

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公布说明书

[21] 申请号 200880006978.1

[51] Int. Cl.

A61K 38/46 (2006.01)

A61K 48/00 (2006.01)

C12N 9/22 (2006.01)

C12N 15/10 (2006.01)

C12N 15/52 (2006.01)

C12N 15/62 (2006.01)

[43] 公开日 2010 年 3 月 24 日

[11] 公开号 CN 101678087A

[22] 申请日 2008.1.31

[21] 申请号 200880006978.1

[30] 优先权

[32] 2007.2.1 [33] IB [31] PCT/IB2007/000849

[86] 国际申请 PCT/IB2008/001333 2008.1.31

[87] 国际公布 WO2008/093249 英 2008.8.7

[85] 进入国家阶段日期 2009.9.2

[71] 申请人 赛莱克蒂斯公司

地址 法国罗曼维尔

[72] 发明人 埃曼努埃尔·法雅尔多桑切斯

西尔韦斯特·格里佐 马克·伊萨兰

路易斯·塞拉诺普布尔

弗朗索瓦·斯特里谢

[74] 专利代理机构 北京集佳知识产权代理有限公司

代理人 顾晋伟 韩宏星

权利要求书 5 页 说明书 51 页 序列表 27 页
附图 21 页

[54] 发明名称

专性异二聚体大范围核酸酶及其用途

[57] 摘要

本发明涉及一种专性异二聚体大范围核酸酶，其由第一和第二单体组成，所述单体来自两种不同的同二聚化 LAGLIDADG 内切核酸酶单体（母体单体），并且具有至少一对突变引入到所述母体单体中使得各母体同二聚化 LAGLIDADG 内切核酸酶的两个单体之间形成分子间相互作用的对应残基上；本发明还涉及编码所述大范围核酸酶的载体；以所述载体修饰的动物或植物；以及所述大范围核酸酶及衍生产物用于分子生物学、基因组工程和基因组治疗的用途。

1. 一种专性异二聚体大范围核酸酶，其由第一和第二单体组成，所述单体来自两种不同的同二聚化 LAGLIDADG 内切核酸酶单体（母体单体），并且具有至少一对突变引入到所述母体单体中使得各母体同二聚化 LAGLIDADG 内切核酸酶的两个单体之间形成分子间相互作用的对应残基上，其中所述突变对中的第一突变在所述第一单体中，所述突变对中的第二突变在所述第二单体中，所述突变对破坏由各单体形成功能性同二聚体，而不阻止形成能切割非回文杂合 DNA 靶标的功能性异二聚体，所述非回文杂合 DNA 靶标包含被各自母体同二聚化内切核酸酶切割的 DNA 靶标中不同的一半。

2. 权利要求 1 的专性异二聚体大范围核酸酶，其中所述单体中的第一单体和第二单体分别具有下述突变对中的至少一种：

a) 8 位上的谷氨酸被碱性氨基酸替换，7 位上的赖氨酸被酸性氨基酸替换，

b) 61 位上的谷氨酸被碱性氨基酸替换，96 位上的赖氨酸被酸性氨基酸替换，

c) 97 位上的亮氨酸被芳香族氨基酸替换，54 位上的苯丙氨酸被小氨基酸替换，以及

d) 137 位上的天冬氨酸被碱性氨基酸替换，51 位上的精氨酸被酸性氨基酸替换，

所述位置参照 I-CreI 氨基酸序列 SEQ ID NO: 1 来表示。

3. 权利要求 2 的专性异二聚体大范围核酸酶，其中如 a) 或 b) 中所述，所述单体的 8 位或 61 位上的谷氨酸被碱性氨基酸替换，7 位和 96 位上的赖氨酸残基中的至少一个被精氨酸替换。

4. 权利要求 2 的专性异二聚体大范围核酸酶，其中如 c) 中所述的 97 位上的亮氨酸被芳香族氨基酸替换的单体还包含 54 位上的苯丙氨酸被色氨酸替换。

5. 权利要求 2 的专性异二聚体大范围核酸酶，其中，如 c) 中所述

的 54 位上的苯丙氨酸被小氨基酸替换的单体还包含 58 位上的亮氨酸或 57 位上的赖氨酸被甲硫氨酸替换。

6. 权利要求 2 的专性异二聚体大范围核酸酶, 其中所述酸性氨基酸是谷氨酸。

7. 权利要求 2 或 3 的专性异二聚体大范围核酸酶, 其中所述碱性氨基酸是精氨酸。

8. 权利要求 2 或 4 的专性异二聚体大范围核酸酶, 其中所述芳香族氨基酸是苯丙氨酸。

9. 权利要求 2 或 5 的专性异二聚体大范围核酸酶, 其中所述小氨基酸是甘氨酸。

10. 权利要求 1-9 中任一项的专性异二聚体大范围核酸酶, 其由至少具有突变 D137R 的第一单体和至少具有突变 R51D 的第二单体组成。

11. 权利要求 1-9 中任一项的专性异二聚体大范围核酸酶, 其包含如权利要求 2 的 a)、b)、c) 或 d) 中所定义的至少两对突变。

12. 权利要求 11 的专性异二聚体大范围核酸酶, 其中一个单体的 7 位和 96 位上的赖氨酸残基被酸性氨基酸替换, 另一个单体的 8 位和 61 位上的谷氨酸残基被碱性氨基酸替换。

13. 权利要求 11 或 12 的专性异二聚体大范围核酸酶, 其包含如权利要求 2 的 a)、b) 和 c) 中所定义的 3 对突变。

14. 权利要求 13 的专性异二聚体大范围核酸酶, 其由第一单体和第二单体组成, 所述第一单体至少具有突变 K7R、E8R、E61R、K96R 和 L97F 或者 K7R、E8R、F54W、E61R、K96R 和 L97F, 所述第二单体至少具有突变 K7E、F54G、L58M 和 K96E 或者 K7E、F54G、K57M 和 K96E。

15. 权利要求 1-14 中任一项的专性异二聚体大范围核酸酶, 其由两种 I-CreI 单体变体组成, 其还包含 I-CreI 的 26-40 位以及 44-77 位的不同突变。

16. 权利要求 15 的专性异二聚体大范围核酸酶, 其能切割来自目的基

基因组序列的非回文杂合 DNA 靶标, 所述目的基因组序列对应于被各母体同二聚化内切核酸酶切割的 DNA 靶标中不同的一半。

17. 一种单链大范围核酸酶, 其包含通过肽连接物连接的如权利要求 1-16 中任一项所定义的专性异二聚体大范围核酸酶的第一单体和第二单体。

18. 权利要求 1-16 中任一项的专性异二聚体大范围核酸酶的单体。

19. 一种多核苷酸片段, 其编码权利要求 18 的单体或权利要求 17 的单链大范围核酸酶。

20. 一种重组载体, 其包含权利要求 19 的至少一种多核苷酸片段。

21. 一种包含两种多核苷酸片段的表达载体, 所述多核苷酸片段各自编码权利要求 1-16 中任一项的专性异二聚体大范围核酸酶的两个单体中之一, 所述片段可操纵地与允许产生所述两种单体的调节序列相连接。

22. 一种包含多核苷酸片段的表达载体, 所述多核苷酸片段编码权利要求 17 的单链大范围核酸酶, 并且所述片段可操纵地与允许产生所述单链大范围核酸酶的调节序列相连接。

23. 权利要求 21 或权利要求 22 的载体, 其包含靶向 DNA 构建体, 所述构建体含有与权利要求 1 或 16 中所定义的杂合 DNA 靶标序列的周围区域有同源性的序列。

24. 权利要求 23 的载体, 其中所述靶向 DNA 构建体包含: a) 与权利要求 1 或 16 中所定义的杂合 DNA 靶标序列的周围区域有同源性的序列, 以及 b) 待引入的序列, 其两侧为 a) 中的序列。

25. 一种宿主细胞, 其包含权利要求 19 或权利要求 21 中所定义的一种或两种多核苷酸片段或者权利要求 20-24 中任一项的载体。

26. 一种非人转基因动物, 其包含权利要求 19 或权利要求 21 中所定义的一种或两种多核苷酸片段。

27. 一种转基因植物, 其包含权利要求 19 或权利要求 21 中所定义的一种或两种多核苷酸片段。

28. 权利要求 1-16 中任一项的至少一种专性异二聚体大范围核酸酶、权利要求 17 的一种单链大范围核酸酶、权利要求 19 或权利要求 21 中所定义的一种或两种多核苷酸片段、权利要求 21-24 中任一项的一种载体、权利要求 25 的一种宿主细胞、权利要求 27 的一种转基因植物、权利要求 26 的一种非人转基因哺乳动物用于分子生物学、体内或体外基因工程以及体内或体内基因组工程的用途。

29. 权利要求 28 的用途，用于诱导目的位点中的双链核酸断裂，从而诱导 DNA 重组事件、DNA 丢失或细胞死亡，其中所述目的位点包含所述专性异二聚体大范围核酸酶所切割的杂合 DNA 靶序列。

30. 权利要求 28 或权利要求 29 的用途，其中所述双链核酸断裂是为了以下目的：修复特定序列、修饰特定序列、恢复功能性基因而替换突变基因、弱化或活化内源目的基因、将突变引入目的位点、引入外源基因或其部分、失活或消灭内源基因或其部分、使染色体臂易位或使 DNA 不被修复并降解。

31. 权利要求 28-30 中任一项的用途，其中所述专性异二聚体大范围核酸酶、单链大范围核酸酶、多核苷酸、载体、细胞、转基因植物或非人转基因哺乳动物与权利要求 23 或 24 中所定义的靶向 DNA 构建体相关联。

32. 一种基因工程方法，其包括以下步骤：通过使包含权利要求 1 或权利要求 16 中所述的杂合 DNA 靶标的载体与权利要求 1-16 中任一项的专性异二聚体大范围核酸酶或权利要求 17 的单链大范围核酸酶相接触，使得位于所述载体上的目的位点发生双链核酸断裂，从而诱导与另一载体进行同源重组，所述另一载体与所述专性异二聚体大范围核酸酶的切割位点的周围序列有同源性。

33. 一种基因组工程方法，其包括以下步骤：1) 使权利要求 1 或权利要求 16 中所述的杂合 DNA 靶标的基因组基因座处发生双链断裂，所述断裂通过使所述 DNA 靶标与权利要求 1-16 中任一项的专性异二聚体大范围核酸酶或权利要求 17 的单链大范围核酸酶相接触而实现；2) 将所述断裂的基因组基因座维持在适于与靶向 DNA 构建体进行同源重组的条件下，所述构建体包含待引入到所述基因座中的序列，其两侧是与所靶向基因座有同源性的序列。

34. 一种基因组工程方法，其包括以下步骤：1) 使包含权利要求 1 或权利要求 16 中所定义的至少一种杂合 DNA 靶标的基因组基因座处发生双链断裂，所述断裂通过使所述杂合 DNA 靶标与权利要求 1-16 中任一项的专性异二聚体大范围核酸酶或权利要求 17 的单链大范围核酸酶相接触而实现；2) 将所述断裂的基因组基因座维持在适于与染色体 DNA 进行同源重组的条件下，所述染色体 DNA 与所靶向基因座周围区域有同源性。

35. 一种组合物，其至少包含以下至少之一：权利要求 1-16 中任一项的至少一种专性异二聚体大范围核酸酶、权利要求 17 的一种单链大范围核酸酶、权利要求 19 或权利要求 21 中所定义的一种或两种多核苷酸片段或者权利要求 21-24 中任一项的一种载体。

36. 权利要求 35 的组合物，其还包含含有修复目的位点之序列的靶向 DNA 构建体，所述目的位点的两侧为与所靶向基因座有同源性的序列。

37. 权利要求 1-16 中任一项的至少一种专性异二聚体大范围核酸酶、权利要求 17 的一种单链大范围核酸酶、权利要求 19 或权利要求 21 中所定义的一种或两种多核苷酸片段、权利要求 21-24 中任一项的一种载体在制备用于在有此需求的个体中预防、改善或治疗遗传性疾病的药物中的用途，所述药物以任意方式施用给所述个体。

38. 权利要求 1-16 中任一项的专性异二聚体大范围核酸酶、权利要求 17 的单链大范围核酸酶、权利要求 19 或权利要求 21 中所定义的一种或两种多核苷酸片段、权利要求 21-24 中任一项的载体在制备用于在有此需求的个体中预防、改善或治疗疾病的药物中的用途，所述疾病由具有 DNA 中间体的传染原所引起，所述药物以任意方式施用给所述个体。

39. 权利要求 1-16 中任一项的专性异二聚体大范围核酸酶、权利要求 17 的单链大范围核酸酶、权利要求 19 或权利要求 21 中所定义的一种或两种多核苷酸片段、权利要求 21-24 中任一项的载体用于体外抑制传染原的传播、灭活或消灭传染原的用途，或者用于物品消毒的用途，其中所述传染原具有 DNA 中间体，存在于生物来源产品或旨在用于生物用途的产品中。

40. 权利要求 38 或 39 的用途，其中所述传染原是病毒。

专性异二聚体大范围核酸酶及其用途

本发明涉及来自同二聚体 LAGLIDADG 内切核酸酶单体的专性异二聚体大范围核酸酶 (meganuclease), 编码所述大范围核酸酶的载体, 所述载体修饰的细胞、动物或植物, 以及所述大范围核酸酶和衍生产物用于分子生物学、基因组工程和基因组治疗的用途。

大范围核酸酶定义为具有大的 (12-45bp) 切割位点的序列特异性内切核酸酶, 其可导致活细胞中特定基因座处的 DNA 双链断裂 (double-strand break, DSB) (Thierry 和 Dujon, *Nucleic Acids Res.*, 1992, 20, 5625-5631)。大范围核酸酶已经用于在培养细胞和植物中刺激其靶序列附近的同源重组 (Rouet 等, *Mol. Cell. Biol.*, 1994, 14, 8096-106; Chouluka 等, *Mol. Cell. Biol.*, 1995, 15, 1968-73; Donoho 等, *Mol. Cell. Biol.*, 1998, 18, 4070-8; Elliott 等, *Mol. Cell. Biol.*, 1998, 18, 93-101; Sargent 等, *Mol. Cell. Biol.*, 1997, 17, 267-77; Puchta 等, *Proc. Natl. Acad. Sci. USA*, 1996, 93, 5055-60; Chiurazzi 等, *Plant Cell*, 1996, 8, 2057-2066), 这使得大范围核酸酶诱导的重组成为一种用于基因组工程的有效且稳定的方法。

对大范围核酸酶诱导重组的利用长期以来受到天然大范围核酸酶种类的限制, 当前技术的主要限制是需要事先在目的基因座中引入大范围核酸酶切割位点。因此, 正在大规模研究构建切割所选靶标的重新设计的大范围核酸酶。

这些蛋白质可用于切割天然的染色体序列, 并开辟基因组工程的广泛应用的新前景。例如, 大范围核酸酶可用于在染色体中敲除内源基因或敲入外源基因。其还可以用于基因矫正, 并原则上用于矫正与单基因疾病相关的突变。

近来, 将 Cys2-His2 型锌指蛋白 (ZFP) 的锌指 DNA 结合结构域与 *FokI* 内切核酸酶的催化结构域进行融合以诱导以下多种细胞类型中的重组: 哺乳动物培养细胞 (包括人淋巴细胞)、植物和昆虫 (Smith 等, *Nucleic Acids Res*, 1999, 27, 674-81; Pabo 等, *Annu. Rev. Biochem*, 2001,

70, 313-40; Porteus 和 Baltimore, *Science*, 2003, 300, 763; Urnov 等, *Nature*, 2005, 435, 646-651; Bibikova 等, *Science*, 2003, 300, 764; Durai 等, *Nucleic Acids Res.*, 2005, 33, 5978-5990; Porteus M.H., *Mol. Ther.*, 2006, 13, 438-446)。ZEP 的结合特异性相对容易操纵, 而且现在可得到能结合许多 (g/a)nn(g/a)nn(g/a)nn 序列的一系列新型人工 ZEP(Pabo 等, 出处同前; Segal 和 Barbas, *Curr. Opin. Biotechnol.*, 2001, 12, 632-7; Isalan 等, *Nat. Biotechnol.*, 2001, 19, 656-60)。然而, 保留非常窄的特异性是基因组工程应用的主要问题之一, 目前不清楚 ZFP 是否能满足非常严格的治疗应用需求。此外, 这些融合蛋白已经在果蝇(Bibikova 等, *Science*, 2003, 300, 764; Bibikova 等, *Genetics*, 2002, 161, 1169-1175)和哺乳动物 NIH3T3 细胞(Alwin 等, *Mol. Ther.*, 2005, 12, 610-617; Porteus M.H.和 Baltimore D., *Science*, 2003, 300, 763; Porteus M.H.和 Carroll D., *Nat. Biotechnol.*, 2005, 23, 967-973)中表现出了高毒性, 这很可能是由于频繁的位点外(off-site)切割而引起的基因毒性作用(Porteus, M.H., *Mol. Ther.*, 2006, 13, 438-446)。

在自然界中, 大范围核酸酶基本上由归巢内切核酸酶(homing endonuclease, HE)代表, 其为由可移动遗传元件编码的内切核酸酶家族, 其功能是在称作归巢的过程中启动 DNA 双链断裂(DSB)诱导的重组事件(Chevalier 和 Stoddard, *Nucleic Acids Res.*, 2001, 29, 3757-74; Kostriken 等, *Cell*, 1983, 35, 167-74; Jacquier 和 Dujon, *Cell*, 1985, 41, 383-94)。在细菌、真核生物和古细菌(archaea)(Chevalier 和 Stoddard, *Nucleic Acids Res.*, 2001, 29, 3757-74)中已经鉴定了数百种 HE; 然而在选定的基因中发现 HE 切割位点的可能性非常低。

由于它们的生物功能和其在效率和特异性方面优异的切割特性, HE 提供了产生用于基因组工程的新内切核酸酶的理想骨架。

过去十年中已积累了 LAGLIDADG 家族(四个 HE 家族中最大的一个)的相对充分表征的资料(Chevalier 和 Stoddard, *Nucleic Acids Res.*, 2001, 29, 3757-74)。LAGLIDADG 指的是整个所述家族中唯一的实际上保守的序列, 以一个或两个(常常更多)拷贝存在于所述蛋白质中。具有单个基序的蛋白质(例如 I-CerI)形成同二聚体并且切割回文 DNA 序列或假回文 DNA 序列, 而较大的双基序蛋白质(例如 PI-SceI)是单体并且切割非回文靶标。已经将 9 种不同的 LAGLIDADG 蛋白进

行结晶,其显示出非常显著的核心结构保守性,这与在一级序列水平上缺少相似性形成了对照 (Chevalier 等, Nat. Struct. Biol., 2001, 8, 312-316; Jurica 等, Mol. Cell., 1998, 2, 469-476; Chevalier 等, J. Mol. Biol., 2003, 329, 253-269; Moure 等, J. Mol. Biol., 2003, 334, 685-695; Moure 等, Nat. Struct. Biol., 2002, 9, 764-770; Ichiyanagi 等, J. Mol. Biol., 2000, 300, 889-901; Gimble 等, J. Biol. Chem., 1998, 273, 30524-30529; Bolduc 等, Genes Dev. 2003, 17, 2875-2888; Silva 等, J. Mol. Biol., 1999, 286, 1123-1136; Nakayama 等, J. Mol. Biol., Epub 29 septembre 2006, Spiegel 等, Structure, 2006, 14, 869-880)。

在此核心结构(图1)中,由两个单体或双 LAGLIDADG 蛋白中的两个结构域组成的两个特征性 $\alpha\beta\alpha\beta\alpha$ 折叠以双重对称方式彼此面对。DNA 结合取决于每个结构域的4个 β 链,其折叠成反向平行的 β -片层,并在 DNA 螺旋的大沟上形成鞍。对于与其天然靶标相结合的 I-CreI 结构进行的分析显示,在每个单体中,八个残基(Y33、Q38、N30、K28、Q26、Q44、R68 和 R70)与靶标 DNA 的 ± 3 、4、5、6、7、9 和 10 位的七个碱基建立直接相互作用(Jurica 等, Mol. Cell., 1998, 2, 469-76)。另外,一些残基与若干碱基建立由水介导的接触;例如 S40、K28 和 N30 与+8 和-8 位的碱基对建立接触(Chevalier 等, J. Mol. Biol., 2003, 329, 253-269)。

催化位点位于中央,由来自两个单体的螺旋形成。就在所述催化位点上方,所述两种 LAGLIDADG 肽也在二聚化界面中起到至关重要的作用。除了该核心结构以外,还可发现其它的结构域,例如 PI-SceI(蛋白质内含子)具有一个蛋白质剪切结构域和其它的 DNA 结合结构域(Moure 等, Nat. Struct. Biol., 2002, 9, 764-770; Pingoud 等, Biochemistry, 1998, 37, 8233-8243)。

大范围核酸酶家族内广泛的结构保守性鼓励人们对 HE 进行诱变和构建发生大量修饰的嵌合及单链 HE。

通过将 N 端 I-DmoI 结构域与 I-CreI 单体融合而产生功能性嵌合及单链人工 HE (Epinat 等, Nucleic Acids Res., 2003, 31, 2952-62; Chevalier 等, Mol. Cell., 2002, 10, 895-905; PCT 国际申请 WO 03/078619 和 WO 2004/031346)已证明了 LAGLIDADG 蛋白的可塑性:不同的单体或核心结构域可融合到单个蛋白质中,以获得切割新的(非回文)靶序列的新型

大范围核酸酶。

此外,不同研究组还使用推理性方法来局部改变 I-*CreI* (Seligman 等, *Nucleic Acids Res.*, 2002, 30, 3870-3879; Sussman 等, *J. Mol. Biol.*, 2004, 342, 31-41; Rosen 等, *Nucleic Acids Res.*, 2006, 34, 4791-4800; Arnould 等, *J. Mol. Biol.*, 2006, 355, 443-458 及 PCT 国际申请 WO 2006/097853 和 WO 2006/097784; Smith 等, *Nucleic Acids Res.*, 2006, 34, e149)、I-*SceI* (Doyon 等, *J. Am. Chem. Soc.*, 2006, 128, 2477-2484)、PI-*SceI* (Gimble 等, *J. Mol. Biol.*, 2003, 334, 993-1008) 和 I-*MsoI* (Ashworth 等, *Nature*, 2006, 441, 656-659) 的特异性。

近期工作表明,得到大量的识别多种靶标的经局部修饰的 I-*CreI* 大范围核酸酶变体、通过组合方法将所述变体组装以及得到具有所选特异性的完全重新设计的突变体是可能的 (Arnould 等, *J. Mol. Biol.*, 2006, 355, 443-458; PCT 国际申请 WO 2006/097853 和 WO 2006/097784; Smith 等, *Nucleic Acids Res.*, 2006, 34, e149)。

因此,通过联用半推理性方法和高通量筛选改造了几百种特异性改变的 I-*CreI* 衍生物:

-诱变了 I-*CreI* 的残基 Q44、R68 和 R70 或者 Q44、R68、D75 和 I77,通过筛选鉴定了 ± 3 至5位上特异性改变的 DNA 靶标(5NNN DNA 靶标)的变体组(PCT 国际申请 WO 2006/097784 和 WO 2006/097853; Arnould 等, *J. Mol. Biol.*, 2006, 355, 443-458; Smith 等, *Nucleic Acids Res.*, 2006, 34, e149)。

-诱变了 I-*CreI* 的残基 K28、N30 和 Q38; N30、Y33 和 Q38 或者 K28、Y33、Q38 和 S40,通过筛选鉴定了 ± 8 至10位特异性改变的 DNA 靶标(10NNN DNA 靶标)的变体组(Smith 等, *Nucleic Acids Res.*, 2006, 34, e149)。

I-*CreI* 的残基 28-40 以及 44-77 显示形成两个可分离的功能性亚结构域,其能结合归巢内切核酸酶半位点的不同部分 (Smith 等 *Nucleic Acids Res.*, 2006, 34, e149)。

在相同单体内组合 I-*CreI* 中两个亚结构域的突变使得能够设计出新

的嵌合分子（同二聚体），其能切割含有通过各亚结构域结合的 ± 3 至5位以及 ± 8 至10位的核苷酸的回文组合DNA靶序列（Smith等, *Nucleic Acids Res.*, 2006, 34, e149）。

将两种不同的变体相组合并组装成能切割由融合各变体DNA靶序列的不同的一半而得到的嵌合靶标的功能性异二聚化内切核酸酶（Arnould等, 出处同前; PCT国际申请WO 2006/097854）。有趣的是，新蛋白质保持了正确的折叠和稳定性、高活性以及窄特异性。

所述两个前述步骤的组合允许进行包括4种不同亚结构域的更大的组合方法。可分别修饰不同的亚结构域，并将其组合成能切割来自目的基因的靶标的一种大范围核酸酶变体（异二聚体或单链分子）。在第一步中，将几对新的大范围核酸酶组合在切割来自希望切割的靶标的回文靶标的新分子（“半大范围核酸酶”）中。然后，这种“半大范围核酸酶”的组合可得到切割目的靶标的异二聚体种类。在Smith等的 *Nucleic Acids Res.*, 2006, 34, e149中已经描述了将4组突变组装成切割模型靶序列或来自人RAG1基因的序列的异二聚体内切核酸酶。

这些变体可用于切割天然染色体序列，并在包括基因治疗的几个领域中开辟了具有新前景的道路。例如，大范围核酸酶诱导的重组可用于矫正与单基因遗传病（例如SCID、SCA或CFTR）相关的突变。该策略具有避开与当前的随机插入回补转基因之策略相关的不均等性的优势。

虽然I-SceI归巢内切核酸酶已表明毒性比ZEP的小（Alwin等, *Mol. Ther.*, 2005, 12, 610-617; Porteus M.H.和 Baltimore D., *Science*, 2003, 300, 763; Porteus M.H.和 Carroll D., *Nat. Biotechnol.*, 2005, 23, 967-973），这可能是由于其特异性较高，但是在非常高的剂量下I-SceI仍可能有害（Gouble等, *J. Gene Med.*, 2006, 8, 616-622）。位点外切割可通过形成蛋白质工程副产物而大大提高。迄今为止，大多数改造的内切核酸酶（ZFN和HE）是异二聚体，包括两种分别改造的单体，每种单体结合所述靶标的一半。通过两个单体在相同细胞中共表达而形成异二聚体（Porteus H.M., *Mol. Ther.*, 2006, 13, 438-446; Smith等, *Nucleic acids Res.*, 2006, 34, e149）。

然而，由于识别不同的靶标实际上与两种同二聚体的形成相关（Arnould 等, J. Mol. Biol., 2006, 355, 443-458; Bibikova 等, Genetics, 2002, 161, 1169-1175），所以单个同二聚体有时可导致非常高水平的毒性（Bibikova 等, Genetics, 2002, 161, 1169-1175; Beumer 等, Genetics, 2006, 172, 2391-403）。

因此，仍然限制单 LAGLIDADG 归巢内切核酸酶（例如 I-CreI）的较广泛应用的因素的是下面这一事实：蛋白质为同二聚体。虽然靶向两种不同 DNA 序列的两种不同 I-CreI 变体的共表达导致形成识别杂交 DNA 序列的功能性异二聚体，但是这仍然导致形成 3 种不同酶的混合物，包括两种同二聚体（Arnould 等, J. Mol. Biol., 2006, 355, 443-458; PCT 国际申请 WO 2006/097853 和 WO 2006/097784; Smith 等, Nucleic Acids Res., 2006, 34, e149）。

该特异性丢失可仅通过抑制功能性同二聚体形成来解决。理论上，此结果可通过将两种单体融合在单链分子中来实现（Epinat 等, Nucleic Acids Res., 2005, 33, 5978-5990; 国际申请 WO 03/078619）。然而，所述单链分子并不一定减轻不同分子中 $\alpha\beta\beta\alpha\beta\beta\alpha$ 折叠的相互作用，尤其是在连接物长并且具有柔性时，但是其应当至少促进相同分子的 $\alpha\beta\beta\alpha\beta\beta\alpha$ 折叠之间的相互作用。

专性异二聚体大范围核酸酶的制备提供了解决基因组工程应用中主要的特异性和毒性问题的良好折叠的功能性蛋白质。

此外，单链分子的制备和二聚化界面的重新设计并不是排他性的策略，并且它们可联合使用。

因此，本发明人重新设计了两种 I-CreI 大范围核酸酶单体之间的相互作用表面（界面）以得到专性异二聚体。

所述同二聚体的大部分二聚化界面由以卷曲螺旋排列的两种 α -螺旋（两个单体中的 Lys7 至 Gly19）组成，这使得它们难于重新设计。这些螺旋下面的氨基酸（Asp20 及其前面的）与 DNA 接触，因此决定内切核酸酶的活性（活性位点）和特异性（DNA 识别）。这些功能单独阻止了任意这些残基在设计过程中容易地修饰。因此，这使得减少功能

性同二聚体形成的可能性很小。尽管如此，本发明人已鉴定了二聚体中可在不妨碍其结合能力或其酶活性的情形下而被干扰和改变的包含在界面中的4个相互作用位点。

在各位点中，一个单体(A)中的两个残基(Z和Z')与另一个单体(B)中的对应残基(分别为Z'和Z)建立起有利的相互作用。为了保持异二聚体中的此相互作用并同时减少功能性同二聚体的形成，在一个单体中利用两个残基Z替换残基Z和Z'，而在另一个单体中利用两个残基Z'替换残基Z和Z'。因此，一个单体的一个残基(例如单体(A)中的Z)被功能上与Z'等同的残基替换，在另一个单体(B)中，Z'被功能上与Z等同的残基替换。因此，AA和BB同二聚体相排斥，而AB异二聚体的形成是有利的。

本身被改造以识别不同DNA序列的新单体允许形成功能性异二聚体并防止同二聚体位点切割。此设计显著提高了改造出用于基因组工程的高度特异性试剂的能力，并除去了将重新设计的大范围核酸酶用于基因治疗和其它应用的道路上的最后障碍之一。对于需要将基因毒性降到最低的治疗应用而言，此特异性增强可能使情况完全改观。

本发明涉及一种由第一单体和第二单体(A和B)组成的专性异二聚体大范围核酸酶，所述单体来源于两种不同的同二聚化LAGLIDADG内切核酸酶(母体单体)并且具有至少一对突变引入到所述母体单体中使得各母体同二聚化LAGLIDADG内切核酸酶的两个单体之间产生分子间相互作用的对应残基上，其中所述突变对中的第一突变在所述第一单体中，所述突变对中的第二突变在所述第二单体中，所述突变对破坏由各单体形成功能性同二聚体，而不防止能切割非回文杂合DNA靶标的功能性异二聚体的形成，所述非回文杂合DNA靶标包含各母体同二聚化内切核酸酶所切割的DNA靶标中不同的一半。

每一母体单体具有二聚化界面中的至少两个残基Z和Z'，其与分别来自同一或另一母体单体的残基Z'和Z相互作用(两对相互作用的ZZ'残基)以形成两种同二聚体和一种异二聚体。根据本发明，所述二聚化界面的两对相互作用的残基中之一相互交换，而得到分别具有两个残基Z或Z'的单体A以及具有两个残基Z'或Z的单体B。结果，各自具有两个残基Z或两个残基Z'的A和B单体相比于其母体对应部分较不容

易同二聚体化，而异二聚体 AB 界面上两对相互作用的 ZZ' 残基的存在则有利于 AB 异二聚体的形成。

定义

氨基酸指天然或合成氨基酸，包括前述氨基酸的对映体和立体异构体。

多肽序列中的氨基酸残基在本文中根据单字母代码命名，其中，例如 Q 表示 Gln 或谷氨酰胺残基，R 表示 Arg 或精氨酸残基，D 表示 Asp 或天冬氨酸残基。

酸性氨基酸指天冬氨酸 (D) 和谷氨酸 (E)。

碱性氨基酸指赖氨酸 (K)、精氨酸 (R) 和组氨酸 (H)。

小氨基酸指甘氨酸 (G) 和丙氨酸 (A)。

芳香族氨基酸指苯丙氨酸 (F)、色氨酸 (W) 和酪氨酸 (Y)。

核苷酸按如下命名：使用单字母代码命名核苷的碱基：a 为腺嘌呤，t 为胸腺嘧啶，c 为胞嘧啶，g 为鸟嘌呤。对于简并核苷酸而言，r 表示 g 或 a (嘌呤核苷酸)，k 表示 g 或 t，s 表示 g 或 c，w 表示 a 或 t，m 表示 a 或 c，y 表示 t 或 c (嘧啶核苷酸)，d 表示 g、a 或 t，v 表示 g、a 或 c，b 表示 g、t 或 c，h 表示 a、t 或 c，n 表示 g、a、t 或 c。

“大范围核酸酶”意指具有 12-45 bp 的双链 DNA 靶序列的内切核酸酶。

“同二聚体 LAGLIDADG 归巢内切核酸酶”意指具有单个 LAGLIDADG 基序并且切割回文 DNA 靶序列 (例如 I-CreI 或 I-MsoI 或其功能性变体) 的野生型同二聚体 LAGLIDADG 归巢内切核酸酶。

“LAGLIDADG 归巢内切核酸酶变体”或“变体”意指通过利用不同的氨基酸替换 LAGLIDADG 归巢内切核酸酶序列的至少一个氨基酸而得到的蛋白质。

“功能性变体”意指能切割 DNA 靶标 (优选不被野生型

LAGLIDADG 归巢内切核酸酶切割的新的 DNA 靶标)的 LAGLIDADG 归巢内切核酸酶变体。例如, 这些变体在与 DNA 靶序列相接触或者与所述 DNA 靶标直接或间接相互作用的位置上具有氨基酸变异。

“具有新特异性的大范围核酸酶变体”意指具有与母体归巢内切核酸酶不同的切割靶标模式的变体。等价且无差异地使用的术语“新特异性”、“经修饰的特异性”、“新的切割特异性”、“新的底物特异性”指变体针对 DNA 靶序列中核苷酸的特异性。

“I-CreI”意指具有序列 SWISSPROT P05725 (SEQ ID NO: 1)或 pdb 登录号 1g9y (SEQ ID NO: 43)的野生型 I-CreI。

“结构域”或“核心结构域”意指“LAGLIDADG 归巢内切核酸酶核心结构域”, 它是 LAGLIDADG 家族归巢内切核酸酶的特征性 $\alpha_1\beta_1\beta_2\alpha_2\beta_3\beta_4\alpha_3$ 折叠, 对应于约一百个氨基酸残基的序列。所述结构域包含折叠成反向平行 β -片层的四条 β 链 (β_1 、 β_2 、 β_3 、 β_4), 所述 β -片层与 DNA 靶标的一半相互作用。该结构域能够与另一 LAGLIDADG 归巢内切核酸酶核心结构域(其与所述 DNA 靶标的另一半相互作用)结合, 形成能够切割所述 DNA 靶标的功能性内切核酸酶。例如, 在二聚化归巢内切核酸酶 I-CreI (163 个氨基酸)的情况下, LAGLIDADG 归巢内切核酸酶核心结构域对应于残基 6 至 94。

-“亚结构域”意指 LAGLIDADG 归巢内切核酸酶核心结构域中的区域, 其与归巢内切核酸酶 DNA 靶标半位点的不同部分相互作用。

-“ β -发夹”是指 LAGLIDADG 归巢内切核酸酶核心结构域反向平行 β -片层的两条连续 β -链 ($\beta_1\beta_2$ 或 $\beta_3\beta_4$), 其通过环或转角相连。

-“DNA 靶标”、“DNA 靶序列”、“靶序列”、“靶位点”、“靶标”、“位点”、“识别位点”、“识别序列”、“归巢识别位点”、“归巢位点”、“切割位点”是指被 LAGLIDADG 归巢内切核酸酶识别并切割的 20 到 24 bp 的双链回文、局部回文(假回文)或非回文的多核苷酸序列。这些术语是指内切核酸酶要在其中诱导双链断裂(切割)的独特 DNA 定位, 优选地为基因组定位。所述 DNA 靶标通过双链多核苷酸的一个链的 5'至 3'序列来定义。例如, 通过野生型 I-CreI 切割的回文 DNA 靶序列被定义为定义为序列

5'- t₁₂c₁₁a₁₀a₉a₈a₇c₆g₅t₄c₃g₂t₁a₁c₂g₃a₄c₅g₆t₇l₈l₉l₁₀g₁₁a₁₂ (SEQ ID NO :2)。

“DNA 靶标半位点”、“半切割位点”或“半位点”意指 DNA 靶标中与各个 LAGLIDADG 归巢内切核酸酶核心结构域/单体结合的部分。

“嵌合 DNA 靶标”或“杂合 DNA 靶标”意指两个母体大范围核酸酶靶序列中不同之一半的融合物。另外，所述靶标的至少一半可包含通过至少两个独立的亚结构域结合的核苷酸之组合（组合 DNA 靶标）。

“载体”是指能够转运与之连接的另一核酸的核酸分子。

“同源”是指一条序列与另一序列具有足够导致序列间发生同源重组的同一性，更具体地具有至少 95% 的同一性，优选 97% 的同一性，更优选 99% 的同一性。

“同一性”是指两条核酸分子或多肽之间的序列同一性。可通过比较各序列中可为比较目的而排列的位置来测定同一性。当所比较序列中的一个位置被相同的碱基占据时，则这些分子在该位置上是相同的。核酸或氨基酸序列之间的相似性或同一性程度是核酸序列间共有的位置上相同或匹配的核苷酸数量的函数。可使用多种比对算法和/或程序计算两条序列之间的同一性，包括 FASTA 或 BLAST，它们可作为 GCG 序列分析包(University of Wisconsin, Madison, Wis.)的一部分获得，并可以例如默认设置使用。

“个体”包括哺乳动物以及其它脊椎动物（例如鸟类、鱼和爬行动物）。本文使用的术语“哺乳动物”是指任何这样的脊椎动物，包括单孔类、有袋类和胎盘动物，它们对其幼仔哺乳，并且分娩活的幼仔（真哺乳亚纲(eutharian)或胎盘哺乳动物）或者产卵（后兽亚纲(metatharian)或非胎盘(nonplacental)哺乳动物）。哺乳动物物种的实例包括人和其他灵长类（例如猴、猩猩）、啮齿动物（例如大鼠、小鼠、豚鼠）和反刍动物（例如牛、猪、马）。

“遗传性疾病”指部分或全部、直接或间接由于一种或几种基因异常而引起的任意疾病。所述异常可以是突变、插入或缺失。所述突变可以是准时突变。所述异常可影响基因或其调节序列的编码序列。所述异

常可影响基因组序列的结构或者所编码 mRNA 的结构或稳定性。所述遗传性疾病可以是隐性或显性的。这样的遗传性疾病可包括但不限于囊性纤维化、亨廷顿舞蹈病、家族性高胆固醇血症 (LDL 受体缺陷)、嗜肝母细胞瘤、威尔森氏病、先天性肝性卟啉症、肝脏代谢的遗传病、Lesch-Nyhan 综合征、镰刀形红细胞贫血病、地中海型贫血、着色性干皮病、范可尼贫血、视网膜色素变性、共济失调性毛细血管扩张综合症、Bloom 综合征、视网膜母细胞瘤、杜氏肌营养不良以及 Tay-Sachs 病。

根据本发明,参照 I-CreI 氨基酸序列 SEQ ID NO: 1 表示突变位置。了解了 I-CreI 结构中的突变位置 (pdb 登录号 1g9y), 本领域技术人员利用众所周知的蛋白质结构分析软件 (例如 Pymol) 可容易地推出另一种同二聚体 LAGLIDADG 归巢内切核酸酶中的对应位置。例如, I-CreI 中的 K96 位和 E61 位对应于 I-MsoI 中的 R102 位和 Q64 位。此外, 对于 I-MsoI 而言, 两个功能性亚结构域位于 26-40 位和 44-77 位上, 在 I-CreI 中则分别位于 28-43 位和 47-83 位。

根据所述专性异二聚体大范围核酸酶的一个优选实施方案, 所述第一单体和所述第二单体分别具有下述突变对其中之一:

- a) 8 位上的谷氨酸被碱性氨基酸替换 (单体 A), 7 位上的赖氨酸被酸性氨基酸替换 (单体 B),
- b) 61 位上的谷氨酸被碱性氨基酸替换 (单体 A), 96 位上的赖氨酸被酸性氨基酸替换 (单体 B),
- c) 97 位上的亮氨酸被芳香族氨基酸替换 (单体 A), 54 位上的苯丙氨酸被小氨基酸替换 (单体 B), 以及
- d) 137 位上的天冬氨酸被碱性氨基酸替换 (单体 A), 51 位上的精氨酸被酸性氨基酸替换 (单体 B),

所述位置参照 I-CreI 氨基酸序列 SEQ ID NO: 1 来表示。

根据所述专性异二聚体大范围核酸酶的一个更优选的实施方案, 如 a) 或 b) 中所定义, 所述单体的 8 位或 61 位上的谷氨酸被碱性氨基酸替换, 该单体还包含其 7 位和 96 位上的赖氨酸残基中的至少一个被精

氨酸替换。

根据所述专性异二聚体大范围核酸酶的另一个更优选的实施方案，如 c) 中所定义的 97 位上的亮氨酸被芳香族氨基酸替换的单体还包含 54 位上的苯丙氨酸被色氨酸替换。

根据所述专性异二聚体大范围核酸酶的另一个更优选的实施方案，如 c) 中所定义的 54 位上的苯丙氨酸被小氨基酸替换的单体还包含 58 位上的亮氨酸或 57 位上的赖氨酸被甲硫氨酸替换。

根据所述专性异二聚体大范围核酸酶的另一个更优选的实施方案，所述酸性氨基酸是谷氨酸。

根据所述专性异二聚体大范围核酸酶的另一个更优选的实施方案，所述碱性氨基酸是精氨酸。

根据所述专性异二聚体大范围核酸酶的另一个更优选的实施方案，所述芳香族氨基酸是苯丙氨酸。

根据所述专性异二聚体大范围核酸酶的另一个更优选的实施方案，所述小氨基酸是甘氨酸。

根据所述专性异二聚体大范围核酸酶的另一个更优选的实施方案，其由具有突变 D137R 的第一单体(A)和具有突变 R51D 的第二单体(B)组成。

根据所述专性异二聚体大范围核酸酶的另一个更优选的实施方案，其包含上文 a)、b)、c)或 d)中所定义的至少两对突变，所述突变对其中之一有利地如 c)或 d)中所述。优选地，一个单体的 7 位和 96 位上的赖氨酸残基被替换为酸性氨基酸，另一个单体的 8 位和 61 位上的谷氨酸残基被替换为碱性氨基酸。

更优选地，所述专性异二聚体大范围核酸酶包含如上文 a)、b)和 c)中所定义的 3 对突变。所述专性异二聚体大范围核酸酶有利地由第一单体(A)和第二单体(B)组成，所述第一单体(A)具有至少选自以下的突变：(i) E8R、E8K 或 E8H，E61R、E61K 或 E61H 以及 L97F、L97W 或 L97Y；(ii) K7R、E8R、E61R、K96R 以及 L97F，或者(iii) K7R、

E8R、F54W、E61R、K96R 以及 L97F; 所述第二单体 (B) 具有至少选自以下的突变: (iv) K7E 或 K7D, F54G 或 F54A 以及 K96D 或 K96E; (v) K7E、F54G、L58M 以及 K96E, 或者 (vi) K7E、F54G、K57M 以及 K96E。

根据本发明的专性异二聚体大范围核酸酶来源于野生型同二聚化 LAGLIDADG 归巢内切核酸酶或其功能性变体。野生型同二聚化 LAGLIDADG 归巢内切核酸酶的实例显示在 Lucas 等, *Nucleic Acids Res.*, 2001, 29, 960-969 的表 1 中。野生型同二聚化 LAGLIDADG 归巢内切核酸酶可有利地选自 I-CreI、I-CeuI、I-MsoI 以及 I-CpaI, 优选 I-CreI。

单体 A 和 B 由于二聚化界面外部的一个或多个突变而与野生型单体不同。该额外突变有利地在与 DNA 靶标半位点相互作用的氨基酸残基的位置上。LAGLIDADG 归巢内切核酸酶 DNA 相互作用残基是本领域中众所周知的 (Jurica 等, *Molecular Cell.*, 1998, 2, 469-476; Chevalier 等, *J. Mol. Biol.*, 2003, 329, 253-269)。突变的残基可直接或通过水分子与 DNA 骨架或核苷酸碱基相互作用。优选地, 所述突变修饰大范围核酸酶的切割特异性, 并产生具有新的特异性的大范围核酸酶, 其能切割来自目的基因的 DNA 靶标。更优选地, 所述突变是对应于位于 I-CreI 氨基酸序列的 26-40 位的第一功能性亚结构域中的一个或多个氨基酸的替换, 其改变对 DNA 靶标中 $\pm 8-10$ 位核苷酸的特异性, 和/或所述突变是对应于位于 I-CreI 氨基酸序列的 44-77 位的第二功能性亚结构域中的替换, 其改变对 DNA 靶标中 ± 3 至 5 位核苷酸的特异性, 如前文所述 (PCT 国际申请 WO 2006/097784、WO 2006/097853、WO 2007/060495、WO 2007/049156、WO 2007/049095 以及 WO 2007/057781; Arnould 等, *J. Mol. Biol.*, 2006, 355, 443-458; Smith 等, *Nucleic Acids Res.*, 2006, 34, e149)。所述替换有利地对应于 I-CreI 氨基酸序列的 26、28、30、32、33、38 和/或 40、44、68、70、75 和/或 77 位。为了切割其中 n_{-4} 是 t 或者 n_{+4} 是 a 的 DNA 靶标, 所述变体在 44 位有利地是谷氨酰胺; 为了切割其中 n_{-4} 是 a 或者 n_{+4} 是 t 的 DNA 靶标, 所述变体在 44 位有利地是丙氨酸 (A) 或天冬酰胺; 为了切割其中 n_{-9} 是 g 或者 n_{+9} 是 c 的 DNA 靶标, 所述变体在 38 位有利地是精氨酸 (R) 或赖氨酸 (K)。

根据所述专性异二聚体大范围核酸酶的另一个有利的实施方案,所述单体 A 和 B 是不同的 I-CreI 单体变体,优选在 I-CreI 的 26-40 位和/或 44-77 位具有突变的变体;由这两个单体组成的专性异二聚体大范围核酸酶 AB 能切割非回文 DNA 靶标,其中所述 DNA 靶标的至少+3 至+5 位、+8 至+10 位、-10 至-8 位以及-5 至-3 位上的核苷酸对应于目的基因的 DNA 靶标的+3 至+5 位、+8 至+10 位、-10 至-8 位以及-5 至-3 位上的核苷酸。优选地,所述异二聚体的两个单体都在 26-40 位和/或 44-77 位具有突变。更优选地,两个单体都在 I-CreI 的 26-40 位和/或 44-77 位具有不同的突变。

单体 A 和 B 在与 DNA 靶序列相互作用的其它位置上可包含一个或多个突变。具体而言,其它替换可以在与磷酸骨架相接触的位置上引入,例如在最后的 C-末端环(137-143 位;Prieto 等, *Nucleic Acids Res.*, Epub 22 April 2007)。优选地,所述残基涉及所述 DNA 切割位点的结合与切割。更优选地,所述残基在 I-CreI 的 138、139、142 或 143 位上。两个残基可在一个结构域中突变,只要每个突变在选自 138 位和 139 位的残基对以及 142 位和 143 位的残基对中的不同残基对中即可。所引入的突变修饰最后一个 C 端环的氨基酸与 I-CreI 位点的磷酸骨架之间的相互作用。优选地,138 位或 139 位上的残基被替换为疏水性氨基酸以避免与 DNA 切割位点的磷酸骨架形成氢键。例如,138 位上的残基被替换为丙氨酸或者 139 位上的残基被替换为甲硫氨酸。142 位或 143 位上的残基有利地被替换为小氨基酸(例如甘氨酸)以降低这些氨基酸残基侧链的大小。更优选地,最后一个 C-端环中的替换改变了所述专性异二聚体大范围核酸酶对 I-CreI 位点的 ± 1 至 2、 ± 6 至 7 和/或 ± 11 至 12 位的核苷酸的特异性。

此外,单体 A 和 B 可包含改善所述专性异二聚体大范围核酸酶对目的基因之 DNA 靶序列的结合和/或切割性质的一个或多个其它突变。突变的其它残基可以在单体的整个序列上。突变的实例包括下述突变,参照 I-CreI 氨基酸序列: I24V, R70S, 75 位天冬氨酸的突变为不带电荷的氨基酸,优选天冬酰胺(D75N)或缬氨酸(D75V),以及在单体序列的 C 端一半中的替换,优选在以下位置的替换: 80、82、85、86、87、94、96、100、103、114、115、117、125、129、131、132、147、151、153、154、155、157、159 和 160 位。

在所述专性异二聚体大范围核酸酶的另一个有利的实施方案中，所述突变是原始氨基酸被替换为选自以下的氨基酸：A、D、E、G、H、K、N、P、Q、R、S、T、Y、C、V、L和W。

此外，可将一个或多个残基插入到单体的NH₂端和/或COOH端。例如，将甲硫氨酸残基引入到NH₂端，将标签（表位或多聚组氨酸序列）引入到NH₂端和/或COOH端；所述标签可用于检测和/或纯化所述异二聚体。

本发明还涉及上文所述的专性异二聚体大范围核酸酶的单体A或B。

本发明还涉及一种单链大范围核酸酶（融合蛋白），其包含通过肽连接物（peptidic linker）连接的上文所述的单体A和B。

本发明的主题还包括编码如上文所定义的专性异二聚体大范围核酸酶的至少一个单体或其单链衍生物的多核苷酸片段。

本发明的主题还包括一种重组载体，其包含如上文所定义的至少一种多核苷酸片段。所述载体可有利地包含两种不同的多核苷酸片段，所述多核苷酸片段各自编码本发明的专性异二聚体大范围核酸酶的单体之一。

可在本发明中使用的载体包括但不限于病毒载体、质粒、RNA载体，或者线性或环形DNA或RNA分子，可由染色体、非染色体、半合成或合成的核酸所组成。优选的载体是能够自主复制与之连接的核酸的载体（附加型载体）和/或表达与之连接的核酸的载体（表达载体）。大量的合适载体是本领域技术人员已知的，并可以商品获得。

病毒载体包括逆转录病毒、腺病毒、细小病毒(parvovirus)（例如腺相关病毒）、冠状病毒、负链RNA病毒如正粘病毒(orthomyxovirus)（例如流感病毒）、杆状病毒（例如狂犬病毒和疱疹性口炎病毒(vesicular stomatitis virus)）、副粘病毒(paramyxovirus)（例如麻疹病毒和仙台病毒(Sendai)）、正链RNA病毒如小RNA病毒(picornavirus)和α病毒，以及双链DNA病毒，其包括腺病毒、疱疹病毒（例如1和2型单纯疱疹病毒、EB病毒、巨细胞病毒）和痘病毒（例如牛痘病毒、禽痘病毒

和金丝雀痘病毒)。其他病毒包括例如诺瓦克病毒、披膜病毒(togavirus)、黄病毒(flavivirus)、呼肠病毒(reoviruses)、乳多空病毒(papovavirus)、嗜肝 DNA 病毒(hepadnavirus)和肝炎病毒。逆转录病毒的实例包括:禽白血病-肉瘤、哺乳动物 C 型、B 型病毒、D 型病毒、HTLV-BLV 组、慢病毒、泡沫病毒(spumavirus) (Coffin, J. M., Retroviridae: The viruses and their replication, In Fundamental Virology, 第三版, B. N. Fields, 等编辑, Lippincott-Raven Publishers, Philadelphia, 1996)。

优选的载体包括慢病毒载体,尤其是自失活(self inactivating)的慢病毒载体。

载体可包含选择标记,例如:用于真核细胞培养的新霉素磷酸转移酶、组氨醇脱氢酶、二氢叶酸还原酶、潮霉素磷酸转移酶、单纯疱疹病毒胸苷激酶、腺苷脱氨酶、谷氨酰胺合成酶和次黄嘌呤-鸟嘌呤磷酸核糖基转移酶;用于酿酒酵母(*S. Cerevisiae*)的 TRP1;大肠杆菌(*E. coli*)中的四环素、利福平或氨苄青霉素抗性。

优选地,所述载体是表达载体,其中编码本发明专性异二聚体大范围核酸酶/单链衍生物的序列置于适当的转录和翻译控制元件的控制下,以允许产生或合成所述大范围核酸酶。因此,所述多核苷酸包含在表达盒中。更具体地,载体包含复制起点、与所述编码多核苷酸有效连接的启动子、核糖体结合位点、RNA 剪接位点(使用基因组 DNA 时)、多腺苷酸化位点和转录终止位点。其还可包含增强子。启动子的选择将取决于其中表达多肽的细胞。优选地,编码所述专性异二聚体大范围核酸酶的各个单体的两种多核苷酸包含在一个载体中,所述载体能够驱动两种多核苷酸同时表达。合适的启动子包括组织特异性和/或诱导型启动子。诱导型启动子的实例是:由提高重金属水平诱导的真核金属硫蛋白(metallothionein)启动子,应答于异丙基- β -D-硫代半乳糖苷(IPTG)诱导的原核 lacZ 启动子,和由提高温度诱导的真核热休克启动子。组织特异性启动子的实例是骨骼肌肌酸激酶、前列腺特异性抗原(PSA)、 α -抗胰蛋白酶、人表面活性剂(SP)A 和 B 蛋白、 β -酪蛋白和酸性乳清蛋白基因。

根据所述载体的另一个有利的实施方案,其包括靶向 DNA 构建体,所述靶向 DNA 构建体含有与目的基因组位点周围区域有同源性的序列,所述目的基因组位点含有如上文所定义的杂合 DNA 靶序列。

更优选地，所述靶向 DNA 构建体包含：

- a) 与含有目的基因组位点周围区域有同源性的序列，其中所述目的基因组位点含有如上文所定义的杂合 DNA 靶序列，以及
- b) 两侧是 a) 中序列的待引入序列。

优选使用至少 50 bp、优选大于 100 bp、更优选大于 200 bp 的同源序列。事实上，共有的 DNA 同源性位于断裂位点上游和下游的两侧区域，且待引入 DNA 序列应位于两臂之间。待引入序列包含外源目的基因或者用于使基因或其一部分失活或缺失的序列。

本发明还涉及经上文所定义多核苷酸或载体（优选表达载体）修饰的原核或真核宿主细胞。

本发明还涉及特征在于其所有或部分细胞经上文所定义多核苷酸或载体修饰的非人转基因动物或转基因植物。

本文所用的细胞指原核细胞（例如细菌细胞）或真核细胞（例如动物细胞、植物细胞或酵母细胞）。

本发明的主题还包括上文所定义的专性异二聚体大范围核酸酶/单链大范围核酸酶衍生物、一种或两种多核苷酸（优选包含在表达载体中）、细胞、转基因植物、非人转基因哺乳动物用于分子生物学、体内或体外基因工程、以及体内或体外基因组工程中的非治疗目的用途。

非治疗目的包括例如 (i) 用于蛋白质生产的包装细胞系中特异性基因座的基因打靶；(ii) 用于植株改善和代谢工程的谷类植物中特异性基因座的基因打靶；(iii) 用于除去基因修饰的谷类植物中的标志物的靶向重组，(iv) 用于除去基因修饰的微生物菌株中的标志物的靶向重组（例如用于抗生素生产）。

根据所述用途的一个有利的实施方案，其用于诱导含有杂交 DNA 靶序列之目的位点中的双链断裂，从而诱导 DNA 重组事件、DNA 丢失或细胞死亡。

根据本发明，所述双链断裂是用于：修复特定序列、修饰特定序列、

恢复功能性基因而替换突变基因，弱化或活化内源目的基因、将突变引入到目的位点中、引入外源基因或其部分、失活或消灭内源基因或其部分、使染色体臂易位或者使 DNA 不被修复并降解。

本发明的主题还涉及基因工程方法，其特征在于包括以下步骤：使包含上文所定义的杂合 DNA 靶标的载体上的目的位点中的双链核酸断裂，所述断裂通过使所述载体与上文所定义的专性异二聚体大范围核酸酶/单链大范围核酸酶衍生物相接触而实现，从而诱导与另一载体进行同源重组，所述另一载体与所述专性异二聚体大范围核酸酶的切割位点周围的序列有同源性。

本发明的主题还涉及基因组工程方法，其特征在于包括以下步骤：
1) 使基因座处发生双链断裂，所述断裂通过使杂交 DNA 靶标与所述大范围核酸酶相接触而实现，其中所述基因座含有上文所定义的至少一种杂合 DNA 靶标和专性异二聚体大范围核酸酶/单链大范围核酸酶衍生物；2) 将所述断裂的基因组基因座维持在适于与靶向 DNA 构建体进行同源重组的条件下，所述构建体包含待引入到所述基因座中的序列，其两侧是与所靶向基因座有同源性的序列。

本发明的主题还涉及基因组工程方法，其特征在于包括以下步骤：
1) 使基因座处发生双链断裂，所述断裂通过使切割位点与上文所定义的所述大范围核酸酶相接触而实现，其中所述基因座含有上文所定义的至少一种杂合 DNA 靶标和专性异二聚体大范围核酸酶/单链大范围核酸酶衍生物；2) 将所述断裂的基因组基因座维持在适于与和所述切割位点周围区域有同源性的染色体 DNA 进行同源重组的条件下。

本发明的主题还包括上文所定义的至少一种专性异二聚体大范围核酸酶/单链大范围核酸酶衍生物、一种或两种多核苷酸（优选包含在表达载体中）在制备用于在有此需求的个体中预防、改善或治疗遗传性疾病的药物中的用途，所述药物可以任意方式施用给所述个体。

本发明的主题还包括在有此需求的个体中预防、改善或治疗遗传性疾病的方法，所述方法至少包括以下步骤：以任意方式将上文所定义的组合物施用给所述个体。

在该情况下,如上文所定义的专性异二聚体大范围核酸酶/单链大范围核酸酶衍生物的使用至少包括以下步骤:(a)在个体的体细胞组织中,在包含所述大范围核酸酶的至少一个识别和切割位点的基因之目的位点处诱导双链切割,和(b)向该个体中引入靶向 DNA,其中所述靶向 DNA 包含(1)与切割位点周围区域有同源性的 DNA,和(2)在该靶向 DNA 与染色体 DNA 之间发生重组时修复目的位点的 DNA。在适于将靶向 DNA 引入目的位点的条件下将靶向 DNA 引入所述个体。

根据本发明,所述双链切割通过对个体施用所述大范围核酸酶而整体(*in toto*)诱导,或者通过将所述大范围核酸酶引入到取自个体的体细胞内并在修饰后放回所述个体内而进行离体诱导。

在所述用途的一个优选实施方案中,将所述大范围核酸酶与靶向 DNA 构建体组合,所述靶向 DNA 构建体包含修复基因中突变的序列,其两侧是与所述大范围核酸酶的基因组 DNA 切割位点周围的基因区域有同源性的序列,如上文所述。所述修复突变的序列是具有正确序列的基因片段或者是外显子敲入构建体。

就矫正基因而言,基因的切割发生在突变附近,优选在突变的 500bp 范围内。靶向构建体包含基因片段,其含有基因组 DNA 切割位点两侧至少 200 bp 的同源序列(最小修复基质)用于修复切割,并含有用于修复突变的正确基因序列。因此,用于基因矫正的靶向构建体包含最小修复基质或由其组成;其优选为 200 pb 到 6000 pb,更优选为 1000 pb 到 2000 pb。

为了恢复功能基因,基因的切割发生在突变的上游。优选所述突变是该基因序列中第一个已知的突变,从而该基因中所有下游突变都可同时被矫正。所述靶向构建体包含基因组 DNA 切割位点下游的外显子,它们在框内融合(如同在 cDNA 中一样),并在 3'端具有多聚腺苷酸化位点以终止转录。待引入序列(外显子敲入构建体)的两侧是切割位点周围的内含子或外显子序列,从而允许所改造基因(外显子敲入基因)转录成为能编码功能性蛋白质的 mRNA。例如,外显子敲入构建体的两侧是上游和下游的序列。

本发明的主题还包括上文所定义的专性异二聚体大范围核酸酶/单链大范围核酸酶衍生物、一种或两种多核苷酸(优选包含在表达载体中)在制备用于在有此需求的个体中预防、改善或治疗由具有 DNA 中

间体的传染原引起的疾病的药物中的用途，所述药物通过任意方式施用给所述个体。

本发明的主题还包括在有此需求的个体中预防、改善或治疗由具有 DNA 中间体的传染原引起的疾病的方法，所述方法至少包括以下步骤：通过任意方式将上文所定义的组合物施用给所述个体。

本发明的主题还包括上文所定义的专性异二聚体大范围核酸酶/单链大范围核酸酶衍生物、一种或两种多核苷酸（优选包含在表达载体中）在体外用于抑制传染原的传播、灭活或消灭传染原的用途，或者用于物品消毒的用途，其中所述传染原具有 DNA 中间体，存在于生物来源产品或旨在用于生物用途的产品中。

本发明的主题还包括给产品或物质消毒以除去其中具有 DNA 中间体之传染原的方法，所述方法至少包括以下步骤：将生物来源产品、旨在用于生物用途的产品或物品与上文所定义的组合物接触一定时间，所述时间足以抑制所述传染原的传播、灭活或消灭所述传染原。

在一个具体的实施方案中，所述传染原是病毒。例如，所述病毒是腺病毒（Ad11，Ad21）、疱疹病毒（HSV；VZV；EBV；CMV；6、7 或 8 型疱疹病毒）、嗜肝 DNA 病毒（HBV）、乳头状多瘤空泡病毒（HPV）、痘病毒或者逆转录病毒（HTLV、HIV）。

本发明的主题还包括组合物，其特征在于至少包含上述专性异二聚体大范围核酸酶、一种或两种核苷酸（优选包含在表达载体中），如上文所述。

在所述组合物的一个优选实施方案中，其包含靶向 DNA 构建体，所述构建体包含修复目的位点的序列，其两侧是与如上所述的靶标基因座有同源性的序列。优选地，所述靶向 DNA 构建体被包含在重组载体中或被包含在表达载体中，所述载体包含编码所述大范围核酸酶的多核苷酸，如本发明中所定义。

本发明的主题还包括产品，其至少含有大范围核酸酶或编码所述大范围核酸酶的一种或两种表达载体及上述包含靶向构建体的载体，所述产品作为组合制剂同时、分别或依次使用，用于预防或治疗遗传性疾病。

对治疗的目的而言,以治疗有效量施用所述大范围核酸酶和可药用赋形剂。如果所施用的量是有生理意义的,则这样的组合被称作以“治疗有效量”施用。如果一种药剂的存在导致接受者生理状况可检测的变化,则它是有生理意义的。在本文中,如果一种药剂的存在导致目标疾病的一个或多个症状的严重程度减轻以及对损伤或异常的基因组矫正,则它是有生理意义的。

在本发明用途的一个实施方案中,所述大范围核酸酶基本上是非免疫原性的,即不引发或几乎不引发不利的免疫应答。根据本发明,可使用大量改善或消除这类有害免疫反应的方法。在一个优选的实施方案中,所述大范围核酸酶基本不含 N-甲酰基甲硫氨酸。避免不想要的免疫反应的另一种方法是将大范围核酸酶与聚乙二醇("PEG")或聚丙二醇("PPG") (优选 500 到 20,000 道尔顿的平均分子量(MW))缀合。与 PEG 或 PPG 的缀合(如 Davis 等 (US 4,179,337)所述)例如可提供具有抗病毒活性的非免疫原性生理活性水溶性内切核酸酶缀合物。Saifer 等 (US 5,006,333)中还描述了使用聚乙二醇-聚丙二醇共聚物的类似方法。

大范围核酸酶可作为多肽或者作为编码所述多肽的多核苷酸构建体/载体使用。通过本领域技术人员众所周知的任何适用于特定细胞类型的便利手段而在体外、离体或在体内将其(单独或者与至少一种合适的载体或运载体组合和/或与靶向 DNA 组合)引入细胞中。一旦进入细胞内,所述大范围核酸酶和包含靶标 DNA 和/或编码大范围核酸酶之核酸的载体(如果存在的话)被细胞从胞质输入或转移至核内的作用位点。

专性异二聚体大范围核酸酶/单链大范围核酸酶衍生物(多肽)可有利地与下述物质结合:脂质体、聚乙烯亚胺(PEI)、和/或膜易位肽(Bonetta, The Scientist, 2002, 16, 38; Ford 等, Gene Ther., 2001, 8, 1-4; Wadia 和 Dowdy, Curr. Opin. Biotechnol., 2002, 13, 52-56);在后一种情况下,所述大范围核酸酶的序列与膜易位肽的序列融合(融合蛋白)。

可通过大量方法(例如注射、直接摄入、射弹轰击、脂质体、电穿孔)将包含靶标 DNA 和/或编码大范围核酸酶之核酸的载体引入细胞中。可使用表达载体在细胞中稳定或瞬时地表达大范围核酸酶。在真核细胞中进行表达的技术是本领域技术人员公知的(见 Current Protocols in Human Genetics: 第 12 章 "Vectors For Gene Therapy" & 第 13 章 "Delivery Systems for Gene Therapy")。任选地,可优选在重组蛋白中掺入核定位信

号以确保其在核内表达。

专性异二聚体大范围核酸酶/单链大范围核酸酶衍生物的运用以及利用本发明的所述大范围核酸酶的方法还包括利用编码所述专性异二聚体大范围核酸酶的多核苷酸、载体、细胞、转基因植物或非人转基因哺乳动物，如上文所述。

根据本发明的用途和方法的另一个有利的实施方案，所述专性异二聚体大范围核酸酶/单链大范围核酸酶衍生物、多核苷酸、载体、细胞、转基因植物或非人转基因哺乳动物与上文所定义的靶向 DNA 构建体相关联。优选地，所述编码专性异二聚体大范围核酸酶/单链大范围核酸酶衍生物的单体的载体包含如上文所定义的靶向 DNA 构建体。

根据本领域中众所周知并且可以商购的标准位点定向突变法，本发明的专性异二聚体大范围核酸酶来源于母体同二聚化 LAGLIDADG 内切核酸酶的单体。根据众所周知的重叠 PCR 技术，通过扩增含有如上文所述的每个突变位置的重叠片段可有利地制备它们。

这可以通过利用互补引物组来实现。例如以下三对：SEQ ID NO: 3 与 SEQ ID NO: 4、SEQ ID NO: 5 与 SEQ ID NO: 6、以及 SEQ ID NO: 7 或 SEQ ID NO: 35 与 SEQ ID NO: 8 可用于扩增单体 A 的编码序列 (CDS)，以下三对：SEQ ID NO: 9 与 SEQ ID NO: 10、SEQ ID NO: 11 与 SEQ ID NO: 12、SEQ ID NO: 13 或 SEQ ID NO: 36 与 SEQ ID NO: 14 可用于扩增单体 B 的 CDS。

或者，以下两对：SEQ ID NO: 37 与 SEQ ID NO: 41、SEQ ID NO: 40 与 SEQ ID NO: 42 可用于扩增单体 A 的 CDS，以下两对：SEQ ID NO: 37 与 SEQ ID NO: 38、SEQ ID NO: 39 与 SEQ ID NO: 40 可用于扩增单体 B 的 CDS。

可通过改造能够切割目的基因组 DNA 靶序列之变体的方法来获得母体同二聚化 LAGLIDADG 内切核酸酶的单体，如之前 Smith 等 *Nucleic Acids Res.*, 2006, 34 e149 中所述，所述方法至少包括以下步骤：

(a) 构建第一系列 I-CreI 变体，其在 LAGLIDADG 核心结构域的第一功能性亚结构域中含有至少一个替换，所述第一功能性亚结构域位于 I-CreI 的 26 到 40 位，

(b) 构建第二系列 I-CreI 变体，其在 LAGLIDADG 核心结构域的第二功能

性亚结构域中具有至少一个替换,所述第二功能性亚结构域位于 I-CreI 的 44 到 77 位,

(c)从步骤(a)的第一系列中选择和/或筛选能够切割突变体 I-CreI 位点的变体,在所述位点中(i) I-CreI 位点中-10 到-8 位的核苷酸三联体已被存在于所述基因组靶标中-10 到-8 位的核苷酸三联体所替换,且(ii)+8 到+10 位的核苷酸三联体已被存在于所述基因组靶标的-10 到-8 位的核苷酸三联体的反向互补序列所替换,

(d)从步骤(b)的第二系列中选择和/或筛选能够切割突变体 I-CreI 位点的变体,在所述位点中(i) I-CreI 位点中-5 到-3 位的核苷酸三联体已被存在于所述基因组靶标中-5 到-3 位的核苷酸三联体所替换,且(ii) +3 到+5 位的核苷酸三联体已被存在于所述基因组靶标中-5 到-3 位的核苷酸三联体的反向互补序列所替换,

(e)从步骤(a)的第一系列中选择和/或筛选能够切割突变体 I-CreI 位点的变体,在所述位点中(i) I-CreI 位点中+8 到+10 位的核苷酸三联体已被存在于所述基因组靶标中+8 到+10 位的核苷酸三联体所替换,且(ii) -10 到-8 位的核苷酸三联体已被存在于所述基因组靶标中+8 到+10 位的核苷酸三联体的反向互补序列所替换,

(f)从步骤(b)的第二系列中选择和/或筛选能够切割突变体 I-CreI 位点的变体,在所述位点中(i) I-CreI 位点中+3 到+5 位的核苷酸三联体已被存在于所述基因组靶标中+3 到+5 位的核苷酸三联体所替换,且(ii) -5 到-3 位的核苷酸三联体已被存在于所述基因组靶标中+3 到+5 位的核苷酸三联体的反向互补序列所替换,

(g)将来自步骤(c)和步骤(d)的两个变体中 26 到 40 位及 44 到 77 位中的突变在单个变体中组合,以获得切割下述序列的新的同二聚体 I-CreI 变体,所述序列中(i)-10 到-8 位的核苷酸三联体与存在于所述基因组靶标中-10 到-8 位的核苷酸三联体相同,(ii) +8 到+10 位的核苷酸三联体与存在于所述基因组靶标中-10 到-8 位的核苷酸三联体的反向互补序列相同,(iii) -5 到-3 位的核苷酸三联体与存在于所述基因组靶标中-5 到-3 位的核苷酸三联体相同,(iv) +3 到+5 位的核苷酸三联体与存在于所述基因组靶标中-5 到-3 位的核苷酸三联体的反向互补序列相同,

(h)将来自步骤(e)和步骤(f)的两个变体中 26 到 40 位及 44 到 77 位中的突变

在单个变体中组合,以获得切割下述序列的新的同二聚体 I-CreI 变体,所述序列中(i)+3 到+5 位的核苷酸三联体与存在于所述基因组靶标中+3 到+5 位的核苷酸三联体相同,(ii) -5 到-3 位的核苷酸三联体与存在于所述基因组靶标中+3 到+5 位的核苷酸三联体的反向互补序列相同,(iii)该 I-CreI 位点中+8 到+10 位的核苷酸三联体已被存在于所述基因组靶标中+8 到+10 位的核苷酸三联体所替换,和(iv) -10 到-8 位的核苷酸三联体与存在于所述基因组靶标中+8 到+10 位的核苷酸三联体的反向互补序列相同,

- (i) 将步骤 (g)与步骤(h)中获得的变体相组合以形成异二聚体,以及
- (j) 选择和/或筛选来自步骤(i)的能够切割位于哺乳动物基因中的所述基因组 DNA 靶标的异二聚体。

步骤 (a) 和 (b) 可包括引入其它突变以提高突变体的结合和/或切割性能,尤其是在与 DNA 靶序列接触或者与所述 DNA 靶标直接或间接相互作用的其它位置上。这些步骤可通过制备组合文库来实施,如 PCT 国际申请 WO 2004/067736 和 Arnould 等 (J. Mol. Biol., 2006, 355, 443-458)中所述。

可通过使用体外或体内切割测定来进行步骤(c)、(d)、(e)、(f)和/或(j)中的选择和/或筛选,如 PCT 国际申请 WO 2004/067736、Epinat 等(Nucleic Acids Res., 2003, 31, 2952-2962)、Chames 等(Nucleic Acids Res., 2005, 33, e178)和 Arnould 等(J. Mol. Biol., 2006, 355, 443-458)中所述。优选地,通过重组介导的所述 DNA 双链断裂的修复在下述条件下在体内进行步骤(c)、(d)、(e)、(f) 和/或(j),在所述条件下,由所述变体产生的突变 DNA 靶序列中的双链断裂导致阳性选择标志物或报告基因的激活或者阴性选择标志物或报告基因的失活,如在 PCT 国际申请 WO 2004/067736、Epinat 等(Nucleic Acids Res., 2003, 31, 2952-2962)、Chames 等(Nucleic Acids Res., 2005, 33, e178)和 Arnould 等(J. Mol. Biol., 2006, 355, 443-458)中所述。

可根据公知的重叠 PCR 技术,通过扩增包含两个亚结构域中之每一个的重叠片段来进行步骤(g)和(h)中的(分子内)突变组合。

另外,步骤(g)和/或(h)可进一步包括在整个变体或部分变体(特别是所述变体的 C 端一半(80 到 163 位))上引入随机突变。可根据本领域公知的并可以商品获得的标准突变方法,通过对变体池(pool of variants)产生随机突变文库来进行突变。

步骤 (i) 中的 (分子内) 变体组合是通过共表达步骤 (g) 的一种变体和步骤 (h) 的一种变体从而允许形成异二聚体来进行的。例如, 可通过编码所述变体的一种或两种重组表达载体修饰宿主细胞。然后在允许表达所述变体的条件下培养所述细胞, 从而在所述宿主细胞中形成异二聚体, 如先前在 PCT 国际申请 WO 2006/097854 和 Arnould 等(J. Mol. Biol., 2006, 355, 443- 458)中所述的。

在此情形下, 将本发明中所定义的单体 A 的突变引入到步骤 (j) 中得到的异二聚体的一个单体中, 将本发明中所定义的单体 B 的突变引入到所述异二聚体的另一个单体中。

或者, 可通过引入下述修饰由源自上述改造大范围核酸酶变体之方法的方法获得本发明的专性异二聚体大范围核酸酶:

-在两种类型的原始骨架蛋白 (scaffold protein) 上进行步骤 (a) 和步骤 (b): 具有单体 A 的突变的第一 I-CreI 骨架和具有单体 B 的突变的第二 I-CreI 骨架, 以及

-通过以下步骤进行步骤 (c) - (f) 的选择和/或筛选: 将具有上文所述的单体 A 或 B 的突变的变体文库转化到表达含有相应突变 (分别来自单体 B 或 A) 的 I-CreI 突变体的宿主细胞中, 以允许形成异二聚体, 利用其中 I-CreI 位点的一半在 ± 3 至 5 位或 ± 8 至 10 位被修饰而另一半未被修饰的非回文 DNA 靶标选择功能性异二聚体变体。

通过将来自相同单体 (A) 或 (B) 的两种变体的突变组合在一个单体中来进行步骤 (g) 和 (h)。

通过将步骤 (g) 中所得的来自单体 (A 或 B) 之一的变体与步骤 (h) 中所得的来自另一个单体的变体相组合来进行步骤 (i)。

根据本领域众所周知的方法从本发明的专性异二聚体大范围核酸酶得到能切割目的基因之 DNA 靶标的单链大范围核酸酶 (Epinat 等, Nucleic Acids Res., 2003, 31, 2952-62; Chevalier 等, Mol. Cell., 2002, 10, 895-905; Steuer 等, Chembiochem., 2004, 5, 206-13; PCT 国际申请 WO 03/078619 以及 WO 2004/031346)。任意这些方法均可用于构建本发明的单链大范围核酸酶。

本领域技术人员可通过任何已知方法制备本发明中所定义的编码单体 A 和 B 的多核苷酸序列。例如，它们由 cDNA 模板通过与特定引物的聚合酶链反应进行扩增。优选地，选择所述 cDNA 的密码子以有利于所述蛋白质在所期望的表达体系中表达。

可通过众所周知的重组 DNA 和基因工程技术得到含有所述多核苷酸的重组载体并将其引入到宿主细胞中。

本发明的专性异二聚体大范围核酸酶是通过在适于上文所述的单体 A 和 B 共表达的条件下，在利用一种或两种表达载体修饰的转基因动物/植物或宿主细胞中共表达所述单体来制得的，并通过任何合适的方法从宿主细胞培养物或转基因动物/植物中回收所述专性异二聚体大范围核酸酶。

本发明的单链大范围核酸酶是通过在适于融合蛋白表达的条件下，在利用一种表达载体修饰的转基因动物/植物或宿主细胞中，融合含有上文所述的单体 A 和 B 的融合蛋白来制得的，并通过任何合适的方法从宿主细胞培养物或转基因动物/植物中回收所述单链大范围核酸酶。

本发明的主题还包括上文所述的至少一种专性异二聚体大范围核酸酶/单链大范围核酸酶衍生物作为用于制备其它大范围核酸酶的骨架的用途。例如，为了制备新的第三代归巢内切核酸酶，可在所述单体上进行第三轮诱变和选择/筛选。

除非另有说明，否则本发明的实施将使用本领域内的细胞生物学、细胞培养、分子生物学、转基因生物学、微生物学、重组 DNA 和免疫学的常规技术。这些技术在文献中有详尽解释。参阅，例如 *Current Protocols in Molecular Biology* (Frederick M. AUSUBEL, 2000, Wiley and son Inc, Library of Congress, USA); *Molecular Cloning: 实验室操作手册*, 第三版, (Sambrook 等, 2001, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press); *Oligonucleotide Synthesis* (M. J. Gait 编辑, 1984); Mullis 等 美国专利号 4,683,195; *Nucleic Acid Hybridization* (B. D. Harries & S. J. Higgins 编辑 1984); *Transcription And Translation* (B. D. Hames & S. J. Higgins 编辑 1984); *Culture Of Animal Cells* (R. I. Freshney, Alan R.

Liss, Inc., 1987); Immobilized Cells And Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide To Molecular Cloning (1984); the series, Methods In ENZYMOLOGY (主编 J. Abelson 和 M. Simon, Academic Press, Inc., New York), 尤其是第 154 和 155 卷 (Wu 等编辑)以及第 185 卷, "Gene Expression Technology" (D. Goeddel 编辑); Gene Transfer Vectors For Mammalian Cells (J. H. Miller 和 M. P. Calos 编辑, 1987, Cold Spring Harbor Laboratory); Immunochemical Methods In Cell And Molecular Biology (Mayer 和 Walker 编辑, Academic Press, London, 1987); Handbook Of Experimental Immunology, 第 I-IV 卷 (D. M. Weir 和 C. C. Blackwell 编辑, 1986);以及 Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986)。

除了上述特征以外, 本发明还包括在以下涉及说明以及附图中所体现的其他特征, 其中参考举例说明本发明专性异二聚体大范围核酸酶的实施例, 在附图中:

图 1 描绘了大范围核酸酶 I-Cre-I 与其靶标 DNA 模板的复合物 (PDB:1G9Y) 结构的侧视图。(A)、(B)、(C) 详细展示了同二聚体上两个单体间的三个可修饰的相互作用团块 (patch)。(D) 是所设计的异二聚化界面, 显示出蛋白质中每个相关位置上的氨基酸变化。

图 2 图示说明高盐浓度对切割特异性的影响。将 (A) QAN 或 (B) KTG 蛋白与 QAN 同二聚体位点 (GTT/AAC)、KTG 同二聚体 DNA 位点 (CCT/AGG) 或者杂合位点 QAN/KTG 位点 (GTT/AGG) 一起孵育, NaCl 浓度为 50-225 mM。箭头表示未切割的靶标 DNA (3.2 kb) 或两个消化所得条带 (1.1 和 2.1 kb)。星号 (*) 标记仅含有 DNA 的对照泳道。1 kb 和 100 bp 的梯带 (FERMENTAS) 分别用 M 和 M1 标记。

图 3 举例说明来自 I-CreI 变体 QAN 和 KTG 的突变体的 CDS, 其是为制备所述专性异二聚体大范围核酸酶而设计的。下划线表示三种诱变寡核苷酸的序列。QAN-A1 和 KTG-A1 突变体: 分别为寡核苷酸 SEQ ID NO: 3、35 和 5; QAN-A2 和 KTG-A2 突变体: 分别为寡核苷酸 SEQ ID NO: 3、7 和 5; QAN-B3 和 KTG-B3 突变体: 分别为寡核苷酸 SEQ ID NO: 9、11 和 13; QAN-B4 和 KTG-B4 突变体: 分别为寡核苷酸 SEQ ID NO: 9、11 和 36。

图 4 图示说明了所设计的大范围核酸酶的表达和纯化。箭头指示靶蛋白质带。(A) 野生型同二聚体和突变体单体。泳道: M=标准大范围标志物(BIORAD); 1.纯化的 QANwt; 2. 纯化的 KTGwt; 3. 沉淀物 QAN-A1; 4. 上清液 QAN-A1; 5. 纯化的 QAN-A1; 6. 沉淀物 KTG-B3; 7. 上清液 KTG-B3; 8. 纯化的 KTG-B3; 9. 沉淀物 KTG-A2; 10. 上清 KTG-A2; 11. 纯化的 KTG-A2; 12. 沉淀物 QAN-B3; 13. 上清液 QAN-B3; 14. 纯化的 QAN-B3; 15. 沉淀物 QAN-B4; 16. 上清液 QAN-B4; 17. 纯化的 QAN-B4。(B) KTG-A2/QAN-B3 的共表达和纯化。泳道: 1. 未诱导; 2. 诱导; 3. 沉淀物; 4. 上清液; 5. 透析之前纯化; 6. 透析之后纯化。(C)另外两种设计的共表达: 泳道 1. 可见两个条带, 对应于异二聚体 QAN-A1/KTG-B3; 泳道 2. 仅一个条带可见, 表明 QAN-B4/KTG-A2 不产生异二聚体。

图 5 图示说明单独表达的设计的大范围核酸酶单体变体在不同盐条件下的非特异性 DNA 切割和非切割。将大约 3.75 μ M 的每种纯化蛋白质与含有 QAN 同二聚体位点(GTT/AAC)、KTG 同二聚体 DNA 位点(CCT/AGG)或杂合位点 QAN/KTG 位点(Q-K: GTT/AGG)的 34 nM 纯化质粒(利用 *Xmn*I 预先线性化)一起孵育。NaCl 浓度为(A) 50 mM 或(B) 225 mM。箭头表示未切割的靶标 DNA (3.2kb) 或两个消化所得条带(1.1 和 2.1 kb)。星号(*)标记仅含有 DNA 的对照泳道。1kb 的梯带(FERMENTAS)用 M 进行标记。

图 6 图示说明不同的大范围核酸酶的分析性超速离心结果。当单独表达时, (A)野生型单体形成约 400 kDa 的同二聚体(KTG-wt; QAN-wt), (B) 所设计的非同二聚化 KTG-A2 和 QAN-B3 形成聚集体。(C) 共表达的 KTG-A2 和 QAN-B3 形成完美的异二聚体。(D) 共表达的 QAN-B4 和 KTG-A2 也在一定程度上形成异二聚体, 而 QAN-A1 和 KTG-B3 则不形成二聚体。

图 7 图示说明共表达的野生型(wt)和所设计的专性异二聚体 KTG-A2-QAN-B3 大范围核酸酶的特异性 DNA 切割。(A) 将纯化蛋白与含有 QAN 同二聚体位点(GTT/AAC)、KTG 同二聚体 DNA 位点(CCT/AGG)或杂合位点 QAN/KTG 位点(Q-K: GTT/AGG)的 3nM 纯化质粒(利用 *Xmn*I 预先线性化)一起孵育。因为不同的构建体具有不同的反应最适条件, 因此所述同二聚体以 0.25 μ M 浓度使用(4 h, 37°C), 而

异二聚体以 $0.50\ \mu\text{M}$ 浓度使用(30 分钟, 37°C)。NaCl 浓度为 $225\ \text{mM}$ 。箭头表示未切割的靶标 DNA (3.2kb) 或两个消化所得条带 (1.1 和 $2.1\ \text{kb}$)。1kb 的梯带 (Fermentas) 用 M 进行标记。(B) 将各酶样品在时间进程试验中对其最佳 DNA 位点的相对活性进行比较, 使用 $1\ \mu\text{M}$ 蛋白质和 $6\ \text{nM}$ 纯化的质粒靶标。白色星号标记估计最接近 50% 切割率的样品位置。

图 8 图示说明竞争分析, 以测定当暴露于 Q-K 异二聚体 DNA 位点 ($3.1\ \text{kb}$ PCR 产物) 和 KTG 同二聚体 DNA 位点 ($0.85\ \text{kb}$ PCR 产物) 的等摩尔混合物时 KTG-wt 同二聚体和 KTG-A2/QAN-B3 异二聚体蛋白的相对特异性。各靶标 DNA 以 $5\ \text{nM}$ 的最终浓度使用。不同的 DNA 靶标具有特征性大小并且给出特定性大小的切割产物。(A) 显示 KTG 和 KTG-A2/QAN-B3 酶的相对特异性和非特异性切割的时间进程试验。最终蛋白质浓度 = $1\ \mu\text{M}$ 。(B) 酶滴定分析, 以测定每一种酶对相应 (cognate) 和非相应 DNA 靶标位点实现 50% 切割时蛋白质浓度的差异。将等摩尔量的 DNA 靶标位点 PCR 产物 (Q-K 和 KTG) 相混合并与每种酶的系列稀释液一起孵育, 如所述, 孵育 1 小时。扫描凝胶并利用 ImageJ 和 Kaleidagraph 进行分析。

图 9 是人 RAG1 基因 (GenBank 登录号 NC_000011) 的示意图。框内是外显子序列, 并且示出了外显子-内含子连接。灰色框表示 ORF。RAG1.10 序列用其序列和位置表示。

图 10 表示被 I-CreI 或其一些衍生变体切割的 $22\ \text{bp}$ DNA 靶标 (分别为 SEQ ID NO: 58 至 65)。C 1221 是 I-CreI 靶标。10GTT_P、5CAG_P、10TGG_P 和 5GAG_P 是回文靶标, 其与 C 1221 不同之处在于框内的基序。RAG1.10 是 RAG1 靶标, RAG1.10.2 和 RAG1.10.3 是回文靶标, 其分别来自 RAG1.10 的左侧部分和右侧部分。如图中所示, 10GTT_P、5CAG_P、10TGG_P 和 5GAG_P 的框内基序存在于 RAG 1.10 靶标中。

图 11 表示大范围核酸酶 I-Cre-I 与 DNA 结合的复合物 (PDB: 1G9Y) 之结构的底视图。A. 显示出靶标 DNA 模板的复合物的视图。B. 示出了相同的视图但是省略了 DNA。对于每种单体而言, 两个残基 R51 和 D137 以棒条 (stick) 表示, 氢键以虚线表示。以虚线画出的圆圈圈定了两条 DNA 链被切割的活性位点。

图 12 图示说明了异二聚化组合突变体对 RAG 1.10 的切割。该图显示了利用 RAG1.10、RAG1.10.2 和 RAG1.10.3 靶标对 RAG 1.10.2 和 RAG 1.10.3 切割酶的突变体组合进行的次级筛选。试验形式是 2 点×2 点的形式。形成左栏的两点是突变体，形成右栏的两点是对照，用于评价实验质量。I-SceI 与 I-SceI 靶标在 a 和 d 中，I-SceI 的低活性形式在 b 中，空载体在 c 中。

图 13 表示 pCLS1088（用于哺乳动物细胞中表达大范围核酸酶的质粒）的图。

图 14 表示 pCLS1058 报告基因载体的图。利用杀稻瘟菌素和氨苄西林抗性基因标记报告基因载体。LacZ 串联重复序列具有 800 bp 的同源性，并且通过 1.3 kb 的 DNA 分隔开。它们被 EF1- α 启动子和终止子序列所环绕。靶位点是利用 Gateway 方案(INVITROGEN)克隆的，导致 CmR 和 ccdB 基因被所选的靶位点所替换。

图 15 图示说明了通过在 CHO 细胞中染色体外分析中监测到的同二聚体和异二聚体针对 3 种 RAG1.10 靶标的活性。A. 分别通过 M2、M2 K7E、M2 E8K 和 M3、M3 K7E、M3 E8K 同二聚体切割回文 RAG 1.10.2 和 RAG 1.10.3 靶标。背景对应于利用空白表达载体转染细胞。在相同试验中通过 I-SceI 切割 S 1234 I-SceI 靶标，显示为阳性对照。B. 三种 RAG 1.10 异二聚体针对三种 RAG1.10 靶标的活性。对照与 A 中所述的那些相同。

图 16 图示说明了来自 M2 和 M3 的突变体分别针对 RAG 1.10.2 和 RAG 1.10.3 靶标的同二聚化活性。通过其所突变的位置设计双突变体：例如，M2 7,61 代表 M2 K7E E61R。如图 15A 所示，背景对应于利用空白表达载体转染细胞。在相同试验中通过 I-SceI 切割 S 1234 I-SceI 靶标，显示为阳性对照。

图 17 图示说明了在酵母中对通过共表达 4 种双突变 M2 突变体与 4 种双突变 M3 突变体得到的 16 种异二聚体针对 3 种 RAG1.10 靶标进行筛选。原始 M2/M3 异二聚体对这三种靶标的活性显示为对照。对于每个 4 个点的酵母簇而言，左侧的两个点是实验结果，而右侧的两个点是用于评价实验质量和有效性的各内部对照。红色椭圆表示 4 种专性异二聚体。

图 18 图示说明了在 CHO 细胞中的染色体外分析中通过共表达 4 种双突变 M2 突变体与 4 种双突变 M3 突变体得到的 16 种异二聚体针对 3 种 RAG1.10 靶标的活性。

图 19 图示说明了在 CHO 细胞中的染色体外分析中 M2/M3 原始 RAG1.10 异二聚体和四种专性异二聚体 OH1 至 OH4 针对三种 RAG1.10 靶标的活性。阳性对照和阴性对照与图 15 种所述的那些相同。

图 20 图示说明了在酵母中两种单链分子 SC1 和 SC2 针对 3 种 RAG 1.10 靶标的筛选。SC1 是 M3-RM2-M2 分子，SC2 代表 M3(K7E K96E)-RM2-M2(E8K E61R) 分子。对于每 4 个点的酵母簇而言，左侧的两个点是试验结果，而右侧的两个点是用于评价实验质量和有效性的各内部对照。

实施例 1: 蛋白质设计

1) 材料和方法

利用自动蛋白质设计算法 FoldX(2.6.4 版本)(Guerois 等, J. Mol. Biol., 2002, 320,369-387; Schymkowitz 等, Nucleic Acids Res, 2005, 33, W382-388; Schymkowitz 等, Proc. Natl. Acad. Sci. U.S.A., 2005, 102, 10147-10152) 设计不同的异二聚体。将与 DNA 复合的 I-CreI 大范围核酸酶物的晶体结构(PDB 号: 1g9y.pdb; Chevalier 等, Nat. Struct. Biol., 2001, 8, 312-316) 用作设计模板。首先利用 FoldX 的<RepairPDB>命令优化结构, 以解除任意范德华力障碍。然后将各目的位置突变成丙氨酸, 利用<BuildModel>命令分别生成所有模型(异二聚体和同二聚体等)。最后, 通过<AnalyseComplex>命令分析所述复合物的各模型以计算不同的相互作用能量。

2) 结果

利用在 2.05 Å 的分辨率(PDB:1g9y) 下测定的与其相应 DNA 靶序列结合的 I-CreI 同二聚体的 X 射线结构(Chevalier 等, Nat. Struct. Biol., 2001, 8, 312-316) 来设计 I-CreI 的异二聚化界面。目的是促进异二聚化并且同时防止同二聚体的形成或者至少使得它们热力学不稳定。

同二聚体的大部分二聚化界面由两个 α -螺旋组成(两个单体中的 Lys7 至 Gly19), 其以卷曲螺旋排列, 这使得它们不适合重新设计。这些螺旋下方的氨基酸(Asp20 及之后)与 DNA 接触, 因此决定内切核酸酶的活性(活性位点)和特异性(DNA 识别)。仅这些功能就使得任意这些残基都无法在设计过程中被修饰。因此, 这使得进行异二聚化的可能性很小。仔细地分析了结构之后, 鉴定了参与界面中相互作用的三个区域, 其可在二聚体中被干扰和改变而不妨碍其结合能力或酶活性(图 1)。

其中之一是两个螺旋上方的区域(图 1A), 其中一个单体中的 Lys7 和 Glu8 与另一个单体中的对应残基建立有利的静电相互作用。为了保持异二聚体中的该相互作用并同时减少单体形成, 将这些残基在一个单体(下文称为单体 A)中替换为两个精氨酸, 另一个单体(下文称为单体 B)中替换为两个谷氨酸。这样 AA 和 BB 同二聚体将发生小的静电排斥, 而 AB 异二聚体的形成将是静电上更有利的。

第二部分是按照产生同二聚体的小的静电不平衡(相对于异二聚体)的同样思路来选择的, 但是其被置于所述卷曲螺旋的每一侧; 由各单体的 Lys96 和 Glu61 形成带电荷残基的双簇(double cluster)(图 1B)。为加强所述第一突变位点的静电作用, 将所述第二位点在单体 A 中突变为两个精氨酸, 在单体 B 中突变为两个谷氨酸, 这样, 在各单体中形成带电荷的三角形(A 中呈正电, B 中呈负电)。

第三个目的区域是两个螺旋中部的周围, 其参与相互作用表面并且主要由疏水相互作用和氢键组成, 形成一种微核(minicore)。由于 H 键网络相当复杂并且一路朝活性部位延伸, 所以决定仅仅干扰由一个单体的残基 Tyr12、Phe16、Val45、Trp53、Phe54、Leu55 和 Leu58 以及另一个单体的残基 Leu97 形成的疏水部分(所述残基 Leu97 像帽子一样起到关闭所述疏水性袋的作用)(图 1C)。重新设计这两个袋以便在不干扰所述异二聚体中疏水相互作用(即不产生腔和空间障碍)的情形下引入强的范德华力障碍。为此, 将大残基引入到单体 A 中(分别为 54 位的 Phe 或 Trp 以及 97 位的 Phe), 将小残基引入到单体 B 中(分别为 Gly 和 Leu)。可在 97 位引入甘氨酸到以使得给突变成色氨酸的 54 位更多空间。结果, AA 同二聚体形成巨大的空间位阻, 阻碍了其形成,

BB 同二聚体含有大的腔，使其不稳定。反之，AB 异二聚体的微核将被这些相容的氨基酸有效填充。最终，在单体 B 中 Leu 58 突变成甲硫氨酸，以防止由于小侧链的引入而在所述异二聚体中形成任何腔。注意利用甲硫氨酸替换 Lys 57 将是解决相同问题的另一种方法。

因此，设计了两种类型的单体 A 即 A1 和 A2（取决于 54 位上氨基酸的性质，分别为 Phe 或 Trp）以及 4 种类型的对应单体 B 即 B3 至 B6。与 B3 不同，B4 通过 97 位的甘氨酸发生突变以适应单体 A2 的 Trp 突变(图 1D)。与 B3 和 B4 不同，B5 和 B6 分别通过甲硫氨酸替换 Lys57，而 Leu58 未突变。利用 FoldX 测试了不同的突变以模拟所有的同二聚体（A1:A1、A2:A2、B3:B3、B4:B4、B5:B5 以及 B6:B6）和异二聚体（A1:B3、A2:B3、A2:B4、A1:B5、A2:B5 以及 A2:B6）以及得到不同的相互作用能量（表 I）。

表 I: 利用 FoldX 计算的野生型或设计的同二聚体和异二聚体的相互作用自由能

二聚体	突变体和野生型之间的互相作用能量差异(kcal/mol)
A1_B5	0,911
A2_B3	0,9113
A2_B5	1,0488
A1_B3	1,2375
A2_B4	3,7899
A2_B6	3,9856
B3_B3	7,9921
B5_B5	8,9138
A1_A1	9,8489
A2_A2	12,0421
B4_B4	12,6754
B6_B6	13,1897

最后的构建体（A2:B4 和 A2_B6）与野生型同二聚体相比表现出相互作用能量少量降低，但是仍然显著高于同二聚体。相反，A1:A1、A2:A2、B3:B3、B4:B4、B5:B5 和 B6:B6 同二聚体均变得非常不稳定，

因此预计这些物质仍然是单体。

实施例 2: 具有不同切割特异性的大范围核酸酶的改造以及切割条件的优化

1) 材料和方法

a) 大范围核酸酶的改造

如前文所述改造出具有改变的底物特异性的大范围核酸酶变体 (PCT 国际申请 WO 2004/067736、WO 2006/097784 和 WO 2006/097853; Epinat 等, *Nucleic Acids Res.*, 2003, 31, 2952-2962; Chames 等, *Nucleic Acids Res.*, 2005, 33, e178; Arnould 等, *J. Mol. Biol.*, 2006, 355, 443-458, 以及 Smith 等, *Nucleic Acids Res.*, 2006, 34, e149)。

b) 大范围核酸酶的制备和纯化

将携带 pET (NOVAGEN) I-CreI 突变体的新鲜 BL21(DE3) (STRATAGENE) 转化体在摇床上于 37°C 下在 5ml Luria 液体培养基 (LB 加上 30 µg/ml 卡那霉素) 中培养过夜。将该预培养物扩增至更大的培养物 (1:200)。在 OD₆₀₀ 为 0.6-0.8 时, 将培养瓶在冰上放置 15 分钟以抑制生长。通过在 16°C 下加入 IPTG (最终浓度为 1mM) 18 小时诱导表达, 通过离心 (15 分钟, 16,000 g) 收集细胞。将沉淀物重悬于含有 1 单位/µl 脱氧核糖核酸酶 I 的 30ml 冰冷裂解缓冲液 (50mM Tris HCl, pH 8; 200mM NaCl; 5mM MgCl₂; 10%甘油; 10 mM 咪唑) 中, 其后在 4°C 下进行操作。立即将混悬物冷冻在液氮中, 在 4°C 下于旋转平台 (60 rpm) 上解冻 16 小时。利用 ULTRA TURRAX T25 (JANKEL & KUNKEL, IKA-Labortechnik; 在冰上进行 3 次, 每次 1 分钟) 对所述混悬物进行匀浆, 然后利用 EmulsiFlex-C5 匀浆器 (AVESTIN) 破碎 5 次, 每次 500-1000 psi (磅/平方英寸)。在 150,000 g 下离心裂解物 60 分钟。通过 0.45µm 滤器 (MILLIPORE) 过滤上清液。将 2 倍珠粒体积的 250 mM NiSO₄ 加样到 5 ml Hi-Trap 柱 (AMERSHAM-PHARMACIA) 上并用 3 体积的结合缓冲液 (50mM Tris HCl, pH 8; 300mM NaCl; 1mM DTT; 20%甘油; 10mM 咪唑) 进行冲洗。然后将上清液加样到所述柱上并利用洗涤缓冲液 (含有 50 mM 咪唑的结合缓冲液) 洗涤直到 A_{280nm}

回到其基线水平上。利用洗脱缓冲液（0.3M 咪唑）洗脱蛋白质。收集蛋白质峰并立即加载到透析膜（分子量截留=3.5 kDa, SPECTRA）中，置于 2 升透析缓冲液（50 mM Tris HCl, pH 8; 200 mM NaCl; 1 mM DTT; 1mM EDTA; 50%甘油）中在 4℃ 下透析至少 12 小时。将纯化的蛋白质分成等份试样并快速冷冻在液氮中，贮存于-80℃。

c) DNA 消化分析

如前文所述并进行一些改变来测定靶序列的切割（Epinat 等, *Nucleic Acids Res.*, 2003, 31, 2952-2962）：将共表达的纯化酶在新鲜的透析缓冲液中稀释至 1 µg/µl（在分别纯化所设计的单体的情形下，加入每种单体 1.5µg，将它们各自稀释至 0.5 µg/µl）。将酶贮存于-80℃。利用 3.75 µM 酶、含有适当靶序列的 34 nM 的纯化的 3.2 kb 质粒（利用 *XmnI* 预先线性化）以及浓度为 50 至 300 mM 的 NaCl，在 20 µl 最终反应体积中制备反应混合物。将消化混合物在 37℃ 下于水浴中孵育 60 分钟，然后与 2.5µl 体积的终止缓冲液进行混合，所述缓冲液修改自 Wang 等, *Nucleic Acids Res.*, 1997, 25, 3767-3776 (5 ml 甘油; 2 ml EDTA, 0.5 M; 0.5 ml SDS, 20 %; 0.5 ml 蛋白酶 K, 20 mg/ml; 2.5 ml 溴酚蓝(1% w/v), pH 8)。将样品在 37℃ 下再孵育 30 分钟，然后将每种样品的一半在 1%琼脂糖凝胶上显影。

2) 结果

利用之前得到的两种识别不同 DNA 序列的大范围核酸酶（PCT 国际申请 WO 2006/097853; Arnould 等, *J. Mol. Biol.*, 2006, 355, 443-458）变体来验证特异性异二聚体的正确设计。这两种酶已通过筛选在 44 位、68 位和 70 位具有随机变异的 I-CreI 文库而得到，其中 64 种回文靶标是由 I-CreI 切割的 22bp 回文靶标的±3、±4 和±5 位的替换得到的。

这些变体均具有 Asp75 突变成 Asn 的突变，其降低了由碱性残基 Arg68 和 Arg70 的替换所引起的能量应力（energetic strain）；这些精氨酸通常满足埋藏在 I-CreI 结构中的 Asp75 的氢受体电位。一种具有突变 Q44K、R68T 和 R70G 的大范围核酸酶（以“KTG”表示）识别 DNA 靶标的-5、-4 和-3 位的碱基 cct。具有突变 Q44Q、R68A 和 R70N 的另一种大

范围核酸酶（称为“QAN”）识别 DNA 靶标的-5、-4 和-3 位的碱基 gtt。在全部实施例中，靶标 DNA 序列以 6 碱基密码表示，其中第一个 3 碱基对应于靶序列的-5、-4 和-3 位，另一个 3 碱基对应于同一 DNA 链上的+3、+4 和+5 位；所述两个三联体以斜线 (/) 分开。因此 KTG 酶的靶标是 cct/agg，QAN 酶的靶标是 gtt/aac，异二聚体 QAN-KTG 的混合 DNA 靶标记为 gtt/agg。

对于 WT 大范围核酸酶 I-CreI 而言，据报道消化其靶标 DNA 的理想条件是：20 mM Tris-HCl (pH 8.0-9.0) 与 10mM MgCl₂ (Wang 等, Nucleic Acids Res., 1997, 25, 3767-3776)，据报道 NaCl 离子强度大于 25 mM 以上时该酶被抑制。当利用这些条件以及 KTG 和 QAN 酶时，观察到了亚最佳特异性（图 2）。实际上，在低离子强度(≤50mM NaCl)下，两种酶不仅消化其靶标 DNA 序列，而且还消化混合的 DNA 靶标。这提示，仅一个单体与 DNA 强结合就足以进行消化。增强离子强度提高了酶对其靶标的活性，而且还降低对混合模板的消化：在约 225mM NaCl 下观察到几乎完美的特异性和良好的活性。该行为可解释为离子强度降低对 DNA 的亲合力（因此，如果仅一个单体在二聚体中建立特异性相互作用时，则防止结合）同时还增加酶活性。作为这些试验的结果，选择下述最佳缓冲液来消化大范围核酸酶设计物：25 mM HEPES (pH 8), 5 %甘油, 10 mM MgCl₂ 以及 225 mM NaCl。

实施例 3：所设计的切割人工靶标之突变体的表达与表征

1) 材料和方法

a) 大范围核酸酶突变体的克隆

在多达 6 个氨基酸位置上对两种同二聚化大范围核酸酶 KTG 和 QAN（基于 I-CreI 大范围核酸酶骨架）各自进行突变以形成两种可相容的异二聚化界面，记作 KTG-A2 和 QAN-B3。利用 Quickchange[®]试剂盒 (STRATAGENE, # 200518), 通过 round-the world PCR 引入突变。

利用 3 组互补引物引入 KTG-A2 突变 (K7R、E8R、F54W、E61R、K96R、L97F)：

(i) **A1_RR_F** (SEQ ID NO: 3) 和 **A1_RR_R** (SEQ ID NO:4):

5' caa tac caa ata taa cag gcg gtt cct gct gta cct ggc cg 3' (SEQ ID NO: 3)

5'cgg cca ggt aca gca gga acc gcc tgt tat att tgg tat tg 3'(SEQ ID NO: 4);

(ii) **A1_RF_F** (SEQ ID NO: 5) 和 **A1_RF_R** (SEQ ID NO: 6):

5' tca act gca gcc gtt tct gag att caa aca gaa aca ggc aaa cc 3' (SEQ ID NO: 5)

5' ggt ttg cct gtt tct gtt tga atc tca gaa acg gct gca gtt ga 3' (SEQ ID NO: 6);

(iii) **A2_WLR_F** 3' (SEQ ID NO: 7) 和 **A2_WLR_R** (SEQ ID NO: 8):

5' cca gcg ccg ttg gtg gct gga caa act agt gga tag aat tgg cgt tgg tta cg 3' (SEQ ID NO: 7)

5'cgt aac caa cgc caa ttc tat cca cta gtt tgt cca gcc acc aac ggc gct gg 3'(SEQ ID NO: 8).

利用 3 组互补引物引入 QAN-B3 突变(K7E、F54G、L58M、K96E):

(i) **B3_EE_F** (SEQ ID NO: 9) 和 **B3_EE_R** (SEQ ID NO: 10):

5' caa tac caa ata taa cga aga gtt cct gct gta cct ggc cg 3' (SEQ ID NO: 9), 和

5' cgg cca ggt aca gca gga act ctt cgt tat att tgg tat tg 3' (SEQ ID NO: 10);

(ii) **B3_GME_F** (SEQ ID NO: 11) 和 **B3_GME_R** (SEQ ID NO: 12):

5' cca gcg ccg ttg ggg tct gga caa aat ggt gga tga aat tgg cgt tgg tta cg 3' (SEQ ID NO: 11)

5' cgt aac caa cgc caa ttt cat cca cca ttt tgt cca gac ccc aac ggc gct gg 3' (SEQ ID NO: 12);

(iii) **B3_EL_F** (SEQ ID NO: 13) 和 **B3_EL_R** (SEQ ID NO: 14):

5' tca act gca gcc gtt tct gga act gaa aca gaa aca ggc aaa cc 3' (SEQ ID NO: 13) 和

5' ggt ttg cct gtt tct gtt tca gtt cca gaa acg gct gca gtt ga 3' (SEQ ID NO: 14).

根据制造商说明 (STRATAGENE, Quikchange®) 将所述第一组引物用于 PCR 和转化。将大约 300 个转化体细菌菌落合并到 2ml 培养基中, 通过小量制备 (miniprep) 回收质粒 DNA。利用诱变引物, 将该 DNA 用作第二轮以及其后第三轮 PCR 的模板。通过 DNA 测序验证 5 个三轮突变体。

值得一提的是上述引物对于具有改变的特异性的任意 I-CreI 突变体都是通用的, 因为二聚体界面突变位于 DNA 识别区域之外。

使用相似方法来进行异二聚体对 (QAN-A1、KTG-B3、QAN-B4) 的替代性设计, 将合适的突变引入到用于诱变 PCR 的寡核苷酸中 (图 3)。

b) 大范围核酸酶的制备与纯化

试验方法如实施例 2 中所述。

c) DNA 消化测定

试验方法如实施例 2 中所述, 其中 DNA 消化缓冲液由以下组成: 25 mM HEPES (pH 8)、5 %甘油、10 mM MgCl₂ 以及 50 mM NaCl (低离子强度)或 225 mM NaCl (高离子强度)。

d) 分析性离心

在离心试验中通过监测沉降性能来研究大范围核酸酶和突变体的寡聚状态。对于贮存在缓冲液(50 mM Tris-HCl pH 8.0、225 mM NaCl、1 mM EDTA、1 mM DTT、8%甘油)中的每个样品(0.52 mg/ml 各单体或 1.04 mg/ml 各 WT 同二聚体) 使用 1.04 mg 纯蛋白质。

, 在样品是在装配有四孔 AN-60 转子和双扇形铝中心件的 BECKMAN Optima XL-A 离心机中离心时 (48000 rpm, 4℃), 通过在 280nm 处监测吸光度信号来收集沉降速度谱。通过 Sedfit(Schuck, P., Biophys., 2000, 78, 1606-1619)中实施的 c 方法以及 UltraScan 7.1 软件包 (Demeler.B.,2005, <http://www.ultrascan.uthscsa.edu>) 来测定分子量分布。

根据 Laue 等(*In Analytical Ultracentrifugation in Biochemistry and Polymer Science*, 1992, Harding S. E., Rowe A.J., Horton J. C. Eds, pp. 90-125, Royal Society of Chemistry, Cambridge)公开的数据执行 UltraScan 7.1 操作来进行缓冲液强度和粘度的校正。

根据 Cohn E. J.和 Edsall J. T. (*In Proteins, Amino Acids and Peptides*, 1943, p.157, Reinhold, New York)的方法由蛋白质序列估计大范围核酸酶和突变体的偏微比容 (partial specific volume)。

2) 结果

通过对原始 KTG 和 QAN 酶表达载体进行定点诱变 (STRATAGENE, QuikChange®试剂盒), 然后表达和纯化相应的蛋白质来得到所设计的突变体 A1、A2、B3 和 B4 (图 4)。在突变体的不同组合中, 选择 QAN-A1、KTG-B3、KTG-A2、QAN-B3 和 QAN-B4 突变体。设计这些突变体以覆盖所有设计的异二聚体相互作用 A1:B3、A2:B3 和 A2:B4, 得到异二聚体 QAN-A1:KTG-B3、KTG-A2:QAN-B3 以及 KTG-A2:QAN-B4。

而野生型 KTG 和 QAN 酶产生的大多数蛋白质在可溶级分中, 但在所设计的酶情形下观察到了相反的结果: 大多数所表达的蛋白质以包涵体形式存在于沉淀中, 仅一小部分可被纯化, 并且即便这部分也被其它蛋白质污染 (图 4)。这是所设计的突变体在单独表达时不能同二聚化并因而变得不稳定的第一个标志。

在低和高离子强度 (50 mM 或 225 mM NaCl) 下测试了纯化的 A1、A2、B3 和 B4 酶对 3 种 DNA 靶标的活性 (图 5)。在低盐浓度下, 仅仅可检测到对 QAN-A1 一些特异性 DNA 消化活性; 观察不到其它酶的特异性切割。此外, 在高离子强度下, 当与酶一起孵育时, 不能检测到所预期的两条 DNA 条带, 尽管在一些情形下 DNA 的量下降。

当单独表达非同二聚化单体设计物时, 这些结果被所得到的低产率和低质量的蛋白质所破坏; 即使利用 6 升之大体积的细菌产生的蛋白质也不足 (设计的单体为 0.5-1.5 mg/ml; 而野生型二聚化单体为 30 mg/ml)。

为了考察所纯化设计酶的寡聚状态, 通过分析性超速离心测量了其大小特性 (图 6)。在单独表达的 A1、A2、B3 和 B4 蛋白质的情形下, 观察到了所预期的单体酶的外观。然而, 还检测到较高分子量的聚集体, 包括三聚体和四聚体 (图 6B; 仅显示了 KTG-A2 和 QAN-B3, 但是利用其它设计物得到相似的结果)。因此, 所设计的酶实际上不能同二聚化, 这可能在纯化过程中已经影响到了其稳定性和聚集特性。

为了研究异二聚化的潜能, 将等摩尔量的单独纯化的设计的酶 (QAN-A1、KTG-A2、QAN-B3、KTG-B3、QAN-B4) 以所有可能的组合进行混合。在 KTG-A2/QAN-B3 的情形下, 观察到了对应于二聚体

分子量的主要物质的外观,但是这不是所形成的唯一物质。对于所述对 QAN-A1/KTG-B3 和 KTG-A2/QAN-B4 而言,观察到了分子质量在单体和二聚体之间的新峰的外观以及高分子量聚集体的减少。对于那些将不产生异二聚体的组合而言,没有蛋白质行为的显著变化可观察到。大体上,这些结果表明,单独表达异二聚化单体不是有效的策略,因此在细菌细胞内进行了共表达测定。

实施例 4: 所设计的切割人工靶标之突变体的共表达与表征

1) 材料和方法

a) 所设计的单体的共表达

为了从 QAN-B3 单体中除去 His 标签,利用 *NcoI/NotI*(NEW ENGLAND BIOLABS)将其从母体质粒 pCLS1214 (pET 系列)上切下来。然后将该片段克隆到类似切割的 pCDFDuet1 质粒(NOVAGEN)中。利用该混合物转化 TOP10 超级感受态细胞 (INVITROGEN) 并在 50 μ g/ml 的链霉素-硫酸大观霉素中进行选择。通过 DNA 测序验证克隆。将 BL21(DE3) 超级感受态细胞与 10 ng 各质粒 (pCLS1211-KTG-A2 和 pCDFDuet1-QAN-B3)进行共转化。通过在链霉素-硫酸大观霉素存在下培养所转化的克隆来选择双重转化体 (double transformants)。

b) 大范围核酸酶的纯化

实验方法如实施例 2 中所述。

c) DNA 消化分析

实验方法如实施例 2 中所述,其中 DNA 消化缓冲液由以下组成: 25 mM HEPES (pH 8), 5 %甘油, 10 mM MgCl₂ 以及 225 mM NaCl (高离子强度)。

d) 分析性离心

实验方法如实施例 3 中所述。

2) 结果

实施例 3 中所示的结果提示, 异二聚体设计物可能是功能性的, 但是单体酶的表达导致强的聚集并因此得到部分失活的酶。为了避免该问题, 将单体基因表达盒亚克隆到互补质粒中并且共转化到细菌细胞中, 从而原始的 pET 系列质粒表达一个单体(具有 His 标签), 相容的 pCDFDuet 载体 (INVITROGEN) 表达配对的单体(无 His 标签)。双重抗生素选择保证各细胞均包含两种质粒。

所述共表达的 KTG-A2/QAN-B3 蛋白的表达分析显示出未产生包涵体, 这提示之前的聚集问题已得到解决。所纯化的酶的 SDS-page 分析随后显示了具有大致相同量蛋白质的 2 条带, 这提示我们异二聚体被纯化, 而同二聚体未被纯化。质谱直接证明了两种蛋白质和异二聚化复合物的存在。此外, 所纯化的蛋白质的分析性超速离心给出了在所期望的二聚体分子量处的非常清晰的单一谱(图 6C)。

利用纯化的共表达异二聚体设计物消化各 DNA 靶标, 观察到异二聚体 DNA 靶标 (gtt/agg) 的清晰的特异性切割, 几乎观察不到同二聚体靶标 (cct/agg 和 gtt/aac; 图 7A) 的特异性切割。尽管 KTG-A2/QAN-B3 异二聚体似乎表现出对杂合 DNA 靶标的特异性切割, 但是其活性可能较小。为了排除这一点, 我们比较了 QAN、KTG 和 KTG-A2/QAN-B3 在时间进程中的活性(图 7B)。该实验表明, 就对其底物的 50% 切割而言, 利用相同的表观蛋白质浓度, KTG 和异二聚体具有相似的活性, 而 QAN 则需要 4 倍时间的孵育。

因此, 所述蛋白质设计是成功的, 这使得能特异性形成异二聚化 I-CreI 酶变体, 只要共表达了单体部分即可。

利用共表达的 QAN-A1/KTG-B3 蛋白质重复的相同实验显示了混合的结果: 在纯化后有 2 个条带, 但一个条带的强度大于另一个条带的强度, 具有对异二聚体的特异性切割, 但是与 KTG-A2/QAN-B3 组合相比特异性水平降低(图 6D)。分析性离心显示形成了含有小比例聚集体的二聚体。

最后, 第三个共表达组合 KTG-A2/QAN-B4 导致仅一个条带被纯化, 并且通过分析性离心检测到单体(图 6D)。因此, 该设计不能成功地制备异二聚体, 即使发生了共表达。

很有趣的是，KTG-A2/QAN-B3、QAN-A1/KTG-B3 和 KTG-A2/QAN-B4 之间的二聚体比例和活性与通过 FoldX 预测的能量相关性极好（表 I）。就所预期的通过计算机得到的（*in silico*）能量而言，最佳设计在体外形成最佳异二聚体。

实施例 5：竞争试验

为了测量酶对同二聚体和异二聚体 DNA 位点的相对差别，进行了竞争试验，其中所述酶同时接近等摩尔量的两种底物（图 8）。选择 KTG-A2/QAN-B3 异二聚体和 KTG-wt 用于这些试验，因为它们表现出了针对其各自靶标的相似活性（图 8B）。在时间进程实验中，KTG 优先切割其靶标，但是 KTG-A2/QAN-B3 靶标的相应位点被轻微消化（图 8A）。与之相反，KTG-A2-QAN-B3 异二聚体优先切割 Q-K 异二聚体 DNA，很少切割 KTG 位点。

为了更直接地比较相对切割偏好，对两种 DNA 靶标的等摩尔混合物进行酶滴定（图 8B）。这使得能在竞争性条件下测定对相应靶标和非相应靶标表现出 50%切割率的表现浓度：KTG-wt 就相应靶标而言 = 0.1 μM ；KTG-wt 就非相应靶标而言 = 1.5 μM ；KTG-A2/QAN-B3 就相应靶标而言 = 0.3 μM ；KTG-A2/QAN-B3 就非相应靶标而言 = 3.2 μM 。因此，就两种酶而言，发现相应靶标和非相应靶标的 50%切割率的酶浓度的差异为大约 10 至 15 倍。

总之，这些结果表明，尽管两种母体构建体和突变体衍生物的特异性不是绝对的，但是可设计活性与最佳 wt 母体相似并且对其异二聚体靶标具有明确的切割优先选择性的专性异二聚体大范围核酸酶，而原始同二聚体对其同二聚体靶标具有相反的优先选择性。

实施例 6：通过靶向精氨酸 51-天冬氨酸 137 相互作用制备 RAG1.10 专性异二聚体

之前通过在酵母中共表达能切割来自回文 RAG1.10 的靶标 RAG1.10.2 和 RAG1.10.3 的两种 I-CreI 突变体得到了能切割来自 RAG1 基因之靶标（RAG1.10 序列，图 9）的异二聚体（Smith 等, *Nucleic Acids Res.*, 2006, 34, e149; 图 10）。但是，两种 I-CreI 突变体的共表达导致溶液中存

在3种分子类型：两种同二聚体以及异二聚体。通过引入其它突变制备了专性异二聚体，以在不影响异二聚体稳定性和切割性能的情形下防止形成功能性同二聚体。因此，仅 RAG1.10 靶标将被切割。为得到该专性异二聚体，选择了两个残基精氨酸 51 (R51) 和天冬氨酸 137 (D137)，其在两种 I-CreI 单体之间产生分子间相互作用 R51-D137。这两个残基还接近于活性位点，并参与围绕所述活性位点的水分子壳的配位（图 11）。通过例如在 RAG 1.10.2 切割酶（一种 I-CreI 变体，其切割 RAG 1.10.2 靶标）上产生 R51D 突变（精氨酸被天冬氨酸残基替换）以及在 RAG 1.10.3 切割酶上产生互补的 D137R 突变，在异二聚体中仍然存在吸引性的 D51-R137 相互作用，而在 RAG 1.10.2 同二聚体中存在排斥性的 D51-D137 相互作用，在 RAG 1.10.3 同二聚体中也存在排斥性的 R51-R137 相互作用。推测这些排斥性相互作用可使得同二聚体的形成失去稳定性和/或干扰催化机制。

如 Smith 等, *Nucleic Acids Res.*, 2006, 34, e149 中所述，之前得到了具有序列 KRSNQS/AYS DR (标出了 28、30、32、33、38、40 / 44、68、70、75、77 位的残基) 的 RAG 1.10.2 切割酶(称为 M2)以及具有序列 NNSSRR/YRSQV 的一种 RAG 1.10.3 切割酶(称为 M3)，对于每种突变体引入 R51D 突变或 D137R 突变，以产生 4 种单突变体。

在本实施例和下述实施例中，通过直接重复重组测定法，在酵母或哺乳动物细胞中测量本发明的专性异二聚体大范围核酸酶对目的非回文靶标和所衍生的回文靶标（其被两种同二聚体所切割）的切割活性，如前文所述（PCT 国际申请 WO 2004/067736 和 WO 2006/097853; Epinat 等, *Nucleic Acids Res.*, 2003, 31, 2952-2962; Chames 等, *Nucleic Acids Res.*, 2005, 33, e178 以及 Arnould 等, *J. Mol. Biol.*, 2006, 355, 443-458）。报告基因载体包含在干扰序列内的 LacZ 报告基因（同向重复）和 DNA 靶标序列的两个截短的非功能性拷贝，其在酵母或哺乳动物表达载体中克隆。大范围核酸酶的表达导致基因组嵌合 DNA 靶序列的切割。该切割诱导同向重复之间进行同源重组，产生功能性 LacZ 基因，其表达可通过合适的方法进行监测。

1) 材料和方法

a) R51D 和 D137R 突变的引入

通过对 RAG1.10.2 和 RAG1.10.3 切割酶的 DNA 进行两次 PCR 反应来引入各突变。

对于 R51D 突变而言，第一次 PCR 反应是利用引物 Gal10F 5'-gcaacttagtgctgacacatacagg-3' (SEQ ID NO: 37) 和 R51DRev 5'-ttgtccagaaaccaacgggtcctgggtctttgagtcac-3' (SEQ ID NO: 38) 进行的，第二次反应是利用引物 R51DFor 5'-gtgactcaaaagaccaggaccgttggttcttgacaaac-3' (SEQ ID NO: 39) 和 Gal10R 5'-acaaccttgattggagacttgacc-3' (SEQ ID NO: 40) 进行的。

为了引入 D137R 突变，利用引物 D137RRev 5'-ggttttacgcgtcttagaacgggttcagagctgcaatctg-3' (SEQ ID NO: 41) 和 D137RFor 5'-cagattgcagctctgaaccgttctaagacgcgtaaaacc-3' (SEQ ID NO: 42) 通过相同的方法进行。纯化所有的 PCR 片段并与前述的利用 *NcoI* 和 *EagI* 线性化（对于 RAG1.10.2 切割酶而言）的 pCLS0542 载体（见例如 PCT 国际申请 WO 2006/097853 的图 5）或前述的利用 *NgoMIV* 和 *DraIII* 线性化（对于 RAG1.10.3 切割酶而言）pCLS1107 载体（见例如 PCT 国际申请 WO 2007/093918 的图 14）一起通过高效 LiAc 转化方案来转化酒酵母 (*Saccharomyces Cerevisiae*) 菌株 FYC2-6A (MAT α 、trp1 Δ 63、leu2 Δ 1、his3 Δ 200)。然后通过酵母体内同源重组产生 I-CreI 突变体的完整编码序列。回收酵母 DNA 并用于转化 DH5 α 大肠杆菌 (*E.coli*) 菌株。纯化细菌 DNA，通过测序来检查 R51D 或 D137R 的存在。

b) 突变体共表达

利用在 pCLS0542 表达载体中编码来自原始 RAG1.10.2 切割酶的突变体之 DNA 和在 pCLS1107 表达载体中编码来自原始 RAG1.10.3 切割酶的突变体之 DNA 转化酵母菌株 FYC2-6A。在 -L Glu + G418 培养基上选择转化体。

c) 在酵母中接合共表达大范围核酸酶的克隆并筛选

使用菌落网格(gridder) (QpixII, Genetix) 进行接合。使用低网格密度（约 4 个点/cm²）将突变体在尼龙滤膜覆盖的 YPD 平板上划成网格状。在同一滤膜上进行第二次划网格的过程，从而印上第二层，所述第二层由针对每个靶标的含有不同报告基因的酵母菌株组成。将膜置于富含固体琼脂

YPD 的培养基上，在 30℃ 孵育一夜，以使其接合。接着，将滤膜转移至缺少亮氨酸和色氨酸且加入 G418 并含有半乳糖(1%)作为碳源的合成培养基上，在 37℃ 孵育五天，以选择带有表达载体和靶标载体的二倍体。5 天后，将滤膜置于在 0.5 M 磷酸钠缓冲液 (pH 7.0) 中的含 0.02% X-Gal、0.1% SDS、6% 二甲基甲酰胺(DMF)、7 mM β -巯基乙醇、1% 琼脂糖的固体琼脂糖培养基中，在 37℃ 孵育，以检测 β -半乳糖苷酶活性。通过扫描分析结果，并使用适当的软件进行定量。

2) 结果

R51D RAG 1.10.2 切割酶与 D137R RAG 1.10.3 切割酶的共表达导致 RAG 1.10 靶标的有效切割，而未看到 RAG 1.10.2 和 RAG 1.10.3 靶标的切割 (图 12)。这清楚地表明所述异二聚体对 RAG 1.10 靶标序列有活性，而所述两种同二聚体不切割回文 RAG 1.10.2 和 RAG 1.10.3 靶标。这样的异二聚体符合用于定义专性异二聚体的标准。D137R RAG 1.10.2 切割酶与 R51D RAG 1.10.3 切割酶的共表达还导致形成专性异二聚体，其对 RAG 1.10 靶标的活性小于先前的异二聚体。

为了比较此异二聚体与实施例 1、3 和 4 中所述的异二聚体，如实施例 1 所述计算了 R51D 和 D137 突变体相互作用的自由能。如表 II 所示，所述异二聚体 (S1_S2) 与野生型同二聚体非常相似，而突变体同二聚体 (S1_S1 和 S2_S2) 明显失去稳定性。

表 II: 野生型或设计的同二聚体和异二聚体的 FoldX 计算的相互作用自由能

二聚体	突变体和野生型之间相互作用 能量差异 (kcal/mol)
A1_B5	0,911
A2_B3	0,9113
A2_B5	1,0488
A1_B3	1,2375
S1_S2	1,3827
A2_B4	3,7899
A2_B6	3,9856
S1_S1	5,9036
B3_B3	7,9921
B5_B5	8,9138
S2_S2	9,7611
A1_A1	9,8489
A2_A2	12,0421
B4_B4	12,6754
B6_B6	13,1897

* S1和S2分别代表R51D和D137R突变体

实施例 7: 通过靶标赖氨酸 7-谷氨酸 8 相互作用来制备 RAG1.10 专性异二聚体

在本实施例中，靶向发生在两个单体之间的赖氨酸 7-谷氨酸 8 相互作用。因此，将单突变 K7E 或 E8K 引入到实施例 6 的 M2 或 M3 突变体中。然后将所得的突变体在 CHO 细胞中共表达，利用之前所述的 CHO 细胞中的单链退火 (SSA) 染色体外测定来监测所述异二聚体和所述两种同二聚体针对 3 种 RAG1.10 靶标的活性 (PCT 国际申请 WO 2004/067736 和 WO 2006/097853; Epinat 等, Nucleic Acids Res., 2003, 31, 2952-2962; Chames 等, Nucleic Acids Res., 2005, 33, e178; Arnould 等, J. Mol. Biol., 2006, 355, 443-458)。

1) 材料和方法

a) K7E 和 E8K 突变的引入

已经将来自 I-CreI 的突变体 M2 和 M3 在哺乳动物表达载体中进行了克隆。通过对所述 M2 和 M3 突变体的 DNA 进行两次重叠 PCR 反

应来引入各突变。对于 K7E 突变而言,第一次 PCR 反应是利用引物 CMVFor (5'-cgcaaatgggcggtaggcgtg-3'; SEQ ID NO: 44) 和 K7ERev (5'-gtacagcaggaactcttcgttatatttggtattgg-3'; SEQ ID NO: 45) 进行的,第二次反应是利用引物 K7EFor (5'-aataccaaatataacgaagagttcctgctgtacc-3'; SEQ ID NO: 46) 和 V5epitopeRev (5'-cgtagaatcgagaccgaggagagg-3'; SEQ ID NO: 47) 进行的。对两种 PCR 片段进行凝胶纯化、混合并利用 CMVFor 和 V5epitopeRev 引物进行第三次装配 PCR。所得的 PCR 片段包含具有 K7E 突变之 I-CreI 突变体的开放阅读框。然后纯化 PCR 片段,利用限制性酶 *SacI* 和 *XbaI* 消化并连接到也通过 *SacI* 和 *XbaI* 进行了消化的 pCLS1088 中(图 13)。通过测序(MILLEGEN)验证所得的克隆 M2 K7E 或 M3 K7E。除了利用两种引物 E8KRev (5'-caggtacagcaggaacttttggttatatttg-3'; SEQ ID NO: 48) 和 E8KFor (5'-accaaataatacaaaaagttcctgctgtacctgg-3'; SEQ ID NO: 49) 以外,利用完全相同的方案将 E8K 突变引入到 M2 和 M3 突变体中。

b) 在用于 CHO 细胞中染色体外测定的载体中克隆 RAG1.10、RAG1.10.2 和 RAG1.10.3 靶标

如下克隆目的靶标:从 PROLIGO 订购寡核苷酸,其相应于两侧为 Gateway 克隆序列的靶序列。利用 Gateway 方案(INVITROGEN)将由单链寡核苷酸的 PCR 扩增得到的双链靶标 DNA 克隆到 CHO 报告基因载体(pCLS1058,图 14)中。

c) CHO 细胞中的染色体外检测

根据供应商(QIAGEN)的方案利用 Polyfect 转染试剂转染 CHO 细胞。转染后 72 小时,除去培养基,加入 150 μ L 裂解/解析缓冲液用于 β -半乳糖苷酶液体测定(通常,1 升缓冲液包含 100 ml 裂解缓冲液(Tris-HCl 10 mM pH7.5, NaCl 150 mM, Triton X100 0.1 %, BSA 0.1 mg/ml, 蛋白酶抑制剂)、10ml Mg 100X 缓冲液(MgCl₂ 100mM, β -巯基乙醇 35%)、110 ml ONPG 8 mg/ml 以及 780 ml 磷酸钠 0.1 M pH7.5)。在 37 $^{\circ}$ C 孵育后,于 420nm 处测量吸光度。整个过程在自动化 Velocity11 BioCel 平台上进行。

2) 结果

利用之前所述的染色体外检测于 CHO 细胞中监测单突变体 M2 K7E 或 E8K 以及 M3 K7E 或 E8K 对其各自的 DNA 靶标 RAG 1.10.2 和 RAG 1.10.3 的活性。如图 15A 所示, 7 位或 8 位突变的突变蛋白质的活性与原始蛋白质的相同。第二步, 共表达 M2 和 M3 突变体以形成 3 种可能的异二聚体: M2/M3、M2 K7E/M3 E8K、M2 E8K/M3 K7E。然后监测其对 3 种 RAG1.10 靶标的活性。图 15B 显示, 当共表达 7 位或 8 位突变的 M2 或 M3 突变体时, 可观察到同二聚化活性的降低, 尤其是具有 E8K 突变的突变体。同时, 在 7 位和 8 位具有突变的两种异二聚体与原始 M2/M3 异二聚体相比较而言保持相同的针对天然 RAG1.10 靶标的活性水平。两种异二聚体 M2 K7E / M3 E8K 和 M2 E8K / M3 K7E 不是严格的专性异二聚体, 因为仍然可检测到同二聚化活性, 甚至在蛋白质仅可形成的同二聚体时也未改变。然而, 当蛋白质可形成同二聚体和异二聚体时, 靶标 K7-E8 相互作用显著降低至少一种同二聚化活性。因此, 与功能性同二聚体形成相比较而言, 这种突变非常有利于功能异二聚体形成。

实施例 8: 通过靶向赖氨酸 7-谷氨酸 8 以及谷氨酸 61-赖氨酸 96 相互作用来制备 RAG1.10 专性异二聚体

实施例 7 表明, 来自具有单个 K7E 或 E8K 突变之 M2 和 M3 突变体的 M2/M3 异二聚体 (M2K7E/M3E8K 和 M2E8K/M3K7E) 保留了同二聚化活性。因此, 通过靶向谷氨酸 61-赖氨酸 96 相互作用 (其出现在形成异二聚体的两个单体之间) 将其它突变引入到 M2 或 M3 突变体中。对于每种 M2 或 M3 突变体而言, 产生了 4 种双突变体: K7E E61R、E8K E61R、K7E K96E 和 E8K K96E。然后在 CHO 细胞中监测 8 种双突变体的同二聚化活性。在第二个阶段, 将 4 种 M2 双突变体与 4 种 M3 双突变体共表达, 利用前述的酵母筛选分析和染色体外 CHO 测定来监测所得 16 种异二聚体针对 3 种 RAG1.10 靶标的活性。

1) 材料和方法

a) E61R 和 K96E 突变的引入

通过对在酵母表达载体中克隆的来自 M2 或 M3 的突变体之 DNA 进行两次重叠 PCR 反应来引入各突变。对于 E61R 突变而言, 第一次

PCR 反 应 是 利 用 引 物
Gal10F (5'-GCAACTTTAGTGCTGACACATACAGG-3'; SEQ ID NO: 37 和
E61RRev (5'-gtaaccaacgccaatacgaatccactagttgtcc-3'; SEQ ID NO: 50)进行的, 第二次反应是
利 用 引 物 E61RFor (5'-gacaaactagtgatcgattggcggtgttacg-3'; SEQ ID NO: 51) 和
Gal10R (5'-ACAACCTTGATTGGAGACTTGACC-3'; SEQ ID NO: 40)进行的。为了引入
K96E 突 变 , 我 们 利 用 引 物
K96ERev (5'-cctgtttctgttcagttccagaaacggctgcag-3'; SEQ ID NO: 52) 和
K96EFor (5'-ctgcagccgtttctggaactgaaacagaaacagg-3'; ; SEQ ID NO: 53)通过完全相同的方法
进行。纯化所有的 PCR 片段, 并将其与具有 *NcoI* 和 *EagI* 的 pCLS0542
线性化载体 (对于来自 M2 的突变体而言) 或具有 *NgoMIV* 和 *DraIII*
的 pCLS1107 线性化载体 (对于来自 M3 的突变体而言) 一起通过高效
LiAc 转化方案用于转化酿酒酵母 (*Saccharomyces cerevisiae*) 菌株
FYC2-6A (MAT α 、trp1 Δ 63、leu2 Δ 1、his3 Δ 200)。然后通过酵母中
体内同源重组产生 I-CreI 突变体的完整编码序列。回收酵母 DNA, 用
于转化 DH5 α 大肠杆菌菌株。纯化细菌 DNA, 通过测序来检查 E61R 或
K96E 的存在。

b) 突变体共表达

利用在 pCLS1107 表达载体中编码来自 M3 的突变体之 DNA 转化
在 pCLS0542 表达载体中编码来自 M2 的突变体的酵母菌株。在 -L Glu +
G418 培养基上选择转化体。

c) 将 M2 或 M3 双突变体克隆到哺乳动物表达载体中

利 用 引 物
CCM2For (5'-aagcagagctctctggctaactagagaacccactgcttactggcttatcgaccatggccaatacca
aatataacaaagagttcc-3'; SEQ ID NO: 54) 和
CCMRev (5'- ctgctctagattagtcggccgcccggggaggattcttc-3'; SEQ ID NO: 55)通过 PCR 扩增各
突变体 ORF。利用限制酶 *SacI* 和 *XbaI* 消化 PCR 片段, 然后将其连接
到也通过 *SacI* 和 *XbaI* 消化的载体 pCLS1088 中。通过测序(MilleGen)
来检查所得的克隆。

2) 结果

4种双突变 M2 或 M3 突变体的同二聚化活性显示在图 16 中, 其表明该活性降低至几乎背景水平, 携带突变 K7E 和 K96E 的突变体尤其如此。第二步, 将来自 M2 的所有突变体与来自 M3 的所有突变体共表达。然后利用酵母筛选测定(图 17)和染色体外 CHO 测定(图 18)来监测所得异二聚体针对 3 种 RAG1.10 靶标的活性。图 17 显示, 通过红色椭圆从 16 种可能的异二聚体中标出了在酵母中表现为专性异二聚体的 4 种异二聚体。它们对 RAG1.10 靶标具有强的切割活性, 与 M2/M3 异二聚体的活性相等, 而它们表现出几乎不可检测的同二聚化活性。为了突出专性异二聚体的该概念, 图 18 表明, 双突变 M2 突变体仅在其与双突变 M3 突变体(具有互补突变)共表达时起作用。4 种所得的专性异二聚体被称为 OH1 至 OH4, 分别为:

OH1=M2 K7E E61R/ M3 E8K K96E

OH2=M2 E8K E61R/M3 K7E K96E

OH3=M2 K7E K96E/M3 E8K E61R

OH4=M2 E8K K96E/M3 K7E E61R

最后, 图 19 表明, 4 种专性异二聚体 OH1 至 OH4 在 CHO 细胞中对 RAG1.10 靶标具有与原始 M2/M3 异二聚体相同的切割活性, 而同二聚化活性则降低至几乎不可检测的水平。这些数据显示, 可通过将 K7E/E8K 和 E61R/K96E 突变引入到来自 I-CreI 的突变体中形成异二聚体而产生专性异二聚体。

实施例 9: K7E/E8K 和 E61R/K96E 突变在单链分子设计中的用途

为了进一步证实 K7E/E8K 和 E61R/K96E 突变改善大范围核酸酶特异性的用途, 将这些突变引入到 RAG 单链构建体中。利用称为 RM2 的 32 个氨基酸的长连接物将 M3 突变体的 C 端连接到 M2 突变体的 N 端的 6 位残基上来改造出单链分子 M3-RM2-M2。因为该分子表现出针对 RAG1.10.2 靶标的一些同二聚化活性, 所以将突变 K7E、K96E 引入到 M3 突变体中, 并将突变 E8K、E61R 引入到 M2 突变体中以产生单链分子 M3(K7E K96E)-RM2-M2(E8K E61R), 其还称为 SC_OH。然后利用前述的酵母筛选测定法监测两种单链分子针对 3 种 RAG1.10 靶标

的活性。

1) 材料和方法

SC_OH 单链分子的克隆

对在 pCLS0542 酵母表达载体中克隆的带有 E8K 和 E61R 突变的 M2 突变体进行 PCR 反应。所述 PCR 反应利用反向引物 CreCterSacI (5'-tagacgagctcctacggggaggatttcttctcgcct-3'; SEQ ID NO: 56) 和正向引物 RM2 (5'-tatcgccgggtggatctgataagtataatcaggctctgtctaaatacaaccaagcadtccaagtacaatcaggccctgtctggaggcggtccaacaagagttcctgctgtatcttgctggattt-3' SEQ ID NO: 57)。

纯化 PCR 片段并利用 *EagI* 和 *SacI* 进行消化并连接到所述酵母表达载体中以便也利用 *EagI* 和 *SacI* 消化携带突变 K7E 和 K96E 的 M3 突变体。对克隆进行测序后, 得到 SC_OH 单链分子。

2) 结果

图 20 中所示的两种单链分子 M3-RM2-M2 和 SC_OH 针对 3 种 RAG1.10 靶标的酵母筛选表明, 引入 K7E/E8K 和 E61R/K96E 允许在不降低所述单链对 RAG1.10 靶标的切割活性的情形下消除针对 RAG1.10.2 靶标的同二聚化活性。因此, 可将实施例 8 中所述的突变引入到专性异二聚体设计物、单链分子中以在不影响其对目的 DNA 靶标之活性的情形下提高其特异性。

<110> 赛莱克蒂斯公司
埃曼努埃尔·法雅尔多桑切斯
西尔韦斯特·格里佐
马克·伊萨兰
路易斯·塞拉诺普布尔
弗朗索瓦·斯特里谢

<120> 专性异二聚体大范围核酸酶及其用途

<130> HLP/bv-1546/14PCT

<160> 65

<170> PatentIn version 3.3

<210> 1

<211> 163

<212> PRT

<213> 莱茵衣藻 (*Chlamydomonas reinhardtii*)

<400> 1

Met Asn Thr Lys Tyr Asn Lys Glu Phe Leu Leu Tyr Leu Ala Gly Phe
1 5 10 15

Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Lys Pro Asn Gln Ser
20 25 30

Tyr Lys Phe Lys His Gln Leu Ser Leu Ala Phe Gln Val Thr Gln Lys
35 40 45

Thr Gln Arg Arg Trp Phe Leu Asp Lys Leu Val Asp Glu Ile Gly Val
50 55 60

Gly Tyr Val Arg Asp Arg Gly Ser Val Ser Asp Tyr Ile Leu Ser Glu
65 70 75 80

Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu Lys
85 90 95

Leu Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Trp Arg Leu
100 105 110

Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr Trp
115 120 125

Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr Thr
130 135 140

Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys Lys
145 150 155 160

Ser Ser Pro

<210>	2		
<211>	24		
<212>	DNA		
<213>	人工序列		
<220>			
<223>	I-CreI 靶标	DNA	
<400>	2		
	tcaaaacgtc gtacgacgtt ttga		24
<210>	3		
<211>	41		
<212>	DNA		
<213>	人工序列		
<220>			
<223>	引物		
<400>	3		
	caataccaaa tataacaggc ggttcctgct gtacctggcc g		41
<210>	4		
<211>	41		
<212>	DNA		
<213>	人工序列		
<220>			
<223>	引物		
<400>	4		
	cggccaggta cagcaggaac cgcctgttat atttggtatt g		41
<210>	5		
<211>	44		
<212>	DNA		
<213>	人工序列		
<220>			
<223>	引物		
<400>	5		
	tcaactgcag ccgtttctga gattcaaaca gaaacaggca aacc		44
<210>	6		
<211>	44		
<212>	DNA		
<213>	人工序列		
<220>			
<223>	引物		
<400>	6		
	ggtttgacctg tttctgtttg aatctcagaa acggctgcag ttga		44
<210>	7		
<211>	53		
<212>	DNA		
<213>	人工序列		
<220>			
<223>	引物		

<400> 7	
ccagcgccgt tggctggctgg acaaactagt ggatagaatt ggcgttggtt acg	53
<210> 8	
<211> 53	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 8	
cgtaaccaac gccatttcta tccactagtt tgtccagcca ccaacggcgc tgg	53
<210> 9	
<211> 41	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 9	
caatacaaaa tataacgaag agttcctgct gtacctggcc g	41
<210> 10	
<211> 41	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 10	
cggccaggta cagcaggaac tcttcgttat atttggatt g	41
<210> 11	
<211> 53	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 11	
ccagcgccgt tgggtcttgg acaaatggt ggatgaaatt ggcgttggtt acg	53
<210> 12	
<211> 53	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 12	
cgtaaccaac gccatttcta tccaccattt tgtccagacc ccaacggcgc tgg	53
<210> 13	
<211> 44	
<212> DNA	
<213> 人工序列	

[illegible]

130	135	140	
act tct gaa acc gtt cgt gct gtg ctg gac agc ctg agc gag aag aag			480
Thr Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys			
145	150	155	160
aaa tcc tcc ccg gcg gcc gca ctc gag cac cac cac cac cac tga			528
Lys Ser Ser Pro Ala Ala Leu Glu His His His His His His			
	165	170	175

<210> 16
 <211> 175
 <212> PRT
 <213> 人工序列

<220>
 <223> 合成构建体

<400> 16

Met Ala Asn Thr Lys Tyr Asn Lys Glu Phe Leu Leu Tyr Leu Ala Gly
 1 5 10 15

Phe Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Lys Pro Asn Gln
 20 25 30

Ser Tyr Lys Phe Lys His Gln Leu Ser Leu Thr Phe Gln Val Thr Gln
 35 40 45

Lys Thr Gln Arg Arg Trp Phe Leu Asp Lys Leu Val Asp Glu Ile Gly
 50 55 60

Val Gly Tyr Val Ala Asp Asn Gly Ser Val Ser Asn Tyr Ile Leu Ser
 65 70 75 80

Glu Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu
 85 90 95

Lys Leu Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Glu Gln
 100 105 110

Leu Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr
 115 120 125

Trp Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr
 130 135 140

Thr Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys
 145 150 155 160

Lys Ser Ser Pro Ala Ala Ala Leu Glu His His His His His His
 165 170 175

<210> 17
 <211> 528
 <212> DNA

<213> 人工序列

<220>

<223> KTG 变体

<220>

<221> CDS

<222> (1)..(528)

<400> 17

atg gcc aat acc aaa tat aac aaa gag ttc ctg ctg tac ctg gcc ggc	48
Met Ala Asn Thr Lys Tyr Asn Lys Glu Phe Leu Leu Tyr Leu Ala Gly	
1 5 10 15	

ttt gtg gac ggt gac ggt agc atc atc gct cag att aaa cca aac cag	96
Phe Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Lys Pro Asn Gln	
20 25 30	

tct tat aag ttt aaa cat cag cta agc ttg acc ttt aag gtg act caa	144
Ser Tyr Lys Phe Lys His Gln Leu Ser Leu Thr Phe Val Thr Gln	
35 40 45	

aag acc cag cgc cgt tgg ttt ctg gac aaa cta gtg gat gaa att ggc	192
Lys Thr Gln Arg Arg Trp Phe Leu Asp Lys Leu Val Asp Glu Ile Gly	
50 55 60	

gtt ggt tac gta acg gat ggg gga tcc gtt tcc aac tac atc tta agc	240
Val Gly Tyr Val Thr Asp Gly Gly Ser Val Ser Asn Tyr Ile Leu Ser	
65 70 75 80	

gaa atc aag ccg ctg cac aac ttc ctg act caa ctg cag ccg ttt ctg	288
Glu Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu	
85 90 95	

aaa ctg aaa cag aaa cag gca aac ctg gtt ctg aaa att atc gaa cag	336
Lys Leu Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Glu Gln	
100 105 110	

ctg ccg tct gca aaa gaa tcc ccg gac aaa ttc ctg gaa gtt tgt acc	384
Leu Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr	
115 120 125	

tgg gtg gat cag att gca gct ctg aac gat tct aag acg cgt aaa acc	432
Trp Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr	
130 135 140	

act tct gaa acc gtt cgt gct gtg ctg gac agc ctg agc gag aag aag	480
Thr Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys	
145 150 155 160	

aaa tcc tcc ccg gcg gcc gca ctc gag cac cac cac cac cac tga	528
Lys Ser Ser Pro Ala Ala Ala Leu Glu His His His His His His	
165 170 175	

<210> 18

<211> 175

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 18

Met Ala Asn Thr Lys Tyr Asn Lys Glu Phe Leu Leu Tyr Leu Ala Gly
1 5 10 15

Phe Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Lys Pro Asn Gln
 20 25 30
 Ser Tyr Lys Phe Lys His Gln Leu Ser Leu Thr Phe Lys Val Thr Gln
 35 40 45
 Lys Thr Gln Arg Arg Trp Phe Leu Asp Lys Leu Val Asp Glu Ile Gly
 50 55 60
 Val Gly Tyr Val Thr Asp Gly Gly Ser Val Ser Asn Tyr Ile Leu Ser
 65 70 75 80
 Glu Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu
 85 90 95
 Lys Leu Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Glu Gln
 100 105 110
 Leu Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr
 115 120 125
 Trp Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr
 130 135 140
 Thr Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys
 145 150 155 160
 Lys Ser Ser Pro Ala Ala Ala Leu Glu His His His His His His
 165 170 175

<210> 19
 <211> 528
 <212> DNA
 <213> 人工序列

<220>
 <223> QAN-A1 变体

<220>
 <221> CDS
 <222> (1)..(528)

<400> 19	
atg gcc aat acc aaa tat aac agg cgg ttc ctg ctg tac ctg gcc ggc	48
Met Ala Asn Thr Lys Tyr Asn Arg Arg Phe Leu Leu Tyr Leu Ala Gly	
1 5 10 15	
ttt gtg gac ggt gac ggt agc atc atc gct cag att aaa cca aac cag	96
Phe Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Lys Pro Asn Gln	
20 25 30	
tct tat aag ttt aaa cat cag cta agc ttg acc ttt cag gtg act caa	144
Ser Tyr Lys Phe Lys His Gln Leu Ser Leu Thr Phe Gln Val Thr Gln	
35 40 45	

aag acc cag cgc cgt tgg ttt ctg gac aaa cta gtg gat aga att ggc Lys Thr Gln Arg Arg Trp Phe Leu Asp Lys Leu Val Asp Arg Ile Gly 50 55 60	192
gtt ggt tac gta gct gat aat gga tcc gtt tcc aac tac atc tta agc Val Gly Tyr Val Ala Asp Asn Gly Ser Val Ser Asn Tyr Ile Leu Ser 65 70 75 80	240
gaa atc aag ccg ctg cac aac ttc ctg act caa ctg cag ccg ttt ctg Glu Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu 85 90 95	288
aga ttc aaa cag aaa cag gca aac ctg gtt ctg aaa att atc gaa cag Arg Phe Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Glu Gln 100 105 110	336
ctg ccg tct gca aaa gaa tcc ccg gac aaa ttc ctg gaa gtt tgt acc Leu Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr 115 120 125	384
tgg gtg gat cag att gca gct ctg aac gat tct aag acg cgt aaa acc Trp Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr 130 135 140	432
act tct gaa acc gtt cgt gct gtg ctg gac agc ctg agc gag aag aag Thr Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys 145 150 155 160	480
aaa tcc tcc ccg gcg gcc gca ctc gag cac cac cac cac cac tga Lys Ser Ser Pro Ala Ala Ala Leu Glu His His His His His His 165 170 175	528

<210> 20
 <211> 175
 <212> PRT
 <213> 人工序列

<220>
 <223> 合成构建体

<400> 20

Met Ala Asn Thr Lys Tyr Asn Arg Arg Phe Leu Leu Tyr Leu Ala Gly 1 5 10 15
Phe Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Lys Pro Asn Gln 20 25 30
Ser Tyr Lys Phe Lys His Gln Leu Ser Leu Thr Phe Gln Val Thr Gln 35 40 45
Lys Thr Gln Arg Arg Trp Phe Leu Asp Lys Leu Val Asp Arg Ile Gly 50 55 60
Val Gly Tyr Val Ala Asp Asn Gly Ser Val Ser Asn Tyr Ile Leu Ser 65 70 75 80
Glu Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu 85 90 95
Arg Phe Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Glu Gln

100	105	110
Leu Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr 115 120 125		
Trp Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr 130 135 140		
Thr Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys 145 150 155 160		
Lys Ser Ser Pro Ala Ala Ala Leu Glu His His His His His His 165 170 175		

<210> 21
 <211> 528
 <212> DNA
 <213> 人工序列

<220>
 <223> QAN-A2 变体

<220>
 <221> CDS
 <222> (1)..(528)

<400> 21 atg gcc aat acc aaa tat aac agg cgg ttc ctg ctg tac ctg gcc ggc Met Ala Asn Thr Lys Tyr Asn Arg Arg Phe Leu Leu Tyr Leu Ala Gly 1 5 10 15	48
ttt gtg gac ggt gac ggt agc atc atc gct cag att aaa cca aac cag Phe Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Lys Pro Asn Gln 20 25 30	96
tct tat aag ttt aaa cat cag cta agc ttg acc ttt cag gtg act caa Ser Tyr Lys Phe Lys His Gln Leu Ser Leu Thr Phe Gln Val Thr Gln 35 40 45	144
aag acc cag cgc cgt tgg tgg ctg gac aaa cta gtg gat aga att ggc Lys Thr Gln Arg Arg Trp Trp Leu Asp Lys Leu Val Asp Arg Ile Gly 50 55 60	192
gtt ggt tac gta gct gat aat gga tcc gtt tcc aac tac atc tta agc Val Gly Tyr Val Ala Asp Asn Gly Ser Val Ser Asn Tyr Ile Leu Ser 65 70 75 80	240
gaa atc aag ccg ctg cac aac ttc ctg act caa ctg cag ccg ttt ctg Glu Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu 85 90 95	288
aga ttc aaa cag aaa cag gca aac ctg gtt ctg aaa att atc gaa cag Arg Phe Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Glu Gln 100 105 110	336
ctg ccg tct gca aaa gaa tcc ccg gac aaa ttc ctg gaa gtt tgt acc Leu Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr 115 120 125	384
tgg gtg gat cag att gca gct ctg aac gat tct aag acg cgt aaa acc Trp Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr 130 135 140	432

act	tct	gaa	acc	gtt	cgt	gct	gtg	ctg	gac	agc	ctg	agc	gag	aag	aag	480
Thr	Ser	Glu	Thr	Val	Arg	Ala	Val	Leu	Asp	Ser	Leu	Ser	Glu	Lys	Lys	
145					150					155					160	

aaa	tcc	tcc	ccg	gcg	gcc	gca	ctc	gag	cac	cac	cac	cac	cac	cac	tga	528
Lys	Ser	Ser	Pro	Ala	Ala	Ala	Leu	Glu	His	His	His	His	His	His		
				165					170					175		

<210> 22
 <211> 175
 <212> PRT
 <213> 人工序列

<220>
 <223> 合成构建体

<400> 22

Met	Ala	Asn	Thr	Lys	Tyr	Asn	Arg	Arg	Phe	Leu	Leu	Tyr	Leu	Ala	Gly
1				5					10					15	

Phe	Val	Asp	Gly	Asp	Gly	Ser	Ile	Ile	Ala	Gln	Ile	Lys	Pro	Asn	Gln
		20						25					30		

Ser	Tyr	Lys	Phe	Lys	His	Gln	Leu	Ser	Leu	Thr	Phe	Gln	Val	Thr	Gln
		35					40					45			

Lys	Thr	Gln	Arg	Arg	Trp	Trp	Leu	Asp	Lys	Leu	Val	Asp	Arg	Ile	Gly
	50					55					60				

Val	Gly	Tyr	Val	Ala	Asp	Asn	Gly	Ser	Val	Ser	Asn	Tyr	Ile	Leu	Ser
65					70					75					80

Glu	Ile	Lys	Pro	Leu	His	Asn	Phe	Leu	Thr	Gln	Leu	Gln	Pro	Phe	Leu
				85					90					95	

Arg	Phe	Lys	Gln	Lys	Gln	Ala	Asn	Leu	Val	Leu	Lys	Ile	Ile	Glu	Gln
			100					105						110	

Leu	Pro	Ser	Ala	Lys	Glu	Ser	Pro	Asp	Lys	Phe	Leu	Glu	Val	Cys	Thr
		115					120					125			

Trp	Val	Asp	Gln	Ile	Ala	Ala	Leu	Asn	Asp	Ser	Lys	Thr	Arg	Lys	Thr
	130					135					140				

Thr	Ser	Glu	Thr	Val	Arg	Ala	Val	Leu	Asp	Ser	Leu	Ser	Glu	Lys	Lys
145					150					155					160

Lys	Ser	Ser	Pro	Ala	Ala	Ala	Leu	Glu	His	His	His	His	His	His	
				165					170					175	

<210> 23
 <211> 528
 <212> DNA
 <213> 人工序列

<220>

<223> KTG-A1 变体

<220>

<221> CDS

<222> (1)..(528)

<400> 23

atg gcc aat acc aaa tat aac agg cgg ttc ctg ctg tac ctg gcc ggc	48
Met Ala Asn Thr Lys Tyr Asn Arg Arg Phe Leu Leu Tyr Leu Ala Gly	
1 5 10 15	

ttt gtg gac ggt gac ggt agc atc atc gct cag att aaa cca aac cag	96
Phe Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Lys Pro Asn Gln	
20 25 30	

tct tat aag ttt aaa cat cag cta agc ttg acc ttt aag gtg act caa	144
Ser Tyr Lys Phe Lys His Gln Leu Ser Leu Thr Phe Lys Val Thr Gln	
35 40 45	

aag acc cag cgc cgt tgg ttt ctg gac aaa cta gtg gat aga att ggc	192
Lys Thr Gln Arg Arg Trp Phe Leu Asp Lys Leu Val Asp Arg Ile Gly	
50 55 60	

gtt ggt tac gta acg gat ggg gga tcc gtt tcc aac tac atc tta agc	240
Val Gly Tyr Val Thr Asp Gly Gly Ser Val Ser Asn Tyr Ile Leu Ser	
65 70 75 80	

gaa atc aag ccg ctg cac aac ttc ctg act caa ctg cag ccg ttt ctg	288
Glu Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu	
85 90 95	

aga ttc aaa cag aaa cag gca aac ctg gtt ctg aaa att atc gaa cag	336
Arg Phe Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Glu Gln	
100 105 110	

ctg ccg tct gca aaa gaa tcc ccg gac aaa ttc ctg gaa gtt tgt acc	384
Leu Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr	
115 120 125	

tgg gtg gat cag att gca gct ctg aac gat tct aag acg cgt aaa acc	432
Trp Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr	
130 135 140	

act tct gaa acc gtt cgt gct gtg ctg gac agc ctg agc gag aag aag	480
Thr Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys	
145 150 155 160	

aaa tcc tcc ccg gcg gcc gca ctc gag cac cac cac cac cac cac tga	528
Lys Ser Ser Pro Ala Ala Ala Leu Glu His His His His His His His	
165 170 175	

<210> 24

<211> 175

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 24

Met Ala Asn Thr Lys Tyr Asn Arg Arg Phe Leu Leu Tyr Leu Ala Gly
1 5 10 15

Phe Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Lys Pro Asn Gln
 20 25 30
 Ser Tyr Lys Phe Lys His Gln Leu Ser Leu Thr Phe Lys Val Thr Gln
 35 40 45
 Lys Thr Gln Arg Arg Trp Phe Leu Asp Lys Leu Val Asp Arg Ile Gly
 50 55 60
 Val Gly Tyr Val Thr Asp Gly Gly Ser Val Ser Asn Tyr Ile Leu Ser
 65 70 75 80
 Glu Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu
 85 90 95
 Arg Phe Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Glu Gln
 100 105 110
 Leu Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr
 115 120 125
 Trp Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr
 130 135 140
 Thr Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys
 145 150 155 160
 Lys Ser Ser Pro Ala Ala Ala Leu Glu His His His His His His
 165 170 175

<210> 25
 <211> 528
 <212> DNA
 <213> 人工序列

<220>
 <223> KTG-A2 变体

<220>
 <221> CDS
 <222> (1)..(528)

<400> 25	
atg gcc aat acc aaa tat aac agg cgg ttc ctg ctg tac ctg gcc ggc	48
Met Ala Asn Thr Lys Tyr Asn Arg Arg Phe Leu Leu Tyr Leu Ala Gly	
1 5 10 15	
ttt gtg gac ggt gac ggt agc atc atc gct cag att aaa cca aac cag	96
Phe Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Lys Pro Asn Gln	
20 25 30	
tct tat aag ttt aaa cat cag cta agc ttg acc ttt aag gtg act caa	144
Ser Tyr Lys Phe Lys His Gln Leu Ser Leu Thr Phe Lys Val Thr Gln	
35 40 45	
aag acc cag cgc cgt tgg tgg ctg gac aaa cta gtg gat aga att ggc	192

Lys	Thr	Gln	Arg	Arg	Trp	Trp	Leu	Asp	Lys	Leu	Val	Asp	Arg	Ile	Gly	
50						55					60					
gtt	ggt	tac	gta	acg	gat	ggg	gga	tcc	gtt	tcc	aac	tac	atc	tta	agc	240
Val	Gly	Tyr	Val	Thr	Asp	Gly	Gly	Ser	Val	Ser	Asn	Tyr	Ile	Leu	Ser	
65					70					75					80	
gaa	atc	aag	ccg	ctg	cac	aac	ttc	ctg	act	caa	ctg	cag	ccg	ttt	ctg	288
Glu	Ile	Lys	Pro	Leu	His	Asn	Phe	Leu	Thr	Gln	Leu	Gln	Pro	Phe	Leu	
				85					90					95		
aga	ttc	aaa	cag	aaa	cag	gca	aac	ctg	gtt	ctg	aaa	att	atc	gaa	cag	336
Arg	Phe	Lys	Gln	Lys	Gln	Ala	Asn	Leu	Val	Leu	Lys	Ile	Ile	Glu	Gln	
			100					105					110			
ctg	ccg	tct	gca	aaa	gaa	tcc	ccg	gac	aaa	ttc	ctg	gaa	gtt	tgt	acc	384
Leu	Pro	Ser	Ala	Lys	Glu	Ser	Pro	Asp	Lys	Phe	Leu	Glu	Val	Cys	Thr	
		115					120					125				
tgg	gtg	gat	cag	att	gca	gct	ctg	aac	gat	tct	aag	acg	cgt	aaa	acc	432
Trp	Val	Asp	Gln	Ile	Ala	Ala	Leu	Asn	Asp	Ser	Lys	Thr	Arg	Lys	Thr	
	130					135					140					
act	tct	gaa	acc	gtt	cgt	gct	gtg	ctg	gac	agc	ctg	agc	gag	aag	aag	480
Thr	Ser	Glu	Thr	Val	Arg	Ala	Val	Leu	Asp	Ser	Leu	Ser	Glu	Lys	Lys	
145					150					155					160	
aaa	tcc	tcc	ccg	gcg	gcc	gca	ctc	gag	cac	cac	cac	cac	cac	cac	tga	528
Lys	Ser	Ser	Pro	Ala	Ala	Ala	Leu	Glu	His	His	His	His	His	His	His	
				165					170					175		

<210> 26
 <211> 175
 <212> PRT
 <213> 人工序列

<220>
 <223> 合成构建体

<400> 26

Met	Ala	Asn	Thr	Lys	Tyr	Asn	Arg	Arg	Phe	Leu	Leu	Tyr	Leu	Ala	Gly
1				5					10					15	
Phe	Val	Asp	Gly	Asp	Gly	Ser	Ile	Ile	Ala	Gln	Ile	Lys	Pro	Asn	Gln
			20					25					30		
Ser	Tyr	Lys	Phe	Lys	His	Gln	Leu	Ser	Leu	Thr	Phe	Lys	Val	Thr	Gln
		35					40					45			
Lys	Thr	Gln	Arg	Arg	Trp	Trp	Leu	Asp	Lys	Leu	Val	Asp	Arg	Ile	Gly
		50				55					60				
Val	Gly	Tyr	Val	Thr	Asp	Gly	Gly	Ser	Val	Ser	Asn	Tyr	Ile	Leu	Ser
65					70					75				80	
Glu	Ile	Lys	Pro	Leu	His	Asn	Phe	Leu	Thr	Gln	Leu	Gln	Pro	Phe	Leu
				85					90					95	
Arg	Phe	Lys	Gln	Lys	Gln	Ala	Asn	Leu	Val	Leu	Lys	Ile	Ile	Glu	Gln
			100					105					110		

Leu Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr
115 120 125

Trp Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr
130 135 140

Thr Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys
145 150 155 160

Lys Ser Ser Pro Ala Ala Ala Leu Glu His His His His His His
165 170 175

<210> 27
<211> 528
<212> DNA
<213> 人工序列

<220>
<223> QAN-B3 变体

<220>
<221> CDS
<222> (1)..(528)

<400> 27
atg gcc aat acc aaa tat aac gaa gag ttc ctg ctg tac ctg gcc ggc 48
Met Ala Asn Thr Lys Tyr Asn Glu Glu Phe Leu Leu Tyr Leu Ala Gly
1 5 10 15
ttt gtg gac ggt gac ggt agc atc atc gct cag att aaa cca aac cag 96
Phe Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Lys Pro Asn Gln
20 25 30
tct tat aag ttt aaa cat cag cta agc ttg acc ttt cag gtg act caa 144
Ser Tyr Lys Phe Lys His Gln Leu Ser Leu Thr Phe Gln Val Thr Gln
35 40 45
aag acc cag cgc cgt tgg ggt ctg gac aaa atg gtg gat gaa att ggc 192
Lys Thr Gln Arg Arg Trp Gly Leu Asp Lys Met Val Asp Glu Ile Gly
50 55 60
gtt ggt tac gta gct gat aat gga tcc gtt tcc aac tac atc tta agc 240
Val Gly Tyr Val Ala Asp Asn Gly Ser Val Ser Asn Tyr Ile Leu Ser
65 70 75
gaa atc aag ccg ctg cac aac ttc ctg act caa ctg cag ccg ttt ctg 288
Glu Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu
85 90 95
gaa ctg aaa cag aaa cag gca aac ctg gtt ctg aaa att atc gaa cag 336
Glu Leu Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Glu Gln
100 105 110
ctg ccg tct gca aaa gaa tcc ccg gac aaa ttc ctg gaa gtt tgt acc 384
Leu Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr
115 120 125
tgg gtg gat cag att gca gct ctg aac gat tct aag acg cgt aaa acc 432
Trp Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr
130 135 140

act tct gaa acc gtt cgt gct gtg ctg gac agc ctg agc gag aag aag 480
 Thr Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys
 145 150 155 160

aaa tcc tcc ccg gcg gcc gca ctc gag cac cac cac cac cac cac tga 528
 Lys Ser Ser Pro Ala Ala Ala Leu Glu His His His His His His
 165 170 175

<210> 28
 <211> 175
 <212> PRT
 <213> 人工序列

<220>
 <223> 合成构建体

<400> 28

Met Ala Asn Thr Lys Tyr Asn Glu Glu Phe Leu Leu Tyr Leu Ala Gly
 1 5 10 15

Phe Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Lys Pro Asn Gln
 20 25 30

Ser Tyr Lys Phe Lys His Gln Leu Ser Leu Thr Phe Gln Val Thr Gln
 35 40 45

Lys Thr Gln Arg Arg Trp Gly Leu Asp Lys Met Val Asp Glu Ile Gly
 50 55 60

Val Gly Tyr Val Ala Asp Asn Gly Ser Val Ser Asn Tyr Ile Leu Ser
 65 70 75 80

Glu Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu
 85 90 95

Glu Leu Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Glu Gln
 100 105 110

Leu Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr
 115 120 125

Trp Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr
 130 135 140

Thr Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys
 145 150 155 160

Lys Ser Ser Pro Ala Ala Ala Leu Glu His His His His His His
 165 170 175

<210> 29
 <211> 528
 <212> DNA
 <213> 人工序列

<220>

<223> QAN-B4 变体

<220>

<221> CDS

<222> (1)..(528)

<400> 29

atg	gcc	aat	acc	aaa	tat	aac	gaa	gag	ttc	ctg	ctg	tac	ctg	gcc	ggc	48
Met	Ala	Asn	Thr	Lys	Tyr	Asn	Glu	Glu	Phe	Leu	Leu	Tyr	Leu	Ala	Gly	
1				5					10					15		

ttt	gtg	gac	ggg	gac	ggg	agc	atc	atc	gct	cag	att	aaa	cca	aac	cag	96
Phe	Val	Asp	Gly	Asp	Gly	Ser	Ile	Ile	Ala	Gln	Ile	Lys	Pro	Asn	Gln	
		20					25					30				

tct	tat	aag	ttt	aaa	cat	cag	cta	agc	ttg	acc	ttt	cag	gtg	act	caa	144
Ser	Tyr	Lys	Phe	Lys	His	Gln	Leu	Ser	Leu	Thr	Phe	Gln	Val	Thr	Gln	
		35				40					45					

aag	acc	cag	cgc	cgt	tgg	ggg	ctg	gac	aaa	atg	gtg	gat	gaa	att	ggc	192
Lys	Thr	Gln	Arg	Arg	Trp	Gly	Leu	Asp	Lys	Met	Val	Asp	Glu	Ile	Gly	
	50				55						60					

gtt	ggg	tac	gta	gct	gat	aat	gga	tcc	gtt	tcc	aac	tac	atc	tta	agc	240
Val	Gly	Tyr	Val	Ala	Asp	Asn	Gly	Ser	Val	Ser	Asn	Tyr	Ile	Leu	Ser	
65				70					75					80		

gaa	atc	aag	ccg	ctg	cac	aac	ttc	ctg	act	caa	ctg	cag	ccg	ttt	ctg	288
Glu	Ile	Lys	Pro	Leu	His	Asn	Phe	Leu	Thr	Gln	Leu	Gln	Pro	Phe	Leu	
			85					90						95		

gaa	ggg	aaa	cag	aaa	cag	gca	aac	ctg	gtt	ctg	aaa	att	atc	gaa	cag	336
Glu	Gly	Lys	Gln	Lys	Gln	Ala	Asn	Leu	Val	Leu	Lys	Ile	Ile	Glu	Gln	
		100					105					110				

ctg	ccg	tct	gca	aaa	gaa	tcc	ccg	gac	aaa	ttc	ctg	gaa	gtt	tgt	acc	384
Leu	Pro	Ser	Ala	Lys	Glu	Ser	Pro	Asp	Lys	Phe	Leu	Glu	Val	Cys	Thr	
		115				120						125				

tgg	gtg	gat	cag	att	gca	gct	ctg	aac	gat	tct	aag	acg	cgt	aaa	acc	432
Trp	Val	Asp	Gln	Ile	Ala	Ala	Leu	Asn	Asp	Ser	Lys	Thr	Arg	Lys	Thr	
	130				135						140					

act	tct	gaa	acc	gtt	cgt	gct	gtg	ctg	gac	agc	ctg	agc	gag	aag	aag	480
Thr	Ser	Glu	Thr	Val	Arg	Ala	Val	Leu	Asp	Ser	Leu	Ser	Glu	Lys	Lys	
145				150					155					160		

aaa	tcc	tcc	ccg	gcg	gcc	gca	ctc	gag	cac	cac	cac	cac	cac	cac	tga	528
Lys	Ser	Ser	Pro	Ala	Ala	Ala	Leu	Glu	His	His	His	His	His	His		
				165					170					175		

<210> 30

<211> 175

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 30

Met	Ala	Asn	Thr	Lys	Tyr	Asn	Glu	Glu	Phe	Leu	Leu	Tyr	Leu	Ala	Gly
1				5					10					15	

Phe Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Lys Pro Asn Gln
20 25 30

Ser Tyr Lys Phe Lys His Gln Leu Ser Leu Thr Phe Gln Val Thr Gln
35 40 45

Lys Thr Gln Arg Arg Trp Gly Leu Asp Lys Met Val Asp Glu Ile Gly
50 55 60

Val Gly Tyr Val Ala Asp Asn Gly Ser Val Ser Asn Tyr Ile Leu Ser
65 70 75 80

Glu Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu
85 90 95

Glu Gly Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Glu Gln
100 105 110

Leu Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr
115 120 125

Trp Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr
130 135 140

Thr Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys
145 150 155 160

Lys Ser Ser Pro Ala Ala Ala Leu Glu His His His His His His
165 170 175

<210> 31
<211> 528
<212> DNA
<213> 人工序列

<220>
<223> KTG-B3 变体

<220>
<221> CDS
<222> (1)..(528)

<400> 31	
atg gcc aat acc aaa tat aac gaa gag ttc ctg ctg tac ctg gcc ggc	48
Met Ala Asn Thr Lys Tyr Asn Glu Glu Phe Leu Leu Tyr Leu Ala Gly	
1 5 10 15	
ttt gtg gac ggt gac ggt agc atc atc gct cag att aaa cca aac cag	96
Phe Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Lys Pro Asn Gln	
20 25 30	
tct tat aag ttt aaa cat cag cta agc ttg acc ttt aag gtg act caa	144
Ser Tyr Lys Phe Lys His Gln Leu Ser Leu Thr Phe Lys Val Thr Gln	
35 40 45	
aag acc cag cgc cgt tgg ggt ctg gac aaa atg gtg gat gaa att ggc	192
Lys Thr Gln Arg Arg Trp Gly Leu Asp Lys Met Val Asp Glu Ile Gly	

50	55	60	
gtt ggt tac gta acg gat ggg gga tcc gtt tcc aac tac atc tta agc Val Gly Tyr Val Thr Asp Gly Gly Ser Val Ser Asn Tyr Ile Leu Ser 65 70 75 80			240
gaa atc aag ccg ctg cac aac ttc ctg act caa ctg cag ccg ttt ctg Glu Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu 85 90 95			288
gaa ctg aaa cag aaa cag gca aac ctg gtt ctg aaa att atc gaa cag Glu Leu Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Glu Gln 100 105 110			336
ctg ccg tct gca aaa gaa tcc ccg gac aaa ttc ctg gaa gtt tgt acc Leu Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr 115 120 125			384
tgg gtg gat cag att gca gct ctg aac gat tct aag acg cgt aaa acc Trp Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr 130 135 140			432
act tct gaa acc gtt cgt gct gtg ctg gac agc ctg agc gag aag aag Thr Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys 145 150 155 160			480
aaa tcc tcc ccg gcg gcc gca ctc gag cac cac cac cac cac tga Lys Ser Ser Pro Ala Ala Ala Leu Glu His His His His His His 165 170 175			528

<210> 32
 <211> 175
 <212> PRT
 <213> 人工序列

<220>
 <223> 合成构建体

<400> 32

Met Ala Asn Thr Lys Tyr Asn Glu Glu Phe Leu Leu Tyr Leu Ala Gly
 1 5 10 15

Phe Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Lys Pro Asn Gln
 20 25 30

Ser Tyr Lys Phe Lys His Gln Leu Ser Leu Thr Phe Lys Val Thr Gln
 35 40 45

Lys Thr Gln Arg Arg Trp Gly Leu Asp Lys Met Val Asp Glu Ile Gly
 50 55 60

Val Gly Tyr Val Thr Asp Gly Gly Ser Val Ser Asn Tyr Ile Leu Ser
 65 70 75 80

Glu Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu
 85 90 95

Glu Leu Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Glu Gln
 100 105 110

Leu Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr
115 120 125

Trp Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr
130 135 140

Thr Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys
145 150 155 160

Lys Ser Ser Pro Ala Ala Ala Leu Glu His His His His His His
165 170 175

<210> 33
<211> 528
<212> DNA
<213> 人工序列

<220>
<223> KTG-B4 变体

<220>
<221> CDS
<222> (1)..(528)

<400> 33	
atg gcc aat acc aaa tat aac gaa gag ttc ctg ctg tac ctg gcc ggc	48
Met Ala Asn Thr Lys Tyr Asn Glu Glu Phe Leu Leu Tyr Leu Ala Gly	
1 5 10 15	
ttt gtg gac ggt gac ggt agc atc atc gct cag att aaa cca aac cag	96
Phe Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Lys Pro Asn Gln	
20 25 30	
tct tat aag ttt aaa cat cag cta agc ttg acc ttt aag gtg act caa	144
Ser Tyr Lys Phe Lys His Gln Leu Ser Leu Thr Phe Lys Val Thr Gln	
35 40 45	
aag acc cag cgc cgt tgg ggt ctg gac aaa atg gtg gat gaa att ggc	192
Lys Thr Gln Arg Arg Trp Gly Leu Asp Lys Met Val Asp Glu Ile Gly	
50 55 60	
gtt ggt tac gta acg gat ggg gga tcc gtt tcc aac tac atc tta agc	240
Val Gly Tyr Val Thr Asp Gly Gly Ser Val Ser Asn Tyr Ile Leu Ser	
65 70 75 80	
gaa atc aag ccg ctg cac aac ttc ctg act caa ctg cag ccg ttt ctg	288
Glu Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu	
85 90 95	
gaa ggg aaa cag aaa cag gca aac ctg gtt ctg aaa att atc gaa cag	336
Glu Gly Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Glu Gln	
100 105 110	
ctg ccg tct gca aaa gaa tcc ccg gac aaa ttc ctg gaa gtt tgt acc	384
Leu Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr	
115 120 125	
tgg gtg gat cag att gca gct ctg aac gat tct aag acg cgt aaa acc	432
Trp Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr	
130 135 140	
act tct gaa acc gtt cgt gct gtg ctg gac agc ctg agc gag aag aag	480

Thr	Ser	Glu	Thr	Val	Arg	Ala	Val	Leu	Asp	Ser	Leu	Ser	Glu	Lys	Lys	
145					150					155					160	
aaa	tcc	tcc	ccg	gcg	gcc	gca	ctc	gag	cac	cac	cac	cac	cac	cac	tga	528
Lys	Ser	Ser	Pro	Ala	Ala	Ala	Leu	Glu	His	His	His	His	His	His		
				165					170					175		

<210> 34
 <211> 175
 <212> PRT
 <213> 人工序列

<220>
 <223> 合成构建体

<400> 34

Met	Ala	Asn	Thr	Lys	Tyr	Asn	Glu	Glu	Phe	Leu	Leu	Tyr	Leu	Ala	Gly	
1				5					10					15		

Phe	Val	Asp	Gly	Asp	Gly	Ser	Ile	Ile	Ala	Gln	Ile	Lys	Pro	Asn	Gln	
		20						25					30			

Ser	Tyr	Lys	Phe	Lys	His	Gln	Leu	Ser	Leu	Thr	Phe	Lys	Val	Thr	Gln	
		35					40					45				

Lys	Thr	Gln	Arg	Arg	Trp	Gly	Leu	Asp	Lys	Met	Val	Asp	Glu	Ile	Gly	
	50					55					60					

Val	Gly	Tyr	Val	Thr	Asp	Gly	Gly	Ser	Val	Ser	Asn	Tyr	Ile	Leu	Ser	
65					70					75					80	

Glu	Ile	Lys	Pro	Leu	His	Asn	Phe	Leu	Thr	Gln	Leu	Gln	Pro	Phe	Leu	
				85					90					95		

Glu	Gly	Lys	Gln	Lys	Gln	Ala	Asn	Leu	Val	Leu	Lys	Ile	Ile	Glu	Gln	
			100					105					110			

Leu	Pro	Ser	Ala	Lys	Glu	Ser	Pro	Asp	Lys	Phe	Leu	Glu	Val	Cys	Thr	
		115					120					125				

Trp	Val	Asp	Gln	Ile	Ala	Ala	Leu	Asn	Asp	Ser	Lys	Thr	Arg	Lys	Thr	
	130					135					140					

Thr	Ser	Glu	Thr	Val	Arg	Ala	Val	Leu	Asp	Ser	Leu	Ser	Glu	Lys	Lys	
145					150					155					160	

Lys	Ser	Ser	Pro	Ala	Ala	Ala	Leu	Glu	His	His	His	His	His	His	His	
				165					170						175	

<210> 35
 <211> 37
 <212> DNA
 <213> 人工序列

<220>

<223> 引物

<400> 35
ctggacaaac tagtggatag aattggcggtt ggttacg 37

<210> 36
<211> 44
<212> DNA
<213> 人工序列

<220>
<223> 引物

<400> 36
tcaactgcag ccgtttctgg aagggaaaca gaaacaggca aacc 44

<210> 37
<211> 26
<212> DNA
<213> 人工序列

<220>
<223> 引物

<400> 37
gcaacttttag tgctgacaca tacagg 26

<210> 38
<211> 39
<212> DNA
<213> 人工序列

<220>
<223> 引物

<400> 38
tttgtccaga aaccaacggt cctgggtctt ttgagtcac 39

<210> 39
<211> 40
<212> DNA
<213> 人工序列

<220>
<223> 引物

<400> 39
gtgactcaaa agacccagga ccgttggttt ctggacaaac 40

<210> 40
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 引物

<400> 40
acaaccttga ttggagactt gacc 24

<210> 41
<211> 39

<212> DNA
<213> 人工序列

<220>
<223> 引物

<400> 41
ggttttacgc gtcttagaac ggttcagagc tgcaatctg

39

<210> 42
<211> 39
<212> DNA
<213> 人工序列

<220>
<223> 引物

<400> 42
cagattgcag ctctgaaccg ttctaagacg cgtaaaacc

39

<210> 43
<211> 163
<212> PRT
<213> 莱茵衣藻

<400> 43

Met Asn Thr Lys Tyr Asn Lys Glu Phe Leu Leu Tyr Leu Ala Gly Phe
1 5 10 15

Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Lys Pro Asn Gln Ser
20 25 30

Tyr Lys Phe Lys His Gln Leu Ser Leu Thr Phe Gln Val Thr Gln Lys
35 40 45

Thr Gln Arg Arg Trp Phe Leu Asp Lys Leu Val Asp Glu Ile Gly Val
50 55 60

Gly Tyr Val Arg Asp Arg Gly Ser Val Ser Asp Tyr Ile Leu Ser Glu
65 70 75 80

Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu Lys
85 90 95

Leu Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Glu Gln Leu
100 105 110

Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr Trp
115 120 125

Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr Thr
130 135 140

Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys Lys
145 150 155 160

Ser Ser Pro

<210>	44	
<211>	21	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	CMVfor 引物	
<400>	44	
	cgcaaattggg cggtaggcgt g	21
<210>	45	
<211>	35	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	K7ERev 引物	
<400>	45	
	gtacagcagg aactcttcgt tatatttggt attgg	35
<210>	46	
<211>	34	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	K7EFor 引物	
<400>	46	
	aataccaaat ataacgaaga gttcctgctg tacc	34
<210>	47	
<211>	24	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	V5 表位 Rev 引物	
<400>	47	
	cgtagaatcg agaccgagga gagg	24
<210>	48	
<211>	32	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	E8Krev 引物	
<400>	48	
	caggtagc aggaactttt tggtatattt gg	32
<210>	49	
<211>	34	
<212>	DNA	

<213>	人工序列	
<220>		
<223>	E8KFor 引物	
<400>	49	
	accaaata acaaaaagtt cctgctgtac ctgg	34
<210>	50	
<211>	34	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	E61RRev 引物	
<400>	50	
	gtaaccaacg ccaatacgat ccactagttt gtcc	34
<210>	51	
<211>	34	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	E61RFor 引物	
<400>	51	
	gacaaactag tggatcgat tggcgttggt tacg	34
<210>	52	
<211>	34	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	K96ERev 引物	
<400>	52	
	cctgtttctg tttcagttcc agaaacggct gcag	34
<210>	53	
<211>	34	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	K96EFor 引物	
<400>	53	
	ctgcagccgt ttctggaact gaaacagaaa cagg	34
<210>	54	
<211>	84	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	CCM2For 引物	
<400>	54	
	aagcagagct ctctggctaa ctagagaacc cactgcttac tggcttatcg accatggcca	60

ataccaaata taacaaagag ttcc	84
<210> 55	
<211> 38	
<212> DNA	
<213> 人工序列	
<220>	
<223> CCMRev 引物	
<400> 55	
ctgctctaga ttagtcggcc gccggggagg atttcttc	38
<210> 56	
<211> 38	
<212> DNA	
<213> 人工序列	
<220>	
<223> CreCterSacI 引物	
<400> 56	
tagacgagct cctacgggga ggattttcttc ttctcgct	38
<210> 57	
<211> 131	
<212> DNA	
<213> 人工序列	
<220>	
<223> RM2 引物	
<400> 57	
tatcggcccg tggatctgat aagtataatc aggctctgtc taaatacaac caagcactgt	60
ccaagtacaa tcaggccctg tctggtggag gcggttccaa caaagagttc ctgctgtatc	120
ttgctggatt t	131
<210> 58	
<211> 22	
<212> DNA	
<213> 人工序列	
<220>	
<223> C1221 靶标	
<400> 58	
caaaacgtcg tacgacgttt tg	22
<210> 59	
<211> 22	
<212> DNA	
<213> 人工序列	
<220>	
<223> 10GTT_P 靶标	
<400> 59	
cgttacgtcg tacgacgtaa cg	22
<210> 60	

<211>	22	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	5CAG_P 靶标	
<400>	60	
	caaaaccagg tacctggttt tg	22
<210>	61	
<211>	22	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	10TGG_P 靶标	
<400>	61	
	ctggacgtcg tacgacgtcc ag	22
<210>	62	
<211>	22	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	5GAG_P 靶标	
<400>	62	
	caaaacgagg tacctcgttt tg	22
<210>	63	
<211>	22	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	RAG1.10 靶标	
<400>	63	
	tgttctcagg tacctcagcc ag	22
<210>	64	
<211>	22	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	RAG1.10.2 靶标	
<400>	64	
	tgttctcagg tacctgagaa ca	22
<210>	65	
<211>	22	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	RAG1.10.3 靶标	
<400>	65	

ctggctgagg tacctcagcc ag

22

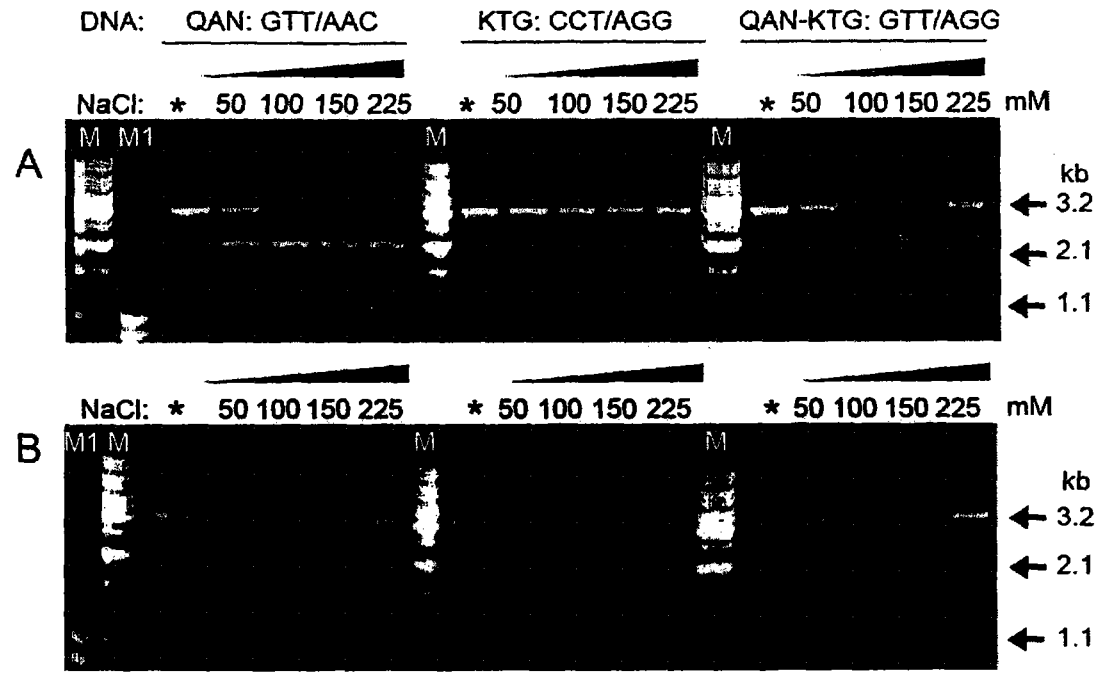


图2

QAN (SEQ ID NO : 15)

ATGgccaataaccaaatataacaagagttcctgctgtacctggccggctttgtggacggtgacggtagcatcatcgctcagattaaacc
aaaccagtccttataagtttaaacatcagctaagcttgacctttcaggtgactcaaaagaccagcgccgttggtttctggacaaactag
tggatgaaattggcggttggttacgtagctgataatggatccgtttccaactacatcttaagcgaaatcaagccgctgcacaacttctg
actcaactgcagccgtttctgaaactgaaacagaaacaggcaaacctgggtctgaaaattatcgaacagctgccgtctgcaaaagaatc
cccggaacaaattcctggaagtttgacctgggtggatcagattgcagctctgaacgattctaaagacgctgaaaccacttctgaaaccg
ttcgtgctgtgctggacagcctgagcgagaagaagaatcctccccgCGggccgcactcgagcaccaccaccaccaccactga

KTG (SEQ ID NO : 17)

ATGgccaataaccaaatataacaagagttcctgctgtacctggccggctttgtggacggtgacggtagcatcatcgctcagattaaacc
aaaccagtccttataagtttaaacatcagctaagcttgacctttcaggtgactcaaaagaccagcgccgttggtttctggacaaactag
tggatgaaattggcggttggttacgtaacggatgggggatccgtttccaactacatcttaagcgaaatcaagccgctgcacaacttctg
actcaactgcagccgtttctgaaactgaaacagaaacaggcaaacctgggtctgaaaattatcgaacagctgccgtctgcaaaagaatc
cccggaacaaattcctggaagtttgacctgggtggatcagattgcagctctgaacgattctaaagacgctgaaaccacttctgaaaccg
ttcgtgctgtgctggacagcctgagcgagaagaagaatcctccccgCGggccgcactcgagcaccaccaccaccaccactga

QAN_A1 (SEQ ID NO : 19)

ATGgccaataaccaaatataacaGGCGgttcctgctgtacctggccggctttgtggacggtgacggtagcatcatcgctcagattaaacc
aaaccagtccttataagtttaaacatcagctaagcttgacctttcaggtgactcaaaagaccagcgccgttggtttctggacaaactag
tggatAGaattggcggttggttacgtagctgataatggatccgtttccaactacatcttaagcgaaatcaagccgctgcacaacttctg
actcaactgcagccgtttctgaGaTtCaaacagaaacaggcaaacctgggtctgaaaattatcgaacagctgccgtctgcaaaagaatc
cccggaacaaattcctggaagtttgacctgggtggatcagattgcagctctgaacgattctaaagacgctgaaaccacttctgaaaccg
ttcgtgctgtgctggacagcctgagcgagaagaagaatcctccccgCGggccgcactcgagcaccaccaccaccaccactga

QAN_A2 (SEQ ID NO : 21)

ATGgccaataaccaaatataacaGGCGgttcctgctgtacctggccggctttgtggacggtgacggtagcatcatcgctcagattaaacc
aaaccagtccttataagtttaaacatcagctaagcttgacctttcaggtgactcaaaagaccagcgccgttggtGGtgggacaaactag
tggatAGaattggcggttggttacgtagctgataatggatccgtttccaactacatcttaagcgaaatcaagccgctgcacaacttctg
actcaactgcagccgtttctgaGaTtCaaacagaaacaggcaaacctgggtctgaaaattatcgaacagctgccgtctgcaaaagaatc
cccggaacaaattcctggaagtttgacctgggtggatcagattgcagctctgaacgattctaaagacgctgaaaccacttctgaaaccg
ttcgtgctgtgctggacagcctgagcgagaagaagaatcctccccgCGggccgcactcgagcaccaccaccaccaccactga

KTG_A1 (SEQ ID NO : 23)

ATGgccaataaccaaatataacaGGCGgttcctgctgtacctggccggctttgtggacggtgacggtagcatcatcgctcagattaaacc
aaaccagtccttataagtttaaacatcagctaagcttgacctttcaggtgactcaaaagaccagcgccgttggtttctggacaaactag
tggatAGaattggcggttggttacgtaacggatgggggatccgtttccaactacatcttaagcgaaatcaagccgctgcacaacttctg
actcaactgcagccgtttctgaGaTtCaaacagaaacaggcaaacctgggtctgaaaattatcgaacagctgccgtctgcaaaagaatc
cccggaacaaattcctggaagtttgacctgggtggatcagattgcagctctgaacgattctaaagacgctgaaaccacttctgaaaccg
ttcgtgctgtgctggacagcctgagcgagaagaagaatcctccccgCGggccgcactcgagcaccaccaccaccaccactga

KTG_A2 (SEQ ID NO : 25)

ATGgccaataaccaaatataacaGGCGgttcctgctgtacctggccggctttgtggacggtgacggtagcatcatcgctcagattaaacc
aaaccagtccttataagtttaaacatcagctaagcttgacctttcaggtgactcaaaagaccagcgccgttggtGGtctggacaaactag
tggatAGaattggcggttggttacgtaacggatgggggatccgtttccaactacatcttaagcgaaatcaagccgctgcacaacttctg
actcaactgcagccgtttctgaGaTtCaaacagaaacaggcaaacctgggtctgaaaattatcgaacagctgccgtctgcaaaagaatc
cccggaacaaattcctggaagtttgacctgggtggatcagattgcagctctgaacgattctaaagacgctgaaaccacttctgaaaccg
ttcgtgctgtgctggacagcctgagcgagaagaagaatcctccccgCGggccgcactcgagcaccaccaccaccaccactga

QAN_B3 (SEQ ID NO : 27)

ATGgccaataaccaaatataaacGaagagttcctgctgtacctggccggctttgtggacggtgacggtagcatcatcgctcagattaaacc
aaaccagtccttataagtttaaacatcagctaagcttgacctttcaggtgactcaaaagaccagcgccgttggtGGtctggacaaaAtGg
tggatgaaattggcggttggttacgtagctgataatggatccgtttccaactacatcttaagcgaaatcaagccgctgcacaacttctg
actcaactgcagccgtttctgGaactgaaacagaaacaggcaaacctgggtctgaaaattatcgaacagctgccgtctgcaaaagaatc
cccggaacaaattcctggaagtttgacctgggtggatcagattgcagctctgaacgattctaaagacgctgaaaccacttctgaaaccg
ttcgtgctgtgctggacagcctgagcgagaagaagaatcctccccgCGggccgcactcgagcaccaccaccaccaccactga

图3

QAN_B4 (SEQ ID NO : 29)

ATGgccaataccaaatataacGaagagttcctgctgtacctggccggccttctgtggacggtgacggtagcatcatcgctcagattaaacc
aaaccagtccttataagtttaaacatcagctaagcttgacctttcagggtgactcaaaagacccagcgccgttggGGtctggacaaaAtGg
tggatgaaattggcgttgggttacgtagctgataatggatccgtttccaactacatcttaagcgaaatcaagccgctgcacaacttcctg
actcaactgcagccgtttctgGaaGGgaaacagaaacaggcaaacctgggtctgaaaattatcgaacagctgccgtctgcaaaagaatc
ccgggacaaattcctggaagtttgacctgggtggatcagattgcagctctgaacgattctaagacgctgtaaaaccacttctgaaaccg
ttcgtgctgtgctggacagcctgagcgagaagaagaatcctccccgGCggccgcactcgagcaccaccaccaccactga

KTG_B3 (SEQ ID NO : 31)

ATGgccaataccaaatataacGaagagttcctgctgtacctggccggccttctgtggacggtgacggtagcatcatcgctcagattaaacc
aaaccagtccttataagtttaaacatcagctaagcttgacctttaagggtgactcaaaagacccagcgccgttggGGtctggacaaaAtGg
tggatgaaattggcgttgggttacgtaacggatgggggatccgtttccaactacatcttaagcgaaatcaagccgctgcacaacttcctg
actcaactgcagccgtttctgGaaGGgaaacagaaacaggcaaacctgggtctgaaaattatcgaacagctgccgtctgcaaaagaatc
ccgggacaaattcctggaagtttgacctgggtggatcagattgcagctctgaacgattctaagacgctgtaaaaccacttctgaaaccg
ttcgtgctgtgctggacagcctgagcgagaagaagaatcctccccgGCggccgcactcgagcaccaccaccaccactga

KTG_B4 (SEQ ID NO : 33)

ATGgccaataccaaatataacGaagagttcctgctgtacctggccggccttctgtggacggtgacggtagcatcatcgctcagattaaacc
aaaccagtccttataagtttaaacatcagctaagcttgacctttaagggtgactcaaaagacccagcgccgttggGGtctggacaaaAtGg
tggatgaaattggcgttgggttacgtaacggatgggggatccgtttccaactacatcttaagcgaaatcaagccgctgcacaacttcctg
actcaactgcagccgtttctgGaaGGgaaacagaaacaggcaaacctgggtctgaaaattatcgaacagctgccgtctgcaaaagaatc
ccgggacaaattcctggaagtttgacctgggtggatcagattgcagctctgaacgattctaagacgctgtaaaaccacttctgaaaccg
ttcgtgctgtgctggacagcctgagcgagaagaagaatcctccccgGCggccgcactcgagcaccaccaccaccactga

图3(续)

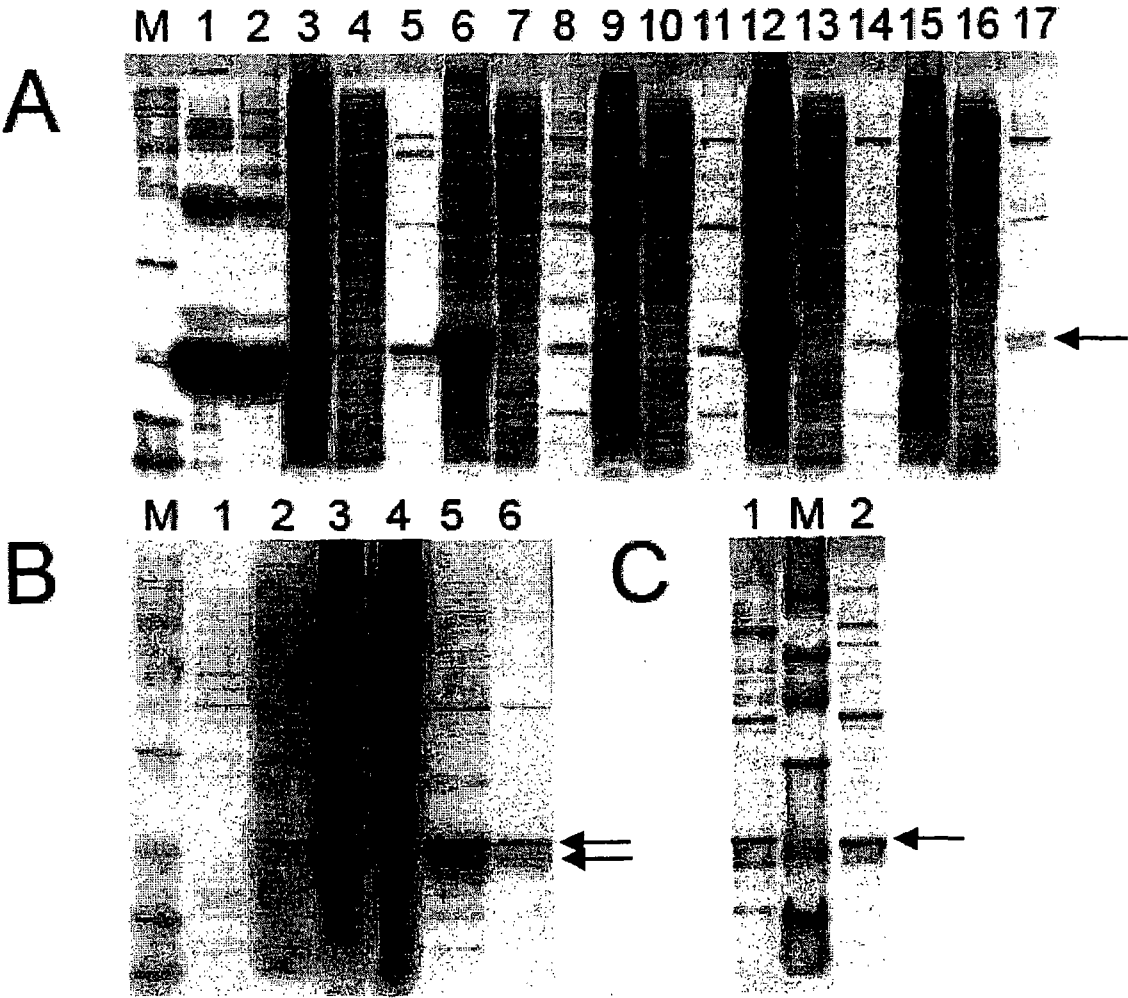


图 4

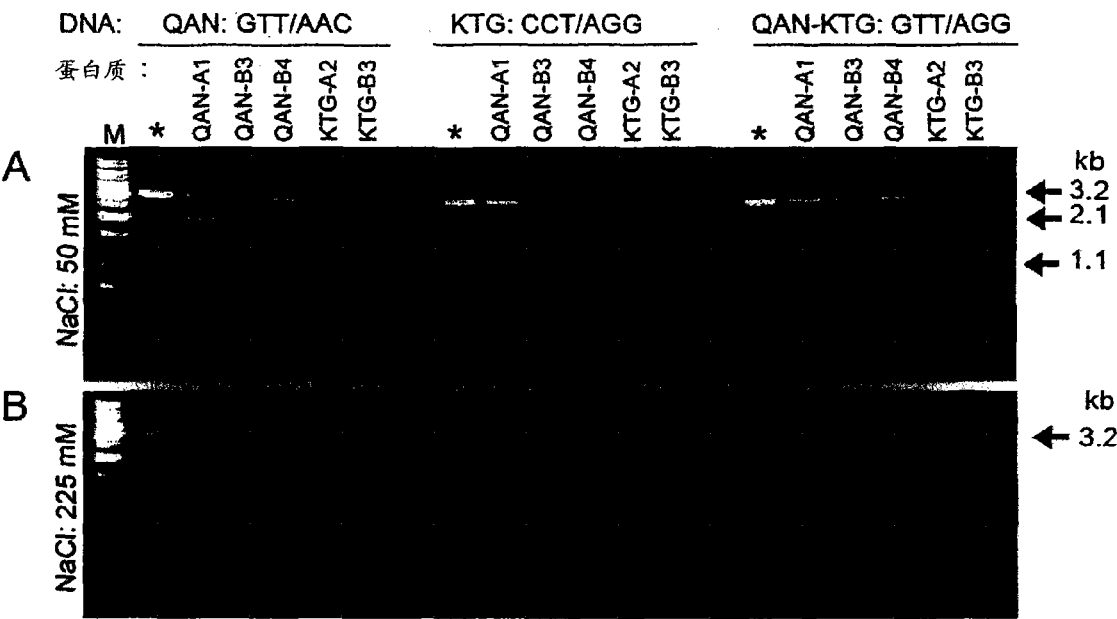


图 5

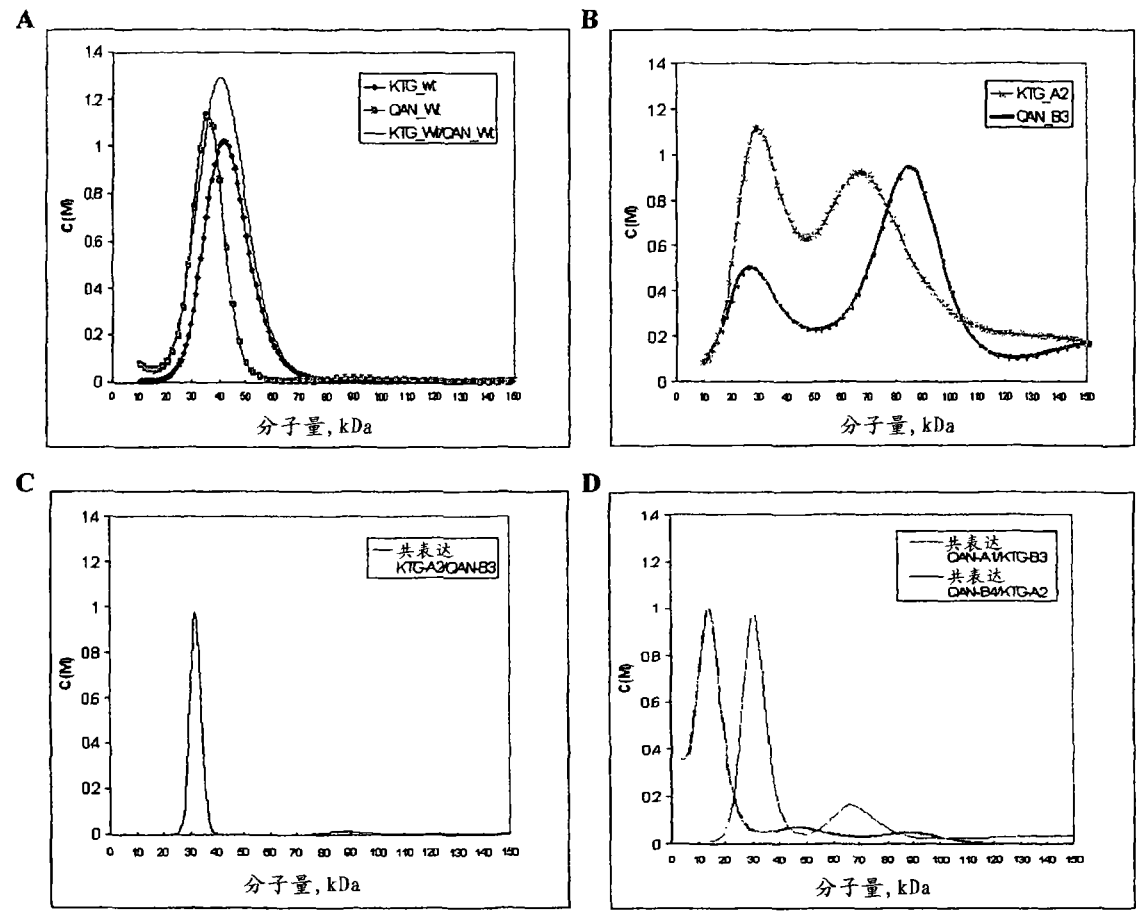


图6

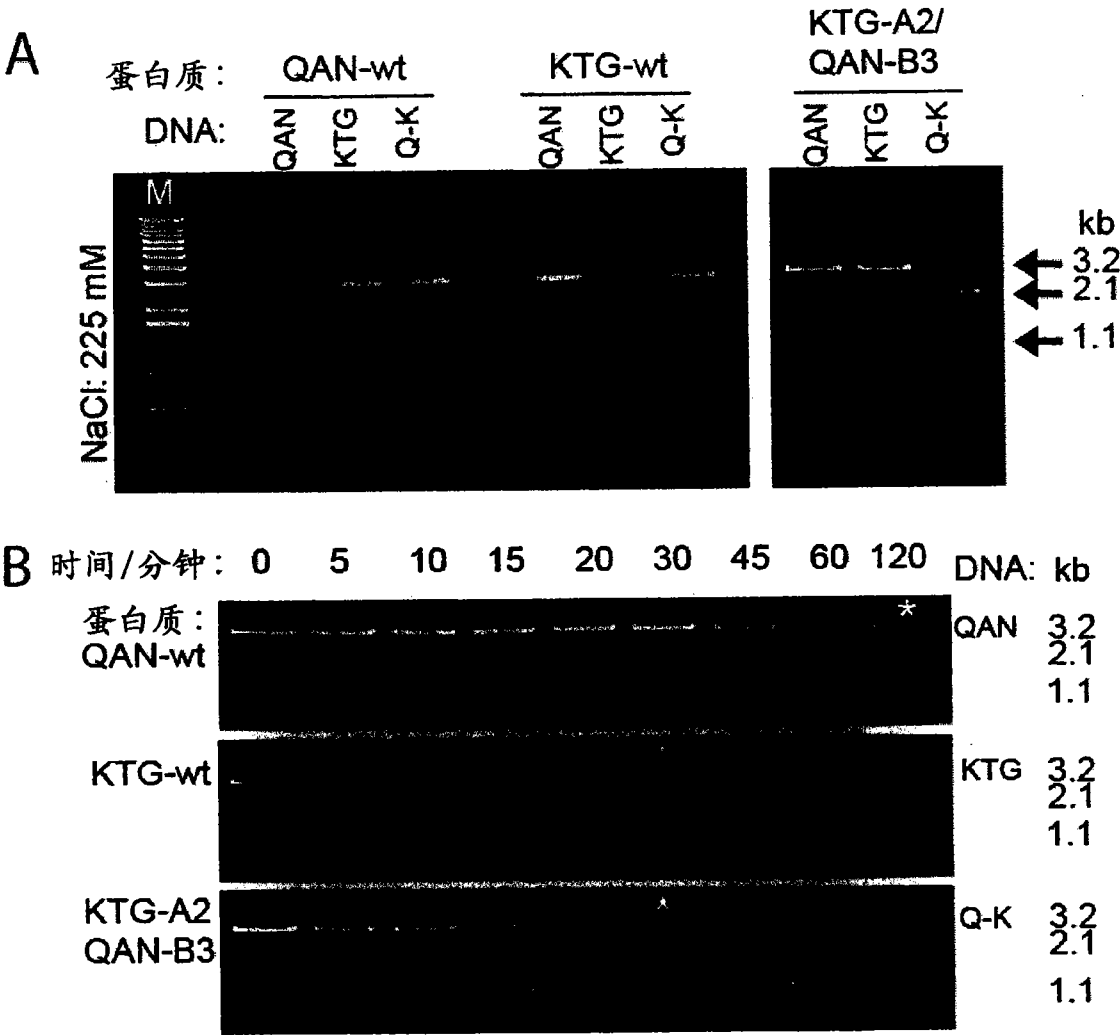


图7

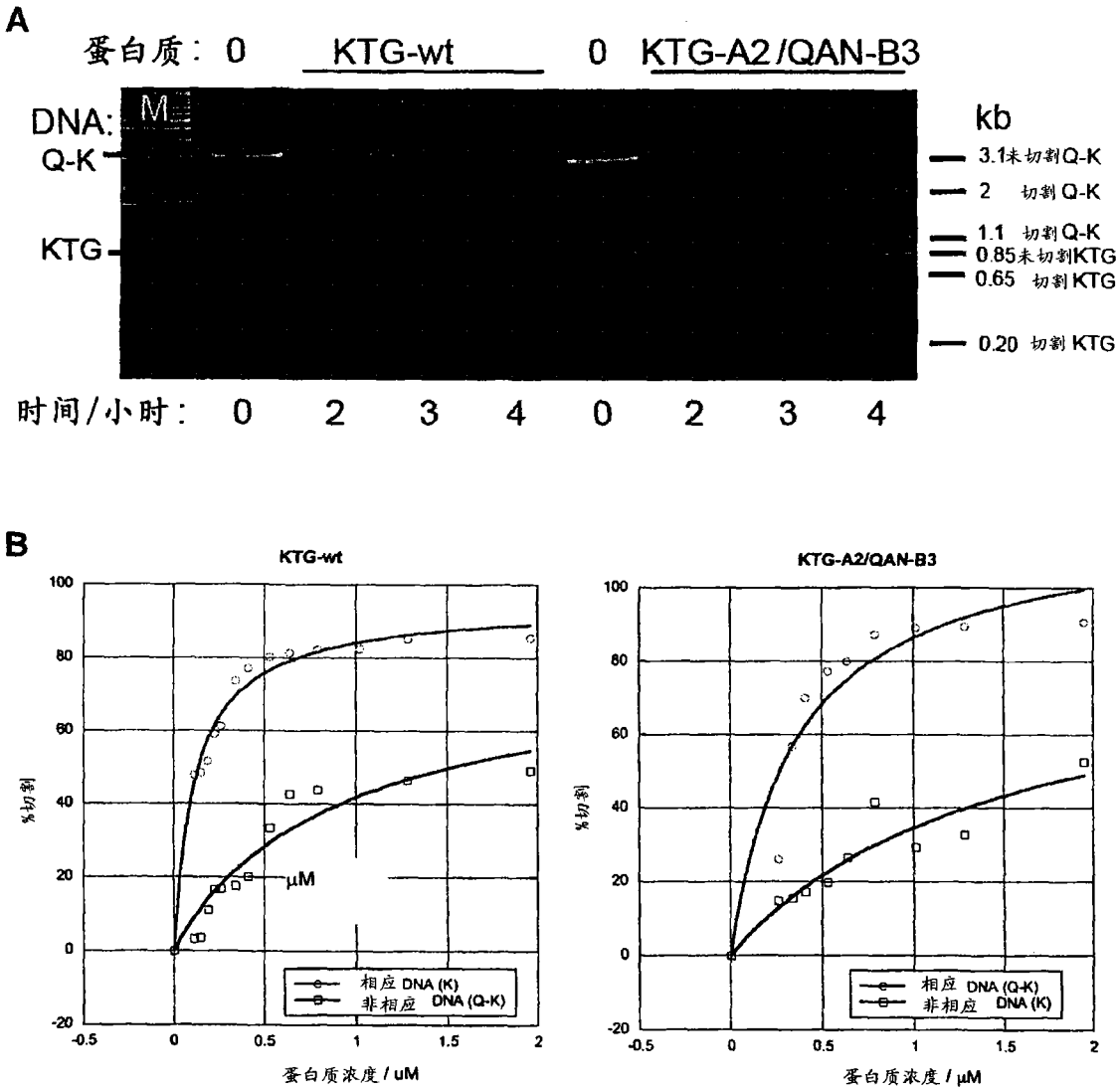


图 8

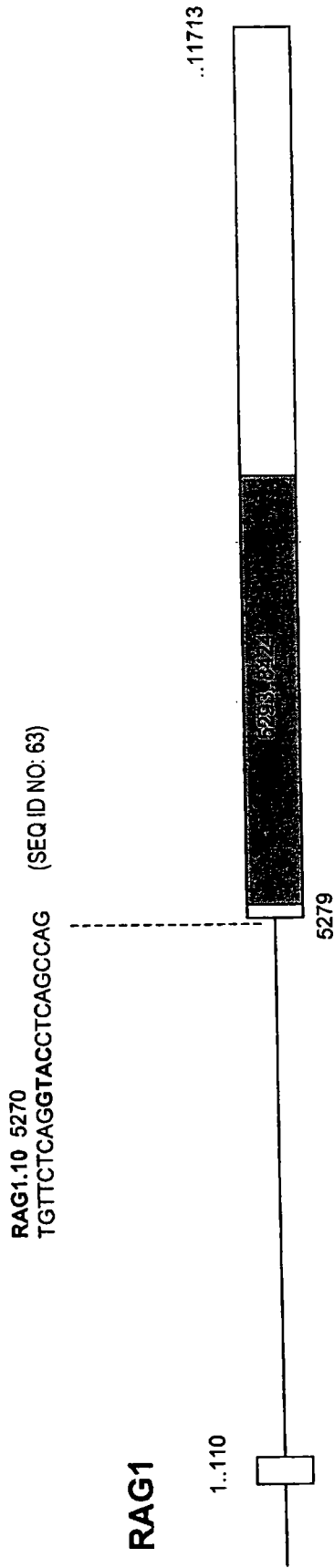
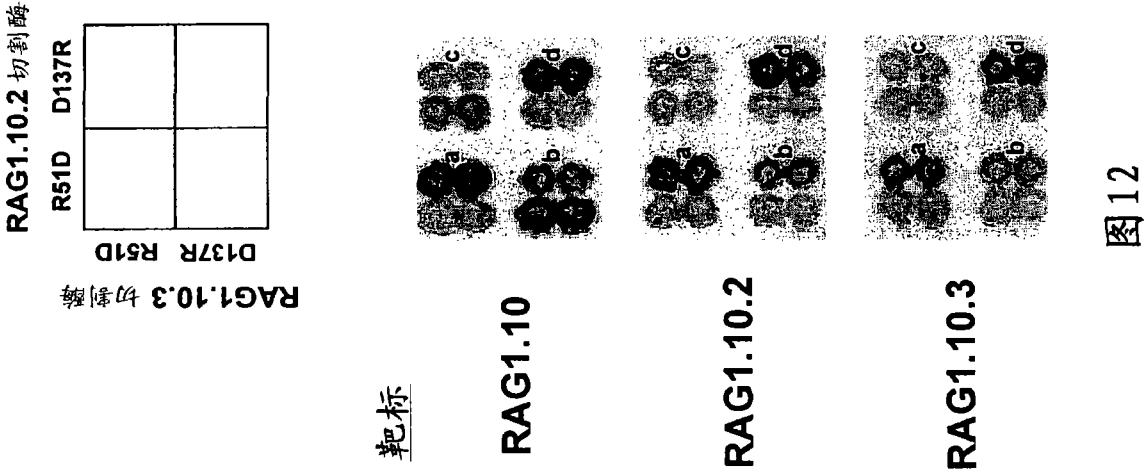


图9



图11



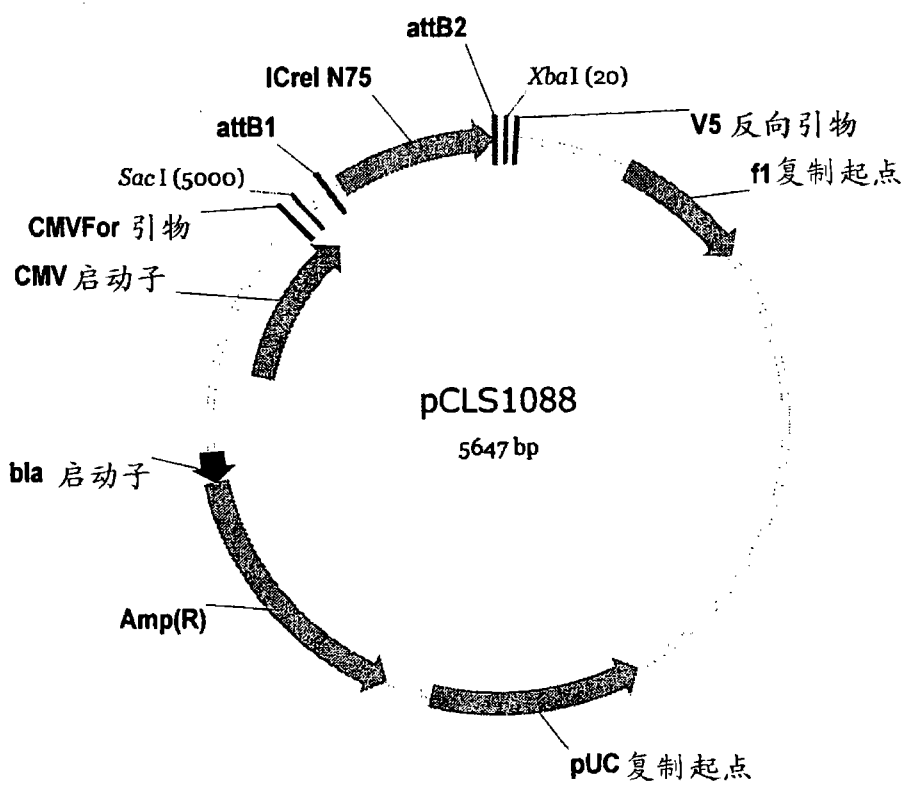


图13

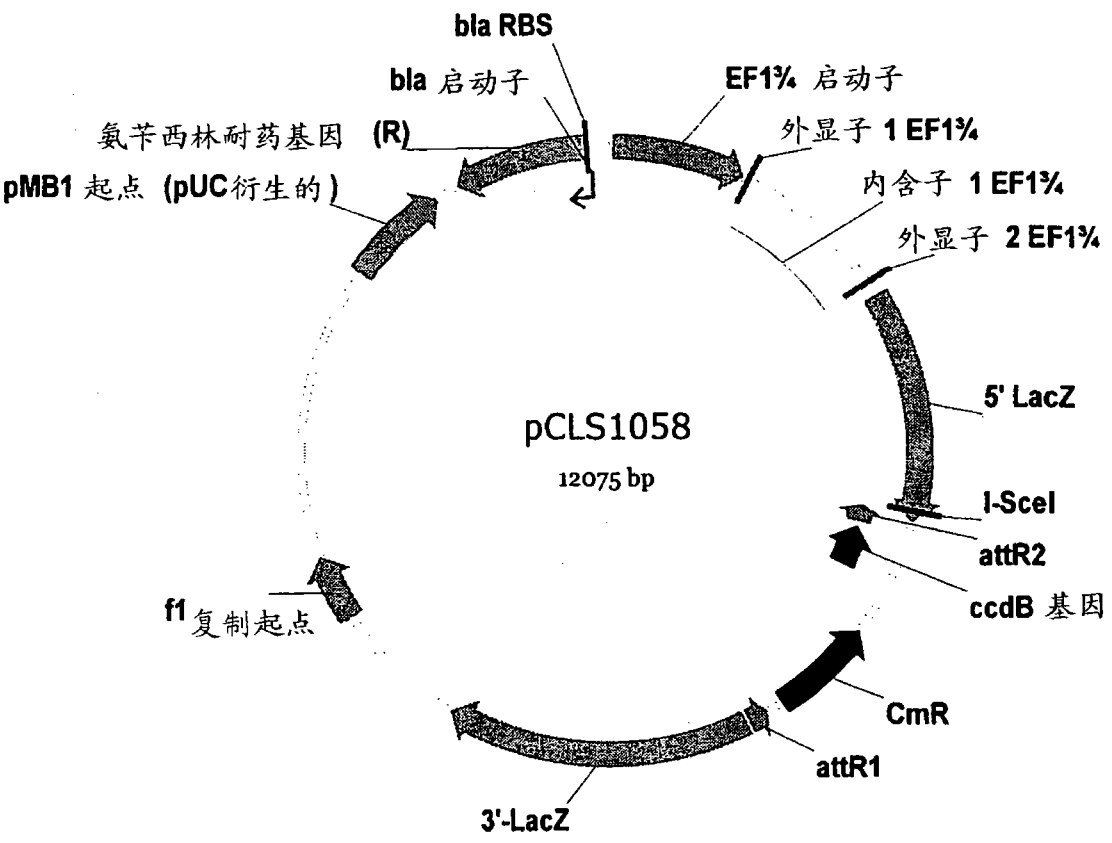


图14

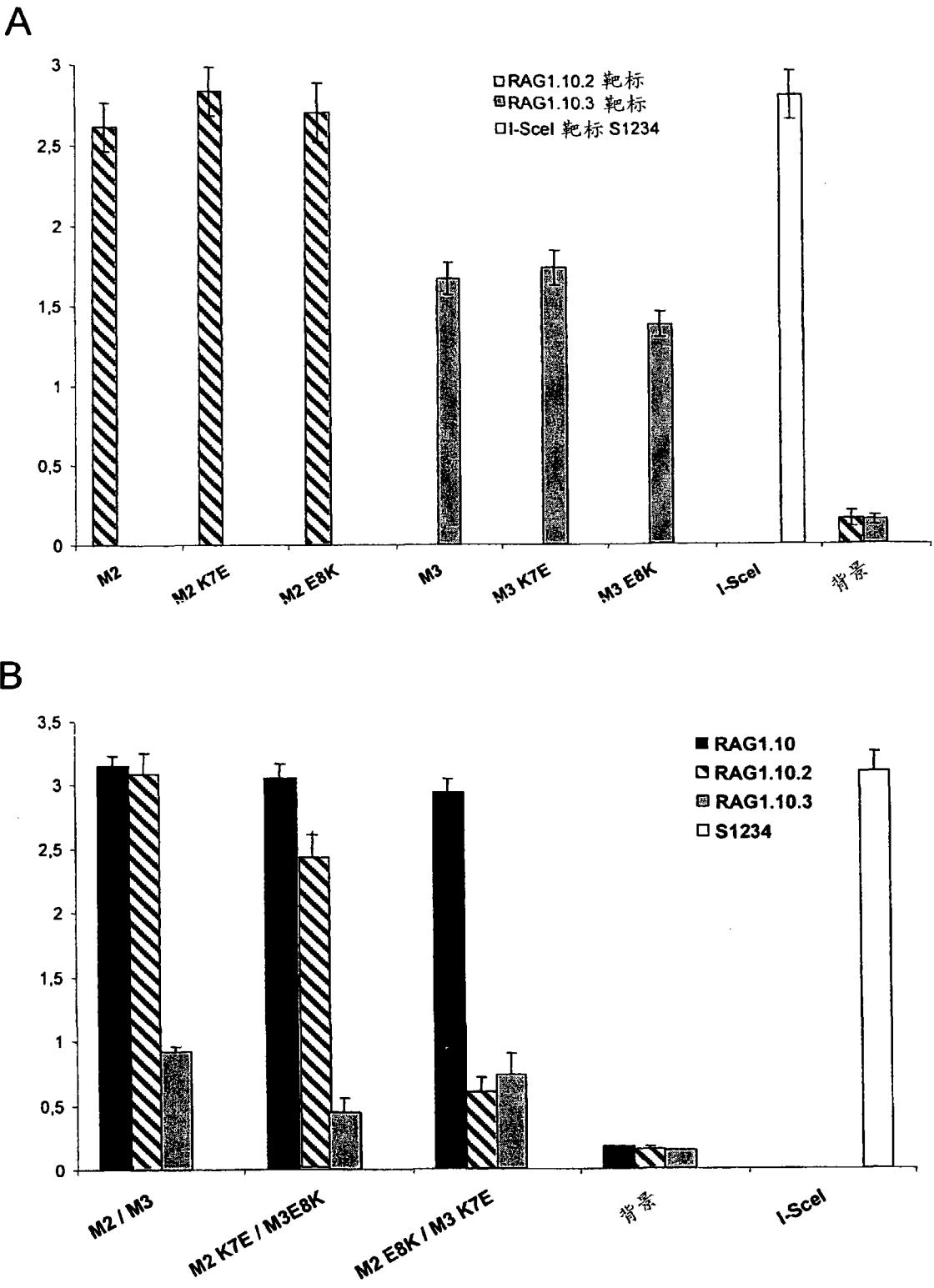


图15

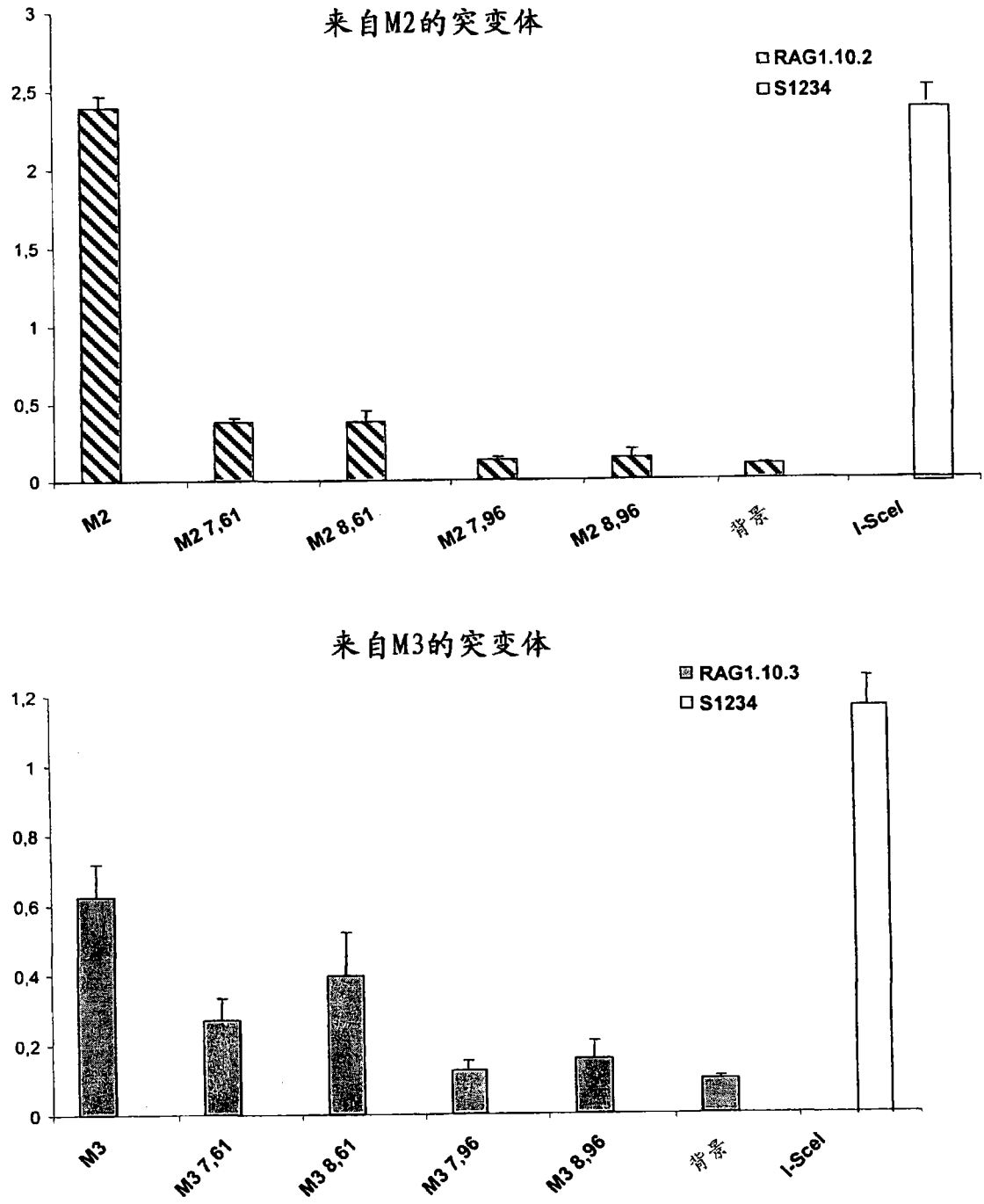


图16

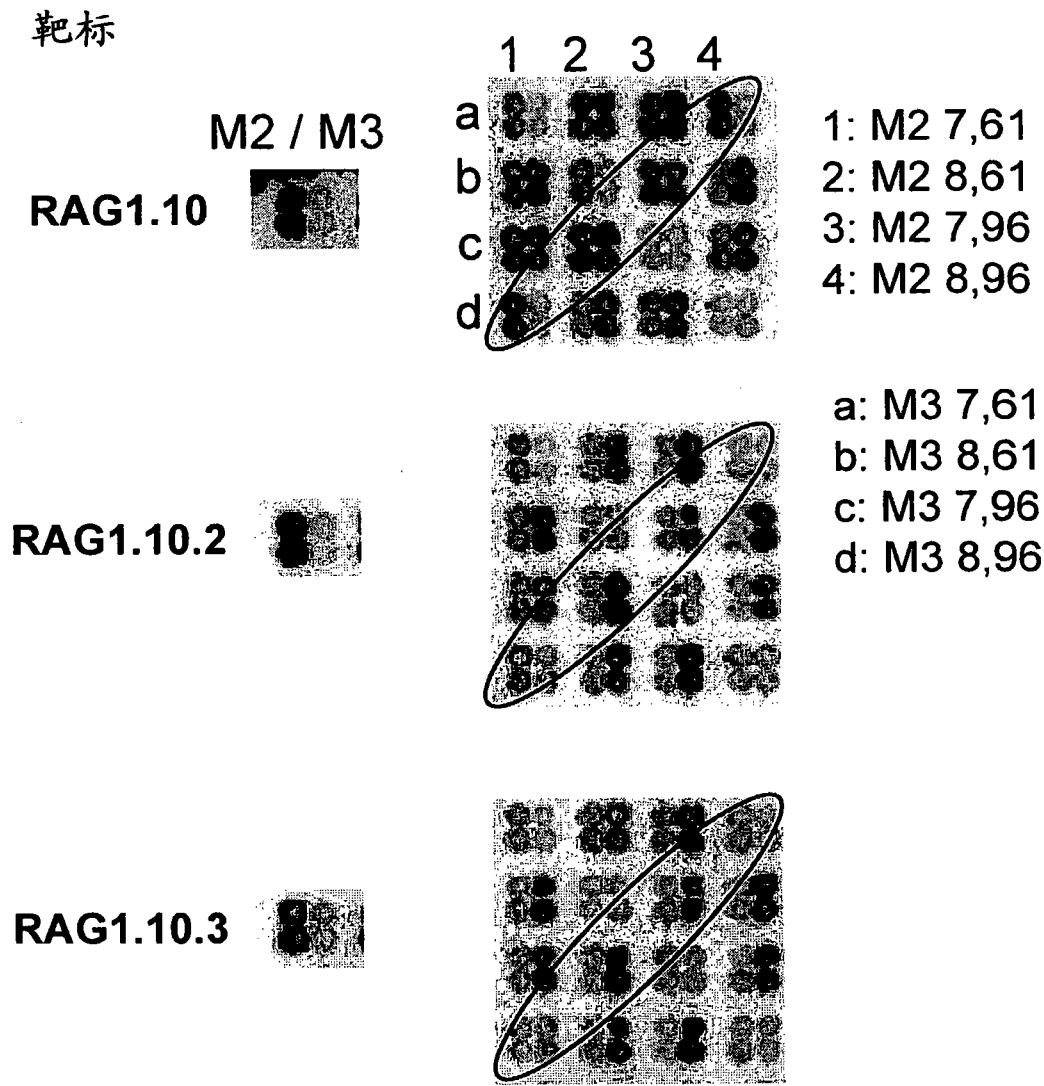


图17

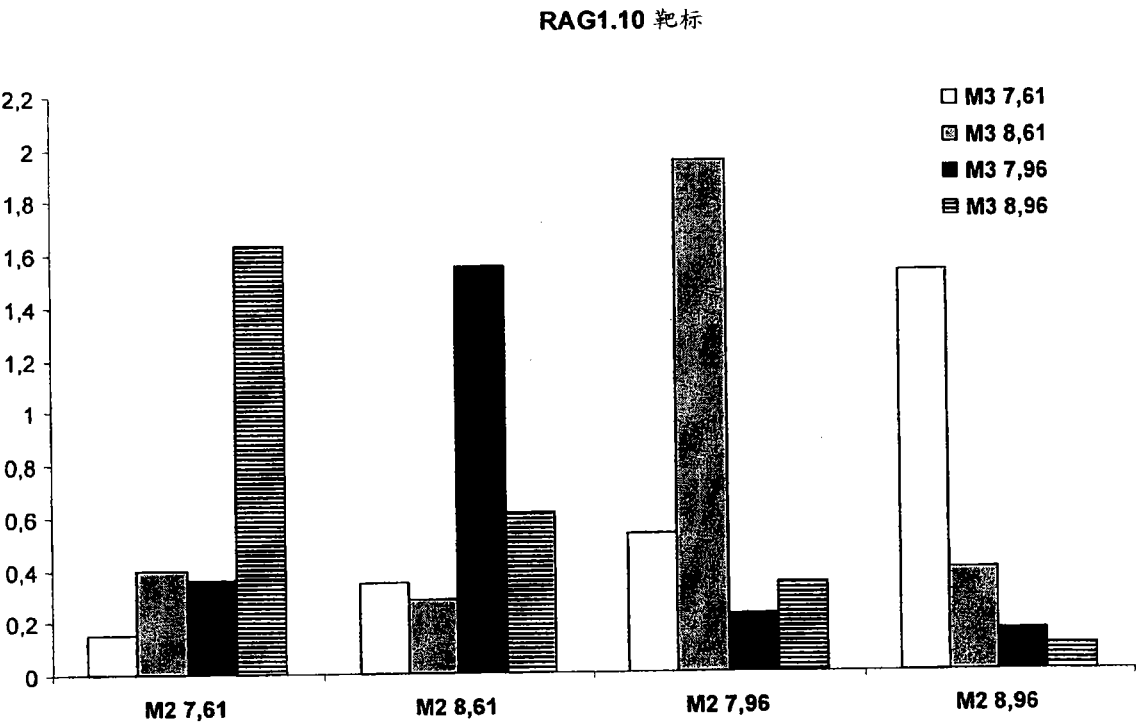


图 18

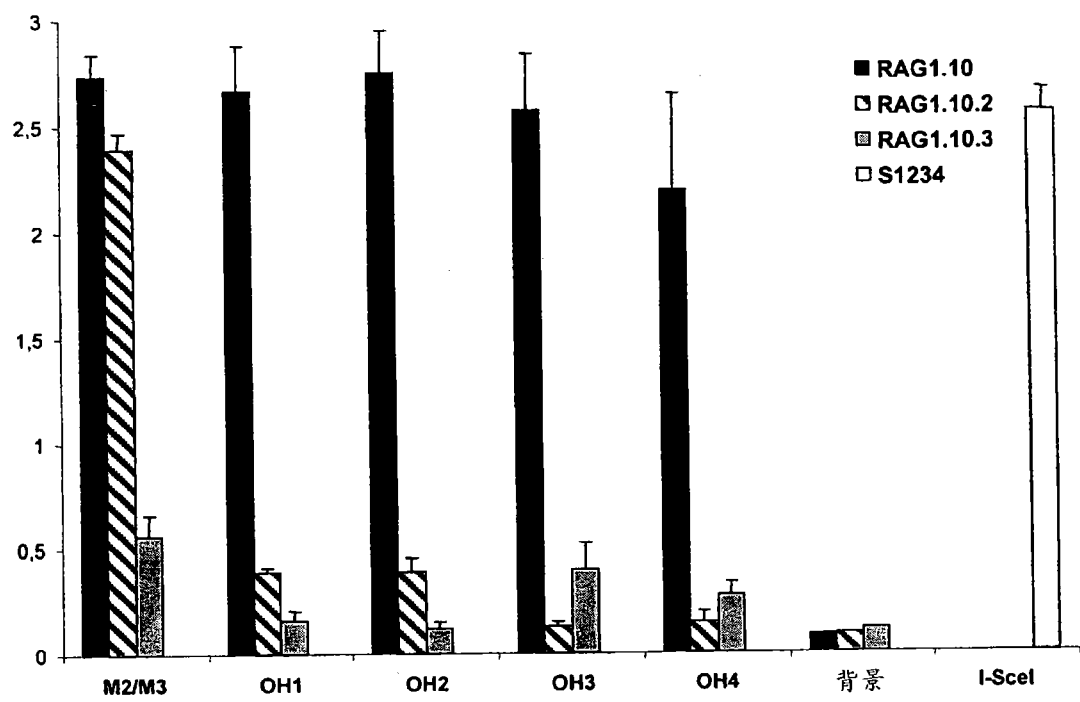


图19

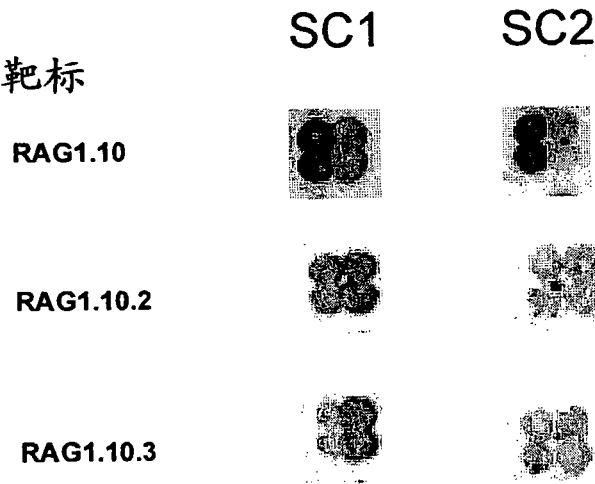


图 20