



(12) 发明专利

(10) 授权公告号 CN 102869778 B

(45) 授权公告日 2015. 05. 20

(21) 申请号 201180018149. 7

C07K 14/00(2006. 01)

(22) 申请日 2011. 03. 28

C07K 19/00(2006. 01)

C12R 1/39(2006. 01)

(30) 优先权数据

61/319, 152 2010. 03. 30 US

PCT/US2010/030573 2010. 04. 09 US

61/325, 235 2010. 04. 16 US

(85) PCT国际申请进入国家阶段日

2012. 10. 08

(86) PCT国际申请的申请数据

PCT/US2011/030227 2011. 03. 28

(87) PCT国际申请的公布数据

W02011/126811 EN 2011. 10. 13

(73) 专利权人 菲尼克斯公司

地址 美国加利福尼亚州

(72) 发明人 D·M·雷奥尔阿克 L·丘 H·金

H·W·塔尔伯特

(74) 专利代理机构 北京市金杜律师事务所

11256

代理人 陈文平 徐志明

(51) Int. Cl.

C12N 15/78(2006. 01)

C12N 15/31(2006. 01)

(56) 对比文件

US 2008269070 A1, 2008. 10. 30,

CN 101641441 A, 2010. 02. 03,

CN 101031655 A, 2007. 09. 05,

Guilin Yang 等. Expression of recombinant Clostridium difficile toxin A and B in Bacillus megaterium. 《BMC Microbiology》. 2008, 第 8 卷 (第 192 期),

审查员 白鸽

权利要求书1页 说明书67页

序列表53页 附图32页

(54) 发明名称

重组毒素蛋白的高水平表达

(57) 摘要

本发明涉及在细菌宿主中生产重组毒素蛋白的领域。本发明尤其涉及从细菌宿主中获得高水平重组 CRM 197、白喉毒素、百日咳毒素、破伤风类毒素 C 片段、霍乱毒素 B、霍乱全毒素和假单胞菌外毒素 A 的生产方法。



1. 一种在假单胞菌宿主细胞中生产重组毒素蛋白的方法,所述方法包括:
将编码毒素蛋白的核苷酸序列连接到表达载体中;
用该表达载体转化假单胞菌宿主细胞;以及
在适合重组毒素蛋白表达的培养基中培养已转化的假单胞菌宿主细胞;
其中所述重组毒素蛋白为艰难梭菌毒素 B,并且所述宿主细胞:在 HtpX 的表达上有缺陷;在 DegP1 的表达上有缺陷;在 HslU、HslV、Prc1 和 Prc2 的表达上有缺陷;在 Lon、Ia 和 DegP2 的表达上有缺陷;或者所述宿主细胞在 Lon、Prc1、DegP2、AprA 的表达上有缺陷并且过表达 DegP2S219A;和
其中所述重组蛋白以 0.2 克/升至 12 克/升的可溶性和/或活性毒素蛋白的产率生产。
2. 权利要求 1 的方法,其中可溶性和/或活性毒素蛋白的产率为 0.5g/L 至 12g/L。
3. 权利要求 1 或 2 的方法,其中所述编码毒素蛋白的核苷酸序列与分泌信号编码序列融合,当所述分泌信号编码序列表达时,指引所述毒素蛋白向周质转移。
4. 权利要求 1 或 2 的方法,其中所述宿主细胞在至少一种蛋白酶的表达上有缺陷,或者其中所述宿主细胞过表达至少一种折叠调节因子,或者两者的组合。
5. 权利要求 1 或 2 的方法,其中所述重组毒素蛋白与它的天然分泌前导序列融合。
6. 权利要求 1 或 2 的方法,进一步包括在活性测定中测定所述重组毒素蛋白的活性,其中所生产的可溶性毒素蛋白中 40%至 100%经确定具有活性。
7. 权利要求 6 的方法,其中所述活性测定为免疫学测定、受体结合测定、细胞毒性测定或酶测定。
8. 权利要求 1 或 2 的方法,其中所述表达载体包含可操作地连接到蛋白质编码序列上的 lac 衍生启动子,而且其中所述培养包括用浓度为 0.02 至 1.0mM 的 IPTG 诱导该启动子,诱导时的细胞密度为 40-200 吸光度单位(AU)的光密度,培养物的 pH 为 6 至 7.5,而且生长温度为 20℃至 35℃。
9. 权利要求 1 或 2 的方法,其中所述宿主细胞是假单胞菌属细胞。
10. 权利要求 1 或 2 的方法,其中所述宿主细胞是荧光假单胞菌。
11. 权利要求 1 或 2 的方法,其中所述核苷酸序列已为了在假单胞菌宿主细胞中表达而进行了优化。
12. 权利要求 9 的方法,其中所述核苷酸序列已为了在假单胞菌属宿主细胞中表达而进行了优化。
13. 权利要求 10 的方法,其中所述核苷酸序列已为了在荧光假单胞菌宿主细胞中表达而进行了优化。
14. 权利要求 3 的方法,其中所述表达载体进一步包含邻近所述分泌信号的编码序列的标签序列。
15. 权利要求 1 或 2 的方法,其中所述表达载体进一步包含邻近所述毒素蛋白的编码序列的标签序列。

重组毒素蛋白的高水平表达

[0001] 优先权要求

[0002] 本申请根据 35 U. S. C. §119(e) 要求 2010 年 4 月 16 日提交的美国临时申请系列号 61/325,235、2010 年 4 月 9 日提交的 PCT/US 10/30573 和 2010 年 3 月 30 日提交的美国临时申请系列号 61/319,152 的优先权。这些申请的内容特此通过引用整体并入本文。

[0003] 序列表

[0004] 本申请包含通过 EFS-Web 以 ASCII 格式提交的序列表,该序列表特此通过引用整体并入本文。创建于 2011 年 3 月 16 日的所述 ASCII 拷贝被命名为 38194201.txt,其大小为 156,975 个字节。

[0005] 发明背景

[0006] 微生物毒素蛋白在医学中用作针对产生毒素的微生物的疫苗接种的免疫原以及用作其他疫苗的载体蛋白质和佐剂,并且在科学研究中用作研究分子途径的工具。

[0007] 白喉毒素 (DT) 是由白喉杆菌 (*Corynebacterium diphtheriae*) 的产毒菌株合成和分泌的蛋白质性毒素。产毒菌株含有携带毒素基因的噬菌体溶素原。DT 被合成为 535 个氨基酸的多肽,其经历蛋白水解以形成成熟的毒素。成熟的毒素包含由二硫键连接的两个亚基 A 和 B。由完整 DT 的 C 末端部分形成的 B 亚基使 DT 能够结合细胞膜并通过细胞膜进入到胞质中。一旦进入细胞,由完整 DT 的 N 末端部分形成的酶的 A 亚基催化延伸因子 2 (EF-2) 的 ADP 核糖基化。其结果是,EF-2 被灭活、蛋白质合成停止和细胞死亡。白喉毒素是高度细胞毒性的,单个分子可以致死细胞,且 10ng/kg 的剂量可以杀死动物和人。

[0008] CRM197 蛋白是 DT 的无毒的、免疫学交叉反应的形式。已研究了其作为 DT 增强剂或疫苗抗原的潜在用途。CRM197 是通过由产毒的 β 棒状杆菌噬菌体的亚硝基胍诱变构建的非产毒噬菌体 β 197^{tox-}感染的白喉杆菌产生的。CRM197 蛋白具有与 DT 相同的分子量,但在 A 亚基中有单碱基变化 (鸟嘌呤变为腺嘌呤) 的差异。这一单碱基变化导致氨基酸置换 (谷氨酸置换为甘氨酸),并消除了 DT 的毒性性质。

[0009] 使用 CRM197 作为载体蛋白的结合多糖疫苗已批准人体使用。疫苗包括:Menveo[®] (Novartis Vaccines and Diagnostics) (标明用于预防由脑膜炎球菌亚群 A、C、Y 和 W-135 引起的侵袭性脑膜炎球菌疾病的疫苗)、Menjugate (Novartis Vaccines) (脑膜炎双球菌 C 群结合疫苗)、和 Prevnar[®] (Wyeth Pharmaceuticals, Inc.) (靶向于肺炎链球菌的七种血清型的儿童肺炎疫苗) 和 HibTITER[®] (Wyeth) (流感嗜血杆菌 b 型疫苗)。此外,CRM197 具有用作白喉杆菌疫苗接种的增强抗原的潜力并正在研究作为用于其它疫苗中的载体蛋白。

[0010] 用于批准的治疗和研究用途的 CRM197 高水平表达的方法尚未见报道。CRM197 已在例如白喉杆菌、枯草芽孢杆菌和大肠杆菌中以数十 mg/L 的水平表达。单剂 Prevnar 结合疫苗含有大约 20 μ g CRM197。因此,用于经济地以约 1g/L 或以上的水平生产 CRM197 的方法将极大地促进疫苗研究和生产。

[0011] 由霍乱弧菌 (*Vibrio cholera*) 产生的霍乱毒素 (CTX) 是一种导致以腹泻和呕吐为特征的感染的细菌性病原体,也是一种 ADP-核糖基化毒素。CTX 是由六个蛋白质亚基组

成的低聚复合物：单一拷贝的霍乱毒素 A 亚基 (CTA) 和五个拷贝的霍乱毒素 B 亚基 (CTB)。这五个各重 12kDa 的 B 亚基形成了一个五元环。A 亚基具有 A1 部分 CTA1 和 A2 链 CTA2，CTA1 是将 G 蛋白进行 ADP-核糖基化的球状酶，而 CTA2 形成紧靠地位于 B 亚基环中心孔中的伸展的 α 螺旋。该环结合到宿主细胞表面上的 GM1 神经节苷脂受体上，导致整个复合物的内化。一旦内化，CTA1 链通过二硫键的还原释放。然后 CTA1 被激活并且催化腺苷酸环化酶的 ADP 核糖基化。由此导致的腺苷酸环化酶活性的增强提高了环 AMP 的合成，这导致大量的流体和电解质流出并且导致腹泻。

[0012] CTX 的 B 亚基尽管是相对无害的，但仍保留了其结合到 GM1 神经节苷脂受体上的能力。因此，CTB 可用于促进化学地或遗传地偶联的外来抗原的粘膜摄取。已证实其诱导粘膜免疫和系统免疫，并且是用于可食用疫苗的生产的候选物。由于其结合偏好，CTB 也可用作神经元示踪物。

[0013] 百日咳毒素 (PTX) 是由导致百日咳病的人呼吸道细菌性病原体百日咳博德特氏菌 (*Bordetella pertussis*) 产生的外毒素和毒力因子。百日咳全毒素是具有 AB₅ 结构的多亚基复合物。酶活性的 A 亚基 (S1) 是修饰哺乳动物细胞内几种异三聚体 G 蛋白的 α 亚基的 ADP-核糖基转移酶，而 B 低聚物 (S2、S3、两个拷贝的 S4 和 S5) 结合细胞上的糖结合物受体。该毒素的五个亚基由百日咳类毒素操纵子表达。

[0014] 已研究了用于保护性疫苗和用作疫苗佐剂的百日咳毒素的无毒变异体。也需要百日咳毒素蛋白来用于研究，例如用于对 G 蛋白信号传导途径的研究。

[0015] 由破伤风梭菌 (*Clostridium tetani*) 产生的破伤风毒素是具有 150kDa 分子量的神经毒素。它由以下两部分组成：100kDa 重链或 B-链和 50kDa 轻链或 A-链。这些链通过二硫键连接。B 链结合到位于神经元膜上的二唾液酸神经节苷脂 (GD2 和 GD1b) 上。A 链 (一种锌内肽酶) 攻击囊泡相关膜蛋白 (VAMP)。

[0016] A 链的作用通过降解小突触泡蛋白阻止了受影响的神经元释放抑制性神经递质 GABA (γ -氨基丁酸) 和甘氨酸。其后果是最小刺激引起肌肉的危险的过度活动—感官刺激对运动反射的抑制失败。这导致主动肌和拮抗肌肌肉组织的全身性收缩，被称为强直性痉挛。

[0017] 破伤风毒素 C 片段 (Tet C 或 TTC) 是通过破伤风毒素的蛋白酶切割 (例如用木瓜蛋白酶) 或通过片段的重组表达而产生的 50kD 的多肽。它对应于 C-末端的 451 个氨基酸 (氨基酸位置 865-1315)。

[0018] C 片段已表明为无毒性的。因为它以高特异性和亲和力与神经元结合，TTC 可用作神经元药物递送的靶向分子或用于研究目的。TTC 蛋白也可潜在地用作疫苗载体蛋白并且在疫苗中用来避免破伤风梭菌感染。

[0019] 艰难梭菌 (*Clostridium difficile*) 毒素 B (TcdB) 是由艰难梭菌产生的毒力因子，它导致医院获得性腹泻和假膜性结肠炎。TcdB 和第二大梭菌毒素 TcdA 参与了假膜性结肠炎的发生。

[0020] TcdB 是约 270kD 的葡糖基化毒素，且可分成酶、易位和受体结合域。TcdB 的前 546 个氨基酸包含酶区域，接着是推定的易位和受体结合域。TcdB 具有作为艰难梭菌感染的保护性疫苗以及用于诊断检验及其开发的潜在用途。

[0021] 铜绿假单胞菌 (*Pseudomonas aeruginosa*) 的外毒素 A (ETA 或 PE) 是一种 II 型

ADPRT。就像它的家族成员白喉毒素和霍乱毒素一样,它通过细胞延伸因子 2 的 ADP- 核糖基化来抑制蛋白质合成。铜绿假单胞菌外毒素 A 作为单体存在,由 613 个氨基酸的单一多肽链 (66kd) 组成。

[0022] ETA 可潜在地用作疫苗偶联物。ETA 的无毒性突变体已作为用于防御金黄色葡萄球菌 (*Staphylococcus aureus*)、疟疾和伤寒沙门氏菌 (*Salmonella Typhi*) 的疫苗接种的疫苗偶联物进行了研究。

[0023] 以足够满足不断扩大的需求的量来生产这些毒素已成为重大的挑战。当在常规蛋白质过表达系统中生产时,由于降解、不正确的折叠或这两者,取决于毒素的具体特征,例如大小和二级结构,毒素蛋白仅以非常低的浓度以活性形式回收。因此,需要以低成本生产大量可溶性和 / 或活性形式的这些毒素的方法。

发明内容

[0024] 本发明涉及在假单胞菌宿主细胞中生产重组毒素蛋白的方法,所述方法包括:将编码毒素蛋白的核苷酸序列接合到表达载体中;用该表达载体转化假单胞菌宿主细胞;和在适合重组毒素蛋白表达的培养基中培养已转化的假单胞菌宿主细胞;其中该重组毒素蛋白是 CRM197、白喉毒素、霍乱全毒素、霍乱毒素 B、百日咳毒素、破伤风毒素 C 片段、艰难梭菌毒素 B 或铜绿假单胞菌外毒素 A。

[0025] 在实施方式中,重组毒素蛋白是霍乱毒素 B、霍乱全毒素、百日咳毒素、破伤风毒素 C 片段、艰难梭菌毒素 B 或铜绿假单胞菌外毒素 A。

[0026] 在其他实施方式中,重组毒素蛋白是霍乱毒素 B、霍乱全毒素、百日咳毒素、破伤风毒素 C 片段或艰难梭菌毒素 B。

[0027] 在其他实施方式中,重组毒素蛋白是 CRM197、白喉毒素、霍乱全毒素、霍乱毒素 B、百日咳毒素、破伤风毒素 C 片段或艰难梭菌毒素 B。

[0028] 在某些实施方式中,重组蛋白以约 0.2 克 / 升至约 12 克 / 升的可溶性和 / 或活性毒素蛋白的产率生产。在特定实施方式中,可溶性和 / 或活性毒素蛋白的产率为约 0.2g/L、约 0.3g/L、约 0.4g/L、约 0.5g/L、约 0.6g/L、约 0.7g/L、约 0.8g/L、约 0.9g/L、约 1g/L、约 1.5g/L、约 2g/L、约 2.5g/L、约 3g/L、约 3.5g/L、约 4g/L、约 4.5g/L、约 5g/L、约 5.5g/L、约 6g/L、约 6.5g/L、约 7g/L、约 7.5g/L、约 8g/L、约 8.5g/L、约 9g/L、约 9.5g/L、约 10g/L、约 10.5g/L、约 11g/L、约 12g/L、约 0.2g/L- 约 0.5g/L、约 0.2g/L- 约 1g/L、约 0.2- 约 2g/L、约 0.3g/L- 约 0.6g/L、约 0.3g/L- 约 1g/L、约 0.3- 约 2g/L、约 0.4- 约 0.7g/L、约 0.4- 约 1g/L 约 0.4- 约 2g/L、约 0.4- 约 3g/L、约 0.5g/L- 约 1g/L、约 0.5g/L- 约 2g/L、约 0.5g/L- 约 3g/L、约 0.5g/L- 约 4g/L、约 0.5g/L- 约 5g/L、约 0.5g/L- 约 6g/L、约 0.5g/L- 约 7g/L、约 0.5g/L- 约 8g/L、约 0.5g/L- 约 9g/L、约 0.5g/L- 约 10g/L、约 0.5g/L- 约 11g/L、约 0.5g/L- 约 12g/L、约 1g/L- 约 2g/L、约 1g/L- 约 3g/L、约 1g/L- 约 4g/L、约 1g/L- 约 5g/L、约 1g/L- 约 6g/L、约 1g/L- 约 7g/L、约 1g/L- 约 8g/L、约 1g/L- 约 9g/L、约 1g/L- 约 10g/L、约 1g/L- 约 11g/L、约 1g/L- 约 12g/L、约 2g/L- 约 3g/L、约 2g/L- 约 4g/L、约 2g/L- 约 5g/L、约 2g/L- 约 6g/L、约 2g/L- 约 7g/L、约 2g/L- 约 8g/L、约 2g/L- 约 9g/L、约 2g/L- 约 10g/L、约 2g/L- 约 11g/L、约 2g/L- 约 12g/L、约 3g/L- 约 4g/L、约 3g/L- 约 5g/L、约 3g/L- 约 6g/L、约 3g/L- 约 7g/L、约 3g/L- 约 8g/L、约 3g/L- 约 9g/L、约 3g/L- 约 10g/L、约 3g/L- 约 11g/L

L、约 3g/L-约 12g/L、约 4g/L-约 5g/L、约 4g/L-约 6g/L、约 4g/L-约 7g/L、约 4g/L-约 8g/L、约 4g/L-约 9g/L、约 4g/L-约 10g/L、约 4g/L-约 11g/L、约 4g/L-约 12g/L、约 5g/L-约 6g/L、约 5g/L-约 7g/L、约 5g/L-约 8g/L、约 5g/L-约 9g/L、约 5g/L-约 10g/L、约 5g/L-约 11g/L、约 5g/L-约 12g/L、约 6g/L-约 7g/L、约 6g/L-约 8g/L、约 6g/L-约 9g/L、约 6g/L-约 10g/L、约 6g/L-约 11g/L、约 6g/L-约 12g/L、约 7g/L-约 8g/L、约 7g/L-约 9g/L、约 7g/L-约 10g/L、约 7g/L-约 11g/L、约 7g/L-约 12g/L、约 8g/L-约 9g/L、约 8g/L-约 10g/L、约 8g/L-约 11g/L、约 8g/L-约 12g/L、约 9g/L-约 10g/L、约 9g/L-约 11g/L、约 9g/L-约 12g/L、约 10g/L-约 11g/L、约 10g/L-约 12g/L 或约 11g/L-约 12g/L。

[0029] 在实施方式中,将编码毒素蛋白的核苷酸序列与分泌信号编码序列融合,分泌信号编码序列表达时指引毒素蛋白向周质转移。在实施方式中,宿主细胞在至少一种蛋白酶的表达上存在缺陷或者宿主细胞过表达至少一种折叠调节因子,或它们的组合。

[0030] 在实施方式中,重组毒素蛋白是 CRM197,和宿主细胞在 HslU、HslV、Prc1、DegP1、DegP2 和 AprA 的表达上存在缺陷。在相关实施方式中,重组毒素蛋白与分泌前导序列即 Azu、IbpS31 A、CupA2、PbpA20V 或 Pbp 融合。在实施方式中,重组毒素蛋白是 CRM197,且宿主细胞在 HslU 和 HslV 或 Prc1 或 DegP1 或 DegP2 或 AprA 的表达上存在缺陷。在特定实施方式中,重组毒素蛋白是 CRM197,且宿主细胞在沙雷氏菌溶素 (Serralysin)、HslU、HslV、Prc1、DegP1、DegP2 或 AprA 的表达上存在缺陷,或者宿主细胞过表达 DsbA、DsbB、DsbC 和 DsbD。在实施方式中,宿主细胞过表达 DsbA、DsbB、DsbC 和 DsbD,且重组毒素蛋白与分泌前导序列 Azu 融合。在实施方式中,宿主细胞在沙雷氏菌溶素的表达上存在缺陷,且重组毒素蛋白与分泌前导序列 Pbp 或 Azu 融合。在实施方式中,宿主细胞在 HslU 和 HslV 的表达上存在缺陷,且重组毒素蛋白与分泌前导序列 Pbp 或 Azu 融合。在实施方式中,重组毒素蛋白是 CRM197,且宿主细胞是野生型,其中重组毒素蛋白与分泌前导序列 Pbp 或 Azu 融合。在实施方式中,重组毒素蛋白是 CRM197 并且重组毒素蛋白与分泌前导序列 Azu、Pbp、IbpS31A、CupA2 或 PbpA20V 融合。

[0031] 在其他实施方式中,重组毒素蛋白是霍乱毒素 B,且宿主细胞在 Lon、La 和 AprA 的表达上存在缺陷,或者宿主细胞在 HslU、HslV、Prc1、DegP1、DegP2 和 AprA 的表达上存在缺陷。在相关实施方式中,宿主细胞在 Lon、La 和 AprA 的表达上存在缺陷,而且其中重组毒素蛋白与分泌前导序列 Pbp A20V 融合。

[0032] 在其他实施方式中,重组毒素蛋白是百日咳毒素 SI E129A R9K,且宿主细胞在以下的表达上存在缺陷:Lon、La 和 AprA ;GrpE、DnaK 和 DnaJ ;HtpX ;RXF01590 ;或 ppiB(RXF05345)。在相关实施方式中,重组毒素蛋白与它的天然分泌前导序列融合。

[0033] 在其他实施方式中,重组毒素蛋白是破伤风毒素 C,且宿主细胞在 HslU、HslV、Prc1、DegP1、DegP2 和 AprA 的表达上存在缺陷。在相关实施方式中,重组毒素蛋白与分泌前导序列 DsbC、Pbp A20V 或 CupA2 融合。

[0034] 在其他实施方式中,重组毒素蛋白是破伤风毒素 C,且宿主细胞在 Lon、La 和 AprA 的表达上存在缺陷。在相关实施方式中,重组毒素蛋白与分泌前导序列 DspA 融合。

[0035] 在其他实施方式中,重组毒素蛋白是破伤风毒素 C,且宿主细胞在 GrpE、DnaK 和 DnaJ 的表达上存在缺陷。在相关实施方式中,重组毒素蛋白与分泌前导序列 NikA 融合。

[0036] 在其他实施方式中,重组毒素蛋白是艰难杆菌毒素 B,且宿主细胞在以下的表达上

存在缺陷 :HtpX ;DegP1 ;HslU、HslV、Prc1 和 Prc2 ;或 Lon 和 DegP2,或者宿主细胞既在 Lon、Prc1、DegP2 和 AprA 的表达上存在缺陷又过表达 DegP2 S219A。

[0037] 在实施方式中,在活性测定中测定重组毒素蛋白的活性,其中所生产的可溶性毒素蛋白中约 40%至约 60%确定为有活性。在相关实施方式中,活性测定为免疫学测定、受体结合测定或酶测定。

[0038] 在本发明的实施方式中,表达载体包含可操作地连接到蛋白质编码序列上的 lac 衍生启动子,而且其中培养包括用浓度为约 0.02 至约 1.0mM 的 IPTG 诱导该启动子,诱导时细胞密度是约 40 至约 200 吸光度单位 (AU) 的光密度,培养物的 pH 值为约 6 至约 7.5,生长温度为约 20 至约 35 摄氏度。

[0039] 在实施方式中,宿主细胞是假单胞菌属细胞。在相关实施方式中,宿主细胞是荧光假单胞菌 (*Pseudomonas fluorescens*)。

[0040] 在本发明的实施方式中,为了在假单胞菌宿主细胞中表达而对核苷酸序列进行了优化。在相关实施方式中,为了在假单胞菌属宿主细胞中表达而对核苷酸序列进行了优化。在其他相关实施方式中,为了在荧光假单胞菌宿主细胞中表达而对核苷酸序列进行了优化。

[0041] 在实施方式中,百日咳毒素是野生型或 S1 E129A R9K。在实施方式中,铜绿假单胞菌外毒素 A 是野生型、CRM66 或 rERA。

[0042] 在本发明的实施方式中,表达载体进一步包含邻近分泌信号的编码序列的标签序列。在实施方式中,表达载体进一步包含邻近毒素蛋白的编码序列的标签序列。

[0043] 本发明还提供根据本文所述方法生产的重组毒素蛋白。在实施方式中,重组毒素蛋白是 CRM197、白喉毒素、霍乱全毒素、霍乱毒素 B、百日咳毒素、破伤风毒素 C 片段、艰难梭菌毒素 B 或铜绿假单胞菌外毒素 A。在实施方式中,外毒素 A 是野生型、CRM66 或 rERA。在某些实施方式中,重组毒素蛋白在荧光假单胞菌的菌株中生产,该菌株在本文中确认为生产高产率的毒素或生产高质量的毒素。在某些实施方式中,重组毒素蛋白在荧光假单胞菌的菌株中生产,该菌株在本文中描述为生产最高产率的毒素蛋白。在其他实施方式中,重组毒素蛋白在本文描述为用于毒素发酵生产的菌株的菌株中生产。

[0044] 援引并入

[0045] 本说明书中提到的所有公开、专利和专利申请通过引用并入本文,其程度如同特别地和单独地指示各单个公开、专利或专利申请通过引用并入本文。

附图说明

[0046] 本发明的新特征在所附的权利要求书中具体阐述。通过参考以下对在其中利用到本发明原理的说明性实施方式加以阐述的详细描述和附图,将获得对本发明的特征和优点的更好的理解。

[0047] 图 1. CRM197 的高通量表达分析。利用毛细管凝胶电泳 (SDS-CGE) 分析利用图 1B 所示的 DNA 序列表达的 CRM197 蛋白。在由 SDS-CGE 数据生成的凝胶样图像中显示了所测试的 40 个 CRM197 表达菌株的可溶性部分。如表 10 所述的菌株名称列于各个泳道上方。荧光假单胞菌表达的 CMR197 在 SDS-CGE 上作为 ~ 58kDa 的单一条带 (左侧箭头) 迁移。第一个和最后一个泳道内的分子量标记是 16、20、29、48、69 和 119kDa。

[0048] 图 2. 霍乱毒素 B 的高通量表达分析。利用毛细管凝胶电泳 (SDS-CGE) 分析应用 SEQ ID NO :23 中所示的 DNA 序列表达的霍乱毒素 B 蛋白。在由 SDS-CGE 数据生成的凝胶样图像中显示了所测试的 40 个霍乱毒素表达菌株的可溶性部分。如表 11 所述的菌株名称列于各个泳道上方。诱导的 CTB 在 SDS-CGE 上作为 ~ 11.5kDa 的单一一条带 (左侧箭头) 迁移。第一个和最后一个泳道内的分子量标记是 16、20、29、48、69 和 119kDa。

[0049] 图 3. 百日咳类毒素操纵子。显示了具有 4210 个碱基对的 BPETOX_S1-R9K & E129A。

[0050] 图 4. 百日咳类毒素的 DNA 序列。图中显示了带翻译结果的百日咳毒素 S1 R9K E129A DNA 序列 (SEQ ID NO :24)。该序列来源于 Genbank 项目 M13223。亚基 S1-S5 和信号序列在序列上方标明。S1 中的 R9K 和 E129A 突变用下划线标注。编码的蛋白质按出现顺序分别披露为 SEQ ID NO :25、26、28、29 和 27。

[0051] 图 5. 百日咳类毒素亚基的氨基酸序列。分泌信号用下划线标注。A. S1 亚基 (R9K E129A) (SEQ ID NO :25)。B. S2 亚基 (SEQ ID NO :26)。C. S3 亚基 (SEQ ID NO :27)。D. S4 亚基 (SEQ ID NO :28)。E. S5 亚基 (SEQ ID NO :29)。

[0052] 图 6. 百日咳类毒素表达样品的 Western 印迹分析。菌株名称已列于各个泳道上方。诱导的 Ptx 以 11-26kDa 范围内的多个条带 (S1 :26.1Kda, S2 :20.9Kda, S3 :21.8Kda, S4 (2x) :12Kda, S5 :11Kda) 进行迁移。A. 还原的样品。B. 非还原的样品。两个图面 :泳道 1- 分子量标记 (15、20、25、37、50、75、100、150、250kDa) ;泳道 2- 空的 ;泳道 3- 菌株 321 ;泳道 4- 菌株 322 ;泳道 5- 菌株 323 ;泳道 6- 菌株 324 ;泳道 7- 菌株 325 ;泳道 8- 菌株 326 ;泳道 9- 菌株 327 ;泳道 10- 菌株 328。

[0053] 图 7. 破伤风毒素 C 片段表达。利用毛细管凝胶电泳 (SDS-CGE) 分析在荧光假单胞菌中表达的破伤风毒素 C 片段。在由 SDS-CGE 数据生成的凝胶样图像中显示了来自所测试的 40 个破伤风毒素表达菌株的可溶性部分。如表 15 所述的菌株名称列于各个泳道上方。诱导的破伤风毒素 C 片段在 SDS-CGE 上以 ~ 51.6kDa 的单一一条带 (左侧箭头) 迁移。第一个和最后一个泳道内的分子量标记是 16、20、29、48、69 和 119kDa。

[0054] 图 8. TcdB 表达。利用毛细管凝胶电泳 (SDS-CGE) 分析在荧光假单胞菌中表达的 TcdB。在由 SDS-CGE 数据生成的凝胶样图像中显示了来自所测试的 24 个 TcdB 表达菌株的可溶性部分。如表 18 所述的菌株名称以及空提取物和参考标准 (List Biologicals) 都列于各个泳道上方。诱导的 TcdB 在 SDS-CGE 上以 ~ 300kDa 的单一一条带 (左侧箭头) 迁移。第一个和最后一个泳道内的分子量标记是 16、20、29、48、69 和 119kDa。

[0055] 图 9. 外毒素 A 氨基酸序列。图中显示了铜绿假单胞菌外毒素 A 的氨基酸序列 (SEQ ID NO :34)。图中显示了三种外毒素 A 蛋白 :野生型、CRM66 和 rEPA。在变异体 CRM66 中, His426 (粗体, 下划线标注的文字) 如序列上方所示被 Tyr 取代。在 rEPA 中, Glu553 (粗体, 下划线标注的文字) 如序列上方所示缺失。

[0056] 图 10. 可溶性霍乱毒素 B 在荧光假单胞菌发酵培养物中的产生。SDS-CGE 分析。泳道 1 — 16、20、29、48、69 和 119kDa 分子量标记。泳道 2 和 4 — 分别是表达霍乱毒素 B 的 PS538-088 U5 和 U6 发酵物的诱导前样品, 泳道 3 和 5 分别是表达霍乱毒素 B 的 PS538-088 U5 和 U6 发酵物的诱导后样品, 通过右侧箭头指示。

[0057] 图 11. 可溶性破伤风毒素 C 片段在荧光假单胞菌发酵培养物中的产生。A. SDS-CGE

分析。泳道 1-16、20、29、48、69 和 119kDa 分子量标记。泳道 2、3 和 4 分别是表达破伤风毒素 C 片段的 PS538-529U1、PS538-546U5 和 PS538-547U7 发酵物的诱导后样品,通过右侧箭头指示。B. Western 印迹分析。通过 Western 印迹法评估来自菌株 PS538-538 (U1 和 U2)、PS538-548 (U3 和 U4)、PS538-558 (U5 和 U6) 和 PS538-568 (U7 和 U8) 的发酵样品。发酵单位和诱导后小时数 (I0、I8、I24) 已在各个泳道上方标出。分子量 (MW) 标准显示在印迹左侧,破伤风毒素 C 参考标准 (Std ;List Biological, Cat# 193) 显示在右侧。先用源自兔的多克隆抗破伤风毒素 C 片段 (Abcam, Cat# :ab34890),然后用源自山羊的抗兔 IgG 过氧化物酶 (Pierce, Cat# :31460) 探测印迹。免疫纯金属强化 DAB (Immunopure Metal Enhanced DAB) (Pierce 34065) 用于检测。

[0058] 图 12. 可溶性艰难梭菌 B 毒素蛋白在荧光假单胞菌发酵培养物中的产生。泳道 1-16、20、29、48、69 和 119kDa 分子量标记。标记大小也根据它们在泳道 1 中的迁移显示在右侧相对应的位置上。泳道 2、3 和 4 分别是表达艰难梭菌 B 毒素蛋白的 PS538-671 U5 和 PS538-647 U7 发酵物的诱导后样品,通过右侧箭头指示。

[0059] 图 13. 野生型百日咳类毒素的 DNA 序列。图中显示了带翻译结果的野生型百日咳毒素的 DNA 序列 (SEQ ID NO :35)。序列来自 Genbank 项目 M13223。亚基 S1-S5 和信号序列已显示在序列上方。编码的蛋白质按出现顺序分别显示为 SEQ ID NO :41-45。

[0060] 图 14. 霍乱全毒素的氨基酸序列和 DNA 序列。A. CTA 氨基酸序列 (SEQ ID NO :38),具有分泌前导序列 (下划线标注) (AE003852 ;蛋白质 ID AAF94614.1)。B. CTB 氨基酸序列 (SEQ ID NO :39),具有分泌前导序列 (下划线标注) (GeneBank AE003852 ;蛋白质 ID AAF94613.1)。C. CTX DNA 序列 (SEQ ID NO :40),显示了 A 和 B 亚基,并显示出翻译结果 (Genbank AE003852)。编码的蛋白质按出现顺序分别显示为 SEQ ID NO :38 和 39。

[0061] 图 15. 在荧光假单胞菌发酵培养物中的可溶性 rEPA 生产的 SDS-CGE 凝胶样图像。利用毛细管凝胶电泳 (SDS-CGE) 分析在荧光假单胞菌发酵培养物中表达的可溶性 rEPA。在由 SDS-CGE 数据生成的凝胶样图像中显示了所测试的在诱导后 0 和 24 小时由表达菌株 PS538-1633 (u1 和 u2)、PS538-1640 (u3 和 u5) 和 PS538-1670 (u6、u7 和 u8) 发酵产生的可溶性部分。Mw = 分子量标准 (16、20、29、48 和 69 千道尔顿)。

[0062] 图 16. 在荧光假单胞菌发酵培养物中的可溶性 rEPA 产生趋势。通过对处于各自发酵物 (u1、u2、u3、u6、u7 和 u8) 中的菌株 (PS538-1633、PS538-1640 和 PS538-1670) 进行 SDS-CGE 分析而确定的可溶性 rEPA 表达水平对诱导后时间作图。

[0063] 图 17. 在荧光假单胞菌发酵培养物中的可溶性 rEPA 产生的 Western 印迹分析。利用 Western 印迹分析来分析在荧光假单胞菌发酵培养物中表达的可溶性 rEPA。利用对于铜绿假单胞菌外毒素 A 特异性的抗体在 Western 印迹分析中显示诱导后 0 和 24 小时来自表达菌株 PS538-1633 (u1)、PS538-1640 (u3 和 u5) 和 PS538-1670 (u6 和 u8) 发酵物的可溶性部分。Mw = 分子量标准。std = rEPA 标准品。

[0064] 图 18. 在荧光假单胞菌发酵培养物中的可溶性 CRM197 产生的 SDS-CGE 凝胶样图像。利用毛细管凝胶电泳 (SDS-CGE) 来分析在荧光假单胞菌发酵培养物中表达的 CRM197。在由 SDS-CGE 数据生成的凝胶样图像中显示了在所测试的诱导后各个不同时间 (0、16、21 和 23 小时) 由表达菌株 PS538-772 (u1 和 u2)、PS538-776 (u3 和 u5) 和 PS538-782 (u6 和 u7) 的各种发酵物所产生的可溶性部分。Mw = 分子量标准 (16、20、29、48、68 和 119 千道尔

顿)。

[0065] 图 19. 在荧光假单胞菌发酵培养物中的可溶性 CRM197 产生趋势。如通过对处于各自发酵物 (u1、u2、u3、u6 和 u7) 中的不同菌株 (PS538-772、PS538-776 和 PS538-782) 进行 SDS-CGE 所确定的可溶性 CRM197 表达水平对诱导后时间作图。

[0066] 图 20. 在荧光假单胞菌发酵培养物中的可溶性 CRM197 产生的 Western 印迹分析。利用 Western 印迹分析来分析在荧光假单胞菌发酵培养物中表达的 CRM197。利用白喉毒素特异性抗体在 Western 印迹分析中显示在所测试的诱导后各个不同时间 (0、16、21 和 23 小时) 由表达菌株 PS538-772 (u1 和 u2)、PS538-776 (u3 和 u5) 和 PS538-782 (u6 和 u7) 的各种发酵物的可溶性部分。Mw = 分子量标准 (37、50、75、100、150 和 250 千道尔顿)。STD = CRM197 标准。

[0067] 发明详述

[0068] 毒素

[0069] ADP-核糖基化毒素

[0070] ADP-核糖基化毒素 (ADPRT) 促进烟酰胺和 NAD 的 N-核糖之间的 N-糖基键切割并且将 ADP-核糖部分转移至靶蛋白。ADPRT 根据他们各自的靶标分类为四个家族。I 型 ADPRT 以异聚 GTP-结合蛋白为靶标。它们包括霍乱毒素 (CTX)、百日咳毒素 (PTX) 和大肠杆菌不耐热肠毒素 (LT)。II 型 ADPRT (白喉毒素和假单胞菌外毒素 A) 修饰延伸因子 2 (EF2)。III 型 ADPRT (肉毒梭菌 (*Clostridium botulinum*) C3 胞外酶) 使小 GTP-结合蛋白 ADP-核糖基化。IV 型 ADPRT 使肌动蛋白 ADP-核糖基化。这些肌动蛋白特异性的 ADPRT 包括二元毒素家族, 该家族包括肉毒梭菌 C2 毒素、产气荚膜梭菌 (*C. perfringens*) ι -毒素、艰难梭菌毒素 (一种不同于 TcdA 和 TcdB 的毒素, 由 Popoff 等人, 1988, "Actin-specific ADP-ribosyltransferase produced by a *Clostridium difficile* strain," *Infection and Immunity* 56(9):2299-2306 描述, 其通过引用并入本文)、螺状梭菌 (*C. spiroforme*) 毒素和蜡状芽孢杆菌 (*Bacillus cereus*) 营养期杀虫蛋白 (VIP)。

[0071] 已经确定来自各类型的 ADPRT 的数个酶组分的结构带有或不带有 NAD, 并且在例如通过引用并入本文的 Tsuge 等人, 2008, "Structural basis of actin recognition and arginine ADP-ribosylation by *Clostridium perfringens*-toxin," *PNAS* 105(21):7399-7404 中讨论。典型的肌动蛋白特异性 ADPRT 具有两个相似的结构域: 对催化活性至关重要的 C 结构域; 以及对与结合和易位亚基的相互作用很重要的 N 结构域。相比之下, 来自沙门氏菌的 SpvB 和 III 型 ADPRT C3 仅具有一个 ADP-核糖基转移酶结构域并且缺乏 N 末端衔接子结构域。在所有的 IV 型 ADPRT 中, 包括两个关键的谷氨酸残基的 EXE 基序存在于催化中心处。EXE 基序的前一谷氨酸被认为是 ADP-核糖基转移酶的关键残基, 它在肌动蛋白中由 Arg-177 去质子化。后一谷氨酸与 N-核糖上的 O'2 形成氢键, 这被认为稳定了羰基碳鎓 (oxocarbenium) 阳离子。

[0072] Barth 等人, 2004, "Binary Bacterial Toxins: Biochemistry, Biology, and Application of Common *Clostridium* and *Bacillus* Proteins," *Microbiology and Molecular Biology Reviews* 68(3):373-402; Mueller-Dieckmann 等人, "Structure of mouse ADP-ribosylhydrolase 3 (mARH3)," *Acta Cryst F* 64:156-162; Kulich 等人, 1995, "Expression of Recombinant Exoenzyme S of *Pseudomonas aeruginosa*," *Infection*

and Immunity 63(1):1-8; Sakurai 等人, 2009, "Clostridium perfringens Iota-Toxin: Structure and Function," Toxins 1:208-228; 以及 Schirmer 等人, 2002, "The ADP-ribosylating Mosquitocidal Toxin from Bacillus sphaericus," The Journal of Biological Chemistry 277(14):11941-11948 进一步描述了 ADPRT, 它们全部通过引用并入本文。

[0073] 在本发明的实施方式中, 产生选自包括 ADPRT 的组的重组毒素蛋白。在实施方式中, ADPRT 的组由 CTX (CTA 和 / 或 CTB)、PTX、DT (CRM197 和 / 或 WT) 和假单胞菌外毒素 A 组成。在实施方式中, ADPRT 的组由 CTX (CTA 和 / 或 CTB)、PTX 和假单胞菌外毒素 A 组成。在其他实施方式中, 产生选自包括 I 型 ADPRT 的组的重组毒素蛋白。在实施方式中, I 型 ADPRT 的组由 CTX (CTA 和 / 或 CTB) 和 PTX 组成。在其他实施方式中, 产生选自包括 II 型 ADPRT 的组的重组毒素蛋白。在实施方式中, II 型 ADPRT 的组由 DT (CRM197 和 / 或 WT) 和假单胞菌外毒素 A 组成。在其他实施方式中, 产生选自包括 IV 型 ADPRT 的组的重组毒素蛋白。在实施方式中, IV 型 ADPRT 是 TcdB。

[0074] CRM197 和 DT

[0075] 交叉反应物质 197 (CRM197) 是由具有错义突变的 DT 基因产生的白喉毒素 (DT) 变异体。DT 是一种 ADP-核糖基化毒素; CRM197 缺乏 DT 的 ADP-核糖基转移酶 (ADPRT) 活性, 且因此是无毒的。CRM197 的基因具有单碱基置换, 从而导致在残基 52 处引入谷氨酸而不是甘氨酸。(参见, 例如, Bishai 等人, 1987, "High-Level Expression of a Proteolytically Sensitive Diphtheria toxin Fragment in Escherichia coli," J. Bact. 169(11): 5140-51, Giannini 等人, 1984, "The Amino-Acid Sequence of Two Non-Toxic Mutants of Diphtheria toxin: CRM45 and CRM197," Nucleic Acids Research 12(10):4063-9, 以及 GenBank 登录号 1007216A, 全部通过引用并入本文。)

[0076] CRM197 蛋白可通过本领域已知的方法或通过白喉杆菌或其他微生物中表达而低水平制备。天然存在的或野生型白喉毒素可从产毒素菌株获得, 该产毒素菌株可从包括美国典型培养物保藏中心 (American Type Culture Collection) 在内的多种公共资源获得。用于在白喉杆菌中生产 CRM197 蛋白的质粒系统在例如美国专利号 5,614,382, "Plasmid for Production of CRM Protein and Diphtheria toxin" 中有描述, 其通过引用整体并入本文。

[0077] 核苷酸序列可利用重组 DNA 技术 (例如, Sambrook 等人, Molecular Cloning, a Laboratory Manual, Cold Spring Harbor Laboratory Press, 1989 描述) 制备, 且也基于棒状杆菌噬菌体 β 所携带的白喉毒素野生型结构基因的已知 DT 核苷酸序列, 通过定点诱变来制备。(参见, 例如, Greenfield 等人, 1993, "Nucleotide Sequence of the Structural Gene for Diphtheria toxin Carried by Corynebacteriophage 18," Proc Nat Acad Sci 80:6953-7, 通过引用并入本文。) 如本文别处所述, 核苷酸序列可以进行优化。

[0078] 在本发明的实施方式中, 利用实施例 1 中所述的任何宿主菌株结合实施例 1 中所述的任何表达载体 (质粒) 来生产 CRM197 或 DT。在实施方式中, 为了在假单胞菌宿主细胞中表达而优化核酸序列。在实施方式中, 所用的表达载体包含表达与重组 CRM197 或 DT 蛋白融合的表 8 和表 3 中所述的任何分泌前导序列的构建体。在实施方式中, 使用天然分泌

前导序列。在某些实施方式中,CRM197 或 DT 蛋白表达有标签,例如,纯化标签。在实施方式中,本发明的方法用于以约 0.5g/L 至至少约 12g/L 的产率生产 CRM197 或 DT。

[0079] 霍乱毒素

[0080] 由霍乱弧菌产生的霍乱毒素 (CTX) 也是一种 ADP-核糖基化毒素。霍乱毒素 (CTX) 是由六个蛋白质亚基组成的低聚复合物:单一拷贝的霍乱毒素 A 亚基 (CTA) 和五个拷贝的霍乱毒素 B 亚基 (CTB)。五个各重 12kDa 的 B 亚基形成五元环。A 亚基具有 A1 部分 CTA1 和 A2 链 CTA2,CTA1 是使 G 蛋白 ADP-核糖基化的球状酶,而 CTA2 形成了一个紧贴地位于 B 亚基环中心孔中的伸展的 α 螺旋。该环结合到位于宿主细胞表面上的 GM1 神经节苷脂受体上,从而导致整个复合物的内化。一旦内化,CTA1 链由于二硫键的还原而释放。随后 CTA1 被激活并且催化腺苷酸环化酶的 ADP 核糖基化。由此导致的腺苷酸环化酶活性的增强提高了环 AMP 的合成,这导致大量的流体和电解质流出并且导致腹泻。

[0081] CTX 的 B 亚基尽管是相对无害的,但保留了其结合到 GM1 神经节苷脂受体上的能力。因此,CTB 可用于促进化学地或遗传地偶联的外来抗原的粘膜摄取。已证实其诱导粘膜免疫和系统免疫,并且是用于可食用疫苗生产的候选物。由于其结合偏好,CTB 也可用作神经元示踪物。

[0082] CTB 的应用及其结构特征在例如 Nozoye 等人,2009,“Production of *Ascaris suum* As14 Protein and Its Fusion Protein with Cholera Toxin B Subunit in Rice Seeds,”*Parasitology* 995-1000;Harakuni 等人,2005,“Heteropentameric Cholera Toxin B Subunit Chimeric Molecules Genetically Fused to a Vaccine Antigen Induce Systemic and Mucosal Immune Responses:a Potential New Strategy to Target Recombinant Vaccine Antigens to Mucosal Immune Systems,”*Infection and Immunity* 73(9):5654-5665;Price 等人,2005,“Intranasal Administration of Recombinant *Neisseria gonorrhoeae* Transferrin Binding Proteins A and B Conjugated to the Cholera Toxin B Subunit Induces Systemic and Vaginal Antibodies in Mice,”*Infection and Immunity* 73(7):3945-3953;和 Sun 等人,1999,“Intranasal Administration of a *Schistosoma mansoni* Glutathione S-Transferase-Cholera Toxoid Conjugate Vaccine Evokes Antiparasitic and Antipathological Immunity in Mice,”*J. Immunol.* 163:1045-1052 中已有描述,它们全部通过引用并入本文。

[0083] 在本发明的实施方式中,利用本文实施例 1 中所述的任何宿主菌株结合实施例 3 中所述的任何表达载体来生产 CTB 或 CTX。在实施方式中,为了在假单胞菌宿主细胞中表达而优化了核酸序列。在实施方式中,所用的表达载体包含表达与重组 CTB 或 CTX 蛋白融合的表 8 和表 3 中所述的任何分泌前导序列的构建体。在实施方式中,使用天然分泌前导序列。在某些实施方式中,CTB 或 CTX 蛋白表达有标签,例如,纯化标签。在实施方式中,本发明的方法用于以约 0.2g/L 至至少约 5g/L 的产率生产 CTB 或 CTX。

[0084] 百日咳毒素

[0085] 百日咳毒素是由导致百日咳病的人呼吸道细菌性病原体百日咳博德特菌产生的外毒素和毒力因子。百日咳全毒素是具有 AB₅ 结构的多亚基复合物。酶活性的 A 亚基 (S1) 是修饰哺乳动物细胞内的几种异源三聚体 G 蛋白 (主要是 G_i 蛋白) 的 α 亚基的 ADP-核糖基转移酶,和 B 低聚物 (S2、S3、两个拷贝的 S4 和 S5) 结合细胞上的糖结合物受体。S1 在

进入细胞后被蛋白水解加工。通过引用并入本文的 Carbonetti 等人,2005,“Proteolytic Cleavage of Pertussis Toxin S1 Subunit is Not Essential for Its Activity in Mammalian Cells,”BMC Microbiology 5:7 报告,S1 的加工对它在哺乳动物细胞中的细胞毒性活性并非必不可少的。

[0086] 已研究百日咳毒素的无毒性变异体以供在疫苗中使用。利用本发明的方法生产的百日咳毒素蛋白预期用于疫苗中以对抗百日咳。百日咳毒素也作为疫苗佐剂进行了测试,例如,如 Roberts 等人,1995,“A Mutant Pertussis Toxin Molecule That Lacks ADP-Ribosyltransferase Activity, PT-9K/129G, Is an Effective Mucosal Adjuvant for Intranasally Delivered Proteins”Infection and Immunity 63(6):2100-2108 所述,其通过引用并入本文。此外,百日咳毒素也可用于研究目的,例如用于 G 蛋白信号传导途径的研究(例如,McCoy 等人,2010,“PAR1 and PAR2 couple to overlapping and distinct sets of G proteins and linked signaling pathways to differentially regulate cell physiology,”Molecular Pharmacology Fast Forward MOL 62018,通过引用并入本文),以及用作佐剂来强化对诸如实验性自身免疫性脑脊髓炎(EAE)、实验性自身免疫性睾丸炎、实验性自身免疫性葡萄膜炎等自身免疫性疾病的诱导。(Su 等人,2001,“Pertussis Toxin Inhibits Induction of Tissue-Specific Autoimmune Disease by Disrupting G Protein-Coupled Signals,”J Immunol 167:250-256,通过引用并入本文)。

[0087] 如图 3 所示,该毒素的五个亚基由百日咳类毒素操纵子表达。包括某些变异体在内的百日咳毒素蛋白的表达和结构在以上引用的报告中以及由 Burnette 等人,1992,“Properties of Pertussis Toxin B Oligomer Assembled In Vitro from Recombinant Polypeptides Produced by Escherichia coli,”Infection and Immunity 60(6):2252-2256;美国专利号 5,085,862,“Genetic detoxification of pertussis toxin;”和 Kaslow 等人,1987,“Structure-Activity Analysis of the Activation of Pertussis Toxin,”Biochemistry 26(1):123-7 描述;它们全部通过引用整体并入本文。

[0088] 如本文所用的百日咳毒素或 PTX 指百日咳毒素突变体 S1 R9K E129A 或野生型蛋白质。野生型百日咳毒素和百日咳毒素突变体 S1 R9K E129A 在例如以下文献中描述:Roberts 等人,1995(以上引用);美国专利号 7,427,404 和美国专利号 7,666,436,名称均为“Pertussis Toxin Mutants, Bordetella Strains Capable of Producing Such Mutants and Their Use in the Development of Antipertussis Vaccines;”美国专利号 5,935,580,“Recombinant Mutants for Inducing Specific Immune Responses;”美国专利号 7,169,399,“Non-Toxic Double Mutant Forms of Pertussis Toxin as Adjuvants;”美国专利号 5,785,971 和美国专利号 5,427,788,名称均为“Pertussis Toxin and Use in Vaccines;”和美国专利号 5,773,600,“DNA Encoding Pertussis Toxin Muteins”,它们全部通过引用整体并入本文。

[0089] 在本发明的实施方式中,利用本文实施例 1、5 和 7 中所述的任何宿主菌株生产百日咳毒素突变体 S1 E129A 或野生型百日咳毒素。在实施方式中,所用的表达载体包含表达与重组 PTX 蛋白融合的表 8 和表 3 中所述的任何分泌前导序列的构建体。在实施方式中,使用天然分泌前导序列。在实施方式中,如本文别处所述,为了在选定的假单胞菌宿主中表

达而优化任何或所有亚基编码序列。在某些实施方式中,亚基由两个或多个构建体表达,例如,通过根据本领域公知的方法亚克隆单个序列。在某些实施方式中,PTX 蛋白表达有标签,例如纯化标签。在实施方式中,本发明的方法用于以约 0.2g/L 至至少约 5g/L 的产率生产 PTX 或 PTX 的各单个亚基。

[0090] 破伤风毒素 C 片段

[0091] 由破伤风梭菌产生的破伤风毒素是具有 150kDa 的分子量的神经毒素。它由两部分组成:100kDa 重链或 B-链和 50kDa 的轻链或 A-链。这些链由二硫键连接。B 链结合到位于神经元膜上的二唾液酸神经节苷脂 (GD2 和 GD1b) 上。A 链 (一种锌内肽酶) 攻击囊泡相关膜蛋白 (VAMP)。

[0092] A 链的作用通过降解小突触泡蛋白来阻止受影响的神经元释放出抑制性神经递质 GABA (γ -氨基丁酸) 和甘氨酸。其后果是最小的刺激引起危险的肌肉过度活动—感官刺激对运动反射的抑制失败。这导致主动肌和拮抗肌肌肉组织的全身性收缩,被称为强直性痉挛。

[0093] 破伤风毒素 C 片段 (Tet C 或 TTC) 是通过破伤风毒素的蛋白酶裂解 (例如,用木瓜蛋白酶) 或通过该片段的重组表达而产生的 50kD 多肽。它对应于 C-末端的 451 个氨基酸 (氨基酸位置 865-1315)。C 片段的重组表达已在例如美国专利号 5,443,966,“Expression of Tetanus Toxin Fragment C,” WO/2005/000346,“Carrier Proteins for Vaccines,” 和 6,010,871,“Modification of Peptide and Protein” 中公开,它们全部通过引用整体并入本文。

[0094] 已证明 C 片段是无毒性的并且能够在小鼠和豚鼠中刺激保护性免疫应答。美国专利号 5,443,966 描述了破伤风毒素的序列以及 C 片段在大肠杆菌中的生产。重组 TTC 在酵母中的表达已在例如美国专利号 5,571,694,“Expression of Tetanus Toxin Fragment C in Yeast” 中描述,其通过引用整体并入本文。

[0095] 因为 TTC 以高特异性和亲和力结合神经元, TTC 可用作神经元药物递送的靶向分子或用于研究目的。此类用途在例如 Townsend 等人,2007,“Tetanus toxin C fragment conjugated nanoparticles for targeted drug delivery to neurons,” Biomaterials 28(34):5176-5184 中描述,其通过引用并入本文。

[0096] TTC 蛋白也可潜在地用作疫苗载体蛋白,如例如 WO/2005/000346 中所述,并且已研究在疫苗中使用以避免破伤风梭菌感染。

[0097] 在本发明的实施方式中,利用本文实施例 1 中所述的任何宿主菌株结合实施例 8 中所述的任何表达载体生产 TTC。在实施方式中,为了在假单胞菌宿主细胞中表达而优化核酸序列。在实施方式中,所用的表达载体具有表达与重组 TTC 蛋白融合的表 8 和表 3 中所述的任何分泌前导序列的构建体。在某些实施方式中, TTC 蛋白表达有标签,例如,纯化标签。在实施方式中,使用天然分泌前导序列。在实施方式中,本发明的方法用于以约 0.5g/L 至至少约 12g/L 的产率生产 TTC。

[0098] 艰难梭菌毒素 B

[0099] 艰难梭菌毒素 B (TcdB) 是由导致医院获得性腹泻和假膜性结肠炎的艰难梭菌产生的毒力因子。TcdB 和第二大梭菌毒素 TcdA 参与了假膜性结肠炎的发生。

[0100] TcdB 是约 270kD 的糖基化毒素,可分成酶结构域、易位结构域和受体结合结构域。

TcdB 的前 546 个氨基酸含有酶区域,之后是推定的易位和受体结合结构域。已报道酶活性需要氨基末端的 546 个残基,因为该片段的氨基或羧基末端缺失降低了活性。在该酶区域内,已证明色氨酸 102 对 UDP- 葡萄糖结合至关重要。LCT 内的保守的 DXD 基序对 LCT 葡萄糖基转移酶活性至关重要。涉及 TcdB 和 TcsL 酶结构域的嵌合体分析的研究提示残基 364 至 516 赋予了底物特异性。

[0101] TcdB 的结构及其表达以及作为针对艰难梭菌感染的保护性疫苗的潜在用途在例如以下文献中讨论:美国专利号 7,226,597,“Mutants of Clostridium Difficile Toxin B and Methods of Use;” Jank 等人,2008,“Structure and mode of action of clostridial glucosylating toxins:the ABCD model,”Trends in Microbiology 16(5):222-229;Sullivan 等人,1982,“Purification and Characterization of Toxins A and B of Clostridium difficile,”Infection and Immunity 35(3):1032-1040;和 Yang 等人,2008,“Expression of recombinant Clostridium difficile toxin A and B in Bacillus megaterium,”BMC Microbiology 8:192,它们全部通过引用整体并入本文。

[0102] 在本发明的实施方式中,利用本文实施例 1、5 和 7 中所述的任何宿主菌株来生产 TcdB。在实施方式中,为了在假单胞菌宿主细胞中表达而优化核酸序列。在实施方式中,所用的表达载体包含表达与重组 TcdB 蛋白融合的表 8 和表 3 中所述的任何分泌前导序列的构建体。在实施方式中,使用天然分泌前导序列。在某些实施方式中,TcdB 蛋白表达有标签,例如,纯化标签。在实施方式中,本发明的方法用于以约 0.5g/L 至至少约 10g/L 的产率生产 TcdB。

[0103] 铜绿假单胞菌外毒素 A

[0104] 铜绿假单胞菌的外毒素 A(ETA 或 PE) 是 II 型 ADPRT。它是能够将催化结构域转位至哺乳动物细胞中并且通过细胞延伸因子 2 的 ADP- 核糖基化抑制蛋白质合成的分泌的细菌毒素家族的成员。该蛋白质作为单体存在,由 613 个氨基酸(66Kd) 的单一多肽链组成。以 3.0-Å 分辨率确定的外毒素 A 的 X 射线晶体结构显示出主要由反平行的 β 结构组成并且包含该分子的大约一半的氨基末端结构域;由 α 螺旋组成的中间结构域;以及包含该分子的大约三分之一的羧基末端结构域。羧基末端结构域是毒素的 ADP- 核糖基转移酶。其他两个结构域推测参与细胞受体结合和膜转位。

[0105] 该毒素通过位于细胞表面上的特定受体结合到细胞上,然后毒素-受体复合物内化到细胞中。最后,ETA 被转移到细胞溶质,在其中它酶促抑制蛋白质合成。由于弱碱比如 NH_4^+ (它提高酸性囊泡中的 pH) 防止细胞中毒,据认为转移过程从酸性区室发生。一旦暴露在酸性条件下,PE 的疏水结构域进入膜内,从而导致通道的形成,酶结构域以伸展形式通过该通道进入细胞溶质中。PE 的活性和毒性减弱的突变体在例如以下文献中描述:美国专利号 4,892,827,“Recombinant Pseudomonas Exotoxins:Construction of an Active Immunotoxin with Low Side Effects” 和 Lukac 等人,1988,“Toxoid of Pseudomonas aeruginosa Exotoxin A Generated by Deletion of an Active-Site Residue”Infection and Immunity 56(12):3095-3098,二者通过引用整体并入本文。

[0106] 外毒素 A 突变体 rEPA 作为疫苗偶联物的应用在例如以下文献中描述:Fattom 等人,1993,“Laboratory and Clinical Evaluation of Conjugate Vaccines Composed of Staphylococcus aureus Type 5 and Type 8 Capsular Polysaccharides Bound to

Pseudomonas aeruginosa Recombinant Exoprotein A” *Infection and Immunity* 61(3) : 1023-1032 ;Qian 等 人,2007,“Conjugating recombinant proteins to *Pseudomonas aeruginosa* ExoProtein A:a strategy for enhancing immunogenicity of malaria vaccine candidates” *Vaccine* 25(20) :3923-3933 ;以及 Lin 等人,2001. “The Efficacy of a *Salmonella* Typhi Vi Conjugate Vaccine in Two-To-Five-Year-Old Children” *N Engl J Med* 344(17) :1263-1269,均通过引用并入本文。

[0107] 本文中使用的铜绿假单胞菌外毒素A指铜绿假单胞菌外毒素A突变体CRM66、缺失rEPA或野生型蛋白质。在本发明的实施方式中,利用本文实施例1、5和7中所述的任何宿主菌株并且利用具有表达与重组外毒素A蛋白融合的表8和表3中所述的任何分泌前导序列的构建体的表达载体来生产外毒素A。在实施方式中,为了在假单胞菌宿主细胞中表达而优化核酸序列。在实施方式中,使用天然分泌前导序列。在某些实施方式中,ETA蛋白表达有标签,例如,纯化标签。在实施方式中,本发明的方法用于以约0.5g/L至至少约12g/L的产率生产外毒素A。

[0108] 用本发明的方法生产的示例性毒素蛋白在表1中列出。可以理解,这个列表不是限制性的。在本发明的实施方式中,为了在选择的假单胞菌宿主细胞中表达,可以优化在此用于利用本发明方法生产本文所述的毒素的任何核酸序列。如本文别处所述,任何给定序列的优化存在多种选择。所述的任何选项预期用于优化利用本发明方法生产的毒素的序列。本文提供的优化序列是在本发明方法中有用的优化序列的非限制性实例。

[0109] 表1. 示例性毒素蛋白

[0110]

靶标	示例性序列	来源
----	-------	----

[0111]

	资源/参考	
CRM197	GenBank 登录号 1007216A	白喉杆菌 NCTC 13129
白喉毒素(WT)	GenBank NC_002935.2 GenBank CAA00374.1	白喉杆菌
霍乱全毒素	GenBank NC_002505.1; NP231099.1 和 NP23110.1	霍乱弧菌
霍乱毒素 B	GenBank ACH70471 (E1 Tor 菌株)	霍乱弧菌 O1 生物变种 El tor
百日咳毒素	GenBank M13223.1, 具有 S1 突变	百日咳博德特菌
破伤风毒素 C 片段	GenBank 1A8D_A	破伤风梭菌
艰难梭菌毒素 B VPI (TcdB)	GenBank CAA63562	艰难梭菌
铜绿假单胞菌外毒素 A	GenBank NP_249839	铜绿假单胞菌 PAO1

[0112] 密码子优化

[0113] 在异源表达系统中,优化步骤可以提高宿主产生外源蛋白的能力。蛋白质表达是由一系列包括那些影响转录、mRNA 加工及翻译的稳定性和起始的因素的许多因素控制的。多核苷酸优化步骤可以包括提高宿主产生外源蛋白能力的步骤,以及辅助研究人员有效地设计表达构建体的步骤。优化策略可以包括,例如,翻译起始区的修饰、mRNA 结构元件的改变和不同密码子偏倚性的使用。优化核酸序列以提高在细菌宿主中异源蛋白的表达的方法在本领域中是已知的,并在文献中描述。例如,用于假单胞菌宿主菌株中表达的密码子优化描述在美国专利申请公开号 2007/0292918,“Codon Optimization Method”中,其全部内容通过引用并入本文。

[0114] 因此,优化可以处理异源基因的多种序列特征中的任意一种。作为具体的实例,稀有密码子引起的翻译中止可导致异源蛋白表达的降低。稀有密码子引起的翻译中止包括在目标多核苷酸中存在很少在宿主生物体中使用的密码子可能对蛋白质的翻译有负面影响,因为它们在可用的 tRNA 池中的缺少。提高宿主生物体中最佳翻译的方法包括 可导致稀有宿主密码子从合成的多核苷酸序列中去除的密码子优化。

[0115] 替代的翻译起始也可以导致异源蛋白表达的降低。替代的翻译起始可包含偶然地包含能够作为核糖体结合位点 (RBS) 发挥功能的基序的合成多核苷酸序列。这些位点可引起从基因内部位点起始截短蛋白的翻译。降低产生截短蛋白(其可能难以在纯化过程中去除)的可能性的一种方法包括从优化的多核苷酸序列中消除推定的内部 RBS 序列。

[0116] 重复引起的聚合酶滑脱可以导致异源蛋白表达的下降。重复引起的聚合酶滑脱涉及表明会引起可产生移码突变的 DNA 聚合酶的滑脱或口吃(stuttering)的核苷酸序列重复。这样的重复序列也能引起 RNA 聚合酶的滑脱。在具有高 G+C 含量偏倚性的生物体中,可以有更程度的由 G 或 C 核苷酸重复构成的重复序列。因此,降低引起 RNA 聚合酶滑脱的可能性的一种方法包括改变 G 或 C 核苷酸的延伸重复序列。

[0117] 干扰二级结构也可能导致异源蛋白表达的下降。二级结构可以隔离 RBS 序列或起始密码子,且与蛋白质表达的下降相关。茎-环结构也可以参与转录中止和减弱。优化的多核苷酸序列在核苷酸序列的 RBS 和基因编码区中可以含有最少的二级结构以允许转录和翻译的改善。

[0118] 另一种可能影响异源蛋白表达的特征是限制性位点的存在。可以通过除去可能干扰随后转录单位亚克隆到宿主表达载体中的限制性位点优化多核苷酸序列。

[0119] 例如,可以通过识别由宿主异源表达所需的氨基酸序列开始优化过程。可以从该氨基酸序列设计候选多核苷酸或 DNA 序列。在设计合成 DNA 序列时,密码子选择的频率可以与宿主表达有机体的密码子选择进行比较和稀有宿主密码子可以从合成序列中除去。此外,合成的候选 DNA 序列可以被修饰以除去不想要的限制性酶切位点,以及添加或移除任何所需的信号序列、接头或非翻译区。可以分析合成 DNA 序列中可能会干扰的翻译过程的二级结构的存在,如 G/C 重复 序列和茎-环结构。在候选 DNA 序列合成之前,可以检验优化的序列设计以确认该序列正确编码所需的氨基酸序列。最后,可以使用 DNA 合成技术合成候选 DNA 序列,如本领域中已知的那些合成技术。

[0120] 在本发明的另一个实施方式中,可以使用宿主生物体如荧光假单胞菌中的通用密码子选择来优化异源多核苷酸序列的表达。可以评估在宿主表达系统中被视为对于特定

氨基酸优选的稀有密码子的百分比和分布。5%和10%的选择率值可作为确定稀有密码子的临界值。例如,表1中列出的密码子在荧光假单胞菌 MB214 基因组中的计算出现率小于5%,因而通常避免用于在荧光假单胞菌宿主中表达的优化基因中。

[0121] 表2. 在荧光假单胞菌 MB214 中出现率低于5%的密码子

[0122]

氨基酸	使用的密码子	% 出现率
G Gly	GGA	3.26
I Ile	ATA	3.05
L Leu	CTA	1.78
	CTT	4.57
	TTA	1.89
	TTG	1.89
R Arg	AGA	1.39
	AGG	2.72
	CGA	4.99
S Ser	TCT	4.28

[0123] 本发明考虑使用所产生毒素的任何编码序列,包括已针对在所使用的假单胞菌宿主细胞中表达进行优化的任何序列。预想使用的序列可以进行任何程度的优化,包括但不限于,优化以消除:在假单胞菌宿主细胞中出现率小于5%的密码子、在假单胞菌宿主细胞中出现率小于10%的密码子、稀有密码子引起的翻译中止、推定的内部RBS序列、G或C核苷酸的延伸重复序列、干扰性二级结构、限制性位点或它们的组合。

[0124] 此外,在本发明的实施中任何有用的分泌前导序列的氨基酸序列可以由任何合适的核酸序列编码。

[0125] 表达系统

[0126] 用于在假单胞菌宿主细胞以及可用于本发明的方法中的宿主细胞中表达异源蛋白质的方法(包括可用的调控序列(例如,启动子、分泌前导序列和核糖体结合位点))在下列文献中有描述,例如,题目均为“Method for Rapidly Screening Microbial Hosts to Identify Certain Strains with Improved Yield and/or Quality in the Expression of Heterologous Proteins”的美国专利申请公开号2008/0269070和美国专利申请系列号12/610,207、题目为“Expression of Mammalian Proteins in Pseudomonas Fluorescens”的美国专利申请公开号2006/0040352和题目为“Process for Improved Protein Expression by Strain Engineering”的美国专利申请公开号2006/0110747,通过引用将所有这些文献全文引入。这些出版物中,还描述了可用于实施本发明方法的细菌宿主菌株,其已被工程化以过表达折叠调节因子或其中为了增加异源蛋白质的表达,已引入蛋白酶突变(包括缺失)。

[0127] 前导序列

[0128] 前导序列在下列文献中有详细描述:美国专利申请公开号2008/0193974和2010/0048864,题目均为“Bacterial Leader Sequences for Increased Expression”及美国专利申请公开号2006/0008877,题目为“Expression systems with Sec-secretion”,其全部通过引用引入本文,以及美国专利申请公开号2008/0269070和美国专利申请系列

号 12/610, 207。

[0129] 在实施方式中,编码分泌前导序列的序列与编码毒素蛋白的序列融合。在实施方式中,分泌前导序列是周质分泌前导序列。在实施方式中,分泌前导序列是天然分泌前导序列。

[0130] 表 3. 示例性分泌前导序列

[0131]

分泌前导 序列	氨基酸序列	SEQ ID NO:
DsbA	MRNLILSAALVTASLFGMTAQA	3
Azu	MFAKLVAVSLLTLASGQLLA	4
Ibp-S31A	MIRDNRLKTSLLRGLTLTLLSLTLLSPAHA	5
Tpr	MNRSSALLAFVFLSGCQAMA	6
CupB2	MLFRTL LASLTF AVIAGLPSTAH A	7

[0132]

CupA2	MSCTRAFKPLLLIGLATLMCSHAFA	8
NikA	MRLAALPLLLAPLFIAPMAVA	9
Pbp A20V	MKLKRLMAAMTFVAAGVATVNAVA	10
DsbC	MRLTQIIAAAAIALVSTFALA	11
TolB	MRNLLRGMLVVICCMAGIAAA	12
Pbp	MKLKRLMAAMTFVAAGVATANAVA	13
Lao	MQNYKKFLLAAAVSMAFSATAMA	14
CupC2	MPPRSIAACLGLLGLLMATQAAA	15
PorE	MKKSTLAVAVTLGAIAQQAGA	16
Pbp	MKLKRLMAAMTFVAAGVATANAVA	17
FlgI	MKFKQLMAMALLALS AVAQA	18
ttg2C	MQNRTVEIGVGLFLLAGILALLLALRVSGLS A	19
CRM197 天然前导 序列	MSRKLFASXLIGALLGIGAPPSAHA	20

[0133] 应该了解,可用于本发明方法中的分泌前导序列并不限于在表 2 中所公开的那些。

[0134] 在实施方式中,分泌前导序列是 Azu、IbpS31A、CupA2 或 PbpA20V。在其它实施方式中,分泌前导序列是 Azu、IbpS31A、CupA2、PbpA20V 或 Pbp。

[0135] 天然 CRM197 通过分泌前导序列从白喉杆菌转运到细胞外间隙,该分泌前导序列被切割以留下 GADD 的氨基末端序列 (SEQ ID NO :21)。为了在荧光假单胞菌中表达后保持 CRM197 的天然氨基酸末端并确保形成二硫键,将蛋白质靶向于周质间隙。

[0136] 启动子

[0137] 根据本发明使用的启动子可以是组成型启动子或调控启动子。常见的可用调控型启动子的实例包括那些源自 lac 启动子 (即 lacZ 启动子) 家族的启动子,尤其是在

DeBoer 的美国专利号 4, 551, 433 中描述的 tac 和 trc 启动子, 以及 Ptac16、Ptac17、PtacII、PlacUV5 和 T7lac 启动子。在一个实施方式中, 所述启动子不是源自于宿主细胞生物体。在某些实施方式中, 所述启动子是源自于大肠杆菌生物体。

[0138] 根据本发明的方法, 诱导型启动子序列可用于调控毒素的表达。在实施方式中, 可用于本发明方法中的诱导型启动子包括那些源自 lac 启动子 (即 lacZ 启动子) 家族的启动子, 尤其是在 DeBoer 的美国专利号 4, 551, 433 中描述的 tac 和 trc 启动子, 以及 Ptac16、Ptac17、PtacII、PlacUV5 和 T7lac 启动子。在一个实施方式中, 所述启动子不是源自于宿主细胞生物体。在某些实施方式中, 所述启动子是源自于大肠杆菌生物体。

[0139] 可用于根据本发明的表达系统中非 lac 型启动子的常见实例包括, 例如, 表 4 中所列出的那些启动子。

[0140] 表 4. 非 lac 启动子的实例

[0141]

启动子	诱导物
P _R	高温
P _L	高温
P _m	烷基 - 或卤代 - 苯甲酸盐
P _u	烷基 - 或卤代 - 甲苯
P _{sal}	水杨酸盐

[0142] 参见, 例如 J. Sanchez-Romero & V. De Lorenzo (1999) Manual of Industrial Microbiology and Biotechnology (A. Demain & J. Davies, eds.) pp. 460-74 (ASM Press, Washington, D. C.) ; H. Schweizer (2001) Current Opinion in Biotechnology, 12 : 439-445 和 R. Slater & R. Williams (2000 Molecular Biology and Biotechnology (J. Walker & R. Rapley, eds.) pp. 125-54 (The Royal Society of Chemistry, Cambridge, UK))。对于所选择的细菌宿主细胞天然的启动子的核苷酸序列的启动子也可用于控制编码靶多肽的转基因的表达, 例如, 假单胞菌邻氨基苯甲酸或苯甲酸操纵子启动子 (Pant, Pben)。也可使用串联启动子, 其中超过一个启动子与另一个启动子共价连接, 无论序列相同或不同, 例如, Pant-Pben 串联启动子 (启动子间杂合体 (interpromoter hybrid)) 或 Plac-Plac 串联启动子, 或无论源自相同或不同的生物体。

[0143] 为了控制启动子作为其部分的基因的转录, 调控启动子利用了启动子调控蛋白。在本发明中使用调控启动子的情况下, 相应的启动子 调控蛋白也是根据本发明的表达系统的部分。启动子调控蛋白的实例包括: 激活剂蛋白, 例如, 大肠杆菌分解代谢物激活剂蛋白、MalT 蛋白; AraC 家族转录激活因子; 阻遏蛋白, 例如, 大肠杆菌 LacI 的蛋白; 和双重功能的调控蛋白, 如大肠杆菌 NagC 蛋白。本领域已知许多调控启动子 / 启动子调控蛋白对。在一个实施方式中, 靶蛋白和目标异源蛋白的表达构建体是在相同调控元件的控制下。

[0144] 启动子调控蛋白与效应子化合物 (即, 可逆或不可逆地与调控蛋白结合的化合物) 相互作用, 以使该蛋白质能够脱离或结合处于该启动子控制下的基因的至少一个 DNA 转录调控区域, 从而允许或阻断启动基因转录的转录酶的作用。效应子化合物被分类为诱导物或共阻遏物, 且这些化合物包括天然的效应子化合物和安慰诱导物化合物。本领域已知很多调控启动子 / 启动子调控蛋白 / 效应子化合物三元组合。虽然效应子化合物可以用在整个细胞培养或发酵过程中, 但在其中在宿主细胞生物质生长到所需数量或密度后才使用调控启动子的优选实施方式中, 将效应子化合物加到培养物中直接或间接导致编码目标

蛋白质或多肽的所需基因的表达。

[0145] 在其中利用 lac 家族启动子的实施方式中, lacI 基因也可以存在于系统中。lacI 基因(其正常情况下是组成型表达的基因)编码 Lac 阻遏蛋白 LacI 蛋白,其结合 lac 家族启动子的 lac 操纵子。因此,在使用 lac 家族启动子的情况下, lacI 基因也可以包括并在表达系统中表达。

[0146] 在假单胞菌中可用的启动子系统被描述在文献中,例如,美国专利申请公开号 2008/0269070,也在上引用献。

[0147] 其他调控元件

[0148] 在实施方式中,产生过程中可溶性蛋白质存在于细胞质或细胞周质中。用于蛋白靶向的分泌前导序列在本文其他地方及文献美国专利申请公开号 2008/0193974、美国专利申请公开号 2006/0008877 和美国专利申请系列号 12/610,207 中进行了描述。

[0149] 其他元件包括但不限于,转录增强子序列、翻译增强子序列、其他启动子、激活剂、转录起始和终止信号、转录终止子、顺反子调节子、多顺反子调节子、标签序列,比如核苷酸序列“标签”和“标签”多肽编码序列,它们促进了对所表达的多肽的鉴别、分离、纯化和/或离析。

[0150] 在实施方式中,表达载体进一步包含与分泌信号的编码序列或与感兴趣的蛋白质或多肽的编码序列邻近的标签序列。在一个实施方式中,该标签序列允许蛋白质的纯化。此标签序列可以是亲和标签,比如六-组氨酸亲和标签(SEQ ID NO:46)。在另一个实施方式中,亲和标签可以是谷胱甘肽-S-转移酶分子。标签也可以是荧光分子,比如 YFP 或 GFP,或此类荧光蛋白的类似物。标签也可以是抗体分子的一部分,或对于可用于纯化的已知结合伴侣的已知抗原或配体。

[0151] 除蛋白编码序列以外,可用于实施本发明方法的表达构建体还可以包括下列可操作地与其连接的调控元件:启动子、核糖体结合位点(RBS)、转录终止子和翻译起始信号和终止信号。根据,例如美国专利申请公开号 2008/0269070 和美国专利申请系列号 12/610,207,可以从可用作表达系统中的宿主的任何物种获得有用的 RBS。已知很多特定的和多样的共有 RBS,例如 D. Frishman 等, Gene 234(2):257-65(8Jul. 1999) 和 B. E. Suzek 等, Bioinformatics 17(12):1123-30(2001 年 12 月)中或其引用的文献中所描述的那些。此外,可以使用天然的或合成的 RBS,例如, EP 0207459(合成的 RBS); O. Ikehata 等, Eur. J. Biochem. 181(3):563-70(1989)(AAGGAAG 的天然 RBS 序列)中所描述的那些。本领域已知可用于本发明中的方法、载体及翻译和转录元件以及其它元素的进一步实例,并在例如 Gilroy 等的美国专利号 5,055,294 和 Gilroy 等的美国专利号 5,128,130; Rammner 等的美国专利号 5,281,532; Barnes 等的美国专利号 4,695,455 和 4,861,595; Gray 等美国专利号 4,755,465 及 Wilcox 的美国专利号 5,169,760,以及许多其它通过引用整体引入的出版物中有描述。

[0152] 宿主菌株

[0153] 设想细菌宿主(包括假单胞菌属)和密切相关的细菌生物体用来实施本发明的方法。在某些实施方式中,假单胞菌宿主细胞为荧光假单胞菌。宿主细胞也可以是大肠杆菌细胞。

[0154] 在实施本发明方法中有用的宿主细胞和构建体可利用本领域已知的和在例如

通过引用整体并入本文的美国专利申请公开号 2009/0325230, "Protein Expression Systems" 等文献中所述的试剂和方法来鉴别或制备。该公开描述了通过将核酸构建体引入到包含染色体 lacI 基因插入序列的营养缺陷型荧光假单胞菌宿主细胞内来产生重组多肽。该核酸构建体包含编码与能够引导核酸在宿主细胞内表达的启动子可操作地连接的重组多肽的核苷酸序列, 且也包含编码营养缺陷型选择标记的核苷酸序列。营养缺陷型选择标记是将原养型恢复成营养缺陷型宿主细胞的多肽。在实施方式中, 细胞对于脯氨酸、尿嘧啶或其组合是营养缺陷的。在实施方式中, 宿主细胞来源于 MB101 (ATCC 保藏号 PTA-7841)。美国专利申请公开号 2009/0325230, "Protein Expression Systems" 和 Schneider 等人, 2005, "Auxotrophic markers pyrF and proC can replace antibiotic markers on protein production plasmids in high-cell-density *Pseudomonas fluorescens* fermentation" *Biotechnol. Progress* 21(2):343-8 中描述了通过删除菌株 MB101 中的 pyrF 基因而构建的尿嘧啶营养缺陷型生产宿主菌株, 这两篇文献通过引用整体并入本文。pyrF 基因是从菌株 MB214 (ATCC 保藏号 PTA-7840) 克隆而来以生成能弥补 pyrF 缺失而恢复原养型的质粒。在特定实施方式中, 使用在荧光假单胞菌宿主细胞中的 pyrF-proC 双重营养缺陷选择标记系统。如所述的 PyrF 生产宿主菌株可用作引入其他期望的基因组改变 (包括本文所述的可用于实施本发明方法的那些) 的背景。

[0155] 在实施方式中, 宿主细胞是假单胞菌目的细胞。在宿主细胞是假单胞菌目细胞时, 它可以是假单胞菌科的成员, 包括假单胞菌属。γ 变形杆菌宿主包括大肠杆菌种的成员和荧光假单胞菌种的成员。

[0156] 其他假单胞菌生物体也可能是有用的。假单胞菌和密切相关的物种包括革兰氏阴性变形杆菌亚群 I, 该亚群 I 包括属于在 R. E. Buchanan 和 N. E. Gibbons (编), *Bergey's Manual of Determinative Bacteriology*, pp. 217-289 (第 8 版, 1974) (The Williams & Wilkins Co., Baltimore, Md., USA) (下文称为 "Bergey (1974)") 中被描述为 "革兰氏阴性需氧杆菌和球菌" 的科和 / 或属的变形杆菌的组。表 5 显示了这些生物体的科和属。

[0157] 表 5. "革兰氏阴性需氧杆菌和球菌" 部分列出的科和属 (Bergey, 1974)

[0158]

I 科. 假单胞菌科	葡糖杆菌属 假单胞菌属 黄单胞菌属 动胶菌属
II 科. 固氮菌科	氮单胞菌属 固氮菌属 拜叶林克氏菌属 德克斯氏菌属
III 科. 根瘤菌科	土壤杆菌属 根瘤菌属
IV 科. 甲基单胞菌科	甲基球菌属 甲基单胞菌属
V 科. 盐杆菌科	盐杆菌属 盐球菌属
其他属	醋杆菌属 产碱杆菌属 博德特菌属 布鲁杆菌属 弗朗西斯氏菌属 栖热菌属

[0159] 假单胞菌和密切相关的细菌通常是定义为“革兰氏(-) 蛋白菌亚群 1”或“革兰氏阴性需氧杆菌和球菌”(Buchanan 和 Gibbons(编)(1974)Bergey's Manual of Determinative Bacteriology, pp. 217-289) 的组的一部分。假单胞菌宿主菌株在例如上文引用的美国专利申请公开号 2006/0040352 等文献中有描述。

[0160] “革兰氏阴性变形杆菌亚群 1”也包括根据分类中使用的标准而分类到该类别下的变形杆菌。该类别也包括以前归类到该部分但现已不再归入该部分的群, 比如, 食酸菌属(Acidovorax)、短波单胞菌属(Brevundimonas)、伯克霍尔德氏菌属(Burkholderia)、产氢噬胞菌属(Hydrogenophaga)、海洋单胞菌属(Oceanimonas)、雷尔氏菌属(Ralstonia) 和寡养单胞菌属(Stenotrophomonas); 通过重新分组属于黄单胞菌属(Xanthomonas)(之前称为其种)的生物体而建立的鞘氨醇单胞菌属(Sphingomonas)(和由其衍生的芽单胞菌属(Blastomonas)); 通过重新分组属于如 Bergey(1974) 所定义的醋杆菌属(Acetobacter)的生物体而建立的酸单胞菌属(Acidomonas)。另外, 宿主可以包括来自假单胞菌属、Pseudomonas enalia(ATCC 14393)、黑色假单胞菌(Pseudomonas nigrifaciens)(ATCC 19375) 和腐败假单胞菌(Pseudomonas putrefaciens)(ATCC 8071)(其已被分别重新分类为 Alteromonas haloplanktis、产黑交替单胞菌(Alteromonas nigrifaciens) 和腐败交替

单胞菌 (*Alteromonas putrefaciens*)) 的细胞。类似地,例如,食酸假单胞菌 (*Pseudomonas acidovorans*) (ATCC 15668) 和 睾 丸 酮 假 单 胞 菌 (*Pseudomonas testosteroni*) (ATCC 11996) 之后被分别重新分类为食酸丛毛单胞菌 (*Comamonas acidovorans*) 和 睾 丸 酮 丛毛单胞菌 (*Comamonas testosteroni*);黑色假单胞菌 (ATCC 19375) 和 *Pseudomonas piscicida* (ATCC 15057) 被分别重新分类为假交替单胞菌 (*Pseudoalteromonas nigrifaciens*) 和杀鱼假交替单胞菌 (*Pseudoalteromonas piscicida*)。“革兰氏阴性变形杆菌亚群 1”也包括被分类为属于任何以下科的变形杆菌:假单胞菌科、固氮菌科(现常被称为同义词—假单胞菌科的“固氮菌群”)、根瘤菌科和甲基单胞菌科(现常被称为同义词—“甲基球菌科”)。因此,除了本文另外描述的那些属以外,属于“革兰氏阴性变形杆菌亚群 1”的更多变形杆菌属包括:1) 嗜氮根瘤菌属的固氮菌群细菌;2) 纤维弧菌属 (*Cellvibrio*)、寡源杆菌属 (*Oligella*) 和 *Teredinibacter* 的假单胞菌科细菌;3) 螯合杆菌属 (*Chelatobacter*)、剑菌属 (*Ensifer*)、韧皮部杆菌属 (*Liberibacter*) (也称为“柑橘黄龙病菌 (*Candidatus Liberibacter*)”) 和中华根瘤菌属 (*Sinorhizobium*) 的根瘤菌科细菌;以及 4) 甲基杆菌属 (*Methylobacter*)、喜热嗜甲基菌属 (*Methylocaldum*)、甲基微菌属 (*Methylomicrobium*)、甲基八迭球菌属 (*Methylosarcina*) 和甲基球菌属 (*Methylosphaera*) 的甲基球菌科细菌。

[0161] 宿主细胞可以选自“革兰氏阴性变形菌亚群 16”。“革兰氏阴性变形菌亚组 16”被定义为以下的假单胞菌种(在括号中显示示例性菌株的 ATCC 或其它保藏号)的变形菌群:*Pseudomonas abietaniphila* (ATCC 700689)、铜绿假单胞菌 (*Pseudomonas aeruginosa*) (ATCC10145)、产碱假单胞菌 (*Pseudomonas alcaligenes*) (ATCC 14909)、鳃败血假单胞菌 (*Pseudomonas anguilliseptica*) (ATCC 33660)、香茅醇假单胞菌 (*Pseudomonas citronellolis*) (ATCC 13674)、变黄假单胞菌 (*Pseudomonas flavescens*) (ATCC 51555)、门多萨假单胞菌 (*Pseudomonas mendocina*) (ATCC 25411)、硝基还原假单胞菌 (*Pseudomonas nitroreducens*) (ATCC 33634)、食油假单胞菌 (*Pseudomonas oleovorans*) (ATCC 8062)、类产碱假单胞菌 (*Pseudomonas pseudoalcaligenes*) (ATCC 17440)、食树脂假单胞菌 (*Pseudomonas resinovorans*) (ATCC 14235)、稻草假单胞菌 (*Pseudomonas straminea*) (ATCC 33636)、伞菊假单胞菌 (*Pseudomonas agarici*) (ATCC 25941)、嗜碱假单胞菌 (*Pseudomonas alcaliphila*)、*Pseudomonas alginovora*、*Pseudomonas andersonii*、铁角蕨假单胞菌 (*Pseudomonas asplenii*) (ATCC 23835)、*Pseudomonas azelaica* (ATCC 27162)、*Pseudomonas beyerinckii* (ATCC 19372)、*Pseudomonas borealis*、北城假单胞菌 (*Pseudomonas boreopolis*) (ATCC 33662)、油菜假单胞菌 (*Pseudomonas brassicacearum*)、*Pseudomonas butanovora* (ATCC 43655)、*Pseudomonas cellulosa* (ATCC 55703)、桔黄假单胞菌 (*Pseudomonas aurantiaca*) (ATCC 33663)、绿针假单胞菌 (*Pseudomonas chlororaphis*) (ATCC 9446、ATCC 13985、ATCC 17418、ATCC 17461)、莓实假单胞菌 (*Pseudomonas fragi*) (ATCC 4973)、海雀假单胞菌 (*Pseudomonas lundensis*) (ATCC 49968)、腐臭假单胞菌 (*Pseudomonas taetrolens*) (ATCC 4683)、青紫菖假单胞菌 (*Pseudomonas cissicola*) (ATCC 33616)、晕斑假单胞菌 (*Pseudomonas coronafaciens*)、*Pseudomonas fiterpeniphila*、伸长假单胞菌 (*Pseudomonas elongata*) (ATCC 10144)、弯曲假单胞菌 (*Pseudomonas flectens*) (ATCC 12775)、生氮假单胞菌 (*Pseudomonas*

azotoformans)、布氏假单胞菌 (*Pseudomonas brenneri*)、*Pseudomonas cedrella*、起皱假单胞菌 (*Pseudomonas corrugata*) (ATCC 29736)、*Pseudomonas extremorientalis*、荧光假单胞菌 (*Pseudomonas fluorescens*) (ATCC 35858)、*Pseudomonas gessardii*、*Pseudomonas libanensis*、孟氏假单胞菌 (*Pseudomonas mandelii*) (ATCC 700871)、边缘假单胞菌 (*Pseudomonas marginalis*) (ATCC 10844)、米氏假单胞菌 (*Pseudomonas migulae*)、霉味假单胞菌 (*Pseudomonas mucidolens*) (ATCC 4685)、*Pseudomonas orientalis*、罗氏假单胞菌 (*Pseudomonas rhodesiae*)、类黄假单胞菌 (*Pseudomonas synxantha*) (ATCC 9890)、托拉斯假单胞菌 (*Pseudomonas tolaasii*) (ATCC 33618)、威隆假单胞菌 (*Pseudomonas veronii*) (ATCC 700474)、弗雷德里克斯堡假单胞菌 (*Pseudomonas frederiksbergensis*)、弯曲假单胞菌 (*Pseudomonas geniculata*) (ATCC 19374)、*Pseudomonas gingeri*、*Pseudomonas graminis*、格氏假单胞菌 (*Pseudomonas grimontii*)、*Pseudomonas halodenitrificans*、嗜盐假单胞菌 (*Pseudomonas halophila*)、栖木槿假单胞菌 (*Pseudomonas hibiscicola*) (ATCC 19867)、赫替假单胞菌 (*Pseudomonas huttiensis*) (ATCC 14670)、噬氢假单胞菌 (*Pseudomonas hydrogenovora*)、*Pseudomonas jessenii* (ATCC 700870)、*Pseudomonas kilonensis*、柳叶刀假单胞菌 (*Pseudomonas lanceolata*) (ATCC 14669)、亚麻假单胞菌 (*Pseudomonas lini*)、划界假单胞菌 (*Pseudomonas marginata*) (ATCC 25417)、臭味假单胞菌 (*Pseudomonas mephitica*) (ATCC 33665)、脱氮假单胞菌 (*Pseudomonas denitrificans*) (ATCC 19244)、穿孔假单胞菌 (*Pseudomonas pertucinogena*) (ATCC 190)、皮克特假单胞菌 (*Pseudomonas pictorum*) (ATCC 23328)、*Pseudomonas psychrophila*、黄褐假单胞菌 (*Pseudomonas filva*) (ATCC 31418)、蒙氏假单胞菌 (*Pseudomonas monteili*) (ATCC 700476)、摩氏假单胞菌 (*Pseudomonas mosselii*)、栖稻假单胞菌 (*Pseudomonas oryzihabitans*) (ATCC 43272)、变形假单胞菌 (*Pseudomonas plecoglossicida*) (ATCC 700383)、恶臭假单胞菌 (*Pseudomonas putida*) (ATCC 12633)、*Pseudomonas reactans*、多刺假单胞菌 (*Pseudomonas spinosa*) (ATCC 14606)、巴利阿里假单胞菌 (*Pseudomonas balearica*)、浅黄假单胞菌 (*Pseudomonas luteola*) (ATCC 43273)；施氏假单胞菌 (*Pseudomonas stutzeri*) (ATCC 17588)、扁桃假单胞菌 (*Pseudomonas amygdali*) (ATCC 33614)、榛色假单胞菌 (*Pseudomonas avellanae*) (ATCC 700331)、番木瓜假单胞菌 (*Pseudomonas caricapapayae*) (ATCC 33615)、菊苣假单胞菌 (*Pseudomonas cichorii*) (ATCC 10857)、天仙果假单胞菌 (*Pseudomonas ficuserectae*) (ATCC 35104)、褐鞘假单胞菌 (*Pseudomonas fuscovaginae*)、苦楝假单胞菌 (*Pseudomonas meliae*) (ATCC 33050)、丁香假单胞菌 (*Pseudomonas syringae*) (ATCC 19310)、绿黄假单胞菌 (*Pseudomonas viridiflava*) (ATCC 13223)、*Pseudomonas thermocarboxydovorans* (ATCC 35961)、耐热假单胞菌 (*Pseudomonas thermotolerans*)、赛维瓦尔假单胞菌 (*Pseudomonas thivervalensis*)、温哥华假单胞菌 (*Pseudomonas vancouverensis*) (ATCC 700688)、*Pseudomonas wisconsinensis* 和厦门假单胞菌 (*Pseudomonas xiamenensis*)。在一个实施方式中，宿主细胞为荧光假单胞菌。

[0162] 宿主细胞也可以选自“革兰氏阴性变形菌亚群 17”。“革兰氏阴性变形菌亚群 17”被定义为本领域中已知为“发荧光的假单胞菌”的变形菌群，包括，例如属于下列假单胞菌种的那些：生氮假单胞菌、布氏假单胞菌、*Pseudomonas cedrella*、起皱假单胞菌、

Pseudomonas extremorientalis、荧光假单胞菌、*Pseudomonas gessardii*、*Pseudomonas libanensis*、孟氏假单胞菌、边缘假单胞菌、米氏假单胞菌、霉味假单胞菌、*Pseudomonas orientalis* ;*Pseudomonas rhodesiae*、产黄假单胞菌、托拉斯假单胞菌和威隆假单胞菌。

[0163] 在实施方式中,假单胞菌宿主细胞为 HslU、HslV、Prc1、DegP1、DegP2、AprA 或它们的组合的表达缺陷的。在实施方式中,宿主细胞为蛋白酶 HslU、HslV、Ppc1、DegP1、DegP2 和 AprA 表达缺陷的,且过表达 DegP2S219A。这种菌株的实例在本文中描述为宿主菌株 2。这些蛋白酶在本领域中已知,并在例如美国专利申请公开号 2006/0110747 中有描述。AprA,一种细胞外 serralysin 型金属蛋白酶金属蛋白酶,由 Maunsell 等,2006,“Complex regulation of AprA metalloprotease in *Pseudomonas fluorescens* M114:evidence for the involvement of iron, the ECF sigma factor, PbrA and pseudobactin M114 siderophore, *Microbiology* 152(Pt 1):29-42 及美国专利申请公开号 2008/0193974 和 2010/0048864 中有描述。

[0164] 在其它实施方式中,假单胞菌宿主细胞过表达 DsbA、DsbB、DsbC 和 DsbD。DsbA、B、C 和 D 是二硫键异构酶,在例如美国专利申请公开号 2008/0269070 和美国专利申请序列号 12/610,207 中描述。

[0165] 在其它实施方式中,假单胞菌的宿主细胞是野生型的,即没有蛋白酶表达缺陷,且也不过表达任何折叠调节因子。

[0166] 蛋白酶表达缺陷的宿主细胞可以具有导致相对于野生型宿主,该蛋白酶的正常活性或表达水平下降的任何修饰。例如,错义或无义突变可导致非活性的蛋白质的表达,且基因缺失可以导致完全没有蛋白质的表达。该基因的上游调节区域的变化可以导致蛋白表达减少或不表达。其他的基因缺陷可影响蛋白质的翻译。蛋白酶的表达也可以是缺陷的,如果对于加工该蛋白酶所需的蛋白质的活性是缺陷的。

[0167] 表 6. 荧光假单胞菌菌株 MB214 蛋白酶

[0168]

类	家族	RXF	基因	评定的功能	位置
天冬氨酸肽酶	A8 (信号肽酶 II 家族)	RXF05383.2	脂蛋白信号肽酶	(ec	细胞质膜

[0169]

		3.4.23.36)
A24 (IV 型 prepilin 肽酶家族)		
RXF05379.1	4 型 prepilin 肽酶 pild	细胞膜 (ec 3.4.99.-)
半胱氨酸肽酶		
C15 (焦谷氨酰肽酶 I 家族)		
RXF02161.1	吡咯烷酮羧酸酯肽酶	细胞质 (ec 3.4.19.3)
C40		
RXF01968.1	侵袭相关蛋白, P60	信号肽
RXF04920.1	侵袭相关蛋白, P60	细胞质
RXF04923.1	磷酸酶-相关蛋白 papq	信号肽
C56 (PfpI 内肽酶家族)		
RXF01816.1	蛋白酶 I (ec 3.4.-.-)	非分泌的
金属肽酶		
M1		
RXF08773.1	细胞膜丙氨酸氨肽酶(ec	非分泌的 3.4.11.2)
M3		
RXF00561.2 prlC	寡肽酶 A (ec 3.4.24.70)	细胞质
RXF04631.2	Zn-依赖性寡肽酶	细胞质
M4 (嗜热菌蛋白酶家族)		
RXF05113.2	细胞外金属蛋白酶前体	细胞外 (ec 3.4.24.-)
M41 (FtsH 内肽酶家族)		
RXF05400.2	细胞分裂蛋白 ftsH (ec	细胞膜 3.4.24.-)
M10		
RXF04304.1	Serralysin (ec 3.4.24.40)	细胞外
RXF04500.1	Serralysin (ec 3.4.24.40)	细胞外
RXF01590.2	Serralysin (ec 3.4.24.40)	细胞外
RXF04497.2	Serralysin (ec 3.4.24.40)	细胞外
RXF04495.2	Serralysin (ec 3.4.24.40)	细胞外
RXF02796.1	Serralysin (ec 3.4.24.40)	细胞外
M14 (羧肽酶 A 家族)		
RXF09091.1	锌 - 羧肽酶前体 (ec	细胞质 3.4.17.-)

[0170]

M16 (溶加压素家族)		
RXF03441.1	辅酶 pqq 合成蛋白 F (ec 3.4.99.-)	非分泌的
RXF01918.1	锌蛋白酶 (ec 3.4.99.-)	信号肽
RXF01919.1	锌蛋白酶(ec 3.4.99.-)	周质
RXF03699.2	加工肽酶 (ec 3.4.24.64)	信号肽
M17 (亮氨酰氨肽酶家族)		
RXF00285.2	胞液氨肽酶(ec 3.4.11.1)	非分泌的
M18		
RXF07879.1	天冬氨酰氨肽酶 (ec 3.4.11.21)	细胞质
M20		
RXF00811.1	dapE 琥珀酰-二氨基庚二酸脱琥珀酰酶(ec 3.5.1.18)	细胞质
RXF04052.2	Xaa-His 二肽酶 (ec 3.4.13.3)	信号肽
RXF01822.2	羧肽酶 G2 前体 (ec 3.4.17.11)	信号肽
RXF09831.2: :RXF04892.1	N-酰基-L-氨基酸氨基水解酶(ec 3.5.1.14)	信号肽
M28 (氨肽酶 Y 家族)		
RXF03488.2	碱性磷酸酶同功酶转化蛋白前体 (ec 3.4.11.-)	外膜
M42 (谷氨酰氨肽酶家族)		
RXF05615.1	解锁氨肽酶(deblocking aminopeptidase) (ec 3.4.11.-)	非分泌的
M22		
RXF05817.1	O-唾液酸蛋白内肽酶(ec 3.4.24.57)	细胞外
RXF03065.2	糖蛋白酶蛋白家族	非分泌的
M23		
RXF01291.2	细胞壁内肽酶, 家族 M23/M37	信号肽
RXF03916.1	与金属内肽酶相关的膜蛋白	信号肽
RXF09147.2	细胞壁内肽酶, 家族	信号肽

[0171]

		M23/M37	
M24			
	RXF04693.1	甲硫氨酸氨肽酶 (ec 3.4.11.18)	细胞质
	RXF03364.1	甲硫氨酸氨肽酶 (ec 3.4.11.18)	非分泌的
	RXF02980.1	Xaa-Pro 氨肽酶 (ec 3.4.11.9)	细胞质
	RXF06564.1	Xaa-Pro 氨肽酶 (ec 3.4.11.9)	细胞质
M48 (Ste24 内肽酶家族)			
	RXF05137.1	热休克蛋白 HtpX	细胞膜
	RXF05081.1	锌金属蛋白酶 (ec 3.4.24.-)	信号肽
M50 (S2P 蛋白酶家族)			
	RXF04692.1	膜金属蛋白酶	细胞膜
丝氨酸氨肽酶			
S1 (糜蛋白酶家族)			
	RXF01250.2	蛋白酶 do (ec 3.4.21.-)	周质
	RXF07210.1	蛋白酶 do (ec 3.4.21.-)	周质
S8 (枯草杆菌蛋白酶家族)			
	RXF06755.2	丝氨酸蛋白酶 (ec 3.4.21.-)	非分泌的
	RXF08517.1	丝氨酸蛋白酶 (ec 3.4.21.-)	细胞外
	RXF08627.2	细胞外丝氨酸蛋白酶 (ec 3.4.21.-)	信号肽
	RXF06281.1	细胞外丝氨酸蛋白酶前体 (ec 3.4.21.-)	非分泌的
	RXF08978.1	细胞外丝氨酸蛋白酶 (ec 3.4.21.-)	外膜
	RXF06451.1	丝氨酸蛋白酶 (ec 3.4.21.-)	信号肽
S9 (脯氨酰寡肽酶家族)			
	RXF02003.2	蛋白酶 ii (ec 3.4.21.83)	周质
	RXF00458.2	水解酶	非分泌的
S11 (D-Ala-D-Ala 羧肽酶 A 家族)			

[0172]

RXF04657.2		D-丙氨酰-D-丙氨酸-内 周质 肽酶(ec 3.4.99.-)
RXF00670.1		D-丙氨酰-D-丙氨酸羧肽 细胞膜 酶(ec 3.4.16.4)
S13 (D-Ala-D-Ala 肽酶 C 家族)		
RXF00133.1		D-丙氨酰-内消旋-二氨 外膜 基庚二酸内肽酶 (ec 3.4.-.)
RXF04960.2		D-丙氨酰-内消旋-二氨 信号肽 基庚二酸内肽酶 (ec 3.4.-.)
S14 (ClpP 内肽酶家族)		
RXF04567.1	clpP	atp 依赖性 Clp 蛋白酶蛋 非分泌的 白 水解 亚 基 (ec 3.4.21.92)
RXF04663.1	clpP	atp 依赖性 Clp 蛋白酶蛋 细胞质 白 水解 亚 基 (ec 3.4.21.92)
S16 (lon 蛋白酶家族)		
RXF04653.2		atp 依赖性蛋白酶 La (ec 细胞质 3.4.21.53)
RXF08653.1		atp 依赖性蛋白酶 La (ec 细胞质 3.4.21.53)
RXF05943.1		atp 依赖性蛋白酶 (ec 细胞质 3.4.21.53)
S24 (LexA 家族)		
RXF00449.1		LexA 阻 遏 物 (ec 非分泌的 3.4.21.88)
RXF03397.1		LexA 阻 遏 物 (ec 细胞质 3.4.21.88)
S26 (信号肽酶 I 家族)		
RXF01181.1		信 号 肽 酶 I (ec 细胞膜 3.4.21.89)
S33		
RXF05236.1	pip3	脯氨酸亚氨基肽酶 (ec 非分泌的 3.4.11.5)

[0173]

	RXF04802.1	pip1	脯氨酸亚氨基肽酶 (ec 3.4.11.5)	非分泌的
	RXF04808.2	pip2	脯氨酸亚氨基肽酶 (ec 3.4.11.5)	细胞质
S41 (C-末端 加工肽酶家族)				
	RXF06586.1		尾特异性蛋白酶 (ec 3.4.21.-)	信号肽
	RXF01037.1		尾特异性蛋白酶 (ec 3.4.21.-)	信号肽
S45				
	RXF07170.1	pacB2	青霉素酰化酶 (ec 3.5.1.11)	信号肽
	RXF06399.2	pacB1	青霉素酰化酶 ii (ec 3.5.1.11)	信号肽
S49 (蛋白酶 IV 家族)				
	RXF06993.2		可能的蛋白酶 sohb (ec 3.4.-.)	非分泌的
	RXF01418.1		蛋白酶 iv (ec 3.4.-.)	非分泌的
S58 (DmpA 氨肽酶家族)				
	RXF06308.2		D-氨肽酶 (ec 3.4.11.19)	细胞膜
苏氨酸肽酶				
T1 (蛋白酶体家族)				
	RXF01961.2	hslV	atp 依赖性蛋白酶 hslV (ec 3.4.25.-)	细胞质
T3 (γ-谷氨酰转移酶家族)				
	RXF02342.1	ggt1	γ -谷氨酰转肽酶 (ec 2.3.2.2)	周质
	RXF04424.2	ggt2	γ -谷氨酰转肽酶 (ec 2.3.2.2)	周质
未分类肽酶				
U32				
	RXF00428.1		蛋白酶 (ec 3.4.-.)	细胞质
	RXF02151.2		蛋白酶 (ec 3.4.-.)	细胞质
U61				
	RXF04715.1		胞壁酰四肽羧肽酶 (ec 3.4.17.13)	非分泌的
U62				
	RXF04971.2	pmbA	PmbA 蛋白	细胞质

RXF04968.2	TldD 蛋白	细胞质
非 MEROPS 蛋白酶		
RXF00325.1	阻遏蛋白 C2	非分泌的
RXF02689.2	微粒体二肽酶 (ec 3.4.13.19)	细胞质
RXF02739.1	膜二肽酶(3.4.13.19)	信号肽
RXF03329.2	假定细胞溶质蛋白	细胞质
RXF02492.1	Xaa-Pro 二肽酶 (ec 3.4.13.9)	细胞质
RXF04047.2	caax 氨基末端蛋白酶家	细胞膜族
RXF08136.2	蛋白酶(转谷氨酰胺酶样蛋白)	细胞质
RXF09487.1	锌金属蛋白酶 (ec 3.4.24.-)	非分泌的

[0175] 某些蛋白酶可以有蛋白酶和伴侣蛋白样活性。当这些蛋白酶对蛋白质的产率和 / 或质量造成负面影响时, 删除他们可能是有利的, 且当他们的伴侣蛋白活性对蛋白产率和 / 或质量产生正面影响时, 可以将他们过表达。这些蛋白酶包括, 但不限于: Hsp100 (Clp/Hsl) 家族成员 RXF04587.1 (clpA)、RXF08347.1、RXF04654.2 (clpX)、RXF04663.1、RXF01957.2 (hslU)、RXF01961.2 (hslV); 肽酰 - 脯氨酰 - 顺 - 反式异构酶家族成员 RXF05345.2 (ppiB); 金属肽酶 M20 家族成员 RXF04892.1 (氨基水解酶); 金属肽酶 M24 家族成员 RXF04693.1 (甲硫氨酸氨肽酶) 和 RXF03364.1 (甲硫氨酸氨肽酶) 及丝氨酸肽酶 S26 信号肽酶 I 家族成员 RXF01181.1 (信号肽酶)。

[0176] 表 7. 荧光假单胞菌菌株 MB214 蛋白质折叠调节因子

[0177]

ORF ID	基因	功能	家族	位置
GroES/EL				
RXF02095.1	groES	伴侣蛋白	Hsp10	细胞质
RXF06767.1:: Rxf02090	groEL	伴侣蛋白	Hsp60	细胞质

[0178]

RXF01748.1	ibpA	小的热休克蛋白 (sHSP) IbpA PA3126 ; 用作 GroESL 折叠的支架	Hsp20	细胞质
RXF03385.1	hscB	伴侣蛋白 hscB	Hsp20	细胞质
<u>Hsp70 (DnaK/J)</u>				
RXF05399.1	dnaK	伴侣蛋白	Hsp70	周质
RXF06954.1	dnaK	伴侣蛋白	Hsp70	细胞质
RXF03376.1	hscA	伴侣蛋白	Hsp70	细胞质
RXF03987.2	cbpA	弯曲 dna-结合蛋白, dnaJ 样活性	Hsp40	细胞质
RXF05406.2	dnaJ	伴侣蛋白 dnaJ	Hsp40	细胞质
RXF03346.2	dnaJ	分子伴侣 (DnaJ 家族)	Hsp40	非分泌的
RXF05413.1	grpE	热休克蛋白 GrpE PA4762	GrpE	细胞质
<u>Hsp100 (Clp/Hsl)</u>				
RXF04587.1	clpA	atp-依赖性 clp 蛋白酶 atp-结合亚基 clpA	Hsp100	细胞质
RXF08347.1	clpB	ClpB 蛋白	Hsp100	细胞质
RXF04654.2	clpX	atp-依赖性 clp 蛋白酶 atp-结合亚基 clpX	Hsp100	细胞质
RXF04663.1	clpP	atp-依赖性 Clp 蛋白酶蛋白水解亚基(ec 3.4.21.92)	MEROPS 肽酶家族 S14	细胞质
RXF01957.2	hslU	atp-依赖性 hsl 蛋白酶 atp-结合亚基 hslU	Hsp100	细胞质
RXF01961.2	hslV	atp-依赖性 hsl 蛋白酶蛋白水解亚基	MEROPS 肽酶亚族 T1B	细胞质
<u>Hsp33</u>				
RXF04254.2	yrfI	33 kDa 伴侣蛋白(热休克蛋白 33 同源物) (HSP33).	Hsp33	细胞质
<u>Hsp90</u>				

[0179]

RXF05455.2	htpG	伴侣蛋白 htpG	Hsp90	细胞质
<u>SecB</u>				
RXF02231.1	secB	分泌特异伴侣蛋白 SecB	SecB	非分泌的
<u>二硫键异构酶</u>				
RXF07017.2	dsbA	二硫化物异构酶	DSBA 氧化 - 还原酶	细胞质
RXF08657.2	dsbA/ dsbC/ dsbG/ fernA	二硫化物异构酶	DSBA 氧化 - 还原酶	细胞质
RXF01002.1	dsbA/ dsbC	二硫化物异构酶	DSBA 氧化 - 还原酶/硫氧还蛋白	周质
RXF03307.1	dsbC	二硫化物异构酶	谷氧还蛋白/硫氧还蛋白	周质
RXF04890.2	dsbG	二硫化物异构酶	谷氧还蛋白/硫氧还蛋白	周质
RXF03204.1	dsbB	二硫键形成蛋白 B (二硫化物氧化还原酶).	DSBA 氧化 - 还原酶	周质
RXF04886.2	dsbD	巯基:二硫键互换蛋白 dsbD	DSBA 氧化 - 还原酶	周质
<u>肽酰-脯氨酰顺-反式异构酶</u>				
RXF03768.1	ppiA	肽酰-脯氨酰顺-反式异构酶 A (ec 5.2.1.8)	PPIase: 亲环蛋白型	周质

[0180]

RXF05345.2	ppiB	肽酰-脯氨酰顺-反式异构酶 B	PPIase: 细胞质 亲环蛋白 型
RXF06034.2	fkIB	肽酰-脯氨酰顺-反式异构酶 FkIB	PPIase: 外膜 FKBP 型
RXF06591.1	fkIB/ fkbP	fk506 结合蛋白肽酰-脯氨酰顺-反式异构酶(EC 5.2.1.8)	PPIase: 周质 FKBP 型
RXF05753.2	fkIB; fkbP	肽酰-脯氨酰顺-反式异构酶(ec 5.2.1.8)	PPIase: 外膜 FKBP 型
RXF01833.2	slyD	肽酰-脯氨酰顺-反式异构酶 SlyD	PPIase: 非分泌 FKBP 型的
RXF04655.2	tig	触发因子, ppiase(ec 5.2.1.8)	PPIase: 细胞质 FKBP 型
RXF05385.1	yaad	可能的 FKBP-型 16 kDa 肽酰-脯氨酰顺-反式异构酶(EC 5.2.1.8) (PPIase) (旋转异构酶).	PPIase: 非分泌 FKBP 型的
RXF00271.1		肽酰-脯氨酰顺-反式异构酶(ec 5.2.1.8)	PPIase: 非分泌 FKBP 型的
<u>pili 组装伴侣蛋白 (papD 样)</u>			
RXF06068.1	cup	伴侣蛋白 cup	pili 组装 周质 papD
RXF05719.1	ecpD	伴侣蛋白 ecpD	pili 组装 信号肽 papD
RXF05319.1	ecpD	Hnr 蛋白	pili 组装 周质 chaperone
RXF03406.2	ecpD; csuC	伴侣蛋白 ecpD	pili 组装 信号肽 papD
RXF04296.1	ecpD; cup	伴侣蛋白 ecpD	pili 组装 周质 papD
RXF04553.1	ecpD; cup	伴侣蛋白 ecpD	pili 组装 周质 papD
RXF04554.2	ecpD; cup	伴侣蛋白 ecpD	pili 组装 周质 papD

[0181]

RXF05310.2	ecpD; cup	伴侣蛋白 ecpD	pili 组装 周质 papD
RXF05304.1	ecpD; cup	伴侣蛋白 ecpD	pili 组装 周质 papD
RXF05073.1	gltF	格兰氏阴性 pili 组装伴侣 周质功能	pili 组装 信号肽 papD
II 型分泌复合物			
RXF05445.1	YacJ	组氨醇-磷酸氨基转移酶 (ec 2.6.1.9)	II 类磷酸 细胞膜 吡哆醛依 赖性转氨 酶家族, 组氨醇- 磷酸转氨 酶亚家族
RXF05426.1	SecD	蛋白转位酶亚基 secD	II 型分泌 细胞膜 复合物
RXF05432.1	SecF	蛋白转位酶亚基 secF	II 型分泌 细胞膜 复合物
二硫键还原酶			
RXF08122.2	trxC	硫氧还蛋白 2	二硫键还 细胞质 原酶
RXF06751.1	Gor	谷胱甘肽还原酶 (EC 1.8.1.7) (GR) (GRase) PA2025	二硫键还 细胞质 原酶
RXF00922.1	gshA	谷氨酸-半胱氨酸连接酶 (ec 6.3.2.2) PA5203	二硫键还 细胞质 原酶

[0182] 高通量筛选

[0183] 在一些实施方式中,可进行高通量筛选以确定用于表达可溶性重组毒素蛋白的最佳条件。在筛选中可变的条件包括,例如,宿主细胞、宿主细胞的遗传背景(例如,不同蛋白酶或缺失)、表达构建体中的启动子类型、与编码重组蛋白的序列融合的秘密前导序列类型、生长温度、当使用诱导型启动子时进行诱导的 OD、当使用 lacZ 启动子时用于诱导的 IPTG 浓度、蛋白质诱导的持续时间、向培养物中加入诱导剂后的生长温度、培养物搅拌速度、质粒保持的选择的方法、容器中的培养物体积和细胞裂解方法。

[0184] 在一些实施方式中,提供了宿主菌株的文库(或“阵列”),其中文库中的各个菌株(或“宿主细胞群”)已进行遗传修饰以调节宿主细胞中的一个或多个靶基因的表达。与阵列中表型明显不同的宿主细胞的其他群体相比,可基于表达的目标蛋白质的数量、质量和/或位置来鉴定和或选择“最佳宿主菌株”或“最佳表达系统”。因此,最佳宿主菌株是根据期

望的标准产生目标多肽的菌株。尽管期望的规格将根据所产生的多肽而变化,但是该规格包括蛋白质的质量和 / 或数量,例如,蛋白质是隔离的还是分泌的,数量如何,蛋白质是否适当地或期望地加工和 / 或折叠,等等。在实施方式中,提高的或期望的质量可以是具有分泌前导序列的高保真切割和低水平降解的毒素蛋白的产生。在实施方式中,最佳宿主菌株或最佳表达系统获得某一绝对水平的产率或相对于指示菌株(即用于对照的菌株)所获得的产率的某一水平的产率,其特征为可溶性异源蛋白质的量或数量、可回收的异源蛋白质的量或数量、适当加工的异源蛋白质的量或数量、适当折叠的异源蛋白质的量或数量、活性异源蛋白质的量或数量和 / 或异源蛋白质的总量和数量。

[0185] 筛选微生物宿主以确定异源蛋白质表达的产率和 / 或数量提高的菌株的方法在例如美国专利申请公开号 20080269070 中有描述。

[0186] 发酵形式

[0187] 根据本发明的表达系统可以以任何发酵模式培养。例如,分批、补料分批、半连续和连续发酵模式可以用于本发明中。

[0188] 在实施方式中,发酵培养基可以选自丰富培养基、基本培养基和矿物盐培养基。在其它实施方式中,可以选择基本培养基或矿物盐培养基。在某些实施方式中,可以选择矿物盐培养基。

[0189] 矿物盐培养基由矿物盐和碳源组成,例如,葡萄糖、蔗糖或甘油。矿物盐培养基的实例包括,例如, M9 培养基、假单胞菌培养基(ATCC 179)及 Davis 和 Mingioli 培养基(见 B D Davis & E S Mingioli(1950) J. Bact. 60 :17-28)。用于制备矿物盐培养基的矿物盐包括选自以下的那些:例如,磷酸钾、硫酸盐或氯化铵、硫酸镁或氯化镁及微量矿物质,如氯化钙、硼酸盐及铁、铜、锰和锌的硫酸盐。通常情况下,矿物盐培养基中没有有机氮源,如蛋白胨、胰蛋白胨、氨基酸或酵母提取物。相反,使用无机氮源,且其可以选自例如,铵盐、氨水以及气态氨。矿物盐培养基中通常会包含葡萄糖或甘油作为碳源。在矿物盐培养基相比,基本培养基也可以包含矿物盐和碳源,但可补充,例如,低水平的氨基酸、维生素、蛋白胨或其它成分,尽管这些成分以非常低的水平添加。可以使用本领域,例如,美国专利申请公开号 2006/0040352、上述引用和引入的文献中描述的方法制备培养基。可用于本发明的方法中的培养过程和矿物盐培养基的细节见 Riesenber, D 等,1991, " High cell density cultivation of Escherichia coli at controlled specific growth rate," J. Biotechnol. 20(1) :17-27 中的描述。

[0190] 在实施方式中,可以在生物反应器培养中进行生产。通过添加氨,培养物可以在例如,含有矿物盐培养基的最多 2 升的生物反应器中生长,并维持温度为 32°C 和 pH 值为 6.5。通过增加搅拌并进入到发酵罐中的鼓泡空气和氧气流使溶解氧保持过量。甘油可以在整个发酵过程中输送到培养基物以维持在过量水平。在实施方式中,保持这些条件直到达到用于诱导的培养细胞目标密度,例如,575nm(A_{575}) 的光密度,此时添加 IPTG 以开始靶蛋白生产。可以理解,可以各自改变诱导时的细胞密度、IPTG 浓度、pH 值和温度以确定最佳表达条件。在实施方式中,诱导时的细胞密度变化范围是 40 至 200 个吸光度单位(AU) 的 A_{575} 。IPTG 浓度的变化范围为 0.02 至 1.0mM、pH 变化范围为 6-7.5、温度变化范围为 20 到 35°C。16-24 小时后,通过离心从来自各生物反应器收获培养物,且细胞沉淀可以在 -80°C 下冷冻。可通过例如 SDS-CGE 分析样品的产物形成。

[0191] 发酵可以以任何规模进行。根据本发明的表达系统可用于任何规模的重组蛋白表达。因此,可以使用,例如微升规模、毫升规模、厘升规模、分升规模的发酵体积,且可以使用1升规模和更大的发酵体积。

[0192] 在实施方式中,发酵体积约为1升或大于1升。在实施方式中,发酵体积为约1升至约100升。在实施方式中,发酵体积为约1升、约2升、约3升、约4升、约5升、约6升、约7升、约8升、约9升或约10升。在实施方式中,发酵体积为约1升至约5升、约1升至约10升、约1升至约25升、约1升至约50升、约1升至约75升、约10升至约25升、约25升至约50升或约50升至约100升。在其它实施方式中,发酵体积大约为或大于5升、10升、15升、20升、25升、50升、75升、100升、200升、500升、1000升、2000升、5000升、10000升或50,000升。

[0193] 细菌生长条件

[0194] 可用于所提供的发明的方法中的生长条件可包括约4℃至约42℃的温度和约5.7至约8.8的pH。当使用具有lacZ启动子的表达构建体时,可通过向培养物中加IPTG至终浓度为约0.01mM至约1.0mM来诱导表达。

[0195] 可以使用pH缓冲剂和本领域技术人员已知的方法来维持培养物的pH。在培养过程中对pH的控制也可使用氨水来实现。在实施方式中,培养物的pH为约5.7至约8.8。在某些实施方式中,pH为约5.7、5.8、5.9、6.0、6.1、6.2、6.3、6.4、6.5、6.6、6.7、6.8、6.9、7.0、7.1、7.2、7.3、7.4、7.5、7.6、7.7、7.8、7.9、8.0、8.1、8.2、8.3、8.4、8.5、8.6、8.7或8.8。在其他实施方式中,pH为约5.7至5.9、5.8至6.0、5.9至6.1、6.0至6.2、6.1至6.3、6.2至6.5、6.4至6.7、6.5至6.8、6.6至6.9、6.7至7.0、6.8至7.1、6.9至7.2、7.0至7.3、7.1至7.4、7.2至7.5、7.3至7.6、7.4至7.7、7.5至7.8、7.6至7.9、7.7至8.0、7.8至8.1、7.9至8.2、8.0至8.3、8.1至8.4、8.2至8.5、8.3至8.6、8.4至8.7或8.5至8.8。在另外其它的实施方式中,pH为约5.7至6.0、5.8至6.1、5.9至6.2、6.0至6.3、6.1至6.4或6.2至6.5。在某些实施方式中,pH为约5.7至约6.25。

[0196] 在实施方式中,生长温度维持在约4℃至约42℃。在某些实施方式中,生长温度为约4℃、约5℃、约6℃、约7℃、约8℃、约9℃、约10℃、约11℃、约12℃、约13℃、约14℃、约15℃、约16℃、约17℃、约18℃、约19℃、约20℃、约21℃、约22℃、约23℃、约24℃、约25℃、约26℃、约27℃、约28℃、约29℃、约30℃、约31℃、约32℃、约33℃、约34℃、约35℃、约36℃、约37℃、约38℃、约39℃、约40℃、约41℃或约42℃。在其他实施方式中,生长温度维持在约25℃至约27℃、约25℃至约28℃、约25℃至约29℃、约25℃至约30℃、约25℃至约31℃、约25℃至约32℃、约25℃至约33℃、约26℃至约28℃、约26℃至约29℃、约26℃至约30℃、约26℃至约31℃、约26℃至约32℃、约27℃至约29℃、约27℃至约30℃、约27℃至约31℃、约27℃至约32℃、约26℃至约33℃、约28℃至约30℃、约28℃至约31℃、约28℃至约32℃、约29℃至约31℃、约29℃至约32℃、约29℃至约33℃、约30℃至约32℃、约30℃至约33℃、约31℃至约33℃、约31℃至约32℃、约30℃至约33℃或约32℃至约33℃。在其他实施方式中,温度在培养过程中改变。在一个实施方式中,在向培养物中加入用于诱导构建体表达的试剂例如IPTG之前,温度维持在约30℃。在加入诱导剂后,温度降至约25℃。

[0197] 诱导

[0198] 如本文别处所述,诱导型启动子可用于表达构建体中以控制重组毒素蛋白的表达,例如, lac 启动子。在 lac 启动子衍生物或家族成员例如 tac 启动子的情况下,效应化合物是诱导物,比如像 IPTG (异丙基-β-D-1-硫代吡喃半乳糖苷,也称为“异丙基硫代半乳糖苷”)的安慰诱导物。在实施方式中,使用 lac 启动子衍生物,并且当细胞密度达到由约 80 至约 160 的 OD₅₇₅ 确定的水平时,通过加入 IPTG 至终浓度为约 0.01mM 至约 1.0mM 来诱导重组蛋白的表达。在实施方式中,在重组蛋白的培养物诱导时,OD₅₇₅ 可以为约 80、约 90、约 100、约 110、约 120、约 130、约 140、约 150、约 160、约 170、约 180。在其他实施方式中,OD₅₇₅ 为约 80 至约 100、约 100 至约 120、约 120 至约 140、约 140 至约 160。在其他实施方式中,OD₅₇₅ 为约 80 至约 120、约 100 至约 140 或约 120 至约 160。在其他实施方式中,OD₅₇₅ 为约 80 至约 140 或约 100 至约 160。细胞密度可通过其他方法测量并且以其他单位表示,例如,每单位体积的细胞。例如,荧光假单胞菌培养物的约 80 至约 160 的 OD₅₇₅ 等同于约 8×10^{10} 至约 1.6×10^{11} 菌落形成单位/mL 或 35-70g/L 细胞干重。在实施方式中,培养物诱导时的细胞密度等同于如本文所述根据 OD₅₇₅ 的吸光度而得出的细胞密度,与用于测定细胞密度的方法或测量单位无关。本领域技术人员将知道如何为任何细胞培养物进行适当的转换。

[0199] 在实施方式中,培养物的最终 IPTG 浓度为约 0.01mM、约 0.02mM、约 0.03mM、约 0.04mM、约 0.05mM、约 0.06mM、约 0.07mM、约 0.08mM、约 0.09mM、约 0.1mM、约 0.2mM、约 0.3mM、约 0.4mM、约 0.5mM、约 0.6mM、约 0.7mM、约 0.8mM、约 0.9mM 或约 1mM。在其他实施方式中,培养物的最终 IPTG 浓度为约 0.08mM 至约 0.1mM、约 0.1mM 至约 0.2mM、约 0.2mM 至约 0.3mM、约 0.3mM 至约 0.4mM、约 0.2mM 至约 0.4mM、约 0.08 至约 0.2mM 或约 0.1 至 1mM。

[0200] 在其中使用非 lac 型启动子的实施方式中,如本文中和文献中所述,可使用其他诱导物或效应物。在一个实施方式中,该启动子是组成型启动子。

[0201] 在加入诱导剂后,培养物可以生长一段时间,例如约 24 小时,在这段时间内重组蛋白得到表达。在加入诱导剂后,培养物可以生长约 1 小时、约 2 小时、约 3 小时、约 4 小时、约 5 小时、约 6 小时、约 7 小时、约 8 小时、约 9 小时、约 10 小时、约 11 小时、约 12 小时、约 13 小时、约 14 小时、约 15 小时、约 16 小时、约 17 小时、约 18 小时、约 19 小时、约 20 小时、约 21 小时、约 22 小时、约 23 小时、约 24 小时、约 36 小时或约 48 小时。在向培养物中加入诱导剂后,培养物可以生长约 1 至 48 小时、约 1 至 24 小时、约 10 至 24 小时、约 15 至 24 小时或约 20 至 24 小时。可通过离心浓缩细胞培养物,并且将培养物沉淀再悬浮在适合后续裂解程序的缓冲液或溶液中。

[0202] 在实施方式中,利用用于高压机械细胞破碎的设备(其可商购,例如, Microfluidics 微射流器 (Microfluidics Microfluidizer)、Constant 细胞破碎机 (Constant Cell Disruptor)、Niro-Soavi 匀浆器或 APV-Gaulin 匀浆器)将细胞破碎。例如,可以利用超声处理使表达重组蛋白的细胞破碎。任何本领域已知的适合于裂解细胞的方法可用于释放可溶性部分。例如,在实施方式中,可使用化学和/或酶细胞裂解试剂,比如细胞壁分解酶和 EDTA。本发明的方法中也设想使用冷冻的或事先储存的培养物。培养物在裂解之前可以进行 OD- 标准化。例如,可以将细胞标准化为约 10、约 11、约 12、约 13、约 14、约 15、约 16、约 17、约 18、约 19 或约 20 的 OD₆₀₀。

[0203] 可利用任何合适的设备和方法来进行离心。为了将可溶性部分与不溶性部分中分离而对细胞培养物或裂解产物进行离心在本领域中是众所周知的。例如,已裂解的细胞可

以在 20,800x g 下离心 20 分钟 (4℃下),并用手动的或自动的液体处理来除去上清液。沉淀(不溶性)部分再悬浮于缓冲溶液例如 pH 7.4 的磷酸盐缓冲盐水(PBS)中。例如,可利用诸如连接到悬臂式搅拌器上的叶轮、磁力搅拌棒、摇摆振荡器等设备来进行再悬浮。

[0204] “可溶性部分”,即裂解液离心后获得的可溶性上清液,以及“不溶性部分”,即裂解液离心后获得的沉淀物,是通过培养物的裂解和离心而获得的。这两部分也可分别称作“第一可溶性部分”和“第一不溶性部分”。

[0205] 产物的评估

[0206] 本领域中已知许多用于表征蛋白质的测定方法。本文中预期任何用于表征重组毒素蛋白的产率或质量的合适方法的使用。

[0207] 蛋白质产率

[0208] 可以通过本领域技术人员已知的方法测定本文所述的任何纯化部分中的蛋白质产率,例如,通过毛细管凝胶电泳(CGE)和 Western 印迹分析。如本文所述的和本领域中已知的,活性测定也可以提供涉及蛋白质产率的信息。在实施方式中,本领域已知的这些方法或任何其他方法用于评价蛋白质的适当加工,例如,适当的分泌前导序列切割。

[0209] 可用的蛋白质产率的量度包括,例如,每体积培养物的重组蛋白质量(例如,克或毫克蛋白质/升培养物)、细胞裂解后得到的不溶性沉淀中测定的重组蛋白质的百分比或分数(例如,提取的上清液中重组蛋白质的量/不溶性成分中蛋白质的量)、活性蛋白的百分比或分数(例如,活性蛋白的量/分析中使用的蛋白质量)、总细胞蛋白质(tcp)的百分比或分数、蛋白质量/细胞和干生物质的百分比或比例。在实施方式中,如本文所述的蛋白质产率的测量基于所获得的可溶性蛋白质的量或活性蛋白质的量或此两者。

[0210] 在其中产率用培养物体积表示的实施方式中,可以考虑培养物细胞密度,尤其是在不同培养物之间进行产率比较时。

[0211] 在实施方式中,本发明的方法可用于获得约 0.2 克/升至约 12 克/升的可溶性和/或活性和/或适当加工的(例如,分泌前导序列正确切割的)重组毒素蛋白或亚基蛋白质产率。在实施方式中,产率为约 0.5 克/升至约 12 克/升。在某些实施方式中,重组蛋白或亚基蛋白质产率为约 0.2g/L、约 0.3g/L、约 0.4g/L、约 0.5g/L、约 0.6g/L、约 0.7g/L、约 0.8g/L、约 0.9g/L、约 1g/L、约 1.5g/L、约 2g/L、约 2.5g/L、约 3g/L、约 3.5g/L、约 4g/L、约 4.5g/L、约 5g/L、约 5.5g/L、约 6g/L、约 6.5g/L、约 7g/L、约 7.5g/L、约 8g/L、约 8.5g/L、约 9g/L、约 9.5g/L、约 10g/L、约 10.5g/L、约 11g/L、约 12g/L、约 0.2g/L 至约 0.5g/L、约 0.2g/L 至约 1g/L、约 0.2 至约 2g/L、约 0.3g/L 至约 0.6g/L、约 0.3g/L 至约 1g/L、约 0.3 至约 2g/L、约 0.4 至约 0.7g/L、约 0.4 至约 1g/L 约 0.4 至约 2g/L、约 0.4 至约 3g/L、约 0.5g/L 至约 1g/L、约 0.5g/L 至约 1g/L、约 0.5g/L 至约 2g/L、约 0.5g/L 至约 3g/L、约 0.5g/L 至约 4g/L、约 0.5g/L 至约 5g/L、约 0.5g/L 至约 6g/L、约 0.5g/L 至约 7 g/L、约 0.5g/L 至约 8g/L、约 0.5g/L 至约 9g/L、约 0.5g/L 至约 10g/L、约 0.5g/L 至约 11g/L、约 0.5g/L 至约 12g/L、约 1g/L 至约 2g/L、约 1g/L 至约 3g/L、约 1g/L 至约 4g/L、约 1g/L 至约 5g/L、约 1g/L 至约 6g/L、约 1g/L 至约 7g/L、约 1g/L 至约 8g/L、约 1g/L 至约 9g/L、约 1g/L 至约 10g/L、约 1g/L 至约 11g/L、约 1g/L 至约 12g/L、约 2g/L 至约 3g/L、约 2g/L 至约 4g/L、约 2g/L 至约 5g/L、约 2g/L 至约 6g/L、约 2g/L 至约 7g/L、约 2g/L 至约 8g/L、约 2g/L 至约 9g/L、约 2g/L 至约 10g/L、约 2g/L 至约 11g/L、约 2g/L 至约 12g/L、约 3g/L 至约 4g/L、约 3g/L 至约 5g/L、约 3g/L

L 至约 6g/L、约 3g/L 至约 7g/L、约 3g/L 至约 8g/L、约 3g/L 至约 9g/L、约 3g/L 至约 10g/L、约 3g/L 至约 11g/L、约 3g/L 至约 12g/L、约 4g/L 至约 5g/L、约 4g/L 至约 6g/L、约 4g/L 至约 7g/L、约 4g/L 至约 8g/L、约 4g/L 至约 9g/L、约 4g/L 至约 10g/L、约 4g/L 至约 11g/L、约 4g/L 至约 12g/L、约 5g/L 至约 6g/L、约 5g/L 至约 7g/L、约 5g/L 至约 8g/L、约 5g/L 至约 9g/L、约 5g/L 至约 10g/L、约 5g/L 至约 11g/L、约 5g/L 至约 12g/L、约 6g/L 至约 7g/L、约 6g/L 至约 8g/L、约 6g/L 至约 9g/L、约 6g/L 至约 10g/L、约 6g/L 至约 11g/L、约 6g/L 至约 12g/L、约 7g/L 至约 8g/L、约 7g/L 至约 9g/L、约 7g/L 至约 10g/L、约 7g/L 至约 11g/L、约 7g/L 至约 12g/L、约 8g/L 至约 9g/L、约 8g/L 至约 10g/L、约 8g/L 至约 11g/L、约 8g/L 至约 12g/L、约 9g/L 至约 10g/L、约 9g/L 至约 11g/L、约 9g/L 至约 12g/L、约 10g/L 至约 11g/L、约 10g/L 至约 12g/L、或约 11g/L 至约 12g/L。

[0212] 在实施方式中,所产生的重组毒素蛋白或亚基蛋白质的量为总细胞蛋白质的约 1% 至 75%。在某些实施方式中,所产生的毒素蛋白或亚基蛋白质的量为总细胞蛋白质的约 1%、约 2%、约 3%、约 4%、约 5%、约 10%、约 15%、约 20%、约 25%、约 30%、约 35%、约 40%、约 45%、约 50%、约 55%、约 60%、约 65%、约 70%、约 75%、约 1% 至约 5%、约 1% 至约 10%、约 1% 至约 20%、约 1% 至约 30%、约 1% 至约 40%、约 1% 至约 50%、约 1% 至约 60%、约 1% 至约 75%、约 2% 至约 5%、约 2% 至约 10%、约 2% 至约 20%、约 2% 至约 30%、约 2% 至约 40%、约 2% 至约 50%、约 2% 至约 60%、约 2% 至约 75%、约 3% 至约 5%、约 3% 至约 10%、约 3% 至约 20%、约 3% 至约 30%、约 3% 至约 40%、约 3% 至约 50%、约 3% 至约 60%、约 3% 至约 75%、约 4% 至约 10%、约 4% 至约 20%、约 4% 至约 30%、约 4% 至约 40%、约 4% 至约 50%、约 4% 至约 60%、约 4% 至约 75%、约 5% 至约 10%、约 5% 至约 20%、约 5% 至约 30%、约 5% 至约 40%、约 5% 至约 50%、约 5% 至约 60%、约 5% 至约 75%、约 10% 至约 20%、约 10% 至约 30%、约 10% 至约 40%、约 10% 至约 50%、约 10% 至约 60%、约 10% 至约 75%、约 20% 至约 30%、约 20% 至约 40%、约 20% 至约 50%、约 20% 至约 60%、约 20% 至约 75%、约 30% 至约 40%、约 30% 至约 50%、约 30% 至约 60%、约 30% 至约 75%、约 40% 至约 50%、约 40% 至约 60%、约 40% 至约 75%、约 50% 至约 60%、约 50% 至约 75%、约 60% 至约 75% 或约 70% 至约 75%。

[0213] 在某些实施方式中,多种蛋白质可产自相同宿主细胞。例如,在实施方式中,百日咳毒素的所有五个亚基由在单一培养物中生长的相同宿主细胞产生。在这样的实施方式中,所观测的浓度、% 总细胞蛋白质或活性是针对各单个毒素亚基或针对结合在一起的所有亚基的。也就是说,在实施方式中,本发明的方法用于获得约 1 克 / 升至约 12 克 / 升的百日咳毒素蛋白 S1、S2、S3、S4 或 S5 亚基的产率。在实施方式中,所产生的 S1、S2、S3、S4 或 S5 亚基蛋白质的量为总细胞蛋白质的 1% -75%。或者,本发明的方法用于获得约 1 克 / 升至约 12 克 / 升的 S1、S2、S3、S4 和 S5 亚基蛋白质的产率。在实施方式中,所产生的 S1、S2、S3、S4 或 S5 亚基蛋白质的量为总细胞蛋白质的 1% -75%。在某些实施方式中,所获得的各亚基的量,以克 / 升或 % 总细胞蛋白质表示,大致是相同的。

[0214] 虽然与质量相关,蛋白质的“溶解性”和“活性”一般通过不同的方式来测定。蛋白质的溶解性,特别是疏水蛋白,通常与蛋白质的折叠相关;不溶性表示疏水性氨基酸残基不适当地位于折叠蛋白的外侧。蛋白质的活性(它可以使用例如,下面所描述的方法进行评价)是适当蛋白质构象的另一个指标。本文所用“可溶性的、活性的或这两者”或者“可

溶性的和/或活性的”是指通过本领域技术人员已知的和本文中描述的方法确定为可溶的、活性的或同时为可溶的和活性的蛋白质。给定蛋白质的“活性”可以包括结合活性,例如,以与受体、特异性抗体或另一已知底物结合或通过酶活性(如果相关)表现的活性。活性水平可以例如,就绝对或相对意义来描述,如相比于标准或对照样品或作为基准使用的任何样品的活性时。

[0215] 用于评价毒素的活性分析在本领域中已知并且在文献中描述。活性分析包括免疫学或抗体结合分析,例如,Western 印迹分析和 ELISA,以及受体结合分析,例如,可通过白喉毒素受体 (proHB-EGF) 结合分析来评估 CRM197。在这些分析中有用的抗体可以商购。活性分析也包括酶活性分析。利用本领域已知的且在本文别处针对铜绿假单胞菌外毒素 A 描述的方法,可通过免疫学方法和也可通过 ADP 核糖基化活性来测定野生型 DT。

[0216] 例如,可以按照例如通过引用并入本文的美国专利号 6,140,082,“Expression of Gene Products from Genetically Manipulated Strains of *Bordetella*”所述来进行 CTB 的 Western 印迹分析。该专利描述了 CTB 在博德特菌中的表达。来自培养上清液的蛋白质通过 SDS-PAGE 解析,或者在解析前煮沸以使 CTB 五聚体转化成单体形式。将蛋白质转移到尼龙膜上,并用山羊抗类霍乱原 IgG 抗体(抗-CTB, List Biologicals #GAC-01C)进行探测。采用 dig 化学发光(Boehringer Mannheim),使用碱性磷酸酶偶联的驴抗山羊 IgG 进行检测。既包含 CTA 又包含 CTB 的霍乱毒素标准品(Sigma)用于比较。

[0217] PTX 的 Western 印迹分析可如本文实施例中所述,利用市售的抗体来进行。单克隆抗体可获自,例如,Abcam, Cambridge, MA。

[0218] 破伤风毒素 C 片段可如例如通过引用并入本文的美国专利号 5,443,966,“Expression of tetanus toxin fragment C”中所述,通过 Western 印迹分析或通过 ELISA 进行评价。抗体可从多种商业来源获得,例如,Abcam, Cambridge, MA。

[0219] 如本领域所述,TcdB 活性可通过 Western 印迹分析或其他检测分析进行评价。例如,可以利用本领域中描述的,例如在通过引用整体并入本文的美国专利号 7,226,597 中所描述的葡萄糖基水解酶/葡萄糖基化分析方法测定酶活性。具体地,葡萄糖基化反应可在含有 50mM n-2 羟乙基哌嗪-n'-2-乙烷磺酸、100mM KCl、1mM MnCl₂、1mM MgCl₂、100 μg/ml BSA、0.2mM GDP、40 μM [¹⁴C]UDP-葡萄糖(303Ci/mol; ICN Pharmaceuticals)、100 μM UDP-葡萄糖和 3pmol TcdB 或 10pmol 的各融合蛋白的反应混合物中进行。使该试验在 37℃ 下孵育过夜,和用 AG1-X2 阴离子交换树脂分离切割的葡萄糖,并且在液体闪烁计数器中计数。

[0220] 铜绿假单胞菌外毒素 A 的活性可用免疫学方法,例如,Western 印迹分析进行评估。由于 ETA 是一种 ADP-核糖基化毒素,例如,如通过引用并入本文的美国专利号 4,892,827 中所述,可测定其 ADP-核糖基化活性。具体地,富含延伸因子(EF-2)的兔网织红细胞制剂或小麦胚芽提取物用作 EF-2 的来源。试验(500 μl 总体积)包含约 10pmol EF-2、37pmol [¹⁴C]-NAD(0.06 μCi)、0.25-1.25 μg ETA 和缓冲液(40mM DTT、1mM EDTA 和 50mM Tris, pH 8.1)。将活性测定为 30 分钟内转移到 EF-2 上的 NAD 的 pmol 数。建立已知浓度的 PE 的标准曲线,并用来确定大肠杆菌提取物中 PE 的活性。在 37℃ 下孵育 30 分钟后,向各个测定混合物中加入 0.5ml 12% TCA。然后将测定混合物置于冰浴中 15 分钟,紧接着在 4℃ 下以 3,000x g 离心 10 分钟。用 1ml 的 6% TCA 洗涤沉淀物,并如上所述离心。然后在液体闪烁计数器中测定沉淀物的 ¹⁴C 放射性作为 ADP-核糖基化活性的指标。

[0221] 因此,活性的测量值可表示,例如,抗体或受体结合能力、底物 结合能力(关于柱材料)或酶活性。

[0222] 在实施方式中,活性表示为与分析的总量相比,提取物上清液中活性重组毒素蛋白的%。这是基于相对于在分析中使用的重组毒素蛋白的总量,确定为活性的重组毒素蛋白的量。在其它实施方式中,活性以与标准品(例如,天然蛋白质)相比该蛋白质的%活性水平表示。这是基于相对于标准样品中活性蛋白质的量,上清提取物样品中活性重组毒素蛋白的量(其中,来自各样品的相同量的蛋白质被用于分析中)。

[0223] 在实施方式中,约 40%至约 100%的毒素蛋白或亚基被确定为活性的。在实施方式中,约 40%、约 50%、约 60%、约 70%、约 80%、约 90%或约 100%的重组毒素蛋白或亚基确定为活性的。在实施方案中、约 40%至约 50%、约 50%至约 60%、约 60%至约 70%、约 70%至约 80%、约 80%至约 90%、约 90%至约 100%、约 50%至约 100%、约 60%至约 100%、约 70%至约 100%、约 80%至约 100%、约 40%至约 90%、约 40%至约 95%、约 50%至约 90%、约 50%至约 95%、约 50%至约 100%、约 60%至约 90%、约 60%至约 95%、约 60%至约 100%、约 70%的至约 90%、约 70%至约 95%、约 70%至约 100%或约 70%至约 100%的重组毒素蛋白或亚基被确定为活性的。

[0224] 在其它实施方式中,约 75%至约 100%的重组毒素蛋白或亚基确定为活性的。在实施方式中,约 75%至约 80%、约 75%至约 85%、约 75%至约 90%、约 75%至约 95%、约 80%至约 85%、约 80%至约 90%、约 80%至约 95%、约 80%至约 100%、约 85%至约 90%、约 85%至约 95%、约 85%至约 100%、约 90%至约 95%、约 90%至约 100%或约 95%至约 100%的重组毒素蛋白或亚基确定为活性的。

[0225] 确认诱导的蛋白质的身份的方法在本领域中也是已知的。例如,蛋白质可以通过使用 MALDI-TOF 质谱仪的肽质量指纹图谱、N-末端测序分析或肽谱法进行分析。

[0226] 虽然本发明的优选实施方式中已经显示和描述,本领域技术人员 将会很明白,这些实施方式仅作为实施例的方式提供。在不脱离本发明的情况下,本领域的技术人员将会想到多种变化、改变和替换。应当了解在本发明的实施过程中,对本文描述的本发明的实施方式可以进行各种改动。意图的是下面的权利要求限定本发明的范围,从而涵盖这些权利要求及其等同物范围内的方法和结构。

实施例

[0227] 实施例 1:重组 CRM197 蛋白的高通量表达

[0228] 构建 CRM197 表达菌株,并利用毛细管凝胶电泳(SDS-CGE)分析该菌株产生的可溶性 CRM197 蛋白的量。根据获得的数据,选择某些菌株用于大规模表达。

[0229] CRM197 表达菌株的构建和生长

[0230] 使用荧光假单胞菌优先的密码子构建 CRM197 编码序列以编码 CRM197 氨基酸序列。SEQ ID NO:1 显示了由表达的合成优化 CRM197 基因编码的氨基酸序列,SEQ ID NO:2 显示了所表达的合成优化 CRM197 基因的 DNA 序列。

[0231] 构建携带已优化的 CRM197 序列的质粒,该 CRM197 序列与表 8 中所示的 10 种荧光假单胞菌分泌前导序列融合。CRM197 编码序列与荧光假单胞菌分泌前导序列的编码序列同框地融合,以使蛋白质靶向至周质以便以正确折叠的和活性的形式回收。

[0232] 表 8. 用于 CRM197 表达筛选的分泌前导序列

[0233]

	分泌前导序列
1	DsbA
2	Azu
3	Ibp-S31A
4	Tpr
5	CupB2
6	CupA2
7	NikA
8	Pbp A20V
9	DsbC
10	TolB

[0234] 在荧光假单胞菌宿主中测试包含与重组 CRM197 编码序列融合的十种分泌前导序列的构建体。列于表 9 中的四种宿主用各表达质粒进行测试。用所示质粒对宿主细胞进行电穿孔,并使宿主细胞再悬浮于含有微量矿物质和 5%甘油的 HTP 生长培养基中,然后转移至含有 400 μ l M9 盐 1%葡萄糖培养基和微量元素的 96 孔深孔板中。96 孔板在 30℃下振荡孵育 48 小时。将各 10 微升的四十种种子培养物一式三份转移到 96 孔深孔板内,各孔含有 500 μ l 补充有微量元素和 5%甘油的 HTP 培养基,并如前所述孵育 24 小时。

[0235] 表 9. 用于 CRM197 表达筛选的宿主菌株

[0236]

宿主菌株	基因型	类型
1	lon, la, aprA	PD
2	<i>hs/UV prc1 degP1 degP2 aprA</i> 缺失; 过表达 DegP2 S219A	PD+FMO
3	dsbABCD	FMO
4	grpE, dnaKJ	FMO

[0237] PD = 蛋白酶缺失 (所列的蛋白酶缺失);FMO = 折叠调节因子过表达子 (所列的折叠调节因子是过表达的)。

[0238] 向各孔中加入异丙基- β -D-硫代吡喃半乳糖苷 (IPTG) 达到终浓度 0.3mM 以诱导靶蛋白的表达。向各孔中加入甘露醇 (Sigma,M1902) 达到终浓度 1%以在折叠调节因子过表达菌株中诱导折叠调节因子的表达,且温度下降至 25℃。诱导后 24 小时,使用 PBS 以 400 μ l 的体积将细胞标准化至 OD600 = 15。冷冻样品用于通过超声和离心进行后续处理,以生成可溶性和不溶性部分。

[0239] 样品制备和 SDS-CGE 分析

[0240] 通过超声处理 OD 标准化的培养物后进行离心来制备可溶性和不溶性细胞部分。解冻冷冻的标准化的培养液 (400 μ l),并超声处理 3.5 分钟。以 20,800 $\times$ g 离心裂解物 20 分钟 (4℃),并用手动或自动液体处理除去上清液 (可溶性部分)。冷冻沉淀 (不溶性部分),然后解冻以用 20,080 $\times$ g 在 4℃下再离心 20 分钟,从而除去残留的上清液。然后再将

沉淀重悬于 400 μ L 的 1X 磷酸盐缓冲盐水 (PBS), pH 值为 7.4 中。在 pH 值为 7.4 的 1X 磷酸盐缓冲盐水 (PBS) 中进一步稀释可溶性和不溶性样品用于 SDS-CGE 分析。在二硫苏糖醇 (DTT) 存在下制备可溶性和不溶性样品用于 SDS 毛细管凝胶电泳 (CGE) (Caliper Life Sciences, Protein Express LabChip Kit, Part 760301)。

[0241] 图 1 显示了表明来自各种菌株的可溶性部分的还原性 SDS-CGE 分析结果的代表性凝胶样图像。表 10 显示了构建的各 CRM197 表达菌株的 3 个重复的平均可溶性 CRM197 产率和标准差。也示出了宿主菌株和对各菌株进行筛选的分泌前导序列。

[0242] 分泌前导序列和宿主菌株都显示出对 CRM197 表达的显著影响。在 0.5mL 规模下的表达从没有可检测到的产率到超过 1.2g/L 不等,在宿主菌株 2 的背景下观察到最高表达水平。在 PS538-776 中观察到的产率为 1263mg/L,在 PS538-772 中为 1241mg/L,两者都大大超过平均产率 340mg/L。在相同宿主菌种中既观察到高产率又观察到低产率,这取决于所使用的前导序列,且在不同宿主菌株中使用相同前导序列既观察到高产率又观察到低产率。

[0243] 选择 PS538-772、PS538-773、PS538-776、PS538-778、PS538-782 用于在大规模发酵中评价。

[0244] 表 10. CRM197 表达菌株的平均 CRM197 产率

[0245]

菌株编号	在美国专利申请系列号 61/325,235 中对应的菌株编号	宿主菌株	前导序列	平均产率 (mg/L)	标准差 (3 个重复)
PS538-761	PS538-731	1	DsbA	205	162

[0246]

PS538-762	PS538-732	1	Azu	427	186
PS538-763	PS538-733	1	Ibp-S31A	361	183
PS538-764	PS538-734	1	Tpr	298	106
PS538-765	PS538-735	1	CupB2	105	109
PS538-766	PS538-736	1	CupA2	175	99
PS538-767	PS538-737	1	NikA	314	85
PS538-768	PS538-738	1	Pbp A20V	291	204
PS538-769	PS538-739	1	DsbC	148	91
PS538-770	PS538-740	1	TolB	213	36
PS538-771	PS538-741	2	DsbA	407	218
PS538-772	PS538-742	2	Azu	1241	372
PS538-773	PS538-743	2	Ibp-S31A	1107	219
PS538-774	PS538-744	2	Tpr	280	285
PS538-775	PS538-745	2	CupB2	192	219
PS538-776	PS538-746	2	CupA2	1263	474
PS538-777	PS538-747	2	NikA	699	259
PS538-778	PS538-748	2	Pbp A20V	914	416
PS538-779	PS538-749	2	DsbC	567	141
PS538-780	PS538-750	2	TolB	382	217
PS538-781	PS538-751	3	DsbA	591	230
PS538-782	PS538-752	3	Azu	1094	543
PS538-783	PS538-753	3	Ibp-S31A	323	143
PS538-784	PS538-754	3	Tpr	419	70
PS538-785	PS538-755	3	CupB2	75	74
PS538-786	PS538-756	3	CupA2	309	214
PS538-787	PS538-757	3	NikA	52	73
PS538-788	PS538-758	3	Pbp A20V	356	295
PS538-789	PS538-759	3	DsbC	319	117
PS538-790	PS538-760	3	TolB	69	88

[0247]

PS538-791	PS538-761	4	DsbA	270	106
PS538-792	PS538-762	4	Azu	0	14
PS538-793	PS538-763	4	Ibp-S31A	0	6
PS538-794	PS538-764	4	Tpr	0	0
PS538-795	PS538-765	4	CupB2	18	39
PS538-796	PS538-766	4	CupA2	118	134
PS538-797	PS538-767	4	NikA	0	9
PS538-798	PS538-768	4	Pbp A20V	0	0
PS538-799	PS538-769	4	DsbC	0	0
PS538-800	PS538-770	4	TolB	0	0

[0248] 实施例 2 :重组 CRM197 蛋白的大规模表达

[0249] 在 2 升发酵罐中由荧光假单胞菌菌株 PS538-772、PS538-776 和 PS538-782 生产重组 CRM197 蛋白。如本文所述以及如 Riesenber, D. 等人, 1991 所述, 培养物在含有矿物盐培养基的 2 升发酵罐中生长, 并通过加入氨维持在 32°C 和 pH 6.5。通过增强搅拌和喷射空气和氧到发酵罐内的流动使溶解氧维持过量。在整个发酵过程中, 将甘油递送到培养物中以维持过量水平。维持这些条件, 直到达到用于诱导的目标培养细胞密度 (575nm 下的光密度 (A_{575})), 此时加入 IPTG 以启动 CRM197 的产生。诱导时的细胞密度可以在 40-200 吸光度单位 (AU) 的 A_{575} 范围内变化。IPTG 浓度可在 0.02-0.4mM 的范围内变化。pH 为 6-7.5, 温度为 20-35°C。16-24 小时后, 通过离心收集各个生物反应器的培养物, 将细胞沉淀在 -80°C 下冷冻。通过 SDS-CGE 分析样品的产物形成。

[0250] 对多种发酵条件进行评价, 得出通过 SDS-CGE 确定的 1-2g/L 的最高 CRM197 表达 (见图 18 和图 19)。利用白喉毒素特异性抗体, 通过 Western 印迹分析来确认被诱导的蛋白质的身份 (图 20)。

[0251] 实施例 3 :重组白喉毒素 B 蛋白的高通量表达

[0252] 白喉毒素 B 表达菌株的构建和生长

[0253] 使用荧光假单胞菌优先的密码子构建白喉毒素 B 编码序列以编码白喉毒素 B 氨基酸序列。SEQ ID NO :22 显示了由表达的合成的白喉毒素 B 基因编码的氨基酸序列, 和 SEQ ID NO :23 显示了所表达的合成的优化白喉毒素 B 基因的 DNA 序列。

[0254] 构建了携带已优化的白喉毒素 B 序列的质粒, 该序列与用于 CRM197 的相同的 10 种荧光假单胞菌分泌前导序列编码序列 (在表 8 中示出) 融合。包含分泌前导序列, 以使蛋白质靶向至周质, 以便以正确折叠的和活性的形式回收。

[0255] 表达与重组白喉毒素 B 蛋白融合的十种分泌前导序列的构建体在荧光假单胞菌宿主中进行测试。用各表达质粒测试表 9 中列出的四种宿主。用所示质粒对宿主细胞进行电穿孔, 并如上文对于 CRM197 高通量表达所述以 96 孔格式生长和诱导。制备样品, 并如上文对于 CRM197 高通量表达样品所述通过 SDS-CGE 进行分析。

[0256] 图 2 显示了显示来自各菌株的可溶性部分的还原性 SDS-CGE 分析结果的代表性凝胶样图像。表 11 显示了构建的各霍乱毒素 B 表达菌株的 3 个重复的平均可溶性霍乱毒素 B 产率和标准差。

[0257] 分泌前导序列和宿主菌株都显示出对霍乱毒素 B 表达的显著影响。在 0.5mL 规模下的表达从没有可检测到的产率到超过 0.2g/L 不等,在 hslUV prc1 degP1 degP2 aprA 缺失 /DegP2 S219A 过表达宿主背景下观察到最高表达水平。与前导序列 6 (CupA2) 和 8 (PbpA20V) 融合的霍乱毒素 B 的表达在全部四种菌株中似乎一致地为高水平。

[0258] 表 11. 霍乱毒素 B 表达总结

[0259]

菌株编号	宿主菌株	质粒	前导序列	平均产率 (mg/L)	标准差 (3个重复)
PS538-081	1	p538-021	DsbA	25	8
PS538-082	1	p538-022	Azu	1	8
PS538-083	1	p538-023	Ibp-S31A	0	5
PS538-084	1	p538-024	Tpr	35	14

[0260]

PS538-085	1	p538-025	CupB2	10	9
PS538-086	1	p538-026	CupA2	138	18
PS538-087	1	p538-027	NikA	0	5
PS538-088	1	p538-028	Pbp A20V	213	23
PS538-089	1	p538-029	DsbC	0	6
PS538-090	1	p538-030	TolB	0	4
PS538-091	2	p538-021	DsbA	133	62
PS538-092	2	p538-022	Azu	83	56
PS538-093	2	p538-023	Ibp-S31A	50	44
PS538-094	2	p538-024	Tpr	61	55
PS538-095	2	p538-025	CupB2	62	19
PS538-096	2	p538-026	CupA2	147	57
PS538-097	2	p538-027	NikA	31	28
PS538-098	2	p538-028	Pbp A20V	223	78
PS538-099	2	p538-029	DsbC	41	24
PS538-100	2	p538-030	TolB	6	5
PS538-101	3	p538-021	DsbA	1	7
PS538-102	3	p538-022	Azu	1	2
PS538-103	3	p538-023	Ibp-S31A	19	17
PS538-104	3	p538-024	Tpr	28	36
PS538-105	3	p538-025	CupB2	5	9
PS538-106	3	p538-026	CupA2	40	12
PS538-107	3	p538-027	NikA	5	10
PS538-108	3	p538-028	Pbp A20V	45	19
PS538-109	3	p538-029	DsbC	0	6
PS538-110	3	p538-030	TolB	0	3
PS538-111	4	p538-021	DsbA	0	5
PS538-112	4	p538-022	Azu	0	3

[0261]

PS538-113	4	p538-023	Ibp-S31A	0	2
PS538-114	4	p538-024	Tpr	13	3
PS538-115	4	p538-025	CupB2	2	4
PS538-116	4	p538-026	CupA2	15	16
PS538-117	4	p538-027	NikA	0	2
PS538-118	4	p538-028	Pbp A20V	35	15
PS538-119	4	p538-029	DsbC	0	2
PS538-120	4	p538-030	TolB	0	2

[0262] 实施例 4：重组霍乱毒素 B 蛋白的大规模表达

[0263] 在荧光假单胞菌 Pfēnex Expression Technology™菌株 PS538-088 和 PS538-091 中生产重组霍乱毒素 B 蛋白。如本文所述,以及如 Riesenber, D. 等人, 1991 所述, 选择的菌株在包含矿物盐培养基的 2 升发酵罐中生长, 并通过加入氨维持在 32°C 和 pH 6.5。通过增加搅拌和喷射空气和氧到发酵罐内的流动使溶解氧维持在过量水平。在整个发酵过程中, 将甘油递送到培养物中以维持过量水平。维持这些条件, 直到达到用于诱导的目标培养细胞密度 (575nm 下的光密度 (A_{575})), 此时加入 IPTG 以启动靶蛋白的生产。加入 IPTG 以启动 CTB 的生产。16-24 小时后, 通过离心收集各生物反应器的培养物, 并将细胞沉淀在 -80°C 下冷冻。

[0264] 对多种发酵条件进行评价, 得出通过 SDS-CGE 确定的 0.6-1.0g/L 的最高 CTB 表达。表现最佳的发酵培养物在约 80-1600D 时在 pH6.5-7.2 和 32°C 的条件下用 0.2mM IPTG 进行诱导。通过 SDS-CGE 确定可溶性 CTB 浓度 (见图 14 和表 12)。利用 MALDI-TOF 质谱分析通过肽质量指纹证实被诱导的蛋白质的身份。

[0265] 表 12. 可溶性霍乱毒素 B 滴度

[0266]

菌株	发酵物	产物	产物浓度 (g/L)
PS538-088	U5	CTB	0.94±0.03
PS538-088	U6	CTB	0.59±0.01
PS538-091	U3	CTB	0.81±0.09

[0267] 实施例 5：重组百日咳毒素蛋白的高通量表达

[0268] 百日咳类毒素 S1 E129A R9K 表达菌株的构建和生长

[0269] 具有 S1 突变 E129A 和 R9K 的、编码亚基 S1、S2、S3、S4 和 S5 的百日咳类毒素操纵子的序列用于重组百日咳毒素的表达。图 3 显示了该操纵子的图谱。图 4 显示了带有翻译结果的该操纵子的 DNA 序列 (SEQ ID NO:24)。图 5 显示了 S1、S2、S3、S4 和 S5 的单个氨基酸序列。

[0270] 构建体在表 13 中所示的 8 种荧光假单胞菌宿主中表达。用 p538-081 对宿主细胞进行电穿孔, 并如上文对于 CRM197 高通量表达所述以 96 孔格式生长和诱导。制备样品, 并如上文对于 CRM197 高通量表达样品所述通过 SDS-CGE 进行分析。

[0271] 表 13. 百日咳毒素 S1 E129A R9K 表达菌株

[0272]

菌株编号	宿主	基因型	质粒	类型
PS538-321	1	lon, la, aprA	p538-081	PD
PS538-322	2	hslUV, prc1, degP1, degP2 和 aprA	p538-081	PD
PS538-323	3	dsbABCD	p538-081	FMO
PS538-324	4	grpE, dnaKJ	p538-081	FMO
PS538-325	5	htpX	p538-081	PD
PS538-326	6	RXF01590	p538-081	PD
PS538-327	7	lon, la, aprA 缺失; 过表达 grpE 和 dnaKJ	p538-081	PD +FMO
PS538-328	8	ppiB (RXF05345)	p538-081	FMO

[0273] PD = 蛋白酶缺失 (所列出的蛋白酶缺失); FMO = 折叠调节因子过表达子 (所列出的折叠调节因子过表达)

[0274] 表达的百日咳毒素的 Western 印迹分析

[0275] 通过 Western 印迹法分析上述 8 种培养物的可溶性部分以评价百日咳类毒素的表达。20 微升的可溶性部分 (2X 稀释, 还原及非还原的) 在 Bio-Rad 12% Bis-Tris 凝胶上在 1X Bio Rad MES 电泳缓冲液中电泳。对于还原性 Western 分析, 加入 1X XT 还原剂。使用含有 20% 甲醇的 1X NuPAGE 转移缓冲液 (Invitrogen, NP0006-1), 在 100V 下经 60 分钟将蛋白质从 SDS-PAGE 转移到 0.2 μ m 硝酸纤维素膜上 (Bio Rad, 162 0232)。在室温下在 PBS (Pierce, 37528) 中的 Blocker™ 1% 酪蛋白中封闭膜 1 小时。为了检测, 倒出稀释液, 然后加入包含各 1 : 1000 稀释的抗百日咳博德特菌毒素 S4 和 S1 单克隆抗体 (Abcam, cat# ab37686 和 #37547) 的组合的更多稀释液。将印迹在 4°C 振荡下孵育过夜。用 PBS-Tween 洗涤印迹 3 次, 每次 5 分钟, 然后在室温下在包含 1 : 5,000 稀释的源自山羊的抗小鼠 IgG- 过氧化酶 (Sigma, Cat#A4416) 的更多稀释液中孵育 1 小时。用 PBS- 吐温 (Sigma, P3563) 洗涤印迹 3 次, 每次 5 分钟, 然后用免疫纯的金属强化的 DAB 底物 (Pierce, 34065) 显色。在还原性和非还原性条件下, 用抗 S1 和抗 S4 抗体检测多种亚基 (图 6)。所观察的表达类毒素的还原和非还原样品的带型与 Sekura 等人 (J. Biological Chemistry 258 :14647, 1983) 报道的对于从菌株 165 纯化的百日咳毒素所观察的带型一致。

[0276] 实施例 6 : 重组百日咳毒素蛋白的大规模表达

[0277] 重组百日咳毒素蛋白在荧光假单胞菌 Pf_{enex} Expression Technology™ 菌株 PS538-321、PS538-324、PS538-325、PS538-326 和 PS538-328 中生产。如以上对于 CTB 大规模表达所述, 选择的菌株在 2 升发酵罐中生长, 用 IPTG 诱导, 并制备样品用于分析。通过 SDS-CGE 分析样品的产物形成, 并通过 Western 印迹分析其活性。

[0278] 实施例 7 :重组野生型百日咳类毒素的高通量表达

[0279] 百日咳类毒素表达菌株的构建和生长

[0280] 编码亚基 S1、S2、S3、S4 和 S5 的野生型百日咳毒素操纵子的序列与 S1 一起用于重组百日咳类毒素的表达。图 13 显示了带有翻译结果的野生型操纵的 DNA 序列 (SEQ ID NO :35)。

[0281] 构建体在表 14 中所示的荧光假单胞菌宿主中表达。检测所列出的不具有过表达质粒的各种菌株 :a) 如前所述 (不具有过表达质粒) ;b) 包含 GrpE DnaKJ 过表达质粒 ;以及 c) 包含 DsbABCD 过表达质粒。用 PTX WT 表达质粒对宿主细胞进行电穿孔,并如上文对于 PTXS1 R9K E129A 高通量表达所述以 96 孔格式生长和诱导。同样如上所述制备样品并通过 SDS-CGE 对其进行分析。

[0282] 表 14. 百日咳类毒素野生型表达菌株

[0283]

宿主*	基因型	类型
9	hslUV prc2	PD
10	hslUV degP1	PD
11	la	PD
12	hslUV prc1 prc2	PD
13	lon la prc1 prc2	PD
14	RXF01590	PD
1	lon la aprA	PD
7	lon la prc1 degP2 aprA; 过表达 GrpE DnaKJ	PD + FMO
15	RXF02151 RXF00428	PD
16	lon la degP2	PD
17	过表达 DsbAB	FMO
18	过表达 DsbCD	FMO
19	prc1 degP2; 过表达 degP2 S219A	PD + FMO
20	prc1 degP2 clp1 aprA; 过表达 degP2 S219A	PD + FMO
21	prc1 degP2 lon aprA; 过表达 degP2 S219A	PD + FMO
22	prc1 degP2 degP1 aprA; 过表达 degP2 S219A	PD + FMO
23	lon prc1 degP2 degP1 aprA; 过表达 degP2 S219A	PD + FMO
2	hslUV prc1 degP2 degP1 aprA; 过表达 degP2 S219A	PD + FMO

[0284]

25	lon la degP2 prc1 aprA; 过表达 degP2 S219A	PD + FMO
26	degP2; 过表达 SecB	PD + FMO
27	degP2; 过表达 FkbP	PD + FMO
28	degP2; 过表达 GroELES	PD + FMO
29	lon la aprA; 过表达 SecB	PD + FMO
30	lon la aprA; 过表达 FkbP	PD + FMO
31	lon la aprA; 过表达 GroELES	PD + FMO
32	dsbC	PD
33	dsbC; 过表达 DsbAB	PD + FMO
3	过表达 DsbABCD	FMO
35	lexA aprA	PD
36	过表达 SlyD	FMO
37	lon hslUV	PD
38	Wt	---
39	aprA	PD
4	过表达 GrpE DnaKJ	FMO
5	htpX	PD
40	lon	PD
41	prc1	PD
42	hslUV	PD
43	degP2	PD
44	degP1	PD
45	prc2	PD
46	RXF6451	PD
6	RXF1590	PD
48	RXF4692	PD
49	hslUV mic	PD
50	RXF2161	PD
51	RXF00133	PD

[0285]

52	RXF2796	PD
53	RXF4968	PD
54	过表达 DsbC	FMO
55	过表达 DsbAC	FMO
56	过表达 LepB	FMO
57	过表达 SecB	FMO
58	过表达 ClpA	FMO
59	过表达 FklB2	FMO
60	过表达 DnaK-样	FMO
61	过表达 FkbP	FMO
62	过表达 PpiA	FMO
8	过表达 PpiB	FMO
63	过表达 HscA	FMO
64	过表达 GshA	FMO
65	过表达 Gor	FMO
66	过表达 TrxC	FMO
67	过表达 DsbG	FMO
68	过表达 Ppi	FMO
69	过表达 GroELES	FMO
70	prc1 aprA; 过表达 GrpE DnaKJ	PD + FMO
71	分泌过多	
72	过表达 DsbD	FMO
73	分泌过多	
74	分泌过多	
75	prc1 prc2	PD
76	hslUV clpA	PD

[0286] * 检测所列出的不具有过表达质粒的各菌株 :a) 如前所述 (不具有过表达质粒); b) 包含 GrpE DnaKJ 过表达质粒; 以及 c) 包含 DsbABCD 过表达质粒。PD = 蛋白酶缺失 (所列出的蛋白缺失); FMO = 折叠调节因子过表达子 (所列出的折叠调节因子过表达)。

[0287] 分泌过多的菌株, 又称为过囊泡化菌株, 例如在 W02010/008764, “Pseudomonas Fluorescens Strains for Production of Extracellular Recombinant Protein” 中描

述,其通过引用整体并入本文。

[0288] 实施例 8 :重组破伤风毒素 C 片段蛋白的高通量表达

[0289] 破伤风毒素 C 表达菌株的构建和生长

[0290] 用荧光假单胞菌优先的密码子构建破伤风毒素 C 编码序列以编码破伤风毒素 C 氨基酸序列。SEQ ID NO :30 显示了由表达的合成破伤风毒素 C 基因编码的氨基酸序列,和 SEQ ID NO :31 显示了表达的合成优化破伤风毒素 C 基因的 DNA 序列。

[0291] 构建携带已优化的破伤风毒素 C 序列的质粒,该序列与用于 CRM197 相同的 10 种荧光假单胞菌分泌前导序列编码序列(表 8 所示)融合。包含分泌前导序列,以将蛋白质靶向至周质,以便以正确折叠的和活性的形式回收。

[0292] 在荧光假单胞菌宿主中测试表达与重组破伤风毒素 C 蛋白融合的这 10 种分泌前导序列的构建体。用各前导区测试表 9 中列出的四种宿主。用所示质粒对宿主细胞进行电穿孔,并如上文对于 CRM197 高通量表达所述以 96 孔格式生长和诱导。如上文对于 CRM197 高通量表达样品所述制备样品并通过 SDS-CGE 对其进行分析。

[0293] 图 7 显示了表明来自各菌株的可溶性部分的还原性 SDS-CGE 分析结果的代表性凝胶样图像。表 15 显示了构建的各种破伤风毒素 C 表达菌株的 3 个重复的平均可溶性破伤风毒素 C 产率和标准差。破伤风毒素 C 片段似乎在所测试的大多数菌株中都很好地表达,在 hslUV prc1 degP1 degP2 aprA 缺失 /DegP2 S219A 过表达宿主中有高达 600mg/L 的最高产率。选择菌株 PS538-529、PS538-538、PS538-544、PS538-546、PS538-547、PS538-548、PS538-558、PS538-565 和 PS538-568 用于进一步评价。

[0294] 表 15. 破伤风毒素 C 表达总结

[0295]

菌株编号	宿主	质粒	前导区	平均产率 (mg/L)	标准差 (3个重复)
PS538-529	1	p538-132	DsbA	261	75
PS538-530	1	p538-133	Azu	200	82
PS538-531	1	p538-134	Ibp-S31A	165	64
PS538-532	1	p538-135	Tpr	207	107
PS538-533	1	p538-136	CupB2	205	128
PS538-534	1	p538-137	CupA2	200	117
PS538-535	1	p538-138	NikA	174	96
PS538-536	1	p538-139	Pbp A20V	311	156
PS538-537	1	p538-140	DsbC	188	97
PS538-538	1	p538-141	TolB	129	63
PS538-539	2	p538-132	DsbA	486	89
PS538-540	2	p538-133	Azu	495	93
PS538-541	2	p538-134	Ibp-S31A	568	68
PS538-542	2	p538-135	Tpr	589	364
PS538-543	2	p538-136	CupB2	534	318
PS538-544	2	p538-137	CupA2	504	134
PS538-545	2	p538-138	NikA	444	145
PS538-546	2	p538-139	Pbp A20V	637	280
PS538-547	2	p538-140	DsbC	574	68
PS538-548	2	p538-141	TolB	438	61
PS538-549	3	p538-132	DsbA	174	37
PS538-550	3	p538-133	Azu	180	58
PS538-551	3	p538-134	Ibp-S31A	88	58
PS538-552	3	p538-135	Tpr	247	134
PS538-553	3	p538-136	CupB2	199	39

[0296]

PS538-554	3	p538-137	CupA2	165	69
PS538-555	3	p538-138	NikA	97	90
PS538-556	3	p538-139	Pbp A20V	297	112
PS538-557	3	p538-140	DsbC	151	52
PS538-558	3	p538-141	TolB	35	13
PS538-559	4	p538-132	DsbA	39	39
PS538-560	4	p538-133	Azu	40	43
PS538-561	4	p538-134	Ibp-S31A	36	40
PS538-562	4	p538-135	Tpr	35	39
PS538-563	4	p538-136	CupB2	54	26
PS538-564	4	p538-137	CupA2	42	36
PS538-565	4	p538-138	NikA	44	37
PS538-566	4	p538-139	Pbp A20V	37	40
PS538-567	4	p538-140	DsbC	39	43
PS538-568	4	p538-141	TolB	45	38

[0297] 实施例 9 : 重组破伤风毒素 C 片段蛋白的大规模表达

[0298] 重组破伤风毒素 C 蛋白在荧光假单胞菌 Pf ϵ nex Expression Technology™ 菌株 PS538-529、PS538-538、PS538-544、PS538-546、PS538-547、PS538-548、PS538-558、PS538-565 和 PS538-568 中生产。如以上对于 CRM197 所述,选择的菌株在包含矿物盐培养基的 2 升发酵罐中生长。

[0299] 对多种发酵条件进行评价,获得通过 SDS-CGE 确定的来自菌株 PS538-529、PS538-546 和 PS538-547 的 6-10g/L 的最高可溶性 TTC 表达(见图 11A 和表 16)。表现最佳的发酵培养物在约 1600D 时在 pH 7.2 和 32°C 下用 0.2mM IPTG 诱导。利用 MALDI-TOF 质谱法和 Western 印迹分析,通过肽质量指纹来确认被诱导的蛋白质的身份。质谱和 Western 印迹分析表明,在这些表达条件下,PS538-529、PS538-546 和 PS538-547 的分泌前导序列(分别为 DsbA、Pbp A20V 和 DsbC)并非由 100% 的表达的蛋白质加工。然而,TolB 前导序列被确定为由分泌的蛋白质精确切割(数据未显示)。利用以上概述的条件,以 2L 发酵规模筛选 TolB-TTC 表达菌株 PS538-538、PS538-548、PS538-558 和 PS538-568,以鉴定能够以分泌前导序列的高保真切割和低水平降解产生 TTC 的菌株。通过 Western 印迹分析观察到菌株 PS538-538、PS538-548 和 PS538-558 产生相似质量和产率的材料(图 11B)。

[0300] 表 16. 可溶性破伤风毒素 C(TTC) 滴度

[0301]

菌株	发酵物	产物	产物浓度 (g/L)
PS538-529	U1- 图 11A	TTC	5.7 $\pm$ 1.3

PS538-546	U7- 图 11A	TTC	9.5 ± 1.1
PS538-547	U5- 图 11A	TTC	6.2 ± 1.9
PS538-538	U1- 图 11B	TTC	2.5 ± 0.09
PS538-538	U2- 图 11B	TTC	1.8 ± 0.2
PS538-548	U3- 图 11B	TTC	5.3 ± 0.6
PS538-548	U4- 图 11B	TTC	4.5 ± 0.2
PS538-558	U5- 图 11B	TTC	1.1 ± 0.8
PS538-558	U6- 图 11B	TTC	1.9 ± 0.1
PS538-568	U7- 图 11B	TTC	0.2 ± 0.01
PS538-568	U8- 图 11B	TTC	0.2 ± 0.01

[0302] 实施例 10 :重组艰难梭菌 B 蛋白的高通量表达

[0303] TcdB 表达菌株的构建和生长

[0304] 用荧光假单胞菌优先的密码子构建 TcdB 编码序列以编码 TcdB 氨基酸序列。SEQ ID NO :32 显示了由表达的合成 TcdB 基因编码的氨基酸序列,和 SEQ ID NO :33 显示了表达的合成优化 TcdB 基因的 DNA 序列。

[0305] 携带已优化的 TcdB 序列的质粒在具有表 17 中所列出的基因型的荧光假单胞菌宿主中进行测试。用细胞质表达质粒 p538-211 对宿主 细胞进行电穿孔,并如上文对于 CRM197 高通量表达所述以 96 孔格式生长和诱导。如以上对于 CRM197 高通量表达样品所述制备样品并通过 SDS-CGE 对其进行分析。

[0306] 表 17. TcdB 宿主菌株

[0307]

宿主菌株	基因型	类型
37	hslUV lon	PD
38		WT
4	dnaKJ grpE	FMO
5	htpX	PD
40	lon	PD
41	prc1	PD
42	hslUV	PD
43	degP2	PD
44	degP1	PD
45	prc2	PD
47	RXF01590	PD
49	hslUV mic	PD
53	RXF04968	PD
55	dsbAC	FMO
61	fkbp	FMO
66	trxC	FMO
72	dsbD	FMO
76	hslUV clpA	PD
12	hslUV prc1 prc2	PD
1	lon la aprA	PD
16	lon la degP2	PD
2	hslUV prc1 degP1 degP2 aprA 缺失; 过表达 degP2 S219A	PD + FMO
3	dsbABCD	FMO

[0308]

21	lon prc1 degP2 aprA 缺失, degP2 S219A 过表达	PD + FMO
----	--	----------

[0309] 图 8 显示了显示来自 24 种测试的菌株中每一种的可溶性部分的还原性 SDS-CGE 分析结果的代表性凝胶样图像。表 18 显示了构建的各 TcdB 表达菌株的 3 个重复的平均

可溶性 TcdB 产率和标准差。选择菌种 PS538-654、PS538-659、PS538-669、PS538-671 和 PS538-674 用于进一步评价。

[0310] 表 18. TcdB 表达总结

[0311]

菌株编号	宿主菌株	质粒	平均产率 (mg/L)	标准差 (3个重复)
PS538-651	37	p538-211	103	7
PS538-652	38	p538-211	55	4
PS538-653	4	p538-211	57	1
PS538-654	5	p538-211	166	13
PS538-655	40	p538-211	88	3
PS538-656	41	p538-211	68	5
PS538-657	42	p538-211	90	14
PS538-658	43	p538-211	68	2
PS538-659	44	p538-211	109	8
PS538-660	45	p538-211	78	4
PS538-661	6	p538-211	98	15
PS538-662	49	p538-211	106	10
PS538-663	53	p538-211	91	6
PS538-664	55	p538-211	45	4
PS538-665	61	p538-211	63	6
PS538-666	66	p538-211	56	8
PS538-667	72	p538-211	70	8
PS538-668	76	p538-211	80	6

[0312]

PS538-669	12	p538-211	117	39
PS538-670	1	p538-211	108	18
PS538-671	16	p538-211	247	65
PS538-672	2	p538-211	32	6
PS538-673	3	p538-211	52	2
PS538-674	21	p538-211	145	12

[0313] 实施例 11 : 重组艰难梭菌毒素 B 蛋白的大规模表达

[0314] 重组艰难梭菌毒素 B 蛋白在荧光假单胞菌 Pf_{enex} Expression Technology™ 菌株 PS538-654、PS538-659、PS538-669、PS538-671 和 PS538-674 中生产。如以上对于 CTB 大规模表达所述, 选择的菌株在 2 升发酵罐中生长, 用 IPTG 进行诱导, 并制备样品以供分析。

[0315] 对多种发酵条件进行评价, 得到通过 SDS-CGE 确定的接近 2g/L 的最高艰难梭菌 B 毒素表达。表现最佳的发酵培养物在约 1600D 时在 pH 6.5 和 32°C 下用 0.08mM IPTG 进行诱导。通过 SDS-CGE 确定可溶性艰难梭菌 B 毒素的浓度 (见图 12 和表 19)。通过 Western 印迹法确认诱导的蛋白的身份。

[0316] 表 19. 可溶性艰难梭菌 B 毒素 (TcdB) 滴度

[0317]

菌株	发酵物	产物	产物浓度 (g/L)
PS538-671	U5	TcdB	1.6±0.4
PS538-671	U6	TcdB	2.1±0.2
PS538-674	U7	TcdB	1.8±0.2

[0318] 实施例 12 : 重组外毒素 A 蛋白的高通量表达

[0319] 铜绿假单胞菌外毒素 A 表达菌株的构建和生长

[0320] 用铜绿假单胞菌优先的密码子构建铜绿假单胞菌外毒素 A 突变 rEPA 编码序列以编码 rEPA 氨基酸序列。图 13 显示了表达的合成 rEPA 基因的氨基酸序列和 DNA 序列。

[0321] 构建编码如图 13 所示的携带任一缺失突变 rEPA 的优化序列的质粒, 该序列与用于 CRM197 的相同的 10 种荧光假单胞菌分泌前导序列编码序列 (表 8 所示) 融合。包含分泌前导序列编码序列, 以将蛋白质靶向至周质, 以便以正确折叠的和活性的形式回收。

[0322] 在表 20 中所列出的 8 种荧光假单胞菌宿主中测试表达与 rEPA 蛋白融合的 10 种分泌前导序列的构建体。用所示质粒对宿主细胞进行电穿孔, 并如上文对于 CRM197 高通量表达所述以 96 孔格式生长和诱导。如上文对于 CRM197 高通量表达样品所述制备样品并通过 SDS-CGE 对其进行分析。最高产率范围为 4.7-6.7g/L 可溶性 rEPA。

[0323] 表 20. 外毒素 A 宿主菌株

[0324]

宿主菌株	基因型	类型
5	htpX	PD
6	沙雷氏菌溶素	PD
7	Lon La1 aprA ; grpE dnaKJ 过表达	PD/FMO
8	ppiB	PD
1	Lon-, La-, aprA-	PD
2	hslUV-, prc1-,degP1-,degP2-,aprA- degP2 S219A 过表达	PD +FMO
3	dsbABCD 过表达	FMO
4	grpE, dnaKJ 过表达	FMO

[0325] 用所示质粒对宿主细胞进行电穿孔,并如上文对于 CRM197 高通量表达所述以 96 孔格式生长和诱导。如上文对于 CRM197 高通量表达样品所述制备样品并通过 SDS-CGE 对其进行分析。最高产率范围为 1.6-2.2g/L 可溶性外毒素 A 蛋白。表 21 显示了为进一步测试而选择的各种表达菌株的可溶性 rEPA 产率。

[0326] 表 21. rEPA HTP 表达总结

菌株编号	宿主	质粒	分泌前导序列	容积产率 (g/L)
PS538-1670	3	p538-250	DsbC	6.7
PS538-1663	3	p538-243	Ibp-s31a	5.7
PS538-1633	1	p538-243	Ibp-s31a	5.7
PS538-1640	1	p538-249	Pbp-A20V	4.7
PS538-1662	3	p538-242	Azu	4.2
PS538-1632	1	p538-242	Azu	3.2
PS538-1671	4	p538-241	DsbA	2.9
PS538-1665	3	p538-245	Tpr	2.7
PS538-1667	3	p538-247	CupA2	2.6
PS538-1674	4	p538-244	ToIB	2.3
PS538-1672	4	p538-242	Azu	2.2
PS538-1676	4	p538-246	CupB2	2.2
PS538-1677	4	p538-247	CupA2	2.1
PS538-1635	1	p538-245	Tpr	2.0
PS538-1675	4	p538-245	Tpr	2.0
PS538-1673	4	p538-243	Ibp-s31a	2.0
PS538-1680	4	p538-250	DsbC	1.9
PS538-1679	4	p538-249	Pbp-A20V	1.7
PS538-1669	3	p538-249	Pbp-A20V	1.6
PS538-1678	4	p538-248	NikA	1.5
PS538-1652	2	p538-242	Azu	1.5
PS538-1653	2	p538-243	Ibp-s31a	1.4
PS538-1660	2	p538-250	DsbC	1.4
PS538-1637	1	p538-247	CupA2	1.3
PS538-1666	3	p538-246	CupB2	1.1
PS538-1636	1	p538-246	CupB2	1.0
PS538-1634	1	p538-244	ToIB	1.0
PS538-1627	8	p538-247	CupA2	0.8
PS538-1631	1	p538-241	DsbA	0.8
PS538-1622	8	p538-242	Azu	0.8

[0327]

PS538-1661	3		p538-241	DsbA	0.7
PS538-1603	5		p538-243	Ibp-s31a	0.6
PS538-1630	8		p538-250	DsbC	0.6
PS538-1602	5		p538-242	Azu	0.6
PS538-1605	5		p538-245	Tpr	0.6
PS538-1623	8		p538-243	Ibp-s31a	0.6
PS538-1664	3		p538-244	TolB	0.5
PS538-1668	3		p538-248	NikA	0.5
PS538-1610	5		p538-250	DsbC	0.5
PS538-1606	5		p538-246	CupB2	0.4
PS538-1659	2		p538-249	Pbp-A20V	0.4
PS538-1607	5		p538-247	CupA2	0.4
PS538-1626	8		p538-246	CupB2	0.4
PS538-1625	8		p538-245	Tpr	0.4
PS538-1638	1		p538-248	NikA	0.3
PS538-1609	5		p538-249	Pbp-A20V	0.3
PS538-1604	5		p538-244	TolB	0.3
PS538-1629	8		p538-249	Pbp-A20V	0.3
PS538-1657	2		p538-247	CupA2	0.2
PS538-1651	2		p538-241	DsbA	0.2
PS538-1601	5		p538-241	DsbA	0.2
PS538-1624	8		p538-244	TolB	0.2
PS538-1621	5		p538-241	DsbA	0.2
PS538-1608	5		p538-248	NikA	0.2
PS538-1654	2		p538-244	TolB	0.2
PS538-1628	8		p538-248	NikA	0.1
PS538-1658	2		p538-248	NikA	0.1
PS538-1655	2		p538-245	Tpr	0.1

[0328]

PS538-1641	7		p538-241	DsbA	0.1
PS538-1611	6		p538-241	DsbA	NQ
PS538-1612	6		p538-242	Azu	NQ
PS538-1613	6		p538-243	Ibp-s31a	NQ
PS538-1614	6		p538-244	To1B	NQ
PS538-1615	6		p538-245	Tpr	NQ
PS538-1616	6		p538-246	CupB2	NQ
PS538-1617	6		p538-247	CupA2	NQ
PS538-1618	6		p538-248	NikA	NQ
PS538-1619	6		p538-249	Pbp-A20V	NQ
PS538-1620	6		p538-250	DsbC	NQ
PS538-1642	7		p538-242	Azu	NQ
PS538-1643	7		p538-243	Ibp-s31a	NQ
PS538-1644	7		p538-244	To1B	NQ
PS538-1645	7		p538-245	Tpr	NQ
PS538-1646	7		p538-246	CupB2	NQ
PS538-1647	7		p538-247	CupA2	NQ
PS538-1648	7		p538-248	NikA	NQ
PS538-1649	7		p538-249	Pbp-A20V	NQ
PS538-1650	7		p538-250	DsbC	NQ
PS538-1656	2		p538-246	CupB2	NQ

[0329] NQ = 不可量化

[0330] 实施例 13 : 重组铜绿假单胞菌外毒素 A 蛋白的大规模表达

[0331] 重组铜绿假单胞菌外毒素 A 蛋白 (rEPA) 在 2 升发酵罐中在荧光假单胞菌菌株 PS538-1633、PS538-1640 和 PS538-1670 生产。如本文所述以及在例如 Riesenber, D. 等人, 1991 中所述, 培养物在包含矿物盐培养基的 2 升发酵罐中生长, 并通过加入氨使其维持在 32°C 和 pH 6.5。通过增加搅拌和喷射空气和氧向发酵罐内的流动使溶解氧维持在过量水平。在整个发酵过程中, 将甘油递送到培养物中以维持过量水平。维持这些条件, 直到达到用于诱导的目标培养细胞密度 (575nm 下的光密度 (A_{575})), 此时加入 IPTG 以启动 rEPA 的生产。诱导时的细胞密度可从 40 到 200 吸光度单位 (AU) 的 A_{575} 下变化。IPTG 浓度可在 0.02-0.4mM 的范围内变化。pH 为 6-7.5, 和温度为 20-35°C。16-24 小时后, 通过离心收集各生物反应器的培养物, 并将细胞沉淀物在 -80°C 下冷冻。通过 SDS-CGE 分析样品的产物形成。

[0332] 对多种发酵条件进行评价, 得到通过 SDS-CGE 确定的高达 32g/L 的最佳 rEPA 表达 (图 15 和图 16)。利用铜绿假单胞菌外毒素 A 特异性抗体通过 Western 印迹分析确认诱导的蛋白质的身份 (图 17)。表 22 显示了所获得的产率。

[0333] 表 22. rEPA 发酵分析

[0334]

菌株编号	发酵物	产率 (g/L)
PS538-1633	U1	15.5+/-0.7
PS538-1633	U2	11.1+/-0.6
PS538-1640	U3	20.1+/-1.7
PS538-1640	U5	31.9+/-1.6
PS538-1670	U6	20.0+/-0.7
PS538-1670	U7	14.6+/-1.1
PS538-1670	U8	31.0+/-1.7

[0335] 实施例 14. 重组野生型白喉毒素蛋白的高通量表达

[0336] 野生型白喉毒素表达菌株的构建和生长

[0337] 利用荧光假单胞菌优先的密码子构建白喉毒素编码序列以编码野生型白喉毒素氨基酸序列。SEQ ID NO :36 显示了表达的合成白喉毒素基因的氨基酸序列, 和 SEQ ID NO :37 显示了表达的合成优化白喉毒素基因的 DNA 序列。

[0338] 构建携带编码白喉毒素的优化序列的质粒, 该序列与用于 CRM197 的 10 种荧光假单胞菌分泌前导序列编码序列 (表 8 所示) 融合。包含分泌前导序列编码序列, 以将蛋白质靶向至周质, 以便以正确折叠的和活性的形式回收。

[0339] 在荧光假单胞菌宿主中测试表达与重组白喉毒素蛋白融合的 10 种分泌前导序列的构建体。用各前导序列测试表 9 中所列出的 4 种宿主。用所示质粒对宿主细胞进行电穿孔, 并如上文对于 CRM197 高通量表达所述以 96 孔格式生长和诱导。如上文对于 CRM197 高通量表达样品所述制备样品并通过 SDS-CGE 对其进行分析。

[0340] 实施例 15 : 重组野生型白喉毒素蛋白的大规模表达

[0341] 重组野生型白喉毒素蛋白在选择的荧光假单胞菌 Pfënex Expression Technology™菌株中生产。如以上对于 CRM197 大规模表达所述, 选择的菌株在 2 升发酵罐中生长, 用 IPTG 诱导, 并制备样品以供分析。通过 SDS-CGE 分析样品。

[0342] 实施例 16 : 重组霍乱全毒素蛋白的高通量表达

[0343] CTX 表达菌株的构建和生长

[0344] 利用荧光假单胞菌优先的密码子构建 CTX 编码序列以编码 CTX 氨基酸序列。编码序列基于图 14 中所示的 CTX 基因的氨基酸序列和 DNA 序列。

[0345] 构建携带已优化的 CTX 序列的质粒, 该 CTX 序列与用于 CRM197 的 10 种荧光假单胞菌分泌前导序列编码序列 (在表 8 中示出) 融合。包含分泌前导序列以将蛋白质靶向至周质, 以便以正确折叠的和活性的形式回收。

[0346] 在荧光假单胞菌宿主中测试表达与重组 CTX 蛋白融合的 10 种分泌前导序列的构建体。用各表达质粒测试表 9 中所列出的 4 种宿主。用所示质粒对宿主细胞进行电穿孔, 并如上文对于 CRM197 高通量表达所述使宿主细胞以 96 孔格式生长和诱导。如上文对于 CRM197 高通量表达样品所述制备样品并通过 SDS-CGE 对其进行分析。

[0347] 实施例 17: 重组霍乱全毒素蛋白的大规模表达

[0348] 重组霍乱全毒素蛋白在选择的荧光假单胞菌 Pfēnex Expression Technology™ 菌株中生产。如以上对于 CRM197 大规模表达所述, 选择的菌株在 2 升发酵罐中生长, 用 IPTG 诱导, 并制备样品以供分析。通过 SDS-CGE 分析样品。

[0349] 表 23. 序列表

[0350]

SEQ ID NO	说明
1	CRM197 氨基酸序列
2	CRM197 DNA 序列, 已优化的
3	DsbA 分泌前导序列
4	Azu
5	Ibp-S31A
6	Tpr
7	CupB2
8	CupA2
9	NikA
10	Pbp A20V
11	DsbC
12	TolB
13	Pbp
14	Lao
15	CupC2
16	PorE
17	Pbp
18	FlgI
19	ttg2C
20	CRM197 天然前导序列

[0351]

21	切割产物 GADD
22	霍乱毒素 B 氨基酸序列
23	霍乱毒素 B DNA 序列, 已优化的, 带有翻译结果: 编码 SEQ ID NO: 22 的氨基酸序列
24	百日咳毒素 S1 R9K E129A DNA 序列
25	百日咳毒素 S1 R9K E129A 氨基酸序列
26	百日咳毒素 S2 氨基酸序列
27	百日咳毒素 S3 氨基酸序列
28	百日咳毒素 S4 氨基酸序列
29	百日咳毒素 S5 氨基酸序列
30	破伤风毒素 C 氨基酸序列
31	破伤风毒素 C DNA 序列, 已优化的
32	TcdB 氨基酸序列
33	TcdB DNA 序列, 已优化的
34	外毒素 A 氨基酸序列
35	野生型百日咳类毒素 DNA 序列
36	野生型白喉毒素氨基酸序列
37	野生型白喉毒素 DNA 序列, 已优化的
38	霍乱毒素 A 氨基酸序列
39	霍乱毒素 B 氨基酸序列
40	霍乱全毒素(CTX) DNA 序列
41	野生型百日咳毒素 S1 氨基酸序列
42	百日咳毒素 S2 氨基酸序列
43	百日咳毒素 S4 氨基酸序列
44	百日咳毒素 S5 氨基酸序列
45	百日咳毒素 S3 氨基酸序列
46	六-组氨酸亲和标签

[0001]

序列表

<110> RETALLACK, DIANE M.
CHEW, LAWRENCE
JIN, HONGFAN
TALBOT, HENRY W.

<120> 重组毒素蛋白的高水平表达

<130> 38194-735. 201

<140>
<141>

<150> 61/325, 235
<151> 2010-04-16

<150> PCT/US10/30573
<151> 2010-04-09

<150> 61/319, 152
<151> 2010-03-30

<160> 46

<170> PatentIn version 3.5

<210> 1
<211> 535
<212> PRT
<213> 白喉杆菌

<400> 1
Gly Ala Asp Asp Val Val Asp Ser Ser Lys Ser Phe Val Met Glu Asn
1 5 10 15
Phe Ser Ser Tyr His Gly Thr Lys Pro Gly Tyr Val Asp Ser Ile Gln
20 25 30
Lys Gly Ile Gln Lys Pro Lys Ser Gly Thr Gln Gly Asn Tyr Asp Asp
35 40 45
Asp Trp Lys Glu Phe Tyr Ser Thr Asp Asn Lys Tyr Asp Ala Ala Gly
50 55 60
Tyr Ser Val Asp Asn Glu Asn Pro Leu Ser Gly Lys Ala Gly Gly Val
65 70 75 80
Val Lys Val Thr Tyr Pro Gly Leu Thr Lys Val Leu Ala Leu Lys Val
85 90 95
Asp Asn Ala Glu Thr Ile Lys Lys Glu Leu Gly Leu Ser Leu Thr Glu
100 105 110
Pro Leu Met Glu Gln Val Gly Thr Glu Glu Phe Ile Lys Arg Phe Gly
115 120 125
Asp Gly Ala Ser Arg Val Val Leu Ser Leu Pro Phe Ala Glu Gly Ser
130 135 140
Ser Ser Val Glu Tyr Ile Asn Asn Trp Glu Gln Ala Lys Ala Leu Ser
145 150 155 160
Val Glu Leu Glu Ile Asn Phe Glu Thr Arg Gly Lys Arg Gly Gln Asp
165 170 175

[0002]

Ala Met Tyr	Glu Tyr	Met Ala Gln	Ala Cys	Ala Gly	Asn Arg	Val Arg
	180		185		190	
Arg Ser Val	Gly Ser	Ser Leu	Ser Cys	Ile Asn	Leu Asp	Trp Asp Val
	195		200		205	
Ile Arg Asp	Lys Thr	Lys Thr	Lys Ile	Glu Ser	Leu Lys	Glu His Gly
	210		215		220	
Pro Ile Lys	Asn Lys	Met Ser	Glu Ser	Pro Asn	Lys Thr	Val Ser Glu
	225		230		235	240
Glu Lys Ala	Lys Gln	Tyr Leu	Glu Glu	Phe His	Gln Thr	Ala Leu Glu
	245			250		255
His Pro Glu	Leu Ser	Glu Leu	Lys Thr	Val Thr	Gly Thr	Asn Pro Val
	260		265			270
Phe Ala Gly	Ala Asn	Tyr Ala	Ala Trp	Ala Val	Asn Val	Ala Gln Val
	275		280		285	
Ile Asp Ser	Glu Thr	Ala Asp	Asn Leu	Glu Lys	Thr Thr	Ala Ala Leu
	290		295		300	
Ser Ile Leu	Pro Gly	Ile Gly	Ser Val	Met Gly	Ile Ala	Asp Gly Ala
	305		310		315	320
Val His His	Asn Thr	Glu Glu	Ile Val	Ala Gln	Ser Ile	Ala Leu Ser
	325			330		335
Ser Leu Met	Val Ala	Gln Ala	Ile Pro	Leu Val	Gly Glu	Leu Val Asp
	340		345			350
Ile Gly Phe	Ala Ala	Tyr Asn	Phe Val	Glu Ser	Ile Ile	Asn Leu Phe
	355		360		365	
Gln Val Val	His Asn	Ser Tyr	Asn Arg	Pro Ala	Tyr Ser	Pro Gly His
	370		375		380	
Lys Thr Gln	Pro Phe	Leu His	Asp Gly	Tyr Ala	Val Ser	Trp Asn Thr
	385		390		395	400
Val Glu Asp	Ser Ile	Ile Arg	Thr Gly	Phe Gln	Gly Glu	Ser Gly His
	405			410		415
Asp Ile Lys	Ile Thr	Ala Glu	Asn Thr	Pro Leu	Pro Ile	Ala Gly Val
	420		425			430
Leu Leu Pro	Thr Ile	Pro Gly	Lys Leu	Asp Val	Asn Lys	Ser Lys Thr
	435		440		445	
His Ile Ser	Val Asn	Gly Arg	Lys Ile	Arg Met	Arg Cys	Arg Ala Ile
	450		455		460	
Asp Gly Asp	Val Thr	Phe Cys	Arg Pro	Lys Ser	Pro Val	Tyr Val Gly
	465		470		475	480

[0003]

Asn Gly Val His Ala Asn Leu His Val Ala Phe His Arg Ser Ser Ser
485 490 495

Glu Lys Ile His Ser Asn Glu Ile Ser Ser Asp Ser Ile Gly Val Leu
500 505 510

Gly Tyr Gln Lys Thr Val Asp His Thr Lys Val Asn Ser Lys Leu Ser
515 520 525

Leu Phe Phe Glu Ile Lys Ser
530 535

<210> 2

<211> 1605

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述: 合成多核苷酸

<220>

<221> CDS

<222> (1).. (1605)

<400> 2

ggg gcg gac gat gtg gtg gat tcc tcc aag tcg ttt gtc atg gaa aat 48
Gly Ala Asp Asp Val Val Asp Ser Ser Lys Ser Phe Val Met Glu Asn
1 5 10 15

ttc tcg tcg tac cat ggc act aag cca ggc tac gtg gat agc att caa 96
Phe Ser Ser Tyr His Gly Thr Lys Pro Gly Tyr Val Asp Ser Ile Gln
20 25 30

aag ggc atc cag aag ccc aag agc ggt act cag ggg aac tat gac gac 144
Lys Gly Ile Gln Lys Pro Lys Ser Gly Thr Gln Gly Asn Tyr Asp Asp
35 40 45

gac tgg aag gaa ttt tac agc acc gac aat aag tac gat gct gcg ggc 192
Asp Trp Lys Glu Phe Tyr Ser Thr Asp Asn Lys Tyr Asp Ala Ala Gly
50 55 60

tat agc gtg gac aac gaa aac cca ttg tcg ggc aag gcc ggt ggc gtg 240
Tyr Ser Val Asp Asn Glu Asn Pro Leu Ser Gly Lys Ala Gly Gly Val
65 70 75 80

gtg aag gtg acc tat cct ggt ctg acg aaa gtt ctg gcg ttg aaa gtg 288
Val Lys Val Thr Tyr Pro Gly Leu Thr Lys Val Leu Ala Leu Lys Val
85 90 95

gac aac gcc gag act atc aag aaa gaa ttg ggc ttg agt ttg acc gag 336
Asp Asn Ala Glu Thr Ile Lys Lys Glu Leu Gly Leu Ser Leu Thr Glu
100 105 110

ccg ctg atg gaa cag gtg ggt acc gaa gaa ttc att aaa cgt ttt ggc 384
Pro Leu Met Glu Gln Val Gly Thr Glu Glu Phe Ile Lys Arg Phe Gly
115 120 125

gac ggc gcg tcg cgc gtg gtc ctg tcg ttg ccg ttc gcc gaa ggc tcg 432
Asp Gly Ala Ser Arg Val Val Leu Ser Leu Pro Phe Ala Glu Gly Ser
130 135 140

tcg tcg gtg gaa tat atc aac aac tgg gaa cag gcc aag gcg ctg tcg 480
Ser Ser Val Glu Tyr Ile Asn Asn Trp Glu Gln Ala Lys Ala Leu Ser
145 150 155 160

gtg gaa ctg gaa att aac ttc gaa acg cgg ggc aaa cgg ggc cag gac 528
Val Glu Leu Glu Ile Asn Phe Glu Thr Arg Gly Lys Arg Gly Gln Asp
165 170 175

gcc atg tac gaa tac atg gcg cag gcg tgc gcc ggc aac cgg gtg cgg 576

[0004]

Ala Met Tyr Glu Tyr Met Ala Gln Ala Cys Ala Gly Asn Arg Val Arg	
180 185 190	
cgc agc gtg ggc agt tcc ttg agt tgc atc aat ctg gac tgg gac gtc	624
Arg Ser Val Gly Ser Ser Leu Ser Cys Ile Asn Leu Asp Trp Asp Val	
195 200 205	
atc cgc gat aag acg aag acg aaa atc gag tgc ctc aaa gag cac ggc	672
Ile Arg Asp Lys Thr Lys Thr Lys Ile Glu Ser Leu Lys Glu His Gly	
210 215 220	
ccg atc aaa aac aaa atg agc gag tgc ccg aat aaa acg gtg tcc gag	720
Pro Ile Lys Asn Lys Met Ser Glu Ser Pro Asn Lys Thr Val Ser Glu	
225 230 235 240	
gag aag gcg aag caa tac ctg gag gaa ttc cac cag acg gct ctg gag	768
Glu Lys Ala Lys Gln Tyr Leu Glu Glu Phe His Gln Thr Ala Leu Glu	
245 250 255	
cac ccg gag ctg agc gaa ctc aaa acc gtt acc ggt acg aac ccg gtg	816
His Pro Glu Leu Ser Glu Leu Lys Thr Val Thr Gly Thr Asn Pro Val	
260 265 270	
ttt gcc ggg gca aac tat gca gct tgg gcc gtc aac gtg gcc caa gtg	864
Phe Ala Gly Ala Asn Tyr Ala Ala Trp Ala Val Asn Val Ala Gln Val	
275 280 285	
atc gac tcc gaa acg gcc gac aac ctg gaa aag act acc gcc gcg ttg	912
Ile Asp Ser Glu Thr Ala Asp Asn Leu Glu Lys Thr Thr Ala Ala Leu	
290 295 300	
tgc atc ctc ccg ggc atc ggg agc gtc atg ggt att gcc gat ggt gcg	960
Ser Ile Leu Pro Gly Ile Gly Ser Val Met Gly Ile Ala Asp Gly Ala	
305 310 315 320	
gtg cat cac aac acc gaa gag atc gtc gcg cag tgc atc gca ttg tcc	1008
Val His His Asn Thr Glu Glu Ile Val Ala Gln Ser Ile Ala Leu Ser	
325 330 335	
tcc ctg atg gtc gcc caa gct atc ccg ctg gtc ggc gag ctg gtc gat	1056
Ser Leu Met Val Ala Gln Ala Ile Pro Leu Val Gly Glu Leu Val Asp	
340 345 350	
atc ggc ttt gcc gct tat aac ttt gtt gaa tgc atc att aac ctg ttc	1104
Ile Gly Phe Ala Ala Tyr Asn Phe Val Glu Ser Ile Ile Asn Leu Phe	
355 360 365	
cag gtg gtg cat aac agc tac aac cgg cca gcg tac tgc ccc ggt cac	1152
Gln Val Val His Asn Ser Tyr Asn Arg Pro Ala Tyr Ser Pro Gly His	
370 375 380	
aag acc cag ccc ttt ctc cac gac ggc tat gcc gtg tgc tgg aac acc	1200
Lys Thr Gln Pro Phe Leu His Asp Gly Tyr Ala Val Ser Trp Asn Thr	
385 390 395 400	
gtg gag gac agc atc atc cgc acc ggt ttc cag ggc gag agc ggc cat	1248
Val Glu Asp Ser Ile Ile Arg Thr Gly Phe Gln Gly Glu Ser Gly His	
405 410 415	
gac atc aaa att acc gcg gaa aac acg ccc ttg ccg atc gct ggc gtg	1296
Asp Ile Lys Ile Thr Ala Glu Asn Thr Pro Leu Pro Ile Ala Gly Val	
420 425 430	
ttg ctc ccg acg atc ccg ggt aag ctc gac gtc aac aag tcc aag acc	1344
Leu Leu Pro Thr Ile Pro Gly Lys Leu Asp Val Asn Lys Ser Lys Thr	
435 440 445	
cat atc agc gtc aat ggc cgc aag atc cgc atg cgc tgt cgg gcc att	1392
His Ile Ser Val Asn Gly Arg Lys Ile Arg Met Arg Cys Arg Ala Ile	
450 455 460	
gat ggc gac gtc acg ttt tgc cgg ccg aag agt ccc gtc tat gtc ggg	1440
Asp Gly Asp Val Thr Phe Cys Arg Pro Lys Ser Pro Val Tyr Val Gly	
465 470 475 480	
aac ggt gtc cat gcc aac ctg cac gtc gca ttc cac cgg agc agc tgc	1488

[0005]

Asn Gly Val His Ala Asn Leu His Val Ala Phe His Arg Ser Ser Ser	
485 490 495	
gaa aag atc cac agc aat gag atc agc agc gac agc atc ggc gtg ctg	1536
Glu Lys Ile His Ser Asn Glu Ile Ser Ser Asp Ser Ile Gly Val Leu	
500 505 510	
ggg tat caa aag acg gtc gat cac acc aag gtg aac agt aaa ctg agc	1584
Gly Tyr Gln Lys Thr Val Asp His Thr Lys Val Asn Ser Lys Leu Ser	
515 520 525	
ttg ttc ttt gaa atc aag tcg	1605
Leu Phe Phe Glu Ile Lys Ser	
530 535	

<210> 3
 <211> 22
 <212> PRT
 <213> 人工序列

<220>
 <223> 人工序列的描述: 合成肽

<400> 3	
Met Arg Asn Leu Ile Leu Ser Ala Ala Leu Val Thr Ala Ser Leu Phe	
1 5 10 15	

Gly Met Thr Ala Gln Ala
20

<210> 4
 <211> 20
 <212> PRT
 <213> 人工序列

<220>
 <223> 人工序列的描述: 合成肽

<400> 4	
Met Phe Ala Lys Leu Val Ala Val Ser Leu Leu Thr Leu Ala Ser Gly	
1 5 10 15	

Gln Leu Leu Ala
20

<210> 5
 <211> 31
 <212> PRT
 <213> 人工序列

<220>
 <223> 人工序列的描述: 合成多核苷酸

<400> 5	
Met Ile Arg Asp Asn Arg Leu Lys Thr Ser Leu Leu Arg Gly Leu Thr	
1 5 10 15	

Leu Thr Leu Leu Ser Leu Thr Leu Leu Ser Pro Ala Ala His Ala
20 25 30

<210> 6
 <211> 21
 <212> PRT
 <213> 人工序列

<220>
 <223> 人工序列的描述: 合成肽

[0006]

<400> 6
Met Asn Arg Ser Ser Ala Leu Leu Leu Ala Phe Val Phe Leu Ser Gly
1 5 10 15

Cys Gln Ala Met Ala
20

<210> 7
<211> 24
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述: 合成肽

<400> 7
Met Leu Phe Arg Thr Leu Leu Ala Ser Leu Thr Phe Ala Val Ile Ala
1 5 10 15

Gly Leu Pro Ser Thr Ala His Ala
20

<210> 8
<211> 25
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述: 合成肽

<400> 8
Met Ser Cys Thr Arg Ala Phe Lys Pro Leu Leu Leu Ile Gly Leu Ala
1 5 10 15

Thr Leu Met Cys Ser His Ala Phe Ala
20 25

<210> 9
<211> 21
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述: 合成肽

<400> 9
Met Arg Leu Ala Ala Leu Pro Leu Leu Leu Ala Pro Leu Phe Ile Ala
1 5 10 15

Pro Met Ala Val Ala
20

<210> 10
<211> 24
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述: 合成肽

<400> 10
Met Lys Leu Lys Arg Leu Met Ala Ala Met Thr Phe Val Ala Ala Gly

[0007]

1	5	10	15
---	---	----	----

Val Ala Thr Val Asn Ala Val Ala
20

<210> 11
<211> 21
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成肽

<400> 11
Met Arg Leu Thr Gln Ile Ile Ala Ala Ala Ala Ile Ala Leu Val Ser
1 5 10 15

Thr Phe Ala Leu Ala
20

<210> 12
<211> 21
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成肽

<400> 12
Met Arg Asn Leu Leu Arg Gly Met Leu Val Val Ile Cys Cys Met Ala
1 5 10 15

Gly Ile Ala Ala Ala
20

<210> 13
<211> 24
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成肽

<400> 13
Met Lys Leu Lys Arg Leu Met Ala Ala Met Thr Phe Val Ala Ala Gly
1 5 10 15

Val Ala Thr Ala Asn Ala Val Ala
20

<210> 14
<211> 23
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成肽

<400> 14
Met Gln Asn Tyr Lys Lys Phe Leu Leu Ala Ala Ala Val Ser Met Ala
1 5 10 15

Phe Ser Ala Thr Ala Met Ala

[0008]

20

<210> 15
<211> 23
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述: 合成肽

<400> 15
Met Pro Pro Arg Ser Ile Ala Ala Cys Leu Gly Leu Leu Gly Leu Leu
1 5 10 15

Met Ala Thr Gln Ala Ala Ala
20

<210> 16
<211> 21
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述: 合成肽

<400> 16
Met Lys Lys Ser Thr Leu Ala Val Ala Val Thr Leu Gly Ala Ile Ala
1 5 10 15

Gln Gln Ala Gly Ala
20

<210> 17
<211> 24
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述: 合成肽

<400> 17
Met Lys Leu Lys Arg Leu Met Ala Ala Met Thr Phe Val Ala Ala Gly
1 5 10 15

Val Ala Thr Ala Asn Ala Val Ala
20

<210> 18
<211> 21
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述: 合成肽

<400> 18
Met Lys Phe Lys Gln Leu Met Ala Met Ala Leu Leu Leu Ala Leu Ser
1 5 10 15

Ala Val Ala Gln Ala
20

<210> 19

[0009]

<211> 33
 <212> PRT
 <213> 人工序列

<220>
 <223> 人工序列的描述：合成多肽

<400> 19
 Met Gln Asn Arg Thr Val Glu Ile Gly Val Gly Leu Phe Leu Leu Ala
 1 5 10 15

Gly Ile Leu Ala Leu Leu Leu Leu Ala Leu Arg Val Ser Gly Leu Ser
 20 25 30

Ala

<210> 20
 <211> 25
 <212> PRT
 <213> 白喉杆菌

<220>
 <221> MOD_RES
 <222> (9)..(9)
 <223> 任何氨基酸

<400> 20
 Met Ser Arg Lys Leu Phe Ala Ser Xaa Leu Ile Gly Ala Leu Leu Gly
 1 5 10 15

Ile Gly Ala Pro Pro Ser Ala His Ala
 20 25

<210> 21
 <211> 4
 <212> PRT
 <213> 白喉杆菌

<400> 21
 Gly Ala Asp Asp
 1

<210> 22
 <211> 102
 <212> PRT
 <213> 霍乱弧菌

<400> 22
 Thr Pro Gln Asn Ile Thr Asp Leu Cys Ala Glu Tyr His Asn Thr Gln
 1 5 10 15

Ile His Thr Leu Asn Asp Lys Ile Phe Ser Tyr Thr Glu Ser Leu Ala
 20 25 30

Gly Lys Arg Glu Met Ala Ile Ile Thr Phe Lys Asn Gly Ala Thr Phe
 35 40 45

Gln Val Glu Val Pro Gly Ser Gln His Ile Asp Ser Gln Lys Lys Ala
 50 55 60

Ile Glu Arg Met Lys Asp Thr Leu Arg Ile Ala Tyr Leu Thr Glu Ala
 65 70 75 80

[0010]

Lys Val Glu Lys Leu Cys Val Trp Asn Asn Lys Thr Pro His Ala Ile
 85 90 95
 Ala Ala Ile Ser Met Ala
 100
 <210> 23
 <211> 306
 <212> DNA
 <213> 人工序列
 <220>
 <223> 人工序列的描述: 合成多核苷酸
 <220>
 <221> CDS
 <222> (1).. (306)
 <400> 23
 acg ccg caa aat atc acc gac ctg tgc gca gaa tat cac aat acc caa 48
 Thr Pro Gln Asn Ile Thr Asp Leu Cys Ala Glu Tyr His Asn Thr Gln
 1 5 10 15
 atc cat act ctg aac gac aaa atc ttc agc tac acc gag agc ctg gct 96
 Ile His Thr Leu Asn Asp Lys Ile Phe Ser Tyr Thr Glu Ser Leu Ala
 20 25 30
 ggc aag cgc gag atg gcg atc att acg ttc aaa aac ggt gcg acc ttt 144
 Gly Lys Arg Glu Met Ala Ile Ile Thr Phe Lys Asn Gly Ala Thr Phe
 35 40 45
 cag gtg gaa gtc ccc ggc agt cag cac atc gat tcc cag aaa aag gcc 192
 Gln Val Glu Val Pro Gly Ser Gln His Ile Asp Ser Gln Lys Lys Ala
 50 55 60
 att gaa cgg atg aag gac acc ctc cgt atc gcc tac ttg acc gaa gcc 240
 Ile Glu Arg Met Lys Asp Thr Leu Arg Ile Ala Tyr Leu Thr Glu Ala
 65 70 75 80
 aag gtg gag aag ctg tgc gtt tgg aac aac aaa acc ccg cac gcc atc 288
 Lys Val Glu Lys Leu Cys Val Trp Asn Asn Lys Thr Pro His Ala Ile
 85 90 95
 gcg gcc atc tcg atg gcc 306
 Ala Ala Ile Ser Met Ala
 100
 <210> 24
 <211> 3119
 <212> DNA
 <213> 百日咳博德特氏菌
 <220>
 <221> CDS
 <222> (1).. (807)
 <220>
 <221> CDS
 <222> (850).. (1527)
 <220>
 <221> CDS
 <222> (1587).. (1979)
 <220>
 <221> CDS
 <222> (1994).. (2353)
 <220>
 <221> CDS

[0011]

<222> (2439).. (3119)

<400> 24

atg cgt tgc act cgg gca att cgc caa acc gca aga aca ggc tgg ctg Met Arg Cys Thr Arg Ala Ile Arg Gln Thr Ala Arg Thr Gly Trp Leu 1 5 10 15	48
acg tgg ctg gcg att ctt gcc gtc acg gcg ccc gtg act tcg ccg gca Thr Trp Leu Ala Ile Leu Ala Val Thr Ala Pro Val Thr Ser Pro Ala 20 25 30	96
tgg gcc gac gat cct ccc gcc acc gta tac aaa tat gac tcc cgc ccg Trp Ala Asp Asp Pro Pro Ala Thr Val Tyr Lys Tyr Asp Ser Arg Pro 35 40 45	144
ccg gag gac gtt ttc cag aac gga ttc acg gcg tgg gga aac aac gac Pro Glu Asp Val Phe Gln Asn Gly Phe Thr Ala Trp Gly Asn Asn Asp 50 55 60	192
aat gtg ctc gac cat ctg acc gga cgt tcc tgc cag gtc ggc agc agc Asn Val Leu Asp His Thr Gly Arg Ser Cys Gln Val Gly Ser Ser 65 70 75 80	240
aac agc gct ttc gtc tcc acc agc agc agc cgg cgc tat acc gag gtc Asn Ser Ala Phe Val Ser Thr Ser Ser Arg Arg Tyr Thr Glu Val 85 90 95	288
tat ctc gaa cat cgc atg cag gaa gcg gtc gag gcc gaa cgc gcc ggc Tyr Leu Glu His Arg Met Gln Glu Ala Val Glu Ala Glu Arg Ala Gly 100 105 110	336
agg ggc acc ggc cac ttc atc ggc tac atc tac gaa gtc cgc gcc gac Arg Gly Thr Gly His Phe Ile Gly Tyr Ile Tyr Glu Val Arg Ala Asp 115 120 125	384
aac aat ttc tac ggc gcc gcc agc tcg tac ttc gaa tac gtc gac act Asn Asn Phe Tyr Gly Ala Ala Ser Ser Tyr Phe Glu Tyr Val Asp Thr 130 135 140	432
tat ggc gac aat gcc ggc cgt atc ctc gcc ggc gcg ctg gcc acc tac Tyr Gly Asp Asn Ala Gly Arg Ile Leu Ala Gly Ala Leu Ala Thr Tyr 145 150 155 160	480
cag agc gca tat ctg gca cac cgg cgc att ccg ccc gaa aac atc cgc Gln Ser Ala Tyr Leu Ala His Arg Arg Ile Pro Pro Glu Asn Ile Arg 165 170 175	528
agg gta acg cgg gtc tat cac aac ggc atc acc ggc gag acc acg acc Arg Val Thr Arg Val Tyr His Asn Gly Ile Thr Gly Glu Thr Thr Thr 180 185 190	576
acg gag tat tcc aac gct cgc tac gtc agc cag cag act cgc gcc aat Thr Glu Tyr Ser Asn Ala Arg Tyr Val Ser Gln Gln Thr Arg Ala Asn 195 200 205	624
ccc aac ccc tac aca tcg cga agg tcc gta gcg tcg atc gtc ggc aca Pro Asn Pro Tyr Thr Ser Arg Arg Ser Val Ala Ser Ile Val Gly Thr 210 215 220	672
ttg gtg cgc atg gcg ccg gtg ata ggc gct tgc atg gcg cgg cag gcc Leu Val Arg Met Ala Pro Val Ile Gly Ala Cys Met Ala Arg Gln Ala 225 230 235 240	720
gaa agc tcc gag gcc atg gca gcc tgg tcc gaa cgc gcc ggc gag gcg Glu Ser Ser Glu Ala Met Ala Ala Trp Ser Glu Arg Ala Gly Glu Ala 245 250 255	768
atg gtt ctc gtg tac tac gaa agc atc gcg tat tcg ttc tagacctggc Met Val Leu Val Tyr Tyr Glu Ser Ile Ala Tyr Ser Phe 260 265	817
ccagccccgc ccaactccgg taattgaaca gc atg ccg atc gac cgc aag acg Met Pro Ile Asp Arg Lys Thr 270 275	870
ctc tgc cat ctc ctg tcc gtt ctg ccg ttg gcc ctc ctc gga tct cac	918

[0012]

Leu Cys His 280	Leu Leu Ser Val 285	Leu Pro Leu Ala Leu Leu Gly Ser His 290	
gtg gcg cgg gcc tcc acg cca ggc atc gtc att ccg ccg cag gaa cag Val Ala Arg Ala Ser Thr Pro Gly Ile Val Ile Pro Pro Gln Glu Gln 295 300 305			966
att acc cag cat ggc agc ccc tat gga cgc tgc gcg aac aag acc cgt Ile Thr Gln His Gly Ser Pro Tyr Gly Arg Cys Ala Asn Lys Thr Arg 310 315 320			1014
gcc ctg acc gtg gcg gaa ttg cgc ggc agc ggc gat ctg cag gag tac Ala Leu Thr Val Ala Glu Leu Arg Gly Ser Gly Asp Leu Gln Glu Tyr 325 330 335 340			1062
ctg cgt cat gtg acg cgc ggc tgg tca ata ttt gcg ctc tac gat ggc Leu Arg His Val Thr Arg Gly Trp Ser Ile Phe Ala Leu Tyr Asp Gly 345 350 355			1110
acc tat ctc ggc ggc gaa tat ggc ggc gtg atc aag gac gga aca ccc Thr Tyr Leu Gly Gly Glu Tyr Gly Thr Val Ile Lys Asp Gly Thr Pro 360 365 370			1158
ggc ggc gca ttc gac ctg aaa acg acg ttc tgc atc atg acc acg cgc Gly Gly Ala Phe Asp Leu Lys Thr Thr Phe Cys Ile Met Thr Thr Arg 375 380 385			1206
aat acg ggt caa ccc gca acg gat cac tac tac agc aac gtc acc gcc Asn Thr Gly Gln Pro Ala Thr Asp His Tyr Tyr Ser Asn Val Thr Ala 390 395 400			1254
act cgc ctg ctc tcc agc acc aac agc agg cta tgc gcg gtc ttc gtc Thr Arg Leu Leu Ser Ser Thr Asn Ser Arg Leu Cys Ala Val Phe Val 405 410 415 420			1302
aga agc ggg caa ccg gtc att ggc gcc tgc acc agc ccg tat gac ggc Arg Ser Gly Gln Pro Val Ile Gly Ala Cys Thr Ser Pro Tyr Asp Gly 425 430 435			1350
aag tac tgg agc atg tac agc cgg ctg cgg aaa atg ctt tac ctg atc Lys Tyr Trp Ser Met Tyr Ser Arg Leu Arg Lys Met Leu Tyr Leu Ile 440 445 450			1398
tac gtg gcc ggc atc tcc gta cgc gtc cat gtc agc aag gaa gaa cag Tyr Val Ala Gly Ile Ser Val Arg Val His Val Ser Lys Glu Glu Gln 455 460 465			1446
tat tac gac tat gag gac gca acg ttc gag act tac gcc ctt acc ggc Tyr Tyr Asp Tyr Glu Asp Ala Thr Phe Glu Thr Tyr Ala Leu Thr Gly 470 475 480			1494
atc tcc atc tgc aat cct gga tca tcc tta tgc tgagacgctt cccactcga Ile Ser Ile Cys Asn Pro Gly Ser Ser Leu Cys 485 490 495			1547
accaccgccc cgggacaggc cggcgcccg cggtcgcgc gtg cgc gcc ctg gcg Val Arg Ala Leu Ala 500			1601
tgg ttg ctg gca tcc ggc gcg atg acg cat ctt tcc ccc gcc ctg gcc Trp Leu Leu Ala Ser Gly Ala Met Thr His Leu Ser Pro Ala Leu Ala 505 510 515			1649
gac gtt cct tat gtg ctg gtg aag acc aat atg gtg gtc acc agc gta Asp Val Pro Tyr Val Leu Val Lys Thr Asn Met Val Val Thr Ser Val 520 525 530			1697
gcc atg aag ccg tat gaa gtc acc ccg acg cgc atg ctg gtc tgc ggc Ala Met Lys Pro Tyr Glu Val Thr Pro Thr Arg Met Leu Val Cys Gly 535 540 545			1745
atc gcc gcc aaa ctg ggc gcc gcg gcc agc agc ccg gac gcg cac gtg Ile Ala Ala Lys Leu Gly Ala Ala Ala Ser Ser Pro Asp Ala His Val 550 555 560			1793
ccg ttc tgc ttc ggc aag gat ctc aag cgt ccc ggc agc agt ccc atg			1841

[0013]

Pro Phe Cys Phe Gly Lys Asp Leu Lys Arg Pro Gly Ser Ser Pro Met 565 570 575 580	
gaa gtc atg ttg cgc gcc gtc ttc atg caa caa cgg ccg ctg cgc atg Glu Val Met Leu Arg Ala Val Phe Met Gln Gln Arg Pro Leu Arg Met 585 590 595	1889
ttt ctg ggt ccc aag caa ctc act ttc gaa ggc aag ccc gcg ctc gaa Phe Leu Gly Pro Lys Gln Leu Thr Phe Glu Gly Lys Pro Ala Leu Glu 600 605 610	1937
ctg atc cgg atg gtc gaa tgc agc ggc aag cag gat tgc ccc Leu Ile Arg Met Val Glu Cys Ser Gly Lys Gln Asp Cys Pro 615 620 625	1979
tgaaggcgaa cccc atg cat acc atc gca tcc atc ctg ttg tcc gtg ctc Met His Thr Ile Ala Ser Ile Leu Leu Ser Val Leu 630 635	2029
ggc ata tac agc ccg gct gac gtc gcc ggc ttg ccg acc cat ctg tac Gly Ile Tyr Ser Pro Ala Asp Val Ala Gly Leu Pro Thr His Leu Tyr 640 645 650	2077
aag aac ttc act gtc cag gag ctg gcc ttg aaa ctg aag ggc aag aat Lys Asn Phe Thr Val Gln Glu Leu Ala Leu Lys Leu Lys Gly Lys Asn 655 660 665 670	2125
cag gag ttc tgc ctg acc gcc ttc atg tcg ggc aga agc ctg gtc cgg Gln Glu Phe Cys Leu Thr Ala Phe Met Ser Gly Arg Ser Leu Val Arg 675 680 685	2173
gcg tgc ctg tcc gac gcg gga cac gag cac gac acg tgg ttc gac acc Ala Cys Leu Ser Asp Ala Gly His Glu His Asp Thr Trp Phe Asp Thr 690 695 700	2221
atg ctt ggc ttt gcc ata tcc gcg tat gcg ctc aag agc cgg atc gcg Met Leu Gly Phe Ala Ile Ser Ala Tyr Ala Leu Lys Ser Arg Ile Ala 705 710 715	2269
ctg acg gtg gaa gac tcg ccg tat ccg ggc act ccc ggc gat ctg ctc Leu Thr Val Glu Asp Ser Pro Tyr Pro Gly Thr Pro Gly Asp Leu Leu 720 725 730	2317
gaa ctg cag atc tgc ccg ctc aac gga tat tgc gaa tgaacccttc Glu Leu Gln Ile Cys Pro Leu Asn Gly Tyr Cys Glu 735 740 745	2363
cggaggtttc gacgtttccg cgcaatccgc ttgagacgat ctccgccct ggttccattc	2423
cgggaaacacc gcaac atg ctg atc aac aac aag aag ctg ctt cat cac att Met Leu Ile Asn Asn Lys Lys Leu Leu His His Ile 750 755	2474
ctg ccc atc ctg gtg ctc gcc ctg ctg ggc atg cgc acg gcc cag gcc Leu Pro Ile Leu Val Leu Ala Leu Leu Gly Met Arg Thr Ala Gln Ala 760 765 770	2522
gtt gcg cca ggc atc gtc atc ccg ccg aag gca ctg ttc acc caa cag Val Ala Pro Gly Ile Val Ile Pro Pro Lys Ala Leu Phe Thr Gln Gln 775 780 785 790	2570
ggc ggc gcc tat gga cgc tgc ccg aac gga acc cgc gcc ttg acc gtg Gly Gly Ala Tyr Gly Arg Cys Pro Asn Gly Thr Arg Ala Leu Thr Val 795 800 805	2618
gcc gaa ctg cgc ggc aac gcc gaa ttg cag acg tat ttg cgc cag ata Ala Glu Leu Arg Gly Asn Ala Glu Leu Gln Thr Tyr Leu Arg Gln Ile 810 815 820	2666
acg ccc ggc tgg tcc ata tac ggt ctc tat gac ggt acg tac ctg ggc Thr Pro Gly Trp Ser Ile Tyr Gly Leu Tyr Asp Gly Thr Tyr Leu Gly 825 830 835	2714
cag gcg tac ggc ggc atc atc aag gac gcg ccg cca ggc gcg ggg ttc Gln Ala Tyr Gly Gly Ile Ile Lys Asp Ala Pro Pro Gly Ala Gly Phe 840 845 850	2762

[0014]

att tat cgc gaa act ttc tgc atc acg acc ata tac aag acc ggg caa Ile Tyr Arg Glu Thr Phe Cys Ile Thr Thr Ile Tyr Lys Thr Gly Gln 855 860 865 870	2810
ccg gct gcg gat cac tac tac agc aag gtc acg gcc acg cgc ctg ctc Pro Ala Ala Asp His Tyr Tyr Ser Lys Val Thr Ala Thr Arg Leu Leu 875 880 885	2858
gcc agc acc aac agc agg ctg tgc gcg gta ttc gtc agg gac ggg caa Ala Ser Thr Asn Ser Arg Leu Cys Ala Val Phe Val Arg Asp Gly Gln 890 895 900	2906
tcg gtc atc gga gcc tgc gcc agc ccg tat gaa ggc agg tac aga gac Ser Val Ile Gly Ala Cys Ala Ser Pro Tyr Glu Gly Arg Tyr Arg Asp 905 910 915	2954
atg tac gac gcg ctg cgg cgc ctg ctg tac atg atc tat atg tcc ggc Met Tyr Asp Ala Leu Arg Arg Leu Leu Tyr Met Ile Tyr Met Ser Gly 920 925 930	3002
ctt gcc gta cgc gtc cac gtc agc aag gag gaa cag tat tac gac tac Leu Ala Val Arg Val His Val Ser Lys Glu Glu Gln Tyr Tyr Asp Tyr 935 940 945 950	3050
gag gac gcc aca ttc cag acc tat gcc ctc acc ggc att tcc ctc tgc Glu Asp Ala Thr Phe Gln Thr Tyr Ala Leu Thr Gly Ile Ser Leu Cys 955 960 965	3098
aac ccg gca gcg tcg ata tgc Asn Pro Ala Ala Ser Ile Cys 970	3119
<210> 25	
<211> 269	
<212> PRT	
<213> 百日咳博德特氏菌	
<400> 25	
Met Arg Cys Thr Arg Ala Ile Arg Gln Thr Ala Arg Thr Gly Trp Leu 1 5 10 15	
Thr Trp Leu Ala Ile Leu Ala Val Thr Ala Pro Val Thr Ser Pro Ala 20 25 30	
Trp Ala Asp Asp Pro Pro Ala Thr Val Tyr Lys Tyr Asp Ser Arg Pro 35 40 45	
Pro Glu Asp Val Phe Gln Asn Gly Phe Thr Ala Trp Gly Asn Asn Asp 50 55 60	
Asn Val Leu Asp His Leu Thr Gly Arg Ser Cys Gln Val Gly Ser Ser 65 70 75 80	
Asn Ser Ala Phe Val Ser Thr Ser Ser Ser Arg Arg Tyr Thr Glu Val 85 90 95	
Tyr Leu Glu His Arg Met Gln Glu Ala Val Glu Ala Glu Arg Ala Gly 100 105 110	
Arg Gly Thr Gly His Phe Ile Gly Tyr Ile Tyr Glu Val Arg Ala Asp 115 120 125	
Asn Asn Phe Tyr Gly Ala Ala Ser Ser Tyr Phe Glu Tyr Val Asp Thr 130 135 140	

[0015]

Tyr Gly Asp Asn Ala Gly Arg Ile Leu Ala Gly Ala Leu Ala Thr Tyr
 145 150 155 160
 Gln Ser Ala Tyr Leu Ala His Arg Arg Ile Pro Pro Glu Asn Ile Arg
 165 170 175
 Arg Val Thr Arg Val Tyr His Asn Gly Ile Thr Gly Glu Thr Thr Thr
 180 185 190
 Thr Glu Tyr Ser Asn Ala Arg Tyr Val Ser Gln Gln Thr Arg Ala Asn
 195 200 205
 Pro Asn Pro Tyr Thr Ser Arg Arg Ser Val Ala Ser Ile Val Gly Thr
 210 215 220
 Leu Val Arg Met Ala Pro Val Ile Gly Ala Cys Met Ala Arg Gln Ala
 225 230 235 240
 Glu Ser Ser Glu Ala Met Ala Ala Trp Ser Glu Arg Ala Gly Glu Ala
 245 250 255
 Met Val Leu Val Tyr Tyr Glu Ser Ile Ala Tyr Ser Phe
 260 265
 <210> 26
 <211> 226
 <212> PRT
 <213> 百日咳博德特氏菌
 <400> 26
 Met Pro Ile Asp Arg Lys Thr Leu Cys His Leu Leu Ser Val Leu Pro
 1 5 10 15
 Leu Ala Leu Leu Gly Ser His Val Ala Arg Ala Ser Thr Pro Gly Ile
 20 25 30
 Val Ile Pro Pro Gln Glu Gln Ile Thr Gln His Gly Ser Pro Tyr Gly
 35 40 45
 Arg Cys Ala Asn Lys Thr Arg Ala Leu Thr Val Ala Glu Leu Arg Gly
 50 55 60
 Ser Gly Asp Leu Gln Glu Tyr Leu Arg His Val Thr Arg Gly Trp Ser
 65 70 75 80
 Ile Phe Ala Leu Tyr Asp Gly Thr Tyr Leu Gly Gly Glu Tyr Gly Gly
 85 90 95
 Val Ile Lys Asp Gly Thr Pro Gly Gly Ala Phe Asp Leu Lys Thr Thr
 100 105 110
 Phe Cys Ile Met Thr Thr Arg Asn Thr Gly Gln Pro Ala Thr Asp His
 115 120 125
 Tyr Tyr Ser Asn Val Thr Ala Thr Arg Leu Leu Ser Ser Thr Asn Ser
 130 135 140
 Arg Leu Cys Ala Val Phe Val Arg Ser Gly Gln Pro Val Ile Gly Ala
 145 150 155 160

[0016]

Cys Thr Ser Pro Tyr Asp Gly Lys Tyr Trp Ser Met Tyr Ser Arg Leu
 165 170 175
 Arg Lys Met Leu Tyr Leu Ile Tyr Val Ala Gly Ile Ser Val Arg Val
 180 185 190
 His Val Ser Lys Glu Glu Gln Tyr Tyr Asp Tyr Glu Asp Ala Thr Phe
 195 200 205
 Glu Thr Tyr Ala Leu Thr Gly Ile Ser Ile Cys Asn Pro Gly Ser Ser
 210 215 220
 Leu Cys
 225
 <210> 27
 <211> 227
 <212> PRT
 <213> 百日咳博德特氏菌
 <400> 27
 Met Leu Ile Asn Asn Lys Lys Leu Leu His His Ile Leu Pro Ile Leu
 1 5 10 15
 Val Leu Ala Leu Leu Gly Met Arg Thr Ala Gln Ala Val Ala Pro Gly
 20 25 30
 Ile Val Ile Pro Pro Lys Ala Leu Phe Thr Gln Gln Gly Gly Ala Tyr
 35 40 45
 Gly Arg Cys Pro Asn Gly Thr Arg Ala Leu Thr Val Ala Glu Leu Arg
 50 55 60
 Gly Asn Ala Glu Leu Gln Thr Tyr Leu Arg Gln Ile Thr Pro Gly Trp
 65 70 75 80
 Ser Ile Tyr Gly Leu Tyr Asp Gly Thr Tyr Leu Gly Gln Ala Tyr Gly
 85 90 95
 Gly Ile Ile Lys Asp Ala Pro Pro Gly Ala Gly Phe Ile Tyr Arg Glu
 100 105 110
 Thr Phe Cys Ile Thr Thr Ile Tyr Lys Thr Gly Gln Pro Ala Ala Asp
 115 120 125
 His Tyr Tyr Ser Lys Val Thr Ala Thr Arg Leu Leu Ala Ser Thr Asn
 130 135 140
 Ser Arg Leu Cys Ala Val Phe Val Arg Asp Gly Gln Ser Val Ile Gly
 145 150 155 160
 Ala Cys Ala Ser Pro Tyr Glu Gly Arg Tyr Arg Asp Met Tyr Asp Ala
 165 170 175
 Leu Arg Arg Leu Leu Tyr Met Ile Tyr Met Ser Gly Leu Ala Val Arg
 180 185 190

[0017]

Val His Val Ser Lys Glu Glu Gln Tyr Tyr Asp Tyr Glu Asp Ala Thr
 195 200 205
 Phe Gln Thr Tyr Ala Leu Thr Gly Ile Ser Leu Cys Asn Pro Ala Ala
 210 215 220
 Ser Ile Cys
 225
 <210> 28
 <211> 131
 <212> PRT
 <213> 百日咳博德特氏菌
 <400> 28
 Val Arg Ala Leu Ala Trp Leu Leu Ala Ser Gly Ala Met Thr His Leu
 1 5 10 15
 Ser Pro Ala Leu Ala Asp Val Pro Tyr Val Leu Val Lys Thr Asn Met
 20 25 30
 Val Val Thr Ser Val Ala Met Lys Pro Tyr Glu Val Thr Pro Thr Arg
 35 40 45
 Met Leu Val Cys Gly Ile Ala Ala Lys Leu Gly Ala Ala Ala Ser Ser
 50 55 60
 Pro Asp Ala His Val Pro Phe Cys Phe Gly Lys Asp Leu Lys Arg Pro
 65 70 75 80
 Gly Ser Ser Pro Met Glu Val Met Leu Arg Ala Val Phe Met Gln Gln
 85 90 95
 Arg Pro Leu Arg Met Phe Leu Gly Pro Lys Gln Leu Thr Phe Glu Gly
 100 105 110
 Lys Pro Ala Leu Glu Leu Ile Arg Met Val Glu Cys Ser Gly Lys Gln
 115 120 125
 Asp Cys Pro
 130
 <210> 29
 <211> 120
 <212> PRT
 <213> 百日咳博德特氏菌
 <400> 29
 Met His Thr Ile Ala Ser Ile Leu Leu Ser Val Leu Gly Ile Tyr Ser
 1 5 10 15
 Pro Ala Asp Val Ala Gly Leu Pro Thr His Leu Tyr Lys Asn Phe Thr
 20 25 30
 Val Gln Glu Leu Ala Leu Lys Leu Lys Gly Lys Asn Gln Glu Phe Cys
 35 40 45
 Leu Thr Ala Phe Met Ser Gly Arg Ser Leu Val Arg Ala Cys Leu Ser
 50 55 60

[0018]

Asp Ala Gly His Glu His Asp Thr Trp Phe Asp Thr Met Leu Gly Phe
 65 70 75 80
 Ala Ile Ser Ala Tyr Ala Leu Lys Ser Arg Ile Ala Leu Thr Val Glu
 85 90 95
 Asp Ser Pro Tyr Pro Gly Thr Pro Gly Asp Leu Leu Glu Leu Gln Ile
 100 105 110
 Cys Pro Leu Asn Gly Tyr Cys Glu
 115 120
 <210> 30
 <211> 451
 <212> PRT
 <213> 破伤风梭菌
 <400> 30
 Lys Asn Leu Asp Cys Trp Val Asp Asn Glu Glu Asp Ile Asp Val Ile
 1 5 10 15
 Leu Lys Lys Ser Thr Ile Leu Asn Leu Asp Ile Asn Asn Asp Ile Ile
 20 25 30
 Ser Asp Ile Ser Gly Phe Asn Ser Ser Val Ile Thr Tyr Pro Asp Ala
 35 40 45
 Gln Leu Val Pro Gly Ile Asn Gly Lys Ala Ile His Leu Val Asn Asn
 50 55 60
 Glu Ser Ser Glu Val Ile Val His Lys Ala Met Asp Ile Glu Tyr Asn
 65 70 75 80
 Asp Met Phe Asn Asn Phe Thr Val Ser Phe Trp Leu Arg Val Pro Lys
 85 90 95
 Val Ser Ala Ser His Leu Glu Gln Tyr Gly Thr Asn Glu Tyr Ser Ile
 100 105 110
 Ile Ser Ser Met Lys Lys His Ser Leu Ser Ile Gly Ser Gly Trp Ser
 115 120 125
 Val Ser Leu Lys Gly Asn Asn Leu Ile Trp Thr Leu Lys Asp Ser Ala
 130 135 140
 Gly Glu Val Arg Gln Ile Thr Phe Arg Asp Leu Pro Asp Lys Phe Asn
 145 150 155 160
 Ala Tyr Leu Ala Asn Lys Trp Val Phe Ile Thr Ile Thr Asn Asp Arg
 165 170 175
 Leu Ser Ser Ala Asn Leu Tyr Ile Asn Gly Val Leu Met Gly Ser Ala
 180 185 190
 Glu Ile Thr Gly Leu Gly Ala Ile Arg Glu Asp Asn Asn Ile Thr Leu
 195 200 205
 Lys Leu Asp Arg Cys Asn Asn Asn Asn Gln Tyr Val Ser Ile Asp Lys
 210 215 220

[0019]

Phe Arg Ile Phe Cys Lys Ala Leu Asn Pro Lys Glu Ile Glu Lys Leu
 225 230 235 240
 Tyr Thr Ser Tyr Leu Ser Ile Thr Phe Leu Arg Asp Phe Trp Gly Asn
 245 250 255
 Pro Leu Arg Tyr Asp Thr Glu Tyr Tyr Leu Ile Pro Val Ala Ser Ser
 260 265 270
 Ser Lys Asp Val Gln Leu Lys Asn Ile Thr Asp Tyr Met Tyr Leu Thr
 275 280 285
 Asn Ala Pro Ser Tyr Thr Asn Gly Lys Leu Asn Ile Tyr Tyr Arg Arg
 290 295 300
 Leu Tyr Asn Gly Leu Lys Phe Ile Ile Lys Arg Tyr Thr Pro Asn Asn
 305 310 315 320
 Glu Ile Asp Ser Phe Val Lys Ser Gly Asp Phe Ile Lys Leu Tyr Val
 325 330 335
 Ser Tyr Asn Asn Asn Glu His Ile Val Gly Tyr Pro Lys Asp Gly Asn
 340 345 350
 Ala Phe Asn Asn Leu Asp Arg Ile Leu Arg Val Gly Tyr Asn Ala Pro
 355 360 365
 Gly Ile Pro Leu Tyr Lys Lys Met Glu Ala Val Lys Leu Arg Asp Leu
 370 375 380
 Lys Thr Tyr Ser Val Gln Leu Lys Leu Tyr Asp Asp Lys Asn Ala Ser
 385 390 395 400
 Leu Gly Leu Val Gly Thr His Asn Gly Gln Ile Gly Asn Asp Pro Asn
 405 410 415
 Arg Asp Ile Leu Ile Ala Ser Asn Trp Tyr Phe Asn His Leu Lys Asp
 420 425 430
 Lys Ile Leu Gly Cys Asp Trp Tyr Phe Val Pro Thr Asp Glu Gly Trp
 435 440 445
 Thr Asn Asp
 450

<210> 31
 <211> 1353
 <212> DNA
 <213> 人工序列

<220>
 <223> 人工序列的描述：合成多核苷酸

<220>
 <221> CDS
 <222> (1)..(1353)

<400> 31

[0020]

aaa aac ctg gac tgt tgg gtt gac aac gaa gaa gat atc gat gtc atc Lys Asn Leu Asp Cys Trp Val Asp Asn Glu Glu Asp Ile Asp Val Ile 1 5 10 15	48
ctg aag aaa tcc acc att ttg aac ctc gac atc aac aat gac atc att Leu Lys Lys Ser Thr Ile Leu Asn Leu Asp Ile Asn Asn Asp Ile Ile 20 25 30	96
tcc gac att agc ggt ttc aac tgc tcc gtg att acg tac cca gat gct Ser Asp Ile Ser Gly Phe Asn Ser Ser Val Ile Thr Tyr Pro Asp Ala 35 40 45	144
cag ctg gtg ccc ggg att aac ggc aag gct atc cac ctc gtc aac aac Gln Leu Val Pro Gly Ile Asn Gly Lys Ala Ile His Leu Val Asn Asn 50 55 60	192
gag tgc tgc gaa gtc atc gtc cat aaa gcg atg gac atc gag tat aac Glu Ser Ser Glu Val Ile Val His Lys Ala Met Asp Ile Glu Tyr Asn 65 70 75 80	240
gac atg ttt aat aat ttc acc gtg tcc ttt tgg ctg cgc gtg ccc aag Asp Met Phe Asn Asn Phe Thr Val Ser Phe Trp Leu Arg Val Pro Lys 85 90 95	288
gtg tcc gcc tcc cac ctg gaa cag tac ggg acc aac gag tac agc atc Val Ser Ala Ser His Leu Glu Gln Tyr Gly Thr Asn Glu Tyr Ser Ile 100 105 110	336
atc agc tgc atg aag aag cac tgc ttg agc atc ggc agc ggc tgg tgc Ile Ser Ser Met Lys Lys His Ser Leu Ser Ile Gly Ser Gly Trp Ser 115 120 125	384
gtt agt ctc aaa ggg aac aac ctg att tgg acc ctg aaa gat agc gcc Val Ser Leu Lys Gly Asn Asn Leu Ile Trp Thr Leu Lys Asp Ser Ala 130 135 140	432
ggc gag gtg cgt cag atc act ttc cgg gac ctg ccg gat aag ttc aac Gly Glu Val Arg Gln Ile Thr Phe Arg Asp Leu Pro Asp Lys Phe Asn 145 150 155 160	480
gcc tac ctg gca aac aaa tgg gtg ttc att acc atc acg aac gac cgc Ala Tyr Lcu Ala Asn Lys Trp Val Phe Ile Thr Ile Thr Asn Asp Arg 165 170 175	528
ctg agt agc gcg aat ctc tac atc aat ggc gtg ctg atg ggc agc gcg Leu Ser Ser Ala Asn Leu Tyr Ile Asn Gly Val Leu Met Gly Ser Ala 180 185 190	576
gaa atc acg ggc ttg ggt gcc atc cgc gaa gat aac aat atc acc ttg Glu Ile Thr Gly Leu Gly Ala Ile Arg Glu Asp Asn Asn Ile Thr Leu 195 200 205	624
aag ctg gac cgc tgc aac aac aac aac caa tac gtg tcc att gat aag Lys Leu Asp Arg Cys Asn Asn Asn Asn Gln Tyr Val Ser Ile Asp Lys 210 215 220	672
ttc cgc atc ttt tgc aag gcc ctg aac ccg aaa gag atc gaa aag ctc Phe Arg Ile Phe Cys Lys Ala Leu Asn Pro Lys Glu Ile Glu Lys Leu 225 230 235 240	720
tac acc agc tac ttg agt atc acc ttc ctg cgc gac ttt tgg ggt aat Tyr Thr Ser Tyr Leu Ser Ile Thr Phe Leu Arg Asp Phe Trp Gly Asn 245 250 255	768
ccg ttg cgt tat gac acc gag tat tat ctg atc ccc gtg gcc agc agc Pro Leu Arg Tyr Asp Thr Glu Tyr Tyr Leu Ile Pro Val Ala Ser Ser 260 265 270	816
agc aag gac gtc cag ctg aag aac atc acc gac tac atg tac ttg act Ser Lys Asp Val Gln Leu Lys Asn Ile Thr Asp Tyr Met Tyr Leu Thr 275 280 285	864
aac gcg ccc tgc tat acc aat ggc aaa ctg aac att tac tac cgc cgg Asn Ala Pro Ser Tyr Thr Asn Gly Lys Leu Asn Ile Tyr Tyr Arg Arg 290 295 300	912

[0021]

ctc tac aac ggg ctc aag ttc atc atc aaa cgc tat acg ccg aat aat Leu Tyr Asn Gly Leu Lys Phe Ile Ile Lys Arg Tyr Thr Pro Asn Asn 305 310 315 320	960
gaa atc gac agt ttt gtc aag agc ggc gac ttc atc aag ttg tac gtg Glu Ile Asp Ser Phe Val Lys Ser Gly Asp Phe Ile Lys Leu Tyr Val 325 330 335	1008
agc tac aat aac aac gag cac atc gtt ggt tac cct aag gat ggc aac Ser Tyr Asn Asn Asn Glu His Ile Val Gly Tyr Pro Lys Asp Gly Asn 340 345 350	1056
gct ttc aac aac ctc gat cgt atc ctg cgg gtt ggc tac aac gca cca Ala Phe Asn Asn Leu Asp Arg Ile Leu Arg Val Gly Tyr Asn Ala Pro 355 360 365	1104
ggc att ccg ctg tat aag aag atg gaa gcg gtc aaa ctg cgt gac ctg Gly Ile Pro Leu Tyr Lys Lys Met Glu Ala Val Lys Leu Arg Asp Leu 370 375 380	1152
aaa acc tac tcc gtg caa ctg aag ctg tac gac gac aag aat gcc tcg Lys Thr Tyr Ser Val Gln Leu Lys Leu Tyr Asp Asp Lys Asn Ala Ser 385 390 395 400	1200
ttg ggt ctg gtc ggc acg cat aat ggt cag att ggc aac gac ccg aac Leu Gly Leu Val Gly Thr His Asn Gly Gln Ile Gly Asn Asp Pro Asn 405 410 415	1248
cgg gac atc ctg atc gcc agc aac tgg tat ttc aat cac ctg aag gat Arg Asp Ile Leu Ile Ala Ser Asn Trp Tyr Phe Asn His Leu Lys Asp 420 425 430	1296
aag atc ttg ggc tgc gat tgg tat ttc gtc cct acc gat gag ggc tgg Lys Ile Leu Gly Cys Asp Trp Tyr Phe Val Pro Thr Asp Glu Gly Trp 435 440 445	1344
act aat gac Thr Asn Asp 450	1353
<210> 32 <211> 2366 <212> PRT <213> 艰难梭菌	
<400> 32 Met Ser Leu Val Asn Arg Lys Gln Leu Glu Lys Met Ala Asn Val Arg 1 5 10 15	
Phe Arg Thr Gln Glu Asp Glu Tyr Val Ala Ile Leu Asp Ala Leu Glu 20 25 30	
Glu Tyr His Asn Met Ser Glu Asn Thr Val Val Glu Lys Tyr Leu Lys 35 40 45	
Leu Lys Asp Ile Asn Ser Leu Thr Asp Ile Tyr Ile Asp Thr Tyr Lys 50 55 60	
Lys Ser Gly Arg Asn Lys Ala Leu Lys Lys Phe Lys Glu Tyr Leu Val 65 70 75 80	
Thr Glu Val Leu Glu Leu Lys Asn Asn Asn Leu Thr Pro Val Glu Lys 85 90 95	
Asn Leu His Phe Val Trp Ile Gly Gly Gln Ile Asn Asp Thr Ala Ile 100 105 110	
Asn Tyr Ile Asn Gln Trp Lys Asp Val Asn Ser Asp Tyr Asn Val Asn	

[0022]

115	120	125
Val Phe Tyr Asp Ser Asn Ala Phe Leu Ile Asn Thr Leu Lys Lys Thr 130 135 140		
Val Val Glu Ser Ala Ile Asn Asp Thr Leu Glu Ser Phe Arg Glu Asn 145 150 155 160		
Leu Asn Asp Pro Arg Phe Asp Tyr Asn Lys Phe Phe Arg Lys Arg Met 165 170 175		
Glu Ile Ile Tyr Asp Lys Gln Lys Asn Phe Ile Asn Tyr Tyr Lys Ala 180 185 190		
Gln Arg Glu Glu Asn Pro Glu Leu Ile Ile Asp Asp Ile Val Lys Thr 195 200 205		
Tyr Leu Ser Asn Glu Tyr Ser Lys Glu Ile Asp Glu Leu Asn Thr Tyr 210 215 220		
Ile Glu Glu Ser Leu Asn Lys Ile Thr Gln Asn Ser Gly Asn Asp Val 225 230 235 240		
Arg Asn Phe Glu Glu Phe Lys Asn Gly Glu Ser Phe Asn Leu Tyr Glu 245 250 255		
Gln Glu Leu Val Glu Arg Trp Asn Leu Ala Ala Ala Ser Asp Ile Leu 260 265 270		
Arg Ile Ser Ala Leu Lys Glu Ile Gly Gly Met Tyr Leu Asp Val Asp 275 280 285		
Met Leu Pro Gly Ile Gln Pro Asp Leu Phe Glu Ser Ile Glu Lys Pro 290 295 300		
Ser Ser Val Thr Val Asp Phe Trp Glu Met Thr Lys Leu Glu Ala Ile 305 310 315 320		
Met Lys Tyr Lys Glu Tyr Ile Pro Glu Tyr Thr Ser Glu His Phe Asp 325 330 335		
Met Leu Asp Glu Glu Val Gln Ser Ser Phe Glu Ser Val Leu Ala Ser 340 345 350		
Lys Ser Asp Lys Ser Glu Ile Phe Ser Ser Leu Gly Asp Met Glu Ala 355 360 365		
Ser Pro Leu Glu Val Lys Ile Ala Phe Asn Ser Lys Gly Ile Ile Asn 370 375 380		
Gln Gly Leu Ile Ser Val Lys Asp Ser Tyr Cys Ser Asn Leu Ile Val 385 390 395 400		
Lys Gln Ile Glu Asn Arg Tyr Lys Ile Leu Asn Asn Ser Leu Asn Pro 405 410 415		
Ala Ile Ser Glu Asp Asn Asp Phe Asn Thr Thr Thr Asn Thr Phe Ile		

[0023]

420	425	430
Asp Ser Ile Met Ala Glu Ala	Asn Ala Asp Asn Gly Arg Phe Met Met	
435	440	445
Glu Leu Gly Lys Tyr Leu Arg Val Gly Phe Phe Pro Asp Val Lys Thr		
450	455	460
Thr Ile Asn Leu Ser Gly Pro Glu Ala Tyr Ala Ala Ala Tyr Gln Asp		
465	470	475
Leu Leu Met Phe Lys Glu Gly Ser Met Asn Ile His Leu Ile Glu Ala		
485	490	495
Asp Leu Arg Asn Phe Glu Ile Ser Lys Thr Asn Ile Ser Gln Ser Thr		
500	505	510
Glu Gln Glu Met Ala Ser Leu Trp Ser Phe Asp Asp Ala Arg Ala Lys		
515	520	525
Ala Gln Phe Glu Glu Tyr Lys Arg Asn Tyr Phe Glu Gly Ser Leu Gly		
530	535	540
Glu Asp Asp Asn Leu Asp Phe Ser Gln Asn Ile Val Val Asp Lys Glu		
545	550	555
Tyr Leu Leu Glu Lys Ile Ser Ser Leu Ala Arg Ser Ser Glu Arg Gly		
565	570	575
Tyr Ile His Tyr Ile Val Gln Leu Gln Gly Asp Lys Ile Ser Tyr Glu		
580	585	590
Ala Ala Cys Asn Leu Phe Ala Lys Thr Pro Tyr Asp Ser Val Leu Phe		
595	600	605
Gln Lys Asn Ile Glu Asp Ser Glu Ile Ala Tyr Tyr Tyr Asn Pro Gly		
610	615	620
Asp Gly Glu Ile Gln Glu Ile Asp Lys Tyr Lys Ile Pro Ser Ile Ile		
625	630	635
Ser Asp Arg Pro Lys Ile Lys Leu Thr Phe Ile Gly His Gly Lys Asp		
645	650	655
Glu Phe Asn Thr Asp Ile Phe Ala Gly Phe Asp Val Asp Ser Leu Ser		
660	665	670
Thr Glu Ile Glu Ala Ala Ile Asp Leu Ala Lys Glu Asp Ile Ser Pro		
675	680	685
Lys Ser Ile Glu Ile Asn Leu Leu Gly Cys Asn Met Phe Ser Tyr Ser		
690	695	700
Ile Asn Val Glu Glu Thr Tyr Pro Gly Lys Leu Leu Leu Lys Val Lys		
705	710	715
Asp Lys Ile Ser Glu Leu Met Pro Ser Ile Ser Gln Asp Ser Ile Ile		

[0024]

725	730	735
Val Ser Ala Asn Gln Tyr Glu Val	Arg Ile Asn Ser Glu Gly Arg Arg	
740	745	750
Glu Leu Leu Asp His Ser Gly Glu Trp Ile Asn Lys Glu Glu Ser Ile		
755	760	765
Ile Lys Asp Ile Ser Ser Lys Glu Tyr Ile Ser Phe Asn Pro Lys Glu		
770	775	780
Asn Lys Ile Thr Val Lys Ser Lys Asn Leu Pro Glu Leu Ser Thr Leu		
785	790	800
Leu Gln Glu Ile Arg Asn Asn Ser Asn Ser Ser Asp Ile Glu Leu Glu		
805	810	815
Glu Lys Val Met Leu Thr Glu Cys Glu Ile Asn Val Ile Ser Asn Ile		
820	825	830
Asp Thr Gln Ile Val Glu Glu Arg Ile Glu Glu Ala Lys Asn Leu Thr		
835	840	845
Ser Asp Ser Ile Asn Tyr Ile Lys Asp Glu Phe Lys Leu Ile Glu Ser		
850	855	860
Ile Ser Asp Ala Leu Cys Asp Leu Lys Gln Gln Asn Glu Leu Glu Asp		
865	870	875
Ser His Phe Ile Ser Phe Glu Asp Ile Ser Glu Thr Asp Glu Gly Phe		
885	890	895
Ser Ile Arg Phe Ile Asn Lys Glu Thr Gly Glu Ser Ile Phe Val Glu		
900	905	910
Thr Glu Lys Thr Ile Phe Ser Glu Tyr Ala Asn His Ile Thr Glu Glu		
915	920	925
Ile Ser Lys Ile Lys Gly Thr Ile Phe Asp Thr Val Asn Gly Lys Leu		
930	935	940
Val Lys Lys Val Asn Leu Asp Thr Thr His Glu Val Asn Thr Leu Asn		
945	950	955
Ala Ala Phe Phe Ile Gln Ser Leu Ile Glu Tyr Asn Ser Ser Lys Glu		
965	970	975
Ser Leu Ser Asn Leu Ser Val Ala Met Lys Val Gln Val Tyr Ala Gln		
980	985	990
Leu Phe Ser Thr Gly Leu Asn Thr Ile Thr Asp Ala Ala Lys Val Val		
995	1000	1005
Glu Leu Val Ser Thr Ala Leu Asp Glu Thr Ile Asp Leu Leu Pro		
1010	1015	1020
Thr Leu Ser Glu Gly Leu Pro Ile Ile Ala Thr Ile Ile Asp Gly		

[0025]

1025	1030	1035
Val Ser Leu Gly Ala Ala 1040	Ile Lys Glu Leu Ser 1045	Glu Thr Ser Asp 1050
Pro Leu Leu Arg Gln Glu 1055	Ile Glu Ala Lys Ile 1060	Gly Ile Met Ala 1065
Val Asn Leu Thr Thr Ala 1070	Thr Thr Ala Ile Ile 1075	Thr Ser Ser Leu 1080
Gly Ile Ala Ser Gly Phe 1085	Ser Ile Leu Leu Val 1090	Pro Leu Ala Gly 1095
Ile Ser Ala Gly Ile Pro 1100	Ser Leu Val Asn Asn 1105	Glu Leu Val Leu 1110
Arg Asp Lys Ala Thr Lys 1115	Val Val Asp Tyr Phe 1120	Lys His Val Ser 1125
Leu Val Glu Thr Glu Gly 1130	Val Phe Thr Leu Leu 1135	Asp Asp Lys Ile 1140
Met Met Pro Gln Asp Asp 1145	Leu Val Ile Ser Glu 1150	Ile Asp Phe Asn 1155
Asn Asn Ser Ile Val Leu 1160	Gly Lys Cys Glu Ile 1165	Trp Arg Met Glu 1170
Gly Gly Ser Gly His Thr 1175	Val Thr Asp Asp Ile 1180	Asp His Phe Phe 1185
Ser Ala Pro Ser Ile Thr 1190	Tyr Arg Glu Pro His 1195	Leu Ser Ile Tyr 1200
Asp Val Leu Glu Val Gln 1205	Lys Glu Glu Leu Asp 1210	Leu Ser Lys Asp 1215
Leu Met Val Leu Pro Asn 1220	Ala Pro Asn Arg Val 1225	Phe Ala Trp Glu 1230
Thr Gly Trp Thr Pro Gly 1235	Leu Arg Ser Leu Glu 1240	Asn Asp Gly Thr 1245
Lys Leu Leu Asp Arg Ile 1250	Arg Asp Asn Tyr Glu 1255	Gly Glu Phe Tyr 1260
Trp Arg Tyr Phe Ala Phe 1265	Ile Ala Asp Ala Leu 1270	Ile Thr Thr Leu 1275
Lys Pro Arg Tyr Glu Asp 1280	Thr Asn Ile Arg Ile 1285	Asn Leu Asp Ser 1290
Asn Thr Arg Ser Phe Ile 1295	Val Pro Ile Ile Thr 1300	Thr Glu Tyr Ile 1305
Arg Glu Lys Leu Ser Tyr 1310	Ser Phe Tyr Gly Ser 1315	Gly Gly Thr Tyr 1320

[0026]

1310	1315	1320
Ala Leu Ser Leu Ser Gln Tyr 1325	Asn Met Gly Ile Asn 1330	Ile Glu Leu 1335
Ser Glu Ser Asp Val Trp 1340	Ile Ile Asp Val Asp 1345	Asn Val Val Arg 1350
Asp Val Thr Ile Glu Ser 1355	Asp Lys Ile Lys Lys 1360	Gly Asp Leu Ile 1365
Glu Gly Ile Leu Ser Thr 1370	Leu Ser Ile Glu Glu 1375	Asn Lys Ile Ile 1380
Leu Asn Ser His Glu Ile 1385	Asn Phe Ser Gly Glu 1390	Val Asn Gly Ser 1395
Asn Gly Phe Val Ser Leu 1400	Thr Phe Ser Ile Leu 1405	Glu Gly Ile Asn 1410
Ala Ile Ile Glu Val Asp 1415	Leu Leu Ser Lys Ser 1420	Tyr Lys Leu Leu 1425
Ile Ser Gly Glu Leu Lys 1430	Ile Leu Met Leu Asn 1435	Ser Asn His Ile 1440
Gln Gln Lys Ile Asp Tyr 1445	Ile Gly Phe Asn Ser 1450	Glu Leu Gln Lys 1455
Asn Ile Pro Tyr Ser Phe 1460	Val Asp Ser Glu Gly 1465	Lys Glu Asn Gly 1470
Phe Ile Asn Gly Ser Thr 1475	Lys Glu Gly Leu Phe 1480	Val Ser Glu Leu 1485
Pro Asp Val Val Leu Ile 1490	Ser Lys Val Tyr Met 1495	Asp Asp Ser Lys 1500
Pro Ser Phe Gly Tyr Tyr 1505	Ser Asn Asn Leu Lys 1510	Asp Val Lys Val 1515
Ile Thr Lys Asp Asn Val 1520	Asn Ile Leu Thr Gly 1525	Tyr Tyr Leu Lys 1530
Asp Asp Ile Lys Ile Ser 1535	Leu Ser Leu Thr Leu 1540	Gln Asp Glu Lys 1545
Thr Ile Lys Leu Asn Ser 1550	Val His Leu Asp Glu 1555	Ser Gly Val Ala 1560
Glu Ile Leu Lys Phe Met 1565	Asn Arg Lys Gly Asn 1570	Thr Asn Thr Ser 1575
Asp Ser Leu Met Ser Phe 1580	Leu Glu Ser Met Asn 1585	Ile Lys Ser Ile 1590
Phe Val Asn Phe Leu Gln Ser	Asn Ile Lys Phe Ile	Leu Asp Ala

[0027]

1595	1600	1605
Asn Phe 1610	Ile Ile Ser Gly Thr 1615	Thr Ser Ile Gly Gln Phe Glu Phe 1620
Ile Cys 1625	Asp Glu Asn Asp Asn 1630	Ile Gln Pro Tyr Phe Ile Lys Phe 1635
Asn Thr 1640	Leu Glu Thr Asn Tyr 1645	Thr Leu Tyr Val Gly Asn Arg Gln 1650
Asn Met 1655	Ile Val Glu Pro Asn 1660	Tyr Asp Leu Asp Asp Ser Gly Asp 1665
Ile Ser 1670	Ser Thr Val Ile Asn 1675	Phe Ser Gln Lys Tyr Leu Tyr Gly 1680
Ile Asp 1685	Ser Cys Val Asn Lys 1690	Val Val Ile Ser Pro Asn Ile Tyr 1695
Thr Asp 1700	Glu Ile Asn Ile Thr 1705	Pro Val Tyr Glu Thr Asn Asn Thr 1710
Tyr Pro 1715	Glu Val Ile Val Leu 1720	Asp Ala Asn Tyr Ile Asn Glu Lys 1725
Ile Asn 1730	Val Asn Ile Asn Asp 1735	Leu Ser Ile Arg Tyr Val Trp Ser 1740
Asn Asp 1745	Gly Asn Asp Phe Ile 1750	Leu Met Ser Thr Ser Glu Glu Asn 1755
Lys Val 1760	Ser Gln Val Lys Ile 1765	Arg Phe Val Asn Val Phe Lys Asp 1770
Lys Thr 1775	Leu Ala Asn Lys Leu 1780	Ser Phe Asn Phe Ser Asp Lys Gln 1785
Asp Val 1790	Pro Val Ser Glu Ile 1795	Ile Leu Ser Phe Thr Pro Ser Tyr 1800
Tyr Glu 1805	Asp Gly Leu Ile Gly 1810	Tyr Asp Leu Gly Leu Val Ser Leu 1815
Tyr Asn 1820	Glu Lys Phe Tyr Ile 1825	Asn Asn Phe Gly Met Met Val Ser 1830
Gly Leu 1835	Ile Tyr Ile Asn Asp 1840	Ser Leu Tyr Tyr Phe Lys Pro Pro 1845
Val Asn 1850	Asn Leu Ile Thr Gly 1855	Phe Val Thr Val Gly Asp Asp Lys 1860
Tyr Tyr 1865	Phe Asn Pro Ile Asn 1870	Gly Gly Ala Ala Ser Ile Gly Glu 1875
Thr Ile	Ile Asp Asp Lys Asn	Tyr Tyr Phe Asn Gln Ser Gly Val

[0028]

1880	1885	1890
Leu Gln Thr Gly Val Phe Ser 1895	Thr Glu Asp Gly Phe 1900	Lys Tyr Phe 1905
Ala Pro Ala Asn Thr Leu Asp 1910	Glu Asn Leu Glu Gly 1915	Glu Ala Ile 1920
Asp Phe Thr Gly Lys Leu Ile 1925	Ile Asp Glu Asn Ile 1930	Tyr Tyr Phe 1935
Asp Asp Asn Tyr Arg Gly Ala 1940	Val Glu Trp Lys Glu 1945	Leu Asp Gly 1950
Glu Met His Tyr Phe Ser Pro 1955	Glu Thr Gly Lys Ala 1960	Phe Lys Gly 1965
Leu Asn Gln Ile Gly Asp Tyr 1970	Lys Tyr Tyr Phe Asn 1975	Ser Asp Gly 1980
Val Met Gln Lys Gly Phe Val 1985	Ser Ile Asn Asp Asn 1990	Lys His Tyr 1995
Phe Asp Asp Ser Gly Val Met 2000	Lys Val Gly Tyr Thr 2005	Glu Ile Asp 2010
Gly Lys His Phe Tyr Phe Ala 2015	Glu Asn Gly Glu Met 2020	Gln Ile Gly 2025
Val Phe Asn Thr Glu Asp Gly 2030	Phe Lys Tyr Phe Ala 2035	His His Asn 2040
Glu Asp Leu Gly Asn Glu Glu 2045	Gly Glu Glu Ile Ser 2050	Tyr Ser Gly 2055
Ile Leu Asn Phe Asn Asn Lys 2060	Ile Tyr Tyr Phe Asp 2065	Asp Ser Phe 2070
Thr Ala Val Val Gly Trp Lys 2075	Asp Leu Glu Asp Gly 2080	Ser Lys Tyr 2085
Tyr Phe Asp Glu Asp Thr Ala 2090	Glu Ala Tyr Ile Gly 2095	Leu Ser Leu 2100
Ile Asn Asp Gly Gln Tyr Tyr 2105	Phe Asn Asp Asp Gly 2110	Ile Met Gln 2115
Val Gly Phe Val Thr Ile Asn 2120	Asp Lys Val Phe Tyr 2125	Phe Ser Asp 2130
Ser Gly Ile Ile Glu Ser Gly 2135	Val Gln Asn Ile Asp 2140	Asp Asn Tyr 2145
Phe Tyr Ile Asp Asp Asn Gly 2150	Ile Val Gln Ile Gly 2155	Val Phe Asp 2160
Thr Ser Asp Gly Tyr Lys Tyr	Phe Ala Pro Ala Asn	Thr Val Asn

[0029]

2165	2170	2175
Asp Asn Ile Tyr Gly Gln Ala Val Glu Tyr Ser Gly Leu Val Arg		
2180	2185	2190
Val Gly Glu Asp Val Tyr Tyr Phe Gly Glu Thr Tyr Thr Ile Glu		
2195	2200	2205
Thr Gly Trp Ile Tyr Asp Met Glu Asn Glu Ser Asp Lys Tyr Tyr		
2210	2215	2220
Phe Asn Pro Glu Thr Lys Lys Ala Cys Lys Gly Ile Asn Leu Ile		
2225	2230	2235
Asp Asp Ile Lys Tyr Tyr Phe Asp Glu Lys Gly Ile Met Arg Thr		
2240	2245	2250
Gly Leu Ile Ser Phe Glu Asn Asn Asn Tyr Tyr Phe Asn Glu Asn		
2255	2260	2265
Gly Glu Met Gln Phe Gly Tyr Ile Asn Ile Glu Asp Lys Met Phe		
2270	2275	2280
Tyr Phe Gly Glu Asp Gly Val Met Gln Ile Gly Val Phe Asn Thr		
2285	2290	2295
Pro Asp Gly Phe Lys Tyr Phe Ala His Gln Asn Thr Leu Asp Glu		
2300	2305	2310
Asn Phe Glu Gly Glu Ser Ile Asn Tyr Thr Gly Trp Leu Asp Leu		
2315	2320	2325
Asp Glu Lys Arg Tyr Tyr Phe Thr Asp Glu Tyr Ile Ala Ala Thr		
2330	2335	2340
Gly Ser Val Ile Ile Asp Gly Glu Glu Tyr Tyr Phe Asp Pro Asp		
2345	2350	2355
Thr Ala Gln Leu Val Ile Ser Glu		
2360	2365	

<210> 33
 <211> 7098
 <212> DNA
 <213> 人工序列

<220>
 <223> 人工序列的描述: 合成多核苷酸

<220>
 <221> CDS
 <222> (1)..(7098)

<400> 33	
atg tcc ctc gtc aat cgc aag cag ctg gag aag atg gcc aac gtt cgt	48
Met Ser Leu Val Asn Arg Lys Gln Leu Glu Lys Met Ala Asn Val Arg	
1 5 10 15	

ttc cgc acc caa gag gac gaa tac gtc gcc atc ctc gac gcc ctg gaa	96
Phe Arg Thr Gln Glu Asp Glu Tyr Val Ala Ile Leu Asp Ala Leu Glu	
20 25 30	

[0030]

gaa tac cat aac atg agc gaa aac acc gtt gtc gag aag tac ctc aag Glu Tyr His Asn Met Ser Glu Asn Thr Val Val Glu Lys Tyr Leu Lys 35 40 45	144
ctg aag gac atc aac agc ctg acg gac atc tat atc gac acg tac aag Leu Lys Asp Ile Asn Ser Leu Thr Asp Ile Tyr Ile Asp Thr Tyr Lys 50 55 60	192
aag tcc ggt cgc aac aag gca ctc aag aag ttc aaa gag tac ctg gtc Lys Ser Gly Arg Asn Lys Ala Leu Lys Lys Phe Lys Glu Tyr Leu Val 65 70 75 80	240
acc gaa gtg ttg gaa ctc aag aac aac aac ctc acg ccg gtg gaa aag Thr Glu Val Leu Glu Leu Lys Asn Asn Asn Leu Thr Pro Val Glu Lys 85 90 95	288
aac ctg cat ttc gtg tgg att ggc ggc cag atc aac gac acc gcc atc Asn Leu His Phe Val Trp Ile Gly Gly Gln Ile Asn Asp Thr Ala Ile 100 105 110	336
aac tac att aac cag tgg aaa gac gtc aac tcg gac tac aat gtg aat Asn Tyr Ile Asn Gln Trp Lys Asp Val Asn Ser Asp Tyr Asn Val Asn 115 120 125	384
gtg ttc tat gac tcg aac gcc ttt ttg atc aac acg ctg aag aaa acc Val Phe Tyr Asp Ser Asn Ala Phe Leu Ile Asn Thr Leu Lys Lys Thr 130 135 140	432
gtc gtg gaa tcg gcc atc aat gac acc ctg gaa agc ttc cgt gaa aac Val Val Glu Ser Ala Ile Asn Asp Thr Leu Glu Ser Phe Arg Glu Asn 145 150 155 160	480
ctc aac gat cct cgg ttt gac tac aac aag ttt ttc cgc aag cgc atg Leu Asn Asp Pro Arg Phe Asp Tyr Asn Lys Phe Phe Arg Lys Arg Met 165 170 175	528
gag atc atc tat gac aag cag aaa aac ttt atc aac tac tac aaa gcg Glu Ile Ile Tyr Asp Lys Gln Lys Asn Phe Ile Asn Tyr Tyr Lys Ala 180 185 190	576
cag cgc gaa gag aac ccg gag ctg atc atc gac gat atc gtg aaa acc Gln Arg Glu Glu Asn Pro Glu Leu Ile Ile Asp Asp Ile Val Lys Thr 195 200 205	624
tat ctg tcc aac gag tat agt aaa gaa atc gat gag ctg aac acg tat Tyr Leu Ser Asn Glu Tyr Ser Lys Glu Ile Asp Glu Leu Asn Thr Tyr 210 215 220	672
atc gaa gag agt ctg aac aag atc act cag aac agc ggt aac gac gtc Ile Glu Glu Ser Leu Asn Lys Ile Thr Gln Asn Ser Gly Asn Asp Val 225 230 235 240	720
cgg aac ttt gaa gag ttc aaa aac ggc gag tcg ttc aac ctc tac gaa Arg Asn Phe Glu Glu Phe Lys Asn Gly Glu Ser Phe Asn Leu Tyr Glu 245 250 255	768
cag gag ctg gtc gag cgc tgg aac ctg gca gcg gcg tcg gac atc ctg Gln Glu Leu Val Glu Arg Trp Asn Leu Ala Ala Ala Ser Asp Ile Leu 260 265 270	816
cgt atc agc gct ctg aaa gag atc ggc ggc atg tac ctg gac gtg gat Arg Ile Ser Ala Leu Lys Glu Ile Gly Gly Met Tyr Leu Asp Val Asp 275 280 285	864
atg ctc cct ggc atc cag cct gat ctg ttt gaa tcg att gaa aag cca Met Leu Pro Gly Ile Gln Pro Asp Leu Phe Glu Ser Ile Glu Lys Pro 290 295 300	912
tcg agc gtg acc gtc gac ttt tgg gag atg acc aag ctg gag gcg atc Ser Ser Val Thr Val Asp Phe Trp Glu Met Thr Lys Leu Glu Ala Ile 305 310 315 320	960
atg aag tac aaa gaa tac atc ccg gag tat acg agt gaa cac ttt gat Met Lys Tyr Lys Glu Tyr Ile Pro Glu Tyr Thr Ser Glu His Phe Asp 325 330 335	1008

[0031]

atg ctg gac gaa gaa gtg caa tcc tcg ttt gaa agc gtc ctg gcg agc Met Leu Asp Glu Glu Val Gln Ser Ser Phe Glu Ser Val Leu Ala Ser 340 345 350	1056
aag agt gat aag agc gaa atc ttc tcg tcc ttg ggc gat atg gag gcg Lys Ser Asp Lys Ser Glu Ile Phe Ser Ser Leu Gly Asp Met Glu Ala 355 360 365	1104
tcc cca ctg gag gtc aaa atc gcc ttc aac agc aag ggc att atc aat Ser Pro Leu Glu Val Lys Ile Ala Phe Asn Ser Lys Gly Ile Ile Asn 370 375 380	1152
cag ggc ctg att tcg gtc aag gat agc tac tgc agc aac ctg atc gtc Gln Gly Leu Ile Ser Val Lys Asp Ser Tyr Cys Ser Asn Leu Ile Val 385 390 395 400	1200
aag cag atc gag aac cgt tac aag atc ctg aac aac agt ctg aac ccc Lys Gln Ile Glu Asn Arg Tyr Lys Ile Leu Asn Asn Ser Leu Asn Pro 405 410 415	1248
gcc atc agc gaa gat aat gac ttc aat acc acg acg aac acg ttt atc Ala Ile Ser Glu Asp Asn Asp Phe Asn Thr Thr Thr Asn Thr Phe Ile 420 425 430	1296
gac tcc atc atg gcc gaa gcc aac gcg gac aac ggc cgc ttt atg atg Asp Ser Ile Met Ala Glu Ala Asn Ala Asp Asn Gly Arg Phe Met Met 435 440 445	1344
gag ttg ggg aag tac ctg cgc gtg ggc ttc ttc ccg gac gtg aaa acc Glu Leu Gly Lys Tyr Leu Arg Val Gly Phe Phe Pro Asp Val Lys Thr 450 455 460	1392
acg atc aat ctc tcc ggc cca gaa gcg tat gca gcc gca tac caa gat Thr Ile Asn Leu Ser Gly Pro Glu Ala Tyr Ala Ala Ala Tyr Gln Asp 465 470 475 480	1440
ctg ctc atg ttc aaa gag ggc tcg atg aac atc cat ctg att gag gcg Leu Leu Met Phe Lys Glu Gly Ser Met Asn Ile His Leu Ile Glu Ala 485 490 495	1488
gac ttg cgc aac ttc gaa atc tcg aaa acg aac atc agc caa tcg acg Asp Leu Arg Asn Phe Glu Ile Ser Lys Thr Asn Ile Ser Gln Ser Thr 500 505 510	1536
gag cag gaa atg gcg agc ctg tgg tcc ttc gac gat gct cgc gcc aag Glu Gln Glu Met Ala Ser Leu Trp Ser Phe Asp Asp Ala Arg Ala Lys 515 520 525	1584
gcc caa ttt gaa gag tac aaa cgg aac tac ttc gaa ggc tcg ctg ggt Ala Gln Phe Glu Glu Tyr Lys Arg Asn Tyr Phe Glu Gly Ser Leu Gly 530 535 540	1632
gag gat gac aac ttg gac ttc tcg caa aac atc gtc gtg gac aaa gaa Glu Asp Asp Asn Leu Asp Phe Ser Gln Asn Ile Val Val Asp Lys Glu 545 550 555 560	1680
tac ctg ttg gaa aag atc agc tcc ctg gcc cgg agc tcg gaa cgg ggc Tyr Leu Leu Glu Lys Ile Ser Ser Leu Ala Arg Ser Ser Glu Arg Gly 565 570 575	1728
tac atc cac tac att gtt cag ctg caa ggg gat aag atc tcg tat gaa Tyr Ile His Tyr Ile Val Gln Leu Gln Gly Asp Lys Ile Ser Tyr Glu 580 585 590	1776
gcg gcg tgc aat ctc ttc gcc aag acg ccg tac gac tcc gtg ctg ttc Ala Ala Cys Asn Leu Phe Ala Lys Thr Pro Tyr Asp Ser Val Leu Phe 595 600 605	1824
cag aag aac atc gag gac agt gaa atc gcc tac tat tac aac ccc ggt Gln Lys Asn Ile Glu Asp Ser Glu Ile Ala Tyr Tyr Tyr Asn Pro Gly 610 615 620	1872
gac ggc gaa atc caa gaa att gat aag tac aag atc ccg tcc att atc Asp Gly Glu Ile Gln Glu Ile Asp Lys Tyr Lys Ile Pro Ser Ile Ile 625 630 635 640	1920

[0032]

tcc gat cgt ccg aag atc aaa ctc acc ttc att ggc cac ggg aag gac Ser Asp Arg Pro Lys Ile Lys Leu Thr Phe Ile Gly His Gly Lys Asp 645 650 655	1968
gag ttc aac acc gat atc ttc gca ggt ttt gac gtg gat agt ctc tcg Glu Phe Asn Thr Asp Ile Phe Ala Gly Phe Asp Val Asp Ser Leu Ser 660 665 670	2016
acc gag atc gag gcc gcg atc gac ctg gcg aaa gag gac att tcg cca Thr Glu Ile Glu Ala Ala Ile Asp Leu Ala Lys Glu Asp Ile Ser Pro 675 680 685	2064
aag tcg atc gaa atc aac ctg ctg ggc tgc aat atg ttt tcg tat agt Lys Ser Ile Glu Ile Asn Leu Leu Gly Cys Asn Met Phe Ser Tyr Ser 690 695 700	2112
atc aac gtt gaa gaa acc tat ccg ggc aaa ctc ctg ctg aag gtc aag Ile Asn Val Glu Glu Tyr Tyr Pro Gly Lys Leu Leu Lys Val Lys 705 710 715 720	2160
gac aag att agc gaa ctg atg ccg agc atc tcg cag gat agc atc atc Asp Lys Ile Ser Glu Leu Met Pro Ser Ile Ser Gln Asp Ser Ile Ile 725 730 735	2208
gtg agt gct aac cag tat gag gtg cgt atc aac agc gag ggt cgc cgc Val Ser Ala Asn Gln Tyr Glu Val Arg Ile Asn Ser Glu Gly Arg Arg 740 745 750	2256
gag ctg ctc gat cac tcg ggc gag tgg atc aac aaa gaa gag agc atc Glu Leu Leu Asp His Ser Gly Glu Trp Ile Asn Lys Glu Glu Ser Ile 755 760 765	2304
atc aaa gat atc agc agt aaa gaa tac att agt ttc aac ccg aaa gag Ile Lys Asp Ile Ser Ser Lys Glu Tyr Ile Ser Phe Asn Pro Lys Glu 770 775 780	2352
aac aag atc acg gtg aaa tcg aag aat ttg ccg gag ttg agt acc ctg Asn Lys Ile Thr Val Lys Ser Lys Asn Leu Pro Glu Leu Ser Thr Leu 785 790 795 800	2400
ctg cag gag atc cgt aac aat tcc aac tcc agt gat atc gaa ctc gaa Leu Gln Glu Ile Arg Asn Asn Ser Asn Ser Ser Asp Ile Glu Leu Glu 805 810 815	2448
gaa aag gtc atg ctg acc gag tgc gag atc aat gtc atc agc aac atc Glu Lys Val Met Leu Thr Glu Cys Glu Ile Asn Val Ile Ser Asn Ile 820 825 830	2496
gac act cag atc gtg gaa gaa cgc att gaa gag gcc aag aac ctg acc Asp Thr Gln Ile Val Glu Glu Arg Ile Glu Glu Ala Lys Asn Leu Thr 835 840 845	2544
tcg gac tcg atc aac tat atc aaa gac gag ttc aag ctg att gaa tcc Ser Asp Ser Ile Asn Tyr Ile Lys Asp Glu Phe Lys Leu Ile Glu Ser 850 855 860	2592
atc tcg gat gcc ctg tgc gac ctg aag cag cag aac gag ttg gag gat Ile Ser Asp Ala Leu Cys Asp Leu Lys Gln Gln Asn Glu Leu Glu Asp 865 870 875 880	2640
agc cac ttc atc agt ttc gag gat atc tcg gag act gac gag ggc ttc Ser His Phe Ile Ser Phe Glu Asp Ile Ser Glu Thr Asp Glu Gly Phe 885 890 895	2688
agc atc cgt ttc atc aac aaa gaa acc ggt gaa tcc att ttc gtt gaa Ser Ile Arg Phe Ile Asn Lys Glu Thr Gly Glu Ser Ile Phe Val Glu 900 905 910	2736
acg gag aaa acc att ttc agc gaa tac gcc aat cac atc acc gaa gag Thr Glu Ile Thr Ile Phe Ser Glu Tyr Ala Asn His Ile Thr Glu Glu 915 920 925	2784
atc agc aag atc aag ggt act atc ttt gac acg gtg aat ggc aaa ctg Ile Ser Lys Ile Lys Gly Thr Ile Phe Asp Thr Val Asn Gly Lys Leu 930 935 940	2832

[0033]

gtg aag aaa gtc aac ctg gac acc acc cac gag gtc aac acc ctg aac Val Lys Lys Val Asn Leu Asp Thr Thr His Glu Val Asn Thr Leu Asn 945 950 955 960	2880
gcc gcg ttc ttc atc caa agc ctg atc gaa tac aac agc tcc aaa gaa Ala Ala Phe Phe Ile Gln Ser Leu Ile Glu Tyr Asn Ser Ser Lys Glu 965 970 975	2928
tcc ttg agc aac ttg agc gtg gcc atg aaa gtg cag gtg tat gcc cag Ser Leu Ser Asn Leu Ser Val Ala Met Lys Val Gln Val Tyr Ala Gln 980 985 990	2976
ttg ttc tcg acc ggc ctg aac acc att acg gac gcg gca aag gtg gtt Leu Phe Ser Thr Gly Leu Asn Thr Ile Thr Asp Ala Ala Lys Val Val 995 1000 1005	3024
gaa ctg gtc agc acc gcc ttg gac gaa acc atc gac ctg ctg ccg Glu Leu Val Ser Thr Ala Leu Asp Glu Thr Ile Asp Leu Leu Pro 1010 1015 1020	3069
acc ctg agc gag ggc ctg ccc atc atc gcc acg atc att gac ggg Thr Leu Ser Glu Gly Leu Pro Ile Ile Ala Thr Ile Ile Asp Gly 1025 1030 1035	3114
gtc agc ctg ggc gca gcc atc aaa gag ttg agc gag act tcc gac Val Ser Leu Gly Ala Ala Ile Lys Glu Leu Ser Glu Thr Ser Asp 1040 1045 1050	3159
ccg ctg ctg cgc cag gaa atc gaa gct aag atc ggg atc atg gcc Pro Leu Leu Arg Gln Glu Ile Glu Ala Lys Ile Gly Ile Met Ala 1055 1060 1065	3204
gtg aat ctg acc acc gcg acc acc gcg atc att acc tcc agc ctc Val Asn Leu Thr Thr Ala Thr Thr Ala Ile Ile Thr Ser Ser Leu 1070 1075 1080	3249
ggc att gcg tcc ggc ttc tcc atc ctg ctg gtc ccc ttg gcg ggc Gly Ile Ala Ser Gly Phe Ser Ile Leu Leu Val Pro Leu Ala Gly 1085 1090 1095	3294
atc agc gcc ggt atc cct agc ttg gtg aac aac gag ttg gtc ctg Ile Ser Ala Gly Ile Pro Ser Leu Val Asn Asn Glu Leu Val Leu 1100 1105 1110	3339
cgt gat aaa gcg acc aag gtt gtg gat tac ttc aag cat gtc tcc Arg Asp Lys Ala Thr Lys Val Val Asp Tyr Phe Lys His Val Ser 1115 1120 1125	3384
ctg gtc gaa acg gaa ggg gtg ttc acc ctg ctg gac gat aag att Leu Val Glu Thr Glu Gly Val Phe Thr Leu Leu Asp Asp Lys Ile 1130 1135 1140	3429
atg atg ccc caa gac gat ttg gtc atc agc gaa att gac ttt aac Met Met Pro Gln Asp Asp Leu Val Ile Ser Glu Ile Asp Phe Asn 1145 1150 1155	3474
aac aat tcg atc gtc ctc ggg aag tgt gaa atc tgg cgg atg gag Asn Asn Ser Ile Val Leu Gly Lys Cys Glu Ile Trp Arg Met Glu 1160 1165 1170	3519
ggc ggt tcg ggc cac acc gtg acc gat gat atc gac cat ttc ttt Gly Gly Ser Gly His Thr Val Thr Asp Asp Ile Asp His Phe Phe 1175 1180 1185	3564
agt gcg ccc tcc atc acg tac cgc gag ccg cac ctg agc atc tac Ser Ala Pro Ser Ile Thr Tyr Arg Glu Pro His Leu Ser Ile Tyr 1190 1195 1200	3609
gac gtg ctc gag gtg cag aaa gaa gaa ctg gat ctg tcc aaa gat Asp Val Leu Glu Val Gln Lys Glu Glu Leu Asp Leu Ser Lys Asp 1205 1210 1215	3654
ctg atg gtt ctg ccg aac gct ccc aat cgg gtg ttc gct tgg gaa Leu Met Val Leu Pro Asn Ala Pro Asn Arg Val Phe Ala Trp Glu 1220 1225 1230	3699

[0034]

aag ggc tgg acg cca ggc ctc cgc tcg ctg gag aac gac ggg act Thr Gly Trp Thr Pro Gly Leu Arg Ser Leu Glu Asn Asp Gly Thr 1235 1240 1245	3744
aag ttg ctg gat cgc att cgc gac aat tac gaa ggc gaa ttc tac Lys Leu Leu Asp Arg Ile Arg Asp Asn Tyr Glu Gly Glu Phe Tyr 1250 1255 1260	3789
tgg cgc tac ttt gcg ttt atc gcc gac gct ctg att acc acc ctg Trp Arg Tyr Phe Ala Phe Ile Ala Asp Ala Leu Ile Thr Thr Leu 1265 1270 1275	3834
aag ccg cgg tac gag gac act aac atc cgc atc aat ctg gat agc Lys Pro Arg Tyr Glu Asp Thr Asn Ile Arg Ile Asn Leu Asp Ser 1280 1285 1290	3879
aac act cgc agc ttc att gtc ccg atc atc acc act gaa tac att Asn Thr Arg Ser Phe Ile Val Pro Ile Ile Thr Thr Glu Tyr Ile 1295 1300 1305	3924
cgc gaa aag ctg agc tac agc ttt tac ggt agc ggt ggt acc tat Arg Glu Lys Leu Ser Tyr Ser Phe Tyr Gly Ser Gly Thr Tyr 1310 1315 1320	3969
gcc ctg agc ctg tcc cag tac aac atg ggc atc aac att gaa ttg Ala Leu Ser Leu Ser Gln Tyr Asn Met Gly Ile Asn Ile Glu Leu 1325 1330 1335	4014
tcg gag agc gat gtg tgg atc att gac gtg gat aat gtc gtc cgc Ser Glu Ser Asp Val Trp Ile Ile Asp Val Asp Asn Val Val Arg 1340 1345 1350	4059
gat gtg acc att gag agc gac aaa atc aag aaa ggg gat ttg atc Asp Val Thr Ile Glu Ser Asp Lys Ile Lys Lys Gly Asp Leu Ile 1355 1360 1365	4104
gaa ggc atc ctg tcg acc ttg agc atc gaa gag aat aag atc atc Glu Gly Ile Leu Ser Thr Leu Ser Ile Glu Glu Asn Lys Ile Ile 1370 1375 1380	4149
ctg aat agc cac gaa atc aac ttc tcc ggc gaa gtg aac ggc agc Leu Asn Ser His Glu Ile Asn Phe Ser Gly Glu Val Asn Gly Ser 1385 1390 1395	4194
aac ggc ttc gtc agc ctg acc ttt tcc atc ctg gaa ggt atc aac Asn Gly Phe Val Ser Leu Thr Phe Ser Ile Leu Glu Gly Ile Asn 1400 1405 1410	4239
gcc atc att gag gtc gat ctg ctc agc aaa agc tac aag ctg ctc Ala Ile Ile Glu Val Asp Leu Leu Ser Lys Ser Tyr Lys Leu Leu 1415 1420 1425	4284
atc tcg ggc gag ctg aaa atc ctc atg ttg aat tcg aat cac atc Ile Ser Gly Glu Leu Lys Ile Leu Met Leu Asn Ser Asn His Ile 1430 1435 1440	4329
caa cag aaa atc gac tat atc ggc ttc aac agc gag ctg caa aag Gln Gln Lys Ile Asp Tyr Ile Gly Phe Asn Ser Glu Leu Gln Lys 1445 1450 1455	4374
aac atc ccc tac tcg ttt gtg gac agc gaa ggc aaa gag aac ggc Asn Ile Pro Tyr Ser Phe Val Asp Ser Glu Gly Lys Glu Asn Gly 1460 1465 1470	4419
ttc atc aac ggc tcc acc aaa gag ggg ctg ttc gtg agc gag ctg Phe Ile Asn Gly Ser Thr Lys Glu Gly Leu Phe Val Ser Glu Leu 1475 1480 1485	4464
ccg gac gtg gtg ctc atc agc aaa gtg tat atg gac gac agc aag Pro Asp Val Val Leu Ile Ser Lys Val Tyr Met Asp Asp Ser Lys 1490 1495 1500	4509
cct agt ttt ggc tac tac tcc aat aac ctg aaa gat gtt aaa gtg Pro Ser Phe Gly Tyr Tyr Ser Asn Asn Leu Lys Asp Val Lys Val 1505 1510 1515	4554

[0035]

att acc Ile Thr 1520	aag gat aac gtg aac Lys Asp Asn Val Asn 1525	atc ctc acc ggt tac Ile Leu Thr Gly Tyr 1530	tac ctg aag Tyr Leu Lys	4599
gac gac Asp Asp 1535	att aag atc agc ctg Ile Lys Ile Ser Leu 1540	tcc ctg acc ctg caa Ser Leu Thr Leu Gln 1545	gat gaa aag Asp Glu Lys	4644
acc atc Thr Ile 1550	aaa ttg aat agc gtc Lys Leu Asn Ser Val 1555	cat ctc gac gag agt His Leu Asp Glu Ser 1560	ggg gtc gcc Gly Val Ala	4689
gag atc Glu Ile 1565	ttg aag ttc atg aat Leu Lys Phe Met Asn 1570	cgc aag ggt aat acc Arg Lys Gly Asn Thr 1575	aac acg agc Asn Thr Ser	4734
gac tcg Asp Ser 1580	ttg atg tcc ttc ctg Leu Met Ser Phe Leu 1585	gaa agc atg aac atc Glu Ser Met Asn Ile 1590	aag agc att Lys Ser Ile	4779
ttc gtg Phe Val 1595	aac ttc ctg caa agc Asn Phe Leu Gln Ser 1600	aat atc aag ttc att Asn Ile Lys Phe Ile 1605	ttg gac gcg Leu Asp Ala	4824
aac ttt Asn Phe 1610	atc atc tcc ggg acc Ile Ile Ser Gly Thr 1615	acg tcc atc ggc cag Thr Ser Ile Gly Gln 1620	ttt gag ttc Phe Glu Phe	4869
atc tgt Ile Cys 1625	gac gag aac gac aac Asp Glu Asn Asp Asn 1630	att cag ccg tat ttc Ile Gln Pro Tyr Phe 1635	atc aag ttc Ile Lys Phe	4914
aat acc Asn Thr 1640	ttg gaa act aac tac Leu Glu Thr Asn Tyr 1645	acc ctg tac gtt ggc Thr Leu Tyr Val Gly 1650	aac cgg caa Asn Arg Gln	4959
aac atg Asn Met 1655	att gtc gaa ccc aac Ile Val Glu Pro Asn 1660	tat gac ttg gat gac Tyr Asp Leu Asp Asp 1665	agt ggt gat Ser Gly Asp	5004
atc agt Ile Ser 1670	agc acc gtg att aac Ser Thr Val Ile Asn 1675	ttt tcc cag aag tac Phe Ser Gln Lys Tyr 1680	ctg tat ggc Leu Tyr Gly	5049
atc gac Ile Asp 1685	tcc tgc gtg aac aaa Ser Cys Val Asn Lys 1690	gtg gtg atc tcg ccg Val Val Ile Ser Pro 1695	aat atc tac Asn Ile Tyr	5094
acg gac Thr Asp 1700	gaa atc aat atc act Glu Ile Asn Ile Thr 1705	ccc gtc tat gaa acc Pro Val Tyr Glu Thr 1710	aac aac acc Asn Asn Thr	5139
tac ccc Tyr Pro 1715	gag gtg att gtc ttg Glu Val Ile Val Leu 1720	gat gcg aac tac att Asp Ala Asn Tyr Ile 1725	aac gaa aag Asn Glu Lys	5184
att aac Ile Asn 1730	gtg aac att aac gac Val Asn Ile Asn Asp 1735	ctg agc atc cgg tat Leu Ser Ile Arg Tyr 1740	gtg tgg agt Val Trp Ser	5229
aat gac Asn Asp 1745	ggg aac gac ttc att Gly Asn Asp Phe Ile 1750	ctg atg agc acc tcc Leu Met Ser Thr Ser 1755	gaa gaa aac Glu Glu Asn	5274
aaa gtc Lys Val 1760	tcg caa gtc aag atc Ser Gln Val Lys Ile 1765	cgc ttc gtt aac gtt Arg Phe Val Asn Val 1770	ttc aaa gac Phe Lys Asp	5319
aag acc Lys Thr 1775	ttg gcc aac aaa ctc Leu Ala Asn Lys Leu 1780	agc ttc aat ttc tcg Ser Phe Asn Phe Ser 1785	gac aaa cag Asp Lys Gln	5364
gac gtg Asp Val 1790	cct gtg tcg gag atc Pro Val Ser Glu Ile 1795	att ctc agt ttc acc Ile Leu Ser Phe Thr 1800	ccg agc tac Pro Ser Tyr	5409

[0036]

tac gag Tyr Glu 1805	gac ggc ctg atc ggt Asp Gly Leu Ile Gly 1810	tac gac ctg ggc ctg Tyr Asp Leu Gly Leu 1815	gtt agc ctc Val Ser Leu	5454
tac aac Tyr Asn 1820	gaa aag ttc tat atc Glu Lys Phe Tyr Ile 1825	aac aat ttc ggg atg Asn Asn Phe Gly Met 1830	atg gtt tcg Met Val Ser	5499
ggt ctg Gly Leu 1835	atc tat atc aat gac Ile Tyr Ile Asn Asp 1840	agc ttg tac tac ttc Ser Leu Tyr Tyr Phe 1845	aaa cct ccg Lys Pro Pro	5544
gtg aac Val Asn 1850	aat ttg atc acc ggc Asn Leu Ile Thr Gly 1855	ttc gtg acc gtg ggc Phe Val Thr Val Gly 1860	gat gac aag Asp Asp Lys	5589
tac tat Tyr Tyr 1865	ttc aac ccg att aac Phe Asn Pro Ile Asn 1870	ggt ggc gct gcg tcg Gly Gly Ala Ala Ser 1875	att ggc gaa Ile Gly Glu	5634
acc atc Thr Ile 1880	atc gac gac aag aac Ile Asp Asp Lys Asn 1885	tac tat ttc aac cag Tyr Tyr Phe Asn Gln 1890	agc ggt gtg Ser Gly Val	5679
ctg caa Leu Gln 1895	acc ggg gtc ttt agc Thr Gly Val Phe Ser 1900	acg gag gat ggg ttt Thr Glu Asp Gly Phe 1905	aag tat ttc Lys Tyr Phe	5724
gcg cct Ala Pro 1910	gcc aat acc ctg gac Ala Asn Thr Leu Asp 1915	gag aat ttg gag ggc Glu Asn Leu Glu Gly 1920	gaa gcc att Glu Ala Ile	5769
gac ttc Asp Phe 1925	acc ggc aaa ctg att Thr Gly Lys Leu Ile 1930	atc gac gaa aac atc Ile Asp Glu Asn Ile 1935	tac tat ttc Tyr Tyr Phe	5814
gat gac Asp Asp 1940	aac tac cgc ggt gcg Asn Tyr Arg Gly Ala 1945	gtc gaa tgg aaa gag Val Glu Trp Lys Glu 1950	ttg gac ggg Leu Asp Gly	5859
gag atg Glu Met 1955	cac tat ttc tcg cca His Tyr Phe Ser Pro 1960	gag act ggt aag gcc Glu Thr Gly Lys Ala 1965	ttc aag ggc Phe Lys Gly	5904
ctg aac Leu Asn 1970	cag atc ggc gac tac Gln Ile Gly Asp Tyr 1975	aag tac tat ttc aat Lys Tyr Tyr Phe Asn 1980	agc gac ggc Ser Asp Gly	5949
gtg atg Val Met 1985	cag aag ggg ttc gtc Gln Lys Gly Phe Val 1990	agc atc aac gac aac Ser Ile Asn Asp Asn 1995	aag cat tac Lys His Tyr	5994
ttc gac Phe Asp 2000	gat agc ggc gtt atg Asp Ser Gly Val Met 2005	aag gtg ggc tat act Lys Val Gly Tyr Thr 2010	gag atc gac Glu Ile Asp	6039
ggc aag Gly Lys 2015	cac ttc tat ttc gcc His Phe Tyr Phe Ala 2020	gag aac ggc gag atg Glu Asn Gly Glu Met 2025	caa atc ggc Gln Ile Gly	6084
gtg ttc Val Phe 2030	aac acc gag gac ggc Asn Thr Glu Asp Gly 2035	ttc aaa tac ttc gct Phe Lys Tyr Phe Ala 2040	cat cac aac His His Asn	6129
gaa gat Glu Asp 2045	ctc ggt aat gaa gaa Leu Gly Asn Glu Glu 2050	ggt gaa gag att tcc Gly Glu Glu Ile Ser 2055	tat tcg ggc Tyr Ser Gly	6174
atc ctg Ile Leu 2060	aac ttc aac aat aag Asn Phe Asn Asn Lys 2065	atc tac tac ttt gat Ile Tyr Tyr Phe Asp 2070	gac tcg ttc Asp Ser Phe	6219
acc gcc Thr Ala 2075	gtg gtg ggt tgg aag Val Val Gly Trp Lys 2080	gac ttg gag gac ggg Asp Leu Glu Asp Gly 2085	agc aag tat Ser Lys Tyr	6264

[0037]

tac ttc Tyr Phe 2090	gac gaa gat acc gcc Asp Glu Asp Thr Ala 2095	gag gca tac atc ggc Glu Ala Tyr Ile Gly 2100	ttg tcg ctc Leu Ser Leu	6309
atc aac Ile Asn 2105	gac ggt cag tac tac Asp Gly Gln Tyr Tyr 2110	ttc aac gac gat ggc Phe Asn Asp Asp Gly 2115	atc atg cag Ile Met Gln	6354
gtg ggc Val Gly 2120	ttc gtg act atc aac Phe Val Thr Ile Asn 2125	gac aag gtg ttc tac Asp Lys Val Phe Tyr 2130	ttc agt gac Phe Ser Asp	6399
agt ggc Ser Gly 2135	atc att gag tcg ggc Ile Ile Glu Ser Gly 2140	gtg cag aat atc gac Val Gln Asn Ile Asp 2145	gat aac tat Asp Asn Tyr	6444
ttc tat Phe Tyr 2150	atc gat gac aat ggc Ile Asp Asp Asn Gly 2155	att gtg cag atc ggc Ile Val Gln Ile Gly 2160	gtg ttt gat Val Phe Asp	6489
acc tcc Thr Ser 2165	gat ggg tat aag tat Asp Gly Tyr Lys Tyr 2170	ttc gca cca gca aat Phe Ala Pro Ala Asn 2175	acc gtc aat Thr Val Asn	6534
gac aac Asp Asn 2180	atc tac ggc cag gcc Ile Tyr Gly Gln Ala 2185	gtg gag tac agc ggc Val Glu Tyr Ser Gly 2190	ctg gtt cgt Leu Val Arg	6579
gtg ggc Val Gly 2195	gag gac gtt tac tat Glu Asp Val Tyr Tyr 2200	ttc ggt gag act tat Phe Gly Glu Thr Tyr 2205	act atc gaa Thr Ile Glu	6624
acc ggc Thr Gly 2210	tgg atc tat gat atg Trp Ile Tyr Asp Met 2215	gaa aac gaa tcg gat Glu Asn Glu Ser Asp 2220	aag tac tac Lys Tyr Tyr	6669
ttt aac Phe Asn 2225	cca gaa acg aag aaa Pro Glu Thr Lys Lys 2230	gcc tgc aag ggc atc Ala Cys Lys Gly Ile 2235	aac ctc att Asn Leu Ile	6714
gat gac Asp Asp 2240	atc aag tat tac ttt Ile Lys Tyr Tyr Phe 2245	gac gaa aag ggt atc Asp Glu Lys Gly Ile 2250	atg cgc acc Met Arg Thr	6759
ggc ctc Gly Leu 2255	atc tcg ttt gag aac Ile Ser Phe Glu Asn 2260	aac aac tac tat ttc Asn Asn Tyr Tyr Phe 2265	aac gag aat Asn Glu Asn	6804
ggc gaa Gly Glu 2270	atg cag ttt ggg tac Met Gln Phe Gly Tyr 2275	atc aat atc gag gat Ile Asn Ile Glu Asp 2280	aag atg ttc Lys Met Phe	6849
tac ttt Tyr Phe 2285	ggg gag gac ggc gtc Gly Glu Asp Gly Val 2290	atg cag att ggt gtg Met Gln Ile Gly Val 2295	ttt aac acc Phe Asn Thr	6894
ccg gat Pro Asp 2300	ggc ttc aag tac ttc Gly Phe Lys Tyr Phe 2305	gcc cat cag aac act Ala His Gln Asn Thr 2310	ctg gac gaa Leu Asp Glu	6939
aac ttc Asn Phe 2315	gag ggc gaa agc atc Glu Gly Glu Ser Ile 2320	aat tac act ggc tgg Asn Tyr Thr Gly Trp 2325	ctg gac ctg Leu Asp Leu	6984
gat gag Asp Glu 2330	aaa cgc tac tac ttc Lys Arg Tyr Tyr Phe 2335	acc gac gag tac att Thr Asp Glu Tyr Ile 2340	gcc gcc acg Ala Ala Thr	7029
ggc tcc Gly Ser 2345	gtg att atc gac ggc Val Ile Ile Asp Gly 2350	gaa gaa tac tat ttc Glu Glu Tyr Tyr Phe 2355	gat ccc gac Asp Pro Asp	7074
acc gcc Thr Ala 2360	cag ttg gtc att agc Gln Leu Val Ile Ser 2365	gaa Glu		7098

[0038]

```

<210> 34
<211> 613
<212> PRT
<213> 绿脓假单胞菌

<220>
<221> MOD_RES
<222> (426)..(426)
<223> His或Tyr

<220>
<221> MOD_RES
<222> (553)..(553)
<223> 可以存在或不存在

<400> 34
Ala Glu Glu Ala Phe Asp Leu Trp Asn Glu Cys Ala Lys Ala Cys Val
1          5          10          15

Leu Asp Leu Lys Asp Gly Val Arg Ser Ser Arg Met Ser Val Asp Pro
20          25          30

Ala Ile Ala Asp Thr Asn Gly Gln Gly Val Leu His Tyr Ser Met Val
35          40          45

Leu Glu Gly Gly Asn Asp Ala Leu Lys Leu Ala Ile Asp Asn Ala Leu
50          55          60

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

Pro Asn Lys Pro Val Arg Tyr Ser Tyr Thr Arg Gln Ala Arg Gly Ser
85          90          95

Trp Ser Leu Asn Trp Leu Val Pro Ile Gly His Glu Lys Pro Ser Asn
100         105         110

Ile Lys Val Phe Ile His Glu Leu Asn Ala Gly Asn Gln Leu Ser His
115         120         125

Met Ser Pro Ile Tyr Thr Ile Glu Met Gly Asp Glu Leu Leu Ala Lys
130         135         140

Leu Ala Arg Asp Ala Thr Phe Phe Val Arg Ala His Glu Ser Asn Glu
145         150         155         160

Met Gln Pro Thr Leu Ala Ile Ser His Ala Gly Val Ser Val Val Met
165         170         175

Ala Gln Ala Gln Pro Arg Arg Glu Lys Arg Trp Ser Glu Trp Ala Ser
180         185         190

Gly Lys Val Leu Cys Leu Leu Asp Pro Leu Asp Gly Val Tyr Asn Tyr
195         200         205

Leu Ala Gln Gln Arg Cys Asn Leu Asp Asp Thr Trp Glu Gly Lys Ile
210         215         220

Tyr Arg Val Leu Ala Gly Asn Pro Ala Lys His Asp Leu Asp Ile Lys

```

[0039]

225	230	235	240
Pro Thr Val Ile Ser His Arg Leu His Phe Pro Glu Gly Gly Ser Leu	245	250	255
Ala Ala Leu Thr Ala His Gln Ala Cys His Leu Pro Leu Glu Thr Phe	260	265	270
Thr Arg His Arg Gln Pro Arg Gly Trp Glu Gln Leu Glu Gln Cys Gly	275	280	285
Tyr Pro Val Gln Arg Leu Val Ala Leu Tyr Leu Ala Ala Arg Leu Ser	290	295	300
Trp Asn Gln Val Asp Gln Val Ile Arg Asn Ala Leu Ala Ser Pro Gly	305	310	315
Ser Gly Gly Asp Leu Gly Glu Ala Ile Arg Glu Gln Pro Glu Gln Ala	325	330	335
Arg Leu Ala Leu Thr Leu Ala Ala Ala Glu Ser Glu Arg Phe Val Arg	340	345	350
Gln Gly Thr Gly Asn Asp Glu Ala Gly Ala Ala Ser Ala Asp Val Val	355	360	365
Ser Leu Thr Cys Pro Val Ala Ala Gly Glu Cys Ala Gly Pro Ala Asp	370	375	380
Ser Gly Asp Ala Leu Leu Glu Arg Asn Tyr Pro Thr Gly Ala Glu Phe	385	390	395
Leu Gly Asp Gly Gly Asp Ile Ser Phe Ser Thr Arg Gly Thr Gln Asn	405	410	415
Trp Thr Val Glu Arg Leu Leu Gln Ala Xaa Arg Gln Leu Glu Glu Arg	420	425	430
Gly Tyr Val Phe Val Gly Tyr His Gly Thr Phe Leu Glu Ala Ala Gln	435	440	445
Ser Ile Val Phe Gly Gly Val Arg Ala Arg Ser Gln Asp Leu Asp Ala	450	455	460
Ile Trp Arg Gly Phe Tyr Ile Ala Gly Asp Pro Ala Leu Ala Tyr Gly	465	470	475
Tyr Ala Gln Asp Gln Glu Pro Asp Ala Arg Gly Arg Ile Arg Asn Gly	485	490	495
Ala Leu Leu Arg Val Tyr Val Pro Arg Ser Ser Leu Pro Gly Phe Tyr	500	505	510
Arg Thr Gly Leu Thr Leu Ala Ala Pro Glu Ala Ala Gly Glu Val Glu	515	520	525
Arg Leu Ile Gly His Pro Leu Pro Leu Arg Leu Asp Ala Ile Thr Gly			

[0040]

530	535	540	
Pro Glu Glu Glu Gly Gly Arg Leu Glu Thr Ile Leu Gly Trp Pro Leu 545 550 555 560			
Ala Glu Arg Thr Val Val Ile Pro Ser Ala Ile Pro Thr Asp Pro Arg 565 570 575			
Asn Val Gly Gly Asp Leu Asp Pro Ser Ser Ile Pro Asp Lys Glu Gln 580 585 590			
Ala Ile Ser Ala Leu Pro Asp Tyr Ala Ser Gln Pro Gly Lys Pro Pro 595 600 605			
Arg Glu Asp Leu Lys 610			
<210> 35 <211> 3119 <212> DNA <213> 百日咳博德特氏菌			
<220> <221> CDS <222> (1).. (807)			
<220> <221> CDS <222> (850).. (1527)			
<220> <221> CDS <222> (1587).. (1979)			
<220> <221> CDS <222> (1994).. (2353)			
<220> <221> CDS <222> (2439).. (3119)			
<400> 35 atg cgt tgc act cgg gca att cgc caa acc gca aga aca ggc tgg ctg Met Arg Cys Thr Arg Ala Ile Arg Gln Thr Ala Arg Thr Gly Trp Leu 1 5 10 15			48
acg tgg ctg gcg att ctt gcc gtc acg gcg ccc gtg act tgc ccg gca Thr Trp Leu Ala Ile Leu Ala Val Thr Ala Pro Val Thr Ser Pro Ala 20 25 30			96
tgg gcc gac gat cct ccc gcc acc gta tac cgc tat gac tcc cgc ccg Trp Ala Asp Asp Pro Pro Ala Thr Val Tyr Arg Tyr Asp Ser Arg Pro 35 40 45			144
ccg gag gac gtt ttc cag aac gga ttc acg gcg tgg gga aac aac gac Pro Glu Asp Val Phe Gln Asn Gly Phe Thr Ala Trp Gly Asn Asn Asp 50 55 60			192
aat gtg ctc gac cat ctg acc gga cgt tcc tgc cag gtc ggc agc agc Asn Val Leu Asp His Leu Thr Gly Arg Ser Cys Gln Val Gly Ser Ser 65 70 75 80			240
aac agc gct ttc gtc tcc acc agc agc agc cgg cgc tat acc gag gtc Asn Ser Ala Phe Val Ser Thr Ser Ser Ser Arg Arg Tyr Thr Glu Val 85 90 95			288
tat ctc gaa cat cgc atg cag gaa gcg gtc gag gcc gaa cgc gcc ggc Tyr Leu Glu His Arg Met Gln Glu Ala Val Glu Ala Glu Arg Ala Gly			336

[0041]

100								105				110								
agg Arg	ggc Gly	acc Thr 115	ggc Gly	cac His	ttc Phe	atc Ile	ggc Gly 120	tac Tyr	atc Ile	tac Tyr	gaa Glu	gtc Val 125	cgc Arg	gcc Ala	gac Asp	384				
aac Asn 130	aat Asn	ttc Phe	tac Tyr	ggc Gly	gcc Ala	gcc Ala 135	agc Ser	tcg Ser	tac Tyr	ttc Phe	gaa Glu 140	tac Tyr	gtc Val	gac Asp	act Thr	432				
tat Tyr 145	ggc Gly	gac Asp	aat Asn	gcc Ala	ggc Gly 150	cgt Arg	atc Ile	ctc Leu	gcc Ala	ggc Gly 155	gcg Ala	ctg Leu	gcc Ala	acc Thr	tac Tyr 160	480				
cag Gln	agc Ser	gaa Glu	tat Tyr	ctg Leu 165	gca Ala	cac His	cgg Arg	cgc Arg	att Ile 170	ccg Pro	ccc Pro	gaa Glu	aac Asn	atc Ile 175	cgc Arg	528				
agg Arg	gta Val	acg Thr 180	cgg Arg	gtc Val	tat Tyr	cac His	aac Asn	ggc Gly 185	atc Ile	acc Thr	ggc Gly	gag Glu	acc Thr 190	acg Thr	acc Thr	576				
acg Thr	gag Glu	tat Tyr 195	tcc Ser	aac Asn	gct Ala	cgc Arg	tac Tyr 200	gtc Val	agc Ser	cag Gln	cag Gln	act Thr 205	cgc Arg	gcc Ala	aat Asn	624				
ccc Pro	aac Asn 210	ccc Pro	tac Tyr	aca Thr	tcg Ser	cga Arg 215	agg Arg	tcc Ser	gta Val	gcg Ala	tcg Ser 220	atc Ile	gtc Val	ggc Gly	aca Thr	672				
ttg Leu 225	gtg Val	cgc Arg	atg Met	gcg Ala	ccg Pro 230	gtg Val	ata Ile	ggc Gly	gct Ala	tgc Cys 235	atg Met	gcg Ala	cgg Arg	cag Gln	gcc Ala 240	720				
gaa Glu	agc Ser	tcc Ser	gag Glu 245	gcc Ala	atg Met	gca Ala	gcc Ala	tgg Trp	tcc Ser 250	gaa Glu	cgc Arg	gcc Ala	ggc Gly	gag Glu 255	gcg Ala	768				
atg Met	gtt Val	ctc Leu	gtg Val 260	tac Tyr	tac Tyr	gaa Glu	agc Ser	atc Ile 265	gcg Ala	tat Tyr	tcg Ser	ttc Phe	tagacctg	ggc		817				
ccagccccgc ccaactccgg taattgaaca gc								atg Met 270	ccg Pro	atc Ile	gac Asp	cgc Arg	aag Lys 275	acg Thr		870				
ctc Leu	tgc Cys	cat His 280	ctc Leu	ctg Leu	tcc Ser	gtt Val	ctg Leu	ccg Pro 285	ttg Leu	gcc Ala	ctc Leu	ctc Leu	gga Gly 290	tct Ser	cac His	918				
gtg Val	gcg Ala	cgg Arg 295	gcc Ala	tcc Ser	acg Thr	cca Pro	ggc Gly 300	atc Ile	gtc Val	att Ile	ccg Pro	ccg Pro	cag Gln	gaa Glu	cag Gln	966				
att Ile	acc Thr 310	cag Gln	cat His	ggc Gly	agc Ser	ccc Pro 315	tat Tyr	gga Gly	cgc Arg	tgc Cys	gcg Ala	aac Asn	aag Lys	acc Thr	cgt Arg	1014				
gcc Ala 325	ctg Leu	acc Thr	gtg Val	gcg Ala	gaa Glu 330	ttg Leu	cgc Arg	ggc Gly	agc Ser	ggc Gly 335	gat Asp	ctg Leu	cag Gln	gag Glu	tac Tyr 340	1062				
ctg Leu	cgt Arg	cat His	gtg Val	acg Thr 345	cgc Arg	ggc Gly	tgg Trp	tca Ser	ata Ile	ttt Phe	gcg Ala	ctc Leu	tac Tyr	gat Asp 355	ggc Gly	1110				
acc Thr	tat Tyr	ctc Leu	ggc Gly 360	ggc Gly	gaa Glu	tat Tyr	ggc Gly	ggc Gly 365	gtg Val	atc Ile	aag Lys	gac Asp	gga Gly 370	aca Thr	ccc Pro	1158				
ggc Gly	ggc Gly	gca Ala 375	ttc Phe	gac Asp	ctg Leu	aaa Lys	acg Thr 380	acg Thr	ttc Phe	tgc Cys	atc Ile	atg Met 385	acc Thr	acg Thr	cgc Arg	1206				
aat Asn	acg Thr	ggt Gly	caa Gln	ccc Pro	gca Ala	acg Thr	gat Asp	cac His	tac Tyr	tac Tyr	agc Ser	aac Asn	gtc Val	acc Thr	gcc Ala	1254				

[0042]

390	395	400	
act cgc ctg ctc tcc agc acc aac agc agg cta tgc gcg gtc ttc gtc Thr Arg Leu Leu Ser Ser Thr Asn Ser Arg Leu Cys Ala Val Phe Val 405 410 415 420			1302
aga agc ggg caa ccg gtc att ggc gcc tgc acc agc ccg tat gac ggc Arg Ser Gly Gln Pro Val Ile Gly Ala Cys Thr Ser Pro Tyr Asp Gly 425 430 435			1350
aag tac tgg agc atg tac agc cgg ctg cgg aaa atg ctt tac ctg atc Lys Tyr Trp Ser Met Tyr Ser Arg Leu Arg Lys Met Leu Tyr Leu Ile 440 445 450			1398
tac gtg gcc ggc atc tcc gta cgc gtc cat gtc agc aag gaa gaa cag Tyr Val Ala Gly Ile Ser Val Arg Val His Val Ser Lys Glu Glu Gln 455 460 465			1446
tat tac gac tat gag gac gca acg ttc gag act tac gcc ctt acc ggc Tyr Tyr Asp Tyr Glu Asp Ala Thr Phe Glu Thr Tyr Ala Leu Thr Gly 470 475 480			1494
atc tcc atc tgc aat cct gga tca tcc tta tgc tgagacgctt cccactcga Ile Ser Ile Cys Asn Pro Gly Ser Ser Leu Cys 485 490 495			1547
accacgcgcc cgggacaggg cggcgcccg cggtcgcgc gtg cgc gcc ctg gcg Val Arg Ala Leu Ala 500			1601
tgg ttg ctg gca tcc ggc gcg atg acg cat ctt tcc ccc gcc ctg gcc Trp Leu Leu Ala Ser Gly Ala Met Thr His Leu Ser Pro Ala Leu Ala 505 510 515			1649
gac gtt cct tat gtg ctg gtg aag acc aat atg gtg gtc acc agc gta Asp Val Pro Tyr Val Leu Val Lys Thr Asn Met Val Val Thr Ser Val 520 525 530			1697
gcc atg aag ccg tat gaa gtc acc ccg acg cgc atg ctg gtc tgc ggc Ala Met Lys Pro Tyr Glu Val Thr Pro Thr Arg Met Leu Val Cys Gly 535 540 545			1745
atc gcc gcc aaa ctg ggc gcc gcg gcc agc agc ccg gac gcg cac gtg Ile Ala Ala Lys Leu Gly Ala Ala Ala Ser Ser Pro Asp Ala His Val 550 555 560			1793
ccg ttc tgc ttc ggc aag gat ctc aag cgt ccc ggc agc agt ccc atg Pro Phe Cys Phe Gly Lys Asp Leu Lys Arg Pro Gly Ser Ser Pro Met 565 570 575 580			1841
gaa gtc atg ttg cgc gcc gtc ttc atg caa caa cgg ccg ctg cgc atg Glu Val Met Leu Arg Ala Val Phe Met Gln Gln Arg Pro Leu Arg Met 585 590 595			1889
ttt ctg ggt ccc aag caa ctc act ttc gaa ggc aag ccc gcg ctc gaa Phe Leu Gly Pro Lys Gln Leu Thr Phe Glu Gly Lys Pro Ala Leu Glu 600 605 610			1937
ctg atc cgg atg gtc gaa tgc agc ggc aag cag gat tgc ccc Leu Ile Arg Met Val Glu Cys Ser Gly Lys Gln Asp Cys Pro 615 620 625			1979
tgaaggcgaa cccc atg cat acc atc gca tcc atc ctg ttg tcc gtg ctc Met His Thr Ile Ala Ser Ile Leu Leu Ser Val Leu 630 635			2029
ggc ata tac agc ccg gct gac gtc gcc ggc ttg ccg acc cat ctg tac Gly Ile Tyr Ser Pro Ala Asp Val Ala Gly Leu Pro Thr His Leu Tyr 640 645 650			2077
aag aac ttc act gtc cag gag ctg gcc ttg aaa ctg aag ggc aag aat Lys Asn Phe Thr Val Gln Glu Leu Ala Leu Lys Leu Lys Gly Lys Asn 655 660 665 670			2125
cag gag ttc tgc ctg acc gcc ttc atg tgc ggc aga agc ctg gtc cgc Gln Glu Phe Cys Leu Thr Ala Phe Met Ser Gly Arg Ser Leu Val Arg 675 680 685 690			2173

[0043]

675	680	685	
gcg tgc ctg tcc gac gcg gga cac gag cac gac acg tgg ttc gac acc Ala Cys Leu Ser Asp Ala Gly His Glu His Asp Thr Trp Phe Asp Thr 690 695 700			2221
atg ctt ggc ttt gcc ata tcc gcg tat gcg ctc aag agc cgg atc gcg Met Leu Gly Phe Ala Ile Ser Ala Tyr Ala Leu Lys Ser Arg Ile Ala 705 710 715			2269
ctg acg gtg gaa gac tcg ccg tat ccg ggc act ccc ggc gat ctg ctc Leu Thr Val Glu Asp Ser Pro Tyr Pro Gly Thr Pro Gly Asp Leu Leu 720 725 730			2317
gaa ctg cag atc tgc ccg ctc aac gga tat tgc gaa tgaacccttc Glu Leu Gln Ile Cys Pro Leu Asn Gly Tyr Cys Glu 735 740 745			2363
cgagggtttc gacgtttccg cgcaatccgc ttgagacgat ctccgccct gggtccattc cggaacacc gcaac atg ctg atc aac aac aag aag ctg ctt cat cac att Met Leu Ile Asn Asn Lys Lys Leu Leu His His Ile 750 755			2423 2474
ctg ccc atc ctg gtg ctc gcc ctg ctg ggc atg cgc acg gcc cag gcc Leu Pro Ile Leu Val Leu Ala Leu Leu Gly Met Arg Thr Ala Gln Ala 760 765 770			2522
gtt gcg cca ggc atc gtc atc ccg ccg aag gca ctg ttc acc caa cag Val Ala Pro Gly Ile Val Ile Pro Pro Lys Ala Leu Phe Thr Gln Gln 775 780 785 790			2570
ggc ggc gcc tat gga cgc tgc ccg aac gga acc cgc gcc ttg acc gtg Gly Gly Ala Tyr Gly Arg Cys Pro Asn Gly Thr Arg Ala Leu Thr Val 795 800 805			2618
gcc gaa ctg cgc ggc aac gcc gaa ttg cag acg tat ttg cgc cag ata Ala Glu Leu Arg Gly Asn Ala Glu Leu Gln Thr Tyr Leu Arg Gln Ile 810 815 820			2666
acg ccc ggc tgg tcc ata tac ggt ctc tat gac ggt acg tac ctg ggc Thr Pro Gly Trp Ser Ile Tyr Gly Leu Tyr Asp Gly Thr Tyr Leu Gly 825 830 835			2714
cag gcg tac ggc ggc atc atc aag gac gcg ccg cca ggc gcg ggg ttc Gln Ala Tyr Gly Gly Ile Ile Lys Asp Ala Pro Pro Gly Ala Gly Phe 840 845 850			2762
att tat cgc gaa act ttc tgc atc acg acc ata tac aag acc ggg caa Ile Tyr Arg Glu Thr Phe Cys Ile Thr Thr Ile Tyr Lys Thr Gly Gln 855 860 865 870			2810
ccg gct gcg gat cac tac tac agc aag gtc acg gcc acg cgc ctg ctc Pro Ala Ala Asp His Tyr Tyr Ser Lys Val Thr Ala Thr Arg Leu Leu 875 880 885			2858
gcc agc acc aac agc agg ctg tgc gcg gta ttc gtc agg gac ggg caa Ala Ser Thr Asn Ser Arg Leu Cys Ala Val Phe Val Arg Asp Gly Gln 890 895 900			2906
tcg gtc atc gga gcc tgc gcc agc ccg tat gaa ggc agg tac aga gac Ser Val Ile Gly Ala Cys Ala Ser Pro Tyr Glu Gly Arg Tyr Arg Asp 905 910 915			2954
atg tac gac gcg ctg cgg cgc ctg ctg tac atg atc tat atg tcc ggc Met Tyr Asp Ala Leu Arg Arg Leu Leu Tyr Met Ile Tyr Met Ser Gly 920 925 930			3002
ctt gcc gta cgc gtc cac gtc agc aag gag gaa cag tat tac gac tac Leu Ala Val Arg Val His Val Ser Lys Glu Glu Gln Tyr Tyr Asp Tyr 935 940 945 950			3050
gag gac gcc aca ttc cag acc tat gcc ctc acc ggc att tcc ctc tgc Glu Asp Ala Thr Phe Gln Thr Tyr Ala Leu Thr Gly Ile Ser Leu Cys 955 960 965			3098

[0044]

aac ccg gca gcg tcg ata tgc
Asn Pro Ala Ala Ser Ile Cys
970

3119

<210> 36
<211> 535
<212> PRT
<213> 白喉杆菌

<400> 36
Gly Ala Asp Asp Val Val Asp Ser Ser Lys Ser Phe Val Met Glu Asn
1 5 10 15
Phe Ser Ser Tyr His Gly Thr Lys Pro Gly Tyr Val Asp Ser Ile Gln
20 25 30
Lys Gly Ile Gln Lys Pro Lys Ser Gly Thr Gln Gly Asn Tyr Asp Asp
35 40 45
Asp Trp Lys Gly Phe Tyr Ser Thr Asp Asn Lys Tyr Asp Ala Ala Gly
50 55 60
Tyr Ser Val Asp Asn Glu Asn Pro Leu Ser Gly Lys Ala Gly Gly Val
65 70 75 80
Val Lys Val Thr Tyr Pro Gly Leu Thr Lys Val Leu Ala Leu Lys Val
85 90 95
Asp Asn Ala Glu Thr Ile Lys Lys Glu Leu Gly Leu Ser Leu Thr Glu
100 105 110
Pro Leu Met Glu Gln Val Gly Thr Glu Glu Phe Ile Lys Arg Phe Gly
115 120 125
Asp Gly Ala Ser Arg Val Val Leu Ser Leu Pro Phe Ala Glu Gly Ser
130 135 140
Ser Ser Val Glu Tyr Ile Asn Asn Trp Glu Gln Ala Lys Ala Leu Ser
145 150 155 160
Val Glu Leu Glu Ile Asn Phe Glu Thr Arg Gly Lys Arg Gly Gln Asp
165 170 175
Ala Met Tyr Glu Tyr Met Ala Gln Ala Cys Ala Gly Asn Arg Val Arg
180 185 190
Arg Ser Val Gly Ser Ser Leu Ser Cys Ile Asn Leu Asp Trp Asp Val
195 200 205
Ile Arg Asp Lys Thr Lys Thr Lys Ile Glu Ser Leu Lys Glu His Gly
210 215 220
Pro Ile Lys Asn Lys Met Ser Glu Ser Pro Asn Lys Thr Val Ser Glu
225 230 235 240
Glu Lys Ala Lys Gln Tyr Leu Glu Glu Phe His Gln Thr Ala Leu Glu
245 250 255
His Pro Glu Leu Ser Glu Leu Lys Thr Val Thr Gly Thr Asn Pro Val

[0045]

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

<210> 37
 <211> 960
 <212> DNA
 <213> 人工序列

[0046]

<220>

<223> 人工序列的描述: 合成多核苷酸

<220>

<221> CDS

<222> (1).. (960)

<400> 37

```

ggg gcg gac gat gtg gtg gat tcc tcc aag tcg ttt gtc atg gaa aat      48
Gly Ala Asp Asp Val Val Asp Ser Ser Lys Ser Phe Val Met Glu Asn
1                               10                               15

```

```

ttc tcg tcg tac cat ggc act aag cca ggc tac gtg gat agc att caa      96
Phe Ser Ser Tyr His Gly Thr Lys Pro Gly Tyr Val Asp Ser Ile Gln
                20                25                30

```

```

aag ggc atc cag aag ccc aag agc ggt act cag ggc aac tat gac gac      144
Lys Gly Ile Gln Lys Pro Lys Ser Gly Thr Gln Gly Asn Tyr Asp Asp
                35                40                45

```

```

gac tgg aag gga ttt tac agc acc gac aat aag tac gat gct gcg ggc      192
Asp Trp Lys Gly Phe Tyr Ser Thr Asp Asn Lys Tyr Asp Ala Ala Gly
50                55                60

```

```

tat agc gtg gac aac gaa aac cca ttg tcg ggc aag gcc ggt ggc gtg      240
Tyr Ser Val Asp Asn Glu Asn Pro Leu Ser Gly Lys Ala Gly Gly Val
65                70                75                80

```

```

gtg aag gtg acc tat cct ggt ctg acg aaa gtt ctg gcg ttg aaa gtg      288
Val Lys Val Thr Tyr Pro Gly Leu Thr Lys Val Leu Ala Leu Lys Val
                85                90                95

```

```

gac aac gcc gag act atc aag aaa gaa ttg ggc ttg agt ttg acc gag      336
Asp Asn Ala Glu Thr Ile Lys Lys Glu Leu Gly Leu Ser Leu Thr Glu
100                105                110

```

```

ccg ctg atg gaa cag gtg ggt acc gaa gaa ttc att aaa cgt ttt ggg      384
Pro Leu Met Glu Gln Val Gly Thr Glu Glu Phe Ile Lys Arg Phe Gly
115                120                125

```

```

gac ggc gcg tcg cgc gtg gtc ctg tcg ttg ccg ttc gcc gaa ggc tcg      432
Asp Gly Ala Ser Arg Val Val Leu Ser Leu Pro Phe Ala Glu Gly Ser
130                135                140

```

```

tcg tcg gtg gaa tat atc aac aac tgg gaa cag gcc aag gcg ctg tcg      480
Ser Ser Val Glu Tyr Ile Asn Asn Trp Glu Gln Ala Lys Ala Leu Ser
145                150                155                160

```

```

gtg gaa ctg gaa att aac ttc gaa acg cgg ggc aaa cgg ggc cag gac      528
Val Glu Leu Glu Ile Asn Phe Glu Thr Arg Gly Lys Arg Gly Gln Asp
165                170                175

```

```

gcc atg tac gaa tac atg gcg cag gcg tgc gcc ggc aac cgg gtg cgg      576
Ala Met Tyr Glu Tyr Met Ala Gln Ala Cys Ala Gly Asn Arg Val Arg
180                185                190

```

```

cgc agc gtg ggc agt tcc ttg agt tgc atc aat ctg gac tgg gac gtc      624
Arg Ser Val Gly Ser Ser Leu Ser Cys Ile Asn Leu Asp Trp Asp Val
195                200                205

```

```

atc cgc gat aag acg aag acg aaa atc gag tcg ctc aaa gag cac ggc      672
Ile Arg Asp Lys Thr Lys Thr Lys Ile Glu Ser Leu Lys Glu His Gly
210                215                220

```

```

ccg atc aaa aac aaa atg agc gag tcg ccg aat aaa acg gtg tcc gag      720
Pro Ile Lys Asn Lys Met Ser Glu Ser Pro Asn Lys Thr Val Ser Glu
225                230                235                240

```

```

gag aag gcg aag caa tac ctg gag gaa ttc cac cag acg gct ctg gag      768
Glu Lys Ala Lys Gln Tyr Leu Glu Glu Phe His Gln Thr Ala Leu Glu
245                250                255

```

```

cac ccg gag ctg agc gaa ctc aaa acc gtt acc ggt acg aac ccg gtg      816
His Pro Glu Leu Ser Glu Leu Lys Thr Val Thr Gly Thr Asn Pro Val
260                265                270

```

[0047]

```

ttt gcc ggg gca aac tat gca gct tgg gcc gtc aac gtg gcc caa gtg      864
Phe Ala Gly Ala Asn Tyr Ala Ala Trp Ala Val Asn Val Ala Gln Val
      275                      280                      285

atc gac tcc gaa acg gcc gac aac ctg gaa aag act acc gcc gcg ttg      912
Ile Asp Ser Glu Thr Ala Asp Asn Leu Glu Lys Thr Thr Ala Ala Leu
      290                      295                      300

tcg atc ctc ccg ggc atc ggg agc gtc atg ggt att gcc gat ggt gcg      960
Ser Ile Leu Pro Gly Ile Gly Ser Val Met Gly Ile Ala Asp Gly Ala
      305                      310                      315                      320

<210> 38
<211> 258
<212> PRT
<213> 霍乱弧菌

<400> 38
Met Val Lys Ile Ile Phe Val Phe Phe Ile Phe Leu Ser Ser Phe Ser
1                      5                      10                      15

Tyr Ala Asn Asp Asp Lys Leu Tyr Arg Ala Asp Ser Arg Pro Pro Asp
      20                      25                      30

Glu Ile Lys Gln Ser Gly Gly Leu Met Pro Arg Gly Gln Ser Glu Tyr
      35                      40                      45

Phe Asp Arg Gly Thr Gln Met Asn Ile Asn Leu Tyr Asp His Ala Arg
      50                      55                      60

Gly Thr Gln Thr Gly Phe Val Arg His Asp Asp Gly Tyr Val Ser Thr
      65                      70                      75                      80

Ser Ile Ser Leu Arg Ser Ala His Leu Val Gly Gln Thr Ile Leu Ser
      85                      90                      95

Gly His Ser Thr Tyr Tyr Ile Tyr Val Ile Ala Thr Ala Pro Asn Met
      100                     105                     110

Phe Asn Val Asn Asp Val Leu Gly Ala Tyr Ser Pro His Pro Asp Glu
      115                     120                     125

Gln Glu Val Ser Ala Leu Gly Gly Ile Pro Tyr Ser Gln Ile Tyr Gly
      130                     135                     140

Trp Tyr Arg Val His Phe Gly Val Leu Asp Glu Gln Leu His Arg Asn
      145                     150                     155                     160

Arg Gly Tyr Arg Asp Arg Tyr Tyr Ser Asn Leu Asp Ile Ala Pro Ala
      165                     170                     175

Ala Asp Gly Tyr Gly Leu Ala Gly Phe Pro Pro Glu His Arg Ala Trp
      180                     185                     190

Arg Glu Glu Pro Trp Ile His His Ala Pro Pro Gly Cys Gly Asn Ala
      195                     200                     205

Pro Arg Ser Ser Met Ser Asn Thr Cys Asp Glu Lys Thr Gln Ser Leu
      210                     215                     220

```

[0048]

Gly Val Lys Phe Leu Asp Glu Tyr Gln Ser Lys Val Lys Arg Gln Ile
225 230 235 240

Phe Ser Gly Tyr Gln Ser Asp Ile Asp Thr His Asn Arg Ile Lys Asp
245 250 255

Glu Leu

<210> 39
<211> 124
<212> PRT
<213> 霍乱弧菌

<400> 39
Met Ile Lys Leu Lys Phe Gly Val Phe Phe Thr Val Leu Leu Ser Ser
1 5 10 15

Ala Tyr Ala His Gly Thr Pro Gln Asn Ile Thr Asp Leu Cys Ala Glu
20 25 30

Tyr His Asn Thr Gln Ile Tyr Thr Leu Asn Asp Lys Ile Phe Ser Tyr
35 40 45

Thr Glu Ser Leu Ala Gly Lys Arg Glu Met Ala Ile Ile Thr Phe Lys
50 55 60

Asn Gly Ala Ile Phe Gln Val Glu Val Pro Gly Ser Gln His Ile Asp
65 70 75 80

Ser Gln Lys Lys Ala Ile Glu Arg Met Lys Asp Thr Leu Arg Ile Ala
85 90 95

Tyr Leu Thr Glu Ala Lys Val Glu Lys Leu Cys Val Trp Asn Asn Lys
100 105 110

Thr Pro His Ala Ile Ala Ala Ile Ser Met Ala Asn
115 120

<210> 40
<211> 1148
<212> DNA
<213> 霍乱弧菌

<400> 40
atggtaaaga taatatttgt gttttttatt ttcttatcat cattttcata tgcaaatgat 60
gataagtat atcgggcaga ttclagacct cctgatgaaa taaagcagtc aggtgggtctt 120
atgccaagag gacagagtga gtactttgac cgaggtactc aaatgaatat caacctttat 180
gatcatgcaa gaggaactca gacgggattt gttaggcacg atgatggata tgtttccacc 240
tcaattagtt tgagaagtgc ccacttagtg ggtcaaacta tattgtctgg tcattctact 300
tattatatat atgttatagc cactgcaccc aacatgttta acgttaatga tgtattaggg 360
gcatacagtc ctcatccaga tgaacaagaa gtttctgctt taggtgggat tccatactcc 420
caaatatatg gatggtatcg agttcatittt ggggtgcttg atgaacaatt acatcgtaat 480
aggggctaca gagatagata ttacagtaac ttagatattg ctccagcagc agatggttat 540
ggattggcag gtttccctcc ggagcataga gcttggaggg aagagccgtg gattcatcat 600

[0049]

```

gcaccgccgg gttgtgggaa tgctccaaga tcatcgatga gtaatacttg cgatgaaaaa    660
acccaaagtc taggtgtaaa attccttgac gaataccaat ctaaagttaa aagacaaata    720
ttttcaggct atcaatctga tattgataca cataatagaa ttaaggatga attatgatta    780
aattaaaatt tgggtgtttt ttacagttt tactatcttc agcatatgca catggaacac    840
ctcaaaatat tactgatttg tgtgcagaat accacaacac acaaatatat acgctaaatg    900
ataagatatt ttcgtataca gaatctctag ctggaaaaag agagatggct atcattactt    960
ttaagaatgg tgcaattttt caagtagaag taccaggtag tcaacatata gattcacaaa   1020
aaaaagcgat tgaaaggatg aaggataccc tgaggattgc atatcttact gaagctaaag   1080
tcgaaaagtt atgtgtatgg aataataaaa cgcctcatgc gattgccgca attagtatgg   1140
caaattaa                                         1148

```

<210> 41
 <211> 269
 <212> PRT
 <213> 百日咳博德特氏菌

```

<400> 41
Met Arg Cys Thr Arg Ala Ile Arg Gln Thr Ala Arg Thr Gly Trp Leu
1          5          10          15

Thr Trp Leu Ala Ile Leu Ala Val Thr Ala Pro Val Thr Ser Pro Ala
20          25          30

Trp Ala Asp Asp Pro Pro Ala Thr Val Tyr Arg Tyr Asp Ser Arg Pro
35          40          45

Pro Glu Asp Val Phe Gln Asn Gly Phe Thr Ala Trp Gly Asn Asn Asp
50          55          60

Asn Val Leu Asp His Leu Thr Gly Arg Ser Cys Gln Val Gly Ser Ser
65          70          75          80

Asn Ser Ala Phe Val Ser Thr Ser Ser Ser Arg Arg Tyr Thr Glu Val
85          90          95

Tyr Leu Glu His Arg Met Gln Glu Ala Val Glu Ala Glu Arg Ala Gly
100         105         110

Arg Gly Thr Gly His Phe Ile Gly Tyr Ile Tyr Glu Val Arg Ala Asp
115         120         125

Asn Asn Phe Tyr Gly Ala Ala Ser Ser Tyr Phe Glu Tyr Val Asp Thr
130         135         140

Tyr Gly Asp Asn Ala Gly Arg Ile Leu Ala Gly Ala Leu Ala Thr Tyr
145         150         155         160

Gln Ser Glu Tyr Leu Ala His Arg Arg Ile Pro Pro Glu Asn Ile Arg
165         170         175

Arg Val Thr Arg Val Tyr His Asn Gly Ile Thr Gly Glu Thr Thr Thr
180         185         190

Thr Glu Tyr Ser Asn Ala Arg Tyr Val Ser Gln Gln Thr Arg Ala Asn

```

[0050]

195	200	205
Pro Asn Pro Tyr Thr Ser Arg Arg Ser Val Ala Ser Ile Val Gly Thr 210 215 220		
Leu Val Arg Met Ala Pro Val Ile Gly Ala Cys Met Ala Arg Gln Ala 225 230 235 240		
Glu Ser Ser Glu Ala Met Ala Ala Trp Ser Glu Arg Ala Gly Glu Ala 245 250 255		
Met Val Leu Val Tyr Tyr Glu Ser Ile Ala Tyr Ser Phe 260 265		
<210> 42		
<211> 226		
<212> PRT		
<213> 百日咳博德特氏菌		
<400> 42		
Met Pro Ile Asp Arg Lys Thr Leu Cys His Leu Leu Ser Val Leu Pro 1 5 10 15		
Leu Ala Leu Leu Gly Ser His Val Ala Arg Ala Ser Thr Pro Gly Ile 20 25 30		
Val Ile Pro Pro Gln Glu Gln Ile Thr Gln His Gly Ser Pro Tyr Gly 35 40 45		
Arg Cys Ala Asn Lys Thr Arg Ala Leu Thr Val Ala Glu Leu Arg Gly 50 55 60		
Ser Gly Asp Leu Gln Glu Tyr Leu Arg His Val Thr Arg Gly Trp Ser 65 70 75 80		
Ile Phe Ala Leu Tyr Asp Gly Thr Tyr Leu Gly Gly Glu Tyr Gly Gly 85 90 95		
Val Ile Lys Asp Gly Thr Pro Gly Gly Ala Phe Asp Leu Lys Thr Thr 100 105 110		
Phe Cys Ile Met Thr Thr Arg Asn Thr Gly Gln Pro Ala Thr Asp His 115 120 125		
Tyr Tyr Ser Asn Val Thr Ala Thr Arg Leu Leu Ser Ser Thr Asn Ser 130 135 140		
Arg Leu Cys Ala Val Phe Val Arg Ser Gly Gln Pro Val Ile Gly Ala 145 150 155 160		
Cys Thr Ser Pro Tyr Asp Gly Lys Tyr Trp Ser Met Tyr Ser Arg Leu 165 170 175		
Arg Lys Met Leu Tyr Leu Ile Tyr Val Ala Gly Ile Ser Val Arg Val 180 185 190		
His Val Ser Lys Glu Glu Gln Tyr Tyr Asp Tyr Glu Asp Ala Thr Phe 195 200 205		

[0051]

Glu Thr Tyr Ala Leu Thr Gly Ile Ser Ile Cys Asn Pro Gly Ser Ser
210 215 220

Leu Cys
225

<210> 43
<211> 131
<212> PRT
<213> 百日咳博德特氏菌

<400> 43
Val Arg Ala Leu Ala Trp Leu Leu Ala Ser Gly Ala Met Thr His Leu
1 5 10 15

Ser Pro Ala Leu Ala Asp Val Pro Tyr Val Leu Val Lys Thr Asn Met
20 25 30

Val Val Thr Ser Val Ala Met Lys Pro Tyr Glu Val Thr Pro Thr Arg
35 40 45

Met Leu Val Cys Gly Ile Ala Ala Lys Leu Gly Ala Ala Ala Ser Ser
50 55 60

Pro Asp Ala His Val Pro Phe Cys Phe Gly Lys Asp Leu Lys Arg Pro
65 70 75 80

Gly Ser Ser Pro Met Glu Val Met Leu Arg Ala Val Phe Met Gln Gln
85 90 95

Arg Pro Leu Arg Met Phe Leu Gly Pro Lys Gln Leu Thr Phe Glu Gly
100 105 110

Lys Pro Ala Leu Glu Leu Ile Arg Met Val Glu Cys Ser Gly Lys Gln
115 120 125

Asp Cys Pro
130

<210> 44
<211> 120
<212> PRT
<213> 百日咳博德特氏菌

<400> 44
Met His Thr Ile Ala Ser Ile Leu Leu Ser Val Leu Gly Ile Tyr Ser
1 5 10 15

Pro Ala Asp Val Ala Gly Leu Pro Thr His Leu Tyr Lys Asn Phe Thr
20 25 30

Val Gln Glu Leu Ala Leu Lys Leu Lys Gly Lys Asn Gln Glu Phe Cys
35 40 45

Leu Thr Ala Phe Met Ser Gly Arg Ser Leu Val Arg Ala Cys Leu Ser
50 55 60

Asp Ala Gly His Glu His Asp Thr Trp Phe Asp Thr Met Leu Gly Phe
65 70 75 80

[0052]

Ala Ile Ser Ala Tyr Ala Leu Lys Ser Arg Ile Ala Leu Thr Val Glu
85 90 95

Asp Ser Pro Tyr Pro Gly Thr Pro Gly Asp Leu Leu Glu Leu Gln Ile
100 105 110

Cys Pro Leu Asn Gly Tyr Cys Glu
115 120

<210> 45
<211> 227
<212> PRT
<213> 百日咳博德特氏菌

<400> 45
Met Leu Ile Asn Asn Lys Lys Leu Leu His His Ile Leu Pro Ile Leu
1 5 10 15

Val Leu Ala Leu Leu Gly Met Arg Thr Ala Gln Ala Val Ala Pro Gly
20 25 30

Ile Val Ile Pro Pro Lys Ala Leu Phe Thr Gln Gln Gly Gly Ala Tyr
35 40 45

Gly Arg Cys Pro Asn Gly Thr Arg Ala Leu Thr Val Ala Glu Leu Arg
50 55 60

Gly Asn Ala Glu Leu Gln Thr Tyr Leu Arg Gln Ile Thr Pro Gly Trp
65 70 75 80

Ser Ile Tyr Gly Leu Tyr Asp Gly Thr Tyr Leu Gly Gln Ala Tyr Gly
85 90 95

Gly Ile Ile Lys Asp Ala Pro Pro Gly Ala Gly Phe Ile Tyr Arg Glu
100 105 110

Thr Phe Cys Ile Thr Thr Ile Tyr Lys Thr Gly Gln Pro Ala Ala Asp
115 120 125

His Tyr Tyr Ser Lys Val Thr Ala Thr Arg Leu Leu Ala Ser Thr Asn
130 135 140

Ser Arg Leu Cys Ala Val Phe Val Arg Asp Gly Gln Ser Val Ile Gly
145 150 155 160

Ala Cys Ala Ser Pro Tyr Glu Gly Arg Tyr Arg Asp Met Tyr Asp Ala
165 170 175

Leu Arg Arg Leu Leu Tyr Met Ile Tyr Met Ser Gly Leu Ala Val Arg
180 185 190

Val His Val Ser Lys Glu Glu Gln Tyr Tyr Asp Tyr Glu Asp Ala Thr
195 200 205

Phe Gln Thr Tyr Ala Leu Thr Gly Ile Ser Leu Cys Asn Pro Ala Ala
210 215 220

Ser Ile Cys

[0053]

225

<210> 46
<211> 6
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成6xHis标签

<400> 46
His His His His His His
1 5

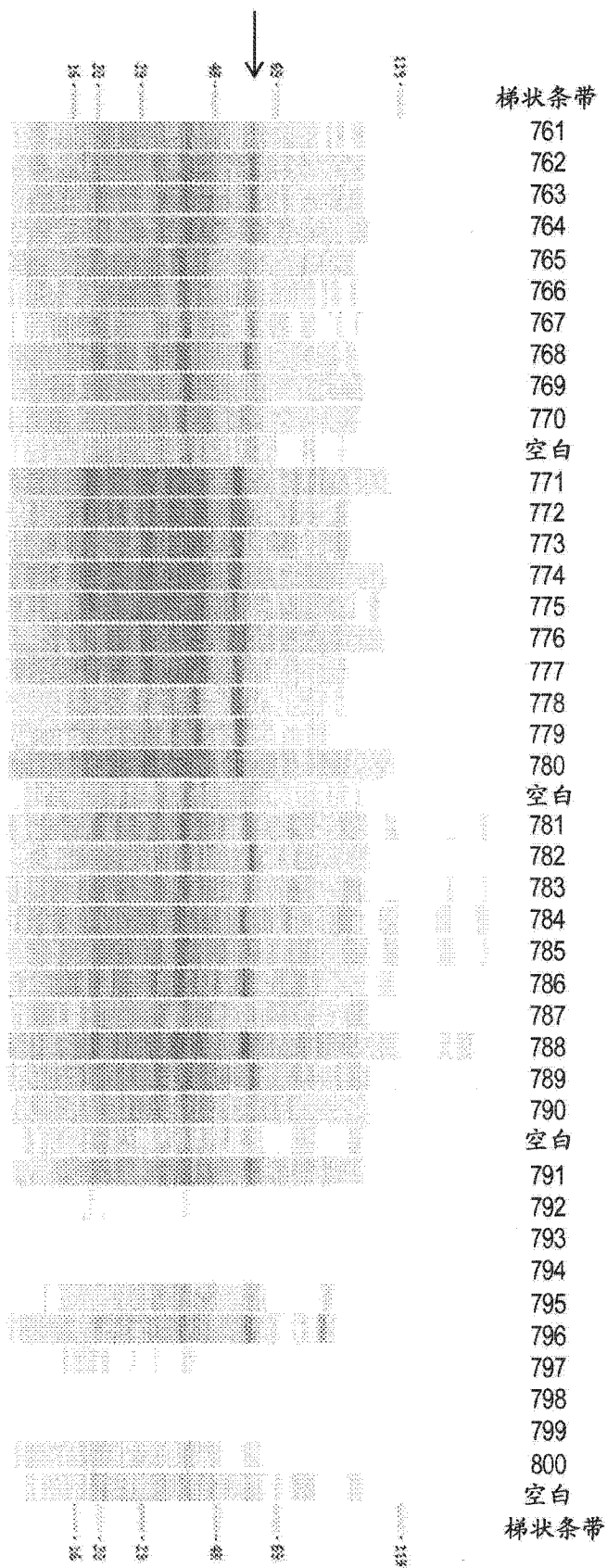


图 1

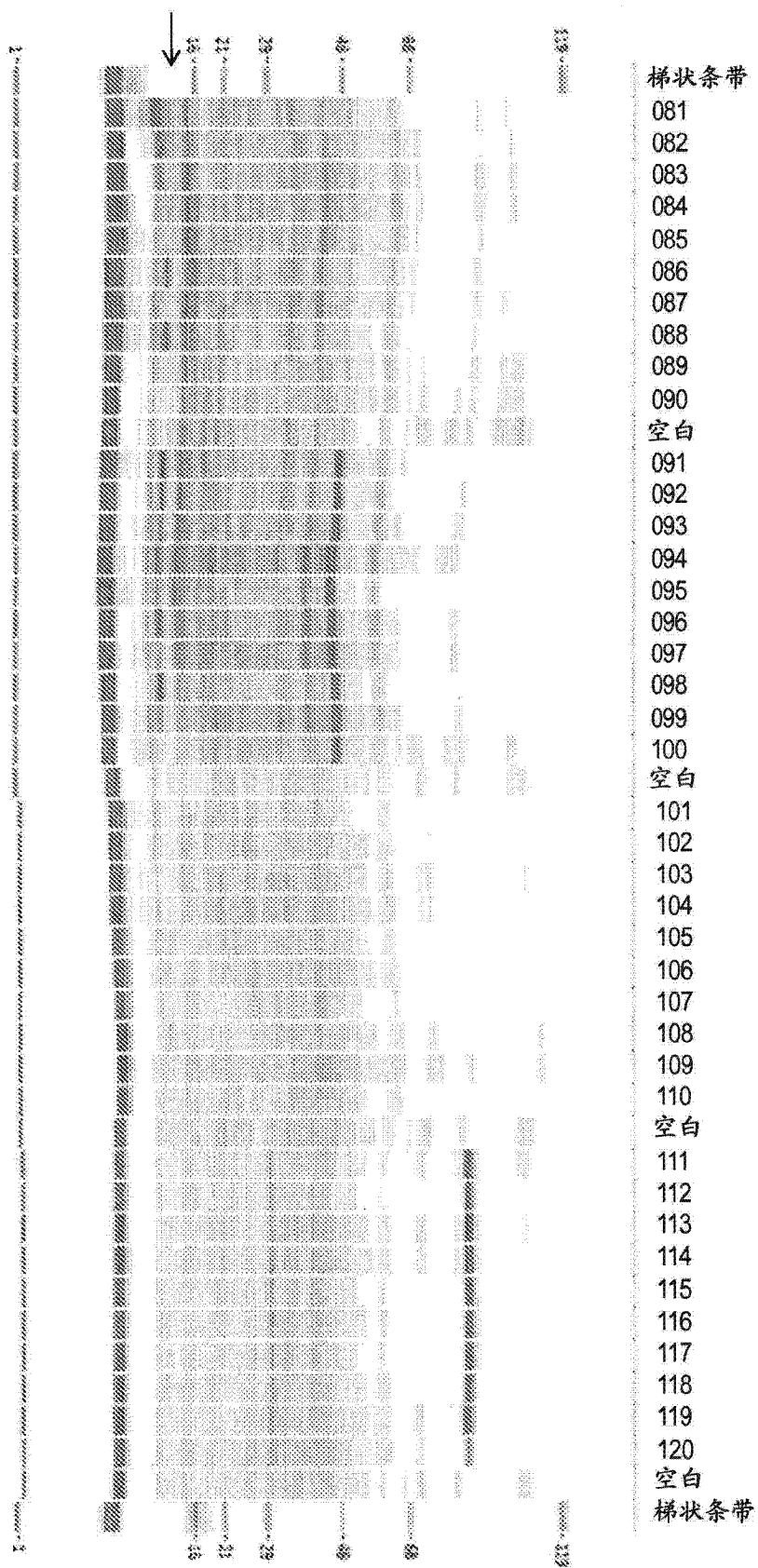


图 2

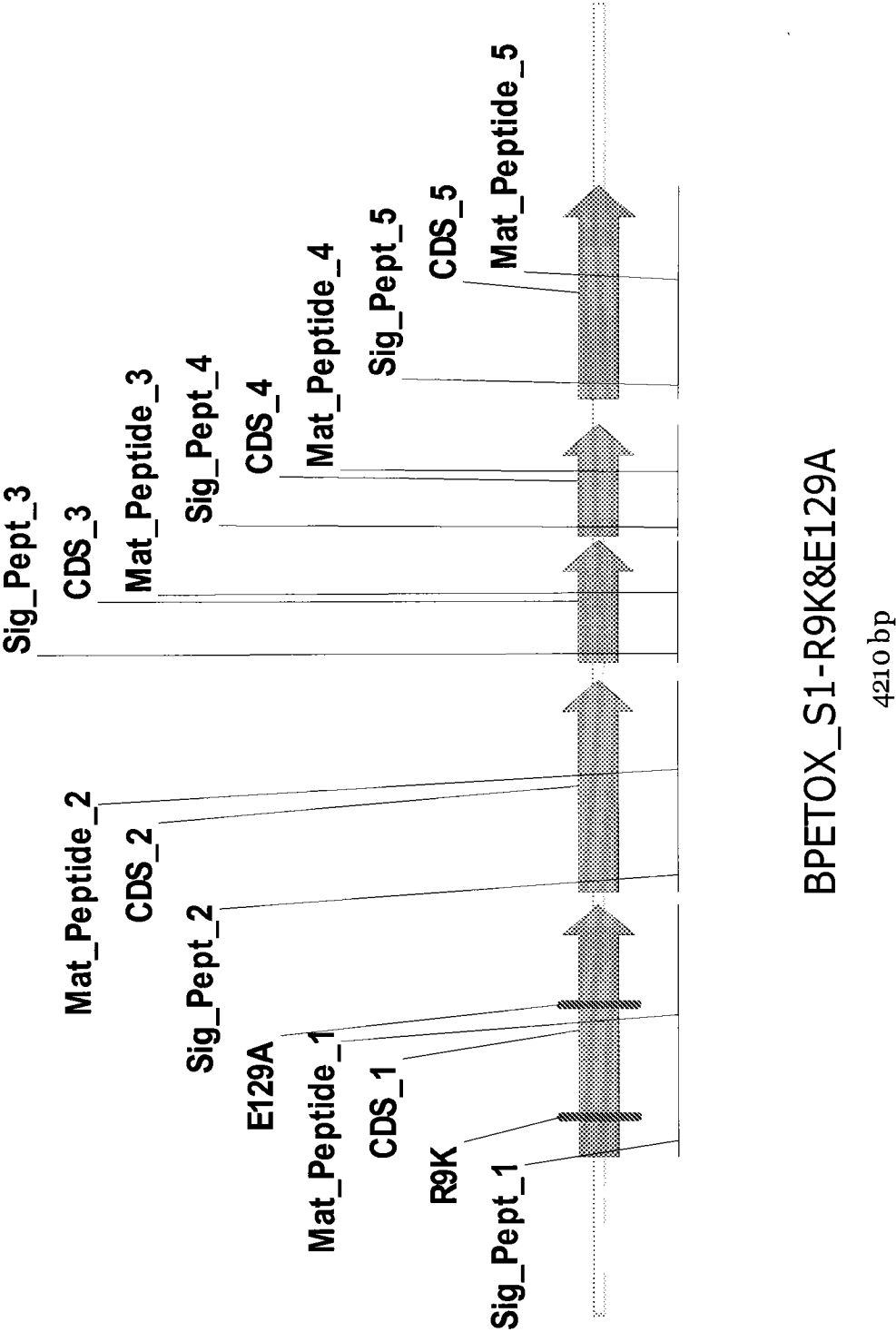


图 3

信号序列

~~~~~

S1

~~~~~

1 Met Arg Cys Thr Arg Ala Ile Arg Gln Thr Ala Arg Thr Gly Trp Leu Thr Trp Leu Ala
 ATG CGT TGC ACT CGG GCA ATT CGC CAA ACC GCA AGA ACA GGC TGG CTG ACG TGG CTG CGC

信号序列

~~~~~

S1

~~~~~

61 Ile Leu Ala Val Thr Ala Pro Val Thr Ser Pro Ala Trp Ala Asp Asp Pro Pro Ala Thr
 ATT CTT GCC GTC ACG GCG CCC GTG ACT TCG CCG GCA TGG GCC GAC GAT CCT CCC GCC ACC

S1

~~~~~

121 Val Tyr Lys Tyr Asp Ser Arg Pro Pro Glu Asp Val Phe Gln Asn Gly Phe Thr Ala Trp  
 GTA TAC AAA TAT GAC TCC CGC CCG CCG GAG GAC GTT TTC CAG AAC GGA TTC ACG GCG TGG

S1

~~~~~

181 Gly Asn Asn Asp Asn Val Leu Asp His Leu Thr Gly Arg Ser Cys Gln Val Gly Ser Ser
 GGA AAC AAC GAC AAT GTG CTC GAC CAT CTG ACC GGA CGT TCC TGC CAG GTC GGC AGC AGC

S1

~~~~~

241 Asn Ser Ala Phe Val Ser Thr Ser Ser Ser Arg Arg Tyr Thr Glu Val Tyr Leu Glu His  
 AAC AGC GCT TTC GTC TCC ACC AGC AGC AGC CGG CGC TAT ACC GAG GTC TAT CTC GAA CAT

S1

~~~~~

301 Arg Met Gln Glu Ala Val Glu Ala Glu Arg Ala Gly Arg Gly Thr Gly His Phe Ile Gly
 CGC ATG CAG GAA GCG GTC GAG GCC GAA CGC GCC GGC AGG GGC ACC GGC CAC TTC ATC GGC

S1

~~~~~

361 Tyr Ile Tyr Glu Val Arg Ala Asp Asn Asn Phe Tyr Gly Ala Ala Ser Ser Tyr Phe Glu  
 TAC ATC TAC GAA GTC CGC GCC GAC AAC AAT TTC TAC GGC GCC GCC AGC TCG TAC TTC GAA

S1

~~~~~

421 Tyr Val Asp Thr Tyr Gly Asp Asn Ala Gly Arg Ile Leu Ala Gly Ala Leu Ala Thr Tyr
 TAC GTC GAC ACT TAT GGC GAC AAT GCC GGC CGT ATC CTC GCC GGC GCG CTG GCC ACC TAC

S1

~~~~~

481 Gln Ser Ala Tyr Leu Ala His Arg Arg Ile Pro Pro Glu Asn Ile Arg Arg Val Thr Arg  
 CAG AGC GCA TAT CTG GCA CAC CGG CGC ATT CCG CCC GAA AAC ATC CGC AGG GTA ACG CGG

S1

~~~~~

541 Val Tyr His Asn Gly Ile Thr Gly Glu Thr Thr Thr Thr Glu Tyr Ser Asn Ala Arg Tyr
 GTC TAT CAC AAC GGC ATC ACC GGC GAG ACC ACG ACC ACG GAG TAT TCC AAC GCT CGC TAC

图 4-1

```

S1
~~~~~
601 Val Ser Gln Gln Thr Arg Ala Asn Pro Asn Pro Tyr Thr Ser Arg Arg Ser Val Ala Ser
    GTC AGC CAG CAG ACT CGC GCC AAT CCC AAC CCC TAC ACA TCG CGA AGG TCC GTA GCG TCG
    S1
~~~~~
661 Ile Val Gly Thr Leu Val Arg Met Ala Pro Val Ile Gly Ala Cys Met Ala Arg Gln Ala
    ATC GTC GGC ACA TTG GTG CGC ATG GCG CCG GTG ATA GGC GCT TGC ATG GCG CGG CAG GCC
    S1
~~~~~
721 Glu Ser Ser Glu Ala Met Ala Ala Trp Ser Glu Arg Ala Gly Glu Ala Met Val Leu Val
    GAA AGC TCC GAG GCC ATG GCA GCC TGG TCC GAA CGC GCC GGC GAG GCG ATG GTT CTC GTG
    S1
~~~~~
781 Tyr Tyr Glu Ser Ile Ala Tyr Ser Phe
    TAC TAC GAA AGC ATC GCG TAT TCG TTC TAG ACC TGG CCC AGC CCC GCC CAA CTC CGG TAA
    信号序列
~~~~~
S2
~~~~~
841 Met Pro Ile Asp Arg Lys Thr Leu Cys His Leu Leu Ser Val Leu Pro Leu
    TTG AAC AGC ATG CCG ATC GAC CGC AAG ACG CTC TGC CAT CTC CTG TCC GTT CTG CCG TTG
    信号序列
~~~~~
S2
~~~~~
901 Ala Leu Leu Gly Ser His Val Ala Arg Ala Ser Thr Pro Gly Ile Val Ile Pro Pro Gln
    GCC CTC CTC GGA TCT CAC GTG GCG CGG GCC TCC ACG CCA GGC ATC GTC ATT CCG CCG CAG
    S2
~~~~~
961 Glu Gln Ile Thr Gln His Gly Ser Pro Tyr Gly Arg Cys Ala Asn Lys Thr Arg Ala Leu
    GAA CAG ATT ACC CAG CAT GGC AGC CCC TAT GGA CGC TGC GCG AAC AAG ACC CGT GCC CTG
    S2
~~~~~
1021 Thr Val Ala Glu Leu Arg Gly Ser Gly Asp Leu Gln Glu Tyr Leu Arg His Val Thr Arg
    ACC GTG GCG GAA TTG CGC GGC AGC GGC GAT CTG CAG GAG TAC CTG CGT CAT GTG ACG CGC
    S2
~~~~~
1081 Gly Trp Ser Ile Phe Ala Leu Tyr Asp Gly Thr Tyr Leu Gly Gly Glu Tyr Gly Gly Val
    GGC TGG TCA ATA TTT GCG CTC TAC GAT GGC ACC TAT CTC GGC GGC GAA TAT GGC GGC GTG
    S2
~~~~~
1141 Ile Lys Asp Gly Thr Pro Gly Gly Ala Phe Asp Leu Lys Thr Thr Phe Cys Ile Met Thr
    ATC AAG GAC GGA ACA CCC GGC GGC GCA TTC GAC CTG AAA ACG ACG TTC TGC ATC ATG ACC
    S2
~~~~~
1201 Thr Arg Asn Thr Gly Gln Pro Ala Thr Asp His Tyr Tyr Ser Asn Val Thr Ala Thr Arg
    ACG CGC AAT ACG GGT CAA CCC GCA ACG GAT CAC TAC TAC AGC AAC GTC ACC GCC ACT CGC

```

图 4-2

图 4-3

图 4-4

信号序列

```

~~~~~
                                S3
~~~~~
Val Ala Pro Gly Ile Val Ile Pro Pro Lys Ala Leu Phe Thr Gln Gln Gly Gly Ala Tyr·
2521 CC GTT GCG CCA GGC ATC GTC ATC CCG CCG AAG GCA CTG TTC ACC CAA CAG GGC GGC GCC T
                                S3
~~~~~
· Gly Arg Cys Pro Asn Gly Thr Arg Ala Leu Thr Val Ala Glu Leu Arg Gly Asn Ala Glu·
2581 AT GGA CGC TGC CCG AAC GGA ACC CGC GCC TTG ACC GTG GCC GAA CTG CGC GGC AAC GCC G
                                S3
~~~~~
Leu Gln Thr Tyr Leu Arg Gln Ile Thr Pro Gly Trp Ser Ile Tyr Gly Leu Tyr Asp Gly·
2641 AA TTG CAG ACG TAT TTG CGC CAG ATA ACG CCC GGC TGG TCC ATA TAC GGT CTC TAT GAC G
                                S3
~~~~~
Thr Tyr Leu Gly Gln Ala Tyr Gly Gly Ile Ile Lys Asp Ala Pro Pro Gly Ala Gly Phe·
2701 GT ACG TAC CTG GGC CAG GCG TAC GGC GGC ATC ATC AAG GAC GCG CCG CCA GGC GCG GGG T
                                S3
~~~~~
Ile Tyr Arg Glu Thr Phe Cys Ile Thr Thr Ile Tyr Lys Thr Gly Gln Pro Ala Ala Asp·
2761 TC ATT TAT CGC GAA ACT TTC TGC ATC ACG ACC ATA TAC AAG ACC GGG CAA CCG GCT GCG G
                                S3
~~~~~
· His Tyr Tyr Ser Lys Val Thr Ala Thr Arg Leu Leu Ala Ser Thr Asn Ser Arg Leu Cys·
2821 AT CAC TAC TAC AGC AAG GTC ACG GCC ACG CGC CTG CTC GCC AGC ACC AAC AGC AGG CTG T
                                S3
~~~~~
Ala Val Phe Val Arg Asp Gly Gln Ser Val Ile Gly Ala Cys Ala Ser Pro Tyr Glu Gly·
2881 GC GCG GTA TTC GTC AGG GAC GGG CAA TCG GTC ATC GGA GCC TGC GCC AGC CCG TAT GAA G
                                S3
~~~~~
Arg Tyr Arg Asp Met Tyr Asp Ala Leu Arg Arg Leu Leu Tyr Met Ile Tyr Met Ser Gly·
2941 GC AGG TAC AGA GAC ATG TAC GAC GCG CTG CGG CGC CTG CTG TAC ATG ATC TAT ATG TCC G
                                S3
~~~~~
· Leu Ala Val Arg Val His Val Ser Lys Glu Glu Gln Tyr Tyr Asp Tyr Glu Asp Ala Thr·
3001 GC CTT GCC GTA CGC GTC CAC GTC AGC AAG GAG GAA CAG TAT TAC GAC TAC GAG GAC GCC A
                                S3
~~~~~
Phe Gln Thr Tyr Ala Leu Thr Gly Ile Ser Leu Cys Asn Pro Ala Ala Ser Ile Cys
3061 CA TTC CAG ACC TAT GCC CTC ACC GGC ATT TCC CTC TGC AAC CCG GCA GCG TCG ATA TGC

```

图 4-5

1 Met Arg Cys Thr Arg Ala Ile Arg Gln Thr Ala Arg Thr Gly Trp
Leu Thr Trp Leu Ala

21 Ile Leu Ala Val Thr Ala Pro Val Thr Ser Pro Ala Trp Ala Asp
Asp Pro Pro Ala Thr

41 Val Tyr Lys Tyr Asp Ser Arg Pro Pro Glu Asp Val Phe Gln Asn
Gly Phe Thr Ala Trp

61 Gly Asn Asn Asp Asn Val Leu Asp His Leu Thr Gly Arg Ser Cys
Gln Val Gly Ser Ser

81 Asn Ser Ala Phe Val Ser Thr Ser Ser Ser Arg Arg Tyr Thr Glu
Val Tyr Leu Glu His

101 Arg Met Gln Glu Ala Val Glu Ala Glu Arg Ala Gly Arg Gly Thr
Gly His Phe Ile Gly

121 Tyr Ile Tyr Glu Val Arg Ala Asp Asn Asn Phe Tyr Gly Ala Ala
Ser Ser Tyr Phe Glu

141 Tyr Val Asp Thr Tyr Gly Asp Asn Ala Gly Arg Ile Leu Ala Gly
Ala Leu Ala Thr Tyr

161 Gln Ser Ala Tyr Leu Ala His Arg Arg Ile Pro Pro Glu Asn Ile
Arg Arg Val Thr Arg

181 Val Tyr His Asn Gly Ile Thr Gly Glu Thr Thr Thr Thr Glu Tyr
Ser Asn Ala Arg Tyr

201 Val Ser Gln Gln Thr Arg Ala Asn Pro Asn Pro Tyr Thr Ser Arg
Arg Ser Val Ala Ser

221 Ile Val Gly Thr Leu Val Arg Met Ala Pro Val Ile Gly Ala Cys
Met Ala Arg Gln Ala

241 Glu Ser Ser Glu Ala Met Ala Ala Trp Ser Glu Arg Ala Gly Glu
Ala Met Val Leu Val

261 Tyr Tyr Glu Ser Ile Ala Tyr Ser Phe

图 5-1

1 Met Pro Ile Asp Arg Lys Thr Leu Cys His Leu Leu Ser Val Leu
Pro Leu Ala Leu Leu

21 Gly Ser His Val Ala Arg Ala Ser Thr Pro Gly Ile Val Ile Pro
Pro Gln Glu Gln Ile

41 Thr Gln His Gly Ser Pro Tyr Gly Arg Cys Ala Asn Lys Thr Arg
Ala Leu Thr Val Ala

61 Glu Leu Arg Gly Ser Gly Asp Leu Gln Glu Tyr Leu Arg His Val
Thr Arg Gly Trp Ser

81 Ile Phe Ala Leu Tyr Asp Gly Thr Tyr Leu Gly Gly Glu Tyr Gly
Gly Val Ile Lys Asp

101 Gly Thr Pro Gly Gly Ala Phe Asp Leu Lys Thr Thr Phe Cys Ile
Met Thr Thr Arg Asn

121 Thr Gly Gln Pro Ala Thr Asp His Tyr Tyr Ser Asn Val Thr Ala
Thr Arg Leu Leu Ser

141 Ser Thr Asn Ser Arg Leu Cys Ala Val Phe Val Arg Ser Gly Gln
Pro Val Ile Gly Ala

161 Cys Thr Ser Pro Tyr Asp Gly Lys Tyr Trp Ser Met Tyr Ser Arg
Leu Arg Lys Met Leu

181 Tyr Leu Ile Tyr Val Ala Gly Ile Ser Val Arg Val His Val Ser
Lys Glu Glu Gln Tyr

201 Tyr Asp Tyr Glu Asp Ala Thr Phe Glu Thr Tyr Ala Leu Thr Gly
Ile Ser Ile Cys Asn

221 Pro Gly Ser Ser Leu Cys

图 5-2

1 Met Leu Ile Asn Asn Lys Lys Leu Leu His His Ile Leu Pro
Ile Leu Val Leu Ala Leu

21 Leu Gly Met Arg Thr Ala Gln Ala Val Ala Pro Gly Ile Val
Ile Pro Pro Lys Ala Leu

41 Phe Thr Gln Gln Gly Gly Ala Tyr Gly Arg Cys Pro Asn Gly
Thr Arg Ala Leu Thr Val

61 Ala Glu Leu Arg Gly Asn Ala Glu Leu Gln Thr Tyr Leu Arg
Gln Ile Thr Pro Gly Trp

81 Ser Ile Tyr Gly Leu Tyr Asp Gly Thr Tyr Leu Gly Gln Ala
Tyr Gly Gly Ile Ile Lys

101 Asp Ala Pro Pro Gly Ala Gly Phe Ile Tyr Arg Glu Thr Phe
Cys Ile Thr Thr Ile Tyr

121 Lys Thr Gly Gln Pro Ala Ala Asp His Tyr Tyr Ser Lys Val
Thr Ala Thr Arg Leu Leu

141 Ala Ser Thr Asn Ser Arg Leu Cys Ala Val Phe Val Arg Asp
Gly Gln Ser Val Ile Gly

161 Ala Cys Ala Ser Pro Tyr Glu Gly Arg Tyr Arg Asp Met Tyr
Asp Ala Leu Arg Arg Leu

181 Leu Tyr Met Ile Tyr Met Ser Gly Leu Ala Val Arg Val His
Val Ser Lys Glu Glu Gln

201 Tyr Tyr Asp Tyr Glu Asp Ala Thr Phe Gln Thr Tyr Ala Leu
Thr Gly Ile Ser Leu Cys

221 Asn Pro Ala Ala Ser Ile Cys

图 5-3

1 Val Arg Ala Leu Ala Trp Leu Leu Ala Ser Gly Ala Met Thr
His Leu Ser Pro Ala Leu

21 Ala Asp Val Pro Tyr Val Leu Val Lys Thr Asn Met·Val Val
Thr Ser Val Ala Met Lys

41 Pro Tyr Glu Val Thr Pro Thr Arg Met Leu Val Cys Gly Ile
Ala Ala Lys Leu Gly Ala

61 Ala Ala Ser Ser Pro Asp Ala His Val Pro Phe Cys Phe Gly
Lys Asp Leu Lys Arg Pro

81 Gly Ser Ser Pro Met Glu Val Met Leu Arg Ala Val Phe Met
Gln Gln Arg Pro Leu Arg

101 Met Phe Leu Gly Pro Lys Gln Leu Thr Phe Glu Gly Lys Pro
Ala Leu Glu Leu Ile Arg

121 Met Val Glu Cys Ser Gly Lys Gln Asp Cys Pro

图 5-4

1 Met His Thr Ile Ala Ser Ile Leu Leu Ser Val Leu Gly Ile
Tyr Ser Pro Ala Asp Val

21 Ala Gly Leu Pro Thr His Leu Tyr Lys Asn Phe Thr Val Gln
Glu Leu Ala Leu Lys Leu

41 Lys Gly Lys Asn Gln Glu Phe Cys Leu Thr Ala Phe Met Ser
Gly Arg Ser Leu Val Arg

61 Ala Cys Leu Ser Asp Ala Gly His Glu His Asp Thr Trp Phe
Asp Thr Met Leu Gly Phe

81 Ala Ile Ser Ala Tyr Ala Leu Lys Ser Arg Ile Ala Leu Thr
Val Glu Asp Ser Pro Tyr

101 Pro Gly Thr Pro Gly Asp Leu Leu Glu Leu Gln Ile Cys Pro
Leu Asn Gly Tyr Cys Glu

图 5-5

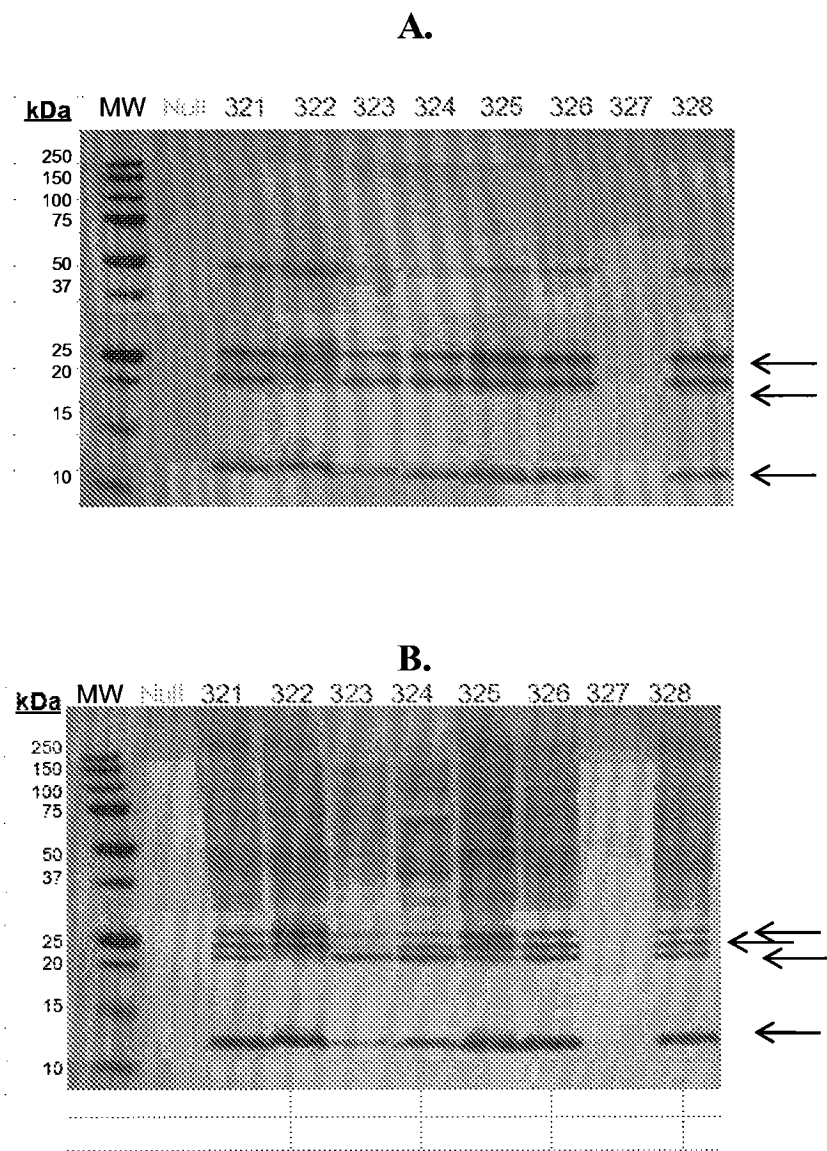


图 6

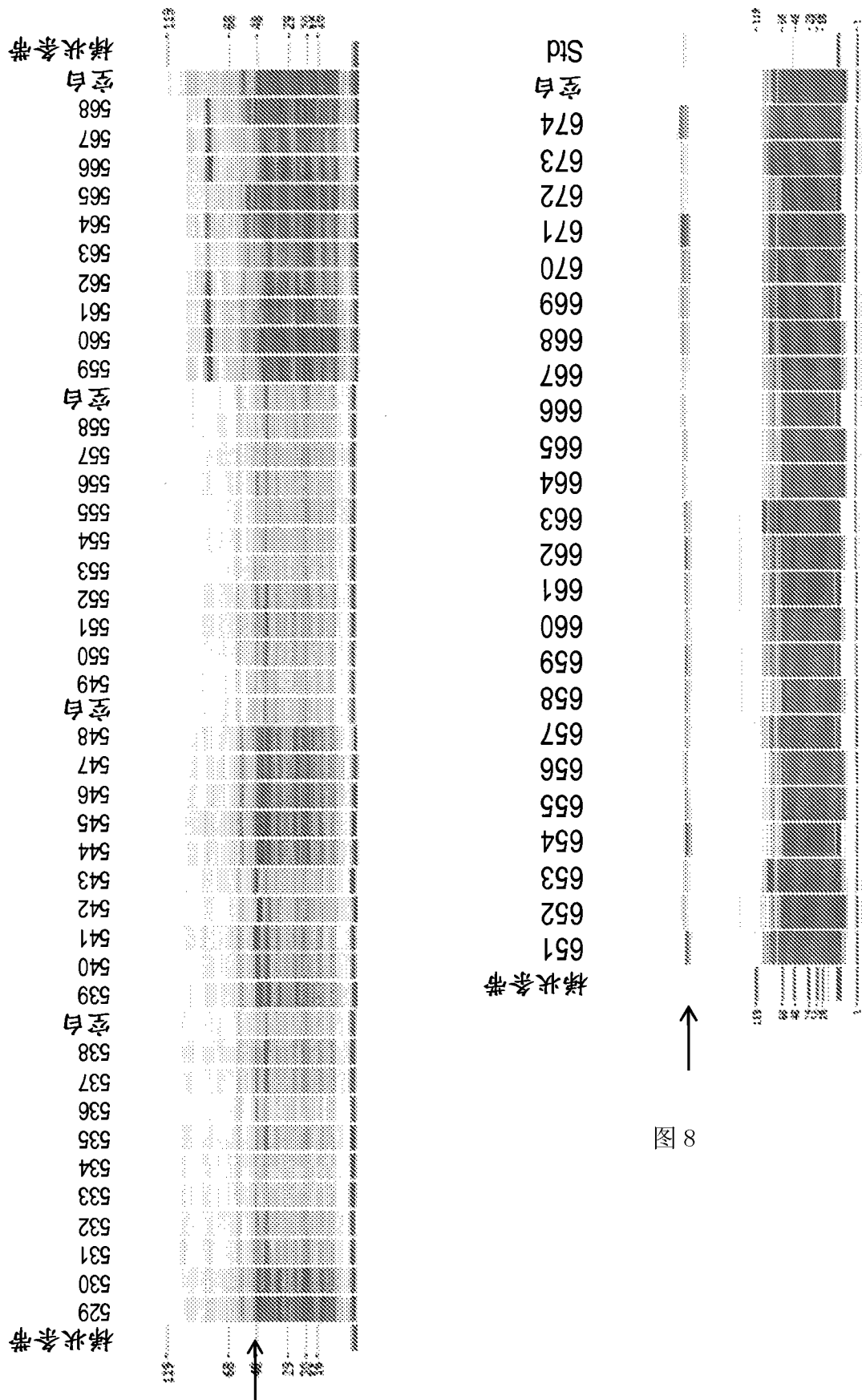


图 8

图 7

假单胞菌外毒素 A 氨基酸序列

1 Ala Glu Glu Ala Phe Asp Leu Trp Asn Glu Cys Ala Lys Ala Cys Val Leu
Asp Leu Lys

21 Asp Gly Val Arg Ser Ser Arg Met Ser Val Asp Pro Ala Ile Ala Asp Thr
Asn Gly Gln

41 Gly Val Leu His Tyr Ser Met Val Leu Glu Gly Gly Asn Asp Ala Leu Lys
Leu Ala Ile

61 Asp Asn Ala Leu Ser Ile Thr Ser Asp Gly Leu Thr Ile Arg Leu Glu Gly
Gly Val Glu

81 Pro Asn Lys Pro Val Arg Tyr Ser Tyr Thr Arg Gln Ala Arg Gly Ser Trp
Ser Leu Asn

101 Trp Leu Val Pro Ile Gly His Glu Lys Pro Ser Asn Ile Lys Val Phe Ile
His Glu Leu

121 Asn Ala Gly Asn Gln Leu Ser His Met Ser Pro Ile Tyr Thr Ile Glu Met
Gly Asp Glu

141 Leu Leu Ala Lys Leu Ala Arg Asp Ala Thr Phe Phe Val Arg Ala His Glu
Ser Asn Glu

161 Met Gln Pro Thr Leu Ala Ile Ser His Ala Gly Val Ser Val Val Met Ala
Gln Ala Gln

181 Pro Arg Arg Glu Lys Arg Trp Ser Glu Trp Ala Ser Gly Lys Val Leu Cys
Leu Leu Asp

201 Pro Leu Asp Gly Val Tyr Asn Tyr Leu Ala Gln Gln Arg Cys Asn Leu Asp
Asp Thr Trp

221 Glu Gly Lys Ile Tyr Arg Val Leu Ala Gly Asn Pro Ala Lys His Asp Leu
Asp Ile Lys

241 Pro Thr Val Ile Ser His Arg Leu His Phe Pro Glu Gly Gly Ser Leu Ala
Ala Leu Thr

261 Ala His Gln Ala Cys His Leu Pro Leu Glu Thr Phe Thr Arg His Arg Gln
Pro Arg Gly

281 Trp Glu Gln Leu Glu Gln Cys Gly Tyr Pro Val Gln Arg Leu Val Ala Leu
Tyr Leu Ala

301 Ala Arg Leu Ser Trp Asn Gln Val Asp Gln Val Ile Arg Asn Ala Leu Ala
Ser Pro Gly

321 Ser Gly Gly Asp Leu Gly Glu Ala Ile Arg Glu Gln Pro Glu Gln Ala Arg
Leu Ala Leu

341 Thr Leu Ala Ala Ala Glu Ser Glu Arg Phe Val Arg Gln Gly Thr Gly Asn
Asp Glu Ala

361 Gly Ala Ala Ser Ala Asp Val Val Ser Leu Thr Cys Pro Val Ala Ala Gly
Glu Cys Ala

图 9-1

381 Gly Pro Ala Asp Ser Gly Asp Ala Leu Leu Glu Arg Asn Tyr Pro Thr Gly Ala
Glu Phe

401 Leu Gly Asp Gly Gly Asp Ile Ser Phe Ser Thr Arg Gly Thr Gln Asn Trp Thr
Val Glu

Tyr - CRM66

421 Arg Leu Leu Gln Ala His Arg Gln Leu Glu Glu Arg Gly Tyr Val Phe Val Gly
Tyr His

441 Gly Thr Phe Leu Glu Ala Ala Gln Ser Ile Val Phe Gly Gly Val Arg Ala Arg
Ser Gln

461 Asp Leu Asp Ala Ile Trp Arg Gly Phe Tyr Ile Ala Gly Asp Pro Ala Leu Ala
Tyr Gly

481 Tyr Ala Gln Asp Gln Glu Pro Asp Ala Arg Gly Arg Ile Arg Asn Gly Ala Leu
Leu Arg

501 Val Tyr Val Pro Arg Ser Ser Leu Pro Gly Phe Tyr Arg Thr Gly Leu Thr Leu
Ala Ala

521 Pro Glu Ala Ala Gly Glu Val Glu Arg Leu Ile Gly His Pro Leu Pro Leu Arg
Leu Asp

Δ - rEPA

541 Ala Ile Thr Gly Pro Glu Glu Glu Gly Gly Arg Leu Glu Thr Ile Leu Gly Trp
Pro Leu

561 Ala Glu Arg Thr Val Val Ile Pro Ser Ala Ile Pro Thr Asp Pro Arg Asn Val
Gly Gly

581 Asp Leu Asp Pro Ser Ser Ile Pro Asp Lys Glu Gln Ala Ile Ser Ala Leu Pro
Asp Tyr

601 Ala Ser Gln Pro Gly Lys Pro Pro Arg Glu Asp Leu Lys

图 9-2

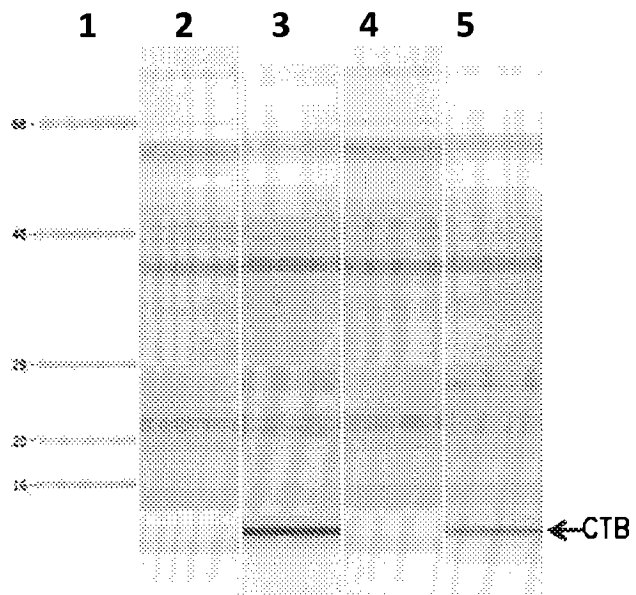


图 10

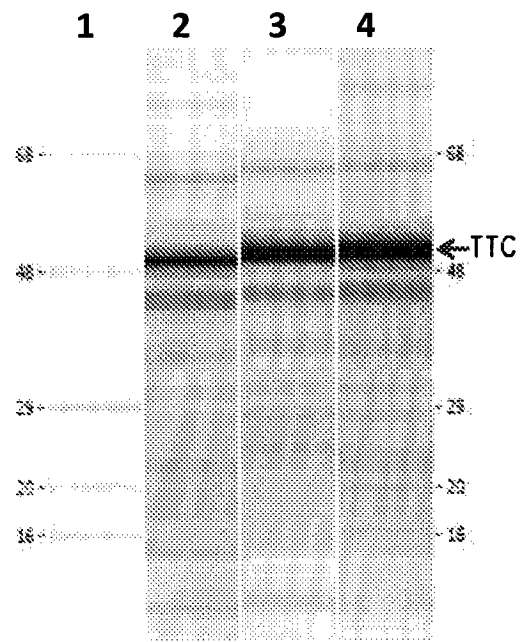


图 11A

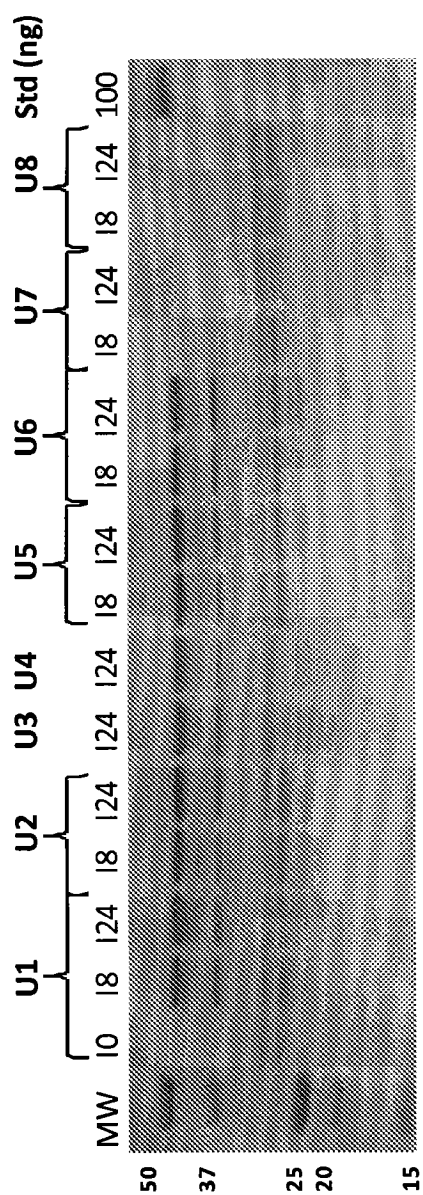


图 11B

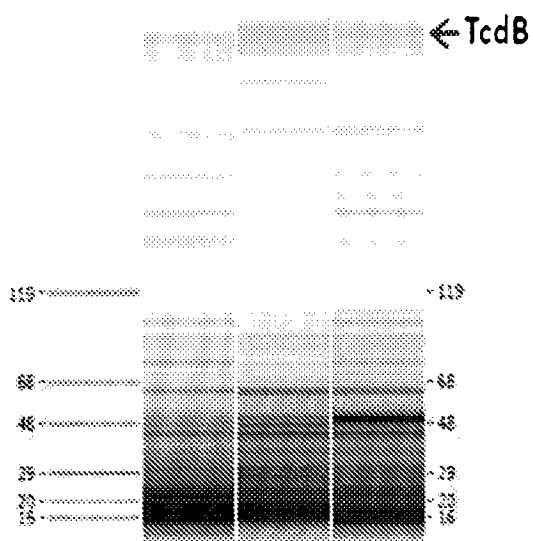


图 12

信号序列

```

~~~~~
S1
~~~~~
1 Met Arg Cys Thr Arg Ala Ile Arg Gln Thr Ala Arg Thr Gly Trp Leu Thr Trp Leu Ala
  ATG CGT TGC ACT CGG GCA ATT CGC CAA ACC GCA ACA ACA GGC TGG CTG ACG TGG CTG GCG
    signal sequence
~~~~~
S1
~~~~~
61 Ile Leu Ala Val Thr Ala Pro Val Thr Ser Pro Ala Trp Ala Asp Asp Pro Pro Ala Thr
  ATT CTT GCC GTC ACG GCG CCC GTG ACT TCG CCG GCA TGG GCC GAC GAT CCT CCC GCC ACC
    S1
~~~~~
121 Val Tyr Arg Tyr Asp Ser Arg Pro Pro Glu Asp Val Phe Gln Asn Gly Phe Thr Ala Trp
  GTA TAC CGC TAT GAC TCC CGC CCG CCG GAG GAC GTT TTC CAG AAC GGA TTC ACG GCG TGG
    S1
~~~~~
181 Gly Asn Asn Asp Asn Val Leu Asp His Leu Thr Gly Arg Ser Cys Gln Val Gly Ser Ser
  GGA AAC AAC GAC AAT GTG CTC GAC CAT CTG ACC GGA CGT TCC TGC CAG GTC GGC AGC AGC
    S1
~~~~~
241 Asn Ser Ala Phe Val Ser Thr Ser Ser Ser Arg Arg Tyr Thr Glu Val Tyr Leu Glu His
  AAC AGC GCT TTC GTC TCC ACC AGC AGC AGC CCG CCG TAT ACC GAG GTC TAT CTC GAA CAT
    S1
~~~~~
301 Arg Met Gln Glu Ala Val Glu Ala Glu Arg Ala Gly Arg Gly Thr Gly His Phe Ile Gly
  CGC ATG CAG GAA GCG GTC GAG GCC GAA CCG GCC GGC AGG GGC ACC GGC CAC TTC ATC GGC
    S1
~~~~~
361 Tyr Ile Tyr Glu Val Arg Ala Asp Asn Asn Phe Tyr Gly Ala Ala Ser Ser Tyr Phe Glu
  TAC ATC TAC GAA GTC CCG GCC GAC AAC AAT TTC TAC GGC GCC GCC AGC TCG TAC TTC GAA
    S1
~~~~~
421 Tyr Val Asp Thr Tyr Gly Asp Asn Ala Gly Arg Ile Leu Ala Gly Ala Leu Ala Thr Tyr
  TAC GTC GAC ACT TAT GGC GAC AAT GCC GGC CGT ATC CTC GCC GGC GCG CTG GCC ACC TAC
    S1
~~~~~
481 Gln Ser Glu Tyr Leu Ala His Arg Arg Ile Pro Pro Glu Asn Ile Arg Arg Val Thr Arg
  CAG AGC GAA TAT CTG GCA CAC CGG CCG ATT CCG CCC GAA AAC ATC CGC AGG GTA ACG CGG
    S1
~~~~~
541 Val Tyr His Asn Gly Ile Thr Gly Glu Thr Thr Thr Thr Thr Glu Tyr Ser Asn Ala Arg Tyr
  GTC TAT CAC AAC GGC ATC ACC GGC GAG ACC ACG ACC ACG GAG TAT TCC AAC GCT CGC TAC

```

图 13-1

```

S1
~~~~~
Val Ser Gln Gln Thr Arg Ala Asn Pro Asn Pro Tyr Thr Ser Arg Arg Ser Val Ala Ser
601 GTC AGC CAG CAG ACT CGC GCC AAT CCC AAC CCC TAC ACA TCG CGA AGG TCC GTA GCG TCG
S1
~~~~~
Ile Val Gly Thr Leu Val Arg Met Ala Pro Val Ile Gly Ala Cys Met Ala Arg Gln Ala
661 ATC GTC GGC ACA TTG GTG CGC ATG GCG CCG GTG ATA GGC GCT TGC ATG GCG CGG CAG GCC
S1
~~~~~
Glu Ser Ser Glu Ala Met Ala Ala Trp Ser Glu Arg Ala Gly Glu Ala Met Val Leu Val
721 GAA AGC TCC GAG GCC ATG GCA GCC TGG TCC GAA CGC GCC GGC GAG GCG ATG GTT CTC GTG
S1
~~~~~
Tyr Tyr Glu Ser Ile Ala Tyr Ser Phe
781 TAC TAC GAA AGC ATC GCG TAT TCG TTC TAG ACC TGG CCC AGC CCC GCC CAA CTC CGG TAA
信号序列
~~~~~
S2
~~~~~
Met Pro Ile Asp Arg Lys Thr Leu Cys His Leu Leu Ser Val Leu Pro Leu
841 TTG AAC AGC ATG CCG ATC GAC CGC AAG ACG CTC TGC CAT CTC CTG TCC GTT CTG CCG TTG
信号序列
~~~~~
S2
~~~~~
Ala Leu Leu Gly Ser His Val Ala Arg Ala Ser Thr Pro Gly Ile Val Ile Pro Pro Gln
901 GCC CTC CTC GGA TCT CAC GTG GCG CGG GCC TCC ACG CCA GGC ATC GTC ATT CCG CCG CAG
S2
~~~~~
Glu Gln Ile Thr Gln His Gly Ser Pro Tyr Gly Arg Cys Ala Asn Lys Thr Arg Ala Leu
961 GAA CAG ATT ACC CAG CAT GGC AGC CCC TAT GGA CGC TGC GCG AAC AAG ACC CGT GCC CTG
S2
~~~~~
Thr Val Ala Glu Leu Arg Gly Ser Gly Asp Leu Gln Glu Tyr Leu Arg His Val Thr Arg
1021 ACC GTG GCG GAA TTG CGC GGC AGC GGC GAT CTG CAG GAG TAC CTG CGT CAT GTG ACG CGC
S2
~~~~~
Gly Trp Ser Ile Phe Ala Leu Tyr Asp Gly Thr Tyr Leu Gly Gly Glu Tyr Gly Gly Val
1081 GGC TGG TCA ATA TTT GCG CTC TAC GAT GGC ACC TAT CTC GGC GGC GAA TAT GGC GGC GTG
S2
~~~~~
Ile Lys Asp Gly Thr Pro Gly Gly Ala Phe Asp Leu Lys Thr Thr Phe Cys Ile Met Thr
1141 ATC AAG GAC GGA ACA CCC GGC GGC GCA TTC GAC CTG AAA ACG ACG TTC TGC ATC ATG ACC
S2
~~~~~
Thr Arg Asn Thr Gly Gln Pro Ala Thr Asp His Tyr Tyr Ser Asn Val Thr Ala Thr Arg
1201 ACG CGC AAT ACG GGT CAA CCC GCA ACG GAT CAC TAC TAC AGC AAC GTC ACC GCC ACT CGC

```

图 13-2

```

S2
~~~~~
1261 Leu Leu Ser Ser Thr Asn Ser Arg Leu Cys Ala Val Phe Val Arg Ser Gly Gln Pro Val
CTG CTC TCC AGC ACC AAC AGC AGG CTA TGC GCG GTC TTC GTC AGA AGC GGG CAA CCG GTC
S2
~~~~~
1321 Ile Gly Ala Cys Thr Ser Pro Tyr Asp Gly Lys Tyr Trp Ser Met Tyr Ser Arg Leu Arg
ATT GGC GCC TGC ACC AGC CCG TAT GAC GGC AAG TAC TGG AGC ATG TAC AGC CGG CTG CGG
S2
~~~~~
1381 Lys Met Leu Tyr Leu Ile Tyr Val Ala Gly Ile Ser Val Arg Val His Val Ser Lys Glu
AAA ATG CTT TAC CTG ATC TAC GTG GCC GGC ATC TCC GTA CGC GTC CAT GTC AGC AAG GAA
S2
~~~~~
1441 Glu Gln Tyr Tyr Asp Tyr Glu Asp Ala Thr Phe Glu Thr Tyr Ala Leu Thr Gly Ile Ser
GAA CAG TAT TAC GAC TAT GAG GAC GCA ACG TTC GAG ACT TAC GCC CTT ACC GGC ATC TCC
S2
~~~~~
1501 Ile Cys Asn Pro Gly Ser Ser Leu Cys
ATC TGC AAT CCT GGA TCA TCC TTA TGC TGA GAC GCT TCC CCA CTC GAA CCA CCG CCC CGG
信号序列
~~~~~
S4
~~~~~
1561 Val Arg Ala Leu Ala Trp Leu Leu Ala Ser Gly
GAC AGG GCG GCG CCC GGC GGT CGC GC GTG CGC GCC CTG GCG TGG TTG CTG GCA TCC GGC
信号序列
~~~~~
S4
~~~~~
1620 Ala Met Thr His Leu Ser Pro Ala Leu Ala Asp Val Pro Tyr Val Leu Val Lys Thr Asn
GCG ATG ACG CAT CTT TCC CCC GCC CTG GCC GAC GTT CCT TAT GTG CTG GTG AAG ACC AAT
S4
~~~~~
1680 Met·Val Val Thr Ser Val Ala Met Lys Pro Tyr Glu Val Thr Pro Thr Arg Met Leu Val
ATG GTG GTC ACC AGC GTA GCC ATG AAG CCG TAT GAA GTC ACC CCG ACG CGC ATG CTG GTC
S4
~~~~~
1740 Cys Gly Ile Ala Ala Lys Leu Gly Ala Ala Ala Ser Ser Pro Asp Ala His Val Pro Phe
TGC GGC ATC GCC GCC AAA CTG GGC GCC GCG GCC AGC AGC CCG GAC GCG CAC GTG CCG TTC
S4
~~~~~
1800 Cys Phe Gly Lys Asp Leu Lys Arg Pro Gly Ser Ser Pro Met Glu Val Met Leu Arg Ala
TGC TTC GGC AAG GAT CTC AAG CGT CCC GGC AGC AGT CCC ATG GAA GTC ATG TTG CGC GCC
S4
~~~~~
1860 Val Phe Met Gln Gln Arg Pro Leu Arg Met Phe Leu Gly Pro Lys Gln Leu Thr Phe Glu
GTC TTC ATG CAA CAA CGG CCG CTG CGC ATG TTT CTG GGT CCC AAG CAA CTC ACT TTC GAA

```

图 13-3

```

S4
~~~~~
Gly Lys Pro Ala Leu Glu Leu Ile Arg Met Val Glu Cys Ser Gly Lys Gln Asp Cys Pro
1920 GGC AAG CCC GCG CTC GAA CTG ATC CGG ATG GTC GAA TGC AGC GGC AAG CAG GAT TGC CCC T
      信号序列
S5
~~~~~
Met His Thr Ile Ala Ser Ile Leu Leu Ser Val Leu Gly Ile Tyr Ser
1981 GAA GGC GAA CCC C ATG CAT ACC ATC GCA TCC ATC CTG TTG TCC GTG CTC GGC ATA TAC AG
      信号序列
S5
~~~~~
Pro Ala Asp Val Ala Gly Leu Pro Thr His Leu Tyr Lys Asn Phe Thr Val Gln Glu Leu
2041 C CCG GCT GAC GTC GCC GGC TTG CCG ACC CAT CTG TAC AAG AAC TTC ACT GTC CAG GAG CT
      S5
~~~~~
Ala Leu Lys Leu Lys Gly Lys Asn Gln Glu Phe Cys Leu Thr Ala Phe Met Ser Gly Arg
2101 G GCC TTG AAA CTG AAG GGC AAG AAT CAG GAG TTC TGC CTG ACC GCC TTC ATG TCG GGC AG
      S5
~~~~~
Ser Leu Val Arg Ala Cys Leu Ser Asp Ala Gly His Glu His Asp Thr Trp Phe Asp Thr
2161 A AGC CTG GTC CCG GCG TGC CTG TCC GAC GCG GGA CAC GAG CAC GAC ACG TGG TTC GAC AC
      S5
~~~~~
Met Leu Gly Phe Ala Ile Ser Ala Tyr Ala Leu Lys Ser Arg Ile Ala Leu Thr Val Glu
2221 C ATG CTT GGC TTT GCC ATA TCC GCG TAT GCG CTC AAG AGC CGG ATC GCG CTG ACG GTG GA
      S5
~~~~~
Asp Ser Pro Tyr Pro Gly Thr Pro Gly Asp Leu Leu Glu Leu Gln Ile Cys Pro Leu Asn
2281 A GAC TCG CCG TAT CCG GGC ACT CCC GGC GAT CTG CTC GAA CTG CAG ATC TGC CCG CTC AA
      S5
~~~~~
Gly Tyr Cys Glu
2341 C GGA TAT TGC GAA TG AAC CCT TCC GGA GGT TTC GAC GTT TCC GCG CAA TCC GCT TGA GAC
      信号序列
S3
~~~~~
Met Leu Ile Asn Asn Lys Lys Leu
2401 GAT CTT CCG CCC TGG TTC CAT TCC GGG AAC ACC GCA AC ATG CTG ATC AAC AAC AAG AAG C
      信号序列
S3
~~~~~
Leu His His Ile Leu Pro Ile Leu Val Leu Ala Leu Leu Gly Met Arg Thr Ala Gln Ala
2461 TG CTT CAT CAC ATT CTG CCC ATC CTG GTG CTC GCC CTG CTG GGC ATG CGC ACG GCC CAG G

```

图 13-4

信号序列

```

~~~~~
                                S3
~~~~~
2521  Val Ala Pro Gly Ile Val Ile Pro Pro Lys Ala Leu Phe Thr Gln Gln Gly Gly Ala Tyr·
      CC GTT GCG CCA GGC ATC GTC ATC CCG CCG AAG GCA CTG TTC ACC CAA CAG GGC GGC GCC T
                                S3
~~~~~
2581  Gly Arg Cys Pro Asn Gly Thr Arg Ala Leu Thr Val Ala Glu Leu Arg Gly Asn Ala Glu·
      AT GGA CGC TGC CCG AAC GGA ACC CGC GCC TTG ACC GTG GCC GAA CTG CGC GGC AAC GCC G
                                S3
~~~~~
2641  Leu Gln Thr Tyr Leu Arg Gln Ile Thr Pro Gly Trp Ser Ile Tyr Gly Leu Tyr Asp Gly·
      AA TTG CAG ACG TAT TTG CGC CAG ATA ACG CCC GGC TGG TCC ATA TAC GGT CTC TAT GAC G
                                S3
~~~~~
2701  Thr Tyr Leu Gly Gln Ala Tyr Gly Gly Ile Ile Lys Asp Ala Pro Pro Gly Ala Gly Phe·
      GT ACG TAC CTG GGC CAG GCG TAC GGC GGC ATC ATC AAG GAC GCG CCG CCA GGC GCG GGG T
                                S3
~~~~~
2761  Ile Tyr Arg Glu Thr Phe Cys Ile Thr Thr Ile Tyr Lys Thr Gly Gln Pro Ala Ala Asp·
      TC ATT TAT CGC GAA ACT TTC TGC ATC ACG ACC ATA TAC AAG ACC GGG CAA CCG GCT GCG G
                                S3
~~~~~
2821  His Tyr Tyr Ser Lys Val Thr Ala Thr Arg Leu Leu Ala Ser Thr Asn Ser Arg Leu Cys·
      AT CAC TAC TAC AGC AAG GTC ACG GCC ACG CGC CTG CTC GCC AGC ACC AAC AGC AGG CTG T
                                S3
~~~~~
2881  Ala Val Phe Val Arg Asp Gly Gln Ser Val Ile Gly Ala Cys Ala Ser Pro Tyr Glu Gly·
      GC GCG GTA TTC GTC AGG GAC GGG CAA TCG GTC ATC GGA GCC TGC GCC AGC CCG TAT GAA G
                                S3
~~~~~
2941  Arg Tyr Arg Asp Met Tyr Asp Ala Leu Arg Arg Leu Leu Tyr Met Ile Tyr Met Ser Gly·
      GC AGG TAC AGA GAC ATG TAC GAC GCG CTG CGG CGC CTG CTG TAC ATG ATC TAT ATG TCC G
                                S3
~~~~~
3001  Leu Ala Val Arg Val His Val Ser Lys Glu Glu Gln Tyr Tyr Asp Tyr Glu Asp Ala Thr·
      GC CTT GCC GTA CGC GTC CAC GTC AGC AAG GAG GAA CAG TAT TAC GAC TAC GAG GAC GCC A
                                S3
~~~~~
3061  Phe Gln Thr Tyr Ala Leu Thr Gly Ile Ser Leu Cys Asn Pro Ala Ala Ser Ile Cys
      CA TTC CAG ACC TAT GCC CTC ACC GGC ATT TCC CTC TGC AAC CCG GCA GCG TCG ATA TGC

```

图 13-5

**A. 霍乱毒素 A 亚基 (AE003852; 蛋白质 IDAAF94614.1),
带下划线标注的分泌前导序列**

1 Met Val Lys Ile Ile Phe Val Phe Phe Ile Phe Leu Ser Ser
 Phe Ser Tyr Ala Asn Asp

21 Asp Lys Leu Tyr Arg Ala Asp Ser Arg Pro Pro Asp Glu Ile
 Lys Gln Ser Gly Gly Leu

41 Met Pro Arg Gly Gln Ser Glu Tyr Phe Asp Arg Gly Thr Gln
 Met Asn Ile Asn Leu Tyr

61 Asp His Ala Arg Gly Thr Gln Thr Gly Phe Val Arg His Asp
 Asp Gly Tyr Val Ser Thr

81 Ser Ile Ser Leu Arg Ser Ala His Leu Val Gly Gln Thr Ile
 Leu Ser Gly His Ser Thr

101 Tyr Tyr Ile Tyr Val Ile Ala Thr Ala Pro Asn Met Phe Asn
 Val Asn Asp Val Leu Gly

121 Ala Tyr Ser Pro His Pro Asp Glu Gln Glu Val Ser Ala Leu
 Gly Gly Ile Pro Tyr Ser

141 Gln Ile Tyr Gly Trp Tyr Arg Val His Phe Gly Val Leu Asp
 Glu Gln Leu His Arg Asn

161 Arg Gly Tyr Arg Asp Arg Tyr Tyr Ser Asn Leu Asp Ile Ala
 Pro Ala Ala Asp Gly Tyr

181 Gly Leu Ala Gly Phe Pro Pro Glu His Arg Ala Trp Arg Glu
 Glu Pro Trp Ile His His

201 Ala Pro Pro Gly Cys Gly Asn Ala Pro Arg Ser Ser Met Ser
 Asn Thr Cys Asp Glu Lys

221 Thr Gln Ser Leu Gly Val Lys Phe Leu Asp Glu Tyr Gln Ser
 Lys Val Lys Arg Gln Ile

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

图 14-1

**B. 霍乱毒素 B 氨基酸序列，带下划线标注的分泌前导序列
(AE003852; 蛋白质 IDAAF94613.1)**

1 Met Ile Lys Leu Lys Phe Gly Val Phe Phe Thr Val Leu Leu Ser
Ser Ala Tyr Ala His

21 Gly Thr Pro Gln Asn Ile Thr Asp Leu Cys Ala Glu Tyr His Asn
Thr Gln Ile Tyr Thr

41 Leu Asn Asp Lys Ile Phe Ser Tyr Thr Glu Ser Leu Ala Gly Lys
Arg Glu Met Ala Ile

61 Ile Thr Phe Lys Asn Gly Ala Ile Phe Gln Val Glu Val Pro Gly
Ser Gln His Ile Asp

81 Ser Gln Lys Lys Ala Ile Glu Arg Met Lys Asp Thr Leu Arg Ile
Ala Tyr Leu Thr Glu

101 Ala Lys Val Glu Lys Leu Cys Val Trp Asn Asn Lys Thr Pro His
Ala Ile Ala Ala Ile

121 Ser Met Ala Asn

图 14-2

C. 霍乱毒素 DNA 序列 (A 和 B 编码区域 Genbank AE003852)

分泌前导序列

```

~~~~~
CTA
~~~~~
Met Val Lys Ile Ile Phe Val Phe Phe Ile Phe Leu Ser Ser Phe Ser Tyr Ala Asn Asp
1 ATG GTA AAG ATA ATA TTT GTG TTT TTT ATT TTC TTA TCA TCA TTT TCA TAT GCA AAT GAT
CTA
~~~~~
Asp Lys Leu Tyr Arg Ala Asp Ser Arg Pro Pro Asp Glu Ile Lys Gln Ser Gly Gly Leu
61 GAT AAG TTA TAT CGG GCA GAT TCT AGA CCT CCT GAT GAA ATA AAG CAG TCA GGT GGT CTT
CTA
~~~~~
Met Pro Arg Gly Gln Ser Glu Tyr Phe Asp Arg Gly Thr Gln Met Asn Ile Asn Leu Tyr
121 ATG CCA AGA GGA CAG AGT GAG TAC TTT GAC CGA GGT ACT CAA ATG AAT ATC AAC CTT TAT
CTA
~~~~~
Asp His Ala Arg Gly Thr Gln Thr Gly Phe Val Arg His Asp Asp Gly Tyr Val Ser Thr
181 GAT CAT GCA AGA GGA ACT CAG ACG GGA TTT GTT AGG CAC GAT GAT GGA TAT GTT TCC ACC
CTA
~~~~~
Ser Ile Ser Leu Arg Ser Ala His Leu Val Gly Gln Thr Ile Leu Ser Gly His Ser Thr
241 TCA ATT AGT TTG AGA AGT GCC CAC TTA GTG GGT CAA ACT ATA TTG TCT GGT CAT TCT ACT
CTA
~~~~~
Tyr Tyr Ile Tyr Val Ile Ala Thr Ala Pro Asn Met Phe Asn Val Asn Asp Val Leu Gly
301 TAT TAT ATA TAT GTT ATA GCC ACT GCA CCC AAC ATG TTT AAC GTT AAT GAT GTA TTA GGG
CTA
~~~~~
Ala Tyr Ser Pro His Pro Asp Glu Gln Glu Val Ser Ala Leu Gly Gly Ile Pro Tyr Ser
361 GCA TAC AGT CCT CAT CCA GAT GAA CAA GAA GTT TCT GCT TTA GGT GGG ATT CCA TAC TCC
CTA
~~~~~
Gln Ile Tyr Gly Trp Tyr Arg Val His Phe Gly Val Leu Asp Glu Gln Leu His Arg Asn
421 CAA ATA TAT GGA TGG TAT CGA GTT CAT TTT GGG GTG CTT GAT GAA CAA TTA CAT CGT AAT
CTA
~~~~~
Arg Gly Tyr Arg Asp Arg Tyr Tyr Ser Asn Leu Asp Ile Ala Pro Ala Ala Asp Gly Tyr
481 AGG GGC TAC AGA GAT AGA TAT TAC AGT AAC TTA GAT ATT GCT CCA GCA GCA GAT GGT TAT
CTA
~~~~~
Gly Leu Ala Gly Phe Pro Pro Glu His Arg Ala Trp Arg Glu Glu Pro Trp Ile His His
541 GGA TTG GCA GGT TTC CCT CCG GAG CAT AGA GCT TGG AGG GAA GAG CCG TGG ATT CAT CAT
CTA
~~~~~
Ala Pro Pro Gly Cys Gly Asn Ala Pro Arg Ser Ser Met Ser Asn Thr Cys Asp Glu Lys
601 GCA CCG CCG GGT TGT GGG AAT GCT CCA AGA TCA TCG ATG AGT AAT ACT TGC GAT GAA AAA

```

图 14-3

```

          CTA
~~~~~
Thr Gln Ser Leu Gly Val Lys Phe Leu Asp Glu Tyr Gln Ser Lys Val Lys Arg Gln Ile
661 ACC CAA AGT CTA GGT GTA AAA TTC CTT GAC GAA TAC CAA TCT AAA GTT AAA AGA CAA ATA
          CDS_1
          ~~~~~
          CTA
~~~~~
          分泌前导序列
          ~~~~~
          CTB
          ~~~~~
          Met Ile Lys·
Phe Ser Gly Tyr Gln Ser Asp Ile Asp Thr His Asn Arg Ile Lys Asp Glu Leu ***
721 TTT TCA GGC TAT CAA TCT GAT ATT GAT ACA CAT AAT AGA ATT AAG GAT GAA TTA TGA TTA
          分泌前导序列
          ~~~~~
          CTB
          ~~~~~
·LLeu Lys Phe Gly Val Phe Phe Thr Val Leu Leu Ser Ser Ala Tyr Ala His Gly Thr Pro·
781 AAT TAA AAT TTG GTG TTT TTT TTA CAG TTT TAC TAT CTT CAG CAT ATG CAC ATG GAA CAC
          CTB
          ~~~~~
·PLeu Asn Ile Thr Asp Leu Cys Ala Glu Tyr His Asn Thr Gln Ile Tyr Thr Leu Asn Asp·
841 CTC AAA ATA TTA CTG ATT TGT GTG CAG AAT ACC ACA ACA CAC AAA TAT ATA CGC TAA ATG
          CTB
          ~~~~~
·ALys Ile Phe Ser Tyr Thr Glu Ser Leu Ala Gly Lys Arg Glu Met Ala Ile Ile Thr Phe·
901 ATA AGA TAT TTT CGT ATA CAG AAT CTC TAG CTG GAA AAA GAG AGA TGG CTA TCA TTA CTT
          CTB
          ~~~~~
·PLys Asn Gly Ala Ile Phe Gln Val Glu Val Pro Gly Ser Gln His Ile Asp Ser Gln Lys·
961 TTA AGA ATG GTG CAA TTT TTC AAG TAG AAG TAC CAG GTA GTC AAC ATA TAG ATT CAC AAA
          CTB
          ~~~~~
·LLys Ala Ile Glu Arg Met Lys Asp Thr Leu Arg Ile Ala Tyr Leu Thr Glu Ala Lys Val·
1021 AAA AAG CGA TTG AAA GGA TGA AGG ATA CCC TGA GGA TTG CAT ATC TTA CTG AAG CTA AAG
          CTB
          ~~~~~
·VLeu Lys Leu Cys Val Trp Asn Asn Lys Thr Pro His Ala Ile Ala Ala Ile Ser Met Ala·
1081 TCG AAA AGT TAT GTG TAT GGA ATA ATA AAA CGC CTC ATG CGA TTG CCG CAA TTA GTA TGG
          CTB
          ~~~~~
·AAsn
1141 CAA ATT AA

```

图 14-4

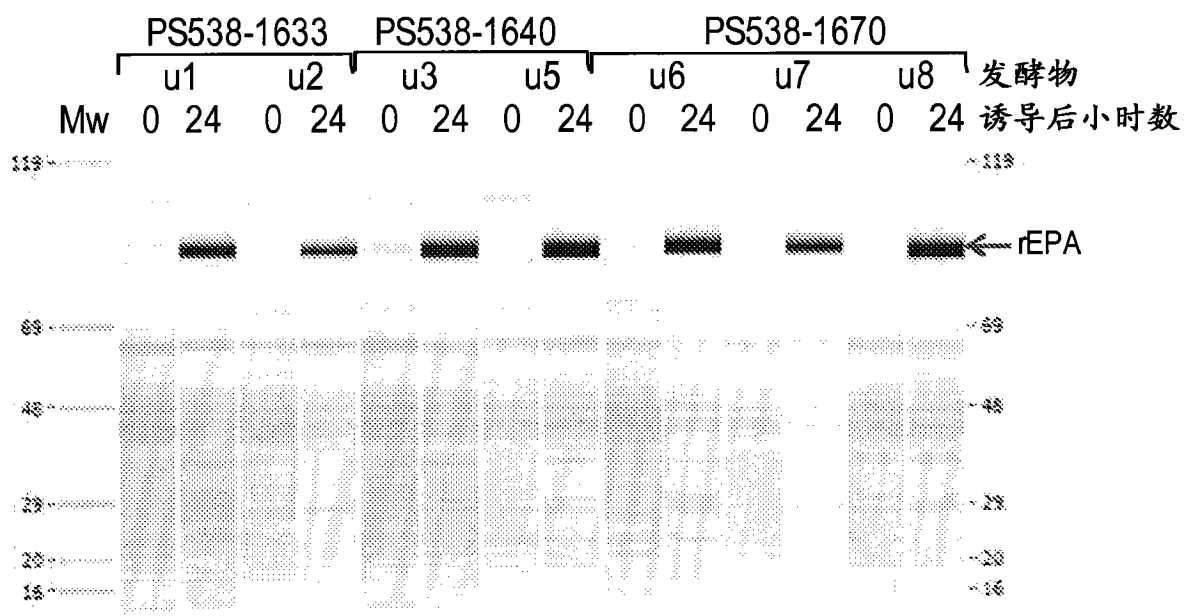


图 15

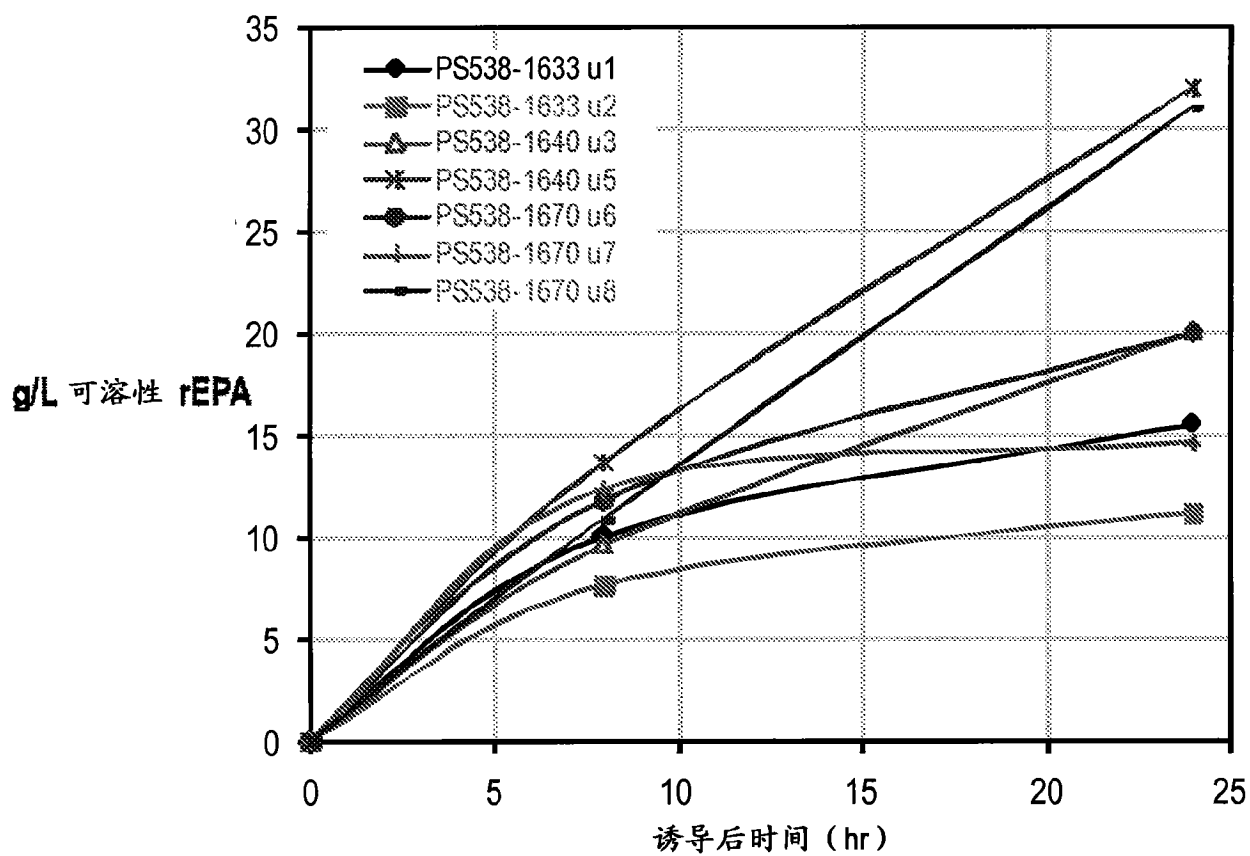


图 16

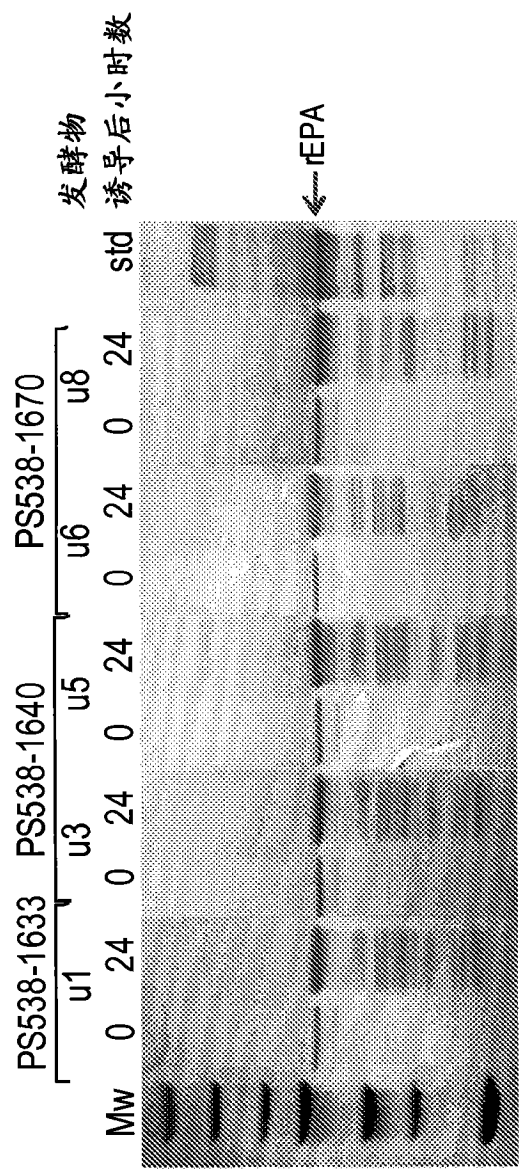


图 17

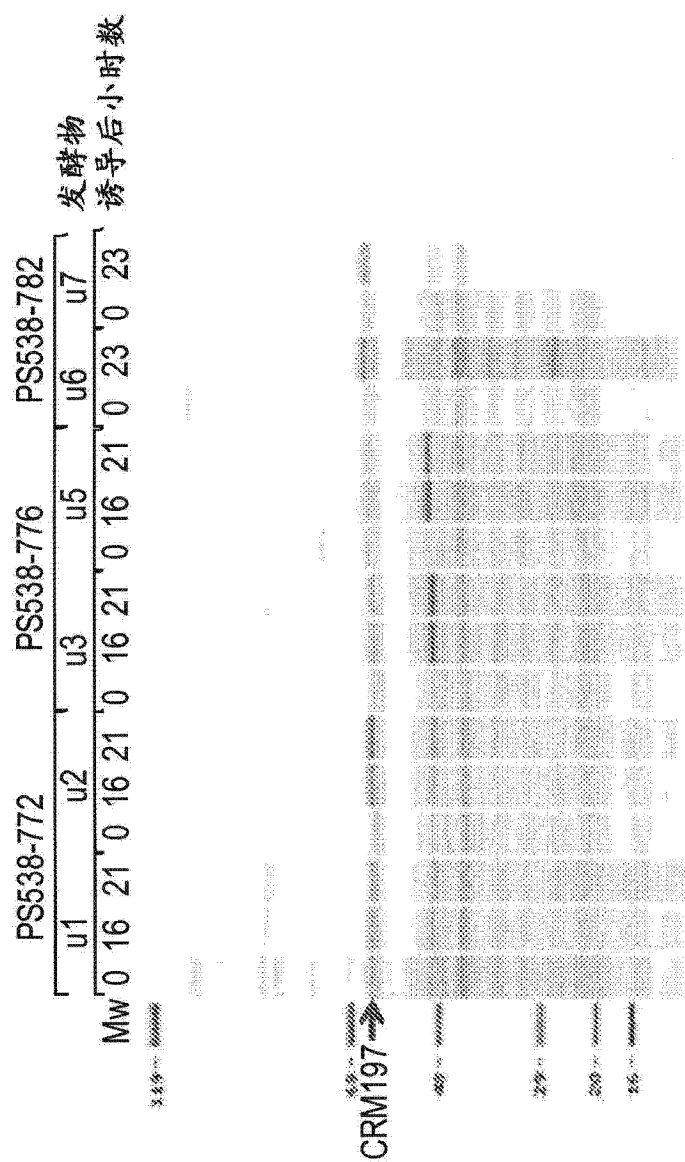


图 18

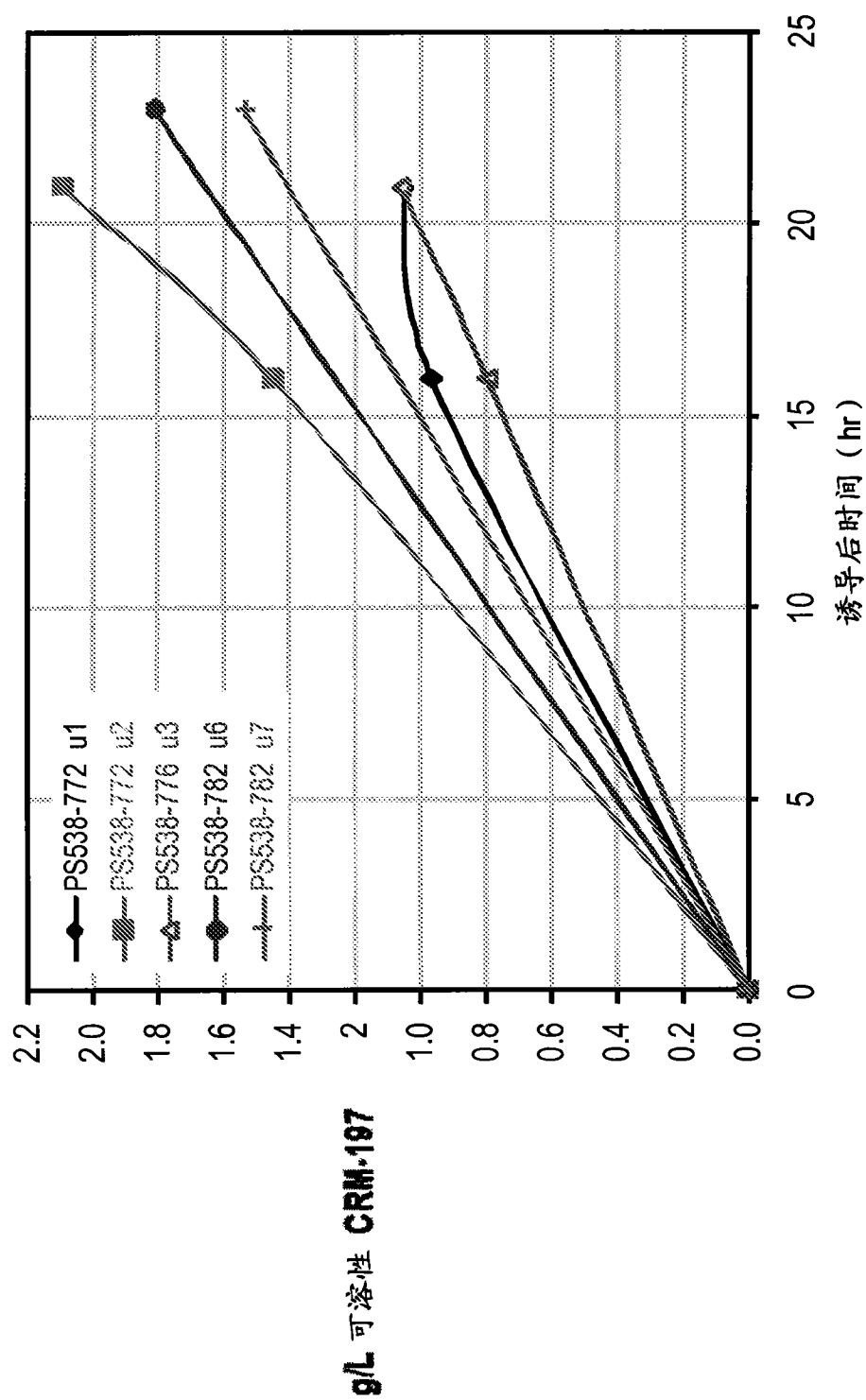


图 19

