



(45)授权公告日 2018.10.26

A

□ pRbA (C200p)
● 对照细胞 (pRbA (C200p) + Casp3)
■ C. neoformans
□ C200p

0 时间

10 分钟

20 分钟

30 分钟

40 分钟

50 分钟

60 分钟

70 分钟

80 分钟

90 分钟

100 分钟

120 分钟

140 分钟

160 分钟

180 分钟

200 分钟

220 分钟

240 分钟

260 分钟

280 分钟

300 分钟

320 分钟

340 分钟

360 分钟

380 分钟

400 分钟

420 分钟

440 分钟

460 分钟

480 分钟

500 分钟

520 分钟

540 分钟

560 分钟

580 分钟

600 分钟

620 分钟

640 分钟

660 分钟

680 分钟

700 分钟

720 分钟

740 分钟

760 分钟

780 分钟

800 分钟

820 分钟

840 分钟

860 分钟

880 分钟

900 分钟

920 分钟

940 分钟

960 分钟

980 分钟

1000 分钟

1020 分钟

1040 分钟

1060 分钟

1080 分钟

1100 分钟

1120 分钟

1140 分钟

1160 分钟

1180 分钟

1200 分钟

1220 分钟

1240 分钟

1260 分钟

1280 分钟

1300 分钟

1320 分钟

1340 分钟

1360 分钟

1380 分钟

1400 分钟

1420 分钟

1440 分钟

1460 分钟

1480 分钟

1500 分钟

1520 分钟

1540 分钟

1560 分钟

1580 分钟

1600 分钟

1620 分钟

1640 分钟

1660 分钟

1680 分钟

1700 分钟

1720 分钟

1740 分钟

1760 分钟

1780 分钟

1800 分钟

1820 分钟

1840 分钟

1860 分钟

1880 分钟

1900 分钟

1920 分钟

1940 分钟

1960 分钟

1980 分钟

2000 分钟

2020 分钟

2040 分钟

2060 分钟

2080 分钟

2100 分钟

2120 分钟

2140 分钟

2160 分钟

2180 分钟

2200 分钟

2220 分钟

2240 分钟

2260 分钟

2280 分钟

2300 分钟

2320 分钟

2340 分钟

2360 分钟

2380 分钟

2400 分钟

2420 分钟

2440 分钟

2460 分钟

2480 分钟

2500 分钟

2520 分钟

2540 分钟

2560 分钟

2580 分钟

2600 分钟

2620 分钟

2640 分钟

2660 分钟

2680 分钟

2700 分钟

2720 分钟

2740 分钟

2760 分钟

2780 分钟

2800 分钟

2820 分钟

2840 分钟

2860 分钟

2880 分钟

2900 分钟

2920 分钟

2940 分钟

2960 分钟

2980 分钟

3000 分钟

3020 分钟

3040 分钟

3060 分钟

3080 分钟

3100 分钟

3120 分钟

3140 分钟

3160 分钟

3180 分钟

3200 分钟

3220 分钟

3240 分钟

3260 分钟

3280 分钟

3300 分钟

3320 分钟

3340 分钟

3360 分钟

3380 分钟

3400 分钟

3420 分钟

3440 分钟

3460 分钟

3480 分钟

3500 分钟

3520 分钟

3540 分钟

3560 分钟

3580 分钟

3600 分钟

3620 分钟

3640 分钟

3660 分钟

3680 分钟

3700 分钟

3720 分钟

3740 分钟

3760 分钟

3780 分钟

3800 分钟

3820 分钟

3840 分钟

3860 分钟

3880 分钟

3900 分钟

3920 分钟

3940 分钟

3960 分钟

3980 分钟

4000 分钟

4020 分钟

4040 分钟

4060 分钟

4080 分钟

4100 分钟

4120 分钟

4140 分钟

4160 分钟

4180 分钟

4200 分钟

4220 分钟

4240 分钟

4260 分钟

4280 分钟

4300 分钟

4320 分钟

4340 分钟

4360 分钟

4380 分钟

4400 分钟

4420 分钟

4440 分钟

4460 分钟

4480 分钟

4500 分钟

4520 分钟

4540 分钟

4560 分钟

4580 分钟

4600 分钟

4620 分钟

4640 分钟

4660 分钟

4680 分钟

4700 分钟

4720 分钟

4740 分钟

4760 分钟

4780 分钟

4800 分钟

4820 分钟

4840 分钟

4860 分钟

4880 分钟

4900 分钟

4920 分钟

4940 分钟

4960 分钟

4980 分钟

5000 分钟

5020 分钟

5040 分钟

5060 分钟

5080 分钟

5100 分钟

5120 分钟

5140 分钟

5160 分钟

5180 分钟

5200 分钟

5220 分钟

5240 分钟

5260 分钟

5280 分钟

5300 分钟

5320 分钟

5340 分钟

5360 分钟

5380 分钟

5400 分钟

5420 分钟

5440 分钟

5460 分钟

5480 分钟

5500 分钟

5520 分钟

5540 分钟

5560 分钟

5580 分钟

5600 分钟

5620 分钟

5640 分钟

5660 分钟

5680 分钟

5700 分钟

5720 分钟

5740 分钟

5760 分钟

5780 分钟

5800 分钟

5820 分钟

5840 分钟

5860 分钟

5880 分钟

5900 分钟

5920 分钟

5940 分钟

5960 分钟

5980 分钟

6000 分钟

6020 分钟

6040 分钟

6060 分钟

6080 分钟

6100 分钟

6120 分钟

6140 分钟

6160 分钟

6180 分钟

6200 分钟

6220 分钟

6240 分钟

6260 分钟

6280 分钟

6300 分钟

6320 分钟

6340 分钟

6360 分钟

6380 分钟

6400 分钟

6420 分钟

6440 分钟

6460 分钟

6480 分钟

6500 分钟

6520 分钟

6540 分钟

6560 分钟

6580 分钟

6600 分钟

6620 分钟

6640 分钟

6660 分钟

6680 分钟

6700 分钟

6720 分钟

6740 分钟

6760 分钟

6780 分钟

6800 分钟

6820 分钟

6840 分钟

6860 分钟

6880 分钟

6900 分钟

6920 分钟

6940 分钟

6960 分钟

6980 分钟

7000 分钟

7020 分钟

7040 分钟

7060 分钟

7080 分钟

7100 分钟

7120 分钟

7140 分钟

7160 分钟

7180 分钟

7200 分钟

7220 分钟

7240 分钟

7260 分钟

7280 分钟

7300 分钟

7320 分钟

734

1. 一种分离的多肽,其包括黄单胞杆菌属衍生的转录激活子样效应因子(TALE)蛋白,所述TALE蛋白具有N-末端结构域(NTD),所述N-末端结构域包括在N₋₁发夹中的氨基酸序列VGKQWSGARAL,只是在所述氨基酸序列中的一个或多个突变选自:Q是Y、Q是S、Q是R、W是R、W是G、W缺失、S是R、S是H、S是A、S是N、以及S是T。

2. 如权利要求1所述的多肽,其中所述NTD包含选自以下序列的氨基酸序列:VGKYRGARAL (SEQ ID NO:4)、VGKSRSRGARAL (SEQ ID NO:5)、VGKYHGARAL (SEQ ID NO:6) 和VGKRGAGARAL (SEQ ID NO:7)。

3. 一种分离的多肽,其包含罗尔斯通菌属衍生的合成的转录激活子样效应因子(TALE)蛋白,所述TALE蛋白具有N-末端结构域(NTD),所述N-末端结构域包含在N₋₁发夹中的氨基酸序列IVDIAR₁QR₂SGDLA,只是在所述氨基酸序列中的一个或多个突变选自:R₁是K、Q是Y、Q是S、Q是R、R₂是W、R₂是G、R₂缺失、S是R、S是H、S是A、S是N、以及S是T。

4. 如权利要求3所述的多肽,其中所述NTD包括选自以下序列的氨基酸序列:IVDIARQWSGDLA (SEQ ID NO:9)、IVDIARYRGDLA (SEQ ID NO:10)、IVDIARSRSGLA (SEQ ID NO:11)、IVDIARYHGDLA (SEQ ID NO:12) 和IVDIARRGAGDLA (SEQ ID NO:13)。

5. 如权利要求1-4中任一项所述的多肽,其进一步包括重组酶结构域或核酸酶结构域。

6. 如权利要求1-2中任一项所述的多肽,其中所述TALE蛋白衍生自AvrXa7、Tal1c或PthXol1。

7. 如权利要求1-2中任一项所述的多肽,其中所述TALE蛋白包含合成的RVD结构域。

8. 如权利要求7所述的多肽,其中所述TALE蛋白包含C-末端截短和/或N-末端截短。

9. 如权利要求7所述的多肽,其中所述TALE蛋白包括在残基120、128或145处截短的AvrXa7 N-末端序列和/或在残基28处截短的AvrXa7 C-末端序列。

10. 如权利要求5所述的多肽,其中所述重组酶选自:Gin, Hin, Tn3, Sin, Beta, Pin, Min, Din, Cin, Gin p, Gin y, Gin 6, Gin e和Gin Z。

11. 如权利要求10所述的多肽,其中所述重组酶是Gin。

12. 如权利要求10所述的多肽,其中所述重组酶是Gin并且所述TALE蛋白是AvrXa7。

13. 一种用于体外位点特异性重组的方法,其包括:

(a) 提供包括用于与权利要求5所述的多肽特异性相互作用的至少两个结合位点的DNA序列;以及

(b) 使所述DNA序列与所述多肽发生反应,其中所述多肽催化与所述多肽特异性相互作用的两个位点之间的位点特异性重组事件,在所述事件中,所述DNA序列的两条链都被切割。

14. 如权利要求13所述的方法,其中所述位点特异性重组事件是倒位。

15. 如权利要求13所述的方法,其中所述位点特异性重组事件是整合。

16. 如权利要求13所述的方法,其中所述位点特异性重组事件是解离。

17. 编码权利要求1或3所述的多肽的核酸分子在制备用于基因治疗的药物中的用途。

18. 一种药物组合物,其包括:

a) 权利要求1或3所述的多肽;以及

b) 药学上可接受的载体。

19. 一种药物组合物,其包括:

- a) 核酸分子,其编码权利要求1或3所述的多肽;以及
- b) 药学上可接受的载体。
- 20. 一种分离的核酸分子,其编码权利要求1-5中任一项所述的多肽。
- 21. 一种表达盒,其包含权利要求20所述的核酸分子。
- 22. 一种载体,其包含权利要求21所述的表达盒。
- 23. 一种宿主细胞,其被用权利要求20所述的核酸分子或权利要求22所述的载体转化或转染。

具有靶向结合特异性的嵌合多肽

[0001] 相关申请的交叉引用

[0002] 本申请根据35U.S.C. §119(e) 要求于2012年9月4日提交的美国序列号为61/696,689的申请、于2013年1月17日提交的美国序列号为61/753,763的申请以及于2013年5月1日提交的美国序列号为61/818,364的申请的优先权的利益,上述申请的全部内容通过引用并入本文。

技术领域

[0003] 本发明总体上涉及生物技术领域,并且更具体地涉及识别特异性DNA序列的嵌合重组酶。

背景技术

[0004] 蛋白质以序列依赖性的方式识别DNA的能力对生命是至关重要的,因为各种蛋白质结构域已经发展到提供序列特异性DNA识别。由这些结构域中的少数几个的DNA识别也是各种各样的生物技术应用的基础。特别是,C₂H₂型锌指蛋白(ZFPs)是第一批被设计以识别用户定义的DNA序列的DNA结合蛋白质之一并且已被不同程度的成功用于许多应用,该应用包括转录调控、基因组工程和后天修饰。ZFPs的模块化装配促进了这些方法。然而,尽管ZFP技术取得了进步且具有前景,但是对某些序列的特异性、高亲和性ZFPs的构建仍然困难并且在选择的情况下,需要使用不易被非专业实验室采用的耗时和劳动密集型的选择系统。

[0005] 转录激活因子样效应因子(TALE)结构域是代表ZFP技术的可能的替代方案的一类天然存在的DNA结合结构域(DBD)。被发现于植物病原体黄单胞杆菌属中的TALE包含一系列的33至35个氨基酸重复序列,该重复序列发挥功能以选择性地结合靶DNA序列。这些重复序列除了两个相邻的重复可变二残基(RVD)之外是相同的,该重复可变二残基通过介导结合到单个的核苷酸而赋予DNA特异性。已经描述了结合到DNA位点的类似数目的碱基对(bp)的超过30个重复序列的阵列。虽然每个RVD的结合中固有简并性,但是最近的报告表明,合成的TALE蛋白质具有足够的特异性以靶向人类基因组内的单个位点。

[0006] 通过嵌合核酸酶(例如,锌指核酸酶(ZFN))引入DNA双链断裂(DSB)可以用来敲除基因功能或者在外源添加的DNA的存在下驱动在目标位点的整合。在过去十年中,ZFN已被广泛研究,并且在某些情况下,正在接近临床应用进行基因治疗。最近,一些团体已经探索了利用将TALE DNA结合结构域与核酸酶(TALEN)融合进行靶向基因组编辑。事实上,许多使用ZFN的工作已使用TALE核酸酶来复制,因为较之ZFN,TALEN可具有关于DNA结合模块化的优点。然而,尽管对ZFN和TALEN进行了令人印象深刻的研究,还仍然存在关于其安全性和特异性的问题。特别是,脱靶裂解事件仍难以检测,脱靶DSB最有可能的结果是引入小插入或小缺失。此外,DSB的修复依赖于随细胞类型而变化的细胞机制。

[0007] 一个实现靶向基因组修饰的替代方法是使用位点特异性重组酶(SSR)。诸如酪氨酸重组酶Cre和Flp之类的SSR是被常规用于操纵在细胞内的染色体结构的有价值的分子生物学工具。因为这些酶依赖于若干复杂的蛋白质-蛋白质和蛋白质-DNA相互作用以协调催

化,SSR表现出显著的靶位点特异性。然而,迄今为止,已经证明许多SSR的特异性的改变非常困难。解离酶型/转化酶型丝氨酸重组酶为酪氨酸重组酶进行基因组工程提供了灵活多样的可变性。在自然界中,这些酶具有以高度模块化的方式协调重组的多域蛋白复合物的功能。然而,几种丝氨酸重组酶突变体已经确定不需要用于重组的辅助因子。此外,许多研究已经表明,丝氨酸重组酶的天然DBD能够被定制设计的ZFP代替,以产生嵌合锌指重组酶(ZFR)。原则上,可以产生能够识别扩展数目的序列的ZFR,但是,由于缺乏能够识别所有可能的DNA三联体的锌指结构域,导致限制这些酶的潜在的模块化靶向能力。

[0008] ZFR是由来源于解离酶系/转化酶系的丝氨酸重组酶的激活的催化结构域和能被定制设计以识别几乎任何DNA序列的锌指DNA结合结构域组成(图30A)。ZFR催化特定ZFR靶位点之间的重组,该特定ZFR靶位点由侧接由重组酶催化结构域识别的中央20bp核心序列的双倒位(two-inverted)的锌指结合位点(ZFB)组成(图30B)。与锌指核酸酶(ZFN)和TAL效应物核酸酶(TALEN)相比,ZFR自动发挥作用并能在不激活细胞DNA损伤应答途径的情况下切除并整合人类和小鼠细胞中的转基因。然而,与常规的位点特异性重组酶一样,ZFR的应用已经受到由重组酶催化结构域强加的序列要求的限制,其决定了ZFR靶位点包含来源于天然丝氨酸解离酶/转化酶重组位点的20-bp核心。

[0009] 诸如Cre-loxP、FLP-FRT和^ΔC31-att之类的位点特异性DNA重组系统已成为基因工程的强大工具。促进这些DNA重排的位点特异性重组酶识别短的(30-bp至40-bp)序列并且通过不需要DNA合成或高能辅助因子的机制协调DNA裂解、链交换以及重新连接。这种简单性使得研究人员能够以非凡的空间和时间的敏感性研究基因功能。然而,由位点特异性重组酶所强加的严格的序列要求已限制它们应用于包含人工引入重组位点的细胞和生物体。为了解决此限制,定向进化已经被用于朝向天然存在的DNA序列改变几种重组酶的序列特异性。尽管取得了进步,但是对于复杂的诱变和选择策略的需要以及关于重新设计的重组酶变体通常表现出宽松的底物特异性的发现已阻碍了这种技术的广泛使用。

[0010] 因此,需要一种催化内源性基因组的靶向和位点特异性重组、尤其是用于基因治疗的更一般化的方法以及用于可以催化这类靶向和位点特异性重组的酶。这种方法对于基因疗法特别有用,但也在分子生物学领域具有许多其它应用,该应用包括在基因克隆中的应用以及在工业微生物和农业植物和动物的修饰中的应用。

发明内容

[0011] 本文公开了用于产生转基因细胞、组织、植物和动物的靶向嵌合多肽,其包括它们的组合物、表达载体、以及使用它们的方法。本发明的组合物、载体和方法在基因治疗技术中也是非常有用的。

[0012] 一方面,本发明提供了嵌合多肽。该多肽包括:a)重组酶、核酸酶或转录因子、或其片段;和b)转录激活因子样效应因子(TALE)蛋白。在各实施方式中,该TALE蛋白是截短的,并且包括C-末端或N-末端截短。在实施方式中,TALE蛋白是AvrXa7、Tal1c和PthXo1。在实施方式中,TALE蛋白包括如SEQ ID NO:2中所列的所有或一部分氨基酸序列。在一些实施方式中,TALE蛋白是在SEQ ID NO:2的氨基酸残基27和268、92和134、120和129、74和147、或87和120之间被截短。在一些实施方式中,TALE蛋白是在SEQ ID NO:2的氨基酸残基28、74、87、92、95、120、124、128、129、147和150处被截短。

[0013] 另一方面,本发明提供产生特异性结合期望的核苷酸的转录激活子样效应因子(TALE)蛋白结合结构域的方法。该方法包括:a)通过使可变二残基(RVD)内的氨基酸残基突变,或者通过使该RVD的1至2个氨基酸残基N-末端或C-末端内的氨基酸残基突变使TALE蛋白结合结构域的氨基酸序列随机化;以及b)选择(a)中的随机化的TALE蛋白结合结构域,其中TALE蛋白结合结构域特异性结合到期望的核苷酸。

[0014] 另一方面,本发明提供分离的多肽,其包括黄单胞杆菌属衍生的转录激活子样效应因子(TALE)蛋白,该TALE蛋白具有包括如SEQ ID NO:3 (VGKQWSGARAL)中所列的氨基酸序列的N-末端结构域(NTD),该氨基酸序列具有选自以下的一个或多个突变或缺失:Q是Y、Q是S、Q是R、W是R、W是G、W缺失、S是R、S是H、S是A、S是N、以及S是T。

[0015] 另一方面,本发明提供分离的多肽,其包含罗尔斯通菌属衍生的转录激活子样效应因子(TALE)蛋白,该TALE蛋白具有包含如SEQ ID NO:8 (IVDIAR₁QR₂SGDLA)中所列的氨基酸序列的N-末端结构域(NTD),该氨基酸序列具有选自以下的一个或多个突变或缺失:R₁是K、Q是Y、Q是S、Q是R、R₂是W、R₂是G、R₂缺失、S是R、S是H、S是A、S是N、以及S是T。

[0016] 在另一个实施方式中,本发明提供产生转录激活子样效应因子(TALE)蛋白N-末端结构域(NTD)的方法。该方法包括:a)通过使NTD内的一个或多个氨基酸残基突变或缺失而使NTD的氨基酸序列随机化,其中该氨基酸序列是SEQ ID NO:14 (VGKXXXGAR)或SEQ ID NO:15 (VDIAXXXXGDLA);以及b)选择(a)中的随机化的TALE蛋白NTD,其中TALE蛋白NTD特异性结合到期望的核苷酸或表现出增强的活性。

[0017] 本文还公开了用于产生转基因细胞、组织、植物和动物的嵌合蛋白,其包括丝氨酸重组酶和一种或多种锌指结合结构域、产生ZFR的方法、它们的组合物、表达载体、以及使用它们的方法。本发明的组合物、载体和方法在基因治疗技术中也是非常有用的。

[0018] 一方面,本发明提供产生具有比对应的野生型重组酶高的催化特异性的多个锌指重组酶(ZFR)蛋白的方法。该方法包括在重组酶催化结构域的相当于Gln 11e120、Thr123、Leu127、Ile136和Gly137或其组合的位置进行随机诱变,使在每个氨基酸的位置2和位置3处的DNA突变;使重组酶催化结构域与多个锌指结合结构域融合以形成ZFR,并且富集具有比对应的野生型重组酶高的催化特异性的ZFR。在一些实施方式中,ZFR对选自GC、GT、CA、TT和AC的DNA靶具有增强的催化活性。在一个实施方式中,重组酶催化结构域在Ile136和/或Gly137处发生诱变。

[0019] 在各方面中,本文所述的嵌合多肽包括来源于如本文所公开的以下物质或者由如本文所公开的以下物质随机诱变的重组酶催化结构域:a) Tn3,也称为EcoTn3;Hin,也称为StyHin;Gin,也称为MuGin;Sin;Beta;Pin;Min;Din;Cin;EcoTn21;SfaTn917;BmeTn5083;Bme53;Cpe;SauSK1;SauSK41;SauTn552;Ran;Aac;Lla;pMER05;Mlo92;Mlo90;Rrh;Pje;Req;PpsTn5501;Pae;Xan;ISXc5;Spy;RhizY4cG;SarpNL1;Sso1SC1904a;Sso1SC1904b;Sso1SC1913;Aam606;MjaM0014;Pab;Hpy1S607;Mtu1S_Y349;MtuRv2792c;MtuRv2799c;MtuRv3828c;MtuRv0921;MceRv0921;TnpX;TndX;WwK;乳球菌(lactococcal)噬菌体TP901-1丝氨酸重组酶;化脓性链球菌噬菌体Φ 370.1丝氨酸重组酶;化脓性链球菌噬菌体Φ FC1丝氨酸重组酶;李斯特菌属噬菌体A118丝氨酸重组酶;天蓝色链霉菌染色体SC3C8.24丝氨酸重组酶;天蓝色链霉菌染色体SC2E1.37丝氨酸重组酶;天蓝色链霉菌染色体SCD78.04c丝氨酸重组酶;天蓝色链霉菌染色体SC8F4.15c丝氨酸重组酶;天蓝色链霉菌染色体SCD12A.23

丝氨酸重组酶;天蓝色链霉菌染色体SCH10.38c丝氨酸重组酶;天蓝色链霉菌染色体SCC88.14丝氨酸重组酶;链霉菌噬菌体 ϕ C31丝氨酸重组酶;链霉菌噬菌体R4丝氨酸重组酶;芽孢杆菌噬菌体 ϕ 105丝氨酸重组酶;芽孢杆菌噬菌体SPBc2丝氨酸重组酶;芽孢杆菌前噬菌体SKIN丝氨酸重组酶;金黄色葡萄球菌ccrA丝氨酸重组酶;金黄色葡萄球菌ccrB丝氨酸重组酶;结核分枝杆菌噬菌体Bxb1丝氨酸重组酶;结核分枝杆菌前噬菌体 ϕ RV1丝氨酸重组酶;YBCK_ECOLI;Y4bA;Bja;Spn;Cac 1956;和Cac 1954;或b) a)的突变蛋白。

[0020] 在又一方面,本发明提供分离的核酸分子,其编码本文所述的嵌合多肽。

[0021] 在又一方面,本发明提供表达盒,其包含编码本文所述的嵌合多肽的核酸分子。

[0022] 在又一方面,本发明提供载体,其包含本文所述的表达盒。

[0023] 在又一方面,本发明提供分离的宿主细胞,其含有本文所述的载体。

[0024] 在又一方面,本发明提供用于位点特异性整合入DNA序列的方法。该方法包括使DNA序列与本发明的嵌合多肽接触,其中该嵌合多肽催化位点特异性整合。

[0025] 在又一方面,本发明提供用于基因治疗的方法。该方法包含向受试者施用包括编码本文所述的嵌合多肽的核酸分子的组合物,其中该核酸分子一经表达,存在于受试者的基因组中的基因就被特异性地去除或失活。

[0026] 在又一方面,本发明提供药物组合物。该组合物包含本文所述的嵌合多肽;以及药学上可接受的载体。在另一方面,该组合物包含编码本文所述的嵌合多肽的核酸分子;以及药学上可接受的载体。

[0027] 在又一方面,本发明提供通过本发明的嵌合多肽催化重组产生的转基因生物。

[0028] 在又一方面,本发明提供用于基因治疗的方法。该方法包含向受试者施用包括具有通过本文所述的位点特异性整合的方法产生的DNA序列的核酸分子的细胞。

[0029] 在另一方面,本发明提供分离的核酸分子,其编码本文所述的嵌合蛋白。

[0030] 在另一方面,本发明提供用于位点特异性重组的方法。该方法包括:a)提供包括用于与本文所述的嵌合蛋白特异性相互作用的至少两个结合位点的DNA序列;和b)使该DNA序列与该嵌合蛋白发生反应,其中该嵌合蛋白催化与该嵌合蛋白特异性相互作用的两个位点之间的位点特异性重组事件,在该事件中,该DNA序列的两条链都被切割。

附图说明

[0031] 图1是一系列关于TALER融合取向和活性的图形表示和图解表示。A)卡通说明用来评估TALER活性的拆分 β 内酰胺酶系统。B)示意性表示每个TALER及其相应的靶位点(1=SEQ ID NO:288;2=SEQ ID NO:289;3=SEQ ID NO:290)的融合取向。C)每个设计的TALER融合对其预期DNA靶的活性。将重组相对背景(空载体作为参照)归一化。D)Gin-Avr对同源(Avr-20G)DNA靶和对非同源(Avr-20T、Avr-20GG、PthXo1-20G)DNA靶的活性。误差线表示标准偏差(s.d.) (n=3)。

[0032] 图2是一系列关于选定的TALER截短的重组谱的图形表示和图解表示。A)示意性说明20-成员TALER截短库的设计。B)选定的TALER变体对含有增加长度(14-bp、20-bp、26-bp、32-bp和44-bp)的核心序列的DNA靶的活性。C)Gin-AvrXa7 Δ 120对含有非同源核心序列或增加长度的核心位点的多样化底物的活性。误差线表示s.d. (n=3)。

[0033] 图3是一系列关于选自增量截短库的TALER变体的图形表示。A)选定的TALER截短

变体的频率。经过3轮选择之后,将递增截短Gin-AvrXa7变体进行分离并使用DNA测序来测定截短长度。B) 递增截短TALER变体(长度介于 $\Delta 92$ 和 $\Delta 134$ 之间)对Avr-32G DNA靶的活性。为了方便参考,最短($\Delta 145$)截短变体和最长($\Delta 74$)截短变体、以及 $\Delta 87$ 截短变体都包括在内。C) Gin-Avr $\Delta 74$ 、Gin-Avr $\Delta 128$ 和Gin-Avr $\Delta 145$ 对多样化同源DNA靶和非同源DNA靶的活性。误差线表示s.d. (n=3)。

[0034] 图4是一系列关于合成TALER的活性的图形表示。A) 合成Gin-Avr15 $\Delta 128$ 、Gin-Avr15 $\Delta 120$ 和Gin-Pht15 $\Delta 120$ 变体对DNA靶Avr-32G或Pth-32G的活性。B) 具有基于Gin-Avr $\Delta 120$ 的长度介于15个重复序列和20个重复序列之间的DBD的合成TALER对Avr-32G和Avr-32T的活性。误差线表示s.d. (n=3)。

[0035] 图5是一系列关于TALER在哺乳动物细胞中的活性的图形表示。图5A和5B均表示HEK293T细胞中萤光素酶表达减少的倍数,其中(A)为TALER或ZFR表达载体(Gin-Avr $\Delta 120$ 和GinC4)在报告质粒(Avr-32G、Avr-44G和C4-20G)的存在下共转染、(B)表示TALER和ZFR表达载体(Gin-Avr $\Delta 120$ +GinC4)相结合与报告质粒(Avr-G-ZF)共转染。误差线表示s.d. (n=3)。

[0036] 图6是AvrXa7 (SEQ ID NO:1 DNA序列;SEQ ID NO:2 氨基酸序列)的设计的N-末端截短的引物的位置的图解表示。星表示 $\Delta 120$ 融合点的位置。

[0037] 图7是AvrXa7靶序列(SEQ ID NO:16-SEQ ID NO:18)的天然野生型RDV结构域和合成RDV结构域的比较的图解表示。

[0038] 图8是AvrXa7蛋白(SEQ ID NO:19)的TALE氨基酸序列和TALER氨基酸序列的图解表示。

[0039] 图9是构建体AvrXa7 DNA序列(SEQ ID NO:20)的图解表示。

[0040] 图10是构建体Gin-Avr $\Delta 74$ 氨基酸序列(SEQ ID NO:21)的图解表示。

[0041] 图11是构建体Gin-Avr $\Delta 87$ 氨基酸序列(SEQ ID NO:22)的图解表示。

[0042] 图12是构建体Gin-Avr $\Delta 120$ 氨基酸序列(SEQ ID NO:23)的图解表示。

[0043] 图13是构建体Gin-Avr $\Delta 120^*$ 氨基酸序列(SEQ ID NO:24)的图解表示。

[0044] 图14是构建体Gin-Avr $\Delta 147$ 氨基酸序列(SEQ ID NO:25)的图解表示。

[0045] 图15是构建体GinAvr15 $\Delta 128$ -合成蛋白氨基酸序列(SEQ ID NO:26)的图解表示。

[0046] 图16是构建体Gin-Avr15 $\Delta 128$ -合成蛋白DNA序列(SEQ ID NO:27)的图解表示。

[0047] 图17是构建体GinAvr15 $\Delta 128$ -合成蛋白氨基酸序列(SEQ ID NO:28)的图解表示。

[0048] 图18是一系列关于TALE N-末端结构域的特异性的图像表示和图形表示。A) TALE (SEQ ID NO:29)结合到其靶DNA的示意图。B) 结构分析表明5'T与N-1发夹(N-0-SEQ ID NO:30;N-1-SEQ ID NO:31和RVD-SEQ ID NO:32)的W232接触。这种发夹与RVD发夹共享显著的序列同源性。C-F) 在C) AvrXa7TALE-R的背景下分析NT-T (wt) NTD,在D) AvrXa7TALE-TF的背景下分析NT-T (wt) NTD,在E) AvrXa7MBPTALE的背景下分析NT-T (wt) NTD,以及在F) CCR5靶TALEN的背景下分析NT-T (wt) NTD。(相比于5'T, *=P<0.05, **=P<0.01, ***=p<0.001)。

[0049] 图19是一系列关于重组酶变体的图形表示和图解表示。A-C) 重组酶选择变体对具有A) 5'G的底物、具有B) 5'A的底物、具有C) 5'C的底物的活性。图19D是优化的TALE NTD (SEQ ID NO:33-SEQ ID NO:36)的比对,说明在N-1发夹中的序列差异性。E) 优化的NTD活性在MBP-TALE AvrXa7的背景下的综合比较。(相比于野生型和5'A/G/C, *=P<0.05, **=P<

0.01,***=p<0.001)。

[0050] 图20是一系列选定的NTD在TALE-TF的背景下的分析的图解表示和图形表示。A)用于转录激活实验的荧光素酶报告质粒上的5xAvr启动子区(SEQ ID NO:37)的示意图。B)具有NT-T、NT-G、NT-aN和**NT-aN**结构域的TALE-TF对具有所示5'残基的底物的相对荧光素酶激活。(相比于NT-T和各自的5'A/G/C/T,*=P<0.05,**=P<0.01,***=p<0.001)。

[0051] 图21是一系列具有野生型NTD和用不同的5'碱基演变的NTD的TALEN对的设计和活性的图解表示和图形表示。A)展开CCR5基因(SEQ ID NO:38-SEQ ID NO:39)以突出用来诱导H32突变的靶位点(SEQ ID NO:40-SEQ ID NO:47)。B)野生型(NT-T)TALEN、具有用非-T5'残基优化的结构域的TALEN、以及dHax3NTD的基因编辑效率。C)具有优化的NTD的TALEN对与具有5'T特异性的TALEN的增强倍数。每个NTD的活性示于每个TALEN对底物上。

[0052] 图22是显示N-末端和C-末端结构域(SEQ ID NO:48-SEQ ID NO:53)的比对的图解表示。

[0053] 图23是说明TALE-重组酶选择方案的示意性表示。使用Not1/StuI限制性内切酶和互补连接作用将NTD库克隆到Avr15TALE-R中。TALE-R的激活导致可以用抗生素(羧苄青霉素)选择和扩增的更频繁的重组事件。得到的输出质粒为酶切的Not1/XbaI并且被连接到TALE-R骨架载体中以便进一步的选择和扩增。

[0054] 图24是发现于库选择(库XXXSGAR(SEQ ID NO:39)和库KXXGAR(SEQ ID NO:291))的变体群的概要的图解表示。

[0055] 图25是显示NT-G(SEQ ID NO:54)与NTD-Brg11(SEQ ID NO:55)、罗尔斯通菌属TALE结构域的排列。排列表明Brg11可以表现出对5'G碱基的特异性。

[0056] 图26是一系列如通过ELISA所分析的MBP-TALE蛋白对靶5'A/G/C/T Avr15发夹寡核苷酸的相对结合亲和力的图形表示。蛋白浓度为~75nM并且板被显色120分钟。

[0057] 图27是一系列在用于右侧的%插入/缺失和插入/缺失群编辑TALEN后PCR扩增的CCR5的细胞分析的图像表示和图形表示。

[0058] 图28是选自图27(从顶部到底部SEQ ID NO:292-SEQ ID NO:332)的TALEN实验的对准插入/缺失测序的图解表示。

[0059] 图29是两个分开的Go1dy TALE-转录因子架构的活性的比较的图形表示,其中每个靶向仅在5'残基不同的相同的5x AvrXa7启动子。

[0060] 图30是一系列有关结合到DNA的锌指重组酶二聚体的结构的图解表示。A)每个锌指重组酶(ZFR)单体(蓝色或橙色)由连接到定制设计的锌指DNA结合结构域的激活的丝氨酸重组酶催化结构域组成。模型是从 γ δ 解离酶和Aart锌指蛋白(分别为PDB ID:1GDT和2I13)的晶体结构产生的。B)ZFR二聚体结合到DNA(SEQ ID NO:333-SEQ ID NO:334)的卡通图。ZFR靶位点由侧接由ZFR催化结构域识别的中央20bp核心序列的双倒位的锌指结合位点(ZFB)组成。锌指蛋白(ZFP)可以被设计为识别“左”或“右”半位点(分别为蓝色框和橙色框)。缩写如下:N表示A、T、C、或者G;R表示G或A;以及Y表示C或T。

[0061] 图31是一系列Gin重组酶催化结构域的特异性的图形表示和图解表示。A-D)于(A, SEQ ID NO:335)在二核苷酸核心处含有每个可能的两个碱基的組合的DNA靶上测量重组、(B, SEQ ID NO:336)在位置3和位置2处含有每个可能的两个碱基的組合的DNA靶上测量重组、(C, SEQ ID NO:337)在位置6、位置5和位置4处含有每个可能的单个碱基取代的DNA靶上

测量重组、以及(D, SEQ ID NO:338)在位置10、位置9、位置8和位置7处含有每个可能的单个碱基取代的DNA靶上测量重组。在每个图的上方,取代的碱基被加以方框表示。重组是通过分裂基因重新组装评估的并且以羧苄青霉素-耐药性转化子与氯霉素-耐药性转化子的比例来测量的(材料和方法)。误差线表示s.d. (n=3)。(E) γ δ 解离酶二聚体与(左)二核苷酸核心、(中间)位置6、位置5和位置4、以及(右)位置10、位置9、位置8和位置7(PDB ID:1GDT)处的DNA之间的相互作用。相互作用的残基显示为紫色棒。碱基的颜色如下:A,黄色;T,蓝色;C,褐色;和G,粉红色。

[0062] 图32是一系列重新设计的Gin重组酶催化特异性的图形表示和图解表示。A) 由该Gin催化结构域识别的规范的20-bp核心。图中方框表示位置3和位置2(SEQ ID NO:339)。B) (顶部) γ δ 解离酶与DNA(PDB ID:1GDT)复合物的结构。为诱变选择的侧臂区域残基显示为紫色棒。(底部) γ δ 解离酶(SEQ ID NO:341)和Gin重组酶(SEQ ID NO:342)催化结构域的序列比对。保守残基显示为橙色阴影。黑色箭头表示为诱变选择的侧臂区域位置。C) 断裂基因重装配选择系统的示意图。活性ZFR变体的表达导致 β 内酰胺酶的阅读框的恢复和宿主细胞对氨苄青霉素的耐药性。实线表示ZFR靶位点的位置和识别性。位置3和位置2标有下划线(SEQ ID NO:340)。D) 重组在位置3和位置2含有GC、GT、CA、TT和AC碱基组合的核心位点的Gin突变体的选择。星号表示选择步骤,其中培养时间从16小时减少到6小时(材料和方法,实施例5)。E) 为在位置3和位置2处的每个可能的两碱基组合所选择的催化结构域(由a指示的 β 、 γ 、 δ 、 ϵ 、和 ζ 、野生型Gin)的重组特异性。预期的DNA靶标有下划线。重组通过断裂基因重装配确定,并且以一式三份进行。

[0063] 图33是一系列说明ZFR在哺乳动物细胞中重组用户定义的序列的能力的图形表示和图解表示。A) 用于评估哺乳动物细胞中的ZFR活性的萤光素酶报告系统的示意图。ZFR靶位点侧接驱动萤光素酶表达的SV40启动子。实线表示用于识别潜在的ZFR靶位点的44-bp共有靶序列。下划线碱基表示锌指靶以及位置3和位置2(SEQ ID NO:343)。B) 萤光素酶在用设计的ZFR对以及它们的同源报告质粒共转染的HEK293T细胞中的表达的减少倍数。减少倍数被归一化至用空载体和报告质粒共转染。示出了每个ZFR靶位点(从顶部到底部SEQ ID NO:344-SEQ ID NO:362)的序列同一性和染色体定位以及每个ZFR对的催化结构域组成。下划线碱基表示位置3和位置2。从三个独立的实验计算标准误差。在表2中提供了ZFR氨基酸序列。C) ZFR对的特异性。为ZFR对1至ZFR对9以及每个非同源报告质粒的GinC4测定萤光素酶表达的减少倍数。重组被归一化至每个ZFR对与它的同源报告质粒的减少倍数。以一式三份进行分析。

[0064] 图34是一系列说明ZFR靶向整合到人类基因组的能力的图形表示和图解表示。A) 供体质粒(顶部)和由ZFR 1(SEQ ID NO:363)、ZFR2(SEQ ID NO:364)和ZFR 3(SEQ ID NO:365)靶向的基因组位点的示意图。打开框表示邻近的外显子。箭头指示转录方向。示出了每个ZFR靶的序列和位置。下划线碱基表示锌指靶以及位置3和位置2。B) ZFR介导的整合的效率。数据被归一化至来自仅用供体质粒转染的细胞的数据。误差线表示标准偏差(n=3)。C) ZFR介导的整合的PCR分析。PCR引物组合以正向方向(中部)或反向方向(底部)扩增(顶部)未修饰的位点或者扩增整合的质粒。D) ZFR 1(SEQ ID NO:366)和ZFR 3(SEQ ID NO:367)的PCR扩增整合的供体的代表性层析谱。箭头指示测序引物的方向。阴影框表示基因组靶序列。

[0065] 图35是重组酶DNA结合残基位于二聚体界面之外的图解表示。 γ δ 解离酶与靶DNA的复合物。催化结构域二聚体为青色。DNA为灰色。侧臂区域残基显示为红色棒。二聚体界面处的残基显示为紫色棒(PDBID:1GDT)。

[0066] 图36是选定的重组酶的序列分析的图解表示。饼图表示在每个靶定侧臂位置处的氨基酸取代的百分比。经过4轮的选择后,从每个库测序>20个克隆。重组TT的克隆的序列分析被描述在别处(1)。

[0067] 图37是显示分离的催化结构域的核心特异性的表。经过4轮的选择后,评估了选定的催化结构域重组在位置3和位置2具有取代基的核心序列的能力。指定的DNA靶标有下划线。示出了重组酶突变体。星号表示选择用于进一步分析的催化结构域。在位置3和位置2的野生型碱基组合为CC。重组通过断裂基因重装配(2)确定,并且以一式三份进行。重组TT取代基的催化结构域被描述在别处(1)。

[0068] 图38是一系列选定的催化结构域的位置特异性的图形表示。重组测定在 α 、 β 、 γ 、 δ 和 ζ 催化结构域与对称取代的靶位点之间进行。对在位置6、位置5和位置4处含有(A (SEQ ID NO:368))>4000随机强碱基(S:G或C)取代以及在位置10、位置9、位置8和位置7处含有(B (SEQ ID NO:369))>10⁶(一种可能为 4.29×10^9)独特碱基组合(N:A、T、C或G)的文库DNA靶点进行重组检测。重组通过断裂基因重装配(2)检测(n=3)。

[0069] 图39是一系列ZFR同二聚体活性的图形表示。用150ng ZFR-L或150ng ZFR-R和2.5ng相应pGL3ZFR报告质粒共转染HEK293T细胞。重组被归一化至用150ng ZFR-L和150ng ZFR-R和2.5ng相应pGL3 ZFR报告质粒共转染。

[0070] 图40是一系列描绘ZFR-修饰的细胞的克隆分析的图像表示。PCR引物组合以正向方向或反向方向扩增未修饰的基因组靶或者扩增整合的质粒。

具体实施方式

[0071] 本发明提供了TALE重组酶(TALER)的首次公开。使用递增截短的TALE结构域文库,鉴定了在细菌和哺乳动物细胞中可以用于重组DNA的优化的TALER架构。任何定制的TALE重复阵列都可以被插入到本文所述的TALER架构,从而显著地扩大了工程化的重组酶在生物技术和医药的应用中的靶向能力。

[0072] 转录激活子样效应因子(TALE)蛋白可以被设计成结合几乎任何DNA序列。用于TALE DNA结合结构域的设计的一般准则建议TALE所结合的DNA序列的5'最末端碱基(N₀碱基)应该是胸腺嘧啶。通过对在该位置的每个DNA碱基分析TALE转录因子(TALE-TF)、TALE重组酶(TALE-R)和TALE核酸酶(TALEN)的活性来对N₀要求进行定量分析。观察到:与含有5'T的靶序列相比,在没有5'T存在的情况下,TALE活性降低高达>1000倍的TALE-TF活性、降低高达100倍的TALE-R活性以及高达10倍的TALEN活性的减少。为了开发识别所有可能的N₀碱基的TALE架构,结构引导的文库设计加上TALE-R活性选择被用来进化新型TALE N末端结构域,以适应任何N₀碱基。分离并表征了G选择性结构域和广义反应性结构域。以TALE-R形式选择的工程化的TALE结构域表现出模块化并且在TALE-TF和TALEN架构有活性。进化的N末端结构域对于作为TALE结合蛋白和设计酶的任何DNA序列提供有效的和不受约束的基于TALE的靶定。

[0073] 此外,为了解决序列要求的限制,描述了重新工程化丝氨酸重组酶催化特异性的

基于知识的方法。这种基于特异性决定DNA结合残基的饱和诱变的策略被用来产生特异性的转变 $>10,000$ 倍的重组酶变体。重要的是,这种方法专门集中在位于重组酶二聚体界面外部的氨基酸残基(图35)。结果,确定重新工程化的催化结构域可以结合形成ZFR异二聚体,并且这些设计的ZFR对以卓越的特异性重组预先确定的DNA序列。总之,这些结果使我们推测由该方法开发的专门的催化结构域的扩展目录可以被用来产生具有定制特异性的ZFR。这里,底物特异性分析和定向进化的组合被用来开发能够识别估计 4×10^8 个独特20-bp核心序列的多样化的Gin重组酶催化结构域。结果表明,从这些重新设计的催化结构域装配的ZFR以高特异性重组用户定义的序列并整合DNA至人类细胞中的靶向内源性位点。这些结果表明,ZFR技术具有用于包括基因组工程和基因治疗的各种各样的应用的潜力。

[0074] 在描述本发明的组合物和方法之前,应当理解的是,本发明不限于所描述的具体组合物、方法和实验条件,因为这些装置、方法和条件可以变化。也应该理解的是,本文所用的术语仅用于描述特定的实施方式,并不是用于限制,由于本发明的范围仅限于所附的权利要求书。

[0075] 如本说明书及所附的权利要求中所使用的,单数形式“一(a)”、“一(an)”和“所述(the)”包括复数形式,除另有明确说明外。因此,例如,本领域技术人员阅读本发明等后,能清楚提到“所述组合物”或“所述方法”时,包括一种或多种组合物和方法,和/或本文描述的类型步骤。

[0076] 除非另有说明,本文使用的所有技术术语和科技术语具有本领域技术人员所理解的一般含义。虽然与本文描述相似或相同的任何方法和材料都能用于进行或测试本发明,但是现在描述的是优选的方法和材料。

[0077] “重组酶”是介导由重组酶识别的特定的DNA序列之间的位点特异性重组的酶家族((Esposito,D.,和Scocca,J.J.,*Nucleic Acids Research* 25,3605-3614(1997);Nunes-Duby,S.E.,等人,*Nucleic Acids Research* 26,391-406(1998);Stark,W.M.,等人,*Trends in Genetics* 8,432-439(1992))。

[0078] 如本文所用,术语“嵌合TALE重组酶”包括但不限于具有TALE结构域的重组酶,所述TALE结构域来自天然存在的TALE蛋白或者合成的TALE蛋白或者具有序列特异性结合活性的结构域。

[0079] 如本文所用,术语“嵌合锌指重组酶”包括但不限于具有锌指结合结构域的重组酶,所述锌指结合结构域来自天然存在的锌指DNA结合蛋白或者合成的锌指结合蛋白或者具有序列特异性结合活性的结构域。

[0080] 如本文所用,术语“锌指”、“锌指核苷酸结合结构域”,或类似术语既指天然存在的锌指,也指人工产生的锌指。此类锌指可以具有多种构架结构,例如,但不限于,C2H2、C4、H4、H3C、C3X、H3X、C2X2和H2X2,其中X为锌连接的氨基酸。在这些构架结构中,如锌指结构的描述中常用的,“C”代表半胱氨酸残基以及“H”代表组氨酸残基。具有构架C2H2的锌指包括,但不限于,如在Barbas等人的国际公开号W02008/006028、Barbas的美国专利号7,101,972、Barbas等人的美国专利号7,067,617、Barbas等人的美国专利号6,790,941、Barbas的美国专利号6,610,512、Barbas等人的美国专利号6,242,568、Barbas等人的美国专利号6,140,466、Barbas的美国专利号6,140,081、Barbas的美国专利申请公开号20060223757、Barbas等人的美国专利申请公开号20060211846、Barbas等人的美国专利申请公开号

20060078880、Barbas的美国专利申请公开号20050148075、Barbas等人的美国专利申请公开号20050084885、Barbas等人的美国专利申请公开号20040224385、Barbas等人的美国专利申请公开号20030059767、和Barbas等人的美国专利申请公开号20020165356中描述的,通过这种引用将它们都并入本文。其它锌指在Rebar等人的美国专利号7,067,317、Liu等人的美国专利号7,030,215、Rebar等人的美国专利号7,026,462、Case等人的美国专利号7,013,219、Cox III等人的美国专利号6,979,539、Case等人的美国专利号6,933,113、Cox III等人的美国专利号6,824,978、Eisenberg等人的美国专利号6,794,136、Eisenberg等人的美国专利号6,785,613、Case等人的美国专利号6,777,185、Choo等人的美国专利号6,706,470、Cox IM等人的美国专利号6,607,882、Case等人的美国专利号6,599,692、Cox II等人的美国专利号6,534,261、Case等人的美国专利号6,503,717、Eisenberg等人的美国专利号6,453,242、Rebar等人的美国专利申请公开号2006/0246588、Rebar等人的美国专利申请公开号2006/0246567、Case等人的美国专利申请公开号2006/0166263、Cox HI等人的美国专利申请公开号2006/0078878、Rebar等人的美国专利申请公开号2005/0257062、Cox III等人的美国专利申请公开号2005/0215502、Cox MI等人的美国专利申请公开号2005/0130304、Case等人的美国专利申请公开号2004/0203064、Case等人的美国专利申请公开号2003/0166141、Case等人的美国专利申请公开号2003/0134318、Eisenberg等人的美国专利申请公开号2003/0105593、Cox IM等人的美国专利申请公开号2003/0087817、Rebar等人的美国专利申请公开号2003/0021776、和Case等人的美国专利申请公开号2002/0081614中描述的,通过这种引用将它们都并入本文。例如,描述在这些专利和专利公开中的一个备选方法涉及使用所谓的“D-能够位点”和锌指组件或锌指DNA结合结构域,其可以结合到此类位点。“D-能够”位点是靶位点区,其允许适当设计的锌指组件或锌指DNA结合结构域结合到靶链的四个碱基而不是三个碱基。此类锌指组件或锌指DNA结合结构域结合到双链DNA靶区段(靶链)的一条链上的三个碱基的三联体和互补链上的第四个碱基。单个锌指与四个碱基靶区段的结合既对靶链的序列施加了限制,也对锌指的氨基酸序列施加了限制。

[0081] 如本文所用,本文中出现的各种氨基酸序列中的氨基酸是根据它们公知的三字母或单字母缩写来表示的。出现在各种DNA片段中的核苷酸,是用本领域常规使用的标准的单字母标识来指定。

[0082] 在肽或蛋白质中,氨基酸的合适的保守替代是本领域技术人员公知的并且通常可以产生所述替代而不改变所得分子的生物活性。本领域技术人员认识到,通常,多肽的非必需区域中的单个氨基酸替代不实质上改变生物活性(见例如,Watson等人Molecular Biology of the Gene, 4th Edition, 1987, Benjamin/Cummings, p. 224)。具体地,这种保守变体具有经修饰的氨基酸序列,使得所述改变不实质上改变该蛋白质的(保守变体的)结构和/或活性,例如,抗体活性、酶活性或者受体活性。这些包括氨基酸序列的保守修饰的变异,即对于蛋白质活性不关键的那些残基的氨基酸替代、添加或缺失,或者用具有相似性质(例如,酸性、碱性、带正电荷或负电荷的、极性或非极性等)残基替代氨基酸,使得甚至关键的氨基酸的替代也不实质上改变结构和/或活性。提供功能上相似的氨基酸的保守替代是本领域公知的。例如,选择保守替代的一个示例性教导包括(原始残基接着是示例性替代):Ala/Gly或Ser;Arg/Lys;Asn/Gln或His;Asp/Glu;Cys/Ser;Gln/Asn;Gly/Asp;Gly/Ala或Pro;His/Asn或Gln;Ile/Leu或Val;Leu/Ile或Val;Lys/Arg或Gln或Glu;Met/Leu或Tyr或

Ile;Phe/Met或Leu或Tyr;Ser/Thr;Thr/Ser;Trp/Tyr;Tyr/Trp或Phe;Val/Ile或Leu。备选
的示例性教导使用下面的六组,每个含有相互为保守替代的氨基酸:(1) 丙氨酸(A或Ala),
丝氨酸(S或Ser),苏氨酸(T或Thr);(2) 天冬氨酸(D或Asp),谷氨酸(E或Glu);(3) 天冬酰胺
(N或Asn),谷氨酰胺(Q或Gln);(4) 精氨酸(R或Arg),赖氨酸(K或Lys);(5) 异亮氨酸(I或
Ile),亮氨酸(L或Leu),甲硫氨酸(M或Met),缬氨酸(V或Val);和(6) 苯丙氨酸(F或Phe),酪
氨酸(Y或Tyr),色氨酸(W或Trp);(也见,例如,Creighton (1984) *Proteins*, W.H. Freeman
and Company;Schulz and Schimer (1979) *Principles of Protein Structure*, Springer-
Verlag)。本领域技术人员将理解上述替代不是仅仅可能的保守替代。例如,对于一些目的,
可以认为所有带电荷的氨基酸相互为保守替代,而不管它们是带正电荷或负电荷。此外,在
编码序列中改变、添加或缺失单个氨基酸或者小百分比的氨基酸的个体替代、缺失或添加
也可以被认为是“保守修饰的变异”(当将被递送的蛋白质的三维结构和功能被这种变异保
留时)。

[0083] 如本文所用,术语“表达载体”指本领域中已知的质粒、病毒、噬菌粒或其他载体,
其已经通过插入或掺入异源DNA (如编码本文的融合蛋白的核酸或本文提供的表达盒) 而被
操作。这种表达载体通常含有用于所插入的核酸在细胞中有效转录的启动子序列。表达载
体通常含有复制原点、启动子,以及允许所转化细胞的表型选择的特定基因。

[0084] 如本文所用,术语“宿主细胞”指其中载体可以增殖并且其DNA可以表达的细胞。该
术语也包括主题宿主细胞的任何后代。可以理解所有后代可以不与亲本细胞相同,因为在
复制期间可以发生突变。当使用术语“宿主细胞”时,包括这种后代。稳定转移(其中外源DNA
在宿主中被持续保持)的方法是本领域已知的。

[0085] 如本文所用,遗传治疗涉及将异源DNA转移到哺乳动物、尤其人的某些细胞、靶细
胞,所述哺乳动物患有该种治疗寻求的病症或状况。将DNA以这样的方式导入所选的靶细胞
使得异源DNA被表达并且产生其编码的治疗产物。替代地,异源DNA可以以某种方式介导编
码治疗产物的DNA的表达,或者它可以编码产物,如肽或RNA,其以某种方式直接或间接介导
治疗产物的表达。遗传治疗也可以用于递送编码基因产物的核酸,其替换它所导入的哺乳
动物或细胞的缺陷基因或者补充所述哺乳动物或细胞产生的基因产物。所导入的核酸可以
编码治疗化合物,如其生长因子抑制剂,或者肿瘤坏死因子或其抑制剂,如其受体,所述治
疗化合物不通常在哺乳动物宿主中产生或者不以治疗有效量或者治疗有用的时间产生。可
以在导入受折磨的宿主细胞前修饰编码治疗产物的异源DNA以便增强或改变其产物或表
达。遗传治疗也可以涉及递送基因表达的抑制剂或阻抑物或其他调节剂。

[0086] 如本文所用,异源DNA是这样的DNA,其编码通常不在其所表达的细胞体内产生的
RNA和蛋白质或者其通过影响转录、翻译或者其它可调节的生物化学过程介导或编码改变
内源DNA的表达的介体。异源DNA也可以被称作外源DNA。本领域技术人员将认识到或认为对
于其所表达的细胞异源或外源的任何DNA都在这里被异源DNA所包括。异源DNA的实例包括
但不限于,编码可追踪的标记蛋白(如赋予药物抗性的蛋白质)的DNA、编码治疗有效性物质
(如抗癌剂、酶和激素)的DNA,和编码其它类型蛋白质(如抗体)的DNA。异源DNA编码的抗体
可以被细胞所分泌或在细胞表面上表达,其中在所述细胞中已经导入了所述异源DNA。

[0087] 因此,本文的异源DNA或外源DNA包括这样的DNA分子,其不存在于在基因组中发现
的对应DNA分子的确切的方向和位置中。它也可以指来自另一生物或物种(即外源的)的DNA

分子。

[0088] 如本文所用,治疗有效产物是异源核酸(通常是DNA)编码的产物,当将所述核酸导入宿主时,表达产物,该产物减轻或消除遗传的或获得性疾病的症状、表现,或者治愈该疾病。通常,通过常规方法将编码期望的基因产物的DNA克隆到质粒载体并导入或显微注射到生产细胞,如包装细胞中,所述方法诸如磷酸钙介导的DNA摄入(见,(1981) *Somat.Cell.Mol.Genet.* 7:603-616)。在生产细胞中扩增后,将含有异源DNA的载体导入所选的靶细胞中。

[0089] 如本文所用,表达载体或递送载体指任何质粒或病毒,其中可以插入外源或异源DNA用于在合适的宿主细胞中表达,即该DNA编码的蛋白质或多肽在宿主细胞的系统中合成。能够指导编码一种或多种蛋白质的DNA区段(基因)表达的载体在本文中被称作“表达载体”。还包括允许使用逆转录酶从产生的mRNA克隆cDNA(互补DNA)的载体。

[0090] 如本文所用,基因指核酸分子,其核苷酸序列编码RNA或多肽。基因可以是RNA或DNA。基因可以包括编码区之前和之后的区域(前导序列和非转录尾区)以及单独的编码区段(外显子)之间的间插序列(内含子)。

[0091] 如本文所用,关于核酸分子或多肽或其它生物分子的术语“分离的”指该核酸或多肽已经与从中得到该多肽或核酸的遗传环境分离。它也可以指该生物分子已经从其天然状态改变。例如,天然存在于活的动物中的多核苷酸或多肽不是“分离的”,但是如该术语在本文中所用与其天然状态的共存物质分离的相同多核苷酸或多肽是“分离的”。从而,认为在重组宿主中产生和/或包含在重组宿主中的多肽或多核苷酸是分离的。还预期为“分离的多肽”或“分离的多核苷酸”是已经从重组宿主细胞或从天然来源纯化、部分或基本纯化的多肽或多核苷酸。例如,可以通过Smith等人(1988) *Gene* 67:3140中所述的一步方法基本上纯化重组产生形式的化合物。术语分离的和纯化的有时可互换使用。

[0092] 从而,“分离的”指核酸没有那些基因的编码序列,所述基因在自然存在的基因组中位于编码目的核酸的基因的直接侧翼。分离的DNA可以是单链或双链的,可以是基因组DNA、cDNA、重组杂种DNA或合成的DNA。它可以与天然DNA序列是相同的,或者可以通过缺失、添加或替代一个或多个核苷酸而与此类序列不同。

[0093] 如那些术语用于指从生物细胞或宿主制备的制备物一样,“分离的”或“纯化的”指含有所述DNA或蛋白质的任何细胞提取物,包括目的DNA或蛋白质的粗提物。例如,在蛋白质的情况下,可以按照单独的技术或者一系列制备或生物化学技术得到纯化制剂,并且目的DNA或蛋白质可以以不同的纯度存在于这些制剂中。具体对于蛋白质,所述方法可以包括例如但不限于,硫酸铵分级分离、凝胶过滤、离子交换层析、亲和层析、密度梯度离心、电聚焦、层析聚焦和电泳。

[0094] “基本上纯的”或“分离的”DNA或蛋白质制剂应该被理解为指这样的制剂,其没有天然中通常伴随着此类DNA或蛋白质的天然存在的物质。“基本上纯的”应该被理解为指“高度”纯化的制剂,其含有至少95%的目的DNA或蛋白质。

[0095] 含有目的DNA或蛋白质的细胞提取物应该被理解为指从表达目的蛋白质或含有目的DNA的细胞得到的匀浆制剂或无细胞的制剂。术语“细胞提取物”意在包括培养基,特别是已经去除细胞的耗尽的培养基。

[0096] 如本文所用,基因的启动子区包括通常位于结构基因的5'的一种或多种调节元

件;可以存在通过间插的核苷酸序列分开的多种调节元件。如果基因将被激活,那么称作转录因子的蛋白质附着到该基因的启动子区。该装配通过使得酶将第二遗传区段从DNA转录为RNA而类似于“通电”。在多数情况下,所得的RNA分子作为特定蛋白质合成的模板;有时RNA自身是终产物。启动子区可以是正常的细胞启动子或者例如,是癌启动子。癌启动子通常是病毒来源的启动子。可以靶定锌指结合多肽的病毒启动子包括但不限于,逆转录病毒长末端重复序列(LTR)、和慢病毒启动子,如来自人T细胞嗜淋巴细胞病毒(HTLV) 1和2的和人免疫缺陷病毒(HIV) 1或2的启动子。

[0097] 如本文所用,术语“截短的”或类似术语是指多肽衍生物,其含有比天然蛋白的完整氨基酸序列少的氨基酸序列,例如,ZFP、TALE或丝氨酸重组酶。

[0098] 如本文所用,多肽“变体”或“衍生物”是指这样的多肽,其为多肽的诱变形式或者通过重组产生的,但仍保留所期望的活性,例如结合配体或核酸分子或调节转录的能力。

[0099] 如本文所用,术语“药学上可接受的”、“生理学耐受的”和其语法变体当指组合物、载体、稀释剂和试剂时,可互换使用并且表示该物质能够施用于人而不产生不期望的生理学作用,如恶心、眩晕、胃不适等等,它们将一定程度上阻止该组合物的施用。

[0100] 如本文所用,术语“载体”指核酸分子,其能够将其已经有效连接的另一核酸在不同的遗传环境之间转运。优选的载体是能够自主复制和表达它们有效连接的DNA区段中存在的结构基因产物的那些载体。因此,载体优选含有复制子和先前描述的可选择的标记。载体包括但不必须局限于表达载体。

[0101] 如本文关于核酸分子,包括DNA片段所用的,短语“有效连接”指序列或区段优选通过常规的磷酸二酯键共价连接到单链或双链形式的DNA的一条链,使得被有效连接的部分如预期的发挥功能。如本领域公知的,本文提供的转录单元或盒有效连接的载体的选择直接取决于所希望的功能性质,例如载体复制和蛋白质表达,和将被转化的宿主细胞,这些是构建重组DNA分子领域中固有的限制。

[0102] 如本文所用,治疗组合物的施用可以通过任何方式实现,并且包括但不限于,经口、皮下、静脉内、肌内、胸骨内、灌注技术、腹膜内施用和肠胃外施用。

[0103] 转化细胞的方法是本领域公知的。通过“转化的”是指细胞中由外源DNA的摄取而产生的可遗传的改变。合适的方法包括病毒感染、转染、偶联、原生质体融合、电穿孔、粒子枪技术、磷酸钙沉淀、直接显微注射等。方法的选择通常取决于被转化的细胞类型和转化发生的环境(即体外、离体或体内)。可以在Ausubel等人,Short Protocols in Molecular Biology,3rd ed.,Wiley&Sons,1995中发现这些方法的一般讨论。

[0104] 术语“核酸分子”和“多核苷酸”可以互换使用并且是指任何长度的核苷酸的聚合形式,或者是脱氧核糖核苷酸或核糖核苷酸,或其类似物。多核苷酸可具有已知或未知的任何三维结构,并且可以执行已知或未知的任何功能。多核苷酸的非限制性实例包括基因、基因片段、外显子、内含子、信使RNA(mRNA)、转移RNA、核糖体RNA、核酶、cDNA、重组多核苷酸、分支多核苷酸、质粒、载体、任何序列的分离的DNA、任何序列的分离的RNA、核酸探针和引物。

[0105] “表达盒”包括能够指导目的基因/编码序列的表达的任何核酸构建体。为了将表达盒转移进靶细胞,可以将这种盒构造成“载体”、“载体构建体”、“表达载体”或“基因转移载体”。因此,该术语包括克隆和表达载体,以及病毒载体。

[0106] 用于测定核酸和氨基酸“序列同一性”的技术在本领域是公知的。通常地,此类技术包括测定基因的mRNA的核苷酸序列和/或测定由其编码的氨基酸序列,以及将这些序列与第二核苷酸或氨基酸序列相比较。一般而言,“同一性”是指两个多核苷酸或多肽序列的确切的核苷酸对核苷酸或氨基酸对氨基酸的分别对应。两个或更多个序列(多核苷酸或氨基酸)可通过测定它们的“百分比同一性”来比较。两个序列的百分比同一性,无论是核酸还是氨基酸序列,是两个比对的序列之间确切匹配的数目除以较短序列的长度,随后乘以100。核酸序列的近似比对利用Smith和Waterman的局部同源性算法,Advances in Applied Mathematics 2:482-489 (1981) 来提供。该算法可通过使用由Dayhoff,Atlas of Protein Sequences and Structure,M.O.Dayhoff编著,5suppl.3:353-358,National Biomedical Research Foundation,Washington,D.C.,USA开发的,和由Gribskov,Nucl.Acids Res.14 (6):6745-6763 (1986) 标准化的评分矩阵来用于氨基酸序列。这种测定序列的百分比同一性的算法的示例性实施方式由“BestFit”实用申请中的Genetics Computer Group (Madison,Wis.) 提供。这种方法的缺省参数被描述在the Wisconsin Sequence Analysis Package Program Manual,Version 8 (1995) (可从Genetics Computer Group,Madison,Wis.中获得) 中。在本发明的背景中,确立百分比同一性的优选的方法是使用MPSRCE程序包,爱丁堡大学拥有其版权,由John F.Collins和Shane S.Sturrok研发,并由IntelliGenetics,Inc. (Mountain View,Calif.) 分销。根据该组程序包,可以使用Smith-Waterman算法,其中评分表使用缺省参数(例如,空位开放罚分12、空位扩展罚分1、以及空位6)。根据产生的数据,“匹配”值反映了“序列同一性”。用于计算序列之间的百分比同一性或相似性的其它合适的程序在本领域通常是已知的,例如,另一个算法程序是使用缺省参数的BLAST。例如,可通过使用下列缺省参数来使用BLASTN和BLASTP:遗传密码=标准;过滤=无;链=两条;截断值=60;期望值=10;矩阵=BLOSUM62;描述=50个序列;排序依据=高分;数据库=非冗余的,GenBank+EMBL+DDBJ+PDB+GenBank CDS翻译+Swiss蛋白+Spupdate+PIR。

[0107] 可选地,通过在同源区域之间形成稳定双链体的条件下杂交多核苷酸,接着通过用单链特异性核酸酶(一种或者多种)消化,并且确定消化片段的大小,可以确定同源性。如使用上述方法所确定的,当序列在分子的限定长度内表现出至少约80%-85%、优选至少约85%-90%、更优选至少约90%-95%、以及最优选至少约95%-98%的序列同一性时,两个DNA或者两个多肽序列彼此“基本同源”。如本文所用,基本同源也指与特定的DNA或多肽序列表现出完全同一性的序列。基本同源的DNA序列可以在例如,该具体系统规定的严格条件下、以DNA杂交(Southern hybridization)实验鉴定。限定合适的杂交条件在本领域的技术范围内。参见例如,Sambrook等人,见上;DNA Cloning,见上;Nucleic Acid Hybridization,见上。

[0108] 正因为如此,本发明提供了核酸和编码本发明的嵌合多肽的氨基酸序列,其是基本同源的并且编码保持同等的生物活性的多肽。

[0109] 两个核酸片段如本文所描述的被视为“选择性杂交”。两个核酸分子之间的序列同一性程度影响此种分子之间的杂交事件的效率和强度。部分等同的核酸序列将至少部分抑制完全等同的序列与靶分子的杂交。完全等同序列的杂交抑制可以使用本领域公知的杂交测定进行评估(例如,DNA印迹、RNA印迹、溶液杂交等,参见Sambrook等人,Molecular

Cloning: A Laboratory Manual, Second Edition, (1989) Cold Spring Harbor, N.Y.). 此种测定可以使用不同程度的选择性来进行, 例如, 使用从低到高严格性不等的条件。如果采用低严格性的条件, 那么使用甚至缺乏部分序列同一性程度的二级探针 (例如, 与靶分子具有小于约30%序列同一性的探针) 也可以评估非特异性结合的不存在, 从而使得在不存在非特异性结合事件的情况下, 二级探针将不与靶杂交。

[0110] 当利用基于杂交的检测系统时, 选择这样的核酸探针, 其与靶核酸序列互补, 并且随后通过选择合适条件, 探针和靶序列彼此“选择性杂交”或结合, 以形成杂交分子。在“中等严格性”下能够与靶序列选择性杂交的核酸分子一般在这样的条件下杂交, 所述条件允许检测与所选择的核酸探针序列具有至少约70%序列同一性、长度至少约10-14个核苷酸的靶核酸序列。严格杂交条件一般允许检测与所选择的核酸探针序列具有超过约90-95%序列同一性、长度至少约10-14个核苷酸的靶核酸序列。用于其中探针和靶具有特定序列同一性程度的探针/靶杂交的杂交条件, 可以如本领域已知的进行测定 (参见例如, Nucleic Acid Hybridization: A Practical Approach, 编辑B.D.Hames和S.J.Higgins, (1985) Oxford; Washington, D.C.; IRL Press)。

[0111] 就用于杂交的严格性条件而言, 本领域众所周知的是, 通过改变例如下述因素众多等价条件可以用于确定特定严格性: 探针和靶序列的长度和性质、各种序列的碱基组成、盐和其它杂交溶液组分的浓度、杂交溶液中阻断剂 (例如, 甲酰胺、硫酸葡聚糖和聚乙二醇) 的存在或不存在、杂交反应温度和时间参数, 以及改变洗涤条件。特定杂交条件组的选择根据本领域的标准方法进行选择 (参见, 例如, Sambrook等人, Molecular Cloning: A Laboratory Manual, Second Edition, (1989) Cold Spring Harbor, N.Y.)。

[0112] 如果第一种多核苷酸与第二种多核苷酸、其cDNA、其互补体的区域具有相同或基本上相同的碱基对序列, 或如果它显示出如上所述的序列同一性, 则认为第一种多核苷酸“衍生自”第二种多核苷酸。

[0113] 如果第一种多肽 (i) 由衍生自第二种多核苷酸的第一种多核苷酸编码, 或 (ii) 显示如上所述的与第二种多肽的序列同一性, 则认为第一种多肽“衍生自”第二种多肽。

[0114] 位点特异性重组酶是用于基因组工程的强大工具。丝氨酸重组酶的解离酶/转化酶家族的过度活化变体在没有辅助因子的情况下发挥功能, 因此可以通过用工程化的锌指蛋白 (ZFP) 取代天然的DNA结合结构域重新靶向目的序列。

[0115] 本文所描述的锌指重组酶是嵌合酶, 其由衍生自丝氨酸重组酶的解离酶/转化酶家族和定制设计的锌指DNA结合结构域的活化的催化结构域组成。从设计的催化结构域装配的ZFR以高度特异性有效重组用户定义的DNA靶, 并且经设计的ZFR将DNA整合入人类细胞中的靶向内源性位点。

[0116] 在一方面, 本发明提供了产生具有比对应的野生型重组酶高的催化特异性的多个锌指重组酶 (ZFR) 蛋白的方法。该方法包括在重组酶催化结构域的相当于就野生型Gin催化结构域而言的Gin Ile120、Thr123、Leu127、Ile136和Gly137或其组合的位置上进行随机诱变, 使在每个氨基酸的位置2和位置3处的DNA突变; 使重组酶催化结构域与多个锌指结合结构域融合以形成ZFR, 并且富集具有比对应的野生型重组酶高的催化特异性的ZFR。在一些实施方式中, ZFR对选自GC、GT、CA、TT和AC的DNA靶具有增高的催化活性。在一个实施方式中, 重组酶催化结构域在Ile136和/或Gly137处发生诱变。

[0117] 如本文所用,野生型Gin催化结构域是指包括所有或部分的具有如下所示的列于SEQ ID NO: 56中的氨基酸序列的多肽的Gin催化结构域: MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGD TLV VWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTS SPMGRFFFYVMGALAEMERELI IERTMAGLAAARNKGRIGRPPKLTKAWEQAGRL LAQGIPRKQVALIYDVALSTLYKKHP。

[0118] 在各种实施方式中,本发明的嵌合多肽包括Gin催化结构域,例如由本发明的方法产生的那些Gin催化结构域。具体的Gin催化结构域包括列于表1中的那些Gin催化结构域。

[0119] 表1.Gin催化结构域

[0120]

Gin 催化结构域		
变体	SEQ ID NO:	序列
Gin α	57	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGD TLV VWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTS SPMGRFFFYVMGALAEMERELI IERTMAGLAAARNKGRIGRPPKSG
Gin β	58	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGD TLV VWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTS SPMGRFFFYVMGALAEMERELI IERTMAGLAAARNKGRIGRPPKS
Gin γ	59	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGD TLV VWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTS SPMGRFFFYVMGALAEMERELI IERTMAGLAAARNKGRIGRPPKSG
Gin δ	60	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGD TLV VWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTS SPMGRFFFYVMGALAEMERELI IERTMAGLAAARNKGRIGRPPKSG
Gin ϵ	61	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGD TLV VWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTS SPMGRFFFYVMGALAEMERELI IERTMAGLAAARNKGRIGRPPKSG
Gin ζ	62	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGD TLV VWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTS SPMGRFFFYVMGALAEMERELI IERTMAGLAAARNKGRIGRPPKSG
靶向侧臂区域位置被高亮显示。随机取代被加粗并加下划线。过度活化的 H106Y 突变被加下划线。		

[0121] 在各种实施方式中,通过本发明的方法产生的ZFR包括可操作地连接到多个锌指结合结构域的Gin催化结构域。由本发明产生的示例性的ZFR包括列于表2中的那些ZFR。

[0122] 表2. ZFR

[0123]

示例性 ZFR 的氨基酸序列		
ZFR-1 左	SEQ ID NO: 63	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGD TLVVWKL DRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFFVVMGALAEMERELI   MAG  AAARNKGR  G RPPKSGTG EKPYKCPEC GKSFS <u>SGDLRRHQ</u> RTHITG EKPYKCPEC G KSFS <u>SQSGDLRRHQ</u> RTHITG EKPYKCPEC GKSFS <u>SGDLRRHQ</u> RTHITG EKPYKCPEC GKSFS <u>SGELVRHQ</u> RTHITG KKTSGQAGQ
ZFR-1 右	SEQ ID NO: 64	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGD TLVVWKL DRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFFVVMGALAEMERELI   MAG  AAARNKGR  G RPPKSGTG EKPYKCPEC GKSFS <u>SHRTLTLN</u> HQ RTHITG EKPYKCPEC G KSFS <u>SQSGDLRRHQ</u> RTHITG EKPYKCPEC GKSFS <u>SQSGDLRRHQ</u> RTHITG EKPYKCPEC GKSFS <u>SQSGDLRRHQ</u> RTHITG KKTSGQAGQ
ZFR-2 左	SEQ ID NO: 65	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGD TLVVWKL DRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFFVVMGALAEMERELI   MAG  AAARNKGR  G RPPKSGTG EKPYKCPEC GKSFS <u>SQSGDLRRHQ</u> RTHITG EKPYKCPEC G KSFS <u>QRAHLERHQ</u> RTHITG EKPYKCPEC GKSFS <u>SGDLRRHQ</u> RTHITG EKPYKCPEC GKSFS <u>RSDCLVRHQ</u> RTHITG KKTSGQAGQ
ZFR-2 右	SEQ ID NO: 66	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGD TLVVWKL DRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFFVVMGALAEMERELI   MAG  AAARNKGR  G RPPKSGTG EKPYKCPEC GKSFS <u>SRSDKLVRHQ</u> RTHITG EKPYKCPEC G KSFS <u>RKDNLKNHQ</u> RTHITG EKPYKCPEC GKSFS <u>SGELVRHQ</u> RTHITG EKPYKCPEC GKSFS <u>SRSDKLVRHQ</u> RTHITG KKTSGQAGQ
ZFR-3 左	SEQ ID NO: 67	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGD TLVVWKL DRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFFVVMGALAEMERELI   MAG  AAARNKGR  G

[0124]

		RPPKSGTGEEKPYKCPECGKSFSSTGNLT ~~TVHQ~~ RTHTGEEKPYKCPECG KSFS ~~DPG~~ ALVRHQ ~~RTH~~ TGEEKPYKCPECGKSFS ~~QSS~~ NLVRHQ ~~RTH~~ TG EKP ~~YK~~ CPECGKSFS ~~RS~~ DHLTNHQ ~~RTH~~ TGKKTSGQAGQ
ZFR-3 右	SEQ ID NO: 68	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGD ~~TLVVWKL~~ DRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFF ~~FF~~ YVMGALAEMEREL ~~TER~~ MAG ~~AAA~~ ARNKGR ~~GG~~ RPPKSGTGEEKPYKCPECGKSFS ~~RKD~~ NLKNHQ ~~RTH~~ TGEEKPYKCPECG KSFS ~~RS~~ DHLTNHQ ~~RTH~~ TGEEKPYKCPECGKSFS ~~DPG~~ NLVRHQ ~~RTH~~ TG EKP ~~YK~~ CPECGKSFS ~~RKD~~ NLKNHQ ~~RTH~~ TGKKTSGQAGQ
ZFR-4 左	SEQ ID NO: 69	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGD ~~TLVVWKL~~ DRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFF ~~FF~~ YVMGALAEMEREL ~~TER~~ MAG ~~AAA~~ ARNKGR ~~GG~~ GRPPKSGTGEEKPYKCPECGKSFSQ ~~RAN~~ LRAHQ ~~RTH~~ TGEEKPYKCPEC GKSFS ~~QSS~~ SLVRHQ ~~RTH~~ TGEEKPYKCPECGKSFS ~~TTG~~ NLT ~~TVHQ~~ RTHT GEKPYKCPECGKSFSQ ~~RAH~~ LERHQ ~~RTH~~ TGKKTSGQAGQ
ZFR-4 右	SEQ ID NO: 70	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGD ~~TLVVWKL~~ DRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFF ~~FF~~ YVMGALAEMEREL ~~TER~~ MAG ~~AAA~~ ARNKGR ~~GG~~ GRPPKSGTGEEKPYKCPECGKSFSQ ~~RAN~~ LRAHQ ~~RTH~~ TGEEKPYKCPEC GKSFS ~~R~~ DELNVHQ ~~RTH~~ TGEEKPYKCPECGKSFSQ ~~LAH~~ LRAHQ ~~RTH~~ T GEKPYKCPECGKSFSQ ~~RAH~~ LERHQ ~~RTH~~ TGKKTSGQAGQ
ZFR-5 左	SEQ ID NO: 71	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGD ~~TLVVWKL~~ DRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFF ~~FF~~ YVMGALAEMEREL ~~TER~~ MAG ~~AAA~~ ARNKGR ~~GG~~ GRPPKSGTGEEKPYKCPECGKSFS ~~R~~ DELNVHQ ~~RTH~~ TGEEKPYKCPEC GKSFS ~~RS~~ DHLTNHQ ~~RTH~~ TGEEKPYKCPECGKSFSQ ~~LAH~~ LRAHQ ~~RTH~~ T GEKPYKCPECGKSFSQ ~~RAH~~ LERHQ ~~RTH~~ TGKKTSGQAGQ
ZFR-5 右	SEQ ID NO: 72	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGD ~~TLVVWKL~~ DRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFF ~~FF~~ YVMGALAEMEREL ~~TER~~ MAG ~~AAA~~ ARNKGR ~~GG~~ RPPKSGTGEEKPYKCPECGKSFS ~~TSG~~ SLVRHQ ~~RTH~~ TGEEKPYKCPECG KSFS ~~RS~~ DKLVRHQ ~~RTH~~ TGEEKPYKCPECGKSFS ~~QSG~~ DLRRHQ ~~RTH~~ TG EKP ~~YK~~ CPECGKSFS ~~TSG~~ ELVRHQ ~~RTH~~ TGKKTSGQAGQ

[0125]

ZFR-6 左	SEQ ID NO: 73	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGSRMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFFVVMGALAEMERELIERTMAGAAARNKGRIGR PPKSGTGEEKPYKCPECGKSFSQLAHLRAHQRTHTGEKPYKCPECGK SFSQLAHLRAHQRTHTGEKPYKCPECGKSFS ~~SDPGHLVRHQRTHTGE~~ KPYKCPECGKSFS ~~SDSGNLRVHQRTHTGKKTSGQAGQ~~
ZFR-6 右	SEQ ID NO: 74	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGSRMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFFVVMGALAEMERELIERTMAGAAARNKGRIGR GRPPKSGTGEEKPYKCPECGKSFSQRAHLERHQRTHTGEKPYKCPEC GKSFSSTTGNLTVHQRTHTGEKPYKCPECGKSFS ~~SDSGNLRVHQRTHT~~ GEKPYKCPECGKSFS ~~SOSSNLVRHQRTHTGKKTSGQAGQ~~
ZFR-7 左	SEQ ID NO: 75	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGSRMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFFVVMGALAEMERELIERTMAGAAARNKGRIGR PPKSGTGEEKPYKCPECGKSFS ~~THLDLIRHQRTHTGEKPYKCPECGKS~~ FSTTGNLTVHQRTHTGEKPYKCPECGKSFS ~~SOSSSLVRHQRTHTGEKPY~~ KPYKCPECGKSFS ~~SRSDNLVRHQRTHTGKKTSGQAGQ~~
ZFR-7 右	SEQ ID NO: 76	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGSRMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFFVVMGALAEMERELIERTMAGAAARNKGRIGR PPKSGTGEEKPYKCPECGKSFS ~~SRSDKLVRHQRTHTGEKPYKCPECGK~~ SFS ~~SRDELNVHQRTHTGEKPYKCPECGKSFS~~ ~~SOSSSLVRHQRTHTGE~~ KPYKCPECGKSFS ~~SRSDHLTNHQRTHTGKKTSGQAGQ~~
ZFR-8 左	SEQ ID NO: 77	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGSRMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFFVVMGALAEMERELIERTMAGAAARNKGRIGR RPPKSGTGEEKPYKCPECGKSFSQRAHLERHQRTHTGEKPYKCPECG KSFSTSGNLVRHQRTHTGEKPYKCPECGKSFS ~~SRSDLVVRHQRTHTG~~ EKPYKCPECGKSFS ~~SHKNALONHQRTHTGKKTSGQAGQ~~
ZFR-8 右	SEQ ID NO: 78	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGSRMKHLISLVGELRERGINFRSLTD

[0126]

		SIDTSSPMGRFFFFVVMGALAEMERELIERTMAGIAAARNKGRIG RPPKSGTGEKPYKCPECGKSFSRRDELNVHQRTHTGEKPYKCPECG KSFSQSSNLVRHQRTHTGEKPYKCPECGKSFSQSSSLVRHQRTHTGE KPYKCPECGKSFSSTIGNLTVHQRTHTGKKTSGQAGQ
ZFR-9 左	SEQ ID NO: 79	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGSRMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFFVVMGALAEMERELIERTMAGIAAARNKGRIG GRPPKSGTGEKPYKCPECGKSFSSTIGNLTVHQRTHTGEKPYKCPECG GKSFSQSSNLVRHQRTHTGEKPYKCPECGKSFSQRAILERTHQRTHT GEKPYKCPECGKSFSQKSSLIHQRTHTGKKTSGQAGQ
ZFR-9 右	SEQ ID NO: 80	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGSRMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFFVVMGALAEMERELIERTMAGIAAARNKGRIGR PDKSGTGEKPYKCPECGKSFSDPGALVRHQRTHTGEKPYKCPECGK SFSQSSSLVRHQRTHTGEKPYKCPECGKSFSQLAHLRAHQRTHTGE KPYKCPECGKSFSQORANLRAHQRTHTGKKTSGQAGQ
侧臂区域突变被高亮显示。特异性-决定 α 螺旋锌指残基被加下划线。		

[0127] 虽然实施例说明了具有Gin催化结构域的ZFR的产生,但是该方法也可以适用于若干其它重组酶的催化结构域。这样的重组酶包括:a) Tn3,也称为EcoTn3;Hin,也称为StyHin;MuGin;Sin;Beta;Pin;Min;Din;Cin;EcoTn21;SfaTn917;BmeTn5083;Bme53;Cpe;SauSK1;SauSK41;SauTn552;Ran;Aac;Lla;pMER05;Mlo92;Mlo90;Rrh;Pje;Req;PpsTn5501;Pae;Xan;ISXc5;Spy;RhizY4cG;SarpNL1;Sso1SC1904a;Sso1SC1904b;Sso1SC1913;Aam606;MjaM0014;Pab;Hpy1S607;Mtu1S_Y349;MtuRv2792c;MtuRv2979c;MtuRv3828c;MtuRv0921;MceRv0921;TnpX;TndX;WwK;乳球菌噬菌体TP901-1丝氨酸重组酶;化脓性链球菌噬菌体 ϕ 370.1丝氨酸重组酶;化脓性链球菌噬菌体 ϕ FC1丝氨酸重组酶;李斯特菌属噬菌体A118丝氨酸重组酶;天蓝色链霉菌染色体SC3C8.24丝氨酸重组酶;天蓝色链霉菌染色体SC2E1.37丝氨酸重组酶;天蓝色链霉菌染色体SCD78.04c丝氨酸重组酶;天蓝色链霉菌染色体SC8F4.15c丝氨酸重组酶;天蓝色链霉菌染色体SCD12A.23丝氨酸重组酶;天蓝色链霉菌染色体SCH10.38c丝氨酸重组酶;天蓝色链霉菌染色体SCC88.14丝氨酸重组酶;链霉菌噬菌体 ϕ C31丝氨酸重组酶;链霉菌噬菌体R4丝氨酸重组酶;芽孢杆菌噬菌体 ϕ 105丝氨酸重组酶;芽孢杆菌噬菌体SPBc2丝氨酸重组酶;芽孢杆菌前噬菌体SKIN丝氨酸重组酶;金黄色葡萄球菌ccrA丝氨酸重组酶;金黄色葡萄球菌ccrB丝氨酸重组酶;结核分枝杆菌噬菌体Bxb1丝氨酸重组酶;结核分枝杆菌前噬菌体 ϕ RV1丝氨酸重组酶;YBCK_ECOLI;Y4bA;Bja;Spn;Cac

1956;和Cac 1954;以及b) a)的突变蛋白。

[0128] 特定结构域的不完善模块化、缺乏结合到所有DNA三联体的高亲和力以及难以构建已经阻碍了ZFP在非专业化的实验室中的广泛使用。从黄单胞菌属发现的转录激活子样效应因子 (TALE) 蛋白中的新型DNA结合结构域提供了ZFP的替代。本文描述的是嵌合TALE重组酶 (TALER): 来自DNA转化酶Gin的过度活化的催化结构域与优化的TALE架构之间的工程化融合。递增截短的TALE变体文库被鉴定以鉴定TALER融合体, 其以与细菌细胞中的锌指重组酶相当的效率和特异性修饰DNA。在实施例中还示出, 在哺乳动物细胞中的TALER重组DNA。本文所述的TALER架构提供用于插入定制的TALE结构域的平台, 从而显著扩大工程化的重组酶的靶向能力以及它们在生物技术和医药中的可能的应用。

[0129] 转录激活子样效应因子 (TALE) 蛋白可以被设计成结合几乎任何DNA序列。用于TALE DNA结合结构域的设计的一般准则建议TALE所结合的DNA序列的5'最末端碱基(N₀碱基) 应该是胸腺嘧啶。通过对在该位置的每个DNA碱基分析TALE转录因子 (TALE-TF)、TALE重组酶 (TALE-R) 和TALE核酸酶 (TALEN) 的活性来对N₀要求进行定量分析。观察到: 与含有5'T的靶序列相比, 在没有5'T存在的情况下, TALE活性降低高达>1000倍的TALE-TF活性、降低高达100倍的TALE-R活性以及高达10倍的TALEN活性的减少。为了开发识别所有可能的N₀碱基的TALE架构, 结构引导的文库设计加上TALE-R活性选择被用来进化新型TALE N末端结构域, 以适应任何N₀碱基。分离并表征G选择性结构域和广义反应性结构域。以TALE-R形式选择的工程化的TALE结构域表现出模块化并且在TALE-TF和TALEN架构有活性。进化的N末端结构域对于作为TALE结合蛋白和设计酶的任何DNA序列提供有效的和不受约束的基于TALE的靶定。

[0130] 在一方面, 本发明提供产生特异性结合期望的核苷酸的转录激活子样效应因子 (TALE) 蛋白结合结构域的方法。如实施例中所示出的, 该方法包括: a) 通过使可变二残基 (RVD) 内的氨基酸残基突变, 或者通过使该RVD的N-末端或C-末端的1至2个氨基酸残基内的氨基酸残基突变使TALE蛋白结合结构域的氨基酸序列随机化; 以及b) 选择 (a) 中的随机化的TALE蛋白结合结构域, 其中TALE蛋白结合结构域特异性结合到期望的核苷酸。

[0131] 本文提供了序列特异性核酸酶、重组酶、核酸酶和转录因子。序列特异性多肽包括定制的TAL效应因子DNA结合结构域。正因为如此, 在另一方面, 本发明提供了嵌合多肽。该多肽包括: a) 重组酶、转录因子或核酸酶; 以及b) 转录激活子样效应因子 (TALE) 蛋白。

[0132] TALE是植物病原菌蛋白, 其通过所述病原体注入植物细胞, 其中所述TALE进入细胞核并作为转录因子启动特定植物基因。TALE的一级结构的氨基酸序列决定其结合的核苷酸序列。因此, 可根据TALE预测靶位置, 且如本文所述, TALE还可被工程化并生成以用于结合具体核苷酸序列。

[0133] 融合TALE编码核酸序列的是编码核酸酶、转录因子或重组酶或其部分的序列。许多这样的蛋白是现有技术中已知的, 其可以在本发明中使用。

[0134] 在各种实施方式中, 嵌合多肽包括重组酶的催化结构域。如上所讨论的, 可以利用若干重组酶的催化结构域。这样的重组酶包括: a) Tn3, 也称为EcoTn3; Hin, 也称为StyHin; Gin, 也称为MuGin; Sin; Beta; Pin; Min; Din; Cin; EcoTn21; SfaTn917; BmeTn5083; Bme53; Cpe; SauSK1; SauSK41; SauTn552; Ran; Aac; Lla; pMER05; Mlo92; Mlo90; Rrh; Pje; Req; PpsTn5501; Pae; Xan; ISXc5; Spy; RhizY4cG; SarpNL1; SsolSC1904a; SsolSC1904b;

SsoISC1913;Aam606;MjaM0014;Pab;Hpy1S607;Mtu1S_Y349;MtuRv2792c;MtuRv2979c;MtuRv3828c;MtuRv0921;MceRv0921;TnpX;TndX;WwK;乳球菌噬菌体TP901-1丝氨酸重组酶;化脓性链球菌噬菌体 ϕ 370.1丝氨酸重组酶;化脓性链球菌噬菌体 ϕ FC1丝氨酸重组酶;李斯特菌属噬菌体A118丝氨酸重组酶;天蓝色链霉菌染色体SC3C8.24丝氨酸重组酶;天蓝色链霉菌染色体SC2E1.37丝氨酸重组酶;天蓝色链霉菌染色体SCD78.04c丝氨酸重组酶;天蓝色链霉菌染色体SC8F4.15c丝氨酸重组酶;天蓝色链霉菌染色体SCD12A.23丝氨酸重组酶;天蓝色链霉菌染色体SCH10.38c丝氨酸重组酶;天蓝色链霉菌染色体SCC88.14丝氨酸重组酶;链霉菌噬菌体 ϕ C31丝氨酸重组酶;链霉菌噬菌体R4丝氨酸重组酶;芽孢杆菌噬菌体 ϕ 105丝氨酸重组酶;芽孢杆菌噬菌体SPBc2丝氨酸重组酶;芽孢杆菌前噬菌体SKIN丝氨酸重组酶;金黄色葡萄球菌ccrA丝氨酸重组酶;金黄色葡萄球菌ccrB丝氨酸重组酶;结核分枝杆菌噬菌体Bxb1丝氨酸重组酶;结核分枝杆菌前噬菌体 ϕ RV1丝氨酸重组酶;YBCK_ECOLI;Y4bA;Bja;Spn;Cac 1956;和Cac 1954;以及b) a)的突变蛋白。在优选的实施方式中,利用高活性的Gin催化结构域。这样的结构域可以使用如本文所述的本发明的方法来产生。

[0135] 如本文所述,TALE包括若干不完善的重复序列,该重复序列决定了其与DNA相互作用的特异性。每个重复序列结合单个碱基,取决于该重复序列的残基12和残基13处的具体双氨基酸序列。因此,通过工程改造TALE内的重复序列可以靶向特定的DNA位点。这种工程改造的TALE可用作例如靶向特定DNA序列的转录因子。

[0136] 如在实施例中所说明的,以如列于表3中的变体及其部分(例如,RVD和NTD)来例举本发明的嵌合蛋白。

[0137] 表3

[0138]

变体	SEQ ID NO.	序列
TALEN (Goldy) NT-TT1 蛋白 序列	81	MRSPKKKRKVQVDLRTLGYSSQQQEKIKPKVRSSTVAQHII EALVGHGFTAHIVALSQHPPAALGTVAVTYQHITATPEAT HEDIVGVGKQWSGARALEALLTDAGELRGPPPLQDITGQLV KIAKRGGVTTAMEAVHASRNALTGAPLNLTDPQVVAIASNG GGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQA LETVQRLLPVLCQDHGLTPDQVVAIASNIGGGKQALETVQR LLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLC QDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLT PDQVVAIASNIGGGKQALETVQRLLPVLCQDHGLTPDQVVAI

[0139]

		ASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGG KQALETVQRLLPVLCQDHGLTPDQVVAIVSHDGGKQALET VQRLLPVLCQDHGLTPDQVVAIVSHDGGKQALETVQRLLP VLCQDHGLTPDQVVAIVSNGGGKQALETVQRLLPVLCQD HGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPD QVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIA SNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGK QALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALES VAQLSRPDPALAAALTNHDLVALACLGGRPAMDVAVKEGLP HAPELIRRVNRRIGERTSHRVAGSQLVKSELEKKSELRHK LKYPHEIYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGK HLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLFIGQA DEMORYVEENQTRNKHINPNEWKVYPSSVTEFKFLFVSG HFKGNYKAQLTRLNHIITNCNGAVLSVEELLIGGEMIKAGT LTLEEVRRKFNNGEINF ----- N-末端结构域 (NTD) 变化如下所示
TALEN RVD 序列		
G1	82	NG-NN-HD-NG-HD-NI-NG-NG-NI-HD-NI-HD-HD-NG-NN-HD- NI 靶向 (TCTTCATTACACCTGCA; SEQ ID NO: 280)
G2	83	HD-NI-NN-NG-HD-NI-NN-NG-NI-NG-HD-NI-NI-NG-NG 靶向 (CAGTCAGTATCAATT; SEQ ID NO: 281)
A1	84	IID-HD-NG-NN-IID-NI-NN-IID-NG-HD-NG-HD-NI-NG-NG-NG- NG 靶向 (CCTGCAGCTCTCATTTT; SEQ ID NO: 282)
A2	85	NI-NG-NG-HD-NG-NG-HD-HD-NI-NN-NI-NG-NG-NN-NI 靶向 (ATTCTCCAGAATTGA; SEQ ID NO: 283)
C2	86	IID-NI-NN-NI-NI-NG-NG-NN-NI-NG-NI-HD-NG-NN-NI-HD-NG 靶向 (CAGAATTGATACTGACT; SEQ ID NO: 284)
T1	87	NG-IID-NI-NG-NG-NI-HD-NI-HD-HD-NG-NN-IID-NI-NN-IID 靶向 (TCATTACACCTGCAGC; SEQ ID NO: 285)

[0140]

T2	88	HD-NG-NG-HD-HD-NI-NN-NI-NI-NG-NG-NN-NI-NG-NI-HD-NG- NN 靶向 (CTTCCAGAATTGATACTG; SEQ ID NO: 286)
N-末端结构域		
NTD = dHax3- TALEN DNA 序列	89	ATGAGATCTCCTAAGAAAAAGAGGAAGATGGTGGACTTGA GGACACTCGGTTATTTCGCAACAGCAACAGGAGAAAAATCAA GCCTAAGGTCAGGAGCACCGTCCGCAACACCACGAGGCG CTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTT TCACAGCACCCTGCGGCGCTTGGGACGCTGGCTGTCAAATA CCAAGATATGATTGCGGCCCTGCCCCAAGCCACGCACGAGG CAATTGTAGGGGTGCGTAAACAGTGGTCGGGAGCGCGAGC ACTTGAGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGC CTCCGCTCCAGCTCGACACCGGGCAGCTGCTGAAGATCGCG AAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCATGCAT CGCGCAATGCACTGACGGGTGCCCCC
NTD = dHax3- TALEN	90	MRSPPKKRKMVDLRTLGYSSQQQEKIKPKVRSSTVAQHIHEAL VGHGFTHAHIVALSQHPAALGTAVVKYQDMIAALPEATHEAL VGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLKIAKRG GVTAVEAVHASRNALTGAP... 重复可变二残基
NTD = NT-βN TALEN DNA 序列	91	ATGAGATCTCCTAAGAAAAAGAGGAAGGTGCAGGTGGATC TACGCACGCTCGGCTACAGTCAGCAGCAGCAAGAGAAGAT CAAACCGAAGGTGCGTTCGACAGTGGCGCAGCACCACGAG GCACTGGTGGGCCATGGGTTTACACACGCGCACATCGTTGC GCTCAGCCAACACCCGCGAGCGTTAGGGACCGTCGCTGTCA CGTATCAGCACATAATCAGGCGTTGCCAGAGGCGACACAC GAAGACATCGTTGGCGTCGGCAAATATCATGGGOCACGCGC TCTGGAGGCCCTTGCTCACGGATGCGGGGGAGTTGAGAGGTC CGCCGTTACAGTTGGACACAGGCCAACTGTGAAOATTGCA AAACGTGGCGGCGTGACCGCAATGGAGGCAGTGCATGCAT CGCGCAATGCACTGACGGGTGCCCCC
NTD = NT-βN TALEN	92	MRSPPKKRKVQVDLRTLGYSSQQQEKIKPKVRSSTVAQHIHEAL VGHGFTHAHIVALSQHPAALGTAVVTYQHITALPEATHEDIV GVGKYHIGARALEALLTDAGELRGPPLQLDTGQLVKIAKRGV TAMEAVHASRNALTGAP... 重复可变二残基

[0141]

NTD NT-G TALEN DNA 序列	93	ATGAGATCTCCTAAGAAAAAGAGGAAGGTGCAGGTGGATC TACGCACGCTCGGCTACAGTCAGCAGCAGCAAGAGAAGAT CAAACCGAAGGTGCGTTTCGACAGTGGCGCAGCACACGAG GCACTGGTGGGCCATGGGTTTACACACGCGCACATCGTTGC GCTCAGCCAACACCCGGCAGCGTTAAGGACCGTCGCTGTCA CGTATCAGCACATAATCACGGCGTTGCCAGAGGCGACACAC GAAGACATCGTTGGCGTCGGCAAATCGCGGTGGGGGCGAC GCGCTCTGGAGGCCTTGCTCACGGATGCGGGGGAGTTGAGA GGTCCGCCGTTACAGTTGGACACAGGCCAACTTGTGAAGAT TGCAAAACGTGGCGGCGTGACCGCAATGGAGGCAGTGCAT GCATCGCGCAATGCACTGACGGGTGCCCCC
NTD NT-G TALEN 蛋白 序列	94	MRSPKKKRRKVQVDLRTLGYSSQQQKEKIKPKVRSTVAQIHHEAL VGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDI GVGKSRSGARALEALLTDAGELRGPPLQLDTGQLVKIAKRG VTAMEAVHASRNALTGAP... 重复可变二残基
NTD = NT-αN TALEN DNA 序列	95	ATGAGATCTCCTAAGAAAAAGAGGAAGGTGCAGGTGGATC TACGCACGCTCGGCTACAGTCAGCAGCAGCAAGAGAAGAT CAAACCGAAGGTGCGTTTCGACAGTGGCGCAGCACACGGG GCACTGGTGGGCCATGGGTTTACACACGCGCACATCGTTGC GCTCAGCCAACACCCGGCAGCGTTAAGGACCGTCGCTGTCA CGTATCAGCACATAATCACGGCGTTGCCAGAGGCGACACAC GAAGACATCGTTGGCGTCGGCAAACGGGGGGCTGGTGCAC GCGCTCTGGAGGCCTTGCTCACGGATGCGGGGGAGTTGAGA GGTCCGCCGTTACAGTTGGACACAGGCCAACTTGTGAAGAT TGCAAAACGTGGCGGCGTGACCGCAATGGAGGCAGTGCAT GCATCGCGCAATGCACTGACGGGTGCCCCC
NTD = NT-αN TALEN 蛋白	96	MRSPKKKRRKVQVDLRTLGYSSQQQKEKIKPKVRSTVAQHIIIGA LVGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDI VGVGKRGAGARALEALLTDAGELRGPPLQLDTGQLVKIAKRG GVTAMEAVHASRNALTGAP... 重复可变二残基
NTD = NT-T I-I TALEN DNA	97	ATGAGATCTCCTAAGAAAAAGAGGAAGGTGCAGGTGGATC TACGCACGCTCGGCTACAGTCAGCAGCAGCAAGAGAAGAT CAAACCGAAGGTGCGTTTCGACAGTGGCGCAGCACACGAG

[0142]

		GC ACTGGTGGGCCATGGGTTTACACACGCGCACATCGTTGC GCTCAGCCAACACCCGGCAGCGTTAGGGACCGTCGCTGTCA CGTATCAGCACATAATCACGGCGTTGCCAGAGGCGACACAC GAAGACATCGTTGGCGTCGGCAAACAGTGGTCCGGCGCAC GCGCCCTGGAGGCCCTTGCTCACGGATGCGGGGGAGTTGAGA GGTCCGCCGTTACAGTTGGACACAGGCCAACTTGTGAAGAT TGCAAAACGTGGCGGCGTGACCGCAATGGAGGCAGTGCAT GCATCGCGCAATGCACTGACGGGTGCCCCC
MIBP-TALE 蛋白序列	98	MKIEEGKLVIWINGDKGYNGLAIEVGGKFEKDTGKVTVEIIPD KLEEKFPQVAATGDGPDHFWAHDREFGGYAQSGLLAETPDKA FQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLPNPPK TWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAF KYENGKYDIKDVGVNAGAKAGLTFVLDLIKHKHMNADTDY SIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKG QPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNK DKPLGAVALKSYEEELAKDPRIAATMENAKGGEIMPNIQMS AFWYAVRTAVINAASGRQTVDEALKDAQTNSSSNNNNNNNN NNLGIEGRISEFGSPARPPRAKPAPRRRSAQPSDASPAAQVDLR TLGYSQQQKEKIKPKVRSVAQHIEALVGHGFTHAHIVALSQ HPAALGTVAVTYQHITALPEATHEDIVGVGK[XXX]GARALE ALLTDAGELLRGPPQLDTGQLVKIAKRGVGTAMEAVIIASRN ALTGAPLNLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGL TPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAI ASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQ LETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLL VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGL TPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAI ASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIVSHDGGKQ ALETVQRLLPVLCQDHGLTPDQVVAIVSHDGGKQALETVQRLL LPVLCQDHGLTPDQVVAIVSNGGKQALETVQRLLPVLCQDH GLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQV VAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGG KQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALESIVAQ LSRPDPALAAALTNDHILVALACLG XXX: NT-T ≈ QWS NT-G ≈ SRS; NT-αN ≈ RGA; NT-βN ≈ Y-H

[0143]

TALE-R 蛋白 序列	99	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDR PGLKRALKRLQKGD TLVVKLDRLGRSMKHLISLVGELRERG INFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIERTMAGLA AARNKGRIGGRPPKSGSPRPRAKPAPERRAAQPSDASPAQV DLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVA LSQHPAALGTVAVTYQHITALPEATHEDIVGVGK[XXX]GARA LEALLTDAGELRGPPQLDTGQLVKIAKRGGVTAMEAVHASR NALTGAPLNLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVA IASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQ ALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLL PVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHG LTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVA IASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIVSHDGGK QALETVQRLLPVLCQDHGLTPDQVVAIVSHDGGKQALETVQR LLPVLCQDHGLTPDQVVAIVSNGGGKQALETVQRLLPVLCQD HGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQV VAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGG KQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALESIVA QLSRPDPALAAALTNDHLVALACLG XXX: NT-T = QWS NT-G = SRS; NT-αN = RGA; NT-βN = Y-H
Avr15 TALE-TF 蛋白序列	100	MAQAASGSPRPRAKPAPRRRAAQPSDASPAQVDLRTLGYSS QQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAAL GTVAVTYQHITALPEATHEDIVGVGK[XXX]GARALEALLTDA GELRGPPQLDTGQLVKIAKRGGVTAMEAVHASRNALTGAPL NLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVV AIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGK QALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLL PVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDH GLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVV AIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGG KQALETVQRLLPVLCQDHGLTPDQVVAIVSHDGGKQALETVQ RLLPVLCQDHGLTPDQVVAIVSHDGGKQALETVQRLLPVLCQ DHGLTPDQVVAIVSNGGGKQALETVQRLLPVLCQDHGLTPDQ VVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHD GGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALET

[0144]

		VQRLLPVLCQDHGLTPDQVVAIASNIGGKQALESIVAQLSRPD PALAALTNDHILVALACLGGRPAMDVKKGLPHAPELIRRVNR RIGERTSHRVADYAQVVRVLEFFQCHSHIPAYAFDEAMTQFGM SGQAGQASPKKKRKVGRADALDDFDLDMLGSDALDDFDLDM LGSDALDDFDLDMLGSDALDDFDLDMLINYPYDVDPDYAS XXX; NT-T = QWS NT-G = SRS; NT-αN = RGA; NT-βN = Y-H
Avr15 RVD 序列 (用于 TALE-R, TALE- TF, MBP-TALE)	101	NI-NG-NI-NI-NI-HD-HD-HD-HD-NI-HD-HD-NI-NI 靶向 (ATAAACCCCTCCAA; SEQ ID NO: 287)

[0145] 在各种实施方式中,嵌合蛋白包括具有C末端或N末端截短的TALE蛋白。例如,TALE蛋白可以包括全部SEQ ID NO:2或者SEQ ID NO:2的一部分。在实施方式中,TALE蛋白在氨基酸残基27与氨基酸残基268、氨基酸残基92与氨基酸残基134、氨基酸残基120与氨基酸残基129、氨基酸残基74与氨基酸残基147、或氨基酸残基87与氨基酸残基120之间是截短的,例如在氨基酸残基28、氨基酸残基74、氨基酸残基87、氨基酸残基92、氨基酸残基95、氨基酸残基120、氨基酸残基124、氨基酸残基128、氨基酸残基129、氨基酸残基147和氨基酸残基150处。

[0146] 在另一个实施方式中,提供了分离的多肽,其包含转录激活因子样效应因子(TALE)蛋白,其中TALE蛋白具有包括如SEQ ID NO:3 (VGKQWSGARAL)中所列的氨基酸序列的N-末端结构域(NTD),该氨基酸序列具有选自以下的一个或多个突变或缺失:Q是Y、Q是S、Q是R、W是R、W是G、W缺失、S是R、S是H、S是A、S是N、以及S是T。

[0147] 在一些实施方式中,NTD包括选自以下的氨基酸序列:VGKYRGARAL (SEQ ID NO:4)、VGKSRSGARAL (SEQ ID NO:5)、VGKYHGARAL (SEQ ID NO:6)和VGKRGAGARAL (SEQ ID NO:7)。

[0148] 在另一个实施方式中,提供了分离的多肽,其包含转录激活因子样效应因子(TALE)蛋白,其中TALE蛋白具有包含如SEQ ID NO:8 (IVDIAR₁QR₂SGDLA)中所列的氨基酸序列的N-末端结构域(NTD),该氨基酸序列具有选自以下的一个或多个突变或缺失:R₁是K、Q是Y、Q是S、Q是R、R₂是W、R₂是G、R₂缺失、S是R、S是H、S是A、S是N、以及S是T。

[0149] 在一些实施方式中,NTD包括选自以下的氨基酸序列:IVDIARQWSGDLA (SEQ ID NO:9)、IVDIARYRGDLA (SEQ ID NO:10)、IVDIARSRSGLA (SEQ ID NO:11)、IVDIARYHGDLA (SEQ ID NO:12)和IVDIARRGAGDLA (SEQ ID NO:13)。

[0150] 在另一个实施方式中,TALE蛋白包括具有阐述如下的氨基酸序列的修饰的N₀结构域:LTPDQLVKIAKRGGTAMEAVHASRNALTGAPLN (SEQ ID NO:102)。在各种实施方式中,TALE蛋白包括经突变的变体,其中SEQ ID NO:102的KRGG (SEQ ID NO:103)是选自LDYE (SEQ ID NO:104)、INLV (SEQ ID NO:105)、YSKK (SEQ ID NO:106)、NMAH (SEQ ID NO:107)、SPTN (SEQ ID

NO:108)、SNTR(SEQ ID NO:109)、LTTT(SEQ ID NO:110)、VADL(SEQ ID NO:111)、MVLS(SEQ ID NO:112)、YNGR(SEQ ID NO:113)、RIPR(SEQ ID NO:114)、YSKI(SEQ ID NO:115)、LTQY(SEQ ID NO:116)、YLSK(SEQ ID NO:117)、LRPN(SEQ ID NO:118)、LFTN(SEQ ID NO:119)、LLTN(SEQ ID NO:120)、EEDK(SEQ ID NO:121)、VTAM(SEQ ID NO:122)、CPSR(SEQ ID NO:123)、LTRV(SEQ ID NO:124)、KGDL(SEQ ID NO:125)、QKAL(SEQ ID NO:126)、LYLL(SEQ ID NO:127)、WISV(SEQ ID NO:128)、GDQV(SEQ ID NO:129)和CPSR(SEQ ID NO:130)。

[0151] 在另一个实施方式中,TALE蛋白包括具有阐述如下的氨基酸序列的修饰的N₁结构域:

[0152]

MRSPKKKRKVQVDLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTAVTYQHIIITALPEATHEDIVGVGXXXXXARALEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNALTGAP(SEQ ID NO:131)。在各种实施方式中,SEQ ID NO:131的XXXXX是KRPAG(SEQ ID NO:132)或KRPSG(SEQ ID NO:133)。此外,该蛋白可以包括表现出增强的活性的E40G突变(参照SEQ ID NO:131)。

[0153] 在另一个实施方式中,TALE蛋白包括具有阐述如下的氨基酸序列的重复序列结构域:

[0154] LTPDVVAISNNGGKQALETVQRLLPVLCQDGH(SEQ ID NO:134)。在各种实施方式中,TALE蛋白包括经突变的变体,其中SEQ ID NO:134的SNNG(SEQ ID NO:135)是选自RGGG(SEQ ID NO:136)、RGGR(SEQ ID NO:137)、RGVR(SEQ ID NO:138)、KGGG(SEQ ID NO:139)、SGGG(SEQ ID NO:140)、GGRG(SEQ ID NO:141)、LGGS(SEQ ID NO:142)、MDNI(SEQ ID NO:143)、RVMA(SEQ ID NO:144)、LASV(SEQ ID NO:145)、VGTG(SEQ ID NO:146)和QGGG(SEQ ID NO:147)。

[0155] 提供以下实施例以进一步说明本发明的优点和特征,但并不旨在限制本发明的范围。虽然它们是可能使用的方法中典型的例子,但是也可使用本领域技术人员已知的其它过程、方法或技术。

[0156] 实施例1

[0157] 嵌合TALE重组酶

[0158] 实验总结

[0159] 该研究提供了第一个TALE重组酶(TALER)的实施例。使用递增截短的TALE结构域文库,鉴定优化的TALER架构,其可用于使DNA重组入细菌和哺乳动物细胞中。任何定制的TALE重复阵列可以被插入本文所述的TALER架构,从而显著扩大设计的重组酶在生物技术和医药应用中的靶向能力。

[0160] 在本实施例中使用了以下的材料和方法。

[0161] 试剂

[0162] 除非另有说明,否则所有的酶均购自New England Biolabs。引物序列提供在表4中。

[0163] 表4. 引物

[0164]

本研究中所用的引物		
用于 pBLA 底物构建的引物		
AvrXa7 lac 靶 F	SEQ ID NO: 148	TTAATTAAGAGTCTAGAAATATAAACCCCTCC AACCAGGTGCTAACTGTAAACCATGGTTTTGGA TTAGCACCTGGTTGGAGGGGGTTTATAAGATCT AGGAGGAATTTAAAATGAG
AvrXa7 lac 靶 R	SEQ ID NO: 149	ACTGACCTAGAGAAGCTTATATAAACCCCTCC AACCAGGTGCTAATCCAAAACCATGGTTTAC AGTTAGCACCTGGTTGGAGGGGGTTTATACTG CAGTTATTTGTACAGTTCATC
AvrXa7 N F	SEQ ID NO: 150	TTAATTAAGAGTCTAGATTAGCACCTGGTTGG AGGGGGTTTATAAGGTTTTGGTACCAAATGTC TATAAACCCCTCCAACCAGGTGCTAAAGATC TAGGAGGAATTTAAAATGAG
AvrXa7 N R	SEQ ID NO: 152	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTG GAGGGGGTTTATAGACATTTGGTACCAAACC

[0165]

		TTATAAACCCCTCCAACCAGGTGCTAACTGC AGTTATTTGTACAGTTCATC
AvrXa7 N RC F	SEQ ID NO: 153	TTAATTAAGAGTCTAGATTAGCACCTGGTTGG AGGGGGGTTTATATCCAAAACCATGGTTTACAG TATAAACCCCTCCAACCAGGTGCTAAAGATC TAGGAGGAATTTAAAATGAG
AvrXa7 N RC R	SEQ ID NO: 154	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTG GAGGGGGTTTATATCCAAAACCATGGTTTACA GTATAAACCCCTCCAACCAGGTGCTAACTGC AGTTATTTGTACAGTTCATC
AvrXa7 N RC +3 F	SEQ ID NO: 155	TTAATTAAGAGTCTAGATTAGCACCTGGTTGG AGGGGGGTTTATAGCTTCCAAAACCATGGTTTA CAGGGTTATAAACCCCTCCAACCAGGTGCTA AAGATCTAGGAGGAATTTAAAATGAG
AvrXa7 N RC +3 R	SEQ ID NO: 277	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTG GAGGGGGTTTATAACCTGTAAACCATGGTTT TGGAAGCTATAAACCCCTCCAACCAGGTGCT AACTGCAGTTATTTGTACAGTTCATC
AvrXa7 N RC +6 F	SEQ ID NO: 156	TTAATTAAGAGTCTAGATTAGCACCTGGTTGG AGGGGGGTTTATAGCTTCATCCAAAACCATGGT TTACAGGGTTTCTATAAACCCCTCCAACCAG GTGCTAAAGATCTAGGAGGAATTTAAAATGAG
AvrXa7 N RC +6 R	SEQ ID NO: 157	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTG GAGGGGGTTTATAGCAACCTGTAAACCATGG TTTGGATGAAGCTATAAACCCCTCCAACCA GGTGCTAACTGCAGTTATTTGTACAGTTCATC
AvrXa7 N RC +12 F	SEQ ID NO: 158	TTAATTAAGAGTCTAGATTAGCACCTGGTTGG AGGGGGGTTTATAGCTTCAGCTTCATCCAAAAC CATGGTTTACAGGGTTCGGTTCCTATAAACCC CCCTCCAACCAGGTGCTAAAGATCTAGGAGGA ATTTAAAATGAG
AvrXa7 N RC +12 R	SEQ ID NO: 278	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTG GAGGGGGTTTATAGCAACCGCAACCTGTAAA CCATGGTTTTGGATGAAGCTGAAGCTATAAAC CCCTCCAACCAGGTGCTAACTGCAGTTATTT GTACAGTTCATC
AvrXa7 N RC -3 F	SEQ ID NO: 160	TTAATTAAGAGTCTAGATTAGCACCTGGTTGG

[0166]

		AGGGGGGTTTATAAAAAACCATGGTTTATATAAA CCCCCTCCAACCAGGTGCTAAAGATCTAGGAG GAATTIAAAATGAG
AvrXa7 N RC -3 R	SEQ ID NO: 161	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTG GAGGGGGTTTATATAAACCATGGTTTATATAA ACCCCTCCAACCAGGTGCTAACTGCAGTTAT TTGTACAGTTCATC
AvrXa7 N RC GG F	SEQ ID NO: 162	TTAATTAAGAGTCTAGATTAGCACCTGGTTGG AGGGGGGTTTATATCCAAAACCGGGGTTTACA GTATAAACCCCTCCAACCAGGTGCTAAAGA TCTAGGAGGAATTTAAATGAG
AvrXa7 N RC GG R	SEQ ID NO: 163	ACTGACCTAGAGAAGCTTTTAGCACCTGGTT GGAGGGGGTTTATACTGTAAACCCCGGTTTT GGATATAAACCCCTCCAACCAGGTGCTAAC TGCAGTTATTTGTACAGTTCATC
AvrXa7 N 20t F	SEQ ID NO: 164	TTAATTAAGAGTCTAGATTAGCACCTGGTTG GAGGGGGTTTATACGAAATATTATAAATTA TCATATAAACCCCTCCAACCAGGTGCTAA AGATCTAGGAGGAATTTAAATGAG
AvrXa7 N RC 20t R	SEQ ID NO: 165	ACTGACCTAGAGAAGCTTTTAGCACCTGGTT GGAGGGGGTTTATATGATAATTTATAATATT TCGTATAAACCCCTCCAACCAGGTGCTAAC TGCAGTTATTTGTACAGTTCATC
AvrXa7 32 GG F	SEQ ID NO: 166	TTAATTAAGAGTCTAGATTAGCACCTGGTTG GAGGGGGTTTATAGCTTCATCCAAAACCGG GGTTTACAGGGTTCCTATAAACCCCTCCAA CCAGGTGCTAAAGATCTAGGAGGAATTTAA AATGAG
AvrXa7 32 GG R	SEQ ID NO: 167	ACTGACCTAGAGAAGCTTTTAGCACCTGGTT GGAGGGGGTTTATAGCAACCCTGTAAACCGG GGTTTTGGATGAAGCTATAAACCCCTCCAA CCAGGTGCTAACTGCAGTTATTTGTACAGTTC ATC
AvrXa7 32t F	SEQ ID NO: 168	TTAATTAAGAGTCTAGATTAGCACCTGGTTG GAGGGGGTTTATAGCTTCACGAAATATTATA AATTATCAGGTTCCTATAAACCCCTCCAAC CAGGTGCTAAAGATCTAGGAGGAATTTAAA

[0167]

		ATGAG
AvrXa7 321 R	SEQ ID NO: 169	ACTGACCTAGAGAAGCTTTTAGCACCTGGTT GGAGGGGGTTTATAGCAACCTGATAATTTAT AATATTTCTGTGAAGCTATAAACCCCTCCAA CCAGGTGCTAACTGCAGTTATTTGTACAGTT CATC
用于 pGL3Pro 靶位点构建的引物		
5' pGL3 SV40 Avr.32G BglII	SEQ ID NO: 170	TTAATTAAGAGAGATCTTTAGCACCTGGTTG GAGGGGGTTTATAGCTTCATCCAAAACCATG GTTTACAGGGTTCCCTATAAACCCCTCCAAC CAGGTGCTAAGCGATCTGCATCTCAATTAGT CAGC
3' pGL3 SV40 Avr.20G HindIII	SEQ ID NO: 171	ACT GAC CTA GAG AAG CTT TTA GCA CCT GGT TGG AGG GGG TTT ATAGCAACC CTG TAA ACC ATG GTT TTG GATGAAGCT ATA AAC CCC CTC CAA CCA GGT GCT AAT TTG CAA AAG CCT AGG CCT CCA AA
5' pGL3 SV40 PH4.20G6Avr BglII	SEQ ID NO: 172	TTAATTAAGAGAGATCTGCCGGGAGGCGTGTC CAAAACCATGGTTTACAGGGTTCCCTATAAAC CCCTCCAACCAGGTGCTAAGCGATCTGCAT CTCAATTAGTCAGC
3' pGL3 SV40 PH4.20G6Avr HindIII	SEQ ID NO: 173	ACT GAC CTA GAG AAG CTT TTA GCA CCT GGT TGG AGG GGG TTT ATAGCAACCCGTGA AACCATGGTTTTGGACACGCCTCCCGCTTTG CAAAAGCCTAGGCCTCCAAA
5' pGL3 SV40 Avr.44G BglII	SEQ ID NO: 174	TTAATTAAGAGAGATCTTTAGCACCTGGTTG GAGGGGGTTTATAGCTTCAGCTTCATCCAAA ACCATGGTTTACAGGOTTCGGTTCCCTATAA ACCCCTCCAACCAGGTGCTAAGCGATCTGC ATCTCAATTAGTCAGC
3' pGL3 SV40 Avr.44G HindIII	SEQ ID NO: 175	ACT GAC CTA GAG AAG CTT TTA GCA CCT GGT TGG AGG GGG TTTATAGCAACCGCAA CCCTG TAA ACC ATG GTT TTG GATGAAGC TGAAGCT ATA AAC CCC CTC CAA CCA GGT GCT AAT TTG CAA AAG CCT AGG CCT CCA AA
用于 BamHI 融合的引物		

Gin_N-端 F	SEQ ID NO: 176	AGTCAGTCGAGAGCTCATGGATCCCGGCTCTA TGCTGATTGGCTATGTAAGG
Gin_N-端 R	SEQ ID NO: 177	ATGCTGATACTAGACTATCCCGATTAGGTGG GCGACC
Gin_C-端 F	SEQ ID NO: 178	AGTCAGTCGAGAGCTCATGCTGATTGGCTATGT AAGG
Gin_C-端 R	SEQ ID NO: 179	TCTAGACTACGGATCCACCGATTACGCGGGC
用于设计的截短的引物		
TalR+28 Xba	SEQ ID NO: 180	ATCGCGTATCTAGACTAGCCGAGGCAGGCCAA GGCGACG
TalR+95 Xba AvrX	SEQ ID NO: 181	ATCGCGTATCTAGACTAGCTCATCTCGAACTGC GTCATG
avr n 1	SEQ ID NO: 182	GTCGCCC CGGTAAATCGGGATCCACTGCAGAT CGGGGGGGGGC
avr n 2	SEQ ID NO: 183	GTCGCCC CGGTAAATCGGGATCCCCCTCGCCTG CGTTCTCGGC
avr n 3	SEQ ID NO: 184	GTCGCCC CGGTAAATCGGGATCCGATTCGATGC CTGCCGTCGG
avr n 4	SEQ ID NO: 185	GTCGCCC CGGTAAATCGGGATCCACCGTGCGT GTCGCTGTCACTG
avr n 5	SEQ ID NO: 186	GTCGCCC CGGTAAATCGGGATCCGTGGATCTAC GCACGCTCGGC
avr n 6	SEQ ID NO: 187	GTCGCCC CGGTAAATCGGGATCCACACACGCG CACATCGTTGC
avr n 7	SEQ ID NO: 188	GTCGCCC CGGTAAATCGGGATCCCACGAAGAC ATCGTTGGCGTCG
avr n 8	SEQ ID NO: 189	GTCGCCC CGGTAAATCGGGATCCAGCGCCTGG AGGCCTTGCTC
avr n 9	SEQ ID NO: 190	GTCGCCC CGGTAAATCGGGATCCTTGGACACA GGCCAACTTCTC
avr n 10	SEQ ID NO: 191	GTCGCCC CGGTAAATCGGGATCCAGCGGCGTG ACCGCAgTGGa
GinNTALPCRfusR	SEQ ID NO: 192	CGATCCCGATTACGCGGGC
用于 pcDNA 克隆的引物		
Nhe-SD-Gin F	SEQ ID NO: 193	ATCGTAGCAGCTAGCGCCACCATGCTGATTGGC TATGTAAG
GinGS R	SEQ ID NO: 194	CGATCCAGACCCCGATTACGCGGGC

[0168]

[0169] 质粒构建

[0170] 为了将BamHI限制位点引入到Gin编码序列的5'或者3'端,分别用引物5'Gin_N端

和3'Gin_N端或者5'Gin_C端和3'Gin_C端将Gin催化结构域进行PCR扩增。将PCR产物连接至pBluescriptII (Fermentas)的SacI和XbaI限制位点中以产生pB-Bam-Gin和pB-Gin-Bam。为了产生C末端和N末端TALER融合,用BamHI将AvrXa7基因从pWAvrXa7释放(由B. Yang博士友情提供, Iowa State University)并且分别连接到pB-Bam-Gin和pB-Gin-Bam (41)的BamHI位点中以建立pB-Avr-Bam-Gin和pB-Gin-Bam-Avr。通过序列分析(图6-16)验证每个TALER的正确构建。

[0171] 为了产生AvrXa7的N-末端截短,利用扩展的高保真PCR系统(Expand High Fidelity PCR System) (Roche)使用5'Avr-n- (1-10)和3'Avr+28或者3'Avr+95引物按照以下程序对AvrXa7进行PCR扩增:94℃3分钟的条件下进行1个循环、94℃1分钟、52℃1分钟、68℃6分钟的条件下进行16个循环、以及68℃1小时的条件下进行最后一个循环。在标准PCR条件下用5'Gin_C端和3'GinNTalPCRfus对Gin催化结构域进行PCR扩增,并且通过利用上述的PCR条件进行重叠PCR而将其融合到截短的AvrXa7变体。以等摩尔比值对纯化的Gin-Avr PCR产物进行混合并用SacI和XbaI对其进行消化。

[0172] 为了产生设计TALE,我们使用带有以下修饰的TALEN试剂盒(Addgene):pTAL1修饰为在Δ120、Δ128、或+28处包含截短。为了实现这一目标,用5'Avr n4或Avr n128和3'TalR Xba+28对AvrXa7 Δ120和AvrXa7 Δ128片段进行PCR扩增,并且将其连接到pTAL1的BamHI限制位点,以产生pTAL Δ120和pTAL Δ128。质粒pTAL Δ120和pTAL Δ128为金门克隆(Golden Gate cloning)保留Esp3I限制性位点。用BamHI和XbaI对克隆到pTAL Δ120和pTAL Δ128的TALE阵列进行消化以连接到pB-Gin-Bam。

[0173] 为了产生哺乳动物TALER表达载体,用5'Nhe-SD-Gin F和3'GinGS R从pB-Gin-Avr中对Gin催化结构域进行PCR扩增并且将其连接到pcDNA 3.1 (Invitrogen)的NheI和BamHI限制性位点。用BamHI和XbaI从pTAL Δ120或pTAL Δ128中对Avr15进行消化,并且将其连接至pcDNA-Gin-Bam以产生pcDNA-Gin-Avr表达载体。

[0174] 如前所述构建pBLA底物质粒。

[0175] 为了产生pGL3报告质粒,用含有引物5'pGL3 SV40 BglIII和3'pGL3 SV40 HindIII的重组位点从pGL3-启动子(Promega)中对SV40启动子进行PCR扩增,并且将其连接入pGL3-启动子的BglIII和HindIII限制性位点。

[0176] 细菌重组分析。

[0177] 如前所述进行细菌重组分析。

[0178] 递增截短文库。

[0179] 利用先前所述的修饰方案产生递增截短文库。简要地说,为了保护Gin编码序列不被外切核酸酶消化,将具有SmaI限制性位点的填充片段插入BamHI以产生pB-Gin-SmaI-Bam-Avr。用NheI将该质粒线性化并与外切核酸酶III在37℃下孵育2.5分钟,随后在75℃下加热灭活25分钟。然后用具有200μM dNTP和5μM [α]-S-dNTP的Klenow片段(3'至5'外切)与pB-Gin-Bam-Avr在37℃下孵育30分钟,随后在80℃下加热灭活25分钟。为了产生截短文库,将pB-Gin-Bam-Avr与外切核酸酶III在37℃下孵育2.5分钟,随后加热灭活以及随后用绿豆核酸酶在30℃下平端1小时。用SmaI消化后,将重组酶编码序列的3'平端连接到TALE片段的平端文库。转化并纯化后,用SacI和XbaI对该质粒进行消化以释放Gin-ΔAvr。

[0180] 哺乳动物报告分析。

[0181] 以每孔 4×10^4 个细胞的密度将HEK293T细胞接种到96孔板上,并使其在湿润的5% CO₂气氛中于37℃下生长。在接种后24小时,根据制造商的说明利用Lipofectamine 2000 (Invitrogen)用150ng pcDNA TALER表达载体、2.5ng pGL3报告质粒以及1ng表达海肾荧光素酶的pRL-CMV对细胞进行转染。在转染后48小时,用被动裂解缓冲液(Promega)使细胞裂解并根据制造商的说明利用双荧光素酶报告基因检测系统(Promega)测定荧光素酶的表达。利用Veritas微孔板检测仪(Turner Biosystems)测量发光。

[0182] 结果。

[0183] TALER架构。

[0184] 已经描述了用于评估和定向进化重组酶活性的定量系统。在该系统(图1A)中,侧接重组位点的GFPuv转基因被插入到编码TEM-1β内酰胺酶的基因。这种改变破坏β内酰胺酶的表达并使包含此质粒(pBLA)的大肠杆菌细胞对氨苄青霉素敏感。但是,活性重组酶从含质粒的底物中的表达导致靶位点与复原的β内酰胺酶阅读框之间的重组。这种修饰使宿主细胞建立对氨苄青霉素的耐药性并且能够实现活性重组酶变体从含质粒的底物中的分离。通过测量质粒纯化和重新转化后的氨苄青霉素-耐药性转化体数量,也可以直接评估重组酶活性。因为嵌合重组酶的活性取决于催化结构域和DBD两者,所以这种分裂基因重装配选择系统也可以用于评估个体DBD的有效性。因此,该系统适合于确定最佳的TALER架构。

[0185] 重要的是,因为DNA转化酶Gin及相关的丝氨酸重组酶的催化结构域具有预先定义的催化特异性,所以TALER融合蛋白不能利用为TALEN描述的设计而被构建。对于γδ解离酶和设计的酶的结构和功能的研究已经表明,C末端E-螺旋介导丝氨酸重组酶DNA识别。在ZFR中,这种螺旋从C末端至N末端、从5'到3'结合DNA。因此,由于TALE以5'至3'的方向结合DNA,因此可以预计仅当TALE结合位点被定位在20-bp核心(图1B)的相反链时才能发生重组。

[0186] 有人选择利用AvrXa7产生TALER,因为这种TALE蛋白先前已经用于产生TALE核酸酶和转录因子。为方便起见,BamHI限制性位点侧接许多TALE,该TALE包括AvrXa7,并且多个基团已经用这种限制性位点以产生合成TALE融合。值得注意的是,BamHI片段保持TALE N末端完整,但会除去C末端的天然效应子结构域。采用这种策略并通过BamHI限制消化产生了Gin-AvrXa7融合。

[0187] 将Gin-AvrXa7克隆到含有由中央20-bp核心序列组成的重组位点和两个侧接26-bp AvrXa7结合位点的pBLA选择载体中,该中央20-bp核心序列由Gin催化结构域识别。正如预期的那样,当AvrXa7结合位点被定位邻近该20-bp核心(图1C)时,Gin-AvrXa7融合无法重组DNA。然而,当AvrXa7结合位点被定位在20-bp核心的相反链上时,重组变得明显(图1C),这表明重组位点取向是催化结构域融合到TALE N-末端的关键组成。为了进一步确立N-末端融合对于重组是必要的,构建出C-末端AvrXa7-Gin变体,其含有预计约束催化结构域活性的非规范融合方向(图1B和表5)。正如预期的那样,确定了这种C-末端AvrXa7融合在细菌细胞中显示出可忽略的活性(图1C)。

[0188] 表5。

[0189]

变体	SEQ ID NO:	序列
Gin-Avr (#1) /Avr20G	195	TTAGCACCTGGTTGGAGGGGGTTTATA TCCAAAACCATGGTTTACAG TATAAACCCCTCCAACCAGGTGCTAA
Gin-Avr (#2)	196	TTAGCACCTGGTTGGAGGGGGTTTATA AGGTTTTGGTACCAAATGTC TATAAACCCCTCCAACCAGGTGCTAA
Avr-Gin (#3)	197	TATAAACCCCTCCAACCAGGTGCTAA CTGTAAACCATGGTTTTTGA TTAGCACCTGGTTGGAGGGGGTTTATA
Avr14G	198	TTAGCACCTGGTTGGAGGGGGTTTATA AAAACCATGGTTTA TATAAACCCCTCCAACCAGGTGCTAA
Avr26G	199	TTAGCACCTGGTTGGAGGGGGTTTATA GCTTCCAAAACCATGGTTTACAGGGT TATAAACCCCTCCAACCAGGTGCTAA
Avr32G	200	TTAGCACCTGGTTGGAGGGGGTTTATA GCTTCATCCAAAACCATGGTTTACAGGGTTCC TATAAACCCCTCCAACCAGGTGCTAA
Avr44G	201	TTAGCACCTGGTTGGAGGGGGTTTATA GCTTCAGCTTCATCCAAAACCATGGTTTACAGGGTTCCGGTTCC TATAAACCCCTCCAACCAGGTGCTAA
Avr20GG	202	TTAGCACCTGGTTGGAGGGGGTTTATA TCCAAAACCGGGGTTTACAG TATAAACCCCTCCAACCAGGTGCTAA
Avr20T	203	TTAGCACCTGGTTGGAGGGGGTTTATA

[0190]

		CGAAATATTATAAAATTATCA TATAAACCCCCCTCCAACCAGGTGCTAA
Avr32GG	204	TTAGCACCTGGTTGGAGGGGGTTTATA GCTTCATCCAAAACCGGGGTTTACAGGGTTCC TATAAACCCCCCTCCAACCAGGTGCTAA
Avr32T	205	TTAGCACCTGGTTGGAGGGGGTTTATA GCTTCACGAAATATTATAAAATTATCAGGTTC TATAAACCCCCCTCCAACCAGGTGCTAA
Avr-G-ZF	206	GCGGGAGGCGTG TCCAAAACCATGGTTTACAGGGTTCC TATAAACCCCCCTCCAACCAGGTGCTAA
PthXo1-20G	207	GTGGTGTACAGTAGGGGGAGATGCA TCCAAAACCATGGTTTACAG TGCATCTCCCCCTACTGTACACCAC
PthXo1-32G	208	GTGGTGTACAGTAGGGGGAGATGCA GCTGCTTCCAAAACCATGGTTTACAGGGTGGT TGCATCTCCCCCTACTGTACACCAC

[0191] 设计截短。

[0192] 尽管上述Gin-AvrXa7融合催化重组,但是该变体的活性比设计的ZFR的活性要低得多。此外,特异性分析显示,Gin-AvrXa7融合无法进行包含非同源DBD位点的识别位点与非天然20-bp核心序列之间的忠实鉴别,表明重组可能不是Gin介导的(图1D)。最近的报告表明,TALEN活性可以在融合蛋白的TALE部分被截短时增强。因此,为了试图提高TALER活性,生成了一系列N末端和C末端AvrXa7截短(图2A)。

[0193] 以大致相等的间隔在始于AvrXa7Thr 27 (Δ 27) 并止于AvrXa7Gly 268 (Δ 268) 的位置上装配十个N-末端截短(图6)。也生成了AvrXa7 Δ 150,其已被报道为TALEN的N-末端截短变体。在位置28(+28)和位置95(+95)处生成两个C末端AvrXa7截短。+28和+95已被报道为在TALEN中的稳定融合点。每个TALE截短变体被融合到Gin催化结构域,并且将这20-成员TALER文库克隆到含AVR-20G识别位点的pBLA选择载体中。在细菌细胞中进行一轮选择(材料和方法)后,对个体的氨苄青霉素-耐药性克隆进行测序,发现所有选定的TALER包含两个N末端截短: Δ 87和 Δ 120中的任一个。每个选定的克隆也是在C末端+28。除了在靠近融合点(Δ 120*)处具有自发的12氨基酸缺失的单个 Δ 120克隆,这些克隆的活性是相当低的(图2B)。在这种分析中,基于Gin的ZFR常规地显示20-40%的重组,但是,在选定的TALER融合中所观察到的最高活性为~7%的重组(Gin-AvrXa7 Δ 120*)。因为TALE DBD比ZF结构域大三倍(不包括所需的侧接肽序列),所以我们推断用于这些TALER构建体的20-bp间隔区可能不是重组的最佳长度。

[0194] 核心序列长度。

[0195] 接下来通过评估含有14 (Avr-14G)、26 (Avr-26G) 和32-bp (Avr-32G) 核心位点的

DNA靶是否能通过选定的TALER重组来研究核心序列长度对重组的影响。为了在重组酶介导的重装配后保持 β 内酰胺酶基因的阅读框,通过 ± 3 -bp来修饰核心半位点(表1)。用每个靶位点变体对上述20-成员TALER文库进行一轮选择。虽然鉴定能够重组最短靶的TALER变体是不可能的,但是鉴定了Avr-14G(数据未示出)、重组Avr-26G和Avr-32G的两个Gin- Δ AvrXa7变体(基于N-末端TALE截短 $\Delta 87$ 和 $\Delta 120$ 以及C-末端截短+28)其。具体地,克隆分析表明,所选择的TALER(Gin-AvrXa7 $\Delta 87$ 和Gin-AvrXa7 $\Delta 120$)重组具有较长核心(例如,26和32-bp)的DNA比重具有较短核心(例如,14和20-bp)的DNA有效至少100倍(图2B)。此外,人们发现,Gin-AvrXa7 $\Delta 120$ 重组含有同源核心序列(Avr-26G和Avr-32G)的靶比重含有非同源核心(Avr-20T、Avr-20GG、Avr-32T和Avr-32GG)的靶有效 >100 倍(图2C)。有趣的是,Gin-AvrXa7 $\Delta 120$ 融合在44-bp核心(Avr-44G)上并不活跃(重组比在Avr-32G上的重组低 ~ 3 倍)(图2C),这表明26和44-bp之间的核心长度对于由Gin-AvrXa7 $\Delta 120$ 在大肠杆菌中的重组来说可能是最佳的。

[0196] 递增截短文库。

[0197] 虽然Gin-AvrXa7 $\Delta 120$ 相比于Gin-AvrXa7表现出增加的重组,但是怀疑Gin-AvrXa7 $\Delta 120$ 可能不是最佳的TALE融合架构,因为:(i)含有Gin催化结构域的ZFR重组DNA比Gin-AvrXa7 $\Delta 120$ 有效 >2 倍,以及(ii)没有从TALE截短变体的综合文库中鉴别出Gin-AvrXa7 $\Delta 120$ 。因此,为了找出更好的融合架构,基于递增截短TALE DBD的文库的产生设计筛选方式。

[0198] 为了实现这一目标,采用如前所述的方案以使未修饰的N末端结构域(Gin)能够融合至截短C-末端片段(AvrXa7)的文库(材料和方法)。通过外切核酸酶消化产生跨越在AvrXa7N-末端截短(Met1)和第一AvrXa7重复序列(Leu 298)之间的区域的N末端AvrXa7截短并融合至Gin催化结构域的未修饰的拷贝(蛋白质变体的理论数目: ~ 300)。因为先前的结果表明,+28是最佳的C-末端截短,所以我们将这种架构并入截短文库。将TALER克隆到含有Avr-32G靶位点的pBLA选择载体,并转化到大肠杆菌($>1 \times 10^5$ 转化子)。序列分析证实跨越目的区域的截短的相等分布(数据未显示)。

[0199] 在三轮选择之后,对个别氨苄青霉素-耐药性克隆进行测序并且鉴定出许多独特的截短变异体(图3A)。与利用20-成员TALE截短文库进行的选择一致,该选择表明最佳的N-末端TALER融合点可能位于接近位置87和位置120的位置,发现所有选定的Gin-AvrXa7变体含有介于位置74($\Delta 74$)和147($\Delta 147$)之间的截短。特别地,73个克隆中的26个克隆(35.6%, $P<0.001$)含有介于位置124($\Delta 124$)和129($\Delta 129$)之间的截短。根据这个占有数,在位置128($\Delta 128$)处的截短最有代表性。

[0200] 为了系统地确定选择的AvrXa7结构域是否增加了TALER活性,我们在大肠杆菌中评估分离的Gin-AvrXa7变体对含有Avr-32G靶位点的DNA底物的性能。我们集中分析在AvrXa7位置92($\Delta 92$)和134($\Delta 134$)之间含有N-末端缺失的克隆。与序列分析一致,结果发现在 $\Delta 120$ 和 $\Delta 129$ 之间含有N-末端截短的TALER重组DNA比基于相对较长或较短的截短的变体重组DNA更有效,尽管 $\Delta 92$ 融合也相当有效(图3B)。三个克隆进一步的特征在于: $\Delta 74$ 和 $\Delta 145$ 被选中是因为它们代表了可能的融合点的界限,并且对 $\Delta 128$ 进行测定是因为它是在选择中发现的最普遍的克隆。对具有从14到44-bp的间隔长度的五个靶连同三个阴性对照(Avr32T、Avr32GG和PthXo1-32G)一起进行分析。经测定,Gin-Avr32G $\Delta 74$ 和Gin-Avr32G

$\Delta 145$ 对于比20-bp长的间隔区具有中度的活性,而Gin-Avr32G $\Delta 128$ 重组DNA具有比得上ZFR GinC4的效率(图3C)。此外,特异性分析表明,Gin-Avr32G $\Delta 74$ 、Gin-Avr32G $\Delta 128$ 和Gin-Avr32G $\Delta 145$ 重组含有同源核心的底物比重组含有非同源核心(Avr-32T、Avr-32GG和PthXol-32G)底物的有效性>100倍(图3C)。总之,这些结果表明,在 $\Delta 120$ 和 $\Delta 129$ 之间含有N-末端缺失的TALE蛋白代表用于融合到重组酶的最佳截短。

[0201] 掺入合成TALE重复阵列。

[0202] 上述本研究使用天然存在的AvrXa7TALE蛋白的天然DBD。为了确定设计的TALE重复阵列是否可以掺入所选定的Gin- Δ AvrXa7框架,生成设计为靶向AvrXa7结合位点的一系列合成TALE蛋白(长度为15至20重复序列)(图7)。利用公开可用的TALEN质粒组(Addgene)构建TALE蛋白。修饰克隆质粒以包含+28C-末端截短和 $\Delta 120$ 或 $\Delta 128$ N-末端截短。将设计TALE融合到Gin催化结构域(表示为Gin-Avr15 $\Delta 120$ 和Gin-Avr15 $\Delta 128$),并克隆到含有Avr-32G或Avr-32T靶位点的pBLA选择载体。

[0203] 在大肠杆菌中的活性分析显示,当Gin-Avr15 $\Delta 120$ 和Gin-Avr15 $\Delta 128$ 融合到活性催化结构域时,都可以用于重组DNA,而且合成重复序列的掺入提供了增加的活性(图4A)。重要的是,每个TALER显示出严格的选择性,重组含有同源核心的靶位点比重组含有非同源核心的靶位点有效>1000倍(图4B)。令人惊讶地,还发现基于 $\Delta 120$ 截短的TALER重组DNA与基于 $\Delta 128$ 架构的TALE一样有效(图4A),这表明设计的TALE不如含有天然AvrXa7DBD的那些对N-末端截短敏感。

[0204] 为了进一步证明本文所述的TALER架构可以被重新编程以靶向任何DNA序列,合成酶被创造设计以靶向由天然存在的TALE蛋白PthXol (Gin-Pth15 $\Delta 120$) 识别的序列。人们发现,Gin-Pth15 $\Delta 120$ 在其同源底物上具有高度活性而且Gin-Pth15 $\Delta 120$ 和Gin-Avr15 $\Delta 120$ 两者在重组具有它们的同源结合位点的靶中表现出>600倍的增加(图4A)。还评估了含有长度介于15和20个重复序列之间的DBD的一系列设计的TALER的活性,并且发现每个融合以相似的高效性和特异性催化重组(图4B),这表明掺入了合成TALE重复阵列的嵌合重组酶可以用于位点特异性重组。

[0205] TALER在哺乳动物细胞中的活性。

[0206] 还确定了TALER是否能在哺乳动物细胞中修饰DNA。为了实现这一目的,我们使用能够快速评估细胞培养中的重组酶活性的附加体报告基因分析。在该分析中,用重组酶表达载体和报告质粒(pGL3)对人胚胎肾(HEK) 293T细胞进行共转染,该报告质粒在侧接重组位点的SV40启动子的控制下包含荧光素酶基因。合适的重组酶的瞬时表达导致SV40启动子切除并减少荧光素酶在细胞中的表达。因此,重组酶活性与荧光素酶表达的减少倍数成正比。

[0207] Gin-Avr15 $\Delta 120$ 与携带Avr-44G识别位点(pGL3-Avr-44G)的报告质粒共转染导致荧光素酶的表达相比于其与pGL3-Avr-44G单独转染减少~20倍(图5A)。尽管事实上在大肠杆菌中Gin-Avr15 $\Delta 120$ 显示出与ZFRGinC4相似的活性,但是我们发现GinC4在与其同源靶质粒,pGL3-C4-20G共转染之后,使荧光素酶表达减少了>80倍(图5A)。这种差异可能是由于相比于pBLA,在pGL3中重组酶靶位点之间的相对较短的插入DNA序列或者在哺乳动物细胞中TALER和ZFR之间的差异表达。然而,这种差异的根本原因仍然不清楚。最后,尽管32-bp被确定为在大肠杆菌中用于TALER的最佳核心序列长度,但是确定了Gin-Avr15 $\Delta 120$ 与pGL3-

Avr-32G的共转染导致荧光素酶的表达仅有6倍的减少(图5A)。这种差异的根本原因仍然不清楚。

[0208] 接下来对ZFR (GinC4) 和TALER (Gin-Avr15 Δ 120) 是否可以在哺乳动物细胞中形成相容的异二聚体进行了研究。为了评估这种可能性,产生了混合的重组位点,其中AvrXa7结合位点和C4锌指结合位点(GCGGGA GGC GTG;SEQ ID NO:279)侧接由Gin催化结构域(pGL3-Avr-G-ZF)识别的核心序列(见表2)。令人惊奇的是,pGL3-Avr-G-ZF与GinC4和Gin-Avr15 Δ 120共转染导致荧光素酶的表达相比于其与pGL3-Avr-G-ZF转染减少>140倍(图5B),而GinC4或者Gin-Avr15 Δ 120与pGL3-Avr-G-ZF的转染导致报告基因表达的减少可以忽略不计。这些结果表明,产生ZF-TALE异二聚体代表一种用于改善嵌合重组酶的靶向能力的潜在有效的方法。

[0209] 讨论。

[0210] 不同于含有极小融合架构的ZFP,TALE DBD在DBD阵列的任一侧上需要天然蛋白框架以发挥作用。在几乎所有已知的TALE识别位点中发现的所谓的第0重复序列和第一重复序列就表示这样的N-末端框架并介导胸腺嘧啶残基在位置0处的结合。最近的晶体结构提供了位置0处胸腺嘧啶的结合的说明,但数据仍不足以确定最小的TALE架构。事实上,直到现在所有的研究已经使用含有比介导位置0处的结合所需要的那些残基多得多的残基的N-末端截短。仍然不确定这部分蛋白在实现适当的DNA结合构象中起什么作用或者什么可以构建最小的TALE结构域。虽然最初尝试基于融合到全长TALE蛋白以产生功能TALE嵌合体,但是最近的研究集中在独特的C末端截短的鉴别上,其在 Δ 150N-末端结构下改善效应子结构域的功能。先前的报告指出,AvrBs3TALE的N-末端残基2-153(Δ 150)的缺失去除了TALE从它的原生细菌移位到靶植物细胞所需要的结构域,但不会影响转录因子活性。

[0211] 然而,开发活性的TALER使得鉴定独特的N-末端TALE变体成为必要。最初对具有C-末端截短+28和+95的N末端TALE进行了广泛的、系统性的调查,并发现只有两个结构域(具有+28的 Δ 87和具有+28的 Δ 120)对进一步的分析显示出足够高的活性。基于AvrXa7N-末端的递增截短的二次分析导致以AvrXa7位置74(Δ 74)到位置145(Δ 145)为中心的截短变体的广泛簇的鉴定。在这个实验中回收的克隆,38%在位置 Δ 119和 Δ 128之间含有截短,并且从在这个区域具有融合的TALER得到的调查数据表现出高活性。特别地,确定基于来自这个区域(Δ 128和 Δ 120)的N-末端截短的TALER可用于在细菌和哺乳动物细胞中重组DNA。 Δ 119和 Δ 128之间的截短变体簇也可以指示这个区域的内在稳定性。

[0212] ZFR通常催化长度介于44至50-bp之间的靶位点的重组。每个靶位点包含中央20-bp核心序列,其由重组酶催化结构域识别,以及两个相邻的ZFP结合位点。然而,TALER的融合方向使得TALE结合位点位于相对于中央核心序列的相反链上成为必要。这种独特的几何特征导致我们调查了用于重组的最低核心序列要求。由于TALE DBD的长度(TALE重复序列比ZFP长3至4倍)和催化结构域与TALE结构域之间的扩展的N-末端接头,我们推断,对重组来说,较长的核心序列(32或44-bp)将会是必要的。事实上,除了携带自发的缺失(Δ 120*)的TALE变体,在本研究中鉴别的大多数的N-末端截短变体显示出对32-bp核心的最佳性能。这些结果与关于TALEN的文献报道相一致,TALEN不像ZFN,其需要显著较长的间隔序列(例如TALEN:17至20-bp,ZFN:5至6-bp)以有效地切割DNA。支持这些观察,我们发现以短核心序列(14-bp)为标准选择独特的N-末端截短变体没有产生任何克隆。

[0213] Gin-AvrXa7 Δ 128被确定为最佳的TALE融合,但使用利用公开可用的TALE装配试剂盒生成的合成TALE蛋白的随后研究表明,基于 Δ 128和 Δ 120的TALER在大肠杆菌中表现出相似的活性。这些设计的TALE是基于嵌合蛋白的,该嵌合蛋白来源于密切相关的和天然存在的Tallc和PthXo1TALE蛋白。虽然这些TALE共享高同源性,但它们是不相同的。尽管在残基12和残基13之外的RVD重复序列中的多态性已被证明对TALE融合活性没有影响,但就我们所知,还没有DBD之外的TALE框架中的差异的系统评价的报道。如递增截短文库的分析所表明的,较小的氨基酸改变可以显著影响特定的融合的活性。因此,我们观察到的Gin-AvrXa7 Δ 120和合成Gin-Avr15 Δ 120之间的活性的差异可能归因于AvrXa7框架和先前使用的TALE框架架构之间的序列变化。

[0214] 受构建合成TALE青睐的四种RVD (NI:A、HD:C、NG:T以及NN:G) 是自然界最普遍的,然而,仍然有待确定这些重复序列是否代表了最特定RVD模块。对于26-重复序列AvrXa7TALE而言,靶定相同序列的合成版本将在RVD组成中具有16种变化(图7)。据推测,因为它们更常见地发现于自然界,因此为合成用途选择的四种RVD可能对它们的同源碱基比其它RVD具有较高的亲和力。如果确实是这样,则可以合理地假定用合成RVD重复序列创建的TALE可具有比使用天然结构域创建的TALE较高的DNA结合亲和力。虽然RVD亲和力的问题没有直接解决,但确定了含有合成重复阵列的TALER比含有天然AvrXa7DBD的构建体更活跃。具有合成DBD的TALER尽管含有明显更少的DBD,仍然显示出比含有天然重复序列的构建体高出大约2倍的活性。此外,使用合成阵列观察到的活性的增益与脱靶重组中的任何增加无关。

[0215] 一些研究已经表明,TALE可以容忍在其靶序列中的一些错配。这些发现并不令人吃惊,因为与特定的碱基正相关的RVD已经显示出本质上容忍非同源碱基。然而,由TALER提供的协同的特异性可以用于规避潜在的局限性。因为催化结构域有助于特异性重组,可以想到的是,也可以生成能够选择性修饰高度同源的基因组序列的设计TALER。事实上,最近已表明,重组酶催化特异性可以被有效地重新编程以靶向非天然核心位点。

[0216] 实施例2

[0217] 新型第0位残基特异性的选择

[0218] 设计了一类新型的基于Ta1的DNA结合蛋白。TAL(转录激活子样)效应子构建了一类新型的具有可预测的特异性的DNA结合蛋白。黄单胞菌属的革兰氏阴性植物致病细菌利用Ta1效应子,该细菌经由III型分泌系统(T3SS)将不同的效应子蛋白的混合物导入植物细胞中,在植物细胞中它们作为毒力基因。TAL的DNA结合特异性是由串联重复序列的中央结构域来确定的。每个重复序列赋予DNA中的一个碱基对(bp)的识别。重复序列模块的重排使得具有期望的DNA结合特异性的蛋白的设计有某些重要的限制。例如,靶定具有Ta1结构域的DNA序列的最约束的特征是Ta1 DNA位点以碱基T开始以及有时候以碱基C开始的要求。已经不可能在-1位置处靶定以G碱基或A碱基开始的结合位点。Ta1重组酶活性选择用来选择Ta1 DNA结合结构域,其通过使突变靶向第-1位和第0位RVD区而没有这种限制。这一发现的实际后果是巨大的,因为现在每个DNA序列都可以被新的Ta1结构域靶向,促成了新的对于TAL转录因子的无限制的接近,从而启动/上调或停止/下调转录、靶向TAL核酸酶以敲除基因功能或者直接同源重组或者靶向我们自己的TAL重组酶或其它TAL酶。

[0219] 对于在(-1)位置处的G特异性,首先利用GinAvr15 Δ 128-合成蛋白的(-1)结构域

之内的NNK密码子策略使氨基酸QWSG (SEQ ID NO:209) 随机化。经过3轮的所得的文库的Ta1重组酶活性选择,选择了在靶区域中具有选定的序列RSNG (SEQ ID NO:210) 和SRSG (SEQ ID NO:211) 的新型Ta1结合结构域。然后相对于由起始克隆识别的亲本T,这些显示出结合靶序列的第0位置处的G。重复选择以随机化下面红色示出的KQW区域,该区域与最初选择的QWSG (SEQ ID NO:212) 有重叠。现在对具有选定的SSR、SRA、SRC和KRC序列的克隆进行选择。在结合研究中,用限定的携带G取代的寡核苷酸对所有选定的Ta1结合结构域进行分析,所有选定的Ta1结合结构域现在显示出优先结合序列G-ATAAACCCCTCCAA (SEQ ID NO:213)。需要注意的是,使用相同的序列进行Ta1重组酶活性选择。起始Ta1结合蛋白GinAvr15 Δ 128结合T-ATAAACCCCTCCAA (SEQ ID NO:214)。对携带所选择的突变的Ta1核酸酶进行的随后测试证实这些序列的G指定使得这种新型的Ta1能够被首次开发。所选择的序列可以移植到来源于其它物种的Ta1。

[0220] 表6

[0221]

选择	
SEQ ID NO: 215	ATHEDIVGVGKQWSGARALEALLTDAGELRGPPLQ (-1 结构域)
SEQ ID NO: 216	ATHEDIVGVGKQWSGARALEALLTDAGELRGPPLQ (用粗体表示随机化的 AA)
SEQ ID NO: 217	KQWSG-起始克隆序列
SEQ ID NO: 218	KRSNG-选定以结合 G
SEQ ID NO: 219	KSRSG-选定以结合 G
SEQ ID NO: 220	ATHEDIVGVG K QWSGARALEALLTDAGELRGPPLQ
SEQ ID NO: 221	K QWSG- WT
	SSR-选定以结合 G
	SRA-选定以结合 G
	SRC-选定以结合 G
	KRC-选定以结合 G

[0222] 还利用这种靶向A的相同的文库进行选择。在这项研究中,选择了序列PRG、PTR以及PKD。在结合研究中,用限定的携带A取代的寡核苷酸对所有选定的Ta1结合结构域进行分析,所有选定的Ta1结合结构域现在显示出优先结合序列A-ATAAACCCCTCCAA (SEQ ID NO: 222)。需要注意的是,使用这种相同的序列进行Ta1重组酶活性选择。起始Ta1结合蛋白

GinAvr15 Δ 128结合T-ATAAACCCCTCCAA (SEQ ID NO:223)。对携带所选择的突变的Ta1核酸酶进行的随后测试证实这些序列的A指定使得这种新型的Ta1能够被首次开发。可以通过随机诱变N-末端结构域或靶向诱变在第0位结构域内的KRGG (SEQ ID NO:223) 序列并在重组酶系统中重新选择来实现随后的对结合活性的优化。

[0223] 实施例3

[0224] 选择

[0225] 对于环境依赖的RVD选择和具有新特异性的RVD的选择,创建了文库,其随机化以下加粗的HD序列。LTPDQVVAIASHDGGKQALETVQRLLPVLCQDHG (原型RVD序列;SEQ ID NO:225)

[0226] 典型地,尽管受限于N、D、H、K和Q氨基酸的文库通常为H残基的成功替代品,但是文库允许所有的氨基酸处在这两个位置。替代地,随机化SHDG (SEQ ID NO:226) 和ASHDGG (SEQ ID NO:227) 区域的较大的文库允许具有环境依赖的特征的独特的RVD特异性的选择。

[0227] 然后Ta1重组酶活性选择迅速允许靶向RVD结构域内的新特异性的选择。所得到的RVD在其序列识别中可以是高度模块化的或者环境依赖的,并且可以接着用于创建Ta1核酸酶和转录因子。

[0228] 这种技术的应用包括对于TAL转录因子的无限制的接近,从而启动/上调或停止/下调转录、靶向TAL核酸酶以敲除基因功能或者直接同源重组或者靶向我们自己的TAL重组酶或其它TAL酶以用作工具和治疗。

[0229] 这一发现的优点和实际效果是巨大的,因为现在每个DNA序列都可以被我们的新的Ta1结构域靶向并且它们的特异性可以很容易地优化。

[0230] 实施例4

[0231] TALE N-末端结构域的定向进化以容纳除胸腺嘧啶外的5'碱基

[0232] 转录激活子样效应因子 (TALE) 蛋白可以被设计成结合几乎任何目的DNA序列。靶向植物无毒基因的天然TALE转录因子 (TALE-TF) 的DNA结合位点具有5'胸腺嘧啶。合成的TALE-TF也有这个要求。最近的结构数据表明,靶序列的N-末端结构域 (NTD) 与5'T之间发生相互作用。最近的TALE核酸酶 (TALEN) 文献调查得出有关靶序列的第一碱基, N₀残基的重要性的矛盾数据。此外,没有关于N₀碱基对TALE重组酶 (TALE-R) 的活性的影响的研究。在这里,在TALE-R、TALE-TF、表达为具有麦芽糖结合蛋白 (MBP-TALE) 和TALEN的融合体的TALE DNA结合结构域的结合区域中,对N₀碱基的影响进行量化。这些TALE平台中的每一个具有不同的N-末端和C-末端架构,但当N₀碱基为胸腺嘧啶时,所有这些都显示出最高的活性。为了简化在这些平台中构建有效的TALE的规则,以及允许在任意DNA序列处的高精度基因工程应用,我们利用我们最近研发的TALE-R系统设计了结构导向的活性选择。鉴定了新型的NTD序列,其对具有5'G的TALE结合位点有高度活化的和选择性的TALE-R活性,并对额外的结构域序列进行选择,其容许任何对5'N₀残基的一般靶定。这些结构域被导入到TALE-TF、MBP-TALE和TALEN架构中并对具有非-T 5'残基的靶序列一贯表现出比野生型NTD所具有的活性更大的活性。新型NTD与金门TALEN装配方案相兼容,现在使得能够有效构建TALE转录因子、重组酶、核酸酶和DNA结合蛋白,其识别任何DNA序列以能够实现在不考虑限制大多数天然TALE蛋白的5'T规则的情况下,对基于TALE-蛋白的DNA的准确和无约束的定位。

[0233] 在这个实施例中使用了下面的材料和方法。

[0234] 寡核苷酸。

[0235] 引物和其它寡核苷酸(以下表4)订购于Integrated DNA Technologies(San Diego,CA)。

[0236] 表7.引物。

[0237]

引物	SEQ ID NO:	序列
KXXG Lib Rev	228	TCTCAACTCCCCCGCCTCCGTGAGCAAGGCCTCCAGAGCGCGTGCC CCMNNMNNTTTGCCGACGCCAACGATGTCTTCGTG
KXXXX Lib Rev	229	TCT CAA CTC CCC CGC CTC CGT GAG CAA GGC CTC CAG AGC GCG TGC MNN MNN MNN MNN TTT GCC GAC GCC AAC GAT GTC TTC GTG
XXXSG Lib Rev	230	CCCGCCTCCGTGAGCAAGGCCTCCAGGGCGCGTGCGCCGGAMNNM NNMNNGCCGACGCCAACGATGTCTTCGTGTGTGCG
KRGG Lib Rev	231	GGC ACC CGT CAG TGC ATT GCG CCA TGC ATG CAC TGC CTC CAC TGC GGT CAC MNN MNN MNN MNN TGC AAT CTT GAG AAG TTG GCC TGT GTC
Goldy TALEN fwd	232	AGAGAGAGAAGAAAATGAGATCTCCTAAGAAAAAGAGGAAGGTGC AGGTGGATCTACGCACGCTCGGCTAC
NTD-dHax3 Fwd	233	AGGAAGAAGAGAAGCATGAGATCTCCTAAGAAAAAGAGGAAGGTG ATGGTGGACTTGAGGACACTCGGTTA
NTD-dHax3 Rev	234	AAGAGAAGAAGAAGAAGCATTGCGCCATGCATGCACTGCCTCTA
pTal127 NotI fwd	235	CCC GCC ACC CAC CGT GC
N-端 SphI	236	TGC TCT ATG CAT GCA CTG CCT CC
pTAL127- SFI Fwd	237	AGA GAA GAG AAG AGA AGG CGC CCG CGG CCC AGG CGG CCT CGG GAT CCC CTC GGC CTC CGC GCG CCA AG
pTAL127-SFI +95 Rev	238	AGA GAG AGA GAG AGA GTC TAG AGG CCG GCC TGG CCG CTC ATC CCG AAC TGC GTC ATG GCC TCA TC

[0238]

pTAL127 Xba +28 Rev	239	GCC CCA GAT CCT GGT ACG CTC TAG AGG
Avr 5'A 生物素发夹	240	5'BiosgATC TTA GCA CCT GGT TGG AGG GGG TTT ATTGG GTT TTC CCAAT AAA CCC CCT CCA ACC AGG TGC TAA GAT
Avr 5'T 生物素发夹	241	5'Biosg/ATC TTA GCA CCT GGT TGG AGG GGG TTT ATAGG GTT TTC CCTAT AAA CCC CCT CCA ACC AGG TGC TAA GAT
Avr 5'G 生物素发夹	242	5'BiosgATC TTA GCA CCT GGT TGG AGG GGG TTT ATCGG GTT TTC CCGAT AAA CCC CCT CCA ACC AGG TGC TAA GAT
Avr 5'C 生物素发夹	243	5'BiosgATC TTA GCA CCT GGT TGG AGG GGG TTT ATGGG GTT TTC CCCAT AAA CCC CCT CCA ACC AGG TGC TAA GAT
CCR5-内 fwd	244	TTAAAAGCCAGGACGGTCAC
CCR5-内 rev	245	TGTAGGGAGCCCAGAAGAGA
CCR5-外 fwd	246	ACAGTTTGCATTCATGGAGGGC
CCR5-外 rev	247	CCGAGCGAGCAAGCTCAGTT
CCR5-插入缺失 fwd	248	CGCGGATCCCCGCCAGTGGGACTTTG
CCR5-插入缺失 rev2	249	CCGGAATTCACCTGTTAGAGCTACTGC
pGL3 NTD 填充片段 fwd	250	AGA GAG AGA GAG AGG CGG CCG CCC TAC CAG GGA TTT CAG TCG ATG TAC ACG TTC
pGL3 NTD 填充片段 rev	251	AAG AAG AAG AAG GAA GAG AAG TAG GCC TGT CAT CGT CGG GAA GAC CTG CGA CAC CTG C
pgl3 5X Avr Xho1	252	ACTGCTATCCGAGTATAAAACCCCTCCAACCAGGTATAAAACCCCT CCAACCAGGTATAAAACCCCTCCAACCAGGTATAAAACCCCTCCAA CCAGGTATAAAACCCCTCCAACCAGGATCTGCGATCTAAGTAAGCT
AvrXa7 32G A F	253	TTAATTAAGAGTCTAGAttagcaactgggtggagggggtttatTgeticaTCCAAAACC ATGGTTTACAGgggttecAATAAACCCCTCCAACCAGGTGCTAAAGAT CTAGGAGGAATTTAAAATGAG
AvrXa7 32G A R	254	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTGGAGGGGGTTTATTgc aaccCTGTAAACCATGGTTTTGGATgaagcAATAAACCCCTCCAACCAG GTGCTAACTGCAGTTATTTGTACAGTTCATC
AvrXa7 32G G F	255	TTAATTAAGAGTCTAGAttagcaactgggtggagggggtttatCgeticaTCCAAAACC ATGGTTTACAGgggttecGATAAACCCCTCCAACCAGGTGCTAAAGAT

[0239]

		CTAGGAGGAATTTAAAATGAG
AvrXa7 32G G R	256	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTGGAGGGGGTTTATCgc aaccCTGTAAACCATGGTTTIGGAtgaagcGATAAACCCCTCCAACCAG GTGCTAACTGCAGTTATTTGTACAGTTCATC
AvrXa7 32G C F	257	TTAATTAAGAGTCTAGAttagcacctggfaggagggtttatGgcttcaTCCAAAACC ATGGTTTACAGgggttccCATAAACCCCTCCAACCAGGTGCTAAAGAT CTAGGAGGAATTTAAAATGAG
AvrXa7 32G C R	258	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTGGAGGGGGTTTATGgc aaccCTGTAAACCATGGTTTIGGAtgaagcCATAAACCCCTCCAACCAG GTGCTAACTGCAGTTATTTGTACAGTTCATC
荧光素酶载体 $\approx$ pgl3 basic. XhoI/SphI		
正向 靶 含有:	SEQ ID NO:	
5x Avr15 n- 1c shpF:	259	actgatactegagcTATAAACCCCTCCAACCAGGcTATAAACCCCTCCAACC AGGcTATAAACCCCTCCAACCAGGcTATAAACCCCTCCAACCAGGcTA TAAACCCCTCCAACCAGGATCTGCGATCTAAGTAAGCT
5x Avr15 0=A n-1c	260	actgatactegagcAATAAACCCCTCCAACCAGGcAATAAACCCCTCCAACC AGGcAATAAACCCCTCCAACCAGGcAATAAACCCCTCCAACCAGGcA ATAAACCCCTCCAACCAGGATCTGCGATCTAAGTAAGCT
5x Avr15 0=C n-1c	261	actgatactegagcCATAAACCCCTCCAACCAGGcCATAAACCCCTCCAACC AGGcCATAAACCCCTCCAACCAGGcCATAAACCCCTCCAACCAGGcC ATAAACCCCTCCAACCAGGATCTGCGATCTAAGTAAGCT
5x Avr15 0=G n-1c	262	actgatactegagcGATAAACCCCTCCAACCAGGcGATAAACCCCTCCAACC AGGcGATAAACCCCTCCAACCAGGcGATAAACCCCTCCAACCAGGcG ATAAACCCCTCCAACCAGGATCTGCGATCTAAGTAAGCT
荧光素酶 逆向 引物:	263	TCAGAAACAGCTCTTCTTCAAATCT

[0240] TALE-R NTD进化质粒的生成。

[0241] 此前报道的TALE-R系统适合于这项研究。简而言之,用HindIII/Spel消化pBCS(含有氯霉素和羧苄青霉素-耐药性基因)。用HindIII/XbaI消化含双重组酶位点的填充片段(Avr X,其中X为N0碱基)并使其连接到载体中以创建分裂 β -内酰胺酶基因。然后用BamHI/SacI消化pBCS AvrX,并用BamHI/SacI消化Gin127-N-填充片段-Avr15并使其连接到载体中以创建Gin127-N-填充片段-Avr15-X。用NotI/StuI消化填充片段以在N₁TALE发夹处演变,以及用NotI/SphI消化填充片段以在N₀TALE发夹处演变。

[0242] TALE NTD演变文库的生成

[0243] 引物ptal127NotI fwd和反向引物KXXG lib rev或KXXXX lib rev被用于生成N₁TALE发夹处的N-末端变体,随后用NotI/StuI对其进行消化,然后使其连接到经消化的Gin127-AvrX。正向引物ptal127NotI fwd和反向引物KRGG lib rev被用于PCR扩增在NoTALE发夹中具有突变的文库。随后用NotI/SphI对其进行消化并使其连接到经NotI/SphI-消化的Gin127-AvrX。

[0244] TALE-R NTD演变分析。

[0245] 第1轮连接进行乙醇沉淀并转化到Top 10F' 电感受态细胞中,然后在SOC中回收1小时。细胞在含100mg/ml氯霉素的100ml超级肉汤(SB) 介质中生长过夜。通过标准程序将DNA分离。得到的质粒DNA(Rd 1输入) 被转化到Top 10F' 电感受态细胞;细胞在含100mg/ml羧苄青霉素和100mg/ml氯霉素的100ml SB中生长过夜。通过标准程序将质粒DNA分离。用NotI/XbaI消化第1轮输出并使其连接到具有互补的粘性末端的Gin127-AvrX载体。当观察到共有序列并且对克隆进行了表征时,这个方案被重复三次至四次。

[0246] 测定N-末端TALEN活性。

[0247] 使用金门方案生成含有每个可能的碱基的四个TALEN对。融合A质粒和融合B质粒经由第二金门反应被直接连接到Goldy TALEN(NΔ152/C+63) 框架。通过用BglIII/NsiI消化pCAG载体并与用BglIII/NsiI消化的PCR扩增的NTD连接来修饰NTD。TALEN对(每个TALEN/孔50-75ng) 被转染到在96孔板的孔中的密度为 1.5×10^4 细胞/孔的HeLa细胞中。转染后,将细胞置于37℃的培养箱中培养24小时,然后将其转移到30℃下持续2天,然后将其转移到37℃下持续24小时。根据公开的方案将基因组DNA分离,并用CellSurveyor分析以及通过测序对DNA突变率进行定量。针对Cell测定,通过巢式PCR扩增基因组DNA,首先用引物CCR5外fwd/CCR5外rev,然后用CCR5内fwd/CCR5内rev。针对插入缺失的测序,用CCR5插入缺失fwd/CCR5插入缺失rev进行第二PCR。然后用BamHI/EcoRI消化片段并使其连接到具有互补消化的pUC19。

[0248] TALE-TF和荧光素酶分析。

[0249] 用引物ptal127SFI fwd和N-端SphI对来自重组酶选择的变体NTD进行PCR扩增。对PCR产物进行扩增,用NotI/StuI消化并使其连接到pTAL127-SFI Avr15,其含有双SFI-1消化位点,促进N-末端修饰的TALE从pTAL127-SFI Avr15转移至pcDNA 3.0VP64。相应的TALE结合位点被克隆到pGL3Basic载体(Promega) 的荧光素酶基因上游。针对每个分析,根据制造商的说明利用Lipofectimine 2000(Life Technology),将5ng的pGL3载体和1ng的pRL海肾萤光素酶控制载体以及100ng的pcDNA共转染至96孔板的孔中的HEK293t细胞中。48小时后,对细胞进行洗涤、裂解并在Veritas微孔板光度计(Turner Biosystems) 上用双荧光素酶报告系统(Promega) 评估荧光素酶活性。转染以一式三份完成,并取平均结果。

[0250] MBP-TALE分析。

[0251] 利用前述的方案进行MBP-TALE结合到生物素化的寡核苷酸的亲和力分析。简单地说,从XL1-Blue细胞中的pMAL MBP-AvrXa7质粒表达AvrXa7TALE结构域并在直链淀粉树脂上纯化。具有修饰的残基的含有靶AvrXa7靶位点的生物素化的寡核苷酸被用来确定在夹心酶联免疫吸附试验设计中的TALE结合活性。靶定MBP取代的抗体被用于分析显色。

[0252] 结果。

[0253] 5'T规则的初步分析。

[0254] 结合到PthXo7 DNA序列的TALE蛋白的最近的晶体结构揭示了在N₁发夹中的W232与DNA底物(该N₀碱基)的接触区域的5'端处的胸腺嘧啶之间的独特的相互作用。这项研究为先前建立的5'T规则提供了结构基础,该5'T规则是在TALE密码首次被破译(图18A和18B)时报道的。出现了关于TALEN的靶序列的第一碱基的重要性的矛盾数据。对于靶DNA中的5'T的要求最初是在TALE-R的背景中使用含有所有可能的5'残基侧接Gin32G核心的四个AvrXa7结合位点的四个分裂β内酰胺酶TALE重组酶选择载体评估的(图18C)。然后用含有识别位点含有每个可能的5'残基的五聚体AvrXa7启动子区的四个荧光素酶报告载体评估由TALE-TF识别的N₀残基(图18D)。对于除了5'T之外的碱基,我们观察到其相对于具有5'T的序列(图18C和18D)TALE-R活性减少高达>100倍和TALE-TF活性减少高达1000倍。尽管据说这些嵌合体的C-末端架构中的变体除去了5'T偏差,尤其是在大大缩短的C-末端结构域(CTD)的存在下,仍然观察到这些减少。酶联免疫吸附试验也表明MBP-TALE DNA结合蛋白对具有非T 5'残基的靶寡核苷酸的亲和力降低(图18E)。最后,检查显示:带有野生型NTD的设计的TALEN对于具有非T 5'核苷酸的靶标的活性相比于对于具有5'T的靶标的活性减少高达10倍(图18F)。结果表明,5'T是在重组酶、转录因子、核酸酶和简单的DNA结合蛋白的背景下使TALE结构域最大程度有效的重要设计参数。

[0255] TALE NTD的演变以适应非T 5'残基

[0256] 为了创建更加灵活的用于DNA识别的系统,有人推测最近开发的TALE-R选择系统可以被利用来演变TALE的NTD,以除去5'T约束(图23)。通过随机化残基G234到残基K230生成文库,并且在几轮选择之后分离相对于每个可能的5'碱基具有活性的TALE-R(图19A-19C)。最活跃的选定的克隆展现出K230和G234的高度保守性;前者可接触DNA磷酸骨架,后者可能影响发夹环的形成(图24)。在文库K230-W232的情况下,经常观察到K230S,但在单独测定的几乎所有变体中都具有比K230R或K230变体低得多的活性。几个观察到具有W232到R232诱变的克隆中的一个(NT-G)表现出从5'T到5'G的选择性的显著转变;该序列类似于最近在这个区域中描述的罗尔斯通菌属TALE蛋白的NTD的序列。在植物转录因子的报告基因调控的背景中的罗尔斯通菌属NTD已被报道在其底物中优选5'G(参见用于蛋白质比对的图25)。如对5'G的NT-G的严格性所显示,残基R232可能特定地接触G碱基。NT-G对5'G的优先比得上野生型结构域对5'T的特异性。不能够得到对5'A或5'C特异的NTD变体,但得到了允许的NTD、NT-αN,其类似于接受具有任意5'残基的底物并保持高的活性的K265-G268 N₀发夹。据推测,这种变体与野生型NTD相比,使与DNA磷酸骨架的非特异性接触增强,增强了TALE-DNA复合物的整体结合而不接触特定的5'残基。据推测,缩短的发夹结构将允许对5'A或5'C残基具有特异性的变体的选择。在Q231-W232具有随机化并具有残基233缺失的文库被设计以缩短推定的DNA结合环。重组酶选择表明高度保守的Q231Y诱变,其在大量克隆中具有高的活性(图19D)。特别地,NT-βN相比于具有野生型NTD的TALE表现出对具有5'A、5'C或5'G的底物的增加的活性,但对5'T底物具有减少的活性(图19E)。

[0257] 演变的TALE NTD的应用。

[0258] 为了评估演变的NTD在设计TALE融合蛋白应用中的可移性,将优化的NTD掺入TALE-TF、MBP-TALE和TALEN。具有NT-G、NT-αN和NT-βN结构域的TALE-TF与具有NT-T结构域的TALE-TF相比时,显示出400-1500倍的荧光素酶靶基因的转录激活的增加,所述荧光素酶靶基因携带操作基因位点而无5'T残基。如在TALE-R选择系统中所观察的,基于NT-G的TF保

留了5'G选择性。基于NT- α N的TF和基于NT- β N的TF对所有5'核苷酸的活性追踪了在重组酶设计中观察到的相对活性(图20)。MBP-TALE相比于野生型MBP-TALE还表现出较大的对具有不含5'T的位点的靶寡核苷酸的相对结合亲和力(图26),其提供进一步的证据表明,所选择的结构域增强了对非胸腺嘧啶5'碱基的识别或容忍。

[0259] 然后将四个优化的NTD导入到Go1dy TALEN框架中。对于这些实验,四个底物被构建在CCR5基因的 Δ 32基因座的范围内(图21A)。每个底物含有不同的5'残基。实验包括具有野生型(NT-T)的TALEN和对5'T具有特异性的dHax3NTD(dHax3是从野油菜黄单胞菌中分离的常用的NTD变体)以确定基准基因编辑活性。设计底物TALEN对以保留尽可能多的RVD同源性(50-90%)以确定变体NTD的活性增强贡献(图21A)。

[0260] 通过测序并通过使用Cel11测定两者分析TALEN的活性。与含有野生型结构域的TALEN的活性相比,所选择的结构域对非T 5'残基表现出介于2倍和9倍的之间的基因编辑活性的增加(图21和图27)。具有野生型或dHax3NTD的TALEN对T1/T2表现出最高的活性。通过具有NT- α N、NT- β N和NT-G的TALEN最有效地处理TALEN对底物G1/G2,相对于NT-T具有2.0-3.5倍的增加。NT- α N对TALEN对A1/A2和TALEN对C1/C2具有的活性比野生型NT-T对TALEN对A1/A2和TALEN对C1/C2具有的活性分别高9倍和2倍。虽然在5'残基处的错配的影响在TALEN中比在TALE-TF和TALE-R框架中更温和,但当用于基因编辑实验中时,优化的NTD大大增加了TALEN活性。

[0261] 讨论

[0262] 大多数,但不是所有的,先前的研究已经表明,需要胸腺嘧啶作为最佳TALE DNA结合结构域的设计中的5'最末端残基。本文所述的分析表明,胸腺嘧啶对构建功能TALE融合蛋白是最佳的,并且在某些情况下是关键的。因此,这一要求对可以被TALE转录因子、核酸酶和重组酶嵌合体有效靶向的序列施加了限制。虽然这一要求理论上对TALEN用于基因敲除的应用施加了微小的限制,但由于其宽的间隔区忍耐度,可以容纳任何5'残基的NTD会进一步简化有效TALE构建的规则并大大提高基因工程和基因询问中需要精确TALE定位的应用(例如,使用TALEN在选定的碱基对处精确切割DNA、经由TALE重组酶进行无缝基因插入和交换、将天然DNA结合蛋白从特定的内源性DNA序列移位以询问它们的功能作用、用于途径工程的正交转录因子的发展、其中转录因子定位是关键的天然和合成基因的协同激活以及许多其它应用)。基于DNA的纳米技术的其它用途包括用特定的DNA结合蛋白修饰DNA纳米结构/DNA折叠(origami)。这里,基于DNA折叠/结构,靶向特定位点被约束,从而能够结合任何位点是至关重要的。具有DNA结合蛋白的这些结构和设备的精心设计可以用来扩展功能的有吸引力的方法。事实上,不难想象,当除去所有靶向约束时的DNA结合蛋白和它们的融合体的许多应用。在这些潜在的应用鼓舞下,我们的目的是开发能够靶向在任何碱基处起始的位点的NTD。

[0263] 最近开发的TALE-R系统被用来演变TALE的NTD以除去5'-T约束。在三轮的选择中,得到了对5'G具有特异性的NTD。进行了许多选择来尝试得到识别5'A或5'C的变体。使G230-K234发夹倒位、扩展K230-G234/ins232发夹、尝试改造K265-G268N₀发夹并评估随机诱变文库。虽然我们确实鉴别出具有缺失的NTD的NT- β N,其以可接受的亲和力识别具有5'A和5'C残基两者的底物,但这些策略都没有产生对具有5'A或5'C的靶序列具有亲和力的NTD。NTD NT-T和NTD NT-G表现出的强的选择优先性以及W232在NT-T中的重要性和R232在NT-G中的

重要性可能是由于这些氨基酸与DNA识别序列的5'末端残基的特异性相互作用。据最近报道,青枯雷尔氏菌TALE严格需要5'G,且与NT-G的序列比对显示这似乎是在类似于在NT-G中的232位置处含有精氨酸的可比较的N-1发夹(图25)。由于NTD Brg11和NT-T之间的高度结构同源性,所以可能通过简单的精氨酸向色氨酸突变修饰罗尔斯通菌属TALE NTD对胸腺嘧啶的优先性或者通过接枝NT- α N或NT- β N结构域至这个相关的蛋白来消除特异性。有趣地注意到,精氨酸-鸟嘌呤相互作用在演变的锌指结构域中是常见的。

[0264] 所选的变体NTD已成功导入TALE-TF、MBP-TALE和TALEN,并且根据重组酶演变系统的数据来看,通常赋予预期的活性和特异性。相对于NT-T对具有非-T 5'残基的AvrXa7启动子位点的活性,具有优化的NTD的TALE-TF使TALE激活增强了400倍和1500倍之间。当掺入TALEN时,我们的具有非T选择性的NTD的活性相对于NT-T结构域对具有5'A、5'C或5'G的底物所具有的活性增强了2-9倍。在TALEN基因编辑中的增加通常与在TALE-R和TALE-TF构建体中观察的活性增加有关。如在用TALEN对A1/A2、C1/C2和T1/T2的测定中的较低活性所证明的,保持了NT-G的特异性和高活性,并且NT- α N和NT- β N的一般高活性也被导入TALEN Δ 152/+63架构中。

[0265] 据最近报道,具有合成TALE RVD结构域的或选截短TALE不需要在DNA底物中的5'T。所报道的 Δ 143,+47截短被构建为GoIdy TALE-TF,并且观察到对AvrXa7底物的活性比 Δ 127,+95截短(已被他人常用并且其是我们的研究中所使用的截短组)大幅降低(图29)。因此,报道的结果中的差异可能是由于所使用的截短架构造成的。

[0266] 总之,在TALE-R、TALE-TF、MBP-TALE和TALEN嵌合体的背景下确定DNA底物中的5'胸腺嘧啶对于结合的重要性以及设计的TALE的活性。靶向诱变和TALE-R选择被应用于工程化TALE NTD,其识别除了胸腺嘧啶以外的碱基作为底物DNA的5'最末端碱基。本文开发的工程化TALE结构域表现出模块化并在TALE-TF和TALEN架构中高度活跃。这些新型的NTD将可以由当前的TALE-R靶向的位点数目扩展了 $\sim$ 15倍,TALE-R对它们的结合位点具有严格的几何要求并且对N₀碱基的鉴别高度敏感。此外,现在它们允许TALE DBD和TALE-TF在任何DNA序列精确定位以促进基因调控、内源DNA结合蛋白位移以及其中精确结合可能是关键的合成生物学应用。虽然基于天然NTD的TALEN表现出对N₀碱基取代的不同程度的容忍度,数据显示,本文报道的新型NTD与天然的基于NTD的TALEN相比,也有利于对任何N₀碱基进行较高效率的基因编辑。

[0267] 实施例5

[0268] 嵌合锌指重组酶

[0269] 使用了下面的材料和方法。

[0270] 分裂基因重组载体(pBLA)来源于pBluescriptII SK(-) (Stratagene),修饰该载体使其含有在lac启动子的控制下的氯霉素耐药性基因和中断的TEM-1 β 内酰胺酶基因。如前所述引入ZFR靶位点。简而言之,用引物GFP-ZFR-XbaI-Fwd和GFP-ZFR-HindIII-Rev对GFPuv(Clontech)进行PCR扩增并将其克隆到pBLA的SpeI和HindIII限制位点中以产生pBLA-ZFR底物。所有的引物序列列于表8中。

[0271] 表8. 引物序列

[0272]

引物	SEQ ID NO:	序列
GFP-ZFR- 20G-XbaI- Fwd	264	TTAATTAAGAGTCTAGAGGAGGCGTGTCCAAAACCATGGTTTACAGCAC <u>GCCCTCCAGATCTAGGAGGAATTTAAATGAG</u>

[0273]

GFP-ZFR- 20G-HindIII- Rev	265	ACTGACCTAGAGAAGCTTGGAGGCGTGTGTAAACCATGGTTTTGGACA <u>CGCCTCCCTGCAGTTATTGTACAGTTCATC</u>
SV40-ZFR-1- BglII-Fwd	266	TTAATTAAGAGAGATCTGCTGATGCAGATACAGAAACCAAGGTTTCTT <u>ACTTGCTGCTGCGCATCTGCATCTCAATTAGTCAGC</u>
CMV-PstI- ZFR-1 Rev	267	CACCACCACGGATCCGCAGCAGCAAGTAAGAAAAACCTTGTTTCTGTAT <u>CTGCATCAGCAATTTCGATAAGCCAGTAAGCAG</u>
5' Gin-HBS- Koz	268	CACCACCACGCGCGCAAGCTTAGATCTGCCCCAGCGGCCACCATGCT GATTGGCTATGTAAGGG
3' Gin-AgeI- Rev	269	CACCACCACACCGGTTCCCGATTAGGTGGGCGAC
ZFR-靶- 1-Fwd	270	GTTCCTGCCAGGATCCACTAG
ZFR-靶- 1-Rev	271	GCATGTGTCCAGATGCATAGG
ZFR-靶- 2-Fwd	272	CACCTTCTCCCAGGATAAAG
ZFR-靶- 2-Rev	273	GTTGGCCTGTATTCTCTGG
ZFR-靶- 3-Fwd	274	AATGAAGTTCCCTTGGCACTTC
ZFR-靶- 3-Rev	275	CTGAAGGGTTTTAAGTGCAGAAG
CMV-Mid Prim-1	276	TGACGTCAATGACGGTAAATGG
ZFR 靶标有下划线		

[0274] 为了生成荧光素酶报告质粒,用引物SV40-ZFR-BglIII-Fwd和SV40-ZFR-HindIII-Rev从pGL3-Prm (Promega) 对SV40启动子进行PCR扩增。用BglIII和HindIII消化PCR产物并使其连接到pGL3-Prm的相同的限制性位点以产生pGL3-ZFR-1、pGL3-ZFR-2、pGL3-ZFR-3……pGL3-ZFR-18。如前所述构建pBPS-ZFR供体质粒,具有以下例外:通过引物3'CMV-PstI-ZFR-1-Rev、3'CMV-PstI-ZFR-2-Rev或3'CMV-PstI-ZFR-3-Rev编码ZFR-1、ZFR-2和ZFR-3重组位点。通过序列分析验证每个质粒的正确构建。

[0275] 重组测定

[0276] 如前所述通过PCR装配ZFR。用SacI和XbaI消化PCR产物并使其连接到pBLA的相同的限制性位点。通过电穿孔将连接转化至大肠杆菌TOP 10F' (Invitrogen)。在SOC培养基中回收1小时后,用具有 $30 \mu\text{g ml}^{-1}$ 氯霉素的SB培养基孵育细胞并于37℃下培养。16小时后,收获细胞;通过Mini-prep (Invitrogen) 分离质粒DNA并用200ng pBLA转化大肠杆菌TOP10F'。在SOC中回收1小时后,将细胞接种在具有 $30 \mu\text{g ml}^{-1}$ 氯霉素或 $30 \mu\text{g ml}^{-1}$ 氯霉素和 $100 \mu\text{g ml}^{-1}$ 羧苄青霉素、氨苄青霉素类似物的固体LB培养基上。重组被确定为LB培养基上的含有氯霉素和羧苄青霉素的菌落数除以LB培养基上的含有氯霉素的菌落数。使用GelDoc XR成像系统 (Bio-Rad) 自动计数测定菌落数。

[0277] 选择

[0278] 如前所述通过重叠延伸PCR构建ZFR文库。突变被引入在具有简并密码子NNK (N:A、T、C或G以及K:G或T) 的位置120、123、127、136和137处,NNK编码所有20种氨基酸。用SacI和XbaI消化PCR产物并使其连接到pBLA的相同限制性位点。用乙醇对连接进行沉淀并将其用于转化大肠杆菌TOP 10F'。文库大小常规确定为 $\sim 5 \times 10^7$ 。在SOC培养基中回收1小时后,用具有 $30 \mu\text{g ml}^{-1}$ 氯霉素的100ml SB培养基于37℃下孵育细胞。16小时后,收获30ml细胞;通过Mini-prep分离质粒DNA并用 $3 \mu\text{g}$ 质粒DNA转化大肠杆菌TOP 10F'。在SOC中回收1小时后,用具有 $30 \mu\text{g ml}^{-1}$ 氯霉素和 $100 \mu\text{g ml}^{-1}$ 羧苄青霉素的100ml SB培养基于37℃下孵育细胞。在16小时,收获细胞,并通过Maxi-prep (Invitrogen) 分离质粒DNA。通过SacI和XbaI消化分离富集的ZFR并使其连接入新鲜pBLA用于进一步的选择。经过4轮的选择后,对单独的羧苄青霉素耐药性克隆进行序列分析。如上所述进行重组测定。

[0279] ZFR构建

[0280] 用引物5'Gin-HBS-koz和3'Gin-AgeI-Rev从各自的pBLA选择载体对重组酶催化结构域进行PCR扩增。用HindIII和AgeI消化PCR产物并使其连接到pBH的相同的限制性位点,以产生SuperZiF-兼容的亚克隆质粒:PBH-Gin-a、P、y、5、S或Z。通过SuperZiF组装锌指并使其连接入PBH-Gin-a、P、y、5、S或Z的AgeI和SpeI限制位点以产生pBH-ZFR-L/R-1、pBH-ZFR-L/R-2、pBH-ZFR-L/R-3……pBH-ZFR-L/R-18 (L:左ZFR;R:右ZFR)。通过SfiI消化从pBH释放ZFR基因并使其连接至pcDNA 3.1 (Invitrogen) 中,以产生pCDNA-ZFR-L/R-1、pCDNA-ZFR-L/R-2、pCDNA-ZFR-L/R-3……pCDNA-ZFR-L/R-18。通过序列分析 (表9) 验证每个ZFR的正确构建。

[0281] 表9. 催化结构域取代和预期的DNA靶

[0282]

i	催化结构域	靶	位置				
			120	123	127	136	137
	A	CC ^a	Ile	Thr	Leu	Ile	Gly
	B	GC	Ile	Thr	Leu	Arg	Phe
	F	GT	Leu	Val	Ile	Arg	Trp
	A	CA	Ile	Val	Leu	Arg	Phe
	ε ^b	AC	Leu	Pro	His	Arg	Phe
	ζ ^c	TT	Ile	Thr	Arg	Ile	Phe

[0283] ^a表示野生型DNA靶。[0284] ^b该ε催化结构域也包含取代E117L和L118S。[0285] ^c该ζ催化结构域也包含取代M124S、R131I和P141R。

[0286] 荧光素酶测定

[0287] 使人类胚胎肾 (HEK) 293和293T细胞 (ATCC) 维持在含有10% (体积/体积) FBS和1% (体积/体积) 抗生素-抗真菌 (抗-抗; Gibco) 的DMEM中。以每孔 4×10^4 个细胞的密度将HEK293T细胞接种到96孔板上, 并使其在湿润的5% CO₂气氛中于37℃下建立。在接种后24小时, 根据制造商的说明利用Lipofectamine 2000 (Invitrogen) 用150ng pcDNA-ZFR-L 1-18、150ng pcDNA-ZFR-R 1-18、2.5ng pGL3-ZFR-1、pGL3-ZFR-2、pGL3-ZFR-3或pGL3-ZFR-18以及1ng pRL-CMV对细胞进行转染。在转染后48小时, 用被动裂解缓冲液 (Promega) 使细胞裂解并使用Veritas微孔板检测仪 (Turner Biosystems) 利用双荧光素酶报告基因检测系统 (Promega) 测定荧光素酶的表达。

[0288] 整合测定

[0289] 以每孔 5×10^5 个细胞的密度将HEK293细胞接种到6孔板上, 并使其在湿润的5% CO₂气氛中于37℃下保持在含血清培养基中。在接种后24小时, 根据制造商的说明利用Lipofectamine 2000用1[^]g pcDNA-ZFR-L-1、1[^]g pcDNA-ZFR-L-2或1[^]g pcDNA-ZFR-L-3和1[^]g pcDNA-ZFR-R-1、1[^]g pcDNA-ZFR-R-2或1[^]g pcDNA-ZFR-R-3以及200ng pBPS-ZFR-1、200ng pBPS-ZFR-2或200ng pBPS-ZFR-3对细胞进行转染。在转染后48小时, 以每孔 5×10^4 个细胞的密度将细胞分开到6孔板上, 并使其保持在具有2[^]g ml⁻¹嘌呤霉素的含血清培养基中。达到100%汇合时收获细胞并用Quick Extract DNA提取液 (Epicentre) 分离基因组DNA。利用扩展的高保真Taq系统 (Expand High Fidelity Taq System) (Roche) 使用以下的引物组合对ZFR靶进行PCR扩增: ZFR-靶-1-Fwd、ZFR-靶-2-Fwd或ZFR-靶-3-Fwd和ZFR-靶-1-Rev、ZFR-靶-2-Rev或ZFR-靶-3-Rev (未修饰靶); ZFR-靶-1-Fwd、ZFR-靶-2-Fwd或ZFR-靶-3-Fwd和CMV-Mid-Prim-1 (正向整合); 以及CMV-Mid-Prim-1和ZFR-靶-1-Rev、ZFR-靶-2-Rev或ZFR-靶-3-Rev (反向整合)。对于克隆分析, 在转染后2天, 将 1×10^5 个细胞分开到100mm培养皿上, 并使其保持在具有2[^]g ml⁻¹嘌呤霉素的含血清培养基中。利用10mm × 10mm开放式 (open-ended) 克隆环使用无菌硅脂 (Millipore) 将单个菌落分离并在培养基中扩增。如上所述, 达到100%汇合时收获细胞并分离基因组DNA, 并用作PCR的模板。对于菌落计数测定,

在转染后2天,以每孔 1×10^4 个细胞的密度将细胞分开到6孔板中,并使其保持在含或不含 $2 \mu\text{g ml}^{-1}$ 嘌呤霉素的含有血清的培养基中,在16天,用0.2%结晶紫溶液对细胞进行着色并通过计数形成在含有嘌呤霉素培养基中的菌落数除以形成在不含嘌呤霉素培养基中的菌落数确定整合效率。使用GelDoc XR成像系统(Bio-Rad)自动计数测定菌落数。

[0290] 结果

[0291] Gin重组酶的特异性分布

[0292] 为了重新工程化丝氨酸重组酶催化特异性,对由这种酶家族识别底物的基础因素进行详细理解。为了实现此目的,对DNA转化酶Gin的催化结构域的活化突变体重组一整套对称取代靶位点的能力进行了评价。Gin催化结构域重组伪对称20-bp核心,它由两个10-bp半位点区组成。因此重组位点的这个集合包含在位置10、位置9、位置8、位置7、位置6、位置5和位置4处的每个可能的单碱基取代以及在位置3和位置2处以及在二核苷酸核心中的每个可能的两碱基组合。重组是由分裂基因重装配(先前描述的方法)测定的,该方法将重组酶活性与抗生素耐药性联系起来。

[0293] 通常,人们发现,Gin容忍(i)在二核苷酸核心处的16种可能的两碱基组合中的12种(AA、AT、AC、AG、TA、TT、TC、TG、CA、CT、GA、GT);(ii)在位置3和位置2处的16种可能的两碱基组合中的4种(CC、CG、GG和TG);(iii)在位置6、位置5或位置4处的单个A至T取代;以及(iv)在位置10、位置9、位置8和位置7处的所有12种可能的单碱基取代(图31A-31D)。此外,人们发现,Gin可以重组在位置10、位置9、位置8和位置7处含有至少 10^6 (可能的 4.29×10^9 中的)独特的碱基组合的靶位点文库(图31D)。

[0294] 这些结果与yS解离酶的晶体结构的观察相一致,其表明:(i)重组酶二聚体对整个二核苷酸核心的相互作用是对称的且以非特异性为主;(ii)在重组酶侧臂区中的进化上保守的Gly-Arg基序与DNA小沟之间的相互作用对位置6、位置5和位置4处的腺嘌呤或胸腺嘧啶施加了要求;以及(iii)侧臂区与小沟之间在位置10、位置9、位置8或位置7处没有序列特异性相互作用(图31E)。这些结果也与集中于确定密切相关的Hin重组酶的DNA结合特性的研究相一致。

[0295] 重新工程化Gin重组酶催化特异性

[0296] 基于这样的发现,即Gin容忍在位置3和位置2处的保守取代(即,CC、CG、GG和TG),对Gin催化特异性是否可以被重新工程化以特异性识别包含不被天然酶容忍的12种碱基组合中的每一种的核心序列(图32A)进行了研究。为了鉴定由Gin参与DNA识别的特定的氨基酸残基,对两种相关的丝氨酸重组酶(y6解离酶及Sin重组酶)在与其各自的DNA靶复合中的晶体结构进行了研究。基于这些模型,鉴定出在位置3和位置2处接触DNA的五种残基:Leu 123、Thr 126、Arg 130、Val 139和Phe 140(根据y5解离酶标号)(图32B)。通过重叠延伸PCR在Gin催化结构域(Ile120、Thr 123、Leu 127、Ile 136和Gly 137)中的等位残基上进行随机诱变并通过使这些催化结构域变体融合至未修饰的“H1”ZFP的拷贝构建ZFP变体的文库。这个文库的理论大小为 3.3×10^7 变体。

[0297] 将该ZFR文库克隆到含有不被天然酶容忍的五种碱基组合(GC、GT、CA、AC或TT)中的一种的底物质粒并通过分裂基因重装配富集活化ZFR(图32C)。经过4轮选择后,我们发现,每个ZFR群对含有GC、GT、CA和TT取代的DNA靶的活性增加>1000倍,以及每个ZFR群对含有AC取代的DNA靶的活性增加>100倍(图32D)。

[0298] 从每个群对单独的重组酶变体进行测序,并发现高水平的氨基酸多样性存在于位置120、位置123和位置127,还发现>80%的选定的克隆在位置136处含有Arg以及在位置137处含有Trp或Phe(图36)。这些结果表明,位置136和位置137在识别非天然核心序列中发挥关键作用。对每个选定的酶重组其靶DNA的能力进行了评估,结果发现,几乎所有的重组酶显示出活性(>10%重组),并朝向其预期的核心序列显示出>1000倍的特异性转变(图37)。就亲本Gin而言,人们发现,几种重组酶容忍在位置3和位置2处的保守取代(即,对GT和CT或AC和AG的交叉反应),表明单个重新工程化的催化结构域可以用于靶向多个核心位点(图37)。

[0299] 为了进一步调查重组酶特异性,对五个Gin变体(以下称Gin p、Gin y、Gin 6、Gin e和Gin Z)重组分布进行测定,这五个Gin变体显示为识别不被亲本酶容忍的12种可能的两碱基组合中的九种(GC、TC、GT、CT、GA、CA、AG、AC和TT)(表1)。Gin p、Gin 6和Gin e重组其预期的核心序列的活性和特异性比得上亲本酶(以下简称为Gin a)重组其预期的核心序列的活性和特异性,并且Gin y和Gin Z能够重组其预期的核心序列的活性和特异性超过了Gin a能够重组其预期的核心序列的活性和特异性(图32E)。每个重组酶对在位置6、位置5和位置4处的腺嘌呤或胸腺嘧啶表现出>1000倍的优先性,并且在位置10、位置9、位置8和位置7处显示无碱基优先性(图38)。这些结果表明,DNA结合臂的诱变没有影响重组酶特异性。不可能选择能够容忍在位置3和位置2处的AA、AT或TA取代的Gin变体。这一结果的一种可能性是包含>4个连续的A-T bp的DNA靶可能会表现出干扰重组酶结合和/或催化的弯曲DNA构象。

[0300] 工程化ZFR以重组用户定义的序列

[0301] 对由重新工程化的催化结构域组成的ZFR是否能够重组预先确定的序列进行了研究。为了检验这种可能性,使用44-bp共有重组位点在人类基因组(GRCh37主参照组)中搜索潜在的ZFR靶位点,所述44-bp共有重组位点被预测为在随机DNA中每400,000bp会发生大约一次(图4A)。这种来源于选定的Gin变体的核心序列分布的ZFR共有靶位点包括大约 7×10^8 (可能的 1.0955×10^{12} 中的)独特的20-bp核心组合和模块化锌指结构域的保守选择,该核心组合预计被21种可能的催化结构域组合所容忍,该保守选择排除每个ZFRS内的5'-CNN-3'和5'-TNN-3'三联体。以ZFR特异性为主要决定因素进行选择,在8个人类染色体(Chr.1、Chr.2、Chr.4、Chr.6、Chr.7、Chr.11、Chr.13和Chr.X)的非蛋白编码基因座鉴定了18种可能的ZFR靶位点。平均来说,每20-bp核心显示出与天然Gin催化结构域所识别的核心序列~46%的序列同一性(图33B)。每个相应的ZFR通过模块化组装来构建(见材料和方法)。

[0302] 为了确定各ZFR对是否可以重组其预期的DNA靶,进行了瞬时报道测定,其使ZFR介导的重组与减少的荧光素酶表达相关联(图33A和图39)。为实现此目的,将ZFR靶位点引入SV40启动子的上游和下游,该SV40启动子驱动荧光素酶报告基因的表达。用每个ZFR对的表达载体及其相应的报告质粒对人胚胎肾(HEK)293T细胞进行共转染。转染后48小时,测定荧光素酶表达。18个ZFR对分析中,使荧光素酶表达减少>75倍的占38%(18个中的7个)以及使荧光素酶表达减少>140倍的占22%(18个中的4个)(图33B)。相比之下,被设计以靶向天然Gin催化结构域所识别的核心序列的ZFR阳性对照GinC4使荧光素酶表达减少107倍。总的来说,我们发现,50%(18个中的9个)的所评估的ZFR对使荧光素酶表达减少至少20倍。重要的

是,实际上,在细菌细胞中显示出显著活性的每一催化结构域(>20%重组)被成功地用于重组哺乳动物细胞中的至少一种天然存在的序列。

[0303] 为了评价ZFR特异性,利用九个最活跃ZFR的表达质粒和各自的非同源报告质粒对分离的HEK293T细胞进行共转染。每个ZFR对对其预期的DNA靶表现出高特异性并且77% (9个中的7个)的所评估的ZFR呈显出与阳性对照GinC4的重组特异性几乎相同的整体重组的特异性(图4C)。为了确定减少的荧光素酶表达是预期的ZFR异二聚体的产物而不是重组态ZFR同二聚体的副产物,对每个ZFR单体对重组的贡献进行了测定。ZFR 1“左”单体与其相应的报告质粒的共转染导致荧光素酶表达的中度减少(对重组的总贡献:~22%),但是绝大多数的(18个中的16个)个别ZFR单体没有显著有助于重组(<10%重组),以及许多(18个中的7个)显示出没有活性(图39)。总之,这些研究表明,ZFR可被工程化以高度特异性重组用户定义的序列。

[0304] 工程化的ZFR介导靶向整合到人类基因组中。

[0305] 接下来对ZFR是否可以将DNA整合到在人体细胞中的内源性基因座进行了评估。为了实现此目的,在SV40启动子的控制下,用ZFR表达载体和相应的DNA供体质粒对HEK293细胞进行共转染,该DNA供体质粒包含特定的ZFR靶位点和嘌呤霉素耐药性基因。对于这种分析,使用了ZFR对1、ZFR对2和ZFR对3,这些ZFR对被设计以分别靶向人类染色体4、X和4上的非蛋白编码基因座(图34A)。在转染后2天,用含有嘌呤霉素的培养基孵育细胞并且通过确定嘌呤霉素耐药性(puro^R)菌落数来测定整合效率。我们发现:(i)供体质粒和相应的ZFR对的共转染与仅用供体质粒的转染相比导致>12倍的puro^R菌落的增加,并且(ii)用两个ZFR共转染与用个别ZFR单体转染相比导致6至9倍的puro^R菌落的增加(图34B)。为了评估ZFR对是否正确地靶向整合,从puro^R群中分离出基因组DNA,并通过PCR扩增每个靶向位点。在由这些ZFR对靶向的每个位点观察到对应于整合为正向和/或反向取向的PCR产物(图34C)。接着,为了确定ZFR介导的整合的整体特异性,从克隆的细胞群中分离出基因组DNA,并通过PCR评价质粒插入。该分析显示,对ZFR对1、ZFR对2和ZFR对3分别为8.3% (12个克隆中的1个)、14.2% (35个克隆中的5个)以及9.1% (11个克隆中管的1个)的靶向效率(图S6)。每个PCR产物的序列分析证实了ZFR介导的整合(图34D)。总之,这些结果表明,ZFR可以被设计以精确地将DNA整合至内源基因座。

[0306] 最后,应该指出的是,发现ZFR-1“左”单体靶向整合入ZFR-1基因座(图34C)。这个与上述荧光素酶报告研究(图39)相一致的结果表明,重组态ZFR同二聚体具有介导脱靶整合的能力。优化的异源二聚体ZFR架构的未来发展和脱靶整合的综合评价应该引起表现更高的靶向效率的ZFR的设计。

[0307] 本文表明,ZFR可以被设计以高特异性重组用户定义的序列,并且ZFR可以将DNA整合入人类细胞中预先确定的内源基因座。通过结合底物特异性分析和定向进化,消除了由ZFR催化结构域施加的几乎所有的序列要求。使用45个预先选择的锌指模块的存档,估计ZFR可以被设计以识别 $>1 \times 10^{22}$ 独特的44-bp DNA序列,其对应于随机序列中每4000bp大约一个潜在的ZFR靶位点。通过选择构建的定制的锌指结构域将进一步延伸靶。本文所述的重新工程化的催化结构域将与最近描述的TAL效应子重组酶相容。这项工作表明生成具有定制特异性的ZFR的可行性,并说明ZFR在包括基因组工程、合成生物学和基因治疗的广泛的应用中的潜在应用。

[0308] 虽然已经参照上述实施例描述了本发明,应当理解的是,修改和变化都包含在本发明的精神和范围之内。因此,本发明仅由以下权利要求书限定。

序列表

<110> 斯科克利普斯研究院
 <120> 具有靶向结合特异性的嵌合多肽
 <130> SCRIP2070-2W0
 <150> US 61/818,364
 <151> 2013-05-01
 <150> US 61/753,763
 <151> 2013-01-17
 <150> US 61/696,689
 <151> 2012-09-04
 <160> 369
 <170> PatentIn version 3.5
 <210> 1
 <211> 864
 <212> DNA
 <213> 人工序列

<220>
 <223> 合成构建体

[0001]

<400> 1
 atggatccca ttcttctcgc caccgcaagt cctgccccgc agcttctgcc cggaccceaa 60
 ccgcataggg ttacagccgc tgcagatcgg gggggggctc cgcctctggt cgcccctgt 120
 gatggcttgc ccctcggcg gacgatgtcc cggacccgge tgccatctcc ccctgcgccc 180
 tgcctctggt tctcggcggg cagcttcagc gatctgtccc gtcagttcga tccgtcgttt 240
 ctltgatacat cgttcttga ttgatgcct gccgtcggca cgcctcctac agcggttgc 300
 ccagcagagt gcgatgaggt gcaatcgggt ctgcgtgcag ccgatgacc gccaccacc 360
 gtgcgtgtcg ctgtcactgc ggcgcggcgc ccgcgcgcca agccggcccc gcgacggcgt 420
 gggcgcaac cctccgacgc ttccgcggcc gcgcaggtgg atctacgcac gctcggtac 480
 agtcagcagc agcaagagaa gatacaaccg aaggctcggt cgacagtggt gcagaccac 540
 gaggcactgg tgggccatgg gtttacacac gcgcacatcg ttgcgtcag ccaacaccg 600
 gcagcgtag ggaccgttgc tgtcacgtat caggacataa tcaggcggtt gccagaggcg 660
 acacacgaag acatcgttgg cgtcggcaaa cagtgggtcc gcgcacgcgc tctggaggcc 720
 ttgctcacgg agcgggggga gttgagaggt ccgccgttac agttggacac aggcccaactt 780
 ctcaagattg caaacgtgg cggcgtgacc gcagtgaggg cagtgcattc atggcgcaat 840
 gcaactgacgg gtgccccctt gaac 864

<210> 2
 <211> 288
 <212> PRT
 <213> 人工序列

<220>
 <223> 合成构建体

<400> 2

Met Asp Pro Ile Arg Ser Arg Thr Pro Ser Pro Ala Arg Glu Leu Leu
 1 5 10 15

Pro Gly Pro Gln Pro Asp Arg Val Gln Pro Thr Ala Asp Arg Gly Gly

	20	25	30
	Ala Pro Pro Ala Gly Gly Pro Leu Asp Gly Leu Pro Ala Arg Arg Thr 35 40 45		
	Met Ser Arg Thr Arg Leu Pro Ser Pro Pro Ala Pro Ser Pro Ala Phe 50 55 60		
	Ser Ala Gly Ser Phe Ser Asp Leu Leu Arg Gln Phe Asp Pro Ser Leu 65 70 75 80		
	Leu Asp Thr Ser Leu Leu Asp Ser Met Pro Ala Val Gly Thr Pro His 85 90 95		
	Thr Ala Ala Ala Pro Ala Glu Cys Asp Glu Val Gln Ser Gly Leu Arg 100 105 110		
	Ala Ala Asp Asp Pro Pro Pro Thr Val Arg Val Ala Val Thr Ala Ala 115 120 125		
	Arg Pro Pro Arg Ala Lys Pro Ala Pro Arg Arg Arg Ala Ala Gln Pro 130 135 140		
	Ser Asp Ala Ser Pro Ala Ala Gln Val Asp Leu Arg Thr Leu Gly Tyr 145 150 155 160		
[0002]	Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg Ser Thr Val 165 170 175		
	Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr His Ala His 180 185 190		
	Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr Val Ala Val 195 200 205		
	Thr Tyr Gln Asp Ile Ile Arg Ala Leu Pro Glu Ala Thr His Glu Asp 210 215 220		
	Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala Leu Glu Ala 225 230 235 240		
	Leu Leu Thr Glu Ala Gly Glu Leu Arg Gly Pro Pro Leu Gln Leu Asp 245 250 255		
	Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val Thr Ala Val 260 265 270		
	Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala Pro Leu Asn 275 280 285		
	<210> 3		
	<211> 11		
	<212> PRT		
	<213> 人工序列		
	<220>		
	<223> 合成构建体		

<220>
<221> MISC_特征
<222> (4)..(4)
<223> Xaa 是 Y, S 或 R

<220>
<221> MISC_特征
<222> (5)..(5)
<223> Xaa 是 R 或 G

<220>
<221> MISC_特征
<222> (6)..(6)
<223> Xaa 是 R, H, A, N 或 T

<400> 3

Val Gly Lys Xaa Xaa Xaa Gly Ala Arg Ala Leu
1 5 10

<210> 4
<211> 10
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 4

Val Gly Lys Tyr Arg Gly Ala Arg Ala Leu
1 5 10

[0003]
<210> 5
<211> 11
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 5

Val Gly Lys Ser Arg Ser Gly Ala Arg Ala Leu
1 5 10

<210> 6
<211> 10
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 6

Val Gly Lys Tyr His Gly Ala Arg Ala Leu
1 5 10

<210> 7
<211> 11
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 7

Val Gly Lys Arg Gly Ala Gly Ala Arg Ala Leu
1 5 10

<210> 8
<211> 13
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<220>
<221> MISC_特征
<222> (7)..(7)
<223> Xaa 是 Y, S 或 R

<220>
<221> MISC_特征
<222> (8)..(8)
<223> Xaa 是 W 或 G

<220>
<221> MISC_特征
<222> (9)..(9)
<223> Xaa 是 R, H, A, N 或 T

<400> 8

Ile Val Asp Ile Ala Lys Xaa Xaa Xaa Gly Asp Leu Ala
1 5 10

<210> 9
<211> 13
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 9

Ile Val Asp Ile Ala Arg Gln Trp Ser Gly Asp Leu Ala
1 5 10

<210> 10
<211> 12
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 10

Ile Val Asp Ile Ala Arg Tyr Arg Gly Asp Leu Ala
1 5 10

<210> 11
<211> 13
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 11

Ile Val Asp Ile Ala Arg Ser Arg Ser Gly Asp Leu Ala
1 5 10

<210> 12
<211> 12
<212> PRT

[0004]

[0005]

<213> 人工序列

<220>

<223> 合成构建体

<400> 12

Ile Val Asp Ile Ala Arg Tyr His Gly Asp Leu Ala
1 5 10

<210> 13

<211> 13

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 13

Ile Val Asp Ile Ala Arg Arg Gly Ala Gly Asp Leu Ala
1 5 10

<210> 14

<211> 9

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<220>

<221> misc_特征

<222> (4)..(6)

<223> Xaa 可以是任何天然氨基酸

<400> 14

Val Gly Lys Xaa Xaa Xaa Gly Ala Arg
1 5

<210> 15

<211> 12

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<220>

<221> misc_特征

<222> (5)..(8)

<223> Xaa 可以是任何天然氨基酸

<400> 15

Val Asp Ile Ala Xaa Xaa Xaa Gly Asp Leu Ala
1 5 10

<210> 16

<211> 48

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 16

Asn Ile His Gly Asn Ile Asn Ile Asn Ser His Asp Asn Asn His Asp

1	5	10	15
His Asp His Asp Asn Ser Asn Asn His Asp His Asp Asn Ser Asn Ser	20	25	30
Asn Asn Asn Asn Asn Ile Asn Gly Asn Asn Asn Ile Asn Asn Ser Asn	35	40	45
<210> 17			
<211> 26			
<212> DNA			
<213> 人工序列			
<220>			
<223> 合成构建体			
<400> 17			
ataaaccccc tccaaccagg tgctaa			26
<210> 18			
<211> 52			
<212> PRT			
<213> 人工序列			
<220>			
<223> 合成构建体			
<400> 18			
Asn Ile Asn Gly Asn Ile Asn Ile Asn Ile His Asp His Asp His Asp	1	5	10
			15
[0006] His Asp His Asp Asn Gly His Asp His Asp Asn Ile Asn Ile His Asp	20	25	30
His Asp Asn Ile Asn Asn Asn Asn Asn Gly Asn Asn His Asp Asn Gly	35	40	45
Asn Ile Asn Ile	50		
<210> 19			
<211> 1446			
<212> PRT			
<213> 人工序列			
<220>			
<223> 合成构建体			
<400> 19			
Met Asp Pro Ile Arg Ser Arg Thr Pro Ser Pro Ala Arg Glu Leu Leu	1	5	10
			15
Pro Gly Pro Gln Pro Asp Arg Val Gln Pro Thr Ala Asp Arg Gly Gly	20	25	30
Ala Pro Pro Ala Gly Gly Pro Leu Asp Gly Leu Pro Ala Arg Arg Thr	35	40	45
Met Ser Arg Thr Arg Leu Pro Ser Pro Pro Ala Pro Ser Pro Ala Phe	50	55	60
Ser Ala Gly Ser Phe Ser Asp Leu Leu Arg Gln Phe Asp Pro Ser Leu			

65	70	75	80
Leu Asp Thr Ser	Leu Asp Ser Met	Pro Ala Val Gly Thr	Pro His
	85	90	95
Thr Ala Ala Ala	Pro Ala Glu Cys Asp	Glu Val Gln Ser Gly	Leu Arg
	100	105	110
Ala Ala Asp Asp	Pro Pro Pro Thr	Val Arg Val Ala Val	Thr Ala Ala
	115	120	125
Arg Pro Pro Arg	Ala Lys Pro Ala Pro	Arg Arg Arg Ala Ala	Gln Pro
	130	135	140
Ser Asp Ala Ser	Pro Ala Ala Gln Val	Asp Leu Arg Thr	Leu Gly Tyr
	145	150	155
Ser Gln Gln Gln	Gln Glu Lys Ile Lys	Pro Lys Val Arg Ser	Thr Val
	165	170	175
Ala Gln His His	Glu Ala Leu Val Gly	His Gly Phe Thr	His Ala His
	180	185	190
Ile Val Ala Leu	Ser Gln His Pro Ala	Ala Leu Gly Thr	Val Ala Val
	195	200	205
Thr Tyr Gln Asp	Ile Ile Arg Ala Leu	Pro Glu Ala Thr	His Glu Asp
	210	215	220
Ile Val Gly Val	Gly Lys Gln Trp Ser	Gly Ala Arg Ala	Leu Glu Ala
	225	230	235
Leu Leu Thr Glu	Ala Gly Glu Leu Arg	Gly Pro Pro Leu	Gln Leu Asp
	245	250	255
Thr Gly Gln Leu	Leu Lys Ile Ala Lys	Arg Gly Gly Val	Thr Ala Val
	260	265	270
Glu Ala Val His	Ala Trp Arg Asn Ala	Leu Thr Gly Ala	Pro Leu Asn
	275	280	285
Leu Thr Pro Asp	Gln Val Val Ala Ile	Ala Ser Asn Ile	Gly Gly Lys
	290	295	300
Gln Ala Leu Glu	Thr Val Gln Arg Leu	Leu Pro Val Leu	Cys Gln Asp
	305	310	315
His Gly Leu Thr	Pro Asp Gln Val Val	Ala Ile Ala Ser	His Gly Gly
	325	330	335
Gly Lys Gln Ala	Leu Glu Thr Val Gln	Arg Leu Leu Pro	Val Leu Cys
	340	345	350
Gln Asp His Gly	Leu Thr Pro Asp Gln	Val Val Ala Ile	Ala Ser Asn
	355	360	365
Ile Gly Gly Lys	Gln Ala Leu Glu Thr	Val Gln Arg Leu	Leu Pro Val

[0007]

	370	375	380
	Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala		
	385	390	395 400
	Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu		
		405	410 415
	Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Ala Gln Val Val Ala		
		420	425 430
	Ile Ala Ser Asn Ser Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg		
		435	440 445
	Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val		
		450	455 460
	Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Leu		
		465	470 475 480
	Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp		
		485	490 495
	Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu		
		500	505 510
	Thr Leu Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr		
		515	520 525
[0008]	Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala		
		530	535 540
	Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly		
		545	550 555 560
	Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys		
		565	570 575
	Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp		
		580	585 590
	His Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser His Asp Gly		
		595	600 605
	Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys		
		610	615 620
	Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn		
		625	630 635 640
	Ser Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val		
		645	650 655
	Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala		
		660	665 670
	Ser Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro		

675	680	685
Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile 690 695 700		
Ala Ser Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu 705 710 715 720		
Pro Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr 725 730 735		
Gln Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala 740 745 750		
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly 755 760 765		
Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys 770 775 780		
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp 785 790 795 800		
His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ser Gly 805 810 815		
Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys 820 825 830		
Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn 835 840 845		
Ser Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val 850 855 860		
Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala 865 870 875 880		
Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu 885 890 895		
Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala 900 905 910		
Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg 915 920 925		
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Ala Gln Val 930 935 940		
Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val 945 950 955 960		
Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Leu Asp 965 970 975		
Gln Val Val Ala Ile Ala Ser Asn Gly Gly Ser Lys Gln Ala Leu Glu		

[0009]

[0010]

980	985	990
Thr Val Gln Arg Leu Leu Pro Val	Leu Cys Gln Asp His	Gly Leu Thr
995	1000	1005
Pro Asp Gln Val Val Ala Ile	Ala Asn Asn Asn Gly	Gly Lys Gln
1010	1015	1020
Ala Leu Glu Thr Val Gln Arg	Leu Leu Pro Val Leu	Cys Gln Asp
1025	1030	1035
His Gly Leu Thr Pro Asp Gln	Val Val Ala Ile Ala	Ser Asn Ile
1040	1045	1050
Gly Gly Lys Gln Ala Leu Glu	Thr Val Gln Arg Leu	Leu Pro Val
1055	1060	1065
Leu Cys Gln Asp His Gly Leu	Thr Leu Asp Gln Val	Val Ala Ile
1070	1075	1080
Ala Ser Asn Gly Gly Lys Gln	Ala Leu Glu Thr Val	Gln Arg Leu
1085	1090	1095
Leu Pro Val Leu Cys Gln Asp	His Gly Leu Thr Pro	Asn Gln Val
1100	1105	1110
Val Ala Ile Ala Ser Asn Ser	Gly Gly Lys Gln Ala	Leu Glu Thr
1115	1120	1125
Val Gln Arg Leu Leu Pro Val	Leu Cys Gln Asp His	Gly Leu Thr
1130	1135	1140
Pro Asn Gln Val Val Ala Ile	Ala Ser Asn Gly Gly	Lys Gln Ala
1145	1150	1155
Leu Glu Ser Ile Val Ala Gln	Leu Ser Arg Pro Asp	Pro Ala Leu
1160	1165	1170
Ala Ala Leu Thr Asn Asp His	Leu Val Ala Leu Ala	Cys Leu Gly
1175	1180	1185
Gly Arg Pro Ala Leu Asp Ala	Val Lys Lys Gly Leu	Pro His Ala
1190	1195	1200
Pro Glu Leu Ile Arg Arg Ile	Asn Arg Arg Ile Pro	Glu Arg Thr
1205	1210	1215
Ser His Arg Val Pro Asp Leu	Ala His Val Val Arg	Val Leu Gly
1220	1225	1230
Phe Phe Gln Ser His Ser His	Pro Ala Gln Ala Phe	Asp Asp Ala
1235	1240	1245
Met Thr Gln Phe Glu Met Ser	Arg His Gly Leu Val	Gln Leu Phe
1250	1255	1260
Arg Arg Val Gly Val Thr Glu	Phe Glu Ala Arg Tyr	Gly Thr Leu

[0011]

1265	1270	1275
Pro Pro Ala Ser Gln Arg Trp Asp Arg Ile Leu Gln Ala Ser Gly		
1280	1285	1290
Met Lys Arg Ala Lys Pro Ser Pro Thr Ser Ala Gln Thr Pro Asp		
1295	1300	1305
Gln Ala Ser Leu His Ala Phe Ala Asp Ser Leu Glu Arg Asp Leu		
1310	1315	1320
Asp Ala Pro Ser Pro Met His Glu Gly Asp Gln Thr Arg Ala Ser		
1325	1330	1335
Ser Arg Lys Arg Ser Arg Ser Asp Arg Ala Val Thr Gly Pro Ser		
1340	1345	1350
Thr Gln Gln Ser Phe Glu Val Arg Val Pro Glu Gln Gln Asp Ala		
1355	1360	1365
Leu His Leu Pro Leu Ser Trp Arg Val Lys Arg Pro Arg Thr Arg		
1370	1375	1380
Ile Gly Gly Gly Leu Pro Asp Pro Gly Thr Pro Ile Ala Ala Asp		
1385	1390	1395
Leu Ala Ala Ser Ser Thr Val Met Trp Glu Gln Asp Ala Ala Pro		
1400	1405	1410
Phe Ala Gly Ala Ala Asp Asp Phe Pro Ala Phe Asn Glu Glu Glu		
1415	1420	1425
Leu Ala Trp Leu Met Glu Leu Leu Pro Gln Ser Gly Ser Val Gly		
1430	1435	1440
Gly Thr Ile		
1445		

<210> 20

<211> 4341

<212> DNA

<213> 人工序列

<220>

<223> 合成构建体

<400> 20

atggatecca ttctgttcgag cagcccaagt cctgcccgcg agcttctgcc cggaccccaa 60

cgggataggg ttccagccgac tgcagatcgg gggggggctc cgcctgtctg cggccccctg 120

gatggettgc ccctcggcg gacgatgtcc cggaccgcgc tgccatctcc ccctgcgcc 180

tcgctgcgt tctcggcggg cagcttcagc gatctgtccc gtcagttega tccgtcgctt 240

cttgatacat cgtttcttga ttgatgcc tgcctcgcca ccgcgcatac agcggtgcc 300

ccagcagagt gcgatgaggt gcaatcggtt ctgcgtgcag ccgatgaccc gccaccacc 360

gtgcgtgtcg ctgtcactgc ggcgcggcgc ccgcgcgcca agccggcccc gcgacggcgt 420

gcggcgcaac cctccgacgc ttccgcggcc gcgcaggtag atctacgcac gctcggtac 480

[0012]

```

agtcagcagc agcaagagaa gatcaaacctg aaggtgcgtt cgacagtggc gcagcaccac 540
gaggcactgg tgggcatgg gtttacacac gcgcacatcg ttgcgtcag ccaacacccg 600
gcagcgtagg ggacgttgc tgtcacgtat caggacataa tcagggcgtt gccagaggcg 660
acacaagaag acatcgttgg cgtcgcaaaa cagtggctcg gcgcacgcgc tctggaggcc 720
ttgtcacagg agcggggga gttagagagt ccccggttac agttggacac aggccaactt 780
ctcaagattg caaacctgg cgcgctgacc gcagtggagg cagtgcctgc atggcgcaat 840
gcactgacgg gtccccct gaacctgacc ccggaccaag tggtagccat cgcagcaat 900
attggcgga agcaggcgct ggagacgga cagcgctgt tgcgggtgt gtgccaggac 960
catggctga cccggacca ggtcgtggcc atgccagcc atggcgcg ccagcaggcg 1020
ctggagacgg tgcaggcgt gttagcgtg ctgtgccagg accatggcct gacccggac 1080
caggtggtgg ccacgccag caatatggc ggcaagcagg cgttagagac ggtgcagcg 1140
ctgttgccgg tctgtgcca ggccatggc ctgaccccg accaggtct ggccatcgcc 1200
agcaafattg gcggcaaga ggcgctggag acggtgcagc ggtgtgtgc ggtgctgtgc 1260
caggaccatg gctgacccc ggccagggt gtggccatg ccagcaatag tggcggaag 1320
caggcgctgg agacggtgca gggctgttgc cgggtctgt gccaggacca tggcctgacc 1380
ccggaccaag tctggccat cgcagcccac gatggcgga agcaggcgct ggagacgtg 1440
cagcgctgt tgcgggtgt gtgccaggac catggcctga cccggacca ggtcgtggcc 1500
atcgccaaca ataacggcg ccagcaggcg ctggagacgc tgcagcgct gttagcgtg 1560
ctgtgccagg accatggcct gacccggac caagtgtgt ccacgccag ccacgatggc 1620
ggcaagcagg cgtgagac ggtagcggc ctgttgcgg tctgtgcca ggaccatggc 1680
ctgaccccg accaggtgt ggccatcgcc agccacgat gcgcaagca ggcgtggag 1740
acggtgcagc ggtgtgtgc ggtgctgtgc caggaccatg gctgacccc ggccaagtg 1800
gtggccatg ccagccagca tggcggaag caggcgctg agacggtgca gcggctgtg 1860
ccggtctgt gccaggacca tggcctgacc ccggaccagg tggtagccat cgcagcaat 1920
agcgcgga agcaggcgct ggagacgga cagcgctgt tgcgggtgt gtgccaggac 1980
catggaatga cccggacca ggtcgtggcc atgccagca atggcgcaa gcaggcgctg 2040
gagacggtac agcggtgtt gccgtgctg tgcaggacc atggcctgac cccggaccag 2100
gtcgtggcca tgcagcaaa tggcggaag caggcgctg agacggtgca gcggctgtt 2160
ccgtacagc ggtgtgtgc ggtgctgtgc caggaccatg gccgaccca ggaccagggtg 2220
gtggccatg ccagccagca tggcggaag caggcgctg agacggtgca gcggctgtg 2280
ccgtgctgt gccaggacca tggcctgacc ccggaccaag tggtagccat cgcagccac 2340
gatggcgga aacaggcgct ggagacggt cagcgctgt tgcgggtgt gtgccaggac 2400
catggcctga cccggacca ggtgtgtg ccacgccagca atagtggcg ccagcaggcg 2460
ctggagacgg tgcagcgct gttagcgtg ctgtgccagg accatggcct gacccggac 2520
caagtgtgt ccacgccag caatagtggc ggcaagcagg cgtgagac ggtagcgg 2580
ctgttgcgg tctgtgcca ggaccatggc ctgaccccg accaggtgt ggccatcgcc 2640
agcaataac gcggcaaga ggcgctggag acggtgcagc ggtgtgtgc ggtgctgtgc 2700
caggaccatg gctgacccc ggaccaggc gtggccatg caacaataa cggcggaag 2760

```

[0013]

```

caggcgctgg agacgggtgca gcggctgttg ccggtgctgt gccaggacca tggcctgacc 2820
ccggcgccagg tgggtggccat cgcagcaat atggcgcgca agcaggcgct ggagacggtg 2880
cagcgctgt tgcgggtgct gtgccaggac catggcctga ccctggacca ggtggtggcc 2940
attgccagca atggcggcag caaacaggcg ctagagacgg tgcagcgct gttgccggtg 3000
ctgtgccagg aecatggcct gaccccgac caagtgttg ccctcgcaa caataacggc 3060
ggcaagcagg cgtggagac ggtgcagcgg ctgttgccg tctgtgcca ggaccatggc 3120
ctgaccccg aecaggctgt ggccatgcc agcaatatg gcgcaagca ggcgctggag 3180
acggtgcagc ggtgttgcc ggtgctgtgc caggaccatg gctgacct ggaccaggtg 3240
gtggccatcg ccagcaatgg cggcangcag gcctggaga cggtgacgg gctgttgccg 3300
gtgctgtgcc aggaccatgg cctgacccc aaccaggttg tgccatcgc cagcaatagt 3360
ggcggaagc aggcctgga gacggtgcag cggctgttc cggctgttg ccaggaccat 3420
ggcctgacc cgaaccagg ggtggccatc gccagcaatg gcgcaagca ggcgctggag 3480
agcattgttg cccagttatc tgcctgat ccggcgttg ccgcttgac caacgaccac 3540
ctgtgcct tggcctgct cggcgacgt cctgccttg atgcagtaa aaagggattg 3600
ccgcacgcgc cgaattgat cagaagaatc atcgccga ttccgaacg cagtcctat 3660
cgggttccc acctcgcca cgtggttgc gtgttgggt tttccagag cactccac 3720
ccagcgcaag cattcgatga cccatgacg cagttaga tgacaggca cggcttgta 3780
cagctcttc gcagagtgg cgtaccgaa ttcaagccc gtagcgaac gctcccca 3840
gcctgcagc gttgggacc tatcctcag gcatcaggga tgaaaaggc caaacgtcc 3900
cctacttcag ctcaaacacc gcatcaggcg tcttgcatg cattcgcca ttgctggag 3960
cgtgacctg atgcgccag cccaatgac gaggagatc agacgcggc aagcagcgt 4020
aaacggtccc gatcgatcg tctgtcacc ggccctcca cacagcaatc ttctgaggtg 4080
cgcgttccc aacagcaaga tgcgctgat ttgccctca gctggagggt aaaacgccc 4140
cgtaccagga tggggggcg cctcccgat cctggtagc ccatcgctgc cgacctgca 4200
gcgtccagca cgtgatgtg ggaacaagt gcggccctc tcgcagggc agcgatgat 4260
ttcccgcat tcaacgaaga ggagctcga tggttgatg agctattgcc tcagtcagc 4320
tcagtcgag ggacgatctg a 4341

```

<210> 21
 <211> 1263
 <212> PRT
 <213> 人工序列

<220>
 <223> 合成构建体

<400> 21

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Gln Gln Ile Phe Glu
20 25 30

Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala
35 40 45

[0014]

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

	Trp	Arg	Asn	Ala	Leu	Thr	Gly	Ala	Pro	Leu	Asn	Leu	Thr	Pro	Asp	Gln	
			355					360						365			
	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	
		370					375					380					
	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	
	385					390					395					400	
	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Gly	Gly	Gly	Lys	Gln	Ala	Leu	
				405						410					415		
	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	
			420					425						430			
	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	
			435					440					445				
	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	
		450					455					460					
	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	
	465					470				475					480		
	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	
				485						490					495		
[0015]	Asp	His	Gly	Leu	Thr	Pro	Ala	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ser	
				500					505					510			
	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	
			515					520					525				
	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	
		530					535					540					
	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Leu	Gln	Arg	Leu	Leu	Pro	
	545					550					555				560		
	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	
				565						570					575		
	Ala	Asn	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Leu	Gln	Arg	Leu	
				580					585					590			
	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	
		595						600					605				
	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	
		610					615					620					
	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	
	625					630					635				640		
	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	
				645							650				655		

[0016]

Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	660	665	670	
Ala	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	675	680	685	
Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	690	695	700	
Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ser	Gly	Gly	Lys	Gln	705	710	715	720
Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	725	730	735	
Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Lys	740	745	750	
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	755	760	765	
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	770	775	780	
Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Gln	Arg	Leu	785	790	795	800
Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Gln	Asp	Gln	Val	Val	805	810	815	
Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	820	825	830	
Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	835	840	845	
Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	850	855	860	
Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	865	870	875	880
Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ser	Gly	Gly	Lys	Gln	Ala	Leu	885	890	895	
Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	900	905	910	
Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ser	Gly	Gly	Lys	Gln	915	920	925	
Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	930	935	940	
Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Asn	Gly	Gly	945	950	955	960

	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	
					965					970						975	
	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Asn	Asn	Asn	
				980					985						990		
	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	
			995					1000						1005			
	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Ala	Gln	Val	Val	Ala	Ile	Ala		
		1010					1015						1020				
	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu		
		1025					1030						1035				
	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Leu	Asp	Gln	Val		
		1040					1045						1050				
	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Ser	Lys	Gln	Ala	Leu	Glu	Thr		
		1055					1060						1065				
	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr		
		1070					1075						1080				
	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Asn	Asn	Asn	Gly	Gly	Lys	Gln		
		1085					1090						1095				
[0017]	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp		
		1100					1105						1110				
	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile		
		1115					1120						1125				
	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val		
		1130					1135						1140				
	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Leu	Asp	Gln	Val	Val	Ala	Ile		
		1145					1150						1155				
	Ala	Ser	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu		
		1160					1165						1170				
	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asn	Gln	Val		
		1175					1180						1185				
	Val	Ala	Ile	Ala	Ser	Asn	Ser	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr		
		1190					1195						1200				
	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr		
		1205					1210						1215				
	Pro	Asn	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Lys	Gln	Ala		
		1220					1225						1230				
	Leu	Glu	Ser	Ile	Val	Ala	Gln	Leu	Ser	Arg	Pro	Asp	Pro	Ala	Leu		
		1235					1240						1245				

Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys Leu Gly
1250 1255 1260

<210> 22
<211> 1250
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 22

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala
35 40 45

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

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

[0018]

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala
115 120 125

Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Arg Lys Ser Gly
130 135 140

Ser Gly Ser Pro Asp Ser Met Pro Ala Val Gly Thr Pro His Thr Ala
145 150 155 160

Ala Ala Pro Ala Glu Cys Asp Glu Val Gln Ser Gly Leu Arg Ala Ala
165 170 175

Asp Asp Pro Pro Pro Thr Val Arg Val Ala Val Thr Ala Ala Arg Pro
180 185 190

Pro Arg Ala Lys Pro Ala Pro Arg Arg Arg Ala Ala Gln Pro Ser Asp
195 200 205

Ala Ser Pro Ala Ala Gln Val Asp Leu Arg Thr Leu Gly Tyr Ser Gln
210 215 220

Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg Ser Thr Val Ala Gln
225 230 235 240

His His Glu Ala Leu Val Gly His Gly Phe Thr His Ala His Ile Val

[0019]

	245	250	255
Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr Val Ala Val Thr Tyr	260	265	270
Gln Asp Ile Ile Arg Ala Leu Pro Glu Ala Thr His Glu Asp Ile Val	275	280	285
Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala Leu Glu Ala Leu Leu	290	295	300
Thr Glu Ala Gly Glu Leu Arg Gly Pro Pro Leu Gln Leu Asp Thr Gly	305	310	315
Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val Thr Ala Val Glu Ala	325	330	335
Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala Pro Leu Asn Leu Thr	340	345	350
Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala	355	360	365
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly	370	375	380
Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Gly Gly Gly Lys	385	390	395
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp	405	410	415
His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly	420	425	430
Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys	435	440	445
Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn	450	455	460
Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val	465	470	475
Leu Cys Gln Asp His Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala	485	490	495
Ser Asn Ser Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu	500	505	510
Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala	515	520	525
Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Leu Gln Arg	530	535	540
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val			

[0020]

545	550	555	560
Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Leu	565	570	575
Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp	580	585	590
Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu	595	600	605
Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr	610	615	620
Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala	625	630	635
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly	645	650	655
Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys	660	665	670
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp	675	680	685
His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ser Gly	690	695	700
Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys	705	710	715
Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn	725	730	735
Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu	740	745	750
Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser	755	760	765
Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val	770	775	780
Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Gln Asp	785	790	795
Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu	805	810	815
Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr	820	825	830
Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala	835	840	845
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly			

[0021]

850	855	860
Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ser Gly Gly Lys		
865	870	875 880
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp		
	885	890 895
His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ser Gly		
	900	905 910
Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys		
	915	920 925
Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn		
	930	935 940
Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val		
	945	950 955 960
Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala		
	965	970 975
Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu		
	980	985 990
Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Ala Gln Val Val Ala		
	995	1000 1005
Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln		
	1010	1015 1020
Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Leu Asp		
	1025	1030 1035
Gln Val Val Ala Ile Ala Ser Asn Gly Gly Ser Lys Gln Ala Leu		
	1040	1045 1050
Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly		
	1055	1060 1065
Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly		
	1070	1075 1080
Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys		
	1085	1090 1095
Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser		
	1100	1105 1110
Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu		
	1115	1120 1125
Pro Val Leu Cys Gln Asp His Gly Leu Thr Leu Asp Gln Val Val		
	1130	1135 1140
Ala Ile Ala Ser Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln		

	1145	1150	1155
	Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asn		
	1160	1165	1170
	Gln Val Val Ala Ile Ala Ser Asn Ser Gly Gly Lys Gln Ala Leu		
	1175	1180	1185
	Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly		
	1190	1195	1200
	Leu Thr Pro Asn Gln Val Val Ala Ile Ala Ser Asn Gly Gly Lys		
	1205	1210	1215
	Gln Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp Pro		
	1220	1225	1230
	Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys		
	1235	1240	1245
	Leu Gly		
	1250		
	<210> 23		
	<211> 1216		
	<212> PRT		
	<213> 人工序列		
[0022]	<220>		
	<223> 合成构建体		
	<400> 23		
	Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp		
	1 5 10 15		
	Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu		
	20 25 30		
	Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala		
	35 40 45		
	Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp		
	50 55 60		
	Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu		
	65 70 75 80		
	Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr		
	85 90 95		
	Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala		
	100 105 110		
	Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala		
	115 120 125		
	Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Arg Lys Ser Gly		
	130 135 140		

	Ser	Gly	Ser	Thr	Val	Arg	Val	Ala	Val	Thr	Ala	Ala	Arg	Pro	Pro	Arg	
	145					150					155					160	
	Ala	Lys	Pro	Ala	Pro	Arg	Arg	Arg	Ala	Ala	Gln	Pro	Ser	Asp	Ala	Ser	
				165						170					175		
	Pro	Ala	Ala	Gln	Val	Asp	Leu	Arg	Thr	Leu	Gly	Tyr	Ser	Gln	Gln	Gln	
				180					185					190			
	Gln	Glu	Lys	Ile	Lys	Pro	Lys	Val	Arg	Ser	Thr	Val	Ala	Gln	His	His	
			195					200					205				
	Glu	Ala	Leu	Val	Gly	His	Gly	Phe	Thr	His	Ala	His	Ile	Val	Ala	Leu	
	210						215					220					
	Ser	Gln	His	Pro	Ala	Ala	Leu	Gly	Thr	Val	Ala	Val	Thr	Tyr	Gln	Asp	
	225				230						235					240	
	Ile	Ile	Arg	Ala	Leu	Pro	Glu	Ala	Thr	His	Glu	Asp	Ile	Val	Gly	Val	
				245						250					255		
	Gly	Lys	Gln	Trp	Ser	Gly	Ala	Arg	Ala	Leu	Glu	Ala	Leu	Leu	Thr	Glu	
			260					265						270			
	Ala	Gly	Glu	Leu	Arg	Gly	Pro	Pro	Leu	Gln	Leu	Asp	Thr	Gly	Gln	Leu	
		275					280					285					
[0023]	Leu	Lys	Ile	Ala	Lys	Arg	Gly	Gly	Val	Thr	Ala	Val	Glu	Ala	Val	His	
	290					295						300					
	Ala	Trp	Arg	Asn	Ala	Leu	Thr	Gly	Ala	Pro	Leu	Asn	Leu	Thr	Pro	Asp	
	305				310						315					320	
	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	
				325					330					335			
	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	
			340					345					350				
	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Gly	Gly	Gly	Lys	Gln	Ala	
		355				360					365						
	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	
	370					375					380						
	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	
	385				390					395						400	
	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	
			405						410					415			
	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	
			420					425					430				
	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	
		435					440					445					

[0024]

Gln Asp His Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn
 450 455 460
 Ser Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
 465 470 475 480
 Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala
 485 490 495
 Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Leu Gln Arg Leu Leu
 500 505 510
 Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala
 515 520 525
 Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Leu Gln Arg
 530 535 540
 Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val
 545 550 555 560
 Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val
 565 570 575
 Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp
 580 585 590
 Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu
 595 600 605
 Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr
 610 615 620
 Pro Ala Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala
 625 630 635 640
 Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly
 645 650 655
 Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ser Gly Gly Lys
 660 665 670
 Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp
 675 680 685
 His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly Gly
 690 695 700
 Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 705 710 715 720
 Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly
 725 730 735
 Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Gln Arg
 740 745 750

[0025]

Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly
 1055 1060 1065
 Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly
 1070 1075 1080
 Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys
 1085 1090 1095
 Gln Asp His Gly Leu Thr Leu Asp Gln Val Val Ala Ile Ala Ser
 1100 1105 1110
 Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 1115 1120 1125
 Val Leu Cys Gln Asp His Gly Leu Thr Pro Asn Gln Val Val Ala
 1130 1135 1140
 Ile Ala Ser Asn Ser Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 1145 1150 1155
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asn
 1160 1165 1170
 Gln Val Val Ala Ile Ala Ser Asn Gly Gly Lys Gln Ala Leu Glu
 1175 1180 1185
 Ser Ile Val Ala Gln Leu Ser Arg Pro Asp Pro Ala Leu Ala Ala
 1190 1195 1200
 Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys Leu Gly
 1205 1210 1215
 <210> 24
 <211> 1204
 <212> PRT
 <213> 人工序列
 <220>
 <223> 合成构建体
 <400> 24
 Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
 1 5 10 15
 Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
 20 25 30
 Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala
 35 40 45
 Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
 50 55 60
 Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
 65 70 75 80
 Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr

[0026]

	85	90	95
	Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala		
	100	105	110
	Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala		
	115	120	125
	Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Arg Lys Ser Gly		
	130	135	140
	Ser Gly Ser Thr Val Arg Val Ala Val Thr Ala Ala Arg Pro Pro His		
	145	150	155
	Ala Val Ala Gly Pro Ala Ala Gln Val Asp Leu Arg Thr Leu Gly Tyr		
	165	170	175
	Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg Ser Thr Val		
	180	185	190
	Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr His Ala His		
	195	200	205
	Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr Val Ala Val		
	210	215	220
	Thr Tyr Gln Asp Ile Ile Arg Ala Leu Pro Glu Ala Thr His Glu Asp		
	225	230	235
[0027]	Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala Leu Glu Ala		
	245	250	255
	Leu Leu Thr Glu Ala Gly Glu Leu Arg Gly Pro Pro Leu Gln Leu Asp		
	260	265	270
	Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val Thr Ala Val		
	275	280	285
	Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala Pro Leu Asn		
	290	295	300
	Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys		
	305	310	315
	Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp		
	325	330	335
	His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Gly Gly		
	340	345	350
	Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys		
	355	360	365
	Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn		
	370	375	380
	Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val		

[0028]

385	390	395	400
Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala	405	410	415
Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu	420	425	430
Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Ala Gln Val Val Ala	435	440	445
Ile Ala Ser Asn Ser Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg	450	455	460
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val	465	470	475
Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Leu	485	490	495
Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp	500	505	510
Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu	515	520	525
Thr Leu Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr	530	535	540
Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala	545	550	555
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly	565	570	575
Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys	580	585	590
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp	595	600	605
His Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser His Asp Gly	610	615	620
Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys	625	630	635
Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn	645	650	655
Ser Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val	660	665	670
Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala	675	680	685
Ser Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro			

690	695	700
Val Leu Cys Gln Asp 705	His Gly Leu Thr Pro 710	Asp Gln Val Val Ala Ile 720
Ala Ser Asn Gly 725	Gly Lys Gln Ala Leu Glu Thr 730	Val Gln Arg Leu Leu 735
Pro Val Gln Arg 740	Leu Leu Pro Val Leu Cys Gln Asp 745	His Gly Leu Thr 750
Gln Asp Gln Val Val 755	Ala Ile Ala Ser His Asp 760	Gly Gly Lys Gln Ala 765
Leu Glu Thr Val Gln Arg 770	Leu Leu Pro Val Leu Cys Gln Asp 775	His Gly 780
Leu Thr Pro Asp Gln 785	Val Val Ala Ile Ala Ser His Asp 790	Gly Gly Lys 800
Gln Ala Leu Glu Thr 805	Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp 810	
His Gly Leu Thr 820	Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ser Gly 825	
Gly Lys Gln Ala Leu Glu Thr 835	Val Gln Arg Leu Leu Pro Val Leu Cys 840	
Gln Asp His Gly Leu Thr 850	Pro Asp Gln Val Val Ala Ile Ala Ser Asn 855	
Ser Gly Gly Lys Gln Ala Leu Glu Thr 865	Val Gln Arg Leu Leu Pro Val 870	
Leu Cys Gln Asp His Gly Leu Thr 885	Pro Asp Gln Val Val Ala Ile Ala 890	
Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr 900	Val Gln Arg Leu Leu 905	
Pro Val Leu Cys Gln Asp His Gly Leu Thr 915	Pro Asp Gln Val Val Ala 920	
Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr 930	Val Gln Arg 935	
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr 945	Pro Ala Gln Val 950	
Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr 965	Val 970	
Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr 980	Leu Asp 985	
Gln Val Val Ala Ile Ala Ser Asn Gly Gly Ser Lys Gln Ala Leu Glu		

[0029]

	995	1000	1005
	Thr Val Gln Arg Leu Leu	Pro Val Leu Cys Gln Asp	His Gly Leu
	1010	1015	1020
	Thr Pro Asp Gln Val Val	Ala Ile Ala Asn Asn Asn	Gly Gly Lys
	1025	1030	1035
	Gln Ala Leu Glu Thr Val	Gln Arg Leu Leu Pro Val	Leu Cys Gln
	1040	1045	1050
	Asp His Gly Leu Thr Pro	Asp Gln Val Val Ala Ile	Ala Ser Asn
	1055	1060	1065
	Ile Gly Gly Lys Gln Ala	Leu Glu Thr Val Gln Arg	Leu Leu Pro
	1070	1075	1080
	Val Leu Cys Gln Asp His	Gly Leu Thr Leu Asp Gln	Val Val Ala
	1085	1090	1095
	Ile Ala Ser Asn Gly Gly	Lys Gln Ala Leu Glu Thr	Val Gln Arg
	1100	1105	1110
	Leu Leu Pro Val Leu Cys	Gln Asp His Gly Leu Thr	Pro Asn Gln
	1115	1120	1125
	Val Val Ala Ile Ala Ser	Asn Ser Gly Gly Lys Gln	Ala Leu Glu
	1130	1135	1140
[0030]	Thr Val Gln Arg Leu Leu	Pro Val Leu Cys Gln Asp	His Gly Leu
	1145	1150	1155
	Thr Pro Asn Gln Val Val	Ala Ile Ala Ser Asn Gly	Gly Lys Gln
	1160	1165	1170
	Ala Leu Glu Ser Ile Val	Ala Gln Leu Ser Arg Pro	Asp Pro Ala
	1175	1180	1185
	Leu Ala Ala Leu Thr Asn	Asp His Leu Val Ala Leu	Ala Cys Leu
	1190	1195	1200
	Gly		
	<210> 25		
	<211> 1190		
	<212> PRT		
	<213> 人工序列		
	<220>		
	<223> 合成构建体		
	<400> 25		
	Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp		
	1 5 10 15		
	Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu		
	20 25 30		

	Asp	Lys	Leu	Ser	Gly	Thr	Arg	Thr	Asp	Arg	Pro	Gly	Leu	Lys	Arg	Ala	
		35						40					45				
	Leu	Lys	Arg	Leu	Gln	Lys	Gly	Asp	Thr	Leu	Val	Val	Trp	Lys	Leu	Asp	
		50					55					60					
	Arg	Leu	Gly	Arg	Ser	Met	Lys	His	Leu	Ile	Ser	Leu	Val	Gly	Glu	Leu	
	65				70					75					80		
	Arg	Glu	Arg	Gly	Ile	Asn	Phe	Arg	Ser	Leu	Thr	Asp	Ser	Ile	Asp	Thr	
				85						90				95			
	Ser	Ser	Pro	Met	Gly	Arg	Phe	Phe	Phe	Tyr	Val	Met	Gly	Ala	Leu	Ala	
			100						105					110			
	Glu	Met	Glu	Arg	Glu	Leu	Ile	Ile	Glu	Arg	Thr	Met	Ala	Gly	Leu	Ala	
		115						120					125				
	Ala	Ala	Arg	Asn	Lys	Gly	Arg	Ile	Gly	Gly	Arg	Pro	Arg	Lys	Ser	Gly	
		130					135					140					
	Ser	Gly	Ser	Pro	Ala	Ser	Pro	Ala	Ala	Gln	Val	Asp	Leu	Arg	Thr	Leu	
	145				150						155					160	
	Gly	Tyr	Ser	Gln	Gln	Gln	Gln	Glu	Lys	Ile	Lys	Pro	Lys	Val	Arg	Ser	
				165						170					175		
[0031]	Thr	Val	Ala	Gln	His	His	Glu	Ala	Leu	Val	Gly	His	Gly	Phe	Thr	His	
			180						185					190			
	Ala	His	Ile	Val	Ala	Leu	Ser	Gln	His	Pro	Ala	Ala	Leu	Gly	Thr	Val	
		195						200					205				
	Ala	Val	Thr	Tyr	Gln	Asp	Ile	Ile	Arg	Ala	Leu	Pro	Glu	Ala	Thr	His	
		210					215					220					
	Glu	Asp	Ile	Val	Gly	Val	Gly	Lys	Gln	Trp	Ser	Gly	Ala	Arg	Ala	Leu	
	225				230						235				240		
	Glu	Ala	Leu	Leu	Thr	Glu	Ala	Gly	Glu	Leu	Arg	Gly	Pro	Pro	Leu	Gln	
			245						250					255			
	Leu	Asp	Thr	Gly	Gln	Leu	Leu	Lys	Ile	Ala	Lys	Arg	Gly	Gly	Val	Thr	
		260						265						270			
	Ala	Val	Glu	Ala	Val	His	Ala	Trp	Arg	Asn	Ala	Leu	Thr	Gly	Ala	Pro	
		275						280					285				
	Leu	Asn	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	
		290					295					300					
	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	
	305				310						315				320		
	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	
				325					330					335			

	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	
				340					345						350		
	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	
			355					360					365				
	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	
		370					375						380				
	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	
	385					390					395					400	
	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	
				405						410						415	
	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Ala	Gln	Val	
			420						425						430		
	Val	Ala	Ile	Ala	Ser	Asn	Ser	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	
			435					440						445			
	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	
		450					455						460				
	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	
	465				470						475					480	
[0032]	Thr	Leu	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	
				485						490					495		
	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Asn	Asn	Asn	Gly	Gly	Lys	Gln	Ala	
			500						505					510			
	Leu	Glu	Thr	Leu	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	
		515					520						525				
	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	
		530				535					540						
	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	
	545				550					555					560		
	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	
			565					570						575			
	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	
			580					585						590			
	Gln	Asp	His	Gly	Leu	Thr	Pro	Ala	Gln	Val	Val	Ala	Ile	Ala	Ser	His	
		595						600					605				
	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	
		610					615					620					
	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	
	625					630					635					640	

	Ser	Asn	Ser	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	
				645					650						655		
	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	
			660						665					670			
	Ile	Ala	Ser	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	
			675					680						685			
	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	
		690					695					700					
	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	
	705					710					715					720	
	Leu	Leu	Pro	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	
				725					730						735		
	Leu	Thr	Gln	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	
			740						745						750		
	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	
		755						760					765				
	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	
		770					775					780					
[0033]	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	
	785					790					795					800	
	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	
				805						810					815		
	Ser	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	
				820					825						830		
	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	
		835						840					845				
	Ser	Asn	Ser	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	
		850					855						860				
	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	
	865					870					875					880	
	Ile	Ala	Ser	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	
				885					890						895		
	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	
			900						905						910		
	Val	Ala	Ile	Ala	Asn	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	
		915						920						925			
	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Ala	
		930					935						940				

Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu
945 950 955 960

Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr
965 970 975

Leu Asp Gln Val Val Ala Ile Ala Ser Asn Gly Gly Ser Lys Gln Ala
980 985 990

Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly
995 1000 1005

Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly
1010 1015 1020

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys
1025 1030 1035

Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser
1040 1045 1050

Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
1055 1060 1065

Pro Val Leu Cys Gln Asp His Gly Leu Thr Leu Asp Gln Val Val
1070 1075 1080

[0034]

Ala Ile Ala Ser Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
1085 1090 1095

Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asn
1100 1105 1110

Gln Val Val Ala Ile Ala Ser Asn Ser Gly Gly Lys Gln Ala Leu
1115 1120 1125

Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly
1130 1135 1140

Leu Thr Pro Asn Gln Val Val Ala Ile Ala Ser Asn Gly Gly Lys
1145 1150 1155

Gln Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp Pro
1160 1165 1170

Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys
1175 1180 1185

Leu Gly
1190

<210> 26

<211> 841

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 26

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala
35 40 45

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

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala
115 120 125

Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Arg Lys Ser Gly
130 135 140

[0035]

Ser Gly Ser Pro Ala Leu Arg Pro Pro Arg Ala Lys Pro Ala Pro Arg
145 150 155 160

Arg Arg Ala Ala Gln Pro Ser Asp Ala Ser Pro Ala Ala Gln Val Asp
165 170 175

Leu Arg Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro
180 185 190

Lys Val Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His
195 200 205

Gly Phe Thr His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala
210 215 220

Leu Gly Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala Leu Pro
225 230 235 240

Glu Ala Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp Ser Gly
245 250 255

Ala Arg Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly
260 265 270

Pro Pro Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg
275 280 285

Gly Gly Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu

290	295	300
Thr Gly Ala Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala 305 310 315 320		
Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu 325 330 335		
Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala 340 345 350		
Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg 355 360 365		
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val 370 375 380		
Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val 385 390 395 400		
Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp 405 410 415		
Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu 420 425 430		
Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr 435 440 445		
Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala 450 455 460		
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly 465 470 475 480		
Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys 485 490 495		
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp 500 505 510		
His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly 515 520 525		
Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys 530 535 540		
Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His 545 550 555 560		
Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val 565 570 575		
Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala 580 585 590		
Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu		

[0036]

595	600	605
Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala 610	615	620
Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg 625	630	635 640
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val 645	650	655
Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val 660	665	670
Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp 675	680	685
Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu 690	695	700
Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr 705	710	715 720
Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala 725	730	735
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly 740	745	750
Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys 755	760	765
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp 770	775	780
His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly 785	790	795 800
Gly Lys Gln Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp 805	810	815
Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys 820	825	830
Leu Gly Pro Lys Lys Lys Arg Lys Val 835	840	

[0037]

<210> 27

<211> 2523

<212> DNA

<213> 人工序列

<220>

<223> 合成构建体

<400> 27

atgctgattg gctatgtaag ggatacaaca aatgaccaga atacagacct gcaacgaaac 60

gcctcttggtt gtgcaggatg tgaacaaata ttgaagata aattaagcgg aacaaggaca 120

[0038]	gaccgaccgg gattaaaacg cgetttaaag cgccttcaaa aaggtgacac actggttgte	180
	tggaaactgg atcgectcgg gcgaagcatg aaacatttga ttctctcgt aggggaatta	240
	cgagagcgag ggattaatit tcgcagtcit acigacagta ttgatacgtc atctccaatg	300
	ggcggttttt tcttctacgt tatgggtgcc ctggetgaaa tggaaacgaga actaattatc	360
	gagcgaacga tggctggact tgcctgccgc agaaataaag gccgtattgg aggtcgcccg	420
	cgtaaatcgg ggtctggatc ccccgcgagg ccgcccgcgc ccaagccggc cccgcgacgg	480
	cgtctgcgc aacctccga cgtctgcgc gccgcgcagg tggatctacg cagctcggc	540
	tacagtcagc agcagcaaga gaagatcaaa ccgaaggcgc gttcgacagt ggcgacgac	600
	cacgaggcac tgggtggcca tgggtttaca cagcgcaaca tegtgcgt cagccaacac	660
	ccggcagcgt tagggaccgt cgtctgcgc tatcagcaca taatcacggc gttgccagag	720
	gcgacacacg aagacatcgt tggcgtcggc aaacagtggt ccggcgacgc cgcctggag	780
	gccttgctca cggatcgagg ggagttgaga ggicgcgcgt tacagttgga cacaggccaa	840
	cttgtgaaga ttgcaaaacg tggcggcgtg accgcaatgg aggcagtgca tgcctcgcgc	900
	aatgcactga cgggtgcccc cctggagctg actccggacc aagtgtgtgc tatcgccagc	960
	aacattggcg gcaagcaagc gctcgaaacg gtgcagcggc tgttgcgggt gctgtgccag	1020
	gacctggcc tgaactcgga ccaagtgtg gctatcgcca gcaacggtgg cggcaagcaa	1080
	gcgctgaaa cgggtcagcg gctgttgcgc gtgctgtgcc aggaccatgg cctgaactcg	1140
	gaccaagtgg tggctatcgc cagcaacatt ggccgcaagc aagcgctcga aacggtgcag	1200
	cggctgttgc cgggtctgtg ccaggaccat ggctgactc cggaccaagt ggtggctatc	1260
	gccagcaaca ttggcggcaa gcaagcgctc gaaacggtgc agcggtgtt gccgggtctg	1320
	tgccaggacc atggcctgac tccggaccaa gtgtggcta tcgccagcaa cattggcggc	1380
	aagcaagcgc tcgaaacggt gcagcggctg ttgccgtgc tggccagga ccatggcctg	1440
	actccggacc aagtggtggtc tatcgccagc cagatggcg gcaagcaagc gctcgaaacg	1500
	gtgcagcggc tgttgcgggt gctgtgccag gacctggcc tgaactcgga ccaagtgtg	1560
	gctatcgcca gccacgatgg cggcaagcaa gcgctgaaa cggtcagcg gctgttgcgc	1620
	gtgctgtgcc aggaccatgg cctgaactcg gaccaagtgg tggctatcgc cagccacgat	1680
	ggcggcaagc aagcgctcga aacggtgcag cggctgttgc cgggtctgtg ccaggaccat	1740
	ggctgactc cggaccaagt gttggctatc gcagccaagc atggcggcaa gcaagcgctc	1800
	gaaacggtgc agcggtgtt gccgggtctg tgccaggacc atggcctgac tccggaccaa	1860
	gtgtgtgcta tcgccagcaa ccatggcggc aagcaagcgc tcgaaacggt gcagcggctg	1920
	ttgccgtgct tggccagga ccatggcctg acccggacc aagtgtgtgc tatcgccagc	1980
	aacggtggcg gcaagcaagc gctcgaaacg gtgcagcggc tgttgcgggt gctgtgccag	2040
	gacctggcc tgaactcgga ccaagtgtg gctatcgcca gccacgatgg cggcaagcaa	2100
	gcgctgaaa cggtcagcg gctgttgcgc gtgctgtgcc aggaccatgg cctgaactcg	2160
	gaccaagtgg tggctatcgc cagccacgat ggccgcaagc aagcgctcga aacggtgcag	2220
	cggctgttgc cgggtctgtg ccaggaccat ggctgaccc cggaccaagt gttggctatc	2280
	gccagcaaca ttggcggcaa gcaagcgctc gaaacggtgc agcggtgtt gccgggtctg	2340
	tgccaggacc atggcctgac cccggaccaa gtgtggcta tcgccagcaa cattggcggc	2400

Leu Gly Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala Leu Pro
225 230 235 240

	Glu	Ala	Thr	His	Glu	Asp	Ile	Val	Gly	Val	Gly	Lys	Gln	Trp	Ser	Gly	
					245					250					255		
	Ala	Arg	Ala	Leu	Glu	Ala	Leu	Leu	Thr	Asp	Ala	Gly	Glu	Leu	Arg	Gly	
				260					265					270			
	Pro	Pro	Leu	Gln	Leu	Asp	Thr	Gly	Gln	Leu	Val	Lys	Ile	Ala	Lys	Arg	
			275					280					285				
	Gly	Gly	Val	Thr	Ala	Met	Glu	Ala	Val	His	Ala	Ser	Arg	Asn	Ala	Leu	
		290					295					300					
	Thr	Gly	Ala	Pro	Leu	Asn	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	
	305					310					315				320		
	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	
				325						330					335		
	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	
			340						345					350			
	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	
		355						360					365				
	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	
		370					375					380					
[0040]	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	
	385					390					395				400		
	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	
				405					410					415			
	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	
			420						425					430			
	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	
		435					440					445					
	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	
		450					455					460					
	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	
	465					470				475					480		
	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	
				485					490						495		
	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	
			500						505					510			
	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	
		515						520					525				
	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	
		530					535						540				

	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	
	545					550					555					560	
	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	
				565						570					575		
	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	
			580						585					590			
	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	
		595					600						605				
	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	
		610					615					620					
	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	
	625				630						635					640	
	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	
				645						650					655		
	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	
			660						665						670		
	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	
		675						680						685			
[0041]	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	
	690					695						700					
	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	
	705				710					715						720	
	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	
			725						730						735		
	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	
			740					745					750				
	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	
		755					760						765				
	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	
	770					775						780					
	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	
	785				790						795					800	
	Gly	Lys	Gln	Ala	Leu	Glu	Ser	Ile	Val	Ala	Gln	Leu	Ser	Arg	Pro	Asp	
			805						810						815		
	Pro	Ala	Leu	Ala	Ala	Leu	Thr	Asn	Asp	His	Leu	Val	Ala	Leu	Ala	Cys	
			820					825						830			
	Leu	Gly	Pro	Lys	Lys	Lys	Arg	Lys	Val								
		835						840									

<210> 29
<211> 134
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 29

Val Asp Leu Arg Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile
1 5 10 15

Lys Pro Lys Val Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val
20 25 30

Gly His Gly Phe Thr His Ala His Ile Val Ala Leu Ser Gln His Pro
35 40 45

Ala Ala Leu Gly Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala
50 55 60

Leu Pro Glu Ala Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp
65 70 75 80

Ser Gly Ala Arg Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu
85 90 95

Arg Gly Pro Pro Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala
100 105 110

[0042]

Lys Arg Gly Gly Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn
115 120 125

Ala Leu Thr Gly Ala Pro
130

<210> 30
<211> 34
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 30

Leu Thr Asp Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly Val Thr
1 5 10 15

Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly Ala Pro
20 25 30

Leu Asn

<210> 31
<211> 30
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 31

Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg
1 5 10 15

Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly
20 25 30

<210> 32

<211> 34

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<220>

<221> misc 特征

<222> (12)..(13)

<223> Xaa 可以是任何天然氨基酸

<400> 32

Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Xaa Xaa Gly Gly Lys
1 5 10 15

Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp
20 25 30

His Gly

[0043]

<210> 33

<211> 15

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 33

Asp Ile Val Gly Val Lys Gln Trp Ser Gly Ala Arg Ala Leu Glu
1 5 10 15

<210> 34

<211> 15

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 34

Asp Ile Val Gly Val Lys Ser Arg Ser Gly Ala Arg Ala Leu Glu
1 5 10 15

<210> 35

<211> 15

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 35

Asp Ile Val Gly Val Lys Arg Gly Ala Gly Ala Arg Ala Leu Glu
1 5 10 15

<210> 36
<211> 14
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 36

Asp Ile Val Gly Val Lys Tyr His Gly Ala Arg Ala Leu Glu
1 5 10

<210> 37
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成构建体

<400> 37

ataaaccccc tecaaccagg c

21

<210> 38
<211> 71
<212> DNA
<213> 人工序列

<220>
<223> 合成构建体

[0044]

<400> 38

aagaaggtct tcattacacc tgcagctctc atttccata cagtcagtat caattctgga

60

agaatttcca g

71

<210> 39
<211> 7
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<220>
<221> misc 特征
<222> (1), (3)
<223> Xaa 可以是任何天然氨基酸

<400> 39

Xaa Xaa Xaa Ser Gly Ala Arg
1 5

<210> 40
<211> 18
<212> DNA
<213> 人工序列

<220>
<223> 合成构建体

<400> 40

gtcttcatta cacctgca

18

[0045]	<210>	41	
	<211>	18	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
	<400>	41	
		cttcattaca cctgcagc	18
	<210>	42	
	<211>	18	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
	<400>	42	
		ttcattacac ctgcagct	18
	<210>	43	
	<211>	18	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
	<400>	43	
		acctgcagct ctcatttt	18
[0045]	<210>	44	
	<211>	16	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
	<400>	44	
		gtcagtcata gtttaag	16
	<210>	45	
	<211>	18	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
	<400>	45	
		tcagtcatag ttaagacc	18
[0045]	<210>	46	
	<211>	19	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
	<400>	46	
		tcatagttaa gaccttctt	19
	<210>	47	
	<211>	17	
	<212>	DNA	

<213> 人工序列

<220>

<223> 合成构建体

<400> 47

agttaagacc ttcttaa

17

<210> 48

<211> 159

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<220>

<221> misc 特征

<222> (104)..(107)

<223> Xaa 可以是任何天然氨基酸

<400> 48

Pro Arg Pro Pro Arg Ala Lys Pro Ala Pro Arg Arg Arg Ala Ala Gln
1 5 10 15

Pro Ser Asp Ala Ser Pro Ala Ala Gln Val Asp Leu Arg Thr Leu Gly
20 25 30

Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg Ser Thr
35 40 45

[0046]

Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr His Ala
50 55 60

His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr Val Ala
65 70 75 80

Val Thr Tyr Gln His Ile Ile Thr Ala Leu Pro Glu Ala Thr His Glu
85 90 95

Asp Ile Val Gly Val Gly Lys Xaa Xaa Xaa Xaa Ala Arg Ala Leu Glu
100 105 110

Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro Leu Gln Leu
115 120 125

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

Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly Ala Pro
145 150 155

<210> 49

<211> 134

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<220>

<221> misc 特征

<222> (79).. (82)

<223> Xaa 可以是任何天然氨基酸

<400> 49

Val	Asp	Leu	Arg	Thr	Leu	Gly	Tyr	Ser	Gln	Gln	Gln	Gln	Glu	Lys	Ile
1				5					10					15	

Lys	Pro	Lys	Val	Arg	Ser	Thr	Val	Ala	Gln	His	His	Glu	Ala	Leu	Val
			20					25					30		

Gly	His	Gly	Phe	Thr	His	Ala	His	Ile	Val	Ala	Leu	Ser	Gln	His	Pro
		35					40					45			

Ala	Ala	Leu	Gly	Thr	Val	Ala	Val	Thr	Tyr	Gln	His	Ile	Ile	Thr	Ala
		50					55					60			

Leu	Pro	Glu	Ala	Thr	His	Glu	Asp	Ile	Val	Gly	Val	Gly	Lys	Xaa	Xaa
65					70					75					80

Xaa	Xaa	Ala	Arg	Ala	Leu	Glu	Ala	Leu	Leu	Thr	Asp	Ala	Gly	Glu	Leu
				85						90				95	

Arg	Gly	Pro	Pro	Leu	Gln	Leu	Asp	Thr	Gly	Gln	Leu	Val	Lys	Ile	Ala
			100					105						110	

Lys	Arg	Gly	Gly	Val	Thr	Ala	Met	Glu	Ala	Val	His	Ala	Ser	Arg	Asn
			115				120						125		

[0047]

Ala	Leu	Thr	Gly	Ala	Pro
					130

<210> 50

<211> 134

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 50

Val	Asp	Leu	Arg	Thr	Leu	Gly	Tyr	Ser	Gln	Gln	Gln	Gln	Glu	Lys	Ile
1				5					10					15	

Lys	Pro	Lys	Val	Arg	Ser	Thr	Val	Ala	Gln	His	His	Glu	Ala	Leu	Val
			20					25					30		

Gly	His	Gly	Phe	Thr	His	Ala	His	Ile	Val	Ala	Leu	Ser	Gln	His	Pro
		35					40					45			

Ala	Ala	Leu	Gly	Thr	Val	Ala	Val	Lys	Tyr	Gln	Asp	Met	Ile	Ala	Ala
		50					55					60			

Leu	Pro	Glu	Ala	Thr	His	Glu	Ala	Ile	Val	Gly	Val	Gly	Lys	Gln	Trp
65					70					75					80

Ser	Gly	Ala	Arg	Ala	Leu	Glu	Ala	Leu	Leu	Thr	Val	Ala	Gly	Glu	Leu
				85						90				95	

Arg	Gly	Pro	Pro	Leu	Gln	Leu	Asp	Thr	Gly	Gln	Leu	Leu	Lys	Ile	Ala
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

	100	105	110
	Lys Arg Gly Gly Val Thr Ala Val Glu Ala Val His Ala Trp Arg Asn		
	115	120	125
	Ala Leu Thr Gly Ala Pro		
	130		
	<210> 51		
	<211> 80		
	<212> PRT		
	<213> 人工序列		
	<220>		
	<223> 合成构建体		
	<400> 51		
	Asn Asp His Leu Val Ala Leu Ala Cys Leu Gly Gly Arg Pro Ala Met		
	1 5 10 15		
	Asp Ala Val Lys Lys Gly Leu Pro His Ala Pro Glu Leu Ile Arg Arg		
	20 25 30		
	Val Asn Arg Arg Ile Gly Glu Arg Thr Ser His Arg Val Ala Asp Tyr		
	35 40 45		
	Ala Gln Val Val Arg Val Leu Glu Phe Phe Gln Cys His Ser His Pro		
	50 55 60		
[0048]	Ala Tyr Ala Phe Asp Glu Ala Met Thr Gln Phe Gly Met Ser Gly Gln		
	65 70 75 80		
	<210> 52		
	<211> 46		
	<212> PRT		
	<213> 人工序列		
	<220>		
	<223> 合成构建体		
	<400> 52		
	Asn Asp His Leu Val Ala Leu Ala Cys Leu Gly Gly Arg Pro Ala Met		
	1 5 10 15		
	Asp Ala Val Lys Lys Gly Leu Pro His Ala Pro Glu Leu Ile Arg Arg		
	20 25 30		
	Val Asn Arg Arg Ile Gly Glu Arg Thr Ser His Arg Val Ala		
	35 40 45		
	<210> 53		
	<211> 10		
	<212> PRT		
	<213> 人工序列		
	<220>		
	<223> 合成构建体		
	<400> 53		
	Asn Asp His Leu Val Ala Leu Ala Cys Leu		
	1 5 10		

<210> 54
 <211> 135
 <212> PRT
 <213> 人工序列

<220>
 <223> 合成构建体

<400> 54

Gln Val Asp Leu Arg Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys
 1 5 10 15

Ile Lys Pro Lys Val Arg Ser Thr Val Ala Gln His His Glu Ala Leu
 20 25 30

Val Gly His Gly Phe Thr His Ala His Ile Val Ala Leu Ser Gln His
 35 40 45

Pro Ala Ala Leu Gly Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr
 50 55 60

Ala Leu Pro Glu Ala Thr His Glu Asp Ile Val Gly Val Gly Lys Ser
 65 70 75 80

Arg Ser Gly Ala Arg Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu
 85 90 95

Leu Arg Gly Pro Pro Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile
 100 105 110

[0049]

Ala Lys Arg Gly Gly Val Thr Ala Met Glu Ala Val His Ala Ser Arg
 115 120 125

Asn Ala Leu Thr Gly Ala Pro
 130 135

<210> 55
 <211> 132
 <212> PRT
 <213> 人工序列

<220>
 <223> 合成构建体

<400> 55

Met Ala Ala Leu Gly Tyr Ser Arg Glu Gln Ile Arg Lys Leu Lys Gln
 1 5 10 15

Glu Ser Leu Ser Gly Val Ala Lys Tyr His Ala Pro Leu Thr Arg His
 20 25 30

Gly Phe Thr His Thr Asp Ile Cys Arg Ile Ser Arg Arg Trp Gln Ser
 35 40 45

Leu Arg Met Val Ala Lys Asn Tyr Pro Lys Leu Ile Ala Ala Leu Pro
 50 55 60

Asp Leu Thr Arg Thr His Ile Val Asp Ile Ala Arg Gln Arg Ser Gly
 65 70 75 80

Asp Leu Ala Leu Glu Ala Leu Leu Pro Val Ala Thr Ala Leu Ala Ala
85 90 95

Ala Pro Leu Arg Leu Arg Ala Ser Gln Ile Ala Ile Ile Ala Gln Cys
100 105 110

Gly Glu Arg Pro Ala Ile Leu Ala Leu His Arg Leu Arg Arg Lys Leu
115 120 125

Thr Gly Ala Pro
130

<210> 56

<211> 180

<212> PRT

<213> 人工序列

<220>

<223> 合成构象体

<400> 56

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala
35 40 45

[0050]

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

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala
115 120 125

Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Pro Lys Leu Thr
130 135 140

Lys Ala Glu Trp Glu Gln Ala Gly Arg Leu Leu Ala Gln Gly Ile Pro
145 150 155 160

Arg Lys Gln Val Ala Leu Ile Tyr Asp Val Ala Leu Ser Thr Leu Tyr
165 170 175

Lys Lys His Pro
180

<210> 57

<211> 144
 <212> PRT
 <213> 人工序列

<220>
 <223> 合成构建体

<400> 57

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
 1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
 20 25 30

Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala
 35 40 45

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

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
 65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
 85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
 100 105 110

[0051]

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala
 115 120 125

Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Pro Lys Ser Gly
 130 135 140

<210> 58
 <211> 143
 <212> PRT
 <213> 人工序列

<220>
 <223> 合成构建体

<400> 58

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
 1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
 20 25 30

Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala
 35 40 45

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

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
 65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr

	85	90	95
	Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala		
	100	105	110
	Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Ile Ala		
	115	120	125
	Ala Ala Arg Asn Lys Gly Arg Arg Phe Gly Arg Pro Pro Lys Ser		
	130	135	140
	<210> 59		
	<211> 144		
	<212> PRT		
	<213> 人工序列		
	<220>		
	<223> 合成构建体		
	<400> 59		
	Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp		
	1	5	10 15
	Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu		
	20	25	30
	Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala		
	35	40	45
[0052]	Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp		
	50	55	60
	Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu		
	65	70	75 80
	Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr		
	85	90	95
	Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala		
	100	105	110
	Glu Met Glu Arg Glu Leu Ile Leu Glu Arg Val Met Ala Gly Ile Ala		
	115	120	125
	Ala Ala Arg Asn Lys Gly Arg Arg Trp Gly Arg Pro Pro Lys Ser Gly		
	130	135	140
	<210> 60		
	<211> 144		
	<212> PRT		
	<213> 人工序列		
	<220>		
	<223> 合成构建体		
	<400> 60		
	Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp		
	1	5	10 15
	Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu		

	20	25	30
	Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala 35 40 45		
	Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp 50 55 60		
	Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu 65 70 75 80		
	Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr 85 90 95		
	Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala 100 105 110		
	Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Val Met Ala Gly Leu Ala 115 120 125		
	Ala Ala Arg Asn Lys Gly Arg Arg Phe Gly Arg Pro Pro Lys Ser Gly 130 135 140		
	<210> 61 <211> 144 <212> PRT <213> 人工序列		
[0053]	<220> <223> 合成构建体		
	<400> 61		
	Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp 1 5 10 15		
	Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu 20 25 30		
	Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala 35 40 45		
	Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp 50 55 60		
	Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu 65 70 75 80		
	Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr 85 90 95		
	Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala 100 105 110		
	Glu Met Glu Arg Leu Ser Ile Leu Glu Arg Pro Met Ala Gly His Ala 115 120 125		
	Ala Ala Arg Asn Lys Gly Arg Arg Phe Gly Arg Pro Pro Lys Ser Gly 130 135 140		

<210> 62
<211> 144
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 62

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala
35 40 45

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

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

[0054]

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Ser Ala Gly Arg Ala
115 120 125

Ala Ala Ile Asn Lys Gly Arg Ile Met Gly Arg Pro Arg Lys Ser Gly
130 135 140

<210> 63
<211> 267
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 63

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala
35 40 45

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

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

	Arg	Glu	Arg	Gly	Ile	Asn	Phe	Arg	Ser	Leu	Thr	Asp	Ser	Ile	Asp	Thr	
					85					90					95		
	Ser	Ser	Pro	Met	Gly	Arg	Phe	Phe	Phe	Tyr	Val	Met	Gly	Ala	Leu	Ala	
				100					105					110			
	Glu	Met	Glu	Arg	Glu	Leu	Ile	Ile	Glu	Arg	Thr	Met	Ala	Gly	Leu	Ala	
				115				120					125				
	Ala	Ala	Arg	Asn	Lys	Gly	Arg	Ile	Gly	Gly	Arg	Pro	Pro	Lys	Ser	Gly	
				130				135				140					
	Thr	Gly	Glu	Lys	Pro	Tyr	Lys	Cys	Pro	Glu	Cys	Gly	Lys	Ser	Phe	Ser	
	145					150				155						160	
	Thr	Ser	Gly	Asn	Leu	Val	Arg	His	Gln	Arg	Thr	His	Thr	Gly	Glu	Lys	
				165					170						175		
	Pro	Tyr	Lys	Cys	Pro	Glu	Cys	Gly	Lys	Ser	Phe	Ser	Gln	Ser	Gly	Asp	
				180				185					190				
	Leu	Arg	Arg	His	Gln	Arg	Thr	His	Thr	Gly	Glu	Lys	Pro	Tyr	Lys	Cys	
				195				200					205				
	Pro	Glu	Cys	Gly	Lys	Ser	Phe	Ser	Thr	Ser	Gly	Asn	Leu	Val	Arg	His	
				210			215					220					
[0055]	Gln	Arg	Thr	His	Thr	Gly	Glu	Lys	Pro	Tyr	Lys	Cys	Pro	Glu	Cys	Gly	
	225					230					235					240	
	Lys	Ser	Phe	Ser	Thr	Ser	Gly	Glu	Leu	Val	Arg	His	Gln	Arg	Thr	His	
				245						250					255		
	Thr	Gly	Lys	Lys	Thr	Ser	Gly	Gln	Ala	Gly	Gln						
				260					265								
	<210>	64															
	<211>	267															
	<212>	PRT															
	<213>	人工序列															
	<220>																
	<223>	合成构建体															
	<400>	64															
	Met	Leu	Ile	Gly	Tyr	Val	Arg	Val	Ser	Thr	Asn	Asp	Gln	Asn	Thr	Asp	
	1				5					10					15		
	Leu	Gln	Arg	Asn	Ala	Leu	Val	Cys	Ala	Gly	Cys	Glu	Gln	Ile	Phe	Glu	
				20					25					30			
	Asp	Lys	Leu	Ser	Gly	Thr	Arg	Thr	Asp	Arg	Pro	Gly	Leu	Lys	Arg	Ala	
			35					40					45				
	Leu	Lys	Arg	Leu	Gln	Lys	Gly	Asp	Thr	Leu	Val	Val	Trp	Lys	Leu	Asp	
			50				55					60					
	Arg	Leu	Gly	Arg	Ser	Met	Lys	His	Leu	Ile	Ser	Leu	Val	Gly	Glu	Leu	

65	70	75	80
Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr			
	85	90	95
Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala			
	100	105	110
Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala			
	115	120	125
Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Pro Lys Ser Gly			
	130	135	140
Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser			
	145	150	155
His Arg Thr Thr Leu Thr Asn His Gln Arg Thr His Thr Gly Glu Lys			
	165	170	175
Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Gln Ser Gly Asp			
	180	185	190
Leu Arg Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys			
	195	200	205
Pro Glu Cys Gly Lys Ser Phe Ser Gln Ser Gly Asp Leu Arg Arg His			
	210	215	220
Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly			
	225	230	235
Lys Ser Phe Ser Gln Ser Gly Asp Leu Arg Arg His Gln Arg Thr His			
	245	250	255
Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln			
	260	265	

[0056]

<210> 65
 <211> 267
 <212> PRT
 <213> 人工序列

<220>
 <223> 合成构建体

<400> 65

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp			
1	5	10	15
Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu			
	20	25	30
Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala			
	35	40	45
Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp			
	50	55	60

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
 65 70 75 80
 Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
 85 90 95
 Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
 100 105 110
 Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala
 115 120 125
 Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Pro Lys Ser Gly
 130 135 140
 Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
 145 150 155 160
 Gln Ser Gly Asp Leu Arg Arg His Gln Arg Thr His Thr Gly Glu Lys
 165 170 175
 Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Gln Arg Ala His
 180 185 190
 Leu Glu Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
 195 200 205
 Pro Glu Cys Gly Lys Ser Phe Ser Thr Ser Gly Asn Leu Val Arg His
 210 215 220
 Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
 225 230 235 240
 Lys Ser Phe Ser Arg Ser Asp Glu Leu Val Arg His Gln Arg Thr His
 245 250 255
 Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln
 260 265

[0057]

<210> 66
 <211> 267
 <212> PRT
 <213> 人工序列
 <220>
 <223> 合成构建体
 <400> 66

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

50	55	60
Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu 65 70 75 80		
Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr 85 90 95		
Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala 100 105 110		
Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala 115 120 125		
Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Pro Lys Ser Gly 130 135 140		
Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser 145 150 155 160		
Arg Ser Asp Lys Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys 165 170 175		
Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Lys Asp Asn 180 185 190		
Leu Lys Asn His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys 195 200 205		
Pro Glu Cys Gly Lys Ser Phe Ser Thr Ser Gly Glu Leu Val Arg His 210 215 220		
Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly 225 230 235 240		
Lys Ser Phe Ser Arg Ser Asp Lys Leu Val Arg His Gln Arg Thr His 245 250 255		
Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln 260 265		
<210> 67 <211> 267 <212> PRT <213> 人工序列		
<220> <223> 合成构建体		
<400> 67		
Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp 1 5 10 15		
Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu 20 25 30		
Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala 35 40 45		

[0058]

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
 50 55 60
 Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
 65 70 75 80
 Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
 85 90 95
 Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
 100 105 110
 Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala
 115 120 125
 Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Pro Lys Ser Gly
 130 135 140
 Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
 145 150 155 160
 Thr Thr Gly Asn Leu Thr Val His Gln Arg Thr His Thr Gly Glu Lys
 165 170 175
 Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Asp Pro Gly Ala
 180 185 190
 [0059] Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
 195 200 205
 Pro Glu Cys Gly Lys Ser Phe Ser Gln Ser Ser Asn Leu Val Arg His
 210 215 220
 Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
 225 230 235 240
 Lys Ser Phe Ser Arg Ser Asp His Leu Thr Asn His Gln Arg Thr His
 245 250 255
 Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln
 260 265
 <210> 68
 <211> 267
 <212> PRT
 <213> 人工序列
 <220>
 <223> 合成构建体
 <400> 68
 Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
 1 5 10 15
 Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
 20 25 30
 Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala

[0060]

35	40	45
Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp 50 55 60		
Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu 65 70 75 80		
Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr 85 90 95		
Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala 100 105 110		
Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala 115 120 125		
Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Pro Lys Ser Gly 130 135 140		
Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser 145 150 155 160		
Arg Lys Asp Asn Leu Lys Asn His Gln Arg Thr His Thr Gly Glu Lys 165 170 175		
Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Ser Asp His 180 185 190		
Leu Thr Asn His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys 195 200 205		
Pro Glu Cys Gly Lys Ser Phe Ser Asp Pro Gly Asn Leu Val Arg His 210 215 220		
Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly 225 230 235 240		
Lys Ser Phe Ser Arg Lys Asp Asn Leu Lys Asn His Gln Arg Thr His 245 250 255		
Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln 260 265		
<210> 69		
<211> 267		
<212> PRT		
<213> 人工序列		
<220>		
<223> 合成构建体		
<400> 69		
Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp 1 5 10 15		
Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu 20 25 30		

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

[0061]

<210> 70
 <211> 267
 <212> PRT
 <213> 人工序列

<220>
 <223> 合成构建体

<400> 70

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
 1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu

[0062]

	20	25	30
Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala	35	40	45
Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp	50	55	60
Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu	65	70	75 80
Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr	85	90	95
Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala	100	105	110
Glu Met Glu Arg Glu Leu Ile Leu Glu Arg Val Met Ala Gly Ile Ala	115	120	125
Ala Ala Arg Asn Lys Gly Arg Arg Trp Gly Arg Pro Pro Lys Ser Gly	130	135	140
Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser	145	150	155 160
Gln Arg Ala Asn Leu Arg Ala His Gln Arg Thr His Thr Gly Glu Lys	165	170	175
Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Arg Asp Glu	180	185	190
Leu Asn Val His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys	195	200	205
Pro Glu Cys Gly Lys Ser Phe Ser Gln Leu Ala His Leu Arg Ala His	210	215	220
Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly	225	230	235 240
Lys Ser Phe Ser Gln Arg Ala His Leu Glu Arg His Gln Arg Thr His	245	250	255
Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln	260	265	
<210> 71			
<211> 267			
<212> PRT			
<213> 人工序列			
<220>			
<223> 合成构象体			
<400> 71			
Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp	1	5	10 15

[0063]

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
 20 25 30
 Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala
 35 40 45
 Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
 50 55 60
 Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
 65 70 75 80
 Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
 85 90 95
 Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
 100 105 110
 Glu Met Glu Arg Glu Leu Ile Leu Glu Arg Val Met Ala Gly Ile Ala
 115 120 125
 Ala Ala Arg Asn Lys Gly Arg Arg Trp Gly Arg Pro Pro Lys Ser Gly
 130 135 140
 Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
 145 150 155 160
 Arg Arg Asp Glu Leu Asn Val His Gln Arg Thr His Thr Gly Glu Lys
 165 170 175
 Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Ser Asp His
 180 185 190
 Leu Thr Asn His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
 195 200 205
 Pro Glu Cys Gly Lys Ser Phe Ser Gln Leu Ala His Leu Arg Ala His
 210 215 220
 Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
 225 230 235 240
 Lys Ser Phe Ser Gln Arg Ala His Leu Glu Arg His Gln Arg Thr His
 245 250 255
 Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln
 260 265
 <210> 72
 <211> 267
 <212> PRT
 <213> 人工序列
 <220>
 <223> 合成构建体
 <400> 72
 Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp

[0064]

1	5	10	15
Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu	20	25	30
Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala	35	40	45
Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp	50	55	60
Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu	65	70	75
Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr	85	90	95
Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala	100	105	110
Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala	115	120	125
Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Pro Lys Ser Gly	130	135	140
Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser	145	150	155
Thr Ser Gly Ser Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys	165	170	175
Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Ser Asp Lys	180	185	190
Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys	195	200	205
Pro Glu Cys Gly Lys Ser Phe Ser Gln Ser Gly Asp Leu Arg Arg His	210	215	220
Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly	225	230	235
Lys Ser Phe Ser Thr Ser Gly Glu Leu Val Arg His Gln Arg Thr His	245	250	255
Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln	260	265	
<210> 73			
<211> 267			
<212> PRT			
<213> 人工序列			
<220>			
<223> 合成构建体			
<400> 73			

	Met	Leu	Ile	Gly	Tyr	Val	Arg	Val	Ser	Thr	Asn	Asp	Gln	Asn	Thr	Asp	
	1				5					10					15		
	Leu	Gln	Arg	Asn	Ala	Leu	Val	Cys	Ala	Gly	Cys	Glu	Gln	Ile	Phe	Glu	
				20					25					30			
	Asp	Lys	Leu	Ser	Gly	Thr	Arg	Thr	Asp	Arg	Pro	Gly	Leu	Lys	Arg	Ala	
			35					40					45				
	Leu	Lys	Arg	Leu	Gln	Lys	Gly	Asp	Thr	Leu	Val	Val	Trp	Lys	Leu	Asp	
		50					55					60					
	Arg	Leu	Gly	Arg	Ser	Met	Lys	His	Leu	Ile	Ser	Leu	Val	Gly	Glu	Leu	
	65					70				75						80	
	Arg	Glu	Arg	Gly	Ile	Asn	Phe	Arg	Ser	Leu	Thr	Asp	Ser	Ile	Asp	Thr	
					85					90					95		
	Ser	Ser	Pro	Met	Gly	Arg	Phe	Phe	Phe	Tyr	Val	Met	Gly	Ala	Leu	Ala	
				100					105					110			
	Glu	Met	Glu	Arg	Glu	Leu	Ile	Ile	Glu	Arg	Thr	Ser	Ala	Gly	Arg	Ala	
		115					120						125				
	Ala	Ala	Ile	Asn	Lys	Gly	Arg	Ile	Met	Gly	Arg	Pro	Arg	Lys	Ser	Gly	
		130					135					140					
[0065]	Thr	Gly	Glu	Lys	Pro	Tyr	Lys	Cys	Pro	Glu	Cys	Gly	Lys	Ser	Phe	Ser	
	145					150				155						160	
	Gln	Leu	Ala	His	Leu	Arg	Ala	His	Gln	Arg	Thr	His	Thr	Gly	Glu	Lys	
				165					170						175		
	Pro	Tyr	Lys	Cys	Pro	Glu	Cys	Gly	Lys	Ser	Phe	Ser	Gln	Leu	Ala	His	
				180					185					190			
	Leu	Arg	Ala	His	Gln	Arg	Thr	His	Thr	Gly	Glu	Lys	Pro	Tyr	Lys	Cys	
		195					200						205				
	Pro	Glu	Cys	Gly	Lys	Ser	Phe	Ser	Asp	Pro	Gly	His	Leu	Val	Arg	His	
		210					215					220					
	Gln	Arg	Thr	His	Thr	Gly	Glu	Lys	Pro	Tyr	Lys	Cys	Pro	Glu	Cys	Gly	
	225					230				235					240		
	Lys	Ser	Phe	Ser	Asp	Ser	Gly	Asn	Leu	Arg	Val	His	Gln	Arg	Thr	His	
				245					250					255			
	Thr	Gly	Lys	Lys	Thr	Ser	Gly	Gln	Ala	Gly	Gln						
			260					265									
	<210>	74															
	<211>	267															
	<212>	PRT															
	<213>	人工序列															
	<220>																
	<223>	合成构建体															

<400> 74

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala
35 40 45

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

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

Glu Met Glu Arg Glu Leu Ile Leu Glu Arg Val Met Ala Gly Ile Ala
115 120 125

Ala Ala Arg Asn Lys Gly Arg Arg Trp Gly Arg Pro Pro Lys Ser Gly
130 135 140

[0066]

Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
145 150 155 160

Gln Arg Ala His Leu Glu Arg His Gln Arg Thr His Thr Gly Glu Lys
165 170 175

Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Thr Thr Gly Asn
180 185 190

Leu Thr Val His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
195 200 205

Pro Glu Cys Gly Lys Ser Phe Ser Asp Ser Gly Asn Leu Arg Val His
210 215 220

Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
225 230 235 240

Lys Ser Phe Ser Gln Ser Ser Asn Leu Val Arg His Gln Arg Thr His
245 250 255

Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln
260 265

<210> 75

<211> 267

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 75

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala
35 40 45

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

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Ile Ala
115 120 125

[0067]

Ala Ala Arg Asn Lys Gly Arg Arg Phe Gly Arg Pro Pro Lys Ser Gly
130 135 140

Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
145 150 155 160

Thr His Leu Asp Leu Ile Arg His Gln Arg Thr His Thr Gly Glu Lys
165 170 175

Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Thr Thr Gly Asn
180 185 190

Leu Thr Val His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
195 200 205

Pro Glu Cys Gly Lys Ser Phe Ser Gln Ser Ser Ser Leu Val Arg His
210 215 220

Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
225 230 235 240

Lys Ser Phe Ser Arg Ser Asp Asn Leu Val Arg His Gln Arg Thr His
245 250 255

Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln
260 265

<210> 76

<211> 267

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 76

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala
35 40 45

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

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Ile Ala
115 120 125

[0068]

Ala Ala Arg Asn Lys Gly Arg Arg Phe Gly Arg Pro Pro Lys Ser Gly
130 135 140

Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
145 150 155 160

Arg Ser Asp Lys Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys
165 170 175

Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Arg Asp Glu
180 185 190

Leu Asn Val His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
195 200 205

Pro Glu Cys Gly Lys Ser Phe Ser Gln Ser Ser Ser Leu Val Arg His
210 215 220

Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
225 230 235 240

Lys Ser Phe Ser Arg Ser Asp His Leu Thr Asn His Gln Arg Thr His
245 250 255

Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln
260 265

<210> 77

<211> 267
 <212> PRT
 <213> 人工序列

<220>
 <223> 合成构建体

<400> 77

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
 1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
 20 25 30

Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala
 35 40 45

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

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
 65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
 85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
 100 105 110

[0069] Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala
 115 120 125

Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Pro Lys Ser Gly
 130 135 140

Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
 145 150 155 160

Gln Arg Ala His Leu Glu Arg His Gln Arg Thr His Thr Gly Glu Lys
 165 170 175

Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Thr Ser Gly Asn
 180 185 190

Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
 195 200 205

Pro Glu Cys Gly Lys Ser Phe Ser Arg Ser Asp Glu Leu Val Arg His
 210 215 220

Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
 225 230 235 240

Lys Ser Phe Ser His Lys Asn Ala Leu Gln Asn His Gln Arg Thr His
 245 250 255

Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln
 260 265

<210> 78
<211> 267
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 78

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala
35 40 45

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

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

[0070]

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala
115 120 125

Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Pro Lys Ser Gly
130 135 140

Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
145 150 155 160

Arg Arg Asp Glu Leu Asn Val His Gln Arg Thr His Thr Gly Glu Lys
165 170 175

Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Gln Ser Ser Asn
180 185 190

Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
195 200 205

Pro Glu Cys Gly Lys Ser Phe Ser Gln Ser Ser Ser Leu Val Arg His
210 215 220

Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
225 230 235 240

Lys Ser Phe Ser Thr Thr Gly Asn Leu Thr Val His Gln Arg Thr His
245 250 255

Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln

	260	265
	<210> 79 <211> 267 <212> PRT <213> 人工序列 <220> <223> 合成构造体 <400> 79	
	Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp	
	1 5 10 15	
	Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu	
	20 25 30	
	Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala	
	35 40 45	
	Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp	
	50 55 60	
	Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu	
	65 70 75 80	
	Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr	
	85 90 95	
[0071]	Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala	
	100 105 110	
	Glu Met Glu Arg Glu Leu Ile Leu Glu Arg Val Met Ala Gly Ile Ala	
	115 120 125	
	Ala Ala Arg Asn Lys Gly Arg Arg Trp Gly Arg Pro Pro Lys Ser Gly	
	130 135 140	
	Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser	
	145 150 155 160	
	Thr Thr Gly Asn Leu Thr Val His Gln Arg Thr His Thr Gly Glu Lys	
	165 170 175	
	Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Gln Ser Ser Asn	
	180 185 190	
	Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys	
	195 200 205	
	Pro Glu Cys Gly Lys Ser Phe Ser Gln Arg Ala His Leu Glu Arg His	
	210 215 220	
	Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly	
	225 230 235 240	
	Lys Ser Phe Ser Gln Lys Ser Ser Leu Ile Ala His Gln Arg Thr His	
	245 250 255	

Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln
260 265

<210> 80
<211> 267
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 80

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala
35 40 45

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

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

[0072]

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Ser Ala Gly Arg Ala
115 120 125

Ala Ala Ile Asn Lys Gly Arg Ile Met Gly Arg Pro Arg Lys Ser Gly
130 135 140

Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
145 150 155 160

Asp Pro Gly Ala Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys
165 170 175

Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Gln Ser Ser Ser
180 185 190

Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
195 200 205

Pro Glu Cys Gly Lys Ser Phe Ser Gln Leu Ala His Leu Arg Ala His
210 215 220

Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
225 230 235 240

Lys Ser Phe Ser Gln Arg Ala Asn Leu Arg Ala His Gln Arg Thr His

	245	250	255
	Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln		
	260	265	
<210>	81		
<211>	938		
<212>	PRT		
<213>	人工序列		
<220>			
<223>	合成构建体		
<400>	81		
	Met Arg Ser Pro Lys Lys Lys Arg Lys Val Gln Val Asp Leu Arg Thr		
	1 5 10 15		
	Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg		
	20 25 30		
	Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr		
	35 40 45		
	His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr		
	50 55 60		
	Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala Leu Pro Glu Ala Thr		
	65 70 75 80		
[0073]	His Glu Asp Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala		
	85 90 95		
	Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro Leu		
	100 105 110		
	Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly Val		
	115 120 125		
	Thr Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly Ala		
	130 135 140		
	Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly		
	145 150 155 160		
	Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu		
	165 170 175		
	Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser		
	180 185 190		
	His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro		
	195 200 205		
	Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile		
	210 215 220		
	Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu		
	225 230 235 240		

	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	
					245					250					255		
	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	
				260					265					270			
	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	
		275						280					285				
	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	
		290					295					300					
	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	
	305				310						315					320	
	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	
				325						330					335		
	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	
			340					345						350			
	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	
			355					360					365				
	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	
		370					375						380				
[0074]	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	
	385					390				395					400		
	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	
				405					410						415		
	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Val	Ser	His	Asp	
			420					425					430				
	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	
		435					440					445					
	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Val	Ser	
		450				455					460						
	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	
	465				470					475					480		
	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	
				485					490					495			
	Val	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	
			500					505						510			
	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	
		515						520					525				
	Ala	Ile	Ala	Ser	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	
		530				535						540					

	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	
	545					550					555					560	
	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	
				565						570					575		
	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	
				580					585					590			
	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	
		595						600					605				
	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	
	610					615						620					
	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Asn	Gly	Gly	Lys	Gln	
	625					630					635					640	
	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	
					645					650					655		
	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	
			660						665					670			
	Lys	Gln	Ala	Leu	Glu	Ser	Ile	Val	Ala	Gln	Leu	Ser	Arg	Pro	Asp	Pro	
		675						680					685				
[0075]	Ala	Leu	Ala	Ala	Leu	Thr	Asn	Asp	His	Leu	Val	Ala	Leu	Ala	Cys	Leu	
	690						695					700					
	Gly	Gly	Arg	Pro	Ala	Met	Asp	Ala	Val	Lys	Lys	Gly	Leu	Pro	His	Ala	
	705					710					715					720	
	Pro	Glu	Leu	Ile	Arg	Arg	Val	Asn	Arg	Arg	Ile	Gly	Glu	Arg	Thr	Ser	
				725						730					735		
	His	Arg	Val	Ala	Gly	Ser	Gln	Leu	Val	Lys	Ser	Glu	Leu	Glu	Glu	Lys	
				740					745					750			
	Lys	Ser	Glu	Leu	Arg	His	Lys	Leu	Lys	Tyr	Val	Pro	His	Glu	Tyr	Ile	
		755					760						765				
	Glu	Leu	Ile	Glu	Ile	Ala	Arg	Asn	Ser	Thr	Gln	Asp	Arg	Ile	Leu	Glu	
	770					775						780					
	Met	Lys	Val	Met	Glu	Phe	Phe	Met	Lys	Val	Tyr	Gly	Tyr	Arg	Gly	Lys	
	785					790					795					800	
	His	Leu	Gly	Gly	Ser	Arg	Lys	Pro	Asp	Gly	Ala	Ile	Tyr	Thr	Val	Gly	
					805					810					815		
	Ser	Pro	Ile	Asp	Tyr	Gly	Val	Ile	Val	Asp	Thr	Lys	Ala	Tyr	Ser	Gly	
			820						825					830			
	Gly	Tyr	Asn	Leu	Pro	Ile	Gly	Gln	Ala	Asp	Glu	Met	Gln	Arg	Tyr	Val	
		835					840						845				

Glu Glu Asn Gln Thr Arg Asn Lys His Ile Asn Pro Asn Glu Trp Trp
850 855 860

Lys Val Tyr Pro Ser Ser Val Thr Glu Phe Lys Phe Leu Phe Val Ser
865 870 875 880

Gly His Phe Lys Gly Asn Tyr Lys Ala Gln Leu Thr Arg Leu Asn His
885 890 895

Ile Thr Asn Cys Asn Gly Ala Val Leu Ser Val Glu Glu Leu Leu Ile
900 905 910

Gly Gly Glu Met Ile Lys Ala Gly Thr Leu Thr Leu Glu Glu Val Arg
915 920 925

Arg Lys Phe Asn Asn Gly Glu Ile Asn Phe
930 935

<210> 82
<211> 34
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 82

Asn Gly Asn Asn His Asp Asn Gly His Asp Asn Ile Asn Gly Asn Gly
1 5 10 15

[0076]

Asn Ile His Asp Asn Ile His Asp His Asp Asn Gly Asn Asn His Asp
20 25 30

Asn Ile

<210> 83
<211> 30
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 83

His Asp Asn Ile Asn Asn Asn Gly His Asp Asn Ile Asn Asn Asn Gly
1 5 10 15

Asn Ile Asn Gly His Asp Asn Ile Asn Ile Asn Gly Asn Gly
20 25 30

<210> 84
<211> 34
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 84

His Asp His Asp Asn Gly Asn Asn His Asp Asn Ile Asn Asn His Asp

1 5 10 15
Asn Gly His Asp Asn Gly His Asp Asn Ile Asn Gly Asn Gly Asn Gly
20 25 30

Asn Gly

<210> 85
<211> 30
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 85

Asn Ile Asn Gly Asn Gly His Asp Asn Gly Asn Gly His Asp His Asp
1 5 10 15

Asn Ile Asn Asn Asn Ile Asn Gly Asn Gly Asn Asn Asn Ile
20 25 30

<210> 86
<211> 34
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

[0077]

<400> 86

His Asp Asn Ile Asn Asn Asn Ile Asn Ile Asn Gly Asn Gly Asn Asn
1 5 10 15

Asn Ile Asn Gly Asn Ile His Asp Asn Gly Asn Asn Asn Ile His Asp
20 25 30

Asn Gly

<210> 87
<211> 32
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 87

Asn Gly His Asp Asn Ile Asn Gly Asn Gly Asn Ile His Asp Asn Ile
1 5 10 15

His Asp His Asp Asn Gly Asn Asn His Asp Asn Ile Asn Asn His Asp
20 25 30

<210> 88
<211> 36
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 88

His Asp Asn Gly Asn Gly His Asp His Asp Asn Ile Asn Asn Asn Ile
1 5 10 15

Asn Ile Asn Gly Asn Gly Asn Asn Asn Ile Asn Gly Asn Ile His Asp
20 25 30

Asn Gly Asn Asn
35

<210> 89

<211> 432

<212> DNA

<213> 人工序列

<220>

<223> 合成构建体

<400> 89

atgagatctc ctaagaaaa gaggaagatg gtggacttga ggacactcgg ttattcgaan 60

cagcaacagg agaaaatcaa gcctaaggtc aggagcaccg tcgcgcaaca ccacgaggcg 120

cttgtggggc atggcttcac tcatgecat attgtcgcgc tttcacagca ccctgcggcg 180

cttgggacgg tggctgtcaa ataccaagat atgattgcgg ccctgcccga agccacgcac 240

gaggcaattg taggggtcgg taacagtggt tcgggagcgc gacacttga ggcgctgctg 300

actgtggcgg gtgagcttag ggggectcgg ctccagctcg acaccgggca gctgcigaag 360

atcgcaaga gagggggagt aacagcggta gaggcagtgc atgcacgcg caatgcactg 420

[0078] aegggtgccc cc 432

<210> 90

<211> 144

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 90

Met Arg Ser Pro Lys Lys Lys Arg Lys Met Val Asp Leu Arg Thr Leu
1 5 10 15

Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg Ser
20 25 30

Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr His
35 40 45

Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr Val
50 55 60

Ala Val Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr His
65 70 75 80

Glu Ala Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala Leu
85 90 95

Glu Ala Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu Gln

Leu	Asp	Thr	Gly	Gln	Leu	Leu	Lys	Ile	Ala	Lys	Arg	Gly	Gly	Val	Thr
	115						120					125			

Ala	Val	Glu	Ala	Val	His	Ala	Ser	Arg	Asn	Ala	Leu	Thr	Gly	Ala	Pro
	130					135					140				

<210> 91

<211> 432

<212> DNA

<213> 人工序列

<220>

<223> 合成构建体

<400> 91

```

atgagatctc ctaagaaaaa gaggaagggtg caggtaggac tacgcacgct cggctacagt      60
cagcagcagc aagagaagat caaacccaag gtgcgttcga cagtggcgca gcaccacgag      120
gcactggtgg gccatgggtt tacacacgcg cacatcgttg cgtcagcca acaccggca      180
gcgttaggga ccgtcgtgt caggtatcag cacataatca cggcgttgc agaggcgaca      240
cacgaagaca tcgttgcggt cggcaaatat cctggggcac gcgtcttga gcccttgctc      300
acggatcgcg gggagttgag aggtccgcg ttacagttgg acacagcgca acttgtgaag      360
attgcaaac gtggcggtt gaccgaatg gaggcagtgc atgcalcgcg caatgcactg      420
acgggtgccc cc                                     432

```

[0079]

<210> 92

<211> 144

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 92

Met	Arg	Ser	Pro	Lys	Lys	Lys	Arg	Lys	Val	Gln	Val	Asp	Leu	Arg	Thr
1				5					10				15		

Leu	Gly	Tyr	Ser	Gln	Gln	Gln	Gln	Glu	Lys	Ile	Lys	Pro	Lys	Val	Arg
			20					25					30		

Ser	Thr	Val	Ala	Gln	His	His	Glu	Ala	Leu	Val	Gly	His	Gly	Phe	Thr
		35					40					45			

His	Ala	His	Ile	Val	Ala	Leu	Ser	Gln	His	Pro	Ala	Ala	Leu	Gly	Thr
	50					55					60				

Val	Ala	Val	Thr	Tyr	Gln	His	Ile	Ile	Thr	Ala	Leu	Pro	Glu	Ala	Thr
65					70				75					80	

His	Glu	Asp	Ile	Val	Gly	Val	Gly	Lys	Tyr	His	Gly	Ala	Arg	Ala	Leu
			85					90						95	

Glu	Ala	Leu	Leu	Thr	Asp	Ala	Gly	Glu	Leu	Arg	Gly	Pro	Pro	Leu	Gln
		100					105					110			

Leu	Asp	Thr	Gly	Gln	Leu	Val	Lys	Ile	Ala	Lys	Arg	Gly	Gly	Val	Thr
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

	115	120	125
	Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly Ala Pro		
	130	135	140
<210>	93		
<211>	435		
<212>	DNA		
<213>	人工序列		
<220>			
<223>	合成构建体		
<400>	93		
	atgagatctc ctaagaaaa gaggaagggtg caggtaggac tacgcacgct cggctacagt	60	
	cagcagcagc aagagaagat caaacccaag gtgcgttcga cagtggcgca gcaccacgag	120	
	gcactggtgg gccatgggtt tacacacgcg cacatcgttg cgtcagcca acaccggca	180	
	gggttaggga ccgtcgtgt cactatcag cacataatca cggcgttgc agaggcgaca	240	
	caagaagaca tcgttgccgt cggcaatcg cggtcgggg caccgcctct ggaggccttg	300	
	ctcacgcatg cgggggagtt gagaggtccg ccgttacagt tggacacagg ccaacttgt	360	
	aagattgcaa aacgtggcgg cgtgaccga atggaggcag tgcattgcat gcgcaatgca	420	
	ctgacgggtg ccccc	435	
<210>	94		
<211>	145		
<212>	PRT		
<213>	人工序列		
<220>			
<223>	合成构建体		
<400>	94		
	Met Arg Ser Pro Lys Lys Lys Arg Lys Val Gln Val Asp Leu Arg Thr		
	1 5 10 15		
	Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg		
	20 25 30		
	Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr		
	35 40 45		
	His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr		
	50 55 60		
	Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala Leu Pro Glu Ala Thr		
	65 70 75 80		
	His Glu Asp Ile Val Gly Val Gly Lys Ser Arg Ser Gly Ala Arg Ala		
	85 90 95		
	Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro Leu		
	100 105 110		
	Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly Val		
	115 120 125		
	Thr Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly Ala		

[0080]

	130	135	140
	Pro		
	145		
	<210> 95		
	<211> 435		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成构建体		
	<400> 95		
	atgagatctc ctaagaaaaa gaggaagggtg caggiggatc tacgcacgct cggctacagt	60	
	cagcagcagc aagagaagat caaacccaag gtgcgttcga cagtggcgca gcaccacggg	120	
	gcactgggtg gccatgggtt tacacacgcg cacatcgttg cgcicagcca acacccggca	180	
	gcgttaggga ccgtcgtgt cactatcag cacataatca cggcgttgcc agagcgaca	240	
	cacgaagaca tcgttgcggt cggcaaacgg ggggctgggt cacgcgtctt ggaggccttg	300	
	ctcaccgatg cgggggggtt gagaggtccg ccgttacagt tggacacagg ccaacttggt	360	
	aagattgcaa aacgtggcgg cgtgaccgca atggaggcag tgcatgcata gcgcaatgca	420	
	ctgacgggtg ccccc	435	
	<210> 96		
	<211> 145		
	<212> PRT		
	<213> 人工序列		
[0081]	<220>		
	<223> 合成构建体		
	<400> 96		
	Met Arg Ser Pro Lys Lys Lys Arg Lys Val Gln Val Asp Leu Arg Thr		
	1 5 10 15		
	Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg		
	20 25 30		
	Ser Thr Val Ala Gln His His Gly Ala Leu Val Gly His Gly Phe Thr		
	35 40 45		
	His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr		
	50 55 60		
	Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala Leu Pro Glu Ala Thr		
	65 70 75 80		
	His Glu Asp Ile Val Gly Val Gly Lys Arg Gly Ala Gly Ala Arg Ala		
	85 90 95		
	Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro Leu		
	100 105 110		
	Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly Val		
	115 120 125		
	Thr Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly Ala		

	130	135	140
	Pro		
	145		
	<210> 97		
	<211> 435		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成构建体		
	<400> 97		
	atgagatctc ctaagaaaaa gaggaagggtg caggiggatc tacgcacgct cggctacagt	60	
	cagcagcagc aagagaagat caaacccaag gtgcgttcga cagtggcgca gcaccacgag	120	
	gcactgggtg gccatgggtt tacacacgcg cacatcgttg cgcicagcca acacccggca	180	
	gcgttaggga ccgtcgtgt cactatcag cacataatca cggcgttgcc agagcgagaca	240	
	cacgaagaca tcgttggtt cggcaaacag tggtcggcg cacgcgccct ggaggccttg	300	
	ctcaccgatg cgggggggtt gagaggtcgg ccgttacagt tggacacagg ccaacttggtg	360	
	aagattgcaa aacgtggcgg cgtgaccgca atggaggcag tgcatgcac gcgcaatgca	420	
	ctgacgggtg ccccc	435	
	<210> 98		
	<211> 1080		
	<212> PRT		
	<213> 人工序列		
[0082]	<220>		
	<223> 合成构建体		
	<220>		
	<221> MISC 特征		
	<222> (498)..(498)		
	<223> Xaa 是 Q, S, R 或 Y		
	<220>		
	<221> MISC 特征		
	<222> (499)..(499)		
	<223> Xaa 是 W, R 或 G		
	<220>		
	<221> MISC 特征		
	<222> (500)..(500)		
	<223> Xaa 是 S, A 或 H		
	<400> 98		
	Met Lys Ile Glu Glu Gly Lys Leu Val Ile Trp Ile Asn Gly Asp Lys		
	1 5 10 15		
	Gly Tyr Asn Gly Leu Ala Glu Val Gly Lys Lys Phe Glu Lys Asp Thr		
	20 25 30		
	Gly Ile Lys Val Thr Val Glu His Pro Asp Lys Leu Glu Glu Lys Phe		
	35 40 45		
	Pro Gln Val Ala Ala Thr Gly Asp Gly Pro Asp Ile Ile Phe Trp Ala		
	50 55 60		
	His Asp Arg Phe Gly Gly Tyr Ala Gln Ser Gly Leu Leu Ala Glu Ile		

65	70	75	80
Thr Pro Asp Lys Ala Phe Gln Asp Lys Leu Tyr Pro Phe Thr Trp Asp	85	90	95
Ala Val Arg Tyr Asn Gly Lys Leu Ile Ala Tyr Pro Ile Ala Val Glu	100	105	110
Ala Leu Ser Leu Ile Tyr Asn Lys Asp Leu Leu Pro Asn Pro Pro Lys	115	120	125
Thr Trp Glu Glu Ile Pro Ala Leu Asp Lys Glu Leu Lys Ala Lys Gly	130	135	140
Lys Ser Ala Leu Met Phe Asn Leu Gln Glu Pro Tyr Phe Thr Trp Pro	145	150	155
Leu Ile Ala Ala Asp Gly Gly Tyr Ala Phe Lys Tyr Glu Asn Gly Lys	165	170	175
Tyr Asp Ile Lys Asp Val Gly Val Asp Asn Ala Gly Ala Lys Ala Gly	180	185	190
Leu Thr Phe Leu Val Asp Leu Ile Lys Asn Lys His Met Asn Ala Asp	195	200	205
Thr Asp Tyr Ser Ile Ala Glu Ala Ala Phe Asn Lys Gly Glu Thr Ala	210	215	220
Met Thr Ile Asn Gly Pro Trp Ala Trp Ser Asn Ile Asp Thr Ser Lys	225	230	235
Val Asn Tyr Gly Val Thr Val Leu Pro Thr Phe Lys Gly Gln Pro Ser	245	250	255
Lys Pro Phe Val Gly Val Leu Ser Ala Gly Ile Asn Ala Ala Ser Pro	260	265	270
Asn Lys Glu Leu Ala Lys Glu Phe Leu Glu Asn Tyr Leu Leu Thr Asp	275	280	285
Glu Gly Leu Glu Ala Val Asn Lys Asp Lys Pro Leu Gly Ala Val Ala	290	295	300
Leu Lys Ser Tyr Glu Glu Glu Leu Ala Lys Asp Pro Arg Ile Ala Ala	305	310	315
Thr Met Glu Asn Ala Gln Lys Gly Glu Ile Met Pro Asn Ile Pro Gln	325	330	335
Met Ser Ala Phe Trp Tyr Ala Val Arg Thr Ala Val Ile Asn Ala Ala	340	345	350
Ser Gly Arg Gln Thr Val Asp Glu Ala Leu Lys Asp Ala Gln Thr Asn	355	360	365
Ser Ser Ser Asn Asn Asn Asn Asn Asn Asn Asn Asn Asn Leu Gly Ile			

[0083]

370	375	380
Glu Gly Arg Ile Ser	Glu Phe Gly Ser Pro Ala Arg Pro Pro Arg Ala	
385	390	395 400
Lys Pro Ala Pro Arg Arg Arg Ser Ala Gln Pro Ser Asp Ala Ser Pro		
	405	410 415
Ala Ala Gln Val Asp Leu Arg Thr Leu Gly Tyr Ser Gln Gln Gln Gln		
	420	425 430
Glu Lys Ile Lys Pro Lys Val Arg Ser Thr Val Ala Gln His His Glu		
	435	440 445
Ala Leu Val Gly His Gly Phe Thr His Ala His Ile Val Ala Leu Ser		
	450	455 460
Gln His Pro Ala Ala Leu Gly Thr Val Ala Val Thr Tyr Gln His Ile		
	465	470 475 480
Ile Thr Ala Leu Pro Glu Ala Thr His Glu Asp Ile Val Gly Val Gly		
	485	490 495
Lys Xaa Xaa Xaa Gly Ala Arg Ala Leu Glu Ala Leu Leu Thr Asp Ala		
	500	505 510
Gly Glu Leu Leu Arg Gly Pro Pro Leu Gln Leu Asp Thr Gly Gln Leu		
	515	520 525
Val Lys Ile Ala Lys Arg Gly Gly Val Thr Ala Met Glu Ala Val His		
	530	535 540
Ala Ser Arg Asn Ala Leu Thr Gly Ala Pro Leu Asn Leu Thr Pro Asp		
	545	550 555 560
Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu		
	565	570 575
Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr		
	580	585 590
Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala		
	595	600 605
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly		
	610	615 620
Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys		
	625	630 635 640
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp		
	645	650 655
His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly		
	660	665 670
Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys		

[0084]

675	680	685
Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn 690	695	700
Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val 705	710	715 720
Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala 725	730	735
Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu 740	745	750
Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala 755	760	765
Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg 770	775	780
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val 785	790	795 800
Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val 805	810	815
Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp 820	825	830
Gln Val Val Ala Ile Val Ser His Asp Gly Gly Lys Gln Ala Leu Glu 835	840	845
Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr 850	855	860
Pro Asp Gln Val Val Ala Ile Val Ser His Asp Gly Gly Lys Gln Ala 865	870	875 880
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly 885	890	895
Leu Thr Pro Asp Gln Val Val Ala Ile Val Ser Asn Gly Gly Gly Lys 900	905	910
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp 915	920	925
His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly 930	935	940
Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys 945	950	955 960
Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His 965	970	975
Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val		

[0085]

	980	985	990
	Leu Cys Gln Asp His Gly Leu Thr	Pro Asp Gln Val Val	Ala Ile Ala
	995	1000	1005
	Ser Asn Ile Gly Gly Lys Gln	Ala Leu Glu Thr Val	Gln Arg Leu
	1010	1015	1020
	Leu Pro Val Leu Cys Gln Asp	His Gly Leu Thr Pro	Asp Gln Val
	1025	1030	1035
	Val Ala Ile Ala Ser Asn Ile	Gly Gly Lys Gln Ala	Leu Glu Ser
	1040	1045	1050
	Ile Val Ala Gln Leu Ser Arg	Pro Asp Pro Ala Leu	Ala Ala Leu
	1055	1060	1065
	Thr Asn Asp His Leu Val Ala	Leu Ala Cys Leu Gly	
	1070	1075	1080
	<210>	99	
	<211>	830	
	<212>	PRT	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
[0086]	<220>		
	<221>	MISC 特征	
	<222>	(249)..(249)	
	<223>	Xaa 是 Q, S, R 或 Y	
	<220>		
	<221>	MISC 特征	
	<222>	(250)..(250)	
	<223>	Xaa 是 W, R 或 G	
	<220>		
	<221>	MISC 特征	
	<222>	(251)..(251)	
	<223>	Xaa 是 S, A 或 H	
	<400>	99	
	Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp		
	1	5	10 15
	Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu		
	20	25	30
	Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala		
	35	40	45
	Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp		
	50	55	60
	Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu		
	65	70	75 80
	Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr		
	85	90	95

	Ser	Ser	Pro	Met	Gly	Arg	Phe	Phe	Phe	Tyr	Val	Met	Gly	Ala	Leu	Ala	
				100					105					110			
	Glu	Met	Glu	Arg	Glu	Leu	Ile	Ile	Glu	Arg	Thr	Met	Ala	Gly	Leu	Ala	
		115					120					125					
	Ala	Ala	Arg	Asn	Lys	Gly	Arg	Ile	Gly	Gly	Arg	Pro	Pro	Lys	Ser	Gly	
		130					135					140					
	Ser	Pro	Arg	Pro	Pro	Arg	Ala	Lys	Pro	Ala	Pro	Arg	Arg	Arg	Ala	Ala	
	145					150					155					160	
	Gln	Pro	Ser	Asp	Ala	Ser	Pro	Ala	Ala	Gln	Val	Asp	Leu	Arg	Thr	Leu	
				165						170					175		
	Gly	Tyr	Ser	Gln	Gln	Gln	Gln	Glu	Lys	Ile	Lys	Pro	Lys	Val	Arg	Ser	
			180						185					190			
	Thr	Val	Ala	Gln	His	His	Glu	Ala	Leu	Val	Gly	His	Gly	Phe	Thr	His	
		195					200						205				
	Ala	His	Ile	Val	Ala	Leu	Ser	Gln	His	Pro	Ala	Ala	Leu	Gly	Thr	Val	
		210					215					220					
	Ala	Val	Thr	Tyr	Gln	His	Ile	Ile	Thr	Ala	Leu	Pro	Glu	Ala	Thr	His	
	225				230						235				240		
[0087]	Glu	Asp	Ile	Val	Gly	Val	Gly	Lys	Xaa	Xaa	Xaa	Gly	Ala	Arg	Ala	Leu	
				245					250					255			
	Glu	Ala	Leu	Leu	Thr	Asp	Ala	Gly	Glu	Leu	Arg	Gly	Pro	Pro	Leu	Gln	
			260					265					270				
	Leu	Asp	Thr	Gly	Gln	Leu	Val	Lys	Ile	Ala	Lys	Arg	Gly	Gly	Val	Thr	
		275						280					285				
	Ala	Met	Glu	Ala	Val	His	Ala	Ser	Arg	Asn	Ala	Leu	Thr	Gly	Ala	Pro	
		290				295						300					
	Leu	Asn	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	
	305				310					315					320		
	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	
			325						330					335			
	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	
			340					345					350				
	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	
		355					360						365				
	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	
		370					375					380					
	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	
	385					390					395				400		

	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	
					405					410						415	
	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	
				420					425					430			
	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	
			435					440					445				
	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	
			450				455						460				
	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	
	465				470					475						480	
	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	
				485						490						495	
	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	
			500					505						510			
	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	
			515					520						525			
	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	
		530					535					540					
[0088]	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	
	545					550					555					560	
	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	
				565					570						575		
	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Val	Ser	His	Asp	Gly	
			580					585						590			
	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	
		595					600						605				
	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Val	Ser	His	
		610					615					620					
	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	
	625					630					635				640		
	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Val	
				645						650					655		
	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	
			660						665					670			
	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	
		675						680					685				
	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	
		690					695						700				

Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val
 705 710 715 720
 Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val
 725 730 735
 Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp
 740 745 750
 Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu
 755 760 765
 Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr
 770 775 780
 Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala
 785 790 795 800
 Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp Pro Ala Leu Ala
 805 810 815
 Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys Leu Gly
 820 825 830
 <210> 100
 <211> 839
 <212> PRT
 <213> 人工序列
 [0089] <220>
 <223> 合成构建体
 <220>
 <221> MISC 特征
 <222> (112)..(112)
 <223> Xaa 是 Q, S, R 或 Y
 <220>
 <221> MISC 特征
 <222> (113)..(113)
 <223> Xaa 是 W, R 或 R
 <220>
 <221> MISC 特征
 <222> (114)..(114)
 <223> Xaa 是 S, A 或 H
 <400> 100
 Met Ala Gln Ala Ala Ser Gly Ser Pro Arg Pro Pro Arg Ala Lys Pro
 1 5 10 15
 Ala Pro Arg Arg Arg Ala Ala Gln Pro Ser Asp Ala Ser Pro Ala Ala
 20 25 30
 Gln Val Asp Leu Arg Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys
 35 40 45
 Ile Lys Pro Lys Val Arg Ser Thr Val Ala Gln His His Glu Ala Leu
 50 55 60
 Val Gly His Gly Phe Thr His Ala His Ile Val Ala Leu Ser Gln His

[0090]

65	70	75	80
Pro Ala Ala Leu Gly Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr	85	90	95
Ala Leu Pro Glu Ala Thr His Glu Asp Ile Val Gly Val Gly Lys Xaa	100	105	110
Xaa Xaa Gly Ala Arg Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu	115	120	125
Leu Arg Gly Pro Pro Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile	130	135	140
Ala Lys Arg Gly Gly Val Thr Ala Met Glu Ala Val His Ala Ser Arg	145	150	155
Asn Ala Leu Thr Gly Ala Pro Leu Asn Leu Thr Pro Asp Gln Val Val	165	170	175
Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln	180	185	190
Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln	195	200	205
Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr	210	215	220
Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro	225	230	235
Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu	245	250	255
Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu	260	265	270
Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln	275	280	285
Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His	290	295	300
Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly	305	310	315
Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln	325	330	335
Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp	340	345	350
Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu	355	360	365
Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser			

[0091]

370	375	380
His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro		
385	390	395 400
Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile		
	405	410 415
Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu		
	420	425 430
Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val		
	435	440 445
Ala Ile Val Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln		
	450	455 460
Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln		
	465	470 475 480
Val Val Ala Ile Val Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr		
	485	490 495
Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro		
	500	505 510
Asp Gln Val Val Ala Ile Val Ser Asn Gly Gly Gly Lys Gln Ala Leu		
	515	520 525
Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu		
	530	535 540
Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln		
	545	550 555 560
Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His		
	565	570 575
Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly		
	580	585 590
Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln		
	595	600 605
Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile		
	610	615 620
Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu		
	625	630 635 640
Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser		
	645	650 655
Asn Ile Gly Gly Lys Gln Ala Leu Glu Ser Ile Val Ala Gln Leu Ser		
	660	665 670
Arg Pro Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala		

675	680	685
Leu Ala Cys Leu Gly Gly Arg Pro Ala Met Asp Ala Val Lys Lys Gly		
690	695	700
Leu Pro His Ala Pro Glu Leu Ile Arg Arg Val Asn Arg Arg Ile Gly		
705	710	715
Glu Arg Thr Ser His Arg Val Ala Asp Tyr Ala Gln Val Val Arg Val		
725	730	735
Leu Glu Phe Phe Gln Cys His Ser His Pro Ala Tyr Ala Phe Asp Glu		
740	745	750
Ala Met Thr Gln Phe Gly Met Ser Gly Gln Ala Gly Gln Ala Ser Pro		
755	760	765
Lys Lys Lys Arg Lys Val Gly Arg Ala Asp Ala Leu Asp Asp Phe Asp		
770	775	780
Leu Asp Met Leu Gly Ser Asp Ala Leu Asp Asp Phe Asp Leu Asp Met		
785	790	795
Leu Gly Ser Asp Ala Leu Asp Asp Phe Asp Leu Asp Met Leu Gly Ser		
805	810	815
Asp Ala Leu Asp Asp Phe Asp Leu Asp Met Leu Ile Asn Tyr Pro Tyr		
820	825	830

[0092]

Asp Val Pro Asp Tyr Ala Ser
835

<210> 101
 <211> 30
 <212> PRT
 <213> 人工序列

<220>
 <223> 合成构建体

<400> 101

Asn Ile Asn Gly Asn Ile Asn Ile Asn Ile His Asp His Asp His Asp
1 5 10 15

His Asp His Asp Asn Ile His Asp His Asp Asn Ile Asn Ile
20 25 30

<210> 102
 <211> 33
 <212> PRT
 <213> 人工序列

<220>
 <223> 合成构建体

<400> 102

Leu Thr Pro Asp Gln Leu Val Lys Ile Ala Lys Arg Gly Gly Thr Ala
1 5 10 15

Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly Ala Pro Leu

	20	25	30
	Asn		
	<210> 103		
	<211> 4		
	<212> PRT		
	<213> 人工序列		
	<220>		
	<223> 合成构建体		
	<400> 103		
	Lys Arg Gly Gly		
	1		
	<210> 104		
	<211> 4		
	<212> PRT		
	<213> 人工序列		
	<220>		
	<223> 合成构建体		
	<400> 104		
	Leu Asp Tyr Glu		
	1		
[0093]	<210> 105		
	<211> 4		
	<212> PRT		
	<213> 人工序列		
	<220>		
	<223> 合成构建体		
	<400> 105		
	Ile Asn Leu Val		
	1		
	<210> 106		
	<211> 4		
	<212> PRT		
	<213> 人工序列		
	<220>		
	<223> 合成构建体		
	<400> 106		
	Tyr Ser Lys Lys		
	1		
	<210> 107		
	<211> 4		
	<212> PRT		
	<213> 人工序列		
	<220>		
	<223> 合成构建体		
	<400> 107		
	Asn Met Ala His		
	1		

[0094]

<210> 108
<211> 4
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 108

Ser Pro Thr Asn
1

<210> 109
<211> 4
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 109

Ser Asn Thr Arg
1

<210> 110
<211> 4
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 110

Leu Thr Thr Thr
1

<210> 111
<211> 4
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 111

Val Ala Asp Leu
1

<210> 112
<211> 4
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 112

Met Val Leu Ser
1

<210> 113
<211> 4
<212> PRT
<213> 人工序列

	<220>	
	<223>	合成构建体
	<400>	113
	Tyr Asn Gly Arg	
	1	
	<210>	114
	<211>	4
	<212>	PRT
	<213>	人工序列
	<220>	
	<223>	合成构建体
	<400>	114
	Arg Ile Pro Arg	
	1	
	<210>	115
	<211>	4
	<212>	PRT
	<213>	人工序列
	<220>	
	<223>	合成构建体
	<400>	115
	Tyr Ser Lys Ile	
	1	
[0095]	<210>	116
	<211>	4
	<212>	PRT
	<213>	人工序列
	<220>	
	<223>	合成构建体
	<400>	116
	Leu Thr Gln Tyr	
	1	
	<210>	117
	<211>	4
	<212>	PRT
	<213>	人工序列
	<220>	
	<223>	合成构建体
	<400>	117
	Tyr Leu Ser Lys	
	1	
	<210>	118
	<211>	4
	<212>	PRT
	<213>	人工序列
	<220>	
	<223>	合成构建体
	<400>	118
	Leu Arg Pro Asn	

1

<210> 119
<211> 4
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 119

Leu Phe Thr Asn
1

<210> 120
<211> 4
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 120

Leu Leu Thr Asn
1

<210> 121
<211> 4
<212> PRT
<213> 人工序列

[0096]

<220>
<223> 合成构建体

<400> 121

Glu Glu Asp Lys
1

<210> 122
<211> 4
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 122

Val Thr Ala Met
1

<210> 123
<211> 4
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 123

Cys Pro Ser Arg
1

<210> 124
<211> 4
<212> PRT

<213> 人工序列
<220>
<223> 合成构建体
<400> 124
Leu Thr Arg Val
1

<210> 125
<211> 4
<212> PRT
<213> 人工序列
<220>
<223> 合成构建体
<400> 125

Lys Gly Asp Leu
1

<210> 126
<211> 4
<212> PRT
<213> 人工序列
<220>
<223> 合成构建体
<400> 126

Gln Lys Ala Leu
1

[0097]

<210> 127
<211> 4
<212> PRT
<213> 人工序列
<220>
<223> 合成构建体
<400> 127

Leu Tyr Leu Leu
1

<210> 128
<211> 4
<212> PRT
<213> 人工序列
<220>
<223> 合成构建体
<400> 128

Trp Ile Ser Val
1

<210> 129
<211> 4
<212> PRT
<213> 人工序列
<220>
<223> 合成构建体
<400> 129

Gly Asp Gln Val
1

<210> 130
<211> 4
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 130

Cys Pro Ser Arg
1

<210> 131
<211> 145
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<220>
<221> misc_特征
<222> (89)..(93)
<223> Xaa 可以是任何天然氨基酸

<400> 131

Met Arg Ser Pro Lys Lys Lys Arg Lys Val Gln Val Asp Leu Arg Thr
1 5 10 15

[0098]

Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg
20 25 30

Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr
35 40 45

His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
50 55 60

Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala Leu Pro Glu Ala Thr
65 70 75 80

His Glu Asp Ile Val Gly Val Gly Xaa Xaa Xaa Xaa Xaa Ala Arg Ala
85 90 95

Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro Leu
100 105 110

Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly Val
115 120 125

Thr Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly Ala
130 135 140

Pro
145

<210> 132

<211> 5
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 132

Lys Arg Pro Ala Gly
1 5

<210> 133
<211> 5
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 133

Lys Arg Pro Ser Gly
1 5

<210> 134
<211> 32
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 134

[0099]

Leu Thr Pro Asp Val Val Ala Ile Ser Asn Asn Gly Gly Lys Gln Ala
1 5 10 15

Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp Gly His
20 25 30

<210> 135
<211> 4
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 135

Ser Asn Asn Gly
1

<210> 136
<211> 4
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 136

Arg Gly Gly Gly
1

<210> 137
<211> 4
<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 137

Arg Gly Gly Arg

1

<210> 138

<211> 4

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 138

Arg Gly Val Arg

1

<210> 139

<211> 4

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 139

Lys Gly Gly Gly

1

[0100]

<210> 140

<211> 4

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 140

Ser Gly Gly Gly

1

<210> 141

<211> 4

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 141

Gly Gly Arg Gly

1

<210> 142

<211> 4

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 142

	Leu Gly Gly Ser I
	<210> 143
	<211> 4
	<212> PRT
	<213> 人工序列
	<220>
	<223> 合成构建体
	<400> 143
	Met Asp Asn Ile I
	<210> 144
	<211> 4
	<212> PRT
	<213> 人工序列
	<220>
	<223> 合成构建体
	<400> 144
	Arg Val Met Ala I
	<210> 145
	<211> 4
	<212> PRT
	<213> 人工序列
[0101]	<220>
	<223> 合成构建体
	<400> 145
	Leu Ala Ser Val I
	<210> 146
	<211> 4
	<212> PRT
	<213> 人工序列
	<220>
	<223> 合成构建体
	<400> 146
	Val Gly Thr Gly I
	<210> 147
	<211> 4
	<212> PRT
	<213> 人工序列
	<220>
	<223> 合成构建体
	<400> 147
	Gln Gly Gly Gly I
	<210> 148

	<211> 118	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 148	
	ttaattaaga gtctagaaat ataaaccccc tccaaccagg tgctaactgt aaaccatggt	60
	tttggattag cacctgggtg gaggggggtt ataagatcta ggaggaattt aaaatgag	118
	<210> 149	
	<211> 117	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 149	
	actgacctag agaagettat ataaaccccc tccaaccagg tgctaateca aaaccatggt	60
	ttacagttag cacctgggtg gaggggggtt atacgcagt tatttgtaca gttcatc	117
	<210> 150	
	<211> 116	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 150	
	ttaattaaga gtctagatta gcacctgggt ggaggggggt tataagggtt tggtacccaa	60
[0102]	tgctataaa cccctccaa ccaggtgcta aagatctagg aggaatttaa aatgag	116
	<210> 151	
	<211> 116	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 151	
	ttaattaaga gtctagatta gcacctgggt ggaggggggt tataagggtt tggtacccaa	60
	tgctataaa cccctccaa ccaggtgcta aagatctagg aggaatttaa aatgag	116
	<210> 152	
	<211> 116	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 152	
	actgacctag agaagctttt agcacctggt tggaggggtt ttatagacat ttgtaccaa	60
	aacctataa acccctcca accagggtgt aactgcagtt attgtacag ttcatc	116
	<210> 153	
	<211> 116	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	

[0103]	<400> 153	
	ttaattaaga gtctagatta gcacctgggt ggaggggggt tatatccaa accatgggtt	60
	acagtataaa cccctccaa ccagggtcfa aagatctagg aggaatttaa aatgag	116
	<210> 154	
	<211> 116	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 154	
	actgacctag agaagctttt agcacctgggt tggaggggggt ttatatccaa aacctgggtt	60
	tacagtataa acccctccaa accaggtgct aactgcagtt attgttacag ttcatc	116
	<210> 155	
	<211> 122	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
[0103]	<400> 155	
	ttaattaaga gtctagatta gcacctgggt ggaggggggt tatagcttcc aaaacctggt	60
	tttacagggt tataaacccc ctccaaccag gtgctaaaga tctaggagga atttaaaatg	120
	ag	122
	<210> 156	
	<211> 128	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 156	
	ttaattaaga gtctagatta gcacctgggt ggaggggggt tatagcttca tccaaaacca	60
	tggttttacag ggttctata aacccctcc aaccagggtc taaagatcta ggaggaattt	120
	aaaatgag	128
	<210> 157	
	<211> 128	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
[0103]	<400> 157	
	actgacctag agaagctttt agcacctgggt tggaggggggt ttatagcaac cctgtaacc	60
	atggttttgg atgaagctat aaacccctc caaccagggtg ctaactgcag ttatttgtac	120
	agttcatc	128
	<210> 158	
	<211> 140	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	

<400> 158
 ttaattaaga gtctagatta gcacctgggt ggaggggggt tatagcttca gcttcatcca 60
 aaaccatggt ttacagggtt ccggttccta taaacccct ccaaccaggt gctaaagatc 120
 taggaggaat ttaaatgag 140

<210> 159
 <211> 140
 <212> DNA
 <213> 人工序列

<220>
 <223> 引物

<400> 159
 ttaattaaga gtctagatta gcacctgggt ggaggggggt tatagcttca gcttcatcca 60
 aaaccatggt ttacagggtt ccggttccta taaacccct ccaaccaggt gctaaagatc 120
 taggaggaat ttaaatgag 140

<210> 160
 <211> 110
 <212> DNA
 <213> 人工序列

<220>
 <223> 引物

<400> 160
 ttaattaaga gtctagatta gcacctgggt ggaggggggt taaaaaac atgggttata 60
 taaacccct ccaaccaggt gctaaagatc taggaggaat ttaaatgag 110

[0104]

<210> 161
 <211> 110
 <212> DNA
 <213> 人工序列

<220>
 <223> 引物

<400> 161
 actgacctag agaagctttt agcacctggt tggagggggt ttatataaac catggttttt 60
 ataaacccc tcaaccagg tgcctaactgc agttatttgt acagttcacc 110

<210> 162
 <211> 116
 <212> DNA
 <213> 人工序列

<220>
 <223> 引物

<400> 162
 ttaattaaga gtctagatta gcacctgggt ggaggggggt tataatccaa accgggggtt 60
 acagtataaa cccctccaa ccaggtgcta aagatctagg aggaatttaa aatgag 116

<210> 163
 <211> 116
 <212> DNA
 <213> 人工序列

<220>
 <223> 引物

<400> 163

	actgacctag agaagctttt agcacctggg ttgagggggg ttatactgta aaccccgggt	60
	ttggatataa accccctcca accaggtgct aactgcagtt attgttacag ttcata	116
	<210> 164	
	<211> 116	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 164	
	ttaattaaga gtctagatta gcacctgggt ggaggggggt tatacgaat attataaatt	60
	atcatataaa ccccccctca ccagggtgcta aagatctagg aggaatttaa aatgag	116
	<210> 165	
	<211> 116	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 165	
	actgacctag agaagctttt agcacctggg ttgagggggg ttatatgata attataata	60
	tttcgtataa accccctcca accaggtgct aactgcagtt attgttacag ttcata	116
	<210> 166	
	<211> 128	
	<212> DNA	
	<213> 人工序列	
[0105]	<220>	
	<223> 引物	
	<400> 166	
	ttaattaaga gtctagatta gcacctgggt ggaggggggt tatagcttca tccaaaaccg	60
	gggttttacg ggttctata aacccccctc aaccagggtc taaagatcta ggaggaattt	120
	aaaatgag	128
	<210> 167	
	<211> 128	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 167	
	actgacctag agaagctttt agcacctggg ttgagggggg ttatagcaac cctgtaaacc	60
	ggggttttgg atgaagctat aaacccccctc caaccagggt ctaactgcag ttatttgtac	120
	agttcaltc	128
	<210> 168	
	<211> 128	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 168	
	ttaattaaga gtctagatta gcacctgggt ggaggggggt tatagcttca cgaatatatta	60

	taaattatca ggttcctata aacccectcc aaccagggtgc taaagatcta ggaggaattt	120
	aaaatgag	128
	<210> 169	
	<211> 128	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 169	
	actgacctag agaagcittt agcacctggt tggaggggt ttatagcaac ctgataattt	60
	ataatatttc gtgaagctat aaacccctc caaccagggt ctaactgcag ttatttgtac	120
	agttcatc	128
	<210> 170	
	<211> 128	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 170	
	ttaattaaga gagatcttta gcacctggt ggagggggt tatagcttca tccaaaacca	60
	tggtttacag ggttcctata aacccectcc aaccagggtc taagcgatct gcctctcaat	120
	tagtcagc	128
[0106]	<210> 171	
	<211> 128	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 171	
	actgacctag agaagctttt agcacctggt tggaggggt ttatagcaac cctgtaaacc	60
	atggttttgg atgaagctat aaacccctc caaccagggt ctaatttgca aaagcctagg	120
	cctccaaa	128
	<210> 172	
	<211> 107	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 172	
	ttaattaaga gagatctgcg ggagcggtg ccaaaacctt ggtttacagg gtctctataa	60
	aaccctccca accaggtgct aagcgatctg catctcaatt agtcagc	107
	<210> 173	
	<211> 107	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 173	

	actgacctag agaagctttt agcacctggt tggagggggt ttatagcaac cctgtaaacc	60
	atggttttgg acacgcctcc cgctttgcaa aagcctaggc ctccaaa	107
	<210> 174	
	<211> 140	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 174	
	ttaattaaga gagatcttta gcacctggtt ggaggggggt tatagcttca gcttcatcca	60
	aaaccatggt ttacagggtt ccggttccta taaacccct ccaaccaggt gctaagcgat	120
	ctgcctctca attagtcagc	140
	<210> 175	
	<211> 140	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 175	
	actgacctag agaagctttt agcacctggt tggagggggt ttatagcaac cgcaaccctg	60
	ttaaacctag ttttggatga agctgaagct ataaacccct tccaaccagg tgctaatllg	120
	caaaagccta ggcttccaaa	140
[0107]	<210> 176	
	<211> 52	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 176	
	agtcagtcga gagctcatgg atcccggctc tatgctgatt ggctatgtaa gg	52
	<210> 177	
	<211> 39	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 177	
	atgctgatat ctgactatc ccgatttagg tgggcgacc	39
	<210> 178	
	<211> 37	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 178	
	agtcagtcga gagctcatge tgattggeta tgtaagg	37
	<210> 179	
	<211> 32	
	<212> DNA	

	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 179	
	tctagactac ggatecaccg atttacggg gc	32
	<210> 180	
	<211> 39	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 180	
	atcgcgatc tagactagcc gaggcaggcc aaggcgacg	39
	<210> 181	
	<211> 39	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 181	
	atcgcgatc tagactagct catctcgaac tgcgtcatg	39
	<210> 182	
	<211> 43	
	<212> DNA	
	<213> 人工序列	
[0108]	<220>	
	<223> 引物	
	<400> 182	
	gtgcccgcg taaatcggga tccactgcag atcggggggg ggc	43
	<210> 183	
	<211> 43	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 183	
	gtgcccgcg taaatcggga tccccctcgc ctgegtcttc ggc	43
	<210> 184	
	<211> 43	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 184	
	gtgcccgcg taaatcggga tccgatcga tgcctgccgt cgg	43
	<210> 185	
	<211> 45	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	

	<400> 185 gtcgcccgcg taaatcgga tccacgtgc gtgtcgtgt cactg	45
	<210> 186 <211> 44 <212> DNA <213> 人工序列	
	<220> <223> 引物	
	<400> 186 gtcgcccgcg taaatcgga tccgtggatc tacgcacgt cggc	44
	<210> 187 <211> 43 <212> DNA <213> 人工序列	
	<220> <223> 引物	
	<400> 187 gtcgcccgcg taaatcgga tccacacacg cgcacatgt tgc	43
	<210> 188 <211> 45 <212> DNA <213> 人工序列	
	<220> <223> 引物	
[0109]	<400> 188 gtcgcccgcg taaatcgga tccacagaag acatcgttgg cgteg	45
	<210> 189 <211> 44 <212> DNA <213> 人工序列	
	<220> <223> 引物	
	<400> 189 gtcgcccgcg taaatcgga tccagcgtc tggaggcctt gctc	44
	<210> 190 <211> 44 <212> DNA <213> 人工序列	
	<220> <223> 引物	
	<400> 190 gtcgcccgcg taaatcgga tecttggaca caggccaact tetc	44
	<210> 191 <211> 43 <212> DNA <213> 人工序列	
	<220> <223> 引物	
	<400> 191 gtcgcccgcg taaatcgga tccagcggcg tgaccgcagt gga	43

	<210> 192	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 192	
	ggatcccgat ttacgcgggc	20
	<210> 193	
	<211> 41	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 193	
	atcgtagcag ctagegccac catgctgatt ggetatgtaa g	41
	<210> 194	
	<211> 26	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 194	
	ggatccagac cccgatttac gcgggc	26
[0110]	<210> 195	
	<211> 74	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 195	
	ttagcacctg gttggagggg gtttataacc aaaaccaagg ttacagtat aaacccctc	60
	caaccaggtg ctaa	74
	<210> 196	
	<211> 74	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 196	
	ttagcacctg gttggagggg gtttataagg tttaggiacc aaatgtctat aaacccctc	60
	caaccaggtg ctaa	74
	<210> 197	
	<211> 74	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 197	
	tataaacccc ctccaaccag gtgctaactg taaaccaagg ttttgatta gcacctggtt	60

	ggagggggtt tata	74
	<210> 198	
	<211> 68	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 198	
	ttagcacctg gttggagggg gtttataaaa accatggttt atataaaccc cctecaacca	60
	ggtgctaa	68
	<210> 199	
	<211> 80	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 199	
	ttagcacctg gttggagggg gtttatagct tccaaaacca tggtttacag ggttataaac	60
	ccccccaac caggcgctaa	80
	<210> 200	
	<211> 86	
	<212> DNA	
	<213> 人工序列	
	<220>	
[0111]	<223> 合成构建体	
	<400> 200	
	ttagcacctg gttggagggg gtttatagct tcatccaaaa ccatggttta cagggttcct	60
	ataaaccccc tccaaccagg tgctaa	86
	<210> 201	
	<211> 98	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 201	
	ttagcacctg gttggagggg gtttatagct tcagcttcat ccaaaacctt ggtttacagg	60
	gttccggttc ctataaaccc cctccaacca ggtgctaa	98
	<210> 202	
	<211> 74	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 202	
	ttagcacctg gttggagggg gtttatatcc aaaaccgggg ttacagtat aaacccctc	60
	caaccagggtg cttaa	74
	<210> 203	
	<211> 74	
	<212> DNA	

[0112]	<213>	人工序列	
	<220>		
	<223>	合成构建体	
	<400>	203	
		ttagcacctg gttggagggg gtttatacga aatattataa attatcatat aaaccccctc	60
		caaccaggtg cttaa	74
	<210>	204	
	<211>	86	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
	<400>	204	
		ttagcacctg gttggagggg gtttatacgt tcatccaaaa cgggggttta cagggttcct	60
		ataaaccccc tccaaccagg tgctaa	86
	<210>	205	
	<211>	86	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
	<400>	205	
		ttagcacctg gttggagggg gtttatacgt tcacgaaata ttataaatta tcagggttcct	60
		ataaaccccc tccaaccagg tgctaa	86
[0112]	<210>	206	
	<211>	65	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
	<400>	206	
		gcgggaggcg tgcacaaaac catgggttac aggggttccta taaaccccct ccaaccaggt	60
		gctaa	65
	<210>	207	
	<211>	70	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
	<400>	207	
		gtgggtgtaca gtagggggag atgcatccaa aaccatggtt tacagtgcac ctccccctac	60
		tgtacaccac	70
	<210>	208	
	<211>	82	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
	<400>	208	

	gtgggtgtaca gtagggggag atgcagctgc ttccaaaaac atgggtttaca gggtggttgc	60
	atctcccccct actgtacacc ac	82
	<210> 209	
	<211> 4	
	<212> PRT	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 209	
	Gln Trp Ser Gly	
	1	
	<210> 210	
	<211> 4	
	<212> PRT	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 210	
	Arg Ser Asn Gly	
	1	
	<210> 211	
	<211> 4	
	<212> PRT	
	<213> 人工序列	
[0113]	<220>	
	<223> 合成构建体	
	<400> 211	
	Ser Arg Ser Gly	
	1	
	<210> 212	
	<211> 4	
	<212> PRT	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 212	
	Gln Trp Ser Gly	
	1	
	<210> 213	
	<211> 16	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 213	
	gataaacccc ctecaa	16
	<210> 214	
	<211> 15	
	<212> DNA	

Lys Arg Ser Asn Gly
1 5

<210> 219
<211> 5
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 219

Lys Ser Arg Ser Gly
1 5

<210> 220
<211> 35
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 220

Ala Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala
1 5 10 15

Arg Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly Pro
20 25 30

[0115]

Pro Leu Gln
35

<210> 221
<211> 5
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 221

Lys Gln Trp Ser Gly
1 5

<210> 222
<211> 16
<212> DNA
<213> 人工序列

<220>
<223> 合成构建体

<400> 222

aataaacccc ctccaa

16

<210> 223
<211> 16
<212> DNA
<213> 人工序列

<220>
<223> 合成构建体

<400> 223

tataaacccc ctccaa

16

<210> 224

<211> 4

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 224

Lys Arg Gly Gly

1

<210> 225

<211> 34

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 225

Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys

1

5

10

15

Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp

20

25

30

His Gly

[0116]

<210> 226

<211> 4

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 226

Ser His Asp Gly

1

<210> 227

<211> 6

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 227

Ala Ser His Asp Gly Gly

1

5

<210> 228

<211> 81

<212> DNA

<213> 人工序列

<220>

<223> 引物

[0117]	<220>			
	<221>	misc_特征		
	<222>	(50)..(51)		
	<223>	n 是 a, c, g, 或 t		
	<220>			
	<221>	misc_特征		
	<222>	(53)..(54)		
	<223>	n 是 a, c, g, 或 t		
	<400>	228		
		tctcaactcc cccgcctccg tgagcaaggc ctccagagcg cgtgccccmn mnnntttgce	60	
		gacgccaacg atgtcttctg g	81	
	<210>	229		
	<211>	84		
	<212>	DNA		
	<213>	人工序列		
	<220>			
	<223>	引物		
[0117]	<220>			
	<221>	misc_特征		
	<222>	(47)..(48)		
	<223>	n 是 a, c, g, 或 t		
	<220>			
	<221>	misc_特征		
	<222>	(50)..(51)		
	<223>	n 是 a, c, g, 或 t		
	<220>			
	<221>	misc_特征		
	<222>	(53)..(54)		
	<223>	n 是 a, c, g, 或 t		
	<220>			
	<221>	misc_特征		
	<222>	(56)..(57)		
	<223>	n 是 a, c, g, 或 t		
[0117]	<400>	229		
		tctcaactcc cccgcctccg tgagcaaggc ctccagagcg cgtgcmnnnn mnnnnnnnttt	60	
		gccgacgcca acgatgtctt cgtg	84	
	<210>	230		
	<211>	80		
	<212>	DNA		
	<213>	人工序列		
	<220>			
	<223>	引物		
	<220>			
	<221>	misc_特征		
	<222>	(43)..(44)		
	<223>	n 是 a, c, g, 或 t		
	<220>			
	<221>	misc_特征		
	<222>	(46)..(47)		
	<223>	n 是 a, c, g, 或 t		
[0117]	<220>			
	<221>	misc_特征		
	<222>	(49)..(50)		
	<223>	n 是 a, c, g, 或 t		
	<400>	230		

	cccgccctcgg ttagcaaggc ctccaggcg cgtagcgccgg amnnnnnnnn gccgacgcca	60
	acgatgtctt cgtgtgtcgc	80
<210>	231	
<211>	90	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<220>		
<221>	misc_特征	
<222>	(53)..(54)	
<223>	n 是 a, c, g, 或 t	
<220>		
<221>	misc_特征	
<222>	(56)..(57)	
<223>	n 是 a, c, g, 或 t	
<220>		
<221>	misc_特征	
<222>	(59)..(60)	
<223>	n 是 a, c, g, 或 t	
<220>		
<221>	misc_特征	
<222>	(62)..(63)	
<223>	n 是 a, c, g, 或 t	
<400>	231	
	ggcacccgtc agtgcattgc gccatgcatg cactgectcc actgcggtea cmmnnnnnnn	60
[0118]	mnntgcaatc ttgagaagtt ggccgtgtgc	90
<210>	232	
<211>	71	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	232	
	agagagagaa gaaatgaga tctcctaaga aaaagaggaa ggtagcagggt gatctacgca	60
	cgctcggcta c	71
<210>	233	
<211>	71	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	233	
	aggaagaaga gaagcatgag atctcctaag aaaaagagga aggtgatggt ggacttgagg	60
	acactcggtt a	71
<210>	234	
<211>	44	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	

[0119]	<400> 234 aagagaagaa gaagaagcat tgcgccatgc atgcactgcc tcta	44
	<210> 235 <211> 17 <212> DNA <213> 人工序列	
	<220> <223> 引物	
	<400> 235 cccgccaccc accgtgc	17
	<210> 236 <211> 23 <212> DNA <213> 人工序列	
	<220> <223> 引物	
	<400> 236 tgetctatgc atgcactgcc tcc	23
	<210> 237 <211> 68 <212> DNA <213> 人工序列	
	<220> <223> 引物	
	<400> 237 agagaagaga agagaaggcg ccgcggccc aggcggcctc gggatcccct cggcctccgc	60
	gcgcacaag	68
	<210> 238 <211> 65 <212> DNA <213> 人工序列	
	<220> <223> 引物	
	<400> 238 agagagagag agagagtcta gaggccggcc tggccgcctc tccgaactg cgtcatggcc	60
	tcate	65
	<210> 239 <211> 27 <212> DNA <213> 人工序列	
	<220> <223> 引物	
	<400> 239 gccccagatc ctggtacgct ctagagg	27
	<210> 240 <211> 70 <212> DNA <213> 人工序列	
	<220> <223> 引物	

[0120]	<400> 240	
	atcttagcac ctggttggag ggggtttatt gggttttccc aataaacccc ctccaaccag	60
	gtgctaagat	70
	<210> 241	
	<211> 70	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 241	
	atcttagcac ctggttggag ggggtttata gggttttccc tataaacccc ctccaaccag	60
	gtgctaagat	70
	<210> 242	
	<211> 70	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 242	
	atcttagcac ctggttggag ggggtttatc gggttttccc gataaacccc ctccaaccag	60
	gtgctaagat	70
	<210> 243	
	<211> 70	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 243	
	atcttagcac ctggttggag ggggtttatg gggttttccc cataaacccc ctccaaccag	60
	gtgctaagat	70
	<210> 244	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 244	
	ttaaaagcca ggacggtcac	20
	<210> 245	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 245	
	tgtaggagac ccagaagaga	20
	<210> 246	
	<211> 22	
	<212> DNA	
	<213> 人工序列	

	<220>		
	<223>	引物	
	<400>	246	
		acagttttgca ttcattggagg gc	22
	<210>	247	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	引物	
	<400>	247	
		ccgagcgagc aagctcagtt	20
	<210>	248	
	<211>	27	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	引物	
	<400>	248	
		cgcggtatccc cgcccagttgg gactttg	27
	<210>	249	
	<211>	27	
	<212>	DNA	
	<213>	人工序列	
[0121]	<220>		
	<223>	引物	
	<400>	249	
		ccggaattca cctgttagag ctactgc	27
	<210>	250	
	<211>	54	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	引物	
	<400>	250	
		agagagagag agaggcgccc gccctaccag ggatttcagt cgatgtacac gttc	54
	<210>	251	
	<211>	58	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	引物	
	<400>	251	
		aagaagaaga aggaagagaa gtaggccigt catcgfcggg aagacctgeg acacctgc	58
	<210>	252	
	<211>	138	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	引物	

[0122]	<400> 252	
	actgcfatcc gagtataaac cccctccaac caggtataaa cccctccaa ccaggtataa	60
	acccctcca accaggtata aacccctcc aaccaggtat aaacccctc caaccaggat	120
	ctgcgatcta agtaagct	138
	<210> 253	
	<211> 128	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 253	
	ttaattaaga gtctagatta gcacctggtt ggagggggtt tattgcttca tccaaaacca	60
	tggtttacag ggttccaata aacccctcc aaccaggtgc taaagatcta ggaggaattt	120
	aaaatgag	128
	<210> 254	
	<211> 128	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 254	
	actgacctag agaagctttt agcaccctggt tggagggggt ttatigcaac cctgtaaacc	60
	atggttttgg atgaagcaat aaacccctc caaccagggt ctaactgcag ttatttgtac	120
	agttcacc	128
	<210> 255	
	<211> 128	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 255	
	ttaattaaga gtctagatta gcacctggtt ggagggggtt tategcttca tccaaaacca	60
	tggtttacag ggttccgata aacccctcc aaccaggtgc taaagatcta ggaggaattt	120
	aaaatgag	128
	<210> 256	
	<211> 128	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 256	
	actgacctag agaagctttt agcaccctggt tggagggggt ttatcgcaac cctgtaaacc	60
	atggttttgg atgaagcat aaacccctc caaccagggt ctaactgcag ttatttgtac	120
	agttcacc	128
	<210> 257	
	<211> 128	
	<212> DNA	
	<213> 人工序列	

<220>		
<223>	引物	
<400>	257	
ttaattaaga gictagatta gcacctgggt ggaggggggt tatggcttca tccaaaacca	60	
tggtttacag ggttcccata aacccccctc aaccagggtgc taaagatcta ggaggaattt	120	
aaaatgag	128	
<210>	258	
<211>	128	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	258	
actgacctag agaagctttt agcacctgggt tggaggggggt ttatggcaac cctgtaaacc	60	
atggttttgg atgaagccat aaacccccctc caaccagggtg ctaactgcag ttatttgtac	120	
agtccatc	128	
<210>	259	
<211>	144	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
[0123] <400>	259	
actgctatct cgagctataa accccccctca accaggetat aaacccccctc caaccaggct	60	
ataaaacccc tccaaccagg ctataaaccc cctccaacca ggetataaac cccctccaac	120	
caggatctgc gatetaagta agct	144	
<210>	260	
<211>	144	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	260	
actgctatct cgagcaataa accccccctca accaggcaat aaacccccctc caaccaggea	60	
ataaaacccc tccaaccagg caataaaccc cctccaacca ggcaataaac cccctccaac	120	
caggatctgc gatetaagta agct	144	
<210>	261	
<211>	144	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	261	
actgctatct cgagccataa accccccctca accaggccat aaacccccctc caaccaggcc	60	
ataaaacccc tccaaccagg ccataaaccc cctccaacca ggccataaac cccctccaac	120	
caggatctgc gatetaagta agct	144	

[0124]	<210>	262	
	<211>	144	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	引物	
	<400>	262	
		actgctatct cgagcgataa acccccteca accaggcgat aaacccctc caaccaggcg	60
		ataaaccccc tccaaccagg cgataaaccc cctccaacca ggcgataaac cccctccaac	120
		caggatctgc gatctaagta agct	144
	<210>	263	
	<211>	25	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	引物	
	<400>	263	
		tcagaaacag ctcttcttca aatct	25
[0124]	<210>	264	
	<211>	80	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	引物	
	<400>	264	
		ttaattaaga gtctagagga ggcgtgtcca aaaccatggt ttacagcacg cctccagatc	60
		taggaggaaat ttaaaatgag	80
	<210>	265	
	<211>	80	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	引物	
	<400>	265	
		actgacctag agaagcttgg aggcgtgctg taaacatgg ttttgacac gcctccctgc	60
		agttatttgt acagttcatc	80
[0124]	<210>	266	
	<211>	86	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	引物	
	<400>	266	
		ttaattaaga gagatctgct gatgcagata cagaaccaa ggttttctta cttgctgctg	60
		cgcgatctgc atctcaatta gtcagc	86
	<210>	267	
	<211>	82	
	<212>	DNA	
	<213>	人工序列	
	<220>		

[0125]

<223>	引物	
<400>	267	
	caccaccacg gatcgcagc agcaagtaag aaaaccttgg ttctgtatc tgcacagca	60
	atttcgataa gccagtaagc ag	82
<210>	268	
<211>	65	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	268	
	caccaccacg cgcgcaagct tagatctggc ccaggcggcc accatgctga ttggctatgt	60
	aaggg	65
<210>	269	
<211>	35	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	269	
	caccaccaca ccggttcacc atttaggtgg gegac	35
<210>	270	
<211>	21	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	270	
	gttcttgcca ggatccacta g	21
<210>	271	
<211>	21	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	271	
	gcatgtgtcc agatgcatag g	21
<210>	272	
<211>	20	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	272	
	caccttctcc caggataagg	20
<210>	273	
<211>	20	
<212>	DNA	
<213>	人工序列	
<220>		

	<223> 引物	
	<400> 273 gttggcctgt attcctctgg	20
	<210> 274 <211> 22 <212> DNA <213> 人工序列	
	<220> <223> 引物	
	<400> 274 aatgaagtgc ccttggcact tc	22
	<210> 275 <211> 23 <212> DNA <213> 人工序列	
	<220> <223> 引物	
	<400> 275 ctgaagggtt ttaagtgcag aag	23
	<210> 276 <211> 22 <212> DNA <213> 人工序列	
	<220> <223> 引物	
[0126]	<400> 276 tgacgtcaat gacggtaaat gg	22
	<210> 277 <211> 122 <212> DNA <213> 人工序列	
	<220> <223> 引物	
	<400> 277 actgacctag agaagctttt agcacctggt tggagggggt ttataaccct gtaaaccatg	60
	gttttgggaag ciataaacc cctccaacca ggtgctaact gcagttattt gtacagtcca	120
	tc	122
	<210> 278 <211> 140 <212> DNA <213> 人工序列	
	<220> <223> 引物	
	<400> 278 actgacctag agnagctttt agcacctggt tggagggggt ttatagcaac cgcaccctg	60
	taaaccatgg ttttggatga agctgaagct ataaaccccc tccaaccagg tgcctaactgc	120
	agttatttgt acagttcacc	140
	<210> 279 <211> 12	

[0127]	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 279	
	gcgggaggcg tg	12
	<210> 280	
	<211> 17	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 280	
	tcctcatlacc acctgea	17
	<210> 281	
	<211> 15	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 281	
	cagtcagtat caatt	15
	<210> 282	
	<211> 17	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 282	
	cctgcagctc tcatttt	17
	<210> 283	
	<211> 16	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 283	
	attcttccag aattga	16
	<210> 284	
	<211> 17	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 284	
	cagaattgat actgact	17
	<210> 285	
	<211> 16	
	<212> DNA	
	<213> 人工序列	
	<220>	

	<223> 合成构建体	
	<400> 285	
	tcattacacc tgcagc	16
	<210> 286	
	<211> 18	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 286	
	cttccagaat tgatactg	18
	<210> 287	
	<211> 15	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 287	
	ataaaccccc tccaa	15
	<210> 288	
	<211> 74	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
[0128]	<400> 288	
	ttagcacctg gttggagggg gtttataacc aaaaccatgg ttacagtat aaaccccctc	60
	caaccaggtg ctaa	74
	<210> 289	
	<211> 74	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 289	
	ttagcacctg gttggagggg gtttataagg ttttggtacc aaatgtctat aaaccccctc	60
	caaccaggtg ctaa	74
	<210> 290	
	<211> 74	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 290	
	tataaacccc ctccaaccag gtgctaactg taaaccatgg ttttggatta gcacctggtt	60
	ggaggggggtt tata	74
	<210> 291	
	<211> 6	
	<212> PRT	
	<213> 人工序列	

	<220>		
	<223>	合成构建体	
	<220>		
	<221>	misc 特征	
	<222>	(2)..(3)	
	<223>	Xaa 可以是任何天然氨基酸	
	<400>	291	
	Lys Xaa Xaa Gly Ala Arg		
	1 5		
	<210>	292	
	<211>	83	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
	<400>	292	
	tggaattct tccagaattg atactgactg tatggaaaat gagagctgca ggtgtaatga	60	
	agaccttctt tttagatctt ggt	83	
	<210>	293	
	<211>	77	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
[0129]	<400>	293	
	tggaattct tccataattg atattgactg tatggaaaggc tgcgggtgta atgaatacct	60	
	tctttttgag atctggt	77	
	<210>	294	
	<211>	77	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
	<400>	294	
	tggaattct tccagaattg atactgactg tatggaaaac tgcaggtgta atgaagacct	60	
	tctttttgag atctggt	77	
	<210>	295	
	<211>	77	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
	<400>	295	
	tggaattct tccagaattg atactgactg tatggaaagc tgcaggtgta atgaagacct	60	
	tctttttgag atctggt	77	
	<210>	296	
	<211>	76	
	<212>	DNA	
	<213>	人工序列	

	<220>		
	<223>	合成构建体	
	<400>	296	
		tggaaattct tccagaattg atactgactg tatggaagct gcagggtgaa tgaagacctt	60
		ctttttgaga tctggc	76
	<210>	297	
	<211>	73	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
	<400>	297	
		tggaaattct tccagaattg atactgactg tagagctgca ggtgtaatga agaccttctt	60
		tttgagatct ggt	73
	<210>	298	
	<211>	72	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
	<400>	298	
		tggaaattct tccagaattg atactgactg tatgctgcag gtgtaatgaa gaccttctt	60
		ttgagatctg gt	72
[0130]	<210>	299	
	<211>	70	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
	<400>	299	
		tggaaattct tccagaattg atactgactg tgcctgcaggt gtaatgaaga ccttctttt	60
		gagatctggc	70
	<210>	300	
	<211>	83	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
	<400>	300	
		tggaaattct tccagaattg atactgactg tatggaaaat gagagctgca ggtgtaatga	60
		agaccttctt tttgagatct ggt	83
	<210>	301	
	<211>	79	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
	<400>	301	
		tggaaattct tccagaattg atactgactg tatggaaaga gctgcagggtg taatgaagac	60

	cttcttttttg agatctggt	79
	<210> 302	
	<211> 78	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 302	
	tggaaattct tccagaattg atactgactg tatggaagag ctgcaggtgt aatgaagacc	60
	ttctttttga gatctggt	78
	<210> 303	
	<211> 77	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 303	
	tggaaattct tccagaattg atactgactg tatggagagc tgcaggtgta atgaagacct	60
	tcctttttgag atctggt	77
	<210> 304	
	<211> 74	
	<212> DNA	
	<213> 人工序列	
	<220>	
[0131]	<223> 合成构建体	
	<400> 304	
	tggaaattct tccagaattg atactgactg tgagagctgc aggtgtaatg aagaccttct	60
	ttttgagatc tggc	74
	<210> 305	
	<211> 73	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 305	
	tggaaattct tccagaattg atactgactg tagagctgca ggtgtaatga agaccttctt	60
	tttgagatct ggt	73
	<210> 306	
	<211> 72	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 306	
	tggaaattct tccagaattg atactgactg agagctgcag gtgtaatgaa gaccttcttt	60
	ttgagatctg gt	72
	<210> 307	
	<211> 72	

	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 307	
	tggaattct tccaggattg atactgactg agagctgcag ggtgaatgaa gaccttcttt	60
	ttgagatctg gt	72
	<210> 308	
	<211> 72	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 308	
	tggaattct tccagaattg atactgactg agagctgcag ggtgaatgaa gaccttcttt	60
	ttgagatctg gt	72
	<210> 309	
	<211> 72	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 309	
	tggaattct tccagaattg atactgacta ggagctgcag ggtgaatgaa gaccttcttt	60
[0132]	ttgagatctg gt	72
	<210> 310	
	<211> 83	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 310	
	tggaattct tccagaattg atactgactg tatggaaaat gagagctgca ggtgaatga	60
	agaccttctt ttgagatct ggt	83
	<210> 311	
	<211> 81	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 311	
	tggaattct tccagaattg atactgacta tggaaaatga gagctgcagg tgaatgaag	60
	accttctttt tgagatctgg t	81
	<210> 312	
	<211> 80	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	

[0133]	<400> 312	
	tggaattct tccagaattg atactgataf ggaaatgag agctgcaggt gtaatgaaga	60
	ctttcttttt gagatctggt	80
	<210> 313	
	<211> 74	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 313	
	tggaattct tccagaattg atattgaaaa tgagagctgc aggtgtaatg aagaccttct	60
	ttttgagatc tgg	74
	<210> 314	
	<211> 83	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 314	
	tggaattct tccagaattg atactgactg tatggaaaat gagagctgca ggtgtaatga	60
	agaccttctt tttgagatct ggt	83
[0134]	<210> 315	
	<211> 79	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 315	
	tggaattct tccagaattg atactgactg tatgatgaga gctgcaggtg taatgaagac	60
	ctttctttttg agatctggt	79
	<210> 316	
	<211> 79	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 316	
	tggaattct tccagaattg atactgactg tatggtgaga gctgcaggtg taatgaagac	60
	ctttctttttg agatctggt	79
[0135]	<210> 317	
	<211> 78	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 317	
	tggaattct tccagaattg atactgactg taaatgagag ctgcaggtgt aatgaagacc	60
	ttcttttttga gatctggt	78

	<210> 318	
	<211> 78	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 318	
	tggaaattct tccagaattg atactgactg tatgtgagag ctgcagggtg aatgaagacc	60
	ticttttttga gatctggt	78
	<210> 319	
	<211> 74	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 319	
	tggaaattct tccagaattg atactgaaaa tgagagctgc aggtgtaatg aagaccttct	60
	ttttgagatc tgggt	74
	<210> 320	
	<211> 74	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
[0134]	<400> 320	
	tggaaattct tccagaattg atactgactg tatggactgc aggtgtaatg aagaccttct	60
	ttttgagatc tgggt	74
	<210> 321	
	<211> 74	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 321	
	tggaaattct tccagaattg atactgactg tgagagctgc aggtgtaatg aagaccttct	60
	ttttgagatc tgggt	74
	<210> 322	
	<211> 73	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 322	
	tggaaattct tccagaattg atactgggtat gagagctgca ggtgtaatga agaccttctt	60
	tttgagatct ggt	73
	<210> 323	
	<211> 71	
	<212> DNA	
	<213> 人工序列	
	<220>	

	<223> 合成构建体	
	<400> 323	
	tggaattct tccagaattg atactgactg tagctgcagg tgtaatgaag accttctttt	60
	tgagatctgg t	71
	<210> 324	
	<211> 83	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 324	
	tggaattct tccagaattg atactgactg tatggaaaat gagagctgca ggtgtaatga	60
	agaccttctt tttagatct ggt	83
	<210> 325	
	<211> 75	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 325	
	tggaattct tccagaattg atactgactg tatggaaact gcggtgtaat gaagaccttc	60
	tttttgagat ctggt	75
[0135]	<210> 326	
	<211> 76	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 326	
	tggaattct tccagaattg atactgactg tatgagagct gcaggtgtaa tgaagacctt	60
	ctttttgaga tctggt	76
	<210> 327	
	<211> 74	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 327	
	tggaattct tccagaatta atactgactg tgagagctgc aggtgtaatg aagaccttct	60
	ttttgagate tgggt	74
	<210> 328	
	<211> 73	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 328	
	tggaattct tccagaattg atactgactg gagagctgca ggigtatga agaccttctt	60
	tttgagatct ggt	73

	<210>	329	
	<211>	72	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
	<400>	329	
		tggaaattct tccagaattg atactgactg agagctgcag gtgtaatgaa gaccttcitt	60
		ttgagatctg gt	72
	<210>	330	
	<211>	68	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
	<400>	330	
		tggaaattct tccagaattg atactgactg ctgcaggtgt aatgaagacc ttcttttga	60
		gatctggt	68
	<210>	331	
	<211>	65	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
[0136]	<400>	331	
		tggaaattct tccagaattg atactgactg tatgtgtaat gaagaccctc tttttgagat	60
		ctggt	65
	<210>	332	
	<211>	39	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
	<400>	332	
		tggaaattct tccagaattg atactttttg agatctggt	39
	<210>	333	
	<211>	24	
	<212>	PRT	
	<213>	人工序列	
	<220>		
	<223>	合成构建体	
	<400>	333	
		Arg Asn Asn Arg Asn Asn Arg Asn Asn Arg Asn Asn Asn Asn Tyr Asn	
		1 5 10 15	
		Asn Tyr Asn Asn Tyr Asn Asn Tyr	
		20	
	<210>	334	
	<211>	24	

	<212> PRT	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 334	
	Arg Asn Asn Arg Asn Asn Arg Asn Asn Arg Asn Asn Asn Tyr Asn	
	1 5 10 15	
	Asn Tyr Asn Asn Tyr Asn Asn Tyr	
	20	
	<210> 335	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 335	
	tccaaaacca tggtttacag	20
	<210> 336	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
[0137]	<400> 336	
	tccaaaacca tggtttacag	20
	<210> 337	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 337	
	tccaaaacca tggtttacag	20
	<210> 338	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 338	
	tccaaaacca tggtttacag	20
	<210> 339	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 339	
	tccaaaacca tggtttacag	20

<210> 340
 <211> 38
 <212> DNA
 <213> 人工序列

<220>
 <223> 合成构建体

<220>
 <221> misc_特征
 <222> (17)...(18)
 <223> n 是 a, c, g, 或 t

<220>
 <221> misc_特征
 <222> (21)...(22)
 <223> n 是 a, c, g, 或 t

<400> 340
 ggaggcgtgt ccaaaannat nntttacagc acgcctcc

38

<210> 341
 <211> 144
 <212> PRT
 <213> 人工序列

<220>
 <223> 合成构建体

<400> 341

Met Arg Leu Phe Gly Tyr Ala Arg Val Ser Thr Ser Gln Gln Ser Leu
 1 5 10 15

[0138]

Asp Ile Gln Val Arg Ala Leu Lys Asp Ala Gly Val Lys Ala Asn Arg
 20 25 30

Ile Phe Thr Asp Lys Ala Ser Gly Ser Ser Cys Asp Arg Lys Gly Leu
 35 40 45

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

Lys Leu Asp Arg Leu Gly Arg Asp Thr Ala Asp Met Ile Gln Leu Ile
 65 70 75 80

Lys Glu Phe Asp Ala Gln Gly Val Ser Ile Arg Phe Ile Asp Asp Gly
 85 90 95

Ile Ser Thr Asp Gly Glu Met Gly Lys Met Val Val Thr Ile Leu Ser
 100 105 110

Ala Val Ala Gln Ala Glu Arg Gln Arg Ile Leu Glu Arg Thr Asn Glu
 115 120 125

Gly Arg Gln Glu Ala Met Ala Lys Gly Val Val Phe Gly Arg Lys Arg
 130 135 140

<210> 342
 <211> 141
 <212> PRT
 <213> 人工序列

<220>
 <223> 合成构建体

<400> 342

Met. Leu. Ile. Gly. Tyr. Val. Arg. Val. Ser. Thr. Asn. Asp. Gln. Asn. Thr. Asp
1 5 10 15

Leu. Gln. Arg. Asn. Ala. Leu. Val. Cys. Ala. Gly. Cys. Glu. Gln. Ile. Phe. Glu
20 25 30

Asp. Lys. Leu. Ser. Gly. Thr. Arg. Thr. Asp. Arg. Pro. Gly. Leu. Lys. Arg. Ala
35 40 45

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

Arg. Leu. Gly. Arg. Ser. Met. Lys. His. Leu. Ile. Ser. Leu. Val. Gly. Glu. Leu
65 70 75 80

Arg. Glu. Arg. Gly. Ile. Asn. Phe. Arg. Ser. Leu. Thr. Asp. Ser. Ile. Asp. Thr
85 90 95

Ser. Ser. Pro. Met. Gly. Arg. Phe. Phe. Phe. His. Val. Met. Gly. Ala. Leu. Ala
100 105 110

Glu. Met. Glu. Arg. Glu. Leu. Ile. Ile. Glu. Arg. Thr. Met. Ala. Gly. Leu. Ala
115 120 125

Ala. Ala. Arg. Asn. Lys. Gly. Arg. Ile. Gly. Gly. Arg. Pro. Pro
130 135 140

[0139]

<210> 343

<211> 44

<212> DNA

<213> 人工序列

<220>

<223> 合成构象体

<220>

<221> misc_特征

<222> (2)..(3)

<223> n 是 a, c, g, 或 t

<220>

<221> misc_特征

<222> (5)..(6)

<223> n 是 a, c, g, 或 t

<220>

<221> misc_特征

<222> (8)..(9)

<223> n 是 a, c, g, 或 t

<220>

<221> misc_特征

<222> (11)..(16)

<223> n 是 a, c, g, 或 t

<220>

<221> misc_特征

<222> (21)..(21)

<223> n 是 a, c, g, 或 t

<220>

<221> misc_特征

<222> (24)..(24)

	<223> n 是 a, c, g, 或 t	
	<220>	
	<221> misc_特征	
	<222> (29)..(34)	
	<223> n 是 a, c, g, 或 t	
	<220>	
	<221> misc_特征	
	<222> (36)..(37)	
	<223> n 是 a, c, g, 或 t	
	<220>	
	<221> misc_特征	
	<222> (39)..(40)	
	<223> n 是 a, c, g, 或 t	
	<220>	
	<221> misc_特征	
	<222> (42)..(43)	
	<223> n 是 a, c, g, 或 t	
	<400> 343	
	raarnrnnr nanunnaaab nwnvtttan mnnyynyyn yyny	44
	<210> 344	
	<211> 44	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 344	
	gtgatgcag atacagaaac caaggtttc ttacttgctg ctgc	44
[0140]	<210> 345	
	<211> 44	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 345	
	gtgatggag cagccaatag gtccctttc tcccttag cccc	44
	<210> 346	
	<211> 44	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 346	
	agggaagta atccagaaac caccctttat ccttccctgt cctt	44
	<210> 347	
	<211> 44	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 347	
	ggaaatgta aagtagaaac taaagtttc gcttccatc ttcc	44
	<210> 348	
	<211> 44	

[0141]

<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成构建体	
<400>	348	
	ggaagaagga tgagagaaac taacctttgt ggaacccctg cagc	44
<210>	349	
<211>	44	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成构建体	
<400>	349	
	aacggcagaa gaagaaaaat tatactttct ttccattgt ttcc	44
<210>	350	
<211>	44	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成构建体	
<400>	350	
	gaggtaaata ctgataaat gtgcttttt tccccatta ccct	44
<210>	351	
<211>	44	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成构建体	
<400>	351	
	attgtgatg gagtaaaaat gatcctttta tacatttcta catt	44
<210>	352	
<211>	44	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成构建体	
<400>	352	
	ataggagaaa atttggaag tataattttt cagactactc tttt	44
<210>	353	
<211>	43	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成构建体	
<400>	353	
	acagaagaca ttaagaaaac ctaacttgac ctccatggt tcc	43
<210>	354	
<211>	44	
<212>	DNA	
<213>	人工序列	
<220>		

	<223> 合成构建体	
	<400> 354 ggcaggacag ctaactaatg aaaggtttgg tgtgtgtctg tctt	44
	<210> 355 <211> 44 <212> DNA <213> 人工序列	
	<220> <223> 合成构建体	
	<400> 355 agggatgagg cctcataaag taaagttttt tgtttgtttg ttcc	44
	<210> 356 <211> 44 <212> DNA <213> 人工序列	
	<220> <223> 合成构建体	
	<400> 356 acagtcaaag tatctgaaag ttaacttttt tegtacagtc ttcc	44
	<210> 357 <211> 44 <212> DNA <213> 人工序列	
	<220> <223> 合成构建体	
[0142]	<400> 357 gaaattgtgg acaattaaat tatcctttct gggeccctta ttcc	44
	<210> 358 <211> 44 <212> DNA <213> 人工序列	
	<220> <223> 合成构建体	
	<400> 358 gnaattggaa ggaaaaaaat tatcctttat ggtgtaatac ttat	44
	<210> 359 <211> 44 <212> DNA <213> 人工序列	
	<220> <223> 合成构建体	
	<400> 359 aaaacagctg gctttgaaag gaaactttta acatactacc tggc	44
	<210> 360 <211> 44 <212> DNA <213> 人工序列	
	<220> <223> 合成构建体	
	<400> 360 atagtaagtg ctcaataaat gtccgtttat atcactcatg ttgc	44

	<210> 361	
	<211> 44	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 361	
	aaagatggaa caaacaaaat taaggittag tacattataa ttcc	44
	<210> 362	
	<211> 43	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 362	
	gcgggaggcg tgtccaaacc atggittaca gcacgcctcc cgc	43
	<210> 363	
	<211> 41	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 363	
	gctgatgcag atcgagaaac caaggttttc ttacttgcctg ctgc	44
[0143]	<210> 364	
	<211> 44	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 364	
	gtggatggag cagccaatag gtctctttcc tcccccttag cccc	44
	<210> 365	
	<211> 44	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 365	
	agggaagtca atccagaaac catcttttat cccttctgt cett	44
	<210> 366	
	<211> 35	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 366	
	agatacagaa accgttttct taettgetgc tggcc	35
	<210> 367	
	<211> 38	

	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<400> 367	
	tcagggaagt catcctttat ccccttcctgt ccttagct	38
	<210> 368	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<220>	
	<221> misc_特征	
	<222> (8)..(9)	
	<223> n 是 a, c, g, 或 t	
	<220>	
	<221> misc_特征	
	<222> (12)..(13)	
	<223> n 是 a, c, g, 或 t	
	<400> 368	
	tccasssna tnnssacag	20
[0144]		
	<210> 369	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成构建体	
	<220>	
	<221> misc_特征	
	<222> (1)..(4)	
	<223> n 是 a, c, g, 或 t	
	<220>	
	<221> misc_特征	
	<222> (8)..(9)	
	<223> n 是 a, c, g, 或 t	
	<220>	
	<221> misc_特征	
	<222> (12)..(13)	
	<223> n 是 a, c, g, 或 t	
	<220>	
	<221> misc_特征	
	<222> (17)..(20)	
	<223> n 是 a, c, g, 或 t	
	<400> 369	
	nnnnaaanna tnnnttnnnn	20

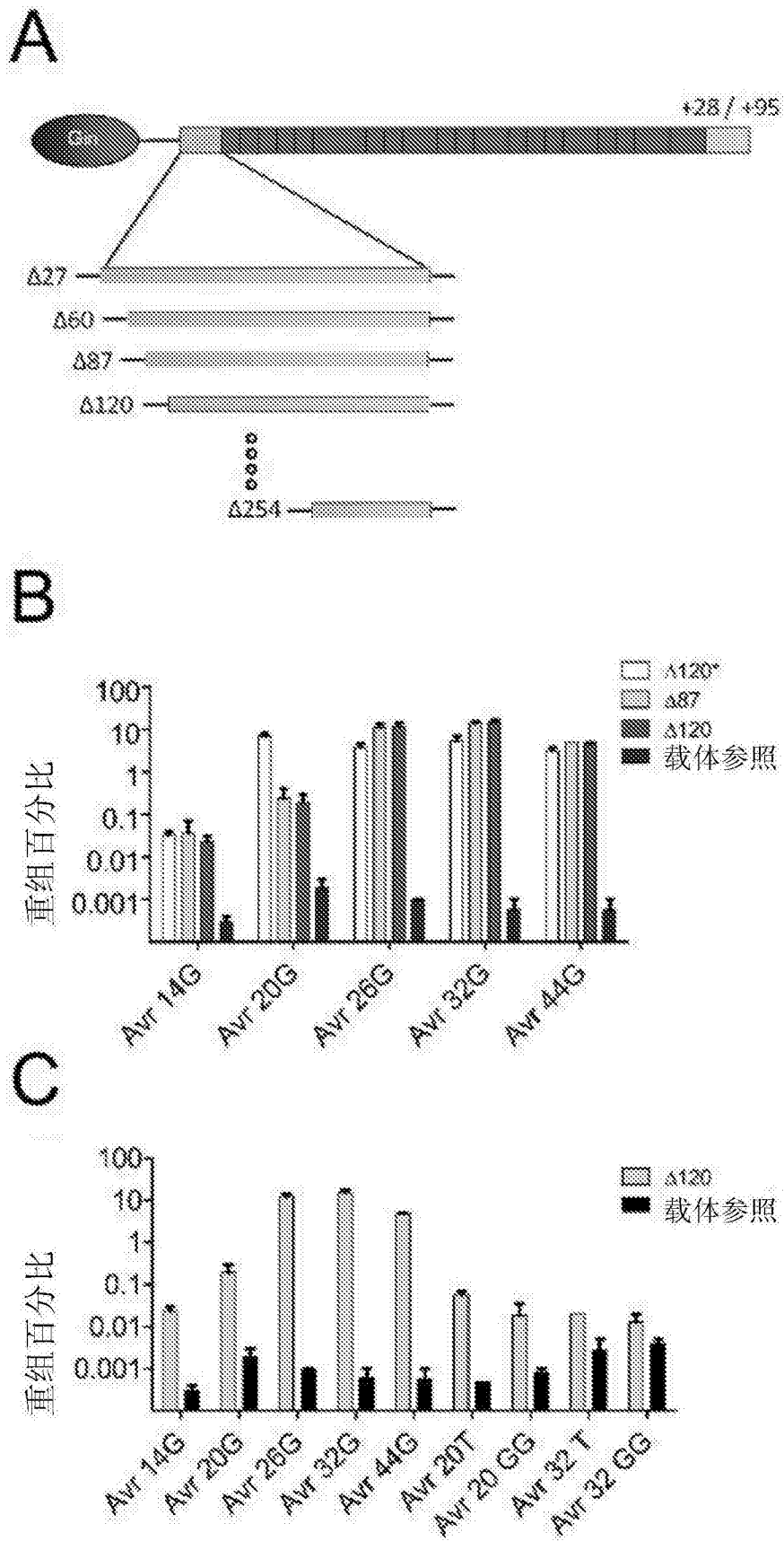


图2

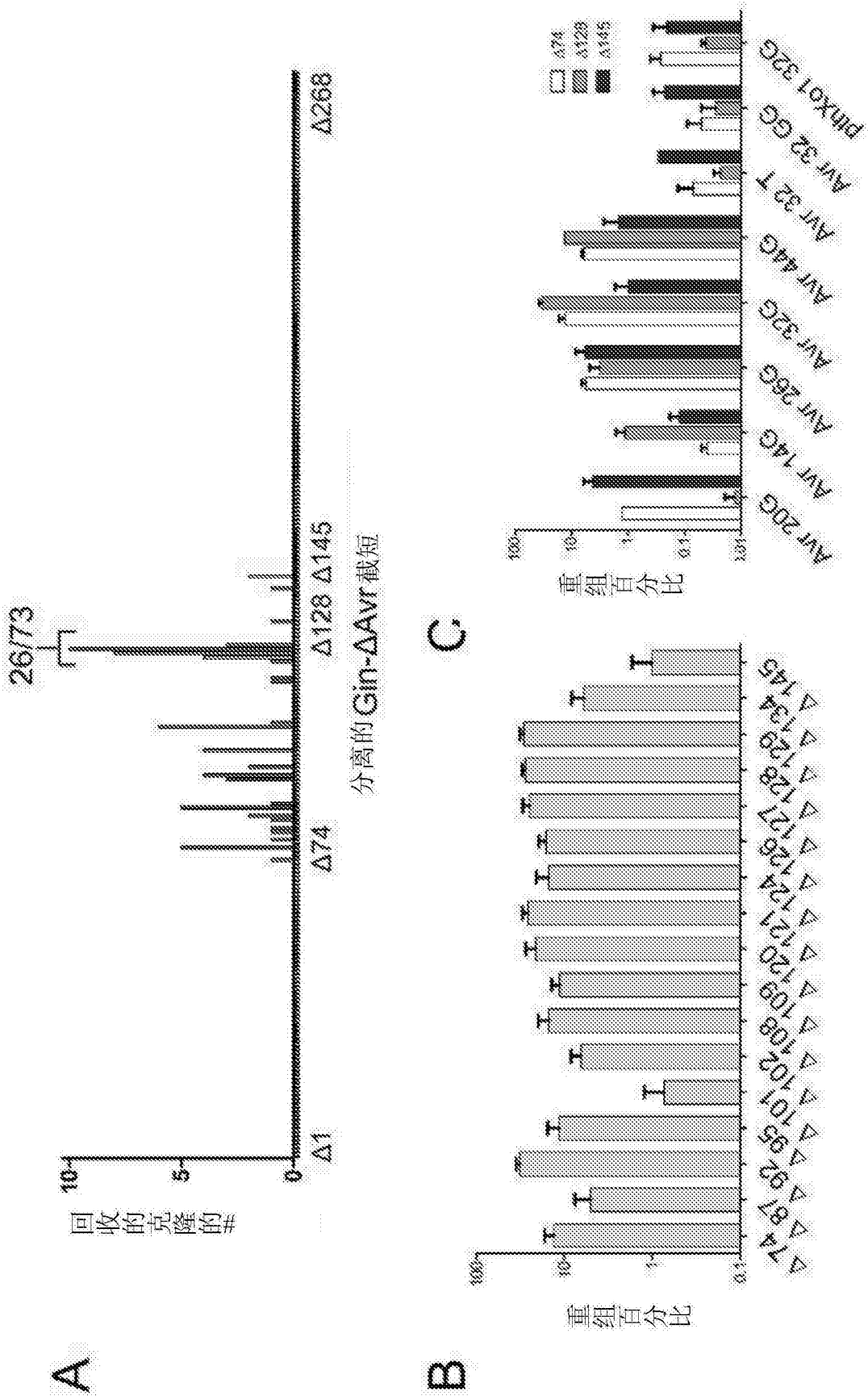


图3

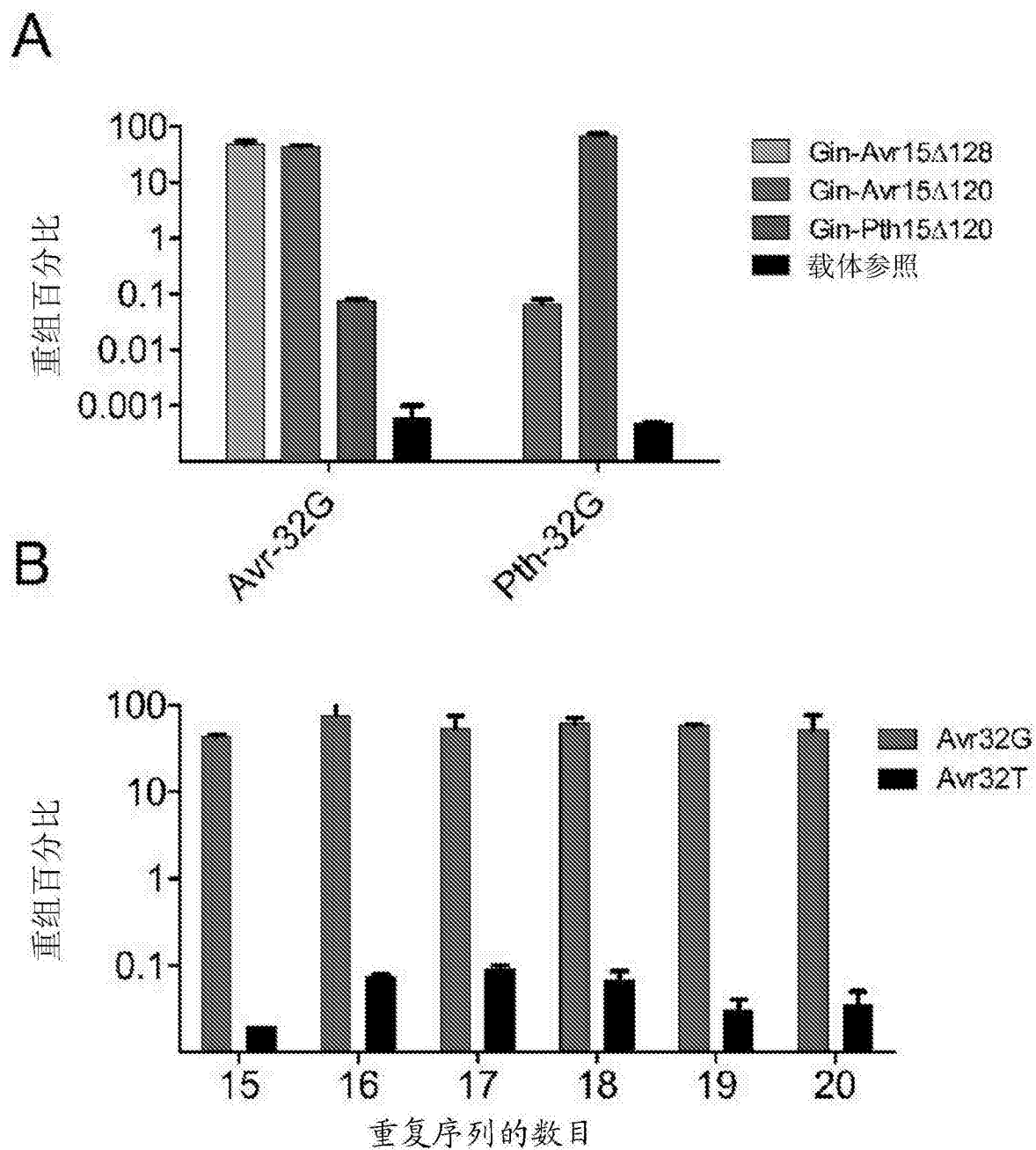


图4

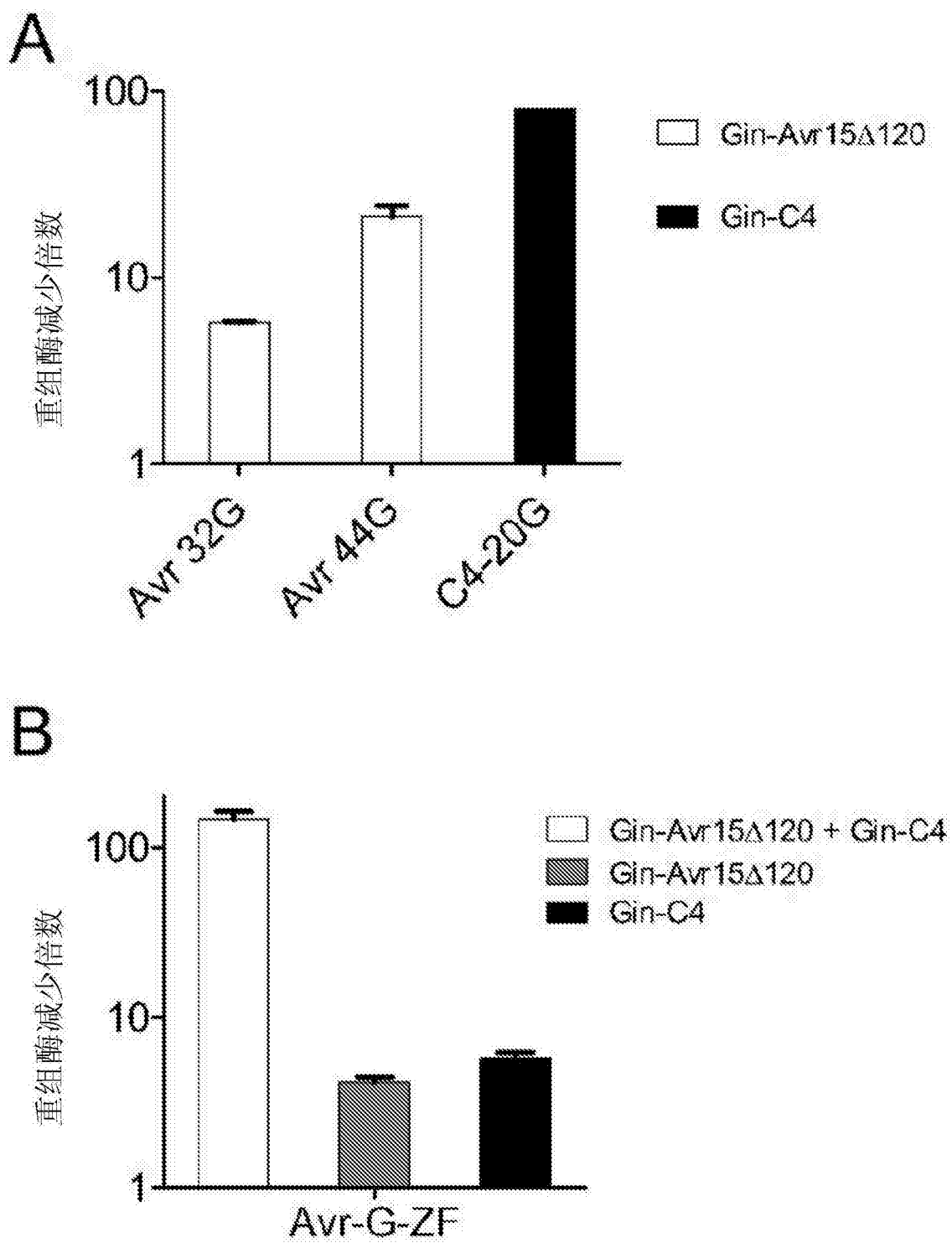


图5



AvrXa7	NI HG NI NI NS HD NN HD HD NS N* N* HD HD NS NS NN NN NI NG NN NI N* NS N*
靶序列	A T A A C C C C T C C A A C C A G G T G C T A A
合成	NI NG NI NI NI HD HD HD HD NG HD HD NI NI HD HD NI NN NN NG NN HD NG NI NI

图7

TALE 和 TALER 序列

AvrXa7 蛋白

MDPIRSRTPSPARELLPGPQPDRVQPTADRGGAPPAGGPLDGLPARRTMSRTRLPSPPAPSPA
 FSAGSFSDLLRQFDPSLLDTSLLDSMPAVGTPHTAAAPAECDEVQSGLRAADDPPTVRVAV
 TAARPPRAKPAPRRRAAQPSDASPAQVLDRLTLGYSQQQKEKPKVVRSTVAQHHEALVGH
 GFTHAHIVALSQHPAALGTVAVTYQDIIRALPEATHEDIVGVGKQWSGARALEALLTEAGEL
 RGPPLQLDTGQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTDPQVVAIASNIGGKQALET
 VQRLLPVLCQDHGLTPDQVVAIASHGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIG
 GKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPAQVV
 AIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETLQRLLPVLCQDHG
 LTPDQVVAIANNGGKQALETLQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLP
 VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPAQVVAIASHDGGKQAL
 ETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASN
 GGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVQRLLPVLCQDHG
 LTQDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLP
 VLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALE
 TVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIAN
 NNGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTLDQ
 VVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIANNGGKQALETVQRLLPVLCQ
 DHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTLDQVVAIASNSGGKQALETVQR
 LPVLCQDHGLTPNQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPNQVVAIASNSGGKQAL
 ESIVAQLSRPDPALAAALNDHLVALACLGGRPALDAVKKGLPHAPELIRNRIRIPERTSHRV
 PDLAHVVRVLGFFQSHSHPAQAFDDAMTQFEMSRHGLVQLFRRVGVTEFEARYGTLPPASQ
 RWDRIQLASGMKRAKPSPTSAQTPDQASLHAFADSLERDLDAAPSPMHEGDQTRASSRKRSR
 SDRAVTGPSTQQSFEVRVPEQQDALHLPLSWRVKRPRTRIGGGPLDPGTPIAADLAASSTVM
 WEQDAAPFAGAADDFAFNEEELAWLMELLPQSGSVGGTI

图8

AvrXa7 基因

ATGGATCCCATTCGTTTCGCGCACGCCAAGTCTCTGCCCGCGAGCTTCTGCCCCGGACCCCA
 ACCGGATAGGGTTCAGCCGACTGCAGATCGGGGGGGGGCTCCGCCTGCTGGCGGCCCC
 CTGGATGGCTTGCCCGCTCGGCGGACGATGTCCCGGACCCGGCTGCCATCTCCCCCTGC
 GCCCTCGCTGCGTTCTCGGCGGGCAGCTTCAGCGATCTGCTCCGTCAGTTCGATCCGTC
 GCTTCTTGATACATCGCTTCTTGATTGATGCTGCGTCGCGCACGCCGCATACAGCGGC
 TGCCCCAGCAGAGTGCGATGAGGTGCAATCGGGTCTGCGTGCAGCCGATGACCCGCCAC
 CCACCGTGCGTGTGCTGCTACT_{gag}GCGCGGCCGCCGCGCGCCAAAGCCGGCCCCGCGAC
 GCGTGCAGCGCAACCCTCCGACGCTTCGCCGGCCGCGCAGGTGGATCTACGCACGCTC
 GGCTACAGTCAGCAGCAGCAAGAGAAGATCAAACCGAAGGTGCGTTCGACAGTGCGCGC
 AGCACCACGAGGCACTGGTGGGCCATGGGTTTACACACGCGCACATCGTTGCGCTCAGC
 CAACACCCGGCAGCGTTAGGGACCGTTGCTGTACGTATCAGGACATAATCAGGGCGTT
 GCCAGAGGCGACACACGAAGACATCGTTGGCGTCCGCAAACAGTGGTCCGGCGCACGC
 GCTCTGGAGGCCTTGCTACGGAGGCGGGGGAGTTGAGAGGTCCGCCGTTACAGTTGGA
 CACAGGCCAACTTCTCAAGATTGCAAAACGTGGCGGCGTGACCGCAGTGGAGGCAGTG
 CATGCATGGCGCAATGCACTGACGGGTGCCCCCTGAACCTGACCCCGGACCAAGTGGT
 GGCCATCGCCAGCAATATTGGCGGCAAGCAGGCGCTGGAGACGGTACAGCGGCTGTTG
 CCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAGGTCTGGCCATCGCCAGCCA
 TGGCGGCGGCAAGCAGGCGCTGGAGACGGTGC

图9

AGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAGGTGGTGGCC
ATCGCCAGCAATATTGGCGGCAAGCAGGCGCTAGAGACGGTGCAGCGGCTGTTGCCGG
TGCTGTGCCAGGCCCATGGCCTGACCCCGGACCAGGTCGTGGCCATCGCCAGCAATATT
GGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACC
ATGGCCTGACCCCGGCCAGGTGGTGGCCATCGCCAGCAATAGTGGCGGCAAGCAGGC
GCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGG
ACCAAGTCGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGGCGCTGGAGACGCTGCA
GCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAGGTGCTGGCCA
TCGCCAACAATAACGGCGGCAAGCAGGCGCTGGAGACGCTGCAGCGGCTGTTGCCGGT
GCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCCATCGCCAGCCACGATG
GCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCA
TGGCCTGACCCCGGACCAGGTGGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGGCG
CTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGC
CCAAGTGGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGGCGCTGGAGACGGTGCAG
CGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAGGTGGTGGCCAT
CGCCAGCAATAGCGGCGGCAAGCAGGCGCTGGAGACGGTACAGCGGCTGTTGCCGGTG
CTGTGCCAGGACCATGGACTGACCCCGGACCAGGTGCTGGCCATCGCCAGCAATGGCG
GCAAGCAGGCGCTGGAGACGGTACAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGG
CCTGACCCCGGACCAGGTGCTGGCCATCGCCAGCAATGGCGGCAAGCAGGCGCTGGAG
ACGGTGCAGCGGCTGTTGCCGGTACAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGG
CCTGACCCAGGACCAGGTGGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGGCGCTG
GAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCA
AGTGGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGG
CTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAGGTGGTGGCCATCGC
CAGCAATAGTGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTG
TGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCCATCGCCAGCAATAGTGGCG
GCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGG
CCTGACCCCGGACCAGGTGGTGGCCATCGCCAGCAATAACGGCGGCAAGCAGGCGCTG
GAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCA
GGTCGTGGCCATCGCCAACAATAACGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGG
CTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGCGCAGGTGGTGGCCATCGC
CAGCAATATTGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGT
GCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCCATTGCCAGCAATGGCGGCAG
CAAACAGGCGCTAGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGC
CTGACCCCGGACCAAGTGGTGGCCATCGCCAACAATAACGGCGGCAAGCAGGCGCTGG
AGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAG
GTCGTGGCCATCGCCAGCAATATTGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGC
TGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAGGTGGTGGCCATCGCC
AGCAATGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCC
AGGACCATGGCCTGACCCCGAACCAGGTGGTGGCCATCGCCAGCAATAGTGGCGGCAA
GCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGA
CCCCGAACCAGGTGGTGGCCATCGCCAGCAATGGCGGCAAGCAGGCGCTGGAGAGCAT
TGTTGCCAGTTATCTCGCCCTGATCCGGCGTTGGCCGCGTTGACCAACGACCACCTCGT
CGCCTTGGCCTGCCTCGGCGGACGTCCTGCCCTGGATGCAGTGAAAAAGGGATTGCCGC
ACGCGCCGGAATTGATCAGAAGAATCAATCGCCGATTCCCAGACGACGTCCTCATCGC
GTTCCCGACCTCGCGCACGTGGTTCCGCTGCTTGGTTTTTCCAGAGCCACTCCACCCA
GCGCAAGCATTTCGATGACGCCATGACGCAGTTCGAGATGAGCAGGCACGGCTTGGTAC
AGCTCTTTCGACAGAGTGGGCGTCACCGAATTCGAAGCCCGCTACGGAACGCTCCCCCA
GCCTCGCAGCGTTGGGACCGTATCCTCCAGGCATCAGGG

图9(连续)

ATGAAAAGGGCCAAACCGTCCCCTACTTCAGCTCAAACACCGGATCAGGCGTCTTTGCA
TGCATTGCGCCGATTGCTGGAGCGTGACCTTGATGCGCCAGCCCAATGCACGAGGGAG
ATCAGACGCGGGCAAGCAGCCGTAAACGGTCCCGATCGGATCGTGCTGTCACCGGCCCC
TCCACACAGCAATCTTTCGAGGTGCGCGTTCCCGAACAGCAAGATGCGCTGCATTTGCC
CCTCAGCTGGAGGGTAAAACGCCCGCGTACCAGGATCGGGGGCGGCCTCCCGGATCCT
GGTACGCCCATCGCTGCCGACCTGGCAGCGTCCAGCACCGTGATGTGGGAACAAGATGC
GGCCCCCTTCGCAGGGGCAGCGGATGATTTCCCGGCATTCAACGAAGAGGAGCTCGCAT
GGTTGATGGAGCTATTGCCTCAGTCAGGCTCAGTCGGAGGGACGATCTGA

图9 (连续)

Gin-AvrA74

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVV
WKLDRLGSRMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIERT
MAGLAAARNKGRIGGRPRKSGSGSPRQFDPSSLDTSLDLSMPAVGTPHTAAAPAECDEVQS
GLRAADDPPPTVRVAVTAARPPRAKPAPRRRAAQPSDASPAQVDLRTLGYSSQQQEKIKP
KVRSTVAQHHEALVGHGFTAHIVALSQHPAALGTVAVTYQDIIRALPEATHEDIVGVGKQ
WSGARALEALLTEAGELRGPPQLDGTGQLLKIARGGVTAVEAVHAWRNALTGAPLNLTP
DQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHGGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQ
RLLPVLCQDHGLTPAQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGK
QALETLQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETLQRLLPVLCQDHGLTPDQVVAI
ASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGL
TPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETV
QRLLPVQRLLPVLCQDHGLTQDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAI
ASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGL
TPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALET
VQRLLPVLCQDHGLTLDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIANNN
GGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTLDQV
VAIASNSGGKQALETVQRLLPVLCQDHGLTPNQVVAIASNSGGKQALETVQRLLPVLCQDHG
LTPNQVVAIASNSGGKQALESIVAQLSRPDPALAAALTNDHLVALACLG

图10

Gin-AvrA87

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGD TLVV
 WKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFFVVMGALAEMERELIERT
 MAGLAAARNKGRIGGRPRKSGSGSPDSMPAVGTPHTAAAPAECDEVQSGLRADDPPPTV
 RVAVTAARPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSSQQQKEKIKPKVRSTVAQHHE
 ALVGHGFTAHIVALSQHPAALGTVAVTYQDIIRALPEATHEDIVGVGKQWSGARALEALL
 TEAGELRGPPQLDGTGQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTDPQVVAIASNIGG
 KQALETVQRLLPVLCQDHGLTPDQVVAIASHGGGKQALETVQRLLPVLCQDHGLTPDQVV
 AIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGL
 TPAQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETLQRLLPV
 LCQDHGLTPDQVVAIANNNGGKQALETLQRLLPVLCQDHGLTPDQVVAIASHDGGKQALE
 TVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPAQVVAIASHD
 GGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQ
 VVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVQRLLPV
 LCQDHGLTQDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALE
 TVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNS
 GGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQ
 VVAIANNNGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQD
 HGLTLDQVVAIASNGGSKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETVQR
 LLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTLDQVVAIASNGGKQA
 LETVQRLLPVLCQDHGLTPNQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPNQVVAIAS
 NGGKQALESIVAQLSRPDPALAAALTNDHLVALACLG

图11

Gin-AvrA120

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGD TLVV
 WKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFFVVMGALAEMERELIERT
 MAGLAAARNKGRIGGRPRKSGSGSTVRVAVTAARPPRAKPAPRRRAAQPSDASPAAQVDL
 RTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTAHIVALSQHPAALGTVAVTYQDIIRA
 LPEATHEDIVGVGKQWSGARALEALLTEAGELRGPPQLDGTGQLLKIARKGGVTAVEAVHA
 WRNALTGAPLNLTDPQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHGGGK
 QALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAI
 ASNIGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNSGGKQALETVQRLLPVLCQDHGLT
 PDQVVAIASHDGGKQALETLQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETLQRLLPV
 LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALET
 VQRLLPVLCQDHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSG
 GKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVV
 AIASNSGGKQALETVQRLLPVQRLLPVLCQDHGLTQDQVVAIASHDGGKQALETVQRLLPV
 LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALET
 VQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNG
 GKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPAQV
 VAIASNIGGKQALETVQRLLPVLCQDHGLTLDQVVAIASNGGSKQALETVQRLLPVLCQDH
 GLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLL
 PVLCQDHGLTLDQVVAIASNGGKQALETVQRLLPVLCQDHGLTPNQVVAIASNSGGKQALE
 TVQRLLPVLCQDHGLTPNQVVAIASNSGGKQALESIVAQLSRPDPALAAALTNDHLVALACLG

图12

Gin-AvrA120*

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGD TLVV
 WKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFFVVMGALAEMERELIERT
 MAGLAAARNKGRIGGRPRKSGSGSTVRVAVTAARPPHAVAGPAAQVDLRTLGYSSQQQEQE
 KIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQDIIRALPEATHEDIVGV
 GKQWSGARALEALLTEAGELRGPPLQLDTGQLLKIARKGGVTAVEAVHAWRNALTGAPLN
 LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHGGGKQALETVQRLLP
 VLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALE
 TVQRLLPVLCQDHGLTPAQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHD
 GGGKQALETLQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETLQRLLPVLCQDHGLTPDQ
 VVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQ
 DHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQ
 RLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQ
 ALETVQRLLPVQRLLPVLCQDHGLTQDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQ
 VVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQD
 HGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLL
 PVLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQ
 ALETVQRLLPVLCQDHGLTLDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIAN
 NNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTLD
 QVVAIASNSGGKQALETVQRLLPVLCQDHGLTPNQVVAIASNSGGKQALETVQRLLPVLCQD
 HGLTPNQVVAIASNSGGKQALESIVAQLSRPDPALAAALTNDHLVALACLG

图13

Gin-AvrA147

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGD TLVV
 WKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFFVVMGALAEMERELIERT
 MAGLAAARNKGRIGGRPRKSGSGSPASPAQVDLRTLGYSSQQQEQEKIKPKVRSTVAQHHE
 ALVGHGFTHAHIVALSQHPAALGTVAVTYQDIIRALPEATHEDIVGVGKQWSGARALEALL
 TEAGELRGPPLQLDTGQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTDPDQVVAIASNIGG
 KQALETVQRLLPVLCQDHGLTPDQVVAIASHGGGKQALETVQRLLPVLCQDHGLTPDQVV
 AIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGL
 TPAQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGGKQALETLQRLLPV
 LCQDHGLTPDQVVAIANNNGGKQALETLQRLLPVLCQDHGLTPDQVVAIASHDGGGKQALE
 TVQRLLPVLCQDHGLTPDQVVAIASHDGGGKQALETVQRLLPVLCQDHGLTPAQVVAIASHD
 GGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQ
 VVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVQRLLPV
 LCQDHGLTQDQVVAIASHDGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGGKQALE
 TVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNS
 GGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQ
 VVAIANNNGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQD
 HGLTLDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETVQR
 LLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTLDQVVAIASNSGGKQ
 ALETVQRLLPVLCQDHGLTPNQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPNQVVAIAS
 NSGGKQALESIVAQLSRPDPALAAALTNDHLVALACLG

图14

GinAvr15A128-合成蛋白

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLOKGD TLVV
WKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFFYVMGALAEMERELIERT
MAGLAAARNKGRIGGRPRKSGSGSPALRPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSSQ
QQQEKIKPKVRSTVAQHHEALVGHGFTAHIVALSQHPAALGTVAVTYQHIITALPEATHED
IVGVGKQWSGARALEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNALTG
APLNLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQ
RLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQ
ALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS
HDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTP
DQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALET
VQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDG
GKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVV
AIASNIGGKQALESIVAQLSRPDPALAAALTNDHLVALACLGPKKKRKV

图15

Gin-Avr15A128-合成 DNA:

ATGCTGATTGGCTATGTAAGGGTATCAACAAATGACCAGAATACAGACCTGCAACGAA
ACGCTCTTGTGTTGTGCAGGATGTGAACAAATATTTGAAGATAAATTAAGCGGAACAAGG
ACAGACCGACCGGGATTAAAACGCGCTTTAAAGCGCCTTCAAAAAGGTGACACACTGG
TTGTCGGAACCTGGATCGCCTCGGGCGAAGCATGAAACATTTGATTTCTCTCGTAGGG
GAATTACGAGAGCGAGGGATTAAATTTTCGCAGTCTTACTGACAGTATTGATACGTCATC
TCCAATGGGGCGTTTTTCTTCTACGTTATGGGTGCCCTGGCTGAAATGGAACGAGAAGT
AATTATCGAGCGAACGATGGCTGGACTTGCTGCCGCCAGAAATAAAGGCCGTATTGGAG
GTCGCCCCGCTAAATCGGGGTCTGGATCCCCCGCGCGGCCGCCGCGGCCAAGCCGGCC
CCGCGACGGCGTGCTGCGCAACCTCCGACGCTTCGCCGGCCGCGCAGGTGGATCTACG
CACGCTCGGCTACAGTCAGCAGCAGCAAGAGAAGATCAAACCGAAGGTGCGTTCGACA
GTGGCGCAGCACCACGAGGCACTGGTGGGCCATGGGTTTACACACGCGCACATCGTTGC
GCTCAGCCAACACCCGGCAGCGTTAGGGACCGTCGCTGTCACGTATCAGCACATAATCA
CGGCGTTGCCAGAGGCGACACACGAAGACATCGTTGGCGTCGGCAAACAGTGGTCCGG
CGCACGCGCCCTGGAGGCCCTTGCTCACGGATGCGGGGGAGTTGAGAGGTCCGCCGTTAC
AGTTGGACACAGGCCAACTTGTGAAGATTGCAAAACGTGGCGGCGTGACCGCAATGGA
GGCAGTGTCATGCATCGCGCAATGCACTGACGGGTGCCCCCTGGAGCTGACTCCGGACC
AAGTGGTGGCTATCGCCAGCAACATTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCG
GCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCG
CCAGCAACGGTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCT
GTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCAACATTGGCG
GCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGG
CCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCAACATTGGCGGCAAGCAAGCGCTCG
AAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAA
GTGGTGGCTATCGCCAGCAACATTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGC
TGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCC
AGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGT
GCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGG
CAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGC
CTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCG
AAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAA
GTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGC
TGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCC
AGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGT
GCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACGGTGGCGG
CAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGC
CTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCG
AAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAA
GTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGC
TGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCC
AGCAACATTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGT
GCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACATTGGCGG
CAAGCAAGCGCTCGAAAGCATTTGTGGCCAGCTGAGCCGGCCTGATCCGGCGTTGGCCG
CGTTGACCAACGACCACCTCGTCGCCTTGGCCTGCCTCGGCCCAAGAAGAAGCGCAAG
GTGTAG

图16

GinAvr15Δ128-合成蛋白:

MLIGYVRVSTNDQNTDLQORNALVCAGCEQIFEDKLSGTRTDRPGLKR
ALKRLQKGDITLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTDSI
DTSSPMGRFFFYVMGALAEMERELIERTMAGLAAARNKGRIGGRPR
KSGSGSPALRPPRAKPPAPRRRAAOPSDASPAAOVDLRTLGYSOQQQE
KIKPKVVRSTVAQHHEALVGHGFETHAHIVALSOHPAALGTVAVTYQHI
ITALPE
ATHEDIVGVGKOWSGARALEALLTDAGELRGPPLQ (-1)
LDTGQLVKIAKRGGVTAMEAVHASRNALTGAPLN (0)T
 LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG (1)A
 LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG (2)T
 LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG (3)A
 LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG (4)A
 LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG (5)A
 LTPDQVVAIASHDGGKQALETVQRLLPVLCQDHG (6)C
 LTPDQVVAIASHDGGKQALETVQRLLPVLCQDHG (7)C
 LTPDQVVAIASHDGGKQALETVQRLLPVLCQDHG (8)C
 LTPDQVVAIASHDGGKQALETVQRLLPVLCQDHG (9)C
 LTPDQVVAIASHDGGKQALETVQRLLPVLCQDHG (10)C
 LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG (11)T
 LTPDQVVAIASHDGGKQALETVQRLLPVLCQDHG (12)C
 LTPDQVVAIASHDGGKQALETVQRLLPVLCQDHG (13)C
 LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG (14)A
 LTPDQVVAIASNIGGKQALE (15)A
SIVAQLSRPDPALAALTNDHLVALACLGPKKKRKV

注:

Gin 重组酶 (斜体)

TALE N-末端框架 (双下划线)

RVD (DNA 结合结构域) (加粗)

TALE C-末端框架 (点下划线)

图17

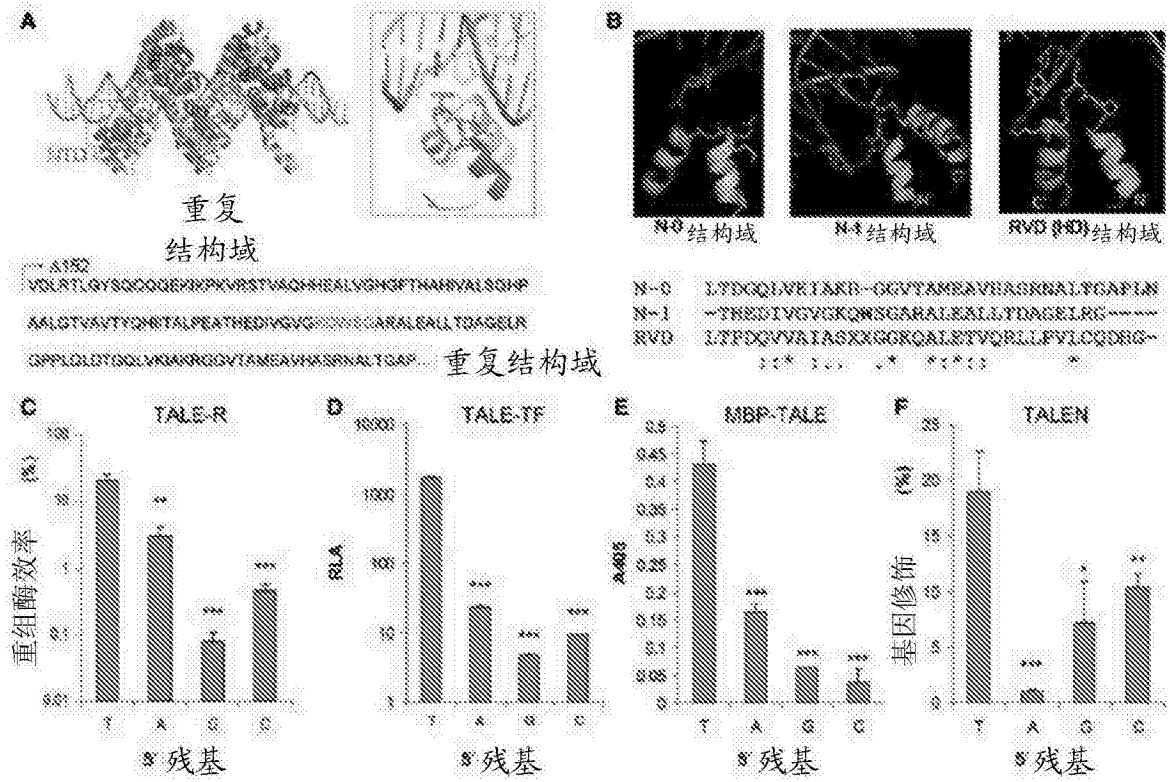


图18

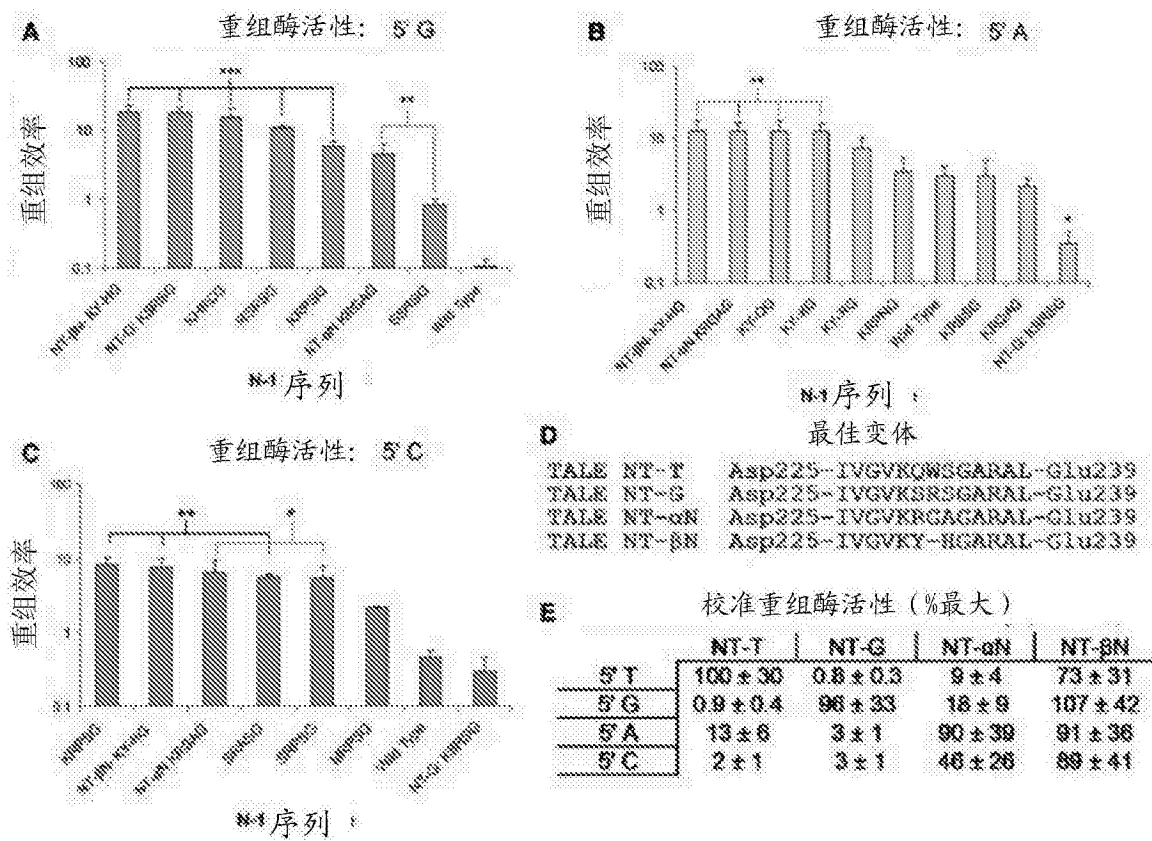


图19

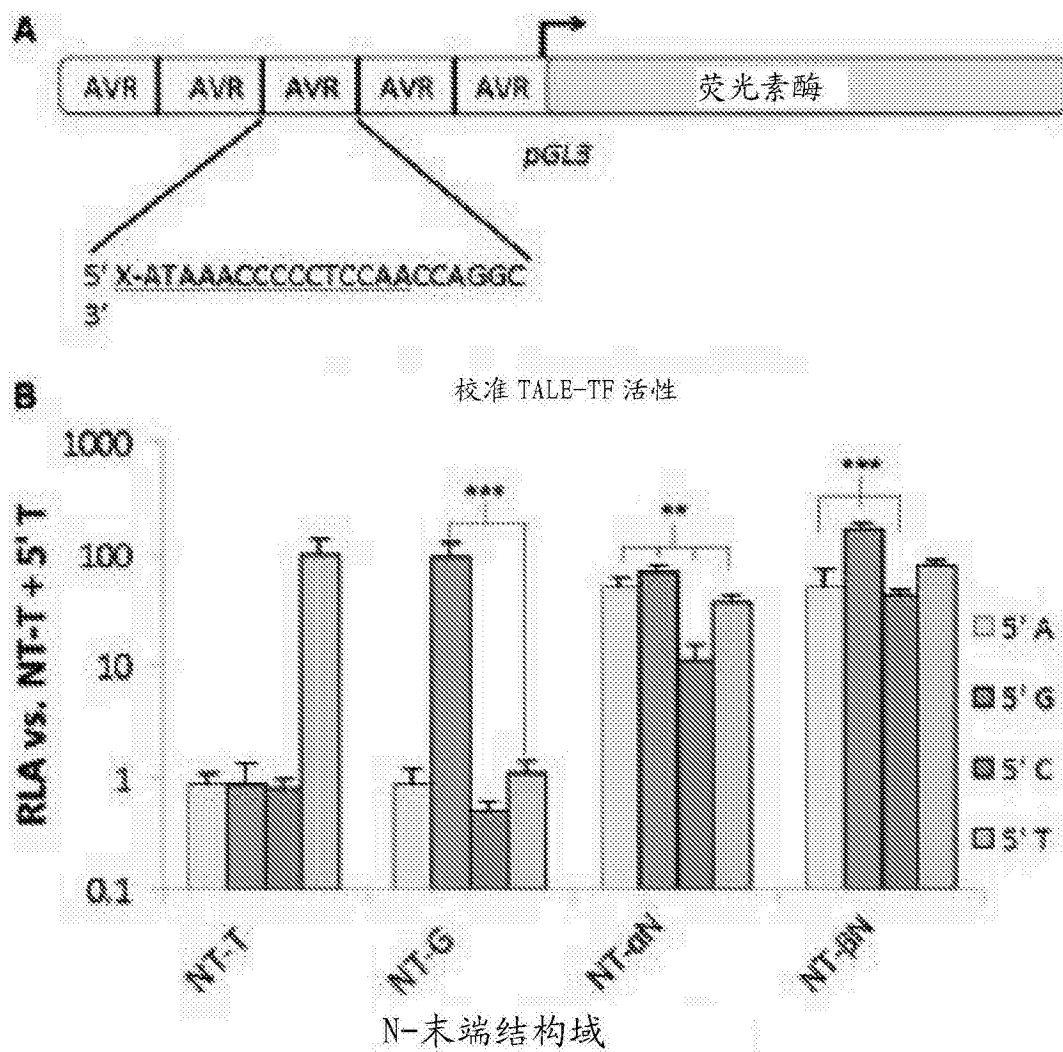


图20

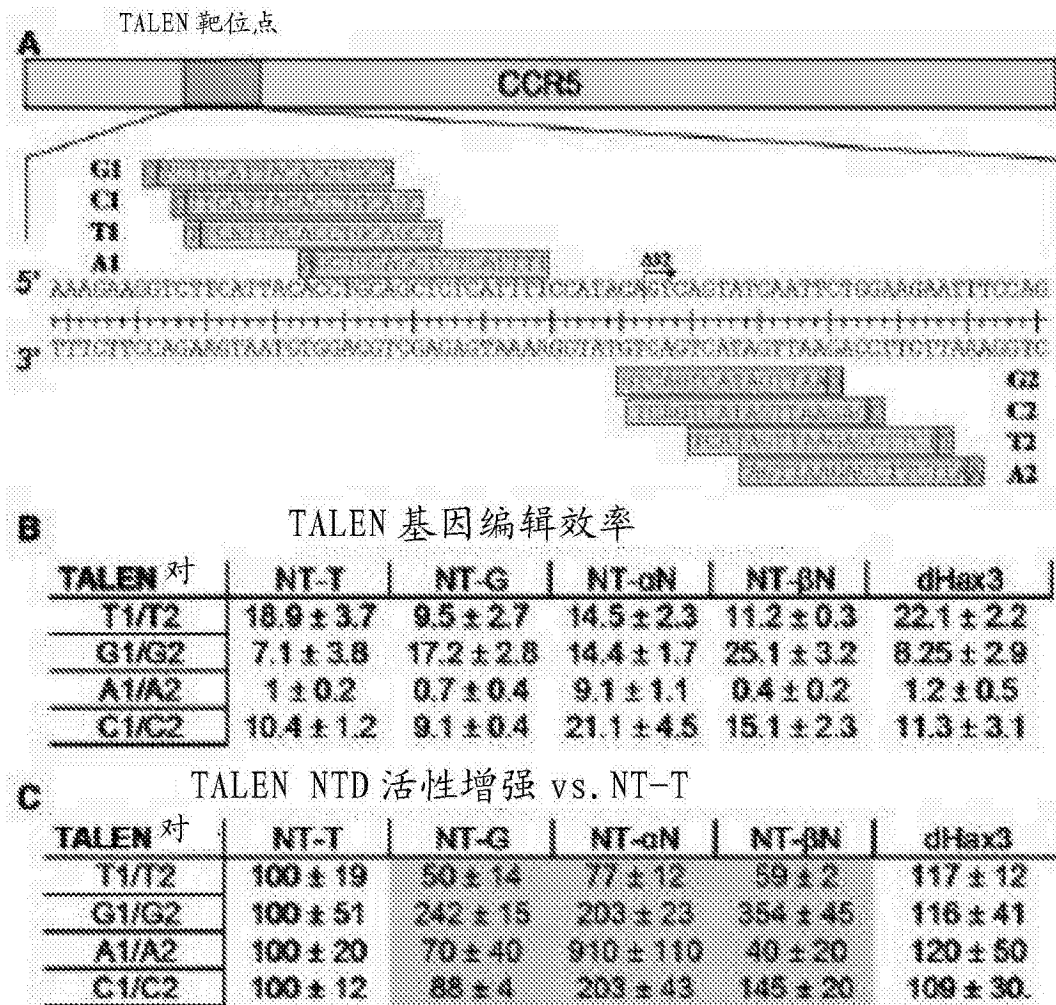


图21

N-末端比对

TALE-R/TALE-MBP/TALE-TF	FRFPFAEFAPRRKAAQFSDASFAGQVDELTLGYSDQQQENIFFEVRSTVAQHNEALVGHG
Goldy TALEN	-----VDELTLGYSDQQQENIFFEVRSTVAQHNEALVGHG
NTD-dHax3	-----VDELTLGYSDQQQENIKPKVRSTVAQHNEALVGHG

TALE-R/TALE-MBP/TALE-TF	FTTHAVIALSQHPAALSTVAVTYVTHIITALFEATHEDIVGVVKKKRAALEALLTDAGE
Goldy TALEN	FTTHAVIALSQHPAALSTVAVTYVTHIITALFEATHEDIVGVVKKKRAALEALLTDAGE
NTD-dHax3	FTTHAVIALSQHPAALSTVAVTYVTHIITALFEATHEDIVGVVKKKRAALEALLTDAGE

TALE-R/TALE-MBP/TALE-TF	LRGPFPLQLDTGQLVFIASRGGVTAHEAVHASNALLTCAP
Goldy TALEN	LRGPFPLQLDTGQLVFIASRGGVTAHEAVHASNALLTCAP
NTD-dHax3	LRGPFPLQLDTGQLVFIASRGGVTAHEAVHASNALLTCAP

RVD 结构域

C-末端比对

TALE-TF	NDHLVALACLGGRFAMDAVYKGLPHAFELIDRYNERTIGERTSHRVADYAQVVRVLEFFQC
Goldy TALEN	NDHLVALACLGGRFAMDAVYKGLPHAFELIDRYNERTIGERTSHRV-----
TALE-R/TALE-MBP	NDHLVALACL-----

TALE-TF	HSHPATAPDEAMTQFGMSGQ	Vfr4
Goldy TALEN	-----	FokI
TALE-R/TALE-MBP	-----	Stop

图22

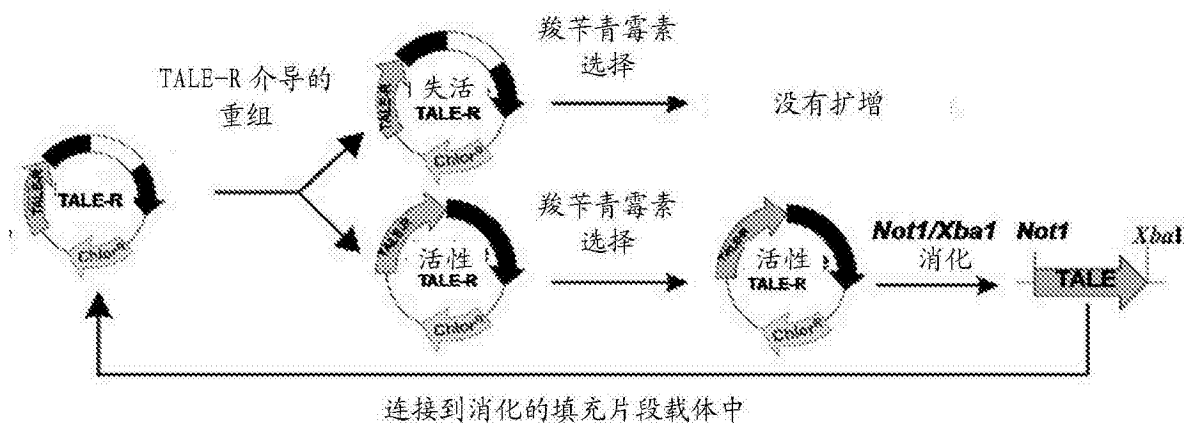


图23

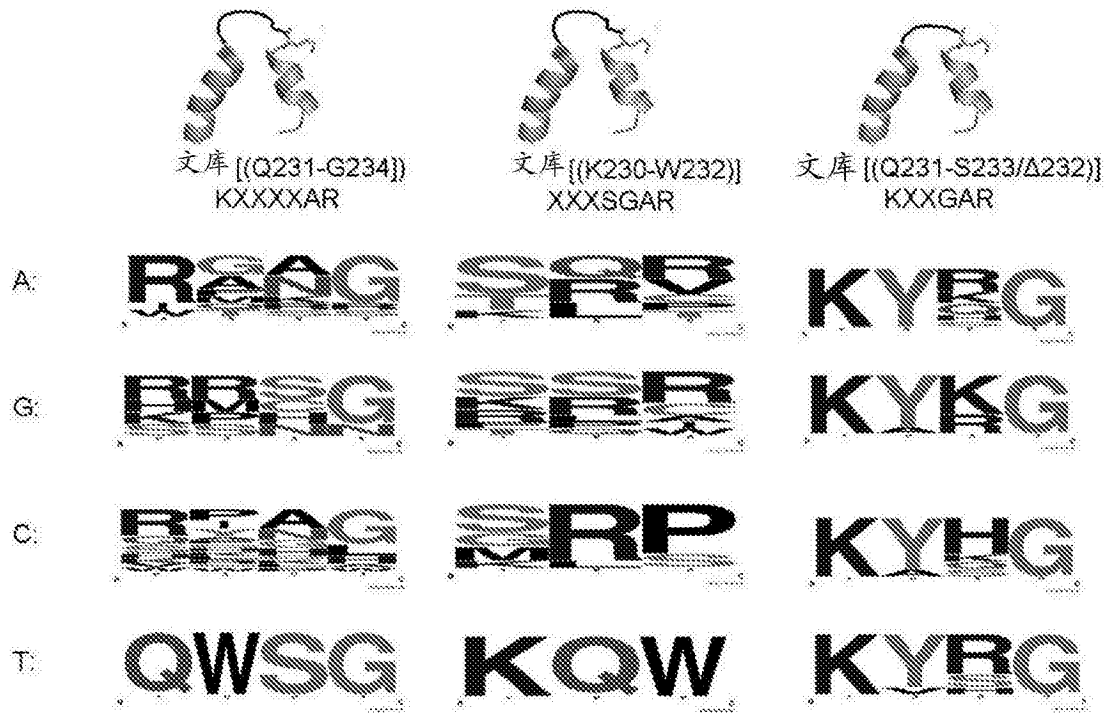


图24

```

TALE-G      QVDLRILGYSSQQQEKIKFVKVSTVAQHSEALVGHGPTTHASIVALSQHFAALGTVAVTYQ 60
Brg11      ---KRALGYSREQIRKLEQESLSGVANYHAPLTRHGFTNYDICKRISRRWQSLRMVAKNYP 57
           : *****: * .*: * : * *: * .*, *****: * :*: * ** .*
TALE-G      HIITALPEATHEDIVGVGSRSGAPALEALLTDAGELEGFFLQLDTGQLVKIAHGGVTA 120
Brg11      ELIAALPOLTRTHIVDIAHSGSGLALEALLPVATALAAAFILRLKASQIAIILQCGGERPA 117
           :*:***: * :*.:.: : * *****. * * ..*: * :*. ** : * .*
TALE-G      MEAVHASRNALTGAF 135
Brg11      ILALSRRLSRRLTGAF 132
           : *: * *. *****

```

图 25

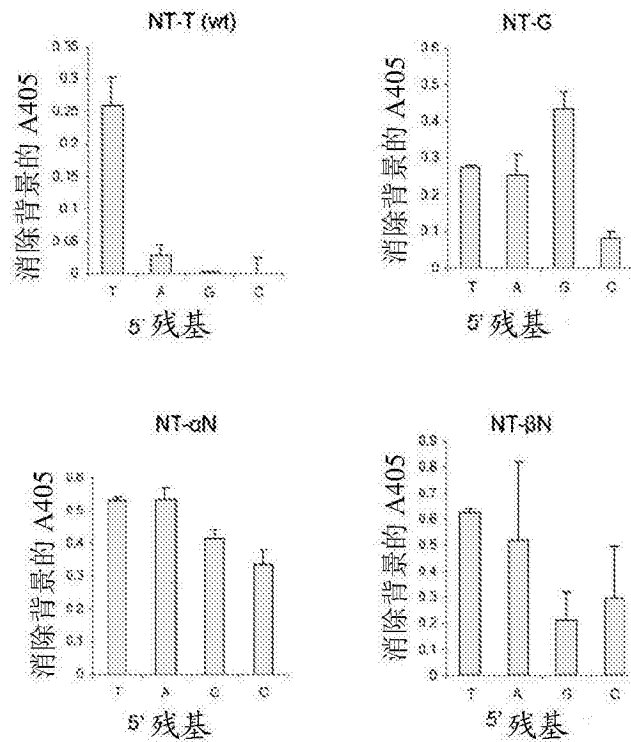
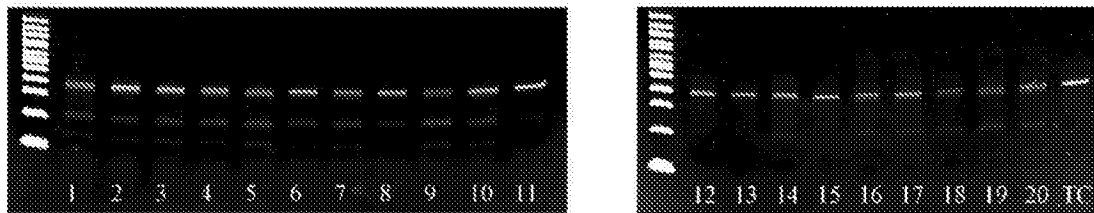


图26



TALEN对	NTD	% 编辑插入缺失	TALEN对	NTD	% 编辑插入缺失
1: T1/T2	NT-T	16.6%	12: A1/A2	NT-G	0.7%
2: T1/T2	NT-G	7.5%	13: A1/A2	NT-βN	0.4%
3: T1/T2	NT-βN	11.3%	14: A1/A2	NT-αN	9.1% (3/30)
4: T1/T2	NT-αN	12.9%	15: A1/A2	NT-dHax3	1.2%
5: T1/T2	NT-dHax3	20.1%	16: C1/C2	NT-T	8.6%
6: G1/G2	NT-T	4.4%	17: C1/C2	NT-G	6.9%
7: G1/G2	NT-G	15.2% (9/30)	18: C1/C2	NT-βN	13.4% (9/30)
8: G1/G2	NT-αN	4.9% (3x Ave = 14.4%)	19: C1/C2	NT-αN	17.9% (8/30)
9: G1/G2	NT-βN	28.9% (10/30)	20: C1/C2	NT-dHax3	9.1%
10: G1/G2	NT-dHax3	6.2%			
11: A1/A2	NT-T	1%			

图27



图28

TALE-TF 架构比较

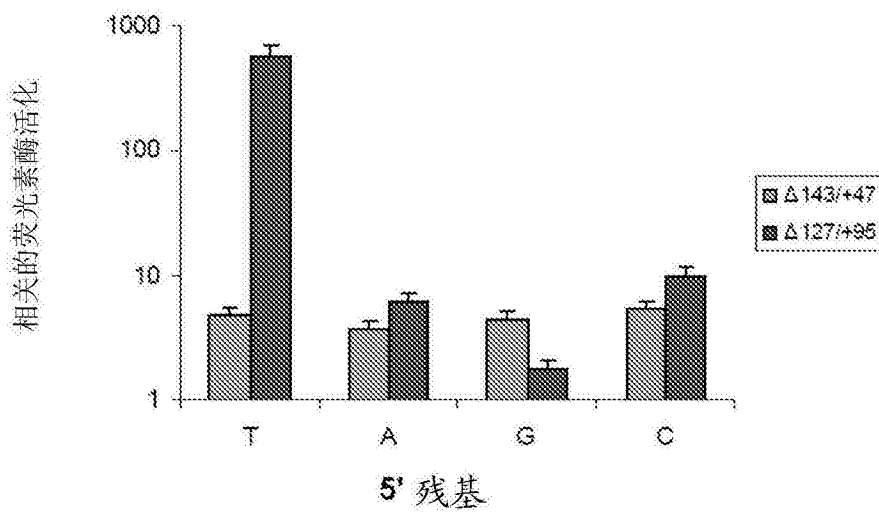


图29

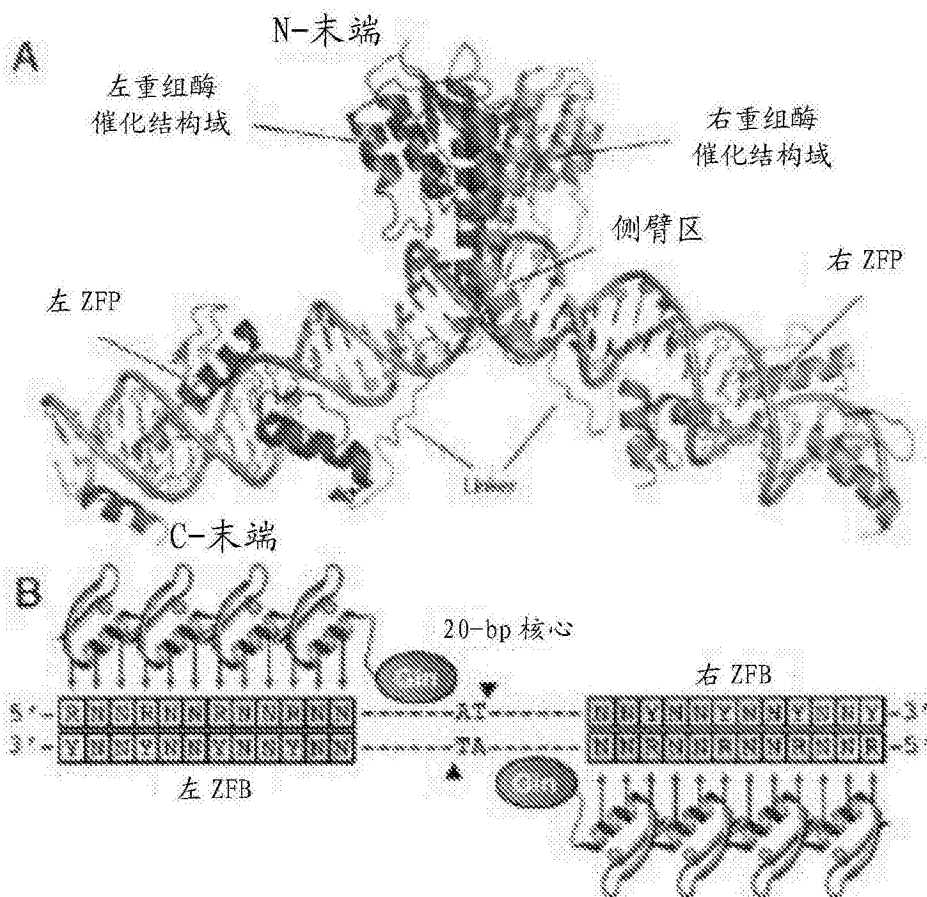


图30

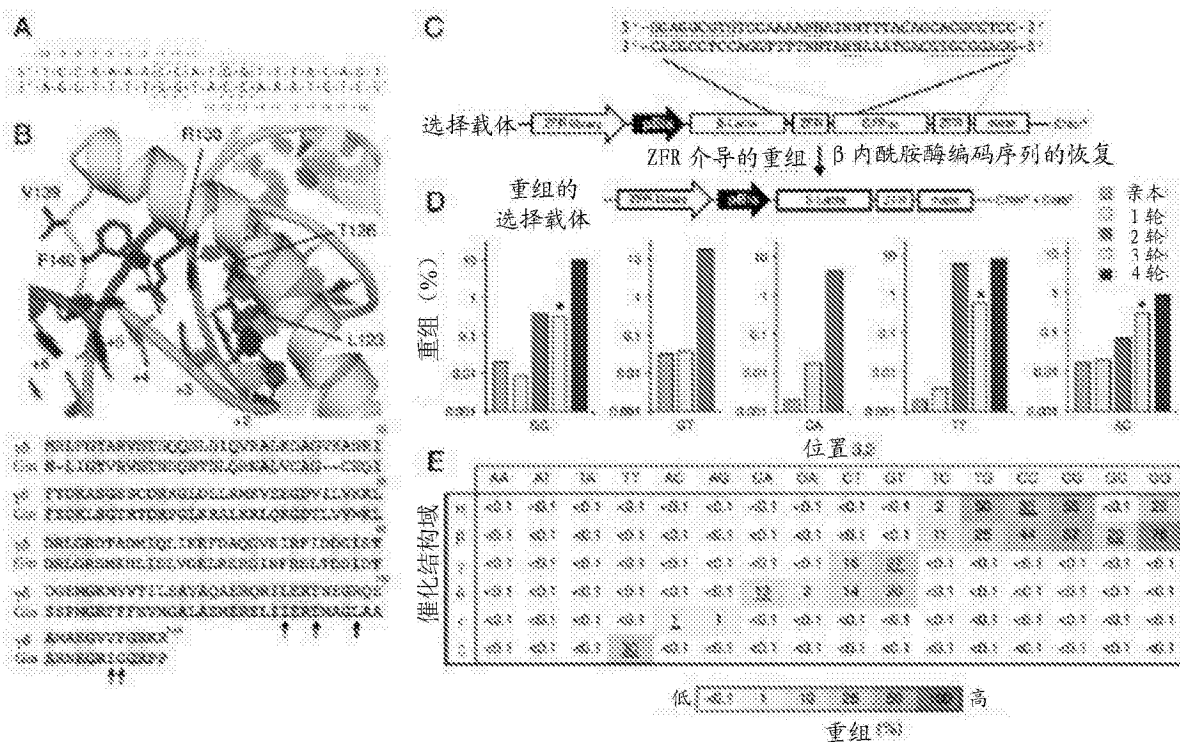


图32

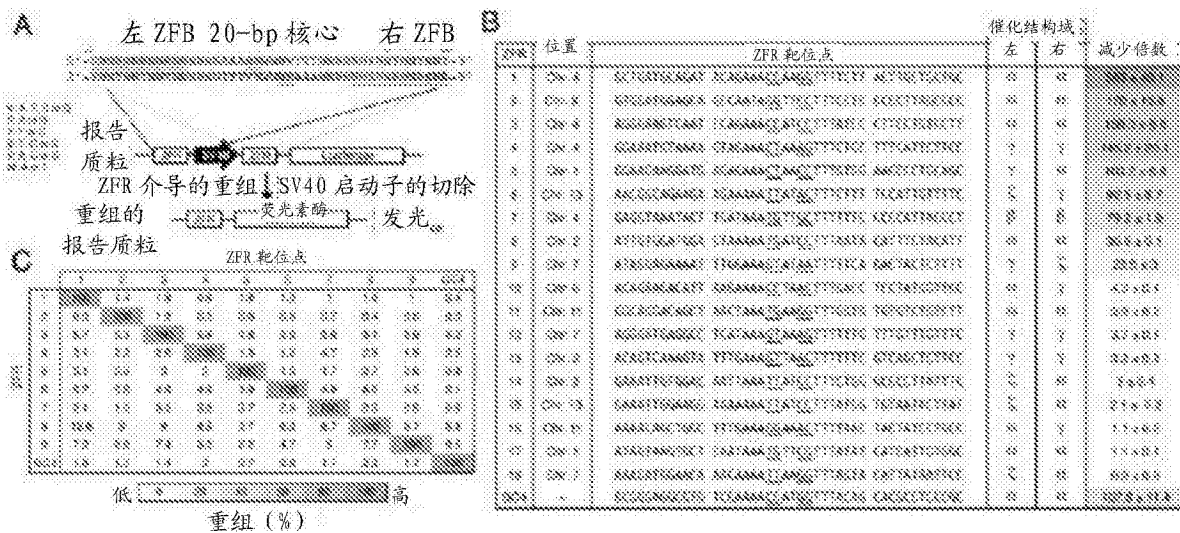


图33

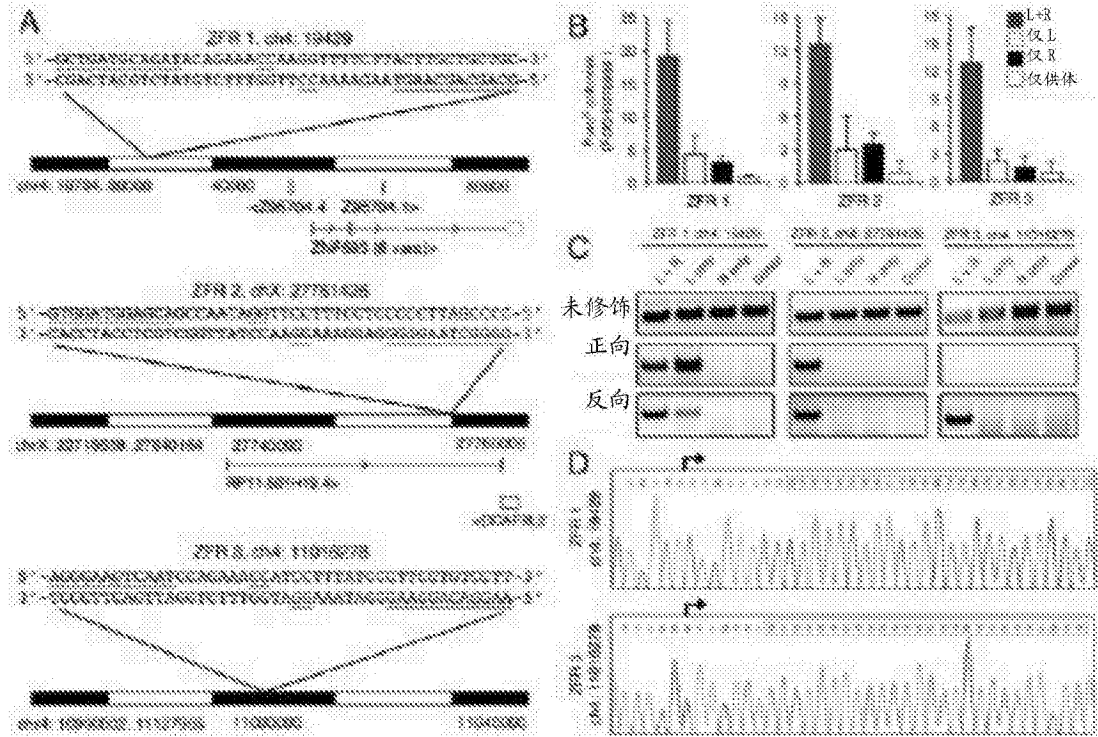


图34

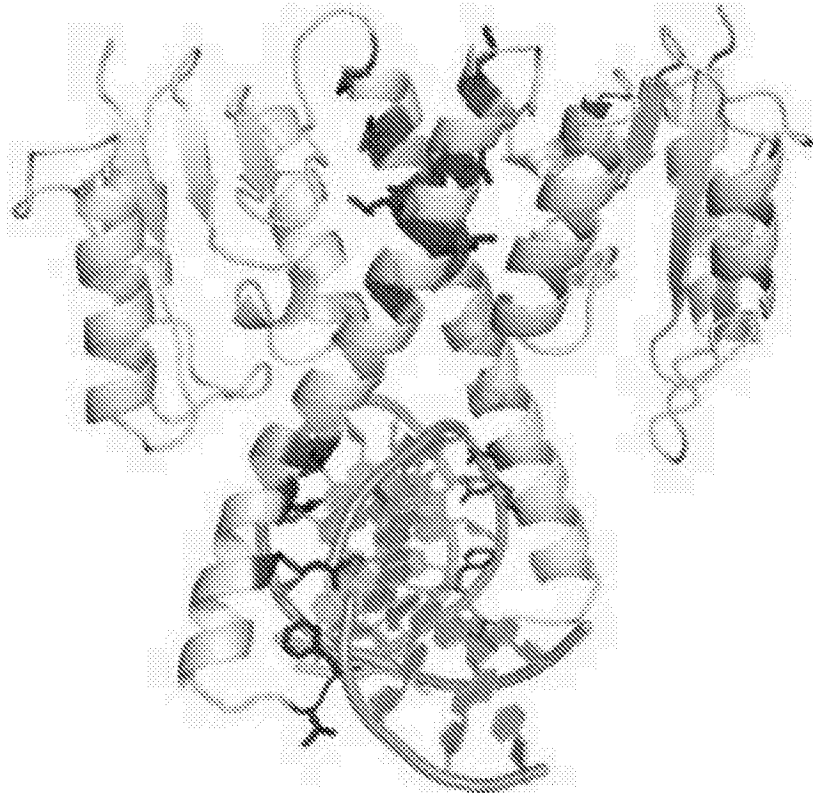


图35

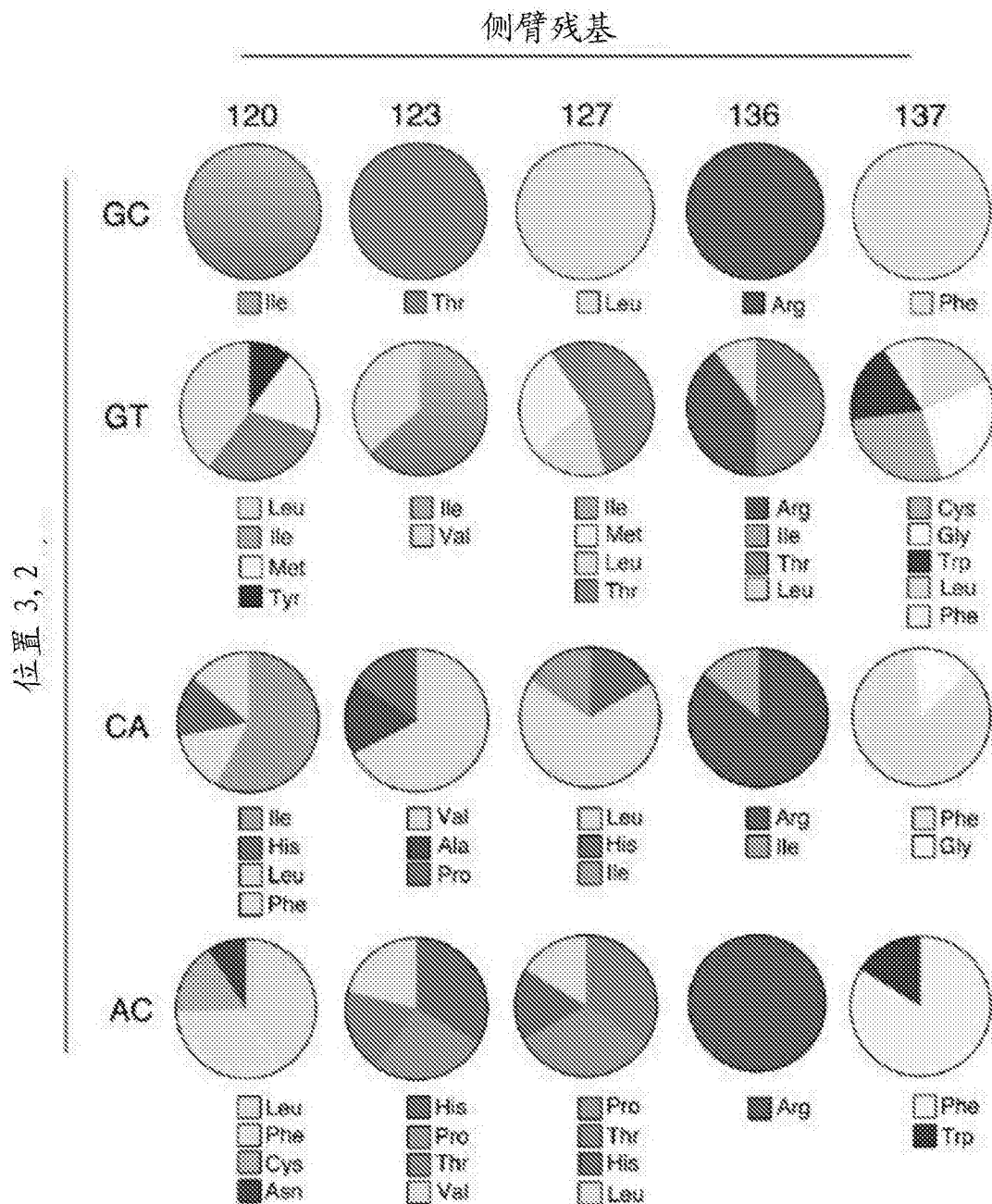


图36

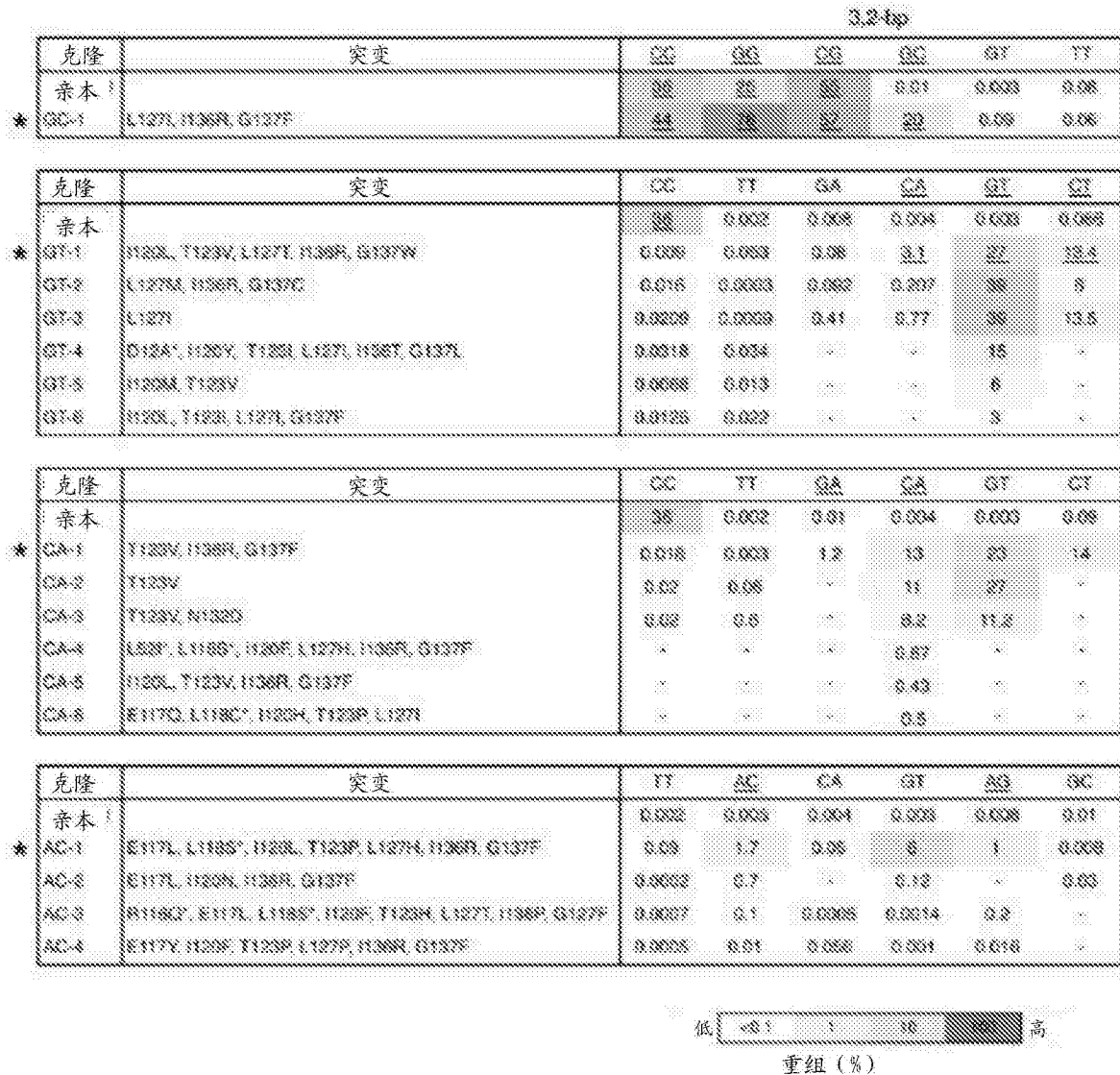


图37

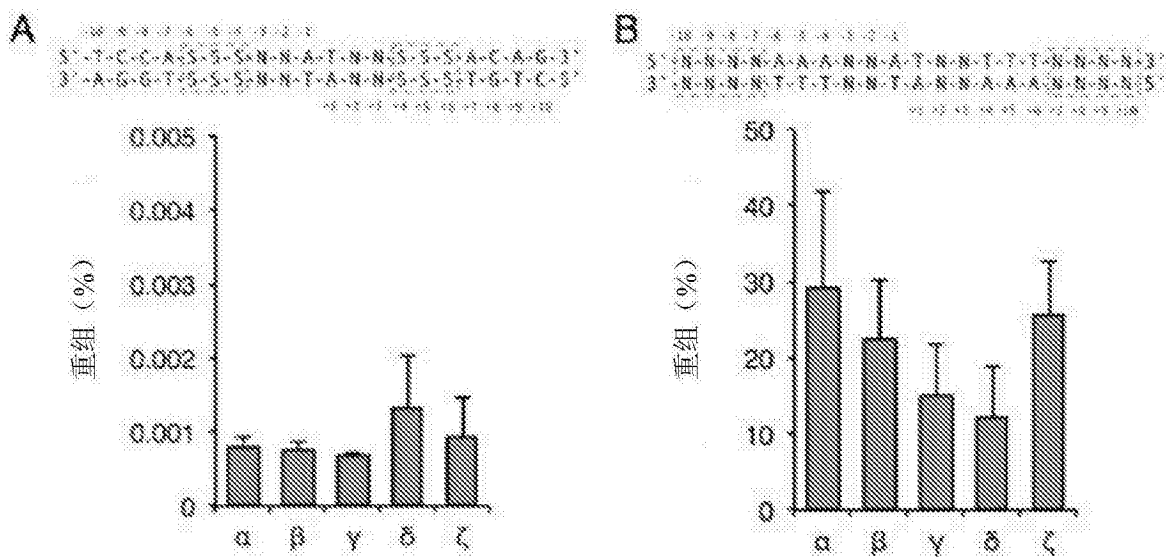


图38

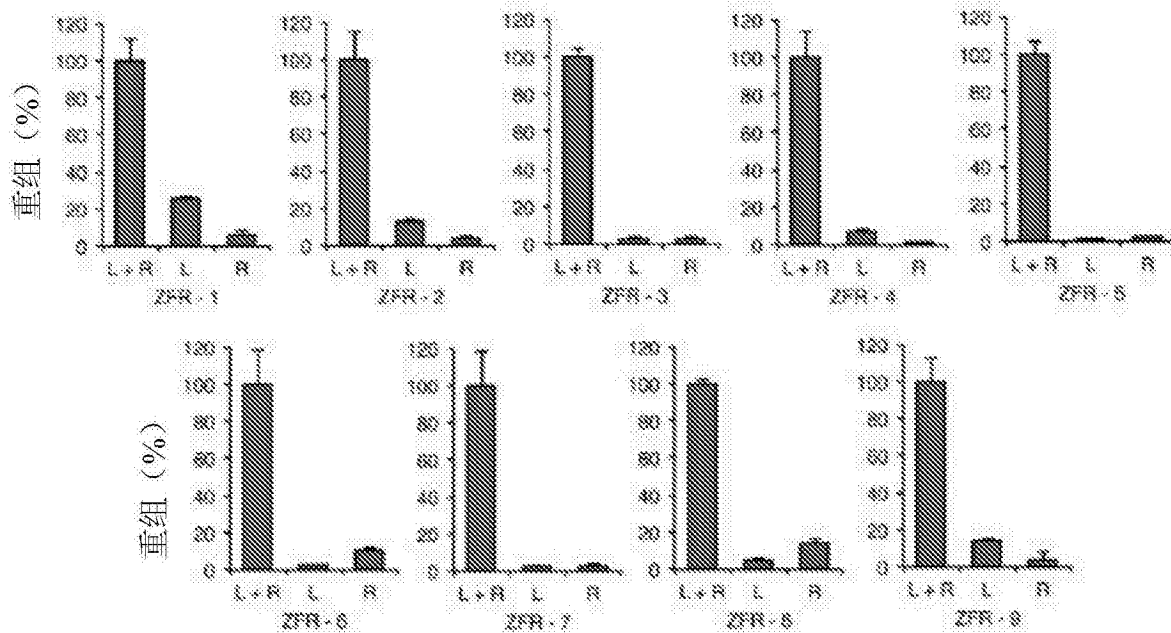


图39

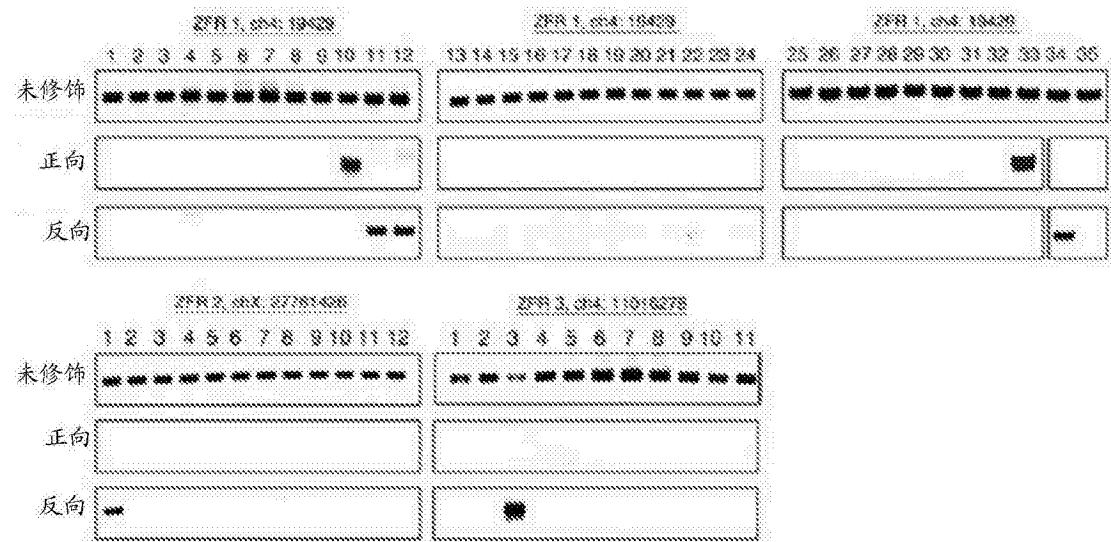


图40