(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 112175927 B
(45)授权公告日 2023.04.18

(21)申请号 201910591202.0
(22)申请日 2019.07.02
(65)同一申请的已公布的文献号
申请公布号 CN 112175927 A
(43)申请公布日 2021.01.05
(73)专利权人 上海科技大学
地址 201210 上海市浦东新区华夏中路393号
(72)发明人 刘馨怡 李广磊 黄行许
(74)专利代理机构 上海光华专利事务所(普通合伙) 31219
专利代理师 严晨 许亦琳
(51)Int.Cl.
C12N 9/78(2006.01)
C12N 9/22(2006.01)
C12N 15/62(2006.01)
C12N 15/90(2006.01)
C12N 15/85(2006.01)
(56)对比文件
CN 101595228 A,2009.12.02
CN 108822217 A,2018.11.16
CN 109804066 A,2019.05.24
WO 2018071868 A1,2018.04.19
审查员 黄炎

权利要求书1页 说明书11页
序列表26页 附图3页

(54)发明名称

一种碱基编辑工具及其用途

(57)摘要

本发明涉及生物技术领域,特别是涉及一种碱基编辑工具及其用途。本发明提供一种融合蛋白,包括ecTadA-ecTadA*二聚体片段和xCas9n片段,所述ecTadA-ecTadA*二聚体片段包括ecTad片段和ecTadA*片段。本发明通过在基于xCas9的碱基编辑器上融合核定位元件及DNA双链断裂位点结合蛋白,开发出能够高效识别PAM序列为NG、GAA及GAT的位点并进行精确的且具有更宽广的基因组识别范围的碱基编辑工具。此外,所述融合蛋白具有高效、精确的单碱基编辑能力,编辑产物纯度高,具有良好的产业化前景。

CN 112175927 B

1.一种融合蛋白,其特征在于,所述融合蛋白的氨基酸序列如SEQ ID NO. 39所示。

2.一种分离的多核苷酸,编码如权利要求1所述的融合蛋白。

3.一种构建体,所述构建体含有如权利要求2所述的分离的多核苷酸。

4.一种表达系统,所述表达系统含有如权利要求3所述的构建体或基因组中整合有外源的如权利要求2所述的多核苷酸。

5.如权利要求4所述的表达系统,其特征在于,所述表达系统的宿主细胞选自真核细胞或原核细胞。

6.如权利要求4所述的表达系统,其特征在于,所述表达系统的宿主细胞选自小鼠细胞、仓鼠细胞、非人灵长类细胞、人细胞。

7.如权利要求4所述的表达系统,其特征在于,所述表达系统的宿主细胞选自小鼠脑神经瘤细胞、中国仓鼠卵巢细胞、非洲绿猴肾细胞、人骨肉瘤细胞人胚胎肾细胞、或人宫颈癌细胞。

8.如权利要求4所述的表达系统,其特征在于,所述表达系统的宿主细胞选自N2a细胞、CHO 细胞、COS-7细胞、U2OS 细胞、HEK293FT细胞、或Hela细胞。

9.如权利要求1所述的融合蛋白、如权利要求2所述的分离的多核苷酸、如权利要求3所述的构建体或如权利要求4~8任一权利要求所述的表达系统在非疾病诊断和治疗目的的基因编辑中的用途。

10.如权利要求1所述的融合蛋白、如权利要求2所述的分离的多核苷酸、如权利要求3所述的构建体或如权利要求4~8任一权利要求所述的表达系统在真核生物非疾病诊断和治疗目的的基因编辑中的用途。

11.一种碱基编辑体系,包括如权利要求1所述的融合蛋白,所述碱基编辑体系还包括sgRNA。

12.一种非疾病诊断和治疗目的的基因编辑方法,包括:通过如权利要求1所述的融合蛋白、或如权利要求11所述的碱基编辑体系进行基因编辑。

一种碱基编辑工具及其用途

技术领域

[0001] 本发明涉及生物技术领域,特别是涉及一种碱基编辑工具及其用途。

背景技术

[0002] CRISPR/Cas技术的出现和发展引领了基因编辑领域的革命浪潮。目前,CRISPR/Cas9系统已被成功应用于多个物种的基因编辑研究,包括DNA的敲除与敲入、分子标记和基因转录调节等。在sgRNA(single guided RNA)引导下,Cas9蛋白到达指定区域并发挥核酸酶活性进行DNA双链切割,造成DNA双链断裂(double-strand breaks,DSBs)。DSBs的产生会激活机体自身的DNA修复机制,包括非同源末端连接修复(Non-Homologous End Join,NHEJ)和同源重组修复(Homologous Directly Repair,HDR)。其中NHEJ是主要的修复通路,修复结果会产生插入缺失(indels)从而造成移码突变(frameshift mutation),导致基因敲除(knockout);HDR介导的修复可以在有同源模板存在的情况下实现基因敲入(knockin),由于HDR效率低且耗时耗力,因此很难通过该方法实现精确编辑。

[0003] 针对上述问题,哈佛大学的David Liu等通过将具有切口酶活性的Cas9nickase(Cas9n)与核苷酸脱氨酶融合,开发了一种能够在不引起DSB的情况下实现单碱基替换的碱基编辑工具(Base editor)。目前有胞嘧啶碱基编辑器(cytosine base editors,CBE,)和腺嘌呤碱基编辑器(Adenine base editors,ABE),分别可以实现C-to-T和A-to-G碱基替换。CRISPR/Cas系统靶向基因组时识别前间隔序列临近基序(protospacer adjacent motif,PAM),因此PAM序列成为限制现有碱基编辑系统应用的障碍之一。有研究通过对来自酿脓链球菌(Streptococcus pyogenes)的Cas9(SpCas9)进行改造使其能够识别PAM为NG、GAA与GAT的序列,改造后的Cas9被命名为xCas9。与基于SpCas9的碱基编辑工具(BE3和ABE)相比,基于xCas9的碱基编辑器(包括xBE3和xABE)能够识别基因组上更多的靶位点,但是xBE3和xABE的编辑效率相对较低,因此通过融合核定位元件与DNA双链断裂结合蛋白,提高xBE3和xABE的编辑效率具有重要的应用意义。

发明内容

[0004] 鉴于以上所述现有技术的缺点,本发明的目的在于提供一种碱基编辑工具及其用途,用于解决现有技术中的问题。

[0005] 为实现上述目的及其他相关目的,本发明一方面提供一种融合蛋白,包括ecTadA-ecTadA*二聚体片段和xCas9n片段,所述ecTadA-ecTadA*二聚体片段包括ecTad片段和ecTadA*片段。

[0006] 在本发明一些实施方式中,所述ecTadA片段的氨基酸序列包括:

[0007] a)如SEQ ID NO.32所示的氨基酸序列;或,

[0008] b)与SEQ ID NO.32具有80%以上序列相似性的氨基酸序列、且具有a)所限定的氨基酸序列的功能,优选为能够与ecTadA*片段形成二聚体、且二聚体具有腺嘌呤脱氨酶活性。

[0009] 在本发明一些实施方式中，所述ecTadA*片段的氨基酸序列包括：

[0010] c)如SEQ ID NO.33所示的氨基酸序列；或，

[0011] d)与SEQ ID NO.33具有80％以上序列相似性的氨基酸序列、且具有c)所限定的氨基酸序列的功能，优选为能够与ecTadA片段形成二聚体、且二聚体具有腺嘌呤脱氨酶活性。

[0012] 在本发明一些实施方式中，所述xCas9n片段的氨基酸序列包括：

[0013] e)如SEQ ID NO.34所示的氨基酸序列；或，

[0014] f)与SEQ ID NO.34具有80％以上序列相似性的氨基酸序列、且具有e)所限定的氨基酸序列的功能，优选为能够识别NG、GAA或GAT作为PAM。

[0015] 在本发明一些实施方式中，所述融合蛋白自N端至C端依次包括ecTadA-ecTadA*二聚体片段和xCas9n片段。

[0016] 在本发明一些实施方式中，所述ecTadA-ecTadA*二聚体片段自N端至C端依次包括ecTad片段和ecTadA*片段。

[0017] 在本发明一些实施方式中，所述融合蛋白还包括核定位信号片段，优选的，所述核定位信号片段位于ecTadA-ecTadA*二聚体片段和xCas9n片段的N端和/或C端，优选的，所述核定位信号片段的氨基酸序列如SEQ ID NO.35、或SEQ ID NO.36所示。

[0018] 在本发明一些实施方式中，所述融合蛋白还包括柔性连接肽片段，优选的，所述柔性连接肽段的氨基酸序列如SEQ ID NO.37、或SEQ ID NO.38所示。

[0019] 在本发明一些实施方式中，所述融合蛋白的氨基酸序列如SEQ ID NO.39所示。

[0020] 本发明另一方面提供一种分离的多核苷酸，编码所述的融合蛋白。

[0021] 本发明另一方面提供一种构建体，所述构建体含有所述的分离的多核苷酸。

[0022] 本发明另一方面提供一种表达系统，所述表达系统含有所述的构建体或基因组中整合有外源的所述的多核苷酸。

[0023] 在本发明一些实施方式中，所述表达系统的宿主细胞选自真核细胞或原核细胞，优选选自小鼠细胞、仓鼠细胞、非人灵长类细胞、人细胞，更优选选自小鼠脑神经瘤细胞、中国仓鼠卵巢细胞、非洲绿猴肾细胞、人骨肉瘤细胞人胚胎肾细胞、或人宫颈癌细胞，更优选选自N2a细胞、CHO细胞、COS-7细胞、U2OS细胞、HEK293FT细胞、或Hela细胞。

[0024] 本发明另一方面提供所述的融合蛋白、所述的分离的多核苷酸、所述的构建体或所述的表达系统在基因编辑中的用途。

[0025] 在本发明一些实施方式中，所述用途具体为在真核生物的基因编辑中的用途。

[0026] 本发明另一方面提供一种碱基编辑体系，包括所述的融合蛋白，所述碱基编辑体系还包括sgRNA。

[0027] 本发明另一方面提供一种基因编辑方法，包括：通过所述的融合蛋白、或所述的碱基编辑体系进行基因编辑。

附图说明

[0028] 图1显示为本发明BPNLS-xABE载体示意图及其在部分测试位点的编辑效率检测。其中，a为优化后的BPNLS-xABE载体示意图；b为优化后的BPNLS-xABE与优化前xABE在内源基因位点进行编辑的Sanger测序峰图；c为应用Sanger测序分析BPNLS-xABE与xABE对内源基因位点编辑的效率；d为对BPNLS-xABE与xABE编辑内源基因位点效率的标准化。

[0029] 图2显示为本发明实施例1中BPNLS-xABE模拟致病位点的编辑效率示意图。其中，a为BPNLS-xABE进行致病突变模拟的部分代表性位点Sanger测序峰图；b为应用Sanger测序分析并量化BPNLS-xABE与xABE模拟致病突变的效率；c为BPNLS-xABE同时模拟两个致病位点突变的效率。

[0030] 图3显示为本发明实施例2中含有威尔逊病致病突变细胞株示意图。其中，a为威尔逊病致病基因突变位点Atp7b^{T1033A}在该基因中的位置与序列信息，以及用于模拟和修复该致病突变的sgRNA序列；b为威尔逊病致病基因突变位点Atp7b^{T1220M}在该基因中的位置与序列信息，以及用于模拟和修复该致病突变的sgRNA序列；c为应用BPNLS-xABE成功构建含有威尔逊病致病突变位点Atp7b^{T1033A}细胞株的Sanger测序峰图；d为应用BPNLS-Gam-xBE3成功构建含有威尔逊病致病突变位点Atp7b^{T1220M}细胞株的Sanger测序峰图。

[0031] 图4显示为本发明实施例3利用BPNLS-Gam-xBE3和BPNLS-xABE分别对Atp7b^{T1033A}和Atp7b^{T1220M}进行修复的示意图。其中，a为应用BPNLS-Gam-xBE3对含有威尔逊病致病突变位点Atp7b^{T1033A}细胞株中突变位点进行修复的Sanger测序峰图；b为威尔逊病致病突变位点Atp7b^{T1033A}在突变细胞株及修复细胞中的量化；c为应用BPNLS-xABE对含有威尔逊病致病突变位点Atp7b^{T1220M}细胞株中突变位点进行修复的Sanger测序峰图；d为威尔逊病致病突变位点Atp7b^{T1220M}在突变细胞株及修复细胞中的量化。

具体实施方式

[0032] 本发明发明人经过大量探索性研究，提供了一种融合蛋白，所述融合蛋白是一种新的腺嘌呤碱基编辑工具，所述融合蛋白可以识别NG、GAA和GAT位点并进行精确的碱基编辑，拓宽了碱基编辑的靶向范围，在此基础上完成了本发明。

[0033] 本发明第一方面提供一种融合蛋白，包括ecTadA-ecTadA*二聚体片段和xCas9n片段，所述ecTadA-ecTadA*二聚体片段包括ecTad片段和ecTadA*片段。所述融合蛋白可以以NG、GAA和GAT为PAM序列，与靶向靶点区域的sgRNA相配合，实现对靶点区域内sgRNA 5’端4-7位的A-to-G的高效碱基编辑，在目标位点具有更高的编辑效率、更高的编辑准确性。

[0034] 本发明所提供的融合蛋白中，所述ecTadA片段的氨基酸序列可以包括：a)如SEQ ID NO.32所示的氨基酸序列；或，b)与SEQ ID NO.32具有80％以上序列相似性的氨基酸序列、且具有a)所限定的氨基酸序列的功能。具体的，所述b)中的氨基酸序列具体指：如SEQ ID NO.32所示的氨基酸序列经过取代、缺失或者添加一个或多个(具体可以是1-50、1-30个、1-20个、1-10个、1-5个、1-3个、1个、2个、或3个)氨基酸而得到的，或者在N-末端和/或C-末端添加一个或多个(具体可以是1-50个、1-30个、1-20个、1-10个、1-5个、1-3个、1个、2个、或3个)氨基酸而得到的，且具有氨基酸如SEQ ID NO.32所示的多肽片段的功能的多肽片段，例如，可以是具有能够与ecTadA*片段形成二聚体、且二聚体具有腺嘌呤脱氨酶活性，更具体可以是将腺嘌呤(adenine，A)脱氨产生次黄嘌呤(hypoxanthine，I)的功能。所述b)中的氨基酸序列可与SEQ ID NO.32具有80％、85％、90％、93％、95％、97％、或99％以上的相似性。所述ecTadA片段源自大肠杆菌(Escherichia coli)。

[0035] 本发明所提供的融合蛋白中，所述ecTadA*片段的氨基酸序列可以包括：c)如SEQ ID NO.33所示的氨基酸序列；或，d)与SEQ ID NO.33具有80％以上序列相似性的氨基酸序列、且具有c)所限定的氨基酸序列的功能。具体的，所述d)中的氨基酸序列具体指：如SEQ

ID NO.33所示的氨基酸序列经过取代、缺失或者添加一个或多个(具体可以是1-50、1-30个、1-20个、1-10个、1-5个、1-3个、1个、2个、或3个)氨基酸而得到的,或者在N-末端和/或C-末端添加一个或多个(具体可以是1-50个、1-30个、1-20个、1-10个、1-5个、1-3个、1个、2个、或3个)氨基酸而得到的,且具有氨基酸如SEQ ID NO.33所示的多肽片段的功能的多肽片段,例如,可以是具有能够与ecTadA片段形成二聚体、且二聚体具有腺嘌呤脱氨酶活性,更具体可以是将腺嘌呤(adenine,A)脱氨产生次黄嘌呤(hypoxanthine,I)的功能。所述d)中的氨基酸序列可与SEQ ID NO.33具有80%、85%、90%、93%、95%、97%、或99%以上的相似性。所述ecTadA*片段源自大肠杆菌(Escherichia coli),由人工定向进化获得。

[0036] 本发明所提供的融合蛋白中,所述xCas9n片段的氨基酸序列可以包括:e)如SEQ ID NO.34所示的氨基酸序列;或,f)与SEQ ID NO.34具有80%以上序列相似性的氨基酸序列、且具有e)所限定的氨基酸序列的功能。具体的,所述f)中的氨基酸序列具体指:如SEQ ID NO.34所示的氨基酸序列经过取代、缺失或者添加一个或多个(具体可以是1-50、1-30个、1-20个、1-10个、1-5个、1-3个、1个、2个、或3个)氨基酸而得到的,或者在N-末端和/或C-末端添加一个或多个(具体可以是1-50个、1-30个、1-20个、1-10个、1-5个、1-3个、1个、2个、或3个)氨基酸而得到的,且具有氨基酸如SEQ ID NO.34所示的多肽片段的功能的多肽片段,例如,可以是能够识别NG、GAA或GAT作为PAM的功能,具体可以是能够将NG、GAA或GAT序列作为PAM,并可以与特异性靶向位点的sgRNA以及ecTadA-ecTadA*二聚体片段相配合,实现靶点区域内sgRNA 5'端4-7位的A-to-G的碱基编辑。所述f)中的氨基酸序列可与SEQ ID NO.34具有80%、85%、90%、93%、95%、97%、或99%以上的相似性。CRISPR/Cas9系统的靶向识别通常需要靶位点旁边具有前间隔序列邻近基序(protospacer adjacent motif,PAM),作为一种最为频繁用于基因组编辑的Cas9酶,来自酿脓链球菌(Streptococcus pyogenes)的Cas9(SpCas9)仅能够识别NGG序列的PAM,这就限制了基因组中能够被靶向的范围,而本发明中的xCas9n片段来源于酿脓链球菌(Streptococcus pyogenes),能够识别NG、GAA或GAT序列作为PAM。

[0037] 本发明所提供的融合蛋白中,所述的取代、缺失或者添加可以是保守氨基酸取代。所述"保守氨基酸取代"具体可以是指氨基酸残基被其他具有相似侧链的氨基酸残基取代的情况。具有相似侧链的氨基酸残基家族对于本领域技术人员来说应该是已知的,例如,可以是包括但不限于碱性侧链(例如赖氨酸,精氨酸,组氨酸),酸性侧链(例如天冬氨酸,谷氨酸),不带电荷的极性侧链(例如,甘氨酸,天冬酰胺,谷氨酰胺,丝氨酸,苏氨酸,酪氨酸,半胱氨酸),非极性侧链(例如丙氨酸,缬氨酸,亮氨酸,异亮氨酸,脯氨酸,苯丙氨酸,甲硫氨酸,色氨酸)异亮氨酸)和芳族侧链(例如酪氨酸,苯丙氨酸,色氨酸,组氨酸)等家族。保守型氨基酸取代更具体可以包括但不限于下表中所列的具体情况,表1(氨基酸相似度矩阵)中的数字表示两个氨基酸之间的相似度,当数字大于等于0时认为是保守氨基酸取代,表2为示例性的保守氨基酸取代的方案。

[0038] 表1

[0039]

	C	G	P	S	A	T	D	E	N	Q	H	K	R	V	M	I	L	F	Y	W
W	-8	-7	-6	-2	-6	-5	-7	-7	-4	-5	-3	-3	2	-6	-4	-5	-2	0	0	17
Y	0	-5	-5	-3	-3	-3	-4	-4	-2	-4	0	-4	-5	-2	-2	-1	-1	7	10	
F	-4	-5	-5	-3	-4	-3	-6	-5	-4	-5	-2	-5	-4	-1	0	1	2	9		

L	-6	-4	-3	-3	-2	-2	-4	-3	-3	-2	-2	-3	-3	2	4	2	6		
I	-2	-3	-2	-1	-1	0	-2	-2	-2	-2	-2	-2	-2	4	2	5			
M	-5	-3	-2	-2	-1	-1	-3	-2	0	-1	-2	0	0	2	6				
V	-2	-1	-1	-1	0	0	-2	-2	-2	-2	-2	-2	-2	4					
R	-4	-3	0	0	-2	-1	-1	-1	0	1	2	3	6						
K	-5	-2	-1	0	-1	0	0	0	1	1	0	5							
H	-3	-2	0	-1	-1	-1	1	1	2	3	6								
Q	-5	-1	0	-1	0	-1	2	2	1	4									
N	-4	0	-1	1	0	0	2	1	2										
E	-5	0	-1	0	0	0	3	4											
D	-5	1	-1	0	0	0	4												
T	-2	0	0	1	1	3													
A	-2	1	1	1	2														
S	0	1	1	1															
P	-3	-1	6																
G	-3	5																	
C	12																		

[0040] 表2

氨基酸残基	**保守替代品**
Alanine	D-Ala, Gly, Aib, β-Ala, L-Cys, D-Cys
Arginine	D-Arg, Lys, D-Lys, Orn D-Orn
Asparagine	D-Asn, Asp, D-Asp, Glu, D-Glu Gln, D-Gln
Aspartic Acid	D-Asp, D-Asn, Asn, Glu, D-Glu, Gln, D-Gln
Cysteine	D-Cys, S-Me-Cys, Met, D-Met, Thr, D-Thr, L-Ser, D-Ser
Glutamine	D-Gln, Asn, D-Asn, Glu, D-Glu, Asp, D-Asp
Glutamic Acid	D-Glu, D-Asp, Asp, Asn, D-Asn, Gln, D-Gln
Glycine	Ala, D-Ala, Pro, D-Pro, Aib, β-Ala
Isoleucine	D-Ile, Val, D-Val, Leu, D-Leu, Met, D-Met
Leucine	Val, D-Val, Met, D-Met, D-Ile, D-Leu, Ile
Lysine	D-Lys, Arg, D-Arg, Orn, D-Orn
Methionine	D-Met, S-Me-Cys, Ile, D-Ile, Leu, D-Leu, Val, D-Val
Phenylalanine	D-Phe, Tyr, D-Tyr, His, D-His, Trp, D-Trp
Proline	D-Pro
Serine	D-Ser, Thr, D-Thr, allo-Thr, L-Cys, D-Cys
Threonine	D-Thr, Ser, D-Ser, allo-Thr, Met, D-Met, Val, D-Val
Tyrosine	D-Tyr, Phe, D-Phe, His, D-His, Trp, D-Trp
Valine	D-Val, Leu, D-Leu, Ile, D-Ile, Met, D-Met

[0041]

[0042]

[0043] 本发明所提供的融合蛋白中，还可以包括核定位信号片段(NLS)，所述核定位信号

片段通常可以与入核载体相互作用,从而可以使蛋白能够被运送进细胞核。所述核定位信号片段可以位于ecTadA-ecTadA*二聚体片段的N端,也可以位于xCas9n片段的C端,也可以位于ecTadA-ecTadA*二聚体片段与xCas9n片段之间。本领域技术人员通常可以选择合适的核定位信号片段的位置和具体序列,例如,ecTadA-ecTadA*二聚体片段的N端可以设有第一核定位信号片段,所述第一核定位信号片段的氨基酸序列可以包括如SEQ ID NO.35所示的氨基酸序列;再例如,xCas9n片段的C端可以设有第二核定位信号片段,所述第二核定位信号片段的氨基酸序列可以包括如SEQ ID NO.36所示的氨基酸序列。

[0044] 本发明所提供的融合蛋白中,还可以包括柔性连接肽段,所述柔性连接肽段可以位于ecTadA-ecTadA*二聚体片段的N端、ecTad片段和ecTadA*片段之间、ecTadA-ecTadA*二聚体片段和xCas9n片段之间、或xCas9n片段的C端。本领域技术人员通常可以选择合适的柔性连接肽段以连接各多肽片段,例如,ecTad片段和ecTadA*片段之间可以设有第一柔性连接肽段,所述第一柔性连接肽段的氨基酸序列可以包括如SEQ ID NO.37所示的氨基酸序列;再例如,ecTadA-ecTadA*二聚体片段和xCas9n片段之间可以设有第二柔性连接肽段,所述第二柔性连接肽段的氨基酸序列可以包括如SEQ ID NO.37所示的氨基酸序列;再例如,xCas9n片段的C端可以设有第三柔性连接肽段,所述第三柔性连接肽段的氨基酸序列可以包括如SEQ ID NO.38所示的氨基酸序列。

[0045] 本发明所提供的融合蛋白中,所述融合蛋白自N端至C端可以依次包括ecTadA-ecTadA*二聚体片段和xCas9n片段,所述ecTadA-ecTadA*二聚体片段自N端至C端可以依次包括ecTadA片段和ecTadA*片段,优选可以依次包括ecTadA片段、第一柔性连接肽段和ecTadA*片段。在本发明一具体实施例中,所述融合蛋白自N端至C端可以依次包括第一核定位信号片段、ecTadA-ecTadA*二聚体片段、第二柔性连接肽段、xCas9n片段,第三柔性连接肽段、第二核定位信号片段,所述融合蛋白的氨基酸序列如SEQ ID NO.39所示。

[0046] 本发明第二方面提供一种分离的多核苷酸,编码本发明第一方面所提供的融合蛋白。

[0047] 本发明第三方面提供一种构建体,所述构建体含有本发明第二方面所提供的分离的多核苷酸。所述构建体通常可以通过将所述分离的多核苷酸插入合适的表达载体中构建获得,本领域技术人员可选择合适的表达载体,例如,所述表达载体可以是包括但不限于pCMV表达载体、pSV2表达载体、pGL3表达载体、pST1347载体、px330载体、pAAV载体等。

[0048] 本发明第四方面提供一种表达系统,所述表达系统含有本发明第三方面所提供的构建体或基因组中整合有外源的本发明第二方面所提供的分离的多核苷酸。所述表达系统可以是宿主细胞,所述宿主细胞可以表达如上所述的融合蛋白,所述融合蛋白可以与sgRNA相配合,从而可以将所述融合蛋白定位到目标区域,实现目标区域的碱基编辑。在本发明另一具体实施例中,所述宿主细胞可以是真核细胞和/或原核细胞,更具体可以是小鼠细胞、仓鼠细胞、非人灵长类细胞、人细胞等,更具体可以是小鼠脑神经瘤细胞、中国仓鼠卵巢细胞、非洲绿猴肾细胞、人骨肉瘤细胞、人胚胎肾细胞、人宫颈癌细胞等,更具体可以是N2a细胞、CHO细胞、COS-7细胞、U2OS细胞、HEK293FT细胞、Hela细胞等。

[0049] 本发明第五方面提供本发明第一方面所提供的融合蛋白、或本发明第二方面所提供的分离的多核苷酸、或本发明第三方面所提供的构建体、或本发明第四方面所提供的表达系统在基因编辑中的用途,优选为真核生物的基因编辑中的用途,所述真核生物具体可

以是后生动物，具体可以是包括但不限于人、非人灵长类、仓鼠、小鼠等。所述用途具体可以是包括但不限于由A到G的碱基编辑(更具体为靶点区域内sgRNA 5’端4-7位的A-to-G的碱基编辑)、编辑剪接受体/供体位点来调节RNA剪接、利用本工具进行模型(例如，疾病模型、细胞模型、动物模型等)的构建或人类疾病的治疗等。在本发明一具体实施例中，所述用途具体可以为靶标基因的由A到G的碱基编辑，具体的编辑位点可以是$ATP7b^{M1169V}$(GenBank:NC_000013.11)、$ATP7b^{I1230V}$、$ATP7b^{W1353R}$、$ATP7b^{Y713C}$、$STK11^{L67P}$(GenBank:NC_000019.10)、$ARHGEF18^{T270A}$(GenBank:NC_000019.10)、$BCS1L^{V327A}$(GenBank:NC_000002.12)、$BCS1L^{M48V}$、$AIFM1^{D237G}$(GenBank:NC_000023.11)、$ALPL^{C201R}$(GenBank:NC_000001.11)等。在本发明一具体实施例中，所述用途具体为细胞株的构建，更具体为含有威尔逊病致病突变细胞株的构建或修复、或PJ’s综合征突变细胞株的构建或修复等，具体的编辑位点可以是$Atp7b^{T1033A}$、$Atp7b^{T1220M}$、$STK11^{D194N}$、$STK11^{R297K}$等。本发明另一具体实施例中，被编辑的对象可以是胚胎、细胞等。

[0050] 本发明第六方面提供一种碱基编辑体系，包括本发明第一方面所提供的融合蛋白，所述碱基编辑体系还包括sgRNA。本领域技术人员可以根据基因的目标编辑区域，选择合适的靶向特异性位点的sgRNA。例如，所述sgRNA的序列通常可以与目标区域至少部分互补，从而可以与所述融合蛋白相配合，将所述融合蛋白定位到目标区域，实现靶点区域内sgRNA 5’端4-7位的A-to-G的碱基编辑，具体可以是腺嘌呤脱氨基反应，即将腺嘌呤(A)编辑为次黄嘌呤(I)。本发明所提供的碱基编辑体系极大地拓宽了基因组可靶向的范围，可以将NG、GAA或GAT序列作为PAM，实现sgRNA靶点区域内5’端4-7位的A-to-G的碱基，且突变具有很高的精准性，邻近脱靶低。在本发明一具体实施例中，所述用途具体为细胞株的构建，更具体为含有威尔逊病致病突变细胞株的构建或修复、或PJ’s综合征突变细胞株的构建或修复等。在本发明一具体实施例中，所述用途具体为细胞株的构建，更具体为含有威尔逊病致病突变细胞株的构建或修复、或PJ’s综合征突变细胞株的构建或修复等，具体的编辑位点可以是$Atp7b^{T1033A}$(GenBank:NC_000013.11)、$Atp7b^{T1220M}$(Genbank:NC_000013.11)、$STK11^{D194N}$(GenBank:NC_000019.10)、STK^{R297K}(GenBank:NC_000019.10)等。本发明另一具体实施例中，被编辑的对象可以是胚胎、细胞等。

[0051] 本发明第七方面提供一种碱基编辑方法，包括：通过本发明第一方面所提供的融合蛋白、或本发明第六方面所提供的碱基编辑体系进行基因编辑。例如，所述基因编辑方法可以包括：在适当条件下培养本发明第四方面所提供的表达系统，从而表达所述融合蛋白，所述融合蛋白可以在与其配合的靶向目标区域的sgRNA存在的条件下，对靶标区域进行碱基编辑。提供所述sgRNA存在的条件的方法对于本领域技术人员来说应该是已知的，例如，可以是在适当条件下培养能够表达所述sgRNA的表达系统，所述表达系统可以是包括含有编码所述sgRNA的多核苷酸的表达载体的宿主细胞、或染色体中整合有编码所述sgRNA的多核苷酸的宿主细胞。在本发明一具体实施例中，所述sgRNA与所述融合蛋白可以在同一宿主细胞中表达，所述宿主细胞即为靶细胞。在本发明另一具体实施例中，所述基因编辑为体外基因编辑。

[0052] 本发明通过在基于xCas9的碱基编辑器上融合核定位元件及DNA双链断裂位点结合蛋白，开发出能够高效识别PAM序列为NG、GAA及GAT的位点并进行精确的且具有更宽广的基因组识别范围的碱基编辑工具。此外，所述融合蛋白具有高效、精确的单碱基编辑能力，

编辑产物纯度高,具有良好的产业化前景。

[0053] 以下通过特定的具体实例说明本发明的实施方式,本领域技术人员可由本说明书所揭露的内容轻易地了解本发明的其他优点与功效。本发明还可以通过另外不同的具体实施方式加以实施或应用,本说明书中的各项细节也可以基于不同观点与应用,在没有背离本发明的精神下进行各种修饰或改变。

[0054] 在进一步描述本发明具体实施方式之前,应理解,本发明的保护范围不局限于下述特定的具体实施方案;还应当理解,本发明实施例中使用的术语是为了描述特定的具体实施方案,而不是为了限制本发明的保护范围;在本发明说明书和权利要求书中,除非文中另外明确指出,单数形式“一个”、“一”和“这个”包括复数形式。

[0055] 当实施例给出数值范围时,应理解,除非本发明另有说明,每个数值范围的两个端点以及两个端点之间任何一个数值均可选用。除非另外定义,本发明中使用的所有技术和科学术语与本技术领域技术人员通常理解的意义相同。除实施例中使用的具体方法、设备、材料外,根据本技术领域的技术人员对现有技术的掌握及本发明的记载,还可以使用与本发明实施例中所述的方法、设备、材料相似或等同的现有技术的任何方法、设备和材料来实现本发明。

[0056] 除非另外说明,本发明中所公开的实验方法、检测方法、制备方法均采用本技术领域常规的分子生物学、生物化学、染色质结构和分析、分析化学、细胞培养、重组DNA技术及相关领域的常规技术。这些技术在现有文献中已有完善说明,具体可参见Sambrook等MOLECULAR CLONING:A LABORATORY MANUAL,Second edition,Cold Spring Harbor Laboratory Press,1989and Third edition,2001;Ausubel等,CURRENT PROTOCOLS IN MOLECULAR BIOLOGY,John Wiley&Sons,New York,1987and periodic updates;the series METHODS IN ENZYMOLOGY,Academic Press,San Diego;Wolffe,CHROMATIN STRUCTURE AND FUNCTION,Third edition,Academic Press,San Diego,1998;METHODS IN ENZYMOLOGY,Vol.304,Chromatin(P.M.Wassarman and A.P.Wolffe,eds.),Academic Press,San Diego,1999;和METHODS IN MOLECULAR BIOLOGY,Vol.119,Chromatin Protocols(P.B.Becker,ed.)Humana Press,Totowa,1999等。

[0057] 实施例中所使用的BPNLS-Gam-xBE3载体和BPNLS-xABE载体的构建方法如下:通过在xBE3载体中引入BPNLS和Gam元件以及在xABE载体中引入BPNLS元件分别对xBE3和xABE进行优化,xBE3和xABE质粒够自Addgene(#108380,#108382)。构建得到的BPNLS-Gam-xBE3和BPNLS-xABE载体序列分别如SEQ ID NO.1、SEQ ID NO.2所示。

[0058] 实施例1

[0059] 本实施例中,利用选定的内源基因sgRNA位点在HEK293T细胞中验证BPNLS-xABE的编辑效率,结果如图1和图2所示。

[0060] 1.1 sgRNA质粒构建

[0061] 选择10个内源基因位点(包括ATP7b^{M1169V},ATP7b^{I1230V},ATP7b^{W1353R},ATP7b^{Y713C},STK11^{L67P},ARHGEF18^{T270A},BCS1L^{V327A},BCS1L^{M48V},AIFM1^{D237G},ALPL^{C201R}等),根据目标序列设计sgRNA并合成oligos,所用到的sgRNA序列如SEQ ID.3～22所示。在每个sgNRA的上游序列5’端加accg序列,下游序列的5’端加aaac序列,因此用于合成的每个sgRNA上游序列为:5’-accgXXXXXXXXXXXXXXXXXXX(20nt)-3’,下游序列为:5’-aaacXXXXXXXXXXXXXXXXXXX

(20nt)-3’。经合成后，上下游序列通过预设程序(95℃，5min；95℃-85℃at-2℃/s；85℃-25℃at-0.1℃/s；hold at 4℃)进行退火，将退火产物连接到经过BsaI(NEB:R0539L)线性化的pGL3-U6-sgRNA(Addgene#51133)载体上，pGL3-U6-sgRNA载体的序列如SEQ ID.31所示。pGL3-U6-sgRNA载体的线性化体系如下：pGL3-U6-sgRNA 3μg；buffer(NEB:R0539L)6μL；BsaI 2μL；ddH2O补齐到60μL，37℃酶切过夜。sgRNA退火产物与线性化载体连接体系如下：T4连接酶buffer(NEB:M0202L)1μL，线性化载体20ng，经退火的oligo片段(10μM)5μL，T4连接酶(NEB:M0202L)0.5μL，ddH2O补齐到10μL，16℃连接过夜。将连接的载体转化大肠杆菌，挑菌鉴定，经过测序确认后对阳性克隆摇菌并提取质粒(Axygene:AP-MN-P-250G)后测定浓度，保存备用。

[0062] 1.2细胞培养与转染

[0063] 将HEK293T细胞(购自ATCC)接种于添加了10％FBS(v/v)的DMEM高糖培养基中(HyClone，SH30022.01B)，其中含1％Penicillin Streptomycin(v/v)(Gibco)，在含有5％CO2的37度细胞培养箱中进行培养。用于转染的细胞，前一天接种于12孔细胞培养板中进行培养，第二天观察细胞，当细胞生长到细胞密度为75％左右时，用含有10％FBS的新鲜无抗生素DMEM培养基换液，培养2小时使细胞状态恢复最佳。转染时，12孔板每孔转染的质粒用量分别是BPNLS-Gam-xBE3或BPNLS-xABE质粒1μg，sgRNA质粒0.6μg。将质粒混合后用50μl的Opti-MEM(Gibco，11058021)培养基稀释，作为试剂A，静置5分钟。同时，将2μl的Lipofectamine 2000转染试剂(Thermo，11668019)用50μl的Opti-MEM培养基稀释并混匀，作为试剂B，静置5分钟。将上述试剂A与试剂B混合并吹打均匀，静置20分钟。静置结束后将混合试剂逐滴加入待转染的12孔板细胞中，放回37度培养箱培养。转染6小时后将培养基换为含有10％FBS的DMEM培养基。转染48小时后在显微镜下观察荧光，并于转染后72小时收取细胞用流式细胞分选仪分选GFP阳性细胞。每个样品按照需求收集5000-10000个GFP阳性细胞，离心后用细胞裂解液裂解并鉴定基因型。细胞裂解液的主要成分为50mM KCl，1.5mM MgCl2，10mM Tris pH 8.0，0.5％Nonidet P-40，0.5％Tween 20，100g/ml protease K。

[0064] 1.3优化后的碱基编辑工具在内源基因位点编辑效率检测

[0065] 根据实验需求设计引物，以1.2所述GFP阳性细胞裂解产物为模板，对靶点附近序列进行PCR扩增，将扩增产物凝胶电泳检测目的条带，随后进行纯化。纯化后的PCR产物用于高通量深度测序或Sanger测序进行编辑效率的鉴定。用于目标位点序列扩增的体系如下：2Xbuffer(Vazyme，P505)25μL；dNTP 1μL；F(10pmol/μL)1μL；R(10pmol/μL)1μL；模板1μL；DNA聚合酶(Vazyme，P505)0.5μL；ddH2O补齐到50μL。扩增产物经过电泳凝胶确定目的条带之后按照下述步骤纯化：在PCR产物中加入三倍体积的PCR-A试剂(Axygen:AP-PCR-250G)，混合均匀后加入纯化柱，12000转/分钟离心1分钟，弃废液；加入700μL Buffer W2，离心1分钟，弃废液；加入700μL Buffer W2，离心1分钟，弃废液；将纯化柱套回收集管中，12000转/分钟空转1分钟；换新的1.5mL收集管，静置2分钟后往纯化柱中加入20μL水洗脱。相关结果如图1和图2所示。

[0066] 实施例2

[0067] 本实施例中，利用本发明中优化后的碱基编辑系统在人细胞系中模拟威尔逊病致病基因相关突变Atp7b^{T1033A}和Atp7b^{T1220M}，并构建分别含有这两种突变的突变细胞株，本方法将利用BPNLS-Gam-xBE3和BPNLS-xABE及相应的sgRNA实现，结果如图3所示。

[0068] 2.1 sgRNA质粒构建

[0069] 在突变位点附近,设计突变sgRNA Atp7b-27和sgRNA ATP7b-16,Atp7b-27上游序列为:5'-accgATTACCCATGGCGTCCCCAG-3'(SEQ ID.23),下游序列为:5'-aaacCTGGGGACGCCATGGGTAAT-3'(SEQ ID.24)。ATP7b-16上游序列为:5'-accgGATCACGGGGGACAACCGGA-3'(SEQ ID.25),下游序列为:5'-aaacTCCGGTTGTCCCCCGTGATC-3'(SEQ ID.26)。合成oligos之后按照1.1所述进行sgRNA质粒构建,经过测序鉴定序列正确后提取质粒保存备用。

[0070] 2.2细胞的培养与转染

[0071] HEK293T细胞按照1.2所述进行培养与转染。构建Atp7b^{T1033A}突变细胞系所用质粒为BPNLS-xABE和sgRNA ATP7b-27,质粒用量分别为1μg和0.5μg。构建Atp7b^{T1220M}突变细胞系所用质粒为BPNLS-Gam-xBE3和sgRNA Atp7b-16,质粒用量分别为1μg和0.5μg。

[0072] 2.3含有威尔逊病致病突变细胞株的构建

[0073] 细胞转染72小时后用胰酶消化重悬后应用流式细胞仪进行分选。选择GFP阳性较强的单细胞,分选至96孔板,37度培养箱培养两周左右,显微镜下挑选单克隆细胞。选取状态良好的单克隆细胞消化重悬,对所有挑选的单克隆细胞编号,一部分用于基因型鉴定,另一部分进行传代扩增。用于基因型鉴定的细胞离心后吸弃上清,对离心管底沉淀按照1.2所述用细胞裂解液进行裂解。根据靶序列设计引物,以细胞裂解产物为模板进行PCR扩增,PCR扩增体系如1.3所述。PCR扩增产物经凝胶电泳鉴定目的条带后进行Sanger测序。根据Sanger测序结果峰图选择致病位点Atp7b^{T1033A}和Atp7b^{T1220M}完全突变的纯合基因型单克隆细胞进行传代扩增及后续实验。相关结果见附图3。

[0074] 实施例3

[0075] 本实施例中,对获得的含有威尔逊病致病突变的纯合突变细胞株,利用本发明所述碱基编辑工具BPNLS-Gam-xBE3和BPNLS-xABEABE分别对Atp7b^{T1033A}和Atp7b^{T1220M}进行修复。本实施例将利用BPNLS-Gam-xBE3和BPNLS-xABEABE质粒结合相应的修复sgRNA进行突变位点的修复,结构如图4所示。

[0076] 3.1 sgRNA质粒构建

[0077] 在突变位点附近,设计修复sgRNA,合成oligos。Atp7b^{T1033A}和Atp7b^{T1220M}致病突变位点对应的修复sgRNA分别为:Atp7b-27-mut-Corretion sgRNA和Atp7b-16-mut-CorretionsgRNA。Atp7b-27-mut-Corretion上游序列为:5'-accgATGGGTAATGGTGCCAGTCT-3'(SEQ ID.27),下游序列为:5'-aaacAGACTGGCACCATTACCCAT-3'(SEQ ID.28)。Atp7b-16-mut-Corretion上游序列为:5'-accgGTCCCCCGTGATCAGAACCA-3'(SEQ ID.29),下游序列为:5'-aaac TGGTTCTGATCACGGGGGAC-3'(SEQ ID.30)。合成oligos之后,按照1.1所述进行sgRNA质粒构建,经过测序鉴定序列正确后提取质粒保存备用。

[0078] 3.2细胞培养与转染

[0079] 将2.3所述构建的含有威尔逊病致病突变位点Atp7b^{T1033A}和Atp7b^{T1220M}的细胞株培养于含有10%FBS的DMEM高糖培养基中进行扩增。致病突变修复时按照1.2所述方法进行细胞转染。修复Atp7b^{T1033A}致病位点需转染质粒BPNLS-Gam-xBE3和Atp7b-27-mut-Corretion质粒,质粒用量分别为1μg和0.6μg,修复Atp7b^{T1220M}致病位点需转染1μg BPNLS-xABE和0.6μg Atp7b-16-mut-Corretion。

[0080] 3.3修复效率的检测

[0081] 转染后72小时收集细胞通过细胞流式仪分选GFP强阳性细胞5000-10000个，离心后弃上清并用细胞裂解液裂解细胞沉淀。设计引物PCR扩增目的片段鉴定基因型。细胞裂解液的主要成分为50mM KCl，1.5mM MgCl2，10mM Tris pH 8.0，0.5％Nonidet P-40，0.5％Tween 20，100g/ml protease K。PCR扩增产物经电泳凝胶确认目的条带后进行Sanger测序，根据Sanger测序峰图检测突变位点是否得到修复，通过对Sanger测序峰图的量化检测突变位点的修复效率。相关结果如图4所示。

[0082] 综上所述，本发明有效克服了现有技术中的种种缺点而具高度产业利用价值。

[0083] 上述实施例仅例示性说明本发明的原理及其功效，而非用于限制本发明。任何熟悉此技术的人士皆可在不违背本发明的精神及范畴下，对上述实施例进行修饰或改变。因此，举凡所属技术领域中具有通常知识者在未脱离本发明所揭示的精神与技术思想下所完成的一切等效修饰或改变，仍应由本发明的权利要求所涵盖。

[0001] 序列表

[0002] <110> 上海科技大学

[0003] <120> 一种碱基编辑工具及其用途

[0004] <160> 39

[0005] <170> SIPOSequenceListing 1.0

[0006] <210> 1

[0007] <211> 9201

[0008] <212> DNA

[0009] <213> 人工序列(Artificial Sequence)

[0010] <400> 1

[0011] cctgtcaagg tatccccacg tcactctgtt tatttacatc gcaaggctgt accaccacgc 60

[0012] tgacccccgc aatcgacaag gcctgcggga tttgatctct tcaggtgtga ctatccaaat 120

[0013] tatgactgag caggagtcag gatactgctg gagaaacttt gtgaattata gcccgagtaa 180

[0014] tgaagcccac tggcctaggt atccccatct gtgggtacga ctgtacgttc ttgaactgta 240

[0015] ctgcatcata ctgggcctgc ctccttgtct caacattctg agaaggaagc agccacagct 300

[0016] gacattcttt accatcgctc ttcagtcttg tcattaccag cgactgcccc cacacattct 360

[0017] ctgggccacc gggttgaaat ctggtggttc ttctggtggt tctagcggca gcgagactcc 420

[0018] cgggacctca gagtccgcca cacccgaaag ttctggtggt tcttctggtg gttctgataa 480

[0019] aaagtattct attggtttag ccatcggcac taattccgtg ggctgggccg tgatcaccga 540

[0020] cgagtacaag gtgcccagca agaaattcaa ggtgctgggc aacaccgacc ggcacagcat 600

[0021] caagaagaac ctgatcggag ccctgctgtt cgacagcggc gaaacagccg aggccacccg 660

[0022] gctgaagaga accgccagaa gaagatacac cagacggaag aaccggatct gctatctgca 720

[0023] agagatcttc agcaacgaga tggccaaggt ggacgacagc ttcttccaca gactggaaga 780

[0024] gtccttcctg gtggaagagg ataagaagca cgagcggcac cccatcttcg gcaacatcgt 840

[0025] ggacgaggtg gcctaccacg agaagtaccc caccatctac cacctgagaa agaaactggt 900

[0026] ggacagcacc gacaaggccg acctgcggct gatctatctg gccctggccc acatgatcaa 960

[0027] gttccggggc cacttcctga tcgagggcga cctgaacccc gacaacagcg acgtggacaa 1020

[0028] gctgttcatc cagctggtgc agacctacaa ccagctgttc gaggaaaacc ccatcaacgc 1080

[0029] cagcggcgtg gacgccaagg ccatcctgtc tgccagactg agcaagagca gacggctgga 1140

[0030] aaatctgatc gcccagctgc ccggcgagaa gaagaatggc ctgttcggaa acctgattgc 1200

[0031] cctgagcctg ggcctgaccc ccaacttcaa gagcaacttc gacctggccg aggataccaa 1260

[0032] actgcagctg agcaaggaca cctacgacga cgacctggac aacctgctgg cccagatcgg 1320

[0033] cgaccagtac gccgacctgt ttctggccgc caagaacctg tccgacgcca tcctgctgag 1380

[0034] cgacatcctg agagtgaaca ccgagatcac caaggccccc ctgagcgcct ctatgatcaa 1440

[0035] gctgtacgac gagcaccacc aggacctgac cctgctgaaa gctctcgtgc ggcagcagct 1500

[0036] gcctgagaag tacaaagaga ttttcttcga ccagagcaag aacggctacg ccggctacat 1560

[0037] tgacggcgga gccagccagg aagagttcta caagttcatc aagcccatcc tggaaaagat 1620

[0038] ggacggcacc gaggaactgc tcgtgaagct gaacagagag gacctgctgc ggaagcagcg 1680

[0039] gaccttcgac aacggcatca tcccccacca gatccacctg ggagagctgc acgccattct 1740

[0040] gcggcggcag gaagattttt acccattcct gaaggacaac cgggaaaaga tcgagaagat 1800

[0041] cctgaccttc cgcatcccct actacgtggg ccctctggcc aggggaaaca gcagattcgc 1860

[0042] ctggatgacc agaaagagcg aggaaaccat cacccctgg aacttcgaga aggtggtgga 1920
[0043] caagggcgct tccgcccaga gcttcatcga gcggatgacc aacttcgata agaacctgcc 1980
[0044] caacgagaag gtgctgccca agcacagcct gctgtacgag tacttcaccg tgtataacga 2040
[0045] gctgaccaaa gtgaaatacg tgaccgaggg aatgagaaag cccgccttcc tgagcggcga 2100
[0046] ccagaaaaag gccatcgtgg acctgctgtt caagaccaac cggaaagtga ccgtgaagca 2160
[0047] gctgaaagag gactacttca agaaaatcga gtgcttcgac tccgtggaaa tctccggcgt 2220
[0048] ggaagatcgg ttcaacgcct ccctgggcac ataccacgat ctgctgaaaa ttatcaagga 2280
[0049] caaggacttc ctggacaatg aggaaaacga ggacattctg gaagatatcg tgctgaccct 2340
[0050] gacactgttt gaggacagag agatgatcga ggaacggctg aaaacctatg cccacctgtt 2400
[0051] cgacgacaaa gtgatgaagc agctgaagcg gcggagatac accggctggg gcaggctgag 2460
[0052] ccggaagctg atcaacggca tccgggacaa gcagtccggc aagacaatcc tggatttcct 2520
[0053] gaagtccgac ggcttcgcca acagaaactt catccagctg atccacgacg acagcctgac 2580
[0054] ctttaaagag gacatccaga aagcccaggt gtccggccag ggcgatagcc tgcacgagca 2640
[0055] cattgccaat ctggccggca gccccgccat taagaagggc atcctgcaga cagtgaaggt 2700
[0056] ggtggacgag ctcgtgaaag tgatgggccg gcacaagccc gagaacatcg tgatcgaaat 2760
[0057] ggccagagag aaccagacca cccagaaggg acagaagaac agccgcgaga gaatgaagcg 2820
[0058] gatcgaagag ggcatcaaag agctgggcag ccagatcctg aaagaacacc ccgtggaaaa 2880
[0059] cacccagctg cagaacgaga agctgtacct gtactacctg cagaatgggc gggatatgta 2940
[0060] cgtggaccag gaactggaca tcaaccggct gtccgactac gatgtggacc atatcgtgcc 3000
[0061] tcagagcttt ctgaaggacg actccatcga caacaaggtg ctgaccagaa gcgacaagaa 3060
[0062] ccggggcaag agcgacaacg tgccctccga agaggtcgtg aagaagatga agaactactg 3120
[0063] gcggcagctg ctgaacgcca agctgattac ccagagaaag ttcgacaatc tgaccaaggc 3180
[0064] cgagagaggc ggcctgagcg aactggataa ggccggcttc atcaagagac agctggtgga 3240
[0065] aacccggcag atcacaaagc acgtggcaca gatcctggac tcccggatga acactaagta 3300
[0066] cgacgagaat gacaagctga tccgggaagt gaaagtgatc accctgaagt ccaagctggt 3360
[0067] gtccgatttc cggaaggatt tccagtttta caaagtgcgc gagatcaaca actaccacca 3420
[0068] cgcccacgac gcctacctga acgccgtcgt gggaaccgcc ctgatcaaaa agtaccctaa 3480
[0069] gctggaaagc gagttcgtgt acggcgacta caaggtgtac gacgtgcgga agatgatcgc 3540
[0070] caagagcgag caggaaatcg gcaaggctac cgccaagtac ttcttctaca gcaacatcat 3600
[0071] gaacttttc aagaccgaga ttaccctggc caacggcgag atccggaagc ggcctctgat 3660
[0072] cgagacaaac ggcgaaaccg gggagatcgt gtgggataag ggccgggatt ttgccaccgt 3720
[0073] gcggaaagtg ctgagcatgc cccaagtgaa tatcgtgaaa aagaccgagg tgcagacagg 3780
[0074] cggcttcagc aaagagtcta tcctgcccaa gaggaacagc gataagctga tcgccagaaa 3840
[0075] gaaggactgg gaccctaaga agtacggcgg cttcgacagc cccaccgtgg cctattctgt 3900
[0076] gctggtggtg gccaaagtgg aaaagggcaa gtccaagaaa ctgaagagtg tgaaagagct 3960
[0077] gctggggatc accatcatgg aaagaagcag cttcgagaag aatcccatcg actttctgga 4020
[0078] agccaagggc tacaaagaag tgaaaaagga cctgatcatc aagctgccta agtactccct 4080
[0079] gttcgagctg gaaaacggcc ggaagagaat gctggcctct gccggcgtgc tgcagaaggg 4140
[0080] aaacgaactg gccctgccct ccaaatatgt gaacttcctg tacctggcca gccactatga 4200
[0081] gaagctgaag ggctcccccg aggataatga gcagaaacag ctgtttgtgg aacagcacaa 4260
[0082] gcactacctg gacgagatca tcgagcagat cagcgagttc tccaagagag tgatcctggc 4320
[0083] cgacgctaat ctggacaaag tgctgtccgc ctacaacaag caccgggata agcccatcag 4380

[0084] agagcaggcc gagaatatca tccacctgtt taccctgacc aatctgggag cccctgccgc 4440

[0085] cttcaagtac tttgacacca ccatcgaccg gaagaggtac accagcacca aagaggtgct 4500

[0086] ggacgccacc ctgatccacc agagcatcac cggcctgtac gagacacgga tcgacctgtc 4560

[0087] tcagctggga ggcgactctg gtggttctac taatctgtca gatattattg aaaaggagac 4620

[0088] cggtaagcaa ctggttatcc aggaatccat cctcatgctc ccagaggagg tggaagaagt 4680

[0089] cattgggaac aagccggaaa gcgatatact cgtgcacacc gcctacgacg agagcaccga 4740

[0090] cgagaatgtc atgcttctga ctagcgacgc ccctgaatac aagccttggg ctctggtcat 4800

[0091] acaggatagc aacggtgaga acaagattaa gatgctctct ggtggttctc ccaagaagaa 4860

[0092] gaggaaagtc taaccggtca tcatcaccat caccattgag tttaaacccg ctgatcagcc 4920

[0093] tcgactgtgc cttctagttg ccagccatct gttgtttgcc cctcccccgt gccttccttg 4980

[0094] accctggaag gtgccactcc cactgtcctt tcctaataaa atgaggaaat tgcatcgcat 5040

[0095] tgtctgagta ggtgtcattc tattctgggg ggtggggtgg ggcaggacag caagggggag 5100

[0096] gattgggaag acaatagcag gcatgctggg gatgcggtgg gctctatggc ttctgaggcg 5160

[0097] gaaagaacca gctggggctc gataccgtcg acctctagct agagcttggc gtaatcatgg 5220

[0098] tcatagctgt ttcctgtgtg aaattgttat ccgctcacaa ttccacacaa catacgagcc 5280

[0099] ggaagcataa agtgtaaagc ctagggtgcc taatgagtga gctaactcac attaattgcg 5340

[0100] ttgcgctcac tgcccgcttt ccagtcggga aacctgtcgt gccagctgca ttaatgaatc 5400

[0101] ggccaacgcg cggggagagg cggtttgcgt attgggcgct cttccgcttc ctcgctcact 5460

[0102] gactcgctgc gctcggtcgt tcggctgcgg cgagcggtat cagctcactc aaaggcggta 5520

[0103] atacggttat ccacagaatc aggggataac gcaggaaaga acatgtgagc aaaaggccag 5580

[0104] caaaaggcca ggaaccgtaa aaaggccgcg ttgctggcgt ttttccatag gctccgcccc 5640

[0105] cctgacgagc atcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc gacaggacta 5700

[0106] taaagatacc aggcgtttcc ccctggaagc tccctcgtgc gctctcctgt tccgaccctg 5760

[0107] ccgcttaccg gatacctgtc cgcctttctc ccttcgggaa gcgtggcgct ttctcatagc 5820

[0108] tcacgctgta ggtatctcag ttcggtgtag gtcgttcgct ccaagctggg ctgtgtgcac 5880

[0109] gaacccccccg ttcagcccga ccgctgcgcc ttatccggta actatcgtct tgagtccaac 5940

[0110] ccggtaagac acgacttatc gccactggca gcagccactg gtaacaggat tagcagagcg 6000

[0111] aggtatgtag gcggtgctac agagttcttg aagtggtggc ctaactacgg ctacactaga 6060

[0112] agaacagtat ttggtatctg cgctctgctg aagccagtta ccttcggaaa aagagttggt 6120

[0113] agctcttgat ccggcaaaca aaccaccgct ggtagcggtg gtttttttgt ttgcaagcag 6180

[0114] cagattacgc gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc tacggggtct 6240

[0115] gacgctcagt ggaacgaaaa ctcacgttaa gggattttgg tcatgagatt atcaaaaagg 6300

[0116] atcttcacct agatcctttt aaattaaaaa tgaagtttta aatcaatcta aagtatatat 6360

[0117] gagtaaactt ggtctgacag ttaccaatgc ttaatcagtg aggcacctat ctcagcgatc 6420

[0118] tgtctatttc gttcatccat agttgcctga ctccccgtcg tgtagataac tacgatacgg 6480

[0119] gagggcttac catctggccc cagtgctgca atgataccgc gagacccacg ctcaccggct 6540

[0120] ccagatttat cagcaataaa ccagccagcc ggaagggccg agcgcagaag tggtcctgca 6600

[0121] actttatccg cctccatcca gtctattaat tgttgccggg aagctagagt aagtagttcg 6660

[0122] ccagttaata gtttgcgcaa cgttgttgcc attgctacag gcatcgtggt gtcacgctcg 6720

[0123] tcgtttggta tggcttcatt cagctccggt tcccaacgat caaggcgagt tacatgatcc 6780

[0124] cccatgttgt gcaaaaaagc ggttagctcc ttcggtcctc cgatcgttgt cagaagtaag 6840

[0125] ttggccgcag tgttatcact catggttatg gcagcactgc ataattctct tactgtcatg 6900

[0126] ccatccgtaa gatgcttttc tgtgactggt gagtactcaa ccaagtcatt ctgagaatag 6960

[0127] tgtatgcggc gaccgagttg ctcttgcccg gcgtcaatac gggataatac cgcgccacat 7020

[0128] agcagaactt taaaagtgct catcattgga aaacgttctt cggggcgaaa actctcaagg 7080

[0129] atcttaccgc tgttgagatc cagttcgatg taacccactc gtgcacccaa ctgatcttca 7140

[0130] gcatctttta ctttcaccag cgtttctggg tgagcaaaaa caggaaggca aaatgccgca 7200

[0131] aaaaagggaa taagggcgac acggaaatgt tgaatactca tactcttcct ttttcaatat 7260

[0132] tattgaagca tttatcaggg ttattgtctc atgagcggat acatatttga atgtatttag 7320

[0133] aaaaataaac aaataggggt tccgcgcaca tttccccgaa aagtgccacc tgacgtcgac 7380

[0134] ggatcgggag atcgatctcc cgatccccta gggtcgactc tcagtacaat ctgctctgat 7440

[0135] gccgcatagt taagccagta tctgctccct gcttgtgtgt tggaggtcgc tgagtagtgc 7500

[0136] gcgagcaaaa tttaagctac aacaaggcaa ggcttgaccg acaattgcat gaagaatctg 7560

[0137] cttagggtta ggcgttttgc gctgcttcgc gatgtacggg ccagatatac gcgttgacat 7620

[0138] tgattattga ctagttatta atagtaatca attacggggt cattagttca tagcccatat 7680

[0139] atggagttcc gcgttacata acttacggta aatggcccgc ctggctgacc gcccaacgac 7740

[0140] ccccgcccat tgacgtcaat aatgacgtat gttcccatag taacgccaat agggactttc 7800

[0141] cattgacgtc aatgggtgga gtatttacgg taaactgccc acttggcagt acatcaagtg 7860

[0142] tatcatatgc caagtacgcc ccctattgac gtcaatgacg gtaaatggcc cgcctggcat 7920

[0143] tatgcccagt acatgacctt atgggacttt cctacttggc agtacatcta cgtattagtc 7980

[0144] atcgctatta ccatggtgat gcggttttgg cagtacatca atgggcgtgg atagcggttt 8040

[0145] gactcacggg gatttccaag tctccacccc attgacgtca atgggagttt gttttggcac 8100

[0146] caaaatcaac gggactttcc aaaatgtcgt aacaactccg ccccattgac gcaaatgggc 8160

[0147] ggtaggcgtg tacggtggga ggtctatata agcagagctg gtttagtgaa ccgtcagatc 8220

[0148] cgctagagat ccgcggccgc taatacgact cactataggg agagccgcca ccatgaagcg 8280

[0149] caccgccgat ggttccgagt tcgaaagccc caaaaaaaag cgcaaggtcg ctaaaccagc 8340

[0150] aaaacgtatc aagagtgccg cagcggctta tgtgccacaa aaccgcgatg cggtgattac 8400

[0151] cgatattaaa cgcatcgggg atttacagcg cgaagcatca cgtctggaaa cggaaatgaa 8460

[0152] tgatgccatc gcggaaatta cggagaaatt tgcggcccgg attgcaccga ttaaaaccga 8520

[0153] tattgaaacc ctttcaaaag gcgttcaggg atggtgtgaa gcgaaccgcg acgaactgac 8580

[0154] gaacggcggc aaagtgaaga cggcgaatct tgtcaccggt gatgtatcgt ggcgggtccg 8640

[0155] tccaccatca gtaagtattc gtggtatgga tgcagtgatg gaaacgctgg agcgtcttgg 8700

[0156] cctgcaacgc tttattcgca cgaagcagga aatcaacaag gaagcgattt tactggaacc 8760

[0157] gaaagcggtc gcaggcgttg ccggaattac agttaaatca ggcattgagg atttttctat 8820

[0158] tattccattt gaacaggaag ccggtattag cggcagcgag actcccggga cctcagagtc 8880

[0159] cgctacaccc gaaagtagct cagagactgg cccagtggct gtggacccca cattgagacg 8940

[0160] gcggatcgag ccccatgagt ttgaggtatt cttcgatccg agagagctcc gcaaggagac 9000

[0161] ctgcctgctt tacgaaatta attgggggg ccggcactcc atttggcgac atacatcaca 9060

[0162] gaacactaac aagcacgtcg aagtcaactt catcgagaag ttcacgacag aaagatattt 9120

[0163] ctgtccgaac acaaggtgca gcattacctg gtttctcagc tggagcccat gcggcgaatg 9180

[0164] tagtagggcc atcactgaat t 9201

[0165] <210> 2

[0166] <211> 8792

[0167] <212> DNA

[0168] <213> 人工序列(Artificial Sequence)

[0169] <400> 2

[0170] tctggtggtt ctcccaagaa gaagaggaaa gtctaaccgg tcatcatcac catcaccatt 60

[0171] gagtttaaac ccgctgatca gcctcgactg tgccttctag ttgccagcca tctgttgttt 120

[0172] gcccctcccc cgtgccttcc ttgaccctgg aaggtgccac tcccactgtc ctttcctaat 180

[0173] aaaatgagga aattgcatcg cattgtctga gtaggtgtca ttctattctg gggggtgggg 240

[0174] tggggcagga cagcaagggg gaggattggg aagacaatag caggcatgct ggggatgcgg 300

[0175] tgggctctat ggcttctgag gcggaaagaa ccagctgggg ctcgataccg tcgacctcta 360

[0176] gctagagctt ggcgtaatca tggtcatagc tgtttcctgt gtgaaattgt tatccgctca 420

[0177] caattccaca caacatacga gccggaagca taaagtgtaa agcctagggt gcctaatgag 480

[0178] tgagctaact cacattaatt gcgttgcgct cactgcccgc tttccagtcg ggaaacctgt 540

[0179] cgtgccagct gcattaatga atcggccaac gcgcggggag aggcggtttg cgtattgggc 600

[0180] gctcttccgc ttcctcgctc actgactcgc tgcgctcggt cgttcggctg cggcgagcgg 660

[0181] tatcagctca ctcaaaggcg gtaatacggt tatccacaga atcaggggat aacgcaggaa 720

[0182] agaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg taaaaaggcc gcgttgctgg 780

[0183] cgtttttcca taggctccgc ccccctgacg agcatcacaa aaatcgacgc tcaagtcaga 840

[0184] ggtggcgaaa cccgacagga ctataaagat accaggcgtt tccccctgga agctccctcg 900

[0185] tgcgctctcc tgttccgacc ctgccgctta ccggatacct gtccgccttt ctcccttcgg 960

[0186] gaagcgtggc gctttctcat agctcacgct gtaggtatct cagttcggtg taggtcgttc 1020

[0187] gctccaagct gggctgtgtg cacgaacccc ccgttcagcc cgaccgctgc gccttatccg 1080

[0188] gtaactatcg tcttgagtcc aacccggtaa gacacgactt atcgccactg gcagcagcca 1140

[0189] ctggtaacag gattagcaga gcgaggtatg taggcggtgc tacagagttc ttgaagtggt 1200

[0190] ggcctaacta cggctacact agaagaacag tatttggtat ctgcgctctg ctgaagccag 1260

[0191] ttaccttcgg aaaaagagtt ggtagctctt gatccggcaa acaaaccacc gctggtagcg 1320

[0192] gtggtttttt tgtttgcaag cagcagatta cgcgcagaaa aaaaggatct caagaagatc 1380

[0193] ctttgatctt ttctacgggg tctgacgctc agtggaacga aaactcacgt taagggattt 1440

[0194] tggtcatgag attatcaaaa aggatcttca cctagatcct tttaaattaa aaatgaagtt 1500

[0195] ttaaatcaat ctaaagtata tatgagtaaa cttggtctga cagttaccaa tgcttaatca 1560

[0196] gtgaggcacc tatctcagcg atctgtctat ttcgttcatc catagttgcc tgactccccg 1620

[0197] tcgtgtagat aactacgata cgggagggct taccatctgg ccccagtgct gcaatgatac 1680

[0198] cgcgagaccc acgctcaccg gctccagatt tatcagcaat aaaccagcca gccggaaggg 1740

[0199] ccgagcgcag aagtggtcct gcaactttat ccgcctccat ccagtctatt aattgttgcc 1800

[0200] gggaagctag agtaagtagt tcgccagtta atagtttgcg caacgttgtt gccattgcta 1860

[0201] caggcatcgt ggtgtcacgc tcgtcgtttg gtatggcttc attcagctcc ggttcccaac 1920

[0202] gatcaaggcg agttacatga tcccccatgt tgtgcaaaaa agcggttagc tccttcggtc 1980

[0203] ctccgatcgt tgtcagaagt aagttggccg cagtgttatc actcatggtt atggcagcac 2040

[0204] tgcataattc tcttactgtc atgccatccg taagatgctt ttctgtgact ggtgagtact 2100

[0205] caaccaagtc attctgagaa tagtgtatgc ggcgaccgag ttgctcttgc ccggcgtcaa 2160

[0206] tacgggataa taccgcgcca catagcagaa ctttaaaagt gctcatcatt ggaaaacgtt 2220

[0207] cttcggggcg aaaactctca aggatcttac cgctgttgag atccagttcg atgtaaccca 2280

[0208] ctcgtgcacc caactgatct tcagcatctt ttactttcac cagcgtttct gggtgagcaa 2340

[0209] aaacaggaag gcaaaatgcc gcaaaaaagg gaataagggc gacacggaaa tgttgaatac 2400

[0210] tcatactctt ccttttttcaa tattattgaa gcatttatca gggttattgt ctcatgagcg 2460

[0211] gatacatatt tgaatgtatt tagaaaaata aacaaatagg ggttccgcgc acatttcccc 2520

[0212] gaaaagtgcc acctgacgtc gacggatcgg gagatcgatc tcccgatccc ctagggtcga 2580

[0213] ctctcagtac aatctgctct gatgccgcat agttaagcca gtatctgctc cctgcttgtg 2640

[0214] tgttggaggt cgctgagtag tgcgcgagca aaatttaagc tacaacaagg caaggcttga 2700

[0215] ccgacaattg catgaagaat ctgcttaggg ttaggcgttt tgcgctgctt cgcgatgtac 2760

[0216] gggccagata tacgcgttga cattgattat tgactagtta ttaatagtaa tcaattacgg 2820

[0217] ggtcattagt tcatagccca tatatggagt tccgcgttac ataacttacg gtaaatggcc 2880

[0218] cgcctggctg accgcccaac gacccccgcc cattgacgtc aataatgacg tatgttccca 2940

[0219] tagtaacgcc aatagggact ttccattgac gtcaatgggt ggagtattta cggtaaactg 3000

[0220] cccacttggc agtacatcaa gtgtatcata tgccaagtac gccccctatt gacgtcaatg 3060

[0221] acggtaaatg gcccgcctgg cattatgccc agtacatgac cttatgggac tttcctactt 3120

[0222] ggcagtacat ctacgtatta gtcatcgcta ttaccatggt gatgcggttt tggcagtaca 3180

[0223] tcaatgggcg tggatagcgg tttgactcac ggggatttcc aagtctccac cccattgacg 3240

[0224] tcaatgggag tttgttttgg caccaaaatc aacgggactt tccaaaatgt cgtaacaact 3300

[0225] ccgccccatt gacgcaaatg ggcggtaggc gtgtacggtg ggaggtctat ataagcagag 3360

[0226] ctggtttagt gaaccgtcag atccgctaga gatccgcggc cgctaatacg actcactata 3420

[0227] gggccaccat gaagcgcacc gccgatggtt ccgagttcga aagccccaaa aaaaagcgca 3480

[0228] aggtcccgag agccgccacc atgtccgaag tcgagttttc ccatgagtac tggatgagac 3540

[0229] acgcattgac tctcgcaaag agggcttggg atgaacgcga ggtgcccgtg ggggcagtac 3600

[0230] tcgtgcataa caatcgcgta atcggcgaag gttggaatag gccgatcgga cgccacgacc 3660

[0231] ccactgcaca tgcggaaatc atggcccttc gacagggagg gcttgtgatg cagaattatc 3720

[0232] gacttatcga tgcgacgctg tacgtcacgc ttgaaccttg cgtaatgtgc gcgggagcta 3780

[0233] tgattcactc ccgcattgga cgagttgtat tcggtgcccg cgacgccaag acgggtgccg 3840

[0234] caggttcact gatggacgtg ctgcatcacc caggcatgaa ccaccgggta gaaatcacag 3900

[0235] aaggcatatt ggcggacgaa tgtgcggcgc tgttgtccga ctttttcgc atgcggaggc 3960

[0236] aggagatcaa ggcccagaaa aaagcacaat cctctactga cagcggcggc agcagcggcg 4020

[0237] gcagcagcgg cagcgagact cccgggacct cagagtccgc cacacccgaa agtagcggcg 4080

[0238] gcagcagcgg cggcagctcc gaagtcgagt tttcccatga gtactggatg agacacgcat 4140

[0239] tgactctcgc aaagagggct cgggatgaac gcgaggtgcc cgtgggggca gtactcgtgc 4200

[0240] ttaacaatcg cgtaatcggc gaaggttgga atagggcgat cggactccac gaccccactg 4260

[0241] cacatgcgga aatcatggcc cttcgacagg gagggcttgt gatgcagaat tatcgactta 4320

[0242] tcgatgcgac gctgtacgtc acgtttgaac cttgcgtaat gtgcgcggga gctatgattc 4380

[0243] actcccgcat tggacgagtt gtattcggtg tccgcaacgc caagacgggt gccgcaggtt 4440

[0244] cactgatgga cgtgctgcat tacccaggca tgaaccaccg ggtagaaatc acagaaggca 4500

[0245] tattggcgga cgaatgtgcg gcgctgttgt gctacttttt tcgcatgccg aggcaggtgt 4560

[0246] tcaatgccca gaaaaaagca caatcctcta ctgacagcgg cggcagcagc ggcggcagca 4620

[0247] gcggcagcga gactcccggg acctcagagt ccgccacacc cgaaagtagc ggcggcagca 4680

[0248] gcggcggcag cgataaaaag tattctattg gtttagccat cggcactaat tccgtgggct 4740

[0249] gggccgtgat caccgacgag tacaaggtgc ccagcaagaa attcaaggtg ctgggcaaca 4800

[0250] ccgaccggca cagcatcaag aagaacctga tcggagccct gctgttcgac agcggcgaaa 4860

[0251] cagccgaggc cacccggctg aagagaaccg ccagaagaag atacaccaga cggaagaacc 4920

[0252] ggatctgcta tctgcaagag atcttcagca acgagatggc caaggtggac gacagcttct 4980

[0253] tccacagact ggaagagtcc ttcctggtgg aagaggataa gaagcacgag cggcacccca 5040

[0254] tcttcggcaa catcgtggac gaggtggcct accacgagaa gtaccccacc atctaccacc 5100

[0255] tgagaaagaa actggtggac agcaccgaca aggccgacct gcggctgatc tatctggccc 5160

[0256] tggcccacat gatcaagttc cggggccact tcctgatcga gggcgacctg aaccccgaca 5220

[0257] acagcgacgt ggacaagctg ttcatccagc tggtgcagac ctacaaccag ctgttcgagg 5280

[0258] aaaaccccat caacgccagc ggcgtggacg ccaaggccat cctgtctgcc agactgagca 5340

[0259] agagcagacg gctggaaaat ctgatcgccc agctgcccgg cgagaagaag aatggcctgt 5400

[0260] tcggaaacct gattgccctg agcctgggcc tgacccccaa cttcaagagc aacttcgacc 5460

[0261] tggccgagga taccaaactg cagctgagca aggacaccta cgacgacgac ctggacaacc 5520

[0262] tgctggccca gatcggcgac cagtacgccg acctgtttct ggccgccaag aacctgtccg 5580

[0263] acgccatcct gctgagcgac atcctgagag tgaacaccga gatcaccaag gccccccctga 5640

[0264] gcgcctctat gatcaagctg tacgacgagc accaccagga cctgaccctg ctgaaagctc 5700

[0265] tcgtgcggca gcagctgcct gagaagtaca aagagatttt cttcgaccag agcaagaacg 5760

[0266] gctacgccgg ctacattgac ggcggagcca gccaggaaga gttctacaag ttcatcaagc 5820

[0267] ccatcctgga aaagatggac ggcaccgagg aactgctcgt gaagctgaac agagaggacc 5880

[0268] tgctgcggaa gcagcggacc ttcgacaacg gcatcatccc ccaccagatc cacctgggag 5940

[0269] agctgcacgc cattctgcgg cggcaggaag atttttaccc attcctgaag gacaaccggg 6000

[0270] aaaagatcga gaagatcctg accttccgca tcccctacta cgtgggccct ctggccaggg 6060

[0271] gaaacagcag attcgcctgg atgaccagaa agagcgagga aaccatcacc ccctggaact 6120

[0272] tcgagaaggt ggtggacaag ggcgcttccg cccagagctt catcgagcgg atgaccaact 6180

[0273] tcgataagaa cctgcccaac gagaaggtgc tgcccaagca cagcctgctg tacgagtact 6240

[0274] tcaccgtgta taacgagctg accaaagtga aatacgtgac cgagggaatg agaaagcccg 6300

[0275] ccttcctgag cggcgaccag aaaaaggcca tcgtggacct gctgttcaag accaaccgga 6360

[0276] aagtgaccgt gaagcagctg aaagaggact acttcaagaa aatcgagtgc ttcgactccg 6420

[0277] tggaaatctc cggcgtggaa gatcggttca acgcctccct gggcacatac cacgatctgc 6480

[0278] tgaaaattat caaggacaag gacttcctgg acaatgagga aaacgaggac attctggaag 6540

[0279] atatcgtgct gaccctgaca ctgtttgagg acagagagat gatcgaggaa cggctgaaaa 6600

[0280] cctatgccca cctgttcgac gacaaagtga tgaagcagct gaagcggcgg agatacaccg 6660

[0281] gctggggcag gctgagccgg aagctgatca acggcatccg ggacaagcag tccggcaaga 6720

[0282] caatcctgga tttcctgaag tccgacggct tcgccaacag aaacttcatc cagctgatcc 6780

[0283] acgacgacag cctgaccttt aaagaggaca tccagaaagc ccaggtgtcc ggccagggcg 6840

[0284] atagcctgca cgagcacatt gccaatctgg ccggcagccc cgccattaag aagggcatcc 6900

[0285] tgcagacagt gaaggtggtg gacgagctcg tgaaagtgat gggccggcac aagcccgaga 6960

[0286] acatcgtgat cgaaatggcc agagagaacc agaccaccca gaagggacag aagaacagcc 7020

[0287] gcgagagaat gaagcggatc gaagagggca tcaaagagct gggcagccag atcctgaaag 7080

[0288] aacaccccgt ggaaaacacc cagctgcaga acgagaagct gtacctgtac tacctgcaga 7140

[0289] atgggcggga tatgtacgtg gaccaggaac tggacatcaa ccggctgtcc gactacgatg 7200

[0290] tggaccatat cgtgcctcag agctttctga aggacgactc catcgacaac aaggtgctga 7260

[0291] ccagaagcga caagaaccgg ggcaagagcg acaacgtgcc ctccgaagag gtcgtgaaga 7320

[0292] agatgaagaa ctactggcgg cagctgctga acgccaagct gattacccag agaaagttcg 7380

[0293] acaatctgac caaggccgag agaggcggcc tgagcgaact ggataaggcc ggcttcatca 7440

[0294] agagacagct ggtggaaacc cggcagatca caaagcacgt ggcacagatc ctggactccc 7500
[0295] ggatgaacac taagtacgac gagaatgaca agctgatccg ggaagtgaaa gtgatcaccc 7560
[0296] tgaagtccaa gctggtgtcc gatttccgga aggatttcca gttttacaaa gtgcgcgaga 7620
[0297] tcaacaacta ccaccacgcc cacgacgcct acctgaacgc cgtcgtggga accgccctga 7680
[0298] tcaaaaagta ccctaagctg gaaagcgagt tcgtgtacgg cgactacaag gtgtacgacg 7740
[0299] tgcggaagat gatcgccaag agcgagcagg aaatcggcaa ggctaccgcc aagtacttct 7800
[0300] tctacagcaa catcatgaac tttttcaaga ccgagattac cctggccaac ggcgagatcc 7860
[0301] ggaagcggcc tctgatcgag acaaacggcg aaaccgggga gatcgtgtgg gataagggcc 7920
[0302] gggattttgc caccgtgcgg aaagtgctga gcatgcccca agtgaatatc gtgaaaaaga 7980
[0303] ccgaggtgca gacaggcggc ttcagcaaag agtctatcct gcccaagagg aacagcgata 8040
[0304] agctgatcgc cagaaagaag gactgggacc ctaagaagta cggcggcttc gacagcccca 8100
[0305] ccgtggccta ttctgtgctg gtggtggcca aagtggaaaa gggcaagtcc aagaaactga 8160
[0306] agagtgtgaa agagctgctg gggatcacca tcatggaaag aagcagcttc gagaagaatc 8220
[0307] ccatcgactt tctggaagcc aagggctaca aagaagtgaa aaaggacctg atcatcaagc 8280
[0308] tgcctaagta ctccctgttc gagctggaaa acggccggaa gagaatgctg gcctctgccg 8340
[0309] gcgtgctgca gaagggaaac gaactggccc tgccctccaa atatgtgaac ttcctgtacc 8400
[0310] tggccagcca ctatgagaag ctgaagggct cccccgagga taatgagcag aaacagctgt 8460
[0311] ttgtggaaca gcacaagcac tacctggacg agatcatcga gcagatcagc gagttctcca 8520
[0312] agagagtgat cctggccgac gctaatctgg acaaagtgct gtccgcctac aacaagcacc 8580
[0313] gggataagcc catcagagag caggccgaga atatcatcca cctgtttacc ctgaccaatc 8640
[0314] tgggagcccc tgccgccttc aagtactttg acaccaccat cgaccggaag aggtacacca 8700
[0315] gcaccaaaga ggtgctggac gccaccctga tccaccagag catcaccggc ctgtacgaga 8760
[0316] cacggatcga cctgtctcag ctgggaggcg ac 8792
[0317] <210> 3
[0318] <211> 24
[0319] <212> DNA
[0320] <213> 人工序列(Artificial Sequence)
[0321] <400> 3
[0322] accgggcact gcggctggag gtgg 24
[0323] <210> 4
[0324] <211> 24
[0325] <212> DNA
[0326] <213> 人工序列(Artificial Sequence)
[0327] <400> 4
[0328] aaacccacct ccagccgcag tgcc 24
[0329] <210> 5
[0330] <211> 24
[0331] <212> DNA
[0332] <213> 人工序列(Artificial Sequence)
[0333] <400> 5
[0334] accggcggtc tcaagcacta ccta 24
[0335] <210> 6

[0336] <211> 24

[0337] <212> DNA

[0338] <213> 人工序列(Artificial Sequence)

[0339] <400> 6

[0340] aaactaggta gtgcttgaga ccgc 24

[0341] <210> 7

[0342] <211> 24

[0343] <212> DNA

[0344] <213> 人工序列(Artificial Sequence)

[0345] <400> 7

[0346] accggagctg cacatactag cccc 24

[0347] <210> 8

[0348] <211> 24

[0349] <212> DNA

[0350] <213> 人工序列(Artificial Sequence)

[0351] <400> 8

[0352] aaacggggct agtatgtgca gctc 24

[0353] <210> 9

[0354] <211> 24

[0355] <212> DNA

[0356] <213> 人工序列(Artificial Sequence)

[0357] <400> 9

[0358] accggcagac ggcagtcact aggg 24

[0359] <210> 10

[0360] <211> 24

[0361] <212> DNA

[0362] <213> 人工序列(Artificial Sequence)

[0363] <400> 10

[0364] aaacccctag tgactgccgt ctgc 24

[0365] <210> 11

[0366] <211> 24

[0367] <212> DNA

[0368] <213> 人工序列(Artificial Sequence)

[0369] <400> 11

[0370] accggtcgca ggacagcttt tcct 24

[0371] <210> 12

[0372] <211> 24

[0373] <212> DNA

[0374] <213> 人工序列(Artificial Sequence)

[0375] <400> 12

[0376] aaacaggaaa agctgtcctg cgac 24

[0377] <210> 13

[0378] <211> 24
[0379] <212> DNA
[0380] <213> 人工序列(Artificial Sequence)
[0381] <400> 13
[0382] accggggaag ctgggtgaat ggag 24
[0383] <210> 14
[0384] <211> 24
[0385] <212> DNA
[0386] <213> 人工序列(Artificial Sequence)
[0387] <400> 14
[0388] aaacctccat tcacccagct tccc 24
[0389] <210> 15
[0390] <211> 24
[0391] <212> DNA
[0392] <213> 人工序列(Artificial Sequence)
[0393] <400> 15
[0394] accggcggag actctggtgc tgtg 24
[0395] <210> 16
[0396] <211> 24
[0397] <212> DNA
[0398] <213> 人工序列(Artificial Sequence)
[0399] <400> 16
[0400] aaaccacagc accagagtct ccgc 24
[0401] <210> 17
[0402] <211> 24
[0403] <212> DNA
[0404] <213> 人工序列(Artificial Sequence)
[0405] <400> 17
[0406] accgggtcag aaataggggg tcca 24
[0407] <210> 18
[0408] <211> 24
[0409] <212> DNA
[0410] <213> 人工序列(Artificial Sequence)
[0411] <400> 18
[0412] aaactggacc ccctatttct gacc 24
[0413] <210> 19
[0414] <211> 24
[0415] <212> DNA
[0416] <213> 人工序列(Artificial Sequence)
[0417] <400> 19
[0418] accggatcca ggtgctgcag aagg 24
[0419] <210> 20

[0420] <211> 24

[0421] <212> DNA

[0422] <213> 人工序列(Artificial Sequence)

[0423] <400> 20

[0424] aaacccttct gcagcacctg gatc 24

[0425] <210> 21

[0426] <211> 24

[0427] <212> DNA

[0428] <213> 人工序列(Artificial Sequence)

[0429] <400> 21

[0430] accggtcact ccaggattcc aata 24

[0431] <210> 22

[0432] <211> 24

[0433] <212> DNA

[0434] <213> 人工序列(Artificial Sequence)

[0435] <400> 22

[0436] aaactattgg aatcctggag tgac 24

[0437] <210> 23

[0438] <211> 24

[0439] <212> DNA

[0440] <213> 人工序列(Artificial Sequence)

[0441] <400> 23

[0442] accgattacc catggcgtcc ccag 24

[0443] <210> 24

[0444] <211> 24

[0445] <212> DNA

[0446] <213> 人工序列(Artificial Sequence)

[0447] <400> 24

[0448] aaacctgggg acgccatggg taat 24

[0449] <210> 25

[0450] <211> 24

[0451] <212> DNA

[0452] <213> 人工序列(Artificial Sequence)

[0453] <400> 25

[0454] accggatcac gggggacaac cgga 24

[0455] <210> 26

[0456] <211> 24

[0457] <212> DNA

[0458] <213> 人工序列(Artificial Sequence)

[0459] <400> 26

[0460] aaactccggt tgtcccccgt gatc 24

[0461] <210> 27

[0462] <211> 24

[0463] <212> DNA

[0464] <213> 人工序列(Artificial Sequence)

[0465] <400> 27

[0466] accgatgggt aatggtgcca gtct 24

[0467] <210> 28

[0468] <211> 24

[0469] <212> DNA

[0470] <213> 人工序列(Artificial Sequence)

[0471] <400> 28

[0472] aaacagactg gcaccattac ccat 24

[0473] <210> 29

[0474] <211> 24

[0475] <212> DNA

[0476] <213> 人工序列(Artificial Sequence)

[0477] <400> 29

[0478] accggtcccc cgtgatcaga acca 24

[0479] <210> 30

[0480] <211> 24

[0481] <212> DNA

[0482] <213> 人工序列(Artificial Sequence)

[0483] <400> 30

[0484] aaactggttc tgatcacggg ggac 24

[0485] <210> 31

[0486] <211> 4927

[0487] <212> DNA

[0488] <213> 人工序列(Artificial Sequence)

[0489] <400> 31

[0490] ggtaccgatt agtgaacgga tctcgacggt atcgatcacg agactagcct cgagcggccg 60

[0491] ccccccttcac cgagggccta tttcccatga ttccttcata tttgcatata cgatacaagg 120

[0492] ctgttagaga gataattgga attaatttga ctgtaaacac aaagatatta gtacaaaata 180

[0493] cgtgacgtag aaagtaataa tttcttgggt agtttgcagt tttaaaatta tgttttaaaa 240

[0494] tggactatca tatgcttacc gtaacttgaa agtatttcga tttcttggct ttatatatct 300

[0495] tgtggaaagg acgaaacacc gtgagaccga gagagggtct cagttttaga gctagaaata 360

[0496] gcaagttaaa ataaggctag tccgttatca acttgaaaaa gtggcaccga gtcggtgctt 420

[0497] tttttaaaga attctcgacc tcgagacaaa tggcagtatt catccacaat tttaaaagaa 480

[0498] aaggggggat tgggggtac agtgcagggg aaagaatagt agacataata gcaacagaca 540

[0499] tacaaactaa agaattacaa aaacaaatta caaaaattca aaattttcgg gtttattaca 600

[0500] gggacagcag agatccactt tggccgcggc tcgaggggt tggggttgcg ccttttccaa 660

[0501] ggcagccctg ggtttgcgca gggacgcggc tgctctgggc gtggttccgg gaaacgcagc 720

[0502] ggcgccgacc ctgggactcg cacattcttc acgtccgttc gcagcgtcac ccggatcttc 780

[0503] gccgctaccc ttgtgggccc cccggcgacg cttcctgctc cgcccctaag tcgggaaggt 840

[0504] tccttgcggt tcgcggcgtg ccggacgtga caaacggaag ccgcacgtct cactagtacc 900

[0505] ctcgcagacg gacagcgcca gggagcaatg gcagcgcgcc gaccgcgatg ggctgtggcc 960

[0506] aatagcggct gctcagcagg gcgcgccgag agcagcggcc gggaaggggc ggtgcgggag 1020

[0507] gcggggtgtg gggcggtagt gtgggccctg ttcctgcccg cgcggtgttc cgcattctgc 1080

[0508] aagcctccgg agcgcacgtc ggcagtcggc tccctcgttg accgaatcac cgacctctct 1140

[0509] ccccaggggg atccatggtg agcaagggcg aggagctgtt caccggggtg gtgcccatcc 1200

[0510] tggtcgagct ggacggcgac gtaaacggcc acaagttcag cgtgtccggc gagggcgagg 1260

[0511] gcgatgccac ctacggcaag ctgaccctga agttcatctg caccaccggc aagctgcccg 1320

[0512] tgccctggcc caccctcgtg accaccctga cctacggcgt gcagtgcttc agccgctacc 1380

[0513] ccgaccacat gaagcagcac gacttcttca agtccgccat gcccgaaggc tacgtccagg 1440

[0514] agcgcaccat cttcttcaag gacgacggca actacaagac ccgcgccgag gtgaagttcg 1500

[0515] agggcgacac cctggtgaac cgcatcgagc tgaagggcat cgacttcaag gaggacggca 1560

[0516] acatcctggg gcacaagctg gagtacaact acaacagcca caacgtctat atcatggccg 1620

[0517] acaagcagaa gaacggcatc aaggtgaact tcaagatccg ccacaacatc gaggacggca 1680

[0518] gcgtgcagct cgccgaccac taccagcaga acacccccat cggcgacggc cccgtgctgc 1740

[0519] tgcccgacaa ccactacctg agcacccagt ccgccctgag caaagacccc aacgagaagc 1800

[0520] gcgatcacat ggtcctgctg gagttcgtga ccgccgccgg gatcactctc ggcatggacg 1860

[0521] agctgtacaa gtaaagcggc cgcgactcta gatcataatc agccatacca catttgtaga 1920

[0522] ggttttactt gctttaaaaa acctcccaca cctccccctg aacctgaaac ataaaatgaa 1980

[0523] tgcaattgtt gttgttaact tgtttattgc agcttataat ggttacaaat aaagcaatag 2040

[0524] catcacaaat ttcacaaata aagcattttt ttcactgcat tctagttgtg gtttgtccaa 2100

[0525] actcatcaat gtatcttagt cgaccgatgc ccttgagagc cttcaaccca gtcagctcct 2160

[0526] tccggtgggc gcggggcatg actatcgtcg ccgcacttat gactgtcttc tttatcatgc 2220

[0527] aactcgtagg acaggtgccg gcagcgctct tccgcttcct cgctcactga ctcgctgcgc 2280

[0528] tcggtcgttc ggctgcggcg agcggtatca gctcactcaa aggcggtaat acggttatcc 2340

[0529] acagaatcag gggataacgc aggaaagaac atgtgagcaa aaggccagca aaaggccagg 2400

[0530] aaccgtaaaa aggccgcgtt gctggcgttt ttccataggc tccgcccccc tgacgagcat 2460

[0531] cacaaaaatc gacgctcaag tcagaggtgg cgaaacccga caggactata aagataccag 2520

[0532] gcgtttcccc ctggaagctc cctcgtgcgc tctcctgttc cgaccctgcc gcttaccgga 2580

[0533] tacctgtccg cctttctccc ttcgggaagc gtggcgcttt ctcaatgctc acgctgtagg 2640

[0534] tatctcagtt cggtgtaggt cgttcgctcc aagctgggct gtgtgcacga acccccgtt 2700

[0535] cagcccgacc gctgcgcctt atccggtaac tatcgtcttg agtccaaccc ggtaagacac 2760

[0536] gacttatcgc cactggcagc agccactggt aacaggatta gcagagcgag gtatgtaggc 2820

[0537] ggtgctacag agttcttgaa gtggtggcct aactacggct acactagaag gacagtattt 2880

[0538] ggtatctgcg ctctgctgaa gccagttacc ttcggaaaaa gagttggtag ctcttgatcc 2940

[0539] ggcaaacaaa ccaccgctgg tagcggtggt ttttttgttt gcaagcagca gattacgcgc 3000

[0540] agaaaaaaag gatctcaaga agatcctttg atcttttcta cggggtctga cgctcagtgg 3060

[0541] aacgaaaact cacgttaagg gattttggtc atgagattat caaaaaggat cttcacctag 3120

[0542] atccttttaa attaaaaatg aagttttaaa tcaatctaaa gtatatatga gtaaacttgg 3180

[0543] tctgacagtt accaatgctt aatcagtgag gcacctatct cagcgatctg tctatttcgt 3240

[0544] tcatccatag ttgcctgact ccccgtcgtg tagataacta cgatacggga gggcttacca 3300

[0545] tctggcccca gtgctgcaat gataccgcgg gacccacgct caccggctcc agatttatca 3360

[0546] gcaataaacc agccagccgg aagggccgag cgcagaagtg gtcctgcaac tttatccgcc 3420
[0547] tccatccagt ctattaattg ttgccgggaa gctagagtaa gtagttcgcc agttaatagt 3480
[0548] ttgcgcaacg ttgttgccat tgctacaggc atcgtggtgt cacgctcgtc gtttggtatg 3540
[0549] gcttcattca gctccggttc ccaacgatca aggcgagtta catgatcccc catgttgtgc 3600
[0550] aaaaaagcgg ttagctcctt cggtcctccg atcgttgtca gaagtaagtt ggccgcagtg 3660
[0551] ttatcactca tggttatggc agcactgcat aattctctta ctgtcatgcc atccgtaaga 3720
[0552] tgcttttctg tgactggtga gtactcaacc aagtcattct gagaatagtg tatgcggcga 3780
[0553] ccgagttgct cttgcccggc gtcaatacgg gataataccg cgccacatag cagaacttta 3840
[0554] aaagtgctca tcattggaaa acgttcttcg gggcgaaaac tctcaaggat cttaccgctg 3900
[0555] ttgagatcca gttcgatgta acccactcgt gcacccaact gatcttcagc atcttttact 3960
[0556] ttcaccagcg tttctgggtg agcaaaaaca ggaaggcaaa atgccgcaaa aaagggaata 4020
[0557] agggcgacac ggaaatgttg aatactcata ctcttccttt ttcaatatta ttgaagcatt 4080
[0558] tatcagggtt attgtctcat gagcggatac atatttgaat gtatttagaa aaataaacaa 4140
[0559] atagggttc cgcgcacatt tccccgaaaa gtgccacctg acgcgccctg tagcggcgca 4200
[0560] ttaagcgcgg cgggtgtggt ggttacgcgc agcgtgaccg ctacacttgc cagcgcccta 4260
[0561] gcgcccgctc ctttcgcttt cttcccttcc tttctcgcca cgttcgccgg ctttccccgt 4320
[0562] caagctctaa atcgggggct ccctttaggg ttccgattta gtgctttacg gcacctcgac 4380
[0563] cccaaaaaac ttgattaggg tgatggttca cgtagtgggc catcgccctg atagacggtt 4440
[0564] tttcgccctt tgacgttgga gtccacgttc tttaatagtg gactcttgtt ccaaactgga 4500
[0565] acaacactca accctatctc ggtctattct tttgatttat aagggatttt gccgatttcg 4560
[0566] gcctattggt taaaaaatga gctgatttaa caaaaattta acgcgaattt taacaaaata 4620
[0567] ttaacgttta caatttccca ttcgccattc aggctgcgca actgttggga agggcgatcg 4680
[0568] gtgcgggcct cttcgctatt acgccagccc aagctaccat gataagtaag taatattaag 4740
[0569] gtacgggagg tacttggagc ggccgcaata aaatatcttt attttcatta catctgtgtg 4800
[0570] ttggtttttt gtgtgaatcg atagtactaa catacgctct ccatcaaaac aaaacgaaac 4860
[0571] aaaacaaact agcaaaatag gctgtcccca gtgcaagtgc aggtgccaga acatttctct 4920
[0572] atcgata 4927
[0573] <210> 32
[0574] <211> 167
[0575] <212> PRT
[0576] <213> 人工序列(Artificial Sequence)
[0577] <400> 32
[0578] Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
[0579] 1 5 10 15
[0580] Thr Leu Ala Lys Arg Ala Trp Asp Glu Arg Glu Val Pro Val Gly Ala
[0581] 20 25 30
[0582] Val Leu Val His Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Pro
[0583] 35 40 45
[0584] Ile Gly Arg His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
[0585] 50 55 60
[0586] Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
[0587] 65 70 75 80

[0588] Tyr Val Thr Leu Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
[0589] 85 90 95
[0590] Ser Arg Ile Gly Arg Val Val Phe Gly Ala Arg Asp Ala Lys Thr Gly
[0591] 100 105 110
[0592] Ala Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His
[0593] 115 120 125
[0594] Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
[0595] 130 135 140
[0596] Leu Ser Asp Phe Phe Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys
[0597] 145 150 155 160
[0598] Lys Ala Gln Ser Ser Thr Asp
[0599] 165
[0600] <210> 33
[0601] <211> 167
[0602] <212> PRT
[0603] <213> 人工序列(Artificial Sequence)
[0604] <400> 33
[0605] Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
[0606] 1 5 10 15
[0607] Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala
[0608] 20 25 30
[0609] Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala
[0610] 35 40 45
[0611] Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
[0612] 50 55 60
[0613] Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
[0614] 65 70 75 80
[0615] Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
[0616] 85 90 95
[0617] Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly
[0618] 100 105 110
[0619] Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His
[0620] 115 120 125
[0621] Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
[0622] 130 135 140
[0623] Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys
[0624] 145 150 155 160
[0625] Lys Ala Gln Ser Ser Thr Asp
[0626] 165
[0627] <210> 34
[0628] <211> 1367
[0629] <212> PRT

[0630] <213> 人工序列(Artificial Sequence)

[0631] <400> 34

[0632] Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly

[0633] 1 5 10 15

[0634] Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys

[0635] 20 25 30

[0636] Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly

[0637] 35 40 45

[0638] Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys

[0639] 50 55 60

[0640] Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr

[0641] 65 70 75 80

[0642] Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe

[0643] 85 90 95

[0644] Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His

[0645] 100 105 110

[0646] Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His

[0647] 115 120 125

[0648] Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser

[0649] 130 135 140

[0650] Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met

[0651] 145 150 155 160

[0652] Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp

[0653] 165 170 175

[0654] Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn

[0655] 180 185 190

[0656] Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys

[0657] 195 200 205

[0658] Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu

[0659] 210 215 220

[0660] Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu

[0661] 225 230 235 240

[0662] Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp

[0663] 245 250 255

[0664] Leu Ala Glu Asp Thr Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp

[0665] 260 265 270

[0666] Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu

[0667] 275 280 285

[0668] Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile

[0669] 290 295 300

[0670] Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met

[0671] 305 310 315 320

[0672] Ile Lys Leu Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
[0673] 325 330 335
[0674] Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
[0675] 340 345 350
[0676] Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
[0677] 355 360 365
[0678] Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
[0679] 370 375 380
[0680] Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
[0681] 385 390 395 400
[0682] Gln Arg Thr Phe Asp Asn Gly Ile Ile Pro His Gln Ile His Leu Gly
[0683] 405 410 415
[0684] Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
[0685] 420 425 430
[0686] Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
[0687] 435 440 445
[0688] Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met
[0689] 450 455 460
[0690] Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Lys Val
[0691] 465 470 475 480
[0692] Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
[0693] 485 490 495
[0694] Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
[0695] 500 505 510
[0696] Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
[0697] 515 520 525
[0698] Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Asp Gln Lys
[0699] 530 535 540
[0700] Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
[0701] 545 550 555 560
[0702] Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
[0703] 565 570 575
[0704] Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
[0705] 580 585 590
[0706] Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
[0707] 595 600 605
[0708] Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
[0709] 610 615 620
[0710] Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
[0711] 625 630 635 640
[0712] Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
[0713] 645 650 655

[0714] Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
[0715] 660 665 670
[0716] Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
[0717] 675 680 685
[0718] Asn Arg Asn Phe Ile Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys
[0719] 690 695 700
[0720] Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
[0721] 705 710 715 720
[0722] Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
[0723] 725 730 735
[0724] Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
[0725] 740 745 750
[0726] His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
[0727] 755 760 765
[0728] Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
[0729] 770 775 780
[0730] Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
[0731] 785 790 795 800
[0732] Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
[0733] 805 810 815
[0734] Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
[0735] 820 825 830
[0736] Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp
[0737] 835 840 845
[0738] Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
[0739] 850 855 860
[0740] Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
[0741] 865 870 875 880
[0742] Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
[0743] 885 890 895
[0744] Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
[0745] 900 905 910
[0746] Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
[0747] 915 920 925
[0748] His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
[0749] 930 935 940
[0750] Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
[0751] 945 950 955 960
[0752] Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
[0753] 965 970 975
[0754] Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
[0755] 980 985 990

[0756] Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
[0757] 995 1000 1005
[0758] Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser
[0759] 1010 1015 1020
[0760] Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn
[0761] 1025 1030 1035 1040
[0762] Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile
[0763] 1045 1050 1055
[0764] Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val
[0765] 1060 1065 1070
[0766] Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met
[0767] 1075 1080 1085
[0768] Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe
[0769] 1090 1095 1100
[0770] Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala
[0771] 1105 1110 1115 1120
[0772] Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro
[0773] 1125 1130 1135
[0774] Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys
[0775] 1140 1145 1150
[0776] Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met
[0777] 1155 1160 1165
[0778] Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys
[0779] 1170 1175 1180
[0780] Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr
[0781] 1185 1190 1195 1200
[0782] Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala
[0783] 1205 1210 1215
[0784] Gly Val Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
[0785] 1220 1225 1230
[0786] Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
[0787] 1235 1240 1245
[0788] Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr
[0789] 1250 1255 1260
[0790] Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile
[0791] 1265 1270 1275 1280
[0792] Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His
[0793] 1285 1290 1295
[0794] Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe
[0795] 1300 1305 1310
[0796] Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr
[0797] 1315 1320 1325

[0798] Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala
[0799] 1330 1335 1340
[0800] Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp
[0801] 1345 1350 1355 1360
[0802] Leu Ser Gln Leu Gly Gly Asp
[0803] 1365
[0804] <210> 35
[0805] <211> 18
[0806] <212> PRT
[0807] <213> 人工序列(Artificial Sequence)
[0808] <400> 35
[0809] Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys Arg
[0810] 1 5 10 15
[0811] Lys Val
[0812] <210> 36
[0813] <211> 7
[0814] <212> PRT
[0815] <213> 人工序列(Artificial Sequence)
[0816] <400> 36
[0817] Pro Lys Lys Lys Arg Lys Val
[0818] 1 5
[0819] <210> 37
[0820] <211> 32
[0821] <212> PRT
[0822] <213> 人工序列(Artificial Sequence)
[0823] <400> 37
[0824] Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr
[0825] 1 5 10 15
[0826] Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser
[0827] 20 25 30
[0828] <210> 38
[0829] <211> 4
[0830] <212> PRT
[0831] <213> 人工序列(Artificial Sequence)
[0832] <400> 38
[0833] Ser Gly Gly Ser
[0834] 1
[0835] <210> 39
[0836] <211> 1799
[0837] <212> PRT
[0838] <213> 人工序列(Artificial Sequence)
[0839] <400> 39

[0840] Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys
[0841] 1 5 10 15
[0842] Arg Lys Val Pro Arg Ala Ala Thr Met Ser Glu Val Glu Phe Ser His
[0843] 20 25 30
[0844] Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Trp Asp
[0845] 35 40 45
[0846] Glu Arg Glu Val Pro Val Gly Ala Val Leu Val His Asn Asn Arg Val
[0847] 50 55 60
[0848] Ile Gly Glu Gly Trp Asn Arg Pro Ile Gly Arg His Asp Pro Thr Ala
[0849] 65 70 75 80
[0850] His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn
[0851] 85 90 95
[0852] Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Leu Glu Pro Cys Val
[0853] 100 105 110
[0854] Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe
[0855] 115 120 125
[0856] Gly Ala Arg Asp Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val
[0857] 130 135 140
[0858] Leu His His Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile
[0859] 145 150 155 160
[0860] Leu Ala Asp Glu Cys Ala Ala Leu Leu Ser Asp Phe Phe Arg Met Arg
[0861] 165 170 175
[0862] Arg Gln Glu Ile Lys Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Ser
[0863] 180 185 190
[0864] Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser
[0865] 195 200 205
[0866] Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser
[0867] 210 215 220
[0868] Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu
[0869] 225 230 235 240
[0870] Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu
[0871] 245 250 255
[0872] Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly
[0873] 260 265 270
[0874] Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly
[0875] 275 280 285
[0876] Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val
[0877] 290 295 300
[0878] Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg
[0879] 305 310 315 320
[0880] Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala
[0881] 325 330 335

[0882] Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val
[0883] 340 345 350
[0884] Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys
[0885] 355 360 365
[0886] Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys Lys Ala
[0887] 370 375 380
[0888] Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser
[0889] 385 390 395 400
[0890] Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly
[0891] 405 410 415
[0892] Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly
[0893] 420 425 430
[0894] Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro
[0895] 435 440 445
[0896] Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys
[0897] 450 455 460
[0898] Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu
[0899] 465 470 475 480
[0900] Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys
[0901] 485 490 495
[0902] Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys
[0903] 500 505 510
[0904] Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu
[0905] 515 520 525
[0906] Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp
[0907] 530 535 540
[0908] Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys
[0909] 545 550 555 560
[0910] Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu
[0911] 565 570 575
[0912] Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly
[0913] 580 585 590
[0914] Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu
[0915] 595 600 605
[0916] Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser
[0917] 610 615 620
[0918] Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg
[0919] 625 630 635 640
[0920] Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly
[0921] 645 650 655
[0922] Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe
[0923] 660 665 670

[0924] Lys Ser Asn Phe Asp Leu Ala Glu Asp Thr Lys Leu Gln Leu Ser Lys
[0925] 675 680 685
[0926] Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp
[0927] 690 695 700
[0928] Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile
[0929] 705 710 715 720
[0930] Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro
[0931] 725 730 735
[0932] Leu Ser Ala Ser Met Ile Lys Leu Tyr Asp Glu His His Gln Asp Leu
[0933] 740 745 750
[0934] Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys
[0935] 755 760 765
[0936] Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp
[0937] 770 775 780
[0938] Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu
[0939] 785 790 795 800
[0940] Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu
[0941] 805 810 815
[0942] Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ile Ile Pro His
[0943] 820 825 830
[0944] Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp
[0945] 835 840 845
[0946] Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu
[0947] 850 855 860
[0948] Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser
[0949] 865 870 875 880
[0950] Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp
[0951] 885 890 895
[0952] Asn Phe Glu Lys Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile
[0953] 900 905 910
[0954] Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu
[0955] 915 920 925
[0956] Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu
[0957] 930 935 940
[0958] Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu
[0959] 945 950 955 960
[0960] Ser Gly Asp Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn
[0961] 965 970 975
[0962] Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile
[0963] 980 985 990
[0964] Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn
[0965] 995 1000 1005

[0966] Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys
[0967] 1010 1015 1020
[0968] Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val
[0969] 1025 1030 1035 1040
[0970] Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu
[0971] 1045 1050 1055
[0972] Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys
[0973] 1060 1065 1070
[0974] Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn
[0975] 1075 1080 1085
[0976] Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys
[0977] 1090 1095 1100
[0978] Ser Asp Gly Phe Ala Asn Arg Asn Phe Ile Gln Leu Ile His Asp Asp
[0979] 1105 1110 1115 1120
[0980] Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln
[0981] 1125 1130 1135
[0982] Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala
[0983] 1140 1145 1150
[0984] Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val
[0985] 1155 1160 1165
[0986] Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala
[0987] 1170 1175 1180
[0988] Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg
[0989] 1185 1190 1195 1200
[0990] Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu
[0991] 1205 1210 1215
[0992] Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr
[0993] 1220 1225 1230
[0994] Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu
[0995] 1235 1240 1245
[0996] Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln
[0997] 1250 1255 1260
[0998] Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser
[0999] 1265 1270 1275 1280
[1000] Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val
[1001] 1285 1290 1295
[1002] Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile
[1003] 1300 1305 1310
[1004] Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu
[1005] 1315 1320 1325
[1006] Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr
[1007] 1330 1335 1340

[1008] Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn
[1009] 1345 1350 1355 1360
[1010] Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile
[1011] 1365 1370 1375
[1012] Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe
[1013] 1380 1385 1390
[1014] Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr
[1015] 1395 1400 1405
[1016] Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu
[1017] 1410 1415 1420
[1018] Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys
[1019] 1425 1430 1435 1440
[1020] Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr
[1021] 1445 1450 1455
[1022] Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu
[1023] 1460 1465 1470
[1024] Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
[1025] 1475 1480 1485
[1026] Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
[1027] 1490 1495 1500
[1028] Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val
[1029] 1505 1510 1515 1520
[1030] Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser
[1031] 1525 1530 1535
[1032] Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly
[1033] 1540 1545 1550
[1034] Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys
[1035] 1555 1560 1565
[1036] Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu
[1037] 1570 1575 1580
[1038] Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp
[1039] 1585 1590 1595 1600
[1040] Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile
[1041] 1605 1610 1615
[1042] Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg
[1043] 1620 1625 1630
[1044] Met Leu Ala Ser Ala Gly Val Leu Gln Lys Gly Asn Glu Leu Ala Leu
[1045] 1635 1640 1645
[1046] Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys
[1047] 1650 1655 1660
[1048] Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu
[1049] 1665 1670 1675 1680

[1050] Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe
[1051] 1685 1690 1695
[1052] Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser
[1053] 1700 1705 1710
[1054] Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
[1055] 1715 1720 1725
[1056] Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
[1057] 1730 1735 1740
[1058] Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys
[1059] 1745 1750 1755 1760
[1060] Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr
[1061] 1765 1770 1775
[1062] Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Ser Gly Gly Ser
[1063] 1780 1785 1790
[1064] Pro Lys Lys Lys Arg Lys Val
[1065] 1795

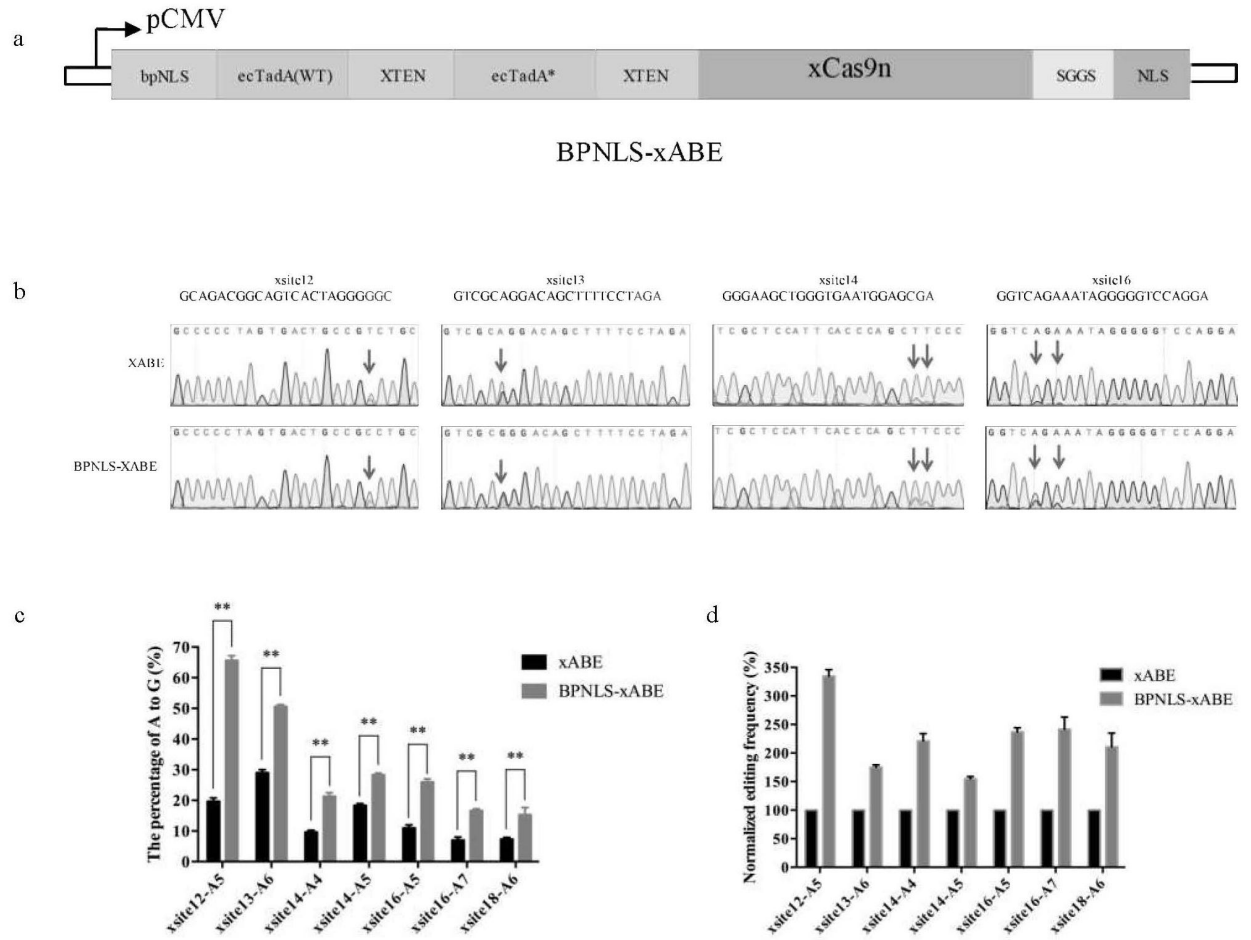
a
pCMV
bpNLS
ecTadA(WT)
XTEN
ecTadA*
XTEN
xCas9n
SGGS
NLS
BPNLS-xABE
b
xsite12
GCAGACGGCAGTCACTAGGGGGC
xsite13
GTCGCAGGACAGCTTTTCCTAGA
xsite14
GGGAAGCTGGGTGAATGGAGCGA
xsite16
GGTCAGAAATAGGGGGTCCAGGA
XABE
BPNLS-XABE
c
The percentage of A to G (%)
xABE
BPNLS-xABE
xsite12-A5
xsite13-A6
xsite14-A4
xsite14-A5
xsite16-A5
xsite16-A7
xsite18-A6
d
Normalized editing frequency (%)
xABE
BPNLS-xABE
xsite12-A5
xsite13-A6
xsite14-A4
xsite14-A5
xsite16-A5
xsite16-A7
xsite18-A6

图1

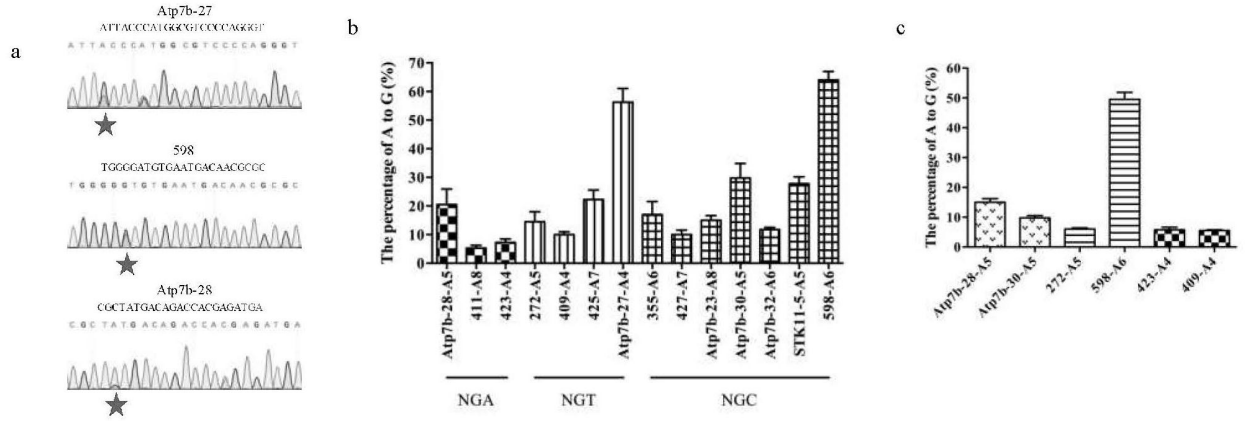
a
Atp7b-27
ATTACCCATGGCGTCCCCAGGGT
598
TGGGGATGTGAATGACAACGCGC
Atp7b-28
CGCTATGACAGACCACGAGATGA
b
The percentage of A to G (%)
Atp7b-28-A5
411-A8
423-A4
272-A5
409-A4
425-A7
Atp7b-27-A4
355-A6
427-A7
Atp7b-23-A8
Atp7b-30-A5
Atp7b-32-A6
STK11-5-A5
598-A6
NGA
NGT
NGC
c
The percentage of A to G (%)
Atp7b-28-A5
Atp7b-30-A5
272-A5
598-A6
423-A4
409-A4

图2

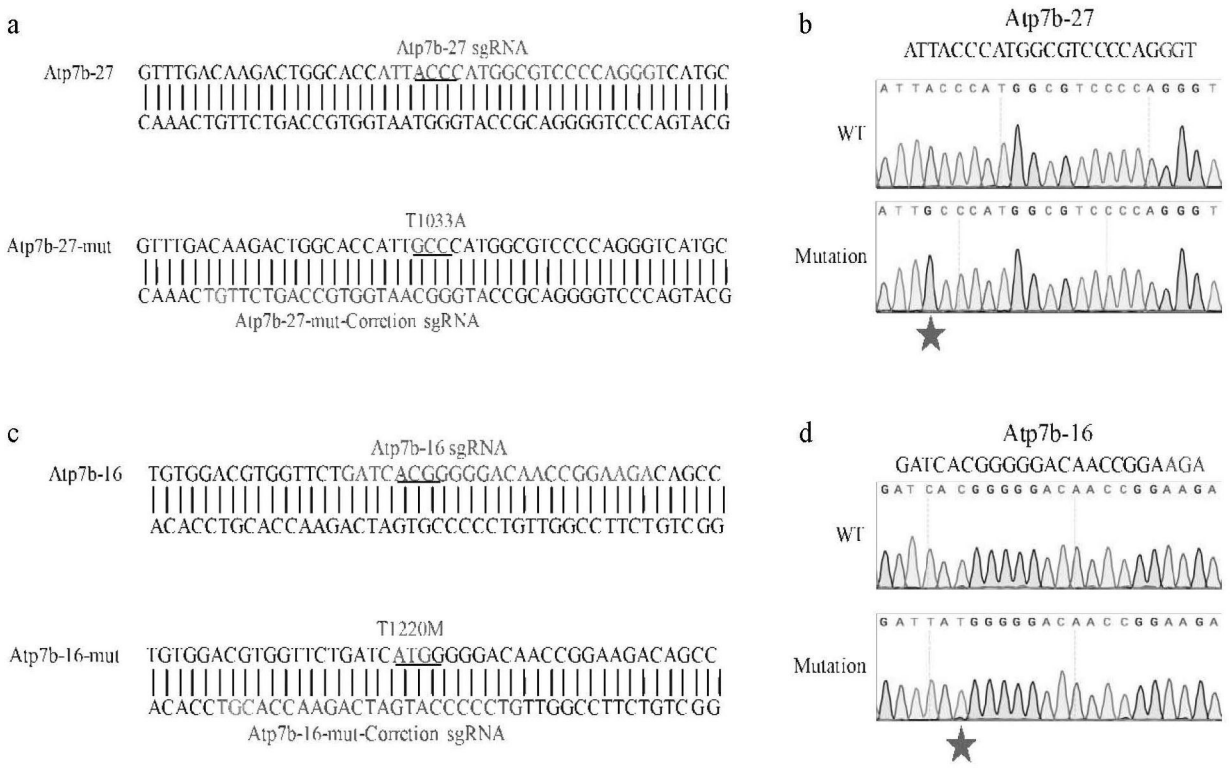
a
Atp7b-27 sgRNA
Atp7b-27
GTTTGACAAGACTGGCACCATTACCCATGGCGTCCCCAGGGTCATGC
CAAACTGTTCTGACCGTGGTAATGGGTACCGCAGGGGTCCCAGTACG
T1033A
Atp7b-27-mut
GTTTGACAAGACTGGCACCATTGCCCATGGCGTCCCCAGGGTCATGC
CAAACTGTTCTGACCGTGGTAACGGGTACCGCAGGGGTCCCAGTACG
Atp7b-27-mut-Correction sgRNA
b
Atp7b-27
ATTACCCATGGCGTCCCCAGGGT
WT
Mutation
c
Atp7b-16 sgRNA
Atp7b-16
TGTGGACGTGGTTCTGATCACGGGGGACAACCGGAAGACAGCC
ACACCTGCACCAAGACTAGTGCCCCCTGTTGGCCTTCTGTCGG
T1220M
Atp7b-16-mut
TGTGGACGTGGTTCTGATCATGGGGGACAACCGGAAGACAGCC
ACACCTGCACCAAGACTAGTACCCCCTGTTGGCCTTCTGTCGG
Atp7b-16-mut-Correction sgRNA
d
Atp7b-16
GATCACGGGGGACAACCGGAAGA
WT
Mutation

图3

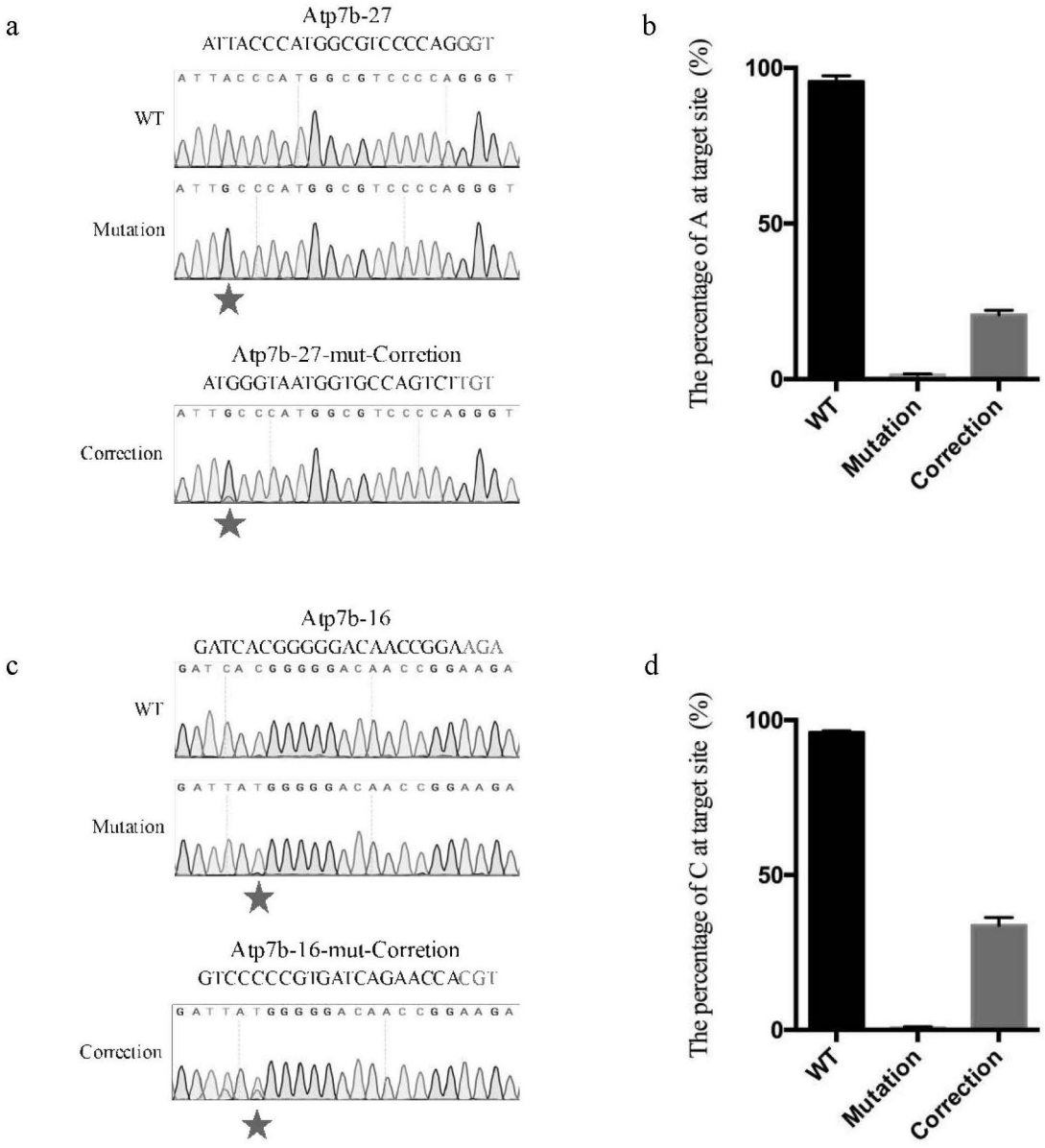
a
Atp7b-27
ATTACCCATGGCGTCCCCAGGGT
WT
Mutation
Atp7b-27-mut-Corretion
ATGGGTAATGGTGCCAGTCTTGT
Correction
b
The percentage of A at target site (%)
100
50
0
WT
Mutation
Correction
c
Atp7b-16
GATCACGGGGGACAACCGGAAGA
WT
Mutation
Atp7b-16-mut-Corretion
GTCCCCCGTGATCAGAACCACGT
Correction
d
The percentage of C at target site (%)
100
50
0
WT
Mutation
Correction

图4