



(12) 发明专利

(10) 授权公告号 CN 101802183 B

(45) 授权公告日 2016.02.24

(21) 申请号 200880106647.5

(22) 申请日 2008.07.14

(30) 优先权数据

60/959,203 2007.07.12 US

(85) PCT国际申请进入国家阶段日

2010.03.11

(86) PCT国际申请的申请数据

PCT/US2008/069997 2008.07.14

(87) PCT国际申请的公布数据

W02009/009797 EN 2009.01.15

(73) 专利权人 新英格兰生物实验室公司

地址 美国马萨诸塞洲

(72) 发明人 朱振宇 A·布兰查德 徐双勇

管胜昔 魏华 张鹏华 孙大鹏

陈少康

(74) 专利代理机构 北京纪凯知识产权代理有限公司

公司 11245

代理人 赵蓉民 杜艳玲

(51) Int. Cl.

C12N 9/22(2006.01)

(56) 对比文件

WO 2007027464 A2, 2007.03.08,

US 2004/0175816 A1, 2004.09.09,

CN 1668740 A, 2005.09.14,

Conlan L.H. 等. Modulating restriction endonuclease activities and specificities using neutral detergents. 《Biotechniques》. 1999, 第27卷(第5期), 955-960.

Mukhopadhyay Partha 等. Protein engineering of BamHI restriction endonuclease: Replacement of Cys54 by Ala enhances catalytic activity. 《Protein Engineering》. 1998, 第11卷(第10期), 931-935.

Xu S.Y. 等. Cofactor Requirements of BamH I Mutant Endonuclease E77K and Its Suppressor Mutants. 《Journal of Bacteriology》. 1991, 第173卷(第16期), 5030-5035.

审查员 吕小蒙

权利要求书1页 说明书56页

序列表81页 附图50页

(54) 发明名称

高保真度限制性内切核酸酶

(57) 摘要

提供针对具有改变性质的酶的组合物和方法,其包括诱变的系统方法和允许选择希望的蛋白质的筛查分析。该方法的实施方式特别适于改变限制性内切核酸酶的特性如星号活性。该组合物包括具有减少星号活性的限制性内切核酸酶,如通过总保真度指数改进因子限定的。

1. 具有人工引入突变的 EcoRI 限制性内切核酸酶, 其中所述人工引入突变对应于 SEQ ID No. 83 中的 K62A ;K62S ;K62L ;R9A ;K15A ;R123A ;K130A ;R131A ;R183A ;S2Y ;D135A ;R187A ;或 K62E。
2. DNA 分子, 其编码权利要求 1 所述的限制性内切核酸酶。
3. 载体, 其包含权利要求 2 所述的 DNA。
4. 宿主细胞, 其含有用于表达限制性内切核酸酶的权利要求 2 所述的 DNA 或权利要求 3 所述的载体。

高保真度限制性内切核酸酶

背景技术

[0001] 限制性内切核酸酶是以序列特异性方式切割双链 DNA 的酶 (Roberts, R. J., Proc Natl Acad Sci U S A, 102:5905-5908(2005); Roberts, et al., Nucleic Acids Res, 31:1805-1812(2003); Roberts, et al., Nucleic Acids Res, 33:D230-232(2005); Alves, et al., Restriction Endonucleases, "Protein Engineering of Restriction Enzymes," ed. Pingoud, Springer-Verlag Berlin Heidelberg, New York, 393-407(2004))。它们在原核生物中是普遍存在的 (Raleigh, et al., Bacterial Genomes Physical Structure and Analysis, Ch. 8, eds. De Bruijn, et al., Chapman & Hall, New York, 78-92(1998)), 其中它们形成限制修饰系统的一部分, 其主要由内切核酸酶和甲基转移酶构成。同源的 (cognate) 甲基转移酶甲基化与其配对内切核酸酶识别的序列相同的特异性序列, 并赋予被修饰 DNA 对内切核酸酶切割的抗性, 以便宿主 DNA 可以被恰当地保护。然而, 当存在外源 DNA 侵入时, 特别是噬菌体 DNA 侵入时, 外源 DNA 将在其可被完全甲基化之前降解。限制修饰系统的主要生物功能是保护宿主免于噬菌体感染 (Arber, Science, 205:361-365(1979))。其他的功能也已被提出, 例如涉及重组和转座 (Carlson, et al., Mol Microbiol, 27:671-676(1998); Heitman, Genet Eng (N Y), 15:57-108(1993); McKane, et al., Genetics, 139:35-43(1995))。

[0002] 大约 3,000 种已知的限制性内切核酸酶对它们的超过 250 种不同靶序列的特异性可以被考虑为它们最令人感兴趣的特性。在发现第一种限制性内切核酸酶的序列特异性性质 (Danna, et al., Proc Natl Acad Sci USA, 68:2913-2917(1971); Kelly, et al., J Mol Biol, 51:393-409(1970)) 之后, 科学家们没有花费太长时间发现了某些限制性内切核酸酶在非最适条件下切割与其限定识别序列相似但不相同的序列 (Polisky, et al., Proc Natl Acad Sci USA, 72:3310-3314(1975); Nasri, et al., Nucleic Acids Res, 14:811-821(1986))。该松弛的特异性 (relaxed specificity) 被称为限制性内切核酸酶的星号活性 (star activity)。已提出, 限制性内切核酸酶和 DNA 之间水介导的相互作用是特异性复合物和星复合物 (star complexes) 之间的主要差异 (Robinson, et al., J Mol Biol, 234:302-306(1993); Robinson, et al., Proc Natl Acad Sci USA, 92:3444-3448(1995); Sidorova, et al., Biophys J, 87:2564-2576(2004))。

[0003] 星号活性是分子生物学反应中的问题。星号活性在克隆载体或其他 DNA 中引入不希望的切割。在例如法医应用的情况中, 在这种情况下某些 DNA 底物需要被限制性内切核酸酶切割以产生独特的指纹图谱, 星号活性将改变切割图案特性, 从而使分析复杂化。避免星号活性在例如链置换扩增技术 (Walker, et al., Proc Natl Acad Sci USA, 89:392-396(1992)) 和基因表达的系列分析 (Velculescu, et al., Science, 270:484-487(1995)) 的应用中也是关键的。

[0004] 发明概述

[0005] 在本发明的一个实施方式中, 提供组合物, 所述组合物包含具有至少一个人工引入突变和至少 2 的总保真度指数 (FI) 改进因子 (overall fidelity index improvement

factor) 的限制性内切核酸酶,在预定的缓冲液中所述限制性内切核酸酶能以与不存在所述人工引入突变的限制性内切核酸酶的切割活性至少相似的切割活性切割底物,所述人工引入突变是靶向突变、饱和诱变或通过 PCR 扩增过程引入的突变的至少一个的产物。

[0006] 在本发明的进一步的实施方式中,人工引入突变中的至少一个是靶向突变,其是由用带相反电荷的残基置换天然存在的残基产生的。丙氨酸或苯丙氨酸可以置换靶位点处的天然存在的残基。

[0007] 在本发明的进一步的实施方式中,上面描述类型的组合物包括选自下列的不存在人工引入突变的限制酶 :BamHI、EcoRI、ScaI、SalI、SphI、PstI、NcoI、NheI、SspI、NotI、SacI、PvuII、MfeI、HindIII、SbfI、EagI、EcoRV、AvrII、BstXI、PciI、HpaI、AgeI、BsmBI、BspQI、SapI、KpnI 和 BsaI。

[0008] 本发明进一步的实施方式包括在表 4 中列出的组分。

[0009] 在本发明的进一步的实施方式中,提供编码在表 4 中列出的酶的任一种的 DNA,包含该 DNA 的载体和从该载体表达蛋白质的宿主细胞。

[0010] 在本发明的一个实施方式中,提供具有下列步骤的方法:(a) 鉴定在具有星号活性的限制性内切核酸酶的氨基酸序列中,哪些氨基酸残基是带电荷的氨基酸;(b) 在编码该限制性内切核酸酶的基因序列中,突变编码一个或多个带电荷的残基的一个或多个密码子;(c) 产生在不同的带电荷的残基具有一个或多个不同密码子突变的基因序列文库;(d) 获得突变的基因序列表达的蛋白质组(a set of proteins);和(e) 测定每一表达蛋白质在预定缓冲液中的 FI 和切割活性。

[0011] 该方法的一个实施方式包括在限定的缓冲液组中对属于该蛋白质组的蛋白质测定总 FI 改进因子的步骤,其中例如,缓冲液组包含 NEB1、NEB2、NEB3 和 NEB4 缓冲液。

[0012] 该方法的一个实施方式包括上面描述的步骤和另外地,在突变带电荷的氨基酸的密码子的同一步骤或随后步骤中,突变编码羟化氨基酸或酰胺氨基酸的密码子。

[0013] 在上面描述的本发明的一个实施方式中,将除了酪氨酸以外的密码子突变为丙氨酸,将酪氨酸突变为苯丙氨酸。

[0014] 在一个进一步的实施方式中,使用对一个或多个突变密码子的饱和诱变,改进总 FI 改进因子。

[0015] 附图简述

[0016] 对于图 1-32:

[0017] * 标记表示其左侧的泳道含有最低浓度的观察到星号活性的酶。

[0018] # 标记指示出不完全切割的泳道,该泳道与含有足够完全切割底物的浓度的酶的泳道相邻并位于其右侧。

[0019] 灰色三角形表示限制性内切核酸酶浓度的连续减少(serial decrease)。

[0020] “U”表示酶单位。

[0021] 在图 1-32 中描述的每一反应中,反应混合物含有 3 μ l 体积的——除非另外规定——来自 New England Biolabs, Inc. (NEB), Ipswich, MA 的缓冲液(参见表 1 和 NEB 目录), 3 μ l 体积的——除非另外规定——来自 NEB, Ipswich, MA 的稀释剂中的指定限制性内切核酸酶(参见表 1 和 NEB 目录),以及不定体积的指定底物(含有 0.6 μ g)和使反应混合物达到总共 30 μ l 的体积的水。在 37°C 下进行反应,温育时间为 1 小时。在 0.8% 琼脂糖

凝胶上分析结果。当反应混合物总体积、底物量、反应温度或温育时间与上述不同时,在图的说明中提供数值。

[0022] 在图 1、5、8、11-18 和 20-32 中,在凝胶的右侧提供理论消化图谱。那些仅具有一个限制性内切核酸酶位点的底物应该从超螺旋形式消化为一个线性带。

[0023] 图 1 显示通过在 NEB3 缓冲液中使用稀释剂 A 对 WT ScaI (1, 200U) 制剂进行两倍连续稀释而消化 1.2 μ l λ DNA 底物 (0.6 μ g)、并在琼脂糖凝胶上检查消化产物,对野生型 (WT) ScaI 的 FI 进行的测定。没有星号活性的限制性内切核酸酶的最高浓度用实心箭头显示;和引起底物完全消化的最小浓度用空心箭头示出。

[0024] 图 2A-D 示出用 WT BamHI 或 BamHI (E86P) 酶消化 0.5 μ l pUC19 底物 (0.5 μ g) 1 小时的结果,所述 WT BamHI 或 BamHI (E86P) 酶是用稀释剂 A 三倍连续稀释的,其起始浓度为 172U 或 512U。中间的泳道是 NEB 1kb 标记 (NewEngland Biolabs, Inc. (NEB), Ipswich, MA)。

[0025] 图 2A 示出使用 NEB1 缓冲液的结果。

[0026] 图 2B 示出使用 NEB2 缓冲液的结果。

[0027] 图 2C 示出使用 NEB3 缓冲液的结果。

[0028] 图 2D 示出使用 NEB4 缓冲液的结果。

[0029] 图 3A-B 示出在两个时间期间 BamHI (E86P) 活性的比较,其在 NEB2 缓冲液使用 0.6 μ l pBR322 底物 (其仅包含 1 个 BamHI 切割位点),使用初始浓度 600U、用稀释剂 A 以 2- 倍连续稀释的酶。

[0030] 图 3A 示出 1 小时的结果。

[0031] 图 3B 示出 14 小时的结果。

[0032] 图 4A-B 示出在琼脂糖凝胶上,在两种不同的缓冲液中温育 14 小时后,使用以稀释剂 A 2 倍连续稀释的 BamHI-HF (E163A/E167T),对 0.6 μ l pBR322 底物的切割。

[0033] 图 4A 示出使用 NEB2 缓冲液和 600U 初始浓度酶的结果。

[0034] 图 4B 示出使用 NEB1 缓冲液和 2,400U 初始浓度酶的结果。

[0035] 图 5A-B 示出在 NEB4 缓冲液中应用 BamHI-HF 和 WT BamHI 的切割反应的比较。该反应在 NEB4 缓冲液中进行,使用 1.2 μ l λ DNA 底物,处于使用稀释剂 A 的 2- 倍连续稀释中。

[0036] 图 5A 示出起始浓度 1,200U 的 WT BamHI,其中 FI 等于 4。

[0037] 图 5B 示出起始浓度 2,400U 的 BamHI-HF,其中 FI \geq 4000。

[0038] 图 6A-D 示出在使用 NEB 稀释剂 C 3- 倍连续稀释下,在 NEB1-4 缓冲液中 WTEcoRI 和 EcoRI (K62A) 的比较。反应混合物包含在 NEB1-4 缓冲液中的 2 μ l λ DNA 底物 (1 μ g)。

[0039] 图 6A 示出两倍连续稀释后,在 NEB2 缓冲液中 120U WT EcoRI 和 240U 的 EcoRI (K62A) 的切割结果。

[0040] 图 6B 示出两倍连续稀释后,在 NEB4 缓冲液中 120U WT EcoRI 和 240U 的 EcoRI (K62A) 的切割结果。

[0041] 图 6C 示出两倍连续稀释后,在 NEB1 缓冲液中 60U WT EcoRI 和 120U 的 EcoRI (K62A) 的切割结果。

[0042] 图 6D 示出两倍连续稀释后,在 NEB3 缓冲液中 120U WT EcoRI 和 60U 的

EcoRI (K62A) 的切割结果。

[0043] 图 7A-C 示出用两种不同的 EcoRI 突变体和 WT EcoRI 的切割结果。示出 100,000U 酶和其在稀释剂 C 中的 10 倍连续稀释物在 10 小时中的消化,其在各种缓冲液中使用 0.6 μ l 的 Litmus28 底物进行。在 Litmus28 底物中仅存在一个 EcoRI 切割位点。

[0044] 图 7A :在 NEB4 缓冲液中的 EcoRI 突变体 K62E。

[0045] 图 7B :在 NEB4 缓冲液中的 EcoRI 突变体 K62A。

[0046] 图 7C :在 EcoRI 缓冲液中的 WT EcoRI (参见 NEB 目录 2007-8)。

[0047] 图 8A-B 示出在 NEB4 缓冲液中 EcoRI-HF 和 WT EcoRI 的比较。反应使用 1.2 μ l λ DNA 底物,其处于使用稀释剂 C 的 2- 倍连续稀释下。

[0048] 图 8A :在 NEB4 缓冲液中,具有 19,200U 起始浓度的 WT EcoRI 显示 FI = 4。

[0049] 图 8B :在 NEB4 缓冲液中,具有 38,400 起始浓度的 UEcoRI-HF 显示 FI = 16,000。

[0050] 图 9A-B 示出 WT ScaI (4.8U)、ScaI (H193A) (9.6U)、ScaI (S201F) (19.2U) 和 ScaI (H193A/S201F) (19.2U) 的连续稀释的比较。每一样品最初稀释 1/10,然后在具有指定百分比甘油的 NEB2 缓冲液中 2- 倍连续稀释。每一反应混合物包含 2 μ l 的 λ DNA 底物 (1 μ g)。

[0051] 图 9A :5%甘油。

[0052] 图 9B :37%甘油。

[0053] 图 10A-D 示出 WT ScaI 和 ScaI-HF (H193A/S201F) 的比较。酶 (如规定的单位浓度) 各自用稀释剂 A 以 2.5- 倍连续稀释。该反应混合物含有 2 μ l λ DNA 底物和 NEB1-4 缓冲液。

[0054] 图 10A :在 NEB2 缓冲液中切割。

[0055] 图 10B :在 NEB4 缓冲液中切割。

[0056] 图 10C :在 NEB1 缓冲液中切割。

[0057] 图 10D :在 NEB3 缓冲液中切割。

[0058] 图 11A-H 示出对 SalI-HF 和 WT SalI 的 FI 测定。两种酶都使用稀释剂 A 以 2- 倍连续稀释。该反应混合物含有 2 μ l HindIII- 消化的 λ DNA 底物。

[0059] 图 11A、B、C 和 D 示出 1,200U、1,200U、300U 和 1,200U 的 SalI-HF 的连续稀释,其分别在 NEB1、2、3 和 4 缓冲液中显示 FI \geq 1,000、FI \geq 2,000、FI \geq 500 和 FI \geq 2,000。

[0060] 图 11E、F、G 和 H 示出 19.2U、150U、9,600U 和 38.4U 的 WT SalI 的连续稀释,其分别在 NEB1、2、3 和 4 缓冲液中显示 FI = 8、FI = 1、FI = 32 和 FI = 1。

[0061] 图 12A-B 示出在 NEB 4 缓冲液中与 1.2 μ l λ DNA 底物反应的、在稀释剂 A 中的 SphI-HF (19,200U) 和在稀释剂 B 中的 WT SphI (143,600U) 的 2- 倍连续稀释的结果。

[0062] 图 12A 示出 WT SphI 的切割。

[0063] 图 12B 示出 SphI-HF 的切割。

[0064] 图 13A-B 示出分别在 NEB3 和 NEB4 缓冲液中,使用 2- 倍连续稀释的 PstI-HF (300U 和 150U) 和 2- 倍连续稀释的 WT PstI (2,400U 和 4,800U) 对 1.2 μ l λ DNA 底物的切割。连续稀释在稀释剂 C 中进行。

[0065] 图 13A 示出 NEB4 缓冲液 (上图) 和 NEB3 缓冲液 (下图) 中 PstI-HF 的切割。

[0066] 图 13B 示出 NEB4 缓冲液 (上图) 和 NEB3 缓冲液 (下图) 中 WT PstI 的切割。

[0067] 图 14A-B 示出分别在 NEB3 和 NEB4 缓冲液中, 使用 2- 倍连续稀释的 NcoI-HF(4, 800U 和 600U) 和 2- 倍连续稀释的 WT NcoI(4, 800U 和 1, 200U) 对 1.2 μ l λ DNA 底物的切割。连续稀释在稀释剂 A 中进行。

[0068] 图 14A 示出 NEB4 缓冲液 (上图) 和 NEB3 缓冲液 (下图) 中 NcoI-HF 的切割。

[0069] 图 14B 示出 NEB4 缓冲液 (上图) 和 NEB3 缓冲液 (下图) 中 WT NcoI 的切割。

[0070] 图 15A-B 示出分别在 NEB3 和 NEB4 缓冲液中, 使用 2- 倍连续稀释的 NheI-HF(287, 200U 和 76.8U) 和 2- 倍连续稀释的 WT NheI(9, 600U 和 300U) 对 1.2 μ l pXbaDNA 底物的切割。连续稀释在稀释剂 A 中进行。

[0071] 图 15A 示出 NEB4 缓冲液 (上图) 和 NEB3 缓冲液 (下图) 中 NheI-HF 的切割。

[0072] 图 15B 示出 NEB4 缓冲液 (上图) 和 NEB3 缓冲液 (下图) 中 WT NheI 的切割。

[0073] 图 16A-B 示出分别在 NEB3 和 NEB4 缓冲液中, 使用 2- 倍连续稀释的 SspI-HF(9, 600U 和 38.4U) 和 2- 倍连续稀释的 WT SspI(19, 200U 和 19, 200U) 对 1.2 μ l λ DNA 底物的切割。连续稀释在稀释剂 C 中进行。

[0074] 图 16A 示出 NEB4 缓冲液 (上图) 和 NEB3 缓冲液 (下图) 中 SspI-HF 的切割。

[0075] 图 16B 示出 NEB4 缓冲液 (上图) 和 NEB3 缓冲液 (下图) 中 WT SspI 的切割。

[0076] 图 17A-B 示出分别在 NEB3 和 NEB4 缓冲液中, 使用 2- 倍连续稀释的 NotI-HF(287, 200U 和 19, 200U) 和 2- 倍连续稀释的 WT NotI(19, 200U 和 76, 800U) 对 1.2 μ l pXba DNA 底物的切割。连续稀释在稀释剂 C 中进行。

[0077] 图 17A 示出 NEB4 缓冲液 (上图) 和 NEB3 缓冲液 (下图) 中 NotI-HF 的切割。

[0078] 图 17B 示出 NEB4 缓冲液 (上图) 和 NEB3 缓冲液 (下图) 中 WT NotI 的切割。

[0079] 图 18A-B 示出分别在 NEB3 和 NEB4 缓冲液中, 使用 2- 倍连续稀释的 SacI-HF(4, 800U 和 76.8U) 和 2- 倍连续稀释的 WT SacI(19, 200U 和 1200U) 对 1.2 μ l pXbaDNA 底物的切割。连续稀释在稀释剂 A 中进行。

[0080] 图 18A 示出 NEB4 缓冲液 (上图) 和 NEB3 缓冲液 (下图) 中 SacI-HF 的切割。

[0081] 图 18B 示出 NEB4 缓冲液 (上图) 和 NEB3 缓冲液 (下图) 中 WT SacI 的切割。

[0082] 图 19A-B 示出分别在 NEB3 和 NEB4 缓冲液中, 使用 2- 倍连续稀释的 PvuII-HF(9, 600U 和 19.2U) 和 2- 倍连续稀释的 WT PvuII(19, 200U 和 300U) 对 0.6 μ l pBR322DNA 底物的切割。连续稀释在稀释剂 A 中进行。

[0083] 图 19A 示出 NEB4 缓冲液 (上图) 和 NEB3 缓冲液 (下图) 中 PvuII-HF 的切割。

[0084] 图 19B 示出 NEB4 缓冲液 (上图) 和 NEB3 缓冲液 (下图) 中 WT PvuII 的切割。

[0085] 图 20A-B 示出分别在 NEB3 和 NEB4 缓冲液中, 使用 2- 倍连续稀释的 MfeI-HF(300U 和 19.2U) 和 2- 倍连续稀释的 WT MfeI(2, 400U 和 38.4U) 对 1.2 μ l λ DNA 底物的切割。连续稀释在稀释剂 A 中进行。

[0086] 图 20A 示出 NEB4 缓冲液 (上图) 和 NEB3 缓冲液 (下图) 中 MfeI-HF 的切割。

[0087] 图 20B 示出 NEB4 缓冲液 (上图) 和 NEB3 缓冲液 (下图) 中 WT MfeI 的切割。

[0088] 图 21A-B 示出分别在 NEB3 和 NEB4 缓冲液中, 使用 2- 倍连续稀释的 HindIII-HF(4, 800U 和 1, 200U) 和 2- 倍连续稀释的 WT HindIII(9, 600U 和 4, 800U) 对 1.2 μ l λ DNA 底物的切割。连续稀释在稀释剂 A 中进行。

[0089] 图 21A 示出 NEB4 缓冲液 (上图) 和 NEB3 缓冲液 (下图) 中 HindIII-HF 的切割。

- [0090] 图 21B 示出 NEB4 缓冲液 (上图) 和 NEB3 缓冲液 (下图) 中 WT HindIII 的切割。
- [0091] 图 22A-B 示出在 NEB4 缓冲液中, 使用稀释剂 C 以 2- 倍连续稀释 SbfI-HF (起始浓度 :76, 800U) 和 WT SbfI (起始浓度 :76, 800U), 对 1.2 μ l λ DNA 底物的切割。
- [0092] 图 22A 示出通过 WT SbfI 的切割。
- [0093] 图 22B 示出通过 SbfI-HF 的切割。
- [0094] 图 23A-B 示出分别在 NEB2 和 NEB1 缓冲液中, 使用 2- 倍连续稀释的 EagI-HF (1, 200U 和 600U) 和 2- 倍连续稀释的 WT EagI (150U 和 38.4U) 对 1.2 μ l pXba DNA 底物的切割, 其应用稀释剂 C。
- [0095] 图 23A 示出 NEB2 缓冲液 (上图) 和 NEB1 缓冲液 (下图) 中 EagI-HF 的切割。
- [0096] 图 23B 示出 NEB2 缓冲液 (上图) 和 NEB1 缓冲液 (下图) 中 WT EagI 的切割。
- [0097] 图 24A-B 示出在 NEB4 缓冲液中, 使用稀释剂 A 以 2 倍连续稀释 EcoRV-HF (起始浓度 :38, 400U) 和 WT EcoRV (起始浓度 :2400U), 对 1.2 μ l pXba DNA 底物的切割。
- [0098] 图 24A 示出通过 WT EcoRV 的切割。
- [0099] 图 24B 示出通过 EcoRV-HF 的切割。
- [0100] 图 25A-B 示出在 NEB4 缓冲液中, 使用稀释剂 A 以 2 倍连续稀释 AvrII-HF (起始浓度 :1, 200U) 和 WT AvrII (起始浓度 :1, 200U), 对 1.2 μ l T7 DNA 底物的切割。
- [0101] 图 25A 示出通过 WT AvrII 的切割。
- [0102] 图 25B 示出通过 AvrII-HF 的切割。
- [0103] 图 26A-B 示出在 NEB4 缓冲液中, 在稀释剂 A 中 2 倍连续稀释 BstXI-HF (起始浓度 :300U) 和 WT BstXI (起始浓度 :38.4U), 对 1.2 μ l λ DNA 底物的切割。该反应在 55 $^{\circ}$ C 进行。
- [0104] 图 26A 示出通过 WT BstXI 的切割。
- [0105] 图 26B 示出通过 BstXI-HF 的切割。
- [0106] 图 27A-B 示出在 NEB4 缓冲液中, 使用稀释剂 A 以 2 倍连续稀释 PciI-HF (起始浓度 :600U) 和 WT PciI (起始浓度 :300U), 对 1.2 μ l pXba DNA 底物的切割。
- [0107] 图 27A 示出通过 WT PciI 的切割。
- [0108] 图 27B 示出通过 PciI-HF 的切割。
- [0109] 图 28A-B 示出在 NEB2 缓冲液中, 使用稀释剂 A 以 2 倍连续稀释 HpaI-HF (起始浓度 :4, 800U) 和 WT HpaI (起始浓度 :9, 600U), 对 1.2 μ l λ DNA 底物的切割。
- [0110] 图 28A 示出通过 WT HpaI 的切割。
- [0111] 图 28B 示出通过 HpaI-HF 的切割。
- [0112] 图 29A-B 示出在 NEB4 缓冲液中, 使用稀释剂 C 以 2- 倍连续稀释 AgeI-HF (起始浓度 :600U) 和 WT AgeI (起始浓度 600U), 对 1.2 μ l pXba DNA 底物的切割。
- [0113] 图 29A 示出通过 WT AgeI 的切割。
- [0114] 图 29B 示出通过 AgeI-HF 的切割。
- [0115] 图 30A-B 示出在 NEB4 缓冲液中, 使用稀释剂 C 以 2- 倍连续稀释 BsmBI-HF (起始浓度 :300U) 和 WT BsmBI (起始浓度 4, 800U), 对 1.2 μ l λ DNA 底物的切割。该反应在 55 $^{\circ}$ C 下进行。
- [0116] 图 30A 示出通过 WT BsmBI 的切割。
- [0117] 图 30B 示出通过 BsmBI-HF 的切割。

[0118] 图 31A-B 分别示出 MluCI 的 DNA 序列 (SEQ ID NO :1) 和蛋白序列 (SEQ IDNO :2), 用于 EcoRI 和 MfeI 的表达。

[0119] 图 32 示出 Hpy166II 的 DNA 序列 (SEQ ID NO :3), 用于 SalI 的表达。

[0120] 图 33A-B 分别示出 MfeI 的 DNA 序列 (SEQ ID NO :4) 和蛋白序列 (SEQ IDNO :5)。

[0121] 图 34A-B 分别示出 BstXI 的 DNA 序列 (SEQ ID NO :6) 和蛋白序列 (SEQ IDNO :7)。

[0122] 图 35A-B 分别示出 M. BstXI 的 DNA 序列 (SEQ ID NO :8) 和蛋白序列 (SEQ IDNO :9)。

[0123] 图 36A-B 分别示出 S. BstXI 的 DNA 序列 (SEQ ID NO :10) 和蛋白序列 (SEQ IDNO :11)。

[0124] 图 37A-B 分别示出 PciI 的 DNA 序列 (SEQ ID NO :12) 和蛋白序列 (SEQ IDNO :13)。

[0125] 图 38A-B 分别示出 M. PciI 的 DNA 序列 (SEQ ID NO :14) 和蛋白序列 (SEQ IDNO :15)。

[0126] 图 39 示出编码 M1. EarI 的 DNA 序列 (SEQ ID NO :17), 所述 M1. EarI 识别 CTCTTC 并且在 N4 胞嘧啶或 N6 腺嘌呤处甲基化, 用于克隆 SapI 和 BspQI。

[0127] 图 40 示出编码 M2. EarI 的 DNA 序列 (SEQ ID NO :18), 所述 M2. EarI 识别 CTCTTC 并且在 N4 胞嘧啶或 N6 腺嘌呤处甲基化, 用于克隆 SapI 和 BspQI。

[0128] 图 41A-B 示出测定 1 μ l WT BspQI 和 1 μ l 突变体 (K279P/R388F) BspQI (起始浓度分别为 512U 和 1,024U) 的 FI 的琼脂糖凝胶, 其通过使用稀释剂 A 和 NEB1 缓冲液加 10% 甘油以 2 倍连续稀释来切割 1 μ l pUC19DNA 底物进行。该反应在 50°C 进行。

[0129] 图 42 示出测定 5 μ l WT SapI 和 5 μ l 突变体 (K273A) SapI (起始浓度分别为 32U 和 16U) 的 FI 的琼脂糖凝胶, 其通过使用稀释剂 A 和 NEB2 缓冲液加 25% DMSO 以 2 倍连续稀释来切割 pUC19DNA 底物进行。

[0130] 图 43A-B 示出通过 5 μ l WT KpnI 和 5 μ l D16N/E132A/D148E KpnI (初始浓度分别为 32U 和 256U) 得到的 pXba 的催化活性和星号活性。使用含有 10mM pH 7.4 的 Tris-HCl、50mM KCl、0.1mM EDTA、1mM DTT 和 50% 甘油的稀释剂在 NEB2 缓冲液中, 以 2- 倍连续稀释的该酶消化 2 μ l pXba DNA 底物 (0.5 μ g), 其中使用水补充至 50 μ l 的总体积。

[0131] 图 45A 示出 WT KpnI 的切割结果。

[0132] 图 45B 示出 D16N/E132A/D148E KpnI 的切割结果。

[0133] 图 44A-D 示出表 1 中的酶的氨基酸序列, 和以前没有公开的表 1 中酶的 DNA 序列。

[0134] 发明详述

[0135] 本发明的实施方式提供选择具有希望特性的限制性内切核酸酶的一般方法。该一般方法依赖于适当的分析测定希望的限制性内切核酸酶是否已经产生。特别地, 该一般方法的一个实施方式提供具有一组步骤的系统筛查方法。通过使用多种限制性内切核酸酶实施数百个反应, 已经推导出该方法。本文提供的实例的大多数涉及鉴定具有降低星号活性但是其切割活性至少与 WT 限制性内切核酸酶相似的限制性内切核酸酶。然而, 期望可以成功地运用同一方法以改变 (修饰) 限制性内切核酸酶的其他性质, 例如涉及在期望缓冲液中的改进的切割活性、热稳定性、在限定条件中的反应速率等。

[0136] 如上所述, 所关心的终点是将具有星号活性的限制性内切核酸酶转化为具有显著减少星号活性的高保真度限制性内切核酸酶。星号活性指单个限制性内切核酸酶在切割特

异性方面的混乱。术语“星号活性降低”和“保真度增加”在本文可互换使用。尽管限制性内切核酸酶特征为其在特定序列处切割 DNA 的性质,但是一些限制性内切核酸酶另外地在该 DNA 的第二(次级)位点处无效率地切割 DNA。该种次级切割可以一贯地发生或者可以仅仅在某些条件例如下列条件之一下出现:浓度增加、某些缓冲液、温度、底物类型、存储和温育时间。

[0137] 通常承认对由构成蛋白质和决定特异性的数以百计的氨基酸产生的复杂环境所知甚少。现有技术的一个方法已经利用晶体学来鉴定酶和其底物之间的接触点。但是,晶体学对于在非自然化学环境中及时冷冻结构方面具有局限性。

[0138] 使用现有的分析技术,已经证明确定蛋白质中任何位点处氨基酸的贡献和底物分子结构所起的作用的规则是难以捉摸的。例如,这里显示,突变限制性内切核酸酶中的一个氨基酸可以引起所有或部分的活性损失。

[0139] 在本文中,没有提出结构解释以说明为什么星号活性可以随着高甘油浓度(> 5% v/v)、高的酶与 DNA 比例(通常每 μg DNA > 100 单位的酶)、低离子强度(< 25mM 的盐)、高 pH(> 8.0)、存在有机溶剂(例如 DMSO、乙醇)和用其他的二价阳离子(Mn^{2+} 、 Co^{2+})取代 Mg^{2+} 而增加。这里承认因为影响星号活性的因素的多样性,将必需在相同反应条件和在相同的预定缓冲液中对 WT 和突变星号活性进行比较,并且开发任何高保真度酶必定能示出所述特性的标准反应条件——即使这些特性也在其他的反应条件下观察到。

[0140] 本发明的实施方式涉及产生具有特定提高性质的修饰的限制性内切核酸酶,即具有增强的切割保真度而没有总切割活性的显著减少或从制造该蛋白质的宿主细胞的显著产量下降。本文已经发展用于发现具有提高性质的突变体的方法是详尽试验的结果,并且生成酶的性质已经在指定条件的背景中限定。本文描述的方法可用来在预定条件下改变任何限制性内切核酸酶的酶促性质,但是不限于所述具体的限定条件。

[0141]

| 限制性内切核酸酶 | 用于产生高保真度限制性内切核酸酶的步骤 |
|------------------|---|
| BamHI (实施例 1) | 同切点酶的比较。 靶向 22 个残基以突变为 Ala。获得 14 个突变体，3 个具有提高的保真度。 对 2 个残基-K30 和 E86 饱和诱变。 回收 E86P 为在选择的缓冲液中具有最大降低的星号活性的优选突变体。添加突变至 E86P。 第二轮突变(Arg、Lys、His、Asp、Glu、Ser、Thr)为 Ala，和 Tyr 突变为 Phe。选择 E167 和 Y165 用于饱和诱变和选择 E167T 和 Y165F。 E163A/E167T 被选为优选的高保真度突变体(BamHI-HF)。 |
| EcoRI (实施例 2) | 同切点酶的比较。 靶向 42 个带电荷的残基以突变为 Ala。没有高保真度突变体。 第二轮突变：靶向另外的 32 个带电荷的残基以突变为 Ala：鉴定 K62A。 对 K62A 饱和诱变。EcoRI(K62E)被选为优选的高保真度突变体(EcoRI-HF)。 |
| ScaI (实施例 3) | 同切点酶的比较。 靶向 58 个带电荷的残基以突变为 Ala。鉴定 4 个突变体。 4 个突变体中优选的突变体是(H193A/S201F)。其被选为优选的高保真度突变体(ScaI-HF)。 |
| SalI (实施例 4) | 靶向 86 个带电荷的残基并突变为 Ala。SalI (R107A)被优先选为优选的高保真度突变体(SalI-HF)。 |
| SphI (实施例 5) | 靶向 71 个带电荷的残基并突变为 Ala。SphI (K100A)被优先选为优选的高保真度突变体(SphI-HF)。 |
| PstI (实施例 6) | 靶向 92 个带电荷的氨基酸并突变为 Ala。PstI (D91A)被优先选为优选的高保真度突变体(PstI-HF)。 |
| NcoI (实施例 7) | 靶向 66 个带电荷的残基并突变为 Ala。NcoI (A2T/R31A)被优先选为优选的高保真度突变体(NcoI-HF)。 |
| NheI (实施例 8) | 靶向 92 个带电荷的残基并突变为 Ala。NheI (E77A)被优先选为优选的高保真度突变体(NheI-HF)。 |

[0142]

| | |
|----------------------|--|
| 限制性内切核酸酶 | 用于产生高保真度限制性内切核酸酶的步骤 |
| SspI (实施例 9) | 靶向 81 个带电荷的残基并突变为 Ala。没有获得优选的突变体。 靶向 95 个另外的带电荷的残基和羟化残基为 Ala,除了 Tyr 以外。Tyr 被突变为 Phe。SspI(Y98F) 被优先选为优选的高保真度突变体 (SspI-HF)。 |
| NotI (实施例 10) | 靶向 97 个带电荷的残基并突变为 Ala。K150A 被优先选为优选的高保真度突变体 (NotI-HF)。 |
| SacI (实施例 11) | 靶向 101 个带电荷的残基并突变为 Ala。SacI(Q117H/R200A) 被优先选为优选的高保真度突变体 (SacI-HF),其中 Q117H 为从模板的遗留突变 (carry over mutation),其没有影响活性。 |
| PvuII (实施例 12) | 靶向 47 个带电荷的残基并突变为 Ala。没有获得优选的突变体。 靶向 19 个羟化残基 -Ser/Thr 和 Tyr。选择 T46A 用于进一步改进。 饱和诱变产生优选的突变体 T46G、T46H、T46K、T46Y。 PvuII(T46G) 被优先选为优选的高保真度突变体 (PvuII-HF)。 |
| MfeI (实施例 13) | 靶向 60 个带电荷的残基并突变为 Ala。没有获得优选的突变体。 靶向 26 个羟化残基并突变为 Ala,除了 Tyr 以外,Tyr 被改为 Phe。 靶向 38 个残基 (Cys、Phe、Met、Asn、Gln、Trp) 并突变为 Ala。鉴定 Mfe(Q13A/F35Y) 为优选的高保真度突变体 (MfeI-HF),其中 F35Y 携带自模板。 |
| HindIII (实施例 14) | 靶向 88 个带电荷的残基并突变为 Ala。没有获得优选的突变体 靶向 103 个残基 (Cys、Met、Asn、Gln、Ser、Thr、Trp) 并突变为 Ala 和 Tyr 被改为 Phe。 鉴定 HindIII(K198A) 为优选的高保真度突变体 (HindIII-HF)。 |
| SbfI (实施例 15) | 靶向 78 个带电荷的残基,将其突变为 Ala。 靶向 41 个残基 (Ser Thr),将其突变为 Ala/Tyr 突变为 Phe。 靶向 Cys、Phe、Met、Asn、Gln、Trp 的 55 个残基突变为 Ala SbfI(K251A) 被选为优选的高保真度突变体 (SbfI-HF)。 |
| EagI (实施例 16) | 靶向 152 个残基 (Asp、Glu、His、Lys、Arg、Ser、thr、Asn 和 Gln 改变为 Ala,和 Tyr 改变为 Phe)。 EagI H43A 被选为优选的高保真度突变体 (EagI-HF)。 |

| | |
|--------------------|---|
| EcoRV (实施例 17) | 靶向 162 个残基 (Cys、Asp、Glu、Phe、his、Lys、Met、Asn、Gln、Arg、Ser、Thr 改变为 Ala 和 Trp 改变为 Phe)。 EcoRV (D19A/E27A) 被选为优选的高保真度突变体 (EcoRV-HF) |
|--------------------|---|

[0143]

| | |
|--------------------|---|
| 限制性内切核酸酶 | 用于产生高保真度限制性内切核酸酶的步骤 |
| AvrII (实施例 18) | 靶向 210 个残基 (Cys、Asp、Glu、Phe、his、Lys、Met、Asn、Gln、Arg、Ser、Thr 改变为 Ala 和 Trp 改变为 Phe)。 AvrII (Y104F) 被选为优选的高保真度突变体 (AvrII-HF)。 |
| BstXI (实施例 19) | 靶向 237 个残基 (Cys、Asp、Glu、Phe、his、Lys、Met、Asn、Gln、Arg、Ser、Thr 改变为 Ala 和 Trp 改变为 Phe)。 BstXI (N65A) 被选为优选的高保真度突变体 (BstXI-HF)。 |
| PciI (实施例 20) | 靶向 151 个残基 (Cys、Asp、Glu、Phe、his、Lys、Met、Asn、Gln、Arg、Ser、Thr 改变为 Ala 和 Trp 改变为 Phe)。 PciI (E78A/S133A) 被选为优选的高保真度突变体。(PciI-HF) 这是自发的, 并且不是该 151 个单独突变的一个。 |
| HpaI (实施例 21) | 靶向 156 个残基 (Cys、Asp、Glu、Phe、his、Lys、Met、Asn、Gln、Arg、Ser、Thr 改变为 Ala 和 Trp 改变为 Phe)。 HpaI (E56A) 被选为优选的高保真度突变体 (HpaI-HF)。 |
| AgeI (实施例 22) | 靶向 149 个残基 (Cys、Asp、Glu、Phe、his、Lys、Met、Asn、Gln、Arg、Ser、Thr 改变为 Ala 和 Trp 改变为 Phe)。 AgeI (R139A) 被选为优选的高保真度突变体 (AgeI-HF)。 |
| BsmBI (实施例 23) | 靶向 358 个残基 (Cys、Asp、Glu、Phe、his、Lys、Met、Asn、Gln、Arg、Ser、Thr 改变为 Ala 和 Trp 改变为 Phe)。 BsmBI (N185Y/R232A) 被选为优选的高保真度突变体 (BsmBI (HF))。 |
| BspQI (实施例 24) | 靶向 122 个残基 (Arg、Lys、His、Glu、Asp、Gln、Asn、Cys)。 在位置 279 用 Phe、Pro、Tyr、Glu、Asp 或 Leu 取代 R。 优选的突变是 R388F 和 K279P。 产生双突变体 BspQI (K279P/R388F) 为优选的高保真度突变体 (BspQI-HF)。 |
| SapI (实施例 25) | 找出相应于 BspQI 中的 R388 和 K279 的 SapI 中的 K273 和 R380。 SapI (K273P/R380F) 被选为优选的高保真度突变体 (SapI-HF)。 |

| | |
|------------------|--|
| KpnI (实施例 26) | 靶向所有残基 (Asp、Glu、Arg、Lys、His、Ser、Thr、Tyr、Asn、Gln、Phe、Trp、Cys、Met) 为 Ala。 在位点 D16 和 D148 进行更多突变。 组合的 D16N/E132A/D148E 被选为优选的高保真度突变体 (KpnI-HF)。 |
| BsaI (实施例 27) | 找到与 BsmBI 中位点相应的 11 个氨基酸。 BsaI (Y231F) 被选为优选的高保真度突变体 (BsaI-HF)。 |

[0144] 该方法从负责同源活性和星号活性的氨基酸是不同的这一认识而得出。本文描述的高保真度限制性内切核酸酶的工程化显示同源活性和星号活性可以分开,并且存在影响这些不同活性的不同的关键氨基酸残基。被发现影响星号活性的氨基酸的位置不一定在蛋白质的活性部位内发现。通过发展在确定 FI (也参见 Wei et al. *Nucleic Acid Res.*, 36, 9, e50 (2008)) 和总保真度指数改进因子的形式方面成功标准, 本文已经第一次确定任何限制性内切核酸酶的切割性质。

[0145] “总保真度指数改进因子 (overall fidelity index improvement factor)”指在选择缓冲液组中具有最大切割活性的突变体的最高 FI 除以具有最大切割活性的相应 WT 内切核酸酶的最高 FI。选择的组可以是大于 1 的任何大小, 但是实际上将含有小于 10 种不同的缓冲液, 并且更优选地含有 4 种缓冲液。该组也可包括小于 4 种缓冲液。另外地而不仅仅对于由 NEB1、NEB2、NEB3 和 NEB4 组成的缓冲液组, 至少 2 的总 FI 改进因子应优选地适用于要求保护的发明的任何突变限制性内切核酸酶。

[0146] 通过将相同量的酶与相同量和类型的底物在相同条件下反应, 并且视觉上比较电泳后凝胶上的切割图, 以便切割产物的量表现在误差的标准容限内是相同的并且其中定量相似性大于 10%, 可以测量“相似的切割活性”。

[0147] “人工的”指“人造的”。

[0148] “标准条件”指从在 NEB1-4 缓冲液中获得的结果计算的总 FI 改进因子。

[0149] 本文描述的一般方法已经用 27 种限制性内切核酸酶示例: AgeI、AvaI、BamHI、BsaI、BsmBI、BspQI、BstXI、EagI、EcoRI、EcoRV、HindIII、HpaI、KpnI、MfeI、NcoI、NheI、NotI、PciI、PstI、PvuII、SacI、SalI、SapI、SbfI、ScaI、SphI 和 SspI 限制性内切核酸酶。然而, 如上所述, 预期该方法对工程化具有显著星号活性的任何限制性内切核酸酶是有效的。

[0150] 该方法的实施方式利用一般的方法产生具有减少星号活性的突变体限制性内切核酸酶。对于某些酶, 已经证明突变被确定在两个同切点酶之间保守的带电荷的残基是有用的 (参见例如实施例 25 中的 SapI)。然而, 一般而言, 该方法包括鉴定内切核酸酶的蛋白序列内所有的带电荷和极性残基的第一步。例如, 带电荷的氨基酸和极性残基包括酸性残基 Glu 和 Asp; 碱性残基 His、Lys 和 Arg; 酰胺残基 Asn 和 Gln; 芳族残基 Phe、Tyr 和 Trp; 和亲核残基 Cys。各残基被靶向并突变为 Ala, 并且针对希望的保真度增加的性质, 筛查这些靶向突变的产物。如果获得的突变体没有一个提供令人满意的结果, 下一步是靶向突变所有的羟化氨基酸, 即 Ser、Thr 和 Tyr, 优选的突变是 Ser 和 Thr 突变为 Ala 和 Tyr 突变为 Phe。也可能同时靶向突变两类残基, 如在实施例 16-23 中所做的。突变为 Ala 可以被

突变为 Val、Leu 或 Ile 所取代。

[0151] 在这些分析之后,如果上述步骤中产生的一种或更多种优选的突变体在选择的检验下仍然具有低于标准的性能,这些突变体可以被选择并再次突变为另外可能的 18 种氨基酸的每一种。这被称为饱和诱变。饱和诱变提供 EcoRI (实施例 2)、部分 BamHI (实施例 1) 和 PvuII (实施例 12) 的优选的高保真度突变体。依赖饱和诱变的结果,下一步将以靶向或随机或者这两种方式将另外的突变引入限制性内切核酸酶。在实施例 11 中,SacI-HF 包括在反向 PCR 期间偶然产生的随机突变。在实施例 20 中,PciI-HF 是随机突变的结果,而不是靶向突变的结果。在实施例 26 中,BspQI-HF 含有被发现在增强保真度中协同作用两个突变。

[0152] 不同的靶向诱变方法例如反向 PCR 的应用可包括在蛋白质的第二位点引入非靶向突变。这些第二突变可偶然地提供需要的性质(参见实施例 20)。检查那些具有多突变的突变酶是希望的,以确定是否所有的突变对观测的效果是必需的。在实施例 11 中,双突变体中 Q117H 对活性没有影响。在实施例 20 中,另外的自发突变表现独自为观察到的提高的保真度的原因,而在实施例 24 中,各个突变协同地作用。

[0153] 在一些情况中,突变可提供除了提高的保真度之外的另外的优点(参见例如 BamHI,其中 E163A 或 P173A 引起酶变得更热不稳定)。

[0154] 根据其在一组标准缓冲液中的功能,选择实施例中提供的突变体的高保真度/减少星号活性的性质。如果选择不同的缓冲液组合物,那么其他的突变可以是优选的。然而,可以运用相同的方法来发现突变体。表 4 列出运用于每一限制性内切核酸酶并在标准缓冲液中提供总 FI 改进因子的突变。

[0155] 提供至少 2 的总 FI 改进因子的高保真度限制性内切核酸酶的工程化包括一个或更多个下列步骤:

[0156] 1. WT 限制性内切核酸酶的星号活性的评估

[0157] 在本发明的一个实施方式中,限制性内切核酸酶的星号活性的程度通过下列方法检测:对于适当的底物,使用高初始浓度的原液(stock)内切核酸酶和其连续稀释物(例如 2 倍或 3 倍稀释),测定内切核酸酶活性。限制性内切核酸酶的初始浓度是不重要的,只要其足以使在至少一个浓度中观察到星号活性,这样通过稀释,不再检测到星号活性。

[0158] 适当的底物含有被同源内切核酸酶活性切割的核苷酸序列,并且在其中可以观察到星号活性。该底物可以是含有限制性内切核酸酶基因的载体或第二 DNA 底物。在表 2 中使用的底物的实例是 pBC4、pXba、T7、lambda 和 pBR322。

[0159] 原液限制性内切核酸酶的浓度最初被选择,以便星号活性可以容易地在 WT 和突变限制性内切核酸酶中识别和分析。根据 2007-08NEB 目录的指导,选择适当的稀释缓冲液例如 NEB 稀释剂 A、B 或 C 进行连续稀释。连续稀释的限制性内切核酸酶与预定浓度的适当的底物以反应容器大小确定的总反应体积进行反应。例如,在微量滴定板中进行多个反应是方便的,在微量滴定板中,30 μ l 反应混合物是每孔适当的体积。因此,实例通常应用 30 μ l 中 0.6 μ g 的底物,其相当于 50 μ l 中的 1 μ g 底物。反应混合物中的底物的量不是关键的,但是优选其在反应之间为恒定。切割反应在预定温度(例如 25 $^{\circ}$ C、30 $^{\circ}$ C、37 $^{\circ}$ C、50 $^{\circ}$ C、55 $^{\circ}$ C 或 65 $^{\circ}$ C)下进行标准时间例如一小时。可以通过任何标准技术例如通过 0.8% 琼脂糖凝胶电泳测定切割产物,以测定上面定义的保真度指数。

[0160] 不是所有的限制性内切核酸酶都具有显著的星号活性,如从它们的FI确定的。然而,如果内切核酸酶具有不超过大约 250 的最高FI 和小于 100 的最低FI,那么该限制性内切核酸酶被分类为具有显著的星号活性。这样的内切核酸酶被选为酶工程的标靶以增加对单一底物的保真度。在一些情况中,具有FI 大于约 500 和FI 小于约 100 的限制性内切核酸酶也被工程化以获得更好的切割活性。

[0161] 下面表 2 列出一些工程化限制性内切核酸酶在工程化之前的FI。所有的样品在 0.8%琼脂糖凝胶上进行分析。

[0162] 表 2

[0163]

| 酶 | 稀释剂 (NEB) *** | 底物 * | 温度℃ | FI-1 ** | FI-2 ** | FI-3 ** | FI-4 ** |
|---------|---------------------|---------|-----|---------------|------------|------------|------------|
| AgeI | C | pXba | 37 | 16(1) | 8(1/2) | 64(1/8) | 8(1/2) |
| AvrII | B | T7 | 37 | 64(1) | 8(1) | 32(1/4) | 32(1) |
| BamHI | A | λ | 37 | 4(1/2) | 4(1) | 32(1) | 4(1/2) |
| BsaI | B | pBC4 | 50 | 8(1/4) | 120(1) | 16(1/4) | 32(1) |
| BsmBI | B | λ | 55 | 1(1/8) | 8(1/2) | 120(1) | 4(1/4) |
| BspQI | B | λ | 50 | 2(1/8) | 16(1) | 32(1) | 4(1/2) |
| BstXI | B | λ | 55 | 2(1/2) | 2(1/2) | 2(1/8) | 4(1) |
| EagI | B | pXba | 37 | 4(1/4) | 8(1/2) | 250(1) | 16(1) |
| EcoRI | C | λ | 37 | 250(1/2) | 4(1) | 250(1) | 4(1) |
| EcoRV | A | pXba | 37 | 32(1/16) | 120(1/2) | 1000(1) | 64(1/4) |
| HindIII | B | λ | 37 | 32(1/4) | 250(1) | 4000(1/4) | 32(1/2) |
| HpaI | A | λ | 37 | 32(1/16) | 1(1/4) | 2(1/8) | 16(1) |
| KpnI | A | pXba | 37 | 16(1) | 16(1/4) | 8(1/16) | 4(1/2) |
| MfeI | A | λ | 37 | 32(1) | 16(1/8) | 8(1/16) | 32(1) |
| NcoI | A | λ | 37 | 120(1) | 32(1) | 120(1/4) | 32(1) |
| NheI | C | pXba | 37 | 32(1) | 120(1/4) | 120(1/8) | 32(1) |
| NotI | C | pXba | 37 | ≥ 32000(1/16) | 64(1) | 500(1) | 32(1/4) |
| PciI | A | pXba | 37 | 2000(1/2) | 16(1/4) | 120(1) | 8(1/8) |

| | | | | | | | |
|-------|---|----------------|----|------------|----------|-----------|------------|
| PstI | C | λ | 37 | 64(1) | 32(1) | 120(1) | 8(1/2) |
| PvuII | A | pBR322 | 37 | 250(1) | 16(1/4) | 8(1/32) | 1/4(1) |
| SacI | A | pXba | 37 | 120(1) | 120(1/2) | 120(1/32) | 32(1/2) |
| SalI | A | λ (H3) | 37 | 8(1/500) | 1(1/16) | 32(1) | 1(1/120) |
| SapI | C | λ | 37 | 16(1/4) | 64(1/2) | 32(1/4) | 16(1) |
| SbfI | A | λ | 37 | 32(1) | 8(1/4) | 8(1/16) | 8(1/2) |
| ScaI | A | λ | 37 | 1/16(1/32) | 1/8(1) | 4(1/2) | 1/64(1/16) |
| SphI | B | λ | 37 | 64(1) | 32(1) | 64(1/4) | 16(1/2) |
| SspI | C | λ | 37 | 64(1) | 16(1) | 32(1/4) | 16(1) |

[0164] *底物： λ 是 λ 噬菌体 DNA； λ (H3) 是 HindIII- 消化的 λ 噬菌体 DNA；pXba 是 pUC19，具有 XbaI- 消化的腺病毒片段；pBC4：较短形式的 pXba；T7：T7DNA

[0165] **FI-1 到 FI-4：NEB 缓冲液 1、2、3 和 4 中的酶的保真度指数。括号中的数是与缓冲液组的任何缓冲液中突变体限制性内切核酸酶的“最好”切割活性相比，在该缓冲液组的指定缓冲液中同一突变体限制性内切核酸酶的相对切割活性的值。

[0166] NEB 缓冲液的组成如下：

[0167] NEB1：10mM Bis Tris 丙烷-HCl、10mM MgCl₂、1mM (25°C 下，pH 7.0)；

[0168] NEB2：50mM NaCl、10mM Tris-HCl、10mM MgCl₂、1mM 二硫苏糖醇 (25°C 下，pH 7.9)；

[0169] NEB3：100mM NaCl、50mM Tris-HCl、10mM MgCl₂、1mM 二硫苏糖醇 (25°C 下，pH7.9)；

[0170] NEB4：50mM 乙酸钾、20mM Tris- 乙酸盐、10mM 乙酸镁、1mM 二硫苏糖醇 (25°C 下，pH7.9)。

[0171] ***NEB 稀释剂的组成如下。(在稀释中使用稀释剂代替水将保持反应中甘油浓度为常数。)

[0172] 稀释剂 A：50mM KCl、10mM Tris-HCl、0.1mM EDTA、1mM 二硫苏糖醇、200mg/ml BSA、50%甘油 (25°C 下，pH7.4)；

[0173] 稀释剂 B：300mM NaCl、10mM Tris-HCl、0.1mM EDTA、1mM 二硫苏糖醇、500mg/ml BSA、50%甘油 (25°C 下，pH7.4)；

[0174] 稀释剂 C：250mM NaCl、10mM Tris-HCl、0.1mM EDTA、1mM 二硫苏糖醇、0.15% Triton X-100、200mg/ml BSA、50%甘油 (25°C 下，pH 7.4)。

[0175] 2. 高表达宿主细胞株的构建

[0176] 如果宿主细胞能过表达针对其寻找减少星号活性的突变体限制性内切核酸酶，这是方便的。如果该限制酶在大肠杆菌中高度表达，那么星号活性可以容易在粗提物中检测到，这简化了高保真度限制性内切核酸酶的筛查。然而，突变的限制性内切核酸酶可以在任何宿主细胞中表达，条件是宿主细胞以某些方式被保护免于从酶切割产生的毒性。这可能

包括：甲基化酶的存在；在提供接近基因组的屏障的细胞区室（例如包含体或周质）中产生；体外合成；在缺少宿主基因组中的切割位点的情况下在乳液（参见美国专利申请序列号 12/035,872）中产生；在经受内含肽介导的连接的构成部件中制造酶，等。

[0177] 为了生产，突变限制性内切核酸酶的过表达可以使用标准克隆技术来实现，例如使用大肠杆菌 (*E. coli*) 宿主，将内切核酸酶插入 pUC19- 衍生的表达载体，这是高拷贝的；和使用相对小的质粒，其能持续表达重组蛋白质。该载体可优选地含有适当的启动子例如 lac 启动子和邻近该启动子定位的多拷贝插入位点。可选地，可以选择需要基因表达的 IPTG 诱导的启动子。如果粗提物中的活性不是足够的，那么在粗提物中限制性内切核酸酶的柱纯化步骤可被进行。

[0178] 3. 限制性内切核酸酶的诱变

[0179] 编码限制性内切核酸酶内的每个带电荷的基团或极性基团的 DNA 可以被单独靶向，并且突变的 DNA 被克隆并制备进行检测。可将多突变引入单独的限制性内切核酸酶基因。限制性内切核酸酶的靶向诱变可通过本领域已知的方法完成。本文使用的一个便利的方法是反向 PCR。在该方法中，合成含有靶向密码子加在该密码子 5' 和 3' 侧上的多个核苷酸（例如 18nt）的互补引物对。通过评估目的氨基酸残基周围的目的内切核酸酶的基因序列，可容易地完成适当引物的选择。通过 REBASE 和 GenBank 提供对基因序列的访问。本文在实施例中描述的内切核酸酶的序列在图 31 到 38 和 44 中提供。PCR 的模板是含有限制性内切核酸酶基因的质粒。优选地，聚合酶为高保真度聚合酶，例如 Vent® 或 Deep Vent™ DNA 聚合酶。通过改变退火温度和 Mg²⁺ 浓度，可以实现大多数突变的成功引入。然后，纯化并优选通过 DpnI 消化 PCR 扩增产物。在本发明的一个实施方式中，将消化的产物转化入感受态宿主细胞（例如大肠杆菌），其已经使用相应的甲基化酶预修饰。来自每个突变体的菌落被挑选并在与 WT 生长条件相似的条件生长（例如，使用相似的生长培养基、药物选择和温度）。针对减少的星号活性，筛查所得限制性内切核酸酶。

[0180] 4. 筛查具有减少的星号活性的突变体限制性内切核酸酶

[0181] 条件例如缓冲液组成、温度和稀释剂应该被限定，以测定突变体限制性内切核酸酶中的星号活性。表 2 和 3 示出重组内切核酸酶在 37°C 下使用三种不同的稀释剂在四种不同的缓冲液中突变之前和之后的 FI。因此，可能确定哪些突变体具有至少 2、超过 10、至少 50 或超过 500 的总的希望提高的保真度指数因子并且选择酶作为优选的高保真度突变体。

[0182] 在本发明的一个实施方式中，首先，在正常缓冲液条件下（不超过 5% 甘油），筛查突变体限制性内切核酸酶的活性。对于那些具有至少大约 10% 的 WT 限制性内切核酸酶活性的突变体，在促进星号活性的星号活性提高条件下——例如高甘油浓度和任选地高 pH 下，也测定活性。优选地，选择在正常的缓冲液中具有最小星号活性但是具有可接受的同源活性的突变体。然后，可提取质粒并测序，以验证该突变体。在一些情况中，星号活性不容易测量——甚至在高甘油和高 pH 条件下。替代地，测量和比较不同缓冲液中的活性，并且 NEB4 与 NEB3 切割活性比率最高的突变体可进一步针对星号活性改进进行检测。

[0183] 5. 对一个单个残基的饱和诱变

[0184] 如前面部分所述的，第一步是将限制性内切核酸酶中的靶氨基酸突变为 Ala。如果该结果不令人满意，那么进行饱和诱变。优选地，这通过两个方法之一进行。一个方法是将意图密码子改为 NNN。诱变后，在正常条件下和促进星号活性的条件下分析多个菌落。

可选地,可以选择不同的密码子,用于进行下列靶氨基酸的每一个的诱变,例如:Ala:GCT; Cys:TGC; Asp:GAC; Glu:GAA; His:CAC; Ile:ATC; Lys:AAA; Leu:CTG; Met:ATG; Asn:AAC; Pro:CCG; Gln:CAG; Arg:CGT; Ser:TCC; Thr:ACC; Val:GTT; Trp:TGG 和 Tyr:TAC。

[0185] 6. 组合

[0186] 如果单一突变不充分减少星号活性,那么可以将超过一个的突变引入限制性内切核酸酶基因。突变组合和饱和诱变可以以任何顺序进行。

[0187] 7. 突变体纯化和改进的评估

[0188] 高保真度突变体可以以多种方式纯化,其包括使用不同的色谱柱。对于常规的质量评估,一个 FPLC 肝素柱足以从该制品除去 DNA 和非特异性核酸酶。包括离子交换柱、疏水柱、大小排阻柱和亲和柱的多个柱可被用于进一步纯化。

[0189] 在四种 NEB 缓冲液中,测量纯化的高保真度限制性内切核酸酶的 FI,并且将其与 WT 限制性内切核酸酶的 FI 进行比较。高保真度限制性内切核酸酶在其最佳缓冲液中的 FI 与 WT 的 FI 的比率是总改进因子。

[0190] 表 3- 示例性限制性内切核酸酶的 FI*

[0191]

| 酶 | 稀释剂 (NEB) | 底物 * | 温度 °C | FI-1 ** | FI-2 ** | FI-3 ** | FI-4 ** |
|------------|--------------|---------|----------|--------------|-------------|--------------|--------------|
| AgeI-HF | C | pXba | 37 | ≥ 500(1) | ≥ 250(1/2) | ≥ 16(1/16) | ≥ 250(1) |
| AvrII-HF | B | T7 | 37 | 500(1) | ≥ 500(1/2) | ≥ 16(1/64) | ≥ 1000(1) |
| BamHI-HF | A | λ | 37 | ≥ 4000(1) | ≥ 4000(1) | ≥ 250(1/16) | ≥ 4000(1) |
| BsaI | B | pBC4 | 50 | ≥ 4000(1/2) | ≥ 8000(1) | 120(1) | ≥ 8000(1) |
| BsmBI | B | λ | 55 | 2(1) | ≥ 500(1) | ≥ 64(1/8) | ≥ 500(1) |
| BspQI-HF | A | pUC19 | 50 | ≥ 1000(1/4) | ≥ 1000(1/4) | ≥ 64(1/64) | ≥ 4000(1) |
| BstXI-HF | A | λ | 55 | ≥ 120(1/2) | ≥ 250(1) | ≥ 16(1/16) | ≥ 250(1) |
| EagI-HF | C | pXba | 37 | 250(1/2) | 250(1) | 250(1/2) | 500(1) |
| EcoRI-HF | C | λ | 37 | 2000(1/8) | 4000(1/4) | 250(1/250) | 16000(1) |
| EcoRV-HF | A | pXba | 37 | ≥ 16000(1/4) | ≥ 64000(1) | ≥ 32000(1/2) | ≥ 64000(1) |
| HindIII-HF | B | λ | 37 | ≥ 16000(1/4) | ≥ 64000(1) | ≥ 16000(1/4) | ≥ 32000(1/2) |
| HpaI-HF | A | λ | 37 | ≥ 32(1/32) | ≥ 2000(1) | 2(1/8) | ≥ 2000(1/2) |
| KpnI-HF | A | pXba | 37 | ≥ 4000(1) | ≥ 1000(1/4) | ≥ 64(1/64) | ≥ 4000(1) |
| MfeI-HF | A | λ | 37 | ≥ 1000(1) | ≥ 250(1/4) | ≥ 16(1/64) | ≥ 500(1/2) |

| | | | | | | | |
|----------|---|----------------|----|-------------------|-------------------|-------------------|-------------------|
| NcoI-HF | A | λ | 37 | $\geq 4000(1/4)$ | $\geq 4000(1/4)$ | $\geq 1000(1/16)$ | $\geq 64000(1)$ |
| NheI-HF | C | pXba | 37 | $\geq 128000(1)$ | $\geq 4000(1/32)$ | $\geq 32(1/2000)$ | $\geq 32000(1/2)$ |
| NotI-HF | C | pXba | 37 | $\geq 8000(1/16)$ | $\geq 128000(1)$ | $\geq 4000(1/64)$ | $\geq 64000(1/2)$ |
| PciI-HF | A | pXba | 37 | NC | $\geq 2000(1)$ | $\geq 2000(1)$ | $\geq 1000(1)$ |
| PstI-HF | C | λ | 37 | $1000(1/8)$ | $4000(1/2)$ | $4000(1/4)$ | $4000(1)$ |
| PvuII-HF | A | pBR322 | 37 | $\geq 250(1/120)$ | $\geq 2000(1/16)$ | $\geq 250(1/120)$ | $500(1)$ |
| SacI-HF | A | pXba | 37 | $\geq 32000(1)$ | $\geq 16000(1/2)$ | $\geq 500(1/64)$ | $\geq 32000(1)$ |
| SaII-HF | A | λ (H3) | 37 | $\geq 8000(1/8)$ | $\geq 64000(1)$ | $\geq 4000(1/16)$ | $\geq 32000(1/2)$ |
| SbfI-HF | C | λ | 37 | $1000(1)$ | $120(1/2)$ | $8(1/32)$ | $250(1)$ |
| ScaI-HF | A | λ | 37 | $4000(1/8)$ | $1000(1)$ | $2000(1/32)$ | $1000(1)$ |
| SphI-HF | B | λ | 37 | $4000(1/8)$ | $2000(1/16)$ | $250(1/250)$ | $8000(1)$ |
| SspI-HF | C | λ | 37 | $\geq 4000(1/2)$ | $120(1/2)$ | $\geq 32(1/128)$ | $500(1)$ |

[0192] *FI 是不示出星号活性的最高浓度与完全消化底物的最低浓度的比率。

[0193] ** 括号中的数是与缓冲液组的任何缓冲液中突变体限制性内切核酸酶的最大切割活性相比,在该缓冲液组的指定缓冲液中同一突变体限制性内切核酸酶的相对切割活性的值。

[0194] 表 4- 提供具有高保真度的限制性内切核酸酶的突变

[0195]

| | |
|----------|--|
| 限制性内切核酸酶 | 具有 ≥ 2 的总改进 FI 因子的突变体的实例 |
| AgeI | R139A ;S201A* |
| AvrII | Y104F ;M29A ;E96A ;K106A ;S127A ;F142A |
| BamHI | E163A/E167T ;K30A ;E86A ;E86P ;K87A ;K87E ;K87V ;K87N ;P144A ; Y165F ;E167A ;E167R ;E167K ;E167L ;E167IK30A/E86A ;E86A/K106A ; K30A/E86A/K106A ;K30A/K87A ;E86P/K87E ;E86A/Y165F ;K30A/E167A ; E163S/E170T/P173A ;E163S/E170T/P173A ;E86P/K87T/K88N/ E163S/E170T/P173A ; E86P/K87R/K88G/E163S/E170T/P173A ;E86P/K87P/K88R/ E163S/E170T/P173A/E211K ;E86P/K87T/K88R/E163S/E170T/P173A/N158S ; E86P/K87S/K88P/E163S/E170T/P173A ;E86P/K87G/K88S/ E163S/E170T/P173A ;E86P/K87R/K88Q/E163S/E170T/P173A ; E86P/K87W/K88V ;E86P/P173A |
| BsaI | Y231F |
| BsmBI | N185Y/R232A ;H230A ;D231A ;R232A ; |
| BspQI | K279P/R388F ;K279A ;K279F ;K279P ;K279Y ;K279E ;K279D R388A ;R388F ;R388Y ;R388L ;K279P/R388F ;K279A/R388A ; D244A |
| BstXI | N65A ;Y57F ;E75A ;N76A ;K199A ; |
| EagI | H43A |
| EcoRI | K62A ;K62S ;K62L ;R9A ;K15A ;R123A ;K130A ;R131A ;R183A ;S2Y ; D135A ;R187A ;K62E |
| EcoRV | D19A ;E27A ;D19A/E27A |
| HindIII | S188P/E190A ;K198A |
| HpaI | Y29F ;E56A |
| KpnI | D148E ;D16N/R119A/D148E ;D2A/D16N/D148E ;D16N/E134A/D148E ; D16N/E132A/D148E |
| MfeI | Y173F ;Q13A/F35Y |

| | |
|-------|--|
| NcoI | D56A ;H143A ;E166A ;R212A ;D268A ;A2T/R31A |
| NheI | E77A |
| NotI | K176A ;R177A ;R253A ;K150A |
| PciI | E78A/S133A |
| PstI | E204G ;K228A ;K228A/A289V ;D91A |
| PvuII | T46A ;T46H ;T46K ;T46Y ;T46G |

[0196]

| | |
|------|--|
| SacI | Q117H/R154A/L284P ;Q117H/R200A |
| SalI | R82A ;K93A ;K101A ;R107A |
| SapI | K273P ;R380A ;K273P/R380A |
| SbfI | K251A |
| ScaI | R18A ;R112A ;E119A ;H193A ;S201F ;H193A/S201F |
| SphI | D91A ;D139A ;D164A ;K100A |
| SspI | H65A ;K74A ;E78A ;E85A ;E89A ;K109A ;E118A ;R177A ;K197A ;Y98F |

[0197] 对每种酶的突变通过分号分开。

[0198] 所有上面和下面引用的参考文献,以及美国临时申请序列号 60/959, 203, 被通过引用并入。

实施例

[0199] 在氨基酸用单字母代码表示时,这拟为标准命名。该代码的解答例如在 NEB 目录 2007/2008 第 280 页上提供。

[0200] 被用来克隆和作为底物的质粒具有如下序列:

[0201] pLaczz2(SEQ ID NO:102)、pSyx20-lacIq(SEQ ID NO:105)、pBC4(SEQ IDNO:103)、pXba(SEQ ID.NO:104)和 pAGR3(SEQ ID NO:106)。pACYC 在 GenBankX0 6403 中描述,T7 在 GenBank NC001604 中描述,pUC18 在 GenBank L09136 中描述,和 pRRS 在 Skoglund et al. Gene, 88 :1-5(1990) 中描述。pSX33 通过将 lacI 基因插入 pLG339 中 EcoRI 位点处进行构建。pLG339 在 Stoker, et al. Gene 19, 335-341(1982) 中描述。

[0202] 本文使用的称为 NEB 缓冲液的所有缓冲液可从 New England Biolabs, Inc. (NEB), Ipswich, MA 获得。

[0203] 实施例 1:工程化高保真度 BamHI

[0204] 1. 提取包含 BamHI 甲基化酶和 BamHI 内切核酸酶的质粒

[0205] 用 pUC18-BamHIR 和 pACYC184-BamHIM 转化感受态大肠杆菌宿主细胞,并且通过使用标准小量制备技术的标准 Qiagen 小量制备方法 (Qiagen, Valencia, CA) 提取 BamHIR。

[0206] 2. 诱变靶的选择

[0207] 克隆并测序 BamHI 和相关的限制性内切核酸酶 OkrAI。如果反应在 37°C 下、在 NEB 缓冲液 (1、2 和 4) 中进行,那么发现 OkrAI 具有显著的星号活性。本分析检验如此假设:作为星号活性原因的氨基酸残基 (一个或更多个) 在 BamHI 和 OkrAI 内切核酸酶之间是相似的。

[0208] BamHI 的完整蛋白序列 (SEQ ID NO :19) 为:

```
[0209] 1MEVEKEFITD EAKELLSKDK LIQQAYNEVK TSICSPIWPA TSKTFTINNT
[0210] 51EKNCNGVVPI KELCYTLLED TYNWYREKPL DILKLEKKGK GPIDVYKEFI
[0211] 101ENSELKRVGM EFETGNISSA HRSMNKLLLG LKHGEIDLAI ILMPIKQLAY
[0212] 151YLTDRVTNFE ELEPYFELTE GQPFIFIGFN AEAYNSNVPL IPKGS DGMSK
[0213] 201RSIKKWKDKV ENK
```

[0214] OkrAI 的完整蛋白序列 (SEQ ID NO :20) 为:

```
[0215] 1MKIKRIEVL I NGSVPGIPM ILNEIQDAIK TVSWPEGNNS FVINPVRKGN
[0216] 51GVKPIKNSCM RHLHQKQWAL EHPVRIKAEM RPGPLDAVKM IGGKAFALAW
[0217] 101ETGNISSSHR AINKMVMGML ERVIIGGVLI LPSRDMYNYL TDRVGNFREL
[0218] 151EPYFSVWRQF NLKDAYLAI V EIEHDSVDAQ VSLIPKGT DG RAIR
```

[0219] 通过 GCG 对 BamHI 和 OkrAI 内切核酸酶的蛋白序列的“最佳拟合 (Bestfit)”相似性分析示出下列结果,其中上面的蛋白序列是 BamHI,下面的蛋白序列是 OkrAI:

[0220] bamhir.pep x okrair.pep

```
[0221] . . . . .
[0222] 22  IQQAYNEVKTSICSPIWPATSKTFTINNTEKNCNGVVPIKELCYTLLEDT 71
[0223] |  | :. . | . | | . . | | | | | | | | |
[0224] 18  IPMILNEIQDAIKTVSWPEGNNSFVINPVRKG. NGVKPIKNSCMRHLHQK 66
[0225] . . . . .
[0226] 72  YNWYREKPLDILKLEKKGKGPIDVYKEFIENSELKRVGMEFETGNISSAH 121
[0227] | | | . | | | : | | : | | | | | | | | | | |
[0228] 67  . GWALEHPVRI. KAEMRP. GPLDAVK. MIGG... KAFALAWETGNISSSH 109
[0229] . . . . .
[0230] 122  RSMNKLLGLKHGEIDLAIILMPIKQLAYYLTDRVTNFELEPYFEL... 168
[0231] |.. |||: . :| : : : | : : | | | | | | | | | | | .
[0232] 110  RAINKMVMGMLERVIIGGVLIILPSRDMYNYLTDRVGNFRELEPYFSVWRQ 159
[0233] . . . . .
[0234] 169  .. TEGQPFIFIGFNAEAYNSNVPLIPKGS DGMSKR 201 (SEQ ID NO :21)
[0235] . : :. . . | | | | | . | | . |
[0236] 160  FNLKDAYLAI VEIEHDSVDAQVSLIPKGT DGRAIR 194 (SEQ ID NO :22)
[0237] 发现 BamHI 中相似的带电荷的残基 (D、E、H、K、R) 是 E28、K30、K52、K61、E77、K84、
```

E86、K88、D94、K97、K106、E111、E113、H121、R122、K126、K146、D154、R155、E161、E163、E170、E182、K193、D196 和 R201。在上面的比较中,这些残基被加下划线。已知的突变体 E77K、D94N、E111K 和 E113K 以前被报道为失活的 (Xu, Shuang-yong et al. J. Bacteriol. 266 : 4425-4429 (1991)), 所以排除它们。最初的诱变选择靶向 22 个共有的带电荷的氨基酸残基, 将其突变为丙氨酸 : E28A、K30A、K52A、K61A、K84A、E86A、K88A、K97A、K106A、H121A、R122A、K126A、K146A、D154A、R155A、E161A、E163A、E170A、E182A、K193A、D196A 和 R201A。

[0238] 3. BamHI 的诱变

[0239] 选择突变的点诱变通过反向 PCR 进行。将相应密码子全部变为 GCA (丙氨酸)。下列引物用于诱变 :

[0240] E28A

[0241] 5' ATTCAACAAGCATACAATGCAGTTAAAACATCTATTGT3' (SEQ ID NO :23)

[0242] 5' ACAAATAGATGTTTTAACTGCATTGTATGCTTGTGAAT3' (SEQ ID NO :24)

[0243] K30A

[0244] 5' CAAGCATACAATGAAGTTGCAACATCTATTTGTTCCACT3' (SEQ ID NO :25)

[0245] 5' AGGTGAACAAATAGATGTGCAACTTCATTGTATGCTTG3' (SEQ ID NO :26)

[0246] K52A

[0247] 5' ACGATTAACAACACCGAAGCAAATTGTAACGGTGTAGTA3' (SEQ ID NO :27)

[0248] 5' TACTACACCGTTACAATTTGCTTCGGTGTGTTAATCGT3' (SEQ ID NO :28)

[0249] K61A

[0250] 5' AACGGTGTAGTACCAATTGCAGAACTATGTTACACCTTA3' (SEQ ID NO :29)

[0251] 5' TAAGGTGTAACATAGTTCTGCAATTGGTACTACACCGTT3' (SEQ ID NO :30)

[0252] K84A

[0253] 5' AACCCCTTGATATACTTGCACTTGAAAAGAAAAAGGT3' (SEQ ID NO :31)

[0254] 5' ACCTTTTTCTTTTCAAGTGCAAGTATATCAAGGGGTTT3' (SEQ ID NO :32)

[0255] E86A

[0256] 5' GATATACTTAACTTGCAAAGAAAAAGGTGGTCCG3' (SEQ ID NO :33)

[0257] 5' CGGACCACCTTTTTCTTTGCAAGTTAAGTATATCAAG3' (SEQ ID NO :34)

[0258] K88A

[0259] 5' AACTTAACTTGAAAAGGCAAAAGGTGGTCCGATTGAT3' (SEQ ID NO :35)

[0260] 5' ATCAATCGGACCACCTTTTGCCTTTTTCAAGTTAAGTAT3' (SEQ ID NO :36)

[0261] K97A

[0262] 5' GGTCCGATTGATGTTTATGCAGAGTTCATAGAAAACAGT3' (SEQ ID NO :37)

[0263] 5' ACTGTTTTCTATGAACTCTGCATAAACATCAATCGGACC3' (SEQ ID NO :38)

[0264] K106A

[0265] 5' ATAGAAAAACAGTGAACCTGCACGTGTAGGTATGGAA3' (SEQ ID NO :39)

[0266] 5' AAATTCCATACCTACACGTGCAAGTTCAGTGTCTTCTAT3' (SEQ ID NO :40)

[0267] H121A

[0268] 5' GGAAATATTAGTTCTGCCGCACGTTCAATGAACAACTT3' (SEQ ID NO :41)

[0269] 5' AAGTTTGTTCATTGAAACGTGCGGCAGAACTAATATCC3' (SEQ ID NO :42)

- [0270] R122A
[0271] 5' AATATTAGTTCTGCCACGCATCAATGAACAACTTCTA3' (SEQ ID NO :43)
[0272] 5' TAGAAGTTTGTTCATTGATGCGTGGGCAGAACTAATATT3' (SEQ ID NO :44)
[0273] K126A
[0274] 5' GCCCACC GTTCAATGAACGCACTTCTATTAGGATTAACAT3' (SEQ ID NO :45)
[0275] 5' ATGTTTTAATCCTAATAGAAGTGCGGTCATTGAACGGTGGGC3' (SEQ ID NO :46)
[0276] K146A
[0277] 5' ATTATCCTTATGCCTATTGCACAATTGGCCTATTATCTT3' (SEQ ID NO :47)
[0278] 5' AAGATAATAGGCCAATTGTGCAATAGGCATAAGGATAAT3' (SEQ ID NO :48)
[0279] D154A
[0280] 5' TTGGCCTATTATCTTACAGCACGTGTTACCAATTTTCGAG3' (SEQ ID NO :49)
[0281] 5' CTCGAAATTGGTAACACGTGCTGTAAGATAATAGGCCAA3' (SEQ ID NO :50)
[0282] R155A
[0283] 5' GCCTATTATCTTACAGATGCAGTTACCAATTTTCGAGGAA3' (SEQ ID NO :51)
[0284] 5' TTCCTCGAAATTGGTAACTGCATCTGTAAGATAATAGGC3' (SEQ ID NO :52)
[0285] E161A
[0286] 5' CGTGTACCAATTTTCGAGGCATTAGAACCTATTTTGAA3' (SEQ ID NO :53)
[0287] 5' TTCAAAATAAGGTTCTAATGCCTCGAAATTGGTAACACG3' (SEQ ID NO :54)
[0288] E163A
[0289] 5' ACCAATTTTCGAGGAATTAGCACCTTATTTGAACTTACT3' (SEQ ID NO :55)
[0290] 5' AGTAAGTTCAAAATAAGGTGCTAATTCCTCGAAATTGGT3' (SEQ ID NO :56)
[0291] E170A
[0292] 5' CCTTATTTTGAACCTTACTGCAGGACAACCATTTATTTTATT3' (SEQ ID NO :57)
[0293] 5' AATAAAAATAAATGGTTGTGCTGCAGTAAGTTCAAAATAAGG3' (SEQ ID NO :58)
[0294] E 182A
[0295] 5' TTTATTTTATTGGATTTAATGCTGCAGCTTATAATTCTAATGTC3' (SEQ ID NO :59)
[0296] 5' GACATTAGAATTATAAGCTGCAGCATTAATCCAATAAAAATAAA3' (SEQ ID NO :60)
[0297] K193A
[0298] 5' AATGTCCCTTTAATTCGCCAGGTTCTGACGGTATGTCA3' (SEQ ID NO :61)
[0299] 5' TGACATACCGTCAGAACC TGCGGAATTAAAGGGACATT3' (SEQ ID NO :62)
[0300] D196A
[0301] 5' TTAATTCCTAAAGGTTCTGCAGGTATGTCAAACGCTCA3' (SEQ ID NO :63)
[0302] 5' TGAGCGTTTTGACATACCTGCAGAACC TTTGGGAATTAA3' (SEQ ID NO :64)
[0303] R201A
[0304] 5' TCTGACGGTATGTCAAAGCATCAATTAAGAAATGGAAA3' (SEQ ID NO :65)
[0305] 5' TTTCCATTTCTTAATTGATGCTTTTGACATACCGTCAGA3' (SEQ ID NO :66)
[0306] PCR 反应的反应体积为 100 μ l, 包含 2 μ l 的每种 PCR 引物、1 μ l pUC18-bamhiR、400 μ M dNTP、4 单位的 Deep Vent™ DNA 聚合酶和 10 μ l 10x 含有 0.2 或 6 μ M MgSO₄ 的 Thermopol 缓冲液加余量的水。

[0307] PCR反应条件是94℃进行5min,然后25个循环的94℃30sec、55℃30sec、72℃4min,并且在72℃下最后延伸时间为7mins。PCR产物在标准Qiagen离心柱(Qiagen, Valencia, CA)上纯化。用20单位的DpnI消化6到16μl的PCR产物1小时。将消化的产物转化入大肠杆菌(pACYC-bamHIM)。

[0308] 6个PCR反应后,获得工程化的22种突变中的14种:E28A、K30A、K61A、E86A、K97A、H121A、K126A、K146A、E161A、E163A、E170A、E182A和R201A。从过夜培养物中的细胞溶胞产物中提取突变体蛋白质,并且将其活性与WT BamHI比较。在有或者没有5%甘油的NEB2缓冲液中,分析正常的酶活性,而在具有39.2%甘油的NEB2中测定星号活性,尽管最初可以使用较低百分比的甘油。用于不同的反应的底物是pBR322、pUC19或λDNA。切割反应在37℃下进行30分钟或1小时。发现突变体K97A、H121A、K126A、E161A、E182A、R201A是失活的(小于1%的WT BamHI活性),而E28A、K146A、E163A、E170A突变体具有与WT酶相似水平的包括星号活性在内的活性。发现与WT BamHI相比,三种突变体K30A、E86A和K126A具有显著减少的星号活性。也发现K30A和E86A具有与WT酶相似的总切割活性,同时其显示出星号活性明显减少。相反,K126A仅具有25%的总WT酶切割活性,并且与对K30A和E86A观察到的星号活性相比具有较不显著的改进。

[0309] 对pUC18-bamHIR质粒再检查显示正常的高拷贝质粒已突变为低拷贝质粒。设计引物对以将bamHIR基因转移入高拷贝质粒:

[0310] 5'GGTGGTGCATGCGGAGGTAAATAAATGGAAGTAGAAAAAGAGTTTATTACTGAT3'(SEQ ID NO:67)

[0311] 5'GGTGGTGGTACCCTATTTGTTTTCAACTTTATCTTTCCATTTCTTAATTGA3'(SEQ ID NO:68)

[0312] 模板是pUC18-bamHIR WT,其在K30A、E86A或K126A处带有突变。PCR混合物包含:5μl模板、2μl每种引物、400μM dNTP、10μl 10x Thermopol缓冲液、4单位2μl Deep Vent™聚合酶、72μl H₂O加0、2、6μl MgSO₄。PCR条件是94℃持续5min,然后94℃30sec、55℃30sec和72℃40sec进行25个循环和7分钟的最终延伸时期。使用SphI和KpnI消化PCR产物,将其连接至具有相同的酶消化对的pUC19。将连接的产物转化入含有pACYC-bamHIM的感受态大肠杆菌。含有BamHIR K30A的pUC19形式的26个菌落和含有E86A的那些的12个被鉴定并生长。检查来自这些培养物的BamHI的活性。所有它们都具有活性。提取来自每个突变的五个菌落的质粒并且来自每个突变的3个的BamHIR质粒被测序。证实质粒pUC19-BamHI(K30A)和pUC19-BamHI(E86A)的身份(identity)。

[0313] 使用pUC19-BamHI(K30A)载体重复那些在pUC18-BamHIR中不成功的突变。PCR混合物包含:1μl模板和含有2μl每种引物、400μM dNTP、10μl 10xThermopol缓冲液、4单位2μl Deep Vent™聚合酶、76μl H₂O加0、2、6μl 100μMMgSO₄的扩增混合物。PCR条件是94℃持续5min,然后94℃30sec、55℃30sec和72℃3分30秒进行25个循环和7分钟的最终延伸时期。使用DpnI消化PCR产物,并且将其转化入用pACYC-BamHIM转化的感受态大肠杆菌。在pUC19底物上检测酶活性。反应混合物为:3μl细胞提取物、3μl NEB2、3μl 50%甘油、0.5μl 0.5μg pUC19、20.5μl H₂O。该反应在37℃下进行1小时。K30A/R122A、K30A/R155A和K30A/K193A是失活的。K30A/K52A和K30A/K88A是K30A活性的大约1/10。K30A/K106A、K30A/D154A和K30A/D196A的正常活性与K30ABamHI的活性相似。这三种突变体与K30A在高浓度甘油(39.2%)下的星号活性的比较显示K30A/D196A具有与

K30A 相似的星号活性, K30A/D154A 甚至具有比 K30A 更多的星号活性, 并且 K30A/K106A 具有比 K30A 更少的星号活性。因为细胞毒性, 分离在 pUC19 载体中的 BamHI 的 K106A 突变的尝试失败。

[0314] 使用反向 PCR, 组合在 K30、E86 和 K106 位点上的突变: K30A/E86A、E86A/K106A、K30A/K106A 和 K30A/E86A/K106A。K30A/E86A 表现为优选的突变体。纯化后, 发现 BamHI 突变体的 FI 在所有 NEB 缓冲液中提高 25%。

[0315] 在 K30 和 E86 位点上, 随机进行进一步诱变:

[0316] 对 K30:

[0317] 5' CAAGCATACAATGAAGTTNNACATCTATTTGTTCCCT3' (SEQ ID NO:69)

[0318] 5' AGGTGAACAAATAGATGTNNNACTTCATTGTATGCTTG3' (SEQ ID NO:70)

[0319] 对 E86:

[0320] 5' GATATACTTAACTTNNNAAGAAAAAAGGTGGTCCG3' (SEQ ID NO:71)

[0321] 5' CGGACCACCTTTTTTCTTNNNAAGTTAAGTATATCAAG3' (SEQ ID NO:72)

[0322] PCR 组合物是: 1 μ l 模板 (pUC19-BamHIR(K30A) 或 pUC19-BamHIR(E86A)) 并且使用如上描述的扩增混合物。进行 PCR: 94°C 5min, 然后 94°C 30sec、55°C 30sec 和 72°C 3 分 30 秒 25 个循环和 7 分钟的最终延伸时期。通过 DpnI 消化 PCR 产物, 并将其转化入大肠杆菌 (pACYC-BamHIM)。

[0323] 对 K30 随机突变选取共 155 个菌落, 和对 E86 位点选取 158 个菌落。将菌落生长过夜并制成细胞提取物。在具有 42.5% 甘油的 NEB2 缓冲液中用 1 μ l 细胞提取物在 37°C 下, 消化 0.5 μ g pUC19, 1 小时。在具有 39.2% 甘油的 NEB2 缓冲液中, 具有明显更少星号活性的细胞提取物在 37°C 下在 1、4、16 倍稀释下对 0.5 μ g pUC19 进行重新分析 30 分钟。对于观察到具有减少星号活性的那些突变体, 提取相应的质粒并且测序以证实突变。总共 3 个克隆 (#12、#66 和 #82) 含有 K30 突变, 并且总过 33 个克隆 (#5、#15、#16、#19、#29、#47、#51、#55、#56、#58、#61、#69、#71、#73、#76、#82、#86、#88、#93、#94、#97、#98、#100、#104、#107、#113、#117、#118、#129、#132、#136、#139 和 #151) 被测序。测序后, 发现 #12 和 #66 含有 K30G 突变, 和 #82 含有 K30N 突变。令人惊奇地, 所有 33 个突变是 E86P 突变, 恰好在不同的密码子中 (CCA、CCT、CCC、CCG)。在这些密码子中, CCG 在大肠杆菌以最高频率发生 (克隆 #98、#136 和 #139)。

[0324] 相应于 K30G、K30N 和 K30A 的细胞提取物被连续稀释为 1、2、4、8、16 和 32 倍, 而 E86P 和 E86A 被连续稀释 1、2、4、8、16、32、64、128 和 256 倍。连续稀释的提取物与 0.5 μ g pUC19 在具有 39.2% 甘油的 NEB2 中在 37°C 下反应 30 分钟。在极端的条件下, E86P 显示比其他突变体更优良。在上至 32 倍消化下, 没有明显星号活性带。E86P 和 K30 突变体 (K30G、K30N 和 K30A) 之间的差异是如此大以致不另外需要在 E86P 突变体中组合这些突变中的任一个。

[0325] 对 1 μ g λ DNA——底物——测定 BamHI (E86P) 活性 (也用于 WT BamHI 活性测定)。在 NEB1 缓冲液中, 在 37°C 下, 进行分析 1 小时。

[0326] 4. BamHI (E86P) 和 WT BamHI 的详细比较

[0327] A. 不同 NEB 缓冲液中 BamHI (E86P) 的活性

[0328] 使用 λ DNA 底物在 37°C 下, 在 NEB1、NEB2、NEB3、NEB4 和 NEB BamHI 缓冲液中测定

纯化的 BamHI (E86P) 的活性 1 小时。BamHI (E86P) 在 NEB1 缓冲液和 NEB2 中是最有活性的，而在 NEB3、NEB4 和 BamHI 缓冲液中具有 50%、50% 和 25% 的活性水平。

[0329] B. 比较 BamHI (E86P) 和 WT BamHI 对 pUC19 的切割活性

[0330] 在 pUC19 中具有一个 GGATCC 位点 (BamHI 位点) 和 6 个 AGATCC 位点 (BamHI 星号活性位点)，所以 pUC19 被选择作为比较 BamHI (E86P) 和 WT BamHI 的优选底物。

[0331] 在不同的缓冲液中，用 NEB 稀释缓冲液 A 以 1、3、9、27、81、243、729、2181、6561 和 19683 倍连续稀释的 WT BamHI 和 BamHI (E86P) 消化 0.5mg pUC19。WT BamHI 示出每种 NEB 正常缓冲液中的星号活性，而 BamHI (E86P) 根本没有示出星号活性带 (图 2-5)。这证明 BamHI (E86P) 已经极大地减少星号活性，同时保持同源切割活性。

[0332] C. BamHI (E86P) 和 WT BamHI 对 λ DNA 底物切割活性的比较

[0333] 为了计算保真度指数，用稀释缓冲液稀释限制酶，并且将甘油浓度恒定保持在 5%。在本文使用的标准反应条件中， λ DNA 底物浓度为 1 μ g，并且总反应体积为 50 μ l。为了保持酶体积为 10%，酶以 5 μ l 的体积加入。这等于在 30 μ l 总体积中 3 μ l 限制酶消化的 0.6 μ g 的底物。在 NEB1、NEB2、NEB3、NEB4 和 NEB BamHI 缓冲液中，从 1 到 32768 的 1 : 2 连续稀释的 3 μ l WT BamHI 和 BamHI (E86P) 在 37 $^{\circ}$ C 下对 0.6mg λ DNA 消化 1 小时。

[0334] 表 5 : 在不同缓冲液中，WT 和突变体 BamHI 的保真度指数

| 缓冲液 | BamHI(E86P) | | WT BamHI | | 改进因子 |
|-----------|-------------|-------------|----------|-------|-------------|
| | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| NEB1 | 100% | ≥ 4000 | 50% | 4 | ≥ 1000 |
| NEB2 | 100% | ≥ 4000 | 100% | 16 | ≥ 250 |
| NEB3 | 50% | ≥ 4000 | 100% | 32 | ≥ 125 |
| NEB4 | 50% | ≥ 4000 | 50% | 4 | ≥ 1000 |
| BamHI 缓冲液 | 25% | ≥ 2000 | 50% | 32 | ≥ 125 |

[0335] 5. 针对高保真度突变体的 BamHI 的进一步改进

[0337] 在一小时水平，BamHI (E86P) 表现为好的高保真度 BamHI 突变体。然而，尽管 E86P 的星号活性在一小时没有检测到，但是当延长反应时间时 (例如过夜或 14 小时)，星号活性带出现。(图 3)。继续进行对改进高保真度 BamHI 的研究。

[0338] 6. 其他带电荷的和极性残基的突变

[0339] 在 SEQ ID NO : 19 中位置 2、4、5、6、10、11、13、14、18、19、20、43、51、62、69、70、76、77、78、81、87、89、94、98、101、104、107、111、113、132、133、135、137、160、167、200、204、205、207、208、209、211、213 处，将其他带电荷的残基 (Arg、Lys、His、Asp、Glu) 突变为 Ala。在 pUC19-BamHI (K30A) 的模板上进行突变。

[0340] 在 SEQ ID NO : 19 的位置 9、17、26、32、36、41、42、44、46、50、65、66、71、72、75、96、103、114、118、119、123、150、151、153、157、165、169、184、186、195、199、202 处，将其他极性残基 (Ser、Thr 和 Tyr) 突变为 Ala，而将 Tyr 突变为 Phe。

[0341] 通过使用相似的突变和筛查方法，发现下列突变具有减少的星号活性 : K30A/K87A、E86P/K87E、E86A/Y165F 和 K30A/E167A。鉴定 E86P/K87E 为在存在另外的 DMSO 情况下具有改进性质的突变体。然而，在正常反应缓冲液中该突变体的活性比 WT BamHI 的活性

低得多。

[0342] 进行下列突变组合 :E86P/Y165F、E86P/E167A、E86P/Y165F/E167A、K30A/Y165F/E167A、K30G/Y165F/E167A、K30A/Y165F/E167A、E86A/Y165F/E167A。它们都具有低活性。

[0343] 到现在为止,发现对 BamHI 星号活性,E167A 和 Y165F 具有强效果,K87A 具有中等效果,以及 K30A 和 E86A 具有弱效果。E86P 是在 1 小时水平而非过夜地减少星号活性的特异性突变。

[0344] 7. 将 E167 和 Y165 突变为所有其他残基

[0345] 在 pUC19-BamHI 中,将 E167 突变为所有其他的残基,这通过将密码子改变为 GCA—Ala、TGC—Cys、GAC—Asp、TTC—Phe、GGT—Gly、CAC—His、ATC—Ile、AAA—Lys、CTG—Leu、ATG—Met、AAC—Asn、CCG—Pro、CAG—Gln、CGT—Arg、TCC—Ser、ACC—Thr、GTT—Val、TGG—Trp 和 TAC—Tyr 进行。

[0346] 在比较所有的突变体之后,优选 E167T 突变,而与 E167A 相比,E167R、E167K、E167L 和 E167I 突变示出在减少星号活性方面的改进。

[0347] 也将 Y165 突变为所有其他氨基酸残基,这通过将相应的密码子改变为 GCT—Ala、TGC—Cys、GAC—Asp、GAA—Glu、GGT—Gly、CAC—His、ATC—Ile、AAA—Lys、CTG—Leu、ATG—Met、AAC—Asn、CCG—Pro、CAG—Gln、CGT—Arg、TCC—Ser、ACC—Thr、GTT—Val、TGG—Trp 进行。

[0348] 在比较所有的突变体之后,Y165F 的存在导致显著的切割活性,而刚刚在上文列出的其他突变示出低活性或没有切割活性。

[0349] 8. 对 BamHI (E167T) 进一步突变

[0350] 在 puc19-BamHI (E167T) 的模板上,将所有带电荷的和极性残基突变为 Ala,方法同上。

[0351] 作为优选突变,将 E163A/E167T 鉴定为 BamHI-HF。

[0352] 9. BamHI-HF 与 WT BamHI 的比较

[0353] 在 E163 处引入突变产生减少热稳定性的 BamHI 突变体,这与突变 P173A 被加入其他负责减少星号活性的突变时一样。

[0354] 与 BamHI (E86P) 不同,BamHI-HF 在 NEB1-4 缓冲液中过夜反应没有显著的星号活性。图 4 示出在 NEB1 和 NEB2 中的结果。因此,选择 BamHI (E163A/E167T) 为优选的高保真度 BamHI。

[0355] 在 37°C 下,在所有四种 NEB 缓冲液中,用稀释剂 A,在 λ DNA 底物上,测量 BamHI-HF 的保真度指数,并且与 WT 酶比较。

[0356] 表 6 :BamHI-HF 和 WT BamHI 的比较

[0357]

| 缓冲液 | BamHI-HF | | WT BamHI | | 改进因子 |
|------|----------|-------|----------|-------|-------|
| | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| NEB1 | 100% | ≥8000 | 50% | 4 | ≥1000 |
| NEB2 | 50% | ≥4000 | 100% | 16 | ≥250 |
| NEB3 | 12.5% | ≥250 | 100% | 32 | ≥8 |
| NEB4 | 50% | ≥4000 | 50% | 4 | ≥1000 |

[0358] 在 NEB1 中, BamHI-HF 具有最高的活性, 保真度指数为 ≥ 8000, 在 NEB2 和 NEB3 中, WT BamHI 具有最高的活性, 并且最高的 FI 为 32。总 FI 改进因子——其为每种突变体和 WT 酶在最好的缓冲液中的 FI 的比率, 是 ≥ 8000/32 = 250 倍。

[0359] 10. BamHI 的另外突变

[0360] 预测 E163A/E167T/P173A 具有优选的星号活性减少, 并且另外是热不稳定的。

[0361] 检测 (E86P/K87S/K88P/E163S/E170T/P173A)。该突变体在比活性方面显示 10 倍减少, 而具有来自宿主细胞的补偿增加的蛋白质产率。

[0362] 共享减少的热稳定性、减少的星号活性和可接受的比活性的其他 BamHI 突变体包括:

[0363] E86P/K87R/K88G/E163S/E170T/P173A

[0364] E86P/K87P/K88R/E163S/E170T/P173A/E211K

[0365] E86P/K87T/K88R/E163S/E170T/P173A/N158S

[0366] E86P/K87S/K88P/E163S/E170T/P173A

[0367] E86P/K87G/K88S/E163S/E170T/P173A

[0368] E86P/K87R/K88Q/E163S/E170T/P173A

[0369] 实施例 2 : 制备高保真度 EcoRI

[0370] 1. EcoRI 的表达

[0371] 对 EcoRI 的 PCR 使用下列引物:

[0372] GGTGGTGCATGCGGAGGTAAATAAATGTCTAATAAAAAACAGTCAAATAGGCTA (SEQ ID NO :73)

[0373] GGTGGTGGTACCTCACTTAGATCTAAGCTGTTCAAACAA (SEQ ID NO :74)

[0374] 然后, 用第二对限制性内切核酸酶——SphI 和 Acc65I——消化 PCR 产物, 并且连接入用同样的第二对限制性内切核酸酶消化的 pUC19。然后, 连接的质粒被转化入用 pACYC-MlucIM 预修饰的感受态大肠杆菌。

[0375] 2. EcoRI 的诱变

[0376] 靶氨基酸残基的最初选择是 EcoRI 与其同切点酶 RsrI 比较的结果, RsrI 的星号活性也是已知的。

[0377] EcoRI 对 RsrI

[0378]

[0379] 4 KKQSNRLTEQHKLSQGVIGIFGDYAKAHD LAVGEVSKLVKKALSNEY PQL 53

[0380] | | . | | : | | | | . | | | . : | | . | | . : : | |

[0381] 10 KGQALRLGIQQELGGGPLSIFGAAAQKH DLSIREVTAGVLT KLAEDFPNL 59

[0382]

[0383] 54 SFRYRDSIKKTEINEALKKIDPDLGGTLFVSNSSIKPDGGIVEVKDDYGE 103

[0384] | . | | : | | | | : | | | | | | | | | | | | | | : |

[0385] 60 EFQLRTSLTKKAIN EKLR SFD PRLGQALFVESASIRPDGGITEVKDRHGN 109

[0386]

[0387] 104 WRVVLVAEAKHQGKDIINIRNGLLVGKRGDQDLMAAGNAIERSHKNI SEI 153

[0388] | | | : | | | . | | | | : | | . | | | | | | | | | | | | | | : | :

[0389] 110 WRVILVGESKHQGN DVEKILAGVLQ GKAKDQDFMAAGNAIERMHKNVLEL 159

[0390]

[0391] 154 ANFMLSESHFPYVLFLEGSNFLTENISITRPDGRVVNLEYNSGILNRLDR 203

[0392] |: || | |||||. ||: |||| ||. : ||||| | : :. ||. |||: ||

[0393] 160 RNYMLDEKHFPYVVFLQGSNFATESFEVTRPDGRVVKIVHDSGMLNRIDR 209

[0394]

[0395] 204 LTAANYGMPINSLCINKFVNHKDKSIMLQAASIYTQDGREWDSKIMFE 253

[0396] . ||.. || | | | | | | | | | | : | . | . | |

[0397] 210 VTASSLSREINQNYCENIVVRAGSFDHMFQIASLYCK. . AAPWTAGEMAE 257

[0398] 254 IMFDISTTSLRVLGRDL 270 (SEQ ID NO :75)

[0399] | :. ||||: : ||

[0400] 258 AMLAVAKTSLRIIADDL 274 (SEQ ID NO :76)

[0401] 除了 D91、E111 和 K113——其为已知的活性中心残基，在两种内切核酸酶中，42 个带电荷的残基是相同或相似的。带电荷的残基如下：

[0402] K4、R9、K15、K29、H31、D32、E37、E49、R56、R58、K63、E68、K71、D74、K89、E96、K98、K99、R105、H114、D118、K130、D133、D135、E144、R145、H147、K148、E152、E160、H162、E170、E177、R183、D185、R200、D202、R203、E253、R264、D269。

[0403] 将所有这些带电荷的残基突变为 Ala (密码子 GCA、GCT、GCC 或 GCG)，并且突变的基因被如下扩增和克隆：

[0404] 扩增混合物与在实施例 1 中使用的扩增混合物相同 (2 μ l 每种 PCR 引物，400mM dNTP，4 单位 Deep Vent DNA 聚合酶，具有另外 0、2、6 μ l MgSO₄ 的 10 μ l 10x Thermopol 缓冲液，并且总反应体积为 100 μ l)，并且加入到 1 μ l pUC19-EcoRI 中。

[0405] PCR 反应条件为 94 $^{\circ}$ C 持续 5min，然后 25 个循环的 94 $^{\circ}$ C 30sec、55 $^{\circ}$ C 30sec、72 $^{\circ}$ C 3 分 30 秒，并且在 72 $^{\circ}$ C 下最后延伸时间为 7min。PCR 后，产物通过标准 Qiagen 离心柱 (Qiagen, Valencia, CA) 纯化。用 20 单位的 DpnI 消化 16 μ l 的 PCR 产物 1 小时。将消化的产物转化入甲基化酶保护的感受态大肠杆菌制品。

[0406] 3. 筛查 EcoRI 高保真度突变体

[0407] 对每一突变选取 3 个菌落，并且在具有氨苄青霉素和氯霉素的 LB 中生长过夜。对 pBR322 和 λ DNA 进行活性分析以确保突变体与 WT EcoRI 至少具有相似的活性。然后，使用 2 倍连续稀释的 3 μ l 的细胞抽提物、12 μ l 50% 甘油、3 μ l NEB1 缓冲液、0.5 μ l pBR322 和 11.5 μ l 水，检测这些突变体，其在 37 $^{\circ}$ C 下反应 1 小时。然而，没有突变改进星号活性的性能。

[0408] 从该结果，推导出在同切点酶之间有效突变不总是被识别为同源残基。

[0409] 4. 对剩余的 32 个带电荷的残基重复诱变

[0410] 如在步骤 2 中所述的，通过靶向氨基酸残基 5、12、14、26、40、43、44、59、62、65、72、76、100、103、117、123、131、192、221、225、226、227、228、242、244、245、247、249、257、268、272 和 277，将所有剩余的 32 个带电荷的残基突变为 Ala。

[0411] 上面的数字相应于 EcoRI 蛋白序列 (SEQ ID NO :83) 中的氨基酸位置。

[0412] 5. 重复选择

[0413] 从含有不同类型突变的每种样品中选取四个菌落，并且在 4ml 具有 CAM 的 LB 中生

长。超声处理后,在 NEB1 缓冲液中在正常甘油条件下,在 λ DNA 底物上检测细胞提取物。通过在 $3 \mu\text{l}$ NEB2 缓冲液中加入 2 倍连续稀释的 $3 \mu\text{l}$ 细胞提取物至 $0.5 \mu\text{l}$ 的 pUC19 和 $23.5 \mu\text{l}$ 50%甘油以在反应混合物中提供 39.2%甘油的终浓度,具有相似活性的那些提取物在 pUC19 底物上再次检测。

[0414] 在所有这些突变体中,发现 K62A 是具有最少星号活性和高 FI 的突变。R9A、K15A、R123A、K130A、R131A、R183A 突变体都示出星号活性的部分减少。令人感兴趣地,含有靶向突变 K5A 的一个克隆示出部分改进。另外地,测序后,发现第二突变 S2Y。分离这两个突变显示该分离物有效的突变是 S2Y。D135A 和 R187AEcoRI 也具有少得多的星号活性。然而,这些突变体的切割活性不是最佳的。

[0415] 6. EcoRI (K62A) 与 WT EcoRI 的比较

[0416] 通过在四种不同 NEB 缓冲液中消化 $0.6 \mu\text{g}$ 的 λ DNA,在使用 NEB 稀释缓冲液 C 进行 3 倍连续稀释中进行并行比较(图 6)。EcoRI (K62A) 具有实质上比 WT EcoRI 小的星号活性。

[0417] 通过测定 EcoRI (K62A) 和 WT EcoRI 的保真度指数量度,进行更定量的比较。保真度指数测量的条件与表 2 相同,其使用 λ DNA 作为底物并使用稀释缓冲液 C。该反应在 37°C 温育 1 小时,并且在 0.8%琼脂糖凝胶上分析消化产物。

[0418] 表 7 :EcoRI (K62A) 和 WT EcoRI 的保真度指数

[0419]

| 缓冲液 | EcoRI(K62A) | | WT EcoRI | | 改进因子 |
|-----------|-------------|-------|----------|-------|------|
| | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| NEB1 | 50% | 32000 | 50% | 250 | 128 |
| NEB2 | 50% | 8000 | 100% | 4 | 2000 |
| NEB3 | 12.5% | 4000 | 100% | 250 | 16 |
| NEB4 | 100% | 32000 | 100% | 4 | 8000 |
| EcoRI 缓冲液 | 12.5% | 4000 | 100% | 250 | 16 |

[0420] 7. EcoRI 的进一步突变

[0421] 尽管对于 λ DNA 底物, EcoRI (K62A) 并不明显具有星号活性,但是使用 Litmus28 底物在 10 小时消化之后观察到星号活性。与 EcoRI 缓冲液中的 WT EcoRI 相比,NEB4 中的 EcoRI (K62A) 具有显著减少的星号活性(图 7)。

[0422] 研究进一步的改进。通过如在实施例 1 中将 K 改变为相应密码子,将 EcoRI (K62) 突变为所有其他的氨基酸残基。K62S 和 K62L 与 K62A 相似。当与 EcoRI (K62A) 相比时, EcoRI (K62E) 具有 ≥ 100 倍总保真度指数改进因子,如在图 6 中所示。EcoRI (K62E) 被命名为 EcoRI-HF。

[0423] 8. EcoRI-HF 和 WT EcoRI 的比较

[0424] 通过在稀释剂 C 中 EcoRI-HF 和 WT EcoRI 上的 FI 测量,进行定量比较。FI 测量的条件与表 2 中相同,其使用 λ DNA 作为底物。反应条件是 37°C 持续 1 小时,并且在 0.8%琼脂糖凝胶上分析结果(图 8)。

[0425] 表 8 :EcoRI-HF 和 WT EcoRI 的比较

| 缓冲液 | EcoRI-HF | | WT EcoRI | | 改进因子 |
|-----------|----------|-------|----------|-------|------|
| | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| NEB1 | 12.5% | 2000 | 50% | 250 | 8 |
| NEB2 | 100% | 4000 | 100% | 4 | 1000 |
| NEB3 | 0.4% | 250 | 100% | 250 | 1 |
| NEB4 | 100% | 16000 | 100% | 4 | 4000 |
| EcoRI 缓冲液 | 0.4% | 250 | 100% | 250 | 1 |

[0427] 发现总保真度指数改进因子为 64 倍 (EcoRI-HF 在 NEB4 中的 16000 比 NEB3 中 WT EcoRI 的 250)。

[0428] 实施例 3 :工程化高保真度 ScaI

[0429] 1. ScaI 的表达

[0430] ScaI 限制性内切核酸酶和甲基化酶的序列在 REBASE 和 GenBank 中描述,并且在图 44 (SEQ ID NO :97) 中提供。将表达这些酶的基因插入质粒以分别产生 pRRS-ScaI 和 pACYC184-ScaIM。然后,将 pACYC184-ScaIM 转化入感受态大肠杆菌宿主细胞。pRRS 载体从 pUC19 得到,并且差异仅仅为存在多克隆位点。将 pRRS-ScaIR 转化入大肠杆菌 (pACYC-ScaIM) 以制造表达菌株。铺板和细胞培养在 30℃ 下进行。

[0431] 2. ScaI 的诱变

[0432] ScaI 具有两种测序的同切点酶 :LlaDI 和 NmeSI。然而,没有已知的关于 LlaDI 或 NmeSI 的星号活性的信息,也没有关于这些酶活性位点的任何信息。因此,最初选择在该蛋白质的位置 4、8、11、12、14、18、25、27、30、37、39、40、43、46、51、57、61、68、72、74、80、86、97、103、108、112、114、119、120、121、127、128、129、133、135、139、140、141、147、152、156、158、159、161、162、171、172、175、179、182、184、187、172、175、192、193、195、200、222、227 处所有 58 个带电荷的残基进行靶向突变。

[0433] 上面的数字相应于 ScaI 蛋白序列 (SEQ ID NO :97) 中的氨基酸位置。

[0434] 引物设计方法和 PCR 方法与对实施例 1 中 BamHI 和实施例 2 中 EcoRI 描述的相似。通过改变退火温度和 DNA 聚合酶,实现诱变。用 DpnI 消化 PCR 产物,并转化入感受态大肠杆菌 (pACYC184-ScaIM)。

[0435] 3. ScaI 高保真度突变体的选择

[0436] 选取来自 ScaI 突变体的每个突变体的 4 个菌落,并且在 30℃ 下、4ml 具有 100 μg/ml Amp 和 33 μg/ml Cam 的 LB 中生长过夜。将每个细胞培养物进行超声处理,并且在 NEB2 缓冲液中对 λ DNA 检测活性。明显具有活性的那些以 10、100 和 1000 倍稀释再次检验。因为 ScaI 具有非常显著的星号活性,所以对于突变体对 WT 限制性内切核酸酶,星号活性带容易比较。具有减少的星号活性的那些用使用 NEB 稀释缓冲液 A 进行 2 倍连续稀释再次检验。对每一突变体测量 FI。对于 NEB 2 缓冲液中的 WT ScaI, FI 为 1/8。发现具有相似活性水平的四种突变体与 WT ScaI 相比具有大量减少的星号活性。突变体 #6-3ScaI 与 WT ScaI 相比具有 2 倍的活性并且 FI 为 4 或 32 倍。#26-2ScaI 与 WT 相比具有 2 倍的活性并且 FI

为 8 或 64 倍；#28-2ScaI 与 WT 相比具有 2 倍的活性并且 FI 为 120 或 1000 倍；#54-3 与 WT 具有相似的活性，并且 FI 为 WT 的 250 或 2000 倍。

[0437] 四种突变体：#6-3、#26-2、#28-2 和 #54-3ScaI 在存在 36.7% 甘油的情况下进一步检测消化 λ DNA 底物。#54-3 在减少星号活性方面示出比其他 3 种突变体更大的提高。

[0438] 在提取质粒后，测序 #6-3，并且发现其具有 R18A 的突变。测序 #26-2，并且发现其具有 R112A 的突变。测序 #28-2，并且发现其具有 E119A 的突变。这些突变被预测到。然而，发现 #54-3 具有双突变体——H193A/S201F。S201F 是在 PCR 期间发生的自发第二突变，并且位于 H193A 突变的引物区域外。

[0439] 为了理解哪个残基主要负责星号活性的减少，使用下列引物，将单一突变 (S201F) 引入 ScaI：

[0440] 5' -GATTGGGTGGCGCAGAAATTTCAAACGGGCCAGCAGTCG-3' (SEQ ID NO :77)

[0441] 5' -CGACTGCTGGCCCGTTTGAATTTCTGCGCCACCCAATC-3' (SEQ ID NO :78)。

[0442] 证实 ScaI (H193A)、ScaI (S201F) 和 ScaI (H193A/S201F) 的序列。在 5% 和 37% 甘油水平下，比较三种突变体和 WT ScaI (图 9)。与仅微弱有助于 FI 的 H193A 相反，S201F 明显有助于 FI。然而，这两种突变在提高 FI 方面表现是加和的。在 5% 甘油中，S201F 没有显示星号活性，但是在 37% 甘油中显示一些星号活性。在 5% 甘油中，H193A 具有一些星号活性，在 37% 甘油中具有显著的星号活性。然而，对于这两种突变的组合，在 5% 或 37% 甘油中，都没有检测到星号活性。该发现不但示出具有羟基的氨基酸可以是星号活性的主要活性残基，而且该突变的正确组合可以推动保真度提高到非常高的水平。如果带电荷的残基的突变不能改进星号活性，这里观察到对 Ser、Thr 和 Tyr 的突变可成功提高保真度指数。将 ScaI (H193A/S201F) 标记为 ScaI-HF。

[0443] 4. ScaI-HF 和 WT ScaI 的比较

[0444] 37°C 下，在不同的 NEB 缓冲液中，用 NEB 稀释缓冲液 A 进行的 2.5 倍连续稀释下，对 4 种不同的 NEB 稀释液中的 1 μ g λ DNA，比较 ScaI-HF 和 WT ScaI，1 小时 (图 10)。

[0445] 表 9 :ScaI-HF 与 WT ScaI 的保真度指数的比较

| 缓冲液 | ScaI-HF | | WT ScaI | | 改进因子 |
|-------------|---------|-------|---------|-------|-------|
| | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| [0446] NEB1 | 12% | 250 | 6% | 1/64 | 16000 |
| NEB2 | 100% | 120 | 100% | 1/8 | 2000 |
| NEB3 | 3% | 2000 | 25% | 4 | 500 |
| [0447] NEB4 | 100% | 250 | 1% | 1/32 | 8000 |

[0448] 在 NEB2 和 NEB4 缓冲液中，ScaI-HF 表现最好，其中最好的 FI 为 250；在 NEB2 缓冲液中，WT ScaI 表现最好，其中 FI 为 1/8。总 FI 改进因子为 $250/(1/8) = 4000$ 。

[0449] 实施例 4 :工程化高保真度 SaII

[0450] 1. SaII 的表达

[0451] SaII 在用 placzz1-SaIIR 和 pACYC-Hpy166IIM 转化的大肠杆菌中表达，其中 placzzI 是 pUC19 质粒，其利用 lac 启动子表达插入到邻近多拷贝位点的限制性内切核酸酶基因。Hpy166IIM 保护 SaII 的外部 4 个碱基。

[0452] 2. SalI 的诱变

[0453] 使用在先前实施例中相似的 PCR 方法, 将 SalI 的 86 个带电荷的残基突变为 Ala : 5、6、8、9、12、13、19、27、31、34、35、37、42、43、45、50、60、63、65、67、73、82、83、84、90、93、97、100、101、103、107、109、111、114、116、119、126、129、131、134、140、143、145、147、148、156、157、164、168、172、173、174、180、181、186、190、191、193、210、218、226、232、235、237、238、244、246、250、256、257、258、259、260、261、264、266、271、275、297、300、304、305、306、308、309、311。

[0454] 上面的数字相应于 SalI 蛋白序列 (SEQ ID NO :94) 中的氨基酸位置。

[0455] 突变体在 30°C 下, 在具有 Amp 和 Cam 的 LB 中生长过夜。

[0456] 3. SalI-HF 的选择

[0457] SalI-HF 的选择如在前面实施例中描述的进行。主要差别是 SalI 的星号活性在粗提物中不能容易地分析——不论在 5% 甘油或更高甘油浓度中。甘油不仅促进 SalI 的星号活性, 而且极大地抑制同源活性。

[0458] 在 5% 甘油和 37% 甘油中, 对 HindIII 消化的 λ DNA 分析活性突变体。在所有 4 种 NEB 缓冲液中, 检测突变体 #22、#26、#29、#31、#43 和 #51 的切割活性。在不同条件和底物中数轮比较后, 发现 #31——SalI (R107A)——是优选的突变体, 其保留高的切割活性, 但是显示大量减少的星号活性。将 SalI (R107A) 标记为 SalI-HF。

[0459] 4. SalI-HF 和 WT SalI 的比较

[0460] 测定 SalI-HF 和 WT SalI 的 FI (图 11)。结果如在表 10 (下面) 中所示出的 :

[0461] 表 10 :SalI-HF 和 WT SalI 的比较

| 缓冲液 | SalI-HF | | WT SalI | | 改进因子 |
|------|---------|-------------|---------|-------|-------|
| | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| NEB1 | 50% | ≥ 1000 | 0.2% | 8 | 16000 |
| NEB2 | 100% | ≥ 2000 | 6% | 1/8 | 2000 |
| NEB3 | 25% | ≥ 500 | 100% | 4 | 500 |
| NEB4 | 100% | ≥ 2000 | 0.8% | 1/32 | 8000 |

[0463] 在 NEB 2 和 NEB 4 缓冲液中, SalI-HF 表现最好, 其中 FI 都 ≥ 2000 ; WT SalI 在 NEB 3 缓冲液中表现最好, 其中 FI 为 4。总 FI 改进因子为 $\geq 2000/4 = \geq 500$ 。

[0464] 实施例 5 : 工程化高保真度 SphI

[0465] 1. SphI 的表达

[0466] 在大肠杆菌 (placzz1-SphIR、pACYC 184-CviAIIM) 中表达 SphI。CviAIIM 保护 SphI 的内部四个碱基。转化的细胞在 37°C 下, 在具有 Amp 和 Cam 的 LB 中生长过夜。

[0467] 2. SphI 的诱变

[0468] 使用在实施例 1 至实施例 4 中描述的方法, 将 SphI 的所有带电荷的残基突变为 Ala。进行共 71 个突变 : 3、5、12、18、21、24、25、30、31、35、43、46、51、54、57、58、60、61、72、75、77、78、87、90、91、95、100、104、107、108、110、113、120、123、124、125、129、130、131、139、140、142、146、147、155、157、159、164、170、172、173、175、178、184、186、190、194、196、197、198、206、207、209、212、215、221、227、230、231、232、235。

[0469] 上面的数字相应于 SphI 蛋白序列 (SEQ ID NO :98) 中的氨基酸位置。

[0470] 3. SphI-HF 的选择

[0471] 每种突变的四个菌落在在 37°C 下,在具有 Amp 和 Cam 的 LB 中生长过夜。活性选择主要针对 NEB2 中的 5% 甘油和 30% 甘油中的 pBR322。使用前面实施例的经验,高保真度 SphI 的选择是简单的。发现 SphI 突变体 D91A、K100A、D139A 和 D164A 显著减少 SphI 中的星号活性。在它们中, K100A 是具有最小星号活性的优选突变。SphI (K100A) 被命名为 SphI-HF。

[0472] 4. SphI-HF 和 WT SphI 的比较

[0473] 在它们各自优选的缓冲液中,并行进行 SphI-HF 和 WT SphI 的比较。使用 NEB 稀释缓冲液 A,将 SphI-HF 2 倍连续稀释,并且在 NEB4 中反应,使用 NEB 稀释缓冲液 B,将 WT SphI 2 倍连续稀释。对 λ DNA 的消化在图 12 中进行比较。

[0474] 表 11 :SphI-HF 和 WT SphI 的 FI 的比较

| 缓冲液 | SphI-HF | | WT SphI | | 改进因子 |
|------|---------|-------------|---------|-------|-----------|
| | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| NEB1 | 50% | ≥ 1000 | 100% | 64 | ≥ 16 |
| NEB2 | 50% | ≥ 1000 | 100% | 32 | ≥ 32 |
| NEB3 | 3% | ≥ 120 | 25% | 64 | ≥ 2 |
| NEB4 | 100% | ≥ 2000 | 50% | 16 | ≥ 64 |

[0476] 在 NEB4 中, SphI-HF 表现最好,其中 $FI \geq 2000$; WT SphI 在 NEB1 或 NEB2 中表现最好,其中优选的 FI 是 64。总 FI 改进因子为 ≥ 32 。

[0477] 实施例 6 :工程化高保真度 PstI

[0478] 1. PstI 的表达

[0479] 从大肠杆菌 (pACYC-HpyCH4VM、pPR594-PstIR) 表达 PstI。HpyCH4VM 保护 PstI 的内部四个碱基。pPR594 是具有 Amp 抗性和 ptac 启动子的表达载体。细胞在 30°C 下,在具有 Amp 和 Cam 的 LB 中生长,然后通过 IPTG 诱导培养物过夜。

[0480] 2. PstI 的诱变

[0481] 使用在先前实施例中描述的方法,将 92 个带电荷的残基突变为 Ala。这些是 :8、10、11、14、25、26、38、40、41、44、45、47、58、61、63、66、67、69、73、74、77、78、82、85、88、91、92、94、95、99、104、105、116、119、127、128、136、142、145、146、150、151、152、156、159、169、170、174、176、179、180、184、188、191、197、202、204、207、212、214、217、218、226、227、228、231、236、237、238、239、240、246、251、257、258、261、263、273、282、284、286、287、295、297、302、305、306、309、313、314、319 和 320。

[0482] 上面的数字相应于 PstI 蛋白序列 (SEQ ID NO :91) 中的氨基酸位置。

[0483] 在用 DpnI 消化 PCR 产物后,将样品转化入感受态大肠杆菌 (pACYC-HpyCH4VM),并且在具有 Amp 和 Cam 的 LB 板上生长。

[0484] 3. PstI-HF 的选择

[0485] PstI-HF 的选择与前面的样品相似。在 5% 甘油下,在 λ DNA 上检测正常活性酶活性,并且在 NEB4 缓冲液和 20% DMSO 的条件下在 pBR322 底物上检测星号活性。DMSO 比

相同浓度的甘油更显著地增强星号活性。在选择期间, #26、#56 和 #65 与 WT 相比, 具有减少的星号活性。当对每一个测序时, 发现突变为 D91A、E204G 和 K228A/A289V。将突变体 #26PstI (D91A) 标记为 PstI-HF。

[0486] 4. PstI-HF 和 WT PstI 的比较

[0487] 在 NEB1-4 缓冲液中, 对 λ DNA 底物, 分别测量 PstI-HF 和 WTPstI 的 FI。稀释缓冲液是 NEB 稀释缓冲液 C。比较如在图 13 中所示, 并且结果在表 12 (下面) 中列出。

[0488] 表 12 :PstI-HF 和 WTPstI 的比较

| 缓冲液 | PstI-HF | | WT PstI | | 改进因子 |
|------|---------|-------------|---------|-------|------------|
| | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| NEB1 | 12.5% | ≥ 250 | 50% | 32 | ≥ 8 |
| NEB2 | 100% | ≥ 2000 | 25% | 16 | ≥ 125 |
| NEB3 | 25% | ≥ 500 | 100% | 120 | ≥ 2 |
| NEB4 | 100% | ≥ 2000 | 50% | 8 | ≥ 250 |

[0490] 在 NEB2 和 NEB4 缓冲液中, PstI-HF 表现最好, 其中优选的 FI 为 ≥ 2000 ; WTPstI 在 NEB 3 缓冲液中表现最好, 其中 FI 为 120。总 FI 改进因子为 $\geq 2000/120 = 16$ 倍。

[0491] 实施例 7 : 工程化高保真度 NcoI

[0492] 1. NcoI 的表达

[0493] 在大肠杆菌 (pSYX20-NcoIM、pRRS-NcoIR) 中实现 NcoI 的表达。pRRS 是 pUC19 衍生质粒, 并且 pSYX20 是与 pRRS 载体相容的低拷贝数质粒。细胞在 30°C 下, 在具有 Amp 和卡那霉素 (Kan) 的 LB 中生长过夜。

[0494] 2. NcoI 的诱变

[0495] 将 NcoI 中所有 66 个带电荷的残基突变为 Ala。这些残基是 :7、8、19、22、27、30、31、32、33、37、39、42、46、55、56、61、62、64、68、69、75、84、88、89、92、93、95、97、100、116、136、144、146、162、166、170、178、183、185、187、188、189、196、199、202、204、209、211、212、213、216、219、227、229、237、241、244、250、251、257、259、261、268、279、282、285。

[0496] 上面的数字相应于 NcoI 蛋白序列 (SEQ ID NO :88) 中的氨基酸位置。

[0497] 方法与在前面实施例中的方法相同, 其使用反向 PCR, 然后进行 DpnI 消化。然后, 将处理的产物转化入大肠杆菌 (pSYX20-NcoIM)。

[0498] 3. NcoI-HF 的选择

[0499] NcoI-HF 的选择与 PstI-HF 的选择相似。如上所述, 在 5% 甘油下, 用 λ DNA 作为底物, 检测活性。使用 pBR322 或 λ 在 19% DMSO 中测定星号活性。发现下面的突变改进星号活性 :A2T/R31A、D56A、H143A、E166A、R212A 和 D268A。在这些突变体中, 选择 NcoI (A2T/R31A) 为 NcoI-HF。

[0500] 4. NcoI-HF 和 WT NcoI 的比较

[0501] 在 NEB1-4 缓冲液中, 对 λ DNA, 分开测定 NcoI-HF 和 WT NcoI 的 FI。比较在图 14 中示出, 并且结果在表 13 (下面) 中列出。

[0502] 表 13 :NcoI-HF 和 WT NcoI 的比较

| [0503] | 缓冲液 | NcoI-HF | | WT NcoI | | 改进因子 |
|--------|------|---------|--------|---------|-------|------|
| | | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| | NEB1 | 25% | ≥4000 | 100% | 120 | ≥32 |
| | NEB2 | 25% | ≥4000 | 100% | 32 | ≥125 |
| | NEB3 | 6.3% | ≥1000 | 25% | 120 | ≥8 |
| | NEB4 | 100% | ≥16000 | 100% | 32 | ≥500 |

[0504] 在 NEB4 中, NcoI-HF 示出最大的星号活性减少, 其中优选的 FI 为 ≥ 16000 ; WT NcoI 在 NEB1、NEB2 和 NEB4 中表现最好, 其中优选的 FI 为 120。总 FI 改进因子为 $\geq 16000/120 = 125$ 。

[0505] 实施例 8: 工程化高保真度 NheI

[0506] 1. NheI 的表达

[0507] NheI 在用 pACYC-NheIM 和 placzz1-NheIR 转化的大肠杆菌中表达。placzz1 是 pUC19 衍生质粒。细胞在 30°C 下, 在具有 Amp 和 Cam 的 LB 中生长过夜。

[0508] 2. NheI 的诱变

[0509] 将 NheI 中所有 92 个带电荷的残基突变为 Ala, 其为下面的残基: 5、6、7、14、17、19、22、25、28、31、38、39、42、47、49、52、56、58、59、60、64、74、75、76、77、80、91、93、104、105、110、112、116、117、123、126、130、131、133、135、137、147、149、152、159、160、165、167、170、171、174、179、183、195、202、205、207、209、210、211、214、216、218、221、225、231、241、243、244、250、252、256、257、259、264、266、267、281、285、287、288、289、291、297、300、307、313、315、318、321、324、325。

[0510] 上面的数字相应于 NheI 蛋白序列 (SEQ ID NO:89) 中的氨基酸位置。

[0511] 方法与在前面实施例中的方法相同, 其使用反向 PCR, 然后进行 DpnI 消化。然后, 将处理的产物转化入大肠杆菌 (pACYC-NheIM)。

[0512] 3. NheI-HF 的选择

[0513] 根据前面的实施例, 进行 NheI-HF 的选择。标准和星号活性分析包含 NEB4 缓冲液以及分别 5% 甘油和 39% 甘油中的 pBR322 作为底物。发现仅一种突变显著改进 NheI。这是 E77A。选择 NheI (E77A) 作为 NheI-HF。

[0514] 4. NheI-HF 和 WT NheI 的比较

[0515] 在 NEB1-4 缓冲液的每一种中, 在 pXba 上, 分开测定 NheI-HF 和 WT NheI 的 FI, pXba 是含有来自腺病毒的 XbaI 消化片段的质粒底物。比较在图 15 中示出, 并且结果在表 14 (下面) 中列出。

[0516] 表 14: NheI-HF 和 WT NheI 的比较

| [0517] | 缓冲液 | NheI-HF | | WT NheI | | 改进因子 |
|--------|------|---------|---------|---------|-------|-------|
| | | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| | NEB1 | 100% | ≥128000 | 100% | 32 | ≥4000 |
| | NEB2 | 3% | ≥4000 | 25% | 120 | ≥32 |
| | NEB3 | 0.05% | ≥32 | 12.5% | 120 | ≥0.25 |
| | NEB4 | 50% | ≥32000 | 100% | 32 | ≥1000 |

[0518] 在 NEB1 缓冲液中, NheI-HF 示出最佳活性, 其中其 FI 为 $\geq 128,000$ 。WT NheI 在 NEB1 和 NEB4 缓冲液具有最大的活性, 其中其最佳 FI 是 32。所以, 总 FI 改进因子为 $\geq 128,000/32 = \geq 4000$ 。

[0519] 实施例 9 : 工程化高保真度 SspI

[0520] 1. SspI 的表达

[0521] 从用 pACYC-SspIM 和 placzz1-SspIR 转化的大肠杆菌表达 SspI。placzz1 是 pUC19 衍生质粒。细胞在 30°C 下, 在具有 Amp 和 Cam 的 LB 中生长过夜。

[0522] 2. SspI 的诱变

[0523] 将 SspI 中所有 81 个带电荷的残基突变为 Ala ; 这些是 : 3、8、12、13、18、19、20、35、40、42、44、47、52、60、62、65、68、69、72、74、76、77、78、79、83、85、88、89、90、96、100、108、109、118、119、127、128、129、131、132、137、144、153、154、155、156、158、165、168、170、172、177、178、179、181、185、186、187、191、194、195、197、202、204、215、222、229、237、240、246、250、256、257、259、260、264、265、267、268、269、274。

[0524] 上面的数字相应于 SspI 蛋白序列 (SEQ ID NO :99) 中的氨基酸位置。

[0525] 方法与在前面实施例中的方法相同, 其使用反向 PCR, 然后进行 DpnI 消化。然后, 将处理的产物转化入大肠杆菌 (pACYC-SspIM)。

[0526] 3. SspI 高保真度突变体的选择

[0527] 在 NEB4 缓冲液以及分别 5% 甘油和 39% 甘油中, 使用 Φ X174 底物, 进行 NheI 的标准同源和星号活性分析。突变体 #16 (H65A)、#20 (K74A)、#23 (E78A)、#26 (E85A)、#28 (E89A)、#33 (K109A)、#34 (E118A)、#52 (R177A)、#62 (K197A)、#67 (D229A) 都示出减少的星号活性。K109A 示出星号活性的最大减少。决定寻求星号活性的进一步改进。

[0528] 4. 进一步突变

[0529] 将最初鉴定为 Tyr 的所有残基突变为 Phe, 而将其他残基 Cys、Phe、Met、Asn、Gln、Ser、Thr 和 Trp 突变为 Ala。该组包括在下列位置处的 95 个残基突变 : 2、6、7、9、10、13、22、25、26、27、29、30、32、33、34、39、41、51、53、55、56、57、58、59、61、63、71、75、81、84、87、91、94、98、104、106、107、110、111、113、114、123、125、134、136、139、140、141、142、143、146、152、157、159、160、164、173、175、180、183、190、192、193、196、198、199、201、205、207、211、214、218、219、220、221、223、225、226、227、228、230、232、233、235、238、239、241、249、254、255、272、275、276、277、280。

[0530] 上面的数字相应于 SspI 蛋白序列 (SEQ ID NO :113) 中的氨基酸位置。

[0531] 通过上面相同的方法, 进行 PCR 和选择。在这些突变体中, 发现 Y98F 具有最少的星号活性, 并且其在这方面比 SspI (K109A) 好。将 SspI (Y98F) 标记为 SspI-HF, 并且作为生产株储存。

[0532] 5. SspI-HF 和 WT SspI 的比较

[0533] 在 NEB1-4 缓冲液中, 使用 λ DNA 底物, 分开测定 SspI-HF 和 WT SspI 的 FI。稀释剂是 NEBC。比较在图 16 中示出, 并且结果在表 15 (下面) 中列出。

[0534] 表 15 : SspI-HF 和 WT SspI 的比较

| 缓冲液 | SspI-HF | | WT SspI | | 改进因子 |
|------|---------|-------|---------|-------|------|
| | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| NEB1 | 50% | ≥4000 | 100% | 64 | ≥64 |
| NEB2 | 50% | 120 | 100% | 16 | ≥8 |
| NEB3 | 0.6% | ≥32 | 25% | 32 | ≥1 |
| NEB4 | 100% | 500 | 100% | 16 | ≥32 |

[0536] 在 NEB4 中, SspI-HF 表现最好, 其中优选的 FI 是 500; WT SspI 在 NEB1、NEB2 和 NEB4 中表现最好, 其中优选的 FI 是 64。总 FI 改进因子为 $500/64 = 8$ 。

[0537] 实施例 10: 工程化高保真度 NotI

[0538] 1. NotI 的表达

[0539] NotI 在 NEB4 缓冲液中具有显著的星号活性, 并且在 NEB3 缓冲液中具有较少的星号活性。工程化 NotI 以在任何 NEB 缓冲液中减少星号活性。NotI 在用 pACYC184-EagIM 和 placzz2-NotIR 转化的感受态大肠杆菌中表达。细胞在 37°C 下, 在具有 Amp 和 Cam 的 LB 中生长过夜。

[0540] 2. NotI 的诱变

[0541] 将 NotI 中所有 97 个带电荷的残基突变为 Ala, 如下列残基: 2、4、8、10、17、21、22、26、31、34、35、36、49、52、57、59、62、72、74、75、77、84、87、96、97、105、117、121、122、125、126、129、130、133、140、141、145、150、152、156、160、165、167、174、176、177、182、187、189、193、194、200、205、208、210、219、224、225、227、236、237、245、251、253、267、271、272、280、283、290、292、294、296、304、306、308、310、314、319、321、323、327、331、335、336、339、353、354、356、358、361、365、367、368、369、370、378、382。

[0542] 上面的数字相应于 NotI 蛋白序列 (SEQ ID NO:90) 中的氨基酸位置。

[0543] 将突变体引入酶的方法与在前面实施例中的方法相同, 其使用反向 PCR, 然后为 DpnI 消化。然后, 将处理的产物转化入含有 pACYC-EagIM 的大肠杆菌。

[0544] 3. NotI-HF 的选择

[0545] 如在前面实施例中描述的, 进行 NotI-HF 的选择。标准同源和星号活性分析使用 pXba 底物, 在 NEB4 缓冲液以及分别 5% 甘油和 NEB ExoI 缓冲液 (67mM 甘氨酸 -KOH, pH 9.5, 6.7mM MgCl₂, 10mM 2-巯基乙醇) 以及 37% 甘油中进行。#37 (K150A)、#44 (K176A)、#45 (R177A)、#63 (R253A) 都示出减少的星号活性。K150A 是优选的减少星号活性的突变。选择 NotI (K150A) 为 NotI-HF。

[0546] 4. NotI-HF 和 WT NotI 的比较

[0547] 在 NEB1-4 缓冲液中, 使用 pXba 底物, 分开测定 NotI-HF 和 WT NotI 的 FI。比较在图 17 中示出, 并且结果在表 16 (下面) 中列出。

[0548] 表 16: NotI-HF 和 WT NotI 的比较

| 缓冲液 | NotI-HF | | WT NotI | | 改进因子 |
|-------------|---------|---------|---------|-------|-------|
| | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| [0549] NEB1 | 6% | ≥8000 | 6% | ≥8000 | ND |
| NEB2 | 100% | ≥128000 | 50% | 250 | ≥512 |
| NEB3 | 1.6% | ≥4000 | 100% | 4000 | ≥1 |
| NEB4 | 50% | ≥64000 | 12.5% | 32 | ≥2000 |

[0550] ND:不可检测,因为两个 FI 都是超出限度的不确定数。

[0551] 在 NEB2 中,NotI-HF 表现最好,其中优选的 FI 是 ≥ 128000 ;WT NotI 在 NEB3 中表现最好,其中优选的 FI 是 4000。总保真度指数改进因子为 $\geq 128000/4000 = \geq 32$ 。工程化 NotI 不但进一步提高 NotI 的 FI,而且改变最佳缓冲液。

[0552] 实施例 11:工程化高保真度 SacI

[0553] 1. SacI 的表达

[0554] SacI 在用 pLG-SacIM 和 pRRS-SacIR 转化的大肠杆菌中表达。pRRS 是 pUC19 衍生质粒,pLG 是低拷贝相容性质粒。细胞在 30°C 下,在具有 Amp 和 Kan 的 LB 中生长过夜。

[0555] 2. SacI 的诱变

[0556] 将 SacI 中所有 101 个带电荷的残基突变为 Ala,如下列残基:6、7、11、15、16、19、24、25、29、30、39、40、42、45、58、61、62、63、65、67、70、71、72、74、75、76、81、85、94、98、104、105、114、116、120、123、127、129、133、134、141、143、144、145、146、150、151、154、169、170、172、181、187、196、197、200、201、211、216、220、221、224、227、228、232、238、240、246、248、250、258、270、271、277、281、288、289、295、296、297、299、303、306、313、314、321、322、324、332、336、337、340、342、344、345、347、349、350、353、357。

[0557] 上面的数字相应于 SacI 蛋白序列 (SEQ IDNO:93) 中的氨基酸位置。

[0558] 方法与在前面实施例中的方法相同,其使用反向 PCR,然后为 DpnI 消化。然后,将处理的产物转化入大肠杆菌 (pLG-SacIM)。

[0559] 3. SacI-HF 的选择

[0560] 使用与前面实施例相似的方法,完成 SacI-HF 的选择。标准活性检查使用 pUC19、在 NEB4 中、具有 5% 甘油,并且星号活性检查在 39% 甘油下、在 NEB4 缓冲液中对 pUC19 进行。#52 SacI (Q117H/R154A/L284P) 和 #60 SacI (Q117H/R200A) 都具有减少的星号活性,和 SacI Q117H/R200A 证明是优选的突变。Q117H 是来自模板的遗留突变,其不影响 SacI 的活性。选择 SacI (Q117H/R200A) 为 SacI-HF。

[0561] 4. SacI-HF 和 WT SacI 的比较

[0562] 在 NEB1-4 缓冲液中,对 pXba 底物,分开测定 SacI-HF 和 WT SacI 的 FI。比较在图 18 中示出,并且结果在表 17(下面)中列出。

[0563] 表 17: SacI-HF 和 WT SacI 的比较

| 缓冲液 | SacI-HF | | WT SacI | | 改进因子 |
|------|---------|-------|---------|-------|------|
| | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| NEB1 | 25% | ≥2000 | 100% | 120 | ≥16 |
| NEB2 | 12.5% | ≥120 | 50% | 120 | ≥1 |
| NEB3 | 0.8% | ≥120 | 3% | 120 | ≥1 |
| NEB4 | 100% | 4000 | 100% | 32 | 120 |

[0565] 在 NEB4 中, SacI-HF 表现最好, 其中 FI 是 4000 ;WT SacI 在 NEB1 和 NEB4 中表现最好, 其中优选的 FI 是 120。总 FI 改进因子为 $4000/120 = 32$ 。

[0566] 实施例 12 : 工程化高保真度 PvuII

[0567] 1. PvuII 的表达

[0568] PvuII 在用 pACYC-PvuIIM 和 placzz2-PvuIIR 转化的大肠杆菌中表达。Placzz2 是 pUC19 衍生质粒 ;pACYC 是低拷贝相容性质粒。细胞在 30°C 下, 在具有 Amp 和 Cam 的 LB 中生长过夜。

[0569] 2. PvuII 的诱变

[0570] 将 PvuII 中所有 47 个带电荷的残基突变为 Ala, 如下列残基 :3、5、8、11、15、18、21、25、26、30、34、38、54、55、58、61、66、68、70、75、78、83、84、85、93、95、105、110、114、116、118、119、121、125、126、128、129、130、134、136、137、138、143、147、151、152 和 155。

[0571] 上面的数字相应于 PvuII 蛋白序列 (SEQ ID NO :92) 中的氨基酸位置。

[0572] 方法与在前面实施例中的方法相同, 其使用反向 PCR, 然后为 DpnI 消化。然后, 将处理的产物转化入大肠杆菌 (pACYC-PvuIIM)。

[0573] 3. PvuII 高保真度突变体的选择

[0574] PvuII-HF 的选择与前面的实施例相似。标准活性检查使用 λ DNA 底物、在 NEB4 中、具有 5% 甘油, 并且星号活性检查在 39% 甘油下、在 NEB4 缓冲液中对 pBR322 进行。没有突变体具有高保真度 PvuII 的资格。

[0575] 4. 另外的诱变步骤

[0576] 另外的诱变步骤是在 PvuII 中将所有 Ser、Thr 突变为 Ala, 将 Tyr 突变为 Phe。突变的位置是 :2、19、46、49、67、71、77、81、82、94、104、113、123、124、132、133、148、154 和 157。

[0577] 方法与在前面实施例中的方法相同, 其使用反向 PCR, 然后为 DpnI 消化。然后, 将处理的产物转化入大肠杆菌 (pACYC-PvuIIM)。

[0578] PvuII (T46A) 表现具有比 WT PvuII 更少的星号活性, 然而, 期望进一步改进。

[0579] 将 T46 突变为所有其他氨基酸残基, 这通过将密码子改变为相应氨基酸进行。在所有这些突变中, T46H、T46K、T46Y、T46G 都比 T46A 更好。选择 T46G 作为 PvuII-HF。

[0580] 5. PvuII-HF 和 WT PvuII 的比较

[0581] 在 NEB1-4 缓冲液中, 使用稀释剂 A, 分开测定 PvuII-HF 和 WT PvuII 对 pBR322 的 FI。比较在图 19 中示出, 并且结果在表 18 (下面) 中列出。

[0582] 表 18 :PvuII-HF 和 WT PvuII 的比较

| 缓冲液 | PvuII-HF | | WT PvuII | | 改进因子 |
|------|----------|--------|----------|-------|------|
| | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| NEB1 | 3.1% | ≥ 250 | 100% | 250 | ≥1 |
| NEB2 | 12.5% | ≥ 1000 | 25% | 16 | ≥64 |
| NEB3 | 0.4% | ≥ 32 | 3.1% | 8 | ≥4 |
| NEB4 | 100% | 500 | 100% | 1/4 | 2000 |

[0584] 在 NEB4 中, PvuII-HF 表现最好, 其中 FI 是 500 ; WTPvuII 在 NEB1 和 NEB4 中表现最好, 其中优选的 FI 是 250。总 FI 改进因子为 $500/250 = 2$ 。尽管总 FI 改进因子对 PvuII 不高, 但是在 NEB4 中 FI 提高 2000 倍。

[0585] 实施例 13 : 工程化高保真度 MfeI

[0586] 1. MfeI 的表达

[0587] MfeI 在用 pACYC-MluCIM 和 pRRS-MfeIR 转化的大肠杆菌中表达。pRRS 是 pUC19 衍生质粒 ; pACYC 是低拷贝相容性质粒。MluCIM 甲基化 AATT, 其是 MfeI 的内部四个核酸序列。细胞在 37°C 下, 在具有 Amp 和 Cam 的 LB 中生长过夜。

[0588] 2. MfeI 的诱变

[0589] 分三批进行 MfeI 的诱变。第一批是所有的带电荷的残基, 将其突变为 Ala, 如下列氨基酸位置 : 5、9、19、24、36、39、44、45、47、48、50、60、61、64、65、72、83、87、90、92、93、98、100、101、103、107、109、110、115、119、120、121、124、132、135、142、143、144、153、155、158、159、161、162、164、165、171、172、175、181、184、187、188、192、195、196、198、199、200 ; 第二批是具有羟基的所有残基 : Ser、Thr 和 Tyr, 其中 Ser 和 Thr 被改变为 Ala, 并且 Tyr 被改变为 Phe。所述残基是在 : 4、7、21、28、38、40、43、53、74、75、76、81、89、91、112、122、127、134、136、157、167、170、173、177、185 和 200。第三批是残基 Cys、Phe、Met、Asn、Gln、Trp, 都改变成 Ala, 所述残基是在 : 10、12、13、25、26、29、31、32、35、51、55、67、68、77、78、84、88、96、102、105、117、123、126、141、148、149、152、168、169、174、176、178、179、180、183、191、193、194。

[0590] 上面的数字相应于 MfeI 蛋白序列 (SEQ ID NO :5) 中的氨基酸位置。

[0591] 方法与在前面实施例中的方法相同, 其使用反向 PCR, 然后为 DpnI 消化。然后, 将处理的产物转化入大肠杆菌 (pACYC-MluCIM)。

[0592] 3. MfeI-HF 的选择

[0593] 使用与前面实施例相似的方法, 完成 MfeI-HF 的选择。在 NEB4 中, 应用 5% 甘油、使用 Φ X174 底物测定切割活性, 并且在 NEB4 中, 应用 39% 甘油、使用 Φ X174 底物测定星号活性。该酶的显著困难是许多突变改进了酶的切割活性且星号活性减少, 但是比 WT 酶需要更高的甘油浓度。MfeI (K50A) 是一个实例, 其在高浓度甘油中具有减少的星号活性和高切割活性, 而在较低甘油浓度中, 所述活性是低的。MfeI (Y173A) 也减少星号活性。优选的突变为 Q13A/F35Y。F35Y 的突变来自模板, 并且 Q13A 是靶向突变。将 MfeI (Q13A/F35Y) 标记为 MfeI-HF。

[0594] 4. MfeI-HF 和 WT MfeI 的比较

[0595] 在 NEB1-4 缓冲液中, 对 λ DNA 底物, 分开测定 MfeI-HF 和 WT MfeI 的 FI, 其中在

NEB 稀释剂 A 中进行稀释。比较在图 20 中示出,并且结果在表 19(下面)中列出。

[0596] 表 19 :MfeI-HF 和 WT MfeI 的比较

| 缓冲液 | MfeI-HF | | WT MfeI | | 改进因子 |
|------|---------|--------|---------|-------|------|
| | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| NEB1 | 100% | ≥ 1000 | 100% | 32 | ≥32 |
| NEB2 | 25% | ≥ 250 | 12.5% | 16 | ≥16 |
| NEB3 | 1.3% | ≥ 16 | 6.3% | 8 | ≥2 |
| NEB4 | 100% | ≥ 500 | 100% | 32 | ≥16 |

[0598] 在 NEB1 和 NEB4 中,MfeI-HF 表现最好,其中优选的 FI 为 ≥ 1000 ;在 NEB1 和 NEB4 中,WT MfeI 表现最好,其中优选的 FI 为 32。总 FI 改进因子是 $\geq 1000/32 = 32$ 倍。

[0599] 实施例 14:工程化高保真度 HindIII

[0600] 1. HindIII 的表达

[0601] HindIII 在用 pUC19-HindIII_{IRM} 转化的大肠杆菌中表达,所述 pUC19-HindIII_{IRM} 含有 HindIII 内切核酸酶和甲基化酶基因。细胞在 30°C 下,在具有 Amp 的 LB 中生长过夜。

[0602] 2. HindIII 的诱变

[0603] 将 HindIII 中 88 个带电荷的残基突变为 Ala。这些是:2、3、7、8、14、20、22、34、37、39、42、45、52、55、61、62、66、69、74、84、87、89、94、100、101、109、111、114、117、120、123、124、126、128、132、134、135、136、137、138、153、158、162、163、171、172、180、182、183、190、197、198、201、202、207、209、214、215、218、222、225、227、228、229、237、238、243、244、245、249、250、251、254、255、261、265、266、267、270、274、275、281、283、286、290、293、296、297。

[0604] 在位置 4、11、15、17、18、19、21、23、26、27、30、31、36、38、46、57、58、59、60、63、64、76、77、80、82、83、88、91、99、102、103、104、112、113、116、118、121、122、125、131、133、139、143、146、147、148、149、151、152、154、155、157、159、160、164、168、169、170、178、184、185、187、188、189、191、193、194、195、199、200、203、204、206、210、211、212、213、216、217、219、220、221、224、230、232、233、236、240、241、246、252、253、256、258、262、263、264、277、278、279、280、284、287、288、294、295、299 处,将所有残基 Cys、Met、Asn、Gln、Ser、Thr、Trp 改变为 Ala,而将 Tyr 改变为 Phe。

[0605] 上面的数字相应于 HindIII 蛋白序列 (SEQ ID NO :85) 中的氨基酸位置。

[0606] 方法与在前面实施例中的方法相同,其使用反向 PCR,然后为 DpnI 消化。然后,将处理的产物转化入大肠杆菌菌株 ER3081。

[0607] 3. HindIII-HF 的选择

[0608] 使用与前面的实施例相似的方法,完成 HindIII-HF 的选择。标准活性检查使用 λ DNA、在 NEB4 中并具有 5% 甘油,并且星号活性在具有 39% 甘油的 NEB4 缓冲液中使用 λ DNA 底物测量。发现 HindIII 的两个突变体具有减少的星号活性。这些是 HindIII (K198A) 和 S188P/E190A。将 HindIII (K198A) 标记为 HindIII-HF。

[0609] 4. HindIII-HF 和 WT HindIII 的比较

[0610] 在 NEB1-4 缓冲液中的每一种中,使用 λ DNA 底物,用稀释剂 B,分开测定 HindIII-HF 和 WT HindIII 的 FI。比较在图 21 中示出,并且结果在表 20(下面)中列出。

[0611] 表 20 :HindIII-HF 和 WT HindIII 的比较

| 缓冲液 | HindIII-HF | | WT HindIII | | 改进因子 |
|------|------------|---------|------------|-------|-------|
| | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| NEB1 | 25% | ≥ 16000 | 25% | 32 | ≥500 |
| NEB2 | 100% | ≥ 64000 | 100% | 250 | ≥250 |
| NEB3 | 12.5% | ≥ 16000 | 25% | 4000 | ≥4 |
| NEB4 | 50% | ≥ 32000 | 50% | 32 | ≥1000 |

[0612] 在 NEB2 中, HindIII-HF 表现最好, 其中优选 FI 是 ≥ 64000 ; WT HindIII 在 NEB2 中表现最好, 其中优选的 FI 是 250。总 FI 改进因子是 $4000/120 = 32$ 。

[0613] 实施例 15 : 工程化高保真度 SbfI

[0614] 1. SbfI 的表达

[0615] SbfI 在用 pUC19-SbfIRM 转化的大肠杆菌中表达。细胞在 30℃ 下, 在具有 Amp 的 LB 中生长过夜。

[0616] 2. SbfI 的诱变

[0617] 将 SbfI 中 78 个带电荷的残基突变为 Ala。这些残基是 :5、8、15、18、23、27、30、34、46、49、50、53、58、63、66、70、71、74、81、82、83、85、86、87、90、94、103、115、120、121、127、132、135、136、143、144、147、150、152、154、164、169、170、183、184、187、188、192、196、204、206、208、213、214、215、218、219、226、228、230、233、237、238、239、241、248、251、253、257、258、259、260、262、266、282、284、285、288、293、297、299、304、305、307、311、316 和 322。

[0618] 也将 SbfI 中的 Ser 和 Thr 残基突变为 Ala。将 Tyr 突变为 Phe。靶向下列位置 : 3、4、5、10、13、16、31、35、38、54、55、56、68、76、78、80、88、109、111、116、119、129、131、137、146、162、174、197、198、201、205、210、224、252、263、270、272、286、298、315、321。

[0619] 也将下列位置的另外 55 个 Cys、Phe、Met、Asn、Gln、Trp 残基突变为 Ala :2、24、26、29、32、51、62、65、67、72、84、91、92、95、97、101、104、106、110、112、114、117、124、134、140、157、160、171、178、179、185、189、193、212、217、225、231、243、245、247、256、265、268、277、279、280、281、283、287、289、290、296、301、313 和 317。

[0620] 上面的数字相应于 SbfI 蛋白序列 (SEQ ID NO :96) 中的氨基酸位置。

[0621] 方法与在前面实施例中的方法相同, 其使用反向 PCR, 然后为 DpnI 消化。然后, 将突变的产物转化入大肠杆菌菌株 ER2984。

[0622] 3. SbfI-HF 的选择

[0623] 如在前面的实施例中描述的, 完成 SbfI-HF 的选择。标准活性检查使用 λ DNA、在 NEB4 中并具有 5% 甘油, 并且星号活性检查针对 λ DNA 在核酸外切酶 I 缓冲液中进行。将 SbfI (K251A) 标记为 SbfI-HF。

[0624] 4. SbfI-HF 和 WT SbfI 的比较

[0625] 在 NEB1-4 缓冲液中, 对 λ DNA, 用稀释剂 C, 分开测定 SbfI-HF 和 WT SbfI 的 FI。比较在图 22 中示出, 并且结果在表 21 (下面) 中列出。

[0626] 表 21 :SbfI-HF 和 WT SbfI 的比较

| 缓冲液 | SbfI-HF | | WT SbfI | | 改进因子 |
|------|---------|-------|---------|-------|------|
| | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| NEB1 | 100% | 1000 | 100% | 16 | 64 |
| NEB2 | 50% | 120 | 50% | 32 | 4 |
| NEB3 | 3.5% | 8 | 25% | 120 | 1/16 |
| NEB4 | 100% | 250 | 12.5% | 8 | 32 |

[0629] 在 NEB1 和 NEB4 中, SbfI-HF 表现最好, 其中优选 FI 是 1000 ; WT SbfI 在 NEB1 中表现最好, 其中优选的 FI 是 8。总 FI 改进因子是 $1000/8 = 125$ 倍。

[0630] 实施例 16 : 工程化高保真度 EagI

[0631] 1. EagI 的表达

[0632] EagI 在用 pBR322-EagIRM 转化的大肠杆菌中表达。细胞在 30°C 下, 在具有 20 μ g/ml 四环素的 LB 中生长过夜。

[0633] 2. EagI 的诱变

[0634] 将 Asp、Glu、His、Lys、Arg、Ser、Thr、Asn 和 Gln 残基突变为 Ala。将 Tyr 突变为 Phe。这些是下列残基 : 2、3、4、5、6、9、13、14、17、19、21、23、27、35、36、37、40、42、43、44、45、46、49、51、53、55、56、58、60、66、67、69、71、72、73、74、75、77、78、80、82、86、87、92、93、94、95、98、99、100、102、103、104、105、112、113、114、116、117、119、122、125、127、132、134、135、137、139、140、141、145、147、148、150、152、154、155、156、157、160、162、163、164、166、169、172、173、176、177、178、179、182、185、187、188、189、193、196、197、201、202、203、204、205、206、208、209、212、217、220、221、222、224、225、230、235、236、237、238、239、240、241、243、245、246、247、248、251、255、257、258、259、260、263、264、265、266、270、272、273、275、276、277、279、280、283、286、288、289、291、295、296。

[0635] 上面的数字相应于 EagI 蛋白序列 (SEQ ID NO :82) 中的氨基酸位置。

[0636] 方法与在前面实施例中的方法相同, 其使用反向 PCR, 然后为 DpnI 消化。然后, 将处理的产物转化入大肠杆菌菌株 ER 3081 中, 并且在具有四环素的 LB 琼脂平板上生长。

[0637] 3. EagI-HF 的选择

[0638] 使用与前面实施例不同的方法, 完成 EagI-HF 的选择, 所述方法使用高浓度甘油、高浓度 DMSO 或高 pH。在粗提物中, 因为表达太低而不能示出星号活性, 纯化每一突变体以检查星号活性会是非常冗长的。根据前面的实施例, 推断与 NEB3 相比, 在 NEB4 中 HF 内切核酸酶倾向具有增加的切割活性。因此, 粗提物中 EagI 的活性在 NEB3 和 NEB4 中都测量; 选择具有最高 NEB4/NEB3 比率的一个。将 EagI (H43A) 标记为 EagI-HF。

[0639] 4. EagI-HF 和 WT EagI 的比较

[0640] 在 NEB1-4 缓冲液的每一个中, 对 pXba 底物, 分开测定 EagI-HF 和 WT EagI 的 FI。比较在图 23 中示出, 并且结果在表 22 (下面) 中列出。

[0641] 表 22 : EagI-HF 和 WT EagI 的比较

| [0642] | 缓冲液 | EagI-HF | | WT EagI | | 改进因子 |
|--------|------|---------|-------|---------|-------|------|
| | | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| | NEB1 | 50% | 250 | 25% | 4 | 64 |
| | NEB2 | 100% | 250 | 50% | 8 | 32 |
| | NEB3 | 50% | 250 | 100% | 250 | 1 |
| | NEB4 | 100% | 500 | 100% | 16 | 16 |

[0643] 在 NEB2 和 NEB4 中, EagI-HF 表现最好, 其中优选 FI 是 500 ;WT EagI 在 NEB3 和 NEB4 中表现最好, 其中优选的 FI 是 250。总 FI 改进因子是 $500/250 = 2$ 。

[0644] 实施例 17 : 工程化高保真度 EcoRV

[0645] 1. EcoRV 的表达

[0646] EcoRV 在用 pACYC-EcoRVM 和 placzz1-EcoRV 转化的大肠杆菌中表达。Placzz1 是 pUC19 衍生质粒, 并且 pACYC 是低拷贝相容性质粒。细胞在 37°C 下, 在具有 Amp 和 Cam 的 LB 中生长过夜。

[0647] 2. EcoRV 的诱变

[0648] 将 Cys、Asp、Glu、Phe、His、Lys、Met、Asn、Gln、Arg、Ser、Thr 和 Trp 残基改变为 Ala。将 Tyr 改变为 Phe。这些是 : 2、4、5、6、9、12、13、14、15、16、17、18、19、21、25、27、29、31、35、36、37、38、41、42、44、45、47、48、49、53、54、57、58、59、61、64、65、67、68、69、70、71、72、74、75、76、78、79、81、82、84、85、86、90、92、93、94、95、96、97、98、99、100、101、102、104、105、106、110、111、112、113、114、115、116、117、118、119、120、123、125、126、127、128、131、132、136、138、139、140、143、144、145、146、147、149、150、151、152、154、155、157、158、161、163、164、167、169、171、172、173、174、179、183、185、186、187、188、191、193、195、196、197、198、199、201、203、206、207、208、209、210、211、212、214、215、216、217、218、219、220、221、222、223、224、226、227、228、229、230、231、232、234、235、236、237、238、239、241、242、244 和 245。

[0649] 上面的数字相应于 EcoRV 蛋白序列 (SEQ ID NO :84) 中的氨基酸位置。

[0650] 方法与在前面实施例中的方法相同, 其使用反向 PCR, 然后为 DpnI 消化。然后, 将处理的产物转化入大肠杆菌 (pACYC-EcoRVM)。

[0651] 3. EcoRV-HF 的选择

[0652] 使用与前面实施例相似的方法, 完成 EcoRV-HF 的选择。标准活性检查在 NEB4 中、应用 5% 甘油使用 pXba 进行, 并且星号活性检查针对 pXba、应用 39% 甘油在核酸外切酶 I 缓冲液中进行。发现 EcoRV (D19A/E27A) 与 WT EcoRV 相比具有减少的星号活性。将该突变体标记为 EcoRV-HF。对于该突变体, E27A 是靶向突变, 并且 D19A 是自发突变。双突变体比 D19A 和 E27A 单突变体在星号活性方面具有更大的减少。

[0653] 4. EcoRV-HF 和 WT EcoRV 的比较

[0654] 在 NEB1-4 缓冲液中的每一种中, 对 pXba 底物, 分开测定 EcoRV-HF 和 WT EcoRV 的 FI。比较在图 24 中示出, 并且结果在表 23 (下面) 中列出。

[0655] 表 23 : EcoRV-HF 和 WT EcoRV 的比较

| 缓冲液 | EcoRV-HF | | WT EcoRV | | 改进因子 |
|-------------|----------|--------|----------|-------|-------|
| | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| [0656] NEB1 | 25% | ≥16000 | 6.3% | 32 | ≥500 |
| NEB2 | 100% | ≥64000 | 50% | 120 | ≥500 |
| NEB3 | 50% | ≥32000 | 100% | 1000 | ≥32 |
| NEB4 | 100% | ≥64000 | 25% | 64 | ≥1000 |

[0657] 在 NEB2 和 NEB4 中, EcoRV-HF 表现最好, 其中优选 FI 是 ≥ 64000 ; WT EcoRV 在 NEB3 中表现最好, 其中优选的 FI 是 1000。总 FI 改进因子为 $\geq 64000/1000 = 64$ 。

[0658] 实施例 18: 工程化高保真度 AvrII

[0659] 1. AvrII 的表达

[0660] AvrII 在用 pUC19-AvrIIRM 转化的大肠杆菌中表达。细胞在 30℃ 下, 在具有 Amp 的 LB 中生长过夜。

[0661] 2. AvrII 的诱变

[0662] 将 Cys、Asp、Glu、Phe、His、Lys、Met、Asn、Gln、Arg、Ser、Thr 和 Trp 残基突变为 Ala。将 Tyr 改变为 Phe。这些是: 2、3、4、6、8、9、10、12、15、17、19、20、22、23、27、29、30、31、32、34、36、40、41、42、43、44、46、47、48、50、51、53、55、56、57、58、59、60、65、68、70、72、74、75、76、77、79、80、82、83、84、86、87、88、94、95、96、97、100、104、105、106、107、108、110、112、113、116、117、119、120、121、122、123、124、126、127、129、130、131、132、134、136、139、142、143、144、145、150、151、152、153、154、156、157、158、161、163、164、165、166、168、169、173、174、177、178、181、182、184、186、187、188、189、190、191、192、195、198、200、202、206、207、211、215、216、220、223、224、226、229、230、231、232、233、234、235、236、237、239、243、244、245、246、248、249、253、255、256、260、262、264、265、266、267、268、269、270、272、274、276、277、278、279、280、281、284、285、286、288、289、290、291、299、302、303、304、305、306、308、310、312、314、315、316、318、321、322、324、325、328、331、333、335、337、338、339、340、342、343、346、347、348、350、351、353、354、355、356、358。

[0663] 上面的数字相应于 AvrII 蛋白序列 (SEQ ID NO :80) 中的氨基酸位置。

[0664] 方法与在前面实施例中的方法相同, 其使用反向 PCR, 然后为 DpnI 消化。然后, 将处理的产物转化入大肠杆菌菌株 ER2984 中。

[0665] 3. AvrII-HF 的选择

[0666] 使用与前面实施例相似的方法, 完成 AvrII-HF 的选择。切割活性在 NEB4 中应用 5% 甘油使用 pBC4 测定, 并且星号活性在具有 39% 甘油的 ExoI 缓冲液中应用 pBC4 测定。突变体 #16 (M29A)、#57 (E96A)、#60 (Y104F)、#62 (K106A)、#154 (S127A)、#170 (F142A) 都示出改进。将 AvrII (Y104F) 标记为 AvrII-HF。

[0667] 4. AvrII-HF 和 WTAvrII 的比较

[0668] 在 NEB1-4 缓冲液中的每一种中, 对 T7DNA 底物, 使用稀释剂 B, 分开测定 AvrII-HF 和 WT AvrII 的 FI。比较在图 25 中示出, 并且结果在表 24 (下面) 中列出。

[0669] 表 24 :AvrII-HF 和 WT AvrII 的比较

| [0670] | 缓冲液 | AvrII-HF | | WT AvrII | | 改进因子 |
|--------|------|----------|-------|----------|-------|------|
| | | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| | NEB1 | 100% | 500 | 100% | 64 | ≥8 |
| | NEB2 | 50% | ≥500 | 100% | 8 | ≥64 |
| | NEB3 | 3.1% | ≥16 | 25% | 32 | ≥0.5 |
| | NEB4 | 100% | 1000 | 100% | 32 | 32 |

[0671] 在 NEB1 和 NEB4 中, AvrII-HF 表现最好, 其中优选 FI 是 1000 ;WT AvrII 在 NEB1 和 NEB4 中表现最好, 其中优选的 FI 是 64。总 FI 改进因子是 $1000/64 = 16$ 。

[0672] 实施例 19 : 工程化高保真度 BstXI

[0673] 1. BstXI 的表达

[0674] BstXI 在用 pACYCBstXIMS 和 pUC19-BstXIR 转化的大肠杆菌中表达。pACYC 是低拷贝相容性质粒。BstXI 必须具有甲基化酶基因和特异性基因以具有甲基化酶功能。细胞在 37°C 下, 在具有 Amp 和 Cam 的 LB 中生长过夜。

[0675] 2. BstXI 的诱变

[0676] 如下, 在 BstXI 中, 进行 237 个氨基酸突变。将 Cys、Asp、Glu、Phe、His、Lys、Met、Asn、Gln、Arg、Ser、Thr、Trp 突变为 Ala。将 Try 突变为 Phe。这些是 :4、6、7、9、11、12、14、15、17、18、20、21、22、23、24、26、27、29、30、31、32、33、34、35、36、37、39、40、42、43、46、48、50、53、54、57、58、59、60、62、63、64、65、66、71、72、73、75、76、78、80、81、82、83、84、86、89、91、93、94、95、96、97、98、103、105、106、108、110、111、112、114、117、118、120、123、124、125、126、127、128、129、130、131、137、138、139、141、142、144、145、146、148、151、152、153、154、155、156、159、162、163、166、168、169、171、172、173、174、175、176、177、178、182、185、188、189、191、193、194、195、196、198、199、201、204、208、209、210、211、212、214、215、216、217、218、219、220、221、223、228、229、230、233、235、236、238、239、240、244、245、248、249、250、253、254、255、258、259、260、261、263、264、265、267、268、269、272、276、277、278、279、280、282、285、286、287、288、289、291、293、294、295、300、301、302、304、305、306、308、309、312、314、317、318、319、320、323、324、325、326、330、331、333、334、335、337、343、344、345、346、347、349、353、355、356、357、358、359、360、362、363、364、365、367、369、371、373、374、376、377、378、379、380、381、382 和 383。

[0677] 上面的数字相应于 BstXI 蛋白序列 (SEQ ID NO :7) 中的氨基酸位置。

[0678] 方法与在前面实施例中的方法相同, 其使用反向 PCR, 然后为 DpnI 消化。然后, 将处理的产物转化入大肠杆菌 (pACYC-BstXIMS)。

[0679] 3. BstXI-HF 的选择

[0680] 使用与前面实施例相似的方法, 完成 BstXI-HF 的选择。切割活性在 NEB4 中应用 5% 甘油使用 pBC4 测定, 并且星号活性在具有 39% 甘油的 NEB4 缓冲液中使用 pBC4DNA 底物测定。突变体 #36 (Y57F)、#44 (N65A)、#48 (E75A)、#49 (N76A) 和 #124 (K199A) 都具有减少的星号活性。将 BstXI (N65A) 标记为 BstXI-HF。

[0681] 4. BstXI-HF 和 WT BstXI 的比较

[0682] 在 NEB1-4 缓冲液中的每一种中, 对 λ DNA 底物, 使用稀释剂 A, 分开测定 BstXI-HF

和 WT BstXI 的 FI。比较在图 26 中示出,并且结果在表 25(下面)中列出。

[0683] 表 25 :BstXI-HF 和 WT BstXI 的比较

| 缓冲液 | BstXI-HF | | WT BstXI | | 改进因子 |
|------|----------|-------|----------|-------|------|
| | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| NEB1 | 50% | ≥120 | 6.3% | 4 | ≥32 |
| NEB2 | 100% | ≥250 | 100% | 32 | ≥8 |
| NEB3 | 6.3% | ≥16 | 100% | 2 | ≥8 |
| NEB4 | 100% | ≥250 | 100% | 32 | ≥32 |

[0685] 在 NEB2 和 NEB4 中,BstXI-HF 表现最好,其中优选 FI 是 ≥ 250 ;WT BstXI 在 NEB2、NEB3 和 NEB4 中表现最好,其中优选的 FI 是 32。总 FI 改进因子为 $\geq 250/32 = 8$ 。

[0686] 实施例 20 :工程化高保真度 PciI

[0687] 1. PciI 的表达

[0688] PciI 在用 pACYC-PciIM 和 placzz1-PciIR 转化的大肠杆菌中表达。placzz1 是 pUC19 衍生质粒,pACYC 是低拷贝相容性质粒。细胞在 37°C 下,在具有 Amp 和 Cam 的 LB 中生长过夜。

[0689] 2. PciI 的诱变

[0690] 利用 Cys、Asp、Glu、Phe、His、Lys、Met、Asn、Gln、Arg、Ser、Thr 突变 PciI 中 151 个氨基酸残基。将 Trp 改变为 Ala,和将 Tyr 改变为 Phe。这些是 :2、3、4、6、8、9、10、11、12、14、17、18、19、21、24、25、26、28、29、30、31、33、34、35、36、38、39、41、44、46、47、49、50、51、54、55、56、58、59、60、63、67、68、69、71、74、75、78、80、81、82、85、86、91、92、95、97、98、101、103、104、107、109、113、114、115、118、119、120、121、122、124、126、127、129、130、131、132、133、135、136、137、138、143、145、146、147、148、149、151、152、153、154、155、157、158、159、161、164、165、167、172、175、178、179、180、182、184、185、186、190、192、193、196、197、198、199、200、202、203、206、207、209、210、215、218、221、222、228、229、230、231、232、233、234、235、237、238、239、241、242、243、244、246、247、248、253、254、255、256。

[0691] 上面的数字相应于 PciI 蛋白序列 (SEQ ID NO :15) 中的氨基酸位置。

[0692] 方法与在前面实施例中的方法相同,其使用反向 PCR,然后为 DpnI 消化。然后,将处理的产物转化入大肠杆菌 (pACYC-PciIM)。

[0693] 3. PciI-HF 的选择

[0694] 使用与前面实施例相似的方法,完成 PciI-HF 的选择。切割活性在 NEB4 中应用 5% 甘油使用 SalI- 切割 pBR322 测定,并且星号活性在具有 39% 甘油的 ExoI 缓冲液中使用 SalI- 切割 pBR322 测定。双突变体 PciI (E78A/S133A) 具有减少的星号活性和强的切割活性。该突变体不是上面描述的靶向突变之一,而是偶然随机事件。

[0695] 4. PciI-HF 和 WT PciI 的比较

[0696] 在 NEB1-4 缓冲液中的每一种中,对 pXba 底物,使用稀释剂 A,分开测定 PciI-HF 和 WT PciI 的 FI。比较在图 27 中示出,并且结果在表 26(下面)中列出。

[0697] 表 26 :PciI-HF 和 WT PciI 的比较

| [0698] | 缓冲液 | PciI-HF | | WT PciI | | 改进因子 |
|--------|------|---------|-------|---------|-------|------|
| | | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| | NEB1 | NC | NC | 50% | 2000 | N/A |
| | NEB2 | 100% | ≥2000 | 25% | 16 | ≥120 |
| | NEB3 | 100% | ≥2000 | 100% | 120 | ≥16 |
| | NEB4 | 100% | ≥1000 | 12.5% | 8 | ≥120 |

[0699] 在 NEB2、NEB3 和 NEB4 中, PciI-HF 表现最好, 其中优选 FI 是 ≥ 2000 ; WT PciI 在 NEB3 中表现最好, 其中优选的 FI 是 120。总 FI 改进因子为 $\geq 2000/120 = 16$ 。

[0700] 实施例 21: 工程化高保真度 HpaI

[0701] 1. HpaI 的表达

[0702] HpaI 在用 pACYC-MseIM 和 placzz1-HpaIR 转化的大肠杆菌中表达。placzz1 是 pUC19 衍生质粒, pACYC 是低拷贝相容性质粒。细胞在 37°C 下, 在具有 Amp 和 Cam 的 LB 中生长过夜。

[0703] 2. HpaI 的诱变

[0704] 利用 Cys、Asp、Glu、Phe、His、Lys、Met、Asn、Gln、Arg、Ser 和 Thr 突变 HpaI 中的 156 个氨基酸残基。将 Trp 改变为 Ala, 和将 Tyr 改变为 Phe。这些是 :7、8、9、13、14、16、17、19、20、21、22、23、26、27、29、30、33、34、35、36、37、38、40、41、42、46、47、48、50、51、56、57、59、60、65、67、69、71、72、74、75、78、79、80、81、82、83、84、85、86、89、91、93、94、95、99、100、104、105、106、108、109、110、113、115、117、119、121、122、123、124、127、128、130、131、133、135、136、137、138、139、141、142、146、147、149、150、152、156、158、159、160、162、164、165、166、167、168、169、170、172、173、176、177、180、181、182、184、185、187、188、190、191、192、193、195、196、197、202、204、206、208、209、211、212、214、215、216、217、218、219、220、221、222、223、224、225、228、230、231、233、234、235、236、237、238、240、241、242、243、244、245、247、248、249。

[0705] 上面的数字相应于 HpaI 蛋白序列 (SEQ ID NO :86) 中的氨基酸位置。

[0706] 方法与在前面实施例中的方法相同, 其使用反向 PCR, 然后为 DpnI 消化。然后, 将处理的产物转化入大肠杆菌 (pACYC-MseIM)。

[0707] 3. HpaI-HF 的选择

[0708] 使用与前面实施例不同的方法, 完成 HpaI-HF 的选择。在 NEB2 缓冲液中, 使用 λ DNA 底物, 测定切割活性和星号活性。在 NEB2 中, 该 HpaI 具有比在 NEB4 中多得多的星号活性, 并且可在 5% 甘油中明显观察到。

[0709] HpaI (Y29F) 和 HpaI (E56A) 都是优选的具有减少星号活性的突变。将 HpaI (E56A) 标记为 HpaI-HF。

[0710] 4. HpaI-HF 和 WT HpaI 的比较

[0711] 在 NEB1-4 缓冲液中的每一种中, 对 λ DNA 底物, 使用稀释剂 A, 分开测定 HpaI-HF 和 WT HpaI 的 FI。比较在图 28 中示出, 并且结果在表 27 (下面) 中列出。

[0712] 表 27 :HpaI-HF 和 WTHpaI 的比较

| 缓冲液 | HpaI-HF | | WT HpaI | | 改进因子 |
|-------------|---------|-------|---------|-------|-------|
| | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| [0713] NEB1 | 3.1% | ≥32 | 6.3% | 32 | ≥1 |
| NEB2 | 100% | ≥2000 | 25% | 1 | ≥2000 |
| NEB3 | 12.5% | 2 | 12.5% | 2 | 1 |
| NEB4 | 50% | ≥2000 | 100% | 16 | ≥120 |

[0714] 在 NEB2 中, HpaI-HF 表现最好, 其中优选 FI 是 ≥ 2000 ; WT PciI 在 NEB4 中表现最好, 其中优选 FI 是 16。总 FI 改进因子为 $\geq 2000/16 = 120$ 。

[0715] 实施例 22: 工程化高保真度 AgeI

[0716] 1. AgeI 的表达

[0717] AgeI 在用 pRRS-AgeIRM 和 psyx20-lacIq 转化的大肠杆菌中表达。pRRS 是 pUC19 衍生质粒, psyx20-lacIq 是低拷贝相容性质粒, 其带有在 lacIq 启动子下表达的 lacI。细胞在 37°C 下, 在具有 Amp 和 Kan 的 LB 中生长至 200Klett 单位, 然后用 0.5mM IPTG 在 25°C 下诱导过夜。AgeI 的表达极难实现, 因为它是不稳定的。

[0718] 2. AgeI 的诱变

[0719] 利用 Cys、Asp、Glu、Phe、His、Lys、Met、Asn、Gln、Arg、Ser 和 Thr 突变 AgeI 中 149 个氨基酸残基。将 Trp 改变为 Ala, 和将 Tyr 改变为 Phe。这些是: 2、4、6、7、9、14、16、18、19、21、22、23、24、25、26、27、29、30、31、32、37、38、40、42、43、44、45、49、51、53、55、56、58、60、62、64、65、67、68、69、72、73、75、77、78、79、82、83、85、86、87、88、90、91、92、94、96、97、102、103、104、105、110、111、114、116、119、120、122、123、128、129、130、134、135、138、139、140、142、144、146、147、148、152、153、155、157、159、166、168、170、173、174、176、177、178、182、183、185、186、188、192、195、198、200、201、206、211、212、214、217、219、220、222、223、224、225、226、227、229、231、233、234、235、237、238、239、240、241、243、245、247、248、250、251、253、255、256、258、260、262、265、266、267、268、269、271、272。

[0720] 方法与在前面实施例中的方法相同, 其使用反向 PCR, 然后为 DpnI 消化。然后, 将处理的产物转化入大肠杆菌 (psyx20-lacIq)。

[0721] 上面的数字相应于 AgeI 蛋白序列 (SEQ ID NO:79) 中的氨基酸位置。

[0722] 3. AgeI-HF 的选择

[0723] 使用与前面实施例相似的方法, 完成 AgeI-HF 的选择。标准活性检查在 NEB4 中应用 5% 甘油使用 pXba 进行, 并且星号活性检查在具有 39% 甘油的 NEB4 缓冲液中对 pXba 进行。因为表达系统的困难性, 在获得有意义的突变体之前, 重复该选择八次。两个突变体——S201A 和 R139A——具有减少的星号活性, 并且将 R139A 标记为 AgeI-HF。

[0724] 4. AgeI-HF 和 WT AgeI 的比较

[0725] 在 NEB1-4 缓冲液中的每一种中, 对 pXba 底物, 使用稀释剂 A, 分开测定 AgeI-HF 和 WT AgeI 的 FI。比较在图 29 中示出, 并且结果在表 28(下面) 中列出。

[0726] 表 28: AgeI-HF 和 WT AgeI 的比较

| 缓冲液 | AgeI-HF | | WT AgeI | | 改进因子 |
|------|---------|-------|---------|-------|-------|
| | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| NEB1 | 100% | ≥500 | 100% | 16 | ≥32 |
| NEB2 | 50% | ≥250 | 50% | 8 | ≥32 |
| NEB3 | 6.3% | ≥16 | 12.5% | 64 | ≥0.25 |
| NEB4 | 100% | ≥250 | 50% | 8 | ≥32 |

[0728] 在 NEB1 和 NEB4 中, AgeI-HF 表现最好, 其中优选 FI 是 ≥ 500 ; WT AgeI 在 NEB3 中表现最好, 其中优选的 FI 是 16。总 FI 改进因子为 $\geq 500/16 = 32$ 。

[0729] 实施例 23: 工程化高保真度 BsmBI

[0730] 1. BsmBI 的表达

[0731] BsmBI 在用 pACYC-BsmAIM、ptaczz2-BsmBIR 和 psyx20-lacIq 转化的大肠杆菌中表达。Ptaczz2 是 pUC19 衍生质粒, 其携带诱导型 ptac 启动子, pACYC 是低拷贝相容性质粒。BsmAIM(GTCTC) 遮盖 (cover) BsmBI (CGTCTC) 特异性。psyx20-lacIq 是低拷贝载体, 其具有强的 lacI 表达。细胞在 37°C 下生长, 然后, 在具有 Amp、Cam 和 Kan 的 LB 中诱导。

[0732] 2. BsmBI 的诱变

[0733] 利用 Cys、Asp、Glu、Phe、His、Lys、Met、Asn、Gln、Arg、Ser 和 Thr 突变 BsmBI 中 358 个氨基酸残基。将 Trp 改变为 Ala, 和将 Tyr 改变为 Phe。这些是: 8、9、12、13、14、15、17、18、19、20、21、24、25、26、27、28、29、31、33、37、38、40、42、43、44、46、47、49、50、51、52、53、54、55、56、57、58、60、61、62、63、64、65、66、67、69、70、72、76、78、79、80、81、82、83、84、88、91、93、95、96、98、99、101、103、104、105、106、109、110、111、113、114、115、117、118、119、120、121、122、124、126、127、128、130、131、132、133、134、135、138、141、143、144、145、147、149、150、154、155、157、158、160、162、163、164、165、166、167、168、169、172、174、175、176、177、179、180、181、182、184、185、186、188、189、191、194、195、197、200、201、203、205、206、207、208、211、212、213、214、215、216、217、220、221、222、223、224、225、226、228、229、230、231、232、233、235、236、237、238、239、240、242、243、247、250、251、252、257、258、260、262、263、264、265、266、268、269、271、273、274、279、280、282、283、284、287、288、289、292、294、295、296、297、299、300、301、302、303、304、305、306、307、309、310、313、314、315、316、317、318、320、321、324、325、326、328、331、332、333、334、335、336、339、340、341、342、343、344、345、348、349、351、352、353、355、356、357、358、360、361、363、364、366、367、368、370、372、375、376、377、379、381、382、384、385、386、388、389、390、393、394、395、396、398、399、400、401、402、403、404、405、406、407、408、409、410、411、412、413、414、418、421、424、425、426、428、430、431、432、433、434、436、437、438、439、442、443、444、445、446、446、448、449、450、451、452、453、454、457、458、460、461、462、464、466、467、468、470、471、472、473、474、477、478、479、480、482、483、484、485、486、487、488、489、491、492、495、496、497、498、499、500、502、503、504、505、506、507、510、511、515、516、517、518、519、522、523。

[0734] 上面的数字相应于 BsmBI 蛋白序列 (SEQ ID NO :81) 中的氨基酸位置。

[0735] 方法与在前面实施例中的方法相同, 其使用反向 PCR, 然后为 DpnI 消化。然后, 将处理的产物转化入大肠杆菌 (pACYC-BsmAIM, psyx20-lacIq)。

[0736] 3. BsmBI-HF 的选择

[0737] 使用与前面实施例相似的方法,完成 BsmBI-HF 的选择。切割在 NEB4 中应用 5% 甘油使用 λ DNA 测定,并且星号活性在具有 39% 甘油的 NEB4 缓冲液中使用 Litmus28i 测定。优选的突变体包括 H230A、D231A 和 N185Y/R232A。标记 N185Y/R232A 为 BsmBI-HF。

[0738] 4. BsmBI-HF 和 WT BsmBI 的比较

[0739] 在 NEB1-4 缓冲液中的每一种中,对 λ DNA 底物,使用稀释剂 A,分开测定 BsmBI-HF 和 WT BsmBI 的 FI。比较在图 30 中示出,并且结果在表 29(下面)中列出。

[0740] 表 29 :BsmBI-HF 和 WT BsmBI 的比较

| 缓冲液 | BsmBI-HF | | WT BsmBI | | 改进因子 |
|------|----------|------------|----------|-------|------------|
| | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| NEB1 | 100% | 32 | 12.5% | 1 | 32 |
| NEB2 | 100% | ≥ 500 | 50% | 8 | ≥ 64 |
| NEB3 | 12.5% | ≥ 64 | 100% | 120 | ≥ 0.5 |
| NEB4 | 100% | ≥ 500 | 25% | 4 | ≥ 120 |

[0742] 在 NEB1、NEB2 和 NEB4 中, BsmBI-HF 表现最好,其中优选 FI 是 ≥ 500 ; WT BsmBI 在 NEB3 中表现最好,其中优选的 FI 是 120。总 FI 改进因子为 $\geq 500/120 = 4$ 。

[0743] 实施例 24 :工程化具有减少星号活性的 BspQI 变体

[0744] 1. BspQI 限制性内切核酸酶的位点定向诱变

[0745] 用 Km^R 和 Amp^R 的 pSX33-EarIM1M2 和 pZZlacI-PspQ 转化大肠杆菌。M. EarI (CTCTTC) 也修饰 BspQI 位点 (GCTCTTC), 并因此使用 pSX33-earIM1M2 (图 17 和 18) 共转化和修饰大肠杆菌染色体。WT 氨基酸序列在图 16 中示出。

[0746] 将 BspQI 中 122 个带电荷的或不带电荷的氨基酸残基 (Arg、Lys、His、Glu、Asp、Gln、Asn、Cys) 通过位点定向诱变改变为 Ala。在下列条件下,进行 PCR :DNA 变性,98°C 持续 30sec,1 个循环 ;DNA 变性 / 引物退火 / 延伸,98°C 持续 10sec,55°C 到 65°C 进行 30sec,72°C 持续 2min,进行 18 个 PCR 循环 ;72°C 持续 15min,1 个循环。在 100 μ l 反应中,2 单位的 Phusion™ DNA 聚合酶 (NEB, Ipswich, MA)、1mM dNTP、10ng 到 100ng 模板 DNA、20 μ l 5x 反应缓冲液、0.04 μ M 引物、无菌水至 100 μ l 总体积。

[0747] 用 DpnI 消化 PCR DNA,以破坏模板 DNA (Dam 甲基化的) 并与 pSX33-earIM1M2 共转化入大肠杆菌。培养各转化体过夜 (5ml LB、50 μ g/ml Km^R 和 100 μ g/ml Amp^R), 并且分成两份。一份 (1.5ml) 通过离心收获,并通过在超声处理缓冲液 (20mM Tris-HCl, pH 7.5, 0.1mM DTT, 50mM NaCl, 10% 甘油) 中超声处理进行裂解。细胞提取物在 50°C 下加热 1 小时,并且通过离心除去变性的大肠杆菌蛋白。在 pUC19DNA 上,分析澄清的溶胞产物的限制性活性和星号活性。

[0748] BspQI 星号活性分析条件 :1 μ g pUC19DNA、5 μ l 10x NEB 缓冲液 1、25% DMSO、2.5 μ l 澄清细胞提取物、无菌去离子水至 50 μ l 总体积,并且在 50°C 下温育 1 小时。通过电泳,在 0.8 到 1% 琼脂糖凝胶中,解析消化的 DNA。

[0749] 收获第二份细胞培养物 (未诱导的), 并通过 Qiagen 离心柱纯化方法 (Qiagen, Valencia, CA) 制备质粒 DNA。通过大-染料双脱氧-终止子-测序方法

(Big-dyedideoxy-terminator sequencing method), 测序 bspQIR 等位基因, 以证实期望的突变。在鉴定减少星号活性突变体之后, 获得新转化体, 并且制造 IPTG- 诱导的培养物。再次分析限制性活性和星号活性, 以证实在所有四种缓冲液中与 WT 酶相比减少的星号活性。

[0750] 在通过位点定向诱变构建的 122 个 BspQI 突变体中, 两种 BspQI 变体——R388A 和 K279A——显示减少的星号活性。在缓冲液 1 和 10% 甘油中, R388A 的星号活性减少大约 16 倍。然而, R388A 在高酶浓度下仍旧显示星号活性。BspQI 变体 K279A 也显示减少的星号活性 (在减少星号活性方面 > 8- 倍改进)。

[0751] 为了进一步在高酶浓度下减少星号活性, 将 R388 和 K279 取代为 Phe、Pro、Tyr、Glu、Asp 或 Leu。分析各种 R388X 和 K279X 突变体的 IPTG- 诱导的细胞提取物的限制性活性和星号活性。发现 R388F 或 K279P 在细胞提取物或纯化酶中显示最小的星号活性。氨基酸取代不影响比活性。

[0752] 为了仍进一步减少 BspQI 星号活性, 通过位点定向诱变, 将两种氨基酸取代组合进一个突变体酶中 (双突变体, K279P/R388F)。该双突变体在具有 10% 甘油的缓冲液 1 和缓冲液 2 中缺少星号活性 (图 41B)。

[0753] 表 30 :BspQI-HF 对 WT BspQI

| 缓冲液 | BspQI-HF | | WT BspQI | | 改进因子 |
|------|----------|-------|----------|-------|-------|
| | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| NEB1 | 25% | ≥1000 | 12.5% | 2 | ≥1000 |
| NEB2 | 25% | ≥1000 | 100% | 16 | ≥64 |
| NEB3 | 1.3% | ≥64 | 100% | 32 | ≥2 |
| NEB4 | 100% | ≥4000 | 50% | 4 | ≥1000 |

[0755] 实施例 25 : 工程化具有减少星号活性的 SapI 变体

[0756] 在 BspQI 中发现保守的 K279 和 R388 氨基酸残基, 其中在 SapI 中相应位置是 K273 和 R380。6x His 标记 SapI 表达克隆首先在 pUC19 中构建。SapI 表达菌株是用 Km^R 和 Amp^R 的 pSX33-earIM1M2 和 pUC-SapI 转化的大肠杆菌。通过位点定向诱变, 将 Lys273 到 Pro (K273P) 和 Arg 380 到 Phe (R380F) 的氨基酸取代引入到 SapI。也构建 SapI 单突变体 R380A。当进行限制性活性和星号活性反应时, SapI 变体 R380A 和 K273P/R380F 都示出减少的星号活性 (图 42)。

[0757] PCR、转化、质粒 DNA 制备和酶活性分析如对 BspQI 描述的进行, 除了在 37°C 下, 测定 SapI 活性。6xHis- 标记 SapI 变体 K273P/R380F 通过 Ni-NTA 柱色谱纯化, 并且示出在存在 25% DMSO 或 5% 甘油情况下, 显示减少的星号活性。

[0758] 实施例 26 : 工程化 KpnI 高保真度突变体

[0759] 包含两种活性的 KpnI 已被改变为具有较低星号活性的突变体 (国际公开号 W007/027464)。下面的实施例描述具有改进的星号活性和与野生型相似的切割活性的新的突变体。

[0760] 将除了催化残基的带电荷的氨基酸残基 (Asp、Glu、Arg、Lys 和 His) 或极性氨基酸 (Ser、Thr、Tyr、Asn、Gln、Phe、Trp、Cys 和 Met) 单独突变为丙氨酸。

[0761] 通过反向 PCR, 使用具有希望突变的引物, 进行诱变。一般而言, 在 100 μL 终体积

中,使用 0.4mM 的每种 4dNTP、1x ThermoPol 缓冲液 (NEB)、20ng 的模板 DNA、40 μ mol 的每种引物和 4U 的 Vent DNA 聚合酶 (NEB),实施反向 PCR。

[0762] 分别含有 Plac 或 Ptac 启动子控制下的 KpnIR 的质粒 pUC19 或 pAGR3 被用作模板。使用下列温度方案进行 PCR:94 $^{\circ}$ C 持续 4 分钟,然后 94 $^{\circ}$ C 持续 30 秒进行变性、55 $^{\circ}$ C 持续 30 秒进行退火和 72 $^{\circ}$ C 持续 5 分钟进行延伸的 25 个循环。循环后在 72 $^{\circ}$ C 下温育 7 分钟,然后用 20U 的 DpnI (NEB) 在 37 $^{\circ}$ C 下处理反应 1 小时,以降解模板 DNA。在 80 $^{\circ}$ C 下失活 DpnI 20 分钟后,使用 2 μ l 反应物转化用 pSYX20-KpnIM 预转化的 50 μ l 化学感受态 NEB5alpha (NEB)。将转化的细菌铺板在含有 100 μ g/ml 氨苄青霉素和 30 μ g/ml 卡那霉素的 LB 板上,并且在 37 $^{\circ}$ C 下温育 12 到 15 小时。每种构建物的 3 到 4 个菌落在 1ml 含有 100 μ g/ml 氨苄青霉素和 30 μ g/ml 卡那霉素的 LB 中,在 37 $^{\circ}$ C 下培养 12 到 15 小时,并伴随 200rpm 的摇动。将培养物离心 (spin down),并且重新悬浮在 0.2ml 的超声处理缓冲液 (20mM Tris-HCl, pH 8.3, 50mM NaCl, 1mM EDTA, 1mM PMSF) 中。将重新悬浮的细胞超声处理 20 秒,然后在 4 $^{\circ}$ C 下以 13,000rpm 离心 5 分钟。进行上清液稀释,并且分析其中 5 μ l 的 KpnI 切割活性。

[0763] 对于筛查突变体,使用 50 μ l 总体积中 5 μ l 以 10- 或 100- 倍稀释的溶胞产物上清液、0.5 μ g 的 pXba DNA (NEB) 和 1x NEBuffer 4,进行活性分析反应。在 37 $^{\circ}$ C 下温育 1 小时后,通过加入 10 μ l 的 6x DNA 上样染料停止分析反应,并且通过 1xTBE 中 0.8% 琼脂糖凝胶电泳分析。分析示出与亲代酶 (KpnI D148E) 相比具有增加的总切割活性的突变体在含有 25% DMSO 的缓冲液中减少的星号活性。

[0764] 在 1x NEBuffer 4 中存在 5% 甘油和 0.2mg/ml BSA 的情况下 (总体积 = 50 μ l),使用 5 μ l 用 1 μ g 的 pXba DNA (NEB) 温育的酶稀释物,进行分析。在 37 $^{\circ}$ C 下温育 1 小时后,用 20 μ g 蛋白激酶 K (NEB) 在 37 $^{\circ}$ C 下处理反应 15 分钟,然后通过 1x TBE 中 0.8% 琼脂糖凝胶电泳分析。使用 50U 的酶和 1 μ g pXba DNA,在含有 20mM TrisHCl、50mM NaCl, pH 7.9、1mM DTT 和增加浓度的 MgSO₄/CaCl₂/MnCl₂ 的缓冲液中,在 37 $^{\circ}$ C 下进行二价金属依赖性分析 1 小时,然后通过 1xTBE 中 0.8% 琼脂糖凝胶电泳。总反应体积为 50 μ l。

[0765] 随机诱变的结果是优选的突变体 KpnI D148E,其示出比野生型酶和前面描述的 KpnI D163I/K165A 突变体 (国际公开号 WO 07/027464) 低的星号活性。然而, KpnID148E 在高酶浓度下显示星号活性。在 D148E 背景中,构建双或三突变体以减少该观察到的星号活性。发现 KpnI (D16N/E132A/D148E) 具有比突变体 D148E 低的星号活性和高的比活性。通过位点定向诱变,引入氨基酸取代 D148 到 E 和 E132 到 A。通过 PCR 引入 D16 到 N 的突变。在标准反应条件中,将纯化酶与底物 DNAPXba (含有 6 个 KpnI 位点),在 37 $^{\circ}$ C 下) 温育 1 小时产生表 31 中示出的减少的星号活性。

[0766] 表 31 :KpnI 突变体的保真度指数

| 缓冲液 | KpnI-HF | | WT KpnI | | 改进因子 |
|------|---------|-------------|---------|-------|-------------|
| | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| NEB1 | 100% | ≥ 4000 | 100% | 16 | ≥ 250 |
| NEB2 | 50% | ≥ 2000 | 25% | 16 | ≥ 64 |
| NEB3 | 6.3% | ≥ 250 | 6.3% | 8 | ≥ 32 |
| NEB4 | 100% | ≥ 4000 | 50% | 4 | ≥ 1000 |

[0767]

[0768] 对于突变体 D16N/E132A/D148E, 上至 4000U, 没有观察到 pXba 的星号活性。当用 KpnI D148E 和 D16N/E132A/D148E 切割时, 对于 pBR322 底物——其不含有 KpnI 位点——观察到的星号活性也被减少。

[0769] 实施例 27: 工程化高保真度 BsaI

[0770] 1. BsaI 的表达

[0771] BsaI 在用 pACYC-BsmAIM、pUC19-BsaIR 和 psyx20-lacIq 转化的大肠杆菌中表达。pACYC 是低拷贝相容性质粒。BsmAIM(GTCTC) 遮盖 BsmBI 特异性 (CGTCTC)。psyx20-lacIq 是具有强 lacI 表达的低拷贝载体。细胞在 37°C 下生长, 然后在具有 Amp、Cam 和 Kan 的 LB 中诱导。

[0772] 2. BsaI 的诱变

[0773] BsaI 的氨基酸与 BsmBI 的氨基酸相似。在相应的前述有效位点周围和相应的前述有效位点处的 11 个氨基酸被突变为 R229A、S230A、Y231F、T232A、T233A、D234A、R235A、R236A、F238A、E239A、Y240F。

[0774] 方法与在前面实施例中的方法相同, 其使用反向 PCR, 然后为 DpnI 消化。然后, 将处理的产物转化入大肠杆菌 (pACYC-BsmAIM、psyx20-lacIq)。

[0775] 3. BsaI-HF 的选择

[0776] 使用与前面实施例相似的方法, 完成 BsaI-HF 的选择。标准活性检查在 NEB4 中应用 5% 甘油使用 λ DNA 进行, 并且星号活性检查在具有 39% 甘油的 NEB4 缓冲液中对 litmus28i 进行。在 11 种设计的突变体中的一种突变体 Y231F 减少星号活性, 并将其标记为 BsaI-HF。

[0777] 4. BsaI-HF 和 WT BsaI 的比较

[0778] 在 NEB1-4 缓冲液中, 对 λ DNA, 使用稀释剂 A, 分开测定 BsaI-HF 和 WT BsaI 的 FI。结果在表 32 (下面) 中列出。

[0779] 表 32: BsaI-HF 和 WT BsaI 的比较

| 缓冲液 | BsaI-HF | | WT BsaI | | 改进因子 |
|------|---------|-------------|---------|-------|------------|
| | 活性 | 保真度指数 | 活性 | 保真度指数 | |
| NEB1 | 50% | ≥ 4000 | 25% | 8 | ≥ 500 |
| NEB2 | 100% | ≥ 8000 | 100% | 120 | ≥ 64 |
| NEB3 | 100% | 120 | 25% | 16 | 8 |
| NEB4 | 100% | ≥ 8000 | 100% | 32 | ≥ 250 |

[0781] 在 NEB2、NEB3 和 NEB4 中, BsaI-HF 表现最好, 其中最好 FI 是 ≥ 8000 ; WT BsaI 在 NEB2 和 NEB4 中表现最好, 其中最好的 FI 是 120。所以, 总 FI 改进因子为 $\geq 8000/120 = \geq 64$ 。

序列表

<110> 新英格兰生物实验室公司

朱振宇

A·布兰查德

徐双勇

管胜昔

魏华

张鹏华

孙大鹏

陈少康

<120> 高保真度限制性内切核酸酶

<130>NEB-293-PCT

<160>106

<170>PatentIn version 3.4

<210>1

<211>1089

<212>DNA

<213> 藤黄微球菌 (*Micrococcus luteus*)

<400>1

```

atgatcaagt acttgggtag caageggacg ctcgtgcccc tcctcgggta categetteg 60
gcctctgaag caacagagge ggttgacctg ttcactggca cgacgcgtgt ggcgcaagag 120
ttcaagegtc gegggetteg agttcttgct aacgacatag cgacgtactc cgaggtttta 180
gcccagtgct atategccac caacggccag gaagttgacc gccgtgcgct cgaggccgct 240
ctggcgggagc tgaacgcctt gcccgcgaa cctggatact tcacggaaac cttctgtgag 300
gcttctcgct acttccagcc caagaacggg gctcgggtgg atgcaatcag gaatgcgac 360
gacgaccggt acgcgactc atggatgcga ccgatcctcc tcacgagctt gatgcttgcg 420
gccgaccgcg tcgactccac taccggagtg cagatggctt acctgaagca gtgggccgcg 480
cgtgcgcaca atgatctaga gttgcggctt ccagacctaa tcgcaggtga cggtgacget 540
gctcgtgagg atgcggtgac tctcgcaaa gagctgcctc gcgtccagct gatgtacett 600
gatectccct ataaccagca caggctactt accaactacc atatttggga gaccctgatt 660
cgttgggatg cccctgagag ttatgggatc gcctgtaagc gcattgactc tcgagatgat 720
gccaccaaga gccctataa tatgaagcgg cgaatgcccc acgagatgcg tcgcctgetg 780
atgaccatca aggcggacct cgcggttgta tcttacaaca atgagtcgtg gattgatccg 840
gagacgatga tgtcgaccct gcgcgatgcg ggatatgagg acgtgcgtct gctcgettte 900
gactataagc gctacgttgg ggctcaaate gggatctaca atccctccgg ggaaaaggte 960
ggtcgtgtga gtcacctccg aaacatcgag tatctctttc ttgcgggacc aacggagcgc 1020
gttgaggtgt gcgccgcgag tgttgaacac cgagcactac ccaaggaacc ggaactcacc 1080
gcgttctag 1089

```

<210>2

<211>362

<212>PRT

<213> 藤黄微球菌

<400>2

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

Asn Asn Glu Ser Trp Ile Asp Pro Glu Thr Met Met Ser Thr Leu Arg
 275 280 285
 Asp Ala Gly Tyr Glu Asp Val Arg Leu Leu Ala Phe Asp Tyr Lys Arg
 290 295 300
 Tyr Val Gly Ala Gln Ile Gly Ile Tyr Asn Pro Ser Gly Glu Lys Val
 305 310 315 320
 Gly Arg Val Ser His Leu Arg Asn Ile Glu Tyr Leu Phe Leu Ala Gly
 325 330 335
 Pro Thr Glu Arg Val Glu Val Cys Ala Ala Ser Val Glu His Arg Ala
 340 345 350
 Leu Pro Lys Glu Pro Glu Leu Thr Ala Phe
 355 360

<210>3

<211>1146

<212>DNA

<213> 幽门螺旋杆菌 (Helicobacter pylori) J166

<400>3

ttggagaatt ttttgaataa ttagatatt aaaaccttag ggcaggtttt caccctaaa 60
 aagatagtgg atttcatgct cactctcaag cacaatcatg ggagtgtttt agagccaagc 120
 gcggcgatg ggagttttt aaagcgctta aaaaaggctg tagggattga aatcgcct 180
 aaaatctgcc ctaaaaatgc ctttgcctg gactttttg actacccttt agaaaatcaa 240
 tttgacacga ttattggcaa tccgcctat gtcaagcaca aggatattgc gccaagcagc 300
 aaagaaaaac tccattacag ctttttgat gaaaggagta atctatactt gtttttcata 360
 gaaaaagcga tcaagcattt aaagcctaaa ggcgattga ttttcatcac cccaagggat 420
 ttttaaaat ccacttctag cgtgaaatta aacgaatgga tttacaaaga aggcacgata 480
 acgcattttt ttgaattagg cgatcaaaag attttcccaa acgcatgcc taattgctg 540
 attttctggt tttgtaaagg tgatttcagt agaatcacca acgatggttt gcaatttgtg 600
 tgcaaaaaag gcattttgta tttctcaac caatcttaca cgcaaaaatt aagcgagggt 660
 ttttaaggta aggtgggggc agtgagcggg tgcgataaga tttttaaaaa tgaaacatac 720
 gggaatttag aatttgtcac ctcaatcacc aaaagaacca atgtttttaga aaaaatgggt 780
 tttgtcaata aacctaatga ttatttactc cagcataaag acagcttgat gcaaagaaag 840
 attaaaaaat tcaatgaaag taattgggtt gaatggggga ggatgcatca catatcccct 900
 aaaaaacgca tttatgttaa cgccaaaacg cgccaaaaaa acccctttt catccaccaa 960
 tgcctaatt atgacggctc tatttttagcg ctattccctt ataaccacaaa tttggattta 1020
 caaacctct gcgataaact caacgctatc aactggcaag aattaggctt tgtgtgcggc 1080
 gggcgttttt tgttttcgca gcgctcttta gaaaacgccc ttttgcctaa agacttttta 1140
 aattag 1146

<210>4

<211>609

<212>DNA

<213> 发酵支原体 (*Mycoplasma fermentans*)

<400>4

```

atgggtaaat ctgaattaag tggaagatta aattggcaag cattggctgg attaaaagct   60
agtggtgctg aacaaaactt atataacgtg tttaacgctg tttttgaagg aactaaatac  120
gttttatacg agaagccaaa gcaccttaaa aatctatacg ctcaagtagt cttacctgat  180
gatgttatta aagaaatfff taatccttta attgatttat caactactca atgggggtgtt  240
tctccagatt tcgcaataga aaatacagaa acgcataaaa ttctttttgg tgaattataa  300
agacaagatg gatgggtaga aggtaaagat cctagtgtcg gcaggggtaa tgcacatgag   360
agatcttgta aattatttac tctgggatta ttaaagctt atagaacaat tgggtgaatt   420
aacgatgaag agatattgcc attctggggtt gtattcgaag gtgatataac acgagatccc  480
aaaagagtaa gagaaattac tttctgggat gaccactatc aagataatta tttcatgtgg  540
cgaccaaatg aatcaggcga aaaattagtt caacacttca atgaaaaatt aaaaaaatat  600
ttagattaa                                     609

```

<210>5

<211>202

<212>PRT

<213> 发酵支原体

<400>5

```

Met Gly Lys Ser Glu Leu Ser Gly Arg Leu Asn Trp Gln Ala Leu Ala
1           5           10           15
Gly Leu Lys Ala Ser Gly Ala Glu Gln Asn Leu Tyr Asn Val Phe Asn
           20           25           30
Ala Val Phe Glu Gly Thr Lys Tyr Val Leu Tyr Glu Lys Pro Lys His
           35           40           45
Leu Lys Asn Leu Tyr Ala Gln Val Val Leu Pro Asp Asp Val Ile Lys
           50           55           60
Glu Ile Phe Asn Pro Leu Ile Asp Leu Ser Thr Thr Gln Trp Gly Val
65           70           75           80
Ser Pro Asp Phe Ala Ile Glu Asn Thr Glu Thr His Lys Ile Leu Phe
           85           90           95
Gly Glu Ile Lys Arg Gln Asp Gly Trp Val Glu Gly Lys Asp Pro Ser
           100          105          110
Ala Gly Arg Gly Asn Ala His Glu Arg Ser Cys Lys Leu Phe Thr Pro
           115          120          125
Gly Leu Leu Lys Ala Tyr Arg Thr Ile Gly Gly Ile Asn Asp Glu Glu
           130          135          140
Ile Leu Pro Phe Trp Val Val Phe Glu Gly Asp Ile Thr Arg Asp Pro
145          150          155          160

```

Lys Arg Val Arg Glu Ile Thr Phe Trp Tyr Asp His Tyr Gln Asp Asn
 165 170 175
 Tyr Phe Met Trp Arg Pro Asn Glu Ser Gly Glu Lys Leu Val Gln His
 180 185 190
 Phe Asn Glu Lys Leu Lys Lys Tyr Leu Asp
 195 200

<210>6

<211>1152

<212>DNA

<213>嗜热脂肪芽胞杆菌 (*Bacillus stearothermophilus*)X1

<400>6

atggctatta cattatgtga cataaatggt tgtagacttg agagaggaca tactggtaaa 60
 cataataaat ttcctgaatt tgtatggact tctcaattta ataaaaaaga tattgataag 120
 gtcaataaag caggatatgc aacaccaaga ggtggggaca aaggagccta tcagaacat 180
 gtttacagaa ataataaagt aattattcct tttgaaaggt tggaaaatgt taatttaaat 240
 aactatcaag atggatatgt tattaggtta ttcctaate agtactttga atcagccggg 300
 gtagttaagc cggaattctt acaaccaaat tcatttgta aagtgggga caatgcattt 360
 attttatate gcacacatte atcttttgag gaattacctc ctctaccaga ctgggaggtt 420
 agacatctaa aaaagaacgg taatatagtt accagaagaa gtaaggacgt aatcgatgct 480
 ggacattatg tcttacgatt atcatcaatt agtaacaaaa aagaaagaaa agagggcct 540
 cctcaaggta tttttgcacc tgaatatgca aatgcagaga ctaattatct gtcaaaagca 600
 tttttagcct ggtaattat taaaactcaa aatagtcctg ataatagaaga acaattccaa 660
 cacttaagag cgatcttaat tagtcataat ctcatcaata tttctcaact tgaagaaaag 720
 gctattctaa agaatggtat cacatgctgc cctttatgcg agcaaattat tttttacgaa 780
 cagctacacg aatgggtttc ttttgaaggt gcgctctggcc ttgcgaattc acaagaacag 840
 gttgaggggtg caactaggte aacatcagtt aatttattcc atatggtacc attagtatat 900
 gaaaccttgg aacacaaacc tgatcaaata gcatggggcc atgccatttg taatactaga 960
 cttggtaaaa gagagtgcct gcctcttagt agactaaaac aagaaggtac gcccgttggt 1020
 cttcttgatg aagattcgaa tcttgaagta ttaggatgga ttagtaaaaga taagcaattt 1080
 attcgtagag aaaatgggga agtttggatt aaaattacag atattgaatt taacgatgac 1140
 tttgaagaat aa 1152

<210>7

<211>383

<212>PRT

<213>嗜热脂肪芽胞杆菌 X1

<400>7

Met Ala Ile Thr Leu Cys Asp Ile Asn Gly Cys Arg Leu Glu Arg Gly
 1 5 10 15
 His Thr Gly Lys His Asn Lys Phe Pro Glu Phe Val Trp Thr Ser Gln

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

Thr Pro Val Gly Leu Leu Asp Glu Asp Ser Asn Leu Glu Val Leu Gly
 340 345 350
 Trp Ile Ser Lys Asp Lys Gln Phe Ile Arg Thr Glu Asn Gly Glu Val
 355 360 365
 Trp Ile Lys Ile Thr Asp Ile Glu Phe Asn Asp Asp Phe Glu Glu
 370 375 380

<210>8

<211>1536

<212>DNA

<213>嗜热脂肪芽胞杆菌 X1

<220>

<221>misc_feature

<223>M. BstXI 的 DNA 序列

<400>8

```

atgatttttg ctgatattga atttgaaaa gaactttttt cagctgctaa taaattaagg 60
ggaaaaattg ctccaagtga gtataagcat tatgttttgc ctttgatatt ccttagatat 120
ttatctctta aataccaaca aagaaggaat gaaattcaac aacagataaa tgattcaagg 180
gatcacaaga aaaatcaaga tgaagtgtta aagatattgg aagacaggac tgaatacacc 240
aaagtaaatg ttttctatat tectgaaaa gctagttggg aatacttatt gaaaaattcc 300
gaaaatgata aaattaaaga aatgatagat tcagctatgg aaatactgga aatgaatat 360
gacgagttaa aaggtgtttt gccaaagata tataaaaact caaatatacc gaatgaagtt 420
attagtgatt tactaaaact attttctcaa gaagtatttt cagcacatga tggaagaaat 480
gttgatttat tggggagagt ttatgaatac ttataagta attttgctac tacagaaggt 540
actagaggtg gtgaatattt tacaccgtct tcaatcgtaa aattattggt agcaatgcta 600
gagcccatta aaggtacagt ttatgatccg gcctgtggga caggaggaat gtttattcag 660
tctaataaat atagagaaaa taatcataac ttgtgttttg taggccagga acaaaacgag 720
cttactatca aattggctaa aatgaatgga attctacatg gaataaatcc tgaattaga 780
caaggtgatt cattattaa tgaccgttat ccagaattga aagctgaaat tgtaatatct 840
aatccaccgt ttaatatgaa ggattgggga gctgaacgcc tgccacttaa tgataagcga 900
ttaataggac cggtaacaaa cagtaatgca aattacatgt ggatacagca ttttctatac 960
catttaaaag atggtggttt agcaggattt gttattgcta atggagcttt gactagtaat 1020
ctggctgctg aaaaaattgt aaggaaacac ttaatagaca atgattatgt agattgtgtt 1080
gttcaattac ctgaaaaaat gttctttggt actggcattc caagtgcttt agtgttttta 1140
agtaagaate gaaatggaag taacggccat gccaaaagag aaaaagaggt tctatttatt 1200
gatgcaagcg ataagggaac attagtgggt aaaaagaata aaatattttt agatgatgaa 1260
ataaaagaaa ttgcagattt atatcattca tttaaatttt taaatgataa tgattataac 1320
catagtgggt tttacaaaaa ggttaacatt gaaaaaatcg tggaaaatga ttataaatta 1380
actccaacte tctatgtagg tgtaaaggaa gagactgaaa tggagaagcc atttagagaa 1440
atgataatag aatataaagc gatattagag caacaatttg aagaatcaaa caaactacag 1500

```

cagaaaatat taaagaattt agaggatta ttatga 1536

<210>9

<211>511

<212>PRT

<213>嗜热脂肪芽胞杆菌 X1

<220>

<221>misc_feature

<223>M. BstXI 的蛋白质序列

<400>9

Met Ile Phe Ala Asp Ile Glu Phe Glu Lys Glu Leu Phe Ser Ala Ala
 1 5 10 15

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

Leu Pro Leu Ile Phe Leu Arg Tyr Leu Ser Leu Lys Tyr Gln Gln Arg
 35 40 45

Arg Asn Glu Ile Gln Gln Gln Ile Asn Asp Ser Arg Asp His Lys Lys
 50 55 60

Asn Gln Asp Glu Val Leu Lys Ile Leu Glu Asp Arg Thr Glu Tyr Thr
 65 70 75 80

Lys Val Asn Val Phe Tyr Ile Pro Glu Lys Ala Ser Trp Glu Tyr Leu
 85 90 95

Leu Lys Asn Ser Glu Asn Asp Lys Ile Lys Glu Met Ile Asp Ser Ala
 100 105 110

Met Glu Ile Leu Glu Asn Glu Tyr Asp Glu Leu Lys Gly Val Leu Pro
 115 120 125

Lys Ile Tyr Lys Asn Ser Asn Ile Pro Asn Glu Val Ile Ser Asp Leu
 130 135 140

Leu Lys Leu Phe Ser Gln Glu Val Phe Ser Ala His Asp Gly Arg Asn
 145 150 155 160

Val Asp Leu Leu Gly Arg Val Tyr Glu Tyr Phe Ile Ser Asn Phe Ala
 165 170 175

Thr Thr Glu Gly Thr Arg Gly Gly Glu Tyr Phe Thr Pro Ser Ser Ile
 180 185 190

Val Lys Leu Leu Val Ala Met Leu Glu Pro Ile Lys Gly Thr Val Tyr
 195 200 205

Asp Pro Ala Cys Gly Thr Gly Gly Met Phe Ile Gln Ser Asn Lys Tyr
 210 215 220

Arg Glu Asn Asn His Asn Leu Cys Phe Val Gly Gln Glu Gln Asn Glu
 225 230 235 240

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

<210>10

<211>1179

<212>DNA

<213>嗜热脂肪芽胞杆菌 X1

<220>

<221>misc_feature

<223>S. BstXI 的 DNA 序列

<400>10

```

atgaaaagta ctttgaagga atataaattg ggtgatatta ccgaagtcgt taatggtgcc   60
actccttcaa ctaaaaagcc tgagtactat gaaaatggta caattccatg gattactcct  120
aaagatttat caggctatta ctttaaatat atatctcatg gtgaacgtaa tataacagag  180
cttggcttaa gaaatagttc agctaagttg ttaccaaaaag gaactgtatt attttcctca  240
agagccccaa taggatacgt agcaatagct gataattggt taactacgaa ccagggattt  300
aaaagtttta tatgtaatga ggagattatt tacaatgaat acctttatta ttttcttatt  360
gctaaaaggg attttattga aacatttgcg aatggggagta cgtttaaaga gctttcatca  420
acttctgcaa agaataatac aatcaatctt cctagtttag aagagcaaaa gaagattgtg  480
acaattttag gggatttgga tagaaagata gaattaaatt ataaaattat tgaaagctta  540
gaaaaaatag cagaaagaac atataaatat tggtttgcg atgaattaa tcaagatgaa  600
cagcacatcc gtaatggatg ggaaactgct aaaattggcg atgtggtgga acttttggga  660
gggggaaccc ctaaaacttc ggaaagtaag tattgggaag atggagatat taattggttt  720
actccttcag atttaacaaa aactagacag ctttttgtac gtgattctca aagaaaaata  780
acaattgatg gacttaataa cagtgcagcg aaattaattc ccccttattc cttgttaatg  840
tcaagtagag ctacaattgg cgagttggca attaatacaag aatctgctac tacaaatcaa  900
gggtttattg tattaatacc aaatgaaaaa atttctatth accaattata cttttgggct  960
aaacttaata agagcaaaat tatttcaatg gcaaatggta gtacttttaa agaaattagt 1020
aagcgggatt ttaaactctt ggagataata ttaccaaaaa atatagacac ttttaattca 1080
attatgcaag attatthtag gaaaattgag gagttaattg atgaaataaa aatcttaaaa 1140
accgcaagag ataatttaat tccaaaactt ataaaatga                               1179

```

<210>11

<211>392

<212>PRT

<213>嗜热脂肪芽胞杆菌 X1

<220>

<221>misc_feature

<223>S. BstXI 的蛋白质序列

<400>11

```

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

```

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

370 375 380
 Asn Leu Ile Pro Lys Leu Ile Lys
 385 390
 <210>12
 <211>770
 <212>DNA
 <213> 柠檬色动性球菌 (planococcus citreus) SE-F45
 <400>12
 atgaaacagt ttgcagatcc ttttgaaaga agattccttg atgcaattga acatcatctt 60
 gatggaattt ctgagaaaat aaaaaagac tttacacaca aaaacttttt aaaagaattg 120
 aatggcctta aaggtgataa agtctatcat gacttaggct ttgataccgc tgaatatact 180
 ctggtacgtc ttataggaag aatgagcata agcgttgga gaaggctggg ggagatatac 240
 gataaagtcc ctcgttatgt tgctgccgcg cgatttggtc ttcaaccaa tcaaattgca 300
 gaagtatttg atggtcttga gttagatata gctttgcgca atagcctttt gtcagatgat 360
 gataaaattc acataaaaaa aataactgaa aagatgtcag gcgaaacata ctcgggaatc 420
 ggaatcgaaa ttcgttataa ctttaatcca aatgacagtt cccgtttaag aaaagacgtc 480
 gatgtagctt ctaaattgtc ggccgcgggg ttatttctcg tttatttaat atttagctct 540
 ctcagtccta ggaatgatgc aatagcccgt cttaaaagag ggggatggag ctttaaacag 600
 gggcaggaag ccttagactt ccttaccgaa cttttaggag tggatattgg gtctgtttta 660
 tctgaccaa taatagccgc agaaactagg gagaaaacat caaaaattat gaagtctata 720
 tttgaatcag aggcattcca atctgttata ccgggagagt ggagtaaact 770
 <210>13
 <211>257
 <212>PRT
 <213> 柠檬色动性球菌 SE-F45
 <400>13
 Met Lys Gln Phe Ala Asp Pro Phe Glu Arg Arg Phe Leu Asp Ala Ile
 1 5 10 15
 Glu His His Leu Asp Gly Ile Ser Glu Lys Ile Lys Lys Asp Phe Thr
 20 25 30
 His Lys Asn Phe Leu Lys Glu Leu Asn Gly Leu Lys Gly Asp Lys Val
 35 40 45
 Tyr His Asp Leu Gly Phe Asp Thr Ala Glu Tyr Thr Leu Val Arg Leu
 50 55 60
 Ile Gly Arg Met Ser Ile Ser Val Gly Arg Arg Leu Gly Glu Ile Tyr
 65 70 75 80
 Asp Lys Val Pro Arg Tyr Val Ala Ala Ala Arg Phe Gly Leu Gln Pro
 85 90 95
 Asn Gln Ile Ala Glu Val Phe Asp Gly Leu Glu Leu Asp Ile Ala Leu

| | | |
|---|-----|-----|
| 100 | 105 | 110 |
| Arg Asn Ser Leu Leu Ser Asp Asp Asp Lys Ile His Ile Lys Lys Ile | | |
| 115 | 120 | 125 |
| Thr Glu Lys Met Ser Gly Glu Thr Tyr Ser Gly Ile Gly Ile Glu Ile | | |
| 130 | 135 | 140 |
| Arg Tyr Asn Phe Asn Pro Asn Asp Ser Ser Arg Leu Arg Lys Asp Val | | |
| 145 | 150 | 155 |
| Asp Val Ala Ser Lys Leu Ser Ala Ala Gly Leu Phe Pro Val Tyr Leu | | |
| 165 | 170 | 175 |
| Ile Phe Ser Ser Leu Ser Pro Arg Asn Asp Ala Ile Ala Arg Leu Lys | | |
| 180 | 185 | 190 |
| Arg Gly Gly Trp Ser Phe Lys Gln Gly Gln Glu Ala Leu Asp Phe Leu | | |
| 195 | 200 | 205 |
| Thr Glu Leu Leu Gly Val Asp Ile Gly Ser Val Leu Ser Asp Pro Ile | | |
| 210 | 215 | 220 |
| Ile Ala Ala Glu Thr Arg Glu Lys Thr Ser Lys Ile Met Lys Ser Ile | | |
| 225 | 230 | 235 |
| Phe Glu Ser Glu Ala Phe Gln Ser Val Ile Pro Gly Glu Trp Ser Lys | | |
| 245 | 250 | 255 |

Leu

<210>14

<211>1347

<212>DNA

<213> 柠檬色动性球菌 SE-F45

<220>

<221>misc_feature

<223>M. PciI 的 DNA 序列

<400>14

```

atgacaaatt tttcgactc agctetaacg agctacgac ttctcgggca tgaattgtc   60
caagattctg aagctgtag ctcgggtcca tatctgggca gctatgacc gatccctgta  120
cgtcggtcta cattcctagc tggactgtca gagaacgttc actcgtgggt tcgtctcaca  180
ccaagtttcg gaccgatct agttcgaaca atcatcaaac agatgaatct tgcgccgcac  240
tcacacatcc atgacccttt ctcaggagcc gggactaccg cgattgagc ttcgtagag   300
ggctatgaag caagctgcgt agaagttaat ccgtttctct acttcgtggg gaaaacatcc  360
atagattggt ctatcaatgc tgatgatgct gcagcgcagc tagaaagcat taaaaataaa  420
tattatagca tgtctgcaac cgctactttg gataacatag ccgacctagg aatagatata  480
ccaaaaatac acaatattca tcggtggtgg agaaacgatg ttcttaaaga tatattagtc  540
ctaaaatctt ctatcagatc ttgcacacaa gataagtatt gttccttttt tgagctagcc  600
ctagctgcag ttctcgttcc agatttgaca aatgtaacgc taggaaaact acaactgcac  660

```

ttgtgtaaaca aagacgataa agagataaac gtctggccta catatgaate tcatgcaaaa 720
 aaaatgattc acgacttgte attaattaat aagcaaaatt tcgaattttt gccaagatt 780
 atttatggtg attcaactca aaaatcaaca tttagcgagg tggcagggat agatgctata 840
 ataacatecc ctcctgaccc taataggtac agctatattt ggaataactcg ccctcacctg 900
 tacattcttg atatgatttc cgaagcaaaa gaggcttcgc aaatagatcg tagaacgatt 960
 ggtggaacat gggggacagc aacttccgaa ttaggaaagg gtatattttc tccaatcaat 1020
 gctgtagtca aagacgcgct tgaaggggtt cacgaaagaa tcgccggttc cgatcaactc 1080
 atggcaaact atgtaactca ttattttaat cgctctttt tacatataga agctataaaa 1140
 ccatcactta atccaaaage aaagcttgct tatgttggtg ggaactcttg gattaagggc 1200
 gaatatgtag cactgacgt aatcttagca aaaattatcg aaggggcttt gccaggctca 1260
 tcaattgatg gtcttcacg tttccgctgc cggaacagtg gaaagaatct ctttgaaact 1320
 atagtttact ccactctccc ggtataa 1347

<210>15

<211>448

<212>PRT

<213> 柠檬色动性球菌 SE-F45

<220>

<221>misc_feature

<223>M. PciI 的蛋白质序列

<400>15

Met Thr Asn Phe Ser His Ser Ala Leu Thr Ser Tyr Asp Leu Leu Gly
 1 5 10 15
 His Glu Ile Val Gln Asp Ser Glu Ala Val Ser Ser Gly Pro Tyr Leu
 20 25 30
 Val Ser Tyr Asp Pro Ile Pro Val Arg Arg Ser Thr Phe Leu Ala Gly
 35 40 45
 Leu Ser Glu Asn Val His Ser Trp Phe Arg Leu Thr Pro Ser Phe Gly
 50 55 60
 Pro Asp Leu Val Arg Thr Ile Ile Lys Gln Met Asn Leu Ala Pro His
 65 70 75 80
 Ser His Ile His Asp Pro Phe Ser Gly Ala Gly Thr Thr Ala Ile Glu
 85 90 95
 Ala Ser Leu Glu Gly Tyr Glu Ala Ser Cys Val Glu Val Asn Pro Phe
 100 105 110
 Leu Tyr Phe Val Gly Lys Thr Ser Ile Asp Trp Ser Ile Asn Ala Asp
 115 120 125
 Asp Ala Ala Ala Gln Leu Glu Ser Ile Lys Asn Lys Tyr Tyr Ser Met
 130 135 140
 Ser Ala Thr Ala Thr Leu Asp Asn Ile Ala Asp Leu Gly Ile Asp Ile

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

<210>16

<211>430

<212>PRT

<213> 球形芽胞杆菌 (*Bacillus sphaericus*)

<400>16

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

Lys Cys Phe Asn Cys Glu Lys Glu Leu Thr Ile Glu Glu Val Arg Leu
 290 295 300
 Asp His Thr Arg Pro Leu Ala Tyr Leu Trp Pro Ile Asp Glu His Ala
 305 310 315 320
 Thr Cys Leu Cys Glu Lys Cys Asn Asn Thr Lys His Asp Met Phe Pro
 325 330 335
 Ile Asp Phe Tyr Gln Gly Asp Glu Asp Lys Leu Arg Arg Leu Ala Arg
 340 345 350
 Ile Thr Gly Leu Asp Tyr Glu Ser Leu Val Lys Arg Asp Val Asn Glu
 355 360 365
 Val Glu Leu Ala Arg Ile Ile Asn Asn Ile Glu Asp Phe Ala Thr Asn
 370 375 380
 Val Glu Ala Arg Thr Phe Arg Ser Ile Arg Asn Lys Val Lys Glu Val
 385 390 395 400
 Arg Pro Asp Thr Asp Leu Phe Glu Ile Leu Lys Ser Lys Asn Ile Asn
 405 410 415
 Leu Tyr Asn Glu Leu Gln Tyr Glu Leu Leu Thr Arg Lys Asp
 420 425 430

<210>17

<211>1245

<212>DNA

<213>产气肠杆菌 (Enterobacter aerogenes)

<400>17

gtgaatcaga aaaatgaaaa atcatttatg cgtttgcaat caacctttag cgggtggcaaa 60
 ggtagtccaa tgcattgattg gtacccatat ttagagggtt attctcccga atttgtgaaa 120
 tgcttgattt cacgatttgc tctaaagcc aaaacaattt tagatccatt ttgtggctct 180
 ggaacaacag ccattgtttc cgtttttagag ggtttaaata attactattg cgaagtaaac 240
 cttttatgcc aatatattat tgaactaaa ctaatagctt taacattaag cgaagaagaa 300
 aaaacaaat tagtaaata actttattct atttctaata aaataactaa tgtactcaaa 360
 ctttctgcaa ccgagacaga tctagagaaa tcatttaaata ccgtttttgg taatacggaaa 420
 ttttttgagg atcacatatt taaagatata cttagttatc aatgttacat tagctctatc 480
 gaagatgaaa atcttaagag acttctgaca atagcaggga ttagatcggt aatcccttcc 540
 tcgttattgg taagacgagg tgatttacga ttcaagacac aaaaagaatt agagaaaggc 600
 aaccagggct ttcgctttca tgtacaaaa agcttagaat taattgccag tgatttatta 660
 gacattacgg aaggtagtgg tttagctacc ttcttatgtg atgatgcaa agaaatatct 720
 gggaataacc tgattgatgc tgtaataaca agcccggcat atttaaattg cacaaattat 780
 tttagaaata ctaaaattga actttggttt atagggaaat taaagaccaa atcagatcta 840
 agacattata gggatttagc tattaccagt ggtattaacg atgtaactaa aggtaaaagc 900
 ttatcttcaa ataatactat tatctcagaa ataccattat tatctgaatg tattaaagaa 960

ctaagcataa aagagtatga tagtcgtatt tcaatgatgg ttgaaaacta cttttgggac 1020
 atgttcaaat tcttatcaaa acteccaaaa ttactaacta atgatgcgac tatctgtata 1080
 gatttaggtg attctgttta ttgtaacgtc tacatcccta cacaagatat tttgaaagaa 1140
 atgatgtcaa agttaggttt tgaagagaac gaaagggtca ttcttcgtga acgaaaatcc 1200
 cgcaatggaa caaagttagt ccagactggt caggttttta aatga 1245

<210>18

<211>1140

<212>DNA

<213> 产气肠杆菌

<400>18

atgaaaaata aatatttttag taaaaaatgg gagcaattca agaaagaatt accccatcaa 60
 tcaggtgaaa tggtaaagag aaattggggc cataactggc actctatgtg ttcataccaa 120
 gggaaactta aaccatcaat agctagatct ttaattgata cattcatgcc atcaagtaag 180
 ggacgtatat tagatgtctt ctcaggtggt ggcaccattc ctttcgaage aagattactt 240
 ggtcactactg catatggatt tgatattagt ccagcagcag ttaatatttc acgcgcaaaa 300
 ctagaagtta taagtaaaaa tgaatccaa gaggtaatta ataaattatc tgattttatt 360
 gagcaaaaca aaaattcaat agattataac gaacataatt taataagggt taatgggtca 420
 attgaatcct attttcatcc tgaactttt aaggaaatac tgtgtgctcg taaattcttt 480
 ttaataaaag gtgaattaaa tgcacttgaa tcgttagtac agtcatgttt attacatatt 540
 ttacatggta atcgtecgta tgcattgagt agaaagtccc atcctattac acctttcgcg 600
 cctactggag attttatata cagtaattta gttataaagt taatcaaaaa agttgaaaga 660
 gtcttgcaaa attctgatgg tatcccagat actggcagca aagtatttta tcaggactct 720
 acaaaaagtt ggcctgaaga agtaaataat ttagatgcaa ttataacatc acctccattt 780
 tatgatagta cccgtttcta ttcagcaaat tggatgcgat tatggttttc tggttgggaa 840
 aaagatgact tccaaacgaa gccaaaagat tttgtggacg aaactcagaa aaaaagcttt 900
 gaaatatatg ataatatatt caacaatct caacaatgct taaaaaaga tggcgttttt 960
 ttaatgcacg ttggcaaaag taaaaaaagt gatatggcag gacaaattgc taaaattggt 1020
 agtaattatc ttagccttat agatatattht gacgaaagtg ttgaacattg cgaaagtcac 1080
 ggaattaaag acaaaggcac gacaacccat catcagtacc ttgtctttac gaaagattag 1140

<210>19

<211>213

<212>PRT

<213> 解淀粉芽胞杆菌 (*Bacillus amyloliquefaciens*)H

<400>19

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

35 40 45
 Asn Thr Glu Lys Asn Cys Asn Gly Val Val Pro Ile Lys Glu Leu Cys
 50 55 60
 Tyr Thr Leu Leu Glu Asp Thr Tyr Asn Trp Tyr Arg Glu Lys Pro Leu
 65 70 75 80
 Asp Ile Leu Lys Leu Glu Lys Lys Lys Gly Gly Pro Ile Asp Val Tyr
 85 90 95
 Lys Glu Phe Ile Glu Asn Ser Glu Leu Lys Arg Val Gly Met Glu Phe
 100 105 110
 Glu Thr Gly Asn Ile Ser Ser Ala His Arg Ser Met Asn Lys Leu Leu
 115 120 125
 Leu Gly Leu Lys His Gly Glu Ile Asp Leu Ala Ile Ile Leu Met Pro
 130 135 140
 Ile Lys Gln Leu Ala Tyr Tyr Leu Thr Asp Arg Val Thr Asn Phe Glu
 145 150 155 160
 Glu Leu Glu Pro Tyr Phe Glu Leu Thr Glu Gly Gln Pro Phe Ile Phe
 165 170 175
 Ile Gly Phe Asn Ala Glu Ala Tyr Asn Ser Asn Val Pro Leu Ile Pro
 180 185 190
 Lys Gly Ser Asp Gly Met Ser Lys Arg Ser Ile Lys Lys Trp Lys Asp
 195 200 205
 Lys Val Glu Asn Lys
 210
 <210>20
 <211>194
 <212>PRT
 <213> 克氏海洋螺菌 (*Oceanospirillum kriegii*)
 <400>20
 Met Lys Ile Lys Arg Ile Glu Val Leu Ile Asn Asn Gly Ser Val Pro
 1 5 10 15
 Gly Ile Pro Met Ile Leu Asn Glu Ile Gln Asp Ala Ile Lys Thr Val
 20 25 30
 Ser Trp Pro Glu Gly Asn Asn Ser Phe Val Ile Asn Pro Val Arg Lys
 35 40 45
 Gly Asn Gly Val Lys Pro Ile Lys Asn Ser Cys Met Arg His Leu His
 50 55 60
 Gln Lys Gly Trp Ala Leu Glu His Pro Val Arg Ile Lys Ala Glu Met
 65 70 75 80
 Arg Pro Gly Pro Leu Asp Ala Val Lys Met Ile Gly Gly Lys Ala Phe

| | | | |
|---|-----|-----|-----|
| | 85 | 90 | 95 |
| Ala Leu Glu Trp Glu Thr Gly Asn Ile Ser Ser Ser His Arg Ala Ile | | | |
| | 100 | 105 | 110 |
| Asn Lys Met Val Met Gly Met Leu Glu Arg Val Ile Ile Gly Gly Val | | | |
| | 115 | 120 | 125 |
| Leu Ile Leu Pro Ser Arg Asp Met Tyr Asn Tyr Leu Thr Asp Arg Val | | | |
| | 130 | 135 | 140 |
| Gly Asn Phe Arg Glu Leu Glu Pro Tyr Phe Ser Val Trp Arg Gln Phe | | | |
| 145 | 150 | 155 | 160 |
| Asn Leu Lys Asp Ala Tyr Leu Ala Ile Val Glu Ile Glu His Asp Ser | | | |
| | 165 | 170 | 175 |
| Val Asp Ala Gln Val Ser Leu Ile Pro Lys Gly Thr Asp Gly Arg Ala | | | |
| | 180 | 185 | 190 |
| Ile Arg | | | |
| <210>21 | | | |
| <211>180 | | | |
| <212>PRT | | | |
| <213> 解淀粉芽胞杆菌 H | | | |
| <220> | | | |
| <221>MISC_FEATURE | | | |
| <222>(1)..(180) | | | |
| <223> 残基 1-180 相应于 BamHI (seq id no. 19) 的蛋白质序列的残基 22-201 | | | |
| <400>21 | | | |
| Ile Gln Gln Ala Tyr Asn Glu Val Lys Thr Ser Ile Cys Ser Pro Ile | | | |
| 1 | 5 | 10 | 15 |
| Trp Pro Ala Thr Ser Lys Thr Phe Thr Ile Asn Asn Thr Glu Lys Asn | | | |
| | 20 | 25 | 30 |
| Cys Asn Gly Val Val Pro Ile Lys Glu Leu Cys Tyr Thr Leu Leu Glu | | | |
| | 35 | 40 | 45 |
| Asp Thr Tyr Asn Trp Tyr Arg Glu Lys Pro Leu Asp Ile Leu Lys Leu | | | |
| | 50 | 55 | 60 |
| Glu Lys Lys Lys Gly Gly Pro Ile Asp Val Tyr Lys Glu Phe Ile Glu | | | |
| 65 | 70 | 75 | 80 |
| Asn Ser Glu Leu Lys Arg Val Gly Met Glu Phe Glu Thr Gly Asn Ile | | | |
| | 85 | 90 | 95 |
| Ser Ser Ala His Arg Ser Met Asn Lys Leu Leu Leu Gly Leu Lys His | | | |
| | 100 | 105 | 110 |
| Gly Glu Ile Asp Leu Ala Ile Ile Leu Met Pro Ile Lys Gln Leu Ala | | | |
| | 115 | 120 | 125 |

Tyr Tyr Leu Thr Asp Arg Val Thr Asn Phe Glu Glu Leu Glu Pro Tyr
 130 135 140
 Phe Glu Leu Thr Glu Gly Gln Pro Phe Ile Phe Ile Gly Phe Asn Ala
 145 150 155 160
 Glu Ala Tyr Asn Ser Asn Val Pro Leu Ile Pro Lys Gly Ser Asp Gly
 165 170 175
 Met Ser Lys Arg
 180

<210>22

<211>177

<212>PRT

<213> 克氏海洋螺菌

<220>

<221>MISC_FEATURE

<222>(1).. (177)

<223> 残基 1-177 相应于 OkrAI (seq id no. 20) 的蛋白质序列的残基 18-194

<400>22

Ile Pro Met Ile Leu Asn Glu Ile Gln Asp Ala Ile Lys Thr Val Ser
 1 5 10 15
 Trp Pro Glu Gly Asn Asn Ser Phe Val Ile Asn Pro Val Arg Lys Gly
 20 25 30
 Asn Gly Val Lys Pro Ile Lys Asn Ser Cys Met Arg His Leu His Gln
 35 40 45
 Lys Gly Trp Ala Leu Glu His Pro Val Arg Ile Lys Ala Glu Met Arg
 50 55 60
 Pro Gly Pro Leu Asp Ala Val Lys Met Ile Gly Gly Lys Ala Phe Ala
 65 70 75 80
 Leu Glu Trp Glu Thr Gly Asn Ile Ser Ser Ser His Arg Ala Ile Asn
 85 90 95
 Lys Met Val Met Gly Met Leu Glu Arg Val Ile Ile Gly Gly Val Leu
 100 105 110
 Ile Leu Pro Ser Arg Asp Met Tyr Asn Tyr Leu Thr Asp Arg Val Gly
 115 120 125
 Asn Phe Arg Glu Leu Glu Pro Tyr Phe Ser Val Trp Arg Gln Phe Asn
 130 135 140
 Leu Lys Asp Ala Tyr Leu Ala Ile Val Glu Ile Glu His Asp Ser Val
 145 150 155 160
 Asp Ala Gln Val Sar Leu Ile Pro Lys Gly Thr Asp Gly Arg Ala Ile
 165 170 175

| | |
|--|----|
| Arg | |
| <210>23 | |
| <211>38 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>23 | |
| attcaacaag catacaatgc agttaaaca tctattgt | 38 |
| <210>24 | |
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>24 | |
| acaaatagat gttttaactg cattgtatgc ttgttgaat | 39 |
| <210>25 | |
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>25 | |
| caagcataca atgaagttgc aacatctatt tgttcacct | 39 |
| <210>26 | |
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>26 | |
| aggtgaacaa atagatgttg caacttcatt gtatgcttg | 39 |
| <210>27 | |
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |

| | |
|---|----|
| <400>27 | |
| acgattaaca acaccgaage aaattgtaac ggtgtagta | 39 |
| <210>28 | |
| <211>40 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>28 | |
| acgattaaca acaccgaage aaattgtaac ggtgtagtat | 40 |
| <210>29 | |
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>29 | |
| aacggtgtag taccaattgc agaactatgt tacacctta | 39 |
| <210>30 | |
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>30 | |
| taaggtgtaa catagttctg caattggtac tacaccggt | 39 |
| <210>31 | |
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>31 | |
| aacccccttg atatacttgc acttgaaaag aaaaaaggt | 39 |
| <210>32 | |
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |

<223> 引物
 <400>32
 accttttttc ttttcaagtg caagtatate aaggggttt 39
 <210>33
 <211>36
 <212>DNA
 <213> 人工
 <220>
 <223> 引物
 <400>33
 gatatactta aacttgcaaa gaaaaaaggt ggtccg 36
 <210>34
 <211>39
 <212>DNA
 <213> 人工
 <220>
 <223> 引物
 <400>34
 cggaccacct tttttctttg caagtttaag tatatcaag 39
 <210>35
 <211>39
 <212>DNA
 <213> 人工
 <220>
 <223> 引物
 <400>35
 atacttaaac ttgaaaaggc aaaaggtggt ccgattgat 39
 <210>36
 <211>40
 <212>DNA
 <213> 人工
 <220>
 <223> 引物
 <400>36
 atcaatcgga ccaccttttg cctttttcaa gttaagtat 40
 <210>37
 <211>39
 <212>DNA
 <213> 人工

| | |
|--|----|
| <220> | |
| <223> 引物 | |
| <400>37 | |
| ggtccgattg atgtttatgc agagttcata gaaaacagt | 39 |
| <210>38 | |
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>38 | |
| actgttttct atgaactctg cataaacate aatcggacc | 39 |
| <210>39 | |
| <211>37 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>39 | |
| atagaaaaac agtgaacttg cacgtgtagg tatggaa | 37 |
| <210>40 | |
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>40 | |
| aaattccata cctacacgtg caagttcact gttttctat | 39 |
| <210>41 | |
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>41 | |
| ggaaatatta gttctgccgc acgttcaatg aacaaactt | 39 |
| <210>42 | |
| <211>39 | |
| <212>DNA | |

<213> 人工
 <220>
 <223> 引物
 <400>42
 aagtttgttc attgaaacgt gcggcagaac taatattcc 39
 <210>43
 <211>39
 <212>DNA
 <213> 人工
 <220>
 <223> 引物
 <400>43
 aatattagtt ctgcccacgc atcaatgaac aaactteta 39
 <210>44
 <211>39
 <212>DNA
 <213> 人工
 <220>
 <223> 引物
 <400>44
 tagaagtttg ttcattgatg cgtgggcaga actaatatt 39
 <210>45
 <211>42
 <212>DNA
 <213> 人工
 <220>
 <223> 引物
 <400>45
 gccaccggtt caatgaacgc acttctatta ggattaaaac at 42
 <210>46
 <211>42
 <212>DNA
 <213> 人工
 <220>
 <223> 引物
 <400>46
 atgttttaat cctaatagaa gtgcggtcat tgaacggtgg gc 42
 <210>47
 <211>39

| | |
|---|----|
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>47 | |
| attatcctta tgcctattgc acaattggcc tattatett | 39 |
| <210>48 | |
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>48 | |
| aagataatag gcccaattgtg caataggcat aaggataat | 39 |
| <210>49 | |
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>49 | |
| ttggcctatt atcttacagc acgtggtacc aatttcgag | 39 |
| <210>50 | |
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>50 | |
| ctcgaaattg gtaacacgtg ctgtaagata ataggccaa | 39 |
| <210>51 | |
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>51 | |
| gcctattate ttacagatgc agttaccaat ttcgaggaa | 39 |
| <210>52 | |

| | |
|---|----|
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>52 | |
| ttcctcgaaa ttgtaactg catctgtaag ataatagc | 39 |
| <210>53 | |
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>53 | |
| cgtgttacca atttcgagc attagaacct tattttgaa | 39 |
| <210>54 | |
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>54 | |
| ttcaaaataa gggttctaatg cctcgaaatt ggtaacacg | 39 |
| <210>55 | |
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>55 | |
| accaatttcg aggaattagc accttatttt gaacttact | 39 |
| <210>56 | |
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>56 | |
| agtaagttca aaataagggtg ctaattcctc gaaattggt | 39 |

| | |
|--|----|
| <210>57 | |
| <211>42 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>57 | |
| ccttatttttg aacttactgc aggacaacca tttattttta tt | 42 |
| <210>58 | |
| <211>42 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>58 | |
| aataaaaata aatggttgctg ctgcagtaag ttcaaaataa gg | 42 |
| <210>59 | |
| <211>45 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>59 | |
| tttattttta ttggatttaa tgctgcagct tataattcta atgctc | 45 |
| <210>60 | |
| <211>45 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>60 | |
| gacattagaa ttataagctg cagcattaa tccaataaaa ataaa | 45 |
| <210>61 | |
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>61 | |

| | |
|---|----|
| aatgtccctt taattcccgc aggtttctgac ggtatgtca | 39 |
| <210>62 | |
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>62 | |
| tgacataccg tcagaacctg cgggaattaa agggacatt | 39 |
| <210>63 | |
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>63 | |
| ttaattccca aaggtttctgc aggtatgtca aaacgctca | 39 |
| <210>64 | |
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>64 | |
| tgagcgtttt gacataacctg cagaaccttt gggaattaa | 39 |
| <210>65 | |
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |
| <400>65 | |
| tctgacggta tgtcaaaagc atcaattaag aaatggaaa | 39 |
| <210>66 | |
| <211>39 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223> 引物 | |

<400>66
 ttccatttc ttaattgatg cttttgacat accgtcaga 39
 <210>67
 <211>54
 <212>DNA
 <213> 人工
 <220>
 <223> 引物
 <400>67
 ggtggtgcat gcggaggtaa ataaatggaa gtagaaaaag agtttattac tgat 54
 <210>68
 <211>51
 <212>DNA
 <213> 人工
 <220>
 <223> 引物
 <400>68
 ggtggtggta ccctatttgt ttccaacttt atctttccat ttcttaattg a 51
 <210>69
 <211>39
 <212>DNA
 <213> 人工
 <220>
 <223> 引物
 <220>
 <221>misc_feature
 <222>(19).. (21)
 <223>n = a, c, g 或 t
 <400>69
 caagcataca atgaagttnn nacatctatt tgttcacct 39
 <210>70
 <211>39
 <212>DNA
 <213> 人工
 <220>
 <223> 引物
 <220>
 <221>misc_feature
 <222>(19).. (21)

<223>n = a, c, g 或 t
 <400>70
 aggtgaacaa atagatgttn naacttcatt gtatgcttg 39
 <210>71
 <211>37
 <212>DNA
 <213> 人工
 <220>
 <223> 引物
 <220>
 <221>misc_feature
 <222>(16).. (18)
 <220>
 <221>misc_feature
 <222>(16).. (18)
 <223>n = a, c, g 或 t
 <400>71
 gatatactta aacttnnaa gaaaaaagg tgggccg 37
 <210>72
 <211>39
 <212>DNA
 <213> 人工
 <220>
 <223> 引物
 <220>
 <221>misc_feature
 <222>(19).. (21)
 <220>
 <221>misc_feature
 <222>(19).. (21)
 <223>n = a, c, g 或 t
 <400>72
 cggaccacct tttttcttn naagtttaag tatatcaag 39
 <210>73
 <211>54
 <212>DNA
 <213> 人工
 <220>
 <223> 引物

<400>73

ggtggtgcat gcggaggtaa ataatgtct aataaaaaac agtcaaatag gcta 54

<210>74

<211>39

<212>DNA

<213> 人工

<220>

<223> 引物

<400>74

ggtggtggta cctcacttag atctaagctg ttcaaaca 39

<210>75

<211>267

<212>PRT

<213> 大肠杆菌 (Escherichia coli)RY13

<220>

<221>MISC_FEATURE

<222>(1).. (267)

<223> 氨基酸残基 1-267 相应于 EcoRI 的蛋白质序列的残基 4-270

<400>75

Lys Lys Gln Ser Asn Arg Leu Thr Glu Gln His Lys Leu Ser Gln Gly

1 5 10 15

Val Ile Gly Ile Phe Gly Asp Tyr Ala Lys Ala His Asp Leu Ala Val

20 25 30

Gly Glu Val Ser Lys Leu Val Lys Lys Ala Leu Ser Asn Glu Tyr Pro

35 40 45

Gln Leu Ser Phe Arg Tyr Arg Asp Ser Ile Lys Lys Thr Glu Ile Asn

50 55 60

Glu Ala Leu Lys Lys Ile Asp Pro Asp Leu Gly Gly Thr Leu Phe Val

65 70 75 80

Ser Asn Ser Ser Ile Lys Pro Asp Gly Gly Ile Val Glu Val Lys Asp

85 90 95

Asp Tyr Gly Glu Trp Arg Val Val Leu Val Ala Glu Ala Lys His Gln

100 105 110

Gly Lys Asp Ile Ile Asn Ile Arg Asn Gly Leu Leu Val Gly Lys Arg

115 120 125

Gly Asp Gln Asp Leu Met Ala Ala Gly Asn Ala Ile Glu Arg Ser His

130 135 140

Lys Asn Ile Ser Glu Ile Ala Asn Phe Met Leu Ser Glu Ser His Phe

145 150 155 160

Pro Tyr Val Leu Phe Leu Glu Gly Ser Asn Phe Leu Thr Glu Asn Ile
 165 170 175
 Ser Ile Thr Arg Pro Asp Gly Arg Val Val Asn Leu Glu Tyr Asn Ser
 180 185 190
 Gly Ile Leu Asn Arg Leu Asp Arg Leu Thr Ala Ala Asn Tyr Gly Met
 195 200 205
 Pro Ile Asn Ser Asn Leu Cys Ile Asn Lys Phe Val Asn His Lys Asp
 210 215 220
 Lys Ser Ile Met Leu Gln Ala Ala Ser Ile Tyr Thr Gln Gly Asp Gly
 225 230 235 240
 Arg Glu Trp Asp Ser Lys Ile Met Phe Glu Ile Met Phe Asp Ile Ser
 245 250 255
 Thr Thr Ser Leu Arg Val Leu Gly Arg Asp Leu
 260 265

<210>76

<211>265

<212>PRT

<213> 类球红细菌 (*Rhodopseudomonas sphaeroides*)

<220>

<221>MISC_FEATURE

<222>(1)..(265)

<223> 氨基酸残基 1-265 相应于 RsrI 的蛋白质序列的残基 10-274

<400>76

Lys Gly Gln Ala Leu Arg Leu Gly Ile Gln Gln Glu Leu Gly Gly Gly
 1 5 10 15
 Pro Leu Ser Ile Phe Gly Ala Ala Ala Gln Lys His Asp Leu Ser Ile
 20 25 30
 Arg Glu Val Thr Ala Gly Val Leu Thr Lys Leu Ala Glu Asp Phe Pro
 35 40 45
 Asn Leu Glu Phe Gln Leu Arg Thr Ser Leu Thr Lys Lys Ala Ile Asn
 50 55 60
 Glu Lys Leu Arg Ser Phe Asp Pro Arg Leu Gly Gln Ala Leu Phe Val
 65 70 75 80
 Glu Ser Ala Ser Ile Arg Pro Asp Gly Gly Ile Thr Glu Val Lys Asp
 85 90 95
 Arg His Gly Asn Trp Arg Val Ile Leu Val Gly Glu Ser Lys His Gln
 100 105 110
 Gly Asn Asp Val Glu Lys Ile Leu Ala Gly Val Leu Gln Gly Lys Ala
 115 120 125

Lys Asp Gln Asp Phe Met Ala Ala Gly Asn Ala Ile Glu Arg Met His
 130 135 140
 Lys Asn Val Leu Glu Leu Arg Asn Tyr Met Leu Asp Glu Lys His Phe
 145 150 155 160
 Pro Tyr Val Val Phe Leu Gln Gly Ser Asn Phe Ala Thr Glu Ser Phe
 165 170 175
 Glu Val Thr Arg Pro Asp Gly Arg Val Val Lys Ile Val His Asp Ser
 180 185 190
 Gly Met Leu Asn Arg Ile Asp Arg Val Thr Ala Ser Ser Leu Ser Arg
 195 200 205
 Glu Ile Asn Gln Asn Tyr Cys Glu Asn Ile Val Val Arg Ala Gly Ser
 210 215 220
 Phe Asp His Met Phe Gln Ile Ala Ser Leu Tyr Cys Lys Ala Ala Pro
 225 230 235 240
 Trp Thr Ala Gly Glu Met Ala Glu Ala Met Leu Ala Val Ala Lys Thr
 245 250 255
 Ser Leu Arg Ile Ile Ala Asp Asp Leu
 260 265

<210>77

<211>39

<212>DNA

<213> 人工

<220>

<223> 引物

<400>77

gattgggtgg cgcagaaatt tcaaacgggc cagcagtcg 39

<210>78

<211>39

<212>DNA

<213> 人工

<220>

<223> 引物

<400>78

cgactgctgg cccgtttgaa atttctgcgc cacccaatc 39

<210>79

<211>272

<212>PRT

<213> 食明胶土壤杆菌 (Agrobacterium gelatinovorum)

<400>79

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

<210>80

<211>358

<212>PRT

<213> 鱼腥藻 (Anabaena variabilis)

<400>80

| | | | |
|---|-----|-----|-----|
| 305 | 310 | 315 | 320 |
| Thr Arg Val Ser Gln Val Ala Glu Ala Ile Arg Ala Tyr Leu Asp Gly | | | |
| | 325 | 330 | 335 |
| Asp Lys Thr Arg Leu Lys Trp Leu Val Asn Phe Sar Ile Glu Glu Ala | | | |
| | 340 | 345 | 350 |
| Asp His Asp Asn Ile Thr | | | |
| | 355 | | |
| <210>81 | | | |
| <211>530 | | | |
| <212>PRT | | | |
| <213>嗜热脂肪芽胞杆菌 (<i>Bacillus stearothermophilus</i>)B61 | | | |
| <400>81 | | | |
| Met Ala Lys Tyr Gly Arg Gly Lys Phe Leu Pro His Gln Asn Tyr Ile | | | |
| 1 | 5 | 10 | 15 |
| Asp Tyr Met His Phe Ile Val Asn His Lys Asn Tyr Ser Gly Met Pro | | | |
| | 20 | 25 | 30 |
| Asn Ala Ile Gly Glu Asp Gly Arg Ile Asn Trp Gln Val Ser Ser Gly | | | |
| | 35 | 40 | 45 |
| Lys Thr Thr Ser Phe Tyr Glu Tyr Tyr Gln Ala Arg Phe Glu Trp Trp | | | |
| | 50 | 55 | 60 |
| Glu Lys Lys Ala Asp Glu Leu Asn Leu Pro Gly Thr Gly Asn Ser Asn | | | |
| 65 | 70 | 75 | 80 |
| Lys Arg Phe Ser Leu Ala Ala Arg Leu Ile His Pro Thr Gly Gln Arg | | | |
| | 85 | 90 | 95 |
| Pro Cys Arg Leu Cys Gly Lys Tyr Gln Tyr Val Gly Tyr Met Tyr Val | | | |
| | 100 | 105 | 110 |
| Ser His Asn Leu Tyr Lys Arg Trp Ser Lys Ile Thr Gly Arg Glu Asp | | | |
| | 115 | 120 | 125 |
| Leu Phe Phe Lys Lys Gln Asn Ile Ile Glu Ala Ala Asn Ile Phe Lys | | | |
| | 130 | 135 | 140 |
| Ser Ile Met Gly Glu Gln Ala Leu Ile Asn Glu Leu Thr Thr Ile Phe | | | |
| 145 | 150 | 155 | 160 |
| Pro Glu Arg Lys Asp Tyr Phe Asn Arg Leu Pro Asn Ile Glu Asp Phe | | | |
| | 165 | 170 | 175 |
| Phe Val Ser Ser Ser His Ile Lys Asn Asn Gly Asn Tyr Ile Ser Pro | | | |
| | 180 | 185 | 190 |
| Gly Phe Met Ala Asn Pro Pro Asp Arg Leu Asp Gly Phe His Asp Tyr | | | |
| | 195 | 200 | 205 |
| Gly Ile Cys Cys Arg Lys Glu Lys Asp Pro Gly Arg His Asp Asp Asn | | | |

| | | |
|---|-----|-----|
| 210 | 215 | 220 |
| Met Arg Leu Tyr Asn His Asp Arg Arg Ala Phe Met Trp Trp Ser Glu | | |
| 225 | 230 | 235 |
| Gly Asp Trp Ala Leu Ala Asp Ala Leu Tyr Asn Lys Ala Gly Ala Gly | | 240 |
| | 245 | 250 |
| Lys Cys Ala Asp Pro Asp Cys Gln Lys Glu Val Glu Lys Ile Ser Pro | | 255 |
| | 260 | 265 |
| Asp His Val Gly Pro Ile Ser Cys Gly Phe Lys Gln Ile Pro Phe Phe | | 270 |
| | 275 | 280 |
| Lys Pro Leu Cys Ala Ser Cys Asn Ser Ala Lys Asn Arg Arg Phe Ser | | 285 |
| 290 | 295 | 300 |
| Tyr Gln Asp Val Lys Glu Leu Leu Lys Tyr Glu Asn Tyr Thr Gly Asp | | |
| 305 | 310 | 315 |
| Ser Val Ala Ser Trp Gln Val Arg Ala Leu Trp Asp Asn Cys Lys His | | 320 |
| | 325 | 330 |
| Leu Val Lys Asn Asp Asp Asp Ser Lys Leu Leu Ser Asn Leu Met Arg | | 335 |
| | 340 | 345 |
| Ser Leu Gln Asp Tyr Tyr Leu Arg Ser Leu Tyr Lys Leu Phe Ser Asn | | 350 |
| 355 | 360 | 365 |
| Gly Phe Ala His Leu Leu Ser Tyr Phe Leu Thr Pro Glu Tyr Ala His | | |
| 370 | 375 | 380 |
| Tyr Lys Ile Thr Phe Glu Gly Leu Asn Thr Ser Thr Leu Glu Tyr Glu | | |
| 385 | 390 | 395 |
| Arg Tyr Tyr Lys Thr Phe Lys Lys Thr Lys Ser Thr Ser Ser Leu Ala | | |
| | 405 | 410 |
| Ala Arg Ile Val Arg Ile Ala Phe Glu Glu Leu Glu Ile Tyr Asn Ser | | 415 |
| | 420 | 425 |
| Lys Asp Ile Asn Glu Arg Lys Leu Ile Lys Phe Asp Thr Ser Ser Trp | | 430 |
| 435 | 440 | 445 |
| Glu Lys Asp Phe Glu Asn Ile Ile Ser Tyr Ala Thr Lys Asn Leu Ser | | |
| 450 | 455 | 460 |
| Leu Asp Glu Glu Ala Ser Lys Trp Asn Lys Val Leu Thr Asp Lys Asn | | |
| 465 | 470 | 475 |
| Leu Ser Ser Thr Glu Lys Asp Lys Lys Ile Ser Ser Leu Leu Glu Asp | | |
| | 485 | 490 |
| Lys Asn Tyr Glu Val Tyr Lys Lys Gln Phe Tyr Ile Leu Lys Asp Leu | | |
| | 500 | 505 |
| Leu Val Glu His Phe Asn Lys Ile Gly Glu Gln Ile Ala Lys Asp Tyr | | 510 |
| 515 | 520 | 525 |

Met Lys
 530
 <210>82
 <211>301
 <212>PRT
 <213> 聚团肠杆菌 (Enterobacter agglomerans)
 <400>82
 Met Lys Lys Arg Arg Asp Leu Val Glu Val Phe Gly Tyr Asn Pro Met
 1 5 10 15
 Asp Leu Ser Pro Glu Val Arg Ala Leu Trp Asn Leu Gly Ala Cys Pro
 20 25 30
 Phe Leu Asn Lys Glu Cys Ile Lys Ile Asn His Asp Gln Thr Ile Ile
 35 40 45
 Tyr Gly Thr Cys Ser Val Thr Ser Pro Tyr Gly Asp Val Ile Ile Cys
 50 55 60
 Pro Asn Arg Leu Tyr Ala Asn Asp Tyr Glu Thr Leu His Lys Val Ser
 65 70 75 80
 Arg Asp Ala Phe Gly Asp Asp Val Pro Phe Leu Thr Tyr Ser Asn Phe
 85 90 95
 Ile Lys Tyr Arg Ala Thr Tyr Lys Asp Cys Ile Val Ala Leu Gly Lys
 100 105 110
 Asn Ser Gly Lys Glu Val Gln Val Gly Arg Ala Leu Ser Met Asp Trp
 115 120 125
 Val Leu Val Arg Ile Thr Asp Gly Glu Leu Lys Glu Tyr Val Gly Val
 130 135 140
 Glu Ile Gln Ser Ile Asp Ile Thr Gly Asn Tyr Arg Asp Ala Trp His
 145 150 155 160
 Ala Tyr Lys Asn Leu Lys Pro Ile Asp Ile Ile Asp Asn Leu Pro Thr
 165 170 175
 Ser Gln His Gly Leu Asn Trp Ala Asn Val His Lys Arg Leu Ile Pro
 180 185 190
 Gln Ile Ile Arg Lys Gly Val Val Tyr Ser Arg Ser Asn Tyr Val Lys
 195 200 205
 Lys Gly Leu Tyr Phe Ile Leu Pro Glu Ile Val Tyr Asn Lys Phe Glu
 210 215 220
 Asp Val Ile Gly Ala Asp Ile Pro Leu Leu Lys Thr Gln Thr Asn Lys
 225 230 235 240
 Ser Ile Thr Val His Thr Tyr Ser Leu Gly Glu Pro Ala Ala Asn Gly
 245 250 255

Glu Gln Arg Lys Leu Ile Ser Glu Arg Glu Ile Ile Phe Asp Leu Asp
 260 265 270
 Glu Phe Ser Lys Arg Phe Thr Thr Gly Pro Asn Leu Pro Lys Gly Asp
 275 280 285
 Asp Leu Asp Ala Val Ile Lys Lys Ala Leu Gly Met Met
 290 295 300
 <210>83
 <211>277
 <212>PRT
 <213> 大肠杆菌 RY13
 <400>83
 Met Ser Asn Lys Lys Gln Ser Asn Arg Leu Thr Glu Gln His Lys Leu
 1 5 10 15
 Ser Gln Gly Val Ile Gly Ile Phe Gly Asp Tyr Ala Lys Ala His Asp
 20 25 30
 Leu Ala Val Gly Glu Val Ser Lys Leu Val Lys Lys Ala Leu Ser Asn
 35 40 45
 Glu Tyr Pro Gln Leu Ser Phe Arg Tyr Arg Asp Ser Ile Lys Lys Thr
 50 55 60
 Glu Ile Asn Glu Ala Leu Lys Lys Ile Asp Pro Asp Leu Gly Gly Thr
 65 70 75 80
 Leu Phe Val Ser Asn Ser Ser Ile Lys Pro Asp Gly Gly Ile Val Glu
 85 90 95
 Val Lys Asp Asp Tyr Gly Glu Trp Arg Val Val Leu Val Ala Glu Ala
 100 105 110
 Lys His Gln Gly Lys Asp Ile Ile Asn Ile Arg Asn Gly Leu Leu Val
 115 120 125
 Gly Lys Arg Gly Asp Gln Asp Leu Met Ala Ala Gly Asn Ala Ile Glu
 130 135 140
 Arg Ser His Lys Asn Ile Ser Glu Ile Ala Asn Phe Met Leu Ser Glu
 145 150 155 160
 Ser His Phe Pro Tyr Val Leu Phe Leu Glu Gly Ser Asn Phe Leu Thr
 165 170 175
 Glu Asn Ile Ser Ile Thr Arg Pro Asp Gly Arg Val Val Asn Leu Glu
 180 185 190
 Tyr Asn Ser Gly Ile Leu Asn Arg Leu Asp Arg Leu Thr Ala Ala Asn
 195 200 205
 Tyr Gly Met Pro Ile Asn Ser Asn Leu Cys Ile Asn Lys Phe Val Asn
 210 215 220

| | | |
|---|-----|-----|
| 195 | 200 | 205 |
| Glu Asp Glu Phe Leu Asp Tyr Trp Arg Asn Tyr Glu Arg Thr Ser Gln | | |
| 210 | 215 | 220 |
| Leu Arg Asn Asp Lys Tyr Asn Asn Ile Ser Glu Tyr Arg Asn Trp Ile | | |
| 225 | 230 | 235 |
| Tyr Arg Gly Arg Lys | | 240 |
| | 245 | |
| <210>85 | | |
| <211>300 | | |
| <212>PRT | | |
| <213> 流感嗜血杆菌 Rd(外-突变体)(Haemophilus influenzae Rd(exo-mutant)) | | |
| <400>85 | | |
| Met Lys Lys Ser Ala Leu Glu Lys Leu Leu Ser Leu Ile Glu Asn Leu | | |
| 1 | 5 | 10 |
| Thr Asn Gln Glu Phe Lys Gln Ala Thr Asn Ser Leu Ile Ser Phe Ile | | |
| | 20 | 25 |
| Tyr Lys Leu Asn Arg Asn Glu Val Ile Glu Leu Val Arg Ser Ile Gly | | |
| | 35 | 40 |
| Ile Leu Pro Glu Ala Ile Lys Pro Ser Ser Thr Gln Glu Lys Leu Phe | | |
| | 50 | 55 |
| Ser Lys Ala Gly Asp Ile Val Leu Ala Lys Ala Phe Gln Leu Leu Asn | | |
| 65 | 70 | 75 |
| Leu Asn Ser Lys Pro Leu Glu Gln Arg Gly Asn Ala Gly Asp Val Ile | | |
| | 85 | 90 |
| Ala Leu Ser Lys Glu Phe Asn Tyr Gly Leu Val Ala Asp Ala Lys Ser | | |
| | 100 | 105 |
| Phe Arg Leu Ser Arg Thr Ala Lys Asn Gln Lys Asp Phe Lys Val Lys | | |
| | 115 | 120 |
| Ala Leu Ser Glu Trp Arg Glu Asp Lys Asp Tyr Ala Val Leu Thr Ala | | |
| | 130 | 135 |
| Pro Phe Phe Gln Tyr Pro Thr Thr Lys Ser Gln Ile Phe Lys Gln Ser | | |
| 145 | 150 | 155 |
| Leu Asp Glu Asn Val Leu Leu Phe Ser Trp Glu His Leu Ala Ile Leu | | |
| | 165 | 170 |
| Leu Gln Leu Asp Leu Glu Glu Thr Asn Ile Phe Pro Phe Glu Gln Leu | | |
| | 180 | 185 |
| Trp Asn Phe Pro Lys Lys Gln Ser Lys Lys Thr Ser Val Ser Asp Ala | | |
| | 195 | 200 |
| Glu Asn Asn Phe Met Arg Asp Phe Asn Lys Tyr Phe Met Asp Leu Phe | | 205 |

| | | | | | |
|---|-----|-----|-----|--|--|
| 210 | 215 | 220 | | | |
| Lys Ile Asp Lys Asp Thr Leu Asn Gln Leu Leu Gln Lys Glu Ile Asn | | | | | |
| 225 | 230 | 235 | 240 | | |
| Phe Ile Glu Glu Arg Ser Leu Ile Glu Lys Glu Tyr Trp Lys Lys Gln | | | | | |
| | 245 | 250 | 255 | | |
| Ile Asn Ile Ile Lys Asn Phe Thr Arg Glu Glu Ala Ile Glu Ala Leu | | | | | |
| | 260 | 265 | 270 | | |
| Leu Lys Asp Ile Asn Met Ser Ser Lys Ile Glu Thr Ile Asp Ser Phe | | | | | |
| | 275 | 280 | 285 | | |
| Ile Lys Gly Ile Lys Ser Asn Asp Arg Leu Tyr Leu | | | | | |
| 290 | 295 | 300 | | | |
| <210>86 | | | | | |
| <211>254 | | | | | |
| <212>PRT | | | | | |
| <213> 副流感嗜血杆菌 (<i>Haemophilus parainfluenzae</i>) | | | | | |
| <400>86 | | | | | |
| Met Lys Tyr Glu Glu Ile Asn Phe Lys Val Pro Val Glu Ser Pro Tyr | | | | | |
| 1 | 5 | 10 | 15 | | |
| Tyr Pro Asn Tyr Ser Gln Cys Val Ile Glu Arg Ile Tyr Ser Ile Leu | | | | | |
| | 20 | 25 | 30 | | |
| Arg Asn Gln Lys Asp Met Gly Asp Asp Arg Ile Ile Ile Asn Thr Asn | | | | | |
| | 35 | 40 | 45 | | |
| Leu Lys Lys Gly Leu Pro Leu Glu Asn Ile Asn Lys Ile Ala Gly Pro | | | | | |
| | 50 | 55 | 60 | | |
| Met Ile Glu Ala Trp Ala Glu Glu Val Phe Ser Gly Ile Arg Asp Asn | | | | | |
| 65 | 70 | 75 | 80 | | |
| Arg Asp Asn Gln Tyr Asn Leu Ile Asn Val Glu Ala Gln Glu Arg Leu | | | | | |
| | 85 | 90 | 95 | | |
| Gly Ile Ser Asp Ile Ile Leu Gln Phe Gln Val Asn Asn Asn Val Ile | | | | | |
| | 100 | 105 | 110 | | |
| Thr Gly Asn Val Asp Val Lys Ala Thr Ser Asn Asp Ile Pro Asp Ser | | | | | |
| | 115 | 120 | 125 | | |
| Gly Lys Ser Pro Asn Ile Thr Ser Phe Ser Arg Ile Arg Thr Ala Tyr | | | | | |
| | 130 | 135 | 140 | | |
| Val Lys Asp Pro Asn Phe Ile Phe Ile Ile Leu Ser Ile Lys His Ser | | | | | |
| 145 | 150 | 155 | 160 | | |
| Val Tyr Val Lys Arg Asn Glu Tyr Thr Asn Leu Met Asp Gly Ile Met | | | | | |
| | 165 | 170 | 175 | | |
| Gln Ile Ile Asp Phe Asn Val Tyr Asp Leu Lys Tyr Ile Ser Asp Ser | | | | | |

| | | | | | |
|---|---|-----|-----|-----|-----|
| | 180 | | 185 | | 190 |
| Asp | Ile Ser Tyr Asn Pro Ala Leu Gly Thr Gly Gln Ile Gln Ile Lys | | | | |
| | 195 | | 200 | | 205 |
| Asp | Ile His Tyr Val Ser Ser Gln Lys Arg Thr Thr Trp Gln Met Cys | | | | |
| | 210 | | 215 | | 220 |
| Gln | Leu Leu Asp Leu Lys Tyr Leu Arg Ser Lys Lys Arg Thr Ile Glu | | | | |
| 225 | | 230 | | 235 | 240 |
| Gln | Phe Tyr Asn Glu Ala Lys Arg Asn Lys Trp Ile Lys Asp | | | | |
| | 245 | | 250 | | |
| <210>87 | | | | | |
| <211>218 | | | | | |
| <212>PRT | | | | | |
| <213> 克雷白氏肺炎菌 (<i>Klebsiella pnermoniae</i>)OK8 | | | | | |
| <400>87 | | | | | |
| Met | Asp Val Phe Asp Lys Val Tyr Ser Asp Asp Asn Asn Ser Tyr Asp | | | | |
| 1 | | 5 | | 10 | 15 |
| Gln | Lys Thr Val Ser Gln Arg Ile Glu Ala Leu Phe Leu Asn Asn Leu | | | | |
| | 20 | | 25 | | 30 |
| Gly | Lys Val Val Thr Arg Gln Gln Ile Ile Arg Ala Ala Thr Asp Pro | | | | |
| | 35 | | 40 | | 45 |
| Lys | Thr Gly Lys Gln Pro Glu Asn Trp His Gln Arg Leu Ser Glu Leu | | | | |
| | 50 | | 55 | | 60 |
| Arg | Thr Asp Lys Gly Tyr Thr Ile Leu Ser Trp Arg Asp Met Lys Val | | | | |
| 65 | | 70 | | 75 | 80 |
| Leu | Ala Pro Gln Glu Tyr Ile Met Pro His Ala Thr Arg Arg Pro Lys | | | | |
| | | 85 | | 90 | 95 |
| Ala | Ala Lys Arg Val Leu Pro Thr Lys Glu Thr Trp Glu Gln Val Leu | | | | |
| | 100 | | 105 | | 110 |
| Asp | Arg Ala Asn Tyr Ser Cys Glu Trp Gln Glu Asp Gly Gln His Cys | | | | |
| | 115 | | 120 | | 125 |
| Gly | Leu Val Glu Gly Asp Ile Asp Pro Ile Gly Gly Gly Thr Val Lys | | | | |
| | 130 | | 135 | | 140 |
| Leu | Thr Pro Asp His Met Thr Pro His Ser Ile Asp Pro Ala Thr Asp | | | | |
| 145 | | 150 | | 155 | 160 |
| Val | Asn Asp Pro Lys Met Trp Gln Ala Leu Cys Gly Arg His Gln Val | | | | |
| | | 165 | | 170 | 175 |
| Met | Lys Lys Asn Tyr Trp Asp Ser Asn Asn Gly Lys Ile Asn Val Ile | | | | |
| | 180 | | 185 | | 190 |
| Gly | Ile Leu Gln Ser Val Asn Glu Lys Gln Lys Asn Asp Ala Leu Glu | | | | |

| | | |
|---|-----|-----|
| 195 | 200 | 205 |
| Phe Leu Leu Asn Tyr Tyr Gly Leu Lys Arg | | |
| 210 | 215 | |
| <210>88 | | |
| <211>288 | | |
| <212>PRT | | |
| <213> 珊瑚红诺卡氏菌 (<i>Nocardia corallina</i>) | | |
| <400>88 | | |
| Met Ala Thr Ala Pro Gly His Leu Leu Gly Gln Ile Ile Gly Asn Val | | |
| 1 | 5 | 10 |
| Met Glu Glu Ala Leu Lys Pro Val Leu Gln Glu Met Ala Asp Arg His | | |
| | 20 | 25 |
| Asp Leu Tyr Leu Asp Ser Lys Gly Leu Arg Pro Gly Val Arg Ser Gly | | |
| | 35 | 40 |
| Ala Leu Val Thr Trp Thr Asp Asp Leu Gly Asn Asn His Asp Leu Asp | | |
| | 50 | 55 |
| Phe Val Leu Glu Arg Gly Gly Ser Ala Thr Lys Ala Gly Asn Pro Ala | | |
| 65 | 70 | 75 |
| Ala Phe Ile Glu Ala Ala Trp Arg Arg Tyr Thr Lys His Ser Lys Ala | | |
| | 85 | 90 |
| Lys Ala Gln Glu Ile Gln Gly Ala Val Leu Pro Val Leu Ala Ala Trp | | |
| | 100 | 105 |
| Asn Asn Val Lys Pro Thr Pro Ala Ala Val Val Ala Gly Gln Trp Thr | | |
| | 115 | 120 |
| Ala Pro Ser Leu Gln Gln Met Arg Ser Asn Gly Phe Val Val Leu His | | |
| | 130 | 135 |
| Leu His Phe Pro Thr Thr Ala Gln Val Phe Gly Gly Asn Gly Ile Asn | | |
| 145 | 150 | 155 |
| Ile Glu Gly Thr Gly Glu Gly Thr Pro Asp Ala Phe Trp Gln Gln Gln | | |
| | 165 | 170 |
| Cys Asp Ala Tyr Thr Ser Lys Ser Glu Ala Asp Lys Asp Ser Leu Ala | | |
| | 180 | 185 |
| Thr Ala Leu Arg Thr Ala His Ala Gln Glu Phe Arg Thr Phe Val Ala | | |
| | 195 | 200 |
| Glu Leu Glu Arg Arg Val Val Arg Ala Ile Asp Tyr Val Val Val Thr | | |
| | 210 | 215 |
| Pro Leu His Gly His Gly Ser Gln Tyr Thr Ser Ile Glu Asn Ala Ile | | |
| 225 | 230 | 235 |
| Glu Ala Val Arg Thr Tyr Ser Cys Gly Glu Glu Ser Ala Pro Phe Leu | | |

| | | | |
|---|-----|-----|-----|
| | 245 | 250 | 255 |
| Arg Phe Glu Ile Arg Ile Ser Tyr Thr Asn Gly Asp Val Ile Gln Ala | | | |
| | 260 | 265 | 270 |
| Thr Phe Gly Ser Ser Ser Asp Ala Ile Glu Phe Leu Asp Thr Phe Asn | | | |
| | 275 | 280 | 285 |
| <210>89 | | | |
| <211>328 | | | |
| <212>PRT | | | |
| <213> 粘膜奈瑟氏菌 (<i>Neisseria mucosa</i>) | | | |
| <400>89 | | | |
| Met Ser Ser Tyr His Asp Asp Leu Asn Ile Leu Asn Val Asp Phe Asn | | | |
| 1 | 5 | 10 | 15 |
| His Leu Arg Leu Thr Glu Leu Ile Lys Leu Ala Asp Gln Ala Glu Pro | | | |
| | 20 | 25 | 30 |
| Phe Tyr Leu Trp Val Glu Lys Ile Phe Arg Gln Val Ser Gly Arg Ala | | | |
| | 35 | 40 | 45 |
| Asp Ser Leu Glu Thr Ile Ile Glu Val Glu Glu Arg Val Val Leu Lys | | | |
| | 50 | 55 | 60 |
| Met Ala Ile Leu Thr Cys Phe Thr Ser Asp Glu Lys Glu Leu Pro Lys | | | |
| 65 | 70 | 75 | 80 |
| Leu Phe Asn Gly Val Gly Val Pro Tyr Pro His Ile Lys Ala Cys Tyr | | | |
| | 85 | 90 | 95 |
| Phe Phe Phe Ala Trp Leu Val Arg Asp Ala Ala Thr Gln Arg Leu Asp | | | |
| | 100 | 105 | 110 |
| Pro Leu Ile Arg Glu Ala Phe Thr Gln Leu Lys Ser Ile His Pro Gln | | | |
| | 115 | 120 | 125 |
| Met Lys Lys Thr Glu Leu Glu Ser Glu Ile Phe Ser Gln Leu Leu Val | | | |
| | 130 | 135 | 140 |
| Asn Tyr Arg Asn Glu Leu Ile His Phe Ser Trp Pro Val Ile Arg Glu | | | |
| 145 | 150 | 155 | 160 |
| Val Leu Ile Ser Arg Leu Glu Gly Ser Arg Arg Ala Ala Arg Gly Ser | | | |
| | 165 | 170 | 175 |
| Tyr Leu Glu Leu Phe Val Arg Thr Ala Leu Ala Gln Ser Ile Thr Tyr | | | |
| | 180 | 185 | 190 |
| Phe Tyr Lys Ile Tyr Gly Asn Tyr Gly Lys Phe Leu Asp Val Lys Ile | | | |
| | 195 | 200 | 205 |
| His Asp Lys Pro Leu Lys Val Lys Asn Arg Thr Tyr Asp Val Val Ala | | | |
| | 210 | 215 | 220 |
| Glu Leu Ile Gly Asn Asn His Asn Thr Gln Tyr Leu Ile Leu Pro Val | | | |

| | | | |
|---|-----|-----|-----|
| | 165 | 170 | 175 |
| Arg Tyr Gly Val Leu Glu Ile Gln Thr Met Asp Phe His Gly Ser Tyr | | | |
| | 180 | 185 | 190 |
| Lys His Ala Val Gly Ala Ile Asp Ile Ala Leu Val Glu Gly Ile Asp | | | |
| | 195 | 200 | 205 |
| Phe His Gly Trp Leu Pro Thr Pro Ala Gly Arg Ala Ala Leu Ser Lys | | | |
| | 210 | 215 | 220 |
| Lys Met Glu Gly Pro Asn Leu Ser Asn Val Phe Lys Arg Thr Phe Tyr | | | |
| 225 | 230 | 235 | 240 |
| Gln Met Ala Tyr Lys Phe Ala Leu Ser Gly His Gln Arg Cys Ala Gly | | | |
| | 245 | 250 | 255 |
| Thr Gly Phe Ala Ile Pro Gln Ser Val Trp Lys Ser Trp Leu Arg His | | | |
| | 260 | 265 | 270 |
| Leu Ala Asn Pro Thr Leu Ile Asp Asn Gly Asp Gly Thr Phe Ser Leu | | | |
| | 275 | 280 | 285 |
| Gly Asp Thr Arg Asn Asp Ser Glu Asn Ala Trp Ile Phe Val Phe Glu | | | |
| | 290 | 295 | 300 |
| Leu Asp Pro Asp Thr Asp Ala Ser Pro Arg Pro Leu Ala Pro His Leu | | | |
| 305 | 310 | 315 | 320 |
| Glu Ile Arg Val Asn Val Asp Thr Leu Ile Asp Leu Ala Leu Arg Glu | | | |
| | 325 | 330 | 335 |
| Ser Pro Arg Ala Ala Leu Gly Pro Ser Gly Pro Val Ala Thr Phe Thr | | | |
| | 340 | 345 | 350 |
| Asp Lys Val Glu Ala Arg Met Leu Arg Phe Trp Pro Lys Thr Arg Arg | | | |
| | 355 | 360 | 365 |
| Arg Arg Ser Thr Thr Pro Gly Gly Gln Arg Gly Leu Phe Asp Ala | | | |
| | 370 | 375 | 380 |
| <210>91 | | | |
| <211>326 | | | |
| <212>PRT | | | |
| <213> 斯氏普罗威登斯菌 (<i>Providencia stuartii</i>) 164 | | | |
| <400>91 | | | |
| Met Lys Glu Leu Lys Leu Lys Glu Ala Lys Glu Ile Leu Lys Ala Leu | | | |
| 1 | 5 | 10 | 15 |
| Gly Leu Pro Pro Gln Gln Tyr Asn Asp Arg Ser Gly Trp Val Leu Leu | | | |
| | 20 | 25 | 30 |
| Ala Leu Ala Asn Ile Lys Pro Glu Asp Ser Trp Lys Glu Ala Lys Ala | | | |
| | 35 | 40 | 45 |
| Pro Leu Leu Pro Thr Val Ser Ile Met Glu Phe Ile Arg Thr Glu Tyr | | | |

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

<210>92

<211>157

<212>PRT

<213> 普通变形杆菌 (Proteus vulgaris)

<400>92

Met Ser His Pro Asp Leu Asn Lys Leu Leu Glu Leu Trp Pro His Ile
 1 5 10 15
 Gln Glu Tyr Gln Asp Leu Ala Leu Lys His Gly Ile Asn Asp Ile Phe
 20 25 30
 Gln Asp Asn Gly Gly Lys Leu Leu Gln Val Leu Leu Ile Thr Gly Leu
 35 40 45
 Thr Val Leu Pro Gly Arg Glu Gly Asn Asp Ala Val Asp Asn Ala Gly
 50 55 60
 Gln Glu Tyr Glu Leu Lys Ser Ile Asn Ile Asp Leu Thr Lys Gly Phe
 65 70 75 80
 Ser Thr His His His Met Asn Pro Val Ile Ile Ala Lys Tyr Arg Gln
 85 90 95
 Val Pro Trp Ile Phe Ala Ile Tyr Arg Gly Ile Ala Ile Glu Ala Ile
 100 105 110
 Tyr Arg Leu Glu Pro Lys Asp Leu Glu Phe Tyr Tyr Asp Lys Trp Glu
 115 120 125
 Arg Lys Trp Tyr Ser Asp Gly His Lys Asp Ile Asn Asn Pro Lys Ile
 130 135 140
 Pro Val Lys Tyr Val Met Glu His Gly Thr Lys Ile Tyr
 145 150 155

<210>93

<211>358

<212>PRT

<213> 不产色链霉菌 (*Streptomyces achromogenes*)

<400>93

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

100 105 110
 Phe Arg Tyr Asp Gln Tyr Ser Glu Ile Val Arg Val Gln Thr Lys Ala
 115 120 125
 Arg Pro Tyr Leu Asp Arg Val Ser Ser Ala Leu Ala Arg Val Asp Glu
 130 135 140
 Glu Asp Tyr Ser Thr Glu Glu Ser Phe Arg Ala Leu Val Ala Val Leu
 145 150 155 160
 Ala Val Cys Ile Ser Val Ala Asn Lys Lys Gln Arg Val Ala Val Gly
 165 170 175
 Ser Ala Ile Val Glu Ala Ser Leu Ile Ala Glu Thr Gln Ser Phe Val
 180 185 190
 Val Ser Gly His Asp Val Pro Arg Lys Leu Gln Ala Cys Val Ala Ala
 195 200 205
 Gly Leu Asp Met Val Tyr Ser Glu Val Val Ser Arg Arg Ile Asn Asp
 210 215 220
 Pro Ser Arg Asp Phe Pro Gly Asp Val Gln Val Ile Leu Asp Gly Asp
 225 230 235 240
 Pro Leu Leu Thr Val Glu Val Arg Gly Lys Ser Val Ser Trp Glu Gly
 245 250 255
 Leu Glu Gln Phe Val Ser Ser Ala Thr Tyr Ala Gly Phe Arg Arg Val
 260 265 270
 Ala Leu Met Val Asp Ala Ala Ser His Val Ser Leu Met Ser Ala Asp
 275 280 285
 Asp Leu Thr Ser Ala Leu Glu Arg Lys Tyr Glu Cys Ile Val Lys Val
 290 295 300
 Asn Glu Ser Val Ser Ser Phe Leu Arg Asp Val Phe Val Trp Ser Pro
 305 310 315 320
 Arg Asp Val His Ser Ile Leu Ser Ala Phe Pro Glu Ala Met Tyr Arg
 325 330 335
 Arg Met Ile Glu Ile Glu Val Arg Glu Pro Gln Leu Asp Arg Trp Ala
 340 345 350
 Glu Ile Phe Pro Glu Thr
 355
 <210>94
 <211>315
 <212>PRT
 <213> 白色链霉菌 (Streptomyces albus)G
 <400>94
 Met Ile Asn Ala Asp Lys Pro His Arg Trp Asn Asp Asp Val Gln Ala

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

<210>95

<211>432

<212>PRT

<213> 糖多孢菌某种 (Saccharopolyspora sp.)

<400>95

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

Lys Ser Gln Val Trp Glu Arg Phe His Arg Lys Cys Phe Asn Cys Arg
 275 280 285
 Lys Asp Leu Lys Leu Ser Glu Val Gln Leu Asp His Thr Arg Pro Leu
 290 295 300
 Ala Tyr Leu Trp Pro Ile Asp Glu His Ala Thr Cys Leu Cys Ala Gln
 305 310 315 320
 Cys Asn Asn Thr Lys Lys Asp Arg Phe Pro Val Asp Phe Tyr Ser Glu
 325 330 335
 Gln Gln Ile Arg Glu Leu Ser Asp Ile Cys Gly Leu Pro Tyr Gln Asp
 340 345 350
 Leu Cys Ala Arg Ser Leu Asn Leu Asp Gln Leu Asp Arg Ile Glu Arg
 355 360 365
 Asn Ile Ala Glu Phe Ser Lys Glu Trp Asp Val Arg Thr Phe Ala Ser
 370 375 380
 Thr Ala Arg Arg Ile Ser Glu Val Tyr Pro Ala Arg Asp Leu Phe Glu
 385 390 395 400
 Thr Leu Lys Lys Glu Ser Glu Ser Ala Tyr Asn Lys Ile Ile Glu Lys
 405 410 415
 Leu Lys Glu Arg Pro Asp Ala Leu Leu Asp Glu Ala Leu Pro Leu Asp
 420 425 430
 <210>96
 <211>323
 <212>PRT
 <213> 链霉菌属菌种 (Streptomyces species)Bf-61
 <400>96
 Met Asn Ser Ser Asp Gly Ile Asp Gly Thr Val Ala Ser Ile Asp Thr
 1 5 10 15
 Ala Arg Ala Leu Leu Lys Arg Phe Gly Phe Asp Ala Gln Arg Tyr Asn
 20 25 30
 Val Arg Ser Ala Val Thr Leu Leu Ala Leu Ala Gly Leu Lys Pro Gly
 35 40 45
 Asp Arg Trp Val Asp Ser Thr Thr Pro Arg Leu Gly Val Gln Lys Ile
 50 55 60
 Met Asp Trp Ser Gly Glu His Trp Ala Lys Pro Tyr Ala Thr Gly Ser
 65 70 75 80
 Arg Glu Asp Phe Arg Lys Lys Thr Leu Arg Gln Trp Val Asp Asn Gly
 85 90 95
 Phe Ala Val Leu Asn Ala Asp Asn Leu Asn Ile Ala Thr Asn Ser Gln
 100 105 110

Leu Asn Glu Tyr Cys Leu Ser Asp Glu Ala Leu Gln Ala Leu Arg Ala
 115 120 125
 Tyr Gly Thr Glu Gly Phe Glu Glu Ser Leu Val Val Phe Leu Asp Glu
 130 135 140
 Ala Ser Lys Ala Val Lys Ala Arg Ala Glu Ala Leu Gln Ala Ala Met
 145 150 155 160
 Ile Ser Val Asp Leu Pro Gly Gly Glu Glu Phe Leu Leu Ser Pro Ala
 165 170 175
 Gly Gln Asn Pro Leu Leu Lys Lys Met Val Glu Glu Phe Val Pro Arg
 180 185 190
 Phe Ala Pro Arg Ser Thr Val Leu Tyr Leu Gly Asp Thr Arg Gly Lys
 195 200 205
 His Ser Leu Phe Glu Arg Glu Ile Phe Glu Glu Val Leu Gly Leu Thr
 210 215 220
 Phe Asp Pro His Gly Arg Met Pro Asp Leu Ile Leu His Asp Glu Val
 225 230 235 240
 Arg Gly Trp Leu Phe Leu Met Glu Ala Val Lys Ser Lys Gly Pro Phe
 245 250 255
 Asp Glu Glu Arg His Arg Ser Leu Gln Glu Leu Phe Val Thr Pro Ser
 260 265 270
 Ala Gly Leu Ile Phe Val Asn Cys Phe Glu Asn Arg Glu Ser Met Arg
 275 280 285
 Gln Trp Leu Pro Glu Leu Ala Trp Glu Thr Glu Ala Trp Val Ala Glu
 290 295 300
 Asp Pro Asp His Leu Ile His Leu Asn Gly Ser Arg Phe Leu Gly Pro
 305 310 315 320
 Tyr Glu Arg
 <210>97
 <211>227
 <212>PRT
 <213> 头状链霉菌 (*Streptomyces caespitosus*)
 <400>97
 Met Ile Asn Asp Gln Leu Pro Arg Trp Val Arg Glu Ala Arg Val Gly
 1 5 10 15
 Thr Arg Thr Gly Gly Pro Ala Met Arg Pro Lys Thr Ser Asp Ser Pro
 20 25 30
 Tyr Phe Gly Trp Asp Ser Glu Asp Trp Pro Glu Val Thr Arg Gln Leu
 35 40 45
 Leu Ser Glu Gln Pro Leu Ser Gly Asp Thr Leu Val Asp Ala Val Leu

| | | | |
|---|-----|-----|-----|
| | 85 | 90 | 95 |
| Val Ile Pro Lys Ala Phe Gly Glu Ala Leu Lys Arg Ala His Gly Gly | | | |
| | 100 | 105 | 110 |
| Lys Cys Ala Val Cys Tyr Gly Asp Phe Ser Glu Arg Glu Leu Gln Cys | | | |
| | 115 | 120 | 125 |
| Asp His Arg Val Pro Phe Ala Ile Ala Gly Asp Lys Pro Lys Leu Val | | | |
| | 130 | 135 | 140 |
| Gln Glu Asp Phe Met Pro Leu Cys Ala Ser Asp Asn Arg Ala Lys Ser | | | |
| 145 | 150 | 155 | 160 |
| Trp Ser Cys Glu Asn Cys Pro Asn Trp Glu Leu Lys Asp Glu Asp Thr | | | |
| | 165 | 170 | 175 |
| Cys Arg Ser Cys Phe Trp Ala Ser Pro Glu Asn Tyr Thr His Val Ser | | | |
| | 180 | 185 | 190 |
| Thr Arg Pro Glu Arg Arg Ile Asn Leu Leu Phe Gln Gly Asp Glu Val | | | |
| | 195 | 200 | 205 |
| Glu Ile Phe Asp Ala Leu Lys Asn Ala Ala Ala Asn Glu Gly Val Ser | | | |
| 210 | 215 | 220 | |
| Leu Thr Glu Ala Thr Lys Arg Lys Leu Ala Asp | | | |
| 225 | 230 | 235 | |
| <210>99 | | | |
| <211>281 | | | |
| <212>PRT | | | |
| <213> 球衣细菌属某种 (<i>Sphaerotilus species</i>) | | | |
| <400>99 | | | |
| Met Ser Lys Ala Ala Tyr Gln Asp Phe Thr Lys Arg Phe Ser Leu Leu | | | |
| 1 | 5 | 10 | 15 |
| Ile Lys Lys His Pro Asn Leu Ile Thr Met Thr Leu Ser Asn Ile Phe | | | |
| | 20 | 25 | 30 |
| Thr Met Arg Leu Ile Gly Asn Lys Thr His Gly Asp Leu Ala Glu Ile | | | |
| | 35 | 40 | 45 |
| Ala Ile Ser Glu Phe Ile Asn Gln Tyr Met Tyr Asp Phe Lys Ser Ile | | | |
| 50 | 55 | 60 | |
| His Val Gly Lys Asp Leu Tyr Arg Ala Lys Ser Lys Glu Glu Asp Ile | | | |
| 65 | 70 | 75 | 80 |
| Thr Val Glu Asn Glu Ile Thr Lys Glu Lys Phe Pro Ile Ser Leu Lys | | | |
| | 85 | 90 | 95 |
| Ala Tyr Gly Asp Gly Pro Leu Gln Leu Ser Thr Asp Lys Asn Phe Leu | | | |
| | 100 | 105 | 110 |
| Met Tyr Pro Leu Leu Glu Glu Ile Gly Ala Phe Ile Asn Ala Lys Glu | | | |

| | | |
|---|-----|-----|
| 115 | 120 | 125 |
| Lys Ile Glu Glu Ile Phe Ala Asn Glu Ala Phe Ser Cys Phe Ser Glu | | |
| 130 | 135 | 140 |
| Ile Asn Val Leu Pro Leu Ile Tyr Asp Glu Lys Arg Gln Arg Cys Asn | | |
| 145 | 150 | 155 |
| Ile Leu Val Phe Asp Ala Ala Arg Ala Arg Ala Glu Thr Ala Tyr Ile | | |
| 165 | 170 | 175 |
| Arg Lys Glu Thr Glu Gly Ser Gly Arg Lys His Pro Ala Tyr Arg Phe | | |
| 180 | 185 | 190 |
| Phe Asp Lys Asn Lys Asn Tyr Ile Cys Glu Val Arg Tyr Gly Asn Ala | | |
| 195 | 200 | 205 |
| Ala Ala Asn Ala Leu Gln Arg Gly Leu Trp Thr Asn Thr Lys Asn Ala | | |
| 210 | 215 | 220 |
| Thr Ser Phe Phe Asp Ser Val Thr Asn Gly Trp Val Asp Tyr Ser His | | |
| 225 | 230 | 235 |
| Asn Leu Val Leu Val Lys Leu Leu Ser His Ala Leu Val Ser Ser Arg | | |
| 245 | 250 | 255 |
| Lys Gly His Glu Ala Ala Leu Glu Glu Ile Lys Lys Asp Ile Leu Gln | | |
| 260 | 265 | 270 |
| Leu Lys Gln Thr Asn Gly Ile Asn Val | | |
| 275 | 280 | |

<210>100

<211>1077

<212>DNA

<213> 鱼腥藻

<400>100

```

atggaagaag accttgattt atctgaaaat atcgaagctg catctgcgga gcttacgact 60
ctttatcagg tagctgctga tgctatgaaa gattatattg aaatctatct tgcgctgagt 120
aaacagtctg atgggttttc aaatattaac aatcttgact taacttctcg taacaggcgt 180
ttggtagtta tacatggact ttcgtagag ttagatccag atacttcgac tccagaggaa 240
attaaacgtg aagctgaacg aatgctagcg atagctcttg atacagagtc agcaattacg 300
gcaggagtat atgaaaaaat gcgtctcttc gcaagctctt tagtagatca gctatttgaa 360
caaacgatg aacttaattc attatcatcg gaatatttgt cagcaaatcc aggatTTTTG 420
ccgtTTTTCC agcagttggc ggggcttaga agtaaatacag agttaaagag agaagtagga 480
aatgcctctg acaatagtat ttctaaagcg gttgcagaga gaatattaga ggcgattata 540
cgtaacttga gaattcgcac tttttccaaa gagaaactat tacaagctgt tgagcctact 600
ttagaaggaa tagtcaggga tctcgttaga aaagtgttat tggaaaatat agttgctgat 660
gctttatctg atttacaagt tcctttcatg cgtgaatcag agtatcaaag ccttaaagga 720
gtgatttatg atttccgcgc tgattttgtg ataccagacg cacaaaatcc aattgctttt 780

```

atcgaggtgc gaaaaagctc tacacgacat gcgtcactct atgccaagga taagatgttt 840
 tcagcgatta attggaaagg aaaaaataaa aggcttttgg gtattttggg tgtggaagga 900
 ccttgacaa gagaaactct tcgcgtcatg gcaaatgtgt ttgattacgt tacacctta 960
 actcgtgttt cccaagttgc agaagctatc agagcatatc tagatgggga taaaacgaga 1020
 ctgaagtggg tagttaattt cagtattgaa gaagcagacc acgacaacat aacctaa 1077

<210>101

<211>1293

<212>DNA

<213> 球形芽孢杆菌

<400>101

atgagacgat tagcaaaaaa ttcacggaac gacagttatt taagtaatag ggattaccag 60
 gaaatcgtga gggaaaatac cactacaata tcgtttccct taaaagaaaa acatactctg 120
 actttaacga aaaaaatagg gctaaatcag actgctggat tcggaggatg gtttttcct 180
 gattcaccat gtttattaac agtaactgta ctatcctctt tcggtacaaa ggtaacttct 240
 aaaaccttta gcctttctaa agattggaat cgtgttgggc ttgcttggat taacgagcat 300
 tcgagtgaca ccataagcat tgtcctagag tttagtgatg tggaaatagt tcatacatgg 360
 ggacttacat gtgatgtttt taatgtccat gaattaatta ttgatgctat agaagatcaa 420
 aataaactaa tagacgtgct aatcaagaa ctttatctc ctgaaacata ttatttaaac 480
 catgactctg atactgattt aattgagaat ttggaatcta cagaagagat aaagatagtt 540
 aaccaaagcc aaaagcaaat ctctttaaaa aatgctgtt attgtcaacg ttatatgcct 600
 gtgaacatat tagttcgttc aaattcatca tttcataaac acaagagtaa gaaaactggg 660
 tttcaaaatg aatgtcgggc ttgtaagaag tggagaataa ataattcatt caatccagtc 720
 agaacaaaag accaactaca tgaatcagca gttattacac gtgaaaaaaaa aatattactt 780
 aaagaacctg aatattaca gaaaatcaaa aatagaaata acggtgaggg cttaaaaagt 840
 attatatgga aaaaatttga taaaaaatgc ttttaattgtg aaaaagaatt aaccattgaa 900
 gaggtacgcc tagaccatac aagaccactt gcttatctgt ggcctatcga tgaacacgca 960
 acttgtttat gtgaaaaatg caacaataca aaacatgata tgtttcctat cgatttttat 1020
 caaggggacg aagacaaatt aagacgttta gctagaatta cggggtaga ttatgaatct 1080
 ctagttaaga gggacgtaaa tgaagttgaa cttgcaagaa taatcaataa cattgaagac 1140
 tttgcaacta atgtagaggc acgtactttt cgctcaataa gaaataaagt aaaagaagta 1200
 cgtcccgata ctgacctatt tgaaattctt aatctaaaa atattaattt atataatgaa 1260
 cttcaatatg aacttcttac ccgtaaggat taa 1293

<210>102

<211>2189

<212>DNA

<213> 人工

<220>

<223> 质粒 placzz2

<400>102

| | | | | | | |
|-------------|-------------|-------------|------------|-------------|-------------|------|
| ctggcgaaag | ggggatgtgc | tgcaaggcga | ttaagttggg | taacgccagg | gttttcccag | 60 |
| tcacgacgtt | gtaaacgac | ggccagtga | ttcgagctcg | gtacccgggg | gcgcgccgga | 120 |
| tccttaatta | agtctagagt | cgactgttta | aacctgcagg | catgcaagct | tggcgtaatc | 180 |
| atggtcatat | gttaacctcc | ggcgtaatca | tggtcatagc | tgtttcctgt | gtgaaattgt | 240 |
| tatccgctca | caattccaca | caacatacga | gccggaagca | taaagtgtaa | agcctgggggt | 300 |
| gcctaataag | tgagctaact | cacattaatt | gcgttgcgct | cactgcccgc | tttccagtcg | 360 |
| ggaaacctgt | cgtgccagca | tgtgagcaaa | aggccagcaa | aaggccagga | accgtaaaaa | 420 |
| ggccgcgttg | ctggcgtttt | tccatagget | ccgccccct | gacgagcate | acaaaaatcg | 480 |
| acgctcaagt | cagaggtggc | gaaacctgac | aggactataa | agataccagg | cgtttcccc | 540 |
| tggaagctcc | ctcgtgcgct | ctcctgttcc | gaccctgccg | cttaccggat | acctgtccgc | 600 |
| ctttctccct | tcgggaagcg | tgccgctttc | tcatagctca | cgctgtaggt | atctcagttc | 660 |
| ggtgtaggtc | gttcgctcca | agctgggctg | tgtgcacgaa | cccccgttc | agcccgaccg | 720 |
| ctgcgcctta | tccggttaact | atcgtcttga | gtccaacctg | gtaagacacg | acttatcgcc | 780 |
| actggcagca | gccactggta | acaggattag | cagagcgagg | tatgtaggcg | gtgctacaga | 840 |
| gttcttgaag | tgggtggccta | actacggcta | cactagaagg | acagtatttg | gtatctgcgc | 900 |
| tctgctgaag | ccagttacct | tcggaaaaag | agttggtagc | tcttgatccg | gcaaacaaac | 960 |
| caccgctggt | agcgggtggt | ttttgtttg | caagcagcag | attacgcgca | gaaaaaaagg | 1020 |
| atctcaagaa | gaccccttga | tctttctac | ggggtctgac | gctcagtgga | acgaaaactc | 1080 |
| acgttaaggg | atcttggtea | tgagattatc | aaaaaggatc | ttcacctaga | tccttttaaa | 1140 |
| ttaaaaaatga | agtttttaaat | caatctaaag | tatatatgag | taaacttggg | ctgacagtta | 1200 |
| ccaatgctta | atcagtgagg | cacctatctc | agcgatctgt | ctatttcggt | catccatagt | 1260 |
| tgccctgactc | cccgtcgtgt | agataactac | gatacgggag | ggcttaccat | ctggccccag | 1320 |
| tgctgcaatg | ataccgcgag | accacgctc | accggctcca | gatttatcag | caataaacca | 1380 |
| gccagccgga | agggccgagc | gcagaagtgg | tccgcaact | ttatccgcct | ccatccagtc | 1440 |
| tattaattgt | tgccgggaag | ctagagtaag | tagttcgcca | gttaatagtt | tgcgcaacgt | 1500 |
| tgttgccatt | gctacaggca | tcgtgggtgc | acgctcgtcg | tttggtaggt | cttcattcag | 1560 |
| ctccggttcc | caacgatcaa | ggcgagttac | atgatcccc | atgttggtgca | aaaaagcggg | 1620 |
| tagctccttc | ggtcctccga | tcgttgctcag | aagtaagttg | gccgcagtg | tatcactcat | 1680 |
| ggttatggca | gcactgcata | attctcttac | tgtcatgcca | tccgtaagat | gcttttctgt | 1740 |
| gactgggtgag | tactcaacca | agtcattctg | agaatagtg | atgcggcgac | cgagttgctc | 1800 |
| ttgcccggcg | tcaatacggg | ataataccgc | gccacatagc | agaactttaa | aagtgtcat | 1860 |
| catttgaaaa | cgttcttcgg | ggcgaaaact | ctcaaggatc | ttaccgctgt | tgagatccag | 1920 |
| ttcgatgtaa | cccactcgtg | cacccaactg | atcttcagca | tcttttactt | tcaccagcgt | 1980 |
| ttctgggtga | gcaaaaacag | gaaggcaaaa | tgccgcaaaa | aagggaataa | gggcgacacg | 2040 |
| gaaatgttga | atactcatac | tcttcctttt | tcaatattat | tgaagcattt | atcagggtta | 2100 |
| ttgtctcatg | agcggataca | tatttgaatg | tatttagaaa | aataaaciaa | taggggttcc | 2160 |
| gcgcacattt | ccccgaaaag | tgccacctg | | | | 2189 |

<210>103

<211>10673

<212>DNA

<213> 人工

<220>

<223> 质粒 pBC4

<400>103

| | | | | | | |
|------------|-------------|-------------|-------------|------------|-------------|------|
| tcgcgcgttt | cggtgatgac | ggtgaaaace | tetgacacat | gcagctcccc | gagacgggtca | 60 |
| cagcttgtct | gtaagecggat | gccgggagca | gacaagcccc | tcagggcgcg | tcagcgggtg | 120 |
| ttggcgggtg | tcggggctgg | cttaactatg | cggcatcaga | gcagattgta | ctgagagtgc | 180 |
| accatatgcg | gtgtgaaata | ccgcacagat | gcgtaaggag | aaaataccgc | atcaggcgcc | 240 |
| attcgccatt | caggctgcgc | aactgttggg | aagggcgatc | ggtgcgggcc | tcttcgctat | 300 |
| tacgccagct | ggcgaaaggg | ggatgtgctg | caaggcgatt | aagttgggta | acgccagggt | 360 |
| tttcccagtc | acgacgttgt | aaaacgacgg | ccagtgaatt | cgagctcggg | acccggggat | 420 |
| cctctagagt | cgaaccccgg | atccggccgt | ccgccgtgat | ccatgcgggt | accgcccgcg | 480 |
| tgtcgaacce | aggtgtgcga | cgtcagacaa | cgggggagcg | ctccttttgg | cttccttcca | 540 |
| ggcgcggcgg | ctgctgcgct | agcttttttg | gccactggcc | gcgcgcggcg | taagcggtta | 600 |
| ggctggaag | cgaagcatt | aagtggctcg | ctccctgtag | ccggagggtt | atthtccaag | 660 |
| ggttgagtcg | caggaccccc | ggttcgagtc | tcgggccggc | cggactgcgg | cgaacggggg | 720 |
| tttgcctccc | cgatcatgaa | gaccccgtt | gcaaattcct | ccgaaacag | ggacgagccc | 780 |
| cttttttgct | tttcccagat | gcatccgggtg | ctgcggcaga | tgcgcccccc | tcctcagcag | 840 |
| cggcaagagc | aagagcagcg | gcagacatgc | agggcacccct | ccccttctcc | taccgcgtca | 900 |
| ggaggggcaa | catccgcggc | tgacgcggcg | gcagatgggtg | attacgaacc | cccgcggcgc | 960 |
| cgggcccggc | actacctgga | cttggaggag | ggcgagggcc | tggcgcggct | aggagcgcgc | 1020 |
| tctcctgagc | gacaccaag | ggtgcagctg | aagcgtgaca | cgcgcgaggc | gtacgtgccg | 1080 |
| cggcagaacc | tgthtgcga | ccgcgagggga | gaggagcccc | aggagatgcg | ggatcgaag | 1140 |
| ttccacgcag | ggcgcgagtt | gcggcatggc | ctgaaccgcg | agcggttgct | gcgcgaggag | 1200 |
| gactttgagc | ccgacgcgcg | gaccgggatt | agtcccgcgc | gcgcacacgt | ggcggccgcc | 1260 |
| gacctggtaa | ccgcgtacga | gcagacgggtg | aaccaggaga | ttaactttca | aaaaagcttt | 1320 |
| aacaaccacg | tgcgcacgct | tgtggcgcgc | gaggaggtgg | ctataggact | gatgcatctg | 1380 |
| tgggactttg | taagcgcgct | ggagcaaaac | ccaaatagca | agccgctcat | ggcgcagctg | 1440 |
| ttccttatag | tgcagcacag | cagggacaac | gaggcattca | gggatgcgct | gctaaacata | 1500 |
| gtagagcccc | agggccgctg | gctgctcgat | ttgataaaca | ttctgcagag | catagtgggtg | 1560 |
| caggagcgcg | gcttgagcct | ggctgacaag | gtggccgcca | ttaactattc | catgctcagt | 1620 |
| ctgggcaagt | tttacgcccg | caagatatac | catacccctt | acgttcccat | agacaaggag | 1680 |
| gtaaagatcg | aggggttcta | catgcgcatg | gcgttgaagg | tgcttacctt | gagcgcagac | 1740 |
| ctgggcgttt | atcgcaacga | gcgcatccac | aaggccgtga | gcgtgagccg | gcggcgcgag | 1800 |
| ctcagcagcc | gcgagctgat | gcacagcctg | caaagggccc | tggctggcac | gggcagcggc | 1860 |
| gatagagagg | ccgagtccta | ctttgacgcg | ggcgctgacc | tgcgctgggc | cccaagccga | 1920 |
| cgcgccctgg | aggcagctgg | ggccggacct | gggctggcgg | tggcacccgc | gcgcgctggc | 1980 |
| aacgtcggcg | gcgtggagga | atatgacgag | gacgatgagt | acgagccaga | ggacggcgag | 2040 |

| | | | | | | |
|-------------|------------|-------------|------------|-------------|-------------|------|
| tactaagcgg | tgatgtttct | gatecagatga | tgcaagacgc | aacggacccg | gcggtgcggg | 2100 |
| cggcgctgca | gagccagccg | tccggcetta | actccacgga | cgactggcgc | caggtcatgg | 2160 |
| accgcatcat | gtegctgact | gcgcgtaace | ctgacgcggt | ccggcagcag | ccgcaggcca | 2220 |
| accggtctctc | cgcaattctg | gaagcgggtg | tcccggcgcg | cgcaaaccce | acgcacgaga | 2280 |
| aggtgctggc | gatecgtaac | gcgctggccg | aaaacagggc | catccggccc | gatgaggccg | 2340 |
| gcctggtcta | cgacgcgctg | cttcagcgcg | tggtctgtta | caacagcggc | aacgtgcaga | 2400 |
| ccaacctgga | ccggctggtg | ggggatgtgc | gcgaggccgt | ggcgcagcgt | gagcgcgcgc | 2460 |
| agcagcaggg | caacctgggc | tccatggttg | cactaaacgc | cttcctgagt | acacagcccc | 2520 |
| ccaacgtgcc | gcggggacag | gaggactaca | ccaactttgt | gagcgcactg | cggtaatgg | 2580 |
| tgactgagac | accgcaaagt | gaggtgtacc | agtccgggce | agactatfff | ttccagacca | 2640 |
| gtagacaagg | cctgcagacc | gtaaacctga | gccaggcttt | caagaacttg | caggggctgt | 2700 |
| ggggggtgcg | ggctcccaca | ggcgaccgcg | cgaccgtgtc | tagcttgctg | acgcccact | 2760 |
| cgcgctgtt | gctgctgcta | atagcgcctt | tcacggacag | tggcagcgtg | tcccgggaca | 2820 |
| catacctagg | tcacttgctg | acactgtacc | gcgaggccat | aggtcaggcg | catgtggacg | 2880 |
| agcatacttt | ccaggagatt | acaagtgtca | gccgcgcgct | ggggcaggag | gacacgggca | 2940 |
| gcctggaggc | aacctgaac | tacctgctga | ccaaccggcg | gcagaagatc | ccctcgttgc | 3000 |
| acagtttaaa | cagcgaggag | gagcgcactt | tgcgctatgt | gcagcagagc | gtgagcctta | 3060 |
| acctgatgcg | cgacggggta | acgcccagcg | tggcgctgga | catgaccgcg | cgcaacatgg | 3120 |
| aaccgggcat | gtatgcctca | aaccggccgt | ttatcaatcg | cctaattggac | tacttgcatc | 3180 |
| gcgcggccgc | cgtgaacccc | gagtatttca | ccaatgccat | cttgaacccg | cactggctac | 3240 |
| cgccccctgg | tttctacacc | gggggatttg | aggtgcccga | gggtaacgat | ggattcctct | 3300 |
| gggacgacat | agacgacagc | gtgttttccc | cgcaaccgca | gaccctgcta | gagttgcaac | 3360 |
| agcgcgagca | ggcagaggcg | gcgctgcgaa | aggaaagctt | ccgcaggcca | agcagcttgt | 3420 |
| ccgatctagg | cgctgcggcc | ccgcggctag | atgcgagtag | cccatttcca | agcttgatag | 3480 |
| ggtcttttac | cagcactcgc | accaccgcgc | cgcgctgct | ggcgaggag | gagtacctaa | 3540 |
| acaactcgct | gctgcagccg | cagcgcgaaa | agaacctgcc | tccggcattt | cccaacaacg | 3600 |
| ggatagagag | cctagtggac | aagatgagta | gatggaagac | gtatgcgcag | gagcacaggg | 3660 |
| atgtgcccgg | cccgcgcccg | cccaccgcgc | gtcaaaggca | cgaccgtcag | cggggctctgg | 3720 |
| tgtgggagga | cgatgactcg | gcagacgaca | gcagcgtcct | ggatttggga | gggagtggca | 3780 |
| accgttttgc | gcaccttcgc | cccaggctgg | ggagaatggt | ttaaaaaaaa | aaaaaaaaaag | 3840 |
| catgatgcaa | aataaaaaac | tcaccaaggc | catggcaccg | agcgttgggt | ttcttgtatt | 3900 |
| ccccttagta | tgcagcgcgc | ggcgatgtat | gaggaaggtc | ctcctccctc | ctacgagagc | 3960 |
| gtggtgagcg | cggcgccagt | ggcggcggcg | ctgggttccc | ccttcgatgc | tcccctggac | 4020 |
| ccgccgtttg | tgcctccgcg | gtacctgcgg | cctaccgggg | ggagaaacag | catccgttac | 4080 |
| tctgagttgg | caccctatt | cgacaccacc | cgtgtgtacc | ttgtggacaa | caagtcaacg | 4140 |
| gatgtggcat | ccctgaacta | ccagaacgac | cacagcaact | ttctaaccac | ggtcattcaa | 4200 |
| aacaatgact | acagcccggg | ggaggcaagc | acacagacca | tcaatcttga | cgaccgttcg | 4260 |
| cactggggcg | gcgacctgaa | aaccatcctg | cataccaaca | tgccaaatgt | gaacgagttc | 4320 |
| atgtttacca | ataagtttaa | ggcgcgggtg | atggtgtcgc | gctcgttac | taaggacaaa | 4380 |

| | | | | | | |
|------------|-------------|-------------|------------|-------------|------------|------|
| caggtggagc | tgaaatatga | gtgggtggag | ttcacgctgc | ccgagggcaa | ctactccgag | 4440 |
| accatgacca | tagaccttat | gaacaacgcg | atcgtggagc | actacttgaa | agtgggcagg | 4500 |
| cagaacgggg | ttctggaaag | cgacateggg | gtaaagtttg | acacccgcaa | cttcagactg | 4560 |
| gggtttgacc | cagtcactgg | tettgtcatg | cctggggtat | atacaaacga | agccttccat | 4620 |
| ccagacatca | ttttgctgcc | aggatgcggg | gtggacttca | cccacagccg | cctgagcaac | 4680 |
| ttgttgggca | tccgcaageg | gcaaccette | caggagggct | ttaggatcac | ctacgatgac | 4740 |
| ctggagggtg | gtaacattcc | cgcactgttg | gatgtggacg | cctaccagge | aagcttaaaa | 4800 |
| gatgacaccg | aacagggcgg | ggatggcgca | ggcggcggca | acaacagtgg | cagcggcgcg | 4860 |
| gaagagaact | ccaacgcggc | agccgcggca | atgcagccgg | tggaggacat | gaacgatcat | 4920 |
| gccattcgcg | gcgacacctt | tgccacacgg | gcgaggaga | agcgcgctga | ggccgaggca | 4980 |
| gcggcagaag | ctgccgcccc | cgctgcgcaa | cccgaggtcg | agaagcctca | gaagaaaccg | 5040 |
| gtgatcaaac | ccctgacaga | ggacagcaag | aaacgcagtt | acaaccta | aagcaatgac | 5100 |
| agcaccttca | cccagtaccg | cagctggtag | cttgcataca | actacggcga | ccctcagacc | 5160 |
| gggatccgct | catggaccct | cctttgcact | cctgacgtaa | cctgcggctc | ggagcaggtc | 5220 |
| tactggtcgt | tgccagacat | gatgcaagac | cccgtgacct | tccgctccac | gagccagatc | 5280 |
| agcaactttc | cgggtggtggg | cgccgagctg | ttgcccgctg | actccaagag | cttctacaac | 5340 |
| gaccaggccg | tctactecca | gctcatccgc | cagtttacct | ctctgaccca | cgtgttcaat | 5400 |
| cgctttcccc | agaaccagat | tttggcgcg | ccgccagccc | ccaccatcac | caccgtcagt | 5460 |
| gaaaacgttc | ctgctctcac | agatcacggg | acgctaccgc | tgcgcaacag | catcggagga | 5520 |
| gtccagcgag | tgaccattac | tgacgccaga | cgccgcacct | gcccctacgt | ttacaaggcc | 5580 |
| ctgggcatag | tctcgccgcg | cgctctatcg | agccgcactt | tttgagcaaa | catgtccatc | 5640 |
| cttatatcgc | ccagcaataa | cacaggctgg | ggcctgcgct | tccaagcaa | gatgtttggc | 5700 |
| ggggcaaaga | agcgctccga | ccaacaccca | gtgcgcgtgc | gcgggcaacta | ccgcgcgccc | 5760 |
| tggggcgcg | acaaacgcgg | ccgcaactggg | cgcaccaccg | tcgatgacgc | cattgacgcg | 5820 |
| gtggtggagg | aggcgcgcaa | ctacacgccc | acgccgccac | cagtgtccac | agtggacgcg | 5880 |
| gccattcaga | ccgtggtgcg | cggagcccgg | cgttatgcta | aaatgaagag | acggcggagg | 5940 |
| cgcgtagcac | gtcgccaccg | ccgcccagccc | ggcactgccc | cccaacgcgc | ggcggcggcc | 6000 |
| ctgcttaacc | gcgcacgtcg | caccggccga | cgggcggcca | tgcgggcccgc | tcgaaggctg | 6060 |
| gccgcgggta | ttgtcactgt | gccccccagg | tccaggcgac | gagcggcccgc | cgcagcagcc | 6120 |
| gcggccatta | gtgctatgac | tcagggtcgc | aggggcaacg | tgtactgggt | gcgcgactcg | 6180 |
| gttagcggcc | tgcgcgtgcc | cgtgcgcacc | cgccccccgc | gcaactagat | tgcaagaaaa | 6240 |
| aactacttag | actcgtactg | ttgtatgtat | ccagcggcgg | cggcgcgcaa | cgaagctatg | 6300 |
| tccaagcgca | aaatcaaaga | agagatgctc | caggtcatcg | cgccggagat | ctatggcccc | 6360 |
| ccgaagaagg | aagagcagga | ttacaagccc | cgaaagctaa | agcgggtcaa | aaagaaaaag | 6420 |
| aaagatgatg | atgatgatga | acttgacgac | gaggtggaac | tgctgcacgc | aaccgcgccc | 6480 |
| aggcggcggg | tacagtggaa | aggctcagcg | gtaagacgtg | ttttgcgacc | cggcaccacc | 6540 |
| gtagttttta | cgcccgggtga | gcgctccacc | cgcacctaca | agcgcgtgta | tgatgaggtg | 6600 |
| tacggcgacg | aggacctgct | tgagcaggcc | aacgagcgcc | tcggggagtt | tgcctacgga | 6660 |
| aagcggcata | aggacatggt | ggcgttgccg | ctggacgagg | gcaaccacaac | acctagccta | 6720 |

| | | | | | | |
|------------|-------------|------------|------------|-------------|-------------|------|
| aagcccgtga | cactgcagca | ggtgctgccc | acgcttgca | cgtccgaaga | aaagcgcggc | 6780 |
| ctaaagecgc | agtctggtga | cttggcacc | accgtgcagc | tgatggtacc | caagcgcag | 6840 |
| cgactggaag | atgtcttgg | aaaaatgacc | gtggagcctg | ggctggagcc | cgaggtccgc | 6900 |
| gtgcgccaa | tcaagcaggt | ggcaccggga | ctgggcgtgc | agaccgtgga | cgttcagata | 6960 |
| cccaccacca | gtagcactag | tattgceact | gccacagagg | gcatggagac | acaaacgtcc | 7020 |
| ccggttgct | cggcggtg | agatgcccgc | gtgcaggcgg | ccgctgcggc | cgctccaaa | 7080 |
| acctctacgg | aggtgcaaac | ggaccegtgg | atgtttcgcg | tttcagcccc | ccggcgcggc | 7140 |
| cgccgttcca | ggaagtacgg | caccgccagc | gcactactgc | ccgaatatgc | cctacatcct | 7200 |
| tccatcgcgc | ctacccccgg | ctatcgtg | tacacctacc | gccccagaag | acgagcgact | 7260 |
| acccgacgcc | gaaccaccac | tggaaccgc | cgccgcgctc | gccgtcgcca | gcccgtgctg | 7320 |
| gccccgattt | ccgtgcgcag | ggtggctcgc | gaaggaggca | ggaccctggt | gctgccaaca | 7380 |
| gcgcgctacc | acccagcat | cgtttaaaag | ccggtctttg | tggttcttgc | agatatggcc | 7440 |
| ctcacctgcc | gcctccgttt | cccggtgccg | ggattccgag | gaagaatgca | ccgtaggagg | 7500 |
| ggcatggccg | gccacggcct | gacgggcggc | atgcgtcgtg | cgcaccaccg | gcggcggcgc | 7560 |
| gcgtgcacc | gtcgcattgc | cgccggtatc | ctgcccctcc | ttattccact | gatcgcggc | 7620 |
| gcgattggcg | ccgtgcccgg | aattgcatcc | gtggccttgc | aggcgcagag | acactgatta | 7680 |
| aaaacaagtt | gcatgtggaa | aaatcaaaat | aaaaagtctg | gagtctcacg | ctcgtttggt | 7740 |
| cctgtaacta | ttttgtagaa | tggaagacat | caactttgcg | tctctggccc | cgcgacacgg | 7800 |
| ctcgcgccc | ttcatgggaa | actggcaaga | tatcggcacc | agcaatatga | gcggtggcgc | 7860 |
| cttcagctgg | ggctcgtgt | ggagcggcat | taaaaatttc | ggttccacca | ttaagaacta | 7920 |
| tggcagcaag | gcctggaaca | gcagcacagg | ccagatgctg | agggacaagt | tgaagagca | 7980 |
| aaatttccaa | caaaaggtgg | tagatggcct | ggcctctggc | attagcgggg | tgggtggacct | 8040 |
| ggccaaccag | gcagtgcaaa | ataagattaa | cagtaagctt | gatccccgcc | ctcccgtaga | 8100 |
| ggagcctcca | ccggcctg | agacagtgtc | tccagagggg | cgtggcgaaa | agcgtccgcg | 8160 |
| gcccagacag | gaagaaactc | tggtgacgca | aatagatgag | cctccctcgt | acgaggaggc | 8220 |
| actaaagcaa | ggcctgcca | ccaccctg | catcgcgccc | atggctaccg | gagtgtggg | 8280 |
| ccagcacaca | cctgtaacgc | tggacctgcc | tcccccgct | gacaccaccg | agaaacctgt | 8340 |
| gctgccaggg | ccgtccgccc | ttgttgtaac | ccgccctagc | cgcgcgctcc | tgcgccgtgc | 8400 |
| cgccagcgg | ccgcgatcga | cctgcaggca | tgcaagcttg | gcgtaatcat | ggtcatagct | 8460 |
| gtttcctgtg | tgaattgtt | atccgctcac | aattccacac | aacatacgag | ccggaagcat | 8520 |
| aaagtgtaaa | gcctgggggtg | cctaattgag | gagctaactc | acattaattg | cgttgcgctc | 8580 |
| actgcccgt | ttccagtcgg | gaaacctgtc | gtgccagctg | cattaatgaa | tggccaacg | 8640 |
| cgcggggaga | ggcggtttgc | gtattggg | ctcttccgct | tcctcgtca | ctgactcgt | 8700 |
| gcgctcggtc | gttcggctgc | ggcgagcgg | atcagctcac | tcaaaggcgg | taatacggtt | 8760 |
| atccacagaa | tcaggggata | acgcaggaaa | gaacatgtga | gcaaaaaggcc | agcaaaaagc | 8820 |
| caggaaccgt | aaaaaggccg | cgttgctggc | gtttttccat | aggctccgcc | cccctgacga | 8880 |
| gcatcaca | aatcagcgt | caagtcagag | gtggcgaaac | ccgacaggac | tataaagata | 8940 |
| ccaggegttt | ccccctggaa | gctcccctgt | gcgctctcct | gttccgacct | tgccgcttac | 9000 |
| cggatacctg | tccgctttc | tcccttcggg | aagcgtggcg | ctttctcaat | gctcacgctg | 9060 |

| | | | | | | |
|-------------|-------------|-------------|------------|-------------|------------|-------|
| taggtatctc | agttcgggtgt | aggteggttcg | ctccaagctg | ggctgtgtgc | acgaaccccc | 9120 |
| cgttcagccc | gaccgctgcg | cettatecgg | taactatcgt | cttgagtcca | acccggtaag | 9180 |
| acacgactta | tcgccactgg | cagcagccac | tggtaacagg | attagcagag | cgaggatgt | 9240 |
| aggcggtgct | acagagttct | tgaagtgggtg | gcctaactac | ggctacacta | gaaggacagt | 9300 |
| atlttggtatc | tgcgctctgc | tgaagccagt | taccttcgga | aaaagagttg | gtagctcttg | 9360 |
| atccggcaaa | caaaccaccg | ctggtagcgg | tggttttttt | gtttgcaage | agcagattac | 9420 |
| gcgcagaaaa | aaaggatctc | aagaagatcc | tttgatcttt | tctacggggg | ctgacgctca | 9480 |
| gtggaacgaa | aactcacggt | aagggatttt | ggtcacgaga | ttatcaaaaa | ggatcttcac | 9540 |
| ctagatcctt | ttaaattaaa | aatgaagttt | taaatcaatc | taaagtatat | atgagtaaac | 9600 |
| ttggctctgac | agttaccaat | gcttaatcag | tgaggcacct | atctcagcga | tctgtctatt | 9660 |
| tcgttcatcc | atagttgcct | gactccccgt | cgtgtagata | actacgatac | gggagggtt | 9720 |
| accatctggc | cccagtgtcg | caatgatacc | gcgagacca | cgctcaccgg | ctccagattt | 9780 |
| atcagcaata | aaccagccag | ccggaagggc | cgagcgcaga | agtggctctg | caactttatc | 9840 |
| cgcctccatc | cagtctatta | attgttgccg | ggaagctaga | gtaagtagtt | cgccagttaa | 9900 |
| tagtttgccg | aacgttgttg | ccattgctac | aggcatcgtg | gtgtcacgct | cgctcgtttg | 9960 |
| tatggcttca | ttcagctccg | gttcccaacg | atcaaggcga | gttacatgat | ccccatggt | 10020 |
| gtgcaaaaaa | gcggttagct | ccttcgggtcc | tccgatcgtt | gtcagaagta | agttggccgc | 10080 |
| agtgttatca | ctcatggtta | tggcagcact | gcataattct | cttactgtca | tgccatccgt | 10140 |
| aagatgcttt | tctgtgactg | gtgagtactc | aaccaagtca | ttctgagaat | agtgtatgcg | 10200 |
| gcgaccgagt | tgctcttgcc | cggcgtcaat | acgggataat | accgcgccac | atagcagaac | 10260 |
| tttaaaagtg | ctcatcattg | gaaaacgttc | ttcggggcga | aaactctcaa | ggatcttacc | 10320 |
| gctgttgaga | tccagttcga | tgtaacccac | tctgtcaccc | aaactgatctt | cagcatcttt | 10380 |
| tactttcacc | agcgtttctg | ggtgagcaaa | aacaggaagg | caaaatgccg | caaaaaagg | 10440 |
| aataaggcgc | acacggaaat | gttgaatact | catactcttc | ctttttcaat | attattgaag | 10500 |
| catttatcag | ggttattgtc | tcatgagcgg | atacatatct | gaatgtatct | agaaaaataa | 10560 |
| acaaataggg | gttccgcgca | catttccccg | aaaagtgcc | cctgacgtct | aagaaacct | 10620 |
| tattatcatg | acattaacct | ataaaaatag | gcgtatcacg | aggccctttc | gtc | 10673 |

<210>104

<211>22563

<212>DNA

<213> 人工

<220>

<223> 质粒 pXba

<400>104

| | | | | | | |
|------------|------------|------------|------------|------------|-------------|-----|
| tcgcgcgttt | cggtgatgac | ggtgaaaacc | tctgacacat | gcagctcccc | gagacgggtca | 60 |
| cagcttgtct | gtaagcggat | gccgggagca | gacaagcccc | tcaggcgcgc | tcagcgggtg | 120 |
| ttggcgggtg | tcggggctgg | cttaactatg | cgcatcaga | gcagattgta | ctgagagtgc | 180 |
| accatagcgc | gtgtgaaata | ccgcacagat | gcgtaaggag | aaaataccgc | atcagcgcgc | 240 |
| atcgcctatt | caggctgcgc | aactgttggg | aaggcgcgac | ggtgcgggcc | tcttcgctat | 300 |

| | | | | | | |
|------------|------------|------------|-------------|-------------|-------------|------|
| tacgccagct | ggcgaaaggg | ggatgtgctg | caaggcgatt | aagttgggta | acgccagggt | 360 |
| tttcccagtc | acgacgttgt | aaaacgacgg | ccagtgaatt | cgagctcggg | acccggggat | 420 |
| ccttctagac | cgtgcaaaag | gagagcctgt | aagcggggcac | tcttccgtgg | tctgggtgat | 480 |
| aaattcgcaa | gggtatcatg | gcggacgacc | ggggttcgaa | ccccggatcc | ggccgtccgc | 540 |
| cgtgatccat | gcggttaccg | cccgcgtgtc | gaaccaggt | gtgcgacgtc | agacaacggg | 600 |
| ggagcgcctc | ttttggcttc | cttccaggcg | cggcggctgc | tgcgctagct | tttttgcca | 660 |
| ctggccgcgc | gcggcgtaag | cggttaggct | ggaaagcgaa | agcattaagt | ggctcgcctc | 720 |
| ctgtagccgg | agggttattt | tccaagggtt | gagtcgcagg | acccccgggt | cgagtctcgg | 780 |
| gccggccgga | ctgcggcgaa | cgggggtttg | cctccccgtc | atgcaagacc | ccgcttgcaa | 840 |
| attcctccgg | aaacagggac | gagccccctt | tttgcctttc | ccagatgcat | ccggtgctgc | 900 |
| ggcagatgcg | ccccctcct | cagcagcggc | aagagcaaga | gcagcggcag | acatgcaggg | 960 |
| caccctcccc | ttctctacc | gcgtcaggag | gggcaacatc | cgcggtgac | gcggcggcag | 1020 |
| atggtgatta | cgaacccccg | cggcgccggg | cccggcacta | cctggacttg | gaggagggcg | 1080 |
| agggcctggc | gcggctagga | gcgccctctc | ctgagcgaca | cccaagggtg | cagctgaagc | 1140 |
| gtgacacgcg | cgaggcgtac | gtgccgcggc | agaacctggt | tcgcgaccgc | gagggagagg | 1200 |
| agcccagga | gatgcgggat | cgaaagtcc | acgcagggcg | cgagttgcgg | catggcctga | 1260 |
| accgcgagcg | gttgctgcgc | gaggaggact | ttgagcccga | cgcgccgacc | gggattagtc | 1320 |
| ccgcgcgcgc | acacgtggcg | gccgccgacc | tggtaacgcg | gtacgagcag | acggtgaacc | 1380 |
| aggagattaa | ctttcaaaaa | agctttaaca | accacgtgcg | cacgcttggt | gcgcgcgagg | 1440 |
| aggtggctat | aggactgatg | catctgtggg | actttgtaag | cgcgctggag | caaaacccaa | 1500 |
| atagcaagcc | gctcatggcg | cagctgttcc | ttatagtgca | gcacagcagg | gacaacgagg | 1560 |
| cattcaggga | tgcgctgcta | aacatagtag | agcccagagg | ccgctggctg | ctcgatttga | 1620 |
| taaacattct | gcagagcata | gtggtgcagg | agcgcagctt | gagcctggct | gacaagggtg | 1680 |
| ccgccattaa | ctattccatg | ctcagtctgg | gcaagtttta | cgcccgaag | atataccata | 1740 |
| ccccttacgt | tcccatagac | aaggaggtaa | agatcgaggg | gttctacatg | cgcatggcgt | 1800 |
| tgaaggtgct | taccttgagc | gacgacctgg | gcgtttatcg | caacgagcgc | atccacaagg | 1860 |
| ccgtgagcgt | gagccggcgg | cgcgagctca | gcgaccgca | gctgatgcac | agcctgcaaa | 1920 |
| gggccctggc | tggcacgggc | agcggcgata | gagaggccga | gtcctacttt | gacgcggggc | 1980 |
| ctgacctgcg | ctgggcccc | agccgacgcg | ccctggaggc | agctggggcc | ggacctgggc | 2040 |
| tggcggtggc | acccgcgcgc | gctggcaacg | tcggcgcgct | ggaggaatat | gacgaggacg | 2100 |
| atgagtacga | gccagaggac | ggcgagtact | aagcgggtgat | gtttctgata | agatgatgca | 2160 |
| agacgcaacg | gacccggcgg | tgcggggcgg | gctgcagagc | cagccgtccg | gccttaactc | 2220 |
| cacggacgac | tggcgccagg | tcatggaccg | catcatgtcg | ctgactgcg | gtaaccctga | 2280 |
| cgcttccgg | cagcagccgc | aggccaaccg | gctctccgca | attctggaag | cggtggctcc | 2340 |
| ggcgcgcgca | aacccacgc | acgagaaggt | gctggcgatc | gtaaacgcgc | tggccgaaaa | 2400 |
| cagggccate | cggcccgatg | aggccggcct | ggtctacgac | gcgctgcttc | agcgcgtggc | 2460 |
| tcgttacaac | agcggcaacg | tgcagaccaa | cctggaccgg | ctgggtggggg | atgtgcgcga | 2520 |
| ggccgtggcg | cagcgtgagc | gcgcgcagca | gcagggcaac | ctgggctcca | tggttgcaact | 2580 |
| aaacgccttc | ctgagtacac | agcccgccaa | cgtgccgcgg | ggacaggagg | actacaccaa | 2640 |

| | | | | | | |
|------------|-------------|-------------|-------------|-------------|------------|------|
| ctttgtgagc | gcactgcggc | taatggtgac | tgagacaccg | caaagtgagg | tgtaccagtc | 2700 |
| cgggccagac | tatTTTTTcc | agaccagtag | acaaggcctg | cagaccgtaa | acctgagcca | 2760 |
| ggctttcaag | aacttgcagg | ggctgtgggg | ggtgcgggct | cccacaggcg | accgcgcgac | 2820 |
| cgtgtctage | ttgctgacgc | ccaactcgcg | cctgttctgt | ctgctaatag | cgcccttcac | 2880 |
| ggacagtggc | agcgtgtccc | gggacacata | cctaggtcac | ttgctgacac | tgtaccgca | 2940 |
| ggccataggt | caggcgcgatg | tggacgagca | tactttccag | gagattacaa | gtgtcagccg | 3000 |
| cgcgctgggg | caggaggaca | cgggcagcct | ggaggcaacc | ctgaactacc | tgctgaccaa | 3060 |
| ccggcggcag | aagateccct | cgttgcacag | tttaaacagc | gaggaggagc | gcatcttgcg | 3120 |
| ctatgtgcag | cagagcgtga | gccttaacct | gatgcgcgac | gggtaacgc | ccagcgtggc | 3180 |
| gctggacatg | accgcgcgca | acatggaacc | ggcatgtat | gcctcaaacc | ggccgtttat | 3240 |
| caatgccta | atggactact | tgcatecgcg | ggccgccgtg | aaccccgagt | atttcaccaa | 3300 |
| tgccatcttg | aacccgcact | ggctaccgcc | ccctggtttc | tacaccgggg | gatttgaggt | 3360 |
| gcccgagggt | aacgatggat | tcctctggga | cgacatagac | gacagcgtgt | tttccccgca | 3420 |
| accgcagacc | ctgctagagt | tgcaacagcg | cgagcaggca | gaggcggcgc | tgcgaaagga | 3480 |
| aagcttccgc | aggccaagca | gcttgtccga | tctaggcgct | gcggccccgc | ggtcagatgc | 3540 |
| gagtagccca | tttccaagct | tgatagggtc | ttttaccagc | actcgcacca | cccgcccgcg | 3600 |
| cctgctgggc | gaggaggagt | acctaaacaa | ctcgtctgtg | cagccgcagc | gcgaaaagaa | 3660 |
| cctgcctccg | gcatttccca | acaacgggat | agagagccta | gtggacaaga | tgagtagatg | 3720 |
| gaagacgtat | gcgcaggagc | acagggatgt | gcccggcccc | cgcccccca | cccgtcgtca | 3780 |
| aaggcacgac | cgtcagcggg | gtctgggtgtg | ggaggacgat | gactcggcag | acgacagcag | 3840 |
| cgctctggat | ttgggaggga | gtggcaacct | gtttgcgcac | cttcgccccca | ggctggggag | 3900 |
| aatgttttaa | aaaaaaaaaa | aaaagcatg | atgcaaaaata | aaaaactcac | caaggccatg | 3960 |
| gcaccgagcg | ttggttttct | tgtattcccc | ttagtatgca | gcgcgcggcg | atgtatgagg | 4020 |
| aaggtcctcc | tcctcctac | gagagcgtgg | tgagcgcggc | gccagtggcg | gcggcgctgg | 4080 |
| gttccccctt | cgatgctccc | ctggaccgcg | cgtttgtgcc | tccgcggtac | ctgcggccta | 4140 |
| ccggggggag | aaacagcadc | cgttactctg | agttggcacc | cctattcgac | accacccgtg | 4200 |
| tgtaccttgt | ggacaacaag | tcaacggatg | tggcatccct | gaactaccag | aacgaccaca | 4260 |
| gcaactttct | aaccacggtc | attcaaaaaca | atgactacag | cccgggggag | gcaagcacac | 4320 |
| agaccatcaa | tcttgacgac | cgttcgact | ggggcggcga | cctgaaaacc | atcctgcata | 4380 |
| ccaacatgcc | aaatgtgaac | gagttcatgt | ttaccaataa | gtttaaggcg | cgggtgatgg | 4440 |
| tgtegcgctc | gcttactaag | gacaaacagg | tggagctgaa | atatgagtgg | gtggagtcca | 4500 |
| cgctgcccga | gggcaactac | tccgagacca | tgaccataga | ccttatgaac | aacgcgatcg | 4560 |
| tggagcacta | cttgaaagtg | ggcaggcaga | acggggttct | ggaaaagcgc | atcggggtaa | 4620 |
| agtttgacac | ccgcaacttc | agactggggg | ttgaccaggt | cactgggtctt | gtcatgcctg | 4680 |
| gggtatatac | aaacgaagcc | ttccatccag | acatcatttt | gctgccagga | tgcgggggtg | 4740 |
| acttaccca | cagccgcctg | agcaacttgt | tgggcatccg | caagcggcaa | cccttcagg | 4800 |
| agggctttag | gateacctac | gatgacctgg | agggtggtaa | cattccccgca | ctgttggatg | 4860 |
| tggacgccta | ccaggcaagc | ttaaaagatg | acaccgaaca | gggcggggat | ggcgcaggcg | 4920 |
| gcggcaacaa | cagtggcagc | ggcgcggaag | agaactccaa | cgccgcagcc | gcggcaatgc | 4980 |

| | | | | | | |
|-------------|------------|------------|------------|-------------|------------|------|
| agccggtgga | ggacatgaac | gatcatgcca | ttcgcggcga | cacctttgcc | acacgggcgg | 5040 |
| aggagaagcg | cgctgaggcc | gaggcagcgg | cagaagctgc | cgcccccgct | gcgcaaccgg | 5100 |
| aggtcgagaa | gcctcagaag | aaaccggtga | tcaaaccctt | gacagaggac | agcaagaaac | 5160 |
| gcagttacaa | cctaataage | aatgacagca | ccttcacca | gtaccgcage | tggtagcttg | 5220 |
| catacaacta | cggcgacct | cagaccggga | tccgctcatg | gacctcctt | tgcactcctg | 5280 |
| acgtaacctg | cggtctggag | caggtctact | ggctggtgcc | agacatgatg | caagaccccg | 5340 |
| tgaccttccg | ctccacgagc | cagatcagca | actttccggt | ggtgggcgcc | gagctgttgc | 5400 |
| ccgtgcactc | caagagcttc | tacaacgacc | aggccgteta | ctcccagctc | atccgccagt | 5460 |
| ttacctctct | gaccacagtg | ttcaatcgct | ttcccagaaa | ccagattttg | gcgccccgc | 5520 |
| cagccccac | catcaccacc | gtcagtgaaa | acgttcctgc | tctcacagat | cacgggacgc | 5580 |
| taccgctgcg | caacagcacc | ggaggagtcc | agcgagtgc | cattactgac | gccagacgcc | 5640 |
| gcacctgccc | ctacgtttac | aaggccctgg | gcatagtctc | gccgcgcgctc | ctatcgagcc | 5700 |
| gcactttttg | agcaaacatg | tccatcctta | tatcgcccag | caataacaca | ggctggggcc | 5760 |
| tgcgcttccc | aagcaagatg | tttggcgggg | caaagaagcg | ctccgaccaa | caccagtg | 5820 |
| gcgtgcgcgg | gcactaccgc | gcgccctggg | gcgcgcacaa | acgcggccgc | actgggcgca | 5880 |
| ccaccgtcga | tgaccgcatt | gacgcggtgg | tggaggaggc | gcgcaactac | acgcccacgc | 5940 |
| cgccaccagt | gtccacagtg | gacgcggcca | ttcagaccgt | ggtgcgcgga | gcccggcgtt | 6000 |
| atgctaaaaat | gaagagacgg | cggaggcgcg | tagcacgtcg | ccaccgccgc | cgaccgggca | 6060 |
| ctgccgcca | acgcgcggcg | gcggccctgc | ttaaccgcgc | acgtcgcacc | ggccgacggg | 6120 |
| cggccatgcg | ggcgcctcga | aggctggccg | cgggtattgt | actgtgccc | cccaggcca | 6180 |
| ggcgacgagc | ggcgcgcgca | gcagccgcgg | ccattagtgc | tatgactcag | ggtcgcaggg | 6240 |
| gcaacgtgta | ctgggtgctc | gactcgggta | gcggcctgcg | cgtgcccgtg | cgcaaccgcc | 6300 |
| ccccgcgcaa | ctagattgca | agaaaaact | acttagactc | gtactgttgt | atgtatccag | 6360 |
| cggcggcggc | gcgcaacgaa | gctatgtcca | agcgcaaaat | caaagaagag | atgtccagg | 6420 |
| tcatcgcgcc | ggagatctat | ggcccccgca | agaaggaaga | gcaggattac | aagccccgaa | 6480 |
| agctaaagcg | ggtcaaaaag | aaaaagaaag | atgatgatga | tgatgaactt | gacgacgagg | 6540 |
| tggaactgct | gcacgcaacc | gcgcccaggc | ggcgggtaca | gtggaaaggt | cgacgcgtaa | 6600 |
| gacgtgtttt | gcgacccggc | accaccgtag | tttttacgcc | cggtagcgc | tccaccgca | 6660 |
| cctacaagcg | cgtgtatgat | gaggtgtacg | gcgacgagga | cctgcttgag | caggccaacg | 6720 |
| agcgcctcgg | ggagtttgcc | tacggaaagc | ggcataagga | catgttggcg | ttgccgctgg | 6780 |
| acgagggcaa | cccaacacct | agcctaaagc | ccgtgacact | gcagcaggtg | ctgcccacgc | 6840 |
| ttgcaccgtc | cgaagaaaag | cgcgccctaa | agcgcgagtc | tggtagcttg | gcaccaccg | 6900 |
| tgacagctgat | ggtaccaaac | cgccagcgac | tggaagatgt | cttggaaaaa | atgaccgtgg | 6960 |
| agcctgggct | ggagcccag | gtccgcgtgc | ggccaatcaa | gcaggtggca | ccgggactgg | 7020 |
| gcgtgcagac | cgtggacgtt | cagataccca | ccaccagtag | cactagtatt | gccactgcca | 7080 |
| cagagggcac | ggagacacaa | acgtccccgg | ttgcctcggc | ggtggcagat | gccgcggtgc | 7140 |
| aggcggccgc | tgcggccgcg | tccaaaacct | ctacggaggt | gcaaacggac | ccgtggatgt | 7200 |
| ttcgcgtttc | agcccccg | cgccccgc | gttccaggaa | gtacggcacc | gccagcgcac | 7260 |
| tactgcccga | atatgcctta | catccttcca | tcgcgcctac | ccccggctat | cgtggctaca | 7320 |

| | | | | | | |
|-------------|------------|-------------|-------------|-------------|-------------|------|
| cctaccgccc | cagaagacga | gcgactacce | gacgccgaac | caccactgga | acccgccgcc | 7380 |
| gccgtcgccg | tcgccagccc | gtgctggccc | cgatttccgt | gcgcaggggtg | gctcgcgaag | 7440 |
| gaggcaggac | cctggtgctg | ccaacagegc | gctaccaccc | cagcatcggt | taaaagccgg | 7500 |
| tctttgtggt | tcttgcatat | atggcectca | cctgccgcct | ccgtttcccc | gtgccgggat | 7560 |
| tccgaggaag | aatgcaccgt | aggaggggca | tggccggcca | cggcctgacg | ggcggcatgc | 7620 |
| gtcgtgcgca | ccaccggcgg | cggegcgcgt | cgcaccgtcg | catgcgcggc | ggtatcctgc | 7680 |
| ccctccttat | tccactgate | gccgcggcga | ttggcgccgt | gcccgggaatt | gcatccgtgg | 7740 |
| ccttgccaggc | gcagagacac | tgattaaaaa | caagttgcat | gtggaaaaat | caaaataaaa | 7800 |
| agtctggagt | ctcacgctcg | cttggctctg | taactatttt | gtagaatgga | agacatcaac | 7860 |
| tttgcgtctc | tggccccgcg | acacggctcg | cgccccgttca | tgggaaactg | gcaagatata | 7920 |
| ggcaccagca | atatgagcgg | tggcgccttc | agctggggct | cgctgtggag | cggcattaaa | 7980 |
| aatttcggtt | ccaccattaa | gaactatggc | agcaaggcct | ggaacagcag | cacaggccag | 8040 |
| atgctgaggg | acaagttgaa | agagcaaaat | ttccaacaaa | aggtggtaga | tggcctggcc | 8100 |
| tctggcatta | gcggggtggt | ggacctggcc | aaccaggcag | tgcaaaataa | gattaacagt | 8160 |
| aagcttgatc | cccgcctcc | cgtagaggag | cctccaccgg | ccgtggagac | agtgtctcca | 8220 |
| gaggggcgtg | gcaaaaagcg | tccgcggccc | gacaggggag | aaactctggt | gacgcaataa | 8280 |
| gatgagcctc | cctcgtacga | ggaggcacta | aagcaaggcc | tgcccaccac | ccgtcccata | 8340 |
| gccccatgg | ctaccggagt | gctgggcccag | cacacacctg | taacgctgga | cctgcctccc | 8400 |
| cccgtgaca | cccagcagaa | acctgtgctg | ccagggccgt | ccgccgttgt | tgtaaccgca | 8460 |
| cctagccgcg | cgteccctcg | ccgtgccgcc | agcggctccgc | gatcgatgcg | gcccgtagcc | 8520 |
| agtggcaact | ggcaaagcac | actgaacagc | atcgtgggtc | tgggggtgca | atccctgaag | 8580 |
| cgccgacgat | gcttctaaat | agctaactgt | tcgtatgtgt | catgtatgcg | tccatgtcgc | 8640 |
| cgccagagga | gctgctgagc | cgccgtgcgc | ccgctttcca | agatggctac | cccttcgatg | 8700 |
| atgccgcagt | ggtcttacat | gcacatctcg | ggccaggacg | cctcggagta | cctgagcccc | 8760 |
| gggctggtgc | agtttgcccc | cgccaccgag | acgtacttca | gcctgaataa | caagtttaga | 8820 |
| aacccacagg | tggcacctac | gcacgacgta | accacagacc | ggtcccagcg | tttgacgctg | 8880 |
| cggttcatac | ctgtggaccg | cgaggatacc | gcgtactcgt | acaaaagcgc | gttcaccctg | 8940 |
| gctgtgggtg | acaaccgtgt | gcttgatatg | gcttccacct | actttgacat | ccgcggcgtg | 9000 |
| ctggacaggg | ggcctacttt | taagccctac | tccggcactg | cctacaacgc | tctagctccc | 9060 |
| aagggcgctc | ctaactcctg | tgagtgggaa | caaaccgaag | atagcggccc | ggcagttgcc | 9120 |
| gaggatgaag | aagaggaaga | tgaagatgaa | gaagaggaag | aagaagagca | aaacgctcga | 9180 |
| gatcaggcta | ctaagaaaac | acatgtctat | gcccaggctc | ctttgtctgg | agaaaacaatt | 9240 |
| acaaaaagcg | ggctacaaat | aggatcagac | aatgcagaaa | cacaagctaa | acctgtatac | 9300 |
| gcagatcctt | cctatcaacc | agaacctcaa | attggcgaat | ctcagtggaa | cgaagctgat | 9360 |
| gctaatacgg | caggagggag | agtgcctaaa | aaaacaactc | ccatgaaacc | atgctatgga | 9420 |
| tcttatgcca | ggcctacaaa | tccttttggg | ggtcaatccg | ttctggttcc | ggatgaaaaa | 9480 |
| gggggtgcctc | ttccaaaggt | tgacttgcaa | ttcttctcaa | atactacctc | tttgaacgac | 9540 |
| cggcaaggca | atgctactaa | accaaaagtg | gttttgtaca | gtgaagatgt | aaatatggaa | 9600 |
| accccagaca | cacatctgtc | ttacaaacct | ggaaaagggtg | atgaaaattc | taaagctatg | 9660 |

| | | | | | | |
|-------------|------------|------------|------------|-------------|-------------|-------|
| ttgggtcaac | aatctatgcc | aaacagacce | aattacattg | ctttcagggga | caatttttatt | 9720 |
| ggcctaattgt | attataacag | cactggcaac | atgggtgttc | ttgctgggtca | ggcatcgcag | 9780 |
| ctaaatgccg | tggtagattt | gcaagacaga | aacacagagc | tgtcctatca | actcttgctt | 9840 |
| gattccatag | gtgatagaac | cagatatttt | tctatgtgga | atcaggctgt | agacagctat | 9900 |
| gatccagatg | ttagaatcat | tgaaaacat | ggaactgagg | atgaattgcc | aaattattgt | 9960 |
| tttctcttg | ggggtattgg | ggtaactgac | acctatcaag | ctattaagge | taatggcaat | 10020 |
| ggctcaggcg | ataatggaga | tactacatgg | acaaaagatg | aaacttttgc | aacacgtaat | 10080 |
| gaaataggag | tgggtaacaa | ctttgceatg | gaaattaacc | taaatgcaa | cctatggaga | 10140 |
| aatttccttt | actccaatat | tgcgctgtac | ctgccagaca | agctaaaata | caaccccacc | 10200 |
| aatgtggaaa | tatctgacaa | ccccaacacc | tacgactaca | tgaacaagcg | agtgggtggct | 10260 |
| cccgggcttg | tagactgcta | cattaacctt | ggggcgcgct | ggtctctgga | ctacatggac | 10320 |
| aacgttaatc | cctttaacca | ccaccgcaat | gcgggcctcc | gttatcgctc | catgttgttg | 10380 |
| ggaaacggcc | gctacgtgcc | ctttcacatt | caggtgcccc | aaaagttttt | tgccattaaa | 10440 |
| aacctcctcc | tcctgccagg | ctcatataca | tatgaatgga | acttcaggaa | ggatgttaac | 10500 |
| atggttctgc | agagctctct | gggaaacgat | cttagagttg | acggggctag | cattaagttt | 10560 |
| gacagcattt | gtctttacgc | caccttcttc | cccatggccc | acaacacggc | ctccacgctg | 10620 |
| gaagccatgc | tcagaaatga | caccaacgac | cagtccttta | atgactacct | ttccgccgcc | 10680 |
| aacatgctat | accccatacc | cgccaacgcc | accaacgtgc | ccatctccat | cccatcgcgc | 10740 |
| aactgggcag | catttcgcgg | ttgggccttc | acacgcttga | agacaaagga | aaccccttcc | 10800 |
| ctgggatcag | gctacgacce | ttactacacc | tactctggct | ccataccata | ccttgacgga | 10860 |
| accttctate | ttaatcacac | ctttaagaag | gtggccatta | cctttgactc | ttctgttagc | 10920 |
| tggccgggca | acgaccgcct | gcttactccc | aatgagtttg | agattaaacg | ctcagttgac | 10980 |
| ggggaggggct | acaacgtagc | tcagtgaac | atgaccaagg | actggttcct | ggtgcagatg | 11040 |
| ttggccaact | acaatattgg | ctaccagggc | ttctacattc | cagaaaagcta | caaggaccgc | 11100 |
| atgtactcgt | tcttcagaaa | cttccagccc | atgagccggc | aagtggttga | cgatactaaa | 11160 |
| tacaaggagt | atcagcaggt | tggaattctt | caccagcata | acaactcagg | attcgttaggc | 11220 |
| tacctcgctc | ccaccatgcg | cgagggacag | gcttaccocg | ccaacgtgcc | ctaccacta | 11280 |
| ataggcaaaa | ccgcggttga | cagtattacc | cagaaaaagt | ttctttgcca | tcgcaccctt | 11340 |
| tggcgcatec | cattctccag | taactttatg | tccatgggcg | cactcacaga | cctgggcaaa | 11400 |
| aaccttctct | acgccaactc | cgcccacgcg | ctagacatga | cttttgaggt | ggatcccatg | 11460 |
| gacgagccca | cccttcttta | tgttttgttt | gaagtctttg | acgtgggtccg | tgtgcaccag | 11520 |
| ccgcaccgcg | gcgtcatcga | gaccgtgtac | ctgcgcacgc | ccttctcggc | cggcaacgcc | 11580 |
| acaacataaa | agaagcaagc | aacatcaaca | acagctgccg | ccatgggctc | cagtgagcag | 11640 |
| gaactgaaag | ccattgtcaa | agatcttggg | tgtgggcat | atTTTTTggg | cacctatgac | 11700 |
| aagcgtttc | caggctttgt | ttctccacac | aagctcgcct | gcgcatagt | caatacggcc | 11760 |
| ggtcgcgaga | ctgggggctg | acactggatg | gcctttgcct | ggaacccgcg | ctcaaaaaca | 11820 |
| tgtacctct | ttgagccctt | tggcttttct | gaccaacgac | tcaagcaggt | ttaccagttt | 11880 |
| gagtacgagt | cactcctgcg | ccgtagcgcc | attgcttctt | cccccgaccg | ctgtataacg | 11940 |
| ctggaaaagt | ccacccaaag | cgtgcagggg | cccaactcgg | ccgcctgtgg | actattctgc | 12000 |

| | | | | | | |
|-------------|-------------|-------------|-------------|-------------|-------------|-------|
| tgc atgtttc | tccacgcctt | tgccaactgg | ccccaaactc | ccatggatca | caaccccacc | 12060 |
| atgaacctta | ttaccggggt | accecaactcc | atgettaaca | gtccccaggt | acagcccacc | 12120 |
| ctgcgtcgca | accaggaaca | gctctacage | ttcctggagc | gccactcgcc | ctacttccgc | 12180 |
| agccacagtg | cgcagattag | gagcgccact | tctttttgtc | acttgaaaaa | catgtaaaaa | 12240 |
| taatgtacta | ggagacactt | tcaataaagg | caaatgtttt | tatttgtaca | ctctcgggtg | 12300 |
| attatttacc | ccccaccctt | gcegtctgcg | ccgtttaaaa | atcaaagggg | ttctgccgcg | 12360 |
| catcgctatg | cgccactggc | agggacacgt | tgcgatactg | gtgttttagtg | ctccacttaa | 12420 |
| actcaggcac | aaccatccgc | ggcagctcgg | tgaagttttc | actccacagg | ctgcgcacca | 12480 |
| tcaccaacgc | gttttagcagg | tcgggcgccc | atatcttgaa | gtcgcagttg | gggcctccgc | 12540 |
| cctgcgcgcg | cgagttgcga | tacacagggg | tgcagcactg | gaacactatc | agcgcgggt | 12600 |
| ggtgcacgct | ggccagcacg | ctcttgctgg | agatcagatc | cgcgtccagg | tcctccgcgt | 12660 |
| tgctcagggc | gaacggagtc | aactttggta | gctgccttcc | caaaaagggg | gcatgcccag | 12720 |
| gctttgagtt | gcactcgcac | cgtagtggca | tcagaaggtg | accgtgcccg | gtctgggcgt | 12780 |
| taggatacag | cgcctgcatg | aaagccttga | tctgcttaaa | agccacctga | gcctttgcgc | 12840 |
| cttcagagaa | gaacatgccg | caagacttgc | cggaaaactg | attggccgga | caggccgcgt | 12900 |
| catgcacgca | gcaccttgcg | tcgggtgttg | agatctgcac | cacatttcgg | cccaccggg | 12960 |
| tcttcacgat | cttggccttg | ctagactgct | ccttcagcgc | gcgctgcccg | ttttcgctcg | 13020 |
| tcacatccat | ttcaatcacg | tgctccttat | ttatcataat | gctcccgtgt | agacacttaa | 13080 |
| gctgccttc | gatctcagcg | cagcgggtgca | gccacaacgc | gcagcccgtg | ggctcgtggt | 13140 |
| gctttaggtt | tacctctgca | aacgactgca | ggtacgcctg | caggaatcgc | cccatcatcg | 13200 |
| tcacaaaggt | cttgttgctg | gtgaaggcca | gctgcaacce | gcggtgctcc | tcgtttagcc | 13260 |
| aggtcttgca | tacggccgcc | agagcttcca | cttggtcagg | cagtagcttg | aagtttgcct | 13320 |
| ttagatcggt | atccacgtgg | tacttgtcca | tcaacgcgcg | cgcagcctcc | atgcccttct | 13380 |
| cccacgcaga | cacgateggc | aggctcagcg | ggtttatcac | cgtgctttca | ctttccgctt | 13440 |
| cactggacte | ttcttttcc | tcttgctcc | gcataccccg | cgccactggg | tcgtcttcat | 13500 |
| tcagccgccg | caccgtgcgc | ttacctccct | tgccgtgctt | gattagcacc | ggtgggttgc | 13560 |
| tgaaaccac | cattttagtc | gccacatctt | ctctttcttc | ctcgtgtcc | acgatcacct | 13620 |
| ctggggatgg | cgggcgctcg | ggcttgggag | aggggcgctt | ctttttcttt | ttggacgcaa | 13680 |
| tggccaaatc | cgccgtcgag | gtcgatggcc | gcgggctggg | tgtgcgcggc | accagcgcac | 13740 |
| cttgtgacga | gtcttcttcg | tcctcggact | cgagacgccg | cctcagccgc | ttttttgggg | 13800 |
| gcgcgcgggg | aggcggcggc | gacggcgacg | gggacgacac | gtcctccatg | gttgggtggac | 13860 |
| gtcgcgccgc | accgcgtccg | cgctcggggg | tggtttcgcg | ctgctcctct | tcccactgg | 13920 |
| ccatttcctt | ctcctatagg | cagaaaaaga | tcattggagtc | agtcgagaag | gaggacagcc | 13980 |
| taaccgcccc | ctttgagttc | gccaccaccg | cctccaccga | tgccgccaac | gcgcctacca | 14040 |
| ccttccccgt | cgaggcaccc | ccgcttgagg | aggaggaagt | gattatcgag | caggaccag | 14100 |
| gttttgtaag | cgaagacgac | gaggatcgct | cagtaccaac | agaggataaa | aagcaagacc | 14160 |
| aggacgacgc | agaggcaaac | gaggaacaag | tcgggcgggg | ggaccaaagg | catggcgact | 14220 |
| acctagatgt | gggagacgac | gtgctgttga | agcatctgca | gcgccagtc | gccattatct | 14280 |
| gcgacgcggt | gcaagagcgc | agcgatgtgc | ccctcgccat | agcggatgtc | agccttgcct | 14340 |

| | | | | | | |
|------------|-------------|-------------|------------|------------|-------------|-------|
| acgaacgcca | cctgttctca | cegecgctac | cccccaaacg | ccaagaaaac | ggcacatgcg | 14400 |
| agcccaacce | gcgcctcaac | ttctaceccg | tatttgccgt | gccagaggtg | cttgccacct | 14460 |
| atcacatett | tttccaaaac | tgcaagatac | ccctatcctg | ccgtgccaac | cgcagccgag | 14520 |
| cggacaagca | gctggccttg | cggcagggcg | ctgtcatacc | tgatatcgcc | tcgctcgacg | 14580 |
| aagtgccaaa | aatctttgag | ggtcttggac | gcgacgagaa | acgcgcggca | aacgctctgc | 14640 |
| aacaagaaaa | cagegaaaat | gaaagteact | gtggagtgtc | ggtggaactt | gagggtgaca | 14700 |
| acgcgcgcct | agccgtgctg | aaacgcagca | tcgaggteac | ccactttgcc | taccggcac | 14760 |
| ttaacctacc | ccccaaagtt | atgagcacag | tcatgagcga | gctgatcgtg | cgccgtgcac | 14820 |
| gacccttgga | gagggatgca | aacttgcaag | aacaaaccga | ggagggccta | cccgcagttg | 14880 |
| gcgatgagca | gctggcgcgc | tggttgaga | cgcgcgagcc | tgccgacttg | gaggagcgac | 14940 |
| gcaagcta | gatggccgca | gtgcttgta | ccgtggagct | tgagtgcacg | cagcggttct | 15000 |
| ttgctgacct | ggagatgcag | cgcaagctag | aggaaacggt | gcactacacc | tttcgccagg | 15060 |
| gctacgtgcg | ccaggcctgc | aaaatttcca | acgtggagct | ctgcaacctg | gtctcctacc | 15120 |
| ttggaat | gcacgaaaac | cgctcggggc | aaaacgtgct | tcattccacg | ctcaagggcg | 15180 |
| aggcgcgccg | cgactacgtc | cgcgactgcg | tttacttatt | tctgtgctac | acctggcaaa | 15240 |
| cggccatggg | cgtgtggcag | caatgcctgg | aggagcgcaa | cctaaaggag | ctgcagaagc | 15300 |
| tgctaaagca | aaacttgaag | gacctatgga | cggccttcaa | cgagcgctcc | gtggccgcgc | 15360 |
| acctggcgga | cattatcttc | cccgaacgcc | tgcttaaaac | cctgcaacag | ggtctgccag | 15420 |
| acttcaccag | tcaaagcatg | ttgcaaaact | ttaggaactt | tatcctagag | cgttcaggaa | 15480 |
| ttctgcccgc | cacctgctgt | gcgcttccca | gcgactttgt | gcccattaag | taccgtgaat | 15540 |
| gccctccgcc | gctttggggg | caactgctacc | ttctgcagct | agccaactac | cttgcctacc | 15600 |
| actccgacat | catggaagac | gtgagcggtg | acggcctact | ggagtgtcac | tgctcgtgca | 15660 |
| acctatgcac | cccgcaccgc | tccttggctt | gcaattcgca | actgcttagc | gaaagtcaaa | 15720 |
| ttatcggtag | ctttgagctg | cagggtecc | cgctgacga | aaagtccgcg | gctccggggg | 15780 |
| tgaaactcac | tccggggctg | tggaagctgg | cttaccttcg | caaatttgta | cctgaggact | 15840 |
| accacgcca | cgagattagg | ttctacgaag | accaatcccg | cccgccaaat | gcggagctta | 15900 |
| ccgcctgcgt | cattaccag | ggccacatcc | ttggccaatt | gcaagccatc | aacaaaagccc | 15960 |
| gccaagagtt | tctgctacga | aagggacggg | gggtttacct | ggacccccag | tccggcgagg | 16020 |
| agctcaacce | aatccccccg | ccgcgcgagc | cctatcagca | gccgcgggcc | cttgcttccc | 16080 |
| aggatggcac | ccaaaaagaa | gctgcagctg | ccgccgccgc | caccacgga | cgaggaggaa | 16140 |
| tactgggaca | gtcaggcaga | ggaggttttg | gacgaggagg | aggagatgat | ggaagactgg | 16200 |
| gacagcctag | acgaagcttc | cgaggccgaa | gaggtgtcag | acgaaacacc | gtcaccctcg | 16260 |
| gtcgattcc | cctcgccggc | gccccagaaa | ttggcaaccg | ttcccagcat | cgctacaacc | 16320 |
| tccgctcttc | aggcgccgcc | ggcactgcct | gttcgcccgc | ccaaccgtag | atgggacacc | 16380 |
| actggaacca | gggcccggtaa | gtctaagcag | ccgccgccgt | tagcccaaga | gcaacaacag | 16440 |
| cgccaagget | accgctcgtg | gcgcggggac | aagaacgcca | tagttgcttg | cttgcaagac | 16500 |
| tgtgggggca | acatctcttc | cgcccgcgc | ttcttctct | accatcacgg | cgtggccttc | 16560 |
| ccccgtaaca | tctgcat | ctaccgtcat | ctctacagcc | cctactgcac | cggcggcagc | 16620 |
| ggcagcggca | gcaacagcag | cggtcacaca | gaagcaaagg | cgaccggata | gcaagactct | 16680 |

| | | | | | | |
|------------|------------|------------|-------------|-------------|------------|-------|
| gacaaagccc | aagaaatcca | cagcggcggc | agcagcagga | ggaggagcgc | tgcgtctggc | 16740 |
| gccaacgaa | cccgtatcga | cccgcgagct | tagaaatagg | atTTTTccca | ctctgtatgc | 16800 |
| tatatttcaa | caaagcaggg | gccaagaaca | agagctgaaa | ataaaaaaca | ggtctctgcg | 16860 |
| ctccctcacc | cgcagctgcc | tgtatcacia | aagcgaagat | cagcttcggc | gcacgctgga | 16920 |
| agacgcggag | gctctcttca | gcaaatactg | cgcgctgact | cttaaggact | agtttcgctc | 16980 |
| cTTTTctcaa | atttaagegc | gaaaactacg | tcctctccag | cggccacacc | cggcgccagc | 17040 |
| acctgtcgtc | agcgccatta | tgagcaagga | aattcccacg | ccctacatgt | ggagttacca | 17100 |
| gccacaaatg | ggacttgccg | ctggagctgc | ccaagactac | tcaacccgaa | taaactacat | 17160 |
| gagcgcggga | ccccacatga | tatcccgggt | caacggaatc | cgcgcccacc | gaaaccgaat | 17220 |
| tctcctcgaa | cagcgggcta | ttaccaccac | acctcgtaat | aaccttaate | cccgtagttg | 17280 |
| gcccgtgcc | ctggtgtacc | aggaaagtcc | cgtcccacc | actgtggtac | ttcccagaga | 17340 |
| cggccaggcc | gaagttcaga | tgactaacte | aggggcgcag | cttgcgggcg | gctttcgtca | 17400 |
| cagggctcgg | tcgcccgggc | agggtataac | tcacctgaaa | atcagagggc | gaggtattca | 17460 |
| gctcaacgac | gagtcggtga | gctcctctct | tggtctccgt | ccggacggga | catttcagat | 17520 |
| cggcggcgct | ggcgcctctt | catttacgcc | cgtcaggcgc | atcctaacte | tgcacacctc | 17580 |
| gtcctcggag | cgcgcctccg | gaggcattgg | aactctacaa | tttattgagg | agttcgtgcc | 17640 |
| ttcggtttac | ttcaaccctt | ttctggacc | tcccggccac | taccggacc | agtttattcc | 17700 |
| caactttgac | gcggtgaaag | actcggcgga | cggctacgac | tgaatgacca | gtggagaggc | 17760 |
| agagcgactg | cgcctgacac | acctcgacca | ctgccgccc | cacaagtgtc | ttgcccgcgg | 17820 |
| ctccggtgag | ttttgttact | ttgaattgcc | cgaagagcat | atcagagggc | cggcgcacgg | 17880 |
| cgtccggctc | accaccaggg | tagagcttac | acgtagcctg | attcgggagt | ttaccaagcg | 17940 |
| ccccctgcta | gtggagcggg | agcggggctc | ctgtgttctg | accgtggttt | gcaactgtcc | 18000 |
| taaccctgga | ttacatcaag | atctttgttg | tcctctctgt | gctgagtata | ataaatacag | 18060 |
| aaattagaat | ctactggggc | tctgtcgc | atcctgtgaa | cgccaccggt | tttaccacc | 18120 |
| caaagcagac | caaagcaaac | ctcacctccg | gtttgcacaa | gcgggccaat | aagtacctta | 18180 |
| cctggtactt | taacggctct | tcatttgtaa | tttacaacag | tttccagcga | gacgaagtaa | 18240 |
| gtttgccaca | caactttctc | ggcttcaact | acaccgtcaa | gaaaaacacc | accaccacca | 18300 |
| ccctcctcac | ctgccgggaa | cgtacgagtg | cgtcaccggt | tgctgcgccc | acacctacag | 18360 |
| cctgagcgta | accagacatt | actcccattt | ttccaaaaca | ggaggtgagc | tcaactcccg | 18420 |
| gaactcaggt | caaaaaagca | ttttgcgggg | tgctgggatt | ttttaattaa | gtatatgagc | 18480 |
| aattcaagta | actctacaag | cttgtctaat | ttttctggaa | ttggggtcgg | ggttatcctt | 18540 |
| actcttgtaa | ttctgtttat | tcttatacta | gcacttctgt | gccttagggg | tgccgcctgc | 18600 |
| tgacgcacg | tttgtacctt | ttgtcagctt | tttaaacgct | gggggcaaca | tccaagatga | 18660 |
| ggtacatgat | tttaggcttg | ctcgccttg | cggcagctctg | cagcgcctgc | aaaaaggttg | 18720 |
| agtttaagga | accagcttgc | aatgttacat | ttaaatcaga | agctaataaa | tgactactc | 18780 |
| ttataaaatg | caccacagaa | catgaaaagc | ttattattcg | ccacaaagac | aaaattggca | 18840 |
| agtatgctgt | atatgctatt | tggcagccag | gtgacactaa | cgactataat | gtcacagtct | 18900 |
| tccaaggtga | aaatcgtaaa | acttttatgt | ataaatttcc | atTTTTatgaa | atgtgcgata | 18960 |
| ttaccatgta | catgagcaaa | cagtacaagt | tgtggccccc | acaaaagtgt | ttagagaaca | 19020 |

| | | | | | | |
|-------------|-------------|-------------|-------------|------------|-------------|-------|
| ctggcacctt | ttgttccacc | gctetgetta | ttacagcgct | tgctttggta | tgtaccttac | 19080 |
| tttatctcaa | atacaaaagc | agacgcagtt | ttattgatga | aaagaaaatg | ccttgatttt | 19140 |
| ccgcttgctt | gtattcccct | ggacaattta | ctctatgtgg | gatatgctcc | aggcgggcaa | 19200 |
| gattataccc | acaaccttea | aatecaactt | tcctggacgt | tagcgcctga | tttctgccag | 19260 |
| cgcttgcact | gcaaatttga | tcaaaccacg | cttcagcttg | cctgctccag | agatgaccgg | 19320 |
| ctcaaccate | gcgcccacaa | eggactateg | caacaccact | gctaccggac | taacatctgc | 19380 |
| cctaaattta | ccccagttc | atgcctttgt | caatgactgg | gcgagcttgg | acatgtggtg | 19440 |
| gttttccata | gcgcttatgt | ttgtttgctt | tattattatg | tggtttattt | gttgccataa | 19500 |
| gcgcagacgc | gccagacccc | ccatctatag | gcctatcatt | gtgctcaacc | cacacaatga | 19560 |
| aaaaattcat | agattggacg | gtctgaaacc | atgttctctt | cttttacagt | atgattaaat | 19620 |
| gagacatgat | tcctcgagtt | cttatattat | tgacccttgt | tgcgcttttc | tgtgctgtct | 19680 |
| ctacattggc | cgcggtcgct | cacatcgaag | tagattgcat | cccaccttcc | acagtttacc | 19740 |
| tgctttacgg | atltgtcacc | cttatectca | tctgcagcct | cgctactgta | gtcatcgctt | 19800 |
| tcattcagtt | cattgactgg | gtttgtgtgc | gcattgcgta | cctcaggcac | catccgcaat | 19860 |
| acagagacag | gactatagct | gatcttctca | gaattcttta | attatgaaac | ggagtgtcat | 19920 |
| ttttgttttg | ctgatttttt | gcgcccctacc | tgtgctttgc | tcccaaacct | cagcgcctcc | 19980 |
| caaaagacat | atltcttgca | gattcactca | aatatggaac | attcccagct | gtacaacaa | 20040 |
| acagagcgat | ttgtcagaag | cctggttata | cgccatcctc | tctgtcatgg | ttttttgcag | 20100 |
| taccattttt | gccctagcca | tatatccata | ccttgacatt | ggctggaatg | ccatagatgc | 20160 |
| catgaaccac | cctactttcc | cagtgcocgc | tgtcatacca | ctgcaacagg | ttattgcccc | 20220 |
| aatcaatcag | cctcgccccc | cttctcccac | ccccactgag | attagctact | ttaatttgac | 20280 |
| aggtggagat | gactgaatct | ctagagtcga | cctgcaggca | tgcaagcttg | gcgtaatcat | 20340 |
| ggctcatagct | gtttctctgt | tgaattgttt | atccgctcac | aattccacac | aacatacgag | 20400 |
| ccggaagcat | aaagtgtaaa | gcctgggggtg | cctaatgagt | gagctaactc | acattaattg | 20460 |
| cgttgcgctc | actgcccgct | ttccagtcgg | gaaacctgtc | gtgccagctg | cattaatgaa | 20520 |
| tcggccaacg | cgcggggaga | ggcggtttgc | gtattgggcg | ctcttccgct | tcctcgctca | 20580 |
| ctgactcgct | gcgctcggtc | gttcggctgc | ggcgagcggt | atcagctcac | tcaaaggcgg | 20640 |
| taatacggtt | atccacagaa | tcaggggata | acgcaggaaa | gaacatgtga | gcaaaaaggcc | 20700 |
| agcaaaaaggc | caggaaccgt | aaaaaggccg | cgttgctggc | gtttttccat | aggctccgcc | 20760 |
| cccctgacga | gcatcacaaa | aategacgct | caagtccagag | gtggcgaaac | ccgacaggac | 20820 |
| tataaagata | ccaggcgttt | ccccctggaa | gctcccctcgt | gcgctctcct | gttccgaccc | 20880 |
| tgccgcttac | cggatacctg | tccgccttcc | tcccctcggg | aagcgtggcg | ctttctcaat | 20940 |
| gctcacgctg | taggtatctc | agttcgggtg | aggctgcttc | ctccaagctg | ggctgtgtgc | 21000 |
| acgaaccccc | cgttcagccc | gaccgctgcg | ccttatccgg | taactatcgt | cttgagtcca | 21060 |
| acccgtaag | acacgactta | tcgccactgg | cagcagccac | tggtaacagg | attagcagag | 21120 |
| cgaggtatgt | aggcggtgct | acagagttct | tgaagtgggtg | gcctaactac | ggctacacta | 21180 |
| gaaggacagt | atlttggtatc | tgcgctctgc | tgaagccagt | taccttcgga | aaaagagttg | 21240 |
| gtagctcttg | atccggcaaa | caaaccaccg | ctggtagcgg | tggttttttt | gtttgcaagc | 21300 |
| agcagattac | gcgcagaaaa | aaaggatctc | aagaagatcc | tttgatcttt | tctacgggggt | 21360 |

| | | | | | | |
|----------------------|-------------|------------|-------------|-------------|-------------|-------|
| ctgacgctca | gtggaacgaa | aactcacggt | aagggatfff | ggtcatgaga | ttatcaaaaa | 21420 |
| ggatcttcac | ctagatcctt | ttaaattaaa | aatgaagfff | taaatacaate | taaagtatat | 21480 |
| atgagtaaac | ttggtctgac | agttaccaat | gcttaatcag | tgaggcacct | atctcagcga | 21540 |
| tctgtctatt | tegttcatec | atagttgect | gactccccgt | cgtgtagata | actacgatac | 21600 |
| gggagggctt | accatctgge | cccagtgetg | caatgatacc | gcgagaccca | cgctcaccgg | 21660 |
| ctccagatff | atcagcaata | aaccagccag | ccggaagggc | cgagcgcaga | agtggtcctg | 21720 |
| caactttate | cgcctccate | cagtctatta | attggtgccc | ggaagctaga | gtaagtagtt | 21780 |
| cgccagttaa | tagtttgccc | aacggtgttg | ccattgctac | aggcatcgtg | gtgtcacgct | 21840 |
| cgtcgtttgg | tatggettca | ttcagctccg | gttcccacg | atcaaggcga | gttacatgat | 21900 |
| ccccatgff | gtgcaaaaaa | gcggttagct | ccttcgggtc | tccgatcgtt | gtcagaagta | 21960 |
| agttggccgc | agtgttatac | ctcatggtta | tggcagcact | gcataattct | cttactgtca | 22020 |
| tgccatccgt | aagatgctff | tctgtgactg | gtgagtactc | aaccaagtca | ttctgagaat | 22080 |
| agtgtatgcg | gcgaccgagt | tgctcttgcc | cggcgtcaat | acgggataat | accgcgccac | 22140 |
| atagcagaac | tttaaaagtg | ctcatcattg | gaaaacgttc | ttcggggcga | aaactctcaa | 22200 |
| ggatcttacc | gctgttgaga | tccagttcga | tgtaacccac | tcgtgcaccc | aactgatctt | 22260 |
| cagcatctff | tactttcacc | agcgtttctg | ggtgagcaaa | aacaggaagg | caaaatgccg | 22320 |
| caaaaaaggg | aataagggcg | acacggaaat | gttgaatact | catactcttc | ctttttcaat | 22380 |
| attattgaag | catttatcag | ggttattgtc | tcatgagcgg | atacatatff | gaatgtatff | 22440 |
| agaaaaataa | acaaataggg | gttccgcgca | catttccccg | aaaagtgccca | cctgacgtct | 22500 |
| aagaaacat | tattatcatg | acattaacct | ataaaaaatag | gcgtatcacg | aggccctffc | 22560 |
| gtc | | | | | | 22563 |
| <210>105 | | | | | | |
| <211>8350 | | | | | | |
| <212>DNA | | | | | | |
| <213>人工 | | | | | | |
| <220> | | | | | | |
| <223>质粒 psyx20-lacIq | | | | | | |
| <400>105 | | | | | | |
| cagctggcgt | aatagcgaag | aggccccgac | cgatcgccct | tcccacacagt | tgcgcagcct | 60 |
| gaatggcgaa | tgggacgcgc | cctgtagcgg | cgcatthaagc | gcggcgggtg | tgggtggttac | 120 |
| gcgcagcgtg | accgctacac | ttgccagcgc | cctagcgcgc | gctcctttcg | ctttcttccc | 180 |
| ttcctttctc | gccacgttcg | ccggttttcc | ccgtcaagct | ctaaatcggg | ggctccctff | 240 |
| agggttccga | tttagtgctt | tacggcacct | cgaccccaaa | aaacttgatt | agggtgatgg | 300 |
| ttcacgtagt | gggcatcgc | cctgatagac | ggtttttcgc | cctttgacgt | tggagtccac | 360 |
| gttctttaat | agtggactct | tgttccaaac | tggaacaaca | ctcaacccta | tctcgggteta | 420 |
| ttcttttgat | ttataagggga | ttttgccgat | ttcggcctat | tggttaaaaa | atgagctgat | 480 |
| ttaacaaaaa | tttaacgcga | attttaacaa | aattcgaccg | atgcccttga | gagccttcaa | 540 |
| cccagtcage | tccttccggt | gggcgcgggg | catgactatc | gtcgcgcgac | ttatgactgt | 600 |
| cttctttate | atgcaactcg | taggacaggt | gccggcagcg | ctctgggtca | ttttcggcga | 660 |

| | | | | | | |
|-------------|------------|------------|-------------|------------|-------------|------|
| ggaccgcttt | cgctggagcg | cgacgatgat | eggcctgtcg | cttgcggtat | tcggaatcct | 720 |
| gcacgccectc | gctcaagcct | tegteactgg | tcccgccacc | aaacgtttcg | gcgagaagca | 780 |
| ggccattate | gccggcatgg | eggccgacgc | gctgggctac | gtcttgctgg | cgttcgcgac | 840 |
| gcgaggettg | atggccttec | ccattatgat | tettctcgct | tccggcgcca | tcgggatgcc | 900 |
| cgcgttgcag | gccatgctgt | ccaggcaggt | agatgacgac | catcaggac | agcttcaagg | 960 |
| atcgctcgcg | gctcttacca | gcctaacttc | gatcattgga | ccgctgatcg | tcacggcgat | 1020 |
| ttatgccgcc | tcggcgagca | catggaacgg | gttggcatgg | attgtaggcg | ccgccctata | 1080 |
| ccttgtctgc | ctccccgcgt | tgcgtcgcg | tgcattggagc | cgggccacct | cgacctgaat | 1140 |
| ggaagccggc | ggcacctcgc | taacggatc | accactccgc | agaccgcca | taaacgccc | 1200 |
| tgagaagccc | gtgacgggct | tttcttgtat | tatgggtagt | ttccttgcat | gaatccataa | 1260 |
| aaggcgcctg | tagtgccatt | tacccccatt | cactgccaga | gccgtgagcg | cagcgaactg | 1320 |
| aatgtcacga | aaaagacagc | gactcaggtg | cctgatggtc | ggagacaaaa | ggaatattca | 1380 |
| gcgatttgcc | cgagcttgcg | agggtgctac | ttaagccttt | agggttttaa | ggtctgtttt | 1440 |
| gtagaggagc | aaacagcggt | tgcgacatcc | ttttgtaata | ctgcggaact | gactaaagta | 1500 |
| gtgagttata | cacagggtcg | ggatctatc | tttttatctt | tttttattct | ttctttatc | 1560 |
| tataaattat | aaccacttga | atataaaca | aaaaaacaca | caaaggtcta | gcggaattta | 1620 |
| cagagggtct | agcagaattt | acaagtttcc | cagcaaaggt | ctagcagaat | ttacagatac | 1680 |
| ccacaactca | aaggaaaagg | actagtaatt | atcattgact | agcccatctc | aattggtata | 1740 |
| gtgattaana | tcacctagac | caattgagat | gtatgtctga | attagttggt | ttcaaaagca | 1800 |
| atgaactagc | gattagtcgc | tatgacttaa | cggagcatga | aaccaagcta | attttatgct | 1860 |
| gtgtggcact | actcaacccc | acgattgaaa | accctacaag | gaaagaacgg | acggtatcgt | 1920 |
| tcacttataa | ccaatacgct | cagatgatga | acatcagtag | ggaaaatgct | tatggtgtat | 1980 |
| tagctaaagc | aaccagagag | ctgatgacga | gaactgtgga | aatcaggaat | cctttgggta | 2040 |
| aaggctttga | gattttccag | tggacaaact | atgccaagtt | ctcaagcgaa | aaattagaat | 2100 |
| tagtttttag | tgaagagata | ttgccttacc | ttttccagtt | aaaaaaattc | ataaaaatata | 2160 |
| atctggaaca | tgtaagtctt | tttgaaaaca | aatactctat | gaggatttat | gagtggttat | 2220 |
| taaaagaact | aacacaaaag | aaaactcaca | aggcaaata | agagattagc | cttgatgaat | 2280 |
| ttaagtccat | gttaatgctt | gaaaataact | accatgagtt | taaaaggctt | aaccaatggg | 2340 |
| ttttgaaacc | aataagtaaa | gatttaaac | cttacagca | tatgaaattg | gtggttgata | 2400 |
| agcgaggccg | cccgactgat | acgttgattt | tccaagttga | actagataga | caaatggatc | 2460 |
| tcgtaaccga | acttgagaac | aaccagataa | aatgaatgg | tgacaaaata | ccaacaacca | 2520 |
| ttacatcaga | ttcctacct | cataacggac | taagaaaaac | actacacgat | gctttaactg | 2580 |
| caaaaattca | gctcaccagt | tttgaggcaa | aatttttgag | tgacatgcaa | agtaagtatg | 2640 |
| atctcaatgg | ttcgttctca | tggctcaccg | aaaaacaacg | aaccacacta | gagaacatac | 2700 |
| tggctaaata | cggaaggatc | tgaggttctt | atggctcttg | tatctatcag | tgaagcatca | 2760 |
| agactaaca | acaaaagtag | aacaactggt | caccgttaca | tatcaaaggg | aaaactgtcc | 2820 |
| atatgcacag | atgaaaacgg | tgtaaaaaag | atagatacat | cagagctttt | acgagttttt | 2880 |
| ggtgcattca | aagctgttca | ccatgaacag | atcgacaatg | taacagatga | acagcatgta | 2940 |
| acacctata | gaacaggtga | aaccagtaaa | acaaagcaac | tagaacatga | aattgaacac | 3000 |

| | | | | | | |
|-------------|------------|-------------|-------------|-------------|-------------|------|
| ctgagacaac | ttgttacage | tcaacagtca | cacatagaca | gcctgaaaca | ggcgatgctg | 3060 |
| cttategaat | caaagctgcc | gacaacacgg | gagccagtga | cgctcccgt | ggggaaaaaa | 3120 |
| tcatggcaat | tctggaagaa | atagecgttt | cagccggcaa | accggctgaa | gccggatctg | 3180 |
| cgattctgat | aacaaactag | caacaccaga | acagcccgtt | tgcgggcage | aaaaccgta | 3240 |
| cttttgacg | ttccggcggt | ttttgtggc | gagtgggtgt | cgggcgggtc | gcgattattg | 3300 |
| aagcatttat | cagggttatt | gtctcatgag | eggatacata | tttgaatgta | tttagaaaaa | 3360 |
| taaacaata | ggggttccgc | gcacatttcc | ccgaaaagtg | ccacctgacg | tctaagaaac | 3420 |
| cattattatc | atgacattaa | cctataaaaa | taggcgtatc | acgaggccct | ttcgtcttca | 3480 |
| agaattcgcg | cgcaaggcc | aagcggcatg | catttacgtt | gacaccatcg | aatggcgcaa | 3540 |
| aacctttcgc | ggtatggcat | gatagcgccc | ggaagagagt | caattcaggg | tggatgaatgt | 3600 |
| gaaaccagta | acgttatacg | atgtcgcaga | gtatgccggt | gtctcttate | agaccgttcc | 3660 |
| ccgcgtgggtg | aaccaggcca | gccacgttcc | tgcgaaaacg | cgggaaaaag | tggaaagcggc | 3720 |
| gatggcggag | ctgaattaca | ttcccaaccg | cgtggcacia | caactggcgg | gcaaacagtc | 3780 |
| gttgctgatt | ggcgttgcca | cctccagtct | ggccctgcac | gcgccgtcgc | aaattgtcgc | 3840 |
| ggcgattaaa | tctcgcgccc | atcaactggg | tgccagcgtg | gtgggtgtcga | tggtagaacg | 3900 |
| aagcggcgtc | gaagcctgta | aagcggcggt | gcacaatctt | ctcgcgcaac | gcgtcagtgg | 3960 |
| gctgatcatt | aactatccgc | tggatgacca | ggatgccatt | gctgtggaag | ctgcctgcac | 4020 |
| taatgttccg | gcgttatttc | ttgatgtctc | tgaccagaca | cccatcaaca | gtattatttt | 4080 |
| ctccatgaa | gacggtacgc | gactgggcgt | ggagcatctg | gtcgcattgg | gtcaccagca | 4140 |
| aatcgcgctg | ttagegggcc | cattaagtcc | tgtctcggcg | cgtctgcgtc | tggctggctg | 4200 |
| gcataaatat | ctcactcgca | atcaaattca | gccgatagcg | gaacgggaag | gcgactggag | 4260 |
| tgccatgtcc | ggttttcaac | aaaccatgca | aatgctgaat | gagggcatcg | ttcccactgc | 4320 |
| gatgctgggt | gccaacgate | agatggcgct | gggcgcaatg | cgcgccatta | ccgagtccgg | 4380 |
| gctgcgcgtt | ggtgcggata | tctcggtagt | gggatacgac | gataccgaag | acagctcatg | 4440 |
| ttatatcccg | ccgtcaacca | ccatcaaaca | ggattttcgc | ctgctggggc | aaaccagcgt | 4500 |
| ggaccgcttg | ctgcaactct | ctcagggccca | ggcgggtgaag | ggcaatcagc | tgttgcccgt | 4560 |
| ctcactgggtg | aaaagaaaaa | ccaccctggc | gcccataacg | caaaccgcct | ctccccgcgc | 4620 |
| gttggccgat | tcattaatgc | agctggcacg | acaggtttcc | cgactggaaa | gcgggcagtg | 4680 |
| agcgaacgc | aattaatgtg | agttagctca | ctcattagcc | gaattctcat | gtttgacagc | 4740 |
| ttatcatcga | taagctttaa | tgcggtagtt | tatcacagtt | aaattgctaa | cgcagtcagg | 4800 |
| caccgtgtat | gaaatctaac | aatgcgctca | tcgtcatcct | cggcaccgtc | accctggatg | 4860 |
| ctgtaggcat | aggcttggtt | atgccggtac | tgccgggcct | cttgccggat | atcgtccatt | 4920 |
| ccgacagcat | cgccagtcac | tatggcgtgc | tgctagcgtc | atatgcgttg | atgcaatttc | 4980 |
| tatgcgcacc | cgttctcgga | gcactgtccg | accgctttgg | ccgccgcca | gtcctgtctg | 5040 |
| cttcgctact | tggagccact | atcgactacg | cgatcatggc | gaccacacc | gtcctgtgga | 5100 |
| tcctctacgc | eggacgcate | gtggccggca | tcaccggcgc | cacaggtcgc | gttgctggcg | 5160 |
| cctatategc | cgacatcacc | gatggggaag | atcgggctcg | ccacttcggg | ctcatgagcg | 5220 |
| cttgtttcgg | cgtgggtatg | gtggcaggcc | ccgtggccgg | gggactgttg | ggcgccatct | 5280 |
| ccttgcatgc | accattcctt | gcggcggcgg | tgctcaacgg | cctcaaccta | ctactgggct | 5340 |

| | | | | | | |
|------------|-------------|------------|-------------|-------------|-------------|------|
| gcttcctaat | gcaggaatcg | cataagggag | agcgtcgacc | gatgcccttg | agagccttca | 5400 |
| accagtcag | ctccttccgg | tgggcgcggg | gcatgactat | cgtcgccgca | cttatgactg | 5460 |
| tcttctttat | catgcaactc | gtaggacagg | tgccggcagc | gctctgggtc | attttcggcg | 5520 |
| aggaccgctt | tcgctggagc | gcgacgatga | teggcctgtc | gcttgccggt | ttcggaaatct | 5580 |
| tgcacgcctt | cgctcaagcc | ttcgteactg | gtcccgccac | caaacgtttc | ggcgagaagc | 5640 |
| aggccattat | cgccggcatg | gcgcccgacg | cgctgggcta | cgctcttgctg | gcgttcgca | 5700 |
| cgcgaggctg | gatggccttc | cccattatga | ttcttctcgc | ttccggcggc | atcgggatgc | 5760 |
| ccgcgttgca | ggccatgctg | tccaggcagg | tagatgacga | ccatcagggg | cagcttcaag | 5820 |
| gatcgctcgc | ggctcttacc | agcctaactt | cgatcattgg | accgctgate | gtcacggcga | 5880 |
| tttatgccgc | ctcggcgagc | acatggaacg | ggttggcatg | gattgtaggc | gccgccctat | 5940 |
| accttgtctg | cctccccgcg | ttgcgtcgcg | gtgcatggag | ccgggccacc | tcgacctgaa | 6000 |
| tggaagccgg | cggcacctcg | ctaacggatt | caccactcca | agaattggag | ccaatcaatt | 6060 |
| cttgcggaga | actgtgaatg | cgcaaacca | cccttggcag | aacatatcca | tcgctccgc | 6120 |
| catctccagc | agccgcacgc | ggcgcatctc | gggcagcggt | gggtcctggc | cacgggtgcg | 6180 |
| catgatcgtg | ctcctgtcgt | tgaggaccgg | gctaggctgg | cggggttgcc | ttactgggta | 6240 |
| gcagaatgaa | tcaccgatac | gcgagcgaac | gtgaagcgac | tgctgctgca | aaacgtctgc | 6300 |
| gacctgagca | acaacatgaa | tggtcttcgg | tttccgtggt | tcgtaaagtc | tggaaacgcg | 6360 |
| gaagtcagcg | ccctgcacca | ttatgttccg | gatctgcate | gcaggatgct | gctggctacc | 6420 |
| ctgtggaaca | cctacatctg | tattaacgaa | gcgctggcat | tgaccctgag | tgatttttct | 6480 |
| ctggctccgc | cgcatecata | ccgccagttg | tttaccctca | caacgttcca | gtaaccgggc | 6540 |
| atgttcatca | tcagtaacce | gtatcgtgag | catcctctct | cgtttcatcg | gtatcattac | 6600 |
| ccccatgaac | agaaateccc | cttacacgga | ggcatcagtg | accaaacagg | aaaaaacgcg | 6660 |
| ccttaacatg | gcccgttta | tcagaagcca | gacattaacg | cttctggaga | aactcaacga | 6720 |
| gctggacgcg | gatgaacagg | cagacatctg | tgaatcgctt | cacgaccacg | ctgatgagct | 6780 |
| ttaccgcagc | gcgcagggtc | agcctgaata | cgcgtttaat | gaccagcaca | gtcgtgatgg | 6840 |
| caaggtcaga | atagcgctga | ggtctgcctc | gtgaagaagg | tgttgctgac | tcataccagg | 6900 |
| cctgaatcgc | cccatcatcc | agccagaaag | tgagggagcc | acggttgatg | agagctttgt | 6960 |
| tgtaggtgga | ccagttgggtg | atthtgaact | tttgctttgc | cacggaacgg | tctgcgttgt | 7020 |
| cgggaagatg | cgtgatctga | tccttcaact | cagcaaaaagt | tcgatttatt | caacaaagcc | 7080 |
| acgttgtgtc | tcaaaatctc | tgatgttaca | ttgcacaaga | taaaaatata | tcatcatgaa | 7140 |
| caataaaact | gtctgcttac | ataaacagta | atacaagggg | tgttatgagc | catattcaac | 7200 |
| gggaaacgtc | ttgctcgagg | ccgcgattaa | attccaacat | ggatgctgat | ttatatgggt | 7260 |
| ataaatgggc | tcgcgataat | gtcgggcaat | caggtgcgac | aatctatcga | ttgtatggga | 7320 |
| agcccgatgc | gccagagttg | tttctgaaac | atggcaaagg | tagcgttgcc | aatgatgtta | 7380 |
| cagatgagat | ggtcagacta | aactggctga | cggaatttat | gcctcttccg | accatcaagc | 7440 |
| atthtatecg | tactcctgat | gatgcatggt | tactcaccac | tgcatcccc | gggaaaacag | 7500 |
| cattccaggt | attagaagaa | tatcctgatt | caggtgaaaa | tattgttgat | gcgctggcag | 7560 |
| tgthctcgcg | ccggttgcat | tcgattcctg | tttgtaattg | tcctthttaa | agcgatcgcg | 7620 |
| tatttctctt | cgctcagcgc | caatcacgaa | tgaataacgg | tttggttgat | gcgagtgatt | 7680 |

| | |
|--|------|
| ttgatgacga gcgtaatggc tggcctgttg aacaagtctg gaaagaaatg cataagcttt | 7740 |
| tgccattctc accggattca gtegteacte atggtgattt ctacttgat aaccttattt | 7800 |
| ttgacgaggg gaaattaata ggttgattg atggtggacg agtcggaatc gcagaccgat | 7860 |
| accaggatct tgccatecta tggaactgcc tgggtgagtt ttctcctca ttacagaaac | 7920 |
| ggctttttca aaaatatggt attgataate ctgatatgaa taaattgcag tttcatttga | 7980 |
| tgctcgatga gtttttctaa tcagaattgg ttaattgggt gtaacactgg cagagcatta | 8040 |
| cgctgacttg acgggacggc ggctttgttg aataaatcga acttttgctg agttgaagga | 8100 |
| tcagatcacg catcttcccg acaacgcaga cegtccctg gcaaagcaaa agttcaaaat | 8160 |
| caccaactgg tccacctaca acaaagetct catcaaccgt ggctccctca ctttctggct | 8220 |
| ggatgatggg gcgattcagg cctggatga gtcagcaaca ctttcttcac gaggcagacc | 8280 |
| tcagcgctat tctgacctg ccatcacgac tgtgctggtc attaaacgcg tattcaggct | 8340 |
| gacctgccc | 8350 |
| <210>106 | |
| <211>3190 | |
| <212>DNA | |
| <213> 人工 | |
| <220> | |
| <223>pAGR3 | |
| <400>106 | |
| aagtgaccaa acaggaaaa accgccctta acatggcccc ctttatcaga agccagacat | 60 |
| taacgcttct ggagaaactc aacgagctgg acgcggatga acaggcagac atctgtgaat | 120 |
| cgcttcacga ccacgctgat gagctttacc gcagctgcct cgcgcgtttc ggtgatgacg | 180 |
| gtgaaaacct ctgacacatg cagctcccgg agacggtcac agcttgtctg taagcggatg | 240 |
| ccgggagcag acaagcccgt cagggcgctg cagcgggtgt tggcgggtgt cggggcgcag | 300 |
| ccatgaccca gtcacgtagc gatagcggag tgtatactgg cttactatg cggcatcaga | 360 |
| gcagattgta ctgagagtgc accatatgcg gtgtgaaata ccgcacagat gcgtaaggag | 420 |
| aaaataccgc atcaggcgct ctcccgcttc ctgctcact gactcgtgc gctcggctct | 480 |
| tcggctgcgg cgagcggat cagctcactc aaaggcggta atacggttat ccacagaatc | 540 |
| aggggataac gcaggaaaga acatgtgagc aaaaggccag caaaaggcca ggaaccgtaa | 600 |
| aaaggccgcg ttgctggcgt tttccatag gctccgcccc cctgacgagc atcacaaaaa | 660 |
| tcgacgctca agtcagaggt ggcgaaaccg gacaggacta taaagatacc aggcgtttcc | 720 |
| ccctggaage tccctcgtgc gctctcctgt tccgaccctg ccgcttaccg gatacctgct | 780 |
| cgctttctc ctttcgggaa gcgtggcgct ttctcatagc tcacgctgta ggtatctcag | 840 |
| ttcgggtgtag gtcgttcgct ccaagctggg ctgtgtgcac gaaccccccg ttcagcccga | 900 |
| ccgctgcgcc ttatccggtg actatcgtct tgagtccaac ccggtaaagac acgacttate | 960 |
| gccactggca gcagccactg gtaacaggat tagcagagcg aggtatgtag gcggtgctac | 1020 |
| agagtctctg aagtggtagg ctaactacgg ctacactaga aggacagtat ttggtatctg | 1080 |
| cgctctgctg aagccagtta ctttcggaaa aagagttggt agctcttgat ccggcaaaaa | 1140 |
| aaccaccgct ggtagcggtg gttttttgt ttgcaagcag cagattacgc gcagaaaaaa | 1200 |

| | | | | | | |
|------------|-------------|-------------|------------|------------|------------|------|
| aggatctcaa | gaagatecctt | tgatcttttc | tacggggtct | gacgctcagt | ggaacgaaaa | 1260 |
| ctcacgttaa | gggatttttg | tcattgagatt | atcaaaaagg | atcttcacct | agatcctttt | 1320 |
| aaattaaaaa | tgaagtttta | aatcaateta | aagtatatat | gagtaaactt | ggtctgacag | 1380 |
| ttaccaatgc | ttaatcagtg | aggcacctat | ctcagcgate | tgtctatttc | gttcatccat | 1440 |
| agttgcctga | ctccccgtcg | tgtagataac | tacgatacgg | gagggtctac | catctggccc | 1500 |
| cagtgctgca | atgataccgc | gagacccacg | ctcaccggct | ccagatttat | cagcaataaa | 1560 |
| ccagccagcc | ggaagggccg | agcgcagaag | tggctctgca | actttatccg | cctccatcca | 1620 |
| gtctattaat | tgttgccggg | aagctagagt | aagtagttcg | ccagttaata | gtttgcgcaa | 1680 |
| cgttgttgcc | attgctgcag | gcategtggt | gtcacgctcg | tcgtttggta | tggcttcatt | 1740 |
| cagctccggt | tcccaacgat | caaggcgagt | tacatgatcc | cccatgttgt | gcaaaaaagc | 1800 |
| ggttagctcc | ttcggctctc | cgatcgttgt | cagaagtaag | ttggccgcag | tgttatcact | 1860 |
| catggttatg | gcagcactgc | ataattctct | tactgtcatg | ccatccgtaa | gatgcttttc | 1920 |
| tgtgactggt | gagtactcaa | ccaagtcatt | ctgagaatag | tgtatgcggc | gaccgagttg | 1980 |
| ctcttgcccg | gcgtcaacac | gggataatac | cgcgccacat | agcagaactt | taaaagtgct | 2040 |
| catcattgga | aaacgttctt | cggggcgaaa | actctcaagg | atcttaccgc | tgttgagatc | 2100 |
| cagttcgatg | taaccactc | gtgcacccaa | ctgatcttca | gcatecttta | ctttcaccag | 2160 |
| cgtttctggg | tgagcaaaaa | caggaaggca | aatgcccga | aaaaagggaa | taagggcgac | 2220 |
| acggaaatgt | tgaatactca | tactcttctt | tttcaatat | tattgaagca | tttatcaggg | 2280 |
| ttattgtctc | atgagcggat | acatatttga | atgtatttag | aaaaataaac | aataggggt | 2340 |
| tccgcgcaca | tttccccgaa | aagtgccacc | tgacgtctaa | gaaaccatta | ttatcatgac | 2400 |
| attaacctat | aaaaataggc | gtatcacgag | gccctttcgt | cttcaagaat | taattcccaa | 2460 |
| ttccaggcat | caaataaaa | gaaaggctca | gtcgaaagac | tgggcctttc | gttttatctg | 2520 |
| ttgtttgtcg | gtgaacgctc | tcttgagtag | gacaaatccg | ccgggagcgg | atttgaacgt | 2580 |
| tgcgaagcaa | cggcccggag | ggtggcgggg | aggacgcccc | ccataaactg | ccaggaatta | 2640 |
| attccaggca | tcaataaaaa | cgaaaggctc | agtcgaaaga | ctgggccttt | cgttttatct | 2700 |
| gttgtttgtc | ggtgaacgct | ctcctgagta | ggacaaatcc | gccgggagcg | gatttgaacg | 2760 |
| ttgcgaagca | acggcccggg | gggtggcggg | caggacgccc | gccataaact | gccaggaatt | 2820 |
| aattccaggc | atcaataaaa | acgaaaggct | cagtcgaaag | actgggcctt | tcgttttata | 2880 |
| tgtttgtttg | cggtgaacgc | tctcctgagt | aggacaaatc | cgccgggagc | ggatttgaac | 2940 |
| gttgcgaagc | aacggcccgg | agggtggcgg | gcaggacgcc | cgccataaac | tgccaggaat | 3000 |
| taattccagg | catcaataaa | aacgaaaggc | tcagtcgaaa | gactgggcct | ttcgttttat | 3060 |
| ctgtttgttg | tcggtgaacg | ctctcctgag | taggacaaat | ccgccgggag | cggatttgaa | 3120 |
| cgttgcgaag | caacggcccc | gagggtggcg | ggcaggacgc | ccgccataaa | ctgccaggaa | 3180 |
| ttggggatcg | | | | | | 3190 |

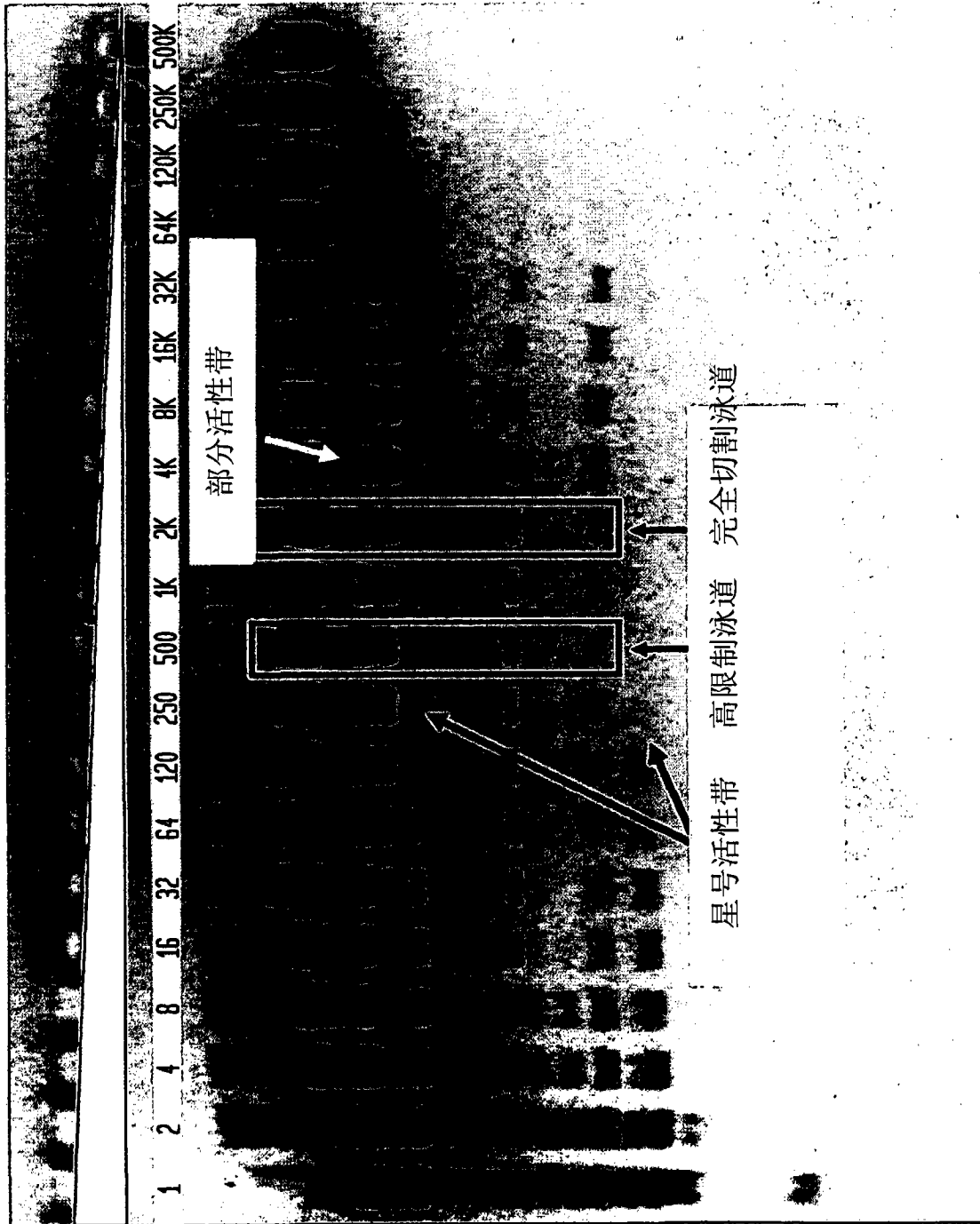


图 1

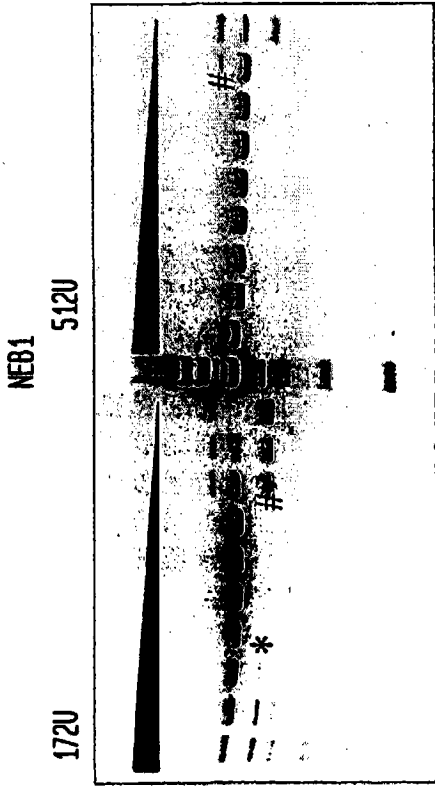


图 2A

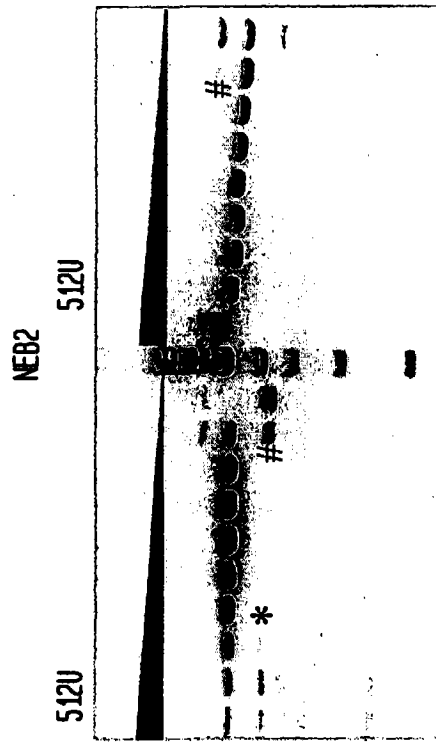


图 2B

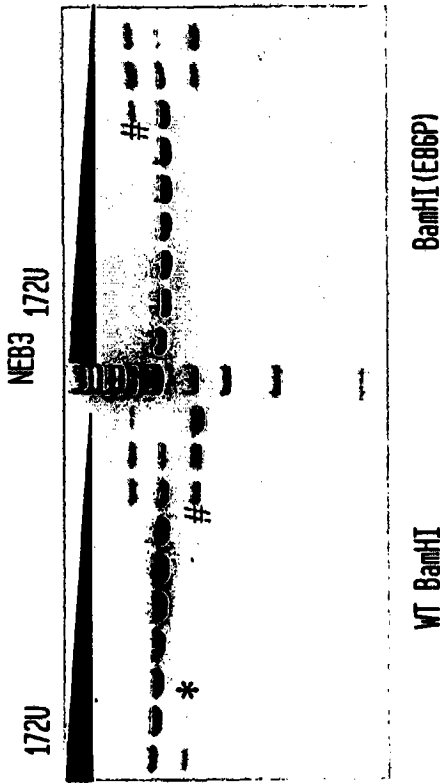


图 2C

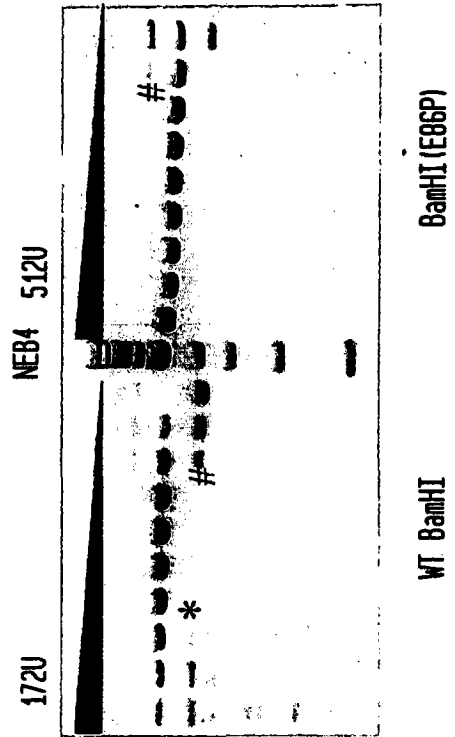


图 2D

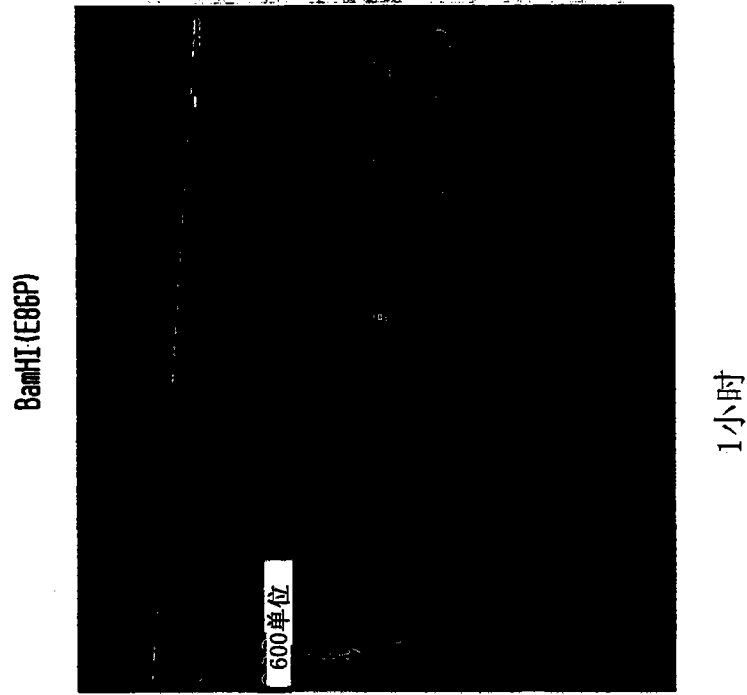


图 3A

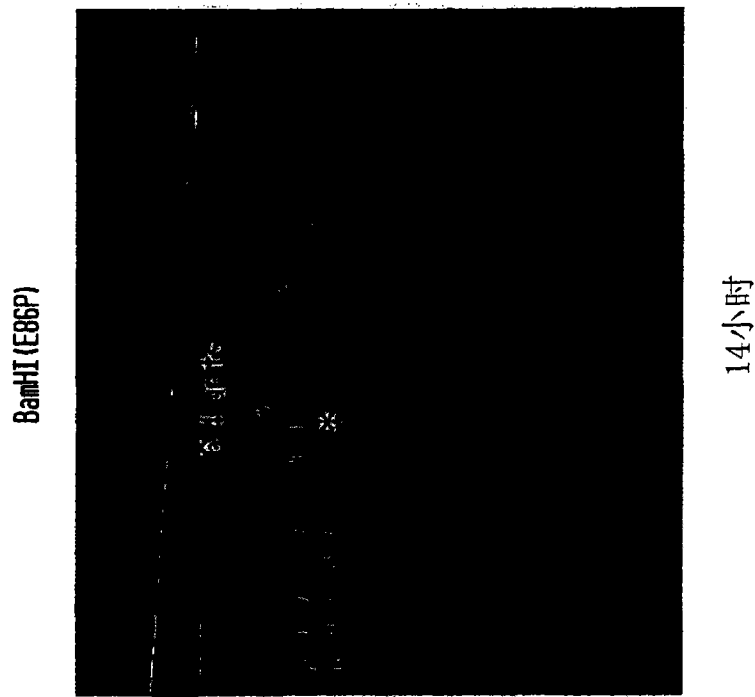


图 3B

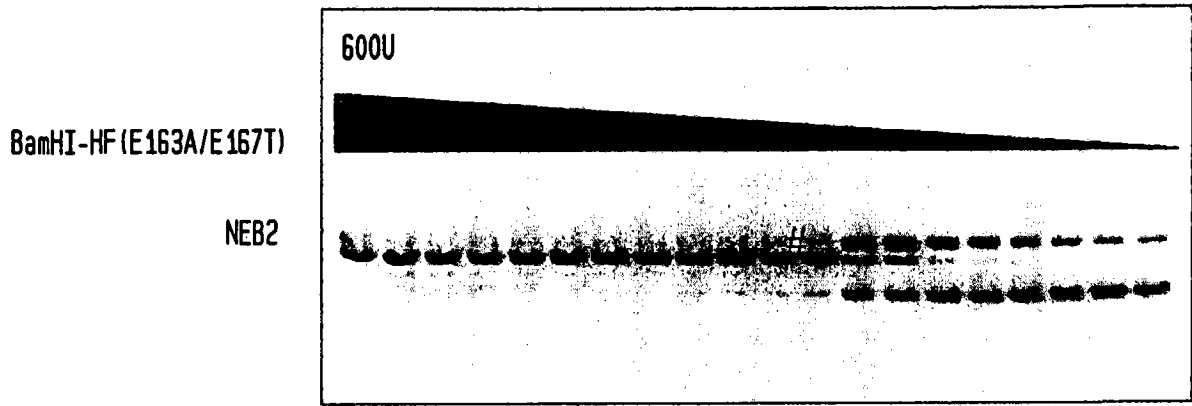


图 4A

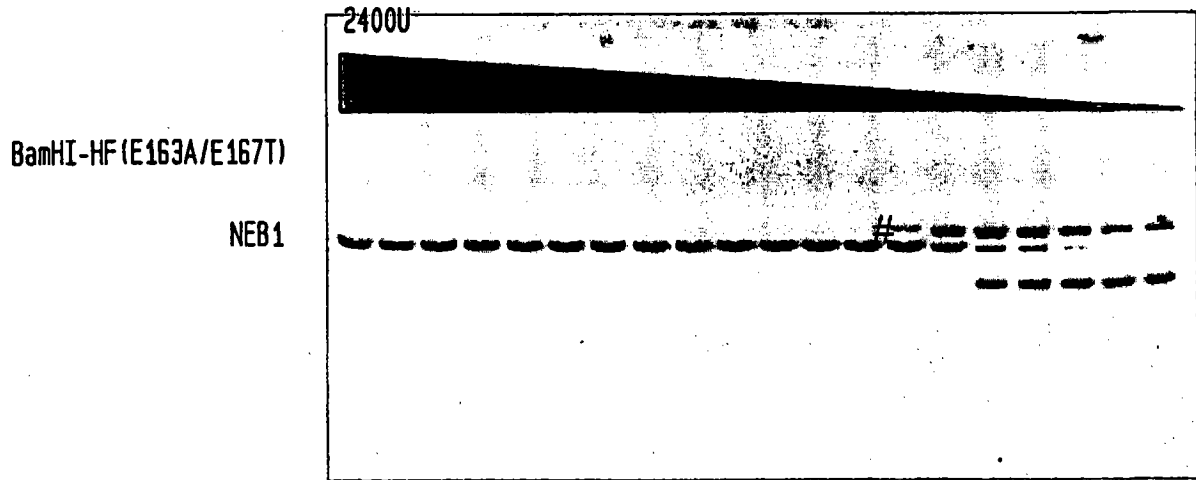
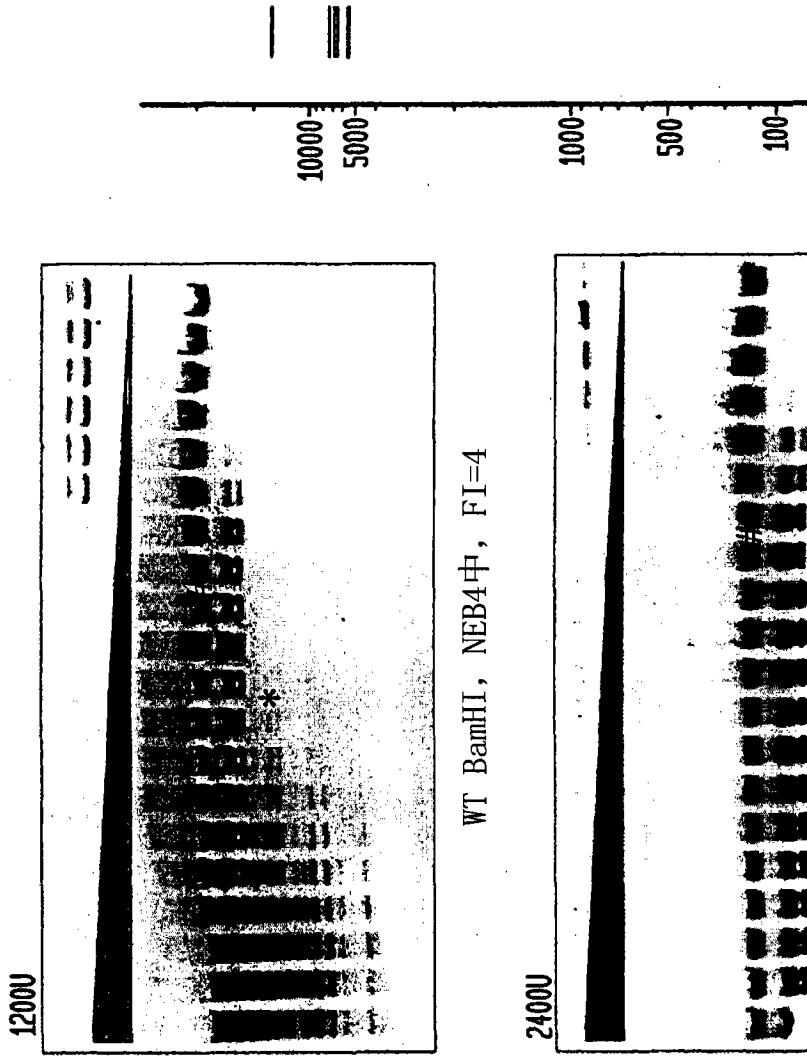
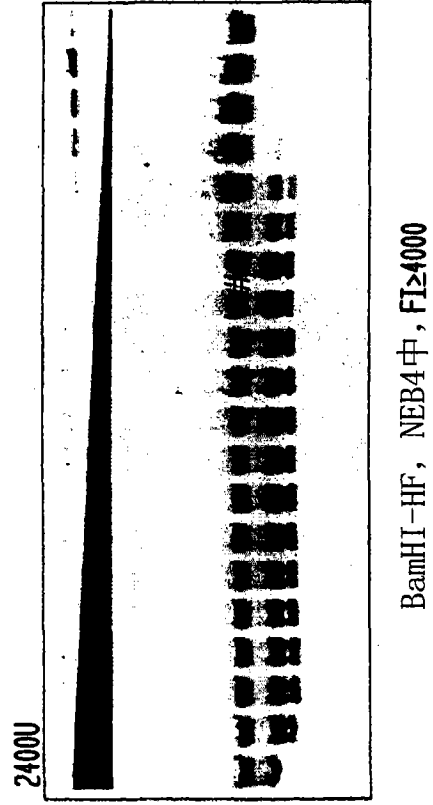


图 4B



BamHI-HF和
WT BamHI的比较

图5A



BamHI-HF和
WT BamHI的比较

图5B

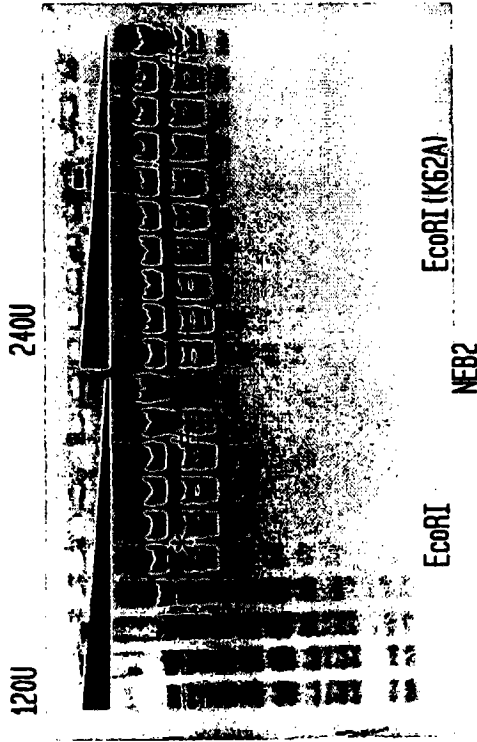


图 6A

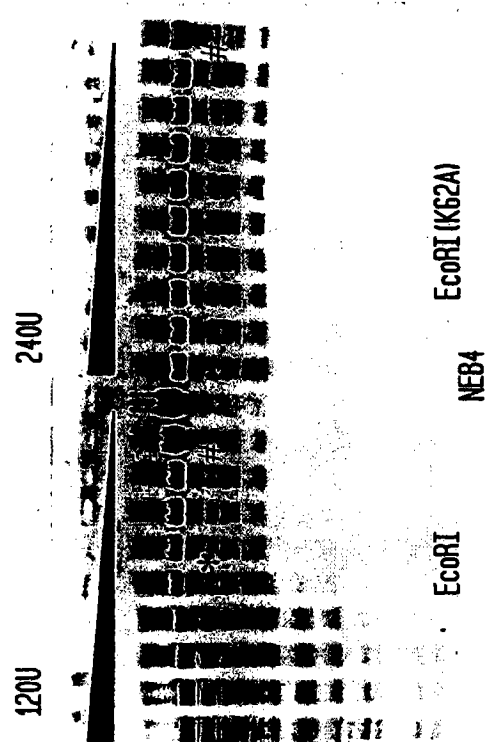


图 6B

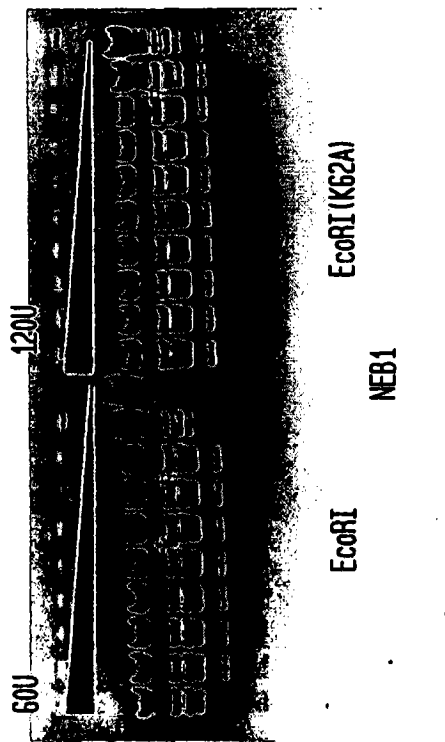


图 6C

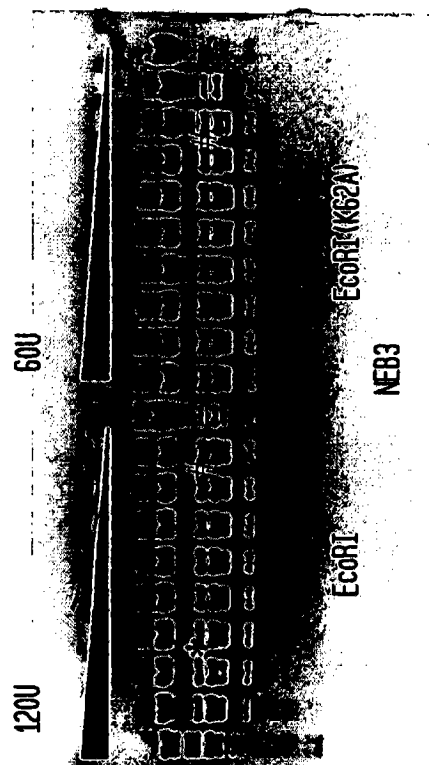


图 6D

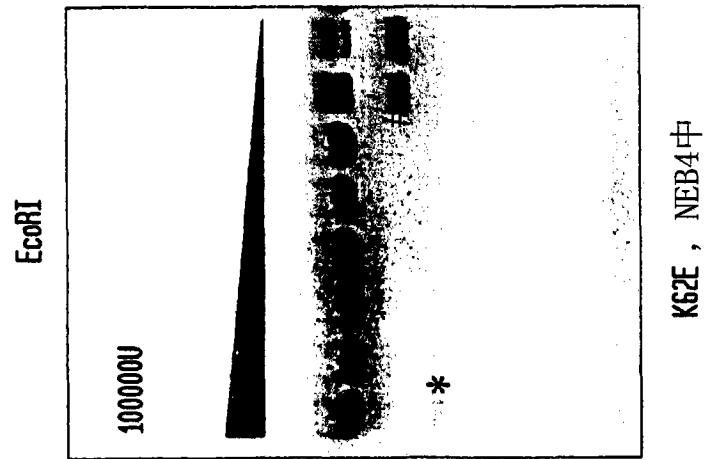


图 7A

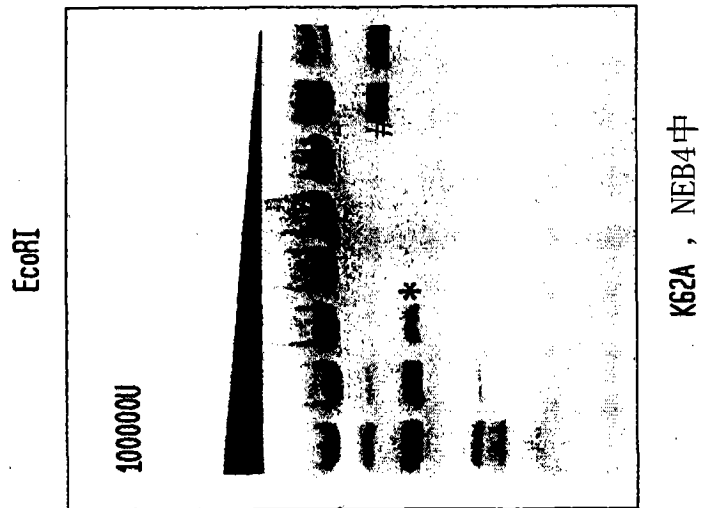


图 7B

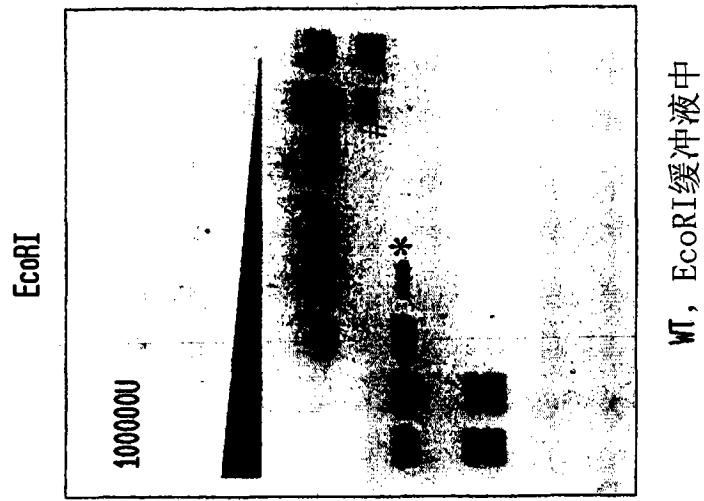
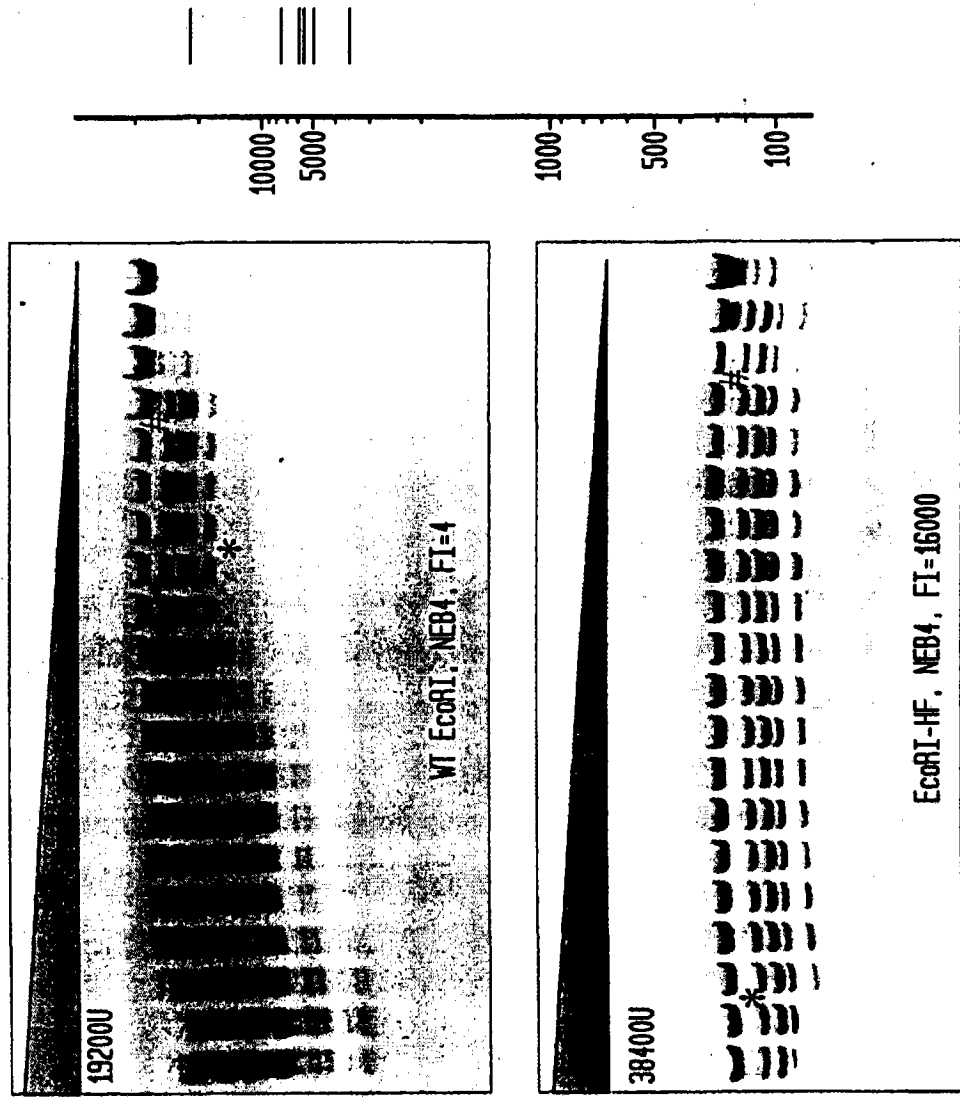


图 7C



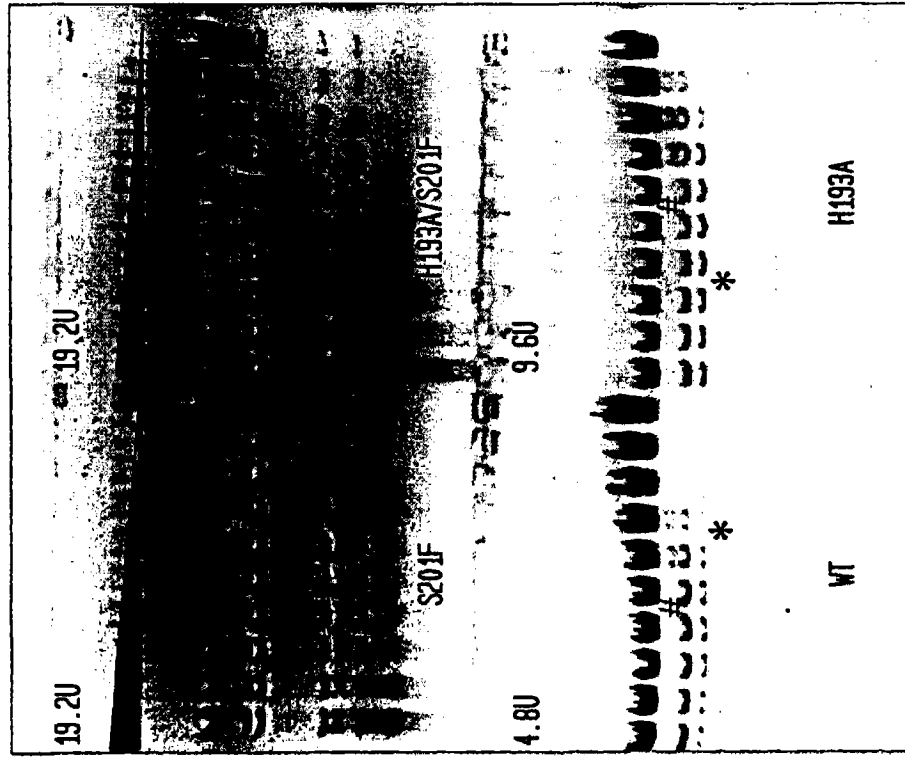
EcoRI-HF和
WT EcoRI的比较

图8A

EcoRI-HF和
WT EcoRI的比较

图8B

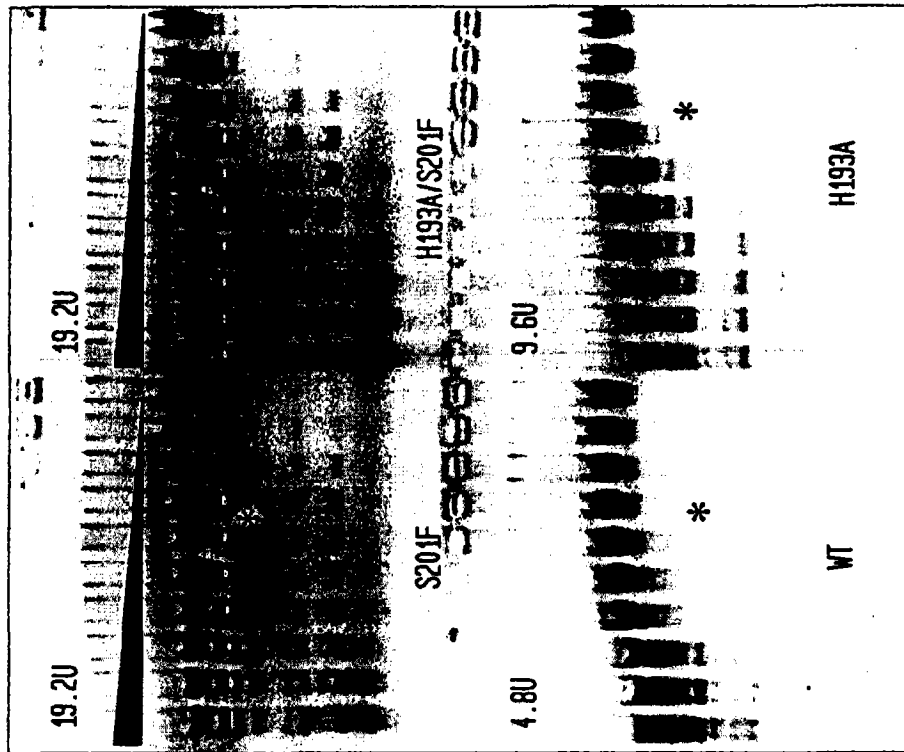
ScaI突变组合



NEB2, 5%甘油

图 9A

ScaI突变组合



NEB2, 37%甘油

图 9B

图10A

ScaI-HF和WT ScaI的ScaI比较

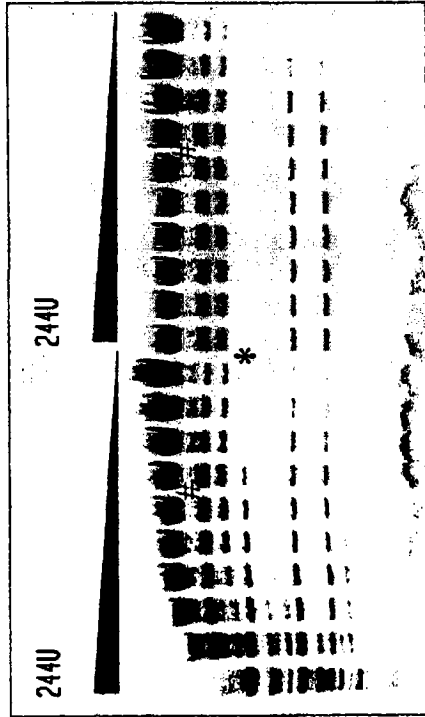


图10B

ScaI-HF和WT ScaI的ScaI比较

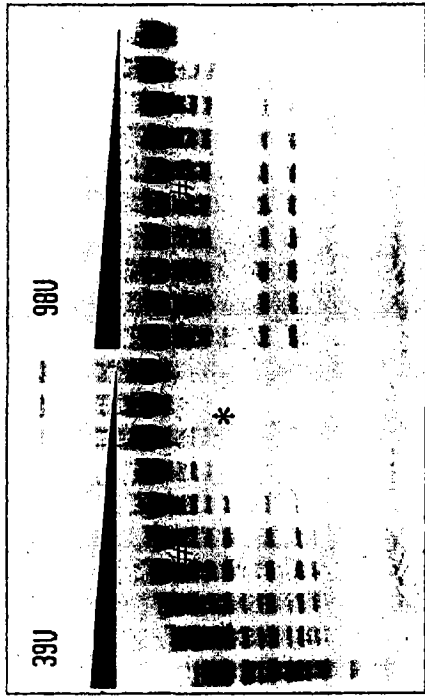


图10C

ScaI-HF和WT ScaI的ScaI比较

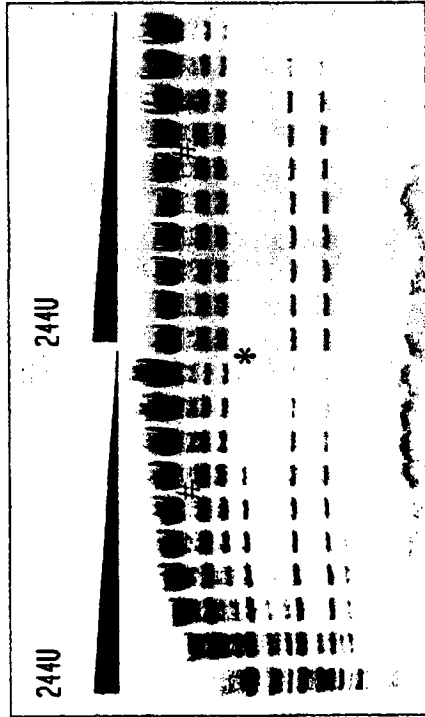
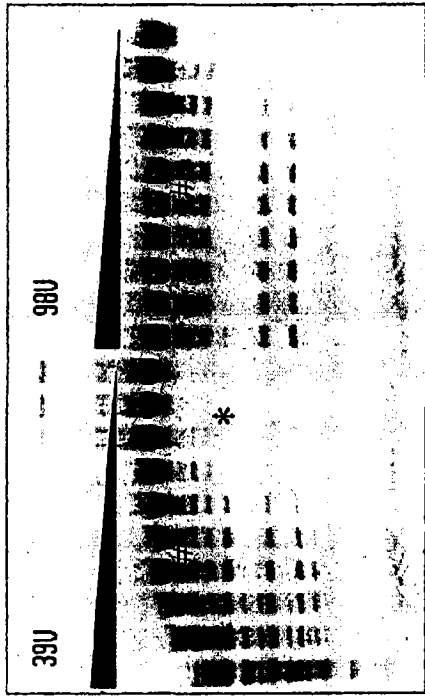


图10D

ScaI-HF和WT ScaI的ScaI比较



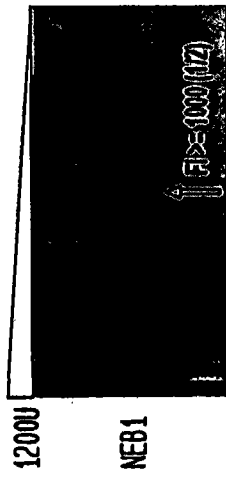


图 11A

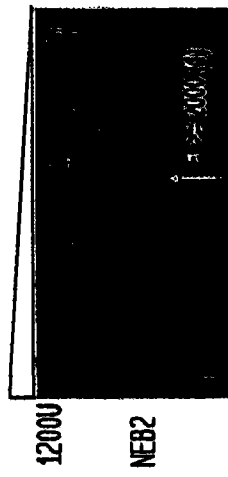


图 11B

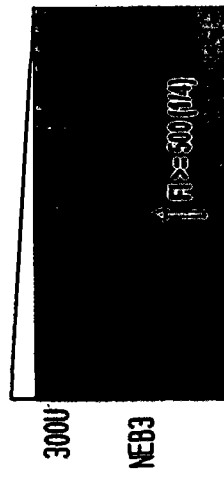


图 11C

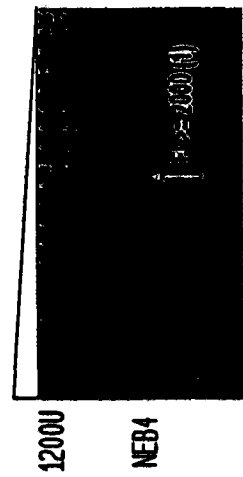


图 11D

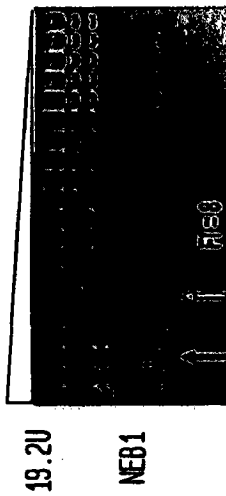
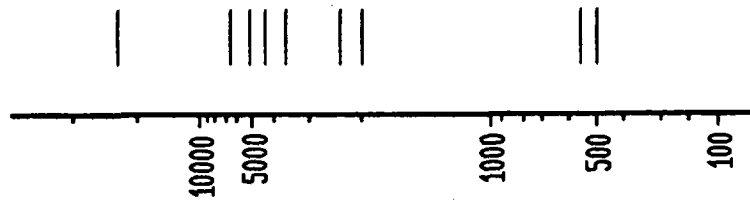


图 11E

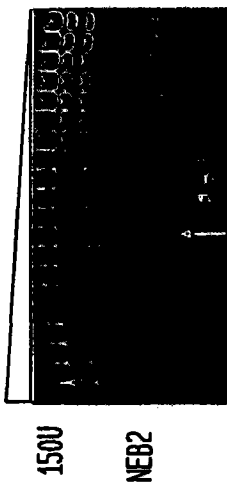


图 11F

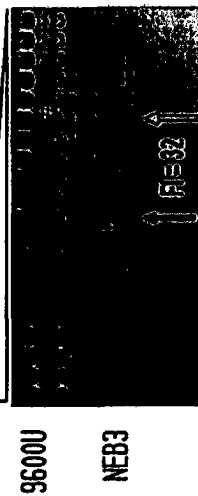


图 11G

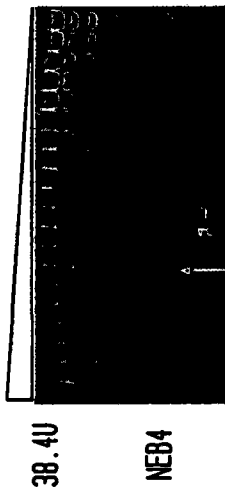


图 11H

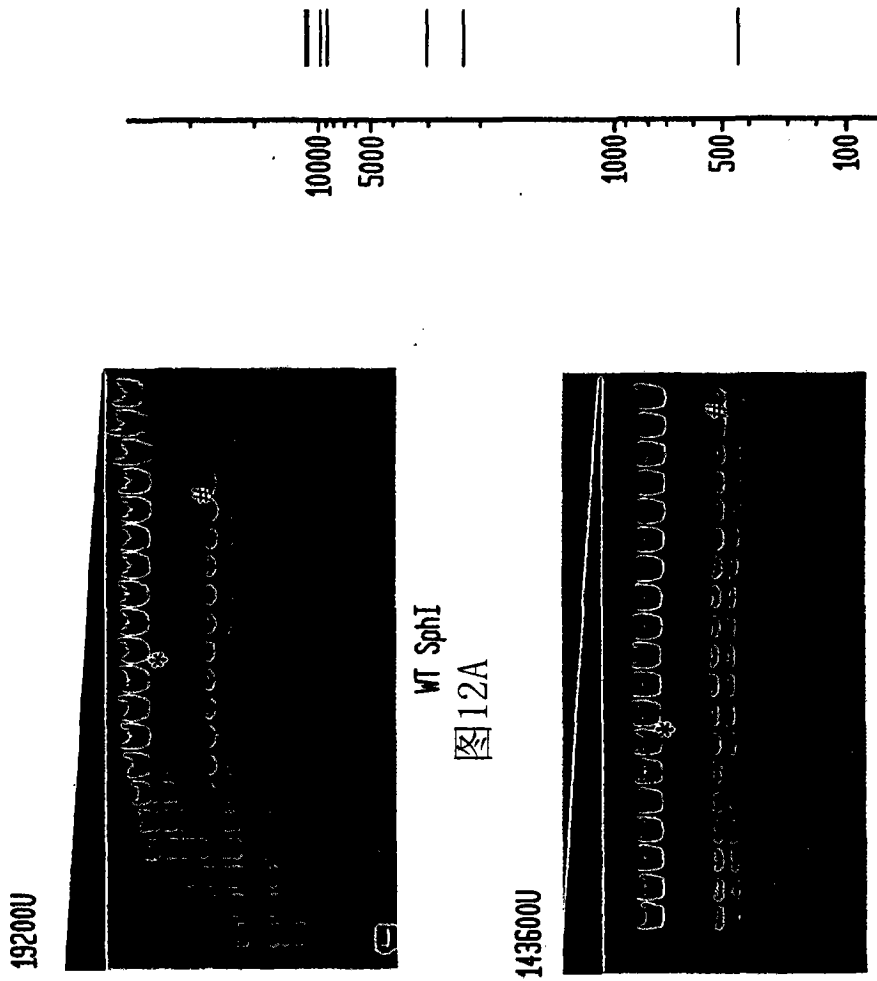


图12A

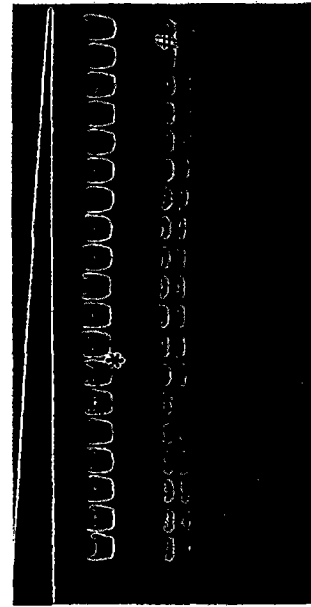


图12B

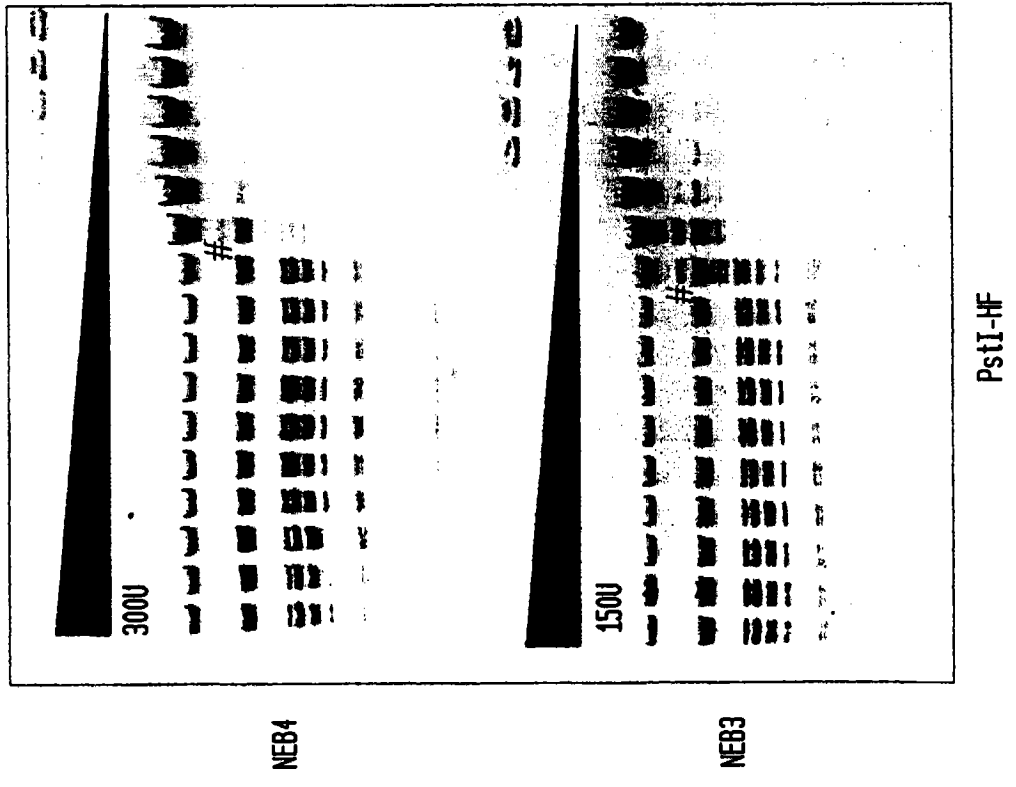


图 13A

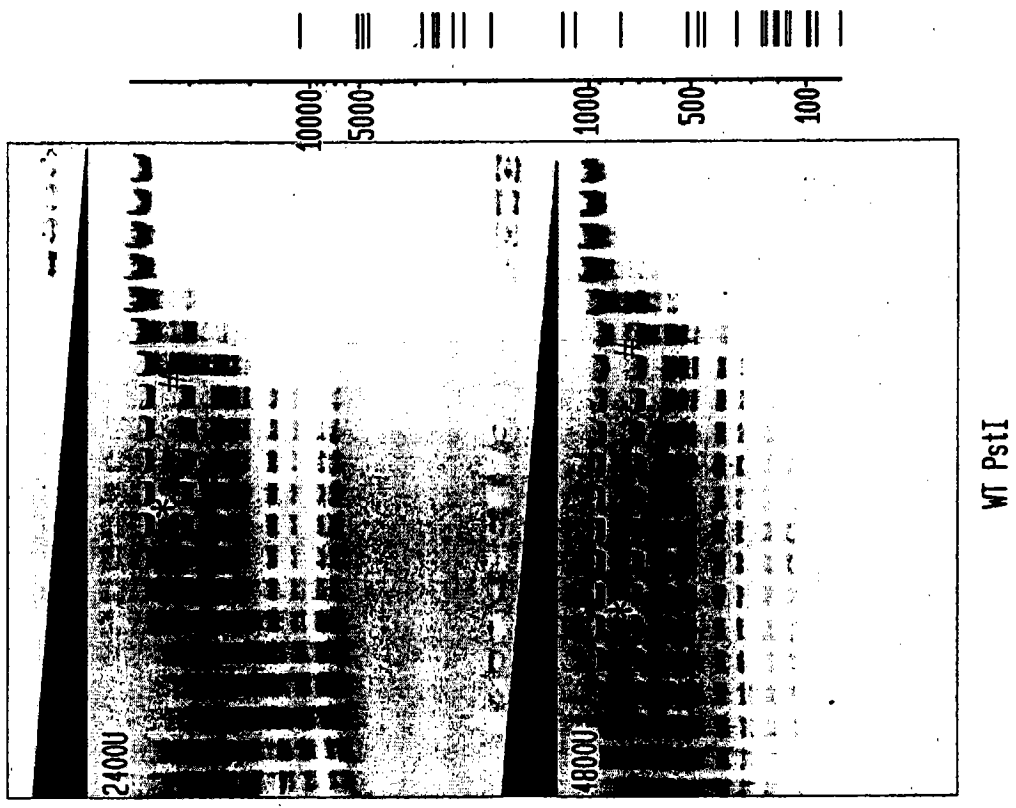


图 13B

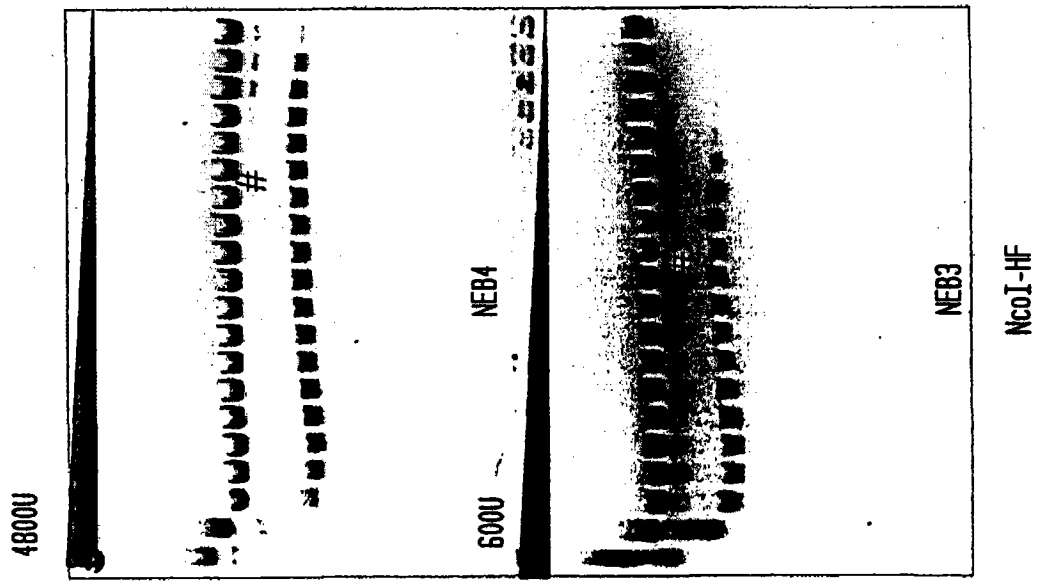


图 14A

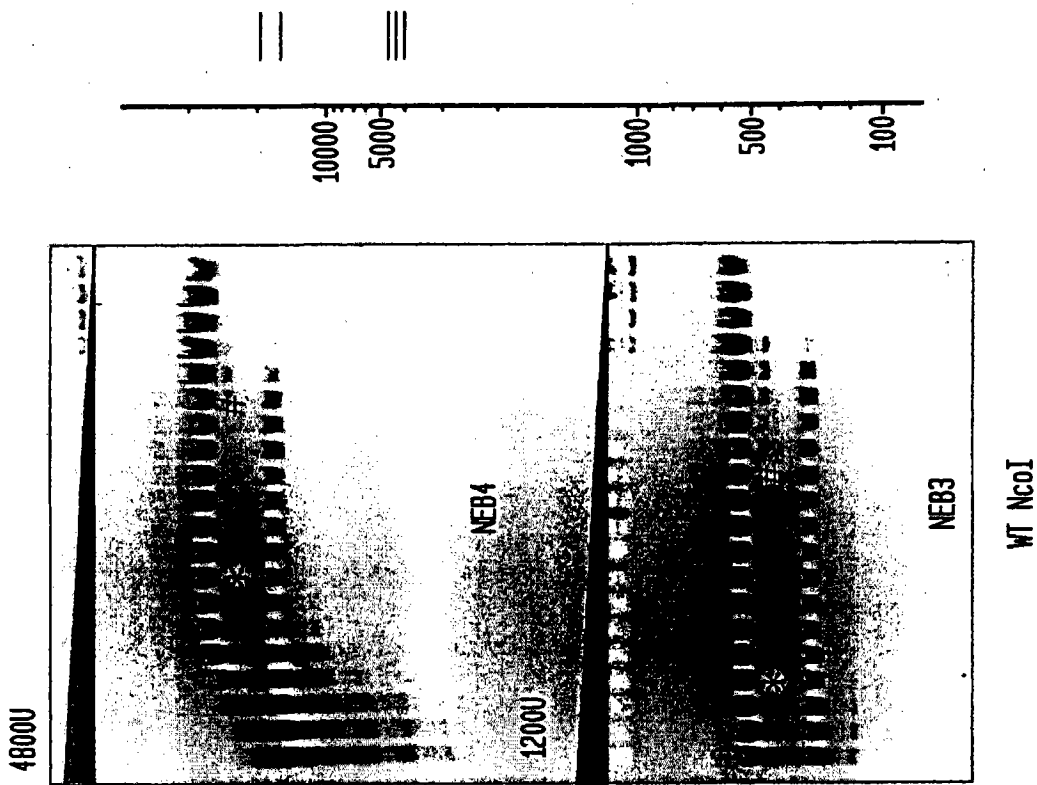


图 14B

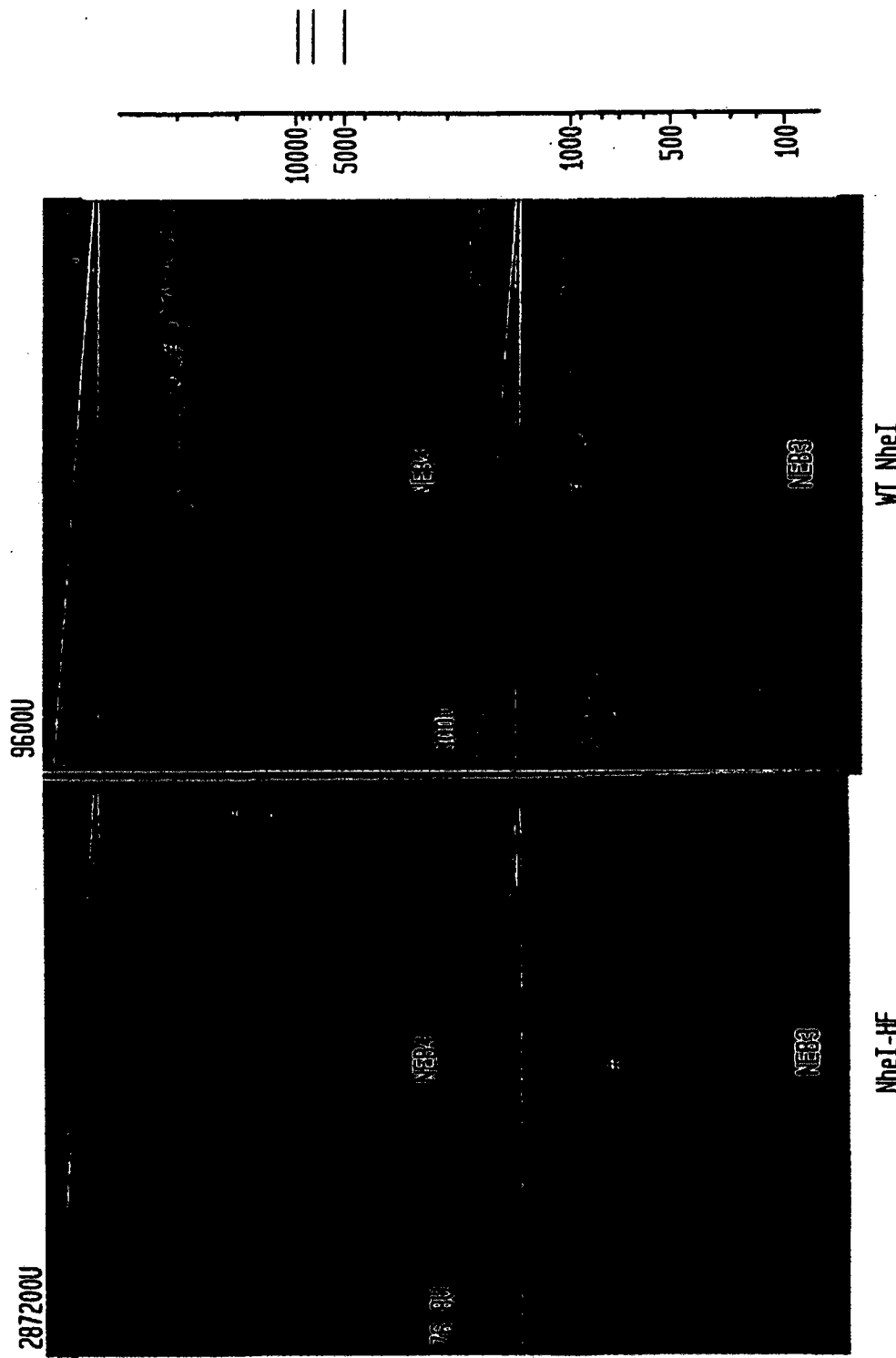


图15B

图15A

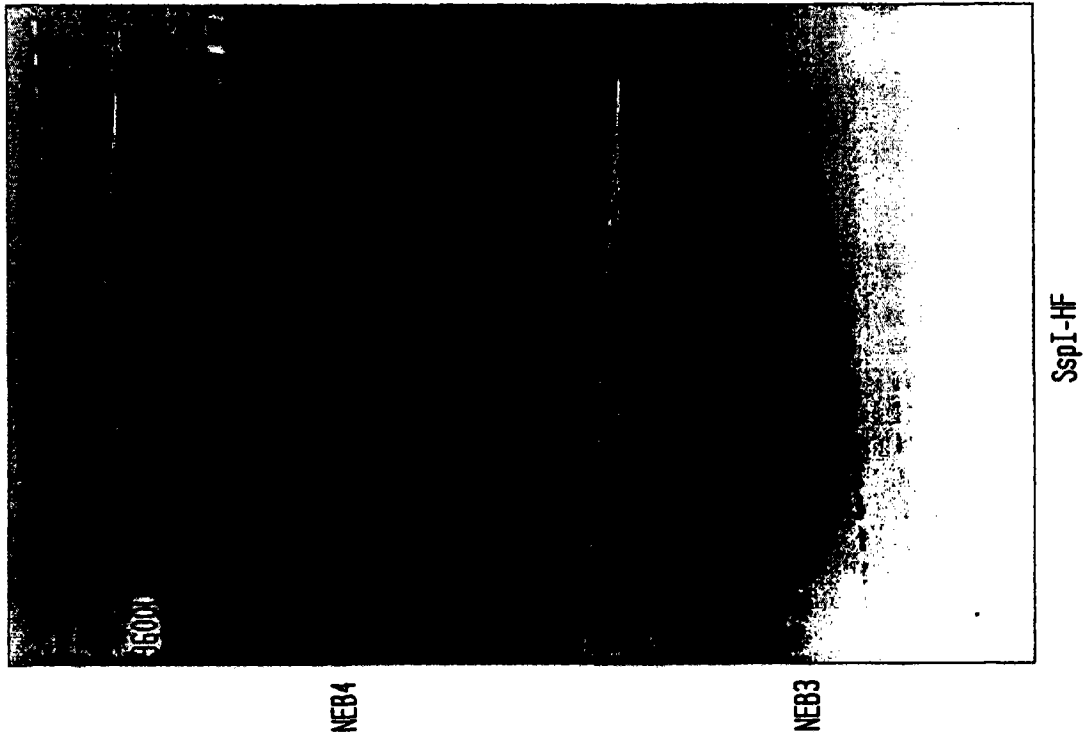


图 16A

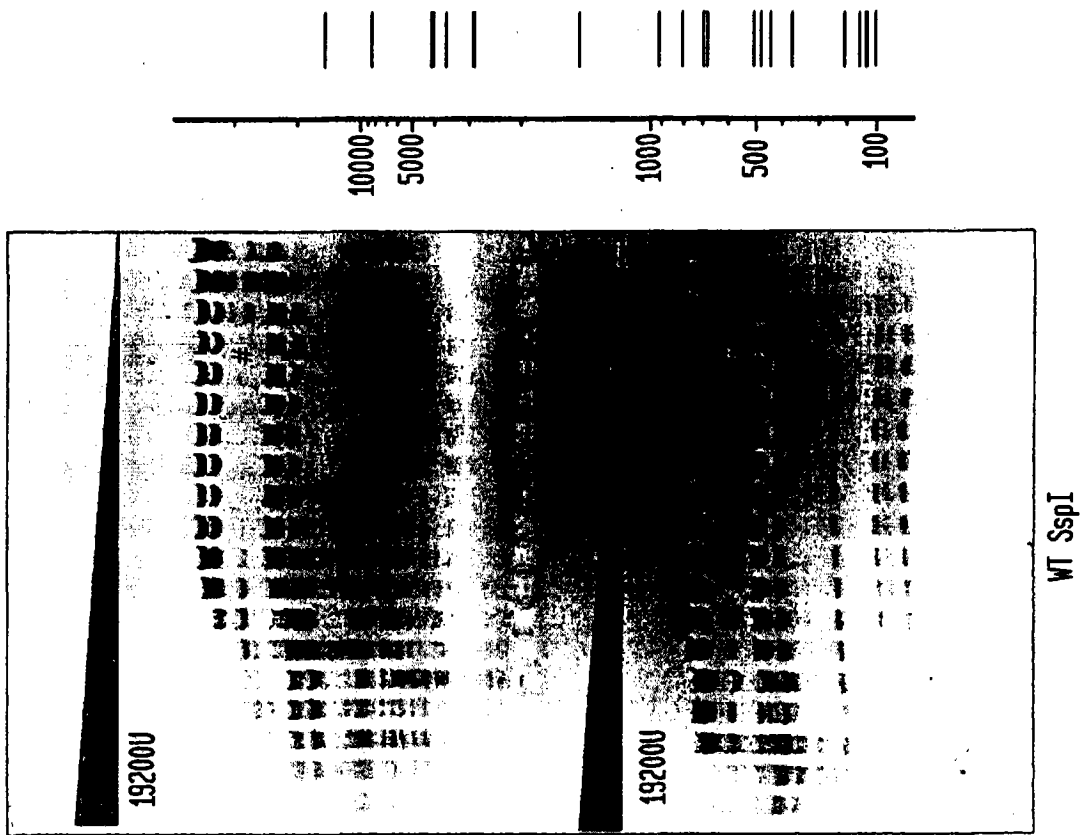


图 16B

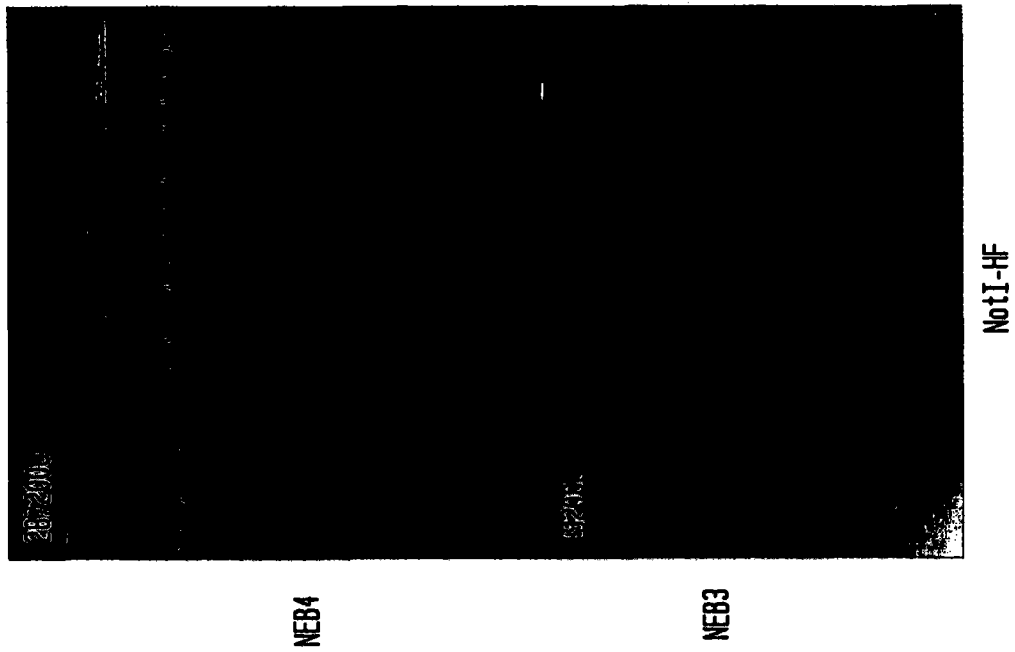


图 17A

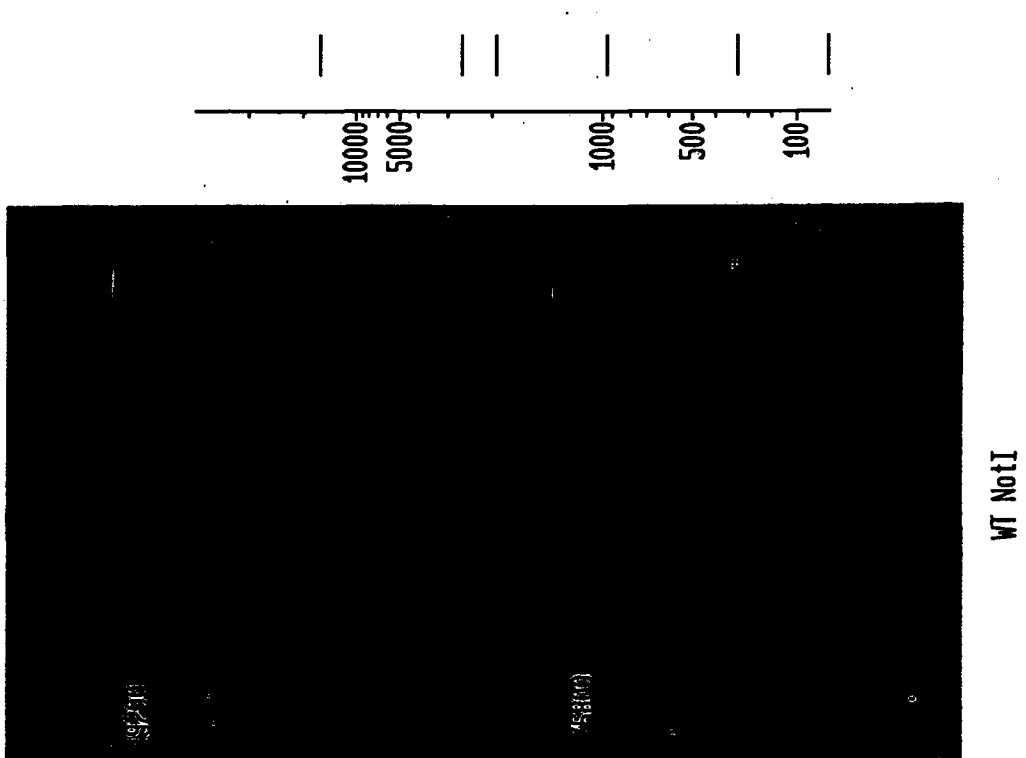


图 17B

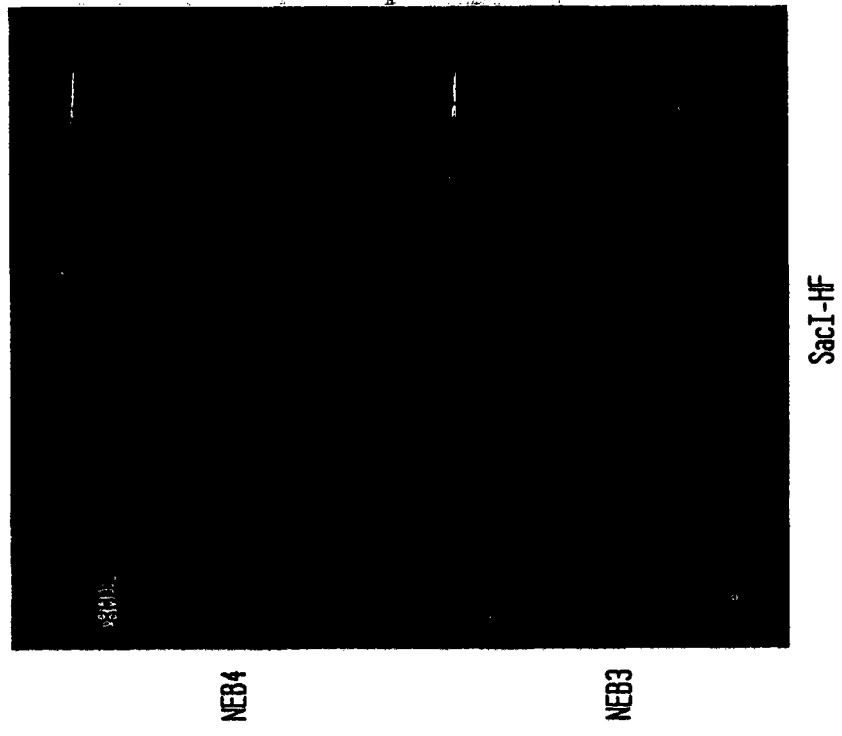


图 18A

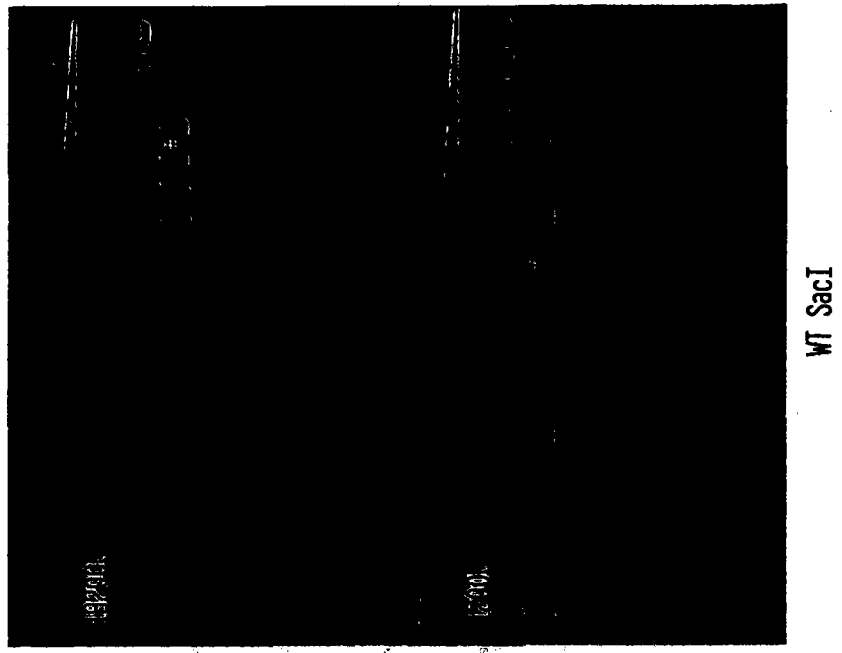
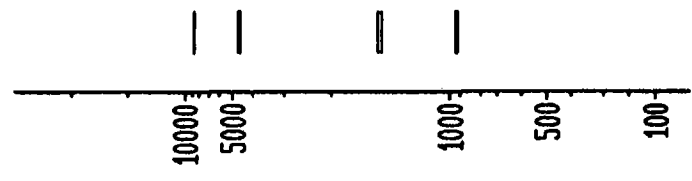


图 18B

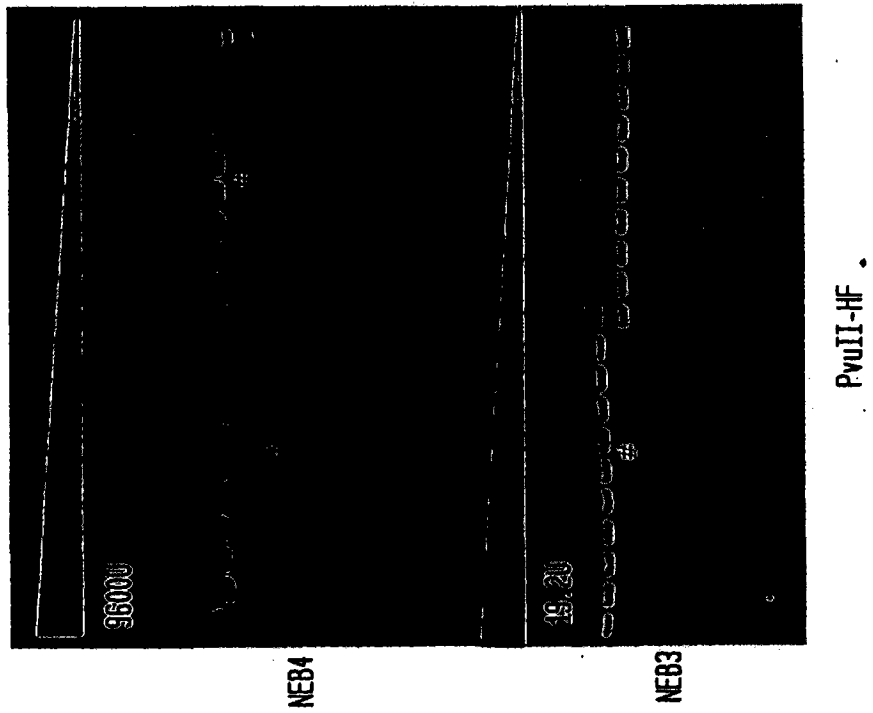


图 19A

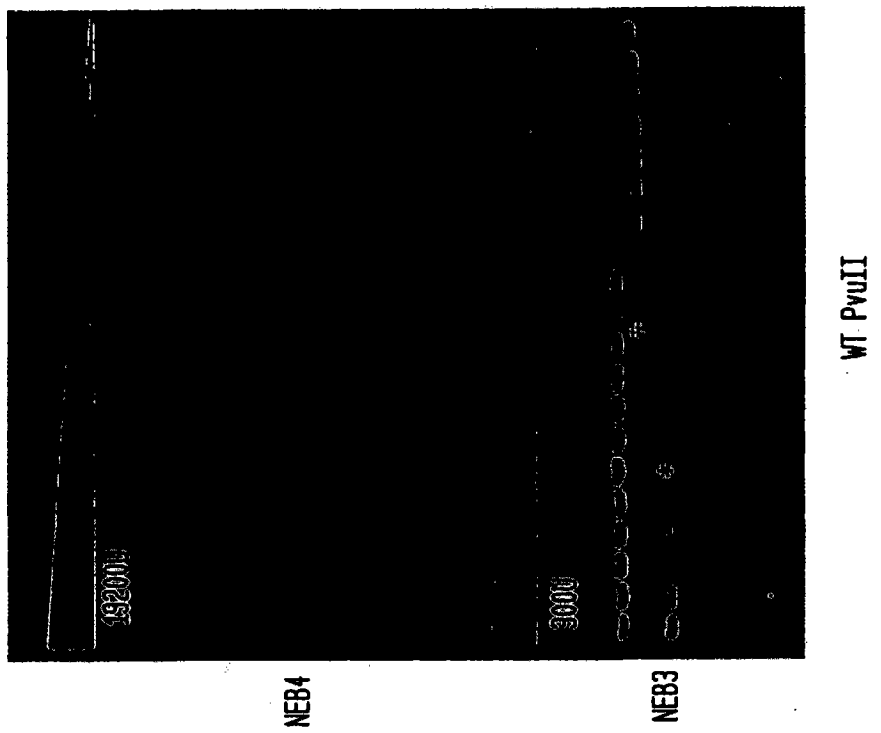


图 19B

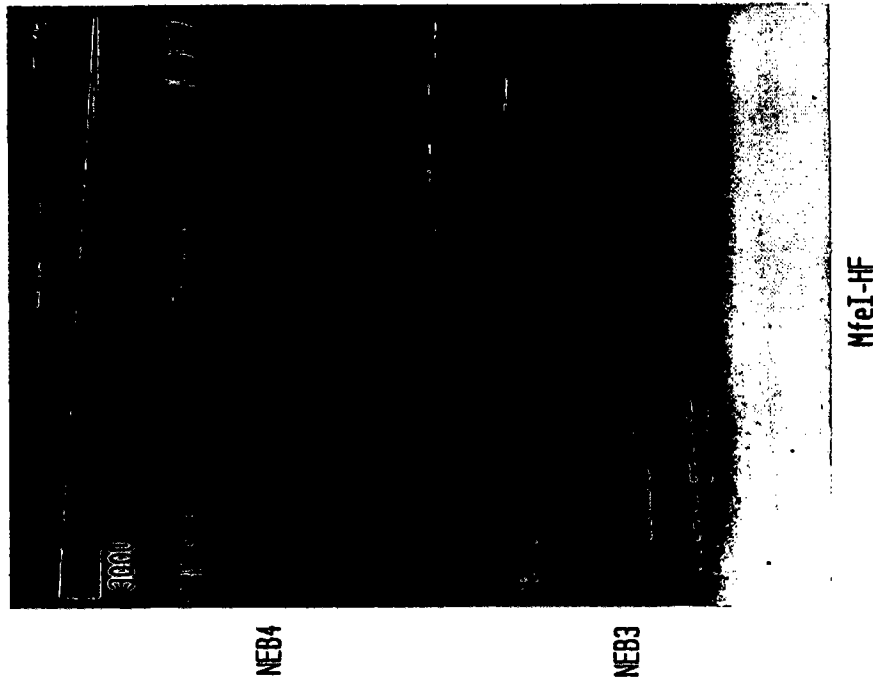


图 20A

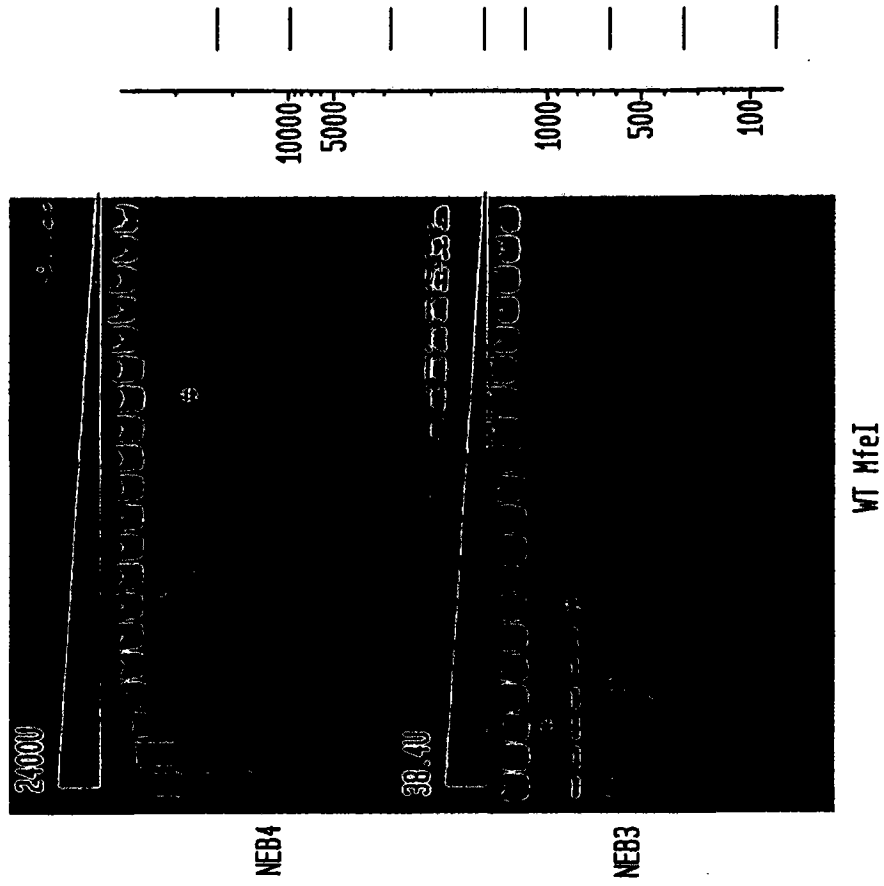


图 20B

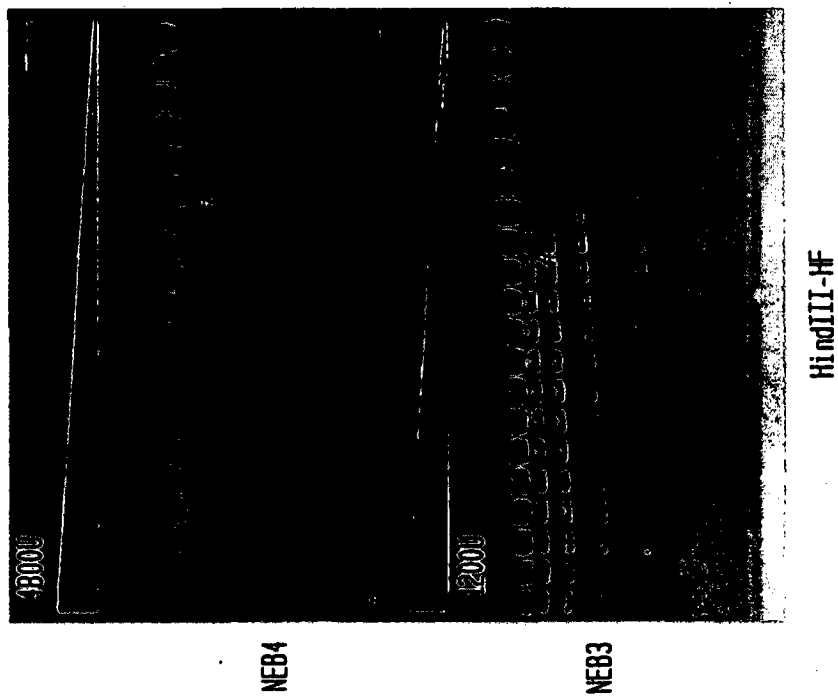


图 21A

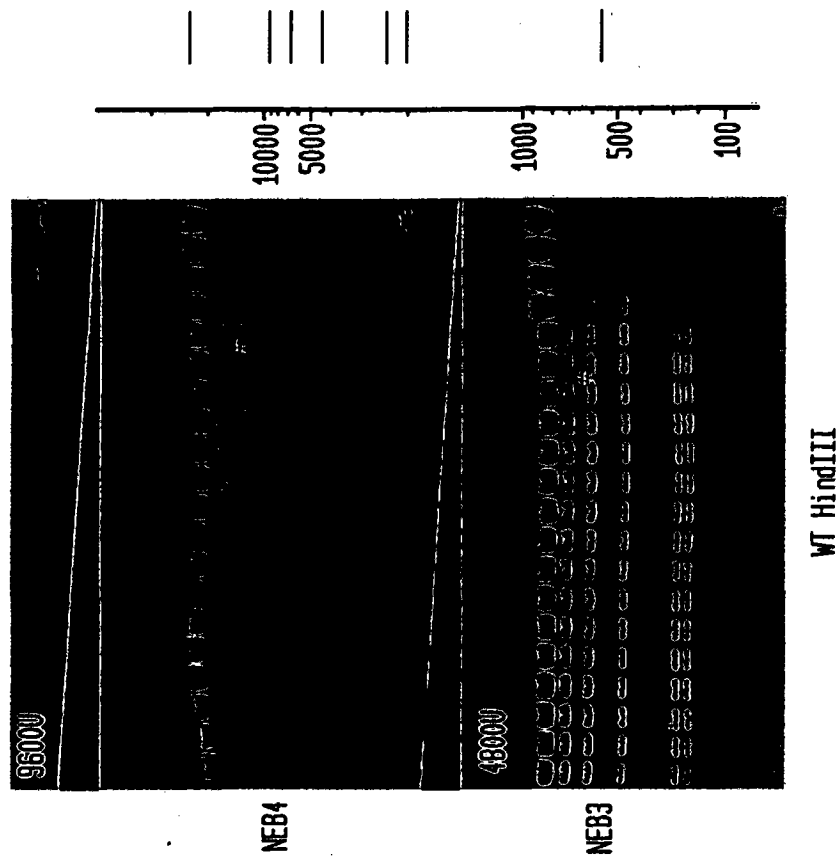


图 21B

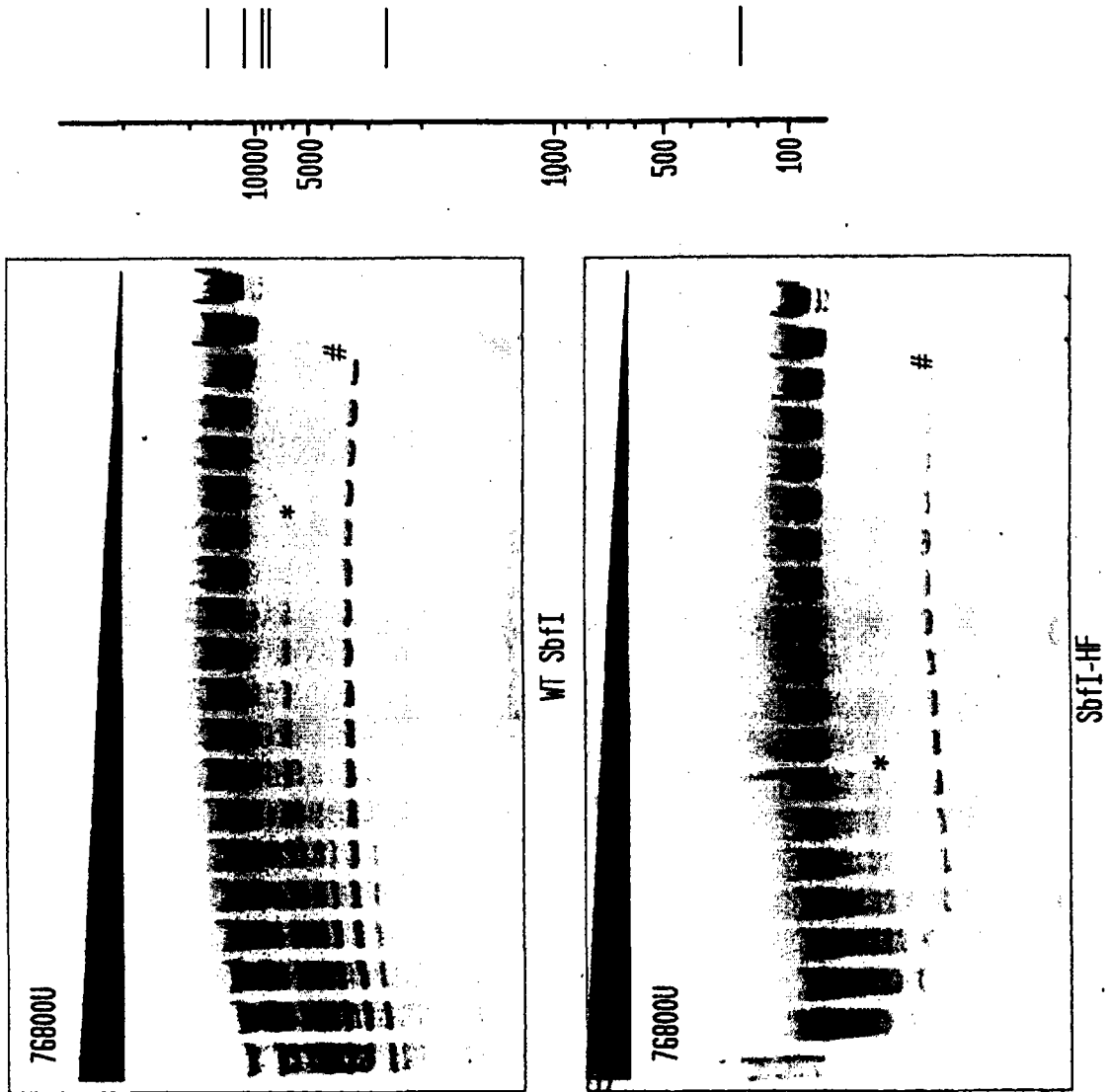


图22A

图22B

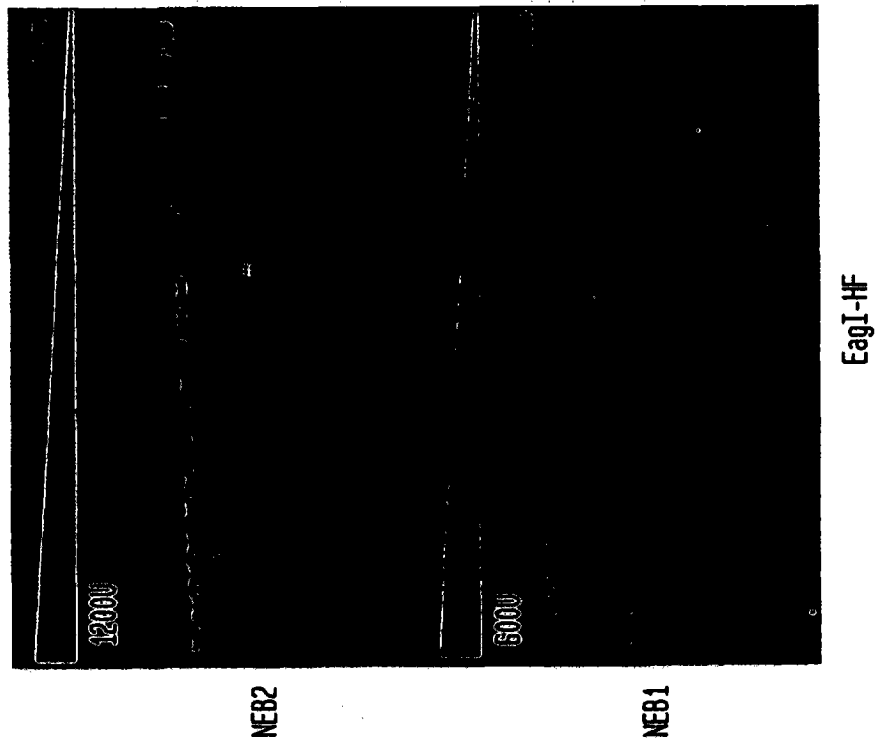


图 23A

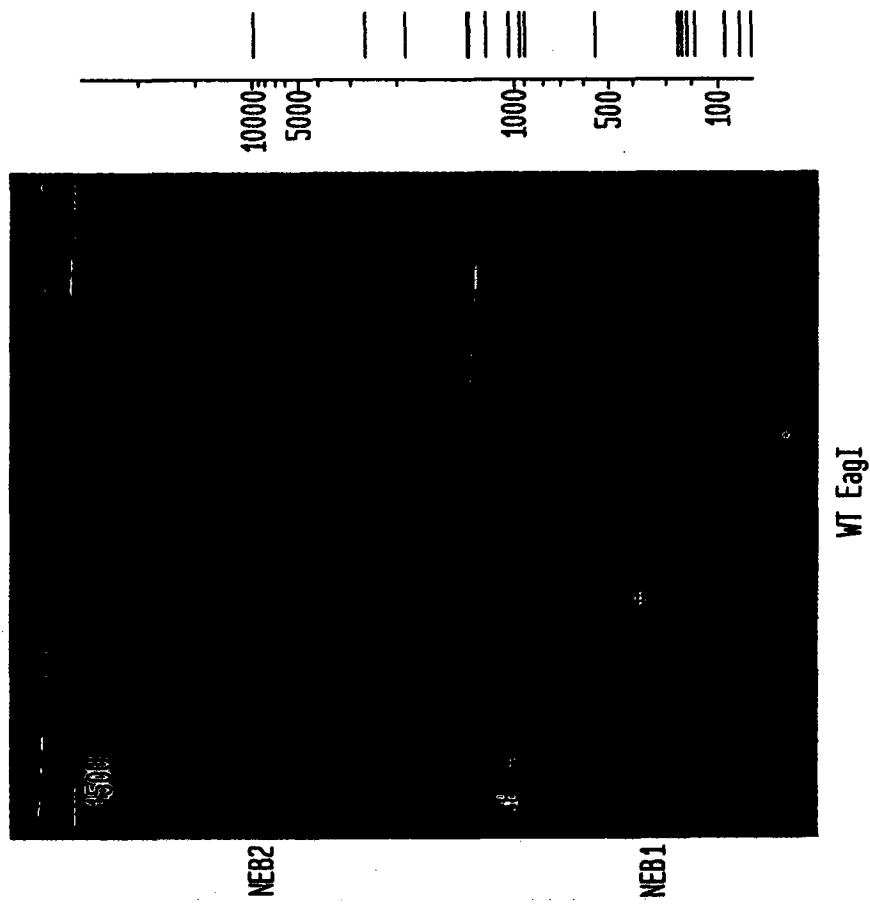


图 23B

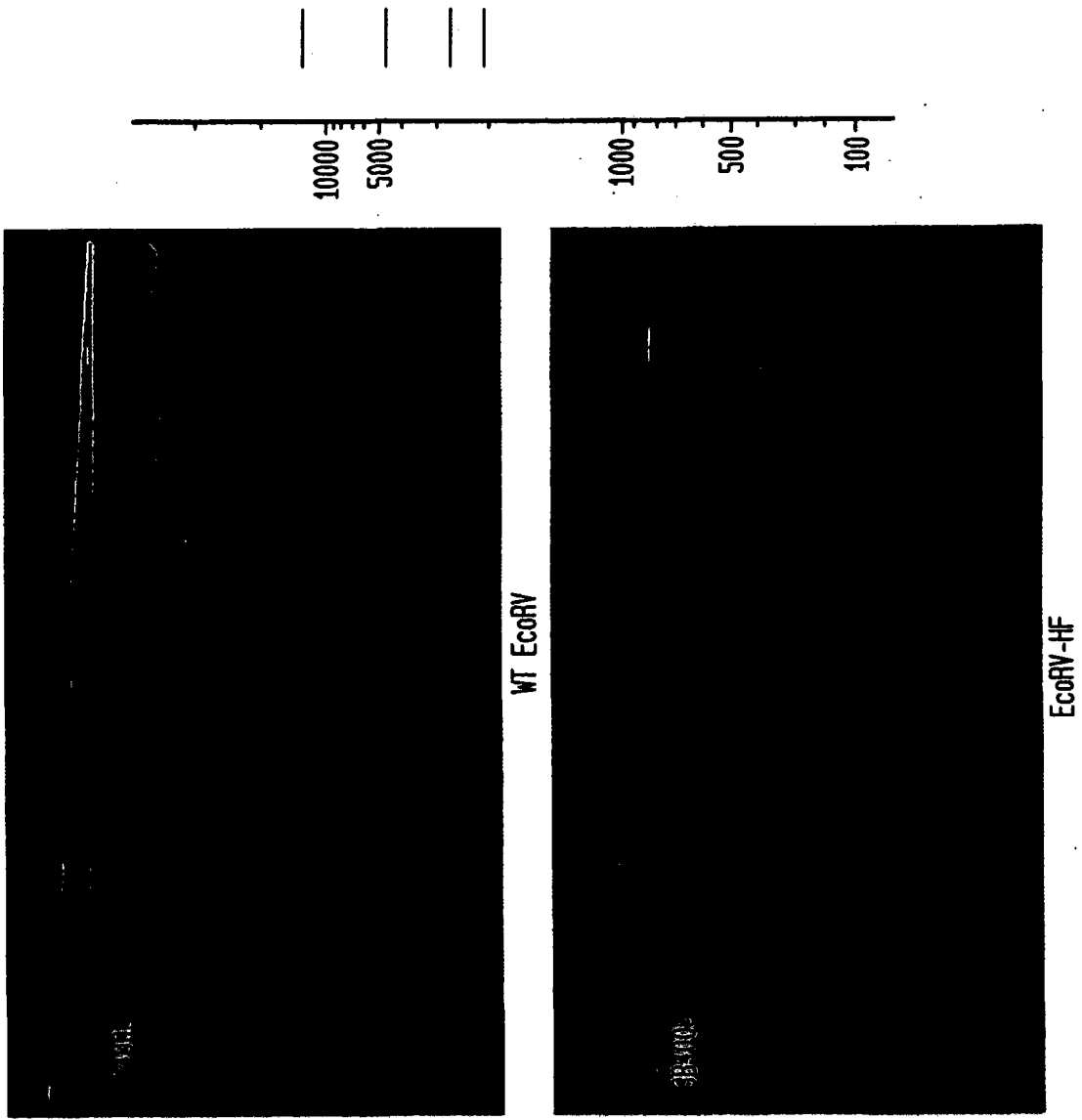


图24A

图24B

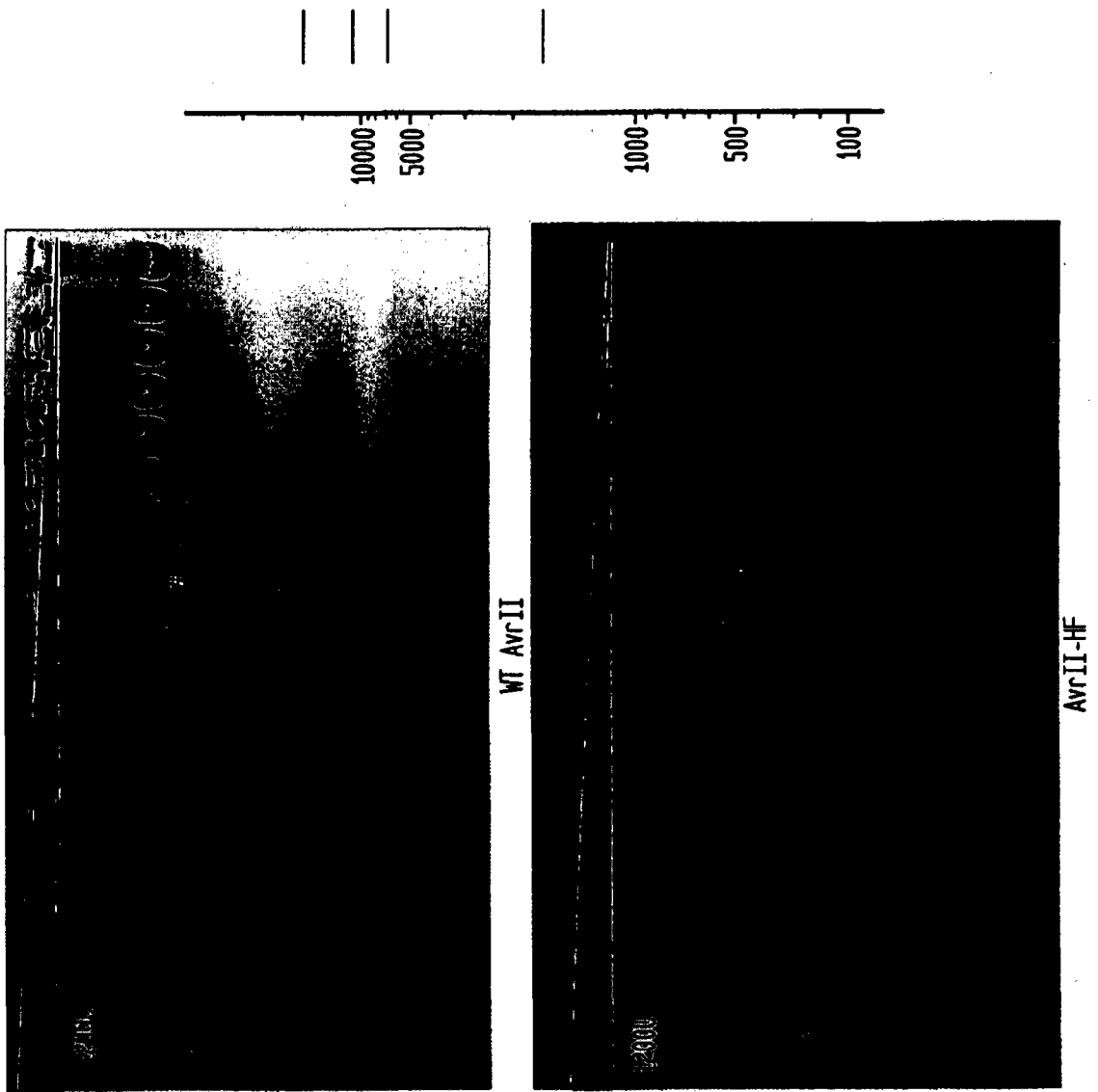


图25A

图25B

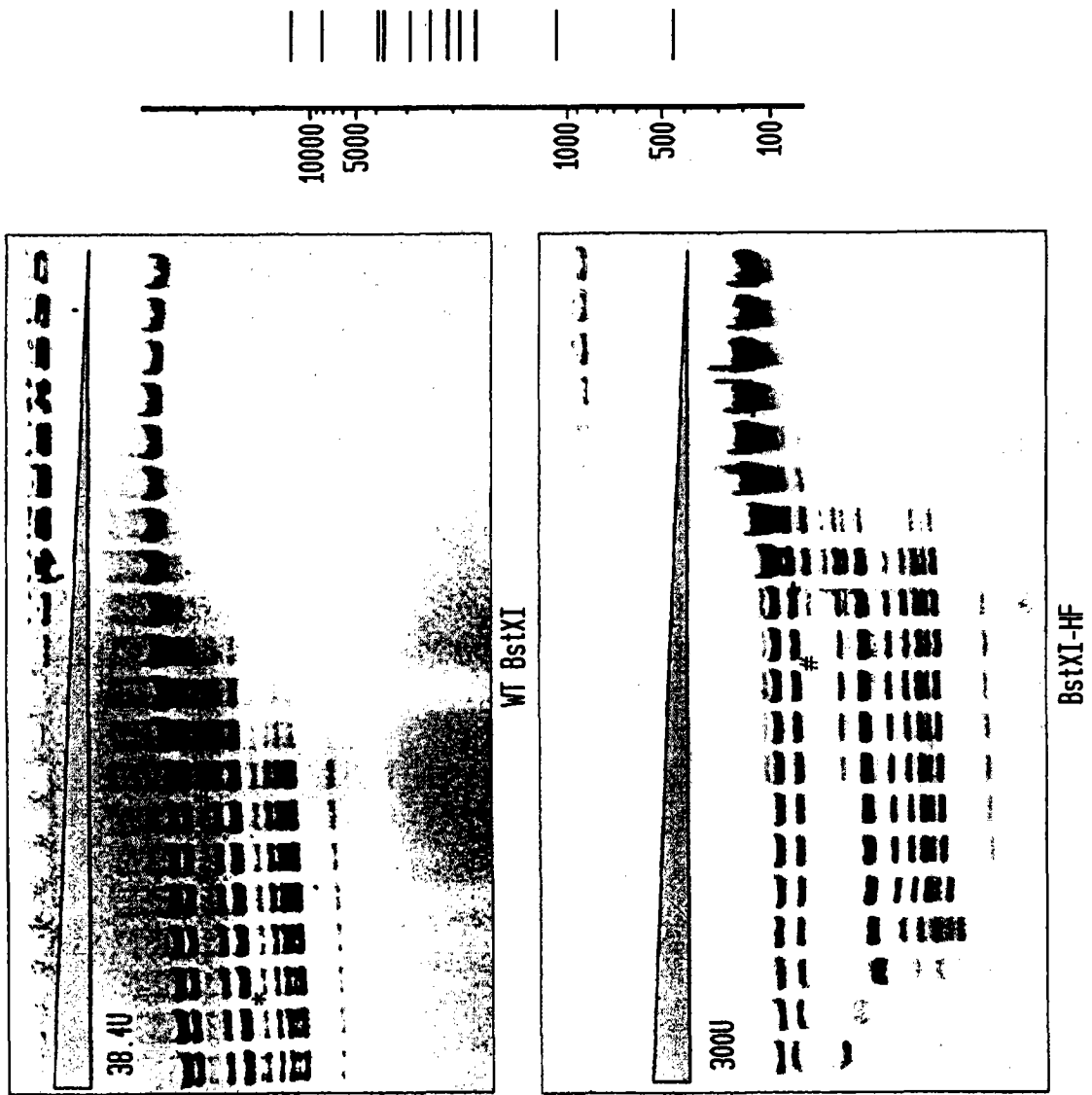


图26A

图26B

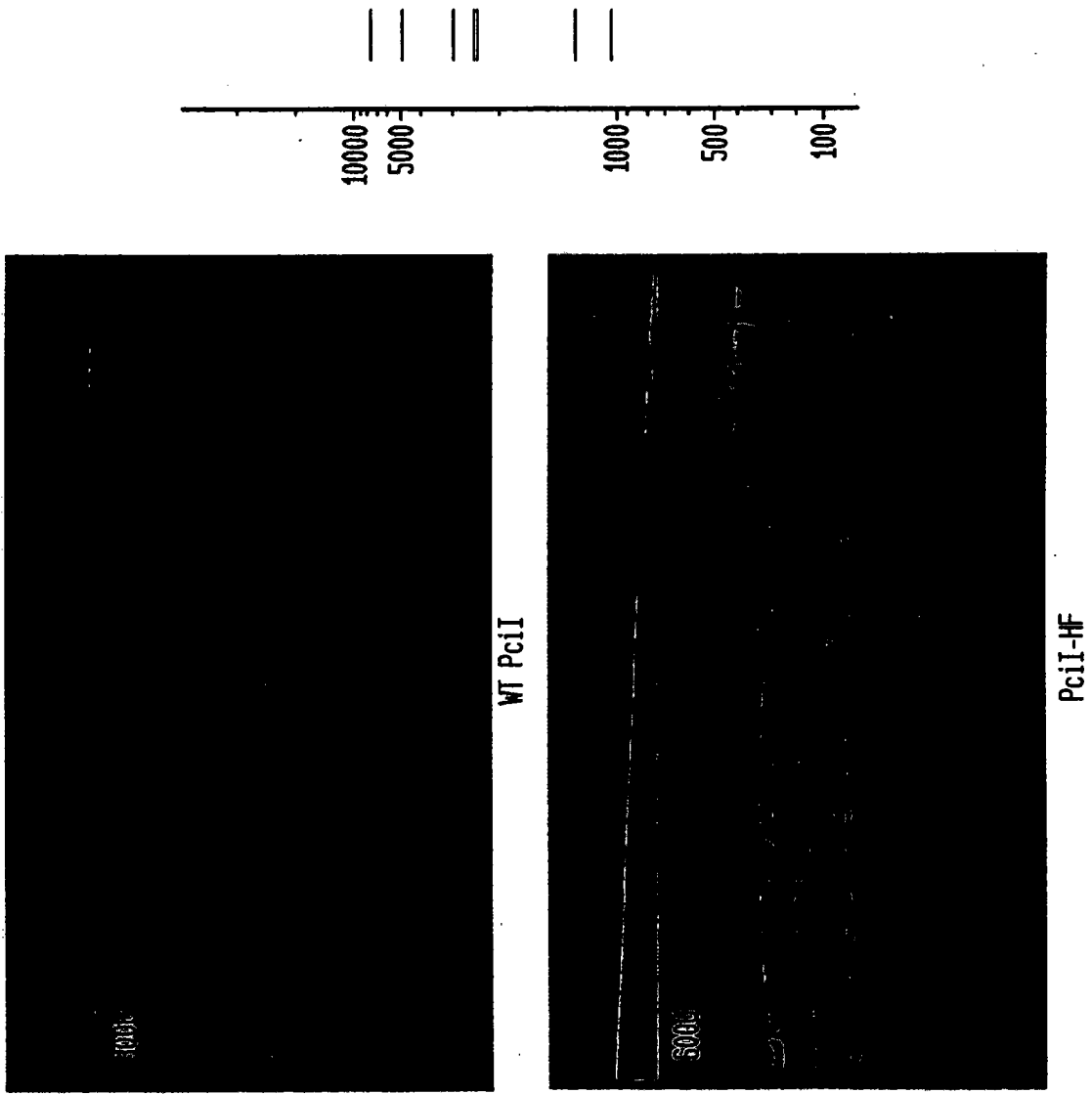


图27A

图27B

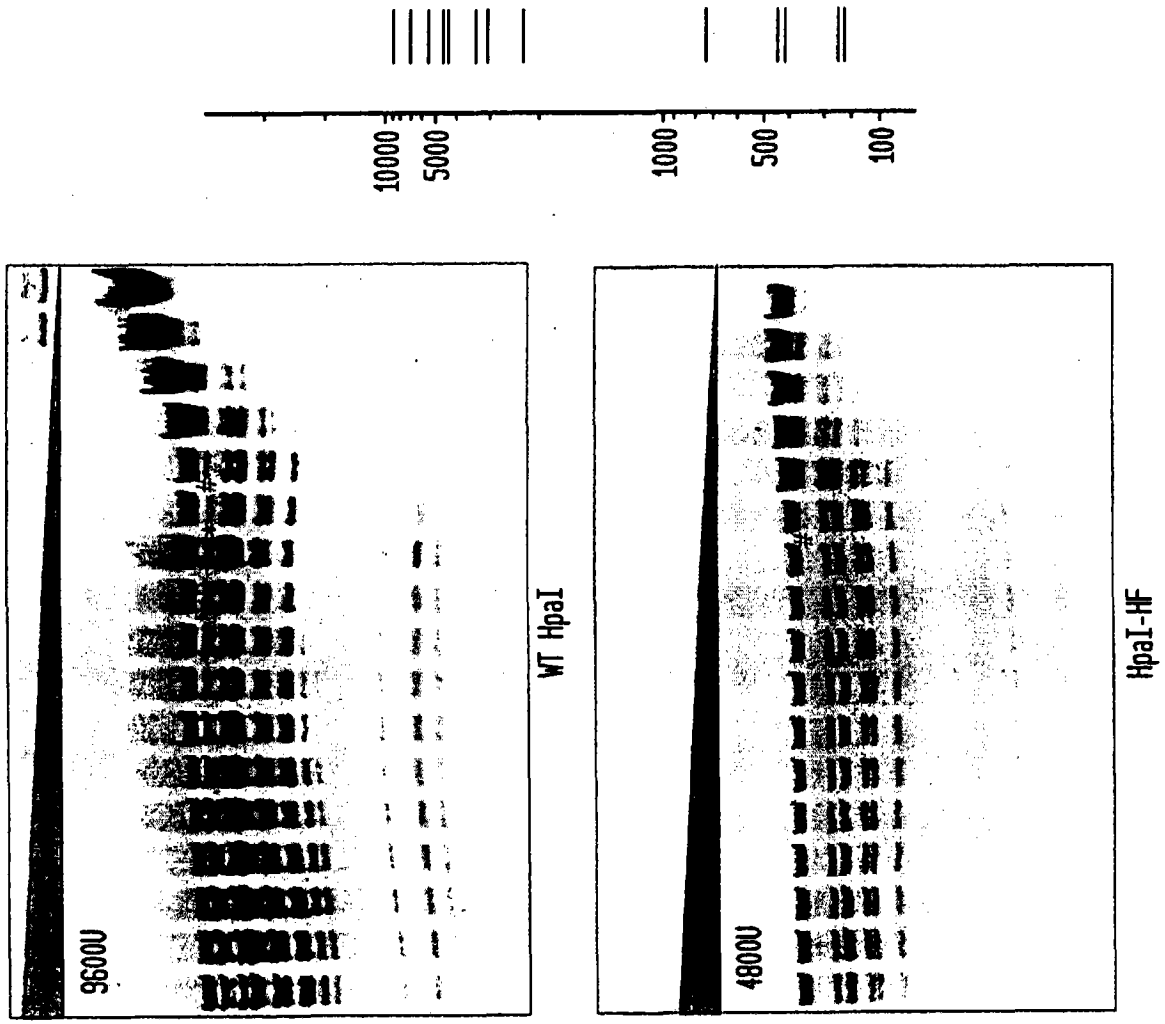


图28A

图28B

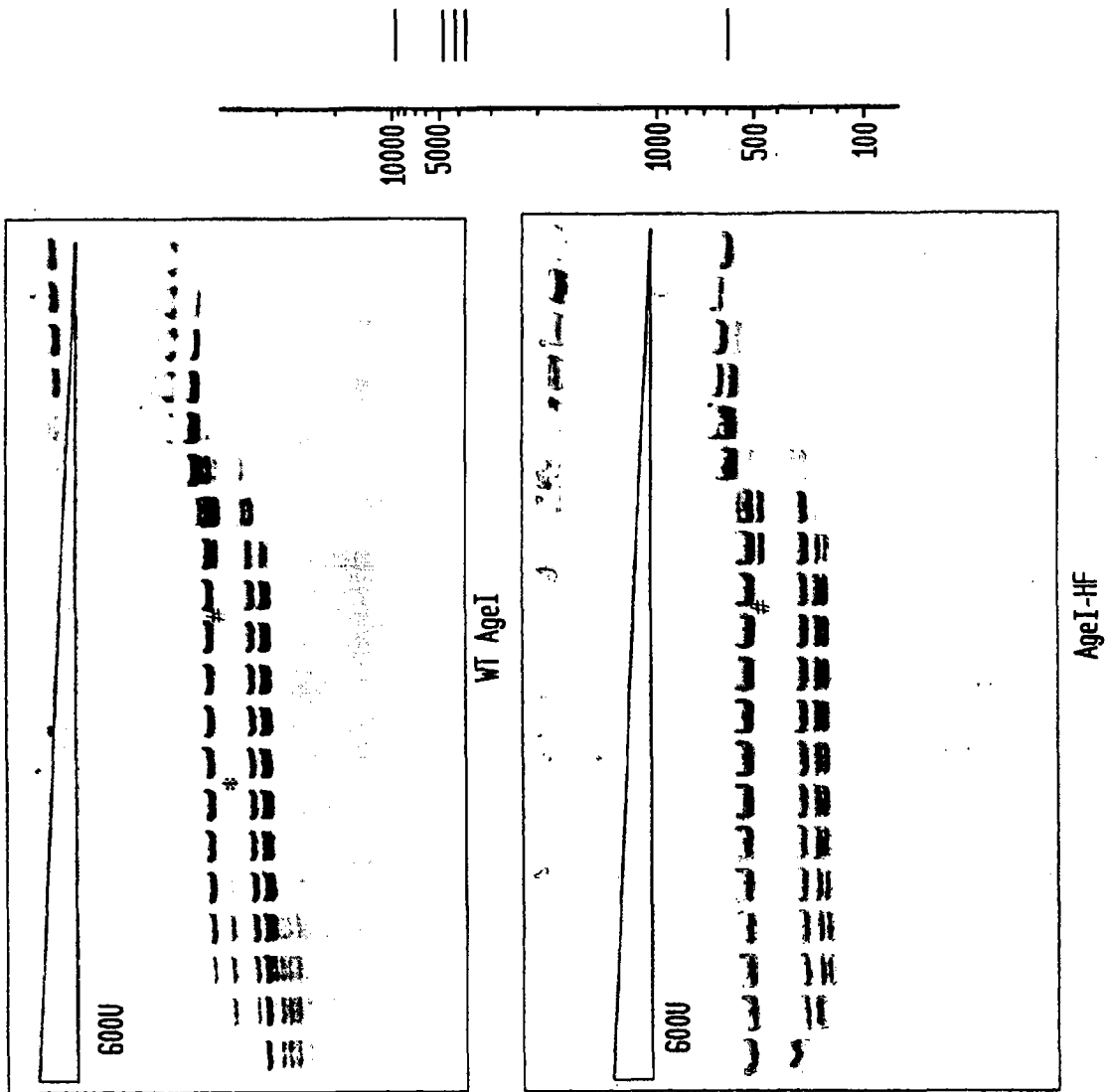


图29A

图29B

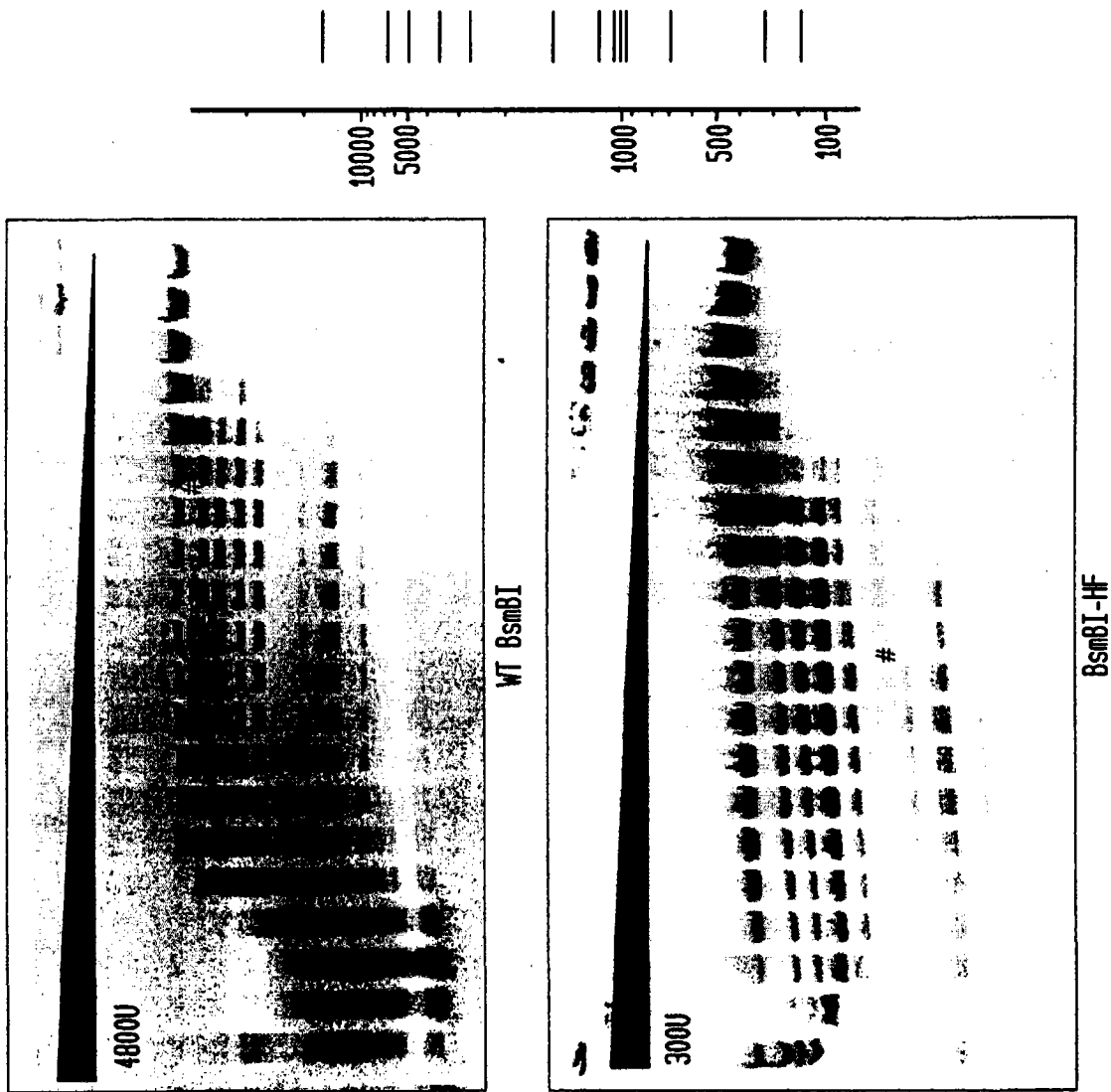


图30A

图30B

| SEQ ID NO:1 | | | | | SEQ ID NO:2 | | |
|-------------|------------|------------|------------|------------|-------------|-------------|-------------|
| 1 | ATGATCAAGT | ACTTGGGTAG | CAAGCGGACG | CTCGTGCCCG | TCCTCGGTGA | MIKYLGSKRT | LVPVLGDIAS |
| 51 | CATCGCTTCG | GCCTCTGAAG | CAACAGAGGC | GGTTGACCTG | TTCACTGGCA | ASEATEAVDL | FTGTTTRVAQE |
| 101 | CGACGCGTGT | GGCGCAAGAG | TTCAAGCGTC | GCGGGCTTCG | AGTTCTTGCT | FKRRGLRVLA | NDIATYSEVL |
| 151 | AACGACATAG | CGACGTACTC | CGAGGTTTTA | GCCCAGTGCT | ATATCGCCAC | AQCYIATNGQ | EVDRRALEAA |
| 201 | CAACGGCCAG | GAAGTTGACC | GCCGTGCGCT | CGAGGCCGCT | CTGGCGGAGC | LAELNALPGE | PGYFTETFCE |
| 251 | TGAACGCCTT | GCCCGGCGAA | CCTGGATACT | TCACGGAAAC | CTTCTGTGAG | ASRYFQPKNG | ARVDAIRNAI |
| 301 | GCTTCTCGCT | ACTTCCAGCC | CAAGAACGGG | GCTCGGGTGG | ATGCAATCAG | DDRYADSWMR | PILLTSLMLA |
| 351 | GAATGCGATC | GACGACCGGT | ACGCGGACTC | ATGGATGCGA | CCGATCCTCC | ADRV DSTTGV | QMAYLKOWAA |
| 401 | TCACGAGCTT | GATGCTTGCG | GCCGACCGCG | TCGACTCCAC | TACCGGAGTG | RAHNDLELRL | PDLIAGDGDA |
| 451 | CAGATGGCTT | ACCTGAAGCA | GTGGGCCGCG | CGTGCGCACA | ATGATCTAGA | AREDAVTLAQ | ELPRVQLMYL |
| 501 | GTTGCGGCTT | CCAGACCTAA | TCGCAGGTGA | CGGTGACGCT | GCTCGTGAGG | DPPYNQHRYP | TNYHIWETLI |
| 551 | ATGCGGTGAC | TCTCGCACAA | GAGCTGCCTC | GCGTCCAGCT | GATGTACCTT | RWDAPESYGI | ACKRIDSRDD |
| 601 | GATCCTCCCT | ATAACCAGCA | CAGGTACTTC | ACCAACTACC | ATATTTGGGA | ATKSPYNMKR | RMPDEMRRLL |
| 651 | GACCTGATT | CGTTGGGATG | CCCCTGAGAG | TTATGGGATC | GCCTGTAAGC | MTKADLAVV | SYNNESWIOP |
| 701 | GCATTGACTC | TCGAGATGAT | GCCACCAAGA | GCCCCTATAA | TATGAAGCGG | ETMMSTLRDA | GYEDVRLAF |
| 751 | CGAATGCCCG | ACGAGATGCG | TCGCCTGCTG | ATGACCATCA | AGGCGGACCT | DYKRYVGAQI | GIYNPSGKEV |
| 801 | CGCGGTTGTA | TCTTACAACA | ATGAGTCGTG | GATTGATCCG | GAGACGATGA | GRVSHLRNIE | YLFLAGPTER |
| 851 | TGTCGACCTT | GCGCGATGCG | GGATATGAGG | ACGTGCGTCT | GCTCGCTTTC | VEVCAASVEH | RALPKEPELT |
| 901 | GACTATAAGC | GCTACGTTGG | GGCTCAAATC | GGGATCTACA | ATCCCTCCGG | AF | |
| 951 | GGAAAAGGTC | GGTCGTGTGA | GTCACCTCCG | AAACATCGAG | TATCTCTTTC | | |
| 1001 | TTGCGGGACC | AACGGAGCGC | GTTGAGGTGT | GCGCCCGGAG | TGTTGAACAC | | |
| 1051 | CGAGCACTAC | CCAAGGAACC | GGAACTCACC | GCGTTCATG | | | |

图 31B

图 31A

SEQ ID NO:3

```

TTGGAGAATT TTTTGAATAA TTTAGATATT AAAACCTTAG GGCAGGTTTT CACCCCTAAA
AAGATAGTGG ATTTTCATGCT CACTCTCAAG CACAATCATG GGAGTGTTTT AGAGCCAAGC
GCGGGCGGATG GGAGTTTTTT AAAGCGCTTA AAAAAGGCTG TAGGGATTGA AATCGATCCT
AAAATCTGCC CTA AAAAATGC CCTTTGCGATG GACTTTTTTG ACTACCCTTT AGAAAATCAA
TTTGACACGA TTATTGGCAA TCCGCCCTAT GTCAAGCACA AGGATATTGC GCCAAGCAGC
AAAGAAAAAC TCCATTACAG CCTTTTTGAT GAAAGGAGTA ATCTATACTT GTTTTTCATA
GAAAAGCGA TCAAGCATT TAAAGCCTAAA GGCGAATTGA TTTTCATCAC CCCAAGGGAT
TTTTTAAAAT CCACTTCTAG CGTGAAATTA AACGAATGGA TTTACAAAGA AGGCACGATA
ACGCATTTTT TTGAATTAGG CGATCAAAG ATTTTCCCAA ACGCCATGCC TAATTGCGTG
ATTTTTCGTT TTTGTAAGG TGATTTAGT AGAATCACC ACGATGGTT GCAATTTGTG
TGCAAAAAAG GCATTTTGTA TTTCTCAAC CAATCTTACA CGCAAAAATT AAGCGAGGTT
TTTAAGGTTA AGGTGGGGC AGTGAGCGGG TCGGATAAGA TTTTTAAAAA TGAAACATAC
GGGAATTTAG AATTTGTCAC CTCAATCACC AAAAGAACCA ATGTTTTAGA AAAAATGGTT
TTTGTCATA AACCTAATGA TTATTTACTC CAGCATAAAG ACAGCTTGAT GCAAAGAAAAG
ATTA AAAAAT TCAATGAAA TAAATGGTTT GAATGGGGGA GGATGCATCA CATATCCCCT
AAAAACGCA TTTATGTTAA CGCCAAAACG CGCCAAAAAA ACCCTTTTT CATCCACCAA
TGCCCTAATT ATGACGGCTC TATTTAGCG CTATTCCTT ATAACCAAAA TTTGGATTTA
CAAAACCTCT GCGATAAACT CAACGCTATC AACTGGCAAG AATTAGGCTT TGTGTGCGGC
GGGCGTTTTT TGTTTTCGCA GCGCTCTTTA GAAAACGCC TTTGCCTAA AGACTTTTTA
AATTAG

```

图 32

SEQ ID NO:4

ATGGGTAAT CTGAATTAAG TGGAAGATTA AATTGGCAAG CATTGGCTGG ATAAAAGCT AGTGGTGCTG
 AACAAACTT ATATAACGTG TTTAACGCTG TTTTGAAGG AACTAAATAC GTTTTATACG AGAAGCCAAA
 GCACCTTAAA AATCTATACG CTCAGTAGT CTTACCTGAT GATGTTATTA AAGAAATTTT TAATCCTTTA
 ATTGATTTAT CAACTACTCA ATGGGGTGT TCTCCAGATT TCGCAATAGA AAATACAGAA ACGCATAAAA
 TTCTTTTTGG TGAAATTAAG AGACAAGATG GATGGGTAGA AGGTAAGAT CCTAGTGCTG GCAGGGGTAA
 TGCACATGAG AGATCTTGTA AATTATTTAC TCCTGGATTA TAAAAGCTT ATAGAACAAT TGGTGAATT
 AACGATGAAG AGATATTGCC ATTCTGGGT GTATTGCAAG GTGATATAAC ACGAGATCCC AAAAGAGTAA
 GAGAAATTAC TTTCTGGTAT GACCACTATC AAGATAATTA TTTTATGTGG CGACCAAATG AATCAGGCGA
 AAAATTAGTT CAACACTTCA ATGAAAAATT AAAAAATAT TTAGATTAA

图 33A

SEQ ID NO:5

MGKSELGRL NWOALAGLKA SGAEONLYNV FNAVFEKTY VLYEKPXHLK NLYAQVLPD DVIKEIFNPL
 IDLSTTQWGV SPDFAIENTE THKILFGEIK RQDQWVEGKD PSAGRGNHE RSCKLFTPGL LKAYRTIGGI
 NDEEILPFVW VFEGDITRDP KRVREITFWY DHYQDNYFMW RPNESGEKLV QHFNEKLKYY LD

图 33B

SEQ ID NO:6

ATGGCTATTA CATTATGTGA CATAAATGGT TGTAGACTTG AGAGAGGACA TACTGGTAAA CATAATAAT TTCCTGAATT
 TGTATGGACT TCTCAATTTA ATAAAAAGA TATTGATAAG GTCAATAAAG CAGGATATGC AACACCAAGA GGTGGGGACA
 AAGGAGCCTA TCAGAACCAT GTTTACAGAA ATAATAAAGT AATTATTCCT TTTGAAAGGT TGGAAAATGT TAATTTAAAT
 AACTATCAAG ATGGATATGT TATTAGTTA TTCCCTAATC AGTACTTTGA ATCAGCCGGG GTAGTTAAGC CGGAATTCTT
 ACAACCAAAT TCATTTGTTA AAGTTGGGGA CAATGCATTT ATTTTATATC GCACACATTC ATCTTTTGAG GAATTACCTC
 CTCTACCAGA CTGGGAGGTT AGACATCTAA AAAAGAACGG TAATATAGTT ACCAGAAGAA GTAAGGACGT AATCGATGCT
 GGACATTATG TCTTACGATT ATCATCAATT AGTAACAAAA AAGAAAGAAA AGAGGGCCCT CCTCAAGGTA TTTTGCACC
 TGAATATGCA AATGCAGAGA CTAATTATCT GTCAAAAGCA TTTTAGCCT GGTAAATTAT TAAAACCTCA AATAGTCCGT
 ATAATGAAGA ACAATTCAA CACTTAAGAG CGATCTTAAT TAGTCATAAT CTCATCAATA TTTCTCAACT TGAAGAAAAG
 GCTATTCTAA AGAATGGTAT CACATGCTGC CCTTTATGCG AGCAAATTAT TTTTACGAA CAGCTACACG AAATGGTTTC
 TTTTGAAGGT GCGTCTGGCC TTGCGAATTC ACAAGAACAG GTTGAGGGTG CAACTAGGTC AACATCAGTT AATTTATTC
 ATATGGTACC ATTAGTATAT GAAACCTTGG AACACAAACC TGATCAAATA GCATGGGGCC ATGCCATTG TAATACTAGA
 CTTGGTCAA GAGAGTGCCT GCCTTTAGT AGACTAAAC AAGAAGGTAC GCCCGTTGGT CTTCTTGATG AAGATTCGAA
 TCTTGAAGTA TTAGGATGGA TTAGTAAAGA TAAGCAATTT ATTCGTACAG AAAATGGGGA AGTTTGGATT AAAATTACAG
 ATATTGAATT TAACGATGAC TTTGAAGAAT AA

图 34A

SEQ ID NO:7

MAITLCDING CRLERHTGK HNKPEFVWT SQFNKKDIDK VNKAGYATPR GGDKGAYQNH VYRNNKVIIP
 FERLENVNLN NYQDGYVIRL FPNQYFESAG VVKPEFLQPN SFVKVGDNAF ILYRTHSSFE ELPLPDWEV
 RHLKKNIGNIV TRRSKDVIDA GHYVLRSSI SNKKERKEGP PQGIFAPEYA NAETNYLSKA FLAWLIKTO
 NSPYNEEQFQ HLRALISHN LINISOLEEK AILKNGITCC PLCEQIIFYE QLHEMVSFEG ASGLANSQEQ
 VEGATRSTSV NLFHMVPLVY ETLEHKPQOI AWGHAICNTR LGQRECLPLS RLKQEGTPVG LLEDSNLEV
 LGWISKDKQF IRTENGEVWI KITDIEFNDD FEE

图 34B

SEQ ID NO:8

ATGATTTTTG CTGATATTGA ATTTGAAAA GAACTTTTTT CAGCTGCTAA TAAATTAAGG GGAAAAATTG CTCCAAGTGA
 GTATAAGCAT TATGTTTTGC CTTTATATT CCTTAGATAT TTATCTCTTA AATACCAACA AAGAAGGAAT GAAATTC AAC
 AACAGATAAA TGATTCAAGG GATCACAAGA AAAATCAAGA TGAAGTGTTA AAGATATTGG AAGACAGGAC TGAATACACC
 AAAGTAAATG TTTTCTATAT TCCTGAAAA GCTAGTTGGG AATACTTATT GAAAAATCC GAAATGATA AAATTAAGA
 AATGATAGAT TCAGCTATGG AAATACTGGA AAATGAATAT GACGAGTAA AAGGTGTTTT GCCAAAGATA TATAAAAAC
 CAAATATACC GAATGAAGTT ATTAGTGATT TACTAAAAC ATTTTCTCAA GAAGTATTTT CAGCACATGA TGAAGAAAT
 GTTGATTAT TGGGAGAGT TTATGAATAC TTTATAAGTA ATTTTGCTAC TACAGAAGGT ACTAGAGGTG GTGAATATTT
 TACACCGTCT TCAATCGTAA AATTATTGGT AGCAATGCTA GAGCCCATTA AAGGTACAGT TTATGATCCG GCCTGTGGGA
 CAGGAGGAAT GTTTATTCAG TCTAATAAAT ATAGAGAAAA TAATCATAAC TTGTGTTTTG TAGGCCAGGA ACAAACGAG
 CTTACTATCA AATTGGCTAA AATGAATGGA ATTCTACATG GAATAAATCC TGAAATTAGA CAAGGTGATT CATTATTA
 TGACCGTTAT CCAGAATTGA AAGCTGAAAT TGTAATATCT AATCCACCGT TTAATATGAA GGATTGGGGA GCTGAAACCC
 TGCCACTTAA TGATAAGCGA TTAATAGGAC CGGTAACAAA CAGTAATGCA AATTACATGT GGATACAGCA TTTTCTATAC
 CATTAAAAAG ATGGTGGTTT AGCAGGATTT GTTATTGCTA ATGGAGCTTT GACTAGTAAT CTGGCTGCTG AAAAAATTGT
 AAGGAAACAC TTAATAGACA ATGATTATGT AGATTGTGTT GTTCAATTAC CTGAAAAAAT GTTCTTTGGT ACTGGCATT
 CAAGTGCTTT AGTGTTTTTA AGTAAGAATC GAAATGGAAG TAACGGCCAT GCCAAAAGAG AAAAAGAGGT TCTATTTATT
 GATGCAAGCG ATAAGGGAAC ATTAGTGGGT AAAAGAATA AAATATTTTT AGATGATGAA ATAAAAGAAA TTGCAGATTT
 ATATCATTCA TTTAAATTTT TAAATGATAA TGATTATAAC CATAGTGGTT TTTACAAAAA GGTTAACATT GAAAAAATCG
 TGGAAATGA TTATAAATTA ACTCCAACCT TCTATGTAGG TGTAAGGAA GAGACTGAAA TGGAGAAGCC ATTAGAGAA
 ATGATAATAG AATATAAAGC GATATTAGAG CAACAATTTG AAGAATCAAA CAACTACAG CAGAAAATAT TAAAGAATTT
 AGAGGGATTA TTATGA

图 35A

SEQ ID NO:9

MIFADIEFEK ELFSAANKLR GKIAPSEYKH YVLPLIFLRY LSLKYQRRN EIQQQINDSR DHKKNQDEV L KILEDRTYET
 KVNIFYIPEK ASWEYLLKNS ENDKIKEMID SAMEILENEY DELKGVLPKI YKNSNIPNEV ISDLLKLSQ EVFSAHDGRN
 VDLLGRVYEEY FISNFATTEG TRGGEYFTPS SIVKLLVAML EPIKGTVYDP ACGTGMFIQ SNKYRENNHN LCFVGOEQNE
 LTIKLAQMNG ILHGINPEIR QGDSLLNDRY PELKAEIVIS NPPFNMKDWG AERLPLNDKR LIGPVTNSNA NYMWIOHFLY
 HLKDGGLAGF VIANGALTSN LAEKIVRKH LIDNDYVDCV VQLPEKMFEG TGIPSALVFL SKNRNGSNHG AKREKEVLFI
 DASDKGLVGG KKNKIFLDDE IKEIADLYHS FKFLNDNDYN HSGFYKKNVI EKIVENDYKL TPTLYVGVKE ETEMEKPFRE
 MIEYKAILE QOFFESNKLO QKILKNLEGL L

图 35B

SEQ ID NO:10

ATGAAAAGTA CTTTGAAGGA ATATAAATG GGTGATATTA CCGAAGTCGT TAATGGTGCC ACTCCTTCAA
 CTA AAAAGCC TGAGTACTAT GAAAATGGTA CAATCCATG GATTACTCCT AAAGATTAT CAGGCTATTA
 CTTTAAATAT ATATCTCATG GTGAACGTAA TATAACAGAG CTTGGTCTAA GAAATAGTTC AGCTAAGTTG
 TTACCAAAG GAACTGTATT ATTTTCTCA AGAGCCCAA TAGGATACGT AGCAATAGCT GATAATTGGT
 TAACTACGAA CCAGGGATT AAAAGTTTA TATGTAATGA GGAGATTATT TACAATGAAT ACCTTTATTA
 TTTTCTTATT GCTAAAAGGG ATTTTATTGA AACATTTGCG AATGGGAGTA CGTTTAAAGA GCTTTCATCA
 ACTTCTGCAA AGAATATACC AATCAATCTT CCTAGTTTAG AAGAGCAAAA GAAGATTGTG ACAATTTTAG
 GGGATTGGA TAGAAAGATA GAATTAATTA ATAAAATTAT TGAAAGCTTA GAAAAATAG CAGAAAGAAC
 ATATAAATAT TGGTTTGTG ATGAATAAA TCAAGATGAA CAGCACATCC GTAATGGATG GGAAACTGCT
 AAAATTGGCG ATGTGGTGA ACTTTTGGGA GGGGAACCC CTA AA ACTTC GAAAGTAAG TATTGGGAAG
 ATGGAGATAT TAATTGGTTT ACTCCTTCAG ATTTAACAAA AACTAGACAG CTTTTTGTAC GTGATTCTCA
 AAGAAAAATA ACAATTGATG GACTTAATAA CAGTGCAGC AAATTAATTC CCCCTTATTC CTTGTTAATG
 TCAAGTAGAG CTACAATTGG CGAGTTGGCA ATTAATCAAG AATCTGCTAC TACAAATCAA GGGTTTATTG
 TATTAATACC AAATGAAAA ATTTCTATTT ACCAATTATA CTTTTGGGCT AAACCTAATA AGAGCAAAAT
 TATTCAATG GCAAATGGTA GTACTTTTAA AGAAATTAGT AAGCGGGATT TAAATCTTT GGAGATAATA
 TTACCAAAA ATATAGACAC TTTAATTCA ATTATGCAAG ATTATTTTAG GAAAATTGAG GAGTTAATTG
 ATGAAATAAA AATCTTAAA ACCGCAAGAG ATAATTTAAT TCCAAACTT ATAAAATGA

图 36A

SEQ ID NO:11

MKSTLKEYKL GDITEVVNGA TPSTKKPEYY ENGTIPWITP KDLSGYFFKY ISHGERNITE
 LGLRNSSAKL LPKGTVLFSS RAPIGYVAIA DNWLTTNOGF KSFICNEEII YNEYLYYFLI
 AKRDFIETFA NGSTFKELSS TSAKNIPINL PSLEEOKKIV TILGDLRKI ELNYKIESL
 EKIAERTYKY WFDVDELNODE QHIRNGWETA KIGDVVELLG GGTPKTSESK YWEDGDINWF
 TPSDLTKTRQ LFRVRSQRKI TIDGLNNSAA KLIPPYSLLM SSRATIGELA INQESATTNO
 GFIVLIPNEK ISIQLYFWA KLNKSKIISM ANGSTFKEIS KRDFKSLEII LPKNIDTFNS
 IMQDYFRKIE ELIDEIKILK TARDNLIPKL IK

图 36B

SEQ ID NO:12

ATGAAACAGT TTGCAGATCC TTTTAAAAGA AGATTCCTTG ATGCAATTGA ACATCATCTT GATGGAATTT
CTGAGAAAAT AAAAAAAGAC TTTACACACA AAAACTTTTT AAAAGAATTG AATGGCCTTA AAGGTGATAA
AGTCTATCAT GACTTAGGCT TTGATACCGC TGAATATACT CTGGTACGTC TTATAGGAAG AATGAGCATA
AGCGTTGGGA GAAGGCTGGG GGAGATATAC GATAAAGTCC CTCGTTATGT TGCTGCCGCG CGATTTGGTC
TTCAACCAA TCAAATTGCA GAAGTATTG ATGGTCTTGA GTTAGATATA GCTTTGCCA ATAGCCTTTT
GTCAGATGAT GATAAAATC ACATAAAAA AATAACTGAA AAGATGTCAG GCGAAACATA CTCGGGAATC
GGAATCGAA TTCGTTATA CTTAATCCA AATGACAGT CCCGTTAAG AAAAGACGC GATGTAGCTT
CTAAATTGTC GGCCCGGGG TTATTCCTG TTTATTTAAT ATTAGCTCT CTCAGTCCTA GGAATGATGC
AATAGCCCGT CTTAAAAGAG GGGGATGGAG CTTAAACAG GGCAGGAAG CCTTAGACTT CTTACCGAA
CTTTAGGAG TGGATATTG GTCTGTTTA TCTGACCAA TAATAGCCG AGAACTAGG GAGAAAACAT
CAAAAATTAT GAAGTCTATA TTTGAATCAG AGGCATCCA ATCTGTTATA CCGGAGAGT GGAGTAACT
ATAG

图 37A

SEQ ID NO:13

MKQFADPFER RFLDAIEHHL DGISEKIKKD FTHKNFLKEL NGLKGDVYH DLGFDTAEYT LVRLIGRMSI
SVGRRLGEIY DKVPRYVAAA RFGLOPNOIA EVFDGLELDI ALRNSLLSDD DKIHKKITE KMSGETYSGI
GIEIRYNFNP NDSSRLRKDV DVASKLSAAG LFPVYLIFSS LSPRNDAIAR LKRGWSFKO GOEALDFLTE
LLGVDIGSVL SDPIIAAETR EKTSKIMKSI FESEAFQSVI PGEWSKL

图 37B

SEQ ID NO:14

ATGACAAATT TTTCCGCACTC AGCTCTAACG AGCTACGATC TTCTCGGGCA TGAATTGTC CAAGATTCTG
 AAGCTGTTAG CTCGGGTCCA TATCTGGTCA GCTATGACCC GATCCCTGTA CGTCGGTCTA CATTCCTAGC
 TGGACTGTCA GAGAACGTTT ACTCGTGGT TCGTCTACA CCAAGTTTCG GACCGGATCT AGTTCGAACA
 ATCATCAAC AGATGAATCT TGCGCCGAC TCACACATCC ATGACCCTTT CTCAGGAGCC GGGACTACCG
 CGATTGAGGC TTCGTTAGAG GGCTATGAAG CAAGCTGCGT AGAAGTTAAT CCGTTTCTCT ACTTCGTGGG
 GAAAACATCC ATAGATTGGT CTATCAATGC TGATGATGCT GCAGCGCAGC TAGAAAGCAT TAAAAATAA
 TATTATAGCA TGCTGCAAC CGCTACTTTG GATAACATAG CCGACCTAGG AATAGATATA CCAAAAATAC
 ACAATATTCA TCGGTGGTGG AGAAACGATG TTCTTAAAGA TATATTAGTC CTAAAATCTT CTATCAGATC
 TTGCACACAA GATAAGTATT GTTCCITTTT TGAGCTAGCC CTAGCTGCAG TTCTCGTTCC AGATTGACA
 AATGTAACGC TAGGAAAAC ACAACTGCAC TTTGTAAACA AAGACGATAA AGAGATAAAC GTCTGGCCTA
 CATATGAATC TCATGCAAAA AAAATGATTC ACGACTTGTG ATTAATTAAT AAGCAAAATT TCGAATTTTT
 GCCCAAGATT ATTTATGGTG ATTCAACTCA AAAATCAACA TTTAGCGAGG TGGCAGGGAT AGATGCTATA
 ATAACATCCC CTCCGTACCC TAATAGGTAC AGCTATATTT GGAATACTCG CCCTCACCTG TACATTCTTG
 ATATGATTTT CGAAGCAAAA GAGGCTTCGC AAATAGATCG TAGAACGATT GGTGGAACAT GGGGGACAGC
 AACTCCGAA TTAGGAAAGG GTATATTTT TCCAATCAAT GCTGTAGTCA AAGACGCGCT TGAAGGGGTT
 CACGAAAGAA TCGCCGTTT CGATCAACTC ATGGCAAAC ATGTAACTCA TTATTTAAT CCGCTCTTTT
 TACATATAGA AGCTATAAAA CCATCACTTA ATCCAAAAGC AAAGCTTGCT TATGTTGTTG GGAACCTTG
 GATTAAGGGC GAATATGTAG CCACTGACGT AATCTTAGCA AAAATTATCG AAGGGGCTTT GCCAGGCTCA
 TCAATTGATG GTCTTCATCG TTTCCGTCGC CGGAACAGTG GAAAGAATCT CTTTGAAACT ATAGTTTACT
 CCACTCTCCC GGTATAA

图 38A

SEQ ID NO:15

MTFNFSHSLT SYDLLGHEIV QDSEAVSSGP YLVSYDPIPV RRSTFLAGLS ENVHSWFRLT PSFGPDLVRT
 IIKQMNLAHP SHIHDPFSGA GTTAIEASLE GYEASCEVN PFLYFVGKTS IDWSINADDA AAQLESIKNK
 YYSMSATATL DNIADLGIDI PKIHNIHRWW RNDVLKDILV LKSSIRSCTO DKYCSFFELA LAAVLVPDLT
 NVTLGKQLQH FVNKDDKEIN VWPTYESHAK KMIHDLSLIN KONFEFLPKI IYGDSTQKST FSEVAGIDAI
 ITSPYPNRY SYIWNTRPHL YILDMISEAK EASQIDRRTI GGTWGTATSE LGKGIFSPIN AVVKDALEGV
 HERIAGSDQL MANYVTHYFN RLFLHIEAIK PSLNPKAKLA YVVGNSWIKG EYVATDVILA KIIEGALPGS
 SIDGLHRFRR RNSGKNLFET IVYSTLPV

图 38B

SEQ ID NO:17

>M1.EarI CTCTTC 1245 nt

```

GTGAATCAGA AAAATGAAAA ATCATTATG CGTTTGCAAT CAACCTTTAG CGGTGGCAAA
GGTAGTCCAA TGCATGATTG GTACCCATAT TTAGAGGGTT ATTCTCCCGA ATTTGTGAAA
TGCTTGATTT CACGATTTGC TCCTAAAGCC AAAACAATTT TAGATCCATT TTGTGGCTCT
GGAACAACAG CCATTGTTTC CGTTTTAGAG GGTTTAAATA ATTACTATTG CGAAGTAAAC
CCTTTATGCC AATATATTAT TGAAACTAAA CTAATAGCTT TAACATTAAG CGAAGAAGAA
AAAACAAAAT TAGTAAATGA ACTTTATTCT ATTTCTAATG AAATAACTAA TGTACTCAAA
CCTTCTGCAA CCGAGACAGA TCTAGAGAAA TCATTTAAAT CCGTTTTTGG TAATACGAAA
TTTTTTGAGG ATCACATATT TAAAGATATA CTTAGTTATC AATGTTACAT TAGCTCTATC
GAAGATGAAA ATCTTAAGAG ACTTCTGACA ATAGCAGGGA TTAGATCGTT AATCCCTTCC
TCGTTATTGG TAAGACGAGG TGATTTACGA TTCAAGACAC AAAAAGAATT AGAGAAAGGC
AACCAGGGCT TTCGCTTTC TGTACAAAA AGCTTAGAAT TAATTGCCAG TGATTTATTA
GACATTACGG AAGGTAGTGG TTTAGCTACC TTCTTATGTG ATGATGCCAA AGAAATATCT
GGGAATAACC TGATTGATGC TGTAATAACA AGCCCGCCAT ATTTAAATGG CACAAATTAT
TTTAGAAATA CAAAATTGA ACTTTGGTTT ATAGGGAAAT TAAAGACCAA ATCAGATCTA
AGACATTATA GGGATTAGC TATTACCAGT GGTATTAACG ATGTAACTAA AGGTAAGGC
TTATCTTCAA ATAATACTAT TATCTCAGAA ATACCATTAT TATCTGAATG TATTAAGAA
CTAAGCATAA AAGAGTATGA TAGTCGTATT TCAATGATGG TTGAAAATA CTTTTGGGAC
ATGTTCAAAT TCTTATCAAA ACTCCAAAA TTAATACTA ATGATGCGAC TATCTGTATA
GATTTAGGTG ATTCTGTTA TTGTAACGTC TACATCCCTA CACAAGATAT TTTGAAAGAA
ATGATGTCAA AGTTAGGTT TGAAGAGAAC GAAAGGGTCA TTCTTCGTGA ACGAAAATCC
CGCAATGGAA CAAAGTTAGT CCAGACTGTT CAGGTTTTTA AATGA

```

图 39

SEQ ID NO:18

>M2.EarI CTCTTC 1140 nt

```

ATGAAAAATA AATATTTTAG TAAAAATGG GAGCAATTCA AGAAAGAATT ACCCATCAA
TCAGGTGAAA TGGTAAAGAG AAATTGGGGC CATAACTGCG ACTCTATGTG TTCATACCAA
GGGAAACTTA AACCATCAAT AGCTAGATCT TTAATTGATA CATTATGCC ATCAAGTAAG
GGACGTATAT TAGATGTCTT CTCAGGTGTT GGCACCATTG CTTTGAAGC AAGATTACTT
GGTCATACTG CATATGGATT TGATATTAGT CCAGCAGCAG TTAATATTTC ACGCGCAAAA
CTAGAAGTTA TAAGTAAAA TGAATCCAA GAGGTAATTA ATAAATTATC TGATTTTATT
GAGCAAAACA AAAATTCAAT AGATTATAAC GAACATAATT TAATAAGTT TAATGGTTCA
ATTGAATCCT ATTTTCATCC TGAACTTTT AAGGAAATAC TGTGTGCTCG TAAATCTTT
TTAATAAAAG GTGAATTAAA TGCATCTGAA TCGTTAGTAC AGTCATGTT ATTACATATT
TTACATGGTA ATCGTCCGTA TGCATTGAGT AGAAAGTCCC ATCCTATTAC ACCTTCGCG
CCTACTGGAG ATTTTATATA CAGTAATTTA GTTATAAAGT TAATCAAAA AGTTGAAAGA
GTCTTGCAAA ATTCTGATGG TATCCAGAT ACTGGCAGCA AAGTATTTTA TCAGGACTCT
ACAAAAAGTT GGCCTGAAGA AGTAAATAAT TTAGATGCAA TTATAACATC ACCTCCATTT
TATGATAGTA CCCGTTTCTA TTCAGCAAAT TGGATGCGAT TATGGTTTTT TGGTTGGGAA
AAAGATGACT TCCAACGAA GCCAAAAGAT TTTGTGGACG AAACCTCAGAA AAAAAGCTTT
GAAATATATG ATAATATATT CAAACAATCT CAACAATGCT TAAAAAAGA TGGCGTTTTT
TTAATGCACG TTGGCAAAAG TAAAAAAGT GATATGGCAG GACAAATTGC TAAATTTGGT
AGTAATTATC TTAGCCTTAT AGATATATTT GACGAAAGTG TTGAACATTG CGAAAGTCAC
GGAATTAAG ACAAAGGCAC GACAACCCAT CATCAGTACC TTGTCTTAC GAAAGATTAG

```

图 40

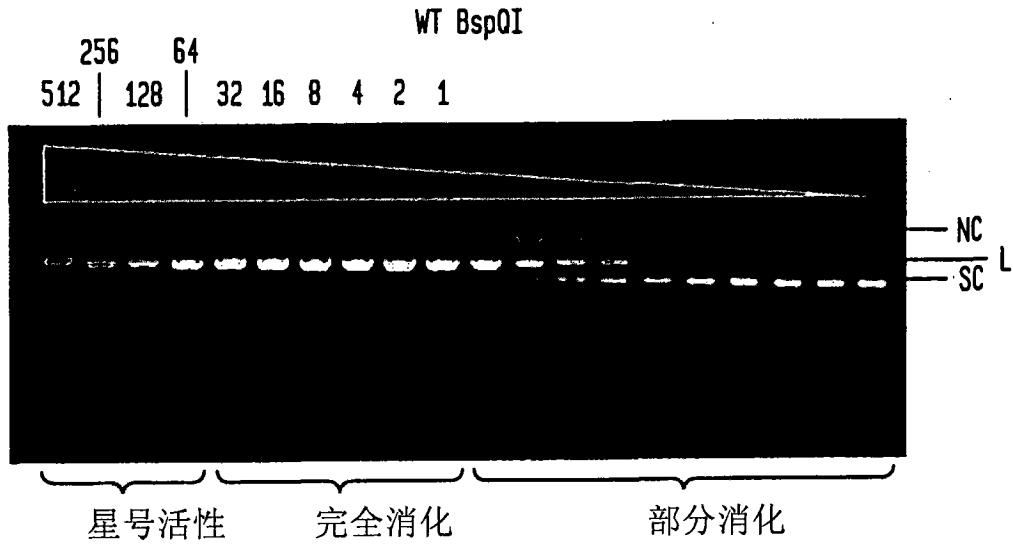


图 41A

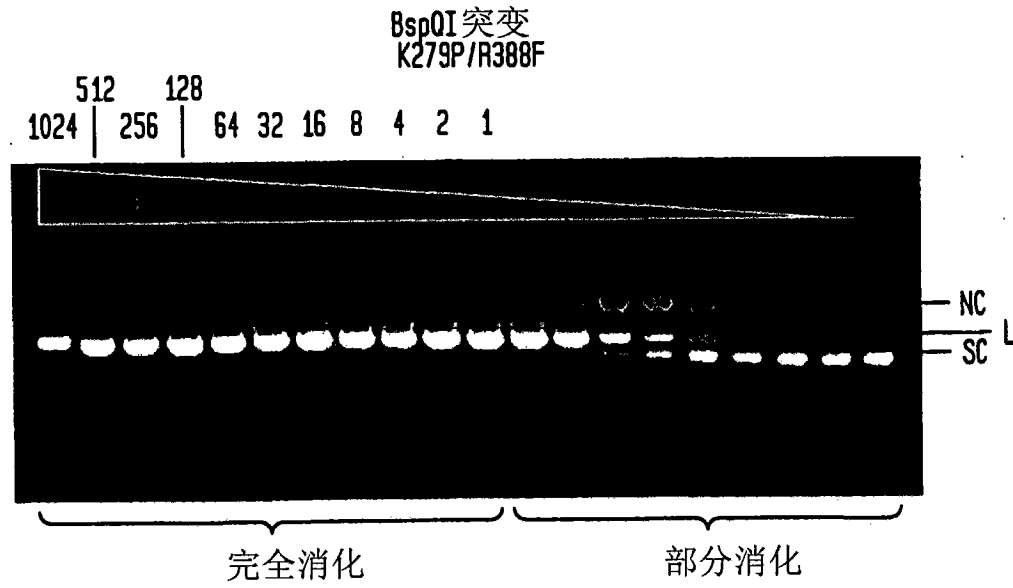


图 41B

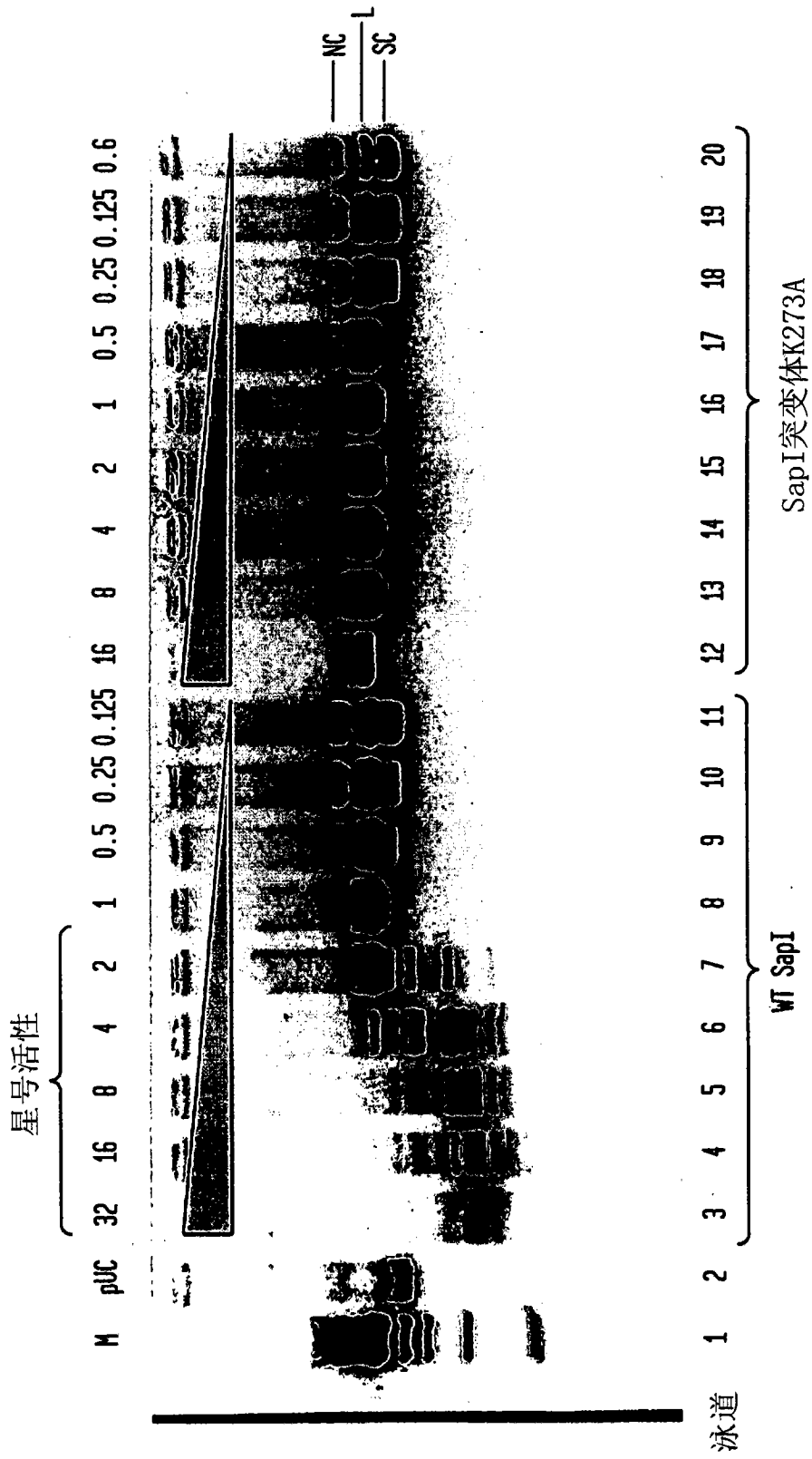


图 42

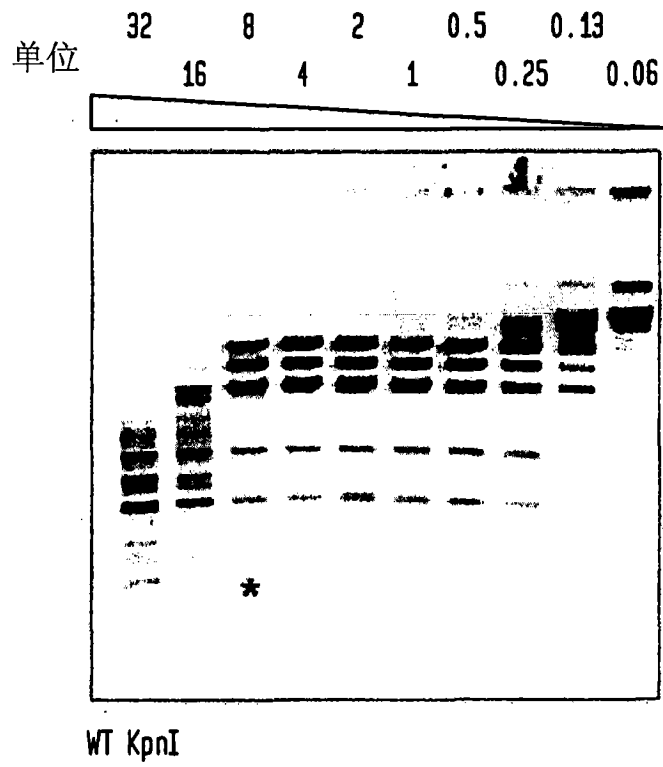


图 43A

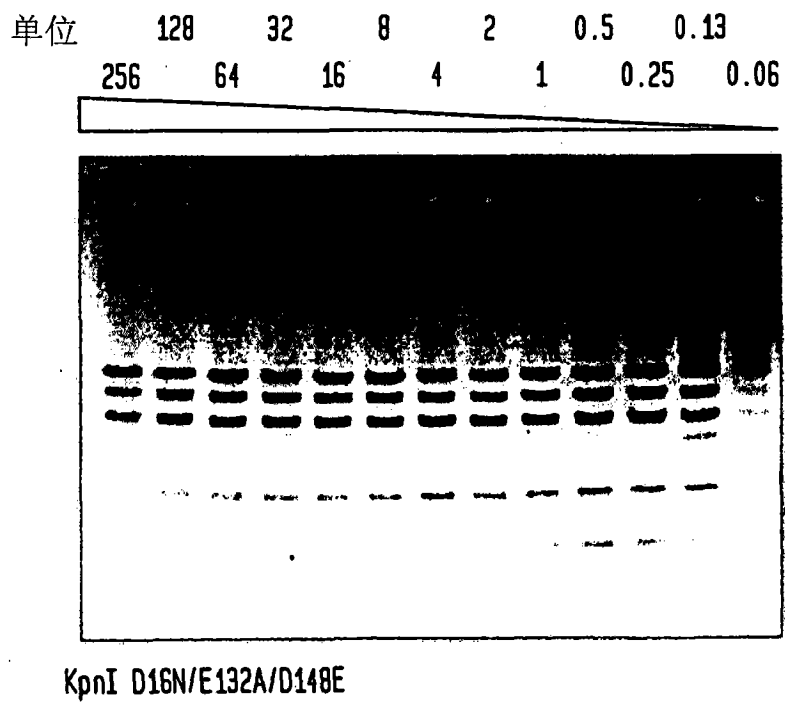


图 43B

>AgeI ACCGGT 272 aa (SEQ ID NO:79)

```
MRLDLDFGRG LVAHVMLDNY SEEQYQOISD YFVPLVKNPK LKSRDAIGQA
FVMATEVCPD ANPSDLWHHV LYRIYIREKI GTDPSQSWVR TSGEAFEVAL
VERYNPVLR HGI RLTALFK GOKGLALTRM GVADRVGSRK VDVHIEKGGG
GRSPDAEGFG VVGGIHAKVS LAERVSDDIP ASRIMMGEGL LSVLSTLDVK
SFPPPHGDLV NRGELGTPDR PSDKRNIEG HGDFSACFSY NLRTPPSNAT
TPSGRHIYVS ASLVRTTSSP TT
```

>AvrII CCTAGG 358 aa (SEQ ID NO:80)

```
MEEDLDLSEN IEAASAELTT LYQVAADAMK DYIEIYLALS KQSDGFSNIN
NLDLTSRNR L VVIHGLSLE LDPDTSTPEE IKREAERMLA IALDTESAIT
AGVYEKMLRF ASSLVQDLFE QDELNSLSS EYLSANPGFL PFFQQLAGLR
SKSELKREVG NASDNSTSKA VAERILERI RNLRIRTFSK EKLLQAVEPT
LEGIVRDLVG KVLLENIVAD ALSDLQVPM RESEYQSLKG VIYDFRADFV
IPDAQPIAF IEVRKSSTRH ASLYAKDKMF SAINWKGNK RLLGILVVEG
PWTRETLRVM ANVFDYVTP L TRVSQVAEAI RAYLDGDKTR LKWL VNF SIE
EADHONIT
```

>AvrII CCTAGG 1077 nt (SEQ ID NO:110)

```
ATGGAAGAAG ACCTTGATTT ATCTGAAAAT ATCGAAGCTG CATCTGCGGA
GCTTACGACT CTTTATCAGG TAGCTGCTGA TGCTATGAAA GATTATATTG
AAATCTATCT TGCGCTGAGT AAACAGTCTG ATGGGTTTTT AAATATTAAC
AATCTTGACT TAAC TTCTCG TAACAGGCGT TTGGTAGTTA TACATGGACT
TTCGTTAGAG TTAGATCCAG A TACTTCGAC TCCAGAGGAA ATTAACGTTG
AAGCTGAACG AATGCTAGCG ATAGCTCTTG ATACAGAGTC AGCAATTACG
GCAGGAGTAT ATGAAAAAAT GCGTCTCTTC GCAAGCTCTT TAGTAGATCA
GCTATTTGAA CAAACGGATG AACTTAATTC ATTATCATCG GAATATTTGT
CAGCAAATCC AGGATTTTTG CCGTTTTTCC AGCAGTTGGC GGGGCTTAGA
AGTAAATCAG AGTTAAAGAG AGAAGTAGGA AATGCCCTCG ACAATAGTAT
TTCTAAAGCG GTTGACAGAGA GAATATTAGA GCGCATTATA CGTAACTTGA
GAATTCGCAC TTTTCCAAA GAGAACTAT TACAAGCTGT TGAGCCTACT
TTAGAAGGAA TAGTCAGGGA TCTCGTAGGA AAAGTGTAT TGGAAAATAT
AGTTGCTGAT GCTTTATCTG ATTTACAAGT TCCTTTCATG CGTGAATCAG
AGTATCAAAG CCTTAAAGGA GTGATTTATG ATTTCCGCGC TGATTTTGTG
ATACCAGACG CACAAAATCC AATTGCTTTT ATCGAGGTGC GAAAAAGCTC
TACACGACAT GCGTCACTCT ATGCCAAGGA TAAGATGTTT TCAGCGATTA
ATTGGAAGG AAAAAATAAA AGGCTTTTGG GTATTTTGGT TGTGGAAGGA
CCTTGGACAA GAGAACTCT TCGCGTCATG GCAAAATGTG TTGATTACGT
TACACCTTTA ACTCGTGT T CCAAGTTGC AGAAGCTATC AGAGCATATC
TAGATGGGGA TAAAACGAGA CTGAAGTGGT TAGTTAATTT CAGTATTGAA
GAAGCAGACC ACGACAACAT AACCTAA
```

图 44A

>BsmBI CGTCTC 530 aa (SEQ ID NO:81)

```

MAKYGRGKFL PHQNYIDYMH FIVNHKNYSG MPNAIGEDGR INWOVSSGKT
TSFYEYYQAR FEWWEKKADE LNLPGTGNNS KRFSLAARLI HPTGORPCRL
CGKYQYVGYM YVSHNLYKRW SKITGREDLF FKKONIEAA NIFKSIMGEQ
ALINELTTIF PERKDYFNRL PNIEDFFVSS SHIKNNGNYI SPGFMANPPD
RLDGFHDYGI CCRKEKDPGR HDDNMRLYNH DRRAFMWSE GDWALADALY
NKAGAGKCAD PDCOKEVEKI SPDHVGPISC GFKQIPFFKP LCASCNSAKN
RRFSYQDVKE LLKYENYTG DSVASWQVRAL WDNCKHLVKN DDDSKLLSNL
MRSLODYLLR SLYKLFNGF AHLLSYFLTP EYAHYKITFE GLNTSTLEYE
RYYKTFKKT STSSLAARIV RIAFELEIY NSKDINERKL IKFDTSSWEK
DFENIISYAT KNLSLDEEAS KWNKVLTKN LSSTEKOKKI SSLLEDKNYE
VYKKQFYILK DLLVEHFNKI GEOIAKOYMK

```

>BspQI GCTCTC 430 aa (SEQ ID NO:16)

```

1 MRRLAKNSRN DSYLSNRDYO EIVRENTTI SFPLKEKHTL TLTKKIGLNO 50
51 TAGFGGWFFP DSPCLLTVTV LSSFGTKVTS KTFSLSKDWN RVGLAWINEH 100
101 SSDTMSIVLE FSDVEIVHTW GLTCDVFNH ELIIDAIEDO NKLIDVNLQE 150
151 HLSPETYLYN HSDTDLIEN LESTEEKIV NQSQKQISLK KCCYCQRYMP 200
201 VNILVRSNSS FHKHKSCKTG FONECRACK WRINNSFNPV RTKDQLHESA 250
251 VITREKILL KEPEILOKIK NRNNGEGLKS IWKKFDKKC FNCEKELTIE 300
301 EVRLDHRPL AYLWPIDEHA TCLCEKCNNT KDMFPIDFY QGDEDKLRL 350
351 ARITGLDYES LVKRDVNEVE LARIINNIED FATNVEARTF RSIRNKVKEV 400
401 RPDIDLFEIL KSKNINLYNE LOYELLTRKD 430

```

>BspQI GCTCTC 1293 nt (SEQ ID NO:101)

```

ATGAGACGAT TAGCAAAAAA TTCACGGAAC GACAGTTATT TAAGTAATG
GGATTACCAG GAAATCGTGA GGGAAAATAC CACTACAATA TCGTTCCCT
TAAAAGAAAA ACATACTCTG ACTTTAACGA AAAAAATAGG GCTAAATCAG
ACTGCTGGAT TCGGAGGATG GTTTTCCCT GATTCACCAT GTTTATTAAC
AGTAACTGTA CTATCCTCTT TCGGTACAAA GGTAACCTCT AAAACCTTTA
GCCTTTCTAA AGATTGGAAT CGTGTGGGC TTGCTTGGAT TAACGAGCAT
TCGAGTGACA CCATAAGCAT TGTCTAGAG TTTAGTGATG TGGAAATAGT
TCATACATGG GGACTTACAT GTGATGTTT TAATGTCCAT GAATTAATTA
TTGATGCTAT AGAAGATCAA AATAAACTAA TAGACGTGCT AAATCAAGAA
CATTTATCTC CTGAAACATA TTATTTAAAC CATGACTCTG ATACTGATTT
AATTGAGAAT TTGGAATCTA CAGAAGAGAT AAAGATAGTT AACCAAAGCC
AAAAGCAAAT CTCTTTAAAA AAATGCTGTT ATTGTCAACG TTATATGCCT
GTGAACATAT TAGTTCGTTT AAATTCATCA TTTCATAAAC ACAAGAGTAA
GAAAACGGT TTTCAAATG AATGTCGGC TTGTAAGAAG TGGAGAATAA
ATAATTCATT CAATCCAGTC AGAACAAAAG ACCAACTACA TGAATCAGCA
GTTATTACAC GTGAAAAAAA AATATTACTT AAAGAACCCTG AAATATTACA
GAAAATCAAA AATAGAAATA ACGGTGAGGG CTAAAAAAGT ATTATATGGA

```

图 44B

```

AAAAATTTGA TAAAAATGC TTTAATTGTG AAAAGAATT AACCATGAA
GAGGTACGCC TAGACCATAC AAGACCACTT GCTTATCTGT GGCCTATCGA
TGAACACGCA ACTTGTTTAT GTGAAAAATG CAACAATACA AAACATGATA
TGTTTCCTAT CGATTTTTAT CAAGGGGACG AAGACAAATT AAGACGTTTA
GCTAGAATTA CGGGGTTAGA TTATGAATCT CTAGTTAAGA GGGACGTAAA
TGAAGTTGAA CTTGCAAGAA TAATCAATAA CATTGAAGAC TTTGCAACTA
ATGTAGAGGC ACGTACTTTT CGCTCAATAA GAAATAAAGT AAAAGAAGTA
CGTCCCGATA CTGACCTATT TGAATTCTT AAATCTAAAA ATATTAATTT
ATATAATGAA CTTCAATATG AACTTCTTAC CCGTAAGGAT TAA

```

>EagI CGGCCG 301 aa (SEQ ID NO:82)

```

MKKRRDLVEV FGYNPMDLSP EVRALWNLGA CPFLNKECIK INHDQTIYIG
TCSVTSPLYGD VIICPNRLYA NOYETLHKVS RDAFGDDVFP LTYSNFIKYR
ATYKDCIVAL GKNSGKEVQV GRALSMDWL VRITDGELKE YVGVEIQSID
ITGNYRDAWH AYKNLKPIDI IDNLPTSQHG LNWANVHKRL IPQIIRKGVV
YSRSNYVKKG LYFILPEIVY NKFEDVIGAD IPLKLTQTNK SITVHTYSLG
EPAANGEQRK LISEREIIFD LDEFKRFIT GPNLPGDDL DAVIKKALGM
M

```

>EcoRI GAATTC 277 aa (SEQ ID NO:83)

```

MSNKKQSNRL TEQHLSQGV IGIFGDYAKA HDLAVGEVSK LVKKALSNEY
PQLSFRYRDS IKKTEINEAL KKIDPDLGGT LFVSNSSIKP DGGIVEVKDD
YGEWRVVLVA EAKHGKDI IIRNGLLVGK RGDQDLMAAG NAIERSHKNI
SEIANFMLSE SHFPYVLFLE GSNFLTENIS ITRPDGRVNV LEYNSGILNR
LDRLTAANYG MPINSNLCIN KFMVHKDKSI MLOAASIYTO GDGREWDSKI
MFEIMFDIST TSLRVLGRDL FEQLTSK

```

>EcoRV GATATC 245 aa (SEQ ID NO:84)

```

MSLRSDLINA LYDENQKYDV CGIISAEGKI YPLGSDTKVL STIFELFSRP
IINKIAEKHG YIVEEPKQON HYPDFTLYKP SEPNNKIAID IKTTYTNKEN
EKIKFTLGGY TSFIRNNTKN IVYPFDQYIA HWIIGYVYTR VATRKSSLKT
YNINELNEIP KPYKGVKVFQ QDKWVIAGDL AGSGNTTNIQ SIHAHYKDFV
EGKGIFDSED EFLDYWRNYE RTSQLRNDKY NNISEYRNWI YRGRK

```

>HindIII AAGCTT 300 aa (SEQ ID NO:85)

```

MKKSALEKLL SLIENLTNOE FKOATNSLIS FIYKLNREVE IELVRSIGIL
PEAIKPSSTQ EKLFKAGDI VLAKAFQLLN LNSKPLEQRG NAGDVIALSK
EFNYGLVADA KSFRLSRTAK NQKDFKVKAL SEWREDKDYA VLTAPFFQYP
TTKSQIFKQS LDENVLLFSW EHLAILLQLD LEETNIFPFE QLWNFPKKQS
KKTSVSDAEN NFMDFNKYF MDLFKIDKDT LNQLLOKEIN FIEERSLIEK
EYWKQINII KNFTREEAIE ALLKDIMSS KIETIDSFIK GIKSNORLYL

```

图 44C

>HpaI GTTAAC 254 aa (SEQ ID NO:86)

```
MKYEEINFKV PVESPYYPNY SQCVIERIYS ILRNQKDMGD DRIINTNLK
KGLPLENINK IAGPMIEAWA EEVFSGIRDN RDNQYNLINV EAQERLGISD
IILQFQVNNN VITGNVDVKA TSNDIPDSGK SPNITSFSRI RTAYVKDPNF
IFIILSIKHS VYVKRNEYTN LMDGIMQIID FNVYDLKYIS DSDISYNPAL
GTGQIQIKDI HYVSSQKRTT WQMCQLLDLK YLRSKKRITIE QFYNEAKRINK
WIKD
```

>KpnI GGTACC 218 aa (SEQ ID NO:87)

```
MDVFDKVVYS DNNSYDQKTV SORIEALFLN NLGKVVTRQQ IIRAATDPKT
GKQOPENWHQR LSELRTDKGY TILSWRDMKV LAPQEYIMPH ATRRPKAAKR
VLPTKETWEQ VLDRANYSCE WQEDGQHCGL VEGDIDPIGG GTVKLTPDHM
TPHSIDPATD VNDPKMWOAL CGRHQVMKKN YWDSNNGKIN VIGILQSVNE
KQKNDALEFL LNYYGLKR
```

>NcoI CCATGG 288 aa (SEQ ID NO:88)

```
MATAPGHLG QIIGNVMEEA LKPVLOEMAD RHDLYLDSKG LPPGVRSAL
VTWDDLGNN HOLDFVLERG GSATKAGNPA AFIEAAWRRY TKHSAKAKAE
IQGAVLPVLA AWNNVKPTPA AVVAGQWTAP SLOQMRSNGF VVLHLHFPTT
AQVFGGNGIN IEGTGEGTPD AFWQQQDAY TSKSEADKDS LATALRTAHA
QEFRTFVAEL ERRVVRAIDY VVVTPLHGKG SQYTSIENAI EAVRTYSCGE
ESAPFLRFEI RISYNGDVI QATFGSSSDA IEFLDTFN
```

>NheI GCTAGC 328 aa (SEQ ID NO:89)

```
MSSYHDDLNI LNVDFNHLRL TELIKLADQA EPFYLWVEKI FRQVSGRADS
LETIIEVEER VVLKMAILTC FTSDEKELPK LFNGVGVYPY HIKACYFFFA
WLVRDAATOR LDPLIREAFT QLKSIHPQMK KTELESEIFS QLLVNYRNL
IHFSWPVIRE VLISRLEGSR RAARGSYLEL FVRTALAQSI TYFYKIYGNV
GKFLDVKIHD KPLKVKNRTY DVVAELIGNN HNTQYLILPV KTRETQGGGH
AHLFTRDIEQ SNNDIRELYP NAVIAPVIA ENWSDTEKDL ENVGYNDFH
FSVNPNRFAG FSDVEQIRLN RLVERILL
```

图 44D

>NotI GCGGCCGC 383 aa (SEQ ID NO:90)

```
MRSOTSVEPE GANFIAEFFG HRVYPEVVST EAARNDQATG TCPFLTAAKL
VETSCVKAET SRGVCVVNTA VDNERYDWLV CPNRALDPLF MSAASRKLFQ
YGPTEPLQFI AAPTLDQAV RDGIREWLDL GVHVVAIFYE KLGELSSK
TDSSPEFSFD WTLAEVESIY PVPKIKRYGV LEIOTMDFHG SYKHAVGAID
IALVEGIDFH GWLPTPAGRA ALSKKMEGPN LSNVFKRTFY QMAYKFALSG
HRCAGTGFA IPQSVWKSLL RHLANPTLID NGDGTFSLGD TRNDSENAWI
FVFELDPDTE ASPRPLAPHL ETRVNVDTLI DLALRESPRA ALGPSGPVAT
FTDKVEARML RFWPKTRRRR STTPGGQRGL FDA
```

>PstI CTGCAG 326 aa (SEQ ID NO:91)

```
MKELKLKEAK EILKALGLPP QQYNDRSGWV LLALANIKPE DSWKEAKAPL
LPTVSIMEFI RTEYGKDYKP NSRETIIRROT LHOFEQARIV DRNRDLPSRA
TNSKDNMNSL NQVIIDILHN YPNGNWKELI QOFLTHVPSL QELYERALAR
DRIPIKLLDG TQISLSPGEH NQLHADIVHE FCPRFVGDHG KILYIGDTAS
SRNEGGKLMV LDSEYLKKG VPPMSHDKLP DVVVYDEKPK WFLIEAVTS
HGPISPKRWL ELEAALSSCT VGKVYVTAFP TRTEFRKNAA NIAWETEVI
ADNPDHMHVF NGDRFLGPHD KKPELS
```

>PvuII CAGCTG 157 aa (SEQ ID NO:92)

```
MSPDLNKLK ELWPHIQEYO DLALKHGIND IFQDNGGKLL QVLLITGLTV
LPGREGNDV DNAGQEYELK SINIDLTKGF STHHMHNPVI IAKYRQVPWI
FAIYRGIAIE AIYRLEPKDC EFYYDKWERK WYSDGHKDIN NPKIPVKYVM
EHGKTIY
```

图 44E

>SacI GAGCTC 358 aa (SEQ ID NO:93)

```

MGITIKKSTA EQVLRKAYEA AASDDVFLED WIFLATSRE VDAPRTYTAA
LVTALLARAC DDRVDPRSIK EKYDDRAFSL RTLCHGVVVP MSVELGFDLG
ATGREPINNO PFFRYDQYSE IVRVQTKARP YLDRVSSALA RVDEEDYSTE
ESFRALVAVL AVCISVANKK QRVAVGSATV EASLIAETQS FVVSGHDVPR
KLQACVAAGL DMVYSEVYSR RINDPSRDFP GDVQVILDGO PLLTVEVRGK
SVSWEGLEQF VSSATYAGFR RVALMVDAAAS HVSLMSADDL TSALERKYEC
IVKVNESVSS FLRDVFWWSP RDVHSILSAF PEAMYRRMIE IEVREPELDR
WAEIFPET

```

>SalI GTCGAC 315 aa (SEQ ID NO:94)

```

MINADKPHRW NDDVQASVRL YNQWFLDAAP KAYRDTRQLT IDEVEQAFQR
TANMITSITPE VLKAHPKTLA TLRMSTAPPI ARDRLVGLSH GSKSLDIME
KGKLPKPMKG DVLDTHLAKM CAVLTDLDDL DLFWYPTGE PAEPRQRELA
ATVVADRLCG AIADPIVRNA QERRQLALIE EWLLARGYK KTHSASLPLN
TMQPGTFSFR QNVVVGSDLP VNIPVDAVIO PHTPHSHKLP ILIEAKSAGD
FTNTNKRRKE EATKIHLQLL KYGNEISLTL FLCGYFNTGY LGYSAEGLD
WWEHRIDDL EAAGA

```

>SapI GCTCTC 432 aa (SEQ ID NO:95)

```

MRRLATORRE DAYKSNRDYO TVHEAQLRV NSTDDNLSL FLLKDISPRE
DSKNIVGFGG FVKPEIATM ALTLTTDIDK QIKSVPLSSN WNRISIVAKF
ASNPSVSITL GFDQTPWVDF WGINSDDIGL SFVSDAVPLE MSMIDSIHIA
PETLYLDHSS ACLLDIDPVE STRFKTGHD PLSLKKCSYC GRLLPIDLER
PGKLSFHKHR AKITNHONEC RSCCKWRINN SFNPMRTIQ LNESALITRE
RKIFLOEPEI LOEIKDRTGA GLKSQVWERF HRKCFNCRKD LKLSEVOLDH
TRPLAYLWPI DEHATCLCAQ CNNTKKDRFP VDFYSEQQIR ELSDICGLPY
QDLCARSLNL DQLDRIERNI AEFKWEQVVR TFASTARRIS EVYPARDLFE
TLKKESESAY NKIEIKLKER PDALLDEALP LD

```

>SbfI CCTGCAGG 323 aa (SEQ ID NO:96)

```

MNSSDGIDGT VASIDTARAL LKRFGFDAQR YNVRSAVTLL ALAGLKPGRD
WVDSTTPRLG VQKIMDWSGE HWAKPYATGS REDFRKKTLL QWVDNGFAVL
NADNLNIATN SOLNEYCLSD EALQALRAYG TEGFEESLVV FLDEASKAVK
ARAEALQAAM ISVDLPGGEE FLLSPAGQNP LLKKMVEEFV PRFAPRSTVL
YLGDRGKHS LFEREIFEEV LGLTFDPHGR MPDLILHDEV RGWFLMEAV
KSKGPFDEER HRSLOELFVT PSAGLIFVNC FENRESMRQW LPELAWETEA
WVAEDPDHLI HLNCSRFLGP YER

```

图 44F

>ScaI AGTACT 227 aa (SEQ ID NO:97)

MINDQLPRW REARVGTRTG GPAMRPKTSQ SPYFGWQSED WPEVTRQLLS
EQPLSGDTLV DAVLASWESI FESRLGSGFH IGTQIRPTQ IMGFLHALI
PLELANGDPS WRADLNSSEK DLVYQPDHXY SIEMKTSSHK DQIFGNRSFG
VENPGKGGKA KGGYYVAVNF EKWSADAPGRL PRIRTIRYGW LDHTDWVAQK
SQTGQQSSLP AVVSNTQLLA IHTGGQR

>SphI GCATGC 235 aa (SEQ ID NO:98)

MTSKDPIVLS ADQIAWLRQL KMSKRAALVR DYILEYGAVT TGKLAELGYS
HPPRAARDLK DAGAGVVTIM YKGPDGRRMA SYAFNGKANE DGAGRVIIPK
AFGEALKRAH GGGKAVCYGD FSERELQCDH RVPFAIAGDK PKLVQEDFMP
LCASDNRAKS WSCENCNPWE LKDEDTCRSC FWASPENYTH VSTRPERRIN
LLFQGDVEI FDALKNAAAN EGVSLTEATK RKLAD

>SspI AATATT 281 aa (SEQ ID NO:99)

MSKAAYQDFT KRFSLLIKKH PNLITMTLSN IFTMRLIGNK THGDIAEIAI
SEFINQYMYD FKSIVHGKDL YRAKSKEEDI TVENEITKEK FPISLKAYGD
GPLQLSTDKN FLMYPLLEEI GAFINAKKI EEIFANEAFS CFSEINVLPL
IYDEKRQRGN ILVFDAARAR AETAYIRKET EGSGRKHAY RFFDKNKNYI
CEVRYGNAAA NALQRGLWTN TKNATSFDS VTNGWVDYSH NLVLVKLLSH
ALVSSRKGHE AALEEIKKDI LQKQTNGIN V

图 44G