

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
C12N 5/00 (2006.01)



[12] 发明专利说明书

专利号 ZL 02824412.5

[45] 授权公告日 2009 年 2 月 25 日

[11] 授权公告号 CN 100463962C

[22] 申请日 2002.12.7 [21] 申请号 02824412.5

[30] 优先权

[32] 2001.12.7 [33] US [31] 60/338,441

[32] 2002.4.26 [33] US [31] 60/376,053

[32] 2002.8.2 [33] US [31] 60/400,904

[32] 2002.8.5 [33] US [31] 60/401,089

[86] 国际申请 PCT/KR2002/002309 2002.12.7

[87] 国际公布 WO2003/048345 英 2003.6.12

[85] 进入国家阶段日期 2004.6.7

[73] 专利权人 图尔金株式会社

地址 韩国大田广域市

[72] 发明人 金晋秀 朴卿顺 李东起 薛媛基

李镐琳 李成一 梁效荣 李良顺

张永纯

[56] 参考文献

WO0140798 A2 2001.6.7

WO 0153480 A1 2001.7.26

. Gene Ther, Vol. 7 No. 12. 2000

审查员 杨振宇

[74] 专利代理机构 永新专利商标代理有限公司

代理人 林晓红

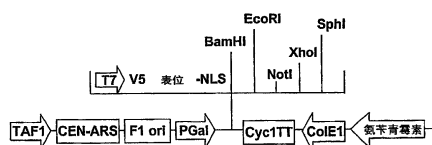
权利要求书 8 页 说明书 202 页 附图 28 页

[54] 发明名称

嵌合蛋白质的表型筛选

[57] 摘要

一方面，本发明涉及筛选编码不同人工嵌合蛋白质的核酸文库以鉴定改变细胞或生物体表型特性的嵌合蛋白质。所述嵌合蛋白质可不通过特定靶基因或通路的先验知识而鉴定。一些嵌合蛋白质包括多个锌指结构域并可以诱导，例如，耐热性，溶剂耐受性，改变的细胞生长，胰岛素产生，分化和药物抗性。



1. 用于鉴定人工转录因子的方法，该方法包括：
提供包含多种核酸的核酸文库，所述多种核酸中的每种核酸编码一种不同的人工转录因子，每种人工转录因子包含一组至少 2 个锌指结构域和激活或抑制转录的调控结构域；
提供具有给定特性的细胞；
将所述核酸文库的成员导入所述细胞中以生成转化细胞；
鉴定使得所述给定特性发生改变的文库成员；和
制备包含编码一种 DNA 结合多肽的序列的编码核酸，所述 DNA 结合多肽包含来自该鉴定的成员的一组锌指结构域，但不包括与该鉴定的成员的调控结构域相同的调控结构域。
2. 权利要求 1 的方法，其中所述 DNA 结合多肽缺少该鉴定的成员的调控结构域。
3. 权利要求 1 的方法，其中所述 DNA 结合多肽包含相对于该鉴定的成员的调控结构域而言是突变了的调控结构域。
4. 权利要求 2 的方法，其中所述 DNA 结合多肽包含与该鉴定的成员的调控结构域相比是相反功能的调控结构域。
5. 权利要求 1 的方法，进一步包括将所述编码核酸导入一种细胞中并评价该细胞的所述给定特性。
6. 权利要求 1 的方法，其中的鉴定步骤包括鉴定具有选自以下一组特性的细胞：对给定环境条件的抗性、分化、去分化、增殖、凋亡、非血清依赖性、病原体抗性和病原体敏感性。
7. 权利要求 1 的方法，其中所述给定特性是所述细胞对毒性试剂的敏感性，且所述鉴定步骤包括在存在毒性试剂的条件下培养所述转化细胞以获得对所述毒性试剂的敏感性发生改变的文库成员。
8. 权利要求 7 的方法，其中所述细胞是真菌细胞，且所述毒性试剂

是抗真菌试剂。

9. 权利要求 7 的方法，其中所述细胞是癌细胞，且所述毒性试剂是抗有丝分裂剂。

10. 权利要求 1 的方法，其中所述细胞是干细胞，所述给定特性是所述干细胞的分化能力，且所述鉴定步骤包括在产生所述人工转录因子因而改变所述干细胞的分化能力的条件下维持该干细胞。

11. 权利要求 10 的方法，其中所述人工转录因子诱导所述干细胞的分化。

12. 权利要求 10 的方法，其中所述人工转录因子增强所述干细胞的自我更新潜力。

13. 权利要求 10 的方法，其中所述干细胞是胚胎干细胞。

14. 权利要求 10 的方法，其中所述干细胞是脊椎动物干细胞或植物干细胞。

15. 权利要求 12 的方法，其中所述干细胞是造血干细胞、神经元祖细胞或肌祖细胞。

16. 包含编码包含 3 个锌指结构域的人工转录因子的序列的核酸，其中所述人工转录因子通过权利要求 1 的方法而鉴定，且所述人工转录因子的表达在至少一种脊椎动物细胞中诱导神经元表型。

17. 权利要求 16 的核酸，其中至少一个锌指结构域具有以下序列：
Cys-X₂₋₅-Cys-X₃-X_a-X-Gln-X_b-X-Ser-Asn-His-X₃₋₅-His (SEQ ID NO: 250)

Cys-X₂₋₅-Cys-X₃-X_a-X-Gln-X_b-X-Ser-Asn-His-X₃₋₅-His (SEQ ID NO: 251); 或

Cys-X₂₋₅-Cys-X₃-X_a-X-Cys-X_b-X-Ser-Asn-His-X₃₋₅-His (SEQ ID NO: 252),

其中 X_a 是苯丙氨酸或酪氨酸；X_b 是一种疏水氨基酸。

18. 权利要求 17 的核酸，其中所述人工转录因子包含以下序列：

Cys-X_{2.5}-Cys-X₃-X_a-X-Gln-X_b-X-Ser-Asn-His-X_{3.5}-His-X_{1.6}-Cys-X_{2.5}-Cys-X₃-X_a-X-Gln-X_b-X-Ser-Asn-His-X_{3.5}-His-X_{1.6}-Cys-X_{2.5}-Cys-X₃-X_a-X-Cys-X_b-X-Ser-Asn-His-X_{3.5}-His (SEQ ID NO: 253), 其中 X_a 是苯丙氨酸或酪氨酸; X_b 是一种疏水氨基酸。

19. 包含编码包含 3 个锌指结构域的人工转录因子的序列的核酸, 其中所述人工转录因子通过权利要求 1 的方法而鉴定, 且其中所述人工转录因子的表达在至少一种脊椎动物细胞中诱导骨发生。

20. 权利要求 19 的核酸, 其中至少一个锌指结构域具有以下序列:

Cys-X_{2.5}-Cys-X₃-X_a-X-Arg-X_b-X-Asp-Lys-His-X_{3.5}-His (SEQ ID NO: 254);

Cys-X_{2.5}-Cys-X₃-X_a-X-Gln-X_b-X-Thr-His-His-X_{3.5}-His (SEQ ID NO: 255);

Cys-X_{2.5}-Cys-X₃-X_a-X-Val-X_b-X-Ser-Thr-His-X_{3.5}-His (SEQ ID NO: 256), 或;

Cys-X_{2.5}-Cys-X₃-X_a-X-Arg-X_b-X-Asp-Lys-His-X_{3.5}-His (SEQ ID NO: 257),

其中 X_a 是苯丙氨酸或酪氨酸; X_b 是一种疏水氨基酸。

21. 权利要求 20 的核酸, 其中所述人工转录因子包含以下氨基酸序列:

Cys-X_{2.5}-Cys-X₃-X_a-X-Arg-X_b-X-Asp-Lys-His-X_{3.5}-His-X_{1.6}-Cys-X_{2.5}-Cys-X₃-X_a-X-Gln-X_b-X-Thr-His-His-X_{3.5}-His-X_{1.6}-Cys-X_{2.5}-Cys-X₃-X_a-X-Val-X_b-X-Ser-Thr-His-X_{3.5}-His-X_{1.6}-Cys-X_{2.5}-Cys-X₃-X_a-X-Arg-X_b-X-Asp-Lys-His-X_{3.5}-His (SEQ ID NO: 258), 其中 X_a 是苯丙氨酸或酪氨酸; X_b 是一种疏水氨基酸。

22. 包含第一组多种核酸和第二组多种核酸的核酸文库, 其中

(a) 第一组多种核酸中的每种核酸编码包含至少 2 个锌指结构域和激活转录的第一个功能性结构域的人工蛋白质嵌合体, 和

(b) 第二组多种核酸中的每种核酸编码包含至少 2 个锌指结构域的人工蛋白质嵌合体，且不包括所述第一个功能性结构域。

23. 权利要求 22 的核酸文库，其中所述第二组多种核酸中的每种核酸编码的人工蛋白质嵌合体包含与所述第一个功能性结构域不同的第二个功能性结构域。

24. 权利要求 23 的核酸文库，其中所述第一个功能性结构域是激活结构域而所述第二个功能性结构域是阻抑结构域。

25. 权利要求 23 的核酸文库，其中所述第一个功能性结构域是激活结构域而所述第二个功能性结构域是与所述第一个功能性结构域具有不同效力的激活结构域。

26. 权利要求 22 的核酸文库，其中所述第一个功能性结构域是组蛋白去乙酰化酶结构域。

27. 用于鉴定蛋白质嵌合体的方法，该方法包括：

提供权利要求 22 的核酸文库；

将所述核酸文库的成员导入具有给定特性的细胞中；和

鉴定使得所述给定特性发生改变的文库成员。

28. 用于鉴定多种蛋白质嵌合体的方法，该方法包括：

提供包含编码第一种人工蛋白质嵌合体的核酸的测试细胞的复制细胞，其中所述第一种人工蛋白质嵌合体的表达使得所述测试细胞的一种给定特性发生改变；

将核酸文库的成员导入每种复制细胞中，该核酸文库包含多种核酸，其中每种核酸编码包含 2 个锌指结构域的不同人工蛋白质嵌合体；和

筛选复制细胞以鉴定所述给定特性被进一步改变的细胞，其中每种复制细胞表达由所述第一种文库成员编码的蛋白质嵌合体 and 由导入的核酸文库成员编码的第二种蛋白质嵌合体。

29. 用于制备修饰的细胞的方法，该方法包括：

提供包含多种核酸的核酸文库，其中每种核酸编码不同的人工蛋白质嵌合体，每种蛋白质嵌合体包含至少 2 个锌指结构域；

鉴定使得测试细胞的一种给定特性发生改变的第一个和第二个文库成员；和

制备可以表达第一种和第二种多肽的细胞，其中所述第一种和第二种多肽分别由所述第一个和第二个鉴定的文库成员编码。

30. 权利要求 29 的方法，其中的鉴定步骤包含所述核酸文库的成员导入所述测试细胞的复制细胞中以提供转化细胞，并鉴定所述给定特性发生改变的的第一种和第二种细胞。

31. 权利要求 29 的方法，该方法进一步包括评价制备的细胞的所述给定特性。

32. 权利要求 29 的方法，其中的制备步骤包含将编码所述第一种多肽的第一个基因和编码所述第二种多肽的第二个基因导入所述细胞。

33. 权利要求 32 的方法，其中所述第一个和第二个基因是同一种核酸的组分。

34. 权利要求 29 的方法，其中的制备步骤包含将包含编码所述第一种多肽的第一个基因的第一种细胞与包含编码所述第二种多肽的第二个基因的第二种细胞融合。

35. 用于鉴定蛋白质嵌合体的方法，该方法包括：

提供包含多种核酸的核酸文库，每种核酸编码包含至少 2 个锌指结构域的不同的人工蛋白质嵌合体；

将所述核酸文库的成员导入具有一种给定特性的细胞；

鉴定使得所述给定特性发生改变的文库成员；和

制备包含多种核酸的第二个文库，其中每种核酸编码

(1) 通过 1 到 6 个氨基酸的取代、插入、或缺失而产生的不同于相应于所述鉴定的成员的人工蛋白质嵌合体的变体，

(2) 包含相应于所述鉴定的成员的人工蛋白质嵌合体的

锌指结构域和一种额外的锌指结构域的蛋白质嵌合体，其中所述额外的锌指结构域在所述第二个文库的成员中是不同的，

(3) 相应于所述鉴定的成员的人工蛋白质嵌合体的变体，其中所述变体具有用其它锌指结构域取代所述锌指结构域的一部分位置以及至少一个与相应于所述鉴定的成员的人工蛋白质嵌合体的相应位置的锌指结构域相同的不变锌指结构域，或

(4) 相应于所述鉴定的成员的人工蛋白质嵌合体的变体，其中，在所述第二个文库的成员中，一或多个锌指结构域位置是不同的，由此相应于所述鉴定的成员的人工蛋白质嵌合体的特定结构域在该位置上出现的频率较其它锌指结构域在该位置上出现的频率更高。

36. 用于鉴定人工转录因子的靶的方法，该方法包括：

提供包含多种核酸的核酸文库，其中每种核酸编码不同的人工转录因子，每个转录因子包含至少 2 个锌指结构域；

将所述文库的成员导入测试细胞的复制细胞中以提供多种转化细胞；

在表达人工转录因子的条件下维持所述的转化细胞；

自所述多种转化细胞中鉴定多种表型发生改变的细胞，其中每种表型发生改变的细胞与所述测试细胞相比具有改变的表型；

在所述多种表型发生改变的细胞的每种细胞中检测转录产物或蛋白质的丰度的分布图以提供每种表型改变的细胞的分布图；和

互相比对分布图以鉴定一或多种转录产物或蛋白质，其中与所述测试细胞相比，所述一或多种转录产物或蛋白质在至少 2 种表型发生改变的细胞中的丰度发生相似的改变。

37. 用于评价细胞的方法，该方法包括：

提供含有一种异源核酸的修饰的细胞，该异源核酸包含编码包含至少 2 个锌指结构域的人工转录因子的序列，其中相对于不表达该人工转录因子的参照细胞而言，该人工转录因子的表达改变所述细胞的表

型特性；

鉴定与所述参照细胞相比，在修饰的细胞中其表达被改变的靶基因；和

评价测试细胞的所述表型特性，其中所述测试细胞含有所述异源核酸并表达所述人工转录因子，且在所述靶基因的活性在所述测试细胞中被改变。

38. 权利要求 37 的方法，其中所述靶基因的活性通过基因突变被改变。

39. 权利要求 37 的方法，其中所述靶基因的活性通过用双链 RNA 或反义寡核苷酸或核酶处理而被改变。

40. 权利要求 37 的方法，其中所述靶基因的活性通过所述靶基因的过表达而改变。

41. 用于评价人工转录因子的方法，该方法包括：

提供含有一种异源核酸的修饰的细胞，该异源核酸包含编码包含至少 2 个锌指结构域的人工转录因子的序列，其中相对于不表达所述人工转录因子的参照细胞而言，所述人工转录因子的表达改变所述细胞的表型特性；和

将文库核酸导入所述修饰的细胞的复制细胞；和

鉴定减弱由所述人工转录因子引起的所述表型特性变化的一种文库核酸。

42. 权利要求 41 的方法，其中所述文库核酸包含 cDNA 文库核酸。

43. 权利要求 41 的方法，其中所述文库核酸包含编码不同人工蛋白质嵌合体的核酸。

44. 权利要求 43 的方法，其中所述不同人工蛋白质嵌合体包括不同的锌指结构域。

45. 用于鉴定蛋白质嵌合体的方法，该方法包括：

提供包含多种核酸的核酸文库，其中每种核酸编码包含至少 2

个锌指结构域的不同的人工蛋白质嵌合体，其中至少一个锌指结构域来自天然产生的蛋白质；

将所述核酸文库的成员导入具有一种给定特性的细胞；和

鉴定一种细胞，其中所述文库的成员使得该细胞中的所述给定特性发生改变。

46. 用于鉴定蛋白质嵌合体的方法，该方法包括：

提供包含多种核酸的核酸文库，其中每种核酸编码包含至少 2 个锌指结构域的不同的人工蛋白质嵌合体，其中所述文库的至少一个成员的至少一个锌指结构域包含表 1 所列的氨基酸序列；

将所述核酸文库的成员导入具有一种给定特性的细胞；和

鉴定一种细胞，其中所述文库的成员使得该细胞中的所述给定特性发生改变。

47. 用于鉴定蛋白质嵌合体的方法，该方法包含：

提供细胞文库，该文库包含多种细胞，其中每种细胞包括编码包含第一个和第二个锌指结构域的多肽的核酸，其中（1）由所述文库的每种核酸编码的多肽的第一个和第二个锌指结构域均与来自天然产生的蛋白质的锌指结构域相同，且所述第一个和第二个锌指结构域或者（i）不存在于同一种天然产生的蛋白质中或者（ii）存在于同一种天然产生的蛋白质中但与在所述多肽中的构型不同，（2）所述第一个锌指结构域在所述文库的核酸中是不同的，和（3）所述第二个锌指结构域在所述文库的核酸中是不同的；

鉴定表型不同于参照细胞的文库细胞；和

从所鉴定的文库细胞回收所述核酸。

嵌合蛋白质的表型筛选

背景技术

大多数基因在转录水平上受结合于基因内、典型地是启动子或增强子区域内的特异 DNA 位点的多肽转录因子调控。这些蛋白质激活或阻抑 RNA 聚合酶在启动子上的转录起始，因而调控靶基因的表达。许多转录因子，无论激活子还是阻抑子，结构上均含有具有例如 DNA 结合，二聚体化，或与转录机制相互作用等特殊功能的独特结构域。转录因子的 DNA 结合部分本身可以由和 DNA 接触的单独结构域组成。许多 DNA 结合结构域，包括锌指结构域，同源结构域，和螺旋-转角-螺旋结构域的三维结构已经被 NMR 和 X-射线晶体学数据确定。效应子结构域例如激活结构域或阻抑结构域在转移到异源转录因子的 DNA 结合结构域时保留了其功能（Brent 和 Ptashne, (1985) Cell 43: 729-36; Dawson 等, (1995) Mol.Cell Biol.15: 6923-31）。

可以产生锌指结构域的嵌合体的人工转录因子。例如，WO 01/60970（Kim 等）描述了确定锌指结构域的特异性和构建识别特异靶位点的人工转录因子的方法。一个关于人工转录因子的申请是去改变特异靶基因的表达。鉴别靶基因中调控区的靶位点，并对人工转录因子进行工程化以识别一个或多个所述的靶位点。当这样的人工转录因子被导入细胞后，他们可以结合相应的靶位点来调节转录。这样控制靶基因表达的策略有时被称为鉴定转录因子的“靶驱动（target-driven）”方法。

发明内容

一方面，本发明描述了一种方法，其包括：（1）提供了一种细胞文库，该文库包含多种细胞，每种细胞均具有表达一种人工嵌合多肽的异

源核酸，该多肽包含第一和第二结合结构域，其中第一和第二结合结构域互相异源，并且所述多种细胞中每个成员的第一和第二结合结构域与其他成员是互异的；和（2）鉴别文库中与参照细胞相比发生特性（trait）改变的细胞。结合结构域可以是，例如，独立折叠组件（module），例如锌指结构域。在许多实施方式中，结合结构域是 DNA 结合结构域。典型的，参照细胞是不含有文库中的核酸或是含有对照核酸的细胞。参照细胞可以是产生细胞文库的亲代细胞或其衍生细胞。

特性可以是任何可探测的表型，例如可被观察到的，选择的，推断的和/或量化的表型。此处使用的嵌合蛋白质包含至少两个互为异源的结合结构域。这两个结合结构域可以来自不同的天然产生的蛋白质。这两个区域也可以来自相同的天然产生的蛋白质，但是与相应的天然产生的蛋白质相比，它们在嵌合蛋白质中位于不同的构型中。

在许多实施方式中，细胞不含有报道基因。换言之，细胞不通过有关由于嵌合多肽的表达使其调控发生改变的靶基因的先验信息而进行筛选。另外，细胞可以含有报道基因作为与特性有关或无关的附加的指示标记。同样，筛选前，可以获知一个或多个靶基因的信息。

在另一实施方式中，特性是产生一个化合物（例如天然或人工化合物）。该化合物可以是抗生素，抗增生药物，止痛剂，蛋白质等等。

在另一实施方式中，特性是对环境条件，例如重金属，盐度，环境毒素，生物毒素，病原体，寄生虫，其它环境极限条件（例如干燥，热，冷）等等的抗性。在一相关的实施方式中，特性是应激抗性（例如，对热，冷，极端 pH，化学药品，如氨，药物，渗透压，和离子辐射）。在另一实施方式中，特性是药物抗性。特性的改变可以是双向的，例如敏感性或增强的抗性。

在另一实施方式中，细胞是植物，动物（例如哺乳动物），真菌，或细菌细胞。对哺乳动物而言，特性可以是细胞增殖，细胞因子、激素或信号分子的产生，细胞信号通路的激活，生理通路（例如葡萄糖动态平

衡，代谢，肥胖）的激活。

DNA 结合结构域可以是，例如，锌指结构域。典型的，文库中核酸的第一个锌指结构域变化多样，文库中核酸的第二个锌指结构域也变化繁多。核酸还可以表达至少一个第三 DNA 结合结构域，例如第三锌指结构域。

每个表达的多肽的锌指结构域可以与不同的天然产生的蛋白质的锌指结构域相同，或者与天然产生的蛋白质不同，例如在 DNA 接触位置的突变体。天然产生的蛋白质可以是任何真核锌指蛋白质：例如，真菌（例如酵母），植物，或动物蛋白质（例如哺乳动物蛋白质，例如人或鼠蛋白质）。每种多肽可以还包含第三，第四，第五，和/或第六个锌指结构域。每个锌指结构域可以是哺乳动物的，例如人的锌指结构域。

任选地，所述多种细胞的核酸编码足够数目的不同的锌指结构域以识别至少 10，20，30，40，或 50 个不同的 3 碱基对的 DNA 位点。在一个实施方式中，核酸编码了足够数目的不同的锌指结构域以识别不超过 30，20，10，或 5 个不同的 3 碱基对的 DNA 位点。

表达自细胞文库的核酸的多肽也可以包括功能性转录调控结构域，例如转录激活、阻抑结构域甲基化结构域，乙酰化结构域，或去乙酰化结构域。而且许多嵌合多肽没有融合到特殊的转录调节结构域中时也是有功能的。编码多肽的核酸可以是可操纵地连接组成性或可诱导的启动子。

该方法可进一步包含从鉴定的细胞中分离核酸。该核酸可被测序。核酸编码的多肽可被分离。该方法还可包含鉴定多肽特异识别的核酸结合位点。结合位点可被鉴定，例如通过对序列数据库，特别是调控序列数据库进行计算机字符串搜索(string search)或分布图搜索(profile search)，或通过体外选择结合多肽的核酸（例如 SELEX）。可分析核酸序列的计算机数据库以确定鉴定的核酸结合位点相似于鉴定的结合位点的出现情况。

该方法还可包含分析鉴定的细胞中一或多个内源基因的表达或一或多个内源表达的多肽的水平/活性,例如使用 mRNA 检测 (mRNA profiling) (例如使用微阵列分析), 2-D 凝胶电泳, 蛋白质配体 (例如抗体) 阵列, 和/或质谱法。并且, 单一或小数目的基因或蛋白质也可被检测。在另一实施方式中, 比较表达受鉴定的嵌合多肽表达改变的基因的调控区来鉴定可以决定因嵌合多肽的表达直接或间接导致的协调调控的候选位点。

该方法还可包括培养细胞以产生可改变的特性。例如, 如果改变的特性是增加产生代谢物, 该方法可包括培养细胞以产生代谢物。该细胞可以来自于文库, 或者是导入了编码嵌合多肽核酸的细胞。嵌合多肽的表达可被调节, 例如通过可诱导的启动子, 以便可细微地区分特性, 或者与另一个条件启动子区分 (例如细胞类型特异启动子)。含有编码嵌合多肽核酸的细胞可被导入生物体 (例如离体处理), 或用于产生转基因生物。

在一实施方式中, 至少一些文库成员编码了具有不同调控结构域的蛋白质。例如, 一些文库成员可以包括激活结构域, 而其它成员包含抑制结构域。例如, 在某种情况下文库中 DNA 结合结构域的特定组合可表示为与激活结构域融合, 在另一种情况下, 与阻抑结构域融合。在另一实施方式中, 一些文库成员包含激活结构域, 然而其它不含有调控结构域。

以下是一些示范表型: 扩增干细胞群 (例如造血干细胞, 神经干细胞, 表皮干细胞, 或脐带血干细胞) 和其它在体外扩增能力有限的细胞; 抑制干细胞的分化; 细胞 (例如分化细胞或干细胞) 的多能性增加; 改变的应激抗性 (例如对环境条件如热, 冷, 极端 pH, 化学试剂如细胞培养中产生的氨, 药物, 盐 (渗透压), 离子辐射等抗性增加或降低); 对离子辐射或毒性试剂 (例如抗肿瘤药物细胞) 敏感性增加, 例如肿瘤细胞中敏感性升高; 支持病毒感染/复制 (例如丙型肝炎病毒的复制); 抗病原体,

例如病毒，细菌，或原生生物的能力；细胞内 RNAi 效率的增加；转化效率的增加；细胞或生物体衰老过程的阻滞或延迟；在无血清，无生长因子的化学合成培养基中生长；包涵体形成的减少或丧失；和细胞蛋白质分泌的增加。

这些方法的典型应用包含：鉴定病原体（例如病原微生物）中必需基因，鉴定微生物的致病所需的（宿主或病原体的）基因，鉴定候选药物靶，发现信号传导通路基因，微生物工程和工业生物技术，增加具有商业价值的代谢物的产量，和调节生长习性（例如提高微生物生长，或者减少肿瘤细胞生长）。

在另一方面，本发明提供了一种培养细胞，其含有(a)编码蛋白质的（内源或外源）基因，和(b)人工转录因子，其中该细胞以比含有该基因但不含有该转录因子的同种细胞更高的水平表达该蛋白质，并且其中转录因子以不是结合可操纵地连接该基因的调控区域的方式影响蛋白质的生产。术语“人工”指非天然产生。术语“基因”指“编码序列”，可以是 cDNA 或基因组（即具有内含子）的，且可以是内源或外源的（瞬时或稳定转染）。

人工转录因子可含有至少包括两个，三个，或四个锌指结构域的嵌合 DNA 结合结构域。人工转录因子还可包括调控结构域（摘要中所列激活和抑制结构域）。至少一或两个单独的锌指结构域可以天然产生（例如哺乳动物，植物，或人）。在一实施方式中，所有锌指结构域都是天然产生的。

人工转录因子可由该细胞的异源基因编码。编码异源转录因子的异源基因可被可诱导的启动子调控。该细胞还可包括至少一个第二人工转录因子。

例如，(i)人工转录因子引起该细胞或缺失第一个基因但在其他方面均与之相同的一种培养细胞产生的由可操纵地连接到一个调控区域（其不是可操纵地连接到第一个基因的调控区域）的第二个基因编码的蛋白

质水平比含有第二个基因但没有该转录因子的同种细胞要高，且(ii)该转录因子通过不同于结合到可操纵地连接到该第二个基因的调控区域的方式影响第二种蛋白质的生产。在另一实施方式中，(i)该细胞进一步包括编码第二种蛋白质的第二个基因，(ii)该细胞产生的第二种蛋白质比含有该第二个基因但不含有该转录因子的同种细胞的水平要高，和(iii) 该转录因子通过不同于结合到可操纵地连接到该第二个基因的调控区域的方式影响该第二种蛋白质的生产。可操纵地连接到第一个基因的调控区域可以和可操纵地连接到第二个基因的调控区域不同。

第一个和第二个转录因子选自于：

a) 含有 SEQ ID NO: 21

(FECKDCGKAFIGKSNLIRHQRTHTGEKPYACPVESCDRRFSDSSNLTRHIRIHTGEKPYACPVESCDRRFSDSSNLTRHIRIH) 的氨基酸序列的多肽；

b) 含有 SEQ ID NO: 22

(SCGICGKSFSDDSSAKRRHCILHTGEKPYVCDVEGCTWKFARSDKLN RHKKRHTGEKPYVCDVEGCTWKFARSDENLRHKKRHTGEKPYECHDCGKSFRQSTHLTRHR RIH) 的氨基酸序列的多肽；

c) 含有 SEQ ID NO: 23

(YECDHCGKSFSQSSHLNVHKRTHHTGEKPYRCEECGKAFRWPSNLTR HKRIHTGEKPYRCEECGKAFRWPSNLTRHKRIHTGEKPFACPECCKRF MRSDNLTQHIKTH) 的氨基酸序列的多肽。

在另一实施方式中，本发明提供了生产蛋白质的方法。该方法包括提供了此处描述的细胞（例如上述）；在可以允许比含有基因但没有转录因子的同种细胞产生的水平要高的水平（例如至少两，十或一百倍）生产蛋白质的条件下培养细胞；检测细胞生产的蛋白质和/或从细胞和/或从细胞周围的培养基中纯化表达的蛋白质。基因可是内源或外源基因。内源基因可是天然产生的基因或是相对于天然产生的基因经遗传改变的基

因（例如插入或修饰调控序列）。内源基因的实例包括编码激素，细胞表面受体，抗体，生长因子，粘附因子，神经递质，和酶的基因。外源基因可被可操纵地连接到病毒启动子，例如 CMV 或腺病毒启动子。细胞可以是哺乳动物细胞。

该方法还可包括导入细胞到对象中。方法还可包括用药物学可接受的载体配制纯化的蛋白质。

该方法可包括一个或多个下述特征：细胞中转录因子的量可有效提高可操纵地连接到 CMV 启动子的基因所编码的萤光素酶标记蛋白质的生产达至少 1.1 倍，或 2 倍；细胞中转录因子的量可有效提高可操纵地连接到 SV40 启动子的基因所编码的分泌的碱性磷酸酶标记蛋白质的生产达至少 2, 5, 7, 10 倍；转录因子直接改变许多内源基因的表达；转录因子改变细胞分裂的速度；转录因子竞争结合 PB08、K_F02 或 K_D10 特异识别的天然产生的 DNA 结合位点，并且对 DNA 位点具有低于 50nM 的解离常数；转录因子特异识别与 PB08、K_F02 或 K_D10 特异识别的 DNA 结合位点部分重叠的 DNA 位点；转录因子包括两个连续锌指结构域，其选自于 QSNR-DSNR；DSNR-DSNR；DSAR-RDKR；RDKR-RDER；RDER-QTHR；QSHV-WSNR；WSNR-WSNR；WSNR-RDNQ，QSNR1-QSNK；QSNK-CSNR1，其中每 4 个字母标识识别锌指结构域的 -1, +2, +3, 和 +6 位 DNA 接触残基的氨基酸；转录因子含有 SEQ ID NO: 21, 22 或 23 的氨基酸序列。

转录因子可含有 SEQ ID NO: 21, 22 或 23 的 1 到 8 个氨基酸取代、插入或删除的氨基酸序列。取代可在不同于 DNA 接触残基的位置，例如金属协同半胱氨酸和 -1 位之间。取代可以是连续取代。转录因子包含图 17, 18 和 19 的核酸序列编码的氨基酸。

在另一实施方式中，转录因子含有下述三个锌指结构域：

a) Cys-X₂₋₅-Cys-X₃-X_a-X-Gln-X-Ser-Asn-X_b-X-Arg-His-X₃₋₅-His (SEQ ID NO: 24)；

b) Cys-X_{2.5}-Cys-X₃-X_a-X-Asp-X-Ser-Asn-X_b-X-Arg-His-X_{3.5}-His (SEQ ID NO: 25); 和

c) Cys-X_{2.5}-Cys-X₃-X_a-X-Asp-X-Ser-Asn-X_b-X-Arg-His-X_{3.5}-His (SEQ ID NO: 26), 例如在不变位置具有一到三个取代。

在另一实施方式中, 转录因子含有至少 3 个下述 4 个锌指结构域(如此顺序):

a) Cys-X_{2.5}-Cys-X₃-X_a-X-Asp-X-Ser-Ala-X_b-X-Arg-His-X_{3.5}-His (SEQ ID NO: 27);

b) Cys-X_{2.5}-Cys-X₃-X_a-X-Arg-X-Asp-Lys-X_b-X-Arg-His-X_{3.5}-His (SEQ ID NO: 28);

c) Cys-X_{2.5}-Cys-X₃-X_a-X-Arg-X-Asp-Glu-X_b-X-Arg-His-X_{3.5}-His (SEQ ID NO: 29); 和

d) Cys-X_{2.5}-Cys-X₃-X_a-X-Gln-X-Thr-His-X_b-X-Arg-His-X_{3.5}-His (SEQ ID NO: 30), 例如, 在不变位置具有一到三个取代。

在另一实施方式中, 转录因子含有至少 3 个下述 4 个锌指结构域(如此顺序):

a) Cys-X_{2.5}-Cys-X₃-X_a-X-Gln-X-Ser-His-X_b-X-Val-His-X_{3.5}-His (SEQ ID NO: 31);

b) Cys-X_{2.5}-Cys-X₃-X_a-X-Trp-X-Ser-Asn-X_b-X-Arg-His-X_{3.5}-His (SEQ ID NO: 32);

c) Cys-X_{2.5}-Cys-X₃-X_a-X-Tip-X-Ser-Asn-X_b-X-Arg-His-X_{3.5}-His (SEQ ID NO: 33); 和

d) Cys-X_{2.5}-Cys-X₃-X_a-X-Arg-X-Asp-Asn-X_b-X-Lys-His-X_{3.5}-His (SEQ ID NO: 34), 例如, 在不变位置具有一到三个取代。上述序列中, X_a是任意氨基酸, 或任选地为苯丙氨酸或酪氨酸; 和 X_b是任意氨基酸, 或任选地为一种疏水氨基酸。

另一方面, 本发明提供了一种生产靶蛋白质的方法, 包括在细胞中

（例如体外或体内）表达编码该靶蛋白质的核酸，其中的细胞包含异源人工转录因子，相对于没有该异源转录因子的细胞，其可提高蛋白质的产量。在一个实施方式中，该异源转录因子通过不同于直接调控编码靶蛋白质的基因的转录的机制引起产量升高。例如，异源转录因子不结合直接调控编码靶基因的核酸的转录调控区域。转录因子可作为蛋白质或通过导入和转录编码其的核酸导入细胞。该方法可含有其它此处描述的特征。

在另一个方面，本发明提供了一种鉴定转录因子的方法。该方法包括：提供含有多种核酸的核酸文库，每种核酸编码一种不同的人工转录因子，每个转录因子含有至少 2 个互为嵌合的锌指结构域；导入文库成员到一种细胞；鉴定第一种靶蛋白质生产提高的细胞，其中第一种靶蛋白质由可操纵地连接到第一个转录调控序列的基因编码；评估文库成员提高生产第二种靶蛋白质的能力，该第二种靶蛋白质由可操纵地连接到不同于该第一个转录调控序列的第二个转录调控序列的基因编码。

该方法可以包含一个或多个下述特征：第一种和第二种靶蛋白质相同；文库成员导入的细胞每种都含有编码可操纵地连接到第二个转录调控序列的第二种靶蛋白质的基因；细胞是真核细胞；第一个和/或第二个转录调控序列包括病毒调控序列；该方法进一步包括制备含有编码提高第一种和第二种靶蛋白质生产的文库成员编码的转录因子的基因的宿主细胞；该方法进一步包括生产第三种靶蛋白质，其不同于宿主细胞的第一种和第二种靶蛋白质；并且评估，包括评估鉴定的细胞。第二种或第三种靶蛋白质可以是，例如，分泌蛋白质，例如，促红细胞生成素，血小板生成素，生长因子，白细胞介素，或化学因子。在另一实施方式中，靶蛋白质是酶，例如催化代谢物生成通路反应的酶。也可包括此处描述的其它特征。

在一相关方面，本发明提供了一种鉴定蛋白质嵌合体的方法。该方法包括：提供含有多种核酸的核酸文库，每种核酸编码含有至少两个锌

指结构域的不同的人工蛋白质嵌合体；提供在特定条件下生产指定水平的第一种靶蛋白质的检测细胞；其中第一种靶蛋白质是可操纵地连接到第一个转录调控序列的基因编码的；将每个成员导入检测细胞的复制体以提供多种转化细胞；从所述多种转化细胞或其子代细胞中鉴定在特定条件下，生产与指定水平不同的水平的第一种靶蛋白质的细胞；并且评估文库成员在鉴定的细胞中提高第二种靶蛋白质生产的能力，该第二种靶蛋白质是由可操纵地连接到不同于第一个转录调控序列的一个第二个转录调控序列的基因编码的。也可包括此处描述的其它特征。

另一方面，本发明提供了一种宿主细胞，其遗传物质包括：编码可提高靶蛋白质生产的人工转录因子的异源基因，其中该靶蛋白质的生产与缺少该异源基因但其他方面均相同的对照宿主细胞相比至少提高了30%。在一个实施方式中，转录因子不直接调控靶蛋白质的转录。在另一实施方式中，转录因子直接调控靶蛋白质的转录。宿主细胞可以包括一段序列或一个报道构建体，如可操纵地连接到启动子（例如病毒启动子）的编码 lacZ，分泌的碱性磷酸酶（SEAP），GFP，萤光素酶等的序列。也可包括此处描述的其它特征。

另一方面，本发明提供了一种宿主细胞，其遗传物质包括编码含有至少 2, 3, 或 4 个锌指结构域，结合天然产生的 DNA 位点，例如具有小于 50nM 的平衡解离常数并且与 PB08, K_F02, 或 K_D10 竞争结合天然产生的 DNA 结合位点的多肽的基因。细胞可以包括此处描述的其它特征。

另一方面，本发明提供了一种宿主细胞，其遗传物质包括编码靶蛋白质的第一个异源基因，和编码可提高靶蛋白质生产的人工转录因子的第二个异源基因，其中该靶蛋白质的生产与缺少该第二个异源基因但其他方面均相同的对照宿主细胞相比至少提高了 30%%，其中转录因子不直接调控靶蛋白质的转录。细胞可包括此处描述的其它特征。

在另一方面，本发明提供了一个分离的多肽，其包括：序列：

X_a-X-Cys-X₂₋₅-Cys-X₃-X_a-X₅-X_b-X-Arg-His-X₃₋₅-His-X₁₋₆-X_a-X-Cys-X₂₋₅-Cys-X₃-X_a-X₅-X_b-X-Arg-His-X₃₋₅-His-X₁₋₆-X_a-X-Cys-X₂₋₅-Cys-X₃-X_a-X₅-X_b-X-Arg-His-X₃₋₅-His (SEQ ID NO: 35), 而且在有效浓度时在入 293 细胞中其具有一或多个下述效果: a)提高编码可操纵地连接到 CMV 启动子的萤光素酶的基因表达至少 2 倍; b)提高编码可操纵地连接到 SV40 启动子的 SEAP 的基因表达至少 2 倍和 c)提高或降低细胞分裂速度至少 50%。

在另一方面, 本发明提供了改变后生动物细胞 (metazoan cell) 分化状态的方法, 该方法包括: 在细胞中表达有效量的人工转录因子以改变细胞的分化状态。在一个实施方式中, 分化状态特征为神经表型 (例如轴突伸展, 突触形成, 或神经标记表达) 或成骨细胞表型 (例如成骨细胞标记表达)。在一个实施方式中, 分化状态被改变以提高细胞的多能性, 例如使其少分化, 例如功能上作为干细胞或前体细胞。在一个实施方式中, 分化状态从一个分化状态改变为另一个 (例如从成肌细胞状态到成骨细胞状态, 从神经元状态到胶质细胞状态, 等等)。在一个实施方式中, 人工转录因子引起轴突伸展。细胞可以是, 例如, 干细胞, 神经元细胞, 神经鞘细胞, 或神经祖细胞。人工转录因子可以是此处描述的转录因子, 例如 Neuro1-p65 或 Neuro1 相关分子 (见下)。人工转录因子可以和 Neuro1-p65 竞争结合天然 DNA 结合位点。人工转录因子可以结合与 Neuro1-p65 相同的位点或重叠 Neuro1-p65 结合的 DNA 位点。

在另一实施方式中, 人工转录因子诱导了成骨细胞特异性标记, 例如在起先不是成骨细胞, 例如成肌细胞中诱导了成骨细胞特异性标记。

在另一实施方式中, 本发明提供了鉴定转录因子的方法。该方法包括: 提供含有多种核酸的核酸文库, 每种核酸编码不同的人工转录因子, 每个转录因子包含至少 2 个锌指结构域; 导入文库成员到细胞中; 和鉴定改变了分化状态的细胞。该方法包括一或多个下述特征: 细胞是干细胞; 细胞是神经元细胞, 神经鞘细胞, 或神经祖细胞; 和分化状态包括神经元长出 (outgrowth) 或轴突形成。也可包括其它特征。

在另一方面，本发明提供了鉴定多种转录因子的方法。该方法包括：提供含有多种核酸的核酸文库，每种核酸编码一个人工转录因子，每个转录因子含有至少 2 个锌指结构域；鉴定文库的第一个成员，其改变了细胞的给定特性；和筛选细胞以鉴定细胞的给定特性被进一步改变的，其中每种筛选的细胞表达文库的第一个成员编码的转录因子和同一核酸文库或另一人工转录因子核酸文库的第二个成员编码的转录因子。该方法可以包括在存在第一个和第二个转录因子时进行更多轮筛选。有时，该方法包括鉴定文库第二个成员，其反向改变第一个成员影响的表型状态。该方法可以包括此处描述的其它特征。

另一方面，本发明提供了制备修饰细胞的方法，该方法包括提供含有多种核酸的核酸文库，每种核酸编码不同的人工转录因子，每个转录因子包括至少 2 个锌指结构域；鉴定第一个和第二个文库成员，其改变了细胞的给定特性；和制备表达第一种和第二种多肽的细胞，第一种和第二种多肽由第一个和第二个鉴定的文库成员分别编码。该方法也可延伸到更多成员，例如第三个成员。该方法也可包括评估制备细胞的给定特性。

该方法也可包括一或多个下述特征：制备步骤包括将编码第一种多肽的第一个基因和编码第二种多肽的第二个基因导入细胞；第一个和第二个基因是同一核酸的成分；其中制备步骤包括将具有编码第一种多肽的第一个基因的细胞与具有编码第二种多肽的第二个基因的细胞融合；给定的特性是生产代谢物；给定的特性是生产靶多肽；生产包括分泌。也可应用此处描述的其它特征。

下述方法与使用病毒筛选生物 ZFP 文库有关。鉴定转录因子的方法包括：提供含有多种核酸的核酸文库，每种核酸编码不同的人工转录因子，每个转录因子包含至少 2 个锌指结构域；组装每个文库成员到可感染哺乳动物的病毒或病毒颗粒以形成多种病毒或病毒颗粒；将所述多种病毒或病毒颗粒导入非人哺乳动物对象中；和鉴定改变了表型的对象。

例如,每个对象具有可检测的异常;多种病毒或病毒颗粒被分成池(pool),每个池被导入对象之一;所述多种病毒或病毒颗粒包括分离的样品,每个样品具有包装其中的核酸文库的一种核酸,并且单个样品被导入每个对象。

一方面,本发明提供了修饰的细胞,其包含编码人工转录因子的异源核酸,该人工转录因子使得修饰的细胞相对于参照细胞而言具有应激抗性,所述参照细胞基本上与修饰的细胞相同但缺少该异源核酸和人工转录因子。例如,人工转录因子包含至少 2 个锌指结构域。一或多个锌指结构域可是天然产生的,例如表 1 的天然产生的结构域。示范的人工转录因子包含具有一或多个此处描述的连续的基序的转录因子,例如此处描述的耐热或耐溶剂的蛋白质。

修饰的细胞可以是原核或真核细胞。应激抗性可包含一或多个下述特性:抗热,抗溶剂,抗重金属,抗渗透压,抗极端 pH,抗化学物,抗冷,和抗基因毒剂,抗放射性。应激抗性使抗性细胞可在非抗性细胞死亡或不能生长的条件下存活或生长。例如,修饰的细胞可表达人工转录因子而在比缺少人工转录因子的基本上相同的培养细胞更广的范围内抵抗刺激。本发明也提供了这样的人工蛋白质,而且蛋白质改变了细胞相对于不含有该核酸的同种细胞的对毒性试剂的敏感性。

另一方面,该方法包括生产细胞产物的方法。该方法包括提供含有编码人工转录因子的异源核酸的修饰的细胞;在产生人工转录因子的条件下维持修饰的细胞;和回收培养细胞产生的产物,其中产物不是人工转录因子。例如,人工转录因子可产生应激抗性,或另一个此处描述的性质,例如,改变的蛋白质生产,改变的代谢物生产等等。例如,人工转录因子包含至少 2 个锌指结构域。一或多个锌指结构域可是天然产生的,例如,表 1 的天然产生的结构域。示范的人工转录因子包含此处描述的具有一或多个连续基序(例如至少 2, 3, 或 4 个连续的基序,或至少 3 个同样的基序,包括非连续的)的转录因子。

示范的产物包括代谢物或蛋白质（例如，内源或异源蛋白质。例如，修饰的细胞包含编码不是人工转录因子的异源蛋白质的核酸，而且产物是异源蛋白质。另一实施方式中，修饰的细胞进一步包含第二个编码异源蛋白质的核酸，异源蛋白质参与代谢物的生产。修饰的细胞可维持在 20°C 和 40°C 之间或高于 37°C。一个实施方式中，修饰的细胞维持在抑制缺少人工转录因子的基本同种的细胞生长的条件下。

一个实施方式中，锌指结构域包含一套相应于表 15 所列的锌指结构域的 DNA 接触残基的 DNA 接触残基。相关实施方式中，人工转录因子包含一组至少 3 个锌指结构域，其中该组的每个锌指结构域的 DNA 接触残基分别对应于表 15 的一行中所列的任何 3 个连续锌指结构域的 DNA 接触残基。一个实施方式中，人工转录因子与含有表 15 的一行中所列的一组锌指结构域相竞争。

另一方面，本发明提供了包含编码靶蛋白质的基因（例如，内源或异源基因）和含有编码人工蛋白质嵌合体序列的异源核酸的细胞，该人工蛋白质嵌合体（1）使该细胞生产的蛋白质量相对于不含有异源核酸细胞的生产量升高，并且（2）不结合直接调控编码靶蛋白质的基因转录的转录调控区。例如，人工转录因子包括至少 2 个锌指结构域，和，例如，结合 DNA。一或多个锌指结构域可以是天然产生的，例如，表 1 的天然产生的结构域。示范的人工转录因子包括具有一或多个此处描述的连续基序的转录因子。一个实施方式中，细胞是真核细胞，而且人工蛋白质嵌合体与 PB08，K_F02，或 K_D10 竞争结合基因组 DNA 位点。

一个实施方式中，人工蛋白质嵌合体改变了（例如，增加或减少）细胞的细胞周期进展的速率。本发明还提供了此种人工转录因子。

细胞可被用于生产蛋白质的方法中，该方法包括：提供细胞，和在人工蛋白质嵌合体提高了该细胞相对于不含有异源核酸细胞生产的靶蛋白质的量的条件下维持细胞。例如，蛋白质是分泌蛋白质，细胞质蛋白质，或核蛋白质。

另一方面，本发明提供了含有编码分泌蛋白质的内源基因和包括编码人工转录因子的序列的异源核酸的细胞，该人工转录因子提高了该细胞相对于不含有异源核酸的细胞生产的分泌蛋白质的量。一个实施方式中，细胞是真核细胞，且分泌蛋白质是胰岛素。一个实施方式中，人工转录因子特异地结合也是 08_D04_p65 特异性结合的内源 DNA 位点。另一实施方式中，人工转录因子特异性结合内源 DNA 位点而且人工转录因子与 08_D04_p65 竞争结合该内源 DNA 位点。例如，人工转录因子包括如下氨基酸序列：

$CX_{(2-5)}CXXXBXRSHJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXRDXHJXTHX_{(3-5)}H$ (SEQ ID NO: 45); 或

$CX_{(2-5)}CXXXBXRDXHJXTHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXVXSJXRHX_{(3-5)}H$ (SEQ ID NO: 46)

其中 B 是苯丙氨酸或酪氨酸；和 J 是一种疏水氨基酸。

细胞可被用于生产分泌蛋白质（例如胰岛素）的方法中。例如，细胞维持在人工转录因子提高细胞相对于不含有异源核酸的细胞生产的胰岛素量的条件下。同样，本发明提供了含有至少 2 个锌指结构域的人工转录因子，其中人工转录因子诱导了哺乳动物细胞内源胰岛素基因的表达，该细胞在没有人工转录因子时是不表达内源胰岛素基因的。

另一方面，本发明提供了人工转录因子，其改变细胞相对于不含有该核酸的细胞对毒性试剂（例如，药物，例如，抗真菌药物，例如，酮康唑）的敏感性。敏感性可被增加或降低。一个实施方式中，细胞是真菌细胞。例如，人工转录因子包括至少 2 个锌指结构域，例如，至少 3 个。一或多个锌指结构域可以是天然产生的，例如，表 1 的天然产生的结构域。示范的人工转录因子包括具有此处描述的一或多个连续基序的转录因子。例如，人工转录因子结合内源 DNA 位点和人工转录因子表 5 所列锌指蛋白质竞争结合内源 DNA 位点。

另一方面，本发明提供了改变真菌细胞药物抗性的方法。该方法包

括改变一种蛋白质的表达或活性，该蛋白质与所述细胞内的 AQY1，YJR147W，YLL052C，YLL053C 或 YPL091W 具有至少 70% 的相同性。表达可被改变，例如使用转录因子。

另一方面，本发明提供了鉴定改变细胞对毒性试剂敏感性的人工嵌合蛋白质的方法，该方法包括：提供包含多种核酸的核酸文库，每种核酸编码一种嵌合蛋白质，该嵌合蛋白质包括一组至少 3 个锌指结构域，其中至少 2 个邻近的锌指结构域在天然产生的蛋白质中是不互相邻近的；导入文库成员到检测细胞的复制细胞以产生转化细胞；在存在毒性试剂下培养转化细胞；和从转化细胞中鉴定相对于测试细胞改变了对毒性试剂敏感性的细胞。例如，测试细胞是真菌细胞，而毒性试剂是抗真菌试剂。另一实施方式中，测试细胞是癌细胞，而毒性试剂是抗有丝分裂试剂。每种核酸编码的嵌合蛋白质可包括转录调控结构域。

该方法还可包括构建编码第二种嵌合蛋白质的核酸，该嵌合蛋白质包含一组所选细胞的文库成员编码的嵌合蛋白质的锌指结构域，但不包含鉴定细胞的文库成员编码的嵌合蛋白质的转录调控结构域。该方法还包括构建编码第二种嵌合蛋白质的核酸，该嵌合蛋白质包括(i)一组所选细胞的文库成员编码的嵌合蛋白质的锌指结构域，和(ii)转录调控结构域，其不同于鉴定细胞中的由文库成员编码的嵌合蛋白质的转录调控结构域。

另一方面，本发明提供了包含含有 3 个锌指结构域的编码人工转录因子的序列的核酸，其中人工转录因子的表达在至少一个脊椎动物细胞中诱导了神经元表型。一个实施方式中，至少一个锌指结构域具有序列：

Cys-X₂₋₅-Cys-X₃-X_a-X-Gln-X_b-X-Ser-Asn-His-X₃₋₅-His (SEQ ID NO: 250)

Cys-X₂₋₅-Cys-X₃-X_a-X-Gln-X_b-X-Ser-Asn-His-X₃₋₅-His (SEQ ID NO: 251); 或

Cys-X₂₋₅-Cys-X₃-X_a-X-Cys-X_b-X-Ser-Asn-His-X₃₋₅-His (SEQ ID NO:

252),

其中 X_a 是苯丙氨酸或酪氨酸; 和 X_b 是疏水氨基酸。例如, 人工转录因子包括序列: Cys- $X_{2.5}$ -Cys- X_3 - X_a -X-Gln- X_b -X-Ser-Asn-His- $X_{3.5}$ -His- $X_{1.6}$ -Cys- $X_{2.5}$ -Cys- X_3 - X_a -X-Gln- X_b -X-Ser-Asn-His- $X_{3.5}$ -His- $X_{1.6}$ -Cys- $X_{2.5}$ -Cys- X_3 - X_a -X-Cys- X_b -X-Ser-Asn-His- $X_{3.5}$ -His (SEQ ID NO: 253), 其中 X_a 是苯丙氨酸或酪氨酸; 和 X_b 是一种疏水氨基酸。

一个方法包括提供含有核酸的脊椎动物细胞(例如, 哺乳动物细胞, 人); 和在产生人工转录因子和诱导轴突形成的条件下维持脊椎动物细胞。一个实施方式中, 脊椎动物细胞在产生人工转录因子前是干细胞。

另一方面, 本发明提供了含有包含 3 个锌指结构域的人工转录因子的序列的核酸, 其中人工转录因子的表达在至少一个脊椎动物细胞中诱导了骨生成。例如, 至少 1 个锌指结构域具有序列:

Cys- $X_{2.5}$ -Cys- X_3 - X_a -X-Arg- X_b -X-Asp-Lys-His- $X_{3.5}$ -His (SEQ ID NO: 254);

Cys- $X_{2.5}$ -Cys- X_3 - X_a -X-Gln- X_b -X-Thr-His-His- $X_{3.5}$ -His (SEQ ID NO: 255);

Cys- $X_{2.5}$ -Cys- X_3 - X_a -X-Val- X_b -X-Ser-Thr-His- $X_{3.5}$ -His (SEQ ID NO: 256)或;

Cys- $X_{2.5}$ -Cys- X_3 - X_a -X-Arg- X_b -X-Asp-Lys-His- $X_{3.5}$ -His (SEQ ID NO: 257),

其中 X_a 是苯丙氨酸或酪氨酸; 和 X_b 是一种疏水氨基酸; 或人工转录因子包含氨基酸序列:

Cys- $X_{2.5}$ -Cys- X_3 - X_a -X-Arg- X_b -X-Asp-Lys-His- $X_{3.5}$ -His- $X_{1.6}$ -Cys- $X_{2.5}$ -Cys- X_3 - X_a -X-Gln- X_b -X-Thr-His-His- $X_{3.5}$ -His- $X_{1.6}$ -Cys- $X_{2.5}$ -Cys- X_3 - X_a -X-Val- X_b -X-Ser-Thr-His- $X_{3.5}$ -His- $X_{1.6}$ -Cys- $X_{2.5}$ -Cys- X_3 - X_a -X-Arg- X_b -X-Asp-Lys-His- $X_{3.5}$ -His (SEQ ID NO: 258), 其中 X_a 是苯丙氨酸或酪氨酸; 和 X_b 是疏水氨基酸。

一个方法包含提供含有该核酸的脊椎动物细胞；和在产生人工转录因子和诱导骨生成的条件下维持脊椎动物细胞。

另一方面，本发明提供了改变干细胞分化能力的方法。该方法包括：提供干细胞和含有编码具有多个锌指结构域的转录因子的序列的核酸，其中人工转录因子改变了干细胞的分化能力；导入核酸到干细胞；和在产生人工转录因子因而改变干细胞分化能力的条件下维持干细胞。例如，人工转录因子诱导干细胞的分化。另一实施方式中，人工转录因子提高了干细胞的自我恢复潜力。干细胞可是胚胎干细胞，脊椎动物干细胞，植物干细胞，造血干细胞，神经元祖细胞或肌肉祖细胞。

另一方面，本发明提供了一个方法，包括：提供了含有多种核酸的核酸文库，每种核酸编码不同的人工转录因子，每个人工转录含有一组至少 2 个锌指结构域和激活或抑制转录的调控结构域；提供了具有给定特性的细胞；导入核酸文库成员到细胞；鉴定改变了特性的文库成员；和制备含有编码 DNA 结合多肽序列的编码核酸，该 DNA 结合多肽包含一组鉴定的锌指结构域，但不包含与鉴定的成员的调控结构域相同的调控结构域。例如，DNA 结合多肽缺少鉴定的成员的调控结构域。另一实施方式中，DNA 结合多肽包含与鉴定的成员的调控结构域相比突变了了的调控结构域。一个实施方式中，DNA 结合多肽包含与鉴定的成员相比具有相反功能的调控结构域。

该方法还可包含导入该编码核酸到细胞并分析细胞的给定特性。鉴定的步骤可包括鉴定具有选自下组特性的细胞：对一定环境条件的抗性；分化；去分化；增殖；凋亡；非血清依赖性；病原抗性；和病原敏感性。

另一方面，本发明提供了编码人工转录因子的分离的核酸，该人工转录因子增加了产生该人工转录因子的细胞内的一种异源的、过表达的蛋白质的溶解性。例如，人工转录因子包括大量锌指结构域。细胞可以是细菌细胞或真核细胞。一个实施方式中，蛋白质是哺乳动物蛋白质，

例如, AKT 蛋白质。一个实施方式中, 所述多个锌指结构域包括结构域: QSTR-DSAR-RDHT-WSNR 或 VSTR-DGNV-QSNR-QSNK。本发明还提供了含有权利要求 99 到 104 任一项的异源核酸的修饰的细胞。本发明还提供了生产异源靶蛋白质的方法。该方法包括提供修饰的细胞, 其中修饰的细胞包含含有编码异源靶蛋白质的第二种核酸; 和在产生人工转录因子和异源靶蛋白质的条件下维持修饰的细胞。修饰的细胞可是培养的细胞或体内细胞, 例如, 在对象内。

另一方面, 本发明提供了将改变的特性从第一种细胞转移到第二种细胞中的方法。该方法包括: 提供包含多种核酸的核酸文库, 每种核酸编码不同的含有至少 2 个锌指结构域的人工蛋白质嵌合体; 导入核酸文库成员到具有给定特性的第一种细胞以提供转化细胞; 从文库成员改变了给定特性的转化细胞中鉴定改变了的细胞; 从鉴定的、改变的细胞中回收核酸文库成员; 导入该核酸文库成员到第二种细胞, 其中第二种细胞通过不同于所给定特性的表型特性与第一种细胞相区别; 和评价含有核酸文库成员和表达核酸文库成员编码的人工蛋白质嵌合体的第二种细胞。例如, 第一种和第二种细胞是真核细胞, 例如, 酵母细胞或哺乳动物细胞。第一种和第二种细胞可通过增殖特性或分化特性区分。例如, 第一种细胞是癌细胞, 和第二种细胞是非癌细胞, 或反过来。评价的第二种细胞可通过给定特性的改变来评价。

另一方面, 本发明提供了含有编码含有 3 个锌指结构域的人工转录因子序列的核酸。例如, 人工转录因子的表达改变了至少一个真核细胞的特性, 其中特性选自于病毒复制, 病毒产生, 和病毒感染; 人工转录因子的表达改变了真核细胞调控和真核细胞共培养的或在真核细胞条件培养基中培养的干细胞的能力; 人工转录因子的表达改变了哺乳动物培养细胞(例如, CHO 细胞)糖基化分泌蛋白质(例如, 抗体)的能力; 和人工转录因子的表达改变了细胞摄入外源核酸的能力。

另一方面, 本发明提供了鉴定转录因子的方法。该方法包括: 提供

含有多种核酸的核酸文库，每种核酸编码不同的人工转录因子，每个转录因子含有至少 2 个锌指结构域；导入核酸文库成员到具有给定特性的细胞中；鉴定改变了给定特性的文库成员；和制备含有多种核酸的第二个文库，每种核酸编码(1)与对应于鉴定的成员的人工蛋白质嵌合体区别为 1 到 6 个氨基酸发生取代，插入，或缺失的变体，(2)蛋白质嵌合体，其含有对应于鉴定的成员的人工蛋白质嵌合体的锌指结构域和另外的锌指结构域，其中另外的锌指结构域在第二个文库成员中是不同的，(3)对应于鉴定的成员的人工蛋白质嵌合体的变体，其中变体具有部分锌指结构域位置被其它锌指结构域取代和至少一个不变锌指结构域，该不变锌指结构域与对应于鉴定的成员的人工蛋白质嵌合体的相应位置的锌指结构域相同，和/或(4)对应于鉴定的成员的人工蛋白质嵌合体的变体，其中，第二个文库成员中，一或多个锌指结构域位置不同，所以在该位置对应于鉴定的成员的人工蛋白质嵌合体的特定结构域出现的频率要高于在该位置的其它锌指结构域。该方法可包括其它此处描述的特征。

另一方面，本发明提供了鉴定转录因子的方法，该方法包括：提供含有多种核酸的核酸文库，每种核酸编码不同的人工转录因子，每个转录因子含有具有至少 2 个锌指结构域的 DNA 结合成分和第一个激活或抑制转录的调控结构域；导入核酸文库成员到具有给定特性的细胞中；鉴定改变了给定特性的文库成员；和制备编码鉴定的成员的 DNA 结合成分和与第一个调控结构域不同的第二个调控结构域的核酸。相关的方法包括鉴定缺少调控结构域的转录因子，并添加一个；或从含有的转录因子中除去调控结构域。一些转录因子可以在没有转录调控结构域时起作用。本发明还提供了包含大量锌指结构域和第一个调控结构域的人工转录因子。人工转录因子当在细胞中产生时会产生第一个特性，但当第一个调控结构域失活并含有与第一个功能相反的第二个调控结构域时会产生第二个特性。

一个实施方式中，第一个调控结构域激活了转录而第二个调控结构

域抑制转录。另一实施方式中，第一个调控结构域抑制转录而第二个调控结构域激活转录。另一实施方式中，第一个调控结构域激活转录到第一种程度而第二个调控结构域激活转录到比第一种程度至少低 50%的程度。该方法还可包括导入制备的核酸到测试细胞并评价该测试细胞，例如，评价测试细胞的给定特性。一个实施方式中，鉴定的成员增加了细胞分裂的速度，而该制备的核酸编码降低细胞分裂速度的转录因子。另一实施方式中，鉴定的成员引起了对化合物的抗性，而该制备的核酸编码了引起对化合物抗性的转录因子。

另一实施方式中，本发明提供了鉴定转录因子的方法，该方法包括：提供含有第一种和第二种多种核酸的核酸文库，其中每个第一种核酸编码含有至少 2 个锌指结构域和激活转录的调控结构域的不同的人工转录因子，和每个第二种核酸编码含有至少 2 个锌指结构域和抑制转录的调控结构域的不同的人工转录因子；导入核酸文库成员到具有给定特性的细胞中；和鉴定改变了给定特性的文库成员。

本发明还提供了包括评价第一种细胞中的嵌合人工锌指蛋白质的表型，在第二种细胞中表达蛋白质，和评价第二种细胞的表型的方法。例如，该方法可包括鉴定改变第一种细胞（例如酵母株活人细胞系如 293 细胞）给定表型的转录因子文库的成员，然后在不同细胞中表达该成员（例如不同的酵母株或 HeLa 细胞）。第一种细胞比其它株或细胞系更易于筛选。

另一方面，本发明提供了改变真菌株对抗真菌试剂敏感性的方法。该方法包括：导入编码包含至少 3 个锌指结构域的人工嵌合蛋白质的核酸到真菌细胞中；和在可以让细胞表达导入的核酸的条件下维持真菌细胞。例如，真菌细胞是酵母细胞，例如，念珠菌，毕赤酵母，*Hansenula*，组织胞浆菌，或隐球菌。一个实施方式中，人工嵌合蛋白质含有选自 K1 到 K11 的蛋白质的锌指结构的氨基酸序列（见实施例 3）。

该方法包括一或多个下述特征：人工嵌合蛋白质含有转录调控结构

域（例如，激活或阻抑结构域）；在细胞中人工嵌合蛋白质的表达改变了水转运子（water transporter）的转录水平；在酿酒酵母细胞中人工嵌合蛋白质的表达改变了 YLL053C 基因或 PDR5 基因的转录水平；人工嵌合蛋白质与选自 K1 到 K11 的多肽竞争结合特异的 DNA 位点。

另一方面，本发明提供了改变真菌株对抗真菌试剂敏感性的方法。该方法包括改变一种蛋白质的活性或表达，该蛋白质含有至少 50 个氨基酸的氨基酸序列，所述氨基酸序列与 YLL053C，AQY1，YJR147W，YLL052C，或 YPL091W 具有至少 30%，50%，60%，70%，80%，90%，或 95% 相同。例如，真菌细胞是酵母细胞，例如，念珠菌，毕赤酵母，*Hansenula*，组织胞浆菌，或隐球菌。一个实施方式中，活性或表达升高以降低敏感性，即增加抗性。另一实施方式中，活性或表达降低以增加敏感性。改变活性或表达可以包括表达人工转录因子，将细胞和包含至少 20 个核苷酸的与 YLL053C，AQY1，YJR147W，YLL052C，或 YPL091W 基因互补的序列的双链 RNA（dsRNA）接触，或将细胞和化学化合物接触。

相关的方法包括筛选与这种蛋白质相互作用的测试化合物（例如，小分子有机化合物），例如，鉴定 YLL053C/AQY1-相关蛋白质的抑制剂或筛选测试化合物改变 AQY1，YJR147W，YLL052C，YLL053C 或 YPL091W 的活性或表达的能力。

另一方面，本发明提供了鉴定改变真菌株对抗真菌试剂敏感性的人工嵌合蛋白质的方法。该方法包括：提供含有多种核酸的核酸文库，每种核酸含有编码人工，嵌合蛋白质的编码序列，该蛋白质含有一组至少 3 个锌指结构域，其中至少 2 个邻近的锌指结构域在天然产生的蛋白质中是不互相邻近的；导入每种核酸到真菌细胞以产生转化的真菌细胞；在存在抗真菌试剂的条件下维持（例如，培养）转化的真菌细胞；和从转化的真菌细胞中选择细胞，其对抗真菌试剂的敏感性相对于对照真菌细胞已经改变了。

该方法包括一或多个下述特征：对照真菌细胞不是转化的或用参照核酸转化；每种核酸编码的嵌合蛋白质含有转录调控结构域；真菌细胞是病原性真菌细胞，例如，念珠菌，组织胞浆菌，或隐球菌。

本方法还可包括，例如，构建编码第二种嵌合蛋白质的核酸，该蛋白质 (i)在选择的细胞中含有文库成员编码的嵌合蛋白质的锌指结构域，和(ii)在选择的细胞中含有不同于文库成员编码的嵌合蛋白质转录调控结构域的转录调控结构域。该方法可包括其它此处描述的特征。

另一方面，本发明提供了鉴定抵消真菌细胞对抗真菌试剂抗性的试剂的方法。该方法包括：将测试化合物与含有 YJR147W, YLL052C, YLL053C 或 YPL091W 的多肽接触；和评价测试试剂和多肽的相互作用，其中相互作用表明测试试剂可被用作抵消抗真菌试剂抗性的试剂。该方法还可包括将真菌细胞和测试化合物和抗真菌试剂接触，例如，和评价细胞存活率或生长。细胞可是抗抗真菌试剂（例如，酮康唑）的细胞。

另一方面，本发明提供了鉴定人工转录因子调控的靶基因的方法。该方法包括：提供含有多种核酸的核酸文库，每种核酸编码不同的人工转录因子，每个转录因子含有至少 2 个锌指结构域；导入每个成员到测试的复制细胞中以提供多种转化细胞；从多种转化细胞中鉴定多种表型改变的细胞，其中每种表型改变的细胞具有相对于测试细胞改变了的表型；和鉴定一或多个转录子或蛋白质，其丰度相对于测试细胞在至少 2 个表型改变的细胞中改变相似。

一个实施方式中，该方法还包括，在测试细胞中，改变转录子或蛋白质的活性，其丰度在至少 2 个表型改变的细胞中改变相似，例如，遗传改变（例如突变或过表达）或其它（例如，RNA 干涉，反义，或抗体结合）。有些情况下，转录子或蛋白质的活性改变。

一个实施方式中，相似改变的转录子或蛋白质通过在每个表型改变的细胞中检测转录子或蛋白质丰度的分布图而提供每个表型改变的细胞的分布图；和相互比较分布图来鉴定。可通过使用，例如，核酸或蛋白

质阵列, SAGE 标签, 差异展示, 或减法杂交获得分布图。

一个实施方式中, 细胞是癌细胞, 例如, 人癌细胞。一或多个鉴定的转录子可是在存在任何人工转录因子时细胞缺少的转录子。

鉴定的转录子编码的多肽可被用作靶多肽以筛选发现和目的物相互作用的化合物。测试化合物被评价以确定是否它们增加或抑制靶多肽的活性。一个实施方式中, 测试化合物是具有小于 10, 5 或 2kD 分子量的小分子。

对于此处描述的所有方法, 可以使用编码嵌合锌指蛋白质的核酸文库。术语“文库”指相似的但不相同的生物分子的物理集合 (physical collection)。集合可以, 例如, 共同在一个容器中, 或在分开的容器中或固体支持物的分开的位置上物理地分开 (成组或单独)。集合中可以存在单个文库成员的重复。一个文库可以包括至少 10^1 , 10^2 , 10^3 , 10^5 , 10^7 , 或 10^9 不同的成员, 或少于 10^{13} , 10^{12} , 10^{10} , 10^9 , 10^7 , 10^5 , 或 10^3 不同的成员。

第一个示范文库包括多种核酸, 每种核酸编码一种多肽, 其含有至少第一个, 第二个, 和第三个锌指结构域。如此处所用, “第一个, 第二个和第三个”表示在多肽中以任意顺序排列的三个分开的结构域: 即每个结构域可以在另一个或另两个的 N 末端或 C 末端。第一个锌指结构域在多种核酸中不同。第二个锌指结构域在多种核酸中不同。文库中至少存在 10 个不同的第一个锌指结构域。一个实施方式中, 至少 0.5%, 1%, 2%, 5%, 10%, 或 25% 的文库成员具有如下一个或两个特点: (1) 在体内, 每个抑制至少一个 p1G 报道质粒的转录至少 2.5 倍; 和 (2) 每个结合至少一个靶位点, 解离常数低于 7, 5, 3, 2, 1, 0.5, 或 0.05nM。第一个和第二个锌指结构域可与天然产生的蛋白质不同或者构型中的位置与它们在天然产生的蛋白质中的相对位置不同。例如, 第一个和第二个锌指结构域在多肽中可以邻近, 但是在天然产生的蛋白质中被一或多个插入的锌指结构域分开。

第二个示范文库包括多种核酸，每种核酸编码一种多肽，其包括至少第一个和第二个锌指结构域。每种多肽的第一个和第二个锌指结构域（1）和不同的天然产生的蛋白质的锌指结构域相同（一般不发生在相同的天然产生的蛋白质中或位于与它们在天然产生的蛋白质的相对位置不同的构型中），（2）与天然产生的蛋白质的结构域具有不多于4个，3个，2个，或1个氨基酸残基的不同，或（3）是天然产生的蛋白质的非邻近的锌指结构域。相同的锌指结构域指从第一个金属配位残基（metal coordinating residue）（代表为半胱氨酸）到最后一个金属配位残基（代表为组氨酸）每个氨基酸都相同的锌指结构域。第一个锌指结构域在多种核酸中不同，而且第二个锌指结构域在多种核酸中也不同。天然产生的蛋白质可以是任何真核锌指蛋白质：例如，真菌（例如，酵母），植物，或动物蛋白质（例如，哺乳动物蛋白质，如人或鼠蛋白质）。每种多肽还可包括第三个，第四个，第五个，和/或第六个锌指结构域。每个锌指结构域可是哺乳动物的锌指结构域，例如，人的锌指结构域。

也可使用其它类型的文库，例如，包括突变的锌指结构域。

一些实施方式中，编码锌指蛋白质的核酸文库或这些蛋白质文库可包括具有不同调控结构域的成员。例如，文库可包括至少10%的成员具有激活结构域，和至少另10%的成员具有阻抑结构域。另一实施方式中，至少10%具有激活结构域或阻抑结构域；另有至少10%不具有调控结构域。另一实施方式中，一些包括激活结构域；其它具有阻抑结构域；另外的根本没有调控结构域。也可使用其它百分比，例如至少20%，25%，30%，40%，50%，60%。

1. Neuro1-p65 和相关分子

另一方面，本发明提供了人工多肽，其包括序列：

CX₍₂₋₅₎CXXXBXQXSNJXRHX₍₃₋₅₎HX₍₁₋₆₎BXCX₍₂₋₅₎CXXXBXQXSNJX
KHX₍₃₋₅₎HX₍₁₋₆₎BXCX₍₂₋₅₎CXXXBXCXSNJXRHX₍₃₋₅₎H (SEQ ID NO: 36),

其中 B 是任意氨基酸，或任选地为苯丙氨酸或酪氨酸；和 J 是任意氨基酸，或任选地为疏水氨基酸。这个序列也可简写为：QSNR-QSNK-CSNR。其它示范人工多肽包括：

$CX_{(2-5)}CXXXBXQXSJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXQXSJXKHX_{(3-5)}H$ (SEQ ID NO : 37), 和

$CX_{(2-5)}CXXXBXQXSJXKHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXCXSJXRHX_{(3-5)}H$ (SEQ ID NO : 38)。

多肽可以，例如，当在小鼠 Neuro2a 细胞中有效浓度时诱导轴突。例如，分离的多肽可以包括与 SEQ ID NO: 2 中锌指序列相同的或者与 SEQ ID NO: 2 中锌指结构域不同之处不多于 8, 6, 4, 3, 或 2 个取代的氨基酸序列。取代可以是保守取代。分离的多肽可具有与 SEQ ID NO: 2 的锌指序列具有至少 80%, 85%, 90%, 95%, 或 97% 相同性的序列。一个实施方式中，多肽可特异性结合靶 DNA 位点。例如，多肽可以和 Neuro1-p65 嵌合 ZFP (SEQ ID NO: 2) 竞争结合靶 DNA 位点，例如，具有 K_d 小于 10nM 的位点。多肽还可包括转录调控结构域，例如，激活或阻抑结构域。多肽可包括 1 个, 2 个, 或 3 个或多个另外的锌指结构域。

还提供了编码上述多肽的核酸。例如，含有 SEQ ID NO: 2 氨基酸的分离多肽可被含有 SEQ ID NO: 1 的核酸序列编码。提供的核酸可包括可操纵连接的调控序列，例如，启动子序列，增强子序列，绝缘子序列 (insulator sequence)，非翻译调控区，polyA 添加位点，等等。一个实施方式中，编码核酸是可操纵地连接到条件启动子例如可诱导的启动子或细胞类型特异的启动子。核酸可被包括在一个载体或整合到染色体中。

另外，本发明提供了宿主细胞（例如，哺乳动物宿主细胞），其包括上述的多肽。宿主细胞可包括上述核酸和表达该核酸。例如，宿主细胞可以是神经元细胞，其延展轴突（例如，至少部分的是由于多肽的结果）。

2. Osteo1-p65 和相关分子

另一实施方式中，本发明提供了人工多肽，其包括序列：

$CX_{(2-5)}CXXXBXRDKJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXQXTHJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXRDKJXRHX_{(3-5)}H$ (SEQ ID NO : 39),

其中, B 是任意氨基酸, 或任选地为苯丙氨酸或酪氨酸; 和 J 是任意氨基酸, 或任选地为一种疏水氨基酸。这个序列也可简写为: RDKR-QTHR-VSTR-RDKR。其它示范人工多肽包括:

$CX_{(2-5)}CXXXBXRDKJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXQXTHJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXVXSTJXRHX_{(3-5)}H$ (SEQ ID NO : 40);

$CX_{(2-5)}CXXXBXQXTHJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXVXSTJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXRDKJXRHX_{(3-5)}H$ (SEQ ID NO : 41);

$CX_{(2-5)}CXXXBXRDKJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXQXTHJXRHX_{(3-5)}H$ (SEQ ID NO : 42); 和

$CX_{(2-5)}CXXXBXVXSTJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXRDKJXRHX_{(3-5)}H$ (SEQ ID NO : 43)。

多肽可以, 例如, 当以有效浓度存在于 C2C12 成肌细胞系细胞中时, 诱导碱性磷酸酶或其它成骨细胞分化的指示物。例如, 分离的多肽可包括与 SEQ ID NO: 4 中的锌指序列相同的或与 SEQ ID NO: 4 中锌指序列不同之处不多于 8, 6, 4, 3, 或 2 个取代的氨基酸序列。取代可是保守取代。分离的多肽可以具有与 SEQ ID NO: 4 的锌指序列具有至少 80%, 85%, 90%, 95%, 或 97%相同性的序列。一个实施方案中, 多肽特异的结合靶 DNA 位点。例如, 多肽可以和 Osteo1-p65 嵌合 ZFP (SEQ ID NO: 4) 竞争结合靶 DNA 位点, 例如, 具有小于 10nM K_d 的位点。多肽还可包括转录调控结构域, 例如, 激活或阻抑结构域。多肽可包括 1, 2, 或 3 或多个另外的锌指结构域。

还提供了编码上述多肽的核酸。例如, 含有 SEQ ID NO:4 的氨基酸的分离的多肽可被含有 SEQ ID NO: 3 序列的核酸序列编码。提供的核酸可包括可操纵连接的调控序列, 例如, 启动子序列, 增强子序列, 绝

缘子序列，非翻译调控区，polyA 添加位点，等等。一个实施方案中，编码核酸可操纵地连接到条件启动子如可诱导的启动子或细胞类型特异的启动子中。核酸可被包含在载体或整合到染色体中。

另外，本发明提供了宿主细胞（例如，哺乳动物细胞），其包括上述多肽。宿主细胞可包括上述的核酸和表达该核酸。例如，宿主细胞可是具有成骨细胞表型的干细胞，例如，至少部分的是由于人工嵌合多肽的结果。

3. 08_D04_p65 和相关分子

另一方面，本发明提供了人工多肽，其包括序列：

$CX_{(2-5)}CXXXBXRSHJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXRDXHJXTHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXVXSSJXRHX_{(3-5)}H$ (SEQ ID NO: 44),

其中 B 是任意氨基酸，或任选地为苯丙氨酸或酪氨酸；和 J 是任意氨基酸，或任选地为一种疏水氨基酸。这个序列也简写为：RSHR-RDHT-VSSR。其它示范人工多肽包括：

$CX_{(2-5)}CXXXBXRSHJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXRDXHJXTHX_{(3-5)}H$ (SEQ ID NO: 45); 和

$CX_{(2-5)}CXXXBXRDXHJXTHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXVXSSJXRHX_{(3-5)}H$ (SEQ ID NO: 46)

多肽可以，例如，诱导胰岛素表达，例如，当其在人 293 细胞中以有效浓度存在时。例如，分离的多肽可以包括与 SEQ ID NO: 6 中锌指序列相同的或与 SEQ ID NO: 6 中锌指结构域不同之处不多于 8, 6, 4, 3, 或 2 个取代的氨基酸序列。取代可以是保守取代。分离的多肽可以具有与 SEQ ID NO: 6 的锌指序列具有至少 80%, 85%, 90%, 95%, 或 97% 相同性的序列。一个实施方案中，多肽可以特异性结合靶 DNA 位点。例如，多肽可以与 08_D04_p65 嵌合 ZFP (SEQ ID NO: 6) 竞争结合靶 DNA 位点，例如，具有 K_d 小于 10nM 的位点。多肽还可包括转录调控结构域，例如，激活或阻抑结构域。多肽可包括 1, 2, 或 3 个另外的锌指结构域。

还提供了编码上述多肽的核酸。例如，包含 SEQ ID NO: 6 的氨基酸的分离的多肽，其可被含有 SEQ ID NO: 5 序列的核酸序列编码。提供的核酸可包括可操纵连接的调控序列，例如，启动子序列，增强子序列，绝缘子序列，非翻译调控区，polyA 添加位点，等等。一个实施方案中，编码核酸是可操纵地连接到条件启动子如可诱导的启动子或细胞类型特异的启动子上。核酸可被包括在载体或整合到染色体中。

另外，本发明提供了宿主细胞（例如，哺乳动物宿主细胞），其包含上述的多肽。宿主细胞可包括上述核酸和表达该核酸。例如，宿主细胞可以是表达胰岛素基因的人细胞，例如，至少部分的由于人工嵌合多肽的后果。本发明还提供了生产胰岛素的方法，其包括体外培养此处所述的细胞，或导入上述多肽或编码多肽的核酸到对象的细胞中而在对象中生产胰岛素。

4. P_B08 和相关分子

另一方面，本发明提供了人工多肽，其包括序列：

$CX_{(2-5)}CXXXBXQXSNJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXDXSNJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXDXSNJXRHX_{(3-5)}H$ (SEQ ID NO: 47),

其中 B 是任意氨基酸，或任选地为苯丙氨酸或酪氨酸；和 J 是任意氨基酸，或任选地为一种疏水氨基酸。这个序列也简写为：QSNR-DSNR-DSNR。其它示范人工多肽包括：

$CX_{(2-5)}CXXXBXQXSNJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXDXSNJXRHX_{(3-5)}H$ (SEQ ID NO: 48); 和

$CX_{(2-5)}CXXXBXDXSNJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXDXSNJXRHX_{(3-5)}H$ (SEQ ID NO: 49)。

多肽可以，例如，提高异源多肽的表达，例如，当其以有效浓度存在于 293 细胞中时提高 SV40-SEAP 报道构建体编码的报道多肽的表达。例如，分离的多肽可以包括与 SEQ ID NO: 8 中的锌指序列相同的或与 SEQ ID NO: 8 中的锌指结构域的不同之处不多于 8, 6, 4, 3, 或 2 个

取代的氨基酸序列。取代可以是保守取代。分离的多肽可具有与 SEQ ID NO: 8 的锌指序列具有至少 80%, 85%, 90%, 95%, 或 97% 相同性的序列。一个实施方案中, 多肽可以特异性结合靶 DNA 位点。例如, 多肽可以与 P_B08 嵌合 ZFP (SEQ ID NO: 8) 竞争结合靶 DNA 位点, 例如, 具有 K_d 小于 10nM 的位点。多肽还可包括转录调控结构域, 例如, 激活或阻抑结构域。多肽可包括 1, 2, 或 3 个另外的锌指结构域。

还提供了编码上述多肽的核酸。例如, 含有 SEQ ID NO: 8 氨基酸的分离的多肽可以被含有 SEQ ID NO: 7 序列的核酸序列编码。提供的核酸可包括可操纵连接的调控序列, 例如, 启动子序列, 增强子序列, 绝缘子序列, 非翻译调控区, polyA 添加位点, 等等。一个实施方案中, 编码核酸是可操纵地连接到条件启动子如可诱导的启动子或细胞类型特异的启动子中。核酸可被包含在载体或整合到染色体中。

另外, 本发明提供了宿主细胞 (例如, 哺乳动物宿主细胞), 其包括上述多肽。宿主细胞包括上述核酸和表达该核酸。例如, 宿主细胞可是具有提高了异源多肽生产的哺乳动物细胞。

5. K_D10 和相关分子

另一方面, 本发明提供了人工多肽, 其包含序列:

$CX_{(2-5)}CXXXBXQXSHJXVHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXWXSNIJ$
 $RHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXWXSNIJRXHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXX$
 $XBXRDNJXQHX_{(3-5)}H$ (SEQ ID NO: 50),

其中 B 是任意氨基酸, 或任选地为苯丙氨酸或酪氨酸; 和 J 是任意氨基酸, 或任选地为一种疏水氨基酸。这个序列也简写为: QSHV-WSNR-WSNR-RDNQ。其它示范人工多肽包括:

$CX_{(2-5)}CXXXBXWXSNIJRXHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXWXSNIJ$
 $XRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXRDNJXQHX_{(3-5)}H$ (SEQ ID NO: 51);
 和

$CX_{(2-5)}CXXXBXQXSHJXVHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXWXSNIJ$

RHX₍₃₋₅₎HX₍₁₋₆₎BXCX₍₂₋₅₎CXXXBXWXSNIJXRHX₍₃₋₅₎H (SEQ ID NO: 52)。

多肽当在哺乳动物细胞中以有效浓度存在时可降低细胞增殖，例如，降低至少 30%，40%，50%，或 60%。细胞增殖可通过在预先确定的孵育时间后的细胞数计算。例如，分离的多肽可包括与 SEQ ID NO: 10 的锌指序列相同的或与 SEQ ID NO: 10 的锌指结构域的不同之处不多于 8, 6, 4, 3, 或 2 个取代的氨基酸序列。取代可是保守取代。分离的多肽可具有与 SEQ ID NO: 10 的锌指序列具有至少 80%，85%，90%，95%，或 97%相同性的序列。一个实施方案中，多肽可特异的结合靶 DNA 位点。例如，多肽可与 K_D10 嵌合 ZFP (SEQ ID NO: 10) 竞争结合靶 DNA 位点，例如，具有 K_d 小于 10nM 的位点。多肽还可包括转录调控结构域，例如，激活或阻抑结构域。多肽可包括 1, 2, 或 3 或多个另外的锌指结构域。

还提供了编码上述多肽的核酸。例如，包含 SEQ ID NO: 10 的氨基酸的分离的多肽可被含有 SEQ ID NO: 9 序列的核酸序列编码。提供的核酸可包括可操纵连接的调控序列，例如，启动子序列，增强子序列，绝缘子序列，非翻译调控区，polyA 添加位点，等等。一个实施方案中，编码核酸是可操纵地连接到条件启动子如可诱导的启动子或细胞类型特异的启动子。核酸可被包含在载体或整合到染色体中。

另外，本发明提供了宿主细胞（例如，真核，例如哺乳动物宿主细胞），其包含此处所述的核酸，例如，编码此处所述多肽的核酸。宿主细胞也可表达该核酸。例如，宿主细胞可以是哺乳动物细胞，其与不含有该多肽的同种细胞相比，具有降低的增殖速度。

6. K_{F02} 和相关分子

另一方面，本发明提供了一种人工多肽，其包括序列：

CX₍₂₋₅₎CXXXBXDXSAJXRHX₍₃₋₅₎HX₍₁₋₆₎BXCX₍₂₋₅₎CXXXBXRDXDKJX
RHX₍₃₋₅₎HX₍₁₋₆₎BXCX₍₂₋₅₎CXXXBXRDXDEJXRHX₍₃₋₅₎HX₍₁₋₆₎BXCX₍₂₋₅₎
CXXXBXQXTHJXRHX₍₃₋₅₎H (SEQ ID NO: 53)，

其中 B 是任意氨基酸，或任选地为苯丙氨酸或酪氨酸；和 J 是任意氨基酸，或任选地为一种疏水氨基酸。这个序列也简写为：DSAR-RDKR-RDER-QTHR。其它示范人工多肽包括：

CX₍₂₋₅₎CXXXBXRDKJXRHX₍₃₋₅₎HX₍₁₋₆₎BXCX₍₂₋₅₎CXXXBXRDEJXRHX₍₃₋₅₎HX₍₁₋₆₎BXCX₍₂₋₅₎CXXXBXQXTHJXRHX₍₃₋₅₎ (SEQ ID NO: 54); 和 CX₍₂₋₅₎CXXXBXDXSAJXRHX₍₃₋₅₎HX₍₁₋₆₎BXCX₍₂₋₅₎CXXXBXRDKJXRHX₍₃₋₅₎HX₍₁₋₆₎BXCX₍₂₋₅₎CXXXBXRDEJXRHX₍₃₋₅₎H (SEQ ID NO: 55)。

多肽可以提高细胞增殖，例如，在哺乳动物细胞中，例如，293 细胞中，提高至少 50%，100%，或 120%。例如，分离的多肽可包括与 SEQ ID NO: 12 的锌指序列相同的或与 SEQ ID NO: 12 的锌指结构域的不同之处不多于 8，6，4，3，或 2 个取代的氨基酸序列。取代可以是保守取代。分离的多肽可具有与 SEQ ID NO: 12 的锌指序列具有至少 80%，85%，90%，95%，或 97%相同性的序列。一个实施方案中，多肽可以特异性结合靶 DNA 位点。例如，多肽和 K_F02 嵌合 ZFP (SEQ ID NO: 12) 竞争结合靶 DNA 位点，例如，具有 K_d 小于 10nM 的位点。多肽还可包括转录调控结构域，例如，激活或阻抑结构域。多肽可包括 1，2，或 3 或多个另外的锌指结构域。

还提供了编码上述多肽的核酸。例如，含有 SEQ ID NO: 12 氨基酸的分离的多肽可被含有 SEQ ID NO: 11 序列的核酸序列编码。提供的核酸可包括可操纵连接的调控序列，例如，启动子序列，增强子序列，绝缘子序列。非翻译调控区，polyA 添加位点，等等。一个实施方案中，编码核酸是可操纵地连接到条件启动子如可诱导的启动子或细胞类型特异的启动子上。核酸可被包含在载体或整合到染色体中。

另外，本发明提供了宿主细胞（例如，哺乳动物宿主细胞），其包括上述核酸。宿主细胞可包括上述核酸和表达该核酸。例如，宿主细胞可是哺乳动物细胞，其与不含有该多肽的同种细胞相比，具有提高的细胞增殖速度。

7. K12_A11 和相关分子

另一方面，本发明提供了一种人工多肽，其包括序列：

$CX_{(2-5)}CXXXBXRDXDHJXTHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXQXSNJX$
 $VHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXQXTHJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXX$
 $BXQXFNJXRHX_{(3-5)}H$ (SEQ ID NO: 56),

其中 B 是任意氨基酸，或任选地为苯丙氨酸或酪氨酸；和 J 是任意氨基酸，或任选地为一种疏水氨基酸。这个序列也简写为：RDHT-QSNV-QTHR-QFNR。其它示范人工多肽包括：

$CX_{(2-5)}CXXXBXRDXDHJXTHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXQXSNJX$
 $VHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXQXTHJXRHX_{(3-5)}H$ (SEQ ID NO: 57);
 和

$CX_{(2-5)}CXXXBXQXSNJXVHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXQXTHJX$
 $RHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXQXFNJXRHX_{(3-5)}H$ (SEQ ID NO: 58)

多肽可提高异源基因表达，例如，在哺乳动物细胞中，可操纵地连接到一个强启动子如病毒启动子的基因。例如，分离的多肽可包括与 SEQ ID NO: 260 的锌指序列相同的或与 SEQ ID NO: 260 的锌指结构域的不同之处不多于 8, 6, 4, 3, 或 2 个取代的氨基酸序列。取代可以是保守取代。分离的多肽可具有与 SEQ ID NO: 260 的锌指序列具有至少 80%, 85%, 90%, 95%, 或 97% 相同性的序列。一个实施方案中，多肽可特异性结合靶 DNA 位点。例如，多肽可与 K12_A11 嵌合 ZFP (SEQ ID NO: 260) 竞争结合靶 DNA 位点，例如，具有 K_d 小于 10nM 的位点。多肽还可包括转录调控结构域，例如，激活或阻抑结构域。多肽可包括 1, 2, 或 3 或多个另外的锌指结构域。因为 K12_A11 嵌合 ZFP 包括一个阻抑结构域，所以不太可能通过直接结合异源基因启动子来提高异源基因表达。

还提供了编码上述多肽的核酸。例如，含有 SEQ ID NO: 260 氨基酸的分离的多肽可被含有 SEQ ID NO: 259 序列的核酸序列编码。提供

的核酸可包括可操纵连接的调控序列，例如，启动子序列，增强子序列，绝缘子序列，非翻译调控区，polyA 添加位点，等等。一个实施方案中，编码核酸是可操纵地连接到条件启动子如可诱导的启动子或细胞类型特异的启动子中。核酸可被包含在载体或整合到染色体中。

另外，本发明提供了宿主细胞（例如，哺乳动物宿主细胞），其包括上述核酸。宿主细胞可包括上述核酸并表达该核酸。

8. K44-16-E12 和相关分子

另一方面，本发明提供了人工多肽，其包括序列：

$CX_{(2-5)}CXXXBXQXSHJXVHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXQXSSJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXQXTHJXRHX_{(3-5)}H$ (SEQ ID NO: 59),

其中 B 是任意氨基酸，或任选地为苯丙氨酸或酪氨酸；和 J 是任意氨基酸，或任选地为一种疏水氨基酸。这个序列也简写为：QSHV-QSSR-QTHR。其它示范多肽包括：

$CX_{(2-5)}CXXXBXQXSSJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXQXTHJXRHX_{(3-5)}H$ (SEQ ID NO: 60); 和

$CX_{(2-5)}CXXBXQXSHJXVHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}XBQXSSJXRHX_{(3-5)}H$ (SEQ ID NO: 61)。

多肽可提高异源基因的表达，例如，在哺乳动物细胞中，基因被可操纵地连接强启动子如病毒启动子。例如，分离的多肽可包括与 SEQ ID NO: 262 的锌指序列相同的或与 SEQ ID NO: 262 的锌指结构域的不同之处不多于 8, 4, 3, 或 2 个取代的氨基酸序列。取代可是保守取代。分离的多肽可具有与 SEQ ID NO: 262 的锌指序列具有至少 80%, 85%, 90%, 95%, 或 97% 相同性的序列。一个实施方案中，多肽可特异性结合靶 DNA 位点。例如，多肽可与 K44-16-E12 嵌合 ZFP (SEQ ID NO: 262) 竞争结合靶 DNA 位点，例如，具有 K_d 小于 10nM 的位点。多肽还可包括转录调控结构域，例如，激活或阻抑结构域。多肽可包括 1, 2, 或 3 或多个另外的锌指结构域。因为 K44-16-E12 嵌合 ZFP 包括阻抑结构域，不太可

能通过直接结合异源基因启动子来提高异源基因的表达。

还提供了编码上述多肽的核酸。例如，含有 SEQ ID NO: 262 氨基酸的分离的多肽可被含有 SEQ ID NO: 261 序列的核酸序列编码。提供的核酸可包括可操纵连接的调控序列，例如，启动子序列，增强子序列，绝缘子序列，非翻译调控区，polyA 添加位点，等等。一个实施方案中，编码核酸是可操纵地连接到条件启动子如可诱导的启动子或细胞类型特异的启动子中。核酸可被包含在载体或整合到染色体中。

另外，本发明提供了宿主细胞（例如，哺乳动物宿主细胞），其包括上述核酸。宿主细胞可包括上述核酸并表达该核酸。例如，宿主细胞可是哺乳动物细胞（例如，293 细胞，其具有提高的异源蛋白质生产（例如，被可操纵连接到 CMV 启动子的基因编码的分泌和细胞内报道蛋白质）的表型）。

9. K13_B08 和相关分子

另一方面，本发明提供了人工多肽，其包括序列：

$CX_{(2-5)}CXXXBXQXSNJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXQXSSJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXQXTHJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXRDKJXRHX_{(3-5)}H$ (SEQ ID NO: 62),

其中 B 是任意氨基酸，或任选地为苯丙氨酸或酪氨酸；和 J 是任意氨基酸，或任选地为一种疏水氨基酸。这个序列也简写为：QSNR-QSSR-QTHR-RDKR。其它示范多肽包括：

$CX_{(2-5)}CXXXBXQXSSJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXQXTHJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXRDKJXRHX_{(3-5)}H$ (SEQ ID NO: 63);
和

$CX_{(2-5)}CXXXBXQXSNJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXQXSSJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXQXTHJXRHX_{(3-5)}H$ (SEQ ID NO: 64)。

多肽可引起可操纵连接到 CMV 启动子的异源基因编码的 SEAP 蛋白质的生产提高。例如，分离的多肽可包括与 SEQ ID NO: 264 的锌指序

列相同的或与 SEQ ID NO: 264 的锌指结构域的不同之处不多于 8, 4, 3, 或 2 个取代的氨基酸序列。取代可是保守取代。分离的多肽可具有与 SEQ ID NO: 264 的锌指序列具有至少 80%, 85%, 90%, 95%, 或 97% 相同性的序列。一个实施方案中, 多肽可特异性结合靶 DNA 位点。例如, 多肽可与 K13_B08 嵌合 ZFP (SEQ ID NO: 264) 竞争结合靶 DNA 位点, 例如, 具有 K_d 小于 10nM 的位点。多肽还可包括转录调控结构域, 例如, 激活或阻抑结构域。多肽可包括 1, 2, 或 3 或多个另外的锌指结构域。

还提供了编码上述多肽的核酸。例如, 含有 SEQ ID NO: 264 氨基酸的分离的多肽可被含有 SEQ ID NO: 263 序列的核酸序列编码。提供的核酸可包括可操纵连接的调控序列, 例如, 启动子序列, 增强子序列, 绝缘子序列, 非翻译调控区, polyA 添加位点, 等等。一个实施方案中, 编码核酸是可操纵地连接到条件启动子如可诱导的启动子或细胞类型特异的启动子中。核酸可被包含在载体或整合到染色体中。

另外, 本发明提供了宿主细胞 (例如, 哺乳动物宿主细胞), 其包括上述核酸。宿主细胞可包括上述核酸并表达该核酸。

10. F104_p65 和相关分子

另一方面, 本发明提供了人工多肽, 其包括序列:

$CX_{(2-5)}CXXXBXRDXHIXTHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXRXSHJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXQXSHJXRHX_{(3-5)}H$ (SEQ ID NO: 65),

其中 B 是任意氨基酸, 或任选地为苯丙氨酸或酪氨酸; 和 J 是任意氨基酸, 或任选地为一种疏水氨基酸。这个序列也简写为: RDHT-RSHR-QSHR。其它示范多肽包括:

$CX_{(2-5)}CXXXBXRXSHJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXQXSHJXRHX_{(3-5)}H$ (SEQ ID NO: 66); 和

$CX_{(2-5)}CXXXBXRDXHJXTHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXRXSHJXRHX_{(3-5)}H$ (SEQ ID NO: 67)。

例如, 分离的多肽可包括与 SEQ ID NO: 18 的锌指序列相同的或与

SEQ ID NO: 18 的锌指结构域的不同之处不多于 8, 4, 3, 或 2 个取代的氨基酸序列。取代可是保守取代。分离的多肽可具有与 SEQ ID NO: 18 的锌指序列具有至少 80%, 85%, 90%, 95%, 或 97% 相同性的序列。一个实施方案中, 多肽可特异性结合靶 DNA 位点。例如, 多肽可与 F104_p65 嵌合 ZFP (SEQ ID NO: 18) 竞争结合靶 DNA 位点, 例如, 具有 K_d 小于 10nM 的位点。多肽还可包括转录调控结构域, 例如, 激活或阻抑结构域。多肽可包括 1, 2, 或 3 或多个另外的锌指结构域。

还提供了编码上述多肽的核酸。例如, 含有 SEQ ID NO: 18 氨基酸的分离的多肽可被含有 SEQ ID NO: 17 序列的核酸序列编码。提供的核酸可包括可操纵连接的调控序列, 例如, 启动子序列, 增强子序列, 绝缘子序列, 非翻译调控区, polyA 添加位点, 等等。一个实施方案中, 编码核酸是可操纵地连接到条件启动子如可诱导的启动子或细胞类型特异的启动子中。核酸可被包含在载体或整合到染色体中。

另外, 本发明提供了宿主细胞 (例如, 哺乳动物宿主细胞), 其包括上述核酸。宿主细胞可包括上述核酸并表达该核酸。

11. F121_p65 和相关分子

另一方面, 本发明提供了人工多肽, 其包括序列:

$CX_{(2-5)}CXXXBXQXSHJXTHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXRSHJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXRDXHJXTHX_{(3-5)}H$ (SEQ ID NO: 68),

其中 B 是任意氨基酸, 或任选地为苯丙氨酸或酪氨酸; 和 J 是任意氨基酸, 或任选地为一种疏水氨基酸。这个序列也简写为: QSHT-RSHR-RDHT。其它示范多肽包括:

$CX_{(2-5)}CXXXBXRSHJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXRDXHJXTHX_{(3-5)}H$ (SEQ ID NO: 69); 和

$CX_{(2-5)}CXXXBXQXSHJXTHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXRSHJXRHX_{(3-5)}H$ (SEQ ID NO: 70)。

例如, 分离的多肽可包括与 SEQ ID NO: 20 的锌指序列相同的或与

SEQ ID NO: 20 的锌指结构域的不同之处不多于 8, 4, 3, 或 2 个取代的氨基酸序列。取代可是保守取代。分离的多肽可具有与 SEQ ID NO: 20 的锌指序列具有至少 80%, 85%, 90%, 95%, 或 97% 相同性的序列。一个实施方案中, 多肽可特异性结合靶 DNA 位点。例如, 多肽可与 F121_p65 嵌合 ZFP (SEQ ID NO: 20) 竞争结合靶 DNA 位点, 例如, 具有 K_d 小于 10nM 的位点。多肽还可包括转录调控结构域, 例如, 激活或阻抑结构域。多肽可包括 1, 2, 或 3 或多个另外的锌指结构域。可应用多肽, 例如, 调控胰岛素样生长因子 2 的表达。

还提供了编码上述多肽的核酸。例如, 含有 SEQ ID NO: 20 氨基酸的分离的多肽可被含有 SEQ ID NO: 19 序列的核酸序列编码。提供的核酸可包括可操纵连接的调控序列, 例如, 启动子序列, 增强子序列, 绝缘子序列, 非翻译调控区, polyA 添加位点, 等等。一个实施方案中, 编码核酸是可操纵地连接到条件启动子如可诱导的启动子或细胞类型特异的启动子中。核酸可被包含在载体或整合到染色体中。

另外, 本发明提供了宿主细胞 (例如, 哺乳动物宿主细胞), 其包括上述核酸。宿主细胞可包括上述核酸并表达该核酸。

12. K44-11-D01 和 K44-11-G12 和相关分子

另一方面, 本发明提供了人工多肽, 其包括序列:

$CX_{(2-5)}CXXXBXQXSHJXVHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXQXSNJXI$
 $HX_{(3-5)}XCX_{(2-5)}CXXXBXQXTHJXRHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXCXSN$
 $JXRHX_{(3-5)}H$ (SEQ ID NO: 265),

$CX_{(2-5)}CXXXBXQXSHJXVHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXVXSTJX$
 $RHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXXXBXRNDNJXQHX_{(3-5)}HX_{(1-6)}BXCX_{(2-5)}CXX$
 $XBQXTHJXRHX_{(3-5)}H$ (SEQ ID NO: 266),

其中 B 是任意氨基酸, 或任选地为苯丙氨酸或酪氨酸; 和 J 是任意氨基酸, 或任选地为一种疏水氨基酸。这些序列也简写为: QSHV-SQNI-QTHR-CSNR 和 QSHV-VSTR-RDNQ-QTHR。

例如，分离的多肽可包括与 SEQ ID NO: 14 或 16 的锌指序列相同的或与 SEQ ID NO: 14 或 16 的锌指结构域的不同之处不多于 8, 4, 3, 或 2 个取代的氨基酸序列。取代可是保守取代。分离的多肽可具有与 SEQ ID NO: 14 或 16 的锌指序列具有至少 80%, 85%, 90%, 95%, 或 97% 相同性的序列。一个实施方案中，多肽可特异性结合靶 DNA 位点。例如，多肽可与 K44-11-D01 或 K44-11-G12 嵌合 ZFP (SEQ ID NO: 14 或 16) 竞争结合靶 DNA 位点，例如，具有 K_d 小于 10nM 的位点。多肽还可包括转录调控结构域，例如，激活或阻抑结构域。多肽可包括 1, 2, 或 3 或多个另外的锌指结构域。可应用多肽，例如，调控蛋白质生产。

还提供了编码上述多肽的核酸。例如，含有 SEQ ID NO: 14 或 16 氨基酸的分离的多肽可被含有 SEQ ID NO: 13 或 15 序列的核酸序列编码。提供的核酸可包括可操纵连接的调控序列，例如，启动子序列，增强子序列，绝缘子序列，非翻译调控区，polyA 添加位点，等等。一个实施方案中，编码核酸是可操纵地连接到条件启动子如可诱导的启动子或细胞类型特异的启动子中。核酸可被包含在载体或整合到染色体中。

另外，本发明提供了宿主细胞（例如，哺乳动物宿主细胞），其包括上述核酸。宿主细胞可包括上述核酸并表达该核酸。

如此处所用，“解离常数”指肽结合一个含有 9 碱基对靶位点的 28 碱基对的双链 DNA 的平衡解离常数。解离常数是通过使用凝胶迁移分析确定，该方法使用室温时结合束缚在 20mM Tris pH7.7, 120mM NaCl, 5mM $MgCl_2$, 20 μ M $ZnSO_4$, 10%甘油, 0.1%NP40, 5mM DTT, 和 0.10mg/mL BSA(牛血清白蛋白)中的纯化的蛋白质。其它细节在实施例 10 中和 Rebar 和 Pabo (1994) Science 263: 671-673 中提供。

如此处所用，术语“筛选”指评价文库成员以发现具有特定特性的一个或多个特定成员的过程。直接筛选中，评价文库中每个成员。例如，评价每种细胞以确定是否伸展轴突。另一种筛选中，称为“选择”，不

是直接评价每个成员。这个评价是将文库成员置于只有具有特定特性的成员能保留下来的环境中。选择可以通过存活（例如，药物抗性）或结合到某种表面（例如，粘附底物）而介导。这种选择包括在术语“筛选”之内。

术语“碱基接触位置”、“DNA 接触位置”、或“核酸接触位置”指锌指结构域的 4 个氨基酸位置，其对应于 ZIF268 的 73 位精氨酸，75 位天门冬氨酸，76 位谷氨酸，和 79 位精氨酸的氨基酸位置。

```

Glu Arg Pro Tyr Ala Cys Pro Val Glu Ser Cys Asp Arg Arg Phe Ser
1           5           10           15
Arg Ser Asp Glu Leu Thr Arg His Ile Arg Ile His Thr Gly Gln Lys
          20           25           30
Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Arg Ser Asp His
          35           40           45
Leu Thr Thr His Ile Arg Thr His Thr Gly Glu Lys Pro Phe Ala Cys
          50           55           60
Asp Ile Cys Gly Arg Lys Phe Ala Arg Ser Asp Glu Arg Lys Arg His
65           70           75           80
Thr Lys Ile His Leu Arg Gln Lys Asp (SEQ ID NO : 71)
          85

```

这些位置也分别称为-1, 2, 3, 和 6 位。为在要研究的序列中鉴定与碱基接触位置相对应的位置，将要研究的序列与感兴趣的锌指结构域对比，使研究序列的半胱氨酸和组氨酸残基与 Zif268 锌指 3 的那些残基对比。欧洲生物信息学院的 ClustalW WWW Service (Thompson 等, (1994) Nucleic Acids Res.22: 4673-4680) 提供了一个对比序列的传统方法。

保守氨基酸取代指具有相似侧链的残基的互换。例如，具有脂肪族侧链的一组氨基酸是甘氨酸，丙氨酸，缬氨酸，亮氨酸，和异亮氨酸；具有脂肪羟基侧链的一组氨基酸是丝氨酸和苏氨酸；具有含氨基侧链的

一组氨基酸是天冬酰胺和谷氨酰胺；具有芳香族侧链的一组氨基酸是苯丙氨酸，酪氨酸，和色氨酸；具有碱性侧链的一组氨基酸是赖氨酸，精氨酸，和组氨酸；具有酸性侧链的一组氨基酸是天门冬氨酸和谷氨酸；和具有含硫侧链的一组氨基酸是半胱氨酸和蛋氨酸。依赖于环境，同组的氨基酸可以互换。一些另外的保守氨基酸取代组是：缬氨酸-亮氨酸-异亮氨酸；苯丙氨酸-酪氨酸；赖氨酸-精氨酸；丙氨酸-缬氨酸；天门冬氨酸-谷氨酸；和天冬酰胺-谷氨酰胺。

术语“异源多肽”指或者是具有非天然产生的序列的多肽（例如，杂合多肽）或者是具有与天然产生的多肽相同但存在于其天然时不存在的环境中的序列的多肽。

术语“杂合”和“嵌合”指非天然产生的多肽，其含有的氨基酸序列衍生自（i）至少两个不同的天然产生的序列，或相同天然产生的序列的非连续区域，其中在杂合体中非连续区域变为连续的；（ii）至少一个人工序列（即，不是天然产生的序列）和至少一个天然产生的序列；或（iii）至少两个人工序列（相同或不同）。人工序列的实施例包括天然产生的序列的突变体和重新设计的序列。“人工序列”不存在于天然产生的序列中。对于此处描述的任何人工序列（例如，蛋白质或核酸），本发明还指具有相同元件的序列，但其不存在于基因组已被测序的下述生物中：人，小鼠，拟南芥，果蝇，大肠杆菌，酿酒酵母，和水稻。具有如此序列的分子可在前述任一生物的细胞中被表达为异源分子。

本发明还包括用此处描述的方法制备的序列（不用称为“人工”），例如，结合编码不同锌指结构域核酸序列的方法或表型筛选的方法。本发明还提供了含有如此序列的细胞。

如此处所用，术语“在严格条件下的杂交”指在 6×氯化钠/柠檬酸钠（SSC），45℃，然后在 0.2×SSC，0.1% SDS，65℃洗两次的杂交条件。

术语“结合倾向性”指多肽选择一种核酸结合位点而不是另一个的区别特性。例如，当多肽对于两个不同的核酸结合位点是限制数量时，在

此处描述的体内或体外分析中，更多量的多肽将结合优选位点，而不是另一个位点。

“参照细胞”指任何标准细胞。一个实施方式中，参照细胞时表达锌指蛋白质细胞的亲代细胞，例如，与锌指蛋白质表达细胞基本相同，但不产生锌指蛋白质的细胞。

在其它优点中，许多方法和化合物涉及鉴定和使用新的和有用的嵌合蛋白质，例如，嵌合转录因子。一些实施方案可能包括一个或多个下述优点：i) 内源基因可上或下调。任意给定的人工嵌合转录因子可通过融合到适当的转录激活结构域而被转变为转录激活子或者通过融合到阻抑结构域而转变为转录阻抑子。而且，即使没有转录调控结构域，嵌合转录因子可是潜在的阻抑子，例如，当它们结合 TATA 盒和起始元件附近的位点时。而且，可能筛选一个具有不同 DNA 结合特异性的激活子和阻抑子的文库。ii) 基因表达可被精细调控。依赖于 DNA 结合亲和力，嵌合转录因子可引起一系列效果，例如，中等到强激活和阻抑。这可导致多样的表型，其通过完全失活或高水平过表达特定靶基因不可能获得。例如，某些情况下，广泛特异性是有利的，比如需要几个基因的协同作用而产生特定表型。另一些情况下，需要有限的特异性。一个灵活控制特异性的方法是添加或去除锌指结构域。理论上，3 指的 ZFP 比 6 指的 ZFP 调控的基因多。iii) ZFP 文库方法是广泛应用的。因为转录调控机制是高度保守的而且因为所有已知的生物都使用 DNA 和转录生存，结合 DNA 的嵌合蛋白质可被用来调控任何需要的细胞。而且，如此处所述，许多方法不要求细胞的预先信息（例如，基因组序列）以鉴定有用的嵌合蛋白质。iv) 人工嵌合蛋白质可被用作工具去研究细胞内通路。例如，可以鉴定选择的克隆中负责表型变化的靶基因，例如，如此处所述。v) ZFP-TF 可以模拟主要调控蛋白质的功能，如主要的调控转录因子。例如，ZFP-TF 可以结合与主要调控相同的位点，或重叠位点。vi) 基因表达水平改变，因而 ZFP-TF 生成的表型程度可通过改变细胞中 ZFP-TF 的表达

水平而被精确的控制。

此处引用的所有专利，专利申请，和参考文献被并入参考。下述专利申请：WO01/60970 (Kim 等)；U. S. Serial No. 60/338, 441, 申请日 December 7, 2001；U. S. Serial No. 60/313,402, 申请日 August 17, 2001；U. S. Serial No. 60/374,355, 申请日 April 22, 2002；U. S. Serial No. 60/376,053, 申请日 April 26, 2002；U. S. Serial No. 60/400,904, 申请日 August 2, 2002；U. S. Serial No. 60/401,089, 申请日 August 5, 2002；和 U. S. Serial No. 10/223,765, 申请日 August 19, 2002，在此并入参考。本发明的一或多个实施方案的细节在随后的图和下述说明书中阐明。此处所述的任何特征可和此处所述其它相合的特征组合。本发明的其它特征，目的，和优点通过本说明书和附图和权利要求书将是显而易见的。

图例说明

图 1 是酵母质粒 pYTC-Lib 的区域图。

图 2 是示范的靶驱动方法的示意图。

图 3A 是制备示范锌指蛋白质文库的方法示意图。

图 3B 是示范的表型驱动的方法的示意图。

图 4 描述了瞬时转染的 ZFP 的表达分布图。左图，单独表达 p65 激活结构域的细胞与对照比较。右图，表达锌指蛋白质 F121_p65 的细胞与对照比较。每个偏离对角线的点表示表达被明显改变的基因。

图 5 描述了稳定转染的 ZFP 的表达分布图。

图 6A 描述了 K5 ZFP 的 DNA 结合结构域突变体和效应子结构域改变对药物抗性的作用。

图 6B 描述了 K5 和 YLL053C 过表达的作用。

图 7 描述了构建多样 3 指文库的一个方法。

图 8 是诱导轴突形成的对照和 ZFP 表达细胞的图。图 8A 没有维甲酸 (RA) 处理的 pcDNA3 (空载体) 转染的 Neuro2A 细胞。图 8B 没有

RA 处理的 Neuro1-p65 (也称作 08_D01-p65) 的表达。图 8C 使用了 10 μ M RA 的 pcDNA3 (空载体) 转染的 Neuro2A 细胞。图 8D 使用了 10 μ M RA 的 Neuro1-p65 的表达。

图 9 描述了胰岛素基因过表达的细胞的比较微阵列数据。嵌合 ZFP 是 08_D04_p65。其序列列于图 16。

图 10A 是列出 ZFP DNA 结合特异性的删节表。

图 10B 是说明锌指结构域和识别位点序列对应性的示意图。

图 11 是鉴定人工 ZFP 靶的方法的流程图。

图 12 是鉴定诱导分化的 ZFP 的方法的示意图。

图 13 描述了存在 F104-p65 ZFP 的细胞的不同时间点的比较微阵列数据。

图 14 列出了诱导轴突的蛋白质 Neuro1-p65 的编码核酸 (SEQ ID NO: 1) 和氨基酸 (SEQ ID NO: 2) 序列。Neuro1-p65 包括锌指结构域 QSNR1-QSNK-CSNR1 和 p65 激活结构域。具有至少 2 或 3 个与 DNA 接触残基有相同形式的结构域: QSNR1-QSNK-CSNR1-p65 的连续锌指结构域的其它人工蛋白质也可以诱导轴突。

图 15 列出了诱导骨发生的蛋白质 Osteo1-p65 的编码核酸 (SEQ ID NO: 3) 和氨基酸 (SEQ ID NO: 4) 序列。Osteo1-p65 包括锌指结构域 RDKR-QTHR1-VSTR-RDKR 和 p65 激活结构域。具有至少 2 或 3 个与 DNA 接触残基有相同形式的结构域: RDKR-QTHR1-VSTR-RDKR-p65 的连续锌指结构域的其它人工蛋白质也可以诱导骨发生。

图 16 列出了提高胰岛素生成 (见图 9) 的蛋白质 08-D04_p65 的编码核酸 (SEQ ID NO: 5) 和氨基酸 (SEQ ID NO: 6) 序列。08-D04_p65 包括锌指结构域 RSHR-RDHT-VSSR 和 p65 激活结构域。具有至少 2 或 3 个与 DNA 接触残基有相同形式的结构域: RSHR-RDHT-VSSR-p65 的连续锌指结构域的其它人工蛋白质也可以提高胰岛素生成。

图 17 列出了提高 SV40-SEAP 的蛋白质 P_B08 的编码核酸 (SEQ ID

NO: 7) 和氨基酸 (SEQ ID NO: 8) 序列。P_B08 包括锌指结构域 QSNR1-DSNR-DSNR 和 p65 激活结构域。具有至少 2 或 3 个与 DNA 接触残基有相同形式的结构域: QSNR1-DSNR-DSNR-p65 的连续锌指结构域的其它人工蛋白质也可以提高 SV40-SEAP。

图 18 列出了降低细胞增殖的蛋白质 K_D10 的编码核酸 (SEQ ID NO: 9) 和氨基酸 (SEQ ID NO: 10) 序列。K_D10 包括锌指结构域 QSHV-WSNR-WSNR-RDNQ 和 kid 阻抑结构域。具有至少 2 或 3 个与 DNA 接触残基有相同形式的结构域: QSHV-WSNR-WSNR-RDNQ 的连续锌指结构域的其它人工蛋白质也可以降低细胞增殖。

图 19 列出了增强细胞增殖的蛋白质 K_F02 的编码核酸 (SEQ ID NO: 11) 和氨基酸 (SEQ ID NO: 12) 序列。K_F02 包括锌指结构域 DSAR2-RDKR-RDER1-QHR1 和 kid 阻抑结构域。具有至少 2 或 3 个与 DNA 接触残基有相同形式的结构域: DSAR2-RDKR-RDER1-QHR1-kid 的连续锌指结构域的其它人工蛋白质也可以增强细胞增殖。

图 20 列出了增强蛋白质表达的蛋白质 K44_11_D01 的编码核酸 (SEQ ID NO: 13) 和氨基酸 (SEQ ID NO: 14) 序列。K44_11_D01 包括锌指结构域 QSHV-QSNI-QTHR1-CSNR1 和 kid 阻抑结构域。具有至少 2 或 3 个与 DNA 接触残基有相同形式的结构域: QSHV-QSNI-QTHR1-CSNR1 的连续锌指结构域的其它人工蛋白质也可以增强蛋白质表达。

图 21 列出了增强蛋白质表达的蛋白质 K44-11-G12 的编码核酸 (SEQ ID NO: 15) 和氨基酸 (SEQ ID NO: 16) 序列。K44-11-G12 包括锌指结构域 QSHV-VSTR-RDNQ-QTHR1 和 kid 阻抑结构域。具有至少 2 或 3 个与 DNA 接触残基有相同形式的结构域: QSHV-VSTR-RDNQ-QTHR1 的连续锌指结构域的其它人工蛋白质也可以增强蛋白质表达。

图 22 列出了 F104_p65 的编码核酸 (SEQ ID NO: 17) 和氨基酸 (SEQ ID NO: 18) 序列。F104_p65 包括锌指结构域 RDHT-RSHR-QSHR2-p65。

图 23 列出了 F121_p65 的编码核酸 (SEQ ID NO: 19) 和氨基酸 (SEQ

ID NO: 20) 序列。F121_p65 包括锌指结构域 QSHT-RSHR-RDHT-p65。具有至少 2 或 3 个与 DNA 接触残基有相同形式的结构域: QSHT-RSHR-RDHT 的连续锌指结构域的其它人工蛋白质也可以提高胰岛素样生长因子 2 的转录。

图 24 列出了 K12_A11 的编码核酸(SEQ ID NO: 259)和氨基酸(SEQ ID NO: 260)序列, 其包括锌指结构域 RDHT-QSNV2-QTHR1-QFNR-kid。

图 25 列出了 K44-16-E12 的编码核酸(SEQ ID NO: 261)和氨基酸(SEQ ID NO: 262)序列, 其包括锌指结构域 QSHV-QSSR1-QTHR1-kid。

图 26 列出了 K13_B08 的编码核酸(SEQ ID NO: 263)和氨基酸(SEQ ID NO: 264)序列, 其包括锌指结构域 QSNR1-QSSR1-QTHR1-RDKR-kid。

图 14 至 26 的 DNA 序列可包括编码 HA 标签(灰色下划线)的序列, 和编码核定位信号(框中)的序列。起始和终止密码子有下划线。有阴影的字母表示编码调控结构域的序列, 例如, p65 或 kid 结构域。小写字母表示连接子序列。对于氨基酸序列, 斜体和有阴影的字母表示锌指结构域, 而有阴影但不是斜体的字母表示调控结构域, 例如, Kid 或 p65 结构域。

图 14 至 24 的 DNA 序列包括编码 HA 标签(灰色下划线)的序列, 和编码核定位信号(框中)的序列。起始和终止密码子有下划线。有阴影的字母表示编码调控结构域的序列, 例如, p65 或 kid 结构域。小写字母表示连接子序列。对于氨基酸序列, 斜体和有阴影的字母表示锌指结构域, 而有阴影但不是斜体的字母表示调控结构域, 例如, Kid 或 p65 结构域。

具体实施方式

本发明的一方面, 筛选编码不同人工嵌合蛋白质的核酸文库以鉴定改变细胞或生物体表型特性的嵌合蛋白质。嵌合蛋白质可在未知特定靶

基因或通路的知识下被鉴定。

一个实施方式中，文库的每种核酸编码含有多个锌指结构域的人工多肽。可制备编码不同嵌合蛋白质的核酸文库，例如，如下部分所描述。文库成员被导入培养基中的细胞或生物体的细胞中。经过表达被编码的蛋白质后，鉴定改变了表型特点的细胞或生物体。这些表型改变的细胞的至少一个的文库核酸被回收，因此鉴定产生表型作用的人工嵌合多肽。

尽管这个方法在锌指结构域的嵌合体文中仅是一般性描述，但其易于适合其它结构性结构域，包括其它 DNA 结合结构域和细胞信号结构域。

文库构建：1、示范结构性结构域

构建核酸文库，其包括编码和表达人工蛋白质的核酸，该人工蛋白质是一个或多个结构性结构域的嵌合体。某些方面，结构性结构域是特异性不同的核酸结合结构域所以文库编码了大量具有不同结合特异性的蛋白质。

已知多种结构性结构域与核酸高亲和性和高特异性结合。回顾可以识别双链 DNA 的结构域基序，见，例如，Pabo 和 Sauer (1992) *Annu. Rev. Biochem.* 61 : 1053-95 ; Patikoglou 和 Burley (1997) *Annu. Rev. Biophys. Biomol. Struct.* 26: 289-325; Nelson (1995) *Curr Opin Genet Dev.* 5: 180-9。一些非限制性的核酸结合结构域的实施例包括：

锌指：锌指是大约 30 个氨基酸的小多肽结构域，其中存在四个氨基酸，半胱氨酸或组氨酸，空间距离大约可以使其调节一个锌离子（综述见，例如，Klug 和 Rhodes, (1987) *Trends Biochem. Sci.* 12: 464-469 (1987); Evans 和 Hollenberg, (1988) *Cell* 52: 1-3; Payre 和 Vincent, (1988) *FEBS Lett.* 234: 245-250; Miller 等, (1985) *EMBO J.* 4: 1609-1614; Berg, (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85: 99-102; Rosenfeld 和 Margalit, (1993) *J. Biomol. Struct. Dyn.* 11: 557-570)。因此，锌指结构域可以按照残基调节锌离子的特性而被划分，例如，Cys₂-His₂类，Cys₂-Cys₂类，Cys₂-CysHis

类，等等。Cys₂-His₂ 类的锌指调节残基一般空间排列是：X_a-X-C-X₂₋₅-C-X₃-X_a-X₅-ψ-X₂-H-X₃₋₅-H (SEQ ID NO: 72)，其中 ψ (psi) 是疏水氨基酸 (Wolfe 等, (1999) Annu. Rev. Biophys. Biomol. Struct. 3: 183-212)，其中“X”代表任意氨基酸，其中 X_a 是苯丙氨酸或酪氨酸，下标表示氨基酸数，连字符连接的数字表示一般间隔氨基酸的数目。典型的，间隔氨基酸折叠以形成与 α-螺旋压紧的反平行 β-折叠，尽管这个反平行 β-折叠可以是短的，非理想的，或非存在的。折叠确定了锌指调节侧链的位置，所以它们在一个适于调节锌离子的四面体构形中。碱基接触残基在锌指的 N 末端和在前的茎环区中。

为方便，锌指结构域的主要 DNA 接触残基编号：-1, 2, 3 和 6，基于以下实例：

-1 1 2 3 4 5 6

X_a-X-C-X₂₋₅-C-X₃-X_a-X-C-X-S-N-X_b-X-R-H-X₃₋₅-H (SEQ ID NO: 73)，

其中 X_a 典型是苯丙氨酸或酪氨酸，和 X_b 典型是疏水氨基酸。如以上实施例所指出，DNA 接触残基是 Cys (C)，Ser (S)，Asn (N) 和 Arg (R)。上述基序可简写为 CSNR。如此处所用，这个简写指含有相应基序的一类序列。该类序列还包括包含特定多肽序列的一种序列，典型的序列列于对应于该基序的表 1 或 2 中。当两个序列具有相同的基序时，可使用数字表示序列。

锌指蛋白质典型地由 3 个或多个锌指结构域的串连排列组成。例如，基序连续列出的锌指结构域未被其它折叠结构域间隔，但是可包括一个连接子，例如，此处所述的结构域间的一个灵活的连接子。

锌指结构域 (ZFP) 是最普通的真核 DNA 结合基序之一，发现于从酵母到高等植物到人的种类中。据估计，单在人基因组中至少存在几千个锌指结构域，大约至少 4500。锌指结构域可从锌指蛋白质中分离出来。非限制性的锌指蛋白质的实例包括 CF2-II, Kruppel, WT1, 碱性核蛋白 (Basonuclin), BCL-6/LAZ-3, 红细胞类 Kruppel 样转录因子, Sp1, Sp2,

Sp3, Sp4, 转录阻抑子 YY1, EGR1/Krox24, EGR2/Krox20, EGR3/Pilot, EGR4/AT133, Evi-1, GLI1, GLI2, GLI3, HIV-EP1/ZNF40, HIV-EP2, KR1, ZfX, ZfY, 和 ZNF7。

下述的计算方法可用来鉴定所有测序的基因组或核酸数据库编码的锌指结构域。任何这样的锌指结构域都可使用。另外, 已经设计了人工锌指结构域, 例如, 使用计算的方法 (例如, Dahiyat 和 Mayo, (1997) Science 278: 82-7)。

还值得注意的是至少一些锌指结构域结合不是 DNA 的配体, 例如, RNA 或蛋白质。因此, 锌指结构域或锌指结构域和其它类型的结构域的嵌合体可用来识别多种靶化合物, 不只是 DNA。

WO 01/60970, U. S. Serial No. 60/374,355, 申请日 April 22,2002, 和 U. S. Serial No. 10/223,765, 申请日 August 19,2002 描述了可用来构建人工锌指蛋白质的示范锌指结构域。见表 1, 见下。

同源结构域。同源结构域是真核结构域, 其组成为接触 DNA 小沟的 N 末端臂, 然后是接触大沟的 3 个 α -螺旋 (综述见, 例如, (1991) Biochemistry 30: 11357-67)。第三个 α -螺旋位于大沟并含有 DNA 接触侧链。同源结构域具有特征性的高保守的位于导致第三个 α -螺旋的转角的基序。该基序包括不变的色氨酸, 其位于结构域的疏水核心。这个基序存在于 Prosite 数据库, 是 PDOC00027([L/I/V/M/F/Y/G]-[A/S/L/V/R]-X(2)-[L/I/V/M/S/T/A/C/N]-X-[L/I/V/M]-X(4)-[L/I/V]-[R/K/N/Q/E/S/T/A/I/Y]-[L/I/V/F/S/T/N/K/H]-W-[F/Y/V/C]-X-[N/D/Q/T/A/H]-X(5)-[R/K/N/A/I/M/W]。同源结构域在决定细胞特性和在生物体发育中提供定位信息的转录因子中经常发现。这个经典的同源结构域发现于基因组中的簇中所以同源结构域在簇中的顺序大约对应于它们沿身体轴线的表达模式。同源结构域可通过和同源结构域对比而鉴定, 例如, Hox-1, 或通过和同源结构域分布图或同源结构域的隐藏的马尔可夫模型 (HMM; 见下) 对比, 例如, Pfam 数据库的 PF00046 或 SMART 数据库的“HOX”,

或通过上述的 Prosite 基序 PDOC00027。

螺旋-转角-螺旋蛋白质。这个 DNA 结合基序是普通的真核转录因子。它有很多亚家族，例如，LacI 家族，AraC 家族，只命名了一些。命名中两个螺旋指压紧和定位 DNA 大沟中的第二个 α -螺旋的第一个 α -螺旋。这些结构域可通过和 HMM，例如，SMART 数据库中可得到的 HTH_ARAC，HTH_ARSR，HTH_ASNC，HTH_CRP，HTH_DEOR，HTH_DTXR，HTH_GNTR，HTH_ICLR，HTH_LACI，HTH_LUXR，HTH_MARR，HTH_MERR，和 HTH_XRE 的分布图对比而鉴定。

文库构建：2、鉴定结构性结构域

多个方法可用于鉴定结构性结构域。编码鉴定的结构域的核酸用来构建核酸文库。进而，编码这些结构域的核酸还可变化（例如，突变）以提供文库编码的另外的结构域。

计算方法。为鉴定另外的天然产生的结构性结构域，已知结构性结构域的氨基酸序列可与已知序列的数据库比较，例如，蛋白质或核酸序列的注解数据库。另一实施方式中，非典型序列例如，没注解的基因组，EST 或全长 cDNA 序列的数据库；典型序列的数据库，例如，SwissProt 或 PDB；和结构域数据库，例如，Pfam，ProDom（Corpet 等（2000）Nucleic Acids Res.28: 267-269），和 SMART（Simple Modular Architecture Research Tool，Letunic 等（2002）Nucleic Acids Res 30, 242-244）可以提供结构性结构域序列。核酸序列数据库可被翻译成所有 6 个阅读框架以和要查询的氨基酸序列比较。标记为编码候选核酸结合结构域的核酸序列可从适当的核酸来源扩增，例如，基因组 DNA 或细胞 RNA。这样的核酸序列可被克隆到表达载体中。基于计算机的结构域鉴定过程可使用寡核苷酸合成仪和自动系统以高产量平台来生产编码结构域的核酸。克隆的编码候选结构域的核酸也可储存在宿主表达载体中并可容易的转移至表达载体中，例如，转移至具有其它结构域（相同的或不同类型）的翻译融合载体中，或通过限制性酶介导的亚克隆或通过位点特异的，

重组酶介导的亚克隆（见 US 专利 No.5, 888, 732）。高产量平台可用来产生含有编码不同候选嵌合体的多孔微量滴定板（multiple 微滴定 plates）。

从起始序列或分布图鉴定结构域的详细方法为本领域熟知。见，例如，Prosite (Hofmann 等, (1999) *Nucleic Acids Res.* 27: 215-219), FASTA, BLAST (Altschul 等, (1990) *J.Mol.Biol.* 215: 403-10) 等。可做简单的字符串搜索来发现与要查询的序列或要查询的分布图具有相同性的氨基酸序列，例如，使用 Perl 扫描文本文件。

如此鉴定的序列可与开始输入的序列具有大约 30%，40%，50%，60%，70%，80%，90%，或更高的相同性。

与查询结构域相似的结构域可在公共数据库中发现，例如，使用 Altschul 等 (1990) *J.Mol.Biol.* 215: 403-10 的 XBLAST 程序 (2.0 版本)。例如，BLAST 蛋白质搜索可使用如下参数的 XBLAST 进行：分数 (score) = 50，字长 (wordlength) = 3。可在查询或搜索的序列中引入缺口，如 Altschul 等, (1997) *Nucleic Acids Res.* 25(17): 3389-3402 中所述。XBLAST 和 Gapped BLAST 程序的缺省参数在 National Center for Biotechnology Information (NCBI), National Institutes of Health, Bethesda MD 可获得。

Prosite profiles PS00028 和 PS50157 可用来鉴定锌指结构域。SWISSPROT 发布了 80, 000 个蛋白质序列，这些分布图分别发现了 3189 和 2316 个锌指结构域。分布图可通过多种不同技术产生的相关蛋白质的多序列对比构建。Gribskov 和同事 (Gribskov 等, (1990) *Meth.Enzymol.* 183: 146-159) 运用符号比较表 (symbol comparison table) 将具有残基频率分布的多序列对比转化为每个位置的权 (weight)。见，例如，PROSITE 数据库和 Luethy 等的结果, (1994) *Protein Sci.* 3: 139-1465。

表示感兴趣的 DNA 结合结构域的隐藏的马尔可夫模型 (HMM's) 可从这样模型的数据库产生或获得，例如，Pfam 数据库, release2.1。可使

用 HMM 搜索数据库, 例如, 使用缺省的参数, 以发现另外的结构域(见, Bateman 等 (2002) *Nucleic Acids Research* 30: 276-280)。另外, 用户可以优化参数。阈值分数可以选择过滤序列数据库以使分数高于阈值的序列表示为候选结构域。Pfam 数据库的描述可见于 Sonhammer 等, (1997) *Proteins* 28 (3): 405-420, 和 HMMs 的详细说明可见于, 例如, Gribskov 等, (1990) *Meth.Enzymol.* 183: 146-159; Gribskov 等, (1987) *Proc.Natl.Acad.Sci.USA* 84: 4355-4358; Krogh 等, (1994) *J.Mol.Biol.* 235: 1501-1531; 和 Stultz 等, (1993) *Protein Sci.* 2: 305-314。

HMM's 的 SMART 数据库 (Simple Modular Architecture Research Tool, Schultz 等, (1998) *Proc.Natl.Acad.Sci.USA* 95: 5857 和 Schultz 等, (2000) *Nucl.Acids Res* 28: 231) 提供了通过 HMMer2 搜索程序 (Durbin 等, (1998) *Biological sequence analysis: probabilistic models of proteins and nucleic acids*. Cambridge University Press) 的隐藏的马尔可夫模型检测鉴定的锌指结构域 (ZnFC2H2; ZnFC2C2; ZnF~C2HC; ZnF C3H1; ZnFC4; ZnF-CHCC; ZnFGATA; 和 ZnF NFX)。

基于杂交的方法。收集的编码多种结构性结构域的核酸可被分析以检测编码保守氨基-和羧基-末端边界序列的序列。简并的寡核苷酸可被设计为与编码这样保守的边界序列的序列杂交。而且, 这样的简并的寡核苷酸的有效性可通过比较它们的组成和已知基因组序列的可能退火位点的频率来估计。如希望, 几轮设计可用来优化简并的寡核苷酸。

比较已知的 Cys₂-His₂ 锌指, 例如, 揭示了天然序列中邻近指间的连接子区的共有序列 (Agata 等, (1998) *Gene* 213: 55-64)。与编码保守的连接子区的核酸相退火的简并的寡核苷酸用来扩增锌指结构域。扩增的编码结构域的核酸用来构建编码锌指嵌合序列的核酸。

文库构建: 3、编码结构性结构域的核酸

用来组成文库的核酸可通过多种方法获得。文库的一些组成核酸可编码天然产生的结构域。另外, 一些组成核酸是通过突变或其它随机化方法获得的变体。组成核酸, 典型的编码单一结构域, 可互相连接而产

生编码不同结构域融合体的核酸。

结构域天然结构的分离。结构域文库可通过从真核生物体如人的基因组 DNA 或 cDNA 中分离编码结构域的核酸序列而构建。多个方法可以做到。例如，计算机搜索获得的氨基酸序列可用来鉴定结构域，如上所述。编码每个结构域的核酸可被分离和插入适于在细胞中表达的载体中，例如，含有启动子，激活结构域，和选择标记的载体。另一实施方式中，与保守基序杂交的简并寡核苷酸用来扩增，例如，通过 PCR，大量含有基序的相关结构域。例如，Kruppel 样 Cys₂His₂ 锌指可用 Agata 等，(1998) Gene 213: 55-64 的方法扩增。这个方法还保留了天然产生的锌指结构域连接子肽序列，例如，具有：Thr-Gly-(Glu/Gln)- (Lys/Arg)-Pro- (Tyr/Phe) (SEQ ID NO: 74)的序列。而且，筛选限于感兴趣的区域，不筛选非选择的基因组或 cDNA 序列的文库，显著降低了文库的复杂性和减少了由于完全筛选大文库固有的困难而带来的丢失所需序列的可能性。

人基因组包含大量锌指结构域，许多是为定性和鉴定的。估计有成千上万的基因编码具有锌指结构域的蛋白质 (Pellegrino 和 Berg, (1991) Proc. Natl. Acad. Sci. USA 88: 671-675)。这些人锌指结构域代表了可以构建新的 DNA 结合蛋白质的大量不同的结构域。许多示范的人锌指结构域描述于 WO 01/60970, U. S. Serial No. 60/374,355, 申请日 April 22,2002, 和 U. S. Serial No. 10/223, 765, 申请日 August 19, 2002。见下表 1。

表 1

ZFD	氨基酸	SEQ ID NO:	取代靶位
CSNR1	YKCKQCGKAFGCPNLRHGRTH	75	GAA>GAC>GAG
CSNR2	YQCNICGKCFSCNSNLHRHQRTH	76	GAA>GAC>GAG
DSAR	YSCGICGKSFSDDSAKRRHCILH	77	GTC
DSCR	YTCSDCGKAFRDKSCLNRHRRTH	78	GCC
HSNK	YKCKECCGKAFNHSSNFNKHRIH	79	GAC
HSSR	FKCPVCGKAFRHSSSLVRHQRTH	80	GTT
ISNR	YRCKYCDRSFSISSNLQRHVNRH	81	GAA>GAT>GAC
ISNV	YECDHCGKAFSIGSNLNVHRRH	82	AAT
KSNR	YGCHLCGKAFSKSSNLRRHEMHH	83	GAG
QAHR	YKCKECCGQAFRQRAHLIRHHKLH	84	GGA
QFNR	YKCHQCGKAFFQSFNLRHERTH	85	GAG
QGNR	FQCNQCGASFQKGNLLRHILH	86	GAA

QSHR1	YACHLCGKAFTQSSHLRRHEKTH	87	GGA>GAA>AGA
QSHR2	YKCGQCGKFYSQVSHLTRHQKIH	88	GGA
QSHR3	YACHLCGKAFTQCSHLRRHEKTH	89	GGA>GAA
QSHR4	YACHLCAKAFIQCSHLRRHEKTH	90	GGA>GAA
QSHR5	YVCRECGRGFRQSHSLVRHKRTH	91	GGA>AGA>GAA>CGA
QSHR	YKCEECGKAFTQSSHLTTHKIIH	92	AGA, CGA>TGA>GGA
QSHV	YECDHCGKSFSQSSHLNVHKRTH	93	CGA>AGA>TGA
QSN1	YMCSECGRGFSQKSNLIIHQORTH	94	AAA, CAA
QSNK	YKCEECGKAFTQSSNLTKHKKIH	95	GAA>TAA>AAA
QSNR1	FECKDCGKAFTQKSNLIRHQORTH	96	GAA
QSNR2	YVCRECRRGFSQKSNLIRHQORTH	97	GAA
QSNR3	YECEKCGKAFNQSSNLTRHKKSH	98	GAA
QSNV1	YECNTCRKTFSQKSNLIVHQORTH	99	AAA>CAA
QSNV2	YVCSKCGKAFTQSSNLTVHQKIH	100	AAA>CAA
QSNV3	YKCECGKNFTQSSNLIVHKRIH	101	AAA
QSNV4	YECDVCGKTFTQKSNLGVHQORTH	102	AAA
QSNT	YECVQCGKGFTQSSNLITHQORVH	103	AAA
QSSR1	YKCPDCGKSFSQSSSLIRHQORTH	104	GTA>GCA
QSSR2	YECQDCGRAFNQSSSLGRHKRTH	105	GTA
QSSR3	YECNECGKFFSQSSSLIRHRRSH	106	GTA>GCA
QSTR	YKCEECGKAFNQSSTLTRHKIVH	107	GTA>GCA
QSTV	YECNECGKAFQNSTLRVHQRIH	108	ACA
QTHQ	YECHDCGKSFRQSTHLTQHRRIH	109	AGA>CGA, TGA
QTHR1	YECHDCGKSFRQSTHLTRHRRIH	110	GGA>AGA, GAA
QTHR2	HKCLECGKCFSONTHLTRHQRT	111	GGA
RDER1	YVCDVEGCTWKFARSDENRHKKRH	112	GCG>GTG, GAC
RDER2	YHCDWDGCGWKFARSDLTRHYRKH	113	GCG>GTG
RDER3	YRCSWEGCEWRFARSDLTRHFRKH	114	GCG>GTG
RDER4	FSCSWKGCERRFARSDLSRHRTH	115	GCG>GTG
RDER5	FACSWQDCNKKFARSDLARHYRTH	116	GCG
RDER6	YHCNWDGCGWKFARSDLTRHYRKH	117	GCG>GTG
RDHR1	FLCQYCAQRFGRKDHLTRHMKKSH	118	GAG, GGG
RDHT	FQCKTCQRKFSRSDHLKTHTRTH	119	AGG, CGG, GGG, TGG
RDKI	FACEVCGVRFTRNDKLKIHMRRKH	120	GGG
RDKR	YVCDVEGCTWKFARSDKLNRRHKKRH	121	GGG>AGG
RSHR	YKCEECGKAFNRRSHLTRHQRIH	122	GGG
RSNR	YICRKCGRGFSRKSNIIRHQORTH	123	GAG>GTG
RTNR	YLCSECDKCFRSTNLIRHRRTH	124	GAG
SSNR	YECKECGKAFSSGSNFTRHQRIH	125	GAG>GAC
VENV	YECDHCGKAFSVSSNLNVHRRIH	126	AAT>CAT>TAT
VSSR	YTCKQCGKAFSVSSSLRRHETTH	127	GTT>GTG>GTA
VSTR	YECNYCGKTFVSSSTLIRHQRIH	128	GCT>GCG
WSNR	YRCEECGKAFRWPSNLTRHKRIH	129	GGT>GGA

如果每个锌指结构域识别唯一一个 3-到 4-碱基对的序列, 需要结合每个可能的 3-到 4-碱基对序列的结构域总数只是 64 到 256 个(4^3 到 4^4)。可能人基因组的天然序列包含足够数目的唯一的锌指结构域包括所有可能的识别位点。这些锌指结构域是珍贵的构建人工嵌合 DNA 结合蛋白质的资源。核酸文库可包括编码含有天然产生的锌指结构域, 该结构域的人工突变体, 和其组合的蛋白质的核酸。

突变的结构域。在一个实施方式中, 文库包括编码至少一个天然产生序列的人工变体的结构性结构域的核酸。一个实施方式中, 这样的变体结构域从简并模式的文库中组装。在核酸结合结构域的情况下, 接近核酸结合界面或接近这样的位置的位置可用来产生诱变。突变的测试锌指结构域, 例如, 通过使用模式化的简并文库, 在任何突变的位置可局限于一套可能的氨基酸。简并密码子可用来在各个位置编码分布图。例如, 可获得只编码疏水残基, 脂肪残基, 或亲水残基的密码子。选择文库选择编码折叠多肽的全长克隆。Cho 等 ((2000) J.Mol.Biol. 297 (2): 309-19) 提供了使用简并寡核苷酸的这样的简并文库的产生方法, 也提供了选择编码全长多肽的文库核酸的方法。这样的核酸可容易的插入表达质粒中, 例如, 使用方便的限制性酶切位点。

在给定的位置选择合适的密码子和每个核苷酸的相关比例可通过简单的检查代表遗传密码子的表, 或计算代数确定。例如, Cho 等, 在前, 描述了接受蛋白质序列的所需的分布图和输出编码该序列的优选的寡核苷酸设计的计算机程序。

另见 Zhang 等, (2000) J.Biol.Chem. 275: 33850-33860; Rebar 和 Pabo (1994) Science 263: 671-673; Segal (1999) Proc.Natl.Acad.Sci USA 96: 2758; Gogus 等, (1996) Proc.Natl.Acad.Sci. USA. 93: 2159-2164; Drier 等, (2001) J.Biol.Chem. 276: 29466-29478; Liu 等, (2001) J.Biol.Chem. 276 (14): 11323-11334; 和 Hsu 等, (1992) Science 257: 1946-50 关于一些可获得的锌指结构域。

文库构建：4、嵌合锌指蛋白质的文库

编码多种嵌合锌指蛋白质的核酸文库可通过系列连接，例如，如实施例 1 所述而形成。可构建该文库而每种核酸编码具有至少 3，4，或 5 个锌指结构域的蛋白质。一些实施方式中，特别是大文库，每个锌指编码片段被设计为随机编码任一个锌指结构域。可选择一套锌指结构域代表具有一定范围特异性的结构域，例如，包含 30，40，50 或更多个 64 种可能的 3 碱基对位点。该套可包括至少 12，15，20，25，30，40 或 50 个不同的锌指结构域。一些或所有这些结构域可是从天然产生的蛋白质种分离的结构域。而且，因为每个 3 碱基对位点很少或不需要多于一个的锌指结构域，可能使用小数目组成结构域产生一个文库，例如，少于 500，200，100，或甚至少于 64 的总组成结构域。

一个示范文库包括编码具有 3 个指和在每个指位置的 30 个可能结构域的嵌合锌指蛋白质的核酸。在其完全的表示形式中，这个文库包括 27,000 个序列（即， 30^3 的结果）。该文库可通过每步添加选自于编码所有 30 个可能的结构域的核酸文库中核酸的系列连接构建。

一个实施方式中，该文库储存为随机收集。另一实施方式中，可分离单个成员，储存于可设地址的位置（例如，阵列的），和测序。高通量筛选 4 万到 5 万个构建的文库成员后，错失的嵌合结合可单独组装以获得完全覆盖。一旦被阵列化，例如，在微滴定盘中，每个单独的成员可随后回收以进一步分析，例如，进行表型筛选。例如，等量的每个阵列化的成员可被集中和转化细胞。选择和定性具有所需表型的细胞。另一实施方式中，每个成员是单独转化细胞的，而且定性细胞，例如，使用核酸为阵列来决定是否内源基因的转录被改变（见“检测 Regulatory Properties of a Chimeric Zinc Finger Protein”，下）。

导入核酸文库到细胞

文库核酸可使用多种方法导入细胞。一个实施方式中，文库储存为包含每个文库核酸的多个复制子的随机的池（pool）。等份的池转化入细

胞。另一实施方案中，单个的文库成员分别储存（例如，在酶标板单独的孔或阵列的单独地点中）及单独导入细胞。

另一实施方案中，文库成员储存于相对于文库整体而言具有降低的复杂性的池中。例如，每个池可包括 10^5 或 10^6 不同成员的文库中的 10^3 不同的文库成员。当一个池被鉴定为具有引起特定作用的成员，该池被展开以鉴定介导表型作用的单个文库成员。这个方法当在回收改变的细胞很困难时是有用的，例如，筛选引起凋亡的嵌合蛋白质。

文库核酸可通过多种方法被导入细胞。示范方法包括电穿孔（见，例如，U.S.5, 384, 253）；微弹轰击技术（microprojectile bombardment techniques）（见，例如，U.S.5, 550, 318；和 5, 610, 042；和 WO 94/09699）；脂质体介导的转染（例如，使用 LIPOFECTAMINE™（Invitrogen）或 SUPERFECT™（QIAGEN GmbH）；见，例如，Nicolau 等，Methods Enzymol., 149: 157-176, 1987）；磷酸钙或 DEAE 葡聚糖介导的转化（见，例如，Rippe 等，（1990）Mol.Cell Biol., 10: 689-695）；直接微注射或超声装载（sonication loading）；受体介导的转染（见，例如，EP273 085）；和土壤杆菌介导的转化（见，例如，U.S.5, 563, 055 和 5, 591, 616）。术语“转化”，入此处所用，涵盖任何导入外源核酸进入细胞的方法。

还可能应用病毒颗粒体外或体内运输文库核酸进入细胞。一个实施方案中，病毒包装用来运送文库核酸进入生物体细胞中。另一实施方案中，文库核酸体外导入细胞，其后细胞转移至生物体中。

导入文库核酸后，文库核酸被表达以至文库编码的嵌合蛋白质被细胞产生。文库核酸的恒定区域可提供必要的调控和支持序列以可以表达。这样的序列可包括转录启动子，转录终止子，剪接位点供者和受者，非翻译调控区（如 polyA 添加位点），细菌的复制起点，指示文库核酸存在的标记或可核酸的选择标记。

筛选编码嵌合蛋白质的核酸文库

筛选中，细胞或生物体被评价以鉴定具有改变了的表型的个体。这

个过程可适合于感兴趣的表型。由于可能的表型数目巨大，筛选的可能性也很大。进行大量的遗传筛选和选择以鉴定突变体或导致特定表型的过表达的天然产生的基因。这些方法的任何一个可适于鉴定编码嵌合蛋白质的核酸文库的有用的成员。筛选，或选择可包括评价含有文库核酸的每种细胞或生物体，例如，评价特定处理后存活或抵抗该处理的细胞或生物体。

评价细胞的示范方法包括显微镜（例如，光学，共聚焦，荧光，扫描电子，和发射电子），基于荧光的细胞分选，差速离心，差异结合，免疫分析，酶促分析，生长分析，和体内分析。

一些筛选涉及特定的环境条件。鉴定对环境敏感或有抗性的细胞。

一些筛选要求监测细胞的特定行为（例如，化学向性，形态变化，或凋亡），或生物体的特定行为（例如，植物的光向性，果蝇的交配行为，等等）。一个实施方案中，细胞或生物体可被直接评价，例如，通过目测，例如，使用显微镜和任选地的计算机软件去自动检测改变的细胞。另一实施方案中，细胞或生物体可使用分析或其它与所需表型相关的指示剂来评价。

一些筛选涉及细胞繁殖。鉴定与参照细胞（例如，正常细胞）繁殖速度不同的细胞。另外，对繁殖信号（例如生长因子或其它有丝分裂原）反应改变的细胞可被鉴定。这些细胞可对信号更或更不敏感。

也可使用涉及细胞分化的筛选。筛选和应用嵌合锌指蛋白质可用来调节多种细胞的分化和繁殖能力，包括干细胞，如 ES 细胞和体细胞，人和其它的。锌指蛋白质可指导 ES 细胞分化到特定品系，如神经元祖细胞或造血干细胞。也可能筛选可以指导分化干细胞为确定的有丝分裂后的细胞类型的锌指蛋白质，例如，指导 ES 细胞和/或神经干细胞分化为多巴胺能或胆碱能神经元。

在其它评价分化的表型中，可能观察标记基因和标记蛋白质的表达。这样的标记的实施例包括：

- FLK1 对于内皮细胞 (Cho 等, (2001) Blood 98: 3635-42; Nishikawa 等, Development 125: 1747-1757),
- 血管平滑肌细胞特异性肌球蛋白重链对于平滑肌(Drab 等, (1997) FASEB J 11: 905-15)
- 骨特异性碱性磷酸酶 (BAP) 和骨钙蛋白 (osteocalcin) 对于成骨细胞, (Demers 等, (2000) Cancer 88: 2919-26)
- CD4, CD8, 和 CD45 对于白细胞(Ody 等, (2000) Blood 96: 3988-90, Martin 等, (2000) Blood 96: 2511-9)
- Flk-2 和 CD34 对于造血干细胞, (Julie 等, Proc.Natl. Acad.Sci. USA , 2001, Vol. 98, Issue 25, 14541-14546, Woodward & Jenkinson. Eur J Immunol 2001 Nov; 31 (11): 3329-38, George AA 等, Blood 2001 Jun 15; 97(12) : 3925-30)
- CFU 对于造血干细胞、MSC 祖细胞 (Frimberger 等, Exp Hematol 2001 May; 29 (5): 643-52)
- Muc-18 (CD146) 对于成纤维细胞, (Filshie 等, (1998) Leukemia 12: 414-21)
- II 型胶原, IV 型胶原和软骨细胞表达的蛋白-68 对于软骨细胞 (Carlberg 等, (2001) Differentiation 67: 128-38, Steck 等, (2001) Biochem J 353 : 169-74)
- 脂肪细胞脂质结合蛋白 (ALBP) 和脂肪酸转运蛋白对于脂肪细胞(Amri 等, (1995) J. Biol.Chem. 270: 2367-2371, Bastie 等, (1999) J Biol Chem 274: 21920-5, Frohnert 等, (1999) J. Biol.Chem. 274, 3970-3977, Teboul 等, (2001) Biochem. J. 360: 305-312)
- CD133 对于神经干细胞 (Uchida N 等, (2000) Proc.Natl.Acad. Sci. USA 97: 14720-5)
- GFAP 对于星形细胞 (Dai 等, (2001) Genes Dev 15: 1913-25)
- 微管相关蛋白-2 对于神经元(Roy 等, (2000) Nat Med 6: 271-7)

还可能筛选哺乳动物细胞的其它特点，如抗肿瘤生成，改变的凋亡，和抗病毒表型。例如，通过选择对病毒感染或病毒产生具有抗性的细胞，可能鉴定可用作抗病毒药的人工嵌合蛋白质。

通过使用与通路活性或失活性相关的探针或通过通路的活性或失活性相关的可见指示可以检测细胞信号通路中的相似变化。

一些筛选涉及产生感兴趣的化合物，例如，代谢化合物，分泌蛋白质，和翻译后修饰的蛋白质。例如，可以鉴定生产更多量化化合物的细胞。另一实施方式中，可以鉴定生产更少量化化合物的细胞，例如，非所需的副产物。感兴趣的细胞可用多种方法鉴定，包括使用应答细胞，微阵列，化学探索分析，和免疫分析。

更多特定实施方案的实施例包括：

1) 蛋白质溶解度：大肠杆菌中，很多异源蛋白质被表达为包涵体。我们鉴定了提高在大肠杆菌中表达的人蛋白质溶解部分的嵌合锌指蛋白质。见实施例 12。因而，本发明提供了改变（例如，提高）表达的（例如，在细胞中过表达）异源蛋白质溶解度的人工转录因子或嵌合锌指蛋白质。

2) 糖基化：包含抗体的治疗蛋白质经常在 CHO 细胞中生产。然而，这样的蛋白质没有最佳的糖基化模式。一个实施方案中，编码嵌合蛋白质的文库被筛选以鉴定 CHO 细胞，该细胞被修饰以便分泌蛋白质，例如，抗体，包含一或多个（例如，所有）糖基化而特征化 B 细胞生产的抗体。因而，本发明提供了改变分泌蛋白质，例如，CHO 细胞分泌的蛋白质，例如，CHO 细胞分泌的抗体的糖基化的人工转录因子或嵌合锌指蛋白质。

3) 病毒滴度：一个实施方案中，编码嵌合蛋白质的文库被筛选以鉴定在细胞培养中提高或降低病毒滴度的嵌合蛋白质。病毒可被用为输送工具，例如，基因输送工具。例如，治疗病毒被发展为治疗某种类型的癌症（例如，腺病毒）。提高病毒滴度对制备治疗病毒有用。另一方面，抑制细胞培养中和体内病毒产生对治疗病毒病有用。因而，本发明提供

了改变（例如，提高或降低）细胞，例如，真核或哺乳动物细胞中病毒生产的人工转录因子或嵌合锌指蛋白质。

4) 转化效率：很多真核细胞系或原核生物体中的遗传工程被低转染或转化效率而限制。可选择人工转录因子，其可修饰细胞以使转染或转化效率提高。选择这样的因子可通过用报道分子或标记在限定的浓度下进行，然后选择吸收了该报道分子的那些细胞。因而，本发明提供了改变（例如，提高）DNA 摄入效率或细胞 DNA 摄入过程的耐受性的人工转录因子或嵌合锌指蛋白质。

5) 饲养细胞。可能鉴定修饰培养细胞特性的的人工转录因子和其它嵌合体以便培养细胞可以支持干细胞的繁殖或分化，例如，因此生产饲养细胞。培养细胞可是人或哺乳动物细胞。该细胞被筛选（例如，通过将文库成员分池（pooling））以鉴定引起在相同环境（例如，相同孔）中培养的干细胞繁殖或分化的细胞。人工转录因子可激活分泌到培养基中的关键细胞因子和生长因子。该培养基可用来诱导分化或繁殖（例如，通过支持自我更新）干细胞。因而，本发明提供了改变（例如，提高）哺乳动物细胞对条件培养基的能力或改变干细胞行为，例如调控干细胞的繁殖或分化的人工转录因子或嵌合锌指蛋白质。

还可能使用一种细胞筛选人工转录因子和然后在第二种细胞中表达该人工转录因子。这个过程可用来一般性的转移转录因子诱导的第一种细胞的表型变化到第二种细胞。例如，我们确定了两种很不同的细胞系总中特定锌指蛋白质的表达分布图；人胚胎肾 293 细胞，其是非癌性的，和人宫颈癌细胞 HeLa。我们发现表达分布图惊人的相似。同样的，在酵母中我们证明了在一个株中 ZFP 诱导的表型可被转移到不同的株中。

干细胞

此处描述的方法，可以通常应用于任何细胞，还特别对调控来自任何后生生物体的干细胞的行为有用。干细胞是具有自我更新能力和分化潜能的细胞。自我更新可被延长，甚至无限。干细胞可产生高度分化的

后代 (Watt 和 Hogan (2000) Science 287: 1427-1430) 最近成功的在培养的人胚胎干细胞中提供了基于细胞的治疗的细胞源。然而, 维持, 复制, 和分化干细胞可以, 至少在一些情况下, 是困难的。ES 细胞, 例如, 具有体外随机分化的趋向。

嵌合转录因子文库可用来鉴定可以控制细胞分化, 例如, 干细胞分化的蛋白质。例如, 一个可以鉴定指导干细胞分化为确定的有丝分裂后的细胞亚型 (例如, 多巴胺能或胆碱能神经元) 的嵌合转录因子。

可以鉴定提高自我更新潜能的蛋白质, 其阻止了分化, 或指导了分化的程度和特征。这些蛋白质一般通过导入编码人工锌指蛋白质的核酸到干细胞或干细胞亲代细胞中而被鉴定, 和评价细胞表型。

控制干细胞分化和繁殖潜能的方法可以在其它方面提供给分化细胞大量的供给, 和对特别细胞类型的分化调控。这样的控制可适合治疗应用或其它应用 (例如, 发展转基因动物, 体外细胞培养, 等等)。

其它实施方式中, 可以鉴定指导胚胎干 (ES) 细胞分化为限定系的嵌合转录因子。因此, 可以从 ES 细胞产生神经元祖细胞或造血干细胞, 还可能鉴定嵌合蛋白质, 例如, 嵌合 ZFP, 其引起 1) 分化的细胞以适应不同的分化状态或 2) 分化的细胞以适应非分化状态, 例如, 因此产生干细胞或多能祖细胞。

鉴定方法不要求有关特定靶基因的信息。靶基因在筛选后被鉴定, 例如, 通过转录或蛋白质分布图来鉴定被鉴定的嵌合转录因子改变了表达或活性的基因。鉴定 ZFP-TFs 调控的基因将提前发现细胞分化。

细胞产物的产生

本发明提供了改变细胞生产细胞产物, 例如, 蛋白质或代谢产物, 能力的人工转录因子 (例如, 嵌合锌指蛋白质)。细胞产物可是内源或异源分子。例如, 可以鉴定提高细胞生产蛋白质, 例如特定的蛋白质 (例如, 特定的内源蛋白质), 过表达的蛋白质, 异源蛋白质, 或错误折叠的蛋白质能力的人工转录因子。

一个实施方案中，筛选细胞生产报道蛋白质的能力，例如，可被酶促方法或荧光方法检测的蛋白质。一个实施方式中，报道蛋白质在参照细胞中过表达时是不溶的。例如，可筛选细菌细胞的减少包涵体的人工转录因子。另一实施方式中，报道蛋白质是分泌的，例如，被原核或真核细胞分泌。可筛选分泌产量较高的，或提高了翻译后修饰，例如，糖基化，磷酸化，或蛋白水解过程的细胞。

一个实施方案中，筛选改变（例如，提高或降低）两个不同报道蛋白质活性的细胞。报道蛋白质可有不同的，例如，活性，定位（例如，分泌的/细胞质的/核的），大小，溶解度，等电点，寡聚状态，翻译后调控，翻译调控，和转录调控（例如，编码它们的基因可被不同的调控序列调控）。本发明包括改变这些特性区别的至少两个不同的报道基因的人工转录因子（例如，锌指蛋白质），和选择性调控报道基因的锌指蛋白质，或这些特性定义的一类报道基因。

因为表型筛选方法可用来分离人工转录因子，不需要知道锌指蛋白质如何介导提高了蛋白质产生。可被证明的可能机制是，包括以下一或多个的改变：翻译机制，转录产物加工，转录，分泌，蛋白质降解，应激抗性，催化活性，例如，代谢物生产。一个实施方式中，人工转录因子可调节代谢通路中一或多个酶的表达和因此提高细胞产物入代谢物或蛋白质的生产。

重复设计（iterative design）

一旦鉴定了嵌合 DNA 结合蛋白质，其改变细胞表型的能力可通过多种策略进一步提高。小文库，例如，具有大约 6 到 200 或 50 到 2000 成员的，或大文库可用来优化特定鉴定的嵌合蛋白质的特性。

在第一个重复设计的示范实施方案中，应用诱变技术改变原始的嵌合 DNA 结合蛋白质。应用该技术来构建第二个文库，其成员包括原始蛋白质的变体，例如，从第一个文库鉴定的蛋白质。这些技术的实例包括：错误倾向的 PCR（Leung 等.（1989）Technique 1: 11-15），重组，使用

随机切割的 DNA 穿梭 (Stemmer (1994) Nature 389-391), Coco 等 (2001) Nature Biotech. 19: 354, 定点诱变 (Zollner 等 (1987) Nucl Acids Res 10: 6487-6504), 结合简并寡核苷酸 (Griffiths 等 (1994) EMBO J 13: 3245) 的盒诱变 (Reidhaar-Olson (1991) Methods Enzymol. 208: 564-586); 系列连接, 从预制的和阵列化的文库中筛选出特异的文库成员, 重组 (例如, 有性 PCR 和“DNA 穿梭TM” (Maxygen, Inc., CA)), 或这些方法的组合。

一个实施方案中, 构建一个文库, 其突变了一套氨基酸位置。例如, 对于嵌合锌指蛋白质, 该套氨基酸位置可位于 DNA 接触残基的附近, 但不是 DNA 接触残基自身。另一实施方案中, 嵌合蛋白质中每个被编码的结构域在文库中是不同的, 但是在比开始文库更限制的程度下, 从中鉴定了嵌合 DNA 结合蛋白质。对嵌合锌指蛋白质, 编码特定结构域的核酸在其它锌指结构域中是不同的, 已知其识别特异性与原始嵌合蛋白质的结构域的相似。

一些技术包括从编码至少两个已知具有特定功能特性的嵌合 DNA 结合蛋白质的结构域的核酸中产生新的嵌合 DNA 结合蛋白质。包括 DNA 穿梭和标准结构域交换的这些技术产生了新的结构域的组合。见, 例如, U.S. 专利 6, 291, 242。DNA 穿梭还可导入另外的点突变到仅仅交换的结构域中。穿梭反应使用编码诱导所需表型的嵌合蛋白质的核酸序列。核酸被穿梭。第二个文库从穿梭产物中产生并筛选诱导所需表型的成员, 例如, 在相似或更严格的条件下。如果开始的文库是完全的以致筛选所有可能的结构域组合, 分离自相同开始文库的结构域的 DNA 穿梭可能没有结果。DNA 穿梭可能在覆盖完全的情况下和全面筛选不实际的情况下有用。

第二个示范的重复设计的实施方案中, 产生了所需表型的嵌合 DNA 结合蛋白质被不同的每个结构域所改变。结构域可按顺序被改变, 例如, 一个接一个, 或同时改变多个。以下实施例提到包括 3 个锌指结构域的原始嵌合蛋白质: 指 I, II, 和 III 和产生所需表型的那个。构建第二个文

库以便第二个文库的每种核酸成员如最初鉴定的蛋白质那样编码相同的指 II 和指 III。然而，该文库包括与原始蛋白质的指 I 不同的指 I 的核酸成员。区别可能是单一核苷酸，其改变了被编码的嵌合蛋白质的氨基酸序列或可能是更基本的差别。可构建第二个文库，例如，所以指 I 的碱基接触残基是不同的，或者指 I 的碱基碱基接触残基保留了而临近的残基是不同的。第二个文库还可包括大量足够的锌指结构域以识别至少 20, 30, 40, 或 60 个三核苷酸位点。

筛选第二个文库以鉴定改变细胞或生物体表型的成员。改变的程度可与原始蛋白质产生的相似或比原始蛋白质产生的更大。

目前，或以后，可构建第三个文库，其指 II 不同，和构建第四个文库，其指 III 不同。可能不需要进一步改良所有结构域都不同的嵌合蛋白质，如果嵌合蛋白质或已经鉴定的突变体已经足够了。其它情况下，需要再优化每个结构域。

如果其它结构域目前不同，每个特定文库的改良的变体可互相组合以产生另一个文库。相似的筛选这个文库。

重复设计的第三个示范实施方案中，方法包括添加，取代，和缺失一个结构域，例如，锌指结构域或调控结构域。另外的锌指结构域可提高嵌合蛋白质的特异性和提高结合亲和力。某些情况下，提高的结合亲和力可以提高嵌合蛋白质产生的表型。另外的调控结构域，例如，第二个激活结构域或动员辅助因子的结构域，也可以提高嵌合蛋白质产生的表型。缺失可以提高或扩大嵌合蛋白质的活性的特异性，依赖于被缺失的结构域的贡献，等等。

在重复设计的第四个示范实施方案中，方法包括在细胞中共表达原始的嵌合蛋白质和第二个嵌合 DNA 结合蛋白质。第二种嵌合蛋白质也可通过筛选编码不同嵌合体的核酸文库来鉴定。一个实施方案中，第二种嵌合蛋白质通过筛选在细胞中表达原始嵌合蛋白质的文库来鉴定。另一实施方案中，第二种嵌合蛋白质是独立鉴定的。

嵌合锌指蛋白质的检测调控特性

改变细胞表型的嵌合转录因子可进一步通过鉴定其直接或间接调控的内源基因而定性。典型的，嵌合转录因子是在细胞中产生的。在适当的时间，例如，表型改变发生之前，之中，或之后，分析细胞以确定细胞中或围绕细胞的培养基中的转录产物或蛋白质的水平。例如，可从细胞中收获 mRNA 并用核酸微阵列分析。

核酸微阵列可通过多种方法构造，例如，光刻方法（见，例如，U.S. 专利 5, 510, 270），机械方法（例如，U.S. 专利 5, 384, 261 描述的 directed-flow 法），和基于针点的（pin based）方法（例如，U.S. 5, 288, 514 所描述）。用每个地址处的独特捕获探针合成阵列，每个捕获探针适于检测特定表达基因的核酸。

可通过常规方法分离 mRNA，例如，包括 Dnase 处理以除去基因组 DNA 和与 oligo-dT 偶联的固体底物杂交（例如，如 Current Protocols in Molecular Biology, John Wiley & Sons, N. Y 所述）。洗底物，并洗脱下 mRNA。分离的 mRNA 于是被逆转录和任选地扩增，例如，通过 rtPCR，例如，如（U.S. 专利 4, 683, 202）所述。核酸可在扩增或逆转录中被标记，例如，通过整合标记核苷酸。优选的标记的实施例包括荧光标记，例如，红色荧光染料 Cy5（Amersham）或绿色荧光染料 Cy3（Amersham）。另外，核酸可用生物素标记，并在杂交后用标记的链霉亲和素检测，例如，链霉亲和素-藻红蛋白（Molecular Probes）。

标记的核酸于是与阵列接触。另外，对照核酸或参照核酸与相同的阵列接触。对照核酸或参照核酸可用不同于样品核酸标记的标记物标记，例如，使用不同最大发射波长的标记物。标记的核酸在杂交条件下与阵列接触。洗阵列，并成像以检测阵列每个地址的荧光。

产生和评价分布图的一般方法包括检测阵列每个地址的杂交。每个地址杂交的程度是通过数值表示的并储存于，例如，载体，一维矩阵，或一维点阵中。载体 x 对阵列的每个地址都有一个值。例如，在特定地

址杂交程度的数值是储存于变量 x_a 中。可调整数值, 例如, 对局部背景水平, 样品量, 和其它变量。从参照样品中制备核酸并与同一或不同阵列杂交。载体 y 是与载体 x 相同的构建体。样品表达分布图和参照分布图可相比较, 例如, 使用两个载体的函数的数学方程式。比较可评价维数值, 例如, 分数代表两个分布图的相似性。每个或两个载体可通过矩阵被转化以给阵列不同检测到的基因加上加权值。

表达数据可以储存在数据库中, 例如, 关系数据库如 SQL 数据库(例如, Oracle 或 Sybase 数据库环境)。数据库可有多个表格。例如, 原始数据可储存在一个表格中, 其中每栏对应于被阵列化的基因, 例如, 地址或阵列, 和每排对应于样品。分开的表格可储存鉴定物和样品信息, 例如, 使用的阵列的批号, 日期, 和其它质量控制信息。

被相似调控的基因可通过簇表达数据鉴定以鉴定共调控的基因。这样的簇可指示被嵌合锌指蛋白质协同调控的基因。基因可使用分等级的簇而被分簇(见, 例如, Sokal 和 Michener(1958) Univ. Kans. Sci. Bull. 38: 1409), 贝叶斯簇, k-means 簇, 和自组织图(见, Tamayo 等 (1999) Proc. Natl. Acad. Sci. USA 96: 2907)。

样品表达分布图 (expression profile) 对参照表达分布图(例如, 对照细胞)的相似性也可, 例如, 通过比较样品表达水平的对数值和预测值或参照表达值的对数值而确定, 并通过对分布图中所有基因预测值的加权因子调整比较值。

也可测量具有活性嵌合蛋白质的细胞中的蛋白质表达分布图。蛋白质分布图一个的示范方法包括 2-D 凝胶电泳和质谱去定性各个蛋白质种类。2-D 凝胶上的各个“点”被蛋白水解, 然后在质谱中分析。这个方法可鉴定蛋白质组成和, 在很多情况下, 鉴定翻译修饰。

蛋白质和核酸表达分布图方法可不仅提供关于嵌合蛋白质特性的信息, 也可提供关于细胞中天然机制的信息。例如, 通过嵌合蛋白质的表达而上调的蛋白质或核酸可以是嵌合蛋白质表达引起的表型变化的天然

效应子。

另外，其它方法可用来鉴定人工嵌合蛋白质直接或间接调控的靶基因和蛋白质。一个实施方式中，定性了补偿（例如，抑制）人工嵌合蛋白质表型作用的改变。这些改变包括遗传改变如染色体基因突变和特定基因过表达，及其它改变，如 RNA 干涉（例如，通过双链 RNA）。

一特定实施方式中，分离了当在细胞中条件表达时，例如，病原体细菌或真菌，引起生长缺陷或致命性的嵌合 ZFP。这样的 ZFP 可通过转化细胞 ZFP 文库而鉴定，该文库包括编码 ZFP 的核酸，该核酸的表达受可诱导的启动子控制。转化体培养在非诱导培养基中并在诱导和非诱导板上复制。在非诱导板上正常生长，但在诱导板上显示生长缺陷的菌落被鉴定为“条件致命性”或“条件生长缺陷”菌落。

（a）使用 cDNA 文库鉴定靶基因

cDNA 表达文库被转化入上述的“条件致命性”或“条件生长缺陷”株中。把转化体在诱导板上铺板。分离存活的菌落，尽管存在和表达了引起缺陷的 ZFP。定性补偿了缺陷的 cDNA 的核酸序列。这些 cDNA 可是介导缺陷的嵌合 ZFP 调控的直接或间接靶基因的转录产物。

（b）使用二级 ZFP 文库鉴定靶基因

鉴定了抑制第一种嵌合蛋白质作用的第二种嵌合蛋白质。鉴定了第二种嵌合蛋白质（存在或不存在第一种嵌合蛋白质）的靶。

例如，ZFP 文库被转化入“条件致命性”或“条件生长缺陷”菌落（其包含引起缺陷的第一个嵌合 ZFP）。把转化体在诱导板上铺板。通过导入的 ZFP 的表达可以存活的菌落被鉴定为“被抑制的菌株”。第二个靶基因可以通过 DNA 微阵列分析定性。比较分析可在 4 个菌株中进行：1）没有 ZFP；2）只有第一个 ZFP；3）只有第二个 ZFP；和 4）有第一个和第二个 ZFP。例如，被第一个和第二个嵌合 ZFP 以相反方向调控的基因是介导生长缺陷表型的候选靶。这个方法可用于任何表型，不仅是生长缺陷。

(c) 通过表达分布图分析鉴定的共调控基因

通过表达分布图鉴定了嵌合 ZFP 的候选靶。随后，为确定是否候选靶介导嵌合 ZFP 的表型，候选靶可独立的过表达或抑制（例如，通过遗传缺失或 RNA 干涉）。另外，可以应用这个分析到多个候选靶，因为至少一些情况下多于一个候选物需要被干扰以引起表型。这个方法的实施例提供在实施例 3 中（酮康唑抗性）。

(d) 时间过程分析

细胞暴露于嵌合 ZFP 后，通过基因表达对于时间的变化鉴定嵌合 ZFP 的靶。例如，编码嵌合 ZFP 的基因可隶属于一个可诱导的启动子。一个示范的可诱导启动子被小分子如强力霉素调控。编码嵌合 ZFP 的基因被导入细胞。在诱导了可诱导启动子后的不同时间后，从细胞中获得 mRNA 样品。见，例如，图 13，其描绘了 ZFP F104-p65 诱导过程中被激活和阻抑的基因。

(e) 使用蛋白质转导和 cDNA 微阵列技术鉴定哺乳动物细胞的 ZFP-TFs 的主要靶基因。

可以通过转导导入嵌合蛋白质到细胞中。将蛋白质提供至胞外环境，而细胞将蛋白质转移给自身。因此，细胞不必要包括编码嵌合蛋白质的基因。这个方法可以不必考虑外源核酸的整合和扩增等问题。可以精确控制蛋白质的水平。一个实施方案中，嵌合 ZFP 被融合到 Tat 或 VP22 的蛋白质转导结构域中。

为分析培养细胞中转导的嵌合蛋白质的作用，加入嵌合蛋白质到培养基中，例如，融合到蛋白质转导结构域中。可通过加入蛋白质合成抑制剂如放线菌酮提高调控的靶基因的检测。因此，主要靶基因的翻译被阻断，则主要靶基因编码的蛋白质调控的基因可被检测出。主要靶基因的特性可在 DNA 微阵列分析中发现。

应用放线菌酮鉴定主要靶基因也可用于当嵌合蛋白质被细胞中异源核酸编码的时候。例如，诱导的异源核酸表达少于 30, 20, 15, 10, 或

5 分钟时, 加入放线菌酮。

(f) 活性分析

潜在的靶基因的功能可通过抑制靶基因的活性来评价, 例如, 用双链 RNA (dsRNA) 的 RNA 干涉 (RNAi), 反义技术, 核酶, 或靶向性遗传突变。靶基因活性被降低的细胞或生物体可被评价及与没有 RNAi 处理的对照细胞或生物体比较。另一实施方式中, 表达了被认为调控靶基因的人工锌指蛋白质的细胞或生物体用 RNAi 处理。人工锌指蛋白质诱导表型的能力在存在和不存在 RNAi 时评价。一些情况下, 如果潜在的靶基因确实是关键靶, RNAi 处理失活潜在靶可以减弱人工锌指蛋白质诱导的表型。

dsRNA 可通过双向转录一个盒产生, 例如, 通过在该盒的每一侧都含有 T7 启动子。选择盒的这个插入以便其包含与潜在的靶基因互补的序列。另见, HiScribe™ RNAi Transcription Kit (New England Biolabs, MA) 和 Fire, A. (1999) Trends Genet. 15, 358-363。dsRNA 可被消化为较小的片段。见, 例如, US Patent Application 2002-0086356。dsRNA 可用来使哺乳动物细胞的基因不表达。见, 例如, Clemens 等 (2000) Proc.Natl. Sci.USA 97, 6499-6503; Billy, E. 等 (2001) Proc. Natl. Sci. USA 98,14428-14433; Elbashir 等(2001) Nature. 411 (6836): 494-8; Yang, D. (2002) Proc. Natl. Acad. Sci. USA 99, 9942-9947。

靶 DNA 位点鉴定

对于嵌合 DNA 结合蛋白质, 多种方法可用来确定产生感兴趣表型的嵌合 DNA 结合蛋白质的靶位点。这样的方法可用来, 单独或组合, 以发现这样的位点。

一个实施方案中, 表达分布图的信息用来鉴定嵌合锌指蛋白质识别的靶位点。比较嵌合锌指蛋白质共调控的基因的调控区域以鉴定与所有或很多调控区域共同的基序。

另一实施方案中, 生物化学手段用来确定嵌合锌指蛋白质结合什么

样的 DNA 位点。例如，染色质免疫沉淀实验可用来分离嵌合锌指蛋白质结合的核酸。分离的核酸通过 PCR 扩增和测序。见，例如，Gogus 等 (1996) Proc. Natl. Acad. Sci. USA. 93: 2159-2164。SELEX 方法是可使用的示范方法。进而，关于嵌合锌指蛋白质的各个锌指结构域的结合特异性的信息可用来预测靶位点。预测可被验证或可用来指导对其它结果的解释（例如，染色质免疫沉淀，共调控基因的 *silico* 分析，和 SELEX）。

另一实施方案中，潜在的靶位点是基于关于每个组成的锌指的结合特异性的信息而推断的。例如，如图 10A 和 10B 所示，含有锌指结构域的嵌合体，从 N-到 C-末端，：CSNR，RSNR，和 QSNR 预期识别靶位点 5'-GAAGAGGACC-3'(SEQ ID NO: 130)。结构域 CSNR，RSNR，和 QSNR 具有以下分别的 DNA 结合特异性 GAC，CAG，和 GAA。预期的靶位点的形成是通过考虑 C 末端到 N 末端顺序的结构域和连接它们的识别特异性而获得一条链上 5'到 3'顺序的靶位点。

尽管大多数情况下，嵌合锌指蛋白质很可能具有转录调节子的功能，有些情况下可能嵌合锌指蛋白质通过结合 RNA 或蛋白质靶介导表型作用。一些天然产生的锌指蛋白质实际上结合这些巨大分子。

嵌合转录因子的另外特征

对于编码嵌合核酸结合结构域的文库，被编码的多肽也可包括一个或多个以下特征。这些特征可以在所有成员中稳定或可以变化。一个实施方式中，一些核酸编码的多肽含有激活结构域，然而其它含有阻抑结构域，或没有转录调控结构域。

激活结构域。可用于本发明中的转录激活结构域包括但不限于酵母的 Glc4 激活结构域和单纯疱疹病毒的 VP16 结构域。证实结构域激活转录的能力可通过融合该结构域到已知 DNA 结合结构域再确定是否可操纵的连接到被已知 DNA 结合结构域识别的位点的报道基因被融合蛋白质激活。

示范激活结构域是 p65 的以下结构域：

YLPDTDDRHRIEEKRKRTYETFKSIMKKSPFSGPTDPRPPRRRIAVPSRS
 SASVPKPAPQYPYFTSSLSTINYDEFPTMVFPSPGQISQASALAPAPPQVL
 PQAPAPAPAPAMVSALAQAAPVPVLAPGPPQAVAPPAPKPTQAGEGT
 LSEALLQLQFDDDELGALLGNSTDPVFTDLASVDNSEFQQLLNQGIP
 VAPHTTEPMLMEYPEAITRLVTAQRPPDPAPAPLGAPGLPNGLLSGDED
 FSSIADMDFSALLSQ (SEQ ID NO: 131)

示范的 Gla4 激活结构域的序列如下:

NFNQSGNIADSSLSFTFTNSSNGPNLITTQTNSQALSQPIASSNVHDNF
 MNNEITASKIDDGNNSKPLSPGWDQTAYNAFGITTGMFNTTTMDDV
 YNYLFDDDEDTPPNPK KEISMAYPYDVPDYAS (SEQ ID NO: 132)

细菌中, 激活结构域功能可被一个结构域模仿, 该结构域补充了野生型 RNA 聚合酶 α -亚基 C 末端结构域或突变 α -亚基 C 末端结构域, 例如, C 末端结构域融合到蛋白质相互作用结构域。

阻抑结构域。如需要, 可以将阻抑结构域而不是激活结构域融合到 DNA 结合结构域。真核阻抑结构域的实施例包括 Kid, UME6, ORANGE, groucho, 和 WRPW 的阻抑结构域(见, 例如, Dawson 等, (1995) Mol. Cell Biol. 15: 6923-31)。证实结构域抑制转录的能力可通过融合该结构域到已知 DNA 结合结构域并确定是否可操纵连接到被已知 DNA 结合结构域识别的位点的报道基因被融合蛋白质抑制。

示范阻抑结构域是 UME6 蛋白质的以下结构域:

NSASSSTKLDDDLGTAAAVLSNMRSSPYRTHDKPISNVNDMNNTNAL
 GVPASRPHSSSFPSKGVLRPILLRIHNSEQQPIFESNNSTACI (SEQ ID
 NO: 133)

另一示范阻抑结构域来自 Kid 蛋白质:

VSVTFEDVAVLFTRDEWKKLDLSQRSLYREVMLENYSNLASMAGFLF
 TKPKVIS LLQQGEDPW (SEQ ID NO: 134)

其它嵌合转录因子既不包括激活又不包括阻抑结构域。可能这样的

转录因子通过取代或和结合的内源转录因子（例如，激活子或阻抑子）竞争而改变转录。

肽连接子。DNA 结合结构域可通过多种连接子相连。使用和设计连接子为本领域熟知。特定的有用的连接子是核酸编码的肽连接子。因此，可以构建编码第一个 DNA 结合结构域，肽连接子，和第二个 DNA 结合结构域的合成基因。这个设计可重复以构建大的，合成的，多结构域的 DNA 结合结构域。PCT WO 99/45132 和 Kim 和 Pabo ((1998) Proc.Natl. Acad.Sci. USA 95: 2812-7)描述了适合连接锌指结构域的肽连接子。

可获得形成随机卷曲， α -螺旋或 β -折叠三级结构的其它肽连接子。形成适当灵活连接子的多肽为本领域熟知（见，例如，Robinson 和 Sauer (1998) Proc Natl Acad Sci USA. 95: 5929-34）。灵活连接子典型的包括甘氨酸，因为这个缺少侧链的氨基酸是唯一可以自由转动的氨基酸。丝氨酸或苏氨酸可分布于连接子以提高亲水性。另外，可使用能与 DNA 磷酸骨架相互作用的氨基酸以提高结合亲和力。明智的使用这些氨基酸可以平衡提高的亲和力和丧失的序列特异性。如果作为连接子需要严格的范围，可以使用 α -螺旋连接子，如 Pantoliano 等 (1991) Biochem. 30: 10117-10125 描述的螺旋连接子。连接子也可用计算机模型设计（见 U. S. 4,946, 778）。分子模型软件可商业获得（例如，从 Molecular Simulations, Inc., San Diego, CA）。连接子可任选地优化，例如，使用蛋白质工程领域的标准诱变技术和适当的生物物理测试，和此处描述的功能分析来降低抗原性和/或提高稳定性。

对使用锌指结构域的实施方案，锌指间天然产生的肽可用来连接锌指。天然产生的连接子的实施例是：Thr-Gly-(Glu 或 Gln)-(Lys 或 Arg)-Pro-(Tyr 或 Phe) (SEQ ID NO: 74) (Agata 等，见上)。通常，连接子被选择或基于连接天然产生蛋白质的锌指的序列。

二聚体结构域。连接 DNA 结合结构域的另一方法是使用二聚体结构域，特别是异二聚体结构域（见，例如，Pomerantz 等(1998) Biochemistry

37: 965-970)。这个实施方案中, DNA 结合结构域位于不同的多肽链中。例如, 第一种多肽编码 DNA 结合结构域 A, 连接子, 和结构域 B, 而第二种多肽编码结构域 C, 连接子, 和结构域 D。技术人员可以从许多熟知的二聚体结构域中选择二聚体结构域。如果不需要同源二聚体, 可以使用倾向异源二聚体的结构域。一个特定的可接受的二聚体结构域是卷曲-卷曲基序, 例如, 二聚体平行或反平行卷曲-卷曲。也可获得优选形成异源二聚体的卷曲-卷曲序列 (Lumb 和 Kim, (1995) *Biochemistry* 34: 8642-8648)。另一类二聚体结构域是被小分子或信号事件引起二聚体化。例如, FK506 的二聚体形式可用来二聚体化两个 FK506 结合蛋白质 (FKBP) 结构域。这样的二聚体结构域可用来提供另外的调控水平。

对非 DNA 应用的嵌合蛋白质

还可能修改此处的实施例以建立编码含有不同非 DNA 结合结构域, 例如, 细胞内信号转导结构域 (例如, SH2, SH3, PDZ, Che 结构域, 或激酶结构域) 的嵌合蛋白质的核酸文库。文库编码的嵌合蛋白质可被细胞表达, 并鉴定具有改变了表型特性的细胞。例如, 可以鉴定降低或提高细胞繁殖速度的信号结构域不同组合形成的嵌合蛋白质。

表达锌指蛋白质

此处描述的方法可包括使用分子生物学, 生物化学, 经典遗传学, 和重组遗传学领域的常规技术。揭示了本发明中应用的一般方法的基础教材包括 Sambrook 等 *Molecular cloning, A Laboratory Manual* (2nd ed. 1989); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); 和 *Current Protocols in Molecular Biology* (Ausubel 等, eds., 1994)。

除了此处描述的其它方法, 编码锌蛋白质的核酸可用合成的寡核苷酸作为连接子构建合成基因来构建。另一实施方式中, 使用合成的寡核苷酸和/或引物扩增编码一或多个锌指结构域的序列, 例如, 从 RNA 或 DNA 模板, 人工或合成的锌指结构域。见 U. S. Patents 4,683, 195 和 4,683, 202; *PCR Protocols: A Guide to Methods and Applications* (Innis 等, eds,

1990)。如聚合酶链式反应(PCR)的方法可用来直接从 mRNA, cDNA, 基因组 cDNA, 或锌指蛋白质文库扩增核酸序列。简并寡核苷酸可设计来使用此处提供的序列扩增同源物。限制性内切酶位点可整合进引物中。

可通过本领域已知的技术,例如 mRNA 的逆转录和扩增,分离总 RNA 或 polyA⁺ RNA, northern 印渍, 点印渍, 原位杂交, RNase 保护, 核酸阵列技术, 例如如此的技术分析锌指蛋白质的基因表达。

编码人工锌指蛋白质的多核苷酸在转化进原核或真核细胞中复制和/或表达前可克隆至载体中。这些载体典型的是原核载体, 例如, 质粒, 噬菌体或穿梭载体, 或真核载体。

蛋白质表达。为获得编码人工锌指蛋白质的多核苷酸的重组表达(例如, 高水平), 可以亚克隆相关编码核酸到含有指导转录的强启动子, 转录/翻译终止子, 和为翻译起始的核糖体结合位点的表达载体中。适合的细菌启动子为本领域熟知和描述, 例如, Sambrook 等, 和 Ausubel 等, 见上。可获得为表达的细菌表达系统, 例如, 大肠杆菌, 杆菌属, 和沙门氏菌属(Palva 等, (1983) Gene 22: 229-235; Mosbach 等, (1983) Nature 302: 543-545)。可商业获得这些表达系统的试剂盒。也可商业获得本领域熟知的哺乳动物细胞, 酵母(例如, 酿酒酵母, 裂殖酵母, 毕赤酵母, 和 Hanseula), 和昆虫细胞的真核表达系统。

用于指导异源核酸表达的启动子的选择依赖域于特定的应用。启动子优选的与异源转录起点的位置与其天然状态时转录起点具有相同的距离。但是如本领域所知, 这个距离可以变化而不丧失启动子功能。

除了启动子, 表达载体典型的包含宿主细胞中表达所需的所有额外元件的转录单位或表达盒。典型的表达盒因此包含可操纵的连接到编码核酸序列的启动子和转录产物有效多腺苷化, 核糖体结合位点, 和翻译终止所需的信号。该盒的另外元件可包括增强子和, 如果基因组 DNA 用作结构基因, 具有功能性剪接供者位点和受者位点的内含子。

除了启动子序列, 表达盒还应包含结构基因的转录终止区下游以提

供有效的终止。终止区可从启动子序列的同一基因中获得或从不同基因中获得。

用来转运遗传信息到细胞的特定表达载体不是特别重要的。可以使用用于在真核或原核细胞中表达的任何常规载体。标准的细菌表达载体包括质粒如 pBR322 为基础的质粒, pSKF, pET23D, 和融合表达系统如 MBP, GST, 和 LacZ。还可将表位标签加入重组蛋白质中以提供常规的分离方法, 例如, c-myc-, 或六组氨酸标签。

表达载体可包含真核病毒的调控元件, 例如, SV40 载体, 刺瘤病毒载体, 和衍生于 Epstein-Barr 病毒的载体。其它示范真核载体包括 pMSG, pAV009/A⁺, pMTO10/A⁺, pMAMneo-5, 杆状病毒 pDSVE, 和任何其它可以表达在 CMV 启动子, SV40 早期启动子, SV40 晚期启动子, 金属流因启动子, 鼠乳腺癌病毒启动子, 劳氏肉瘤病毒启动子, 多角体蛋白启动子, 或其它所示启动子的指导下可以在真核细胞中有效表达的载体。

真核载体表达蛋白质还可使用可诱导的启动子调控。用可诱导的启动子, 表达水平与诱导剂如四环素或蜕化素的浓度相关, 其整合这些试剂的反应元件到启动子中。通常, 只有存在诱导试剂时可由可诱导的启动子获得高水平的表达; 基本表达水平很低。如果感兴趣的蛋白质的表达对真核细胞是致命的, 经常选择可诱导的表达载体。

一些表达系统具有提供了基因扩增的标记如胸苷激酶和二氢叶酸还原酶。另外, 不涉及基因扩增的高产出表达系统也适用, 如使用在多角蛋白启动子或其它强杆状病毒启动子指导下的昆虫细胞的杆状病毒载体, 其具有线粒体呼吸链蛋白质编码序列和糖酵解蛋白质编码序列。

典型的包含在表达载体中的元件还包括大肠杆菌中起作用的复制子, 其是编码抗生素抗性的基因以允许选择吸收了重组质粒的细菌, 和在质粒非基本区域的单一性的限制位点以可以插入真核序列。可选择原核序列以便它们不干扰真核细胞中 DNA 的复制。

标准的转染方法用来产生表达大量锌指蛋白质的细菌, 哺乳动物,

酵母或昆虫细胞系，该蛋白质随后使用标准技术纯化（见，例如，Colley 等, J.Biol.Chem. 264: 17619-17622 (1989); Guide to Protein Purification, in Methods in Enzymology, vol. 182 (Deutscher, eds., 1990)）。真核和原核细胞的转化按照标准技术进行（见，例如，Morrison, J.Bact. 132: 349-351 (1977); Clark-Curtiss & Curtiss, Methods in Enzymology 101: 347-362 (Wu 等, eds, 1983)）。

可以使用任何熟知的导入外源核苷酸序列进入宿主细胞的方法。这些包括使用磷酸钙转染，原生质体融合，电穿孔，脂质体，微注射，原生质载体，病毒载体和任何其它已知的导入克隆的基因组 DNA，cDNA，合成 DNA 或其它外源遗传物质进入宿主细胞的方法（见，例如，Sambrook 等，见上）。不需要使用的特定的遗传工程过程能够成功的导入至少一个基因进入宿主细胞。

表达载体导入细胞后，转染的细胞在适合表达或激活表达的条件下培养。然后可从细胞提取物，细胞膜组分或囊泡，或培养基中分离蛋白质。

具有适当调控序列的表达载体可用来在模型生物体，例如果蝇，线虫，斑马鱼，爪蟾，或小鼠中表达编码人工锌指的异源基因。见，例如，Riddle 等, eds., C.elegans II. Plainview (NY): Cold Spring Harbor Laboratory Press; 1997。

蛋白质纯化。锌指蛋白质可从任何适合的表达系统，例如，上述那些系统产生的物质中纯化出来。

锌指蛋白质可用标准技术纯化非常纯，包括用如硫酸铵等物质选择沉淀；柱层析，亲和纯化，免疫纯化方法，和其它的方法（见，例如，Scopes, Protein Purification: Principles and Practice (1982); U. S. Patent No. 4,673, 641; Ausubel 等，见上；和 Sambrook 等，见上）。例如，锌指蛋白质可包括用于纯化的亲和标签，例如，和其它步骤组合。

重组蛋白质被转化细菌大量表达，典型的是启动子诱导后；但是表

达可是组成性的。用 IPTG 诱导启动子是可诱导的启动子系统的一个实施例。细菌按照本领域的标准方法生长。新鲜或冻存的细菌细胞用来分离蛋白质。细菌表达的蛋白质可以形成不溶的聚集物（“包涵体”）。几个方法适于从包涵体中纯化蛋白质。见，例如，Sambrook 等，见上；Ausubel 等，见上。如果蛋白质可溶或运至外周胞质，可从细胞裂解物或外周质制备物中获得。

差速离心。可使用盐溶或盐析以选择沉淀锌指蛋白质或污染蛋白质。一个示范盐是硫酸铵。硫酸铵基于蛋白质的溶解度沉淀蛋白质。蛋白质越疏水，在较低硫酸铵浓度下其越易于沉淀。典型的方法包括加入饱和硫酸铵到蛋白质溶液以便所得硫酸铵浓度在 20%-30%之间。这个浓度沉淀许多较疏水的蛋白质。分析沉淀物以确定感兴趣的蛋白质被沉淀了还是在上清中。加入硫酸铵到上清至可以沉淀感兴趣的蛋白质的已知浓度。然后在缓冲液中溶解沉淀，如需要通过透析或过滤浓缩法去除多余的盐。

柱色谱。基于蛋白质的大小，净表面电荷，疏水性，和配体亲和性，可以从其它蛋白质中分离锌指蛋白质。另外，蛋白质的抗体可偶联在柱基质中而免疫纯化蛋白质。所有这些方法为本领域熟知。色谱技术可在任何范围及使用很多不同制造者（例如，Pharmacia Biotech）的设备进行。见，一般的，Scopes, *Protein Purification: Principles and Practice* (1982)。

类似的常规蛋白质纯化过程可用来回收生产细胞中人工锌指蛋白质表达而改变（例如，提高）了其生产的蛋白质。

以基因和细胞为基础的治疗方法

本发明的一个方面，如上述，提供了编码人工锌指蛋白质的分离的 DNA 分子。这些分离的 DNA 分子可插入为基因治疗目的的多种 DNA 构建体和载体。如此处所用，“载体”是可以转运共价连接其上的另一核酸分子的核酸分子。载体包括质粒，粘粒，人工染色体，和病毒元件。载体可在宿主细胞中复制或整合入宿主 DNA。病毒载体包括，例如，复制缺陷的逆转录病毒，腺病毒和腺相关病毒。基因治疗载体是设计为施用

给对象，例如，哺乳动物的载体，所以对象的细胞可以表达载体含有的治疗基因。

基因治疗载体可包含调控元件，信号序列，5'非翻译区，聚腺苷化位点，和3'调控区。例如，5'调控元件，增强子或启动子可以调控DNA编码治疗多肽的转录。调控可是组织特异性的。例如，调控可以限制所需基因转录至脑细胞中，例如，皮质神经元或胶质细胞；造血细胞；内皮细胞。另外，调控元件可包含在应答外源药物，例如，胆固醇，四环素等的物质中。因此，治疗锌指多肽（例如，调控VEGF的多肽）表达的水平和时间可被控制。

可将用于输送的基因治疗载体制备为裸核酸，如病毒或失活病毒的组分，或脂质体或其它输送工具的内含物。另外，基因输送试剂，例如，病毒载体，可从产生基因输送系统的重组细胞中产生。适当的病毒载体包括逆转录病毒，例如，莫罗尼逆转录病毒，腺病毒，腺相关病毒，和慢病毒，例如，单纯疱疹病毒（HSV）。HSV对感染神经系统细胞是潜在有用的。

基因治疗载体可施用给对象，例如，通过静脉注射，通过局部施用（见U. S. Patent 5,328, 470）或通过stereotactic注射（见例如，Chen等（1994）Proc. Natl. Acad. Sci. USA 91: 3054-3057）。基因治疗试剂可进一步配方化，例如，通过缓释物质减慢或延长试剂的释放。提供重组治疗3-结构域多肽的方法是插入一个基因治疗载体到对象的骨髓细胞中。细胞被感染，例如，用逆转录病毒基因治疗载体，并在培养基中生长。同时，辐射处理对象以耗竭对象的骨髓细胞。然后将感染的培养细胞补充到对象的骨髓。监测对象恢复及治疗多肽的产生。

基于细胞的治疗方法包括将编码人工锌指蛋白质的可操纵的连接到启动子的核酸导入培养的细胞中。可选择人工锌指蛋白质以调控培养细胞的内源基因或在培养细胞中产生所需的表型。进而，还可能使用核酸重组，例如插入转基因，例如，编码调控内源基因的人工锌指蛋白质的

转基因修饰细胞，例如，干细胞。修饰的干细胞可施用于对象。体外培养干细胞的方法描述于，例如，US Application 2002-0081724。一些实施方式中，干细胞被诱导在对象中分化和表达转基因。例如，干细胞可分化为肝，脂肪，或骨骼肌细胞。干细胞可衍生自产生所需组织类型，例如，肝，脂肪，或骨骼肌细胞的细胞系。

另一实施方案中，表达或可以表达人工锌指蛋白质重组细胞，例如，如此处所述，可用于对象中的取代疗法。例如，编码可操纵的连接到启动子（例如，可诱导的启动子，例如，类固醇激素受体调控的启动子）的人工锌指蛋白质的核酸被导入人或非人，例如，哺乳动物，例如，猪重组细胞中。培养细胞并装入生物相容物质的胶囊中，如聚赖氨酸藻酸盐，及随后植入对象。见，例如，Lanza (1996) Nat.Biotechnol. 14: 1107; Joki 等 (2001) Nat.Biotechnol. 19: 35; 和 U. S. Patent No. 5,876,742。在人工锌指蛋白质调控编码分泌蛋白质的内源基因的实施方式中，对象中分泌多肽的产生可通过施用试剂（例如，类固醇激素）给对象而调控。

另一实施方案中，体外培养表达或能够表达人工锌指蛋白质的重组细胞。重组细胞产生的蛋白质可从细胞或从包围细胞的培养基中被回收（例如，纯化）。

改变蛋白质产生的靶

一个实施方案中，筛选核酸文库以鉴定改变一个或多个特定靶蛋白质的产生，合成或活性的人工锌指蛋白质。改变可提高或降低靶蛋白质的活性或丰度。筛选的表型可与一个或多种靶蛋白质的改变的生产或活性相关或是其自身生产或活性的水平。例如，可能筛选核酸文库中在内源靶基因的调控序列（例如，启动子）的控制下激活或抑制报道基因（如编码萤光素酶，LacZ，或 GFP 的基因）表达的人工转录因子。

一些示范靶蛋白质包括：细胞表面蛋白质（例如，糖基化表面蛋白质），癌相关蛋白质，细胞因子，化学因子，肽激素，神经递质，细胞表面受体（例如，细胞表面受体激酶，7 次跨膜受体，病毒受体和共同受体，胞外基质结合蛋白质，细胞结合蛋白质，病原（例如，细菌抗原，疟疾

抗原, 等等) 抗原)。另外的蛋白质靶包括酶如烯醇酶, 细胞色素 P450, 酰基转移酶, 甲基酶, TIM barrel 酶, 异构酶, 酰基转移酶等等。

更特异的实施例包括: 整联蛋白, 细胞粘附分子或“CAMs”如钙粘着蛋白, 选择素, N-CAM, E-CAM, U-CAM, I-CAM 等等; 蛋白酶(例如, 枯草菌素, 胰蛋白酶, 糜蛋白酶; 血浆酶原激活子, 如尿激酶或人类组织类型特异纤溶酶原激活剂); 蛙皮素; 因子 IX, 凝血酶; CD-4; 血小板衍生长因子; 胰岛素样生长因子-I 和 II; 神经生长因子; 纤维细胞生长因子(例如, aFGF 和 bFGF); 表皮生长因子(EGF); VEGFa; 转化生长因子(TGF, 例如, TGF- α 和 TGF- β); 胰岛素样生长因子结合蛋白质; 促红细胞生成素; 血小板生成素; 粘液素; 人血清白蛋白; 生长激素(例如, 人生长激素); 胰岛素原, 胰岛素 A 链胰岛素 B 链; 甲状旁腺激素; 促甲状旁腺素; 甲状旁腺素; 卵泡刺激素; 降钙素; 心钠肽 A, B 或 C; 黄体生成素; 胰高血糖素; 因子 VIII; 造血生长因子; 肿瘤坏死因子(例如, TNF- α 和 TNF- β); 脑啡肽酶; 繆氏抑制物; 促性腺激素相关肽; 组织因子蛋白质; 抑制素; 活化素; 血管内皮生长因子; 激素或生长因子受体; 类风湿因子; 骨诱导因子; 干扰素, 例如, 干扰素- α , β , γ ; 集落刺激因子(CSF), 例如, M-CSF, GM-CSF, 和 G-CSF; 细胞白介素(ILs) 例如, IL-1, IL-2, IL-3, IL-4 等; 促衰变因子; 和免疫球蛋白。一些实施方案中, 靶和疾病相关, 例如, 癌。

本发明在下述实施方式中将详细描述。然而, 应注意这些实施例不是限制本发明的范围。

实施例 1: 构建 ZFP 文库

一个实施方式中, 使用锌指蛋白质(ZFP)表达文库通过调控基因表达改变了酿酒酵母的多种表型。这些示范文库的锌指蛋白质组成为 3 或 4 个锌指结构域(ZFD)并分别识别 9 到 12 碱基对的 DNA 序列。不需要预先了解靶基因就可鉴定嵌合锌指蛋白质。从文库中 ZFP 产生了 3 类不同的转录因子: 分离的 ZFP, 当它们结合启动子区附近的位点时自身作

为有效的转录阻抑子；ZFP 也表达为融合到转录激活结构域或阻抑结构域的融合以分别产生转录激活子或阻抑子。

我们使用了 40 个不同的锌指结构域作为模块构建 3-指或 4-指锌指蛋白质。在其上限，3-指 ZFP 文库组成为 64, 000 ($=40^3$) 个序列和 4-指文库组成为二百六十万 ($=40^4$) 个序列。

这些 ZFP 表达质粒文库被转化进酵母细胞。每个转化细胞中，表达不同的 ZFP 转录因子并调控基因组中未明的靶基因。这个基因表达模式的改变导致表型变化。通过筛选大量转化细胞，可以分离具有所需表型的克隆。另外，在鉴定了导入转化体的锌指蛋白质后，可通过基因组范围的基因表达分布图分析（例如，通过 DNA 微阵列分析）或通过对靶 DNA 序列进行计算机（in silico）预测而鉴定被调控的靶基因。

（1）酵母株

用于这个实验的酿酒酵母株是 YPH499a (MATa, ade2-101, ura3-52, lys2-801, trp1-63, his3-200, leu2-1, GAL+). 酵母细胞的转化通过使用醋酸锂转化方法（见，例如，Gietz 等，(1992) Nucl. Acids Res 20: 1245）进行。

（2）构建质粒 p3

用于构建锌指蛋白质文库的亲代载体是质粒 p3。p3 是通过修饰 pcDNA3 载体 (Invitrogen, San Diego CA) 而构建的。pcDNA3 载体用 HindIII 和 XhoI 消化。将具有相一致的突出端的合成的寡核苷酸双链连接到消化的 pcDNA3。该双链含有编码红血球凝聚素 (HA) 标签和核定位信号的核酸。该双链还包括：BamHI, EcoRI, NotI, 和 BglII 的限制位点。载体的 SV40 复制起点的 XmaI 位点用 XmaI 消化破坏，填平消化的 XmaI 限制位点的末端，并再连接末端。

（3）构建质粒 pYCT-Lib

我们使用 pYCT-Lib 作为在酵母中条件表达锌指蛋白质的亲代载体。pYCT-Lib 是含有可诱导的 GAL1 启动子的酵母穿梭载体（图 1）。其它特征可包括：(i) 编码核定位信号 (NLS) 和红血球凝聚素标签 (HA) 的

序列和(ii)含有在合成的缺少色氨酸的最低培养基中选择含有质粒细胞的 TRP1 基因的序列。

T7 启动子位点到 SphI 位点区域的多位点接头可包括：
TAATACGACTCACTATAGGGAATATTAAGCTAAGCTCACCATGGGTA
AGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTACACAAGCTATG
GGTGCTCCTCCAAAAAAGAAGAGAAAGGTAGCTGGATCCACTAGT
AACGGCCGCCAGTGTGCTGGAATTCTGCAGATATCCATCACACTGG
CGGCGCTCGAGGCATGCATCTA (SEQ ID NO: 135)

如下构建 pYCT-Lib。酵母表达质粒 pYESTrp2 (Invitrogen, San Diego CA) 用 NgoM4 消化然后用 PstI 部分消化以从载体中除去 2μ ori 片段。NgoM4-PstI 消化的载体的 5.0kb DNA 片段凝胶电泳纯化并和扩增自 pRS313 (正向引物: 5'-CGATCTGCAGGG TCCTTTTCATCACGTGCT-3' (SEQ ID NO: 136), 反向引物: 5'-CGATCGATGCCGGCGGACGGATCGCTTGCCT (SEQ ID NO: 137)) 的 CEN-ARS 片段连接。

编码 B42 激活结构域的 DNA 片段通过用 NcoI 和 BamHI 消化除去并用编码 V5 表位标签和核定位信号的 DNA 片段取代。后一个 DNA 片段是 PCR 扩增自 pYESTrp2 (正向引物: 5'-AATTCCATGGGTAAGCCTATCCCTAACC-3' (SEQ ID NO: 138), 反向引物: 5'-AATTGGATCCAGCTACCTTTCTCTTCTT-3' (SEQ ID NO: 139)) 并连接入 NcoI 和 BamHI 位点。所得质粒命名为 pYCT-Lib (图 1)。

(4) 构建质粒 pYCT-Lib-Gal4

为产生 pYCT-Lib-Gal4, Gal4 激活结构域用 PCR 扩增自酵母基因组 DNA (正向引物: 5'-AAGGAAGGAAGGAAGCGGC CGCAGCCAATTTTAATCAAAGTGG-3' (SEQ ID NO: 140), 反向引物: 5'-ACATACATGCATGCGCCGTTACTAGTGGATCC-3' 序列 (SEQ ID NO: 141)) 并插入 pYCT-Lib NotI 和 SphI 识别位点之间以产生 pYCT-Lib-Gal4。编码 Gal4 激活结构域的示范序列和连接序列包括:

GGCCGCCAGTGTGCTGGAATTCTGCAGATATCCATCACACTGGCGG
CCGCAGCCAATTTTAATCAAAGTGGGAATATTGCTGATAGCTCATTG
TCCTTCACTTTTCACTAACAGTAGCAACGGTCCGAACCTCATAACAA
CTCAAACAAATTCTCAAGCGCTTTCACAACCAATTGCCTCCTCTAAC
GTTTCATGATAACTTCATGAATAATGAAATCACGGCTAGTAAAATTGA
TGATGGTAATAATTCAAACCACTGTCACCTGGTTGGACGGACCAA
ACTGCGTATAACGCGTTTGGAATCACTACAGGGATGTTTAATACCAC
TACAATGGATGATGTATATAACTATCTATTCGATGATGAAGATACCCC
ACCAAACCCAAAAAAGAGATCTCTATGGCTTACCCATACGATGTT
CCAGATTACGCTAGCTAAGGATCCACTAGTAACGGCGCATGCATCTA
GAGGGCC (SEQ ID NO: 142)

(5) 构建质粒 pYCT-Lib-Ume6

为产生 pYCT-Lib-Ume6, 编码酿酒酵母 Ume6 的 508 至 594 氨基酸的 DNA 片段扩增自酵母基因组 DNA (正向引物: 5'-AAGGAAGGAAGGAAGCGGCCGCAAATTCTGCATCTTCATCTACC-3' (SEQ ID NO: 143), 反向引物: 5'-ACATACATGCATGCTGTAGAATTGTTGCTTTTCG-3' (SEQ ID NO: 144)) 并插入 pYCT-Lib 的 NotI 和 SphI 识别位点之间。这个 87 个氨基酸的区域功能是作为转录阻抑结构域 (Kadosh 和 Struhl (1997) Cell 89: 365-371)。编码 Ume6 阻抑结构域和连接序列的示范序列包括:

GGCCGCCAGTGTGCTGGAATTCTGCAGATATCCATCACACTGGCGG
CCGCAAATTCTGCATCTTCATCTACCAAACCTAGACGACGACTTGGGT
ACAGCAGCAGCAGTGCTATCAAACATGAGATCATCCCCATATAGAA
CTCATGATAAACCCATTTCCAATGTCAATGACATGAATAACACAAAT
GCGCTCGGTGTGCCGGCTAGTAGGCCTCATTCGTCATCTTTTCCATC
AAAGGGTGTCTTAAGACCAATTCTGTTACGTATCCATAATTCCGAAC
AACAACCCATTTTCGAAAGCAACAATTCTACAGCATGCATCTAGAG
GGCC (SEQ ID NO: 145)

(6) 文库构建

3-指蛋白质文库 (“3-F 文库”), 编码具有 3 个 ZFD 阵列的锌指蛋白质, 它可从编码 40 个不同 ZFD 或“指”的核酸来构建。一个四指蛋白质文库 (“4-F 文库”) 从编码 27 个不同 ZFD 的核酸构建 (表 2, 如下)。

表 2: 构建 3-指或 4-指 ZFP 文库的锌指结构域

结构域	来源	靶位点	氨基酸序列	SEQ ID NO:	文库
DSAR	Mutated ¹	GTC	FMCTWSYCGKRFTDRSALARHKRTH	146	3F
DSHR	Mutated ¹	GGC	HICHIQCGKVVYGDRSHLTRHLRWH	147	3F
DSKR	Mutated ¹	GGT	FACPBCKPRFMDSSKLSRHKTH	148	3F
DSNR	Mutated ²	GAC	YACPVESCDRRFSDSSNLTRHRIH	149	3F
DSSR	Mutated ³	GCC	HICHIQCGKVVYGDRSSLTRHLRWH	150	3F
HSSK	Human	GAC	YKCKECGKAFNHSSNFNKHRIH	151	3F
HSSR	Human	GTT	FKCPVCGKAFRHSSSLVRHQTH	152	3F
ISNR	Human	GAA>GAT>GAC	YRCKYCDRSPSSNLQRHVRIH	153	3F
KSNR	Human	GAG	YGCILCGKAFSKSSNLRRHEMIH	154	3F
QAGR	Human	GGA	YKCKBCGQAFRQRAHLIRHHKLIH	155	3F
QFNR	Human	GAG	YKCHQCGKAFIQSFNLRRHERTH	156	3F
QGNR	Human	GAA	FQCNQCGASFTQKGNLLRHKLIH	157	3F
QNTQ	Drosophila ⁴	ATA	YTCGYCGKSFTQSNILKQHTRIH	158	3F
QSHR3	Human	GGA>AGA>GAA>CGA	YVCRBCQGRGFRQSHSLVRHKRTH	159	3F-
QSHV	Human	CGA>AGA>TGA	YBCDHCGKSFSQSSHLNVHCRTH	160	3F
QSN1	Human	AAA, CAA	YMCSECGRGFSQKSNLIHQTH	161	3F
QSNK	Human	GAA>TAA>AAA	YKCEBCGKAFQSSNLTKHKLIH	162	3F
QSTR	Human	GTA>GCA	YKCEBCGKAFNQSSTLTRHKIVH	163	3F
QTHR1	Human	GGA>AGA, GAA>TGA, CGA	YBCHDCGKSFRQSTHLTRHRIH	164	3F
RDHR1	Human	GAG, GGG	FLCQYCAQRFGKDHILTRHMKKSH	165	3F
RDKR	Human	GGG>AGG	YVCDVEGCTWKFARSDKLNRRHKRTH	166	3F
RDNQ	Mutated ⁴	AAG	FACPECPKRFMRSDNLTRHRIH	167	3F
SADR	Mutated ⁵	AGA	FQCRICMRNFSSPADLTRHIRT	168	3F
SSNR	Human	GAG>GAC	YECCECGKAFSSGNSNTRHRIH	169	3F
TIDR	Mutated ⁵	ACT	FQCRICMRNFSTHIDLIRHIRT	170	3F
VSNV	Human	AAT>CAT>TAT	YBCDHCGKAFSVSSNLNVHRIH	171	3F
VSTR	Human	GCT>GCG	YBCNYCGKTPSVSSILRHQRIH	172	3F
CSNR1	Human	GAA>GAC>GAG	YKCKQCGKAFGCPNLRHGRTH	173	3F-, 4F-
DGNV	Mutated ⁵	AAC	FQCRICMRNFSDGNLVRHIRT	174	3F-, 4F-
QSHR3	Human	GGA>GAA	YACHLCGKAFQCSHLRRHEKTH	175	3F-
QSHR	Human	AGA, CGA>TGA>GGA	YKCEBCGKAFRQSSHLTTHKLIH	176	3F-, 4F-
QSNR1	Human	GAA	FECKDCGKAFIQKSNLIRHQTH	177	3F-, 4F-
QSNV2	Human	AAA, CAA	YVCSKCGKAFQSSNLTVHQKLIH	178	3F-, 4F-
QSSR1	Human	GTA>GCA	YKCPDCGKSFSQSSILRHQTH	179	3F-, 4F-
QTHQ	Human	CGA>TGA, AGA	YBCHDCGKSFRQSTHLTQHRRIH	180	3F-, 4F-
RDER1	Human	GCG>GTG, GAG	YVCDVEGCTWKFARSDENLRHKRTH	181	3F-, 4F-
RDHT	Human	TGG, AGG, CCG, GGG	FQCKTCQKRFSDHLKTHTRTH	182	3F-, 4F-
RSHR	Human	GGG	YKCMBCGKAFNRKSHLTRHQRIH	183	3F-, 4F-
RSNR	Human	GAG>GTG	YICRKCGRGFSRKENLIRHQTH	184	3F-, 4F-
VDYK	Drosophila ⁷	TAT, GAT	FHCYCEKSFVKDYLTIRKTH	185	3F-, 4F-
VSSR	Human	GTT>GCT>GTG>GTA	YTCKQCGKAFSVSSSLRRHETTH	186	3F-, 4F-
DGAR	Mutated ³	GTC	FQCRICMRNFSDPGALVRHIRT	187	4F-
DGHR	Mutated ³	GGC	FQCRICMRNFSDPGHLVRHIRT	188	4F-
DGNR	Mutated ³	GAC	FQCRICMRNFSDPGNLKRHIRT	189	4F-
DRDR	Mutated ³	GCC	FQCRICMRNFSDCDLARHIRT	190	4F-
MHHE	Mutated ⁸	TGT	YACPVESCDRRFMSHHLKEHIRT	191	4F-
QASA	Mutated ⁸	ATA	FQCRICMRNFSSQASLNAHIRT	192	4F-
QGDR	Mutated ³	GCA, GCC	FQCRICMRNFSSQSDLRHIRT	193	4F-
QSDR	Mutated ⁸	GCT	FQCRICMRNFSSQSDLRHIRT	194	4F-
QGTR	Mutated ⁸	ACA	FQCRICMRNFSSQGTLTRHIRT	195	4F-
RDIN	Mutated ³	AAG	FQCRICMRNFSSDTLSNHIRT	196	4F-
TDKR	Mutated ³	GGG, GGT	FQCRICMRNFSTADKLSRHIRT	197	4F-
TGNR	Mutated ³	GAT>GAA	FQCRICMRNFSTGNLVRHIRT	198	4F-
TIDR	Mutated ⁵	ACT	FQCRICMRNFSTHIDLIRHIRT	199	4F-

表 2 第 2 栏的上标指 1) Zhang 等, (2000) J.Biol.Chem. 275: 33850-33860; 2) Rebar 和 Pabo (1994) Science 263: 671-673; 3) Segal (1999) Proc.Natl. Acad. Sci. USA 96: 2758; 4) Gogus 等, (1996) Proc. Natl. Acad. Sci. USA. 93: 2159-2164; 5) Drier 等, (2001) J. Biol.Chem. 276: 29466-29478; 6) Liu 等 (2001) J.Biol.Chem. 276 (14): 11323-11334; 7) Hsu 等, (1992) Science 257: 1946-50。

图 7 描绘了构建多种 3 指文库的方法。编码每个 ZFD 的核酸片段被各个克隆到 p3 载体中以形成“单指”载体。等量的每个“单指”载体组合形成池。每份等量的池用 AgeI 和 XhoI 消化以获得消化的载体片段。这些载体片段用磷酸酶处理 30 分钟。另一等份的池用 XmaI 和 XhoI 消化以获得编码单个指的片段。得自 AgeI 和 XhoI 消化的池的消化载体核酸与通过 XmaI 和 XhoI 消化从载体释放的核酸片段连接。连接产生各编码 2 个锌指结构域的载体。转化到大肠杆菌中后, 获得大约 1.4×10^4 个独立的转化体, 因而形成 2-指文库。2-指文库的插入区域大小通过 PCR 分析 40 个菌落证实。正确大小的插入存在于 95% 的文库成员中。

为制备 3-指文库, 编码一个指的 DNA 片段插入编码 2 个指的质粒中。2-指文库用 AgeI 和 XhoI 消化。消化的质粒, 其保留了编码 2 个锌指结构域的核酸序列, 与编码单个指 (如上通过用 XmaI 和 XhoI 消化制备) 的核酸片段池连接。这个连接的产物转化至大肠杆菌中以获得大约 2.4×10^5 个独立的转化体。插入区域的验证证实了文库成员主要包括编码 3 个锌指结构域的序列。

为制备 4-指文库, 编码 2 个指的 DNA 片段插入编码 2 个指的质粒中。2-指文库用 XmaI 和 XhoI 消化以获得编码 2 个锌指结构域的核酸片段。2-指文库也用 AgeI 和 XhoI 消化以获得消化的质粒池。消化的质粒, 其保留了编码 2 个锌指结构域的核酸序列, 与编码 2 个锌指结构域的核酸片段连接以产生大量编码不同 4 指蛋白质组合的质粒。这个连接的产物转化入大肠杆菌并产出大约 7×10^6 个独立的转化体。

(7) 构建酵母表达文库

3-指 (3-F) 和 4-指 (4-F) 文库亚克隆至 pYCT-Lib, pYCT-Gal4, 和 pYCT-Ume6 的 EcoRI 和 NotI 位点。这些亚克隆过程产生了 6 个编码具有和没有转录调控结构域的 3 和 4 指 ZFP 的不同的文库。在大肠杆菌中扩增后, 每个文库使用醋酸锂转化入酵母株 YPH499a 中。转化产出了大约 1.5×10^7 个菌落。文库插入区域的大小通过 PCR 分析 50 个菌落证实。95% 的文库成员包含正确的插入大小。转化体重悬于 TE 缓冲液中并储存于 -80°C 的甘油中。

实施例 2: 半乳糖培养基中的生长缺陷转化体

筛选 3-F 文库以鉴定损害酵母细胞生长的嵌合锌指蛋白质。这个筛选使用 GAL 启动子去条件表达嵌合锌指蛋白质。以前的研究使用了对酵母 cDNA 和基因组 DNA 序列的 GAL 启动子去鉴定致命过表达的基因 (Liu 等, (1992) Genetics 132: 665-673; Ramer 等 (1992) Proc.Natl. Acad. Sci. USA 89: 11589-11593; Espinet 等, (1995) Yeast 11: 25-32; Akada 等, (1997) Mol.Gen.Genet 254: 267-274; Stevenson 等, (2001) Proc. Natl. Acad. Sci. USA. 98: 3946- 3951)。

酵母株 YPH499a 用 3-F 文库的核酸转化。转化体在含有缺少色氨酸和含有葡萄糖的合成基本培养基平板上在 30°C 生长 2 天。每个这些葡萄糖平板 replica-plated 到半乳糖平板和第二个葡萄糖平板上。复制的平板在 30°C 生长过夜。鉴定不在半乳糖上生长但是生长在葡萄糖上的菌落。这些菌落从葡萄糖平板上回收并通过在半乳糖培养基 streaking 选择的菌落再验证。质粒转化入 YPH499a 以证实生长缺陷是锌指蛋白质过表达引起的。载体质粒 pYTC 和 pYTC-Gal4 都不编码任何锌指结构域, 作为对照分析。

如表 3 所示, 3-F 和 4-F 文库的 0.7%到 2.8%的转化体不能生长在半乳糖培养基中。这些百分比明显大于用 pYTC-Lib 载体对照 (0.1%) 和

pYTC-Lib-Gal4 载体对照 (0.2%) 获得的相似数值。

表 3: 半乳糖培养基上生长缺陷转化体的比率

文库	葡萄糖培养基上的集落数 (A)	半乳糖培养基上的生长缺陷集落数 (B)	生长缺陷突变体比例 (B/A×100)
3- 指	5,820	42	0.7
3- 指 + Gal4AD	7,428	206	2.8
3- 指 + Ume6RD	8,400	78	0.93
PYTC-Lib 载体	2,690	3	0.1
PYTC-Lib-Gal4 载体	2,750	6	0.2

从在是半乳糖生长缺陷的 10 个菌落中回收质粒。质粒 (L1 到 L10) 再转化入酵母细胞。再测试所有 10 个质粒。用回收的质粒转化的细胞不能在半乳糖培养基生长但能在葡萄糖培养基上生长。这 10 个质粒编码的锌指蛋白质通过 DNA 测序定性 (表 4)。这些蛋白质的潜在靶 DNA 结合位点从关于组成锌指结构域的结合特异性的信息推断。

表 4: 分离自生长缺陷转化体的质粒编码的 ZFP

No	ZFD名称 (N至C)			潜在靶序列
	1	2	3	
L1	RSHR	CSNR1	RDHT	5'-NGG GAV GGG-3' (SEQ ID NO:200)
L2	RSNR	RDHT	TDKR	5'-GGK NGG GAG-3' (SEQ ID NO:201)
L3	RDHT	QSHR3	RDHT	5'-NGG GRA NTT-3' (SEQ ID NO:202)
L4	QSDR	RDHT	RSNR	5'-GAANGG GCT-3' (SEQ ID NO:203)
L5	CSNR1	RDTN	VSSR	5'-GTD AAG GAV-3' (SEQ ID NO:204)
L6	RDHT	RDHT	RDER1	5'-GHG NGG KGG-3' (SEQ ID NO:205)
L7	QSNR1	RDTN	QTHQ	5'-HGA AAG GAA-3' (SEQ ID NO:206)
L8	TGNR	RDER1	RDHT	5'-NGG GHG GAW-3' (SEQ ID NO:207)
L9	QSSR1	RDHT	QTHQ	5'-HGA NGG GYA-3' (SEQ ID NO:208)
L10	QSSR1	RDHT	QSNR1	5'-GAA NGG GYA-3' (SEQ ID NO:209)

实施例 3：抗真菌药物抗性

酮康唑是一种口服吸收的咪唑类抗真菌药。它可以用来处理某些粘液。酮康唑阻断了酵母和其它真菌的麦角固醇的生物合成（Burden 等，(1989) *Phytochemistry* 28: 1791-1804）并对细胞代谢有额外作用（Kelly 等，(1992) In Fernandes, P. B. (Ed.) *New Approaches for Antifungal Drug*, Birkhäuser, Boston, pp.155-187）。

为检查测试株 PRH499a 对酮康唑的抑制真菌的反应，将 YPH499a 株的 10^7 种细胞铺在含有不同药物浓度的合成培养基的板上。我们发现 35 μ M 酮康唑抑制了 YPH499a 细胞的生长并用这个浓度筛选酮康唑-抗性酵母菌落。

1×10^7 个含有 3-指和 4-指文库质粒的酵母细胞在含有 2% 半乳糖的合成液体培养基中 30 $^{\circ}$ C 培养 3 小时以诱导锌指蛋白质表达，然后铺在含有 35 μ M 酮康唑（ICN Biomedicals）的合成半乳糖琼脂板上。30 $^{\circ}$ C 孵育 4 天后，大约 120 个克隆在含有 35 μ M 酮康唑的半乳糖培养基上形成菌落。挑出这些酮康唑抗性酵母菌落并划线接种在含有 35 μ M 酮康唑的新鲜合成的半乳糖琼脂板上。从 120 个抗性克隆中随机选择 23 个克隆。对 23 个克隆中的每一个，通过质粒拯救鉴定抗性表型。分离质粒，转化入大肠杆菌中扩增，并再转化入酵母株 YPH499a（Ausubel 等 (Eds) (1995) *Current Protocols in Molecular Biology* John Wiley and Sons Ltd, New York）。等数量的再转化体铺在含有或没有 35 μ M 酮康唑的合成半乳糖琼脂板上。再转化体也铺在含有或没有酮康唑的合成葡萄糖琼脂板上以证实药物抗性是由锌指蛋白质的半乳糖诱导的表达诱导的。

一个实施方式中，每个转化体的 5×10^4 种细胞被连续稀释（ 10^{-1} ， 10^{-2} ，和 10^{-3} 倍）并铺在添加或未添加 35 μ M 酮康唑的半乳糖或葡萄糖培养基中。30 $^{\circ}$ C，4 天后监测细胞的生长并与同一板的对照比较。对照包括编码不产生酮康唑抗性的锌指蛋白质的质粒和没有编码锌指蛋白质插入序列

的 pYTC-Lib 质粒。所有 23 个情况下，测试回收的质粒。另外，酮康唑抗性只在细胞铺在半乳糖培养基上的时候观察到，证实锌指蛋白质的表达产生了酮康唑抗性。

分离自酮康唑抗性转化体的质粒测试并预测它们在酵母基因组中的预期靶序列（表 5）。鉴定了 11 个独特克隆（表 5）。

表 5：产生酮康唑抗性的 ZFP

No.	ZFD名称 (N至C)				功能性结构域	潜在靶序列	分离物数量 ⁽¹⁾
	#1	#2	#3	#4			
K1	QSHV	QFNR	RSHR	-	UME6	5'-GGG GAG HGA-3' (SEQ ID NO:210)	3
K2	RSNR	RSNR	QSSR1	QSHT	UME6	5'-HGA GYA GAG GAG-3' (SEQ ID NO:211)	1
K3	RSNR	RSNR	QGTR	QSHR5	UME6	5'-GRA ACA GAG GAG-3' (SEQ ID NO:212)	2
K4	RSNR	RSNR	QGTR	QTHQ	UME6	5'-HGA ACA GAG GAG-3' (SEQ ID NO:213)	1
K5	VSSR	DGNV	VSSR	VDYK	GAL4	5'-KAT GTD AAC GTD-3' (SEQ ID NO:214)	2
K6	MHHE	QSNR1	VSSR	QGDR	GAL4	5'-GCA GTD GAA TGT-3' (SEQ ID NO:215)	3
K7	DGNV	QSHT	QSSR1	DGHR	GAL4	5'-GGC GYA HGA AAC-3' (SEQ ID NO:216)	3
K8	DGAR	RDTN	QTHQ	RDTN	-	5'-AAG HGA AAG GTG-3' (SEQ ID NO:217)	1
K9	RDHT	QTHQ	QSHT	DGNV	-	5'-AAC HGA HGA NGG-3' (SEQ ID NO:218)	1
K10	RDHT	QTHQ	QSHT	-	-	5'-HGA HGA NGG-3' (SEQ ID NO:219)	5
K11	RDHT	QSHV	QSHV	-	-	5'-HGA HGA NGG-3' (SEQ ID NO:220)	1

以下列出了这些蛋白质的氨基酸序列（锌指结构域有下划线和转录调控结构域是黑体）。K1: QSHV-QFNR- RSHR-Ume6 包括以下氨基酸序列:

MGKPIPNPLLGLNSTQAMGAPPKKKRKVGIRIPGEKPYECDHCGKSFSQSSHLNVHKRTHHTGEKP
YKCHQCGKAFIQSFNLRRHERTHHTGEKPYKCMCEGKAFNRRSHLTRHORI**HAAAANSASSSTKLD**
DDLGTAAAVLSNMRSSPYRTHDKPISNVNDMNTNALGVPASRPHSSSFPSKGVLRPILLRIENS
EQQPIFESNNSTACI (SEQ ID NO:221)

K2: RSNR-RSNR-QSSR1-QSHT-Ume6 包括以下氨基酸序列:

MGKPIPNPLLGLNSTQAMGAPPKKKRKVGIRIPGEKPYICRKCGRGFSRKSNLIRHORTHHTGEKP
 YICRKCGRGFSRKSNLIRHORTHHTGEKPYKCPDCGKSFSQSSSLIRHORTHHTGEKPYKCEECGKA
FROSHLTHKLIHAAAANSASSSTKLD**DDLGTAAAVLSNMRSSPYRTHDKPISNVNDMNTNAL**
GVPASRPHSSSFPSKGVLRPILLRIHNS**EQQPIFESNNSTACI** (SEQ ID NO:222)

K3: RSNR-RSNR-QGTR-QSHR5-Ume6 包括以下氨基酸序列:

MGKPIPNPLLGLNSTQAMGAPPKKKRKVGIRIPGEKPYICRKCGRGFSRKSNLIRHORTHHTGEKP
YICRKCGRGFSRKSNLIRHORTHHTGEKPFOCRICMRNFSQSGTLTRHIRTHHTGEKPYVCRECGRG
FROSHLVRHKRTHAAAANSASSSTKLD**DDLGTAAAVLSNMRSSPYRTHDKPISNVNDMNTNAL**
GVPASRPHSSSFPSKGVLRPILLRIHNS**EQQPIFESNNSTACI** (SEQ ID NO:223)

K4: RSNR-RSNR-QGTR-QTHQ-Ume6 包括以下氨基酸序列:

MGKPIPNPLLGLNSTQAMGAPPKKKRKVGIRIPGEKPYICRKCGRGFSRKSNLIRHORTHHTGEKP
YICRKCGRGFSRKSNLIRHORTHHTGEKPFOCRICMRNFSQSGTLTRHIRTHHTGEKPYECDCGKS
FRQSTHLTHRRIHAAAANSASSSTKLD**DDLGTAAAVLSNMRSSPYRTHDKPISNVNDMNTNAL**
GVPASRPHSSSFPSKGVLRPILLRIHNS**EQQPIFESNNSTACI** (SEQ ID NO:224)

K5: VSSR-DGNV-VSSR-VDYK-Gal4 包括以下氨基酸序列:

MGKPIPNPLLGLNSTQAMGAPPKKKRKVGIRIPGEKPYTCKQCGKAFFSVSSSLRRHETHTHTGEKP
FOCRICMRNFSDSGNLRVHIRTHHTGEKPYTCKQCGKAFFSVSSSLRRHETHTHTGEKPFHCGYCEKS
FSVKDYLTKIRTHAAAANFNQSGNIADSSLSFTFNSSNGENLITTOINSQALSQPIASSNVHDNFMNELLAS
KIDGNSKPLSPGWIDQIAYNAFGITTMFNITIMDDVYNLFDEDTPPNPKKELSMAYPYDVPDYAS (SEQ
 ID NO:225)

K6: MHHE-QSNR1-VSSR-QGDR-Gal4 包括以下氨基酸序列:

MGKPIPNPLLGLNSTQAMGAPPKKKRKVGIRIPGEKPYACPVESCDRRFMSHHLKEHIRTHTGE
 KPFECKDCGKAFTQKSNLIRHORTHHTGEKPYTCKQCGKAFFSVSSSLRRHETHTHTGEKPFOCRICM
RNFSQSGDLRRHIRTHAAAANFNQSGNIADSSLSFTFNSSNGENLITTOINSQALSQPIASSNVHDNFMNELLAS
ITASKIDGNSKPLSPGWIDQIAYNAFGITTMFNITIMDDVYNLFDEDTPPNPKKELSMAYPYDVPDYAS
 (SEQ ID NO:226)

K7: DGNV-QSHT-QSSRI-DGHR-Gal4 包括以下氨基酸序列:

MGKPIPNPLLGLNSTQAMGAPPKKRKRKVGIRIPGEKPFQCRICMRNFSDSGNLRVHIRTHTGEKP
YKCEECGKAFROSSHLTTHKIIHTGEKPYKCPDCGKSFSQSSSLIRHORTHHTGEKPFQCRICMRN
FSDPGHLVRHIRTHAAAANFNQGNIDSSLSFTFTSSNGPNLFTTQINSQALSQPIASSNVEDNFMNETTA
SKIDDGNSKFLSPGWIDQTAYNAFGITTGENTITMDVYNLFDEKTPPNPKKELSMAYFTWPDYAS (SEQ
 ID NO:227)

K8: DGAR-RDTN-QTHQ-RDTN 包括以下氨基酸序列:

MGKPIPNPLLGLNSTQAMGAPPKKRKRKVGIRIPGEKPFQCRICMRNFSDPGALVRHIRTHTGEKP
FQCRICMRNFSRSDTLNHNIRHTHTGEKPYECHDCGKSFRQSTHLTOHRRIRHTGEKPFQCRICMRN
FSRSDTLNHNIRTHAAAARGMHLEGRIM (SEQ ID NO:228)

K9: RDHT-QTHQ-QSHT-DGNV 包括以下氨基酸序列:

MGKPIPNPLLGLNSTQAMGAPPKKRKRKVGIRIPGEKPFQCKTCORKFSRSDHLKTHTRTHTGEKP
YECHDCGKSFRQSTHLTOHRRIRHTGEKPYKCEECGKAFROSSHLTTHKIIHTGEKPFQCRICMRN
FSDSGNLRVHIRTHAAAARGMHLEGRIM (SEQ ID NO:229)

K10: RDHT-QTHQ-QSHT 包括以下氨基酸序列:

MGKPIPNPLLGLNSTQAMGAPPKKRKRKVGIRIPGEKPFQCKTCORKFSRSDHLKTHTRTHTGEKP
YECHDCGKSFRQSTHLTOHRRIRHTGEKPYKCEECGKAFROSSHLTTHKIIHAAAARGMHLEGRIM
 (SEQ ID NO:230)

K11: RDHT-QSHV-QSHV 包括以下氨基酸序列:

MGKPIPNPLLGLNSTQAMGAPPKKRKRKVGIRIPGEKPFQCKTCORKFSRSDHLKTHTRTHTGEKP
YECDCGKSFSQSSHLNVHVKRTHHTGEKPYECDCGKSFSQSSHLNVHVKRTHAAAARGMHLEGRIM
 (SEQ ID NO:231)

如上述的任何蛋白质, 包含具有相同普通基序 (即, 4 个 DNA 接触残基) 和以相同连续顺序排列的 2, 3, 或多个锌指结构域的其他锌指蛋白质也可以让真菌细胞产生药物抗性。

一些克隆包括了相似的锌指结构域构型。K2, K3, 和 K4 的第一个和第二个指互同。这些克隆中, 4 个指中有 3 个与其它蛋白质相应位置的那些指相同。这些 ZFP 不太可能只是随机相关的 ($P < 1.6 \times 10^{-5}$) 因此, 在体内 K2, K3, 和 K4 ZFP 可结合相同靶位点并调控相同靶基因。

另外, K10 克隆编码的锌指蛋白质与 K9 和 K11 克隆编码的那些蛋白质密切相关。在 K9, K10, 和 K11 中发现的 QTHQ, QSHT, 和 QSHV 指可识别相同的 3-bp DNA 位点: 5'-HGA-3'。鉴于它们的结构相似性, K9, K10, 和 K11 ZFP 在体内可结合相同的靶位点并调控相同的靶基因。两组相关的克隆 (即, K2, K3, 和 K4; 和 K9, K10, 和 K11) 的每个中, 所有 ZFP 包括相同类型的调控特征。K2, K3, 和 K4 克隆每个包括 Ume6 阻抑结构域。K9, K10, 和 K11 克隆每个不具有专门的转录调控结构域的功能。

当 2 个或多个 ZFP 共转化入细胞时可以产生协同或加合作用。例如, 当 K4 合 K5 ZFP 共表达时, 酵母细胞具有酮康唑的完全抗性。组合产生了大约 1000 倍增强的表型。

构建了改变 DNA 接触残基或去除或取代调控结构域的 ZFP 突变体。一个突变体 (VSSR-DGAV-VSSR-VDYK-GAL4AD) 中, K5 的第二个锌指的天冬酰胺 DNA 接触残基突变为丙氨酸。凝胶迁移分析证明这个突变的 ZFP 对预期 DNA 位点的 DNA 结合亲和力至少降低了 10 倍。这个突变的 K5 蛋白质不使酵母细胞产生药物抗性。另一突变体中, K5 锌指蛋白质的 Gal4 激活结构域通过在编码激活结构域的 DNA 序列前插入终止密码子而删除。这个蛋白质不产生酮康唑抗性。对其它酮康唑抗性 ZFP 可获得相似结果。

当融合到 K5 ZFP 的激活结构域用 Ume6 阻抑结构域取代时, Ume6-形式蛋白质的表达逆转了酮康唑抗性表型。表达 Ume6-形式的细胞比对照细胞对酮康唑更敏感。这个结果显示选择转录因子可通过选择恶化表型的转录因子, 然后改变附上的调控结构域 (例如, 通过转换功能的方向性) 来产生具有所需表型作用的转录因子。筛选与所需表型相对的表型可能比筛选所需表型自身更容易。转换功能方向性的取代实例包括用另一种类型的调控结构域取代一种类型的调控结构域和去除调控结构域。K5 ZFP 的情况下, 通过筛选提高药物抗性的蛋白质并用转录阻抑结

构域取代其转录激活结构域而获得了提高药物敏感性的蛋白质。

表 6:

酮康唑浓度 (μM)	对照		K5		YLL053C	
	平均值	S. D.	平均值	S. D.	平均值	S. D.
0	100	0	100	0	100	0
5	82.7	4.1569	89.9	15.544	84.167	16.003
10	0.031	0.0044	90.6	5.8207	28.433	7.0002
15	0.0123	0.0021	82.467	7.7106	15.7	11.95
20	0.001	0	50.367	4.7014	2.4	0.1
25			4.2	3.1193	0.0233	0.0058
30			0.0433	0.0058	0.0033	0.0006
35			0.0167	0.0058	0.0013	0.0006

DNA 微阵列分析用来鉴定和药物抗性表型相关的基因。我们认为产生相同表型的 ZFP 可能调控那些表达直接或间接和表型相关的基因组。3 个 ZFP—K5, K6, 和 K7—被选择来做表达分布图分析。所有 3 个转录因子含有 Gal4 激活结构域。6400 个酵母开放阅读框架中, 10 个被至少 2 个不同的 ZFP 转录因子激活了超过 2 倍, 及 4 个开放阅读框架倍所有 3 个测试的 ZFP 转录因子激活。4 个激活的开放阅读框架是: YLL053C, YJR147W, YLL052C, 和 YPL091W。

表 7 列出了被超过一个嵌合 ZFP 调控的基因数。

K5	K6	K7	通常被调节 的基因数量
+	-	-	39
+	+	-	1
+	-	+	2
+	+	+	4
-	+	-	95
-	+	+	3
-	-	+	126

已知泵出酮康唑的基因 PDR5，通过 2 个 ZFP，即 K6 和 K7，而不是通过 K5 被激活。这个结果提示 K5 通过非 PDR5 依赖的机制产生酮康唑抗性而且至少 2 个不同的通路可以产生酵母的酮康唑抗性。一个通路依赖 PDR 的激活，和另一个不依赖 PDR5。

为鉴定和药物抗性表型相关的新基因，我们评价了过表达被所有测试的 ZFP 转录因子激活的 4 个基因中一个基因的细胞的药物抗性表型。我们发现基因之一--YLL053C—当自身过表达时诱导了酮康唑抗性。见表 6，上。YLL053C 与白色念珠菌的质膜和水通道蛋白质同源。YLL053C 的氨基酸序列如下：

```
MWFPQIIAGMAAGGAASAMTPGKVLFTNALGLGCSRSRGLFLEMFGTAVLCLTVLMTAVE
KRETNFMAALFIGISLFMAHMAALTGYTGTGVNPARSLGAAVAARYFFPHYHWIYWISPLLG
AFLAWSVWQLLQILDYTTYVNAEKAAGQKKED (SEQ ID NO:232)
```

来自白色念珠菌的示范同源通道蛋白质(AQY1)的氨基酸序列如下：

```
MVAESSIDNTPNDVEAQRPVYEPKYDDSVNVSPKKNHIAFLGEFFGTFFIWLWVAFVIA
QIANQDPTIPDKGSDPMQLIMISFGFGFGVMMGVFMFFRVSGGNLNPVTLTLVLAQAVP
PIRGLFMMVAQMIAGMAAAGAASAMTPGPPIAFTNGLGGGASKARGVFLEAFGTCILCLTV
LMAVEKSRATFMAPFVIGISLFLGHLICVYYTGAGLNPARSFGPCVAARSFPVYHWIYW
VGPILGSVIAFAIWKIFKILKYETCNPGQSDA (SEQ ID NO:233)
```

YLL053C 基因产物可以通过如 PDR5 基因产物似的泵出酮康唑而产生抗性。这些数据证明和表型改变相关的基因可通过分析细胞的基因表达分布图进行鉴定。基因鉴定还可进一步应用引起相同表型（例如，酮康唑抗性）的不同的嵌合 ZFP。

实施例 4：耐热转化体筛选

我们筛选了编码产生耐热酵母细胞的蛋白质的核酸的嵌合锌指文库。

含有 3-指文库核酸的 1×10^7 个酵母细胞在 30℃ 在含 2% 半乳糖的 SD 合成液体培养基培养 3 小时并在 52℃（空气温度）缓慢旋转孵育 2 小时。

热处理后，培养物铺在半乳糖培养基上并在 30℃ 孵育 5 天。生长的酵母菌落悬浮于半乳糖液体培养基中并在 52℃ 孵育 2 小时以证实耐热表型。从这些细胞中分离质粒并如上述再转化至酵母。转化体在 30℃ 在半乳糖液体培养基培养 3 小时并在 52℃ 孵育 2 小时。孵育的转化体以 4 倍稀释铺在 SD 半乳糖琼脂板上。为证实耐热性是突变体的锌指蛋白质的表达诱导的，再转化体培养在具有上述条件的 SD 葡萄糖液体培养基中并铺在 SD 葡萄糖琼脂板上。具有 pYTC 载体和编码随机选择的 3-指蛋白质的质粒的转化体用作对照。

耐热酵母细胞从转化了锌指蛋白质表达质粒的细胞中鉴定。携带 pYTC-Lib 质粒或编码随机选择的锌指蛋白质的质粒的野生型细胞用作阴性对照。总共 10^7 种细胞生长在半乳糖液体培养基中。半乳糖诱导锌指蛋白质表达。细胞在 52℃ 热处理 2 小时并铺在半乳糖基本琼脂板上。在 30℃ 孵育 5 天后，在板上生长了 26 个菌落。我们从这些菌落中回收质粒并再转化它们至 YPH499a 中。我们分离了诱导多种程度耐热性的 9 个克隆。典型的，当在半乳糖培养基而不是葡萄糖培养基中培养时，这些克隆表达的锌指蛋白质使达到 10% 的细胞在热处理后存活。只有大约 0.3% 的转化了对照质粒的细胞在相同条件下存活。新生长的再转化细胞在实验条件下比冻存细胞存活的好。这些结果证明特定锌指蛋白质的表达可以诱导酵母细胞的耐热性。通过 DNA 测序回收的文库质粒定性 ZFP（表 8）。

表 8：耐热转化体分离的质粒编码的 ZFP

No.	ZFD基序的名称（N至C）			功能性 结构域	潜在的靶序列
	1	2	3		
H1	ISNR	QSNi	RDNQ	UME6	5'-AAG MAA GAH-3'
H2	QNTQ	QNTQ	HSNK	UME6	5'-GAC ATA ATA-3'
H3	QSHR5	DSHR	DSKR	GAL4	5'-GGT GGC GRA-3'
H4	RDKR	QSTR	QSHR5	GAL4	5'-GRA GYA RGG-3'
H5	RDHR1	QSSR1	QSHR5	UME6	5'-GRA GYA GRG-3'
H6	QAHR	RSHR	RSHR	UME6	5'-GGG GGG GGA-3'
H7	DSNR	RDHT	QNTQ	-	5'-ATA NGG GAC-3'
H8	ISNR	RSNR	RSNR	-	5'-GAG GAG GAH-3'
H9	RDHT	QSNK	QSTR	-	5'-GYA DAA NGG-3'

实施例 5：轴突形成

我们从我们的核酸文库筛选了编码在小鼠神经瘤母细胞系 Neuro2A，一个可以分化为神经元细胞的细胞系中诱导轴突发生的锌指蛋白质的核酸。我们鉴定了一个诱导神经元分化的嵌合 ZFP，可通过细胞形态变化和表达神经元标记基因证实。

材料和方法

文库构建

见实施例 1，见上。如述，3-指和 4-指 ZFP-Tf 文库分别使用 40 和 25 个锌指结构域构建。3-指和 4-指文库分别识别大约 9-和 12-bp DNA 结合位点。ZFP 表达为与 p65 转录激活结构域和 KRAB 阻抑结构域的融合蛋白质。

Neuro2A 细胞的细胞培养和神经元分化

小鼠神经瘤 Neuro2A 细胞在 37℃，含有 95%空气和 5%CO₂潮湿气体中，维持在具有 10%FBS 和抗生素的 MEM-a 培养基中。细胞在每个

96-孔培养板接种 8.0×10^3 种细胞并使用 LIPOFECTAMINE PLUS™ (Invitrogen, CA) 试剂, 按照厂商方法, 用 50ng ZFP, 和 20ng LacZ 报道质粒转染。用或不用维甲酸 (RA) ($10 \mu\text{M}$) 分别进行体外分化, 转化后并用 G418 (1mg/ml) 处理 24 小时以减少未转染细胞数。细胞再培养 96 小时并固定, 对 β -半乳糖苷酶染色, 并照相。 β -半乳糖苷酶阳性细胞中, 如果轴突伸展长度至少是胞体直径的 2 倍时认为细胞发生分化。

结果和讨论

筛选诱导轴突发生的 ZFP-TFs

为筛选可以诱导神经元分化的 ZFP-TFs, Neuro2A 细胞用文库质粒和含有 LacZ 基因的报道质粒瞬时转染。LacZ 表达用来目测转染细胞的形态。因为分化细胞比非分化生长慢因而后者将在细胞数上占优势, 转染后我们用 G418 处理细胞培养 24 小时以减少未转染的细胞数。5 天后, 细胞固定并 LacZ 染色。我们然后通过形态特征定性细胞。特别的, 我们鉴定了具有增加了轴突长度和厚度的细胞。

我们鉴定了几个改变轴突发生的 ZFP-TFs。这些 ZFP-TFs 以不同程度影响轴突发生。ZFP-TFs, 命名为 Neuro1-p65, 通过测量其产生的轴突的长度和厚度, 其对分化有最明显的作用。当细胞用 $10 \mu\text{M}$ RA (维甲酸) 处理时, 在 Neuro1-p65 转染的 Neuro2A 细胞中观察到了明显的轴突发生加速。

编码 Neuro1-p65 和其氨基酸序列的核酸序列示于图 14。Neuro1-p65 核酸编码含有锌指结构域 QSNRI-QSNK-CSNRI 的嵌合 ZFP。其它具有相同基序 ZFP 和/或与 Neuro1-p65 结合位点至少部分重叠位点结合的 ZFP 也预期调节分化。基于锌指 DNA 结合位点目录 (见表 6), Neuro1 的预测结合位点是: 5'-GACGAAGAA-3'。

Neuro1-p65 需要 p65 激活结构域以诱导轴突形成。融合到 KRAB 转录阻抑结构域或没有任何效应子结构域的相同的锌指结构域不引起轴突发生。见图 8 和表 9。而且 Neuro1-p65 的 DNA 结合能力是关键, 因为预

计使 Neuro1-p65 的锌指结构域丧失 DNA 结合能力的突变 (Neuro1-p65mut) 消除了它支持轴突发生的能力。

实时PCR用来定性在细胞Neuro1-p65表达过程中神经元标记基因的表达水平。相似的,核酸微阵列用来比较RA处理的细胞和ZFP-TF处理的细胞间在分化的不同时间点的基因表达模式。Neuro1-p65可以激活神经瘤母细胞至少体外分化所需的通路。

表 9: 具有轴突细胞的百分比

构建体	无RA	RA 10uM
pCDNA3 (载体)	5.77 ± 1.99	14.93 ± 7.69
08_D1-p65	42.39 ± 5.56	41.59 ± 6.46
08_D1mut-p65	5.25 ± 0.71	NA
08_D1 alone	5.85 ± 4.44	NA
08_D1-KRAB	4.46 ± 0.57	NA

实施例 6: 骨发生

C2C12 细胞系衍生自成肌细胞系,但是在加入骨形态形成蛋白质-2 (BMP-2) 时可分化为成骨细胞 (Katagiri, T 等, (1994)J. Cell. Biol. 127, 1755)。几个天然产生的转录因子被鉴定为控制这个过程的候选因子 (Lee, K.-S.等, (2000) Mol. Cell. Biol. 20: 8783; Nakashima, K.等, (2002) Cell 108: 17)。

我们筛选了在不存在BMP-2时诱导C2C12成肌细胞转分化为成骨细胞的嵌合锌指蛋白质。我们瞬时转染了编码嵌合锌指蛋白质的核酸文库到C2C12细胞中并筛选了在不存在BMP-2时进行转分化的细胞。转染7天后,转化的细胞对碱性磷酸酶(ALP)染色,其是成骨细胞的标记 (Katagiri, T.等, (1994)J. Cell. Biol. 127: 1755)。从大约2000个ZFP-TFs的筛选中,我们鉴定了一个激活子, Osteo1-p65, 其在大约30%的细胞中诱导强ALP染色。ALP染色的细胞百分比与LacZ染色探出的转染细胞的百分比相似。从这些结果,我们推断Osteo1-p65转录因子可以在不存

在 BMP-2 时引起 C2C12 成肌细胞转分化为成骨细胞。

作为正对照, 用 1 μ g/ml 的 BMP-2 处理细胞。几乎 100% 的正对照细胞在 BMP-2 处理时对 ALP 强染色。没有 BMP-2 处理, 转染了对照载体的阴性对照细胞只有背景染色。Osteo1-p65 是组成为 RDKR-QTHR1-VSTR-RDKR 锌指结构域(从 N-到 C-末端)的 4-指蛋白质。见图 15。这个蛋白质预计识别 DNA 元件 5'-GGGGCWRGAGGG-3' (SEQ ID NO: 234)。

对于这些实验, 小鼠成肌细胞系, C2C12, 在 37 $^{\circ}$ C, 含有 95% 空气和 5%CO₂ 潮湿气体中, 维持在含有 4.5g/L, 10%FBS, 和抗生素的 DMEM 培养基中。细胞以每个 96-孔板 1.0×10^4 种细胞接种。使用 LIPOFECTAMINE PLUSTM (Invitrogen), 按照厂商方法, 用 50ng 文库核酸转染细胞。转染 24 小时后, 用含 2%FBS 的 DMEM 代替生长培养基, 细胞再培养 6 天。

为检查 C2C12 细胞分化为成骨细胞, ALP 染色如 Katagiri, T.等, (1994) J. Cell. Biol. 127: 1755 描述的进行。

简言之(指实施例 5 和 6), 我们对人工转录因子文库筛选了可以诱导 2 个不同细胞分化过程, 神经发生和骨发生的蛋白质。再者, 该方法不要求特定干细胞的生物学知识。

实施例 7: 胰岛素调控

我们稳定转染了编码嵌合锌指蛋白质的核酸到人 293 细胞系, 并使用 DNA 微阵列实验分析了转化的细胞以定性每个嵌合锌指蛋白质调控的基因(见图 4, 5 和 9)。我们鉴定了提高人胰岛素基因表达超过 60 倍的嵌合锌指蛋白质(08D04-p65)。

为测试 08_D04-p65 是否可以在不同的人细胞系中诱导胰岛素基因表达, 我们瞬时转染了编码 08_D04-p65 的核酸到 HeLa 细胞中。HeLa 细胞中核酸的表达引起了胰岛素基因表达提高 80 倍。

因此, 08_D04-p65, 其衍生物, 和相似功能的锌指蛋白质可以用作糖尿病的治疗物。编码 08_D04-p65 的 DNA 或相似功能的锌指蛋白质可以通过病毒输送或在胶囊(例如, 脂质体)中运送给糖尿病人。一旦 DNA 输送进细胞, 可表达锌指蛋白质以诱导产生胰岛素。一些实施方式中, 编码锌指蛋白质的核酸可操纵的连接到一个可诱导的启动子上, 例如, Tet 诱导的启动子。使用强力霉素作为诱导剂可使胰岛素产生的水平被小的化学物调控。因为胰岛素诱导的锌指蛋白质, 如 08_D04-p65, 可以在不同的人细胞系中有功能, 它可以在胰腺细胞(例如, beta 细胞或非 beta 细胞)和非胰腺细胞中起作用。

还可能鉴定诱导胰岛素加工酶表达的人工转录因子。工程化细胞以表达这样的转录因子和诱导胰岛素基因的转录因子, 例如, 08_D04-p65。另外, 胰岛素诱导锌指蛋白质还可用来进行离体细胞治疗。一个实施方式中, 病人的细胞通过导入编码胰岛素诱导锌指蛋白质的核酸被体外修饰。被修饰的细胞然后被植回病人或另一个对象中。

材料和方法

DNA 微阵列分析表达 ZFP-TFs 的细胞系

编码锌指蛋白质的核酸如厂商方法描述被稳定导入 FlpTRex-293 细胞系 (Invitrogen)。暂时的, pLFD-p65 或 pLFD-Kid 载体的 HindIII-XhoI 片段被亚克隆至 pCDNA5/FRT/TO (Invitrogen CA)。这个片段包含编码锌指蛋白质的序列。所得质粒和 pOG44 (Invitrogen) 共转染至 FlpTRex-293 细胞中。获得了在强力霉素诱导时表达 ZFP-TFs 的稳定整合体。

含有 7458 个人 EST 克隆的 DNA 微阵列得自 Genomictree (Korea)。稳定表达 ZFP-TFs 的 FlpTRex-293 细胞在有(+Dox)或没有(-Dox) 1 μ g/ml 强力霉素下生长 48 小时。从每个样品中制备总 RNA。-Dox 样品的 RNA 用作参照 (Cy3), 和+Dox 的用作实验样品 (Cy5)。微阵列分析按照厂商的方法进行。中间数之和小于 500 的点标为红色, 不分析。

编码某种锌指蛋白质的核酸也在 HeLa 细胞中表达。含有 08_D04 ZFP

的 pLFD-p65 使用 LIPOFECTAMINE™ 2000(Invitrogen)瞬时转染入 HeLa 细胞中。pLFD-p65 载体, 其不编码锌指蛋白质, 也平行转染入细胞中作为对照。

实施例 8

我们筛选了编码嵌合锌指的文库以鉴定编码改变哺乳动物细胞中报道基因表达的蛋白质的核酸。

构建文库

2 个 ZFP 文库如 4 指蛋白质和 3 指蛋白质构建在 P3, 修饰的 pcDNA3 载体中。这些文库的构建见上述。

转染 ZFP 质粒

人胚胎肾 293 细胞在含有 95%空气和 5%CO₂ 潮湿气体中, 用 100μl 含有 10%FBS 的 DMEM 培养, 转染前 1 天接种在 96 孔板上。分别从每个文库的各个菌落中分离质粒并使用 LIPOFECTAMINE PLUS™ (Invitrogen)如厂商建议转染进 293 细胞中。为进行 MTT 分析, 50ng 含有融合到功能性结构域的锌指蛋白质基因的质粒被转染进以 5×10^3 种细胞/孔培养的细胞中。为进行 SEAP 分析, 10ng Pseap2-对照 (Clontech) 质粒和 50ng 含有融合到功能性结构域的锌指蛋白质基因的质粒被共转染进以 1×10^4 种细胞/孔培养的细胞中。转染的细胞在分析前在 37℃ 孵育 3 天。

SEAP 分析

转染后 3 天, 培养基转移至新管中并在 65℃ 加热 30 分钟以失活内源的碱性磷酸酶。25μl 培养基和等体积的 2×SEAP 分析缓冲液 (2M 二乙醇胺 (pH9.8) 10ml, 1M MgCl₂ 10μl, 1M 高精氨酸 200μl, 正氮苯基磷酸 44.52mg) 混合。37℃ 在指定的孵育时间, 在 405nm 测量吸收。

萤光素酶分析

对萤光素酶分析, 5ng CMV-萤光素酶和 pRL-SV40 质粒和 50ng ZFP 编码质粒共转染进接种在 96 孔板的孔中的 293 细胞中。孵育 3 天后, 收

获细胞并通过双萤光素酶报道分析系统（Dual Luciferase reporter assay system (Promega)）如厂商建议进行萤光素酶分析。

MTT 分析

MTT 分析用来测量细胞生长。代谢活性细胞还原的四唑 MTT 变为黄色。测量黄色的光吸收表示活细胞数。使用 MTT 分析试剂盒 (Trevigen) 进行 MTT 分析。在转染后 3 天进行 MTT 分析。简要的，10 μ l MTT 溶液直接加入培养基并在 CO₂ 孵箱中孵育 2 小时。然后，加入 100 μ l MTT 试剂盒中的去垢剂溶液并在使用 Power-Wave 340x (Bio-Tek Instrument) 测量 570nm 的吸收前在暗处孵育 2 小时。

为计数细胞，293 细胞在转染前以 1.5×10^4 种细胞/孔接种于 24 孔板。用 200ng ZFP 编码核酸或对照载体质粒转染 3 天后，用 PBS 洗细胞并悬浮于 30 μ l TE (10 mM Tris 1 mM EDTA, pH 8.0) 缓冲液中并用血球计数器计数。

结果

蛋白质生产性的变化

我们测试了是否文库核酸可以改变异源核酸的表达。SEAP（分泌性碱性磷酸酶）用作报道蛋白质以测试是否 ZFP 可以影响细胞内产生的外源蛋白质的量。表 10 显示了从 81 个孔得到的代表性结果，每个孔含有转染了编码文库随机的嵌合 ZFP 的核酸的细胞和 3 个孔含有转染了对照载体的参照细胞。几个孔展示了相对其它含有转染了编码其它随机 ZFP 的核酸的细胞的孔和相对含有转染了对照载体的细胞的孔提高了的吸收。

表 10: 81 种随机 ZFP 存在下的报道分子活性。

ZFP	吸光度 (405nm)	ZFP	吸光度 (405nm)	ZFP	吸光度 (405nm)
A01	0.29	C07	0.12	F01	0.15
A02	0.14	C08	0.16	F02	0.13
A03	0.14	C09	0.18	F03	0.13
A04	0.13	C10	0.17	F04	0.12
A05	0.19	C11	0.18	F05	0.13
A06	0.13	C12	0.15	F06	0.12
A07	0.12	D01	0.17	F07	0.12
A08	0.13	D02	0.15	F08	0.14
A09	0.11	D03	0.14	F09	0.19
A10	0.15	D04	0.14	F10	0.12
A11	0.15	D05	0.12	F11	0.20
A12	0.12	D06	0.13	F12	0.12
B01	0.15	D07	0.21	G01	0.15
B02	0.78	D08	0.12	G02	0.16
B03	0.15	D09	0.14	G03	0.14
B04	0.19	D10	0.15	G04	0.49
B05	0.13	D11	0.20	G05	0.26
B06	0.15	D12	0.12	G06	0.15
B07	0.14	E01	0.16	G07	0.22
B08	2.40	E02	0.13	G08	0.65
B09	0.11	E03	0.12	G09	0.18
B10	0.12	E04	0.41	vector	0.37
B11	0.13	E05	0.12	vector	0.39
B12	0.11	E06	0.15	vector	0.34
C01	0.20	E07	0.12		
C02	0.14	E08	0.12		
C03	0.31	E09	0.13		
C04	0.13	E10	0.34		
C05	0.13	E11	0.22		
C06	0.13	E12	0.11		

P_B08, 融合到 p65 结构域的锌指蛋白质, 显示了最高的 SEAP 活性。这个 ZFP 通过在 3 个分别的转染中用 SEAP 质粒转染测试。表 11 说明了一个这样的再测试。这个分析中 P_B08 的 P 值是 0.008。P_B08 比亲代质粒的 SEAP 活性提高了大约 16 倍。

表 11: 293 细胞中 SV40-SEAP 报道分子

表达质粒	平均值	S. D.
P_B08	16.51	0.54
载体 (HANLS)	1	0.17

因为 SEAP 分析用于分析外源蛋白质的生产, 观察提高的 SEAP 活性的原因是不同的, 例如, 直接或间接激活用于 pSEAP2-对照质粒的 SEAP 基因表达的 SV40 启动子。其它原因包括: 在分泌通路中有功能的蛋白质提高了活性或激活蛋白质生产的一般机制。SEAP 提高活性的机制可通过确定 P_B08 对其它蛋白质生产的作用来阐明, 例如, 使用其它报道蛋白质 (例如, 非分泌的报道分子) 和连接于其它启动子的报道分子。

为研究这点, 在 293 细胞中使用 CMV 启动子 (CMV-SEAP) 控制的 SEAP 报道分子进行了广泛筛选。分离了 2 个 ZFP, K44_11_D01 和 K44_11_G12, 其提高了 SEAP 蛋白质的活性。这 2 个 ZFP 进一步使用 2 个其它报道质粒, SV40-SEAP 和 CMV 启动子控制的萤光素酶报道分子 (CMV-Luc) 进行测试。表 12 列出了一些结果。K44_11_D01 和 K44_11_G12 比亲代质粒 (载体) 的 SEAP 活性大约提高了 3 倍。

K44_11_D01 显示激活 SV40-SEAP, 但不能激活 CMV-萤光素酶, 提示激活可能与 SEAP 特异的机制相关, 如分泌。相反, K44_11_G12 显示激活 CMV-萤光素酶和 SV40-SEAP 大约到 2-3 倍, 提示机制对启动子或报道分子不特异。因此, K44_11_G12 可以用于提高细胞中一般蛋白质的生产, 例如, 真核细胞中。

表 12: 293F 细胞中的 CMV-SEAP, CMV-Luc 和 SV40-SEAP 报道分子。

表达质粒	CMV-SEAP	平均值 CMV-Luc	SV40-SEAP	CMV-SEAP	S. D. CMV-Luc	SV40-SEAP
K44_11_D01	3.49	0.87	3.66	0.46	0.26	0.28
K44_11_G12	2.82	2.75	2.14	0.50	0.38	0.15
载体 (HA-NLS)	1	1	1	0.22	0.19	0.17

锌指蛋白质 K44_16_E12 (图 25) 和 K12_A11 (图 24) 也可以改变蛋白质生产, 例如, 如 CMV-SEAP 和 SMV-Luc 表达所示。

作为类似的实施例, 周期蛋白 D1 的过表达据报导可以引起一些蛋白质生产提高 (US patent No. 6,210, 924 B1)。可以直接或间接激活周期蛋白 D1 表达的锌指蛋白质可以上调蛋白质生产, 尽管锌指蛋白质的作用可能不同, 范围更广, 或更有效力。造成与周期蛋白 D1 过表达相同生理作用的其它锌指蛋白质可通过这个分析分离。这样的锌指蛋白质可用于提高感兴趣的蛋白质的生产。

细胞生长速度的变化

2 个 ZFP 融合到 kid 结构域, K_D10 和 K_F02, 选择作为 ZFP 调控细胞生长的代表, 因为基于 MTT 分析结果, 与对照相比它们显示了最大的变化。为证实 MTT 分析, 293 细胞中转染了 K_D10 和 K_F02 后计数细胞。结果与用亲代载体转染孔的细胞数比较, 并转化为相对百分比并示于表 13。与亲代载体相比, 与对照比较, K_D10 对细胞生长显示了大约 3 倍的抑制而 K_F02 显示了 2.7 倍的激活。K_D10 和 K_F02 的 P 值分别是 0.001 和 0.01。关于序列信息, 见图 18 和 19。

表 13: 2 个 ZFP 的 MTT 分析结果。

	相对细胞数	
	平均值	S. D.
K_F02	279.3	6.9
K_D10	27.4	4.3
Vector	99.99	3.4

细胞生长速度的变化

我们鉴定了可以改变细胞生长和繁殖速度的锌指蛋白质。这些重要的表型与疾病如癌和病毒感染及与发育过程有关联。使用锌指蛋白质控制细胞生长和繁殖的能力包括, 例如, 控制凋亡, 细胞分化, 宿主细胞防御 (例如, 抗病毒感染), 和 p53 介导的信号传导。

产生一个或多个这些表型的锌指蛋白质可以直接或间接调控那些调控细胞生长或信号传导的基因。这里描述的特异的锌指蛋白质产生了对细胞生长可探索的作用如 MTT 分析。例如，2 个锌指蛋白质显示了与对照相比至少 2 倍的不同。鉴于这个分析是基于观察转染 3 天后瞬时转染的细胞。2 倍不同是显著的。另外，细胞数目的变化证实使用 MTT 分析发现的不同对这个分析不是特别的。

因为细胞生长是细胞生理学的一个重要参数，这个实验可扩展到大量筛选以分离足够数目的调控细胞生长的锌指蛋白质。cDNA 微阵列可以用来分析表达每个锌指蛋白质细胞的基因表达分布图。这个分布图可鉴定参与细胞繁殖或凋亡的新的基因和通路。

这些实验显示锌指蛋白质可以诱导哺乳动物细胞的不同表型，例如，在这个例子是改变细胞繁殖。

实施例 9：溶剂耐受的细菌细胞

我们筛选了表达人工锌指蛋白质的细菌细胞，细胞具有有机溶剂抗性是因为人工嵌合锌指蛋白质的结果。鉴定了 3 个不同的锌指蛋白质使大肠杆菌产生己烷耐受性的能力（表 13）。己烷耐受性通过比较表达锌指蛋白质—HT-1，HT-2，和 HT-3 的之一的转化体的存活率和对照细胞的存活率进行评价。对照细胞包括一个空载体（C1）或 ZFP-1。ZFP-1 构建体编码不产生己烷抗性和含有 RDER-QSSR-DSKR 的锌指蛋白质。表达产生己烷抗性的锌指蛋白质的细菌细胞显示提高了己烷耐受性有 200 倍。

表 13：己烷抗性锌指蛋白质。

表达构建体	名称	存活率
对照	C1	0.14
对照	ZFP-1	0.05
己烷抗性ZFP	HT-1	21.4
己烷抗性ZFP	HT-2	1.85
己烷抗性ZFP	HT-3	28.6

这些锌指蛋白质的表达质粒包括 IPTG-诱导启动子。表达每个产生己烷抗性的 ZFP 的转化体在存在或不存在 IPTG 下定性。表达 HT-2 的转化体在存在 IPTG 时显示了较高的己烷耐受性,而 HT-1 或 HT-3 表达细胞甚至在不存在 IPTG 时也是己烷耐受的。

表 14: 在大肠杆菌中产生己烷耐受性的锌指蛋白质。

名称	F1	F2	F3	F4	推定的DNA靶	出现的数量 (##)
HT-1	RSHR	HSSR	ISNR		GAH GTT GGG	5
HT-2	ISNR	RDHT	QTHR	VSTR	GCT GRA NGG GAH (SEQ ID NO:235)	3
HT-3	QNTQ	CSNR	ISNR		GAH GAV ATA	1

(##) 在第三轮己烷耐受性筛选后能够生长的 9 个集落中存在 ZFP。

构建编码大肠杆菌表达的 ZFP 的核酸文库。为在大肠杆菌中条件表达锌指蛋白质,我们将克隆在 pYTC-Lib 酵母载体中的锌指蛋白质亚克隆到 pZL1。pZL1 是通过修饰 pBT-LGF2 (Clontech, Palo Alto, CA) 而构建。Plac-UV5 通过 PCR 从 pBT-LGF2 中扩增(正向引物: 5'-GACA ACC GGT CAT CGA TAA GCT AAT TCT CAC-3' (SEQ ID NO: 236); 反向引物: 5'-TTG TCC ATG GAC GCT GTT TCC TG GTG AAA-3' (SEQ ID NO: 237))。PCR 产物克隆到 pYTC Lib 载体的 AgeI 和 NcoI 位点间。载体命名为 pYTC-lac。pYTC-lac 用 ClaI 和 NotI 消化以亚克隆含有以下元件的 DNA 片段: Plac 启动子-V5 表位-MCS。消化的 DNA 片段在凝胶电泳后纯化并亚克隆到 pBT-LGF2 的 ClaI 和 NotI 位点。所得载体命名为 pZL1。构建于 pYTC-Lib 的 3F-或 4F-文库用 EcoRI 和 NotI 消化。消化的 DNA 片段用凝胶纯化并亚克隆到 pZL1 的 EcoRI 和 NotI 位点,因而提供了大肠杆菌表达文库。

筛选溶剂耐受性。大肠杆菌株 DH5 α 用可以原核表达的 3-指或 4-指 ZFP 核酸文库转化。转化体在含氯霉素 (34 μ g/ml) LB 中培养过夜。过夜

培养物在 1ml 含有 1mM IPTG 和氯霉素的新鲜 LB 培养基中稀释至 1:500 以诱导 ZFP 表达。在 30℃ 孵育 3 小时后，加入己烷到 1.5% 并迅速振荡以形成己烷和大肠杆菌培养物的乳状液。混合物在 37℃ 振荡（250rpm）孵育 3 小时并铺在含有氯霉素（c）的 LB 板上。从生长菌落的池中纯化质粒并转化 DH5 α 。转化体用己烷如上述处理。选择己烷耐受性重复额外 2 次。第三轮选择后，从 20 个在含有氯霉素（34 μ g/ml）的 LB 板上生长的单个菌落中回收质粒。这些质粒再转化 DH5 α 。每个转化体再如上述测试己烷耐受性。诱导己烷耐受性的质粒倍测序以定性编码的锌指蛋白质。

实施例 10：耐热细菌细胞

我们筛选了产生细胞抗热性的锌指蛋白质。编码不同锌指蛋白质的核酸文库转化入大肠杆菌细胞中。细胞暴露于热，和回收具有抗热性的细胞。第三轮表型筛选后从 23 个选择的单个菌落中纯化质粒。鉴定了 10 个不同的锌指蛋白质（表 15）并通过比较 ZFP 转化体和对照细胞，C1 或 ZFP-2 在处理后的存活率分析了热耐受性的提高。C1 或 ZFP-2 分别代表空载体或无关 ZFP（QTHQ-RSHR-QTHR1）的转化体。超过 99.99% 的野生型细胞在 50℃ 热处理 2 小时后死亡。相反，大约 6% 的转化了某种 ZFP-TFs 的细胞在这些极端条件下存活，700 倍提高的热耐受性表型——即，在刺激条件下存活的表达 ZFP-TFs 的细胞的百分比（6.3%）除以在相同条件下存活的 C1 的百分比（0.0085%）。因为 IPTG 诱导 ZFP 的表达，分析了诱导热耐受性的 ZFP 转化体在存在和不存在 IPTG 时的表型。T1 到 T10 转化体在存在 IPTG 时显示了较高的热耐受性。

表 15: 产生热耐受性的 ZFP。

名称	F1	F2	F3	F4	推定的DNA靶	出现数量
T-1	QSHV	VSNV	QSNK	QSNK	5' DAA DAA AAT HGA 3' (SEQ ID NO:238)	6
T-2	RDHT	QSHV	QTHR1	QSSR1	5' GYA GRA HGA NGG K 3' (SEQ ID NO:239)	3
T-3	WSNR	QSHV	VSNV	QSHV	5' HGA AAT HGA GGT 3' (SEQ ID NO:240)	1
T-4	QTHR1	RSHR	QTHR1	QTHR1	5' GRA GRA GGG GRA 3' (SEQ ID NO:241)	1
T-5	DSAR	RDHT	QSHV	QTHR1	5' GRA HGA NGG GTC 3' (SEQ ID NO:242)	2
T-6	QTHQ	RSHR	QTHR1	QTHR1	5' GRA GRA GGG HGA 3' (SEQ ID NO:243)	1
T-7	QSHV	VSNV	QSNR1	CSNR1	5' GAV GAA AAT HGA 3' (SEQ ID NO:244)	3
T-8	VSNV	QTHR1	QSSR1	RDHT	5' NGG GYA GRA AAT 3' (SEQ ID NO:245)	2
T-9	RDHT	QSHV	QTHR1	QSNR1	5' GAA GRA HGA NGG K 3' (SEQ ID NO:246)	2
T-10	DSAR	RDHT	QSNK	QTHR1	5' GRA DAA NGG GTC 3' (SEQ ID NO:247)	2

筛选热耐受性。如实施例 9 所述制备原核表达文库。大肠杆菌株 DH5 α 用 3-指或 4-指 ZFP 文库转化并在含氯霉素 (34 μ g/ml) 的 LB 上培养过夜。过夜培养物在含 1 μ M IPTG 和氯霉素 (34 μ g/ml) 的 1ml 新鲜 LB 培养基中稀释至 1: 500 以诱导 ZFP 表达。在 30 $^{\circ}$ C 孵育 3 小时后, 100 μ l 培养物转移到微离心管中在 50 $^{\circ}$ C 水浴孵育 2 小时。培养物铺在含氯霉素 (34 μ g/ml) 的 LB 板上。从生长菌落池中纯化质粒并转化 DH5 α 。用再转化体重复选择热耐受性。第三轮选择后, 从 30 个生长在 LB+氯霉素 (34 μ g/ml) 的板上的单个菌落中纯化质粒并再转化 DH5 α 。如上述分析每个转化体的热耐受性。可以诱导热耐受性的质粒被测序以鉴定 ZFP。

实施例 11: ZFP F121-P65 调控的基因

锌指蛋白质 F121_p65 在人胚胎肾 293 细胞中表达。这些细胞的转录产物被检测并与相应的不表达 F121_p65 的对照细胞的分布图比较。F121_p65 上调的转录产物的实施例列于表 16。F121_p65 在其它基因中提高了胰岛素样生长因子 2 的转录。

表 16: F121_p65 上调的基因

名称	ID	中值比例 (ROM)
胰岛素样生长因子 2	H59614	17.71
胰岛素样生长因子 2	H59614	16.88
蛋白酪氨酸磷酸酶, 受体	R45961	16.45
胰岛素样生长因子 2	N54596	14.40
推定的基因产物	H09111	12.35
jun B 原癌基因	N94468	11.61
细胞维甲酸结合蛋白 2	AA598508	10.30
蛋白磷酸酶 2, 调节亚基	R59165	9.89
活化 T 细胞核因子	AA679278	9.64
FK506-结合蛋白 8 (38kD)	R95418	8.66
蛋白磷酸酶 2, 调节亚基	R59165	8.60
钙粘着蛋白 13, H-钙粘着蛋白 (心)	R41787	7.34
富半胱氨酸蛋白 2	AA485427	6.71
与 Zfp-36 同源的锌指蛋白	R38383	6.46
肿瘤坏死因子 (配体) 超家族	AI347622	6.15
钙粘着蛋白 13, H-钙粘着蛋白 (心)	R41787	5.90
脑特异性蛋白 p25 α	AA133959	5.71
生长停滞和 DNA 损伤诱导型	AA404666	5.55
淋巴细胞抗原 6 复合体, 基因座 H	AI929550	5.50
己酮糖激酶 (果糖激酶)	T61256	5.28
细胞周期素 D1 (PRAD1: 甲状旁腺腺瘤病)	AA487486	5.27
—	AA045731	5.27
I 型跨膜蛋白的 Fn14	AI221536	5.06

实施例 12: 大肠杆菌中外源蛋白质溶解度的变化

我们筛选了使重组蛋白质, 大肠杆菌表达的 Akt1 的溶解度提高的锌指蛋白质。哺乳动物 Akt1 基因克隆在 pET21b 载体中并且 GFP 连接到 Akt1 ORF 的 C-末端。编码不同锌指蛋白质的核酸文库转化入大肠杆菌细胞中。Akt1: GFP 构建体导入 ZFP 转化体中。通过 FACS 排序分析 Akt1 提高的溶解度, 因为可溶 Akt1 表达细胞产生更多的荧光 GFP。8 轮表型筛选后, 选择单个菌落并通过 western 印迹分析比较 ZFP 表达和 C1 表达细胞之间的可溶组分中 Akt1 的量来分析溶解度的提高。western 印迹结果通过成像分析软件, QUANTITY ONE (Bio-Rad, Hercules, CA) 量化。C1 代表空载体 pZL1 的转化体。鉴定了 2 个不同的提高了大肠杆菌表达

的重组 Akt1 蛋白质的溶解度的锌指蛋白质。

C1 中可溶和不可溶组分之间的 Akt1 量的比率是 32.5: 67.5。相反, 转化了某种锌指蛋白质的细胞显示了可溶组分中 Akt1 的比率提高了 2 倍一即, 细胞表达 ZFP-TFs 的可溶组分中 Akt1 蛋白质的量的百分比 (66.1%) 除以 C1 中的量 (32.5%)。

表 19: 提高大肠杆菌中 Akt1 蛋白质溶解度的 ZFP。

表达构建体	名称	Akt1 蛋白数的比例 (%)	
		可溶成分	不溶成分
对照	C1	32.5	67.5
溶解性增加的ZFP	S-1	66.1	33.9
溶解性增加的ZFP	S-2	45.7	54.3

表 20: 使大肠杆菌 Akt1 蛋白质溶解度提高的 ZFP。

名称	F1	F2	F3	F4	推定的DNA靶
S-1	QSTR	DSAR	RDHT	WSNR	GGT NGG GTC GYA (SEQ ID NO:248)
S-2	VSTR	DGNV	QSNR	QSNK	DAA GAA AAC GCT (SEQ ID NO:249)

筛选溶解度提高了的重组蛋白质。

如实施例 9 所述制备原核表达文库。用 ZFP 文库和 Akt1 表达载体共转化大肠杆菌株 DH5 α 并在含氯霉素 (34 μ g/ml) 和氨苄青霉素 (50 μ g/ml) 的 LB 上培养过夜。过夜培养物在含 1mM IPTG 和氯霉素 (34 μ g/ml) 和氨苄青霉素 (50 μ g/ml) 的 1ml 新鲜 LB 培养基中稀释至 1: 500 以诱导 ZFP 表达。30 $^{\circ}$ C 孵育 3 小时后, 用 FACS 在 FACS Vantage 流式细胞仪上分析 10⁶ 种细胞并分选 5% 到 10% 表达高荧光的细胞。GFP 表达或 pET21b 空载体转化的细胞用作正或阴性对照以设定背景荧光。

分析的细胞在含氯霉素 (34 μ g/ml) 和氨苄青霉素 (50 μ g/ml) 的 LB 培养基中过夜培养。从培养物中纯化质粒并转化入 DH5 α 。用再转化体重复筛选提高溶解度的重组蛋白质。第八轮选择后, 从可以在 LB+氯霉素

(34 μ g/ml) + 氨苄青霉素 (50 μ g/ml) 的板上生长的单个菌落中纯化质粒并转化入 DH5 α 。用 western 印迹分析法分析每个转化体提高的 Akt1 的溶解度。测序可以提高 Akt1 溶解度的质粒以鉴定 ZFP。

描述了许多本发明的实施方案。然而，应理解可以进行不偏离本发明精神和范围的多种修饰。因此，其它实施方案也在以下权利要求的范围内。

序列表

<110> 图尔金株式会社

<120> 嵌合蛋白质的表型筛选

<130> PCA21281-TGI

<140> PCT/KR2002/002309

<141> 2002-12-07

<150> US 60/338,441

<151> 2001-12-07

<150> US 60/376,053

<151> 2002-04-26

<150> US 60/400,904

<151> 2002-08-02

<150> US 60/401,089

<151> 2002-08-05

<160> 266

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 1137

<212> DNA

<213> 人工序列

<220>

<223> 质粒序列

<221> CDS

<222> (1)...(1134)

<400> 1

atg	gtg	tac	ccc	tac	gac	gtg	ccc	gac	tac	gcc	gaa	ttg	cct	cca	aaa	48
Met	Val	Tyr	Pro	Tyr	Asp	Val	Pro	Asp	Tyr	Ala	Glu	Leu	Pro	Pro	Lys	
1				5					10					15		

aag	aag	aga	aag	gta	ggg	atc	cga	att	ccc	ggg	gaa	aaa	ccg	ttt	gag	96
Lys	Lys	Arg	Lys	Val	Gly	Ile	Arg	Ile	Pro	Gly	Glu	Lys	Pro	Phe	Glu	
			20				25						30			

tgt	aaa	gat	tgc	ggg	aaa	gct	ttc	att	cag	aag	tca	aac	ctc	atc	aga	144
Cys	Lys	Asp	Cys	Gly	Lys	Ala	Phe	Ile	Gln	Lys	Ser	Asn	Leu	Ile	Arg	
		35					40					45				

cac	cag	aga	act	cac	acc	ggg	gaa	aaa	ccg	tac	aag	tgt	gaa	gaa	tgt	192
His	Gln	Arg	Thr	His	Thr	Gly	Glu	Lys	Pro	Tyr	Lys	Cys	Glu	Glu	Cys	
	50					55					60					

ggc	aaa	gct	ttt	acc	caa	tcc	tca	aac	ctt	act	aaa	cat	aag	aaa	att	240
Gly	Lys	Ala	Phe	Thr	Gln	Ser	Ser	Asn	Leu	Thr	Lys	His	Lys	Lys	Ile	
65					70					75					80	

cat	acc	ggg	gaa	aaa	ccg	tat	aaa	tgt	aag	caa	tgt	ggg	aaa	gct	ttt	288
His	Thr	Gly	Glu	Lys	Pro	Tyr	Lys	Cys	Lys	Gln	Cys	Gly	Lys	Ala	Phe	
				85				90						95		

gga	tgt	ccc	tca	aac	ctt	cga	agg	cat	gga	agg	act	cac	acc	ggt	gaa	336
Gly	Cys	Pro	Ser	Asn	Leu	Arg	Arg	His	Gly	Arg	Thr	His	Thr	Gly	Glu	

100				105				110				
aaa gcg gcc gct	aaa ttc tac	ctg cca gat	aca gac gat	cgt cac cgg								384
Lys Ala Ala Ala	Lys Phe Tyr	Leu Pro Asp Thr	Asp Asp Arg	His Arg								
115		120	125									
att gag gag aaa	cgt aaa agg	aca tat gag	acc ttc aag	agc atc atg								432
Ile Glu Glu Lys	Arg Lys Arg	Thr Tyr Glu Thr	Phe Lys Ser	Ile Met								
130		135	140									
aag aag agt cct	ttc agc gga	ccc acc gac	ccc cgg cct	cca cct cga								480
Lys Lys Ser Pro	Phe Ser Gly	Pro Thr Asp	Pro Arg Pro	Pro Arg								
145	150	155	160									
cgc att gct gtg	cct tcc cgc	agc tca gct	tct gtc ccc	aag cca gca								528
Arg Ile Ala Val	Pro Ser Arg	Ser Ser Ala	Ser Val Pro	Lys Pro Ala								
165		170	175									
ccc cag ccc tat	ccc ttt acg	tca tcc ctg	agc acc atc	aac tat gat								576
Pro Gln Pro Tyr	Pro Phe Thr	Ser Ser Leu	Ser Thr Ile	Asn Tyr Asp								
180		185	190									
gag ttt ccc acc	atg gtg ttt	cct tct ggg	cag atc agc	cag gcc tcg								624
Glu Phe Pro Thr	Met Val Phe	Pro Ser Gly	Gln Ile Ser	Gln Ala Ser								
195		200	205									
gcc ttg gcc ccg	gcc cct ccc	caa gtc ctg	ccc cag gct	cca gcc cct								672
Ala Leu Ala Pro	Ala Pro Pro	Gln Val Leu	Pro Gln Ala	Pro Ala Pro								
210		215	220									
gcc cct gct cca	gcc atg gta	tca gct ctg	gcc cag gcc	cca gcc cct								720
Ala Pro Ala Pro	Ala Met Val	Ser Ala Leu	Ala Gln Ala	Pro Ala Pro								
225	230	235	240									
gtc cca gtc cta	gcc cca ggc	cct cct cag	gct gtg gcc	cca cct gcc								768
Val Pro Val Leu	Ala Pro Gly	Pro Pro Gln	Ala Val Ala	Pro Pro Ala								
245		250	255									
ccc aag ccc acc	cag gct ggg	gaa gga acg	ctg tca gag	gcc ctg ctg								816
Pro Lys Pro Thr	Gln Ala Gly	Glu Gly Thr	Leu Ser Glu	Ala Leu Leu								
260		265	270									
cag ctg cag ttt	gat gat gaa	gac ctg ggg	gcc ttg ctt	ggc aac agc								864
Gln Leu Gln Phe	Asp Asp Glu	Asp Leu Gly	Ala Leu Leu	Gly Asn Ser								
275		280	285									
aca gac cca gct	gtg ttc aca	gac ctg gca	tcc gtc gac	aac tcc gag								912
Thr Asp Pro Ala	Val Phe Thr	Asp Leu Ala	Ser Val Asp	Asn Ser Glu								
290		295	300									
ttt cag cag ctg	ctg aac cag	ggc ata cct	gtg gcc ccc	cac aca act								960
Phe Gln Gln Leu	Leu Asn Gln	Gly Ile Pro	Val Ala Pro	His Thr Thr								
305	310	315	320									
gag ccc atg ctg	atg gag tac	cct gag gct	ata act cgc	cta gtg aca								1008
Glu Pro Met Leu	Met Glu Tyr	Pro Glu Ala	Ile Thr Arg	Leu Val Thr								
325		330	335									
gcc cag agg ccc	ccc gac cca	gct cct gct	cca ctg ggg	gcc ccg ggg								1056
Ala Gln Arg Pro	Pro Asp Pro	Ala Pro Ala	Pro Leu Gly	Ala Pro Gly								
340		345	350									
ctc ccc aat ggc	ctc ctt tca	gga gat gaa	gac ttc tcc	tcc att gcg								1104
Leu Pro Asn Gly	Leu Leu Ser	Gly Asp Glu	Asp Phe Ser	Ser Ile Ala								
355		360	365									

gac atg gac ttc tca gcc ctg ctg agt cag taa
 Asp Met Asp Phe Ser Ala Leu Leu Ser Gln
 370 375

1137

<210> 2
 <211> 378
 <212> PRT
 <213> 人工序列

<220>
 <223> 质粒序列

<400> 2
 Met Val Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Glu Leu Pro Pro Lys
 1 5 10 15
 Lys Lys Arg Lys Val Gly Ile Arg Ile Pro Gly Glu Lys Pro Phe Glu
 20 25 30
 Cys Lys Asp Cys Gly Lys Ala Phe Ile Gln Lys Ser Asn Leu Ile Arg
 35 40 45
 His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Glu Glu Cys
 50 55 60
 Gly Lys Ala Phe Thr Gln Ser Ser Asn Leu Thr Lys His Lys Lys Ile
 65 70 75 80
 His Thr Gly Glu Lys Pro Tyr Lys Cys Lys Gln Cys Gly Lys Ala Phe
 85 90 95
 Gly Cys Pro Ser Asn Leu Arg Arg His Gly Arg Thr His Thr Gly Glu
 100 105 110
 Lys Ala Ala Ala Lys Phe Tyr Leu Pro Asp Thr Asp Asp Arg His Arg
 115 120 125
 Ile Glu Glu Lys Arg Lys Arg Thr Tyr Glu Thr Phe Lys Ser Ile Met
 130 135 140
 Lys Lys Ser Pro Phe Ser Gly Pro Thr Asp Pro Arg Pro Pro Pro Arg
 145 150 155 160
 Arg Ile Ala Val Pro Ser Arg Ser Ser Ala Ser Val Pro Lys Pro Ala
 165 170 175
 Pro Gln Pro Tyr Pro Phe Thr Ser Ser Leu Ser Thr Ile Asn Tyr Asp
 180 185 190
 Glu Phe Pro Thr Met Val Phe Pro Ser Gly Gln Ile Ser Gln Ala Ser
 195 200 205
 Ala Leu Ala Pro Ala Pro Pro Gln Val Leu Pro Gln Ala Pro Ala Pro
 210 215 220
 Ala Pro Ala Pro Ala Met Val Ser Ala Leu Ala Gln Ala Pro Ala Pro
 225 230 235 240
 Val Pro Val Leu Ala Pro Gly Pro Pro Gln Ala Val Ala Pro Pro Ala
 245 250 255
 Pro Lys Pro Thr Gln Ala Gly Glu Gly Thr Leu Ser Glu Ala Leu Leu
 260 265 270
 Gln Leu Gln Phe Asp Asp Glu Asp Leu Gly Ala Leu Leu Gly Asn Ser
 275 280 285
 Thr Asp Pro Ala Val Phe Thr Asp Leu Ala Ser Val Asp Asn Ser Glu
 290 295 300
 Phe Gln Gln Leu Leu Asn Gln Gly Ile Pro Val Ala Pro His Thr Thr
 305 310 315 320
 Glu Pro Met Leu Met Glu Tyr Pro Glu Ala Ile Thr Arg Leu Val Thr
 325 330 335
 Ala Gln Arg Pro Pro Asp Pro Ala Pro Ala Pro Leu Gly Ala Pro Gly
 340 345 350
 Leu Pro Asn Gly Leu Leu Ser Gly Asp Glu Asp Phe Ser Ser Ile Ala
 355 360 365
 Asp Met Asp Phe Ser Ala Leu Leu Ser Gln
 370 375

<210> 3
 <211> 1233

<212> DNA
<213> 人工序列

<220>
<223> 质粒序列

<221> CDS
<222> (1)...(1230)

```

<400> 3
atg gtg tac ccc tac gac gtg ccc gac tac gcc gaa ttg cct cca aaa      48
Met Val Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Glu Leu Pro Pro Lys
  1             5             10             15

aag aag aga aag gta ggg atc cga att ccc ggg gaa aaa ccg tat gta      96
Lys Lys Arg Lys Val Gly Ile Arg Ile Pro Gly Glu Lys Pro Tyr Val
             20             25             30

tgc gat gta gag gga tgt acg tgg aaa ttt gcc cgc tca gat aag ctc     144
Cys Asp Val Glu Gly Cys Thr Trp Lys Phe Ala Arg Ser Asp Lys Leu
             35             40             45

aac aga cac aag aaa agg cac acc ggg gaa aaa ccg tat gag tgt cac     192
Asn Arg His Lys Lys Arg His Thr Gly Glu Lys Pro Tyr Glu Cys His
             50             55             60

gat tgc gga aag tcc ttt agg cag agc acc cac ctc act cgg cac cgg     240
Asp Cys Gly Lys Ser Phe Arg Gln Ser Thr His Leu Thr Arg His Arg
             65             70             75             80

agg atc cac acc ggg gaa aaa ccg tat gag tgt aat tac tgt gga aaa     288
Arg Ile His Thr Gly Glu Lys Pro Tyr Glu Cys Asn Tyr Cys Gly Lys
             85             90             95

acc ttt agt gtg agc tca acc ctt att aga cat cag aga atc cac acc     336
Thr Phe Ser Val Ser Ser Thr Leu Ile Arg His Gln Arg Ile His Thr
             100             105             110

ggg gaa aaa ccg tat gta tgc gat gta gag gga tgt acg tgg aaa ttt     384
Gly Glu Lys Pro Tyr Val Cys Asp Val Glu Gly Cys Thr Trp Lys Phe
             115             120             125

gcc cgc tca gat aag ctc aac aga cac aag aaa agg cac acc ggt gaa     432
Ala Arg Ser Asp Lys Leu Asn Arg His Lys Lys Arg His Thr Gly Glu
             130             135             140

aaa gcg gcc gct aaa ttc tac ctg cca gat aca gac gat cgt cac cgg     480
Lys Ala Ala Ala Lys Phe Tyr Leu Pro Asp Thr Asp Asp Arg His Arg
             145             150             155             160

att gag gag aaa cgt aaa agg aca tat gag acc ttc aag agc atc atg     528
Ile Glu Glu Lys Arg Lys Arg Thr Tyr Glu Thr Phe Lys Ser Ile Met
             165             170             175

aag aag agt cct ttc agc gga ccc acc gac ccc cgg cct cca cct cga     576
Lys Lys Ser Pro Phe Ser Gly Pro Thr Asp Pro Arg Pro Pro Pro Arg
             180             185             190

cgc att gct gtg cct tcc cgc agc tca gct tct gtc ccc aag cca gca     624
Arg Ile Ala Val Pro Ser Arg Ser Ser Ala Ser Val Pro Lys Pro Ala
             195             200             205

ccc cag ccc tat ccc ttt acg tca tcc ctg agc acc atc aac tat gat     672
Pro Gln Pro Tyr Pro Phe Thr Ser Ser Leu Ser Thr Ile Asn Tyr Asp
             210             215             220

```

gag ttt ccc acc atg gtg ttt cct tct ggg cag atc agc cag gcc tcg Glu Phe Pro Thr Met Val Phe Pro Ser Gly Gln Ile Ser Gln Ala Ser 225 230 235 240	720
gcc ttg gcc ccg gcc cct ccc caa gtc ctg ccc cag gct cca gcc cct Ala Leu Ala Pro Ala Pro Pro Gln Val Leu Pro Gln Ala Pro Ala Pro 245 250 255	768
gcc cct gct cca gcc atg gta tca gct ctg gcc cag gcc cca gcc cct Ala Pro Ala Pro Ala Met Val Ser Ala Leu Ala Gln Ala Pro Ala Pro 260 265 270	816
gtc cca gtc cta gcc cca ggc cct cct cag gct gtg gcc cca cct gcc Val Pro Val Leu Ala Pro Gly Pro Pro Gln Ala Val Ala Pro Pro Ala 275 280 285	864
ccc aag ccc acc cag gct ggg gaa gga acg ctg tca gag gcc ctg ctg Pro Lys Pro Thr Gln Ala Gly Glu Gly Thr Leu Ser Glu Ala Leu Leu 290 295 300	912
cag ctg cag ttt gat gat gaa gac ctg ggg gcc ttg ctt ggc aac agc Gln Leu Gln Phe Asp Asp Glu Asp Leu Gly Ala Leu Leu Gly Asn Ser 305 310 315 320	960
aca gac cca gct gtg ttc aca gac ctg gca tcc gtc gac aac tcc gag Thr Asp Pro Ala Val Phe Thr Asp Leu Ala Ser Val Asp Asn Ser Glu 325 330 335	1008
ttt cag cag ctg ctg aac cag ggc ata cct gtg gcc ccc cac aca act Phe Gln Gln Leu Leu Asn Gln Gly Ile Pro Val Ala Pro His Thr Thr 340 345 350	1056
gag ccc atg ctg atg gag tac cct gag gct ata act cgc cta gtg aca Glu Pro Met Leu Met Glu Tyr Pro Glu Ala Ile Thr Arg Leu Val Thr 355 360 365	1104
gcc cag agg ccc ccc gac cca gct cct gct cca ctg ggg gcc ccg ggg Ala Gln Arg Pro Pro Asp Pro Ala Pro Ala Pro Leu Gly Ala Pro Gly 370 375 380	1152
ctc ccc aat ggc ctc ctt tca gga gat gaa gac ttc tcc tcc att gcg Leu Pro Asn Gly Leu Leu Ser Gly Asp Glu Asp Phe Ser Ser Ile Ala 385 390 395 400	1200
gac atg gac ttc tca gcc ctg ctg agt cag taa Asp Met Asp Phe Ser Ala Leu Leu Ser Gln 405 410	1233

<210> 4
 <211> 382
 <212> PRT
 <213> 人工序列

<220>
 <223> 质粒序列

<400> 4
 Met Val Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Glu Leu Pro Pro Lys
 1 5 10 15
 Lys Lys Arg Lys Val Gly Ile Arg Ile Pro Gly Glu Lys Pro Phe Glu
 20 25 30
 Cys Lys Asp Cys Gly Lys Ala Phe Ile Gln Lys Ser Asn Leu Ile Arg
 35 40 45
 His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Ala Cys Pro Val Glu

50	55	60
Ser Cys Asp Arg Arg Phe Ser Asp Ser Ser Asn Leu Thr Arg His Ile		
65	70	75
Arg Ile His Thr Gly Glu Lys Pro Tyr Ala Cys Pro Val Glu Ser Cys		
	85	90
Asp Arg Arg Phe Ser Asp Ser Ser Asn Leu Thr Arg His Ile Arg Ile		
	100	105
His Thr Gly Glu Lys Ala Ala Ala Lys Phe Tyr Leu Pro Asp Thr Asp		
	115	120
Asp Arg His Arg Ile Glu Glu Lys Arg Lys Arg Thr Tyr Glu Thr Phe		
	130	135
Lys Ser Ile Met Lys Lys Ser Pro Phe Ser Gly Pro Thr Asp Pro Arg		
145	150	155
Pro Pro Pro Arg Arg Ile Ala Val Pro Ser Arg Ser Ser Ala Ser Val		
	165	170
Pro Lys Pro Ala Pro Gln Pro Tyr Pro Phe Thr Ser Ser Leu Ser Thr		
	180	185
Ile Asn Tyr Asp Glu Phe Pro Thr Met Val Phe Pro Ser Gly Gln Ile		
	195	200
Ser Gln Ala Ser Ala Leu Ala Pro Ala Pro Pro Gln Val Leu Pro Gln		
	210	215
Ala Pro Ala Pro Ala Pro Ala Pro Ala Met Val Ser Ala Leu Ala Gln		
225	230	235
Ala Pro Ala Pro Val Pro Val Leu Ala Pro Gly Pro Pro Gln Ala Val		
	245	250
Ala Pro Pro Ala Pro Lys Pro Thr Gln Ala Gly Glu Gly Thr Leu Ser		
	260	265
Glu Ala Leu Leu Gln Leu Gln Phe Asp Asp Glu Asp Leu Gly Ala Leu		
	275	280
Leu Gly Asn Ser Thr Asp Pro Ala Val Phe Thr Asp Leu Ala Ser Val		
	290	295
Asp Asn Ser Glu Phe Gln Gln Leu Leu Asn Gln Gly Ile Pro Val Ala		
305	310	315
Pro His Thr Thr Glu Pro Met Leu Met Glu Tyr Pro Glu Ala Ile Thr		
	325	330
Arg Leu Val Thr Ala Gln Arg Pro Pro Asp Pro Ala Pro Ala Pro Leu		
	340	345
Gly Ala Pro Gly Leu Pro Asn Gly Leu Leu Ser Gly Asp Glu Asp Phe		
	355	360
Ser Ser Ile Ala Asp Met Asp Phe Ser Ala Leu Leu Ser Gln		
370	375	380

<210> 5

<211> 1137

<212> DNA

<213> 人工序列

<220>

<223> 质粒序列

<221> CDS

<222> (1)...(1134)

<400> 5

atg gtg tac ccc tac gac gtg ccc gac tac gcc gaa ttg cct cca aaa 48

Met Val Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Glu Leu Pro Pro Lys

1

5

10

15

aag aag aga aag gta ggg atc cga att ccc ggg gaa aaa ccg tat aag 96

Lys Lys Arg Lys Val Gly Ile Arg Ile Pro Gly Glu Lys Pro Tyr Lys

20

25

30

tgc atg gag tgt ggg aag gct ttt aac cgc agg tca cac ctc aca cgg 144

Cys Met Glu Cys Gly Lys Ala Phe Asn Arg Arg Ser His Leu Thr Arg

35

40

45

cac cag cgg att cac acc ggg gaa aaa ccg ttc cag tgt aaa act tgt His Gln Arg Ile His Thr Gly Glu Lys Pro Phe Gln Cys Lys Thr Cys 50 55 60	192
cag cga aag ttc tcc cgg tcc gac cac ctg aag acc cac acc agg act Gln Arg Lys Phe Ser Arg Ser Asp His Leu Lys Thr His Thr Arg Thr 65 70 75 80	240
cat acc ggg gaa aaa ccg tat aca tgt aaa cag tgt ggg aaa gcc ttc His Thr Gly Glu Lys Pro Tyr Thr Cys Lys Gln Cys Gly Lys Ala Phe 85 90 95	288
agt gtt tcc agt tcc ctt cga aga cat gaa acc act cac acc ggt gaa Ser Val Ser Ser Ser Leu Arg Arg His Glu Thr Thr His Thr Gly Glu 100 105 110	336
aaa gcg gcc gct aaa ttc tac ctg cca gat aca gac gat cgt cac cgg Lys Ala Ala Ala Lys Phe Tyr Leu Pro Asp Thr Asp Asp Arg His Arg 115 120 125	384
att gag gag aaa cgt aaa agg aca tat gag acc ttc aag agc atc atg Ile Glu Glu Lys Arg Lys Arg Thr Tyr Glu Thr Phe Lys Ser Ile Met 130 135 140	432
aag aag agt cct ttc agc gga ccc acc gac ccc cgg cct cca cct cga Lys Lys Ser Pro Phe Ser Gly Pro Thr Asp Pro Arg Pro Pro Pro Arg 145 150 155 160	480
cgc att gct gtg cct tcc cgc agc tca gct tct gtc ccc aag cca gca Arg Ile Ala Val Pro Ser Arg Ser Ser Ala Ser Val Pro Lys Pro Ala 165 170 175	528
ccc cag ccc tat ccc ttt acg tca tcc ctg agc acc atc aac tat gat Pro Gln Pro Tyr Pro Phe Thr Ser Ser Leu Ser Thr Ile Asn Tyr Asp 180 185 190	576
gag ttt ccc acc atg gtg ttt cct tct ggg cag atc agc cag gcc tcg Glu Phe Pro Thr Met Val Phe Pro Ser Gly Gln Ile Ser Gln Ala Ser 195 200 205	624
gcc ttg gcc ccg gcc cct ccc caa gtc ctg ccc cag gct cca gcc cct Ala Leu Ala Pro Ala Pro Pro Gln Val Leu Pro Gln Ala Pro Ala Pro 210 215 220	672
gcc cct gct cca gcc atg gta tca gct ctg gcc cag gcc cca gcc cct Ala Pro Ala Pro Ala Met Val Ser Ala Leu Ala Gln Ala Pro Ala Pro 225 230 235 240	720
gtc cca gtc cta gcc cca ggc cct cct cag gct gtg gcc cca cct gcc Val Pro Val Leu Ala Pro Gly Pro Pro Gln Ala Val Ala Pro Pro Ala 245 250 255	768
ccc aag ccc acc cag gct ggg gaa gga acg ctg tca gag gcc ctg ctg Pro Lys Pro Thr Gln Ala Gly Glu Gly Thr Leu Ser Glu Ala Leu Leu 260 265 270	816
cag ctg cag ttt gat gat gaa gac ctg ggg gcc ttg ctt ggc aac agc Gln Leu Gln Phe Asp Asp Glu Asp Leu Gly Ala Leu Leu Gly Asn Ser 275 280 285	864
aca gac cca gct gtg ttc aca gac ctg gca tcc gtc gac aac tcc gag Thr Asp Pro Ala Val Phe Thr Asp Leu Ala Ser Val Asp Asn Ser Glu 290 295 300	912

```

ttt cag cag ctg ctg aac cag ggc ata cct gtg gcc ccc cac aca act      960
Phe Gln Gln Leu Leu Asn Gln Gly Ile Pro Val Ala Pro His Thr Thr
305                      310                      315                      320

gag ccc atg ctg atg gag tac cct gag gct ata act cgc cta gtg aca      1008
Glu Pro Met Leu Met Glu Tyr Pro Glu Ala Ile Thr Arg Leu Val Thr
                      325                      330                      335

gcc cag agg ccc ccc gac cca gct cct gct cca ctg ggg gcc ccg ggg      1056
Ala Gln Arg Pro Pro Asp Pro Ala Pro Ala Pro Leu Gly Ala Pro Gly
                      340                      345                      350

ctc ccc aat ggc ctc ctt tca gga gat gaa gac ttc tcc tcc att gcg      1104
Leu Pro Asn Gly Leu Leu Ser Gly Asp Glu Asp Phe Ser Ser Ile Ala
                      355                      360                      365

gac atg gac ttc tca gcc ctg ctg agt cag taa      1137
Asp Met Asp Phe Ser Ala Leu Leu Ser Gln
                      370                      375

```

<210> 6
 <211> 378
 <212> PRT
 <213> 人工序列

<220>
 <223> 质粒序列

```

<400> 6
Met Val Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Glu Leu Pro Pro Lys
 1      5      10      15
Lys Lys Arg Lys Val Gly Ile Arg Ile Pro Gly Glu Lys Pro Tyr Lys
 20      25      30
Cys Met Glu Cys Gly Lys Ala Phe Asn Arg Arg Ser His Leu Thr Arg
 35      40      45
His Gln Arg Ile His Thr Gly Glu Lys Pro Phe Gln Cys Lys Thr Cys
 50      55      60
Gln Arg Lys Phe Ser Arg Ser Asp His Leu Lys Thr His Thr Arg Thr
 65      70      75      80
His Thr Gly Glu Lys Pro Tyr Thr Cys Lys Gln Cys Gly Lys Ala Phe
 85      90      95
Ser Val Ser Ser Ser Leu Arg Arg His Glu Thr Thr His Thr Gly Glu
100      105      110
Lys Ala Ala Ala Lys Phe Tyr Leu Pro Asp Thr Asp Asp Arg His Arg
115      120      125
Ile Glu Glu Lys Arg Lys Arg Thr Tyr Glu Thr Phe Lys Ser Ile Met
130      135      140
Lys Lys Ser Pro Phe Ser Gly Pro Thr Asp Pro Arg Pro Pro Pro Arg
145      150      155      160
Arg Ile Ala Val Pro Ser Arg Ser Ser Ala Ser Val Pro Lys Pro Ala
165      170      175
Pro Gln Pro Tyr Pro Phe Thr Ser Ser Leu Ser Thr Ile Asn Tyr Asp
180      185      190
Glu Phe Pro Thr Met Val Phe Pro Ser Gly Gln Ile Ser Gln Ala Ser
195      200      205
Ala Leu Ala Pro Ala Pro Pro Gln Val Leu Pro Gln Ala Pro Ala Pro
210      215      220
Ala Pro Ala Pro Ala Met Val Ser Ala Leu Ala Gln Ala Pro Ala Pro
225      230      235      240
Val Pro Val Leu Ala Pro Gly Pro Pro Gln Ala Val Ala Pro Pro Ala
245      250      255
Pro Lys Pro Thr Gln Ala Gly Glu Gly Thr Leu Ser Glu Ala Leu Leu
260      265      270
Gln Leu Gln Phe Asp Asp Glu Asp Leu Gly Ala Leu Leu Gly Asn Ser
275      280      285

```

```

Thr Asp Pro Ala Val Phe Thr Asp Leu Ala Ser Val Asp Asn Ser Glu
 290      295      300
Phe Gln Gln Leu Leu Asn Gln Gly Ile Pro Val Ala Pro His Thr Thr
 305      310      315      320
Glu Pro Met Leu Met Glu Tyr Pro Glu Ala Ile Thr Arg Leu Val Thr
      325      330      335
Ala Gln Arg Pro Pro Asp Pro Ala Pro Ala Pro Leu Gly Ala Pro Gly
      340      345      350
Leu Pro Asn Gly Leu Leu Ser Gly Asp Glu Asp Phe Ser Ser Ile Ala
      355      360      365
Asp Met Asp Phe Ser Ala Leu Leu Ser Gln
      370      375

```

<210> 7

<211> 1149

<212> DNA

<213> 人工序列

<220>

<223> 质粒序列

<221> CDS

<222> (1)...(1146)

<400> 7

```

atg gtg tac ccc tac gac gtg ccc gac tac gcc gaa ttg cct cca aaa      48
Met Val Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Glu Leu Pro Pro Lys
 1          5          10          15

aag aag aga aag gta ggg atc cga att ccc ggg gaa aaa ccg ttt gag      96
Lys Lys Arg Lys Val Gly Ile Arg Ile Pro Gly Glu Lys Pro Phe Glu
      20          25          30

tgt aaa gat tgc ggg aaa gct ttc att cag aag tca aac ctc atc aga      144
Cys Lys Asp Cys Gly Lys Ala Phe Ile Gln Lys Ser Asn Leu Ile Arg
      35          40          45

cac cag aga act cac acc ggg gaa aaa ccg tat gct tgc cct gtc gag      192
His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Ala Cys Pro Val Glu
      50          55          60

tcc tgc gat cgc cgc ttt tct gat tcg tcg aac ctt acc cgc cat atc      240
Ser Cys Asp Arg Arg Phe Ser Asp Ser Ser Asn Leu Thr Arg His Ile
      65          70          75          80

cgc atc cac acc ggg gaa aaa ccg tat gct tgc cct gtc gag tcc tgc      288
Arg Ile His Thr Gly Glu Lys Pro Tyr Ala Cys Pro Val Glu Ser Cys
      85          90          95

gat cgc cgc ttt tct gat tcg tcg aac ctt acc cgc cat atc cgc atc      336
Asp Arg Arg Phe Ser Asp Ser Ser Asn Leu Thr Arg His Ile Arg Ile
      100          105          110

cac acc ggt gaa aaa gcg gcc gct aaa ttc tac ctg cca gat aca gac      384
His Thr Gly Glu Lys Ala Ala Lys Phe Tyr Leu Pro Asp Thr Asp
      115          120          125

gat cgt cac cgg att gag gag aaa cgt aaa agg aca tat gag acc ttc      432
Asp Arg His Arg Ile Glu Glu Lys Arg Lys Arg Thr Tyr Glu Thr Phe
      130          135          140

aag agc atc atg aag aag agt cct ttc agc gga ccc acc gac ccc cgg      480
Lys Ser Ile Met Lys Lys Ser Pro Phe Ser Gly Pro Thr Asp Pro Arg
      145          150          155          160

```

cct cca cct cga cgc att gct gtg cct tcc cgc agc tca gct tct gtc Pro Pro Pro Arg Arg Ile Ala Val Pro Ser Arg Ser Ser Ala Ser Val 165 170 175	528
ccc aag cca gca ccc cag ccc tat ccc ttt acg tca tcc ctg agc acc Pro Lys Pro Ala Pro Gln Pro Tyr Pro Phe Thr Ser Ser Leu Ser Thr 180 185 190	576
atc aac tat gat gag ttt ccc acc atg gtg ttt cct tct ggg cag atc Ile Asn Tyr Asp Glu Phe Pro Thr Met Val Phe Pro Ser Gly Gln Ile 195 200 205	624
agc cag gcc tcg gcc ttg gcc ccg gcc cct ccc caa gtc ctg ccc cag Ser Gln Ala Ser Ala Leu Ala Pro Ala Pro Pro Gln Val Leu Pro Gln 210 215 220	672
gct cca gcc cct gcc cct gct cca gcc atg gta tca gct ctg gcc cag Ala Pro Ala Pro Ala Pro Ala Pro Ala Met Val Ser Ala Leu Ala Gln 225 230 235 240	720
gcc cca gcc cct gtc cca gtc cta gcc cca ggc cct cct cag gct gtg Ala Pro Ala Pro Val Pro Val Leu Ala Pro Gly Pro Pro Gln Ala Val 245 250 255	768
gcc cca cct gcc ccc aag ccc acc cag gct ggg gaa gga acg ctg tca Ala Pro Pro Ala Pro Lys Pro Thr Gln Ala Gly Glu Gly Thr Leu Ser 260 265 270	816
gag gcc ctg ctg cag ctg cag ttt gat gat gaa gac ctg ggg gcc ttg Glu Ala Leu Leu Gln Leu Gln Phe Asp Asp Glu Asp Leu Gly Ala Leu 275 280 285	864
ctt ggc aac agc aca gac cca gct gtg ttc aca gac ctg gca tcc gtc Leu Gly Asn Ser Thr Asp Pro Ala Val Phe Thr Asp Leu Ala Ser Val 290 295 300	912
gac aac tcc gag ttt cag cag ctg ctg aac cag ggc ata cct gtg gcc Asp Asn Ser Glu Phe Gln Gln Leu Leu Asn Gln Gly Ile Pro Val Ala 305 310 315 320	960
ccc cac aca act gag ccc atg ctg atg gag tac cct gag gct ata act Pro His Thr Thr Glu Pro Met Leu Met Glu Tyr Pro Glu Ala Ile Thr 325 330 335	1008
cgc cta gtg aca gcc cag agg ccc ccc gac cca gct cct gct cca ctg Arg Leu Val Thr Ala Gln Arg Pro Pro Asp Pro Ala Pro Ala Pro Leu 340 345 350	1056
ggg gcc ccg ggg ctc ccc aat ggc ctc ctt tca gga gat gaa gac ttc Gly Ala Pro Gly Leu Pro Asn Gly Leu Leu Ser Gly Asp Glu Asp Phe 355 360 365	1104
tcc tcc att gcg gac atg gac ttc tca gcc ctg ctg agt cag Ser Ser Ile Ala Asp Met Asp Phe Ser Ala Leu Leu Ser Gln 370 375 380	1146
taa	1149
<210> 8	
<211> 382	
<212> PRT	
<213> 人工序列	
<220>	
<223> 质粒序列	

<400> 8

```

Met Val Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Glu Leu Pro Pro Lys
 1          5          10          15
Lys Lys Arg Lys Val Gly Ile Arg Ile Pro Gly Glu Lys Pro Phe Glu
 20          25          30
Cys Lys Asp Cys Gly Lys Ala Phe Ile Gln Lys Ser Asn Leu Ile Arg
 35          40          45
His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Ala Cys Pro Val Glu
 50          55          60
Ser Cys Asp Arg Arg Phe Ser Asp Ser Ser Asn Leu Thr Arg His Ile
 65          70          75          80
Arg Ile His Thr Gly Glu Lys Pro Tyr Ala Cys Pro Val Glu Ser Cys
 85          90          95
Asp Arg Arg Phe Ser Asp Ser Ser Asn Leu Thr Arg His Ile Arg Ile
 100          105          110
His Thr Gly Glu Lys Ala Ala Ala Lys Phe Tyr Leu Pro Asp Thr Asp
 115          120          125
Asp Arg His Arg Ile Glu Glu Lys Arg Lys Arg Thr Tyr Glu Thr Phe
 130          135          140
Lys Ser Ile Met Lys Lys Ser Pro Phe Ser Gly Pro Thr Asp Pro Arg
 145          150          155          160
Pro Pro Pro Arg Arg Ile Ala Val Pro Ser Arg Ser Ser Ala Ser Val
 165          170          175
Pro Lys Pro Ala Pro Gln Pro Tyr Pro Phe Thr Ser Ser Leu Ser Thr
 180          185          190
Ile Asn Tyr Asp Glu Phe Pro Thr Met Val Phe Pro Ser Gly Gln Ile
 195          200          205
Ser Gln Ala Ser Ala Leu Ala Pro Ala Pro Pro Gln Val Leu Pro Gln
 210          215          220
Ala Pro Ala Pro Ala Pro Ala Pro Ala Met Val Ser Ala Leu Ala Gln
 225          230          235          240
Ala Pro Ala Pro Val Pro Val Leu Ala Pro Gly Pro Pro Gln Ala Val
 245          250          255
Ala Pro Pro Ala Pro Lys Pro Thr Gln Ala Gly Glu Gly Thr Leu Ser
 260          265          270
Glu Ala Leu Leu Gln Leu Gln Phe Asp Asp Glu Asp Leu Gly Ala Leu
 275          280          285
Leu Gly Asn Ser Thr Asp Pro Ala Val Phe Thr Asp Leu Ala Ser Val
 290          295          300
Asp Asn Ser Glu Phe Gln Gln Leu Leu Asn Gln Gly Ile Pro Val Ala
 305          310          315          320
Pro His Thr Thr Glu Pro Met Leu Met Glu Tyr Pro Glu Ala Ile Thr
 325          330          335
Arg Leu Val Thr Ala Gln Arg Pro Pro Asp Pro Ala Pro Ala Pro Leu
 340          345          350
Gly Ala Pro Gly Leu Pro Asn Gly Leu Leu Ser Gly Asp Glu Asp Phe
 355          360          365
Ser Ser Ile Ala Asp Met Asp Phe Ser Ala Leu Leu Ser Gln
 370          375          380

```

<210> 9

<211> 630

<212> DNA

<213> 人工序列

<220>

<223> 质粒序列

<221> CDS

<222> (1)...(627)

<400> 9

```

atg gtg tac ccc tac gac gtg ccc gac tac gcc gaa ttg cct cca aaa
Met Val Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Glu Leu Pro Pro Lys

```

48

1					5					10					15					
aag	aag	aga	aag	gta	ggg	atc	cga	att	ccc	ggg	gaa	aaa	ccg	tat	gag			96		
Lys	Lys	Arg	Lys	Val	Gly	Ile	Arg	Ile	Pro	Gly	Glu	Lys	Pro	Tyr	Glu					
				20					25					30						
tgt	gat	cac	tgt	gga	aaa	tcc	ttt	agc	cag	agc	tct	cat	ctg	aat	gtg			144		
Cys	Asp	His	Cys	Gly	Lys	Ser	Phe	Ser	Gln	Ser	Ser	His	Leu	Asn	Val					
				35					40					45						
cac	aaa	aga	act	cac	acc	ggg	gaa	aaa	ccg	tac	aga	tgt	gag	gaa	tgt			192		
His	Lys	Arg	Thr	His	Thr	Gly	Glu	Lys	Pro	Tyr	Arg	Cys	Glu	Glu	Cys					
				50					55					60						
ggc	aaa	gcc	ttt	agg	tgg	ccc	tca	aac	ctt	act	aga	cat	aag	aga	att			240		
Gly	Lys	Ala	Phe	Arg	Trp	Pro	Ser	Asn	Leu	Thr	Arg	His	Lys	Arg	Ile					
				65					70					75					80	
cac	acc	ggg	gaa	aaa	ccg	tac	aga	tgt	gag	gaa	tgt	ggc	aaa	gcc	ttt			288		
His	Thr	Gly	Glu	Lys	Pro	Tyr	Arg	Cys	Glu	Glu	Cys	Gly	Lys	Ala	Phe					
				85					90					95						
agg	tgg	ccc	tca	aac	ctt	act	aga	cat	aag	aga	att	cac	acc	ggg	gaa			336		
Arg	Trp	Pro	Ser	Asn	Leu	Thr	Arg	His	Lys	Arg	Ile	His	Thr	Gly	Glu					
				100					105					110						
aaa	ccg	ttt	gcc	tgc	cct	gag	tgt	cct	aag	cgc	ttc	atg	aga	tcc	gac			384		
Lys	Pro	Phe	Ala	Cys	Pro	Glu	Cys	Pro	Lys	Arg	Phe	Met	Arg	Ser	Asp					
				115					120					125						
aac	ctg	acc	cag	cat	atc	aag	acc	cac	acc	ggt	gaa	aaa	gcg	gcc	gct			432		
Asn	Leu	Thr	Gln	His	Ile	Lys	Thr	His	Thr	Gly	Glu	Lys	Ala	Ala	Ala					
				130					135					140						
aaa	ttc	gtg	tca	gtg	aca	ttt	gaa	gat	gtg	gct	gtg	ctc	ttt	act	cgg			480		
Lys	Phe	Val	Ser	Val	Thr	Phe	Glu	Asp	Val	Ala	Val	Leu	Phe	Thr	Arg					
				145					150					155					160	
gac	gag	tgg	aag	aag	ctg	gat	ctg	tct	cag	aga	agc	ctg	tac	cgt	gag			528		
Asp	Glu	Trp	Lys	Lys	Leu	Asp	Leu	Ser	Gln	Arg	Ser	Leu	Tyr	Arg	Glu					
				165					170					175						
gtg	atg	ctg	gag	aat	tac	agc	aac	ctg	gcc	tcc	atg	gca	gga	ttc	ctg			576		
Val	Met	Leu	Glu	Asn	Tyr	Ser	Asn	Leu	Ala	Ser	Met	Ala	Gly	Phe	Leu					
				180					185					190						
ttt	acc	aaa	cca	aag	gtg	atc	tcc	ctg	ttg	cag	caa	gga	gag	gat	ccc			624		
Phe	Thr	Lys	Pro	Lys	Val	Ile	Ser	Leu	Leu	Gln	Gln	Gly	Glu	Asp	Pro					
				195					200					205						
tgg	taa																630			
Trp																				
<210> 10																				
<211> 209																				
<212> PRT																				
<213> 人工序列																				
<220>																				
<223> 质粒序列																				
<400> 10																				
Met	Val	Tyr	Pro	Tyr	Asp	Val	Pro	Asp	Tyr	Ala	Glu	Leu	Pro	Pro	Lys					
1				5					10				15							

Lys Lys Arg Lys Val Gly Ile Arg Ile Pro Gly Glu Lys Pro Tyr Glu
 20 25 30
 Cys Asp His Cys Gly Lys Ser Phe Ser Gln Ser Ser His Leu Asn Val
 35 40 45
 His Lys Arg Thr His Thr Gly Glu Lys Pro Tyr Arg Cys Glu Glu Cys
 50 55 60
 Gly Lys Ala Phe Arg Trp Pro Ser Asn Leu Thr Arg His Lys Arg Ile
 65 70 75 80
 His Thr Gly Glu Lys Pro Tyr Arg Cys Glu Glu Cys Gly Lys Ala Phe
 85 90 95
 Arg Trp Pro Ser Asn Leu Thr Arg His Lys Arg Ile His Thr Gly Glu
 100 105 110
 Lys Pro Phe Ala Cys Pro Glu Cys Pro Lys Arg Phe Met Arg Ser Asp
 115 120 125
 Asn Leu Thr Gln His Ile Lys Thr His Thr Gly Glu Lys Ala Ala Ala
 130 135 140
 Lys Phe Val Ser Val Thr Phe Glu Asp Val Ala Val Leu Phe Thr Arg
 145 150 155 160
 Asp Glu Trp Lys Lys Leu Asp Leu Ser Gln Arg Ser Leu Tyr Arg Glu
 165 170 175
 Val Met Leu Glu Asn Tyr Ser Asn Leu Ala Ser Met Ala Gly Phe Leu
 180 185 190
 Phe Thr Lys Pro Lys Val Ile Ser Leu Leu Gln Gln Gly Glu Asp Pro
 195 200 205
 Trp

<210> 11
 <211> 642
 <212> DNA
 <213> 人工序列

<220>
 <223> 质粒序列

<221> CDS
 <222> (1)...(639)

<400> 11
 atg gtg tac ccc tac gac gtg ccc gac tac gcc gaa ttg cct cca aaa 48
 Met Val Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Glu Leu Pro Pro Lys
 1 5 10 15
 aag aag aga aag gta ggg atc cga att ccc ggg gaa aaa ccg tac tcc 96
 Lys Lys Arg Lys Val Gly Ile Arg Ile Pro Gly Glu Lys Pro Tyr Ser
 20 25 30
 tgt ggc att tgt ggc aaa tcc ttc tct gac tcc agt gcc aaa agg aga 144
 Cys Gly Ile Cys Gly Lys Ser Phe Ser Asp Ser Ser Ala Lys Arg Arg
 35 40 45
 cac tgc att cta cac acc ggg gaa aaa ccg tat gta tgc gat gta gag 192
 His Cys Ile Leu His Thr Gly Glu Lys Pro Tyr Val Cys Asp Val Glu
 50 55 60
 gga tgt acg tgg aaa ttt gcc cgc tca gat aag ctc aac aga cac aag 240
 Gly Cys Thr Trp Lys Phe Ala Arg Ser Asp Lys Leu Asn Arg His Lys
 65 70 75 80
 aaa agg cac acc ggg gaa aaa ccg tat gta tgc gat gta gag gga tgt 288
 Lys Arg His Thr Gly Glu Lys Pro Tyr Val Cys Asp Val Glu Gly Cys
 85 90 95
 acg tgg aaa ttt gcc cgc tca gat gag ctc aac aga cac aag aaa agg 336
 Thr Trp Lys Phe Ala Arg Ser Asp Glu Leu Asn Arg His Lys Lys Arg

100	105	110	
cac acc ggg gaa aaa ccg tat	gag tgt cac gat tgc gga aag tcc ttt		384
His Thr Gly Glu Lys Pro Tyr	Glu Cys His Asp Cys Gly Lys Ser Phe		
115	120	125	
agg cag agc acc cac ctc act	cgg cac cgg agg atc cac acc ggt gaa		432
Arg Gln Ser Thr His Leu Thr	Arg His Arg Arg Ile His Thr Gly Glu		
130	135	140	
aaa gcg gcc gct aaa ttc gtg	tca gtg aca ttt gaa gat gtg gct gtg		480
Lys Ala Ala Ala Lys Phe Val	Ser Val Thr Phe Glu Asp Val Ala Val		
145	150	155	160
ctc ttt act cgg gac gag tgg	aag aag ctg gat ctg tct cag aga agc		528
Leu Phe Thr Arg Asp Glu Trp	Lys Lys Leu Asp Leu Ser Gln Arg Ser		
165	170	175	
ctg tac cgt gag gtg atg ctg	gag aat tac agc aac ctg gcc tcc atg		576
Leu Tyr Arg Glu Val Met Leu	Glu Asn Tyr Ser Asn Leu Ala Ser Met		
180	185	190	
gca gga ttc ctg ttt acc aaa	cca aag gtg atc tcc ctg ttg cag caa		624
Ala Gly Phe Leu Phe Thr Lys	Pro Lys Val Ile Ser Leu Leu Gln Gln		
195	200	205	
gga gag gat ccc tgg taa			642
Gly Glu Asp Pro Trp			
210			
<210> 12			
<211> 213			
<212> PRT			
<213> 人工序列			
<220>			
<223> 质粒序列			
<400> 12			
Met Val Tyr Pro Tyr Asp Val	Pro Asp Tyr Ala Glu Leu Pro Pro Lys		
1	5	10	15
Lys Lys Arg Lys Val Gly Ile	Arg Ile Pro Gly Glu Lys Pro Tyr Ser		
20	25	30	
Cys Gly Ile Cys Gly Lys Ser	Phe Ser Asp Ser Ser Ala Lys Arg Arg		
35	40	45	
His Cys Ile Leu His Thr Gly	Glu Lys Pro Tyr Val Cys Asp Val Glu		
50	55	60	
Gly Cys Thr Trp Lys Phe Ala	Arg Ser Asp Lys Leu Asn Arg His Lys		
65	70	75	80
Lys Arg His Thr Gly Glu Lys	Pro Tyr Val Cys Asp Val Glu Gly Cys		
85	90	95	
Thr Trp Lys Phe Ala Arg Ser	Asp Glu Leu Asn Arg His Lys Lys Arg		
100	105	110	
His Thr Gly Glu Lys Pro Tyr	Glu Cys His Asp Cys Gly Lys Ser Phe		
115	120	125	
Arg Gln Ser Thr His Leu Thr	Arg His Arg Arg Ile His Thr Gly Glu		
130	135	140	
Lys Ala Ala Ala Lys Phe Val	Ser Val Thr Phe Glu Asp Val Ala Val		
145	150	155	160
Leu Phe Thr Arg Asp Glu Trp	Lys Lys Leu Asp Leu Ser Gln Arg Ser		
165	170	175	
Leu Tyr Arg Glu Val Met Leu	Glu Asn Tyr Ser Asn Leu Ala Ser Met		
180	185	190	
Ala Gly Phe Leu Phe Thr Lys	Pro Lys Val Ile Ser Leu Leu Gln Gln		
195	200	205	

Gly Glu Asp Pro Trp
210

<210> 13
<211> 630
<212> DNA
<213> 人工序列

<220>
<223> 质粒序列

<221> CDS
<222> (1)...(627)

```

<400> 13
atg gtg tac ccc tac gac gtg ccc gac tac gcc gaa ttg cct cca aaa      48
Met Val Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Glu Leu Pro Pro Lys
 1              5              10              15

aag aag aga aag gta ggg atc cga att ccc ggg gaa aaa ccg tat gag      96
Lys Lys Arg Lys Val Gly Ile Arg Ile Pro Gly Glu Lys Pro Tyr Glu
              20              25              30

tgt gat cac tgt gga aaa tcc ttt agc cag agc tct cat ctg aat gtg      144
Cys Asp His Cys Gly Lys Ser Phe Ser Gln Ser Ser His Leu Asn Val
              35              40              45

cac aaa aga act cac acc ggg gaa aaa ccg tac atg tgc agt gag tgt      192
His Lys Arg Thr His Thr Gly Glu Lys Pro Tyr Met Cys Ser Glu Cys
              50              55              60

ggg cga ggc ttc agc cag aag tca aac ctc atc ata cac cag agg aca      240
Gly Arg Gly Phe Ser Gln Lys Ser Asn Leu Ile Ile His Gln Arg Thr
 65              70              75              80

cac acc ggg gaa aaa ccg tat gag tgt cac gat tgc gga aag tcc ttt      288
His Thr Gly Glu Lys Pro Tyr Glu Cys His Asp Cys Gly Lys Ser Phe
              85              90              95

agg cag agc acc cac ctc act cgg cac cgg agg atc cac acc ggg gaa      336
Arg Gln Ser Thr His Leu Thr Arg His Arg Arg Ile His Thr Gly Glu
              100              105              110

aaa ccg tat aaa tgt aag caa tgt ggg aaa gct ttt gga tgt ccc tca      384
Lys Pro Tyr Lys Cys Lys Gln Cys Gly Lys Ala Phe Gly Cys Pro Ser
              115              120              125

aac ctt cga agg cat gga agg act cac acc ggt gaa aaa gcg gcc gct      432
Asn Leu Arg Arg His Gly Arg Thr His Thr Gly Glu Lys Ala Ala Ala
              130              135              140

aaa ttc gtg tca gtg aca ttt gaa gat gtg gct gtg ctc ttt act cgg      480
Lys Phe Val Ser Val Thr Phe Glu Asp Val Ala Val Leu Phe Thr Arg
145              150              155              160

gac gag tgg aag aag ctg gat ctg tct cag aga agc ctg tac cgt gag      528
Asp Glu Trp Lys Lys Leu Asp Leu Ser Gln Arg Ser Leu Tyr Arg Glu
              165              170              175

gtg atg ctg gag aat tac agc aac ctg gcc tcc atg gca gga ttc ctg      576
Val Met Leu Glu Asn Tyr Ser Asn Leu Ala Ser Met Ala Gly Phe Leu
              180              185              190

ttt acc aaa cca aag gtg atc tcc ctg ttg cag caa gga gag gat ccc      624
Phe Thr Lys Pro Lys Val Ile Ser Leu Leu Gln Gln Gly Glu Asp Pro

```

195	200	205	630
tgg taa Trp			
<210> 14 <211> 209 <212> PRT <213> 人工序列			
<220> <223> 质粒序列			
<400> 14 Met Val Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Glu Leu Pro Pro Lys 1 5 10 15 Lys Lys Arg Lys Val Gly Ile Arg Ile Pro Gly Glu Lys Pro Tyr Glu 20 25 30 Cys Asp His Cys Gly Lys Ser Phe Ser Gln Ser Ser His Leu Asn Val 35 40 45 His Lys Arg Thr His Thr Gly Glu Lys Pro Tyr Met Cys Ser Glu Cys 50 55 60 Gly Arg Gly Phe Ser Gln Lys Ser Asn Leu Ile Ile His Gln Arg Thr 65 70 75 80 His Thr Gly Glu Lys Pro Tyr Glu Cys His Asp Cys Gly Lys Ser Phe 85 90 95 Arg Gln Ser Thr His Leu Thr Arg His Arg Arg Ile His Thr Gly Glu 100 105 110 Lys Pro Tyr Lys Cys Lys Gln Cys Gly Lys Ala Phe Gly Cys Pro Ser 115 120 125 Asn Leu Arg Arg His Gly Arg Thr His Thr Gly Glu Lys Ala Ala Ala 130 135 140 Lys Phe Val Ser Val Thr Phe Glu Asp Val Ala Val Leu Phe Thr Arg 145 150 155 160 Asp Glu Trp Lys Lys Leu Asp Leu Ser Gln Arg Ser Leu Tyr Arg Glu 165 170 175 Val Met Leu Glu Asn Tyr Ser Asn Leu Ala Ser Met Ala Gly Phe Leu 180 185 190 Phe Thr Lys Pro Lys Val Ile Ser Leu Leu Gln Gln Gly Glu Asp Pro 195 200 205 Trp			
<210> 15 <211> 630 <212> DNA <213> 人工序列			
<220> <223> 质粒序列			
<221> CDS <222> (1)...(627)			
<400> 15 atg gtg tac ccc tac gac gtg ccc gac tac gcc gaa ttg cct cca aaa Met Val Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Glu Leu Pro Pro Lys 1 5 10 15 aag aag aga aag gta ggg atc cga att ccc ggg gaa aaa ccg tat gag Lys Lys Arg Lys Val Gly Ile Arg Ile Pro Gly Glu Lys Pro Tyr Glu 20 25 30 tgt gat cac tgt gga aaa tcc ttt agc cag agc tct cat ctg aat gtg 144			

Cys	Asp	His	Cys	Gly	Lys	Ser	Phe	Ser	Gln	Ser	Ser	His	Leu	Asn	Val	
		35					40					45				
cac	aaa	aga	act	cac	acc	ggg	gaa	aaa	ccg	tat	gag	tgt	aat	tac	tgt	192
His	Lys	Arg	Thr	His	Thr	Gly	Glu	Lys	Pro	Tyr	Glu	Cys	Asn	Tyr	Cys	
	50					55					60					
gga	aaa	acc	ttt	agt	gtg	agc	tca	acc	ctt	att	aga	cat	cag	aga	atc	240
Gly	Lys	Thr	Phe	Ser	Val	Ser	Ser	Thr	Leu	Ile	Arg	His	Gln	Arg	Ile	
65					70				75						80	
cac	acc	ggg	gaa	aaa	ccg	ttt	gcc	tgc	cct	gag	tgt	cct	aag	cgc	ttc	288
His	Thr	Gly	Glu	Lys	Pro	Phe	Ala	Cys	Pro	Glu	Cys	Pro	Lys	Arg	Phe	
				85					90					95		
atg	aga	tcc	gac	aac	ctg	acc	cag	cat	atc	aag	acc	cac	acc	ggg	gaa	336
Met	Arg	Ser	Asp	Asn	Leu	Thr	Gln	His	Ile	Lys	Thr	His	Thr	Gly	Glu	
			100					105					110			
aaa	ccg	tat	gag	tgt	cac	gat	tgc	gga	aag	tcc	ttt	agg	cag	agc	acc	384
Lys	Pro	Tyr	Glu	Cys	His	Asp	Cys	Gly	Lys	Ser	Phe	Arg	Gln	Ser	Thr	
	115						120					125				
cac	ctc	act	cgg	cac	cgg	agg	atc	cac	acc	ggt	gaa	aaa	gcg	gcc	gct	432
His	Leu	Thr	Arg	His	Arg	Arg	Ile	His	Thr	Gly	Glu	Lys	Ala	Ala	Ala	
	130					135					140					
aaa	ttc	gtg	tca	gtg	aca	ttt	gaa	gat	gtg	gct	gtg	ctc	ttt	act	cgg	480
Lys	Phe	Val	Ser	Val	Thr	Phe	Glu	Asp	Val	Ala	Val	Leu	Phe	Thr	Arg	
145					150				155						160	
gac	gag	tgg	aag	aag	ctg	gat	ctg	tct	cag	aga	agc	ctg	tac	cgt	gag	528
Asp	Glu	Trp	Lys	Lys	Leu	Asp	Leu	Ser	Gln	Arg	Ser	Leu	Tyr	Arg	Glu	
				165					170					175		
gtg	atg	ctg	gag	aat	tac	agc	aac	ctg	gcc	tcc	atg	gca	gga	ttc	ctg	576
Val	Met	Leu	Glu	Asn	Tyr	Ser	Asn	Leu	Ala	Ser	Met	Ala	Gly	Phe	Leu	
			180					185					190			
ttt	acc	aaa	cca	aag	gtg	atc	tcc	ctg	ttg	cag	caa	gga	gag	gat	ccc	624
Phe	Thr	Lys	Pro	Lys	Val	Ile	Ser	Leu	Leu	Gln	Gln	Gly	Glu	Asp	Pro	
		195					200					205				
tgg	taa															630
Trp																
<210> 16																
<211> 209																
<212> PRT																
<213> 人工序列																
<220>																
<223> 质粒序列																
<400> 16																
Met	Val	Tyr	Pro	Tyr	Asp	Val	Pro	Asp	Tyr	Ala	Glu	Leu	Pro	Pro	Lys	
1				5					10					15		
Lys	Lys	Arg	Lys	Val	Gly	Ile	Arg	Ile	Pro	Gly	Glu	Lys	Pro	Tyr	Glu	
			20					25					30			
Cys	Asp	His	Cys	Gly	Lys	Ser	Phe	Ser	Gln	Ser	Ser	His	Leu	Asn	Val	
		35					40					45				
His	Lys	Arg	Thr	His	Thr	Gly	Glu	Lys	Pro	Tyr	Glu	Cys	Asn	Tyr	Cys	
	50					55					60					
Gly	Lys	Thr	Phe	Ser	Val	Ser	Ser	Thr	Leu	Ile	Arg	His	Gln	Arg	Ile	

```

65          70          75          80
His Thr Gly Glu Lys Pro Phe Ala Cys Pro Glu Cys Pro Lys Arg Phe
85          90          95
Met Arg Ser Asp Asn Leu Thr Gln His Ile Lys Thr His Thr Gly Glu
100        105        110
Lys Pro Tyr Glu Cys His Asp Cys Gly Lys Ser Phe Arg Gln Ser Thr
115        120        125
His Leu Thr Arg His Arg Arg Ile His Thr Gly Glu Lys Ala Ala Ala
130        135        140
Lys Phe Val Ser Val Thr Phe Glu Asp Val Ala Val Leu Phe Thr Arg
145        150        155        160
Asp Glu Trp Lys Lys Leu Asp Leu Ser Gln Arg Ser Leu Tyr Arg Glu
165        170        175
Val Met Leu Glu Asn Tyr Ser Asn Leu Ala Ser Met Ala Gly Phe Leu
180        185        190
Phe Thr Lys Pro Lys Val Ile Ser Leu Leu Gln Gln Gly Glu Asp Pro
195        200        205
Trp

```

<210> 17
 <211> 1137
 <212> DNA
 <213> 人工序列

<220>
 <223> 质粒序列

<221> CDS
 <222> (1)...(1134)

```

<400> 17
atg gtg tac ccc tac gac gtg ccc gac tac gcc gaa ttg cct cca aaa      48
Met Val Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Glu Leu Pro Pro Lys
1          5          10          15

aag aag aga aag gta ggg atc cga att ccc ggg gaa aaa ccg ttc cag      96
Lys Lys Arg Lys Val Gly Ile Arg Ile Pro Gly Glu Lys Pro Phe Gln
20        25        30

tgt aaa act tgt cag cga aag ttc tcc cgg tcc gac cac ctg aag acc      144
Cys Lys Thr Cys Gln Arg Lys Phe Ser Arg Ser Asp His Leu Lys Thr
35        40        45

cac acc agg act cat acc ggg gaa aaa ccg tat aag tgc atg gag tgt      192
His Thr Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Met Glu Cys
50        55        60

ggg aag gct ttt aac cgc agg tca cac ctc aca cgg cac cag cgg att      240
Gly Lys Ala Phe Asn Arg Arg Ser His Leu Thr Arg His Gln Arg Ile
65        70        75        80

cac acc ggg gaa aaa ccg tat aaa tgc ggc cag tgt ggg aag ttc tac      288
His Thr Gly Glu Lys Pro Tyr Lys Cys Gly Gln Cys Gly Lys Phe Tyr
85        90        95

tcg cag gtc tcc cac ctc acc cgc cac cag aaa atc cac acc ggt gaa      336
Ser Gln Val Ser His Leu Thr Arg His Gln Lys Ile His Thr Gly Glu
100       105       110

aaa gcg gcc gct aaa ttc tac ctg cca gat aca gac gat cgt cac cgg      384
Lys Ala Ala Ala Lys Phe Tyr Leu Pro Asp Thr Asp Asp Arg His Arg
115       120       125

att gag gag aaa cgt aaa agg aca tat gag acc ttc aag agc atc atg      432

```

Ile	Glu	Glu	Lys	Arg	Lys	Arg	Thr	Tyr	Glu	Thr	Phe	Lys	Ser	Ile	Met	
130						135					140					
aag	aag	agt	cct	ttc	agc	gga	ccc	acc	gac	ccc	cgg	cct	cca	cct	cga	480
Lys	Lys	Ser	Pro	Phe	Ser	Gly	Pro	Thr	Asp	Pro	Arg	Pro	Pro	Pro	Arg	
145					150					155					160	
cgc	att	gct	gtg	cct	tcc	cgc	agc	tca	gct	tct	gtc	ccc	aag	cca	gca	528
Arg	Ile	Ala	Val	Pro	Ser	Arg	Ser	Ser	Ala	Ser	Val	Pro	Lys	Pro	Ala	
				165					170					175		
ccc	cag	ccc	tat	ccc	ttt	acg	tca	tcc	ctg	agc	acc	atc	aac	tat	gat	576
Pro	Gln	Pro	Tyr	Pro	Phe	Thr	Ser	Ser	Leu	Ser	Thr	Ile	Asn	Tyr	Asp	
			180					185					190			
gag	ttt	ccc	acc	atg	gtg	ttt	cct	tct	ggg	cag	atc	agc	cag	gcc	tcg	624
Glu	Phe	Pro	Thr	Met	Val	Phe	Pro	Ser	Gly	Gln	Ile	Ser	Gln	Ala	Ser	
		195					200					205				
gcc	ttg	gcc	ccg	gcc	cct	ccc	caa	gtc	ctg	ccc	cag	gct	cca	gcc	cct	672
Ala	Leu	Ala	Pro	Ala	Pro	Pro	Gln	Val	Leu	Pro	Gln	Ala	Pro	Ala	Pro	
	210					215					220					
gcc	cct	gct	cca	gcc	atg	gta	tca	gct	ctg	gcc	cag	gcc	cca	gcc	cct	720
Ala	Pro	Ala	Pro	Ala	Met	Val	Ser	Ala	Leu	Ala	Gln	Ala	Pro	Ala	Pro	
	225				230					235					240	
gtc	cca	gtc	cta	gcc	cca	ggc	cct	cct	cag	gct	gtg	gcc	cca	cct	gcc	768
Val	Pro	Val	Leu	Ala	Pro	Gly	Pro	Pro	Gln	Ala	Val	Ala	Pro	Pro	Ala	
				245					250					255		
ccc	aag	ccc	acc	cag	gct	ggg	gaa	gga	acg	ctg	tca	gag	gcc	ctg	ctg	816
Pro	Lys	Pro	Thr	Gln	Ala	Gly	Glu	Gly	Thr	Leu	Ser	Glu	Ala	Leu	Leu	
			260				265						270			
cag	ctg	cag	ttt	gat	gat	gaa	gac	ctg	ggg	gcc	ttg	ctt	ggc	aac	agc	864
Gln	Leu	Gln	Phe	Asp	Asp	Glu	Asp	Leu	Gly	Ala	Leu	Leu	Gly	Asn	Ser	
			275				280					285				
aca	gac	cca	gct	gtg	ttc	aca	gac	ctg	gca	tcc	gtc	gac	aac	tcc	gag	912
Thr	Asp	Pro	Ala	Val	Phe	Thr	Asp	Leu	Ala	Ser	Val	Asp	Asn	Ser	Glu	
	290					295					300					
ttt	cag	cag	ctg	ctg	aac	cag	ggc	ata	cct	gtg	gcc	ccc	cac	aca	act	960
Phe	Gln	Gln	Leu	Leu	Asn	Gln	Gly	Ile	Pro	Val	Ala	Pro	His	Thr	Thr	
	305				310				315						320	
gag	ccc	atg	ctg	atg	gag	tac	cct	gag	gct	ata	act	cgc	cta	gtg	aca	1008
Glu	Pro	Met	Leu	Met	Glu	Tyr	Pro	Glu	Ala	Ile	Thr	Arg	Leu	Val	Thr	
				325					330					335		
gcc	cag	agg	ccc	ccc	gac	cca	gct	cct	gct	cca	ctg	ggg	gcc	ccg	ggg	1056
Ala	Gln	Arg	Pro	Pro	Asp	Pro	Ala	Pro	Ala	Pro	Leu	Gly	Ala	Pro	Gly	
			340					345					350			
ctc	ccc	aat	ggc	ctc	ctt	tca	gga	gat	gaa	gac	ttc	tcc	tcc	att	gcg	1104
Leu	Pro	Asn	Gly	Leu	Leu	Ser	Gly	Asp	Glu	Asp	Phe	Ser	Ser	Ile	Ala	
		355					360					365				
gac	atg	gac	ttc	tca	gcc	ctg	ctg	agt	cag	taa						1137
Asp	Met	Asp	Phe	Ser	Ala	Leu	Leu	Ser	Gln							
	370					375										

<210> 18

<211> 378

<212> PRT
<213> 人工序列

<220>
<223> 质粒序列

<400> 18
Met Val Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Glu Leu Pro Pro Lys
1 5 10 15
Lys Lys Arg Lys Val Gly Ile Arg Ile Pro Gly Glu Lys Pro Phe Gln
20 25 30
Cys Lys Thr Cys Gln Arg Lys Phe Ser Arg Ser Asp His Leu Lys Thr
35 40 45
His Thr Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Met Glu Cys
50 55 60
Gly Lys Ala Phe Asn Arg Arg Ser His Leu Thr Arg His Gln Arg Ile
65 70 75 80
His Thr Gly Glu Lys Pro Tyr Lys Cys Gly Gln Cys Gly Lys Phe Tyr
85 90 95
Ser Gln Val Ser His Leu Thr Arg His Gln Lys Ile His Thr Gly Glu
100 105 110
Lys Ala Ala Ala Lys Phe Tyr Leu Pro Asp Thr Asp Asp Arg His Arg
115 120 125
Ile Glu Glu Lys Arg Lys Arg Thr Tyr Glu Thr Phe Lys Ser Ile Met
130 135 140
Lys Lys Ser Pro Phe Ser Gly Pro Thr Asp Pro Arg Pro Pro Pro Arg
145 150 155 160
Arg Ile Ala Val Pro Ser Arg Ser Ser Ala Ser Val Pro Lys Pro Ala
165 170 175
Pro Gln Pro Tyr Pro Phe Thr Ser Ser Leu Ser Thr Ile Asn Tyr Asp
180 185 190
Glu Phe Pro Thr Met Val Phe Pro Ser Gly Gln Ile Ser Gln Ala Ser
195 200 205
Ala Leu Ala Pro Ala Pro Pro Gln Val Leu Pro Gln Ala Pro Ala Pro
210 215 220
Ala Pro Ala Pro Ala Met Val Ser Ala Leu Ala Gln Ala Pro Ala Pro
225 230 235 240
Val Pro Val Leu Ala Pro Gly Pro Pro Gln Ala Val Ala Pro Pro Ala
245 250 255
Pro Lys Pro Thr Gln Ala Gly Glu Gly Thr Leu Ser Glu Ala Leu Leu
260 265 270
Gln Leu Gln Phe Asp Asp Glu Asp Leu Gly Ala Leu Leu Gly Asn Ser
275 280 285
Thr Asp Pro Ala Val Phe Thr Asp Leu Ala Ser Val Asp Asn Ser Glu
290 295 300
Phe Gln Gln Leu Leu Asn Gln Gly Ile Pro Val Ala Pro His Thr Thr
305 310 315 320
Glu Pro Met Leu Met Glu Tyr Pro Glu Ala Ile Thr Arg Leu Val Thr
325 330 335
Ala Gln Arg Pro Pro Asp Pro Ala Pro Ala Pro Leu Gly Ala Pro Gly
340 345 350
Leu Pro Asn Gly Leu Leu Ser Gly Asp Glu Asp Phe Ser Ser Ile Ala
355 360 365
Asp Met Asp Phe Ser Ala Leu Leu Ser Gln
370 375

<210> 19
<211> 1137
<212> DNA
<213> 人工序列

<220>
<223> 质粒序列

<221> CDS

<222> (1)...(1134)

<400> 19

atg	gtg	tac	ccc	tac	gac	gtg	ccc	gac	tac	gcc	gaa	ttg	cct	cca	aaa	48
Met	Val	Tyr	Pro	Tyr	Asp	Val	Pro	Asp	Tyr	Ala	Glu	Leu	Pro	Pro	Lys	
1				5					10					15		
aag	aag	aga	aag	gta	ggg	atc	cga	att	ccc	ggg	gaa	aaa	ccg	tac	aaa	96
Lys	Lys	Arg	Lys	Val	Gly	Ile	Arg	Ile	Pro	Gly	Glu	Lys	Pro	Tyr	Lys	
			20					25					30			
tgt	gaa	gaa	tgt	ggc	aaa	gcc	ttt	agg	cag	tcc	tca	cac	ctt	act	aca	144
Cys	Glu	Glu	Cys	Gly	Lys	Ala	Phe	Arg	Gln	Ser	Ser	His	Leu	Thr	Thr	
			35				40					45				
cat	aag	ata	att	cat	acc	ggg	gaa	aaa	ccg	tat	aag	tgc	atg	gag	tgt	192
His	Lys	Ile	Ile	His	Thr	Gly	Glu	Lys	Pro	Tyr	Lys	Cys	Met	Glu	Cys	
	50					55					60					
ggg	aag	gct	ttt	aac	cgc	agg	tca	cac	ctc	aca	cgg	cac	cag	cgg	att	240
Gly	Lys	Ala	Phe	Asn	Arg	Arg	Ser	His	Leu	Thr	Arg	His	Gln	Arg	Ile	
65					70				75						80	
cac	acc	ggg	gaa	aaa	ccg	ttc	cag	tgt	aaa	act	tgt	cag	cga	aag	ttc	288
His	Thr	Gly	Glu	Lys	Pro	Phe	Gln	Cys	Lys	Thr	Cys	Gln	Arg	Lys	Phe	
				85				90						95		
tcc	cgg	tcc	gac	cac	ctg	aag	acc	cac	acc	agg	act	cat	acc	ggt	gaa	336
Ser	Arg	Ser	Asp	His	Leu	Lys	Thr	His	Thr	Arg	Thr	His	Thr	Gly	Glu	
			100					105					110			
aaa	gcg	gcc	gct	aaa	ttc	tac	ctg	cca	gat	aca	gac	gat	cgt	cac	cgg	384
Lys	Ala	Ala	Ala	Lys	Phe	Tyr	Leu	Pro	Asp	Thr	Asp	Asp	Arg	His	Arg	
	115						120					125				
att	gag	gag	aaa	cgt	aaa	agg	aca	tat	gag	acc	ttc	aag	agc	atc	atg	432
Ile	Glu	Glu	Lys	Arg	Lys	Arg	Thr	Tyr	Glu	Thr	Phe	Lys	Ser	Ile	Met	
	130					135					140					
aag	aag	agt	cct	ttc	agc	gga	ccc	acc	gac	ccc	cgg	cct	cca	cct	cga	480
Lys	Lys	Ser	Pro	Phe	Ser	Gly	Pro	Thr	Asp	Pro	Arg	Pro	Pro	Pro	Arg	
145				150					155						160	
cgc	att	gct	gtg	cct	tcc	cgc	agc	tca	gct	tct	gtc	ccc	aag	cca	gca	528
Arg	Ile	Ala	Val	Pro	Ser	Arg	Ser	Ser	Ala	Ser	Val	Pro	Lys	Pro	Ala	
				165					170					175		
ccc	cag	ccc	tat	ccc	ttt	acg	tca	tcc	ctg	agc	acc	atc	aac	tat	gat	576
Pro	Gln	Pro	Tyr	Pro	Phe	Thr	Ser	Ser	Leu	Ser	Thr	Ile	Asn	Tyr	Asp	
			180					185					190			
gag	ttt	ccc	acc	atg	gtg	ttt	cct	tct	ggg	cag	atc	agc	cag	gcc	tcg	624
Glu	Phe	Pro	Thr	Met	Val	Phe	Pro	Ser	Gly	Gln	Ile	Ser	Gln	Ala	Ser	
	195						200					205				
gcc	ttg	gcc	ccg	gcc	cct	ccc	caa	gtc	ctg	ccc	cag	gct	cca	gcc	cct	672
Ala	Leu	Ala	Pro	Ala	Pro	Pro	Gln	Val	Leu	Pro	Gln	Ala	Pro	Ala	Pro	
	210					215					220					
gcc	cct	gct	cca	gcc	atg	gta	tca	gct	ctg	gcc	cag	gcc	cca	gcc	cct	720
Ala	Pro	Ala	Pro	Ala	Met	Val	Ser	Ala	Leu	Ala	Gln	Ala	Pro	Ala	Pro	
225					230					235					240	
gtc	cca	gtc	cta	gcc	cca	ggc	cct	cct	cag	gct	gtg	gcc	cca	cct	gcc	768
Val	Pro	Val	Leu	Ala	Pro	Gly	Pro	Pro	Gln	Ala	Val	Ala	Pro	Pro	Ala	

245	250	255	
ccc aag ccc acc cag gct ggg gaa gga acg ctg tca gag gcc ctg ctg Pro Lys Pro Thr Gln Ala Gly Glu Gly Thr Leu Ser Glu Ala Leu Leu 260 265 270			816
cag ctg cag ttt gat gat gaa gac ctg ggg gcc ttg ctt ggc aac agc Gln Leu Gln Phe Asp Asp Glu Asp Leu Gly Ala Leu Leu Gly Asn Ser 275 280 285			864
aca gac cca gct gtg ttc aca gac ctg gca tcc gtc gac aac tcc gag Thr Asp Pro Ala Val Phe Thr Asp Leu Ala Ser Val Asp Asn Ser Glu 290 295 300			912
ttt cag cag ctg ctg aac cag ggc ata cct gtg gcc ccc cac aca act Phe Gln Gln Leu Leu Asn Gln Gly Ile Pro Val Ala Pro His Thr Thr 305 310 315 320			960
gag ccc atg ctg atg gag tac cct gag gct ata act cgc cta gtg aca Glu Pro Met Leu Met Glu Tyr Pro Glu Ala Ile Thr Arg Leu Val Thr 325 330 335			1008
gcc cag agg ccc ccc gac cca gct cct gct cca ctg ggg gcc ccg ggg Ala Gln Arg Pro Pro Asp Pro Ala Pro Ala Pro Leu Gly Ala Pro Gly 340 345 350			1056
ctc ccc aat ggc ctc ctt tca gga gat gaa gac ttc tcc tcc att gcg Leu Pro Asn Gly Leu Leu Ser Gly Asp Glu Asp Phe Ser Ser Ile Ala 355 360 365			1104
gac atg gac ttc tca gcc ctg ctg agt cag taa Asp Met Asp Phe Ser Ala Leu Leu Ser Gln 370 375			1137
<210> 20			
<211> 378			
<212> PRT			
<213> 人工序列			
<220>			
<223> 质粒序列			
<400> 20			
Met Val Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Glu Leu Pro Pro Lys 1 5 10 15			
Lys Lys Arg Lys Val Gly Ile Arg Ile Pro Gly Glu Lys Pro Tyr Lys 20 25 30			
Cys Glu Glu Cys Gly Lys Ala Phe Arg Gln Ser Ser His Leu Thr Thr 35 40 45			
His Lys Ile Ile His Thr Gly Glu Lys Pro Tyr Lys Cys Met Glu Cys 50 55 60			
Gly Lys Ala Phe Asn Arg Arg Ser His Leu Thr Arg His Gln Arg Ile 65 70 75 80			
His Thr Gly Glu Lys Pro Phe Gln Cys Lys Thr Cys Gln Arg Lys Phe 85 90 95			
Ser Arg Ser Asp His Leu Lys Thr His Thr Arg Thr His Thr Gly Glu 100 105 110			
Lys Ala Ala Ala Lys Phe Tyr Leu Pro Asp Thr Asp Asp Arg His Arg 115 120 125			
Ile Glu Glu Lys Arg Lys Arg Thr Tyr Glu Thr Phe Lys Ser Ile Met 130 135 140			
Lys Lys Ser Pro Phe Ser Gly Pro Thr Asp Pro Arg Pro Pro Pro Arg 145 150 155 160			
Arg Ile Ala Val Pro Ser Arg Ser Ser Ala Ser Val Pro Lys Pro Ala 165 170 175			

```

Pro Gln Pro Tyr Pro Phe Thr Ser Ser Leu Ser Thr Ile Asn Tyr Asp
      180      185      190
Glu Phe Pro Thr Met Val Phe Pro Ser Gly Gln Ile Ser Gln Ala Ser
      195      200      205
Ala Leu Ala Pro Ala Pro Pro Gln Val Leu Pro Gln Ala Pro Ala Pro
      210      215      220
Ala Pro Ala Pro Ala Met Val Ser Ala Leu Ala Gln Ala Pro Ala Pro
225      230      235      240
Val Pro Val Leu Ala Pro Gly Pro Pro Gln Ala Val Ala Pro Pro Ala
      245      250      255
Pro Lys Pro Thr Gln Ala Gly Glu Gly Thr Leu Ser Glu Ala Leu Leu
      260      265      270
Gln Leu Gln Phe Asp Asp Glu Asp Leu Gly Ala Leu Leu Gly Asn Ser
      275      280      285
Thr Asp Pro Ala Val Phe Thr Asp Leu Ala Ser Val Asp Asn Ser Glu
      290      295      300
Phe Gln Gln Leu Leu Asn Gln Gly Ile Pro Val Ala Pro His Thr Thr
305      310      315      320
Glu Pro Met Leu Met Glu Tyr Pro Glu Ala Ile Thr Arg Leu Val Thr
      325      330      335
Ala Gln Arg Pro Pro Asp Pro Ala Pro Ala Pro Leu Gly Ala Pro Gly
      340      345      350
Leu Pro Asn Gly Leu Leu Ser Gly Asp Glu Asp Phe Ser Ser Ile Ala
      355      360      365
Asp Met Asp Phe Ser Ala Leu Leu Ser Gln
      370      375

```

<210> 21
 <211> 83
 <212> PRT
 <213> 人工序列

<220>
 <223> 质粒序列

```

<400> 21
Phe Glu Cys Lys Asp Cys Gly Lys Ala Phe Ile Gln Lys Ser Asn Leu
 1      5      10      15
Ile Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Ala Cys Pro
      20      25      30
Val Glu Ser Cys Asp Arg Arg Phe Ser Asp Ser Ser Asn Leu Thr Arg
      35      40      45
His Ile Arg Ile His Thr Gly Glu Lys Pro Tyr Ala Cys Pro Val Glu
      50      55      60
Ser Cys Asp Arg Arg Phe Ser Asp Ser Ser Asn Leu Thr Arg His Ile
      65      70      75      80
Arg Ile His

```

<210> 22
 <211> 110
 <212> PRT
 <213> 人工序列

<220>
 <223> 质粒序列

```

<400> 22
Ser Cys Gly Ile Cys Gly Lys Ser Phe Ser Asp Ser Ser Ala Lys Arg
 1      5      10      15
Arg His Cys Ile Leu His Thr Gly Glu Lys Pro Tyr Val Cys Asp Val
      20      25      30
Glu Gly Cys Thr Trp Lys Phe Ala Arg Ser Asp Lys Leu Asn Arg His
      35      40      45
Lys Lys Arg His Thr Gly Glu Lys Pro Tyr Val Cys Asp Val Glu Gly

```

```

      50              55              60
Cys Thr Trp Lys Phe Ala Arg Ser Asp Glu Leu Asn Arg His Lys Lys
65              70              75              80
Arg His Thr Gly Glu Lys Pro Tyr Glu Cys His Asp Cys Gly Lys Ser
      85              90              95
Phe Arg Gln Ser Thr His Leu Thr Arg His Arg Arg Ile His
      100              105              110

```

<210> 23
 <211> 107
 <212> PRT
 <213> 人工序列

<220>
 <223> 质粒序列

```

<400> 23
Tyr Glu Cys Asp His Cys Gly Lys Ser Phe Ser Gln Ser Ser His Leu
1              5              10              15
Asn Val His Lys Arg Thr His Thr Gly Glu Lys Pro Tyr Arg Cys Glu
      20              25              30
Glu Cys Gly Lys Ala Phe Arg Trp Pro Ser Asn Leu Thr Arg His Lys
      35              40              45
Arg Ile His Thr Gly Glu Lys Pro Tyr Arg Cys Glu Glu Cys Gly Lys
      50              55              60
Ala Phe Arg Trp Pro Ser Asn Leu Thr Arg His Lys Arg Ile His Thr
65              70              75              80
Gly Glu Lys Pro Phe Ala Cys Pro Glu Cys Pro Lys Arg Phe Met Arg
      85              90              95
Ser Asp Asn Leu Thr Gln His Ile Lys Thr His
      100              105

```

<210> 24
 <211> 26
 <212> PRT
 <213> 人工序列

<220>
 <223> 纯化的多肽

<221> 变体
 <222> (1)...(26)
 <223> Xaa = 任何氨基酸

```

<400> 24
Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa Ser Asn
1              5              10              15
Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
      20              25

```

<210> 25
 <211> 26
 <212> PRT
 <213> 人工序列

<220>
 <223> 纯化的多肽

<221> 变体
 <222> (1)...(26)
 <223> Xaa = 任何氨基酸

```

<400> 25
Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Asp Xaa Ser Asn
1              5              10              15

```

Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
20 25

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

<220>
<223> 纯化的多肽

<221> 变体
<222> (1)...(26)
<223> Xaa = 任何氨基酸

<400> 26
Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Asp Xaa Ser Asn
1 5 10 15
Xaa Xaa Arg His Xaa Xaa Xaa Xaa His
20 25

<210> 27
<211> 26
<212> PRT
<213> 人工序列

<220>
<223> 纯化的多肽

<221> 变体
<222> (1)...(26)
<223> Xaa = 任何氨基酸

<400> 27
Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Asp Xaa Ser Ala
1 5 10 15
Xaa Xaa Arg His Xaa Xaa Xaa Xaa His
20 25

<210> 28
<211> 26
<212> PRT
<213> 人工序列

<220>
<223> 纯化的多肽

<221> 变体
<222> (1)...(26)
<223> Xaa = 任何氨基酸

<400> 28
Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Arg Xaa Asp Lys
1 5 10 15
Xaa Xaa Arg His Xaa Xaa Xaa Xaa His
20 25

<210> 29
<211> 26
<212> PRT
<213> 人工序列

<220>
<223> 纯化的多肽

<221> 变体
 <222> (1)...(26)
 <223> Xaa = 任何氨基酸

<400> 29
 Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Arg Xaa Asp Glu
 1 5 10 15
 Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
 20 25

<210> 30
 <211> 26
 <212> PRT
 <213> 人工序列

<220>
 <223> 纯化的多肽

<221> 变体
 <222> (1)...(26)
 <223> Xaa = 任何氨基酸

<400> 30
 Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa Thr His
 1 5 10 15
 Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
 20 25

<210> 31
 <211> 26
 <212> PRT
 <213> 人工序列

<220>
 <223> 纯化的多肽

<221> 变体
 <222> (1)...(26)
 <223> Xaa = 任何氨基酸

<400> 31
 Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa Ser His
 1 5 10 15
 Xaa Xaa Val His Xaa Xaa Xaa Xaa Xaa His
 20 25

<210> 32
 <211> 26
 <212> PRT
 <213> 人工序列

<220>
 <223> 纯化的多肽

<221> 变体
 <222> (1)...(26)
 <223> Xaa = 任何氨基酸

<400> 32
 Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Trp Xaa Ser Asn
 1 5 10 15
 Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
 20 25

<210> 33

<211> 26
 <212> PRT
 <213> 人工序列

<220>
 <223> 纯化的多肽

<221> 变体
 <222> (1)...(26)
 <223> Xaa = 任何氨基酸

<400> 33
 Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Trp Xaa Ser Asn
 1 5 10 15
 Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
 20 25

<210> 34
 <211> 26
 <212> PRT
 <213> 人工序列

<220>
 <223> 纯化的多肽

<221> 变体
 <222> (1)...(26)
 <223> Xaa = 任何氨基酸

<400> 34
 Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Arg Xaa Asp Asn
 1 5 10 15
 Xaa Xaa Lys His Xaa Xaa Xaa Xaa Xaa His
 20 25

<210> 35
 <211> 96
 <212> PRT
 <213> 人工序列

<220>
 <223> 纯化的多肽

<221> 变体
 <222> (1)...(96)
 <223> Xaa = 任何氨基酸

<400> 35
 Xaa Xaa Cys Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Xaa Xaa
 1 5 10 15
 Xaa Xaa Xaa Xaa Arg His Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa
 20 25 30
 Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa
 35 40 45
 Xaa Xaa Xaa Xaa Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa
 50 55 60
 Xaa Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa
 65 70 75 80
 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
 85 90 95

<210> 36
 <211> 94
 <212> PRT
 <213> 人工序列

<220>

<223> 纯化的多肽

<221> 变体

<222> (1)...(94)

<223> Xaa = 任何氨基酸

<400> 36

```

Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa Ser Asn
 1             5             10             15
Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
          20             25             30
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa
          35             40             45
Ser Asn Xaa Xaa Lys His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa
 50             55             60
Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa
65             70             75             80
Cys Xaa Ser Asn Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa Xaa His
          85             90

```

<210> 37

<211> 60

<212> PRT

<213> 人工序列

<220>

<223> 纯化的多肽

<221> 变体

<222> (1)...(60)

<223> Xaa = 任何氨基酸

<400> 37

```

Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa Ser Asn
 1             5             10             15
Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
          20             25             30
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa
          35             40             45
Ser Asn Xaa Xaa Lys His Xaa Xaa Xaa Xaa Xaa His
 50             55             60

```

<210> 38

<211> 60

<212> PRT

<213> 人工序列

<220>

<223> 纯化的多肽

<221> 变体

<222> (1)...(60)

<223> Xaa = 任何氨基酸

<400> 38

```

Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa Ser Asn
 1             5             10             15
Xaa Xaa Lys His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
          20             25             30
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa
          35             40             45
Ser Asn Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
 50             55             60

```

<210> 39
 <211> 128
 <212> PRT
 <213> 人工序列

<220>
 <223> 纯化的多肽

<221> 变体
 <222> (1)...(128)
 <223> Xaa = 任何氨基酸

<400> 39
 Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Arg Xaa Asp Lys
 1 5 10 15
 Xaa Xaa Arg His Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
 20 25 30
 Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa
 35 40 45
 Thr His Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa
 50 55 60
 Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa
 65 70 75 80
 Val Xaa Ser Thr Xaa Xaa Arg His Xaa Xaa Xaa Xaa His Xaa Xaa
 85 90 95
 Xaa Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa
 100 105 110
 Xaa Xaa Arg Xaa Asp Lys Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
 115 120 125

<210> 40
 <211> 94
 <212> PRT
 <213> 人工序列

<220>
 <223> 纯化的多肽

<221> 变体
 <222> (1)...(94)
 <223> Xaa = 任何氨基酸

<400> 40
 Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Arg Xaa Asp Lys
 1 5 10 15
 Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
 20 25 30
 Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa
 35 40 45
 Thr His Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa
 50 55 60
 Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa
 65 70 75 80
 Val Xaa Ser Thr Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
 85 90

<210> 41
 <211> 94
 <212> PRT
 <213> 人工序列

<220>
 <223> 纯化的多肽

<221> 变体
 <222> (1)...(94)
 <223> Xaa = 任何氨基酸

<400> 41
 Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa Thr His
 1 5 10 15
 Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa
 20 25 30
 Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Val Xaa
 35 40 45
 Ser Thr Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa
 50 55 60
 Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa
 65 70 75 80
 Arg Xaa Asp Lys Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
 85 90

<210> 42
 <211> 60
 <212> PRT
 <213> 人工序列

<220>
 <223> 纯化的多肽

<221> 变体
 <222> (1)...(60)
 <223> Xaa = 任何氨基酸

<400> 42
 Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Arg Xaa Asp Lys
 1 5 10 15
 Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa
 20 25 30
 Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Gln Xaa
 35 40 45
 Thr His Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
 50 55 60

<210> 43
 <211> 60
 <212> PRT
 <213> 人工序列

<220>
 <223> 纯化的多肽

<221> 变体
 <222> (1)...(60)
 <223> Xaa = 任何氨基酸

<400> 43
 Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Val Xaa Ser Thr
 1 5 10 15
 Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa
 20 25 30
 Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Arg Xaa
 35 40 45
 Asp Lys Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
 50 55 60

<210> 44
 <211> 94
 <212> PRT

<213> 人工序列

<220>

<223> 纯化的多肽

<221> 变体

<222> (1)...(94)

<223> Xaa = 任何氨基酸

<400> 44

```

Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Arg Xaa Ser His
 1          5          10          15
Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
 20          25          30
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Arg Xaa
 35          40          45
Asp His Xaa Xaa Thr His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa
 50          55          60
Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Xaa
 65          70          75          80
Val Xaa Ser Ser Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa Xaa His
      85          90

```

<210> 45

<211> 60

<212> PRT

<213> 人工序列

<220>

<223> 纯化的多肽

<221> 变体

<222> (1)...(60)

<223> Xaa = 任何氨基酸

<400> 45

```

Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Arg Xaa Ser His
 1          5          10          15
Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
 20          25          30
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Arg Xaa
 35          40          45
Asp His Xaa Xaa Thr His Xaa Xaa Xaa Xaa Xaa His
 50          55          60

```

<210> 46

<211> 60

<212> PRT

<213> 人工序列

<220>

<223> 纯化的多肽

<221> 变体

<222> (1)...(60)

<223> Xaa = 任何氨基酸

<400> 46

```

Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Arg Xaa Asp His
 1          5          10          15
Xaa Xaa Thr His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
 20          25          30
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Val Xaa
 35          40          45
Ser Ser Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa Xaa His

```

50 55 60

<210> 47
<211> 94
<212> PRT
<213> 人工序列

<220>
<223> 纯化的多肽

<221> 变体
<222> (1)...(94)
<223> Xaa = 任何氨基酸

<400> 47
Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa Ser Asn
1 5 10 15
Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
20 25 30
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Asp Xaa
35 40 45
Ser Asn Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa
50 55 60
Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Xaa
65 70 75 80
Asp Xaa Ser Asn Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
85 90

<210> 48
<211> 60
<212> PRT
<213> 人工序列

<220>
<223> 纯化的多肽

<221> 变体
<222> (1)...(60)
<223> Xaa = 任何氨基酸

<400> 48
Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa Ser Asn
1 5 10 15
Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
20 25 30
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Asp Xaa
35 40 45
Ser Asn Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
50 55 60

<210> 49
<211> 60
<212> PRT
<213> 人工序列

<220>
<223> 纯化的多肽

<221> 变体
<222> (1)...(60)
<223> Xaa = 任何氨基酸

<400> 49
Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Asp Xaa Ser Asn
1 5 10 15

```

Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
      20      25      30
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Asp Xaa
      35      40      45
Ser Asn Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
      50      55      60

```

<210> 50
 <211> 128
 <212> PRT
 <213> 人工序列

<220>
 <223> 纯化的多肽

<221> 变体
 <222> (1)...(128)
 <223> Xaa = 任何氨基酸

```

<400> 50
Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa Ser His
 1      5      10      15
Xaa Xaa Val His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
      20      25      30
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Trp Xaa
      35      40      45
Ser Asn Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa
      50      55      60
Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa
65      70      75      80
Trp Xaa Ser Asn Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa
      85      90      95
Xaa Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa
      100      105      110
Xaa Xaa Arg Xaa Asp Asn Xaa Xaa Gln His Xaa Xaa Xaa Xaa Xaa His
      115      120      125

```

<210> 51
 <211> 94
 <212> PRT
 <213> 人工序列

<220>
 <223> 纯化的多肽

<221> 变体
 <222> (1)...(94)
 <223> Xaa = 任何氨基酸

```

<400> 51
Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Trp Xaa Ser Asn
 1      5      10      15
Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
      20      25      30
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Trp Xaa
      35      40      45
Ser Asn Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa
      50      55      60
Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa
65      70      75      80
Arg Xaa Asp Asn Xaa Xaa Gln His Xaa Xaa Xaa Xaa Xaa His
      85      90

```

<210> 52
 <211> 94

<212> PRT

<213> 人工序列

<220>

<223> 纯化的多肽

<221> 变体

<222> (1)...(94)

<223> Xaa = 任何氨基酸

<400> 52

Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Gln	Xaa	Ser	His
1				5					10					15	
Xaa	Xaa	Val	His	Xaa	Xaa	Xaa	Xaa	Xaa	His	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa
			20						25					30	
Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Trp	Xaa
		35						40					45		
Ser	Asn	Xaa	Xaa	Arg	His	Xaa	Xaa	Xaa	Xaa	His	Xaa	Xaa	Xaa	Xaa	Xaa
	50					55				60					
Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa
65				70						75					80
Trp	Xaa	Ser	Asn	Xaa	Xaa	Arg	His	Xaa	Xaa	Xaa	Xaa	Xaa	His		
			85						90						

<210> 53

<211> 128

<212> PRT

<213> 人工序列

<220>

<223> 纯化的多肽

<221> 变体

<222> (1)...(128)

<223> Xaa = 任何氨基酸

<400> 53

Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Asp	Xaa	Ser	Ala
1				5					10					15	
Xaa	Xaa	Arg	His	Xaa	Xaa	Xaa	Xaa	Xaa	His	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa
			20						25					30	
Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Arg	Xaa
		35						40					45		
Asp	Lys	Xaa	Xaa	Arg	His	Xaa	Xaa	Xaa	Xaa	His	Xaa	Xaa	Xaa	Xaa	Xaa
	50					55				60					
Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa
65				70						75					80
Arg	Xaa	Asp	Glu	Xaa	Xaa	Arg	His	Xaa	Xaa	Xaa	Xaa	Xaa	His	Xaa	Xaa
			85						90					95	
Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa
			100						105					110	
Xaa	Xaa	Gln	Xaa	Thr	His	Xaa	Xaa	Arg	His	Xaa	Xaa	Xaa	Xaa	Xaa	His
		115						120					125		

<210> 54

<211> 93

<212> PRT

<213> 人工序列

<220>

<223> 纯化的多肽

<221> 变体

<222> (1)...(93)

<223> Xaa = 任何氨基酸

<400> 54

```

Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Arg Xaa Asp Lys
 1           5           10           15
Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
 20           25           30
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Arg Xaa
 35           40           45
Asp Glu Xaa Xaa Arg His Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa
 50           55           60
Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa
 65           70           75           80
Gln Xaa Thr His Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa
 85           90

```

<210> 55

<211> 94

<212> PRT

<213> 人工序列

<220>

<223> 纯化的多肽

<221> 变体

<222> (1)...(94)

<223> Xaa = 任何氨基酸

<400> 55

```

Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Asp Xaa Ser Ala
 1           5           10           15
Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
 20           25           30
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Arg Xaa
 35           40           45
Asp Lys Xaa Xaa Arg His Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa
 50           55           60
Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa
 65           70           75           80
Arg Xaa Asp Glu Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
 85           90

```

<210> 56

<211> 128

<212> PRT

<213> 人工序列

<220>

<223> 纯化的多肽

<221> 变体

<222> (1)...(128)

<223> Xaa = 任何氨基酸

<400> 56

```

Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Arg Xaa Asp His
 1           5           10           15
Xaa Xaa Thr His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
 20           25           30
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Gln Xaa
 35           40           45
Ser Asn Xaa Xaa Val His Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa
 50           55           60
Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa
 65           70           75           80
Gln Xaa Thr His Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa

```

				85					90					95			
Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa		
			100					105					110				
Xaa	Xaa	Gln	Xaa	Phe	Asn	Xaa	Xaa	Arg	His	Xaa	Xaa	Xaa	Xaa	Xaa	His		
		115					120					125					

<210> 57
 <211> 94
 <212> PRT
 <213> 人工序列

<220>
 <223> 纯化的多肽

<221> 变体
 <222> (1)...(94)
 <223> Xaa = 任何氨基酸

Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Arg	Xaa	Asp	His		
1				5					10					15			
Xaa	Xaa	Thr	His	Xaa	Xaa	Xaa	Xaa	Xaa	His	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa		
			20					25					30				
Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Gln	Xaa		
		35				40					45						
Ser	Asn	Xaa	Xaa	Val	His	Xaa	Xaa	Xaa	Xaa	His	Xaa	Xaa	Xaa	Xaa	Xaa		
	50				55					60							
Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa		
65				70				75						80			
Gln	Xaa	Thr	His	Xaa	Xaa	Arg	His	Xaa	Xaa	Xaa	Xaa	Xaa	His				
				85				90									

<210> 58
 <211> 94
 <212> PRT
 <213> 人工序列

<220>
 <223> 纯化的多肽

<221> 变体
 <222> (1)...(94)
 <223> Xaa = 任何氨基酸

Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Gln	Xaa	Ser	Asn		
1				5					10					15			
Xaa	Xaa	Val	His	Xaa	Xaa	Xaa	Xaa	Xaa	His	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa		
			20					25					30				
Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Gln	Xaa		
		35				40					45						
Thr	His	Xaa	Xaa	Arg	His	Xaa	Xaa	Xaa	Xaa	His	Xaa	Xaa	Xaa	Xaa	Xaa		
	50				55					60							
Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa		
65				70				75						80			
Gln	Xaa	Phe	Asn	Xaa	Xaa	Arg	His	Xaa	Xaa	Xaa	Xaa	Xaa	His				
				85				90									

<210> 59
 <211> 94
 <212> PRT
 <213> 人工序列

<220>
 <223> 纯化的多肽

<221> 变体

<222> (1)...(94)

<223> Xaa = 任何氨基酸

<400> 59

```

Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa Ser His
 1          5          10          15
Xaa Xaa Val His Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
 20          25          30
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa
 35          40          45
Ser Ser Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa
 50          55          60
Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa
 65          70          75          80
Gln Xaa Thr His Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
          85          90

```

<210> 60

<211> 60

<212> PRT

<213> 人工序列

<220>

<223> 纯化的多肽

<221> 变体

<222> (1)...(60)

<223> Xaa = 任何氨基酸

<400> 60

```

Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa Ser Ser
 1          5          10          15
Xaa Xaa Arg His Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
 20          25          30
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa
 35          40          45
Thr His Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
 50          55          60

```

<210> 61

<211> 60

<212> PRT

<213> 人工序列

<220>

<223> 纯化的多肽

<221> 变体

<222> (1)...(60)

<223> Xaa = 任何氨基酸

<400> 61

```

Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa Ser His
 1          5          10          15
Xaa Xaa Val His Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
 20          25          30
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa
 35          40          45
Ser Ser Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
 50          55          60

```

<210> 62

<211> 128

<212> PRT
<213> 人工序列

<220>
<223> 纯化的多肽

<221> 变体
<222> (1)...(128)
<223> Xaa = 任何氨基酸

<400> 62
Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa Ser Asn
1 5 10 15
Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
20 25 30
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa
35 40 45
Ser Ser Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa
50 55 60
Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa
65 70 75 80
Gln Xaa Thr His Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa
85 90 95
Xaa Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa
100 105 110
Xaa Xaa Arg Xaa Asp Lys Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
115 120 125

<210> 63
<211> 94
<212> PRT
<213> 人工序列

<220>
<223> 纯化的多肽

<221> 变体
<222> (1)...(94)
<223> Xaa = 任何氨基酸

<400> 63
Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa Ser Ser
1 5 10 15
Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
20 25 30
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa
35 40 45
Thr His Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa
50 55 60
Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa
65 70 75 80
Arg Xaa Asp Lys Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
85 90

<210> 64
<211> 94
<212> PRT
<213> 人工序列

<220>
<223> 纯化的多肽

<221> 变体
<222> (1)...(94)
<223> Xaa = 任何氨基酸

<400> 64
 Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa Ser Asn
 1 5 10 15
 Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
 20 25 30
 Xaa Xaa Cys Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Gln Xaa
 35 40 45
 Ser Ser Xaa Xaa Arg His Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa
 50 55 60
 Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa
 65 70 75 80
 Gln Xaa Thr His Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
 85 90

<210> 65
 <211> 94
 <212> PRT
 <213> 人工序列

<220>
 <223> 纯化的多肽
 <221> 变体
 <222> (1)...(94)
 <223> Xaa = 任何氨基酸

<400> 65
 Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Arg Xaa Asp His
 1 5 10 15
 Xaa Xaa Thr His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
 20 25 30
 Xaa Xaa Cys Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Arg Xaa
 35 40 45
 Ser His Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa
 50 55 60
 Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa
 65 70 75 80
 Gln Xaa Ser His Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
 85 90

<210> 66
 <211> 60
 <212> PRT
 <213> 人工序列

<220>
 <223> 纯化的多肽
 <221> 变体
 <222> (1)...(60)
 <223> Xaa = 任何氨基酸

<400> 66
 Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Arg Xaa Ser His
 1 5 10 15
 Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
 20 25 30
 Xaa Xaa Cys Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Gln Xaa
 35 40 45
 Ser His Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
 50 55 60

<210> 67
 <211> 60

<212> PRT
<213> 人工序列

<220>
<223> 纯化的多肽

<221> 变体
<222> (1)...(60)
<223> Xaa = 任何氨基酸

<400> 67
Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Arg Xaa Asp His
1 5 10 15
Xaa Xaa Thr His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
20 25 30
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Arg Xaa
35 40 45
Ser His Xaa Xaa Arg His Xaa Xaa Xaa Xaa His
50 55 60

<210> 68
<211> 94
<212> PRT
<213> 人工序列

<220>
<223> 纯化的多肽

<221> 变体
<222> (1)...(94)
<223> Xaa = 任何氨基酸

<400> 68
Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa Ser His
1 5 10 15
Xaa Xaa Thr His Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
20 25 30
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Arg Xaa
35 40 45
Ser His Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa
50 55 60
Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa
65 70 75 80
Arg Xaa Asp His Xaa Xaa Thr His Xaa Xaa Xaa Xaa His
85 90

<210> 69
<211> 60
<212> PRT
<213> 人工序列

<220>
<223> 纯化的多肽

<221> 变体
<222> (1)...(60)
<223> Xaa = 任何氨基酸

<400> 69
Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Arg Xaa Ser His
1 5 10 15
Xaa Xaa Arg His Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
20 25 30
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Arg Xaa
35 40 45

Asp His Xaa Xaa Thr His Xaa Xaa Xaa Xaa Xaa His
50 55 60

<210> 70
<211> 60
<212> PRT
<213> 人工序列

<220>
<223> 纯化的多肽

<221> 变体
<222> (1)...(60)
<223> Xaa = 任何氨基酸

<400> 70
Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa Ser His
1 5 10 15
Xaa Xaa Thr His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa
20 25 30
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Arg Xaa
35 40 45
Ser His Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
50 55 60

<210> 71
<211> 89
<212> PRT
<213> 人工序列

<220>
<223> 质粒序列

<400> 71
Glu Arg Pro Tyr Ala Cys Pro Val Glu Ser Cys Asp Arg Arg Phe Ser
1 5 10 15
Arg Ser Asp Glu Leu Thr Arg His Ile Arg Ile His Thr Gly Gln Lys
20 25 30
Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Arg Ser Asp His
35 40 45
Leu Thr Thr His Ile Arg Thr His Thr Gly Glu Lys Pro Phe Ala Cys
50 55 60
Asp Ile Cys Gly Arg Lys Phe Ala Arg Ser Asp Glu Arg Lys Arg His
65 70 75 80
Thr Lys Ile His Leu Arg Gln Lys Asp
85

<210> 72
<211> 28
<212> PRT
<213> 人工序列

<220>
<223> 配位残基

<221> 变体
<222> 1, 13
<223> Xaa = Phe 或 Tyr

<221> 变体
<222> 19
<223> Xaa = 疏水残基

<221> 变体
<222> 2, 4-8, 10-12, 14-18, 20-21, 23-27

<223> Xaa = 任何氨基酸

<400> 72

```
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
 1           5           10           15
Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa His
                20           25
```

<210> 73

<211> 28

<212> PRT

<213> 人工序列

<220>

<223> 纯化的多肽

<221> 变体

<222> 1, 13

<223> Xaa = Phe 或 Tyr

<221> 变体

<222> 19

<223> Xaa = 疏水残基

<221> 变体

<222> 2, 4-8, 10-12, 14, 16, 20, 23-27

<223> Xaa = 任何氨基酸

<400> 73

```
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa
 1           5           10           15
Ser Asn Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
                20           25
```

<210> 74

<211> 6

<212> PRT

<213> 真核细胞

<220>

<221> 变体

<222> 3

<223> Xaa = Glu 或 Gln

<221> 变体

<222> 4

<223> Xaa = Lys 或 Arg

<221> 变体

<222> 6

<223> Xaa = Tyr 或 Phe

<400> 74

```
Thr Gly Xaa Xaa Pro Xaa
 1           5
```

<210> 75

<211> 23

<212> PRT

<213> 智人

<400> 75

```
Tyr Lys Cys Lys Gln Cys Gly Lys Ala Phe Gly Cys Pro Ser Asn Leu
 1           5           10           15
Arg Arg His Gly Arg Thr His
```

20

<210> 76
 <211> 23
 <212> PRT
 <213> 智人

<400> 76
 Tyr Gln Cys Asn Ile Cys Gly Lys Cys Phe Ser Cys Asn Ser Asn Leu
 1 5 10 15
 His Arg His Gln Arg Thr His
 20

<210> 77
 <211> 23
 <212> PRT
 <213> 智人

<400> 77
 Tyr Ser Cys Gly Ile Cys Gly Lys Ser Phe Ser Asp Ser Ser Ala Lys
 1 5 10 15
 Arg Arg His Cys Ile Leu His
 20

<210> 78
 <211> 23
 <212> PRT
 <213> 智人

<400> 78
 Tyr Thr Cys Ser Asp Cys Gly Lys Ala Phe Arg Asp Lys Ser Cys Leu
 1 5 10 15
 Asn Arg His Arg Arg Thr His
 20

<210> 79
 <211> 23
 <212> PRT
 <213> 智人

<400> 79
 Tyr Lys Cys Lys Glu Cys Gly Lys Ala Phe Asn His Ser Ser Asn Phe
 1 5 10 15
 Asn Lys His His Arg Ile His
 20

<210> 80
 <211> 23
 <212> PRT
 <213> 智人

<400> 80
 Phe Lys Cys Pro Val Cys Gly Lys Ala Phe Arg His Ser Ser Ser Leu
 1 5 10 15
 Val Arg His Gln Arg Thr His
 20

<210> 81
 <211> 24
 <212> PRT
 <213> 智人

<400> 81
 Tyr Arg Cys Lys Tyr Cys Asp Arg Ser Phe Ser Ile Ser Ser Asn Leu
 1 5 10 15

Gln Arg His Val Arg Asn Ile His
20

<210> 82
<211> 23
<212> PRT
<213> 智人

<400> 82
Tyr Glu Cys Asp His Cys Gly Lys Ala Phe Ser Ile Gly Ser Asn Leu
1 5 10 15
Asn Val His Arg Arg Ile His
20

<210> 83
<211> 23
<212> PRT
<213> 智人

<400> 83
Tyr Gly Cys His Leu Cys Gly Lys Ala Phe Ser Lys Ser Ser Asn Leu
1 5 10 15
Arg Arg His Glu Met Ile His
20

<210> 84
<211> 23
<212> PRT
<213> 智人

<400> 84
Tyr Lys Cys Lys Glu Cys Gly Gln Ala Phe Arg Gln Arg Ala His Leu
1 5 10 15
Ile Arg His His Lys Leu His
20

<210> 85
<211> 23
<212> PRT
<213> 智人

<400> 85
Tyr Lys Cys His Gln Cys Gly Lys Ala Phe Ile Gln Ser Phe Asn Leu
1 5 10 15
Arg Arg His Glu Arg Thr His
20

<210> 86
<211> 23
<212> PRT
<213> 智人

<400> 86
Phe Gln Cys Asn Gln Cys Gly Ala Ser Phe Thr Gln Lys Gly Asn Leu
1 5 10 15
Leu Arg His Ile Lys Leu His
20

<210> 87
<211> 23
<212> PRT
<213> 智人

<400> 87
Tyr Ala Cys His Leu Cys Gly Lys Ala Phe Thr Gln Ser Ser His Leu

1 5 10 15
 Arg Arg His Glu Lys Thr His
 20

<210> 88
 <211> 23
 <212> PRT
 <213> 智人

<400> 88
 Tyr Lys Cys Gly Gln Cys Gly Lys Phe Tyr Ser Gln Val Ser His Leu
 1 5 10 15
 Thr Arg His Gln Lys Ile His
 20

<210> 89
 <211> 23
 <212> PRT
 <213> 智人

<400> 89
 Tyr Ala Cys His Leu Cys Gly Lys Ala Phe Thr Gln Cys Ser His Leu
 1 5 10 15
 Arg Arg His Glu Lys Thr His
 20

<210> 90
 <211> 23
 <212> PRT
 <213> 智人

<400> 90
 Tyr Ala Cys His Leu Cys Ala Lys Ala Phe Ile Gln Cys Ser His Leu
 1 5 10 15
 Arg Arg His Glu Lys Thr His
 20

<210> 91
 <211> 23
 <212> PRT
 <213> 智人

<400> 91
 Tyr Val Cys Arg Glu Cys Gly Arg Gly Phe Arg Gln His Ser His Leu
 1 5 10 15
 Val Arg His Lys Arg Thr His
 20

<210> 92
 <211> 23
 <212> PRT
 <213> 智人

<400> 92
 Tyr Lys Cys Glu Glu Cys Gly Lys Ala Phe Arg Gln Ser Ser His Leu
 1 5 10 15
 Thr Thr His Lys Ile Ile His
 20

<210> 93
 <211> 23
 <212> PRT
 <213> 智人

<400> 93

Tyr Glu Cys Asp His Cys Gly Lys Ser Phe Ser Gln Ser Ser His Leu
 1 5 10 15
 Asn Val His Lys Arg Thr His
 20

<210> 94
 <211> 23
 <212> PRT
 <213> 智人

<400> 94
 Tyr Met Cys Ser Glu Cys Gly Arg Gly Phe Ser Gln Lys Ser Asn Leu
 1 5 10 15
 Ile Ile His Gln Arg Thr His
 20

<210> 95
 <211> 23
 <212> PRT
 <213> 智人

<400> 95
 Tyr Lys Cys Glu Glu Cys Gly Lys Ala Phe Thr Gln Ser Ser Asn Leu
 1 5 10 15
 Thr Lys His Lys Lys Ile His
 20

<210> 96
 <211> 23
 <212> PRT
 <213> 智人

<400> 96
 Phe Glu Cys Lys Asp Cys Gly Lys Ala Phe Ile Gln Lys Ser Asn Leu
 1 5 10 15
 Ile Arg His Gln Arg Thr His
 20

<210> 97
 <211> 23
 <212> PRT
 <213> 智人

<400> 97
 Tyr Val Cys Arg Glu Cys Arg Arg Gly Phe Ser Gln Lys Ser Asn Leu
 1 5 10 15
 Ile Arg His Gln Arg Thr His
 20

<210> 98
 <211> 23
 <212> PRT
 <213> 智人

<400> 98
 Tyr Glu Cys Glu Lys Cys Gly Lys Ala Phe Asn Gln Ser Ser Asn Leu
 1 5 10 15
 Thr Arg His Lys Lys Ser His
 20

<210> 99
 <211> 23
 <212> PRT
 <213> 智人

<400> 99

```

Tyr Glu Cys Asn Thr Cys Arg Lys Thr Phe Ser Gln Lys Ser Asn Leu
 1           5           10           15
Ile Val His Gln Arg Thr His
                20

```

<210> 100

<211> 23

<212> PRT

<213> 智人

<400> 100

```

Tyr Val Cys Ser Lys Cys Gly Lys Ala Phe Thr Gln Ser Ser Asn Leu
 1           5           10           15
Thr Val His Gln Lys Ile His
                20

```

<210> 101

<211> 23

<212> PRT

<213> 智人

<400> 101

```

Tyr Lys Cys Asp Glu Cys Gly Lys Asn Phe Thr Gln Ser Ser Asn Leu
 1           5           10           15
Ile Val His Lys Arg Ile His
                20

```

<210> 102

<211> 23

<212> PRT

<213> 智人

<400> 102

```

Tyr Glu Cys Asp Val Cys Gly Lys Thr Phe Thr Gln Lys Ser Asn Leu
 1           5           10           15
Gly Val His Gln Arg Thr His
                20

```

<210> 103

<211> 23

<212> PRT

<213> 智人

<400> 103

```

Tyr Glu Cys Val Gln Cys Gly Lys Gly Phe Thr Gln Ser Ser Asn Leu
 1           5           10           15
Ile Thr His Gln Arg Val His
                20

```

<210> 104

<211> 23

<212> PRT

<213> 智人

<400> 104

```

Tyr Lys Cys Pro Asp Cys Gly Lys Ser Phe Ser Gln Ser Ser Ser Leu
 1           5           10           15
Ile Arg His Gln Arg Thr His
                20

```

<210> 105

<211> 23

<212> PRT

<213> 智人

<400> 105
 Tyr Glu Cys Gln Asp Cys Gly Arg Ala Phe Asn Gln Asn Ser Ser Leu
 1 5 10 15
 Gly Arg His Lys Arg Thr His
 20

<210> 106
 <211> 23
 <212> PRT
 <213> 智人

<400> 106
 Tyr Glu Cys Asn Glu Cys Gly Lys Phe Phe Ser Gln Ser Ser Ser Leu
 1 5 10 15
 Ile Arg His Arg Arg Ser His
 20

<210> 107
 <211> 23
 <212> PRT
 <213> 智人

<400> 107
 Tyr Lys Cys Glu Glu Cys Gly Lys Ala Phe Asn Gln Ser Ser Thr Leu
 1 5 10 15
 Thr Arg His Lys Ile Val His
 20

<210> 108
 <211> 23
 <212> PRT
 <213> 智人

<400> 108
 Tyr Glu Cys Asn Glu Cys Gly Lys Ala Phe Ala Gln Asn Ser Thr Leu
 1 5 10 15
 Arg Val His Gln Arg Ile His
 20

<210> 109
 <211> 23
 <212> PRT
 <213> 智人

<400> 109
 Tyr Glu Cys His Asp Cys Gly Lys Ser Phe Arg Gln Ser Thr His Leu
 1 5 10 15
 Thr Gln His Arg Arg Ile His
 20

<210> 110
 <211> 23
 <212> PRT
 <213> 智人

<400> 110
 Tyr Glu Cys His Asp Cys Gly Lys Ser Phe Arg Gln Ser Thr His Leu
 1 5 10 15
 Thr Arg His Arg Arg Ile His
 20

<210> 111
 <211> 22
 <212> PRT

<213> 智人

<400> 111

His Lys Cys Leu Glu Cys Gly Lys Cys Phe Ser Gln Asn Thr His Leu
 1 5 10 15
 Thr Arg His Gln Arg Thr
 20

<210> 112

<211> 25

<212> PRT

<213> 智人

<400> 112

Tyr Val Cys Asp Val Glu Gly Cys Thr Trp Lys Phe Ala Arg Ser Asp
 1 5 10 15
 Glu Leu Asn Arg His Lys Lys Arg His
 20 25

<210> 113

<211> 25

<212> PRT

<213> 智人

<400> 113

Tyr His Cys Asp Trp Asp Gly Cys Gly Trp Lys Phe Ala Arg Ser Asp
 1 5 10 15
 Glu Leu Thr Arg His Tyr Arg Lys His
 20 25

<210> 114

<211> 25

<212> PRT

<213> 智人

<400> 114

Tyr Arg Cys Ser Trp Glu Gly Cys Glu Trp Arg Phe Ala Arg Ser Asp
 1 5 10 15
 Glu Leu Thr Arg His Phe Arg Lys His
 20 25

<210> 115

<211> 25

<212> PRT

<213> 智人

<400> 115

Phe Ser Cys Ser Trp Lys Gly Cys Glu Arg Arg Phe Ala Arg Ser Asp
 1 5 10 15
 Glu Leu Ser Arg His Arg Arg Thr His
 20 25

<210> 116

<211> 25

<212> PRT

<213> 智人

<400> 116

Phe Ala Cys Ser Trp Gln Asp Cys Asn Lys Lys Phe Ala Arg Ser Asp
 1 5 10 15
 Glu Leu Ala Arg His Tyr Arg Thr His
 20 25

<210> 117

<211> 25

<212> PRT
<213> 智人

<400> 117
Tyr His Cys Asn Trp Asp Gly Cys Gly Trp Lys Phe Ala Arg Ser Asp
1 5 10 15
Glu Leu Thr Arg His Tyr Arg Lys His
20 25

<210> 118
<211> 24
<212> PRT
<213> 智人

<400> 118
Phe Leu Cys Gln Tyr Cys Ala Gln Arg Phe Gly Arg Lys Asp His Leu
1 5 10 15
Thr Arg His Met Lys Lys Ser His
20

<210> 119
<211> 23
<212> PRT
<213> 智人

<400> 119
Phe Gln Cys Lys Thr Cys Gln Arg Lys Phe Ser Arg Ser Asp His Leu
1 5 10 15
Lys Thr His Thr Arg Thr His
20

<210> 120
<211> 23
<212> PRT
<213> 智人

<400> 120
Phe Ala Cys Glu Val Cys Gly Val Arg Phe Thr Arg Asn Asp Lys Leu
1 5 10 15
Lys Ile His Met Arg Lys His
20

<210> 121
<211> 25
<212> PRT
<213> 智人

<400> 121
Tyr Val Cys Asp Val Glu Gly Cys Thr Trp Lys Phe Ala Arg Ser Asp
1 5 10 15
Lys Leu Asn Arg His Lys Lys Arg His
20 25

<210> 122
<211> 23
<212> PRT
<213> 智人

<400> 122
Tyr Lys Cys Met Glu Cys Gly Lys Ala Phe Asn Arg Arg Ser His Leu
1 5 10 15
Thr Arg His Gln Arg Ile His
20

<210> 123

<211> 23
 <212> PRT
 <213> 智人

<400> 123
 Tyr Ile Cys Arg Lys Cys Gly Arg Gly Phe Ser Arg Lys Ser Asn Leu
 1 5 10 15
 Ile Arg His Gln Arg Thr His
 20

<210> 124
 <211> 23
 <212> PRT
 <213> 智人

<400> 124
 Tyr Leu Cys Ser Glu Cys Asp Lys Cys Phe Ser Arg Ser Thr Asn Leu
 1 5 10 15
 Ile Arg His Arg Arg Thr His
 20

<210> 125
 <211> 23
 <212> PRT
 <213> 智人

<400> 125
 Tyr Glu Cys Lys Glu Cys Gly Lys Ala Phe Ser Ser Gly Ser Asn Phe
 1 5 10 15
 Thr Arg His Gln Arg Ile His
 20

<210> 126
 <211> 23
 <212> PRT
 <213> 智人

<400> 126
 Tyr Glu Cys Asp His Cys Gly Lys Ala Phe Ser Val Ser Ser Asn Leu
 1 5 10 15
 Asn Val His Arg Arg Ile His
 20

<210> 127
 <211> 23
 <212> PRT
 <213> 智人

<400> 127
 Tyr Thr Cys Lys Gln Cys Gly Lys Ala Phe Ser Val Ser Ser Ser Leu
 1 5 10 15
 Arg Arg His Glu Thr Thr His
 20

<210> 128
 <211> 23
 <212> PRT
 <213> 智人

<400> 128
 Tyr Glu Cys Asn Tyr Cys Gly Lys Thr Phe Ser Val Ser Ser Thr Leu
 1 5 10 15
 Ile Arg His Gln Arg Ile His
 20

<210> 129
 <211> 23
 <212> PRT
 <213> 智人

<400> 129
 Tyr Arg Cys Glu Glu Cys Gly Lys Ala Phe Arg Trp Pro Ser Asn Leu
 1 5 10 15
 Thr Arg His Lys Arg Ile His
 20

<210> 130
 <211> 10
 <212> DNA
 <213> 人工序列

<220>
 <223> 靶位点

<400> 130
 gaagaggacc

10

<210> 131
 <211> 260
 <212> PRT
 <213> 智人

<400> 131
 Tyr Leu Pro Asp Thr Asp Asp Arg His Arg Ile Glu Glu Lys Arg Lys
 1 5 10 15
 Arg Thr Tyr Glu Thr Phe Lys Ser Ile Met Lys Lys Ser Pro Phe Ser
 20 25 30
 Gly Pro Thr Asp Pro Arg Pro Pro Arg Arg Ile Ala Val Pro Ser
 35 40 45
 Arg Ser Ser Ala Ser Val Pro Lys Pro Ala Pro Gln Pro Tyr Pro Phe
 50 55 60
 Thr Ser Ser Leu Ser Thr Ile Asn Tyr Asp Glu Phe Pro Thr Met Val*
 65 70 75 80
 Phe Pro Ser Gly Gln Ile Ser Gln Ala Ser Ala Leu Ala Pro Ala Pro
 85 90 95
 Pro Gln Val Leu Pro Gln Ala Pro Ala Pro Ala Pro Ala Met
 100 105 110
 Val Ser Ala Leu Ala Gln Ala Pro Ala Pro Val Pro Val Leu Ala Pro
 115 120 125
 Gly Pro Pro Gln Ala Val Ala Pro Pro Ala Pro Lys Pro Thr Gln Ala
 130 135 140
 Gly Glu Gly Thr Leu Ser Glu Ala Leu Leu Gln Leu Gln Phe Asp Asp
 145 150 155 160
 Glu Asp Leu Gly Ala Leu Leu Gly Asn Ser Thr Asp Pro Ala Val Phe
 165 170 175
 Thr Asp Leu Ala Ser Val Asp Asn Ser Glu Phe Gln Gln Leu Leu Asn
 180 185 190
 Gln Gly Ile Pro Val Ala Pro His Thr Thr Glu Pro Met Leu Met Glu
 195 200 205
 Tyr Pro Glu Ala Ile Thr Arg Leu Val Thr Ala Gln Arg Pro Pro Asp
 210 215 220
 Pro Ala Pro Ala Pro Leu Gly Ala Pro Gly Leu Pro Asn Gly Leu Leu
 225 230 235 240
 Ser Gly Asp Glu Asp Phe Ser Ser Ile Ala Asp Met Asp Phe Ser Ala
 245 250 255
 Leu Leu Ser Gln
 260

<210> 132
 <211> 127

<212> PRT
<213> 酿酒酵母

<400> 132
Asn Phe Asn Gln Ser Gly Asn Ile Ala Asp Ser Ser Leu Ser Phe Thr
1 5 10 15
Phe Thr Asn Ser Ser Asn Gly Pro Asn Leu Ile Thr Thr Gln Thr Asn
20 25 30
Ser Gln Ala Leu Ser Gln Pro Ile Ala Ser Ser Asn Val His Asp Asn
35 40 45
Phe Met Asn Asn Glu Ile Thr Ala Ser Lys Ile Asp Asp Gly Asn Asn
50 55 60
Ser Lys Pro Leu Ser Pro Gly Trp Thr Asp Gln Thr Ala Tyr Asn Ala
65 70 75 80
Phe Gly Ile Thr Thr Gly Met Phe Asn Thr Thr Thr Met Asp Asp Val
85 90 95
Tyr Asn Tyr Leu Phe Asp Asp Glu Asp Thr Pro Pro Asn Pro Lys Lys
100 105 110
Glu Ile Ser Met Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Ser
115 120 125

<210> 133
<211> 90
<212> PRT
<213> 酿酒酵母

<400> 133
Asn Ser Ala Ser Ser Ser Thr Lys Leu Asp Asp Asp Leu Gly Thr Ala
1 5 10 15
Ala Ala Val Leu Ser Asn Met Arg Ser Ser Pro Tyr Arg Thr His Asp
20 25 30
Lys Pro Ile Ser Asn Val Asn Asp Met Asn Asn Thr Asn Ala Leu Gly
35 40 45
Val Pro Ala Ser Arg Pro His Ser Ser Ser Phe Pro Ser Lys Gly Val
50 55 60
Leu Arg Pro Ile Leu Leu Arg Ile His Asn Ser Glu Gln Gln Pro Ile
65 70 75 80
Phe Glu Ser Asn Asn Ser Thr Ala Cys Ile
85 90

<210> 134
<211> 63
<212> PRT
<213> 智人

<400> 134
Val Ser Val Thr Phe Glu Asp Val Ala Val Leu Phe Thr Arg Asp Glu
1 5 10 15
Trp Lys Lys Leu Asp Leu Ser Gln Arg Ser Leu Tyr Arg Glu Val Met
20 25 30
Leu Glu Asn Tyr Ser Asn Leu Ala Ser Met Ala Gly Phe Leu Phe Thr
35 40 45
Lys Pro Lys Val Ile Ser Leu Leu Gln Gln Gly Glu Asp Pro Trp
50 55 60

<210> 135
<211> 207
<212> DNA
<213> 人工序列

<220>
<223> 质粒序列

<400> 135
taatacgact cactataggg aatattaagc taagctcacc atgggtaagc ctatccctaa

60

```

ccctctcctc ggtctcgatt ctacacaagc tatgggtgct cctccaaaaa agaagagaaa 120
ggtagctgga tccactagta acggccgcca gtgtgctgga attctgcaga tatccatcac 180
actggcggcg ctcgaggcat gcatcta 207

```

```

<210> 136
<211> 30
<212> DNA
<213> 人工序列

```

```

<220>
<223> 引物

```

```

<400> 136
cgatctgcag ggtccttttc atcacgtgct 30

```

```

<210> 137
<211> 31
<212> DNA
<213> 人工序列

```

```

<220>
<223> 引物

```

```

<400> 137
cgatcgatgc cggcggacgg atcgcttgcc t 31

```

```

<210> 138
<211> 28
<212> DNA
<213> 人工序列

```

```

<220>
<223> 引物

```

```

<400> 138
aattccatgg gtaagcctat ccctaacc 28

```

```

<210> 139
<211> 28
<212> DNA
<213> 人工序列

```

```

<220>
<223> 引物

```

```

<400> 139
aattggatcc agctaccitt ctcttctt 28

```

```

<210> 140
<211> 43
<212> DNA
<213> 人工序列

```

```

<220>
<223> 引物

```

```

<400> 140
aaggaaggaa ggaagcggcc gcagccaatt ttaatcaaag tgg 43

```

```

<210> 141
<211> 32
<212> DNA
<213> 人工序列

```

```

<220>

```

<223> 引物

<400> 141

acatacatgc atgcgccggtt actagtggat cc

32

<210> 142

<211> 474

<212> DNA

<213> 人工序列

<220>

<223> 质粒序列

<400> 142

ggccgccaagt	gtgctggaat	tctgcagata	tccatcacac	tggcggccgc	agccaatttt	60
aatcaaagt	ggaatattgc	tgatagctca	ttgtccttca	ctttcactaa	cagtagcaac	120
ggtccgaacc	tcataacaac	tcaaacaat	tctcaagcgc	tttcacaacc	aattgcctcc	180
tctaacgttc	atgataactt	catgaataat	gaaatcacgg	ctagtaaaat	tgatgatggg	240
aataattcaa	aaccactgtc	acctgggttg	acggaccaaa	ctgcgtataa	cgcgtttgga	300
atcactacag	ggatgtttta	taccactaca	atggatgatg	tatataacta	tctattcgat	360
gatgaagata	ccccaccaa	ccccaaaaa	gagatctcta	tggtttaccc	atacgatgtt	420
ccagattacg	ctagctaagg	atccactagt	aacggcgcat	gcattctagag	ggcc	474

<210> 143

<211> 44

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 143

aaggaaggaa ggaagcggcc gcaaattctg catcttcac c tacc

44

<210> 144

<211> 33

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 144

acatacatgc atgctgtaga attgttgctt tcg

33

<210> 145

<211> 330

<212> DNA

<213> 人工序列

<220>

<223> 质粒序列

<400> 145

ggccgccaagt	gtgctggaat	tctgcagata	tccatcacac	tggcggccgc	aaattctgca	60
tcttcactcta	ccaaactaga	cgacgacttg	ggtacagcag	cagcagtgct	atcaaaccatg	120
agatcatccc	catatagaac	tcatgataaa	cccatttcca	atgtcaatga	catgaataac	180
acaaatgcgc	tcgggtgtgcc	ggctagtagg	cctcatctgt	catcttttcc	atcaaagggt	240
gtcttaagac	caattctgtt	acgtatccat	aattccgaac	aacaacccat	tttcgaaagc	300
aacaattcta	cagcatgcat	ctagagggcc				330

<210> 146

<211> 25

<212> PRT

<213> 智人

<400> 146

```

Phe Met Cys Thr Trp Ser Tyr Cys Gly Lys Arg Phe Thr Asp Arg Ser
 1           5           10           15
Ala Leu Ala Arg His Lys Arg Thr His
           20           25

```

<210> 147

<211> 25

<212> PRT

<213> 智人

<400> 147

```

His Ile Cys His Ile Gln Gly Cys Gly Lys Val Tyr Gly Asp Arg Ser
 1           5           10           15
His Leu Thr Arg His Leu Arg Trp His
           20           25

```

<210> 148

<211> 23

<212> PRT

<213> 智人

<400> 148

```

Phe Ala Cys Pro Glu Cys Pro Lys Arg Phe Met Asp Ser Ser Lys Leu
 1           5           10           15
Ser Arg His Ile Lys Thr His
           20

```

<210> 149

<211> 25

<212> PRT

<213> 智人

<400> 149

```

Tyr Ala Cys Pro Val Glu Ser Cys Asp Arg Arg Phe Ser Asp Ser Ser
 1           5           10           15
Asn Leu Thr Arg His Ile Arg Ile His
           20           25

```

<210> 150

<211> 25

<212> PRT

<213> 智人

<400> 150

```

His Ile Cys His Ile Gln Gly Cys Gly Lys Val Tyr Gly Asp Arg Ser
 1           5           10           15
Ser Leu Thr Arg His Leu Arg Trp His
           20           25

```

<210> 151

<211> 23

<212> PRT

<213> 智人

<400> 151

```

Tyr Lys Cys Lys Glu Cys Gly Lys Ala Phe Asn His Ser Ser Asn Phe
 1           5           10           15
Asn Lys His His Arg Ile His
           20

```

<210> 152

<211> 23

<212> PRT

<213> 智人

<400> 152

Phe	Lys	Cys	Pro	Val	Cys	Gly	Lys	Ala	Phe	Arg	His	Ser	Ser	Ser	Leu
1				5					10					15	
Val	Arg	His	Gln	Arg	Thr	His									
			20												

<210> 153

<211> 24

<212> PRT

<213> 智人

<400> 153

Tyr	Arg	Cys	Lys	Tyr	Cys	Asp	Arg	Ser	Phe	Ser	Ile	Ser	Ser	Asn	Leu
1				5					10					15	
Gln	Arg	His	Val	Arg	Asn	Ile	His								
			20												

<210> 154

<211> 23

<212> PRT

<213> 智人

<400> 154

Tyr	Gly	Cys	His	Leu	Cys	Gly	Lys	Ala	Phe	Ser	Lys	Ser	Ser	Asn	Leu
1				5					10					15	
Arg	Arg	His	Glu	Met	Ile	His									
			20												

<210> 155

<211> 23

<212> PRT

<213> 智人

<400> 155

Tyr	Lys	Cys	Lys	Glu	Cys	Gly	Gln	Ala	Phe	Arg	Gln	Arg	Ala	His	Leu
1				5					10					15	
Ile	Arg	His	His	Lys	Leu	His									
			20												

<210> 156

<211> 23

<212> PRT

<213> 智人

<400> 156

Tyr	Lys	Cys	His	Gln	Cys	Gly	Lys	Ala	Phe	Ile	Gln	Ser	Phe	Asn	Leu
1				5					10					15	
Arg	Arg	His	Glu	Arg	Thr	His									
			20												

<210> 157

<211> 23

<212> PRT

<213> 智人

<400> 157

Phe	Gln	Cys	Asn	Gln	Cys	Gly	Ala	Ser	Phe	Thr	Gln	Lys	Gly	Asn	Leu
1				5					10					15	
Leu	Arg	His	Ile	Lys	Leu	His									
			20												

<210> 158

<211> 23

<212> PRT
<213> 黑腹果蝇

<400> 158
Tyr Thr Cys Ser Tyr Cys Gly Lys Ser Phe Thr Gln Ser Asn Thr Leu
1 5 10 15
Lys Gln His Thr Arg Ile His
20

<210> 159
<211> 23
<212> PRT
<213> 智人

<400> 159
Tyr Val Cys Arg Glu Cys Gly Arg Gly Phe Arg Gln His Ser His Leu
1 5 10 15
Val Arg His Lys Arg Thr His
20

<210> 160
<211> 23
<212> PRT
<213> 智人

<400> 160
Tyr Glu Cys Asp His Cys Gly Lys Ser Phe Ser Gln Ser Ser His Leu
1 5 10 15
Asn Val His Lys Arg Thr His
20

<210> 161
<211> 23
<212> PRT
<213> 智人

<400> 161
Tyr Met Cys Ser Glu Cys Gly Arg Gly Phe Ser Gln Lys Ser Asn Leu
1 5 10 15
Ile Ile His Gln Arg Thr His
20

<210> 162
<211> 23
<212> PRT
<213> 智人

<400> 162
Tyr Lys Cys Glu Glu Cys Gly Lys Ala Phe Thr Gln Ser Ser Asn Leu
1 5 10 15
Thr Lys His Lys Lys Ile His
20

<210> 163
<211> 23
<212> PRT
<213> 智人

<400> 163
Tyr Lys Cys Glu Glu Cys Gly Lys Ala Phe Asn Gln Ser Ser Thr Leu
1 5 10 15
Thr Arg His Lys Ile Val His
20

<210> 164

<211> 23
 <212> PRT
 <213> 智人

<400> 164
 Tyr Glu Cys His Asp Cys Gly Lys Ser Phe Arg Gln Ser Thr His Leu
 1 5 10 15
 Thr Arg His Arg Arg Ile His
 20

<210> 165
 <211> 24
 <212> PRT
 <213> 智人

<400> 165
 Phe Leu Cys Gln Tyr Cys Ala Gln Arg Phe Gly Arg Lys Asp His Leu
 1 5 10 15
 Thr Arg His Met Lys Lys Ser His
 20

<210> 166
 <211> 25
 <212> PRT
 <213> 智人

<400> 166
 Tyr Val Cys Asp Val Glu Gly Cys Thr Trp Lys Phe Ala Arg Ser Asp
 1 5 10 15
 Lys Leu Asn Arg His Lys Lys Arg His
 20 25

<210> 167
 <211> 23
 <212> PRT
 <213> 智人

<400> 167
 Phe Ala Cys Pro Glu Cys Pro Lys Arg Phe Met Arg Ser Asp Asn Leu
 1 5 10 15
 Thr Gln His Ile Lys Thr His
 20

<210> 168
 <211> 23
 <212> PRT
 <213> 智人

<400> 168
 Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Ser Pro Ala Asp Leu
 1 5 10 15
 Thr Arg His Ile Arg Thr His
 20

<210> 169
 <211> 23
 <212> PRT
 <213> 智人

<400> 169
 Tyr Glu Cys Lys Glu Cys Gly Lys Ala Phe Ser Ser Gly Ser Asn Phe
 1 5 10 15
 Thr Arg His Gln Arg Ile His
 20

<210> 170
 <211> 23
 <212> PRT
 <213> 智人

<400> 170
 Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Thr His Ile Asp Leu
 1 5 10 15
 Ile Arg His Ile Arg Thr His
 20

<210> 171
 <211> 23
 <212> PRT
 <213> 智人

<400> 171
 Tyr Glu Cys Asp His Cys Gly Lys Ala Phe Ser Val Ser Ser Asn Leu
 1 5 10 15
 Asn Val His Arg Arg Ile His
 20

<210> 172
 <211> 23
 <212> PRT
 <213> 智人

<400> 172
 Tyr Glu Cys Asn Tyr Cys Gly Lys Thr Phe Ser Val Ser Ser Thr Leu
 1 5 10 15
 Ile Arg His Gln Arg Ile His
 20

<210> 173
 <211> 23
 <212> PRT
 <213> 智人

<400> 173
 Tyr Lys Cys Lys Gln Cys Gly Lys Ala Phe Gly Cys Pro Ser Asn Leu
 1 5 10 15
 Arg Arg His Gly Arg Thr His
 20

<210> 174
 <211> 23
 <212> PRT
 <213> 智人

<400> 174
 Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Asp Ser Gly Asn Leu
 1 5 10 15
 Arg Val His Ile Arg Thr His
 20

<210> 175
 <211> 23
 <212> PRT
 <213> 智人

<400> 175
 Tyr Ala Cys His Leu Cys Gly Lys Ala Phe Thr Gln Cys Ser His Leu
 1 5 10 15
 Arg Arg His Glu Lys Thr His
 20

<210> 176
 <211> 23
 <212> PRT
 <213> 智人

<400> 176
 Tyr Lys Cys Glu Glu Cys Gly Lys Ala Phe Arg Gln Ser Ser His Leu
 1 5 10 15
 Thr Thr His Lys Ile Ile His
 20

<210> 177
 <211> 23
 <212> PRT
 <213> 智人

<400> 177
 Phe Glu Cys Lys Asp Cys Gly Lys Ala Phe Ile Gln Lys Ser Asn Leu
 1 5 10 15
 Ile Arg His Gln Arg Thr His
 20

<210> 178
 <211> 23
 <212> PRT
 <213> 智人

<400> 178
 Tyr Val Cys Ser Lys Cys Gly Lys Ala Phe Thr Gln Ser Ser Asn Leu
 1 5 10 15
 Thr Val His Gln Lys Ile His
 20

<210> 179
 <211> 23
 <212> PRT
 <213> 智人

<400> 179
 Tyr Lys Cys Pro Asp Cys Gly Lys Ser Phe Ser Gln Ser Ser Ser Leu
 1 5 10 15
 Ile Arg His Gln Arg Thr His
 20

<210> 180
 <211> 23
 <212> PRT
 <213> 智人

<400> 180
 Tyr Glu Cys His Asp Cys Gly Lys Ser Phe Arg Gln Ser Thr His Leu
 1 5 10 15
 Thr Gln His Arg Arg Ile His
 20

<210> 181
 <211> 25
 <212> PRT
 <213> 智人

<400> 181
 Tyr Val Cys Asp Val Glu Gly Cys Thr Trp Lys Phe Ala Arg Ser Asp
 1 5 10 15
 Glu Leu Asn Arg His Lys Lys Arg His

20 25

<210> 182
 <211> 23
 <212> PRT
 <213> 智人

<400> 182
 Phe Gln Cys Lys Thr Cys Gln Arg Lys Phe Ser Arg Ser Asp His Leu
 1 5 10 15
 Lys Thr His Thr Arg Thr His
 20

<210> 183
 <211> 23
 <212> PRT
 <213> 智人

<400> 183
 Tyr Lys Cys Met Glu Cys Gly Lys Ala Phe Asn Arg Arg Ser His Leu
 1 5 10 15
 Thr Arg His Gln Arg Ile His
 20

<210> 184
 <211> 23
 <212> PRT
 <213> 智人

<400> 184
 Tyr Ile Cys Arg Lys Cys Gly Arg Gly Phe Ser Arg Lys Ser Asn Leu
 1 5 10 15
 Ile Arg His Gln Arg Thr His
 20

<210> 185
 <211> 22
 <212> PRT
 <213> 黑腹果蝇

<400> 185
 Phe His Cys Gly Tyr Cys Glu Lys Ser Phe Ser Val Lys Asp Tyr Leu
 1 5 10 15
 Thr Lys Ile Arg Thr His
 20

<210> 186
 <211> 23
 <212> PRT
 <213> 智人

<400> 186
 Tyr Thr Cys Lys Gln Cys Gly Lys Ala Phe Ser Val Ser Ser Ser Leu
 1 5 10 15
 Arg Arg His Glu Thr Thr His
 20

<210> 187
 <211> 23
 <212> PRT
 <213> 智人

<400> 187
 Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Asp Pro Gly Ala Leu
 1 5 10 15

Val Arg His Ile Arg Thr His
20

<210> 188
<211> 23
<212> PRT
<213> 智人

<400> 188
Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Asp Pro Gly His Leu
1 5 10 15
Val Arg His Ile Arg Thr His
20

<210> 189
<211> 23
<212> PRT
<213> 智人

<400> 189
Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Asp Pro Gly Asn Leu
1 5 10 15
Lys Arg His Ile Arg Thr His
20

<210> 190
<211> 23
<212> PRT
<213> 智人

<400> 190
Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Asp Cys Arg Asp Leu
1 5 10 15
Ala Arg His Ile Arg Thr His
20

<210> 191
<211> 25
<212> PRT
<213> 智人

<400> 191
Tyr Ala Cys Pro Val Glu Ser Cys Asp Arg Arg Phe Ser Met Ser His
1 5 10 15
His Leu Lys Glu His Ile Arg Thr His
20 25

<210> 192
<211> 23
<212> PRT
<213> 智人

<400> 192
Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Gln Gln Ala Ser Leu
1 5 10 15
Asn Ala His Ile Arg Thr His
20

<210> 193
<211> 23
<212> PRT
<213> 智人

<400> 193
Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Gln Ser Gly Asp Leu

1 5 10 15
 Arg Arg His Ile Arg Thr His
 20

<210> 194
 <211> 23
 <212> PRT
 <213> 智人

<400> 194
 Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Gln Ser Ser Asp Leu
 1 5 10 15
 Val Arg His Ile Arg Thr His
 20

<210> 195
 <211> 23
 <212> PRT
 <213> 智人

<400> 195
 Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Gln Arg Gly Thr Leu
 1 5 10 15
 Thr Arg His Ile Arg Thr His
 20

<210> 196
 <211> 23
 <212> PRT
 <213> 智人

<400> 196
 Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Arg Ser Asp Thr Leu
 1 5 10 15
 Ser Asn His Ile Arg Thr His
 20

<210> 197
 <211> 23
 <212> PRT
 <213> 智人

<400> 197
 Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Thr Ala Asp Lys Leu
 1 5 10 15
 Ser Arg His Ile Arg Thr His
 20

<210> 198
 <211> 23
 <212> PRT
 <213> 智人

<400> 198
 Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Thr Ser Gly Asn Leu
 1 5 10 15
 Val Arg His Ile Arg Thr His
 20

<210> 199
 <211> 23
 <212> PRT
 <213> 智人

<400> 199

<210> 204	
<211> 9	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<400> 204	
gtdaaggav	9
<210> 205	
<211> 9	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<221> 其他_特征	
<222> (1)...(9)	
<223> n = A,T,C 或 G	
<400> 205	
ghgnggkgg	9
<210> 206	
<211> 9	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<400> 206	
hgaaaggaa	9
<210> 207	
<211> 9	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<221> 其他_特征	
<222> (1)...(9)	
<223> n = A,T,C 或 G	
<400> 207	
nggghggaw	9
<210> 208	
<211> 9	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<221> 其他_特征	
<222> (1)...(9)	
<223> n = A,T,C 或 G	
<400> 208	

hganggggya	9
<210> 209	
<211> 9	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<221> 其他_特征	
<222> (1)...(9)	
<223> n = A,T,C 或 G	
<400> 209	
gaanggggya	9
<210> 210	
<211> 9	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<400> 210	
ggggagghga	9
<210> 211	
<211> 12	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<400> 211	
hgagyagagg ag	12
<210> 212	
<211> 12	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<400> 212	
graacagagg ag	12
<210> 213	
<211> 12	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<400> 213	
hgaacagagg ag	12
<210> 214	
<211> 12	
<212> DNA	
<213> 人工序列	

<220>	
<223> 靶序列	
<400> 214	
katgtdaacg td	12
<210> 215	
<211> 12	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<400> 215	
gcagtdgaat gt	12
<210> 216	
<211> 12	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<400> 216	
ggcgyahgaa ac	12
<210> 217	
<211> 12	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<400> 217	
aaghgaaagg tg	12
<210> 218	
<211> 12	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<221> 其他_特征	
<222> (1)...(12)	
<223> n = A,T,C 或 G	
<400> 218	
aachgahgan gg	12
<210> 219	
<211> 9	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<221> 其他_特征	
<222> (1)...(9)	

<223> n = A,T,C 或 G

<400> 219
hgahgangg

9

<210> 220
<211> 9
<212> DNA
<213> 人工序列

<220>
<223> 靶序列

<221> 其他_特征
<222> (1)...(9)
<223> n = A,T,C 或 G

<400> 220
hgahgangg

9

<210> 221
<211> 210
<212> PRT
<213> 人工序列

<220>
<223> 质粒序列

<400> 221
Met Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asn Ser Thr Gln
1 5 10 15
Ala Met Gly Ala Pro Pro Lys Lys Lys Arg Lys Val Gly Ile Arg Ile
20 25 30
Pro Gly Glu Lys Pro Tyr Glu Cys Asp His Cys Gly Lys Ser Phe Ser
35 40 45
Gln Ser Ser His Leu Asn Val His Lys Arg Thr His Thr Gly Glu Lys
50 55 60
Pro Tyr Lys Cys His Gln Cys Gly Lys Ala Phe Ile Gln Ser Phe Asn
65 70 75 80
Leu Arg Arg His Glu Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
85 90 95
Met Glu Cys Gly Lys Ala Phe Asn Arg Arg Ser His Leu Thr Arg His
100 105 110
Gln Arg Ile His Ala Ala Ala Ala Asn Ser Ala Ser Ser Ser Thr Lys
115 120 125
Leu Asp Asp Asp Leu Gly Thr Ala Ala Ala Val Leu Ser Asn Met Arg
130 135 140
Ser Ser Pro Tyr Arg Thr His Asp Lys Pro Ile Ser Asn Val Asn Asp
145 150 155 160
Met Asn Asn Thr Asn Ala Leu Gly Val Pro Ala Ser Arg Pro His Ser
165 170 175
Ser Ser Phe Pro Ser Lys Gly Val Leu Arg Pro Ile Leu Leu Arg Ile
180 185 190
His Asn Ser Glu Gln Gln Pro Ile Phe Glu Ser Asn Asn Ser Thr Ala
195 200 205
Cys Ile
210

<210> 222
<211> 238
<212> PRT
<213> 人工序列

<220>
<223> 质粒序列

<400> 222

```

Met Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asn Ser Thr Gln
 1          5          10          15
Ala Met Gly Ala Pro Pro Lys Lys Lys Arg Lys Val Gly Ile Arg Ile
 20          25          30
Pro Gly Glu Lys Pro Tyr Ile Cys Arg Lys Cys Gly Arg Gly Phe Ser
 35          40          45
Arg Lys Ser Asn Leu Ile Arg His Gln Arg Thr His Thr Gly Glu Lys
 50          55          60
Pro Tyr Ile Cys Arg Lys Cys Gly Arg Gly Phe Ser Arg Lys Ser Asn
 65          70          75          80
Leu Ile Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
 85          90          95
Pro Asp Cys Gly Lys Ser Phe Ser Gln Ser Ser Ser Leu Ile Arg His
100          105          110
Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Glu Glu Cys Gly
115          120          125
Lys Ala Phe Arg Gln Ser Ser His Leu Thr Thr His Lys Ile Ile His
130          135          140
Ala Ala Ala Ala Asn Ser Ala Ser Ser Ser Thr Lys Leu Asp Asp Asp
145          150          155          160
Leu Gly Thr Ala Ala Val Leu Ser Asn Met Arg Ser Ser Pro Tyr
165          170          175
Arg Thr His Asp Lys Pro Ile Ser Asn Val Asn Asp Met Asn Asn Thr
180          185          190
Asn Ala Leu Gly Val Pro Ala Ser Arg Pro His Ser Ser Ser Phe Pro
195          200          205
Ser Lys Gly Val Leu Arg Pro Ile Leu Leu Arg Ile His Asn Ser Glu
210          215          220
Gln Gln Pro Ile Phe Glu Ser Asn Asn Ser Thr Ala Cys Ile
225          230          235

```

<210> 223

<211> 238

<212> PRT

<213> 人工序列

<220>

<223> 质粒序列

<400> 223

```

Met Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asn Ser Thr Gln
 1          5          10          15
Ala Met Gly Ala Pro Pro Lys Lys Lys Arg Lys Val Gly Ile Arg Ile
 20          25          30
Pro Gly Glu Lys Pro Tyr Ile Cys Arg Lys Cys Gly Arg Gly Phe Ser
 35          40          45
Arg Lys Ser Asn Leu Ile Arg His Gln Arg Thr His Thr Gly Glu Lys
 50          55          60
Pro Tyr Ile Cys Arg Lys Cys Gly Arg Gly Phe Ser Arg Lys Ser Asn
 65          70          75          80
Leu Ile Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Phe Gln Cys
 85          90          95
Arg Ile Cys Met Arg Asn Phe Ser Gln Arg Gly Thr Leu Thr Arg His
100          105          110
Ile Arg Thr His Thr Gly Glu Lys Pro Tyr Val Cys Arg Glu Cys Gly
115          120          125
Arg Gly Phe Arg Gln His Ser His Leu Val Arg His Lys Arg Thr His
130          135          140
Ala Ala Ala Ala Asn Ser Ala Ser Ser Ser Thr Lys Leu Asp Asp Asp
145          150          155          160
Leu Gly Thr Ala Ala Ala Val Leu Ser Asn Met Arg Ser Ser Pro Tyr
165          170          175
Arg Thr His Asp Lys Pro Ile Ser Asn Val Asn Asp Met Asn Asn Thr

```

```

                180                185                190
Asn Ala Leu Gly Val Pro Ala Ser Arg Pro His Ser Ser Ser Phe Pro
      195                200                205
Ser Lys Gly Val Leu Arg Pro Ile Leu Leu Arg Ile His Asn Ser Glu
      210                215                220
Gln Gln Pro Ile Phe Glu Ser Asn Asn Ser Thr Ala Cys Ile
      225                230                235

```

<210> 224

<211> 238

<212> PRT

<213> 人工序列

<220>

<223> 质粒序列

<400> 224

```

Met Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asn Ser Thr Gln
 1                5                10                15
Ala Met Gly Ala Pro Pro Lys Lys Lys Arg Lys Val Gly Ile Arg Ile
      20                25                30
Pro Gly Glu Lys Pro Tyr Ile Cys Arg Lys Cys Gly Arg Gly Phe Ser
      35                40                45
Arg Lys Ser Asn Leu Ile Arg His Gln Arg Thr His Thr Gly Glu Lys
      50                55                60
Pro Tyr Ile Cys Arg Lys Cys Gly Arg Gly Phe Ser Arg Lys Ser Asn
      65                70                75                80
Leu Ile Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Phe Gln Cys
      85                90                95
Arg Ile Cys Met Arg Asn Phe Ser Gln Arg Gly Thr Leu Thr Arg His
      100                105                110
Ile Arg Thr His Thr Gly Glu Lys Pro Tyr Glu Cys His Asp Cys Gly
      115                120                125
Lys Ser Phe Arg Gln Ser Thr His Leu Thr Gln His Arg Arg Ile His
      130                135                140
Ala Ala Ala Ala Asn Ser Ala Ser Ser Ser Thr Lys Leu Asp Asp Asp
      145                150                155                160
Leu Gly Thr Ala Ala Ala Val Leu Ser Asn Met Arg Ser Ser Pro Tyr
      165                170                175
Arg Thr His Asp Lys Pro Ile Ser Asn Val Asn Asp Met Asn Asn Thr
      180                185                190
Asn Ala Leu Gly Val Pro Ala Ser Arg Pro His Ser Ser Ser Phe Pro
      195                200                205
Ser Lys Gly Val Leu Arg Pro Ile Leu Leu Arg Ile His Asn Ser Glu
      210                215                220
Gln Gln Pro Ile Phe Glu Ser Asn Asn Ser Thr Ala Cys Ile
      225                230                235

```

<210> 225

<211> 274

<212> PRT

<213> 人工序列

<220>

<223> 质粒序列

<400> 225

```

Met Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asn Ser Thr Gln
 1                5                10                15
Ala Met Gly Ala Pro Pro Lys Lys Lys Arg Lys Val Gly Ile Arg Ile
      20                25                30
Pro Gly Glu Lys Pro Tyr Thr Cys Lys Gln Cys Gly Lys Ala Phe Ser
      35                40                45
Val Ser Ser Ser Leu Arg Arg His Glu Thr Thr His Thr Gly Glu Lys
      50                55                60

```

```

Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Asp Ser Gly Asn
65      70      75      80
Leu Arg Val His Ile Arg Thr His Thr Gly Glu Lys Pro Tyr Thr Cys
85      90      95
Lys Gln Cys Gly Lys Ala Phe Ser Val Ser Ser Ser Leu Arg Arg His
100     105     110
Glu Thr Thr His Thr Gly Glu Lys Pro Phe His Cys Gly Tyr Cys Glu
115     120     125
Lys Ser Phe Ser Val Lys Asp Tyr Leu Thr Lys Ile Arg Thr His Ala
130     135     140
Ala Ala Ala Asn Phe Asn Gln Ser Gly Asn Ile Ala Asp Ser Ser Leu
145     150     155     160
Ser Phe Thr Phe Thr Asn Ser Ser Asn Gly Pro Asn Leu Ile Thr Thr
165     170     175
Gln Thr Asn Ser Gln Ala Leu Ser Gln Pro Ile Ala Ser Ser Asn Val
180     185     190
His Asp Asn Phe Met Asn Asn Glu Ile Thr Ala Ser Lys Ile Asp Asp
195     200     205
Gly Asn Asn Ser Lys Pro Leu Ser Pro Gly Trp Thr Asp Gln Thr Ala
210     215     220
Tyr Asn Ala Phe Gly Ile Thr Thr Gly Met Phe Asn Thr Thr Thr Met
225     230     235     240
Asp Asp Val Tyr Asn Tyr Leu Phe Asp Asp Glu Asp Thr Pro Pro Asn
245     250     255
Pro Lys Lys Glu Ile Ser Met Ala Tyr Pro Tyr Asp Val Pro Asp Tyr
260     265     270
Ala Ser

```

```

<210> 226
<211> 277
<212> PRT
<213> 人工序列

```

```

<220>
<223> 质粒序列

```

```

<400> 226
Met Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asn Ser Thr Gln
1      5      10      15
Ala Met Gly Ala Pro Pro Lys Lys Lys Arg Lys Val Gly Ile Arg Ile
20     25     30
Pro Gly Glu Lys Pro Tyr Ala Cys Pro Val Glu Ser Cys Asp Arg Arg
35     40     45
Phe Ser Met Ser His His Leu Lys Glu His Ile Arg Thr His Thr Gly
50     55     60
Glu Lys Pro Phe Glu Cys Lys Asp Cys Gly Lys Ala Phe Ile Gln Lys
65     70     75     80
Ser Asn Leu Ile Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr
85     90     95
Thr Cys Lys Gln Cys Gly Lys Ala Phe Ser Val Ser Ser Ser Leu Arg
100    105    110
Arg His Glu Thr Thr His Thr Gly Lys Pro Phe Gln Cys Arg Ile
115    120    125
Cys Met Arg Asn Phe Ser Gln Ser Gly Asp Leu Arg Arg His Ile Arg
130    135    140
Thr His Ala Ala Ala Ala Asn Phe Asn Gln Ser Gly Asn Ile Ala Asp
145    150    155     160
Ser Ser Leu Ser Phe Thr Phe Thr Asn Ser Ser Asn Gly Pro Asn Leu
165    170    175
Ile Thr Thr Gln Thr Asn Ser Gln Ala Leu Ser Gln Pro Ile Ala Ser
180    185    190
Ser Asn Val His Asp Asn Phe Met Asn Asn Glu Ile Thr Ala Ser Lys
195    200    205
Ile Asp Asp Gly Asn Asn Ser Lys Pro Leu Ser Pro Gly Trp Thr Asp

```

210		215		220
Gln Thr Ala Tyr Asn Ala Phe Gly Ile Thr Thr Gly Met Phe Asn Thr				
225		230		235
Thr Thr Met Asp Asp Val Tyr Asn Tyr Leu Phe Asp Asp Glu Asp Thr				
	245		250	255
Pro Pro Asn Pro Lys Lys Glu Ile Ser Met Ala Tyr Pro Tyr Asp Val				
	260	265		270
Pro Asp Tyr Ala Ser				
275				

<210> 227
 <211> 275
 <212> PRT
 <213> 人工序列

<220>
 <223> 质粒序列

<400> 227
Met Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asn Ser Thr Gln
1 5 10 15
Ala Met Gly Ala Pro Pro Lys Lys Lys Arg Lys Val Gly Ile Arg Ile
20 25 30
Pro Gly Glu Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser
35 40 45
Asp Ser Gly Asn Leu Arg Val His Ile Arg Thr His Thr Gly Glu Lys
50 55 60
Pro Tyr Lys Cys Glu Glu Cys Gly Lys Ala Phe Arg Gln Ser Ser His
65 70 75 80
Leu Thr Thr His Lys Ile Ile His Thr Gly Glu Lys Pro Tyr Lys Cys
85 90 95
Pro Asp Cys Gly Lys Ser Phe Ser Gln Ser Ser Ser Leu Ile Arg His
100 105 110
Gln Arg Thr His Thr Gly Glu Lys Pro Phe Gln Cys Arg Ile Cys Met
115 120 125
Arg Asn Phe Ser Asp Pro Gly His Leu Val Arg His Ile Arg Thr His
130 135 140
Ala Ala Ala Ala Asn Phe Asn Gln Ser Gly Asn Ile Ala Asp Ser Ser
145 150 155 160
Leu Ser Phe Thr Phe Thr Asn Ser Ser Asn Gly Pro Asn Leu Ile Thr
165 170 175
Thr Gln Thr Asn Ser Gln Ala Leu Ser Gln Pro Ile Ala Ser Ser Asn
180 185 190
Val His Asp Asn Phe Met Asn Asn Glu Ile Thr Ala Ser Lys Ile Asp
195 200 205
Asp Gly Asn Asn Ser Lys Pro Leu Ser Pro Gly Trp Thr Asp Gln Thr
210 215 220
Ala Tyr Asn Ala Phe Gly Ile Thr Thr Gly Met Phe Asn Thr Thr Thr
225 230 235 240
Met Asp Asp Val Tyr Asn Tyr Leu Phe Asp Asp Glu Asp Thr Pro Pro
245 250 255
Asn Pro Lys Lys Glu Ile Ser Met Ala Tyr Pro Tyr Asp Val Pro Asp
260 265 270
Tyr Ala Ser
275

<210> 228
 <211> 158
 <212> PRT
 <213> 人工序列

<220>
 <223> 质粒序列

<400> 228

```

Met Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asn Ser Thr Gln
 1      5      10      15
Ala Met Gly Ala Pro Pro Lys Lys Lys Arg Lys Val Gly Ile Arg Ile
 20      25      30
Pro Gly Glu Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser
 35      40      45
Asp Pro Gly Ala Leu Val Arg His Ile Arg Thr His Thr Gly Glu Lys
 50      55      60
Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Arg Ser Asp Thr
 65      70      75      80
Leu Ser Asn His Ile Arg Thr His Thr Gly Glu Lys Pro Tyr Glu Cys
 85      90      95
His Asp Cys Gly Lys Ser Phe Arg Gln Ser Thr His Leu Thr Gln His
100      105      110
Arg Arg Ile His Thr Gly Glu Lys Pro Phe Gln Cys Arg Ile Cys Met
115      120      125
Arg Asn Phe Ser Arg Ser Asp Thr Leu Ser Asn His Ile Arg Thr His
130      135      140
Ala Ala Ala Ala Arg Gly Met His Leu Glu Gly Arg Ile Met
145      150      155

```

<210> 229
 <211> 158
 <212> PRT
 <213> 人工序列

<220>
 <223> 质粒序列

```

<400> 229
Met Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asn Ser Thr Gln
 1      5      10      15
Ala Met Gly Ala Pro Pro Lys Lys Lys Arg Lys Val Gly Ile Arg Ile
 20      25      30
Pro Gly Glu Lys Pro Phe Gln Cys Lys Thr Cys Gln Arg Lys Phe Ser
 35      40      45
Arg Ser Asp His Leu Lys Thr His Thr Arg Thr His Thr Gly Glu Lys
 50      55      60
Pro Tyr Glu Cys His Asp Cys Gly Lys Ser Phe Arg Gln Ser Thr His
 65      70      75      80
Leu Thr Gln His Arg Arg Ile His Thr Gly Glu Lys Pro Tyr Lys Cys
 85      90      95
Glu Glu Cys Gly Lys Ala Phe Arg Gln Ser Ser His Leu Thr Thr His
100      105      110
Lys Ile Ile His Thr Gly Glu Lys Pro Phe Gln Cys Arg Ile Cys Met
115      120      125
Arg Asn Phe Ser Asp Ser Gly Asn Leu Arg Val His Ile Arg Thr His
130      135      140
Ala Ala Ala Ala Arg Gly Met His Leu Glu Gly Arg Ile Met
145      150      155

```

<210> 230
 <211> 130
 <212> PRT
 <213> 人工序列

<220>
 <223> 质粒序列

```

<400> 230
Met Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asn Ser Thr Gln
 1      5      10      15
Ala Met Gly Ala Pro Pro Lys Lys Lys Arg Lys Val Gly Ile Arg Ile
 20      25      30
Pro Gly Glu Lys Pro Phe Gln Cys Lys Thr Cys Gln Arg Lys Phe Ser

```

```

      35      40      45
Arg Ser Asp His Leu Lys Thr His Thr Arg Thr His Thr Gly Glu Lys
  50      55      60
Pro Tyr Glu Cys His Asp Cys Gly Lys Ser Phe Arg Gln Ser Thr His
  65      70      75      80
Leu Thr Gln His Arg Arg Ile His Thr Gly Glu Lys Pro Tyr Lys Cys
      85      90      95
Glu Glu Cys Gly Lys Ala Phe Arg Gln Ser Ser His Leu Thr Thr His
      100      105      110
Lys Ile Ile His Ala Ala Ala Ala Arg Gly Met His Leu Glu Gly Arg
      115      120      125
Ile Met
  130

```

<210> 231

<211> 130

<212> PRT

<213> 人工序列

<220>

<223> 质粒序列

<400> 231

```

Met Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asn Ser Thr Gln
  1      5      10      15
Ala Met Gly Ala Pro Pro Lys Lys Lys Arg Lys Val Gly Ile Arg Ile
      20      25      30
Pro Gly Glu Lys Pro Phe Gln Cys Lys Thr Cys Gln Arg Lys Phe Ser
      35      40      45
Arg Ser Asp His Leu Lys Thr His Thr Arg Thr His Thr Gly Glu Lys
      50      55      60
Pro Tyr Glu Cys Asp His Cys Gly Lys Ser Phe Ser Gln Ser Ser His
      65      70      75      80
Leu Asn Val His Lys Arg Thr His Thr Gly Glu Lys Pro Tyr Glu Cys
      85      90      95
Asp His Cys Gly Lys Ser Phe Ser Gln Ser Ser His Leu Asn Val His
      100      105      110
Lys Arg Thr His Ala Ala Ala Ala Arg Gly Met His Leu Glu Gly Arg
      115      120      125
Ile Met
  130

```

<210> 232

<211> 152

<212> PRT

<213> 酿酒酵母

<400> 232

```

Met Trp Phe Pro Gln Ile Ile Ala Gly Met Ala Ala Gly Gly Ala Ala
  1      5      10      15
Ser Ala Met Thr Pro Gly Lys Val Leu Phe Thr Asn Ala Leu Gly Leu
      20      25      30
Gly Cys Ser Arg Ser Arg Gly Leu Phe Leu Glu Met Phe Gly Thr Ala
      35      40      45
Val Leu Cys Leu Thr Val Leu Met Thr Ala Val Glu Lys Arg Glu Thr
      50      55      60
Asn Phe Met Ala Ala Leu Pro Ile Gly Ile Ser Leu Phe Met Ala His
      65      70      75      80
Met Ala Leu Thr Gly Tyr Thr Gly Thr Gly Val Asn Pro Ala Arg Ser
      85      90      95
Leu Gly Ala Ala Val Ala Ala Arg Tyr Phe Pro His Tyr His Trp Ile
      100      105      110
Tyr Trp Ile Ser Pro Leu Leu Gly Ala Phe Leu Ala Trp Ser Val Trp
      115      120      125
Gln Leu Leu Gln Ile Leu Asp Tyr Thr Thr Tyr Val Asn Ala Glu Lys

```

130 135 140
 Ala Ala Gly Gln Lys Lys Glu Asp
 145 150

 <210> 233
 <211> 273
 <212> PRT
 <213> 白色念珠菌

 <400> 233
 Met Val Ala Glu Ser Ser Ser Ile Asp Asn Thr Pro Asn Asp Val Glu
 1 5 10 15
 Ala Gln Arg Pro Val Tyr Glu Pro Lys Tyr Asp Asp Ser Val Asn Val
 20 25 30
 Ser Pro Leu Lys Asn His Met Ile Ala Phe Leu Gly Glu Phe Phe Gly
 35 40 45
 Thr Phe Ile Phe Leu Trp Val Ala Phe Val Ile Ala Gln Ile Ala Asn
 50 55 60
 Gln Asp Pro Thr Ile Pro Asp Lys Gly Ser Asp Pro Met Gln Leu Ile
 65 70 75 80
 Met Ile Ser Phe Gly Phe Gly Phe Gly Val Met Met Gly Val Phe Met
 85 90 95
 Phe Phe Arg Val Ser Gly Gly Asn Leu Asn Pro Ala Val Thr Leu Thr
 100 105 110
 Leu Val Leu Ala Gln Ala Val Pro Pro Ile Arg Gly Leu Phe Met Met
 115 120 125
 Val Ala Gln Met Ile Ala Gly Met Ala Ala Ala Gly Ala Ala Ser Ala
 130 135 140
 Met Thr Pro Gly Pro Ile Ala Phe Thr Asn Gly Leu Gly Gly Gly Ala
 145 150 155 160
 Ser Lys Ala Arg Gly Val Phe Leu Glu Ala Phe Gly Thr Cys Ile Leu
 165 170 175
 Cys Leu Thr Val Leu Met Met Ala Val Glu Lys Ser Arg Ala Thr Phe
 180 185 190
 Met Ala Pro Phe Val Ile Gly Ile Ser Leu Phe Leu Gly His Leu Ile
 195 200 205
 Cys Val Tyr Tyr Thr Gly Ala Gly Leu Asn Pro Ala Arg Ser Phe Gly
 210 215 220
 Pro Cys Val Ala Ala Arg Ser Phe Pro Val Tyr His Trp Ile Tyr Trp
 225 230 235 240
 Val Gly Pro Ile Leu Gly Ser Val Ile Ala Phe Ala Ile Trp Lys Ile
 245 250 255
 Phe Lys Ile Leu Lys Tyr Glu Thr Cys Asn Pro Gly Gln Asp Ser Asp
 260 265 270
 Ala

<210> 234
 <211> 12
 <212> DNA
 <213> 人工序列

<220>
 <223> 识别元件

<400> 234
 ggggcwrigag gg

12

<210> 235
 <211> 12
 <212> DNA
 <213> 人工序列

<220>
 <223> 靶序列

<221> 其他_特征	
<222> (1)...(12)	
<223> n = A,T,C 或 G	
<400> 235	
gctgranggg ah	12
<210> 236	
<211> 31	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 236	
gacaaccggt catcgataag ctaattctca c	31
<210> 237	
<211> 29	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 237	
ttgtccatgg acgctgtttc ctggtgaaa	29
<210> 238	
<211> 12	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<400> 238	
daadaaaath ga	12
<210> 239	
<211> 13	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<221> 其他_特征	
<222> (1)...(13)	
<223> n = A,T,C 或 G	
<400> 239	
gyagrahgan ggk	13
<210> 240	
<211> 12	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<400> 240	

hgaaathgag gt	12
<210> 241	
<211> 12	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<400> 241	
gragragggg ra	12
<210> 242	
<211> 12	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<221> 其他_特征	
<222> (1)...(12)	
<223> n = A,T,C 或 G	
<400> 242	
grahganggg tc	12
<210> 243	
<211> 12	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<400> 243	
gragragggh ga	12
<210> 244	
<211> 12	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<400> 244	
gavgaaaath ga	12
<210> 245	
<211> 12	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<221> 其他_特征	
<222> (1)...(12)	
<223> n = A,T,C 或 G	
<400> 245	
ngggyagraa at	12

<210> 246	
<211> 13	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<221> 其他_特征	
<222> (1)...(13)	
<223> n = A,T,C 或 G	
<400> 246	
gaagrahgan ggk	13
<210> 247	
<211> 12	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<221> 其他_特征	
<222> (1)...(12)	
<223> n = A,T,C 或 G	
<400> 247	
gradaanggg tc	12
<210> 248	
<211> 12	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<221> 其他_特征	
<222> (1)...(12)	
<223> n = A,T,C 或 G	
<400> 248	
ggtnggggtcg ya	12
<210> 249	
<211> 12	
<212> DNA	
<213> 人工序列	
<220>	
<223> 靶序列	
<400> 249	
daagaaaacg ct	12
<210> 250	
<211> 24	
<212> PRT	
<213> 人工序列	
<220>	
<223> 纯化的多肽	
<221> 变体	

<222> 11
 <223> Xaa = Phe 或 Tyr

 <221> 变体
 <222> 14
 <223> Xaa = 疏水残基

 <221> 变体
 <222> 2-6, 8-10, 12, 15, 19-23
 <223> Xaa = 任何氨基酸

 <400> 250
 Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa Xaa Ser
 1 5 10 15
 Asn His Xaa Xaa Xaa Xaa Xaa His
 20

<210> 251
 <211> 24
 <212> PRT
 <213> 人工序列

 <220>
 <223> 纯化的多肽

 <221> 变体
 <222> 11
 <223> Xaa = Phe 或 Tyr

 <221> 变体
 <222> 14
 <223> Xaa = 疏水残基

 <221> 变体
 <222> 2-6, 8-10, 12, 15, 19-23
 <223> Xaa = 任何氨基酸

 <400> 251
 Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa Xaa Ser
 1 5 10 15
 Asn His Xaa Xaa Xaa Xaa Xaa His
 20

<210> 252
 <211> 24
 <212> PRT
 <213> 人工序列

 <220>
 <223> 纯化的多肽

 <221> 变体
 <222> 11
 <223> Xaa = Phe 或 Tyr

 <221> 变体
 <222> 14
 <223> Xaa = 疏水残基

 <221> 变体
 <222> 2-6, 8-10, 12, 15, 19-23
 <223> Xaa = 任何氨基酸

 <400> 252
 Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Ser

1 5
Asn His Xaa Xaa Xaa Xaa Xaa His
 20

<210> 253
<211> 84
<212> PRT
<213> 人工序列

<220>
<223> 纯化的多肽

<221> 变体
<222> 11, 41, 71
<223> Xaa = Phe 或 Tyr

<221> 变体
<222> 14, 44, 74
<223> Xaa = 疏水残基

<221> 变体
<222> 2-6, 8-10, 12, 15, 19-23, 25-30, 32-36, 38-40, 42, 45, 49-53,
55-60, 62-66, 68-70, 72, 75, 79-83
<223> Xaa = 任何氨基酸

[illegible]

<210> 254
<211> 24
<212> PRT
<213> 人工序列

<220>
<223> 纯化的多肽

<221> 变体
<222> 11
<223> Xaa = Phe 或 Tyr

<221> 变体
<222> 14
<223> Xaa = 疏水残基

<221> 变体
<222> 2-6, 8-10, 12, 15, 19-23
<223> Xaa = 任何氨基酸

```
<400> 254
Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Arg Xaa Xaa Asp
 1          5          10          15
Lys His Xaa Xaa Xaa Xaa Xaa His
      20
```

<210> 255
 <211> 24
 <212> PRT
 <213> 人工序列

<220>
 <223> 纯化的多肽

<221> 变体
 <222> 11
 <223> Xaa = Phe 或 Tyr

<221> 变体
 <222> 14
 <223> Xaa = 疏水残基

<221> 变体
 <222> 2-6, 8-10, 12, 15, 19-23
 <223> Xaa = 任何氨基酸

<400> 255
 Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa Xaa Thr
 1 5 10 15
 His His Xaa Xaa Xaa Xaa Xaa His
 20

<210> 256
 <211> 24
 <212> PRT
 <213> 人工序列

<220>
 <223> 纯化的多肽

<221> 变体
 <222> 11
 <223> Xaa = Phe 或 Tyr

<221> 变体
 <222> 14
 <223> Xaa = 疏水残基

<221> 变体
 <222> 2-6, 8-10, 12, 15, 19-23
 <223> Xaa = 任何氨基酸

<400> 256
 Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Val Xaa Xaa Ser
 1 5 10 15
 Thr His Xaa Xaa Xaa Xaa Xaa His
 20

<210> 257
 <211> 24
 <212> PRT
 <213> 人工序列

<220>
 <223> 纯化的多肽

<221> 变体
 <222> 11
 <223> Xaa = Phe 或 Tyr

<221> 变体

<222> 14

<223> Xaa = 疏水残基

<221> 变体

<222> 2-6, 8-10, 12, 15, 19-23

<223> Xaa = 任何氨基酸

<400> 257

Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Arg Xaa Xaa Asp

1

5

10

15

Lys His Xaa Xaa Xaa Xaa Xaa His

20

<210> 258

<211> 114

<212> PRT

<213> 人工序列

<220>

<223> 纯化的多肽

<221> 变体

<222> 11, 41, 71, 101

<223> Xaa = Phe 或 Tyr

<221> 变体

<222> 14, 44, 74, 104

<223> Xaa = 疏水残基

<221> 变体

<222> 2-6, 8-10, 12, 15, 19-23, 25-30, 32-36, 38-40, 42, 45, 49-53,
55-60, 62-66, 68-70, 72, 75, 79-83, 85-90, 92-96, 98-100, 102,
105, 109-113

<223> Xaa = 任何氨基酸

<400> 258

Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Arg Xaa Xaa Asp

1

5

10

15

Lys His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa Cys Xaa

20

25

30

Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Gln Xaa Xaa Thr His His

35

40

45

Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa

50

55

60

Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Val Xaa Xaa Ser Thr His Xaa Xaa

65

70

75

80

Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa

85

90

95

Cys Xaa Xaa Xaa Xaa Xaa Arg Xaa Xaa Asp Lys His Xaa Xaa Xaa Xaa

100

105

110

Xaa His

<210> 259

<211> 630

<212> DNA

<213> 人工序列

<220>

<223> 质粒序列

<221> CDS

<222> (1)...(627)

<400> 259

atg gtg tac ccc tac gac gtg ccc gac tac gcc gaa ttg cct cca aaa	48
Met Val Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Glu Leu Pro Pro Lys	
1 5 10 15	
aag aag aga aag gta ggg atc cga att ccc ggg gaa aaa ccg ttc cag	96
Lys Lys Arg Lys Val Gly Ile Arg Ile Pro Gly Glu Lys Pro Phe Gln	
20 25 30	
tgt aaa act tgt cag cga aag ttc tcc cgg tcc gac cac ctg aag acc	144
Cys Lys Thr Cys Gln Arg Lys Phe Ser Arg Ser Asp His Leu Lys Thr	
35 40 45	
cac acc agg act cat acc ggg gaa aaa ccg tat gtt tgc tca aaa tgt	192
His Thr Arg Thr His Thr Gly Glu Lys Pro Tyr Val Cys Ser Lys Cys	
50 55 60	
ggg aaa gcc ttc act cag agt tca aat ctg act gta cat caa aaa atc	240
Gly Lys Ala Phe Thr Gln Ser Ser Asn Leu Thr Val His Gln Lys Ile	
65 70 75 80	
cac acc ggg gaa aaa ccg tat gag tgt cac gat tgc gga aag tcc ttt	288
His Thr Gly Glu Lys Pro Tyr Glu Cys His Asp Cys Gly Lys Ser Phe	
85 90 95	
agg cag agc acc cac ctc act cgg cac cgg agg atc cac acc ggg gaa	336
Arg Gln Ser Thr His Leu Thr Arg His Arg Arg Ile His Thr Gly Glu	
100 105 110	
aaa ccg tat aag tgt cat caa tgt ggg aaa gcc ttt att caa tcc ttt	384
Lys Pro Tyr Lys Cys His Gln Cys Gly Lys Ala Phe Ile Gln Ser Phe	
115 120 125	
aac ctt cga aga cat gag aga act cac acc ggt gaa aaa gcg gcc gct	432
Asn Leu Arg Arg His Glu Arg Thr His Thr Gly Glu Lys Ala Ala Ala	
130 135 140	
aaa ttc gtg tca gtg aca ttt gaa gat gtg gct gtg ctc ttt act cgg	480
Lys Phe Val Ser Val Thr Phe Glu Asp Val Ala Val Leu Phe Thr Arg	
145 150 155 160	
gac gag tgg aag aag ctg gat ctg tct cag aga agc ctg tac cgt gag	528
Asp Glu Trp Lys Lys Leu Asp Leu Ser Gln Arg Ser Leu Tyr Arg Glu	
165 170 175	
gtg atg ctg gag aat tac agc aac ctg gcc tcc atg gca gga ttc ctg	576
Val Met Leu Glu Asn Tyr Ser Asn Leu Ala Ser Met Ala Gly Phe Leu	
180 185 190	
ttt acc aaa cca aag gtg atc tcc ctg ttg cag caa gga gag gat ccc	624
Phe Thr Lys Pro Lys Val Ile Ser Leu Leu Gln Gln Gly Glu Asp Pro	
195 200 205	
tgg taa	630
Trp	

<210> 260

<211> 209

<212> PRT

<213> 人工序列

<220>

<223> 质粒序列

<400> 260

Met Val Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Glu Leu Pro Pro Lys
 1 5 10 15
 Lys Lys Arg Lys Val Gly Ile Arg Ile Pro Gly Glu Lys Pro Phe Gln
 20 25 30
 Cys Lys Thr Cys Gln Arg Lys Phe Ser Arg Ser Asp His Leu Lys Thr
 35 40 45
 His Thr Arg Thr His Thr Gly Glu Lys Pro Tyr Val Cys Ser Lys Cys
 50 55 60
 Gly Lys Ala Phe Thr Gln Ser Ser Asn Leu Thr Val His Gln Lys Ile
 65 70 75 80
 His Thr Gly Glu Lys Pro Tyr Glu Cys His Asp Cys Gly Lys Ser Phe
 85 90 95
 Arg Gln Ser Thr His Leu Thr Arg His Arg Arg Ile His Thr Gly Glu
 100 105 110
 Lys Pro Tyr Lys Cys His Gln Cys Gly Lys Ala Phe Ile Gln Ser Phe
 115 120 125
 Asn Leu Arg Arg His Glu Arg Thr His Thr Gly Glu Lys Ala Ala Ala
 130 135 140
 Lys Phe Val Ser Val Thr Phe Glu Asp Val Ala Val Leu Phe Thr Arg
 145 150 155 160
 Asp Glu Trp Lys Lys Leu Asp Leu Ser Gln Arg Ser Leu Tyr Arg Glu
 165 170 175
 Val Met Leu Glu Asn Tyr Ser Asn Leu Ala Ser Met Ala Gly Phe Leu
 180 185 190
 Phe Thr Lys Pro Lys Val Ile Ser Leu Leu Gln Gln Gly Glu Asp Pro
 195 200 205
 Trp

<210> 261
 <211> 546
 <212> DNA
 <213> 人工序列

<220>
 <223> 质粒序列

<221> CDS
 <222> (1)...(543)

<400> 261
 atg gtg tac ccc tac gac gtg ccc gac tac gcc gaa ttg cct cca aaa 48
 Met Val Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Glu Leu Pro Pro Lys
 1 5 10 15
 aag aag aga aag gta ggg atc cga att ccc ggg gaa aaa ccg tat gag 96
 Lys Lys Arg Lys Val Gly Ile Arg Ile Pro Gly Glu Lys Pro Tyr Glu
 20 25 30
 tgt gat cac tgt gga aaa tcc ttt agc cag agc tct cat ctg aat gtg 144
 Cys Asp His Cys Gly Lys Ser Phe Ser Gln Ser Ser His Leu Asn Val
 35 40 45
 cac aaa aga act cac acc ggg gaa aaa ccg tat aag tgc cct gat tgt 192
 His Lys Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Asp Cys
 50 55 60
 ggg aag agt ttt agt cag agt tcc agc ctc att cgc cac cag cgg aca 240
 Gly Lys Ser Phe Ser Gln Ser Ser Ser Leu Ile Arg His Gln Arg Thr
 65 70 75 80
 cac acc ggg gaa aaa ccg tat gag tgt cac gat tgc gga aag tcc ttt 288
 His Thr Gly Glu Lys Pro Tyr Glu Cys His Asp Cys Gly Lys Ser Phe
 85 90 95

agg cag agc acc cac ctc act cgg cac cgg agg atc cac acc ggt gaa	336
Arg Gln Ser Thr His Leu Thr Arg His Arg Arg Ile His Thr Gly Glu	
100 105 110	
aaa gcg gcc gct aaa ttc gtg tca gtg aca ttt gaa gat gtg gct gtg	384
Lys Ala Ala Ala Lys Phe Val Ser Val Thr Phe Glu Asp Val Ala Val	
115 120 125	
ctc ttt act cgg gac gag tgg aag aag ctg gat ctg tct cag aga agc	432
Leu Phe Thr Arg Asp Glu Trp Lys Lys Leu Asp Leu Ser Gln Arg Ser	
130 135 140	
ctg tac cgt gag gtg atg ctg gag aat tac agc aac ctg gcc tcc atg	480
Leu Tyr Arg Glu Val Met Leu Glu Asn Tyr Ser Asn Leu Ala Ser Met	
145 150 155 160	
gca gga ttc ctg ttt acc aaa cca aag gtg atc tcc ctg ttg cag caa	528
Ala Gly Phe Leu Phe Thr Lys Pro Lys Val Ile Ser Leu Leu Gln Gln	
165 170 175	
gga gag gat ccc tgg taa	546
Gly Glu Asp Pro Trp	
180	

<210> 262
 <211> 181
 <212> PRT
 <213> 人工序列

<220>
 <223> 质粒序列

<400> 262

Met Val Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Glu Leu Pro Pro Lys	
1 5 10 15	
Lys Lys Arg Lys Val Gly Ile Arg Ile Pro Gly Glu Lys Pro Tyr Glu	
20 25 30	
Cys Asp His Cys Gly Lys Ser Phe Ser Gln Ser Ser His Leu Asn Val	
35 40 45	
His Lys Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Asp Cys	
50 55 60	
Gly Lys Ser Phe Ser Gln Ser Ser Ser Leu Ile Arg His Gln Arg Thr	
65 70 75 80	
His Thr Gly Glu Lys Pro Tyr Glu Cys His Asp Cys Gly Lys Ser Phe	
85 90 95	
Arg Gln Ser Thr His Leu Thr Arg His Arg Arg Ile His Thr Gly Glu	
100 105 110	
Lys Ala Ala Ala Lys Phe Val Ser Val Thr Phe Glu Asp Val Ala Val	
115 120 125	
Leu Phe Thr Arg Asp Glu Trp Lys Lys Leu Asp Leu Ser Gln Arg Ser	
130 135 140	
Leu Tyr Arg Glu Val Met Leu Glu Asn Tyr Ser Asn Leu Ala Ser Met	
145 150 155 160	
Ala Gly Phe Leu Phe Thr Lys Pro Lys Val Ile Ser Leu Leu Gln Gln	
165 170 175	
Gly Glu Asp Pro Trp	
180	

<210> 263
 <211> 636
 <212> DNA
 <213> 人工序列

<220>
 <223> 质粒序列

<221> CDS

<222> (1)...(633)

<400> 263

```

atg gtg tac ccc tac gac gtg ccc gac tac gcc gaa ttg cct cca aaa      48
Met Val Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Glu Leu Pro Pro Lys
  1              5              10              15

```

```

aag aag aga aag gta ggg atc cga att ccc ggg gaa aaa ccg ttt gag      96
Lys Lys Arg Lys Val Gly Ile Arg Ile Pro Gly Glu Lys Pro Phe Glu
              20              25              30

```

```

tgt aaa gat tgc ggg aaa gct ttc att cag aag tca aac ctc atc aga     144
Cys Lys Asp Cys Gly Lys Ala Phe Ile Gln Lys Ser Asn Leu Ile Arg
              35              40              45

```

```

cac cag aga act cac acc ggg gaa aaa ccg tat aag tgc cct gat tgt     192
His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Asp Cys
              50              55              60

```

```

ggg aag agt ttt agt cag agt tcc agc ctc att cgc cac cag cgg aca     240
Gly Lys Ser Phe Ser Gln Ser Ser Ser Leu Ile Arg His Gln Arg Thr
  65              70              75              80

```

```

cac acc ggg gaa aaa ccg tat gag tgt cac gat tgc gga aag tcc ttt     288
His Thr Gly Glu Lys Pro Tyr Glu Cys His Asp Cys Gly Lys Ser Phe
              85              90              95

```

```

agg cag agc acc cac ctc act cgg cac cgg agg atc cac acc ggg gaa     336
Arg Gln Ser Thr His Leu Thr Arg His Arg Arg Ile His Thr Gly Glu
              100              105              110

```

```

aaa ccg tat gta tgc gat gta gag gga tgt acg tgg aaa ttt gcc cgc     384
Lys Pro Tyr Val Cys Asp Val Glu Gly Cys Thr Trp Lys Phe Ala Arg
              115              120              125

```

```

tca gat aag ctc aac aga cac aag aaa agg cac acc ggt gaa aaa gcg     432
Ser Asp Lys Leu Asn Arg His Lys Lys Arg His Thr Gly Glu Lys Ala
              130              135              140

```

```

gcc gct aaa ttc gtg tca gtg aca ttt gaa gat gtg gct gtg ctc ttt     480
Ala Ala Lys Phe Val Ser Val Thr Phe Glu Asp Val Ala Val Leu Phe
  145              150              155              160

```

```

act cgg gac gag tgg aag aag ctg gat ctg tct cag aga agc ctg tac     528
Thr Arg Asp Glu Trp Lys Lys Leu Asp Leu Ser Gln Arg Ser Leu Tyr
              165              170              175

```

```

cgt gag gtg atg ctg gag aat tac agc aac ctg gcc tcc atg gca gga     576
Arg Glu Val Met Leu Glu Asn Tyr Ser Asn Leu Ala Ser Met Ala Gly
              180              185              190

```

```

ttc ctg ttt acc aaa cca aag gtg atc tcc ctg ttg cag caa gga gag     624
Phe Leu Phe Thr Lys Pro Lys Val Ile Ser Leu Leu Gln Gln Gly Glu
              195              200              205

```

```

gat ccc tgg taa
Asp Pro Trp
  210

```

<210> 264

<211> 211

<212> PRT

<213> 人工序列

<220>

<223> 质粒序列

<400> 264

```

Met Val Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Glu Leu Pro Pro Lys
 1          5          10          15
Lys Lys Arg Lys Val Gly Ile Arg Ile Pro Gly Glu Lys Pro Phe Glu
          20          25          30
Cys Lys Asp Cys Gly Lys Ala Phe Ile Gln Lys Ser Asn Leu Ile Arg
          35          40          45
His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Asp Cys
          50          55          60
Gly Lys Ser Phe Ser Gln Ser Ser Ser Leu Ile Arg His Gln Arg Thr
65          70          75          80
His Thr Gly Glu Lys Pro Tyr Glu Cys His Asp Cys Gly Lys Ser Phe
          85          90          95
Arg Gln Ser Thr His Leu Thr Arg His Arg Arg Ile His Thr Gly Glu
          100          105          110
Lys Pro Tyr Val Cys Asp Val Glu Gly Cys Thr Trp Lys Phe Ala Arg
          115          120          125
Ser Asp Lys Leu Asn Arg His Lys Lys Arg His Thr Gly Glu Lys Ala
          130          135          140
Ala Ala Lys Phe Val Ser Val Thr Phe Glu Asp Val Ala Val Leu Phe
145          150          155          160
Thr Arg Asp Glu Trp Lys Lys Leu Asp Leu Ser Gln Arg Ser Leu Tyr
          165          170          175
Arg Glu Val Met Leu Glu Asn Tyr Ser Asn Leu Ala Ser Met Ala Gly
          180          185          190
Phe Leu Phe Thr Lys Pro Lys Val Ile Ser Leu Leu Gln Gln Gly Glu
          195          200          205
Asp Pro Trp
210

```

<210> 265

<211> 128

<212> PRT

<213> 人工序列

<220>

<223> 纯化的多肽

<221> 变体

<222> (1)...(128)

<223> Xaa = 任何氨基酸

<400> 265

```

Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa Ser His
 1          5          10          15
Xaa Xaa Val His Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
          20          25          30
Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Gln Xaa
          35          40          45
Ser Asn Xaa Xaa Ile His Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa
          50          55          60
Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa
65          70          75          80
Gln Xaa Thr His Xaa Xaa Arg His Xaa Xaa Xaa Xaa His Xaa Xaa Xaa
          85          90          95
Xaa Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa
          100          105          110
Xaa Xaa Cys Xaa Ser Asn Xaa Xaa Arg His Xaa Xaa Xaa Xaa Xaa His
          115          120          125

```

<210> 266

<211> 128

<212> PRT

<213> 人工序列

<220>

<223> 纯化的多肽

<221> 变体

<222> (1)...(128)

<223> Xaa = 任何氨基酸

<400> 266

Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Gln	Xaa	Ser	His
1				5					10					15	
Xaa	Xaa	Val	His	Xaa	Xaa	Xaa	Xaa	Xaa	His	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa
			20					25					30		
Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Val	Xaa
		35					40					45			
Ser	Thr	Xaa	Xaa	Arg	His	Xaa	Xaa	Xaa	Xaa	Xaa	His	Xaa	Xaa	Xaa	Xaa
	50				55						60				
Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa
65					70					75					80
Arg	Xaa	Asp	Asn	Xaa	Xaa	Gln	His	Xaa	Xaa	Xaa	Xaa	Xaa	His	Xaa	Xaa
			85						90					95	
Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa	Xaa	Xaa	Cys	Xaa	Xaa	Xaa
			100					105					110		
Xaa	Xaa	Gln	Xaa	Thr	His	Xaa	Xaa	Arg	His	Xaa	Xaa	Xaa	Xaa	Xaa	His
		115						120					125		

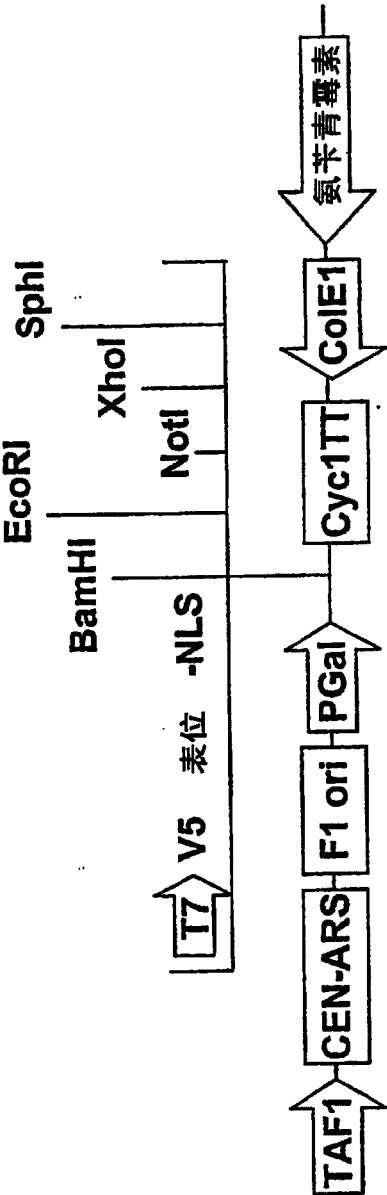


图1

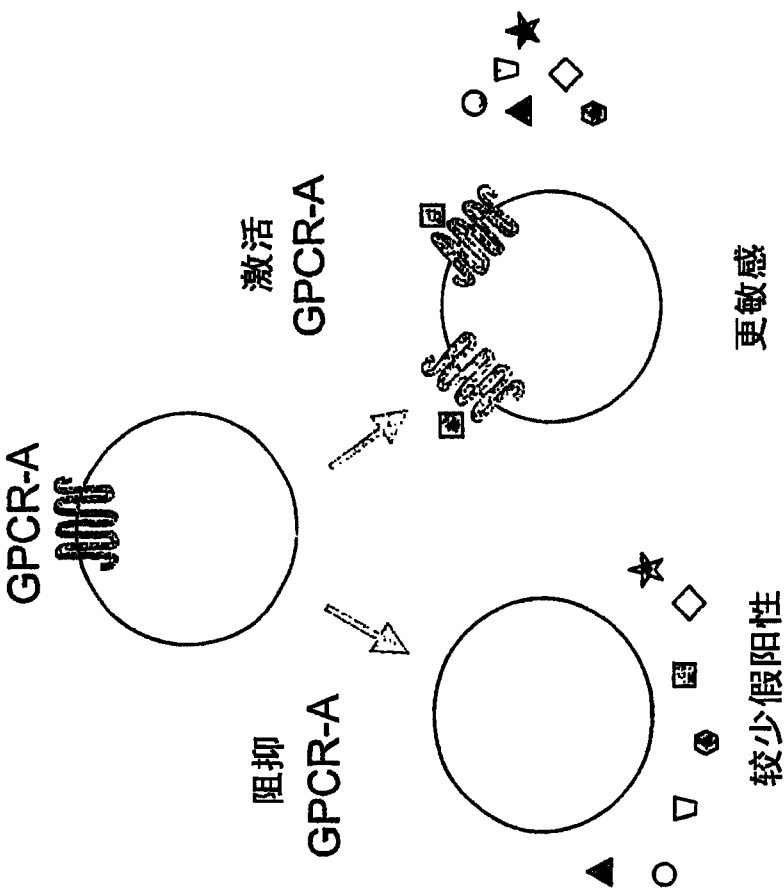


图2

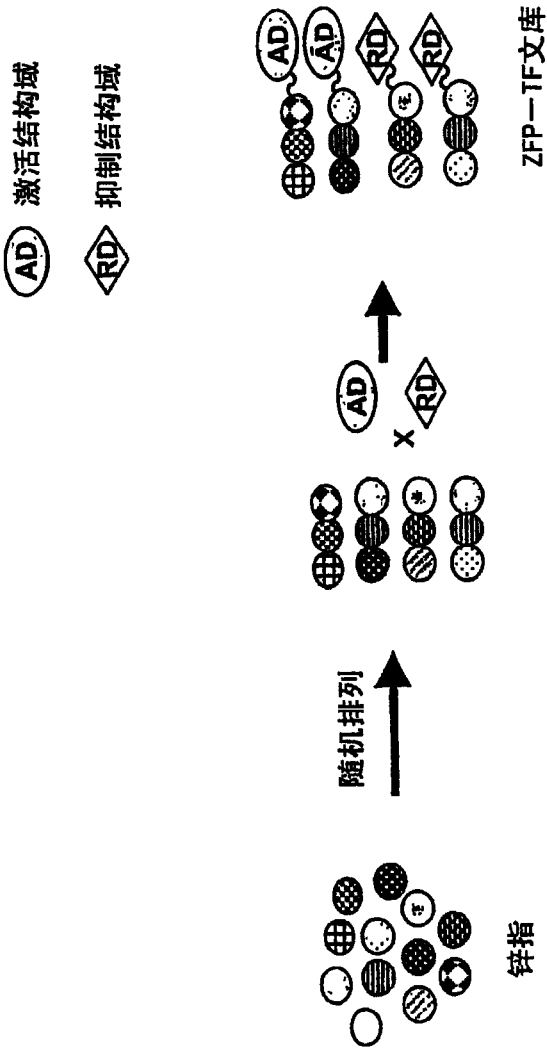


图3A

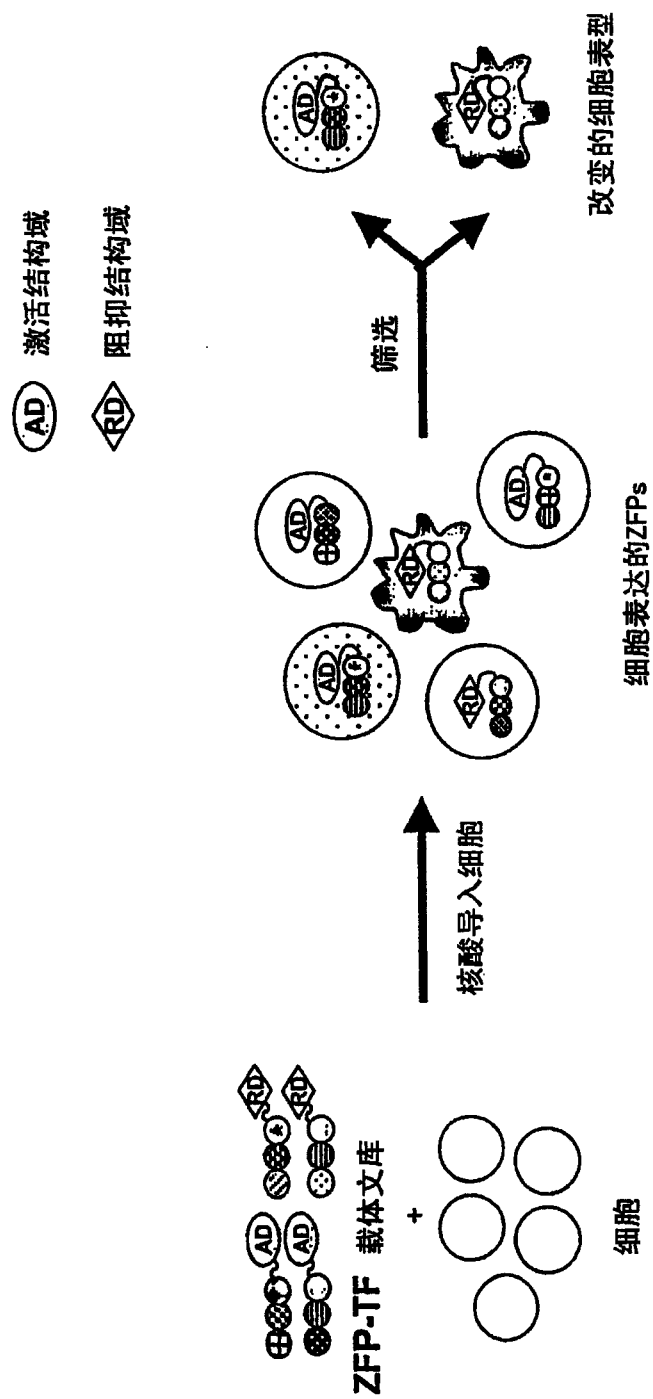


图3B

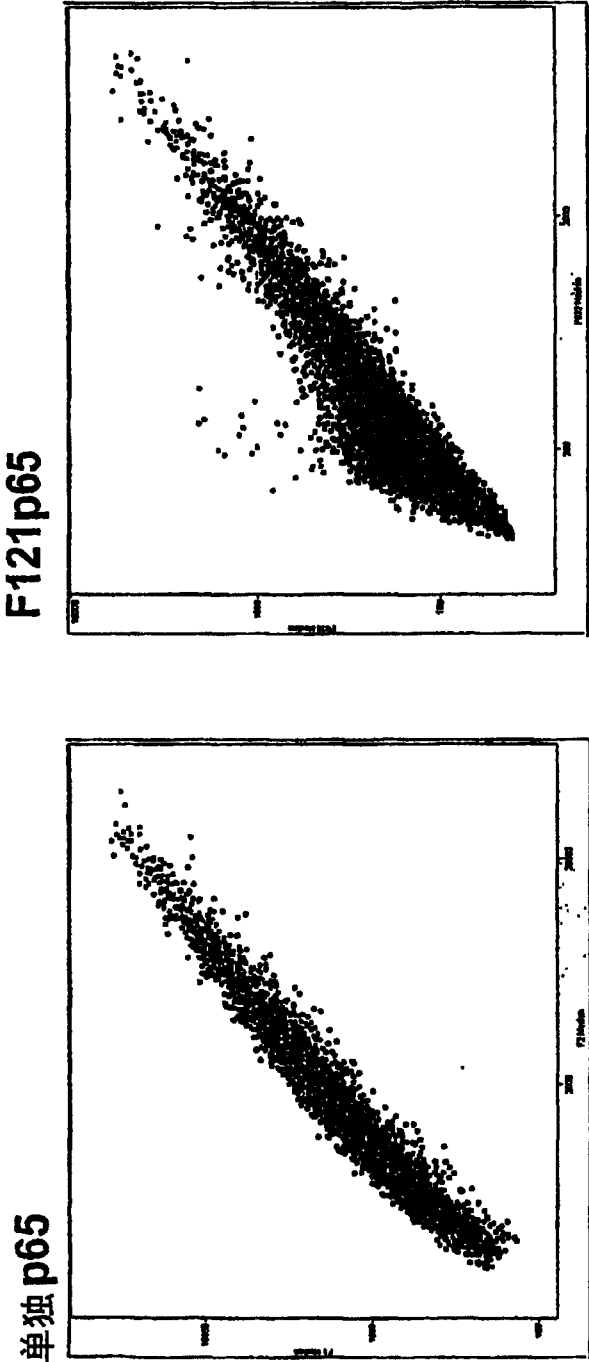


图4

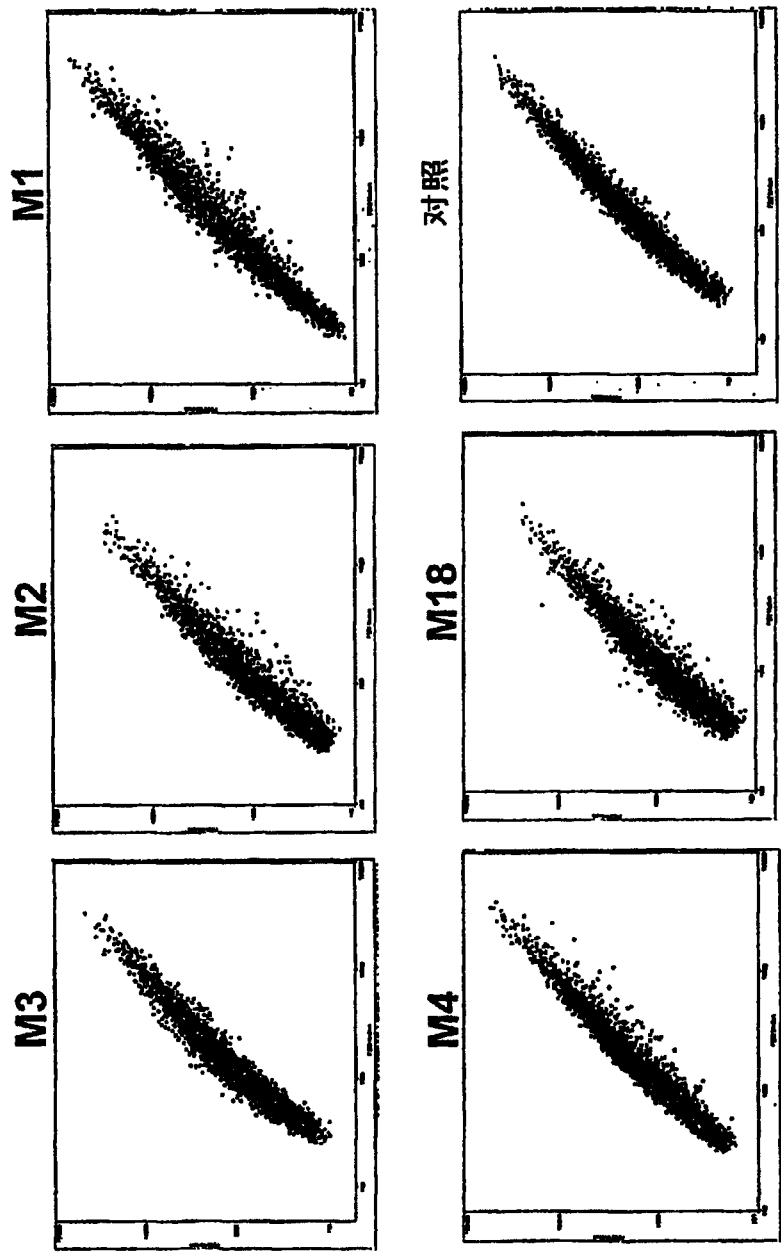
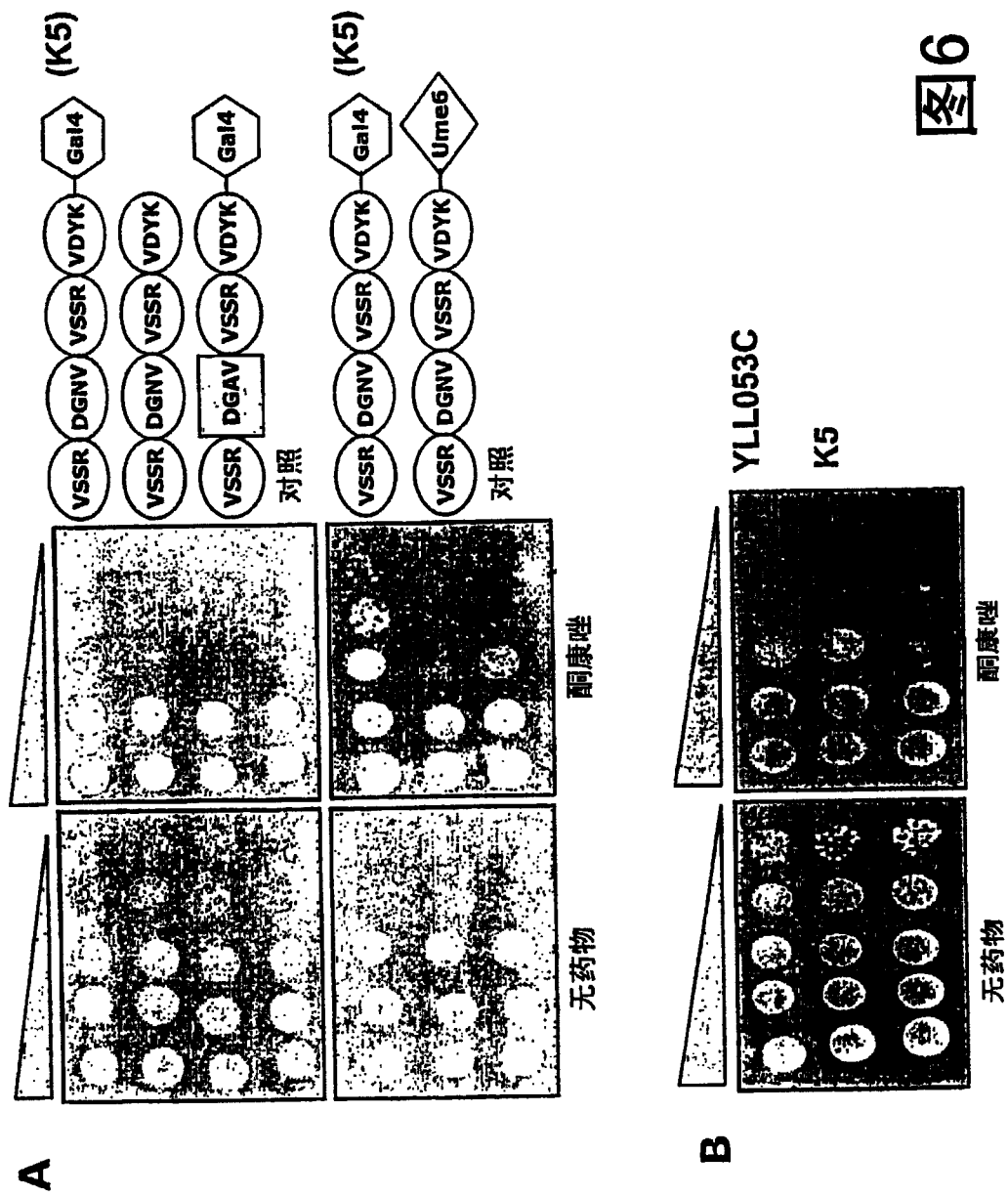


图5



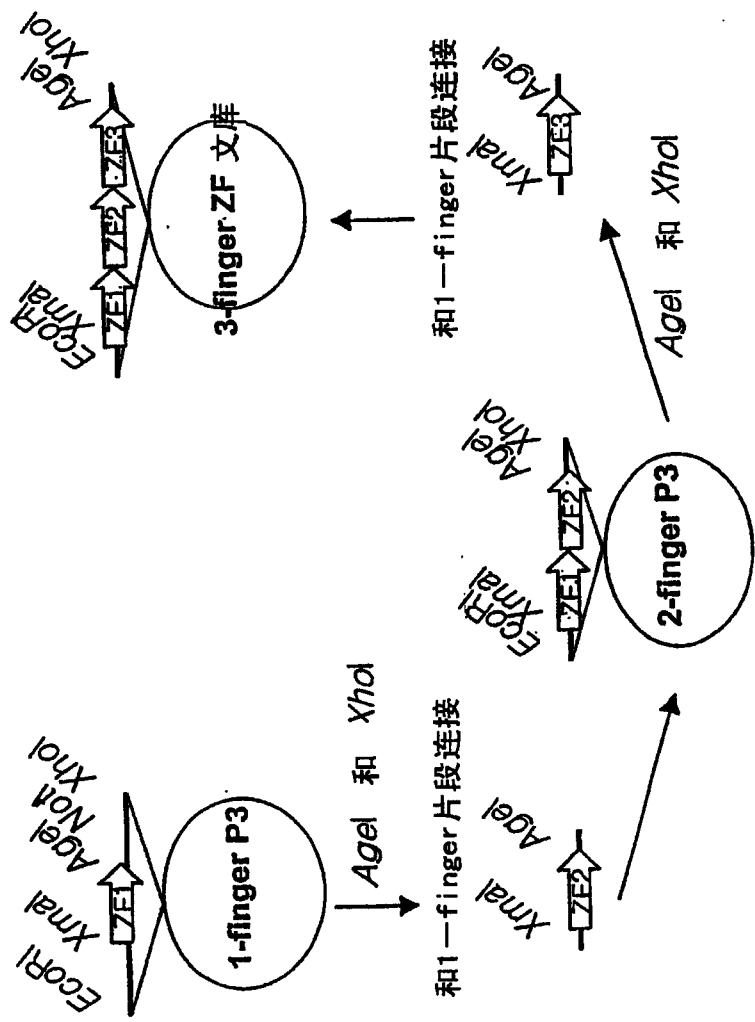


图7

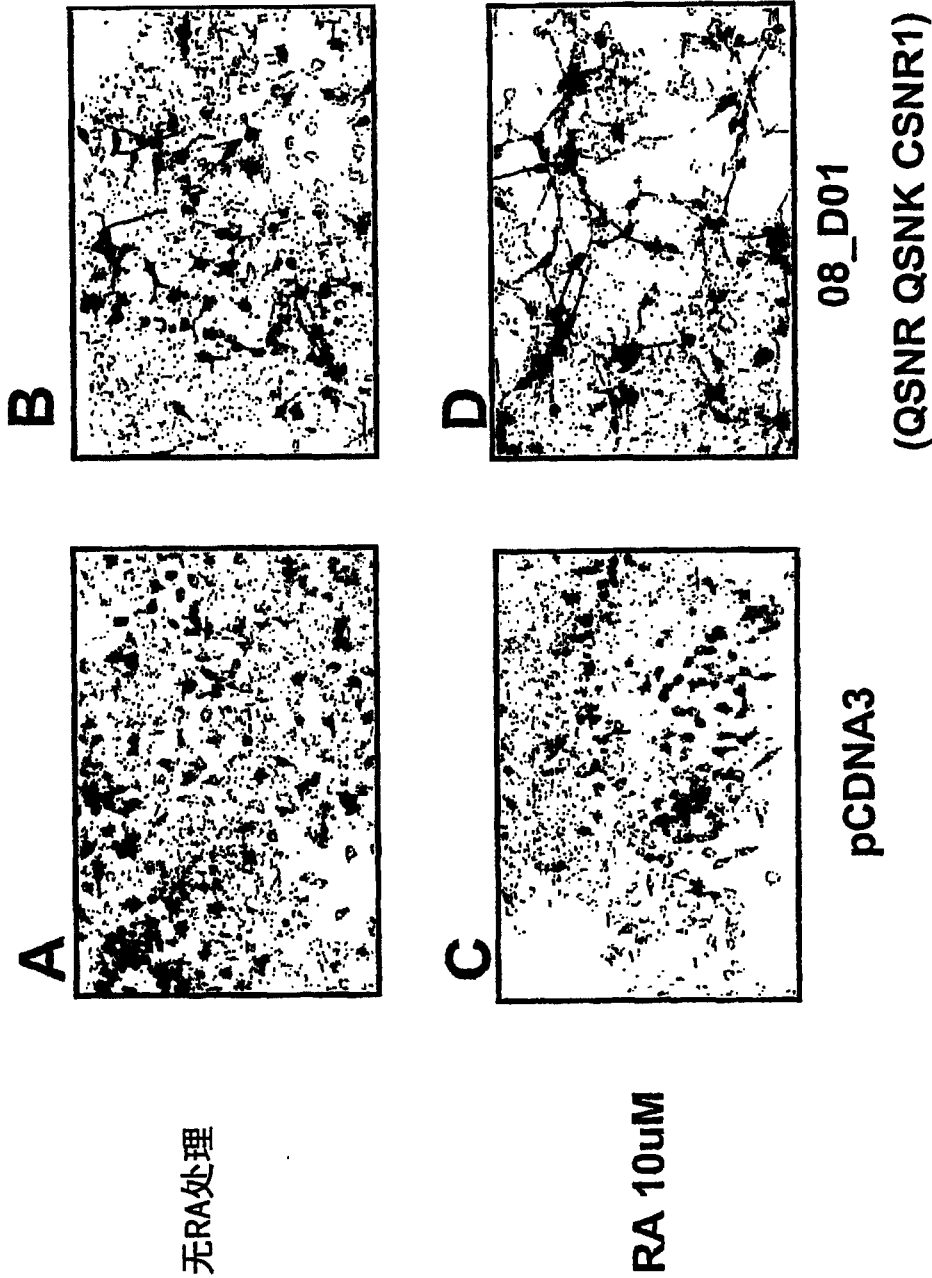


图8

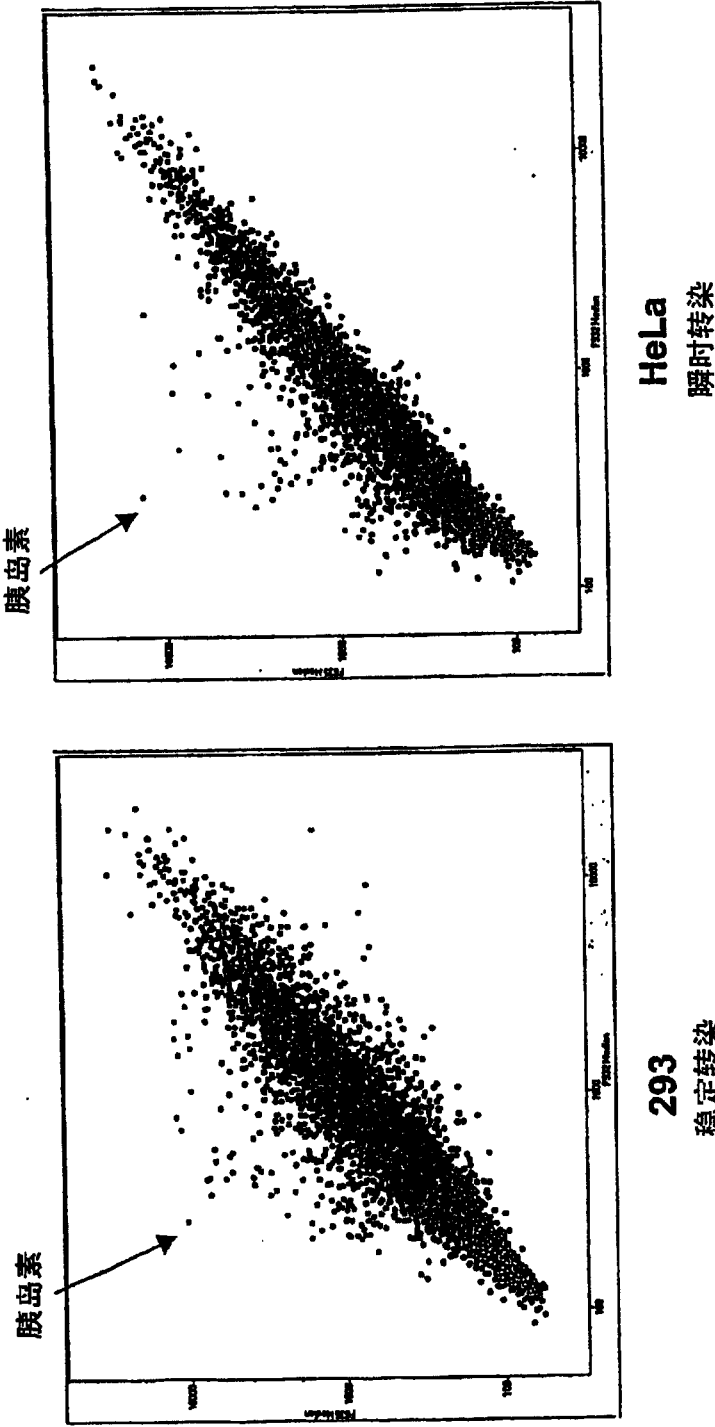


图9

锌指—DNA相互作用的部分目录

3—bp序列	锌指
GAA	QSNR
GAC	CSNR
GAG	RSNR
GAT	ISNR
GCA	QSSR
GCC	-
GCG	RDER
GCT	VSTR

图10A

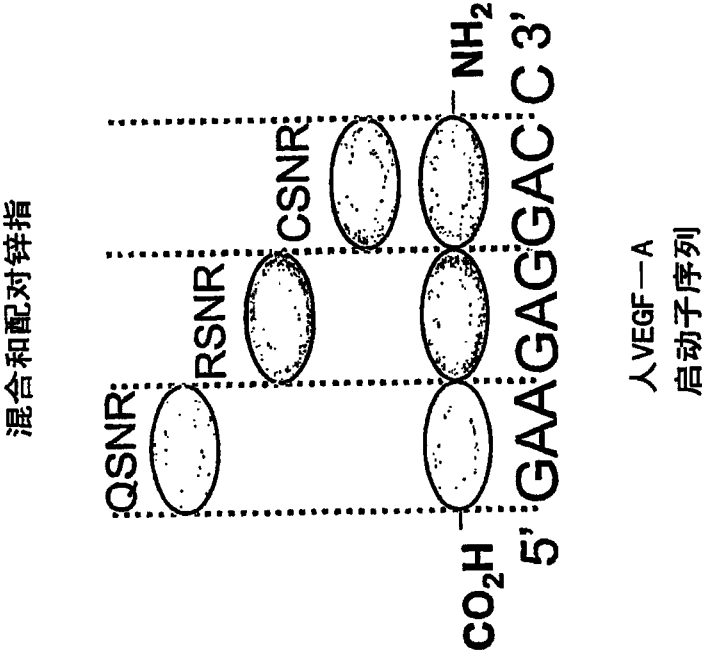


图10B

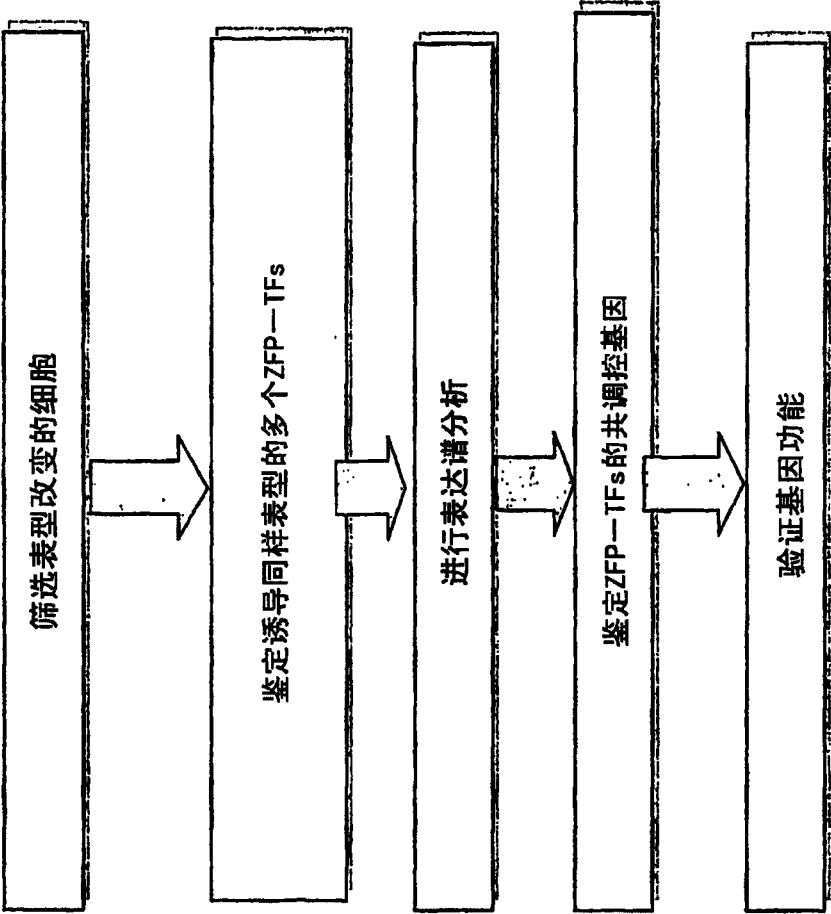


图11

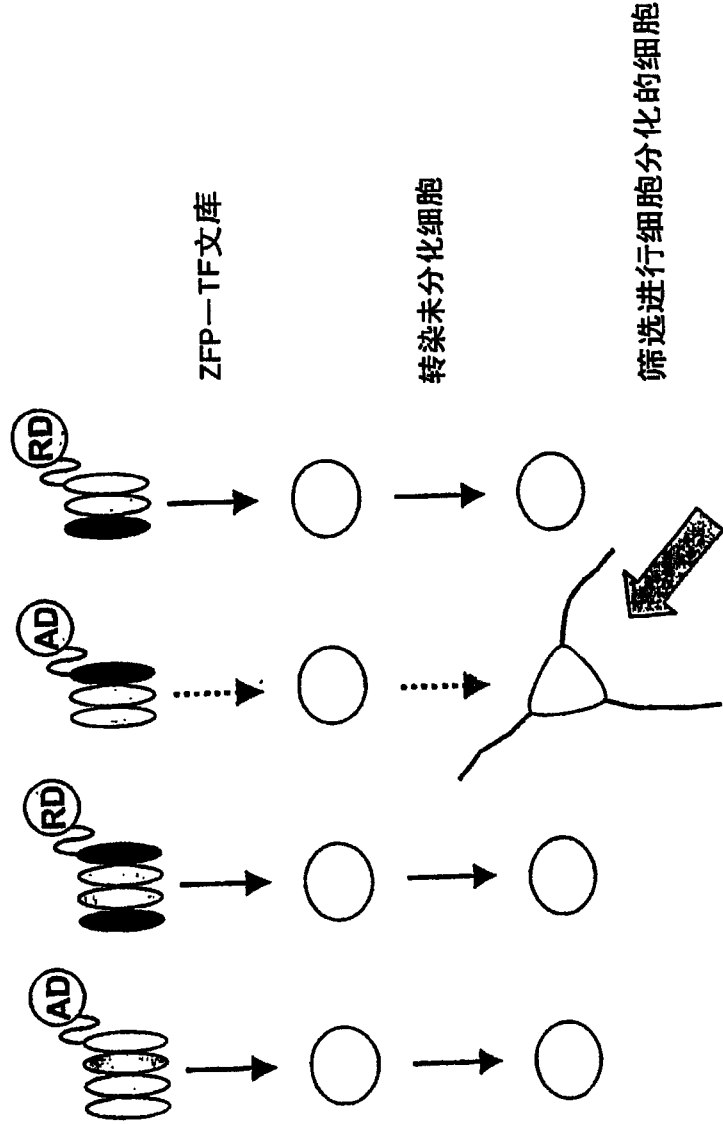


图12

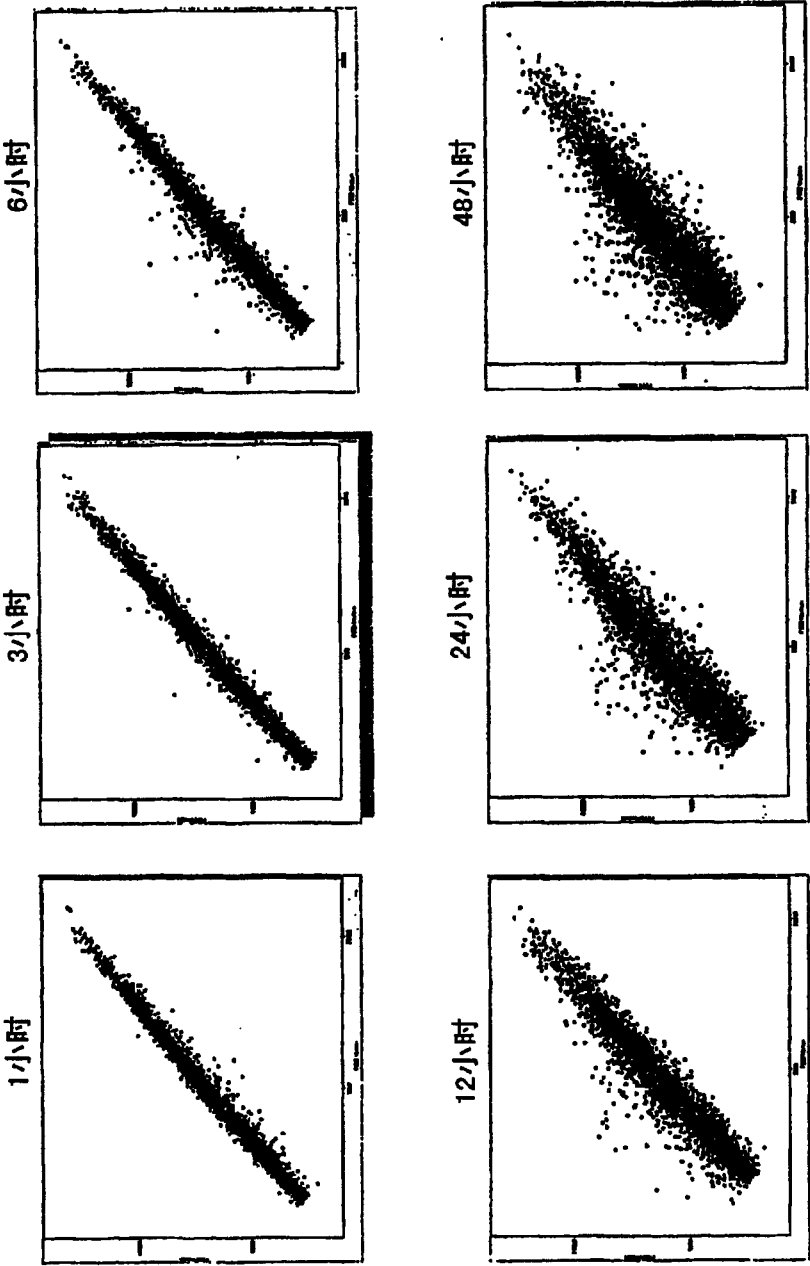


图13

NEURO1-P65 (也称作 08_D1)

QSNR1-QSNK-CSNR1-p65

核酸序列

ATGGTGTACCCCTACGACGTGCCCGACTACGCCGAATTGCCTCCAAAAAGAAGAGAAAGGTAGGG
 ATCCgaattccccggggaaaaaccgTTTGAGTGTAAGATTGCGGGAAAGCTTTCATTTCAGAAGTCA
 AACCTCATCAGACACCAGAGAACTCACaccggggaaaaaccgTACAAGTGTGAAGAATGTGGCAAA
 GCTTTTACCCAATCCTCAAACCTTACTAAACATAAGAAAATTCATaccggggaaaaaccgTATAAA
 TGTAAAGCAATGTGGGAAAGCTTTTGGATGTCCCTCAAACCTTCGAAGGCATGGAAGGACTCACacc
 ggtgaaaaagcgggccgctaaattcTACCTGCCAGATACAGACGATCGTCACCGGATTGAGGAGAAA
 CGTAAAAGGACATATGAGACCTTCAAGAGCATCATGAAGAAGAGTCCTTTCAGCGGACCCACCGAC
 CCCCCGCCTCCACCTCGACGCATTGCTGTGCCTTCCCGCAGCTCAGCTTCTGTCCCCAAGCCAGCA
 CCCAGCCCTATCCCTTTACGTCATCCCTGAGCACCATCAACTATGATGAGTTTCCCACCATGGTG
 TTTCTTCTGGGCAGATCAGCCAGGCCTCGGCCTTGGCCCCGGCCCCCTCCCCAAGTCCTGCCCCAG
 GCTCCAGCCCCCTGCCCTGCTCCAGCCATGGTATCAGCTCTGGCCCAGGCCCCAGCCCCCTGTCCCA
 GTCCTAGCCCCAGGCCCTCCTCAGGCTGTGGCCCCACCTGCCCCCAAGCCCACCCAGGCTGGGGAA
 GGAACGCTGTCAGAGGCCCTGCTGCAGCTGCAGTTTGATGATGAAGACCTGGGGGCCTTGCTTGGC
 AACAGCACAGACCCAGCTGTGTTACAGACCTGGCATCCGTCGACAACTCCGAGTTTCAGCAGCTG
 CTGAACCAGGGCATACTGTGGCCCCCACAACTGAGCCCATGCTGATGGAGTACCCTGAGGCT
 ATAACGCTAGTGACAGCCAGAGGCCCCCGACCCAGCTCCTGCTCCACTGGGGGCCCCGGGG
 CTCCCCAATGGCCTCCTTTCAGGAGATGAAGACTTCTCCTCCATTGCGGACATGGACTTCTCAGCC
 CTGCTGAGTCAG TAA (SEQ ID NO:1)

蛋白质序列

MVYPYDVPDYAELPPKKRKVGIRIPGEKPFECKDCGKA~~FI~~QKSNLIRHQ~~R~~TH~~T~~GEKPYKCEE~~C~~GK
 AFTQSSNLTKHKKIHTGEKPYKCKQCGKA~~F~~GCP~~S~~NLRRHGRTH~~T~~GEKAAKFYLPDTDDRHRIEEK
 RKRTYETFKSIMKSPFSGPTDPRPPPRRIAVPSRSSASVPKPA~~P~~QYPFTSSLSTINYDEFPTMV
 FPSGQISQASALAPAPPQVLPQAPAPAPAMVSALAQA~~P~~VPVLPAGPPQAVAPPAPKPTQAGE
 GTLSEALLQLQFDD~~E~~DLGALLGNSTDPVFTDLASVDNSEFQQLLNQGI~~P~~VAPHTTEPMLMEYPEA
 ITRLVTAQRPPDPAPAPLGAPGLPNGLLSGDEDFSSIADMDFSALLSQ (SEQ ID NO:2)

图14

OSTEO1-P65 (也称作 S045-15-D11)

RDKR-QTHR1-VSTR-RDKR-p65

核酸序列

ATGGTGTACCCCTACGACGTGCCCGACTACGCCGAATTGCCTCCAAAAAGAAGAGAAAGGTAGGG
 ATCCGaattcccggggaaaaaccgTATGTATGCGATGTAGAGGGATGTACGTGGAAATTTGCCCGC
 TCAGATAAGCTCAACAGACACAAGAAAAGGCACaccggggaaaaaccgTATGAGTGTACGATTGC
 GGAAAGTCCTTTAGGCAGAGCACCCACCTCACTCGGCACCGGAGGATCCACaccggggaaaaaccg
 TATGAGTGTAACTACTGTGGAAAAACCTTTAGTGTGAGCTCAACCCTTATTAGACATCAGAGAATC
 CACaccggggaaaaaccgTATGTATGCGATGTAGAGGGATGTACGTGGAAATTTGCCCGCTCAGAT
 AAGCTCAACAGACACAAGAAAAGGCACaccggtgaaaaagcggcgctaattcTACCTGCCAGAT
 ACAGACGATCGTCACCGGATTGAGGAGAAACGTAAAAGGACATATGAGACCTTCAAGAGCATCATG
 AAGAAGAGTCCTTTCAGCGGACCCACCGACCCCGGCCTCCACCTCGACGCATTGCTGTGCCTTCC
 CGCAGCTCAGCTTCTGTCCCCAAGCCAGCACCCCAGCCCTATCCCTTTACGTCATCCCTGAGCACC
 ATCAACTATGATGAGTTTCCACCATGGTGTTCCTTCTGGGCAGATCAGCCAGGCCTCGGCCTTG
 GCCCCGCCCCCTCCCCAAGTCCTGCCCCAGGCTCCAGCCCCTGCCCTGCTCCAGCCATGGTATCA
 GCTCTGGCCCAGGCCCCAGCCCCGTGCCAGTCCTAGCCCCAGGCCCTCCTCAGGCTGTGGCCCCA
 CCTGCCCCCAAGCCCACCCAGGCTGGGGAAGGAACGCTGTCAGAGGCCCTGCTGCAGCTGCAGTTT
 GATGATGAAGACCTGGGGGCCTTGCTTGGAACAGCACAGACCCAGCTGTGTTTACAGACCTGGCA
 TCCGTCGACAACTCCGAGTTTCAGCAGCTGCTGAACCAGGGCATACTGTGGCCCCCCACACAAC
 GAGCCCATGCTGATGGAGTACCCTGAGGCTATAACTCGCCTAGTGACAGCCAGAGGCCCCCGAC
 CCAGCTCCTGCTCCACTGGGGGCCCCGGGGCTCCCCAATGGCCTCCTTTCAGGAGATGAAGACTTC
 TCCTCCATTGCGGACATGGACTTCTCAGCCCTGCTGAGTCAG TAA (SEQ ID NO:3)

蛋白质序列

MVYPYDVPDYAELPPKKRKVGIRIPGEKPFECKDCGKAFIQKSNLIRHQRTHTGEKPYACPVESC
 DRRFSDSSNLTRHIRIHTGEKPYACPVESCDRRFSDSSNLTRHIRIHTGEKAAAFYLPDTPDRHR
 IEEKRKRTYETFKSIMKKSPPFSGPTDPRPPPRRIAVPSRSSASVPKPAPQYPFTSSLSTINYDEF
 PTMVFPSPQISQASALAPAPPQVLPQAPAPAPAPAMVSALAQAPAPVPVLAPGPPQAVAPPAPKPT
 QAGEGTLSEALLQLQFDDLDLALLGNSTDPVFTDLASVDNSEFQQLLNQGIPVAPHTTEPMLME
 YPEAITRLVTAQRPPDPAPAPPLGAPGLPNGLLSGDEDFSSIADMDFSALLSQ (SEQ ID NO:4)

图15

08_D04_p65:

RSHR-RDHT-VSSR-p65

核酸序列

ATGGTGTACCCCTACGACGTGCCCGACTACGCCGAATTGCCTCCAAAAAAGAAGAGAAAGGTAGGG
ATCCGaatccccggggaaaaaccgTATAAGTGCATGGAGTGTGGGAAGGCTTTTAACCGCAGGTCA
CACCTCACACGGCACCAGCGGATTCACaccggggaaaaaccgTTCAGTGTAAACTTGTGAGCGA
AAGTTCTCCCGGTCCGACCACCTGAAGACCCACACCAGGACTCATaccggggaaaaaccgTATACA
TGTAACAGTGTGGGAAAGCCTTCAGTGTTCAGTTCCTTCGAAGACATGAAACCACTCACacc
ggtgaaaaagcgggccgctaaattcTACCTGCCAGATACAGACGATCGTCACCGGATTGAGGAGAAA
CGTAAAAGGACATATGAGACCTTCAAGAGCATCATGAAGAAGAGTCCTTTCAGCGGACCCACCGAC
CCCCGGCCTCCACCTCGACGCATTGCTGTGCCTTCCCGCAGCTCAGCTTCTGTCCCCAAGCCAGCA
CCCCAGCCCTATCCCTTTACGTCATCCCTGAGCACCATCAACTATGATGAGTTTCCCACCATGGTG
TTTCCTTCTGGGCAGATCAGCCAGGCCTCGGCCTTGGCCCCGGCCCCCTCCCCAAGTCCTGCCCCAG
GCTCCAGCCCCCTGCCCCCTGCTCCAGCCATGGTATCAGCTCTGGCCCAGGCCCCAGCCCCCTGTCCCA
GTCCTAGCCCCAGGCCCTCCTCAGGCTGTGGCCCCACCTGCCCCCAAGCCACCCAGGCTGGGGAA
GGAACGCTGTGAGAGGCCCTGCTGCAGCTGCAGTTTGATGATGAAGACCTGGGGGCCTTGCTTGGC
AACAGCACAGACCAGCTGTGTTACAGACCTGGCATCCGTGACAACTCCGAGTTTCAGCAGCTG
CTGAACCAGGGCATACTGTGGCCCCCACACAACTGAGCCCATGCTGATGGAGTACCCTGAGGCT
ATAACTCGCCTAGTGACAGCCAGAGGCCCCCGACCCAGCTCCTGCTCCACTGGGGGCCCCGGGG
CTCCCCAATGGCCTCCTTTCAGGAGATGAAGACTTCTCCTCCATTGCGGACATGGACTTCTCAGCC
CTGCTGAGTCAG TAA (SEQ ID NO:5)

蛋白质序列

MVYPYDVPDYAELPPKKKRVGIRIPGEKPYKMECGKAFNRRSHLTRHQRIHTGEKPFQCKTCQR
KFSRSDHLKTHTRHTGEKPYTCKQCGKAFSVSSSLRRHETHTGEKAAAFYLPDDDRHRIEEK
RKRTYETFKSIMKKS PFSGPTDPRPPPRRIAVPSRSSASVPKPAPQYPFTSSLSTINYDEFPTMV
FPSGQISQASALAPAPPQVLPQAPAPAPAPAMVSALAQAPAPVPVLAPGPPQAVAPPAPKPTQAGE
GTLSEALLQLQFDDDELGALLGNSTDPVFTDLASVDNSEFQQLLNQGI PVAPHTTTPMLMEYPEA
ITRLVTAQRPPDPAPAPLGA PGLPNGLLSGDEDFSSIADMDFSALLSQ (SEQ ID NO:6)

图16

名称: P_B08

QSNR1-DSNR-DSNR-p65

核酸序列

ATGGTGTACCCCTACGACGTGCCCCGACTACGCCGAATTG CCTCCAAAAAAGAAGAGAAAGGTAGGG
 ATCCGaattccccggggaaaaaccgTTTGAGTGTAAAGATTGCGGGAAAGCTTTCATTTCAGAAGTCA
 AACCTCATCAGACACCAGAGAACTCACaccggggaaaaaccgTATGCTTGCCCTGTCGAGTCCTGC
 GATCGCCGCTTTTCTGATTTCGTCGAACCTTACCCGCCATATCCGCATCCACaccggggaaaaaccg
 TATGCTTGCCCTGTCGAGTCCTGCGATCGCCGCTTTTCTGATTTCGTCGAACCTTACCCGCCATATC
 CGCATCCACaccggtgaaaaagcgccgctaaattcTACCTGCCAGATACAGACGATCGTCACCGG
 ATTGAGGAGAAACGTAAAAGGACATATGAGACCTTCAAGAGCATCATGAAGAAGAGTCCTTTCAGC
 GGACCCACCGACCCCGGCCTCCACCTCGACGCATTGCTGTGCCTTCCCGCAGCTCAGCTTCTGTC
 CCCAAGCCAGCACCCCGCCCTATCCCTTTACGTCATCCCTGAGCACCATCAACTATGATGAGTTT
 CCCACCATGGTGTTCCTTCTGGGCAGATCAGCCAGGCCTCGGCCTTGGCCCCGGCCCTCCCCAA
 GTCCTGCCCCAGGCTCCAGCCCCTGCCCCTGCTCCAGCCATGGTATCAGCTCTGGCCCAGGCCCA
 GCCCCTGTCCCAGTCCTAGCCCCAGGCCCTCCTCAGGCTGTGGCCCCACCTGCCCCAAGCCCACC
 CAGGCTGGGGAAGGAACGCTGTCAGAGGCCCTGCTGCAGCTGCAGTTTGATGATGAAGACCTGGGG
 GCCTTGCTTGGAACAGCACAGACCCAGCTGTGTTACAGACCTGGCATCCGTCGACAACCTCCGAG
 TTTCAGCAGCTGCTGAACCAGGGCATACTGTGGCCCCCACACAACTGAGCCCATGCTGATGGAG
 TACCCTGAGGCTATAACTCGCCTAGTGACAGCCAGAGGCCCCCGACCCAGCTCCTGCTCCACTG
 GGGGCCCCGGGGCTCCCCAATGGCCTCCTTTCAGGAGATGAAGACTTCTCCTCCATTGCGGACATG
 GACTTCTCAGCCCTGCTGAGTCAG TAA (SEQ ID NO:7)

蛋白质序列

MVYPYDVPDYAELPPKKRKGIRIPGEKPFECKDCGKA~~FI~~QKSNLIRHQ~~R~~TH~~T~~GEKPYACP~~V~~ESC
 DRRFSDSSNLTRHIRIHTGEKPYACP~~V~~ESCDRRFSDSSNLTRHIRIHTGEKAAAKFYLPDTDDRHR
 IEEKRKRTYETFKSIMKSPFSGPTDPRPPPRRIAVPSRSSASVPKPAPQYPYPTSSSLSTINYDEF
 PTMVFPSGQISQASALAPAPPQVLPQAPAPAPAPAMVSALAQAPAPVPVLPAGPPQAVAPPAPKPT
 QAGEGTLSEALLQLQFDDDLGALLGNSTDPVFTDLASVDNSEFQQLLNQGI~~P~~VAPHTTEPMLME
 YPEAITRLVTAQRPPDPAPAPL~~G~~APGLPNGLLSGDEDFSSIADMDFSALLSQ (SEQ ID NO:8)

图17

名称: K_D10
QSHV-WSNR-WSNR-RDNQ-kid

核酸序列

ATGGTGTACCCCTACGACGTGCCCCGACTACGCCGAATTGCTCCAAAAAGAAGAGAAAGGTAGGG
ATCCgaattccccggggaaaaaccgTATGAGTGTGATCACTGTGGAAAATCCTTTAGCCAGAGCTCT
CATCTGAATGTGCACAAAAGAACTCACaccggggaaaaaccgTACAGATGTGAGGAATGTGGCAA
GCCTTTAGGTGGCCCTCAAACCTTACTAGACATAAGAGAATTCACaccggggaaaaaccgTACAGA
TGTGAGGAATGTGGCAAAGCCTTTAGGTGGCCCTCAAACCTTACTAGACATAAGAGAATTCACacc
ggggaaaaaccgTTTGCCTGCCCTGAGTGTCTAAGCGCTTCATGAGATCCGACAACCTGACCCAG
CATATCAAGACCCACaccggtgaaaaagcgccgctaaattcGTGTCAGTGACATTTGAAGATGTG
GCTGTGCTCTTTACTCGGGACGAGTGGAAGAAGCTGGATCTGTCTCAGAGAAGCCTGTACCGTGAG
GTGATGCTGGAGAATTACAGCAACCTGGCCTCCATGGCAGGATTCCTGTTTACCAAACCAAAGGTG
ATCTCCCTGTTGCAGCAAGGAGAGGATCCCTGGTAA (SEQ ID NO:9)

蛋白质序列

MVYPYDVPDYAELPPKKKRVGIRIPGEKPYECDHCGKSFSQSSHLNVHKRTHHTGEKPYRCEECKG
AFRWPSNLTRHKRIHTGEKPYRCEECKGAFRWPSNLTRHKRIHTGEKPFACPECPKRFMRSDNLTQ
HIKTHTEKAAAKFVSVTFEDVAVLFTRDEWKKLDLSQRSYREVMLENYSNLASMAFLFTKPKV
ISLLQQGEDPW (SEQ ID NO:10)

图18

名称：K_F02

DSAR2-RDKR-RDER1-QTHR1-kid

核酸序列

ATGGTGTACCCCTACGACGTGCCCCGACTACGCCGAATTGCCTCCAAAAAGAAGAGAAAGGTAGGG
ATCCGaattcccggggaaaaaccgTACTCCTGTGGCATTGTGTGGCAAATCCTTCTCTGACTCCAGT
GCCAAAAGGAGACACTGCATTCTACACaccggggaaaaaccgTATGTATGCGATGTAGAGGGATGT
ACGTGGAAATTTGCCCGCTCAGATAAGCTCAACAGACACAAGAAAAGGCACaccggggaaaaaccg
TATGTATGCGATGTAGAGGGATGTACGTGGAAATTTGCCCGCTCAGATGAGCTCAACAGACACAAG
AAAAGGCACaccggggaaaaaccgTATGAGTGTACGATTGCGGAAAGTCCTTTAGGCAGAGCACC
CACCTCACTCGGCACCGGAGGATCCACaccggtgaaaaagcgccgctaaattcGTGTCAGTGACA
TTTGAAGATGTGGCTGTGCTCTTTACTCGGGACGAGTGGAAGAAGCTGGATCTGTCTCAGAGAAGC
CTGTACCGTGAGGTGATGCTGGAGAATTACAGCAACCTGGCCTCCATGGCAGGATTCCTGTTTACC
AAACCAAAGGTGATCTCCCTGTTGCAGCAAGGAGAGGATCCCTGGTAA (SEQ ID NO:11)

蛋白质序列

MVYPYDVPDYAELPPKKRKGIRIPGEKPYSCGICGKSFS~~SDSSAKRRHCIL~~HTGEKPYVCDVEGC
TWKFARSDKLN~~RHKKRHT~~GEKPYVCDVEGCTWKFARS~~DELNRHKKRHT~~GEKPYECHDCGKSFRQST
HLTRHRR~~IHTGEKAAAKFVS~~VTTFEDVAVLFTRDEWKKLDLSQRS~~LYREV~~MLENYSNLAS~~MAG~~FLFT
KPKVISLLQGEDPW (SEQ ID NO:12)

图19

名称: K44_11_D01
QSHV-QSNI-QTHR1-CSNR1-kid

核酸序列

ATGGTGTACCCCTACGACGTGCCGACTACGCCGAATTGCCTCCAAAAAGAAGAGAAAGGTAGGG
ATCCgaattcccggggaaaaaccgTATGAGTGTGATCACTGTGGAAAATCCTTTAGCCAGAGCTCT
CATCTGAATGTGCACAAAAGAACTCACaccggggaaaaaccgTACATGTGCAGTGAGTGTGGGCGA
GGCTTCAGCCAGAAGTCAAACCTCATCATACACCAGAGGACACACaccggggaaaaaccgTATGAG
TGTCACGATTGCGGAAAGTCCTTTAGGCAGAGCACCCACCTCACTCGGCACCGGAGGATCCACacc
ggggaaaaaccgTATAAATGTAAGCAATGTGGGAAAGCTTTTGGATGTCCCTCAAACCTTCGAAGG
CATGGAAGGACTCACaccggtgaaaaagcgccgctaaattcGTGTCAGTGACATTTGAAGATGTG
GCTGTGCTCTTTACTCGGGACGAGTGGAAGAAGCTGGATCTGTCTCAGAGAAGCCTGTACCGTGAG
GTGATGCTGGAGAATTACAGCAACCTGGCCTCCATGGCAGGATTCCTGTTTACCAAACCAAAGGTG
ATCTCCCTGTTGCAGCAAGGAGAGGATCCCTGGTAA (SEQ ID NO:13)

蛋白质序列

MVYPYDVPDYAELPPKKRKVGIRIPGEKPYECDHCGKSFSQSSHLNVHKRTHTEKPYMCSECGR
GFSQKSNLIHQRTHTGEKPYECHDCGKSFRQSTHLTRHRIHTGEKPYKCKQCGKAFGCPSNLRR
HGRTHTEKAAAKFVSVTFEDVAVLFTREWKLDLSQRSLYREVMLENYSNLASMAGFLFTKPKV
ISLLQQGEDPW (SEQ ID NO:14)

图20

名称: K 44_11_G12
QSHV-VSTR-RDNQ-QTHR1-kid

核酸序列

ATGGTGTACCCCTACGACGTGCCCCGACTACGCCGAATTGCCTCCAAAAAGAAGAGAAAGGTAGGG
ATCCgaattccccggggaaaaaccgTATGAGTGTGATCACTGTGGAAAATCCTTTAGCCAGAGCTCT
CATCTGAATGTGCACAAAGAACTCACaccggggaaaaaccgTATGAGTGTAATTACTGTGGAAAA
ACCTTTAGTGTGAGCTCAACCCTTATTAGACATCAGAGAATCCACaccggggaaaaaccgTTTGCC
TGCCCTGAGTGTCTTAAGCGCTTCATGAGATCCGACAACCTGACCCAGCATATCAAGACCCACacc
ggggaaaaaccgTATGAGTGTACGATTGCGGAAAGTCCTTTAGGCAGAGCACCCACCTCACTCGG
CACCGGAGGATCCACaccggtgaaaaagcggcgcgctaaattcGTGTCAGTGACATTTGAAGATGTG
GCTGTGCTCTTTACTCGGGACGAGTGGAAAGAAGCTGGATCTGTCTCAGAGAAGCCTGTACCGTGAG
GTGATGCTGGAGAATTACAGCAACCTGGCCTCCATGGCAGGATTCCTGTTTACCAAACCAAAGGTG
ATCTCCCTGTTGCAGCAAGGAGAGGATCCCTGGTAA (SEQ ID NO:15)

蛋白质序列

MVYPYDVPDYAELPPKKRKGIRIPGEKPYECDHCGKSFSQSSHLNVHKRTHTGKPYECNYCGK
TFSVSSTLIRHQRIHTGEKPFACPECPKRFRMSDNLQHIKTHTGKPYECHDCGKSFRQSTHLTR
HRRIHTGEKAAAKFVSVTFEDVAVLFTRDEWKKLDLSQRSLYREVMLENYSNLASMAGFLFTKPKV
ISLLQQGEDPW (SEQ ID NO:16)

图21

F104_p65

RDHT-RSHR-QSHR2-p65

核酸序列

ATGGTGTACCCCTACGACGTGCCCGACTACGCCGAATTGCTCCAAAAAGAAGAGAAAGGTAAGGG
 ATCCGaattcccggggaaaaaccgTTCCAGTGTAAACTTGTTCAGCGAAAGTTCTCCCGGTCCGAC
 CACCTGAAGACCCACACCAGGACTCATaccggggaaaaaccgTATAAGTGCATGGAGTGTGGGAAG
 GCTTTTAACCGCAGGTACACCTCACACGGCACCAGCGGATTCACaccggggaaaaaccgTATAAA
 TCGGGCCAGTGTGGGAAGTTCTACTCGCAGGTCTCCACCTCACC CGCCACCAGAAAATCCACacc
 ggtgaaaaagcgcccgctaaattctACCTGCCAGATACAGACGATCGTCACCGGATTGAGGAGAAA
 CGTAAAAGGACATATGAGACCTTCAAGAGCATCATGAAGAAGAGTCCTTTTCAGCGGACCCACCGAC
 CCGCGGCTCCACCTCGACGCATTGCTGTGCCTTCCCGCAGCTCAGCTTCTGTCCCCAAGCCAGCA
 CCCCAGCCCTATCCCTTTACGTCATCCCTGAGCACCATCAACTATGATGAGTTTCCCACCATGGTG
 TTTCTTCTGGGCAGATCAGCCAGGCCTCGGCCTTGGCCCCGGCCCCCTCCCCAAGTCCTGCCCCAG
 GCTCCAGCCCCCTGCCCCTGCTCCAGCCATGGTATCAGCTCTGGCCCAGGCCCCAGCCCCCTGTCCCA
 GTCCTAGCCCCAGGCCCTCCTCAGGCTGTGGCCCCACCTGCCCCCAAGCCACCCAGGCTGGGGAA
 GGAACGCTGTGAGAGGCCCTGCTGCAGCTGCAGTTTGATGATGAAGACCTGGGGGCCTTGCTTGGC
 AACAGCACAGACCCAGCTGTGTTACAGACCTGGCATCCGTCGACAACTCCGAGTTTCAGCAGCTG
 CTGAACCAGGGCATACTGTGGCCCCCACAACTGAGCCCATGCTGATGGAGTACCCTGAGGCT
 ATAACCTGCCTAGTGACAGCCAGAGGCCCCCGACCCAGCTCCTGCTCCACTGGGGGCCCCGGGG
 CTCCCCAATGGCCTCCTTTTCAGGAGATGAAGACTTCTCCTCCATTGCGGACATGGACTTCTCAGCC
 CTGCTGAGTCAG TAA (SEQ ID NO:17)

蛋白质序列

MVYPYDVPDYAELPPKKRKGIRIPGEKPFQCKTCQRKFSRSDHLKTHTRHTGEKPYKCMCEGK
 AFNRRSHLTRHQRIHTGEKPYKCGQCGKFYSQVSHLTRHQKIHTGEKAAAFYLPDTPDRHRIEEK
 RKRTYETFKSIMKSPFSGPTDPRPPPRRIAVPSRSSASVPKPAPQYPFTSSLSTINYDEFPTMV
 FPSGQISQASALAPAPPQVLPQAPAPAPAPAMVSALAQAPAPVPVLAPGPPQAVAPPAPKPTQAGE
 GTLSEALLQLQFDDDLGALLGNSTDPVFTDLASVDNSEFQQLLNQGI PVAPHTTEPMLMEYPEA
 ITRLVTAQRPPDPAPAPLGAPGLPNGLLSGDEDFSSIADMDFSALLSQ (SEQ ID NO:18)

图22

F121_p65**QSHT-RSHR-RDHT-p65****核酸序列**

ATGGTGTACCCCTACGACGTGCCCCGACTACGCCGAATTGCTCCAAAAAGAAGAGAAAGGTAGGG
 ATCCGaattcccggggaaaaaccgTACAAATGTGAAGAATGTGGCAAAGCCTTTAGGCAGTCCTCA
 CACCTTACTACACATAAGATAATTCATaccgggggaaaaaccgTATAAGTGCATGGAGTGTGGGAAG
 GCTTTTAACCGCAGGTCACACCTCACACGGCACCAGCGGATTTCACaccgggggaaaaaccgTTCCAG
 TGTAAAACTTGTACGCGAAAGTTCTCCCGGTCCGACCACCTGAAGACCCACACCAGGACTCATacc
 ggtgaaaaagcgcccgctaaattcTACCTGCCAGATACAGACGATCGTCACCGGATTGAGGAGAAA
 CGTAAAAGGACATATGAGACCTTCAAGAGCATCATGAAGAAGAGTCCTTTCAGCGGACCCACCGAC
 CCCCCGCCTCCACCTCGACGCATTGCTGTGCCTTCCCGCAGCTCAGCTTCTGTCCCCAAGCCAGCA
 CCCAGCCCTATCCCTTTACGTCATCCCTGAGCACCATCAACTATGATGAGTTTCCCACCATGGTG
 TTTCTTCTGGGCAGATCAGCCAGGCCTCGGCCTTGGCCCCGGCCCCCTCCCCAAGTCCTGCCCCAG
 GCTCCAGCCCCTGCCCCCTGCTCCAGCCATGGTATCAGCTCTGGCCCAGGCCCCAGCCCCCTGTCCCA
 GTCCTAGCCCCAGGCCCTCCTCAGGCTGTGGCCCCACCTGCCCCCAAGCCCACCCAGGCTGGGGAA
 GGAACGCTGTCAGAGGCCCTGCTGCAGCTGCAGTTTGATGATGAAGACCTGGGGGCCTTGCTTGGC
 AACAGCACAGACCCAGCTGTGTTACAGACCTGGCATCCGTCGACAACTCCGAGTTTCAGCAGCTG
 CTGAACCAGGGCATACTGTGGCCCCCACACAAGTGGAGCCCATGCTGATGGAGTACCCTGAGGCT
 ATAACTCGCCTAGTGACAGCCAGAGGCCCCCCGACCCAGCTCCTGCTCCACTGGGGGCCCCGGGG
 CTCCCCAATGGCCTCCTTTCAGGAGATGAAGACTTCTCCTCCATTGCGGACATGGACTTCTCAGCC
 CTGCTGAGTCAG TAA (SEQ ID NO:19)

蛋白质序列

MVYPYDVPDYAELPPKKRKVGIRIPGEKPYKCEECGKAFRQSSHLTTHKI IHTGEKPYKMECGK
 AFNRRSHLTRHQRIHTGEKPFQCKTCQRKFSRSDHLKTHTRTHTGEKAAAFYLPDTDDRHRIEEK
 RKRTYETFKSIMKSPFSGPTDPRPPPRRIAVPSRSSASVPKPAQPYPFTSSLSTINYDEFPTMV
 FPSGQISQASALAPAPPQVLPQAPAPAPAPAMVSALAQAPAPVPVLAPGPPQAVAPPAPKPTQAGE
 GTLSEALLQLQFDDDLGALLGNSTDPVFTDLASVDNSEFQQLLNQGI PVAPHTTEPMLMEYPEA
 ITRLVTAQRPPDPAPAPLGPGLPNGLLSGDEDFSSIADMDFSALLSQ (SEQ ID NO:20)

图23

K12_A11

RDHT-QSNV2-QTHR1-QFNR-kid

核酸序列

ATGGTGTACCCCTACGACGTGCCCGACTACGCCGAATTG CCTCCAAAAAGAAGAGAAAGGTAGGG
ATCCgaattcccgggggaaaaaccgTTCCAGTGTAAACTTGTGAGCGAAAGTTCTCCCGGTCCGAC
CACCTGAAGACCCACACCAGGACTCATaccgggggaaaaaccgTATGTTTGCTCAAAATGTGGGAAA
GCCTTCACTCAGAGTTCAAATCTGACTGTACATCAAAAAATCCACaccgggggaaaaaccgTATGAG
TGTCACGATTGCGGAAAGTCCTTTAGGCAGAGCACCCACCTCACTCGGCACCGGAGGATCCACacc
ggggaaaaaccgTATAAGTGTCAATGTGGGAAAGCCTTTATTCAATCCTTTAACCTTCGAAGA
CATGAGAGAACTCACaccggtgaaaaagcgccgctaaattcGTGTCAGTGACATTTGAAGATGTG
GCTGTGCTCTTTACTCGGGACGAGTGGAAGAAGCTGGATCTGTCTCAGAGAAGCCTGTACCGTGAG
GTGATGCTGGAGAATTACAGCAACCTGGCCTCCATGGCAGGATTCCTGTTTACCAAACCAAAGGTG
ATCTCCCTGTTGCAGCAAGGAGAGGATCCCTGGTAA (SEQ ID NO:259)

蛋白质序列

MVYPYDVPDYAELPPKKKRKVGIRIPGEKPFQCKTCQRKFSRSDHLKTHTRTHTGEKPYVCSKCGK
AFTQSSNLTVHQIHTGEKPYECHDCGKSFRQSTHLTRHRIHTGEKPYKCHQCGKAFIQSFNLRR
HERTHTGEKAAAKFVSVTFEDVAVLFTRDEWKLDLSQRSLYREVMLENYSNLASMAGFLFTKPKV
ISLLQQGEDPW (SEQ ID NO:260)

图24

K44-16-E12

QSHV-QSSRI-QTHR1-kid

核酸序列

ATGGTGTACCCCTACGACGTGCCCGACTACGCCGAATTGCCTCCAAAAAGAAGAGAAAGGTAGGG
ATCCgaattccccggggaaaaaccgTATGAGTGTGATCACTGTGGAAAATCCTTTAGCCAGAGCTCT
CATCTGAATGTGCACAAAAGAACTCACaccgggggaaaaaccgTATAAGTGCCCTGATTGTGGGAAG
AGTTTTTAGTCAGAGTTCCAGCCTCATTGCGCCACCAGCGGACACACaccgggggaaaaaccgTATGAG
TGTCACGATTGCGGAAAGTCCTTTAGGCAGAGCACCCACCTCACTCGGCACCGGAGGATCCACacc
ggtgaaaaagcgggcgctaaattcGTGTCAGTGACATTTGAAGATGTGGCTGTGCTCTTTACTCGG
GACGAGTGGAAGAAGCTGGATCTGTCTCAGAGAAGCCTGTACCGTGAGGTGATGCTGGAGAATTAC
AGCAACCTGGCCTCCATGGCAGGATTCCTGTTTACCAAACCAAAGGTGATCTCCCTGTTGCAGCAA
GGAGAGGATCCCTGGTAA (SEQ ID NO:261)

蛋白质序列

MVYPYDVPDYAELPPKKRKVGIRIPGEKPYECDHCGKSFSQSSHLNVHKRTHTGEKPYKCPDCGK
SFSQSSSLIRHQRTHTGEKPYECHDCGKSFRQSTHLTRHRRITHTGEKAAAKFVSVTFEDVAVLFTR
DEWKKLDLSQRSLYREVMLENYSNLASMAGFLFTKPKVISLLQQGEDPW (SEQ ID NO:262)

图25

K13_B08

QSNR1-QSSR1-QTHR1-RDKR-kid

核酸序列

ATGGTGTACCCCTACGACGTGCCCGACTACGCCGAATTGCCTCCAAAAAGAAGAGAAAGGTAGGG
ATCCgaattcccggggaaaaaccgTTTGAGTGTAAGATTGCGGGAAAGCTTTCATTTCAGAAGTCA
AACCTCATCAGACACCAGAGAACTCACaccggggaaaaaccgTATAAGTGCCCTGATTGTGGGAAG
AGTTTTAGTCAGAGTTCCAGCCTCATTGCGCACACAGCGGACACACaccggggaaaaaccgTATGAG
TGTCACGATTGCGGAAAGTCCTTTAGGCAGAGCACCCACCTCACTCGGCACCGGAGGATCCACacc
ggggaaaaaccgTATGTATGCGATGTAGAGGGATGTACGTGGAAATTTGCCCGCTCAGATAAGCTC
AACAGACACAAGAAAAGGCACaccggtgaaaaagcgccgctaattcGTGTCAGTGACATTTGAA
GATGTGGCTGTGCTCTTTACTCGGGACGAGTGGAAGAAGCTGGATCTGTCTCAGAGAAGCCTGTAC
CGTGAGGTGATGCTGGAGAATTACAGCAACCTGGCCTCCATGGCAGGATTCCTGTTTACCAAACCA
AAGGTGATCTCCCTGTTGCAGCAAGGAGAGGATCCCTGGTAA (SEQ ID NO:263)

蛋白质序列

MVYPYDVPDYAELPPKKKRKVGIRIPGEKPFECKDCGKAFIQKSNLIRHQRTHTGEKPYKCPDCGK
SFSQSSSLIRHQRTHTGEKPYECHDCGKSFRQSTHLTRHRIHTGEKPYVCDVEGCTWKFARSDKL
NRHKKRHTGEKAAAKFVSVTFEDVAVLFTRDEWKKLDLSQRSLYREVMLENYSNLASMAFLFTKP
KVISLLQQGEDPW (SEQ ID NO:264)

图26