactgagcagaagcgcaatgtgatcaggatcatccccaccacaactacaatgctgccatcaa
caagtacaacatgacattgccctgctggagctggatgagccccctggctcctcaacagctatgtgacc
cccatctgcattgctgacaaggagtacaccaacatcttctcaagtttggtctgtggctatgtctctg
gctggggcagagtgttccacaaaggcaggtctgccctggtgctccagtacctgagagtggccctgg
ggacagggccacctgcctcttgagcaccaagttcaccatctacaacaacatgttctgtgctggcttc
catgaggggaggaagagacagctgccagggggactctggaggaccccatgtcactgaggtggagggca
cctccttccctcactggcatcatctcctggggagaggagtgtgccatgaaaggcaaatatggcatcta
caccaaagtctccagatatgtcaactggatcaaggagaagaccaagctgacctgatgagcatgccta
gagctcgtgatcagcctcgactgtgccttctagtgtccagccatctgttgtttgccccctccccgt
gccttcccttgacctggaagggtgccactcccactgtccttccctaataaaaatgaggaaattgcatcg
cattgtcTgagtaggtgtcattctattctggggggtgggggtggggcaggacagcaagggggaggatt
gggaagacaatagcaggcatgctggggaattaattaagctcgcaaggaacccctagtgtgaggtt

(续)

图25A

(续)

ggccactccctctctgcgcgctcgctcgctcactgagggcgggcgaccaaaggctcgcccgacgcccg
ggctttgccccggggcgctcagtgagcgagcgagcgcgagagagggagtgcccaagacgatttaaa
tgacaagcttggcgtaatcatgggtcatagctgtttcctgtgtgaaattgttatccgctcacaattcc
acacaacatacagagccggaagcataaagtgtaaagcctgggggtgcctaatactgagtgagctaactcaca
ttaattgctgttgcgctcactgcccgtttccagtcgggaaacctgtcggtgccagctgcattaatgaa
tcggccaacgcgcggggagagggcggtttgcgtattggggcgctcttccgcttccctcgctcactgactc
gctgcgctcggtcggttcggctgcggcgagcggtatcagctcactcaaaggcggttaatacggttatcc
acagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgta
aaaaggccgcgttgctggcggttttccataggctccgccccctgacgagcatcacaaaaatcgacg
ctcaagtacagaggtggcgaaacccgacaggactataaagataccaggcggtttccccctggaagctcc
ctcgctgcgctctcctgttccgacctgcccgttaccggatacctgtccgcctttctcccttcgggaa
gcggtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcggttcgctccaagct
gggctgtgtgcacgaaccccccggttcagcccagccgctgcgccttatccggtaactatcgctcttgag
tccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcga
ggtatgtaggcggtgctacagagttcttgaagtgggtggcctaactacggctacactagaagaacagt
atttgggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggc
aaacaaaccacgctggtagcggtgggttttttgtttgcaagcagcagattacgcgcagaaaaaaag
gatctcaagaagatcctttgatctttctacggggctctgacgctcagtggaacgaaaaactcacgtta
agggattttgggtcatgagattatcaaaaaggatcttcacctagatccttttaaatataaaatgaagt
tttaaatcaatctaaagtatatatgagtaaacttgggtcagagttaccaatgcttaatcagtgagg
cacctatctcagcgatctgtctatttcgttcacatagttgcctgactccccgctcggtgtagataac
tacgatacggggagggcttaccatctggccccagtgctgcaatgataccgcgagaccacgctcaccg
gctccagatttatcagcaataaaaccagccagccggaaggccgagcgcagaagtgggtccctgcaactt
tatccgcctccatccagctctattaattgttgcgggaagctagagtaagtagttcgccagtttaatag
tttgcgcaacgttggttgcattgctacagggcatcggtgtcacgctcgctcggttggtatggcttca
ttcagctccggttcccaacgatcaaggcgagttacatgatccccatggttgcaaaaaagcggtta
gctccttcggtcctccgatcggttgctcagaagtaagttggccgagtggttatcactcatggttatggc
agcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactca
accaagtcatcttgagaatagtgatgctggcgaccgagttgctcttgccggcgctcaatacgggata
ataccgcgccacatagcagaactttaaaagtgtcatcattggaaaacggttcttcggggcgaaaact
ctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcggtgcacccaactgatcttca
gcatcttttactttcaccagcggttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagg
gaataagggcgacacggaaatgttgaaatactcatactcttcccttttcaatattattgaagcattta
tcagggttattgtctcatgagcggatacatatttgaaatgtatttagaaaaataaacaataggggtt
ccgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccattattatcatgacattaacct
ataaaaaataggcggtatcacgaggcccttctgctcgcgcgttccggtgatgacggtgaaaacctctg
acacatgcagctcccggagacgggtcacagcttgtctgtaagcggtgcccgggagcagacaagcccgt
cagggcgcgctcagcgggtgttggcggtgtcggggtggtgcttaactatgcggcatcagagcagattg
tactgagagtgacccatattcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcacag
gcgccatttcgccattcaggtgcgcaactgttggaaggcgatcggtgcgggcctcTtcgctatta
cgccagctggcgaaagggggatgtgctgcaaggcgattaagttgggtaacgccagggttttccagct
cacgacgttgtaaaacgacggccagtgtaattcctcgagatttaaatgacg

(SEQ ID NO:40)

图25B