

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公布说明书

[21] 申请号 200480043720.0

[51] Int. Cl.

C07K 5/00 (2006.01)

C07K 5/04 (2006.01)

C12N 9/00 (2006.01)

[43] 公开日 2007年7月4日

[11] 公开号 CN 1993377A

[22] 申请日 2004.7.30

[21] 申请号 200480043720.0

[86] 国际申请 PCT/US2004/024686 2004.7.30

[87] 国际公布 WO2006/022664 英 2006.3.2

[85] 进入国家阶段日期 2007.1.30

[71] 申请人 卡吉尔公司

地址 美国明尼苏达州

[72] 发明人 奥利·J·耶森 拉维·R·戈卡恩

史蒂文·J·戈特

奥尔加·V·谢利丰诺娃

汉斯·H·廖 布赖恩·J·布拉佐

[74] 专利代理机构 中原信达知识产权代理有限责任
公司

代理人 杨青 樊卫民

权利要求书9页 说明书81页 序列表76页
附图10页

[54] 发明名称

丙氨酸2,3氨基变位酶

[57] 摘要

本发明公开了丙氨酸2,3-氨基变位酶的序列,本发明还公开了具有丙氨酸2,3-氨基变位酶活性的细胞以及选择这些细胞的方法。本发明还公开了制备 β -丙氨酸,泛酸盐,3-羟丙酸,以及其他有机化合物的方法。

1. 包含丙氨酸 2,3-氨基变位酶活性的分离多肽，其中所述多肽含有突变的赖氨酸 2,3-氨基变位酶的氨基酸序列，且其中突变的赖氨酸 2,3-氨基变位酶氨基酸序列含有 P/S11T; N19Y; L/K/R/T26I; E/R30K; L/V32A; K36E; S/T/C52R; L/T53P/H; Y63F; E/N/D71G; H/I/S85Q; Q/L/E86R; Q/L95M; K/M/Q125L; M128V; Y132H; Q/S141R; A/D/S/M144G; D179N; K/Q187R; I192V; L228M; D331G/H; M/Q342T 或 K/Q/T398E 取代，或其组合，其中所述编号是基于牙龈卟啉单胞菌 (*Porphyromonas gingivalis*) 赖氨酸 2,3 氨基变位酶。

2. 权利要求 1 的分离多肽，其中所述突变的赖氨酸 2,3-氨基变位酶氨基酸序列是突变的枯草芽孢杆菌 (*Bacillus subtilis*)，斯氏梭菌 (*Clostridium sticklandii*)，具核梭杆菌 (*Fusobacterium nucleatum*) 或牙龈卟啉单胞菌 (*Porphyromonas gingivalis*) 赖氨酸 2,3-氨基变位酶。

3. 权利要求 1 的分离多肽，其中所述突变的赖氨酸 2,3-氨基变位酶氨基酸序列含有 N19Y, L/T53P/H, H/I/S85Q, D331G/H 和 M/Q342T 取代。

4. 权利要求 1 的分离多肽，其中所述突变的赖氨酸 2,3-氨基变位酶氨基酸序列含有 N19Y, E/R30K, L/T53P/H, H/I/S85Q, I192V, D331G/H 和 M/Q342T 取代。

5. 权利要求 1 的分离多肽，其中所述突变的赖氨酸 2,3-氨基变位酶氨基酸序列含有 N19Y, L/K/R/T26I; E/R30K, L/T53P/H, H/I/S85Q, I192V, D331G/H 和 M/Q342T 取代。

6. 权利要求 1 的分离多肽，其中所述突变的赖氨酸 2,3-氨基变位酶氨基酸序列含有 E/R30K, Y63F, Q/L/E86R, Q/L95M, M128V,

A/D/S/M144G, L228M, D331G/H 和 K/Q/T398E 取代。

7. 权利要求 1 的分离多肽, 其中所述突变的赖氨酸 2,3-氨基变位酶氨基酸序列含有 E/R30K, C52R, Q/L95M; M128V 和 D331G/H 取代。

8. 权利要求 1 的分离多肽, 其中所述突变的赖氨酸 2,3-氨基变位酶氨基酸序列含有 E/R30K, K36E, Y63F, Q/L/E86R, Q/L95M, M128V, A/D/S/M144G, D179N, L228M, D331G/H 和 K/Q/T398E 取代。

9. 权利要求 1 的分离多肽, 其中所述突变的赖氨酸 2,3-氨基变位酶氨基酸序列含有 E/R30K, Q/L95M, M128V 和 D331G/H 取代。

10. 权利要求 1 的分离多肽, 其中所述突变的赖氨酸 2,3-氨基变位酶氨基酸序列含有 P/S11T, E/R30K, Q/L95M, M128V, Q/S141R, K/Q187R 和 D331G/H 取代。

11. 权利要求 1 的分离多肽, 其中所述突变的赖氨酸 2,3-氨基变位酶氨基酸序列含有 E/R30K, L/V32A, L/T53P/H, E/N/D71G, Q/L95M; K/M/Q125L, M128V 和 D331G/H 取代。

12. 权利要求 1 的分离多肽, 其中所述突变的赖氨酸 2,3-氨基变位酶氨基酸序列含有 Q/L95M, M128V 和 D331G/H 取代。

13. 权利要求 1 的分离多肽, 其中所述突变的赖氨酸 2,3-氨基变位酶氨基酸序列含有 Q/L95M, M128V; Y132H 和 D331G/H 取代。

14. 权利要求 1 的分离多肽, 其中所述突变的赖氨酸 2,3-氨基变位酶氨基酸序列含有至少 3 个取代。

15. 权利要求 1 的分离多肽,其中所述突变的赖氨酸 2,3-氨基变位酶氨基酸序列含有 3-11 个取代。

16. 权利要求 1 的分离多肽,其中所述多肽含有与 SEQ ID NO: 19, 21, 43, 45, 47, 49 或 51 具有至少 90%同一性的序列。

17. 权利要求 1 的分离多肽,其中所述多肽含有与 SEQ ID NO: 19, 21, 43, 45, 47, 49 或 51 具有至少 95%同一性的序列。

18. 权利要求 1 的分离多肽,其中所述多肽含有 SEQ ID NO: 19, 21, 43, 45, 47, 49 或 51。

19. 权利要求 17 的多肽,其中所述多肽含有 1-10 个保守氨基酸取代。

20. 含有编码如权利要求 1 所述分离多肽的核酸序列的分离核酸。

21. 可操作连接于启动子序列的如权利要求 20 所述的分离核酸。

22. 权利要求 20 的分离核酸,其中所述核酸含有与 SEQ ID NO: 18, 20, 42, 44, 46, 48 或 50 具有至少 90%同一性的序列。

23. 权利要求 20 的分离核酸,其中所述核酸含有与 SEQ ID NO: 18, 20, 42, 44, 46, 48 或 50 具有至少 95%同一性的序列。

24. 权利要求 20 的分离核酸,其中所述核酸序列包括导致 1-10 保守氨基酸取代的一个或多个取代。

25. 权利要求 20 的分离核酸,其中所述核酸包括 SEQ ID NO: 18, 20, 42, 44, 46, 48 或 50。

26. 含有如权利要求 20 所述的分离核酸的载体。
27. 含有如权利要求 20 所述的分离核酸的重组核酸。
28. 用权利要求 27 所述的重组核酸转化的细胞。
29. 权利要求 28 的细胞，其中所述细胞是原核细胞。
30. 权利要求 29 的细胞，其中所述原核细胞是乳杆菌(Lactobacillus)，乳球菌(Lactococcus)，芽孢杆菌(Bacillus)或埃希氏肠杆菌细胞(Escherichia)。
31. 权利要求 28 的细胞，其中所述细胞是植物细胞，细菌细胞，酵母细胞或真菌细胞。
32. 含有如权利要求 31 所述的细胞的植物。
33. 含有如权利要求 27 所述的重组核酸的转基因植物。
34. 权利要求 28 的细胞，其中所述细胞含有丙氨酸 2,3-氨基变位酶活性并可从 α -丙氨酸制备 β -丙氨酸。
35. 权利要求 28 的细胞，其中所述分离的核酸序列含有与 SEQ ID NO: 18, 20, 42, 44, 46, 48 或 50 具有至少 90% 同一性的序列。
36. 权利要求 28 的细胞，其中所述分离核酸序列含有 SEQ ID NO: 18, 20, 42, 44, 46, 48 或 50。
37. 权利要求 28 的细胞，其中所述细胞产生 3-羟丙酸(3-HP)。

38. 权利要求 37 的细胞，其中所述细胞还包括：

丙酮酸/2-酮戊二酸盐氨基转移酶活性；

β -丙氨酸/2-酮戊二酸盐氨基转移酶活性；和

3-羟丙酸盐脱氢酶活性。

39. 权利要求 38 的细胞，其中所述细胞还具有脂酶或酯酶活性。

40. 权利要求 39 的细胞，其中所述细胞产生 3-HP 的酯。

41. 权利要求 40 的细胞，其中所述 3-HP 的酯是 3-羟丙酸甲酯，3-羟丙酸乙酯，3-羟丙酸丙酯，3-羟丙酸丁酯或 3-羟丙酸 2-乙基己基酯。

42. 权利要求 38 的细胞，其中所述细胞还含有聚羟酸合成酶活性。

43. 权利要求 42 的细胞，其中所述细胞产生聚合的 3-HP。

44. 权利要求 38 的细胞，其中所述细胞还包括编码具有醇脱氢酶活性的肽的核酸分子，编码具有醛脱氢酶活性的肽的核酸分子或二者皆有。

45. 权利要求 44 的细胞，其中所述细胞产生 1,3-丙二醇。

46. 权利要求 28 的细胞，其中所述细胞还含有：

α -酮泛解酸盐羟甲基转移酶活性；

α -酮泛解酸盐还原酶活性；和

泛酸盐合酶活性。

47. 权利要求 46 的细胞，其中所述细胞产生泛酸盐。

48. 权利要求 46 的细胞，其中所述细胞还包括：

泛酸盐激酶活性；

4'-磷酸泛酰基-L-半胱氨酸合成酶活性；

4'-磷酸泛酰半胱氨酸脱羧酶活性；

ATP:4'-磷酸泛酰巯基乙胺腺嘌呤基转移酶活性；和

脱磷酸-CoA 激酶活性。

49. 权利要求 48 的细胞，其中所述细胞产生辅酶 A(CoA)。

50. 含有至少一种外源核酸分子的转化细胞，其中至少有一种外源核酸分子含有编码权利要求 1 所述多肽的核酸序列。

51. 权利要求 50 的转化的细胞，其中所述细胞从 α -丙氨酸产生 β -丙氨酸。

52. 权利要求 51 的细胞，其中所述细胞产生 3-HP，1,3-丙二醇，泛酸盐，CoA 或其组合。

53. 制备含有丙氨酸 2,3-氨基变位酶活性的多肽的方法，其包括在允许细胞产生含有丙氨酸 2,3-氨基变位酶活性的多肽的条件下培养权利要求 28 所述的细胞。

54. 从 α -丙氨酸制备 β -丙氨酸的方法，该方法包括在允许细胞从 α -丙氨酸制备 β -丙氨酸的条件下培养权利要求 28 所述的细胞。

55. 权利要求 54 的方法，其中所述细胞含有至少一种编码丙氨酸 2,3-氨基变位酶的外源核酸分子，其中所述丙氨酸 2,3-氨基变位酶能从 α -丙氨酸产生 β -丙氨酸。

56. 权利要求 54 的方法，其中所述细胞是原核细胞。

57. 权利要求 56 的方法，其中所述细胞含有功能性缺失 panD。

58. 制备 3-HP 的方法，该方法包括在所述细胞产生 3-HP 的条件下培养权利要求 38 所述的细胞。

59. 权利要求 58 的方法，其中所述细胞含有至少一种编码丙氨酸 2,3-氨基变位酶从而使得从 β -丙氨酸产生 3-HP 的外源核酸，其中所述丙氨酸 2,3-氨基变位酶能从 α -丙氨酸产生 β -丙氨酸。

60. 制备 3-HP 的酯的方法，该方法包括在细胞产生 3-HP 的酯的条件下培养权利要求 39 所述的细胞。

61. 制备聚合的 3-HP 的方法，该方法包括在细胞产生聚合的 3-HP 的条件下培养权利要求 42 所述的细胞。

62. 制备 1,3-丙二醇的方法，该方法包括在细胞产生 1,3-丙二醇的条件下培养权利要求 44 所述的细胞。

63. 制备泛酸盐的方法，该方法包括在细胞产生泛酸盐的条件下培养权利要求 46 所述的细胞。

64. 制备 CoA 的方法，该方法包括在细胞产生 CoA 的条件下培养权利要求 48 所述的细胞。

65. 制备 3-HP 的方法，该方法包括：

从权利要求 28 所述的细胞中纯化 β -丙氨酸；

将所述 β -丙氨酸与含有 β -丙氨酸/2-酮戊二酸盐氨基转移酶活性的多肽接触从而形成 3-酮丙酸盐；且

将 3-酮丙酸盐与含有 3-羟丙酸盐脱氢酶活性的多肽相接触从而制

备 3-HP。

66. 制备 3-HP 的方法，该方法包括：

使用编码含有丙酮酸盐/2-酮戊二酸盐氨基转移酶活性多肽的核酸分子，使用编码含有 β -丙氨酸/2-酮戊二酸盐氨基转移酶活性多肽的核酸分子，和使用编码含有 3-羟丙酸盐脱氢酶活性多肽的核酸分子转染权利要求 28 所述的细胞；且

培养所转染的细胞从而使该转染的细胞制备 3-HP。

67. 从 3-HP 制备 1,3-丙二醇的方法，该方法包括：

使用权利要求 65 的方法制备 3-HP；

将所述 3-HP 与含有醇脱氢酶活性的多肽和含有醛脱氢酶活性的多肽接触从而制备 1,3-丙二醇。

68. 制备 1,3-丙二醇的方法，该方法包括：

使用编码含有丙酮酸盐/2-酮戊二酸盐氨基转移酶活性的多肽的核酸分子，使用编码含有 β -丙氨酸/2-酮戊二酸盐氨基转移酶活性的多肽的核酸分子，使用编码含有 3-羟丙酸盐脱氢酶活性的多肽的核酸分子，使用编码含有醛脱氢酶活性的多肽的核酸，和使用编码含有醇脱氢酶活性的多肽的核酸转染权利要求 28 所述的细胞；且

培养所转染的细胞从而允许该转染的细胞制备 1,3-丙二醇。

69. 制备泛酸盐的方法，该方法包括：

从权利要求 28 所述的细胞中纯化 β -丙氨酸；以及

将所述 β -丙氨酸与 α -酮泛解酸盐羟甲基转移酶， α -酮泛解酸盐还原酶和泛酸盐合酶相接触从而制备泛酸盐。

70. 制备泛酸盐的方法，该方法包括：

使用编码含有 α -酮泛解酸盐羟甲基转移酶活性的多肽的核酸分子，使用编码含有 α -酮泛解酸盐还原酶活性的多肽的核酸分子，和使

用编码含有泛酸盐合酶活性的多肽的核酸分子转染权利要求 28 所述的细胞；以及

培养所转染的细胞从而允许该转染的细胞制备泛酸盐。

71. 制备 CoA 的方法，该方法包括：

从权利要求 28 所述的细胞中纯化 β -丙氨酸；以及

将所述 β -丙氨酸与 α -酮泛解酸盐羟甲基转移酶， α -酮泛解酸盐还原酶和泛酸盐合酶相接触从而制备泛酸盐；以及

将所述泛酸盐与泛酸盐激酶，4'-磷酸泛酰基-1-半胱氨酸合成酶，4'-磷酸泛酰半胱氨酸脱羧酶，ATP:4'-磷酸泛酰巯基乙胺腺嘌呤基转移酶和脱磷酸-CoA 激酶接触从而制备 CoA。

72. 制备 CoA 的方法，该方法包括：

使用编码含有 α -酮泛解酸盐羟甲基转移酶活性的多肽的核酸分子，使用编码含有 α -酮泛解酸盐还原酶活性的多肽的核酸分子，和使用编码含有泛酸盐合酶活性的多肽的核酸分子，使用编码含有泛酸盐激酶活性的多肽的核酸分子，使用编码含有 4'-磷酸泛酰基-1-半胱氨酸合成酶活性的多肽的核酸分子，使用编码含有 4'-磷酸泛酰半胱氨酸脱羧酶活性的多肽的核酸分子，使用编码含有 ATP:4'-磷酸泛酰巯基乙胺腺嘌呤基转移酶活性的多肽的核酸分子，和使用编码含有脱磷酸-CoA 激酶活性的多肽的核酸分子转染权利要求 28 所述的细胞；且

培养所转染的细胞从而允许该转染的细胞制备泛酸盐

73. 特异性结合于权利要求 1 所述多肽的特异性结合剂。

丙氨酸 2,3 氨基变位酶

发明领域

本公开涉及丙氨酸 2,3-氨基变位酶的核酸和氨基酸序列，具有丙氨酸 2,3-氨基变位酶活性的细胞，所述丙氨酸 2,3-氨基变位酶能将 α -丙氨酸转化为 β -丙氨酸，以及使用这些细胞制备 β -丙氨酸，泛酸，3-羟丙酸和其他有机化合物的方法。

发明背景

可使用诸如有机酸、酯和聚醇的有机化学试剂来合成塑料材料和其他制品。为了满足日益增长的对有机化学试剂的需要，现正在开发利用基于碳水化合物类而非烃类原材料的更有效和更划算的制备方法。例如，某些细菌已被用来大量生产用于聚乳酸生产中的酸。

3-羟丙酸(3-HP)是一种有机酸。已记载有数种化学合成途径来制备 3-HP，且其生物催化途径也已公开(Suthers 等人的 WO 01/16346)。3-HP 具有特异性合成的用途，且可通过化工领域中的公知技术转化为具有重要商业价值的中间体，例如，通过脱水转化为丙烯酸，通过氧化转化为丙二酸，通过与醇的酯化反应转化为酯，以及通过还原转化为 1,3-丙二醇。

发明内容

可从 PEP 或丙酮酸盐，通过关键的 β -丙氨酸中间体(图 1)生物催化产生化合物 3-羟丙酸(3-HP)。在细胞中可通过 5,6-二氢尿嘧啶和 N-氨基甲酰- β -丙氨酸，N-乙酰- β -丙氨酸，鹅肌肽或天冬氨酸从肌肽， β -丙氨酰精氨酸， β -丙氨酰赖氨酸，尿嘧啶合成 β -丙氨酸(图 1 和图 2)。然而，这些途径相对而言效率较低，因为它们都需要价值高于 3-HP 的较为难得的前体或起始化合物。因此，如果 α -丙氨酸可以被直接转化为 β -

丙氨酸的话(图 1), 那么采用生物催化途径产生 3-HP 的效率会更高。

本说明书中公开了具有丙氨酸 2,3 氨基变位酶生物学活性的新型突变的赖氨酸 2,3 氨基变位酶核酸和蛋白质序列。在一个实施例中, 突变的赖氨酸 2,3 氨基变位酶包括一个或多个以下取代: P/S11T; N19Y; L/K/R/T26I; E/R30K; L/V32A; K36E; S/T/C52R; L/T53P/H; Y63F; E/N/D71G; H/I/S85Q; Q/L/E86R; Q/L95M; K/M/Q125L; M128V; Y132H; Q/S141R; A/D/S/M144G; D179N; K/Q187R; I192V; L228M; D331G/H; M/Q342T; 或 K/Q/T398E, 其中数字之前的字母代表在赖氨酸 2,3 氨基变位酶中所发现的氨基酸的单字母氨基酸编码, 数字代表氨基酸位置(基于牙龈卟啉单胞菌赖氨酸 2,3 氨基变位酶(SEQ ID NO: 52 的编码), 参见图 7), 而数字之后的字母代表在丙氨酸 2,3 氨基变位酶中所发现的氨基酸的单字母氨基酸编码。本领域所属技术人员应该理解, 根据需要突变的赖氨酸 2,3 氨基变位酶序列, 实际上的第一个氨基酸和氨基酸数字会发生改变, 而且, 可以通过序列比对确定同源序列中的所述位置。例如, 如图 7 所示, 牙龈卟啉单胞菌赖氨酸 2,3 氨基变位酶的位置 11 对应于具核梭杆菌赖氨酸 2,3 氨基变位酶(SEQ ID NO: 10)的位置 12, 斯氏梭菌赖氨酸 2,3 氨基变位酶(SEQ ID NO: 33)的位置 9, 和枯草芽孢杆菌赖氨酸 2,3 氨基变位酶(SEQ ID NO: 59)的位置 8。可根据图 7 所提供的信息和其他公众可获得的赖氨酸 2,3 氨基变位酶序列(示例包括, 但不限于, 炭疽杆菌(*Bacillus anthracis*) str. Sterne 的 GenBank Accession Nos. YP_028406; 创伤弧菌(*Vibrio vulnificus*)YJ016 的 BAC95867; *Moorella thermoacetica* 的 ZP_00330962 ; 流感嗜血杆菌(*Haemophilus influenzae*)的 ZP_00322274 ; *Methanosarcina barkeri* str. Fusaro 的 ZP_00298043; *Microbulbifer degradans* 的 ZP_00314982 和破伤风梭菌(*Clostridium tetani*)E88 的 NP_781545), 使用本领域中的公知方法对每个剩余的位点进行类似的分析。

使用标准的分子生物学方法可对任意的赖氨酸 2,3 氨基变位酶进行突变产生丙氨酸 2,3-氨基变位酶。例如, 可对来自于诸如芽孢杆菌,

梭菌, 埃希氏菌, 梭形杆菌, 嗜血杆菌, 甲烷八叠球菌, Microbulbifer, Moorella, 卟啉单胞菌, 嗜热厌氧菌或弧菌的原核生物的赖氨酸 2,3 氨基变位酶进行突变, 从而包含有一种或多种以下取代, P/S11T; N19Y; L/K/R/T26I; E/R30K; L/V32A; K36E; S/T/C52R; L/T53P/H; Y63F; E/N/D71G; H/I/S85Q; Q/L/E86R; Q/L95M; K/M/Q125L; M128V; Y132H; Q/S141R; A/D/S/M144G; D179N; K/Q187R; I192V; L228M; D331G/H; M/Q342T 或 K/Q/T398E, 例如至少 2 种, 至少 3 种, 至少 4 种, 至少 5 种, 至少 6 种, 至少 8 种, 至少 9 种, 至少 10 种, 至少 11 种, 至少 12 种, 至少 13 种, 至少 14 种, 至少 15 种, 至少 16 种, 至少 17 种, 至少 18 种, 至少 19 种, 至少 20 种, 至少 21 种, 至少 22 种, 至少 23 种, 至少 24 种或不多于 3 种, 不多于 4 种, 不多于 5 种, 不多于 6 种, 不多于 8 种, 不多于 9 种, 不多于 10 种, 不多于 11 种, 不多于 12 种, 不多于 13 种, 不多于 14 种, 不多于 15 种, 不多于 16 种, 不多于 17 种, 不多于 18 种, 不多于 19 种, 不多于 20 种, 不多于 21 种, 不多于 22 种或不多于 23 种这样的突变。这些取代的特定组合包括, 但不限于: (1) N19Y, L/T53P/H, H/I/S85Q, D331G/H, 和 M/Q342T; (2) N19Y, E/R30K, L/T53P/H, H/I/S85Q, I192V, D331G/H, 和 M/Q342T; (3) N19Y, L/K/R/T26I; E/R30K, L/T53P/H, H/I/S85Q, I192V, D331G/H, 和 M/Q342T; (4) E/R30K, Y63F, Q/L/E86R, Q/L95M, M128V, A/D/S/M144G, L228M, D331G/H, 和 K/Q/T398E; (5) E/R30K, K36E, Y63F, Q/L/E86R, Q/L95M, M128V, A/D/S/M144G, D179N, L228M, D331G/H, 和 K/Q/T398E; (6) E/R30K, Q/L95M, M128V, 和 D331G/H; (7) P/S11T, E/R30K, Q/L95M, M128V, Q/S141R, K/Q187R, 和 D331G/H; (8) E/R30K, L/V32A, L/T53P/H, E/N/D71G, Q/L95M; K/M/Q125L, M128V, 和 D331G/H; (9) E/R30K, C52R, Q/L95M; M128V, 和 D331G/H; (10) Q/L95M, M128V, 和 D331G/H; (11) Q/L95M, M128V; Y132H, 和 D331G/H; 和(12) Q/L95M, M128V, 和 D331G/H。

丙氨酸 2,3-氨基变位酶分子的特别示例包括如 SEQ ID NOS: 18, 20, 42, 44, 46, 48 和 50 所示的核酸序列, SEQ ID NOS: 19, 21, 43,

45, 47, 49 和 51 所示的与其对应的氨基酸序列, 以及保留了能将 α -丙氨酸转变为 β -丙氨酸的这些序列的变体、片段、融合体和多态形。所公开的丙氨酸 2,3-氨基变位酶序列可用于转化细胞, 从而使得所转化的细胞具有丙氨酸 2,3-氨基变位酶活性, 能使所述细胞从 α -丙氨酸生成 β -丙氨酸。本公开内容还包括了能够特异性结合于丙氨酸 2,3-氨基变位酶的结合剂。

本发明还公开了具有丙氨酸 2,3-氨基变位酶活性的细胞, 所述丙氨酸 2,3-氨基变位酶能将 α -丙氨酸转化为 β -丙氨酸。这样的细胞可以是真核或原核细胞, 例如酵母细胞, 植物细胞, 乳杆菌属, 乳球菌属, 芽胞杆菌属, 或埃希氏细胞。在一个实施例中, 可使用能够向转化细胞传递丙氨酸 2,3-氨基变位酶活性的突变赖氨酸 2,3-氨基变位酶来转化细胞。在另一个实施例中, 可使用 SEQ ID NO: 18, 20, 42, 44, 46, 48 或 50(或保留了丙氨酸 2,3-氨基变位酶活性的所述序列片段, 融合体或变体)来转化细胞。所公开的细胞可用来产生核酸分子, 肽和有机化合物。所述肽可用来催化有机化合物的形成, 或可用作抗原来构建特异性的结合剂。

本发明还公开了制备细胞, 其含有至少一种外源核酸, 例如编码丙氨酸 2,3-氨基变位酶的核酸。在一个实施例中, 所述核酸序列包括 SEQ ID NO: 8, 20, 42, 44, 46, 48 或 50 (或保留了丙氨酸 2,3-氨基变位酶活性的片段, 变体或融合体)。在另一个实施例中, 所述核酸序列编码了如 SEQ ID NO: 19, 21, 43, 45, 47, 49 或 51 (或保留了丙氨酸 2,3-氨基变位酶活性的片段, 变体或融合蛋白质)所示的氨基酸序列。制备细胞可用来表达多肽, 所述多肽具有诸如丙酮酸盐/2-酮戊二酸盐氨基转移酶、 β -丙氨酸/2-酮戊二酸盐氨基转移酶和 3-羟丙酸盐脱氢酶的酶活性, 其能够从 3-氧丙酸盐制备 3-羟丙酸盐。本发明还公开了制备由上述核酸序列所编码的多肽的方法。

本发明还公开了鉴定具有丙氨酸 2,3-氨基变位酶活性的细胞的方法

法。在一个实施例中，所述方法包括在既不含有 β -丙氨酸也不含有泛酸盐的培养基中培养功能性缺失 panD 的细胞。例如，所述细胞可以从碳源、氧源、氢源和氮源培养基中制备 α -丙氨酸，而不含有 β -丙氨酸。可对能够在所述培养基中生长的细胞进行鉴定，其中细胞生长说明了该细胞能够从 α -丙氨酸制备 β -丙氨酸，这说明了所述细胞具有丙氨酸 2,3-氨基变位酶活性。相反，未观察到细胞生长说明该细胞不能从 α -丙氨酸制备 β -丙氨酸，这说明了所述细胞不具有丙氨酸 2,3-氨基变位酶活性。在一个实施例中，在对培养细胞进行选择之前，可以使用一种或多种突变的赖氨酸 2,3-氨基变位酶转化细胞。

本发明还公开了制备具有丙氨酸 2,3-氨基变位酶活性的肽的方法。在一个实施例中，该方法包括培养具有至少一种外源核酸分子的细胞，所述核酸分子编码能够从 α -丙氨酸制备 β -丙氨酸的丙氨酸 2,3-氨基变位酶(例如 SEQ ID NO: 8, 20, 42, 44, 46, 48 或 50, 或保留了丙氨酸 2,3-氨基变位酶活性的片段，变体和融合体)。

在此还公开了使用具有丙氨酸 2,3-氨基变位酶活性的公开细胞从 β -丙氨酸制备 3-HP 的方法。在一个实施例中，可使用一种或多种从 β -丙氨酸转化为 3-HP 必须的酶来转染所述细胞。在另一个实施例中，所述方法包括从所述细胞中纯化 β -丙氨酸，然后将所述 β -丙氨酸与将 β -丙氨酸转化为 3-HP 所必须的多肽相接触。

在此所公开的所述细胞，丙氨酸 2,3-氨基变位酶核酸和氨基酸序列(例如 SEQ ID NO:8, 20, 42, 44, 46, 48 或 50 或保留了丙氨酸 2,3-氨基变位酶活性的片段，变体和融合体)，及方法都可用于制备泛酸盐，3-HP 及其衍生物，例如辅酶 A(CoA)，以及其他有机物，例如，1,3-丙二醇，丙烯酸，聚丙烯酸酯，丙烯酸酯的酯，聚合 3-HP，3-HP 的共聚物和其他化合物，例如丁酸酯，戊酸酯和其他化合物，3-HP 的酯，及丙二酸及其酯。3-HP 在生物学和商业上都是很重要的。例如，营养产业可将 3-HP 用于食品，饲料添加剂或防腐剂，而以上提到的所述衍生

物也可从 3-HP 进行制备。

编码丙氨酸 2,3-氨基变位酶(例如 SEQ ID NO:8, 20, 42, 44, 46, 48 或 50 或保留了丙氨酸 2,3-氨基变位酶活性的片段, 变体或融合体)的核酸分子可用于对细胞进行操作, 从而使其具有制备 3-HP 的能力, 以及制备其他有机化合物, 例如上述所列的化合物的能力。丙氨酸 2,3-氨基变位酶多肽(例如 SEQ ID NOS: 19, 21, 43, 45, 47, 49 和 51, 以及保留了具有将 α -丙氨酸互变为 β -丙氨酸能力的这些序列的变体, 片段或融合体)可用于无细胞的系统来生产 3-HP 和其他有机化合物, 如上述所列。在此所公开的细胞可用于培养系统从而制备大量 3-HP, 以及其他如上所述的有机化合物。

本发明的一方面提供了这样的细胞, 除了具有丙氨酸 2,3-氨基变位酶活性, 所述细胞还包括其他的酶活性, 例如丙酮酸盐/2-酮戊二酸盐氨基转移酶活性、 β -丙氨酸/2-酮戊二酸盐氨基转移酶活性和 3-羟丙酸盐脱氢酶的酶活性。此外, 这些细胞还可包括聚羟酸合成酶活性或脂酶或酯酶活性。

在另一个实施例中, 含有丙氨酸 2,3-氨基变位酶活性、丙酮酸盐/2-酮戊二酸盐氨基转移酶活性、 β -丙氨酸/2-酮戊二酸盐氨基转移酶活性和 3-羟丙酸盐脱氢酶活性的细胞可以制备这样的产物, 例如, 3-HP 或 3-HP 的酯, 例如 3-羟丙酸甲基酯, 3-羟丙酸乙基酯, 3-羟丙酸丙基酯或 3-羟丙酸丁基酯。因此, 本公开还提供了一种或多种这些产物的制备方法。在一些实施例中, 本方法包括在允许产生所述产物的条件下培养包括丙氨酸 2,3-氨基变位酶活性、丙酮酸盐/2-酮戊二酸盐氨基转移酶活性、 β -丙氨酸/2-酮戊二酸盐氨基转移酶活性和 3-羟丙酸盐脱氢酶活性的细胞。这些细胞还可包括脂酶或酯酶活性。

本发明的另一方面提供了细胞, 除了具有丙氨酸 2,3-氨基变位酶活性之外, 其具有丙酮酸盐/2-酮戊二酸盐氨基转移酶活性、 β -丙氨酸/2-

酮戊二酸盐氨基转移酶活性, 3-羟丙酸盐脱氢酶活性, 以及聚羟酸合成酶活性。这样的细胞可用来, 例如, 产生诸如聚合 3-HP 和 3-HP 共聚物的产物, 以及其他的化合物, 例如丁酸酯, 戊酸酯和其他化合物。

本发明还公开了产生 1,3-丙二醇的细胞及其使用方法。可使用具有酶活性的多肽从 3-HP 制备 1,3-丙二醇。当从 3-HP 制备 1,3-丙二醇时, 可以使用具有乙醛脱氢酶活性(例如来自 1.2.1.-类的酶)的多肽和具有乙醇脱氢酶活性(例如来自 1.1.1.-类的酶)的多肽的组合, 例如醛脱氢酶(NAD(P)+) (EC 1.2.1.-)和醇脱氢酶(EC 1.1.1.1)。

在一些实施例中, 可在体外(细胞外)生成产物。在其他实施例中, 可使用体外和体内(细胞内)方法的组合制备产物。在又一种实施例中, 可在体内制备产物。对于涉及体内步骤的方法, 所述细胞可从培养细胞或整个生物体, 例如转基因植物, 非人类的哺乳动物, 或单细胞生物体, 例如酵母和细菌(例如乳杆菌属, 乳球菌属, 芽孢杆菌属和埃希氏肠杆菌细胞)中分离得到。在下文中, 这些细胞这称为制备细胞。由这些制备细胞制备的产物可以是有机物, 例如 β -丙氨酸, 3-HP, 泛酸盐, 及其衍生物, 例如有机酸, 聚醇(例如 1,3-丙二醇), 辅酶 A(CoA), 以及在此所述的丙氨酸 2,3-氨基变位酶。

泛酸盐是对许多动物的生长和健康具有重要作用的维生素, 参与脂肪酸的合成和分解。维生素的缺乏会导致常见的临床不适。因此, 可以治疗有效量对具有泛酸缺乏的患者给药使用本公开方法制备的泛酸盐。本发明公开了生成泛酸盐的细胞和使用所公开的细胞从 β -丙氨酸制备泛酸盐的方法。用于制备泛酸盐和/或 CoA 的制备细胞可用来表达 α -酮泛解酸盐羟甲基转移酶(E.C.2.1.2.11), α -酮泛解酸盐还原酶(E.C.1.1.1.169), 和泛酸盐合酶(E.C.6.3.2.1), 从而制备泛酸盐或另外的泛酸盐激酶(E.C.2.7.1.33), 4'-磷酸羟泛酰基-L-半胱氨酸合成酶(E.C.6.3.2.5), 4'-磷酸泛酰半胱氨酸脱羧酶(E.C.4.1.1.36), ATP:4'-磷酸泛酰巯基乙胺腺嘌呤基转移酶(E.C.2.7.7.3), 和脱磷酸-CoA 激酶(E.C.

2.7.1.24), 从而制备辅酶 A。

附图说明

图 1 所示为通过 β -丙氨酸中间体生成 3-HP 及其衍生物的途径示意图, 以及从 α -丙氨酸制备 β -丙氨酸的途径示意图。

图 2 所示为公知的代谢途径, 通过该途径 β -丙氨酸得以产生和消耗。

图 3 所示为通过天冬氨酸盐脱羧酶(panD)或通过丙氨酸 2,3-氨基变位酶生成辅酶 A 和泛酸盐的替换途径。

图 4 所示为对获得了丙氨酸 2,3 氨基变位酶活性蛋白质, 来自具核梭杆菌(Fnkam; SEQ ID NO: 10)和两种变体序列(Fnaam, SEQ ID NO: 19; 和 Fnaam2, SEQ ID NO: 21)的赖氨酸 2,3-氨基变位酶蛋白质序列进行的比对。在赖氨酸 2,3 氨基变位酶蛋白质序列中的取代用粗体标示。

图 5 所示为对获得了具有丙氨酸 2,3 氨基变位酶活性的蛋白质, 来自斯氏梭菌(Cskam; SEQ ID NO: 33)的赖氨酸 2,3-氨基变位酶蛋白质序列, 来自具有 E28K, L93M, M126V 和 D329H 取代的赖氨酸 2,3-氨基变位酶蛋白质序列(Cscodm; SEQ ID NO: 41), 和三种变体序列(Cscodm mut8, SEQ ID NO: 43; Cscodm mut12, SEQ ID NO: 45; 和 Cscodm mut15, SEQ ID NO: 47)进行的比对。在赖氨酸 2,3-氨基变位酶蛋白质序列中的取代用粗体标示。

图 6 所示为对获得了具有丙氨酸 2,3 氨基变位酶活性的蛋白质, 来自牙龈卟啉单胞菌的赖氨酸 2,3-氨基变位酶蛋白质序列(Pgkam; SEQ ID NO: 52), 和三种变体序列(Pgaam SEQ ID NO: 6; Pgaam2, SEQ ID NO: 49; 和 Pgaam2 L26I, SEQ ID NO: 51)的蛋白质序列进行的比对。在赖氨酸 2,3 氨基变位酶蛋白质序列中的取代用粗体标示。

图 7A-D 所示为对来自赖氨酸 2,3-氨基变位酶(来自牙龈卟啉单胞菌(Pgkam, SEQ ID NO: 52); 具核梭杆菌(Fnkam; SEQ ID NO: 10); 斯氏梭菌(Cskam; SEQ ID NO: 33)和枯草芽孢杆菌(Bskam; SEQ ID NO: 59)) 以及丙氨酸 2,3-氨基变位酶(来自牙龈卟啉单胞菌(Pgaam SEQ ID

NO: 6; Pgaam2, SEQ ID NO: 49; 和 Pgaam2 L26I, SEQ ID NO: 51); 具核梭杆菌(Fnaam, SEQ ID NO: 19; 和 Fnaam2, SEQ ID NO: 21); 斯氏梭菌(Cscodm mut8, SEQ ID NO: 43; Cscodm mut12, SEQ ID NO: 45; 和 Cscodm mut15, SEQ ID NO: 47)和枯草芽孢杆菌(Bsaam, SEQ ID NO: 2 and Bsaam2co, SEQ ID NO: 4))以及来自具核梭杆菌(Fncodm, SEQ ID NO: 14)和斯氏梭菌(Cscodm, SEQ ID NO: 41)的中间体序列进行的比对。以粗体所示的残基是指具有丙氨酸 2,3-氨基变位酶活性的赖氨酸 2,3-氨基变位酶变体中的突变。

序列表

后附序列表中所列示的核酸和氨基酸序列使用的都是针对核苷酸碱基的标准单字母缩写, 和针对氨基酸的三字母编码。每种核酸序列都仅表示出一条链, 但可根据所示出链的任意参考得出所述互补链。

SEQ ID NO: 1 是来自枯草芽孢杆菌的丙氨酸 2,3 氨基变位酶核酸序列。

SEQ ID NO: 2 是由 SEQ ID NO: 1 编码的蛋白质序列。

SEQ ID NO: 3 是来自枯草芽孢杆菌的丙氨酸 2,3 氨基变位酶核酸序列。

SEQ ID NO: 4 是由 SEQ ID NO: 3 编码的蛋白质序列。

SEQ ID NO: 5 是来自牙龈卟啉单胞菌的丙氨酸 2,3 氨基变位酶核酸序列。

SEQ ID NO: 6 是由 SEQ ID NO: 5 编码的蛋白质序列。

SEQ ID NOS: 7 和 8 是用于克隆具核梭杆菌赖氨酸 2,3-氨基变位酶的核酸 PCR 引物。

SEQ ID NO: 9 是来自具核梭杆菌的赖氨酸 2,3-氨基变位酶核酸序列。

SEQ ID NO:10 是由 SEQ ID NO: 9 编码的蛋白质序列。

SEQ ID NO:11 是来自具核梭杆菌的部分密码子优化的赖氨酸 2,3-氨基变位酶核酸序列。

SEQ ID NO:12 是由 SEQ ID NO: 11 编码的蛋白质序列。

SEQ ID NO:13 是来自具核梭杆菌的突变的部分密码子优化的赖氨酸 2,3-氨基变位酶核酸序列。

SEQ ID NO:14 是由 SEQ ID NO:13 编码的蛋白质序列。

SEQ ID NOS: 15-17 是用于对来自具核梭杆菌的赖氨酸 2,3-氨基变位酶核酸序列进行突变的核酸引物。

SEQ ID NO:18 是来自具核梭杆菌的丙氨酸 2,3 氨基变位酶核酸序列。

SEQ ID NO:19 是由 SEQ ID NO:18 编码的蛋白质序列。

SEQ ID NO:20 是来自具核梭杆菌的丙氨酸 2,3 氨基变位酶核酸序列。

SEQ ID NO:21 是由 SEQ ID NO:20 编码的蛋白质序列。

SEQ ID NOS: 22-27 是用于克隆来自斯氏梭菌的赖氨酸 2,3-氨基变位酶核酸序列的核酸 PCR 引物。

SEQ ID NOS: 28-31 是用于基因组步移从斯氏梭菌克隆赖氨酸 2,3-氨基变位酶核酸序列的核酸引物。

SEQ ID NO:32 是来自斯氏梭菌的赖氨酸 2,3-氨基变位酶核酸序列。

SEQ ID NO:33 是由 SEQ ID NO:32 编码的蛋白质序列。

SEQ ID NO:34 是来自斯氏梭菌的部分密码子优化的赖氨酸 2,3-氨基变位酶核酸序列。

SEQ ID NO:35 是由 SEQ ID NO:34 编码的蛋白质序列。

SEQ ID NOS: 36-39 是用来对来自斯氏梭菌的部分密码子优化的赖氨酸 2,3-氨基变位酶核酸序列进行突变的核酸引物。

SEQ ID NO:40 是来自斯氏梭菌的突变的赖氨酸 2,3-氨基变位酶核酸序列。

SEQ ID NO:41 是由 SEQ ID NO:40 编码的蛋白质序列。

SEQ ID NO:42 是来自斯氏梭菌的丙氨酸 2,3-氨基变位酶核酸序列。

SEQ ID NO:43 是由 SEQ ID NO:42 编码的蛋白质序列。

SEQ ID NO:44 是来自斯氏梭菌的丙氨酸 2,3-氨基变位酶核酸序列。

SEQ ID NO:45 是由 SEQ ID NO:44 编码的蛋白质序列。

SEQ ID NO:46 是来自斯氏梭菌的丙氨酸 2,3-氨基变位酶核酸序列。

SEQ ID NO:47 是由 SEQ ID NO:46 编码的蛋白质序列。

SEQ ID NO:48 是来自牙龈卟啉单胞菌的丙氨酸 2,3-氨基变位酶核酸序列。

SEQ ID NO:49 是由 SEQ ID NO:48 编码的蛋白质序列。

SEQ ID NO:50 是来自牙龈卟啉单胞菌的丙氨酸 2,3-氨基变位酶核酸序列。

SEQ ID NO:51 是由 SEQ ID NO:50 编码的蛋白质序列。

SEQ ID NO:52 是来自牙龈卟啉单胞菌的赖氨酸 2,3-氨基变位酶核酸序列。

SEQ ID NOS: 53 和 54 是用于扩增 pKD3 的 CAT 基因的 PCR 引物。

SEQ ID NOS: 55 和 56 是用于确认所述 CAT 基因正确插入 panD 座的 PCR 引物。

SEQ ID NOS: 57 和 58 是用于扩增 pKD3 的 CAT 基因的引物的核酸序列。

SEQ ID NO: 59 是来自枯草芽孢杆菌的赖氨酸 2,3-氨基变位酶蛋白质序列。

若干实施方案的详细描述

以下对术语和方法的解释旨在对本公开进行更详细地描述，同时指导那些本领域所属技术人员实施本发明。除非文中特别指出，否则单数形式“a”，“an”和“the”是指一个或多于一个的情况。例如，术语“含有核酸分子”包括了单种或多种核酸分子，并相当于短语“包括至少一种核酸分子”。除非文中特别指出，否则术语“或”是指规定可选元素中的单种元素或两种或更多种元素的组合。在本文中，“包含”是指“包括”。因此，“包含 A 或 B”是指“包括 A，B 或 A 和

B”，且未将其他元素排除在外。

除非另有解释，否则在此使用的所有技术和科学术语都与本发明归属领域技术人员通常理解的意思相同。尽管可采用与在此描述相类似的方法和材料来实施或检测本发明，但下文描述了适合的方法和材料。所述材料，方法和实施例仅仅是说明性而非限制性的。根据以下的详细描述和权利要求，本发明的其他特征和优点是显而易见的。

丙氨酸 2,3-氨基变位酶：例如在细胞中，能够将 α -丙氨酸转化为 β -丙氨酸，或反之的酶。其包括了来自诸如原核生物的任何生物体的任意丙氨酸 2,3-氨基变位酶基因，cDNA，RNA 或蛋白质。在一个实施例中，丙氨酸 2,3-氨基变位酶是具有丙氨酸 2,3-氨基变位酶活性的突变的赖氨酸 2,3-氨基变位酶。可从任意生物体，例如原核生物，例如枯草芽孢杆菌(*Bacillus subtilis*)，牙龈卟啉单胞菌(*Porphyromonas gingivalis*)，具核梭杆菌(*Fusobacterium nucleatum*)，斯氏梭菌(*Clostridium sticklandii*) 或大肠杆菌 (*Escherichia coli*) 使用本领域公知的任意方法获得赖氨酸 2,3-氨基变位酶(或在遗传数据库中被标注为赖氨酸 2,3 氨基变位酶的基因)并对其进行突变。

在特定的实施例中，丙氨酸 2,3-氨基变位酶核酸序列包括如 SEQ ID NO: 1, 3, 5, 18, 20, 42, 44, 46, 48 或 50 所示的序列，以及保留了编码具有丙氨酸 2,3-氨基变位酶活性的肽或蛋白质的能力的片段，变体或融合体所示的序列。在其他实施例中，丙氨酸 2,3-氨基变位酶蛋白质包括 SEQ ID NO: 2, 4, 6, 19, 21, 43, 45, 47, 49 或 51 所示的氨基酸序列，以及保留了丙氨酸 2,3-氨基变位酶活性的序列片段，变体或融合体所示的氨基酸序列。

在其他实施例中，丙氨酸 2,3-氨基变位酶序列包括全长的野生型序列，例如 SEQ ID NO: 2, 4, 6, 19, 21, 43, 45, 47, 49 或 51，以及保留了将 α -丙氨酸转化为 β -丙氨酸能力的短序列，例如至少 9 个连

续的氨基酸(例如, SEQ ID NO: 2, 4, 6, 19, 21, 43, 45, 47, 49 或 51 的至少 10 个, 至少 11 个, 至少 12 个, 至少 13 个, 至少 14 个, 至少 15 个, 至少 20 个, 至少 25 个, 至少 30 个, 至少 35 个, 至少 40 个, 至少 45 个, 至少 50 个, 至少 55 个, 至少 60 个, 至少 70 个, 至少 80 个, 至少 90 个, 至少 100 个, 至少 150 个, 至少 200 个, 至少 250 个, 至少 300 个, 至少 350 个, 至少 400 个, 至少 410 个或至少 450 个连续的氨基酸。丙氨酸 2,3-氨基变位酶序列的特定片段包括, 但不限于 SEQ ID NO: 2 或 4 的氨基酸 50-390, SEQ ID NO: 2 或 4 的氨基酸 101-339, SEQ ID NO: 2 或 4 的氨基酸 15-390, 和 SEQ ID NO: 2 或 4 的氨基酸 15-340(SEQ ID NOS: 6, 19, 21, 43, 45, 47, 49 和 51 (例如 SEQ ID NO: 6, 49 或 51 的氨基酸 42-342)中的对应片段也包括在本发明的范围之内, 并可使用图 7 来进行确定)。只要所得多肽具有丙氨酸 2,3-氨基变位酶活性, 这些片段(或全长多肽)就可融合于其他多肽序列。本说明书包括了丙氨酸 2,3-氨基变位酶等位基因变体, 以及保留了将 α -丙氨酸转化为 β -丙氨酸能力的任意变体, 片段或融合序列。

丙氨酸 2,3-氨基变位酶活性: 丙氨酸 2,3-氨基变位酶的将 α -丙氨酸转化为 β -丙氨酸或反之的能力。在一个实施例中, 这样的活性存在于细胞中。在另一个实施例中, 这样的活性存在于体外。可以使用任意公职的分析方法, 例如以下实施例中所述的筛选分析和酶分析方法来测定这样的活性。此外, 可通过将所述酶与 α -丙氨酸或 β -丙氨酸孵育来鉴定丙氨酸 2,3-氨基变位酶的活性, 并且通过高效液相色谱(例如, Abe 等人 *J. Chromatography B*, 712:43-9, 1998 所述的方法)来确定所述反应产物。在一个实施例中, 在功能性缺失了 panD 基因的大肠杆菌突变体中是将 α -丙氨酸转化为 β -丙氨酸的能力。

在一个实施例中, 突变赖氨酸 2,3 氨基变位酶具有如 SEQ ID NO: 51 所示肽序列的丙氨酸 2,3 氨基变位酶活性的至少 10%, 至少 20%, 至少 50%, 至少 60%, 至少 70%, 至少 80%, 至少 90%或甚至至少 100%。

抗体：包括能够特异性结合于抗原(与抗原进行免疫反应)的抗原结合位点的分子。其示例包括多克隆抗体，单克隆抗体，人源化单克隆抗体，或其免疫有效部分。

本发明还包括了免疫球蛋白分子及其免疫活性部分。天然发生的抗体(例如 IgG)包括四条多肽链，由二硫键交联的两条重链(H)和两条轻链(L)。然而，也可通过自然发生抗体的部分来实施抗体的抗原结合功能。单克隆抗体的免疫有效部分包括，但不限于：Fab, Fab', F(ab')₂, Fabc 和 Fv 部分(其综述可参见 Better 和 Horowitz, *Methods. Enzymol.* 1989, 178:476-96)。抗原结合片段的其他示例包括，但不限于：(i) 由 VL, VH, CL 和 CH1 结构域组成的 Fab 片段；(ii) 由 VH 和 CH1 结构域组成的 Fd 片段；(iii) 由抗体单臂的 VL 和 VH 结构域组成的 Fv 片段；(iv) 由 VH 结构域组成的 dAb 片段；(v) 分离的互补决定区(CDR)；和 (vi) F(ab')₂ 片段，含有在铰链区由二硫键桥接的两个 Fab 片段的二价片段。此外，尽管所述 Fv 片段的两个结构域是由不同的基因编码的，但可通过重组方法使用合成连接物将其制备成单条蛋白质链(也称为单链 Fv(scFv))。本发明也包括了这样的单链抗体。

“特异性结合”是指特异性制剂(“特异结合剂”)与特定被分析物特异性反应，例如，与抗体特异性免疫反应，或与特定的肽序列特异性结合的能力。所述结合是非随机的结合反应，例如在抗体分子和抗原决定簇之间的反应。抗体的结合特异性通常是根椐该抗体区别结合特异性抗原和不相关抗原能力的参考点确定的，并因此能够在两种不同的抗原之间进行区别，特别是当这两种抗原具有独特的表位时。特异性结合于特定表位的抗体称为“特异性抗体”。

本发明还包括了可制备成丙氨酸 2,3-氨基变位酶多肽(例如 SEQ ID NO: 19, 21, 43, 45, 47, 49 或 51)，丙氨酸 2,3-氨基变位酶多肽片段(可由图 7 确定的片段，例如 SEQ ID NO: 2 或 4 的氨基酸 50-390，例如 SEQ ID NO: 2 或 4 的氨基酸 101-339 或 SEQ ID NO: 2 或 4 的氨基

酸 15-390, 例如 SEQ ID NO: 2 或 4 的氨基酸 15-331; 或 SEQ ID NOS: 6, 19, 21, 43, 45, 47, 49, 和 51 (例如 SEQ ID NO: 6, 49 或 51 的氨基酸 42-342)的对应片段)或所述序列的变体或融合体的单克隆或多克隆抗体。任意的, 对多肽抗原上的一个或多个表位产生反应的抗体可以特异性检测所述多肽。换言之, 针对一个特定多肽的抗体可以识别并结合于所述特定的多肽, 并几乎不会识别或结合于其他多肽。可以通过任意的标准免疫分析方法确定特异性结合于特定多肽的抗体; 例如, Western 印迹(例如, 参见 Sambrook 等人 (ed.), *Molecular Cloning: A Laboratory Manual*, 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989)。

为了通过 Western 印迹能确定抗体制剂(例如, 在小鼠中产生的针对丙氨酸 2,3-氨基变位酶多肽的制剂, 例如 SEQ ID NO: 19, 21, 43, 45, 47, 49 或 51)特异性地检测到适当的多肽(例如丙氨酸 2,3-氨基变位酶), 可从细胞中提取总细胞蛋白质并通过 SDS-聚丙烯酰胺凝胶电泳进行分离。可将所分离的总细胞蛋白质随后转移至膜上(例如硝酸纤维素), 然后将所述抗体制剂与所述膜共同孵育。在对膜进行洗涤去除非特异性结合的抗体之后, 可使用耦于诸如碱性磷酸酶的酶的适当第二抗体(例如, 抗小鼠抗体)来测定特异性结合的抗体的存在, 因为 5-溴-4-氯-3-吲哚磷酸/硝基蓝四唑的应用可导致由免疫固定碱性磷酸酶产生的深蓝色化合物。

可以从转染细胞, 转化细胞或野生型细胞中获得适于用作免疫原的基本上纯的多肽。可对最终制剂中的多肽浓度进行调节, 例如, 通过 Amicon 过滤装置对浓度进行调节, 达到每微升几微克的水平。此外, 从全长多肽直至少至 9 个氨基酸残基多肽的多肽都可用作免疫原。这样的多肽可从细胞培养物中制备, 也可以使用标准方法化学合成, 或可通过将大的多肽剪切成可纯化的较小多肽来进行制备。在存在有主要组织相容性复合物(MHC)分子, 例如 MHC I 类或 MHC II 类分子的情况下, 当长度低至 9 个氨基酸残基的多肽向免疫系统进行呈递时, 所

述多肽是具有免疫原性的。因此，具有至少 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 410, 450, 500 或更多个丙氨酸 2,3-氨基变位酶多肽的连续氨基酸残基的肽可用作制备抗体的免疫原。

根据 Kohler & Milstein (Nature 256:495, 1975)的经典方法或其衍生方法，可从鼠杂交瘤中制备出针对于任意在此所公开多肽的单克隆抗体。

可通过用所述多肽(或其片段，融合物或变体)免疫适当的动物制备出在此公开的含有针对于任意多肽的异源表位的抗体的多克隆抗血清，可以对其不进行修饰或进行修饰以提高免疫原性。适用于兔子的有效免疫方案可参见 Vaitukaitis 等人 (J. Clin. Endocrinol. Metab. 33:988-91, 1971)。

抗体片段可用来替代整个抗体，并可在原核宿主细胞中较容易地表达。制备和使用单克隆抗体免疫有效部分，也称为“抗体片段”的方法是本领域公知的，并可见于 Better & Horowitz (Methods Enzymol. 178:476-96, 1989), Glockshuber 等人(Biochemistry 29:1362-7, 1990), 美国专利第 5,648,237 号("Expression of Functional Antibody Fragments"), 美国专利第 4,946,778 号("Single Polypeptide Chain Binding Molecules"), 美国专利第 5,455,030 号("Immunotherapy Using Single Chain 多肽 Binding Molecules"), 在此将其引为参考。

抗原：能够在动物体内刺激抗体产生或 T-细胞应答的化合物，组合物或物质，包括向动物给药，例如，注射或吸收的组合物。抗原与特异性的体液或细胞免疫性产物，包括通过异源免疫原诱导的产物发生反应。术语“抗原”包括所有相关的抗原表位。

cDNA(互补 DNA)：缺乏内部、非编码片段(内含子)和决定转录的

调节序列的 DNA 片段。可通过逆转录从细胞中提取的信使 RNA 在实验室中合成 cDNA。

保守性取代：对具有类似生化性质的氨基酸残基进行一个或多个氨基酸的取代(例如，1, 2, 5 或 10 个残基)。通常，保守性取代对所得多肽的活性只有很小的影响甚至没有影响。例如，保守性取代可以是在丙氨酸 2,3-氨基变位酶多肽中，基本上并未影响将 α -丙氨酸转化为 β -丙氨酸能力的氨基酸取代。

在特定的实施例中，保守性取代可以是在丙氨酸 2,3-氨基变位酶多肽中的氨基酸取代，例如，并未明显改变所述蛋白质将 α -丙氨酸转化为 β -丙氨酸能力的、在 SEQ ID NO: 19, 21, 43, 45, 47, 49 或 51 中的保守性取代。在以下实施例中公开了可用于测定丙氨酸 2,3-氨基变位酶活性的方法。丙氨酸扫描可用来鉴定在丙氨酸 2,3-氨基变位酶中的哪一个氨基酸残基能够耐受氨基酸取代。在一个实施例中，当丙氨酸，保守或氨基酸(例如那些以下所示)，或非保守性氨基酸，被一种或多种天然氨基酸取代时，丙氨酸 2,3-氨基变位酶活性的降低未超过 25%，例如未超过 20%，例如未超过 10%。

在一个实施例中，所述肽中包括一个保守性取代，例如在 SEQ ID NOS: 2, 4, 6, 19, 21, 43, 45, 47, 49 或 51 中的任一保守性取代。在另一个实施例中，所述多肽中可包括 10 个或更少的保守性取代，例如 5 个或更少。可使用，例如，诸如位点定向的突变或 PCR 的标准方法，通过对编码所述多肽的核苷酸序列进行操作，制备出含有一个或多个保守性取代的多肽。或者，可使用标准的多肽合成方法制备出含有一个或多个保守性取代的多肽。

取代变体是那些在其氨基酸序列中至少有一个氨基酸残基被去除或在其位置上插入了不同残基的序列。可用于在蛋白质的原始氨基酸中进行取代的氨基酸例子以及被认为是保守性取代的实例包括：Ser 取

代 Ala; Lys 取代 Arg; Gln 或 His 取代 Asn; Glu 取代 Asp; Ser 取代 Cys; Asn 取代 Gln; Asp 取代 Glu; Pro 取代 Gly; Asn 或 Gln 取代 His; Leu 或 Val 取代 Ile; Ile 或 Val 取代 Leu; Arg 或 Gln 取代 Lys; Leu 或 Ile 取代 Met; Met, Leu 或 Tyr 取代 Phe; Thr 取代 Ser; Ser 取代 Thr; Tyr 取代 Trp; Trp 或 Phe 取代 Tyr; 以及 Ile 或 Leu 取代 Val。

关于保守性取代的进一步信息可以在其他地方找到,如 Ben-Bassat 等人, (J. Bacteriol. 169:751-7, 1987), O'Regan 等人, (Gene 77:237-51, 1989), Sahin-Toth 等人, (Protein Sci 3:240-7, 1994), Hochuli 等人, (Bio/Technology 6:1321-5, 1988), WO 00/67796 (Curd 等人)以及遗传学和分子生物学的标准教材中。

缺失: 核酸序列, 例如 DNA, 在每侧相连时所述区域的去除。

可检测的: 能够确证存在或出现。例如, 如果从 β -丙氨酸产生的信号强至足够被检测时, 从 α -丙氨酸制备 β -丙氨酸就是可检测的。

DNA: 脱氧核糖核酸。DNA 是长链聚合物, 其包含了大多数活生物体(一些病毒具有含有核糖核酸 RNA 的基因)的遗传物质。在 DNA 聚合物中的重复单位是四种不同的核苷酸, 每一种核苷酸都含有结合于脱氧核糖的四种碱基之一, 即腺嘌呤, 鸟嘌呤, 胞嘧啶和胸腺嘧啶, 其上还连接有磷酸基团。在 DNA 分子中的核苷酸三联体, 称为密码子, 编码多肽中的氨基酸。术语密码子也可用于从 DNA 序列转录为 mRNA 序列中的对应(互补)的三个核苷酸序列。

外源的: 参照核酸分子和特定的细胞, 本说明书中的术语“外源的”是指未从特定所述天然发现细胞产生的任意核酸分子。因此, 一旦被引入到细胞内, 就认为非自然发生的核酸分子对于所述细胞外源性的。天然发生的核酸分子也可能对特定的细胞是外源性的。例如, 对个体 Y 的细胞而言, 一旦将从个体 X 的细胞分离到的全部染色体导

入 Y 的细胞，那么个体 X 的染色体就是外源的。

功能性缺失：对基因序列进行的突变、部分或完全缺失，插入，或其他变异，其能够减少或抑制所述基因产物的产生，或使得所述基因产物不具有功能。例如，在大肠杆菌中的功能性 panD 缺失阻止了通过 panD 基因编码的天冬氨酸脱羧酶从天冬氨酸盐制备 β -丙氨酸。大肠杆菌中的 panD 功能性缺失灭活了天冬氨酸盐脱羧酶，其导致了在生长培养基中缺乏 β -丙氨酸或泛酸时，大肠杆菌的生长受到抑制。

功能性等价物：具有相同的功能。在涉及丙氨酸 2,3-氨基变位酶的内容中，功能性等价分子包括保留了丙氨酸 2,3-氨基变位酶功能的不同分子。例如，可通过丙氨酸 2,3-氨基变位酶的序列改变提供功能性等价物，其中具有一处或多处序列改变的所述多肽保留了未改变多肽的功能，从而保留了将 α -丙氨酸转化为 β -丙氨酸的能力。

序列改变的例子包括，但不限于，取代(保守性和非保守性)，缺失，突变，移码，和插入。在一个实施例中，给定一个多肽结合于抗体，则结合于相同抗体的多肽就是功能性等价物。因此，功能性等价物包括与多肽具有相同结合特异性的肽，且其可用作替换所述多肽的制剂(例如在泛酸和 3-HP 制备中)。在一个实施例中，功能性等价物包括这样的多肽，其中所述结合序列是不连续的，其中所述抗体结合于线性表位。因此，如果所述肽序列为 MNTVNTRKKF (SEQ ID NO: 19 的氨基酸 1-10)，则包括不连续表位的功能性等价物可表示如下(**=任意数量的插入氨基酸)：NH₂-**-M**N**T**V**N**T**R**K**K**F-COOH。在这个实施例中，如果所述多肽的三维结构是能够结合于 SEQ ID NO: 19 的氨基酸 1-10 的单克隆抗体的三维结构，那么所述多肽就是 SEQ ID NO: 19 的氨基酸 1-10 的功能性等价物。

杂交：互补的单链 DNA 或 RNA 能够形成双链分子的能力。在一

些实施例中，杂交可用来确定两个或更多核酸序列之间的互补。核酸杂交技术可用来在本发明的范围之内获得分离的核酸。简言之，任意与丙氨酸 2,3-氨基变位酶(例如与 SEQ ID NO: 1, 3, 5, 18, 20, 42, 44, 46, 48 或 50 或其变体或片段)具有同源性的核酸分子可用作探针，在适当至较高严谨的条件下，通过杂交来鉴定类似的核酸分子。一旦完成鉴定，可随后将所述核酸纯化，测序，并进行分析从而确定其是否是具有丙氨酸 2,3-氨基变位酶活性的丙氨酸 2,3-氨基变位酶。

可分别通过 Southern 或 Northern 分析进行杂交来鉴定杂交于探针的 DNA 或 RNA 序列。所述探针可以用例如，生物素，荧光团，地高辛，酶，或诸如 ^{32}P 的放射性同位素标记的。可将待分析的 DNA 或 RNA 在琼脂糖或丙烯酰胺凝胶电泳上分离，转移至硝酸纤维素、尼龙或其他适当的膜上，并使用标准技术与探针进行杂交，所述标准技术在本领域是公知的，例如在 Sambrook 等人, (1989) *Molecular Cloning, second edition*, Cold Spring Harbor Laboratory, Plainview, NY 中的 7.39-7.52 部分所述。通常，探针的长度为至少约 20 个核苷酸。例如，包括丙氨酸 2,3-氨基变位酶(例如 SEQ ID NO: 1, 3, 5, 18, 20, 42, 44, 46, 48 或 50 的 20 个连续核苷酸)的 20 个连续核苷酸的探针可用来鉴定同一或类似的核酸。此外，还可以使用大于或小于 20 个核苷酸长的探针。

本说明书还提供了长度为至少约 12 个核苷酸(例如长度为至少约 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, 60, 100, 250, 500, 750, 1000, 1400, 2000, 3000, 4000 或 5000 个核苷酸)，并且能够在杂交条件下，与丙氨酸 2,3-氨基变位酶核酸序列(例如 SEQ ID NO: 1, 3, 5, 18, 20, 42, 44, 46, 48 或 50)的有义链或反义链杂交的分离的核酸序列。所述杂交条件可以使温和或较高的严谨杂交条件。

温和的严谨杂交条件是在约 42°C ，在含有 25 mM KPO_4 (pH 7.4)，5X SSC，5X Denhart's 溶液，50 $\mu\text{g}/\text{mL}$ 变性超声波降解的鲑精 DNA，50% 甲酰胺，10% 硫酸葡聚糖，和 1-15 ng/mL 探针(约 5×10^7 cpm/ μg)

的杂交液中进行杂交,而使用含有 0.2X SSC 和 0.1%十二烷基磺酸钠的洗涤液在约 50°C 进行洗涤。

较高的严谨杂交条件是在约 42°C, 在含有 25 mM KPO₄ (pH 7.4), 5X SSC, 5X Denhart's 溶液, 50 μg/mL 变性超声波降解的鲑精 DNA, 50% 甲酰胺, 10% 硫酸葡聚糖, 和 1-15 ng/mL 探针(约 5×10⁷ cpm/μg) 的杂交液中进行杂交,而使用含有 0.2X SSC 和 0.1%十二烷基磺酸钠的洗涤液在约 65°C 进行洗涤。

分离的: 可将“分离的”生物成分(例如核酸分子或蛋白质)与该成分自然存在的生物体细胞的其他生物成分(例如其他染色体和染色体外 DNA 或 RNA, 和蛋白质)基本上分离和纯化出来。“分离”的核酸分子和蛋白质包括通过标准纯化方法纯化得到的核酸分子和蛋白质。该术语还包括了通过在宿主细胞中重组表达制备的核酸分子和蛋白质, 以及化学合成的核酸, 蛋白质和多肽。

在一个实施例中, 分离的序列是指未与两条序列直接毗邻的自然发生的核酸分子, 其是从所来源的生物体自然发生的基因组中直接毗邻(一个在 5'端且一个在 3'端)的序列。例如, 分离的核酸分子可以是, 但不限于, 任意长度的重组 DNA 分子, 条件是天然发生基因组中直接侧翼于所述重组 DNA 分子的正常存在的核酸序列之一被去除或缺失。因此, 分离的核酸分子包括, 但不限于, 以独立于其他序列存在的单独分子(例如, 由 PCR 或限制性内切酶处理产生的 cDNA 或基因组 DNA 片段), 以及整合入载体、自我复制质粒, 病毒(例如逆转录病毒, 腺病毒或疱疹病毒), 或整合入原核或真核生物基因组 DNA 的重组 DNA。此外, 分离的核酸还可包括作为部分杂交或融合核酸序列的重组 DNA 分子。

在一个实施例中, 参照核酸分子, 术语“分离的”还包括任意的非天然存在的核酸序列, 因为非天然存在的核酸序列在自然界中未找

到，且其在天然发生的基因组中不含有直接连续的序列。例如，非天然发生的核酸分子，例如工程化的核酸分子被认为是分离的核酸。可以使用常见的分子克隆或化学核酸合成技术制备工程化的核酸分子。分离的非天然发生的核酸分子可以独立于其他序列，或整合入载体，自我复制的质粒，病毒(例如逆转录病毒，腺病毒或疱疹病毒)，或原核或真核生物的基因组 DNA。此外，非天然发生的核酸分子还可包括作为部分杂交或融合核酸序列的核酸分子。

赖氨酸 2,3-氨基变位酶：能够将 α -丙氨酸转化为 β -丙氨酸的酶。其包括来自任意生物体物的任意赖氨酸 2,3-氨基变位酶基因，cDNA，RNA 或蛋白质，例如，来自枯草芽孢杆菌(*Bacillus subtilis*)，近端梭菌(*Clostridium subterminale*)，牙龈卟啉单胞菌(*Porphyromonas gingivalis*)，具核梭杆菌(*Fusobacterium nucleatum*)，斯氏梭菌(*Clostridium sticklandii*) 或大肠杆菌 (*Escherichia coli*)。本说明书包括赖氨酸 2,3-氨基变位酶等位变体，以及保留了将 α -丙氨酸转化为 β -丙氨酸活性的任意变体，片段或融合序列。在一个实施例中，其包括由公共序列数据库，如 GenBank 和 EMBL 中如赖氨酸 2,3-氨基变位酶注释的基因所编码的肽。赖氨酸 2,3-氨基变位酶蛋白质的特定示例(及对应的核酸序列)包括以下可公开获得的序列：炭疽杆菌(*Bacillus anthracis*)str. Sterne 的 GenBank Accession Nos. YP_028406；创伤弧菌(*Vibrio vulnificus*) YJ016 的 BAC95867；*Moorella thermoacetica* 的 ZP_00330962；流感嗜血杆菌(*Haemophilus influenzae*)的 ZP_00322274；*Methanosarcina barkeri* str. Fusaro 的 ZP_00298043；*Microbulbifer degradans* 的 ZP_00314982 和破伤风梭菌(*Clostridium tetani*) E88 的 NP_781545。

突变的赖氨酸 2,3-氨基变位酶：含有一个或多个导致多肽具有赖氨酸 2,3-氨基变位酶活性的氨基酸取代的赖氨酸 2,3-氨基变位酶分子。这些突变的示例包括，但不限于，以下的一种或多种：P/S11T；N19Y；L/K/R/T26I；E/R30K；L/V32A；K36E；S/T/C52R；L/T53P/H；Y63F；E/N/D71G；H/I/S85Q；Q/L/E86R；Q/L95M；K/M/Q125L；M128V；

Y132H; Q/S141R; A/D/S/M144G; D179N; K/Q187R; I192V; L228M; D331G/H; M/Q342T; 或 K/Q/T398E, 例如至少 2 种, 至少 3 种, 至少 4 种, 至少 5 种, 至少 6 种, 至少 8 种, 至少 9 种, 至少 10 种, 至少 11 种, 至少 12 种, 至少 13 种, 至少 14 种, 至少 15 种, 至少 16 种, 至少 17 种, 至少 18 种, 至少 19 种, 至少 20 种, 至少 21 种, 至少 22 种, 至少 23 种, 至少 24 或不超过 3 种, 不超过 4 种, 不超过 5 种, 不超过 6 种, 不超过 8 种, 不超过 9 种, 不超过 10 种, 不超过 11 种, 不超过 12 种, 不超过 13 种, 不超过 14 种, 不超过 15 种, 不超过 16 种, 不超过 17 种, 不超过 18 种, 不超过 19 种, 不超过 20 种, 不超过 21 种, 不超过 22 种或不超过 23 种这样的取代。这些取代的特定组合包括, 但不限于: ((1) N19Y, L/T53P/H, H/I/S85Q, D331G/H 和 M/Q342T; (2) N19Y, E/R30K, L/T53P/H, H/I/S85Q, I192V, D331G/H 和 M/Q342T; (3) N19Y, L/K/R/T26I; E/R30K, L/T53P/H, H/I/S85Q, I192V, D331G/H 和 M/Q342T; (4) E/R30K, Y63F, Q/L/E86R, Q/L95M, M128V, A/D/S/M144G, L228M, D331G/H 和 K/Q/T398E; (5) E/R30K, K36E, Y63F, Q/L/E86R, Q/L95M, M128V, A/D/S/M144G, D179N, L228M, D331G/H 和 K/Q/T398E; (6) E/R30K, Q/L95M, M128V 和 D331G/H; (7) P/S11T, E/R30K, Q/L95M, M128V, Q/S141R, K/Q187R 和 D331G/H; (8) E/R30K, L/V32A, L/T53P/H, E/N/D71G, Q/L95M; K/M/Q125L, M128V 和 D331G/H; (9) E/R30K, C52R, Q/L95M; M128V 和 D331G/H; (10) Q/L95M, M128V 和 D331G/H; (11) Q/L95M, M128V; Y132H 和 D331G/H; 和(12) Q/L95M, M128V 和 D331G/H。

核酸: 涵盖了 RNA 和 DNA, 其包括, 但不限于, cDNA, 基因组 DNA, 和合成(例如化学合成的)DNA。核酸可以是双链或单链的。为单链时, 该核酸可以是有义链或反义链。此外, 核酸可以是环状或线性的。

寡核苷酸: 至少 9 个核苷酸的线性多核苷酸(例如 DNA 或 RNA)序列, 例如至少 12 个, 至少 15 个, 至少 18 个, 至少 20 个, 至少 24

个, 至少 25 个, 至少 27 个, 至少 30 个, 至少 50 个, 至少 100 个或至少 200 个核苷酸长。在特定的实施例中, 寡核苷酸包括 SEQ ID NO: 1, 3, 5, 18, 20, 42, 44, 46, 48 或 50 (或其互补链) 的至少 9 个连续核苷酸, 例如 SEQ ID NO: 1, 3, 5, 18, 20, 42, 44, 46, 48 或 50 (或其互补链) 的至少 12 个, 至少 15 个, 至少 18 个, 至少 20 个, 至少 24 个, 至少 25 个, 至少 27 个, 至少 30 个, 至少 50 个, 至少 100 个或至少 200 个核苷酸长。

可操作性连接的: 当第一核酸序列与第二核酸序列放置成具有功能关系时, 第一核酸序列可以可操作性连接于第二核酸序列。例如, 如果启动子会影响编码序列的转录或表达, 那么可将该启动子可操作性地连接于该编码序列。一般而言, 可操作连接的 DNA 序列是连续的, 并且在有必要连接两个蛋白质编码区域时, 需要具有相同的读框。

ORF(开放阅读框): 没有终止密码子的一系列编码氨基酸的核苷酸三联体(密码子)。这些序列通常可被翻译成肽。

泛酸盐或泛酸: 应用于化妆品、医药和食品中的重要商业维生素。在本说明书中, 术语泛酸和泛酸盐可以交替使用, 不仅可指游离酸, 还可指 D-泛酸的盐, 例如钙盐, 钠盐, 铵盐或钾盐。可通过化学合成或生物技术使用本发明的细胞和方法从 β -丙氨酸制备泛酸盐。

测定泛酸盐量的方法是公知的(例如, 参见 Rieping 等人的美国专利第 6,184,006 号, 和 Eggeling 等人的美国专利第 6,177,264 号)。例如, 可使用植物乳杆菌(*Lactobacillus plantarum*)泛酸盐分析(检测菌株: *Lactobacillus plantarum* ATCC 8014, Cat. No. 3211-30-3; 培养基: Bacto 泛酸分析培养基(DIFCO Laboratories, Michigan, USA), cat. No. 0604-15-3)对 D-泛酸盐进行定量测定。这一指示菌株仅可在指示培养基中存在泛酸盐的条件下生长, 并在所述培养基中显示出可光度测量的线性生长趋势对泛酸浓度的影响。泛酸的半钙盐(Sigma Catalog Number P

2250)用作校准。在 580nm 波长测定光密度。

多肽修饰：本公开包括丙氨酸 2,3-氨基变位酶多肽，以及合成的实施方案。此外，可使用本发明所述方法利用具有丙氨酸 2,3-氨基变位酶活性的类似物(非肽有机分子)，衍生物(从公开的肽序列开始获得的化学官能化的肽分子)和变体(同源体)。在此公开的肽包括能够自然发生或以其他方式产生的，或者是 L-或者是 D-氨基酸的氨基酸序列。

可通过多种化学方法对多肽进行修饰，从而制备与未修饰多肽具有本质相同活性的衍生物，并可任选具有其他需要的性质。例如，所述蛋白质的羧基基团，无论在羧基端或侧链上，都可以药物可接受阳离子盐的形式提供，或酯化形成 C_1-C_{16} 的酯，或转化成通式为 NR_1R_2 的酰胺，其中 R_1 和 R_2 独立地为 H 或 C_1-C_{16} 烷基，或组合形成杂环，例如 5-或 6-元环。所述肽的氨基，无论在氨基端或侧链上，都可以是诸如 HCl, HBr, 乙酸, 苯甲酸, 甲苯磺酸, 马来酸, 酒石酸和其他有机酸的药物可接受加成盐形式，或可被修饰成 C_1-C_{16} 烷基或二烷基氨基或进一步转变为酰胺。

使用公知的技术，可将肽侧链的羟基基团转化为 C_1-C_{16} 烷氧基或 C_1-C_{16} 酯。可使用一种或多种卤素原子，例如，F, Cl, Br 或 I, 或 C_1-C_{16} 烷基， C_1-C_{16} 烷氧基，羧酸及其酯，或这些羧酸的酰胺来取代肽侧链中的苯基或酚环。肽侧链中的亚甲基可延伸成同源的 C_2-C_4 亚烷基。可使用任意一种公知的保护基团，例如乙酰胺基团保护巯基。本领域所属技术人员应该意识到，向本公开的多肽中引入环结构的方法可以对得到增强稳定性的结构进行选择并提供该结构的构象限制。例如，可将所述多肽加上 C-或 N-末端的半胱氨酸，从而当所述氧化肽含有二硫键，生成环肽。其他的环肽方法包括形成硫酯和羧基-和氨基-末端酰胺和酯。

肽模拟和有机模拟实施方案也属于本公开的范围之内，当这些肽

模拟和有机模拟化学成分的三维排列模拟了所述肽主链和组成氨基酸侧链的三维排列时,可得到具有可检测丙氨酸 2,3-氨基变位酶活性的本发明蛋白质的多肽和有机模拟物。对计算机建模应用而言,药效团是理想的,具有其生物学活性必须结构的三维界定。可对多肽-和有机模拟物进行设计从而将每个药效团都适合当前的计算机建模软件(使用计算机辅助药物设计或 CADD)。参见 Walters, "Computer-Assisted Modeling of Drugs", in Klegerman & Groves, eds., 1993, Pharmaceutical Biotechnology, Interpharm Press: Buffalo Grove, IL, pp. 165-174 和 Principles of Pharmacology Munson (ed.) 1995, Ch. 102 中关于 CADD 中技术使用的描述。本公开的范围同样包括了使用这些技术制备的模拟物。在一个实施例中,通过丙氨酸 2,3-氨基变位酶或其变体,片段或融合物产生了能够模拟丙氨酸 2,3-氨基变位酶活性的模拟物。

多核苷酸:任意长度的线性核酸序列。因此,多核苷酸包括至少 15 个,至少 25 个,至少 50 个,至少 75 个,至少 100 个,至少 200 个,至少 300 个,至少 400 个,至少 500 个,至少 600 个,至少 700 个,至少 800 个,至少 900 个,至少 1000 个,至少 1100 个或甚至至少 1200 个核苷酸长度的分子。在特定的实施例中,多核苷酸包括 SEQ ID NO: 1, 3, 5, 18, 20, 42, 44, 46, 48 或 50 (或其互补链)的至少 15 个连续核苷酸,例如 SEQ ID NO: 1, 3, 5, 18, 20, 42, 44, 46, 48 或 50 (或其互补链)的至少 25 个,至少 50 个,至少 75 个,至少 100 个,至少 200 个,至少 300 个,至少 400 个,至少 500 个,至少 600 个,至少 700 个,至少 800 个,至少 900 个,至少 1000 个,至少 1100 个或甚至至少 1200 个连续核苷酸。

探针和引物:“探针”包括含有可检测标记或报告分子的分离核酸分子。示例性的标记包括放射性同位素,配体,化学发光剂,荧光团,和酶。对进行标记的方法和对各种目的的标记进行选择的指导已有讨论,例如,参见 Sambrook 等人 (ed.), Molecular Cloning: A Laboratory Manual 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory

Press, Cold Spring Harbor, N. Y., 1989, 和 Ausubel 等人 (ed.) *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience, New York (with periodic updates), 1987。

“引物”通常是具有 10 个或更多核苷酸(例如具有从约 10 个核苷酸至约 100 个核苷酸的核酸分子)的核酸分子。可通过核酸杂交将引物退火成互补的靶核酸链, 从而在引物和靶核酸链之间形成杂交, 然后通过, 例如, DNA 聚合酶沿靶核酸链进行延伸。引物对可用来对核酸序列进行扩增, 例如, 通过聚合酶链反应(PCR)或其他核酸扩增方法。

制备和使用探针和引物的方法已有描述, 例如, 参见 Sambrook 等人(ed.), *Molecular Cloning: A Laboratory Manual*, 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989; Ausubel 等人(ed.), *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience, New York (with periodic updates), 1987; 和 Innis 等人, *PCR Protocols: A Guide to Methods and Applications*, Academic Press: San Diego, 1990。可从已知序列衍生出 PCR 引物对, 例如, 通过使用针对此目的的计算机程序, 例如 Primer (Version 0.5, © 1991, Whitehead Institute for Biomedical Research, Cambridge, Mass.)。本领域所属技术人员应该理解, 特定探针或引物的特异性会随着长度的增加而增加, 但探针或引物的长度却可从全长序列直至短至 5 个连续核苷酸的序列。因此, 例如, 与具有 15 个连续核苷酸的相对应引物相比, 具有 20 个连续核苷酸的引物对靶的退火具有更高的特异性。因此, 为了获得更大的特异性, 可以选择包括, 例如, 25 个, 30 个, 35 个, 40 个, 50 个, 60 个, 70 个, 80 个, 90 个, 100 个, 150 个, 200 个, 250 个, 300 个, 350 个, 400 个, 450 个, 500 个, 550 个, 600 个, 650 个, 700 个, 750 个, 800 个, 850 个, 900 个, 950 个, 1000 个, 1050 个, 1100 个, 1150 个, 1200 个, 1250 个, 1300 个, 1350 个, 1400 个, 1450 个, 1500 个或更多个连续核苷酸的探针或引物。

启动子：能够指导核酸转录的核酸控制序列排列。启动子包括邻近转录起始位点的必须核酸序列，例如，在聚合酶 II 型启动子情况下的 TATA 元件。启动子还可任选地包括远端增强子或抑制物元件，其可位于离转录起始位点数千碱基对的位置上。

纯化的：术语纯化的并不需要绝对纯净；相反，其是个相对的术语。因此，例如，纯化的肽制剂是这样的一种制剂，其中所述肽或蛋白质比细胞环境内的该肽和蛋白质更富集，因此所述多肽是基本上从伴随其存在的细胞成分(核酸，脂类，碳水化合物，和其他肽)中分离出来的。在另一个实施例中，纯化的肽制剂是这样的一种制剂，其中所述多肽基本上没有污染物，就像那些在肽的化学合成之后所表现出的那样。

在一个实施例中，当至少 50%的样本由丙氨酸 2,3-氨基变位酶肽组成时，例如，当至少 60%，70%，80%，85%，90%，92%，95%，98%或 99%或更高部分的样本由丙氨酸 2,3-氨基变位酶肽组成时，该丙氨酸 2,3-氨基变位酶肽是纯化的。可用于纯化多肽的方法示例包括，但不限于，在 Sambrook 等人 (Molecular Cloning A Laboratory Manual, Cold Spring Harbor, New York, 1989, Ch 17)中所公开的方法。可通过，例如，蛋白质样本的聚丙烯酰胺凝胶电泳，然后对所述聚丙烯酰胺凝胶染色后的单个多肽条带进行观察；高压液相色谱；测序；或其他常规方法测定蛋白质的纯度。

重组体：重组核酸分子是这样的分子，其具有非自然发生的序列或由两种独立序列片段人工组合而成的序列。可使用本领域中公知的方法，例如化学合成和对核酸分子的分离片段进行人工操作，例如通过遗传工程技术来获得这种人工组合。重组体还可用来描述经过人工操作，但却含有与从中分离到所述核酸的生物体中找到的相同调节序列和编码区的核酸分子。

序列同一性/相似性：两个或多个核酸序列之间，或两个或多个氨基酸序列之间的同一性/相似性，是以序列之间的同一性或相似性的术语来表达的。可以通过同一性的百分比来测定序列同一性；百分比越高，所述序列同一性越高。可以通过相似性的百分比(其考虑的是保守氨基酸取代)来测定序列相似性；百分比越高，序列的相似性越高。当使用标准方法进行比对时，核酸或氨基酸序列的同源物或直系同源物(ortholog)具有相对较高的序列同一性/相似性。与相对更远相关的物种(例如人和秀丽线虫序列)相比，当直系同源(orthologous)蛋白质或 cDNA 是从更加紧密相关的物种中(例如人类和小鼠序列)衍生而来的时候，其同源性更加显著。

对序列比对进行比较的方法是本领域所公知的。已有多种程序和比对算法，参见 Smith & Waterman, *Adv Appl Math* 2:482, 1981; Needleman & Wunsch, *J Mol Biol* 48:443, 1970; Pearson & Lipman, *Proc Natl Acad Sci USA* 85:2444, 1988; Higgins & Sharp, *Gene*, 73:237-44, 1988; Higgins & Sharp, *CABIOS* 5:151-3, 1989; Corpet et al, *Nuc Acids Res.* 16:10881-90, 1988, Huang 等人 *Computer Appls in the Biosciences* 8, 155-65, 1992, 和 Pearson 等人, *Meth Mol Biol* 24:307-31, 1994. Altschul 等人, *J Mol Biol* 215:403-10, 1990, 对序列比对方法和同源性计算进行了详细的描述。

可从若干渠道获得 NCBI Basic Local Alignment Search Tool (BLAST) (Altschul 等人, *J Mol Biol* 215:403-10, 1990), 包括 National Center for Biological Information (NCBI, National Library of Medicine, Building 38A, Room 8N805, Bethesda, MD 20894)和通过互联网, 其可与序列分析程序 blastp, blastn, blastx, tblastn 和 tblastx 相关联使用。还可以在 NCBI 网站找到其他的信息。

BLASTN 可用于比较核酸序列, 而 BLASTP 则用于比较氨基酸序

列。为了比较两个核酸序列，可对选项进行如下设置：-i 设置成含有需要比较的第一核酸序列的文件(例如 C:\seq1.txt)；-j 设置成含有需要比较的第二核酸序列的文件(例如 C:\seq2.txt)；-p 设置成 blastn；-o 设置成任意需要的文件名(例如 C:\output.txt)；-q 设置成-1；-r 设置成 2；其他全部选项保留其默认设置。例如，以下命令可用来生成含有两个核酸序列比较的输出文件：C:\BI2seq -i c:\seq1.txt -j c:\seq2.txt -p blastn -o c:\output.txt -q -1 -r 2。

为了比较两个氨基酸序列，可对 BI2seq 的选项进行如下设置：-i 设置成含有需要比较的第一氨基酸序列的文件(例如 C:\seq1.txt)；-j 设置成含有需要比较的第二氨基酸序列的文件(例如 C:\seq2.txt)；-p 设置成 blastn；-o 设置成任意需要的文件名(例如 C:\output.txt)；其他全部选项保留其默认设置。例如，以下命令可用来生成含有两个氨基酸序列比较的输出文件：C:\BI2seq -i c:\seq1.txt -j c:\seq2.txt -p blastn -o c:\output.txt。如果两个比较的序列具有同源性，那么所设计的输出文件就会将那些同源性区域表示为比对序列。如果两个比较的序列没有同源性，那么所设计的输出文件就不会表现出比对序列。

一旦完成比对，可以通过计数两个序列中存在的同一的核苷酸或氨基酸残基位置数量确定匹配的数量。可通过用匹配的数量除以在鉴定序列中设定的序列长度，或除以联接的长度(例如来自鉴定序列中设定序列的 100 个连续核苷酸或氨基酸残基)，然后将所得数值乘以 100 就可确定序列同一性的百分数。例如，当与具有 1554 个核苷酸的检测序列进行比对具有 1166 个匹配的核酸序列与所述检测序列的同一性为 75%(1166/1554*100=75.0)。将序列同一性的百分数四舍五入至 0.1。例如，75.11, 75.12, 75.13 和 75.14 都被四舍五入为 75.1，而 75.15, 75.16, 75.17, 75.18 和 75.19 则被四舍五入成 75.2。长度值永远都是整数。在另一个实施例中，含有 20 个核苷酸区域的靶序列与含有来自鉴定序列的具有 20 个连续核苷酸的序列比对如下，其还有与鉴定序列 75%序列同一性的区域(即：15/20*100=75)。

```

      1                               20
靶序列      CTGTCAGTCA
      | | | | | | | | | |
鉴定序列      CTGCCAGTGA

```

为了对氨基酸数量大于 30 的氨基酸序列进行比较，可将默认的 BLOSUM62 矩阵设置成默认参数(缺口存在值为 11，而每个残基缺口值为 1)来使用 Blast 2 序列函数。同源序列的特征在于其拥有在氨基酸序列全长比对基础上，采用诸如 nr 或 swissprot 的数据库进行的 NCBI Basic Blast 2.0, gapped blastp 进行的序列同一性计数至少为 70%。采用 blastn 程序进行的查询搜索可通过 DUST(Hancock and Armstrong, 1994, Comput. Appl. Biosci. 10:67-70)进行过滤。其他的程序使用 SEG。此外，还可以进行手工比对。当采用这样的方法进行评估时，具有较高相似性的蛋白质会表现出增加的同源性百分数，例如，至少 75%，80%，85%，90%，95%或 99%序列同一性。

当比对短肽(低于约 30 个氨基酸)时，可将 PAM30 矩阵设置成默认参数(开放缺口 9，延伸缺口 1 罚分)，来使用 Blast2 序列函数进行比对。当使用本发明进行分析时，与参考序列具有较高相似性的蛋白质会表现出增加的同源性百分数，例如，至少 60%，70%，75%，80%，85%，90%，95%，98%，99%序列同一性。当对小于全序列的序列进行序列同一性比较时，在 10-20 个氨基酸的短窗口内，同源序列通常拥有至少 75%的序列同一性，并根据其与参考序列的同源性，其拥有的序列同一性可至少为 85%，90%，95%或 98%。在这些短窗口上测定序列同一性的方法在 NCBI 的网站上已有描述。

两个核酸分子紧密关联的一个标志是这两个分子可在严谨条件下相互杂交。严谨条件是序列依赖的，并在不同的环境参数下各不相同。能够在严谨条件下与丙氨酸 2,3-氨基变位酶基因序列杂交的核酸分子通常都可以上述条件下分别杂交于基于丙氨酸 2,3-氨基变位酶全基因

或该基因选定部分的探针。

由于遗传密码的简并性，未表现出较高同一性的核酸序列也可以编码同一或相似(保守的)的氨基酸序列。可以使用这一简并性在核酸序列中进行改变，从而制备出全部都编码基本相同蛋白质的多种核酸分子。这样的同源核酸序列可以，例如，拥有通过本发明的方法确定的至少 60%，70%，80%，90%，95%，98%或 99%的序列同一性。

本领域所属技术人员应该理解，在此提供的序列同一性范围仅具有参考作用；也有可能获得了非常显著的同源序列却落在所提供的范围之外。

两个核酸序列基本上同一的另一个(不是必须具有叠加作用的)标志是第一核酸编码的多肽可与由第二核酸编码的多肽具有免疫交叉反应性。

特异性结合剂：基本上仅结合于限定靶，例如肽靶的制剂。例如，丙氨酸 2,3-氨基变位酶结合剂包括抗-丙氨酸 2,3-氨基变位酶抗体和基本上仅结合于丙氨酸 2,3-氨基变位酶的其他制剂(例如肽或药)。丙氨酸 2,3-氨基变位酶蛋白(或其片段)的抗体可用来对这样的蛋白质进行纯化或鉴定。

转化的：转化的细胞是向其中引入了核酸分子的细胞，例如通过分子生物学技术。转化包括可向这样的细胞中引入核酸分子的全部技术，包括，但不限于用病毒载体进行转染，结合，用质粒载体进行转化，和通过电穿孔、脂质体转染和颗粒枪加速将裸 DNA 导入。

变体，片段或融合体：公开的丙氨酸 2,3-氨基变位酶序列包括保留了丙氨酸 2,3-氨基变位酶活性的变体，片段或融合体。编码丙氨酸 2,3-氨基变位酶蛋白质(例如 SEQ ID NO: 1, 3, 5, 18, 20, 42, 44,

46, 48 或 50), 融合丙氨酸 2,3-氨基变位酶蛋白质, 或丙氨酸 2,3-氨基变位酶蛋白质片段或变体的 DNA 序列可被工程化从而允许在真核细胞, 细菌, 昆虫或植物中表达蛋白质。为了得到表达, 可对 DNA 序列进行改变并可操作地连接于其他调控序列。含有调节序列和所述蛋白质的终产物可称为载体。这样的载体可被导入真核细胞, 细菌, 昆虫或植物细胞。一旦进入细胞, 所述载体可允许蛋白质产生。

融合蛋白质包括丙氨酸 2,3-氨基变位酶(或其变体或片段), 例如 SEQ ID NO.: 19, 21, 43, 45, 47, 49 或 51, 连接到不显著抑制丙氨酸 2,3 氨基变位酶活性, 例如将 α -丙氨酸转化为 β -丙氨酸的能力的其它氨基酸序列上的蛋白质。在一个实施例中, 其他氨基酸序列不大于约 10, 12, 15, 20, 25, 30 或 50 个氨基酸的长度。此外, 可将间隔序列放置于丙氨酸 2,3-氨基变位酶序列和其他氨基酸序列之间。这样的间隔序列可以是至少 4 个, 至少 6 个或至少 10 个氨基酸长。

本领域的普通技术人员可意识到 DNA 序列可在多个方面被改变, 只要不影响该编码的蛋白质的生物活性。例如, 可使用 PCR 在编码丙氨酸 2,3-氨基变位酶的 DNA 序列中产生变化。这种变体可以是用于表达该蛋白质的宿主细胞所偏好的密码子, 或促进表达其它序列变化而优化的变体。

载体: 被导入细胞从而产生转化细胞的核酸分子。载体可包括允许其在细胞中, 例如复制原点进行复制的核酸序列。载体还可包括一种或多种可选择的标记基因和其它本领域已知的遗传元件。

丙氨酸 2, 3-氨基变位酶的核酸和多肽

在此公开了具有丙氨酸 2,3-氨基变位酶活性的多肽。在一个实施例中, 该多肽是突变的赖氨酸 2,3-氨基变位酶序列。在一个实施例中, 突变的赖氨酸 2,3-氨基变位酶包括一种或多种以下取代: P/S11T; N19Y; L/K/R/T26I; E/R30K; L/V32A; K36E; S/T/C52R; L/T53P/H;

Y63F; E/N/D71G; H/I/S85Q; Q/L/E86R; Q/L95M; K/M/Q125L; M128V; Y132H; Q/S141R; A/D/S/M144G; D179N; K/Q187R; I192V; L228M; D331G/H; M/Q342T; 或 K/Q/T398E, 其中数字之前的字母代表在赖氨酸 2,3 氨基变位酶中所发现的氨基酸的单字母氨基酸编码, 所述数字代表氨基酸位置(根据牙龈卟啉单胞菌赖氨酸 2,3 氨基变位酶(SEQ ID NO: 52)编号, 参见图 7), 而数字之后的字母代表在丙氨酸 2,3 氨基变位酶中所发现氨基酸的单字母氨基酸编码。根据需要突变的赖氨酸 2,3 氨基变位酶序列, 实际上的第一个氨基酸和氨基酸数字都会发生改变。然而, 本领域所属技术人员应该可以通过序列比对确定同源序列中的所述位置。例如, 如图 7 所示, 牙龈卟啉单胞菌赖氨酸 2,3 氨基变位酶的位置 128 对应于具核梭杆菌赖氨酸 2,3 氨基变位酶 (SEQ ID NO: 10)的位置 129, 斯氏梭菌(*Clostridium sticklandii*)赖氨酸 2,3 氨基变位酶 (SEQ ID NO: 33)的位置 126, 和枯草芽孢杆菌(*Bacillus subtilis*)赖氨酸 2,3 氨基变位酶 (SEQ ID NO: 59)的位置 136。可根据图 7 所提供的信息和其他公众可获得的赖氨酸 2,3 氨基变位酶序列, 使用本领域中的公知方法对剩余 24 个位点的每一个都进行类似的分析。

使用标准的分子生物学方法都可对任意的赖氨酸 2,3 氨基变位酶进行突变, 产生丙氨酸 2,3-氨基变位酶。例如, 可对来自于诸如芽孢杆菌, 梭菌, 埃希氏, 梭形杆菌, 嗜血杆菌, 甲烷八叠球菌, *Microbulbifer*, *Moorella*, 卟啉单胞菌, 嗜热厌氧菌或弧菌的原核生物的赖氨酸 2,3 氨基变位酶进行突变, 从而包含有一种或多种以下取代, P/S11T; N19Y; L/K/R/T26I; E/R30K; L/V32A; K36E; S/T/C52R; L/T53P/H; Y63F; E/N/D71G; H/I/S85Q; Q/L/E86R; Q/L95M; K/M/Q125L; M128V; Y132H; Q/S141R; A/D/S/M144G; D179N; K/Q187R; I192V; L228M; D331G/H; M/Q342T 或 K/Q/T398E, 例如至少 2 种, 至少 3 种, 至少 4 种, 至少 5 种, 至少 6 种, 至少 8 种, 至少 9 种, 至少 10 种, 至少 11 种, 至少 12 种, 至少 13 种, 至少 14 种, 至少 15 种, 至少 16 种, 至少 17 种, 至少 18 种, 至少 19 种, 至少 20 种, 至少 21 种, 至少 22 种, 至少 23 种, 至少 24 种或不多于 3 种, 不多于 4 种, 不多于 5 种,

不多于 6 种, 不多于 8 种, 不多于 9 种, 不多于 10 种, 不多于 11 种, 不多于 12 种, 不多于 13 种, 不多于 14 种, 不多于 15 种, 不多于 16 种, 不多于 17 种, 不多于 18 种, 不多于 19 种, 不多于 20 种, 不多于 21 种, 不多于 22 种或不多于 23 种这样的突变。这些取代的特定组合包括, 但不限于: (1) N19Y, L/T53P/H, H/I/S85Q, D331G/H, 和 M/Q342T; (2) N19Y, E/R30K, L/T53P/H, H/I/S85Q, I192V, D331G/H, 和 M/Q342T; (3) N19Y, L/K/R/T26I; E/R30K, L/T53P/H, H/I/S85Q, I192V, D331G/H, 和 M/Q342T; (4) E/R30K, Y63F, Q/L/E86R, Q/L95M, M128V, A/D/S/M144G, L228M, D331G/H, 和 K/Q/T398E; (5) E/R30K, K36E, Y63F, Q/L/E86R, Q/L95M, M128V, A/D/S/M144G, D179N, L228M, D331G/H, 和 K/Q/T398E; (6) E/R30K, Q/L95M, M128V, 和 D331G/H; (7) P/S11T, E/R30K, Q/L95M, M128V, Q/S141R, K/Q187R, 和 D331G/H; (8) E/R30K, L/V32A, L/T53P/H, E/N/D71G, Q/L95M; K/M/Q125L, M128V, 和 D331G/H; (9) E/R30K, C52R, Q/L95M; M128V, 和 D331G/H; (10) Q/L95M, M128V, 和 D331G/H; (11) Q/L95M, M128V; Y132H, 和 D331G/H; 和(12) Q/L95M, M128V, 和 D331G/H。

具有丙氨酸 2,3-氨基变位酶活性肽的特别示例如 SEQ ID NOS: 19, 21, 43, 45, 47, 49 和 51 所示。然而, 本公开还包括了保留了丙氨酸 2,3-氨基变位酶活性的 SEQ ID NOS: 19, 21, 43, 45, 47, 49 和 51 的变体, 融合体和片段。在特定的实施例中, SEQ ID NOS: 19, 21, 43, 45, 47, 49 和 51 的变体, 融合体和片段具有 SEQ ID NOS: 41 的至少 50%的丙氨酸 2,3-氨基变位酶活性, 例如 SEQ ID NOS: 41 的至少 60%, 至少 70%, 至少 80%, 至少 90%, 至少 95%, 至少 98%或甚至至少 100%的丙氨酸 2,3-氨基变位酶活性。在其他实施例中, SEQ ID NOS: 19, 21, 43, 45, 47, 49 和 51 的变体, 融合体和片段具有 SEQ ID NOS: 51 的至少 50%的丙氨酸 2,3-氨基变位酶活性, 例如 SEQ ID NOS: 51 的至少 60%, 至少 70%, 至少 80%, 至少 90%, 至少 95%, 至少 98%或甚至至少 100%的丙氨酸 2,3-氨基变位酶活性。

片段的示例可包括,但不限于:SEQ ID NO: 2 或 4(SEQ ID NOS: 6, 19, 21, 43, 45, 47, 49 和 51(例如 SEQ ID NO: 6, 49 或 51 的氨基酸 42-342)的对应片段也包括在本发明中,并可使用图 7 进行测定)的氨基酸 1-390, 15-390, 50-390, 50-350, 60-350, 15-340, 75-340, 19-331 或 100-339。本发明还提供了丙氨酸 2,3-氨基变位酶肽,其含有 SEQ ID NO: 19, 21, 43, 45, 47, 49 或 51 的至少 15 个连续氨基酸,并保留了丙氨酸 2,3-氨基变位酶活性。丙氨酸 2,3-氨基变位酶肽还可包括 SEQ ID NO: 19, 21, 43, 45, 47, 49 或 5 的至少 16 个,至少 17 个,至少 18 个,至少 19 个,至少 20 个,至少 21 个,至少 22 个,至少 23 个,至少 24 个,至少 25 个,至少 26 个,至少 27 个,至少 28 个,至少 29 个,至少 30 个,至少 50 个,至少 75 个,至少 100 个,至少 150 个,至少 200 个,至少 250 个,至少 300 个或更多个连续的氨基酸残基。

变体包括一个或多个氨基酸取代,例如一个或多个保守氨基酸取代,一个或多个非保守氨基酸取代或其组合。变体还包括一个或多个氨基酸缺失或插入(或其组合,例如单个缺失与多个插入),例如,不超过 50 个氨基酸,不超过 20 个氨基酸,不超过 10 个氨基酸,不超过 5 个氨基酸,或不超过 2 个氨基酸的加入或缺失,例如 1-5 个氨基酸,1-10 个氨基酸或 2-20 个氨基酸的加入或缺失。在一个实施例中,变体丙氨酸 2,3 氨基变位酶具有不超过 5 个,不超过 10 个,不超过 20 个,不超过 30 个,不超过 40 个或不超过 50 个的保守性取代。非保守性取代那些取代,其中所述氨基酸具有更本质的区别,例如其作用可表现在维持:(a)在取代区域所述多肽主链的结构,例如,作为片状或螺旋构象;(b)在靶位点所述多肽的电荷或疏水性;或(c)侧链的体积。通常预期可在多肽功能中产生最大改变的取代是那些取代,其中:(a)亲水残基,例如丝氨酸或苏氨酸,被疏水残基,例如亮氨酸,异亮氨酸,苯丙氨酸,缬氨酸或丙氨酸所取代;(b)半胱氨酸或脯氨酸被任意其他残基取代;(c)具有正电性的侧链,例如赖氨酸,精氨酸或组氨酸的残基,被负电性残基,例如谷氨酸或天冬氨酸取代;或(d)具有大体积侧链的残基,例如苯丙氨酸,被没有侧链的残基,例如甘氨酸所取代。可使用

本发明公开的分析方法对肽催化转化 α -丙氨酸为 β -丙氨酸的能力进行分析，从而评估这些氨基酸取代(或其他缺失或加入)对具有丙氨酸 2,3-氨基变位酶活性的肽所产生的影响。

可使用标准方法，例如位点定向的突变或 PCR 对编码丙氨酸 2,3-氨基变位酶多肽的核苷酸序列进行操作从而产生变体丙氨酸 2,3-氨基变位酶多肽的序列。

在依然保留了丙氨酸 2,3-氨基变位酶活性的条件下，可进行取代的示例包括，但不限于：SEQ ID NO: 21 (在其他公开序列中的对应取代，例如 SEQ ID NO: 6, 49 或 51 中的 Y289W；SEQ ID NO: 19 中的 Y290W，SEQ ID NO: 41, 43, 45 或 47 中的 Y287W；和 SEQ ID NO: 2 或 4 中的 Y297W，也包括在本发明的范围内，并可通过使用图 7 进行测定)中的 V108L, T240S, D295E, Y290F 或其组合，以及其组合。只要该氨基酸序列编码的多肽保留了丙氨酸 2,3-氨基变位酶活性，变体丙氨酸 2,3-氨基变位酶多肽就可与 SEQ ID NOS: 2, 4, 6, 19, 21, 43, 45, 47, 49 或 51 中的任意序列具有至少 60，至少 65，至少 70，至少 75，至少 80，至少 85，至少 90，至少 95，至少 97，至少 98 或至少 99% 的序列同一性。

可使用公开的丙氨酸 2,3-氨基变位酶序列产生融合蛋白(及对应的核酸序列)。融合蛋白可包括全长丙氨酸 2,3-氨基变位酶蛋白序列(例如 SEQ ID NO: 2, 4, 6, 19, 21, 43, 45, 47, 49 或 51)，或连接于第二氨基酸序列、具有丙氨酸 2,3-氨基变位酶活性的序列变体或片段。在一些实施例中，所述第二氨基酸序列包括至少 5 个氨基酸，例如至少 10 个氨基酸，至少 20 个氨基酸，至少 30 个氨基酸，至少 50 个氨基酸，至少 75 个氨基酸，至少 100 个氨基酸，至少 150 个氨基酸或甚至至少 200 个氨基酸。在特定的实施例中，所述丙氨酸 2,3-氨基变位酶序列和第二氨基酸序列通过间隔序列相连接。间隔序列的特定实施例包括一个或多个丙氨酸或甘氨酸残基，或其他非极性氨基酸或中性极性氨基

酸。在一些实施例中，间隔物不超过 50 个氨基酸，例如不超过 20 个氨基酸，不超过 10 个氨基酸，不超过 5 个氨基酸，例如 5-50 氨基酸。

本发明还公开了编码具有丙氨酸 2,3-氨基变位酶活性的分离核酸分子，例如包括 SEQ ID NO: 18, 20, 42, 44, 46, 48 或 50 的序列。然而，本发明还包括了保留了编码具有丙氨酸 2,3-氨基变位酶活性的 SEQ ID NOS: 18, 20, 42, 44, 46, 48 和 50 序列的变体，融合体和片段。在一个实施例中，可将编码具有丙氨酸 2,3-氨基变位酶活性的肽的分离核酸分子可操作地连接于启动子序列，并可以是载体的一部分。所述核酸可以是可用于转化细胞和制备转化细胞或转基因非人类哺乳动物(例如小鼠，大鼠和兔子)的重组核酸。

本发明还公开了具有至少一种编码具有丙氨酸 2,3-氨基变位酶活性多肽的外源核酸分子(例如 SEQ ID NO: 18, 20, 42, 44, 46, 48 或 50 或其保留了丙氨酸 2,3-氨基变位酶活性的序列片段，融合体或变体)的转化细胞。在一个实施例中，这样的转化细胞可从 α -丙氨酸制备 β -丙氨酸。在其他实施例中，所述细胞可制备 3-HP，泛酸盐，CoA，或有机化合物，例如 1,3-丙二醇。转化细胞可以是真核或原核细胞，例如细菌细胞，植物细胞或酵母细胞。

编码丙氨酸 2,3-氨基变位酶的核酸序列可含有编码该酶的完整核酸序列，以及保留了所希望酶活性的部分序列。例如，丙氨酸 2,3-氨基变位酶核酸分子可包含丙氨酸 2,3-氨基变位酶核酸序列(例如 SEQ ID NO: 18, 20, 42, 44, 46, 48 或 50)的至少 15 个连续核苷酸。应该理解，本发明还提供了包含至少 16 个，至少 17 个，至少 18 个，至少 19 个，至少 20 个，至少 21 个，至少 22 个，至少 23 个，至少 24 个，至少 25 个，至少 26 个，至少 27 个，至少 28 个，至少 29 个，至少 30 个，至少 40 个，至少 50 个，至少 75 个，至少 100 个，至少 200 个，至少 500 个，至少 1000 个，至少 1200 个或更多个核苷酸的分离核酸分子，例如 SEQ ID NO: 18, 20, 42, 44, 46, 48 或 50 中的至少那么

多连续核苷酸。在一些实施例中，SEQ ID NO: 18, 20, 42, 44, 46, 48 或 50 (或其互补链)的片段并不编码具有丙氨酸 2,3-氨基变位酶的蛋白，但却是可用作探针或引物的较短片段。

本发明公开了变体丙氨酸 2,3-氨基变位酶的核酸序列。只要所编码的多肽保留丙氨酸 2,3-氨基变位酶活性(或可作为探针或引物行使功能)，变体就可含有单插入，单缺失，单取代，多插入，多缺失，多取代，或其任意组合(例如单缺失和多插入)。只要所述核酸编码的肽保留了所希望的酶活性，例如丙氨酸 2,3-氨基变位酶活性，这样的分离核酸分子就可与丙氨酸 2,3-氨基变位酶序列(例如 SEQ ID NO: 18, 20, 42, 44, 46, 48 或 50)具有至少 60%，至少 70，至少 75，至少 80，至少 85，至少 90，至少 95，至少 97，至少 98 或至少 99%的序列同一性。

例如，可将以下变体制备成丙氨酸 2,3-氨基变位酶核酸序列：对 SEQ ID NO: 13, 18 或 20 而言，将 96 或 384 位点上的"t"用"c" "a"或"g"取代；将 438 或 480 位点上的"a"用"c" "t"或"g"取代；将 1056 位点上的"t"用"g" "a"或"c"取代；对 SEQ ID NO: 42, 44 或 46 而言，144 位点上的"a"可用"g"取代；672 位点上的"a"可用"c" "t"或"g"取代；而 1107 位点上的"t"；可用"c"取代；对 SEQ ID NO: 48 和 50 而言，576 位点上的"a"可用"g" "t"或"c"取代；864 位点上的"c"可用"t"取代；而 1200 位点上的"t"；可用"g" "a"或"c"取代。使用序列表，还可对本发明公开的其他丙氨酸 2,3-氨基变位酶氨基酸序列进行类似的取代。

可利用遗传编码的简并性来改变丙氨酸 2,3-氨基变位酶序列的编码区对编码序列进行改变，即，虽然对所述核酸序列进行了显著的改变，但其却编码了与天然氨基酸序列一致或基本上相似的氨基酸序列多肽。例如，可使用针对特定物种的密码子偏好性和密码子使用表，对分离核酸分子进行工程化，所述工程化利用了该特定物种密码子使用的偏好性。由于遗传密码的简并性，丙氨酸可由四种核苷酸密码子三联体编码：GCT, GCA, GCC 和 GCG。因此，可在不影响所述编码

多肽的氨基酸序列或所述多肽性质的条件下，在丙氨酸位点上对任意的这些密码子的开放阅读框的核酸序列进行改变。基于遗传密码的简并性，可使用本发明所述的标准 DNA 突变技术，或核酸序列的合成从核酸序列中衍生出核酸变体。因此，本发明还包括了编码相同多肽，但由于密码子简并性在核酸序列上有差异的核酸分子。因此，可对本发明公开的丙氨酸 2,3-氨基变位酶进行设计，从而使其具有特殊感兴趣生物体所优选使用的密码子(例如，如以下实施例所述)。

本发明还包括了编码丙氨酸 2,3-氨基变位酶的变体，融合体和片段的核酸(例如那些以上所公开的内容)。本发明还提供了编码丙氨酸 2,3-氨基变位酶的分离的核酸序列，其中所述序列的长度为至少 12 个核苷酸(例如长度为至少 13 个，至少 14 个，至少 15 个，至少 16 个，至少 17 个，至少 18 个，至少 19 个，至少 20 个，至少 25 个，至少 30 个，至少 40 个，至少 50 个，至少 60 个，至少 100 个，至少 250 个，至少 500 个，至少 750 个，至少 1000 个，至少 1200 个或至少 1500 个核苷酸)，并可在杂交条件下，杂交于编码所述酶的核酸分子的有义链或反义链。杂交条件可以是温和或较高的严谨杂交条件。

可通过标准 DNA 突变技术，例如，M13 引物突变，制备丙氨酸 2,3-氨基变位酶多肽和编码这些肽的核酸分子。这些技术的详细内容可参见 Sambrook 等人 (ed.), *Molecular Cloning: A Laboratory Manual* 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring, Harbor, N.Y., 1989, Ch. 15。核酸分子可含有编码区的改变从而适合该分子所导入的特定生物体的密码子使用偏好。

具有丙氨酸 2,3-氨基变位酶活性的细胞

本发明公开了具有丙氨酸 2,3-氨基变位酶活性的细胞。这些细胞可从 α -丙氨酸制备 β -丙氨酸。在一个实施例中，由于自然发生的突变或在该细胞染色体中诱导的突变，例如将所述细胞暴露于化学或 UV 诱变条件，这些细胞会具有丙氨酸 2,3-氨基变位酶活性。包括丙氨酸 2,3-

氨基变位酶活性的细胞可以是真核或原核细胞。这些细胞的示例包括，但不限于，乳杆菌(*Lactobacillus*)，乳球菌(*Lactococcus*)，芽孢杆菌，*Escherichia*，*Geobacillus*，*Corynebacterium*，梭菌，真菌，植物和酵母细胞。在一个实施例中，植物细胞是植物，例如转基因植物的一部分。

在一个实施例中，具有丙氨酸 2,3-氨基变位酶活性的细胞是转化细胞。这些细胞可包括编码丙氨酸 2,3-氨基变位酶的至少一种外源核酸分子，例如保留了编码具有丙氨酸 2,3-氨基变位酶活性蛋白能力的 SEQ ID NO: 18, 20, 42, 44, 46, 48 或 50，或其变体，片段或融合体的序列。在一个实施例中，所述外源核酸分子是突变的赖氨酸 2,3-氨基变位酶，例如突变的原核赖氨酸 2,3-氨基变位酶。在特定的实施例中，突变的原核赖氨酸 2,3-氨基变位酶是突变的枯草芽孢杆菌(*Bacillus subtilis*)，牙龈卟啉单胞菌(*Porphyromonas gingivalis*)，具核梭杆菌(*Fusobacterium nucleatum*)或斯氏梭菌(*Clostridium sticklandii*)赖氨酸 2,3-氨基变位酶。可使用本领域公知的方法，例如，通过在 BLAST 上检索类似序列或通过杂交方法来鉴定其他赖氨酸 2,3-氨基变位酶。在特定的实施例中，突变的赖氨酸 2,3-氨基变位酶是突变的枯草芽孢杆菌，具核梭杆菌，牙龈卟啉单胞菌或斯氏梭菌的赖氨酸 2,3-氨基变位酶。在特定的实施例中，突变的赖氨酸 2,3-氨基变位酶是在 E30K; K36E; Y63F; Q86R; L95M; M128V; D144G; D179N; L228M; D331H 或 K398E 位点上具有取代的突变具核梭杆菌赖氨酸 2,3-氨基变位酶。在又一个实施例中，突变的赖氨酸 2,3-氨基变位酶是在 P/S11T; E30K; V32A; C52R; L53P/H; E71G; Q95M; Q125L; M128V; Q141R; K87R 或 D331G/H 位点上具有取代的突变斯氏梭菌赖氨酸 2,3-氨基变位酶。在又一个实施例中，突变的赖氨酸 2,3-氨基变位酶是在 N19Y; L26I; E30K; L53P/H; H85Q; I192V; D331G/H 或 M342T 位点上具有取代的突变牙龈卟啉单胞菌的赖氨酸 2,3-氨基变位酶。在所有这些实施例中，在特定赖氨酸 2,3-氨基变位酶中的取代可以表现为单独的，或其任意组合。

本发明还公开了包括丙氨酸 2,3-氨基变位酶活性以及其他酶活性

的细胞。这些细胞可用于制备 β -丙氨酸, 3-HP, 泛酸盐, CoA, 和有机酸, 聚醇, 例如 1,3-丙二醇, 聚合的 3-HP, 3-HP 的共聚物和其他化合物, 例如丁酸酯, 戊酸酯, 和其他化合物, 以及 3-HP 的酯。

例如, 本发明公开了能够从 3-氧丙酸盐产生 3-羟丙酸盐、具有丙氨酸 2,3 氨基变位酶活性以及丙酮酸盐/2-酮戊二酸盐氨基转移酶活性, β -丙氨酸/2-酮戊二酸盐氨基转移酶活性, 和 3-羟丙酸盐脱氢酶活性的细胞。在这些实施例中, 所述细胞可制备 3-HP。在其他实施例中, 这些细胞还包括脂酶或酯酶活性。这些细胞可用于制备 3-HP 的酯, 例如, 3-羟丙酸甲酯, 3-羟丙酸乙酯, 3-羟丙酸丙酯, 3-羟丙酸丁酯或 3-羟丙酸-2-乙基己基酯。

在其他实施例中, 具有丙氨酸 2,3 氨基变位酶活性的细胞还包括丙酮酸盐/2-酮戊二酸盐氨基转移酶活性, β -丙氨酸/2-酮戊二酸盐氨基转移酶活性, 和 3-羟丙酸盐脱氢酶活性; 以及聚羟酸合成酶活性, 这些细胞可用于制备聚合 3-HP。

或者, 具有丙氨酸 2,3 氨基变位酶活性的细胞还包括丙酮酸盐/2-酮戊二酸盐氨基转移酶活性, β -丙氨酸/2-酮戊二酸盐氨基转移酶活性, 3-羟丙酸盐脱氢酶活性和乙醇脱氢酶活性。这些细胞可用于制备聚 1,3-丙二醇。

在一个实施例中, 具有丙氨酸 2,3 氨基变位酶活性的细胞还具有 α -酮泛解酸盐羟甲基转移酶(E.C. 2.1.2.11), α -酮泛解酸盐还原酶(E.C. 1.1.1.169), 和泛酸盐合酶(E.C. 6.3.2.1)活性。这样的细胞可用于制备泛酸盐。或者, 细胞也具有泛酸盐激酶(E.C. 2.7.1.33), 4'-磷酸羟泛酰基-1-半胱氨酸合成酶(E.C. 6.3.2.5), 4'-磷酸泛酰半胱氨酸脱羧酶(E.C. 4.1.1.36), ATP:4'-磷酸泛酰巯基乙胺腺嘌呤基转移酶(E.C. 2.7.7.3), 和脱磷酸-CoA 激酶(E.C. 2.7.1.24)。这些细胞可用于制备辅酶 A(CoA)。

可通过表达编码具有所希望活性的酶的核酸分子，或通过直接供应所述酶来提供酶活性。

鉴定细胞具有丙氨酸 2,3-氨基变位酶活性的方法

本发明公开了鉴定细胞具有丙氨酸 2,3-氨基变位酶活性的方法。所述方法包括在含有 α -丙氨酸，但不含有 β -丙氨酸或泛酸盐的培养基中，培养功能性缺失 panD 的原核细胞，或在不含有 β -丙氨酸或泛酸盐，细胞却能从碳源、氧源、氢源和氮源制备 α -丙氨酸的培养基中培养功能性缺失 panD 的原核细胞，并且鉴定能够生长于 β -丙氨酸或泛酸盐缺陷培养基中的细胞。在特定的实施例中，还对该细胞进行了 panF 的功能性缺失，从而使得所述细胞不能从补充有泛酸盐的培养基中摄取泛酸盐。所述细胞的生长说明该细胞可从 α -丙氨酸制备 β -丙氨酸，这说明了所述细胞具有丙氨酸 2,3-氨基变位酶活性。相反，如果细胞不能在 β -丙氨酸或泛酸盐缺陷培养基中生长或存活，说明该细胞不能从 α -丙氨酸制备 β -丙氨酸，这说明了所述细胞不具有丙氨酸 2,3-氨基变位酶活性。

在一个实施例中，可使用一种或多种突变的氨基变位酶，例如包括突变赖氨酸 2,3-氨基变位酶的文库对功能性缺失 panD 的细胞进行转化。在特定的实施例中，在培养和筛选细胞之前，可用突变的赖氨酸 2,3-氨基变位酶的文库转化细胞。已对来自 *Clostridium subterminale* SB4 (Chirpich 等人, J. Biol. Chem. 245:1778-89, 1970)和枯草芽孢杆菌 (*Bacillus subtilis*)(Chen 等人, Biochem. J. 348:539-49, 2000)的赖氨酸 2,3-氨基变位酶进行了描述，并证明其可催化赖氨酸与 β -赖氨酸的相互转化。氨基变位酶突变体，例如赖氨酸 2,3-氨基变位酶突变体，可根据其具有的丙氨酸 2,3-氨基变位酶活性进行筛选。此外，虽然之前未对具有丙氨酸 2,3-氨基变位酶活性的多肽进行描述，这样的酶也有可能存在于自然界中。因此，可使用包括编码丙氨酸 2,3-氨基变位酶基因，以及通过能使细胞在含有 α -丙氨酸，或碳源，氧源，氢源和氮源，但不含有 β -丙氨酸或泛酸盐培养基中的这种细胞传递生长能力的分离的基因

的文库来转化功能性缺失 panD 的细胞,可使得所述细胞产生 α -丙氨酸。

在其他实施例中,本方法还包括在鉴定细胞生长于所述培养基之后鉴定突变氨基变位酶中的突变,其中所述突变的氨基变位酶可使细胞具有丙氨酸 2,3-氨基变位酶活性。为了鉴定突变,可对所述氨基变位酶核酸或氨基酸进行测序,并与未突变的氨基变位酶序列进行比较,从而鉴定出使细胞具有丙氨酸 2,3-氨基变位酶活性的突变。这些突变随后被转移给其他的赖氨酸 2,3-氨基变位酶,如图 7 所示,从而产生具有丙氨酸 2,3-氨基变位酶活性的氨基变位酶变体。

制备含有丙氨酸 2,3-氨基变位酶活性肽的方法

本发明公开了制备具有丙氨酸 2,3-氨基变位酶活性肽的方法。本发明包括在允许该细胞产生丙氨酸 2,3-氨基变位酶肽的条件下培养具有丙氨酸 2,3-氨基变位酶活性的已公开的细胞。在一个实施例中,该方法包括培养具有一个或多个编码丙氨酸 2,3-氨基变位酶(例如包括保留了丙氨酸 2,3-氨基变位酶活性的 SEQ ID NO: 18, 20, 42, 44, 46, 48 或 50, 或其变体,融合体或片段)外源核酸分子的细胞,从而制备丙氨酸 2,3-氨基变位酶。

本发明公开了从 α -丙氨酸制备 β -丙氨酸的方法。在一个实施例中,该方法包括在允许该细胞从 α -丙氨酸制备 β -丙氨酸的条件下培养具有丙氨酸 2,3-氨基变位酶活性的已公开的细胞。在一个实施例中,该方法包括培养具有一个或多个编码丙氨酸 2,3-氨基变位酶的外源核酸分子的细胞,从而制备出可从 α -丙氨酸制成 β -丙氨酸的丙氨酸 2,3-氨基变位酶。在一个实施例中,所述外源核酸是这样的序列,其包括保留了丙氨酸 2,3-氨基变位酶活性的 SEQ ID NO: 18, 20, 42, 44, 46, 48 或 50, 或其变体,融合体或片段。

在特定的实施例中,该细胞功能性缺失了 panD, 或 panD 和 panF。

制备 3-HP, 泛酸盐及其衍生物的途径

本发明公开了通过丙氨酸 2,3-氨基变位酶, 例如使用公开的丙氨酸 2,3-氨基变位酶序列以及公开的具有丙氨酸 2,3-氨基变位酶活性的细胞从 α -丙氨酸制备 β -丙氨酸的相关方法和材料。此外, 本发明还公开了从 β -丙氨酸制备泛酸盐和 3-HP, 以及 CoA 和有机物, 例如 1,3-丙二醇, 聚合的 3-HP, 3-HP 的共聚物和其他化合物, 例如丁酸酯, 戊酸酯或其他化合物, 以及 3-HP 的酯的方法和相关材料。具体地, 本发明提供了丙氨酸 2,3-氨基变位酶核酸分子(例如 SEQ ID NO: 18, 20, 42, 44, 46, 48 或 50), 肽(例如 SEQ ID NO: 2, 4, 6, 19, 21, 43, 45, 47, 49 或 51), 宿主细胞以及从 α -丙氨酸制备 β -丙氨酸的方法及材料, 其用来更有效地制备 β -丙氨酸, 泛酸盐和 3-HP, 及其衍生物, 例如 CoA 和有机物, 例如 1,3-丙二醇, 聚合的 3-HP, 和 3-HP 的酯。

还可使用若干代谢途径从由 α -丙氨酸制备得到的 β -丙氨酸来制备有机化合物(图 1 和图 3)。

3-HP 及其衍生物的途径

如图 1 所示, 可使用具有 β -丙氨酸/2-酮戊二酸盐氨基转移酶活性的多肽来从 β -丙氨酸制备 3-HP, 该多肽能从 β -丙氨酸生成 3-氧丙酸盐。可使用具有 3-HP 脱氢酶(EC 1.1.1.59 或.31)活性的多肽将 3-氧丙酸盐转化为 3-HP。

从 β -丙氨酸制备 3-HP 的衍生物可参见图 1 所示。可通过使用具有聚羟酸合成酶活性(EC 2.3.1.-)的多肽将所得的 3-HP 转化成聚合的 3-HP。或者, 可使用具有氧化还原酶活性或还原酶活性的多肽将 3-HP 转化成 1,3-丙二醇。

可使用具有脂酶或酯酶活性(EC 3.1.1.-)的多肽将所得的 3-HP 转化成 3-HP 的酯。或者, 可使用具有醛脱氢酶活性的多肽和具有醇脱氢酶活性的多肽组合, 从 3-HP 制备 1,3-丙二醇。

泛酸盐及其衍生物的途径

如图 3 所示, 可使用具有 α -酮泛解酸盐羟甲基转移酶(E. C. 2.1.2.11), α -酮泛解酸盐还原酶(E.C. 1.1.1.169), 和泛酸盐合酶(E.C. 6.3.2.1)活性、能将 β -丙氨酸转化为泛酸盐的肽从 β -丙氨酸制备泛酸盐。

可根据如下所述从 β -丙氨酸制备泛酸盐的衍生物。可使用具有泛酸盐激酶(E.C. 2.7.1.33), 4'-磷酸羟泛酰基-1-半胱氨酸合成酶(E.C. 6.3.2.5), 4'-磷酸泛酰半胱氨酸脱羧酶(E.C. 4.1.1.36), ATP:4'-磷酸泛酰巯基乙胺腺嘌呤基转移酶(E.C. 2.7.7.3), 和脱磷酸-CoA 激酶(E.C. 2.7.1.24)活性的多肽将所得的泛酸盐转化成 CoA。

酶

可从多种物种得到具有赖氨酸 2,3-氨基变位酶活性的多肽以及编码这种多肽的核酸, 这些物种包括, 但不限于: *C. subterminale*, 大肠杆菌, 枯草芽孢杆菌, 牙龈卟啉单胞菌, 具核梭杆菌, 和斯氏梭菌。例如, 对牙龈卟啉单胞菌而言, 具有赖氨酸 2,3-氨基变位酶活性的氨基酸序列可见于 SEQ ID NO: 52, 对斯氏梭菌而言, 可见于 SEQ ID NO: 33; 对枯草芽孢杆菌而言, 可见于 SEQ ID NO: 59; 而对具核梭杆菌而言, 则可见于 SEQ ID NO: 10。

本发明公开了编码具有丙氨酸 2,3-氨基变位酶活性肽的核酸分子。其示例包括, 但不限于, 牙龈卟啉单胞菌的 SEQ ID NOS: 48 和 50 (对应的氨基酸序列可参见 SEQ ID NOS: 49 和 51), 具核梭杆菌的 SEQ ID NOS: 18 和 20(对应的氨基酸序列可参见 SEQ ID NOS: 19 和 21), 和斯氏梭菌的 SEQ ID NOS: 42, 44 和 46(对应的氨基酸序列可参见 SEQ ID NOS: 43, 45 和 47)。此外, 也可使用本发明所述的方法获得其他具有丙氨酸 2,3-氨基变位酶活性的肽, 以及编码这些多肽的核酸。例如, 丙氨酸 2,3-氨基变位酶变体可编码具有上述丙氨酸 2,3-氨基变位酶活性的肽。

可从多个物种中获得具有 β -丙氨酸/2-酮戊二酸盐氨基转移酶活性, 3-羟丙酸盐脱氢酶活性的多肽, 以及编码这些多肽的核酸。

可从多个物种中获得具有聚羟酸合成酶活性的多肽以及编码这些多肽的核酸, 这些物种包括, 但不限于, *Rhodobacter sphaeroides*, *Comamonas acidovorans*, *Ralstonia eutropha* 和 *Pseudomonas oleovorans*。例如, 可从 *R. sphaeroides* 获得编码具有聚羟酸合酶活性多肽的核酸, 且其具有如 GenBank accession number X97200 所列出的序列。关于聚羟酸合酶的其他信息可参见 Song 等人(*Biomacromolecules* 1:433-9, 2000)。

如上所述, 可通过两种不同的多肽来实施醛: NAD(+)氧化还原酶活性和醇 NAD(+)氧化还原酶活性, 或通过单种多肽来实施, 例如来自于大肠杆菌的多功能醛-醇脱氢酶(EC 1.2.1.10)(Goodlove 等人 *Gene*85:209-14, 1989; GenBank Accession No. M33504)。

可从多个物种, 包括, 但不限于, 啤酒酵母中获得具有醛脱氢酶(NAD(P)+)(EC 1.2.1.-)活性的多肽以及编码这些多肽的核酸。例如, 可从啤酒酵母中获得编码具有乙醛脱氢酶活性 d 多肽的核酸, 且其具有如 GenBank Accession No. Z75282 所列出的序列(Tessier 等人 *FEMS Microbiol. Lett.* 164:29-34, 1998)。

可从多个物种, 包括, 但不限于, 运动发酵单胞菌(*Z. mobilis*)中获得具有醇脱氢酶(EC 1.1.1.1)活性的多肽以及编码这些多肽的核酸。例如, 可从运动发酵单胞菌(*Z. mobilis*)中获得编码具有醇脱氢酶活性多肽的核酸, 且其具有如 GenBank accession No. M32100 所列出的序列。

可从多个物种, 包括, 但不限于, 铁红假丝酵母(*Candida rugosa*), 热带假丝酵母(*Candida tropicalis*), 和白假丝酵母(*Candida albicans*)中获

得具有脂酶活性的多肽以及编码这些多肽的核酸。例如，可从铁红假丝酵母获得编码具有脂酶活性多肽的核酸，且其具有如 GenBank accession No. A81171 所列出的序列。

可从多个物种，包括，但不限于大肠杆菌中获得具有 α -酮泛解酸盐羟甲基转移酶和泛酸合酶活性的多肽以及编码这些多肽的核酸。例如，可从大肠杆菌中获得编码具有 α -酮泛解酸盐羟甲基转移酶和泛酸盐合酶活性肽的核酸，且其具有如 GenBank accession No. L17086 所列出的序列。

可从多个物种，包括，但不限于，大肠杆菌中获得具有 α -酮泛解酸盐还原酶，泛酸盐激酶，4'-磷酸羟泛酰基-1-半胱氨酸合成酶，4'-磷酸泛酰半胱氨酸脱羧酶，ATP:4'-磷酸泛酰巯基乙胺腺嘌呤基转移酶，和脱磷酸-CoA 激酶活性的多肽以及编码这些多肽的核酸。例如，可从大肠杆菌中获得编码具有 α -酮泛解酸盐还原酶，泛酸盐激酶，4'-磷酸羟泛酰基-1-半胱氨酸合成酶，4'-磷酸泛酰半胱氨酸脱羧酶，ATP:4'-磷酸泛酰巯基乙胺腺嘌呤基转移酶，和脱磷酸-CoA 激酶活性多肽的核酸，且其具有如 GenBank accession No. NC000913 所列出的序列。

术语“具有酶活性的多肽”是指在反应完成后自身不受到破坏或改变的条件下，能够催化其他物质进行化学反应的任意多肽。一般地，具有酶活性的多肽催化从一种或多种底物形成一种或多种产物。这些多肽可具有任意类型的酶活性，包括，但不限于，与诸如丙氨酸 2,3-氨基变位酶，聚羟酸合酶， β -丙氨酸/2-酮戊二酸盐氨基转移酶，3-羟丙酸盐脱氢酶，脂酶，酯酶，乙酰醛：NAD(+)氧化还原酶，醇：NAD(+)氧化还原酶，醛脱氢酶，醇脱氢酶，合酶，合成酶，脱羧酶， α -酮泛解酸盐羟甲基转移酶， α -酮泛解酸盐还原酶，泛酸盐合酶，泛酸盐激酶，4'-磷酸羟泛酰基-1-半胱氨酸合成酶，4'-磷酸泛酰半胱氨酸脱羧酶，ATP:4'-磷酸泛酰巯基乙胺腺嘌呤基转移酶，和脱磷酸-CoA 激酶，乙酰醛：NAD(+)氧化还原酶，醇：NAD(+)氧化还原酶，醛脱氢酶(NAD(P)+)

和醇脱氢酶的酶活性或与之关联的酶活性。

制备 3-HP，泛酸盐及其衍生物的方法

可在细胞内(体内)或细胞外(体外，例如在容器或柱子中)实施如图 1 和 3 所示途径中的每一步。此外，可通过体内合成与体外合成的组合生产有机产物。而且，通过化学反应或酶反应实施体外合成的步骤。

例如，本说明书中提供的细胞或微生物可用于实现图 1 和 3 所描述的步骤，或者可将含有指定酶活性多肽的提取物用于实施图 1 和 3 所描述的步骤。此外，可使用化学处理来实施图 1 和 3 所提供的转化。例如，可通过化学氢化将 3-氧基丙酸酯转化为 3-HP。其他的化学处理包括，但不限于，将 3-HP 转化为 3-HP 酯的酯交换反应。

多肽的表达

可在宿主细胞或宿主细胞组合中单独地制备本发明中所述的多肽，例如图 1 中所列示的酶。而且，具有特殊酶活性的肽可以是天然发生的多肽或非天然发生的肽。天然发生的肽是具有如自然界中发现氨基酸序列的任意肽，包括野生型和聚合多肽。可从任意物种，包括，但不限于，动物(例如哺乳动物)，植物，真菌，和细菌物种中获得天然发生的肽。非天然发生的多肽是不具有如自然界中发现氨基酸序列的任意多肽。因此，非天然发生的多肽可以是天然发生多肽的突变版本，或是工程化的多肽。例如，具有丙氨酸 2,3-氨基变位酶活性的非天然发生多肽可以是具有至少一些丙氨酸 2,3-氨基变位酶活性(例如 SEQ ID NO:19, 21, 43, 45, 47, 49 或 51)的赖氨酸 2,3-氨基变位酶活性的天然发生多肽的突变版本。可使用本领域的公知方法，通过，例如，序列添加，缺失，取代或其组合对多肽进行突变。

可将遗传修饰的细胞用于实施本发明所述途径中的一个或多个步骤，或可将遗传修饰的细胞用于制备在体外进行后续使用的公开多肽。例如，单个微生物可含有各个编码实施图 1 和 3 所述步骤必须多肽的

外源核酸。这些细胞可含有任意数量的外源核酸分子。例如，一个特定的细胞可含有一个，两个，三个或四个不同的外源核酸分子，其中的每一个都编码如图 1 所示的将丙酮酸转化为 3-HP 所必需的多肽，或者某个特定的细胞可内源性地产生将丙酮酸盐转化为 α -丙氨酸所必需的多肽，同时还含有编码将 α -丙氨酸转化为 3-HP 所必需多肽的外源核酸。

此外，单个的外源核酸分子可编码一个或多个的多肽。例如，单个的外源核酸分子可含有编码两个，三个，或甚至四个不同多肽的序列。而且，本说明书中的细胞可含有单拷贝，或多拷贝(例如约 5, 10, 20, 35, 50, 75, 100 或 150 个拷贝)的特定外源核酸分子，例如特定的酶。本发明所述的细胞可含有不止一种特定的外源核酸。例如，特定的细胞可含有约 50 拷贝的外源核酸分子 X 以及约 75 拷贝的外源核酸分子 Y。

在另一个实施例中，细胞可含有编码具有丙氨酸 2,3-氨基变位酶活性多肽，例如 SEQ ID NO:18, 20, 42, 44, 46, 48 或 50 的外源核酸分子。这些细胞可具有任意可检测水平的丙氨酸 2,3-氨基变位酶活性，包括可通过产生 β -丙氨酸的代谢产物，例如泛酸盐检测的活性。例如，含有编码具有丙氨酸 2,3-氨基变位酶活性的多肽的外源核酸分子的细胞可具有比活性为每小时每克干细胞重形成的 β -丙氨酸大于 $1\mu\text{g}$ (例如每小时每克干细胞重形成的 β -丙氨酸大于约 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 200, 250, 300, 350, 400, 500 或更多 μg)的丙氨酸 2,3-氨基变位酶活性。或者，细胞可具有的丙氨酸 2,3-氨基变位酶活性使细胞提取物可具有每分钟每毫克总蛋白形成的 β -丙氨酸大于约 1ng 的比活性(例如每分钟每毫克总蛋白形成的 β -丙氨酸大于约 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 200, 250, 300, 350, 400, 500 或更多 ng)。

可使用本发明中的任意方法鉴定和获得编码具有酶活性的多肽的

核酸分子。例如，可使用常规的分子克隆或化学核酸合成方法和技术，包括 PCR 来鉴定和获得编码具有酶活性多肽的核酸分子。此外，也可使用标准核酸测序技术和基于遗传密码将核酸序列翻译成氨基酸序列的软件程序来确定，特定的核酸是否具有与公知的具有酶活性的多肽任意的同源性。可使用诸如 MEGALIGN(DNASTAR, Madison, WI, 1997) 的序列比对软件对不同的序列进行比对。

此外，可使用常规的分子克隆技术(例如位点定向突变)对编码已知具有酶活性多肽的核酸分子进行突变。可能的突变包括，但不限于，缺失，插入，和取代，以及缺失，插入和核苷酸取代的组合。而且，可使用核酸和氨基酸数据库(例如，GenBank 或 EMBL)来鉴定编码具有酶活性多肽的核酸序列。简言之，与具有酶活性的多肽具有一些同源性的任意氨基酸序列，或与编码具有酶活性的多肽的序列具有一些同源性的任意核酸序列都可用作检索项在 GenBank 中进行检索。随后可对鉴定的多肽进行分析从而确定其是否表现出了酶活性。

此外，可使用核酸杂交技术来鉴定和获得编码具有酶活性的多肽的核酸分子。简言之，任意编码已知酶活性多肽的核酸分子，或其片段，都可用作探针从而在温和至严谨条件下通过杂交鉴定类似的核酸分子。然后可将这些类似的核酸分子分离，测序，并进行分析从而确定所编码的多肽是否具有酶活性。

还可使用表达克隆技术来鉴定和获得编码具有酶活性的多肽的核酸分子。例如，可使用公知与特定酶活性多肽相互作用的底物来筛选含有该酶活性多肽的噬菌体显示文库。噬菌体显示文库的制备可参见(Burritt 等人, Anal. Biochem. 238:1-13, 1990)，或可通过从诸如 Novagen(Madison, WI)的供应商处获得。

此外，也可使用多肽测序技术来鉴定和获得编码具有酶活性多肽的核酸分子。例如，可通过凝胶电泳对纯化的多肽进行分离，并通过

例如，氨基酸微测序技术测定其氨基酸序列。一旦完成测定，可使用所述氨基酸序列设计简并寡核苷酸引物。可使用简并寡核苷酸引物，通过 PCR 获得编码所述多肽的核酸。一旦得到，可对所述核酸进行测序，克隆入适当的表达载体，并导入微生物中。

可使用任意方法将外源核酸分子导入细胞。例如，热击，脂质体转染，电穿孔，结合，原生质体融合以及基因枪递送是向细菌和酵母细胞中导入核酸的常见方法(例如，参见 Ito 等人, *J. Bacteriol.* 153:163-8, 1983; Durrens 等人, *Curr. Genet.* 18:7-12, 1990; Sambrook 等人, *Molecular cloning: A laboratory manual*, Cold Spring Harbour Laboratory Press, New York, USA, second edition, 1989; 和 Becker and Guarente, *Methods in Enzymology* 194:182-7, 1991)。其他表达来自外源核酸分子的氨基酸序列的方法包括，但不限于，构建核酸，从而使得调节元件能够启动编码该多肽的核酸序列的表达。通常，调节元件是在转录水平调节其他 DNA 序列表达的 DNA 序列。因此，调节元件包括，但不限于，启动子，增强子等。可使用任意类型的启动子来表达来自外源核酸分子的氨基酸序列。启动子的示例包括，但不限于，组成型启动子，组织特异型启动子，和响应或不响应于特殊刺激(例如，光，氧气，化学浓度)的启动子。向哺乳动物细胞转移核酸的方法也是公知的，例如使用病毒载体。

包含于本发明特定细胞内的外源核酸分子可以任意形式维持于该细胞中。例如，外源核酸分子可整合入该细胞的基因组或以游离体状态维持。换言之，细胞可以是稳定的或瞬时的转化子。微生物可含有单个或多个拷贝(例如约 5, 10, 20, 35, 50, 75, 100 或 150 拷贝)的特定外源核酸分子，例如编码酶的核酸。

由宿主细胞制备有机酸和相关产物

可使用本发明提供的核酸和氨基酸序列以及细胞来制备 β -丙氨酸，泛酸盐和 3-HP，及其衍生物，例如 CoA，和有机物，例如 1,3-丙

二醇, 3-HP 的酯, 和聚合的 3-HP。这些细胞可来自于任意物种, 例如那些在 NIH 分类网页上所列出的物种。这些细胞可以是真核或原核的。例如, 遗传修饰的细胞可以是哺乳动物细胞(例如人, 鼠和牛细胞)植物细胞(例如玉米, 小麦, 水稻和大豆细胞), 真菌细胞(例如曲霉和根霉细胞), 酵母细胞, 或细菌细胞(例如, 乳杆菌(*Lactobacillus*), 乳球菌(*Lactococcus*), 芽孢杆菌, 埃希氏肠杆菌, 和梭菌细胞)。在一个实施例中, 细胞是微生物。术语“微生物”是指任意的显微生物体, 包括, 但不限于, 细菌, 藻类, 真菌和原生动物。因此, 大肠杆菌, 枯草芽孢杆菌, 地衣芽孢杆菌(*B. licheniformis*), 啤酒酵母(*S. cerevisiae*), 乳酸克鲁维酵母(*Kluveromyces lactis*), *Candida blankii*, 铁红假丝酵母(*Candida rugosa*)和巴斯德毕赤酵母(*Pichia pastoris*)是示例性的微生物, 并可进行如本发明所述的使用。在另一个实施例中, 所述细胞是诸如植物, 例如转基因植物的较大生物体的一部分。可从 β -丙氨酸制备 3-HP, 泛酸盐, 或其他有机物的植物示例包括, 但不限于, 遗传工程化的植物作物, 例如玉米, 水稻, 小麦和大豆。

在一个实施例中, 可对细胞进行遗传修饰从而获得特定的有机物。在一个实施方案中, 细胞可从 β -丙氨酸如图 1 和 3 所示途径制备 3-HP 或泛酸盐。在又一个实施例中, 所述细胞可制备 3-HP 或泛酸盐的衍生物, 例如 CoA, 和其他有机物, 例如 1,3-丙二醇, 3-HP 的酯, 和聚合的 3-HP。

可进行遗传修饰来合成特定有机物的细胞可包括一个或多个编码具有特定酶活性的多肽的核酸分子。例如, 微生物可含有编码具有 3- β -丙氨酸/2-酮戊二酸盐氨基转移酶活性的外源核酸。在这种情况下, β -丙氨酸可被转化成 2-氧基丙酸盐, 其可导致 3-HP 的产生。细胞可被给予编码具有酶活性多肽的外源核酸分子, 所述酶活性可催化产生通常该细胞不能生成的化合物。或者, 细胞可被给予编码具有酶活性多肽的外源核酸分子, 所述酶活性可催化产生通常该细胞可以生成的化合物。在这种情况下, 与没有进行遗传修饰的类似细胞相比, 所述遗传

修饰的细胞可生成更多的所述化合物，或可更有效地生成所述化合物。

在另一个实施例中，公开了含有编码具有酶活性多肽的外源核酸分子，其可形成 3-HP，泛酸盐，或其衍生物。所产生的产物可从细胞中分泌出来，从而免除了破坏细胞膜来回收所述有机物。在一个实施例中，所述细胞可产生 3-HP，泛酸盐，或其衍生物，其产物的浓度为至少约 100mg/L(例如，至少约 1g/L，5 g/L，10 g/L，25 g/L，50 g/L，75 g/L，80 g/L，90 g/L，100 g/L 或 120 g/L)。当对特定细胞所产生的诸如 3-HP，泛酸盐，和/或其衍生物的化合物产量进行确定时，可以使用任意的方(例如，参见 *Applied Environmental Microbiology* 59(12):4261-5, 1993)。本发明范围内的细胞可利用多种碳源。

细胞可含有一个或多个编码具有酶活性多肽的外源核酸分子，其可形成 3-HP，泛酸盐，及其衍生物，例如 CoA，1,3-丙二醇，3-HP 酯，和含有 3-HP 的聚合物和共聚物。鉴定细胞含有外源核酸分子的方法是公知的。这些方法包括，但不限于，PCR 和核酸杂交技术，例如 Northern 和 Southern 分析(参见本发明书中的杂交)。还可以使用免疫组织化学和生物化学技术，通过检测由特定核酸分子编码多肽的表达，来确定细胞是否含有特定的核酸分子。例如，可使用对多肽具有特异性的抗体来测定特定的细胞是否含有编码所述多肽的核酸。此外，可以使用生物化学技术，通过检测所产生的有机物是具有酶活性多肽表达的结果，来测定细胞是否含有编码具有酶活性多肽的特定核酸分子。例如，向通常不表达所述多肽的细胞中导入编码具有 3-羟丙酸盐脱氢酶活性多肽的外源核酸之后，所测定的 3-HP 说明了，所述细胞不仅含有了导入的外源核酸分子，而且还表达了由所导入外源核酸分子编码的多肽。测定特异性酶活性或特异性有机物存在的方法是公知的，例如，对诸如 3-HP 的有机物存在的测定可参见 Sullivan and Clarke (*J. Assoc. Offic. Agr. Chemists*, 38:514-8, 1955)。

具有降低多肽活性的细胞

本发明公开了具有降低多肽活性的遗传修饰的细胞。关于细胞和特定多肽活性，在此使用的术语“降低”或“下降”是指与相同种属可比较细胞相比，所测定的活性水平降低。例如，如果可比较的微生物具有至少一些酶活性 X，那么缺乏酶活性 X 的特定微生物就具有降低的酶活性 X。

细胞可具有任意类型降低的多肽活性，包括，但不限于，酶，转录因子，转运因子(transporter)，受体，信号分子等。例如，细胞可含有这样的外源核酸分子，该分子可破坏具有 panD 活性多肽的调节和编码序列。破坏 panD 可阻止细胞生成 β -丙氨酸。

降低的多肽活性还可以是降低的多肽浓度，降低的多肽特异活性或其组合的结果。可使用许多不同的方法使得细胞具有降低的多肽活性。例如，可使用常规的突变或敲除技术对细胞进行工程化，从而使其具有破坏的调节序列或多肽编码序列。(Methods in Yeast Genetics (1997 edition), Adams, Gottschling, Kaiser, and Sterns, Cold Spring Harbor Press, 1998; Datsenko and Wanner, Proc. Natl. Acad. Sci. USA 97: 6640-5, 2000)。或者，可使用反义技术来降低特定多肽的活性。例如，可对细胞进行工程化，从而使其含有编码能够抑制多肽翻译的反义分子。术语“反义分子”包括任意的核酸分子或核酸类似物(例如多肽核酸)，其含有对应于内源多肽编码链的序列。反义分子还可具有侧翼序列(例如调节序列)。因此，反义分子可以是核酶或反义寡核苷酸。核酶可具有任意的常规结构，包括，但不限于，发卡，锤头，或斧头结构，从而提供剪切 RNA 的分子。此外，还可使用基因沉默来降低特定多肽的活性。

可使用任意方法鉴定具有降低多肽活性的细胞。例如，可使用本发明所述的那些酶活性分析方法来鉴定具有降低的酶活性的细胞。

通过体外技术制备有机酸和相关产物

可单独使用具有酶活性的纯化多肽或与细胞组合使用来制备泛酸盐, 3-HP, 或其衍生物, 例如 CoA, 和有机物, 例如 1,3-丙二醇, 3-HP 的酯, 和聚合的 3-HP。例如, 可使用包括具有 3-羟丙酸盐脱氢酶活性的基本上纯的多肽制剂来催化 3-HP 的形成。

此外, 可单独使用含有酶活性多肽的无细胞提取物或与纯化多肽或细胞组合, 来制备 3-HP, 或其衍生物。例如, 可使用包括具有丙氨酸 2,3-氨基变位酶活性多肽的无细胞提取物来从 α -丙氨酸形成 β -丙氨酸, 而使用具有催化从 β -丙氨酸形成 3-HP 反应所必须酶活性多肽的微生物来从 β -丙氨酸制备 3-HP。在其他实施例中, 可使用含有 α -酮泛解酸盐羟甲基转移酶(E. C. 2.1.2.11), α -酮泛解酸盐还原酶(E.C. 1.1.1.169), 和泛酸盐合酶(E.C. 6.3.2.1)的无细胞提取物来从 β -丙氨酸形成泛酸盐。可使用任意的制备方法制备无细胞提取物。例如, 可使用渗透打击, 超声波, 或过滤后反复冻融和/或离心来从完整的细胞制备无细胞提取物。

可使用细胞, 纯化的多肽, 或无细胞提取物来制备 3-HP, 其随后, 被化学处理以制备其他化合物。例如, 可使用微生物来制备 3-HP, 而使用化学方法将 3-HP 修饰成衍生物, 例如聚合的 3-HP 或 3-HP 的酯。类似的, 可使用化学方法来制备特定的化合物, 随后, 使用本发明的细胞, 基本上纯的多肽或无细胞提取物将其转化为 3-HP 或其他有机物(例如 1,3-丙二醇, 丙烯酸, 聚丙烯酸酯, 丙烯酸酯, 3-HP 的酯, 和聚合 3-HP)。例如, 可使用化学方法来制备丙烯酰-CoA, 而可使用微生物将丙烯酰-CoA 转化为 3-HP。

类似的, 可使用细胞, 纯化的多肽, 或无细胞提取物来制备泛酸盐, 其随后, 被化学处理生成其他化合物。例如, 可使用微生物制备泛酸盐, 而使用化学方法将泛酸盐修饰成诸如 CoA 的衍生物。类似的, 可使用化学方法来制备特定的化合物, 随后, 使用本发明的细胞, 基本上纯的多肽, 或无细胞提取物将其转化为泛酸盐或其他化合物(例如

CoA)。例如，可使用化学方法制备泛酸盐，而使用微生物将泛酸转化为 CoA。

产生有机酸的细胞的发酵

本发明公开了通过在培养基中培养制备细胞，例如微生物，产生泛酸盐，3-HP，或其衍生物来制备泛酸盐，3-HP，或其衍生物的方法。通常，培养基和培养条件是使所述微生物生长至足够密度并有效产生所述产物的培养基和培养条件。对大规模制备方法而言，可使用在别处描述的任意方法 (Manual of Industrial Microbiology and Biotechnology, 2nd Edition, Editors: Demain and Davies, ASM Press; and Principles of Fermentation Technology, Stanbury and Whitaker, Pergamon)。

简言之，向含有适当培养基，例如，葡萄糖碳源的大罐(例如 100 加仑，200 加仑，500 加仑或更大的罐)中接种特定的微生物。接种后，对微生物进行孵育从而允许生产生物质。一旦达到所需要的生物质，可将含有所述微生物的肉汤转移至第二个罐中。第二个罐可具有任意尺寸。例如，第二个罐可以大于，小于，或与第一个罐具有相同的尺寸。通常，第二个罐比第一个要大，从而可向来自第一个罐的肉汤中加入额外的培养基。此外，在第二个罐内的培养基可以与用于第一个罐中的培养基相同或不同。例如，第一个罐可包含具有木糖的培养基，而第二个罐则可包含具有葡萄糖的培养基。

一旦转移完成，可对微生物进行孵育从而允许产生泛酸盐，3-HP 或其衍生物。一旦产生完成，可使用任意方法分离所形成的产物。例如，可使用常见的分离技术从肉汤中除去所述生物质，并可使用常见的分离方法(例如，提取，蒸馏，和离子交换方法)从无微生物的肉汤中获得泛酸盐，3-HP，或其衍生物。或者，可以在其产生时就对产物进行分离，或可在产物生成阶段完成之后再从肉汤中分离产物。

从公开的生物合成途径制备的产物

可将从图 1 和图 3 所提供的任意步骤制备的化合物化学转变为其他有机化合物。例如，可将 3-HP 脱氢形成 1,3-丙二醇，其是一种有价值的聚酯单体。可使用任意方法对诸如 3-HP 的有机酸进行氢化，就像那些用于对琥珀酸或乳酸进行氢化的方法。例如，可使用金属催化剂对 3-HP 进行氢化。在其他实施例中，3-HP 可被氢化形成丙烯酸。可使用任意的的方法实施氢化反应。例如，可在存在催化剂(例如金属或无机酸催化剂)的条件下加热 3-HP 形成丙烯酸。还可在体外或体内使用具有氧化还原酶活性(例如 EC 1.1.1.-类的酶)的多肽生成 1,3-丙二醇。

在其他实施例中，可使用泛酸盐形成辅酶 A。可使用具有泛酸盐激酶(E.C. 2.7.1.33)，4'-磷酸羟泛酰基-l-半胱氨酸合成酶(E.C. 6.3.2.5)，4'-磷酸泛酰半胱氨酸 脱羧酶(E.C. 4.1.1.36)，ATP:4'-磷酸泛酰巯基乙胺腺嘌呤基转移酶(E.C. 2.7.7.3)，和脱磷-CoA 激酶(E.C. 2.7.1.24)活性的多肽制备辅酶 A。

1,3-丙二醇的制备

本发明公开了制备 1,3-丙二醇的方法和用于此种制备的细胞。可从 3-HP-CoA 或 3-HP 制备 1,3-丙二醇。可对制备 3-HP-CoA 或 3-HP 的细胞或微生物进行工程化，通过克隆编码具有氧化还原酶/脱氢酶类活性酶的基因来制备 1,3-丙二醇。

例如，可在存在具有乙酰醛：NAD(+)氧化还原酶和醇：NAD(+)氧化还原酶活性酶的条件下将 3-HP-CoA 转化成 1,3-丙二醇。这些转化可在体内，体外或其组合条件下进行。可通过单种多肽或两种不同的多肽来实现这些活性。单种酶包括来自于大肠杆菌的多功能醛-醇脱氢酶(EC 1.2.1.10)(Goodlove 等人基因 85:209-14, 1989; GenBank Accession No. M33504)。还描述了具有单种乙酰醛：NAD(+)氧化还原酶(EC 1.2.1.10)或醇：NAD(+)氧化还原酶活性酶(EC 1.1.1.1)活性的酶。已经分离和测序了编码来自大肠杆菌的乙酰醛脱氢酶(GenBank Accession

No. Y09555)和来自运动发酵单胞菌(*Z. mobilis*)的醇脱氢酶(GenBank Accession No. M32100)的基因。可通过公知的分子生物学技术,将编码这些酶的基因克隆入产生 3-HP-CoA 的有机体或细胞中。这些酶在产生 3-HP-CoA 的生物体或细胞中的表达可以发挥将 3-HP-CoA 转化为 1,3-丙二醇的能力。可使用公知的技术,例如倾向错误的 PCR 或突变大肠杆菌菌株来改变或改善这些酶对于 3-HP-CoA 的底物特异性。

可通过将 3-HP 与具有醛脱氢酶(NAD(P)⁺)(EC 1.2.1.-)和醇脱氢酶(EC 1.1.1.1)活性的酶接触来实现 3-HP 向 1,3-丙二醇的转化。这样的转化可以在体内,体外,或其组合的条件下完成。例如,这些基因在产生 3-HP 微生物或细胞中的克隆和表达可以发挥所述细胞或微生物将 3-HP 转化为 1,3-丙二醇的活性。可使用如上所述的公知技术来改变或改善这些酶对于 3-HP 的底物特异性。

可使用高效液相色谱(HPLC)对发酵过程或体外化验中的 1,3-丙二醇形成进行分析。可使用 Bio-Rad 87H 离子交换柱完成色谱分离。以 0.6ml/分钟的流速通过 0.01N 硫酸的流动相,并将柱子的温度保持在 45-65°C。可使用折射率探测器测定样本中 1,3-丙二醇的存在(Skraly 等人, *Appl. Environ. Microbiol.* 64:98-105, 1998)。

实施例 1

对具核梭杆菌(*Fusobacterium nucleatum*)赖氨酸 2,3-氨基变位酶(kam 基因)的克隆和密码子改进

本实施例描述了用于从具核梭杆菌中克隆赖氨酸 2,3-氨基变位酶(E.C. 5.4.3.2),并随后对在大肠杆菌中表达基因的密码子使用进行优化的方法。本领域所属技术人员应该理解,类似的方法可用于从任意希望的生物体中克隆赖氨酸 2,3-氨基变位酶。

具核梭杆菌赖氨酸 2,3-氨基变位酶对赖氨酸具有较高的专一活性(Barker 等人 *J. Bacteriol.* 152(1):201-7, 1982)。具核梭杆菌可将赖氨酸

作为唯一碳源和氮源加以利用，因而需要赖氨酸降解途径中的酶具有较高的活性。为了克隆编码赖氨酸 2,3-氨基变位酶的具核梭杆菌 kam 基因，可使用以下方法。在 ATCC 培养基 1053 中(加强的梭菌培养基)扩增具核梭杆菌具核亚种(American Type Culture Collection, Manassas, VA, catalog number ATCC 25586)。通过在 Coy 厌氧舱中向所述培养基中通入厌氧气体来制备培养基(85% N₂, 5%CO₂, 10%H₂)。在密闭的厌氧管中在 37°C 进行培养。收集 10ml 培养物，并使用 Mekalanos 的方法分离基因组 DNA(Cell 35(1)-253-63, 1983)。

用于通过 PCR 扩增 kam 基因的引物是基于完整的具核梭杆菌基因组序列(GenBank Accession No: NC 003454)，还加上了限制性位点从而允许将所述 PCR 产物克隆入质粒中。使用 PCR 引物：CCGGCCCATATGAATACAGTTAATACTAG (SEQ ID NO:7) 和 CGCCGCGGATCCTTATTTAAACAATCTCTCCCTGTTCG (SEQ ID NO:8)来克隆入载体 pET11A (Novagen)的 NdeI 和 BamHI 位点。克隆的具核梭杆菌 kam 基因如 SEQ ID NO: 9 所示(其对应的氨基酸序列如 SEQ ID NO: 10 所示)。

为了在大肠杆菌中具有改善的表达，对具核梭杆菌 kam 基因进行了部分密码子优化。通过使用含有对精氨酸更优选密码子的引物，在扩增反应中通过整合替换了稀有的精氨酸密码子。两轮引物整合得到了变化密码子被替换的克隆。对两条 DNA 链所设计的引物长度为 22-23 个核苷酸，且在该引物中部具有一个或两个密码子替换。精氨酸的密码子 AGG, AGA 和 CGA 变为 CGT 或 CGC。为使得密码子优化最大化，可使用来自三个不同克隆的模板，和最小化扩增错误的校对 DNA 聚合酶来扩增 kam 基因的重叠片段。将所述片段凝胶纯化，并作为模板与该基因头尾同源的引物一起用于第二轮扩增。

在第二轮装配中，对最优化克隆的三个重叠片段进行扩增，从而确保用作扩增引物的引物实现了引物的整合。将所述片段凝胶纯化，

并作为模板与该基因头尾同源的引物一起用于第二轮扩增。使用针对这些引物所设计的限制性位点,将所述产物克隆入 pET11 载体(Novagen) 和 pPRO-Nde。质粒 pPRO-Nde 是 pPROLar.A122((Clontech Laboratones, Inc., Palo Alto, CA)的衍生物, 其中使用 Stratagene 的 QuikChange Site-Directed Mutagenesis 试剂盒通过寡核苷酸定向的突变在起始 ATG 密码子上构建了 NdeI 位点。在装配克隆中, 稀有精氨酸的密码子从 28 个降低至 8 个。部分优化的具核梭杆菌 kam 基因如 SEQ ID NO: 11 所示(其对应的氨基酸序列如 SEQ ID NO: 12 所示)。

本领域所属技术人员应该理解也可以使用其他的密码子优化方法, 例如使用重叠引物或多位点定向突变(Stratagene)的基因装配。

实施例 2

具核梭杆菌 kam 基因的体外突变

本实施例描述了用于对实施例 1 中所述部分优化具核梭杆菌 kam 基因(SEQ ID NO: 11)进行突变, 从而鉴定具有丙氨酸 2,3-氨基变位酶活性的突变体赖氨酸 2,3-氨基变位酶序列的方法。

通过定向的突变, 将之前在枯草芽孢杆菌(*Bacillus subtilis*)氨基变位酶(参见 WO 03/062173 和 SEQ ID NOS: 1-4)中获得的 3 个突变转移给了在具核梭杆菌密码子优化 kam 基因的同源位点。这些突变包括在 SEQ ID NO: 12 (其导致了如 SEQ ID NO: 14 所示序列)中的 L96M, M129V 和 D332H 取代, 其中第一个氨基酸是野生型序列, 该数字是氨基酸的位点, 而第二个氨基酸则是在丙氨酸 2,3-氨基变位酶中观察到的残基。可使用 Stratagene QuikChange[®] Multi Site-Directed Mutagenesis 试剂盒(Stratagene)产生这些突变。用于产生这些突变的引物是: FnL96M 5'Phos/CATCAATCTGATGCTGATATGTTGGATCCTCTACATGAAG (SEQ ID NO: 15) ; FnM129V 5'Phos/AACAGACATGTGTTCTGTATACTGTGCGCCACTGCACTC (SEQ ID NO: 16) ; 和 FnD332H

5'Phos/GTACCAACATTTGTTGTGCATGCACCTGGTGGTG (SEQ ID NO: 17)。具有三个定向突变(Fncodm)的部分密码子优化具核梭杆菌 kam 基因序列如 SEQ ID NO. 13 所示(而所得的蛋白质序列如 SEQ ID NO: 14 所示)。

在液体生长检测中对所得的 Fncodm 克隆进行检测。将重悬的克隆用于接种 2ml 玻璃管中的含有和不含有泛酸盐(25 μ M)的 1.4ml M9 基本培养基。向培养基中添加 0.4%葡萄糖, 2mg/ml L-丙氨酸, 100 μ M IPTG, 50 μ M Fe(NH₄)₂(SO₄)₂, 痕量元素, 和 25 μ g/mL 卡那霉素。当泛酸盐对照达到 OD₆₀₀ 为约 0.7 时, 读取培养物的 OD。通过比较无泛酸盐的生长情况和有泛酸盐的生长情况来测定表现, 发现 Fncodm 几乎没有或没有丙氨酸 2,3-氨基变位酶活性。

为了鉴定提高丙氨酸 2,3 氨基变位酶活性的突变, 使用倾向错误的 PCR 方法在体外向 Fncodm 基因(SEQ ID NO: 13)中引入随机突变。可使用类似的方法向编码赖氨酸 2,3 氨基变位酶的任意 kam 基因, 例如来自耐辐射奇球菌(*Deinococcus radiodurans*) (GenBank Accession No: RDR02336), 近端梭菌(*Clostridium subterminale*) (GenBank Accession No: AF159146)或牙龈卟啉单胞菌(*Porphyromonas gingivalis*)(不完全基因组, The Institute for Genomic Research)的 kam 基因中引入突变。

诱变 PCR 是根据 Cadwell 和 Joyce(PCR Methods Appl. 2:28-33,1992)的方法进行的。该方法使用了多种诱变缓冲液的稀释液, 所述缓冲液含有 21.2mM MgCl₂, 2.0 mM MnCl₂, 3.2 mM dTTP 和 3.2 mM dCTP。除了含有 1.5 mM MgCl₂, 0.25 μ M 正向和反向载体引物, 200 μ M 的各种 dNTP, 5 ng 的 pPRO-Fncodm 模板 DNA, 和 10 单位 Taq DNA 聚合酶(Roche)的 1X Taq PCR 缓冲液, 还向各个 PCR 反应中加入 6.25 和 9.38 μ L 诱变缓冲液。PCR 程序包括 94 $^{\circ}$ C 2 分钟的起始变性; 30 个循环的 94 $^{\circ}$ C 30 秒, 50 $^{\circ}$ C 1 分钟, 和 72 $^{\circ}$ C 2.25 分钟; 和 72 $^{\circ}$ C 7 分钟的最终延伸。

PCR 完成之后, 用限制性酶 NotI 和 NdeI 消化所述 PCR 产物。将每个处理的等量 DNA 连接于载体 pPRO-Nde, 并转化大肠杆菌 Electromax™ DH10B™ 细胞。从单克隆提取质粒 DNA 并测序从而获得对突变率的估计(0.57%)。将多次转化涂布于含有 25 μ g/ml 卡那霉素(LKB25)的 LB 平板获得约 53,000 个克隆。从平板上刮下克隆, 并使用 Qiagen MiniSpin Plasmid 方法制备质粒 DNA。在对选定宿主进行转化之前, 使用乙酸胺和乙醇沉淀质粒 DNA 从而增加其浓度。

实施例 3

具有丙氨酸 2,3-氨基变位酶活性的克隆的鉴定

本实施例描述了用于鉴定具有丙氨酸 2,3-氨基变位酶活性的突变赖氨酸 2,3-氨基变位酶克隆的方法。

之前已公开了鉴定具有丙氨酸 2,3-氨基变位酶活性细胞的方法(参见 WO 03/062173, 在此将其全文引用作为参考)。简言之, 该方法包括在不含有 β -丙氨酸或泛酸盐的培养基中培养功能性缺失 panD 的细胞。PanD 基因编码天冬氨酸盐脱羧酶, 其催化天冬氨酸向 β -丙氨酸的生产。大肠杆菌中 panD 的功能性缺失使得天冬氨酸盐脱羧酶失活, 从而导致了在制备辅酶 A 中由于对 β -丙氨酸的需求而导致的大肠杆菌生长抑制。能够在缺乏 β -丙氨酸和泛酸盐培养基中生长的细胞说明, 其可以从 α -丙氨酸制备 β -丙氨酸, 其说明了该细胞具有丙氨酸 2,3-氨基变位酶活性。

将在实施例 2 中产生的突变 Fncodm 文库转化大肠杆菌 Δ panD: : CAT 菌株的电击感受态细胞(参见 WO 03/062173 和本说明书中的实施例 10)。一小时之后在 SOC 培养基中回收转化子, 然后加入 100 μ M IPTG 和 25 μ g/ml 卡那霉素, 再生长 3 小时。然后将细胞离心, 用 0.85%NaCl 洗涤, 并重悬于添加了 0.4% 葡萄糖, 100 μ M IPTG, 50 μ M Fe(NH₄)₂(SO₄)₂, 2mg/ml α -L-丙氨酸, 痕量元素, 和 25 μ g/mL 卡那霉素

(Sigma, St. Louis, MO)的 500 μ L M9 基本培养基中。用 30 μ L 重悬细胞接种 2ml 玻璃管中的 1.33ml M9 培养基。6 天之后, 将生长的培养物接种液体 LBK25 培养基。生长 5.5 小时后, 提取质粒 DNA, 并用其转化 Δ panD: : CAT 菌株的新鲜感受态细胞。生长细胞的再次转化减少了能够由于宿主基因组中产生的突变而生长的突变体数, 并确保了所述生长是由于具有质粒的效果。

一小时之后在 SOC 培养基中回收转化子, 用 0.85%NaCl 洗涤, 并重悬于 1ml NaCl 中。对 LBK25 培养平板进行菌落计数, 并将剩余的重悬液(保存于 4 $^{\circ}$ C)涂布于 LBK25 平板从而获得每平板约 250 个克隆。将单个的克隆接种于 LBK25 和 M9 基本培养基。使用 2ml 管中的 1.4ml M9 基本培养基(如上所述), 在液体生长检测中对在 M9 培养基中生长旺盛的克隆进行检测。使用重悬的克隆来接种含有和不含有泛酸盐(25 μ M)的培养基。当泛酸盐对照达到 OD₆₀₀ 为约 0.7 时, 读取培养物的 OD。根据在没有泛酸盐中的生长比率高于在含有泛酸盐中的生长比率来鉴定克隆。使用标准分子生物学方法对生长旺盛克隆的质粒 DNA 进行测序。观察到所有的克隆都具有相同的序列(SEQ ID NO: 18, 而对应的氨基酸序列如 SEQ ID NO: 19 所示)。

编码丙氨酸 2,3-氨基变位酶(Fnnam)的突变具核梭杆菌 kam 基因序列如 SEQ ID NO. 18 所示, 而相应的氨基酸序列如 SEQ ID NO: 14 所示。除了在实施例 2 种产生的 3 个定向突变, 与具核梭杆菌 kam 基因序列(图 4)相比, 在突变的序列中还观察到了 6 个氨基酸改变。这些额外的取代是 E31K, Y64F, Q87R, D145G, L229M 和 K401E。然而, 所有这些突变都不是丙氨酸 2,3-氨基变位酶活性所必须的。

将 Fnaam 突变体的质粒 DNA 作为模板, 可根据以上描述构建第二轮诱变文库。因为起始克隆现已具有丙氨酸 2,3-氨基变位酶活性, 所以选择就更加依赖于液体富集技术而不是在固体培养板上涂布。采用液体富集, 可以使用 Δ panD/ Δ panF 选择宿主。panF 基因编码泛酸通透

酶，其能够允许泛酸的浓缩性摄取。PanF 缺失突变体的使用防止了失活克隆与具有活性丙氨酸 2,3-氨基变位酶所产生泛酸盐之间的可能交叉给料。将由约 48,000 克隆的 DNA 组成诱变文库用于转化 Δ panD/ Δ panF 大肠杆菌 BW25113 菌株。使用一半的回收物来接种 30ml M9 基本培养基，其中添加了 0.4%葡萄糖，100 μ M IPTG，20 μ M 柠檬酸铁，2mg/ml α -L-丙氨酸，痕量元素，和 25 μ g/mL 卡那霉素。将培养基放置于 50ml 管中，并不时地混合。8 天后可见良好的生长，并在 M9 基本培养基上划线。

如上所述在液体生长检测中检测菌落，其表现比 Δ panD/ Δ panF 菌株中的 Fnaam 突变体约好 4 倍。对新突变体的质粒 DNA 进行测序，并将其用来重新转化 Δ panD 菌株。对重新转化子的液体生长检测确认了所述生长优势是由所述质粒所赋予的。在这种具核梭杆菌丙氨酸 2,3-氨基变位酶(Fnaam2, SEQ ID NO: 21)中还有两处额外的氨基酸取代(K37E 和 D180N)。所述 DNA 序列如 SEQ ID NO: 20 所示。改进的突变体以 0.5mg/ml 的 α -L-丙氨酸在液体生长检测中比 Fnaam 突变体的表现好约 3 倍。

实施例 4

斯氏梭菌(*Clostridium sticklandii*)赖氨酸 2,3-氨基变位酶的克隆和密码子改进

本实施例描述了用于克隆斯氏梭菌(*Clostridium sticklandii*)赖氨酸 2,3-氨基变位酶的方法，以及对在大肠杆菌中表达的密码子进行优化的方法。虽然对涉及斯氏梭菌赖氨酸 2,3 和赖氨酸 5,6-氨基变位酶的赖氨酸降解途径已有描述，但该基因序列依然未知。

在 ATCC 培养基 1053 中对斯氏梭菌(ATCC 12662)进行扩增，并根据对具核梭杆菌的描述提取基因组 DNA。根据公知赖氨酸 2,3-氨基变位酶基因的保守蛋白区设计了三条简并正向 PCR 引物(F1:

5'-YTWAGAATGGCWATWACWCC-3' (SEQ ID NO: 22) ; F2: 5'-AGAAARCARGCWATWCCWAC-3' (SEQ ID NO: 23); 和 F3: 5'-GGWYTWACWCAYAGATAYCC-3' (SEQ ID NO: 24)); 和三条简并反向 PCR 引物(R1 : 5'-TAWGTWGTWATWACWCCYTC-3' (SEQ ID NO: 25); R2: 5'-TCWACWACRAAWGTWGGWAC-3' (SEQ ID NO: 26); 和 R3: 5'-CCWCCWCCWGGWGCRTCWAC-3' (SEQ ID NO: 27))。使用了斯氏梭菌密码子偏好表从蛋白序列设计引物。

所述引物被用于全部逻辑组合的 PCR 中, 这些 PCR 使用了 Taq 聚合酶和 1ng 斯氏梭菌基因组 DNA/ μ L 反应混合物。使用降落 (touchdown) PCR 程序进行 PCR, 该程序包括 4 个循环 56°C 退火, 4 个循环 54°C, 4 个循环 52°C, 和 20 个循环 50°C。每个循环都使用在 94°C 的起始 45 秒变性步骤和在 72°C 的两分钟延伸, 以及在 72°C 5 分钟的最终延伸。由于在所述引物 3'末端的简并性, 在反应中使用的 PCR 引物的量比常规的 PCR 引物量增加了 3 倍。此外, 可以使用含有每一种引物的单独 PCR 反应来鉴定从单种简并引物中得到的 PCR 产物。

将 25 μ L 的各种 PCR 产物进行 1.5%琼脂糖凝胶分离。引物组合 F1-R1(950bp),F1-R3 (900 bp)和 F3-R3 (730 bp)产生了较强的条带, 其具有根据其他物种基因所得的大约预期大小。引物组合 F3-R1 (800 bp), F1-R2 (900 bp)和 F3-R2 (750 bp)产生了较弱的条带。未观察到 F2-R1, F2-R2, F2-R3 引物组合或其他单独引物对照获得的条带。使用 Qiagen Gel Extraction 方法(Qiagen Inc., Valencia, CA)提取和纯化了 F1-R1 和 F1-R3 片段。将 4 μ l 纯化条带连接入 TOPO 4.0 载体, 并使用 TOPO 克隆方法(Invitrogen, Carlsbad, CA)通过热击法转化入 TOP10 大肠杆菌细胞。将转化物涂布于含有 100 μ g/ml 氨苄青霉素和 50 μ g/ml X-gal(5-溴-4-氯-3-吡啶- β -D-半乳糖苷)的 LB 平板上。将单独的、白色的菌落放置于小量缓冲液中, 并将等分试样用于启动 5ml 用于提取质粒 DNA 的培养液。将剩余物在 95°C 加热 10 分钟, 并使用 2 μ L 进行 PCR 反应从而确认所希望插入的存在。使用 QiaPrep Spin Miniprep 试剂盒, 从具

有所希望插入的多个克隆中获得质粒 DNA。用 M13R 和 M13F 引物对 F1-R1 克隆的插入进行测序。序列分析说明这一片段与已知的赖氨酸 2,3-氨基变位酶同源。

基因组步行来获得完整的编码序列

使用从上述 F1-R1 片段获得的序列，避免了对应于简并引物的区域，设计出了上游和下游方向的用于基因组步行的引物。引物序列为：用于 GSP1F 的 5'-CCTTTCAGTTGGAATTGAGCACTTTAGAAC-3' (SEQ ID NO: 28) ， 用于 GSP2F 的 5'-GATACTGCGTTCCTACATTTGTTGTGGATG-3' (SEQ ID NO: 29)，用于 GSP1R 的 5'-CGCTGCTCTATG TAGCTCTAAAGAAAGAG-S' (SEQ ID NO: 30) ， 和 用于 GSP2R 的 5'-CAGCTTGCTTTCTTACAGGGTCATTTGG-3' (SEQ ID NO: 31)，其中 GSP1F 和 GSP2F 是朝向下流的引物，GSP1R 和 GSP2R 是朝向上游的引物，而 GSP2F 和 GSP2R 则分别是位于 GSP1F 和 GSP1R 内部的巢式引物。除了限制酶使用 Hinc II, Ssp I, EcoR V, Pvu II 和 Dra I, 可根据 Clontech's Universal Genome Walking 试剂盒(Clontech Laboratories, Inc., Palo Alto, CA)的操作手册进行染色体步行。

在 Perkin Elmer 9700 热循环仪上进行第一轮 PCR，条件为 94°C 起始变性 15 秒；7 个循环的 94°C 5 秒和 70°C 3 分钟；32 个循环的 94°C 5 秒和 65°C 3 分钟，以及 65°C 4 分钟的最终延伸。第二轮 PCR 的条件为 94°C 起始变性 15 秒；5 个循环的 94°C 5 秒和 70°C 3 分钟；26 个循环的 94°C 5 秒和 65°C 3 分钟，以及 65°C 4 分钟的最终延伸。

将 20 μ L 各轮次的 PCR 产物在 1.5%琼脂糖凝胶上进行电泳分离。采用 Ssp I, EcoR V, Pvu II 和 Dra I 文库的正向引物和采用 Ssp I, Hinc II, Pvu II 和 Dra I 文库的反向引物获得了扩增产物。如上所述，对使用 Dra I 正向反应(1.2kb)和 Hinc II 反向反应(1.3kb)的第二轮产物进行凝胶纯化，克隆和插入大小的筛选。用 M13R 和 M13F 引物对具有所希

望插入大小的多个克隆的质粒 DNA 进行测序。将该序列与原始基因片段的序列进行比对，从而确定所述斯氏梭菌赖氨酸 2,3-氨基变位酶同源物的全序列。然后使用 PCR 扩增从基因组 DNA 中将全基因重新克隆入载体中。诸如斯氏梭菌 kam 基因(SEQ ID NO: 32, 其对应的蛋白质序列如 SEQ ID NO: 33 所示)的新基因可用作突变的起始模板，从而获得丙氨酸 2,3-氨基变位酶活性。

对如 SEQ ID NO: 32 所示的斯氏梭菌 kam 基因进行部分密码子优化以适于在大肠杆菌中的表达。如实施例 2 所述，通过使用含有对精氨酸更优选密码子的引物，在扩增反应中通过整合替换了稀有的精氨酸密码子。两轮引物整合之后，将所述 PCR 产物克隆入 pPRO-Nde 载体。测序说明，与含有 22 个稀有精氨酸密码子的野生型基因相比，最好的克隆含有 8 个稀有精氨酸密码子。部分优化的斯氏梭菌 kam 基因(Csco)如 SEQ ID NO: 34 所示(未改变的蛋白质序列如 SEQ ID NO: 35 所示)。

实施例 5

斯氏梭菌(*Clostridium sticklandii*)赖氨酸 2,3-氨基变位酶(kam 基因)的体外突变

本实施例描述了用于对实施例 4 中生成的部分优化的斯氏梭菌 kam 基因(SEQ ID NO: 34)进行突变，从而鉴定具有丙氨酸 2,3-氨基变位酶活性突变体的方法。

通过定向的突变，将之前在枯草芽孢杆菌，牙龈卟啉单胞菌或具核梭杆菌丙氨酸 2,3-氨基变位酶(例如，参见 SEQ ID NOS: 1-6 和 18-21 和 WO 03/062173)中获得的四种突变转移到斯氏梭菌 Csco 基因中(SEQ ID NO: 35)。这些突变包括 E28K, L93M, M126V 和 D329H 取代。可使用 QuikChange[®] Multi Site-Directed Mutagenesis 试剂盒(Stratagene)制备这些突变。用于制备这些突变的引物是：CsE28K: 5'Phos/GAAATG GCAAGTAAGAAATCGTATAAAGACTGTTGAAGAACTTAA (SEQ ID

NO: 36); CsL93M: 5'Phos/TCGCGCAGCGTCTGATATGGAAGACCCA
CTTCATG (SEQ ID NO: 37); CsM126V: 5'Phos/GACTGATCAATGTTC
AGTATACTGCCGCCACTGTACTCGT (SEQ ID NO: 38); 和 CsD329H:
5'Phos/GTTCCTACATTTGTTGTGCATGCACCTGGTGGTG (SEQ ID
NO: 39)。

对在 Cscodm 基因(SEQ ID NO: 41)中具有全部四种定向突变的四种克隆进行了如实施例 2 所述的液体生长检测, 该检测证明了低水平的丙氨酸 2,3-氨基变位酶活性。在对生长检测培养物进行亚培养之后, 所述 Cscodm 菌株表现出比 Cscodm 母体菌株多约三倍的生长, 说明了较低水平的丙氨酸 2,3-氨基变位酶活性。

为了在体外向 Cscodm 基因(SEQ ID NO: 40)中引入额外的突变, 如实施例 2 所述使用了倾向于错误的 PCR 方法, 除了该 PCR 程序由 94°C 起始变性 2 分钟; 30 个循环的 94°C 30 秒, 55°C 45 秒, 和 72°C 2.25 分钟; 以及 72°C 7 分钟的最终延伸组成。

PCR 完成之后, 用 NotI 和 NdeI 消化 PCR 产物。将每个处理的等量 DNA 连接于载体 pPRO-Nde, 并转化大肠杆菌 Electromax™ DH10B™ 细胞。将多次转化涂布于 LKB25 平板获得约 54,000 个克隆。从平板上刮下克隆, 并使用 Qiagen MiniSpin Plasmid 方法制备质粒 DNA。在对选定宿主进行转化之前, 使用乙酸胺和乙醇沉淀质粒 DNA 从而增加其浓度。

实施例 6

具有丙氨酸 2,3-氨基变位酶活性克隆的鉴定

本实施例描述了用于鉴定具有丙氨酸 2,3-氨基变位酶活性的突变赖氨酸 2,3-氨基变位酶克隆的方法。

将在实施例 5 中产生的突变 Cscodm 文库转化大肠杆菌 BW25113

的 $\Delta panD/\Delta gabT/\Delta yeiA$ 菌株的电击感受态细胞。可制备 *gabT* 和 *yeiA* 突变, 从而消除由细胞途径形成或处理 β -丙氨酸。在 SOC 培养基中一小时之后回收转化子, 离心, 用 0.85%NaCl 洗涤, 并重悬于 1ml NaCl 中。将一半回收物用于接种 25ml M9 基本培养基, 其中添加了 0.4%葡萄糖, 100 μ M IPTG, 20 μ M 柠檬酸铁, 2mg/ml α -L-丙氨酸, 微量元素, 和 25 μ g/mL 卡那霉素(Sigma, St. Louis, MO)。

三天后, 将生长的培养物在 LBK 培养基上划线。将单个克隆接种于 LBK 和 M9 基本培养基。使用 2ml 管中的 1.6ml M9 基本培养基(如上所述), 在液体生长检测中对在 M9 培养基中厌氧生长旺盛的克隆进行检测。使用重悬的克隆来接种含有和不含有泛酸盐(25 μ M)的培养基。当泛酸盐对照达到 OD₆₀₀ 为约 0.6 时, 读取培养物的 OD。如表 1 所示, 有三个克隆在没有泛酸盐中的生长比率高于在含有泛酸盐中的生长比率。对这三个克隆的质粒 DNA 进行测序, 并将其用于再次转化 $\Delta panD$ 菌株。对再次转化的菌株重复生长检测, 从而确保所述生长优势是由所述质粒赋予的, 而不是宿主效应。

表 1: 存在斯氏梭菌丙氨酸 2,3-氨基变位酶基因条件下的生长

克隆	比率
Cscodm(SEQ ID NOS: 40 和 41)	0.12
Cscdm mut8(SEQ ID NOS: 42 和 43)	0.45
Cscdm mut12(SEQ ID NOS: 44 和 45)	0.68
Cscdm mut15(SEQ ID NOS: 46 和 47)	0.48

编码丙氨酸 2,3-氨基变位酶的三个突变体斯氏梭菌 *kam* 基因序列如 SEQ ID NOS: 42, 44 和 45 所示, 其对应的氨基酸序列如 SEQ ID NOS: 43, 45 和 47 所示。除了四种定向的突变(参见 SEQ ID NO: 41), 与斯氏梭菌 *kam* 基因序列(FIG. 5)相比, 还在突变序列中观察到了 1-4 种氨基酸改变。这些突变包括 S9T, V30A, C50R, L51H, E69G, Q123L, Q139R 和 K185R。然而, 所有这些突变对获得丙氨酸 2,3-氨基变位酶

活性而言都不是必须的。

实施例 7

牙龈卟啉单胞菌(*Porphyromonas gingivalis*)丙氨酸 2,3-氨基变位酶的体外突变

可使用实施例 2 所述的方法,对已有描述的(WO 03/062173) 牙龈卟啉单胞菌丙氨酸 2,3-氨基变位酶 (SEQ ID NO: 5, 其对应蛋白质序列如 SEQ ID NO: 6 所示)进行突变 PCR,除了该 PCR 程序由 94°C 起始变性 2 分钟; 30 个循环的 94°C 30 秒, 55°C 1 分钟, 和 72°C 2.25 分钟; 以及 72°C 7 分钟的最终延伸组成。

PCR 完成之后,用 NotI 和 NdeI 消化 PCR 产物。将每个处理的等量 DNA 连接于载体 pPRO-Nde, 并转化大肠杆菌 Electromax™ DH10B™ 细胞。从单克隆提取质粒 DNA 并测序从而获得对突变率的估计。平均突变率为 0.3%。将多次转化涂布于 LKB25 平板获得约 80,000 个克隆。从平板上刮下克隆, 并使用 Qiagen MiniSpin Plasmid 方法制备质粒 DNA。在对选定宿主进行转化之前, 使用乙酸胺和乙醇沉淀质粒 DNA 从而增加其浓度。

实施例 8

对具有改进丙氨酸 2,3 氨基变位酶克隆的选择和鉴定

将在实施例 7 中产生的突变 Pgaam 质粒文库转化大肠杆菌 BW25113 的 Δ panD 菌株电击感受态细胞。一小时之后在 SOC 培养基中回收转化子, 离心, 用 0.85%NaCl 洗涤, 并重悬于 1ml NaCl 中。用 10 μ L 接种 1.8ml 玻璃管中的 1.33ml M9 基本培养基, 其中添加了 0.4% 葡萄糖, 100 μ M IPTG, 50 μ M Fe(NH₄)₂(SO₄)₂, 2mg/ml α -L-丙氨酸, 微量元素, 和 25 μ g/mL 卡那霉素(Sigma, St. Louis, MO)。三天之后, 将 100 μ L 生长培养物用于接种新鲜管中的相同培养基。过夜培养之后, 将培养物在 M9 基本培养基上划线, 并放置于厌氧舱中。将生长出来的 6 个克隆接种于 LBK, 并使用 2ml 管中的 1.4ml M9 基本培养基(如上所

述)进行液体生长检测。使用重悬的克隆来接种含有和不含有泛酸盐 (25 μ M)的培养基。当泛酸盐对照达到 OD₆₀₀ 为约 0.7 时, 读取培养物的 OD。根据在没有泛酸盐中的生长比率高于在含有泛酸盐中的生长比率来鉴定克隆。

如表 2 所示, Pgaam2 克隆在没有泛酸盐中的生长比率高于在含有泛酸盐中的生长比率。对这个克隆的质粒 DNA 进行测序, 并将其用于再次转化 Δ panD 大肠杆菌菌株。对再次转化的菌株重复生长检测, 从而确保所述生长优势是由所述质粒赋予的, 而不是宿主效应(表 3, “再次转化的”)。

表 2: 存在牙龈卟啉单胞菌野生型和突变丙氨酸 2,3-氨基变位酶条件下的生长

克隆	比率
Pgkam(SEQ ID NOS: 52)	0.10
Pgaam(SEQ ID NOS: 5 和 6)	0.33
Pgaam2(SEQ ID NOS: 48 和 49)	0.73
Pgaam(再次转化)	0.46
Pgaam2(再次转化)	0.84

在改进的牙龈卟啉单胞菌丙氨酸氨基变位酶(Pgaam2, SEQ ID NO: 49)中存在有 E30K 和 I192V 的氨基酸取代。其 DNA 序列如 SEQ ID NO: 48 所示。

为了增加 Pgaam2 的丙氨酸 2,3-氨基变位酶活性(SEQ ID NO: 49), 可对若干荷负电的氨基酸进行突变。因为荷负电的氨基酸可与荷正电的天然底物, 赖氨酸相互作用, 一些或全部负电荷的清除可以增强以丙氨酸作为底物的活性, 因为其较小且不带电。可使用 Stratagene QuikChange® Multi Site-Directed Mutagenesis 试剂盒进行定向突变, 对其进行以下修饰: 使用 50 μ L 反应, 并在 Dpn I 处理之后使用 NH₄OAc

和乙醇沉淀。在用 10 μ l 水进行重悬之后，用 2.5 μ l 转化 Δ panD 大肠杆菌菌株。将三分之一的转化回收物用来接种 14ml 管中的 8ml M9 基本培养基。该培养基中添加了 0.4%葡萄糖，100 μ M IPTG，100mM MOPS pH 7.0，50 μ M 柠檬酸铁，1mg/ml α -L-丙氨酸，痕量元素，和 25 μ g/mL 卡那霉素。生长 5 天后，在 M9 基本培养基上涂布以获得单菌落。将单菌落接种于 LBK 培养基，并使用 2ml 管中的 1.4ml M9 基本培养基(如上所述，1mg/ml α -L-丙氨酸)进行液体生长检测。使用重悬的克隆来接种含有和不含有泛酸盐(25 μ M)的培养基。当泛酸盐对照达到 OD₆₀₀ 为约 0.7 时，读取培养物的 OD。将阳性菌落(Pgaam2 L26I)的质粒 DNA 再次转化 Δ panD 菌株，并在液体生长检测中重新检测再次转化的菌株(表 3)。

表 3: 存在牙龈卟啉单胞菌丙氨酸 2,3-氨基变位酶基因条件下的生长

克隆	比率
Pgaam2(SEQ ID NOS: 48 和 49)	0.21
Pgaam2 L26I(SEQ ID NOS: 50 和 51)	0.53

序列结果说明，所述突变位于突变引物之一 3'末端的紧邻下游核苷酸上。因此，其可能是由于复制错误或引物合成错误导致的。改进的牙龈卟啉单胞菌丙氨酸 2,3-氨基变位酶(Pgaam2 L26I)的 DNA 序列如 SEQ ID NOS: 50 所示，其对应的氨基酸序列如 SEQ ID NOS: 51 所示。

实施例 9

丙氨酸 2,3-氨基变位酶活性的体外分析

本实施例描述了用于确定 SEQ ID NO: 49 的丙氨酸 2,3-氨基变位酶活性的方法。本领域所属技术人员应该意识到，类似的方法还可用来测定针对在此公开的任意丙氨酸 2,3-氨基变位酶蛋白，例如 SEQ ID NO: 19, 21, 43, 45, 47 或 51，以及保留了丙氨酸 2,3-氨基变位酶活性的序列变体，片段，或融合体的丙氨酸 2,3-氨基变位酶活性。

可使用制造商描述的方法，使用具有 C-末端 strep-tag 的 pASK-IBA3 载体(IBA, St. Louis, MO)，在大肠杆菌 BL21(DE3)中表达了具有丙氨酸 2,3-氨基变位酶活性的蛋白。将携带有 strep-tag 氨基变位酶克隆的细胞生长于 Terrific Broth 中，其中添加了 100 μ g/ml 氨苄青霉素和 40 μ g/ml 柠檬酸钠铁，并在 OD₆₀₀=1-2 时用 0.2 μ g/ml 无水四环素诱导表达。将孵育温度降至 25°C，并在过夜生长之后离心收集细胞。

以 2ml 缓冲液/g 细胞沉淀的比率将细胞沉淀重悬于缓冲液(50mM HEPPS pH8, 25 μ M 磷酸吡哆醛, 0.1mM α -L-丙氨酸)，并将所述细胞悬液进行超声波处理 (9-15W, 2 x 1.5 分钟)。将匀质的细胞悬液离心，并将可溶性的无细胞提取物(CFE)小心地倾倒入并保存。

根据已知的 Chen 等人对于赖氨酸 2,3-氨基变位酶进行的描述 (Biochem. J. 348:539-549, 2000)，进行了丙氨酸 2,3-氨基变位酶活性分析，除了该分析含有替代赖氨酸的 α -L-丙氨酸。通过用 0.4ml 还原孵育缓冲液 (4.6mL H₂O, 0.25mL 1M HEPPS pH8, 0.05mL 100 mM Fe(NH₄)₂(SO₄)₂, 0.05mL 50mM 磷酸吡哆醛, 0.05mL 100mM 二硫苏糖醇；所有溶液都是在厌氧条件下制备的)处理对 0.1ml CFE 进行预还原，预孵育可在 37°C 进行 4 小时。向预孵育的酶中，顺序加入厌氧制备的以下成分：0.21mL H₂O, 0.05mL 1M HEPPS pH8, 0.02mL 5-10mM S-腺苷甲硫氨酸, 0.02mL 100mM 亚硫酸氢钠, 0.2mL 500mM α -L-丙氨酸。通过与等体积的 90%甲酸混合，在多个时间点上猝灭所述反应，使用互作色谱 AA511 柱的 HPLC 监测 β -丙氨酸的形成，并使用 O-邻苯二醛进行后续柱衍生。使用 Pickering Laboratories 缓冲液 Na 328 和 Na 740 来展开层析，以 0.5ml/分钟流速，柱温度 60°C，并将反应的色谱与那些标准量的 β -丙氨酸的色谱进行比较来进行定量。

使用这样的分析，Pgaam2 丙氨酸 2,3-氨基变位酶 (SEQ ID NO: 49) 的比活性被测定为 0.03 \pm 0.01 单位/mg 蛋白(1 单位=每分钟产生 1 μ mol β -丙氨酸)。

实施例 10

大肠杆菌 Δ panD::CAT 菌株的构建

为了对编码能实施丙氨酸 2,3-氨基变位酶反应的多肽的基因进行鉴定, 必须对所需要的活性进行有效的筛选和选择。因此, 可通过识别使用 β -丙氨酸合成依次以辅酶 A(CoA)和酰基载体蛋白(ACP)为成分的泛酸的大肠杆菌而开发了筛选方法。CoA 和 ACP 是活生物体中重要的酰基基团, 都生长至关重要。在大肠杆菌中, 形成 β -丙氨酸的主要途径是通过由 panD 基因编码的天冬氨酸盐脱羧酶(E.C. 4.1.1.11)催化反应中的天冬氨酸盐开始的(图 3)。PanD 的功能性缺失突变导致了 β -丙氨酸的营养缺乏和生长抑制, 其可通过外源性的加入泛酸盐或 β -丙氨酸, 或通过从其他来源产生 β -丙氨酸得以缓解。

将两株 β -丙氨酸合成缺陷的大肠杆菌用于筛选。菌株 DV1((#6865, E. coli Genetic Stock Center, New Haven CT; Vallari 和 Rock, J. Bacteriol. 164: 136-42, 1985)是通过化学突变制备的大肠杆菌突变株, 其具有使得两个基因都丧失功能的 panF 和 panD 双重宿主(染色体)突变。PanF 基因编码从培养基中摄取泛酸盐, 因此 panF 和 panD 的组合对用于生长的 β -丙氨酸提供了更为苛刻的需求。因此, 虽然 DV1 菌株是已知的, 但其用于对具有丙氨酸 2,3-氨基变位酶活性的细胞进行筛选是之前所不知道的。

其他的选择菌株, BW25113 Δ panD::CAT, 含有 panD 座的缺失, 从而阻止了能够在无外源 β -丙氨酸条件下生长时 panD 突变的回复突变。这个菌株, 具有由插入 panD 座的 CAT 基因所赋予的氯霉素抗性标记的插入, 是使用 E. Coli Genetic Stock Center 的大肠杆菌菌株 BW25113/pKD46 和 BW 25141/pKD3, 使用 Datsenko 和 Wanner 的基因失活方法(Proc. Natl. Acad. Sci USA 97: 6640-5, 2000)构建的。

可使用引物 TATCAATTCGTTACAGGCGATACATGGCACGCTT

CGGCGCGTGTAGGCTGGAGCTGCTTC (SEQ ID NO: 53)和 GATGTC GCGGCTGGTGAGTAACCAGCCGCAGGGATAACAACATATGAATA TCCTCCTTAG (SEQ ID NO: 54)对 pKD3 的 CAT 基因进行扩增。PCR 反应包括在 300 μ l 终体积中的 30 μ l 10X 浓缩 PCR 缓冲液(Roche Molecular Biochemicals), 质粒 pKD3, 0.2 mM 各种 dNTP, 0.2 μ M 各种引物和 15 单位 Taq 聚合酶(Roche Molecular Biochemicals)。PCR 反应在 95 $^{\circ}$ C 孵育 30 秒, 之后是 30 个循环的 95 $^{\circ}$ C 30 秒, 45 $^{\circ}$ C 30 秒, 72 $^{\circ}$ C 1 分钟, 然后 72 $^{\circ}$ C 10 分钟。用乙醇沉淀 PCR 产物, 用 DpnI 消化, 用 QIAquick PCR Purification 试剂盒(Qiagen)进行纯化, 并转化入表达重组功能的 BW25113/pKD46 中。将转化自涂布于含有 25 μ g/ml 氯霉素和 5 μ M β -丙氨酸的 LB 平板上。

在 43 $^{\circ}$ C, 添加有 5 μ M β -丙氨酸的非选择性 LB 培养基中单克隆纯化氯霉素抗性的转化物, 并检测单克隆的氯霉素抗性保持, 氨基青霉素抗性的丧失(显示 pKD46 的熟化 (curing)), 以及在 M9-葡萄糖基本培养基中生长 β -丙氨酸的需求。CAT 基因向 panD 座正确插入的确认是通过使用侧翼于插入座(TTACCGAGCAGCGTTCAGAG, SEQ ID NO: 55; 和 CACCTGGCGGTGACAACCAT, SEQ ID NO: 56)的引物, 对所得的 Δ panD::CAT 菌株进行菌落 PCR 实现的。当野生型 panD 座预期产生 713 个碱基对的 PCR 产物时, Δ panD::CAT 构建体产生了 1215-碱基对的产物。如前所述(Datsenko 和 Wanner, Proc. Natl. Acad. Sci. USA 97: 6640-5, 2000), 构建了 Δ panD::CAT 菌株的衍生物, 其中通过质粒 pCP20 编码的 FLP 重组酶活性去除了插入的 CAT 基因。这样的菌株称为 Δ panD。

存在于大肠杆菌中生成 β -丙氨酸的第二条途径是根据尿嘧啶代谢的还原途径(West, Can. J. Microbiol. 44: 1106-9, 1998, 图 2)。在这条途径中, 尿嘧啶被二氢嘧啶脱氢酶(E.C. 1.3.1.2)转化成了二氢尿嘧啶。通过二氢嘧啶酶(E.C. 3.5.2.2), 二氢尿嘧啶随后被转化成 N-氨基甲酰- β -丙氨酸, 其依次被 N-氨基甲酰- β -丙氨酸-酰胺水解酶(E.C. 3.5.1.6)水解

成 β -丙氨酸, CO_2 , 和 NH_3 。为了阻止通过这一途径形成 β -丙氨酸, 通过上述 Datsenko 和 Wanner 的方法, 可将编码二氢嘧啶脱氢酶的基因, *yeiA* (GenBank Accession No. AAC75208), 插入性缺失。使用引物 GCGGCGTGAAGTTTCCCAACCCGTTCTGCCTCTCTTCTTCGTGTAG GCTGGAGCTGCTTC (SEQ ID NO: 57)和 TTACAACGTTACCGGGTG TTCTTTCTCGCCTTTCTTAAACCATATGAATATCCTCCTTAG (SEQ ID NO: 58)对 pKD3 的 CAT 基因进行扩增。

如上所述分离氯霉素抗性的插入突变体, 并将所述抗性标记转到入 ΔpanD 菌株从而生成双突变体 $\Delta\text{panD}/\Delta\text{yeiA}::\text{CAT}$ 。如以上实施例所述, 生成大肠杆菌 BW25115 $\Delta\text{panD}::\text{CAT}$, ΔpanD 或 $\Delta\text{panD}/\Delta\text{yeiA}::\text{CAT}$ 的电击感受态细胞, 并将其用作转化突变体赖氨酸 2,3-氨基变位酶 DNA 文库的宿主。

实施例 11

不使用突变的赖氨酸 2,3-氨基变位酶来选择丙氨酸 2,3-氨基变位酶活性

鉴定具有丙氨酸 2,3-氨基变位酶活性细胞的其他方法是不使用突变的赖氨酸 2,3-氨基变位酶文库对其进行转化, 而是在如上所述的培养基上涂布细胞, 例如如实施例 10 所述的 DV1 或 $\Delta\text{panD}::\text{CAT}$ 细胞。如上所述对这些细胞进行选择, 并如实施例 3, 5, 6, 8 和 9 所述验证丙氨酸 2,3-氨基变位酶活性的存在。

可在涂布之前对细胞进行突变, 例如通过将细胞暴露于 UV 辐射或化学物质(例如 EMS)。这可以允许分离出具有一个或多个其他基因突变的突变体, 其导致了该细胞具有丙氨酸 2,3-氨基变位酶活性。

或者, 可在涂布之前不对细胞进行改变(例如, 不转化, 不突变)。这一方法允许分离出具有丙氨酸 2,3-氨基变位酶活性天然存在的菌株。

实施例 12

从 β -丙氨酸制备泛酸盐

可使用具有 α -酮泛解酸盐羟甲基转移酶(E. C. 2.1.2.11), α -酮泛解酸盐还原酶(E.C. 1.1.1.169), 和泛酸盐合酶(E.C. 6.3.2.1)活性的多肽从 β -丙氨酸制备泛酸盐(图 3)。

使用本发明所述的克隆方法, 可对 α -酮泛解酸盐羟甲基转移酶(E. C. 2.1.2.11), α -酮泛解酸盐还原酶(E.C. 1.1.1.169), 和泛酸盐合酶(E.C. 6.3.2.1)进行分离, 测序, 表达和检测。本领域所属技术人员应该理解, 类似的方法可用于从任意生物体中获得任意这些多肽的序列。

实施例 13

重组表达

根据公众可获得的酶 cDNA 和氨基酸序列, 以及在此所公开的丙氨酸 2,3-氨基变位酶, 及其变体, 片段和融合体, 通过标准的实验室技术可对任意蛋白, 例如丙氨酸 2,3-氨基变位酶进行表达和纯化。本领域所属技术人员应该理解, 酶及其片段可在任意感兴趣的细胞或生物体中重组地制备, 并在使用之前, 例如在制备 3-HP, 泛酸盐及其衍生物之前进行纯化。

制备重组蛋白的方法时本领域所公知的。因此, 本发明的范围包括了任意蛋白或其片段, 例如丙氨酸 2,3-氨基变位酶的重组表达。例如, 参见 Johnson 等人的美国专利第 5,342,764 号; Pausch 等人的美国专利第 5,846,819 号; Fleer 等人的美国专利第 5,876,969 号和 Sambrook 等人 (Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, New York, 1989, Ch. 17)。

简言之, 可将编码蛋白的部分、全长、或变体 cDNA 序列连接入表达载体, 例如细菌表达载体。可通过在所述 cDNA 序列的上游放置

启动子来制备蛋白。启动子的示例包括，但不限于，lac, trp, tac, trc, 主要操纵子和 λ 噬菌体的启动子区，fd衣壳蛋白的控制区，SV40的早期和晚期启动子，衍生自多瘤病毒，腺病毒，逆转录病毒，杆状病毒和猿猴病毒的启动子，3-磷酸甘油酸激酶的启动子，酵母酸性磷酸酶的启动子的启动子，酵母 α -结合因子的启动子及其组合。

适于制备完整天然蛋白的载体包括 pKC30(Shimatake and Rosenberg, 1981, Nature 292:128), pKK177-3 (Amann and Brosius, 1985, Gene 40:183)和 pET-3 (Studier and Moffatt, 1986, J. Mol. Biol 189: 113)。DNA 序列可转移至其他克隆载体，例如其他质粒，噬菌体，粘粒，动物病毒和酵母人工染色体(YAC) (Burke 等人, 1987, Science 236:806-12)中。这些载体可被导入多种宿主，包括体细胞，和简单或复杂生物体，例如细菌，真菌(Timberlake and Marshall, 1989, Science 244:1313-7), 无脊椎动物，植物(Gasser and Fraley, 1989, Science 244:1293), 和哺乳动物(Pursel 等人, 1989, Science 244:1281-8), 其通过导入异源 cDNA 而成为转基因的。

对哺乳动物细胞中的表达而言，可将 cDNA 序列连接于异源启动子，例如猿猴病毒 SV40, pSV2 载体的启动子(Mulligan and Berg, 1981, Proc. Natl. Acad. Sci. USA 78:2072-6), 并导入细胞，例如猴 COS-1 细胞(Gluzman, 1981, Cell 23:175-82), 从而获得瞬时或长期表达。可通过生化选择，例如新霉素(Southern and Berg, 1982, J. Mol. Appl. Genet. 1:327-41)和 mycophoenolic 酸(Mulligan and Berg, 1981, Proc. Natl. Acad. Sci. USA 78:2072-6), 在哺乳动物细胞中获得嵌合基因构建体的稳定整合。

将 DNA 转移入真核细胞，例如人类或其他哺乳动物细胞是常规的技术。载体可作为纯 DNA, 通过，例如，磷酸钙沉淀(Graham and vanDer Eb, 1973, Virology 52:466), 磷酸锶(Brash 等人, 1987, Mol Cell Biol. 7:2013), 电穿孔(Neumann 等人, 1982, EMBO J. 1:841), 脂质体转染

(Feigner 等人, 1987, Proc. Natl. Acad. Sci USA 84:7413), DEAE 葡聚糖(McCuthan 等人, 1968, 7. Natl. Cancer Inst. 41 :351), 显微注射(Mueller 等人, 1978, Cell 15:579), 原生质体融合(Schafner, 1980, Proc. Natl. Acad. Sci. USA 77:2163-7)或基因枪(pellet guns) (Klein 等人, 1987, Nature 327:70)导入受体细胞中(转染)。或者, 可通过使用病毒载体, 例如逆转录病毒(Bernstein 等人, 1985, Gen. Engrg. 7:235), 例如腺病毒(Ahmad 等人, 1986, J. Virol. 57:267)或疱疹病毒(Spaete 等人, 1982, Cell 30:295)进行转染将 cDNA 导入。

实施例 14

多肽合成与纯化

可通过本领域公知的任意数量的手人工或自动方法化学地合成在此公开的酶, 例如丙氨酸 2,3-氨基变位酶及其变体, 融合体和片段。例如, 可使用 Applied Biosystems Model 431A 多肽合成仪, 使用 9-芴甲氧羰基(Fmoc)氨基末端保护, 与二环己基碳二亚胺/羟基苯并三唑或 2-(1H-苯并-三唑-1-基)-1,1,3,3-六氟磷酸四甲基脒鎓/羟基苯并三唑(HBTU/HOBT)结合, 并使用对-羟基甲基苯氧甲基聚苯乙烯(HMP)或适于羧基末端酸的 Sasrin 树脂或适于羧基末端酸的 Rink 酰胺树脂实施 0.25 毫摩尔(mmole)规模的固相多肽合成(SPPS)。

可通过三苯甲基化作用和溶于三氟乙酸中的三苯甲醇, 然后通过 Fmoc 衍生, 从适当的前体氨基酸制备 Fmoc-衍生的氨基酸, 如 Atherton 等人所述(Solid Phase Peptide Synthesis, IRL Press: Oxford, 1989)。使用 1%TFA 的二氯甲烷溶液可对 Sasrin 树脂结合的肽进行剪切, 从而生成被保护的肽。在适当的条件下, 使用二苯基磷酰基叠氮化物, 通过在新生肽中氨基末端游离胺与羧基末端游离酸的反应, 可将被保护肽前体在氨基末端和羧基末端之间环化, 其中所述氨基酸侧链是被保护的。

HMP 或 Rink 酰胺树脂-结合的产物通常被剪切, 且可以使用含有

三氟乙酸(TFA)的溶液,也可任意含有水、苯硫基甲烷和乙二硫醇,以100:5:5:2.5的比率于室温0.5-3小时,对含有受保护侧链的环化肽脱保护。

可使用制备性的高压液相色谱(HPLC),例如使用Water Delta-PAK C18柱和用乙腈缓和的含0.1%TFA的水梯度洗脱,来纯化粗制的多肽。柱洗提后,从洗脱级分中蒸发乙腈,然后冻干该洗脱级分。如此生产和纯化的各种产物的同一性可通过快速原子轰击质谱法(FABMS)或电喷射质谱法(ESMS)进行验证。

鉴于我们公开的原理可应用于许多可能的实施方案,应该理解举例说明的实施方案只是公开的特定实施例而不应该作为本发明范围的限制。相应的,所公开内容的范围与下列权利要求是一致的。因此我们要求保护所有在这些权利要求的范围和精神内的发明。

<110> 卡吉尔公司 (Cargill Incorporated)

<120> 丙氨酸 2,3 氨基变位酶 (Alanine 2,3 aminomutases)

<130> SCT070223-66

<160> 59

<170> PatentIn version 3.2

<210> 1
 <211> 1416
 <212> DNA
 <213> Bacillus subtilis

<220>
 <221> CDS
 <222> (1)..(1416)

<400> 1

atg aaa aac aaa tgg tat aaa ccg aaa cgg cat tgg aag gag atc gag	48
Met Lys Asn Lys Trp Tyr Lys Pro Lys Arg His Trp Lys Glu Ile Glu	
1 5 10 15	
tta tgg aag gac gtt ccg gaa gag aaa tgg aac gat tgg ctt tgg cag	96
Leu Trp Lys Asp Val Pro Glu Glu Lys Trp Asn Asp Trp Leu Trp Gln	
20 25 30	
ctg aca cac act gta aga acg tta gat gat tta aag aaa gtc att aat	144
Leu Thr His Thr Val Arg Thr Leu Asp Asp Leu Lys Lys Val Ile Asn	
35 40 45	
ctg acc gag gat gaa gag gaa ggc gtc aga att tct acc aaa acg atc	192
Leu Thr Glu Asp Glu Glu Glu Gly Val Arg Ile Ser Thr Lys Thr Ile	
50 55 60	
ccc tta aat att aca cct tac tat gct tct tta atg gac ccc gac aat	240
Pro Leu Asn Ile Thr Pro Tyr Tyr Ala Ser Leu Met Asp Pro Asp Asn	
65 70 75 80	
ccg aga tgc ccg gta cgc atg cag tct gtg ccg ctt tct gaa gaa atg	288
Pro Arg Cys Pro Val Arg Met Gln Ser Val Pro Leu Ser Glu Glu Met	
85 90 95	
cac aaa aca aaa tac gat atg gaa gac ccg ctt cat gag gat gaa gat	336
His Lys Thr Lys Tyr Asp Met Glu Asp Pro Leu His Glu Asp Glu Asp	
100 105 110	
tca ccg gta ccc ggt ctg aca cac cgc tat ccc gac cgt gtg ctg ttt	384
Ser Pro Val Pro Gly Leu Thr His Arg Tyr Pro Asp Arg Val Leu Phe	
115 120 125	
ctt gtc acg aat caa tgt tcc gtg tac tgc cgc tac tgc aca aga agg	432

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

355	360	365	
ggt gtg att acg tca tat ccg gaa cca gag aat tat atc ccc aat cag			1152
Gly Val Ile Thr Ser Tyr Pro Glu Pro Glu Asn Tyr Ile Pro Asn Gln			
370	375	380	
gca gac gcc tat ttt gag tcc gtt ttc cct gaa acc gct gac aaa aag			1200
Ala Asp Ala Tyr Phe Glu Ser Val Phe Pro Glu Thr Ala Asp Lys Lys			
385	390	395	400
gag ccg atc ggg ctg agt gcc att ttt gct gac aaa gaa gtt tcg ttt			1248
Glu Pro Ile Gly Leu Ser Ala Ile Phe Ala Asp Lys Glu Val Ser Phe			
405	410	415	
aca cct gaa aat gta gac aga atc aaa agg aga gag gca tac atc gca			1296
Thr Pro Glu Asn Val Asp Arg Ile Lys Arg Arg Glu Ala Tyr Ile Ala			
420	425	430	
aat ccg gag cat gaa aca tta aaa gat cgg cgt gag aaa aga gat cag			1344
Asn Pro Glu His Glu Thr Leu Lys Asp Arg Arg Glu Lys Arg Asp Gln			
435	440	445	
ctc aaa gaa aag aaa ttt ttg gcg cag cag aaa aaa cag aaa gag act			1392
Leu Lys Glu Lys Lys Phe Leu Ala Gln Gln Lys Lys Gln Lys Glu Thr			
450	455	460	
gaa tgc gga ggg gat tct tca tga			1416
Glu Cys Gly Gly Asp Ser Ser			
465	470		
<210> 2			
<211> 471			
<212> PRT			
<213> Bacillus subtilis			
<400> 2			
Met Lys Asn Lys Trp Tyr Lys Pro Lys Arg His Trp Lys Glu Ile Glu			
1	5	10	15
Leu Trp Lys Asp Val Pro Glu Glu Lys Trp Asn Asp Trp Leu Trp Gln			
	20	25	30
Leu Thr His Thr Val Arg Thr Leu Asp Asp Leu Lys Lys Val Ile Asn			
	35	40	45
Leu Thr Glu Asp Glu Glu Glu Gly Val Arg Ile Ser Thr Lys Thr Ile			
	50	55	60
Pro Leu Asn Ile Thr Pro Tyr Tyr Ala Ser Leu Met Asp Pro Asp Asn			
65	70	75	80

Glu Gly Ile Gly His Phe Arg Ala Pro Val Ser Lys Gly Leu Glu Ile
305 310 315 320

Ile Glu Gly Leu Arg Gly His Thr Ser Gly Tyr Ala Val Pro Thr Phe
325 330 335

Val Val His Ala Pro Gly Gly Gly Gly Lys Ile Ala Leu Gln Pro Asn
340 345 350

Tyr Val Leu Ser Gln Ser Pro Asp Lys Val Ile Leu Arg Asn Phe Glu
355 360 365

Gly Val Ile Thr Ser Tyr Pro Glu Pro Glu Asn Tyr Ile Pro Asn Gln
370 375 380

Ala Asp Ala Tyr Phe Glu Ser Val Phe Pro Glu Thr Ala Asp Lys Lys
385 390 395 400

Glu Pro Ile Gly Leu Ser Ala Ile Phe Ala Asp Lys Glu Val Ser Phe
405 410 415

Thr Pro Glu Asn Val Asp Arg Ile Lys Arg Arg Glu Ala Tyr Ile Ala
420 425 430

Asn Pro Glu His Glu Thr Leu Lys Asp Arg Arg Glu Lys Arg Asp Gln
435 440 445

Leu Lys Glu Lys Lys Phe Leu Ala Gln Gln Lys Lys Gln Lys Glu Thr
450 455 460

Glu Cys Gly Gly Asp Ser Ser
465 470

<210> 3
<211> 1416
<212> DNA
<213> Bacillus subtilis

<220>
<221> CDS
<222> (1)..(1416)

<400> 3
atg aaa aac aaa tgg tat aaa ccg aaa cgg cat tgg aag gag atc gag
Met Lys Asn Lys Trp Tyr Lys Pro Lys Arg His Trp Lys Glu Ile Glu

48

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

aac aca agc atc gaa atg aca gaa gaa tcc gtt gag gca tgt gaa aag Asn Thr Ser Ile Glu Met Thr Glu Glu Ser Val Glu Ala Cys Glu Lys 245 250 255	768
ctg gtg aac gcg gga gtg ccg gtc gga aat cag gct gtc gta tta gca Leu Val Asn Ala Gly Val Pro Val Gly Asn Gln Ala Val Val Leu Ala 260 265 270	816
ggt att aat gat tcg gtt cca att atg aaa aag ctc atg cat gac ttg Gly Ile Asn Asp Ser Val Pro Ile Met Lys Lys Leu Met His Asp Leu 275 280 285	864
gta aaa atc aga gtc cgt cct tat tat att tac caa tgt gat ctg tca Val Lys Ile Arg Val Arg Pro Tyr Tyr Ile Tyr Gln Cys Asp Leu Ser 290 295 300	912
gaa gga ata ggg cat ttc cgt gct cct gtt tcc aaa ggt ttg gag atc Glu Gly Ile Gly His Phe Arg Ala Pro Val Ser Lys Gly Leu Glu Ile 305 310 315 320	960
att gaa ggg ctg aga ggt cat acc tca ggc tat gcg gtt cct acc ttt Ile Glu Gly Leu Arg Gly His Thr Ser Gly Tyr Ala Val Pro Thr Phe 325 330 335	1008
gtc gtt cac gca cca ggc gga gga ggt aaa atc gcc ctg cag ccg aac Val Val His Ala Pro Gly Gly Gly Gly Lys Ile Ala Leu Gln Pro Asn 340 345 350	1056
tat gtc ctg tca caa agt cct gac aaa gtg atc tta aga aat ttt gaa Tyr Val Leu Ser Gln Ser Pro Asp Lys Val Ile Leu Arg Asn Phe Glu 355 360 365	1104
ggt gtg att acg tca tat ccg gaa cca gag aat tat atc ccc aat cag Gly Val Ile Thr Ser Tyr Pro Glu Pro Glu Asn Tyr Ile Pro Asn Gln 370 375 380	1152
gca gac gcc tat ttt gag tcc gtt ttc cct gaa acc gct gac aaa aag Ala Asp Ala Tyr Phe Glu Ser Val Phe Pro Glu Thr Ala Asp Lys Lys 385 390 395 400	1200
gag ccg atc ggg ctg agt gcc att ttt gct gac aaa gaa gtt tcg ttt Glu Pro Ile Gly Leu Ser Ala Ile Phe Ala Asp Lys Glu Val Ser Phe 405 410 415	1248
aca cct gaa aat gta gac aga atc aaa cgg cgt gag gca tac atc gca Thr Pro Glu Asn Val Asp Arg Ile Lys Arg Arg Glu Ala Tyr Ile Ala 420 425 430	1296
aat ccg gag cat gaa aca tta aaa gat cgg cgt gag aaa aga gat cag Asn Pro Glu His Glu Thr Leu Lys Asp Arg Arg Glu Lys Arg Asp Gln 435 440 445	1344
ctc aaa gaa aag aaa ttt ttg gcg cag cag aaa aaa cag aaa gag act Leu Lys Glu Lys Lys Phe Leu Ala Gln Gln Lys Lys Gln Lys Glu Thr 450 455 460	1392

gaa tgc gga ggg gat tct tca tga 1416
 Glu Cys Gly Gly Asp Ser Ser
 465 470

<210> 4
 <211> 471
 <212> PRT
 <213> Bacillus subtilis

<400> 4

Met Lys Asn Lys Trp Tyr Lys Pro Lys Arg His Trp Lys Glu Ile Glu
 1 5 10 15

Leu Trp Lys Asp Val Pro Glu Glu Lys Trp Asn Asp Trp Leu Trp Gln
 20 25 30

Leu Thr His Thr Val Arg Thr Leu Asp Asp Leu Lys Lys Val Ile Asn
 35 40 45

Leu Thr Glu Asp Glu Glu Glu Gly Val Arg Ile Ser Thr Lys Thr Ile
 50 55 60

Pro Leu Asn Ile Thr Pro Tyr Tyr Ala Ser Leu Met Asp Pro Asp Asn
 65 70 75 80

Pro Arg Cys Pro Val Arg Met Gln Ser Val Pro Leu Ser Glu Glu Met
 85 90 95

His Lys Thr Lys Tyr Asp Met Glu Asp Pro Leu His Glu Asp Glu Asp
 100 105 110

Ser Pro Val Pro Gly Leu Thr His Arg Tyr Pro Asp Arg Val Leu Phe
 115 120 125

Leu Val Thr Asn Gln Cys Ser Val Tyr Cys Arg His Cys Thr Arg Arg
 130 135 140

Arg Phe Ser Gly Gln Ile Gly Met Gly Val Pro Lys Lys Gln Leu Asp
 145 150 155 160

Ala Ala Ile Ala Tyr Ile Arg Glu Thr Pro Glu Ile Arg Asp Cys Leu
 165 170 175

Ile Ser Gly Gly Asp Gly Leu Leu Ile Asn Asp Gln Ile Leu Glu Tyr

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

Thr Pro Glu Asn Val Asp Arg Ile Lys Arg Arg Glu Ala Tyr Ile Ala
420 425 430

Asn Pro Glu His Glu Thr Leu Lys Asp Arg Arg Glu Lys Arg Asp Gln
435 440 445

Leu Lys Glu Lys Lys Phe Leu Ala Gln Gln Lys Lys Gln Lys Glu Thr
450 455 460

Glu Cys Gly Gly Asp Ser Ser
465 470

<210> 5
<211> 1251
<212> DNA
<213> Porphyromonas gingivalis

<220>
<221> CDS
<222> (1)..(1251)

<400> 5
atg gca gaa agt cgt aga aag tat tat ttc cct gat gtc acc gat gag 48
Met Ala Glu Ser Arg Arg Lys Tyr Tyr Phe Pro Asp Val Thr Asp Glu
1 5 10 15
caa tgg tac gac tgg cat tgg cag gtc ctc aat cga att gag acg ctc 96
Gln Trp Tyr Asp Trp His Trp Gln Val Leu Asn Arg Ile Glu Thr Leu
20 25 30
gac cag ctg aaa aag tac gtt aca ctc acc gct gaa gaa gaa gag gga 144
Asp Gln Leu Lys Lys Tyr Val Thr Leu Thr Ala Glu Glu Glu Glu Gly
35 40 45
gta aaa gaa tcg ccc aaa gta ctc cga atg gct atc aca cct tat tat 192
Val Lys Glu Ser Pro Lys Val Leu Arg Met Ala Ile Thr Pro Tyr Tyr
50 55 60
ttg agt ttg ata gac ccc gag aat cct aat tgt ccg att cgt aaa caa 240
Leu Ser Leu Ile Asp Pro Glu Asn Pro Asn Cys Pro Ile Arg Lys Gln
65 70 75 80
gcc att cct act caa cag gaa ctg gta cgt gct cct gaa gat cag gta 288
Ala Ile Pro Thr Gln Gln Glu Leu Val Arg Ala Pro Glu Asp Gln Val
85 90 95
gac cca ctt agt gaa gat gaa gat tcg ccc gta ccc gga ctg act cat 336
Asp Pro Leu Ser Glu Asp Glu Asp Ser Pro Val Pro Gly Leu Thr His
100 105 110

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

Gly Lys Ile Pro Val Thr Pro Asn Tyr Val Val Ser Gln Ser Pro Arg
 340 345 350

cat gtg gtt ctt cgc aat tat gaa ggt gtt atc aca acc tat acg gag 1104
 His Val Val Leu Arg Asn Tyr Glu Gly Val Ile Thr Thr Tyr Thr Glu
 355 360 365

ccg gag aat tat cat gag gag tgc gat tgt gag gac tgt cga gcc ggt 1152
 Pro Glu Asn Tyr His Glu Glu Cys Asp Cys Glu Asp Cys Arg Ala Gly
 370 375 380

aag cat aaa gag ggt gta gct gca ctt tcc gga ggt cag cag ttg gct 1200
 Lys His Lys Glu Gly Val Ala Ala Leu Ser Gly Gly Gln Gln Leu Ala
 385 390 395 400

atc gag cct tcc gac tta gct cgc aaa aaa cgc aag ttt gat aag aac 1248
 Ile Glu Pro Ser Asp Leu Ala Arg Lys Lys Arg Lys Phe Asp Lys Asn
 405 410 415

tga 1251

<210> 6
 <211> 416
 <212> PRT
 <213> Porphyromonas gingivalis

<400> 6

Met Ala Glu Ser Arg Arg Lys Tyr Tyr Phe Pro Asp Val Thr Asp Glu
 1 5 10 15

Gln Trp Tyr Asp Trp His Trp Gln Val Leu Asn Arg Ile Glu Thr Leu
 20 25 30

Asp Gln Leu Lys Lys Tyr Val Thr Leu Thr Ala Glu Glu Glu Glu Gly
 35 40 45

Val Lys Glu Ser Pro Lys Val Leu Arg Met Ala Ile Thr Pro Tyr Tyr
 50 55 60

Leu Ser Leu Ile Asp Pro Glu Asn Pro Asn Cys Pro Ile Arg Lys Gln
 65 70 75 80

Ala Ile Pro Thr Gln Gln Glu Leu Val Arg Ala Pro Glu Asp Gln Val
 85 90 95

Asp Pro Leu Ser Glu Asp Glu Asp Ser Pro Val Pro Gly Leu Thr His
 100 105 110

Arg Tyr Pro Asp Arg Val Leu Phe Leu Ile Thr Asp Lys Cys Ser Met
 115 120 125

Tyr Cys Arg His Cys Thr Arg Arg Arg Phe Ala Gly Gln Lys Asp Ala
 130 135 140

Ser Ser Pro Ser Glu Arg Ile Asp Arg Cys Ile Asp Tyr Ile Ala Asn
 145 150 155 160

Thr Pro Thr Val Arg Asp Val Leu Leu Ser Gly Gly Asp Ala Leu Leu
 165 170 175

Val Ser Asp Glu Arg Leu Glu Tyr Ile Leu Lys Arg Leu Arg Glu Ile
 180 185 190

Pro His Val Glu Ile Val Arg Ile Gly Ser Arg Thr Pro Val Val Leu
 195 200 205

Pro Gln Arg Ile Thr Pro Gln Leu Val Asp Met Leu Lys Lys Tyr His
 210 215 220

Pro Val Trp Leu Asn Thr His Phe Asn His Pro Asn Glu Val Thr Glu
 225 230 235 240

Glu Ala Val Glu Ala Cys Glu Arg Met Ala Asn Ala Gly Ile Pro Leu
 245 250 255

Gly Asn Gln Thr Val Leu Leu Arg Gly Ile Asn Asp Cys Thr His Val
 260 265 270

Met Lys Arg Leu Val His Leu Leu Val Lys Met Arg Val Arg Pro Tyr
 275 280 285

Tyr Ile Tyr Val Cys Asp Leu Ser Leu Gly Ile Gly His Phe Arg Thr
 290 295 300

Pro Val Ser Lys Gly Ile Glu Ile Ile Glu Asn Leu Arg Gly His Thr
 305 310 315 320

Ser Gly Tyr Ala Val Pro Thr Phe Val Val Gly Ala Pro Gly Gly Gly
 325 330 335

Gly Lys Ile Pro Val Thr Pro Asn Tyr Val Val Ser Gln Ser Pro Arg

340	345	350	
His Val Val Leu Arg Asn Tyr Glu Gly Val Ile Thr Thr Tyr Thr Glu			
355	360	365	
Pro Glu Asn Tyr His Glu Glu Cys Asp Cys Glu Asp Cys Arg Ala Gly			
370	375	380	
Lys His Lys Glu Gly Val Ala Ala Leu Ser Gly Gly Gln Gln Leu Ala			
385	390	395	400
Ile Glu Pro Ser Asp Leu Ala Arg Lys Lys Arg Lys Phe Asp Lys Asn			
	405	410	415
<210> 7			
<211> 29			
<212> DNA			
<213> Artificial			
<220>			
<223> PCR primer			
<400> 7			
ccggcccata tgaatacagt taatactag			29
<210> 8			
<211> 37			
<212> DNA			
<213> Artificial			
<220>			
<223> PCR primer			
<400> 8			
cgccgcggat ccttatttaa acaatctctc cctgtcg			37
<210> 9			
<211> 1278			
<212> DNA			
<213> Fusobacterium nucleatum			
<220>			
<221> CDS			
<222> (1)..(1278)			
<400> 9			
atg aat aca gtt aat act aga aaa aaa ttt ttc cca aat gta act gat			48
Met Asn Thr Val Asn Thr Arg Lys Lys Phe Phe Pro Asn Val Thr Asp			
1	5	10	15

gaa gaa tgg aat gat tgg aca tgg caa gta aaa aac aga ctt gaa agt Glu Glu Trp Asn Asp Trp Thr Trp Gln Val Lys Asn Arg Leu Glu Ser 20 25 30	96
gtt gaa gat tta aaa aaa tat gtt gat tta agt gaa gaa gaa aca gaa Val Glu Asp Leu Lys Lys Tyr Val Asp Leu Ser Glu Glu Glu Thr Glu 35 40 45	144
ggg gtt gta aga act ctt gaa act tta aga atg gca atc act cca tat Gly Val Val Arg Thr Leu Glu Thr Leu Arg Met Ala Ile Thr Pro Tyr 50 55 60	192
tac ttc tca ttg ata gat ttg aat agt gat aga tgc cca ata aga aag Tyr Phe Ser Leu Ile Asp Leu Asn Ser Asp Arg Cys Pro Ile Arg Lys 65 70 75 80	240
caa gct ata cct act ata caa gaa ata cat caa tct gat gct gat ttg Gln Ala Ile Pro Thr Ile Gln Glu Ile His Gln Ser Asp Ala Asp Leu 85 90 95	288
tta gat cct cta cat gaa gat gaa gac tct cca gta cca gga tta act Leu Asp Pro Leu His Glu Asp Glu Asp Ser Pro Val Pro Gly Leu Thr 100 105 110	336
cat aga tat cca gat aga gtt tta ctt cta ata aca gac atg tgt tct His Arg Tyr Pro Asp Arg Val Leu Leu Leu Ile Thr Asp Met Cys Ser 115 120 125	384
atg tat tgt aga cac tgc act cgt aga aga ttt gct ggg tca agt gat Met Tyr Cys Arg His Cys Thr Arg Arg Arg Phe Ala Gly Ser Ser Asp 130 135 140	432
gat gct atg cct atg gat aga att gac aaa gca ata gaa tat att gca Asp Ala Met Pro Met Asp Arg Ile Asp Lys Ala Ile Glu Tyr Ile Ala 145 150 155 160	480
aaa act cca caa gta agg gat gta ttg tta tca gga gga gat gca ctt Lys Thr Pro Gln Val Arg Asp Val Leu Leu Ser Gly Gly Asp Ala Leu 165 170 175	528
cta gtt tct gat aaa aaa tta gaa agc ata atc caa aaa cta aga gca Leu Val Ser Asp Lys Lys Leu Glu Ser Ile Ile Gln Lys Leu Arg Ala 180 185 190	576
ata cct cat gtt gaa ata ata aga ata gga agt aga aca cca gtt gtt Ile Pro His Val Glu Ile Ile Arg Ile Gly Ser Arg Thr Pro Val Val 195 200 205	624
tta cct caa aga att act cct gaa tta tgt aat atg tta aag aaa tat Leu Pro Gln Arg Ile Thr Pro Glu Leu Cys Asn Met Leu Lys Lys Tyr 210 215 220	672
cat cca att tgg ttg aat act cat ttt aac cac cct caa gaa gta aca His Pro Ile Trp Leu Asn Thr His Phe Asn His Pro Gln Glu Val Thr 225 230 235 240	720

cca gaa gct aaa aaa gct tgt gaa atg ttg gca gat gca gga gtt cca 768
 Pro Glu Ala Lys Lys Ala Cys Glu Met Leu Ala Asp Ala Gly Val Pro
 245 250 255

tta gga aat caa act gta cta tta aga gga ata aat gac agt gta cct 816
 Leu Gly Asn Gln Thr Val Leu Leu Arg Gly Ile Asn Asp Ser Val Pro
 260 265 270

gta atg aaa agg tta gta cat gat tta gta atg atg aga gta aga cct 864
 Val Met Lys Arg Leu Val His Asp Leu Val Met Met Arg Val Arg Pro
 275 280 285

tat tat att tac caa tgt gac tta tct atg gga ctt gaa cac ttc aga 912
 Tyr Tyr Ile Tyr Gln Cys Asp Leu Ser Met Gly Leu Glu His Phe Arg
 290 295 300

aca cca gtt tct aaa ggt ata gaa att att gaa gga tta aga gga cat 960
 Thr Pro Val Ser Lys Gly Ile Glu Ile Ile Glu Gly Leu Arg Gly His
 305 310 315 320

aca tct gga tat gca gta cca aca ttt gtt gtt gat gca cct ggt ggt 1008
 Thr Ser Gly Tyr Ala Val Pro Thr Phe Val Val Asp Ala Pro Gly Gly
 325 330 335

gga gga aaa act cca gta atg cct caa tat gta att tct caa tct cct 1056
 Gly Gly Lys Thr Pro Val Met Pro Gln Tyr Val Ile Ser Gln Ser Pro
 340 345 350

cat aga gta gtt tta aga aac ttt gaa gga gtt ata aca act tat aca 1104
 His Arg Val Val Leu Arg Asn Phe Glu Gly Val Ile Thr Thr Tyr Thr
 355 360 365

gaa cca gaa aat tat aca cat gaa cct tgt tat gat gaa gaa aaa ttt 1152
 Glu Pro Glu Asn Tyr Thr His Glu Pro Cys Tyr Asp Glu Glu Lys Phe
 370 375 380

gaa aaa atg tat gaa ata agt gga gtt tat atg cta gat gaa gga tta 1200
 Glu Lys Met Tyr Glu Ile Ser Gly Val Tyr Met Leu Asp Glu Gly Leu
 385 390 395 400

aaa atg tca cta gaa cct agc cac tta gca aga cat gaa aga aat aaa 1248
 Lys Met Ser Leu Glu Pro Ser His Leu Ala Arg His Glu Arg Asn Lys
 405 410 415

aag aga gca gaa gct gaa ggg aaa aaa taa 1278
 Lys Arg Ala Glu Ala Glu Gly Lys Lys
 420 425

<210> 10
 <211> 425
 <212> PRT
 <213> Fusobacterium nucleatum

<400> 10

Met Asn Thr Val Asn Thr Arg Lys Lys Phe Phe Pro Asn Val Thr Asp

Pro Glu Ala Lys Lys Ala Cys Glu Met Leu Ala Asp Ala Gly Val Pro
 245 250 255

Leu Gly Asn Gln Thr Val Leu Leu Arg Gly Ile Asn Asp Ser Val Pro
 260 265 270

Val Met Lys Arg Leu Val His Asp Leu Val Met Met Arg Val Arg Pro
 275 280 285

Tyr Tyr Ile Tyr Gln Cys Asp Leu Ser Met Gly Leu Glu His Phe Arg
 290 295 300

Thr Pro Val Ser Lys Gly Ile Glu Ile Ile Glu Gly Leu Arg Gly His
 305 310 315 320

Thr Ser Gly Tyr Ala Val Pro Thr Phe Val Val Asp Ala Pro Gly Gly
 325 330 335

Gly Gly Lys Thr Pro Val Met Pro Gln Tyr Val Ile Ser Gln Ser Pro
 340 345 350

His Arg Val Val Leu Arg Asn Phe Glu Gly Val Ile Thr Thr Tyr Thr
 355 360 365

Glu Pro Glu Asn Tyr Thr His Glu Pro Cys Tyr Asp Glu Glu Lys Phe
 370 375 380

Glu Lys Met Tyr Glu Ile Ser Gly Val Tyr Met Leu Asp Glu Gly Leu
 385 390 395 400

Lys Met Ser Leu Glu Pro Ser His Leu Ala Arg His Glu Arg Asn Lys
 405 410 415

Lys Arg Ala Glu Ala Glu Gly Lys Lys
 420 425

<210> 11
 <211> 1278
 <212> DNA
 <213> Fusobacterium nucleatum

<220>

```

<221> CDS
<222> (1)..(1278)

<400> 11
atg aat aca gtt aat act cgt aaa aaa ttt ttc cca aat gta act gat      48
Met Asn Thr Val Asn Thr Arg Lys Lys Phe Phe Pro Asn Val Thr Asp
1                               5                               10                               15

gaa gaa tgg aat gat tgg aca tgg caa gta aaa aac cgc ctt gaa agt      96
Glu Glu Trp Asn Asp Trp Thr Trp Gln Val Lys Asn Arg Leu Glu Ser
                20                               25                               30

gtt gaa gat tta aaa aaa tat gtt gat tta agt gaa gaa gaa aca gaa      144
Val Glu Asp Leu Lys Lys Tyr Val Asp Leu Ser Glu Glu Glu Thr Glu
                35                               40                               45

ggg gtt gta cgc act ctt gaa act tta cgt atg gca atc act cca tat      192
Gly Val Val Arg Thr Leu Glu Thr Leu Arg Met Ala Ile Thr Pro Tyr
                50                               55                               60

tac ttc tca ttg ata gat ttg aat agt gat cgc tgc cca ata cgt aag      240
Tyr Phe Ser Leu Ile Asp Leu Asn Ser Asp Arg Cys Pro Ile Arg Lys
        65                               70                               75                               80

caa gct ata cct act ata caa gaa ata cat caa tct gat gct gat ttg      288
Gln Ala Ile Pro Thr Ile Gln Glu Ile His Gln Ser Asp Ala Asp Leu
                85                               90                               95

tta gat cct cta cat gaa gat gaa gac tct cca gta cca gga tta act      336
Leu Asp Pro Leu His Glu Asp Glu Asp Ser Pro Val Pro Gly Leu Thr
                100                              105                              110

cat cgc tat cca gat cgt gtt tta ctt cta ata aca gac atg tgt tct      384
His Arg Tyr Pro Asp Arg Val Leu Leu Leu Ile Thr Asp Met Cys Ser
                115                              120                              125

atg tat tgt cgc cac tgc act cgt cgc aga ttt gct ggg tca agt gat      432
Met Tyr Cys Arg His Cys Thr Arg Arg Arg Phe Ala Gly Ser Ser Asp
                130                              135                              140

gat gct atg cct atg gat aga att gac aaa gca ata gaa tat att gca      480
Asp Ala Met Pro Met Asp Arg Ile Asp Lys Ala Ile Glu Tyr Ile Ala
        145                               150                               155                               160

aaa act cca caa gta agg gat gta ttg tta tca gga gga gat gca ctt      528
Lys Thr Pro Gln Val Arg Asp Val Leu Leu Ser Gly Gly Asp Ala Leu
                165                               170                               175

cta gtt tct gat aaa aaa tta gaa agc ata atc caa aaa cta cgc gca      576
Leu Val Ser Asp Lys Lys Leu Glu Ser Ile Ile Gln Lys Leu Arg Ala
                180                               185                               190

ata cct cat gtt gaa ata atc aga ata gga agt cgt aca cca gtt gtt      624
Ile Pro His Val Glu Ile Ile Arg Ile Gly Ser Arg Thr Pro Val Val
                195                               200                               205

tta cct caa aga att act cct gaa tta tgt aat atg tta aag aaa tat      672

```

Leu	Pro	Gln	Arg	Ile	Thr	Pro	Glu	Leu	Cys	Asn	Met	Leu	Lys	Lys	Tyr		
	210					215					220						
cat	cca	att	tgg	ttg	aat	act	cat	ttt	aac	cac	cct	caa	gaa	gta	acg		720
His	Pro	Ile	Trp	Leu	Asn	Thr	His	Phe	Asn	His	Pro	Gln	Glu	Val	Thr		240
	225				230				235								
cca	gaa	gct	aaa	aaa	gct	tgt	gaa	atg	ttg	gca	gat	gca	gga	gtt	cca		768
Pro	Glu	Ala	Lys	Lys	Ala	Cys	Glu	Met	Leu	Ala	Asp	Ala	Gly	Val	Pro		255
				245					250					255			
tta	gga	aat	caa	act	gta	cta	tta	aga	gga	ata	aat	gac	agt	gta	cct		816
Leu	Gly	Asn	Gln	Thr	Val	Leu	Leu	Arg	Gly	Ile	Asn	Asp	Ser	Val	Pro		270
			260					265						270			
gta	atg	aaa	agg	tta	gta	cat	gat	tta	gta	atg	atg	cgt	gta	cgc	cct		864
Val	Met	Lys	Arg	Leu	Val	His	Asp	Leu	Val	Met	Met	Arg	Val	Arg	Pro		285
		275					280					285					
tat	tat	att	tac	caa	tgt	gac	tta	tct	atg	gga	ctc	gaa	cac	ttc	cgc		912
Tyr	Tyr	Ile	Tyr	Gln	Cys	Asp	Leu	Ser	Met	Gly	Leu	Glu	His	Phe	Arg		300
		290				295						300					
aca	cca	gtt	tct	aaa	ggt	ata	gaa	att	att	gaa	gga	tta	cgt	gga	cat		960
Thr	Pro	Val	Ser	Lys	Gly	Ile	Glu	Ile	Ile	Glu	Gly	Leu	Arg	Gly	His		320
					310					315							
aca	tct	gga	tat	gca	gta	cca	aca	ttt	gtt	gtt	gat	gca	cct	ggt	ggt		1008
Thr	Ser	Gly	Tyr	Ala	Val	Pro	Thr	Phe	Val	Val	Asp	Ala	Pro	Gly	Gly		335
				325					330					335			
gga	gga	aaa	act	cca	gta	atg	cct	caa	tat	gta	att	tct	caa	tct	cct		1056
Gly	Gly	Lys	Thr	Pro	Val	Met	Pro	Gln	Tyr	Val	Ile	Ser	Gln	Ser	Pro		350
			340					345					350				
cat	cgt	gta	gtt	tta	cgc	aac	ttt	gaa	gga	gtt	ata	aca	act	tat	aca		1104
His	Arg	Val	Val	Leu	Arg	Asn	Phe	Glu	Gly	Val	Ile	Thr	Thr	Tyr	Thr		365
		355				360							365				
gaa	cca	gaa	aat	tat	aca	cat	gaa	cct	tgt	tat	gat	gaa	gaa	aaa	ttt		1152
Glu	Pro	Glu	Asn	Tyr	Thr	His	Glu	Pro	Cys	Tyr	Asp	Glu	Glu	Lys	Phe		380
		370				375					380						
gaa	aaa	atg	tat	gaa	ata	agt	gga	gtt	tat	atg	cta	gat	gaa	gga	tta		1200
Glu	Lys	Met	Tyr	Glu	Ile	Ser	Gly	Val	Tyr	Met	Leu	Asp	Glu	Gly	Leu		400
		385			390					395							
aaa	atg	tca	cta	gaa	cct	agc	cac	tta	gca	cgt	cat	gaa	cgc	aat	aaa		1248
Lys	Met	Ser	Leu	Glu	Pro	Ser	His	Leu	Ala	Arg	His	Glu	Arg	Asn	Lys		415
				405					410					415			
aag	aga	gca	gaa	gct	gaa	ggg	aaa	aaa	taa								1278
Lys	Arg	Ala	Glu	Ala	Glu	Gly	Lys	Lys									425
			420					425									

<210> 12

<211> 425
 <212> PRT
 <213> Fusobacterium nucleatum

 <400> 12

 Met Asn Thr Val Asn Thr Arg Lys Lys Phe Phe Pro Asn Val Thr Asp
 1 5 10 15

 Glu Glu Trp Asn Asp Trp Thr Trp Gln Val Lys Asn Arg Leu Glu Ser
 20 25 30

 Val Glu Asp Leu Lys Lys Tyr Val Asp Leu Ser Glu Glu Glu Thr Glu
 35 40 45

 Gly Val Val Arg Thr Leu Glu Thr Leu Arg Met Ala Ile Thr Pro Tyr
 50 55 60

 Tyr Phe Ser Leu Ile Asp Leu Asn Ser Asp Arg Cys Pro Ile Arg Lys
 65 70 75 80

 Gln Ala Ile Pro Thr Ile Gln Glu Ile His Gln Ser Asp Ala Asp Leu
 85 90 95

 Leu Asp Pro Leu His Glu Asp Glu Asp Ser Pro Val Pro Gly Leu Thr
 100 105 110

 His Arg Tyr Pro Asp Arg Val Leu Leu Leu Ile Thr Asp Met Cys Ser
 115 120 125

 Met Tyr Cys Arg His Cys Thr Arg Arg Arg Phe Ala Gly Ser Ser Asp
 130 135 140

 Asp Ala Met Pro Met Asp Arg Ile Asp Lys Ala Ile Glu Tyr Ile Ala
 145 150 155 160

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

 Leu Val Ser Asp Lys Lys Leu Glu Ser Ile Ile Gln Lys Leu Arg Ala
 180 185 190

 Ile Pro His Val Glu Ile Ile Arg Ile Gly Ser Arg Thr Pro Val Val
 195 200 205

Leu Pro Gln Arg Ile Thr Pro Glu Leu Cys Asn Met Leu Lys Lys Tyr
 210 215 220
 His Pro Ile Trp Leu Asn Thr His Phe Asn His Pro Gln Glu Val Thr
 225 230 235 240
 Pro Glu Ala Lys Lys Ala Cys Glu Met Leu Ala Asp Ala Gly Val Pro
 245 250 255
 Leu Gly Asn Gln Thr Val Leu Leu Arg Gly Ile Asn Asp Ser Val Pro
 260 265 270
 Val Met Lys Arg Leu Val His Asp Leu Val Met Met Arg Val Arg Pro
 275 280 285
 Tyr Tyr Ile Tyr Gln Cys Asp Leu Ser Met Gly Leu Glu His Phe Arg
 290 295 300
 Thr Pro Val Ser Lys Gly Ile Glu Ile Ile Glu Gly Leu Arg Gly His
 305 310 315 320
 Thr Ser Gly Tyr Ala Val Pro Thr Phe Val Val Asp Ala Pro Gly Gly
 325 330 335
 Gly Gly Lys Thr Pro Val Met Pro Gln Tyr Val Ile Ser Gln Ser Pro
 340 345 350
 His Arg Val Val Leu Arg Asn Phe Glu Gly Val Ile Thr Thr Tyr Thr
 355 360 365
 Glu Pro Glu Asn Tyr Thr His Glu Pro Cys Tyr Asp Glu Glu Lys Phe
 370 375 380
 Glu Lys Met Tyr Glu Ile Ser Gly Val Tyr Met Leu Asp Glu Gly Leu
 385 390 395 400
 Lys Met Ser Leu Glu Pro Ser His Leu Ala Arg His Glu Arg Asn Lys
 405 410 415
 Lys Arg Ala Glu Ala Glu Gly Lys Lys
 420 425

```

<210> 13
<211> 1278
<212> DNA
<213> Fusobacterium nucleatum

<220>
<221> CDS
<222> (1)..(1278)

<400> 13
atg aat aca gtt aat act cgt aaa aaa ttt ttc cca aat gta act gat      48
Met Asn Thr Val Asn Thr Arg Lys Lys Phe Phe Pro Asn Val Thr Asp
1          5          10          15

gaa gaa tgg aat gat tgg aca tgg caa gta aaa aac cgc ctt gaa agt      96
Glu Glu Trp Asn Asp Trp Thr Trp Gln Val Lys Asn Arg Leu Glu Ser
          20          25          30

gtt gaa gat tta aaa aaa tat gtt gat tta agt gaa gaa gaa aca gaa      144
Val Glu Asp Leu Lys Lys Tyr Val Asp Leu Ser Glu Glu Glu Thr Glu
          35          40          45

ggg gtt gta cgc act ctt gaa act tta cgt atg gca atc act cca tat      192
Gly Val Val Arg Thr Leu Glu Thr Leu Arg Met Ala Ile Thr Pro Tyr
          50          55          60

tac ttc tca ttg ata gat ttg aat agt gat cgc tgc cca ata cgt aag      240
Tyr Phe Ser Leu Ile Asp Leu Asn Ser Asp Arg Cys Pro Ile Arg Lys
65          70          75          80

caa gct ata cct act ata caa gaa ata cat caa tct gat gct gat atg      288
Gln Ala Ile Pro Thr Ile Gln Glu Ile His Gln Ser Asp Ala Asp Met
          85          90          95

ttg gat cct cta cat gaa gat gaa gac tct cca gta cca gga tta act      336
Leu Asp Pro Leu His Glu Asp Glu Asp Ser Pro Val Pro Gly Leu Thr
          100          105          110

cat cgc tat cca gat cgt gtt tta ctt cta ata aca gac atg tgt tct      384
His Arg Tyr Pro Asp Arg Val Leu Leu Leu Ile Thr Asp Met Cys Ser
          115          120          125

gta tac tgt cgc cac tgc act cgt cgc aga ttt gct ggg tca agt gat      432
Val Tyr Cys Arg His Cys Thr Arg Arg Arg Phe Ala Gly Ser Ser Asp
          130          135          140

gat gct atg cct atg gat aga att gac aaa gca ata gaa tat att gca      480
Asp Ala Met Pro Met Asp Arg Ile Asp Lys Ala Ile Glu Tyr Ile Ala
145          150          155          160

aaa act cca caa gta agg gat gta ttg tta tca gga gga gat gca ctt      528
Lys Thr Pro Gln Val Arg Asp Val Leu Leu Ser Gly Gly Asp Ala Leu
          165          170          175

cta gtt tct gat aaa aaa tta gaa agc ata atc caa aaa cta cgc gca      576
Leu Val Ser Asp Lys Lys Leu Glu Ser Ile Ile Gln Lys Leu Arg Ala

```

180				185				190								
ata	cct	cat	gtt	gaa	ata	atc	aga	ata	gga	agt	cgt	aca	cca	gtt	gtt	624
Ile	Pro	His	Val	Glu	Ile	Ile	Arg	Ile	Gly	Ser	Arg	Thr	Pro	Val	Val	
		195					200							205		
tta	cct	caa	aga	att	act	cct	gaa	tta	tgt	aat	atg	tta	aag	aaa	tat	672
Leu	Pro	Gln	Arg	Ile	Thr	Pro	Glu	Leu	Cys	Asn	Met	Leu	Lys	Lys	Tyr	
		210					215									
cat	cca	att	tgg	ttg	aat	act	cat	ttt	aac	cac	cct	caa	gaa	gta	acg	720
His	Pro	Ile	Trp	Leu	Asn	Thr	His	Phe	Asn	His	Pro	Gln	Glu	Val	Thr	
		225				230					235				240	
cca	gaa	gct	aaa	aaa	gct	tgt	gaa	atg	ttg	gca	gat	gca	gga	gtt	cca	768
Pro	Glu	Ala	Lys	Lys	Ala	Cys	Glu	Met	Leu	Ala	Asp	Ala	Gly	Val	Pro	
			245								250				255	
tta	gga	aat	caa	act	gta	cta	tta	aga	gga	ata	aat	gac	agt	gta	cct	816
Leu	Gly	Asn	Gln	Thr	Val	Leu	Leu	Arg	Gly	Ile	Asn	Asp	Ser	Val	Pro	
			260												270	
gta	atg	aaa	agg	tta	gta	cat	gat	tta	gta	atg	atg	cgt	gta	cgc	cct	864
Val	Met	Lys	Arg	Leu	Val	His	Asp	Leu	Val	Met	Met	Arg	Val	Arg	Pro	
			275												285	
tat	tat	att	tac	caa	tgt	gac	tta	tct	atg	gga	ctc	gaa	cac	ttc	cgc	912
Tyr	Tyr	Ile	Tyr	Gln	Cys	Asp	Leu	Ser	Met	Gly	Leu	Glu	His	Phe	Arg	
		290				295									300	
aca	cca	gtt	tct	aaa	ggt	ata	gaa	att	att	gaa	gga	tta	cgt	gga	cat	960
Thr	Pro	Val	Ser	Lys	Gly	Ile	Glu	Ile	Ile	Glu	Gly	Leu	Arg	Gly	His	
						310									320	
aca	tct	gga	tat	gca	gta	cca	aca	ttt	ggt	gtg	cat	gca	cct	ggt	ggt	1008
Thr	Ser	Gly	Tyr	Ala	Val	Pro	Thr	Phe	Val	Val	His	Ala	Pro	Gly	Gly	
						325									335	
gga	gga	aaa	act	cca	gta	atg	cct	caa	tat	gta	att	tct	caa	tct	cct	1056
Gly	Gly	Lys	Thr	Pro	Val	Met	Pro	Gln	Tyr	Val	Ile	Ser	Gln	Ser	Pro	
			340												350	
cat	cgt	gta	gtt	tta	cgc	aac	ttt	gaa	gga	ggt	ata	aca	act	tat	aca	1104
His	Arg	Val	Val	Leu	Arg	Asn	Phe	Glu	Gly	Val	Ile	Thr	Thr	Tyr	Thr	
			355												365	
gaa	cca	gaa	aat	tat	aca	cat	gaa	cct	tgt	tat	gat	gaa	gaa	aaa	ttt	1152
Glu	Pro	Glu	Asn	Tyr	Thr	His	Glu	Pro	Cys	Tyr	Asp	Glu	Glu	Lys	Phe	
			370												380	
gaa	aaa	atg	tat	gaa	ata	agt	gga	ggt	tat	atg	cta	gat	gaa	gga	tta	1200
Glu	Lys	Met	Tyr	Glu	Ile	Ser	Gly	Val	Tyr	Met	Leu	Asp	Glu	Gly	Leu	
						390									400	
aaa	atg	tca	cta	gaa	cct	agc	cac	tta	gca	cgt	cat	gaa	cgc	aat	aaa	1248
Lys	Met	Ser	Leu	Glu	Pro	Ser	His	Leu	Ala	Arg	His	Glu	Arg	Asn	Lys	
						405									415	

aag aga gca gaa gct gaa ggg aaa aaa taa 1278
 Lys Arg Ala Glu Ala Glu Gly Lys Lys
 420 425

<210> 14
 <211> 425
 <212> PRT
 <213> Fusobacterium nucleatum
 <400> 14

Met Asn Thr Val Asn Thr Arg Lys Lys Phe Phe Pro Asn Val Thr Asp
 1 5 10 15

Glu Glu Trp Asn Asp Trp Thr Trp Gln Val Lys Asn Arg Leu Glu Ser
 20 25 30

Val Glu Asp Leu Lys Lys Tyr Val Asp Leu Ser Glu Glu Glu Thr Glu
 35 40 45

Gly Val Val Arg Thr Leu Glu Thr Leu Arg Met Ala Ile Thr Pro Tyr
 50 55 60

Tyr Phe Ser Leu Ile Asp Leu Asn Ser Asp Arg Cys Pro Ile Arg Lys
 65 70 75 80

Gln Ala Ile Pro Thr Ile Gln Glu Ile His Gln Ser Asp Ala Asp Met
 85 90 95

Leu Asp Pro Leu His Glu Asp Glu Asp Ser Pro Val Pro Gly Leu Thr
 100 105 110

His Arg Tyr Pro Asp Arg Val Leu Leu Leu Ile Thr Asp Met Cys Ser
 115 120 125

Val Tyr Cys Arg His Cys Thr Arg Arg Arg Phe Ala Gly Ser Ser Asp
 130 135 140

Asp Ala Met Pro Met Asp Arg Ile Asp Lys Ala Ile Glu Tyr Ile Ala
 145 150 155 160

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

Leu Val Ser Asp Lys Lys Leu Glu Ser Ile Ile Gln Lys Leu Arg Ala
 180 185 190

Ile Pro His Val Glu Ile Ile Arg Ile Gly Ser Arg Thr Pro Val Val
 195 200 205

Leu Pro Gln Arg Ile Thr Pro Glu Leu Cys Asn Met Leu Lys Lys Tyr
 210 215 220

His Pro Ile Trp Leu Asn Thr His Phe Asn His Pro Gln Glu Val Thr
 225 230 235 240

Pro Glu Ala Lys Lys Ala Cys Glu Met Leu Ala Asp Ala Gly Val Pro
 245 250 255

Leu Gly Asn Gln Thr Val Leu Leu Arg Gly Ile Asn Asp Ser Val Pro
 260 265 270

Val Met Lys Arg Leu Val His Asp Leu Val Met Met Arg Val Arg Pro
 275 280 285

Tyr Tyr Ile Tyr Gln Cys Asp Leu Ser Met Gly Leu Glu His Phe Arg
 290 295 300

Thr Pro Val Ser Lys Gly Ile Glu Ile Ile Glu Gly Leu Arg Gly His
 305 310 315 320

Thr Ser Gly Tyr Ala Val Pro Thr Phe Val Val His Ala Pro Gly Gly
 325 330 335

Gly Gly Lys Thr Pro Val Met Pro Gln Tyr Val Ile Ser Gln Ser Pro
 340 345 350

His Arg Val Val Leu Arg Asn Phe Glu Gly Val Ile Thr Thr Tyr Thr
 355 360 365

Glu Pro Glu Asn Tyr Thr His Glu Pro Cys Tyr Asp Glu Glu Lys Phe
 370 375 380

Glu Lys Met Tyr Glu Ile Ser Gly Val Tyr Met Leu Asp Glu Gly Leu
 385 390 395 400

Lys Met Ser Leu Glu Pro Ser His Leu Ala Arg His Glu Arg Asn Lys

405	410	415
Lys Arg Ala Glu Ala Glu Gly Lys Lys 420 425		
<210> 15		
<211> 40		
<212> DNA		
<213> Artificial		
<220>		
<223> mutagenesis primer		
<400> 15		
catcaatctg atgctgatat gttggatcct ctacatgaag		40
<210> 16		
<211> 38		
<212> DNA		
<213> Artificial		
<220>		
<223> mutagenesis primer		
<400> 16		
aacagacatg tgttctgtat actgtcgcca ctgcactc		38
<210> 17		
<211> 34		
<212> DNA		
<213> Artificial		
<220>		
<223> mutagenesis primer		
<400> 17		
gtaccaacat ttgttgtgca tgcacctggt ggtg		34
<210> 18		
<211> 1278		
<212> DNA		
<213> Fusobacterium nucleatum		
<220>		
<221> CDS		
<222> (1)..(1278)		
<400> 18		
atg aat aca gtt aat act cgt aaa aaa ttt ttc cca aat gta act gat		48
Met Asn Thr Val Asn Thr Arg Lys Lys Phe Phe Pro Asn Val Thr Asp		
1 5 10 15		

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

cca gaa gct aaa aaa gct tgt gaa atg ttg gca gat gca gga gtt cca 768
 Pro Glu Ala Lys Lys Ala Cys Glu Met Leu Ala Asp Ala Gly Val Pro
 245 250 255

tta gga aat caa act gta cta tta aga gga ata aat gac agt gta cct 816
 Leu Gly Asn Gln Thr Val Leu Leu Arg Gly Ile Asn Asp Ser Val Pro
 260 265 270

gta atg aaa agg tta gta cat gat tta gta atg atg cgt gta cgc cct 864
 Val Met Lys Arg Leu Val His Asp Leu Val Met Met Arg Val Arg Pro
 275 280 285

tat tat att tac caa tgt gac tta tct atg gga ctc gaa cac ttc cgc 912
 Tyr Tyr Ile Tyr Gln Cys Asp Leu Ser Met Gly Leu Glu His Phe Arg
 290 295 300

aca cca gtt tct aaa ggt ata gaa att att gaa gga tta cgt gga cat 960
 Thr Pro Val Ser Lys Gly Ile Glu Ile Ile Glu Gly Leu Arg Gly His
 305 310 315 320

aca tct gga tat gca gta cca aca ttt gtt gtg cat gca cct ggt ggt 1008
 Thr Ser Gly Tyr Ala Val Pro Thr Phe Val Val His Ala Pro Gly Gly
 325 330 335

gga gga aaa act cca gta atg cct caa tat gta att tct caa tct cct 1056
 Gly Gly Lys Thr Pro Val Met Pro Gln Tyr Val Ile Ser Gln Ser Pro
 340 345 350

cat cgt gta gtt tta cgc aac ttt gaa gga gtt ata aca act tat aca 1104
 His Arg Val Val Leu Arg Asn Phe Glu Gly Val Ile Thr Thr Tyr Thr
 355 360 365

gaa cca gaa aat tat aca cat gaa cct tgt tat gat gaa gaa aaa ttt 1152
 Glu Pro Glu Asn Tyr Thr His Glu Pro Cys Tyr Asp Glu Glu Lys Phe
 370 375 380

gaa aaa atg tat gaa ata agt gga gtt tat atg cta gat gaa gga tta 1200
 Glu Lys Met Tyr Glu Ile Ser Gly Val Tyr Met Leu Asp Glu Gly Leu
 385 390 395 400

gaa atg tca cta gaa cct agc cac tta gca cgt cat gaa cgc aat aaa 1248
 Glu Met Ser Leu Glu Pro Ser His Leu Ala Arg His Glu Arg Asn Lys
 405 410 415

aag aga gca gaa gct gaa ggg aaa aaa taa 1278
 Lys Arg Ala Glu Ala Glu Gly Lys Lys
 420 425

<210> 19

<211> 425

<212> PRT

<213> Fusobacterium nucleatum

<400> 19

Met Asn Thr Val Asn Thr Arg Lys Lys Phe Phe Pro Asn Val Thr Asp

Pro Glu Ala Lys Lys Ala Cys Glu Met Leu Ala Asp Ala Gly Val Pro
 245 250 255

Leu Gly Asn Gln Thr Val Leu Leu Arg Gly Ile Asn Asp Ser Val Pro
 260 265 270

Val Met Lys Arg Leu Val His Asp Leu Val Met Met Arg Val Arg Pro
 275 280 285

Tyr Tyr Ile Tyr Gln Cys Asp Leu Ser Met Gly Leu Glu His Phe Arg
 290 295 300

Thr Pro Val Ser Lys Gly Ile Glu Ile Ile Glu Gly Leu Arg Gly His
 305 310 315 320

Thr Ser Gly Tyr Ala Val Pro Thr Phe Val Val His Ala Pro Gly Gly
 325 330 335

Gly Gly Lys Thr Pro Val Met Pro Gln Tyr Val Ile Ser Gln Ser Pro
 340 345 350

His Arg Val Val Leu Arg Asn Phe Glu Gly Val Ile Thr Thr Tyr Thr
 355 360 365

Glu Pro Glu Asn Tyr Thr His Glu Pro Cys Tyr Asp Glu Glu Lys Phe
 370 375 380

Glu Lys Met Tyr Glu Ile Ser Gly Val Tyr Met Leu Asp Glu Gly Leu
 385 390 395 400

Glu Met Ser Leu Glu Pro Ser His Leu Ala Arg His Glu Arg Asn Lys
 405 410 415

Lys Arg Ala Glu Ala Glu Gly Lys Lys
 420 425

<210> 20
 <211> 1278
 <212> DNA
 <213> Fusobacterium nucleatum

<220>

```

<221> CDS
<222> (1)..(1278)

<400> 20
atg aat aca gtt aat act cgt aaa aaa ttt ttc cca aat gta act gat      48
Met Asn Thr Val Asn Thr Arg Lys Lys Phe Phe Pro Asn Val Thr Asp
1          5          10          15

gaa gaa tgg aat gat tgg aca tgg caa gta aaa aac cgc ctt aaa agt      96
Glu Glu Trp Asn Asp Trp Thr Trp Gln Val Lys Asn Arg Leu Lys Ser
          20          25          30

gtt gaa gat tta gaa aaa tat gtt gat tta agt gaa gaa gaa aca gaa      144
Val Glu Asp Leu Glu Lys Tyr Val Asp Leu Ser Glu Glu Glu Thr Glu
          35          40          45

ggg gtt gta cgc act ctt gaa act tta cgt atg gca atc act cca ttt      192
Gly Val Val Arg Thr Leu Glu Thr Leu Arg Met Ala Ile Thr Pro Phe
          50          55          60

tac ttc tca ttg ata gat ttg aat agt gat cgc tgc cca ata cgt aag      240
Tyr Phe Ser Leu Ile Asp Leu Asn Ser Asp Arg Cys Pro Ile Arg Lys
65          70          75          80

caa gct ata cct act ata cga gaa ata cat caa tct gat gct gat atg      288
Gln Ala Ile Pro Thr Ile Arg Glu Ile His Gln Ser Asp Ala Asp Met
          85          90          95

ttg gat cct cta cat gaa gat gaa gac tct cca gta cca gga tta act      336
Leu Asp Pro Leu His Glu Asp Glu Asp Ser Pro Val Pro Gly Leu Thr
          100          105          110

cat cgc tat cca gat cgt gtt tta ctt cta ata aca gac atg tgt tct      384
His Arg Tyr Pro Asp Arg Val Leu Leu Leu Ile Thr Asp Met Cys Ser
          115          120          125

gta tac tgt cgc cac tgc act cgt cgc aga ttt gct ggg tca agt gat      432
Val Tyr Cys Arg His Cys Thr Arg Arg Arg Phe Ala Gly Ser Ser Asp
          130          135          140

ggt gct atg cct atg gat aga att gac aaa gca ata gaa tat att gca      480
Gly Ala Met Pro Met Asp Arg Ile Asp Lys Ala Ile Glu Tyr Ile Ala
145          150          155          160

aaa act cca caa gta agg gat gta ttg tta tca gga gga gat gca ctt      528
Lys Thr Pro Gln Val Arg Asp Val Leu Leu Ser Gly Gly Asp Ala Leu
          165          170          175

cta gtt tct aat aaa aaa tta gaa agc ata atc caa aaa cta cgc gca      576
Leu Val Ser Asn Lys Lys Leu Glu Ser Ile Ile Gln Lys Leu Arg Ala
          180          185          190

ata cct cat gtt gaa ata atc aga ata gga agt cgt aca cca gtt gtt      624
Ile Pro His Val Glu Ile Ile Arg Ile Gly Ser Arg Thr Pro Val Val
          195          200          205

tta cct caa aga att act cct gaa tta tgt aat atg tta aag aaa tat      672

```

Leu	Pro	Gln	Arg	Ile	Thr	Pro	Glu	Leu	Cys	Asn	Met	Leu	Lys	Lys	Tyr		
	210					215					220						
cat	cca	att	tgg	atg	aat	act	cat	ttt	aac	cac	cct	caa	gaa	gta	acg		720
His	Pro	Ile	Trp	Met	Asn	Thr	His	Phe	Asn	His	Pro	Gln	Glu	Val	Thr		
	225				230				235						240		
cca	gaa	gct	aaa	aaa	gct	tgt	gaa	atg	ttg	gca	gat	gca	gga	gtt	cca		768
Pro	Glu	Ala	Lys	Lys	Ala	Cys	Glu	Met	Leu	Ala	Asp	Ala	Gly	Val	Pro		
				245					250					255			
tta	gga	aat	caa	act	gta	cta	tta	aga	gga	ata	aat	gac	agt	gta	cct		816
Leu	Gly	Asn	Gln	Thr	Val	Leu	Leu	Arg	Gly	Ile	Asn	Asp	Ser	Val	Pro		
			260					265					270				
gta	atg	aaa	agg	tta	gta	cat	gat	tta	gta	atg	atg	cgt	gta	cgc	cct		864
Val	Met	Lys	Arg	Leu	Val	His	Asp	Leu	Val	Met	Met	Arg	Val	Arg	Pro		
		275					280					285					
tat	tat	att	tac	caa	tgt	gac	tta	tct	atg	gga	ctc	gaa	cac	ttc	cgc		912
Tyr	Tyr	Ile	Tyr	Gln	Cys	Asp	Leu	Ser	Met	Gly	Leu	Glu	His	Phe	Arg		
	290					295					300						
aca	cca	gtt	tct	aaa	ggt	ata	gaa	att	att	gaa	gga	tta	cgt	gga	cat		960
Thr	Pro	Val	Ser	Lys	Gly	Ile	Glu	Ile	Ile	Glu	Gly	Leu	Arg	Gly	His		
	305				310					315					320		
aca	tct	gga	tat	gca	gta	cca	aca	ttt	gtt	gtg	cat	gca	cct	ggt	ggt		1008
Thr	Ser	Gly	Tyr	Ala	Val	Pro	Thr	Phe	Val	Val	His	Ala	Pro	Gly	Gly		
				325					330					335			
gga	gga	aaa	act	cca	gta	atg	cct	caa	tat	gta	att	tct	caa	tct	cct		1056
Gly	Gly	Lys	Thr	Pro	Val	Met	Pro	Gln	Tyr	Val	Ile	Ser	Gln	Ser	Pro		
			340					345					350				
cat	cgt	gta	gtt	tta	cgc	aac	ttt	gaa	gga	ggt	ata	aca	act	tat	aca		1104
His	Arg	Val	Val	Leu	Arg	Asn	Phe	Glu	Gly	Val	Ile	Thr	Thr	Tyr	Thr		
		355				360						365					
gaa	cca	gaa	aat	tat	aca	cat	gaa	cct	tgt	tat	gat	gaa	gaa	aaa	ttt		1152
Glu	Pro	Glu	Asn	Tyr	Thr	His	Glu	Pro	Cys	Tyr	Asp	Glu	Glu	Lys	Phe		
	370					375					380						
gaa	aaa	atg	tat	gaa	ata	agt	gga	gtt	tat	atg	cta	gat	gaa	gga	tta		1200
Glu	Lys	Met	Tyr	Glu	Ile	Ser	Gly	Val	Tyr	Met	Leu	Asp	Glu	Gly	Leu		
	385				390					395					400		
gaa	atg	tca	cta	gaa	cct	agc	cac	tta	gca	cgt	cat	gaa	cgc	aat	aaa		1248
Glu	Met	Ser	Leu	Glu	Pro	Ser	His	Leu	Ala	Arg	His	Glu	Arg	Asn	Lys		
				405					410					415			
aag	aga	gca	gaa	gct	gaa	ggg	aaa	aaa	taa								1278
Lys	Arg	Ala	Glu	Ala	Glu	Gly	Lys	Lys									
			420					425									

<210> 21

<211> 425
 <212> PRT
 <213> Fusobacterium nucleatum

 <400> 21

 Met Asn Thr Val Asn Thr Arg Lys Lys Phe Phe Pro Asn Val Thr Asp
 1 5 10 15

 Glu Glu Trp Asn Asp Trp Thr Trp Gln Val Lys Asn Arg Leu Lys Ser
 20 25 30

 Val Glu Asp Leu Glu Lys Tyr Val Asp Leu Ser Glu Glu Glu Thr Glu
 35 40 45

 Gly Val Val Arg Thr Leu Glu Thr Leu Arg Met Ala Ile Thr Pro Phe
 50 55 60

 Tyr Phe Ser Leu Ile Asp Leu Asn Ser Asp Arg Cys Pro Ile Arg Lys
 65 70 75 80

 Gln Ala Ile Pro Thr Ile Arg Glu Ile His Gln Ser Asp Ala Asp Met
 85 90 95

 Leu Asp Pro Leu His Glu Asp Glu Asp Ser Pro Val Pro Gly Leu Thr
 100 105 110

 His Arg Tyr Pro Asp Arg Val Leu Leu Leu Ile Thr Asp Met Cys Ser
 115 120 125

 Val Tyr Cys Arg His Cys Thr Arg Arg Arg Phe Ala Gly Ser Ser Asp
 130 135 140

 Gly Ala Met Pro Met Asp Arg Ile Asp Lys Ala Ile Glu Tyr Ile Ala
 145 150 155 160

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

 Leu Val Ser Asn Lys Lys Leu Glu Ser Ile Ile Gln Lys Leu Arg Ala
 180 185 190

 Ile Pro His Val Glu Ile Ile Arg Ile Gly Ser Arg Thr Pro Val Val
 195 200 205

Leu Pro Gln Arg Ile Thr Pro Glu Leu Cys Asn Met Leu Lys Lys Tyr
 210 215 220

His Pro Ile Trp Met Asn Thr His Phe Asn His Pro Gln Glu Val Thr
 225 230 235 240

Pro Glu Ala Lys Lys Ala Cys Glu Met Leu Ala Asp Ala Gly Val Pro
 245 250 255

Leu Gly Asn Gln Thr Val Leu Leu Arg Gly Ile Asn Asp Ser Val Pro
 260 265 270

Val Met Lys Arg Leu Val His Asp Leu Val Met Met Arg Val Arg Pro
 275 280 285

Tyr Tyr Ile Tyr Gln Cys Asp Leu Ser Met Gly Leu Glu His Phe Arg
 290 295 300

Thr Pro Val Ser Lys Gly Ile Glu Ile Ile Glu Gly Leu Arg Gly His
 305 310 315 320

Thr Ser Gly Tyr Ala Val Pro Thr Phe Val Val His Ala Pro Gly Gly
 325 330 335

Gly Gly Lys Thr Pro Val Met Pro Gln Tyr Val Ile Ser Gln Ser Pro
 340 345 350

His Arg Val Val Leu Arg Asn Phe Glu Gly Val Ile Thr Thr Tyr Thr
 355 360 365

Glu Pro Glu Asn Tyr Thr His Glu Pro Cys Tyr Asp Glu Glu Lys Phe
 370 375 380

Glu Lys Met Tyr Glu Ile Ser Gly Val Tyr Met Leu Asp Glu Gly Leu
 385 390 395 400

Glu Met Ser Leu Glu Pro Ser His Leu Ala Arg His Glu Arg Asn Lys
 405 410 415

Lys Arg Ala Glu Ala Glu Gly Lys Lys
 420 425

<210> 22
 <211> 20
 <212> DNA
 <213> Artificial

 <220>
 <223> primer

 <220>
 <221> misc_feature
 <222> (1)..(20)
 <223> Y is t/u or c; w is a or t/u.

 <400> 22
 ytwagaatgg cwatwacwcc 20

 <210> 23
 <211> 20
 <212> DNA
 <213> Artificial

 <220>
 <223> primer

 <220>
 <221> misc_feature
 <222> (1)..(20)
 <223> R is g or a, W is a or t/u.

 <400> 23
 agaaarcarg cwatwccwac 20

 <210> 24
 <211> 20
 <212> DNA
 <213> Artificial

 <220>
 <223> primer

 <220>
 <221> misc_feature
 <222> (1)..(20)
 <223> Y is t/u or c; w is a or t/u.

 <400> 24
 ggwytwacwc ayagataycc 20

 <210> 25
 <211> 20
 <212> DNA

<213> Artificial
 <220>
 <223> primer

 <220>
 <221> misc_feature
 <222> (1)..(20)
 <223> Y is t/u or c; w is a or t/u.

 <400> 25
 tawgtwgtwa twacwccytc 20

 <210> 26
 <211> 20
 <212> DNA
 <213> Artificial

 <220>
 <223> primer

 <220>
 <221> misc_feature
 <222> (1)..(20)
 <223> r is g or a; w is a or t/u.

 <400> 26
 tcwacwacra awgtwggwac 20

 <210> 27
 <211> 20
 <212> DNA
 <213> Artificial

 <220>
 <223> primer

 <220>
 <221> misc_feature
 <222> (1)..(20)
 <223> r is g or a; w is a or t/u.

 <400> 27
 ccwccwccwg gwgertcwac 20

 <210> 28
 <211> 30
 <212> DNA
 <213> Artificial

 <220>

```

<223> primer

<400> 28
cctttcagtt ggaattgagc actttagaac          30

<210> 29
<211> 30
<212> DNA
<213> Artificial

<220>
<223> primer

<400> 29
gatactgcgt tcctacattt gttgtggatg          30

<210> 30
<211> 29
<212> DNA
<213> Artificial

<220>
<223> primer

<400> 30
cgctgctcta tgtagctcta aagaaagag          29

<210> 31
<211> 28
<212> DNA
<213> Artificial

<220>
<223> primer

<400> 31
cagottgctt tcttacaggg tcatttgg          28

<210> 32
<211> 1245
<212> DNA
<213> Clostridium sticklandii

<220>
<221> CDS
<222> (1)..(1245)

<400> 32
atg agt tta aag gat aag ttt ttt tca cat gta agc caa gaa gat tgg          48
Met Ser Leu Lys Asp Lys Phe Phe Ser His Val Ser Gln Glu Asp Trp
1           5           10           15

```

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

Lys Arg Ala Cys Glu Leu Leu Ala Asp Ala Gly Ile Pro Leu Gly Asn
 245 250 255
 caa agt gtg ctt ctt gca ggt gta aat gat tgc atg cac gtt atg aaa 816
 Gln Ser Val Leu Leu Ala Gly Val Asn Asp Cys Met His Val Met Lys
 260 265 270
 .
 aaa cta gta aat gat tta gtt aaa ata aga gta aga cct tac tat att 864
 Lys Leu Val Asn Asp Leu Val Lys Ile Arg Val Arg Pro Tyr Tyr Ile
 275 280 285
 .
 tat caa tgt gac ctt tca gtt gga att gag cac ttt aga act cca gtt 912
 Tyr Gln Cys Asp Leu Ser Val Gly Ile Glu His Phe Arg Thr Pro Val
 290 295 300
 .
 gca aag gga ata gaa ata att gaa ggc tta aga gga cat act tca gga 960
 Ala Lys Gly Ile Glu Ile Ile Glu Gly Leu Arg Gly His Thr Ser Gly
 305 310 315 320
 .
 tac tgc gtt cct aca ttt gtt gtg gat gca cct ggt ggt gga gga aaa 1008
 Tyr Cys Val Pro Thr Phe Val Val Asp Ala Pro Gly Gly Gly Gly Lys
 325 330 335
 .
 act cca gtt atg cca aac tat gtt att tca caa aat cac aat aaa gtt 1056
 Thr Pro Val Met Pro Asn Tyr Val Ile Ser Gln Asn His Asn Lys Val
 340 345 350
 .
 att tta cgt aac ttt gaa ggt gta att aca act tac gat gag cct gat 1104
 Ile Leu Arg Asn Phe Glu Gly Val Ile Thr Thr Tyr Asp Glu Pro Asp
 355 360 365
 .
 cat tat act ttc cac tgt gac tgt gat gta tgc act gga aaa aca aat 1152
 His Tyr Thr Phe His Cys Asp Cys Asp Val Cys Thr Gly Lys Thr Asn
 370 375 380
 .
 gtt cat aag gtt gga gta gct gga ctt ctt aat gga gag aca gcg aca 1200
 Val His Lys Val Gly Val Ala Gly Leu Leu Asn Gly Glu Thr Ala Thr
 385 390 395 400
 .
 ctt gaa cca gag ggt ttg gaa aga aaa caa aga gga cat cac taa 1245
 Leu Glu Pro Glu Gly Leu Glu Arg Lys Gln Arg Gly His His
 405 410

<210> 33
 <211> 414
 <212> PRT
 <213> Clostridium sticklandii

<400> 33

Met Ser Leu Lys Asp Lys Phe Phe Ser His Val Ser Gln Glu Asp Trp
 1 5 10 15

Asn Asp Trp Lys Trp Gln Val Arg Asn Arg Ile Glu Thr Val Glu Glu
 20 25 30

Leu Lys Lys Tyr Ile Pro Leu Thr Pro Glu Glu Glu Glu Gly Val Lys
 35 40 45

Arg Cys Leu Asp Thr Leu Arg Met Ala Ile Thr Pro Tyr Tyr Leu Ser
 50 55 60

Leu Ile Asp Val Glu Asn Pro Asn Asp Pro Val Arg Lys Gln Ala Val
 65 70 75 80

Pro Leu Ser Leu Glu Leu His Arg Ala Ala Ser Asp Gln Glu Asp Pro
 85 90 95

Leu His Glu Asp Gly Asp Ser Pro Val Pro Gly Leu Thr His Arg Tyr
 100 105 110

Pro Asp Arg Val Leu Leu Leu Met Thr Asp Gln Cys Ser Met Tyr Cys
 115 120 125

Arg His Cys Thr Arg Arg Arg Phe Ala Gly Gln Thr Asp Ser Ala Val
 130 135 140

Asp Thr Lys Gln Ile Asp Ala Ala Ile Glu Tyr Ile Lys Asn Thr Pro
 145 150 155 160

Gln Val Arg Asp Val Leu Leu Ser Gly Gly Asp Ala Leu Leu Ile Ser
 165 170 175

Asp Glu Lys Leu Glu Tyr Thr Ile Lys Arg Leu Arg Glu Ile Pro His
 180 185 190

Val Glu Val Ile Arg Ile Gly Ser Arg Val Pro Val Val Met Pro Gln
 195 200 205

Arg Ile Thr Pro Glu Leu Val Ser Met Leu Lys Lys Tyr His Pro Val
 210 215 220

Trp Leu Asn Thr His Phe Asn His Pro Asn Glu Ile Thr Glu Glu Ser
 225 230 235 240

Lys Arg Ala Cys Glu Leu Leu Ala Asp Ala Gly Ile Pro Leu Gly Asn
 245 250 255

Gln Ser Val Leu Leu Ala Gly Val Asn Asp Cys Met His Val Met Lys
 260 265 270

Lys Leu Val Asn Asp Leu Val Lys Ile Arg Val Arg Pro Tyr Tyr Ile
 275 280 285

Tyr Gln Cys Asp Leu Ser Val Gly Ile Glu His Phe Arg Thr Pro Val
 290 295 300

Ala Lys Gly Ile Glu Ile Ile Glu Gly Leu Arg Gly His Thr Ser Gly
 305 310 315 320

Tyr Cys Val Pro Thr Phe Val Val Asp Ala Pro Gly Gly Gly Gly Lys
 325 330 335

Thr Pro Val Met Pro Asn Tyr Val Ile Ser Gln Asn His Asn Lys Val
 340 345 350

Ile Leu Arg Asn Phe Glu Gly Val Ile Thr Thr Tyr Asp Glu Pro Asp
 355 360 365

His Tyr Thr Phe His Cys Asp Cys Asp Val Cys Thr Gly Lys Thr Asn
 370 375 380

Val His Lys Val Gly Val Ala Gly Leu Leu Asn Gly Glu Thr Ala Thr
 385 390 395 400

Leu Glu Pro Glu Gly Leu Glu Arg Lys Gln Arg Gly His His
 405 410

<210> 34
 <211> 1245
 <212> DNA
 <213> Clostridium sticklandii

<220>
 <221> CDS
 <222> (1)..(1245)

<400> 34
 atg agt tta aag gat aag ttt ttt tca cat gta agc caa gaa gat tgg 48
 Met Ser Leu Lys Asp Lys Phe Phe Ser His Val Ser Gln Glu Asp Trp
 1 5 10 15
 aat gat tgg aaa tgg caa gta aga aat cgt ata gaa act gtt gaa gaa 96

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

245				250				255					
caa agt gtg ctt ctt gca ggt gta aat gat tgc atg cac gtt atg aaa													816
Gln Ser Val Leu Leu Ala Gly Val Asn Asp Cys Met His Val Met Lys													
		260					265					270	
aaa cta gta aat gat tta gtt aaa ata cgc gta cgt cct tac tat att													864
Lys Leu Val Asn Asp Leu Val Lys Ile Arg Val Arg Pro Tyr Tyr Ile													
		275					280					285	
tat caa tgt gac ctt tca gtt gga att gag cac ttt cgc act cca gtt													912
Tyr Gln Cys Asp Leu Ser Val Gly Ile Glu His Phe Arg Thr Pro Val													
		290					295					300	
gca aag gga ata gaa ata att gaa ggc tta aga gga cat act tca gga													960
Ala Lys Gly Ile Glu Ile Ile Glu Gly Leu Arg Gly His Thr Ser Gly													
		305					310					315	320
tac tgc gtt cct aca ttt gtt gtg gat gca cct ggt ggt gga gga aaa													1008
Tyr Cys Val Pro Thr Phe Val Val Asp Ala Pro Gly Gly Gly Gly Lys													
							325						335
act cca gtt atg cca aac tat gtt att tca caa aat cac aat aaa gtt													1056
Thr Pro Val Met Pro Asn Tyr Val Ile Ser Gln Asn His Asn Lys Val													
		340					345						350
att tta cgt aac ttt gaa ggt gta att aca act tac gat gag cct gat													1104
Ile Leu Arg Asn Phe Glu Gly Val Ile Thr Thr Tyr Asp Glu Pro Asp													
		355					360						365
cat tat act ttc cac tgt gac tgt gat gta tgc act gga aaa aca aat													1152
His Tyr Thr Phe His Cys Asp Cys Asp Val Cys Thr Gly Lys Thr Asn													
		370					375						380
gtt cat aag gtt gga gta gct gga ctt cta aat gga gag aca gcg aca													1200
Val His Lys Val Gly Val Ala Gly Leu Leu Asn Gly Glu Thr Ala Thr													
		385					390						400
ctt gaa cca gag ggt ttg gaa aga aaa caa aga gga cat cac taa													1245
Leu Glu Pro Glu Gly Leu Glu Arg Lys Gln Arg Gly His His													
		405					410						
<210> 35													
<211> 414													
<212> PRT													
<213> Clostridium sticklandii													
<400> 35													
Met Ser Leu Lys Asp Lys Phe Phe Ser His Val Ser Gln Glu Asp Trp													
1				5					10				15
Asn Asp Trp Lys Trp Gln Val Arg Asn Arg Ile Glu Thr Val Glu Glu													
				20					25				30

Leu Lys Lys Tyr Ile Pro Leu Thr Pro Glu Glu Glu Glu Gly Val Lys
 35 40 45

Arg Cys Leu Asp Thr Leu Arg Met Ala Ile Thr Pro Tyr Tyr Leu Ser
 50 55 60

Leu Ile Asp Val Glu Asn Pro Asn Asp Pro Val Arg Lys Gln Ala Val
 65 70 75 80

Pro Leu Ser Leu Glu Leu His Arg Ala Ala Ser Asp Gln Glu Asp Pro
 85 90 95

Leu His Glu Asp Gly Asp Ser Pro Val Pro Gly Leu Thr His Arg Tyr
 100 105 110

Pro Asp Arg Val Leu Leu Leu Met Thr Asp Gln Cys Ser Met Tyr Cys
 115 120 125

Arg Tyr Cys Thr Arg Arg Arg Phe Ala Gly Gln Thr Asp Ser Ala Val
 130 135 140

Asp Thr Lys Gln Ile Asp Ala Ala Ile Glu Tyr Ile Lys Asn Thr Pro
 145 150 155 160

Gln Val Arg Asp Val Leu Leu Ser Gly Gly Asp Ala Leu Leu Ile Ser
 165 170 175

Asp Glu Lys Leu Glu Tyr Thr Ile Lys Arg Leu Arg Glu Ile Pro His
 180 185 190

Val Glu Val Ile Arg Ile Gly Ser Arg Val Pro Val Val Met Pro Gln
 195 200 205

Arg Ile Thr Pro Glu Leu Val Ser Met Leu Lys Lys Tyr His Pro Val
 210 215 220

Trp Leu Asn Thr His Phe Asn His Pro Asn Glu Ile Thr Glu Glu Ser
 225 230 235 240

Lys Arg Ala Cys Glu Leu Leu Ala Asp Ala Gly Ile Pro Leu Gly Asn
 245 250 255

Gln Ser Val Leu Leu Ala Gly Val Asn Asp Cys Met His Val Met Lys
 260 265 270
 Lys Leu Val Asn Asp Leu Val Lys Ile Arg Val Arg Pro Tyr Tyr Ile
 275 280 285
 Tyr Gln Cys Asp Leu Ser Val Gly Ile Glu His Phe Arg Thr Pro Val
 290 295 300
 Ala Lys Gly Ile Glu Ile Ile Glu Gly Leu Arg Gly His Thr Ser Gly
 305 310 315 320
 Tyr Cys Val Pro Thr Phe Val Val Asp Ala Pro Gly Gly Gly Gly Lys
 325 330 335
 Thr Pro Val Met Pro Asn Tyr Val Ile Ser Gln Asn His Asn Lys Val
 340 345 350
 Ile Leu Arg Asn Phe Glu Gly Val Ile Thr Thr Tyr Asp Glu Pro Asp
 355 360 365
 His Tyr Thr Phe His Cys Asp Cys Asp Val Cys Thr Gly Lys Thr Asn
 370 375 380
 Val His Lys Val Gly Val Ala Gly Leu Leu Asn Gly Glu Thr Ala Thr
 385 390 395 400
 Leu Glu Pro Glu Gly Leu Glu Arg Lys Gln Arg Gly His His
 405 410

<210> 36

<211> 45

<212> DNA

<213> Artificial

<220>

<223> mutagenesis primer

<400> 36

gaaatggcaa gtaagaaatc gtataaagac tgttgaagaa cttaa

45

<210> 37

<211> 35

<212> DNA

<213> Artificial

<220>
 <223> mutagenesis primer

 <400> 37
 tcgcgcgagcg tctgatatgg aagacccact tcatg 35

<210> 38
 <211> 40
 <212> DNA
 <213> Artificial

<220>
 <223> mutagenesis primer

 <400> 38
 gactgatcaa tgttcagtat actgccgccca ctgtactcgt 40

<210> 39
 <211> 34
 <212> DNA
 <213> Artificial

<220>
 <223> mutagenesis primer

 <400> 39
 gttcctacat ttgttgtagca tgcacctggt ggtg 34

<210> 40
 <211> 1245
 <212> DNA
 <213> Clostridium sticklandii

<220>
 <221> CDS
 <222> (1)..(1245)

<400> 40
 atg agt tta aag gat aag ttt ttt tca cat gta agc caa gaa gat tgg 48
 Met Ser Leu Lys Asp Lys Phe Phe Ser His Val Ser Gln Glu Asp Trp
 1 5 10 15

aat gat tgg aaa tgg caa gta aga aat cgt ata aag act gtt gaa gaa 96
 Asn Asp Trp Lys Trp Gln Val Arg Asn Arg Ile Lys Thr Val Glu Glu
 20 25 30

ctt aaa aaa tat att cca ctt act cca gaa gaa gaa gaa ggg gta aaa 144
 Leu Lys Lys Tyr Ile Pro Leu Thr Pro Glu Glu Glu Glu Gly Val Lys
 35 40 45

cgc tgt ctt gat aca tta cgt atg gct att act cca tac tat cta tog 192
 Arg Cys Leu Asp Thr Leu Arg Met Ala Ile Thr Pro Tyr Tyr Leu Ser
 50 55 60

tat caa tgt gac ctt tca gtt gga att gag cac ttt cgc act cca gtt	912
Tyr Gln Cys Asp Leu Ser Val Gly Ile Glu His Phe Arg Thr Pro Val	
290 295 300	
gca aag gga ata gaa ata att gaa ggc tta aga gga cat act tca gga	960
Ala Lys Gly Ile Glu Ile Ile Glu Gly Leu Arg Gly His Thr Ser Gly	
305 310 315 320	
tac tgc gtt cct aca ttt gtt gtg cat gca cct ggt ggt gga gga aaa	1008
Tyr Cys Val Pro Thr Phe Val Val His Ala Pro Gly Gly Gly Lys	
325 330 335	
act cca gtt atg cca aac tat gtt att tca caa aat cac aat aaa gtt	1056
Thr Pro Val Met Pro Asn Tyr Val Ile Ser Gln Asn His Asn Lys Val	
340 345 350	
att tta cgt aac ttt gaa ggt gta att aca act tac gat gag cct gat	1104
Ile Leu Arg Asn Phe Glu Gly Val Ile Thr Thr Tyr Asp Glu Pro Asp	
355 360 365	
cat tat act ttc cac tgt gac tgt gat gta tgc act gga aaa aca aat	1152
His Tyr Thr Phe His Cys Asp Cys Asp Val Cys Thr Gly Lys Thr Asn	
370 375 380	
gtt cat aag gtt gga gta gct gga ctt cta aat gga gag aca gcg aca	1200
Val His Lys Val Gly Val Ala Gly Leu Leu Asn Gly Glu Thr Ala Thr	
385 390 395 400	
ctt gaa cca gag ggt ttg gaa aga aaa caa aga gga cat cac taa	1245
Leu Glu Pro Glu Gly Leu Glu Arg Lys Gln Arg Gly His His	
405 410	
<210> 41	
<211> 414	
<212> PRT	
<213> Clostridium sticklandii	
<400> 41	
Met Ser Leu Lys Asp Lys Phe Phe Ser His Val Ser Gln Glu Asp Trp	
1 5 10 15	
Asn Asp Trp Lys Trp Gln Val Arg Asn Arg Ile Lys Thr Val Glu Glu	
20 25 30	
Leu Lys Lys Tyr Ile Pro Leu Thr Pro Glu Glu Glu Gly Val Lys	
35 40 45	
Arg Cys Leu Asp Thr Leu Arg Met Ala Ile Thr Pro Tyr Tyr Leu Ser	
50 55 60	
Leu Ile Asp Val Glu Asn Pro Asn Asp Pro Val Arg Lys Gln Ala Val	

Ala Lys Gly Ile Glu Ile Ile Glu Gly Leu Arg Gly His Thr Ser Gly
305 310 315 320

Tyr Cys Val Pro Thr Phe Val Val His Ala Pro Gly Gly Gly Gly Lys
325 330 335

Thr Pro Val Met Pro Asn Tyr Val Ile Ser Gln Asn His Asn Lys Val
340 345 350

Ile Leu Arg Asn Phe Glu Gly Val Ile Thr Thr Tyr Asp Glu Pro Asp
355 360 365

His Tyr Thr Phe His Cys Asp Cys Asp Val Cys Thr Gly Lys Thr Asn
370 375 380

Val His Lys Val Gly Val Ala Gly Leu Leu Asn Gly Glu Thr Ala Thr
385 390 395 400

Leu Glu Pro Glu Gly Leu Glu Arg Lys Gln Arg Gly His His
405 410

<210> 42
<211> 1245
<212> DNA
<213> Clostridium sticklandii

<220>
<221> CDS
<222> (1)..(1245)

<400> 42
atg agt tta aag gat aag ttt ttt aca cat gta agc caa gaa gat tgg 48
Met Ser Leu Lys Asp Lys Phe Phe Thr His Val Ser Gln Glu Asp Trp
1 5 10 15
aat gat tgg aaa tgg caa gta aga aat cgt ata aag act gtt gaa gaa 96
Asn Asp Trp Lys Trp Gln Val Arg Asn Arg Ile Lys Thr Val Glu Glu
20 25 30
ctt aaa aaa tat att cca ctt act cca gaa gaa gaa gaa ggg gta aaa 144
Leu Lys Lys Tyr Ile Pro Leu Thr Pro Glu Glu Glu Glu Gly Val Lys
35 40 45
cgc tgt ctt gat aca tta cgt atg gct att act cca tac tat cta tcg 192
Arg Cys Leu Asp Thr Leu Arg Met Ala Ile Thr Pro Tyr Tyr Leu Ser
50 55 60

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

Tyr Gln Cys Asp Leu Ser Val Gly Ile Glu His Phe Arg Thr Pro Val
 290 295 300

gca aag gga ata gaa ata att gaa ggc tta aga gga cat act tca gga 960
 Ala Lys Gly Ile Glu Ile Ile Glu Gly Leu Arg Gly His Thr Ser Gly
 305 310 315 320

tac tgc gtt cct aca ttt gtt gtg cat gca cct ggt ggt gga gga aaa 1008
 Tyr Cys Val Pro Thr Phe Val Val His Ala Pro Gly Gly Gly Gly Lys
 325 330 335

act cca gtt atg cca aac tat gtt att tca caa aat cac aat aaa gtt 1056
 Thr Pro Val Met Pro Asn Tyr Val Ile Ser Gln Asn His Asn Lys Val
 340 345 350

att tta cgt aac ttt gaa ggt gta att aca act tac gat gag cct gat 1104
 Ile Leu Arg Asn Phe Glu Gly Val Ile Thr Thr Tyr Asp Glu Pro Asp
 355 360 365

cat tat act ttc cac tgt gac tgt gat gta tgc act gga aaa aca aat 1152
 His Tyr Thr Phe His Cys Asp Cys Asp Val Cys Thr Gly Lys Thr Asn
 370 375 380

gtt cat aag gtt gga gta gct gga ctt cta aat gga gag aca gcg aca 1200
 Val His Lys Val Gly Val Ala Gly Leu Leu Asn Gly Glu Thr Ala Thr
 385 390 395 400

ctt gaa cct gag ggt ttg gaa aga aaa caa aga gga cat cac taa 1245
 Leu Glu Pro Glu Gly Leu Glu Arg Lys Gln Arg Gly His His
 405 410

<210> 43
 <211> 414
 <212> PRT
 <213> Clostridium sticklandii

<400> 43

Met Ser Leu Lys Asp Lys Phe Phe Thr His Val Ser Gln Glu Asp Trp
 1 5 10 15

Asn Asp Trp Lys Trp Gln Val Arg Asn Arg Ile Lys Thr Val Glu Glu
 20 25 30

Leu Lys Lys Tyr Ile Pro Leu Thr Pro Glu Glu Glu Glu Gly Val Lys
 35 40 45

Arg Cys Leu Asp Thr Leu Arg Met Ala Ile Thr Pro Tyr Tyr Leu Ser
 50 55 60

Leu Ile Asp Val Glu Asn Pro Asn Asp Pro Val Arg Lys Gln Ala Val
 65 70 75 80

Ala Lys Gly Ile Glu Ile Ile Glu Gly Leu Arg Gly His Thr Ser Gly
305 310 315 320

Tyr Cys Val Pro Thr Phe Val Val His Ala Pro Gly Gly Gly Gly Lys
325 330 335

Thr Pro Val Met Pro Asn Tyr Val Ile Ser Gln Asn His Asn Lys Val
340 345 350

Ile Leu Arg Asn Phe Glu Gly Val Ile Thr Thr Tyr Asp Glu Pro Asp
355 360 365

His Tyr Thr Phe His Cys Asp Cys Asp Val Cys Thr Gly Lys Thr Asn
370 375 380

Val His Lys Val Gly Val Ala Gly Leu Leu Asn Gly Glu Thr Ala Thr
385 390 395 400

Leu Glu Pro Glu Gly Leu Glu Arg Lys Gln Arg Gly His His
405 410

<210> 44
<211> 1245
<212> DNA
<213> Clostridium sticklandii

<220>
<221> CDS
<222> (1)..(1245)

<400> 44
atg agt tta aag gat aag ttt ttt tca cat gta agc caa gaa gat tgg 48
Met Ser Leu Lys Asp Lys Phe Phe Ser His Val Ser Gln Glu Asp Trp
1 5 10 15
aat gat tgg aaa tgg caa gta aga aat cgt ata aag act gct gaa gaa 96
Asn Asp Trp Lys Trp Gln Val Arg Asn Arg Ile Lys Thr Ala Glu Glu
20 25 30
ctt aaa aaa tat att cca ctt act cca gaa gaa gaa gaa ggg gta aaa 144
Leu Lys Lys Tyr Ile Pro Leu Thr Pro Glu Glu Glu Glu Gly Val Lys
35 40 45
cgc tgt cat gat aca tta cgt atg gct att act cca tac tat cta tcg 192
Arg Cys His Asp Thr Leu Arg Met Ala Ile Thr Pro Tyr Tyr Leu Ser
50 55 60
cta att gat gta gga aat cca aat gac cct gta aga aag caa gct gta 240

Leu 65	Ile	Asp	Val	Gly	Asn 70	Pro	Asn	Asp	Pro	Val 75	Arg	Lys	Gln	Ala	Val 80	
cct	ctt	tct	tta	gag	ctg	cat	cgc	gca	gcg	tct	gat	atg	gaa	gac	cca	288
Pro	Leu	Ser	Leu	Glu	Leu	His	Arg	Ala	Ala	Ser	Asp	Met	Glu	Asp	Pro	
				85					90					95		
ctt	cat	gaa	gat	gga	gat	tct	cca	gtt	cca	gga	ctt	aca	cat	cgc	tat	336
Leu	His	Glu	Asp	Gly	Asp	Ser	Pro	Val	Pro	Gly	Leu	Thr	His	Arg	Tyr	
				100				105						110		
cct	gat	cgt	gtt	ctt	ctt	tta	atg	act	gat	cta	tgt	tca	gta	tac	tgc	384
Pro	Asp	Arg	Val	Leu	Leu	Leu	Met	Thr	Asp	Leu	Cys	Ser	Val	Tyr	Cys	
				115			120							125		
cgc	cac	tgt	act	cgt	aga	cgc	ttc	gct	ggc	caa	aca	gat	tct	gct	gtt	432
Arg	His	Cys	Thr	Arg	Arg	Arg	Phe	Ala	Gly	Gln	Thr	Asp	Ser	Ala	Val	
				130			135					140				
gat	acg	aag	caa	ata	gat	gct	gcg	att	gaa	tat	atc	aaa	aat	act	cca	480
Asp	Thr	Lys	Gln	Ile	Asp	Ala	Ala	Ile	Glu	Tyr	Ile	Lys	Asn	Thr	Pro	
					150					155					160	
caa	gta	aga	gac	gtt	cta	ctt	tca	gga	gga	gat	gca	cta	tta	atc	tca	528
Gln	Val	Arg	Asp	Val	Leu	Leu	Ser	Gly	Gly	Asp	Ala	Leu	Leu	Ile	Ser	
				165					170					175		
gat	gaa	aag	ctt	gag	tac	aca	atc	aaa	aga	ctt	cgt	gaa	ata	cca	cac	576
Asp	Glu	Lys	Leu	Glu	Tyr	Thr	Ile	Lys	Arg	Leu	Arg	Glu	Ile	Pro	His	
				180				185					190			
gtt	gag	gtt	att	cgt	att	gga	tca	cgt	gta	cca	gtt	gta	atg	cca	caa	624
Val	Glu	Val	Ile	Arg	Ile	Gly	Ser	Arg	Val	Pro	Val	Val	Met	Pro	Gln	
				195			200						205			
cgt	att	aca	cca	gaa	cta	gtt	tct	atg	ctt	aaa	aag	tat	cat	cca	gta	672
Arg	Ile	Thr	Pro	Glu	Leu	Val	Ser	Met	Leu	Lys	Lys	Tyr	His	Pro	Val	
				210			215					220				
tgg	tta	aat	aca	cac	ttc	aac	cat	cct	aat	gaa	att	act	gaa	gag	tct	720
Trp	Leu	Asn	Thr	His	Phe	Asn	His	Pro	Asn	Glu	Ile	Thr	Glu	Glu	Ser	
					230					235				240		
aaa	cgt	gca	tgt	gag	tta	ctt	gct	gat	gca	ggc	att	cct	ctt	gga	aat	768
Lys	Arg	Ala	Cys	Glu	Leu	Leu	Ala	Asp	Ala	Gly	Ile	Pro	Leu	Gly	Asn	
				245					250					255		
caa	agt	gtg	ctt	cta	gca	ggc	gta	aat	gat	tgc	atg	cac	gtt	atg	aaa	816
Gln	Ser	Val	Leu	Leu	Ala	Gly	Val	Asn	Asp	Cys	Met	His	Val	Met	Lys	
				260				265						270		
aaa	cta	gta	aat	gat	tta	gtt	aaa	ata	cgc	gta	cgt	cct	tac	tat	att	864
Lys	Leu	Val	Asn	Asp	Leu	Val	Lys	Ile	Arg	Val	Arg	Pro	Tyr	Tyr	Ile	
				275			280					285				
tat	caa	tgt	gac	ctt	tca	gtt	gga	att	gag	cac	ttt	cgc	act	cca	gtt	912
Tyr	Gln	Cys	Asp	Leu	Ser	Val	Gly	Ile	Glu	His	Phe	Arg	Thr	Pro	Val	

290	295	300	
gca aag gga ata gaa ata att gaa ggc tta aga gga cat act tca gga			960
Ala Lys Gly Ile Glu Ile Ile Glu Gly Leu Arg Gly His Thr Ser Gly			
305	310	315	320
tac tgc gtt cct aca ttt gtt gtg cat gca cct ggt ggt gga gga aaa			1008
Tyr Cys Val Pro Thr Phe Val Val His Ala Pro Gly Gly Gly Gly Lys			
	325	330	335
act cca gtt atg cca aac tat gtt att tca caa aat cac aat aaa gtt			1056
Thr Pro Val Met Pro Asn Tyr Val Ile Ser Gln Asn His Asn Lys Val			
	340	345	350
att tta cgt aac ttt gaa ggt gta att aca act tac gat gag cct gat			1104
Ile Leu Arg Asn Phe Glu Gly Val Ile Thr Thr Tyr Asp Glu Pro Asp			
	355	360	365
cat tat act ttc cac tgt gac tgt gat gta tgc act gga aaa aca aat			1152
His Tyr Thr Phe His Cys Asp Cys Asp Val Cys Thr Gly Lys Thr Asn			
	370	375	380
gtt cat aag gtt gga gta gct gga ctt cta aat gga gag aca gcg aca			1200
Val His Lys Val Gly Val Ala Gly Leu Leu Asn Gly Glu Thr Ala Thr			
385	390	395	400
ctt gaa cca gag ggt ttg gaa aga aaa caa aga gga cat cac taa			1245
Leu Glu Pro Glu Gly Leu Glu Arg Lys Gln Arg Gly His His			
	405	410	
<210> 45			
<211> 414			
<212> PRT			
<213> Clostridium sticklandii			
<400> 45			
Met Ser Leu Lys Asp Lys Phe Phe Ser His Val Ser Gln Glu Asp Trp			
1	5	10	15
Asn Asp Trp Lys Trp Gln Val Arg Asn Arg Ile Lys Thr Ala Glu Glu			
	20	25	30
Leu Lys Lys Tyr Ile Pro Leu Thr Pro Glu Glu Glu Glu Gly Val Lys			
	35	40	45
Arg Cys His Asp Thr Leu Arg Met Ala Ile Thr Pro Tyr Tyr Leu Ser			
	50	55	60
Leu Ile Asp Val Gly Asn Pro Asn Asp Pro Val Arg Lys Gln Ala Val			
65	70	75	80

Pro Leu Ser Leu Glu Leu His Arg Ala Ala Ser Asp Met Glu Asp Pro
 85 90 95

Leu His Glu Asp Gly Asp Ser Pro Val Pro Gly Leu Thr His Arg Tyr
 100 105 110

Pro Asp Arg Val Leu Leu Leu Met Thr Asp Leu Cys Ser Val Tyr Cys
 115 120 125

Arg His Cys Thr Arg Arg Arg Phe Ala Gly Gln Thr Asp Ser Ala Val
 130 135 140

Asp Thr Lys Gln Ile Asp Ala Ala Ile Glu Tyr Ile Lys Asn Thr Pro
 145 150 155 160

Gln Val Arg Asp Val Leu Leu Ser Gly Gly Asp Ala Leu Leu Ile Ser
 165 170 175

Asp Glu Lys Leu Glu Tyr Thr Ile Lys Arg Leu Arg Glu Ile Pro His
 180 185 190

Val Glu Val Ile Arg Ile Gly Ser Arg Val Pro Val Val Met Pro Gln
 195 200 205

Arg Ile Thr Pro Glu Leu Val Ser Met Leu Lys Lys Tyr His Pro Val
 210 215 220

Trp Leu Asn Thr His Phe Asn His Pro Asn Glu Ile Thr Glu Glu Ser
 225 230 235 240

Lys Arg Ala Cys Glu Leu Leu Ala Asp Ala Gly Ile Pro Leu Gly Asn
 245 250 255

Gln Ser Val Leu Leu Ala Gly Val Asn Asp Cys Met His Val Met Lys
 260 265 270

Lys Leu Val Asn Asp Leu Val Lys Ile Arg Val Arg Pro Tyr Tyr Ile
 275 280 285

Tyr Gln Cys Asp Leu Ser Val Gly Ile Glu His Phe Arg Thr Pro Val
 290 295 300

Ala Lys Gly Ile Glu Ile Ile Glu Gly Leu Arg Gly His Thr Ser Gly
305 310 315 320

Tyr Cys Val Pro Thr Phe Val Val His Ala Pro Gly Gly Gly Gly Lys
325 330 335

Thr Pro Val Met Pro Asn Tyr Val Ile Ser Gln Asn His Asn Lys Val
340 345 350

Ile Leu Arg Asn Phe Glu Gly Val Ile Thr Thr Tyr Asp Glu Pro Asp
355 360 365

His Tyr Thr Phe His Cys Asp Cys Asp Val Cys Thr Gly Lys Thr Asn
370 375 380

Val His Lys Val Gly Val Ala Gly Leu Leu Asn Gly Glu Thr Ala Thr
385 390 395 400

Leu Glu Pro Glu Gly Leu Glu Arg Lys Gln Arg Gly His His
405 410

<210> 46
<211> 1245
<212> DNA
<213> Clostridium sticklandii

<220>
<221> CDS
<222> (1)..(1245)

<400> 46
atg agt tta aag gat aag ttt ttt tca cat gta agc caa gaa gat tgg 48
Met Ser Leu Lys Asp Lys Phe Phe Ser His Val Ser Gln Glu Asp Trp
1 5 10 15

aat gat tgg aaa tgg caa gta aga aat cgt ata aag act gtt gaa gaa 96
Asn Asp Trp Lys Trp Gln Val Arg Asn Arg Ile Lys Thr Val Glu Glu
20 25 30

ctt aaa aaa tat att cca ctt act cca gaa gaa gaa gaa ggg gta aaa 144
Leu Lys Lys Tyr Ile Pro Leu Thr Pro Glu Glu Glu Glu Gly Val Lys
35 40 45

cgc cgt ctt gat aca tta cgt atg gct att act cca tac tat cta tcg 192
Arg Arg Leu Asp Thr Leu Arg Met Ala Ile Thr Pro Tyr Tyr Leu Ser
50 55 60

cta att gat gta gaa aat cca aat gac cct gta aga aag caa gct gta 240
Leu Ile Asp Val Glu Asn Pro Asn Asp Pro Val Arg Lys Gln Ala Val

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

gca aag gga ata gaa ata att gaa ggc tta aga gga cat act tca gga 960
 Ala Lys Gly Ile Glu Ile Ile Glu Gly Leu Arg Gly His Thr Ser Gly
 305 310 315 320

tac tgc gtt cct aca ttt gtt gtg cat gca cct ggt ggt gga gga aaa 1008
 Tyr Cys Val Pro Thr Phe Val Val His Ala Pro Gly Gly Gly Gly Lys
 325 330 335

act cca gtt atg cca aac tat gtt att tca caa aat cac aat aaa gtt 1056
 Thr Pro Val Met Pro Asn Tyr Val Ile Ser Gln Asn His Asn Lys Val
 340 345 350

att tta cgt aac ttt gaa ggt gta att aca act tac gat gag cct gat 1104
 Ile Leu Arg Asn Phe Glu Gly Val Ile Thr Thr Tyr Asp Glu Pro Asp
 355 360 365

cat tat act ttc cac tgt gac tgt gat gta tgc act gga aaa aca aat 1152
 His Tyr Thr Phe His Cys Asp Cys Asp Val Cys Thr Gly Lys Thr Asn
 370 375 380

gtt cat aag gtt gga gta gct gga ctt cta aat gga gag aca gcg aca 1200
 Val His Lys Val Gly Val Ala Gly Leu Leu Asn Gly Glu Thr Ala Thr
 385 390 395 400

ctt gaa cca gag ggt ttg gaa aga aaa caa aga gga cat cac taa 1245
 Leu Glu Pro Glu Gly Leu Glu Arg Lys Gln Arg Gly His His
 405 410

<210> 47

<211> 414

<212> PRT

<213> Clostridium sticklandii

<400> 47

Met Ser Leu Lys Asp Lys Phe Phe Ser His Val Ser Gln Glu Asp Trp
 1 5 10 15

Asn Asp Trp Lys Trp Gln Val Arg Asn Arg Ile Lys Thr Val Glu Glu
 20 25 30

Leu Lys Lys Tyr Ile Pro Leu Thr Pro Glu Glu Glu Glu Gly Val Lys
 35 40 45

Arg Arg Leu Asp Thr Leu Arg Met Ala Ile Thr Pro Tyr Tyr Leu Ser
 50 55 60

Leu Ile Asp Val Glu Asn Pro Asn Asp Pro Val Arg Lys Gln Ala Val
 65 70 75 80

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

```

ccg gta tct aaa gga atc gaa att atc gaa aat ttg cgc gga cac acc      960
Pro Val Ser Lys Gly Ile Glu Ile Ile Glu Asn Leu Arg Gly His Thr
305                      310                      315                      320

tcg ggc tat gca gtt cct acc ttt gtg gta ggt gct ccg ggg ggt ggt      1008
Ser Gly Tyr Ala Val Pro Thr Phe Val Val Gly Ala Pro Gly Gly Gly
                      325                      330                      335

ggt aag ata cct gta acg ccg aac tat gtt gta tct cag tcc cca cga      1056
Gly Lys Ile Pro Val Thr Pro Asn Tyr Val Val Ser Gln Ser Pro Arg
                      340                      345                      350

cat gtg gtt ctt cgc aat tat gaa ggt gtt atc aca acc tat acg gag      1104
His Val Val Leu Arg Asn Tyr Glu Gly Val Ile Thr Thr Tyr Thr Glu
                      355                      360                      365

ccg gag aat tat cat gag gag tgc gat tgt gag gac tgt cga gcc ggt      1152
Pro Glu Asn Tyr His Glu Glu Cys Asp Cys Glu Asp Cys Arg Ala Gly
                      370                      375                      380

aag cat aaa gag ggt gta gct gca ctt tcc gga ggt cag cag ttg gct      1200
Lys His Lys Glu Gly Val Ala Ala Leu Ser Gly Gly Gln Gln Leu Ala
385                      390                      395                      400

atc gag cct tcc gac tta gct cgc aaa aaa cgc aag ttt gat aag aac      1248
Ile Glu Pro Ser Asp Leu Ala Arg Lys Lys Arg Lys Phe Asp Lys Asn
                      405                      410                      415

tga                                                                    1251

<210> 49
<211> 416
<212> PRT
<213> Porphyromonas gingivalis

<400> 49

Met Ala Glu Ser Arg Arg Lys Tyr Tyr Phe Pro Asp Val Thr Asp Glu
1                      5                      10                      15

Gln Trp Tyr Asp Trp His Trp Gln Val Leu Asn Arg Ile Lys Thr Leu
20                      25                      30

Asp Gln Leu Lys Lys Tyr Val Thr Leu Thr Ala Glu Glu Glu Glu Gly
35                      40                      45

Val Lys Glu Ser Pro Lys Val Leu Arg Met Ala Ile Thr Pro Tyr Tyr
50                      55                      60

Leu Ser Leu Ile Asp Pro Glu Asn Pro Asn Cys Pro Ile Arg Lys Gln
65                      70                      75                      80

```

Ala Ile Pro Thr Gln Gln Glu Leu Val Arg Ala Pro Glu Asp Gln Val
85 90 95

Asp Pro Leu Ser Glu Asp Glu Asp Ser Pro Val Pro Gly Leu Thr His
100 105 110

Arg Tyr Pro Asp Arg Val Leu Phe Leu Ile Thr Asp Lys Cys Ser Met
115 120 125

Tyr Cys Arg His Cys Thr Arg Arg Arg Phe Ala Gly Gln Lys Asp Ala
130 135 140

Ser Ser Pro Ser Glu Arg Ile Asp Arg Cys Ile Asp Tyr Ile Ala Asn
145 150 155 160

Thr Pro Thr Val Arg Asp Val Leu Leu Ser Gly Gly Asp Ala Leu Leu
165 170 175

Val Ser Asp Glu Arg Leu Glu Tyr Ile Leu Lys Arg Leu Arg Glu Val
180 185 190

Pro His Val Glu Ile Val Arg Ile Gly Ser Arg Thr Pro Val Val Leu
195 200 205

Pro Gln Arg Ile Thr Pro Gln Leu Val Asp Met Leu Lys Lys Tyr His
210 215 220

Pro Val Trp Leu Asn Thr His Phe Asn His Pro Asn Glu Val Thr Glu
225 230 235 240

Glu Ala Val Glu Ala Cys Glu Arg Met Ala Asn Ala Gly Ile Pro Leu
245 250 255

Gly Asn Gln Thr Val Leu Leu Arg Gly Ile Asn Asp Cys Thr His Val
260 265 270

Met Lys Arg Leu Val His Leu Leu Val Lys Met Arg Val Arg Pro Tyr
275 280 285

Tyr Ile Tyr Val Cys Asp Leu Ser Leu Gly Ile Gly His Phe Arg Thr
290 295 300

Pro Val Ser Lys Gly Ile Glu Ile Ile Glu Asn Leu Arg Gly His Thr
305 310 315 320

Ser Gly Tyr Ala Val Pro Thr Phe Val Val Gly Ala Pro Gly Gly Gly
325 330 335

Gly Lys Ile Pro Val Thr Pro Asn Tyr Val Val Ser Gln Ser Pro Arg
340 345 350

His Val Val Leu Arg Asn Tyr Glu Gly Val Ile Thr Thr Tyr Thr Glu
355 360 365

Pro Glu Asn Tyr His Glu Glu Cys Asp Cys Glu Asp Cys Arg Ala Gly
370 375 380

Lys His Lys Glu Gly Val Ala Ala Leu Ser Gly Gly Gln Gln Leu Ala
385 390 395 400

Ile Glu Pro Ser Asp Leu Ala Arg Lys Lys Arg Lys Phe Asp Lys Asn
405 410 415

<210> 50
<211> 1251
<212> DNA
<213> Porphyromonas gingivalis

<220>
<221> CDS
<222> (1)..(1251)

<400> 50
atg gca gaa agt cgt aga aag tat tat ttc cct gat gtc acc gat gag 48
Met Ala Glu Ser Arg Arg Lys Tyr Tyr Phe Pro Asp Val Thr Asp Glu
1 5 10 15

caa tgg tac gac tgg cat tgg cag gtc atc aat cga att aag acg ctc 96
Gln Trp Tyr Asp Trp His Trp Gln Val Ile Asn Arg Ile Lys Thr Leu
20 25 30

gac cag ctg aaa aag tac gtt aca ctc acc gct gaa gaa gaa gag gga 144
Asp Gln Leu Lys Lys Tyr Val Thr Leu Thr Ala Glu Glu Glu Glu Gly
35 40 45

gta aaa gaa tcg ccc aaa gta ctc cga atg gct atc aca oct tat tat 192
Val Lys Glu Ser Pro Lys Val Leu Arg Met Ala Ile Thr Pro Tyr Tyr
50 55 60

ttg agt ttg ata gac ccc gag aat cct aat tgt ccg att cgt aaa caa 240
Leu Ser Leu Ile Asp Pro Glu Asn Pro Asn Cys Pro Ile Arg Lys Gln

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

```

ccg gta tct aaa gga atc gaa att atc gaa aat ttg cgc gga cac acc      960
Pro Val Ser Lys Gly Ile Glu Ile Ile Glu Asn Leu Arg Gly His Thr
305                310                315                320

tcg ggc tat gca gtt cct acc ttt gtg gta ggt gct ccg ggg ggt ggt      1008
Ser Gly Tyr Ala Val Pro Thr Phe Val Val Gly Ala Pro Gly Gly Gly
                325                330                335

ggg aag ata cct gta acg ccg aac tat gtt gta tct cag tcc cca cga      1056
Gly Lys Ile Pro Val Thr Pro Asn Tyr Val Val Ser Gln Ser Pro Arg
                340                345                350

cat gtg gtt ctt cgc aat tat gaa ggt gtt atc aca acc tat acg gag      1104
His Val Val Leu Arg Asn Tyr Glu Gly Val Ile Thr Thr Tyr Thr Glu
                355                360                365

ccg gag aat tat cat gag gag tgc gat tgt gag gac tgt cga gcc ggt      1152
Pro Glu Asn Tyr His Glu Glu Cys Asp Cys Glu Asp Cys Arg Ala Gly
                370                375                380

aag cat aaa gag ggt gta gct gca ctt tcc gga ggt cag cag ttg gct      1200
Lys His Lys Glu Gly Val Ala Ala Leu Ser Gly Gly Gln Gln Leu Ala
385                390                395                400

atc gag cct tcc gac tta gct cgc aaa aaa cgc aag ttt gat aag aac      1248
Ile Glu Pro Ser Asp Leu Ala Arg Lys Lys Arg Lys Phe Asp Lys Asn
                405                410                415

tga                                                                    1251

<210> 51
<211> 416
<212> PRT
<213> Porphyromonas gingivalis

<400> 51

Met Ala Glu Ser Arg Arg Lys Tyr Tyr Phe Pro Asp Val Thr Asp Glu
1                5                10                15

Gln Trp Tyr Asp Trp His Trp Gln Val Ile Asn Arg Ile Lys Thr Leu
                20                25                30

Asp Gln Leu Lys Lys Tyr Val Thr Leu Thr Ala Glu Glu Glu Glu Gly
                35                40                45

Val Lys Glu Ser Pro Lys Val Leu Arg Met Ala Ile Thr Pro Tyr Tyr
                50                55                60

Leu Ser Leu Ile Asp Pro Glu Asn Pro Asn Cys Pro Ile Arg Lys Gln
65                70                75                80

```

Ala Ile Pro Thr Gln Gln Glu Leu Val Arg Ala Pro Glu Asp Gln Val
85 90 95

Asp Pro Leu Ser Glu Asp Glu Asp Ser Pro Val Pro Gly Leu Thr His
100 105 110

Arg Tyr Pro Asp Arg Val Leu Phe Leu Ile Thr Asp Lys Cys Ser Met
115 120 125

Tyr Cys Arg His Cys Thr Arg Arg Arg Phe Ala Gly Gln Lys Asp Ala
130 135 140

Ser Ser Pro Ser Glu Arg Ile Asp Arg Cys Ile Asp Tyr Ile Ala Asn
145 150 155 160

Thr Pro Thr Val Arg Asp Val Leu Leu Ser Gly Gly Asp Ala Leu Leu
165 170 175

Val Ser Asp Glu Arg Leu Glu Tyr Ile Leu Lys Arg Leu Arg Glu Val
180 185 190

Pro His Val Glu Ile Val Arg Ile Gly Ser Arg Thr Pro Val Val Leu
195 200 205

Pro Gln Arg Ile Thr Pro Gln Leu Val Asp Met Leu Lys Lys Tyr His
210 215 220

Pro Val Trp Leu Asn Thr His Phe Asn His Pro Asn Glu Val Thr Glu
225 230 235 240

Glu Ala Val Glu Ala Cys Glu Arg Met Ala Asn Ala Gly Ile Pro Leu
245 250 255

Gly Asn Gln Thr Val Leu Leu Arg Gly Ile Asn Asp Cys Thr His Val
260 265 270

Met Lys Arg Leu Val His Leu Leu Val Lys Met Arg Val Arg Pro Tyr
275 280 285

Tyr Ile Tyr Val Cys Asp Leu Ser Leu Gly Ile Gly His Phe Arg Thr
290 295 300

Pro Val Ser Lys Gly Ile Glu Ile Ile Glu Asn Leu Arg Gly His Thr
305 310 315 320

Ser Gly Tyr Ala Val Pro Thr Phe Val Val Gly Ala Pro Gly Gly Gly
325 330 335

Gly Lys Ile Pro Val Thr Pro Asn Tyr Val Val Ser Gln Ser Pro Arg
340 345 350

His Val Val Leu Arg Asn Tyr Glu Gly Val Ile Thr Thr Tyr Thr Glu
355 360 365

Pro Glu Asn Tyr His Glu Glu Cys Asp Cys Glu Asp Cys Arg Ala Gly
370 375 380

Lys His Lys Glu Gly Val Ala Ala Leu Ser Gly Gly Gln Gln Leu Ala
385 390 395 400

Ile Glu Pro Ser Asp Leu Ala Arg Lys Lys Arg Lys Phe Asp Lys Asn
405 410 415

<210> 52

<211> 416

<212> PRT

<213> Porphyromonas gingivalis

<400> 52

Met Ala Glu Ser Arg Arg Lys Tyr Tyr Phe Pro Asp Val Thr Asp Glu
1 5 10 15

Gln Trp Asn Asp Trp His Trp Gln Val Leu Asn Arg Ile Glu Thr Leu
20 25 30

Asp Gln Leu Lys Lys Tyr Val Thr Leu Thr Ala Glu Glu Glu Glu Gly
35 40 45

Val Lys Glu Ser Leu Lys Val Leu Arg Met Ala Ile Thr Pro Tyr Tyr
50 55 60

Leu Ser Leu Ile Asp Pro Glu Asn Pro Asn Cys Pro Ile Arg Lys Gln
65 70 75 80

Ala Ile Pro Thr His Gln Glu Leu Val Arg Ala Pro Glu Asp Gln Val

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

<400> 55
 ttaccgagca gcggttcagag 20

<210> 56
 <211> 20
 <212> DNA
 <213> Artificial

<220>
 <223> primer

<400> 56
 cacctggcgg tgacaacat 20

<210> 57
 <211> 60
 <212> DNA
 <213> Artificial

<220>
 <223> primer

<400> 57
 gcggcgtgaa gtttccaac cgtttctgcc tctcttcttc gtgtaggctg gagctgcttc 60

<210> 58
 <211> 60
 <212> DNA
 <213> Artificial

<220>
 <223> primer

<400> 58
 ttacaacggt accgggtggt ctttctcgcc tttcttaaac catatgaata tctctcttag 60

<210> 59
 <211> 471
 <212> PRT
 <213> Bacillus subtilis

<400> 59
 Met Lys Asn Lys Trp Tyr Lys Pro Lys Arg His Trp Lys Glu Ile Glu
 1 5 10 15

Leu Trp Lys Asp Val Pro Glu Glu Lys Trp Asn Asp Trp Leu Trp Gln
 20 25 30

Leu Thr His Thr Val Arg Thr Leu Asp Asp Leu Lys Lys Val Ile Asn

Gly Ile Asn Asp Ser Val Pro Ile Met Lys Lys Leu Met His Asp Leu
 275 280 285

Val Lys Ile Arg Val Arg Pro Tyr Tyr Ile Tyr Gln Cys Asp Leu Ser
 290 295 300

Glu Gly Ile Gly His Phe Arg Ala Pro Val Ser Lys Gly Leu Glu Ile
 305 310 315 320

Ile Glu Gly Leu Arg Gly His Thr Ser Gly Tyr Ala Val Pro Thr Phe
 325 330 335

Val Val Asp Ala Pro Gly Gly Gly Gly Lys Ile Ala Leu Gln Pro Asn
 340 345 350

Tyr Val Leu Ser Gln Ser Pro Asp Lys Val Ile Leu Arg Asn Phe Glu
 355 360 365

Gly Val Ile Thr Ser Tyr Pro Glu Pro Glu Asn Tyr Ile Pro Asn Gln
 370 375 380

Ala Asp Ala Tyr Phe Glu Ser Val Phe Pro Glu Thr Ala Asp Lys Lys
 385 390 395 400

Glu Pro Ile Gly Leu Ser Ala Ile Phe Ala Asp Lys Glu Val Ser Phe
 405 410 415

Thr Pro Glu Asn Val Asp Arg Ile Lys Arg Arg Glu Ala Tyr Ile Ala
 420 425 430

Asn Pro Glu His Glu Thr Leu Lys Asp Arg Arg Glu Lys Arg Asp Gln
 435 440 445

Leu Lys Glu Lys Lys Phe Leu Ala Gln Gln Lys Lys Gln Lys Glu Thr
 450 455 460

Glu Cys Gly Gly Asp Ser Ser
 465 470

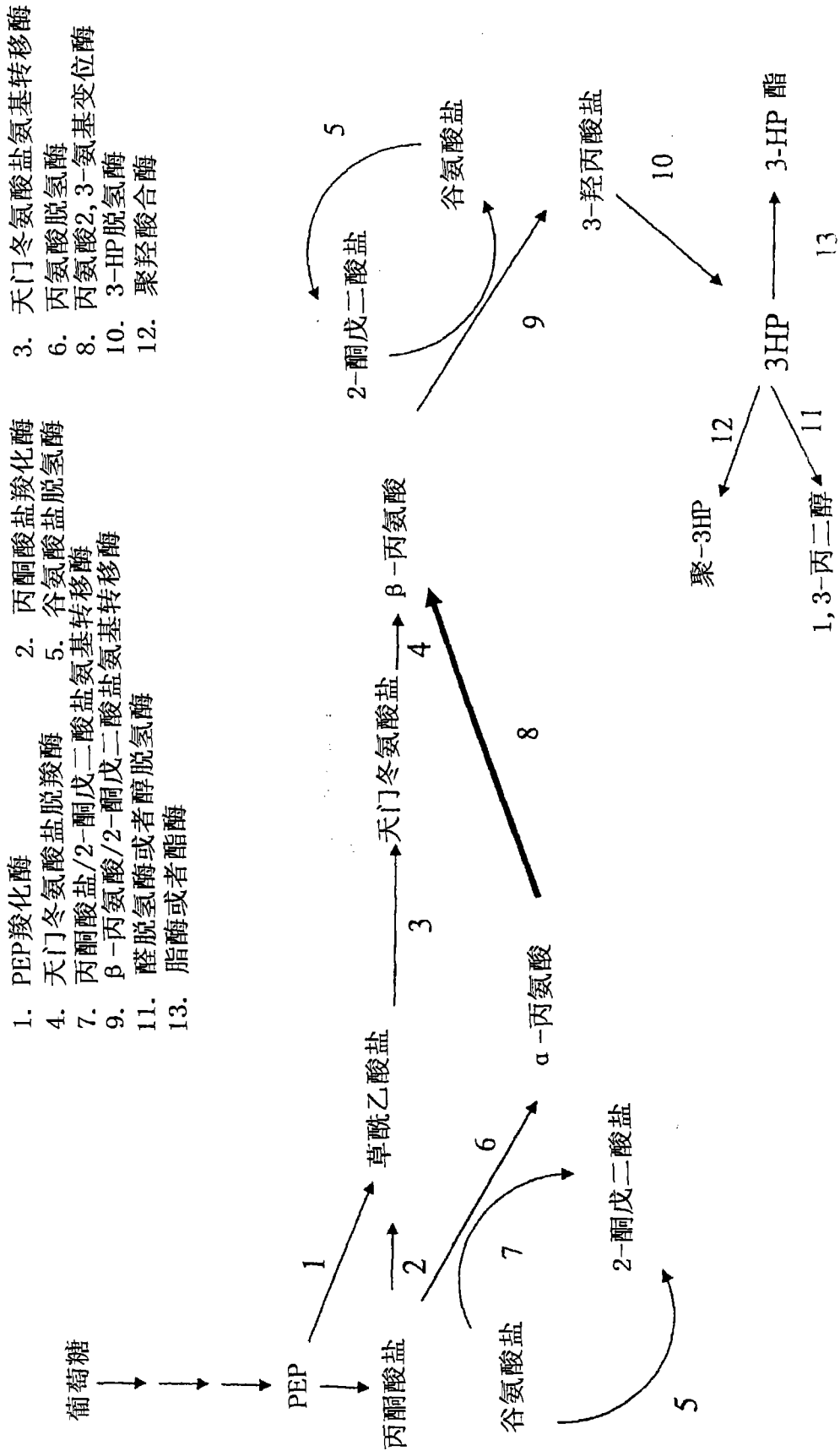


图1

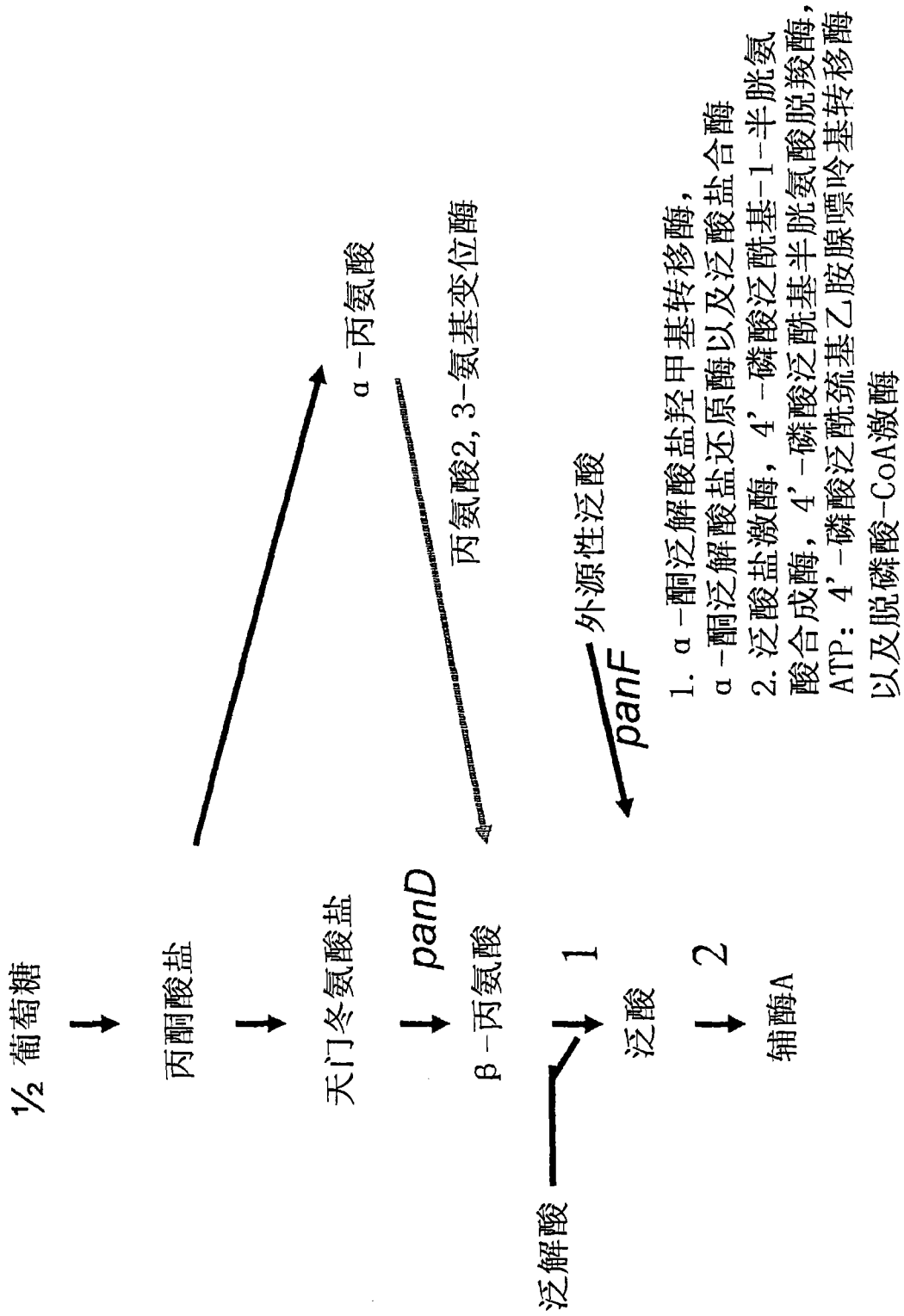


图3

Fnkam	1 mntvnrkkffpnvtdeewndwtwqvknrleavedlkkyvdlseeetegv
Fnaam	1 mntvnrkkffpnvtdeewndwtwqvknrksvedlkkyvdlseeetegv
Fnaam2	1 mntvnrkkffpnvtdeewndwtwqvknrksvedlekyvdlseeetegv
Fnkam	51 vrtletlrmaitypyfslidlndrcpirkqaipitqeihqsdadlldpl
Fnaam	51 vrtletlrmaitypyfslidlndrcpirkqaipitireihqsdadmldpl
Fnaam2	51 vrtletlrmaitypyfslidlndrcpirkqaipitireihqsdadmldpl
Fnkam	101 hededspvpglthrypdrvlllitdmcsmycrhctrrrfagssddampmd
Fnaam	101 hededspvpglthrypdrvlllitdmcsvycrhtctrrrfagssdgampmd
Fnaam2	101 hededspvpglthrypdrvlllitdmcsvycrhtctrrrfagssdgampmd
Fnkam	151 ridkaieyiaktpqvrdrvllsggdallvsdkklesiiqklraiphveiir
Fnaam	151 ridkaieyiaktpqvrdrvllsggdallvsdkklesiiqklraiphveiir
Fnaam2	151 ridkaieyiaktpqvrdrvllsggdallvsnklesiiqklraiphveiir
Fnkam	201 igsrtpvvlpqritpelcnmlkkyhpiwlntfhnhpqevtpeakkaceml
Fnaam	201 igsrtpvvlpqritpelcnmlkkyhpiwmnthfnhpqevtpeakkaceml
Fnaam2	201 igsrtpvvlpqritpelcnmlkkyhpiwmnthfnhpqevtpeakkaceml
Fnkam	251 adagvplgnqtvllrgindsvpvmkrvlvhdvmmrvrpyyiyqcdlsmgl
Fnaam	251 adagvplgnqtvllrgindsvpvmkrvlvhdvmmrvrpyyiyqcdlsmgl
Fnaam2	251 adagvplgnqtvllrgindsvpvmkrvlvhdvmmrvrpyyiyqcdlsmgl
Fnkam	301 ehfrtpvskgieiieglrghtsgyavptfvvdapgggktpvmpqyvisq
Fnaam	301 ehfrtpvskgieiieglrghtsgyavptfvvhapgggktpvmpqyvisq
Fnaam2	301 ehfrtpvskgieiieglrghtsgyavptfvvhapgggktpvmpqyvisq
Fnkam	351 sphrvvlrnfegvittytepenythepecydeekfekmyeisgvmldegl
Fnaam	351 sphrvvlrnfegvittytepenythepecydeekfekmyeisgvmldegl
Fnaam2	351 sphrvvlrnfegvittytepenythepecydeekfekmyeisgvmldegl
Fnkam	401 kmslepshlarhernkkraeaegkk
Fnaam	401 emslepshlarhernkkraeaegkk
Fnaam2	401 emslepshlarhernkkraeaegkk

图4

Cskam	1	mslkdkffshvsqedwndkwqvrnrietveelkkyipltpeeeegvkrc
Cscodm	1	mslkdkffshvsqedwndkwqvrnriktveelkkyipltpeeeegvkrc
Cscodm mut8	1	mslkdkffthvsqedwndkwqvrnriktveelkkyipltpeeeegvkrc
Cscodm mut12	1	mslkdkffshvsqedwndkwqvrnriktaeelkkyipltpeeeegvkrc
Cscodm mut15	1	mslkdkffshvsqedwndkwqvrnriktveelkkyipltpeeeegvkrr
Cskam	51	ldtlrmaitypylslidvenpndpvrkqavplslelhraasdedplhed
Cscodm	51	ldtlrmaitypylslidvenpndpvrkqavplslelhraasdedplhed
Cscodm mut8	51	ldtlrmaitypylslidvenpndpvrkqavplslelhraasdedplhed
Cscodm mut12	51	hdtlrmaitypylslidvgnpndpvrkqavplslelhraasdedplhed
Cscodm mut15	51	ldtlrmaitypylslidvenpndpvrkqavplslelhraasdedplhed
Cskam	101	gdsppvplthrypdrvlllmtdqcsmycrhctrrrfagqtdsavidtkqid
Cscodm	101	gdsppvplthrypdrvlllmtdqcsvycrhcrrrfagqtdsavidtkqid
Cscodm mut8	101	gdsppvplthrypdrvlllmtdqcsvycrhcrrrfagrtdsavidtkqid
Cscodm mut12	101	gdsppvplthrypdrvlllmtdlcsvycrhcrrrfagqtdsavidtkqid
Cscodm mut15	101	gdsppvplthrypdrvlllmtdqcsvycrhcrrrfagqtdsavidtkqid
Cskam	151	aaeyikntpqvrdvllsggdallisdekleytikrlreiphvevirigs
Cscodm	151	aaeyikntpqvrdvllsggdallisdekleytikrlreiphvevirigs
Cscodm mut8	151	aaeyikntpqvrdvllsggdallisdekleytirrlreiphvevirigs
Cscodm mut12	151	aaeyikntpqvrdvllsggdallisdekleytikrlreiphvevirigs
Cscodm mut15	151	aaeyikntpqvrdvllsggdallisdekleytikrlreiphvevirigs
Cskam	201	rvpvmpqritpelvsmkkkyhpwlnthfnhpneiteeskracellada
Cscodm	201	rvpvmpqritpelvsmkkkyhpwlnthfnhpneiteeskracellada
Cscodm mut8	201	rvpvmpqritpelvsmkkkyhpwlnthfnhpneiteeskracellada
Cscodm mut12	201	rvpvmpqritpelvsmkkkyhpwlnthfnhpneiteeskracellada
Cscodm mut15	201	rvpvmpqritpelvsmkkkyhpwlnthfnhpneiteeskracellada
Cskam	251	giplgnqsvllagvndcmhvmkklvndlvkirvrpyyiyqcdlsvgiehf
Cscodm	251	giplgnqsvllagvndcmhvmkklvndlvkirvrpyyiyqcdlsvgiehf
Cscodm mut8	251	giplgnqsvllagvndcmhvmkklvndlvkirvrpyyiyqcdlsvgiehf
Cscodm mut12	251	giplgnqsvllagvndcmhvmkklvndlvkirvrpyyiyqcdlsvgiehf
Cscodm mut15	251	giplgnqsvllagvndcmhvmkklvndlvkirvrpyyiyqcdlsvgiehf
Cskam	301	rtpvakgieieglrghtsgycvptfvvdapggggktpvmpnyvisqnhn
Cscodm	301	rtpvakgieieglrghtsgycvptfvvhapggggktpvmpnyvisqnhn
Cscodm mut8	301	rtpvakgieieglrghtsgycvptfvvhapggggktpvmpnyvisqnhn
Cscodm mut12	301	rtpvakgieieglrghtsgycvptfvvhapggggktpvmpnyvisqnhn
Cscodm mut15	301	rtpvakgieieglrghtsgycvptfvvhapggggktpvmpnyvisqnhn
Cskam	351	kvilrnfeqvittydepdhytfhcdcdvctgktnvhkvgvagllngetat
Cscodm	351	kvilrnfeqvittydepdhytfhcdcdvctgktnvhkvgvagllngetat
Cscodm mut8	351	kvilrnfeqvittydepdhytfhcdcdvctgktnvhkvgvagllngetat
Cscodm mut12	351	kvilrnfeqvittydepdhytfhcdcdvctgktnvhkvgvagllngetat
Cscodm mut15	351	kvilrnfeqvittydepdhytfhcdcdvctgktnvhkvgvagllngetat
Cskam	401	lepeglerkqrghh
Cscodm	401	lepeglerkqrghh
Cscodm mut8	401	lepeglerkqrghh
Cscodm mut12	401	lepeglerkqrghh
Cscodm mut15	401	lepeglerkqrghh

图5

Pgkam	1	maesrrkyyfpdvtdeqwndwhwqvlrietldqlkkyvtltaeeeeegvk
Pgaam	1	maesrrkyyfpdvtdeqwydwhwqvlrietldqlkkyvtltaeeeeegvk
Pgaam2	1	maesrrkyyfpdvtdeqwydwhwqvlriektldqlkkyvtltaeeeeegvk
Pgaam2L26I	1	maesrrkyyfpdvtdeqwydwhwqvinriektldqlkkyvtltaeeeeegvk
Pgkam	51	esklvrlmaitpyylslidpenpncpirkqaipthqelvrappedqvdpis
Pgaam	51	espkvlrmaitypyylslidpenpncpirkqaipthqelvrappedqvdpis
Pgaam2	51	espkvlrmaitypyylslidpenpncpirkqaipthqelvrappedqvdpis
Pgaam2L26I	51	espkvlrmaitypyylslidpenpncpirkqaipthqelvrappedqvdpis
Pgkam	101	ededspvpglthrypdrvflitdkcsmycrhctrrrfagqkdasspser
Pgaam	101	ededspvpglthrypdrvflitdkcsmycrhctrrrfagqkdasspser
Pgaam2	101	ededspvpglthrypdrvflitdkcsmycrhctrrrfagqkdasspser
Pgaam2L26I	101	ededspvpglthrypdrvflitdkcsmycrhctrrrfagqkdasspser
Pgkam	151	idrcidyiantptvrdvllsggdallvsderleyilkrreiphveivri
Pgaam	151	idrcidyiantptvrdvllsggdallvsderleyilkrreiphveivri
Pgaam2	151	idrcidyiantptvrdvllsggdallvsderleyilkrrevphveivri
Pgaam2L26I	451	idrcidyiantptvrdvllsggdallvsderleyilkrrevphveivri
Pgkam	201	gsrtpvvlpqrirtpqlvdmkkyhpvwlntfhfhpnevteeaveacerma
Pgaam	201	gsrtpvvlpqrirtpqlvdmkkyhpvwlntfhfhpnevteeaveacerma
Pgaam2	201	gsrtpvvlpqrirtpqlvdmkkyhpvwlntfhfhpnevteeaveacerma
Pgaam2L26I	201	gsrtpvvlpqrirtpqlvdmkkyhpvwlntfhfhpnevteeaveacerma
Pgkam	251	nagiplgnqtvllrgindcthvmlkrvlhllvkmrvrpyyiyvcdlslgig
Pgaam	251	nagiplgnqtvllrgindcthvmlkrvlhllvkmrvrpyyiyvcdlslgig
Pgaam2	251	nagiplgnqtvllrgindcthvmlkrvlhllvkmrvrpyyiyvcdlslgig
Pgaam2L26I	251	nagiplgnqtvllrgindcthvmlkrvlhllvkmrvrpyyiyvcdlslgig
Pgkam	301	hfrtpvskgieiienlrghstgyavptfvvdapgggkipvmpnyvvsqs
Pgaam	301	hfrtpvskgieiienlrghstgyavptfvvgapgggkipvtpnyvvsqs
Pgaam2	301	hfrtpvskgieiienlrghstgyavptfvvgapgggkipvtpnyvvsqs
Pgaam2L26I	301	hfrtpvskgieiienlrghstgyavptfvvgapgggkipvtpnyvvsqs
Pgkam	351	prhvvlrnyegvittytepenyheecdcedcragkhkegvaalsggqqla
Pgaam	351	prhvvlrnyegvittytepenyheecdcedcragkhkegvaalsggqqla
Pgaam2	351	prhvvlrnyegvittytepenyheecdcedcragkhkegvaalsggqqla
Pgaam2L26I	351	prhvvlrnyegvittytepenyheecdcedcragkhkegvaalsggqqla
Pgkam	401	iepsdlarkkrkfdkn-
Pgaam	401	iepsdlarkkrkfdkn-
Pgaam2	401	iepsdlarkkrkfdkn-
Pgaam2L26I	401	iepsdlarkkrkfdkn-

图6

```

Pgkam      1 -maesrrkyyfp-----dvtdeqwndwhwqvlrxietldqlkky
Pgaam      1 -maesrrkyyfp-----dvtdeqwydwhwqvlrxietldqlkky
Pgaam2     1 -maesrrkyyfp-----dvtdeqwydwhwqvlrxietldqlkky
Pgaam2L26I 1 -maesrrkyyfp-----dvtdeqwydwhwqvlrxietldqlkky
Fnkam      1 mntvntrkkkffp-----nvtdeewndwtwqvkxrlesvedlkky
Fncodm     1 mntvntrkkkffp-----nvtdeewndwtwqvkxrlesvedlkky
Fnaam      1 mntvntrkkkffp-----nvtdeewndwtwqvkxrlesvedlkky
Fnaam2     1 mntvntrkkkffp-----nvtdeewndwtwqvkxrlesvedlkky
Cskam      1 ---mslkdkkffs-----hvsqedwndwkwqvrnriktveelkky
Cscodm     1 ---mslkdkkffs-----hvsqedwndwkwqvrnriktveelkky
Cscodm mut8 1 ---mslkdkkffs-----hvsqedwndwkwqvrnriktveelkky
Cscodm mut12 1 ---mslkdkkffs-----hvsqedwndwkwqvrnriktveelkky
Cscodm mut15 1 ---mslkdkkffs-----hvsqedwndwkwqvrnriktveelkky
Bskam      1 ---mknkwykprhwkeielwkdvpeekwndwlvqlthtvtlddlkky
Bsaam      1 ---mknkwykprhwkeielwkdvpeekwndwlvqlthtvtlddlkky
Bsaam2co   1 ----mknkwykprhwkeielwkdvpeekwndwlvqlthtvtlddlkky

Pgkam      39 vlttaeeegvkeslklvrmaitpyylslidpenpncpirkqaipthqel
Pgaam      39 vlttaeeegvkespkvlrmaitpyylslidpenpncpirkqaipthqel
Pgaam2     39 vlttaeeegvkespkvlrmaitpyylslidpenpncpirkqaipthqel
Pgaam2L26I 39 vlttaeeegvkespkvlrmaitpyylslidpenpncpirkqaipthqel
Fnkam      40 vdlseeetegvrtletlrmaitpyyflslidlnsdrpcpirkqaipthqei
Fncodm     40 vdlseeetegvrtletlrmaitpyyflslidlnsdrpcpirkqaipthqei
Fnaam      40 vdlseeetegvrtletlrmaitpyyflslidlnsdrpcpirkqaipthqei
Fnaam2     40 vdlseeetegvrtletlrmaitpyyflslidlnsdrpcpirkqaipthqei
Cskam      37 ipltpeeeegvkrcltdlrmaitpyylslidvenpndpvrkqavplslel
Cscodm     37 ipltpeeeegvkrcltdlrmaitpyylslidvenpndpvrkqavplslel
Cscodm mut8 37 ipltpeeeegvkrcltdlrmaitpyylslidvenpndpvrkqavplslel
Cscodm mut12 37 ipltpeeeegvkrcltdlrmaitpyylslidvenpndpvrkqavplslel
Cscodm mut15 37 ipltpeeeegvkrcltdlrmaitpyylslidvenpndpvrkqavplslel
Bskam      47 inltedeeegvristkktipltitpyyaslmpdnprcpvrmsqsvplseem
Bsaam      47 inltedeeegvristkktipltitpyyaslmpdnprcpvrmsqsvplseem
Bsaam2co   47 inltedeeegvristkktipltitpyyaslmpdnprcpvrmsqsvplseem

Pgkam      89 vrapedqvdplsededspvpglthrypdrvflitdkcsmycrhctrrrf
Pgaam      89 vrapedqvdplsededspvpglthrypdrvflitdkcsmycrhctrrrf
Pgaam2     89 vrapedqvdplsededspvpglthrypdrvflitdkcsmycrhctrrrf
Pgaam2L26I 89 vrapedqvdplsededspvpglthrypdrvflitdkcsmycrhctrrrf
Fnkam      90 hqsdadlldplhededspvpglthrypdrvllitdmcsmycrhctrrrf
Fncodm     90 hqsdadlldplhededspvpglthrypdrvllitdmcsmycrhctrrrf
Fnaam      90 hqsdadlldplhededspvpglthrypdrvllitdmcsmycrhctrrrf
Fnaam2     90 hqsdadlldplhededspvpglthrypdrvllitdmcsmycrhctrrrf
Cskam      87 hraasdmedplhedgdspvpglthrypdrvllitdmcsmycrhctrrrf
Cscodm     87 hraasdmedplhedgdspvpglthrypdrvllitdmcsmycrhctrrrf
Cscodm mut8 87 hraasdmedplhedgdspvpglthrypdrvllitdmcsmycrhctrrrf
Cscodm mut12 87 hraasdmedplhedgdspvpglthrypdrvllitdmcsmycrhctrrrf
Cscodm mut15 87 hraasdmedplhedgdspvpglthrypdrvllitdmcsmycrhctrrrf
Bskam      97 hktkydledplhededspvpglthrypdrvflvtngqcsmycryctrrrf
Bsaam      97 hktkydmedplhededspvpglthrypdrvflvtngqcsmycryctrrrf
Bsaam2co   97 hktkydmedplhededspvpglthrypdrvflvtngqcsmycryctrrrf

```

图7A

Pgkam	139	agqkdasspseridrcidyantptvrdvllsggdallvsderleyilkr
Pgaam	139	agqkdasspseridrcidyantptvrdvllsggdallvsderleyilkr
Pgaam2	139	agqkdasspseridrcidyantptvrdvllsggdallvsderleyilkr
Pgaam2L26I	139	agqkdasspseridrcidyantptvrdvllsggdallvsderleyilkr
Fnkam	140	agssddampmdridkaieyiaktppqvrdivllsggdallvsdkklesiik
Fncodm	140	agssddampmdridkaieyiaktppqvrdivllsggdallvsdkklesiik
Fnaam	140	agssdgampmdridkaieyiaktppqvrdivllsggdallvsdkklesiik
Fnaam2	140	agssdgampmdridkaieyiaktppqvrdivllsggdallvsdkklesiik
Cskam	137	agqtdsavdtkqidaaieyikntppqvrdivllsggdallisdekleytikr
Cscodm	137	agqtdsavdtkqidaaieyikntppqvrdivllsggdallisdekleytikr
Cscodm mut8	137	agrtdsavdtkqidaaieyikntppqvrdivllsggdallisdekleytirr
Cscodm mut12	137	agqtdsavdtkqidaaieyikntppqvrdivllsggdallisdekleytikr
Cscodm mut15	137	agqtdsavdtkqidaaieyikntppqvrdivllsggdallisdekleytikr
Bskam	147	sgqigmgvppkkqldaaiayiretpeirdclisggdglindqileyilke
Bsaam	147	sgqigmgvppkkqldaaiayiretpeirdclisggdglindqileyilke
Bsaam2co	147	sgqigmgvppkkqldaaiayiretpeirdclisggdglindqileyilke
Pgkam	189	lreiphveivrigsrtpvvlpqrtpqlvdmlkkyhpvwlntfhfhpnev
Pgaam	189	lreiphveivrigsrtpvvlpqrtpqlvdmlkkyhpvwlntfhfhpnev
Pgaam2	189	lrevphveivrigsrtpvvlpqrtpqlvdmlkkyhpvwlntfhfhpnev
Pgaam2L26I	189	lrevphveivrigsrtpvvlpqrtpqlvdmlkkyhpvwlntfhfhpnev
Fnkam	190	lraiphveivrigsrtpvvlpqrtpelcnmlkkyhpiwlntfhfhpnev
Fncodm	190	lraiphveivrigsrtpvvlpqrtpelcnmlkkyhpiwlntfhfhpnev
Fnaam	190	lraiphveivrigsrtpvvlpqrtpelcnmlkkyhpiwlntfhfhpnev
Fnaam2	190	lraiphveivrigsrtpvvlpqrtpelcnmlkkyhpiwlntfhfhpnev
Cskam	187	lreiphveivrigsrtpvvlpqrtpelvsmlkkyhpvwlntfhfhpnei
Cscodm	187	lreiphveivrigsrtpvvlpqrtpelvsmlkkyhpvwlntfhfhpnei
Cscodm mut8	187	lreiphveivrigsrtpvvlpqrtpelvsmlkkyhpvwlntfhfhpnei
Cscodm mut12	187	lreiphveivrigsrtpvvlpqrtpelvsmlkkyhpvwlntfhfhpnei
Cscodm mut15	187	lreiphveivrigsrtpvvlpqrtpelvsmlkkyhpvwlntfhfhpnei
Bskam	197	lrsiphlevirigtrapvvpqrtdhlceilkkyhpvwlntfhfntsiem
Bsaam	197	lrsiphlevirigtrapvvpqrtdhlceilkkyhpvwlntfhfntsiem
Bsaam2co	197	lrsiphlevirigtrapvvpqrtdhlceilkkyhpvwlntfhfntsiem
Pgkam	239	teeaveacermanagiplgnqtvllrgindcthvmlkrlvhlvlkrmvrpy
Pgaam	239	teeaveacermanagiplgnqtvllrgindcthvmlkrlvhlvlkrmvrpy
Pgaam2	239	teeaveacermanagiplgnqtvllrgindcthvmlkrlvhlvlkrmvrpy
Pgaam2L26I	239	teeaveacermanagiplgnqtvllrgindcthvmlkrlvhlvlkrmvrpy
Fnkam	240	tpeakkacemladagvplgnqtvllrgindsvpvmkrlvhdlvmmvrpy
Fncodm	240	tpeakkacemladagvplgnqtvllrgindsvpvmkrlvhdlvmmvrpy
Fnaam	240	tpeakkacemladagvplgnqtvllrgindsvpvmkrlvhdlvmmvrpy
Fnaam2	240	tpeakkacemladagvplgnqtvllrgindsvpvmkrlvhdlvmmvrpy
Cskam	237	teeskracelladagiplgnqsvllagvndcmhvmkklvndlvlkivrpy
Cscodm	237	teeskracelladagiplgnqsvllagvndcmhvmkklvndlvlkivrpy
Cscodm mut8	237	teeskracelladagiplgnqsvllagvndcmhvmkklvndlvlkivrpy
Cscodm mut12	237	teeskracelladagiplgnqsvllagvndcmhvmkklvndlvlkivrpy
Cscodm mut15	237	teeskracelladagiplgnqsvllagvndcmhvmkklvndlvlkivrpy
Bskam	247	teesveaceklvnagvpgnqavvlagindsvpimkklmhdvlkivrpy
Bsaam	247	teesveaceklvnagvpgnqavvlagindsvpimkklmhdvlkivrpy
Bsaam2co	247	teesveaceklvnagvpgnqavvlagindsvpimkklmhdvlkivrpy

图7B

Pgkam	289	yyvcdlslgighfrtpvskgieiienlrghtsgyavptfvvdapggggk
Pgaam	289	yyvcdlslgighfrtpvskgieiienlrghtsgyavptfvvgapggggk
Pgaam2	289	yyvcdlslgighfrtpvskgieiienlrghtsgyavptfvvgapggggk
Pgaam2L26I	289	yyvcdlslgighfrtpvskgieiienlrghtsgyavptfvvgapggggk
Fnkam	290	yyqcdlsmglehfrtpvskgieieglrghtsgyavptfvvdapggggk
Fncodm	290	yyqcdlsmglehfrtpvskgieieglrghtsgyavptfvvhapggggk
Fnaam	290	yyqcdlsmglehfrtpvskgieieglrghtsgyavptfvvhapggggk
Fnaam2	290	yyqcdlsmglehfrtpvskgieieglrghtsgyavptfvvhapggggk
Cskam	287	yyqcdlsvgiehfrtpvakgieieglrghtsgycvptfvvdapggggk
Cscodm	287	yyqcdlsvgiehfrtpvakgieieglrghtsgycvptfvvhapggggk
Cscodm mut8	287	yyqcdlsvgiehfrtpvakgieieglrghtsgycvptfvvhapggggk
Cscodm mut12	287	yyqcdlsvgiehfrtpvakgieieglrghtsgycvptfvvhapggggk
Cscodm mut15	287	yyqcdlsvgiehfrtpvakgieieglrghtsgycvptfvvhapggggk
Bskam	297	yyqcdlsegighfrapvskgleieglrghtsgyavptfvvdapggggk
Bsaam	297	yyqcdlsegighfrapvskgleieglrghtsgyavptfvvhapggggk
Bsaam2co	297	yyqcdlsegighfrapvskgleieglrghtsgyavptfvvhapggggk
Pgkam	339	ipvmpnyvvsqsprhvvlrnyegvittytepeny-----heec-dced
Pgaam	339	ipvtpnyvvsqsprhvvlrnyegvittytepeny-----heec-dced
Pgaam2	339	ipvtpnyvvsqsprhvvlrnyegvittytepeny-----heec-dced
Pgaam2L26I	339	ipvtpnyvvsqsprhvvlrnyegvittytepeny-----heec-dced
Fnkam	340	tpvmpqyvisqsphrvvlrnfegvittytepenyt-----hepcydeek
Fncodm	340	tpvmpqyvisqsphrvvlrnfegvittytepenyt-----hepcydeek
Fnaam	340	tpvmpqyvisqsphrvvlrnfegvittytepenyt-----hepcydeek
Fnaam2	340	tpvmpqyvisqsphrvvlrnfegvittytepenyt-----hepcydeek
Cskam	337	tpvmpnyvisqnhnkvilrnfegvittydepdhy-----tfhc-dcdv
Cscodm	337	tpvmpnyvisqnhnkvilrnfegvittydepdhy-----tfhc-dcdv
Cscodm mut8	337	tpvmpnyvisqnhnkvilrnfegvittydepdhy-----tfhc-dcdv
Cscodm mut12	337	tpvmpnyvisqnhnkvilrnfegvittydepdhy-----tfhc-dcdv
Cscodm mut15	337	tpvmpnyvisqnhnkvilrnfegvittydepdhy-----tfhc-dcdv
Bskam	347	ialqpnnyvlsqspdkvilrnfegvitsypepenyipnqadayfes-vfpe
Bsaam	347	ialqpnnyvlsqspdkvilrnfegvitsypepenyipnqadayfes-vfpe
Bsaam2co	347	ialqpnnyvlsqspdkvilrnfegvitsypepenyipnqadayfes-vfpe
Pgkam	381	c--ragkhkegvaalsggqqlaiepsdlar-----
Pgaam	381	c--ragkhkegvaalsggqqlaiepsdlar-----
Pgaam2	381	c--ragkhkegvaalsggqqlaiepsdlar-----
Pgaam2L26I	381	c--ragkhkegvaalsggqqlaiepsdlar-----
Fnkam	384	f--ekmyeisgvymldeglmslepshlar-----
Fncodm	384	f--ekmyeisgvymldeglmslepshlar-----
Fnaam	384	f--ekmyeisgvymldeglmslepshlar-----
Fnaam2	384	f--ekmyeisgvymldeglmslepshlar-----
Cskam	379	ctgktnvhkvgvagllngetatlepegler-----
Cscodm	379	ctgktnvhkvgvagllngetatlepegler-----
Cscodm mut8	379	ctgktnvhkvgvagllngetatlepegler-----
Cscodm mut12	379	ctgktnvhkvgvagllngetatlepegler-----
Cscodm mut15	379	ctgktnvhkvgvagllngetatlepegler-----
Bskam	396	t--adkkepiglsaifadkevsftpenvdrikreayianpehetlkdr
Bsaam	396	t--adkkepiglsaifadkevsftpenvdrikreayianpehetlkdr
Bsaam2co	396	t--adkkepiglsaifadkevsftpenvdrikreayianpehetlkdr

图7C

Pgkam	409	-----kkrkf-----dkn-----
Pgaam	409	-----kkrkf-----dkn-----
Pgaam2	409	-----kkrkf-----dkn-----
Pgaam2L26I	409	-----kkrkf-----dkn-----
Fnkam	412	-----hernk-----kraeaegkk
Fncodm	412	-----hernk-----kraeaegkk
Fnaam	412	-----hernk-----kraeaegkk
Fnaam2	412	-----hernk-----kraeaegkk
Cskam	409	-----kqrgh-----h-----
Cscodm	409	-----kqrgh-----h-----
Cscodm mut8	409	-----kqrgh-----h-----
Cscodm mut12	409	-----kqrgh-----h-----
Cscodm mut15	409	-----kqrgh-----h-----
Bskam	444	ekrdqlkekfflaqqkkqketecggdss-----
Bsaam	444	ekrdqlkekfflaqqkkqketecggdss-----
Bsaam2co	444	ekrdqlkekfflaqqkkqketecggdss-----

图7D