



(12) 发明专利申请

(10) 申请公布号 CN 102203123 A

(43) 申请公布日 2011. 09. 28

(21) 申请号 200980143919. 3

代理人 董巍 谢枸

(22) 申请日 2009. 10. 31

(51) Int. Cl.

(30) 优先权数据

08400044. 7 2008. 10. 31 EP

61/198, 023 2008. 10. 31 US

C07K 14/395 (2006. 01)

C12P 21/00 (2006. 01)

C12N 15/81 (2006. 01)

C12N 9/10 (2006. 01)

C07K 14/44 (2006. 01)

(85) PCT申请进入国家阶段日

2011. 05. 03

(86) PCT申请的申请数据

PCT/EP2009/007816 2009. 10. 31

(87) PCT申请的公布数据

W02010/049177 EN 2010. 05. 06

(71) 申请人 隆萨股份公司

地址 瑞士巴塞尔

(72) 发明人 乔尼·赫勒纽斯 克里斯廷·纽珀特

马库斯·艾比

法劳西·帕尔赛-纳赛比

亚历山大·丹尼尔·弗雷

(74) 专利代理机构 北京市磐华律师事务所

11336

权利要求书 8 页 说明书 81 页

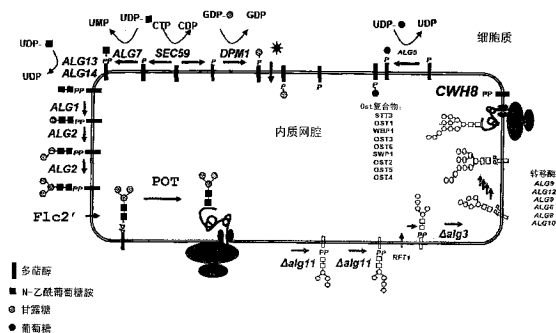
序列表 73 页 附图 26 页

(54) 发明名称

在宿主细胞中制造糖基化蛋白的新方法

(57) 摘要

本发明改善了真核生物中的糖蛋白制造及蛋白糖基化工程,特别是类人复合物或杂交糖蛋白在例如酵母等低级真核生物中的制造。本发明提供了经糖基化修饰的真核宿主细胞,其能够制造可用作免疫球蛋白及其他治疗用蛋白的糖基化优化蛋白,并提供了能够制造具有与人细胞中制造的糖蛋白类似的多糖结构的糖蛋白的细胞。本发明进一步提供了可由所述细胞制造的具有类人多糖结构的蛋白质及其新组合物。



1. 一种被修饰以表达脂质连接寡糖 (LLO) 翻转酶活性物的细胞, 所述脂质连接寡糖翻转酶活性物能够从细胞器的胞质面向其腔面高效地翻转包含 1 个甘露糖残基的脂质连接寡糖, 能够高效地翻转包含 2 个甘露糖残基的脂质连接寡糖, 并且能够高效地翻转包含 3 个甘露糖残基的脂质连接寡糖。

2. 如权利要求 1 所述的细胞, 其特征在于, 所述脂质连接寡糖翻转酶具有翻转选自 Man1GlcNAc2, Man2GlcNAc2 和 Man3GlcNAc2 的脂质连接寡糖的活性。

3. 如前述权利要求中任一项所述的细胞, 其中, 所述脂质连接寡糖翻转酶活性物通过一种或多种核酸分子的表达来提供, 所述核酸分子选自:

a) 包含下述一个或多个序列的核酸分子: SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, 和 SEQ ID NO:17; SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, 和 SEQ ID NO:29;

b) 包含下述一个或多个序列的编码聚氨基酸的核酸分子: SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14; SEQ ID NO:16 和 SEQ ID NO:18; SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, 和 SEQ ID NO:30; 及

c) a) 或 b) 所述的核酸分子的片段、变体、类似物或衍生物。

4. 如前述权利要求中任一项所述的细胞, 其中, 所述的细胞器为内质网 (ER)。

5. 如前述权利要求中任一项所述的细胞, 其特征在于, 所述细胞包含至少一种编码异源 (糖) 蛋白的核酸, 并表达所述 (糖) 蛋白。

6. 如前述权利要求中任一项所述的细胞, 其特征在于, 所述细胞缺乏或具有被抑制、减少或耗竭的 Rft1 型脂质连接寡糖翻转酶活性物, 其中, 所述 Rft1 型脂质连接寡糖翻转酶活性物的特征在于, 对带有 1 至 3 个甘露糖残基的脂质连接寡糖的翻转活性小于对带有 5 个甘露糖残基的脂质连接寡糖的翻转活性。

7. 如权利要求 6 所述的细胞, 其中, 所述细胞为 rft1 基因或 rft1 同源基因的敲除突变株。

8. 如前述权利要求中任一项所述的细胞, 其特征在于, 所述细胞缺乏或具有被抑制、减少或耗竭的一种或多种定位于内质网的糖基转移酶活性物。

9. 如权利要求 8 所述的细胞, 其特征在于, 所述定位于内质网的糖基转移酶为甘露糖基转移酶。

10. 如前述权利要求中任一项所述的细胞, 其特征在于, 所述细胞缺乏或具有被抑制、减少或耗竭的一种或多种定位于内质网的脂质连接单糖 (LLM) 翻转酶活性物。

11. 如前述权利要求中任一项所述的细胞, 其特征在于, 所述细胞缺乏或具有被抑制、减少或耗竭的 Alg11 型活性物。

12. 如权利要求 11 所述的细胞, 其中, 所述细胞为 alg11 基因或 alg11 同源基因的敲除突变株。

13. 如前述权利要求中任一项所述的细胞, 其特征在于, 所述细胞缺乏或具有被抑制、减少或耗竭的 Alg11 型活性物, 并且进一步缺乏或具有被抑制、减少或耗竭的一种或多种脂质连接单糖 (LLM) 翻转酶型活性物。

14. 如权利要求 13 所述的细胞, 其中, 所述细胞为 alg11 基因或 alg11 同源基因及一种

或多种编码脂质连接单糖 (LLM) 翻转酶活性物的基因的敲除突变株。

15. 如前述权利要求中任一项所述的细胞,其特征在於,所述细胞缺乏或具有被抑制、减少或耗竭的 Alg11 型活性物,并且进一步缺乏或具有被抑制、减少或耗竭的 Alg3 型活性物。

16. 如权利要求 16 所述的细胞,其中,所述细胞为 alg11 基因或 alg11 同源基因及 alg3 基因或 alg3 同源基因的敲除突变株。

17. 如前述权利要求中任一项所述的细胞,其特征在於,所述细胞缺乏或具有被抑制、减少或耗竭的 Alg11 型活性物,并且进一步缺乏或具有被抑制、减少或耗竭的 β -D-甘露糖基转移酶或 DPM1 型活性物。

18. 如权利要求 17 所述的细胞,其中,所述细胞为 alg11 基因或 alg11 同源基因及 dpm1 基因或 dpm1 同源基因的敲除突变株。

19. 如前述权利要求中任一项所述的细胞,其特征在於,所述细胞缺乏或具有被抑制、减少或耗竭的 Alg2 型活性物。

20. 如权利要求 19 所述的细胞,其中,所述细胞为 alg2 基因或 alg2 同源基因的敲除突变株。

21. 如前述权利要求中任一项所述的细胞,其特征在於,所述细胞包含一种或多种编码寡糖基转移酶活性物的核酸分子,所述寡糖基转移酶活性物的特征在於,其能够转移除 Glc3Man9GlcNAc2 之外的寡糖至蛋白质。

22. 如权利要求 21 所述的细胞,其特征在於,所述寡糖基转移酶活性物为原生动物的寡糖基转移酶 (POT) 活性物。

23. 如权利要求 22 所述的细胞,其特征在於,所述原生动物的寡糖基转移酶活性物来源于弓形虫 (*Toxoplasma gondii* (Tg)), 硕大利什曼虫 (*Leishmania major* (Lm)); 婴儿利什曼虫 (*Leishmania infantum* (Li)), 巴西利什曼虫 (*Leishmania braziliensis* (Lb)), 墨西哥利什曼虫 (*Leishmania Mexicana* (Lmx)), 杜氏利什曼虫 (*Leishmania donovani* (Ld)), 利什曼虫圭亚那亚种 (*Leishmania guyanensis* (Lg)), 热带利什曼虫 (*Leishmania tropica* (Lt)), 克氏锥虫 (*Trypanosoma cruzi* (Tc)), 或布氏锥虫 (*Trypanosoma brucei* (Tb))。

24. 如权利要求 23 所述的细胞,其特征在於,所述原生动物的寡糖基转移酶活性物选自 TbStt3Bp 型活性物, TbStt3Cp 型活性物, LmStt3Ap 型活性物, LmStt3Bp 型活性物, 及 LmStt3Dp 型活性物。

25. 如权利要求 23 所述的细胞,其特征在於,所述原生动物的寡糖基转移酶活性物选自 TbStt3B 型活性物, TbStt3C 型活性物; LmStt3A 型活性物, LmStt3B 型活性物, LmStt3C 型活性物, 及 LmStt3D 型活性物; LiStt3-1 型活性物, LiStt3-2 型活性物, 及 LiStt3-3 型活性物; LbStt3-1 型活性物, LbStt3-2 型活性物, 及 LbStt3-3 型活性物。

26. 一种被修饰以表达寡糖基转移酶活性物的细胞,所述寡糖基转移酶活性物能够高效地转移包含 3 个甘露糖残基、4 个甘露糖残基和 / 或 5 个甘露糖残基的寡糖至蛋白质。

27. 如权利要求 26 所述的细胞,其特征在於,所述寡糖基转移酶活性物为单元原生动物的寡糖基转移酶 (POT)。

28. 如权利要求 27 所述的细胞,其特征在於,所述原生动物的寡糖基转移酶活性物来源

于弓形虫 (*Toxoplasma gondii* (Tg)), 硕大利什曼虫 (*Leishmania major* (Lm)); 婴儿利什曼虫 (*Leishmania infantum* (Li)), 巴西利什曼虫 (*Leishmania braziliensis* (Lb)), 墨西哥利什曼虫 (*Leishmania Mexicana* (Lmx)), 杜氏利什曼虫 (*Leishmania donovani* (Ld)), 利什曼虫圭亚那亚种 (*Leishmania guyanensis* (Lg)), 热带利什曼虫 (*Leishmania tropica* (Lt)), 克氏锥虫 (*Trypanosoma cruzi* (Tc)), 或布氏锥虫 (*Trypanosoma brucei* (Tb))。

29. 如权利要求 28 所述的细胞, 其特征在于, 所述原生动物寡糖基转移酶活性物选自 TbStt3Bp 型活性物, TbStt3Cp 型活性物, LmStt3Ap 型活性物, LmStt3Bp 型活性物, 及 LmStt3Dp 型活性物。

30. 如权利要求 28 所述的细胞, 其特征在于, 所述原生动物寡糖基转移酶选自 TbStt3B 型活性物, TbStt3C 型活性物; LmStt3A 型活性物, LmStt3B 型活性物, LmStt3C 型活性物, 及 LmStt3D 型活性物; LiStt3-1 型活性物, LiStt3-2 型活性物, 及 LiStt3-3 型活性物; LbStt3-1 型活性物, LbStt3-2 型活性物, 及 LbStt3-3 型活性物。

31. 如权利要求 26-30 中任一权利要求所述的细胞, 其特征还在于, 所述细胞缺乏或具有被抑制、减少或耗竭的一种或多种定位于内质网的糖基转移酶活性物。

32. 如权利要求 31 所述的细胞, 其特征在于, 所述定位于内质网的糖基转移酶为甘露糖基转移酶。

33. 如权利要求 26-32 中任一权利要求所述的细胞, 其特征还在于, 所述细胞缺乏或具有被抑制、减少或耗竭的一种或多种定位于内质网的脂质连接单糖 (LLM) 翻转酶活性物。

34. 如权利要求 26-33 中任一权利要求所述的细胞, 其特征还在于, 所述细胞缺乏或具有被抑制、减少或耗竭的 Alg11 型活性物。

35. 如权利要求 34 所述的细胞, 其中, 所述细胞为 alg11 基因或 alg11 同源基因的敲除突变株。

36. 如权利要求 26-35 中任一权利要求所述的细胞, 其特征还在于, 所述细胞缺乏或具有被抑制、减少或耗竭的 Alg11 型活性物, 并且进一步缺乏或具有被抑制、减少或耗竭的一种或多种脂质连接单糖 (LLM) 翻转酶型活性物。

37. 如权利要求 36 所述的细胞, 其中, 所述细胞为 alg11 基因或 alg11 同源基因及一个或多个编码脂质连接单糖 (LLM) 翻转酶活性物的基因的敲除突变株。

38. 如权利要求 26-37 中任一权利要求所述的细胞, 其特征还在于, 所述细胞缺乏或具有被抑制、减少或耗竭的 Alg11 型活性物, 并且进一步缺乏或具有被抑制、减少或耗竭的 Alg3 型活性物。

39. 如权利要求 38 所述的细胞, 其中, 所述细胞为 alg11 基因或 alg11 同源基因及 alg3 基因或 alg3 同源基因的敲除突变株。

40. 如权利要求 26-39 中任一权利要求所述的细胞, 其特征还在于, 所述细胞缺乏或具有被抑制、减少或耗竭的 Alg11 型活性物, 并且进一步缺乏或具有被抑制、减少或耗竭的 β -D-甘露糖基转移酶或 DPM 型活性物 1。

41. 如权利要求 40 所述的细胞, 其中, 所述细胞为 alg11 基因或 alg11 同源基因及 dpm1 基因或 dpm1 同源基因的敲除突变株。

42. 如权利要求 26-41 中任一权利要求所述的细胞, 其特征还在于, 所述细胞显示出

Rft1 型脂质连接寡糖翻转酶活性物。

43. 如权利要求 42 所述的细胞,所述细胞与野生型细胞相比,超表达 Rft1 型活性物。

44. 如前述权利要求中任一项所述的细胞,其特征在于,所述细胞缺乏或具有被抑制、减少或耗竭的一种或多种定位于高尔基体的甘露糖基转移酶活性物。

45. 如权利要求 44 所述的细胞,其特征在于,所述定位于高尔基体的甘露糖基转移酶活性物选自 Och1 型活性物,Mnn1 型活性物,Mnn2 型活性物,Mnn4 型活性物,Mnn5 型活性物,Mnn9 型活性物,Mnn10 型活性物,及 Mnn11 型活性物。

46. 如权利要求 45 所述的细胞,其中,所述细胞为选自 och1, mnn1, mnn2, mnn4, mnn5, mnn9, mnn10, mnn11, 及它们的同源基因中的至少一种基因的敲除突变株。

47. 如权利要求 44 所述的细胞,其特征在于,所述细胞缺乏或具有被抑制、减少或耗竭的 Mnn1 型活性物。

48. 如权利要求 47 所述的细胞,其中,所述细胞为 mnn1 基因和 / 或 mnn1 同源基因的敲除突变株。

49. 如权利要求 44 所述的细胞,其特征在于,所述细胞缺乏或具有被抑制、减少或耗竭的 Och1 型活性物。

50. 如权利要求 49 所述的细胞,其中,所述细胞为 och1 基因或 och1 同源基因的敲除突变株。

51. 如前述权利要求中任一项所述的细胞,其特征在于,所述细胞表达一种或多种定位于高尔基体的异源酶或其催化域,所述定位于高尔基体的异源酶或其催化域选自:

甘露糖基 (α -1,3)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶 (GnTI);

甘露糖基 (α -1,6)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶 (GnTII);

β -1,4-甘露糖基-糖蛋白 4- β -N-乙酰葡萄糖氨基转移酶 (GnTIII);

甘露糖基 (α -1,3)-糖蛋白 β -1,4-N-乙酰葡萄糖氨基转移酶 (GnTIV);

甘露糖基 (α -1,6)-糖蛋白 β -1,6-N-乙酰葡萄糖氨基转移酶 (GnTV);

甘露糖基 (α -1,6)-糖蛋白 β -1,4-N-乙酰葡萄糖氨基转移酶 (GnTVI);

β -N-乙酰葡萄糖氨基糖肽 β -1,4-半乳糖基转移酶 (GalT);

α (1,6) 岩藻糖基转移酶 (FucT);

β -半乳糖苷 α -2,6-唾液酸转移酶 (ST);

UDP-N-乙酰葡萄糖胺 2-差向异构酶 (NeuC);

唾液酸合酶 (NeuB);

CMP-Neu5Ac 合酶;

N-酰基神经氨酸-9-磷酸合酶;

N-酰基神经氨酸-9-磷酸酶;

UDP-N-乙酰葡萄糖胺转运体;

UDP-半乳糖转运体;

GDP-岩藻糖转运体;

CMP-唾液酸转运体;

核苷酸二磷酸酶;

GDP-D-甘露糖 4,6-脱水酶;及

GDP-4- 酮 - 去氧 -D- 甘露糖 -3,5- 差向异构酶 -4- 还原酶。

52. 如前述权利要求中任一项所述的细胞,其特征在于,所述细胞选自:包括真菌细胞的低级真核细胞及包括哺乳动物细胞、哺乳动物细胞系、植物细胞和昆虫细胞的高级真核细胞。

53. 一种或多种分离的核酸分子,其能够编码或提供权利要求 1 或 2 所述的脂质连接寡糖翻转酶活性物。

54. 如权利要求 53 所述的核酸分子,其特征在于,所述分子选自权利要求 3 所述的核酸分子中的一种或多种。

55. 一种用于在真核宿主细胞中表达的表达盒,包含权利要求 53 或 54 所述的任一种核酸分子的一个或多个拷贝,所述一个或多个拷贝与编码启动子的核酸分子和编码终止子的核酸分子中的至少一种核酸分子结合。

56. 如权利要求 55 所述的表达盒,还包含编码权利要求 21-25 中任一项所述的寡糖基转移酶活性物的核酸分子的一个或多个拷贝。

57. 一种用于真核宿主细胞转化的载体,包含权利要求 53 或 54 所述的任一种核酸分子的一个或多个拷贝和 / 或权利要求 55 或 56 所述的表达盒的一个或多个拷贝。

58. 一种用于制造细胞的方法,所述细胞具体能够在所述细胞的细胞器中合成具有 Man3GlcNAc2 多糖结构的脂质连接寡糖,所述方法包括下述步骤:

用至少一种编码脂质连接寡糖翻转酶活性物的构造或结构转化所述细胞,所述构造或结构选自:

权利要求 53 或 54 所述的核酸分子;

权利要求 55 或 56 所述的表达盒;及

权利要求 57 所述的载体;

以使得所述细胞能够表达由所述构造或结构编码的脂质连接寡糖翻转酶活性物。

59. 如权利要求 58 所述的方法,其中,所述构造或结构还编码寡糖基转移酶活性物,并选自:

权利要求 56 所述的表达盒,及

包含权利要求 57 所述的表达盒的一个或多个拷贝的载体,

以使得所述细胞能够表达由所述构造或结构编码的脂质连接寡糖翻转酶活性物和寡糖基转移酶活性物。

60. 如权利要求 58 或 59 所述的方法,还包括步骤:

在所述细胞内减少或耗竭至少一种酶活性物,所述酶活性物选自:

Alg2 型活性物;

Alg11 型活性物;

Alg3 型活性物;

DPM1 型活性物;

脂质连接单糖 (LLM) 翻转酶型活性物。

61. 一种或多种分离的细胞,所述细胞能够在细胞器中合成具有 Man1GlcNAc2, Man2GlcNAc2 和 / 或 Man3GlcNAc2 多糖结构的脂质连接寡糖,并且能够转移所述多糖结构至在所述细胞中表达的新生蛋白质,其特征在于,所述细胞根据权利要求 58-60 中任一项所

述的方法来制造。

62. 一种用于制造糖蛋白或糖蛋白组合物的方法,包括如下步骤:

提供权利要求 1-52 或权利要求 61 中任一项所述的细胞;

在所述细胞内允许制造所述糖蛋白或糖蛋白组合物的条件下,于培养基中培养所述细胞;及

必要时,从所述细胞和/或所述培养基中分离所述糖蛋白或糖蛋白组合物。

63. 一种用于制造糖蛋白或糖蛋白组合物的试剂盒,包括:

权利要求 1-52 或权利要求 61 中任一项所述的细胞;及

培养基,用于培养所述细胞以制造所述糖蛋白。

64. 如权利要求 1-52 或权利要求 61 中任一项所述的细胞,其具体经设计用来制造具有 GlcNAcMan3-5GlcNAc2 结构的糖蛋白。

65. 一种具有 GlcNAcMan3-5GlcNAc2 多糖结构的糖蛋白,由权利要求 45 所述的细胞制造。

66. 如权利要求 1-52 或权利要求 61 中任一项所述的细胞,其具体经设计用来制造具有 GlcNAc2Man3GlcNAc2 结构的糖蛋白。

67. 一种具有 GlcNAc2Man3GlcNAc2 多糖结构的糖蛋白,由权利要求 66 所述的细胞制造。

68. 如权利要求 1-52 或权利要求 61 中任一项所述的细胞,其具体经设计用来制造具有 GlcNAc3Man3GlcNAc2- 平分型多糖结构的糖蛋白。

69. 一种具有 GlcNAc3Man3GlcNAc2- 平分型多糖结构的糖蛋白,由权利要求 68 所述的细胞制造。

70. 如权利要求 1-52 或权利要求 61 中任一项所述的细胞,其具体经设计用来制造具有 Gal2GlcNAc2Man3GlcNAc2 多糖结构的糖蛋白。

71. 一种具有 Gal2GlcNAc2Man3GlcNAc2 多糖结构的糖蛋白,由权利要求 70 所述的细胞制造。

72. 一种具有 Gal2GlcNAc2Man3GlcNAc2 或 Gal1GlcNAc2Man3GlcNAc2 多糖结构的糖蛋白组合物,由权利要求 70 所述的一种或多种细胞制造。

73. 如权利要求 1-52 或权利要求 61 中任一项所述的细胞,其具体经设计用来制造具有 Gal2GlcNAc2Man3GlcNAc2Fuc 多糖结构的糖蛋白。

74. 一种具有 Gal2GlcNAc2Man3GlcNAc2Fuc 多糖结构的糖蛋白,由权利要求 73 所述的细胞制造。

75. 如权利要求 1-52 或权利要求 61 中任一项所述的细胞,其具体经设计用来制造具有 Gal2GlcNAc3Man3GlcNAc2- 平分型多糖结构的糖蛋白。

76. 一种具有 Gal2GlcNAc3Man3GlcNAc2- 平分型多糖结构的糖蛋白,由权利要求 75 所述的细胞制造。

77. 如权利要求 1-52 或权利要求 61 中任一项所述的细胞,其具体经设计用来制造具有 Gal2GlcNAc3Man3GlcNAc2Fuc- 平分型多糖结构的糖蛋白。

78. 一种具有 Gal2GlcNAc3Man3GlcNAc2Fuc- 平分型多糖结构的糖蛋白,由权利要求 77 所述的细胞制造。

79. 如权利要求 1-52 或权利要求 61 中任一项所述的细胞,其具体经设计用来制造具有 NeuAc2Gal2GlcNAc2Man3GlcNAc2 多糖结构的糖蛋白。

80. 一种具有 NeuAc2Gal2GlcNAc2Man3GlcNAc2 多糖结构的糖蛋白,由权利要求 79 所述的细胞制造。

81. 如权利要求 1-52 或权利要求 61 中任一项所述的细胞,其具体经设计用来制造具有 NeuAc2Gal2GlcNAc2Man3GlcNAc2Fuc 多糖结构的糖蛋白。

82. 一种具有 NeuAc2Gal2GlcNAc2Man3GlcNAc2Fuc 多糖结构的糖蛋白,由权利要求 81 所述的细胞制造。

83. 如权利要求 1-52 或权利要求 61 中任一项所述的细胞,其具体经设计用来制造具有 NeuAc2Gal2GlcNAc3Man3GlcNAc2- 平分型多糖结构的糖蛋白。

84. 一种具有 NeuAc2Gal2GlcNAc3Man3GlcNAc2- 平分型多糖结构的糖蛋白,由权利要求 83 所述的细胞制造。

85. 如权利要求 1-52 或权利要求 61 中任一项所述的细胞,其具体经设计用来制造具有 NeuAc2Gal2GlcNAc3Man3GlcNAc2Fuc- 平分型多糖结构的糖蛋白。

86. 一种具有 NeuAc2Gal2GlcNAc3Man3GlcNAc2Fuc- 平分型多糖结构的糖蛋白,由权利要求 85 所述的细胞制造。

87. 如权利要求 1-52 或权利要求 61 中任一项所述的细胞,其具体经设计用来制造具有 GlcNAc3Man3GlcNAc2 多糖结构的糖蛋白。

88. 一种具有 GlcNAc3Man3GlcNAc2 多糖结构的糖蛋白,由权利要求 87 所述的细胞制造。

89. 如权利要求 1-52 或权利要求 61 中任一项所述的细胞,其具体经设计用来制造具有 Gal3GlcNAc3Man3GlcNAc2 多糖结构的糖蛋白。

90. 一种具有 Gal3GlcNAc3Man3GlcNAc2 多糖结构的糖蛋白,由权利要求 89 所述的细胞制造。

91. 如权利要求 1-52 或权利要求 61 中任一项所述的细胞,其具体经设计用来制造具有 NeuAc3Gal3GlcNAc3Man3GlcNAc2 多糖结构的糖蛋白。

92. 一种具有 NeuAc3Gal3GlcNAc3Man3GlcNAc2 多糖结构的糖蛋白,由权利要求 91 所述的细胞制造。

93. 如权利要求 1-52 或权利要求 61 中任一项所述的细胞,其具体经设计用来制造具有 NeuAc3Gal3GlcNAc3Man3GlcNAc2Fuc 多糖结构的糖蛋白。

94. 一种具有 NeuAc3Gal3GlcNAc3Man3GlcNAc2Fuc 多糖结构的糖蛋白,由权利要求 93 所述的细胞制造。

95. 一种或多种分离的糖蛋白,选自下述一种或多种:

由权利要求 1-52 或权利要求 61 中任一项所述的细胞制造的糖蛋白;

可利用权利要求 62 所述的方法制造的糖蛋白;及

权利要求 63,67,69,71,72,74,76,78,80,82,84,86,88,90,92,及 94 中任一项所述的糖蛋白。

96. 一种糖蛋白组合物,包含两种或多种不同的如权利要求 95 所述的糖蛋白。

97. 一种或多种重组的治疗用蛋白,其如权利要求 95 或 96 所述。

98. 一种或多种免疫球蛋白,其如权利要求 96 或 97 所述。

99. 一种药物组合物,包含权利要求 95-98 中任一项或多项所述的糖蛋白或糖蛋白组合物中的一种或多种,并且包含至少一种药学上可接受的载体或辅药。

100. 如权利要求 95-99 中任一权利要求所述的糖蛋白或糖蛋白组合物,应用在可通过施用所述糖蛋白或糖蛋白组合物来治疗的疾病的治疗方法中。

101. 一种治疗失调的方法,所述失调可通过施用权利要求 95-99 中任一项或多项所述的糖蛋白或糖蛋白组合物中的一种或多种来治疗,所述方法包括下述步骤:

给对象施用如上所述的糖蛋白或糖蛋白组合物,其中,所述对象患有或疑似患有可通过施用所述糖蛋白或糖蛋白组合物治疗的疾病。

在宿主细胞中制造糖基化蛋白的新方法

技术领域

[0001] 本发明涉及真核生物中糖蛋白制造及蛋白糖基化工程的领域,特别是涉及类人复合物或杂交糖基化蛋白在例如酵母等低级真核生物中的制造。本发明进一步涉及经糖基化修饰的真核宿主细胞,其能够制造作为免疫球蛋白及其他治疗用蛋白对人类特别有效的糖基化优化蛋白。本发明还涉及基因工程改造的真核非人细胞,其能够制造与人细胞中制造的糖蛋白类似的具有多糖结构的糖蛋白。因此,本发明进一步涉及可由所述细胞制造的具有类人多糖结构的蛋白质及其新组合物。

背景技术

[0002] 大多数基于蛋白质的生物药物会暴露出一些可以对其治疗应用有关的蛋白质特性产生极大影响的翻译后修饰的形态。蛋白质糖基化是最普通的修饰(大约50%的人蛋白质是糖基化蛋白质)。糖基化可以通过在蛋白质组合物中的蛋白质上产生不同的多糖结构而向该组合物中引入重要的异质性。这样的多糖结构通过糖基化系统的各种酶的作用来形成,随着糖蛋白运输内质网(ER)和高尔基体(糖基化级联)。蛋白质多糖结构的性质对蛋白质的折叠、稳定、寿命、运输、药效学、药物动力学及免疫原性均产生影响。多糖结构对蛋白质的初级功能活性有很大影响。糖基化可以影响蛋白质局部结构并可能帮助引导多肽链的折叠。一种重要的多糖结构就是所谓的N-多糖。N-多糖通过寡糖共价键合新生多肽链共有序列NXS/T中的天冬酰胺残基的氨基(N)而产生。N-多糖可能进一步参与分选或引导蛋白质到达其最终靶点:抗体的N-多糖,例如,可能与补体成分相互作用。N-多糖还可以稳定糖蛋白,例如,可通过提高其溶解度、在其表面屏蔽疏水位点、防止蛋白水解及引导链内稳定化作用。糖基化可调节蛋白质的半衰期,例如,在人体内,N-多糖的末端唾液酸的存在可通过在血流中循环而增加蛋白质的半衰期。

[0003] 寡糖的合成发生在内质网膜的两侧。糖基化级联起始于内质网膜的胞质面上连接有脂质的寡糖(lipid-linked oligosaccharide;LLO)的生成。首先,合成一个具有确定结构(Man3GlcNAc2)的与脂质连接的核心寡糖,然后添加所述寡糖至胞质面上的脂质多萜醇连接的Man3GlcNAc2以形成低聚七糖Man5GlcNAc2多糖结构。随后将该LLO转移(翻转)到内质网的腔面,进一步地,在该处加工所述庚-寡多糖以形成包含3个葡萄糖、9个甘露糖和2个N-乙酰葡萄糖胺残基(Glc3Man9GlcNAc2)结构的支链寡糖单元。所述Glc3Man9GlcNAc2结构通过若干糖基转移酶的作用而形成。每个单独的糖基转移酶都对特定的寡糖底物显示出很强的优先选择性,这导致可以基本线性、逐步地生物合成该支链寡糖。然后所述Glc3Man9GlcNAc2结构从多萜醇脂质转移至新生多肽。图1示出了野生型酵母的内质网上的LLO操作过程。

[0004] 所述内质网糖基化途径的两个步骤不与糖基转移酶的作用直接相关:(1)Man5GlcNAc2-LLO从内质网膜的胞质面向腔面的翻转和(2)从脂质连接物到新生多肽的寡糖基转移。

[0005] 翻转是在一种不依赖于ATP的双向翻转酶的催化下完成的。在酵母中,翻转酶活

性是由一种包含约 10 个横跨内质网膜的跨膜域的多起源膜蛋白即“Rft1”支持或给予的。同源蛋白质的基因也出现在其他真核生物的基因组中。

[0006] 不希望受理论的束缚,完整的寡糖 Glc3Man9GlcNAc2 是寡糖基转移酶 (OT 或 OST) 最佳的底物,所述寡糖基转移酶将全部寡糖从供体 LL0 转移至新生蛋白质或多肽的 Asn-X-Ser/Thr 共有序列中选定的天冬酰胺残基的氨基上。在大部分生物中,寡糖基转移酶是一种包含七或八个不同蛋白质的多聚体复合物,其中一个蛋白质 (Stt3p) 是催化性亚基。一旦糖蛋白适当地完成折叠和寡聚体化,就会移动到高尔基复合体中。然后,进一步修剪和修饰 N- 连接多糖,并且添加新的糖类例如在人细胞中生成杂交型或复合型多糖。

[0007] 糖基转移酶和糖苷酶布满内质网和高尔基体的内 (腔) 表面,因此在蛋白质通过内质网和高尔基网时提供了一个允许糖蛋白序列加工的催化表面。事实上,顺面、中间和反面高尔基体的多区室结构及反面高尔基网 (TGN) 提供了可以产生糖基化反应的有序序列的不同场所。随着糖蛋白从在内质网中合成到在晚期高尔基体或 TGN 中完全成熟,其连续暴露于不同的糖苷酶、甘露糖苷酶和糖基转移酶,使得可以合成特定的寡糖多糖结构。

[0008] 不同的生物提供不同的糖基化酶 (糖基转移酶和糖苷酶)。因此,蛋白质的多糖结构的最终组成可根据宿主的不同而发生明显变化。例如,典型地,例如酵母和丝状真菌等低级真核生物在高尔基体中添加大量甘露糖残基而产生“高甘露糖”型糖蛋白;然而,在哺乳动物细胞中,多糖结构可能在高尔基体中经过修剪而去除若干个九甘露糖残基,并且可通过附加的糖残基进一步延长,这在低级真核生物的 N- 多糖例如唾液酸或岩藻糖中并不典型。

[0009] 制造重组蛋白的可能性已经彻底改变了对患有多种不同疾病的患者的治疗。大部分治疗用蛋白需要通过添加多糖结构来进行修饰。所述糖基化对于蛋白质的正确折叠、长循环及最佳活性 (大多数情况下) 可能是必要的。哺乳动物细胞如常用的中国仓鼠卵巢细胞 (CHO 细胞) 可以产生与人多糖结构类似的复合多糖结构,然而,来源于例如 CHO 细胞的多糖结构与人源多糖结构不同,因为 CHO 细胞 a) 唾液酸化程度较低,b) 额外地使寡糖与普通唾液酸 (NeuAc) 合并成另一种非人唾液酸 (NeuGc),及 c) 包括人细胞中不存在的末端结合的 α -1-3 半乳糖。现在使用的用于制造重组蛋白的哺乳动物表达系统的缺点在于:(1) 制造率低,(2) 发酵程序昂贵,(3) 菌株设计复杂,及 (4) 存在病毒污染的风险。

[0010] 与哺乳动物细胞相反,酵母细胞是用于工业发酵的强健生物体,而且可以在成分明确的培养基中进行高密度培养。虽然糖基化在酵母和真菌中与在哺乳动物和人有很大不同,但一些常见的要素是可以共享的。第一步,LL0 在内质网中向新生蛋白质的转移在包括酵母、真菌、植物和人类的所有真核生物中是高度保存的。但是,随后 N- 多糖在高尔基体中的加工过程在酵母和哺乳动物中有显著不同。在酵母中,该过程包括若干甘露糖的添加。所述甘露糖基化由残留于高尔基体 (例如 Och1, Mnn1, Mnn2, 等) 中的甘露糖基转移酶催化,连续向 N- 多糖添加甘露糖。

[0011] 对于生物制药产业来说,采用一种具有重现性和一致性的糖式模型来制造治疗用蛋白仍然是一个相当大的挑战。特别是,在酵母中制造的治疗用糖蛋白可能会在高级真核生物尤其是动物和人中引起不必要的免疫应答,从而导致在酵母及其同类中制造的治疗用蛋白的治疗价值低。对于几种重要的治疗剂类别,包括血液因子、抗凝血剂、溶血栓药、抗体、激素、刺激因子及细胞因子,例如肿瘤坏死因子 (TNF) 家族的调节蛋白、促红细胞生成

素 (EPO)、促性腺激素、免疫球蛋白 G (IgG)、粒细胞-巨噬细胞集落刺激因子及干扰素,均已观察到糖基化对若干治疗用蛋白的分泌、稳定性、免疫原性及活性会产生影响。

[0012] 最近正在开发例如毕赤酵母 (*Pichia pastoris*)、解脂耶氏酵母 (*Yarrowia lipolytica*) 及酿酒酵母 (*Saccharomyces cerevisiae*) 等许多酵母以发挥此类系统的优点而消除其在糖基化方面的缺点。已对若干菌株进行基因组学研究以在蛋白质上产生明确的类人多糖结构。

发明内容

[0013] 本发明的目的在于提供制造例如脂质和蛋白质等糖基化分子,特别是重组糖蛋白并且优选实例是免疫球蛋白的手段及方法。本发明的目的还在于提供一种可采用所述手段和方法制造的、具有明确的多糖结构例如特别是类人或杂交或复合多糖结构的糖蛋白及其新组合物。本发明的一个特别目的在于提供具有类人多糖结构的 N-糖基化蛋白特别是免疫球蛋白,其可用于治疗人类疾病,疗效高并且不会引起不必要的副作用。

[0014] 本发明所基于的技术问题是主要通过提供一种新的与脂质连接的寡糖(以下有时简称为脂质连接寡糖,缩写为 LLO) 翻转酶活性物 (LLO 翻转酶活性) 而得到解决的。所述新的翻转酶活性物主要特征在于,其能够高效地翻转含有包含一个甘露糖残基的多糖结构特别是 Man1GlcNAc2 的 LLO;能够高效地翻转含有包含两个甘露糖残基的多糖结构特别是 Man2GlcNAc2 的脂质连接寡糖;还能够高效地并且尤其高活性地翻转含有包含三个甘露糖残基的多糖结构特别是 Man3GlcNAc2 的脂质连接寡糖。

[0015] 本发明提供了一种与已知的翻转酶活性物特别是 Rft-1 型活性物相反的新型“LLO 翻转酶活性物”,其对于所要被翻转的寡糖基结构显示出“松弛”的特异性。不希望受理论的束缚,已知的翻转酶,例如低级真核生物的之前已经被表征的翻转酶活性物,对于要被翻转的脂质连接寡糖的特定多糖结构显示出很高的特异性。更特别地, Rft-1 型活性物(同义词: YBL020W; Man5GlcNAc2-PP-Do1 翻转酶) 主要能够翻转包含五个甘露糖残基特别是 Man5GlcNAc2 多糖结构的脂质连接寡糖,但基本不能翻转包含 Man1GlcNAc2 多糖结构的脂质连接寡糖。

[0016] 本文中所述术语“高效”主要是指酶活性或转移活性以足以实现本发明所述的范围和目标内的宿主细胞的技术目的的数量或比例呈现。例如,据认为,“高效”的转移或合成不是类似于或反映为了制造本发明所述的糖蛋白而在宿主细胞中进行的酶合成级步骤中的化合物流动中的主要限速步骤。

[0017] 本发明还通过提供一种修饰或基因工程细胞或宿主细胞,特别是一种包含和表达所述新型脂质连接寡糖翻转酶活性物的真核细胞,解决了其潜在的技术问题。

[0018] 发明人惊奇地发现,有可能提供所述对于要被翻转的脂质连接寡糖的多糖结构具有松弛的特异性的新型“LLO 翻转酶”。该新型 LLO 翻转酶有利地使得能够在细胞内的细胞器官膜特别是内质网膜上进行糖基化加工的基因工程。

[0019] 根据本发明的第一个方面,提供了一种具有松弛的特异性的新型 LLO 翻转酶,其可用作一种在宿主细胞中修饰和控制糖基化的有价值的手段。在优选实施方案中,将所述宿主细胞的修饰与在膜的胞质面和/或细胞器的腔面构建 LLO 结构的过程中的至少一种或多种遗传修饰相结合(见图 1)。

[0020] 在更优选的实施方案中,在所述构建过程末期,由细胞器介导的向新生多肽进行寡糖基转移时,这些修饰进一步与寡糖基转移酶的遗传修饰相结合。这些修饰的复合体系有利地使得能够提供新修饰的宿主细胞,其尤其能够在细胞内的细胞器更特别是内质网中,具体地合成包含 1、2 或 3 个甘露糖残基的多糖结构,特别是 Man1GlcNAc2, Man2GlcNAc2 或 Man3GlcNAc2。

[0021] 在本发明的一个优选方面,所述细胞被进一步修饰来使其缺乏或具有被抑制、减少或耗尽的一个或多个定位于细胞器或内质网的糖基转移酶活性物,特别是甘露糖基转移酶活性物,并且更特别是来替换表达异源糖基转移酶活性物及其他为蛋白质的杂交或复合 N-糖基化所必需的酶。

[0022] 根据本发明的第二个方面,提供了一种细胞,该细胞被替代性地或附加性地修饰以包含或表达一个或多个定位于细胞器或内质网的修饰、特别是异源的寡糖基转移酶 (OT) 活性物,所述寡糖基转移酶 (OT) 活性物对于被从 LLO 转移到蛋白质的多糖结构具有松弛的特异性。特别地,此类寡糖基转移酶对 Man1GlcNAc2, Man2GlcNAc2 或 Man3GlcNAc2 多糖结构的转移活性高。在本文中,术语“高”意味着 Man1GlcNAc2, Man2GlcNAc2 或 Man3GlcNAc2 将被转移至新生蛋白质中的至少 20%,至少 40%,至少 60%,优选至少 80%,最优选至少 90%。所述细胞的特征可以进一步在于:所述细胞包含一个或多个编码寡糖基转移酶活性物的核酸分子,特征在于所述寡糖基转移酶活性物不优先转移 Glc3Man9GlcNAc2 到蛋白质,但能够转移除 Glc3Man9GlcNAc2 之外的寡糖到蛋白质,优选带有 1 到 9 个甘露糖残基的寡糖,最优选为 Man1GlcNAc2, Man2GlcNAc2 或 Man3GlcNAc2。更特别地,所述细胞的特征在于,所述寡糖基转移酶活性物不仅转移 Glc3Man9GlcNAc2 到蛋白质,还能够高效地转移除 Glc3Man9GlcNAc2 之外的寡糖到蛋白质,优选带有 1 到 9 个甘露糖残基的寡糖 (Man1GlcNAc2, Man2GlcNAc2, Man3GlcNAc2, Man4GlcNAc2, Man5GlcNAc2, Man6GlcNAc2, Man7GlcNAc2, Man8GlcNAc2, Man9GlcNAc2),最优选为 Man1GlcNAc2, Man2GlcNAc2 和 / 或 Man3GlcNAc2。

[0023] 更特别地,所述寡糖基转移酶 (OT) 活性物是一个单独的单位或原生型寡糖基转移酶,其以一个单独蛋白质单位的形式产生寡糖基转移酶活性。在一个更特别的实施方案中,所述寡糖基转移酶 (OT) 活性物来源于原生动物体,即原生动物寡糖基转移酶 (POT)。这一方案的所述细胞优选的特征进一步在于,所述原生动物寡糖基转移酶活性物来源于弓形虫 (*Toxoplasma gondii* (Tg)), 硕大利什曼虫 (*Leishmania major* (Lm)); 婴儿利什曼虫 (*Leishmania infantum* (Li)), 巴西利什曼虫 (*Leishmania braziliensis* (Lb)), 墨西哥利什曼虫 (*Leishmania Mexicana* (Lmx)), 杜氏利什曼虫 (*Leishmania donovani* (Ld)), 利什曼虫圭亚那亚种 (*Leishmania guyanensis* (Lg)), 热带利什曼虫 (*Leishmania tropica* (Lt)), 克氏锥虫 (*Trypanosoma cruzi* (Tc)), 及布氏锥虫 (*Trypanosoma brucei* (Tb))。本发明还涉及与所述 POT 有关或来源于所述 POT 的同源或人造结构,其功能是在细胞中产生 POT 活性。

[0024] 在本发明的一个特别方面,所述细胞被进一步修饰而使其缺乏或具有被抑制、减少或耗尽的一个或多个定位于高尔基体的甘露糖基转移酶活性物。

[0025] 本发明所述的细胞优选包含一个或多个核酸分子,所述核酸分子编码一个或多个糖蛋白,特别是异源和重组糖蛋白,并且能制造所述糖蛋白或由一个或多个所述糖蛋白构

成的组合物。本发明还提供了制造所述糖蛋白或糖蛋白组合物的方法或工艺,其中所述方法的主要特征在于,提供并使用本发明所述的细胞来制造所述糖蛋白。本发明还提供了可利用本发明所述细胞制造的或利用本发明所述细胞制造的糖蛋白,特别是新的糖蛋白组合物。

[0026] 根据本发明所述的细胞,与该宿主细胞未经修饰的野生型菌株相比,Man1至Man3型脂质连接寡糖的腔内浓度更高。特别地,腔内浓度增加至少5%,10%,15%,20%,25%,30%,40%,50%,70%,或90%,更特别地,增加至少100%,200%,500%,700%,1000%,1500%,2000%或更多。因此,与该宿主细胞未经修饰的野生型菌株相比,所述细胞表现出糖基化效率提高。特别地,尤其是对于基于Man3的结构,糖基化率提高至少5%,10%,15%,20%,25%,30%,40%,50%,70%,或90%,更特别地,提高至少100%,200%,500%,700%,1000%,1500%,2000%或更多。

[0027] 关于内质网敲除突变株,即在内质网中具有被修饰的糖基化特别是采用algl修饰途径的菌株,与未经修饰的内质网敲除突变株相比,根据本发明方法修饰的此类突变株表现出生长率增加和/或温度敏感性降低。特别地,内质网敲除突变株中的生长率增加至少5%,10%,15%,20%,25%,30%,40%,50%,70%,或90%,更特别地,增加至少100%,200%,500%,700%,1000%,1500%,2000%或更多。

[0028] 本发明的一个特别方面涉及一种分离的脂质连接寡糖翻转酶及编码所述翻转酶的分离的核酸分子。根据本发明所述的翻转酶是一种包含至少一个跨膜域及至少一个细胞内膜的定位序列的蛋白质,并且是膜结合的。所述翻转酶的进一步特征在于能跨膜翻转脂质连接寡糖的Man1GlcNAc2,Man2GlcNAc2,或Man3GlcNAc2结构,例如将所述Man1GlcNAc2,Man2GlcNAc2,或Man3GlcNAc2结构从所述细胞器的胞质面翻转至腔面。根据下文进一步描述的方法可分离出所述脂质连接寡糖翻转酶。本发明进一步涉及用于在细胞中表达所述翻转酶活性物的表达盒及表达载体。

[0029] 本发明更特别的方面涉及所述脂质连接寡糖翻转酶的用途,优选将其与对多糖结构具有松弛特异性的寡糖基转移酶,例如特别是原生动物的寡糖基转移酶(POT)结合使用,或者涉及本发明所述的任意一种细胞在制造糖蛋白或包含这种糖蛋白的组合物方面的用途。本发明的其他方面涉及由包含本发明所述细胞的试剂盒制造的糖蛋白及该试剂盒,和它们在制造所述糖蛋白中的应用。

[0030] 更特别地,在第一个方面,本发明提供了一种细胞或宿主细胞,所述细胞或宿主细胞被修饰以表达脂质连接寡糖翻转酶活性物,所述翻转酶活性物能高效地将包含1至3个甘露糖残基的所有脂质连接寡糖从胞内细胞器的胞质面翻转至腔面。

[0031] 在本发明的一个特别方面,所述细胞的特征还在于,所述脂质连接寡糖翻转酶具有高效翻转脂质连接寡糖的活性,所述脂质连接寡糖选自由Man1GlcNAc2,Man2GlcNAc2和Man3GlcNAc2组成的组。

[0032] 在本发明的一个优选方面,所述细胞的特征还在于,所述脂质连接寡糖翻转酶活性物由一个或多个核酸分子的表达来赋予,所述核酸分子选自:

[0033] a) 包含下述一个或多个序列或者由下述一个或多个序列组成的核酸分子:SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:7,SEQ ID NO:9,SEQ ID NO:11,SEQ ID NO:13,SEQ ID NO:15,及SEQ ID NO:17;SEQ ID NO:21,SEQ ID NO:23,SEQ ID NO:25,SEQ ID

NO :27, 及 SEQ ID NO :29 ;

[0034] b) 包含下述一个或多个序列的编码聚氨基酸的核酸分子:SEQ ID NO :2, SEQ ID NO :4, SEQ ID NO :6, SEQ ID NO :8, SEQ ID NO :10, SEQ ID NO :12, SEQ ID NO :14 ;SEQ ID NO 16 及 SEQ ID NO :18 ;SEQ ID NO :22, SEQ ID NO :24, SEQ ID NO :26, SEQ ID NO :28, 及 SEQ ID NO :30 ;及

[0035] c)a) 或 b) 所述的核酸分子的片段、变体、类似物或衍生物。

[0036] 如以上任一方面中所述的细胞,其特征还可以在于,所述细胞器为内质网 (ER)。

[0037] 如以上任一方面中所述的细胞,其特征还可以在于,所述细胞包含至少一种编码异源 (糖) 蛋白的核酸,并且优先表达所述 (糖) 蛋白。

[0038] 所述细胞的特征还在于,所述细胞缺乏或具有被抑制、减少或耗竭的 Rft-1 型脂质连接寡糖翻转酶活性物。优选地,此方面的所述细胞的特征还在于,所述 Rft-1 型脂质连接寡糖翻转酶的特征在于,其对带有少于 5 个甘露糖残基的脂质连接寡糖的翻转活性小于其对带有 5 个甘露糖残基的脂质连接寡糖的翻转活性。更特别地,所述 Rft-1 型脂质连接寡糖翻转酶的特征在于,其对带有少于 5 个甘露糖残基的脂质连接寡糖的翻转活性低于其对带有 5 个甘露糖残基的脂质连接寡糖的翻转活性,其中,“小于”是指,与带有 5 个甘露糖残基的脂质连接寡糖的数量相比,小于其 10%,20%,50%,80%的带有少于 5 个甘露糖残基的脂质连接寡糖被翻转。

[0039] 该特别方案中所述的细胞进一步优选的特征在于,所述细胞为 rft1 基因或 rft1 同源基因的敲除突变株。

[0040] 进一步地,所述细胞的特征还可以在于,所述细胞缺乏或具有被抑制、减少或耗竭的一个或多个定位于内质网的糖基转移酶活性物。进一步优选地,所述细胞的特征在于,所述定位于内质网的糖基转移酶为甘露糖基转移酶。

[0041] 进一步地,所述细胞的特征还可以在于,所述细胞缺乏或具有被抑制、减少或耗竭的一个或多个定位于内质网的脂质连接单糖 (LLM) 翻转酶活性物。

[0042] 进一步地,所述细胞的特征还可以在于,所述细胞缺乏或具有被抑制、减少或耗竭的 Alg-11 型活性物。进一步优选地,所述细胞的特征还可以在于,所述细胞为 alg11 基因或 alg11 同源基因的敲除突变株。

[0043] 进一步地,所述细胞的特征在于,所述细胞缺乏或具有被抑制、减少或耗竭的 Alg-11 型活性物,并且进一步缺乏或具有被抑制、减少或耗竭的一个或多个脂质连接单糖 (LLM) 翻转酶活性物。进一步优选地,所述细胞的特征在于,所述细胞为 alg11 基因或 alg11 同源基因及一个或多个编码脂质连接单糖 (LLM) 翻转酶活性物的基因的敲除突变株。

[0044] 进一步地,所述细胞的特征可以在于,所述细胞缺乏或具有被抑制、减少或耗竭的 Alg11 型活性物,并且进一步缺乏或具有被抑制、减少或耗竭的 Alg3 型活性物。进一步优选地,所述细胞的特征在于,所述细胞为 alg11 基因或 alg11 同源基因及 alg3 基因或 alg3 同源基因的敲除突变株。

[0045] 进一步地,所述细胞的特征可以在于,所述细胞缺乏或具有被抑制、减少或耗竭的 Alg11 型活性物,并且进一步缺乏或具有被抑制、减少或耗竭的 β -D- 甘露糖基转移酶或 DPM1 型活性物。进一步优选地,所述细胞的特征在于,所述细胞为 alg11 基因或 alg11 同源基因及 dpm1 基因或 dpm1 同源基因的敲除突变株。

[0046] 进一步地,所述细胞的特征可以在于,所述细胞缺乏或具有被抑制、减少或耗竭的 Alg2 型活性物。进一步优选地,所述细胞的特征在于,所述细胞为 alg2 基因或 alg2 同源基因的敲除突变株。

[0047] 进一步地,所述细胞的特征可以在于,所述细胞包含一种或多种编码寡糖基转移酶活性物的核酸分子,所述寡糖基转移酶活性物的特征在于,不优先转移 Glc3Man9GlcNAc2 至蛋白质,而是还能够转移除 Glc3Man9GlcNAc2 之外的寡糖至蛋白质,优选具有 1-9 个甘露糖残基的寡糖,最优选 Man1GlcNAc2, Man2GlcNAc2, 和 / 或 Man3GlcNAc2。更特别地,所述细胞的特征在于,所述酶活性物不仅转移 Glc3Man9GlcNAc2 至蛋白质,还能够高效地转移除 Glc3Man9GlcNAc2 之外的寡糖至蛋白质,优选所述寡糖具有 1-9 个甘露糖残基 (Man1GlcNAc2, Man2GlcNAc2, Man3GlcNAc2, Man4GlcNAc2, Man5GlcNAc2, Man6GlcNAc2, Man7GlcNAc2, Man8GlcNAc2, Man9GlcNAc2), 最优选 Man1GlcNAc2, Man2GlcNAc2, 和 / 或 Man3GlcNAc2。

[0048] 进一步优选地,前述方面中所述的细胞的特征在于,所述原动物寡糖基转移酶活性物选自 TbStt3Bp 型活性物, TbStt3Cp 型活性物, LmStt3Ap 型活性物, LmStt3Bp 型活性物, 及 LmStt3Dp 型活性物。

[0049] 进一步地,所述细胞的特征可以在于,所述细胞缺乏或具有被抑制、减少或耗竭的一种或多种定位于高尔基体的甘露糖基转移酶。

[0050] 前述方面中任一或多项所述的细胞,特别地,其特征在于,所述定位于高尔基体的甘露糖基转移酶选自 Och1 型活性物及 Mnn 甘露糖基转移酶家族,尤其是 Mnn1 型活性物、Mnn2 型活性物、Mnn4 型活性物、Mnn5 型活性物、Mnn9 型活性物、Mnn10 型活性物及 Mnn11 型活性物。进一步优选地,所述细胞的特征在于,所述细胞为选自 och1, mnn1, mnn2, mnn4, mnn5, mnn9, mnn10, mnn11 和 / 或它们的同源基因中的至少一种基因的敲除突变株。

[0051] 前述方面中任一或多项所述的细胞,特别地,其特征在于,所述定位于高尔基体的甘露糖基转移酶选自 Ktr 甘露糖基转移酶家族,尤其是 Ktr1 型活性物、Ktr2 型活性物、Ktr3 型活性物、Ktr5 型活性物、Ktr6 型活性物及 Ktr7 型活性物。进一步优选地,所述细胞的特征在于,所述细胞为选自 ktr1, ktr2, ktr3, ktr4, ktr5, ktr6, ktr7 和 / 或它们的同源基因中的至少一种基因的敲除突变株。

[0052] 前述方面中任一或多项所述的细胞,特别地,其特征在于,所述定位于高尔基体的甘露糖基转移酶选自 Van 甘露糖基转移酶家族,尤其是 Van1 型活性物及 Vrg4 型活性物。进一步优选地,所述细胞的特征在于,所述细胞为选自 van1, vrg4 和 / 或它们的同源基因中的至少一种基因的敲除突变株。

[0053] 进一步优选地,前述方面中任一项所述的细胞,其特征在于,所述细胞缺乏或具有被抑制、减少或耗竭的 Mnn2 型活性物,并且进一步缺乏或具有被抑制、减少或耗竭的 Mnn5 型活性物。进一步优选地,所述细胞的特征在于,所述细胞为 mnn2 基因或 mnn2 同源基因及 mnn5 基因或 mnn5 同源基因的敲除突变株。

[0054] 进一步地,所述细胞的特征可以在于,所述细胞缺乏或具有被抑制、减少或耗竭的 Och1 型活性物。进一步优选地,所述细胞的特征在于,所述细胞为 och1 基因或 och1 同源基因的敲除突变株。

[0055] 进一步地,所述细胞的特征可以在于,所述细胞表达一种或多种定位于高尔基体

的异源酶或其催化域,优选地,所述定位于高尔基体的异源酶或其催化域选自:

- [0056] 甘露糖基 (α -1,3)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶 (GnTI);
- [0057] 甘露糖基 (α -1,6)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶 (GnTII);
- [0058] β -1,4-甘露糖基-糖蛋白 4- β -N-乙酰葡萄糖氨基转移酶 (GnTIII);
- [0059] 甘露糖基 (α -1,3)-糖蛋白 β -1,4-N-乙酰葡萄糖氨基转移酶 (GnTIV);
- [0060] 甘露糖基 (α -1,6)-糖蛋白 β -1,6-N-乙酰葡萄糖氨基转移酶 (GnTV);
- [0061] 甘露糖基 (α -1,6)-糖蛋白 β -1,4-N-乙酰葡萄糖氨基转移酶 (GnTVI);
- [0062] β -N-乙酰葡萄糖氨基糖肽 β -1,4-半乳糖基转移酶 (GalT);
- [0063] α (1,6) 岩藻糖基转移酶 (FucT);
- [0064] β -半乳糖苷 α -2,6-唾液酸转移酶 (ST);
- [0065] UDP-N-乙酰葡萄糖胺 2-差向异构酶 (NeuC);
- [0066] 唾液酸合酶 (NeuB);
- [0067] CMP-Neu5Ac 合酶;
- [0068] N-酰基神经氨酸-9-磷酸合酶;
- [0069] N-酰基神经氨酸-9-磷酸酶;
- [0070] UDP-N-乙酰葡萄糖胺转运体;
- [0071] UDP-半乳糖转运体;
- [0072] GDP-岩藻糖转运体;
- [0073] CMP-唾液酸转运体;
- [0074] 核苷酸二磷酸酶;
- [0075] GDP-D-甘露糖 4,6-脱水酶;及
- [0076] GDP-4-酮-去氧-D-甘露糖-3,5-差向异构酶-4-还原酶。

[0077] 进一步地,所述细胞的特征可以在于,所述细胞选自:包括真菌细胞在内的低级真核细胞及包括哺乳动物细胞、哺乳动物细胞系、植物细胞和昆虫细胞在内的高级真核细胞。

[0078] 第三方面,本发明提供了一种或多种分离的核酸分子,其能够编码或提供本发明第一方面中所述的脂质连接寡糖翻转酶活性物。在一个优选方面,所述核酸分子的特征在于,所述分子选自本发明前述方面中任一项所述的核酸分子的一种或多种。

[0079] 第四方面,本发明提供了一种用于在真核宿主细胞中表达的表达盒,包含本发明前述方面中任一项所述的任一种核酸分子的一个或多个拷贝,所述一个或多个拷贝与编码启动子的核酸分子和编码终止子的核酸分子中的至少一种核酸分子结合在一起。

[0080] 在其一个优选方面,所述表达盒还包含编码本发明前述方面中任一项所述的寡糖基转移酶的核酸分子的一个或多个拷贝。

[0081] 第五方面,本发明提供了一种用于真核宿主细胞转化的载体,所述载体包含选自以下拷贝中的一种或多种:本发明前述方面中任一项所述的任一种核酸分子的拷贝和本发明前述方面中任一项所述的表达盒的一个或多个拷贝。

[0082] 第六方面,本发明提供了一种用于制造细胞的方法,所述细胞能够在其细胞内的细胞器如内质网中特别合成具有 Man1GlcNAc2, Man2GlcNAc2 或 Man3GlcNAc2 多糖结构的脂质连接寡糖,所述方法至少包括下述步骤:

[0083] 用至少一种编码脂质连接寡糖翻转酶活性物的构造或结构转化所述细胞,所述构

造或结构选自：

[0084] 本发明前述方面中任一项所述的核酸分子；

[0085] 本发明前述方面中任一项所述的表达盒；及

[0086] 本发明前述方面中任一项所述的载体；

[0087] 以使得所述细胞能够表达由所述构造或结构编码的脂质连接寡糖翻转酶活性物。

[0088] 在其一个优选方面，所述构造还编码寡糖基转移酶活性物，以使得所述细胞能够表达由所述结构编码的脂质连接寡糖翻转酶活性物和寡糖基转移酶活性物。

[0089] 在一个或多个前述方面的优选方面，所述方法还包括在细胞内减少或耗尽以下的至少一种酶活性物的步骤，所述酶活性物选自：

[0090] Alg2 型活性物；

[0091] Alg11 型活性物；

[0092] Alg3 型活性物；

[0093] DPM1 型活性物；以及

[0094] 脂质连接单糖 (LLM) 翻转酶型活性物。

[0095] 第七方面，本发明提供了一种或多种分离的细胞，所述细胞能够在细胞内的细胞器中特别合成具有 Man1GlcNAc2, Man2GlcNAc2 和 / 或 Man3GlcNAc2 多糖结构的脂质连接寡糖，并且能够将所述多糖结构转移至在所述细胞中表达的新生蛋白质，其特征在于，所述细胞可以由或实际上由本发明前述方面中任一项所述的方法来制造。

[0096] 第八方面，本发明提供了一种用于制造糖蛋白或糖蛋白组合物的方法，包括如下步骤：

[0097] 提供本发明前述方面中任一项所述的细胞；

[0098] 在使所述细胞内允许制造所述糖蛋白或糖蛋白组合物的条件下，于培养基中培养所述细胞；及

[0099] 必要时，从所述细胞和 / 或所述培养基中分离所述糖蛋白或糖蛋白组合物。

[0100] 第九方面，本发明提供了一种用于制造糖蛋白或糖蛋白组合物的试剂盒，包括：

[0101] 本发明前述方面中任一项所述的细胞；及

[0102] 培养基，用于培养所述细胞以制造所述糖蛋白。

[0103] 第十方面，本发明提供了一种糖蛋白或糖蛋白组合物，其特征在于，所述糖蛋白或糖蛋白组合物的多糖结构选自：

[0104] GlcNAcMan3-5GlcNAc2，

[0105] GlcNAc2Man3GlcNAc2，

[0106] GlcNAc3Man3GlcNAc2- 平分型

[0107] Gal2GlcNAc2Man3GlcNAc2，

[0108] Gal2GlcNAc2Man3GlcNAc2Fuc，

[0109] Gal2GlcNAc3Man3GlcNAc2- 平分型，

[0110] Gal2GlcNAc3Man3GlcNAc2Fuc- 平分型，

[0111] NeuAc2Gal2GlcNAc2Man3GlcNAc2，

[0112] NeuAc2Gal2GlcNAc2Man3GlcNAc2Fuc，

[0113] NeuAc2Gal2GlcNAc3Man3GlcNAc2- 平分型，

- [0114] NeuAc2Gal2GlcNAc3Man3GlcNAc2Fuc- 平分型,
- [0115] GlcNAc3Man3GlcNAc2,
- [0116] Gal3GlcNAc3Man3GlcNAc2,
- [0117] Gal3GlcNAc3Man3GlcNAc2Fuc,
- [0118] NeuAc3Gal3GlcNAc3Man3GlcNAc2, 及
- [0119] NeuAc3Gal3GlcNAc3Man3GlcNAc2Fuc
- [0120] 第十一方面, 本发明提供了一种宿主细胞, 能够特别制造本发明第九方面所述的一种或多种糖蛋白或糖蛋白组合物。
- [0121] 第十二方面, 本发明提供了一种糖蛋白, 选自:
- [0122] 可由本发明前述方面中任一项所述的细胞制造的糖蛋白;
- [0123] 可利用本发明前述方面中任一项所述的方法制造的糖蛋白; 及
- [0124] 本发明第十方面所述的糖蛋白。
- [0125] 一个优选方面为一种糖蛋白组合物, 包含两种或多种第十方面所述的糖蛋白。
- [0126] 一个优选方面为一种或多种重组蛋白。一个优选方面为一种或多种治疗活性蛋白。
- [0127] 一个优选方面为一种或多种免疫球蛋白。
- [0128] 第十三方面, 本发明提供了一种药物组合物, 包含本发明前述方面中任一项所述的糖蛋白中的一种或多种糖蛋白, 并且优选地, 包含至少一种药学上可接受的载体或辅药(佐剂)。
- [0129] 第十四方面, 本发明提供了一种治疗失调的方法, 所述失调可通过施用本发明前述方面中任一或多项所述的糖蛋白或其组合物中的一种或多种来治疗, 所述方法包括下述步骤:
- [0130] 给对象施用如上所述的糖蛋白或其组合物, 其中, 所述对象患有或疑似患有可通过施用所述糖蛋白或其组合物治疗的疾病。

附图说明

- [0131] 图 1 显示了酵母中生物合成脂质连接寡糖 (LLO) 途径的图示。脂质连接寡糖合成起始于内质网的外膜, 当生成 Man5GlcNAc2 (M5) 结构时, 所述脂质连接寡糖被翻转进入内质网腔并完成所述脂质连接寡糖合成。所述寡糖由 OT (OST) 转移至蛋白质。
- [0132] 图 2 显示了来源于 $\Delta alg11$ 突变菌株 (YG1365) (图 2A) 及 $\Delta alg3 \Delta alg11$ 突变菌株 (YG1363) (图 2B) 的 [3H]-甘露糖标记的脂质连接寡糖的 HPLC 图, 示出 Man3GlcNAc2 结构 (M3) 在 YG1363 中的产生。
- [0133] 图 3 显示了来源于 $\Delta alg111$ 突变菌株 (YG1365) (图 3A) 及 $\Delta alg3 \Delta alg11$ 突变菌株 (YG1363) (图 3B) 的 [3H]-甘露糖标记的蛋白连接寡糖的 HPLC 图。所述内质网合成的 Man3GlcNAc2 LLO 结构 (M3) 在高尔基体区室中被进一步延伸成 Man4GlcNAc2 (M4) 及 Man5GlcNAc2 (M5)。
- [0134] 图 4 显示了从来源于野生型菌株 (WT) (图 4A)、 $\Delta alg11$ 突变菌株 (YG1365) (图 4B) 及 $\Delta alg3 \Delta alg11$ 突变菌株 (YG1363) (图 4C) 的细胞壁蛋白质中分离的 2-AB- 标记 N- 多糖的 MALDI-TOF MS 谱。每个 N- 多糖峰在各峰下方已有注释, 代表 Man3GlcNAc2 至

Man12GlcNAc2 多糖结构 (M3 至 M12)。每个标记的结构都由 2 个 N-乙酰葡萄糖胺 (GlcNAc) 残基和各自指定数目的甘露糖组成;在 m/z 1053 处的峰代表 M3, 在 m/z 1215 处的峰代表 M4, 在 m/z 1377 处的峰代表 M5。所述内质网合成的 M3 LLO 结构在高尔基体区室中被进一步延伸成 M4 及 M5。

[0135] 图 5A-K 列举了编码 Flc2' 或其片段的核苷酸序列或其转录物的氨基酸序列。(内质网定位信号以下划线印出, 跨膜域以加粗字体印出)

[0136] 图 5A 表示编码 Flc2' 的核苷酸序列 (SEQ ID NO:1); 图 5B 表示 Flc2' 转录物的氨基酸序列 (SEQ ID NO:2);

[0137] 图 5C 表示编码内质网定位信号及编码 flc2' 的编码区的跨膜域 (TM) 1 至 3 (TM1-3) 的核苷酸序列 (SEQ ID NO:3); 图 5D 表示图 5C 的核苷酸序列的转录物的氨基酸序列 (SEQ ID NO:4);

[0138] 图 5E 表示编码内质网定位信号及编码 flc2' 的编码区的跨膜域 (TM) 1 至 2 (TM1-2) 的核苷酸序列 (SEQ ID NO:5); 图 5F 表示图 5E 的核苷酸序列的转录物的氨基酸序列 (SEQ ID NO:6);

[0139] 图 5G 表示编码内质网定位信号及编码 flc2' 的编码区的跨膜域 (TM) 2 至 4 (TM2-4) 的核苷酸序列 (SEQ ID NO:7); 图 5H 表示图 5G 的核苷酸序列的转录物的氨基酸序列 (SEQ ID NO:8);

[0140] 图 5I 表示编码内质网定位信号及编码 flc2' 的编码区的跨膜域 (TM) 3 至 4 (TM3-4) 的核苷酸序列 (SEQ ID NO:9); 图 5K 表示图 5I 的核苷酸序列的转录物的氨基酸序列 (SEQ ID NO:10);

[0141] 图 5L 表示代表 flc2' 的内源性启动子的核苷酸序列 (SEQ ID NO:61); 下划线部分标识了起始密码子。

[0142] 图 6A 显示了将野生型菌株与携带空载体、Rft1 (oe RFT1) 或 Flc2' 表达质粒 (oe Flc2') 的 $\Delta rft1$ 突变菌株相比较的斑点分析。每排由指定菌株的一系列稀释物组成。质粒运载的 Flc2' 可以补足 Rft1 缺失; 图 6B 显示了将野生型菌株与 $\Delta alg11$ 突变菌株相比较的各自的斑点分析; 图 6C 显示了将野生型菌株与 $\Delta alg2-1$ 突变菌株相比较的各自的斑点分析。

[0143] 图 7A 及 B 显示了携带空载体、Rft1、Flc2'、包含跨膜域 3 (TM 3)、跨膜域 1 和 3 (TM1-3) 或跨膜域 3 和 4 (TM 3-4) 的 Flc2' 片段或 Flc2 表达质粒的 $\Delta rft1$ 突变菌株的斑点分析。每排由指定菌株的一系列稀释物组成。质粒运载的 Flc2' 可以补足 Rft1 缺失。相反, 全长 Flc2 的超表达 (oe Flc2) 不能补足生长缺陷, 因此引起对内源性翻转酶活性物缺乏的补偿。

[0144] 图 7C 显示了羧肽酶 Y 在野生型酵母菌株或携带空质粒 (YEp352) 或用于 Rft1 的超表达的质粒中任一种及 Flc2' 翻转酶的 $\Delta rft1$ 突变菌株中的 N-糖基化。代表 CPY 的全部糖基化 (mCPY) 及低糖基化形式的带以 -1, -2, -3 及 -4 显示。YEp26.2 代表在 HCSS 中鉴定的初始克隆。

[0145] 图 8 显示了来源于 $\Delta rft1$ 突变菌株的 [3H]-甘露糖标记的脂质连接寡糖的 HPLC 图。图 8A: 携带空载体 YEp352 的 $\Delta rft1$ 突变菌株; 图 8B: 携带 Rft1 表达构造的 $\Delta rft1$ 突变菌株; 图 8C: 携带 Flc2' 表达构造的 $\Delta rft1$ 突变菌株。

[0146] 图 9 描述了羧肽酶 Y 在携带空质粒 (YEp352)、用于 F1c2' 的超表达的质粒或 Rft1 翻转酶的野生型酵母或 $\Delta alg3 \Delta alg11$ 突变菌株中的 N-糖基化结果。

[0147] 图 10 描述了羧肽酶 Y (CPY) 和 β -1,3-葡聚糖转移酶 (Gas1p) 在野生型酵母 (YG1509) 或表达 F1c2' 翻转酶、LmStt3D、或由 F1c2' 翻转酶和 LmStt3D 组成的组合的突变酵母株 YG1365 ($\Delta alg11$) 及 YG1363 ($\Delta alg3 \Delta alg11$) 中 N-糖基化的蛋白免疫印迹 (Western blot) 结果。指出了代表 CPY 及 Gas1p 的全部糖基化 (mCPY) 及低糖基化形式的带。

[0148] 图 11 描述了羧肽酶 Y 在携带空载体 (例如, YEp352) 或用于 F1c2'、POT 或 F1c2' 和 POT 的超表达的质粒的 $\Delta alg11$ 突变菌株中的 N-糖基化。代表 CPY 的全部糖基化 (mCPY) 及低糖基化形式的带以 -1, -2, -3 及 -4 显示。

[0149] 图 12 显示了根据本发明所述的用于在以酵母为例的低级真核生物中的 N-连接糖基化的优选复合系统的图示。更详细地, 脂质连接寡糖的合成发生在内质网的胞质面; 所述合成通过由 Sec59p 将磷酸残基向多萜醇转移而发起, 并且通过数个单糖转移酶在内质网胞质面和腔面的连续动作来延伸寡糖供体, 最终形成脂质连接的 Glc3Man9GlcNAc2。脂质连接的 Glc3Man9GlcNAc2 作为内源性多亚基酵母寡糖基复合物 (Ost 复合物) 的底物; 在复合系统中, $alg3$ 及 $alg11$ 基因缺失 ($\Delta alg11$, $\Delta alg3$) 导致脂质连接 Man3GlcNAc2 的产生。剩余的转移酶仍然存在于细胞中, 但是, 对脂质连接 GlcNAc2Man3 底物是无活性的。添加了本发明所述的新的脂质连接寡糖翻转酶 (F1c2') 及原生动物寡糖基转移酶 (POT 硕大利什曼虫 Stt3D)。在一个可替代的实施方案中, 脂质连接 Man3GlcNAc2 的产生是由 dpm1 基因的缺失而赋予的, 所述基因产物在内质网膜的胞质面制造脂质连接甘露糖 (DPM1)。在一个可替代的实施方案中, 脂质连接 Man3GlcNAc2 是由单糖翻转酶的缺失而产生的, 所述翻转酶将多萜醇连接的甘露糖翻转入内质网腔 (星号)。脂质连接甘露糖为定位于内质网腔的寡糖基转移酶提供供体。通过与 $alg11$ 突变结合, 这种细胞还能制造脂质连接 Man3GlcNAc2。多余未用的转移酶、翻转酶 (Rft1)、酵母 Ost 复合物的组分及未合成的结构以灰体示出。

[0150] 图 13 描述了一个优选实施方案 F1c2' 表达质粒 YEp352F1c2' 的核苷酸序列 (SEQ ID NO:31)。

[0151] 图 14 描述了另一个优选实施方案 LmStt3D 和 F1c2' 共表达质粒 pAX306f 的核苷酸序列 (SEQ ID NO:32)。

[0152] 图 15A 显示了酵母 F1c2 蛋白的截短译本 F1c2' (跨膜域 1-4) 的图示。图 15B 描述了 $\Delta rft1$ 突变菌株的斑点分析, 所述突变菌株或携带空载体 (v. c.), 或携带用于 F1c2' (oe F1c2*) 或截短部分 (TMD1-2, TMD1-3, TMD3-4) 的超表达的载体、或 F1c2' 的单个跨膜域 1, 3 或 4 (TMD1, TMD3, TMD4)。带有跨膜域 3 和 4 (TMD3-4) 及跨膜域 4 (TMD4) 的截短部分被露出以补足 Rft1 的缺失至与全长 F1c2' (=跨膜域 1 至 4) 相似的水平。图 15C 描述了羧肽酶 Y (CPY) 在 $\Delta rft1$ 突变酵母菌株中的 N-糖基化结果, 所述突变菌株或携带空载体 (v. c.), 或携带质粒, 所述质粒用于 F1c2' 的超表达 (oe F1c2*) 或仅包含 F1c2' 跨膜域 4 (F1c2*-TMD4) 的 F1c2' 的截短译本的超表达。指出了代表 CPY 的全部糖基化 (mCPY) 及低糖基化形式的带。跨膜域 4 (F1c2*-TMD4) 的超表达能单独补足 $\Delta rft1$ 突变酵母菌株中的糖基化缺陷。

[0153] 图 16A 描述了羧肽酶 Y (CPY) 在 $\Delta rft1$ 突变酵母菌株中的 N-糖基化结果, 所述突变菌株或携带空载体 (v. c.), 或携带用于 Rft1 (oe Rft1), F1c2' (oe F1c2*) 或内源性

Flc2(oe Flc2) 的超表达的质粒。指出了代表 CPY 的全部糖基化 (mCPY) 及低糖基化形式的带。Flc2 的超表达不能补足 Rft1 缺失时被观察到的低糖基化表型。图 16B 描述了携带空载体 (v. c.)、用于 Rft1(oe Rft1), Flc2*(oe Flc2*) 或 Flc2 的超表达的质粒的 $\Delta rft1$ 细胞的生长分析。该生长分析证实了 Flc2* 能补足 Rft1 缺陷, 及全长 Flc2 不能补足 Rft1 缺陷。

[0154] 图 17ABC 显示了 [3H]-甘露糖标记的脂质连接寡糖的 HPLC 图, 该脂质连接寡糖是从携带空载体 (v. c.) (图 17A) 的 $\Delta alg11$ 突变菌株 (YG1365)、携带用于 Rft1(oe Rft1) (图 17B) 或 Flc2' (oe Flc2*) (图 17C) 的超表达的质粒的 $\Delta alg11$ 突变菌株 (YG1365) 中分离的。检测到的脂质连接寡糖物种为 Man2GlcNac2 (Man2, M2), Man3GlcNac2 (Man3, M3), Man5GlcNac2 (Man5, M5), Man6GlcNac2 (Man6, M6), 及 Man7GlcNac2 (Man7, M7)。M2 及 M3 寡糖定位于内质网膜的胞质面 (cytopl.), M5 至 M7 寡糖定位于内质网膜的腔面 (lumenal)。细胞质的脂质连接寡糖物种之于内腔的脂质连接寡糖物种的相对量可用于表示被表达蛋白质的翻转酶活性物。

[0155] 图 18A 描述了携带空载体 (v. c.)、携带用于 Rft1(oe Rft1) 或 Flc2' (oe Flc2*) 的超表达的质粒的 $\Delta alg11 \Delta alg3$ 突变酵母菌株的生长分析。图 18B 描述了各自细胞的斑点分析。所述生长分析及斑点分析显示出 Flc2' 或 Rft1 的超表达能够改善 $\Delta alg11 \Delta alg3$ 突变酵母菌株的生长。图 18C 描述了羧肽酶 Y (CPY) 在 $\Delta alg11 \Delta alg3$ 突变酵母菌株中的 N-糖基化结果, 所述突变菌株或携带空载体 (v. c.), 或携带用于 Rft1(oe Rft1) 或 Flc2' (oe Flc2*) 的超表达的质粒。指出了代表 CPY 的全部糖基化 (mCPY) 及低糖基化形式的带。Rft1 或 Flc2' 的超表达改善了 CPY 的 N-糖基化。

[0156] 图 19A 描述了携带空载体 (v. c.)、携带用于 Rft1(oe Rft1) 或 Flc2' (oe Flc2*) 的超表达的质粒的 $\Delta alg11$ 突变酵母菌株的生长分析。图 19B 描述了各细胞的斑点分析。所述生长分析及斑点分析显示出 Flc2' 或 Rft1 的超表达能够改善 $\Delta alg11$ 突变酵母菌株的生长。图 19C 描述了羧肽酶 Y (CPY) 在 $\Delta alg11$ 突变酵母菌株中的 N-糖基化结果, 所述突变菌株或携带空载体 (v. c.), 或携带用于 Rft1(oe Rft1) 或 Flc2' (oe Flc2*) 的超表达的质粒。指出了代表 CPY 的全部糖基化 (mCPY) 及低糖基化形式的带。Rft1 或 Flc2' 的超表达改善了 CPY 的 N-糖基化。

[0157] 图 20A 显示了具有温度敏感性 Alg2 蛋白的 $alg2-1$ 菌株中的脂质连接寡糖合成的图示。Alg2 催化甘露糖向 Man1GlcNAc2 (M1) 结构的两步连续添加而产生 Man2GlcNAc2 (M2) 及 Man3GlcNAc2 (M3)。该突变降低 Alg2 活性, 所述 Alg2 反过来减少大于 M1 的脂质连接寡糖种类的合成。但是, Alg2 的剩余活性足以维持常规的脂质连接寡糖合成, 从而导致 Glc3Man9GlcNAc2 结构的产生。M1 和 M2 结构的翻转与由 Alg2 催化的延伸反应互相竞争。如果 M1 和 M2 结构被翻转至内质网腔内, 这些结构在内质网腔中将不作为甘露糖基转移酶的底物, 并且不被进一步延长。最终, 来源于不同脂质连接寡糖供体的寡糖被转移到蛋白质的 N-糖基化共有序列的 Asn 残基上。图 20B 显示了 MALDI-TOF 质谱的图示, 具有预期的峰 Man1GlcNAc2 (M1), Man2GlcNAc2 (M2) 和高甘露糖结构 Man8GlcNAc2 至 Man12GlcNAc2 (M8-M12)。根据 NLO 种类的峰强度, 可以计算出每个结构的相对丰度。M1 类的相对增加表明 M1 的翻转在由 Alg2 催化的 Man1GlcNAc2 (M1) 的延伸反应中占优势地位。

[0158] 图 21A 描述了羧肽酶 Y (CPY) 在 $\Delta alg11$ 突变酵母细胞中的 N-糖基化结果, 所述

突变细胞携带用于 F1c2' (oe F1c2*) 或原生动寡糖基转移酶 POT (oe POT) 及由 F1c2' 和 POT 组成的组合 (oe F1c2* & POT) 的超表达的载体。指出了代表 CPY 的全部糖基化 (mCPY) 及低糖基化形式的带。图 21B 显示了羧肽酶 Y (CPY) 在 $\Delta alg11 \Delta alg3$ 突变酵母细胞中的 N-糖基化结果, 所述突变细胞携带用于 F1c2' (oe F1c2*)、POT (oe POT) 及由 F1c2' 和 POT 组成的组合 (oe F1c2* & POT) 的超表达的载体。指出了代表 CPY 的全部糖基化 (mCPY) 及低糖基化形式的带。POT 和 F1c2' 的共表达在 $\Delta alg11$ 及 $\Delta alg11 \Delta alg3$ 这两种酵母菌株中均抑制低糖基化表型向更高程度变化。

[0159] 图 22AB 描述了从来源于 $\Delta alg3 \Delta alg11$ 酵母突变菌株 (图 22A) 及 $\Delta alg11 \Delta alg3 \Delta mnn1$ 酵母突变菌株 (图 22B) 的细胞壁蛋白质中分离的 2-AB- 标记 N-多糖的 MALDI-TOF MS 谱。每个 N-多糖峰在各自峰上面进行了注释, 代表 Man3GlcNAc2 (M3) 至 Man6GlcNAc2 (M6)。除甘露糖之外, 每个指明的结构包含 2 个附加的 Gn 残基。在 m/z 1053 处的峰代表 M3, 在 m/z 1215 处的峰代表 M4, 在 m/z 1377 处的峰代表 M5, 在 m/z 1539 处的峰代表 M6。所述内质网合成的 Man3GlcNAc2 LL0 结构在高尔基体区室中被进一步延伸成 Man4GlcNAc2、Man5GlcNAc2 及非常少量的 Man6GlcNAc2。正如 Man5 峰大幅降低所显示的, mnn1 的缺失部分地破坏了内质网合成的 Man3GlcNAc2 结构在高尔基体区室中的加工。

具体实施方式

[0160] 本发明主要涉及具有被修饰的脂质连接寡糖的宿主细胞, 其可以通过一组糖基转移酶、糖转运体及甘露糖苷酶的异源表达被进一步修饰, 而成为制造哺乳动物例如人的治疗糖蛋白的宿主株。所述方法提供了一种基因工程的宿主细胞, 其能被用于表达及靶向参与糖基化的任何所需基因。创造或挑选出具有被修饰的脂质连接寡糖的宿主细胞。在所述基因工程宿主细胞中形成的 N-多糖具有 Man1GlcNAc2, Man2GlcNAc2, 和 / 或 Man3GlcNAc2 核心结构, 然后所述核心结构可以通过一种或多种酶例如糖基转移酶、糖转运体及甘露糖苷酶的异源表达被进一步修饰, 来制造类人糖蛋白。对于治疗用蛋白的制造, 该方法可以适用于以基因工程形成细胞系, 在所述细胞系中可以获得任何期望的糖基化结构。

[0161] 除非另有定义, 本发明中所使用的科技术语应当具有本领域普通技术人员所普遍理解的含义。此外, 除非上下文另有要求, 单数词应当包括复数, 且复数词应当包括单数。本发明所述的方法和技术通常根据本领域公知的常规方法来实现。一般来说, 本发明所使用的命名法及所述的生物化学、酶学、分子和细胞生物学、微生物学、遗传学、蛋白质与核酸化学和杂交中的技术均为本领域公知和常用的。除非另有指明, 本发明所述的方法和技术通常根据本领域公知的及本说明书中引用和讨论的各种概括的和具体的参考文献中所述的常规方法来实现。参见, 例如, Sambrook 等的《Molecular Cloning: A Laboratory Manual》(第二版, Cold Spring Harbor Laboratory Press, 纽约州冷泉港 (1989)); Ausubel 等的《Current Protocols in Molecular Biology》(Greene Publishing Associates (1992 年及 2002 年的添加本)); Harlow and Lane Antibodies: A Laboratory Manual》(Cold Spring Harbor Laboratory Press, 纽约州冷泉港 (1990)); 《Introduction to Glycobiology》(Maureen E. Taylor, Kurt Drickamer, Oxford Univ. Press (2003)); 《Worthington Enzyme Manual》(Worthington Biochemical Corp., 新泽西州弗里霍尔德); 《Handbook of Biochemistry: Section A Proteins》(第 I 卷, 1976 年, CRC Press); 《Handbook of

Biochemistry :Section A Proteins》(第 II 卷,1976 年, CRC Press);《Essentials of Glycobiology》(Cold Spring Harbor Laboratory Press(1999))。对于本文中所用的命名法及所述的生物化学与分子生物学的实验过程和技术,均为本领域公知和常用的。

[0162] 提供新的脂质连接寡糖翻转酶

[0163] 在本发明正文中,“脂质连接寡糖翻转酶活性物”或“翻转酶”被定义为将脂质连接寡糖 (LLO) 尤其是多萜醇连接寡糖从胞质面穿过膜向细胞器腔面移位的功能,所述脂质连接寡糖被结合在细胞内的细胞器的膜上,主要是被结合在该膜的胞质面。具体而言,所述细胞内细胞器为内质网 (ER)。该脂质连接寡糖的移位过程的特征在于“翻转”。在一个优选实施方案中,所述翻转酶活性物以内质网为靶点。不希望受理论的束缚,术语“翻转酶”和“翻转”还指用于支持另一种潜在翻转酶蛋白以产生翻转酶活性的支持作用。

[0164] 已经惊奇地发现,所述新的脂质连接寡糖翻转酶是可以分离并在细胞的糖基化级联中具有功能的,并且它们能够补偿内源性脂质连接寡糖翻转酶活性物例如 Rft1 型活性物的减少或缺乏。此外,还惊奇地发现,本发明所述的脂质连接寡糖翻转酶活性物能够在改变的糖基化级联中起作用。所述改变包括产生具有较少量寡糖的脂质连接寡糖,例如,包含少于 5 个甘露糖残基的脂质连接寡糖,并将其跨过内质网膜翻转。在野生型细胞中制造或翻转这样的脂质连接寡糖结构通常是不占优势的。已经惊奇地发现,所述新的脂质连接寡糖翻转酶在跨过细胞内的细胞器膜翻转包含少于 5 个甘露糖残基的脂质连接寡糖尤其是 Man1GlcNAc2, Man2GlcNAc2, Man3GlcNAc2, 或 Man4GlcNAc2 的脂质连接寡糖方面具有高效的活性。所述新的脂质连接寡糖翻转酶在翻转包含 Man5GlcNAc2 的脂质连接寡糖中显示出高活性,在翻转包含 Man4GlcNAc2 的脂质连接寡糖中显示出高活性,在翻转包含 Man3GlcNAc2 的脂质连接寡糖中显示出高活性,在翻转包含 Man3GlcNAc2 的脂质连接寡糖中仍显示出高活性,在翻转包含 Man2GlcNAc2 的脂质连接寡糖中显示出高活性,并且在翻转包含 Man1GlcNAc2 的脂质连接寡糖中仍显示出高活性。发现所述新的脂质连接寡糖翻转酶对于要被翻转的寡糖基结构显示出松弛的特异性。

[0165] 不希望受理论的束缚,本文所用的术语“活性”,特别是对于脂质连接寡糖翻转酶而言,涉及具体对要被转运或合成的特定化合物或分子进行转运、转移或合成的比率。对于分子的跨膜转运,通过估算要被转运通过生物屏障,尤其是被翻转通过或透过细胞内的细胞器膜的具体分子或结构的净通量来估算以转运率表示的转运活性。所述净通量具体通过流入率和流出率来计算。据发现,所述净通量在很大程度上可以取决于被转运分子的分子结构。净通量及相应的转运活性对于每个被转运或翻转的单独结构可以是特定的。不希望受理论的束缚,可以通过确定混入存在于内质网胞质面的脂质连接寡糖的带标记甘露糖的数量,并以该数量除以混入脂质连接寡糖的带标记甘露糖 (优选 [3H]-甘露糖) 的总数来计算所述翻转酶活性。可替代地,所述脂质连接寡糖翻转酶活性可以使用“人工”囊泡来确定。例如,Rft1 型脂质连接寡糖翻转酶对具有 Man5GlcNAc2 结构的脂质连接寡糖的翻转活性很高,但发现其对于具有 Man1GlcNAc2 结构的脂质连接寡糖的翻转活性即使有的话也很低。因此,Rft1 型脂质连接寡糖翻转酶对翻转 Man5GlcNAc2 结构显示出很高的特异性。相反,本发明所述的新的脂质连接寡糖翻转酶对具有 Man1GlcNAc2 结构的脂质连接寡糖的翻转活性很高,并且对具有 Man2GlcNAc2 或 Man3GlcNAc2 结构的脂质连接寡糖的翻转活性也很高。本发明所述的新的脂质连接寡糖翻转酶显示出对特定多糖结构特异性较低,因此显

示出松弛的或低特异性的翻转酶活性。

[0166] 可以通过采用附加的实施例详细描述的高拷贝抑制基因筛选 (HCSS) 方式分离出 (在一个优选实例中, 是从酵母细胞中分离出) 编码本发明所述的脂质连接寡糖翻转酶的基因或“人工”基因。简而言之, 在高拷贝抑制基因筛选中可以使用其内源性脂质连接寡糖翻转酶已被失活的细胞, 例如, 携带 *rft1* 基因缺失的酵母细胞。然后, 所述细胞可以用基因组 DNA 文库, 例如基因组酵母 DNA 文库转化, 从高拷贝质粒例如也携带可选择的标记的 *Yep352* 中被表达。在糖基化级联中有缺陷的细胞将产生低糖基化蛋白, 并且温度和渗透敏感性较高。因此, 检测高拷贝抑制基因筛选中获得的选定细胞在不存在渗透稳定剂例如山梨醇时的生长能力。然后进一步分析阳性菌落的温度敏感性及使所表达的蛋白糖基化的能力。

[0167] 本发明还涉及一种或多种分离的核酸, 该核酸编码具有新的脂质连接寡糖翻转酶活性的脂质连接寡糖翻转酶多肽的; 包括所述分离的核酸的载体及包含所述载体的细胞。

[0168] 在一个具体实施方案中, 本发明提供了一种“人工”的新的脂质连接寡糖翻转酶活性物, 其为 *f1c2'* 的转录物。所述“人工”基因 *f1c2'* 来源于 *f1c2* 基因 (同义名: *YAL053W*; 定位于酵母 1 号染色体上; 碱基 45900-48251)。所述 *F1c2* 转录物是推定的 FAD 转运体, 其定位于内质网膜并且具有使 FAD 进入到内质网中的功能。内源性 *F1c2* 蛋白质不作为翻转酶起作用并且不转运脂质连接寡糖。

[0169] 所述“人工”基因 *f1c2'* 主要是 *f1c2* 的 3' 端的截短译本。*f1c2'* 的全部序列列于 SEQ ID NO: 1 (图 5A) 中, 并代表酵母 1 号染色体, 碱基 45900-47222。所述 *f1c2'* 的转录物产生由 452 个氨基酸组成的蛋白质, 其包含 4 个完整的跨膜域和 1 个截去五分之一的跨膜域 (SEQ ID NO: 2; 图 5B)。C 末端的从氨基酸 442 至 452 的 11 个氨基酸从克隆过程中产生。意想不到的是, *F1c2'* 即 *F1c2* 的 N 末端片段, 能够在 $\Delta rft1$ 突变菌株中补偿缺乏的翻转酶活性, 尽管全长 *F1c2* 本身根本不显示翻转酶活性。更特别的是, 据发现, 所述 *F1c2'* 翻转酶对 *Man1* 结构表现出很大的亲和力, 并以高比例 (速率) 翻转所述 *Man1* 结构。

[0170] 本发明提供了编码本发明所述的新的脂质连接寡糖翻转酶的数个“人工”基因或基因构造。这些“人工”基因或基因构造全部来源于 *f1c2* 基因。具体地说, 提供了“人工”*f1c2'* 的片段及这些片段的一个或多个构造物。本发明不限于这些序列。本发明特别地涉及显示出本文所表征和描述的新的脂质连接寡糖翻转酶型功能的“人工”基因或基因构造。发明人惊奇地发现, 定位于细胞内的细胞器膜的并且翻转脂质连接寡糖进入细胞器内腔的“人工”跨膜蛋白能被翻译或是可获得的。这些蛋白显示出所述新的脂质连接寡糖翻转酶活性, 其主要特征在于对本文所述的脂质连接寡糖的多糖结构具有松弛特异性。

[0171] 经过本文所提出的开创性“原理证明”之后, 主要是以来源于 *f1c2* 的“人工”基因或基因构造的形式, 本领域技术人员通过简单地继续采取如下所述的筛选方法, 可以很容易地提供编码脂质连接寡糖翻转酶类似功能的“人工”基因或基因构造。

[0172] 可替代地或附加地, 本发明提供了基于并且尤其包括 *rft1* 基因或多核苷酸的基因构造, 所述多核苷酸编码 *Rft1* 或 *Rft1* 型活性物以在细胞中产生脂质连接寡糖翻转酶活性, 尤其是在通过超表达 *rft1* 的方式 *Rft1* 以高浓度存在的经过基因修饰的细胞中; 本发明还提供了制造所述细胞的手段。

[0173] 在一个优选实施方案中, 所述脂质连接寡糖翻转酶活性体现在一种或多种蛋白或

类蛋白结构上,例如多单元转运体上。

[0174] 根据本发明,提供了一种分离的或“基本上纯的”核酸分子或其功能性类似物,其能够编码或提供如上文所述的翻转酶活性物。在优选实施方案中,所述核酸分子选自如下所述的核酸分子中的一种或多种。

[0175] 术语“多核苷酸”或“核酸分子”是指在长度上至少有 10 个碱基的核苷酸的聚合物形式。所述术语包括 DNA 分子(例如,cDNA 或基因组 DNA 或合成 DNA)和 RNA 分子(例如,mRNA 或合成 RNA),以及含有非天然核苷酸类似物、非原有核苷酸间的键或同时含有两者的 DNA 或 RNA 的类似物。所述核酸可以处于任何一种拓扑构象。例如,核酸可以是单链的、双链的、三链的、四链的、部分双链的、分支的、发夹结构的、环状的或是处于挂锁构象。所述术语包括单链和双链形式的 DNA。

[0176] 分离的或“基本上纯的”核酸或多核苷酸(例如,RNA,DNA 或混合聚合体)是一种基本上与在其天然宿主细胞中自然伴随原有多核苷酸存在的其他细胞成分例如,核糖体、聚合酶,及其自然结合的基因组序列分离开的核酸或多核苷酸。所述术语包含(1)已从其天然存在环境中脱离的核酸或多核苷酸,(2)不与多核苷酸的全部或一部分结合的核酸或多核苷酸,其中,所述“分离的多核苷酸”是在自然界中发现的,(3)可操作地连接至在自然状态下未连接的多核苷酸的核酸或多核苷酸,或(4)在自然状态下不存在的核酸或多核苷酸。

[0177] 所用的术语“分离的”也可以指重组或克隆 DNA 分离体、化学合成的多核苷酸类似物或利用异源系统生物合成的多核苷酸类似物。但是,“分离的”不一定要求所述核酸或多核苷酸实体上从其原来环境中脱离。例如,生物体基因组中的以下内源性核酸序列被视为“分离的”:如果异源序列(即,与该内源性核酸序列不是自然相邻的序列)被置于邻近所述内源性核酸序列的位置而导致该内源性核酸序列的表达改变。举例来说,非原有启动子序列可以被替换(例如,通过同源重组)成为人细胞基因组中基因的原有启动子,从而导致该基因具有被修改的表达模式。因为所述基因可以与天然位于其侧面的至少一些序列中分开,因而成为“分离的”。如果核酸包含任何对基因组中的相应核酸不自然发生的修饰,其也被认为是“分离的”。例如,如果一种内源性编码序列包括插入序列、缺失序列或通过“人工方式”例如通过人为干预诱导的点突变,则该内源性编码序列被认为是“分离的”。“分离的核酸”还包括在异源部位整合入宿主细胞染色体的核酸,以游离基因存在的核酸构造。此外,“分离的核酸”可以基本上不含其他细胞材料,或当采用重组技术制造时可以基本不含培养基,或当采用化学合成时可以基本上不含化学前体或其他化学品。

[0178] 本发明的一个主要方面涉及来源于 f1c2 且编码所述脂质连接寡糖翻转酶活性物的核酸分子。在此方面的优选实施方案中,所述核酸分子至少携带内质网定位信号的序列和一种或多种跨膜区的序列。

[0179] 在优选实施方案中,宿主细胞中的所述脂质连接寡糖翻转酶活性物通过一种或多种核酸分子的表达来提供,所述核酸分子选自:

[0180] 包含下述一种或多种序列的或者由下述一种或多种序列组成的核酸分子:SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:7,SEQ ID NO:9,SEQ ID NO:11,SEQ ID NO:13,SEQ ID NO:15,及 SEQ ID NO:17;

[0181] 包含下述一种或多种序列的或者由下述一种或多种序列组成的编码聚氨基酸的

核酸分子:SEQ ID NO:2,SEQ ID NO:4,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:14;SEQ ID NO 16 及 SEQ ID NO:18;

[0182] 包含下述一种或多种序列的或者由下述一种或多种序列组成的核酸分子:SEQ ID NO:21,SEQ ID NO:23,SEQ ID NO:25,SEQ ID NO:27,及SEQ ID NO:29,特别地,当被融合于编码内质网定位信号的一种或多种核酸分子时,优选选自 SEQ ID NO:19 及编码包含 HDEL 基序和 / 或 KKxx 基序的聚氨基酸序列的核苷酸序列中的一种。

[0183] 包含下述一种或多种序列的或者由下述一种或多种序列组成的编码聚氨基酸的核酸分子:SEQ ID NO:22,SEQ ID NO:24,SEQ ID NO:26,SEQ ID NO:28,和 SEQ ID NO:30,特别是还包含一种或多种内质网定位信号的核酸分子,优选选自 SEQ ID NO:20 及包含 HDEL 基序和 / 或 KKxx 基序的聚氨基酸序列中的一种;及

[0184] 上述核酸分子的提供本发明脂质连接寡糖翻转酶活性的片段、变体、类似物或衍生物。

[0185] 本文所用的术语“片段”是指多核苷酸的一个节段。片段可具有末端(5' - 或 3' - 端)和 / 或内部缺失。一般来说,多核苷酸的片段在长度上具有至少 4 个核苷酸,具体而言,至少 5 个,至少 6 个,至少 7 个,至少 8 个,至少 9 个,至少 10 个,至少 12 个,至少 15 个,至少 18 个,至少 25 个,至少 30 个,至少 35 个,至少 40 个,至少 50 个,至少 60 个,至少 65 个,至少 70 个,至少 75 个,至少 80 个,至少 85 个,至少 90 个,至少 100 个或更多个核苷酸。

[0186] 本文所用的术语“缺失”是指核苷酸序列的以下变体:所述变体中,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,或 20 个多核苷酸段(由两个或多个核苷酸组成的多核苷酸段)从该核苷酸序列中失去或被除去。

[0187] 本文所用的术语“添加”是指核苷酸序列的以下变体:所述变体中,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,或 20 个多核苷酸段(由两个或多个核苷酸组成的多核苷酸段)被加入或融合至该核苷酸序列内。添加变体也包括融合分子。

[0188] 应当理解的是,在上述修饰的优选变体尤其是通过添加或缺失一个或多个核苷酸而进行修饰的优选变体中,通过添加或缺失 3 个或 3 的整数倍个核苷酸来避免移码。

[0189] 本文所用的术语“类似物”或“相似物”主要是指结构上与天然存在的 RNA 及 DNA 类似(相似)的化合物。核酸是核苷酸的链状物,由以下三部分组成:磷酸主链、折叠形成戊糖(核糖或脱氧核糖)及四种核酸碱基之一。类似物可以具有这些改变中的任意一种,其中,典型地,核酸碱基类似物提供不同的碱基配对和碱基堆积性质,例如能与全部四种标准碱基配对的通用碱基,而磷酸糖主链类似物影响链的性质,例如 PNA(Petersson B et al., Crystal structure of a partly self-complementary peptide nucleic acid(PNA) oligomer showing a duplex-triplex network. J Am Chem Soc. 2005 Feb 9;127(5):1424-30),其二级结构与 DNA 显著不同,并且可以形成三重结构(三链螺旋)。

[0190] 一个优选实施方案为一种或多种分离的核酸分子,选自:(a) 具有如上特征的核酸分子和 (b) 在高严格条件下杂交成 (a) 核酸分子的补体的核酸分子。高严格条件通常被定义为等同于在 45℃ 于 6X 氯化钠 / 柠檬酸钠 (SSC) 中杂交,随后于 65℃ 在 0.2XSSC、0.1% SDS 中清洗。

[0191] 所述实施方案的优选变型为分离的核酸分子,其包含与本文所述的任一核酸序列

至少有 80% 相同的序列或由所述序列组成。

[0192] “内质网定位信号”是指一种肽序列，其引导具有该肽序列的蛋白质被转运至并被保留在内质网中。这种内质网定位序列通常见于驻留在内质网中并起作用的蛋白质中。对于本领域技术人员来说，内质网定位或“保留”信号是可以获得的，例如，*S. cerevisiae* 内质网蛋白 MNS1 的最初 21 个氨基酸残基 (Martinet et al. *Biotechnology Letters* 20: 1171-1177, 1998)。本发明使用的优选内质网定位信号为肽 HDEL (SEQ ID NO: 31)。所述发现于许多酵母蛋白的 C-末端的 HDEL 肽序列，作为保留 / 回收信号作用于内质网 (Pelham *EMBO J.* 7: 913-918, 1988)。具有 HDEL 序列的蛋白质被膜结合受体 (Erd2p) 结合，然后进入逆向转运途径从高尔基体返回到内质网。

[0193] 可替代地，KKxx 序列可以提供内质网定位 (Jackson *J. Cell Biol.* 121: 317)。该基序存在于若干内源性内质网膜蛋白上。所述序列可以存在于所述蛋白的 N 末端或者 C 末端，并且从内质网后区室中被回收。

[0194] 本发明的主要方面是提供适当的宿主细胞 (见下文) 的修饰或基因工程的方法和手段，并在所述细胞中提供改变的和更合适的 N-糖基化。

[0195] 因此，本发明还提供一种表达盒或其功能性类似物，用于在真核宿主细胞中表达如上所述的新的脂质连接寡糖翻转酶活性物，其包含任一种如上所述的核酸分子的一个或多个拷贝。所述载体中的核酸序列可以被操作性地连接到表达调控序列上。优选地，所述核酸分子的一种或多种与编码启动子的核酸分子和编码终止子的核酸分子中的至少一种结合存在。

[0196] 本文所用的“启动子”是指一种能使基因被转录的 DNA 序列。所述启动子被 RNA 聚合酶识别，然后开始转录。启动子包含被 RNA 聚合酶直接结合的 DNA 序列或者参与募集 RNA 聚合酶的 DNA 序列。启动子序列还可以包含“增强子区”，所述“增强子区”为一个或多个 DNA 区域，可以在基因簇中与蛋白质 (即，反式作用因子，更类似于一组转录因子) 结合来增强基因的转录水平 (因此得名)。典型地，当位于编码区的 5' 端时，增强子还可以是与启动子序列分离的，并且可以是例如基因的固有区域或该基因的 3' 端至编码区。

[0197] 根据本发明，所述启动子优选为基因的内源性启动子。在一个优选实施方案中，所述基因位于高拷贝数质粒上，优选为引起超表达的高拷贝数质粒上。在另一个优选实施方案中，所述基因位于低拷贝数质粒上。所述启动子可以为异源启动子。在一个具体的变型中，所述启动子为组成型启动子。在另一个具体的变型中，所述启动子为诱导型启动子。根据本发明所述的一种具体的启动子可进行所述核酸分子的一个或多个拷贝的超表达。在优选实施方案中，与从内源性启动子的表达相比，所述分子被超表达 2 倍，更优选为 5 倍、10 倍、20 倍、50 倍、100 倍、200 倍、500 倍、1000 倍，最优选为 2000 或更多倍。例如，当宿主细胞为巴斯德毕赤酵母 (*Pichia pastoris*) 时，适合的启动子包括但不限于，*aox1*、*aox2*、*das*、*gap*、*pex8*、*ypt1*、*fld1* 及 *p40*；当宿主细胞为酿酒酵母 (*Saccharomyces cerevisiae*) 时，适合的启动子包括但不限于，*gal1*、交配因子 *a*、*cyc-1*、*pgk1*、*adh2*、*adh*、*tef*、*gpd*、*met25*、*galL*、*galS*、*ctr1*、*ctr3* 及 *cup1*；当宿主细胞例如是哺乳动物细胞时，适合的启动子包括但不限于，CMV、SV40、肌动蛋白启动子、*rps21*、劳氏肉瘤病毒基因组大基因组长端重复序列 (RSV)、金属硫蛋白、胸苷激酶或干扰素基因启动子。

[0198] “终止子”或 3' 末端序列是结构基因的终止密码子，其功能为稳定基因的 mRNA

转录产物,在所述转录产物上例如诱发聚腺苷酸化的序列等可被操作性地连接。可以从 *Pichia* 或其他甲基营养型酵母或其他酵母或高级真菌或其他真核生物中获得 3' 末端序列。可用于本发明实际应用的 *Pichia pastoris* 3' 末端序列的实例包括选自 *aox1* 基因、*p40* 基因、*his4* 基因和 *f1d1* 基因的末端序列。

[0199] 根据本发明,还提供了一种用于真核宿主细胞转化的载体,包含任一种如上所述的核酸分子的一个或多个拷贝或如上所述的表达盒的一个或多个拷贝。

[0200] 本文所用的术语“载体”意指能转运与其连接的另一种核酸的核酸分子。载体的一种类型为质粒,其是指附加的 DNA 片段可以接入其内部的一种环状双链 DNA。其他载体包括粘粒、细菌人工染色体 (BAC) 及酵母人工染色体 (YAC)。载体的另一种类型为病毒载体,其中,附加的 DNA 片段可以被接入病毒基因组(下文将详细讨论)。某些载体能在它们被引入的宿主细胞中自主复制(例如,具有在宿主细胞中起作用的复制起点的载体)。其他载体可以在被引入宿主细胞后整合入宿主细胞的基因组,从而随着宿主细胞被复制。此外,特定的优选载体能引导与其操作性连接的基因的表达。这种载体在本文中称为“重组表达载体”(或简称为“表达载体”)。

[0201] 优选地,本发明所述的载体包括选择性标记基因。此类系统的实例包括可用于补足 *his4* *Pichia* 菌株的 *Saccharomyces cerevisiae* 或 *Pichia pastoris his4* 基因,可用于补足 *Pichia pastoris arg* 突变菌株的 *S. cerevisiae* 或 *Pichia pastoris arg4* 基因,或分别可用于补足 *Pichia pastoris ura3* 或 *ade1* 突变菌株的 *Pichia pastoris ura3* 和 *ade1* 基因。其他作用于 *Pichia pastoris* 的选择性标记基因包括 *zeo^R* 基因、*g418^R* 基因、杀稻瘟菌素抗性基因及其类似基因。

[0202] 本发明所述的载体还包括自主复制序列 (ARS)。所述载体还包括作用于细菌的选择性标记基因及负责在细菌中复制和染色体外保持的序列。在替代的实施方案中,营养缺陷标记提供所述选择。细菌选择性标记基因的实例包括氨苄西林抗性 (*amp^r*)、四环素抗性 (*tet^r*)、新霉素抗性、潮霉素抗性 & 博来霉素抗性 (*zeo^R*) 基因。

[0203] 本发明还提供了用于直接基因整合的各种手段。本发明所述的编码在细胞中被表达的蛋白质的核苷酸序列,可以被置于整合载体或者置于复制型载体(例如复制型环状质粒)中。通常,整合载体包括至少第一插入 DNA 片段、选择性标记基因和第二插入 DNA 片段的连续排布序列。所述第一和第二插入 DNA 片段每个在长度上各为大约 200 个核苷酸,并且具有与被转化物种的部分基因组 DNA 同源的核苷酸序列。在所述载体的第一和第二插入 DNA 片段之间而限于所述标记基因之前或之后,插入有包含表达所涉及的结构基因的核酸序列。整合载体能在酵母转化之前被线性化,以促进有关的核酸序列向宿主细胞基因组的整合。

[0204] 本发明还提供了一种聚氨基酸分子,具体地说,提供了一种或多种蛋白质,其能翻转脂质连接的、截短的或完整的前体寡糖 (LLO),特别是 *Man1GlcNAc2*, *Man2GlcNAc2* 和 / 或 *Man3GlcNAc2*。术语“聚氨基酸分子”“多肽”“蛋白质”可替换使用,并且不考虑其长度或是否翻译后修饰,均意味着氨基酸的任何肽连接链。

[0205] 在本发明的一个具体并优选的实施方案中,所述分子包含编码 *F1c2'* 的跨膜域 4 (TM4) 的片段或其同源功能性结构,或基本上由其组成。在本发明一个具体并优选的实施方案中,所述分子包含编码 *F1c2'* 的跨膜域 3-4 (TM3-4) 的片段或其同源功能性结构,或基

本上由其组成。

[0206] 所述分子可以包含编码 F1c2' 的跨膜域 1(TM1) 的片段或其同源功能性结构,或基本上由其组成。所述分子还可以包含编码 F1c2' 的跨膜域 2(TM3) 的片段或其同源功能性结构,或基本上由其组成。在一个具体并优选的实施方案中,所述分子包含编码 F1c2' 的跨膜域 1-2(TM1-2) 的片段或其同源功能性结构,或基本上由其组成。在另一个实施方案中,所述分子包含编码 F1c2' 的跨膜域 2-4(TM2-4) 的片段或其同源功能性结构,或基本上由其组成。

[0207] 所述分子可以包含编码 F1c2' 的跨膜域 3(TM3) 的片段或其同源功能性结构,或基本上由其组成。在一个具体实施方案中,所述分子包含编码 F1c2' 的跨膜域 1-3(TM1-3) 的片段或其同源功能性结构,或基本上由其组成。在另一个实施方案中,所述分子包含编码 F1c2' 的跨膜域 2-3(TM2-3) 的片段或其同源功能性结构,或基本上由其组成。

[0208] 在一个主要方面,所述聚氨基酸为上述来源于 f1c2' 并包括 f1c2' 的“人工”构造中的一种或多种“人工”构造的转录物。在一个优选实施方案中,所述转录物包含下述一种或多种序列或由下述一种或多种序列组成:SEQ ID NO:2,SEQ ID NO:4,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:14;SEQ ID NO:16 和 SEQ ID NO:18。

[0209] 在另一个优选实施方案中,所述转录物包含下述一种或多种序列或由下述一种或多种序列组成:SEQ ID NO:22,SEQ ID NO:24,SEQ ID NO:26,SEQ ID NO:28,和 SEQ ID NO:30,其被融合至内质网定位信号,优选地,选自 SEQ ID NO:20 及包含 HDEL 基序和 KKxx 基序的聚氨基酸序列中的一种。

[0210] 在另一个优选实施方案中,所述聚氨基酸分子为上述转录物中的一种或多种转录物的片段、类似物及衍生物。本文所用的转录物的“片段”、“类似物”及“衍生物”是指具有生物活性的变体,其可以包含添加、缺失或置换。

[0211] 优选地,带有置换的变体具有不多于 50 个的保守氨基酸置换,具体而言,不多于 1,2,3,4,5,6,7,8,9,10,12,15,20,25,30,35,40 或 45 个保守氨基酸置换。“保守置换”应理解为一种氨基酸置换为具有相似特征的另一种氨基酸。保守置换包括在下述组内的置换:缬氨酸、丙氨酸和甘氨酸;亮氨酸、缬氨酸和异亮氨酸;天冬氨酸和谷氨酸;天冬酰胺和谷氨酰胺;丝氨酸、半胱氨酸和苏氨酸;赖氨酸和精氨酸;及,苯丙氨酸和酪氨酸。非极性疏水性氨基酸包括丙氨酸、亮氨酸、异亮氨酸、缬氨酸、脯氨酸、苯丙氨酸、色氨酸及甲硫氨酸。极性中性氨基酸包括甘氨酸、丝氨酸、苏氨酸、半胱氨酸、酪氨酸、天冬酰胺和谷氨酰胺。带正电荷的(碱性)氨基酸包括精氨酸、赖氨酸和组氨酸。带负电荷的(酸性)氨基酸包括天冬氨酸和谷氨酸。相反的,非保守置换是一种氨基酸置换为具有相似特征的另一种氨基酸。

[0212] 本发明所述的聚氨基酸分子显示出或提供了本文中所述的脂质连接寡糖翻转酶活性。具体地说,其特征在于能翻转脂质连接的、截短的或完整的前体寡糖(LL0),尤其是 Man1GlcNAc2, Man2GlcNAc2 和 / 或 Man3GlcNAc2。

[0213] 不希望受理论的束缚,可以采用本领域技术人员公知的方法来测定脂质连接寡糖翻转酶或其片段、变体、类似物或衍生物的活性或特异性。不希望受理论的束缚,一种估算本发明所述的脂质连接寡糖翻转酶活性的优选方法可以包括下述步骤:生长和培养正在表达蛋白质的细胞,所述蛋白质为推定的脂质连接寡糖翻转酶;在特定温度下暴露所述细胞

于标记的甘露糖底物中一定时间,所述甘露糖底物具体为放射性标记的 [3H]- 甘露糖 (标记);及分离甘露糖标记的脂质连接寡糖;及分析 [3H]- 甘露糖标记的脂质连接寡糖中的寡糖含量。可以按照本发明实施例中详细描述的方法分离 [3H]- 标记的脂质连接寡糖。可以采用适当的检测方法例如质谱法 (例如, MALDI-TOF-MS) 或高效液相色谱法 (HPLC) 来分析 [3H]- 甘露糖标记的脂质连接寡糖中的寡糖含量。然后,可以通过确定混入存在于内质网胞质面的脂质连接寡糖的 [3H]- 甘露糖的数量,并将所述数量除以混入脂质连接寡糖的 [3H]- 甘露糖的总数来计算所述翻转酶活性。不能翻转具有特定多糖结构的脂质连接寡糖的细胞将积聚胞质脂质连接寡糖。例如,当具有 Man5GlcNAc2 结构的脂质连接寡糖被翻转并且在内质网腔中被甘露糖基转移酶进一步修饰时,可检测到本发明所述的推定的脂质连接寡糖翻转酶。

[0214] 在野生型细胞中,具有 Man5GlcNAc2 结构的脂质连接寡糖是脂质连接寡糖翻转酶例如 Rft1 翻转酶的底物。表达功能性翻转酶的野生型细胞会主要产生腔脂质连接寡糖,其被进一步加工为具有 Glc3Man9GlcNAc2 结构的最终脂质连接寡糖。然而,缺失脂质连接寡糖翻转酶或具有脂质连接寡糖翻转酶缺陷的细胞,例如 rft1 敲除细胞会主要产生存在于内质网胞质面并且可测量的具有 Man5GlcNAc2 结构的脂质连接寡糖,表明阻碍所述脂质连接寡糖移位进入内质网腔,换言之,阻碍所述脂质连接寡糖进一步加工形成最终的具有 Glc3Man9GlcNAc2 结构的内质网腔脂质连接寡糖。

[0215] 可替代地,可以使用“人工”囊泡来确定所述脂质连接寡糖翻转酶的活性或特异性。可以通过从细胞中提取内质网膜来生成此类囊泡。从这种被耗竭的膜中重组内源性脂质连接寡糖翻转酶例如 Rft1,并用新的脂质连接寡糖翻转酶组装所述囊泡以允许确定所述新蛋白质的翻转酶活性。添加 [3H]- 甘露糖进行标记,胞质甘露糖基转移酶活性使所述 [3H]- 甘露糖合并入胞质面的脂质连接寡糖。然后,所述脂质连接寡糖可借助一种活性脂质连接寡糖翻转酶翻转进入内质网腔。通过用一种 Endo H 酶处理所述囊泡,修剪暴露于囊泡表面的脂质连接寡糖,仅在多萜醇脂质上留下末端 GlcNAc 残基并且从而从囊泡表面去除放射性标记。通过定量测定囊泡腔中出现在 Endo H 酶处理的 [3H]- 甘露糖及未经 Endo H 酶处理的囊泡中的放射性活度,能够计算翻转数量;其中,所测得的放射性越小,脂质连接寡糖翻转酶对特定脂质连接寡糖的活性或特异性越小。

[0216] 可以通过使用至少一种胞质甘露糖基转移酶缺乏或具有缺陷的细胞来确定脂质连接寡糖翻转酶对特定类型脂质连接寡糖的特异性或活性,其根据寡糖结构的不同而不同。例如,Alg2 型活性具有缺陷的细胞会产生具有 Man1GlcNAc2 或 Man2GlcNAc2 结构的脂质连接寡糖;然而,Alg11 型活性和可选地,Alg3 型活性缺乏或具有缺陷的细胞会产生具有 Man3GlcNAc2 结构的脂质连接寡糖。此类突变细胞或其重组的内质网膜囊泡可以被用于测定和确定新分离的脂质连接寡糖翻转酶的活性及特异性。

[0217] 本文中进一步详细描述翻转酶被测定具有翻转酶活性,其对于翻转脂质连接的低甘露糖寡糖,即具有 Man1-3GlcNAc2 结构的寡糖具有很小的特异性,或基本无特异性,其中,所述 Man1-3GlcNAc2 结构为 Man1GlcNAc2、Man2GlcNAc2 或 Man3GlcNAc2 结构。相反,内源性脂质连接寡糖翻转酶尤其是 Rft1,对具有 Man5GlcNAc2 结构的脂质连接寡糖具有最高的翻转酶活性或特异性,其次是对具有 Man3GlcNAc2 结构的脂质连接寡糖。更特别地,本发明所述的翻转酶对脂质连接寡糖显示出相对于内源性 Rft1 反向的特异性,对小的脂质连

接寡糖例如 Man1GlcNAc2 具有最高的特异性,对 Man5GlcNAc2 具有最低的特异性。

[0218] 合适的宿主细胞

[0219] 内质网中脂质连接寡糖向新生蛋白质的转移在包括酵母、真菌、植物、动物和人类的所有真核生物中都是高度保守的。因此,如上文详细描述的本发明的细胞在原则上可以是任意一种真核细胞,包括低级真核细胞、真菌细胞,还包括植物细胞、昆虫细胞或哺乳动物细胞。

[0220] 本发明所述的“宿主细胞”旨在涉及一种已经导入重组载体的细胞。应当理解的是,该术语意为不仅是指特殊的主体细胞,还指这种细胞的子代。由于突变或环境影响,特定的修饰可能发生在随后的几代中,因此,虽然实际上这样的子代可能与其亲代细胞不相同,但是仍然包括在本文所用的术语“宿主细胞”的范围内。重组宿主细胞可以为生长在培养基中的分离的细胞或细胞系,或者可以为存在于活组织或生物体的细胞。用于制造异源糖蛋白的术语“细胞”或“宿主细胞”是指能导入 / 转染或者导入 / 转染核酸,例如编码异源糖蛋白的核酸的细胞。这种细胞不仅包括用于载体或质粒增殖的原核细胞,还包括真核细胞。

[0221] 在优选实施方案中,所述宿主细胞为哺乳动物细胞。优选地,所述细胞选自并优选为无限增殖化的,杂交瘤细胞、骨髓瘤细胞或人细胞的细胞系,骨髓瘤细胞优选为大鼠骨髓瘤细胞和小鼠骨髓瘤细胞。

[0222] 在其更多优选的变型中,所述细胞选自但不限于,CHO 细胞尤其是 CHO K-1 和 CHO DG44、BHK 细胞、NS0 细胞、SP2/0 细胞、HEK293 细胞、HEK293EBNA 细胞、PER. C6 细胞、COS 细胞、3T3 细胞、YB2 细胞、HeLa 细胞和 Vero 细胞。在优选变型中,所述细胞选自 DHFR 缺陷 CHO 细胞,例如 dhfr⁻CHO (Proc. Natl. Acad. Sci. USA, Vol. 77, p. 4216-4220, 1980) 和 CHO K-1 (Proc. Natl. Acad. Sci. USA, Vol. 60, p. 1275, 1968)。

[0223] 在另一个优选实施方案中,所述宿主细胞为两栖动物细胞。优选地,所述细胞选自但不限于, *Xenopus laevis* 卵母细胞 (Nature, Vol. 291, p. 358-360, 1981)。

[0224] 在另一个优选实施方案中,所述宿主细胞为昆虫细胞。优选地,所述细胞选自但不限于, Sf9、Sf21 和 Tn5。

[0225] 在另一个优选实施方案中,所述宿主细胞为植物细胞。优选地,所述细胞选自但不限于,来源于烟草 (*Nicotiana tabacum*)、水生植物浮萍 (*Lemna minor*) 或苔藓 *Physcomitrella patens* (球蒴藓) 的细胞。这些细胞作为制造多肽的体系是公知的,并且还可以作为愈伤组织培养。

[0226] 在目前最优选的实施方案中,所述宿主细胞为低级真核细胞。本发明所述的低级真核细胞包括但不限于,单细胞、多细胞和丝状真菌,优选选自: *Pichia* sp. *Candida* sp. *Saccharomyces* sp.、*Saccharomycodes* sp.、*Saccharomycopsis* sp.、*Schizosaccharomyces* sp.、*Zygosaccharomyces* sp. *Yarrowia* sp.、*Hansenula* sp.、*Kluyveromyces* sp.、*Trichoderma* sp.、*Aspergillus* sp. 和 *Fusarium* sp. 及菌物界,优选选自子囊菌尤其是 *Chrysosporium lucknowense*, 和担子菌尤其是 *Coniophora* sp. 和 *Arxula* sp.

[0227] 在更优选的变型中,所述细胞选自但不限于, *P. pastoris*, *P. stiptis*, *P. methanolica*, *P. bovis*, *P. canadensis*, *P. fermentans*, *P. membranaefaciens*,

P. pseudopolymorpha, *P. quercuum*, *P. robertsii*, *P. saitoi*, *P. silvestrisi*, *P. strasburgensis*; *P. finlandica*, *P. trehalophila*, *P. koclamae*, *P. opuntiae*, *P. thermotolerans*, *P. salictaria*, *P. guercuum*, *P. pijperi*; *C. albicans*, *C. amphixiae*, *C. atlantica*, *C. corydalis*, *C. dosseyi*, *C. fructus*, *C. glabrata*, *C. fermentati*, *C. krusei*, *C. lusitaniae*, *C. maltosa*, *C. membranifaciens*, *C. utilis*; *S. bayanus*, *S. cerevisiae*, *S. bisporus*, *S. delbrueckii*, *S. fermentati*, *S. fragilis*, *S. mellis*, *S. rosei*; *Saccharomyces ludwigii*, *Saccharomycopsis capsularis*; *Schizosaccharomyces pombe*, *Schizosaccharomyces octosporus*, *Zygosaccharomyces bisporus*, *Zygosaccharomyces mellis*, *Zygosaccharomyces rouxii*; *Yarrowia lipolytica*, *Hansenula polymorpha*, *Kluyveromyces* sp., *Trichoderma reesei*, *A. nidulans*, *A. candidus*, *A. carneus*, *A. clavatus*, *A. fumigatus*, *A. niger*, *A. oryzae*, *A. versicolor*, *Fusarium gramineum*, *Fusarium venenatum*, 及 *Neurospora crassa* 和 *Arxula adeninivorans*。

[0228] Rft1 翻转酶缺乏的宿主细胞

[0229] 本文所述的及表 1 和 2 中所指的所有酶及基因均根据它们各自在酵母 *S. cerevisiae* 中的基因座位来命名。本领域技术人员能够提供它们各自存在于其他生物体包括原核生物中的相应活性物。可供选择的来源的实例为酿酒酵母, 毕赤酵母, 曲霉菌 (*Aspergillus*), 假丝酵母 (*Candida*), 及类似菌株。在已知的酶的同源物的基础上, 本领域技术人员可以设计 PCR 引物或者可以用编码此类酶的基因或基因片段来作为探针在靶标生物体的 DNA 文库中识别出同源物。可替代地, 本领域技术人员可以在有关生物体中补足特别的表型。

[0230] 可替代地, 如果一种特别相关的真菌的全部基因组序列是已知的, 本领域技术人员通过检索公众可获得的 DNA 数据文库可以简单地鉴定该基因, 所述 DNA 数据文库可从例如 NCBI, Swissprot 等一些来源获得。例如, 通过用来源于 *S. cerevisiae* 的已知基因检索给定的基因组序列或数据文库, 本领域技术人员能在这样一个基因组中识别出具有高同源性的基因, 其以高度确定性编码具有相似或相同活性的基因。例如, 使用这些方法中的任意一种方法已经在 *P. pastoris* 中识别出来源于 *S. cerevisiae* 的已知甘露糖基转移酶的同源物; 这些基因与 *S. cerevisiae* 中的蛋白质的甘露糖基化涉及的基因具有相似的功能, 因此, 可以应用它们的缺失以在 *P. pastoris* 或具有相似糖基化途径的任何其他真菌中操纵糖基化模式。

[0231] 在上述实施方案的优选变型中, 通过例如敲除 *rft1* 和 / 或 *rft1* 同源基因的方式, 进一步修饰或以基因工程改造所述宿主细胞, 使其缺乏或减少或耗竭 (内源性) 脂质连接寡糖翻转酶, 尤其是 *Rft1*。更特别地, 所述细胞为 *rft1* 基因的敲除突变株。本发明还涉及制造所述细胞的方法。

[0232] 因此, 本发明涉及基因工程细胞, 其中, 通过采用一种或多种方式, 使其中至少一种内源性酶活性缺乏或失活, 所述方法选自倒位引起的抑制、反义结构引起的抑制、缺失引起的抑制、转录水平上的抑制、翻译水平上的抑制及其他方法。这些对于分子生物学领域的技术人员来说都是公知的。

[0233] 在本发明正文中, 术语“敲除”或“敲除突变株”是指该基因或转录物根本不存在

的全部敲除系统,及该基因或转录物仍然存在但是沉默的或低浓度的部分敲除突变株,均导致转录物在细胞中没有发挥充分影响。

[0234] 一旦确定给定的靶基因序列,基因敲除的建立在酵母和真菌分子生物学界是一种完善的技术,并且任何一个本领域普通技术人员都能实施(例如,见 R. Rothsteins, (1991) *Methods in Enzymology*, vol. 194, p. 281)。事实上,对于这类宿主,良性转化的利用性及基因中断技术都可能影响宿主生物体的选择。如果一些转移酶必须被敲除,已经开发了允许重复使用标记物例如 URA3 标记物来连续去除所有不期望的内源性转移酶或本文所引用的其他酶的方法。该技术已经被其他人完善,但基本上包括使用攻击反选择性标记物侧面的两种重复 DNA 序列。所述标记物的存在在转化株的后续选择中是有用的;例如,可以应用在酵母 *ura3*, *his4*, *suc2*, *g418*, *bla*, 或 *shble* 基因中。例如,可以使用 *ura3* 作为标记物来确保选择一种已经整合了构造的转化株。通过用直接重复序列攻击 *ura3* 标记物,技术人员可以首先对已经整合了所述构造并因此中断靶基因的转化株进行选择。分离所述转化株及它们的表型之后,技术人员可以在第二轮中对具有 5' FOA 抗性的转化株进行反选择。能在包含 5' FOA 的平板上生存的菌落已经通过涉及上述重复序列的交叉事件重新失去 *ura3* 标记物。因此,该方法允许重复使用相同的标记物,并且不需要额外标记物的情况下促进多基因的中断。

[0235] 本文所用的适用于核酸或多肽的术语“野生型”是指存在于生物有机体(自然界中存在的生物有机体)中的核酸或由所述生物有机体产生的多肽。

[0236] 本文中适用于宿主细胞中的核酸或由宿主细胞制造的多肽的术语“异源”是指不是来源于与宿主细胞同种的细胞的任何核酸或多肽(例如,基因 N-糖基化的蛋白质)。因此,本文所用的“同源”核酸或蛋白质,是存在于与宿主细胞同种的细胞中或由所述细胞制造的核酸或蛋白质。

[0237] 更特别地,本文所用的与核酸及具体的宿主细胞有关的术语“异源”是指不存在于(并且不能从其获得)自然界中发现的具体细胞中的任何核酸。因此,非天然存在的核酸一旦被导入宿主细胞,就被认为对于该宿主细胞是异源的。值得注意的是,非天然存在的核酸可以包含自然界中发现的核酸序列或核酸序列片段,假如所述核酸作为一个整体在自然界中不存在。例如,在表达载体内包含基因组 DNA 序列的核酸分子是非天然存在的核酸,因此一旦被导入宿主细胞,就被认为对于该宿主细胞是异源的,因为所述核酸分子作为一个整体(基因组 DNA 加上载体 DNA)在自然界中不存在。因此,任何一种作为整体在自然界中不存在的载体、自主复制质粒或病毒(例如,逆转录病毒、腺病毒或疱疹病毒)均被认为是非天然存在的核酸。由此可见,通过 PCR 或限制性核酸内切酶处理制造的基因组 DNA 片段,及 cDNAs 均被认为是非天然存在的核酸,因为未发现它们在自然界中以分离的分子存在。

[0238] 由此还可以得知,任何一种在自然界中未发现的在一种排列中包含启动子序列和多肽编码序列(例如, cDNA 或基因组 DNA)的核酸是非天然存在的核酸。一种非天然存在的核酸对于具体的细胞来说可以是异源的。例如,从酵母 x 的细胞中分离的完整染色体一旦被导入酵母 y 的细胞,对于酵母 y 的细胞来说就是一种异源核酸。

[0239] 进一步缺乏定位于内质网的甘露糖基转移酶的宿主细胞

[0240] 本发明所述的翻转酶在突变细胞中被表达时支持生长和稳定,所述突变细胞缺乏定位于内质网的多糖的合成途径(例如通过基因工程)的一种或多种酶活性物,尤其是缺

乏具有甘露糖基转移酶活性的一种或多种酶,所述甘露糖基转移酶活性转移甘露糖残基至多糖结构例如具有 Man1-3GlcNAc2 结构的脂质连接寡糖。

[0241] 在一个优选实施方案中,所述细胞经特别设计或选择用来合成新生糖蛋白,所述糖蛋白具有适合进一步在高尔基体中进行糖基化加工的 Man1GlcNAc2 结构。

[0242] 在另一个优选实施方案中,所述细胞经特别设计或选择用来合成新生糖蛋白,所述糖蛋白具有适合进一步在高尔基体中进行糖基化加工的 Man2GlcNAc2 结构。

[0243] 在另一个优选实施方案中,所述细胞经特别设计或选择用来合成新生糖蛋白,所述糖蛋白具有适合进一步在高尔基体中进行糖基化加工的 Man3GlcNAc2 结构。

[0244] 在一个优选方面,本发明所述的被修饰后表达上述新脂质连接寡糖翻转酶的宿主细胞被进一步修饰或基因工程改造,以缺乏一种或多种定位于细胞内的细胞器的糖基转移酶活性物。这些优选实施方案的基本构思是减少和控制脂质连接寡糖在细胞器表面和 / 或细胞器内的糖基化,尤其是甘露糖基化。提供本发明所述的宿主细胞能有选择地控制糖基化并且有可能特备提供下述改进的实施方案,所述宿主细胞被修饰后表达上述具有松弛特异性的新脂质连接寡糖翻转酶活性物并因此能翻转低甘露糖尤其是 Man1-3 多糖结构至内腔。

[0245] 优选地,在所述宿主细胞中被敲除、减少或耗竭的定位于内质网的糖基转移酶为甘露糖基转移酶 (见表 1)。在所述宿主细胞的优选方案中,Alg2、Alg3 和 Alg11 型活性物中的一种或多种被敲除、减少或耗竭。在更优选的变型中,这些实施方案进一步缺乏或被减少或耗竭 β -D-甘露糖基转移酶 (Dpm1) 和脂质连接单糖 (LLM) 翻转酶活性物中的一种或多种。

[0246] 表 1 定位于内质网的糖基转移酶活性物

[0247]

名称	功能	EC 号	同义词
DPM1	多萜醇磷酸 β -D-甘露糖基转移酶	2.4.1.83	多萜醇磷酸甘露糖基合酶 多萜醇磷酸甘露糖基转移酶 甘露糖基磷酸多萜醇合酶 甘露糖基磷酸基多萜醇合酶
Alg2	α -1,3-甘露糖基转移酶	2.4.1.-	YGL065C
Alg11	α -1,2-甘露糖基转移酶	2.4.1.-	YNL048L
Alg3	多萜醇磷酸-甘露糖-糖脂 α -甘露糖基转移酶	2.4.1.130	AlgC

[0248] 在一个具体实施方案中,所述宿主细胞为缺乏 Alg2 型活性物的突变菌株。更特别地,所述细胞为 alg2 基因和 / 或 alg2 同源基因的敲除突变株。特别地,所述宿主细胞能合成具有 Man1GlcNAc2 和 Man2GlcNAc2 结构的脂质连接寡糖。本发明还涉及制造所述细胞的方法。

[0249] 在另一个具体实施方案中,所述宿主细胞为缺乏 Alg11 型活性物的突变菌株。更特别地,所述细胞为 alg11 基因和 / 或 alg11 同源基因的敲除突变株。特别地,所述细胞能合成具有 Man3GlcNAc2、Man6GlcNAc2 和 Man7GlcNAc2 结构的脂质连接寡糖。在其一个优选的变型中,所述宿主细胞为同时缺乏 Alg11 型活性物和脂质连接单糖 (LLM) 翻转酶活性物的突变菌株。更特别地,所述细胞为 alg11 基因和 / 或 alg11 同源基因及一个或多个编码脂质连接单糖 (LLM) 翻转酶活性物的基因的敲除突变株。特别地,所述细胞能主要合成具有 Man3GlcNAc2 结构的脂质连接寡糖。本发明还涉及制造所述细胞的方法。

[0250] 在其另一个优选的变型中,所述宿主细胞为同时缺乏 Alg11 型活性物和 β -D-甘露糖基转移酶 (DPM1) 型活性物的突变菌株。更特别地,所述细胞为 alg11 基因或 alg11 同源基因及 dpm1 基因或 dpm1 同源基因的敲除突变株。特别地,所述细胞能主要合成具有 Man3GlcNAc2 结构的脂质连接寡糖。本发明还涉及制造所述细胞的方法。

[0251] 在另一个具体实施方案中,所述宿主细胞为缺乏 Alg3 型活性物的突变菌株。更特别地,所述细胞为 alg3 基因和 / 或 alg3 同源基因的敲除突变株。在一个更优选的实施方案中,所述细胞为同时缺乏 Alg3 型活性物和 Alg11 型活性物的突变菌株。更特别地,所述细胞为 alg3 基因和 alg11 基因和 / 或其任何同源基因的敲除突变株。特别地,所述细胞能合成具有 Man3GlcNAc2 结构的脂质连接寡糖。本发明还涉及制造所述细胞的方法。

[0252] 在上述实施方案的优选变型中,所述宿主细胞被进一步修饰或基因工程改造,以缺乏或被减少或耗竭至少一种定位于高尔基体的甘露糖基转移酶活性物。本发明还涉及制造所述细胞的方法。

[0253] 表达 POT- 复合系统的宿主细胞

[0254] 本发明一个特别优选的实施方案涉及一种寡糖基转移酶 (OST 或 OT) 和 / 或修饰的寡糖基转移酶的表达,优选为异源的寡糖基转移酶。所述寡糖基转移酶为糖基转移酶。所述寡糖基转移酶是一种膜蛋白或蛋白复合物,转移脂质连接寡糖中的寡糖至新生蛋白质。在野生型细胞中,脂质连接寡糖的 Glc3Man9GlcNAc2 结构会被转移并被附加到将被糖基化的蛋白质的天冬酰胺 (Asn) 残基上。所述被寡糖基转移酶催化的反应在 N- 连接糖基化途径中是重要的步骤。

[0255] 酵母和脊椎动物的寡糖基转移酶是由 7 或 8 个亚单位 (酵母中为 Ost1p, Ost2p, Ost3p/Ost6p, Ost4p, Ost5p, Stt3p, Wbp1p, 和 Swp1p; 在哺乳动物细胞中为核糖体结合蛋白 I, DAD1, N33/IAP, OST4, Stt3A/Stt3B, Ost48, 和核糖体结合蛋白 II) 组成的复合异源寡聚蛋白。与酵母或脊椎动物的多蛋白复合物相反,除了仅包含催化 Stt3 亚单位的 Trypanosoma sp. 和 Leishmania sp 之外,原生动物的基因组具有 2-4 个亚单位,编码 3 或 4 种完整的旁系同源基因。原生动物寡糖基转移酶 (POT) 与酵母和脊椎动物的寡糖基转移酶在它们对不同脂质连接寡糖结构的特异性上是有区别的。

[0256] 不希望受理论的束缚,内源性寡糖基转移酶对转移具有高甘露糖多糖结构的脂质连接寡糖可以是高度特异的,所述高甘露糖多糖结构是野生型细胞的内质网中典型具有

的。因此,内源性寡糖基转移酶对转移具有 Glc3Man9GlcNAc2 结构的脂质连接寡糖可以是高度特异的。在本发明所述的宿主细胞中,内质网中的甘露糖基化被抑制,并且被修饰的细胞主要制造具有 Man1-3GlcNAc2 结构的脂质连接寡糖。内源性寡糖基转移酶例如酵母多萜基二磷酸寡糖蛋白糖基转移酶(亚单位:Wbpl, Ost1, Ost2, Ost3, Ost4, Ost5, Ost6, Swp1, Stt3p)对此类低甘露糖脂质连接寡糖的活性可能很低。例如,预期酵母寡糖基转移酶(见图 1)对具有 Man1GlcNAc2, Man2GlcNAc2, Man3GlcNAc2, Man4GlcNAc2 或 Man5GlcNAc2 结构的脂质连接寡糖的活性可能很低。不希望受理论的束缚,内源性寡糖基转移酶的存在可能影响限速步骤,并可能在糖基化级联中引起“瓶颈”,因为低甘露糖多糖向新生蛋白质的转移以非常受限的速率发生,如果真发生的话。

[0257] 因此,在另一个方面,本方面还提供了一种或多种被修饰的或优选的异源的寡糖基转移酶,及特别是表达或超表达这些被修饰的或优选的异源的寡糖基转移酶中的一种或多种寡糖基转移酶的细胞。可替代地或附加地,根据本发明提供的宿主细胞被修饰或基因工程改造,来表达或包含一种或多种被修饰的或优选的异源的寡糖基转移酶活性物,其特征在于,所述寡糖基转移酶活性物不优先转移 Glc3Man9GlcNAc2 至蛋白质,而是还能够转移除 Glc3Man9GlcNAc2 之外的寡糖至蛋白质,所述寡糖优选是具有 1-9 个甘露糖残基的寡糖,最优选 Man1GlcNAc2, Man2GlcNAc2, 和 / 或 Man3GlcNAc2。换言之,本发明提供了具有至少一种定位于内质网的寡糖基转移酶活性物的宿主细胞,所述寡糖基转移酶活性物对不同类型的将被转移至蛋白质的多糖结构显示出松弛特异性。具体地说,这种活性物在本文中被称作“类 POT 活性物”或“POT 活性物”。

[0258] 在一个具体实施方案中,提供一种在本发明的宿主细胞中使用的原生动物寡糖基转移酶(POT),其对转移低甘露糖结构,尤其是 Man1GlcNAc2, Man2GlcNAc2 或 Man3GlcNAc2, 显示出相当大的活性。

[0259] 在更优选的变型中,所述原生动物寡糖基转移酶为一种原生动物的酵母寡糖基转移酶的 Stt3 亚单位的同源物,具体的所述原生动物选自但不限于, *Toxoplasma* sp., *Leishmania* sp., 和 *Trypanosoma* sp.。优选地,所述原生动物选自但不限于, *Toxoplasma gondii*(Tg), *Leishmania major*(Lm); *Leishmania infantum*(Li), *Leishmania braziliensis*(Lb), *Leishmania mexicana*(Lmx), *Leishmania donovani*(Ld), *Leishmania guyanensis*(Lg), *Leishmania tropica*(Lt), *Trypanosoma cruzi*(Tc), 和 *Trypanosoma brucei*(Tb)。在具体实施方案中,所述原生动物寡糖基转移酶选自下述旁系同源基因中的一种或多种: *Trypanosoma brucei* 的 TbStt3Bp 和 TbStt3Cp; *Leishmania infantum* 的 LiStt3-1, LiStt3-2, 和 LiStt3-3; *Leishmania braziliensis* 的 LbStt3-1, LbStt3-2, 和 LbStt3-3; 和 *Leishmania major* 及其同源结构的 LmStt3A, LmStt3B, LmStt3C, 和 LmStt3D。在另一个实施方案中,所述原生动物寡糖基转移酶选自下述一种或多种: *Trypanosoma brucei* 的 TbStt3Bp 和 TbStt3Cp; 及 *Leishmania major* 的 LmStt3Ap, LmStt3Bp, 和 LmStt3Dp。

[0260] 因此,本发明还涉及本发明所述的宿主细胞,其包含一种或多种编码一种或多种原生动物寡糖基转移酶的核酸。表达所述原生动物寡糖基转移酶或类原生动物寡糖基转移酶活性物的启动子可以是一种内源性启动子,内源性是针对所述活性物被表达时所在的细胞而言。启动子可以进行所述核酸分子的一个或多个拷贝的超表达。可以使用例如 ADH,

Tef 或 GPD 的启动子在酵母中表达所述原生动动物寡糖基转移酶活性物或类原生动动物寡糖基转移酶活性物。在一个优选实施方案中,编码所述原生动动物寡糖基转移酶活性物或类原生动动物寡糖基转移酶活性物的基因位于一种高拷贝数质粒上,所述质粒优选引起超表达。在优选实施方案中,与从低拷贝数质粒或从单拷贝染色体整合的表达相比,所述分子被超表达 2 倍,更优选为 5 倍、10 倍、20 倍、50 倍、100 倍、200 倍、500 倍、1000 倍,最优选为 2000 或更多倍。表达所述原生动动物寡糖基转移酶活性物或类原生动动物寡糖基转移酶活性物的启动子可以为 adh, Tef 或 gpd,例如在一种高拷贝数质粒上。

[0261] 本发明还涉及制造这些细胞的方法。

[0262] 脂质连接单糖 (LLM) 敲除 - 原生动动物寡糖基转移酶 (POT) 复合系统

[0263] 本发明提供了一种被修饰的或基因工程改造的宿主细胞,在下文中被称为“复合系统”。本发明所述的复合系统是指一种宿主细胞,其特别能合成具有低甘露糖多糖结构的脂质连接寡糖,并转移所述低甘露糖多糖至一种或多种在所述细胞中表达的新生蛋白质;所述细胞:

[0264] (i) 被修饰以在细胞内的细胞器中合成具有低甘露糖多糖结构,尤其是 Man1GlcNAc2, Man2GlcNAc2 或 Man3GlcNAc2 结构的脂质连接寡糖;尤其是通过敲除至少一种定位于细胞器的甘露糖基转移酶,并且根据情况敲除脂质连接单糖 (LLM) 翻转酶的方式来完成所述修饰,如本文中所详细描述;及

[0265] (ii) 进一步被修饰后表达外源性 / 异源性寡糖基转移酶,其对将被转移至新生蛋白质的低甘露糖多糖结构显示出松弛的底物特异性,尤其是与内源性寡糖基转移酶的底物特异性相比,其中,所述外源性 / 异源性寡糖基转移酶为原生动动物寡糖基转移酶 (POT)。

[0266] 在一个具体实施方案中,对将被转移至新生蛋白质的低甘露糖多糖结构显示出松弛的底物特异性的寡糖基转移酶为原生动动物寡糖基转移酶 (POT)。在一个具体实施方案中,将在本发明所述的宿主细胞中被表达或超表达的原生动动物寡糖基转移酶为 *Leishmania major* 的旁系同源基因 (paralogue) LmStt3B 或其同源结构。在另一个具体实施方案中,将在本发明所述的宿主细胞中被表达或超表达的原生动动物寡糖基转移酶为 *Leishmania major* 的旁系同源基因 LmStt3C 或其同源结构。在另一个具体实施方案中,将在本发明所述的宿主细胞中被表达或超表达的原生动动物寡糖基转移酶为 *Leishmania major* 的旁系同源基因 LmStt3D 或其同源结构。

[0267] 在另一个具体实施方案中,将在本发明所述的宿主细胞中被表达或超表达的原生动动物寡糖基转移酶为 *Leishmania braziliensis* 的旁系同源基因 LbStt3-1 或其同源结构。在所述宿主细胞中被表达或超表达的原生动动物寡糖基转移酶还可以为 *Leishmania braziliensis* 的旁系同源基因 LbStt3-2 或其同源结构。在另一个具体实施方案中,将在本发明所述的宿主细胞中被表达或超表达的原生动动物寡糖基转移酶为 *Leishmania braziliensis* 的旁系同源基因 LbStt3-3 或其同源结构。

[0268] 在另一个具体实施方案中,将在本发明所述的宿主细胞中被表达或超表达的原生动动物寡糖基转移酶为 *Leishmania infantum* 的旁系同源基因 LiStt3-1 或其同源结构。在另一个具体实施方案中,将在本发明所述的宿主细胞中被表达或超表达的原生动动物寡糖基转移酶为 *Leishmania infantum* 的旁系同源基因 LiStt3-2 或其同源结构。在所述宿主细胞中被表达或超表达的原生动动物寡糖基转移酶还可以为 *Leishmania infantum* 的旁系同源

基因 LiStt3-3 或其同源结构。

[0269] 在另一个具体实施方案中,将在本发明所述的宿主细胞中被表达或超表达的原生动物寡糖基转移酶为 *Trypanosoma brucei* (布氏锥虫) 的旁系同源基因 TbStt3A 或其同源结构。在另一个具体实施方案中,将在本发明所述的宿主细胞中被表达或超表达的原生动物寡糖基转移酶为 *Trypanosoma brucei* 的旁系同源基因 TbStt3B 或其同源结构。在另一个具体实施方案中,将在本发明所述的宿主细胞中被表达或超表达的原生动物寡糖基转移酶为 *Trypanosoma brucei* 的旁系同源基因 TbStt3C 或其同源结构。

[0270] 在本发明的具体实施方案中,提供了一种表达盒或其功能性类似物,用于表达一种或多种对低甘露糖多糖结构具有松弛的底物特异性的原生动物寡糖基转移酶,例如尤其是上述原生动物寡糖基转移酶中的一种或多种。所述表达盒包含编码寡糖基转移酶的任一种核酸分子的一个或多个拷贝,所述寡糖基转移酶对低甘露糖多糖结构具有松弛的底物特异性,选自上述原生动物寡糖基转移酶。

[0271] 在其一个具体的变型中,还提供了一种用于转化真核宿主细胞的载体,包含编码上述原生动物寡糖基转移酶中的一种或多种的核酸分子的一个或多个拷贝。所述载体中的核酸序列能被操作性地连接到表达调控序列。优选地,所述核酸分子中的一种或多种与编码启动子的核酸分子及编码终止子的核酸分子中的至少一种核酸分子结合存在。表达原生动物寡糖基转移酶的启动子可以为 ADH, Tef 或 GPD,例如在一种高拷贝数质粒上。

[0272] 在更优选的实施方案中,本发明提供了一种转基因突变细胞,其表达 *Leishmania major* 的旁系同源基因 LmStt3D 或其同源结构。在其一个具体的变型中,LmStt3D 在一种低拷贝载体的所述细胞中被表达。在其另一个具体的变型中,LmStt3D 在一种高拷贝载体的所述细胞中被表达。

[0273] 在另一个优选实施方案中,所提供的细胞表达 *Leishmania braziliensis* 的旁系同源基因 LbStt3-3 或其同源结构。在其一个具体的变型中,LbStt3-3 在一种低拷贝载体的所述细胞中被表达。在其另一个具体的变型中,LbStt3-3 在一种高拷贝载体的所述细胞中被表达。

[0274] 在另一个优选实施方案中,所提供的细胞表达 *Leishmania braziliensis* 的旁系同源基因 LbStt3-1 或其同源结构。在其一个具体的变型中,LbStt3-1 在一种高拷贝载体的所述细胞中被表达。

[0275] 在另一个优选实施方案中,所提供的细胞表达 *Leishmania infantum* 的旁系同源基因 LiStt3-2 或其同源结构。在其一个具体的变型中,LiStt3-2 在一种低拷贝载体的所述细胞中被表达。

[0276] 在另一个优选实施方案中,所提供的细胞表达 *Trypanosoma brucei* 的旁系同源基因 TbStt3B 或其同源结构。在其一个具体的变型中,TbStt3B 在一种高拷贝载体的所述细胞中被表达。

[0277] 在另一个优选实施方案中,所提供的细胞表达 *Trypanosoma brucei* 的旁系同源基因 TbStt3C 或其同源结构。在其一个具体的变型中,TbStt3C 在一种高拷贝载体的所述细胞中被表达。

[0278] 在所述复合系统的一个具体实施方案中,所述细胞为一种突变菌株,其 (i) 至少缺乏 Alg2 型活性物,及 (ii) 表达或超表达原生动物寡糖基转移酶型活性物。更特别地,所

述细胞 (i) 为 *alg2* 基因和 / 或 *alg2* 同源基因的敲除突变株, 及 (ii) 表达一种或多种上述原生动物寡糖基转移酶型活性物。本发明还涉及制造所述细胞的方法。

[0279] 在所述复合系统的一个具体实施方案中, 所述细胞为一种突变菌株, 其 (i) 至少缺乏 *Alg11* 型活性物, 及 (ii) 表达或超表达原生动物寡糖基转移酶型活性物。更特别地, 所述细胞 (i) 为 *alg11* 基因和 / 或 *alg11* 同源基因的敲除突变株, 及 (ii) 表达一种或多种上述原生动物寡糖基转移酶型活性物。在一个优选实施方案中, 本发明提供了一种表达 *Leishmania major* 的旁系同源基因 *LmStt3D* 的 *alg11* 基因和 / 或 *alg11* 同源基因的敲除突变株。在本发明一个具体的变型中, *LmStt3D* 在一种低拷贝载体中被表达。在另一个优选实施方案中, 所述突变细胞表达 *Leishmania braziliensis* 的旁系同源 *LbStt3-3*。在一个具体的变型中, *LbStt3-3* 在一种低拷贝载体中被表达。本发明还涉及制造这些细胞的方法。

[0280] 在所述复合系统的另一个具体实施方案中, 所述细胞为一种突变菌株, 其 (i) 至少缺乏 *Alg3* 型活性物和 *Alg11* 型活性物, 及 (ii) 表达或超表达原生动物寡糖基转移酶型活性物。更特别地, 所述细胞 (i) 为 *alg3* 基因和 *alg11* 基因和 / 或其任一种同源基因的敲除突变株, 及 (ii) 表达一种或多种上述原生动物寡糖基转移酶型活性物。在一个优选实施方案中, 本发明提供了一种表达 *Leishmania major* 的旁系同源基因 *LmStt3D* 的 *alg3* 基因和 *alg11* 基因和 / 或其任一种同源基因的敲除突变株。在其一个具体的变型中, *LmStt3D* 在一种低拷贝载体中被表达。在另一个优选实施方案中, 所述突变细胞表达 *Leishmania braziliensis* 的旁系同源基因 *LbStt3-3*。在其一个具体的变型中, *LbStt3-3* 在一种低拷贝载体中被表达。在另一个优选实施方案中, 所述突变细胞表达 *Trypanosoma brucei* 的旁系同源基因 *TbStt3B* 或 *TbStt3C*。在一个具体的变型中, *TbStt3B* 或 *TbStt3C* 在一种高拷贝载体中被表达。本发明还涉及制造这些细胞的方法。

[0281] 在所述复合系统的另一个具体实施方案中, 所述细胞为一种突变菌株, 其 (i) 至少缺乏 *Alg11* 型活性物和脂质连接单糖 (LLM) 翻转酶型活性物, 及 (ii) 表达或超表达原生动物寡糖基转移酶型活性物。更特别地, 所述细胞 (i) 为 *alg11* 基因和 / 或 *alg11* 同源基因及一个或多个编码脂质连接单糖 (LLM) 翻转酶型活性物的基因的敲除突变株, 及 (ii) 表达一种或多种上述原生动物寡糖基转移酶型活性物。本发明还涉及制造这些细胞的方法。

[0282] 在所述复合系统的另一个具体实施方案中, 所述细胞为一种突变菌株, 其 (i) 至少缺乏 *Alg11* 型活性物和 β -D-甘露糖基转移酶 (*DPM1*) 型活性物, 及 (ii) 表达或超表达原生动物寡糖基转移酶型活性物。更特别地, 所述细胞 (i) 为 *alg11* 基因和 / 或 *dpml* 基因和 / 或其同源基因的敲除突变株, 及 (ii) 表达一种或多种上述原生动物寡糖基转移酶型活性物。本发明还涉及制造这些细胞的方法。

[0283] 不希望受理论的束缚, 在优选变型中, 要求对内源性寡糖基转移酶无敲除突变。然而, 在一个优选变型中, 内源性寡糖基转移酶在细胞中是不存在或是被抑制的。因此, 提供了一种细胞, 其中一种或多种编码内源性寡糖基转移酶亚单位的基因被敲除。在包括酵母细胞的优选变型中, 所述内源性寡糖基转移酶的至少一种亚单位选自 *Wbp1p*, *Ost1p*, *Ost2p*, *Ost3p*, *Ost4p*, *Ost5p*, *Ost6p*, *Swp1p*, 和 *Stt3p*。在一个优选实施方案中, 所述细胞为 *wbp1* 基因和 *stt3* 基因的敲除突变株。在另一个优选实施方案中, 所述细胞为 *ost1* 基因和 *ost2* 基因的敲除突变株。

[0284] 在一个具体的变型中,所述宿主细胞为 Stt3p 的突变菌株,更特别地,所述宿主细胞为酵母菌株 YG543,其具有 stt3-7 等位基因的温度敏感性表现型 (Spirig et al. Mol. Gen. Genet. 256, p. 628-637, 1997)。

[0285] 脂质连接单糖敲除 - 脂质连接寡糖翻转酶 - 原生动物寡糖基转移酶复合系统

[0286] 根据另一个方面,本发明提供了一种宿主细胞,其特别能合成具有低甘露糖多糖结构的脂质连接寡糖,并转移所述低甘露糖多糖至一种或多种在所述细胞中表达的新生蛋白质;所述细胞:

[0287] (i) 被修饰以在细胞内的细胞器中合成具有低甘露糖多糖结构,尤其是 Man1GlcNAc2, Man2GlcNAc2 或 Man3GlcNAc2 结构的脂质连接寡糖;尤其是通过敲除至少一种定位于细胞器的甘露糖基转移酶,并且根据情况敲除脂质连接单糖 (LLM) 翻转酶的方式来形成,如本文中所述详细描述;

[0288] (ii) 被修饰以表达新的对低甘露糖脂质连接寡糖具有松弛特异性的脂质连接寡糖翻转酶,如本文中所述详细描述;及

[0289] (iii) 进一步被修饰以表达寡糖基转移酶,其对将被转移至新生蛋白质的低甘露糖多糖结构显示出松弛的底物特异性,优选地,所述寡糖基转移酶为原生动物寡糖基转移酶 (POT),更特别选自上述原生动物寡糖基转移酶。

[0290] 本发明提供了一种表达盒或其功能性类似物,用于表达上述新的脂质连接寡糖翻转酶活性物,及对低甘露糖多糖结构具有松弛的底物特异性的寡糖基转移酶例如原生动物寡糖基转移酶。所述表达盒包含编码上述脂质连接寡糖翻转酶活性物的任一种核酸分子的一个或多个拷贝,及编码寡糖基转移酶例如上述原生动物寡糖基转移酶的任一种核酸分子的一个或多个拷贝,所述寡糖基转移酶对低甘露糖多糖结构具有松弛的底物特异性。

[0291] 在其一个具体的变型中,还提供了一种用于转化真核宿主细胞的载体,包含编码上述核酸分子中的任一种的一个或多个拷贝,或包含上述表达盒的一个或多个拷贝。所述载体中的核酸序列能被操作性地连接到表达调控序列。优选地,所述核酸分子中的一种或多种与编码启动子的核酸分子及编码终止子的核酸分子中的至少一种核酸分子结合存在。表达原生动物寡糖基转移酶的启动子可以为 ADH, Tef 或 GPD,例如在一种高拷贝数质粒上。

[0292] 一个关于为宿主细胞提供新的脂质连接寡糖翻转酶活性物和原生动物寡糖基转移酶活性物的载体的优选实施方案示于图 14 中。所述核酸序列提供于 SEQ ID NO :32 中。

[0293] 本文所用的术语“来源于 f1c2'”还包括包含 f1c2' (SEQ ID NO :1) 的完整序列的分子,并且在更优选的变型中,包括包含 f1c2' 的一个或多个片段的分子,所述一个或多个片段编码一个或多个 F1c2 分子的跨膜域。在本发明一个具体并优选的实施方案中,所述分子包含编码 F1c2' 的跨膜域 4 (TM4) 的片段或其同源功能性结构,或基本上由其组成。在本发明一个具体并优选的实施方案中,所述分子包含编码 F1c2' 的跨膜域 3-4 (TM3-4) 的片段或其同源功能性结构,或基本上由其组成。

[0294] 所述分子可以包含编码 F1c2' 的跨膜域 1 (TM1) 的片段或其同源功能性结构,或基本上由其组成。所述分子还可以包含编码 F1c2' 的跨膜域 2 (TM3) 的片段或其同源功能性结构,或基本上由其组成。在本发明一个具体并优选的实施方案中,所述分子包含编码 F1c2' 的跨膜域 1-2 (TM1-2) 的片段或其同源功能性结构,或基本上由其组成。在本发明的另一个实施方案中,所述分子包含编码 F1c2' 的跨膜域 2-4 (TM2-4) 的片段或其同源功能性结构,

或基本上由其组成。

[0295] 所述分子可以包含编码 Flc2' 的跨膜域 3(TM3) 的片段或其同源功能性结构,或基本上由其组成。在本发明的一个具体实施方案中,所述分子包含编码 Flc2' 的跨膜域 1-3(TM1-3) 的片段或其同源功能性结构,或基本上由其组成。在本发明的另一个实施方案中,所述分子包含编码 Flc2' 的跨膜域 2-3(TM2-3) 的片段或其同源功能性结构,或基本上由其组成。

[0296] 在所述复合系统的一个具体实施方案中,所述细胞为一种突变菌株,其 (i) 至少缺乏 Alg2 型活性物;(ii) 表达本发明所述的新的脂质连接寡糖翻转酶活性物;及 (iii) 表达原生动动物寡糖基转移酶活性物。更特别地,所述细胞 (i) 为 alg2 基因和 / 或 alg2 同源基因的敲除突变株;(ii) 表达一种或多种提供脂质连接寡糖翻转酶活性物的核酸分子;及 (iii) 表达一种或多种上述原生动动物寡糖基转移酶活性物。在一个更具体的实施方案中,所述细胞表达一种或多种来源于 flc2' 的提供脂质连接寡糖翻转酶活性物的核酸分子,如上文中所详细描述。在另一个变型中,所述细胞表达一种或多种来源于 rft1 的提供脂质连接寡糖翻转酶活性物的核酸分子,如上文中所描述的。特别地,所述细胞能合成具有 Man1GlcNAc2 和 Man2GlcNAc2 结构的脂质连接寡糖,并能转移所述结构至新生蛋白质。本发明还涉及制造所述细胞的方法。

[0297] 在另一个优选实施方案中,所述细胞为一种突变菌株,其 (i) 至少缺乏 Alg11 型活性物, (ii) 表达本发明所述的新的脂质连接寡糖翻转酶活性物,及 (iii) 表达原生动动物寡糖基转移酶活性物。更特别地,所述细胞 (i) 为 alg11 基因和 / 或 alg11 同源基因的敲除突变株, (ii) 表达一种或多种提供脂质连接寡糖翻转酶活性物的核酸分子,及 (iii) 表达一种或多种上述原生动动物寡糖基转移酶活性物。在一个更优选的实施方案中,所述细胞表达一种或多种来源于 flc2' 的提供脂质连接寡糖翻转酶活性物的核酸分子,如上文中所详细描述。在另一个变型中,所述细胞表达一种或多种来源于 rft1 的提供脂质连接寡糖翻转酶活性物的核酸分子,如上文中所描述的。特别地,所述细胞能合成具有 Man3GlcNAc, Man6GlcNAc2, Man7GlcNAc2 和 / 或 Man8GlcNAc2 结构的脂质连接寡糖,并能转移所述结构至新生蛋白质。本发明还涉及制造所述细胞的方法。

[0298] 在一个最优选的实施方案中,所述细胞为一种突变菌株,其 (i) 至少缺乏 Alg3 型活性物和 Alg11 型活性物, (ii) 表达本发明所述的新的脂质连接寡糖翻转酶活性物,及 (iii) 表达原生动动物寡糖基转移酶活性物。更特别地,所述细胞 (i) 为 alg3 基因和 alg11 基因或其任一种同源基因的敲除突变株, (ii) 表达一种或多种提供脂质连接寡糖翻转酶活性物的核酸分子,及 (iii) 表达一种或多种上述原生动动物寡糖基转移酶活性物。在一个更具体的实施方案中,所述细胞表达一种或多种来源于 flc2' 的提供脂质连接寡糖翻转酶活性物的核酸分子,如上文中所详细描述。在另一个变型中,所述细胞表达一种或多种来源于 rft1 的提供脂质连接寡糖翻转酶活性物的核酸分子,如上文中所描述的。特别地,所述细胞能合成具有 Man3GlcNAc2 结构的脂质连接寡糖,并能转移所述结构至新生蛋白质。本发明所述的一个优选突变细胞表达 *Leishmania major* 的旁系同源基因 LmStt3D。在一个具体的变型中,LmStt3D 在一种低拷贝载体中被表达。在另一个优选实施方案中,所述突变细胞表达 *Leishmania braziliensis* 的旁系同源基因 LbStt3-3。在一个具体的变型中,LbStt3-3 在一种低拷贝载体中被表达。在另一个优选实施方案中,所述突变细胞表达 *Trypanosoma*

brucei 的旁系同源基因 TbStt3B 或 TbStt3C。在一个具体的变型中, TbStt3B 或 TbStt3C 在一种高拷贝载体中被表达。本发明还涉及制造这些细胞的方法。

[0299] 在另一个优选实施方案中, 所述细胞为一种突变菌株, 其 (i) 至少缺乏 Alg11 型活性物和脂质连接单糖 (LLM) 翻转酶活性物, (ii) 表达本发明所述的新的脂质连接寡糖翻转酶活性物, 及 (iii) 表达原生动物寡糖基转移酶活性物。更特别地, 所述细胞 (i) 为 alg11 基因和 / 或 alg11 同源基因及一个或多个编码脂质连接单糖 (LLM) 翻转酶活性物的基因的敲除突变株, (ii) 表达一种或多种提供脂质连接寡糖翻转酶活性物的核酸分子, 及 (iii) 表达一种或多种上述原生动物寡糖基转移酶活性物。在一个更具体的实施方案中, 所述细胞表达一种或多种来源于 flc2' 的提供脂质连接寡糖翻转酶活性物的核酸分子, 如上文中所详细描述。可替代地或附加地, 所述细胞表达一种或多种来源于 rft1 的提供脂质连接寡糖翻转酶活性物的核酸分子, 如上文中所描述的。本发明所述的一个优选突变细胞表达 Leishmania major 的旁系同源基因 LmStt3D。在一个具体的变型中, LmStt3D 在一种低拷贝载体中被表达。在另一个优选实施方案中, 所述突变细胞表达 Leishmania braziliensis 的旁系同源基因 LbStt3-3。在一个具体的变型中, LbStt3-3 在一种低拷贝载体中被表达。在另一个优选实施方案中, 所述突变细胞表达 Trypanosoma brucei 的旁系同源基因 TbStt3B 或 TbStt3C。在一个具体的变型中, TbStt3B 或 TbStt3C 在一种高拷贝载体中被表达。本发明还涉及制造这些细胞的方法。特别地, 所述细胞能合成具有 Man3GlcNAc2 结构的脂质连接寡糖, 并能转移所述结构至新生蛋白质。本发明还涉及制造所述细胞的方法。

[0300] 在另一个优选实施方案中, 所述细胞为一种突变菌株, 其 (i) 至少缺乏 Alg11 型活性物和 β -D-甘露糖基转移酶 (DPM1) 型活性物, (ii) 表达本发明所述的新的脂质连接寡糖翻转酶活性物, 及 (iii) 表达原生动物寡糖基转移酶活性物。更特别地, 所述细胞 (i) 为 alg11 基因和 / 或 alg11 同源基因及 dpm1 基因和 / 或 dpm1 同源基因的敲除突变株, (ii) 表达一种或多种提供脂质连接寡糖翻转酶活性物的核酸分子, 及 (iii) 表达一种或多种上述原生动物寡糖基转移酶活性物。所述细胞表达一种或多种来源于 flc2' 的提供脂质连接寡糖翻转酶活性物的核酸分子, 如上文中所详细描述的。可替代地或附加地, 所述细胞表达一种或多种来源于 rft1 的提供脂质连接寡糖翻转酶活性物的核酸分子, 如上文中所描述的。本发明所述的一个优选突变细胞表达 Leishmania major 的旁系同源基因 LmStt3D。在一个具体的变型中, LmStt3D 在一种低拷贝载体中被表达。在另一个优选实施方案中, 所述突变细胞表达 Leishmania braziliensis 的旁系同源基因 LbStt3-3。在一个具体的变型中, LbStt3-3 在一种低拷贝载体中被表达。在另一个优选实施方案中, 所述突变细胞表达 Trypanosoma brucei 的旁系同源基因 TbStt3B 或 TbStt3C。在一个具体的变型中, TbStt3B 或 TbStt3C 在一种高拷贝载体中被表达。本发明还涉及制造这些细胞的方法。特别地, 所述细胞能合成具有 Man3GlcNAc2 结构的脂质连接寡糖, 并能转移所述结构至新生蛋白质。本发明还涉及制造所述细胞的方法。

[0301] 具体地讲, 提供了一种细胞, 其中一种或多种编码内源性寡糖基转移酶亚单位的基因被敲除。在包括酵母细胞的优选变型中, 内源性寡糖基转移酶的至少一种亚单位选自 Wbplp, Ost1p, Ost2p, Ost3p, Ost4p, Ost5p, Ost6p, Swplp, 和 Stt3p。在一个优选实施方案中, 所述细胞为 wbp1 基因和 stt3 基因的敲除突变株。在另一个优选实施方案中, 所述细胞为 ost1 基因和 ost2 基因的敲除突变株。

[0302] 在本发明更优选的实施方案中,任何一种如上所述的细胞可以还包含至少一种编码异源糖蛋白的核酸。表达异源糖蛋白的启动子可以为内源性启动子,内源性是针对所述酶被表达时所在的细胞而言的。在另一个优选实施方案中,启动子为一种异源启动子、诱导型或构建型启动子,其超表达核酸分子的一个或多个拷贝。特别地,这些细胞能主要合成具有 Man1-3GlcNAc2 结构的脂质连接寡糖,并能转移所述结构至所述异源蛋白质。

[0303] 不希望受理论的束缚,上述指定的敲除缺失菌株仅能够在内质网上或内质网中制造低甘露糖脂质连接寡糖,尤其是 Man3GlcNAc2,其然后被附加到内质网中的蛋白质上。在一些情况下可以发现,随后在高尔基体中甘露糖基转移酶作用下添加附加的甘露糖残基,可能在蛋白质上形成 Man4GlcNAc2 和 Man5GlcNAc2 结构。为了减少不期望的 Man4GlcNAc2 和 Man5GlcNAc2 结构的数量,本发明提供了避免其发生的措施。一个优选的措施为在如上所述的本发明任何一种细胞中缺失一种或多种编码定位于高尔基体的甘露糖基转移酶的基因。

[0304] 本发明与现有技术的在先教导明确相反,其中,通过使用同源或异源甘露糖苷酶修剪或分裂高甘露糖(例如,Man8GlcNAc2 或 Man 9GlcNAc2)或高甘露糖基化糖形,得到期望的低甘露糖基化多糖。因此,在一个优选实施方案中,本发明涉及不表现出有效的甘露糖苷酶活性或根本无甘露糖苷酶活性的细胞。

[0305] 具有修饰的高尔基体-糖基化的宿主细胞

[0306] 在内质网由寡糖基转移酶产生的主要糖蛋白可以在高尔基体中进一步糖基化,如下文详细描述。本发明更主要的方面在于提供用于在本发明所述的宿主细胞中修饰基于高尔基体的糖基化的手段和方法。上文中详细描述的基于内质网的糖基化的修饰及下文中详细描述的基于高尔基体的糖基化的修饰是密不可分的。本发明有利地提供了具有低甘露糖多糖结构的主要糖蛋白,为随后在高尔基体中的修饰糖基化形成理想的底物

[0307] 进一步缺乏定位于高尔基体的甘露糖基转移酶的宿主细胞

[0308] 在优选实施方案中,本发明所述的宿主细胞被进一步修饰或基因工程改造,以缺乏或被抑制或耗竭一种或多种,至少两种,优选至少三种、至少四种或至少五种定位于高尔基体的甘露糖基转移酶。优选地,所述甘露糖基转移酶选自:Och1p, Mnn1p, Mnn2p, Mnn4p, Mnn5p, Mnn9p, Mnn10p, 及 Mnn11p, 及其同源物(见表 2)。优选地,所述细胞为至少一种基因的敲除突变株,所述基因选自:och1, mnn1, mnn2, mnn4, mnn5, mnn9, mnn10, 和 mnn11 基因及其同源基因。同源基因还包括相同或相关基因家族的其他成员。

[0309] 表 2 定位于高尔基体的甘露糖基转移酶

[0310]

名称	功能	EC 号	同义词
Och1	α -1, 6-甘露糖基转移酶	2. 4. 1. 232	YGL048C

Mnn1	α -1, 3-甘露糖基转移酶	2. 4. 1. -	YER001W
Mnn2	α 1, 2-甘露糖基转移酶	2. 4. 1. -	YBR015C, TTP1, CRV4, LDB8
Mnn4	甘露糖基磷酸转移酶调节子	2. 4. 1. -	YKL201C
Mnn5	α 1, 2-甘露糖基转移酶	2. 4. 1. -	YJL186W
Mnn6	甘露糖基磷酸转移酶	2. 4. 1. -	KTR6, YPL053C
Mnn8	α -1, 6 甘露糖基转移酶复合物	2. 4. 1. -	ANP1
Mnn9	高尔基体甘露糖基转移酶复合物的亚单位	2. 4. 1. -	YPL050C
Mnn10	高尔基体甘露糖基转移酶复合物的亚单位	2. 4. 1. -	YDR245W, BED1, SLC2, REC41
Mnn11	高尔基体甘露糖基转移酶复合物的亚单位	2. 4. 1. -	YJL183W

Ktr1	A-1, 2-甘露糖基转移酶	2. 4. 1. -	YOR099W
Ktr2	甘露糖基转移酶	2. 4. 1. -	YKR061W
Ktr3	假定的 α -1, 2-甘露糖基转移酶	2. 4. 1. -	YBR205W
Ktr4	假定的甘露糖基转移酶	2. 4. 1. -	YBR199W
Ktr5	假定的甘露糖基转移酶	2. 4. 1. -	YNL029C
Ktr6	可能的甘露糖基转移酶	2. 4. 1. -	YPL053C (Mnn6)
Ktr7	假定的甘露糖基转移酶	2. 4. 1. -	YIL085C

Van1	甘露聚糖聚合酶 I 组分		YML115C
Vrg4	高尔基体 GDP-甘露糖转运体		YGL225W

[0311] 所述细胞可以为 och1, or mnn1, mnn2, mnn4, mnn5, mnn9, mnn10, mnn11 和 / 或它们的同源基因中至少一种基因的敲除突变株。所述细胞还可以为 ktr1, ktr2, ktr3, ktr4, ktr5, ktr6, ktr7 和 / 或它们的同源基因中至少一种基因的敲除突变株。所述细胞还可以为 van1, vrg4 和 / 或它们的同源基因中至少一种基因的敲除突变株。

[0312] 在一个优选实施方案中, 本发明所述的细胞尤其是上述复合系统, 进一步缺乏至少一种 Och1, 更特别是 α -1, 6-甘露糖基转移酶。更特别地, 所述细胞进一步为 och1 基因的敲除突变株。例如, 本发明所述的复合系统能在 *Pichia pastoris* 的高甘露糖基化-阴性 (Och1) 突变菌株的基础上进行基因工程改造。在一个优选实施方案中, 本发明所述的细胞尤其是上述复合系统, 至少缺乏由 mnn1 基因或其同源基因提供的 α -1, 3-甘露糖基转移酶。更特别是至少 mnn1 基因或其同源基因的敲除突变株。所述细胞还可以缺乏一种或多种

上述甘露糖基转移酶,更特别地,所述细胞为一种或多种编码所述甘露糖基转移酶的基因,尤其是选自 *mn9*, *mn5*, *van1* 及其同源基因中的一种或多种基因的敲除突变株。

[0313] 在一个优选实施方案中,所述细胞为至少缺乏 *Alg11* 及 *Mnn1* 的突变菌株。更特别地,所述细胞为至少 *alg11* 及 *mn1* 的敲除突变株。一个优选的实施方案为突变细胞,优选为酵母细胞,其为一种复合系统,所述复合系统:

[0314] (i) 被修饰以表达至少一种新的脂质连接寡糖翻转酶活性物,尤其是由一种或多种本发明所述的核酸分子编码的脂质连接寡糖翻转酶活性物,并且所述复合系统是 *alg11* 或其同源基因的敲除突变株;

[0315] (ii) 是至少 *mn1* 或其同源基因的敲除突变株;及

[0316] (iiia) 进一步表达或超表达至少一种上述原生动物寡糖基转移酶活性物;及,可替代地或附加地,

[0317] (iiib) 进一步表达或超表达至少一种上述脂质连接寡糖翻转酶活性物。

[0318] 特别地,所述细胞能主要合成具有 *Man3GlcNAc2* 结构的脂质连接寡糖,并能转移所述结构至新生蛋白质。本发明还涉及制造所述细胞的方法。

[0319] 在一个优选实施方案中,所述细胞为至少缺乏 *Alg3*、*Alg11* 及 *Mnn1* 的突变菌株。更特别地,所述细胞为至少 *alg11*、*alg3* 及 *mn1* 的敲除突变株。一个优选的实施方案为突变细胞,优选为酵母细胞,其为一种复合系统,所述复合系统:

[0320] (i) 被修饰以表达至少一种新的脂质连接寡糖翻转酶活性物,尤其是由一种或多种本发明所述的核酸分子编码的脂质连接寡糖翻转酶活性物,并且所述复合系统是 *alg3* 及 *alg11* 或其它它们的同源基因的敲除突变株;

[0321] (ii) 是至少 *mn1* 或其同源基因的敲除突变株;及

[0322] (iiia) 进一步表达或超表达至少一种上述原生动物寡糖基转移酶活性物;及,可替代地或附加地,

[0323] (iiib) 进一步表达或超表达至少一种上述脂质连接寡糖翻转酶活性物。

[0324] 特别地,所述细胞能主要合成具有 *Man3GlcNAc2*, *Man6GlcNAc2*, *Man7GlcNAc2*, 或 *Man8GlcNAc2* 结构的脂质连接寡糖,并能转移所述结构至新生蛋白质。本发明还涉及制造所述细胞的方法。

[0325] 在另一个优选实施方案中,所述细胞为至少缺乏 *Alg11*、*DPM1* 及 *Mnn1* 的突变菌株。更特别地,所述细胞为至少 *alg11*、*dpm1* 及 *mn1* 的敲除突变株。一个优选的实施方案为突变细胞,优选为酵母细胞,其为一种复合系统,所述复合系统:

[0326] (i) 被修饰以表达至少一种新的脂质连接寡糖翻转酶活性物,尤其是由一种或多种本发明所述的核酸分子编码的脂质连接寡糖翻转酶活性物,并且所述复合系统是 *dpm1* 及 *alg11* 或其它它们的同源基因的敲除突变株;

[0327] (ii) 是至少 *mn1* 或其同源基因的敲除突变株;及

[0328] (iiia) 进一步表达或超表达至少一种上述原生动物寡糖基转移酶活性物;及,可替代地或附加地,

[0329] (iiib) 进一步表达或超表达至少一种上述脂质连接寡糖翻转酶活性物。

[0330] 在具体实施方案中,这些细胞表达或超表达一种或多种来源于 *f1c2'* 的核酸分子,如上文中所详细描述,所述核酸分子提供新的脂质连接寡糖翻转酶活性物。可替代地或

附加地,所述细胞表达一种或多种来源于 *rft1* 的核酸分子,如上所述,其提供脂质连接寡糖翻转酶活性物。

[0331] 在具体实施方案中,这些细胞表达或超表达 *Leishmania major* 的旁系同源基因 *LmStt3D*。在一个具体的变型中,*LmStt3D* 在一种低拷贝载体中被表达。在另一个优选实施方案中,所述突变细胞表达 *Leishmania braziliensis* 的旁系同源基因 *LbStt3-3*。在一个具体的变型中,*LbStt3-3* 在一种低拷贝载体中被表达。在另一个优选实施方案中,所述突变细胞表达 *Trypanosoma brucei* 的旁系同源基因 *TbStt3B* 或 *TbStt3C*。在一个具体的变型中,*TbStt3B* 或 *TbStt3C* 在一种高拷贝载体中被表达。本发明还涉及制造这些细胞的方法。

[0332] 在具体实施方案中,这些细胞还是内源性寡糖基转移酶活性物的敲除突变株,尤其是通过敲除 *ost1* 和 *ost2* 和 / 或 *wbp1* 和 *stt3* 和 / 或它们各自的同源基因。

[0333] 通过异源糖基转移酶表达对基于高尔基体的糖基化的具体调控

[0334] 如下文中详细描述,本发明所述的核酸分子或聚氨基酸分子的优选实施方案被用于制造修饰的宿主细胞,所述细胞被指定用于制造如下文所述的糖蛋白或糖蛋白组合物。

[0335] 本发明所述的细胞可以被进一步基因工程改造以改变高尔基体中的糖基化级联,其在不同的真核生物中明显不同,因此,所述糖蛋白根据它们被表达及被分离的细胞类型不同,其多糖结构会有区别。例如,低级真核生物通常制造含有 N- 多糖的高甘露糖。因此,本发明的另一个目标是提供一种可用于在除人细胞之外的细胞中制造糖蛋白的细胞,及能够在除人细胞之外的细胞中制造糖蛋白的方法,所述糖蛋白具有特定类型 N- 多糖结构,例如人多糖结构。因此,这种细胞将会在高尔基体糖基化途径被进一步基因修饰,允许细胞执行模拟糖蛋白例如在人类中的加工的酶反应的序列。在这些基因工程细胞中表达的重组蛋白制造与其人类配对物更相似的糖蛋白,如果不能基本上相同。如果使用如上列举的通常制造含有 N- 多糖的高甘露糖的低级真核细胞,所述细胞被修饰以制造 N- 多糖例如 *Man3GlcNAc2* 或 *Man5GlcNAc2* 或人类糖基化途径中的其他结构。优选实施例包括但不限于,包含一种或多种多糖结构的重组蛋白,所述多糖结构选自:

[0336] *GlcNAcMan3-5GlcNAc2*,

[0337] *GlcNAc2Man3GlcNAc2*,

[0338] *GlcNAc3Man3GlcNAc2*- 平分型

[0339] *Gal2GlcNAc2Man3GlcNAc2*,

[0340] *Gal2GlcNAc2Man3GlcNAc2Fuc*,

[0341] *Gal2GlcNAc3Man3GlcNAc2*- 平分型,

[0342] *Gal2GlcNAc3Man3GlcNAc2Fuc*- 平分型,

[0343] *NeuAc2Gal2GlcNAc2Man3GlcNAc2*,

[0344] *NeuAc2Gal2GlcNAc2Man3GlcNAc2Fuc*,

[0345] *NeuAc2Gal2GlcNAc3Man3GlcNAc2*- 平分型,

[0346] *euAc2Gal2GlcNAc3Man3GlcNAc2Fuc*- 平分型,

[0347] *GlcNAc3Man3GlcNAc2*,

[0348] *Gal3GlcNAc3Man3GlcNAc2*,

[0349] *Gal3GlcNAc3Man3GlcNAc2Fuc*,

[0350] NeuAc3Gal3GlcNAc3Man3GlcNAc2, 及

[0351] NeuAc3Gal3GlcNAc3Man3GlcNAc2Fuc

[0352] 本文所用的 GlcNAc 为 N- 乙酰葡萄糖胺, Gal 为半乳糖, Fuc 为岩藻糖, NeuAc 为 N- 乙酰神经氨酸或唾液酸。在优选实施方案中, 本文所用的所有多糖结构中都缺少岩藻糖, 除非特别例证存在岩藻糖 (Fuc)。

[0353] 根据本发明, 优选地, 通过缺乏特定酶的菌株的基因工程和 / 或选择, 创造了低级真核生物尤其是真菌细胞例如酵母的糖蛋白中不期望的高甘露糖型结构特征。优选地, 通过基因工程改造表达异源酶的宿主细胞, 所述异源酶制造不被创造高甘露糖型的酶识别的多糖结构, 其被选择以在期望具有活性的低级真核细胞例如真菌中存在时, 发挥最佳活性, 或者以达到最佳活性的细胞器为靶点, 及它们的组合, 其中, 基因工程的真核生物表达制造“类人”糖蛋白所必需的多重异源酶。

[0354] 在优选实施方案中, 本发明还涉及高尔基体中一种或多种异源酶的整合, 所述酶能够制造“类人”N- 多糖。在优选实施方案中, 本发明提供了基因工程细胞, 在高尔基体中包含至少一种异源糖基转移酶和 / 或一种或多种选自表 3、4 和 5 中所列举的糖基转移酶相关酶。

[0355] 类人糖基化的主要特征在于, 包含 N- 乙酰葡萄糖胺、半乳糖、岩藻糖和 / 或 N- 乙酰神经氨酸的复合 N- 多糖结构。其他唾液酸如存在于其他动物如仓鼠的 N- 多糖中的 N- 糖基神经氨酸, 在人类中是不存在的。还有特殊的寡糖基连接例如末端结合的 α -1-3 半乳糖对于啮齿动物是典型的, 但在人细胞中是不存在的。

[0356] 表 3 异源糖基转移酶、转运体和相关酶

[0357]

名称	功能/酶	定位	EC 号	同义词(s)	基因, 模板
GnTI	甘露糖基(α -1,3)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶	高尔基体	2.4.1.101	GlcNAc 转移酶 1, α -1,3-甘露糖基-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶	Mgat1
GnTII	甘露糖基(α -1,6)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶	高尔基体	2.4.1.143	GlcNAc 转移酶 2, N-乙酰葡萄糖氨基转移酶 II, UDP-GlcNAc: 甘露糖苷 α -1-6 乙酰葡萄糖氨基转移酶, A-1,6-甘露糖基-糖蛋白 2- β -N-乙酰葡萄糖氨基转移酶	Mgat2
GnTIII	β -1,4-甘露糖基-糖蛋白 4- β -N-乙酰葡萄糖氨基转移酶	高尔基体	2.4.1.144	GlcNAc 转移酶 3, N-乙酰葡萄糖氨基转移酶 III	Mgat3
GnTIV	甘露糖基(α -1,3)-糖蛋白 β -1,4-N-乙酰葡萄糖氨基转移酶	高尔基体	2.4.1.145	GlcNAc 转移酶 4, N-乙酰葡萄糖氨基转移酶 IV, A-1,3-甘露糖基-糖蛋白 4- β -N-乙酰葡萄糖氨基转移酶, 同工酶 A 和 B	Mgat4
GnTV	甘露糖基(α -1,6)-糖蛋白 β -1,6-N-乙酰葡萄糖氨基转移酶	高尔基体	2.4.1.155	GlcNAc 转移酶 5, N-乙酰葡萄糖氨基转移酶 V, A-1,6-甘露糖基-糖蛋白 6- β -N-乙酰葡萄糖氨基转移酶	Mgat5
GnTVI	α -1,6-甘露糖基-糖蛋白 4- β -N-乙酰葡萄糖氨基转移酶	高尔基体	2.4.1.201	GlcNAc 转移酶 6, N-乙酰葡萄糖氨基转移酶 VI	Mgat6
GalT	β -N-乙酰葡萄糖氨基糖肽 β -1,4-半乳糖基转移酶	高尔基体	2.4.1.38	Gal-转移酶 8, UDP-Gal 转移酶	B4galT1
FucT	α (1,6) 岩藻糖基转移酶	高尔基体	2.4.1.68	Fuc-t 转移酶 8, GDP-Fuc 转移酶	Fut8
ST	β -半乳糖苷 α -2,6-唾液酸转移酶	高尔基体	2.4.99.1	唾液酸转移酶, CMP-N-乙酰神经氨酸- β -半乳糖	ST6gal1

[0358]

				胺- α -2,6-唾液 酸转移酶	
	UDP-N-乙酰葡萄糖胺 2-差向异构酶	细胞溶质	5.1.3. 14	UDP-GlcNAc-2-差 向异构酶	NeuC
	唾液酸合酶	细胞溶质			NeuB
	CMP-NeuNAc 合成酶	细胞溶质	2.7.7. 43		Cmas NeuA
	N-酰基神经氨酸-9- 磷酸合酶		2.5.1. 57		
	N-酰基神经氨酸-9- 磷酸酶		3.1.3. 29		
	UDP-GlcNAc 转运体	高尔基体			S1c35A3
	UDP-半乳糖转运体	高尔基体			S1c35A2
	GDP-岩藻糖转运体	高尔基体			S1c35C1
	CMP-唾液酸转运体	高尔基体			S1c35A1
	核苷酸二磷酸酶	高尔基体			
	GDP-D-甘露糖 4,6- 脱水酶	细胞溶质	4.2.1. 47		Gmds
	GDP-4-酮-去氧-D-甘 露糖-3,5-差向异构 酶-4-还原酶	细胞溶质	1.1.1. 271	GDP L-岩藻糖合 酶, FX 蛋白	Tsta3

[0359]

表 4 用于优选双触角型多糖的基于高尔基体的合成的异源酶

N-乙酰葡萄糖氨基糖基化	平分型 GlcNAc	半乳糖基化	岩藻糖基化	唾液酸化
GlcNAcMan3-5GlcNAc2				
甘露糖基 (α-1, 3-)-糖蛋白 β-1, 2-N-乙酰葡萄糖氨基转移酶 (GnTI)				
UDP-N-乙酰葡萄糖胺转运体				
GlcNAc2Man3GlcNAc2				
甘露糖基 (α-1, 3-)-糖蛋白 β-1, 2-N-乙酰葡萄糖氨基转移酶 (GnTI)				
UDP-N-乙酰葡萄糖胺转运体				
甘露糖基 (α-1, 6-)-糖蛋白 β-1, 2-N-乙酰葡萄糖氨基转移酶 (GnTII)				
GlcNAc3Man3GlcNAc2-平分型				
甘露糖基 (α-1, 3-)-糖蛋白 β-1, 2-N-乙酰葡萄糖氨基转移酶 (GnTI)	β-1, 4-甘露糖基-糖蛋白 4-β-N-乙酰葡萄糖氨基转移酶 (GnTIII)			
UDP-N-乙酰葡萄糖胺转运体				
甘露糖基 (α-1, 6-)-糖蛋白 β-1, 2-N-乙酰葡萄糖氨基转移酶 (GnTII)				
Gal12GlcNAc2Man3GlcNAc2				
甘露糖基 (α-1, 3-)-糖蛋白 β-1, 2-N-乙酰葡萄糖氨基转移酶 (GnTI)		β-N-乙酰葡萄糖氨基糖肽 β-1, 4-半乳糖基转移酶 (GalT)		
UDP-N-乙酰葡萄糖胺转运体		UDP-半乳糖转运体		
甘露糖基 (α-1, 6-)-糖蛋白 β-1, 2-N-乙酰葡萄糖氨基转移酶 (GnTII)				

[0360]

续.

N-乙酰葡萄糖氨基化	平分型 GlcNAc	半乳糖基化	岩藻糖基化	唾液酸化
Ga12G1cNAc2Man3G1cNAc2Fuc				
β -1,2-N-乙酰葡萄糖氨基转移酶 (GnTI) UDP-N-乙酰葡萄糖胺转运体 甘露糖基 (α -1,6)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶 (GnTII)		β -N-乙酰葡萄糖氨基糖 肽 β -1,4-半乳糖基转移酶 (GalT) UDP-半乳糖转运体	GDP-D-甘露糖 4,6-脱水酶 GDP-4-酮-去氧-D-甘露糖-3,5-差向异构酶-4-还原酶 GDP-岩藻糖转运体 α (1,6) 岩藻糖基转移酶 (FucT)	
Ga12G1cNAc3Man3G1cNAc2-平分型				
β -1,2-N-乙酰葡萄糖氨基转移酶 (GnTI) UDP-N-乙酰葡萄糖胺转运体 甘露糖基 (α -1,6)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶 (GnTII)	β -1,4-甘露糖基-糖蛋白 4- β -N-乙酰葡萄糖氨基转移酶 (GnTIII)	β -N-乙酰葡萄糖氨基糖 肽 β -1,4-半乳糖基转移酶 (GalT) UDP-半乳糖转运体		
Ga12G1cNAc3Man3G1cNAc2Fuc-平分型				
β -1,2-N-乙酰葡萄糖氨基转移酶 (GnTI) UDP-N-乙酰葡萄糖胺转运体 甘露糖基 (α -1,6)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶 (GnTII)	β -1,4-甘露糖基-糖蛋白 4- β -N-乙酰葡萄糖氨基转移酶 (GnTIII)	β -N-乙酰葡萄糖氨基糖 肽 β -1,4-半乳糖基转移酶 (GalT) UDP-半乳糖转运体	GDP-D-甘露糖 4,6-脱水酶 GDP-4-酮-去氧-D-甘露糖-3,5-差向异构酶-4-还原酶 GDP-岩藻糖转运体 α (1,6) 岩藻糖基转移酶 (FucT)	

[0361]

续.

N-乙酰葡萄糖氨基糖基化	平分型 GlcNAc	半乳糖基化	岩藻糖基化	唾液酸化
NeuAc2Gal12GlcNAc2Man3GlcNAc2				
β-1,2-N-乙酰葡萄糖氨基糖基转移酶 (GnTI) UDP-N-乙酰葡萄糖胺转运体 甘露糖基(α-1,6-)-糖蛋白 β-1,2-N-乙酰葡萄糖氨基糖基转移酶 (GnTII)		β-N-乙酰葡萄糖氨基糖基转移酶 β-1,4-半乳糖基转移酶 (GalT) UDP-半乳糖转运体		β-半乳糖苷 α-2,6-唾液酸转移酶 (ST) UDP-N-乙酰葡萄糖胺 2-差向异构酶 (NeuC) 唾液酸合酶 (NeuB) 或: N-酰基神经氨酸-9-磷酸合酶 N-酰基神经氨酸-9-磷酸酶 CMP-Neu5Ac 合成酶 CMP-唾液酸转运体
NeuAc2Gal12GlcNAc2Man3GlcNAc2Fuc				
β-1,2-N-乙酰葡萄糖氨基糖基转移酶 (GnTI) UDP-N-乙酰葡萄糖胺转运体 甘露糖基(α-1,6-)-糖蛋白 β-1,2-N-乙酰葡萄糖氨基糖基转移酶 (GnTII)		β-N-乙酰葡萄糖氨基糖基转移酶 β-1,4-半乳糖基转移酶 (GalT) UDP-半乳糖转运体	GDP-D-甘露糖 4,6-脱水酶 GDP-4-酮-去氧-D-甘露糖-3,5-差向异构酶-4-还原酶 GDP-岩藻糖转运体 α(1,6)岩藻糖基转移酶 (FucT)	β-半乳糖苷 α-2,6-唾液酸转移酶 (ST) UDP-N-乙酰葡萄糖胺 2-差向异构酶 (NeuC) 唾液酸合酶 (NeuB) 或: N-酰基神经氨酸-9-磷酸合酶 N-酰基神经氨酸-9-磷酸酶 CMP-Neu5Ac 合成酶 CMP-唾液酸转运体

[0362]

续.

N-乙酰葡萄糖氨基糖基化	平分型 GlcNAc	半乳糖基化	岩藻糖基化	唾液酸化
NeuAc2Gal12GlcNAc3Man3GlcNAc2-平分型				
β-1, 2-N-乙酰葡萄糖氨基转移酶 (GnTI) UDP-N-乙酰葡萄糖胺转运体 甘露糖基 (α-1, 6-)-糖蛋白 β-1, 2-N-乙酰葡萄糖氨基转移酶 (GnTII)	β-1, 4-甘露糖基-糖蛋白 4-β-N-乙酰葡萄糖氨基转移酶 (GnTIII)	β-N-乙酰葡萄糖氨基糖肽 β-1, 4-半乳糖基转移酶 (GalT) UDP-半乳糖转运体		β-半乳糖苷 α-2, 6-唾液酸转移酶 (ST) UDP-N-乙酰葡萄糖胺 2-差向异构酶 (NeuC) 唾液酸合酶 (NeuB) 或: N-酰基神经氨酸-9-磷酸合酶 N-酰基神经氨酸-9-磷酸酶 CMP-Neu5Ac 合成酶 CMP-唾液酸转运体

续.

N-乙酰葡萄糖氨基糖基化	平分型 GlcNAc	半乳糖基化	岩藻糖基化	唾液酸化
NeuAc2Gal12GlcNAc3Man3GlcNAc2Fuc-平分型				
β-1, 2-N-乙酰葡萄糖氨基转移酶 (GnTI) UDP-N-乙酰葡萄糖胺转运体 甘露糖基 (α-1, 6-)-糖蛋白 β-1, 2-N-乙酰葡萄糖氨基转移酶 (GnTII)	β-1, 4-甘露糖基-糖蛋白 4-β-N-乙酰葡萄糖氨基转移酶 (GnTIII)	β-N-乙酰葡萄糖氨基糖肽 β-1, 4-半乳糖基转移酶 (GalT) UDP-半乳糖转运体	GDP-D-甘露糖 4, 6-脱水酶 GDP-4-酮-去氧-D-甘露糖-3, 5-差向异构酶-4-还原酶 GDP-岩藻糖转运体 α (1, 6) 岩藻糖基转移酶 (FucT)	β-半乳糖苷 α-2, 6-唾液酸转移酶 (ST) UDP-N-乙酰葡萄糖胺 2-差向异构酶 (NeuC) 唾液酸合酶 (NeuB) 或: N-酰基神经氨酸-9-磷酸合酶 N-酰基神经氨酸-9-磷酸酶 CMP-Neu5Ac 合成酶 CMP-唾液酸转运体

[0363]

表 5 用于优选三触角型多糖的基于高尔基体的合成的异源酶

N-乙酰葡萄糖氨基糖基化		半乳糖基化		岩藻糖基化		唾液酸化	
GlcNAc3Man3GlcNAc2							
β-1, 2-N-乙酰葡萄糖氨基转移酶 (GnTI) UDP-N-乙酰葡萄糖胺转运体 甘露糖基 (α-1, 6-)-糖蛋白 β-1, 2-N-乙酰葡萄糖氨基转移酶 (GnTII) 甘露糖基 (α-1, 3-)-糖蛋白 β-1, 4-N-乙酰葡萄糖氨基转移酶 (GnTIV)							
Gal3GlcNAc3Man3GlcNAc2							
β-1, 2-N-乙酰葡萄糖氨基转移酶 (GnTI) UDP-N-乙酰葡萄糖胺转运体 甘露糖基 (α-1, 6-)-糖蛋白 β-1, 2-N-乙酰葡萄糖氨基转移酶 (GnTII) 甘露糖基 (α-1, 3-)-糖蛋白 β-1, 4-N-乙酰葡萄糖氨基转移酶 (GnTIV)		β-N-乙酰葡萄糖氨基糖肽 β-1, 4-半乳糖基转移酶 (GalT) UDP-半乳糖转运体					
Gal3GlcNAc3Man3GlcNAc2Fuc							
β-1, 2-N-乙酰葡萄糖氨基转移酶 (GnTI) UDP-N-乙酰葡萄糖胺转运体 甘露糖基 (α-1, 6-)-糖蛋白 β-1, 2-N-乙酰葡萄糖氨基转移酶 (GnTII) 甘露糖基 (α-1, 3-)-糖蛋白 β-1, 4-N-乙酰葡萄糖氨基转移酶 (GnTIV)		β-N-乙酰葡萄糖氨基糖肽 β-1, 4-半乳糖基转移酶 (GalT) UDP-半乳糖转运体		GDP-D-甘露糖 4, 6-脱水酶 GDP-4-酮-去氧-D-甘露糖-3, 5-差向异构酶-4-还原酶 GDP-岩藻糖转运体 α (1, 6) 岩藻糖基转移酶 (FucT)			

[0364]

续.			
NeuAc3Gal13GlcNAc3Man3GlcNAc			
β-1,2-N-乙酰葡萄糖胺转移酶 (GnTI) UDP-N-乙酰葡萄糖胺转运体 甘露糖基(α-1,6-)-糖蛋白 β-1,2-N-乙酰葡萄糖胺转移酶 (GnTII) 甘露糖基(α-1,3-)-糖蛋白 β-1,4-N-乙酰葡萄糖胺转移酶 (GnTIV)	β-N-乙酰葡萄糖胺基糖肽 β-1,4-半乳糖基转移酶 (GalT) UDP-半乳糖转运体		2,6-唾液酸转移酶 (ST) UDP-N-乙酰葡萄糖胺 2-差向异构酶 (NeuC) 唾液酸合酶 (NeuB) 或: N-乙酰神经氨酸-9-磷酸合酶+ N-乙酰神经氨酸-9-磷酸酶 CMP-Neu5Ac 合成酶 CMP-唾液酸转运体
NeuAc3Gal13GlcNAc3Man3GlcNAcFuc			
β-1,2-N-乙酰葡萄糖胺转移酶 (GnTI) UDP-N-乙酰葡萄糖胺转运体 甘露糖基(α-1,6-)-糖蛋白 β-1,2-N-乙酰葡萄糖胺转移酶 (GnTII) 甘露糖基(α-1,3-)-糖蛋白 β-1,4-N-乙酰葡萄糖胺转移酶 (GnTIV)	β-N-乙酰葡萄糖胺基糖肽 β-1,4-半乳糖基转移酶 (GalT) UDP-半乳糖转运体	GDP-D-甘露糖 4,6-脱水酶 GDP-4-酮-去氧-D-甘露糖 -3,5-差向异构酶-4-还原酶 GDP-岩藻糖转运体 α (1,6) 岩藻糖基转移酶 (FucT)	2,6-唾液酸转移酶 (ST) UDP-N-乙酰葡萄糖胺 2-差向异构酶 (NeuC) 唾液酸合酶 (NeuB) 或: N-乙酰神经氨酸-9-磷酸合酶+ N-乙酰神经氨酸-9-磷酸酶 CMP-Neu5Ac 合成酶 CMP-唾液酸转运体

[0365] 所述基因工程的主要目标是制造稳健的蛋白制造菌株,其能够在工业发酵工艺中完成具有限定的类人多糖结构。多重基因向宿主(例如真菌)染色体中的整合需要仔细的

计划。所述基因工程菌株将最可能被一系列不同的基因转化,并且这些基因将必须以稳定的模式被转化来确保整个发酵工艺中期望的被维持。所述酶的任何结合都要被工程整合入蛋白表达宿主细胞。

[0366] 采用 DNA 序列信息,技术人员能使用本领域技术人员公知的标准技术来克隆编码 GnT 的 DNA 分子,编码一种或多种 GnT(或编码其催化的片段)的核酸分子可以在启动子及其他表达调控序列的转录调控下被插入到合适的表达载体中,所述表达调控序列能在本发明所述的选定宿主细胞中,例如,如本文所述的 *Piichia* sp., *Kluyveromyces* sp., *Saccharomyces* sp., *Yarrowia* sp. 和 *Aspergillus* sp. 等真菌宿主中发起转录,这样,这些哺乳动物 GnT 酶的一种或多种可以在选定的宿主细胞中被主动表达来制造类人复合糖蛋白。

[0367] 所述工程菌株将被不同的糖基化相关基因稳定地转化,以确保在整个发酵工艺中期望的活性被维持。下述酶的任何结合都要被工程整合入表达宿主。并行地,不期望的糖基化反应中涉及的一些宿主基因将要被缺失。

[0368] 在优选实施方案中,转化编码异源糖基化酶的基因的一个子集,至少两种基因(又名文库)进入宿主生物体,起初引起一个基因混合群体。然后,从所述混合群体中选择具有期望的糖基化表型的转化株。在一个优选实施方案中,所述宿主生物体为低级真核生物,并且所述宿主糖基化途径通过一种或多种人或动物糖基化酶的稳定表达被修饰,所述糖基化酶制造与人多糖结构相似或相同的 N-多糖。在一个特别优选的实施方案中,所述基因的子集或“DNA 文库”包括编码糖基化酶的融合基因构造,所述糖基化酶带有针对糖基化涉及的各种细胞内场所尤其是内质网、顺面高尔基体、中间高尔基体或反面高尔基体的靶向序列。

[0369] 在一些情况下,所述 DNA 文库可以直接从现有或野生型基因中集合。但是,在一个优选实施方案中,所述 DNA 文库从两个或多个子文库的融合中集合得到。通过所述子文库的框架内连接,使创造大量新的编码有用的靶向糖基化的基因构造成为可能。例如,一种有用的子文库包括编码以下记载的酶和酶活性的任一组合的 DNA 序列。

[0370] 优选地,所述酶为人源的酶,虽然其他真核或原核酶,更尤其哺乳动物、原生动物、植物、细菌或真菌酶也有用。在一个优选实施方案中,基因被截短来提供编码酶催化域的片段。通过去除内源性靶向序列,所述酶可以在其他细胞内场所被重定向及表达。通过特殊环境的常识可以指导此类催化域的选择,在所述特殊环境下所述催化域随后被激活。另一种有用的子文库包括编码信号肽的 DNA 序列,其导致蛋白质向内质网、高尔基体或反面高尔基网内的特殊场所的定位。这些信号肽可以选自宿主生物体及其他相关或不相关的生物体。典型地,内质网或高尔基体的膜结合蛋白可以包括,例如,编码胞质尾区(ct)、跨膜域(tmd)和干区(sr)的 N-末端序列。所述 ct、tmd 及 sr 序列足以单独地或以组合的形式将蛋白质固定在细胞器的内(腔)膜上。因此,所述信号序列的子文库的一个优选实施例包括这些蛋白质中的 ct、tmd 和 / 或 sr 序列。在一些情况下,期望提供具有不同长度的 sr 序列的子文库。这可以通过 PCR 使用结合在编码胞质区的 DNA 的 5' 端的启动子并使用一系列结合在干区不同部位的反启动子来实现。所述信号序列的其他有用来源还包括回收信号肽。

[0371] 除开放阅读框序列之外,通常优选提供带有此类启动子、转录终止子、增强子、核

糖体结合位点及其他功能性序列的每个文库的构造,其对于确保基因在转化进入宿主生物体后有效的转录及翻译可能是必需的。

[0372] 据此,本发明进一步涉及,根据本发明所述的宿主细胞进一步被基因工程改造或修饰来表达至少一种优选的异源酶或其催化域,所述酶或其催化域示于表 3、4 和 5 中,并且优选地,选自基于高尔基体的异源酶:

[0373] 甘露糖基 (α -1,3)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶或 N-乙酰葡萄糖氨基转移酶 I (GnTI);

[0374] 甘露糖基 (α -1,6)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶或 N-乙酰葡萄糖氨基转移酶 II (GnTII);

[0375] β -1,4-甘露糖基-糖蛋白 4- β -N-乙酰葡萄糖氨基转移酶或 N-乙酰葡萄糖氨基转移酶 III (GnTIII);

[0376] 甘露糖基 (α -1,3)-糖蛋白 β -1,4-N-乙酰葡萄糖氨基转移酶或 N-乙酰葡萄糖氨基转移酶 IV (GnTIV);

[0377] 甘露糖基 (α -1,6)-糖蛋白 β -1,6-N-乙酰葡萄糖氨基转移酶或 N-乙酰葡萄糖氨基转移酶 V (GnTV);

[0378] α -1,6-甘露糖基-糖蛋白 4- β -N-乙酰葡萄糖氨基转移酶或 N-乙酰葡萄糖氨基转移酶 VI (GnTVI);

[0379] β -N-乙酰葡萄糖氨基糖肽 β -1,4-半乳糖基转移酶或半乳糖基转移酶 (GalT);

[0380] α (1,6) 岩藻糖基转移酶或岩藻糖基转移酶 (FucT);

[0381] β -半乳糖苷 α -2,6-唾液酸转移酶或唾液酸转移酶 (ST);

[0382] 这些酶可以进一步被下述中的一种或多种所支持:UDP-GlcNAc 转移酶;UDP-GlcNAc 转运体;UDP-半乳糖基转移酶;UDP-半乳糖转运体;GDP-岩藻糖基转移酶;GDP-岩藻糖转运体;CMP-唾液酸转移酶;CMP-唾液酸转运体;及核苷酸二磷酸酶。

[0383] 不言而喻,本文中所述至少一种酶或催化域优选包含至少一种用于细胞内膜或细胞器的定位序列。在优选实施方案中,所述细胞内膜或细胞器为高尔基体。

[0384] 在优选变型中,N-乙酰葡萄糖氨基转移酶 V (GnTV) 和 / 或 N-乙酰葡萄糖氨基转移酶 VI (GnTVI) 在被修饰的细胞中是不存在或缺乏的。在这些变型中,由这两种酶中的一或两种催化的修饰不是必需的,或是从基于高尔基体的修饰中排除在外的。

[0385] GlcNAcMan3-5GlcNAc2 结构合成的实施方案

[0386] 在一个优选实施方案中,被修饰的宿主细胞提供用于基于高尔基体的加工的酶,优选为异源酶,其选自:

[0387] 甘露糖基 (α -1,3)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶 (GnTI),尤其是 Mgat1 型转录物。

[0388] 所述细胞还可以包含一种酶,优选为异源酶,其选自:

[0389] UDP-N-乙酰葡萄糖胺转运体,尤其是 Slc35A3 型转录物。

[0390] 在一个最优选的实施方案中,所述细胞至少包含或只包含这两种高尔基体加工相关酶。

[0391] 在该实施方案的一个优选变型中,所述细胞表达下述基因中的一种或多种:

[0392] mgat1 和 slc35A3

[0393] 和 / 或它们的同源基因。

[0394] 具体地说,所述细胞能够制造具有 GlcNAcMan3-5GlcNAc2 结构的 N- 多糖。因此,本发明还涉及一种或多种宿主细胞,其被特别设计用来制造具有所述多糖结构的糖蛋白。因此,本发明还涉及一种优选为分离的,具有所述结构的糖蛋白,优选地,所述糖蛋白可以由所述细胞制造或者确实由所述细胞制造。本发明还提供了一种通过使用所述细胞来制作所述糖蛋白的方法或工艺。

[0395] GlcNAc2Man3GlcNAc2 结构合成的实施方案

[0396] 在另一个优选实施方案中,被修饰的宿主细胞提供用于基于高尔基体的加工的酶,优选为异源酶,其选自:

[0397] 甘露糖基 (α -1,3)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTI), 尤其是 Mgat1 型转录物;

[0398] UDP-N- 乙酰葡萄糖胺转运体, 尤其是 Slc35A3 型转录物; 及

[0399] 甘露糖基 (α -1,6)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTII), 尤其是 Mgat2 型转录物。

[0400] 在一个最优选的实施方案中,所述细胞至少包含或只包含这些高尔基体加工相关的酶。

[0401] 在该实施方案的一个优选变型中,所述细胞表达下述基因中的一种或多种:

[0402] mgat1, mgat2, 和 slc35A3

[0403] 和 / 或它们的同源基因。

[0404] 具体地说,所述细胞能够制造具有 GlcNAc2Man3GlcNAc2 结构的 N- 多糖。因此,本发明还涉及一种或多种宿主细胞,其被特别设计用来制造具有所述多糖结构的糖蛋白。因此,本发明还涉及一种优选为分离的,具有所述结构的糖蛋白,优选地,所述糖蛋白可以由所述细胞制造或者确实由所述细胞制造。本发明还提供了一种通过使用所述细胞来制作所述糖蛋白的方法或工艺。

[0405] GlcNAc3Man3GlcNAc2- 平分型结构合成的实施方案

[0406] 在另一个优选实施方案中,被修饰的宿主细胞提供用于基于高尔基体的加工的酶,优选为异源酶,其选自:

[0407] 甘露糖基 (α -1,3)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTI), 尤其是 Mgat1 型转录物;

[0408] UDP-N- 乙酰葡萄糖胺转运体, 尤其是 Slc35A3 型转录物;

[0409] 甘露糖基 (α -1,6)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTII), 尤其是 Mgat2 型转录物; 及

[0410] β -1,4- 甘露糖基 - 糖蛋白 4- β -N- 乙酰葡萄糖氨基转移酶 (GnTIII), 尤其是 Mgat3 型转录物。

[0411] 在一个最优选的实施方案中,所述细胞至少包含或只包含这些高尔基体加工相关的酶。

[0412] 在该实施方案的一个优选变型中,所述细胞表达下述基因中的一种或多种:

[0413] mgat1, mgat2, mgat3, 和 slc35A3

[0414] 和 / 或它们的同源基因。

[0415] 具体地说,所述细胞能够制造具有 GlcNAc2Man3GlcNAc2- 平分型结构的 N- 多糖。因此,本发明还涉及一种或多种宿主细胞,其被特别设计用来制造具有所述多糖结构的糖蛋白。因此,本发明还涉及一种优选为分离的,具有所述结构的糖蛋白,优选地,所述糖蛋白可以由所述细胞制造或者确实由所述细胞制造。本发明还提供了一种通过使用所述细胞来制作所述糖蛋白的方法或工艺。

[0416] Gal2GlcNAc2Man3GlcNAc2 结构合成的实施方案

[0417] 在另一个优选实施方案中,被修饰的宿主细胞提供用于基于高尔基体的加工的酶,优选为异源酶,其选自:

[0418] 甘露糖基 (α -1,3)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTI), 尤其是 Mgat1 型转录物;

[0419] UDP-N- 乙酰葡萄糖胺转运体, 尤其是 Slc35A3 型转录物;

[0420] 甘露糖基 (α -1,6)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTII), 尤其是 Mgat2 型转录物;

[0421] β -N- 乙酰葡萄糖氨基糖肽 β -1,4- 半乳糖基转移酶 (GalT), 尤其是 B4galT1 型转录物; 及

[0422] UDP- 半乳糖转运体, 尤其是 Slc35A2 型转录物。

[0423] 在一个最优选的实施方案中,所述细胞至少包含或只包含这些高尔基体加工相关的酶。

[0424] 在该实施方案的一个优选变型中,所述细胞表达下述基因中的一种或多种:

[0425] mgat1, mgat2, mgat3, b4galT1, 和 slc35a2

[0426] 和 / 或它们的同源基因。

[0427] 具体地说,所述细胞能够制造具有 Gal2GlcNAc2Man3GlcNAc2 结构的 N- 多糖。因此,本发明还涉及一种或多种宿主细胞,其被特别设计用来制造具有所述多糖结构的糖蛋白。因此,本发明还涉及一种优选为分离的,具有所述结构的糖蛋白,优选地,所述糖蛋白可以由所述细胞制造或者确实由所述细胞制造。本发明还提供了一种通过使用所述细胞来制作所述糖蛋白的方法或工艺。

[0428] Gal2GlcNAc2Man3GlcNAc2Fuc 结构合成的实施方案

[0429] 在另一个优选实施方案中,被修饰的宿主细胞提供用于基于高尔基体的加工的酶,优选为异源酶,其选自:

[0430] 甘露糖基 (α -1,3)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTI), 尤其是 Mgat1 型转录物;

[0431] UDP-N- 乙酰葡萄糖胺转运体, 尤其是 Slc35A3 型转录物;

[0432] 甘露糖基 (α -1,6)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTII), 尤其是 Mgat2 型转录物;

[0433] β -N- 乙酰葡萄糖氨基糖肽 β -1,4- 半乳糖基转移酶 (GalT), 尤其是 B4galT1 型转录物;

[0434] UDP- 半乳糖转运体, 尤其是 Slc35A2 型转录物;

[0435] GDP-D- 甘露糖 4,6- 脱水酶, 尤其是 Gmds 型转录物;

[0436] GDP-4- 酮 - 去氧 -D- 甘露糖 -3,5- 差向异构酶 -4- 还原酶, 尤其是 Tsta3 型转录

物；

[0437] GDP- 岩藻糖转运体，尤其是 Slc35C1 型转录物；及

[0438] α (1,6) 岩藻糖基转移酶 (FucT)，尤其是 Fut8 型转录物。

[0439] 在一个最优选的实施方案中，所述细胞至少包含或只包含这些高尔基体加工相关的酶。

[0440] 在该实施方案的一个优选变型中，所述细胞表达下述基因中的一种或多种：

[0441] mgat1, mgat2, slc35a3, mgat3, b4galt1, slc35a2, gmds, tsta3, slc35c1 和 fut8

[0442] 和 / 或它们的同源基因。

[0443] 具体地说，所述细胞能够制造具有 Gal2GlcNAc2Man3GlcNAc2Fuc 结构的 N- 多糖。因此，本发明还涉及一种或多种宿主细胞，其被特别设计用来制造具有所述多糖结构的糖蛋白。因此，本发明还涉及一种优选为分离的，具有所述结构的糖蛋白，优选地，所述糖蛋白可以由所述细胞制造或者确实由所述细胞制造。本发明还提供了一种通过使用所述细胞来制作所述糖蛋白的方法或工艺。

[0444] Gal2GlcNAc3Man3GlcNAc2- 平分型结构合成的实施方案

[0445] 在另一个优选实施方案中，被修饰的宿主细胞提供用于基于高尔基体的加工的酶，优选为异源酶，其选自：

[0446] 甘露糖基 (α -1,3-) - 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTI)，尤其是 Mgat1 型转录物；

[0447] UDP-N- 乙酰葡萄糖胺转运体，尤其是 Slc35A3 型转录物；

[0448] 甘露糖基 (α -1,6-) - 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTII)，尤其是 Mgat2 型转录物；

[0449] β -1,4- 甘露糖基 - 糖蛋白 4- β -N- 乙酰葡萄糖氨基转移酶 (GnTIII)，尤其是 Mgat3 型转录物；

[0450] β -N- 乙酰葡萄糖氨基糖肽 β -1,4- 半乳糖基转移酶 (GalT)，尤其是 B4galt1 型转录物；及

[0451] UDP- 半乳糖转运体，尤其是 Slc35A2 型转录物。

[0452] 在一个最优选的实施方案中，所述细胞至少包含或只包含这些高尔基体加工相关的酶。

[0453] 在该实施方案的一个优选变型中，所述细胞表达下述基因中的一种或多种：

[0454] mgat1, mgat2, mgat3, slc35a3, b4galt1, 和 slc35a2

[0455] 和 / 或它们的同源基因。

[0456] 具体地说，所述细胞能够制造具有 Gal2GlcNAc3Man3GlcNAc2- 平分型结构的 N- 多糖。因此，本发明还涉及一种或多种宿主细胞，其被特别设计用来制造具有所述多糖结构的糖蛋白。因此，本发明还涉及一种优选为分离的，具有所述结构的糖蛋白，优选地，所述糖蛋白可以由所述细胞制造或者确实由所述细胞制造。本发明还提供了一种通过使用所述细胞来制作所述糖蛋白的方法或工艺。

[0457] Gal2GlcNAc3Man3GlcNAc2Fuc- 平分型结构合成的实施方案

[0458] 在另一个优选实施方案中，被修饰的宿主细胞提供用于基于高尔基体的加工的酶，优选为异源酶，其选自：

[0459] 甘露糖基 (α -1,3)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶 (GnTI), 尤其是 Mgat1 型转录物;

[0460] UDP-N-乙酰葡萄糖胺转运体, 尤其是 Slc35A3 型转录物;

[0461] 甘露糖基 (α -1,6)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶 (GnTII), 尤其是 Mgat2 型转录物;

[0462] β -1,4-甘露糖基-糖蛋白 4- β -N-乙酰葡萄糖氨基转移酶 (GnTIII), 尤其是 Mgat3 型转录物;

[0463] β -N-乙酰葡萄糖氨基糖肽 β -1,4-半乳糖基转移酶 (GalT), 尤其是 B4gal1 型转录物;

[0464] UDP-半乳糖转运体, 尤其是 Slc35A2 型转录物;

[0465] GDP-D-甘露糖 4,6-脱水酶, 尤其是 Gmds 型转录物;

[0466] GDP-4-酮-去氧-D-甘露糖-3,5-差向异构酶-4-还原酶, 尤其是 Tsta3 型转录物;

[0467] GDP-岩藻糖转运体, 尤其是 Slc35C1 型转录物; 及

[0468] α (1,6) 岩藻糖基转移酶 (FucT), 尤其是 Fut8 型转录物。

[0469] 在一个最优选的实施方案中, 所述细胞至少包含或只包含这些高尔基体加工相关的酶。

[0470] 在该实施方案的一个优选变型中, 所述细胞表达下述基因中的一种或多种:

[0471] mgat1, mgat2, mgat3, slc3533, b4gal1, slc35a2, gmds, tsta3, slc35c1 和 fut8

[0472] 和 / 或它们的同源基因。

[0473] 具体地说, 所述细胞能够制造具有 Gal2GlcNAc3Man3GlcNAc2Fuc- 平分型结构的 N-多糖。因此, 本发明还涉及一种或多种宿主细胞, 其被特别设计用来制造具有所述多糖结构的糖蛋白。因此, 本发明还涉及一种优选为分离的, 具有所述结构的糖蛋白, 优选地, 所述糖蛋白可以由所述细胞制造或者确实由所述细胞制造。本发明还提供了一种通过使用所述细胞来制作所述糖蛋白的方法或工艺。

[0474] NeuAc2Gal2GlcNAc2Man3GlcNAc2 结构合成的实施方案

[0475] 在另一个优选实施方案中, 被修饰的宿主细胞提供用于基于高尔基体的加工的酶, 优选为异源酶, 其选自:

[0476] 甘露糖基 (α -1,3)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶 (GnTI), 尤其是 Mgat1 型转录物;

[0477] UDP-N-乙酰葡萄糖胺转运体, 尤其是 Slc35A3 型转录物;

[0478] 甘露糖基 (α -1,6)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶 (GnTII), 尤其是 Mgat2 型转录物;

[0479] β -N-乙酰葡萄糖氨基糖肽 β -1,4-半乳糖基转移酶 (GalT), 尤其是 B4gal1 型转录物;

[0480] UDP-半乳糖转运体, 尤其是 Slc35A2 型转录物;

[0481] β -半乳糖苷 α -2,6-唾液酸转移酶 (ST), 尤其是 ST6gal1 型转录物;

[0482] UDP-N-乙酰葡萄糖胺 2-差向异构酶 (NeuC), 尤其是 NeuC 型转录物;

[0483] 唾液酸合酶 (NeuB), 尤其是 NeuB 型转录物;

- [0484] CMP-Neu5Ac 合成酶,尤其是 Slc35A1 型转录物;及
- [0485] CMP- 唾液酸转运体,尤其是 NeuA/Cmas 型转录物。
- [0486] 在一个最优选的实施方案中,所述细胞至少包含或只包含这些高尔基体加工相关的酶。
- [0487] 在其一个可替代的变型中,被修饰的细胞提供 N- 酰基神经氨酸 -9- 磷酸合酶和 N- 酰基神经氨酸 -9- 磷酸酶来代替唾液酸合酶,更特别地,被修饰的宿主细胞提供用于基于高尔基体的加工的酶,优选为异源酶,其选自:
- [0488] 甘露糖基 (α -1,3)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTI),尤其是 Mgat1 型转录物;
- [0489] UDP-N- 乙酰葡萄糖胺转运体,尤其是 Slc35A3 型转录物;
- [0490] 甘露糖基 (α -1,6)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTII),尤其是 Mgat2 型转录物;
- [0491] β -N- 乙酰葡萄糖氨基糖肽 β -1,4- 半乳糖基转移酶 (GalT),尤其是 B4gal1 型转录物;
- [0492] UDP- 半乳糖转运体,尤其是 Slc35A2 型转录物;
- [0493] β - 半乳糖苷 α -2,6- 唾液酸转移酶 (ST),尤其是 ST6gal1 型转录物;
- [0494] UDP-N- 乙酰葡萄糖胺 2- 差向异构酶 (NeuC),尤其是 NeuC 型转录物;
- [0495] N- 酰基神经氨酸 -9- 磷酸合酶;
- [0496] N- 酰基神经氨酸 -9- 磷酸酶;
- [0497] CMP-Neu5Ac 合成酶,尤其是 Slc35A1 型转录物;及
- [0498] CMP- 唾液酸转运体,尤其是 NeuA/Cmas 型转录物。
- [0499] 在一个最优选的实施方案中,所述细胞至少包含或只包含这些高尔基体加工相关的酶。
- [0500] 在该实施方案的一个优选变型中,所述细胞表达下述基因中的一种或多种:
- [0501] mgat1,mgat2,slc35a3,b4gal1,slc35a2,st6gal1,neuC,neuB,slc35a1,和 neuC/cmas
- [0502] 和 / 或它们的同源基因。
- [0503] 具体地说,所述细胞能够制造具有 NeuAc2Gal2GlcNAc2Man3GlcNAc2 结构的 N- 多糖。因此,本发明还涉及一种或多种宿主细胞,其被特别设计用来制造具有所述多糖结构的糖蛋白。因此,本发明还涉及一种优选为分离的,具有所述结构的糖蛋白,优选地,所述糖蛋白可以由所述细胞制造或者确实由所述细胞制造。本发明还提供了一种通过使用所述细胞来制作所述糖蛋白的方法或工艺。
- [0504] NeuAc2Gal2GlcNAc3Man3GlcNAc2- 平分型结构合成的实施方案
- [0505] 在另一个优选实施方案中,被修饰的宿主细胞提供用于基于高尔基体的加工的酶,优选为异源酶,其选自:
- [0506] 甘露糖基 (α -1,3)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTI),尤其是 Mgat1 型转录物;
- [0507] UDP-N- 乙酰葡萄糖胺转运体,尤其是 Slc35A3 型转录物;
- [0508] 甘露糖基 (α -1,6)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTII),尤其是

Mgat2 型转录物；

[0509] β -1,4-甘露糖基-糖蛋白 4- β -N-乙酰葡萄糖氨基转移酶 (GnTIII), 尤其是 Mgat3 型转录物；

[0510] β -N-乙酰葡萄糖氨基糖肽 β -1,4-半乳糖基转移酶 (GalT), 尤其是 B4gal1 型转录物；

[0511] UDP-半乳糖转运体, 尤其是 Slc35A2 型转录物；

[0512] β -半乳糖苷 α -2,6-唾液酸转移酶 (ST), 尤其是 ST6gal1 型转录物；

[0513] UDP-N-乙酰葡萄糖胺 2-差向异构酶 (NeuC), 尤其是 NeuC 型转录物；

[0514] 唾液酸合酶 (NeuB), 尤其是 NeuB 型转录物；

[0515] CMP-Neu5Ac 合成酶, 尤其是 Slc35A1 型转录物；及

[0516] CMP-唾液酸转运体, 尤其是 NeuA/Cmas 型转录物。

[0517] 在一个最优选的实施方案中, 所述细胞至少包含或只包含这些高尔基体加工相关的酶。

[0518] 在其一个可替代的变型中, 被修饰的细胞提供 N-酰基神经氨酸-9-磷酸合酶和 N-酰基神经氨酸-9-磷酸酶来代替唾液酸合酶, 更特别地, 被修饰的宿主细胞提供用于基于高尔基体的加工的酶, 优选为异源酶, 其选自：

[0519] 甘露糖基 (α -1,3)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶 (GnTI), 尤其是 Mgat1 型转录物；

[0520] UDP-N-乙酰葡萄糖胺转运体, 尤其是 Slc35A3 型转录物；

[0521] 甘露糖基 (α -1,6)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶 (GnTII), 尤其是 Mgat2 型转录物；

[0522] β -1,4-甘露糖基-糖蛋白 4- β -N-乙酰葡萄糖氨基转移酶 (GnTIII), 尤其是 Mgat3 型转录物；

[0523] β -N-乙酰葡萄糖氨基糖肽 β -1,4-半乳糖基转移酶 (GalT), 尤其是 B4gal1 型转录物；

[0524] UDP-半乳糖转运体, 尤其是 Slc35A2 型转录物；

[0525] β -半乳糖苷 α -2,6-唾液酸转移酶 (ST), 尤其是 ST6gal1 型转录物；

[0526] UDP-N-乙酰葡萄糖胺 2-差向异构酶 (NeuC), 尤其是 NeuC 型转录物；

[0527] N-酰基神经氨酸-9-磷酸合酶；

[0528] N-酰基神经氨酸-9-磷酸酶；

[0529] CMP-Neu5Ac 合成酶, 尤其是 Slc35A1 型转录物；及

[0530] CMP-唾液酸转运体, 尤其是 NeuA/Cmas 型转录物。

[0531] 在一个最优选的实施方案中, 所述细胞至少包含或只包含这些高尔基体加工相关的酶。

[0532] 在该实施方案的一个优选变型中, 所述细胞表达下述基因中的一种或多种：

[0533] mgat1, mgat2, slc35a3, mgat3, b4gal1, slc35a2, st6gal1, neuC, neuB, slc35a1, 和 neuC/cmas

[0534] 和 / 或它们的同源基因。

[0535] 具体地说, 所述细胞能够制造具有 NeuAc2Gal2GlcNAc2Man3GlcNAc2- 平分型结构

的 N- 多糖。因此,本发明还涉及一种或多种宿主细胞,其被特别设计用来制造具有所述多糖结构的糖蛋白。因此,本发明还涉及一种优选为分离的,具有所述结构的糖蛋白,优选地,所述糖蛋白可以由所述细胞制造或者确实由所述细胞制造。本发明还提供了一种通过使用所述细胞来制作所述糖蛋白的方法或工艺。

[0536] NeuAc2Gal12GlcNAc2Man3GlcNAc2Fuc 结构合成的实施方案

[0537] 在另一个优选实施方案中,被修饰的宿主细胞提供用于基于高尔基体的加工的酶,优选为异源酶,其选自:

[0538] 甘露糖基 (α -1,3)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTI), 尤其是 Mgat1 型转录物;

[0539] UDP-N- 乙酰葡萄糖胺转运体, 尤其是 Slc35A3 型转录物;

[0540] 甘露糖基 (α -1,6)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTII), 尤其是 Mgat2 型转录物;

[0541] β -N- 乙酰葡萄糖氨基糖肽 β -1,4- 半乳糖基转移酶 (GalT), 尤其是 B4gal1 型转录物;

[0542] UDP- 半乳糖转运体, 尤其是 Slc35A2 型转录物;

[0543] GDP-D- 甘露糖 4,6- 脱水酶, 尤其是 Gmds 型转录物;

[0544] GDP-4- 酮 - 去氧 -D- 甘露糖 -3,5- 差向异构酶 -4- 还原酶, 尤其是 Tsta3 型转录物;

[0545] GDP- 岩藻糖转运体, 尤其是 Slc35C1 型转录物;

[0546] α (1,6) 岩藻糖基转移酶 (FucT), 尤其是 Fut8 型转录物;

[0547] β - 半乳糖苷 α -2,6- 唾液酸转移酶 (ST), 尤其是 ST6gal1 型转录物;

[0548] UDP-N- 乙酰葡萄糖胺 2- 差向异构酶 (NeuC), 尤其是 NeuC 型转录物;

[0549] 唾液酸合酶 (NeuB), 尤其是 NeuB 型转录物;

[0550] CMP-Neu5Ac 合成酶, 尤其是 Slc35A1 型转录物; 及

[0551] CMP- 唾液酸转运体, 尤其是 NeuA/Cmas 型转录物。

[0552] 在一个最优选的实施方案中,所述细胞至少包含或只包含这些高尔基体加工相关的酶。

[0553] 在其一个可替代的变型中,被修饰的细胞提供 N- 酰基神经氨酸 -9- 磷酸合酶和 N- 酰基神经氨酸 -9- 磷酸酶来代替唾液酸合酶,更特别地,被修饰的宿主细胞提供用于基于高尔基体的加工的酶,优选为异源酶,其选自:

[0554] 甘露糖基 (α -1,3)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTI), 尤其是 Mgat1 型转录物;

[0555] UDP-N- 乙酰葡萄糖胺转运体, 尤其是 Slc35A3 型转录物;

[0556] 甘露糖基 (α -1,6)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTII), 尤其是 Mgat2 型转录物;

[0557] β -N- 乙酰葡萄糖氨基糖肽 β -1,4- 半乳糖基转移酶 (GalT), 尤其是 B4gal1 型转录物;

[0558] UDP- 半乳糖转运体, 尤其是 Slc35A2 型转录物;

[0559] GDP-D- 甘露糖 4,6- 脱水酶, 尤其是 Gmds 型转录物;

- [0560] GDP-4-酮-去氧-D-甘露糖-3,5-差向异构酶-4-还原酶,尤其是 Tsta3 型转录物;
- [0561] GDP-岩藻糖转运体,尤其是 Slc35C1 型转录物;
- [0562] α (1,6) 岩藻糖基转移酶 (FucT),尤其是 Fut8 型转录物;
- [0563] β -半乳糖苷 α -2,6-唾液酸转移酶 (ST),尤其是 ST6gal1 型转录物;
- [0564] UDP-N-乙酰葡萄糖胺 2-差向异构酶 (NeuC),尤其是 NeuC 型转录物;
- [0565] N-酰基神经氨酸-9-磷酸合酶;
- [0566] N-酰基神经氨酸-9-磷酸酶;
- [0567] CMP-Neu5Ac 合成酶,尤其是 Slc35A1 型转录物;及
- [0568] CMP-唾液酸转运体,尤其是 NeuA/Cmas 型转录物。
- [0569] 在一个最优选的实施方案中,所述细胞至少包含或只包含这些高尔基体加工相关的酶。
- [0570] 在该实施方案的一个优选变型中,所述细胞表达下述基因中的一种或多种:
- [0571] mgat1, mgat2, slc35a3, b4gal1, slc35a2, gmds, tsta3, slc35c1, fut8, st6gal1, neuC, neuB, slc35a1, 和 neuC/cmas
- [0572] 和/或它们的同源基因。
- [0573] 具体地说,所述细胞能够制造具有 NeuAc2Gal2GlcNAc2Man3GlcNAc-2Fuc 结构的 N-多糖。因此,本发明还涉及一种或多种宿主细胞,其被特别设计用来制造具有所述多糖结构的糖蛋白。因此,本发明还涉及一种优选为分离的,具有所述结构的糖蛋白,优选地,所述糖蛋白可以由所述细胞制造或者确实由所述细胞制造。本发明还提供了一种通过使用所述细胞来制作所述糖蛋白的方法或工艺。
- [0574] NeuAc2Gal2GlcNAc3Man3GlcNAc2Fuc-平分型结构合成的实施方案
- [0575] 在另一个优选实施方案中,被修饰的宿主细胞提供用于基于高尔基体的加工的酶,优选为异源酶,其选自:
- [0576] 甘露糖基 (α -1,3)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶 (GnTI),尤其是 Mgat1 型转录物;
- [0577] UDP-N-乙酰葡萄糖胺转运体,尤其是 Slc35A3 型转录物;
- [0578] 甘露糖基 (α -1,6)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶 (GnTII),尤其是 Mgat2 型转录物;
- [0579] β -1,4-甘露糖基-糖蛋白 4- β -N-乙酰葡萄糖氨基转移酶 (GnTIII),尤其是 Mgat3 型转录物;
- [0580] β -N-乙酰葡萄糖氨基糖肽 β -1,4-半乳糖基转移酶 (GalT),尤其是 B4gal1 型转录物;
- [0581] UDP-半乳糖转运体,尤其是 Slc35A2 型转录物;
- [0582] GDP-D-甘露糖 4,6-脱水酶,尤其是 Gmds 型转录物;
- [0583] GDP-4-酮-去氧-D-甘露糖-3,5-差向异构酶-4-还原酶,尤其是 Tsta3 型转录物;
- [0584] GDP-岩藻糖转运体,尤其是 Slc35C1 型转录物;
- [0585] α (1,6) 岩藻糖基转移酶 (FucT),尤其是 Fut8 型转录物;

- [0586] β -半乳糖苷 α -2,6-唾液酸转移酶 (ST), 尤其是 ST6gal1 型转录物;
- [0587] UDP-N-乙酰葡萄糖胺 2-差向异构酶 (NeuC), 尤其是 NeuC 型转录物;
- [0588] 唾液酸合酶 (NeuB), 尤其是 NeuB 型转录物;
- [0589] CMP-Neu5Ac 合成酶, 尤其是 Slc35A1 型转录物; 及
- [0590] CMP-唾液酸转运体, 尤其是 NeuA/Cmas 型转录物。
- [0591] 在一个最优选的实施方案中, 所述细胞至少包含或只包含这些高尔基体加工相关的酶。
- [0592] 在其一个可替代的变型中, 被修饰的细胞提供 N-酰基神经氨酸-9-磷酸合酶和 N-酰基神经氨酸-9-磷酸酶来代替唾液酸合酶, 更特别地, 被修饰的宿主细胞提供用于基于高尔基体的加工的酶, 优选为异源酶, 其选自:
- [0593] 甘露糖基 (α -1,3)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶 (GnTI), 尤其是 Mgat1 型转录物;
- [0594] UDP-N-乙酰葡萄糖胺转运体, 尤其是 Slc35A3 型转录物;
- [0595] 甘露糖基 (α -1,6)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶 (GnTII), 尤其是 Mgat2 型转录物;
- [0596] β -1,4-甘露糖基-糖蛋白 4- β -N-乙酰葡萄糖氨基转移酶 (GnTIII), 尤其是 Mgat3 型转录物;
- [0597] β -N-乙酰葡萄糖氨基糖肽 β -1,4-半乳糖基转移酶 (GalT), 尤其是 B4gal1 型转录物;
- [0598] UDP-半乳糖转运体, 尤其是 Slc35A2 型转录物;
- [0599] GDP-D-甘露糖 4,6-脱水酶, 尤其是 Gmds 型转录物;
- [0600] GDP-4-酮-去氧-D-甘露糖-3,5-差向异构酶-4-还原酶, 尤其是 Tsta3 型转录物;
- [0601] GDP-岩藻糖转运体, 尤其是 Slc35C1 型转录物;
- [0602] α (1,6) 岩藻糖基转移酶 (FucT), 尤其是 Fut8 型转录物;
- [0603] β -半乳糖苷 α -2,6-唾液酸转移酶 (ST), 尤其是 ST6gal1 型转录物;
- [0604] UDP-N-乙酰葡萄糖胺 2-差向异构酶 (NeuC), 尤其是 NeuC 型转录物;
- [0605] N-酰基神经氨酸-9-磷酸合酶;
- [0606] N-酰基神经氨酸-9-磷酸酶;
- [0607] CMP-Neu5Ac 合成酶, 尤其是 Slc35A1 型转录物; 及
- [0608] CMP-唾液酸转运体, 尤其是 NeuA/Cmas 型转录物。
- [0609] 在一个最优选的实施方案中, 所述细胞至少包含或只包含这些高尔基体加工相关的酶。
- [0610] 在该实施方案的一个优选变型中, 所述细胞表达下述基因中的一种或多种:
- [0611] mgat1, mgat2, slc35a3, b4gal1, mgat3, slc35a2, gmds, tsta3, slc35c1, fut8, st6gal1, neuC, neuB, slc35a1, 和 neuC/cmas
- [0612] 和 / 或它们的同源基因。
- [0613] 具体地说, 所述细胞能够制造具有 NeuAc2Gal2GlcNAc2Man3GlcNAc-2Fuc- 平分型结构的 N-多糖。因此, 本发明还涉及一种或多种宿主细胞, 其被特别设计用来制造具有所

述多糖结构的糖蛋白。因此,本发明还涉及一种优选为分离的,具有所述结构的糖蛋白,优选地,所述糖蛋白可以由所述细胞制造或者确实由所述细胞制造。本发明还提供了一种通过使用所述细胞来制作所述糖蛋白的方法或工艺。

[0614] GlcNAc3Man3GlcNAc2 结构合成的实施方案

[0615] 在另一个优选实施方案中,被修饰的宿主细胞提供用于基于高尔基体的加工的酶,优选为异源酶,其选自:

[0616] 甘露糖基 (α -1,3)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶 (GnTI),尤其是 Mgat1 型转录物;

[0617] UDP-N-乙酰葡萄糖胺转运体,尤其是 Slc35A3 型转录物;

[0618] 甘露糖基 (α -1,6)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶 (GnTII),尤其是 Mgat2 型转录物;及

[0619] 甘露糖基 (α -1,3)-糖蛋白 β -1,4-N-乙酰葡萄糖氨基转移酶 (GnTIV),尤其是 Mgat4 型转录物;

[0620] 在一个最优选的实施方案中,所述细胞至少包含或只包含这些高尔基体加工相关的酶。

[0621] 在该实施方案的一个优选变型中,所述细胞表达下述基因中的一种或多种:

[0622] mgat1, mgat2, mgat4, 和 slc35A3

[0623] 和/或它们的同源基因。

[0624] 具体地说,所述细胞能够制造具有 GlcNAc3Man3GlcNAc2 结构的 N-多糖。因此,本发明还涉及一种或多种宿主细胞,其被特别设计用来制造具有所述多糖结构的糖蛋白。因此,本发明还涉及一种优选为分离的,具有所述结构的糖蛋白,优选地,所述糖蛋白可以由所述细胞制造或者确实由所述细胞制造。本发明还提供了一种通过使用所述细胞来制作所述糖蛋白的方法或工艺。

[0625] Gal3GlcNAc3Man3GlcNAc2 结构合成的实施方案

[0626] 在另一个优选实施方案中,被修饰的宿主细胞提供用于基于高尔基体的加工的酶,优选为异源酶,其选自:

[0627] 甘露糖基 (α -1,3)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶 (GnTI),尤其是 Mgat1 型转录物;

[0628] UDP-N-乙酰葡萄糖胺转运体,尤其是 Slc35A3 型转录物;

[0629] 甘露糖基 (α -1,6)-糖蛋白 β -1,2-N-乙酰葡萄糖氨基转移酶 (GnTII),尤其是 Mgat2 型转录物;

[0630] 甘露糖基 (α -1,3)-糖蛋白 β -1,4-N-乙酰葡萄糖氨基转移酶 (GnTIV),尤其是 Mgat4 型转录物;

[0631] β -N-乙酰葡萄糖氨基糖肽 β -1,4-半乳糖基转移酶 (GalT),尤其是 B4gal1t1 型转录物;及

[0632] UDP-半乳糖转运体,尤其是 Slc35A2 型转录物。

[0633] 在一个最优选的实施方案中,所述细胞至少包含或只包含这些高尔基体加工相关的酶。

[0634] 在该实施方案的一个优选变型中,所述细胞表达下述基因中的一种或多种:

[0635] mgat1, mgat2, maga4, slc35a3, b4galt1 和 slc35a2

[0636] 和 / 或它们的同源基因。

[0637] 具体地说, 所述细胞能够制造具有 Gal3GlcNAc3Man3GlcNAc2 结构的 N- 多糖。因此, 本发明还涉及一种或多种宿主细胞, 其被特别设计用来制造具有所述多糖结构的糖蛋白。因此, 本发明还涉及一种优选为分离的, 具有所述结构的糖蛋白, 优选地, 所述糖蛋白可以由所述细胞制造或者确实由所述细胞制造。本发明还提供了一种通过使用所述细胞来制作所述糖蛋白的方法或工艺。

[0638] Gal3GlcNAc3Man3GlcNAc2Fuc 结构合成的实施方案

[0639] 在另一个优选实施方案中, 被修饰的宿主细胞提供用于基于高尔基体的加工的酶, 优选为异源酶, 其选自:

[0640] 甘露糖基 (α -1,3)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTI), 尤其是 Mgat1 型转录物;

[0641] UDP-N- 乙酰葡萄糖胺转运体, 尤其是 Slc35A3 型转录物;

[0642] 甘露糖基 (α -1,6)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTII), 尤其是 Mgat2 型转录物;

[0643] 甘露糖基 (α -1,3)- 糖蛋白 β -1,4-N- 乙酰葡萄糖氨基转移酶 (GnTIV), 尤其是 Mgat4 型转录物;

[0644] β -N- 乙酰葡萄糖氨基糖肽 β -1,4- 半乳糖基转移酶 (GalT), 尤其是 B4galt1 型转录物; 及

[0645] UDP- 半乳糖转运体, 尤其是 Slc35A2 型转录物;

[0646] GDP-D- 甘露糖 4,6- 脱水酶, 尤其是 Gmds 型转录物;

[0647] GDP-4- 酮 - 去氧 -D- 甘露糖 -3,5- 差向异构酶 -4- 还原酶, 尤其是 Tsta3 型转录物;

[0648] GDP- 岩藻糖转运体, 尤其是 Slc35C1 型转录物; 及

[0649] α (1,6) 岩藻糖基转移酶 (FucT), 尤其是 Fut8 型转录物。

[0650] 在一个最优选的实施方案中, 所述细胞至少包含或只包含这些高尔基体加工相关的酶。

[0651] 在该实施方案的一个优选变型中, 所述细胞表达下述基因中的一种或多种:

[0652] mgat1, mgat2, maga4, slc35a3, b4galt1, slc35a2, gmds, tsta3, slc35c1 和 fut8

[0653] 和 / 或它们的同源基因。

[0654] 具体地说, 所述细胞能够制造具有 Gal3GlcNAc3Man3GlcNAc2Fuc 结构的 N- 多糖。因此, 本发明还涉及一种或多种宿主细胞, 其被特别设计用来制造具有所述多糖结构的糖蛋白。因此, 本发明还涉及一种优选为分离的, 具有所述结构的糖蛋白, 优选地, 所述糖蛋白可以由所述细胞制造或者确实由所述细胞制造。本发明还提供了一种通过使用所述细胞来制作所述糖蛋白的方法或工艺。

[0655] NeuAc3Gal3GlcNAc3Man3GlcNAc2 结构合成的实施方案

[0656] 在另一个优选实施方案中, 被修饰的宿主细胞提供用于基于高尔基体的加工的酶, 优选为异源酶, 其选自:

[0657] 甘露糖基 (α -1,3)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTI), 尤其是

Mgat1 型转录物；

[0658] UDP-N- 乙酰葡萄糖胺转运体,尤其是 Slc35A3 型转录物；

[0659] 甘露糖基 (α -1,6)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTII),尤其是 Mgat2 型转录物；

[0660] 甘露糖基 (α -1,3)- 糖蛋白 β -1,4-N- 乙酰葡萄糖氨基转移酶 (GnTIV),尤其是 Mgat4 型转录物；

[0661] β -N- 乙酰葡萄糖氨基糖肽 β -1,4- 半乳糖基转移酶 (GalT),尤其是 B4gal1 型转录物；

[0662] UDP- 半乳糖转运体,尤其是 Slc35A2 型转录物；

[0663] β - 半乳糖苷 α -2,6- 唾液酸转移酶 (ST),尤其是 ST6gal1 型转录物；

[0664] UDP-N- 乙酰葡萄糖胺 2- 差向异构酶 (NeuC),尤其是 NeuC 型转录物；

[0665] 唾液酸合酶 (NeuB),尤其是 NeuB 型转录物；

[0666] CMP-Neu5Ac 合成酶,尤其是 Slc35A1 型转录物；及

[0667] CMP- 唾液酸转运体,尤其是 NeuA/Cmas 型转录物。

[0668] 在一个最优选的实施方案中,所述细胞至少包含或只包含这些高尔基体加工相关的酶。

[0669] 在其一个可替代的变型中,被修饰的细胞提供 N- 酰基神经氨酸-9- 磷酸合酶和 N- 酰基神经氨酸-9- 磷酸酶来代替唾液酸合酶,更特别地,被修饰的宿主细胞提供用于基于高尔基体的加工的酶,优选为异源酶,其选自：

[0670] 甘露糖基 (α -1,3)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTI),尤其是 Mgat1 型转录物；

[0671] UDP-N- 乙酰葡萄糖胺转运体,尤其是 Slc35A3 型转录物；

[0672] 甘露糖基 (α -1,6)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTII),尤其是 Mgat2 型转录物；

[0673] 甘露糖基 (α -1,3)- 糖蛋白 β -1,4-N- 乙酰葡萄糖氨基转移酶 (GnTIV),尤其是 Mgat4 型转录物；

[0674] β -N- 乙酰葡萄糖氨基糖肽 β -1,4- 半乳糖基转移酶 (GalT),尤其是 B4gal1 型转录物；

[0675] UDP- 半乳糖转运体,尤其是 Slc35A2 型转录物；

[0676] β - 半乳糖苷 α -2,6- 唾液酸转移酶 (ST),尤其是 ST6gal1 型转录物；

[0677] UDP-N- 乙酰葡萄糖胺 2- 差向异构酶 (NeuC),尤其是 NeuC 型转录物；

[0678] N- 酰基神经氨酸-9- 磷酸合酶；

[0679] N- 酰基神经氨酸-9- 磷酸酶；

[0680] CMP-Neu5Ac 合成酶,尤其是 Slc35A1 型转录物；及

[0681] CMP- 唾液酸转运体,尤其是 NeuA/Cmas 型转录物。

[0682] 在一个最优选的实施方案中,所述细胞至少包含或只包含这些高尔基体加工相关的酶。

[0683] 在该实施方案的一个优选变型中,所述细胞表达下述基因中的一种或多种：

[0684] mgat1, mgat2, slc35a3, b4gal1, mgat4, slc35a2, st6gal1, neuC, neuB, slc35a1,

和 neuC/cmas

[0685] 和 / 或它们的同源基因。

[0686] 具体地说,所述细胞能够制造具有 NeuAc3Gal3GlcNAc3Man3GlcNAc2 结构的 N- 多糖。因此,本发明还涉及一种或多种宿主细胞,其被特别设计用来制造具有所述多糖结构的糖蛋白。因此,本发明还涉及一种优选为分离的,具有所述结构的糖蛋白,优选地,所述糖蛋白可以由所述细胞制造或者确实由所述细胞制造。本发明还提供了一种通过使用所述细胞来制作所述糖蛋白的方法或工艺。

[0687] NeuAc3Gal3GlcNAc3Man3GlcNAc2Fuc 结构合成的实施方案

[0688] 在另一个优选实施方案中,被修饰的宿主细胞提供用于基于高尔基体的加工的酶,优选为异源酶,其选自:

[0689] 甘露糖基 (α -1,3)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTI), 尤其是 Mgat1 型转录物;

[0690] UDP-N- 乙酰葡萄糖胺转运体, 尤其是 Slc35A3 型转录物;

[0691] 甘露糖基 (α -1,6)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTII), 尤其是 Mgat2 型转录物;

[0692] 甘露糖基 (α -1,3)- 糖蛋白 β -1,4-N- 乙酰葡萄糖氨基转移酶 (GnTIV), 尤其是 Mgat4 型转录物;

[0693] β -N- 乙酰葡萄糖氨基糖肽 β -1,4- 半乳糖基转移酶 (GalT), 尤其是 B4gal1 型转录物;

[0694] UDP- 半乳糖转运体, 尤其是 Slc35A2 型转录物;

[0695] GDP-D- 甘露糖 4,6- 脱水酶, 尤其是 Gmds 型转录物;

[0696] GDP-4- 酮 - 去氧 -D- 甘露糖 -3,5- 差向异构酶 -4- 还原酶, 尤其是 Tsta3 型转录物;

[0697] GDP- 岩藻糖转运体, 尤其是 Slc35C1 型转录物;

[0698] α (1,6) 岩藻糖基转移酶 (FucT), 尤其是 Fut8 型转录物;

[0699] β - 半乳糖苷 α -2,6- 唾液酸转移酶 (ST), 尤其是 ST6gal1 型转录物;

[0700] UDP-N- 乙酰葡萄糖胺 2- 差向异构酶 (NeuC), 尤其是 NeuC 型转录物;

[0701] 唾液酸合酶 (NeuB), 尤其是 NeuB 型转录物;

[0702] CMP-Neu5Ac 合成酶, 尤其是 Slc35A1 型转录物; 及

[0703] CMP- 唾液酸转运体, 尤其是 NeuA/Cmas 型转录物。

[0704] 在一个最优选的实施方案中,所述细胞至少包含或只包含这些高尔基体加工相关的酶。

[0705] 在其一个可替代的变型中,被修饰的细胞提供 N- 酰基神经氨酸 -9- 磷酸合酶和 N- 酰基神经氨酸 -9- 磷酸酶来代替唾液酸合酶,更特别地,被修饰的宿主细胞提供用于基于高尔基体的加工的酶,优选为异源酶,其选自:

[0706] 甘露糖基 (α -1,3)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTI), 尤其是 Mgat1 型转录物;

[0707] UDP-N- 乙酰葡萄糖胺转运体, 尤其是 Slc35A3 型转录物;

[0708] 甘露糖基 (α -1,6)- 糖蛋白 β -1,2-N- 乙酰葡萄糖氨基转移酶 (GnTII), 尤其是

Mgat2 型转录物；

[0709] 甘露糖基 (α -1,3)-糖蛋白 β -1,4-N-乙酰葡萄糖氨基转移酶 (GnTIV), 尤其是 Mgat4 型转录物；

[0710] β -N-乙酰葡萄糖氨基糖肽 β -1,4-半乳糖基转移酶 (GalT), 尤其是 B4gal1 型转录物；

[0711] UDP-半乳糖转运体, 尤其是 Slc35A2 型转录物；

[0712] GDP-D-甘露糖 4,6-脱水酶, 尤其是 Gmds 型转录物；

[0713] GDP-4-酮-去氧-D-甘露糖-3,5-差向异构酶-4-还原酶, 尤其是 Tsta3 型转录物；

[0714] GDP-岩藻糖转运体, 尤其是 Slc35C1 型转录物；

[0715] α (1,6) 岩藻糖基转移酶 (FucT), 尤其是 Fut8 型转录物；

[0716] β -半乳糖苷 α -2,6-唾液酸转移酶 (ST), 尤其是 ST6gal1 型转录物；

[0717] UDP-N-乙酰葡萄糖胺 2-差向异构酶 (NeuC), 尤其是 NeuC 型转录物；

[0718] N-酰基神经氨酸-9-磷酸合酶；

[0719] N-酰基神经氨酸-9-磷酸酶；

[0720] CMP-Neu5Ac 合成酶, 尤其是 Slc35A1 型转录物；及

[0721] CMP-唾液酸转运体, 尤其是 NeuA/Cmas 型转录物。

[0722] 在一个最优选的实施方案中, 所述细胞至少包含或只包含这些高尔基体加工相关的酶。

[0723] 在该实施方案的一个优选变型中, 所述细胞表达下述基因中的一种或多种：

[0724] mgat1, mgat2, slc35a3, b4gal1, mgat4, slc35a2, gmds, tsta3, slc35c1, fut8, st6gal1, neuC, neuB, slc35a1, 和 neuC/cmas

[0725] 和 / 或它们的同源基因。

[0726] 具体地说, 所述细胞能够制造具有 NeuAc3Gal2GlcNAc2Man3GlcNAc2Fuc 结构的 N-多糖。因此, 本发明还涉及一种或多种宿主细胞, 其被特别设计用来制造具有所述多糖结构的糖蛋白。因此, 本发明还涉及一种优选为分离的, 具有所述结构的糖蛋白, 优选地, 所述糖蛋白可以由所述细胞制造或者确实由所述细胞制造。本发明还提供了一种通过使用所述细胞来制作所述糖蛋白的方法或工艺。

[0727] 用于制作糖蛋白的方法或工艺

[0728] 本发明还提供了一种通过使用任意一种本发明所述的宿主细胞来制作糖蛋白的方法或工艺。不希望受理论的束缚, 本发明所述的细胞能够制造大量在所述糖蛋白上带有 Man1GlcNAc2, Man2GlcNAc2 或 Man3GlcNAc2 结构的 N-多糖。所述糖蛋白可以为同源或异源蛋白。因此, 任意一种如上所述的宿主细胞都优选包含至少一种编码异源糖蛋白的核酸。同源蛋白主要是指来源于宿主细胞本身的蛋白质, 而被“外来的”克隆基因编码的蛋白质为所述宿主细胞的异源蛋白。更特别地, 任何一种编码本发明所述的异源蛋白的核酸都能被密码子优化以在感兴趣的宿主细胞中表达。例如, 一种编码 *Trypanosoma brucei* 的原生动物寡糖基转移酶的核酸能被密码子优化以在酵母细胞例如 *Saccharomyces cerevisiae* 中表达。

[0729] 根据本发明所述的宿主细胞能够制造复合 N-连接寡糖及杂交寡糖。在例如人

促红细胞生成素 (hEPO) 等治疗用蛋白的生理学中已经暗含了分支的复合 N- 多糖。已经表明了具有双触角型结构的人促红细胞生成素具有低活性,然而具有三触角型结构的人促红细胞生成素从血流中清除缓慢,因此导致更高的活性 (Misaizu T et al. (1995) Blood December 1 ;86(11) :4097-104)。

[0730] 多糖结构意味着一种结合于蛋白核的寡糖。高甘露糖结构包括多于 5 个甘露糖,而主要由只延伸至少于 5 个甘露糖部分的甘露糖组成的多糖结构为低甘露糖多糖结构,例如 Man3GlcNAc2。更特别地,本文所用的术语“多糖”或“糖蛋白”是指一种 N- 连接寡糖,例如,一种通过天冬酰胺 -N- 乙酰葡萄糖胺键附加到多肽的天冬酰胺残基上的 N- 连接寡糖。N- 多糖具有一个普通的 Man3GlcNAc2 (“Man”是指甘露糖;“Glc”是指葡萄糖;“NAc”是指 N- 乙酰基;GlcNAc 是指 N- 乙酰葡萄糖胺) 的戊糖核。N- 多糖针对包含被添加到所述 Man3GlcNAc2 (“Man3”) 核心结构的外周糖 (例如,岩藻糖及唾液酸) 在内的分支 (触角) 数目而发生变化。N- 多糖根据它们的分支组成 (例如,高甘露糖、复合或杂交) 来进行分类。一种糖形代表携带一种特异 N- 多糖的糖基化蛋白。因此,糖形代表携带不同 N- 多糖的糖基化蛋白。一种“高甘露糖”型 N- 多糖具有 5 个或更多个甘露糖。

[0731] 所述核心结构 Man3GlcNAc2 是 N- 多糖的所有分类中共有的。所述核心结构通过延伸序列接在每个分支上,并由细胞型特异的己糖终止。N- 多糖结构的三种普通类型可以被定义为:(1) 高甘露糖多糖,主要包括在延伸序列内并且还作为终止部分的甘露糖。(2) 与此相反,由不同的己糖组成的复合多糖。在人类中,它们通常包括 N- 乙酰神经氨酸作为终止己糖。及 (3) 杂交多糖,在单一多糖内包括聚甘露糖基和复合延伸序列两种。

[0732] 典型地,一种“复合”型 N- 多糖具有至少一个附加在“三甘露糖”核的 1,3 甘露糖臂的 GlcNAc 及至少一个附加在“三甘露糖”核的 1,6 甘露糖臂的 GlcNAc。所述“三甘露糖核”为具有一个 Man3 结构的戊糖核。复合 N- 多糖还可以具有半乳糖 (“Gal”) 残基,可以用唾液酸或衍生物 (“NeuAc”,这里“Neu”是指神经氨酸,“Ac”是指乙酰基) 任意修饰。复合 N- 多糖还可以具有包含“平分型”GlcNAc 和核心岩藻糖 (“Fuc”) 的链内替换。一种“杂交”型 N- 多糖具有至少一个在三甘露糖核的 1,3 甘露糖臂末端的 GlcNAc 及零或多个在三甘露糖核的 1,6 甘露糖臂上的甘露糖。

[0733] 本发明的另一方面是涉及一种工艺,用于制作具有低甘露糖多糖结构的糖蛋白或包含一种或多种具有低甘露糖多糖结构的糖蛋白的糖蛋白组合物。

[0734] 在一个优选实施方案中,所述蛋白为异源蛋白。在其一个优选变型中,所述异源蛋白为重组蛋白。本发明的一个优选实施方案为一种组合物,包含由或可以由本发明所述的细胞制造的异源和 / 或重组蛋白,其中,所述组合物包含大量具有 Man1-3GlcNAc2 中的一种多糖结构的糖蛋白。

[0735] “重组蛋白”、“异源蛋白”和“外源蛋白”可替换使用来指代一种通过重组 DNA 技术制造的多肽,其中,通常情况下,编码多肽的 DNA 被插入到合适的表达载体内,所述表达载体反过来被用于转化宿主细胞来制造所述异源蛋白。就是说,多肽是从异源核酸中被表达的。

[0736] 在一个优选变型中,提供了一种工艺,用于制作具有 Man3GlcNAc2 多糖结构的糖蛋白或包含至少一种具有 Man3GlcNAc2 多糖结构的糖蛋白的糖蛋白组合物。在另一个优选变型中,提供了一种工艺,用于制作具有 Man2GlcNAc2 多糖结构的糖蛋白或包含至少一种

具有 Man2GlcNAc2 多糖结构的糖蛋白的糖蛋白组合物。在另一个优选变型中,还提供了一种工艺,用于制作具有 Man21GlcNAc2 多糖结构的糖蛋白或包含至少一种具有 Man1GlcNAc2 多糖结构的糖蛋白的糖蛋白组合物。在另一个优选变型中,还提供了一种工艺,用于制作具有 Man4GlcNAc2 多糖结构的类人糖蛋白或包含至少一种具有 Man4GlcNAc2 多糖结构的糖蛋白的糖蛋白组合物。在另一个优选变型中,还提供了一种工艺,用于制作具有 Man5GlcNAc2 多糖结构的类人糖蛋白或包含至少一种具有 Man5GlcNAc2 多糖结构的糖蛋白的糖蛋白组合物。

[0737] 所述工艺至少包含下述步骤:提供本发明所述的突变细胞。优选在液体培养基中培养所述细胞,并优选在所述细胞内允许或最优选为支持所述糖蛋白或糖蛋白组合物的制造条件下,培养所述细胞。必要时,可以从所述细胞和/或所述培养基中分离所述糖蛋白或糖蛋白组合物。优选地,使用本领域公知的方法及手段进行分离。

[0738] 本发明还提供新的糖蛋白及其组合物,可以由或是由本发明所述的细胞或方法来制造。这种组合物的特征还在于,包含选自 Man1GlcNAc2, Man2GlcNAc2, 和 Man3GlcNAc2 的多糖核心结构,优选为 Man3GlcNAc2 结构。本发明还可以提供特征在于包含选自 Man4GlcNAc2 和 Man5GlcNAc2 的多糖结构的组合物,所述多糖结构可以通过高尔基体中 Man1GlcNAc2, Man2GlcNAc2, 或 Man3GlcNAc2 核的进一步甘露糖基化来制造。

[0739] 在优选实施方案中,一种或多种所述多糖结构在组合物中以至少 40%或更多的量存在,更优选为至少 50%或更多,更优选为至少 60%或更多,更优选为至少 70%或更多,更优选为至少 80%或更多,更优选为至少 90%或更多,更优选为至少 95%或更多,最优选至 99%或 100%。不言而喻,对这种蛋白组合物普遍的其他物质及副产物被排除在所述计算结果之外。在一个最优选的实施方案中,基本上所有由所述细胞制造的多糖结构都带有 Man3GlcNAc2 结构。在另一个优选实施方案中,基本上所有由所述细胞制造的糖形都带有 Man4GlcNAc2 和/或 Man5GlcNAc2 结构。

[0740] 作为高尔基体修饰的结果,正如上文中所详细描述,携带复合及杂交 N-多糖的糖蛋白是能获得的。所述糖蛋白包含多糖结构,其选自但不限于:

- [0741] GlcNAcMan3-5GlcNAc2,
- [0742] GlcNAc2Man3GlcNAc2,
- [0743] GlcNAc3Man3GlcNAc2- 平分型
- [0744] Gal2GlcNAc2Man3GlcNAc2,
- [0745] Gal2GlcNAc2Man3GlcNAc2Fuc,
- [0746] Gal2GlcNAc3Man3GlcNAc2- 平分型,
- [0747] Gal2GlcNAc3Man3GlcNAc2Fuc- 平分型,
- [0748] NeuAc2Gal2GlcNAc2Man3GlcNAc2,
- [0749] NeuAc2Gal2GlcNAc2Man3GlcNAc2Fuc,
- [0750] NeuAc2Gal2GlcNAc3Man3GlcNAc2- 平分型,
- [0751] NeuAc2Gal2GlcNAc3Man3GlcNAc2Fuc- 平分型,
- [0752] GlcNAc3Man3GlcNAc2,
- [0753] Gal3GlcNAc3Man3GlcNAc2,
- [0754] Gal3GlcNAc3Man3GlcNAc2Fuc,

[0755] NeuAc3Gal3GlcNAc3Man3GlcNAc2, 及

[0756] NeuAc3Gal3GlcNAc3Man3GlcNAc2Fuc

[0757] 在优选实施方案中, 一种或多种上述多糖结构在糖蛋白或糖蛋白组合物中以至少 40% 或更多的量存在, 优选为至少 50% 或更多, 更优选为至少 60% 或更多, 更优选为至少 70% 或更多, 更优选为至少 80% 或更多, 更优选为至少 90% 或更多, 更优选为至少 95% 或更多, 最优选为 99% 至全部糖蛋白。不言而喻, 对这种蛋白组合物普遍的其他物质及副产物被排除在所述计算结果之外。在一个最优选的实施方案中, 基本上所有由本发明所述宿主细胞制造的糖蛋白都带有一种或多种上述多糖结构。

[0758] 在一些实施方案中, 本发明所述糖蛋白的 N- 糖基化形式可以为同源或基本上同源的。具体地讲, 一种特殊多糖结构的片段在糖蛋白中占至少约 20% 或更多, 约 30% 或更多, 约 40% 或更多, 优选至少约 50% 或更多, 更优选至少约 60% 或更多, 更优选约 70% 或更多, 更优选约 80% 或更多, 更优选约 90% 或更多, 更优选约 95% 或更多, 最优选为 99% 至全部糖蛋白。

[0759] 本发明的优选实施方案为新的由或可以由宿主细胞制造的糖蛋白组合物, 带有上述多糖结构的两种或多种不同的糖蛋白。不希望受理论的束缚, 在一个优选实施方案中, 本发明所述的特殊宿主细胞能够同时制造两种或多种不同的糖蛋白, 结果得到不同结构的糖蛋白的混合物。这还涉及糖基化的中间形式。必须注意的是, 在本发明最优选的变型中, 所述宿主细胞提供了一种特殊多糖结构, 主要地或甚至是纯粹地 (大于 90%; 优选大于 95%, 最优选 99% 或更多) 达到必要的延伸。

[0760] 在另一个优选实施方案中, 优选将两种或更多本发明所述的不同宿主细胞进行共培养来制造两种或更多不同的 N- 多糖结构, 结果得到不同结构的糖蛋白的混合物。

[0761] 适用于 N- 多糖分析的仪器包括, 例如 ABI PRISM® 377 DNA 顺序分析仪 (Applied Biosystems)。数据分析可以使用例如 GENESCAN® 3.1 软件 (Applied Biosystems) 进行。N- 多糖分析的其他方法包括, 例如, 质谱法 (例如, MALDI-TOF-MS)、正相或反相高效液相色谱法 (HPLC) 及离子交换色谱法 (例如, 当多糖未被标记时采用脉冲安培检测, 如果多糖被适当标记则采用 UV 吸收或荧光检测)。

[0762] 一个优选实施方案是一种例如 IgG 的重组免疫球蛋白, 可以由本发明的细胞制造, 包含 Gal2GlcNAc2Man3GlcNAc2 结构的 N- 多糖。

[0763] 另一个更优选的实施例为重组人促红细胞生成素 (rhuEPO), 可以由本发明的细胞制造, 包含 NeuAc3Gal3GlcNAc3Man3GlcNAc2Fuc 结构的三种 N- 多糖。

[0764] 在优选实施方案中, 糖蛋白或糖蛋白组合物可以, 但不必须从所述宿主细胞中分离出来。在优选实施方案中, 糖蛋白或糖蛋白组合物可以, 但不必须从所述宿主细胞中进一步纯化。本文所用的术语“分离的”指一种已经从组分中被分开或纯化的分子或其片段, 例如, 所述组分为天然伴随其存在的蛋白质或其他天然存在的生物学或有机分子。典型地, 本发明所述的分离糖蛋白或糖蛋白组合物在一种药剂中构成同类分子总量的至少 60% (重量), 例如, 在一个样品中为同类分子总量的 60%。例如, 一种分离糖蛋白在药剂或样品中构成了总蛋白的至少 60% (重量)。在一些实施方案中, 药剂中的分离糖蛋白构成同类分子总量的至少 75%, 至少 90%, 或至少 99% (重量)。

[0765] 所述基因工程宿主细胞可以用于制造新的具有治疗的糖蛋白或其组合物的方法

中。

[0766] 由或可以由上述优选实施方案中的宿主细胞制造的优选糖蛋白或糖蛋白组合物包括但不限于,血液因子、抗凝血剂、溶血栓药、抗体、其抗原结合片段、激素、生长因子、刺激因子、趋化因子及细胞因子,更具体地,肿瘤坏死因子(TFN)家族的调节蛋白、促红细胞生成素(EPO)、促性腺激素、免疫球蛋白、粒细胞-巨噬细胞集落刺激因子、干扰素及酶。最优选的糖蛋白或糖蛋白组合物选自:促红细胞生成素(EPO)、干扰素- $[\alpha]$ 、干扰素- $[\beta]$ 、干扰素- $[\gamma]$ 、干扰素- $[\omega]$ 、及粒细胞-CSF、因子VIII、因子IX、人蛋白C、可溶性IgE受体 $[\alpha]$ -链、免疫球蛋白G(IgG)、IgG的Fab、IgM、尿激酶、胃促胰酶、尿素胰蛋白酶抑制剂、IGF结合蛋白、表皮生长因子、生长激素释放因子、annexin V融合蛋白、血管生成抑素、血管内皮生长因子-2、骨髓抑制因子-1、骨保护素、葡糖脑苷脂酶、半乳糖脑苷脂酶、 α -L-艾杜糖苷酸酶、 β -D-半乳糖苷酶、 β -葡萄糖苷酶、 β -氨基己糖苷酶、 β -D-甘露糖苷酶、 α -L-岩藻糖苷酶、芳基硫酸酯酶B、芳基硫酸酯酶A、 α -N-乙酰半乳糖胺酶、天冬氨酰葡糖胺酶、艾杜糖-2-硫酸酯酶、 α -氨基葡萄糖苷-N-乙酰转移酶、 β -D-葡萄糖苷酸酶、透明质酸酶、 α -L-甘露糖苷酶、 α -神经氨酸酶、磷酸转移酶、酸性脂肪酶、酸性神经酰胺酶、鞘磷脂酶、硫酸酯酶、组织蛋白酶K及脂蛋白脂酶。

[0767] 本发明的另一个实施方案为重組的治疗用蛋白或多个这种蛋白,构成一种或多种上述糖蛋白,尤其是具有上述低甘露糖多糖结构的糖蛋白。优选地,所述治疗用蛋白可以由本发明所述的细胞来制造。

[0768] 本发明的一个优选实施方案为一个或多个免疫球蛋白。另一个优选实施方案为抗体或包含一个或多个上述免疫球蛋白的抗体组合物。术语“免疫球蛋白”是指具有氨基酸序列的任一分子,由于特别与抗原相互影响并且其中,所述分子的任一链包括抗体可变区的功能性操纵区,无限制地包括所述分子的任一种天然存在或重組形式例如嵌合抗体或人源化抗体。本文所用的“免疫球蛋白”意指一种蛋白质,由一种或多种本质上由免疫球蛋白基因编码的多肽组成。优选地,本发明所述的免疫球蛋白包括片段,优选为包含一个或多个糖基化位点片段。所述片段意指具有抗原-抗体反应的抗体的片段,并包括 $F(ab')_2$, Fab' , Fab , Fv ,及重組 Fv 。

[0769] 本发明另一个优选实施方案为一种药物组合物,包含下述中的一种或多种:一种或多种本发明所述的上述糖蛋白或糖蛋白组合物,一种或多种本发明所述的上述重組治疗用蛋白,一种或多种本发明所述的上述免疫球蛋白,及一种或多种本发明所述的上述抗体。如果必要或可行,所述组合物还包含至少一种药学上可接受的载体或辅药。

[0770] 本发明所述的糖蛋白能在药物组合物中被明确地表示。除上述物质中的一种外,这些组合物可以包含药学上可接受的辅料、载体、缓冲剂、稳定剂或本领域技术人员公知的其他材料。所述材料应为无毒的并应当不影响成分的功效。所述载体或其他材料的确切性质可以取决于给药途径,例如,口服、静脉内、经皮或皮下、鼻腔、肌内、腹膜内或贴片途径。

[0771] 用于口服给药的药物组合物可以以片剂、胶囊、粉末或液体的形式。片剂可以包括固体载体例如明胶或辅药。液体药物组合物通常包括液体载体例如水、石油、动物或植物油、矿物油或合成油等。还可以包括生理盐水、葡萄糖或其他糖溶液或乙二醇类例如乙二醇、丙二醇或聚乙二醇。对于静脉内、经皮或皮下注射,或在疼痛处的注射,所述成分将以胃肠外可接受的水性溶液的形式注射,所述水性溶液无热原并具有适当的pH、等渗性及稳定

性。本领域相关技术的技术人员能很好地使用例如等渗介质来制备适当的溶液。根据要求还可以包括防腐剂、稳定剂、缓冲剂、抗氧化剂和 / 或其他添加剂。

[0772] 不管给个体施用的是本发明所述的多肽、肽、或核酸分子、其他药学上有效的化合物,优选以足以对个体显示出效果的“预防有效量”或“治疗有效量”(根据情况可以,虽然预防可以被认为是治疗)进行给药。实际给药量及给药的速度和时程将取决于被治疗的疾病的性质和严重性。治疗处方例如剂量的决定等,在全科医生和其他专科医生的责任范围内,并且通常考虑被治疗的病症、个体患者的条件、传递部位、给药方法及从业人员公知的其他因素。

[0773] 另一方面,本发明提供了一种治疗疾病的方法,所述疾病可以通过施用一种或多种上述经确认的糖蛋白或其组合物来治疗,所述方法包括步骤:给实验对象施用如上所述的糖蛋白或其组合物,其中,其中,所述实验对象患有或疑似患有可通过施用糖蛋白或其组合物治疗的疾病。在一个优选实施方案中,所述方法还包括步骤:(a) 提供实验对象和 / 或 (b) 确定所述实验对象是否患有可通过施用所述糖蛋白或其组合物治疗的疾病。所述实验对象为动物例如人。所述疾病为例如癌、免疫紊乱、炎症或代谢紊乱。

[0774] 根据本发明,还提供了一种用于制造糖蛋白的试剂盒,所述试剂盒至少包括:本发明所述的一种或多种能够制造重组蛋白的宿主细胞,及优选地,用于培养所述细胞以制造所述重组蛋白的培养基。

[0775] 序列表

[0776] SEQ ID NO :1 代表编码 f1c2 基因 (图 5A) 的截短片段 f1c2' 的核苷酸序列。

[0777] ATGATCTTCCTAAACACCTTCGCAAGGTGCCTTTTAACGTGTTTCGTACTGTGCAGCGG TACAGCAC GTTCCTCTGACACAAACGACACTACTCCGGCGTCTGCAAAGCATTTGCAGACCACT TCTTTATTGACGTGTATGG ACAATTCGCAATTAACGGCATCATTCTTTGATGTGAAATTTTACC CCGATAATAATACTGTTATCTTTGATATTG ACGCTACGACGACGCTTAATGGGAACGTCACTGT GAAGGCTGAGCTGCTTACTTACGGACTGAAAGTCCTGGATA AGACTTTTGATTTATGTTCCCTTG GGCCAAGTATCGCTTTCCCCCTAAGTGCTGGGCGTATTGATGTCATGTCCA CACAGGTGATCG AATCATCCATTACCAAGCAATTTCCCGGCATTGCTTACACCATTCCAGATTTGGACGCACAAG T ACGTGTGGTGGCATAACGCTCAGAATGACACGGAATTCGAAACTCCGCTGGCTTGTGTCCAGGCT ATCTTGAGT AACGGGAAGACAGTGCAAAACAAAGTATGCGGCCTGGCCATTGCCGCTATCTCAG GTGTCCGTGTACTTACCTCA GGGTTTGTGTCTGTGATCGGTTACTCAGCCACTGCTGCTCACAT TGCGTCCAACTCCATCTCATTGTTTCATATAC TTCCAAAATCTAGCTATCACTGCAATGATGGGT GTCTCAAGGGTCCACCCATTGCTGCCGCGTGGACGCAGAAT TTCCAATGGTCCATGGGTATCA TCAATACAACTTCATGCAAAAGATTTTTGATTGGTACGTACAGGCCACTAAT GGTGTCTCAAA TGTGTGTGGTAGCTAACAAGGACGTCTTGTCCATTAGTGTGCAAAAACGTGCTATCTCTATGGC A TCGTCTAGTGATTACAATTTTGACACCATTTTAGACGATTCGGATCTGTACACCACTTCTGAGA AGGATCCAA GCAATTACTCAGCCAAGATTCTCGTGTTAAGAGGTATAGAAAGAGTTGCTTATTT GGCTAATATTGAGCTATCTA ATTTCTTTTTGACCGGTATTGTGTTTTTTCTATTCTTCCTATTT GTAGTTGTCGTCTCTTTGATTTTCTTTAAGG CGCTATTGGAAGTTCTTACAAGAGCAAGAATAT TGAAAGAGACTTCCAATTTCTTCCAATATAGGAAGAAGTGGG GGAGTATTATCAAAGGCACCCT TTTCAGATTATCTATCATCGCCTCCCTCAAGTTTCTTCTTGGCGATTTGGG AATTTACTCAG GTCAACTCTCCAGCGATTGTTGTTGATGCGGTAGTAATATTACTGATCGATCCTCTAGAGTCGA CCTGCAGGCATGCAAGCTAG

[0778] SEQ ID NO :2 代表 f1c2' (图 5B) 的氨基酸序列。

[0779] MIFLNTFARCLLTCFVLCSTARSSDTNDTPASAKHLQTTSLTLCMDNSQLTASFFDV KFYPDNNT VIFDIDATTTLNGNVTVKAELLYGLKVLDKTFDLCSLGQVSLSPLSAGRIDVMSTQVIESSITKQFPGIAYTIPDLDAQVRVVAYAQNDEFETPLACVQAILSNGKTVQTKYAAWPIA AISGVGVLTSGFVSVIGYSATAAHIASNSISLFIYFQNLAITAMMGVSRVPPIAAAWTQNFQWS MGIINTNFMQKIFDWWYVQATNGVSNVVVANKDVLSSISVQKRAISMASSSDYNFDTILDDSDLYT TSEKDPSNYSKILVLRGIERVAYLANIELSNFFLTGIVFFLFFLVVVVSLIFFKALLEVLTR ARILKETSNNFFQYRKNWGSIIKGTFLRLSIIAFPQVSLLAIWFEFTQVNSPAIVDVAVILLIDPLESTCRHAS

[0780] SEQ ID NO :3 代表编码内质网定位信号和编码 f1c2' 编码区的跨膜域 (TM) 1 至 3(TM1-3) 的核苷酸序列。

[0781] ATGATCTTCCTAAACACCTTCGCAAGGTGCCTTTTAACGTGTTTCGTACTGTGCAGCGG TACAGCACG TTCCTCTGACACAAACGACACTACTCCGGCGTCTGCAAAGCATTTGCAGACCACT TCTTTATTGACGTGTATGGACAATTCGCAATTAACGGCATCATTCTTTGATGTGAAATTTTACC CCGATAATAATACTGTTATCTTTGATATTGACGCTACGACGACGCTTAATGGGAACGTCACTGT GAAGGCTGAGCTGCTTACTTACGGACTGAAAGTCCTGGATAAGAC TTTTGATTTATGTTTCCTTG GGCCAAGTATCGCTTTCCCCCTAAGTGCTGGGCGTATTGATGTCATGTCCACACAGGTGATCG AATCATCCATTACCAAGCAATTTCCCGGCATTGCTTACACCATTCCAGATTTGGACGCACAAGT ACGTGTGGTGGCATAACGCTCAGAATGACACGGAATTCGAAACTCCGCTGGCTTGTGTCCAGGCT ATCTTGAGTAACGGGAAGACAGTGCAAAACAAAGTATGCGGCCTGGCCATTGCCGCTATCTCAG GTGTCCGTGTACTTACCTCAGGGTTTGTGTCTGTGATCGTTACTCAGCCACTGCTGCTCATAT TGCGTCCAATCCATCTCATTGTTTCATATACTTCCAAAATCTAGCTATCACTGCAATGATGGGT GTCTCAAGGGTTCCACCCATTGCTGCCGCGTGGACGCAGAATTTCCAATGGTCCATGGGTATCA TCAATACAACTTCATGCAAAAGATTTTGTATTGGTACGTACAGGCCACTAATGGTGTCTCAAA TGTTGTGGTAGCTAACAAGGACGTCTTGTCCATTAGTGTGCAAAAACGTGCTATCTCTATGGCA TCGTCTAGTGAT TACAATTTTGACACCATTTTAGACGATTCGGATCTGTACACCACTTCTGAGA AGGATCCAAGCAATTACTCAGCCAAGATTCTCGTGTTAAGAGGTATAGAAAGAGTTGCTTATTT GGCTAATATTGAGCTATCTAATTTCTTTTTGACCGGTATTGTGTTTTTTCTATTCTTCTATTT GTAGTTGTCGTCTCTTTGATTTTCTTTAAGTAG

[0782] SEQ ID NO :4 代表内质网定位信号和 f1c2' 的跨膜域 (TM) 1 至 3(TM1-3) 的氨基酸序列。

[0783] MIFLNTFARCLLTCFVLCSTARSSDTNDTPASAKHLQTTSLTLCMDNSQLTASFFDV KFYPDNNTVIFDIDATTTLNGNVTVKAELLYGLKVLDKTFDLCSLGQVSLSPLSAGRIDVMST QVIESSITKQFPGIAYTIPDLDAQVRVVAYAQNDEFETPLACVQAILSNGKTVQTKYAAWPIA AISGVGVLTSGFVSVIGYSATAAHIASNSISLFIYFQNLAITAMMGVSRVPPIAAAWTQNFQWS MGIINTNFMQKIFDWWYVQATNGVSNVVVANKDVLSSISVQKRAISMASSSDYNFDTILDDSDLYT TSEKDPSNYSKILVLRGIERVAYLANIELSNFFLTGIVFFLFFLVVVVSLIFFK

[0784] SEQ ID NO :5 代表编码内质网定位信号和编码 f1c2' 编码区的跨膜域 (TM) 1 至 2(TM1-2) 的核苷酸序列。

[0785] ATGATCTTCCTAAACACCTTCGCAAGGTGCCTTTTAACGTGTTTCGTACTGTGCAGCGG TACAGCACG TTCCTCTGACACAAACGACACTACTCCGGCGTCTGCAAAGCATTTGCAGACCACT TCTTTATTGACGTGTATGGACAATTCGCAATTAACGGCATCATTCTTTGATGTGAAATTTTACC CCGATAATAATACTGTTATCTTTGATATTGACGCTACGACGACGCTTAATGGGAACGTCACTGT GAAGGCTGAGCTGCTTACTTACGGACTGAAAGTCCTGGATAAGAC TTTTGATTTATGTTTCCTTG GGCCAAGTATCGCTTTCCCCCTAAGTGCTGGGCGTATTGATGTCATGTCCACACAGGTGATCG AATCATCCATTACCAAGCAATTTCCCGGCATTGCTTACACCATTCCAGATTTGGACGCACAAGT ACGT

GTGGTGGCATACGCTCAGAATGACACGGAATTCGAAACTCCGCTGGCTTGTGTCCAGGCT ATCTTGAGTAACGGGA
AGACAGTGCAAACAAAGTATGCGGCCTGGCCATTGCCGCTATCTCAG GTGTCGGTGTACTTACCTCAGGGTTTGT
GTCTGTGATCGGTTACTCAGCCACTGCTGCTCACAT TCGTCCAACCTCCATCTCATTGTTTCATATACTTCCAAAAT
CTAGCTATCACTGCAATGATGGGT GTCTCAAGGGTTCACCCATTGCTGCCGCGTGGACTAG

[0786] SEQ ID NO :6 代表内质网定位信号和 f1c2' 的跨膜域 (TM) 1 至 2(TM1-2) 的氨基酸序列。

[0787] MIFLNTFARCLLTCFVLCSGTARSSDTNDTPASAKHLQTTSLLTCMDNSQLTASFFDV KFYPDNNTV
IFDIDATTTLNGNVTVKAELLTYGLKVLDKTFDLCSLGQVSLSPLSAGRIDVMST QVIESSITKQFPGIAYTIPDL
DAQVRVVAYAQNDFETPLACVQAILSNGKTVQTKYAAWPIA AISGVGVLTSGFVSVIGYSATAAHIASNSISLF
IYFQNLAITAMMGVSRVPPIAAAWT

[0788] SEQ ID NO :7 代表编码内质网定位信号和编码 f1c2' 编码区的跨膜域 (TM) 2 至 4(TM2-4) 的核苷酸序列。

[0789] ATGATCTTCCTAAACACCTTCGCAAGGTGCCTTTTAACGTGTTTCGTACTGTGCAGCGG TACAGCACG
TTCCTCTGACACAAACGACATTGCGTCCAACCTCCATCTCATTGTTTCATATACTTC CAAAATCTAGCTATCACTGCA
ATGATGGGTGTCTCAAGGGTTCACCCATTGCTGCCGCGTGGA CGCAGAATTTCCAATGGTCCATGGGTATCATCA
ATACAACTTCATGCAAAAGATTTTTGATTG GTACGTACAGGCCACTAATGGTGTCTCAAATGTTGTGGTAGCTAA
CAAGGACGTCTTGCCATT AGTGTGCAAAAACGTGCTATCTCTATGGCATCGTCTAGTGATTACAATTTTGACACC
ATTTTAG ACGATTCGGATCTGTACACCACTTCTGAGAAGGATCCAAGCAATTACTCAGCCAAGATTCTCGT GTTA
AGAGGTATAGAAAGAGTTGCTTATTTGGCTAATATTGAGCTATCTAATTTCTTTTTGACC GGTATTGTGTTTTTTC
TATTCTTCCTATTTGTAGTTGTCGTCTCTTTGATTTTCTTTAAGGCGC TATTGGAAGTTCTTACAAGAGCAAGAAT
ATTGAAAGAGACTTCCAATTTCTTCCAATATAGGAA GAACTGGGGGAGTATTATCAAAGGCACCCTTTTCAGATTA
TCTATCATCGCCTTCCCTCAAGTT TCTCTTCTGGCGATTTGGGAATTTACTCAGGTCAACTCTCCAGCGATTGTTG
TTGATGCGGTAG TAATATTACTGATCGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTAG

[0790] SEQ ID NO :8 代表内质网定位信号和 f1c2' 的跨膜域 (TM) 2 至 4(TM2-4) 的氨基酸序列。

[0791] MIFLNTFARCLLTCFVLCSGTARSSDTNDIASNSISLFIYFQNLAITAMMGVSRVPPIA AAWTQNFQW
SMGIINTNFMQKIFDWYVQATNGVSNVVVANKDVLISVQKRAISMSSSDYNFD TILDDSDLYTTSEKDPSNYS
KILVLRGIERVAYLANIELSNFFLTGIVFFLFFLFVVVVSLIF FKALLEVLTRARILKETSNNFFQYRKNWGSIIKG
TLFRLSIIAFPQVSLLAIWFTQVNSPAIVV DAVVILLIDPLESTCRHAS

[0792] SEQ ID NO :9 代表编码内质网定位信号和编码 f1c2' 编码区的跨膜域 (TM) 3 至 4(TM3-4) 的核苷酸序列。

[0793] ATGATCTTCCTAAACACCTTCGCAAGGTGCCTTTTAACGTGTTTCGTACTGTGCAGCGG TACAGCACG
TTCCTCTGACACAAACGACTTCTTTTTGACCGGTATTGTGTTTTTCTATTCTTC CTATTTGTAGTTGTGCTCTCT
TTGATTTTCTTTAAGGCGCTATTGGAAGTTCTTACAAGAGCAA GAATATTGAAAGAGACTTCCAATTTCTTCCAAT
ATAGGAAGAACTGGGGGAGTATTATCAAAGG CACCCTTTTCAGATTATCTATCATCGCCTTCCCTCAAGTTTCTCT
TCTGGCGATTTGGGAATTT ACTCAGGTCAACTCTCCAGCGATTGTTGTTGATGCGGTAGTAATATTACTGATCGAT
CCTCTAG AGTCGACCTGCAGGCATGCAAGCTAG

[0794] SEQ ID NO :10 代表内质网定位信号和 f1c2' 的跨膜域 (TM) 3 至 4(TM3-4) 的氨基酸序列。

[0795] MIFLNTFARCLLTCFVLCSGTARSSDTNDFFLTGIVFFLFFLFV VVSLIFFKALLEVL TRARILKET SNFFQYRKNWGSIIKGTFLRSLIIAFPQVSL LAIWEFTQVNSPAIVVDVAVILLI DPLESTCRHAS

[0796] SEQ ID NO :11 代表编码内质网定位信号和编码 f1c2' 编码区的跨膜域 (TM) 1 (TM1) 的核苷酸序列。

[0797] ATGATCTTCCTAAACACCTTCGCAAGGTGCCTTTTAACGTGTTTCGTACTGTGCAGCGG TACAGCACG TTCCTCTGACACAAACGACACTACTCCGGCGTCTGCAAAGCATTTGCAGACCACT TCTTTATTGACGTGTATGGAC AATTCGCAATTAACGGCATCATTCTTTGATGTGAAATTTTACC CCGATAATAATACTGTTATCTTTGATATTGACG CTACGACGACGCTTAATGGGAACGTCAGTGT GAAGGCTGAGCTGCTTACTTACGGACTGAAAGTCCTGGATAAGAC TTTTGATTTATGTTCTTG GGCCAAGTATCGCTTCCCCCTAAGTGCTGGGCGTATTGATGTCATGTCCACACAG GTGATCG AATCATCCATTACCAAGCAATTTCCCGGCATTGCTTACACCATTCCAGATTTGGACGCACAAGT ACGT GTGGTGGCATA CGCTCAGAATGACACGGAATTCGAAACTCCGCTGGCTTGTGTCCAGGCT ATCTTGAGTAACGGGA AGACAGTGCAAACAAAGTATGCGGCCTGGCCATTGCCGCTATCTCAG GTGTGCGGTGTACTTACCTCAGGGTTTGT GTCTGTGATCGGTTACTCATAG

[0798] SEQ ID NO :12 代表内质网定位信号和 f1c2' 的跨膜域 (TM) 1 (TM1) 的氨基酸序列。

[0799] MIFLNTFARCLLTCFVLCSGTARSSDTNDTPASAKHLQTTSL LTCMDNSQLTASFFDV KFYPDNNTV IFDIDATTTLNGNVTVKAELLTYGLKVL DKTDFDLCSLGQVSLSPLSAGRIDVMST QVIESSITKQFPGIAYTIPDL DAQVRVVAYAQN DTEFETPLACVQAILSNKTVQTKYAAWPIA AISGVGVLTSGFVSVIGYS

[0800] SEQ ID NO :13 代表编码内质网定位信号和编码 f1c2' 编码区的跨膜域 (TM) 2 (TM2) 的核苷酸序列。

[0801] ATGATCTTCCTAAACACCTTCGCAAGGTGCCTTTTAACGTGTTTCGTACTGTGCAGCGG TACAGCACG TTCCTCTGACACAAACGACATTGCGTCCAACCTCCATCTCATTGTT CATATACTTC CAAAATCTAGCTATCACTGCA ATGATGGGTGTCTCAAGGGTTCCACCCATTGCTGCCGCGTGGA CTAG

[0802] SEQ ID NO :14 代表内质网定位信号和 f1c2' 的跨膜域 (TM) 2 (TM2) 的氨基酸序列。

[0803] MIFLNTFARCLLTCFVLCSGTARSSDTNDIASNSISLFIYFQNLAITAMMGVSRVPPIA AAWT

[0804] SEQ ID NO :15 代表编码内质网定位信号和编码 f1c2' 编码区的跨膜域 (TM) 3 (TM3) 的核苷酸序列。

[0805] ATGATCTTCCTAAACACCTTCGCAAGGTGCCTTTTAACGTGTTTCGTACTGTGCAGCGG TACAGCACG TTCCTCTGACACAAACGACTTCTTTTTGACCGGTATTGTGTTTTTCTATTCTTC CTATTTGTAGTTGTGCTCTCT TTGATTTTCTTTAAGTAG

[0806] SEQ ID NO :16 代表内质网定位信号和 f1c2' 的跨膜域 (TM) 3 (TM3) 的氨基酸序列。

[0807] MIFLNTFARCLLTCFVLCSGTARSSDTNDFFLTGIVFFLFFLFV VVSLIFFK

[0808] SEQ ID NO :17 代表编码内质网定位信号和编码 f1c2' 编码区的跨膜域 (TM) 4 (TM4) 的核苷酸序列。

[0809] ATGATCTTCCTAAACACCTTCGCAAGGTGCCTTTTAACGTGTTTCGTACTGTGCAGCGG TACAGCACG TTCCTCTGACACAAACGACGGCACCCTTTTCAGATTATCTATCATCGCCTTCCCT CAAGTTTCTCTTCTGGCGATT TGGGAATTTACTCAGGTCAACTCTCCAGCGATTGTTGTTGATG CGGTAGTAATATTACTGATCGATCCTCTAGAGT CGACCTGCAGGCATGCAAGCTAG

[0810] SEQ ID NO :18 代表内质网定位信号和 f1c2' 的跨膜域 (TM) 4 (TM4) 的氨基酸序列。

[0811] MIFLNTFARCLLTCFVLCSGTARSSDTNDGTLFRLSIIAFPQVSL LAIWEFTQVNSPAI

VVDAVVILLIDPLESTCRHAS

[0812] SEQ ID NO :19 代表编码 f1c2' 编码区的内质网定位信号的核苷酸序列。

[0813] ATGATCTTCCTAAACACCTTCGCAAGGTGCCTTTTAAACGTGTTTCGTA CTGTGCAGCGG
TACAGCACGTTCC

[0814] SEQ ID NO :20 代表 F1c2' 的内质网定位信号的氨基酸序列。

[0815] MIFLNTFARCLLTCFVLCSGTARS

[0816] SEQ ID NO :21 代表编码 f1c2' 的第一个跨膜域 (TM1) 的核苷酸序列。

[0817] GCCTGGCCCATTTGCCGCTATCTCAGGTGTCGGTGTACTTACCTCAGGGTTTGTGTCTGT
GATCGGTTAC

[0818] SEQ ID NO :22 代表 F1c2' 的第一个跨膜域 (TM1) 的氨基酸序列。

[0819] AWPIAAISGVGVLTSGFVSVIGY

[0820] SEQ ID NO :23 代表编码 f1c2' 的第二个跨膜域 (TM2) 的核苷酸序列。

[0821] ATTGCGTCCAACTCCATCTCATTGTTTCATATACTTCCAAAATCTAGCTATCACTGCAAT GATGGGTGT
CTCAAGGGTTCCACCCATTGCTGCCGCGTG

[0822] SEQ ID NO :24 代表 F1c2' 的第二个跨膜域 (TM2) 的氨基酸序列。

[0823] IASNSISLFIYFQNLAITAMMGVSRVPPIAAAW

[0824] SEQ ID NO :25 代表编码 f1c2' 的第三个跨膜域 (TM3) 的核苷酸序列。

[0825] TTCTTTTTGACCGGTATTGTGTTTTTTCTATTCTTCCTATTTGTAGTTGTCGTCTCTTT
GATTTTCTTT

[0826] SEQ ID NO :26 代表 F1c2' 的第三个跨膜域 (TM3) 的氨基酸序列。

[0827] FFLTGIVFFLFFLVVVVSLIFF

[0828] SEQ ID NO :27 代表编码 f1c2' 的第四个跨膜域 (TM4) 的核苷酸序列。

[0829] GGCACCCTTTTTCAGATTATCTATCATCGCCTTCCCTCAAGTTTCTTCTTGCGGATTTG G

[0830] SEQ ID NO :28 代表 F1c2' 的第四个跨膜域 (TM4) 的氨基酸序列。

[0831] GTLFRLSIIAFPQVSLIAIW

[0832] SEQ ID NO :29 代表编码 f1c2' 的第五个跨膜域 (TM5) 的核苷酸序列。

[0833] GTAGTAATATTACTGAT

[0834] SEQ ID NO :30 代表 F1c2' 的第五个跨膜域 (TM5) 的氨基酸序列。

[0835] VVILLI

[0836] SEQ ID NO :31 代表 F1c2' 表达质粒 YEp352F1c2' (图 13) 的核苷酸序列,所述质粒带有包含 ACS 启动子 (1..399) 的 F1c2' 表达盒、具有潜在终止密码子 (1753..1758) 的 f1c2' (400..1722)。

[0837] SEQ ID NO :32 代表 LmStt3D 及 F1c2' 共表达质粒 pAX306f (图 14) 的核苷酸序列,所述质粒包括:包含 ACS 启动子 (1..399) 的 F1c2' 表达盒、f1c2' ORF (400..1722)、在终止密码子后为 CYC1 终止子 (6904..7155),并且还包括在反方向包含 LmStt3D ORF (补体) (7192..9762) 及强组成型 GPD 启动子 (补体) (9781..10435) 的 POT LmStt3D 表达盒; LmStt3D 的 ATG 正好在 GPD 启动子之后。

[0838] SEQ ID NO :33 代表编码 *Leishmania braziliensis* 的旁系同源基因 LbStt3-1 的核苷酸序列。

- [0839] SEQ ID NO :34 代表 LbStt3-1 的氨基酸序列。
- [0840] SEQ ID NO :35 代表编码 *Leishmania braziliensis* 的旁系同源基因 LbStt3-2 的核苷酸序列。
- [0841] SEQ ID NO :36 代表 LbStt3-2 的氨基酸序列。
- [0842] SEQ ID NO :37 代表编码 *Leishmania braziliensis* 的旁系同源基因 LbStt3-3 的核苷酸序列。
- [0843] SEQ ID NO :38 代表 LbStt3-3 的氨基酸序列。
- [0844] SEQ ID NO :39 代表编码 *Leishmania infantum* 的旁系同源基因 LiStt3-1 的核苷酸序列。
- [0845] SEQ ID NO :40 代表 LiStt3-1 的氨基酸序列。
- [0846] SEQ ID NO :41 代表编码 *Leishmania infantum* 的旁系同源基因 LiStt3-2 的核苷酸序列。
- [0847] SEQ ID NO :42 代表 LiStt3-2 的氨基酸序列。
- [0848] SEQ ID NO :43 代表编码 *Leishmania infantum* 的旁系同源基因 LiStt3-3 的核苷酸序列。
- [0849] SEQ ID NO :44 代表 LiStt3-3 的氨基酸序列。
- [0850] SEQ ID NO :45 代表编码 *Leishmania major* 的旁系同源基因 LmStt3A 的核苷酸序列。
- [0851] SEQ ID NO :46 代表 LmStt3A 的氨基酸序列。
- [0852] SEQ ID NO :47 代表编码 *Leishmania major* 的旁系同源基因 LmStt3B 的核苷酸序列。
- [0853] SEQ ID NO :48 代表 LmStt3B 的氨基酸序列。
- [0854] SEQ ID NO :49 代表编码 *Leishmania major* 的旁系同源基因 LmStt3C 的核苷酸序列。
- [0855] SEQ ID NO :50 代表 LmStt3C 的氨基酸序列。
- [0856] SEQ ID NO :51 代表编码 *Leishmania major* 的旁系同源基因 LmStt3D 的核苷酸序列。
- [0857] SEQ ID NO :52 代表 LmStt3D 的氨基酸序列。
- [0858] SEQ ID NO :53 代表编码 *Trypanosoma brucei* 的旁系同源基因 TbStt3A 的核苷酸序列。
- [0859] SEQ ID NO :54 代表 TbStt3A 的氨基酸序列。
- [0860] SEQ ID NO :55 代表编码 *Trypanosoma brucei* 的旁系同源基因 TbStt3B 的核苷酸序列。
- [0861] SEQ ID NO :56 代表 TbStt3B 的氨基酸序列。
- [0862] SEQ ID NO :57 代表编码 *Trypanosoma brucei* 的旁系同源基因 TbStt3C 的核苷酸序列。
- [0863] SEQ ID NO :58 代表 TbStt3C 的氨基酸序列。
- [0864] SEQ ID NO :59 代表编码 *Trypanosoma cruzi* 的旁系同源基因 TbStt3 的核苷酸序列。

[0865] SEQ ID NO:60 代表 TbStt3 的氨基酸序列。

[0866] SEQ ID NO:61 代表编码 flc2' 的内源性启动子部分的核苷酸序列。

[0867] 实施例

[0868] 实施例 1 具有 Man3GlcNAc2 结构的糖蛋白的制造

[0869] 1.1 酵母培养基及方法

[0870] 除非另外说明,所有菌株都在 YPD 培养基上生长。菌株 YG1137 在 YPGal 上保持。除非另外说明,菌株 YCN1(Δ rft1), YG1363(Δ alg3 Δ alg11), YG1365(Δ alg11), 及 YG1830(alg2-1) 在补充有 1M 山梨醇的培养基中生长。

[0871] 1.2 菌株构建

[0872] 通过整合包括 *S. cerevisiae* HIS3 位点的 PCR 产物,在 SS328XSS330 中取代整个 Alg11 开放阅读框。使转化的酵母菌株 YG1141(MATa/ α ade2-201/ade2-201 ura3-52/ura3-52 his3 Δ 200/his3 Δ 200 tyr1/+lys2-801/+ Δ alg11::HIS3/+) 产生孢子,并分裂四分体以获得 Δ alg11 单倍体,及与 YG248(MATa Δ alg3::HIS3 ade2-101 his3 Δ 200 lys2-801 ura3-52) 配对的 YG1361(MAT α ade2-201 ura3-52 his3 Δ 200 Δ alg11::HIS3)。使得到的二倍体 YG1362(MATa/ α ade2-201/ade2-201 ura3-52/ura3-52 his3 Δ 200/his3 Δ 200 lys2-801/+ Δ alg3::HIS3 Δ alg11::HIS3/+) 在包括 1M 山梨醇的平板上产生孢子,以获得单倍体菌株 YG1365(MAT α ade2-101 ura3-52 his3 Δ 200 Δ alg11::HIS3) 和 YG1363(MAT α ade2-101 ura3-52 his3 Δ 200 lys2-801 Δ alg3::HIS3 Δ alg11::HIS3)。通过在二倍体菌株中用 HIS3 框架取代 rft1 基因、所述得到的二倍体杂交菌株的孢子形成及所述得到的单倍体 Δ rft1::HIS3 菌株 (YCN1) 的选择来制造 Δ rft1 菌株。

[0873] 1.3 蛋白分析

[0874] 如上所述进行蛋白提取及免疫印迹分析。抗 CPY 抗体被稀释 3000 倍。

[0875] 1.4 脂质和蛋白连接寡糖分析

[0876] 如上所述,标记、提取并分析脂质连接寡糖。简单地说,在 YPD 中培育酵母细胞 (50ml 培养物在 546nm 处的吸光度为 1),并在用有机溶剂溶解细胞之前于包括 [3H]-甘露糖的介质中孵化。使用有机溶剂提取脂质连接寡糖并通过温和酸水解释放寡糖。使用流通计数的氨基柱,采用 HPLC 法分析释放的寡糖。由总计数划分的每分钟计数的数目被包括在内。一个样本中总信号的百分比为使用两种测定方法的平均值。在脂质连接寡糖提取之后,从细胞碎片中纯化 N-连接寡糖。将碎片中的蛋白质溶解 (在 100 $^{\circ}$ C 10min) 于 0.2ml 1% SDS, 50mmol/l Tris-HCl, 1% - 巯基乙醇中。离心 (2min, 15,000g) 后补足上清液至 0.25ml 中 NP40 为 1% (v/v), 并使用 PNGaseF (2 单位, 于 37 $^{\circ}$ C 过夜) 消化掉蛋白连接寡糖。用 0.75ml 甲醇沉淀蛋白并于 15,000g 条件下旋转样品 20min。将上清液干燥并重新混悬于 0.2ml 70 : 30 乙腈 : 水中, 取 0.1ml 采用上述的 HPLC 进行分析。

[0877] 1.5 MALDI-TOF-MS

[0878] 对于来源于细胞壁蛋白质的 N-多糖的分析,使用玻璃珠在 10mmol/l Tris 中破碎细胞,并在含有 2M 硫脲、7mol/l 尿素、pH 8.0 的 2% SDS 50mmol/l Tris 及 10mmol/l DTT 的缓冲液中减少不溶的细胞壁碎片。在含有 25mmol/l 碘代乙酰胺的相同缓冲液中于 37 $^{\circ}$ C 剧烈振摇 1h 进行烷基化。离心收集所述细胞壁碎片并在 50mmol/l NH_4CO_3 中清洗得到的球状物。

[0879] 在含有 1x 变性缓冲液、pH7.5 的 50mmol/l 磷酸缓冲液及 1% NP-40 的缓冲液中使用 1 μ l PNGase F, 于 37°C 过夜, 释放 N- 多糖。用 C18 和碳柱纯化 N- 多糖, 并浓缩含有 N- 多糖的洗脱液。用 2- 氨基苯甲酰胺标记 N- 多糖并使用碳柱进行最后纯化。使用 Autoflex MALDI-TOF MS (Bruker Daltonics, **Fällanden**, Switzerland) 在正离子模式下并在反射模式操作下获得纯化的 N- 多糖制剂的质谱。测定 m/z 的范围为 800 3000。

[0880] 1.6 高拷贝抑制基因筛选

[0881] 对于高拷贝抑制基因筛选, 将 1 μ g 包含部分消化的被导入载体 YEp352 (Hill et al., 1986) 中的酵母染色体 DNA 的基因组文库 (Stagljar et al., 1994) 通过电穿孔转化进入 1x10⁹ YNC1 (Δ rft1) 细胞, 并用 1M 山梨醇于 25°C 在缺乏尿嘧啶的基本培养基上选择转化株。通过于 33°C 在 YPD 和 YPDS 上影印培养来检测转化株的生长情况。检测阳性菌落 (于 33°C 生长在 YPD 和 YPDS 上) 对支持 Δ rft1 在 33、35 及 37°C 下生长的能力。通过提取总酵母 DNA 来分离表现出全部或部分抑制的菌落的质粒 DNA, 并将其用于 E. coli 菌株 DH5 中的质粒扩增。重新转化回收的质粒并检测它们对支持 Δ rft1 在 YPD 上 33、35 及 37°C 下生长的能力。进一步分析 64 C. tes 对改善 Δ rft1 细胞中的糖基化的能力。用 M13 (GTA AAA CGA CGG CCA GT) 及 M13rev (GAG CGG ATA ACA ATT) 引物确定选择性高拷贝抑制基因质粒的顺序。

[0882] 1.7 斑点分析

[0883] 为评价酵母菌株或酵母突变菌株例如表达 Rft1 或 Flc2' 或其片段的 Δ rft1, Δ alg11 或 Δ alg2 突变菌株的生长, 对这些菌株进行斑点分析。将菌株过夜培养并调节培养物至相等的细胞密度。将系列稀释物涂在琼脂平板上, 并将所述平板孵化 3 天至指定温度。

[0884] 在液体介质中的生长分析按如下方法进行, 将用单一菌落接种的预培养液在 5ml 缺少尿嘧啶的 SD 介质中培育 48h 来维持质粒, 并用 1mol/l 山梨醇补足。在 600nm 测定细胞密度。以达到起始细胞密度 0.05 相等的细胞量接种 25ml 相同的介质来进行生长分析。将细胞在回转摇床上以 200rpm 于 23°C 或 30°C 培育 48h。在指定时间点测量细胞密度。

[0885] 1.8 Man3GlcNAc2 结构的产生

[0886] 脂质连接寡糖 (LLO) 作为内质网 (ER) 中寡糖基转移酶的底物的代表, 转移装配好的糖脂 N- 糖基化共有序列的天冬酰胺残基。所述脂质连接寡糖的构建是序列加工, 其中, 来源于被激活的糖供体的糖被添加到生长的脂质连接寡糖结构上。所述脂质连接寡糖合成的详细途径如图 1 所述。通过从细胞中去除特定的转移酶而产生特定的脂质连接寡糖结构。

[0887] 不希望受理论的束缚, 发明人已经发现, 在所述过程中, 蛋白质 Alg3p 及 Alg11p 在脂质连接寡糖结构的构建中起到很重要的作用。通过 Alg11p 的靶向去除, 可以成功地阻止 A- 分支的合成, 从而导致主要产生 Man6GlcNAc2 及 Man7GlcNAc2 结构 (图 2A)。在本发明的宿主细胞中, Man3GlcNAc2 结构能在内质网的胞质面被合成, 然后被翻转进入内质网腔, 作为定位于内质网腔的转移酶的底物。此外, Alg3p, 作为催化发起 B- 分支的 (1,3)- 甘露糖导入的酶, 也被认定在被翻转的脂质连接寡糖底物的加工中起到很重要的主要。Alg3p 的去除不仅阻止 B- 分支的形成, 而且 (1,3)- 甘露糖的存在对于 C- 分支的形成是必要的。因此, 提供了突变酵母菌株或类似菌株, 缺乏 Alg3p 和 Alg11p 这两种。因此, 本发明的宿主细

胞主要制造并优选仅制造低甘露糖,尤其是 Man3GlcNAc2 多糖结构,例如,通过 [3H]-甘露糖标记和制造的脂质连接寡糖结构的 HPLC 图(图 2B)所揭示的。

[0888] 使用 [3H]-甘露糖标记的蛋白连接寡糖(NLO)分析揭示了在 $\Delta alg3 \Delta alg11$ 菌株中的结构大于 Man3GlcNAc2,但小于在 $\Delta alg11$ 菌株中制造的 N-多糖结构(图 3B)。

[0889] 使用分离自细胞壁蛋白质的 2-AB 标记 N-多糖的 MALDI-TOF MS 进一步表征所述结构(图 4)。在野生型酵母中,包含除 GlcNAc2 之外的 8 和 9 个己糖残基的多糖排列存在于还原端(图 4A),与此相反,在 $\Delta alg11$ 菌株中,在还原端主要检测到包含除 GlcNAc2 之外的 5 至 9 个己糖残基的 N-多糖(图 4B),在 $\Delta alg3 \Delta alg11$ 菌株中,检测到 Man3GlcNAc2(m/z 1053)的小碎片及 Man4GlcNAc2(m/z 1215)和 Man5GlcNAc2(m/z 1377)的大碎片(图 4C)。

[0890] 总的说来,脂质连接寡糖和蛋白连接寡糖的分析表明,在 $\Delta alg3 \Delta alg11$ 菌株中,Man3GlcNAc2 在内质网中被制造并被转移至蛋白质,但是所述结构进一步在高尔基体中被修饰。

[0891] 1.9 高拷贝抑制基因筛选-新翻转酶的鉴定

[0892] 高拷贝抑制基因筛选(HCSS)是一种优选和有效的方法,用于选择提供期望表型的基因。

[0893] 为了鉴定能补偿必要的 Rft1 功能缺乏的基因,在 $\Delta rft1$ 菌株中进行了高拷贝抑制基因筛选。基因组酵母 DNA 文库从突变菌株的高拷贝质粒 Yep352 中被表达。

[0894] 用 1mol/l 山梨醇于 25°C 在缺乏尿嘧啶的基本培养基上选择转化株。通过于 33°C 在 YPD 和 YPDS 上影印培养来检测转化株的生长情况。检测阳性菌落(于 33°C 生长在 YPD 和 YPDS 上)对支持 $\Delta rft1$ 在 33、35 及 37°C 下生长的能力。进一步分析 64 C. tes 对改善 $\Delta rft1$ 细胞中的糖基化的能力。

[0895] 一种克隆包括了 flc2 基因的 3' 端截短翻译(fl c2')。Fl c2' 在酵母 1 号染色体上被编码。在 HCSS 筛选中鉴定的截短翻译包括全长基因的碱基 43309 至 44631,所述全长基因包括其天然启动子。Fl c2' 表达质粒(YEp352Fl c2')的序列示于图 13(SEQ ID NO:33)中,或 Fl c2' 的编码序列描述于图 5A 中。Fl c2' 编码包含 4 个完整的和第五截短跨膜域的 452 个氨基酸的蛋白质。从氨基酸 442 至 452 的 C-末端 11 个氨基酸来源于克隆程序(图 5B)。所述 fl c2' 基因序列及其启动子表示于图 5(图 5L)中。

[0896] 1.10 突变宿主细胞

[0897] 对携带 Rft1 或 Fl c2' 表达质粒的 $\Delta rft1$, $\Delta alg11$ 或 alg2-1 突变菌株进行斑点分析。使细胞在 YPD 平板上形成斑点。将所述平板分别于 37、30 或 31.5°C 孵化 3 天。Fl c2' 的超表达导致 $\Delta rft1$ 或 $\Delta alg11$ 菌株生长改善,显示出与表达 Rft1 的突变菌株相同或相似的生长表型(图 6A 及 6B)。Fl c2' 的超表达还导致 alg2-1 菌株生长改善,而 Rft1 的超表达未导致生长改善(图 6C)。

[0898] 所述 $\Delta alg3 \Delta alg11$ 菌株显示出高温敏感性表型和生长缺陷。这些缺陷可以被 Fl c2' 的表达强烈削弱。Fl c2' 的表达强烈改善菌株的生长行为并降低温度敏感性(图 18B)。

[0899] 此外,对携带表达质粒的 $\Delta rft1$ 突变菌株进行了斑点分析,所述质粒编码 Fl c2' 的跨膜域 3(SEQ ID NO:16)或跨膜域 3 和 4(SEQ ID NO:10)。如上所述细胞被形成斑点并于 37°C 孵化 3 天。Fl c2' 的跨膜域 1-3 或 Fl c2' 的跨膜域 3-4 的超表达导致生长改善,而

表达全长 Flc2 的细胞未表现出生长改善 (图 7A 和 7B)。

[0900] 此外,检测了 Flc2' 在 $\Delta rft1$ 菌株中恢复糖基化缺陷的能力。在 SD ura 介质 (缺少尿嘧啶的合成葡萄糖培养基) 中培育野生型菌株和 $\Delta rft1$ 菌株,所述 $\Delta rft1$ 菌株携带空质粒 (YEp352),或携带用于 Rft1 和 Flc2' 的超表达的质粒。在 SDS-PAGE 凝胶分离并通过使用抗 CPY 抗体的免疫印迹法总可溶蛋白质。Flc2' 的超表达恢复羧肽酶 CPY 在 $\Delta rft1$ 菌株中的 N-糖基化至与免疫印迹所揭示的 Rft1 的超表达中所观察到的水平相似的水平 (图 7C)。

[0901] 为了研究 Flc2' 对脂质连接寡糖合成的影响,对携带 Flc2' 表达构造 (图 8C) 的 $\Delta rft1$ 细胞进行了 3H-甘露糖标记。使用携带空载体 YEp352 (图 8A) 的 $\Delta rft1$ 细胞和携带 Rft1 表达构造 (图 8B) 的 $\Delta rft1$ 细胞作为对照。首先用 [3H]-甘露糖标记细胞。通过酸水解从脂质载体中释放出寡糖,并使用 HPLC 纯化和分析所述寡糖。[3H]-甘露糖标记的脂质连接寡糖的 HPLC 图显示,在缺乏功能性翻转酶的情况下,细胞累积 Man5GlcNAc2 (图 8A)。这表明脂质连接寡糖合成在内质网胞质面上由 Alg11p 催化的步骤之后停止,由于没有能翻转 Man5GlcNAc2 进入内质网腔的分子存在。在质粒上提供 rft1 恢复了脂质连接寡糖合成并导致 Glc3Man9GlcNAc2 的累积 (图 8B)。在 $\Delta rft1$ 细胞中 Flc2' 的表达基础上,翻转被恢复,并且除了 Man5GlcNAc2, Glc3Man9GlcNAc2 也在细胞中被累积 (图 8C)。总的说来,该数据表明 Flc2' 在 $\Delta rft1$ 酵母细胞中是作为翻转酶起作用的。

[0902] $\Delta rft1$ 突变菌株中 Flc2' 和 / 或 Rft1 的表达提高了 48h 后培养物的最终的细胞密度,达到相对于对照菌株细胞密度高大约 3 倍的近似程度 (表 6)。与 Flc2' 相反,全长 Flc2 的超表达为补偿在 $\Delta rft1$ 突变菌株中的翻转酶敲除:可以看出相对于 $\Delta rft1$ 菌株的对照,无生长改善。内源性全长 Flc2 不能补偿 $\Delta rft1$ 菌株的生长缺陷。

[0903] Rft1 或 Flc2' 的表达改善了生长并导致 $\Delta alg11$ (图 19A) 和 $\Delta alg3 \Delta alg11$ (图 18A) 突变菌株 48h 后的最终光学细胞密度高于仅携带空质粒的相应的对照菌株。在 $\Delta alg11$ 菌株中,Flc2* 的表达相对于载体对照改善了生长 33%,Rft1 的超表达导致最终的细胞密度增长 49%。在 $\Delta alg3 \Delta alg11$ 突变菌株中,Flc2* 的表达相对于载体对照改善了生长 54%,Rft1 的超表达导致最终的细胞密度增长 74% (表 6)。

[0904] 表 6 概括了超表达 Rft1, Flc2', 全长 Flc2 或携带空载体 (对照) 的酵母菌株的生长分析的结果 (n. d. = 未测出 / 检测)。

[0905] 表 6

[0906]		突变菌株		
	质粒	$\Delta rft1$	$\Delta alg11$	$\Delta alg11 \Delta alg3$
	空载体	3.75	2.44	1.49
	F1c2*	11.70	3.63	2.30
	Rft1	10.20	3.24	2.59
	F1c2	2.60	n. d.	n. d.

[0907] 1.11 Man3GlcNAc2 结构的翻转和转移

[0908] 在 $\Delta alg3 \Delta alg11$ 菌株中分析了 Flc2' 的超表达对羧肽酶 Y (CPY) 的 N-糖基化效率的影响。在 SD ura 介质中培育野生型菌株和 $\Delta alg3 \Delta alg11$ 菌株,所述 $\Delta alg3 \Delta alg11$

菌株携带空质粒 (YEp352), 或携带用于 Flc2' 或 Rft1 的超表达的质粒。在 SDS-PAGE 凝胶分离并通过使用抗 CPY 抗体的免疫印迹法总可溶蛋白质 (图 9)。在野生型细胞中, CPY 不依赖于 Rft1 或 Flc2' 的超表达而被完全地糖基化。但是, 通过 CPY 向高分子量的变化揭示了 $\Delta alg3 \Delta alg11$ 菌株中 Flc2' 或 Rft1 的表达改善了 CPY 的糖基化 (图 9)。

[0909] 1.12 alg2-1 菌株中翻转酶的特异性分析

[0910] 为了确认 Flc2' 对短脂质连接寡糖的和特异性, 选择了一种携带温度敏感性 Alg2 蛋白的酵母菌株。由于较低的 Alg2, 所述菌株主要累积 Man1GlcNAc2 (M1) 和 Man2GlcNAc2 (M2) 结构。但是, 残余的酶导致常规的酵母脂质连接寡糖 Glc3Man9GlcNAc2 的产生。如果 M1 或 M2 被翻转入内质网腔, 这两类脂质连接寡糖不是 Alg 途径中涉及的腔内甘露糖基转移酶的底物。M1 或 M2, 还有 Glc3Man9GlcNAc2 被转移至蛋白质上。所述 Glc3Man9GlcNAc2 结构在内质网及高尔基体中被进一步加工, 而形成包含 8 至 14 个甘露糖残基的蛋白连接寡糖 (NLO) 类。用 Flc2' 和 Rft1 表达载体及空载体对照来转化 Alg2-1 菌株。培育所述菌株至 A600 为 1 并收获所述细胞。分离、减少、烷基化细胞壁蛋白质, 并使用 PNGase F 释放 N- 多糖。纯化、全甲基化 N- 多糖, 并通过 MALDI-TOF MS 在 m/z 为 700 至 4000 范围内分析 N- 多糖。

[0911] 在 MALDI-TOF 谱图中检测到 M1, M2 的预期大小的峰及高甘露糖结构 Man8GlcNAc2 至 Man14GlcNAc2 (M8 至 M14)。在 NLO 类的峰强度的基础上, 计算每个结构的相对丰度。M1 或 M2 类中相对增加表明这些结构的翻转优势于由 Alg2 催化的延长。Flc2' 的表达导致 M1 结构累积了 88.5%。相反, 在表达 Rft1 或携带空载体的 Alg2-1 菌株中, M1 结构对总的 N- 多糖仅贡献了 74.7% 和 78.7% (表 7)。

[0912] 表 7 概括了 N- 多糖在超表达 Rft1 或 Flc2* 或携带空载体的 Alg2-1 菌株中的相对丰度 (%)。

[0913] 表 7

[0914]		突变菌株		
	N-多糖种类	空载体	<i>oeRft1</i>	<i>oeFlc2'</i>
	M1	78.7	74.7	88.5
	M2	19.1	21.7	10.9
	M8 至 M14	2.1	3.5	0.6

[0915] 1.13 $\Delta alg11$ 菌株中翻转酶的特异性分析

[0916] 为了确认 Flc2' 对 Man3GlcNAc2 (M3) 结构的和特异性, 选择了一种 $\Delta alg11$ 酵母菌株。该菌株的使用允许确定脂质连接寡糖结构在内质网膜的胞质面和腔面的相对丰度。由于 alg11 基因的失活, 胞质面上脂质连接寡糖的合成仅进行到 M3 的水平。所述结构如果被翻转入内质网腔, 就会被 Alg3 进一步修饰, 并且随后的甘露糖基转移酶导致 M7 的产生。采用 3H- 甘露糖标记细胞允许使用 HPLC 对不同脂质连接寡糖类的相对丰度进行定量。如果翻转是无效的, 则胞质脂质连接寡糖类累积到内质网膜的胞质面上, 与腔内脂质连接寡糖减少的相对量正好相反。

[0917] Flc2' 和 Rft1 在 $\Delta alg11$ 菌株中的表达降低了胞质脂质连接寡糖类对脂质连接寡糖总量的相对贡献 (图 17A, 17B, 17C), 因此增加腔内脂质连接寡糖类从对照菌株中的大约

43%至超表达 Flc2' 或 Rft1 的两种菌株中的大约 70% (表 8)。

[0918] 表 8 概括了不同脂质连接寡糖类在超表达 Rft1 或 Flc2* 或携带空载体的 $\Delta alg11$ 菌株中的相对丰度 (%)。脂质连接寡糖被指定到胞质组或腔内组。

[0919] 表 8

[0920]		突变菌株		
	脂质连接寡糖类	空载体	<i>oeRft1</i>	<i>oeFlc2'</i>
	胞质脂质连接寡糖	43.5	28.5	31.0
	腔内脂质连接寡糖	56.5	71.5	69.0

[0921] 1.14 $\Delta alg3 \Delta alg11 \Delta mnn1$ 敲除菌株的产生

[0922] 用 $\Delta alg3$ 缺失菌株与 $\Delta mnn1$ 缺失菌株进行杂交。使二倍体杂交 $\Delta alg3 \Delta alg11$ 菌株形成孢子并检测单倍体孢子的 $\Delta alg3$ 和 $mnn1$ 基因是否缺乏。通过 PCR 分析检测双敲除菌株的 $\Delta alg3$ 和 $mnn1$ 基因是否缺乏。然后,进一步用 $\Delta alg3 \Delta alg11$ 菌株与选择的 $\Delta alg3 \Delta mnn1$ 菌株进行杂交,使得到的菌株形成孢子,并分析四分体是否为缺乏 $alg3$, $alg11$ 和 $mnn1$ 基因的菌株。

[0923] 分析双重突变和三重突变的糖形。用 PNGase F 从细胞壁蛋白质中释放 N-多糖,用 2-AB 对 N-多糖进行标记并采用 MALDI-TOF MS 进行分析。来源于 $\Delta alg3 \Delta alg11$ 和三重突变菌株的 N-多糖谱的比较揭示了在 $m/z = 1377$ 处代表 M5 结构的峰减少。这些数据表明通过去除 $mnn1$ 基因,可以终止 NLO 在高尔基体中的修饰。图 22 描述了从来源于 $\Delta alg3 \Delta alg11$ 酵母突变菌株 (图 22A) 及 $\Delta alg11 \Delta alg3 \Delta mnn1$ 酵母突变菌株 (图 22B) 的细胞壁蛋白质中分离的 2-AB- 标记 N-多糖的 MALDI-TOF MS 谱。

[0924] 实施例 2 用于糖基化的复合系统

[0925] 2.1 新的脂质连接寡糖和原生动物寡糖基转移酶在酵母突变菌株中的表达

[0926] 在一个优选实施方案中,提供了一种用于蛋白质糖基化,尤其是在酵母中糖基化的复合系统,包括至少 3 个实体:(i) 作为寡糖基转移酶前体的脂质连接 Man3GlcNAc2 的产生;(ii) 一种翻转酶例如 (Flc2');及 (iii) 显示出松弛底物特异性的原生动物寡糖基转移酶 (POT)。

[0927] 为了将翻转酶和原生动物寡糖基转移酶这两种异源蛋白结合,构建了包括这两部分的载体。

[0928] 结果,在 GPD 启动子和 *cyc1* 终止子调控下的原生动物寡糖基转移酶 (LmStt3D) 以反转录基因的形式被插入到包括 Flc2' 的载体中。携带 LmStt3D, Flc2' 或携带这两种酶的质粒被转化入野生型酵母 (YG1509) 或缺乏 $alg11$ (YG1365) 或缺乏 $alg11$ 和 $alg3$ (YG1363) 的酵母细胞中,并用蛋白免疫印迹法分析 CPY 和 Gas1p 的 N-糖基化 (图 10)。

[0929] 在未缺失定位于内质网的寡糖基转移酶的对照菌株中,CPY 流动性在 Flc2' 或 LmStt3D, 或 Flc2' 和 LmStt3D 两者的表达上是相同的。在缺乏 $alg11$ 并制造脂质连接 GlcNAc2Man5 的酵母菌株 YG1365 或缺乏 $alg11$ 和 $alg3$ 并制造脂质连接 GlcNAc2Man3 的酵母菌株 YG1363 中, Flc2' 和 LmStt3D 的共表达使 CPY 向相对于单独表达 Flc2' 或 Rft1 的细胞较高的分子量变化,表明在 Flc2' 和 LmStt3D 存在下,CPY 更完整的 N-糖基化。在 β -1, 3- 葡聚糖转移酶 (Gas1p) 观察到流动性的相似变化。该 GPI 锚定蛋白定位在细胞壁上,并

且也经历发生在高尔基体中的修饰。

[0930] 在 $\Delta alg11$ 菌株中分析了 Flc2' 和 LmStt3D 的超表达对羧肽酶 Y (CPY) 的 N-糖基化效率的影响, 所述 $\Delta alg11$ 菌株携带空质粒 (YEp352), 或携带用于 Flc2', 或 LmStt3D, 或 Flc2' 和 LmStt3D 的超表达的质粒。在 SD ura 介质中于 23°C 培育所述菌株。在 SDS-PAGE 凝胶分离并通过使用抗 CPY 抗体的免疫印迹法总可溶蛋白质 (图 11)。在超表达 Flc2' 和 LmStt3D 的 $\Delta alg11$ 细胞中, CPY 被完全地糖基化 (mCPY), 然而, 仅超表达 Flc2' 或 POT LmStt3D 的细胞与载体对照相比, 减少 CPY 的低糖基化, 但未达到与 Flc2' 和 POT 共表达相同的程度 (图 11)。

[0931] 在被示意性地示于图 12 中的复合系统中, $alg3$ 及 $alg11$ 两种基因被缺失致脂质连接 Man3GlcNAc2 的产生。剩余的转移酶仍然存在于细胞中, 但是, 对脂质连接 Man3GlcNAc2 底物是无的。在第一个方法中, 添加了新的翻转酶 (例如 Flc2')。第二, 添加了原生动寡糖基转移酶 (POT, 例如 *Leishmania major* Stt3D)。替代脂质连接 Man3GlcNAc2 的产生的方案可以是缺失 $dpm1$ 基因, 所述基因产物在内质网膜的胞质面制造脂质连接甘露糖, 或者缺失翻转多萜醇连接寡糖进入内质网腔的单糖翻转酶。脂质连接甘露糖为定位于内质网腔的寡糖基转移酶提供供体。通过与 $alg11$ 突变结合, 这样一种细胞还能制造脂质连接 Man3GlcNAc2。多余未用的转移酶、翻转酶 (Rft1)、酵母 Ost 复合物的组分及未合成的结构以灰体描述。

[0932] 2.2 原生动寡糖基转移酶在酵母突变菌株中的表达

[0933] 提供了一种用于蛋白质糖基化, 尤其是在酵母中糖基化的复合系统, 包括至少 2 个实体: (i) 作为寡糖基转移酶前体的脂质连接 Man3GlcNAc2 的产生; (ii) 显示出松弛底物特异性的原生动寡糖基转移酶 (POT) 的一种或多种旁系同源基因的表达。

[0934] 构建了包括原生动寡糖基转移酶的载体。L. major 具有 LmStt3A 至 LmStt3D 四个 Stt3 旁系同源基因; L. braziliensis and L. infantum 每种具有分别命名为 Lb3_1 至 Lb3_3, 及 Li3_1 至 Li3_3 的三种不同的 Stt3 旁系同源基因。低拷贝数质粒和高拷贝数质粒上均包括所有单个的原生动寡糖基转移酶基因。此外, 高拷贝数质粒上还包括 Trypanosoma brucei 的旁系同源基因 TbStt3_B 及 TbStt3_C 的原生动寡糖基转移酶基因。

[0935] 每个原生动寡糖基转移酶旁系同源基因在被修饰的 $\Delta alg11$ 突变酵母菌株及 $\Delta alg3 \Delta alg11$ 突变酵母菌株中被表达, 所述菌株导入了原生动寡糖基转移酶质粒。制备所有菌株的细胞提取物, 并采用 CPY 特殊抗体对其进行分析。N-糖基化效率结果的比较揭示每个原生动寡糖基转移酶的效果在不同的突变菌株中有所区别, 表明不同的原生动寡糖基转移酶对脂质连接寡糖底物具有不同的偏爱。原生动寡糖基转移酶从低拷贝数质粒的表达在改善 N-糖基化方面比从高拷贝数质粒的表达更有效, 表明适当的表达水平是关键并且可以被优化。

[0936] 为建立 N-糖基化计分, 分析了一组 Western CPY 印迹 ($n = 2$ to 5), 将 N-糖基化效率与未修饰的 $\Delta alg11$ 及 $\Delta alg3 \Delta alg11$ 参照相比较, 计为 0 (无附加影响) 至 3 (高附加影响)。通过合计单个实验的点并划分重复的总沟数计算 N-糖基化计分。结果概括于表 9 中。

[0937] 表 9

[0938]

POT 质粒	糖基化计分	
低拷贝质粒	$\Delta alg11$	$\Delta alg3alg11$
LmStt3D	2.25	1.33
LbStt3-1	0	1
LbStt3-2	0	0
LbStt3-3	3	2.2
LiStt3-1	0	1
LiStt3-2	2.5	1
LiStt3-3	0	0
高拷贝质粒	$\Delta alg11$	$\Delta_{alg11}^{alg3} \Delta$
LmStt3D	2	1.75
LbStt3-1	1	1
LbStt3-2	0	0
LbStt3-3	1.25	1
LiStt3-1	0	0.5
LiStt3-2	0	1
LiStt3-3	0	0
Tb3-B	0.5	1
Tb3-C	0	1

[0001]

SEQUENCE LISTING (序列表)

<110> Lonza Ltd. (隆萨股份公司)

<120> novel tools for the production of glycosylated proteins in host cells (在宿主细胞中制造糖基化蛋白的新方法)

<130> 202148

<160> 61

<170> PatentIn version 3.4

<210> 1

<211> 1359

<212> DNA

<213> *Saccharomyces cerevisiae*

<400> 1

```

atgatcttcc taaacacctt cgcaagggtgc cttttaacgt gtttcgtact gtgcagcggt      60
acagcacggt cctctgacac aaacgacact actccggcgt ctgcaaagca tttgcagacc      120
acttctttat tgacgtgtat ggacaattcg caattaacgg catcattctt tgatgtgaaa      180
ttttaccccg ataalaatac tgttatcttt gatattgacg ctacgacgac gcttaatggg      240
aacgtcactg tgaaggctga gctgcttact tacggactga aagtcctgga taagactttt      300
gatttatgtt ccttgggccg agtalcgctt tccccctaa gtgctgggcg tattgatgtc      360
atgtccacac aggtgatcga atcatccatt accaagcaat ttcccgcat tgcttacacc      420
attccagatt tggacgcaca agtacgtgtg gtggcatacg ctcagaatga cacggaattc      480
gaaactccgc tggcttgtgt ccaggctatc ttgagtaacg ggaagacagt gcaaacaag      540
tatggcgctt ggccatttgc cgctatctca ggtgtcggtg tactlaccic agggtttgtg      600
tctgtgatcg gttactcagc cactgctgct cacattgcgt ccaactccat ctcatgttc      660
atatacttcc aaaatctagc tatcactgca atgatgggtg tctcaagggt tccacccatt      720
gctgccgcgt ggacgcagaa ttccaatgg tccatgggta tcatcaatac aaacttcatg      780
caaaagattt ttgattggta cgtacaggcc actaatgggt tctcaaatgt tgtggtagct      840
aacaaggacg tcttgtccat tagtgtgcaa aaacgtgcta tctctatggt atcgtctagt      900
gattacaatt ttgacaccat tttagacgat tcggatctgt acaccacttc tgagaaggat      960
ccaagcaatt actcagccaa gattcicgtg ttaagaggta tagaaagagt tgcttatttg     1020
gctaataatt agctatctaa ttcttttttg accggtattg tgttttttct attcttccta     1080
tttgtagttg tcgtctcttt gattttcttt aaggcgcctat tggaagtctt tacaagagca     1140
agaatattga aagagacttc caatttcttc caatatagga agaactgggg gagtattatc     1200
aaaggcaccc ttttcagatt atctatcacc gccttccttc aagtttctct tctggcgatt     1260
tggaatttta ctcagggtcaa ctctccagcg attgtgttg atgcggtagt aatattactg     1320
atcgatcctc tagagtcgac ctgcaggcat gcaagctag                                1359

```

<210> 2

<211> 452

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 2

```

Met Ile Phe Leu Asn Thr Phe Ala Arg Cys Leu Leu Thr Cys Phe Val
 1           5           10           15

```

```

Leu Cys Ser Gly Thr Ala Arg Ser Ser Asp Thr Asn Asp Thr Thr Pro
          20           25           30

```

[0002]

Ala Ser Ala Lys His Leu Gln Thr Thr Ser Leu Leu Thr Cys Met Asp
 35 40 45
 Asn Ser Gln Leu Thr Ala Ser Phe Phe Asp Val Lys Phe Tyr Pro Asp
 50 55 60
 Asn Asn Thr Val Ile Phe Asp Ile Asp Ala Thr Thr Thr Leu Asn Gly
 65 70 75 80
 Asn Val Thr Val Lys Ala Glu Leu Leu Thr Tyr Gly Leu Lys Val Leu
 85 90 95
 Asp Lys Thr Phe Asp Leu Cys Ser Leu Gly Gln Val Ser Leu Ser Pro
 100 105 110
 Leu Ser Ala Gly Arg Ile Asp Val Met Ser Thr Gln Val Ile Glu Ser
 115 120 125
 Ser Ile Thr Lys Gln Phe Pro Gly Ile Ala Tyr Thr Ile Pro Asp Leu
 130 135 140
 Asp Ala Gln Val Arg Val Val Ala Tyr Ala Gln Asn Asp Thr Glu Phe
 145 150 155 160
 Glu Thr Pro Leu Ala Cys Val Gln Ala Ile Leu Ser Asn Gly Lys Thr
 165 170 175
 Val Gln Thr Lys Tyr Ala Ala Trp Pro Ile Ala Ala Ile Ser Gly Val
 180 185 190
 Gly Val Leu Thr Ser Gly Phe Val Ser Val Ile Gly Tyr Ser Ala Thr
 195 200 205
 Ala Ala His Ile Ala Ser Asn Ser Ile Ser Leu Phe Ile Tyr Phe Gln
 210 215 220
 Asn Leu Ala Ile Thr Ala Met Met Gly Val Ser Arg Val Pro Pro Ile
 225 230 235 240
 Ala Ala Ala Trp Thr Gln Asn Phe Gln Trp Ser Met Gly Ile Ile Asn
 245 250 255
 Thr Asn Phe Met Gln Lys Ile Phe Asp Trp Tyr Val Gln Ala Thr Asn
 260 265 270
 Gly Val Ser Asn Val Val Val Ala Asn Lys Asp Val Leu Ser Ile Ser
 275 280 285
 Val Gln Lys Arg Ala Ile Ser Met Ala Ser Ser Ser Asp Tyr Asn Phe
 290 295 300
 Asp Thr Ile Leu Asp Asp Ser Asp Leu Tyr Thr Thr Ser Glu Lys Asp
 305 310 315 320
 Pro Ser Asn Tyr Ser Ala Lys Ile Leu Val Leu Arg Gly Ile Glu Arg
 325 330 335
 Val Ala Tyr Leu Ala Asn Ile Glu Leu Ser Asn Phe Phe Leu Thr Gly
 340 345 350

[0003]

Ile Val Phe Phe Leu Phe Phe Leu Phe Val Val Val Val Ser Leu Ile
 355 360 365
 Phe Phe Lys Ala Leu Leu Glu Val Leu Thr Arg Ala Arg Ile Leu Lys
 370 375 380
 Glu Thr Ser Asn Phe Phe Gln Tyr Arg Lys Asn Trp Gly Ser Ile Ile
 385 390 395 400
 Lys Gly Thr Leu Phe Arg Leu Ser Ile Ile Ala Phe Pro Gln Val Ser
 405 410 415
 Leu Leu Ala Ile Trp Glu Phe Thr Gln Val Asn Ser Pro Ala Ile Val
 420 425 430
 Val Asp Ala Val Val Ile Leu Leu Ile Asp Pro Leu Glu Ser Thr Cys
 435 440 445
 Arg His Ala Ser
 450
 <210> 3
 <211> 1116
 <212> DNA
 <213> *Saccharomyces cerevisiae*
 <400> 3
 atgattctcc taaacacctt cgcaagggtc cttttaacgt gtttcgtact gtgcagcggg 60
 acagcaggtt cctctgacac aaacgacact actccggcgt ctgcaaagca ttgtcagacc 120
 acttctttat tgacgtgtat ggacaattcg caattaacgg catcattctt tgatgtgaaa 180
 ttttaccocg alaataatac tgttatcttt gatattgacg ctacgacgac gcttaatggg 240
 aacgtcactg tgaaggctga gctgcttact tacggactga aagtcctgga taagactttt 300
 gatttatgtt ccttgggcca agtatcgctt tccccctaa gtgctgggcg taltgatgtc 360
 atgtccacac aggtgatcga atcatccatt accaagcaat ttcccgcat tgcttacacc 420
 attccagatt tggacgcaca agtacgtgtg gtggcatacg ctacagaatga cacggaattc 480
 gaaactccgc tggcttgtgt ccaggctatc ttgagtaacg ggaagacagt gcaaacaaag 540
 tatgcggcct ggccatttgc cgctatctca ggtgtcgggt tacttacctc aggglltltg 600
 tctgtgatcg gttactcagc cactgctgct cacatlgcgt ccaactccat ctcatgttc 660
 atatacttcc aaaatctagc tatcactgca atgatgggtg tctcaagggt tccaccatt 720
 gctgccgctt ggacgcagaa ttccaatgg tccatgggta tcatcaatac aaacttcatg 780
 caaaagattt ttgatttgta cgtacaggcc actaatgggt tctcaaatgt tgtggtagct 840
 aacaaggacg tcttgtccat tagtgtgcaa aaacgtgcta tctctatggc atcgtctagt 900
 gattacaatt ttgacacat tttagacgat tcggatctgt acaccacttc tgagaaggat 960
 ccaagcaatt actcagccaa gattctcgtg ttaagaggta tagaaagagt tgcttatttg 1020
 gctaattatt agctatctaa tttctttttg accggtattg tgttttllct attcttccta 1080
 tttgtagttg tcgtctcttt gattttctll aagtag 1116
 <210> 4
 <211> 371
 <212> PRT
 <213> *Saccharomyces cerevisiae*
 <400> 4

[0004]

Met Ile Phe Leu Asn Thr Phe Ala Arg Cys Leu Leu Thr Cys Phe Val
 1 5 10 15
 Leu Cys Ser Gly Thr Ala Arg Ser Ser Asp Thr Asn Asp Thr Thr Pro
 20 25 30
 Ala Ser Ala Lys His Leu Gln Thr Thr Ser Leu Leu Thr Cys Met Asp
 35 40 45
 Asn Ser Gln Leu Thr Ala Ser Phe Phe Asp Val Lys Phe Tyr Pro Asp
 50 55 60
 Asn Asn Thr Val Ile Phe Asp Ile Asp Ala Thr Thr Thr Leu Asn Gly
 65 70 75 80
 Asn Val Thr Val Lys Ala Glu Leu Leu Thr Tyr Gly Leu Lys Val Leu
 85 90 95
 Asp Lys Thr Phe Asp Leu Cys Ser Leu Gly Gln Val Ser Leu Ser Pro
 100 105 110
 Leu Ser Ala Gly Arg Ile Asp Val Met Ser Thr Gln Val Ile Glu Ser
 115 120 125
 Ser Ile Thr Lys Gln Phe Pro Gly Ile Ala Tyr Thr Ile Pro Asp Leu
 130 135 140
 Asp Ala Gln Val Arg Val Val Ala Tyr Ala Gln Asn Asp Thr Glu Phe
 145 150 155 160
 Glu Thr Pro Leu Ala Cys Val Gln Ala Ile Leu Ser Asn Gly Lys Thr
 165 170 175
 Val Gln Thr Lys Tyr Ala Ala Trp Pro Ile Ala Ala Ile Ser Gly Val
 180 185 190
 Gly Val Leu Thr Ser Gly Phe Val Ser Val Ile Gly Tyr Ser Ala Thr
 195 200 205
 Ala Ala His Ile Ala Ser Asn Ser Ile Ser Leu Phe Ile Tyr Phe Gln
 210 215 220
 Asn Leu Ala Ile Thr Ala Met Met Gly Val Ser Arg Val Pro Pro Ile
 225 230 235 240
 Ala Ala Ala Trp Thr Gln Asn Phe Gln Trp Ser Met Gly Ile Ile Asn
 245 250 255
 Thr Asn Phe Met Gln Lys Ile Phe Asp Trp Tyr Val Gln Ala Thr Asn
 260 265 270
 Gly Val Ser Asn Val Val Val Ala Asn Lys Asp Val Leu Ser Ile Ser
 275 280 285
 Val Gln Lys Arg Ala Ile Ser Met Ala Ser Ser Ser Asp Tyr Asn Phe
 290 295 300
 Asp Thr Ile Leu Asp Asp Ser Asp Leu Tyr Thr Thr Ser Glu Lys Asp
 305 310 315 320

[0005]

Pro Ser Asn Tyr Ser Ala Lys Ile Leu Val Leu Arg Gly Ile Glu Arg
325 330 335

Val Ala Tyr Leu Ala Asn Ile Glu Leu Ser Asn Phe Phe Leu Thr Gly
340 345 350

Ile Val Phe Phe Leu Phe Phe Leu Phe Val Val Val Val Ser Leu Ile
355 360 365

Phe Phe Lys
370

<210> 5
<211> 737
<212> DNA
<213> *Saccharomyces cerevisiae*

<400> 5
atgatcttcc taaacacctt cgcaaggtgc cttttaacgt gtttcgtact gtgcagcggt 60
acagcacgtt cctctgacac aaacgacact actccggcgt ctgcaaagca ttgcagacc 120
acttctttat tgacgtgtat ggacaattcg caattaacgg catcattctt tgatgtgaaa 180
ttttaccccg ataataatac tgttatcttt gatattgacg ctacgacgac gcttaatggg 240
aagtcacttg tgaaggtga gctgcttact tacggactga aagtcctgga taagactttt 300
gatttatgtt ccttgggcca agtatcgctt tccccctaa gtgctgggag tattgatgtc 360
atgtccacac aggtgatcga atcatccatt accaagcaat ttcccgcat tgettacacc 420
attccagatt tggacgcaca agtacgtgtg gggcatacgc ctccagaatga caccgaattc 480
gaaactccgc tggcttgtgt ccaggctatc ttgagtaacg ggaagacagt gcaacaaag 540
tatgcggcct ggcccattgc cgtatctcga ggtgtcgggt tacttacctc agggtttgtg 600
tctgtgatcg gttactcagc cactgctgct cacattgcgt ccaactccat ctcatgttcc 660
atatacttcc aaaatctagc tatcactgca atgatgggtg tctcaagggt tccaccatt 720
gtgcccgcgt ggactag 737

<210> 6
<211> 245
<212> PRT
<213> *Saccharomyces cerevisiae*

<400> 6

Met Ile Phe Leu Asn Thr Phe Ala Arg Cys Leu Leu Thr Cys Phe Val
1 5 10 15

Leu Cys Ser Gly Thr Ala Arg Ser Ser Asp Thr Asn Asp Thr Thr Pro
20 25 30

Ala Ser Ala Lys His Leu Gln Thr Thr Ser Leu Leu Thr Cys Met Asp
35 40 45

Asn Ser Gln Leu Thr Ala Ser Phe Phe Asp Val Lys Phe Tyr Pro Asp
50 55 60

Asn Asn Thr Val Ile Phe Asp Ile Asp Ala Thr Thr Thr Leu Asn Gly
65 70 75 80

Asn Val Thr Val Lys Ala Glu Leu Leu Thr Tyr Gly Leu Lys Val Leu
85 90 95

[0006]

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

<210> 7
 <211> 813
 <212> DNA
 <213> *Saccharomyces cerevisiae*

<400> 7
 atgatcttcc taaacacctt cgcaagggtgc cttttaacgt gtttegtact gtgcagcggg 60
 acagcacggt cctctgacac aaacgacatt gcgtccaact ccattctcatt gttcatatac 120
 ttccaaaatc tagctatcac tgcaatgatg ggtgtctcaa ggggtccacc cattgtgtcc 180
 gcgtggacgc agaatttcca atggtccatg ggtatcatca atacaaactt catgcaaaag 240
 atttttgatt ggtacgtaca ggccactaat ggtgtctcaa atgttgtggt agctaacaag 300
 gacgtcttgi ccattagtgt gcaaaaacgt gctatctcta tggcatcgtc tagtgattac 360
 aattttgaca ccattttaga cgattcggat ctgtacacca cttctgagaa ggatccaagc 420
 aattactcag ccaagattct cgtgttaaga ggtatagaaa gagttgctta tttggctaatt 480
 attgagctat ctaatttctt tttagaccgt atttgttttt ttctattctt cctatttgta 540
 gttgtcgtct ctttgatttt ctllaaggcg ctattggaag ttcttacaag agcaagaata 600
 ttgaaagaga ctccaattt ctccaatat aggaagaact gggggagtat tatcaaaggc 660
 acccttttca gattatctat catcgcttc cctcaagttt ctcttctggc gatttgggaa 720
 ttactcagg tcaactctcc agcgattgtt gttgatgcgg tagtaatatt actgatcgat 780
 cctctagagt cgacctgcag gcatgcaagc tag 813

<210> 8
 <211> 270
 <212> PRT
 <213> *Saccharomyces cerevisiae*

[0007]

<400> 8

Met Ile Phe Leu Asn Thr Phe Ala Arg Cys Leu Leu Thr Cys Phe Val
1 5 10 15

Leu Cys Ser Gly Thr Ala Arg Ser Ser Asp Thr Asn Asp Ile Ala Ser
20 25 30

Asn Ser Ile Ser Leu Phe Ile Tyr Phe Gln Asn Leu Ala Ile Thr Ala
35 40 45

Met Met Gly Val Ser Arg Val Pro Pro Ile Ala Ala Ala Trp Thr Gln
50 55 60

Asn Phe Gln Trp Ser Met Gly Ile Ile Asn Thr Asn Phe Met Gln Lys
65 70 75 80

Ile Phe Asp Trp Tyr Val Gln Ala Thr Asn Gly Val Ser Asn Val Val
85 90 95

Val Ala Asn Lys Asp Val Leu Ser Ile Ser Val Gln Lys Arg Ala Ile
100 105 110

Ser Met Ala Ser Ser Ser Asp Tyr Asn Phe Asp Thr Ile Leu Asp Asp
115 120 125

Ser Asp Leu Tyr Thr Thr Ser Glu Lys Asp Pro Ser Asn Tyr Ser Ala
130 135 140

Lys Ile Leu Val Leu Arg Gly Ile Glu Arg Val Ala Tyr Leu Ala Asn
145 150 155 160

Ile Glu Leu Ser Asn Phe Phe Leu Thr Gly Ile Val Phe Phe Leu Phe
165 170 175

Phe Leu Phe Val Val Val Val Ser Leu Ile Phe Phe Lys Ala Leu Leu
180 185 190

Glu Val Leu Thr Arg Ala Arg Ile Leu Lys Glu Thr Ser Asn Phe Phe
195 200 205

Gln Tyr Arg Lys Asn Trp Gly Ser Ile Ile Lys Gly Thr Leu Phe Arg
210 215 220

Leu Ser Ile Ile Ala Phe Pro Gln Val Ser Leu Leu Ala Ile Trp Glu
225 230 235 240

Phe Thr Gln Val Asn Ser Pro Ala Ile Val Val Asp Ala Val Val Ile
245 250 255

Leu Leu Ile Asp Pro Leu Glu Ser Thr Cys Arg His Ala Ser
260 265 270

<210> 9

<211> 405

<212> DNA

<213> *Saccharomyces cerevisiae*

<400> 9

atgatcttcc taaacacctt cgcaagggtgc cttttaacgt gtttcgtact gtcacagcgg 60

acagcagcgtt cctctgacac aaacgacttc ttttgaccg gtattgtgtt tttctattc 120

[0008]

```

ttctatttg tagttgtcgt ctctttgatt ttctttaagg cgctattgga agttcttaca    180
agagcaagaa tattgaaaga gacttccaat ttcttccaat ataggaagaa ctgggggagt    240
attatcaaag gcaccctttt cagattatct atcatcgctt tccetcaagt ttctttctg    300
gcgatttggg aattttactca ggtcaactct ccagcgattg ttgttgatgc ggtagtaata    360
ttactgatcg atcctctaga gtcgacctgc aggcattgcaa gctag                    405

```

<210> 10
 <211> 134
 <212> PRT
 <213> *Saccharomyces cerevisiae*

<400> 10

```

Met Ile Phe Leu Asn Thr Phe Ala Arg Cys Leu Leu Thr Cys Phe Val
1           5           10           15

```

```

Leu Cys Ser Gly Thr Ala Arg Ser Ser Asp Thr Asn Asp Phe Phe Leu
                20                25                30

```

```

Thr Gly Ile Val Phe Phe Leu Phe Phe Leu Phe Val Val Val Val Ser
35                40                45

```

```

Leu Ile Phe Phe Lys Ala Leu Leu Glu Val Leu Thr Arg Ala Arg Ile
50                55                60

```

```

Leu Lys Glu Thr Ser Asn Phe Phe Gln Tyr Arg Lys Asn Trp Gly Ser
65                70                75                80

```

```

Ile Ile Lys Gly Thr Leu Phe Arg Leu Ser Ile Ile Ala Phe Pro Gln
85                90                95

```

```

Val Ser Leu Leu Ala Ile Trp Glu Phe Thr Gln Val Asn Ser Pro Ala
100                105                110

```

```

Ile Val Val Asp Ala Val Val Ile Leu Leu Ile Asp Pro Leu Glu Ser
115                120                125

```

```

Thr Cys Arg His Ala Ser
130

```

<210> 11
 <211> 621
 <212> DNA
 <213> *Saccharomyces cerevisiae*

```

<400> 11
atgatcttcc taaacacctt cgcaagggtc cttttaacgt gtttcgtact gtgcagcggg    60
acagcacggt cctctgacac aaacgacact actccggcgt ctgcaaagca ttgcagacc    120
acttctttat tgacgtgtat ggacaattcg caattaacgg catcaltcct tgatgtgaaa    180
ttttaccccg ataalaatac tgttatcttt gatattgacg ctacgacgac gcttaatggg    240
aacgtcactg tgaaggctga gctgcttact tacggactga aagtccttga taagactttt    300
gatttatgtt ccttgggcca agtatcgctt tccccctaa gtgctgggcg tattgatgtc    360
atgtccacac aggtgatcga atcatccatt accaagcaat ttcccggcat tgcttacacc    420
attccagatt tggacgcaca agtacgtgtg gtggcatacg ctccagaatga cacggaattc    480
gaaactccgc tggttgtgtg ccaggetatc ttgagtaacg ggaagacagt gcaacaaaag    540
tatgcggcct ggcccatgtc cgctatctca ggtgtcgggt tacttacctc agggtttgtg    600

```

[0009]

tctgtgatcg gttactcata g	621
<210> 12	
<211> 206	
<212> PRT	
<213> <i>Saccharomyces cerevisiae</i>	
<400> 12	
Met Ile Phe Leu Asn Thr Phe Ala Arg Cys Leu Leu Thr Cys Phe Val 1 5 10 15	
Leu Cys Ser Gly Thr Ala Arg Ser Ser Asp Thr Asn Asp Thr Thr Pro 20 25 30	
Ala Ser Ala Lys His Leu Gln Thr Thr Ser Leu Leu Thr Cys Met Asp 35 40 45	
Asn Ser Gln Leu Thr Ala Ser Phe Phe Asp Val Lys Phe Tyr Pro Asp 50 55 60	
Asn Asn Thr Val Ile Phe Asp Ile Asp Ala Thr Thr Thr Leu Asn Gly 65 70 75 80	
Asn Val Thr Val Lys Ala Glu Leu Leu Thr Tyr Gly Leu Lys Val Leu 85 90 95	
Asp Lys Thr Phe Asp Leu Cys Ser Leu Gly Gln Val Ser Leu Ser Pro 100 105 110	
Leu Ser Ala Gly Arg Ile Asp Val Met Ser Thr Gln Val Ile Glu Ser 115 120 125	
Ser Ile Thr Lys Gln Phe Pro Gly Ile Ala Tyr Thr Ile Pro Asp Leu 130 135 140	
Asp Ala Gln Val Arg Val Val Ala Tyr Ala Gln Asn Asp Thr Glu Phe 145 150 155 160	
Glu Thr Pro Leu Ala Cys Val Gln Ala Ile Leu Ser Asn Gly Lys Thr 165 170 175	
Val Gln Thr Lys Tyr Ala Ala Trp Pro Ile Ala Ala Ile Ser Gly Val 180 185 190	
Gly Val Leu Thr Ser Gly Phe Val Ser Val Ile Gly Tyr Ser 195 200 205	
<210> 13	
<211> 191	
<212> DNA	
<213> <i>Saccharomyces cerevisiae</i>	
<400> 13	
atgatcttcc taaacacctt cgcaagggtc cttttaacgt gtttcgtact gtgcagcggc	60
acagcacggt cctctgacac aaacgacatt gcgtccaact ccatttcatt gticatatat	120
ttccaaaatc tagctatcac tgcaatgatg ggtgtctcaa gggttccacc cattgctgcc	180
gcgtggacta g	191
<210> 14	
<211> 63	

[0010]

<212> PRT
 <213> *Saccharomyces cerevisiae*
 <400> 14
 Met Ile Phe Leu Asn Thr Phe Ala Arg Cys Leu Leu Thr Cys Phe Val
 1 5 10 15
 Leu Cys Ser Gly Thr Ala Arg Ser Ser Asp Thr Asn Asp Ile Ala Ser
 20 25 30
 Asn Ser Ile Ser Leu Phe Ile Tyr Phe Gln Asn Leu Ala Ile Thr Ala
 35 40 45
 Met Met Gly Val Ser Arg Val Pro Pro Ile Ala Ala Ala Trp Thr
 50 55 60
 <210> 15
 <211> 162
 <212> DNA
 <213> *Saccharomyces cerevisiae*
 <400> 15
 atgatcttcc taaacacctt cgcaaggtgc cttttaacgt gtttctgact gtgcagcggg 60
 acagcacggtt cctctgacac aaacgacttc tttttgacgg gtattgtgtt tttctattc 120
 ttcttatttg tagttgtcgt ctctttgatt ttctttaagt ag 162
 <210> 16
 <211> 53
 <212> PRT
 <213> *Saccharomyces cerevisiae*
 <400> 16
 Met Ile Phe Leu Asn Thr Phe Ala Arg Cys Leu Leu Thr Cys Phe Val
 1 5 10 15
 Leu Cys Ser Gly Thr Ala Arg Ser Ser Asp Thr Asn Asp Phe Phe Leu
 20 25 30
 Thr Gly Ile Val Phe Phe Leu Phe Phe Leu Phe Val Val Val Val Ser
 35 40 45
 Leu Ile Phe Phe Lys
 50
 <210> 17
 <211> 243
 <212> DNA
 <213> *Saccharomyces cerevisiae*
 <400> 17
 atgatcttcc taaacacctt cgcaaggtgc cttttaacgt gtttctgact gtgcagcggg 60
 acagcacggtt cctctgacac aaacgacggc acccttttca gattatctat catgccttc 120
 cctcaagttt ctctcttggc gatttgggaa ttactcagg tcaactctcc agcgattgtt 180
 gttgatgcgg tagtaatatl actgatcgat cctctagagt cgacctgcag gcatgcaagc 240
 tag 243
 <210> 18
 <211> 80
 <212> PRT
 <213> *Saccharomyces cerevisiae*
 <400> 18

[0011]

Met Ile Phe Leu Asn Thr Phe Ala Arg Cys Leu Leu Thr Cys Phe Val
1 5 10 15

Leu Cys Ser Gly Thr Ala Arg Ser Ser Asp Thr Asn Asp Gly Thr Leu
20 25 30

Phe Arg Leu Ser Ile Ile Ala Phe Pro Gln Val Ser Leu Leu Ala Ile
35 40 45

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

Val Ile Leu Leu Ile Asp Pro Leu Glu Ser Thr Cys Arg His Ala Ser
65 70 75 80

<210> 19
<211> 72
<212> DNA
<213> *Saccharomyces cerevisiae*

<400> 19
atgatcttcc taaacacctt cgcaagggtgc cttttaacgt gtttctgtact gtgcagcggt 60
acagcacggtt cc 72

<210> 20
<211> 24
<212> PRT
<213> *Saccharomyces cerevisiae*

<400> 20

Met Ile Phe Leu Asn Thr Phe Ala Arg Cys Leu Leu Thr Cys Phe Val
1 5 10 15

Leu Cys Ser Gly Thr Ala Arg Ser
20

<210> 21
<211> 69
<212> DNA
<213> *Saccharomyces cerevisiae*

<400> 21
gcctggccca ttgccgtat ctcagggtgc ggtgtactta cctcagggtt tgtgtctgtg 60
atcggttac 69

<210> 22
<211> 23
<212> PRT
<213> *Saccharomyces cerevisiae*

<400> 22

Ala Trp Pro Ile Ala Ala Ile Ser Gly Val Gly Val Leu Thr Ser Gly
1 5 10 15

Phe Val Ser Val Ile Gly Tyr
20

<210> 23
<211> 98
<212> DNA
<213> *Saccharomyces cerevisiae*

<400> 23
attgcgtcca actccatctc attgttcata tacttccaaa atctagctat cactgcaatg 60

[0012]

atgggtgtct caagggttcc acccattgct gccgcgtg 98

<210> 24
<211> 33
<212> PRT
<213> *Saccharomyces cerevisiae*

<400> 24
Ile Ala Ser Asn Ser Ile Ser Leu Phe Ile Tyr Phe Gln Asn Leu Ala
1 5 10 15

Ile Thr Ala Met Met Gly Val Ser Arg Val Pro Pro Ile Ala Ala Ala
20 25 30

Trp

<210> 25
<211> 69
<212> DNA
<213> *Saccharomyces cerevisiae*

<400> 25
ttctttttga cgggtattgt gttttttcta ttcttcttat ttgtagttgt cgtctctttg 60
attttcttt 69

<210> 26
<211> 23
<212> PRT
<213> *Saccharomyces cerevisiae*

<400> 26
Phe Phe Leu Thr Gly Ile Val Phe Phe Leu Phe Phe Leu Phe Val Val
1 5 10 15

Val Val Ser Leu Ile Phe Phe
20

<210> 27
<211> 60
<212> DNA
<213> *Saccharomyces cerevisiae*

<400> 27
ggcacccctt tcagattatc tatcatcgcc ttccctcaag ttctcttctt gccgatttgg 60

<210> 28
<211> 20
<212> PRT
<213> *Saccharomyces cerevisiae*

<400> 28
Gly Thr Leu Phe Arg Leu Ser Ile Ile Ala Phe Pro Gln Val Ser Leu
1 5 10 15

Leu Ala Ile Trp
20

<210> 29
<211> 17
<212> DNA
<213> *Saccharomyces cerevisiae*

<400> 29
gtagtaatat tactgat 17

[0013]

<210> 30
 <211> 6
 <212> PRT
 <213> *Saccharomyces cerevisiae*

<400> 30

Val Val Ile Leu Leu Ile
 1 5

<210> 31
 <211> 6941
 <212> DNA
 <213> *Saccharomyces cerevisiae*

<400> 31

```

atcattctgg acgtatgtgc acaigtgatt tgettttgtt tttttaagaa tgcgggtaa      60
taaacagatt gtttttctgg gaggataatc ttttctttt tcctgttggg attctaaaaat    120
taaccttgct gtttcttttt tttttttttt tcgcgcgact actcagccat cttgcatttt    180
taaagaaaaa gataatcatt aatgccttca cgggaatacg tatagaacat tattaaaagt    240
atatgaatgg catatatata tagaacacca cccttggaac acatttatac cccttaaaact    300
aaaacaattt gctgcgctat accgtgtttc agtgtattat aatacattca tttctgtttc    360
attacgatta tattgacgtg ataaaaagat tatatagcca tgatcttcct aaacaccttc    420
gcaagggtgcc tttaaacgtg lttcgtactg tgcagcggta cagcacgttc ctctgacaca    480
aacgacacta ctccggcgtc tgcaaaagcat ttgcagacca cttctttatt gacgtgtatg    540
gacaattcgc aattaacggc atcattcttt gatgtgaaat tttaccccca taataatact    600
gttatctttg atattgacgc tacgacgacg cttaatggga acgtcactgt gaaggctgag    660
ctgcttactt acggactgaa agtcctggat aagacttttg atttatgttc cttgggccaa    720
gtatcgcttt cccccctaag tgctgggcgt attgatgtca tglccacaca ggtgatcgaa    780
tcateccatta ccaagcaatt tcccggcatt gcttacacca ttccagattt ggacgcacaa    840
gtacgtgtgg tggcalacgc tcagaatgac acggaattcg aaactccgct ggccttgtgc    900
caggctatct ttagtaacgg gaagacagtg caaacaagat atgcggcctg gccattgcc     960
gctatctcag gtgtcgggtg acttacctca gggtttgtgt ctgtgatcgg ttactcagcc   1020
actgctgtct acattgcgtc caactccatc tcattgttca talacttcca aaatctagct   1080
atcaactgcaa tgatgggtgt ctcaagggtt ccaccatttg ctgccgctg gacgcagaat   1140
tlccaatggt ccatgggtat catcaatata aacttcatgc aaaagatttt tgatttgtac   1200
gtacaggcca ctaatgtgt ctcaaatgtt gtggtagcta acaaggacgt cttgtccatt   1260
agtgtgcaaa aacgtgctat ctctatggca tcgtctagtg attacaattt tgacaccatt   1320
ttagacgatt cggatctgta caccacttct gagaaggatc caagcaatta ctacgccaag   1380
attctcgtgt taagaggtat agaaagagtt gcttatttgg ctaatattga gctatctaatt  1440
ttctttttga ccggtattgt gttttttcta ttcttctat ttgtagtgt cgtctctttg   1500
attttcttta aggcgtattt ggaagttctt acaagagcaa gaatatlgaa agagacttcc   1560
aatttcttcc aatataggaa gaactggggg agtattatca aaggcaccct tttcagatta   1620
tctatcatcg ctttccctca agtttctctt ctggcgattt gggaatttac tcagggtcaac   1680
tctccacgca ttgttgttga tgcggtagta atattactga tcgatcctct agagtcgacc   1740
tgcaggcatg caagctagct tggcactggc cgtcgtttta caacgtcgtg actgggaaaa   1800
ccctggcggt acccaactta atcgccctgc agcacatccc cccttcgcca gctggcgtaa   1860

```

[0014]

tagcgaagag gcccgccacg atcgcccttc ccaacagttg cgcagcctga atggcgaatg	1920
gcgcctgatg cggatatttc tccttacgca tctgtgcggt atttcacacc gcatagggta	1980
ataactgata taattaaatt gaagctctaa ttgtgtgagt tagtatacat gcatttactt	2040
ataatacagt tttttagttt tgctggccgc atcttctcaa atatgcttcc cagcctgctt	2100
ttctgtaacg ttccacctct accttagcat ccttccctt tgcaaatagt cctcttccaa	2160
caataataat gtcagatcct gtagagacca catcatccac ggttctatac tgttgaccca	2220
atgcgtctcc ctgtcatct aaaccacac cgggtgtcat aatcaaccaa tcgtaacctt	2280
catctcttcc acctatgtct ctttgagcaa taaagccgat aacaaaatct ttgtcgtct	2340
tcgcaatgtc aacagtaccc ttagtatatt ctccagtaga tagggagccc ttgcatgaca	2400
attctgctaa catcaaaagg cctctagggt cctttgttac ttctctgccc gctgcttca	2460
aaccgctaac aatacctggg cccaccacac cgtgtgcatt cgtaatgctt gccattctg	2520
ctatctgta tacaccgcga gagtactgca atttgactgt attaccaatg tcagcaaat	2580
ttctgtcttc gaagagtaaa aaattgtact tggcggataa tgccttttagc ggcttaactg	2640
tgcctccat ggaaaaatca glcaagatat ccacatgtgt ttttagtaaa caaattttgg	2700
gacctaatgc ttcaactaac tccagtaatt ccttggtggt acgaacatcc aatgaagcac	2760
acaagtttgt ttgcttttcg tgcattgatat taaatagctt ggcagcaaca ggactaggat	2820
gagtagcagc acgttcccta tatgtagctt tcgacatgat ttatcttctg ttccggtttt	2880
gttctgtgca gttgggttaa gaatactggg caatttcatt ttcttcaac actacatatg	2940
cgtatatata ccaatctaag tctgtgctcc ttcttctgtt ctctctctg ttccggagatt	3000
accgaatcaa aaaaatttca aagaaccga aatcaaaaaa aagaataaaa aaaaaatgat	3060
gaattgaaaa gctcttgta cccatcattg aattttgaac atccgaacct gggagttttc	3120
cctgaacacg atagtatat tgaacctgta taataatata tagtctagcg ctttacggaa	3180
gacaatgtat gtatttcggt tcctggagaa actatlgcat ctattgcata ggtaatcttg	3240
cacgtcgcatt ccccggttca tttctgcgt ttccatcttg cacttcaata gcattatctt	3300
gttaacgaag catctgtgct tcatttgta gaacaaaaa gcaacgcgag agcgctaatt	3360
tttcaaaaa agaactcgag ctgcattttt acagaacaga aatgcaacgc gaaagcgcta	3420
ttttaccaac gaagaatctg tgcttcattt ttgtaaaaca aaaatgcaac gcgagagcgc	3480
taatttttca aacaaagaat ctgagctgca tttttacaga acagaaatgc aacgcgagag	3540
cgtattttta ccaacaaaga atctatactt cttttttgtt ctacaaaaat gcattcccgag	3600
agcgtatatt ttctaacaaa gcattcttaga ttactttttt tctcctttgt gcgctctata	3660
atgcagtctc ttgataactt ttgcaactgt aggtccgta aggttagaag aaggctactt	3720
tgggtctctat ttctcttcc ataaaaaaag cctgactcca ctcccgctt ttactgatta	3780
ctagcgaagc tgcgggtgca ttttttcaag ataaaggcat ccccgattat attctatacc	3840
gatgtggatt gcgcatactt lgtgaacaga aagtatagc gttgatgatt cttcattggt	3900
cagaaaaatta tgaacggtt ctctatctt gtctctatat actacgtata ggaatgttt	3960
acattttctg attgttttcg attcacctta tgaatagttc ttactacaat tttttgtct	4020
aaagagtaat actagagata aacataaaaa atgtagaggt cgagtttaga tgcaagtcca	4080
aggagcgaaa ggtggatggg taggttata agggatatag cacagagata tatagcaaag	4140
agatactttt gagcaatgtt tgtggaagcg gtattcgcaa tatttttagta gctcgttaca	4200
gtccggtgcg tttttggtt tttgaaagt cgtcttcaga gcgcttttgg ttttcaaaag	4260

[0015]

```

cgctctgaag ttctataact ttctagctag agaataaggaa cttcgggaata ggaacttcaa 4320
agcgtttccg aaaacgagcg ctccgaaaa tgcaacgcga gctgcgcaca tacagctcac 4380
tgttcacgtc gcacctatat ctgcgtgttg cctgtatata tatatacatg agaagaacgg 4440
catagtgcgt gtttatgctt aaatgcgta lgglgcactc tcagtacaat ctgctctgat 4500
gccgcatagt taagccagcc ccgacaccgc ccaacaccgc ctgacgcgcc ctgacgggct 4560
tgtctgtccc cgccatccgc ttacagacaa gctgtgaccg tctccgggag ctgcatgtgt 4620
cagaggtttt caccgtcatc accgaaacgc gcgagacgaa agggcctcgt gatacgcta 4680
ttttatagg ttaatgtcat gataataatg gtttcttaga cgtcaggtag cacttttcgg 4740
ggaaatgtgc gcggaacccc tattgttta ttttctaaa tacattcaaa tatgtatccg 4800
ctcatgagac aataaccctg ataaatgctt caataatatt gaaaaaggaa gagtatgagt 4860
attcaacatt tccgtgtcgc ccttattccc tttttgcgg cattttgcct tccgtgtttt 4920
gctcaccag aaacgttgtt gaaagtaaaa gatgctgaag atcagttggg tgcacgagtg 4980
ggttacatcg aactggatct caacagcggg aagatccttg agagttttcg ccccgaagaa 5040
cgttttccaa tgatgagcac ttttaaagtt ctgctatgtg gcgcggtatt atcccgatt 5100
gacgcggggc aagagcaact cggtcgccgc atacactatt ctcaaatga cttggttgag 5160
tactcaccag tcacagaaaa gcatcttacg gatggcalga cagtaagaga attatgcagt 5220
gtgccataa ccatgagtga taacactgcg gccaaactac ttctgacaac gatcggagga 5280
ccgaaggagc taaccgcttt ttgcacaac atgggggagc atgtaactcg cctlgatcgt 5340
tgggaacccg agctgaatga agccatacca aacgacgagc gtgacaccac gatgcctgta 5400
gcaatggcaa caacgttgcg caaactatta actggcgaac tacttactct agcttcccgg 5460
caacaattaa tagactggat ggaggcggat aaagtgcag gaccacttct gcgctcgcc 5520
cttccggtg gctggtttat tgctgataaa tctggagccg gtgagcgtgg gtctcgggt 5580
atcattgcag cactggggcc agatgglaag cctcccgta tctagttat ctacacgacg 5640
gggagtcagg caactatgga tgaacgaaat agacagatcg ctgagatagg tgcctcactg 5700
attaagcatt ggtaactgtc agaccaagtt tactcalata tacttttagat tgatttaaaa 5760
cttcattttt aatttaaaag gatctaggtg aagatccttt ttgataatct catgaccaaa 5820
atccctaac gtgagttttc gttccactga gcgtcagacc ccgtagaaaa galcaaagga 5880
tcllcttgag atcctttttt tctgcgcgta atctgctgct tgcaaacaaa aaaaccaccg 5940
ctaccagcgg tggtttgttt gccggatcaa gagctaccaa ctctttttcc gaaggtaact 6000
ggcllcagca gagegcagat accaaatact gtcttcttag tgtagccgta gttaggccac 6060
cacttcaaga actctgtage accgcctaca tacctcgtc tgctaactct gttaccagtg 6120
gtcgtcgcca gtggcgataa gtctgtctt accgggttg actcaagacg atagttaccg 6180
gataaggcgc agcggtcggg ctgaacgggg ggttcgtgca cacagcccag cttggagcga 6240
acgacctaca ccgaactgag atacctacag cgtgagcatt gagaaagcgc cacgcttccc 6300
gaaggagagaa aggcggacag gtatccgta agcggcaggg tcggaacagg agagcgcacg 6360
agggagcttc caggggaaa cgcttggtat ctttatagtc ctgtcgggtt tcgccacctc 6420
tgacttgagc gtcgattttt gtgatgtcgt tcaggggggc ggagcctatg gaaaaacgcc 6480
agcaacgcgg cctttttacg gttcclgccc ttttgcggc cttttgctca catgttcttt 6540
cctgcgttat cccctgatcc tgtggataac cgtattaccg cctttgagtg agctgatacc 6600
gtcgcgcgca gccgaacgac cgagcgcagc gagtcagtga gcgaggaagc ggaagagcgc 6660

```

[0016]

ccaatacgca aaccgcctct ccccgcgctg tggccgattc attaatccag ctggcacgac 6720
 aggtttcccg actggaaagc gggcagtgag cgcaacgcaa ttaatgtgag ttacctcaact 6780
 cattaggcac cccaggtctt acactttatg ctcccgctc gtatgttggt tggaattgtg 6840
 agcggataac aatttcacac aggaacacgc tatgacctg attacgaatt cgagctcggg 6900
 acccggggat cctctagagt cgacctgcag gcatgcaagc t 6941

<210> 32
 <211> 10475
 <212> DNA
 <213> *Saccharomyces cerevisiae*

<400> 32
 atcattctgg acgtatgtgc acatgtgatt tgcctttgtt tttttaagaa tgcggglaa 60
 taaacagatt gttttcttgg gaggataatc tttcttttt tccgtgttgt attctaaaat 120
 taaccttgct gtttcttttt ttttttttt tgcgcgact actcagccat ctgcatitt 180
 taaagaaaaa gataatcatt aatgccttca cgggaatacg tatagaacat tattaanaat 240
 atatgaatgg catatatata lagaacacca ccttggaac acatttatac ccttaaaact 300
 aaaacaattt gctgcgctat accgtgttgc agtgtattat aatacattca ttctgtttc 360
 attacgatta tattgacgtg ataaaaagat tatatagcca tgatcttctt aaacaccttc 420
 gcaagggtcc ttttaacgtg tttcgtactg tgcagcggtg cagcacgttc ctclgacaca 480
 aacgacacta ctccggcgtc tgcaaacgat ttgcagacca cttctttatt gacgtgtatg 540
 gacaattcgc aattaacggc atcattcttt gatglgaaat tttacccga taataatact 600
 gttatctttg atattgacgc tacgacgacg cttaatggga acgtcactgt gaaggctgag 660
 ctgcttactt acggactgaa agtcctggat aagacttttg atttatgttc cttgggcca 720
 gtategcttt ccccccctag tgcctggcgt attgatgtca tgccacaca ggtgatcgaa 780
 tcatccatta ccaagcaatt tcccgccatt gcttacacca ttccagattt ggacgcacaa 840
 gtacgtgtgg tggcatacgc tcagaatgac acggaattcg aaactccgtt ggcttgtgtc 900
 caggctatct tgagtaacgg gaagacagtg caaacaagt atgcggcctg gccattgcc 960
 gctatctcag gtgtcgggtg acttacctca gggtttgtgt ctgtgatcgg ttactcagcc 1020
 actgctgctc acattgcgtc caactccatc tcattgttca tatacttcca aaacttagct 1080
 atcactgcaa tgatgggtgt ctcaagggtt ccacccattg ctgcgcgtg gacgcagaat 1140
 ttccaatggt ccatgggtat catcaatata aactlcatgc aaaagatttt tgattggtac 1200
 gtacaggcca ctaatgtgt ctcaaatgtt gtggtagcta acaaggacgt cttgtccatt 1260
 agtgtgcaaa aacgtgctat ctctatggca tcgtctagt attacaattt tgacaccatt 1320
 ttagacgatt cggatctgta caccacttct gagaaggatc caagcaatta ctacgccaag 1380
 attctcgtgt taagaggtat agaaagattt gcttatttgg ctaatatga gctatcta 1440
 ttctttttga ccgtattgtt gtttttctta ttcttctat ttgtagtgt cgtctcttg 1500
 attttcttta aggcgtattt ggaagtctct acaagagcaa gaatattgaa agagacttcc 1560
 aatttcttcc aatataggaa gaactggggg agtattatca aaggcaccct ttacagatta 1620
 tctatcatcg ctttccctca agtttctct ctggcgattt gggaatttac tcaggccaac 1680
 tctccacgca ttgtgttga tgcggtagta atattaciga tcatctctt agagtcgacc 1740
 tgcaggcatg caagctagct tggcactggc cgtcgtttta caacgtcgtg actgggaaaa 1800
 ccctggcggt acccaactta atgccttgc agcacatccc ccttcgcca gctggcgtaa 1860

[0017]

tagcgaagag gccgcaccg atcgcccttc ccaacagttg cgcagcctga atggcgaattg	1920
gcgcctgatg cggatatttc tccttacgca tetgtgcgtt atttcacacc gcatagggtg	1980
ataactgata taattaaatt gaagcctctaa ttgtgagtt tagtatacat gcatttactt	2040
ataatacagt tttttagttt tgcctggccgc atcttctcaa atatgcttcc cagcctgctt	2100
tictgtaacg ttcacctctt accttagcat cccctccctt tgcaaatagt cctcttccaa	2160
caataataat gtcagatcct gtagagacca calcalccac ggttctatac tgttgaccca	2220
atgcgtctcc ctgtcatctt aaaccacac cgggtgtcat aatcaaccaa tcgtaacctt	2280
catctcttcc acctatgtct ctttgagcaa taaagccgat acaaaaatct ttgtcgtctt	2340
tcgcaatgtc aacagtaccc ttagtatatt ctccagtaga tagggagccc ttgcatgaca	2400
attctgctaa catcaaaagg cctctaggtt cctttgttac ttcttctgcc gctgcttca	2460
aaccgtaac aataaccggg cccaccacac cgtgtgcatt cgtaatgtct gccattctg	2520
ctattctgta tacaccgca gagtactgca atttgactgt attaccaatg tcagcaaat	2580
tictgtcttc gaagagtaaa aaattgtact tggcggataa tgcctttagc ggcttaactg	2640
tgcctccat ggaataatca gtcaagatat ccacatgtgt ttttagtaaa caaattttg	2700
gacctaatgc ttcaactaac tccagtaatt ccttggtggt acgaacatcc aatgaagcac	2760
acaagtttgi ttgcttttcg tgcattgatat taaatagctt ggagcaaca ggactaggat	2820
gagtagcagc acgttcttta tatgtactt tcgacatgat ttatcttctg ttcggttttt	2880
gttctgtgca gttgggttaa gaatactggg caatttcatt tttcttcaac actacatatg	2940
cgtatatata ccaactaag tctgtgtccc ttccttctgt ctctcttctg ttcgagatt	3000
accgaatcaa aaaaatttca aagaaccga aatcaaaaaa aagaataaaa aaaaatgat	3060
gaattgaaaa gctcttgta cccatcttg aattttgaac atccgaacct ggagttttc	3120
cctgaacag atagtatatt tgaacctgta taataatata tagtctagcg ctttacggaa	3180
gacaatgtat gtatttcggt tccaggagaa actattgcat ctattgcata ggtaacttg	3240
cacgtcgcac ccccggttca tttctgcgt ttccatcttg cacttcaata gcatacttt	3300
gttaacgaag catctgtgtt tcattttgta gaacaaaaat gcaacgcgag agcgttaatt	3360
tttcaacaa agaacttgag ctgcattttt acagaacaga aatgcaacgc gaaagcgcta	3420
ttttaccaac gaagaatctg tgcttcattt ttgtaaaaca aaaatgcaac gcgagagcgc	3480
taatttttca aacaaagaat ctgagctgca tttttacaga acagaaatgc aacgcgagag	3540
cgtattttta ccaacaaaga atctatactt cttttttgtt ctacaaaaat gcaccccgag	3600
agcgtatttt ttctaacaaa gcactttaga ttactttttt tctcctttgt gcgtcttala	3660
atgcagtctc ttgataactt tttgcactgt aggtccgtta aggttagaag aaggctactt	3720
tggtgtctat tttctcttc ataaaaaag cctgacacca ctccccgct ttactgatta	3780
ctagcgaagc tgcgggtgca ttttttcaag ataaaggcat ccccgattat attctatacc	3840
gatgtggatt gcgcatactt tgtgaacaga aagtgatagc gttgatgatt cttcattggt	3900
cagaaaatta tgaacggttt cttctatttt gtctctatat actacgtata ggaaatgttt	3960
acattttcgt attgttttcg attcactcta tgaatagtcc ttactacaat ttttttgtt	4020
aaagagtaat actagagata aacataaaaa atgtagaggt cgagttttaga tgcaagtcca	4080
aggagcgaag ggtggaagg taggttatat agggatatag cacagagata tatagcaag	4140
agatactttt gagcaatgtt tgtggaagcg gtattcgcaa ttttttagta gctcgttaca	4200
gtccgggtcgt tttttggttt ttgaaagtg cgtcttcaga gcgcttttgg ttttcaaaag	4260

[0018]

cgctctgaag ttctataact ttctagctag agaataggaa ctctggaata ggaacttcaa	4320
agcgtttccg aaaacgagcg ctccgaaaa tgcaacgga gctgcgcaca tacagctcac	4380
tgttcacgct gcacctatat ctgcgtgttg cctgtatata tatatacatg agaagaacgg	4440
catagtgcgt gtttatgctt aaatgcgtta tggtagcctc tcagtacaat ctgctctgat	4500
gccgcatagt taagccagcc ccgacacccg ccaacacccg ctgacgcgcc ctgacgggct	4560
tgtctgctcc cgccatccgc ttacagacaa gctgtgaccg tctccgggag ctgcatgtgt	4620
cagaggittt caccgtcatc accgaaacgc gcgagacgaa agggcctcgt gatacgccta	4680
tttttatagg ttaatgtcat gataataatg gtttcttaga cgtcagggtg cacttttcgg	4740
ggaaatgtgc gcggaacccc tatllgltta tttttctaaa tacattcaaa tatgtatccg	4800
ctcatgagac aataaccctg ataaatgctt caataatatt gaaaaaggaa gagtatgagt	4860
attcaacatt tccgtgtcgc ccttattccc ttttttgcgg cattttgcc tctgttttt	4920
gctcaccgag aaacgctggt gaaagtaaaa gatgctgaag atcagttggg tgcacgagtg	4980
ggttacatcg aactggatct caacagcggg aagatccttg agagttttcg ccccgaagaa	5040
cgttttccaa tgaagagcac ttttaaagtt ctgctatgtg gcgcggtatt atcccgatatt	5100
gacgccgggc aagagcaact cggtcgccgc atacactatt ctgagaatga ctgggttag	5160
tactcaccag tcacagaaaa gcattctacg gatggcatga cagtaagaga attatgcagt	5220
gctgccataa ccatgagtga taacactgcg gccaaacttac ttctgacaac gatcggagga	5280
ccgaaggagc taaccgcttt ttgcacaaac atgggggagc atgtaactcg ccttgatcgt	5340
tgggaaccgg agctgaatga agccatacca aacgacgagc gtgacaccac gatgcctgta	5400
gcantggcaa caacgttgcg caaactatta actggcgaac tacttactct agcttcccg	5460
caacaattaa tagactggat ggaggcggat aaagttgcag gaccacttct gcgctcgcc	5520
cttcggctg gtgtgtttat tctgataaaa tctggagccg gtgagcgtgg gtctcgggt	5580
atcattgcag cactggggcc agatgglaag ccttcccgtg tctgtattat ctacacgacg	5640
gggagtcagg caactatgga tgaacgaaat agacagatcg ctgagatagg tgcctcactg	5700
attaagcatt ggtaactgtc agaccaagtt tactatata tactttagat tgatttaaaa	5760
cttcattttt aatttaaaag gatctagggt aagatccttt ttgataatct catgacaaa	5820
atcccttaac gtgagttttc gttccactga gcgtcagacc ccgtagaaaa gatcaaagga	5880
tcctcttag atcctttttt tctgcgcgta atctgctgct tgcaaacaaa aaaaccaccg	5940
ctaccagcgg tggtttgttt gccggatcaa gagctaccaa ctctttttcc gaaggtaact	6000
ggcttcagca gagcgcagat accaaatact gtccctctag tctagccgta gttaggccac	6060
cacttcaaga actctgtagc accgcctaca taactcgtc tgctaatact gttaccagt	6120
gctgctgcca gtggcgataa gtcgtgtctt accgggttg actcaagacg atagttaccg	6180
gataaggcgc agcggtcggg ctgaacgggg ggttcgtgca cacagcccag ctggagcga	6240
acgacctaca ccgaactgag alacctacag cgtgagcatt gagaaagcgc cagccttccc	6300
gaagggagaa agcggacag gtatccgcta agcggcaggg tcggaacagg agagcgacg	6360
aggagcttc cagggggaaa cgcctgggtat ctttatagtc ctgtcgggtt tccccacctc	6420
tgacttgagc gtcgattttt gtgatgctcg tcaggggggc ggagcctatg gaaaaacgcc	6480
agcaacgcgg cttttttacg glicctggcc ttttctggc cttttgtca catgttctt	6540
cctgcgttat cccctgattc tgtggataac cgtattaccg cctttgagtg agctgatacc	6600
gctcgcgcga gccgaacgac cgagcgacgc gactcagtga gcgaggaagc ggaagagcgc	6660

[0019]

ccaatacgca aaccgcctct ccccgcgctg tggccgattc attaatccag ctggcacgac	6720
aggtttcccg actggaaagc gggcagtgag cgcaacgcaa ttaatgtgag ttacctcaact	6780
cattagggcac ccaggtttt acactttatg cticcggctc gtatgttggtg tggattgtg	6840
agcggataac aatttcacac aggaacacgc tatgacctg attacgaatt cgagctcgg	6900
accggccgca aattaaagcc ttcgagcgtc ccaaacctt ctcaagcaag gttttcagta	6960
taatgttaca tgcgtacacg cgtctgtaca gaaaaaaaag aaaaattga aatataaata	7020
acgtttctta tactaacata actataaaaa aataaatagg gacctagact tcaggttg	7080
taactccttc cttttcgggt agagcggatg tggggggagg gcgtgaatgt aagcgtgaca	7140
taactaatta catgactcga gttacgcata gtcaggaaca tcgtatgggt accgtcgggt	7200
ctcgttttca cgcacccggc gcatgtactc ctcttggtac gaccgcacat tgtttttccg	7260
atcggcggtt gtcacctgat cgaaggggac gcggtgtgcc agcatctcct ggaiclcctt	7320
cggcgccggg tactgcccgg ggcagatcca cgaccaggc gggtagcaca cgcgggttg	7380
cgggtcagca acccactttt tgctctccgc gctcacgttc atgaccttga agatgcgcac	7440
caggccgtac ttcgacgagt acacctctg aaagagggac gggttcacct tcacgccctt	7500
tcttttcccg gcctcgtgca ggttgtagcag cagcgacgcc cgcacatcg gcgttggcg	7560
actgtagtca tttctgtgaa agccgaattg ctggcacagc gggcgcgcgg ggcagatgtc	7620
gtggtacacg ctgttgccga tgcgcgccat gtgcggtagc ttcatcaggt cggcgtctg	7680
cccagcccag atcaggacgt agtcggccat gtggcgcacc agcgagtgcg cctccgccac	7740
gggcgacgtc agcatcttgc cgatcgtggc gatgtgctcg tggttccagg tgttgccatc	7800
ggccagcgag gtgcgggtgc cgatgcctgt gatctggtag ccgtagtccc accaggccaa	7860
aacgcgcgcg tctctggcg tgctgctgcg cagccactcg taggccttga ggtagtcatc	7920
caccaatagg ttcataggtt tgctgtggc acggttttgc acgacggcg cgaanaaat	7980
catcggattt gacgactgct ccgcaactt tgltagtggt gacgcgaact cggaaactgaa	8040
gaagctcacc gcggttgctg tgacgagagc ccacatagcg atggacagga ccatgcgatg	8100
gccccagca agagagctac cggcaagac gtcgcaaat gcgcgcgcgg tcgttgcggt	8160
cttagcgta tctcgccgc tgcccttgcc agccccctc tggtagcggt gcgctgctt	8220
ctgctgcttt ttagcctttg tcgcatcact gtcccagaaa ctgagttgca cggctgcttc	8280
cagaattgtc ccgacgaaaa tgccagtga cagacacgca cggggcccg agagaagcag	8340
cagccagacc atgcgagtg tgaagtagta cagggcgcg gagttcagta gccagaatac	8400
cttcgacggg gtagtgta cgaacgtcga gacagcaagc acaatggagc ccaaaccaca	8460
tgtcacgcgc cacacgtgaa gaaaagccca catcgctcc gggctggcgg gttgatgttc	8520
ggcgaccgag tcgaccagcg gatlgccagt gcgcgtgtgc tccacgaaca gcgcacgcac	8580
acggaccgag agggggccga agtaccctg cggtgccagc accgagattg caagcgagc	8640
cacgccagcc atcacgtga agacgcgcac gcggatcttg aagttcgcgc gagagcgaac	8700
ctcgacacgc gcgcgtgcgc gcagcacctc gcacacctgc agtccacaca ggaagacaag	8760
caccagcagc gcacccagct gctccagcga ctgaagggc gacatcccca ctggcggcac	8820
gcacacggcg atggcggtgc ccacagcgtg gaacagcgtg tatgcacgca gcagcgacgg	8880
gttgtacgtg ttgcgggccc agtccacctc cgtatgatg ccggcatgca tggcaacct	8940
gttagcagc aaaatgtagc cggccacgc cggcccatg tagccgtagg cgacaccggt	9000
gaggacaccg atgggccagc aggaccgcgt gcgcagcgag gcacccagc agtagaaagt	9060

[0020]

gaggagcatg gctgcgacgg cgaatgcactc gttgtcgaac tcacccgcca tggaccgcat 9120
 cagggtggct gggatgatgg agaaggagag lgcggcagcg gcccccgcta ctgtcgaccc 9180
 actggcttcg laggtgcaaa acgccagagt agcggtagcg atggcgccaa accacgtgg 9240
 catcagcagc cacacgttgt tgagagacat cggcatgccg gcagccgcca gtgcgcggtg 9300
 aatggcgacg gcagtgaagt gcaggcccgg gtacgtggtg gagccgacgg ggcggcccag 9360
 cgggtaccag ctcatgtagt cgaaccagct gaagaaggcg gaccagccgt gcgtggacat 9420
 gtactcggca gcgcgglagl tgaaccacgg gtgcgaactc tggatcaggt atccgtaaat 9480
 ctgaacggag atcatgcgaa ccgtgaaggc ttggaagcag ctggcggcta agacgaagag 9540
 tgccaccacg gtgaggacga agtgtactgg ccagaaagga aacggaaaga tgccgatgaa 9600
 atcccttcca tctgttagcg ttttggcgag caaaatcacc ttgcgcggtg gagatgcgtt 9660
 cttggttgtt gaggcggcat ctccggcttg ggccgaagct tcacgagatg cggttgccgc 9720
 agagccggag tcgccaatg aatttccctt ccgttgcgc atgttgttac tagttctaga 9780
 atccgtcgaa actaagttct ggtgttttaa aactaaaaa aagactaact ataaaagtag 9840
 aatttaagaa gtttaagaaa tagatttaca gaattacaat caatacciac cgtctttata 9900
 tactatttag tcaagtaggg gaataaattc agggaaactg tttaacctt tttttcagc 9960
 tttttccaaa tcagagagag cagaaggtaa tagaaggtgt aagaaaaatga gatagataca 10020
 tgcgtgggtc aattgccttg tgtcatcatt tactccagcg aggttgcac actccattga 10080
 ggttgtgccc gttttttgcc tgtttgtgcc cctgttcctt glagttgcgc taagagaatg 10140
 gacctatgaa ctgatggttg gtgaagaaaa caatattttg gtgctgggat tctttttttt 10200
 tctggatgcc agcttaaaaa gcgggctcca ttatatattg tggatgccag gaataaactg 10260
 ttacccaga cacctacgat gttatatatt ctgtgtaacc cgccccctat ttgggcatg 10320
 tacgggttac agcagaatta aaaggctaatt tttttgacta aataaagtta ggaataacac 10380
 tactattaat tattttacga ttctttgaaa tggcgagtat tgataatgat aaactgaggg 10440
 ggatcctcta gagtcgacct gcaggcatgc aagct 10475

<210> 33
 <211> 2580
 <212> DNA
 <213> Leishmania braziliensis

<400> 33
 atccgatca agaaccagcg caaaggatgc gaggagggtg accccaaccc ctctccaca 60
 cccgcagcag agccactggc aaacgcagaa ggcacgcaga gggataccgc tgaagggact 120
 cctatggagc caccagcga gacgtacctc ttcaactgcc gcgccgcacc gtactcgaag 180
 ctgatatacg tctacaaaag tatcatgttc acattgattc tctacgcgat ccgcttagcg 240
 taccagactc gcattctatc cgttcagact tatggttaca tcatccaga gttcgaccgc 300
 tggttcaact accgcgccgc cgagtacatg tccgcgcacg gctggtcgcg cttcttcagc 360
 tggttcgact acatgagctg gtaccccgctg ggccgccttg ttggcaccac cacgtaccgc 420
 ggcttcgacg tcaccgccgt tgccatccac cgcgcatttg cagctgccgg ggtgccgatg 480
 tctctcaaca acgtgttgtg gctgacccc gcgtggtatg gtccatcgc tactgctatc 540
 atggccctca tggccttcga aacgactggc tcgacgctg tttctgcatg ggctgcactc 600
 ctcttctcca tcatccagc acacctgatg cgggtccatg cgggcgagtt cgacaacgag 660
 tgcacgccc ttgcagccat gctcctcacc ttctacttgt gggtacgctc gctgcgcacg 720
 cgggtgctct ggcccatcgg catcctcacc ggtatcgctt acggctacat ggtggcggcg 780

[0021]

```

tggggcggat acatttttgt gctcaacatg gttgccatgc acgccggcat atcatcgatg 840
gtcgactggg ctgcgaacac gtacaacccg tcgctgctgc gcgcatacgc gctgtttctac 900
gttgctggca ccgccatgc caccgcgctg ccgcctgtgg ggatgtcgcc cttcaggctcg 960
ctggagcagc tgggtgcgct ggtggtgctc ctcttcctgt gcgggctgca ggccctgcgag 1020
gtgtttcgcg caccggccga cgtcgagggt cgctcccgcg cgaacttcaa gatccgcatg 1080
cgtgccitca gcgtgatggc tggcgtgggt gcgcttgcaa tcgcggtgct gtcgccgacc 1140
gggtactttg gccccctcac ggctcgtgtg cgtgcgctgt tcatgaagca caccgacact 1200
ggcaatccgc tggtcgactc ggtcgtgag caccaccccg cagacgcgct cgcctacctg 1260
caatatattga acattgtgta cgttttgtgg gtatttagca tccctgtgca gctgatcctg 1320
cccccccta acctgtatgc gattctcttt ctccctgtgt acagttgcat ggctactat 1380
ttcagcactc gcattggtgcg ctgtctctg clggctggcc cagtggcggtg ccttagcggg 1440
agtttgatga gtggtacgt gacgaagtgg tgctttcaac agctgttctg ggacgacaac 1500
ctgcgcaccg ccgatatggc ggcggtggt gatactccgt ttccacaaga ggaccacccc 1560
aacagcggtg caccgcgccg acggaaccag cagaacaga aggcgaccca ggctcctgcc 1620
agaggctcaa gcacaggcga cgaagaacga cgttacacat cactaatccc ttttgacttc 1680
cgcaaggaga tcaagatgaa ccgctggccg accggaaaaa agcaagccac gttcatcatc 1740
tctgccacca tcgtaccgt tcttcgctt gccttgtct actacttctc atgcacttcc 1800
atggcaact cttgtcgag cccacagatc ctgtacaaa cccgtatggg gggcaagacg 1860
atcatggtgg ctgactatct cgagtcatac gagtggctgc gcgacaacac gccagcggac 1920
gcgcgcgtgc tgtcctggtg ggactacggc taccagatca caggcatcgg caaccgcacc 1980
tcgtggccg atggcaacac ctggaaccac gagcacatcg ccaccatcgg caagatgctg 2040
acgtcgccg tggcggaggc gcactcactg gtgcgccaca tggcggacta cgtcctcacc 2100
tgggctggg agggcggaga cttgatgaag tcgcgcaca tggcgcgat tggcaacagc 2160
gtgtaccacg acatctgcc caacgacccg ctttgccagc atttcggctt ttacgaagac 2220
tacagtcgcc caaaaccgat gatgcgcgcg tcgctgctgt acaacctgca cgaggccgga 2280
cgaagcgcgg gtgtgaaggt ggaccgctc ctctttcagg aagtgtactc atccaagtac 2340
ggcctggtgc gcatttcaa ggtcatgaac gtgagcgcgg agagcaagaa gtgggtggct 2400
gacccgcaa accgcgtgtg ccaccgcct gggtcgtgga tctgccccgg gcagtaccgg 2460
ccggcgaagg agatccagga gatgctggcg caccgcgtcc ctttgacca gatgggcaag 2520
aagcacgacg acacgcaca ggcgcgcaig gcacgcagca gaacactggg cgaggcttga 2580

```

<210> 34
 <211> 859
 <212> PRT
 <213> Leishmania braziliensis

<400> 34

Met Pro Ile Lys Asn Gln Arg Lys Gly Cys Glu Glu Gly Asn Pro Asn
 1 5 10 15

Pro Ser Ser Thr Pro Ala Ala Glu Pro Leu Ala Asn Ala Glu Gly Thr
 20 25 30

Gln Arg Asp Thr Ala Glu Gly Thr Pro Met Glu Pro Pro Ser Glu Thr
 35 40 45

[0022]

```

Tyr Leu Phe Asn Cys Arg Ala Ala Pro Tyr Ser Lys Leu Ile Tyr Val
 50                               55                               60

Tyr Lys Gly Ile Met Phe Thr Leu Ile Leu Tyr Ala Ile Arg Leu Ala
 65                               70                               75                               80

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

Glu Phe Asp Pro Trp Phe Asn Tyr Arg Ala Ala Glu Tyr Met Ser Ala
                        100                               105                               110

His Gly Trp Ser Ala Phe Phe Ser Trp Phe Asp Tyr Met Ser Trp Tyr
                        115                               120                               125

Pro Leu Gly Arg Pro Val Gly Thr Thr Thr Tyr Pro Gly Leu Gln Leu
 130                               135                               140

Thr Ala Val Ala Ile His Arg Ala Leu Ala Ala Ala Gly Val Pro Met
 145                               150                               155                               160

Ser Leu Asn Asn Val Cys Val Leu Ile Pro Ala Trp Tyr Gly Ala Ile
                        165                               170                               175

Ala Thr Ala Ile Met Ala Leu Met Ala Phe Glu Thr Thr Gly Ser Ile
                        180                               185                               190

Ala Val Ser Ala Trp Ala Ala Leu Leu Phe Ser Ile Ile Pro Ala His
                        195                               200                               205

Leu Met Arg Ser Met Ala Gly Glu Phe Asp Asn Glu Cys Ile Ala Val
 210                               215                               220

Ala Ala Met Leu Leu Thr Phe Tyr Leu Trp Val Arg Ser Leu Arg Thr
 225                               230                               235                               240

Arg Cys Ser Trp Pro Ile Gly Ile Leu Thr Gly Ile Ala Tyr Gly Tyr
                        245                               250                               255

Met Val Ala Ala Trp Gly Gly Tyr Ile Phe Val Leu Asn Met Val Ala
                        260                               265                               270

Met His Ala Gly Ile Ser Ser Met Val Asp Trp Ala Arg Asn Thr Tyr
                        275                               280                               285

Asn Pro Ser Leu Leu Arg Ala Tyr Ala Leu Phe Tyr Val Val Gly Thr
 290                               295                               300

Ala Ile Ala Thr Arg Val Pro Pro Val Gly Met Ser Pro Phe Arg Ser
 305                               310                               315                               320

Leu Glu Gln Leu Gly Ala Leu Val Val Leu Leu Phe Leu Cys Gly Leu
                        325                               330                               335

Gln Ala Cys Glu Val Phe Arg Ala Arg Ala Asp Val Glu Val Arg Ser
                        340                               345                               350

Arg Ala Asn Phe Lys Ile Arg Met Arg Ala Phe Ser Val Met Ala Gly
                        355                               360                               365

```

[0023]

Val Gly Ala Leu Ala Ile Ala Val Leu Ser Pro Thr Gly Tyr Phe Gly
 370 375 380
 Pro Leu Thr Ala Arg Val Arg Ala Leu Phe Met Lys His Thr His Thr
 385 390 395 400
 Gly Asn Pro Leu Val Asp Ser Val Ala Glu His His Pro Ala Asp Ala
 405 410 415
 Leu Ala Tyr Leu Gln Tyr Leu Asn Ile Val Tyr Val Leu Trp Val Phe
 420 425 430
 Ser Ile Pro Val Gln Leu Ile Leu Pro Thr Pro Asn Leu Tyr Ala Ile
 435 440 445
 Leu Phe Leu Leu Val Tyr Ser Cys Met Ala Tyr Tyr Phe Ser Thr Arg
 450 455 460
 Met Val Arg Leu Leu Leu Leu Ala Gly Pro Val Ala Cys Leu Ser Gly
 465 470 475 480
 Ser Leu Met Ser Gly Thr Leu Thr Lys Trp Cys Phe Gln Gln Leu Phe
 485 490 495
 Trp Asp Asp Asn Leu Arg Thr Ala Asp Met Ala Ala Ala Gly Asp Thr
 500 505 510
 Pro Phe Ser Gln Glu Asp His Pro Asn Ser Gly Ala Arg Ala Arg Arg
 515 520 525
 Asn Gln Gln Lys Gln Lys Ala Thr Gln Ala Pro Ala Arg Gly Ser Ser
 530 535 540
 Thr Gly Asp Glu Glu Arg Arg Tyr Thr Ser Leu Ile Pro Phe Asp Phe
 545 550 555 560
 Arg Lys Glu Ile Lys Met Asn Arg Trp Pro Thr Gly Lys Lys Gln Ala
 565 570 575
 Thr Phe Ile Ile Ser Ala Thr Ile Cys Thr Val Leu Pro Leu Ala Phe
 580 585 590
 Val Tyr Tyr Phe Ser Cys Thr Ser Met Ala Asn Ser Leu Ser Ser Pro
 595 600 605
 Gln Ile Leu Tyr Gln Thr Arg Met Gly Gly Lys Thr Ile Met Val Ala
 610 615 620
 Asp Tyr Leu Glu Ser Tyr Glu Trp Leu Arg Asp Asn Thr Pro Ala Asp
 625 630 635 640
 Ala Arg Val Leu Ser Trp Trp Asp Tyr Gly Tyr Gln Ile Thr Gly Ile
 645 650 655
 Gly Asn Arg Thr Ser Leu Ala Asp Gly Asn Thr Trp Asn His Glu His
 660 665 670
 Ile Ala Thr Ile Gly Lys Met Leu Thr Ser Pro Val Ala Glu Ala His
 675 680 685

[0024]

Ser Leu Val Arg His Met Ala Asp Tyr Val Leu Ile Trp Ala Gly Gln
 690 695 700
 Gly Gly Asp Leu Met Lys Ser Pro His Met Ala Arg Ile Gly Asn Ser
 705 710 715 720
 Val Tyr His Asp Ile Cys Pro Asn Asp Pro Leu Cys Gln His Phe Gly
 725 730 735
 Phe Tyr Glu Asp Tyr Ser Arg Pro Lys Pro Met Met Arg Ala Ser Leu
 740 745 750
 Leu Tyr Asn Leu His Glu Ala Gly Arg Ser Ala Gly Val Lys Val Asp
 755 760 765
 Pro Ser Leu Phe Gln Glu Val Tyr Ser Ser Lys Tyr Gly Leu Val Arg
 770 775 780
 Ile Phe Lys Val Met Asn Val Ser Ala Glu Ser Lys Lys Trp Val Ala
 785 790 795 800
 Asp Pro Ala Asn Arg Val Cys His Pro Pro Gly Ser Trp Ile Cys Pro
 805 810 815
 Gly Gln Tyr Pro Pro Ala Lys Glu Ile Gln Glu Met Leu Ala His Arg
 820 825 830
 Val Pro Phe Asp Gln Met Gly Lys Lys His Asp Asp Thr His Lys Ala
 835 840 845
 Arg Met Ala Arg Ser Arg Thr Leu Gly Glu Ala
 850 855
 <210> 35
 <211> 2319
 <212> DNA
 <213> Leishmania braziliensis
 <400> 35
 atgtactgcc taaacaaggc ctaacgcatt cgcattgttt ccgttcagct ttatggctac 60
 atcatccacg agttcgaccc gtgggtcaac taccgcgcgc ccgagtacat gtccgcgcac 120
 ggctggtcgc cttcttcag ctgggtcgac tacatgagct ggtacccgct gggccgcccc 180
 gttggcacca ccacgtaccc gggcctgcag ctcaccgcgc ttgccatcca ccgcgcattg 240
 gcagctgccg ggggtccgat gtctctcaac aacgtgtgcg tgctgatccc cgcgtggtat 300
 ggtgccatcg ctactgctct agaagcgcta atgatctatg agtgaacgg ctccggaatt 360
 accgctgcca tcggagcitt tatctttatg attctccccg cacacctgat gcggtccatg 420
 gccggcgagt tcgacaacga gtgcacgcgc gttgcagcca tgctccctac ctctacttg 480
 tgggtacgct cgtgcgcac gcgggtgctg tggcccatcg gcacccctac cggtatcgcc 540
 tacggctaca tggggcgccg gtggggcgga tacatttttg tgctcaacat ggttgccatg 600
 caccgccgga tatcatgat ggtcgactgg gctcgcaaca cgtacaaccc gtcgctgctg 660
 cgcgcatacg cgtgtttcta cgttgtcgcc accgccatcg ccacgcgcgt gccgcctgtg 720
 gggatgtcgc ctttcaggtc gctggagcag ctgggtgcgc tggcgggtgct cctcttctg 780
 tgcgggctgc aggcctgcga ggigtgtcgc gcacgggcgc acgtcgaggt tcgctccgc 840
 gcgaacttca agatccgat gcgtgccttc agcgtgatgg ctggcgtggg tgcgcttgca 900

[0025]

```

atcgcggtgc tgtcgccgac cgggtacttt ggccccctca cggtctcgtgt gcgtgcgctg 960
ttcatggagc acacgcgcac tggcaatccg ctggtcgact cggtcgctga gcaccgcaaa 1020
acgaaccacac aggcgtacga gtactttctg gactttacct attcgatgtg gatgctggga 1080
gcagtgttgc agttgctcgg tgcagccgtt ggctcacgaa aggaggcgcg gctgttcatg 1140
gggctgtact cactcgccac ctactacttc tcagatcgca tgtcacggct gatggtactt 1200
gcggggcctg cggtgcgcg gatagcagcg gaaatcttgg gcatcccata cgagtgggtg 1260
tggacgcagc tgacgggatg ggcatctccg aacacctccg ccagagagcg taaaagcaag 1320
gaggacggtc cctgcaagac aaaaagaaat cagagacaga ccgtcgccac aaaactagat 1380
catggggcgc gggctagggc tacggccgct gtcaagtcca tggagacggc tctggagcgt 1440
gttctcttgg tgtttcgagc tgccatgcc ataggcatca ttggggccac tgttgaaca 1500
ccgtacgict atcagttcca ggctcgttgc attcaatctt cctattcttt tgcgtgcccg 1560
cgtatcatgt tccacacgca gctgcgcacc ggcgaaacag tgattgtaa ggactacgtg 1620
gaagcatacg agtggctcgc gcacaacacg ccagcggacg cgcgctgct gtcttggtg 1680
gactacggct accagatcac aggtatcggc aaccgcacct cgctggccga tggcaacacc 1740
tggaaccaag agcacatgc caccatcggc aagatgctga cgtcgccgt ggccggagcg 1800
cactcactgg tgcgccacat ggccgactac gtctctatct gggctgggca gggcggagac 1860
ttgatgaagt cgccgcacat ggcgcgatt ggcaacagcg tgtaccaga catctgcccc 1920
aacgacccgc ttgccagca ttctggcttt tacgaagact acagtcgcc aaaaccgatg 1980
atgcgcgcgt cgctgctgta caacctgcac gaggccggac gaagcgcggg tgtgaagggtg 2040
gaccgcgtcc tctttcagga agtgtactca tccaagtacg gcctggtgcg catcttcaag 2100
gtcatgaacg tgagcgcgga gagcaagaag tgggtggctg acccgcaaaa ccgctgtgct 2160
caccgcctg ggtcgtggat ctgccccggg cagtaccgc cggcgaagga gatccaggag 2220
atgctggcgc accgcgtccc cttgaccag atgggcaaga agcacgacga cagcacaag 2280
gcgcgcatgg cagcagcag aacactgggc gaggcttga 2319

```

<210> 36
 <211> 772
 <212> PRT
 <213> Leishmania braziliensis

<400> 36

Met Tyr Cys Leu Asn Lys Ala Tyr Arg Ile Arg Met Phe Ser Val Gln
 1 5 10 15

Leu Tyr Gly Tyr Ile Ile His Glu Phe Asp Pro Trp Phe Asn Tyr Arg
 20 25 30

Ala Ala Glu Tyr Met Ser Ala His Gly Trp Ser Ala Phe Phe Ser Trp
 35 40 45

Phe Asp Tyr Met Ser Trp Tyr Pro Leu Gly Arg Pro Val Gly Thr Thr
 50 55 60

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

Ala Ala Ala Gly Val Pro Met Ser Leu Asn Asn Val Cys Val Leu Ile
 85 90 95

[0026]

```

Pro Ala Trp Tyr Gly Ala Ile Ala Thr Ala Leu Glu Ala Leu Met Ile
    100                      105                      110

Tyr Glu Cys Asn Gly Ser Gly Ile Thr Ala Ala Ile Gly Ala Phe Ile
    115                      120                      125

Phe Met Ile Leu Pro Ala His Leu Met Arg Ser Met Ala Gly Glu Phe
    130                      135                      140

Asp Asn Glu Cys Ile Ala Val Ala Ala Met Leu Leu Thr Phe Tyr Leu
    145                      150                      155                      160

Trp Val Arg Ser Leu Arg Thr Arg Cys Ser Trp Pro Ile Gly Ile Leu
    165                      170                      175

Thr Gly Ile Ala Tyr Gly Tyr Met Val Ala Ala Trp Gly Gly Tyr Ile
    180                      185                      190

Phe Val Leu Asn Met Val Ala Met His Ala Gly Ile Ser Ser Met Val
    195                      200                      205

Asp Trp Ala Arg Asn Thr Tyr Asn Pro Ser Leu Leu Arg Ala Tyr Ala
    210                      215                      220

Leu Phe Tyr Val Val Gly Thr Ala Ile Ala Thr Arg Val Pro Pro Val
    225                      230                      235                      240

Gly Met Ser Pro Phe Arg Ser Leu Glu Gln Leu Gly Ala Leu Ala Val
    245                      250                      255

Leu Leu Phe Leu Cys Gly Leu Gln Ala Cys Glu Val Phe Arg Ala Arg
    260                      265                      270

Ala Asp Val Glu Val Arg Ser Arg Ala Asn Phe Lys Ile Arg Met Arg
    275                      280                      285

Ala Phe Ser Val Met Ala Gly Val Gly Ala Leu Ala Ile Ala Val Leu
    290                      295                      300

Ser Pro Thr Gly Tyr Phe Gly Pro Leu Thr Ala Arg Val Arg Ala Leu
    305                      310                      315                      320

Phe Met Glu His Thr Arg Thr Gly Asn Pro Leu Val Asp Ser Val Ala
    325                      330                      335

Glu His Arg Lys Thr Asn Pro Gln Ala Tyr Glu Tyr Phe Leu Asp Phe
    340                      345                      350

Thr Tyr Ser Met Trp Met Leu Gly Ala Val Leu Gln Leu Leu Gly Ala
    355                      360                      365

Ala Val Gly Ser Arg Lys Glu Ala Arg Leu Phe Met Gly Leu Tyr Ser
    370                      375                      380

Leu Ala Thr Tyr Tyr Phe Ser Asp Arg Met Ser Arg Leu Met Val Leu
    385                      390                      395                      400

Ala Gly Pro Ala Ala Ala Ala Ile Ala Ala Glu Ile Leu Gly Ile Pro
    405                      410                      415

```

[0027]

Tyr Glu Trp Cys Trp Thr Gln Leu Thr Gly Trp Ala Ser Pro Asn Thr
 420 425 430
 Ser Ala Arg Glu Arg Lys Ser Lys Glu Asp Gly Pro Cys Lys Thr Lys
 435 440 445
 Arg Asn Gln Arg Gln Thr Val Ala Thr Lys Leu Asp His Gly Ala Arg
 450 455 460
 Ala Arg Ala Thr Ala Ala Val Lys Phe Met Glu Thr Ala Leu Glu Arg
 465 470 475 480
 Val Pro Leu Val Phe Arg Ala Ala Ile Ala Ile Gly Ile Ile Gly Ala
 485 490 495
 Thr Val Gly Thr Pro Tyr Val Tyr Gln Phe Gln Ala Arg Cys Ile Gln
 500 505 510
 Ser Ser Tyr Ser Phe Ala Val Pro Arg Ile Met Phe His Thr Gln Leu
 515 520 525
 Arg Thr Gly Glu Thr Val Ile Val Lys Asp Tyr Val Glu Ala Tyr Glu
 530 535 540
 Trp Leu Arg Asp Asn Thr Pro Ala Asp Ala Arg Val Leu Ser Trp Trp
 545 550 555 560
 Asp Tyr Gly Tyr Gln Ile Thr Gly Ile Gly Asn Arg Thr Ser Leu Ala
 565 570 575
 Asp Gly Asn Thr Trp Asn His Glu His Ile Ala Thr Ile Gly Lys Met
 580 585 590
 Leu Thr Ser Pro Val Ala Glu Ala His Ser Leu Val Arg His Met Ala
 595 600 605
 Asp Tyr Val Leu Ile Trp Ala Gly Gln Gly Gly Asp Leu Met Lys Ser
 610 615 620
 Pro His Met Ala Arg Ile Gly Asn Ser Val Tyr His Asp Ile Cys Pro
 625 630 635 640
 Asn Asp Pro Leu Cys Gln His Phe Gly Phe Tyr Glu Asp Tyr Ser Arg
 645 650 655
 Pro Lys Pro Met Met Arg Ala Ser Leu Leu Tyr Asn Leu His Glu Ala
 660 665 670
 Gly Arg Ser Ala Gly Val Lys Val Asp Pro Ser Leu Phe Gln Glu Val
 675 680 685
 Tyr Ser Ser Lys Tyr Gly Leu Val Arg Ile Phe Lys Val Met Asn Val
 690 695 700
 Ser Ala Glu Ser Lys Lys Trp Val Ala Asp Pro Ala Asn Arg Val Cys
 705 710 715 720
 His Pro Pro Gly Ser Trp Ile Cys Pro Gly Gln Tyr Pro Pro Ala Lys
 725 730 735

[0028]

Glu Ile Gln Glu Met Leu Ala His Arg Val Pro Phe Asp Gln Met Gly
740 745 750

Lys Lys His Asp Asp Thr His Lys Ala Arg Met Ala Arg Ser Arg Thr
755 760 765

Leu Gly Glu Ala
770

<210> 37
<211> 2565
<212> DNA
<213> Leishmania braziliensis

<400> 37
atgggtaaga agaaagcaat tccgtcgggc agcgtcggcc ctgcgacaac cacctcccgt 60
gaagctccag gcaaagacga aggtgcctcc caaccgccca agactgcagc tctgccggtg 120
aagccctttg tgttgcccaa cacgtcgaca gacgaggagg agtttgttgg catctttccc 180
tgccctttct ggccagtgcg atttgcctac acagtgaagg cactcgtcct cttgggtgac 240
agctgtatcc gcgccttcac gattcgcatt ctatccgttc agctttatgg ctacatcctc 300
cacgagttcg acccgtggtt caactaccgc gccgccgagt acatgtccgc gcacggctgg 360
tccgccttct tcagctggtt cgactacatg agctgggtacc cgtcgggccg ccccggttggc 420
accaccacgt acccgggcc tgcagtcacc gccgttgcca tccaccgcgc attggcggtc 480
gccgggggtg cgtatgtctc caacaacgtg tgcgtgtga tccccgctg gtatgglgcc 540
atcgctactg ctatcctggc ctttgcgtc tacgaggcca gtaggtcaat gtagcgggcg 600
gctgttctg cactctcatt ctccatcatt ccagcacacc tgatgcggtc catggcgggc 660
gagttcgaca acgagtgcct cgccttgcca gccatgcctc tcaccttcta cttgtgggta 720
cgctcgctgc gcacgcggtg ctctggccc atcgccatcc tcaccggtat cgcctacggc 780
tacatggttg cggcgtgggg cggatacatt tttgtgtca acatggttgc catgcacgcc 840
ggcatatcat cgtatgtcga ctgggctcgc aacacgtaca acccgtcgtc gctgcgcgca 900
tacgcgtgt tctacgttgt cggcacccgc atcgccacgc gcgtgccgcc tgtggggatg 960
tcgcccttca ggtcgttga gcagctgggt gcgtggcggt tgctcctctt cctgtgcggg 1020
ctcgagccct gcgaggtgtt tcgcgcacgg gccgacgtc aggttcgctc ccgcgcgaac 1080
ttcaagatcc gcattcgtgc cttcagcgtg atggctggcg tgggtgcgtc tgcaatcgcg 1140
gtgctgtgc cgaccgggta ctllggcccc ctcacggctc gtgtgcgtgc gctgttcatt 1200
gagcacacgc gcactggcaa tccgtggtc gactcggctc ctgagcacca ccccgccagt 1260
cctgagggca tgtggacatt tcttcacgtg tcggcggtga cttggggttt gggtccatt 1320
gttcttcttg tgcgttgcgt ggtggactac tcccgccaa agctcttttg gctgatgaac 1380
tctggtgccg tgtactattt cagcacccgc atgtcacgac tgcgtctct caccggcccc 1440
gctgcgtgtc tgccactgg ctgtttcgtg gggacattac tggaagcggc gatacagttc 1500
accttctgtt ccagcgtgc aacaaaggcc aaaaaacagc aagagacaca acttcaccaa 1560
aaggggcgcc gcaagcatag cgaccggagt aactctaaga atgcactgac tgtgcgtaca 1620
ttggggcgac tcttgaggag tacctctctg gcattgggtc atcgcatggt gctctgcttc 1680
gctatgltgg ctcttgttat tacagtcgag gtgtgcctct tgggttccga ttactcttc 1740
catgcaacga tgtttgcaag gcagacgtc aaccgcgtga ttgtctttgc aaccgtgctg 1800
cgagaccgag ctaccggcaa gccaacacag gtattggtgg atgactacct gcgcagctat 1860

[0029]

```

ctctggctgc gcgacaacac gccagaaat gcgcgcgtgc tgtcctggtg ggactacggc 1920
taccagatca caggatcgg caaccgcacc tcgctggccg atggcaacac ctggaaccac 1980
gagcacatcg ccaccatcgg caagatgctg acgtcgcccg tggcggaggc gcactcactg 2040
gtgcgccaca tggcggacta cgctctcatc tgggctgggc agggcggaga cttgatgaag 2100
tcgcgcgaca tggcgcgcat tggcaacagc gtgtaccacg acatctgccc caacgacccg 2160
ctttgccagc atttcggctt ttacaagaac gatcgcaatc gcccaaaacc gatgatgcgc 2220
gcgtcgtgc tgtacaacct gcacgaggcc ggacgaagcg cgggtgtgaa ggtggacccg 2280
tccccttttc aggaagtgtc ctcatccaag tacggcctgg tgcgcattct caaggtcatg 2340
aacglgagcg cggagagcaa gaagtgggtg gctgacccgg caaaccgctg gtgccaccg 2400
cctgggtcgt ggatctgccc cgggcagtac ccgcggcgga aggagatcca ggagatgctg 2460
gcgcaccgcg tcccctttga ccatgtgaac agcttcagtc ggaaaaagge cgggtcttat 2520
catgaagaat acatgcgccg gatgcgtgaa gacgaggacc gatga 2565

```

<210> 38
 <211> 854
 <212> PRT
 <213> Leishmania brucei

<400> 38

Met Gly Lys Lys Lys Ala Ile Pro Ser Gly Ser Val Gly Pro Ala Thr
1 5 10 15

Thr Thr Ser Arg Glu Ala Pro Gly Lys Asp Glu Gly Ala Ser Gln Pro
20 25 30

Ala Lys Thr Ala Ala Leu Pro Val Lys Pro Phe Val Leu Pro Asn Thr
35 40 45

Leu Thr Asp Glu Glu Glu Phe Val Gly Ile Phe Pro Cys Pro Phe Trp
50 55 60

Pro Val Arg Phe Val Ile Thr Val Met Ala Leu Val Leu Leu Gly Ala
65 70 75 80

Ser Cys Ile Arg Ala Phe Thr Ile Arg Met Leu Ser Val Gln Leu Tyr
85 90 95

Gly Tyr Ile Ile His Glu Phe Asp Pro Trp Phe Asn Tyr Arg Ala Ala
100 105 110

Glu Tyr Met Ser Ala His Gly Trp Ser Ala Phe Phe Ser Trp Phe Asp
115 120 125

Tyr Met Ser Trp Tyr Pro Leu Gly Arg Pro Val Gly Thr Thr Thr Tyr
130 135 140

Pro Gly Leu Gln Leu Thr Ala Val Ala Ile His Arg Ala Leu Ala Ala
145 150 155 160

Ala Gly Val Pro Met Ser Leu Asn Asn Val Cys Val Leu Ile Pro Ala
165 170 175

Trp Tyr Gly Ala Ile Ala Thr Ala Ile Leu Ala Leu Cys Ala Tyr Glu
180 185 190

[0030]

Val	Ser	Arg	Ser	Met	Val	Ala	Ala	Ala	Val	Ala	Ala	Leu	Ser	Phe	Ser	195	200	205
Ile	Ile	Pro	Ala	His	Leu	Met	Arg	Ser	Met	Ala	Gly	Glu	Phe	Asp	Asn	210	215	220
Glu	Cys	Ile	Ala	Val	Ala	Ala	Met	Leu	Leu	Thr	Phe	Tyr	Leu	Trp	Val	225	230	235
Arg	Ser	Leu	Arg	Thr	Arg	Cys	Ser	Trp	Pro	Ile	Gly	Ile	Leu	Thr	Gly	245	250	255
Ile	Ala	Tyr	Gly	Tyr	Met	Val	Ala	Ala	Trp	Gly	Gly	Tyr	Ile	Phe	Val	260	265	270
Leu	Asn	Met	Val	Ala	Met	His	Ala	Gly	Ile	Ser	Ser	Met	Val	Asp	Trp	275	280	285
Ala	Arg	Asn	Thr	Tyr	Asn	Pro	Ser	Leu	Leu	Arg	Ala	Tyr	Ala	Leu	Phe	290	295	300
Tyr	Val	Val	Gly	Thr	Ala	Ile	Ala	Thr	Arg	Val	Pro	Pro	Val	Gly	Met	305	310	315
Ser	Pro	Phe	Arg	Ser	Leu	Glu	Gln	Leu	Gly	Ala	Leu	Ala	Val	Leu	Leu	325	330	335
Phe	Leu	Cys	Gly	Leu	Gln	Ala	Cys	Glu	Val	Phe	Arg	Ala	Arg	Ala	Asp	340	345	350
Val	Glu	Val	Arg	Ser	Arg	Ala	Asn	Phe	Lys	Ile	Arg	Met	Arg	Ala	Phe	355	360	365
Ser	Val	Met	Ala	Gly	Val	Gly	Ala	Leu	Ala	Ile	Ala	Val	Leu	Ser	Pro	370	375	380
Thr	Gly	Tyr	Phe	Gly	Pro	Leu	Thr	Ala	Arg	Val	Arg	Ala	Leu	Phe	Met	385	390	395
Glu	His	Thr	Arg	Thr	Gly	Asn	Pro	Leu	Val	Asp	Ser	Val	Ala	Glu	His	405	410	415
His	Pro	Ala	Ser	Pro	Glu	Ala	Met	Trp	Thr	Phe	Leu	His	Val	Cys	Gly	420	425	430
Val	Thr	Trp	Gly	Leu	Gly	Ser	Ile	Val	Leu	Leu	Val	Ser	Leu	Leu	Val	435	440	445
Asp	Tyr	Ser	Ser	Ala	Lys	Leu	Phe	Trp	Leu	Met	Asn	Ser	Gly	Ala	Val	450	455	460
Tyr	Tyr	Phe	Ser	Thr	Arg	Met	Ser	Arg	Leu	Leu	Leu	Leu	Thr	Gly	Pro	465	470	475
Ala	Ala	Cys	Leu	Ser	Thr	Gly	Cys	Phe	Val	Gly	Thr	Leu	Leu	Glu	Ala	485	490	495
Ala	Ile	Gln	Phe	Thr	Phe	Trp	Ser	Ser	Asp	Ala	Thr	Lys	Ala	Lys	Lys	500	505	510

[0031]

Gln Gln Glu Thr Gln Leu His Gln Lys Gly Ala Arg Lys His Ser Asp
 515 520 525
 Arg Ser Asn Ser Lys Asn Ala Leu Thr Val Arg Thr Leu Gly Asp Val
 530 535 540
 Leu Arg Ser Thr Ser Leu Ala Trp Gly His Arg Met Val Leu Cys Phe
 545 550 555 560
 Ala Met Trp Ala Leu Val Ile Thr Val Ala Val Cys Leu Leu Gly Ser
 565 570 575
 Asp Phe Thr Ser His Ala Thr Met Phe Ala Arg Gln Thr Ser Asn Pro
 580 585 590
 Leu Ile Val Phe Ala Thr Val Leu Arg Asp Arg Ala Thr Gly Lys Pro
 595 600 605
 Thr Gln Val Leu Val Asp Asp Tyr Leu Arg Ser Tyr Leu Trp Leu Arg
 610 615 620
 Asp Asn Thr Pro Arg Asn Ala Arg Val Leu Ser Trp Trp Asp Tyr Gly
 625 630 635 640
 Tyr Gln Ile Thr Gly Ile Gly Asn Arg Thr Ser Leu Ala Asp Gly Asn
 645 650 655
 Thr Trp Asn His Glu His Ile Ala Thr Ile Gly Lys Met Leu Thr Ser
 660 665 670
 Pro Val Ala Glu Ala His Ser Leu Val Arg His Met Ala Asp Tyr Val
 675 680 685
 Leu Ile Trp Ala Gly Gln Gly Gly Asp Leu Met Lys Ser Pro His Met
 690 695 700
 Ala Arg Ile Gly Asn Ser Val Tyr His Asp Ile Cys Pro Asn Asp Pro
 705 710 715 720
 Leu Cys Gln His Phe Gly Phe Tyr Lys Asn Asp Arg Asn Arg Pro Lys
 725 730 735
 Pro Met Met Arg Ala Ser Leu Leu Tyr Asn Leu His Glu Ala Gly Arg
 740 745 750
 Ser Ala Gly Val Lys Val Asp Pro Ser Leu Phe Gln Glu Val Tyr Ser
 755 760 765
 Ser Lys Tyr Gly Leu Val Arg Ile Phe Lys Val Met Asn Val Ser Ala
 770 775 780
 Glu Ser Lys Lys Trp Val Ala Asp Pro Ala Asn Arg Val Cys His Pro
 785 790 795 800
 Pro Gly Ser Trp Ile Cys Pro Gly Gln Tyr Pro Pro Ala Lys Glu Ile
 805 810 815
 Gln Glu Met Leu Ala His Arg Val Pro Phe Asp His Val Asn Ser Phe
 820 825 830

[0032]

Ser Arg Lys Lys Ala Gly Ser Tyr His Glu Glu Tyr Met Arg Arg Met
835 840 845

Arg Glu Glu Gln Asp Arg
850

<210> 39
<211> 2592
<212> DNA
<213> Leishmania infantum

<400> 39
atgcccgccca agaatacaaca caaagggggc ggagacggca accccgatcc tacctccaca 60
cccgaagcgg cgtcgacaaa tgtgacaagc acaaacgacg gtgccgccgt cgattcttcc 120
gtgccaccgt ccggcgagac atacctcttt cattgccgcg ccgccccgta ctggaagcta 180
tcglacgcct tcaaagggtat catggccgtc ctgattctct gcgcccttcg ctggcgctac 240
caggttcgcc tgcctccgt tcagatttac ggatacctga tccacgagtt cgacccgtgg 300
ttcaactacc gcgtgccga gtacatgtcc acgcacggct ggctccgctt cttcagctgg 360
ttcgactaca tgagctggta cccgtgggc cgccctgttg gtccaccac gtaccgggc 420
ctgcagctca ctgcgtgc cattcaccgc gcgtggcgg ctgccggcat gccgatgtct 480
ctcaacaacg tgtcgtgct gatgccggc tggtttggcg ccacgcacac cgctactctg 540
gctctcatag cattcgaagt gagcgaatcc atctgtatgg cggcgtgggc cgcactctcc 600
ttctctatca tcccgccca cctgatgcgg tccatggcgg gtgagttcga caacgagtgc 660
attgccgtcg cagccatgct cctgacctc tactgctggg tgcgtcgtct gcgcacgcgg 720
tctcgtggc ccacgggtg cctcaccggt gtcgcctacg gctacatggt ggcggcgtgg 780
ggcggetaca ttttcgtgct caacatgggt gccatgcatg ccggcataac atcgatgggtg 840
gactgggccc gcaacaagta caaccgctg ctgctgcgtg catacacgct gttctacgta 900
gtcggcaccg ccacgcctg gtgcgtgcg ccaglgggga tgcgcctt caagtcgctg 960
gagcagctgg gtgcactgct ggtgcttgct ttctgtgtg gactgcaggc gtgcgaggtg 1020
tttcgcgcac gcgcgggtgt cgaggttcgc tctcgcgcga acttcaagat ccgcgtgcgc 1080
gtcttcagcg tgatggctgg cgtggctgcg ctgcgatcg cgggtgctggc accgacgggg 1140
tacttcgggc ccttttcggt ccgtgtgcgt gcgtgttcg tggagcacac gcgcactggc 1200
aatccgctgg tcgactcgtt cgccgagcac catcctgccg acgcgctcgc gtatctgaac 1260
tatittgcaca tcgtttattt tatgtggata tlcagcttcc cgggtgcagct catcctgccg 1320
agccgaaacc agtacgggt tctctttgtc tttgtctaca gttcatggc ctactacttc 1380
agcaccgcga tgggtcgctt gctcattctg gctggcccg cgccgtgcct cggcgcaagt 1440
gaggtaggtg gaacgtgat ggagtgggc tttcaacagc tgttctggga cgacggcatg 1500
cggaccgccg atatggtagc agccggtgac atgccttacc aaaaacaaga ccattgccagt 1560
agaggtgcag gcgcccga gaagcagcag aaacagaagc cgcgccaggt ttccgcgagg 1620
gactccagca ctacgagcga ggagcgtcct tacaggacac tgatccccgt cgacttcgc 1680
agggacgcgc aatatgaaccg ctggtcagcc ggaaagacaa acgcccctt catcgtggct 1740
ctcacgatcg gtgttctttt accgattgcg tttgtcttcc acttctcgtg cgtcagctca 1800
gcgtactect ttgtggccc gcgtatcgtg ttccagaagc agctgcgcac cggcgagcaa 1860
gtgatagtga aggactacct cgaggcctac gagtggctgc gcgacaacac gccagaggac 1920
gcgcgcattt tggcctggtg ggactacggc taccagatca caggcatcgg caaccgcacc 1980

[0033]

```

tcgtggccg atggcaacac ctggaaccac gagcacatcg ccaccatcgg caagatgctg 2040
acgtcgccc tggcggaggc gcactcgtg gtgcgccaca tggccgacta cgtcctaate 2100
tggtctgggc agagcggcga cctgatgaag tcaccgcaca tggcgcgcgt cggcaacagt 2160
gtgtaccacg acatctgccc ccacgaccgg ctgtgccagc aatttggtt ttacagaaat 2220
gattacagtc gccaacacc gatgatgcgg gcgtcgtgc tgtacaacct gcacgaggtc 2280
gggaaaacaa agggcgtgaa ggtggaccgg tctctctttc aggaggtgta ctggtccaag 2340
tacggcctgg tgcgcgtctt caaggtcatg aacgtgagcg aggagagcaa gaagtgggtt 2400
gtgtaccggg caaacccgtg gtgccaccgg cctgggtcgt ggatctgccc cgggcagtag 2460
ccgcggcga aggagatcca ggagatgctg gcacaccggg tccccctcga tcaggtgggc 2520
aaggacaaga aggacaagga ggcgtaccac aaggcgtaca tggaacgcag cagaacgctg 2580
ggtgaagttt ga 2592

```

<210> 40
 <211> 863
 <212> PRT
 <213> Leishmania infantum

<400> 40

Met Pro Ala Lys Asn Gln His Lys Gly Gly Gly Asp Gly Asn Pro Asp
 1 5 10 15

Pro Thr Ser Thr Pro Glu Ala Ala Ser Thr Asn Val Thr Ser Thr Asn
 20 25 30

Asp Gly Ala Ala Val Asp Ser Ser Val Pro Pro Ser Gly Glu Thr Tyr
 35 40 45

Leu Phe His Cys Arg Ala Ala Pro Tyr Ser Lys Leu Ser Tyr Ala Phe
 50 55 60

Lys Gly Ile Met Ala Val Leu Ile Leu Cys Ala Leu Arg Ser Ala Tyr
 65 70 75 80

Gln Val Arg Leu Leu Ser Val Gln Ile Tyr Gly Tyr Leu Ile His Glu
 85 90 95

Phe Asp Pro Trp Phe Asn Tyr Arg Ala Ala Glu Tyr Met Ser Thr His
 100 105 110

Gly Trp Ser Ala Phe Phe Ser Trp Phe Asp Tyr Met Ser Trp Tyr Pro
 115 120 125

Leu Gly Arg Pro Val Gly Ser Thr Thr Tyr Pro Gly Leu Gln Leu Thr
 130 135 140

Ala Val Ala Ile His Arg Ala Leu Ala Ala Ala Gly Met Pro Met Ser
 145 150 155 160

Leu Asn Asn Val Cys Val Leu Met Pro Ala Trp Phe Gly Ala Ile Ala
 165 170 175

Thr Ala Thr Leu Ala Leu Ile Ala Phe Glu Val Ser Glu Ser Ile Cys
 180 185 190

Met Ala Ala Trp Ala Ala Leu Ser Phe Ser Ile Ile Pro Ala His Leu

[0034]

195	200	205
Met Arg Ser Met Ala Gly Glu Phe Asp Asn Glu Cys Ile Ala Val Ala 210 215 220		
Ala Met Leu Leu Thr Phe Tyr Cys Trp Val Arg Ser Leu Arg Thr Arg 225 230 235 240		
Ser Ser Trp Pro Ile Gly Val Leu Thr Gly Val Ala Tyr Gly Tyr Met 245 250 255		
Val Ala Ala Trp Gly Gly Tyr Ile Phe Val Leu Asn Met Val Ala Met 260 265 270		
His Ala Gly Ile Ser Ser Met Val Asp Trp Ala Arg Asn Thr Tyr Asn 275 280 285		
Pro Ser Leu Leu Arg Ala Tyr Thr Leu Phe Tyr Val Val Gly Thr Ala 290 295 300		
Ile Ala Val Cys Val Pro Pro Val Gly Met Ser Pro Phe Lys Ser Leu 305 310 315 320		
Glu Gln Leu Gly Ala Leu Leu Val Leu Val Phe Leu Cys Gly Leu Gln 325 330 335		
Ala Cys Glu Val Phe Arg Ala Arg Ala Gly Val Glu Val Arg Ser Arg 340 345 350		
Ala Asn Phe Lys Ile Arg Val Arg Val Phe Ser Val Met Ala Gly Val 355 360 365		
Ala Ala Leu Ala Ile Ala Val Leu Ala Pro Thr Gly Tyr Phe Gly Pro 370 375 380		
Leu Ser Val Arg Val Arg Ala Leu Phe Val Glu His Thr Arg Thr Gly 385 390 395 400		
Asn Pro Leu Val Asp Ser Val Ala Glu His His Pro Ala Asp Ala Leu 405 410 415		
Ala Tyr Leu Asn Tyr Leu His Ile Val Tyr Phe Met Trp Ile Phe Ser 420 425 430		
Phe Pro Val Gln Leu Ile Leu Pro Ser Arg Asn Gln Tyr Ala Val Leu 435 440 445		
Phe Val Phe Val Tyr Ser Phe Met Ala Tyr Tyr Phe Ser Thr Arg Met 450 455 460		
Val Arg Leu Leu Ile Leu Ala Gly Pro Ala Ala Cys Leu Gly Ala Ser 465 470 475 480		
Glu Val Gly Gly Thr Leu Met Glu Trp Cys Phe Gln Gln Leu Phe Trp 485 490 495		
Asp Asp Gly Met Arg Thr Ala Asp Met Val Ala Ala Gly Asp Met Pro 500 505 510		
Tyr Gln Lys Gln Asp His Ala Ser Arg Gly Ala Gly Ala Arg Gln Lys		

[0035]

515	520	525
Gln Gln Lys Gln Lys Pro Arg Gln Val Phe Ala Arg Asp Ser Ser Thr 530 535 540		
Ser Ser Glu Glu Arg Pro Tyr Arg Thr Leu Ile Pro Val Asp Phe Arg 545 550 555 560		
Arg Asp Ala Gln Met Asn Arg Trp Ser Ala Gly Lys Thr Asn Ala Ala 565 570 575		
Leu Ile Val Ala Leu Thr Ile Gly Val Leu Leu Pro Ile Ala Phe Val 580 585 590		
Phe His Phe Ser Cys Val Ser Ser Ala Tyr Ser Phe Ala Gly Pro Arg 595 600 605		
Ile Val Phe Gln Thr Gln Leu Arg Thr Gly Glu Gln Val Ile Val Lys 610 615 620		
Asp Tyr Leu Glu Ala Tyr Glu Trp Leu Arg Asp Asn Thr Pro Glu Asp 625 630 635 640		
Ala Arg Ile Leu Ala Trp Trp Asp Tyr Gly Tyr Gln Ile Thr Gly Ile 645 650 655		
Gly Asn Arg Thr Ser Leu Ala Asp Gly Asn Thr Trp Asn His Glu His 660 665 670		
Ile Ala Thr Ile Gly Lys Met Leu Thr Ser Pro Val Ala Glu Ala His 675 680 685		
Ser Leu Val Arg His Met Ala Asp Tyr Val Leu Ile Trp Ala Gly Gln 690 695 700		
Ser Gly Asp Leu Met Lys Ser Pro His Met Ala Arg Ile Gly Asn Ser 705 710 715 720		
Val Tyr His Asp Ile Cys Pro His Asp Pro Leu Cys Gln Gln Phe Gly 725 730 735		
Phe Tyr Arg Asn Asp Tyr Ser Arg Pro Thr Pro Met Met Arg Ala Ser 740 745 750		
Leu Leu Tyr Asn Leu His Glu Val Gly Lys Thr Lys Gly Val Lys Val 755 760 765		
Asp Pro Ser Leu Phe Gln Glu Val Tyr Ser Ser Lys Tyr Gly Leu Val 770 775 780		
Arg Val Phe Lys Val Met Asn Val Ser Glu Glu Ser Lys Lys Trp Val 785 790 795 800		
Ala Asp Pro Ala Asn Arg Val Cys His Pro Pro Gly Ser Trp Ile Cys 805 810 815		
Pro Gly Gln Tyr Pro Pro Ala Lys Glu Ile Gln Glu Met Leu Ala His 820 825 830		
Arg Val Pro Phe Asp Gln Val Gly Lys Asp Lys Lys Asp Lys Glu Ala		

[0036]

835	840	845
Tyr His Lys Ala Tyr Met Glu Arg Ser Arg Thr Leu Gly Glu Val		
850	855	860
<210> 41 <211> 2355 <212> DNA <213> Leishmania infantum		
<400> 41 atgctgctct tgttcttctc ctttctgtac tgccctgaaa atgcatatgg cgttcgcctg 60 ctctccgttc agatttacgg atacctgac caccagttcg acccgtgggt caactaccgc 120 gctgccgagt acatgtccac gcacggctgg tccgccttct tcagctgggt cgactacatg 180 agctgggtacc cgttgggccc ccttgttggc tccaccacgt acccgggccc gcagctcact 240 gccgtcgcca ttcaccgcgc gctggcggct gccggcatgc cgatgtctct caacaacgtg 300 tgcgtgctga tgcggcgctg gtttggcgcc atgccaccg ctactatggc tggcatgacg 360 tatgagatga gcggatcagg cattaccgct gccatcgag cttttatctt tatgatctc 420 ccagcccacc tgatgcggtc catggcgggt gaggtcgaca acgagtgcac tgcctcgca 480 gccatgctcc tgaccttcta ctgtgggtg cgctcgctgc gcacgcggtc ctctgggccc 540 atcgggtgctc tcaccgggtg cgcctacggc tacatgggtg cggcgtgggg cggctacatt 600 ttctgtctca acatggttgc catgcatgcc ggcatatcat cgatgggtga ctgggcccgc 660 aacacgtaca acccgtcgct gctgcgtgca tacacgtgt tctacgtct cggcaccgcc 720 atcgccgtgt gcgtgccgcc agtggggatg tcgccctca agtcgctgga gcagctgggt 780 gcactgctgg tgcctgtctt cctgtgtgga ctgcaggcgt gcgaggtgtt tcgcgcacgc 840 gccggtgtcg aggttcgctc tcgcgcgaac ttcaagatcc gcgtgcgcgt cttcagcgtg 900 atggctggcg tggctgcgct tgcgctcgcg gtgctggcac cgacggggta cttcgggccc 960 ctttcggtcc gttgctgtc gctgttcgtg gaggacacgc gcactggcaa tccgtgggtc 1020 gactcggtcg ccgagcaccg caagacgagt ccggaagcgt acgcattttt tctggacttc 1080 acctaccga tgtggctgct cggcacagta ttgcagctgt tcggtgcagg gatggggtca 1140 cggaaggagg cgcggttgtt tatggggctg tactcactcg ccacctacta cttttcagat 1200 cgtatgtcac gcttgatggt gctggctggc ccggcggctg ccgctatggc ggcaggaatc 1260 ctgggcatcg tgtacgaatg gtgttggcgc cagctgaccg gctgggcac tcgagcctc 1320 tctgtgttgc gcagcaaagg caccgacggc ttgacaacc atataggaaa aactcagacg 1380 cagtccagca ccgcaaaccg taaccgtggt gtgcgagctc atgctgttgc cgctgtaaag 1440 tcgatgaagg cagctgtgga ctttcttccg ctggtgttgc gagctggcgt cgctgtggcc 1500 atccttgcgt tcaccgttgg tacgccgtac gtctctcagt tccaggttcg ttgtattcag 1560 tctgcgtact ctttgcgtgg ccgcgtatc gtgttccaga cgcagctgcg caccggcgag 1620 caagtgatag tgaaggacta cctcgaggcc tacgagtggc tgcgcgacaa caccgagag 1680 gacgcgcga ttttggcctg gtgggactac ggctaccaga tcacaggcat cggcaaccgc 1740 acctcgctgg ccgatggcaa caccctggaac caccgacaca tcgccaccat cggcaagatg 1800 ctgacgtcgc ccgtggcggg ggcgcacatg ctgggtgcgc acatggccga ctacgtccta 1860 atctgggctg ggcagagcgg cgacctgatg aagtcaccgc acatggcgcg catcggaac 1920 agtgtglacc acgacatctg cccccacgac ccgtgtgccc agcaatttgg cttttacaga 1980 aatgattaca gtgcgccaac accgatgatg cgggcgtcgc tgctgtacaa cctgcacgag 2040		

[0037]

```

gtcgggaaaa caaagggcgt gaaggtggac ccgtctctct ttcaggaggt gtactcgtcc 2100
aagtagggcc tggtagcgct cttcaaggtc atgaacgtga gcgaggagag caagaagtgg 2160
gttgctgacc cggcaaaccg cgtgtgccac ccgcctgggt cgtggatctg ccccgggcag 2220
taccgcggcg cgaaggagat ccaggagatg ctggcacacc gcgtccctt cgatcagatg 2280
ggcaaggaca agaaggacaa ggaggcgtac cacaaggcgt acatggaaaa gtcgaagaag 2340
gtagtcgagt tciga 2355

```

```

<210> 42
<211> 784
<212> PRT
<213> Leishmania infantum

```

```

<400> 42

```

```

Met Leu Leu Leu Phe Phe Ser Phe Leu Tyr Cys Leu Lys Asn Ala Tyr
1 5 10 15

```

```

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

```

```

Phe Asp Pro Trp Phe Asn Tyr Arg Ala Ala Glu Tyr Met Ser Thr His
35 40 45

```

```

Gly Trp Ser Ala Phe Phe Ser Trp Phe Asp Tyr Met Ser Trp Tyr Pro
50 55 60

```

```

Leu Gly Arg Pro Val Gly Ser Thr Thr Tyr Pro Gly Leu Gln Leu Thr
65 70 75 80

```

```

Ala Val Ala Ile His Arg Ala Leu Ala Ala Ala Gly Met Pro Met Ser
85 90 95

```

```

Leu Asn Asn Val Cys Val Leu Met Pro Ala Trp Phe Gly Ala Ile Ala
100 105 110

```

```

Thr Ala Thr Met Ala Gly Met Thr Tyr Glu Met Ser Gly Ser Gly Ile
115 120 125

```

```

Thr Ala Ala Ile Ala Ala Phe Ile Phe Met Ile Leu Pro Ala His Leu
130 135 140

```

```

Met Arg Ser Met Ala Gly Glu Phe Asp Asn Glu Cys Ile Ala Val Ala
145 150 155 160

```

```

Ala Met Leu Leu Thr Phe Tyr Cys Trp Val Arg Ser Leu Arg Thr Arg
165 170 175

```

```

Ser Ser Trp Pro Ile Gly Val Leu Thr Gly Val Ala Tyr Gly Tyr Met
180 185 190

```

```

Val Ala Ala Trp Gly Gly Tyr Ile Phe Val Leu Asn Met Val Ala Met
195 200 205

```

```

His Ala Gly Ile Ser Ser Met Val Asp Trp Ala Arg Asn Thr Tyr Asn
210 215 220

```

```

Pro Ser Leu Leu Arg Ala Tyr Thr Leu Phe Tyr Val Val Gly Thr Ala
225 230 235 240

```

[0038]

Ile Ala Val Cys Val Pro Pro Val Gly Met Ser Pro Phe Lys Ser Leu
 245 250 255
 Glu Gln Leu Gly Ala Leu Leu Val Leu Val Phe Leu Cys Gly Leu Gln
 260 265 270
 Ala Cys Glu Val Phe Arg Ala Arg Ala Gly Val Glu Val Arg Ser Arg
 275 280 285
 Ala Asn Phe Lys Ile Arg Val Arg Val Phe Ser Val Met Ala Gly Val
 290 295 300
 Ala Ala Leu Ala Ile Ala Val Leu Ala Pro Thr Gly Tyr Phe Gly Pro
 305 310 315 320
 Leu Ser Val Arg Val Arg Ala Leu Phe Val Glu His Thr Arg Thr Gly
 325 330 335
 Asn Pro Leu Val Asp Ser Val Ala Glu His Arg Lys Thr Ser Pro Glu
 340 345 350
 Ala Tyr Ala Phe Phe Leu Asp Phe Thr Tyr Pro Met Trp Leu Leu Gly
 355 360 365
 Thr Val Leu Gln Leu Phe Gly Ala Gly Met Gly Ser Arg Lys Glu Ala
 370 375 380
 Arg Leu Phe Met Gly Leu Tyr Ser Leu Ala Thr Tyr Tyr Phe Ser Asp
 385 390 395 400
 Arg Met Ser Arg Leu Met Val Leu Ala Gly Pro Ala Ala Ala Ala Met
 405 410 415
 Ala Ala Gly Ile Leu Gly Ile Val Tyr Glu Trp Cys Trp Ala Gln Leu
 420 425 430
 Thr Gly Trp Ala Ser Pro Ser Leu Ser Ala Val Gly Ser Lys Gly Thr
 435 440 445
 Asp Gly Phe Asp Asn His Ile Gly Lys Thr Gln Thr Gln Ser Ser Thr
 450 455 460
 Ala Asn Arg Asn Arg Gly Val Arg Ala His Ala Val Ala Ala Val Lys
 465 470 475 480
 Ser Met Lys Ala Ala Val Asp Leu Leu Pro Leu Val Leu Arg Ala Gly
 485 490 495
 Val Ala Val Ala Ile Leu Ala Val Thr Val Gly Thr Pro Tyr Val Ser
 500 505 510
 Gln Phe Gln Val Arg Cys Ile Gln Ser Ala Tyr Ser Phe Ala Gly Pro
 515 520 525
 Arg Ile Val Phe Gln Thr Gln Leu Arg Thr Gly Glu Gln Val Ile Val
 530 535 540
 Lys Asp Tyr Leu Glu Ala Tyr Glu Trp Leu Arg Asp Asn Thr Pro Glu
 545 550 555 560

[0039]

Asp Ala Arg Ile Leu Ala Trp Trp Asp Tyr Gly Tyr Gln Ile Thr Gly
 565 570 575
 Ile Gly Asn Arg Thr Ser Leu Ala Asp Gly Asn Thr Trp Asn His Glu
 580 585 590
 His Ile Ala Thr Ile Gly Lys Met Leu Thr Ser Pro Val Ala Glu Ala
 595 600 605
 His Ser Leu Val Arg His Met Ala Asp Tyr Val Leu Ile Trp Ala Gly
 610 615 620
 Gln Ser Gly Asp Leu Met Lys Ser Pro His Met Ala Arg Ile Gly Asn
 625 630 635 640
 Ser Val Tyr His Asp Ile Cys Pro His Asp Pro Leu Cys Gln Gln Phe
 645 650 655
 Gly Phe Tyr Arg Asn Asp Tyr Ser Arg Pro Thr Pro Met Met Arg Ala
 660 665 670
 Ser Leu Leu Tyr Asn Leu His Glu Val Gly Lys Thr Lys Gly Val Lys
 675 680 685
 Val Asp Pro Ser Leu Phe Gln Glu Val Tyr Ser Ser Lys Tyr Gly Leu
 690 695 700
 Val Arg Val Phe Lys Val Met Asn Val Ser Glu Glu Ser Lys Lys Trp
 705 710 715 720
 Val Ala Asp Pro Ala Asn Arg Val Cys His Pro Pro Gly Ser Trp Ile
 725 730 735
 Cys Pro Gly Gln Tyr Pro Pro Ala Lys Glu Ile Gln Glu Met Leu Ala
 740 745 750
 His Arg Val Pro Phe Asp Gln Met Gly Lys Asp Lys Lys Asp Lys Glu
 755 760 765
 Ala Tyr His Lys Ala Tyr Met Glu Lys Ser Lys Lys Val Val Glu Phe
 770 775 780
 <210> 43
 <211> 2382
 <212> DNA
 <213> Leishmania infantum
 <400> 43
 atgtcctcgc agactcgtag catcatctac tectcctgct ttgcagtggc catggccatt 60
 gccctcccta tcgcgtacga catgcgcgic cgctccatcg gcgtgtacgg gtacctcttc 120
 cacagaagtg acccggtggt caactaccgc gctgccgagt acatgtccac gcacggctgg 180
 tccgecttct tcagctggtt cgactacatg agctgggtacc cgctggggcc cctgtttggc 240
 tccaccacgt acccgggcct gcagctcact gccgtcgcca ttcaccgcgc gctggcggtt 300
 gccggcatgc cgatgtctct caacaacglg tgcgtgtcga tgccggcatg gttttcactt 360
 gtctcttcag cgatgggtggc tctgttggcg catgagttga gcggcaatat ggcggtggcc 420
 agcatctcgt ctattttgtt tagtgtgatt ccagcccacc tgatgcggte catggcgggt 480
 gaggttcgaca acgagtgcac tgccgtcgca gccatgctcc tgaccttcta ctgctgggtg 540

[0040]

```

cgctcgctgc gcacgcggtc ctcgtagccc atcgggtgtcc tcaccgggtgt cgcctacggc    600
tacatgggtgg cggcgtgggg cggctacatt ttcgtgtcca acatggttgc catgcatgcc    660
ggcataatcat cgatggtgga ctgggccccgc aacacgtaca acccgtcgct gctgcgtgca    720
tacacgctgt tctacgtcgt cggcaccgcc atcgccgtgt gcgtgccgcc agtggggatg    780
tcgcccttca agtcgttga gcagctgggt gcaactgtgg tgcttctctt cattttcgggt    840
cagtctgtgt gtgaggccca gcgcagacga ttggaaatcg cgcgcttttc aaaggagggc    900
gtggcgctgc tcatccgcat ctacgcagcc ttcttcgttg gtatcgttgc cgtggccacc    960
attgccccgg ccggattctt caagccgctc tccctgcaag cgagcgcgat aatcactggc   1020
gtatctcgta ccggaacac actcgtagac actctgattg cgcaagacgc gtccaacctt   1080
ctcatagtgt ggcagctttt tctctttccc gtctttggtt gggtagccgg catgagcgcc   1140
ttccttacag agttggtccg gaattacacc tacacaaaga gtttcatgct gatgtacggc   1200
gtggttggtc tgtacttcgc cagccaatct gtccgaatga tggtagatgat ggcccccggt   1260
gcgtgcatct tcactgcctt ttgtttccgc tgggcaactgg actacctcct cgggtcgttg   1320
ttttgggctg agatccacc ttgctttgac accgacgcgc agcgcggggc gcagcaacag   1380
accgcggagg aggcggaggc agagaccaag cgtaaggagg aagagtacaa caccatgcag   1440
gtcaagaaga tgacgacgcg catgttcccc ttcatgtttt tgctcttact gtttcgtctt   1500
tcggggttca tcgaagatgt ggccggcgata tcgcgcgaga tggaggcacc gggatatagt   1560
tttcccagtg gacaggtgca gggcgtgtcg gaaaaaagg tggacgacta ctattcgggg   1620
tacctgtacc tcgcgcgaca caccagagag gacgcgcga ttttggcctg gtgggactac   1680
ggctaccaga tcacaggcat cggcaaccgc acctcgttgg ccgatggcaa cacctggaac   1740
cacgagcaca tcgccaccat cggcaagatg ctgacgtcgc ccgtggcgga ggcgcactcg   1800
ctggtagcgc acatggccga ctatgttctg atttttgccg gagacacgta cttttccgac   1860
ctgaatcgct caccgcacat ggccgcgcatc ggcaacagtg tgtaccgga catctgcccc   1920
cacgacccgc tgttagtctg gttcgtgita cagaaaagac cgaagctgc tgcagcgaag   1980
cgcagtcgac acgtcagcgt tgacgaacta gaggaggagg acaatgcaga gcacgtggta   2040
tacgagccgl catcactcat ggccaagtcc ctcatatata atctgcactc agcaggggtg   2100
gtgaaggggg tcacgtgaa tgagacgctc ttccagcagc tcttcacctc ggcgcaaggt   2160
ctcatagcga tcttcaaggt catgaacgig agcgaggaga gcaagaagtg ggttgcgtgac   2220
ccggcaaacc gcgtgtgcca cccgcctggg tcgtggatct gccccgggca gtaccgcccg   2280
gcgaaggaga tccaggagat gctggcacac caacacacca acttcaagga ctttcttgat   2340
gccatgaacg acttgagcgc ggagcaggcg ctgaacaagg ag                               2382

```

<210> 44
 <211> 794
 <212> PRT
 <213> Leishmania infantum

<400> 44

Met Ser Ser Gln Thr Arg Ser Ile Ile Tyr Ser Ser Cys Phe Ala Val
 1 5 10 15

Ala Met Ala Ile Ala Leu Pro Ile Ala Tyr Asp Met Arg Val Arg Ser
 20 25 30

Ile Gly Val Tyr Gly Tyr Leu Phe His Arg Ser Asp Pro Trp Phe Asn

[0041]

35	40	45
Tyr Arg Ala Ala Glu Tyr Met Ser Thr His Gly Trp Ser Ala Phe Phe 50 55 60		
Ser Trp Phe Asp Tyr Met Ser Trp Tyr Pro Leu Gly Arg Pro Val Gly 65 70 75 80		
Ser Thr Thr Tyr Pro Gly Leu Gln Leu Thr Ala Val Ala Ile His Arg 85 90 95		
Ala Leu Ala Ala Ala Gly Met Pro Met Ser Leu Asn Asn Val Cys Val 100 105 110		
Leu Met Pro Ala Trp Phe Ser Leu Val Ser Ser Ala Met Val Ala Leu 115 120 125		
Leu Ala His Glu Leu Ser Gly Asn Met Ala Val Ala Ser Ile Ser Ser 130 135 140		
Ile Leu Phe Ser Val Ile Pro Ala His Leu Met Arg Ser Met Ala Gly 145 150 155 160		
Glu Phe Asp Asn Glu Cys Ile Ala Val Ala Ala Met Leu Leu Thr Phe 165 170 175		
Tyr Cys Trp Val Arg Ser Leu Arg Thr Arg Ser Ser Trp Pro Ile Gly 180 185 190		
Val Leu Thr Gly Val Ala Tyr Gly Tyr Met Val Ala Ala Trp Gly Gly 195 200 205		
Tyr Ile Phe Val Leu Asn Met Val Ala Met His Ala Gly Ile Ser Ser 210 215 220		
Met Val Asp Trp Ala Arg Asn Thr Tyr Asn Pro Ser Leu Leu Arg Ala 225 230 235 240		
Tyr Thr Leu Phe Tyr Val Val Gly Thr Ala Ile Ala Val Cys Val Pro 245 250 255		
Pro Val Gly Met Ser Pro Phe Lys Ser Leu Glu Gln Leu Gly Ala Leu 260 265 270		
Leu Val Leu Leu Phe Ile Phe Gly Gln Ser Val Cys Glu Ala Gln Arg 275 280 285		
Arg Arg Leu Glu Ile Ala Arg Phe Ser Lys Glu Gly Val Ala Leu Leu 290 295 300		
Ile Arg Ile Tyr Ala Ala Phe Phe Val Gly Ile Val Ala Val Ala Thr 305 310 315 320		
Ile Ala Pro Ala Gly Phe Phe Lys Pro Leu Ser Leu Gln Ala Ser Ala 325 330 335		
Ile Ile Thr Gly Val Ser Arg Thr Gly Asn Thr Leu Val Asp Thr Leu 340 345 350		
Ile Ala Gln Asp Ala Ser Asn Leu Leu Ile Val Trp Gln Leu Phe Leu		

[0042]

355	360	365
Phe Pro Val Phe Gly Trp 370	Val Ala Gly Met Ser 375	Ala Phe Leu Thr Glu 380
Leu Val Arg Asn Tyr Thr 385	Tyr Thr Lys Ser Phe Met 390	Leu Met Tyr Gly 400
Val Val Gly Leu Tyr Phe 405	Ala Ser Gln Ser Val Arg 410	Met Met Val Met 415
Met Ala Pro Val Ala Cys 420	Ile Phe Thr Ala Leu Leu 425	Phe Arg Trp Ala 430
Leu Asp Tyr Leu Leu Gly 435	Ser Leu Phe Trp Ala Glu 440	Met Pro Pro Cys 445
Phe Asp Thr Asp Ala Gln 450	Arg Gly Arg Gln Gln Gln 455	Thr Ala Glu Glu 460
Ala Glu Ala Glu Thr Lys 465	Arg Lys Glu Glu Glu Tyr 470	Asn Thr Met Gln 475
Val Lys Lys Met Thr Thr 485	Arg Met Leu Pro Phe Met 490	Phe Leu Leu Leu 495
Leu Phe Arg Leu Ser Gly 500	Phe Ile Glu Asp Val Ala 505	Ala Ile Ser Arg 510
Glu Met Glu Ala Pro Gly 515	Ile Val Phe Pro Ser Gly 520	Gln Val Gln Gly 525
Val Ser Glu Lys Lys Val 530	Asp Asp Tyr Tyr Ser Gly 535	Tyr Leu Tyr Leu 540
Arg Asp Asn Thr Pro Glu 545	Asp Ala Arg Ile Leu Ala 550	Trp Trp Asp Tyr 555
Gly Tyr Gln Ile Thr Gly 565	Ile Gly Asn Arg Thr Ser 570	Leu Ala Asp Gly 575
Asn Thr Trp Asn His Glu 580	His Ile Ala Thr Ile Gly 585	Lys Met Leu Thr 590
Ser Pro Val Ala Glu Ala 595	His Ser Leu Val Arg His 600	Met Ala Asp Tyr 605
Val Leu Ile Phe Ala Gly 610	Asp Thr Tyr Phe Ser Asp 615	Leu Asn Arg Ser 620
Pro His Met Ala Arg Ile 625	Gly Asn Ser Val Tyr Arg 630	Asp Ile Cys Pro 635
His Asp Pro Leu Cys Ser 645	Arg Phe Val Leu Gln Lys 650	Arg Pro Lys Ala 655
Ala Ala Ala Lys Arg Ser 660	Arg His Val Ser Val Asp 665	Glu Leu Glu Glu 670
Glu Asp Asn Ala Glu His 680	Val Val Tyr Glu Pro Ser 685	Ser Leu Met Ala 690

[0043]

675	680	685
Lys Ser Leu Ile Tyr His Leu His Ser Ala Gly Val Val Lys Gly Val 690 695 700		
Thr Leu Asn Glu Thr Leu Phe Gln His Val Phe Thr Ser Ala Gln Gly 705 710 715 720		
Leu Ile Arg Ile Phe Lys Val Met Asn Val Ser Glu Glu Ser Lys Lys 725 730 735		
Trp Val Ala Asp Pro Ala Asn Arg Val Cys His Pro Pro Gly Ser Trp 740 745 750		
Ile Cys Pro Gly Gln Tyr Pro Pro Ala Lys Glu Ile Gln Glu Met Leu 755 760 765		
Ala His Gln His Thr Asn Phe Lys Asp Leu Leu Asp Ala Met Asn Asp 770 775 780		
Leu Glu Arg Glu Gln Ala Leu Asn Lys Glu 785 790		
<210> 45 <211> 2559 <212> DNA <213> Leishmania infantum		
<400> 45		
atgccggcca agaaccagca caaaggaggc ggagacggcg accccgaccc tacctccaca	60	
cctgcacggg agtcgacaaa agtgacgaac acaagcgacg gtgccgccgt cgattccacc	120	
ctgccaccgt ccgacgagac atacctcttc cattgccgag cggccccgta ctggaagctg	180	
tcttacgcct tcaaagggtat catgaccgtc ctgattctgt gcgccattcg ctggcgctac	240	
caggttcgcc tgatctccgt tcagatttac ggataccatga tccacgagtt cgaccctggg	300	
ttcaactacc gcgtgccga gtacatgtcc acgcaaggct ggtcgcctt ctccagctgg	360	
ttcgactaca tgagctggta cccgtgggc cgccccgtcg gctccaccac gtaccggggc	420	
ctgcagctca ctgccgtgc cattcaccgc gcactggcag ctgccggcat gccgatgtct	480	
ctcaacaacg tgtcgtgct gatgccagcg tggtttggcg ccacgctac cgtactctg	540	
gctctcatag cattcgaagt gagcgaatcg atctgtatgg ctgcgtgggc cgcactctcc	600	
ttctccatca tcccagccca cctgatgcgg tccatggcgg gtgagttcga caacgagtcg	660	
atgccgtcg cagccatgct cctcactttc tactgctggg tgcgtcgtct gcgcacggcg	720	
tcctcgtggc ccacgggtgt cctcaccggt gtgcctacg gctacatggc ggcggcgtgg	780	
ggcggtaca tttcgtgct caacatggtt gccatgcat cggcctatc atcgatggtg	840	
gactggggcc gcaacacgta caaccgtcg ctgctgcgtg catacacgct gtctacgtc	900	
gtgggcaccg ccacgcccgt gtgcgtgccg ccagtgggga tgcgcctt caagtcgctg	960	
gagcagctgg gtgcgtgct ggtgctgtc ttctgtgtg gactgcaggt gtgcgaggtg	1020	
ctgcgcgac gcgccgtgt cgaggttcgc tctcgcgca acttcaagat ccgcgtgcgc	1080	
gtcttcagcg tgatggctgg cgtggctgcg ctgcaatct cggctgtggc accgacgggg	1140	
tacttcgggc cctctcgggt ccgtgtgcgt gcgtgttcg tggagcacac gcgcaactggc	1200	
aatccgtgg tcgactcggg cgcgcgacac caccctgcgg acgcgtcgc gtatctgaac	1260	
tatttgcaca ttgttcaatt gatgtggata tgcagcttgc cgggtgcagct catccttcca	1320	

[0044]

```

agccgaaacc agtacgcggt tctctttgtc ttggtctaca gctttatggc ctactacttc 1380
agcacccgca tgggtgcgctt gctcattctg gctggccctg tggcgtgcct cggcgcaagc 1440
gaggtagggtg gaacgctgat ggagtgggtgc tttaacacgc tgttctggga caacggcatg 1500
cggaccgccc atattgtagc agccgggtgac atgccttacc aaaaagacga ccacaccagc 1560
agaggtgcag gcgcccagaca gaagcagcag aaacagaagc cgggccagggt ttccgcgagg 1620
ggctccagca ctagcagcga ggagcgtcct tacaggacac tgatccccgt cgacttccgc 1680
agggacgccc aatatgaaccg ctggtcggcc ggaaagacaa acgcccctt catcgtggct 1740
ctcacgatcg gtgttctttt accgcttgcg ttgtcttcc acctctcatg catcagctca 1800
gcgtactcct tcgctggccc gcgtatcgtg ttccagacgc agctgcacac cggcgagcaa 1860
gtgatagtga aggactacct cgaggcctac gagtggctgc gcgacagcac gccagaggac 1920
gcgcgcgctt tggcctgggtg ggactacggc taccagatca caggcatcgg caaccgcacc 1980
tcgctggccc atggcaacac ctggaaccac gagcacatcg ccacgatcgg caagatgctg 2040
acgtgccccg tggcggaggc gcactcgtcg gtgcgccaca tggccgacta cgtcctgac 2100
tgggctgggc agagcggcga cctgatgaag tcaccgcaca tggcgcgcgt cggcaacagc 2160
gtgtaccacg acatctgccc cgacgacccg ctgtgccagc aattcggtt tcacagaaat 2220
gactacagtc gcccaacgcc gatgatgcgg gcgtcgtgc tgtacaacct gcacgaggcc 2280
gggaaaacaa agggcgtgaa ggtgaacccg tccctcttcc aggaggtgta ctgctcgaag 2340
tacggcctgg tgcgcatctt caaggtcatg aacgtgagcg cggagagcaa aaagtgggtt 2400
gctgacccgg caaacccgt gtgccacccg cctgggtcgt ggatctgccc cgggcagtac 2460
ccgccggcga aggagatcca ggagatgctg gcacaccgcg tccccctcga tcagatggac 2520
aagcacaagc agcacaagga gacgcaccac aaggcgtag 2559

```

<210> 46
 <211> 852
 <212> PRT
 <213> Leishmania infantum

<400> 46

Met Pro Ala Lys Asn Gln His Lys Gly Gly Gly Asp Gly Asp Pro Asp
 1 5 10 15

Pro Thr Ser Thr Pro Ala Ala Glu Ser Thr Lys Val Thr Asn Thr Ser
 20 25 30

Asp Gly Ala Ala Val Asp Ser Thr Leu Pro Pro Ser Asp Glu Thr Tyr
 35 40 45

Leu Phe His Cys Arg Ala Ala Pro Tyr Ser Lys Leu Ser Tyr Ala Phe
 50 55 60

Lys Gly Ile Met Thr Val Leu Ile Leu Cys Ala Ile Arg Ser Ala Tyr
 65 70 75 80

Gln Val Arg Leu Ile Ser Val Gln Ile Tyr Gly Tyr Leu Ile His Glu
 85 90 95

Phe Asp Pro Trp Phe Asn Tyr Arg Ala Ala Glu Tyr Met Ser Thr His
 100 105 110

Gly Trp Ser Ala Phe Phe Ser Trp Phe Asp Tyr Met Ser Trp Tyr Pro

[0045]

115	120	125
Leu Gly Arg Pro Val Gly Ser Thr Thr Tyr Pro Gly Leu Gln Leu Thr 130 135 140		
Ala Val Ala Ile His Arg Ala Leu Ala Ala Ala Gly Met Pro Met Ser 145 150 155 160		
Leu Asn Asn Val Cys Val Leu Met Pro Ala Trp Phe Gly Ala Ile Ala 165 170 175		
Thr Ala Thr Leu Ala Leu Ile Ala Phe Glu Val Ser Glu Ser Ile Cys 180 185 190		
Met Ala Ala Trp Ala Ala Leu Ser Phe Ser Ile Ile Pro Ala His Leu 195 200 205		
Met Arg Ser Met Ala Gly Glu Phe Asp Asn Glu Cys Ile Ala Val Ala 210 215 220		
Ala Met Leu Leu Thr Phe Tyr Cys Trp Val Arg Ser Leu Arg Thr Arg 225 230 235 240		
Ser Ser Trp Pro Ile Gly Val Leu Thr Gly Val Ala Tyr Gly Tyr Met 245 250 255		
Ala Ala Ala Trp Gly Gly Tyr Ile Phe Val Leu Asn Met Val Ala Met 260 265 270		
His Ala Gly Ile Ser Ser Met Val Asp Trp Ala Arg Asn Thr Tyr Asn 275 280 285		
Pro Ser Leu Leu Arg Ala Tyr Thr Leu Phe Tyr Val Val Gly Thr Ala 290 295 300		
Ile Ala Val Cys Val Pro Pro Val Gly Met Ser Pro Phe Lys Ser Leu 305 310 315 320		
Glu Gln Leu Gly Ala Leu Leu Val Leu Val Phe Leu Cys Gly Leu Gln 325 330 335		
Val Cys Glu Val Leu Arg Ala Arg Ala Gly Val Glu Val Arg Ser Arg 340 345 350		
Ala Asn Phe Lys Ile Arg Val Arg Val Phe Ser Val Met Ala Gly Val 355 360 365		
Ala Ala Leu Ala Ile Ser Val Leu Ala Pro Thr Gly Tyr Phe Gly Pro 370 375 380		
Leu Ser Val Arg Val Arg Ala Leu Phe Val Glu His Thr Arg Thr Gly 385 390 395 400		
Asn Pro Leu Val Asp Ser Val Ala Glu His His Pro Ala Asp Ala Leu 405 410 415		
Ala Tyr Leu Asn Tyr Leu His Ile Val His Leu Met Trp Ile Cys Ser 420 425 430		
Leu Pro Val Gln Leu Ile Leu Pro Ser Arg Asn Gln Tyr Ala Val Leu		

[0046]

435	440	445
Phe Val Leu Val Tyr Ser 450	Phe Met Ala Tyr Tyr 455	Phe Ser Thr Arg Met 460
Val Arg Leu Leu Ile 465	Leu Ala Gly Pro Val 470	Ala Cys Leu Gly Ala Ser 475 480
Glu Val Gly Gly Thr 485	Leu Met Glu Trp Cys 490	Phe Gln Gln Leu Phe Trp 495
Asp Asn Gly Met Arg Thr 500	Ala Asp Met Val Ala 505	Ala Gly Asp Met Pro 510
Tyr Gln Lys Asp Asp His Thr 515	Ser Arg Gly Ala Gly 520	Ala Arg Gln Lys 525
Gln Gln Lys Gln Lys Pro 530	Gly Gln Val Ser Ala 535	Arg Gly Ser Ser Thr 540
Ser Ser Glu Glu Arg Pro 545 550	Tyr Arg Thr Leu Ile 555	Pro Val Asp Phe Arg 560
Arg Asp Ala Gln Met Asn Arg Trp Ser 565	Ala Gly Lys Thr Asn Ala Ala 570 575	
Leu Ile Val Ala Leu Thr Ile 580	Gly Val Leu Leu Pro Leu 585 590	Ala Phe Val
Phe His Leu Ser Cys Ile Ser 595	Ser Ala Tyr Ser Phe 600 605	Ala Gly Pro Arg
Ile Val Phe Gln Thr Gln 610	Leu His Thr Gly Glu 615 620	Gln Val Ile Val Lys
Asp Tyr Leu Glu Ala Tyr 625 630	Glu Trp Leu Arg Asp 635	Ser Thr Pro Glu Asp 640
Ala Arg Val Leu Ala Trp Trp Asp Tyr 645	Gly Tyr Gln Ile Thr Gly 650 655	Ile
Gly Asn Arg Thr Ser Leu Ala Asp 660	Gly Asn Thr Trp Asn His 665 670	Glu His
Ile Ala Thr Ile Gly Lys Met 675	Leu Thr Ser Pro Val 680 685	Ala Glu Ala His
Ser Leu Val Arg His Met 690	Ala Asp Tyr Val Leu 695 700	Ile Trp Ala Gly Gln
Ser Gly Asp Leu Met Lys 705 710	Ser Pro His Met Ala 715	Arg Ile Gly Asn Ser 720
Val Tyr His Asp Ile Cys Pro Asp Asp 725	Pro Leu Cys Gln Gln 730 735	Phe Gly
Phe His Arg Asn Asp Tyr Ser Arg 740	Pro Thr Pro Met Met 745 750	Arg Ala Ser
Leu Leu Tyr Asn Leu His Glu Ala Gly Lys Thr Lys Gly Val Lys Val		

[0047]

755	760	765
Asn Pro Ser Leu Phe Gln Glu Val Tyr Ser Ser Lys Tyr Gly Leu Val 770 775 780		
Arg Ile Phe Lys Val Met Asn Val Ser Ala Glu Ser Lys Lys Trp Val 785 790 795 800		
Ala Asp Pro Ala Asn Arg Val Cys His Pro Pro Gly Ser Trp Ile Cys 805 810 815		
Pro Gly Gln Tyr Pro Pro Ala Lys Glu Ile Gln Glu Met Leu Ala His 820 825 830		
Arg Val Pro Phe Asp Gln Met Asp Lys His Lys Gln His Lys Glu Thr 835 840 845		
His His Lys Ala 850		
<210> 47 <211> 2322 <212> DNA <213> Leishmania infantum		
<400> 47		
atgctgctct tgttcttctc ctttctgtac tgcctaaaaa atgcgtatgg ccttcgcatg	60	
atctccgttc agatttacgg atacctgac caccagttcg acccgtggtt caactaccgc	120	
gctgccgagt acatgtccac gcacggcctg tccgccttct tcagctggtt cgactacatg	180	
agctgggtacc cgtggggcgc ccccgteggc tccaccacgt acccgggcct gcagctcact	240	
gccgtcgcca ttcaccgcgc actggcggct gccggcatgc cgatgtctct caacaacgtg	300	
tgcgtgctga tgcacgcgtg gtttggcgcc atcgtaccg ctactctggc tctcatgacg	360	
tatgagatga gcggatcagg cattgccgt gccattgcag cttttatctt ctccatcatc	420	
ccagcccacc tgatgcggtc catggcgggt gatttcgaca acgagtgcac cgcctcgca	480	
gccatgctcc tcactttcta ctgctgggtg cgtctgctgc gcacgggtc ctctggcccc	540	
atcgggtgcc tcaccgggtg cgcctacggc tacatggcgg cggcgtgggg cggctacatt	600	
ttcgtgctca acatggttgc catgcatgcc ggcatatcat cgatggtgga ctgggccccg	660	
aacacgtaca acccgtcgct gctgcgtgca tacacgctgt tctacgtcgt gggcaccgcc	720	
atcgccgtgt gcgtgccgcc agtggggatg tcgcccttca agtcgctgga gcagctgggt	780	
gcgtgctgg tcttgtctt cctgtgtgga ctgcagggtg gcgaggtgct gcgcgcacgc	840	
gccgggtgct aggttcgctc tcgcgcgaac tccaagatcc gcgtgcgcgt cttcagcgtg	900	
atggctggcg tggctgcgct tgcaatctcg gtgctggcac cgacggggta cttcgggccc	960	
ctctcggtcc gtgtgcgtgc gctgttcgtg gagcacacgc gcactggcaa tccgctggtc	1020	
gactcggctc ccgagcaccg catgacgagc ccgaaggcgt acgcattttt tctggacttc	1080	
acctaccggg tttggctgct cggcacagla ttgcagttgt taggtgcatt catggggtcg	1140	
cgaaggagg cgcggttggt tatgggtctg cactcactcg ccacctacta ctttcagat	1200	
cgtatgtcac gcttgatagt gctggcaggc cccgcggctg ccgctatgac ggcaggaatc	1260	
ctgggccttg tgtatgagt gtgttggcgc cagctgactg gctgggcgtc tccgggcctc	1320	
tctgctgctg gcagcggagg catggacgac ttgacaaca agcgaggcca aactcagata	1380	
cagtcagca ccgcaaacgc caaccgtggg gtgcgtgctc atgctattgc cgtgttaaag	1440	

[0048]

```

tcgatcaagg caggtgtgaa ccttcttcct ctgggtgttg gagtcggcgt cgctgtggct 1500
atccttgtctg tcaccgttgg tacgccgtac gtctcgcagt tccaggctcg ttgtattcag 1560
tgtgcgtact ccttcgctgg ccgcgcgcac gtgttcagg cgcagctgca caccggcgag 1620
caagtgatag tgaaggacta cctcgaggcc tacgagtgcc tgcgcgacag cacgccagag 1680
gaocgcgcgcg ttttggcctg gtgggactac ggctaccaga tcacaggcat cggcaaccgc 1740
acctcgtctg ccgatggcaa cacctggaac cagcagcaca tcgccacgat cggcaagatg 1800
ctgacgtcgc ccgtggcggg ggcgactcgc ctgggtgcgc acatggccga ctacgtcctg 1860
atctgggctg ggcagagcgg cgacctgatg aagtcaccgc acatggcgcg catcggcaac 1920
agcgtgtacc acgacatctg ccccgacgac ccgctgtgcc agcaattcgg ctttcacaga 1980
aatgactaca gtgcaccaac gccgatgatg cgggcgtcgc tgctgtacaa cctgcacgag 2040
gccgggaaaa caaagggcgt gaaggtgaac ccgtccctct ttcaggaggt gtactcgtcg 2100
aagtacggcc tgggtgcgcat cttcaaggtc atgaacgtga gcgcggagag caaaaagtgg 2160
gttgctgacc cggcaaaccc cggtgtgccac ccgcctgggt cgtggatctg ccccgggcag 2220
taccgcgccg cgaaggagat ccaggagatg ctggcacacc gcgtccctt cgatcagatg 2280
gacaagcaca agcagcaca ggagacgcac cacaaggcgt ag 2322

```

<210> 48
 <211> 773
 <212> PRT
 <213> Leishmania infantum

<400> 48

```

Met Leu Leu Leu Phe Phe Ser Phe Leu Tyr Cys Leu Lys Asn Ala Tyr
1           5           10           15

```

```

Gly Leu Arg Met Ile Ser Val Gln Ile Tyr Gly Tyr Leu Ile His Glu
20           25           30

```

```

Phe Asp Pro Trp Phe Asn Tyr Arg Ala Ala Glu Tyr Met Ser Thr His
35           40           45

```

```

Gly Trp Ser Ala Phe Phe Ser Trp Phe Asp Tyr Met Ser Trp Tyr Pro
50           55           60

```

```

Leu Gly Arg Pro Val Gly Ser Thr Thr Tyr Pro Gly Leu Gln Leu Thr
65           70           75           80

```

```

Ala Val Ala Ile His Arg Ala Leu Ala Ala Ala Gly Met Pro Met Ser
85           90           95

```

```

Leu Asn Asn Val Cys Val Leu Met Pro Ala Trp Phe Gly Ala Ile Ala
100          105          110

```

```

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

```

```

Ala Ala Ala Ile Ala Ala Phe Ile Phe Ser Ile Ile Pro Ala His Leu
130          135          140

```

```

Met Arg Ser Met Ala Gly Glu Phe Asp Asn Glu Cys Ile Ala Val Ala
145          150          155          160

```

```

Ala Met Leu Leu Thr Phe Tyr Cys Trp Val Arg Ser Leu Arg Thr Arg

```

[0049]

165							170							175						
Ser	Ser	Trp	Pro 180	Ile	Gly	Val	Leu	Thr 185	Gly	Val	Ala	Tyr	Gly 190	Tyr	Met					
Ala	Ala	Ala 195	Trp	Gly	Gly	Tyr	Ile 200	Phe	Val	Leu	Asn	Met 205	Val	Ala	Met					
His	Ala 210	Gly	Ile	Ser	Ser	Met 215	Val	Asp	Trp	Ala	Arg 220	Asn	Thr	Tyr	Asn					
Pro 225	Ser	Leu	Leu	Arg	Ala 230	Tyr	Thr	Leu	Phe	Tyr 235	Val	Val	Gly	Thr	Ala 240					
Ile	Ala	Val	Cys 245	Val	Pro	Pro	Val	Gly	Met 250	Ser	Pro	Phe	Lys	Ser 255	Leu					
Glu	Gln	Leu	Gly 260	Ala	Leu	Leu	Val	Leu 265	Val	Phe	Leu	Cys	Gly 270	Leu	Gln					
Val	Cys	Glu 275	Val	Leu	Arg	Ala	Arg 280	Ala	Gly	Val	Glu	Val 285	Arg	Ser	Arg					
Ala	Asn 290	Phe	Lys	Ile	Arg	Val 295	Arg	Val	Phe	Ser	Val 300	Met	Ala	Gly	Val					
Ala 305	Ala	Leu	Ala	Ile	Ser 310	Val	Leu	Ala	Pro	Thr 315	Gly	Tyr	Phe	Gly	Pro 320					
Leu	Ser	Val	Arg	Val 325	Arg	Ala	Leu	Phe	Val 330	Glu	His	Thr	Arg	Thr 335	Gly					
Asn	Pro	Leu	Val 340	Asp	Ser	Val	Ala	Glu 345	His	Arg	Met	Thr 350	Ser	Pro	Lys					
Ala	Tyr	Ala 355	Phe	Phe	Leu	Asp	Phe 360	Thr	Tyr	Pro	Val	Trp 365	Leu	Leu	Gly					
Thr 370	Val	Leu	Gln	Leu	Leu	Gly 375	Ala	Phe	Met	Gly	Ser 380	Arg	Lys	Glu	Ala					
Arg 385	Leu	Phe	Met	Gly	Leu 390	His	Ser	Leu	Ala	Thr 395	Tyr	Tyr	Phe	Ala	Asp 400					
Arg	Met	Ser	Arg	Leu 405	Ile	Val	Leu	Ala	Gly 410	Pro	Ala	Ala	Ala	Ala 415	Met					
Thr	Ala	Gly	Ile 420	Leu	Gly	Leu	Val	Tyr 425	Glu	Trp	Cys	Trp	Ala 430	Gln	Leu					
Thr	Gly	Trp 435	Ala	Ser	Pro	Gly	Leu 440	Ser	Ala	Ala	Gly	Ser 445	Gly	Gly	Met					
Asp 450	Asp	Phe	Asp	Asn	Lys	Arg 455	Gly	Gln	Thr	Gln	Ile 460	Gln	Ser	Ser	Thr					
Ala 465	Asn	Arg	Asn	Arg	Gly 470	Val	Arg	Ala	His	Ala 475	Ile	Ala	Ala	Val	Lys 480					
Ser	Ile	Lys	Ala	Gly	Val	Asn	Leu	Leu	Pro	Leu	Val	Leu	Arg	Val	Gly					

[0050]

485	490	495
Val Ala Val Ala Ile Leu Ala Val Thr Val Gly Thr Pro Tyr Val Ser 500 505 510		
Gln Phe Gln Ala Arg Cys Ile Gln Ser Ala Tyr Ser Phe Ala Gly Pro 515 520 525		
Arg Ile Val Phe Gln Ala Gln Leu His Thr Gly Glu Gln Val Ile Val 530 535 540		
Lys Asp Tyr Leu Glu Ala Tyr Glu Trp Leu Arg Asp Ser Thr Pro Glu 545 550 555 560		
Asp Ala Arg Val Leu Ala Trp Trp Asp Tyr Gly Tyr Gln Ile Thr Gly 565 570 575		
Ile Gly Asn Arg Thr Ser Leu Ala Asp Gly Asn Thr Trp Asn His Glu 580 585 590		
His Ile Ala Thr Ile Gly Lys Met Leu Thr Ser Pro Val Ala Glu Ala 595 600 605		
His Ser Leu Val Arg His Met Ala Asp Tyr Val Leu Ile Trp Ala Gly 610 615 620		
Gln Ser Gly Asp Leu Met Lys Ser Pro His Met Ala Arg Ile Gly Asn 625 630 635 640		
Ser Val Tyr His Asp Ile Cys Pro Asp Asp Pro Leu Cys Gln Gln Phe 645 650 655		
Gly Phe His Arg Asn Asp Tyr Ser Arg Pro Thr Pro Met Met Arg Ala 660 665 670		
Ser Leu Leu Tyr Asn Leu His Glu Ala Gly Lys Thr Lys Gly Val Lys 675 680 685		
Val Asn Pro Ser Leu Phe Gln Glu Val Tyr Ser Ser Lys Tyr Gly Leu 690 695 700		
Val Arg Ile Phe Lys Val Met Asn Val Ser Ala Glu Ser Lys Lys Trp 705 710 715 720		
Val Ala Asp Pro Ala Asn Arg Val Cys His Pro Pro Gly Ser Trp Ile 725 730 735		
Cys Pro Gly Gln Tyr Pro Pro Ala Lys Glu Ile Gln Glu Met Leu Ala 740 745 750		
His Arg Val Pro Phe Asp Gln Met Asp Lys His Lys Gln His Lys Glu 755 760 765		
Thr His His Lys Ala 770		

<210> 49
 <211> 2373
 <212> DNA
 <213> Leishmania infantum

[0051]

```

<400> 49
atgccctcgc aaactcgtac cctcatctac tctcctgct ttgcggtagc catggccatt    60
gccctcccta tcgcgtacga catgcgtgic cgcctccatcg gcgtgtacgg gtacctcttc    120
cacagcagtg acccgtgggt caactaccgc gctgccgagt acatgtccac gcacggctgg    180
tccgcttctt tcagctgggt cgactacatg agctgggtacc cgtggggcgg ccccgctggc    240
tccaccacgt acccgggcct gcagctcact gccgtcgcca ttcaccgcgc actggcggt    300
gccggcatgc cgtgtctct caacaacgtg tgcgtgtga tgccagcgtg gttttcactt    360
gtctcttcag cgaaggcggc actgtggcg catgagatga gcggcaatat ggcggtagcc    420
agcatctcgt ctatcttatt cagtgtgggt ccagcccacc tgatgcggtc catggcgggt    480
gagttcgaca acgagtgtat cgcgctcgca gccatgctcc tcaccttcta ctgctgggtg    540
cgctcgtcgc gcacgcggtc ctggtggccc atcgggtgccc tcaccgggtg cgcctacggc    600
tacatggcgg cggcgtgggg cggctacatt ttctgtctca acatggttgc catgcatgcc    660
ggcatatcat cgtatgtgga ctggggccc aacacgtaca acccgtcgtc gctgcgtgca    720
tacacgtgtg ttacgtcgt gggcaccgcc atcggcgtg gcgtgccgcc agtggggatg    780
tcgcccttca agtcgtgga gcagctgggt gcgtgtctgg tgcgtgtctt cattttcgtt    840
cagtctgtgt gtgaggccca gcgcagacga ttgggaatcg cgcgccttcc aaaggagggc    900
gtggcgctgc tcatccgat cgacgcagcc ttcttcgtcg gtatcgttgc cgtggccacc    960
attgccccgg ctggtattct caagccgctc tccctgcaag cgaacgcgat aatcactggc    1020
gtatctcgta ccggaacac actcgtagac attctgttgc cgaagacgc gtccaacct    1080
ctcatgggtg ggcagcttct tctctttccc ttcttaggtt ggggtggcgg catgagcgcc    1140
ttccttagag agttgatccg gaactacacc taacgaaga gtttcacct gatgtacggc    1200
gtggtcggtg tgtacttcgc cagccagtct gtccgaatga tggtagatgat ggcgccgltg    1260
gcgtgcatct ttactgcctt ctgttccgc tgggcactgg actacctcct cgggtctttg    1320
ttttgggctg agatgccacc ttcccttgac accgacgcac agcgtggggc gcagcaacag    1380
accgccgagg agtcggaggc agagaccaag cgtaaggagg aagagtaca caccatgcag    1440
gtcaagaaga tgcggtgctg catgttgcct ttcatgcgtg tgccttact gtttcgtctt    1500
tcgggggtca tcgaagatgt ggcggcgata tcgcgaaga tggaggcgcc gggtatagtt    1560
tttcccagtg aacaggtgca aggcgtgtcg gaaaaaagg tcgacgacta ctatgcgggg    1620
tacctgtatc tgcgcgacag cagccagag gacgcgcgcg ttttggcctg glgggactac    1680
ggctaccaga tcacaggcat cggcaaccgc acctcgttgc ccgatggcaa cacctggaac    1740
cacgagcaca tcgccagat cggcaagatg ctgacgtcgc ccgtggcgga ggcgcactcg    1800
ctgggtgcgc acatggcga ctatgttctg attctgctg gagacacata tttttccgac    1860
ctgaatcgtc caccgatgat ggccgcgcat ggcaacagcg tgtaccacga catctgcccc    1920
gacgacccac tttgtagtca gttcgtgttg cagaaaagac cgaagctgc tgcagcgaag    1980
cgcagtcggc acgtcagcgt tgacgcacia gaggaggatg aactgcaga gcataatgta    2040
tacgagccgt catcactcat agccaagtgc ctcatatata acctgcactc cacaggggtg    2100
gtgacggggg tcacgtgaa tgagacgtc ttccagcagc tcttcacctc accgcagggt    2160
ctcatgcgca tcttcaaggt catgaacgtg agcacggaga gcaaaaagtg ggttctgtac    2220
tcggcaaac ccgtgtgcca cccgcctggg tcgtggatct gcccgggca gtaccgcgcg    2280
gcgaaggaga tccaggagat gctggcacac caacacacca acttcaagga cttcttgat    2340
cccagaacga cttggagcgg gagcaggcgc tga    2373

```

[0052]

```

<210> 50
<211> 790
<212> PRT
<213> Leishmania infantum

<400> 50
Met Pro Ser Gln Thr Arg Ser Leu Ile Tyr Ser Ser Cys Phe Ala Val
1          5          10          15

Ala Met Ala Ile Ala Leu Pro Ile Ala Tyr Asp Met Arg Val Arg Ser
20          25          30

Ile Gly Val Tyr Gly Tyr Leu Phe His Ser Ser Asp Pro Trp Phe Asn
35          40          45

Tyr Arg Ala Ala Glu Tyr Met Ser Thr His Gly Trp Ser Ala Phe Phe
50          55          60

Ser Trp Phe Asp Tyr Met Ser Trp Tyr Pro Leu Gly Arg Pro Val Gly
65          70          75          80

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

Ala Leu Ala Ala Ala Gly Met Pro Met Ser Leu Asn Asn Val Cys Val
100         105         110

Leu Met Pro Ala Trp Phe Ser Leu Val Ser Ser Ala Met Ala Ala Leu
115         120         125

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

Ile Leu Phe Ser Val Val Pro Ala His Leu Met Arg Ser Met Ala Gly
145         150         155         160

Glu Phe Asp Asn Glu Cys Ile Ala Val Ala Ala Met Leu Leu Thr Phe
165         170         175

Tyr Cys Trp Val Arg Ser Leu Arg Thr Arg Ser Ser Trp Pro Ile Gly
180         185         190

Val Leu Thr Gly Val Ala Tyr Gly Tyr Met Ala Ala Ala Trp Gly Gly
195         200         205

Tyr Ile Phe Val Leu Asn Met Val Ala Met His Ala Gly Ile Ser Ser
210         215         220

Met Val Asp Trp Ala Arg Asn Thr Tyr Asn Pro Ser Leu Leu Arg Ala
225         230         235         240

Tyr Thr Leu Phe Tyr Val Val Gly Thr Ala Ile Ala Val Cys Val Pro
245         250         255

Pro Val Gly Met Ser Pro Phe Lys Ser Leu Glu Gln Leu Gly Ala Leu
260         265         270

Leu Val Leu Val Phe Ile Phe Gly Gln Ser Val Cys Glu Ala Gln Arg
275         280         285

```

[0053]

Arg Arg Leu Gly Ile Ala Arg Leu Ser Lys Glu Gly Val Ala Leu Leu
 290 295 300
 Ile Arg Ile Asp Ala Ala Phe Phe Val Gly Ile Val Ala Val Ala Thr
 305 310 315 320
 Ile Ala Pro Ala Gly Phe Phe Lys Pro Leu Ser Leu Gln Ala Asn Ala
 325 330 335
 Ile Ile Thr Gly Val Ser Arg Thr Gly Asn Thr Leu Val Asp Ile Leu
 340 345 350
 Leu Ala Gln Asp Ala Ser Asn Leu Leu Met Val Trp Gln Leu Phe Leu
 355 360 365
 Phe Pro Phe Leu Gly Trp Val Ala Gly Met Ser Ala Phe Leu Arg Glu
 370 375 380
 Leu Ile Arg Asn Tyr Thr Tyr Ala Lys Ser Phe Ile Leu Met Tyr Gly
 385 390 395 400
 Val Val Gly Met Tyr Phe Ala Ser Gln Ser Val Arg Met Met Val Met
 405 410 415
 Met Ala Pro Val Ala Cys Ile Phe Thr Ala Leu Leu Phe Arg Trp Ala
 420 425 430
 Leu Asp Tyr Leu Leu Gly Ser Leu Phe Trp Ala Glu Met Pro Pro Ser
 435 440 445
 Phe Asp Thr Asp Ala Gln Arg Gly Arg Gln Gln Gln Thr Ala Glu Glu
 450 455 460
 Ser Glu Ala Glu Thr Lys Arg Lys Glu Glu Glu Tyr Asn Thr Met Gln
 465 470 475 480
 Val Lys Lys Met Ser Val Arg Met Leu Pro Phe Met Leu Leu Leu Leu
 485 490 495
 Leu Phe Arg Leu Ser Gly Phe Ile Glu Asp Val Ala Ala Ile Ser Arg
 500 505 510
 Lys Met Glu Ala Pro Gly Ile Val Phe Pro Ser Glu Gln Val Gln Gly
 515 520 525
 Val Ser Glu Lys Lys Val Asp Asp Tyr Tyr Ala Gly Tyr Leu Tyr Leu
 530 535 540
 Arg Asp Ser Thr Pro Glu Asp Ala Arg Val Leu Ala Trp Trp Asp Tyr
 545 550 555 560
 Gly Tyr Gln Ile Thr Gly Ile Gly Asn Arg Thr Ser Leu Ala Asp Gly
 565 570 575
 Asn Thr Trp Asn His Glu His Ile Ala Thr Ile Gly Lys Met Leu Thr
 580 585 590
 Ser Pro Val Ala Glu Ala His Ser Leu Val Arg His Met Ala Asp Tyr
 595 600 605

[0054]

Val Leu Ile Ser Ala Gly Asp Thr Tyr Phe Ser Asp Leu Asn Arg Ser 610 615 620	
Pro Met Met Ala Arg Ile Gly Asn Ser Val Tyr His Asp Ile Cys Pro 625 630 635 640	
Asp Asp Pro Leu Cys Ser Gln Phe Val Leu Gln Lys Arg Pro Lys Ala 645 650 655	
Ala Ala Ala Lys Arg Ser Arg His Val Ser Val Asp Ala Leu Glu Glu 660 665 670	
Asp Asp Thr Ala Glu His Met Val Tyr Glu Pro Ser Ser Leu Ile Ala 675 680 685	
Lys Ser Leu Ile Tyr His Leu His Ser Thr Gly Val Val Thr Gly Val 690 695 700	
Thr Leu Asn Glu Thr Leu Phe Gln His Val Phe Thr Ser Pro Gln Gly 705 710 715 720	
Leu Met Arg Ile Phe Lys Val Met Asn Val Ser Thr Glu Ser Lys Lys 725 730 735	
Trp Val Ala Asp Ser Ala Asn Arg Val Cys His Pro Pro Gly Ser Trp 740 745 750	
Ile Cys Pro Gly Gln Tyr Pro Pro Ala Lys Glu Ile Gln Glu Met Leu 755 760 765	
Ala His Gln His Thr Asn Phe Lys Asp Leu Leu Asp Pro Arg Thr Thr 770 775 780	
Trp Ser Gly Ser Arg Arg 785 790	
<210> 51	
<211> 2574	
<212> DNA	
<213> Leishmania infantum	
<400> 51	
atgggcaagc ggaagggaaa ttcatgggc gactccggct ctgcggcaac cgcatctcgt 60	
gaagcttcgg cccaagccga agatgccgcc tcacaaacca agaccgcac tcaccggcg 120	
aaggtgattt tgctgccc aaacgtaaca gatgagaagg atttcacgg catctttccg 180	
tttctttct ggccagtaca ctctgtctc accgtgggtg cactcttctg cttagccgcc 240	
agctgcttc aagccttcac ggcttcgatg atctccgtc agatttacgg atacctgac 300	
cacgagttcg acccggtgtt caactaccgc gctgccgagt acatgtccac gcacggctgg 360	
tcgccttct teagctggtt cgactacatg agctgggtacc cgtggggccg ccccgtcggc 420	
tcaccacgt accggggcct gcagctcact gccgtcgcca ttcaccgcgc actggcggt 480	
gccggcatgc cgatgtctct caacaacgtg tgcgtgtga tgccagcgtg gtttggcgcc 540	
atcgctaccg ctactctggc gttttgcacc tacgaagcca gtgggtcgac agtagcgcg 600	
gccgtgccg cactctcctt ctccateatc ccagcccacc tgatggcggtc catggcgggt 660	
gagttcgaca acgagtgcat cgccgtcgca gccatgtcc tcactttcta ctgcgggtg 720	
cgctcgctgc gcacgggtc ctctggccc atcgggtgtc tcaccggtgt cgcctacggc 780	

[0055]

```

tacatggcgg cggcgtgggg cggctacatt ttcgtgctca acatgggtgc catgcatgcc 840
ggcatatcat cgatgggtga ctgggcccg c aacacgtaca acccgctcgt gctgcgtgca 900
tacacgtgtg tctacgtcgt gggcaccgcc atcgccgtgt gcgtgccgcc agtggggatg 960
tcgcccttca agtcgctgga gcagctgggt gcgctgctgg tgcttgtctt cctgtgtgga 1020
ctgcaggtgt gcgaggtgct gcgcgcacgc gccgggtgctg aggttcgctc tcgcgcgaac 1080
ttcaagatcc gcgtgcgcgt cltcagcgtg atggctggcg tggtgcgct tgcaatctcg 1140
gtgctggcac cgacggggtta cttcggggcc ctctcggtcc gtgtgcgtgc gctgttcgtg 1200
gagcacacgc gcactggcaa tccgctggtc gactcggctg ccgaacatca acccgccagc 1260
ccggaggcga tgtgggcttt tcttcacgtg tgcggcgtga catggggttt gggtccatt 1320
gtgcttctg tctcgacgtt cgtgcactac tccccgtcga aggtattctg gctactgaac 1380
tccggcgccg tgtactactt cagcactcgc atggctcggc tgctgttct ctccggcccc 1440
gtgcgtgic tgtccactgg cattttcgtc gggacaattc tggaagcagc cgtgcaactc 1500
agtttctggg acagtgatgc gacaaaggct aaaaagcagc agaagcaggc gcaacgccac 1560
cagagggggg ctggcaaggg cagcggccga gatgacgcta agaacgcaac gaccgcgcgc 1620
gcattttgcg acgtctttgc cggtagctct ctgtcttggg gccatcgcat ggtcctgtcc 1680
atcgctatgt gggctctcgt cagacaacc gcggtgagct tcttcagttc cgagttcgcg 1740
tcccactcaa caaagtttgc ggagcagtcg tcaaatecga tgattgttlt cgcggccgctc 1800
gtgcaaaacc gtgccacagg caagcctatg aacctattgg tggatgacta cctcaaggcc 1860
tacgagtggc tcgcgcacag cagccagag gacgcgcgcg ttttggcctg gtgggactac 1920
ggctaccaga tcacaggcat cggcaaccgc acctcgtcgg ccgatggcaa caccitggaac 1980
cagcagcaca tcgccacgat cggcaagatg ctgacgtcgc ccgtggcgga ggccgactcg 2040
ctgggtgcgc acatggccga ctacgtcctg atctgggctg ggcagagcgg cgacctgatg 2100
aagtcaccgc acatggcgcg catcggcaac agcgtgtacc acgacatctg ccccgacgac 2160
ccgctgtgcc agcaattcgg ctttcacaga aatgactaca gtcgcccac gccgatgatg 2220
cgggcgtcgc tgctgtacaa cctgcacgag gccgggaaaa gaaagggcgt gaagggtgaa 2280
ccgtccctct ttcaggaggt gtactcgtcg aagtacggcc tgggtgcgcat cttcaaggte 2340
atgaacgtga gcgcggagag caaaaagtgg gttgctgacc cggcaaacgc cgtgtgccac 2400
ccgcclgggt cgtggatctg cccggggcag tacccgccgg cgaaggagat ccaggagatg 2460
ctggcacacc gcgtccctt cgatcaggtg acaaacgccg atcggaaaaa caatgtcggg 2520
tcgtaccagg aggagtacat gcgccgcatg cgtgaaagcg agaaccgacg gtaa 2574

```

<210> 52
 <211> 857
 <212> PRT
 <213> Leishmania infantum

<400> 52

Met Gly Lys Arg Lys Gly Asn Ser Leu Gly Asp Ser Gly Ser Ala Ala
1 5 10 15

Thr Ala Ser Arg Glu Ala Ser Ala Gln Ala Glu Asp Ala Ala Ser Gln
20 25 30

Thr Lys Thr Ala Ser Pro Pro Ala Lys Val Ile Leu Leu Pro Lys Thr
35 40 45

[0056]

Leu Thr Asp Glu Lys Asp Phe Ile Gly Ile Phe Pro Phe Pro Phe Trp	50	55	60
Pro Val His Phe Val Leu Thr Val Val Ala Leu Phe Val Leu Ala Ala	65	70	75
Ser Cys Phe Gln Ala Phe Thr Val Arg Met Ile Ser Val Gln Ile Tyr	85	90	95
Gly Tyr Leu Ile His Glu Phe Asp Pro Trp Phe Asn Tyr Arg Ala Ala	100	105	110
Glu Tyr Met Ser Thr His Gly Trp Ser Ala Phe Phe Ser Trp Phe Asp	115	120	125
Tyr Met Ser Trp Tyr Pro Leu Gly Arg Pro Val Gly Ser Thr Thr Tyr	130	135	140
Pro Gly Leu Gln Leu Thr Ala Val Ala Ile His Arg Ala Leu Ala Ala	145	150	155
Ala Gly Met Pro Met Ser Leu Asn Asn Val Cys Val Leu Met Pro Ala	165	170	175
Trp Phe Gly Ala Ile Ala Thr Ala Thr Leu Ala Phe Cys Thr Tyr Glu	180	185	190
Ala Ser Gly Ser Thr Val Ala Ala Ala Ala Ala Ala Leu Ser Phe Ser	195	200	205
Ile Ile Pro Ala His Leu Met Arg Ser Met Ala Gly Glu Phe Asp Asn	210	215	220
Glu Cys Ile Ala Val Ala Ala Met Leu Leu Thr Phe Tyr Cys Trp Val	225	230	235
Arg Ser Leu Arg Thr Arg Ser Ser Trp Pro Ile Gly Val Leu Thr Gly	245	250	255
Val Ala Tyr Gly Tyr Met Ala Ala Ala Trp Gly Gly Tyr Ile Phe Val	260	265	270
Leu Asn Met Val Ala Met His Ala Gly Ile Ser Ser Met Val Asp Trp	275	280	285
Ala Arg Asn Thr Tyr Asn Pro Ser Leu Leu Arg Ala Tyr Thr Leu Phe	290	295	300
Tyr Val Val Gly Thr Ala Ile Ala Val Cys Val Pro Pro Val Gly Met	305	310	315
Ser Pro Phe Lys Ser Leu Glu Gln Leu Gly Ala Leu Leu Val Leu Val	325	330	335
Phe Leu Cys Gly Leu Gln Val Cys Glu Val Leu Arg Ala Arg Ala Gly	340	345	350
Val Glu Val Arg Ser Arg Ala Asn Phe Lys Ile Arg Val Arg Val Phe	355	360	365

[0057]

Ser Val Met Ala Gly Val Ala Ala Leu Ala Ile Ser Val Leu Ala Pro
 370 375 380
 Thr Gly Tyr Phe Gly Pro Leu Ser Val Arg Val Arg Ala Leu Phe Val
 385 390 395 400
 Glu His Thr Arg Thr Gly Asn Pro Leu Val Asp Ser Val Ala Glu His
 405 410 415
 Gln Pro Ala Ser Pro Glu Ala Met Trp Ala Phe Leu His Val Cys Gly
 420 425 430
 Val Thr Trp Gly Leu Gly Ser Ile Val Leu Ala Val Ser Thr Phe Val
 435 440 445
 His Tyr Ser Pro Ser Lys Val Phe Trp Leu Leu Asn Ser Gly Ala Val
 450 455 460
 Tyr Tyr Phe Ser Thr Arg Met Ala Arg Leu Leu Leu Leu Ser Gly Pro
 465 470 475 480
 Ala Ala Cys Leu Ser Thr Gly Ile Phe Val Gly Thr Ile Leu Glu Ala
 485 490 495
 Ala Val Gln Leu Ser Phe Trp Asp Ser Asp Ala Thr Lys Ala Lys Lys
 500 505 510
 Gln Gln Lys Gln Ala Gln Arg His Gln Arg Gly Ala Gly Lys Gly Ser
 515 520 525
 Gly Arg Asp Asp Ala Lys Asn Ala Thr Thr Ala Arg Ala Phe Cys Asp
 530 535 540
 Val Phe Ala Gly Ser Ser Leu Ala Trp Gly His Arg Met Val Leu Ser
 545 550 555 560
 Ile Ala Met Trp Ala Leu Val Thr Thr Thr Ala Val Ser Phe Phe Ser
 565 570 575
 Ser Glu Phe Ala Ser His Ser Thr Lys Phe Ala Glu Gln Ser Ser Asn
 580 585 590
 Pro Met Ile Val Phe Ala Ala Val Val Gln Asn Arg Ala Thr Gly Lys
 595 600 605
 Pro Met Asn Leu Leu Val Asp Asp Tyr Leu Lys Ala Tyr Glu Trp Leu
 610 615 620
 Arg Asp Ser Thr Pro Glu Asp Ala Arg Val Leu Ala Trp Trp Asp Tyr
 625 630 635 640
 Gly Tyr Gln Ile Thr Gly Ile Gly Asn Arg Thr Ser Leu Ala Asp Gly
 645 650 655
 Asn Thr Trp Asn His Glu His Ile Ala Thr Ile Gly Lys Met Leu Thr
 660 665 670
 Ser Pro Val Ala Glu Ala His Ser Leu Val Arg His Met Ala Asp Tyr
 675 680 685

[0058]

Val Leu Ile Trp Ala Gly Gln Ser Gly Asp Leu Met Lys Ser Pro His
690 695 700

Met Ala Arg Ile Gly Asn Ser Val Tyr His Asp Ile Cys Pro Asp Asp
705 710 715 720

Pro Leu Cys Gln Gln Phe Gly Phe His Arg Asn Asp Tyr Ser Arg Pro
725 730 735

Thr Pro Met Met Arg Ala Ser Leu Leu Tyr Asn Leu His Glu Ala Gly
740 745 750

Lys Arg Lys Gly Val Lys Val Asn Pro Ser Leu Phe Gln Glu Val Tyr
755 760 765

Ser Ser Lys Tyr Gly Leu Val Arg Ile Phe Lys Val Met Asn Val Ser
770 775 780

Ala Glu Ser Lys Lys Trp Val Ala Asp Pro Ala Asn Arg Val Cys His
785 790 795 800

Pro Pro Gly Ser Trp Ile Cys Pro Gly Gln Tyr Pro Pro Ala Lys Glu
805 810 815

Ile Gln Glu Met Leu Ala His Arg Val Pro Phe Asp Gln Val Thr Asn
820 825 830

Ala Asp Arg Lys Asn Asn Val Gly Ser Tyr Gln Glu Glu Tyr Met Arg
835 840 845

Arg Met Arg Glu Ser Glu Asn Arg Arg
850 855

<210> 53
<211> 2406
<212> DNA
<213> Trypanosoma brucei

<400> 53
atgacgaaag gtgggaaagt agctgtgact aagggtcag cacagagtga tgggtgctggt 60
gagggaggga tgagtaaggc caagtcattc actacgttcg tcgccactgg cgggtggttct 120
ttgcctgcct gggcgctaaa ggctgtaagc acgattgtga gtgcagtgat tcttatatac 180
tctgtccatc gtgcttacga tatacgactt acttctgtcc gtctttatgg tgagcttatt 240
cacgagttcg acccttggtt caattaccgt gcaacgcagt acctcagcga caacgggtgg 300
cgtgcttttt tccaatggta cgactacatg agctgggtacc cgttggccg accggtgggc 360
acaaccatct tccccggaat gcagcttacc ggtgtagcca ttcatcgtgt gctggaaatg 420
ctcgggcgag gtatgtccat caacaatata tgtgtgtaca ttctgcatg gttcggtagt 480
attgccactg tgttggtgc tctcattgcg tacgaatcat ctaattcgct cagtgtcatg 540
gcgtttactg cgtacttttt ttccatcgta cctgcacacc tgatgcgac aatggctggt 600
gaatttgaca atgagtgtgt tgcaatggcg gcgatgcttc tgacgttcta catgtgggta 660
cgatcggttac gcagctcaag ttctggccc attggcgctt tagctggtgt ggcatacggg 720
tacatggtgt ccacgtgggg tgggtatatt ttctgtctga acatggttagc cttccacgct 780
tctgtatgtg tactgcttga ttgggctcgt gggatataca gcgtcagttt gctgagggcg 840
tattcactgt tttttgtcat tggcactgcc ctgtcgattt gcgtaccgcc agtggagtgg 900

[0059]

```

acgcccttcc ggctcgtgga gcaactgaca gcattgtttg tcttcgtttt catgtgggca 960
ctccactact cggaatacct gcgtgagcgt gcccgagcgc ccattcactc ttctaaagea 1020
cttcagatcc gtgcccgcgt ttcatgggc accctctcct tgctgttgat tgtggcaagt 1080
ttgcttgcgc cgttcggatt cttcaaacct acagcgtacc gcgtccgtgc gttgttcgtg 1140
aaacatacgc gtaccggaaa tcccctcgtg gattctgttg ctgagcatcg gccgacgact 1200
gccggggcgt atctgcgcta ctctcatgtt tgttaccctc ttgggggtg cgggtgggctc 1260
tctatgttgg tattcatgaa aaaggaccgc tggcgcgcca ttgtttttct tgettcactt 1320
tccactgita cgatgtattt cagcggccgt atgtcgcgat tacttctgtt agcgggtccc 1380
gcagcaacgg ctctgcgcgg catgttcata ggggggcctt ttgatctggc gctgtcacag 1440
tttggtgatt tgcatagccc aaaagatgcc tccggcgatt ccgatcccgc gggagggtcg 1500
aagcgggcaa agggcaaagt tgtaatgag ccttccaaaa gagccatctt tagtcaccgc 1560
tggtttcaac gtttagtgca atcgttgcgc gtcccgtac gacgtggtat cgcggttggtg 1620
gtgctcgtat gtctcttcgc caatcccatg agacactcat tcgaaaagtc ctgcgagaaa 1680
atggcacatg cactttcact tccaaggatc attgccgtga ctgatctacc caatggagag 1740
agagtcctcg ccgatgatta ctacgtgtcg tacttgtggc tgcgaaacaa taeccttgaa 1800
gaigccccta ttctctcatg gtgggactac gggatatcaa tcaactggaat tggcaatcgc 1860
acaacccttg cggacggtaa cacatggagt cacaagcaca tagcaactat tgggaagatg 1920
cttacatccc ctgtgaagga gtcacatgct cttatacgcc atctcgtga ttatgtgctg 1980
atatgggccg gtgaggatcg gggcgattta cttaaatcgc cacacatggc tcggatagcc 2040
aacagtgtat atcgcgatat gtgttcagaa gacgatccta gatgcaggca gttcggcttt 2100
gaggggagtg acctcaataa gcctacgcct atgatgcagc ggtccctatt atacaatctg 2160
cacaggtttg gtacggatgg cggaagaca caactlgata agaacatgtt tcagctcgcc 2220
tacgtgtcaa agtatggtt ggtgaagatc tacaagtggt tgaatgtgag tgaagagagc 2280
aaggcgtggg tcgcagacct aaagaaccgc gtatgcgacc cgcccggatc ttggatatgc 2340
gccggccagt acccgccagc gaaggagatc caagacatgt tagcgaagcg gttccattac 2400
gaatga 2406

```

<210> 54
 <211> 801
 <212> PRT
 <213> Trypanosoma brucei

<400> 54

Met Thr Lys Gly Gly Lys Val Ala Val Thr Lys Gly Ser Ala Gln Ser
1 5 10 15

Asp Gly Ala Gly Glu Gly Gly Met Ser Lys Ala Lys Ser Ser Thr Thr
20 25 30

Phe Val Ala Thr Gly Gly Gly Ser Leu Pro Ala Trp Ala Leu Lys Ala
35 40 45

Val Ser Thr Ile Val Ser Ala Val Ile Leu Ile Tyr Ser Val His Arg
50 55 60

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

[0060]

His	Glu	Phe	Asp	Pro	Trp	Phe	Asn	Tyr	Arg	Ala	Thr	Gln	Tyr	Leu	Ser	85	90	95
Asp	Asn	Gly	Trp	Arg	Ala	Phe	Phe	Gln	Trp	Tyr	Asp	Tyr	Met	Ser	Trp	100	105	110
Tyr	Pro	Leu	Gly	Arg	Pro	Val	Gly	Thr	Thr	Ile	Phe	Pro	Gly	Met	Gln	115	120	125
Leu	Thr	Gly	Val	Ala	Ile	His	Arg	Val	Leu	Glu	Met	Leu	Gly	Arg	Gly	130	135	140
Met	Ser	Ile	Asn	Asn	Ile	Cys	Val	Tyr	Ile	Pro	Ala	Trp	Phe	Gly	Ser	145	150	155
Ile	Ala	Thr	Val	Leu	Ala	Ala	Leu	Ile	Ala	Tyr	Glu	Ser	Ser	Asn	Ser	165	170	175
Leu	Ser	Val	Met	Ala	Phe	Thr	Ala	Tyr	Phe	Phe	Ser	Ile	Val	Pro	Ala	180	185	190
His	Leu	Met	Arg	Ser	Met	Ala	Gly	Glu	Phe	Asp	Asn	Glu	Cys	Val	Ala	195	200	205
Met	Ala	Ala	Met	Leu	Leu	Thr	Phe	Tyr	Met	Trp	Val	Arg	Ser	Leu	Arg	210	215	220
Ser	Ser	Ser	Ser	Trp	Pro	Ile	Gly	Ala	Leu	Ala	Gly	Val	Ala	Tyr	Gly	225	230	235
Tyr	Met	Val	Ser	Thr	Trp	Gly	Gly	Tyr	Ile	Phe	Val	Leu	Asn	Met	Val	245	250	255
Ala	Phe	His	Ala	Ser	Val	Cys	Val	Leu	Leu	Asp	Trp	Ala	Arg	Gly	Ile	260	265	270
Tyr	Ser	Val	Ser	Leu	Leu	Arg	Ala	Tyr	Ser	Leu	Phe	Phe	Val	Ile	Gly	275	280	285
Thr	Ala	Leu	Ala	Ile	Cys	Val	Pro	Pro	Val	Glu	Trp	Thr	Pro	Phe	Arg	290	295	300
Ser	Leu	Glu	Gln	Leu	Thr	Ala	Leu	Phe	Val	Phe	Val	Phe	Met	Trp	Ala	305	310	315
Leu	His	Tyr	Ser	Glu	Tyr	Leu	Arg	Glu	Arg	Ala	Arg	Ala	Pro	Ile	His	325	330	335
Ser	Ser	Lys	Ala	Leu	Gln	Ile	Arg	Ala	Arg	Ile	Phe	Met	Gly	Thr	Leu	340	345	350
Ser	Leu	Leu	Leu	Ile	Val	Ala	Ser	Leu	Leu	Ala	Pro	Phe	Gly	Phe	Phe	355	360	365
Lys	Pro	Thr	Ala	Tyr	Arg	Val	Arg	Ala	Leu	Phe	Val	Lys	His	Thr	Arg	370	375	380
Thr	Gly	Asn	Pro	Leu	Val	Asp	Ser	Val	Ala	Glu	His	Arg	Pro	Thr	Thr	385	390	395

[0061]

Ala Gly Ala Tyr Leu Arg Tyr Phe His Val Cys Tyr Pro Leu Trp Gly
 405 410 415
 Cys Gly Gly Leu Ser Met Leu Val Phe Met Lys Lys Asp Arg Trp Arg
 420 425 430
 Ala Ile Val Phe Leu Ala Ser Leu Ser Thr Val Thr Met Tyr Phe Ser
 435 440 445
 Ala Arg Met Ser Arg Leu Leu Leu Leu Ala Gly Pro Ala Ala Thr Ala
 450 455 460
 Cys Ala Gly Met Phe Ile Gly Gly Leu Phe Asp Leu Ala Leu Ser Gln
 465 470 475 480
 Phe Gly Asp Leu His Ser Pro Lys Asp Ala Ser Gly Asp Ser Asp Pro
 485 490 495
 Ala Gly Gly Ser Lys Arg Ala Lys Gly Lys Val Val Asn Glu Pro Ser
 500 505 510
 Lys Arg Ala Ile Phe Ser His Arg Trp Phe Gln Arg Leu Val Gln Ser
 515 520 525
 Leu Pro Val Pro Leu Arg Arg Gly Ile Ala Val Val Val Leu Val Cys
 530 535 540
 Leu Phe Ala Asn Pro Met Arg His Ser Phe Glu Lys Ser Cys Glu Lys
 545 550 555 560
 Met Ala His Ala Leu Ser Ser Pro Arg Ile Ile Ala Val Thr Asp Leu
 565 570 575
 Pro Asn Gly Glu Arg Val Leu Ala Asp Asp Tyr Tyr Val Ser Tyr Leu
 580 585 590
 Trp Leu Arg Asn Asn Thr Pro Glu Asp Ala Arg Ile Leu Ser Trp Trp
 595 600 605
 Asp Tyr Gly Tyr Gln Ile Thr Gly Ile Gly Asn Arg Thr Thr Leu Ala
 610 615 620
 Asp Gly Asn Thr Trp Ser His Lys His Ile Ala Thr Ile Gly Lys Met
 625 630 635 640
 Leu Thr Ser Pro Val Lys Glu Ser His Ala Leu Ile Arg His Leu Ala
 645 650 655
 Asp Tyr Val Leu Ile Trp Ala Gly Glu Asp Arg Gly Asp Leu Leu Lys
 660 665 670
 Ser Pro His Met Ala Arg Ile Gly Asn Ser Val Tyr Arg Asp Met Cys
 675 680 685
 Ser Glu Asp Asp Pro Arg Cys Arg Gln Phe Gly Phe Glu Gly Gly Asp
 690 695 700
 Leu Asn Lys Pro Thr Pro Met Met Gln Arg Ser Leu Leu Tyr Asn Leu
 705 710 715 720

[0062]

His Arg Phe Gly Thr Asp Gly Gly Lys Thr Gln Leu Asp Lys Asn Met
725 730 735

Phe Gln Leu Ala Tyr Val Ser Lys Tyr Gly Leu Val Lys Ile Tyr Lys
740 745 750

Val Val Asn Val Ser Glu Glu Ser Lys Ala Trp Val Ala Asp Pro Lys
755 760 765

Asn Arg Val Cys Asp Pro Pro Gly Ser Trp Ile Cys Ala Gly Gln Tyr
770 775 780

Pro Pro Ala Lys Glu Ile Gln Asp Met Leu Ala Lys Arg Phe His Tyr
785 790 795 800

Glu

<210> 55
<211> 2466
<212> DNA
<213> Trypanosoma brucei

<400> 55
atgacgaaag gtgggaaagt agctgtgact aagggcicag cacagagtga tgggtciggt 60
gaggaggagga tgagtaaggc caagtcaccc actacgttcg togcactgg cgggtgttct 120
ttgcctgcct gggcgctaaa ggcgtgaagc acggttgtga gtgcagtgat tcttatatac 180
tclgtccatc gtgcttacga tatacgactt acttctgtcc gtctttatgg tgagcttatt 240
cacgagttcg acccttgggt caattaccgt gcaacgcagt acctcagcga caacgggtgg 300
cgtgcctttt tccaatggta cgactacatg agctgggtacc cgttggccg accggtgggc 360
acaaccatct tcccgggaat gcagcttacc ggtgtagcca ttcacgtgt gctggaaatg 420
ctcgggcgag gtatgtccat caacaatgc tgtgtglaca ttcctgcag gttcggtagt 480
attgccactg tgttggtgc tctcattgag tacgaatcat ctaattcgct cagtgtcatg 540
gcgtttactg cgtactttt ttcacgtga cctgcacacc tgatgcgac aatggctggt 600
gaatttgaca atgagtgtgt tgcaatggcg gcgatgcttc tgacgttcta catgtgggta 660
cgatcgttac gcagctcaag ttcgtggccc attggcgctt tagctggtgt ggcatacggg 720
tacalgggtg ccacgigggg tggttatatt ttcgtgctga acatggtagc cttccacgct 780
tctgtatgtg tactgcttga ttgggtcgtt gggacataca gcgtcagttt gctgaggcgg 840
tattcactgt ttttgtcat tggcactgcc ctltgcattt gcgtaccacc agtggagtgg 900
acgcccctcc ggtcgttga gcaactgaca gcattgtttg tcttcgtttt catgtgggca 960
ctccactact cggaatacct gcgtgagcgt gcccgagcgc ccattcacic ttclaaagca 1020
cttcagatcc gtgccegcac tttcatgggc accctctcct tgctgttgat ttagctatc 1080
tacctatatt cgacaggata cttcagccg tttctctctc gtgtccgtgc gttgttcgtg 1140
aaacatacgc gtaccgaaa tcccctcgtg gattctgtgg ctgagcacca tccggcgtcg 1200
aatgatgatt tctttggta ccttcatgta tgttacaacg gctggataat tggcttttc 1260
ttcatgtccg tgtcgtgctt tttccattgt acaccgggaa tgcgttccct gcgtgtgtac 1320
tctatacttg cgtactactt ctctctcaag atgagtcgtc tgcgttact tctgcacct 1380
gtggcttcca tactcactgg ctatgtgtg ggatctattg ttgacctgc agcagattgt 1440
ttcgcgcgtt ccggaacaga gcacgcgcac agtaaggagc atcaagggaa agcccgtggt 1500

[0063]

```

aaggacaga aggaacaaat cactgtcgag tgtgggtgcc ataatccctt ctacaaatta 1560
tggtgcaatt cattttcctc ccgcctggta gttgtaagt tttttgtcgt tgttgtcctt 1620
tccatctgtg gaccacatt tcttgggtct aacttccgga tatattctga gcaattcgca 1680
gacagcatgt cgagcccccga gattalaatg agggcaactg tcggtggacg acgagttatc 1740
ttggatgatt actacgtgtc gtaattgtgg ctgcgaaaca atacgcctga agatgccctg 1800
attctctcat ggtgggacta cgggtatcaa atcactggaa ttggcaatcg cacaaccctt 1860
gcggatggta acacatggaa tcacgagcac atagcaacta ttgggaagat gcttacatcc 1920
cctgtgaagg agtcacatgc tcttatacgc catctcgctg attatgtgct gatatgggcc 1980
ggttatgatg gcagcgattt acttaaateg ccacacatgg ctcgatagg caacagtgtg 2040
tatcgcgata tatgctcaga ggatgatccg ctgtgtacgc agttcgggtt ttatagtgtt 2100
gacttcagta aacctacgcc tatgatgcag cggtccttat tatacaatct gcacaggttt 2160
ggtagcgatg gcgggaagac acaactggat aagaacatgt ttcagctcgc ctacgtgtca 2220
aagtatggtt tgggaagat ctacaagggt atgaatgtga gtgaagagag caaggcgtgg 2280
gttcagacc caaagaaccg taagtgcgat gcacctggat ctgggatatg caccggccag 2340
taccgcccag cgaaggagat ccaagacatg ttagcgaaga ggattgacta cgaacaactc 2400
gaggatttca accgccgcaa tcgaagtgcg gcttattatc gtgcgtatat gcgtcagatg 2460
ggttag 2466

```

<210> 56
 <211> 821
 <212> PRT
 <213> Trypanosoma brucei

<400> 56

Met Thr Lys Gly Gly Lys Val Ala Val Thr Lys Gly Ser Ala Gln Ser
1 5 10 15

Asp Gly Ala Gly Glu Gly Gly Met Ser Lys Ala Lys Ser Ser Thr Thr
20 25 30

Phe Val Ala Thr Gly Gly Gly Ser Leu Pro Ala Trp Ala Leu Lys Ala
35 40 45

Val Ser Thr Val Val Ser Ala Val Ile Leu Ile Tyr Ser Val His Arg
50 55 60

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

His Glu Phe Asp Pro Trp Phe Asn Tyr Arg Ala Thr Gln Tyr Leu Ser
85 90 95

Asp Asn Gly Trp Arg Ala Phe Phe Gln Trp Tyr Asp Tyr Met Ser Trp
100 105 110

Tyr Pro Leu Gly Arg Pro Val Gly Thr Thr Ile Phe Pro Gly Met Gln
115 120 125

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

Met Ser Ile Asn Asn Ile Cys Val Tyr Ile Pro Ala Trp Phe Gly Ser

[0064]

145	150	155	160
Ile Ala Thr Val	Leu Ala Ala Leu Ile	Ala Tyr Glu Ser Ser	Asn Ser
	165	170	175
Leu Ser Val Met	Ala Phe Thr Ala Tyr Phe Phe Ser	Ile Val Pro Ala	
	180	185	190
His Leu Met Arg Ser Met Ala Gly Glu Phe Asp Asn	Glu Cys Val Ala		
	195	200	205
Met Ala Ala Met Leu Leu Thr Phe Tyr Met Trp Val Arg Ser Leu Arg			
	210	215	220
Ser Ser Ser Ser Trp Pro Ile Gly Ala Leu Ala Gly Val Ala Tyr Gly			
	225	230	235
Tyr Met Val Ser Thr Trp Gly Gly Tyr Ile Phe Val Leu Asn Met Val			
	245	250	255
Ala Phe His Ala Ser Val Cys Val Leu Leu Asp Trp Ala Arg Gly Thr			
	260	265	270
Tyr Ser Val Ser Leu Leu Arg Ala Tyr Ser Leu Phe Phe Val Ile Gly			
	275	280	285
Thr Ala Leu Ala Ile Cys Val Pro Pro Val Glu Trp Thr Pro Phe Arg			
	290	295	300
Ser Leu Glu Gln Leu Thr Ala Leu Phe Val Phe Val Phe Met Trp Ala			
	305	310	315
Leu His Tyr Ser Glu Tyr Leu Arg Glu Arg Ala Arg Ala Pro Ile His			
	325	330	335
Ser Ser Lys Ala Leu Gln Ile Arg Ala Arg Ile Phe Met Gly Thr Leu			
	340	345	350
Ser Leu Leu Leu Ile Val Ala Ile Tyr Leu Phe Ser Thr Gly Tyr Phe			
	355	360	365
Arg Pro Phe Ser Ser Arg Val Arg Ala Leu Phe Val Lys His Thr Arg			
	370	375	380
Thr Gly Asn Pro Leu Val Asp Ser Val Ala Glu His His Pro Ala Ser			
	385	390	395
Asn Asp Asp Phe Phe Gly Tyr Leu His Val Cys Tyr Asn Gly Trp Ile			
	405	410	415
Ile Gly Phe Phe Phe Met Ser Val Ser Cys Phe Phe His Cys Thr Pro			
	420	425	430
Gly Met Ser Phe Leu Leu Leu Tyr Ser Ile Leu Ala Tyr Tyr Phe Ser			
	435	440	445
Leu Lys Met Ser Arg Leu Leu Leu Leu Ser Ala Pro Val Ala Ser Ile			
	450	455	460
Leu Thr Gly Tyr Val Val Gly Ser Ile Val Asp Leu Ala Ala Asp Cys			

[0065]

465	470	475	480
Phe Ala Ala Ser Gly Thr Glu His Ala Asp Ser Lys Glu His Gln Gly	485	490	495
Lys Ala Arg Gly Lys Gly Gln Lys Glu Gln Ile Thr Val Glu Cys Gly	500	505	510
Cys His Asn Pro Phe Tyr Lys Leu Trp Cys Asn Ser Phe Ser Ser Arg	515	520	525
Leu Val Val Gly Lys Phe Phe Val Val Val Val Leu Ser Ile Cys Gly	530	535	540
Pro Thr Phe Leu Gly Ser Asn Phe Arg Ile Tyr Ser Glu Gln Phe Ala	545	550	555
Asp Ser Met Ser Ser Pro Gln Ile Ile Met Arg Ala Thr Val Gly Gly	565	570	575
Arg Arg Val Ile Leu Asp Asp Tyr Tyr Val Ser Tyr Leu Trp Leu Arg	580	585	590
Asn Asn Thr Pro Glu Asp Ala Arg Ile Leu Ser Trp Trp Asp Tyr Gly	595	600	605
Tyr Gln Ile Thr Gly Ile Gly Asn Arg Thr Thr Leu Ala Asp Gly Asn	610	615	620
Thr Trp Asn His Glu His Ile Ala Thr Ile Gly Lys Met Leu Thr Ser	625	630	635
Pro Val Lys Glu Ser His Ala Leu Ile Arg His Leu Ala Asp Tyr Val	645	650	655
Leu Ile Trp Ala Gly Tyr Asp Gly Ser Asp Leu Leu Lys Ser Pro His	660	665	670
Met Ala Arg Ile Gly Asn Ser Val Tyr Arg Asp Ile Cys Ser Glu Asp	675	680	685
Asp Pro Leu Cys Thr Gln Phe Gly Phe Tyr Ser Gly Asp Phe Ser Lys	690	695	700
Pro Thr Pro Met Met Gln Arg Ser Leu Leu Tyr Asn Leu His Arg Phe	705	710	715
Gly Thr Asp Gly Gly Lys Thr Gln Leu Asp Lys Asn Met Phe Gln Leu	725	730	735
Ala Tyr Val Ser Lys Tyr Gly Leu Val Lys Ile Tyr Lys Val Met Asn	740	745	750
Val Ser Glu Glu Ser Lys Ala Trp Val Ala Asp Pro Lys Asn Arg Lys	755	760	765
Cys Asp Ala Pro Gly Ser Trp Ile Cys Thr Gly Gln Tyr Pro Pro Ala	770	775	780
Lys Glu Ile Gln Asp Met Leu Ala Lys Arg Ile Asp Tyr Glu Gln Leu			

[0066]

785	790	795	800
Glu Asp Phe Asn Arg Arg Asn Arg Ser Asp Ala Tyr Tyr Arg Ala Tyr	805	810	815
Met Arg Gln Met Gly	820		
<210> 57			
<211> 2466			
<212> DNA			
<213> Trypanosoma brucei			
<400> 57			
atgacgaaag gtgggaaagt agctgtgact aagggtcag cacagagtga tgggtgctggt			60
gagggaggga tgagtaaggc caagtcaccc actacgttcg tcgccactgg cgggtggttct			120
ttgcctgcct gggcgctaaa ggctgtaagc acggttgta gtgcagtgat tcttatatac			180
tctgtccatc gtgcttaaga tatacgactt acttctgtcc gtctttatgg tgagcttatt			240
cacgagttcg acccttggtt caattaccgt gcaacgcagt acctcagcga caacgggtgg			300
cgtgcttttt tccaatggta cgactacatg agctgggtacc cgcttgccg accggtgggc			360
acaaccatct tccccgaat gcagcttacc ggtgtagcca ttcacgtgt gtctgaaatg			420
ctcgggcgag gtatgtccat caacaatata tgtgtgtaca ttctgcatg gttcggtagt			480
attgccactg tgttgctgc tctcattgcg tacgagtcac ctaattcgct cagtgtcatg			540
gcgtttactg cgtacttttt ttccatcgta cctgcacacc tgatgcgac aatggctggt			600
gaatttgaca atgagtgtgt tgcaatggcg gcgaigcttc tgacgttcta catgtgggta			660
cgatcggtac gcagctcaag ttctggccc attggcgctt tagctgggtg gccatacggg			720
tacatggtgt ccacgtgggg tggttatatt ttctgtctga acatggtagc ctccacgct			780
tctgtatgtg tactgcttga ttgggtcgt gggacataca gcgtcagitt gctgagggcg			840
tattcactgt tttttgcat tggcaactgcc ctgcgattt gcgtaccacc agtggagtgg			900
acgcccttcc ggctgctgga gcaacigaca gcattgtitg tcttcgtttt catgtgggca			960
ctccactact cggaatacct gcgtgagcgt gcccgagcgc ccattcactc ttctaaagca			1020
cttcagatcc gtgcccgcat ttccatgggc accctctcct tgcgttgat tgtggctatc			1080
tacctatttt cgacaggata ctccaggtcg tttctcttc gtgtccgtgc gttgttcgtg			1140
aaacatacgc gtaccggaaa tcccctcgtg gattctgtgg ctgagcatcg gccgacgact			1200
gccggggcct tcttacgtca tcttcatgtg tgttacaacg gctggataat tggtttttc			1260
ttcatgtccg tgcgtgctt ttccattgt acaccgggaa tgcgttccct gctgttgtac			1320
tctatacttg cgtactactt ctctctcaag atgagtcgtc tgcgttact ttctgcacct			1380
gtggcttcca tactcactgg ctatgttgtg ggatctattg ttgacctcg agcagattgt			1440
ttcgccgctt ccggaacaga gcacgccgac agtaaggagc atcaagggaa agcccggtgt			1500
aagggacaaa agagacaaat cactgtcgag tgtgggtgcc ataaccctt ctacaaatta			1560
tgggtcaatt cattttctc ccgctggta gtggtaagt ttttgtcgt tgtgttctt			1620
tccatctgtg gaccacatt tcttgggtct gagttlaggg ctcatlaga acgtttctc			1680
gtaagtgttg ctatccgcg taltatctg agtattagcg actccgtaa gttggttctt			1740
gccgatgatt actacgtgtc gtacttgtgg ctgcgaaaca atacgctga agatgccgt			1800
attctctcat ggtgggacta cgggtatcaa atcactggaa ttggcaatcg cacaaccctt			1860
gcggacggta acacatggaa tcacgagcac atagcaacta ttgggaagat gcttacatcc			1920

[0067]

```

cctgtgaagg agtcacatgc tcctatacgc catctcgtcg attatgtgct gatatgggcc 1980
ggtgaggatc ggggcgattt acgtaagtca cggcatatgg ctcgatagg caacagtgtg 2040
tatcgcgata tgtgttcaga agacgatccg ctgtgtacgc agttcgggtt ttatagtgg 2100
gacttcaata aacctacgcc tatgatgcag cggtccttat tatacaatct gcacagggtt 2160
ggtacggatg gcgggaagac acaactggat aagaacatgt ttcagctcgc ctacgtgtca 2220
aagtatggtt tggatgaagt ctacaaggat atgaatgtga gtgaagagag caaggcgtgg 2280
gttgacagacc caaagaaccg taagtgcgat gcacctggat cttggatatg cgccggccag 2340
taccgcccag cgaaggagat ccaagacatg ttagcgaaga ggattgacta cgaacaactc 2400
gaggatttca atcgccgcaa tcgaagtgc gcttattatc gtgcgtatat gcgtcagatg 2460
ggtttag
2466

```

```

<210> 58
<211> 821
<212> PRT
<213> Trypanosoma brucei

```

```

<400> 58

```

```

Met Thr Lys Gly Gly Lys Val Ala Val Thr Lys Gly Ser Ala Gln Ser
1 5 10 15

```

```

Asp Gly Ala Gly Glu Gly Gly Met Ser Lys Ala Lys Ser Ser Thr Thr
20 25 30

```

```

Phe Val Ala Thr Gly Gly Gly Ser Leu Pro Ala Trp Ala Leu Lys Ala
35 40 45

```

```

Val Ser Thr Val Val Ser Ala Val Ile Leu Ile Tyr Ser Val His Arg
50 55 60

```

```

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

```

```

His Glu Phe Asp Pro Trp Phe Asn Tyr Arg Ala Thr Gln Tyr Leu Ser
85 90 95

```

```

Asp Asn Gly Trp Arg Ala Phe Phe Gln Trp Tyr Asp Tyr Met Ser Trp
100 105 110

```

```

Tyr Pro Leu Gly Arg Pro Val Gly Thr Thr Ile Phe Pro Gly Met Gln
115 120 125

```

```

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

```

```

Met Ser Ile Asn Asn Ile Cys Val Tyr Ile Pro Ala Trp Phe Gly Ser
145 150 155 160

```

```

Ile Ala Thr Val Leu Ala Ala Leu Ile Ala Tyr Glu Ser Ser Asn Ser
165 170 175

```

```

Leu Ser Val Met Ala Phe Thr Ala Tyr Phe Phe Ser Ile Val Pro Ala
180 185 190

```

```

His Leu Met Arg Ser Met Ala Gly Glu Phe Asp Asn Glu Cys Val Ala
195 200 205

```

[0068]

Met	Ala	Ala	Met	Leu	Leu	Thr	Phe	Tyr	Met	Trp	Val	Arg	Ser	Leu	Arg	210	215	220
Ser	Ser	Ser	Ser	Trp	Pro	Ile	Gly	Ala	Leu	Ala	Gly	Val	Ala	Tyr	Gly	225	230	235
Tyr	Met	Val	Ser	Thr	Trp	Gly	Gly	Tyr	Ile	Phe	Val	Leu	Asn	Met	Val	245	250	255
Ala	Phe	His	Ala	Ser	Val	Cys	Val	Leu	Leu	Asp	Trp	Ala	Arg	Gly	Thr	260	265	270
Tyr	Ser	Val	Ser	Leu	Leu	Arg	Ala	Tyr	Ser	Leu	Phe	Phe	Val	Ile	Gly	275	280	285
Thr	Ala	Leu	Ala	Ile	Cys	Val	Pro	Pro	Val	Glu	Trp	Thr	Pro	Phe	Arg	290	295	300
Ser	Leu	Glu	Gln	Leu	Thr	Ala	Leu	Phe	Val	Phe	Val	Phe	Met	Trp	Ala	305	310	315
Leu	His	Tyr	Ser	Glu	Tyr	Leu	Arg	Glu	Arg	Ala	Arg	Ala	Pro	Ile	His	325	330	335
Ser	Ser	Lys	Ala	Leu	Gln	Ile	Arg	Ala	Arg	Ile	Phe	Met	Gly	Thr	Leu	340	345	350
Ser	Leu	Leu	Leu	Ile	Val	Ala	Ile	Tyr	Leu	Phe	Ser	Thr	Gly	Tyr	Phe	355	360	365
Arg	Ser	Phe	Ser	Ser	Arg	Val	Arg	Ala	Leu	Phe	Val	Lys	His	Thr	Arg	370	375	380
Thr	Gly	Asn	Pro	Leu	Val	Asp	Ser	Val	Ala	Glu	His	Arg	Pro	Thr	Thr	385	390	395
Ala	Gly	Ala	Phe	Leu	Arg	His	Leu	His	Val	Cys	Tyr	Asn	Gly	Trp	Ile	405	410	415
Ile	Gly	Phe	Phe	Phe	Met	Ser	Val	Ser	Cys	Phe	Phe	His	Cys	Thr	Pro	420	425	430
Gly	Met	Ser	Phe	Leu	Leu	Leu	Tyr	Ser	Ile	Leu	Ala	Tyr	Tyr	Phe	Ser	435	440	445
Leu	Lys	Met	Ser	Arg	Leu	Leu	Leu	Leu	Ser	Ala	Pro	Val	Ala	Ser	Ile	450	455	460
Leu	Thr	Gly	Tyr	Val	Val	Gly	Ser	Ile	Val	Asp	Leu	Ala	Ala	Asp	Cys	465	470	475
Phe	Ala	Ala	Ser	Gly	Thr	Glu	His	Ala	Asp	Ser	Lys	Glu	His	Gln	Gly	485	490	495
Lys	Ala	Arg	Gly	Lys	Gly	Gln	Lys	Arg	Gln	Ile	Thr	Val	Glu	Cys	Gly	500	505	510
Cys	His	Asn	Pro	Phe	Tyr	Lys	Leu	Trp	Cys	Asn	Ser	Phe	Ser	Ser	Arg	515	520	525

[0069]

Leu Val Val Gly Lys Phe Phe Val Val Val Val Leu Ser Ile Cys Gly
 530 535 540
 Pro Thr Phe Leu Gly Ser Glu Phe Arg Ala His Cys Glu Arg Phe Ser
 545 550 555 560
 Val Ser Val Ala Asn Pro Arg Ile Ile Ser Ser Ile Arg His Ser Gly
 565 570 575
 Lys Leu Val Leu Ala Asp Asp Tyr Tyr Val Ser Tyr Leu Trp Leu Arg
 580 585 590
 Asn Asn Thr Pro Glu Asp Ala Arg Ile Leu Ser Trp Trp Asp Tyr Gly
 595 600 605
 Tyr Gln Ile Thr Gly Ile Gly Asn Arg Thr Thr Leu Ala Asp Gly Asn
 610 615 620
 Thr Trp Asn His Glu His Ile Ala Thr Ile Gly Lys Met Leu Thr Ser
 625 630 635 640
 Pro Val Lys Glu Ser His Ala Leu Ile Arg His Leu Ala Asp Tyr Val
 645 650 655
 Leu Ile Trp Ala Gly Glu Asp Arg Gly Asp Leu Arg Lys Ser Arg His
 660 665 670
 Met Ala Arg Ile Gly Asn Ser Val Tyr Arg Asp Met Cys Ser Glu Asp
 675 680 685
 Asp Pro Leu Cys Thr Gln Phe Gly Phe Tyr Ser Gly Asp Phe Asn Lys
 690 695 700
 Pro Thr Pro Met Met Gln Arg Ser Leu Leu Tyr Asn Leu His Arg Phe
 705 710 715 720
 Gly Thr Asp Gly Gly Lys Thr Gln Leu Asp Lys Asn Met Phe Gln Leu
 725 730 735
 Ala Tyr Val Ser Lys Tyr Gly Leu Val Lys Ile Tyr Lys Val Met Asn
 740 745 750
 Val Ser Glu Glu Ser Lys Ala Trp Val Ala Asp Pro Lys Asn Arg Lys
 755 760 765
 Cys Asp Ala Pro Gly Ser Trp Ile Cys Ala Gly Gln Tyr Pro Pro Ala
 770 775 780
 Lys Glu Ile Gln Asp Met Leu Ala Lys Arg Ile Asp Tyr Glu Gln Leu
 785 790 795 800
 Glu Asp Phe Asn Arg Arg Asn Arg Ser Asp Ala Tyr Tyr Arg Ala Tyr
 805 810 815
 Met Arg Gln Met Gly
 820
 <210> 59
 <211> 2397
 <212> DNA

[0070]

<213> Trypanosoma cruzi	
<400> 59	
atggacacag cacaattaac actgtgtgga aaatatccac tcgattattc tacagcaagg	60
cgggtcatct ccatactcaa tgtattttatt attgctcttg ccattttaccg cgcctacagc	120
attcgtatga ttccatttcg tgtgtacggc aaggtcatcc acgaatttga tccgtggttc	180
aacttccgag cgtcggagta cctcgacgag cacgggtggg atgccttttt ccactggtag	240
gactacatga gttggtatcc gcttggtcgg cccgtgggca caaccatctt tcctggacta	300
cagatcacaa gcgtccttat tcgcagggcc ctttccatgc tgggaatgag catgacgatg	360
aacgatgtgt gttgcttgat tccggcatgg tttggctcag tggcgactgt attggctgcg	420
cttctggcgt acgagacgtg gggtcgtttt tcaggggctg ccattgacggc gggtcttttt	480
gccatcctac ccgccacct aatgcgtccc atggctggcg aatacgacaa tgagtgtata	540
gcaatggcgg cgatgctgct cacattctac ctgtgggtgc gctcgttgcg caatgcaggc	600
tcattggcca ttggtgtgct tacggggctg gcgtacggct acatggtgtc gacctgggga	660
ggcttcatct ttgtcctgaa catggtggcg ctgcacgccg cgggtgtcgt ttttctgac	720
tggatgcgcg gcaggtaacg tgccagtctg ctgtgggcgt actccctgtt ctttttggtg	780
ggcacggcga ttgcacgtg tgtgccgccg gtgggtggga cccattcaa gtccttgtag	840
cagttgatgg cgttctggtt gttcattttc atgtgggcat tgcacttctc cgagatattg	900
cggcgccgcg ctgatgtccc gattcgtccc accaaggcac tgcggatccg tgcacgggta	960
tttatgatca cctgcggagt gcttgtgtg gctgcagcgc tgcgtggcacc acagggatac	1020
ttcgggcccc tctctcccg cgtccgtgct ttgtttgtgc agcacaccg caccggcaac	1080
cctctcgtag attccgtcgc ggagcaccgt ccattctagt ggggcgcttt gtggaggctt	1140
cttcatctgt gttgtccgct gtgggtgata ggaatgattt cgcgaatact gtcggagaaa	1200
aacgaaaatt taagggaac laccillal g alllggiact ccattatggt attttatttc	1260
ggctgccgca tgcacgctt gatlltgta aciggaccag ttgcagcacc atactccgga	1320
agagtaallg gaggccliat ggactgggcg gtcaggcttc ttttttgac gaatgtagag	1380
tcgatgaaaa gcaaagggtc cccaacgatt cgaagcaaaa aacttgaaaa aaaggggcat	1440
ctatctaata acaatgagcg ttctttacaa aaccgttttc aagacgctgc caatttgtg	1500
ccccacggaa tacgtgtcac aatcgcaatg cttgtgtttg cagcaattct tttcaatccg	1560
atggcccgat cgtataacga agattcaata aagatggcac acacactatc taatccacgg	1620
attatgtggt attcgtatgac cgagcagaac acccctgtac ttgtagacga ctattacgtc	1680
tcgtacctgt ggctgcgcaa caacacaccg gcggacgccg gcattcttgc atggtgggac	1740
tacggtacc agatcacggg gatcgggaat cgcacagccc ttgcggacgg caacacgtg	1800
aatcacgagc acattgccac aattggnaag ctgcttacgt cgcctgtgga gaagggccac	1860
ttgtctattc ggacacctgc cgactatgta ctgatatgga ccggcagccg ggcgaggagc	1920
ttgatgaagt cgcgcacat ggcacgcatt ggcaacagtg tgtaccgca catctgcccc	1980
gaggacgacc cgttggcgc caactllggg tllgaggact acgacctlaag tglcccacg	2040
ccgatgatgc ggatgtcgtt gctgtacaac ctgcatgtct ctggggagag cccagtcgg	2100
gcgatcgaca atatgttcag gcttgcttac aggtcgcgcc accgcctggt gaagatctac	2160
aaggtgatga atgtgagcgc ggagagcaag gcgtgggttg cggacccgaa gaaccgcaag	2220
tgcagcgcg cagggtcgtg gctgtgcaat gggcagtagc cgcagcgaa ggagatccag	2280
gagatgctgg cgaggcgcct cgactacggc cagctggagg acttcaaccg cggcaaacga	2340

[0071]

```

gatgacgcgt actaccgtgc gtacatgcgt cgcacccgca atgaaggcgc tggctag      2397

<210> 60
<211> 798
<212> PRT
<213> Trypanosoma cruzi

<400> 60
Met Asp Thr Ala Gln Leu Thr Leu Cys Gly Lys Tyr Pro Leu Asp Tyr
1          5          10          15

Ser Thr Ala Arg Arg Val Ile Ser Ile Leu Asn Val Phe Ile Ile Ala
          20          25          30

Leu Ala Ile Tyr Arg Ala Tyr Ser Ile Arg Met Ile Ser Ile Arg Val
          35          40          45

Tyr Gly Lys Val Ile His Glu Phe Asp Pro Trp Phe Asn Phe Arg Ala
          50          55          60

Ser Glu Tyr Leu Asp Glu His Gly Trp Asp Ala Phe Phe His Trp Tyr
65          70          75          80

Asp Tyr Met Ser Trp Tyr Pro Leu Gly Arg Pro Val Gly Thr Thr Ile
          85          90          95

Phe Pro Gly Leu Gln Ile Thr Ser Val Leu Ile Arg Arg Ala Leu Ser
          100         105         110

Met Leu Gly Met Ser Met Thr Met Asn Asp Val Cys Cys Leu Ile Pro
          115         120         125

Ala Trp Phe Gly Ser Val Ala Thr Val Leu Ala Ala Leu Leu Ala Tyr
          130         135         140

Glu Thr Trp Gly Ser Phe Ser Gly Ala Ala Met Thr Ala Gly Leu Phe
145         150         155         160

Ala Ile Leu Pro Ala His Leu Met Arg Ser Met Ala Gly Glu Tyr Asp
          165         170         175

Asn Glu Cys Ile Ala Met Ala Ala Met Leu Leu Thr Phe Tyr Leu Trp
          180         185         190

Val Arg Ser Leu Arg Asn Ala Gly Ser Trp Pro Ile Gly Val Leu Thr
          195         200         205

Gly Leu Ala Tyr Gly Tyr Met Val Ser Thr Trp Gly Gly Phe Ile Phe
210         215         220

Val Leu Asn Met Val Ala Leu His Ala Ala Val Cys Val Phe Ala Asp
225         230         235         240

Trp Met Arg Gly Arg Tyr Asp Ala Ser Leu Leu Trp Ala Tyr Ser Leu
          245         250         255

Phe Phe Leu Val Gly Thr Ala Ile Ala Thr Cys Val Pro Pro Val Gly
          260         265         270

Trp Thr Pro Phe Lys Ser Leu Glu Gln Leu Met Ala Leu Leu Val Phe

```

[0072]

275	280	285
Ile Phe Met Trp Ala Leu His Phe Ser Glu Ile Leu Arg Arg Arg Ala 290 295 300		
Asp Val Pro Ile Arg Ser Thr Lys Ala Leu Arg Ile Arg Ala Arg Val 305 310 315 320		
Phe Met Ile Thr Cys Gly Val Leu Val Leu Ala Ala Ala Leu Leu Ala 325 330 335		
Pro Gln Gly Tyr Phe Gly Pro Leu Ser Ser Arg Val Arg Ala Leu Phe 340 345 350		
Val Gln His Thr Arg Thr Gly Asn Pro Leu Val Asp Ser Val Ala Glu 355 360 365		
His Arg Pro Ser Ser Gly Gly Ala Leu Trp Arg Leu Leu His Leu Cys 370 375 380		
Cys Pro Leu Trp Leu Ile Gly Met Ile Ser Gln Ile Leu Ser Gly Glu 385 390 395 400		
Asn Glu Asn Leu Arg Ala Thr Thr Phe Met Ile Trp Tyr Ser Ile Met 405 410 415		
Val Phe Tyr Phe Gly Cys Arg Met Ser Arg Leu Ile Leu Leu Thr Gly 420 425 430		
Pro Val Ala Ala Ser Tyr Ser Gly Arg Val Ile Gly Gly Leu Met Asp 435 440 445		
Trp Ala Val Arg Leu Leu Phe Trp Thr Asn Val Glu Ser Met Lys Ser 450 455 460		
Lys Gly Ser Pro Thr Ile Arg Ser Lys Lys Leu Glu Lys Lys Gly His 465 470 475 480		
Leu Ser Asn Asn Asn Glu Arg Ser Leu Gln Asn Arg Phe Gln Asp Ala 485 490 495		
Ala Asn Leu Trp Pro His Gly Ile Arg Val Thr Ile Ala Met Leu Val 500 505 510		
Phe Ala Ala Leu Leu Phe Asn Pro Met Ala Arg Ser Tyr Asn Glu Asp 515 520 525		
Ser Ile Lys Met Ala His Thr Leu Ser Asn Pro Arg Ile Met Trp Tyr 530 535 540		
Ser Met Thr Glu Gln Asn Thr Pro Val Leu Val Asp Asp Tyr Tyr Val 545 550 555 560		
Ser Tyr Leu Trp Leu Arg Asn Asn Thr Pro Ala Asp Ala Arg Ile Leu 565 570 575		
Ala Trp Trp Asp Tyr Gly Tyr Gln Ile Thr Gly Ile Gly Asn Arg Thr 580 585 590		
Ser Leu Ala Asp Gly Asn Thr Trp Asn His Glu His Ile Ala Thr Ile		

[0073]

595	600	605
Gly Lys Leu Leu Thr Ser Pro Val Ala Lys Ala His Leu Leu Ile Arg 610 615 620		
His Leu Ala Asp Tyr Val Leu Ile Trp Thr Gly Ser Arg Ala Glu Asp 625 630 635 640		
Leu Met Lys Ser Pro His Met Ala Arg Ile Gly Asn Ser Val Tyr Arg 645 650 655		
Asp Ile Cys Pro Glu Asp Asp Pro Leu Cys Ser Asn Phe Gly Phe Glu 660 665 670		
Asp Tyr Asp Leu Ser Arg Pro Thr Pro Met Met Arg Met Ser Leu Leu 675 680 685		
Tyr Asn Leu His Val Ser Gly Glu Ser Pro Ser Pro Ala Ile Asp Asn 690 695 700		
Met Phe Arg Leu Ala Tyr Arg Ser Arg His Gly Leu Val Lys Ile Tyr 705 710 715 720		
Lys Val Met Asn Val Ser Ala Glu Ser Lys Ala Trp Val Ala Asp Pro 725 730 735		
Lys Asn Arg Lys Cys Asp Ala Pro Gly Ser Trp Leu Cys Thr Gly Gln 740 745 750		
Tyr Pro Pro Ala Lys Glu Ile Gln Glu Met Leu Ala Arg Arg Ile Asp 755 760 765		
Tyr Gly Gln Leu Glu Asp Phe Asn Arg Gly Lys Arg Asp Asp Ala Tyr 770 775 780		
Tyr Arg Ala Tyr Met Arg Arg Ile Arg Asn Glu Gly Arg Gly 785 790 795		
<210> 61 <211> 402 <212> DNA <213> <i>Saccharomyces cerevisiae</i>		
<400> 61 atcattctgg acgtatgtgc acatgtgatt tgcttttggt tttttaagaa tgcgggtaa 60 taaacagatt gtttttctgg gaggataatc ttttctttt tcctgttggt attctaaaat 120 taaccttgct gtttctttt tttttttt tcgcgcgact actcagccat ctgcatttt 180 taaagaaaaa gataatcatt aatgccttca cggaataacg tatagaacat tattaaggt 240 atatgaatgg catatatata tagaacacca cccttggaac acatttatac cccttaaac 300 aaaacaattt gctgcgctat accgtgtttc agtgtattat aatacattca tttctgtttc 360 attacgatta tattgacgtg ataaaaagat tatatagcca tg 402		

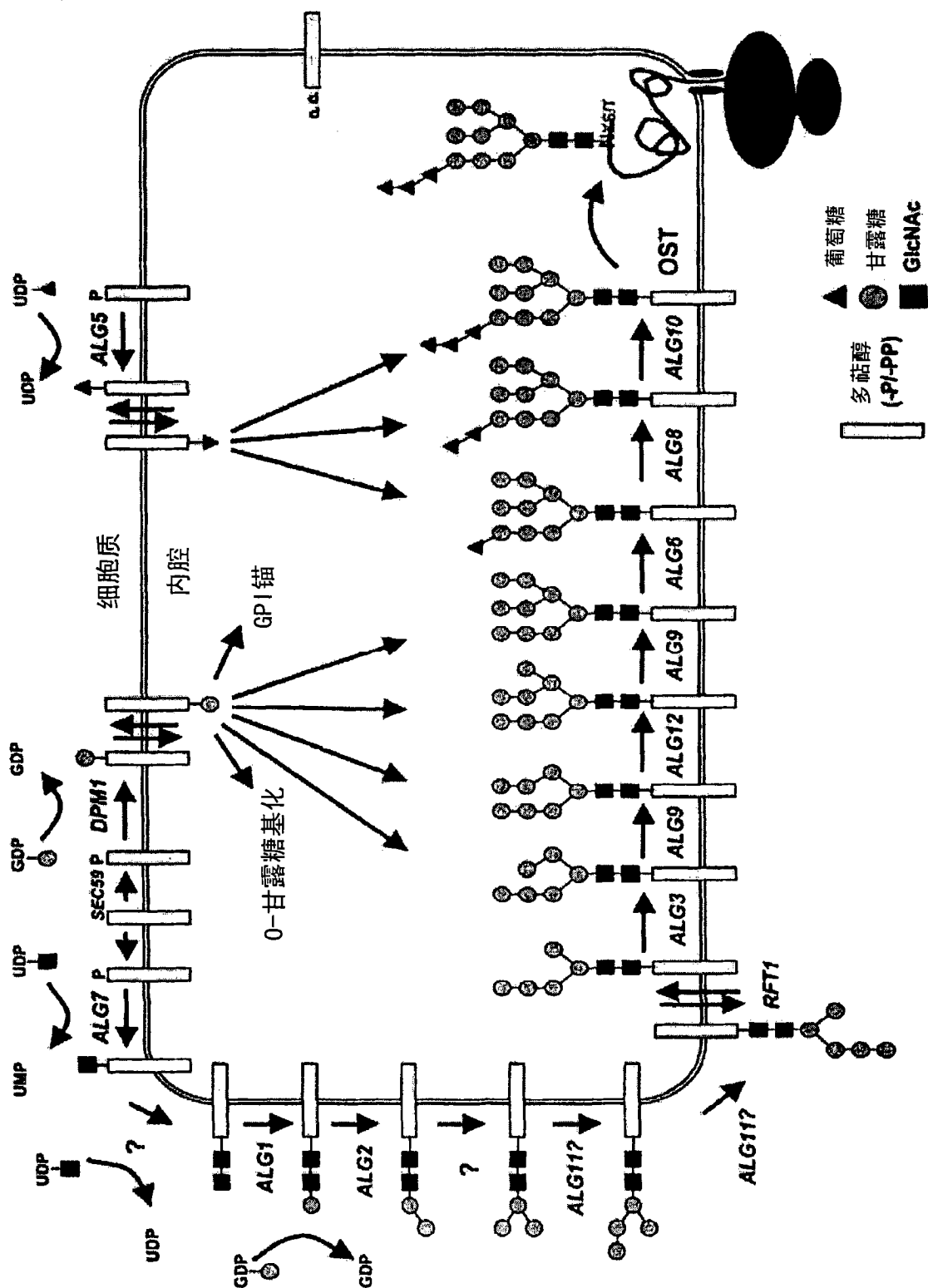


图 1

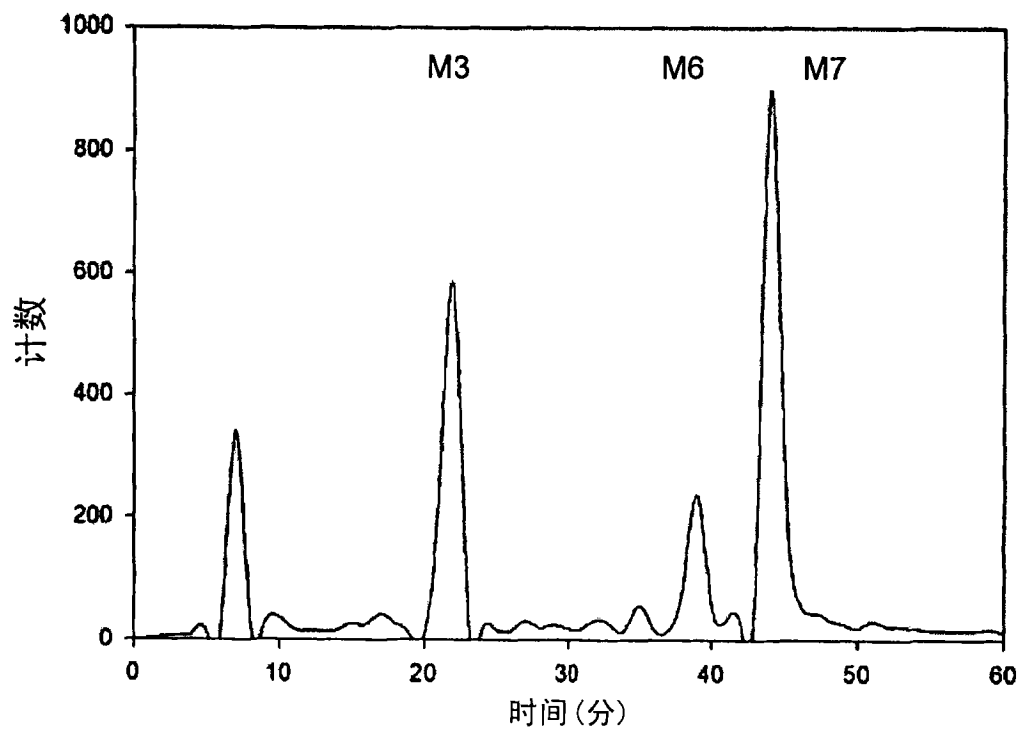


图 2A

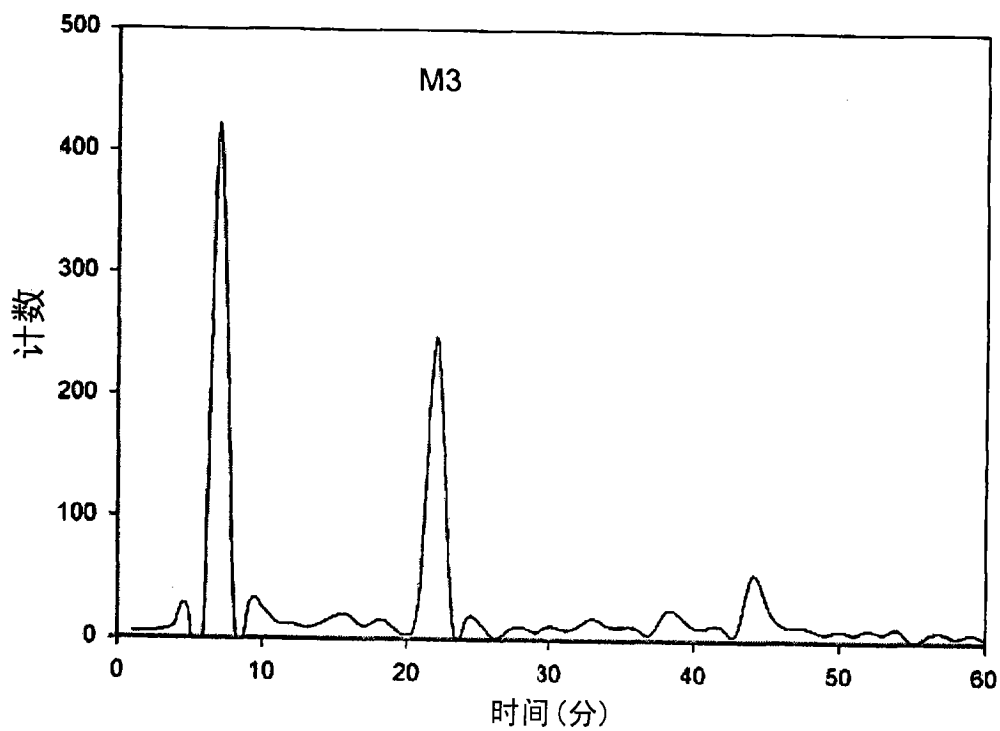


图 2B

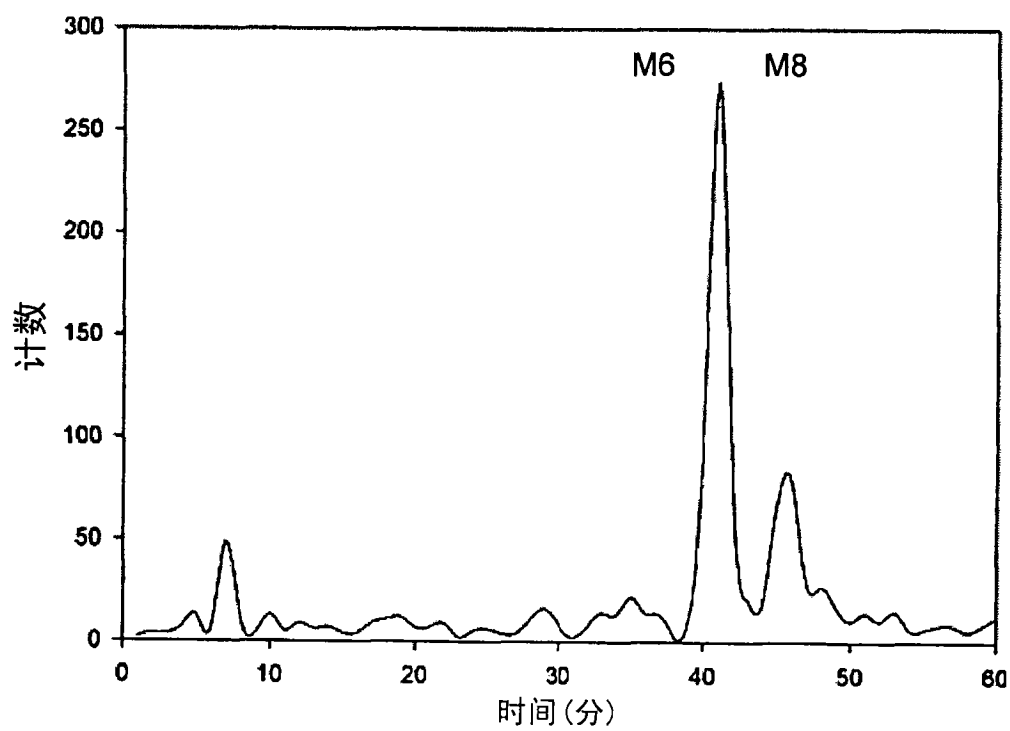


图 3A

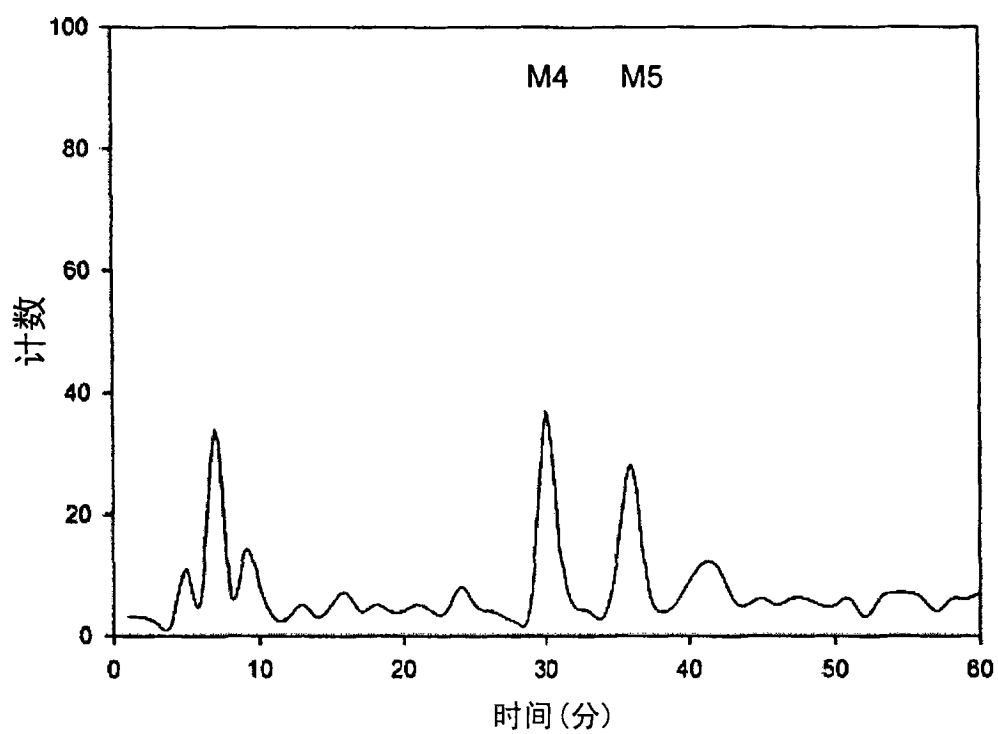
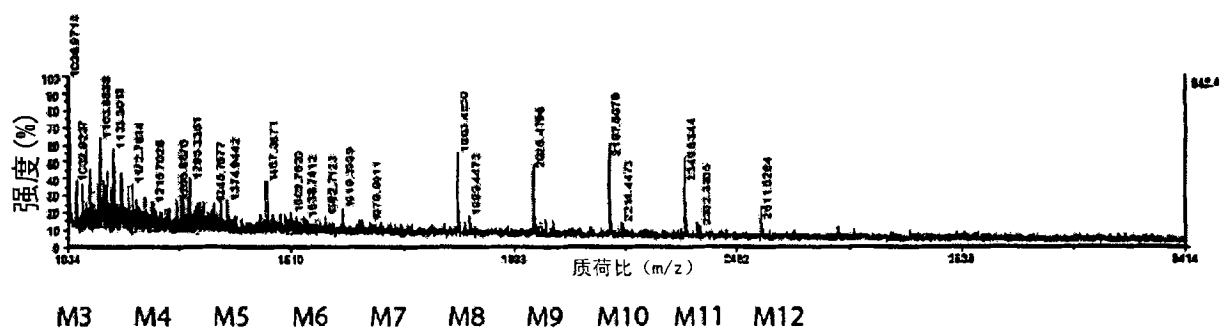


图 3B



ATGATCTTCCTAAACACCTTCGCAAGGTGCCTTTTAACGTGTTTCGTACTGTGCAGCGGTA
CAGCACGTTCCCTCTGACACAAACGACACTACTCCGGCGTCTGCAAAGCATTTCAGACCAC
TTCTTTATTGACGTGTATGGACAATTTCGCAATTAACGGCATCATTCTTTGATGTGAAATTT
TACCCCGATAATAATACTGTTATCTTTGATATTGACGCTACGACGACGCTTAATGGGAACG
TCACTGTGAAGGCTGAGCTGCTTACTTACGGACTGAAAGTCCTGGATAAGACTTTTGATTT
ATGTTCCCTGGGCCAAGTATCGCTTTCCCCCCTAAGTGCTGGGCGTATTGATGTCATGTCC
ACACAGGTGATCGAATCATCCATTACCAAGCAATTTCCCGGCATTGCTTACACCATTCCAG
ATTTGGACGCACAAGTACGTGTGGTGGCATAACGCTCAGAATGACACGGAATTCGAAACTCC
GCTGGCTTGTGTCCAGGCTATCTTGAGTAACGGGAAGACAGTGCAAACAAAGTATGCGGCC
TGGCCATTGCCGCTATCTCAGGTGTCGGTGTACTTACCTCAGGGTTTGTGTCTGTGATCG
GTTACTCAGCCACTGCTGCTCACATTGCGTCCAACCTCCATCTCATTGTTTCATATACTTCCA
AAATCTAGCTATCACTGCAATGATGGGTGCTCAAGGGTTCCACCCATTGCTGCCGCGTGG
ACGCAGAATTTCCAATGGTCCATGGGTATCATCAATACAACTTCATGCAAAAGATTTTGTG
ATTGGTACGTACAGGCCACTAATGGTGTCTCAAATGTTGTGGTAGCTAACAAGGACGTCTT
GTCCATTAGTGTGCAAAAACGTGCTATCTCTATGGCATCGTCTAGTGATTACAATTTTGAC
ACCATTTTAGACGATTTCGGATCTGTACACCACTTCTGAGAAGGATCCAAGCAATTACTCAG
CCAAGATTCTCGTGTTAAGAGGTATAGAAAGAGTTGCTTATTTGGCTAATATTGAGCTATC
TAATTTCTTTT**GACCGTATTGTGTTTTTCTATTCTTCCTATTTGTAGTTGTCGTCTCT**
TTGATTTCTTTAAGGCGCTATTGGAAGTTCTTACAAGAGCAAGAATATTGAAAGAGACTT
CCAATTTCTTCCAATATAGGAAGAACTGGGGGAGTATTATCAAAGGCACCC**TTTTCAGATT**
ATCTATCATCGCCTTCCCTCAAGTTCTCTTCTGGCGATTTGGGAATTTACTCAGGTCAAC
TCTCCAGCGATTGTTGTTGATGCGGTAGTAATATTACTGATCGATCCTCTAGAGTCGACCT
GCAGGCATGCAAGCTAG

图 5A

MI FLNTFARCLLT CFVLC SG TARSSD TNDTPASAKHLQTTSLLT CMDNSQLTASF FDVKF
YPDNNTVIFDIDATTTPLNGNVT VKAELLTYGLKVL DKTFDLCSLGQVSLSP LSAGRIDVMS
TQVIESSITKQFP GIAYTIPDLDAQVRVVAY AQNDTEFETPLACVQAILSN GKTVQTKYAA
WPIAAISGVGVLTSGFVSVIGYSATAAHIASNSISLFIYFQNLAITAMMGVSRVPPIAAAW
TQNFQWSMGIINTNFMQKIFDWYVQATNGVSNVVVANKDVLSISVQKRAISM ASSSDYNFD
TILDDSDLYTTSEKDPSNYS AKILVLRGI ERVAYLANIELSN**FFLTGIVFFLFFLFVWVVS**
LIFFKALLEVLTRARILKETSNEFFQYRKNWGSIIKGT**LFRLSIIAF**PQVSL**LAIWEFTQVN**
SPAIVVD**AVVILLIDPLE**STCRHAS

图 5B

ATGATCTTCCTAAACACCTTCGCAAGGTGCCTTTTAAACGTGTTTCGTACTGTGCAGCGGTA
CAGCACGTTTCCTCTGACACAAACGACACTACTCCGGCGTCTGCAAAGCATTTGCAGACCAC
TTCTTTATTGACGTGTATGGACAATTTCGCAATTAACGGCATCATTCTTTGATGTGAAATTT
TACCCCGATAATAATACTGTTATCTTTGATATTGACGCTACGACGACGCTTAATGGGAACG
TCACTGTGAAGGCTGAGCTGCTTACTTACGGACTGAAAGTCCTGGATAAGACTTTTGATTT
ATGTTCCCTGGGCCAAGTATCGCTTTCCCCCTAAGTGCTGGGCGTATTGATGTCATGTCC
ACACAGGTGATCGAATCATCCATTACCAAGCAATTTCCCGGCATTGCTTACACCATTCCAG
ATTTGGACGCACAAGTACGTGTGGTGGCATAACGCTCAGAATGACACGGAATTCGAAACTCC
GCTGGCTTGTGTCCAGGCTATCTTGAGTAACGGGAAGACAGTGCAAACAAAGTATGCGGCC
TGGCCCATTTGCCGCTATCTCAGGTGTCCGGTGTACTTACCTCAGGGTTTGTGTCTGTGATCG
GTTACTCAGCCACTGCTGCTCACATTGCGTCCAACCTCCATCTCATTGTTTCATATACTTCCA
AAATCTAGCTATCACTGCAATGATGGGTGTCTCAAGGGTTCCACCCATTGCTGCCGCGTGG
ACGCAGAATTTCCAATGGTCCATGGGTATCATCAATACAACTTCATGCAAAAGATTTTGTG
ATTGGTACGTACAGGCCACTAATGGTGTCTCAAATGTTGTGGTAGCTAACAAGGACGTCTT
GTCCATTAGTGTGCAAAAACGTGCTATCTCTATGGCATCGTCTAGTGATTACAATTTTGAC
ACCATTTTAGACGATTCGGATCTGTACACCATTCTGAGAAGGATCCAAGCAATTACTCAG
CCAAGATTCTCGTGTAAAGAGGTATAGAAAGAGTTGCTTATTTGGCTAATATTGAGCTATC
TAATTTCTTTTTGACCGGTATTGTGTTTTTCTATTCTTCCTATTTGTAGTTGTCGTCTCT
TTGATTTTCTTTAAGTAG

图 5C

MI FLNTFARCLLT CFVLC SG TARSSD TNDTTPASAKHLQTTSLT CMDNSQLTASFFDVKF
YPDNNVTIFDIDATTTLNGNVTVKAELLYGLKVLDKTFDLCSLGQVSLSPLSAGRIDVMS
TQVIESSITKQFPGIAYTIPDLDAQVRVVAYAQNDETFETPLACVQAILSNGKTVQTKYAA
WP~~IAAISGVGVLTSGFVSVIGYSATAAHIASNSISLFYFQNLAITAMMGVSRVPPIAAAW~~
TQNFQWSMGIINTNFMQKIFDWYVQATNGVSNVVVANKDVLISVQKRAISMSSSDYNFD
TILDDSDLYTTSEKDPSNYSAKILVLRGIERVAYLANIELSNFFLTGIVFFLEFFLFVVVVS
LIFFK

图 5D

ATGATCTTCCTAAACACCTTCGCAAGGTGCCTTTTAAACGTGTTTCGTACTGTGCAGCGGTA
CAGCACGTTTCCTCTGACACAAACGACACTACTCCGGCGTCTGCAAAGCATTTGCAGACCAC
TTCTTTATTGACGTGTATGGACAATTTCGCAATTAACGGCATCATTCTTTGATGTGAAATTT
TACCCCGATAATAATACTGTTATCTTTGATATTGACGCTACGACGACGCTTAATGGGAACG
TCACTGTGAAGGCTGAGCTGCTTACTTACGGACTGAAAGTCCTGGATAAGACTTTTGATTT
ATGTTCCCTGGGCCAAGTATCGCTTTCCCCCTAAGTGCTGGGCGTATTGATGTCATGTCC
ACACAGGTGATCGAATCATCCATTACCAAGCAATTTCCCGGCATTGCTTACACCATTCCAG
ATTTGGACGCACAAGTACGTGTGGTGGCATAACGCTCAGAATGACACGGAATTCGAAACTCC
GCTGGCTTGTGTCCAGGCTATCTTGAGTAACGGGAAGACAGTGCAAACAAAGTATGCGGCC
TGGCCCATTTGCCGCTATCTCAGGTGTCCGGTGTACTTACCTCAGGGTTTGTGTCTGTGATCG
GTTACTCAGCCACTGCTGCTCACATTGCGTCCAACCTCCATCTCATTGTTTCATATACTTCCA
AAATCTAGCTATCACTGCAATGATGGGTGTCTCAAGGGTTCCACCCATTGCTGCCGCGTGG
ACTAG

图 5E

MIFLNTFARCLLTCFVLC**SGTARSSD**TND**TPASAKHLQ**TT**SL**TCMDNSQLTASFFDVKF
YPDNN**TVIFDIDAT**TT**LNGNV**TVKA**ELLTYGLK**VLDK**TFDLCS**LGQVSL**SPLSAGRI**DVMS
TQV**IESSITKQ**FP**GIAYTI**PD**LDAQVR**VV**AYAQND**TEFETPLACVQ**AILSNGK**TVQTKYAA
WPLAAISGVGV**LTS**GFV**SVIGYS**ATAAH**IASNSISL**FIY**FQNL**AITAMMGVSRVP**PPIAAW**
T

图 5F

ATGATCTTCCTAAACACCTTCGCAAGGTGCCTTTTAACGTGTTTCG**TACTGTGCAGCGGTA**
CAGCACGTTCCCTCTGACACAAACGACATTGCGTCCA**ACTCCATCTCATTGTT**CATATACTT
CCAAAATCTAGCTATCACTGCAATGATGGGTG**TCTCAAGGGTTCACCCATTGCTGCCGCG**
TGGACGCAGAAATTTCCAATGGTCCATGGGTATCATCAATACAACTTCATGCAAAAGATTT
TTGATTGGTACGTACAGGCCACTAATGGTGTCTCAAATGTTGTGGTAGCTAACAAGGACGT
CTTGCCATTAGTGTGCAAAAACGTGCTATCTCTATGGCATCGTCTAGTGATTACAATTTT
GACACCATTTTAGACGATTCGGATCTGTACACCACTTCTGAGAAGGATCCAAGCAATTACT
CAGCCAAGATTCTCGTGTTAAGAGGTATAGAAAGAGTTGCTTATTTGGCTAATATTGAGCT
ATCTAATTTCTTTTTGACCGGTATTGTGTTTTTCTATTCTTCCTATTTGTAGTTGT**CGTC**
TCTTTGATTTTCTTTAAGGCGCTATTGGAAGTTCTTACAAGAGCAAGAATATTGAAAGAGA
CTTCCAATTTCTTCCAATATAGGAAGAACTGGGGGAGTATTATCAAAGGCACCCTTTT**CAG**
ATTATCTATCATCGCCTTCCCTCAAGTTTCTCTTCTGGCGATT**TGGGAATTTACTCAGGTC**
AACTCTCCAGCGATTGTTGTTGATGCGGTAGTAATATTACTGATCGATCCTCTAGAGTCGA
CCTGCAGGCATGCAAGCTAG

图 5G

MIFLNTFARCLLTCFVLC**SGTARSSD**TND**IASNSISL**FIY**FQNL**AITAMMGVSRVP**PPIAAA**
WTQNFQWSMGII**NTNFMQ**KIFDWYVQATNGVSNVVVANKDVLSISVQKRAISM**ASSSDYNF**
DTILDDSDLYTTSEKDP**SNYSAKIL**VLRGIERVAYLANIEL**SNFFLTGIVFFLEF**LVVVV
SLIFFKALLEVLTRARILKETS**NFFQYRKNWGSIIKGT**LFRLSIIAF**PQVSL**LAIWEFTQV
NSPAIVVD**AVVILLIDPLE**STCRHAS

图 5H

ATGATCTTCCTAAACACCTTCGCAAGGTGCCTTTTAACGTGTTTCG**TACTGTGCAGCGGTA**
CAGCACGTTCCCTCTGACACAAACGACTTCTTTTTGACCGGTATTGTGTTTTTCTATTCTT
CCTATTTGTAGTTGT**CGTCTCTTTGATTTTCTTTAAGGCGCTATTGGAAGTTCTTACAAGA**
GCAAGAATATTGAAAGAGACTTCCAATTTCTTCCAATATAGGAAGAACTGGGGGAGTATTA
TCAAAGGCACCCTTTT**CAGATTATCTATCATCGCCTTCCCTCAAGTTTCTCTTCTGGCGAT**
TTGGGAATTTACTCAGGTCAACTCTCCAGCGATTGTTGTTGATGCGGTAGTAATATTACTG
ATCGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTAG

图 5I

MIFLNTFARCLLTCFVLC**SGTARSSD**TND**FFLTGIVFFLEF**LVVVV**SLIFFK**ALLEVLTR
ARILKETS**NFFQYRKNWGSIIKGT**LFRLSIIAF**PQVSL**LAIWEFTQVNSPAIVVD**AVVILL**
IDPLESTCRHAS

图 5K

ATCATTCTGGACGTATGTGCACATGTGATTTGCTTTTGTGTTTTTTAAGAATGTCGGGTAAT
 AAACAGATTGTTTTCTGGGAGGATAATCTTTCTTTTTTCTGTTGGTATTCTAAAATTA
 ACCTTGCTGTTTCTTTTTTTTTTTTTTTCGCGCGACTACTCAGCCATCTTGCAATTTTTAA
 AGAAAAAGATAATCATTAAATGCCTTCACGGGAATACGTATAGAACATTATTTAAAGTATAT
 GAATGGCATATATATATAGAACACCACCCTTGAAAAACATTTATACCCCTTAAACTAAAAC
 AATTTGCTGCGCTATACCGTGTTTCAGTGTATTATAATACATTCATTTCTGTTTCATTACG
 ATTATATTGACGTGATAAAAAAGATTATATAGCCATG

图 5L

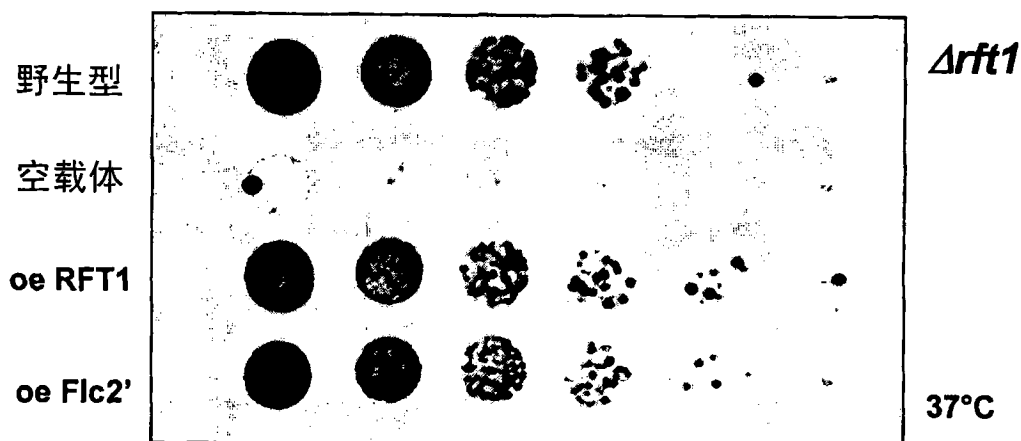


图 6A

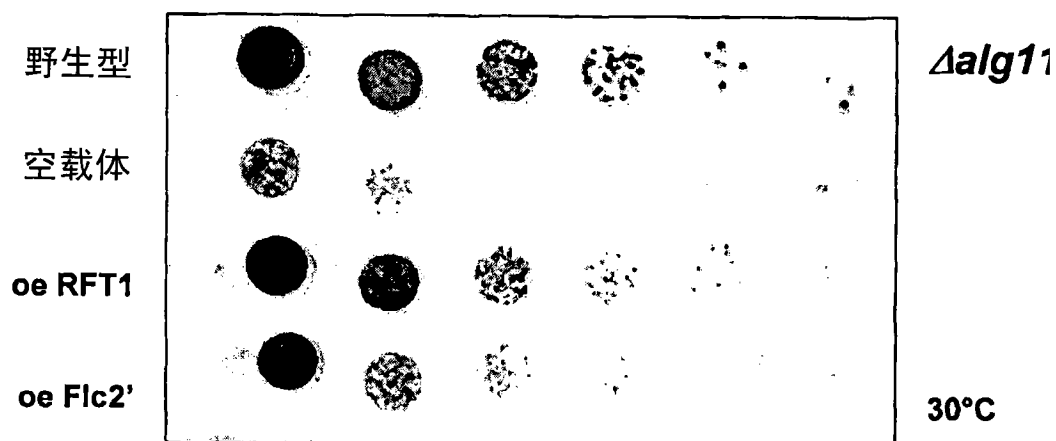


图 6B

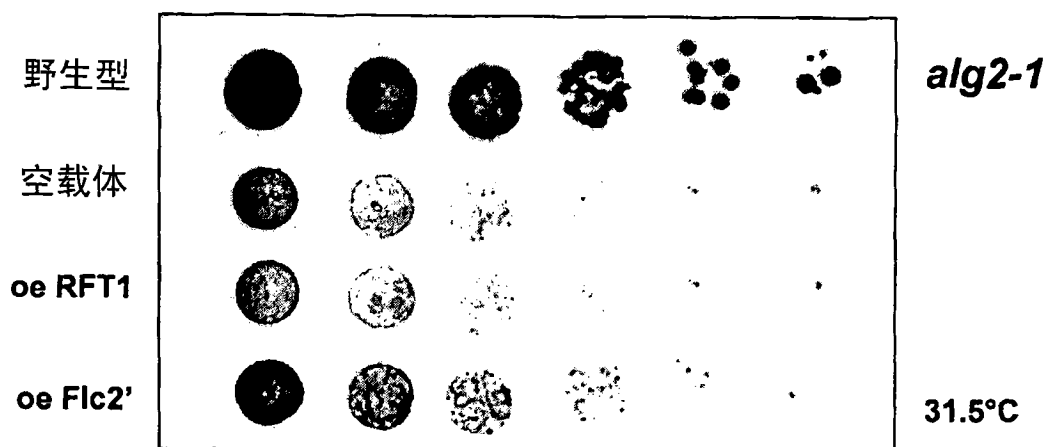


图 6C

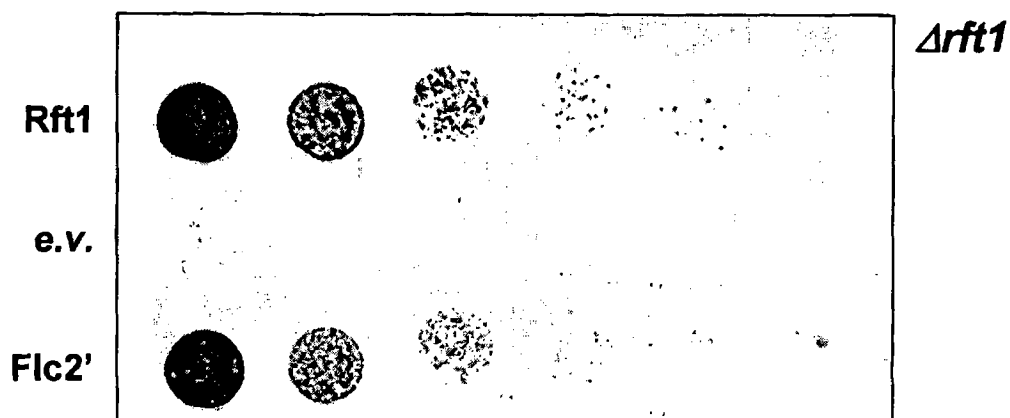


图 7A

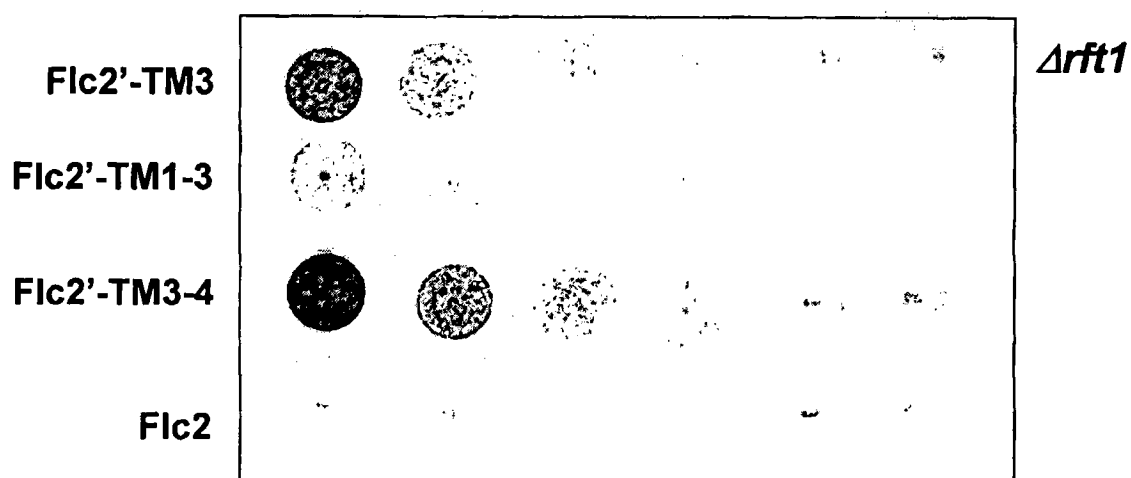


图 7B

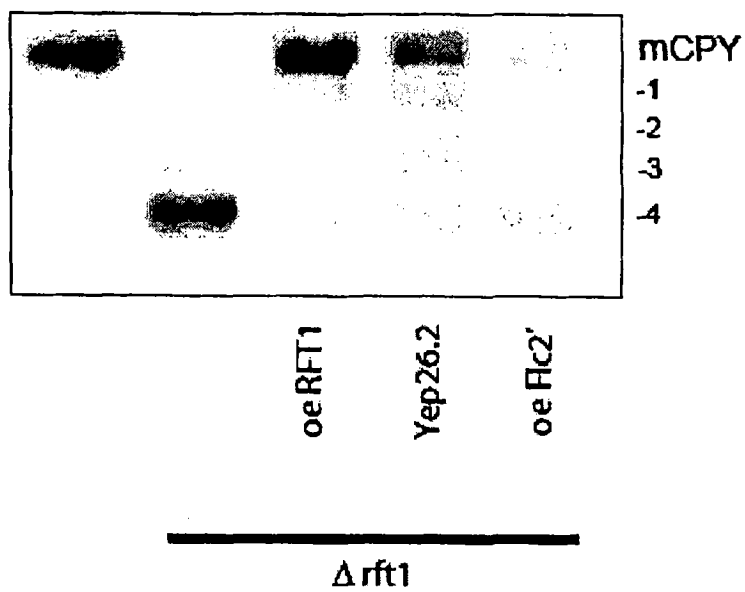


图 7C

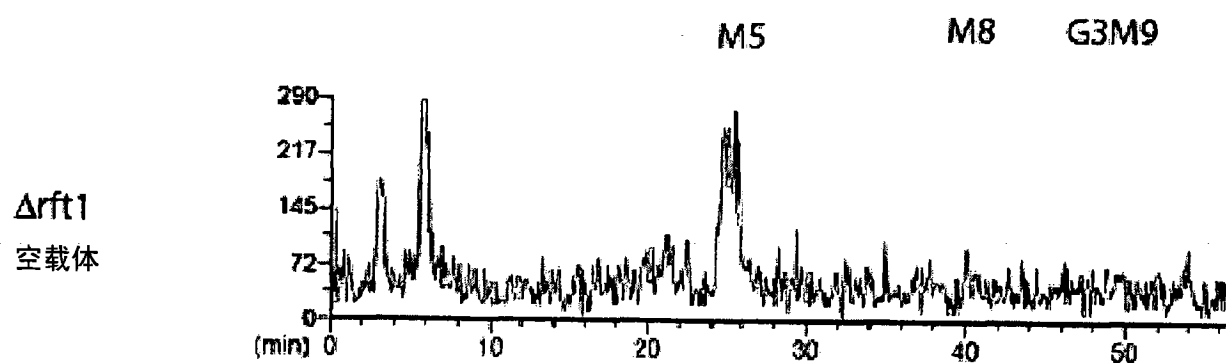


图 8A

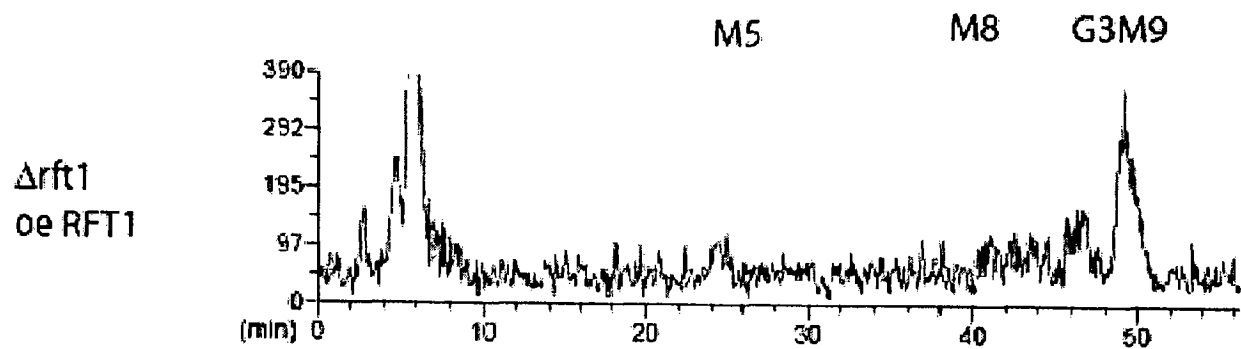


图 8B

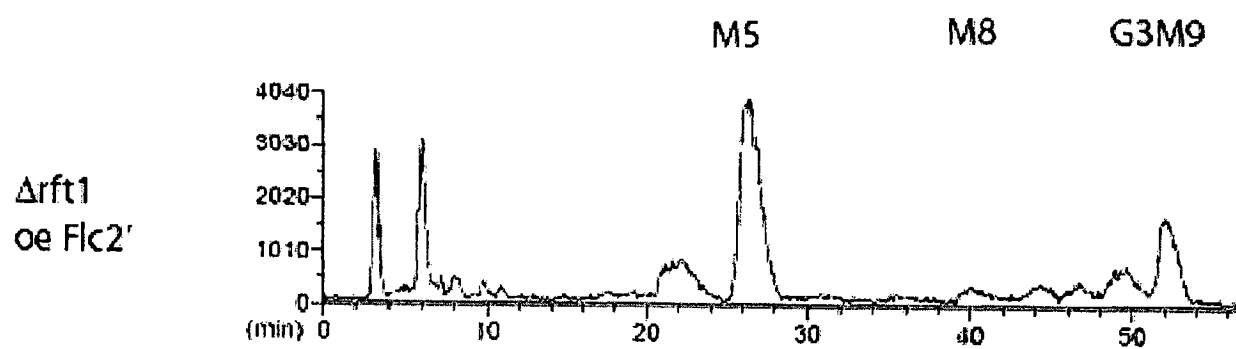


图 8C

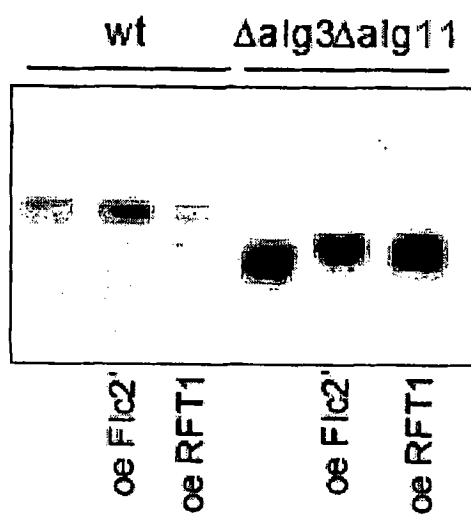


图 9

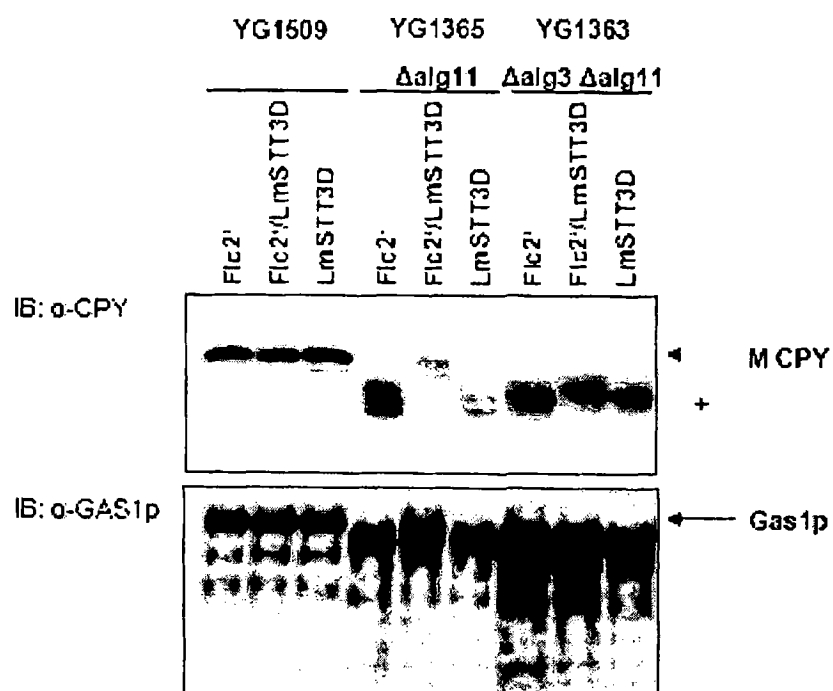


图 10

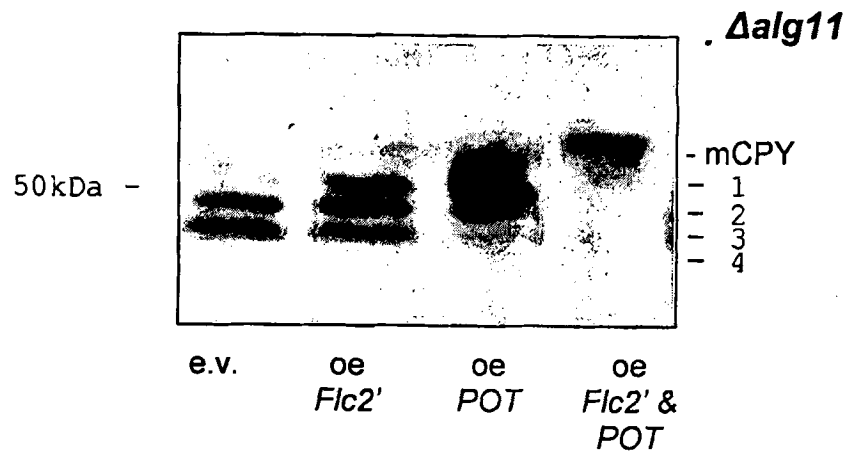


图 11

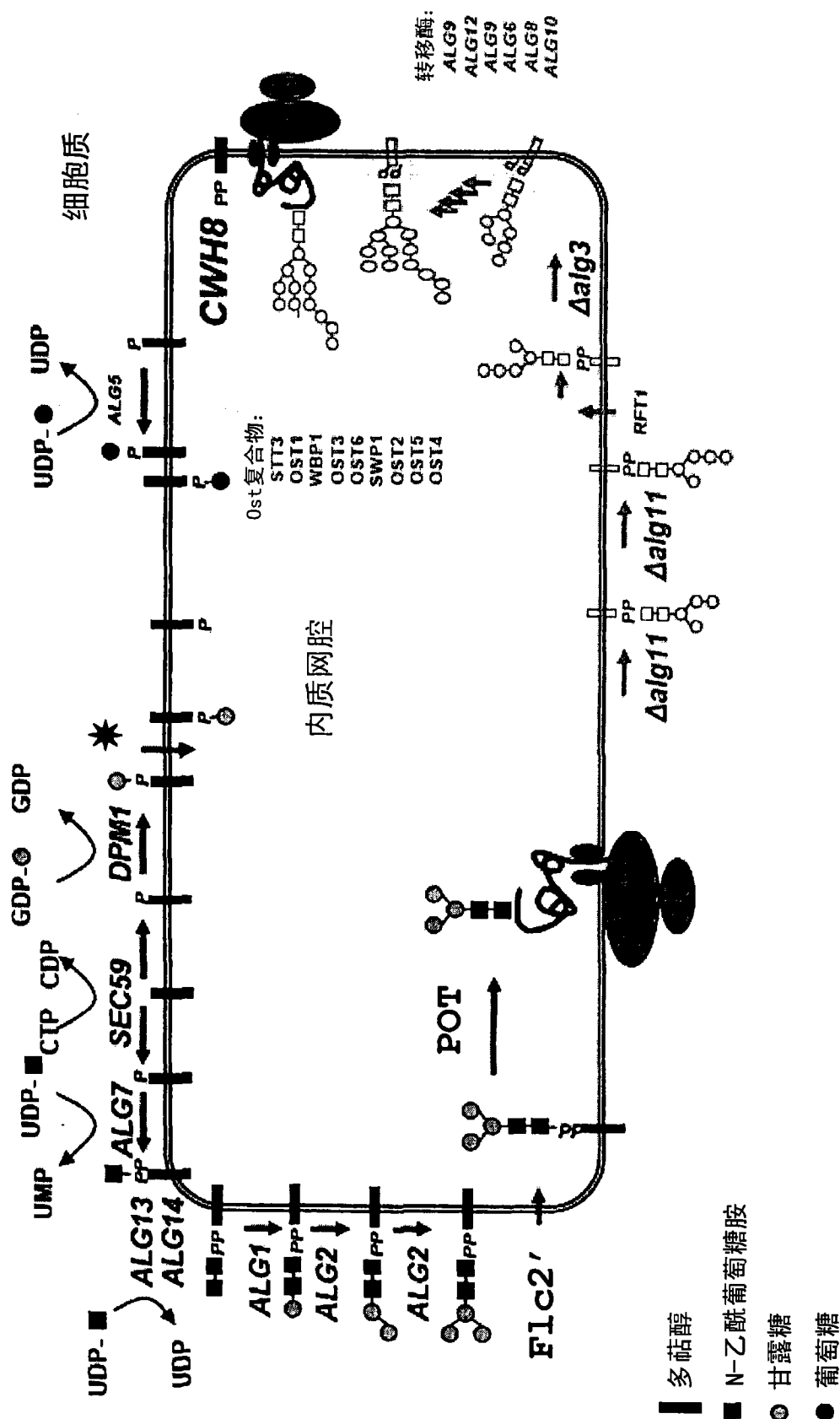


图 12

LOCUS YEp352FLC2 6941 bp DNA CIRCULAR SYN 14-MAR-2008
 DEFINITION Ligation of YEp352* into FLC2*
 ACCESSION YEp352FLC2
 FEATURES Location/Qualifiers
 misc_feature 1..399
 /gene="ACS promoter"
 /product="acetyl-CoA synthetase promoter"
 /SECDrawAs="Region"
 CDS 400..1758
 /gene="FLC2* with potential stop codon"
 /SECDrawAs="Gene"
 CDS 400..1722
 /gene="FLC2 3'-truncated"
 /SECDrawAs="Gene"
 misc_signal 400..471
 /gene="signal sequence"
 /SECDrawAs="Label"
 misc_signal 946..1014
 /gene="1. TMD"
 /SECDrawAs="Label"
 misc_signal 1033..1131
 /gene="2. TMD"
 /SECDrawAs="Label"
 misc_signal 1441..1509
 /gene="3. TMD"
 /SECDrawAs="Label"
 misc_signal 1603..1662
 /gene="4. TMD"
 /SECDrawAs="Label"
 misc_signal 1705..1722
 /gene="part of 5. TMD"
 /SECDrawAs="Label"
 CDS complement (2054..2857)
 /gene="URA3"
 /SECDrawAs="Gene"
 CDS 4855..5715
 /gene="AmpR"
 /SECDrawAs="Gene"
 BASE COUNT 1834 a 1571 c 1479 g 2057 t
 1 atcattctgg acgtatgtgc acatgtgatt tgcttttggt tttttaagaa tgtcgggtaa
 61 taaacagatt gtttttctgg gaggataatc ttttcttttt tcctgttggt attctaaaaat
 121 taaccttgct gtttcttttt tttttttttt tcgcgcgact actcagccat cttgcattttt
 181 taaagaaaaa gataatcatt aatgccttca cgggaatacg tatagaacat tattaaaagt
 241 atatgaatgg catatatata tagaacacca cccttggaac acatttatac cccttaaaact
 301 aaaacaattt gctgcgctat accgtgtttc agtgtattat aatacattca tttctgtttc
 361 attacgatta tattgacgtg ataaaaagat tatatagcca tgatcttcct aaacaccttc
 421 gcaaggtgcc ttttaacgtg tttcgtactg tgcagcggta cagcacgttc ctctgacaca
 481 aacgacacta ctccggcgtc tgcaaagcat ttgcagacca cttctttatt gacgtgtatg
 541 gacaattcgc aattaacggc atcattcttt gatgtgaaat tttaccccga taataatact
 601 gttatctttg atattgacgc tacgacgacg cttaatggga acgtcactgt gaaggctgag
 661 ctgcttactt acggactgaa agtcctggat aagacttttg atttatgttc cttgggccaa
 721 gtatcgcttt cccccctaag tgctgggcgt attgatgtca tgtccacaca ggtgatcgaa
 781 tcatccatta ccaagcaatt tcccggcatt gcttacacca ttccagattt ggacgcacaa
 841 gtacgtgtgg tggcatagc tcagaatgac acggaattcg aaactccgct ggcttgtgtc
 901 caggctatct tgagtaacgg gaagacagtg caaacaagat atgcggcctg gccattgccc
 961 gctatctcag gtgtcgtgtg acttacctca gggtttgtgt ctgtgatcgg ttacttagcc
 1021 actgctgctc acattgcgtc caactccatc tcattgttca tatactcca aaatctagct
 1081 atcactgcaa tgatgggtgt ctcaagggtt ccacccattg ctgccgcgtg gacgcagaat
 1141 ttccaatggt ccatgggtat catcaataga aacttcatgc aaaagatttt tgatttgtac
 1201 gtacaggcca ctaatggtgt ctcaaatggt gtggtagcta acaaggacgt cttgtccatt
 1261 agtgtgcaaa aacgtgctat ctctatggca tcgtctagtg attacaattt tgacaccatt
 1321 ttagacgatt cggtatctgt caccacttct gagaaggatc caagcaatta ctcagccaag
 1381 attctcgtgt taagaggtat agaaagagtt gcttatttgg ctaatattga gctatcta
 1441 ttctttttga ccggtattgt gttttttcta ttcttcctat ttgtagtgtg cgtctctttg

cont.

```
1501 attttcttta aggcgctatt ggaagttctt acaagagcaa gaatattgaa agagacttcc
1561 aatttcttcc aatataggaa gaactggggg agtattatca aaggcaccct ttctagatta
1621 tctatcatcg ccttccctca agtttctctt ctggcgattt gggaatttac tcagggtcaac
1681 tctccagcga ttgttgttga tgcggtagta atattactga tcgatcctct agagtcgacc
1741 tgcaggcatg caagctagct tggcactggc cgtcgtttta caacgctcgt actgggaaaa
1801 ccctggcggt acccaactta atcgccctgc agcacatccc cccttcgcca gctggcgtaa
1861 tagcgaagag gcccgccacc atcgcccttc ccaacagttg cgcagcctga atggcgaatg
1921 ggcgctgatg cggatatttc tccttacgca tctgtgcggt atttcacacc gcatagggtt
1981 ataactgata taattaaatt gaagctctaa tttgtgagtt tagtatacat gcatttactt
2041 ataatacagt tttttagttt tgctggccgc atcttctcaa atatgcttcc cagcctgctt
2101 ttctgtaacg ttcaccctct acctagcat cccttccctt tgcaaatagt cctcttccaa
2161 caataataat gtcagatcct gtagagacca catcatccac ggttctatac tgttgaccca
2221 atgctctccc cttgtcatct aaaccacac cgggtgctat aatcaaccaa tcgtaacctt
2281 catctcttcc acccatgtct ctttgagcaa taaagccgat aacaaaatct ttgtcgtctt
2341 tcgcaatgtc aacagtaccc ttagtatatt ctccagtaga tagggagccc ttgcatgaca
2401 attctgctaa catcaaaagg cctctaggtt cctttgttac ttcttctgcc gcctgcttca
2461 aaccgctaac aatacctggg cccaccacac cgtgtgcatt cgtaatgtct gccattcttg
2521 ctattctgta tacaccgca gactactgca atttgactgt attaccaatg tcagcaaatg
2581 ttctgtcttc gaagagtaaa aaattgtact tggcggataa tgcccttagc ggcttaactg
2641 tgccctccat ggaaaaatca gtcaagatat ccacatgtgt ttttagtaaa caaattttgg
2701 gacctaatgc ttcaactaac tccagtaatt ccttggtggt acgaacatcc aatgaagcac
2761 acaagtttgt ttgcttttcg tgcagtatat taaatagctt ggcagcaaca ggactaggat
2821 atgtagcagc acgttcctta tatgtagctt tcgacatgat ttatcttcgt ttccggtttt
2881 gttctgtgca gttgggttaa gaatactggg caatttcatg ttcttcaac actacatatg
2941 cgtatatata ccaatctaag tctgtgtccc ttcttctggt ctctcttctg ttccggagatt
3001 accgaatcaa aaaaatttca aagaaaccga aatcaaaaaa aagaataaaa aaaaaatgat
3061 gaattgaaaa gctcttgtaa cccatcattg aattttgaac atccgaacct gggagttttc
3121 cctgaacacag atagtatat tgaacctgta taataatata tagtctagcg ctttacggaa
3181 gacaatgtat gtatttcggt tcctggagaa actattgcat ctattgcata ggtaatcttg
3241 cacgtcgcat ccccggttca tttctgcgt ttccatcttg cacttcaata gcatactttt
3301 gttaacgaag catctgtgct tcattttgta gaacaaaaat gcaacgcgag agcgctaatt
3361 tttcaacaaa agaactctgag ctgcattttt acagaacaga aatgcaacgc gaaagcgcta
3421 ttttaccacac gaagaatctg ttgctcattt ttgtaaaaca aaaatgcaac cgcgagcgc
3481 taatttttca aacaaagaat ctgagctgca tttttacaga acagaaatgc aacgcgagag
3541 cgctatttta ccaacaaaga atctatactt cttttttggt ctacaaaaat gcatcccag
3601 agcgctattt ttctaacaaa gcactctaga ttactttttt tctcctttgt gcgctctata
3661 atgcagtctc ttgataactt tttgcactgt aggttcggtta aggttagaag aaggctactt
3721 tgggtgtctat tttctcttcc ataaaaaaag cctgactcca cttccgcgt ttactgatta
3781 ctagcgaagc tgcgggtgca ttttttcaag ataaaggcat ccccgattat attctatacc
3841 gatgtggatt ggcgatactt tgtgaacaga aagtgatagc gttgatgatt cttcattggt
3901 cagaaaaatta tgaacggttt cttctatttt gtctctatat actacgtata ggaaatggtt
3961 acattttcgt attgttttcg attcaactta tgaatagttc ttactacaat ttattgtctt
4021 aaagagtaat actagagata aacataaaaa atgtagaggt cgagtttaga tgcaagtcca
4081 aggagcgaaa ggtggatggg taggttatat agggatatag cacagagata tatagcaaa
4141 agatactttt gagcaatggt tgtggaagcg gtattcgcaa tatttttaga gctcgttaca
4201 gtcggtgctg tttttggttt tttgaaagtg cgtcttcaga gcgcttttgg ttttcaaaag
4261 cgctctgaag ttctatact ttctagctag agaataggaa cttcggaata ggaacttcaa
4321 agcgtttccg aaaacgagcg cttccgaaaa tgcaacgcga gctgcgcaca tacagctcac
4381 tgttcacgtc gcacctatat ctgctgtgtg cctgtatata tatatacatg agaagaacgg
4441 catagtgcgt gtttatgctt aaatgcgtta tgggtgactc tcagtacaat ctgctctgat
4501 gccgcatagt taagccagcc ccgacaccg ccaacaccg ctgacgcgcc ctgacgggct
4561 tgtctgctcc cggcatccgc ttacagacaa gctgtgaccg tctccgggag ctgcatgtgt
4621 cagaggtttt caccgtcatc accgaaacgc gcgagacgaa agggcctcgt gatacgctta
4681 tttttatagg ttaatgtcat gataataatg gtttcttaga cgtcaggtgg cacttttcgg
4741 ggaaatgtgc cgggaacccc tatttggtta tttttctaaa tacattcaaa tatgtatccg
4801 ctcatgagac aataaccctg ataaatgctt caataatatt gaaaaaggaa gagtatgagt
4861 attcaacatt tccgtgtcgc cttatttccc ttttttgcgg cattttgcct tccgtttttt
4921 gctcaccag aaacgctggt gaaagtaaaa gatgctgaag atcagttggg tgcacgagt
4981 ggttacatcg aactggtct caacagcggg aagatccttg agagttttcg ccccgagaa
5041 cgttttccaa tgatgagcac ttttaaagtt ctgctatgtg gcgcggtatt atcccgtatt
5101 gacgcggggc aagagcaact cggtcgccc atacactatt ctcagaatga cttggttgag
5161 tactcaccag tcacagaaaa gcactttacg gatggcatga cagtaagaga attatgcagt
```

cont.

```
5221 gctgccataa ccatgagtgga taacactgcg gccaaacttac ttctgacaac gatcggagga
5281 ccgaaggagc taaccgcttt ttgacacaac atggggggtc atgtaactcg ccttgatcgt
5341 tgggaaccgg agctgaatga agccatacca aacgacgagc gtgacaccac gatgcctgta
5401 gcaatggcaa caacgttgcg caaactatta actggcgaaac tacttactct agcttcccgg
5461 caacaattaa tagactggat ggaggcggt aaagttgcag gaccacttct gcgctcggcc
5521 cttccggctg gctggtttat tgctgataaa tctggagccg gtgagcgtgg gtctcgcggt
5581 atcattgcag cactggggcc agatggtaag ccctcccgta tcgtagttat ctacacgacg
5641 gggagtcagg caactatgga tgaacgaaat agacagatcg ctgagatagg tgcctcactg
5701 attaagcatt ggttaactgtc agaccaagtt tactcatata tacttttagat tgatttaaaa
5761 cttcatTTTT aatttaaaag gatctaggtg aagatccttt ttgataatct catgaccaa
5821 atcccttaac gtgagttttc gttccactga gcgtcagacc ccgtagaaaa gatcaaagga
5881 tcttcttgag atcctttttt tctgcgcgta atctgctgct tgcaaacaaa aaaaccaccg
5941 ctaccagcgg tggtttggtt gccggatcaa gagctaccaa ctctttttcc gaaggtaact
6001 ggcttcagca gagcgcagat accaaatact gtccttctag tgtagccgta gttaggccac
6061 cacttcaaga actctgtagc accgcctaca tacctcgctc tgctaatacct gttaccagt
6121 gctgctgcca gtggcgataa gtcgtgtctt accgggttgg actcaagacg atagtaccg
6181 gataaggcgc agcggtcggg ctgaacgggg gggttcgtgca cacagcccag cttggagcga
6241 acgacctaca ccgaactgag atacctacag cgtgagcatt gagaaagcgc cacgcttccc
6301 gaaggggagaa aggcggacag gtatccggtg agcggcaggg tcggaacagg agagcgcacg
6361 agggagcttc cagggggaaa cgcctggtat cttttagtgc ctgtcgggtt tcgccacctc
6421 tgacttgagc gtcgattttt gtgatgctcg tcaggggggc ggagcctatg gaaaaacgcc
6481 agcaacgcgg ccttttttac gttcctggcc ttttgctggc cttttgctca catgttcttt
6541 cctgcgttat ccctgattc tgtggataac cgtattaccg cttttgagtg agctgatacc
6601 gctcgcgcga gccgaacgac cgagcgcagc gaggcagtga gcgaggaagc ggaagagcgc
6661 ccaatacgca aaccgcctct ccccgcgctt tggccgattc attaatccag ctggcacgac
6721 aggtttcccg actggaaagc gggcagtgag cgcaacgcaa ttaatgtgag ttacctcact
6781 cattaggcac cccaggcttt acactttatg cttccggctc gtatgttgtg tgggaattgtg
6841 agcggataac aatttcacac aggaaacagc tatgaccatg attacgaatt cgagctcggg
6901 acccggggat cctctagagt cgacctgcag gcatgcaagc t
```

图 13

LOCUS pAX306 10475 bp DNA CIRCULAR SYN 31-MAR-2008
 DEFINITION Ligation of LmSTT3D into SpeI XhoI of pAX305
 ACCESSION pAX306
 misc_feature 1..399
 / gene="ACS promoter"
 / product="acetyl-CoA synthetase promoter"
 CDS 400..1722
 / gene="FLC2 3'-truncated"
 misc_signal 400..471
 / gene="signal sequence"
 CDS 400..1758
 / gene="FLC2* with pot.stop"
 misc_signal 946..1014
 / gene="1. TMD"
 misc_signal 1033..1131
 / gene="2. TMD"
 misc_signal 1441..1509
 / gene="3. TMD"
 misc_signal 1603..1662
 / gene="4. TMD"
 misc_signal 1705..1722
 / gene="part of 5. TMD"
 CDS complement (2054..2857)
 / gene="URA3"
 CDS 4855..5715
 / gene="AmpR"
 misc_feature complement (6904..7155)
 / gene="CYC1 term"
 / product="cyc1 terminator"
 CDS complement (7165..7191)
 / gene="2x HA tag"
 CDS complement (7192..9762)
 / gene="LmSTT3D"
 misc_feature complement (9781..10435)
 / gene="GPD promoter"
 BASE COUNT 2669 a 2505 c 2488 g 2813 t
 1 atcattctgg acgtatgtgc acatgtgatt tgcttttgtt tttttaagaa tgtcgggtaa
 61 taaacagatt gtttttctgg gaggataatc ttttcttttt tcctgttggt attctaaaat
 121 taaccttgct gtttcttttt tttttttttt tcgcgcgact actcagccat cttgcatttt
 181 taaagaaaaa gataatcatt aatgccttca cggaataacg tatagaacat tattaaaagt
 241 atatgaatgg catatatata tagaacacca cccttggaaa acatttatac cccttaaact
 301 aaaacaattt gctgcgctat accgtgtttc agtgtattat aatacattca tttctgtttc
 361 attacgatta tattgacgtg ataaaaagat tatatagcca tgatcttcct aaacaccttc
 421 gcaagggtgc ttttaacgtg tttcgtactg tgcagcggta cagcacgttc ctctgacaca
 481 aacgacacta ctccggcgtc tgcaaagcat ttgcagacca cttctttatt gacgtgtatg
 541 gacaattcgc aattaacggc atcattcttt gatgtgaaat tttaccccca taataatact
 601 gttatctttg atattgacgc tacgacgacg cttaatggga acgtcactgt gaaggctgag
 661 ctgcttactt acggactgaa agtcctggat aagacttttg atttatgttc cttgggccaa
 721 gtatcgcttt ccccctaag tgctgggcgt attgatgtca tgtccacaca ggtgatcgaa
 781 tcatccatta ccaagcaatt tcccggcatt gcttacacca ttccagattt ggacgcacaa
 841 gtacgtgtgg tggcatagc tcagaatgac acggaattcg aaactccgct ggcttgtgtc
 901 caggctatct tgagtaacgg gaagacagtg caaacaagat atgcggcctg gccattgccc
 961 gctatctcag gtgtcgggtg acttacctca gggtttgtgt ctgtgatcgg ttaactcagcc
 1021 actgctgctc acattgcgtc caactccatc tcattgttca tatacttcca aaatctagct
 1081 atcactgcaa tgatgggtgt ctcaagggtt ccacccattg ctgccgcgtg gacgcagaat
 1141 ttccaatggt ccatgggtat catcaatata aacttcatgc aaaagatttt tgattggtac
 1201 gtacaggcca ctaatgggtg ctcaaagtgt gtggtagcta acaaggacgt cttgtccatt
 1261 agtgtgcaaa aacgtgctat ctctatggca tcgtctagtg attacaattt tgacaccatt
 1321 ttagacgatt cggtatctgt caccacttct gagaaggatc caagcaatta ctacagccaag
 1381 attctcgtgt taagaggtat agaaagagtt gcttatttgg ctaatatgga gctatctaatt
 1441 ttctttttga ccggtattgt gttttttcta ttcttcctat ttgtagtgtt cgtctctttg
 1501 attttcttta aggcgctatt ggaagtctt acaagagcaa gaatatgaa agagacttcc
 1561 aatttcttcc aatataggaa gaactggggg agtattatca aaggcaccct ttctcagatta
 1621 tctatcatcg cttccctca agtttctctt ctggcgattt gggaatttac tcaggtcaac

cont.

1681 tctccagcga ttgttgttga tgcggtagta atattactga tcgacccctc agagtcgacc
1741 tgcaggcatg caagctagct tggcactggc cgtcgtttta caacgctcgt actgggaaaa
1801 ccctggcggt acccaactta atcgccctgc agcacatccc cccttcgcca gctggcgtaa
1861 tagcgaagag gcccgcacgg atcgcccttc ccaacagttg cgcagcctga atggcgaaatg
1921 gcgcctgatg cggatatttc tccttacgca tctgtgcggt atttcacacc gcatagggtatg
1981 ataactgata taattaaatt gaagctctaa tttgtgagtt tagtatacat gcatttactt
2041 ataatacagt tttttagttt tgctggccgc atcttctcaa atatgcttcc cagcctgctt
2101 ttctgtaacg ttcaccctct accttagcat cccttccctt tgcaaatagt cctcttccaa
2161 caataataat gtcagatcct gtagagacca catcatccac ggttctatac tgttgaccca
2221 atgctgtctc cttgtcatct aaacccacac cgggtgtcat aatcaacca gctaacctt
2281 catctcttcc acccatgtct ctttgagcaa taaagccgat aacaaaatct ttgtcgtctt
2341 tcgcaatgtc aacagtaccc ttagtatatt ctccagtaga tagggagccc ttgcatgaca
2401 attctgtcaa catcaaaagg cctctaggtt cctttgttac ttcttctgcc gcctgcttca
2461 aaccgctaac aatacctggg cccaccacac cgtgtgcatt cgtaatgtct gccattcttg
2521 ctattctgta tacacccgca gactactgca atttgactgt attaccaatg tcagcaaatg
2581 ttctgtcttc gaagagtaaa aaattgtact tggcggataa tgcccttagc ggcttaactg
2641 tgccctccat ggaaaaatca gtcaagatat ccacatgtgt ttttagtaaa caaattttgg
2701 gacctaatgc ttaactaac tccagtaatt ccttgggtgt acgaacatcc aatgaagcac
2761 acaagtttgt ttgcttttgc tgcagtatat taaatagctt ggcagcaaca ggactaggat
2821 gagtagcagc acgttccctt tatgtagctt tcgacatgat ttatcttctg ttcgggtttt
2881 gttctgtgca gttgggttaa gaatactggg caatttctatg tttcttcaac actacatag
2941 cgtatatata ccaatctaag tctgtgctcc ttccttctgt cttccttctg ttcggagatt
3001 accgaatcaa aaaaatttca aagaaaccca aatcaaaaaa aagaataaaa aaaaaatgat
3061 gaattgaaaa gctcttgtaa cccatcattg aattttgaac atccgaacct gggagttttc
3121 cctgaacagc atagtatatt tgaacctgta taataatata tagtctagcg ctttacggaa
3181 gacaatgtat gtatttcggg tccctggagaa actattgcat ctattgcata ggtaactctg
3241 cacgtcgcat ccccggttca ttttctgcgt ttccatcttg cacttcaata gcatacttt
3301 gttaacgaag catctgtgct tcattttgta gaacaaaaat gcaacgcgag agcgtaatt
3361 tttcaacaa agaactcgag ctgcattttt acagaacaga aatgcaacgc gaaagcgcta
3421 tttaccaac gaagaatctg tgcttcattt ttgtaaaaca aaaatgcaac gcgagagcgc
3481 taatttttca aacaagaat ctgagctgca tttttacaga acagaaatgc aacgcgagag
3541 cgctatttta ccaacaaaga atctatactt ctttttgtgt ctacaaaaat gcaccccgag
3601 agcgctatct ttctaacaaa gcatcttaga ttactttttt tctcctttgt gcgctctata
3661 atgcagtctc ttgataactt tttgcactgt aggtccgtta aggttagaag aaggctactt
3721 tgggtgtctat tttctcttcc ataaaaaaag cctgactcca cttcccgctt ttactgatta
3781 cttagcgaagc tgcgggtgca ttttttcaag ataaaggcat ccccgattat attctatacc
3841 gatgtggatt gcgcatactt tgtgaacaga aagtgatagc gttgatgatt ttcattgggt
3901 cagaaaatta tgaacgggtt cttctatttt gtctctatat actacgtata ggaatgttt
3961 acattttcgt attgttttcg attcactcta tgaatagttc ttactacaat ttttttgtct
4021 aaagagtaat actagagata aacataaaaa atgtagaggt cgagttaga tgcaagttca
4081 aggaagcgaaa ggtggatggg taggttata agggatagc cacagagata tatagcaaa
4141 agatactttt gagcaatggt tgtggaagcg gtatttcgca tattttagta gctcctatac
4201 gtccggtgag tttttgggtt tttgaaagtg cgtcttcaga gcgcttttgg tttcaaaaag
4261 cgctctgaag ttctataact ttctagctag agaataggaa cttcggaaata ggaacttcaa
4321 agcgtttccg aaaacgagcg cttccgaaaa tgcaacgcga gctgcgcaca tacagctcac
4381 tgttcacgct gcacctatat ctgctgtgtg cctgtatata tatatacatg agaagaacgg
4441 catagtgcgt gtttatgctt aaatgcgtta tgggtgactc tcagtacaat ctgctctgat
4501 gccgcatagt taagccagcc ccgacacccg ccaacacccg ctgacgcgcc ctgacgggct
4561 tgtctgctcc cggcatccgc ttacagacaa gctgtgaccg tctccgggag ctgcatgtgt
4621 cagagggtttt caccgtcatc accgaaacgc gcgagacgaa agggcctcgt gatacgccta
4681 tttttatagg ttaatgtcat gataataatg gtttcttaga cgtcaggtgg cacttttcgg
4741 ggaaatgtgc gcggaacccc tatttgttta tttttctaaa tacattcaaa tatgtatccg
4801 ctcatgagac aataaccctg ataaatgctt caataatatt gaaaaaggaa gagtatgagt
4861 attcaacatt tccgtgtcgc ccttattccc ttttttgcgg cattttgcct tccgtttttt
4921 gctcaccagc aaacgctggt gaaagtaaaa atgctgaag atcagttggg tgcacgagt
4981 ggttacatcg aactggatct caacagcggg aagatccttg agagttttcg cccgaagaa
5041 cgttttccaa tgatgagcac ttttaaagtt ctgctatgtg gcgcgggtatt atcccgtatt
5101 gacgccgggc aagagcaact cggctcgccgc atacactatt ctcagaatga cttgggtgag
5161 tactcaccag tcacagaaaa gcatcttacg gatggcatga cagtaagaga attatgcagt
5221 gctgccataa ccatgagtga taacactcg gccaaactac ttctgacaac gatcggagga
5281 ccgaaggagc taaccgcttt tttgcacaa atgggggatc atgtaaactc ccttgatcgt
5341 tgggaaccgg agctgaatga agccatacca aacgacgagc gtgacaccac gatgcctgta

cont.

```

5401 gcaatggcaa caacgttgcg caaactatta actggcgaac tacttactct agcttcccgg
5461 caacaattaa tagactggat ggaggcggat aaagtgtcag gaccacttct gcgctcggcc
5521 cttccggctg gctggtttat tgctgataaa tctggagccg gtgagcgtgg gtctcggcgt
5581 atcattgcag cactggggcc agatggtaag ccctcccgtg tcgtagttat ctacacgacg
5641 gggagtcagg caactatgga tgaacgaaat agacagatcg ctgagatagg tgcctcactg
5701 attaagcatt ggtaactgtc agaccaagtt tactcatata tacttttagat tgatttaaaa
5761 cttcattttt aattttaaag gatctagggtg aagatccctt ttgataatct catgaccaa
5821 atcccttaac gtgagtttct gttccactga gcgtcagacc ccgtagaaaa gatcaaagga
5881 tcttcttgag atcctttttt tctgcgcgta atctgctgct tgcaaacaaa aaaaccaccg
5941 ctaccagcgg tggtttggtt gccggatcaa gagctaccaa ctctttttcc gaaggttaact
6001 ggcttcagca gagcgcagat accaaatact gtcttcttag tgtagccgta gttaggccac
6061 cacttcaaga actctgtagc accgcctaca tactcgcgtc tgctaatacct gttaccagt
6121 gctgctgcca gtggcgataa gtcgtgtctt accgggttggt actcaagacg atagttaccg
6181 gataaggcgc agcggctcggg ctgaacgggg ggttcgtgca cacagcccag cttggagcga
6241 acgacctaca ccgaactgag atacctacag cgtgagcatt gagaaagcgc cacgcttccc
6301 gaagggagaa aggcggacag gtatccggta agcggcaggg tcggaacagg agagcgcacg
6361 agggagcttc cagggggaaa cgcttggtat ctttatagtc ctgtcgggtt tcgccacctc
6421 tgacttgagc gtcgattttt gtgatgctcg tcaggggggc ggagcctatg gaaaaacgcc
6481 agcaacgcgg cttttttacg gttcctggcc ttttgctggc cttttgctca catgttcttt
6541 cctgcgttat cccctgattc tgtggataac cgtattaccg cctttgagtg agctgatacc
6601 gctcgcgcga gccgaacgac cgagcgcagc gagtcagtga gcgaggaagc ggaagagcgc
6661 ccaatacgca aaccgcctct cccgcgcgct tggccgattc ataatccag ctggcacgac
6721 aggtttcccg actggaaaag gggcagttag cgcaacgcaa ttaatgtgag ttacctcact
6781 cattaggcac ccaggccttt acactttatg cttccggctc gtatgttgtg tggaattgtg
6841 agcggataac aatttcacac aggaacacgc tatgaccatg attacgaatt cgagctcggg
6901 accggccgca aattaaagcc ttcgagcgtc ccaaaacctt ctcaagcaag gttttcagta
6961 taatgttaca tgcgtacacg cgtctgtaca gaaaaaaaag aaaaatttga aatataaata
7021 acgttcttaa tactaacata actataaaaa aataaatagg gacctagact tcaggttgct
7081 taactccttc cttttcgggt agagcggatg tggggggagg gcgtgaatgt aagcgtgaca
7141 taactaatta catgactcga gttacgcata gtcaggaaca tcgtatgggt accgtcgggt
7201 ctgcgtttca cgcacccggc gcatgtactc ctcctgttac gacccgacat tgtttttccg
7261 atcggcggtt gtcacctgat cgaaggggac gcggtgtgcc agcatctcct ggatctcctt
7321 cgccggcggg tactgcccgg ggcagatcca cgacccaggc ggggtggcaca cgcggtttgc
7381 cgggtcagca acccactttt tgctctccgc gctcacgttc atgaccttga agatgcgcac
7441 caggccgtac ttcgacgagt acacctcctg aaagagggac gggttcacct tcacgccctt
7501 tcttttcccg gcctcgtgca ggttgtacag cagcgacgcc cgcacatcg gcgttgggcg
7561 actgtagtca tttctgtgaa agccgaattg ctggcacagc gggctcgtcg ggcagatgtc
7621 gtgttacacg ctgttgccga tgcgcgccat gtgcggtgac ttcacagggt cgccgctctg
7681 ccagcccag atcaggacgt agtcggccat gtggcgcaac agcagtgctg cctccgccac
7741 gggcgacgtc agcatcttgc cगतcgtggc gatgtgctcg tggttccagg tgttgccatc
7801 ggccagcgag gtgcggttgc cगतcgtgtg gatctgtag ccgtagtccc accaggccaa
7861 aacgcgcgcg tcctctggcg tgctgtcgcg cagccactcg taggccttga ggtagtcatc
7921 caccaatagg ttcataggct tgcctgtggc acggttttgc acgacggccg cgaaaacaat
7981 catcggattt gacgactgct ccgcaaactt tgttgagtgg gacgcgaact cggactgaa
8041 gaagctcacc gcggttgtcg tgacgagagc ccacatagcg atggacagga ccatgcgatg
8101 gccccaagca agagagctac cggcaaaagc gtcgcaaaat gcgcgcgcgg tcgttgcggt
8161 cttagcgtca tctcggccgc tgcccttgcc agccccctc tgggtggcgtt gcgcctgctt
8221 ctgctgcttt ttagcctttg tcgcatcact gtcccagaaa ctgagttgca cggctgcttc
8281 cagaattgtc ccgacgaaaa tgccagtggg cagacacgca gcggggccgg agagaagcag
8341 cagccgagcc atgcgagtgc tgaagttaga cagcgcgccg gatttcagta gccagaatac
8401 cttcgacggg gagtagtgca cgaacgtcga gacagcaagc acaatggagc ccaaaccaca
8461 tgtcacgccg cacacgtgaa gaaaagccca catcgctctc gggctggcgg gttgatgttc
8521 ggcgaccgag tcgaccagcg gattgccagt gcgcgtgtgc tccacgaaca gcgcacgcac
8581 acggaccgag agggggcccga agtaccctgt cggtgccagc accgagattg caagcgcagc
8641 cacgccagcc atcacgctga agacgcgcac cgggatcttg aagtctgcgc gagagcgaac
8701 ctcgacaccg gcgcgtgcgc gcagcacctc gcacacctgc agtccacaca ggaagacaag
8761 caccagcagc gcacccagct gctccagcga cttgaagggc gacatcccca ctggcggcac
8821 gcacacggcg atggcggtgc ccacgacgta gaacagcgtg tatgcacgca gcagcgacgg
8881 gttgtacgtg ttgcgggccc agtccaccat cगतgatatg ccggcatgca tggcaaccat
8941 gttgagcacg aaaatgtagc cggcccacgc tagccgtagg cgacaccggt
9001 gaggacaccg atgggcccag aggaccgcgt gcgcagcgag cgacccagc agtagaaagt
9061 gaggagcatg gctgcgacgg cगतgcactc gttgtcgaa tcaccccgca tggaccgcat

```

cont.

```

9121 caggtgggct gggatgatgg agaaggagag tgcggcagcg gccgccgcta ctgtcgaccc
9181 actggcttcg taggtgcaaa acgccagagt agcggtagcg atggcgccaa accacgctgg
9241 catcagcacg cacacgttgt tgagagacat cggcatgccg gcagccgcca gtgcgcggtg
9301 aatggcgacg gcagtgagct gcaggccccg gtacgtggtg gagccgacgg ggcggccccag
9361 cgggtaccag ctcagttagt cgaaccagct gaagaaggcg gaccagccgt gcgtggacat
9421 gtactcggca gcgcggtagt tgaaccacgg gtcgaactcg tggatcaggt atccgtaaat
9481 ctgaacggag atcatgcgaa ccgtgaaggc ttggaagcag ctggcggcta agacgaagag
9541 tgccaccacg gtgaggacga agtgtagctg ccagaaagga aacggaaaga tgccgatgaa
9601 atcctttctca tctgttagcg ttttgggcag caaaatcacc ttcgccggtg gagatgcggt
9661 cttggtttgt gaggcgccat cttcggcttg gccgaagct tcacgagatg cggttgccgc
9721 agagccggag tcgccaatg aatttccctt ccgcttgccc atgttggttac tagttctaga
9781 atccgtcgaa actaagttct ggtgttttaa aactaaaaaa aagactaact ataaaagtag
9841 aatttaagaa gtttaagaaa tagatttaca gaattacaat caatacctac cgtctttata
9901 tactttattg tcaagtaggg gaataatttc agggaactgg tttcaacctt ttttttcagc
9961 tttttccaaa tcagagagag cagaaggtaa tagaagggtg aagaaaatga gatagataga
10021 tgcgtgggtc aattgccttg tgtcatcatt tactccaggc aggttgcatc actccattga
10081 ggttgtgccc gttttttgcc tgtttgtgcc cctgttctct gtagttgcgc taagagaatg
10141 gacctatgaa ctgatggttg gtgaagaaaa caatatattg gtgctgggat tctttttttt
10201 tctggatgcc agcttaaaaa gcgggctcca ttatatattg tggatgccag gaataaactg
10261 ttcaccacga cacctacgat gttatatatt ctgtgtaacc cgccccctat tttgggcatg
10321 tacgggttac agcagaatta aaaggctaatt tttttgacta aataaagtta ggaaaatcac
10381 tactattaat tatttacgta ttctttgaaa tggcgagtat tgataatgat aaactgaggg
10441 ggatcctcta gatcgacct gcaggcatgc aagct

```

图 14

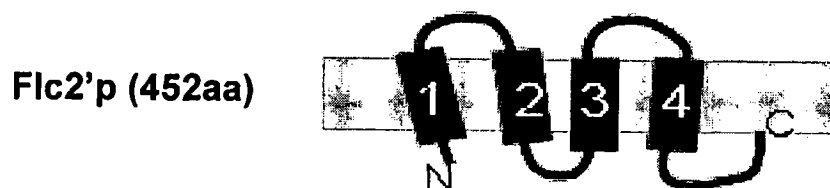


图 15A

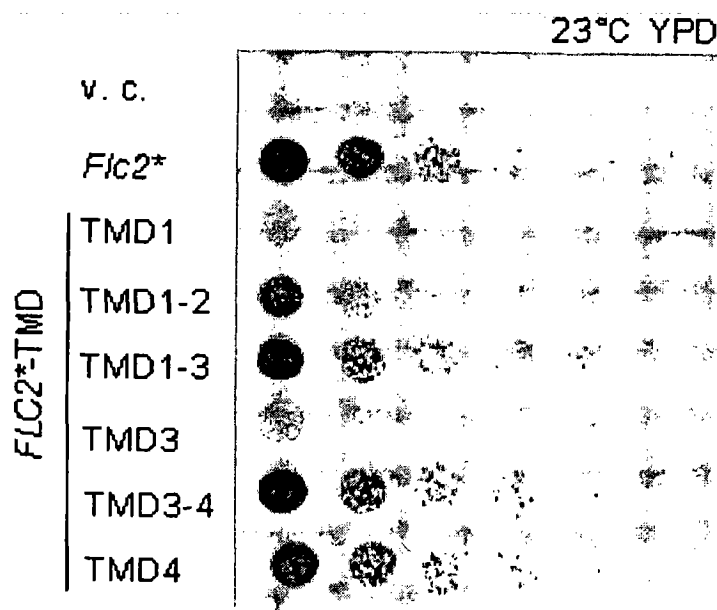


图 15B

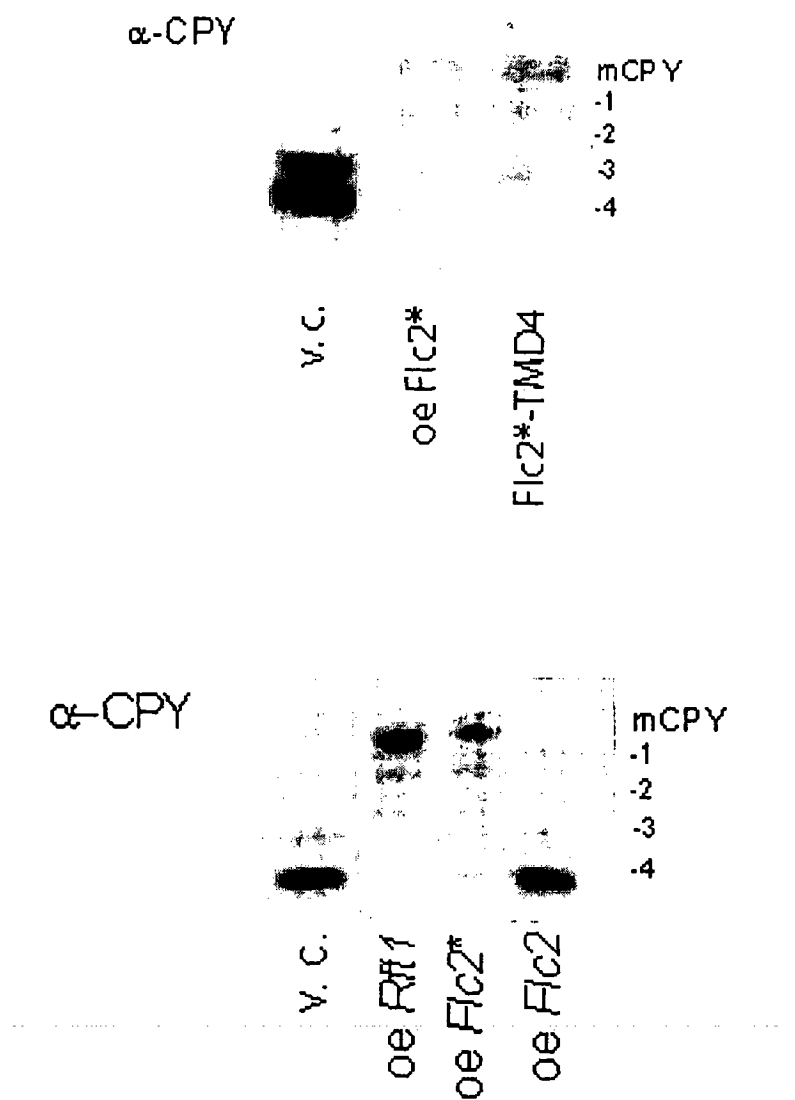


图 16A

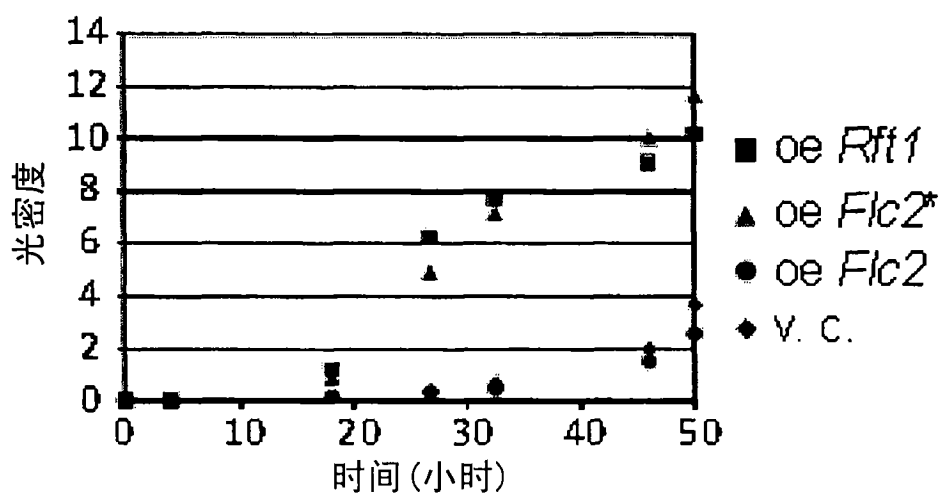


图 16B

$\Delta alg11$

V. C.

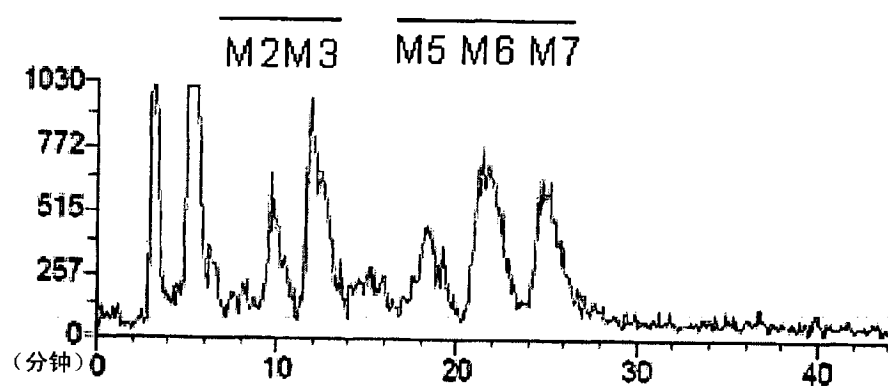


图 17A

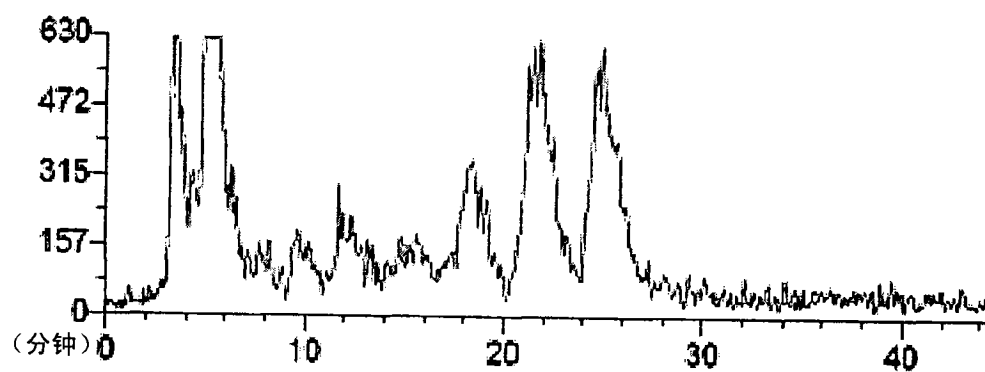
oe *RFT1*

图 17B

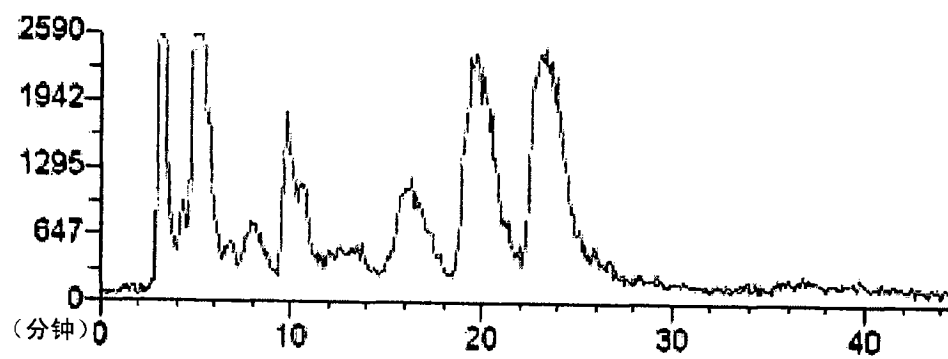
oe *FLC2**

图 17C

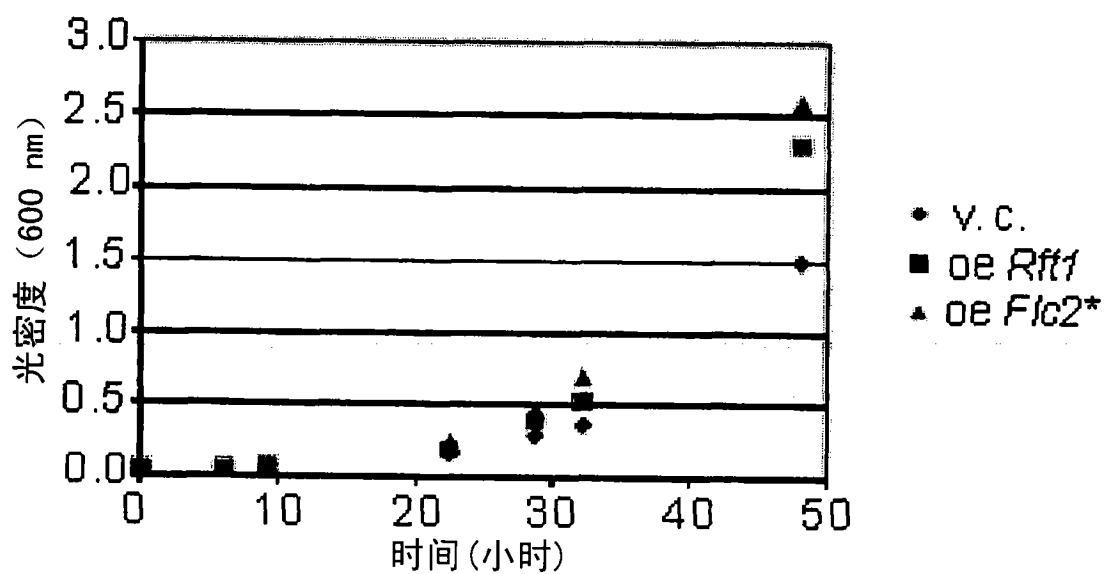


图 18A

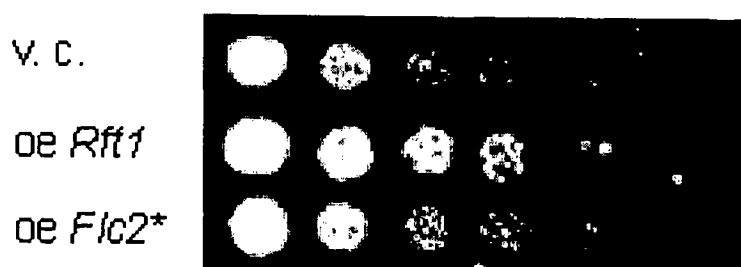


图 18B

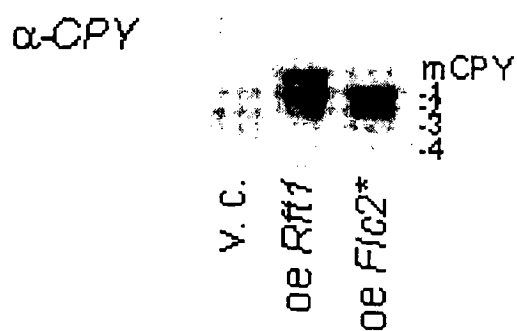


图 18C

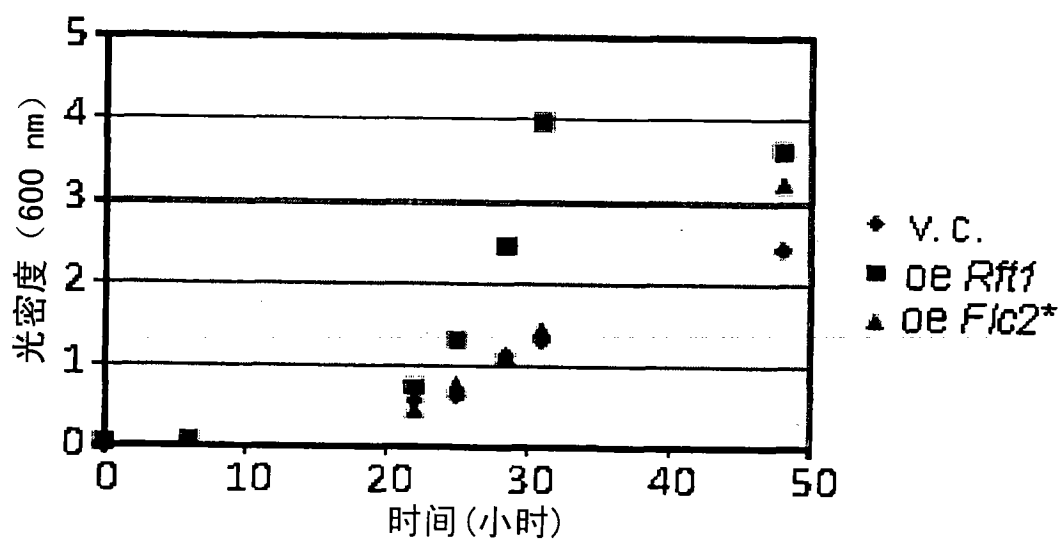


图 19A

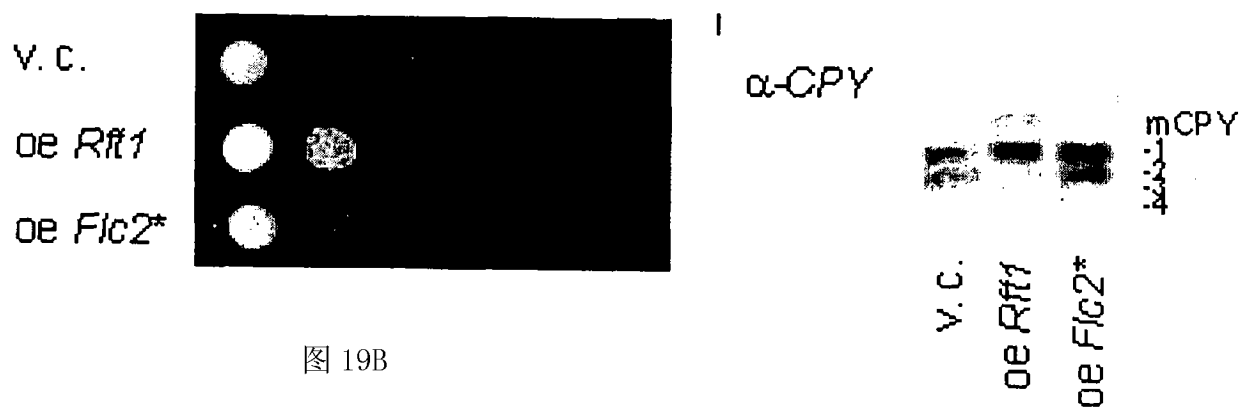


图 19B

图 19C

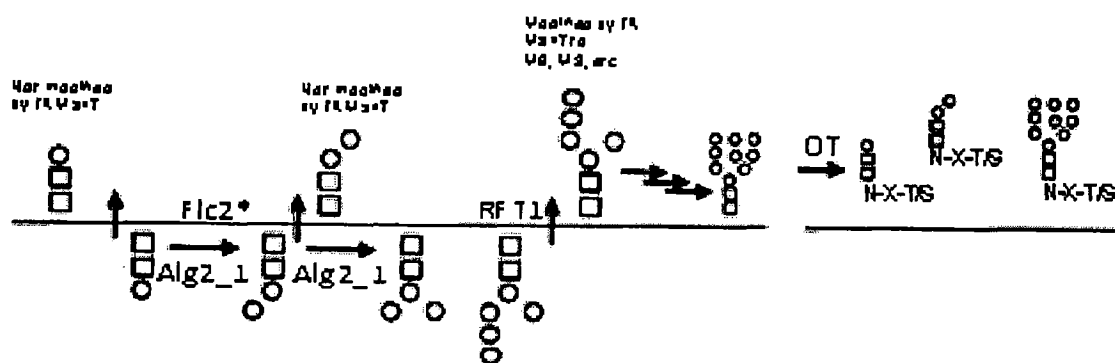


图 20A

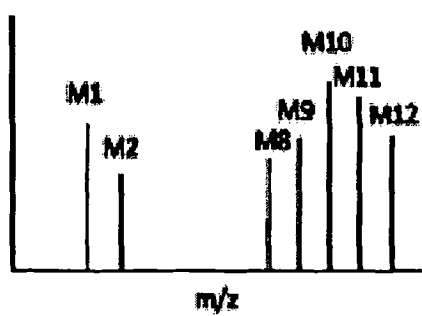


图 20B

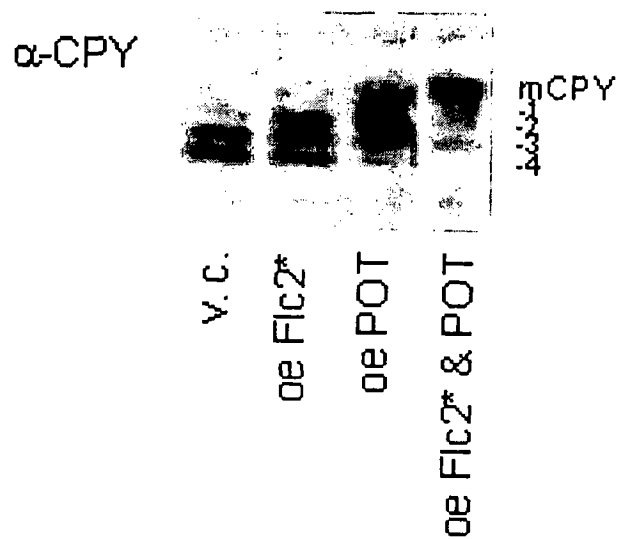


图 21A

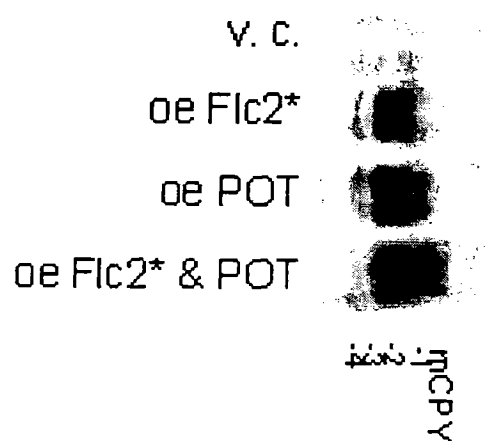


图 21B

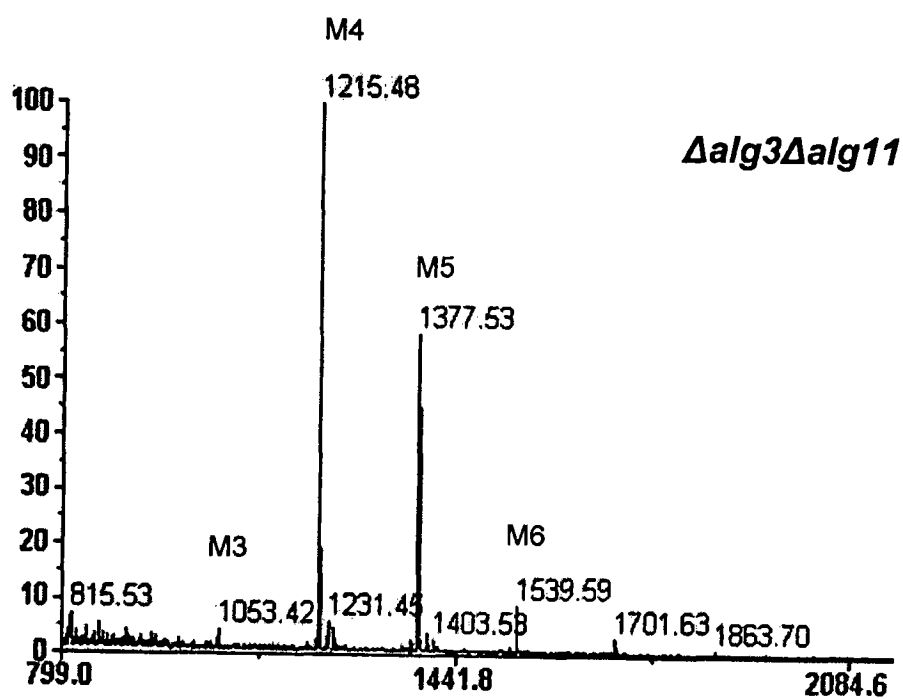


图 22A

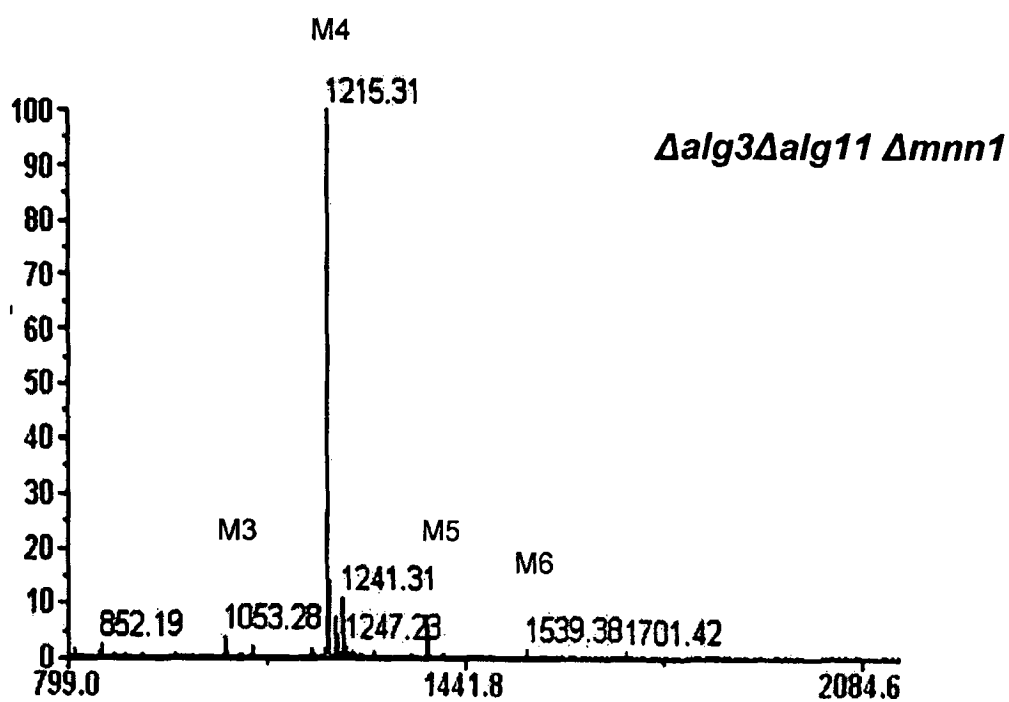


图 22B