



## (12)发明专利

(10)授权公告号 CN 104480140 B

(45)授权公告日 2019.12.31

(21)申请号 201410681757.1

C12N 9/12(2006.01)

(22)申请日 2008.04.03

C12N 9/10(2006.01)

(65)同一申请的已公布的文献号

C07K 14/00(2006.01)

申请公布号 CN 104480140 A

C07K 16/00(2006.01)

(43)申请公布日 2015.04.01

C12N 1/19(2006.01)

(30)优先权数据

C07K 2/00(2006.01)

60/909,904 2007.04.03 US

A61K 38/00(2006.01)

60/940,212 2007.05.25 US

A61K 38/02(2006.01)

(62)分案原申请数据

A61K 38/16(2006.01)

200880018736.4 2008.04.03

A61K 38/46(2006.01)

(73)专利权人 奥克西雷恩英国有限公司

A61P 3/00(2006.01)

地址 英国曼彻斯特

(续)

(72)发明人 尼科.L.M.卡尔沃尔特

沃特.弗维肯

(续)

(74)专利代理机构 北京市柳沈律师事务所

11105

代理人 张文辉

(51)Int.Cl.

C12N 15/81(2006.01)

C12N 9/96(2006.01)

C12N 9/38(2006.01)

C12N 9/40(2006.01)

C12N 9/26(2006.01)

C12N 9/24(2006.01)

C12N 9/20(2006.01)

C12N 9/16(2006.01)

(56)对比文件

WO 03056914 A1,2003.07.17,

WO 2007035930 A2,2007.03.29,

Yasunori Chiba等.Production in yeast of  $\alpha$ -galactosidase A, a lysosomal enzyme applicable to enzyme replacement therapy for Fabry disease.《Glycobiology》.2002, 821-828.Stephanie Barnay-Verdier等.Identification and characterization of two  $\alpha$ -1,6-mannosyltransferases, An11p and Och1p,in the yeast Yarrowia lipolytica..《Microbiology》.2004,2185-2195.

审查员 徐俊

权利要求书4页 说明书61页

序列表23页 附图53页

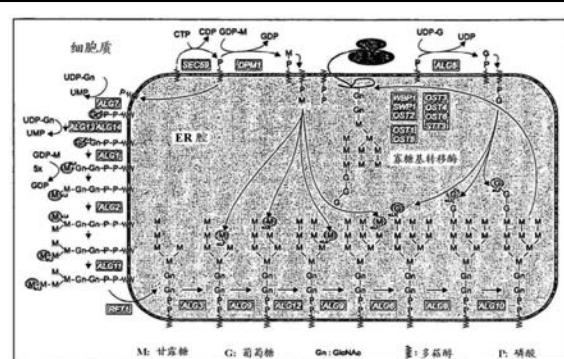
(54)发明名称

分子的糖基化

(57)摘要

本文描述的是可用于产生目标分子的改变的N-糖基化形式的方法和遗传工程化细胞。还描述用于治疗多种病症例如代谢紊乱的方法和N-糖基化改变的分子。

CN 104480140 B



[转续页]

[接上页]

(72)发明人 卡伦.J.马塞尔德波尔克  
    史蒂文.C.J.盖森斯  
    芒娜.古尔法尔  
  
(51)Int.Cl.  
    *A61P 35/00(2006.01)*

*A61P 29/00(2006.01)*  
*A61P 37/00(2006.01)*  
*C12P 21/00(2006.01)*  
*C12R 1/645(2006.01)*

1. 产生目标蛋白的改变的N-糖基化形式的方法,所述方法包括:提供解脂西洋蓍霉细胞,所述细胞经遗传工程化而包含MNN4的过表达;和向细胞中引入编码目标蛋白的核酸,其中所述细胞以改变的糖基化形式产生所述目标蛋白。

2. 权利要求1的方法,其中所述细胞进一步遗传工程化而包含外链的延伸OCH1活性的缺陷,其中所述进一步遗传工程化的细胞包含:

-编码具有所述OCH1活性的蛋白的基因的缺失;或

-引入或表达RNA分子,所述RNA分子干扰具有所述OCH1活性的蛋白的功能性表达。

3. 权利要求1的方法,进一步包含分离所述目标蛋白的改变的N-糖基化形式。

4. 权利要求1的方法,其中所述目标蛋白是外源蛋白。

5. 权利要求1的方法,其中所述目标蛋白是内源蛋白。

6. 权利要求1的方法,其中所述目标蛋白是哺乳动物蛋白。

7. 权利要求6的方法,其中所述目标蛋白是人蛋白。

8. 权利要求1的方法,其中所述目标蛋白是病原体蛋白、溶酶体蛋白、细胞因子、抗体或其抗原结合片段,或融合蛋白。

9. 权利要求8的方法,其中所述融合蛋白是病原体蛋白、溶酶体蛋白、或细胞因子与抗体或其抗原结合片段的融合物。

10. 权利要求1的方法,其中所述目标蛋白是生长因子或趋化因子,或生长因子或趋化因子与抗体或其抗原结合片段的融合物。

11. 权利要求1的方法,其中所述目标蛋白是与溶酶体贮积病相关的蛋白。

12. 权利要求1的方法,其中所述目标蛋白是葡糖脑苷酯酶、 $\alpha$ -半乳糖苷酶或半乳糖脑苷酯酶。

13. 权利要求1的方法,其中所述目标蛋白选自下组: $\alpha$ -L-艾杜糖苷酸酶、 $\beta$ -D-半乳糖苷酶、 $\beta$ -葡糖苷酶、 $\beta$ -己糖胺酶、 $\beta$ -D-甘露糖苷酶、 $\alpha$ -L-岩藻糖苷酶、芳基硫酸酯酶B、芳基硫酸酯酶A、 $\alpha$ -N-乙酰半乳糖胺酶、天冬氨酰葡萄糖胺酶、艾杜糖醛酸-2-硫酸酯酶、 $\alpha$ -氨基葡糖苷-N-乙酰转移酶、 $\beta$ -D-葡糖醛酸酶、透明质酸酶、 $\alpha$ -L-甘露糖苷酶、 $\alpha$ -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、硫酯酶、组织蛋白酶K和脂蛋白脂肪酶。

14. 权利要求1的方法,其中所述细胞进一步遗传工程化而包含至少一种额外的经修饰的N-糖基化活性。

15. 权利要求14的方法,其中所述至少一种额外的经修饰的N-糖基化活性包含N糖基化活性的缺陷。

16. 权利要求15的方法,其中所述缺陷是ALG3活性中的缺陷。

17. 权利要求14的方法,其中所述至少一种额外的经修饰的N-糖基化活性包含缺失编码具有N-糖基化活性的蛋白质的基因。

18. 权利要求14的方法,其中所述至少一种额外的经修饰的N-糖基化活性包含引入或表达RNA分子,所述RNA分子干扰具有N-糖基化活性的蛋白质的功能性表达。

19. 权利要求14的方法,其中所述至少一种额外的经修饰的N-糖基化活性包含具有N-糖基化活性的蛋白质的表达。

20. 权利要求19的方法,其中表达的蛋白质是由细胞中的外源核酸编码的蛋白质。

21. 权利要求19的方法,其中表达的具有N-糖基化活性的蛋白质是 $\alpha$ -甘露糖苷酶。

22. 权利要求1的方法,其中所述细胞进一步遗传工程化以包含至少两种额外的经修饰的N-糖基化活性。

23. 权利要求19-21中任一项的方法,其中所述具有N-糖基化活性的蛋白质是外源蛋白。

24. 权利要求19-21中任一项的方法,其中所述具有N-糖基化活性的蛋白质是哺乳动物蛋白。

25. 权利要求24的方法,其中所述具有N-糖基化活性的蛋白质是人蛋白。

26. 权利要求19-21中任一项的方法,其中所述具有N-糖基化活性的蛋白质是低等真核生物蛋白。

27. 权利要求26的方法,其中所述低等真核生物是真菌、锥虫或原生动物。

28. 权利要求26的方法,其中所述低等真核生物选自下组:布氏锥虫、哈茨木霉和曲霉属。

29. 权利要求1的方法,进一步包括对目标蛋白的改变的糖基化形式的额外加工。

30. 权利要求29的方法,其中所述额外加工发生在体外。

31. 权利要求29的方法,其中所述额外加工发生在体内。

32. 权利要求29的方法,其中所述额外加工包括向目标蛋白的改变的糖基化形式添加异源模块。

33. 权利要求32的方法,其中所述异源模块是聚合物或载体。

34. 权利要求29的方法,其中所述额外加工包括对所述目标蛋白的改变的N-糖基化形式进行酶处理或化学处理。

35. 权利要求34的方法,其中所述额外加工包括用甘露糖苷酶、甘露聚糖酶、磷酸二酯酶、葡糖苷酶或糖基转移酶处理所述目标蛋白的改变的N-糖基化形式。

36. 权利要求34的方法,其中所述额外加工包括用氢氟酸处理所述目标蛋白的改变的N-糖基化形式。

37. 权利要求29的方法,其中所述额外加工包括所述目标蛋白的改变的N-糖基化形式的磷酸化。

38. 具有改变的N-糖基化的分离的蛋白质,其中所述蛋白质是通过权利要求1或3-37中任一项的方法产生的且其中目标蛋白的至少80%的N-聚糖是磷酸化的。

39. 药物组合物,其包含药学上可接受的载体或赋形剂,以及具有N-糖基化的分离的蛋白,其中所述蛋白是通过权利要求1或3-37中任一项的方法产生的且其中目标蛋白的至少80%的N-聚糖是磷酸化的。

40. 分离的解脂西洋蓍霉细胞,所述细胞经遗传工程化而包含MNN4的过表达。

41. 权利要求40的分离的解脂西洋蓍霉细胞,其中所述细胞进一步遗传工程化而包含外链的延伸OCH1活性的缺陷,其中所述进一步遗传工程化的细胞包含:

-缺失编码具有OCH1活性的蛋白的基因;或

-引入或表达RNA分子,所述RNA分子干扰具有OCH1活性的蛋白的功能性表达。

42. 权利要求40的分离的解脂西洋蓍霉细胞,其中所述细胞进一步遗传工程化以包含至少一种额外的经修饰的N-糖基化活性。

43. 权利要求42的分离的解脂西洋蓍霉细胞,其中所述至少一种额外的经修饰的N-糖

基化活性包含N-糖基化活性的缺陷。

44. 权利要求43的分离的解脂西洋蓍霉细胞,其中所述缺陷是ALG3活性中的缺陷。

45. 权利要求42的分离的解脂西洋蓍霉细胞,其中所述修饰包含缺失编码具有N-糖基化活性的蛋白质的基因。

46. 权利要求42的分离的解脂西洋蓍霉细胞,其中所述修饰包含引入或表达RNA分子,所述RNA分子干扰具有N-糖基化活性的蛋白质的功能性表达。

47. 权利要求42的分离的解脂西洋蓍霉细胞,其中所述额外的经修饰的N-糖基化活性包含表达具有N-糖基化活性的蛋白质。

48. 权利要求47的分离的解脂西洋蓍霉细胞,其中所述表达的具有N-糖基化活性的蛋白质是 $\alpha$ -甘露糖苷酶。

49. 权利要求40的分离的解脂西洋蓍霉细胞,其中所述细胞进一步遗传工程化以包含至少两种额外的经修饰的N-糖基化活性。

50. 权利要求47-48中任一项的分离的解脂西洋蓍霉细胞,其中所述具有N-糖基化活性的蛋白质是外源蛋白。

51. 权利要求47-48中任一项的分离的解脂西洋蓍霉细胞,其中所述具有N-糖基化活性的蛋白质是哺乳动物蛋白。

52. 权利要求47-48中任一项的分离的解脂西洋蓍霉细胞,其中所述具有N-糖基化活性的蛋白质是人蛋白。

53. 权利要求47-48中任一项的分离的解脂西洋蓍霉细胞,其中所述具有N-糖基化活性的蛋白质是低等真核生物蛋白。

54. 权利要求53的分离的解脂西洋蓍霉细胞,其中所述低等真核生物是真菌、锥虫或原虫。

55. 权利要求53的分离的解脂西洋蓍霉细胞,其中所述低等真核生物选自下组:布氏锥虫、哈茨木霉和曲霉菌属。

56. 根据权利要求38的蛋白质或根据权利要求39的组合物在制备用于治疗病症的药物中的用途,所述病症可通过向受试者施用所述蛋白质来治疗,其中所述受试者是患有,或疑似患有通过施用所述蛋白质治疗的疾病的受试者。

57. 权利要求56的用途,其中所述病症是代谢紊乱。

58. 权利要求57的用途,其中所述代谢紊乱是溶酶体贮积病。

59. 权利要求58的用途,其中溶酶体贮积病是高歇尔氏病、泰-沙二氏病、旁姆普氏病、尼-皮二氏病或法布里氏病。

60. 权利要求56-59中任一项的用途,其中所述蛋白质是与溶酶体贮积病相关的蛋白质。

61. 权利要求60的用途,其中所述蛋白质是葡糖脑苷酯酶、半乳糖脑苷酯酶或 $\alpha$ -半乳糖苷酶。

62. 权利要求60的用途,其中所述蛋白质选自下组: $\alpha$ -L-艾杜糖苷酸酶、 $\beta$ -D-半乳糖苷酶、 $\beta$ -葡糖苷酶、 $\beta$ -己糖胺酶、 $\beta$ -D-甘露糖苷酶、 $\alpha$ -L-岩藻糖苷酶、芳基硫酸酯酶B、芳基硫酸酯酶A、 $\alpha$ -N-乙酰半乳糖胺酶、天冬氨酰葡萄糖胺酶、艾杜糖醛酸-2-硫酸酯酶、 $\alpha$ -氨基葡糖苷-N-乙酰转移酶、 $\beta$ -D-葡糖醛酸酶、透明质酸酶、 $\alpha$ -L-甘露糖苷酶、 $\alpha$ -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、硫酯酶、组织蛋白酶K和脂蛋白脂肪酶。

63. 权利要求56-59中任一项的用途,其中所述受试者已确定患有可通过施用具有改变的N-糖基化的蛋白质来治疗的疾病。

64. 权利要求56-59中任一项的用途,其中所述受试者是人。

65. 产生目标蛋白的改变的N-糖基化形式的方法,所述方法包括使目标蛋白与制备自解脂西洋蓍霉细胞的细胞裂解物接触,所述细胞经遗传工程化而包含MNN4的过表达,其中使所述目标蛋白与所述细胞裂解物接触导致该目标蛋白的改变的N-糖基化形式。

66. 产生目标蛋白的改变的N-糖基化形式的方法,所述方法包括使目标蛋白与一种或多种具有N-糖基化活性的蛋白质接触,其中所述一种或多种具有N-糖基化活性的蛋白质获得自解脂西洋蓍霉细胞,所述细胞经遗传工程化而包含MNN4的过表达,并且其中使所述目标蛋白与所述一种或多种具有N-糖基化活性的蛋白质接触导致该目标蛋白的改变的N-糖基化形式。

67. 解脂西洋蓍霉细胞的培养物,所述解脂西洋蓍霉细胞经遗传工程化而包含MNN4的过表达。

## 分子的糖基化

[0001] 本申请是基于申请日为2008年4月3日，优先权日为2007年4月3日，申请号为200880018736.4 (PCT/IB2008/001814)，发明名称为：“分子的糖基化”的专利申请的分案申请。

### 技术领域

[0002] 本发明涉及获得糖基化的分子，特别是蛋白质和脂质分子的方法。

### 背景

[0004] 高效表达系统是生产大多数目前正在开发中的生物药物(例如，重组蛋白)所需要的。许多这些生物药物的生物学活性依赖于它们的修饰(例如，磷酸化或糖基化)。一种基于酵母的表达系统将遗传操作和微生物发酵的简易性与分泌和修饰蛋白的能力相结合。然而，酵母细胞中产生的重组糖蛋白主要展现为异质性的高甘露糖和超甘露糖聚糖结构 (high-mannose and hyper-mannose glycan structures)，其可能对蛋白质功能、下游加工和后续治疗用途有害，特别是在糖基化扮演重要的生物学角色的情况下。

### 概述

[0006] 本发明至少部分基于：(a)发现解脂西洋蓍霉 (*Yarrowia lipolytica*) 细胞中的单个基因缺失(外部链延长 (Outer CHain elongation) (OCH1) 缺失) 导致基本上均质产生在  $\text{Man}_5\text{GlcNAc}_2$  (结构式IV；图1) 骨架上具有 $\alpha$ -1,2-连接的甘露糖残基的糖基化蛋白；(b)发现靶向解脂西洋蓍霉细胞(具有和不具有OCH1缺失的两种细胞)ER的经工程化的 $\alpha$ -1,2-甘露糖苷酶的过表达导致基本上均质产生具有  $\text{Man}_5\text{GlcNAc}_2\text{N}-$ 聚糖结构 (结构式IV；图1) 的糖基化蛋白；(c)发现使解脂西洋蓍霉细胞中天冬酰胺连接的糖基化3 (ALG3) 酶活性失活导致糖基化聚糖水平的高度增加；和(d)发现解脂西洋蓍霉细胞中解脂西洋蓍霉基因 (MNN4) 的过表达导致 $\alpha$ -1,2-连接的甘露糖残基的超磷酸化 (hyperphosphorylation)。如此，遗传工程化的细胞(例如，解脂西洋蓍霉 (*Yarrowia lipolytica*)、*Arxula adeninivorans*，或其它相关种类的双态性酵母细胞)可以在产生如下目标分子的方法中使用，所述目标分子如与在未经遗传工程化的相同种类的细胞中产生的目标分子的N-糖基化形式相比具有改变的N-糖基化形式。向患有代谢紊乱(例如，溶酶体贮积病 (lysosomal storage disorder))的患者施用N-糖基化的目标分子(例如，N-糖基化蛋白)已被证明可改善所述病症的症状，所描述的方法和细胞可用于制备N-糖基化的目标分子用于治疗，包括但不限于，代谢紊乱，例如溶酶体贮积病。

[0007] 本发明还至少部分基于解脂西洋蓍霉和巴斯德毕赤酵母 (*Pichia pastoris*) HAC1 基因的剪接形式的发现。由HAC1基因编码的蛋白质Hac1p是转录激活因子，其通过与称为未折叠蛋白质应答 (Unfolded Protein Response) (UPR) 元件的DNA序列基序结合而激活数个靶基因的转录。在Hac1p靶基因之中包括编码伴侣蛋白、折叠酶 (foldase) 和负责脂质和肌醇代谢的蛋白质的那些。由于经剪接形式Hac1p是比由未经剪接的HAC1mRNA所编码的形式更强烈的转录激活因子，所以Hac1p转录因子的经剪接形式的过表达能够导致天然和异源蛋白表达水平的增加，以及在ER膜中的增加。由此，Hac1p的经剪接形式能够用于在多种真

核细胞(例如,真菌细胞(例如,解脂西洋蓍霉或任何其它本文描述的酵母细胞)、植物细胞或动物细胞(例如,哺乳动物细胞如人的细胞)中通过同时激活UPR和表达目标分子来增加膜蛋白和分泌蛋白的产生。

[0008] 本发明还基于这样的发现,即在解脂西洋蓍霉中表达时,MNS1甘露糖苷酶的突变形式能够将 $\text{Man}_8\text{GlcNAc}_2$ (结构式I;图4)结构转化成 $\text{Man}_5\text{GlcNAc}_2$ (结构式IV;图4)、 $\text{Man}_6\text{GlcNAc}_2$ (结构式V;图4)和 $\text{Man}_7\text{GlcNAc}_2$ (结构式VI;图4)。由此,表达甘露糖苷酶如MNS1的突变形式的遗传工程化的真核细胞(例如,真菌细胞(例如,解脂西洋蓍霉或任何其它本文描述的酵母细胞)、植物细胞或动物细胞(例如,哺乳动物细胞如人的细胞))能够在产生如下目标分子的方法中使用,所述目标如与在未经遗传工程化的相同种类的细胞中产生的目标分子的N-糖基化形式相比具有改变的N-糖基化形式。因此,所描述的方法和细胞可用于制备N-糖基化的目标分子用于治疗,包括但不限于,代谢紊乱,例如溶酶体贮积病(见下文)。

[0009] 在一个方面,本公开描述了产生目标蛋白的改变的N-糖基化形式的方法。所述方法包括向细胞引入编码目标蛋白的核酸的步骤,其中所述细胞产生N-糖基化形式改变的目标蛋白,并且其中所述细胞是经遗传工程化而含有至少一种经修饰的N-糖基化活性的解脂西洋蓍霉或Arxula adeninivorans细胞(或相关种类的双态性的酵母细胞)。所述方法也可包括提供经遗传工程化而含有至少一种经修饰的N-糖基化活性的解脂西洋蓍霉或Arxula adeninivorans细胞(或相关种类的双态性的酵母细胞)的步骤。所述方法也可包括分离目标蛋白的改变的N-糖基化形式的步骤。

[0010] 在一些实施方案中,目标蛋白可以是内源蛋白或外源蛋白。目标蛋白可以是哺乳动物蛋白如人的蛋白。目标蛋白可以是例如,病原体蛋白、溶酶体蛋白、生长因子、细胞因子、趋化因子、抗体或其抗原结合片段,或融合蛋白。融合蛋白可以是例如,病原体蛋白、溶酶体蛋白、生长因子、细胞因子或趋化因子与抗体或其抗原结合片段的融合物。目标蛋白可以是例如,与溶酶体贮积病(LSD)相关的蛋白质。目标蛋白可以是例如,葡糖脑苷酯酶、半乳糖脑苷酯酶、 $\alpha$ -L-艾杜糖苷酸酶、 $\beta$ -D-半乳糖苷酶、 $\beta$ -葡糖苷酶、 $\beta$ -己糖胺酶、 $\beta$ -D-甘露糖苷酶、 $\alpha$ -L-岩藻糖苷酶、芳基硫酸酯酶B、芳基硫酸酯酶A、 $\alpha$ -N-乙酰半乳糖胺酶( $\alpha$ -N-acetylgalactosaminidase)、天冬氨酰葡萄糖胺酶(aspartylglucosaminidase)、艾杜糖醛酸-2-硫酸酯酶、 $\alpha$ -氨基葡萄糖-N-乙酰转移酶、 $\beta$ -D-葡萄糖醛酸酶( $\beta$ -D-glucuronidase)、透明质酸酶、 $\alpha$ -L-甘露糖苷酶、 $\alpha$ -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(sphingomyelinase)、硫酯酶、组织蛋白酶K或脂蛋白脂肪酶。

[0011] 在一些实施方案中,改变的N-糖基化形式可以含有一种或多种N-聚糖结构如,例如, $\text{Man}_5\text{GlcNAc}_2$ 、 $\text{Man}_8\text{GlcNAc}_2$ 、 $\text{Man}_9\text{GlcNAc}_2$ 、 $\text{Man}_3\text{GlcNAc}_2$ 、 $\text{Glc}_1\text{Man}_5\text{GlcNAc}_2$ 、 $\text{Glc}_2\text{Man}_5\text{GlcNAc}_2$ 。在一些实施方案中,改变的糖基化可以是,例如, $\text{Man}_5\text{GlcNAc}_2$ 、 $\text{Man}_8\text{GlcNAc}_2$ 、 $\text{Man}_9\text{GlcNAc}_2$ 、 $\text{Man}_3\text{GlcNAc}_2$ 、 $\text{Glc}_1\text{Man}_5\text{GlcNAc}_2$ 、 $\text{Glc}_2\text{Man}_5\text{GlcNAc}_2$ 。

[0012] 在一些实施方案中,目标蛋白的改变的N-糖基化形式可以是均质的或基本上均质的。例如,含有改变的糖基化的改变的目标分子的分数可以是至少约20%,至少约30%,至少约40%,至少约45%,至少约50%,至少约55%,至少约60%,至少约65%,至少约70%,至少约75%,至少约80%,至少约85%,至少约90%,或至少约95%或更多。

[0013] 在一些实施方案中,细胞可以经遗传工程化而缺乏至少一种N-糖基化活性。所述

N-糖基化活性可以是，例如，ALG3活性、OCH1活性、MNS1活性或MNN9活性。

[0014] 在一些实施方案中，至少一种修饰可以是：(a) 缺失编码具有N-糖基化活性的蛋白质的基因；(b) 表达具有N-糖基化活性的蛋白质的突变形式；(c) 引入或表达干扰具有N-糖基化活性的蛋白质的功能性表达的RNA分子；(d) 表达具有N-糖基化活性的蛋白质（例如ALG6或 $\alpha$ -甘露糖苷酶（例如，靶向内质网的 $\alpha$ -甘露糖苷酶）。表达的蛋白质可以是由细胞中的内源核酸编码的蛋白质。表达的蛋白质可以是最适pH在7.5以下的（例如，最适pH在5.1以下）的 $\alpha$ -甘露糖苷酶。具有N-糖基化活性的蛋白质可以是外源蛋白。具有N-糖基化活性的蛋白质可以是哺乳动物蛋白（如人的蛋白）或低等真核生物的（例如，真菌、原生动物或锥虫）蛋白。低等真核生物可以选自下组：布氏锥虫（*Typanosoma brucei*）、哈茨木霉（*Trichoderma harzianum*）、曲霉（*Aspergillus*），和任何其它本文描述的低等真核生物。

[0015] 在一些实施方案中，N-糖基化活性可以是葡萄糖基转移酶活性。在一些实施方案中，具有N-糖基化活性的蛋白质是ALG6或 $\alpha$ -甘露糖苷酶。 $\alpha$ -甘露糖苷酶可以靶向内质网。例如，具有N-糖基化活性的蛋白质可以是融合蛋白，其包含 $\alpha$ -甘露糖苷酶多肽和HDEL内质网保留肽（HDEL endoplasmic reticulum retention peptide）。

[0016] 在一些实施方案中，具有N-糖基化活性的蛋白质可以是能够从Man<sub>5</sub>GlcNAc<sub>2</sub>去除葡萄糖残基的蛋白质。例如，具有N-糖基化活性的蛋白质可以是具有 $\alpha$ -1,3-葡糖苷酶活性的蛋白质，例如，但不限于，葡糖苷酶II（例如，葡糖苷酶II的 $\alpha$ 和 $\beta$ 亚基之一或二者）或变聚糖酶（mutanase）。

[0017] 在一些实施方案中，细胞可以经遗传工程化而包含至少两种经修饰的N糖基化活性，例如任何本文描述的经修饰的N-糖基化活性。所述至少两种经修饰的N-糖基化活性可以包含，例如，ALG3活性的缺乏和升高的ALG6活性水平。

[0018] 在一些实施方案中，细胞可以经遗传工程化而包含至少三种经修饰的N-糖基化活性，例如任何本文描述的经修饰的N-糖基化活性。所述至少三种经修饰的N-糖基化活性可以包含，例如，ALG3活性的缺乏；升高的ALG6活性水平；和升高的葡糖苷酶II活性水平。

[0019] 在一些实施方案中，细胞未经遗传工程化而缺乏OCH1活性。

[0020] 在一些实施方案中，修饰可以包含表达能够影响目标蛋白的甘露糖基磷酸化的蛋白质或其生物活性的变体。所述能够影响目标蛋白的甘露糖基磷酸化的蛋白质或其生物活性的变体可以是MNN4、PN01或MNN6。在一些实施方案中，糖蛋白的至少约30%的甘露糖基残基可以是磷酸化的。

[0021] 在一些实施方案中，所述方法可进一步包括对糖蛋白的额外加工。额外加工可以发生在体外或体内。额外加工可以包括将异源模块（moiety）添加到修饰的糖蛋白。异源模块可以是聚合物或载体。额外加工可以包括对目标蛋白的改变的N-糖基化形式酶处理或化学处理。例如，额外加工可以包括用甘露糖苷酶、甘露聚糖酶、磷酸二酯酶、葡糖苷酶或糖基转移酶处理目标蛋白的改变的N-糖基化形式。额外加工可以包括用氢氟酸处理目标蛋白的改变的N-糖基化形式。额外加工可以包括目标蛋白的改变的N-糖基化形式的磷酸化。

[0022] 在另一方面，本公开提供产生目标蛋白的改变的N-糖基化形式的方法。本方法包括如下步骤：提供经遗传工程化而包含至少一种经修饰的N-糖基化活性的真核细胞（例如，真菌细胞、植物细胞或动物细胞）；和向细胞引入编码目标蛋白的核酸，其中所述细胞产生N-糖基化形式改变的目标蛋白。

[0023] 在另一方面,本公开描述了产生目标蛋白的改变的N-糖基化形式的方法。所述方法包括将目标蛋白与从解脂西洋蓍霉或Arxula adeninivorans细胞制备的细胞裂解物接触的步骤,所述解脂西洋蓍霉或Arxula adeninivorans细胞经遗传工程化而包含至少一种经修饰的N-糖基化活性,其中将目标蛋白与所述细胞裂解物接触导致目标蛋白的改变的N-糖基化形式。

[0024] 在又一方面,本公开描述产生目标蛋白的改变的N-糖基化形式的方法,所述方法包括将目标蛋白与一种或多种具有N-糖基化活性的蛋白质接触的步骤,其中所述一种或多种具有N-糖基化活性的蛋白质获得自经遗传修饰而包含至少一种经修饰的N-糖基化活性的解脂西洋蓍霉或Arxula adeninivorans细胞,并且其中将目标分子与一种或多种具有N-糖基化活性的蛋白质接触导致目标蛋白的改变的N-糖基化形式。

[0025] 在另一方面,本公开提供N-糖基化改变的分离的蛋白质,其中所述蛋白质由任一上述方法产生。

[0026] 在又一方面,本公开提供分离的经遗传工程化而包含至少一种经修饰的N-糖基化活性的解脂西洋蓍霉或Arxula adeninivorans细胞(或其它相关种类的双态性的酵母细胞)。所述N-糖基化活性可以是,例如,ALG3活性、OCH1活性、MNS1活性或MNN9活性。所述修饰可以是任何本文所述的那些修饰。例如,所述修饰可以包括:(a)缺失编码具有N-糖基化活性的蛋白质的基因;(b)表达具有N-糖基化活性的蛋白质的突变形式;(c)引入或表达干扰具有N-糖基化活性的蛋白质的功能性表达的RNA分子;或(d)表达具有N-糖基化活性的蛋白质。具有N-糖基化活性的蛋白质可以是例如,ALG6。具有N-糖基化活性的蛋白质可以是哺乳动物蛋白,例如人的蛋白。所述修饰也可以包括表达能够促进经修饰糖蛋白的甘露糖基磷酸化的蛋白质(例如,MNN4或PN01)或其生物学活性的变体。

[0027] 在另一方面,本公开提供治疗病症的方法,所述病症可以通过施用N-糖基化改变的蛋白质来治疗。所述方法包括向受试者施用如下蛋白质的步骤,所述蛋白质可以通过上述任何方法获得,其中所述受试者是患有或疑似患有可通过施用N-糖基化改变的蛋白质来治疗的疾病的患者。所述方法也可以包括如下步骤:(a)提供受试者和/或(b)测定该受试者是否患有可以通过施用N-糖基化改变的蛋白质来治疗的疾病。所述受试者可以是哺乳动物如人。所述病症可以是例如,癌症,免疫病症(例如,炎性状态)或代谢紊乱。代谢紊乱可以是任何本文描述的那些代谢紊乱,例如,溶酶体贮积病(LSD)如高歇尔氏病(Gaucher disease)、泰-沙二氏病(Tay-Sachs disease)、旁姆普氏病(Pompe disease)、尼-皮二氏病(Niemann-Pick disease)或法布里氏病(Fabrydisease)。所述蛋白质可以是与LSD相关的蛋白质,例如,所述蛋白质可以是,例如,葡糖脑苷酯酶、 $\alpha$ -半乳糖苷酶。所述蛋白质可以是,例如, $\alpha$ -L-艾杜糖苷酸酶、 $\beta$ -D-半乳糖苷酶、 $\beta$ -葡糖苷酶、 $\beta$ -己糖胺酶(hexosaminidase)、 $\beta$ -D-甘露糖苷酶、 $\alpha$ -L-岩藻糖苷酶、芳基硫酸酯酶B、芳基硫酸酯酶A、 $\alpha$ -N-乙酰半乳糖胺酶、天冬氨酰葡萄糖胺酶、艾杜糖醛酸-2-硫酸酯酶、 $\alpha$ -氨基葡萄糖-N-乙酰转移酶、 $\beta$ -D-葡萄糖醛酸酶、透明质酸酶、 $\alpha$ -L-甘露糖苷酶、 $\alpha$ -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(sphinoglycanase)、硫酯酶、组织蛋白酶K或脂蛋白脂肪酶。

[0028] 在另一方面,本公开提供解脂西洋蓍霉或Arxula adeninivorans细胞(或其它相关种类的双态性的酵母细胞)的基本上纯的培养物,其中大量细胞是经遗传工程化而包含至少一种经修饰(例如任何本文所述的修饰)的N-糖基化活性的。所述细胞培养物可以包含

一种或多种细胞亚群,各亚群包含不同的经修饰的糖基化活性。

[0029] 在又一方面,本公开提供:(a)包含SEQ ID N0:1或SEQ ID N0:2的分离的核苷酸序列;(b)包含与SEQ ID N0:1或SEQ ID N0:2至少80%相同的序列的分离的核苷酸序列;或(c)由(a)或(b)的分离的核苷酸序列编码的多肽。在一些实施方案中,分离的核酸序列是SEQ ID N0:1或SEQ ID N0:2。

[0030] 在另一方面,本公开描述分离的核酸,其含有:(a)在高严紧条件下与SEQ ID N0:1或SEQ ID N0:2的补体杂交的核苷酸序列;或(b)核苷酸序列的补体。

[0031] 在又一方面,本公开提供:(a)包含本文所示核酸序列中任一(或由其组成)的分离的核苷酸序列;(b)包含与本为所示的核酸序列中的任一至少80%相同的序列的分离的核苷酸序列;或(c)由(a)或(b)的分离的核苷酸序列编码的多肽。在一些实施方案中,所述分离的核酸序列是本文所示的核酸序列中的任一。

[0032] 在另一方面,本公开描述分离的核酸,其含有:(a)在高严紧条件下与本文所示核酸序列中任一的补体杂交的核苷酸序列;或(b)所述核苷酸序列的补体。

[0033] 在又一方面,本公开提供:(a)包含上述核酸序列中任一的载体或(b)含有(a)的载体的经培养的细胞。所述载体可以是表达载体。载体中的核酸序列可以与表达调控序列可操作地连接。

[0034] 在另一方面,本公开提供产生蛋白质的方法。所述方法包括在允许多肽表达的条件下培养上述细胞中的任一种的步骤。所述方法也可包括如下步骤:在培养细胞之后,从细胞或培养细胞的培养基中分离多肽。所述细胞可以是例如,含有包含上述核酸序列中任一的载体的经培养细胞。

[0035] 本文所述的目标分子(例如,目标蛋白)、具有N-糖基化活性的蛋白质和N-糖基化改变的分子(统称为“本发明的分子”)可以是,但不需要是分离的。术语“分离的”如应用于任何本文所述的本发明的分子是指这样的分子或其片段,其已经从天然与其伴随的成分(例如,蛋白质或其它天然存在的生物分子或有机分子)分开或纯化。应理解的是重组分子(例如,重组蛋白)将总是“分离的”。通常,当本发明的分子按重量占制备物中全部相同类型的分子的至少60%时,例如占样品中全部相同类型的分子的60%时,则它是分离的。例如,当糖基化改变的蛋白质按重量占制备物或样品中全部蛋白质的至少60%时,它是分离的。在一些实施方案中,制备物中的本发明的分子按重量组成至少75%,至少90%,或至少99%的制备物中全部相同类型的分子。

[0036] 如用于本文,目标分子的“改变的N-糖基化形式”是由经遗传工程化的宿主细胞(例如,解脂西洋蓍霉细胞、Arxula adeninivorans细胞或另一相关的双态性酵母细胞种类的细胞)产生的目标分子的N-糖基化形式,其不同于未经遗传工程化的与经遗传工程化的细胞相同种类的细胞中产生的目标分子的N-糖基化形式。由此,目标分子的改变的糖基化形式可以是例如,非N-糖基化的目标分子。此外,目标分子的改变的糖基化形式可以是例如,一种或多种N-联聚糖的磷酸化改变的目标分子的形式。

[0037] 如用于本文,术语“其它相关的双态性酵母细胞种类”是指与解脂西洋蓍霉和Arxula adeninivorans相关的属于双足囊菌科(family Dipodascaceae)的酵母,例如Arxula、双足囊菌属(Dipodascus)(例如白色双足囊菌(D.albidus)、大双足囊菌(D.ingens)、或D.specifer)、半乳糖霉(Galactomyces)(例如里氏半乳糖霉(G.reesii)或

大地半乳糖霉 (*G. geotrichum*) )、*Sporopachyderma*、射盾子囊霉属 (*Stephanoascus*) (例如, 西非射盾子囊霉 (*S. ciferii*))、拟威克酵母属 (*Wickerhamiella*) 和接囊菌属 (*Zygoascus*)。具体而言, 进化枝 *Metchnikowia* (例如, *M. pulcherrima* 或 *M. agaves*) 和射盾子囊霉属 (通过分析菌种如 *Arxula* (例如 *A. adeninivorans* 或 *A. terrestris*) 的 26S-rDNA 序列的 D1/D2 结构域将解脂西洋蓍霉被分配于其中) 中的酵母以及一些念珠菌属菌种 (*Candida species*) (例如, *C. apicola*, 而非白念珠菌 (*C. albicans*)、麦芽糖念珠菌 (*C. maltose*) 或热带念珠菌 (*C. tropicalis*))。

[0038] “多肽”和“蛋白质”可以互换使用并且意指任何肽连接的氨基酸链,无论其长度和翻译后修饰。

[0039] 本公开还提供本文所述的野生型、全长、成熟“目标蛋白”或“具有N-糖基化活性的蛋白质”的 (i) 生物学活性变体和 (ii) 生物学活性片段或其生物学活性变体。全长、成熟、野生型蛋白质或所述蛋白质的片段的生物学活性变体可以含有添加、缺失或取代。具有取代的蛋白质将通常具有不多于 50 个 (例如, 不多于 1、2、3、4、5、6、7、8、9、10、12、15、20、25、30、35、40 或 50 个) 保守氨基酸取代。保守取代是用一种氨基酸取代另一种具有相似特征的氨基酸。保守取代包括在以下组内的取代: 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; 以及苯丙氨酸和酪氨酸。非极性疏水氨基酸包括丙氨酸、亮氨酸、异亮氨酸、缬氨酸、脯氨酸、苯丙氨酸、色氨酸和甲硫氨酸。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。带正电的 (碱性) 氨基酸包括精氨酸、赖氨酸和组氨酸。带负电的 (酸性) 氨基酸包括天冬氨酸和谷氨酸。将上述极性、碱性或酸性组中的一个成员用相同组中的另一成员取代可认为是保守取代。相反, 非保守取代是用一种氨基酸取代另一种具有不相似的特征的氨基酸的取代。

[0040] 缺失变体可以缺乏 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19 或 20 个 (两个或更多个氨基酸的) 氨基酸区段或不连续的单独氨基酸。

[0041] 添加 (添加变体) 包括融合蛋白, 所述融合蛋白含有: (a) 全长、野生型、成熟多肽或其含有至少五个氨基酸的片段; 和 (b) 内部或末端 (C 或 N) 的无关的或异源的氨基酸序列。在这些融合蛋白的背景下, 术语“异源氨基酸序列”是指除 (a) 之外的氨基酸序列。因此含有本文描述的肽和异源氨基酸序列的融合蛋白在序列上不对应于天然存在的蛋白质的全部或部分。例如, 异源序列可以是用于重组蛋白的纯化的序列 (例如, FLAG、多组氨酸 (例如, 六组氨酸)、血凝素 (*hemagluttinin*) (HA)、谷胱甘肽-S-转移酶 (GST) 或麦芽糖结合蛋白 (MBP))。异源序列也可以是用作诊断或检测标志物的蛋白质, 例如, 萤光素酶、绿色荧光蛋白 (GFP) 或氯霉素乙酰转移酶 (CAT)。在一些实施方案中, 融合蛋白含有来自另一蛋白质的信号序列。在某些宿主细胞 (例如, 酵母宿主细胞) 中, 目标蛋白的表达和/或分泌可以通过使用异源信号序列增加。在一些实施方案中, 融合蛋白可以含有可用于例如引发免疫应答 (例如, 用于抗体生成; 见下文) 的载体 (例如, KLH) 或内质网或高尔基体保留信号。异源序列可以具有不同的长度, 并且在一些情况下可以是比异源序列所附接的全长目标蛋白更长的序列。

[0042] “片段”如用于本文是指比全长、不成熟的蛋白质短的多肽的区段。蛋白质的片段可以具有末端 (羧基或氨基末端) 缺失和/或内部缺失。通常, 蛋白质的片段的长度将是至少 4 个 (例如, 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个或至少100个或更多个)氨基酸。

[0043] 目标蛋白或具有N-糖基化活性的蛋白质的生物学活性片段或生物学活性变体具有野生型、全长、成熟蛋白质的活性的至少25% (例如,至少:30%;40%;50%;60%;70%;75%;80%;85%;90%;95%;97%;98%;99%;99.5%,或100%或甚至更高)。在目标蛋白的情况下,相关活性是目标蛋白在经遗传工程化的细胞中经历改变的N-糖基化的能力。在具有N-糖基化活性的蛋白质的情况下,相关活性是N-糖基化活性。

[0044] 依赖于它们的预期用途,蛋白质、其生物学活性片段或生物学活性变体可以是任何物种的,如,例如,真菌(包括酵母)、线虫、昆虫、植物、鸟、爬行动物、或哺乳动物(例如,小鼠、大鼠、兔、仓鼠、沙鼠(gerbil)、犬、猫、山羊、猪、牛、马、鲸、猴或人)。在一些实施方案中,生物学活性片段或生物学活性变体包括所述蛋白质的免疫原性和抗原性片段。免疫原性片段是具有相关全长、未成熟蛋白质在感兴趣的动物中刺激免疫应答(例如,抗体应答或细胞免疫应答)的能力的至少25% (例如,至少30%;40%;50%;60%;70%;75%;80%;85%;90%;95%;97%;98%;99%;99.5%或100%或甚至更多)的片段。蛋白质的抗原性片段是具有相关全长、未成熟的蛋白质被对于所述蛋白质特异性的抗体或对于所述蛋白质特异性的T细胞识别的能力的至少25% (例如,至少30%;40%;50%;60%;70%;75%;80%;85%;90%;95%;97%;98%;99%;99.5%或100%或甚至更多)的片段。

[0045] “N-糖基化活性”如用于本文是指任何如下活性,即(i)能够向目标分子添加N-联聚糖(即,寡糖基转移酶活性);(ii)从目标分子去除N-联聚糖,(iii)修饰目标分子上的一个或多个N-联聚糖,(iv)修饰多萜醇连接的寡糖;或(v)能够辅助(i-iv)之内的活性的活性。同样,N-糖基化活性包括,例如,N-糖苷酶活性、糖苷酶活性、糖基转移酶活性、糖核苷酸合成、修饰或转运蛋白活性。对目标分子上一种或多种N-联聚糖的修饰包括甘露糖基磷酰基转移酶(mannosylphosphoryltransferase)活性、激酶活性或磷酸酶活性,例如,改变目标分子上N-联聚糖的磷酸化状态的甘露糖基磷酰基转移酶、激酶或磷酸酶活性。

[0046] 如用于本文,“遗传工程化”细胞或“经遗传工程化的细胞”和类似术语是指任何人工创造的细胞的遗传改变,其导致与未经遗传工程化的细胞(例如,真菌细胞例如解脂西洋蓍霉细胞、*Arxula adeninivorans*细胞或其它相关种类的双态性的酵母细胞,植物细胞或动物细胞(例如,哺乳动物细胞例如人的细胞))相比细胞中的至少一种经修饰的N-糖基化活性。由此,应该理解的是人工创造的遗传改变不包括,例如,自发突变。人工遗传改变的实例在下文描述(参见“遗传工程化的细胞”)。

[0047] 如用于本文,术语“野生型”如应用于核酸或多肽是指当生物学生物体存在于自然界中时分别在该生物学生物体中存在的或由所述生物学生物体产生的核酸或多肽。

[0048] 术语“异源”如在本文中应用于宿主细胞中的核酸或由宿主细胞产生的多肽是指并非源自与宿主细胞相同种类的细胞的核酸或多肽(例如,具有N-糖基化活性的蛋白质)。因此,如用于本文,“同源”核酸或蛋白质是在与宿主细胞相同种类的细胞中存在的或由与宿主细胞相同种类的细胞产生的那些。

[0049] 术语“外源”如用于本文与核酸和特定宿主细胞有关时是指不存在于(并且不能获得自)在自然界中存在的所述特定细胞的任何核酸。因此,非天然存在的核酸一旦引入宿主细胞则被认为对于所述宿主细胞是外源的。重要的是应注意非天然存在的核酸可以含有在

自然界中存在的核酸序列的核酸亚序列或片段,条件是所述核酸作为整体不存在于自然界中。例如,在表达载体内含有基因组DNA序列的核酸分子是非天然存在的核酸,并且因此一旦引入宿主细胞则对所述宿主细胞是外源的,因为该核酸分子作为整体(基因组DNA加上载体DNA)不存在于自然界中。因此,作为一个整体在自然界中不存在的任何载体、自复制质粒或病毒(例如,逆转录病毒、腺病毒或疱疹病毒)被认为是非天然存在的核酸。由此推断通过PCR或限制性内切核酸酶处理的基因组DNA片段以及cDNA被认为是天然存在的核酸,因为它们作为在自然界中不存在的单独的分子而存在。还由此推断以自然界中不存在的排列含有启动子序列和多肽编码序列的任何核酸是非天然存在的核酸。天然存在的核酸对于特定细胞可以是外源的。例如,从酵母x的细胞分离的完整染色体一旦引入酵母y的细胞则该染色体对于酵母y的细胞是外源核酸。

[0050] 从上文将显而易见的是,“外源”核酸可以是“同源”或“异源”核酸。相反,术语“内源”如用于本文与核酸或基因(或由所述核酸或基因编码的蛋白质)以及特定细胞有关时是指确实存在于(并且可以获得自)在自然界中存在的所述特定细胞中。

[0051] 作为对上述概念的示例,转化入解脂西洋蓍霉细胞中的编码解脂西洋蓍霉ALG6蛋白的表达质粒相对于该细胞而言是外源核酸。然而,编码ALG6蛋白的序列和由其产生的ALG6蛋白与所述细胞是同源的。类似地,转化入解脂西洋蓍霉细胞中的编码*Arxula adeninivorans* ALG6蛋白的表达质粒相对于该细胞而言是外源核酸。然而与前一例子相反,编码所述ALG6蛋白的序列和由其产生的ALG6蛋白与所述细胞是异源的。

[0052] 如用于本文,“启动子”是指使基因能够被转录的DNA序列。启动子由RNA聚合酶识别,然后起始转录。因此,启动子含有或者直接通过RNA聚合酶的募集结合的或在RNA聚合酶的募集中涉及的DNA序列。启动子序列也可以包括“增强子区”,其是能够与蛋白质结合以增强基因簇中基因的转录水平(因此得名)的一个或多个DNA区(即,反式作用因子,更像一组转录因子)。所述增强子尽管通常在编码区的5'端,但是也可以与启动子序列分开,并且可以例如在基因内含子区内或在基因编码区的3'。

[0053] 如用于本文,“可操作地连接”是指并入遗传构建体从而表达调控序列有效地调控感兴趣的编码序列的表达。

[0054] 本文描述的任何核酸序列的变体(例如,SEQ ID NO:1或SEQ ID NO:2中所示的HAC1序列)可以具有与野生型核酸序列同源的序列,例如,与野生型核酸序列为至少约70%(例如,至少约75%,至少约80%,至少约85%,至少约90%,至少约95%或至少约99%)同源(相同)的序列。这样的野生型序列可以分离自自然界或者可以通过重组或合成方法产生。由此野生型序列核酸可以具有天然存在的人的核酸序列、猴核酸序列、鼠类核酸序列或任何其它含有所述感兴趣的野生型核酸的同源物的物种的核酸序列。如用于本文,“同源”或“同源核酸序列”或类似术语是指如下序列,其特征在于在核苷酸水平的至少指定百分比的同源性,并且可与序列同一性互换使用。

[0055] 百分比同源性或同一性可以如下测定,例如通过Gap程序(Wisconsin Sequence Analysis Package,用于UNIX的第八版,Genetics Computer Group,University Research Park,Madison,WI),使用缺省设定,其使用Smith和Waterman((1981) Adv.Appl.Math.2:482-489)的算法。在一些实施方案中,探针和靶(见下文)之间的同源性在约50%至约60%之间。在一些实施方案中,探针和靶核酸之间的同源性是大约55%-65%,65%-75%,约

70%-80%，约75%-85%，约80%-90%，约85%-95%或约90%-100%。

[0056] 术语“探针”如用于本文是指长度可变的核酸序列。在一些实施方案中，探针包含至少10个并且多如6,000个核苷酸序列。在一些实施方案中，探针包含至少12个，至少14个，至少16个，至少18个，至少20个，至少25个，至少50个或至少75个或100个连续核苷酸。较长长度的探针通常获得自天然或重组的来源(与直接的化学合成相反)，对于靶序列是高特异性的，并且与较长的寡聚物相比与靶杂交慢得多。探针可以是单链或双链核酸分子。

[0057] 在一些实施方案中，本文描述的变体核酸可以具有如下序列，所述序列包含与感兴趣的野生型核酸(例如，如SEQ ID N0:1或SEQ ID N0:2中所示的HAC1核酸序列)中的区域、部分、结构域或区段具有部分互补性(例如，至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%互补性)的单链或双链。在一些实施方案中，感兴趣的变体核酸序列可以具有如下序列，所述序列包含与野生型核酸序列中的区域、部分、结构域或区段具有完全互补性(即，100%互补性)的单链或双链。序列“互补性”是指特定含氮碱基之间由于它们的氢键键合特性(即，具有使反向平行双链体能够形成的碱基序列的两条核酸链的性质，其中腺嘌呤和尿嘧啶(或胸腺嘧啶，在DNA或经修饰的RNA的情况下)彼此相反，而鸟嘌呤和胞嘧啶彼此相反)产生的化学亲和力。如此，完全互补序列将是在核苷酸序列形成反向平行双链体时具有完全的一对一碱基序列对应性(即，腺嘌呤对尿嘧啶/胸腺嘧啶且鸟嘌呤对胞嘧啶)的两个序列。

[0058] 杂交也可用作对两个核酸序列之间同源性的测量。本文描述的核酸序列，或其片段或变体，可以用作依据标准杂交技术的杂交探针。感兴趣的特定探针(例如，HAC1核苷酸序列的探针，例如，如SEQ ID NOS:1或2中所示的HAC1核苷酸序列)与来自试验来源(例如，真核细胞)的DNA或RNA的杂交表明对在试验来源中对应于所述探针的DNA或RNA的存在(例如，HAC1核苷酸序列)。杂交条件是本领域那些技术人员已知的，并且可以在Current Protocols in Molecular Biology, John Wiley&Sons, N.Y., 6.3.1-6.3.6, 1991中找到。将中等杂交条件定义为等同于在2X氯化钠/柠檬酸钠(SSC)中在30℃杂交，之后在1X SSC、0.1% SDS中在50℃清洗。将高严紧性条件定义为等同于在6X氯化钠/柠檬酸钠(SSC)中在45℃杂交，之后在0.2X SSC、0.1% SDS中在65℃清洗。

[0059] 除非另有限定，本文使用的全部技术和科学术语具有与本发明所属领域的一个普通技术人员所通常理解的相同的含义。尽管在本发明的实践或测试中可以使用与本文描述的那些相似的或等同的方法和材料，在下文将描述示例性的方法和材料。将本文提及的全部公开文献、专利申请、专利、Genbank®登录号和其它参考文献通过所述整体并入。发生冲突的情况下，以包括定义在内的本申请为准。材料、方法和实例仅作为示例说明，而非意欲进行限制。

[0060] 本发明的其它特征和优势，例如，产生N-糖基化改变的分子的方法，将从以下详述并从权利要求书中清楚可见。

#### [0061] 附图简述

[0062] 图1是描述在酵母内质网的N-聚糖前体合成的示意图。其所编码的蛋白质具有介导指明的酶促转化的活性的基因在带阴影的框中(例如，ALG7；左上)。“UDP”和“UMP”分别指二磷酸尿苷和一磷酸尿苷。“GDP”和“GMP”分别指二磷酸鸟苷和一磷酸鸟苷。“Gn”是指N-乙酰葡萄糖胺。“M”是指单体甘露糖，G是指葡萄糖，Pi是指磷酸。

- [0063] 图2是描述酵母内质网中N-聚糖加工的示意图。
- [0064] 图3是描述酿酒酵母(*S.cerevisiae*)高尔基体中N-聚糖加工的示意图。其编码的蛋白质具有介导指明的酶促转化的活性的基因在带阴影的框中(例如,0CH1;中上)。
- [0065] 图4是描述多种本文所述的N-聚糖结构的示意图。
- [0066] 图5是描述用于破坏解脂西洋蓍霉中0CH1基因的克隆策略的示意图。“PCR”是指聚合酶链式反应。
- [0067] 图6是描述用于MNN9基因破坏片段的克隆策略的示意图。“PCR”是指聚合酶链式反应。
- [0068] 图7是一系列描述获得自野生型解脂西洋蓍霉细胞或糖基化突变体(例如,Δoch1cI9、Δmnn91和Δoch1Δmnn9)细胞和MTLY60菌株细胞的甘露糖蛋白的N-聚糖分析的电泳图谱(electroferogram)。在一些情况下,将所述N-聚糖进一步用α-1,2甘露糖苷酶处理。使用DNA测序仪辅助型、荧光团辅助型糖电泳(DSA-FACE)来进行分析。“M5”,“M6”,“M7”,“M8”,和“M9”是指与基本N-乙酰葡糖胺结构结合(conjugate)的甘露糖残基数。Y轴代表相对荧光单位,表明各个甘露糖结构的量。X轴代表各个复合甘露糖结构通过凝胶的相对迁移率。顶部的电泳图谱是对用作迁移率标准品的寡麦芽糖(oligomaltose)的分析。
- [0069] 图8是描述用于酿酒酵母MNS1表达载体的克隆策略的示意图。“PCR”是指聚合酶链式反应。
- [0070] 图9是一系列描述对获得自MTLY60细胞的分泌型糖蛋白进行的N-聚糖分析的电泳图谱,所述细胞表达野生型(WT)Mns1p或所指明的Mns1p的多种突变形式(即,R273G、R273L或R269S/S272G/R273L)。分析使用DSA-FACE进行。“M5”,“M6”,“M7”,“M8”,“M9”是指与基本N-乙酰葡糖胺结构结合的甘露糖残基数。Y轴代表相对荧光单位,表明各个甘露糖结构的量。X轴代表各个复合甘露糖结构通过凝胶的相对迁移率。顶部的电泳图谱是对用作迁移率标准品的寡麦芽糖的分析。
- [0071] 图10是描述用于MNN4表达载体的克隆策略的示意图。
- [0072] 图11是一系列描述对获得自指明的野生型MTLY60细胞或糖基化突变细胞的分泌型糖蛋白进行的N-聚糖分析的电泳图谱。分析使用DSA-FACE进行。“M5”,“M6”,“M7”,“M8”,“M9”是指与壳二糖核心结构结合的甘露糖残基数。“P”是指含有一个磷酸残基的甘露糖蛋白,而“PP”是指含有两个磷酸残基的甘露糖蛋白。Y轴代表相对荧光单位,表明各个甘露糖结构的量。X轴代表各个复合甘露糖结构通过凝胶的相对迁移率。顶部的电泳图谱是对用作迁移率标准品的寡麦芽糖的分析。
- [0073] 图12是描述用于α-半乳糖苷酶表达载体的克隆策略的示意图。
- [0074] 图13是一系列描述对获得自指明的野生型MTLY60细胞或糖基化突变细胞的多种克隆的甘露糖蛋白和磷酸甘露糖蛋白进行的N-聚糖分析的电泳图谱。“alg3”表明所述细胞是ALG3敲除型。“ALG6过表达”表明所述ALG6的蛋白质产物在细胞中过表达。分析使用DSA-FACE进行。“M5”,“M6”,“M7”,“M8”和“M9”是指与基本N-乙酰葡糖胺结构结合的甘露糖残基数。Y轴代表相对荧光单位,表明各个甘露糖结构的量。X轴代表各个复合甘露糖结构通过聚丙烯酰胺凝胶的相对迁移率。顶部的电泳图谱是对用作迁移率标准品的寡麦芽糖的分析。
- [0075] 图14是一系列描述对获得自指明的野生型MTLY60细胞或糖基化突变细胞的多种克隆的甘露糖蛋白和磷酸甘露糖蛋白进行的N-聚糖分析的电泳图谱。“alg3”表明所述细胞

是ALG3敲除型。“ALG6过表达”表明所述ALG6的蛋白质产物在细胞中过表达。一个峰与RNA酶B标志物的Man<sub>5</sub>GlcNAc<sub>2</sub>运行在相同位置，并且在α-1,2-甘露糖苷酶处理之后迁移了两个葡萄糖单位，并在α-甘露糖苷酶(JB)消化之后迁移4个葡萄糖单位。这与预期的Man<sub>5</sub>GlcNAc<sub>2</sub>结构相符。额外的两个峰在约一个和两个糖单位的距离运行，并且不受α-1,2-甘露糖苷酶消化的影响。在α-甘露糖苷酶(JB)消化时两个峰都迁移一个葡萄糖单位。小迁移是因为加入的酶(例如JB甘露糖苷酶)的较高的盐浓度。分析使用DSA-FACE进行。“M5”, “M6”, “M7”, “M8”和“M9”是指与壳二糖核心结构结合的甘露糖残基数。Y轴代表相对荧光单位，表明各个甘露糖结构的量。X轴代表各个复合甘露糖结构通过凝胶的相对迁移率。顶部的电泳图谱是对用作迁移率标准品的寡麦芽糖的分析。

[0076] 图15是对获得自未折叠蛋白质应答(UPR)诱导的菌株解脂西洋蓍霉的分离的DNA片段(SEQ ID NO:1)序列与基因组HAC1cDNA序列(SEQ ID NO:5)的序列比对。带框的序列对应于非常规剪接的内含子。

[0077] 图16是一系列对巴斯德毕赤酵母和酿酒酵母的预测的5’(顶部)和3’(底部)剪接位点的序列比对。粗体加下划线的核苷酸存在于环结构中。

[0078] 图17A和17B是对获得自DTT-诱导型(I)(SEQ ID NO:2)和非诱导型(NI)(SEQ ID NO:6)巴斯德毕赤酵母培养物的HAC1cDNA的序列比对的两个部分视图。

[0079] 图18是对巴斯德毕赤酵母和酿酒酵母18个氨基酸的C-末端区域的序列比对。保守的氨基酸为粗体并加下划线。

[0080] 图19是描述对KAR2mRNA的相对表达水平进行比较的柱状图。将克隆3、4和5(巴斯德毕赤酵母GSM5细胞)在作为碳源的甲醇上培养。“3+”、“4+”和“5+”是指在作为碳源的甲醇上培养的各个克隆，而“3-”、“4-”和“5-”是指在作为碳源的葡萄糖上培养的各个克隆。Y轴代表使用实时PCR的KAR2基因的相对表达。

[0081] 图20是描述两种巴斯德毕赤酵母克隆(克隆6和8)中Kar2和HAC1的相对表达水平的柱状图。“6+”和“8+”是指在作为碳源的甲醇上培养的各个克隆，而“6-”和“8-”是指在作为碳源的葡萄糖上培养的各个克隆。Y轴代表使用实时PCR的KAR2基因的相对表达。

[0082] 图21是描述用于Y1MNN6表达载体的克隆策略的示意图。

[0083] 图22是一系列描述对获得自指明的单独的Δ och1解脂西洋蓍霉细胞，或表达Y1MNN6的Δ och1解脂西洋蓍霉的多种克隆(Z3、Z4、Z5、U5、U6和U8)的糖蛋白进行的N-聚糖分析的电泳图谱。分析使用DSA-FACE进行。Y轴代表相对荧光单位，表明各个甘露糖结构的量。X轴代表各个复合甘露糖结构通过凝胶的相对迁移率。顶部的电泳图谱是对用作迁移率标准品的寡麦芽糖的分析。

[0084] 图23是描述用于MFManHDEL表达载体的克隆策略的示意图。

[0085] 图24是一系列描述对获得自指明的单独的Δ och1解脂西洋蓍霉细胞，或表达MFManHDEL的Δ och1解脂西洋蓍霉的多种克隆(9、11、10、3、5和6)的糖蛋白进行的N-聚糖分析的电泳图谱。分析使用DSA-FACE进行。Y轴代表相对荧光单位，表明各个甘露糖结构的量。X轴代表各个复合甘露糖结构通过凝胶的相对迁移率。顶部的电泳图谱是对用作迁移率标准品的寡麦芽糖的分析。

[0086] 图25是描述用于LIP2preManHDEL表达载体的克隆策略的示意图。

[0087] 图26是一系列描述对获得自指明的单独的Δ och1解脂西洋蓍霉细胞，或表达

LIP2ManHDEL的 $\Delta$  och1解脂西洋蓍霉的多种克隆(1、5、10和11)的糖蛋白进行的N-聚糖分析的电泳图谱。分析使用DSA-FACE进行。“M5”,“M6”,“M7”,“M8”和“M9”是指与壳二糖核心结构结合的甘露糖残基数。Y轴代表相对荧光单位,表明各个甘露糖结构的量。X轴代表各个复合甘露糖结构通过凝胶的相对迁移率。顶部的电泳图谱是对用作迁移率标准品的寡麦芽糖的分析。

[0088] 图27A和27B是解脂西洋蓍霉(图27A;SEQ ID NO:3)和巴斯德毕赤酵母(图27B;SEQ ID NO:4)的HAC1蛋白的氨基酸序列。

[0089] 图28是考马斯蓝染色的聚丙烯酰胺凝胶的照片,其描述在多种解脂西洋蓍霉细胞(MTLY60、MTLY60 $\Delta$  alg3和MTLY60 $\Delta$  alg3ALG6)培养物中Lip2p过表达的结果。在凝胶中解析了以下样品:泳道1(“梯子”或“梯”),已知分子量的蛋白质的组合;泳道2(“WT”),获得自过表达Lip2p的WT解脂西洋蓍霉细胞(MTLY60)的Lip2p蛋白;泳道3(“WT+PNGase F”),获得自过表达Lip2p的WT解脂西洋蓍霉细胞并用PNGase F酶处理的Lip2p蛋白;泳道4(“alg3-ALG6”),获得自缺乏alg3并且过表达Lip2p和ALG6二者的西洋蓍霉细胞(MTLY60 $\Delta$  alg3ALG6)的Lip2p蛋白;泳道5(“alg3-ALG6+PNGase F”),获得自缺乏alg3并且过表达Lip2p和ALG6二者的西洋蓍霉细胞(MTLY60 $\Delta$  alg3ALG6)并且用PNGase F酶处理的Lip2p蛋白;泳道6(“alg3”),获得自缺乏alg3并且过表达Lip2p的解脂西洋蓍霉细胞(MTLY60 $\Delta$  alg3)的Lip2p蛋白;泳道7(“alg3+PNGase F”),获得自缺乏alg3并且过表达Lip2p的解脂西洋蓍霉细胞(MTLY60 $\Delta$  alg3)并用PNGase F酶处理的Lip2p蛋白;泳道8(“无Lip2p过表达的WT”),获得自MTLY60细胞的蛋白质;和泳道9(“无Lip2p过表达的WT+PNGase F”),获得自MTLY60细胞并用PNGase F酶处理的蛋白质。

[0090] 图29是一系列描述对获得自指明的多种解脂西洋蓍霉细胞(WT(MTLY60); $\Delta$  alg3; $\Delta$  alg3ALG6过表达的;和过表达ALG6连同来自解脂西洋蓍霉(Y1)或布氏锥虫(Tb)的葡糖昔酶II的 $\alpha$ 亚基的 $\Delta$  alg3克隆)的糖蛋白进行的N-聚糖分析的电泳图谱。分析使用DSA-FACE进行。“M5”,“M6”,“M7”,“M8”和“M9”是指与壳二糖核心结构结合的甘露糖残基数。Y轴代表相对荧光单位,表明各个甘露糖结构的量。X轴代表各个复合甘露糖结构通过凝胶的相对迁移率。顶部的电泳图谱是对用作迁移率标准品的寡麦芽糖的分析。底部的电泳图谱是对RNA酶B的分析。

[0091] 图30是一系列描述对获得自指明的多种解脂西洋蓍霉细胞( $\Delta$  alg3; $\Delta$  alg3ALG6过表达的;和过表达ALG6连同含有HDEL序列的来自解脂西洋蓍霉(Y1)的葡糖昔酶II的 $\alpha$ 亚基的 $\Delta$  alg3克隆)的糖蛋白进行的N-聚糖分析的电泳图谱。分析使用DSA-FACE进行。Y轴代表相对荧光单位,表明各个甘露糖结构的量。X轴代表各个复合甘露糖结构通过凝胶的相对迁移率。

[0092] 图31是一系列描述对获得自指明的多种解脂西洋蓍霉细胞( $\Delta$  alg3; $\Delta$  alg3ALG6过表达的;和过表达ALG6连同含有HDEL序列的来自布氏锥虫(Trypanosoma brucei)(Tb)的葡糖昔酶II的 $\alpha$ 亚基的 $\Delta$  alg3克隆)的糖蛋白进行的N-聚糖分析的电泳图谱。分析使用DSA-FACE进行。Y轴代表相对荧光单位,表明各个甘露糖结构的量。X轴代表各个复合甘露糖结构通过凝胶的相对迁移率。

[0093] 图32是一系列描述对获得自指明的用不同浓度变聚糖酶体外处理的alg3ALG6解脂西洋蓍霉细胞的糖蛋白进行的N-聚糖分析的电泳图谱。分析使用DSA-FACE进行。Y轴代表

相对荧光单位,表明各个甘露糖结构的量。X轴代表各个复合甘露糖结构通过凝胶的相对迁移率。顶部的电泳图谱是对用作迁移率标准品的寡麦芽糖的分析。底部的电泳图谱是对RNA酶B的分析。

[0094] 图33是一系列描述对获得自指明的多种解脂西洋蓍霉细胞( $\Delta$  alg3;  $\Delta$  alg3ALG6过表达的;和过表达ALG6连同在Hp4d或TEF启动子调控下表达的来自解脂西洋蓍霉(Y.1.)的葡糖昔酶II的 $\alpha$ 亚基和来自Y.1的葡糖昔酶II的 $\beta$ 亚基的 $\Delta$  alg3克隆)的糖蛋白进行的N-聚糖分析的电泳图谱。Y轴代表相对荧光单位,表明各个甘露糖结构的量。X轴代表各个复合甘露糖结构通过凝胶的相对迁移率。顶部的电泳图谱是对用作迁移率标准品的寡麦芽糖的分析。底部的电泳图谱是对RNA酶B的分析。

[0095] 图34是一系列描述对获得自指明的多种解脂西洋蓍霉细胞( $\Delta$  alg3ALG6过表达的;和过表达ALG6连同在Hp4d或TEF启动子调控下表达的含HDEL的来自解脂西洋蓍霉(Y.1.)的葡糖昔酶II的 $\alpha$ 亚基和来自Y.1的葡糖昔酶II的 $\beta$ 亚基的 $\Delta$  alg3克隆)的糖蛋白进行的N-聚糖分析的电泳图谱。分析使用DSA-FACE进行。Y轴代表相对荧光单位,表明各个甘露糖结构的量。X轴代表各个复合甘露糖结构通过凝胶的相对迁移率。顶部的电泳图谱是对用作迁移率标准品的寡麦芽糖的分析。底部的电泳图谱是对RNA酶B的分析。

[0096] 图35是一系列描述对获得自指明的多种解脂西洋蓍霉细胞( $\Delta$  alg3和过表达在TEF启动子调控下表达的来自解脂西洋蓍霉(Y.1.)的葡糖昔酶II的 $\alpha$ 亚基和来自Y.1的葡糖昔酶II的 $\beta$ 亚基的 $\Delta$  alg3克隆)的糖蛋白进行的N-聚糖分析的电泳图谱。Y轴代表相对荧光单位,表明各个甘露糖结构的量。X轴代表各个复合甘露糖结构通过凝胶的相对迁移率。顶部的电泳图谱是对用作迁移率标准品的寡麦芽糖的分析。底部的电泳图谱是对RNA酶B的分析。

[0097] 图36A和36B描述的是编码黑曲霉(Aspergillus niger)葡糖昔酶II $\alpha$ 成熟形式(缺少信号肽)的cDNA的核苷酸序列,其是为了在解脂西洋蓍霉中表达而经密码子优化的cDNA(SEQ ID NO:7)。

[0098] 图37是描述编码黑曲霉(Aspergillus niger)葡糖昔酶II $\beta$ 成熟形式(缺少信号肽)的cDNA的核苷酸序列,其是为了在解脂西洋蓍霉中表达而经密码子优化的cDNA(SEQ ID NO:8)。

[0099] 图38是一系列对获得自指明的多种解脂西洋蓍霉细胞( $\Delta$  alg3和ALG6过表达的,连同在TEF或hp4d启动子调控下表达的来自黑曲霉(An)的葡糖昔酶II的 $\alpha$ 亚基)的糖蛋白进行的N-聚糖分析的电泳图谱。Y轴代表相对荧光单位,表明各个甘露糖结构的量。X轴代表各个复合甘露糖结构通过凝胶的相对迁移率。顶部的电泳图谱是对用作迁移率标准品的寡麦芽糖的分析。底部的电泳图谱是对RNA酶B的分析。

[0100] 图39A和39B是一对柱状图,其描述在WT(MTLY60)解脂西洋蓍霉细胞中或在含有在hp4d启动子表达调控下的HAC1cDNA的经剪接形式的解脂西洋蓍霉细胞的两个克隆(克隆7和克隆2)中,HAC1(39A)或KAR(39B)基因的相对表达水平(Y轴)。

[0101] 图40是线型图,其描述用空载体转化的野生型巴斯德毕赤酵母GS115细胞与表达Hac1p蛋白的巴斯德毕赤酵母GS115细胞的生长相比的生长。

[0102] 图41是考马斯蓝染色的聚丙烯酰胺凝胶的照片,其比较了来自表达鼠类IL-10(mIL-10)蛋白的巴斯德毕赤酵母GS115细胞细胞培养物的mIL-10蛋白的表达水平与获得自

在诱导型启动子AOX1调控下表达mIL-10和来自巴斯德毕赤酵母的经剪接HAC1蛋白的GS115细胞的培养物的mIL-10蛋白的表达。在凝胶中解析了以下样品：泳道1（“梯子”），已知分子量的蛋白质的组合；泳道2（“参照”），获得自表达参照mIL-10的巴斯德毕赤酵母菌株(GS115)的蛋白质；泳道3（“参照”），获得自表达参照mIL-10的巴斯德毕赤酵母菌株的、在用PNGase F酶处理蛋白质之后的蛋白质；泳道4（“克隆1”），获得自诱导型表达HAC1蛋白的表达mIL-10的巴斯德毕赤酵母细胞的蛋白质；泳道5（“克隆1”），获得自诱导型表达HAC1蛋白的表达mIL-10的巴斯德毕赤酵母细胞的、在用PNGase F酶处理蛋白质之后的蛋白质；泳道6（“克隆2”），获得自诱导型表达HAC1蛋白1的表达mIL-10的巴斯德毕赤酵母细胞的蛋白质；泳道7（“克隆2”），获得自诱导型表达HAC1蛋白的表达mIL-10的巴斯德毕赤酵母细胞的、在用PNGase F酶处理蛋白质之后的蛋白质。

[0103] 图42描述的是编码里氏木霉(*Trichoderma reesei*) $\alpha$ -1,2甘露糖苷酶、为了在解脂西洋蓍霉中表达而经密码子优化、含有LIP2前信号序列的示例性cDNA序列的核苷酸序列(SEQ ID NO:9)。

[0104] 图43描述的是用于解脂西洋蓍霉的GAP启动子的示例性核苷酸序列的核苷酸序列(SEQ ID NO:10)。

[0105] 图44A-44C描述的是用于表达载体pYLHUXdL2preManHDEL的示例性核苷酸序列(SEQ ID NO:11)，其含有编码里氏木霉 $\alpha$ -1,2甘露糖苷酶的、为了在解脂西洋蓍霉中表达而经密码子优化并且含有LIP2前信号序列的cDNA序列。

[0106] 图45A-45C描述的是用于表达载体pYLGUXdL2preManHDEL的示例性核苷酸序列(SEQ ID NO:12)，其含有编码里氏木霉 $\alpha$ -1,2甘露糖苷酶的、为了在解脂西洋蓍霉中表达而经密码子优化并且含有LIP2前信号序列的cDNA序列。

[0107] 图46A-46C描述的是用于表达载体pYLPUXdL2preManHDEL的示例性核苷酸序列(SEQ ID NO:13)，其含有编码里氏木霉 $\alpha$ -1,2甘露糖苷酶的、为了在解脂西洋蓍霉中表达而经密码子优化并且含有LIP2前信号序列的cDNA序列。

[0108] 图47A-47C描述的是用于表达载体pYLTUXdL2preManHDEL的示例性核苷酸序列(SEQ ID NO:14)，其含有编码里氏木霉 $\alpha$ -1,2甘露糖苷酶的、为了在解脂西洋蓍霉中表达而经密码子优化并且含有LIP2前信号序列的cDNA序列。

[0109] 图48是一系列对获得自用指明的不同表达载体转化的解脂西洋蓍霉细胞的糖蛋白进行的N-聚糖分析的电泳图谱：“hp4dL2ManHDEL”(pYLHUXdL2preManHDEL,图44A-44C)；“GAPL2ManHDEL”(pYLGUXdL2preManHDEL,图45A-45C)；“TEF1L2ManHDEL”(pYLTUXdL2preManHDEL,图47A-47C)。Y轴代表相对荧光单位，表明各个甘露糖结构的量。X轴代表各个复合甘露糖结构通过凝胶的相对迁移率。顶部的电泳图谱是对用作迁移率标准品的葡聚糖(dextran)的分析。系列中第二个电泳图谱是对RNA酶B的分析。

[0110] 图49是一系列对获得自含有稳定整合的表达载体pYLTUXdL2preManHDEL(图47A-47C)的解脂西洋蓍霉MTLY60 $\Delta$ och1细胞的糖蛋白进行的N-聚糖分析的电泳图谱。糖蛋白样品获得自24、48、72和96小时的细胞培养物。顶部的电泳图谱是对用作迁移率标准品的葡聚糖(dextran)的分析。系列中第二个电泳图谱是对RNA酶B的分析。

[0111] 图50是用于人葡糖脑苷酯酶的示例性核酸序列(GLCM, Swiss Prot条目号:P04062; SEQ ID NO:15)，其是按照为了在解脂西洋蓍霉中表达而经密码子优化的cDNA以化

学方法合成的。

[0112] 图51是免疫印迹的照片,其描述了在解脂西洋蓍霉菌株MTLY60 (WT;泳道4和6) 和MTLY60 Δ och1 (Δ och1;前三个泳道) 中表达的人葡糖脑苷酯酶的迁移图。蛋白质的分子量(kDa)通过在所述免疫印迹最右侧的分子量标志物来说明。

[0113] 图52是人促红细胞生成素的示例性核酸序列(Epo, Swiss Prot条目号:P01588; SEQ ID NO:16),其是按照为了在解脂西洋蓍霉中表达而经密码子优化的cDNA以化学方法合成的。

[0114] 图53是人α-半乳糖苷酶A的示例性核酸序列(AGAL, Swiss Prot条目号:P06280; SEQ ID NO:17),其是按照为了在解脂西洋蓍霉中表达而经密码子优化的cDNA以化学方法合成的。

[0115] 图54是一系列过表达Hac1p蛋白的经剪接形式的巴斯德毕赤酵母细胞或野生型巴斯德毕赤酵母细胞的电子显微照片。将细胞中堆叠的脂质膜的离散区加框。

[0116] 图55是一系列电泳图谱,其描述对获得自指明的WT解脂西洋蓍霉细胞(po11d)和表达α-1,2-甘露糖苷酶和HDEL序列的融合蛋白的解脂西洋蓍霉细胞的糖蛋白进行的N-聚糖分析。分析使用DSA-FACE进行。“M5”,“M6”,“M7”,“M8”和“M9”是指与壳二糖核心结构结合的甘露糖残基数。Y轴代表相对荧光单位,表明各个甘露糖结构的量。X轴代表各个复合甘露糖结构通过凝胶的相对迁移率。顶部的电泳图谱是对用作迁移率标准品的寡麦芽糖的分析。底部的电泳图谱是对RNA酶B的分析。

[0117] 详述

[0118] 本文描述的方法和遗传工程化的细胞可以用于产生如下目标分子(例如,目标蛋白或目标多糖醇),其与在未经遗传工程化的细胞中产生的所述目标分子的N-糖基化形式相比具有改变的N-糖基化形式。已经证明糖基化目标分子(例如,糖基化蛋白)向患有代谢紊乱(例如,溶酶体贮积病)患者的施用使所述病症的症状改善。因此,所述的方法和细胞可用于制备N-糖基化改变的目标分子,其用于包括但不限于治疗代谢紊乱如溶酶体贮积病。这样的N-糖基化改变的分子也可用于广泛多样的其它领域,例如,食品和饮料工业;药物工业(例如,作为疫苗);农业工业;和化学工业,等等。

[0119] N-糖基化改变的分子

[0120] 如用于本文,目标分子是指通过来自遗传工程化的细胞(例如,真菌细胞如解脂西洋蓍霉或Arxula adeninivorans(或其它相关种类的双态性酵母)细胞;植物细胞;或动物细胞)的一种或多种N-糖基化活性而经受改变的N-糖基化的任何分子。在一些实施方案中,目标分子能够被运输经过解脂西洋蓍霉或Arxula adeninivorans(或其它相关种类的双态性酵母)分泌途径中的一步或多步,导致它们的N-糖基化因宿主细胞机构(machinery)而改变。所述目标分子可以是内源或外源的。

[0121] 目标蛋白、它们的生物学活性片段或其生物学活性变体可以包括如上所述含有添加、缺失或取代的蛋白质。合适的目标蛋白包括病原体蛋白(例如,破伤风类毒素;白喉(diphtheria)类毒素;病毒表面蛋白(例如,巨细胞病毒(CMV)糖蛋白B、H和gCIII;人免疫缺陷病毒1(HIV-1)包膜糖蛋白;劳斯肉瘤病毒(RSV)包膜糖蛋白;单纯疱疹病毒(HSV)包膜糖蛋白;EB病毒(EBV)包膜糖蛋白;水痘带状疱疹病毒(VZV)包膜糖蛋白;人乳头状瘤病毒(HPV)包膜糖蛋白;流感病毒糖蛋白;和肝炎家族表面抗原),溶酶体蛋白(例如,葡糖脑苷酯

酶、脑苷酶或半乳糖脑苷酯酶),胰岛素,胰高血糖素,生长因子,细胞因子,趋化因子,抗体及其片段,或所述蛋白质中任一与抗体或抗体片段的融合物(例如,蛋白质-Fc)。生长因子包括,例如,血管内皮生长因子(VEGF)、胰岛素样生长因子(IGF)、骨形态发生蛋白(BMP)、粒细胞集落刺激因子(G-CSF)、粒细胞-巨噬细胞集落刺激因子(GM-CSF)、神经生长因子(NGF);神经营养蛋白、血小板衍生生长因子(PDGF)、促红细胞生成素(EPO)、血小板生成素(TPO)、Myostatin(GDF-8)、生长分化因子-9(GDF9)、碱性成纤维细胞生长因子(bFGF或FGF2)、表皮生长因子(EGF)、肝细胞生长因子(HGF)。细胞因子包括,例如,白介素(例如,IL-1至IL-33(例如,IL-1、IL-2、IL-3、IL-4、IL-5、IL-6、IL-7、IL-8、IL-9、IL-10、IL-12、IL-13或IL-15))。趋化因子包括,例如,I-309、TCA-3、MCP-1、MIP-1 $\alpha$ 、MIP-1 $\beta$ 、RANTES、C10、MRP-2、MARC、MCP-3、MCP-2、MRP-2、CCF18、MIP-1 $\gamma$ 、Eotaxin、MCP-5、MCP-4、NCC-1、Ck $\beta$ 10、HCC-1、白细胞诱素-1(Leukotactin-1)、LEC、NCC-4、TARC、PARC或Eotaxin-2。还包括肿瘤糖蛋白(例如,肿瘤相关抗原),例如,癌胚抗原(CEA)、人粘蛋白、HER-2/neu和前列腺特异性抗原(PSA)[R.A.Henderson和O.J.Finn,Advances in Immunology,62,第217-56页(1996)]。在一些实施方案中,目标蛋白可以是与溶酶体贮积病相关的目标蛋白,所述目标蛋白包括,例如, $\alpha$ -L-艾杜糖苷酸酶、 $\beta$ -D-半乳糖苷酶、 $\beta$ -葡糖苷酶、 $\beta$ -己糖胺酶、 $\beta$ -D-甘露糖苷酶、 $\alpha$ -L-岩藻糖苷酶、芳基硫酸酯酶B、芳基硫酸酯酶A、 $\alpha$ -N-乙酰半乳糖胺酶、天冬氨酰葡萄糖胺酶、艾杜糖醛酸-2-硫酸酯酶、 $\alpha$ -氨基葡萄糖-N-乙酰转移酶、 $\beta$ -D-葡萄糖醛酸酶、透明质酸酶、 $\alpha$ -L-甘露糖苷酶、 $\alpha$ -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、硫酯酶、组织蛋白酶K和脂蛋白脂肪酶。

[0122] 目标蛋白也可以是融合蛋白。融合蛋白包括,例如,(i)本文描述的任何蛋白或其片段与(ii)抗体或其片段的融合物。如用于本文,术语“抗体片段”是指抗原结合片段,例如,Fab、F(ab')<sub>2</sub>、Fv和单链Fv(scFv)片段。scFv片段是单一多肽链,其包括所述scFv所源自的抗体的重链和轻链可变区二者。此外,双抗体[Poljak(1994)Structure 2(12):1121-1123;Hudson等(1999)J.Immunol.Methods 23(1-2):177-189,将这两篇的内容通过所述整体并入本文]和细胞内抗体[Huston等(2001)Hum.Antibodies 10(3-4):127-142;Wheeler等(2003)Mol.Ther.8(3):355-366;Stocks(2004)Drug Discov.Today 9(22):960-966,将全部这些的内容通过所述整体并入本文]也可以在本发明的方法中使用。

[0123] 目标蛋白也可以与聚合物、载体、佐剂、免疫毒素或可检测的(例如,荧光、发光或放射性)模块中的一种或多种结合。例如,可以将目标蛋白与聚乙二醇结合,所述聚合物模块可以用于例如增加小蛋白质的分子量和/或增加循环滞留期。

[0124] 在一些实施方案中,目标分子可以是或可以含有多萜醇。

[0125] 遗传工程化的细胞

[0126] 本文描述的遗传工程化的细胞具有至少一种经修饰的N-糖基化活性,所述细胞可用于一种或多种具有改变的N-糖基化形式的目标分子的产生。适合于遗传工程化的细胞包括,例如,真菌细胞(例如,解脂西洋蓍霉或本文所述的任何其它相关的双态性酵母细胞),植物细胞或动物细胞(例如,(线虫、昆虫、植物、鸟、爬行动物、或哺乳动物(例如,小鼠、大鼠、兔、仓鼠、沙鼠、犬、猫、山羊、猪、牛、马、鲸、猴或人))。所述细胞可以是原代细胞、无限增殖化细胞或经转化的细胞。所述细胞可以是动物中,例如非人哺乳动物中的那些细胞。这些细胞,在进行本文指定的遗传工程化之前,可以获得自多种商业来源和研究资源机构,如,

例如,美国典型培养物保藏中心(American Type Culture Collection) (Rockville, MD)。目标分子包括蛋白质,例如任何本文所述的目标蛋白(见上文)。目标分子也包括多萜醇。

[0127] 对细胞进行的遗传工程化包括遗传修饰如: (i) 缺失编码具有N-糖基化活性的蛋白质的内源基因; (ii) 引入编码具有N-糖基化活性的蛋白质(例如,内源或外源蛋白)的突变形式的重组核酸(即,表达具有N-糖基化活性的突变蛋白); (iii) 引入或表达RNA分子,所述RNA分子干扰具有N-糖基化活性的蛋白质的功能性表达; (iv) 引入编码具有N-糖基化活性的蛋白质的野生型(例如,内源的或外源的)的重组核酸(即,表达具有N-糖基化活性的蛋白质); 或(v) 改变编码具有N-糖基化活性的蛋白质的一种或多种内源基因的启动子或增强子元件,由此改变它们编码的蛋白质的表达。RNA分子包括,例如,小干扰RNA、短发夹RNA(shRNA)、反义RNA或微小RNA(miRNA)。应该理解的是,第(ii)项包括,例如,用相对于如下被替代的内源基因编码具有更大N-糖基化活性的蛋白质的基因替代内源基因(例如,通过同源重组)。遗传工程化还包括改变编码具有N-糖基化活性的蛋白质的内源基因以产生具有添加(例如,异源序列)、缺失或取代(例如,突变如点突变;保守或非保守突变)的蛋白质。突变可以特异性地引入(例如,定点诱变或同源重组;参加随附实施例)或者可以随机引入(例如,可以对细胞进行化学诱变,如在例如Newman和Ferro-Novick (1987) J. Cell Biol. 105 (4):1587中所述,将其公开内容通过所述整体并入本文)。

[0128] 本文描述的遗传修饰可以导致如下的一种或多种: (i) 在经遗传修饰的细胞中一种或多种N-糖基化活性的增加, (ii) 在经遗传修饰的细胞中一种或多种N-糖基化活性的减少, (iii) 在经遗传修饰的细胞中一种或多种N-糖基化活性的定位或胞内分布的变化, 或(iv) 在经遗传修饰的细胞中一种或多种N-糖基化活性的比例的变化。应该理解的是,N-糖基化活性的量的增加可以归因于一种或多种具有N-糖基化活性的蛋白质的过表达;内源基因的拷贝数的增加(例如,基因重复);或内源基因的启动子或增强子的改变,其刺激由所述基因编码的蛋白质的表达增加。一种或多种N-糖基化活性的减少可以归因于一种或多种蛋白质的突变形式(例如,显性阴性形式)的过表达,所述突变形式具有改变N-糖基化的活性;一种或多种干扰RNA分子的引入或表达,所述干扰RNA分子降低一种或多种具有N-糖基化活性的蛋白质的表达;或一种或多种编码具有N-糖基化活性的蛋白质的内源基因的缺失。

[0129] 缺失或破坏一种或多种内源基因的方法在随附实施例中描述。例如,为了通过同源重组破坏基因,“基因替代”载体可以以一种使其包括可选择的标记基因的方式构建。这些可选择的标记基因可以在5'和3'端与长度足以介导同源重组的基因部分可操作地连接。所述可选择的标记可以是多种基因之一,所述基因或者补足宿主细胞的营养缺陷型(auxotrophy),或者提供抗生素抗性,包括URA3、LEU2和HIS3基因。其它合适的可选择标记包括CAT基因,其赋予酵母细胞氯霉素抗性,或lacZ基因,其导致因表达β-半乳糖苷酶产生的蓝色菌落。然后使用本领域公知的方法(见下文)将基因替代载体的线性化DNA片段引入细胞。线性片段向基因组中的整合和基因的破坏可以基于所述选择标记来测定,并且可以通过例如Southern印迹分析来验证。

[0130] 如在随附实施例中详述的,在可选择的标记在选择中使用之后,可以将其从宿主细胞的基因组去除,通过例如Cre-loxP系统(见下文)。

[0131] 或者,基因替代载体可以以一种使其包括待破坏基因的部分的方式构建,所述部分缺乏任何内源基因启动子序列并且什么也不编码或编码所述基因编码序列的失活片段。

“失活片段”是这样的基因片段，其编码的蛋白质具有，例如，产自所述基因全长编码序列的蛋白质的活性的低于约10% (例如，低于约9%，低于约8%，低于约7%，低于约6%，低于约5%，低于约4%，低于约3%，低于约2%，低于约1%，或0%)。将这种基因部分以如下方式插入载体，即没有已知启动子序列与该基因序列可操作连接，但是终止密码子和转录终止序列与所述基因序列的所述部分可操作地连接。然后可以将这种载体在所述基因序列的所述部分中线性化并且转化入细胞中。通过单同源重组的方式，于是将这种线性化的载体整合入所述基因的内源对应物(counterpart)中。

[0132] 表达载体可以是自主性的或整合型的。

[0133] 可以将重组核酸(例如，编码具有N-糖基化活性的蛋白质的野生型或突变形式的重组核酸)以表达载体的形式引入细胞，所述表达载体例如质粒、噬菌体、转座子、粘粒或病毒颗粒。重组核酸可以保持在染色体外，或者它可以被整合入酵母细胞染色体DNA。表达载体可以含有选择标记基因，其编码细胞在选择条件下生存所需的蛋白质(例如，URA3，其编码尿嘧啶生物合成所需要的酶；或TRP1，其编码色氨酸生物合成所需要的酶)，以允许检测和/或选择那些用期望的核酸转化的细胞(参见，例如，美国专利No.4,704,362)。表达载体也可以包括自主复制序列(ARS)。例如，美国专利No.4,837,148描述了自主复制序列，其为在巴斯德毕赤酵母中保持质粒提供了适合的方法。将美国专利No.4,837,148的公开内容通过所述整体并入本文。

[0134] 整合型载体在例如美国专利No.4,882,279(将其公开内容通过所述整体并入本文)中披露。整合型载体通常包括连续排列的至少有第一可插入DNA片段、可选择的标记基因和第二可插入DNA片段的序列。所述第一和第二可插入DNA片段的长度各为约200个(例如，约250，约300，约350，约400，约450，约500或约1000或更长)核苷酸并且具有与待转化物种基因组DNA中的部分同源的核苷酸序列。将用于表达的含有感兴趣的基因(例如，编码具有N-糖基化活性的蛋白质的基因)的核苷酸序列插入这种载体的第一和第二可插入DNA片段之间，在标记基因之前或是之后。可以在酵母转化之前将整合型载体线性化以促进感兴趣的核苷酸序列整合入宿主细胞基因组。

[0135] 表达载体可以包含(feature)在酵母(例如，解脂西洋蓍霉、*Arxula adeninivorans*，或其它相关的双态性酵母种)启动子调控下的重组核酸，所述启动子使得它们能够在酵母中表达。合适的酵母启动子包括例如，ADC1、TPI1、ADH2、hp4d、POX和Gal10(参见，例如，Guarente等(1982) Proc.Natl.Acad.Sci.USA 79 (23) :7410)启动子。另外的合适启动子在例如Zhu和Zhang(1999) Bioinformatics 15 (7-8) :608-611和美国专利No.6,265,185中记载，将它们每一篇的公开内容通过所述整体并入本文。在将表达载体引入动物细胞(例如哺乳动物细胞)的情况下，所述表达载体可以包含在适合用于在感兴趣的宿主细胞中表达的动物细胞启动子调控下的重组核酸。哺乳动物启动子的实例包括，例如，SV40或巨细胞病毒(CMV)启动子。

[0136] 启动子可以是组成型或诱导型(条件性)的。组成型启动子应理解为这样的启动子，其表达在标准培养条件下是恒定的。诱导型启动子是响应于一种或多种诱导信号(induction cue)的启动子。例如，可以对诱导型启动子进行化学调节(例如，其转录活性受到化学诱导剂的存在或缺失调节的启动子，所述化学诱导剂例如醇、四环素、类固醇、金属或其它小分子)或物理调节(例如，其转录活性受到物理诱导物的存在或缺失调节的启动

子,所述物理诱导物例如光或高温或低温)。诱导型启动子也可以直接通过一种或多种转录因子调节,所述转录因子本身直接受到化学或物理信号的调节。

[0137] 细胞的遗传工程化还包括激活存在于宿主细胞中,但是通常在所述细胞中不表达或在所述细胞中不以显著水平表达的内源基因(例如,编码具有N-糖基化活性的蛋白质的基因)。例如,可以修饰内源基因的调节序列(例如,基因启动子或增强子)从而使可操作连接的编码序列展现增加的表达。同源重组和靶向可以用于使用如下调节序列替代通常与该基因结合的调节序列或使其失效,所述调节序列引起待表达基因以高于在相应的未经遗传工程化的细胞中所显现的水平表达,或引起所述基因展现不同于在相应的未经遗传工程化的细胞中所显现的调节或诱导模式。用于引入基因的调节序列(例如,启动子或增强子)的改变的合适方法记载着,例如,美国专利申请公开No.20030147868中,将其公开内容通过所述整体并入本文。

[0138] 应理解的是,其它遗传工程化的修饰也可以是条件性的。例如,可以将基因条件性地缺失,使用例如位点特异性DNA重组酶如Cre-loxP系统(参见,例如,Gossen等(2002)Ann.Rev.Genetics 36:153-173和美国申请No.20060014264,将它们每一篇的公开内容通过所述完整并入本文)。

[0139] 可以使用多种方法将重组核酸引入本文所述的细胞,所述方法例如原生质球技术或全细胞氯化锂酵母转化方法。其它对于将质粒或线性核酸载体转化入细胞有用的方法记载在,例如,美国专利No.4,929,555;Hinnen等(1978)Proc.Nat.Acad.Sci.USA 75:1929;Ito等(1983)J.Bacteriol.153:163;美国专利No.4,879,231;和Sreekrishna等(1987)Gene 59:115中,将它们每一篇的公开内容通过所述完整并入本文。电穿孔和PEG1000全细胞转化方法也可以使用,如Cregg和Russel,Methods in Molecular Biology:Pichia Protocols,Chapter 3,Human Press,Totowa,N.J.,第27-39页(1998)中所述,将其公开内容通过所述完整并入本文。动物细胞的转染可以包括,例如,使用磷酸钙、电穿孔、热激、脂质体或转染试剂如**FUGENE®**或**LIPOFECTAMINE®**或通过使裸核酸载体与细胞在溶液中接触(参见,例如,Sambrook等,Molecular Cloning:A Laboratory Manual Second Edition vol.1,2and 3.Cold Spring Harbor Laboratory Press:Cold Spring Harbor,New York,USA,Nov.1989;将其公开内容通过所述完整并入本文)将载体引入细胞。

[0140] 可以通过使用合适的技术来选择经转化的酵母细胞,所述技术包括但不限于,在转化之后在缺乏所需生化产物(由于细胞的营养缺陷型)条件下培养营养缺陷型细胞,选择和检测新的表型,或在对于缺乏转化体中所含抗性基因的酵母为毒性的抗生素存在下进行培养。转化体也可以通过将表达盒整合入基因组来选择和/或验证,其可以通过例如Southern印迹或PCR分析来评定。

[0141] 在将载体引入感兴趣的靶细胞之前,可以将载体在细菌细胞如大肠杆菌(E.coli)中培养(例如,扩增)。载体DNA可以通过任何本领域已知的方法从细菌细胞分离,所述方法致使载体DNA从细菌环境中纯化。经纯化的载体DNA可以用酚、氯仿和醚粗提(extract extensively),以确保没有大肠杆菌蛋白存在于质粒DNA制备物中,因为这些蛋白质对于哺乳动物细胞是毒性的。

[0142] 如本文所述,遗传工程化可以用于表达(例如,过表达)多种基因、向多种基因中引入修饰或缺失多种基因,例如,编码具有N-糖基化活性的蛋白质的基因。这些基因包括,例

如,ALG7、ALG13、ALG14、ALG1、ALG2、ALG11、RFT1、ALG3、ALG9、ALG12、ALG6、ALG8、ANL1、ALG10、ALG5、OST3、OST4、OST6、STT3、OST1、OST5、WBP1、SWP1、OST2、DPM1、SEC59、OCH1、MNN9、VAN1、MNN8、MNN10、MNN11、HOC1、MNN2、MNN5、MNN6、KTR1、YUR1、MNN4、KRE2、KTR2、KTR3、MNN1、MNS1、MNN4、PN01、MNN9、葡糖昔酶I、葡糖昔酶II、或内切甘露糖昔酶(endomannosidase)。编码具有N-糖基化活性的蛋白质的基因可以来自任何含有这样的基因的物种(例如,低等真核生物(例如,真菌(包括酵母)或锥虫),植物,或动物(例如,昆虫、鸟、爬行动物或哺乳动物(例如,啮齿动物如小鼠或大鼠、犬、猫、马、山羊、牛、猪、非人灵长类动物或人))。编码具有N-糖基化活性的蛋白质的基因可以从中获得的示例性真菌菌种包括,但不限于,异常毕赤酵母(*Pichia anomala*)、牛肠毕赤酵母(*Pichia bovis*)、加拿大毕赤酵母(*Pichia canadensis*)、*Pichia carsonii*、粉状毕赤酵母(*Pichia farinose*)、发酵毕赤酵母(*Pichia fermentans*)、泌液毕赤酵母(*Pichia fluxuum*)、膜醭毕赤酵母(*Pichia membranaefaciens*)、膜醭毕赤酵母(*Pichia membranaefaciens*)、粗壮念珠菌(*Candida valida*)、白念珠菌(*Candida albicans*)、*Candida ascalaphidarum*、*Candida amphixiae*、南极念珠菌(*Candida Antarctica*)、大西洋念珠菌(*Candida atlantica*)、*Candida atmosphaerica*、*Candida blattae*、*Candida carpophila*、*Candida cerambycidarum*、*Candida chauliodes*、*Candida corydalis*、*Candida dosseyi*、都柏林念珠菌(*Candida dubliniensis*)、*Candida ergatensis*、*Candida fructus*、光滑念珠菌(*Candida glabrata*)、发酵念珠菌(*Candida fermentati*)、季也蒙念珠菌(*Candida guilliermondii*)、希木隆念珠菌(*Candida haemulonii*)、*Candida insectamens*、昆虫念珠菌(*Candida insectorum*)、间型念珠菌(*Candida intermedia*)、*Candida jeffresii*、乳酒念珠菌(*Candida kefyr*)、克鲁斯念珠菌(*Candida krusei*)、葡萄牙念珠菌(*Candida lusitaniae*)、*Candida lyxosiphila*、*Candida maltosa*、醭膜念珠菌(*Candida membranifaciens*)、*Candida milleri*、*Candida oleophila*、奥里戈念珠菌(*Candida oregonensis*)、近平滑念珠菌(*Candida parapsilosis*)、桔念珠菌(*Candida quercitrusa*)、休哈塔念珠菌(*Candida shehatea*)、*Candida temnochilae*、纤细念珠菌(*Candida tenuis*)、热带念珠菌(*Candida tropicalis*)、*Candida tsuchiyaе*、*Candida sinolaborantium*、酱油念珠菌(*Candida sojae*)、维斯念珠菌(*Candida viswanathii*)、产朊念珠菌(*Candida utilis*)、膜醭毕赤酵母(*Pichia silvestris*)、膜醭毕赤酵母、*Pichia chodati*、膜醭毕赤酵母、膜醭毕赤酵母、*Pichia minuscule*、巴斯德毕赤酵母、假多形毕赤酵母(*Pichia pseudopolymorpha*)、栎毕赤酵母(*Pichia quercuum*)、*Pichia robertsii*、斋藤毕赤酵母(*Pichia saitoi*)、*Pichia silvestrisi*、斯地毕赤酵母(*Pichia strasburgensis*)、陆生毕赤酵母(*Pichia terricola*)、*Pichia vanriji*、*Pseudozyma Antarctica*、红冬孢酵母(*Rhodosporidium toruloides*)、红酵母(*Rhodotorula glutinis*)、贝酵母(*Saccharomyces bayanus*)、贝酵母、*Saccharomyces momdshuricus*、葡萄汁酵母(*Saccharomyces uvarum*)、贝酵母、酿酒酵母、二孢酵母(*Saccharomyces bisporus*)、薛瓦酵母(*Saccharomyces chevalieri*)、德尔布酵母(*Saccharomyces delbrueckii*)、少孢酵母(*Saccharomyces exiguous*)、发酵性酵母(*Saccharomyces fermentati*)、脆壁酵母(*Saccharomyces fragilis*)、马克思酵母(*Saccharomyces marxianus*)、蜂蜜酵母(*Saccharomyces mellis*)、罗斯酵母(*Saccharomyces rosei*)、鲁酵

母 (*Saccharomyces rouxii*)、葡萄汁酵母、威尔酵母 (*Saccharomyces willianus*)、路德酵母路德酵母、*Saccharomycopsis capsularis*、扣囊复膜孢酵母 (*Saccharomycopsis fibuligera*)、扣囊复膜孢酵母、大麦曲内孢霉 (*Endomyces hordei*)、*Endomycopsis fibuligera*、*Saturnispora saitoi*、八孢裂殖酵母 (*Schizosaccharomyces octosporus*)、粟酒裂殖酵母 (*Schizosaccharomyces pombe*)、西方许旺酵母 (*Schwanniomyces occidentalis*)、戴尔有孢圆酵母 (*Torulaspora delbrueckii*)、戴尔有孢圆酵母、大连酵母 (*Saccharomyces dairensis*)、戴尔有孢圆酵母、发酵有孢圆酵母 (*Torulaspora fermentati*)、发酵性酵母、戴尔有孢圆酵母、罗斯有孢圆酵母 (*Torulaspora rosei*)、罗斯酵母 (*Saccharomyces rosei*)、戴尔有孢圆酵母、罗斯酵母、戴尔有孢圆酵母、德尔布酵母、戴尔有孢圆酵母、德尔布酵母、*Zygosaccharomyces mongolicus*、(*Dorulaspora globosa*)、球形德巴利酵母 (*Debaryomyces globosus*)、圆球拟酵母 (*Torulopsis globosa*)、丝孢酵母 (*Trichosporon cutaneum*)、三角酵母 (*Trigonopsis variabilis*)、加利福尼亚拟威尔酵母 (*Williopsis californica*)、拟威尔酵母 (*Williopsis saturnus*)、二孢接合酵母 (*Zygosaccharomyces bisporus*)、二孢接合酵母 (*Debaryomyces disporua*)、二孢酵母、二孢接合酵母、二孢酵母、蜂蜜接合酵母 (*Zygosaccharomyces mellis*)、*Zygosaccharomyces priorianus*、(*Zygosaccharomyces rouxiim*)、鲁接合酵母 (*Zygosaccharomyces rouxii*)、(*Zygosaccharomyces barkeri*)、鲁酵母、鲁接合酵母、大接合酵母 (*Zygosaccharomyces major*)、鲁酵母、异常毕赤酵母、牛肠毕赤酵母、加拿大毕赤酵母、*Pichia carsonii*、粉状毕赤酵母、发酵毕赤酵母、泌液毕赤酵母、膜醭毕赤酵母、假多形毕赤酵母、栎毕赤酵母、*Pichia robertsii*、*Pseudozyma Antarctica*、红冬孢酵母、红冬孢酵母、红酵母、贝酵母、贝酵母、二孢酵母、酿酒酵母、薛瓦酵母、德尔布酵母、发酵性酵母、脆壁酵母、路德酵母、粟酒裂殖酵母、西方许旺酵母、戴尔有孢圆酵母、*Torulaspora globosa*、三角酵母、加利福尼亚拟威尔酵母、拟威尔酵母、二孢接合酵母、蜂蜜接合酵母、鲁接合酵母或任何其它本领域已知的或本文描述的真菌(例如,酵母)。示例性低等真核生物还包括曲霉属 (*Aspergillus*) 的多个种,包括但不限于,浅蓝灰曲霉 (*Aspergillus caesiellus*)、亮白曲霉 (*Aspergillus candidus*)、肉色曲霉 (*Aspergillus carneus*)、棒曲霉 (*Aspergillus clavatus*)、弯头曲霉 (*Aspergillus deflectus*)、黄曲霉 (*Aspergillus flavus*)、烟曲霉 (*Aspergillus fumigatus*)、灰绿曲霉 (*Aspergillus glaucus*)、构巢曲霉 (*Aspergillus nidulans*)、黑曲霉、赭曲霉 (*Aspergillus ochraceus*)、米曲霉 (*Aspergillus oryzae*)、寄生曲霉 (*Aspergillus parasiticus*)、青霉状曲霉 (*Aspergillus penicilloides*)、局限曲霉 (*Aspergillus restrictus*)、酱油曲霉 (*Aspergillus sojae*)、聚多曲霉 (*Aspergillus sydowi*)、溜曲霉 (*Aspergillus tamari*)、土曲霉 (*Aspergillus terreus*)、焦曲霉 (*Aspergillus ustus*) 或杂色曲霉 (*Aspergillus versicolor*)。编码具有N-糖基化活性的蛋白质的基因可以从中获得的示例性原虫属包括,但不限于,芽短膜虫属 (*Blastocritidida*)、短膜虫属 (*Critidiida*)、肠锥虫属 (*Endotrypanum*)、匐滴虫属 (*Herpetomonas*)、利什曼原虫属 (*Leishmania*)、细滴虫属 (*Leptomonas*)、植生滴虫属 (*Phytomonas*)、锥虫属 (*Trypanosoma*) (例如,布氏锥虫 (*T. brucei*)、冈比亚锥虫 (*T. gambiense*)、罗德西亚锥虫 (*T. rhodesiense*) 和克氏锥虫 (*T. cruzi*)) 和 *Wallaceina*。

[0143] 应该理解的是,如本文所述,遗传工程化可以用于表达(例如过表达)、引入修饰

至、或缺失多种基因(例如,编码具有N-糖基化活性的蛋白质的基因)和/或本文所述任何基因中一种或多种(例如,2、3、4、5、6、7、8、9、10、11、12、15或20种或更多种)的任意组合。

[0144] 在一些实施方案中,所述经遗传工程化的细胞缺乏ALG3 (**Genbank®** 登录号:XM\_503488;Genolevures Ref:YALI0E03190g) 基因或其基因产物(例如,mRNA或蛋白质)。在一些实施方案中,所述经遗传工程化的细胞表达(例如过表达)ALG6 (**Genbank®** 登录号:XM\_502922,Genolevures Ref:YALI0D17028g) 蛋白。在一些实施方案中,所述经遗传工程化的细胞表达MNN4基因 (**Genbank®** 登录号:XM\_503217,Genolevures Ref:YALI0D24101g)。在一些实施方案中,所述经遗传工程化的细胞缺乏OCH1和/或MNN9基因或其基因产物(例如,mRNA或蛋白质)。在一些实施方案中,所述经遗传工程化的细胞不缺乏OCH1基因或其基因产物(例如,mRNA或蛋白质)。在一些实施方案中,所述经遗传工程化的细胞表达葡糖昔酶II的 $\alpha$ 或 $\beta$ 亚基(或 $\alpha$ 和 $\beta$ 亚基二者),所述葡糖昔酶II例如解脂西洋蓍霉或布氏锥虫的葡糖昔酶II。在一些实施方案中,所述经遗传工程化的细胞表达变聚糖酶(mutantase),例如哈茨木霉的变聚糖酶。在一些实施方案中,所述经遗传工程化的细胞可以具有这些修饰的任意组合。

[0145] 例如,在一些实施方案中,所述经遗传工程化的细胞可以缺乏ALG3(例如,由**Genbank®** 登录号:XM\_503488,Genolevures Ref:YALI0E03190g示例的ALG3基因)基因或其基因产物(例如,mRNA或蛋白质);可以过表达ALG6(例如,如由**Genbank®** 登录号:XM\_502922,Genolevures Ref:YALI0D17028g示例的ALG6)蛋白;可以过表达葡糖昔酶II(例如解脂西洋蓍霉、布氏锥虫或任何其它本文所述的物种的葡糖昔酶II)的 $\alpha$ 和/或 $\beta$ 亚基;可以过表达 $\alpha$ -1,2-甘露糖昔酶;并且过表达如下中的一种或多种(和任意组合):糖昔酶、糖基转移酶、糖-核苷酸转运蛋白、糖-核苷酸修饰酶。在一些实施方案中,所述经遗传工程化的细胞不缺乏OCH1基因或其基因产物(例如,mRNA或蛋白质)。

[0146] 在一些实施方案中,所述经遗传工程化的细胞可以含有甘露糖昔酶活性(例如, $\alpha$ -甘露糖昔酶活性)。所述甘露糖昔酶活性可以靶向内质网。所述甘露糖昔酶可以具有至少7.5以下的最适pH(例如,至少7.4以下,至少7.3以下,至少7.2以下,至少7.1以下,至少7.0以下,至少6.9以下,至少6.8以下,至少6.7以下,至少6.6以下,至少6.5以下,至少6.4以下,至少6.3以下,至少6.2以下,至少6.1以下,至少6.0以下,至少5.9以下,至少5.8以下,至少5.7以下,至少5.6以下,至少5.5以下,至少5.4以下,至少5.3以下,至少5.2以下,至少5.1以下,至少5.0以下,至少4.9以下,至少4.8以下或至少4.7以下)。

[0147] 所述甘露糖昔酶可以是MNS1。

[0148] 例如,所述经遗传工程化的细胞可以过表达甘露糖昔酶(例如, $\alpha$ -1,2-甘露糖昔酶或任何其它本文所述的甘露糖昔酶),但是不缺乏OCH1基因或其基因产物(例如,mRNA或蛋白质)。所述甘露糖昔酶可以是所述蛋白质的野生型形式,或者可以是突变形式,例如含有甘露糖昔酶和HDEL ER-保留氨基酸序列的融合蛋白(参见实施例)。(应该理解可以将任何具有N-糖基化活性的蛋白质工程化改造成包含HDEL序列的融合蛋白)。

[0149] 在一些实施方案中,所述经遗传工程化的细胞可以含有能够促进目标分子的改变的N-糖基化形式进行甘露糖基磷酸化的活性。例如,可以将编码促进N-聚糖进行磷酸化的活性的核酸(例如MNN4,MNN6,PN01)引入遗传工程化的细胞,所述细胞能够增加目标分子的

N-糖基化的磷酸化。

[0150] 在一些实施方案中,所述经遗传工程化的细胞可以含有能够去除甘露糖残基的活性(例如,甘露糖苷酶,如来自Jack Bean的甘露糖苷酶),所述甘露糖残基遮蔽N-糖基化改变的分子的磷酸化。

[0151] 在一些实施方案中,所述经遗传工程化的细胞能够从 $\text{Man}_5\text{GlcNAc}_2$ 去除葡萄糖残基。例如,所述经遗传修饰的细胞可以过表达具有 $\alpha$ -1,3-葡糖苷酶活性的蛋白质,例如,但不限于,变聚糖酶或葡糖苷酶II(例如解脂西洋蓍霉、布氏锥虫或本文所述的任何其它真菌菌种的葡糖苷酶II)的 $\alpha$ 和/或 $\beta$ 亚基。

[0152] 在具有N-糖基化活性的蛋白质源自与表达该蛋白质的细胞不同类型(例如,不同种)的细胞的实施方案中,可以为了在特定的感兴趣的细胞中表达而对编码所述蛋白质的核酸进行密码子优化。例如,可以对来自布氏锥虫的编码具有N-糖基化的蛋白质的核酸进行密码子优化用于在酵母细胞(例如解脂西洋蓍霉)中表达。这种密码子优化可以用于增加蛋白质在感兴趣的细胞中的表达。用于对编码蛋白质的核酸进行密码子优化的方法是本领域已知的,并且记载在例如Gao等(Biotechnol.Prog.(2004)20(2):443-448),Kotula等(Nat.Biotechn.(1991)9,1386-1389)和Bennetzen等(J.Biol.Chem.(1982)257(6):2036-3031)中。

[0153] 也可以对细胞进行遗传工程化以主要产生作为哺乳动物(例如,人)糖基化途径的中间体的N-聚糖。例如,可以将编码具有N-糖基化活性的人的蛋白的一种或多种核酸引入细胞。在一些实施方案中,可以将人的蛋白引入细胞,并且可以抑制(例如,缺失或突变)一种或多种具有N-糖基化活性的内源酵母蛋白。用于“人源化”真菌糖基化途径的技术记载在,例如,Choi等(2003)Proc.Natl.Acad.Sci.USA 100(9):5022-5027;Verveken等(2004)Appl.Environ.Microb.70(5):2639-2646;和Gerngross(2004)Nature Biotech.22(11):1410-1414中,将它们每一篇的公开内容通过所述整体并入本文。

[0154] 在遗传工程化涉及例如改变蛋白质的表达或外源蛋白(包括内源蛋白的突变形式)的表达的情况下,可以使用多种技术来测定经遗传工程化的细胞是否表达该蛋白质。例如,编码所述蛋白质的mRNA或蛋白质本身的存在可以如下检测,分别使用例如Northern印迹或RT-PCR分析或Western印迹分析。具有N-糖基化活性的蛋白质的胞内定位可以通过使用多种技术来分析,所述技术包括亚细胞分级和免疫荧光。

[0155] 另外的遗传修饰和将它们引入本文描述的任何细胞的方法可以根据例如,美国专利Nos.7,029,872;5,272,070;和6,803,225;和美国专利申请公开Nos.20050265988、20050064539、20050170452和20040018588的公开内容修改适用,将它们每一篇的公开内容通过所述整体并入本文。

[0156] 尽管在双态性酵母菌种中为了实现体内产生 $\text{Man}_5\text{GlcNAc}_2$ 和 $\text{Man}_3\text{GlcNAc}_2$ 而进行的工程化步骤可能与在其它酵母菌种中进行的工程化步骤不同,但是本领域技术人员将清楚在双态性酵母中体内产生经修饰的糖蛋白(具有 $\text{Man}_5\text{GlcNAc}_2$ 和 $\text{Man}_3\text{GlcNAc}_2$ 核心N-聚糖结构)的工程化技术可以根据常规实验法从包括但不限于美国专利No.7,326,681和美国公开Nos.20040018590、20060040353和20060286637(将它们每一篇的公开内容通过所述整体并入本文)中公开方法修改获得。由此可以将经修改的方法用于实现糖蛋白的产生,所述糖蛋白经修饰而具有人类型杂合和复合的N-聚糖。这些复合N-聚糖可以在上文指定的核心聚糖

上具有2-5个以GlcNAc残基起始的分支,其可以进一步延伸,例如,用半乳糖、果糖、唾液酸残基延伸。

[0157] 在一些实施方案中,具有N-糖基化活性的突变蛋白或野生型蛋白可以从经遗传工程化的细胞中使用标准技术分离。例如,在所述经遗传工程化的细胞中表达突变蛋白或野生型蛋白之后,可以从该细胞本身或从培养该细胞的培养基分离所述蛋白质。分离蛋白质的方法是本领域已知的并且包括,例如,液相层析(例如,HPLC)、亲和层析(例如,金属螯合或免疫亲和层析)、离子交换层析、疏水相互作用层析、沉淀或差示溶解(differential solubilization)。

[0158] 在一些实施方案中,可以将具有N-糖基化活性的分离的蛋白质冷冻、冻干或固定,并且储存在合适条件下,所述条件允许该蛋白质保持活性。

[0159] 本公开还提供本文描述的任何经遗传工程化的细胞的基本上纯的培养物。如用于本文,经遗传工程化的细胞的“基本上纯的培养物”是这样的细胞培养物,其中所述培养物中活细胞总数的低于约40% (即,低于约:35%;30%;25%;20%;15%;10%;5%;2%;1%;0.5%;0.25%;0.1%;0.01%;0.001%;0.0001%;或甚至更低)是除所述经遗传工程化的细胞之外的活细胞,例如,细菌、真菌(包括酵母)、支原体或原生动物细胞。术语“约”在本文上下文中指相关的百分比可以比指定百分比高或低所述指定百分比的15%。因此,例如,约20%可以是17%-23%。这样的经遗传工程化的细胞的培养物包括细胞和培养、储存或转运培养基。培养基可以是液体、半固体(例如,胶状培养基)或是冷冻的。培养物包括培养在液体培养基中或培养在半固体培养基中/上的细胞、或储存或转运培养基(包括冷冻储存或转运培养基)中储存或转运的细胞。培养物可以在培养容器或储存容器或基底(例如,培养皿、烧瓶或管或储存小瓶或管)中。

[0160] 本文描述的经遗传工程化的细胞可以如下储存,例如,作为冷冻细胞悬液,例如,在含有冷冻保护剂如甘油或蔗糖的缓冲剂中,作为冻干的细胞。或者,它们可以例如作为干燥的细胞制备物储存,通过例如流化床干燥或喷雾干燥,或任何其它合适的干燥方法。

[0161] 产生N-糖基化改变的分子的方法

[0162] 本文描述产生目标分子的改变的N-糖基化形式的方法。所述方法通常涉及使目标分子与来自经遗传工程化的细胞(例如,真菌细胞(例如,解脂西洋蓍霉、*Arxula adeninivorans*或本文所述的任何其它相关的双态性酵母细胞)、植物细胞或动物细胞(例如,线虫、昆虫、植物、鸟、爬行动物或哺乳动物(例如,小鼠、大鼠、兔、仓鼠、沙鼠、犬、猫、山羊、猪、牛、马、鲸、猴或人))的一种或多种N-糖基化活性接触。所述方法可以是基于细胞的或不基于细胞的。

[0163] 基于细胞的方法可以包括向经遗传工程化而具有至少一种经修饰N-糖基化活性的细胞(例如,真菌细胞(例如,解脂西洋蓍霉、*Arxula adeninivorans*或本文所述的任何其它相关的双态性酵母细胞)中引入编码在该细胞中经历N-糖基化的目标分子的核酸,其中所述细胞以改变的N-糖基化形式产生所述目标分子。所述目标分子可以是,例如,蛋白质如任何本文所述的目标蛋白。在目标蛋白是脂质的实施方案中,核酸可以是编码一种或多种促进脂质合成的酶的核酸。

[0164] 通过对细胞进行遗传工程化来产生的修饰的类型在本文描述(参见随附实施例和上文的“遗传工程化的细胞”)。

- [0165] 用于引入核酸的方法是本领域已知的并且记载在随附实施例和上文中。
- [0166] 在遗传工程化的细胞中引入或表达目标分子(例如,目标蛋白)可以导致运输目标分子通过细胞的内质网和/或高尔基体,由此产生目标分子的改变的N-糖基化形式。
- [0167] 在对目标分子进行加工之后(例如,在经遗传修饰的细胞中),目标分子(例如,目标蛋白)的改变的N-糖基化形式可以含有一种或多种N-聚糖结构。例如,目标分子的改变形式可以含有一种或多种特定的N-具体结构如 $\text{Man}_5\text{GlcNAc}_2$ (结构式I或VII;图4), $\text{Man}_8\text{GlcNAc}_2$ (结构式I;图4), $\text{Man}_9\text{GlcNAc}_2$ (结构式II;图4), $\text{Man}_3\text{GlcNAc}_2$ (结构式XIV;图4), $\text{Glc}_1\text{Man}_5\text{GlcNAc}_2$ (结构式VIII;图4),或 $\text{Glc}_2\text{Man}_5\text{GlcNAc}_2$ (结构式IX;图4)(“Man”是甘露糖;“Glc”是葡萄糖;而“GlcNAc”是N-乙酰葡萄糖胺)。
- [0168] 产生自经遗传工程化的细胞的N-糖基化改变的目标分子可以是均质的(即,全部N-糖基化改变的分子含有相同的特定N-聚糖结构)或者可以是基本上均质的。“基本上均质”是指改变的目标分子是由经遗传工程化的细胞产生的改变N-糖基化的目标分子的至少约25%(例如,至少约27%,至少约30%,至少约35%,至少约40%,至少约45%,至少约50%,至少约55%,至少约60%,至少约65%,至少约70%,至少约75%,至少约80%,至少约85%,至少约90%或至少约95%或至少约99%)。
- [0169] 在经遗传工程化的细胞包括一种或多种影响N-聚糖磷酸化的N-糖基化活性的情况下,目标分子的改变的N-糖基化形式可以具有至少约25%(例如,至少约27%,至少约30%,至少约35%,至少约40%,至少约45%,至少约50%,至少约55%,至少约60%,至少约65%,至少约70%,至少约75%或至少约80%)的它的甘露糖基残基是磷酸化的。
- [0170] 在遗传工程化的细胞的任何遗传修饰在诱导信号(例如,化学或物理信号)存在下为诱导型或条件性的情况下,可以任选地将所述遗传工程化的细胞在引入核酸之前、过程中或之后在诱导剂存在下培养。例如,在引入编码目标蛋白的核酸之后,可以将所述细胞暴露于化学诱导剂,所述化学诱导剂能够促进一种或多种具有N-糖基化活性的蛋白质的表达。在多种诱导信号诱导一种或多种具有N-糖基化活性的蛋白质条件性表达的情况下,可以使细胞与多种诱导信号接触。
- [0171] 在通过一种或多种N-糖基化活性加工之后,将改变的目标分子分离。所述改变的目标分子可以保持在酵母细胞内或在细胞裂解时释放,或者所述改变的目标分子可以经由编码序列(外源核酸所固有的或工程化至表达载体中的)提供的机制分泌至培养基中,所述编码序列指导所述分子从细胞分泌。细胞裂解物或培养基中改变的目标分子的存在可以通过多种用于检测分子存在的标准规程来验证。例如,在改变的目标分子是蛋白质的情况下,这样的规程可以包括,但不限于,使用对于所述改变的目标蛋白(或目标蛋白本身)特异性的抗体进行的免疫印迹或放射免疫沉淀,对于所述改变的目标蛋白(或目标蛋白本身)特异性的配体进行的结合,或测试改变的目标蛋白(或目标蛋白本身)的特异性酶活性。
- [0172] 在一些实施方案中,可以将分离的改变的目标分子冷冻、冻干或固定,并且储存在合适条件下,例如,所述条件允许改变的目标分子保持生物学活性。
- [0173] 可以对目标分子的改变的N-糖基化形式进一步进行体内加工(例如,在经遗传工程化的细胞中),或者可以在从经遗传工程化的细胞或细胞培养基分离之后进行体外加工。进一步加工可以包括对改变的目标分子的一种或多种N-聚糖残基的修饰或对改变的目标分子的N-聚糖残基之外的修饰。改变的目标分子的额外的加工可以包括添加(共价或非共

价结合)异源模块如聚合物或载体。进一步的加工也可以包括对改变的目标分子进行的酶或化学处理。酶处理可以包括使改变的目标分子与一种或多种糖苷酶(例如,甘露糖苷酶或甘露聚糖酶)、磷酸二酯酶、磷脂酶、糖基转移酶或蛋白酶接触足以诱导对所述改变的目标分子的修饰的时间。酶处理也可以包括使所述改变的目标分子与能够从 $\text{Man}_5\text{GlcNAc}_2$ 去除一种或多种葡萄糖残基的酶接触,所述酶例如,但不限于,甘露糖苷酶或葡糖苷酶II的 $\alpha$ 和/或 $\beta$ 亚基。化学处理可以例如包括使改变的目标分子与酸(例如氢氟酸)接触足以诱导对所述改变的目标分子的修饰的时间。特定条件下的氢氟酸处理特异性地去除与聚糖以磷酸二酯键连接的甘露糖残基,同时在聚糖上保留磷酸。改变的目标分子可以通过从一个或多个N-聚糖添加或去除磷酸基团来进一步加工。例如,可以使改变的目标分子与甘露糖基激酶或甘露糖基磷酸酶接触。

[0174] 在一些实施方案中,任何文描述的改变的目标分子可以在分离之后与异源模块附接,例如,使用酶或化学方法。“异源模块”是指与改变的目标分子结合(例如,共价或非共价)的任何成分,所述成分不同于原始存在于所述改变的目标分子之上的成分。异源模块包括,例如,聚合物、载体、佐剂、免疫毒素或可检测(例如,荧光、发光或放射性)的模块在一些实施方案中,可以向改变的目标分子添加额外的N-聚糖。

[0175] 应该理解的是,可以,但不需要在遗传工程化的细胞中加工目标分子。例如,本公开还包括无细胞的产生具有改变的N-糖基化形式的目标分子的方法,所述方法包括使目标分子在N-糖基化条件下与制备自细胞(例如,真菌细胞(例如,解脂西洋蓍霉、*Arxula adeninivorans*或本文描述的任何其它相关的双态性酵母细胞)、植物细胞或动物细胞(例如,线虫、昆虫、植物、鸟、爬行动物、或哺乳动物(例如,小鼠、大鼠、兔、仓鼠、沙鼠、犬、猫、山羊、猪、牛、马、鲸、猴或人))的细胞裂解物接触的步骤,所述细胞经遗传工程化而具有至少一种经修饰的N-糖基化活性,其中使目标分子与所述细胞裂解物接触产生目标分子的改变的N-糖基化形式。

[0176] “N-糖基化条件”是指将混合物(例如,目标分子和细胞裂解物的混合物)在允许改变的N-糖基化的条件下温育(如上所述)。

[0177] 用于获得细胞裂解物(在所述裂解物中保留一种或多种N-糖基化活性的活性或完整性)的合适方法可以包括使用合适的缓冲液和/或抑制剂,包括核酸酶、蛋白酶和磷酸酶抑制剂,其保护或最小化细胞裂解物中N-糖基化活性的改变。这些抑制剂包括,例如,螯合剂如乙二胺四乙酸(EDTA)、乙二醇双(P-氨基乙醚)N,N,N1,N1-四乙酸(EGTA),蛋白酶抑制剂如苯甲基碘酰氟化物(PMSF)、抑肽酶、亮抑肽酶、抗痛素等,和磷酸酶抑制剂如磷酸盐、氟化钠、钒酸盐等。抑制剂可以这样选择,即它们不干扰或仅最小程度负面影响N-糖基化活性或感兴趣的活性。用于获得含有酶活性的裂解物的合适的缓冲剂和条件记载在,例如,Ausubel等Current Protocols in Molecular Biology(Supplement 47),John Wiley& Sons,New York(1999);Harlow和Lane,Antibodies:A Laboratory Manual Cold Spring Harbor Laboratory Press(1988);Harlow and Lane,Using Antibodies:A Laboratory Manual,Cold Spring Harbor Press(1999);Tietz Textbook of Clinical Chemistry,3rd ed.Burtis and Ashwood,eds.W.B.Saunders,Philadelphia,(1999)中。

[0178] 可以将细胞裂解物进一步加工以使当地消除或最小化干扰物质的存在。如有需要,可以通过多种本领域技术人员熟知的方法将细胞裂解物分级,所述方法包括亚细胞分

级和层析技术,如离子交换、疏水和反向、大小排阻、亲和、疏水电荷诱导层析等(参见,例如,Scopes, Protein Purification:Principles and Practice, third edition, Springer-Verlag, New York (1993); Burton和Harding, J.Chromatogr.A 814:71-81 (1998))。

[0179] 在一些实施方案中,可以制备细胞裂解物,其中全部细胞器保持完整和/或有功能。例如,裂解物可以含有完整的糙面内质网、完整的光面内质网或完整的高尔基体中的一种或多种。用于制备含有完整的细胞器的裂解物和测试细胞器的功能性的合适方法记载在,例如,Moreau等(1991) J.Biol.Chem. 266 (7) :4329-4333; Moreau等(1991) J.Biol.Chem. 266 (7) :4322-4328; Rexach等(1991) J.Cell Biol. 114 (2) :219-229; 和 Paulik等(1999) Arch.Biochem.Biophys. 367 (2) :265-273中;将它们每一篇的公开内容通过所述整体并入本文。

[0180] 本公开还提供产生具有改变的N-糖基化形式的目标分子的方法,其包括使目标分子在N-糖基化条件下与一种或多种具有N-糖基化活性的分离的蛋白质接触的步骤,其中使所述目标分子与一种或多种具有N-糖基化活性的蛋白质接触产生目标分子的改变的N-糖基化形式,并且其中所述一种或多种具有N-糖基化活性的蛋白质制备自经遗传工程化而具有至少一种经修饰的N-糖基化活性的细胞(例如真菌细胞(例如,解脂西洋蓍霉、Arxula adeninivorans或本文描述的任何其它相关的双态性酵母细胞)、植物细胞或动物细胞(例如,线虫、昆虫、植物、鸟、爬行动物、或哺乳动物(例如,小鼠、大鼠、兔、仓鼠、沙鼠、犬、猫、山羊、猪、牛、马、鲸、猴或人))。

[0181] 可以使用如上所述的标准技术将一种或多种具有N-糖基化活性的蛋白质纯化。可以使目标分子与一种或多种蛋白质在合适的缓冲液中接触足以诱导对目标分子的修饰的时间,如例如, Lee和Park (2002) 30 (6) : 716-720以及Fujita和Takegawa (2001) Biochem.Biophys.Res.Commun. 282 (3) :678-682中所述,将所述公开内容通过所述完整并入本文。

[0182] 在一些实施方案中,可以使目标分子仅与一种具有N-糖基化活性的蛋白质接触。在一些实施方案中,可以使目标分子与超过一种具有N-糖基化活性的蛋白质接触。可以使目标分子与超过一种蛋白质同时或相继接触。在将目标分子与超过一种具有N-糖基化活性的蛋白质相继接触的情况下,可以,但不需要,将目标分子在一个或多个步骤之后纯化。即,可以使目标分子与蛋白活性A接触,然后在将所述分子与蛋白活性B接触之前进行纯化,等等。

[0183] 在无细胞方法的一些实施方案中,在使目标分子与一种或多种N-糖基化活性接触之前将目标分子与固相支撑物连接可能是有利的。这样的连接能够使N-糖基化修饰之后的纯化更简单。合适的固相支撑物包括,但不限于,多孔测定平板、颗粒(例如,磁性或编码颗粒)、柱或膜。

[0184] 用于检测目标分子的N-糖基化(例如,改变的N-糖基化)的方法包括DNA测序仪辅助型(DSA)、荧光团辅助型糖电泳(FACE)(如随附实施例中所述)或表面增强型激光解吸/电离飞行时间质谱(SELDI-TOF MS)。例如,分析可以利用DSA-FACE,其中,例如,将糖蛋白变性,其后固定在例如膜上。然后可以将糖蛋白用合适的还原剂如二硫苏糖醇(DTT)或β-巯基乙醇还原。蛋白质的巯氢基可以使用酸如碘乙酸来羧化。然后,可以使用酶如N-糖苷酶F将

N-聚糖从蛋白质释放。任选地,N-聚糖可以通过还原性氨基化重建和衍生化。然后可以将衍生化的N-聚糖浓缩。适合用于N-聚糖分析的仪器包括,例如,ABI PRISM® 377DNA测序仪(Applied Biosystems)。数据分析可以使用例如GENESCAN® 3.1软件(Applied Biosystems)进行。任选地,可以将分离的甘露糖蛋白进一步用一种或多种酶处理以确认它们的N-聚糖状态。示例性的酶包括,例如, $\alpha$ -甘露糖苷酶或 $\alpha$ -1,2甘露糖苷酶,如随附实施例中所述。另外的N-聚糖分析方法包括,例如,质谱(例如,MALDI-TOF-MS),正相、反相高压液相层析(HPLC)和离子交换层析(例如,当未标记聚糖时使用脉冲电流计检测,而如果对聚糖进行了合适的标记则使用UV吸光度或荧光)。还参见Callewaert等(2001)Glycobiology 11 (4):275-281和Freire等(2006)Bioconjug.Chem.17 (2):559-564,将它们每一篇的公开内容通过所述整体并入本文。

[0185] 可由N-糖基化改变的分子处理的病症

[0186] 本文描述的分离的、N-糖基化改变的分子(例如,N-糖基化改变的蛋白质或多萜醇)可以用于治疗多种疾病,所述疾病可以通过施用一种或多种N-糖基化改变的分子来治疗(例如,N-糖基化改变的蛋白质)。能够通过施用N-糖基化改变的分子(例如,改变的N-糖蛋白或改变的N-糖基化的多萜醇)治疗或预防的一些特定医学状况的实例在下节中评述。

[0187] (i) 代谢紊乱

[0188] 代谢紊乱是影响个体人(或动物)细胞内能力产生的病症。大多数代谢紊乱是遗传性的,尽管一些可以因饮食、毒素、感染等而“获得”。遗传性代谢紊乱也称作代谢的先天性缺陷。概括而言,遗传性代谢紊乱是由遗传缺陷引起的,所述缺陷导致缺失或不正确地构建细胞代谢过程中的一些步骤所必须的酶。最大几类的代谢紊乱是糖代谢的紊乱、氨基酸代谢的紊乱、有机酸代谢的紊乱(有机酸尿症(organic acidurias)),脂肪酸氧化和线粒体代谢的紊乱,卟啉代谢的紊乱,嘌呤或嘧啶代谢的紊乱,类固醇代谢的紊乱,线粒体功能的紊乱,过氧化物酶体功能的紊乱和溶酶体贮积病(LSD)。

[0189] 能够通过使用一种或多种本文描述的N-糖基化改变的分子(或其药物组合物)来治疗的代谢紊乱的实例包括,例如,遗传性血色病(hereditary hemochromatosis)、眼皮肤白化病(oculocutaneous albinism)、蛋白C(protein C deficiency)、I型遗传性血管性水肿(type I hereditary angioedema)、先天性蔗糖酶-异麦芽糖酶缺乏(congenital sucrase-isomaltase deficiency)、克-纳二氏II型(Crigler-Najjar type II)、Laron综合征(Laron syndrome)、遗传性髓过氧化物酶(hereditary Myeloperoxidase)、原发性甲状腺机能减退(primary hypothyroidism)、先天性长QT综合征(congenital long QT syndrome)、酪氨酸结合球蛋白缺乏(tyroxine binding globulin deficiency)、家族性高胆固醇血症(familial hypercholesterolemia)、家族性乳糜微粒血症(familial chylomicronemia)、 $\alpha\beta$ -脂蛋白血症( $\alpha\beta$ -lipoproteinemia)、低原生质脂蛋白A水平(low plasma lipoprotein A levels)、伴有肝损伤的遗传性气肿(hereditary emphysema with liver injury)、先天性甲状腺机能减退(congenital hypothyroidism)、成骨不全(osteogenesis imperfecta)、遗传性血纤维蛋白原过少(hereditary hypofibrinogenemia)、 $\alpha$ -1抗胰凝乳蛋白酶缺乏( $\alpha$ -1antichymotrypsin deficiency)、肾原性尿崩症(nephrogenic diabetes insipidus)、垂体神经部尿崩症(neurohypophyseal diabetes insipidus)、腺苷脱氨酶缺乏(adenosine deaminase deficiency)、佩-迈二氏

病 (Pelizaeus Merzbacher disease)、IIA型冯维勒布兰德病 (von Willebrand disease type IIA)、结合性因子V和VIII缺乏 (combined factors V and VIII deficiency)、迟发性脊椎骨骺发育不全 (spondylo-epiphyseal dysplasia tarda)、无脉络膜 (choroideremia)、I细胞病 (I cell disease)、白顿氏病 (Batten disease)、共济失调性毛细血管扩张症 (ataxia telangiectasias)、常染色体显性多囊性肾病 (ADPKD-autosomal dominant polycystic kidney disease)、微绒毛包含病 (microvillus inclusion disease)、结核性硬化 (tuberous sclerosis)、Lowe眼脑肾综合征 (oculocerebro-renal syndrome of Lowe)、肌萎缩性侧索硬化 (amyotrophic lateral sclerosis)、骨髓增生异常综合征 (myelodysplastic syndrome)、裸淋巴细胞综合征 (Bare lymphocyte syndrome)、丹吉尔病 (Tangier disease)、家族性肝内胆汁郁积 (familial intrahepatic cholestasis)、X连锁的脑白质肾上腺萎缩症 (X-linked adreno-leukodystrophy)、Scott 综合征 (Scott syndrome)、1型和2型海-普二氏综合征 (Hermansky-Pudlak syndrome types 1 and 2)、左尔威格氏综合征 (Zellweger syndrome)、肢根点状软骨发育异常 (rhizomelic chondrodysplasia puncta)、常染色体隐性原发性高草酸尿 (autosomal recessive primary hyperoxaluria)、Mohr-Tranebjaerg 综合征 (Mohr-Tranebjaerg syndrome)、脊髓延髓肌萎缩 (spinal and bulbar muscular atrophy)、原发性睫运动障碍 (primary ciliary dyskinesia) (卡塔格内氏综合征 (Kartagener's syndrome))、巨人症和肢端肥大症 (giantism and acromegaly)、乳溢 (galactorrhea)、艾迪逊氏病 (Addison's disease)、肾上腺性男性化 (adrenal virilism)、库兴氏综合征 (Cushing's syndrome)、酮酸中毒 (ketoacidosis)、原发性或继发性醛固酮增多症 (primary or secondary aldosteronism)、米-迪综合征 (Miller-Dieker syndrome)、无脑回 (lissencephaly)、运动神经元病 (motor neuron disease)、阿史尔氏综合征 (Usher's syndrome)、威-阿二氏综合征 (Wiskott-Aldrich syndrome)、奥佩齐氏综合征 (Optiz syndrome)、亨廷顿氏病 (Huntington's disease)、遗传性胰腺炎 (hereditary pancreatitis)、抗磷脂综合征 (anti-phospholipid syndrome)、重叠结缔组织病 (overlap connective tissue disease)、斯耶格伦氏综合征 (Sjögren's syndrome)、僵体综合征 (stiff-man syndrome)、Brugada 综合征 (Brugada syndrome)、芬兰型先天性肾病综合征 (congenital nephritic syndrome of the Finnish type)、杜-约二氏综合征 (Dubin-Johnson syndrome)、X连锁低磷酸盐血症 (X-linked hypophosphatemia)、佩德里得综合征 (Pendred syndrome)、婴儿期持续性高血糖素低血糖 (persistent hyperinsulinemic hypoglycemia of infancy)、遗传性球形红细胞增多症 (hereditary spherocytosis)、无血浆酮蓝蛋白血症 (aceruloplasminemia)、婴儿神经元蜡样脂褐质沉积症 (infantile neuronal ceroid lipofuscinosis)、假性软骨发育不全和多发性骨骺发育不良 (不全) (pseudoachondroplasia and multiple epiphyseal)、Stargardt样黄斑发育不良 (Stargardt-like macular dystrophy)、X连锁夏-马-图三氏病 (X-linked Charcot-Marie-Tooth disease)、常染色体显性视网膜色素变性 (autosomal dominant retinitis pigmentosa)、Wolcott-Rallison 综合征 (Wolcott-Rallison syndrome)、库兴氏病 (Cushing's disease)、肢带肌营养不良 (limb-girdle muscular dystrophy)、IV型粘多糖贮积病 (mucopolysaccharidosis type IV)、芬兰遗传性家族性淀粉样变性 (hereditary

familial amyloidosis of Finish)、安德森氏病(Anderson disease)、肉瘤(sarcoma)、慢性髓单核细胞性白血病(chronic myelomonocytic leukemia)、心肌病(cardiomyopathy)、面生殖发育异常(faciogenital dysplasia)、Torsion病(Torsion disease)、亨廷顿和脊髓小脑萎缩(Huntington and spinocerebellar ataxias)、hereditary hyperhomosyteinemia、多神经病(polyneuropathy)、下运动神经元病(lower motor neuron disease)、色素性视网膜炎(pigmented retinitis)、血清阴性多关节炎(seronegative polyarthritis)、间质肺纤维化(interstitial pulmonary fibrosis)、雷诺氏现象(Raynaud's phenomenon)、韦格纳氏肉芽肿病(Wegner's granulomatosis)、蛋白尿(preteinuria)、CDG-Ia、CDG-Ib、CDG-Ic、CDG-Id、CDG-Ie、CDG-If、CDG-IIa、CDG-IIb、CDG-IIc、CDG-IId、埃-当二氏综合征(Ehlers-Danlos syndrome)、多发性外生骨疣(multiple exostoses)、格里斯塞利氏综合征(Griselli syndrome)(1型或2型)或X连锁性非特异性脑力发育迟缓(X-linked non-specific mental retardation)。此外,代谢紊乱还可包括溶酶体贮积病例如,但不限于,法布里氏病、法勃氏病(Farber disease)、高歇尔氏病、GM<sub>1</sub>神经节糖苷沉积症(GM<sub>1</sub>-gangliosidosis)、泰-沙二氏病、桑德霍夫氏病(Sandhoff disease)、GM<sub>2</sub>激活病(GM<sub>2</sub>activator disease)、克腊伯氏病(Krabbe disease)、异染性脑白质营养不良(metachromatic leukodystrophy)、尼-皮二氏病(Niemann-Pick disease)(A、B和C型)、赫勒氏病(Hurler disease)、沙伊氏病(Scheie disease)、Hunter病(Hunter disease)、桑菲列普氏病(Sanfilippo disease)、莫尔基奥氏病(Morquio disease)、马-垃病(Maroteaux-Lamy disease)、透明质酸酶缺乏(hyaluronidase deficiency)、天门冬酰氨酸葡萄糖胺尿症(aspartylglucosaminuria)、岩藻糖苷贮积病(fucosidosis)、甘露糖苷过多症(mannosidosis)、谢尔德氏病(Schindler disease)、1型唾液酸缺乏病(sialidosis type 1)、旁姆普氏病、致密性成骨不全(Pycnodynatoses)、蜡样脂褐质沉积症(ceroid lipofuscinosis)、胆固醇酯沉积病(cholesterol ester storage disease)、乌尔曼氏病(Wolman disease)、多发性硫酸酯酶缺乏(Multiple sulfatase deficiency)、乳液唾液异常增多(galactosialidosis)、粘脂贮积病(mucolipidosis)(II、III和IV型)、胱氨酸病(cystinosis)、唾液酸贮积病(sialic acid storage disorder)、伴有马-斯二氏综合征的乳糜微粒留滞病(chylomicron retention disease with Marinesco-Sjögren syndrome)、海-普二氏综合征(Hermansky-Pudlak syndrome)、切-希二氏综合征(Chediak-Higashi syndrome)、Danon病(Danon disease)或Geleophysic发育不良(Geleophysic dysplasia)。

[0190] 代谢紊乱的症状多种多样,并且可以包括如下的一种或多种,例如,贫血、疲劳、容易挫伤(bruising easily)、血小板低、肝增大、脾增大、骨骼脆弱(skeletal weakening)、肺损伤、感染(例如,胸部感染或肺炎)、肾损伤、进行性脑损伤、发作、胎粪过厚、咳嗽、哮鸣、唾液或粘液过量产生、气短(shortness of breath)、腹痛、肠或消化道闭塞(occluded bowel or gut)、生育力问题、鼻内息肉、杵状指/趾甲和皮肤(clubbing of the finger/toe nails and skin)、手或脚疼痛(pain in the hands or feet)、血管角质瘤(angiokeratoma)、排汗减少、角膜和晶状体混浊(corneal and lenticular opacities)、白内障(cataracts)、二尖瓣脱垂和/或回流(mitral valve prolapse and/or regurgitation)、心脏肥大(cardiomegaly)、温度不耐受、行走困难、吞咽困难、进行性视力

丧失、进行性听力丧失、张力减退、巨舌、反射消失、下腰痛、睡眠呼吸暂停、端坐呼吸、嗜睡、脊柱前凸或脊柱侧凸。应理解的是，由于缺陷或缺乏蛋白质和所导致疾病表型（例如，代谢紊乱的症状显现）的不同性质，给定病症将通常仅出现对于特定病症为特征性的症状。例如，患有法博氏病的患者可以出现上述症状中的特定亚组，例如，但不限于，温度不耐受、角膜眩晕（corneal whirling）、疼痛、皮疹、恶心或腹泻。患有高歇尔氏综合征（Gaucher syndrome）的患者可以出现脾大、肝硬变、惊厥、张力过高、呼吸暂停、骨质疏松或皮肤变色。

[0191] 除了施用本文描述的一种或多种N-糖基化改变的分子，代谢紊乱也可以通过适当的营养和维生素（例如，辅因子疗法）、物理疗法和疼痛药物疗法来治疗。

[0192] 依赖于给定代谢紊乱的特定性质，患者可以在任何年龄出现这些症状。在许多情况下，症状可以在儿童期或成人期早期出现。例如，法博氏病的症状可以在年轻时（early age）出现，在10或11周岁的年龄。

[0193] 如用于本文，“有风险发展出代谢紊乱”（例如本文所述的代谢紊乱）的受试者是具有发展出紊乱的倾向（predisposition）的受试者，即，由于酶中的突变而产生的发展出代谢紊乱的遗传倾向，所述酶例如 $\alpha$ -L-艾杜糖苷酸酶、 $\beta$ -D-半乳糖苷酶、 $\beta$ -葡糖苷酶、 $\beta$ -己糖胺酶、 $\beta$ -D-甘露糖苷酶、 $\alpha$ -L-岩藻糖苷酶、芳基硫酸酯酶B、芳基硫酸酯酶A、 $\alpha$ -N-乙酰半乳糖胺酶、天冬氨酰葡萄糖胺酶、艾杜糖醛酸-2-硫酸酯酶、 $\alpha$ -氨基葡萄糖-N-乙酰转移酶、 $\beta$ -D-葡糖醛酸酶、透明质酸酶、 $\alpha$ -L-甘露糖苷酶、 $\alpha$ -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、硫酯酶、组织蛋白酶K或脂蛋白脂肪酶。显然，“有风险发展出代谢紊乱”的受试者不是感兴趣的物种中的全部受试者。

[0194] “疑似患有病症”的受试者是具有病症（例如，代谢紊乱或任何其它本文所述的病症）的一种或多种症状（例如，任何本文所述的那些症状）的受试者。

[0195] (ii) 癌症

[0196] 癌症是一类这样的疾病或病症，其特征在于细胞不受控制的分裂和这些细胞扩散的能力，所述扩散或者通过侵入直接生长进入相邻组织或者通过移植入远距离位点（其中癌细胞经过血流或淋巴系统转运）。癌症可以侵袭各个年龄的人，但是风险有随年龄增加的趋势。癌症的类型可以包括，例如，肺癌、乳腺癌、结肠癌、胰腺癌、肾癌、胃癌、肝癌、骨癌、血液癌（hematological cancer）、神经组织癌（neural tissue cancer）、黑素瘤、甲状腺癌、卵巢癌、睾丸癌、前列腺癌、宫颈癌、阴道癌或膀胱癌。

[0197] 如用于本文，“有风险发展出癌症”的受试者是具有发展出癌症的倾向的受试者，即，发展出癌症的遗传倾向，例如，肿瘤抑制基因中的突变（例如，BRCA1、p53、RB或APC中的突变）或已经暴露于能够导致癌症的条件。因此，当受试者已经暴露于诱变或致癌水平的特定化合物（例如，香烟烟雾中的致癌化合物例如丙烯醛、砷、苯、苯并蒽（Benz{a}anthracene）、苯并芘（Benzo{a}pyrene）、钋-210（氡）、氨基甲酸酯（Urethane）或氯乙烯）时，所述受试者也可以是“有风险发展出癌症”的受试者。此外，当受试者已经暴露于例如大剂量的紫外线或X线辐射或暴露于（例如感染）引起肿瘤/与肿瘤相关的病毒如乳头状瘤病毒、EB病毒（Epstein-Barr virus）、乙型肝炎病毒或人T-细胞白血病-淋巴瘤病毒时，所述受试者可以是“有风险发展出癌症”的受试者。从上文可以清楚的是“有风险发展出癌症”的受试者不是感兴趣的物种中的全部受试者。

[0198] “疑似患有癌症”的受试者是具有癌症的一种或多种症状的受试者。癌症的症状是

本领域技术人员熟知的并且包括,但不限于,乳腺肿块、乳头变化、乳腺囊中、乳腺疼痛、体重减轻、虚弱、过度疲劳、进食困难、食欲丧失、久咳、恶化的呼吸急促 (worsening breathlessness)、咳血、血尿、血便、恶心、呕吐、肝转移、肺转移、骨转移、腹胀、气胀、腹膜腔积液 (fluid in peritoneal cavity)、阴道出血、便秘、腹胀、结肠穿孔、急性腹膜炎 (感染、发热、疼痛)、疼痛、吐血、大量出汗 (heavy sweating)、发热、高血压、贫血、腹泻、黄疸、眩晕、寒战、肌肉痉挛、结肠转移、肺转移、膀胱转移、肝转移、骨转移、肾转移和胰腺转移、吞咽困难等。

[0199] 除了施用一种或多种本文所述的N-糖基化改变的分子,也可以通过化疗剂、电离辐射、免疫治疗剂或超高温治疗剂来治疗癌症。化疗剂包括,例如,顺铂(cisplatin)、卡铂(carboplatin)、丙卡巴肼(procarbazine)、氮芥(mechlorethamine)、环磷酰胺(cyclophosphamide)、喜树碱(camptothecin)、阿霉素(adriamycin)、异环磷酰胺(ifosfamide)、美法仑(melphalan)、苯丁酸氮芥(chlorambucil)、白消安、亚硝基脲、放线菌素D(dactinomycin)、柔红霉素(daunorubicin)、多柔比星(doxorubicin)、博来霉素(bleomycin)、普卡霉素、丝裂霉素(mitomycin)、依托泊苷(etoposide)、verampil、鬼臼毒素(podophyllotoxin)、他莫昔芬(tamoxifen)、紫杉醇(taxol)、(transplatin)、5-氟尿嘧啶(5-fluouracil)、长春新碱(vincristine)、长春碱(vinblastine)和甲氨蝶呤(methotrexate)。

[0200] (iii) 炎性病症

[0201] 如用于本文,“炎性病症”是指其中一种或多种物质(例如,非天然存在于受试者中的物质),藉由白细胞(例如,B细胞、T细胞、巨噬细胞、单核细胞或树突细胞)的作用,不适当引发病理学应答(例如,病理学免疫应答)的过程。因此,所述炎性应答中涉及的这些细胞称作“炎性细胞”。不适当引发的炎性应答可以是其中没有外源物质(例如,抗原、病毒、细菌、真菌)存在于受试者中或受试者上的应答。不适当引发的应答可以是其中自身成分(例如,自体抗原)被炎性细胞靶向的应答(例如,自身免疫病如多发性硬化)。不适当引发的应答也可以是在程度或持续时间方面不适当的应答,例如,过敏反应。因此,不适当引发的应答的原因可以是存在微生物感染(例如,病毒、细菌或真菌的)。炎性病症的类型(例如,自身免疫病)可以包括,但不限于,骨关节炎、类风湿性关节炎(rheumatoid arthritis(RA))、脊椎关节病(spondyloarthropathies)、POEMS综合征(POEMS syndrome)、克朗氏病(Crohn's disease)、多中心性卡斯尔氏病(multicentric Castleman's disease)、系统性红斑狼疮(systemic lupus erythematosus(SLE))、多发性硬化(MS)、肌营养不良(muscular dystrophy(MD))、胰岛素依赖性糖尿病(insulin-dependent diabetes mellitus(IDDM))、皮肌炎(dermatomyositis)、多肌炎(polymyositis)、炎性神经病(inflammatory neuropathies)例如归伯二氏综合征(Guillain Barre syndrome)、脉管炎(vasculitis)例如韦格纳肉芽肿病(Wegener's granulomatosis)、结节性多动脉炎(polyarteritis nodosa)、风湿性多肌痛(polymyalgia rheumatica)、颞动脉炎(temporal arteritis)、舍格伦综合征(Sjogren's syndrome)、贝赛特氏病(Bechet's disease)、丘-施二氏综合征(Churg-Strauss syndrome)或高安氏动脉炎(Takayasu's arteritis)。炎性病症还包括特定类型的变态反应例如鼻炎(rhinitis)、鼻窦炎(sinusitis)、荨麻疹(urticaria)、寻麻疹(hives)、血管性水肿(angioedema)、特应性皮炎(atopic

dermatitis)、食物变态反应(food allergies) (例如, (坚果变态反应(nut allergy))、药物变态反应(drug allergies) (例如, 青霉素)、昆虫变态反应(insect allergies) (例如, 针对蜂螯的变态反应) 或肥大细胞病(mastocytosis)。炎性病症也可包括溃疡性结肠炎(ulcerative colitis) 和哮喘(asthma)。

[0202] “有风险患上炎性病症”的受试者是指具有一种或多种炎性病症的家族史(例如, 一种或多种炎性病症的遗传倾向)的患者或暴露于一种或多种炎症诱导条件的受试者。例如, 受试者可能已经暴露于病毒或细菌超抗原, 例如, 但不限于, 葡萄球菌肠毒素(staphylococcal enterotoxins) (SEs)、酿脓链球菌外毒素(streptococcus pyogenes exotoxin) (SPE)、金黄色葡萄球菌中毒性休克-综合征毒素(staphylococcus aureus toxic shock-syndrome toxin) (TSST-1)、链球菌促有丝分裂外毒素(streptococcal mitogenic exotoxin) (SME) 和链球菌超抗原(streptococcal superantigen) (SSA)。从上文可以清楚的是“有风险发展出炎性病症”的受试者不是感兴趣的物种中的全部受试者。

[0203] “疑似患有炎性病症”的受试者是出现炎性病症的一种或多种症状的受试者。炎性病症的症状是本领域熟知的并且包括, 但不限于, 发红、肿胀(例如, 关节肿胀)、触碰发热的关节(joints that are warm to the touch)、关节痛、僵硬、关节功能丧失、发热、寒战、疲劳、精力丧失(loss of energy)、头痛、食欲丧失、肌肉僵硬、失眠、瘙痒、鼻塞、喷嚏、咳嗽、一种或多种神经学症状如眩晕、发作或疼痛。

[0204] 除了施用一种或多种本文所述的N-糖基化改变的分子之外, 炎性病症也可以通过非类固醇抗炎药(NSAID)、缓解病情的抗风湿药(DMARD)、生物反应调节剂或皮质类固醇来治疗。生物反应调节剂包括, 例如, 抗TNF剂(例如, 可溶性TNF受体或对TNF特异性的抗体如adulimumab、英夫利昔单抗(infliximab)或依那西普(etanercept)

[0205] 使用N-糖基化改变的分子(或其药物组合物)、适用于治疗(例如, 预防或改善其一种或多种症状)任何本文所述的病症的方法在下节中陈述。

#### [0206] 药物组合物和治疗方法

[0207] N-糖基化改变的分子(例如, 目标分子如目标蛋白的改变的N-糖基化形式)可以并入药物组合物, 所述药物组合物含有治疗有效量的所述分子和一种或多种佐剂、赋形剂、载体和/或稀释剂。可接受的稀释剂、载体和赋形剂通常不负面影响受体的体内稳态(例如电解质平衡)。可接受的载体包括生物相容性、惰性或生物可吸收的盐、缓冲剂、寡糖或多糖、聚合物、粘度改善剂(viscosity-improving agent)、防腐剂等。一种示例性载体是生理盐水(0.15M NaCl, pH 7.0-7.4)。另一种示例性载体是50mM磷酸钠、100mM氯化钠。关于配制和施用药物组合物的技术的进一步细节可以在例如Remington's Pharmaceutical Sciences(Maack Publishing Co., Easton, Pa.)中找到。也可以将辅助性活性化合物(supplementary active compound)并入组合物中。

[0208] 含有N-糖基化改变的分子的药物组合物的施用可以是系统的或局部的。可以将药物组合物这样配制, 即令它们适用于胃肠外和/或非胃肠外施用。具体使用方式(administration modality)包括皮下、静脉内、肌肉内、腹膜内、经皮、鞘内、口服、直肠、口腔、局部、鼻、眼、关节内、动脉内、蛛网膜下、支气管、淋巴、阴道和子宫内施用。

[0209] 施用可以通过定期注射所述药物组合物的推注(bolus), 或者可以是不间断或连续地通过从外部储器(reservoir) (例如, IV包) 或内部储器(例如, 生物可腐蚀性植入物

(bioerodable implant)、生物人工器官(bioartificial organ)或植入的产生N-糖基化改变的分子的细胞集落)进行静脉内或腹膜内施用。参见,例如,美国专利Nos.4,407,957、5,798,113和5,800,828,将它们通过所述整体并入本文。药物组合物的施用可以使用合适的递送手段(delivery means)来实现,所述递送手段例如:泵(参见,例如,Annals of Pharmacotherapy,27:912(1993);Cancer,41:1270(1993);Cancer Research,44:1698(1984),通过所述整体并入本文);微囊化(参见,例如,美国专利Nos.4,352,883;4,353,888;和5,084,350,通过所述整体并入本文);连续释放聚合物植入物(continuous release polymer implant)(参见,例如,Sabel,美国专利No.4,883,666,通过所述整体并入本文);大囊化(macroencapsulation)(参见,例如,美国专利Nos.5,284,761、5,158,881、4,976,859和4,968,733,和公开的PCT专利申请W092/19195、W0 95/05452,将它们每一篇的公开内容通过所述整体并入本文);皮下、静脉内、动脉内、肌肉内注射或注射至其它合适部位;或在胶囊、液体、片剂、丸剂(pill)或延长释放制剂(prolonged release formulation)中口服施用。

[0210] 胃肠外递送系统的实例包括乙烯-乙酸乙烯共聚物颗粒、渗透泵(osmotic pump)、可植入输注系统(implantable infusion system)、泵递送(pump delivery)、囊化细胞递送(encapsulated cell delivery)、脂质体递送(liposomal delivery)、针递送注射(needle-delivered injection)、无针注射(needle-less injection)、喷雾器(nebulizer)、雾化器(aerosolizer)、电穿孔和经皮贴片(transdermal patch)。

[0211] 适合于胃肠外施用的制剂通常含有N-糖基化改变的分子的无菌含水制备物,其优选与受体的血液等渗(例如,生理盐水溶液)。制剂可以以单位剂量或多剂量形式存在。

[0212] 适合于口服施用的制剂可以作为离散单位如胶囊、扁囊剂(cachet)、片剂或锭剂(lozenges)存在,其各自含有预定量的N-糖基化改变的分子;或含水液体中或非水液体中的悬液,如糖浆、酏剂、乳剂或饮剂(draught)。

[0213] 可以将适合于局部施用的N糖基化改变的分子(例如,目标分子例如目标蛋白的改变的N-糖基化形式)向哺乳动物(例如,人患者)作为例如乳膏(cream)、喷雾、泡沫、凝胶、软膏(ointment)、油膏(salve)或干燥摩擦剂(dry rub)施用。干燥摩擦剂可以在施用部位再水合。N-糖基化改变的分子也可以直接注入(例如,浸入并干燥)其后可以进行局部应用的绷带、纱布或贴片。也可以将N-糖基化改变的分子以半液态、凝胶(gelled)态或完全液态保持在用于局部施用的绷带、纱布或贴片中(参见,例如,美国专利No.4,307,717,将其内容通过所述整体并入本文)。

[0214] 可以向需要的受试者以本领域技术人员可以确定的给药方案施用治疗有效量的药物组合物。例如,可以向受试者施用组合物,例如,以每剂 $0.01\mu\text{g}/\text{kg}$ - $10,000\mu\text{g}/\text{kg}$ 受试者体重的剂量系统性施用。在另一实例中,剂量是每剂 $1\mu\text{g}/\text{kg}$ - $100\mu\text{g}/\text{kg}$ 受试者体重。在另一实例中,剂量是每剂 $1\mu\text{g}/\text{kg}$ - $30\mu\text{g}/\text{kg}$ 受试者体重,例如,每剂 $3\mu\text{g}/\text{kg}$ - $10\mu\text{g}/\text{kg}$ 受试者体重。

[0215] 为了优化治疗效力,可以首先以不同的给药方案施用N-糖基化改变的分子。单位剂量和方案依赖于多种因素,包括,例如,哺乳动物的种类、其免疫状态、哺乳动物的体重。通常,组织中N-糖基化改变的分子的水平可以使用适合的、作为临床测试流程一部分的筛查测定法监测,例如,来测定给定治疗方案的效力。

[0216] N-糖基化改变的分子的给药频率在医疗工作者(medical practitioner)(例如,

医生或护士)的技术和临床判断范围之内。通常,施用方案通过临床试验建立,所述试验可以建立最适施用参数。然而,从业者可以根据受试者的年龄、健康、重量、性别和医疗状态改变这些治疗方案。给药频率可以依赖于治疗是预防性还是治疗性的而变化。

[0217] 这些N-糖基化改变的分子(例如,目标分子例如目标蛋白的改变的N-糖基化形式)或其药物组合物的毒性和治疗效力可以通过已知的药学方法在例如细胞培养物或实验动物中测定。这些方法可以用于例如测定LD<sub>50</sub>(对群体的50%致死的剂量)和ED<sub>50</sub>(在群体的50%中治疗有效的剂量)。毒性和治疗效果的剂量比是治疗指数并且可以表示为LD<sub>50</sub>/ED<sub>50</sub>的比例。展现高治疗指数的药物组合物是优选的。尽管可以使用展现毒性副作用的药物组合物,应该注意设计使这些化合物靶向患病组织的部位以尽量减小对正常细胞的损害(例如,非靶细胞),并且由此降低副作用。

[0218] 从细胞培养测定法和动物研究获得的数据可以用于配制用于适当受试者(例如,人患者)中的剂量范围。这些药物组合物的剂量通常在如下循环浓度的范围内,所述浓度包括ED<sub>50</sub>且几乎无毒或无毒。剂量可以在这个范围内依赖于采用的剂量形式和使用的施用途径而变化。对于如本文所述使用的药物组合物(例如,用于治疗受试者中的代谢紊乱),最初可以从细胞培养测定法估计治疗有效剂量。可以在动物模型中配制剂量以实现包括在细胞培养物中测定的IC<sub>50</sub>(即,实现对症状的最大抑制的一半的药物组合物浓度)的循环血浆浓度范围。这种信息可以用于更精确地测定在人中有用的剂量。血浆中的水平可以例如通过高效液相层析测量。

[0219] 如本文所定义的,“治疗有效量”的N-糖基化改变的分子是能够在受治疗的受试者中产生期望的医学结果的分子的量(例如,改善代谢紊乱的一种或多种症状或减少癌细胞增殖)。N-糖基化改变的分子的治疗有效量(即,有效剂量)包括毫克或微克量的所述化合物每千克受试者或样品重量(例如,约1微克每千克至约500毫克每千克,约100微克每千克至约5毫克每千克,或约1微克每千克至约50微克每千克)。

[0220] 受试者可以是任何哺乳动物,例如,人(例如,人患者)或非人灵长类(例如,黑猩猩(chimpanzee)、狒狒(baboon)或猴(monkey))、小鼠、大鼠、兔、豚鼠、沙鼠、仓鼠、马、牲畜类型(例如,牛、猪、绵羊或山羊)、犬、猫或鲸。

[0221] 可以将本文描述的N-糖基化改变的分子或其药物组合物与另一治疗作为联合疗法向受试者施用,所述另一治疗例如用于代谢紊乱(例如,溶酶体贮积病)的治疗。例如,联合疗法可以包括向受试者(例如,人患者)施用一种或多种额外的为受试者提供治疗益处的剂(agent),所述受试者患有或有风险发展出(或疑似患有)代谢紊乱(例如,溶酶体贮积病)。由此,化合物或药物组合物和所述一种或多种额外的剂同时施用。或者,可以首先及时(in time)施用N-糖基化改变的分子(例如,蛋白质或多萜醇),并其次及时施用所述一种或两种额外的剂。可以首先及时施用所述一种或两种额外的剂,并其次及时施用N-糖基化改变的分子(例如,蛋白质或多萜醇)。N-糖基化改变的分子可以替代或加强先前施用或当前施用的疗法。例如,当用本发明的N-糖基化改变的分子治疗时,一种或多种额外的剂的施用可以停止或减少,例如,以较低水平施用。也可以保持先前疗法的施用。在一些情况下,可以将先前疗法保持到N-糖基化改变的分子的水平(例如,剂量或进度(schedule))达到足以提供治疗效果的水平。可以联合施用两种疗法。

[0222] 应该理解的是在先前治疗有特定毒性的情况(例如,具有显著副作用谱的用于代

代谢紊乱的治疗)下,N-糖基化改变的分子(例如,蛋白质或多萜醇)的施用可以用于将先前治疗的量抵消和/或减少到足以给出相同或改进的治疗益处、但是没有毒性的水平。

[0223] 在一些情况下,当向受试者施用本发明的N-糖基化改变的分子(例如,蛋白质、多萜醇或与多萜醇连接的脂质)或药物组合物时,将第一种疗法中断。可以监测受试者的第一种预先选择的结果,例如,代谢紊乱的一种或多种症状的改善,例如任何本文所述的那些(见上文)。在一些情况下,在观察到第一种预先选择的结果的情况下,减少或中断使用N-糖基化改变的分子(例如,N-糖基化改变的蛋白质或N-糖基化改变的多萜醇)的治疗。然后在使用N-糖基化改变的分子(例如,蛋白质或多萜醇)的治疗之后,可以监测受试者第二种预先选择的结果,例如,代谢紊乱的症状的恶化。当观察到所述第二种预先选定的结果时,可以恢复或增加向受试者施用N-糖基化改变的分子(例如,蛋白质或多萜醇),或恢复所述第一种治疗的施用,或对受试者施用N-糖基化改变的分子(例如,蛋白质、多萜醇或与多萜醇连接的脂质)和第一种治疗二者或量增加的N-糖基化改变的分子(例如,蛋白质或多萜醇)和所述第一种治疗方案。

[0224] N-糖基化改变的分子(例如,蛋白质或多萜醇)也可以与用于疾病(例如,代谢紊乱)的一种或多种症状的治疗一起施用。例如,N-糖基化改变的分子(例如,蛋白质、多萜醇或与多萜醇连接的脂质)可以与例如疼痛用药(pain medication)共同施用(例如,同时或通过上述联合方案)。

[0225] 应该理解的是在一些实施方案中,N-糖基化改变的分子是其中改变的糖基化使所述分子产生医学相关产物的能力增加的分子。例如,N-糖基化改变的分子可以是能够产生治疗性产物(例如,小分子或治疗性肽)的酶,所述酶的活性通过糖基化增加或优化。这些产物和使用所述产物的方法在本公开的范围之内。

[0226] 本文所述的任何药物组合物可以与施用说明书一起包括在容器、包装或散发器(dispenser)中。

[0227] 本申请具体涉及如下项:

[0228] 1.产生目标蛋白的改变的N-糖基化形式的方法,所述方法包括:

[0229] 提供解脂西洋蓍霉或Arxula adeninivorans细胞,所述细胞经遗传工程化而包含至少一种经修饰的N-糖基化活性;和

[0230] 向细胞中引入编码目标蛋白的核酸,其中所述细胞以改变的糖基化形式产生所述目标蛋白。

[0231] 2.项1的方法,还包括分离所述目标蛋白的改变的N-糖基化形式。

[0232] 3.项1或2的方法,其中所述目标蛋白是外源蛋白。

[0233] 4.项1-3中任一项的方法,其中所述目标蛋白是内源蛋白。

[0234] 5.项1-4中任一项的方法,其中所述目标蛋白是哺乳动物蛋白。

[0235] 6.项5的方法,其中所述目标蛋白是人蛋白。

[0236] 7.项1-6中任一项的方法,其中所述目标蛋白是病原体蛋白、溶酶体蛋白、生长因子、细胞因子、趋化因子、抗体或其抗原结合片段,或融合蛋白。

[0237] 8.项7的方法,其中所述融合蛋白是病原体蛋白、溶酶体蛋白、生长因子、细胞因子或趋化因子与抗体或其抗原结合片段的融合物。

[0238] 9.项1-8中任一项的方法,其中所述目标蛋白是与溶酶体贮积病(LSD)相关的蛋白

质。

[0239] 10. 项1-9中任一项的方法,其中所述目标蛋白是葡糖脑苷酯酶或半乳糖脑苷酯酶。

[0240] 11. 项1-9中任一项的方法,其中所述目标蛋白选自下组: $\alpha$ -L-艾杜糖苷酸酶、 $\beta$ -D-半乳糖苷酶、 $\beta$ -葡糖苷酶、 $\beta$ -己糖胺酶、 $\beta$ -D-甘露糖苷酶、 $\alpha$ -L-岩藻糖苷酶、芳基硫酸酯酶B、芳基硫酸酯酶A、 $\alpha$ -N-乙酰半乳糖胺酶、天冬氨酰葡萄糖胺酶、艾杜糖醛酸-2-硫酸酯酶、 $\alpha$ -氨基葡萄糖-N-乙酰转移酶、 $\beta$ -D-葡萄糖醛酸酶、透明质酸酶、 $\alpha$ -L-甘露糖苷酶、 $\alpha$ -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、硫酯酶、组织蛋白酶K和脂蛋白脂肪酶。

[0241] 12. 项1-11中任一项的方法,其中所述改变的N-糖基化形式包括一种或多种N-聚糖结构。

[0242] 13. 项12的方法,其中所述一种或多种N-聚糖是Man<sub>5</sub>GlcNAc<sub>2</sub>。

[0243] 14. 项12的方法,其中所述一种或多种N-聚糖是Man<sub>8</sub>GlcNAc<sub>2</sub>。

[0244] 15. 项12的方法,其中所述一种或多种N-聚糖是Man<sub>9</sub>GlcNAc<sub>2</sub>。

[0245] 16. 项12的方法,其中所述一种或多种N-聚糖是Man<sub>3</sub>GlcNAc<sub>2</sub>。

[0246] 17. 项12的方法,其中所述一种或多种N-聚糖是Glc<sub>1</sub>Man<sub>5</sub>GlcNAc<sub>2</sub>。

[0247] 18. 项12的方法,其中所述一种或多种N-聚糖是Glc<sub>2</sub>Man<sub>5</sub>GlcNAc<sub>2</sub>。

[0248] 19. 项12-18中任一项的方法,其中所述目标蛋白的改变的N-糖基化形式是均质的。

[0249] 20. 项12-18中任一项的方法,其中所述目标蛋白的改变的N-糖基化形式基本上是均质的。

[0250] 21. 项20的方法,其中含有改变的糖基化的改变的目标分子之部分是至少约20%。

[0251] 22. 项20的方法,其中含有改变的糖基化的改变的目标分子之部分是至少约30%。

[0252] 23. 项20的方法,其中含有改变的糖基化的改变的目标分子之部分是至少约40%。

[0253] 24. 项19-23中任一项的方法,其中所述改变的糖基化是Man<sub>5</sub>GlcNAc<sub>2</sub>。

[0254] 25. 项19-23中任一项的方法,其中所述改变的糖基化是Man<sub>8</sub>GlcNAc<sub>2</sub>。

[0255] 26. 项19-23中任一项的方法,其中所述改变的糖基化是Man<sub>9</sub>GlcNAc<sub>2</sub>。

[0256] 27. 项19-23中任一项的方法,其中所述改变的糖基化是Man<sub>3</sub>GlcNAc<sub>2</sub>。

[0257] 28. 项19-23中任一项的方法,其中所述改变的糖基化是Glc<sub>1</sub>Man<sub>5</sub>GlcNAc<sub>2</sub>。

[0258] 29. 项19-23中任一项的方法,其中所述改变的糖基化是Glc<sub>2</sub>Man<sub>5</sub>GlcNAc<sub>2</sub>。

[0259] 30. 项1-29中任一项的方法,其中所述细胞经遗传工程化而缺乏至少一种N-糖基化活性。

[0260] 31. 项30的方法,其中N-糖基化活性是ALG3活性。

[0261] 32. 项30的方法,其中N-糖基化活性是OCH1活性。

[0262] 33. 项30的方法,其中N-糖基化活性是MNS1活性。

[0263] 34. 项30的方法,其中N-糖基化活性是MNN9活性。

[0264] 35. 项1-34中任一项的方法,其中所述至少一种修饰包含缺失编码具有N-糖基化活性的蛋白质的基因。

[0265] 36. 项1-35中任一项的方法,其中所述至少一种修饰包含表达具有N-糖基化活性

的蛋白质的突变形式。

[0266] 37. 项1-36中任一项的方法,其中所述至少一种修饰包含引入或表达RNA分子,所述RNA分子干扰具有N-糖基化活性的蛋白质的功能性表达。

[0267] 38. 项1-37中任一项的方法,其中所述至少一种修饰包含表达具有N-糖基化活性的蛋白质。

[0268] 39. 项38的方法,其中所述表达的蛋白质是由细胞中的外源核酸编码的蛋白质。

[0269] 40. 项38的方法,其中所述N-糖基化活性是葡萄糖基转移酶活性。

[0270] 41. 项38-40中任一项的方法,其中所述具有N-糖基化活性的蛋白质是ALG6。

[0271] 42. 项38的方法,其中所述具有N-糖基化活性的蛋白质是 $\alpha$ -甘露糖苷酶。

[0272] 43. 项38的方法,其中所述 $\alpha$ -甘露糖苷酶靶向内质网。

[0273] 44. 项42或43的方法,其中所述具有N-糖基化活性的蛋白质是融合蛋白,所述融合蛋白包含 $\alpha$ -甘露糖苷酶多肽和HDEL内质网保留肽。

[0274] 45. 项42的方法,其中所述 $\alpha$ -甘露糖苷酶具有5.1以下的最适pH。

[0275] 46. 项38的方法,其中所述具有N-糖基化活性的蛋白质是能够从 $\text{Man}_5\text{GlcNAc}_2$ 去除葡萄糖残基的蛋白质。

[0276] 47. 项38或46的方法,其中所述具有N-糖基化活性的蛋白质是具有 $\alpha$ -1,3-葡糖苷酶活性的蛋白质。

[0277] 48. 项46或47的方法,其中所述蛋白质是葡糖苷酶II。

[0278] 49. 项46-48中任一项的方法,其中所述蛋白质包含葡糖苷酶II的 $\alpha$ 和/或 $\beta$ 亚基。

[0279] 50. 项46的方法,其中所述蛋白质是变聚糖酶。

[0280] 51. 项1的方法,其中所述细胞经遗传工程化而包含至少两种经修饰的N-糖基化活性。

[0281] 52. 项51的方法,其中所述至少两种经修饰的N-糖基化活性包含ALG3活性的缺乏和升高的ALG6活性水平。

[0282] 53. 项1的方法,其中所述细胞经遗传工程化而包含至少三种经修饰的N-糖基化活性。

[0283] 54. 项51的方法,其中所述至少三种经修饰的N-糖基化活性包含ALG3活性的缺乏;升高的ALG6活性水平;和升高的葡糖苷酶II活性水平。

[0284] 55. 项38-54中任一项的方法,其中所述具有N-糖基化活性的蛋白质是外源蛋白。

[0285] 56. 项38-54中任一项的方法,其中所述具有N-糖基化活性的蛋白质是哺乳动物蛋白。

[0286] 57. 项56的方法,其中所述具有N-糖基化活性的蛋白质是人蛋白。

[0287] 58. 项38-54中任一项的方法,其中所述具有N-糖基化活性的蛋白质是低等真核生物蛋白。

[0288] 59. 项45的方法,其中所述低等真核生物是真菌、锥虫或原生动物。

[0289] 60. 项58的方法,其中所述低等真核生物选自下组布氏锥虫、哈茨木霉和曲霉属。

[0290] 61. 项1-60中任一项的方法,其中修饰包含表达能够影响目标蛋白的甘露糖基磷酸化的蛋白质或其生物学活性变体。

[0291] 62. 项61的方法,其中所述能够影响甘露糖基磷酸化的蛋白质或其生物学活性变

体是MNN4。

[0292] 63. 项61的方法,其中所述能够影响甘露糖基磷酸化的蛋白质或其生物学活性变体是PN01。

[0293] 64. 项61的方法,其中所述能够影响甘露糖基磷酸化的蛋白质或其生物学活性变体是MNN6。

[0294] 65. 项61-64中任一项的方法,其中糖蛋白中至少约30%的甘露糖残基是磷酸化的。

[0295] 66. 项1-65中任一项的方法,其中所述细胞未经遗传工程化而成为OCH1活性缺陷的。

[0296] 67. 项1-66中任一项的方法,还包括对糖蛋白的额外加工。

[0297] 68. 项67的方法,其中所述额外加工发生在体外。

[0298] 69. 项67的方法,其中所述额外加工发生在体内。

[0299] 70. 项67-69中任一项的方法,其中所述额外加工包括向修饰的糖蛋白添加异源模块。

[0300] 71. 项55的方法,其中所述异源模块是聚合物或载体。

[0301] 72. 项67-71中任一项的方法,其中所述额外加工包括对所述目标蛋白的改变的N-糖基化形式进行酶处理或化学处理。

[0302] 73. 项67的方法,其中所述额外加工包括用甘露糖苷酶、甘露聚糖酶、磷酸二酯酶、葡糖苷酶或糖基转移酶处理所述目标蛋白的改变的N-糖基化形式。

[0303] 74. 项72的方法,其中所述额外加工包括用氢氟酸处理所述目标蛋白的改变的N-糖基化形式。

[0304] 75. 项67-74中任一项的方法,其中所述额外加工包括将所述目标蛋白的改变的N-糖基化形式磷酸化。

[0305] 76. 具有改变的N-糖基化的分离的蛋白质,其中所述蛋白质是通过项1-75中任一项的方法产生的。

[0306] 77. 分离的解脂西洋蓍霉或*Arxula adeninivorans*细胞,所述细胞经遗传工程化而包含至少一种经修饰的N-糖基化活性。

[0307] 78. 项77的分离的细胞,其中N-糖基化活性是ALG3活性。

[0308] 79. 项77或78的分离的细胞,其中N-糖基化活性是OCH1活性。

[0309] 80. 项77-79中任一项的分离的细胞,其中N-糖基化活性是MNS1活性。

[0310] 81. 项77-79中任一项的分离的细胞,其中N-糖基化活性是MNN9活性。

[0311] 82. 项77-79中任一项的分离的细胞,其中所述修饰包含缺失编码具有N-糖基化活性的蛋白质的基因。

[0312] 83. 项77-79中任一项的分离的细胞,其中所述修饰包含表达具有N-糖基化活性的蛋白质的突变形式。

[0313] 84. 项77-83中任一项的分离的细胞,其中所述修饰包含引入或表达RNA分子,所述RNA分子干扰具有N-糖基化活性的蛋白质的功能性表达。

[0314] 85. 项77-83中任一项的分离的细胞,其中所述修饰包含表达具有N-糖基化活性的蛋白质。

- [0315] 86. 项85的分离的细胞,其中所述N-糖基化活性是葡萄糖基转移酶活性。
- [0316] 87. 项85或86的分离的细胞,其中所述具有N-糖基化活性的蛋白质是ALG6。
- [0317] 88. 项85的分离的细胞,其中所述具有N-糖基化活性的蛋白质是 $\alpha$ -甘露糖苷酶。
- [0318] 89. 项88的分离的细胞,其中所述 $\alpha$ -甘露糖苷酶靶向内质网。
- [0319] 90. 项88或89的分离的细胞,其中所述具有N-糖基化活性的蛋白质是融合蛋白,所述融合蛋白包含 $\alpha$ -甘露糖苷酶多肽和HDEL内质网保留肽。
- [0320] 91. 项85的分离的细胞,其中所述具有N-糖基化活性的蛋白质是能够从Man<sub>5</sub>GlcNAc<sub>2</sub>去除葡萄糖残基的蛋白质。
- [0321] 92. 项85或91的分离的细胞,其中所述具有N-糖基化活性的蛋白质是具有 $\alpha$ -1,3-葡萄糖苷酶活性的蛋白质。
- [0322] 93. 项91或92的分离的细胞,其中所述蛋白质是葡糖苷酶II。
- [0323] 94. 项91-93中任一项的分离的细胞,其中所述蛋白质包含葡糖苷酶II的 $\alpha$ 和/或 $\beta$ 亚基。
- [0324] 95. 项91的分离的细胞,其中所述蛋白质是变聚糖酶。
- [0325] 96. 项77-95中任一项的分离的细胞,其中所述细胞经遗传工程化而包含至少两种经修饰的N-糖基化活性。
- [0326] 97. 项96的分离的细胞,其中所述至少两种经修饰的N-糖基化活性包含ALG3活性的缺乏和升高的ALG6活性水平。
- [0327] 98. 项77-97中任一项的分离的细胞,其中所述细胞经遗传工程化而包含至少三种经修饰的N-糖基化活性。
- [0328] 99. 项98的分离的细胞,其中至少三种经修饰的N-糖基化活性包含ALG3活性的缺乏;升高的ALG6活性水平;和升高的葡糖苷酶II活性水平。
- [0329] 100. 利要求77-99中任一项的分离的细胞,其中所述细胞未经遗传工程化而成为OCH1活性缺陷的。
- [0330] 101. 项85-100中任一项的分离的细胞,其中所述具有N-糖基化活性的蛋白质是人蛋白。
- [0331] 102. 项77-101中任一项的分离的细胞,其中所述修饰包含表达能够促进经修饰的糖蛋白的甘露糖基磷酸化的蛋白质或其生物学活性变体。
- [0332] 103. 项102的分离的细胞,其中所述能够影响甘露糖基磷酸化的蛋白质或其生物学活性变体是MNN4。
- [0333] 104. 项103的分离的细胞,其中所述影响甘露糖基磷酸化的蛋白质或其生物学活性变体是PN01。
- [0334] 105. 一种治疗病症的方法,所述病症可以通过施用具有改变的N-糖基化的蛋白质来治疗,所述方法包括:
- [0335] 向受试者施用项76的蛋白质,
- [0336] 其中所述受试者是患有或疑似患有可以通过施用具有改变的N-糖基化的蛋白质来治疗的受试者。
- [0337] 106. 项105的方法,其中所述病症是代谢紊乱。
- [0338] 107. 项106的方法,其中所述代谢紊乱是溶酶体贮积病(LSD)。

- [0339] 108. 项107的方法,其中溶酶体贮积病是高歇尔氏病、泰-沙二氏病、旁姆普氏病、尼-皮二氏病或法布里氏病。
- [0340] 109. 项105-108中任一项的方法,其中所述蛋白质是与LSD相关的蛋白质。
- [0341] 110. 项109的方法,其中所述蛋白质是葡糖脑苷酯酶或 $\alpha$ -半乳糖苷酶。
- [0342] 111. 项109的方法,其中所述蛋白质选自下组: $\alpha$ -L-艾杜糖苷酸酶、 $\beta$ -D-半乳糖苷酶、 $\beta$ -葡糖苷酶、 $\beta$ -己糖胺酶、 $\beta$ -D-甘露糖苷酶、 $\alpha$ -L-岩藻糖苷酶、芳基硫酸酯酶B、芳基硫酸酯酶A、 $\alpha$ -N-乙酰半乳糖胺酶、天冬氨酰葡萄糖胺酶、艾杜糖醛酸-2-硫酸酯酶、 $\alpha$ -氨基葡糖苷-N-乙酰转移酶、 $\beta$ -D-葡糖醛酸酶、透明质酸酶、 $\alpha$ -L-甘露糖苷酶、 $\alpha$ -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、硫酯酶、组织蛋白酶K和脂蛋白脂肪酶。
- [0343] 112. 项105-111中任一项的方法,其还包括确定所述受试者是否患有可以通过施用具有改变的N-糖基化的蛋白质来治疗的疾病。
- [0344] 113. 项105-112中任一项的方法,其中所述受试者是人。
- [0345] 114. 产生目标蛋白的改变的N-糖基化形式的方法,所述方法包括使目标蛋白与制备自解脂西洋蓍霉或Arxula adeninivorans细胞的细胞裂解物接触,所述细胞经遗传工程化而包含至少一种经修饰的N-糖基化活性,其中使所述目标蛋白与所述细胞裂解物接触导致该目标蛋白的改变的N-糖基化形式。
- [0346] 115. 产生目标蛋白的改变的N-糖基化形式的方法,所述方法包括使目标蛋白与一种或多种具有N-糖基化活性的蛋白质接触,其中所述一种或多种具有N-糖基化活性的蛋白质获得自解脂西洋蓍霉或Arxula adeninivorans细胞,所述细胞经遗传工程化而包含至少一种经修饰的N-糖基化活性,并且其中使所述目标分子与所述一种或多种具有N-糖基化活性的蛋白质接触导致该目标蛋白的改变的N-糖基化形式。
- [0347] 116. 解脂西洋蓍霉或Arxula adeninivorans细胞的基本上纯的培养物,所述培养物中的大量经遗传工程化而包含至少一种经修饰的N-糖基化活性。
- [0348] 117. 项87的基本上纯的细胞培养物,其中所述细胞培养物含有一种或多种细胞亚群,每个亚群包含不同的经修饰的糖基化活性。
- [0349] 118. 分离的核苷酸序列,其包含SEQ ID NO:1或SEQ ID NO:2。
- [0350] 119. 分离的核苷酸序列,其包含与SEQ ID NO:1或SEQ ID NO:2至少80%相同的序列。
- [0351] 120. 由项118或119的分离的核苷酸序列编码的多肽。
- [0352] 121. 分离的核酸,其包含:
- [0353] (a) 核苷酸序列,其在高严紧条件下与SEQ ID NO:1或SEQ ID NO:2的补体杂交;或
- [0354] (b) 所述核苷酸序列的补体。
- [0355] 122. 包含项118-121中任一项的核酸序列的载体。
- [0356] 123. 包含项118-121中任一项的核酸序列的表达载体。
- [0357] 124. 包含项123的表达载体的培养的细胞或所述细胞的子代,其中所述细胞表达所述多肽。
- [0358] 125. 产生蛋白质的方法,所述方法包括在允许表达所述多肽的条件下培养项124的细胞。
- [0359] 126. 项125的方法,其还包括在培养细胞之后,从细胞或培养细胞的培养基分离多

肽。

[0360] 以下是本发明的实践的实施例。不应将它们理解为以任何方式对本发明范围的限制。

## 实施例

[0361] 实施例1. 质粒、引物和菌株

[0362] 表1含有在本文所述实验中所用载体(例如,表达载体)和缺失盒的构建中使用的全部质粒的列表。在所述实验中使用了解脂西洋蓍霉的MTLY60菌株。

[0363] 表2含有在以下实施例中使用的引物(引物的名称)和引物应用的列表。

[0364] 表1

[0365] 质粒:

JMP62

pYLTsA

pYLHmL

pYLHmA

JMP113

JMP114

pRRQ2

JME 507

JME 509

JME 461

KS-LPR-URA3

KS-LPR-LEU2

Cre ARS68LEU2

[0366] 表2.

<u>引物:</u>	<u>名称:</u>	<u>用途:</u>
TCGCTATCACGTCTCTA GC (SEQ ID NO:18)	Yloch1 prom fw	扩增 YIOCH1 扩增 YIOCH1 P 片段
TCTCTGTATACTTGTAT GTACTG (SEQ ID	Yloch1 ter rev	扩增 YIOCH1 扩增 YIOCH1 T 片段

<u>引物:</u>	<u>名称:</u>	<u>用途:</u>	
NO:19)			
CTAGGGATAACAGGGT AATGGTGTGACGAAGT ATCGAG (SEQ ID NO:20)	YIOCH1 Pfrag rev	扩增包括 I-Sce I 位点的 P 片段	
CATTACCCCTGTTATCCC TAGCGAGATCATGGAC TGG (SEQ ID NO:21)	YIOCH1 Tfrag fw	扩增包括 I-Sce I 位点的 T 片段	
GACGC GGCCGCATGAG CTTCAACATTCCAAA AC (SEQ ID NO:22)	YIMNS1 ORF+Ter (Pfrag) S	扩增 YIMNS1 P 片段(ORF + 终止子)	
CTAGGGATAACAGGGT AATACAAAATTCA GAA ATAAAAAATCTTACAG (SEQ ID NO:23)	YIMNS1 ORF+Ter (Pfrag) AS	扩增 YIMNS1 P 片段(ORF + 终止子) + I-SceI	
[0368]	CATTACCCCTGTTATCCC TAAGTAACATGAGTGC TATGAG (SEQ ID NO:24)	YIMNS1 Tfrag S	扩增 YIMNS1 T 片段 (下游终止子) + I-SceI
	CGCTTAATTAAATGCAT GGAGGTATTGCTG (SEQ ID NO:25)	YIMNS1 Tfrag AS	扩增 YIMNS1 T 片段(下游终止子)
GGTGCTTCGACTATCA GTTTCGGAGGATTGGG TGATTCTTTTATG (SEQ ID NO:26)	ScMNS1 mut 269-273 S	转换成哺乳动物(mam.)高 尔基型甘露聚糖酶的 ScMNS1 突变引物=>Sc 中 概念的验证(proof of concept in Sc)	
CATAAAAAGAACACC CAATCCTCCGAAACTG ATAGTCGAAGCACC (SEQ ID NO:27)	ScMNS1 mut 269-273 AS	转换成哺乳动物高尔基型 甘露聚糖酶的 YIMNS1 突 变引物=>Sc 中概念的验证	

<u>引物:</u>	<u>名称:</u>	<u>用途:</u>
TGAGCGGCCGCTTTCT TACTTCAGAGCTGGAG (SEQ ID NO:28)	YIMNN9 P fw	YIMNN9 KO 引物
GGCTTAATTAATTGGTA GTGATATAATGTAACGC (SEQ ID NO:29)	YIMNN9 T rv	YIMNN9 KO 引物
TAGGGATAACAGGGTA ATCACGACACACATACTC ATTCAAG (SEQ ID NO:30)	YIMNN9 P rv	YIMNN9 KO 引物
ATTACCCTGTTATCCCT AGAAGGAGATGTAGCG TAAG (SEQ ID NO:31)	YIMNN9 T fw	YIMNN9 KO 引物
TGATAAAATAGCTTAGAT ACCACAG (SEQ ID NO:32)	LIP2 rv	用于测序的反向引物
ACATACAACCACACAC ATC (SEQ ID NO:33)	5' hp4d	用于测序的正向引物
GGCGGATCCATGGTGC TGCACCCGTTTC (SEQ ID NO:34)	YIMNN4 BamHI fw	用于扩增 YIMNN4 的正向 引物
GGCCCTAGGCTACTCA AACTCCTCGCGAAC (SEQ ID NO:35)	YIMNN4 AvrII rv	用于扩增 YIMNN4 的反向 引物
GGTCTCGCCAGCGCGC CCACCCCTTTC (SEQ ID NO:36)	HAC1FW06-003	HAC1 剪接位点周围的正 向引物区
CTAGATCAGCAATAAA GTCGTGCTGGGC (SEQ ID NO:37)	HAC1Rv06-001	HAC1 剪接位点周围的反 向引物区

[0369]

<u>引物:</u>	<u>名称:</u>	<u>用途:</u>
[0370] GGATCCATGTCTATCAA GCGAGAAGAG TCC (SEQ ID NO:38)	HAC1Fw06-002	扩增包括起始密码子和BamHI 限制位点的 HAC1 基因
	CCTAGGCTAGATCAGC AATAAAGTCGTGCTGG GC (SEQ ID NO:39)	扩增包括终止密码子和AvrII 限制位点的 HAC1 基因

[0371] 实施例2.解脂西洋蓍霉OCH1和MNN9的破坏

[0372] 敲除解脂西洋蓍霉中OCH1 (GenBank® 登录号: AJ563920) 和MNN9GenBank® 登录号: AF441127) 两种基因的策略如Fickers等((2003) J Microbiol Methods. 55 (3) :727-37) 中所述用于LIP2基因的来制定。OCH1基因所遵循的基因构建策略示于图5。

[0373] 将OCH1K0片段通过限制消化并通过PCR自质粒Y1OCH1PUT TOP0分离,并且转化至解脂西洋蓍霉菌株MTLY60。获得了20株尿嘧啶原养型菌株,并使用引物Yloch1prom fw (SEQ ID NO:18) 和Yloch1ter rev (SEQ ID NO:19) 通过对基因组DNA (gDNA) 进行PCR筛选来分析所述质粒的基因组整合。在测试的20个克隆的2个中扩增到了正确大小的片段(即,2618bp vs. 野生型中的1894bp)。数个克隆含有随机整合的构建体拷贝并因此扩增到了所述两种片段。

[0374] 为了去除URA3基因,将两个阳性克隆用附加型质粒pRRQ2转化,所述质粒pRRQ2含有Cre重组酶的表达盒。通过对gDNA使用引物Yloch1 prom fw和Yloch1 ter rev(见上文)进行PCR来筛选URA3基因的去除。在所述阳性克隆中缺失2328bp片段(包括URA3) 而存在1075bp的1075bp(不含URA3) 片段。

[0375] 对2个阳性克隆进行了Southern印迹分析以检查是否发生了异常DNA整合。将基因组DNA (gDNA) 用EcoRV/HindIII双重消化,进行琼脂糖凝胶电泳,再转移至硝酸纤维素膜。将所述膜用来自质粒Y1OCH1 PT TOP0的500bp SpeI/I-SceI片段探查(probe)。 $\Delta$  och1 PUT中存在1456bp的片段,而 $\Delta$  och1 PT中存在2066bp的片段且野生型菌株中存在2893bp的片段。

[0376] 制定使MNN9失活的构建策略并示于图6。

[0377] 通过NotI/PacI双重消化将破坏片段从质粒Y1MNN9PUT TOP0切下并转化至MTLY60 和 $\Delta$  och1PT克隆9。对于两种菌株获得了数个URA3阳性克隆,在分离了单克隆之后通过对gDNA进行PCR来筛选它们中构建体的正确整合。使用引物Y1MNN9P fw和Y1MNN9T rv. (表2) 在破坏体(disruptant) 菌株中扩增了2349bp的片段,而在非转化体中扩增了2056bp的片段。

[0378] 为了分析通过突变菌株合成的N-聚糖结构,对从甘露糖蛋白衍生的聚糖进行了DSA-FACE(图7)。野生型(MTLY60) 菌株作为主要核心类型聚糖结构主要具有Man<sub>8</sub>GlcNAc<sub>2</sub>(结构式I;图4) 和大量的Man<sub>9</sub>GlcNAc<sub>2</sub>(结构式II;图4),后者最可能含有因Och1p活性产生的额外的甘露糖。另外,可以观察到一些较大的结构。 $\Delta$  och1菌株主要具有Man<sub>8</sub>GlcNAc<sub>2</sub>(结构式I) 和小部分的Man<sub>9</sub>GlcNAc<sub>2</sub>(结构式II;图4),其二者均对α-1,2-甘露糖苷酶处理敏感(表明 $\Delta$  och1α-1,2-man),导致剪切(trimming) 成Man<sub>5</sub>GlcNAc<sub>2</sub>(结构式IV;图4)。 $\Delta$  mnn9菌株积累

比 $\Delta$ och1菌株更多的Man<sub>9</sub>GlcNAc<sub>2</sub>(结构式II;图4),这表明Mnn9p涉及接着Och1p活性发生的聚糖结构的延伸。双突变 $\Delta$ och1 $\Delta$ mnn9展现了一种类似 $\Delta$ och1菌株的糖基化表型。

[0379] 实施例3.MNS1的诱变

[0380] MNS1(ER $\alpha$ -1,2-甘露糖苷酶)涉及将Man<sub>9</sub>GlcNAc<sub>2</sub>剪切Man<sub>8</sub>GlcNAc<sub>2</sub>并且具有严格的底物特异性,因为它仅能剪切与中央臂的 $\alpha$ -1,3-甘露糖连接的 $\alpha$ -1,2-甘露糖(图2)。为了测定能够在何处诱变MNS1基因从而使其底物特异性转换成高尔基型 $\alpha$ -1,2-甘露糖苷酶,将几种ER型甘露糖苷酶的初级序列与高尔基型甘露糖苷酶进行了比较。鉴定了一个在所述两类之间不同的区域。此外,还分析了在高尔基型甘露糖苷酶的催化位点中结晶的寡糖进入酵母MNS1(an oligosaccharide that was crystallised in the catalytic site of the Golgi type mannosidase into the yeast MNS1)以鉴定可能的糖和蛋白质之间的相互作用。令人惊讶的是,使用两种方法鉴定了相同的位点。

[0381] 将来自酿酒酵母(**GenBank®**登录号:Z49631,sgd:YJR131W)的MNS1基因突变从而改变其底物特异性。在相同区域内制造了三种突变版本:两种具有一个突变(R273L和R273G),和一种具有3个突变(R269S/S272G/R273L):

[0382] A) R273L(精氨酸273成为亮氨酸)

[0383] B) R273G(精氨酸273成为甘氨酸)

[0384] C) R269S/S272G/R273L(精氨酸269成为丝氨酸/丝氨酸272成为甘氨酸/精氨酸273成为亮氨酸)。

[0385] 全部突变使用Quick Change(Stratagene)诱变试剂盒制造。

[0386] 使构建体在强组成型TPI1启动子调控下表达3种不同的突变基因。使用寡核苷酸CGACTATCCGGTTCGGATCATGGGTGATTCTTTATGAG(SEQ ID NO:40)和CTCATAAAAAGAACATCACCC AATGATCCGAACCGGATAGTCG(SEQ ID NO:41)来生成突变R273L,并使用寡核苷酸CGACTATCCG GTTCGGATCAGGTGGTGATTCTTTATGAG(SEQ ID NO:42)和CTCATAAAAAGAACATCACACCTGATCCGA ACCGGATAGTCG(SEQ ID NO:43)使用野生型基因作为模板来生成突变R273G。使用寡核苷酸GGTGCTTCGACTATCAGTTCGGAGGATTGGGTGATTCTTTATG(SEQ ID NO:44)和CATAAAAAGAACATCA CCCAATCCTCCGAAACTGATAGTCGAAGCACC(SEQ ID NO:45)施用突变R273L作为模板DNA来获得突变R269S/S272G/R273L。通过使用寡核苷酸CCCGATATCGGATCCATGAAGAACTCTGCGGTATTTC(SEQ ID NO:46)和GGGAAGCTTAACGCGGTTCCAGCGGGATACGGCACCGCGCACCAACGACCAAC CTGTGGTCAG(SEQ ID NO:47)的PCR反应,在突变体和野生型MNS1开放阅读框的3'端添加E-标记物的编码序列从而允许在表达之后的蛋白检测。对构建策略的概括示于图8。

[0387] 将三种构建体以及未突变的基因(作为阴性对照)转化至酿酒酵母菌株XW27(MAT $\alpha$  leu2 ura3 trp1 his3 ade2 lys2 och1::LEU2 mnn1::URA3mnn6::ADE2),使用TRP1作为选择标记,在用XbaI消化质粒之后将构建体引导至酿酒酵母基因组中的TRP1基因座。在后菌株能够合成统一的Man<sub>8</sub>GlcNAc<sub>2</sub>(在其糖蛋白上)。如果经突变的酶是活性的,那么这种Man<sub>8</sub>GlcNAc<sub>2</sub>(结构式I;图4)应该被剪切成Man<sub>5</sub>GlcNAc<sub>2</sub>(结构式IV;图4)、Man<sub>6</sub>GlcNAc<sub>2</sub>(结构式V;图4)和/或Man<sub>7</sub>GlcNAc<sub>2</sub>(结构式VI;图4)。

[0388] 分离了色氨酸原养型菌株,培养在液体SDC-trp培养基中,并制备了甘露糖蛋白。通过DSA-FACE分析源自甘露糖蛋白的N-聚糖。能够从图9看出,来自含有R273G和R269S/S272G/R273L突变的菌株的小量Man<sub>8</sub>GlcNAc<sub>2</sub>(结构式I;图4)被转化成Man<sub>5</sub>GlcNAc<sub>2</sub>(结构式

IV;图4)、 $\text{Man}_6\text{GlcNAc}_2$ (结构式V;图4)和 $\text{Man}_7\text{GlcNAc}_2$ (结构式VI;图4)。其它突变体或野生型基因的表达引起改变的N-糖基化表型。为了评估全部突变体是否同样表达良好,使用对E-标记物(添加至MNS1蛋白的13个氨基酸的表位)特异性的抗体进行了Western印迹分析。全部突变蛋白以及野生型MNS1蛋白等均同样表达良好。

[0389] 实施例4.增加磷酸化

[0390] 解脂西洋蓍霉MNN4的表达

[0391] 为了增加 $\text{Man}_8\text{GlcNAc}_2$ 的磷酸化,将解脂西洋蓍霉MNN4(巴斯德毕赤酵母PN01的同源物)在解脂西洋蓍霉中过表达以促进N-聚糖的核心型磷酸化。

[0392] 使用引物GGCGGATCCATGGTGCTGCACCGTTTC(Y1MNN4 BamHI fw; SEQ ID NO:34)和GGCCCTAGGCTACTCAAACCTCTCGCGAATC(Y1MNN4AvrII rv; SEQ ID NO:35)扩增了解脂西洋蓍霉MNN4(XM\_503217, YALI0D24101g)基因的编码序列。将此开放阅读框(ORF)使用BamHI和AvrII位点克隆至质粒中,其将所述ORF置于质粒pY1HURA3的hp4d启动子调控下,所述质粒pY1HURA3含有作为选择标记的URA3d1基因和用于改善随机整合的zeta序列(图10)。

[0393] 在转化入MTLY60  $\Delta$  och1菌株之前,用Eco47III(用于整合入URA3基因座)、PvuI(用于整合入MNN4基因座)或RsrII/BstBI(用于随机整合)消化含有MNN4表达盒的质粒。通过PCR使用hp4d启动子中的引物和一个LIP2终止子中的引物分析了靶向URA3和MNN4基因座的转化体。通过Southern印迹分析评估了随机整合了构建体的转化体。

[0394] 为了评估甘露糖磷酸化是否增加,我们在YPD培养基中的48小时培养之后通过DSA-FACE毛细管电泳分析了自分泌的糖蛋白衍生的N-聚糖(图11)。 $\text{Man}_8\text{GlcNAc}_2$ (结构式I)的量剧烈降低有利于(in favour of)两种结构,所述两种结构迁移更快(与 $\text{Man}_8\text{GlcNAc}_2$ (结构式I;图4)相比)并且很有可能分别含有单(P)(结构式X或XI;图4)和双(PP)(结构式XII;图4)(图11)。因此,可以总结随机整合的表达盒按顺序比整合入URA3基因座或MNN4基因座的表达盒表现更好。MZ2展现了最高的磷酸化水平。

[0395] 假设两个峰均衍生自 $\text{Man}_8\text{GlcNAc}_2$ (结构式I;图4)峰,测定了转化成磷酸化聚糖的 $\text{Man}_8\text{GlcNAc}_2$ 的量(表3)。

[0396] 表3

N-聚糖结构	高度	面积	%信号 磷酸状态
<b>菌株 och1</b>			
M82P(结构形式 XII)	18	302	1,02826 18.91045*
M8P(结构形式 X 或 XI)	261	5252	17,88219
M8(结构形式 I)	928	23816	81,08955 81.08955*
	29370		100 100

菌株 MU5				%信号
[0398]	M82P (结构形式 XII)	1319	19736	27,16773 81.17283*
	M8P (结构形式 X 或 XI)	2025	39232	54,00509
	M8 (结构形式 I)	539	13677	18,82717 18.82717*
			72645	100 100

菌株 MZ2				%信号
[0399]	M82P (结构形式 XII)	1182	17662	27,75299 83.11282*
	M8P (结构形式 X 或 XI)	1803	35231	55,35984
	M8 (结构形式 I)	419	10747	16,88718 16.88718*
			63640	100 100

[0400] 表3图例:高度和面积是指从电泳图谱测定的峰的高度和峰的面积。“%信号”是指每种聚糖在N-聚糖混合物中的比例。由星号表示的数显示磷酸化的Man<sub>8</sub>Gn<sub>2</sub>的比例(顶部)和非磷酸化的Man<sub>8</sub>Gn<sub>2</sub>的比例(底部)。

[0401] 这些结果表明在过表达Y1MNN4基因的菌株中多于80%的存在于亲本Δ och1中的Man<sub>8</sub>GlcNAc<sub>2</sub>(结构式I;图4)是磷酸化的。

[0402] 实施例5.通过在内质网中与脂质连接的寡糖修饰修饰糖基化

[0403] 材料和方法

[0404] 菌株、培养条件和试剂。将大肠杆菌菌株MC1061或TOP10或DH5α用于重组质粒DNA的扩增并且在37℃在补充有100μg/ml羧苄青霉素或50μg/ml卡那霉素(根据使用的质粒)的Luria-Broth(LB)培养基中在热振荡器(thermal shaker)中培养

[0405] 使用解脂西洋蓍霉MTLY60(ura3leu2)菌株作为亲本菌株。将全部酵母培养在28℃保温箱中。将它们在YPD培养基(2%葡萄糖,2%细菌蛋白胨和1%酵母提取物)或合成葡萄糖完全(SDC)培养基(0.17%w/o氨基酸且不含硫酸铵的YNB,1%葡萄糖,0.5%NH<sub>4</sub>Cl,50mM K/Na磷酸盐缓冲液pH 6.8和0.077%完全补充混合物(Qbiogene Inc,Morgan Irvine,CA))上。为了选择Ura<sup>+</sup>和Leu<sup>+</sup>转化体,分别加入0.077%CSM-ura或CSM-leu。

[0406] 标准遗传技术。解脂西洋蓍霉的转化用感受态细胞如Boisrame等(1996)J.Biol.Chem.271(20):11668-75所述制备,将其公开内容通过所述整体并入本文。使用公开的规程(Epicenter试剂盒目录号:MPY80200;Epicenter Biotechnologies,Madison,WI)分离全部酵母菌株的基因组DNA。所述规程涉及在65℃的非酶促细胞溶解,其后通过沉淀去除蛋白质,并进行核酸沉淀和重悬。PCR扩增在50μl终体积中进行,其中含有5μl 10x缓冲液(200mM Tris-HCl pH8.4和500mM KC1),可变量的MgCl<sub>2</sub>,2.5μM dNTP,50ng模板,50pmol的正确引物和2.5单位的Taq或Pfu DNA聚合酶。使用的循环条件如下:94℃变性10分钟,其后是热启动和30个循环的94℃45秒、在合适的退火温度进行45秒和在72℃每kb延伸1分钟,其后是在72℃延伸10分钟。使用NucleoSpin Extract II(Macherey-Nagel)纯化从凝胶回收的DNA片段(PCR产物或片段)。通过VIB Genetic Service Facility(Antwerp,Belgium)进行DNA测序。

## [0407] 载体构建

[0408] (i) ALG3基因的敲除(基因替代)。将ALG3基因(**GenBank®**登录号:XM\_503488, Genolevures:YALI0E03190g)的启动子片段(P)从解脂西洋蓍霉MTLY60菌株的基因组DNA通过PCR扩增,所述PCR使用5' CAGTGGGCCGCACTCCCTCTTTCACTCACTATTG3' (SEQ ID NO:48)和5' CATTACCCCTGTTATCCCTACGCTCAGATCCAATTGTTTGGTGGTC3' (SEQ ID NO:49)分别作为正向和反向引物,使用Taq聚合酶(Invitrogen)。用T4DNA聚合酶(Fermentas, Ontario, Canada)去除突出的(overhanging) A核苷酸。通过PCR从解脂西洋蓍霉MTLY60菌株扩增ALG3基因的终止子片段(T),所述PCR使用

[0409] 5' GTAGGGATAACAGGGTAATGCTCTCAAGGACGGACCAGATGAGACTGTTATCG3' (SEQ ID NO:50)和

[0410] 5' GACTTTAATTAAACCCATGTGGCACCTAACCCACATCTCCCGTC3' (SEQ ID NO:51)分别作为正向和反向引物,使用校正读码Pfu DNA聚合酶(Fermentas)。由于含有ISceI限制位点的重叠的引物序列,可以通过PCR将两个片段连接,所述PCR使用P-正向引物和T-反向引物。然后将这种共扩增子(co-amplicon)亚克隆至pCR-2.1TOPO TA(Invitrogen)载体中,并且通过测序确认所述共扩增子的序列正确性。然后将共扩增子使用NotI-PacI位点克隆入中间载体(intermediate vector)。

[0411] (ii) ALG6基因的过表达。将ALG60RF(1725bp)连同ALG6基因(**GenBank®**登录号:XM\_502922, Genolevures:YALI0D17028g)的终止子(415bp下游)从解脂西洋蓍霉MTLY60菌株的基因组DNA通过PCR克隆,所述PCR使用5' CAGTGGATCCATGAACCTCCTATTTCACTACCG3' (SEQ ID NO:52)和5' GACTCCTAGGAAGCTCCAGGTTACAAGTTGTTAC3' (SEQ ID NO:53)分别作为正向和反向引物,使用校正读码Pfu DNA聚合酶(Fermentas)。将所述序列克隆至pCR-Blunt II-TOP0(Invitrogen)中并且通过测序确认ALG60RF序列的正确性(如上)。然后,将ALG60RF通过BamHI和AvrII克隆至含有hp4d启动子的载体(pYLHmA)中,然后再通过ALG3的终止子片段中存在的独特的限制位点ClaI和HindIII克隆至中间载体中。

[0412] (iii) 选择标记盒。为了从宿主基因组DNA去除可选择的标记URA3,使用了Cre-lox重组系统,例如,如Fickers等所述((2003) J. Microbiol. Methods 55 (3):727-737所述,将其公开内容通过所述整体并入本文)。在从质粒pRRQ2(hp4d-cre, LEU2)(来自Institut National de Recherche Agronomique (INRA)的馈赠)表达Cre重组酶时,通过两个lox位点之间的重组将标记切除。在两种具有和不具有ALG6过表达盒的构建体中,将由lox位点侧接的URA3选择标记插入在载体的P和T片段之间引入的I-SceI位点,产生“PUT”构建体。

[0413] 甘露糖蛋白的制备。将酵母菌株在28℃保温箱中以250rpm旋转在50ml Falcon管中的10ml标准YPD培养基中过夜培养。然后通过在4℃以4000rpm离心使细胞沉淀。去除上清液,并且将细胞首先用2ml 0.9%NaCl溶液清洗,其后用2ml水清洗两次,再重悬在微量离心管中1.5ml 0.02M柠檬酸钠pH 7中。在将管在121℃高压灭菌90分钟之后,涡旋振荡所述管并且通过离心使细胞碎片沉淀。收集上清液并且用4倍体积的甲醇在4℃以旋转运动过夜沉淀。然后通过离心所述用醇沉淀的材料获得沉淀物。使粒状沉淀(pe11et)干燥并溶解在50μl水中。

[0414] 糖分析。使用ABI 3130DNA测序仪如Callewaert等(2001;见上文)所述进行了DNA

测序仪辅助型(DSA)、荧光团辅助型糖电泳(FACE)。简言之,将糖蛋白在RCM缓冲液(8M尿素,360mM Tris pH 8.6和3.2mM EDTA)中在50℃变性1小时,其后固定在含15μl RCM的IP平板的预润湿PVDF膜上。膜的预润湿用300μl MeOH进行,用300μl水和50μl RCM清洗3次,然后真空去除。用50μl 0.1M二硫苏糖醇将糖蛋白还原1小时,再用300μl水清洗3次。在黑暗中与50μl 0.1M碘乙酸一起温育30分钟,用于将SH基团羧甲基化,其后用300μl水清洗3次。然后将平板与100μl 1%聚乙烯吡咯烷酮360温育1小时以饱和膜上未被占据的结合位点,再用300μl水清洗3次。接下来,通过用50μl 10mM Tris-乙酸pH 8.3中的肽:N-糖苷酶F(PNGase F) x U处理3小时来释放N-聚糖。回收N-聚糖并且用荧光团8-氨基芘-1,3,6-三磺酸酯(8-aminopyrene-1,3,6-trisulfonate)(APTS)通过还原性氨基化来衍生化。这通过在37℃与1.2M柠檬酸中20mM APTS和DMSO中1M NaCNBH<sub>3</sub>的1μl 1:1混合物过夜(ON)温育并通过加入4μl水猝灭来实现。通过在Sephadex G-10树脂上进行大小分级来去除过量的标记。然后将余下的标记的N-聚糖通过蒸发进行浓缩。包括RNA酶B的N-聚糖和寡麦芽糖梯子作为大小标志物。使用Genemapper®软件(Applied Biosystems)进行数据分析。对标记的糖的糖苷酶消化在37℃在100mM NH<sub>4</sub>AC pH5中过夜进行。在过夜消化之后加入额外的Jack bean (JB) 甘露糖苷酶并且再在37℃放置24小时。

[0415] 解脂西洋蓍霉中ALG3基因的破坏

[0416] 为了破坏ALG3基因,生成包括ALG3的终止子和启动子的部分并且具有URA3选择标记盒的载体,并命名为pYLalg3PUT。整合NotI和PacI位点以线性化所述载体并由此去除与大肠杆菌相关的DNA元件。使用在启动子和终止子位点的双同源重组来用URA3可选择的标记替代ALG3,这产生了alg3::URA3突变菌株。应用的敲除策略由Fickers等(2003;见上文)描述并且利用Cre-lox重组系统,其促进有效的标记拯救(marker rescue)。当整合入基因组ALG3重叠群(contig)时,Alg3pa-1,6-甘露糖基转移酶活性应该丧失。这通过分析几个转化体的甘露糖蛋白的糖基化型来监测。将源自甘露糖蛋白的N-聚糖通过DSA-FACE(毛细管电泳)分析并且用选择的外切糖苷酶进行处理以揭示结构。24个转化体中的7个在糖基化谱中给出了变化(将其中三个示于图13)。在全部7个转化体中,能够通过PCR确认敲除盒在基因组中的正确整合。通过分析所述谱发现了三种主要的聚糖结构:(i)一种(结构式VII;图4)与RNA酶B的Man<sub>5</sub>GlcNAc<sub>2</sub>结构(后者为结构式IV;图4)运行在同样大小;(ii)一种距离一个额外的葡萄糖单位;和(iii)一种距离两个额外的葡萄糖单位。(图13)这些结果表明将这些细胞中的ALG3破坏。

[0417] α-1,6-甘露糖基转移酶Alg6p的过表达

[0418] 开发了一种策略,其中将用于第一葡萄糖基转移酶(即Alg6p)的组成性活性过表达盒并入alg3基因替代载体。将这种载体命名为pYLalg3PUT-ALG6。将这种载体的NotI/PacI片段转化入解脂西洋蓍霉MTLY60菌株。以这种方法,实现对ALG3的破坏和ALG6在hp4d启动子调控下的过表达。再次通过PCR确认了基因组中的正确整合。对源自甘露糖蛋白的N-聚糖进行的DSA-FACE分析显示半数的转化体,即,24个中的12个,与WT菌株相比展现出糖基化型的变化。ALG6的过表达导致了轻度克隆变异(图13)。

[0419] N-聚糖结构的鉴定

[0420] 为了进一步揭示来自上述实验的聚糖结构的性质,用选择的外切糖苷酶对源自甘露糖蛋白的聚糖(如上)进行了体外消化。用以下酶分析甘露糖蛋白聚糖:α-1,2-甘露糖苷

酶;α-甘露糖苷酶(JB)和葡萄糖苷酶II。三种观察到的聚糖结构为Man<sub>5</sub>GlcNAc<sub>2</sub>(结构式VII;图4)、GlcMan<sub>5</sub>GlcNAc<sub>2</sub>(结构式VIII;图4)和Glc<sub>2</sub>Man<sub>5</sub>GlcNAc<sub>2</sub>(结构式IX;图4)(图14)。这些结果表明由α-1,6-甘露糖基转移酶(例如,0ch1p)进行的甘露糖延伸非常少甚至没有。

[0421] 为了测定ALG6过表达对于促进N-糖基化位点占据是否是必须的,在三种不同的菌株(西洋蓍霉:MTLY60,MTLY60 Δ alg3和MTLY60 Δ alg3ALG6)中表达了来自解脂西洋蓍霉的脂肪酶2(LIP2)。从INRA获得了在TEF组成型启动子调控下的用于解脂西洋蓍霉LIP2的构建体。将所述表达盒转化至上述菌株并且通过对制备自经转化细胞的上清液进行SDS-PAGE分析(图28)来验证蛋白质的表达。Lip2p蛋白具有2个糖基化位点。将源自alg3-缺陷型(“敲除”)酵母菌株的Lip2p蛋白通过SDS-PAGE解析成三条不同的条带,使用考马斯蓝染色凝胶使所述条带可见(图28)。为了确认凝胶中蛋白质的三种形式是Lip2p蛋白的不同糖基化形式,对获得自alg3缺陷型(“敲除”)酵母菌株的Lip2p蛋白进行用PNGase F(从糖蛋白去除寡糖残基的酶)的处理,然后再进行SDS-PAGE分析,如上所述。用PNGase F处理Lip2p蛋白在考马斯蓝染色后的凝胶上产生了单一条带,表明先前观察到的全部三种形式的蛋白质是相同的Lip2p分子的不同糖基化形式。对于源自alg3ALG6菌株的Lip2p也是如此。然而,还原性糖基化形式的蛋白质量降低。因此,可以总结出ALG6的过表达能够(至少部分)恢复在alg3敲除的突变酵母菌株中减少的N-糖基化位点的占据。

[0422] 去除加帽的葡萄糖结构

[0423] 接下来,为了体内去除单葡萄糖基化的(结构式VIII;图4)和双葡萄糖基化的(结构式IX;图4)Man<sub>5</sub>GlcNAc<sub>2</sub>(结构式VII;图4)结构,将细胞遗传工程化以过表达葡萄糖苷酶II的α-亚基。西洋蓍霉的葡萄糖苷酶II(**GenBank® 登录号:XM\_500574**)的α亚基和布氏锥虫的葡萄糖苷酶II(**GenBank® 登录号:AJ865333**)的α亚基作为两种策略独立地克隆以过表达所述蛋白质。选择布氏锥虫葡萄糖苷酶II的α亚基的原因是它的天然底物是GlcMan<sub>5</sub>GlcNAc<sub>2</sub>(结构式VIII;图4)。将两种基因克隆在组成型hp4d启动子的调控下,并且它们的质粒含有URA3标记。将这些构建体转化入带有和不带有ALG6过表达的alg3突变酵母菌株。

[0424] 从源自含有所述构建体的经培养细胞的分泌蛋白制备寡糖并且通过DSA-FACE分析测定寡糖谱。全部转化体给出相同的DSA-FACE谱,将每种葡萄糖苷酶IIα的两个不同的克隆示于图29。从这些结果总结出西洋蓍霉或锥虫葡萄糖苷酶IIα亚基的过表达对单葡萄糖基化的(结构式VIII;图4)和双葡萄糖基化的(结构式IX;图4)Man<sub>5</sub>GlcNAc<sub>2</sub>(结构式VII;图4)结构仅具有较小的作用。

[0425] 用HDEL序列标记的解脂西洋蓍霉和布氏锥虫葡萄糖苷酶IIα-亚基的表达

[0426] 为了改进西洋蓍霉或锥虫葡萄糖苷酶IIα亚基的表达对从Man<sub>5</sub>GlcNAc<sub>2</sub>体内去除葡萄糖残基的作用,使用分子生物学技术将编码HDEL标记的核酸符合读框地添加至编码所述两种Gls IIα酶中每一种的核酸的3'端。HDEL标记意欲充当从高尔基到ER的恢复机制(retrieval mechanism)。将编码在hp4d启动子调控下的、HDEL标记的来自解脂西洋蓍霉(Y.1.)和布氏锥虫(T.b.)的葡萄糖苷酶II序列的质粒转化至过表达或不过表达ALG6基因的alg3K0菌株。如图30中可见,解脂西洋蓍霉葡萄糖苷酶IIα亚基的过表达对葡萄糖基化结构的量仅具有较小的作用。相反,过表达带有额外的HDEL标记的布氏锥虫α-葡萄糖苷酶II的α亚基导致单葡萄糖峰的降低(参见图31)。

[0427] 用变聚糖酶处理葡萄糖基化聚糖

[0428] 上述结果示范了一种减少 $\text{Man}_5\text{GlcNAc}_2$ 的单葡萄糖基化形式的示例性方法。为了从糖蛋白降低 $\text{Man}_5\text{GlcNAc}_2$ 的双葡萄糖基化形式,作为一种潜在解决方法研究了哈茨木霉的变聚糖酶。从Novozymes (Novozyme 234; Bagsvaerd, Denmark) 获得酶制备物并用于体外消化寡糖。即,将不同浓度的变聚糖酶添加至源自alg3ALG6菌株的寡糖(聚糖: $\text{Man}_5\text{GlcNAc}_2$ 、 $\text{GlcMan}_5\text{GlcNAc}_2$ 和 $\text{Glc}_2\text{Man}_5\text{GlcNAc}_2$ )。如图32的DSA-FACE谱中所示,将在寡糖中观察到的双葡萄糖峰有效降低。

[0429] 接下来,体内过表达哈茨木霉的变聚糖酶。按照为了在解脂西洋蓍霉中表达经密码子优化的cDNA合成了含有HDEL序列的变聚糖酶。将成熟蛋白符合读框地与LIP2前信号序列一起克隆在TEF1启动子的调控下(图33)。将这种构建体转化入过表达和不过表达ALG6的alg3突变酵母菌株。从含有所述构建体的经培养细胞制备了寡糖并且通过DSA-FACE分析测定寡糖谱。预期DSA-FACE谱将显示在所述寡糖中观察到的双葡萄糖峰的降低。从这些结果将总结出变聚糖酶的体内过表达在与不过表达变聚糖酶的细胞相比降低寡糖中观察到的双葡萄糖峰方面有效。

[0430]  $\text{Y1 G1sII}\alpha$ 亚基和 $\beta$ 亚基的共表达

[0431] 已知葡糖昔酶II的 $\alpha$ 亚基和 $\beta$ 亚基形成异源二聚体复合物,其中 $\beta$ -亚基负责将所述复合体恢复到ER并且也涉及底物识别,而 $\alpha$ -亚基含有催化活性。由于仅过表达葡糖昔酶II的 $\alpha$ -亚基对双葡萄糖寡糖结构具有的效果小,所以共表达 $\alpha$ -和 $\beta$ -亚基。

[0432] 从基因组DNA扩增了 $\beta$ -亚基(YALI0B03652g)的开放阅读框,所述基因组DNA使用PCR分离自MTLY60菌株,并且将所述开放阅读框克隆在TEF1和hp4d启动子的调控下。将构建体制成具有LEU2作为选择标记并具有在TEF1和hp4d启动子调控下的葡糖昔酶II $\beta$ -亚基。将这些转化至过表达和不过表达ALG6、且过表达带有和不带有HDEL序列标记的解脂西洋蓍霉葡糖昔酶II $\alpha$ 亚基的alg3敲除菌株。从分泌自细胞的蛋白质制备N-聚糖,并且将所述N-聚糖的DSA-FACE谱示于图33和图34(alg3敲除且过表达ALG6)。从这些谱可以总结,过表达来自解脂西洋蓍霉的葡糖昔酶II $\beta$ 亚基对剪切葡萄糖基化的糖确实具有正面效果。概括而言,当在TEF1启动子下表达时葡糖昔酶II的 $\beta$ 亚基的效力得到改善。当解脂西洋蓍霉葡糖昔酶II $\alpha$ 亚基含有HDEL标记时,葡萄糖基化结构的降低甚至更大(图33和34)。

[0433] 对于不过表达ALG6的alg3缺陷型细胞,对于不同的细胞群体观察到了有关葡萄糖基化结构降低的相似结果(图35)。

[0434] 曲霉G1sII a和b亚基的表达

[0435] 为了从alg3-缺陷型背景中存在的含有葡萄糖的结构去除葡萄糖残基,按照为了在解脂西洋蓍霉中表达经密码子优化的cDNA合成了黑曲霉成熟(缺乏信号肽)葡糖昔酶II $\alpha$ 和 $\beta$ ( $\alpha$ -亚基(SEQ ID NO:7;图36A-36B) $\beta$ -亚基:(SEQ ID NO:8;图37))。将黑曲霉(An)葡糖昔酶 $\alpha$ 亚基克隆在组成型TEF1和hp4d启动子的调控下,并且具有URA3基因作为选择标记。将所述表达盒(在TEF1和hp4d调控下的ORF)转化至解脂西洋蓍霉alg3ALG6菌株。将转化体候选培养在YPD中,并且通过DSA-FACE分析了来自分泌蛋白的聚糖。从图38能够推导出与非转化体(alg3ALG6)相比在转化体菌株中两种葡萄糖基化结构较不丰富。

[0436] 为了进一步降低葡萄糖基化聚糖结构,制备了具有在TEF1启动子或hp4d启动子调控下的黑曲霉葡糖昔酶II $\beta$ -亚基和作为选择标记的LEU2的构建体。将这种构建体转化至表

达An葡糖昔酶II $\alpha$ -亚基的解脂西洋蓍霉alg3ALG6菌株。预期黑曲霉葡糖昔酶II $\beta$ -亚基的表达将导致解脂西洋蓍霉细胞中葡萄糖基化结构的减少。

[0437] 实施例6. 鉴定HAC内含子以及克隆和分离HAC1基因

[0438] 解脂西洋蓍霉HAC1剪接位点。基于解脂西洋蓍霉和真菌里氏木霉以及构巢曲霉中HAC1内含子区之间的序列同源性，鉴定了解脂西洋蓍霉HAC1 (Genbank : XM\_500811 , Genolevures:Yali0B12716g) 的潜在剪接位点。预期5' 和3' 剪接位点定位在特征性环结构中并且计算内含子为29bp长。

[0439] 在所述剪接位点周围开发了引物从而鉴定所述内含子。使用基因特异性引物，从来自UPR(未折叠蛋白质应答)诱导型(通过在二硫苏糖醇(DTT)中培养)和非诱导型培养物(阴性对照)的分离的mRNA合成了第一链cDNA。然后对第一链使用引物HAC1Fw06-003和HAC1Rv06-001进行了PCR。在1.5%琼脂糖凝胶上分析了扩增产物。

[0440] 预期对于非诱导型细胞将扩增+/-400bp的片段，预期对于诱导型细胞将扩增29bp的较小的片段。从非诱导型和UPR诱导型细胞获得了正确大小的片段。对于UPR诱导型培养物获得了额外两种扩增产物。中部的片段与获得自非诱导型培养物的条带大小相同，并且被理解为未剪接的HAC1。将底部的最突出的条带从凝胶纯化并且克隆入测序载体。在测序所述构建体之后，进行了序列比对以鉴定剪接位点(图15)。从所述序列比对能够看出，剪接位点位于从解脂西洋蓍霉和(里氏木霉和构巢曲霉)HAC1序列的比较中预测的位置。所述剪接位点为29bp长。

[0441] 为了分离活性全长HAC1序列，将引物工程化为具有适合于克隆至表达载体中的限制为点。引物序列如下:Hac1y1Rv07-018:CCT AGG TCA CTC CAA TCC CCC AAA CAG GTT GCT GAC GCT CGA CTC ATA GTG AGC TAG ATC AGC AAT AAA GTC G (SEQ ID NO:54) 和 HAC1Fw06-002:GGA TCC ATG TCT ATC AAG CGA GAA GAG TCC (SEQ ID NO:55)。将10ml酵母细胞培养物在5mM DTT存在下温育1.5小时以诱导UPR应答。温育之后，从经DTT处理的细胞分离了RNA并且从所述分离的RNA使用逆转录酶制备了第一链cDNA，并且PCR使用所述cDNA作为模板和上述引物。所述PCR扩增的序列含有经剪接的HAC1，将所述序列使用标准分子生物学技术插入pCR-blunt-TOP0克隆载体并测序。

[0442] 巴斯德毕赤酵母HAC1剪接位点。基于巴斯德毕赤酵母和酿酒酵母HAC1基因的内含子区的序列同源性，在巴斯德毕赤酵母HAC1基因中鉴定了潜在的剪接位点(图16)。预期所述5' 和3' 剪接位点定位在特征性环结构中并且计算所述内含子长度为322bp。

[0443] 在预期的剪接位点周围开发了引物(HAC1Fw06-004和HAC1Rv06-005)以鉴定所述内含子(参见表4)。当内含子被去除时预期将扩增257个核苷酸的片段，若内含子仍然存在则预期扩增579bp片段。从来自UPR诱导型和非诱导型培养物的分离的mRNA合成了第一链cDNA。UPR通过向10ml指数生长的细胞培养物加入5mM DTT来诱导。将所述细胞在DTT存在下培养1.5小时。通过1.5%琼脂糖凝胶电泳分析所述扩增产物。从来自非诱导型和诱导型细胞二者的cDNA获得了约257bp的片段。

[0444] 表4. 引物

引物代码	序列 5'→3'	信息
[0445]	HAC1-Karl GAATTCATGCCCGTAGATTCTT CTC (SEQ ID NO:56)	正向引物 Hac1 基因+ 起始密码子和 EcoRI 位点
	HAC1Fw06-004 GAGTCTCCGGAGGATTCA (SEQ ID NO:57)	正向引物 剪接位点 5'周围的 Hac1 基因区
	HAC1Rv06-005 CCTGGAAGAACATAAAAGTC (SEQ ID NO:58)	反向引物 终止密码子附近的 Hac1 基因
	HAC1Rv06-009 CCTAGGCTATCCTGGAAGAAT ACAAAGTC (SEQ ID NO:59)	反向引物 Hac1 基因+ 终止密码子和 AvrII 位点
	ACTppFw07-007 GGTATTGCTGAGCGTATGC AAA (SEQ ID NO:60)	用于 QPCR 的 Act1 正向引物
	ACTppRv07-003 CCACCGATCCATACGGAGTACT (SEQ ID NO:61)	用于 QPCR 的 Act1 反向引物
	HAC1ppFw07-008 CGACCTGGAATCTGC ACTTCAA (SEQ ID NO:62)	Hac1 正向引物 QPCR
	HAC1ppRV07-004 CGGTACCACCTAAGGCTTCAA (SEQ ID NO:63)	Hac1 反向引物 QPCR
	Kar2ppFw07-009 CCAGCCA ACTGTGTTGATTCAA (SEQ ID NO:64)	Kar2 正向引物 QPCR
引物代码	序列 5'→3'	信息
[0446]	Kar2ppRv07-005 GGAGCTGGTGG AAATACCA GAGTCA (SEQ ID NO:65)	Kar2 反向引物 QPCR

[0447] 为了验证未剪接的巴斯德毕赤酵母HAC1基因的长度,使用引物HAC1-Karl和HAC1Rv06-005对基因组DNA进行了PCR。将获得的片段的长度与从来自诱导型细胞培养物的cDNA获得的PCR产物的长度比较。从基因组DNA扩增的片段比使用相同引物源自cDNA的扩增子长约300bp,这表明内含子存在于基因组DNA序列而从经剪接的mRNA中缺失。

[0448] 从凝胶分离了257bp的cDNA片段并且克隆至测序载体中。对所述片段进行测序并且进行了比对以鉴定剪接位点(图17)。为了分离并克隆经剪接的巴斯德毕赤酵母HAC1基因,开发了具有用于克隆入表达载体的限制酶位点的PCR引物(HAC1-Karl和HAC1Rv06-009)。将10ml培养物用5mM DTT进行1.5小时的UPR诱导。使用逆转录酶从分离的RNA制备了第一链cDNA,然后再使用上述引物对所述cDNA模板DNA进行了PCR。分离了经剪接的HAC1并克隆至pCR-blunt-TOP0克隆载体中用于测序。将经剪接的基因也克隆至表达载体pBLHIS

IX中在甲醇诱导型AOX1启动子的调控下,以获得载体pBLHIS IX ppHAC1spliced。HAC1基因向表达载体中的正确插入使用PCR和限制酶分析来确认。

[0449] 在酿酒酵母中,在剪接时,将未剪接mRNA中C-末端10个氨基酸的编码序列用18个氨基酸的编码序列替代。同样(in accordance),在巴斯德毕赤酵母中揭示了未剪接HAC1中C-末端45个氨基酸在剪接时也由18个氨基酸的编码序列替代,所述18个氨基酸与来自酿酒酵母序列的同源(图18)。

[0450] 实施例7.经剪接HAC1基因向解脂西洋蓍霉中的转化和诱导

[0451] 将解脂西洋蓍霉细胞(MTLY60菌株)用载体“PYHMAXHAC1y1spliced”转化,所述载体含有在hp4d启动子的表达调控下的经剪接的HAC1cDNA(上文)和作为选择标记的URA3基因。所述载体向酵母基因组中的整合使用PCR验证。将用PYHMAXHAC1y1spliced转化的MTLY60菌株在2m1在YPG中的培养物中在28℃培养24小时。将培养的细胞用YNB清洗两次,然后稀释至OD<sub>600</sub>0.6,再在用50mM磷酸盐缓冲液pH:6.8缓冲的YTG中培养24小时。然后将所述细胞稀释至OD<sub>600</sub>0.2并且再培养3代从而收获中指数期(mid-exponential phase)的细胞。向细胞的粒状沉淀加入1ml RNApure™溶液连同1g玻璃珠。通过剧烈振荡破坏细胞。通过加入150μl氯仿并用异丙醇使RNA沉淀来从破碎的细胞提取RNA。将提取的RNA也用DNA酶处理以去除任何共沉淀的DNA杂质。

[0452] 使用iScriptTMcDNA Synthesis Kit(Bio-Rad Laboratories,Hercules,CA)在20 μl总体积的反应中从800ng的RNA制备第一链cDNA。将20ng RNA当量用于实时PCR分析以测定细胞中HAC1mRNA的量。使用SYBR®绿(SYBR® green)作为检测试剂(荧光的)(Eurogentec)运行实时PCR。除了设计用于检测细胞中HAC1mRNA的量的引物,还设计了对作为对照用于所述实时PCR的ACT1(持家基因)和KAR2(UPR应答基因)基因进行定量的引物。使用肌动蛋白(持家基因)作为表达对照从比较性阈循环值计算细胞中各种基因的相对量。通过测量UPR的表达确认了所述细胞对UPR应答的诱导。在组成型启动子调控下表达HAC1的菌株中的KAR2以及HAC1的表达水平比野生型菌株MTLY60高(图39)。

[0453] 实施例8.经剪接的HAC1基因向巴斯德毕赤酵母中的转化和诱导

[0454] 培养基:对于以下实验,使用了三种类型的培养基:BYM(用于酵母的缓冲型培养基:100mM磷酸钾pH:6.0/1.34%无氨基酸的YNB/1%酵母提取物/2%蛋白胨);BGMY(用于酵母的缓冲型甘油-复合培养基:100mM磷酸钾pH:6.0/1.34%无氨基酸的YNB/1%酵母提取物/2%蛋白胨/1%甘油);和

[0455] BMGY(用于酵母的缓冲型甲醇-复合培养基:100mM磷酸钾pH:6.0/1.34%无氨基酸的YNB/1%酵母提取物/2%蛋白胨/0.5%甘油)。

[0456] 按照来自毕赤酵母表达试剂盒(Invitrogen目录号K1710-01)的电穿孔规程转化了巴斯德毕赤酵母细胞。将载体pBLHIS IX ppHAC1spliced在HIS4基因线性化以使所述构建体靶向HIS4基因组进行整合。将10微克的DNA转化入酵母细胞。在分离单菌落之后,使用对基因组DNA进行的PCR(引物HAC1-Kar1和HAC1Rv06-005)验证了构建体的正确整合。从获得自所述经转化细胞的DNA扩增了915kb和1237kb的片段,而在非转化体中(未整合所述构建体的细胞)扩增了1237kb的片段。将由此鉴定为质粒整合阳性的克隆在诱导之前在10ml BMGY培养基中培养24小时。用BYM将细胞清洗一次。向非诱导型培养物加入BMGY,而向诱导型培养物加入BYM。每12小时,为诱导型培养物补料0.5%甲醇(终浓度)。将诱导进行24小

时,之后通过离心收获细胞。为了制备RNA,将细胞与1ml RNApure<sup>TM</sup> (Genhunter Corporation, Nashville, NY) 和1g玻璃珠组合,并且通过剧烈振荡裂解。通过加入150μl氯仿并用异丙醇沉淀来提取RNA。对提取并沉淀的RNA进行DNA酶处理,使用获得自Qiagen的不含RNA酶的DNA酶(目录号79254)。使用寡dT引物和Superscript II逆转录酶(Invitrogen, 目录号18064-014)对400ng的总RNA进行逆转录反应。在实时PCR反应中使用20ng RNA当量。通过Primer Express软件(Applied Biosystems)设计了引物序列(序列参见引物表)。利用SYBR绿色荧光试剂(Eurogentec)的实时PCR在来自BioRad的iCycler仪器中运行。使用持家基因肌动蛋白作为对照从比较性阈循环值计算mRNA的相对量。对UPR的定量通过对UPR-靶基因KAR2的表达分析来进行。当将未经诱导的克隆与用甲醇诱导的相同克隆比较时,获得了高3-7倍的KAR2表达(图19)。

[0457] 通过定量PCR测定了来自另外两个克隆6和克隆8的HAC1mRNA的相对量,并且与Kar2的mRNA的相对量进行了比较。在两个克隆中均观察到了对HAC1的强诱导。KAR2mRNA的相对量似乎与HAC1mRNA的相对量相关,较高的HAC1表达水平导致较高的KAR2表达水平(图20)。

[0458] 使用荧光流式细胞仪(FFC)进行经甲醇诱导的培养物的细胞凋亡研究,并且与未经诱导的培养物的细胞凋亡比较。每次分析测量了数万细胞。在12、36和48小时的诱导之后,在FACScalibur<sup>TM</sup> (Becton Dickinson)上分析细胞。GlycoSwitchM5(GSM5)菌株具有主要为Man<sub>5</sub>GlcNAc<sub>2</sub>(结构式IV;图4)的主要核心型聚糖结构。为了检查对Hac1p的诱导是否对N-聚糖结构有影响,对1ml培养基进行了DSA-FACE分析。在对经剪接Hac1p进行48小时诱导之后获得的聚糖谱与亲本GSM5菌株的谱相似。

[0459] 制作了生长曲线以检查Hac1p诱导是否损害了巴斯德毕赤酵母的生长。与用空载体转化的菌株相比在Hac1p诱导菌株中未见生长缺陷(图22)。

#### [0460] 实施例9.Y1MNN6的表达

[0461] 在酿酒酵母中,MNN6将磷酸甘露糖残基转移至N-聚糖。因此,解脂西洋蓍霉中Y1MNN6的过表达能够导致磷酸化的增加。此外,对解脂西洋蓍霉磷酸化的额外作用通过过表达Y1MNN4和Y1MNN6获得。将Y1MNN6编码区(Genbank<sup>®</sup>登录号XM\_499811, Genolevures Ref: YALI0A06589g)使用PCR引物Y1MNN6BamHI fw (GCGGGATCCATGCACAACGTGCACGAAGC (SEQ ID NO:34)) 和Y1MNN6AvrII rv (GCGCCTAGGCTACCAGTCACTATAGTTCTCC (SEQ ID NO:35)) 从基因组以PCR扩增,并克隆入pYHmAX表达载体中用于在hp4d启动子调控下的表达。将质粒使用zeta序列转化至解脂西洋蓍霉菌株MTLY60以改善随机整合。从培养在无尿嘧啶的培养基上的细胞克隆收集分泌的糖蛋白,并且使用DSA-FACE分析合成所述糖蛋白的聚糖组成。然而,未观察到磷酸化的增加(图22)。

#### [0462] 实施例10.Hac1p表达的效果

[0463] 对异源蛋白分泌评估Hac1p过表达。将含有在诱导型AOX1启动子(pPIChygppHAC1spliced)调控下或在组成型GAP启动子调控下(pGAPhygHAC1ppsliced)的潮霉素抗性标记和经剪接的HAC1cDNA的载体转化至表达在诱导型AOX1启动子调控下的mIL-10蛋白的GS115菌株。将巴斯德毕赤酵母细胞按照来自毕赤酵母表达试剂盒(Invitrogen目录号K1710-01, Invitrogen, Carlsbad, CA)的电穿孔规程转化。将所述载体在AOX1或GAP启动子中线性化以使Hac1p基因的整合分别靶向AOX1或GAP基因座。使用PCR确

认了所述质粒向宿主基因组中的整合。

[0464] 将来自鉴定为阳性的克隆的预培养物(5ml)在YPD中培养24小时。测量了培养物中细胞在600nm(OD<sub>600</sub>)波长的浓度(OD)并且将培养物在24孔平板的各个孔中的2ml BMGY培养基中稀释至OD<sub>600</sub>为1。将培养物在BMGY中培养48小时,用BMY清洗两次,然后在BMMY中诱导24小时。每8-12小时,为培养物再次补料含1%甲醇(终浓度)的培养基。诱导之后,收获细胞的上清液并且使用三氯乙酸(TCA)从1ml上清液沉淀蛋白质。对沉淀的蛋白质进行15%SDS-PAGE。

[0465] 从SDS-PAGE,在不同的克隆之间观察到在至少一种蛋白质(mIL-10)的表达方面的克隆变异。例如,对于组成型(在GAP启动子调控下)表达Hac1p蛋白的克隆,未观察到表达水平的改进,而对于诱导型(AOX1启动子)表达Hac1p的克隆,可以鉴定出两个克隆展现了更高的mIL-10蛋白的表达水平(图40和图41)。将各个克隆的mIL-10表达与由表达mIL-10的参照GS115菌株产生的mIL-10表达比较。

[0466] 对于这些克隆进行了新的诱导。将培养24小时的预培养物在带挡板的摇瓶中的20ml BMGY中稀释至OD 1。将细胞在BMGY中培养48小时,清洗两次,然后在BMMY中诱导。每8-12小时为培养物再次补料含1%甲醇的培养基。诱导之后,收获细胞的上清液并且使用TCA从1ml的所述上清液沉淀蛋白质。在对沉淀的蛋白质进行15%SDS-PAGE之前,用PNGase F处理所述蛋白质(或不处理)以去除全部糖基化(图41)。SDS-PAGE解析出来自表达Hac1p的菌株的上清液的蛋白质含有75kDa的显著条带,该条带在参照菌株中不存在。通过质谱的方法鉴定了这个条带为Kar2p,其是最主要的UPR靶基因。使用细胞因子珠测定法(cytokine bead array)(CBA)能够证明Hac1p和mIL-10蛋白的同时诱导型表达能够导致2倍高的mIL-10蛋白表达(克隆1,图41)。CBA是对经endoH处理的mIL-10蛋白进行的。

[0467] 对异源蛋白的表面表达评估Hac1p过表达。将含有在诱导型AOX1启动子调控下(pPIChygppHAC1spliced)或在组成型GAP启动子调控下(pGAPhygHAC1ppsliced)的潮霉素抗性标记和经剪接的HAC1cDNA的载体转化至分别在诱导型AOX1启动子调控下表达成熟人干扰素-β/α-凝集素融合蛋白、成熟小鼠干扰素γ/α-凝集素融合蛋白、成熟人促红细胞生成素/α-凝集素融合蛋白或α-凝集素和小鼠凝血调节蛋白的凝集素样结构域(lectin-like domain of mouse thrombomodulin)的融合蛋白的GlycoswitchMan5菌株中。按照来自毕赤酵母表达试剂盒(Invitrogen目录号K1710-01)的电穿孔规程转化了巴斯德毕赤酵母细胞。将所述载体在AOX1或GAP启动子中线性化以使Hac1p基因分别靶向AOX1或GAP基因座进行整合。使用PCR确认了所述质粒向宿主基因组中的整合。

[0468] 将来自鉴定为阳性的克隆的预培养物(5ml)在YPD中培养24小时。测量了OD<sub>600</sub>并且将培养物在24孔平板的各个孔中的2ml BMGY培养基中稀释至OD<sub>600</sub>为1。将培养物在BMGY中培养24小时,用去离子水清洗两次,然后在BMMY中诱导(使用含1%甲醇的培养基)24小时。通过直接使用对V5表位特异性的抗体进行免疫染色证明表面表达,所述抗体与V<sub>H</sub>H编码序列在C-末端融合。诱导之后,将在补充有0.1%牛血清清蛋白的1ml PBS(pH7.2)(PBS/BSA)中的10<sup>7</sup>细胞与1μl/ml所述抗V5抗体(1μg/μl; Invitrogen)一起温育,用PBS/BSA清洗,并且与1μl/ml Alexa fluor 488标记的羊抗小鼠IgG(1μg/μl; 分子探针)一起温育。在用PBS/BSA清洗两次之后,通过流式细胞仪分析细胞(表5)。

[0469] 表5.通过流式细胞仪测定的MFI值

表达的蛋白 [0470]	野生型 毕赤酵母	毕赤酵母 +AOX GAP	毕赤酵母 +GAP HAC
	小鼠干扰素-γ	36.6	19.9
人 EPO	59.5	45.8	66.5
干扰素-β	22.6	12.4	14.4
人凝血调节蛋白	95.5	184.1	67.8

[0471] MFI=从流式细胞仪分析获得的平均荧光强度。

[0472] 对于组成型表达Hac1p蛋白的菌株,与仅表达表面蛋白的参照菌株相比,对于全部四种蛋白在表面表达水平方面没有观察到改进或观察到非常小的差异。在表达人干扰素-β的细胞中,观察到了表面表达水平的显著降低。对于过表达诱导型Hac1p(表5)的菌株,可以观察到如下:1)在人干扰素-γ表面表达菌株中,与仅表达人干扰素-βa-凝集素融合物的参照菌株相比能够观察到表面表达水平1.8倍的降低;2)对于表面表达人促红细胞生成素a-凝集素融合蛋白的菌株,与参照菌株相比能够观察到表面表达水平1.3倍的降低;3)在表面表达人干扰素-β的菌株中,与参照菌株相比没有观察到表面表达水平的差异;和4)在表面表达小鼠凝血调节蛋白凝集素样结构域的菌株中,与参照菌株相比能够观察到表面表达水平1.9倍的增加。

[0473] Hac1p的过表达对磷脂合成的影响。为了测定Hac1p产物的过表达(自经剪接的HAC1cDNA产生)是否对巴斯德毕赤酵母中的脂质代谢有影响,用上述经剪接的HAC1cDNA转化细胞并且通过电子显微镜分析测定Hac1p对细胞中脂质代谢的影响。将细胞在BMGY上培养48小时,用PBS清洗一次,然后在BMMY上再培养48小时。每8-12小时将细胞再次培养(next culture)在含有1%甲醇的培养基中。然后按照Baharaeen的方法(Baharaeen等(2004)Mycopathologia)制备用于电子显微镜的细胞。简言之,使含有戊二醛(3%)和低聚甲醛(para-formaldehyde)(1.5%,在pH 7.2的0.05M二甲胂酸钠中缓冲)的初期固定剂与所述细胞在冰上接触2小时。然后将细胞用0.05M二甲胂酸钠清洗三次20分钟。清洗之后,将细胞与6%高锰酸钾溶液在室温接触1小时,然后用0.05M二甲胂酸钠清洗三次20分钟。将实验结果示于图54。巴斯德毕赤酵母中Hac1p产物(自经剪接的HAC1cDNA产生)的过表达导致离散的膜堆叠的区域,如示于图54的电子显微照片(EM)中所示。这些结果证明Hac1p的过表达通过其对脂质代谢中涉及的基因的转录激活而确实对巴斯德毕赤酵母中的脂质代谢有强烈效果。

#### [0474] 实施例11.ManHDEL的表达

[0475] 对于与由Δoch1菌株表达的糖蛋白结合的Man<sub>5</sub>GlcNAc<sub>2</sub>,可以表达α-1,2-甘露糖苷酶以将Man<sub>8</sub>GlcNAc<sub>2</sub>切割成Man<sub>5</sub>GlcNAc<sub>2</sub>(即,高尔基型α-1,2-甘露糖苷酶活性)。这种甘露糖苷酶应该靶向分泌系统。与酿酒酵母前原交配因子(prepro mating factor)融合并且用HDEL序列标记的里氏木霉α-1,2-甘露糖苷酶(Genbank®登录号.AF212153)能够在巴斯德毕赤酵母中以及在里氏木霉和黑曲霉中将Man<sub>8</sub>GlcNAc<sub>2</sub>体内剪切成Man<sub>5</sub>GlcNAc<sub>2</sub>。制备了在解脂西洋蓍霉中在组成型hp4d启动子调控下过表达MFManHDEL(与里氏木霉α-1,2-甘露糖苷酶融合的用HDEL序列标记的酿酒酵母前原α-交配因子)的表达构建体(图23)。在用限制

酶NotI消化质粒pYHmAXManHDEL之后,将所述表达盒转化入细胞,其后使用琼脂糖凝胶电泳分离期望的片段。

[0476] 使用DSA-FACE分析从来自经转化细胞的甘露糖蛋白衍生的聚糖。仅一小部分Man<sub>8</sub>GlcNAc<sub>2</sub>被转化成Man<sub>5</sub>GlcNAc<sub>2</sub>(图24)。Man<sub>8</sub>GlcNAc<sub>2</sub>向Man<sub>5</sub>GlcNAc<sub>2</sub>的不完全转化可能是因为非最适的分泌信号。因此,将所述酿酒酵母分泌信号用源自良好表达的解脂西洋蓍霉LIP2(LIP2pre)的分泌信号替代。LIP2pre序列通过将合成寡核苷酸LIP2pre fw GATCCATG AAGCTTCCACCCTCCTCACAGCCTGCCTACCCCTGGCCGGTAC(SEQ ID NO:66)和Lip2prepro rv GTACCGGCCGGCCGCTCTGGAGAACTGCGGCCTCAGAAGGAGTGATGGGGAAAGGGAGGGCGGC(SEQ ID NO:67)杂交来制备并将所述DNA克隆入pYLHmA载体(在BamHI/AvrII位点),产生以下构建体:pYLUdL2pre。使用寡核苷酸ManHDEL Eco47III fw (GGCAGCGCTACAAACGTGGATCTCCAAC(SEQ ID NO:68))和ManHDEL AvrII rv (GGCCCTAGGTTACAACCTCGTCGTGAGCAAG(SEQ ID NO:69))从pGAPZMFManHDEL PCR扩增了ManHDEL编码序列,并将该序列克隆入pYLUdL2pre。所述构建策略示于图25。在用NotI消化质粒并分离正确片段(见上)之后将所述表达盒(具有在组成型启动子hp4d调控下的L2preManHDEL)转化至解脂西洋蓍霉Δoch1菌株。经由DSA FACE分析从分泌蛋白衍生的聚糖。发生了一些Man<sub>8</sub>GlcNAc<sub>2</sub>向Man<sub>5</sub>GlcNAc<sub>2</sub>的转化,但是所述反应是不完全的(存在Man<sub>8</sub>GlcNAc<sub>2</sub>以及中间产物Man<sub>7</sub>GlcNAc<sub>2</sub>和Man<sub>6</sub>GlcNAc<sub>2</sub>;图26)。

[0477] 为了进一步改进将Man<sub>8</sub>GlcNAc<sub>2</sub>、Man<sub>7</sub>GlcNAc<sub>2</sub>和Man<sub>6</sub>GlcNAc<sub>2</sub>剪切成Man<sub>5</sub>GlcNAc<sub>2</sub>,将所述里氏木霉α-1,2甘露糖苷酶针对在解脂西洋蓍霉中的表达进行密码子优化(SEQ ID NO:9;图42)并与LIP2前信号序列融合。将这种融合构建体在4中不同的启动子调控下表达:(i) hp4d, (ii) GAP (SEQ ID NO:10;图43), (iii) POX2, 和(iv) TEF1。将最终的表达质粒命名为pYLHUXdL2preManHDEL (SEQ ID NO:11;图44A-C) pYLGUXdL2preManHDEL (SEQ ID NO:12;图45A-) pYLPUXdL2preManHDEL (SEQ ID NO:13;图46A-C) pYLTUXdL2preManHDEL (SEQ ID NO:14;图47A-C)。在用NotI切割质粒并分离含有ManHDEL表达盒的片段之后,将全部4种质粒转化至解脂西洋蓍霉MTLY60 Δoch1菌株(在实施例2中描述)。将具有在hp4d、GAP和TEF启动子调控下的ManHDEL(质粒pYLHUXdL2preManHDEL、pYLGUXdL2preManHDEL和pYLTUXdL2preManHDEL)的转化菌株培养在YPD中。

[0478] 通过DSA FACE分析从转化菌株的分泌蛋白原生的聚糖。结果示于图48。或者,将转化体(包括整合了pYLPUXdL2preManHDEL质粒的转化体)培养在含有油酸的培养基中(蛋白质产生条件),并通过DSA-FACE分析聚糖。有关所述载体之一pYLTUXdL2preManHDEL的数据示于图49。从所述数据中能够总结出,通过48小时的培养,几乎全部聚糖转化成Man<sub>5</sub>GlcNAc<sub>2</sub>。

[0479] 实施例12. 用于受POX2启动子调控的基因表达的培养条件

[0480] 培养始于新鲜平板的单菌落,并且在50mL管中的10mL YPD中于28°C在定轨摇床上以250rpm过夜培养。然后,用所述预培养物以0.2的最终OD600接种含有22mL生产培养基(production medium)(包括2.5mL油酸乳剂)的250mL摇瓶。将这种培养物在28°C在定轨摇床上以250rpm温育。在96小时过程中在不同时间点取培养物的样品。

[0481] 油酸乳剂(20%)按照如下方法制备:

[0482] 向无菌50ml容器加入:

[0483] 20ml无菌水;

[0484] 5ml油酸;和

[0485] 125μl Tween 40。

[0486] 导致乳剂形成的超声处理在75Hz进行1分钟。

[0487] 生产培养基由如下组成:

[0488] 1%酵母提取物;

[0489] 2%胰蛋白胨;

[0490] 1%葡萄糖;和

[0491] 50mM磷酸盐pH 6.8。

[0492] 实施例13:人葡糖脑苷酯酶的表达:

[0493] 按照为了在解脂西洋蓍霉中的表达而经密码子优化的cDNA (SEQ ID NO:15;图50)以化学方法合成了人葡糖脑苷酯酶(GLCM, Swiss Prot条目号:P04062)。

[0494] 将成熟蛋白的编码序列融合至LIP2前信号序列的编码序列。将这种融合构建体克隆在油酸诱导型POX2启动子的调控下。将所得质粒命名为pYLPUXL2preGLCM (=pRAN21)。在转化之前,将质粒用NotI消化并且将含有所述表达盒的片段分离并转化至解脂西洋蓍霉菌株MTLY60、MTLY60 Δ och1 (上文实施例2中描述), 和MTLY60 Δ och1ManHDEL (实施例11中描述)。将在这三种菌株中获得的转化体如实施例12中所述培养。如上所述将蛋白质从上清液沉淀, 进行SDS-PAGE并使用大鼠单克隆抗葡糖脑苷酯酶抗体进行免疫印迹(Alessandrini等(2004) J. Invest. Dermatol 123 (6):1030-6)。示例性免疫一进分析示于图51。从图51可以认识到在破坏了och1的菌株中没有出现拖尾(泳道1、2和3), 而在WT细胞中作为拖尾观察到了蛋白质的异质性(泳道4和6)。在获得自表达ManHDEL的酵母菌株的蛋白质中没有观察到蛋白质拖尾。这些结果证明使用遗传工程化的解脂西洋蓍霉细胞MTLY60 Δ och1和MTLY60 Δ och1ManHDEL能够获得更均质的目标蛋白群体。

[0495] 实施例14:人促红细胞生成素的表达

[0496] 以化学方法合成针对在解脂西洋蓍霉中表达而经密码子优化的编码人促红细胞生成素(Epo, Swiss Prot条目号:P01588)的cDNA (SEQ ID NO:16;图52)。将所述成熟蛋白的cDNA编码序列与LIP2前信号序列的编码序列融合。将这种融合构建体克隆在油酸诱导型POX2启动子的调控下。将所得质粒命名为pYLPUXL2prehuEPO。在转化之前将质粒用NotI切割并将含有表达盒的片段分离并转化至解脂西洋蓍霉菌株MTLY60 Δ och1 (在实施例2中描述)。将转化体候选如实施例12中所述培养, 并通在使用来自R&D系统的单克隆小鼠抗人Epo抗体(克隆AE7A5)进行SDS PAGE之后通过Western印迹分析分泌的蛋白质。获得自所述细胞的EPO产物展现了非常均质的糖基化。

[0497] 实施例15:人的α-半乳糖苷酶A的表达:

[0498] 按照为了在解脂西洋蓍霉中表达而经密码子优化的cDNA (SEQ ID NO:17;图53)以化学方法合成了人α-半乳糖苷酶A(AGAL, Swiss Prot条目号:P06280)编码cDNA。

[0499] 将成熟蛋白的cDNA编码序列与LIP2前信号序列的编码序列融合。将这种融合构建体克隆在油酸诱导型POX2启动子的调控下。将所得质粒命名为pYLPUXL2preαGalase。在转化之前将所述质粒用NotI切割并且将含有所述表达盒的片段分离并转化至解脂西洋蓍霉菌株MTLY60和MTLY60 Δ och1MNN4 (在实施例4中描述)。将在这两种菌株中获得的转化体如实施例12中所述培养。在SDS-PAGE分析之后通过免疫印迹分析获得自转化体的胞外蛋白。

使用对 $\alpha$ -半乳糖苷酶A特异性的抗体(获得自Abcam的鸡多克隆抗体(ab28962)和获得自Santa Cruz Biotechnology的兔多克隆抗体(sc-25823))检测表达的人 $\alpha$ -半乳糖苷酶A蛋白。

[0500] 实施例16. 甘露糖苷酶在WT解脂西洋蓍霉中的表达

[0501] 为了测定单独的甘露糖苷酶HDEL的表达是否能够导致对真菌细胞表达的蛋白质更均质的糖基化,将含有编码甘露糖苷酶HDEL的核酸的表达盒(参见实施例11)转化入野生型解脂西洋蓍霉po1d细胞。通过DSA-FACE分析从获得自所述细胞的分泌蛋白衍生的聚糖(图55)。经分析的聚糖主要由Man<sub>5</sub>GlcNAc<sub>2</sub>和小部分的Man<sub>6</sub>GlcNAc<sub>2</sub>组成。这些结果证明,在缺乏对OCH1基因的任何破坏的条件下,单独的甘露糖苷酶HDEL的表达导致对解脂西洋蓍霉表达的蛋白质更均质的糖基化。

[0502] 其它实施方案

[0503] 尽管以结合本发明的详述描述了本发明,但前文所述意在进行示例说明而非对本发明的范围进行限制,本发明的范围由随附权利要求的范围限定。其它方面、优势和修饰形式在随附权利要求的范围之内。

## 序列表

<110> 奥克西雷恩（英国）有限公司 (OXYRANE (UK) LIMITED)

<120> 分子的糖基化

<130> 18990-018W01

<140> PCT/IB2008/001814

<141> 2008-04-03

<150> US 60/909, 904

<151> 2007-04-03

<150> US 60/940, 212

<151> 2007-05-25

<160> 75

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 358

<212> DNA

<213> 解脂西洋著霉(Yarrowia lipolytica)

<400> 1

ggctcgccca ggcgccccac ccttttcaac cacgataatg agaccgtgt ccccgactct 60  
 cctattgtga agaccgagga agtcgactct acaaacttcc tcctccacac ggagtctcc 120  
 tcccccccg aacttagctga gagcactggc tcaggctcgc catcgctgac tcgtcctgc 180  
 gacgaaactg attatcttgtt ggaccgggccc cgtcatccag cagcgtcggc acaagatttc 240  
 wtttcatca aggacgagcc cgttgacgac gagcttgac tgcattggact gtcggatgac 300  
 ttcacccctgt ttgaagacaa caagcagcct gcccagcagc actttattgc tgatctag 358

[0001]

<210> 2

<211> 915

<212> DNA

<213> 巴斯德毕赤酵母(Pichia pastoris)

<400> 2

atgcccgttag atttttctca taagacagct agcccaacttc cacctcgtaa aagagcaaag 60  
 acggaagaag aaaaggagca gcgtcgagt gaacgtatcc tacgtaatag gagagcggcc 120  
 catgttcca gagagaagaa acgaagacac gttgaatttc tgaaaacca cgtcgctgac 180  
 ctggaatctg cacttcaaga atcggccaa gccactaaca agttgaaaga aatacaagat 240  
 atcattgttt caaggttgg agccttagt ggtacctct cagatttggg tttaacagtt 300  
 ccggaaatcg attttccaa attttctgat ttggaaaaa tgtctgact ctcaacttct 360  
 tggaaatcg agaaagcatc tacatccact cgccatctt tgactggagga tctggacgaa 420  
 gatgacgtcg ctgaatatga cgacgaagaa gaggacgaag agttaccac gaaaatgaaa 480  
 gtcttaaacc acaaaaacaa gggcacatct atcaagcagg agaagttgaa tgaacttcca 540  
 ttccttttgtt catccgattt ttcagacgta gatgaagaaa agtcaactct cacacattt 600  
 aagtgtcaac acgaaacaaca acaaccatgt gacaattatg tttctactcc tttagtctt 660  
 ccggaggatt cagttgattt tattaaacca ggtactttaaa aaatagatgc cgatgagaac 720  
 ttcttgttga gttcaaaataa ttacaataa aaacacgaaa atgacaccga ctacattact 780  
 acagctccat cagttccat caatgattt tttaattttt atgacattag cgagtcaat 840  
 cggttgcacatccacgac accatttacc gctaattcat ttgattttaa tgactttgtt 900  
 ttcttccagg aatag 915

<210> 3

<211> 305

<212> PRT

<213> 解脂西洋著霉

<400> 3

Met	Ser	Ile	Lys	Arg	Glu	Glu	Ser	Phe	Thr	Pro	Thr	Pro	Glu	Asp	Leu
1					5				10			15			
Gly	Ser	Pro	Leu	Thr	Ala	Asp	Ser	Pro	Gly	Ser	Pro	Glu	Ser	Gly	Asp
20							25				30				

Lys	Arg	Lys	Lys	Asp	Leu	Thr	Leu	Pro	Leu	Pro	Ala	Gly	Ala	Leu	Pro
35							40					45			
Pro	Arg	Lys	Arg	Ala	Lys	Thr	Glu	Asn	Glu	Lys	Glu	Gln	Arg	Arg	Ile
50							55					60			
Glu	Arg	Ile	Met	Arg	Asn	Arg	Gln	Ala	Ala	His	Ala	Ser	Arg	Glu	Lys
65							70			75			80		
Lys	Arg	Arg	His	Leu	Glu	Asp	Leu	Glu	Lys	Cys	Ser	Glu	Leu	Ser	
							85			90			95		
Ser	Glu	Asn	Asn	Asp	Leu	His	His	Gln	Val	Thr	Glu	Ser	Lys	Lys	Thr
							100			105			110		
Asn	Met	His	Leu	Met	Glu	Gln	His	Tyr	Ser	Leu	Val	Ala	Lys	Leu	Gln
							115			120			125		
Gln	Leu	Ser	Ser	Leu	Val	Asn	Met	Ala	Lys	Ser	Ser	Gly	Ala	Leu	Ala
							130			135			140		
Gly	Val	Asp	Val	Pro	Asp	Met	Ser	Asp	Val	Ser	Met	Ala	Pro	Lys	Leu
145							150			155			160		
Glu	Met	Pro	Thr	Ala	Ala	Pro	Ser	Gln	Pro	Met	Gly	Leu	Ala	Ser	Ala
							165			170			175		
Pro	Thr	Leu	Phe	Asn	His	Asp	Asn	Glu	Thr	Val	Val	Pro	Asp	Ser	Pro
							180			185			190		
Ile	Val	Lys	Thr	Glu	Glu	Val	Asp	Ser	Thr	Asn	Phe	Leu	Leu	His	Thr
							195			200			205		
Glu	Ser	Ser	Ser	Pro	Pro	Glu	Leu	Ala	Glu	Ser	Thr	Gly	Ser	Gly	Ser
							210			215			220		
Pro	Ser	Ser	Thr	Leu	Ser	Cys	Asp	Glu	Thr	Asp	Tyr	Leu	Val	Asp	Arg
							225			230			235		240
Ala	Arg	His	Pro	Ala	Ala	Ser	Ala	Gln	Asp	Ile	Phe	Ile	Lys	Asp	
							245			250			255		
Glu	Pro	Val	Asp	Asp	Glu	Leu	Gly	Leu	His	Gly	Leu	Ser	Asp	Asp	Phe
							260			265			270		
Thr	Leu	Phe	Glu	Asp	Asn	Lys	Gln	Pro	Ala	Gln	His	Asp	Phe	Ile	Ala
							275			280			285		
Asp	Leu	Ala	His	Tyr	Glu	Ser	Ser	Val	Ser	Asn	Leu	Phe	Gly	Gly	Leu
							290			295			300		
[0002]	Glu														
	305														

<210> 4  
<211> 304  
<212> PRT  
<213> 巴斯德毕赤酵母

<400>	4														
Met	Pro	Val	Asp	Ser	Ser	His	Lys	Thr	Ala	Ser	Pro	Leu	Pro	Pro	Arg
1							5			10			15		
Lys	Arg	Ala	Lys	Thr	Glu	Glu	Glu	Lys	Glu	Gln	Arg	Arg	Val	Glu	Arg
							20			25			30		
Ile	Leu	Arg	Asn	Arg	Arg	Ala	Ala	His	Ala	Ser	Arg	Glu	Lys	Lys	Arg
							35			40			45		
Arg	His	Val	Glu	Phe	Leu	Glu	Asn	His	Val	Val	Asp	Leu	Glu	Ser	Ala
							50			55			60		
Leu	Gln	Glu	Ser	Ala	Lys	Ala	Thr	Asn	Lys	Leu	Lys	Glu	Ile	Gln	Asp
							65			70			75		80
Ile	Ile	Val	Ser	Arg	Leu	Glu	Ala	Leu	Gly	Gly	Thr	Val	Ser	Asp	Leu
							85			90			95		
Asp	Leu	Thr	Val	Pro	Glu	Val	Asp	Phe	Pro	Lys	Ser	Ser	Asp	Leu	Glu
							100			105			110		
Pro	Met	Ser	Asp	Leu	Ser	Thr	Ser	Ser	Lys	Ser	Glu	Lys	Ala	Ser	Thr
							115			120			125		
Ser	Thr	Arg	Arg	Ser	Leu	Thr	Glu	Asp	Leu	Asp	Glu	Asp	Asp	Val	Ala
							130			135			140		
Glu	Tyr	Asp	Asp	Glu	Glu	Glu	Asp	Glu	Glu	Leu	Pro	Arg	Lys	Met	Lys
							145			150			155		160
Val	Leu	Asn	Asp	Lys	Asn	Lys	Ser	Thr	Ser	Ile	Lys	Gln	Glu	Lys	Leu
							165			170			175		
Asn	Glu	Leu	Pro	Ser	Pro	Leu	Ser	Ser	Asp	Phe	Ser	Asp	Val	Asp	Glu
							180			185			190		
Glu	Lys	Ser	Thr	Leu	Thr	His	Leu	Lys	Leu	Gln	Gln	Gln	Gln	Gln	Gln

195	200	205
Pro Val Asp Asn Tyr Val Ser Thr Pro Leu Ser Leu Pro Glu Asp Ser		
210	215	220
Val Asp Phe Ile Asn Pro Gly Asn Leu Lys Ile Glu Ser Asp Glu Asn		
225	230	235
Phe Leu Leu Ser Ser Asn Thr Leu Gln Ile Lys His Glu Asn Asp Thr		
245	250	255
Asp Tyr Ile Thr Thr Ala Pro Ser Gly Ser Ile Asn Asp Phe Phe Asn		
260	265	270
Ser Tyr Asp Ile Ser Glu Ser Asn Arg Leu His His Pro Ala Ala Pro		
275	280	285
Phe Thr Ala Asn Ala Phe Asp Leu Asn Asp Phe Val Phe Phe Gln Glu		
290	295	300

<210> 5  
<211> 387  
<212> DNA  
<213> 解脂西洋蓍霉

```

<400> 5
ggcttcgcca gcgcgcac cctcttcaac cacgataatg agaccgttgt ccccgactct 60
cctattgtga agaccgagga agtcgactct acaaacttgc tcttccacac ggagtctcc 120
cccccccccgg aactagcttgc gggcacgttgc tcaggctcgc catgtgcac tctgtcc 180
gacgaaacttattatctgt ggacggggcg cgicatccag cagtgtatgc tgcgcacta 240
actgaccagc acgttcgcca caaaggatca ttttcatcaa ggacgagccc gtgtacgacg 300
actgttgact gcaatggact tcggatgtact tcaccctgtt tgaagacaac aaggcaggctt 360
ccccagcaca ctttatgtct gatctag 387

```

<210> 6  
<211> 1237  
<212> DNA  
<213> 巴斯德毕赤酵母

<210> 7  
<211> 2810  
<212> DNA  
<213> 人工座

<220>  
<223> 以合成方法生成的寡核苷酸

<400> 7  
gtaaaggcacg agaacttcaa gacctgtct cagtcggct tctgttaagcg aaaccgagcc 60  
ttcgcgcacg acggccgtgc ccagggtct tcttggcct ctccctacga gctggactct 120  
tcttctatcc agtcaagaga cggccacgtg cacggcacca tctgttaagtc tggtgtctccc 180

[0004]

aacgagaagg tgaagctgcc cctgggggtg tcttccctgg agictggcgc cgctcgagtg 240  
gtggtgacg aggagaagcg aatgaacggc gacatccagc tgccacacga ctctaaaggcc 300  
cgaaaggagc gataacaacga ggccgagaag tgggtctgg tcggggcct ggagctgtct 360  
aagaccgcca ccctgcgacc cgagaccgag tctgcttca cccgagtgtct gtacggcccc 420  
gacaaccagt tcgaggccgt gatccgacac gccccctct ctggcactt caagcgagat 480  
ggccagaccc acgtgcagct gaacaacaag ggctacacgt aacatgggacca ctggcgaccc 540  
aagggtggagg tggggggca gggtgacagc gagaccagg aggacgagtc tacctgggtgg 600  
gacgacttct tcggggca caccgacacc aagcccccgg gccccggatc tggtggcctg 660  
gacatcacct tccccggta caagcactg ttcggatcc cegagcacegc gacacttctg 720  
tctctgaagg agaccggagg cggcgaggac aaccacgagg agccctacgg aatgtacaac 780  
gcccacgtgt tcgagtaacg gctgccccct cccatgaccc tgcacggcgc catccccctc 840  
atgcaggccc accgaaagga ctctaccgtg ggcgtgttct ggctgaacgc cggcggagacc 900  
tgggtggaca tcgtgaagtc tacctttct cccaaaccccc tggccctggg ctggggagcc 960  
accaccgaca cccagtcac ctggttctct gactgtggcc agctggacgt ttctgttttc 1020  
ctggggccca cccccccagga gatttcaag acctaaggcc agctgaccgg ctacacccag 1080  
ctgccccccagc acttcgcacat tgccattaccac cagtgctgat ggaactacat caccgacgg 1140  
gacgtgaagg aggtggaccc aaacttcgc aagtaccaga tccctatcga ctgtatctgg 1200  
ctggacatcgc agtacacgcg cgacccaaag tacttcacct gggacccctt gtctttttcc 1260  
gaccacccatct ctatggagga geacgtggac gactgtggc gaaactgtgt ctgtatcatc 1320  
gaccacccacaa tcaagaacca ggacaatgtc tctatctgc aggagatgaa ctgtacggac 1380  
ctggccacca agaaacacca cggcgagatt tacgacggct gglttggcc cggctttct 1440  
cactggatcgc acacccatcaa cccctgtcc atcaagtgtt gggtgtctct ttcaagttc 1500  
gacaagttaa agggcacccct gtctaaacgt ttcatctggg acgacatgaa cgaccctct 1560  
gtgttcaacg gccccggagc caccatgtcc aaggacaacc tgacccacgg caactgggag 1620  
caccgagacca tccacaacgt gcacggcatc accctgtgtg acgcccacca cggccctctg 1680  
ctggagcgaa agaaggggcga gatccgacga cccttcattcc tgacccgatc ttactacgccc 1740  
ggtgcccccgc gaaatgtcgc catgtggacc ggcgacaacc agggccaccig ggacccatctg 1800  
gcccgcctcta tccccatgtt gctgaacaac ggaatcgccg gettccccctt cggccgtgtcc 1860  
gacgtggggcg gettcttcca gaaaccctct aaggactgtc tgacccgtatg gtatcaggcc 1920  
ggcatctgtt atctttttctt cggagccaccc gccccatcg acacccggacg acgagaccc 1980  
tacactgtcgc cggagccccc cggactatc atctcttcagg ccattccgact gcatgatccag 2040  
ctgctggccc cctggatcac cgcctttccac gaggcccttg tgaacggcat gcccattctg 2100  
cgaccacccatg actacgcaca cccctgggac gaggccggct tcggccatcga cgaccagctg 2160  
tacctgggtt ctaccggct gctggccaag cccgtgtgt ctgaggaggc caccacccgccc 2220  
gacatctacc tggctgacga cgagaagttac tacgactact tcgactacac ctgttaccag 2280  
ggcgctggca agcgacacac cgtggccctt cccatggaga ccgtggccctt gctgtatcg 2340  
ggcgcccaacg tgatcccccg aaaggaccga ccccgacgt ctctgtccctt gatcgatgg 2400  
gacccttaca ccctgggtt ggtgttggac aagaacggcc agggccgaegg ctctctgtac 2460  
gtggacgacgc gggagaccc tgcactacgg cgagggccctt acatccatcg acgatccgca 2520  
ttccaggatct ctcggggatc gtggggccatc agggcccaaa gaccggccgg 2580  
tacctcaaga ccatggccaa cgtcgagtg gagegagtgg tggctgtgg cccccccaaag 2640  
gatgtggccgg gcaacaccc tgcgttggccatc acgaggaggc ggccttcgtc cgcccttacc 2700  
gccccttacgc agtaccaccc tcagccggac ggcgaacccg cctacggctt ggtgaagaac 2760  
cccaacgtgg gcatcgccaa gaccctggca atcgatctt aatagcttag 2810

<210> 8  
<211> 1651  
<212> DNA  
<213> 人工

<220>  
<223> 以合成方法生成的寡核苷酸

<400> 8

```

acgcgtggcg acgccttc tcgaccggga ggcggtggcc ccgagttcgc caagtttac 60
aggacacca ccacccatc ctgtatctc caccggca tccagattcc ttctctgccc 120
gtgaacgacg actactgtga ctgtccggac ggctctgacg agccccggcac ctctgcctgt 180
gccttctgt ctgcggaaactc tgccctgacc cccggtgagc gaccggatc tgacgaccgt 240
gagctgacct ctggccctgcc cggcttctac tgtaagaaca agggccaa gcccggctac 300
gtgcccttcc agcgagttaa cgatggcatc tgigactacg actgttgttg tgacggatcg 360
gtatggatggg ctgcgaccgg cggaaaccaag tgigaggaca atgttaaaggga gatcgccaag 420
gagtggcgaa agaaggagga gaagggacag aagtctatga cccggctct gaagaagaag 480
aaggaccgtc tggtcggagc eggacggacag cagaaggagg tgaggacaa catcaagcg 540
ctggagggtgg agatccggc ccagggactg aaggctcaacg acctcgaggc cgagctggag 600
gagggtggggc acggaggaggc ctctaaggctc gtcaaggggc gagaccgggg caaggtggac 660
gtgtggctg gcctcgccaa gtctcgatg gaggagctgc gaaacggcc gatggacgtg 720
cgaaaggagc gagatggacac cggaggccgg gtiatggaaac tgcgaagagat ctgtctaa 780
ttcaagggtgg agtacaaaccc caacttcaac gacgaggggcg tgaagcgagc ctgtcgatct 840
tgggaggact acggccggcaa gggcaccctg gaggggccgg tgaacaacgc ccaggaccga 900

```

gacctggacg agatcgccaa gcccgcac gagaaggccg gcatcaactg ggagcagtgg 960  
 gagaacgagg aggacggctg tgaggctgac ctgtgttacc agctggccgc ctacctgccc 1020  
 cccctctgg tggagttcat cgagggcaag gtgcgttgc tgccgaggct gctggaggac 1080  
 aacggcatcc tgcccaaggc cgccgagacc tctacccctg agtctaagggt ggtgtctgag 1140  
 gcccgagagg ccgtgaagtc tgccgagaag gagctggccg acaagcagaa gcacgtgaag 1200  
 gaccacaagt ctgacccctgga gaccgactac ggcgtggct ctatctccg agccctgaag 1260  
 ggegtgtgta tctctaaggta ctctggcgag tacaccctacg agcactgttt tctggaccag 1320  
 accaaggcaga tcccccaaggaa gggcgaggc tctacccgaa tggcaagta caccggcattc 1380  
 ggcctgtgt ctgtggacgt gctgaacccg gccggcggaa tcgtgcctga ggatcgagtg 1440  
 accctgcagt acgccaacgg ccaggctgtt tggAACGGAC ccgcggcattc taccaccgtg 1500  
 atccctgaccc tggcgagga ggacggccatc ctgaagggtgg ccgaggacga gaagtgtgtg 1560  
 tactctatgc acgtgaccc tcccccggc tgtaaaaaaaa ggacggccac 1620  
 cccaaaccgaa aggacgagct gtaatagctt a 1651

&lt;210&gt; 9

&lt;211&gt; 1534

&lt;212&gt; DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 以合成方法生成的寡核苷酸

&lt;400&gt; 9

gagctcagcg ctaccaagcg aggctctccc aaccccaccc gagccgcgc tggtaaggcc 60  
 gccttcaga cctcttggaa cgccttccac cacttcgcct tcccccaacgca cgcacccgcac 120  
 cccgtgtcta actcggttca cgacgagcga aacggctggg gctcttcgtc catcgacggc 180  
 ctggacacccg ccattccgtat gggcgcaccc gacatcgta acaccatccg cgaatcggtg 240  
 ccccgatca acttcacccac caccgcgcgtg gccaaccagg gcatctctgt gttcgagacc 300  
 aacatccgtt acctgggggg cctcggtgtt gcttgcaccc ttctggagg cccttctct 360  
 tctctggcca ccaaccaggc cctgggttac acctcggttc gacaggccca gaccctggcc 420  
 aacggccctga aggtggctt taccacccccc tctggcggtc ccgcggccac cgtgttcttc 480  
 aaccccacccg tgcgcacgtt tctaacaacg tggccgatag cggctctcg 540  
 gtgtggagt ggacccgact gtcgtacccg acccggcaacc cccagttacgc ccacgtggcc 600  
 cagaaggccg agtcttaccc tcttgcaccc aagggttccgc ccggccctg gcccggactg 660  
 atcggcaccc tctgtgttca ctctaaacggc accttccagg acttccgg ctcttggct 720  
 ggcctgtatgg acttccatca cgatgttccg atcaagatgt acctgtacga cccctggcc 780  
 ttgcggccact acaaggaccc atgggtgttgc gccggccact ctaccatcgcc ccacctggcc 840  
 tcacccctt ctacccgaaa ggacccgttcc ttccctgttcc cttaacacgg ccacgttacc 900  
 tcctccaaact ctggacaccc ggccttccttgc gccgggttcc accttccatcc gggggccatc 960  
 ctgtgttcaacg agcagaatgtt catcgacttcc ggcataacgc tggcccttcc ctacttcgccc 1020  
 acctacaacc acaccggccctt tggcatcgcc cccggggctt tgccttgggtt ggtactctgtg 1080  
 accggccgtt gggcccttcc ccccttccctt cagtcgttcc tctacttcc ttccggcttc 1140  
 tgggtgacccg ccccttacta catccgttcc cccggaccc tggagtttcc tttactacgccc 1200  
 taccgatgtt cccggccactt taatggccgtt gacccgttcc tttctggccatc 1260  
 gaggacgcctt gtcgagccgg ctctgccttcc tttcttccatca acgacgttacgc ccaggccaa 1320  
 gggtggccgg ccccttgcgttcc catggatgtt ttctgggttcc cccggccctt gaagttacgccc 1380  
 tacctgtatcc tggccgagga atctgtacgtt caggtgttcc ccaacggccgg caacaaggttc 1440  
 gtgttcaaca ccggggccca ccccttcttcc atccgttcc ttcttcgttcc agggccac 1500  
 ctggcccatgtt acggatgttcc ataggcttcc tacc 1534

&lt;210&gt; 10

&lt;211&gt; 500

&lt;212&gt; DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 解脂西洋著霉 GAP 启动子的示例性核苷酸序列

&lt;400&gt; 10

cggcgactg cgtccgaacc agcicccagca gcgttttttc cggggccattt agccgactgc 60  
 gaccccgcca acgtgttttgc gcccacgcac tcatgttcatg ttgggtttgg gaggccactt 120  
 tttaagttagt acaaggccacc tagctcgatg caagggtgtcc gaaccaaaga aacggctgtca 180  
 gtgggtcaaaa cggggcgaa acggccggaa aaaggccacgg gggcacgaat tgaggccacgc 240  
 cctcgatattt gagacggatc acggcccat tccggccgc aatggcttgc caacggccgg 300  
 tcctttgcac cacatcgatcc taccacccat ctttttttttgc ttttttttttgc 360  
 tataccgttccat gtaggtttgg gccggccatc tccgttgc ttttttttttgc 420  
 ggtctgttccat gccggccatc ttgtatccat ttttttttttgc ttttttttttgc 480  
 gaattaaaca cccatcgatcc 500

<210> 11  
<211> 6549  
<212> DNA  
<213> 人工序列

<220>  
<223> 示例性基序

<400> 11  
cttatctggg cgactgttgcgat atatgtttatg gtaatagttt cggatgtttagttt gaactttatag 60  
atagactgga ctatacgctt atcggccaa attagaaga acgtcaatgg ctctctggc 120  
gaaatttcgtt taacctcgta tagcaggagt tatccgaacg gataattacc ctgttatccc 180  
tagcttatacg aiaccgcttc atgcgttgcgat gtctcacaag tgcggcgttgcag tccggcccccc 240  
acttgtttctt ctttgtgtgt agtgtacgtt cattatcgat acggttgttc cggcccaactt 300  
cgatccggca tgctgagggtg tctcacaagttt gcccgttgcgtt cccggcccca ctgttgttctc 360  
tttgtgtgtgtt gttgtacgtt attatcgatg cctgttgttcc cggccacccatc gatccggcat 420  
cggtgggtt cttcaacaaatgg cctgtgttgcgatc cggcccccacccatc tttgttgttag 480  
tgtacgttaca ttatcgagatc cgttgttccccc gcccacccatc atccggcatg ctgttgtgttc 540  
ctacaaggatc cgtgtccatc cggcccccactt tgcttcttctt tttgttgttagt gtacgtatcat 600  
tatcgagacc gtttgtcccg cccacccatcga tccggcatgc acgtatcaccg gcaaaaatgt 660  
cgatataata caagagcgatc tgccagccac agattttatc tccacacacca acatcacaca 720  
tacaaccaca cacatccaaatgg tggaaaccccg aaactaaggatc tccatgaagc ttccacat 780  
ctcttttacca gcttgcgttca ccctggccgc taccatcgatc ggctctccca accccaccccg 840  
agccggccgtt gtgttgcgttcc ccttccatcgatc ctcttggaaatc gcttacccacc acttgcctt 900  
cccccaatcgatc gacccatcgatc cctgtgttccatc cttgtgttgcac gacggatcgaa acgggttgggg 960  
ctcttttgcgtt atcgacggcc tggacacccgcg atccatcgatc gggacggccg acatcgatgaa 1020  
caccatctgtt cttgtgttgcgatc cccatcgatc tttccacacc acggccgtgg ccaaccagggtt 1080  
catctcttgtt tttccatcgatc acatcgatc cctggccgcg cttgtgttgcgatc tttccacacc 1140  
gttgcgttgcgtt cccttcttctt ctctggccac caaccatcgatc ctggtgaactt ctctgtgttgcg 1200  
acaggcccaatccatcgatc acggccgttac ggtgttttaccaccctt ctggcgttgc 1260  
cgaccccaatccatcgatc acggccgttac ggtgttttaccaccctt ctggcgttgc 1320  
ggccgagatc ggctcttgcgtt tgctggatgtt gacccgttgcgtt ctgtaccatgatc cccggcaaccc 1380  
ccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc 1440  
cgaggccgtt ccggatcgatc tggacacccatc cttgtgttgcac tttccacacc cccatcgatc 1500  
ctcttccggc ttttgtgttgcgtt ccgttgcgtt ctcttccatcgatc gtttccatcgatc tttccacacc 1560  
cttgcgttgcgtt ccctgtgttgcgtt ctcttccatcgatc tttccacacc cccatcgatc 1620  
taccatcgatc caccatcgatc cccatcgatc tttccacacc cccatcgatc tttccacacc 1680  
ttacaacggc cttgtgttgcgtt ctccatcgatc tggacacccatc gtttccatcgatc tttccacacc 1740  
cttcatctgtt gggccatccatcgatc tttccatcgatc gtttccatcgatc tttccacacc 1800  
ggcccttccatcgatc tttccatcgatc gtttccatcgatc tttccatcgatc tttccacacc 1860  
cgccctgggtt gactctgttgcgtt ctccatcgatc tttccatcgatc tttccatcgatc tttccacacc 1920  
ctactcttctt gccggcttgcgtt ggggttgcgtt cccttccatcgatc tttccatcgatc tttccacacc 1980  
ggaggttgcgtt tactacccatcgatc acccgatgttgcgtt ccgttgcgtt aatggccagg accttggcc 2040  
ggaggcccttc tttccatcgatc aggacccatcgatc tttccatcgatc tttccatcgatc tttccacacc 2100  
cgacgttgcgtt cccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccacacc 2160  
cgaggcccttc aatggccaggatc acctgttgcgtt ccgttgcgtt tttccatcgatc tttccacacc 2220  
aaacggccgcg aacaatgttgcgtt tttccatcgatc tttccatcgatc tttccatcgatc tttccacacc 2280  
tttccatcgatc gggccatccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccacacc 2340  
atctaagctt tttatctgttgcgtt tttccatcgatc tttccatcgatc tttccatcgatc tttccacacc 2400  
tatcgtttat tttccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccacacc 2460  
tgtaacactt tttccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccacacc 2520  
tattaactctt ggttgcgtt tttccatcgatc tttccatcgatc tttccatcgatc tttccacacc 2580  
catcatctgtt tttccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccacacc 2640  
aacaactttt attcgttgcgtt tttccatcgatc tttccatcgatc tttccatcgatc tttccacacc 2700  
atcgtgttgcgtt tttccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccacacc 2760  
gttggggaaatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccacacc 2820  
gggttacccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccacacc 2880  
ttagatccccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccacacc 2940  
acccatccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccacacc 3000  
gggttccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccacacc 3060  
ggcgatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccacacc 3120  
tttccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccacacc 3180  
cgccacccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccacacc 3240  
atccgttgcgtt tttccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccacacc 3300  
cttgcgttgcgtt tttccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccacacc 3360  
ctgtatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccacacc 3420  
gttgggttgcgtt tttccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccacacc 3480  
gttgggttgcgtt tttccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccacacc 3540  
gcccataatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccatcgatc tttccacacc 3600

[0007]

<210> 12  
<211> 6485  
<212> DNA  
<213> 人工序列

<220>  
<223> 示例性基序

<400> 12

```

tagcaggagt tatccgaagc gataatttacc ctgttatccc tagcttatacg atacggcgga 60
ctgcgtccga accagctcca gcagegtttt ttccgggcca tttagccgac tgccgccccg 120
ccaacgtgtc ttggcccaag cactcatgtc atgttgtgt tgggaggccaa cttttaagt 180
agcacaaggc accttagtcg cagcaagggt tccgaaccaa agaaggcgct tcgatgtgtc 240
aaacggggcc aaaaaggccg gaaaaagccca cggggccacg aatttgaggca cgccttcgaa 300
tttgagacga gtcacggccc catttcggcc cgaatggct cggcaacgcg cggcttttgc 360
caccacatca ggtttacccca agccaaacctt ttgtttaaa aagcttaaca tattataccg 420

```

[0008]

aacgttaggtt tggcgccgt tgctccgtct gtccaaaggca acattttat aagggtctgc 480  
 atcgccggct caattgaatc ttttttcctc ttctttctc tatattcatt cttaattaa 540  
 acacacatca acaggatcca tgaagcttc caccatcctc ttacacgctt gcgttacccct 600  
 ggcgcgttacc aagcgaggct ctcccaaccc cacccgagcc gcccgtgtga agggccctt 660  
 ccagacctct tggaaacgcctt accaccactt cgcctcccc caccgacgacc tgacccctgt 720  
 gtctaactcg ttgcacgacg aegaaacgg ctggggctct tctgttccatcg acgacgttgg 780  
 caccggcatc ctgtggcg aegacgacat cgttacacc acctgtcgt acgtggccca 840  
 gatcaacttc accaccacgg cctggccaa cacgggatc tctgttccatcg agaccaacat 900  
 ccgataccctg ggccgttgc tgcgttccatcg cgaccgttgc cgaggccctt tctttctct 960  
 ggcaccaac cagacccctgg tgaactctct gctgttccatcg gcccacccgg tggccaaacgg 1020  
 cctgaagggtt gtttttacca cccctctgg cgttccatcg cccacccgtt tcttcaaccc 1080  
 caccgttgcga cgatctggcg ctcttctaa caacgttgcgc gagatggct ctctgttgc 1140  
 ggagtggacc cgactgttgc acctgttgcg caaccccccgg tacggccgg tggccaggaa 1200  
 gggcgatctt tacccatgtt accccaaggg ctctcccgag gcttggccgg gactgttgg 1260  
 caccctgtt tcttccatcg acggacccctt ccaggactct tccggctctt ggttggccct 1320  
 gatggactt ttctacgtt acctgttgcgat gatgttccatcg ttttttttttggcccttgc 1380  
 ccatacaag gaccgttggg tgcgttgcgc egacttccatcg acggccacc tggcccttca 1440  
 cccctctacc cggaaaggacc tgcgttccatcg ccgttccatcg aacggccatgtt 1500  
 caactctggt caccgttgcgtt ctccgttgcgc tggcaacttc atccgttggc gcaactgtt 1560  
 gaacgagcag aagtacatcg acttccgttgcgtt caacgttgcgc tcttccatcg tcggccaccta 1620  
 caaccagacc gcctctggca tggcccccgg gggcttgcgc tgggtggact ctgttgcgg 1680  
 cgttgcggc tctcccccctt ctcttcgttgc accttccatcg tcttccatcg gettctgggt 1740  
 gaccggccccc tactacatcc tgcgttgcgc gacccttggag tcttccatcg acggccatcg 1800  
 agtggccggc gacttcaatgtt ggcaggactt ggccttgggg gcttccatcg ccacgttgg 1860  
 cccgttgcgtt ggcggctctt ctacttccatcg ttttttttttgcgttgcgc 1920  
 cggggccctt gacgttccatcg agtccatcg ttttttttttgcgttgcgc gccccttccatcg 1980  
 gatcttgcgtt gacgttccatcg acgttccatcg ttttttttttgcgttgcgc agtgggtttt 2040  
 caacaccggc gcccacccctt tcttccatcg atcttccatcg ttttttttttgcgttgcgc 2100  
 ccatgttgcgtt ctgttccatcg ttttttttttgcgttgcgc ttttttttttgcgttgcgc 2160  
 caacttccatcg ctcaactatcc tactttatataatgttacatc gtttttttttgcgttgcgc 2220  
 gtatatgttgcgtt tcttccatcg caaatgttgcgtt ttttttttttgcgttgcgc 2280  
 gtcatccgtt acggtaacttc taatccatcg acacccatccatcg aacttccatcg aactgttact 2340  
 cttttgcgtt ctttccatcg actcaatccatcg ttttttttttgcgttgcgc ttttttttttgcgttgcgc 2400  
 tcttccatcg ttttccatcg actcaatccatcg ttttttttttgcgttgcgc ttttttttttgcgttgcgc 2460  
 aatttttttaa taaaatgttgcgtt ttttttttttgcgttgcgc ttttttttttgcgttgcgc 2520  
 tagggtaggtt tttttccatcg tttttccatcg ttttttttttgcgttgcgc ttttttttttgcgttgcgc 2580  
 acacaccacc tttttccatcg acttccatcg ttttttttttgcgttgcgc ttttttttttgcgttgcgc 2640  
 agtggccggc cccgggggggg gggaaagactt ccacgttgcgtt ttttttttttgcgttgcgc 2700  
 tccagccggc tttttccatcg acgttccatcg ttttttttttgcgttgcgc ttttttttttgcgttgcgc 2760  
 aatcgaaatcc tttttccatcg tttttccatcg ttttttttttgcgttgcgc ttttttttttgcgttgcgc 2820  
 tcccgccatcg aagaacttccatcg tttttccatcg ttttttttttgcgttgcgc ttttttttttgcgttgcgc 2880  
 ggegataccg taaagcacgtt ggaagccggc acggccatccatcg ttttttttttgcgttgcgc 2940  
 atcacgggtt gccaacccgtt tttttccatcg tttttccatcg ttttttttttgcgttgcgc 3000  
 gatgttccatcg gaaaaggccg catgttccatcg tttttccatcg ttttttttttgcgttgcgc 3060  
 ggttccatcg agatccatcg tttttccatcg tttttccatcg ttttttttttgcgttgcgc 3120  
 tttttccatcg tttttccatcg tttttccatcg tttttccatcg ttttttttttgcgttgcgc 3180  
 ccgatgttccatcg tttttccatcg tttttccatcg tttttccatcg ttttttttttgcgttgcgc 3240  
 atcaagccgtt tttttccatcg tttttccatcg tttttccatcg ttttttttttgcgttgcgc 3300  
 aaggttgatgtt gacaggatgtt tttttccatcg tttttccatcg ttttttttttgcgttgcgc 3360  
 cgttccatcg acaacgttccatcg tttttccatcg tttttccatcg ttttttttttgcgttgcgc 3420  
 tagccggccgtt tttttccatcg tttttccatcg tttttccatcg ttttttttttgcgttgcgc 3480  
 aagaaacccggg ccccccgttccatcg tttttccatcg tttttccatcg ttttttttttgcgttgcgc 3540  
 ctgttccatcg ctttccatcg tttttccatcg tttttccatcg ttttttttttgcgttgcgc 3600  
 caatccatcg tttttccatcg tttttccatcg tttttccatcg ttttttttttgcgttgcgc 3660  
 tcccttgcgtt tttttccatcg tttttccatcg tttttccatcg ttttttttttgcgttgcgc 3720  
 cccacccatcg tttttccatcg tttttccatcg tttttccatcg ttttttttttgcgttgcgc 3780  
 cggccatccatcg tttttccatcg tttttccatcg tttttccatcg ttttttttttgcgttgcgc 3840  
 tttttccatcg tttttccatcg tttttccatcg tttttccatcg ttttttttttgcgttgcgc 3900  
 ctgttccatcg tttttccatcg tttttccatcg tttttccatcg ttttttttttgcgttgcgc 3960  
 gggcggccgtt gaaatccatcg tttttccatcg tttttccatcg ttttttttttgcgttgcgc 4020  
 tttttccatcg tttttccatcg tttttccatcg tttttccatcg ttttttttttgcgttgcgc 4080  
 ctgttccatcg tttttccatcg tttttccatcg tttttccatcg ttttttttttgcgttgcgc 4140  
 gataacgttccatcg tttttccatcg tttttccatcg tttttccatcg ttttttttttgcgttgcgc 4200  
 gggccatccatcg tttttccatcg tttttccatcg tttttccatcg ttttttttttgcgttgcgc 4260  
 cgttccatcg tttttccatcg tttttccatcg tttttccatcg ttttttttttgcgttgcgc 4320  
 gggccatccatcg tttttccatcg tttttccatcg tttttccatcg ttttttttttgcgttgcgc 4380  
 tttttccatcg tttttccatcg tttttccatcg tttttccatcg ttttttttttgcgttgcgc 4440  
 gtgttccatcg tttttccatcg tttttccatcg tttttccatcg ttttttttttgcgttgcgc 4500  
 tttttccatcg tttttccatcg tttttccatcg tttttccatcg ttttttttttgcgttgcgc 4560

[0009]

<210> 13  
<211> 6996  
<212> DNA  
<213> 人工序列

<220>  
<223> 示例性基序

<400> 13

[0010]

ttaccacccc ctctggcggtg cccgacccca ccgtgttctt caacccacc gtgcgacgat 1500  
 ctggcgctc ttctacaac gttggccgaga tcggctctct ggtgtcgag tggaccgac 1560  
 tgcgtgaccc gaccggcaac ccccagtacg cccagctggc ccagaaggc gaggcttacc 1620  
 tgcgtgaccc caagggtctc cccgaggctc ggcggact gatcgccacc ttctgtcta 1680  
 cccttaacgg cacccttcag gactcttcg gtcgttgcc tggcttgatg gactcttct 1740  
 acggttaccc gatcaagatg tacctgtacg accccggtgc ctgcgttccac tacaaggacc 1800  
 gatgggtgt gcggccggac tcttacatc cccactggc ctctcatacc tctaccggaa 1860  
 aggacctgac cttecttgcc tcttacaac gccagctac ctctccaaac tctggacacc 1920  
 tggttccctt cgccgggtgc aacttcatcc tggccgtat cctgtgtaa gaggcagaagt 1980  
 acatcgactt cggcatcaag ctggcttctt cctacttcg cacttacaac cagacccct 2040  
 ctggcattcg ccccgaggcc ttgcgttccg tggactctgt gaccggcgtt ggccgtctc 2100  
 ccccttcc tcagtcgtgc ttctacttctt ctgcgggtt ctgggtgacc gccccctact 2160  
 acatcctcgcc acccgagacc ctggagtctc tgcgtactacgc ctaccggatg accggcgtact 2220  
 ctaagtggca ggacctggcc tggggggctc tctctgtccat cgaggacgccc tgcgtggcc 2280  
 gctctgccta ctcttctatc aacgacgtga cccaggccaa cggtgtggaa gcctctgacg 2340  
 acatggagtc ttctgggtc gcccggggcc tgaagtgacgc ctacgtatc ttgcggagg 2400  
 aatcgactt gccaacggcc gcaacaactt cgtgttcaac accggggcc 2460  
 accccctctc tatccgtatc tcttctcgac gaggcccat cttggccat gacgactgtt 2520  
 aatagcttag ggtgtctgt gatctataac tatttataac tcttataac ttctacatca 2580  
 actatctact ttaataatg aatatcgttt attctctatg attactgtat atgcgttctt 2640  
 ctaagacaaa tgaatttca tggtaacac tcgctctggg gagttgtca tccgacagg 2700  
 taactctaat ctcccaacac ctattaaact ctgcgttaact gtaacttctc ttggcacgtc 2760  
 galcttactc aattttcttgc tctatcatc gctggatgtt tgcgtatgt ctggcttcaa 2820  
 tacattttt gtttattgtcc caaacaactt tcatgtcactg taagtgatg gttttataac 2880  
 agegttcgtc aattgtcgcc ctatcgatgtt cccggatcc taccgtttagg gtagtgtgtc 2940  
 tcacactacc gagggttacta gagggtggaa aegcatactg ctcggacac accacctgtt 3000  
 cttagactg cagagaaat cggcgttacc tcttcacaaa gccctcgtt cggccggcc 3060  
 gggggggcgaa agaactccag catgagatcc cccgcgttggg gatcatcca gccggcggtcc 3120  
 cggaaaacgaa ttccgaagcc caacccatca tagaaggcc cgggttggaaatcgtt 3180  
 gatggcgatg tggcggtc ttggcggtt atttcgaaacc ccagagtccc gctcagaaga 3240  
 actcgatcaag aaggcgatag aaggcgatgc gtcgttgcgatc gggggggccg ataccgtaaa 3300  
 gcaagggggaa ggggttgcgc catttcggcc caagcttcc agcaatataca cgggttgc 3360  
 acgttatgtc ctgtatcgcc tccggccacac ccaggccggcc acagtgtatg aatccagaaaa 3420  
 agccggccatt ttccacatc atattcgccca agcaggccatc gccatgggtt acgacggat 3480  
 cctcgccgtc gggcatggc gccttgagcc tggcgaaacag ttggcggttgc gcgagccct 3540  
 gatgtcgatca tccatcgatca tccatcgatca caagggccg tttccatccgatc gtacgtgtc 3600  
 gtcgtatcgatc atgtttcgatc tgggtgtca atggcgatgtt agccgtatca agcgatgtca 3660  
 gcccggccat tgcgtatcgcc atgtgttgcgatc ctttctggc agggatcagg tgagatgaca 3720  
 ggagatctgtt ccccggtact tggcccaata gcggccatc ctttccgtt tcaatgttca 3780  
 cgtcgatcgatc agtgcgtccaa ggaacggcccg tctgtggccatc ccacgtatgc cgcgtgtcc 3840  
 cgtccgtccat ttcatttcgatc gcaatggatcc ggtcggttcc gacaaaataga accggggcc 3900  
 cctcgccgtatc cttccatcgatc acggccgttccatc ctttccgtt gatgtgttgc tttccatccg 3960  
 catagccgaa tagccatcc acccaaggcc cccggaaacc tgcgtgtcaat ccattttgtt 4020  
 caatcatcgatc aaacgatccatc catccatgtt ctgtatcaga tcttgcgttcc ctgcgtccatc 4080  
 agatcccttg cggcaagaaa gccatccatc ttttgcgttca ggggttccca accttaccatc 4140  
 agggggccccc agtggcaat tccgggttgc ttgtgttca taaaaccggcc cagtttgcgtt 4200  
 atcgccatgtt aagccacttgc caagtttgcgtt ttttccatccatc ttttccatccatc 4260  
 tccagatagc ccagttatcc acattcatcc ggggttgcgtt ccgttttgcgtt ggactgggtt 4320  
 tctacgttccatc tccatccatc tagcggccatc tggcccttgc tggctgttcc cggccgttgc 4380  
 cttagttatc cgggtgttca taccgttccatc atgcgttcaagg agaaaatacc gcatcggcc 4440  
 ctcttccatccatc tccatccatc ctgtatcgatc gggccgttccatc ttttccatccatc 4500  
 atcagcttccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 4560  
 gaacatgttca gcaaaaaggcc agcaaaaaggcc cggggatccatc tccatccatc tccatccatc 4620  
 gttttccatccatc aggtttccatc ccccttgcgttca gcatcacaatc aatcgatccatc tccatccatc 4680  
 gttggccatccatc ccccttgcgttca tccatccatc tccatccatc tccatccatc tccatccatc 4740  
 ggcgttccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 4800  
 aagggttccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 4860  
 ctccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 4920  
 taactatgttccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 4980  
 tggtaacagg attacgatcc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 5040  
 gcttaactac ggttccatcc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 5100  
 tacccttccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 5160  
 cggtttttgcgttccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 5220  
 ttgtatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 5280  
 gttcgttccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 5340  
 ttcatatgttccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 5400  
 acgcaaaaacgaa gaccgttccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 5460  
 taccgggttccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 5520  
 aactcgatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 5580

tcgacaaagg cggggccccc ctgcagatata ccctgttata cctacataac ttgcgtatgc 5640  
 atacattata cgaagtattt ctgaattccg cccagagacg cattgacgtt ctttctaatt 5700  
 tggaccgata gccgtatagt ccagtctatac tataagtca actaactcgta aactattacc 5760  
 ataacaatata cttcaactgcc ccagataagg ttccgataaa aagtttcgtca gactaaattt 5820  
 atttcgtct cctcttcacc accaaaatgc cctctacga agtcgagct aacgtccaca 5880  
 agtccgcctt tgccgtcgta gtgcgtcaage tcgtggcagc caagaaaaacc aacctgtgtg 5940  
 ctctctggta gtttacacc accaaggagc tcatttgatgc tgccgtatag gtggacatt 6000  
 atgtgtcat gatcaagacc catatcgaca tcatttgacga cttcacctac gccggcactg 6060  
 tgcctccctt caaggaactt gcctttaaagc acgggtttttt ccgttgcagc gagacaaagt 6120  
 tgcagatata tggcaacact gtcaaggacc agtaccgggt tcacccgaatc gccgagtgg 6180  
 ccgatatac cAACGCCAC ggtgtacccg gaaccggaaat cattgtggc ctgcgagctg 6240  
 gtggcgagga aactgtctt gaacagaaga aggaggacgt ctctgactac gagaactccc 6300  
 agtacaagga gttccctagtc ccccttcacc acggagaagct ggccagaggct ctgcctatgc 6360  
 tggccgagct gtcttgcag ggtcttcgtt ccactggcga gtactccaag cagaccattt 6420  
 agtttgcggc atccgaccc gatttgtgg ttggcttcat tgcccgaaac cgacctaagg 6480  
 gggactctgt ggactggcatttctgaccc cgggggtggg tcttgacgac aaggggagacg 6540  
 ctctcgagaca gcaatcccgaa actgttgggg atgtcatgtc taccggaaacg gataatcataa 6600  
 ttggccggc aggtctgtac ggcggaaacc gagatcttat tgaggaggcc aagcgatacc 6660  
 agaaggctgg ctggggaggct taccagaaga ttaactgtta gaggttagac tatggatatg 6720  
 taatattaact gtgtatataag agagcgtgca agtattggggc gctgttcagc ttgtatgtg 6780  
 gtcagacgac ctgtctgtac ggtatgtat ggtactgcac aacctgtgtt tccgcatgg 6840  
 ctgtccaatg gggcatgtt tttgttttctt cggaaatttgtt ataaacttctgt atagcaggag 6900  
 ttatccgaag cgataattac cctgttatcc ctatgttatac gattccccaca agacgaacaa 6960  
 gtataggcc gagagccgag gacgagggtt agtgc 6996

<210> 14  
 <211> 7195  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 示例性基序

[0011]

<400> 14

taattaccct gttatcccta gtttatcgat agagaccggg ttggccgcgt atttgtgtcc 60  
 caaaaaacag ccccaattgc cccaaattgc cccaaattga cccagtagcg ggcccaaccc 120  
 cggcgagacg ccccttcacc ccacatatac aacctcccc ggttccaca cttggcgat 180  
 agggcgtagg gtactgcgtt ctggatatac cgtttgtca gactttgtac tagtttcttt 240  
 gtctggccat cccggtaacc catggcgac gcaaaataga ctactgaaaa ttttttgc 300  
 ttgtgggttgg gacttttagcc aagggtataa aagaccaccc tccccaattt acctttccctc 360  
 ttcttttctc tctctccctt tcaactcaca cccgaaggat ccatgaagct ttccaccatc 420  
 ctcttcacccg cctggcgatc cctggccgtt accaaggcgat gtcctcccaa ccccacccga 480  
 gcccggcgtt tgaaggccgc ctggccatc ttttggaaacg cttaccacca ctggcccttc 540  
 cccacacgac acctgcaccc cgtgtctaaac tgggtcgacg acggacgaaa cggctggggc 600  
 tctctgcca tggacggcc atccgtatgg gggacggccgatcgtgaac 660  
 accatccctgc agtacgtgcc ccagatcaac ttccacccca cccggcgatc caaccaggcc 720  
 atctctgtgt tggagacccaa catccgatac ctggccggcc tgcgtctgc ctacgaccc 780  
 ctggcgaggcc ccttcttc tctggccacc aaccagaccc tggtaactc tctgtgtcg 840  
 caggcccaga ccctggccaa cggcctgaaat gtggctttt ccaccccttc tggcgatggcc 900  
 gacccacccg ttttctcaa ccccacccgg cggacgtatcg gggcccttc taacaacgtg 960  
 gcccggatcg gtcctctgtt gttggatggg accccgatgtt ctgcacgtc cggcaacccc 1020  
 cgtacgcccc agctggccca gaaaggccgc ttttacatgc tgaatcccaaa gggctctccc 1080  
 gaggtcttgc cggactgtat cggcacttc ttttgcaccc ttttgcaccc cttccaggac 1140  
 tttttccgtt ctggatggcc ttttctgtatc agtactgtat caagatgtac 1200  
 ctgtacgacc cctggccctt cggccactac aaggacccat gggatggcc cggcacttc 1260  
 accatcgcccc acctggccctc ttttgcaccc accccgaaagg acctgcaccc ttttgcaccc 1320  
 tacaacggcc agtctaccc ttttgcaccc ggacccatgg ctttgcaccc cggatggcc 1380  
 ttcatccctgg gggccatcc ttttgcaccc ttttgcaccc ttttgcaccc ctttgcaccc 1440  
 gccccttcctt acttgcaccc ttttgcaccc ttttgcaccc ttttgcaccc ctttgcaccc 1500  
 gcccggatcg acttgcaccc ttttgcaccc ttttgcaccc ttttgcaccc ctttgcaccc 1560  
 tactttctgtt cggccatcc ttttgcaccc ttttgcaccc ttttgcaccc ctttgcaccc 1620  
 gagtcttgc ttttgcaccc ttttgcaccc ttttgcaccc ttttgcaccc ctttgcaccc 1680  
 gggcccttctt ctggccatcc ttttgcaccc ttttgcaccc ttttgcaccc ttttgcaccc 1740  
 gacgtgtatcc ttttgcaccc ttttgcaccc ttttgcaccc ttttgcaccc ttttgcaccc 1800  
 gggcccttctt ctggccatcc ttttgcaccc ttttgcaccc ttttgcaccc ttttgcaccc 1860  
 aacggccggca acaaggccgtt ttttgcaccc ttttgcaccc ttttgcaccc ttttgcaccc 1920  
 ttttgcaccc ttttgcaccc ttttgcaccc ttttgcaccc ttttgcaccc ttttgcaccc 1980

[0012]

tctaagctat ttatcactt ttacaacttc tacctcaact atctacttta ataaatgaat 2040  
 atcgtttatt ctctatgatt actgtataatg cgttcctctaa agacaaatcg aattccatgt 2100  
 gtaacactcg ctctggagag tttagtcatcc gacagggtaa ctctaatctc ccaaacacctt 2160  
 attaactctg cgtaactgtt actcttcttg ccacgtcgat cttaactcaat tttccctgctc 2220  
 atcatctgt ggatttgtt ctatcgctcg gctctaaatc atttattgtt tattgcuccaa 2280  
 acaactttca ttgcacgtaa gtgaattttt ttataacage gttcgccaa tgctcgccaa 2340  
 tgcgtccg gctgtctac cgttaggtt gtgtgtctca cactaccgg gttactagag 2400  
 ttggaaagc gatactgcct cggcacacacc acctggctt acgactcgag agagaatcg 2460  
 cgttacctcc tcacaaagcc ctcagtccgg ccccccgggg tggcgaaga actccagcat 2520  
 gagatccccg cgctggagga tcatccagcc ggcgtccccc aaaacgattc cgaagccaa 2580  
 ccttcatag aaggcgggg tggaatcgaa atctcggtat ggcaggttgg gcgtcgctt 2640  
 gtccgttatt tegaacccca gagtcccgct cagaagaact cgtcaagaag gcatagaag 2700  
 gcatcgccgt gccaatcggtt agcggcgata ccgttaaagca cgaggaagcg gtcagccat 2760  
 tcgcgccaa gctcttcage aatatcaccg gttagccaacg ctatgtctt atagcggtcc 2820  
 gecacacccca gccggccaca gtcatgtat ccagaaaagc ggcattttt caccatgata 2880  
 ttccggcaagc aggcatcgcc atgggtcaag acggatctt cgcgtcggtt catggcgcc 2940  
 tttagccctgg cgaacagtgc ggctggcgcc agccccctgtat gctcttcgtc cagatcatcc 3000  
 ttagtcgacaa gaccggccatc catccggat cgtgtcgat cgtgcgtat tttcgcttgg 3060  
 tggcgaatg ggcaggtagc cggatcaagc gtatgcagcc ggcgcatttc atcagecatg 3120  
 atggatactt tctccggcagg agcaagggtt gatgacaggg gatccgtcccc cggcacttcg 3180  
 cccaaatagca gccagttcc tccgcttca gtgacaacgt cgagcacagc tgccaaagga 3240  
 acgcccgtcg tggccagcc cgtatccgcg cgtccctgtt cctgcgttcc attcaggcga 3300  
 cccggacagggtt cgggtcttgc aaaaagaacc gggcccccgtt ggcgttccatc cggaaacacg 3360  
 gggccatcg agccatcgat tgctgtgtt gcccgttccatc agccgaatag ccttcacc 3420  
 caaggccggc gagaacatcg tgcgttccatc tcttgcgttccatc cgtatcgat 3480  
 cctgtcttcc tgcgtatctt tgatccctt cccatcaga tccctggcg caagaagcc 3540  
 atccagtttta ctttgcaggg ctttccaaatc ttaccagggtt ggcgcattcc tggcaattcc 3600  
 gggtcgcttgc ctgtccataa aaccggccatc tcttagtata gccatgttccatc 3660  
 gtcacctgttgc ttctcttgc gcttgcgtt tccctgtccatc agatagccca gtagtgcata 3720  
 ttcatccggg gtcagccatcg ttctgcggg ctggcttctt acgtgttccg ctcccttttag 3780  
 cagcccttgc gcccgttgc ttttgcggg cgttgcgttccatc gttatgcgg ttttgcgg 3840  
 cgcacatcgat cgttgcgttccatc aaatccgc tccggccatc tccctgtccatc 3900  
 actcgctcg ctcggcgtt cggctgcggc gaggccatc agtgcactca aaggccgtaa 3960  
 taggttatac cacaatcgatca ggggataacg cagggaaatcgatca ctttgcggcc 4020  
 aaaaggcccg gaaccgtaaa aaggccgttgc tgctggcggtt tttccatagg ctccggccccc 4080  
 ctgacgagca tcacaaaaatc cgcgttccatc gtcagggatc gcaaaaccccg acaggactat 4140  
 aaagataccca ggcgttccatc ctttgcggat cccctgtccatc tccctgtccatc 4200  
 cgcgttccatc ataccgttccatc gcttgcgttccatc ttttgcggccatc tccatagatc 4260  
 caacgttgcgttccatc ttttgcggccatc ttttgcggccatc tccatagatc 4320  
 aaccccccgtt tcaagccgc tccgtccatc tatccggatc tccatagatc 4380  
 cggtaagaca cgttgcgttccatc ccactggccatc cagccacttgc taacaggattt acgcagatcg 4440  
 ggtatgttgcgttccatc gtttgcgttccatc ttttgcggccatc tccatagatc 4500  
 ggacagtttgcgttccatc ttttgcggccatc ttttgcggccatc tccatagatc 4560  
 gtccttgcgttccatc ttttgcggccatc ttttgcggccatc tccatagatc 4620  
 gattacgcgc agaaaaaaatcgatcaaga agatccgttccatc ttttgcggccatc tccatagatc 4680  
 gatccccccatc gggatttgcgttccatc ttttgcggccatc tccatagatc 4740  
 ctttgcgttccatc ttttgcggccatc ttttgcggccatc tccatagatc 4800  
 aaattttacta cacaccacccatc ttttgcggccatc ttttgcggccatc tccatagatc 4860  
 aaagccctca ttttgcggccatc ttttgcggccatc tccatagatc 4920  
 ggaggatcat ctttgcggccatc ttttgcggccatc tccatagatc 4980  
 cggcggttgcgttccatc ttttgcggccatc ttttgcggccatc tccatagatc 5040  
 accccagatc cccgttccatc ttttgcggccatc ttttgcggccatc tccatagatc 5100  
 atccggggatc gggataccgttccatc ttttgcggccatc ttttgcggccatc tccatagatc 5160  
 ttttgcggccatc ttttgcggccatc ttttgcggccatc tccatagatc 5220  
 gccacatcgat ttttgcggccatc ttttgcggccatc tccatagatc 5280  
 atcccatgttccatc ttttgcggccatc ttttgcggccatc tccatagatc 5340  
 ctttgcgttccatc ttttgcggccatc ttttgcggccatc tccatagatc 5400  
 ggcgttccatc ttttgcggccatc ttttgcggccatc tccatagatc 5460  
 ggttgcgttccatc ttttgcggccatc ttttgcggccatc tccatagatc 5520  
 ggcaggatcgat ttttgcggccatc ttttgcggccatc tccatagatc 5580  
 gtttgcgttccatc ttttgcggccatc ttttgcggccatc tccatagatc 5640  
 ctttgcgttccatc ttttgcggccatc ttttgcggccatc tccatagatc 5700  
 ctgttgcgttccatc ttttgcggccatc ttttgcggccatc tccatagatc 5760  
 ggcgttccatc ttttgcggccatc ttttgcggccatc tccatagatc 5820  
 acctcggttccatc ttttgcggccatc ttttgcggccatc tccatagatc 5880  
 agatccgttccatc ttttgcggccatc ttttgcggccatc tccatagatc 5940  
 gcaaggccatc ccaacccatc ttttgcggccatc ttttgcggccatc tccatagatc 6000  
 ccataaaaatcgatca gtttgcggccatc ttttgcggccatc tccatagatc 6060  
 ctttgcgttccatc ttttgcggccatc ttttgcggccatc tccatagatc 6120

gcaccgtttc tgccggactgg ctttctacgt gttccgcgttc ctttagcgcgc ccttgcgc 6180  
 tgagtgcgtt cgccgcgcgtg aagcttagett atgcgggtgtg aaataccgca cagatgcgta 6240  
 aggagaaaaat accgcacatcg ggcgccttcc gcttccctgc tcactgactc gctgcgcctcg 6300  
 gtcgttgcgc tgccggcgcg ggtatcactgt cactcaagg cggtaatacg gitatccaca 6360  
 gaatcagggg ataacgcagg aaagaacatg tgagcaaaag gcccggaaac gcccaggaaac 6420  
 cgtaaaaagg ccgcgttttgc ggcgttttgc cataggctcc gccccccctga cgacatcac 6480  
 aaaaatgcac gtciaagtc gagggtggcga aaccggcagc gactataaag ataccaggcgc 6540  
 ttcccccttg gaagctccct cgtgcgcctc cctgttccga ccttcgcct taccggatac 6600  
 ctgtccgcct ttccccccttc gggaaagcgtg ggcgcgttctc atagctcagc ctgttaggtat 6660  
 ctcagttcggt tgtaggtcgt tcgcgttccaaag ctggcgctgtg tgacacgaaacc ccccggttcag 6720  
 cccgaccgcgt ggcgccttatac cggtaaataat cgtcttgcgtt ccaaccgggt aagacacgcac 6780  
 ttatcgccac tggcagcgcg cacttgttac acggatttagca gagcggaggta tgtaggcgtt 6840  
 gctacagagt tcttgaagtg gtggcctaactac gtcgttaca ctagaaggac agtattttgtt 6900  
 atctgcgcctc tgcgttgcgc agttacacttc ggaaaaagag ttggtagtc tttatccggc 6960  
 aaacaaaccca cccgtggtagt cggcggtttt ttgtttgcga gcacggatc acgcggagaa 7020  
 aaaaggatc tcaagaagat ccttttgatct ttcttctactg aacgggtatc cccacccggaa 7080  
 ttgcggcgc tgccggaaac cgcgttccagg tggaaacaggaa caccctccctt gcacatcttg 7140  
 gtatatacgtt ataggctgtat gtattcatag tggggttttt cttatataat ttact 7195

<210> 15  
<211> 1518  
<212> DNA  
<213> 人工序列

<220>  
<223> 以合成方法生成的寡核苷酸

<400> 15

<210> 16  
<211> 498  
<212> DNA  
<213> 人工序列

<220>  
<223> 以合成方法生成的寡核苷酸

<400> 16

```
gtcccccc gactgatctg tgactctcgaa gtgtggAAC gataccTGTCTC 60  
gaAGCCGAGA acATCACCAc CGGCTGTC GAGCACTGTT CTCGAACGAA 120  
gtGCCCGACa ccaaggTGAa CTTC TACGCC TGGAAAGCGAA TGGAGGTGGG 180
```

gtggagggtgt ggcagggact ggctctgcgt tctgaggccg tgctgcgagg acaggcttc 240 ctggtaact ctctcagcc ctgggagccc ctgcagctgc acgtggacaa ggcctgtct 300 ggccitgcgtat ctctgaccac cctgctgca gcccctcggt ctcagaagga agccatctct 360 ccccccgacg ccgcctctgc tgccccctcg cgaaccatca cggccgacac ctccgaaag 420 ctgitccgag tgtactctaa ctccctgca ggcgaagctga agctgtacac cggcgaggct 480 tgtcgaacctc gcgaccga 498	
<210> 17	
<211> 1194	
<212> DNA	
<213> 人工序列	
<220>	
<223> 以合成方法生成的寡核苷酸	
<400> 17	
ctggacaacg gcctggcccg aaccccccacc atgggctggc tgcactggga gcgattcatg 60 tgtaacctgg actgtcagga agagcccgac tcttgcatact ctgagaagct gttcatggaa 120 atggccgagc ttagtgggtgc tgagggtcg aaggacccg gctacgagta cctgtgtatc 180 gacgactgtt ggtatggccccc ccagcgagac tctgaggggc gactccaggc cgaccccccag 240 cgattcccccc acggcatccg acagctcggc aactacgtgc acitctaaggg cctgaagctg 300 ggcatctacg ccgacgtggg caacaagacc tggccggct tcccccgtc ttccggctac 360 tacgacatcg acgcccagac ctgcggcggc tggggcggtt acctgtctaa gttcgcacggc 420 tgttactgtg actctctcgaa gacccatggcc gacggataca agcacatgtc tctggccctg 480 aacccgaaccc gccgatctat cgtgtactt tggatgtgc cctgtatcat gtggcccttc 540 cagaageccca actacaccga gatccgacag tactgttaacc actggcgaaa cttccggc 600 atcgacgact cgtggaaatgc tatcaagtc attctggat ggcacatctt caaccaggag 660 cgaatcgtcg acgtcgccgg accccgggaa tggaaacgacc cggacatgtc ggtgtatggc 720 aacttcggcc tggatgtggaa ccaggaggc acccagatgg cctgtgggc tatcatggct 780 gccccctgt tcatgtctaa cgacactcgca cacatcttc cccagggccaa ggcctgtc 840 caggacaagg acgtgtatgc catcaaccag gacccctgg gcaaggcagg ctaccatgtc 900 cgacaggcgca acaacttcga ggtgtgggag cgacccctgt ctggcctggc ctggccgtg 960 gcacatgtatc accgacagga gatccgggaa ccccgatctt acaccatgtc cgtggccctcc 1020 cttggaaagg gcgtggctg taaccccgcc tggatgttc cccagctctt gcccgtgaag 1080 cggaaatgtgg gattctacga gtggacctct cgaactgcgtat ctcacatcaa cccacccggc 1140 accgtgtcg tccagctcgaa gaacaccatcg cagatgttc tgaaggaccc gctg 1194	[0014]
<210> 18	
<211> 19	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 18	
tcgttatcac gtctcttagc 19	
<210> 19	
<211> 23	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 19	
tctctgtata ctgtatgtatct 23	
<210> 20	
<211> 38	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 20	
ctaggataatgcggatataatg gtgtgacgaa gtatcgag 38	

<210> 21		
<211> 36		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 21		
cattaccctg ttatccctag cgagatcatg gactgg		36
<210> 22		
<211> 34		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 22		
gacgcggccg catgagcttc aacattccca aaac		34
<210> 23		
<211> 49		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 23		
ctaggataa caggtataa caaaatttag aaataaaaat actttacag		49
[0015] <210> 24		
<211> 39		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 24		
cattaccctg ttatccctaa gtaacatgag tgctatgag		39
<210> 25		
<211> 30		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 25		
cgcttaatta aatgcatgga ggtattgctg		30
<210> 26		
<211> 46		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 26		
ggtgcttcga ctatcgttt cggaggatg ggtgattctt tttatg		46
<210> 27		
<211> 46		

<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 27		
cataaaaaga atcacccaat cctccgaaac tgatagtcga agcacc		46
<210> 28		
<211> 32		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 28		
tgagcgcccg cttttctact tcagagctgg ag		32
<210> 29		
<211> 34		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 29		
ggcttaatta attggtagtg atataatgtt acgc		34
<210> 30		
<211> 39		
[0016] <212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 30		
tagggataac aggtaatca cgacacatac tcattcaag		39
<210> 31		
<211> 37		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 31		
attaccctgt tatccctaga aggagatgtt gcgttaag		37
<210> 32		
<211> 24		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 32		
tgataaatag cttagatacc acag		24
<210> 33		
<211> 19		
<212> DNA		
<213> 人工序列		

<220>		
<223> 引物		
<400> 33		
acatacaacc acacacatc		19
<210> 34		
<211> 28		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 34		
ggcgatcca tggtgctgca cccgttcc		28
<210> 35		
<211> 31		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 35		
ggcccttaggc tactcaaact cctcgcaat c		31
<210> 36		
<211> 27		
<212> DNA		
<213> 人工序列		
[0017] <220>		
<223> 引物		
<400> 36		
ggctcgcca gcgcgcac ccttttc		27
<210> 37		
<211> 28		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 37		
ctagatcagc aataaagtgc tgctggc		28
<210> 38		
<211> 30		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 38		
ggatccatgt ctatcaagcga agaaggatcc		30
<210> 39		
<211> 34		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		

<400> 39		
cctaggctag atcagcaata aagtgcgtgct gggc		34
<210> 40		
<211> 41		
<212> DNA		
<213> 人工序列		
<220>		
<223> 以合成方法生成的寡核苷酸		
<400> 40		
cgactatccg gttcggatca ttgggtgatt ctttttatga g		41
<210> 41		
<211> 41		
<212> DNA		
<213> 人工序列		
<220>		
<223> 以合成方法生成的寡核苷酸		
<400> 41		
ctcataaaaaaa gaatcaccca atgatccgaa ccggatagtc g		41
<210> 42		
<211> 41		
<212> DNA		
<213> 人工序列		
<220>		
<223> 以合成方法生成的寡核苷酸		
[0018] <400> 42		
cgactatccg gttcggatca ggtgggtgatt ctttttatga g		41
<210> 43		
<211> 41		
<212> DNA		
<213> 人工序列		
<220>		
<223> 以合成方法生成的寡核苷酸		
<400> 43		
ctcataaaaaaa gaatcacacc ac ctgatccgaa ccggatagtc g		41
<210> 44		
<211> 46		
<212> DNA		
<213> 人工序列		
<220>		
<223> 以合成方法生成的寡核苷酸		
<400> 44		
ggtgcttcga ctatcagttt cggaggattt ggtgatttttt tttatg		46
<210> 45		
<211> 46		
<212> DNA		
<213> 人工序列		
<220>		
<223> 以合成方法生成的寡核苷酸		
<400> 45		
cataaaaaaga atcacccaaat cctccgaaac tgatagtcga agcacc		46

<210> 46  
<211> 38  
<212> DNA  
<213> 人工序列

<220>  
<223> 以合成方法生成的寡核苷酸

<400> 46  
cccgatatacg gatccatgaa gaactctgtc ggtatttc 38

<210> 47  
<211> 70  
<212> DNA  
<213> 人工序列

<220>  
<223> 以合成方法生成的寡核苷酸

<400> 47  
gggaagctta acgcggttcc agcgggtccg gatacggcac cggcgcaccc aacgaccaac 60  
ctgtggtcag 70

<210> 48  
<211> 36  
<212> DNA  
<213> 人工序列

<220>  
<223> 以合成方法生成的寡核苷酸

<400> 48  
cagtgcggcc gcactccctc ttttactca ctattg 36

[0019] <210> 49  
<211> 46  
<212> DNA  
<213> 人工序列

<220>  
<223> 以合成方法生成的寡核苷酸

<400> 49  
cattacccttg ttatccctac gctcagatcc aattgttttgc 46

<210> 50  
<211> 53  
<212> DNA  
<213> 人工序列

<220>  
<223> 以合成方法生成的寡核苷酸

<400> 50  
gtagggataa cagggtaatg ctctcaaggaa cgaccaggat gagactgtta tcg 53

<210> 51  
<211> 46  
<212> DNA  
<213> 人工序列

<220>  
<223> 以合成方法生成的寡核苷酸

<400> 51  
gactttaatt aaaccctatg tggcacctca acccacatct cccgtc 46

<210> 52  
<211> 35

<212> DNA  
<213> 人工序列

<220>  
<223> 以合成方法生成的寡核苷酸

<400> 52  
cagtggatcc atgaactctc ctatttcac taccg 35

<210> 53  
<211> 35  
<212> DNA  
<213> 人工序列

<220>  
<223> 以合成方法生成的寡核苷酸

<400> 53  
gactcctagg aagcttccag gttacaagtt gttac 35

<210> 54  
<211> 73  
<212> DNA  
<213> 人工序列

<220>  
<223> 引物

<400> 54  
ccttaggtcac tccaaatcccc caaacagggtt gctgacgctc gactcatagt gagctagatc 60  
agcaataaaag tcg 73

<210> 55  
<211> 30  
[0020] <212> DNA  
<213> 人工序列

<220>  
<223> 引物

<400> 55  
ggatccatgt ctatcaagcg agaagagtcc 30

<210> 56  
<211> 25  
<212> DNA  
<213> 人工序列

<220>  
<223> 引物

<400> 56  
gaattcatgc ccgttagattc ttctc 25

<210> 57  
<211> 20  
<212> DNA  
<213> 人工序列

<220>  
<223> 引物

<400> 57  
gagtcttcgg gaggattcag 20

<210> 58  
<211> 19  
<212> DNA  
<213> 人工序列

<220>		
<223> 引物		
<400> 58		
ccttggaaagaa tacaaaagtc		19
<210> 59		
<211> 30		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 59		
ccttaggctat tccttggaga atacaaaagtc		30
<210> 60		
<211> 22		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 60		
gglatggctg agcgatgc aa		22
<210> 61		
<211> 22		
<212> DNA		
<213> 人工序列		
[0021] <220>		
<223> 引物		
<400> 61		
ccaccgatcc atacggatca ct		22
<210> 62		
<211> 22		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 62		
cgacctggaa tctgcacttc aa		22
<210> 63		
<211> 22		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 63		
cggtaaccacc taaggcttcc aa		22
<210> 64		
<211> 22		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		

<400> 64		
ccagccaaact gtgttggattc aa		22
<210> 65		
<211> 22		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 65		
ggagctggtg gaataccagt ca		22
<210> 66		
<211> 57		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 66		
gatccatgaa gcttccacc atcctttca cagcctgcgc taccctggcc gcggta	57	
<210> 67		
<211> 65		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
[0022] <400> 67		
gtaccggccg gccgcttctg gagaactgcg gcctcagaag gagtgatggg ggaaggagg	60	
gcccc	65	
<210> 68		
<211> 30		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 68		
ggcagcgcta caaaaacgtgg atctcccaac		30
<210> 69		
<211> 30		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 69		
ggcccttaggt tacaactcggt cgtgagcaag		30
<210> 70		
<211> 17		
<212> DNA		
<213> 巴斯德毕赤酵母		
<400> 70		
atccagcagt gatgacg		17
<210> 71		

<211> 17		
<212> DNA		
<213> 酿酒酵母 ( <i>Saccharomyces cerevisiae</i> )		
<400> 71		
atccagccgt gattacg		17
<210> 72		
<211> 15		
<212> DNA		
<213> 巴斯德毕赤酵母		
<400> 72		
ggtctgcagc accat		15
<210> 73		
<211> 15		
<212> DNA		
<213> 酿酒酵母		
<400> 73		
[0023] tgccgaagc gcagt		15
<210> 74		
<211> 18		
<212> PRT		
<213> 巴斯德毕赤酵母		
<400> 74		
Ala Pro Phe Thr Ala Asn Ala Phe Asp Leu Asn Asp Phe Val Phe Phe		
1                   5                   10                   15		
Gln Glu		
<210> 75		
<211> 18		
<212> PRT		
<213> 酿酒酵母		
<400> 75		
Glu Ala Gln Ser Gly Leu Asn Ser Phe Glu Leu Asn Asp Phe Phe Ile		
1                   5                   10                   15		
Thr Ser		

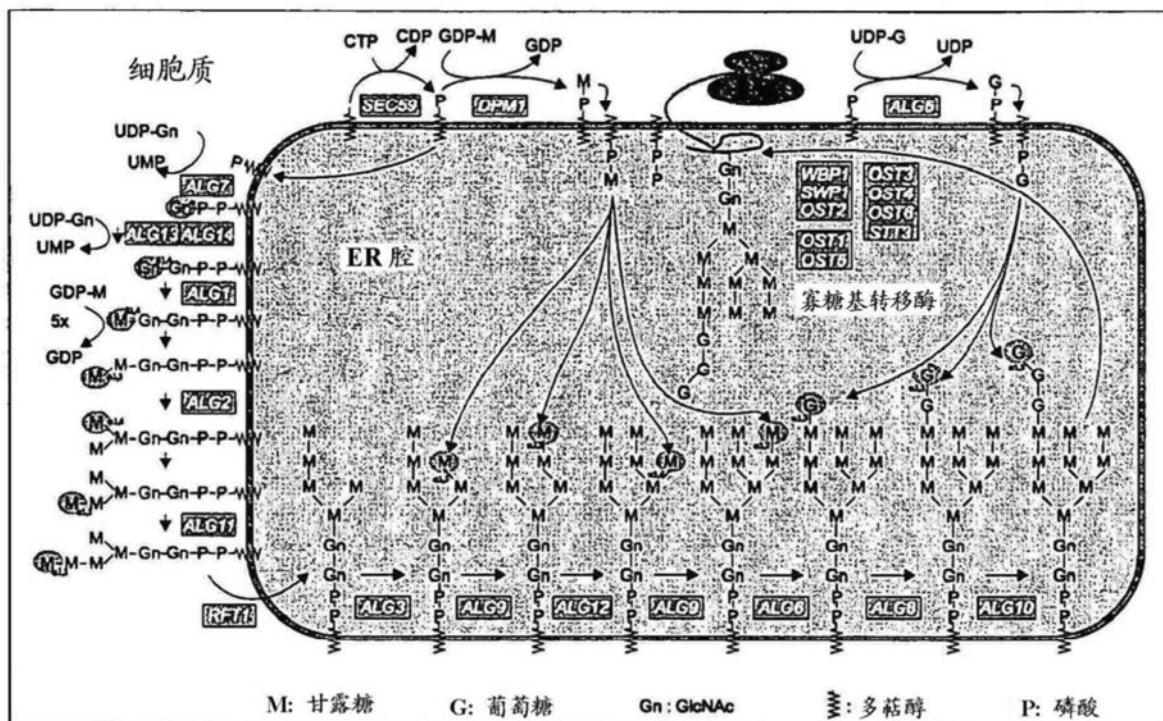


图1

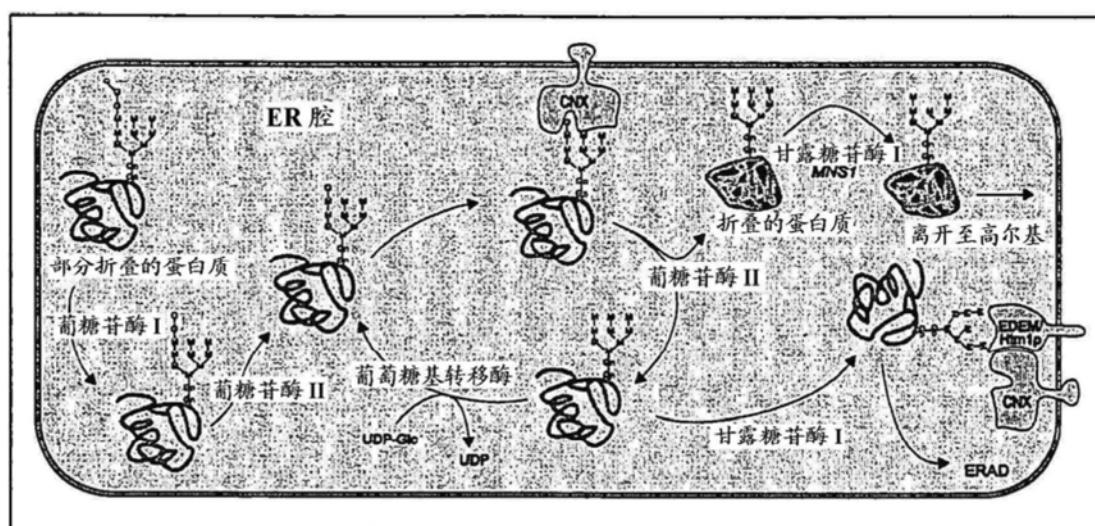


图2

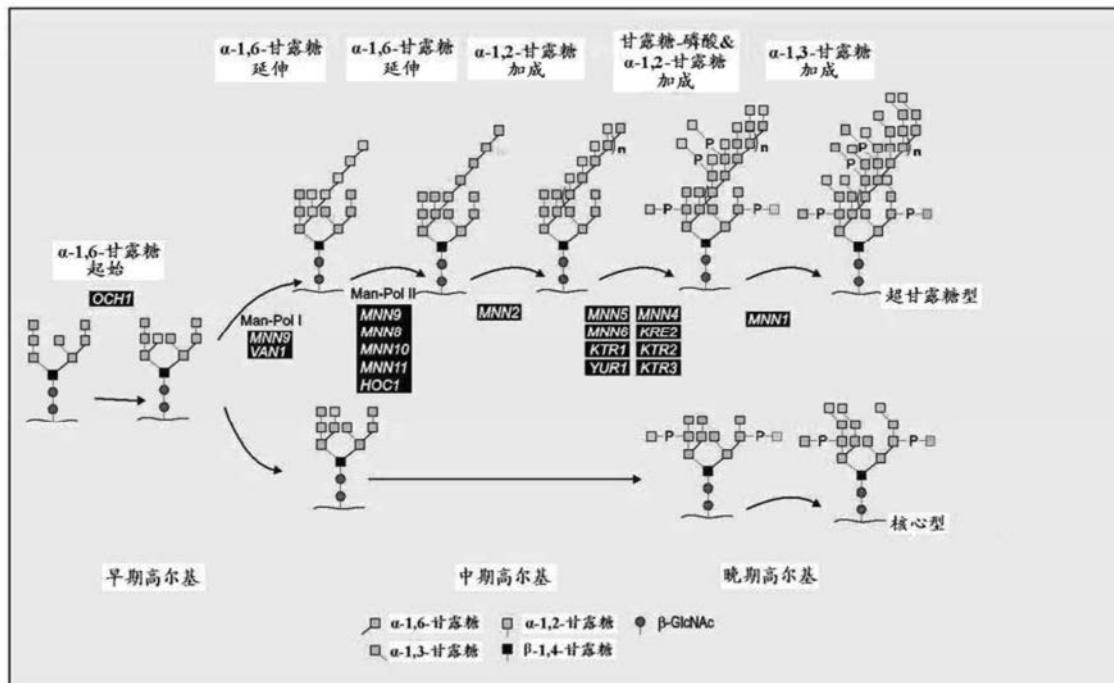
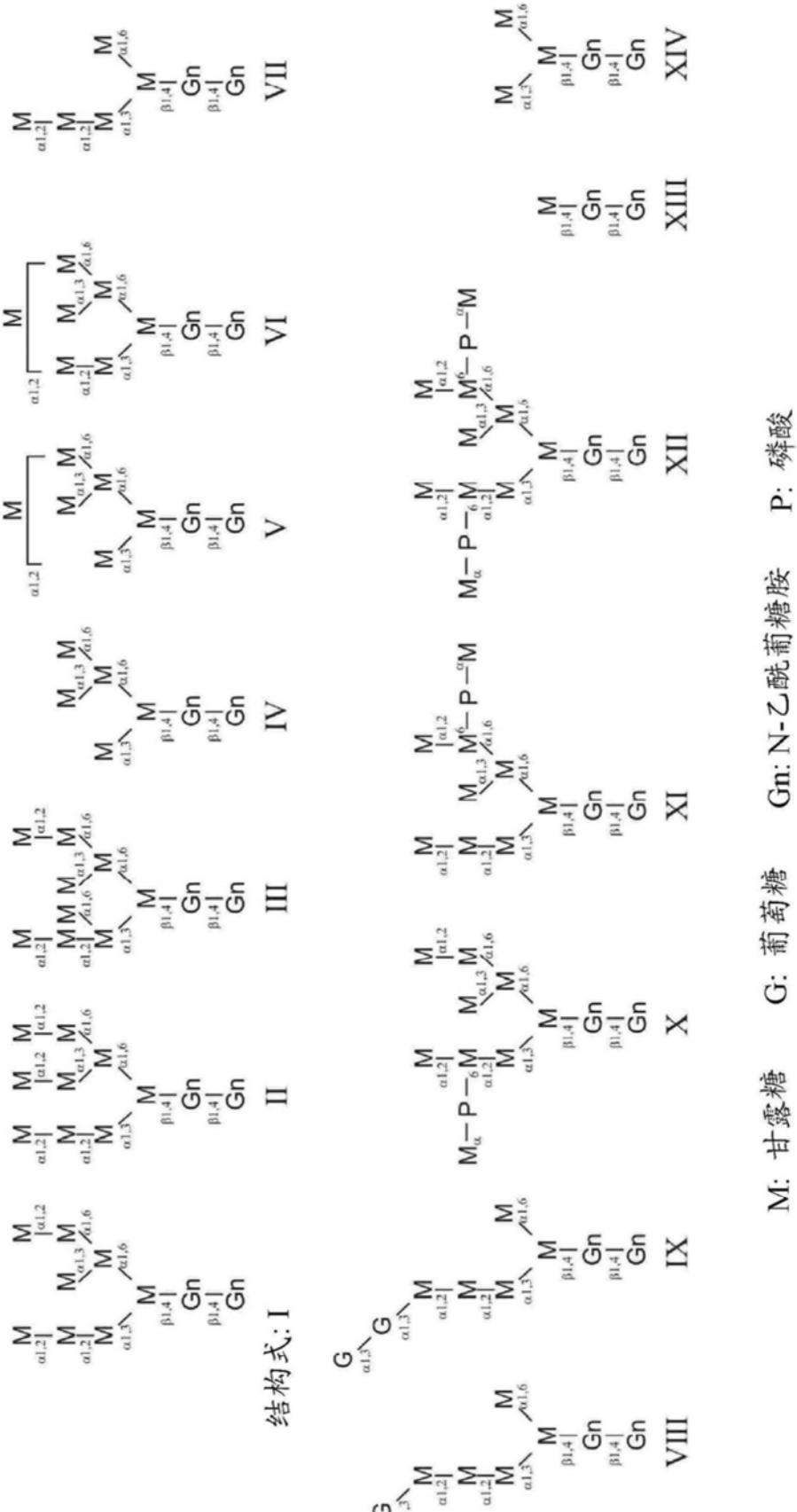
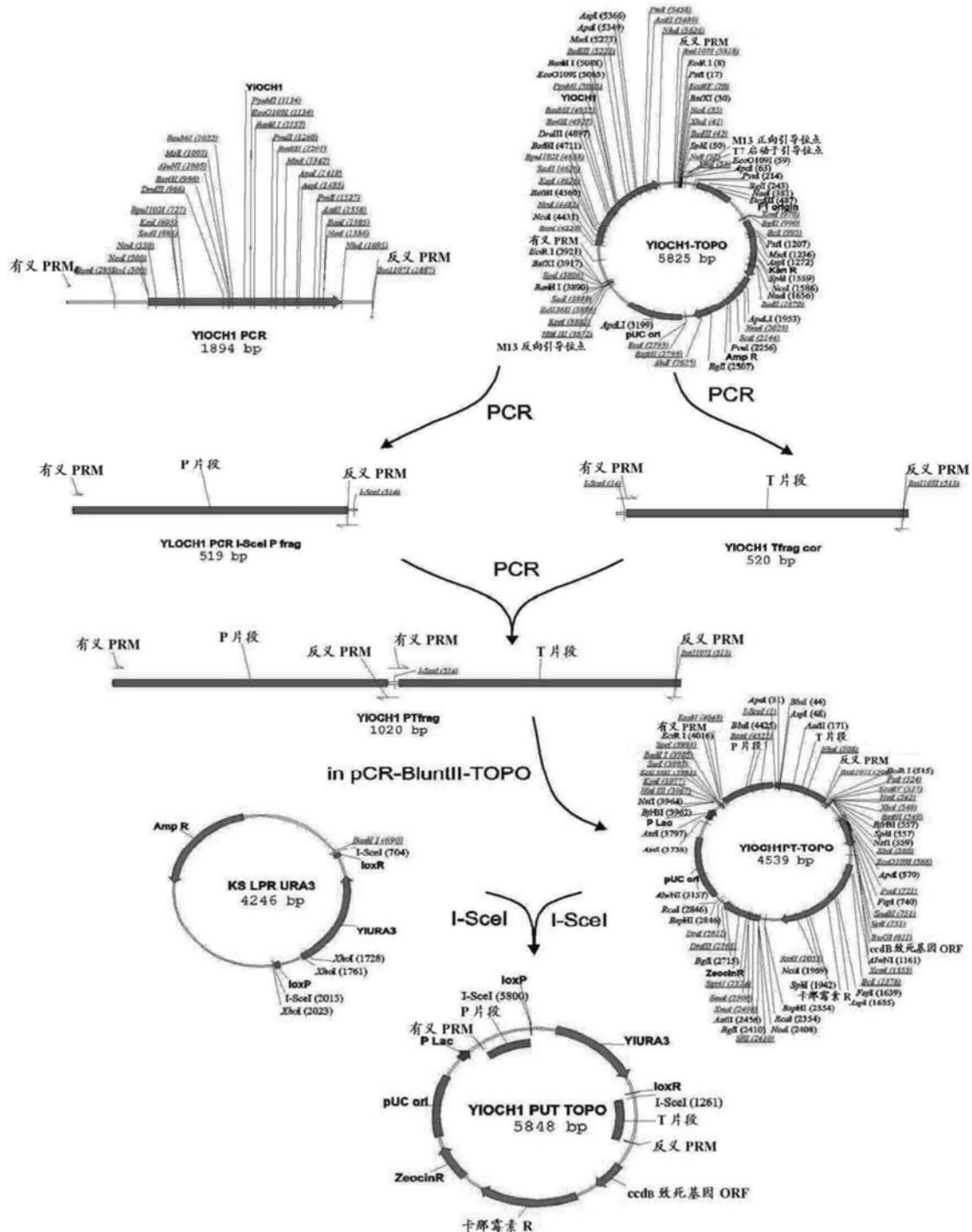


图3



M: 吲露糖 G: 葡萄糖 Gn: N-乙酰葡萄糖胺 P: 磷酸



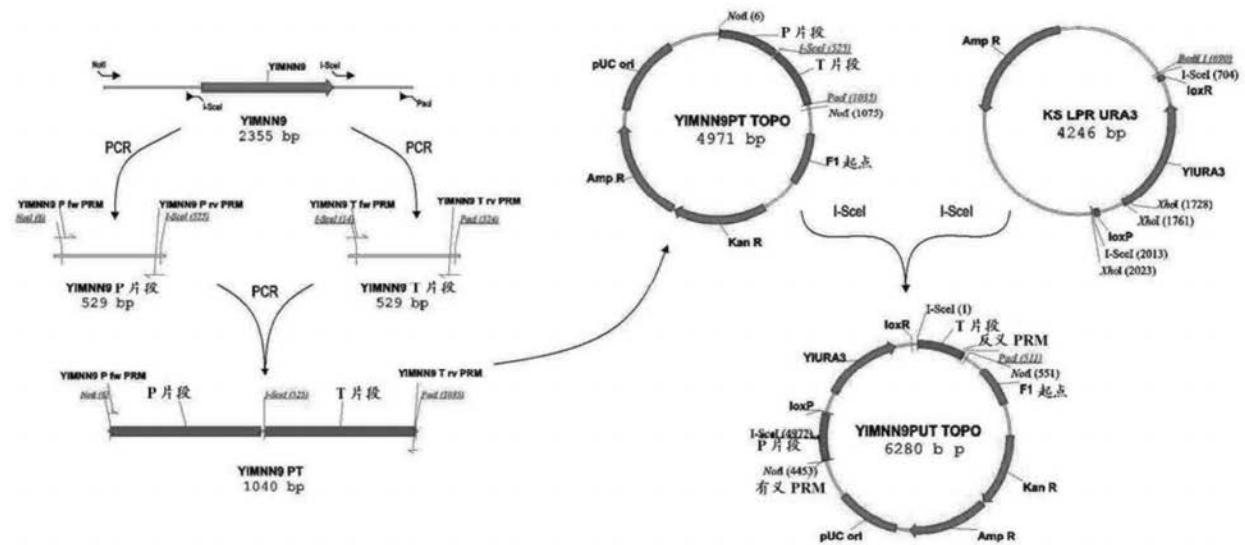


图6

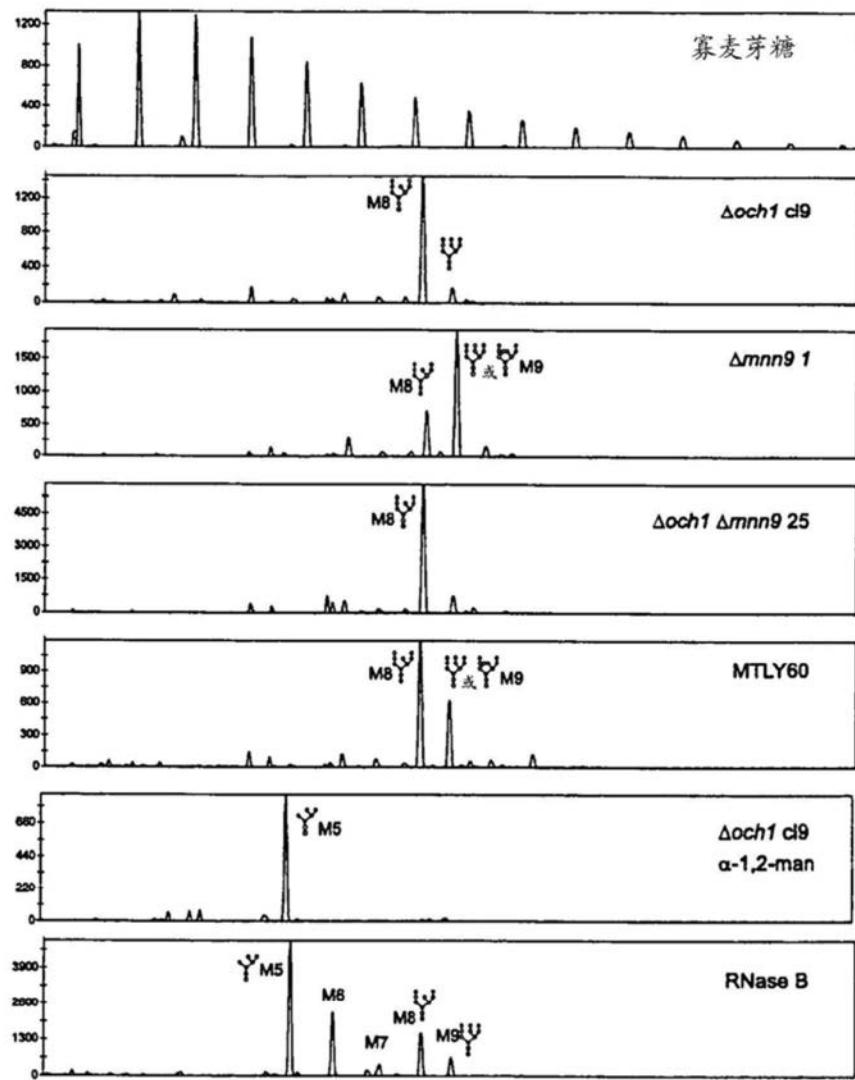


图7

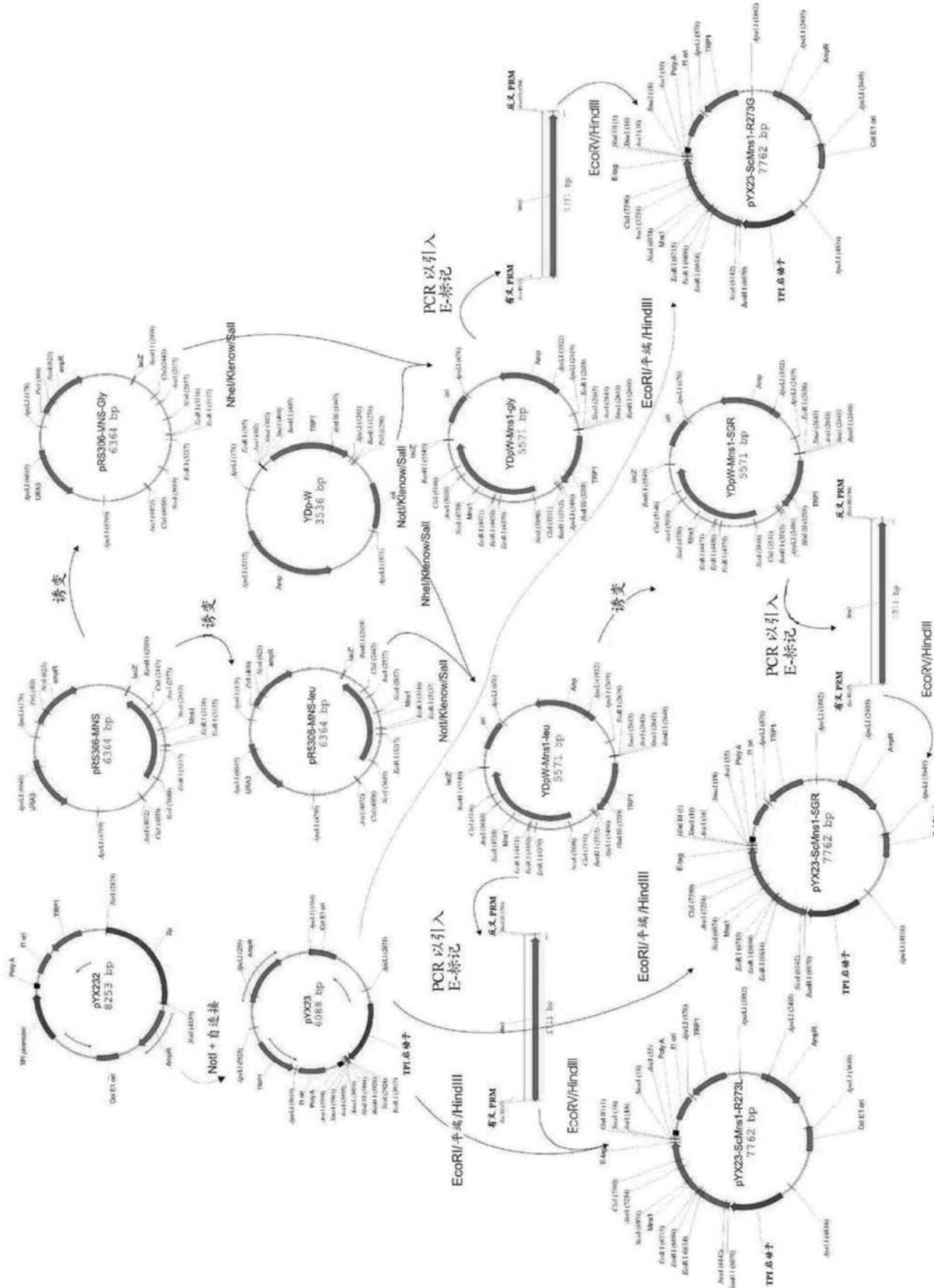


图8

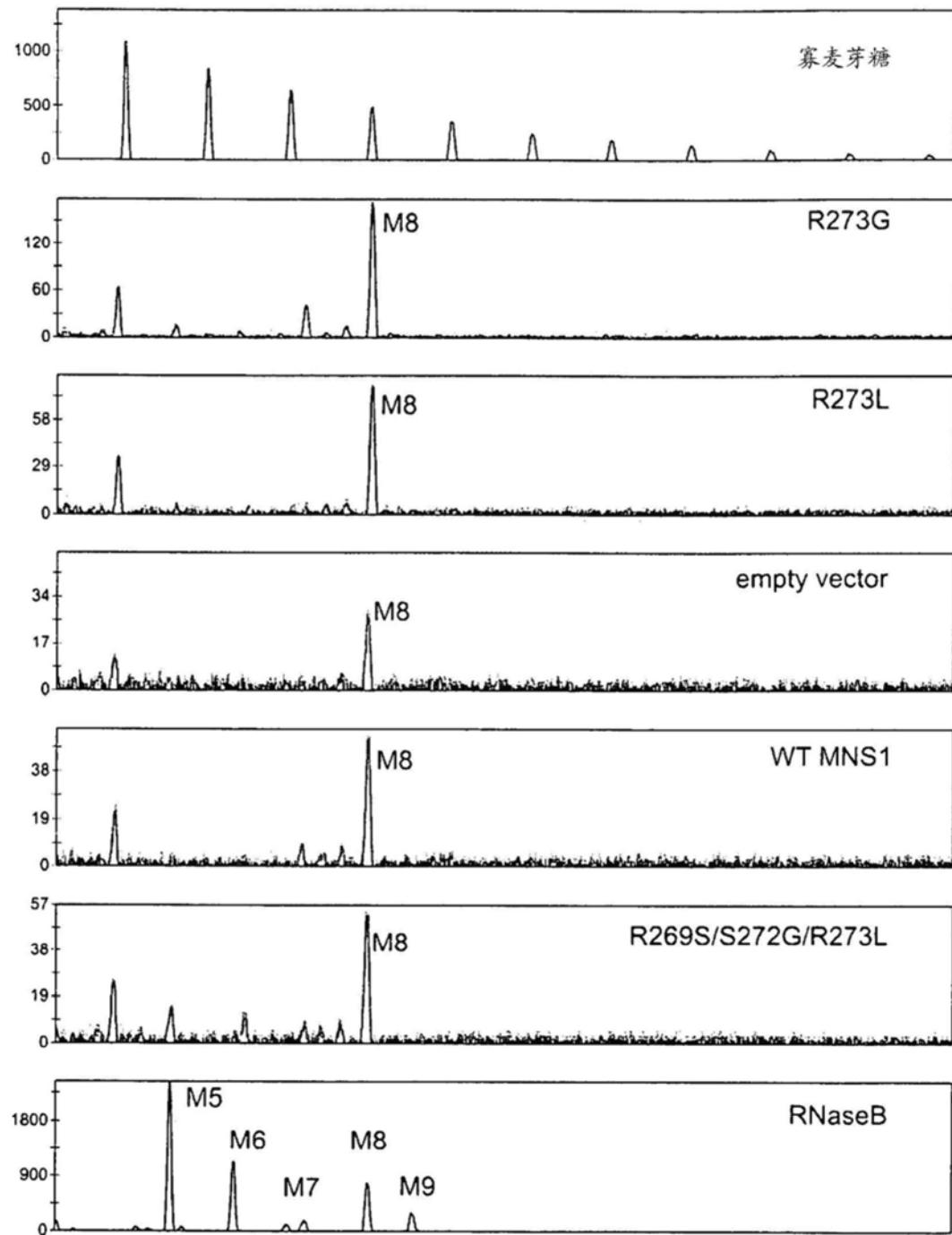


图9

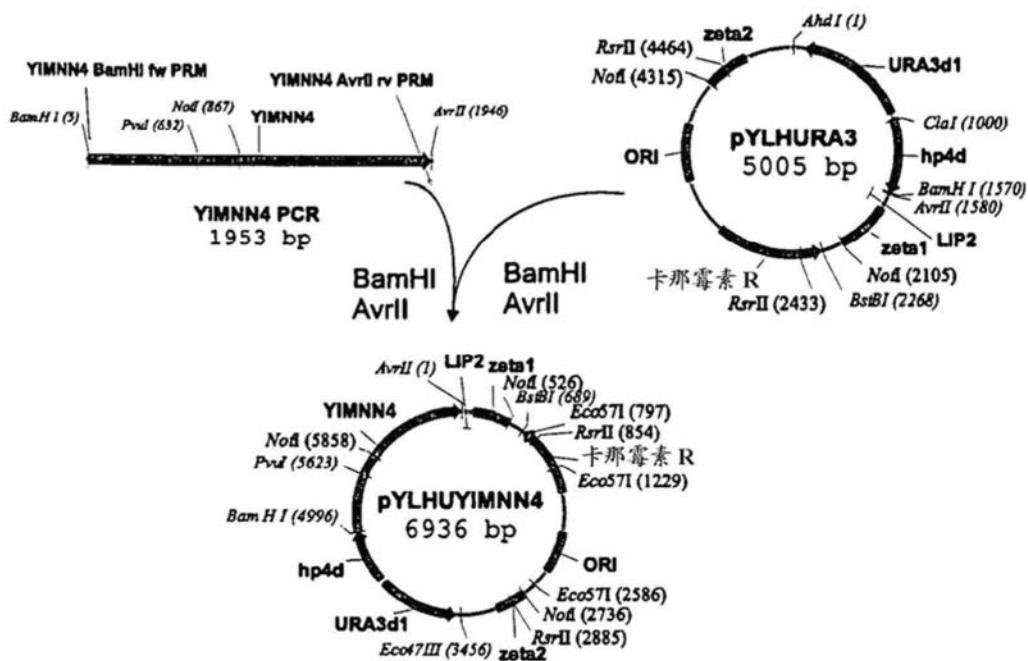


图10

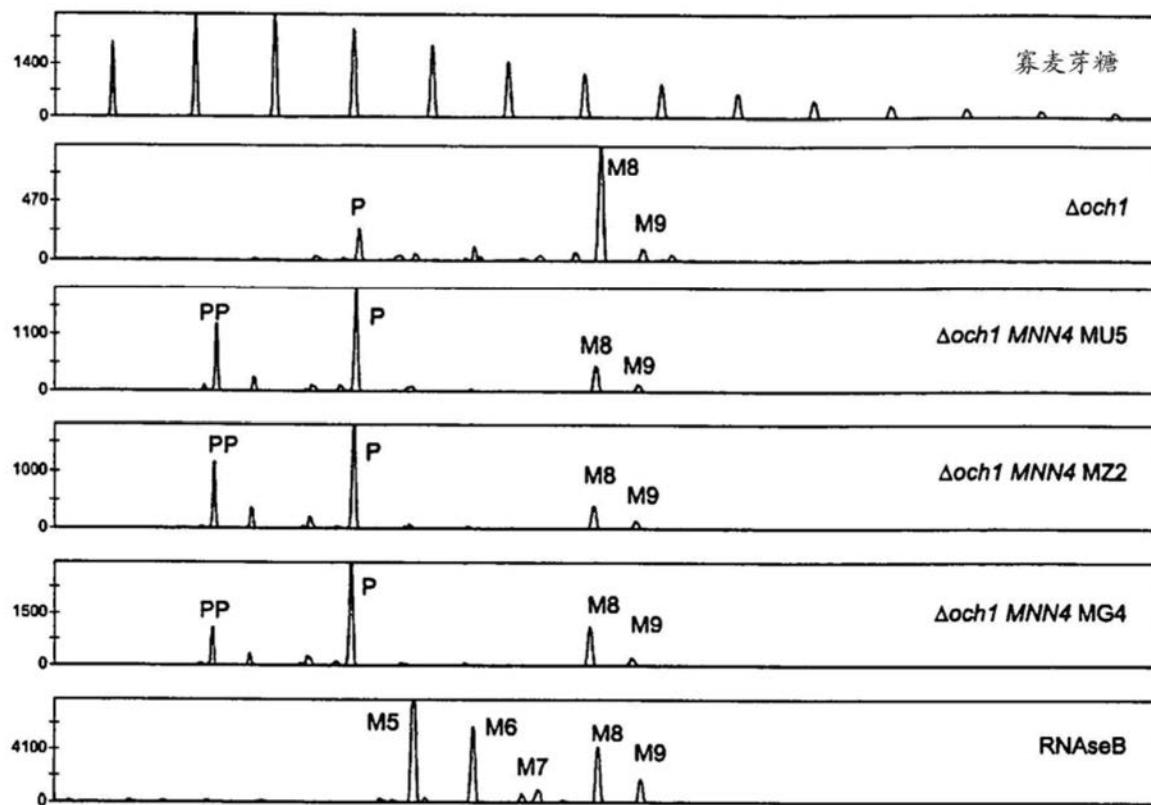


图11

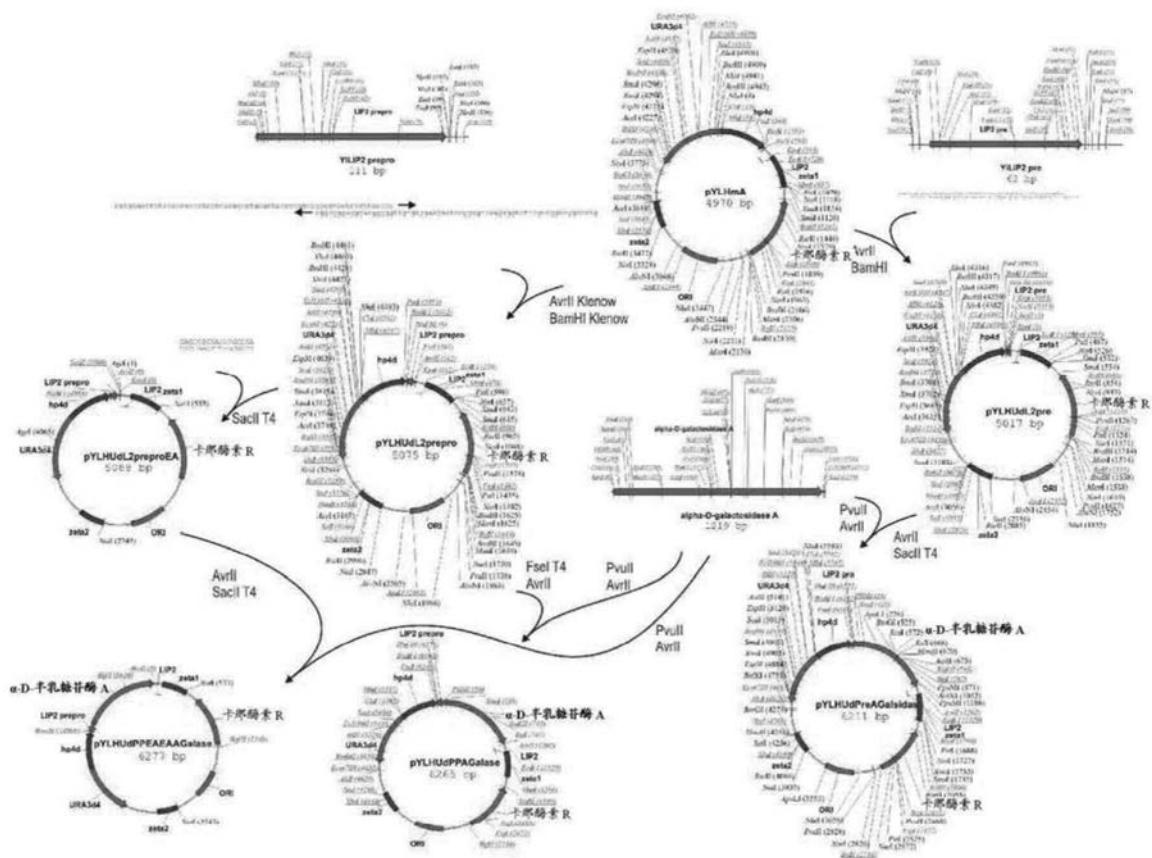


图12

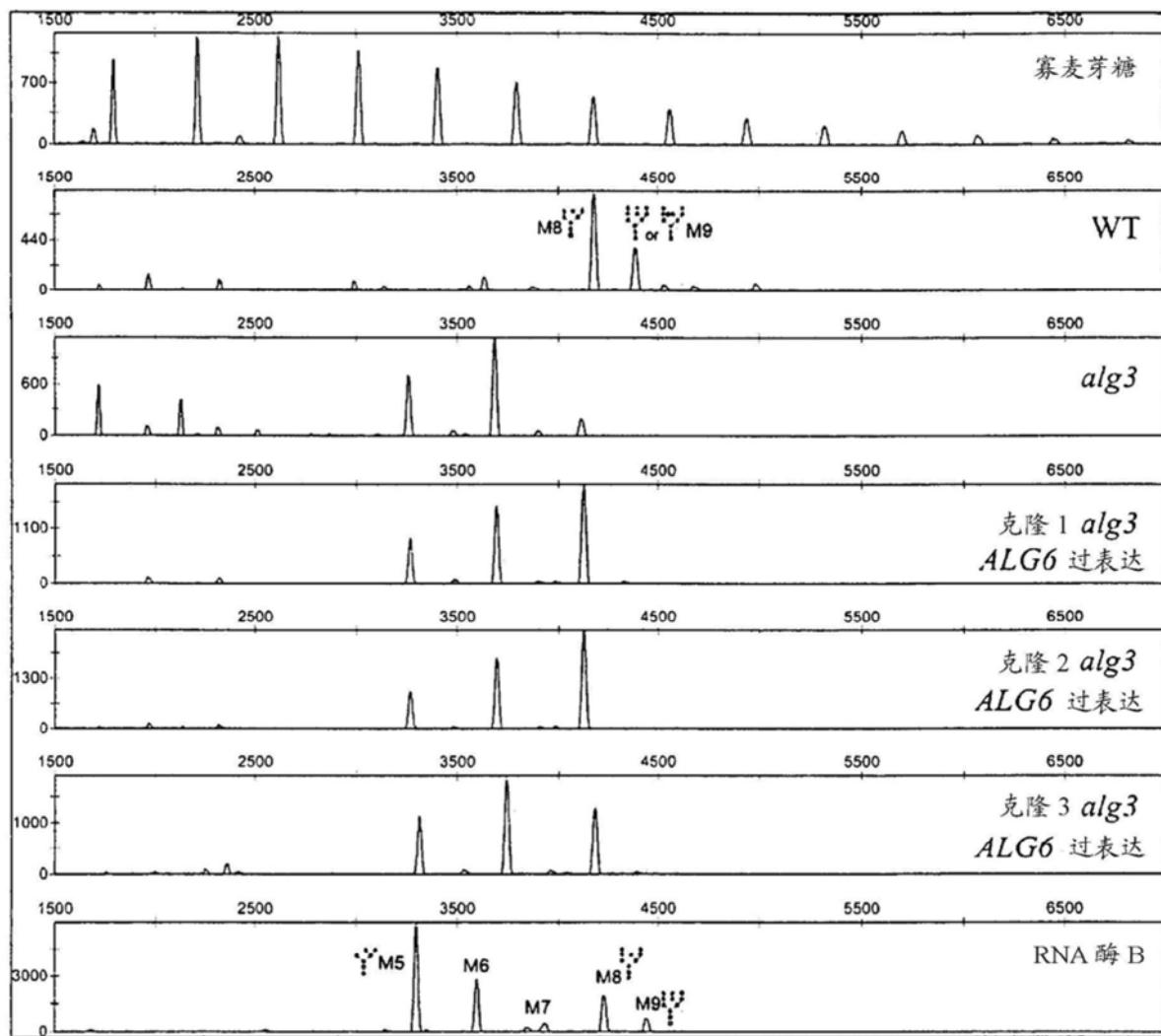


图13

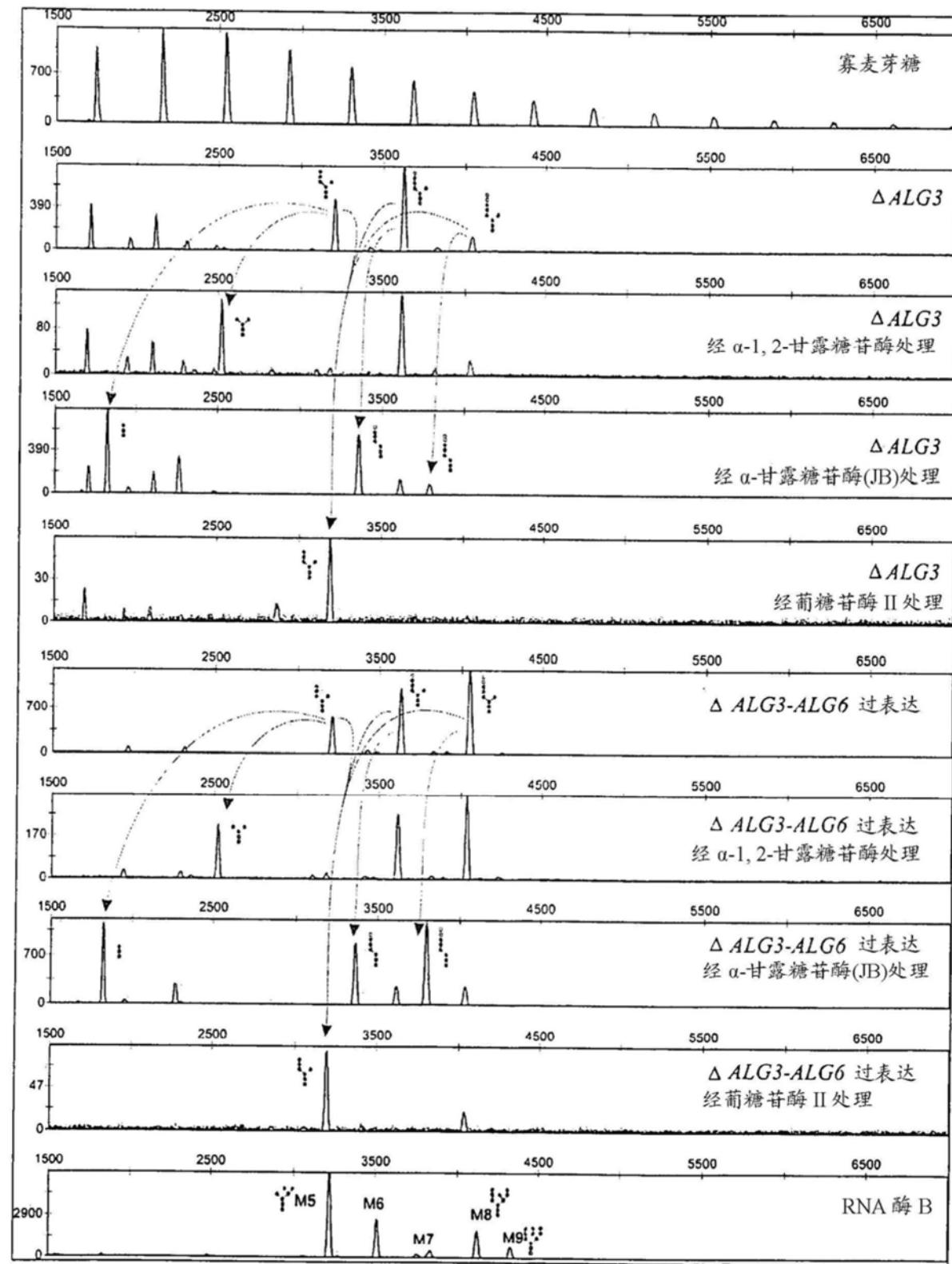


图14

序列	GGTCTCGCCAGCGGCCACCCCTTCACCACGATAATGAGACCGTCGT 
HAC1YL	GGTCTCGCCAGCGGCCACCCCTTCACCACGATAATGAGACCGTCGT
序列	CCCCGACTCTCCTATTGTGAAGACCGAGGAAGTCGACTCTACAAACTTC 
HAC1	CCCCGACTCTCCTATTGTGAAGACCGAGGAAGTCGACTCTACAAACTTC
序列	TCCCTCACACGGAGTCCTCCTCCCCCCCCGAACTAGCTGAGAGCACTGGC 
HAC1	TCCCTCACACGGAGTCCTCCTCCCCCCCCGAACTAGCTGAGAGCACTGGC
序列	TCAGGCTGCCATCGTCGACTCTGTCCTGCGACGAAACTGATTATCTTGT 
HAC1	TCAGGCTGCCATCGTCGACTCTGTCCTGCGACGAAACTGATTATCTTGT
序列	GGACCGGGCGCGTCATCCAG-----C 
HAC1	GGACCGGGCGCGTCATCCAG CAGTGATGACTGTCGCAACTACTGACCAGC
序列	AGCGTCGGACAAGATTTCWTTTCATCAAGGACGAGCCC GTTGACGACG             .
HAC1	AGCGTCGGACAAGATTTCATCAAGGACGAGCCC GTTGACGACG
序列	AGCTTGGACTGCATGGACTGTCGGATGACTTCACCCCTGTTGAAGACAAC 
HAC1	AGCTTGGACTGCATGGACTGTCGGATGACTTCACCCCTGTTGAAGACAAC
序列	AAGCAGCCTGCCAGCACGACTTTATTGCTGATCTAG 
HAC1	AAGCAGCCTGCCAGCACGACTTTATTGCTGATCTAG

图15

## 5' 剪接位点

巴斯德华赤酵母      atcccagcagtgatgacg  
 酿酒酵母                atcccagccgtgattacg  
 \*\*\*\*\* \* \* \* \* \*

## 3' 剪接位点

巴斯德华赤酵母      ggttgcagcaccat  
 酿酒酵母                tgtccgaaggcgcagt  
 \*\*\* \* \* \* \* \*

图16

NI.	1 atgcccgtagattcttctcataagacagactagccacttccacctcgtaa 	50
I.	1 atgcccgtagattcttctcataagacagactagccacttccacctcgtaa 	50
NI.	51 aagagcaaagacggaagaagaagaaaaggagcagcgtcgagtggAACGTATCC 	100
I.	51 aagagcaaagacggaagaagaagaaaaggagcagcgtcgagtggAACGTATCC 	100
NI.	101 tacgtaataggagagcggcccattgttcagagagaagaaacgaagacac 	150
I.	101 tacgtaataggagagcggcccattgttcagagagaagaaacgaagacac 	150
NI.	151 gttgaatttctggaaaaccacgtcgacgtcgacattggatctgcacttcaaga 	200
I.	151 gttgaatttctggaaaaccacgtcgacgtcgacattggatctgcacttcaaga 	200
NI.	201 atcagccaaagccactaacaagttgaaagaaatacaagatatcatttt 	250
I.	201 atcagccaaagccactaacaagttgaaagaaatacaagatatcatttt 	250
NI.	251 caagggttggaaagccttaggtggtaccgtctcagatttgatttaacagtt 	300
I.	251 caagggttggaaagccttaggtggtaccgtctcagatttgatttaacagtt 	300
NI.	301 ccggaaagtcgattttccaaatcttcgttggacccatgtctgatct 	350
I.	301 ccggaaagtcgattttccaaatcttcgttggacccatgtctgatct 	350
NI.	351 ctcacttcttcgaaatcgagaaagcatctacatccactcgcagatctt 	400
I.	351 ctcacttcttcgaaatcgagaaagcatctacatccactcgcagatctt 	400
NI.	401 tgactgaggatctggacgaagatgacgtcgctgaatatgacgacgaagaa 	450
I.	401 tgactgaggatctggacgaagatgacgtcgctgaatatgacgacgaagaa 	450
NI.	451 gaggacgaagagttacccaggaaaatgaaagtcttaacgacaaaaacaa 	500
I.	451 gaggacgaagagttacccaggaaaatgaaagtcttaacgacaaaaacaa 	500
NI.	501 gagcacatcttatcaagcaggagaagttgaatgaacttccatctcctttgt 	550
I.	501 gagcacatcttatcaagcaggagaagttgaatgaacttccatctcctttgt 	550
NI.	551 catccgattttcagacgttagatgaagaaaagtcaactctcacacattta 	600
I.	551 catccgattttcagacgttagatgaagaaaagtcaactctcacacattta 	600
NI.	601 aagttgcaacagcaacaacaacaaccagttagacaattatgtttctactcc 	650
I.	601 aagttgcaacagcaacaacaacaaccagttagacaattatgtttctactcc 	650

图17A

NI.	651 tttgagtcttccggaggattcagttgatttattaacccaggtaactaa	700
I.	651 tttgagtcttccggaggattcagttgatttattaacccaggtaactaa	700
NI.	701 aaatagagtccgatgagaacttcttggatgtcaataactttacaata	750
I.	701 aaatagagtccgatgagaacttcttggatgtcaataactttacaata	750
NI.	751 aaacacgaaaatgacaccgactacattactacagctccatcagggttccat	800
I.	751 aaacacgaaaatgacaccgactacattactacagctccatcagggttccat	800
NI.	801 caatgatttttaattcttatgacatttagcgagtcgaatcggttgcattc	850
I.	801 caatgatttttaattcttatgacatttagcgagtcgaatcggttgcattc	850
NI.	851 atccagcagtgtatgacggattcatcttacacattacagcaggctccatc	900
I.	851 atccag-----	855
NI.	901 ggcttttctttgattggggggggggaaagttctgttagcagggaggcg	950
I.	856 -----	855
NI.	951 cagttcagttggcacatatcagttgacatgcatacgatcagg[gat]tgc	1000
I.	856 -----	855
NI.	1001 tcttgtaactggatgccaaagtcaaccgtttcggcgtacctacttc	1050
I.	856 -----	855
NI.	1051 tacttaactcttgaattgagatgaaacaatcgcccttgcgagacgcg	1100
I.	856 -----	855
NI.	1101 tagtttaatttaagatgaatcaatttgcacaaaagaacaaacaaaaaga	1150
I.	856 -----	855
NI.	1151 taacagcaaacttaagttcgtggctgcagcaccattaccgctaattgc	1200
I.	856 -----cagcaccattaccgctaattgc	878
NI.	1201 atttgatttaaatgactttgtattcttccaggaatag	1237
I.	879 atttgatttaaatgactttgtattcttccaggaatag	915

图17B

巴斯德华赤酵母	1 APFTANAFDLNDVFVFFQE	18
	..... : :     ....	
酿酒酵母	1 EAQSGLNSFELNDFFITS	18

图18

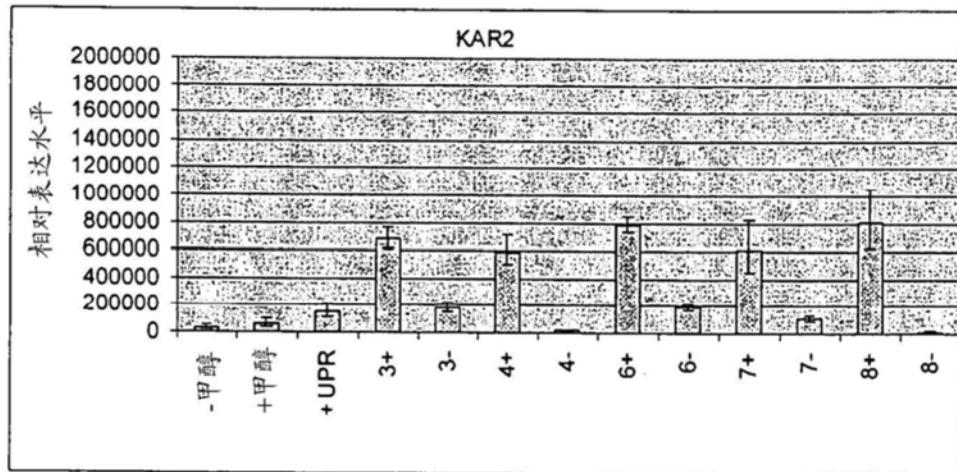


图19

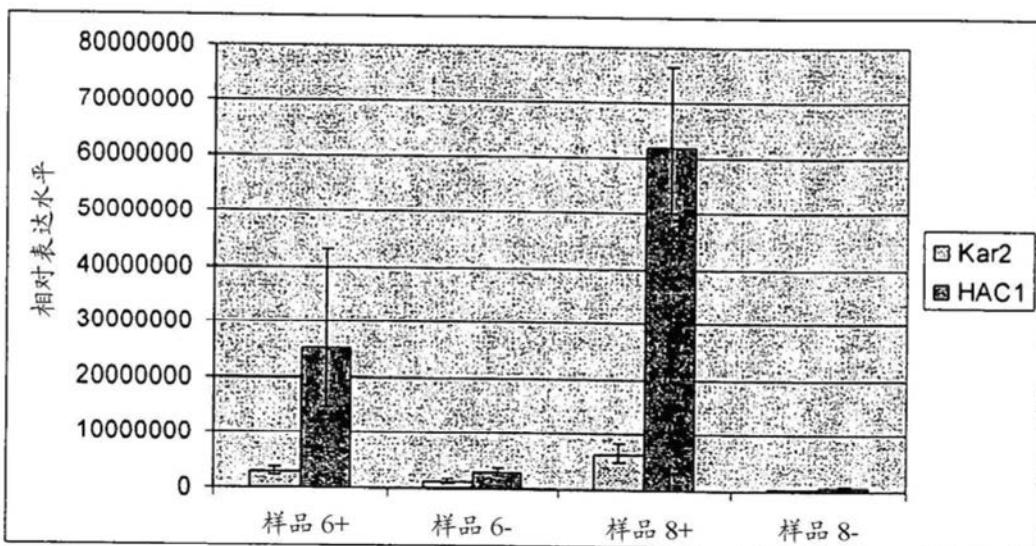


图20

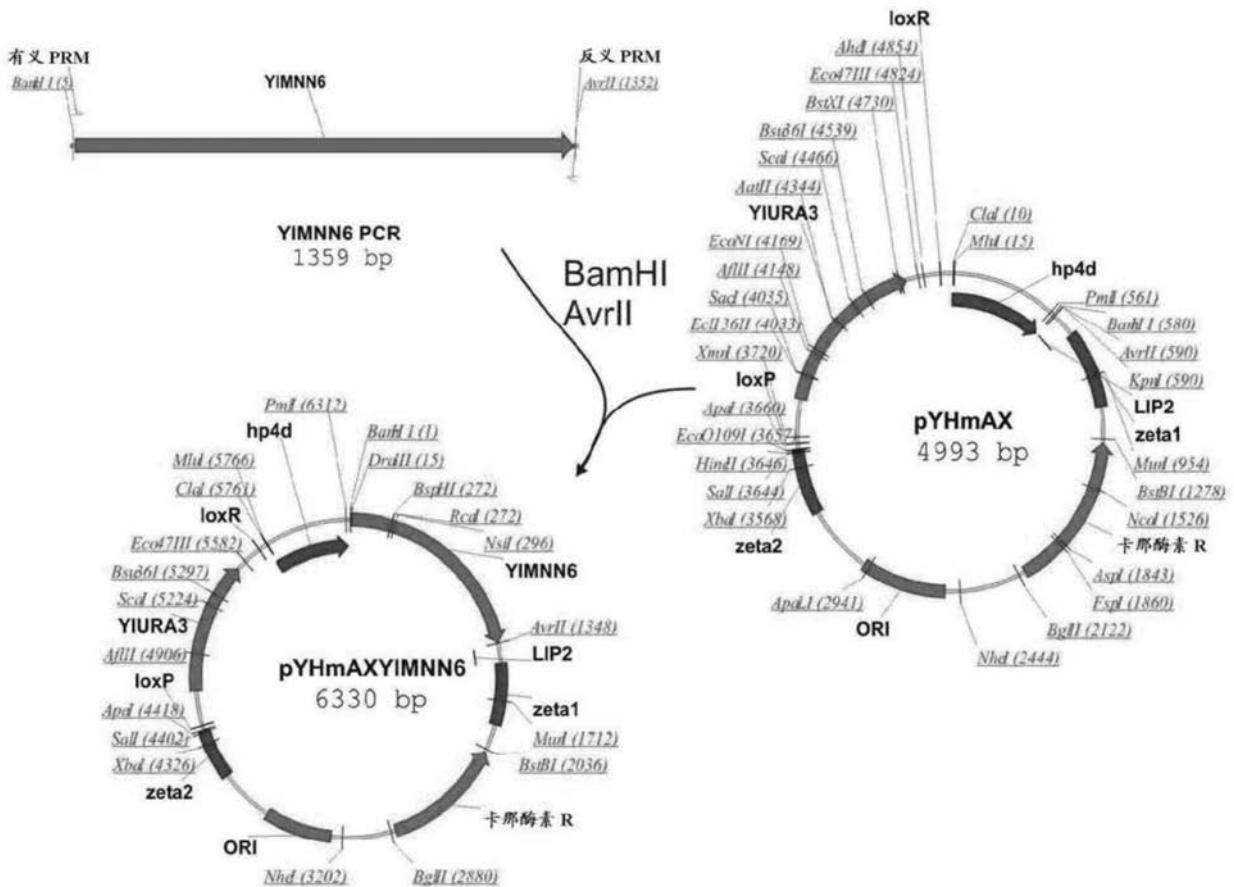


图21

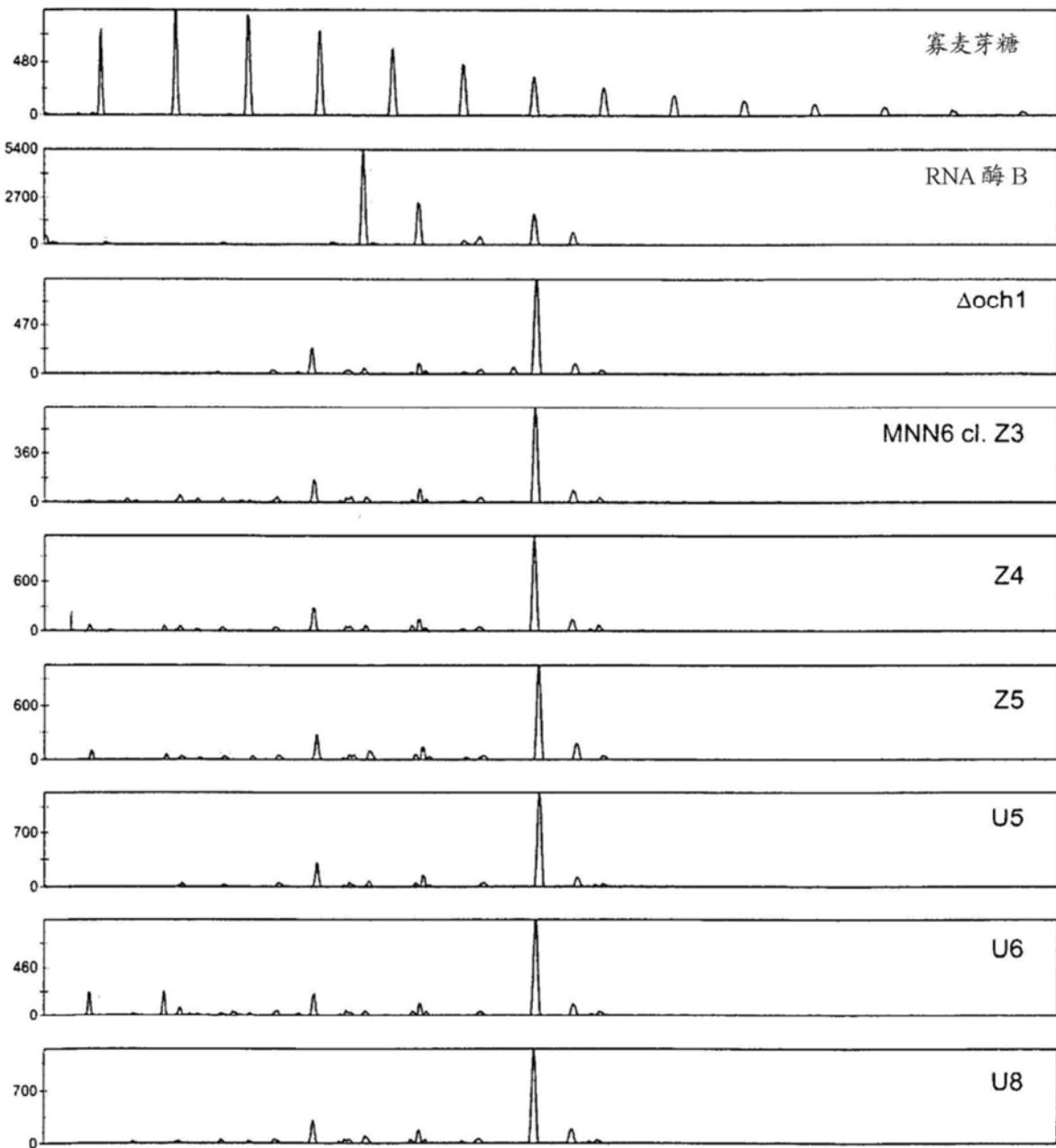


图22

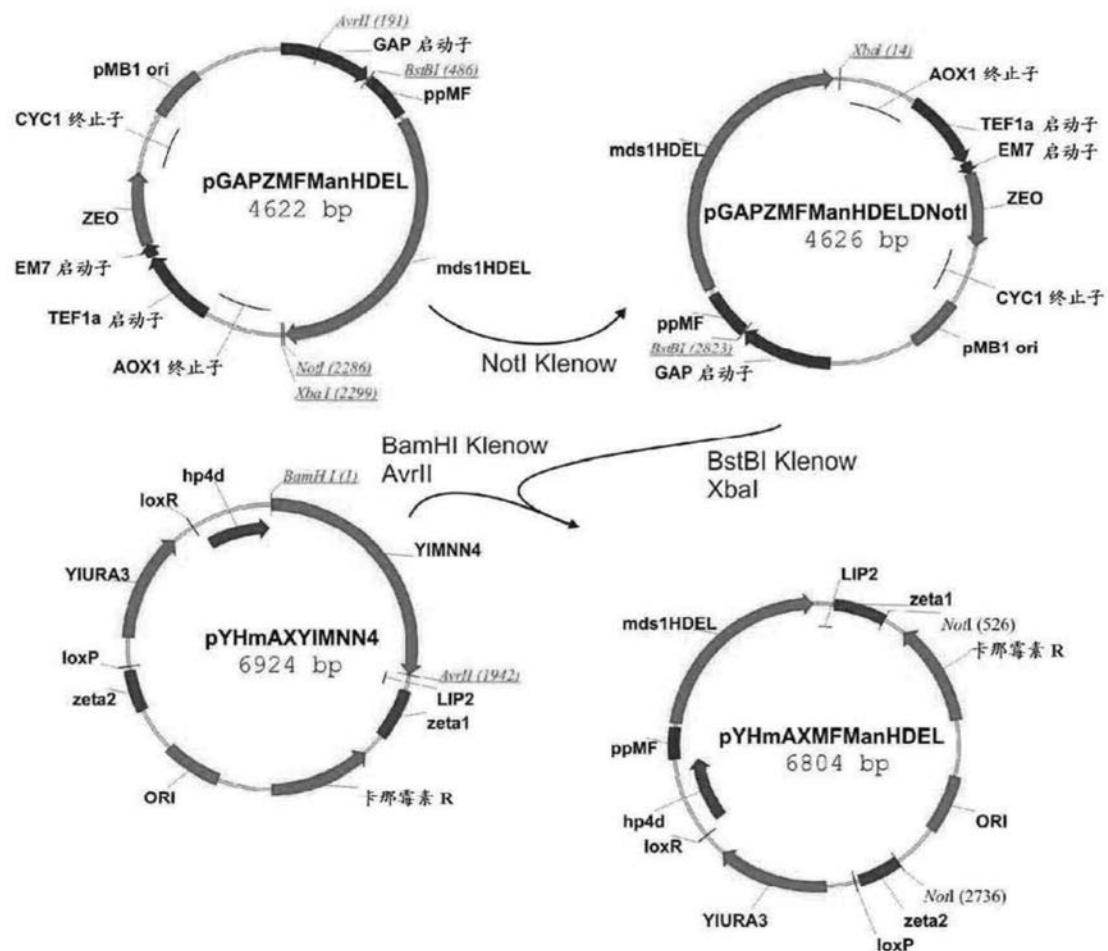


图23

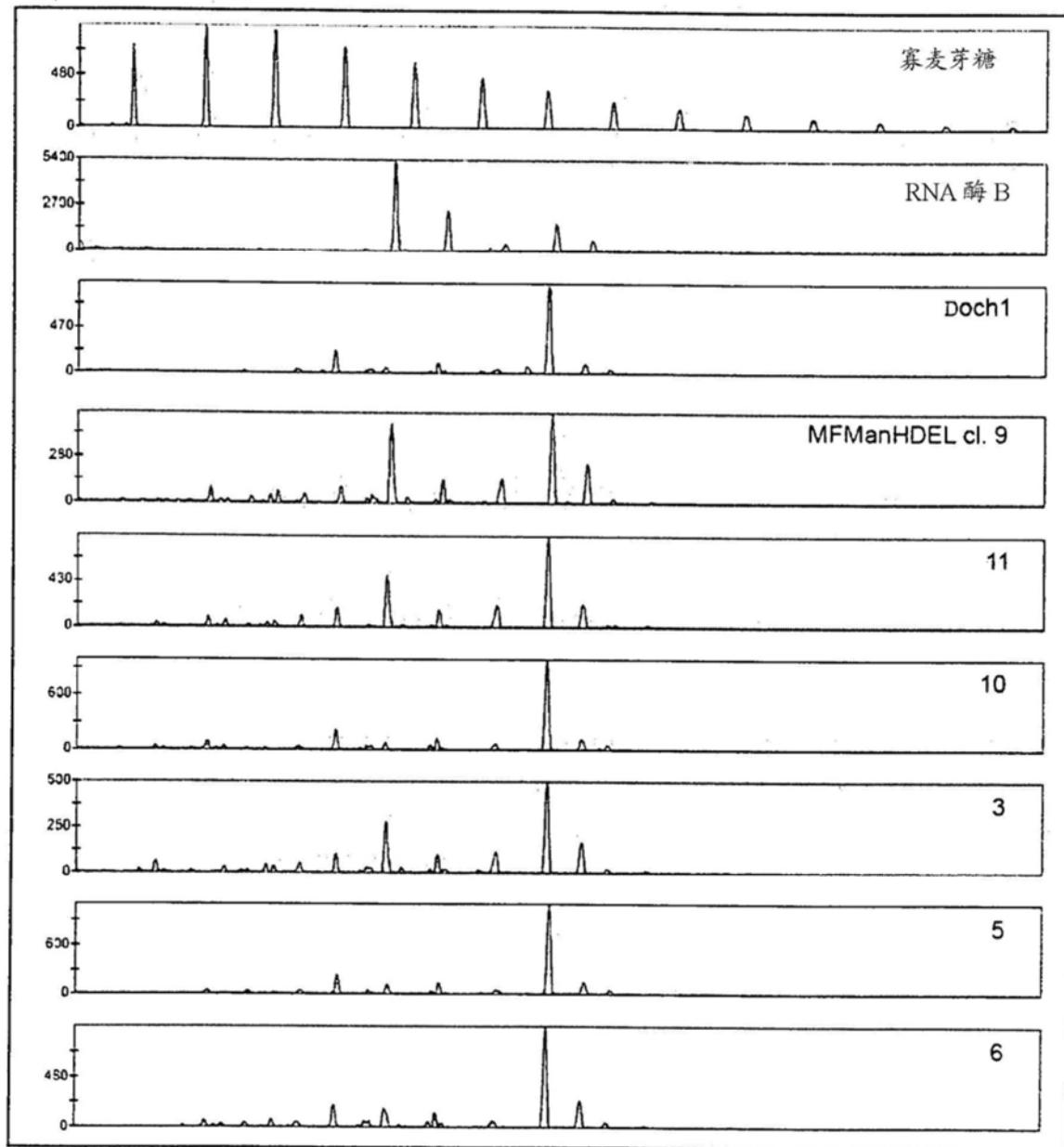


图24

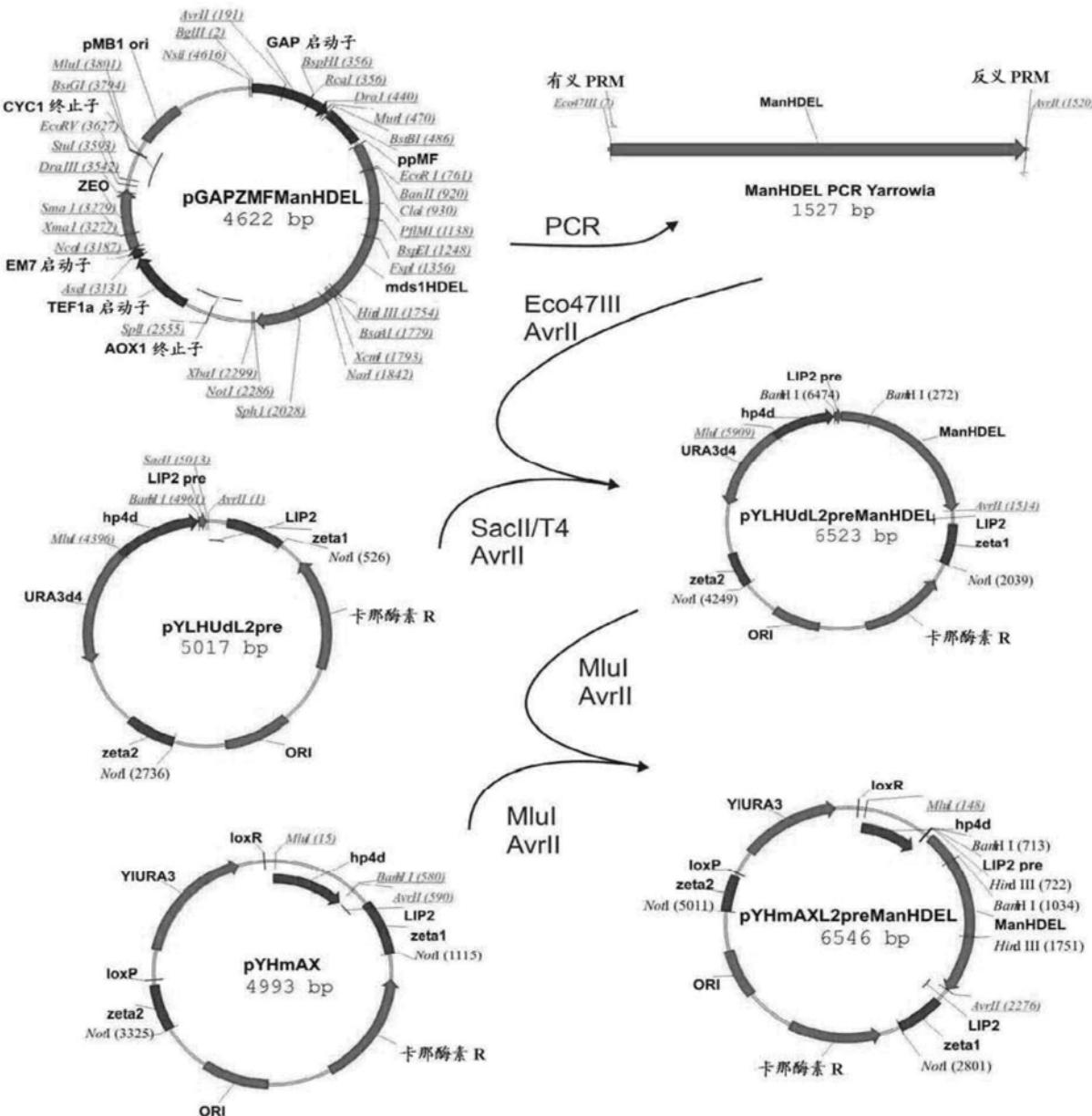


图25

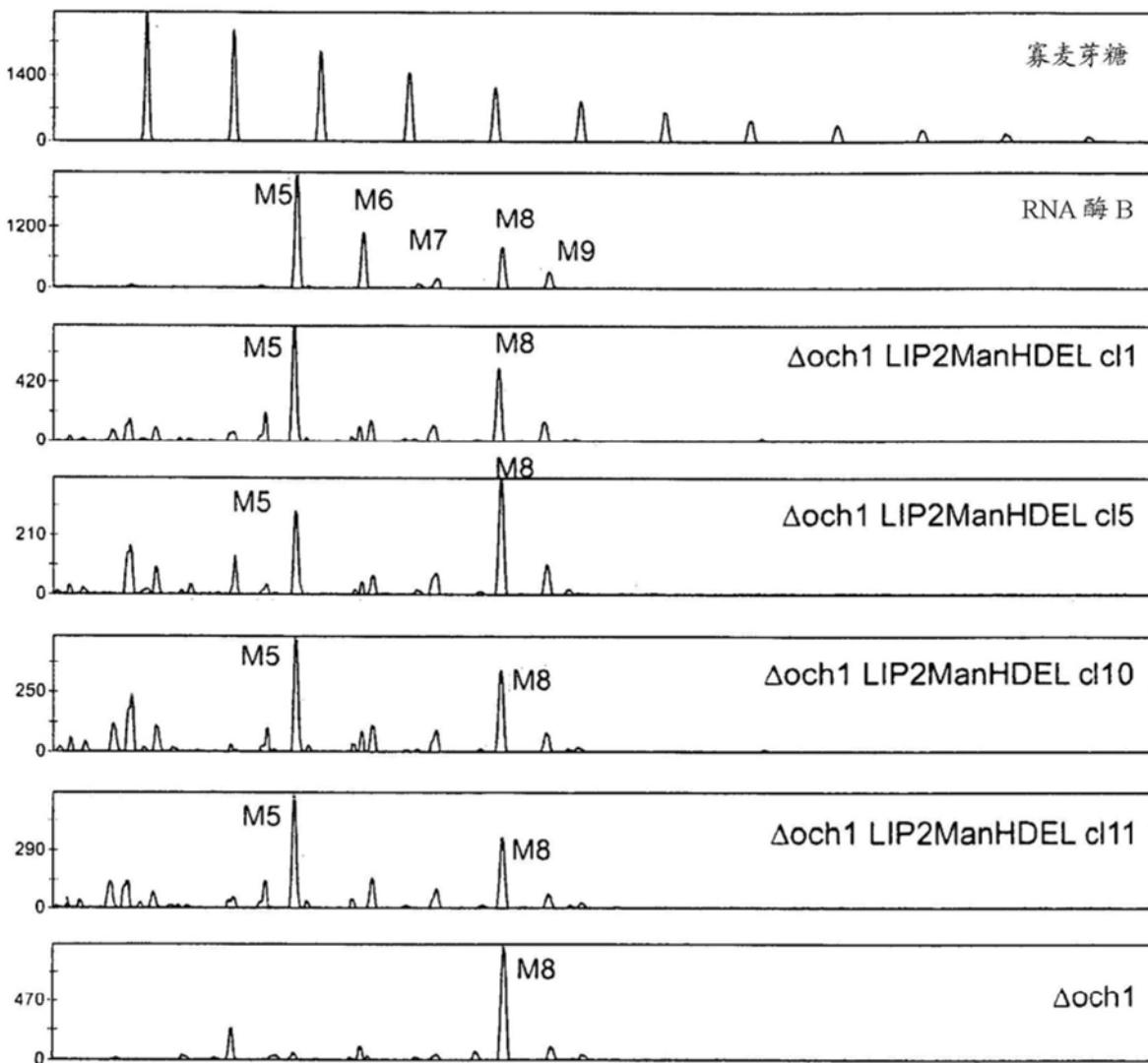


图26

MSIKREESFTPTPEDLGSPLTADSPGSPESGDKRKKDLTLPLPAGALPPRKRAKTENEKEQRRRIERIMRNRQAAHASREKKRRHLEDLEKKCSELSENNDLHHQVTESKKTNMHLMEQHYSLVAKLQLQLSSLVNMAKSSGALAGVDVPDMSDVSMAPKLEMPTAAPSQPMGLASAPTLFNHDNETVVVDSPIVKTEEVNSTNFLHTESSSPPELAESTGSGSPSSTLSCDETDLVDRARHPAASAQDFIFIKD  
EPVDDELGLHGLSDDFTLFEDNKQPAQHDFIADLAHYESSVSNLFGGLE

图27A

MPVDSSHKTASPLPPRKRAKTEEEKEQRRVERILRNRRAA  
HASREKKRRHVEFLLENHVVDLESALQESAKATNKLKEIQDIIV  
SRLEALGGTVSDLDTVPEVDFPKSSDLEPMSDLSTSSKSEKAS  
TSTRRSLTELDDEDDVAEYDDEEDEELPRKMVKVLNDKNKSTS  
IKQEKLNELPSPLSSDFSDVDEEKSTLTHLKLQQQQQPVDNY  
VSTPLSLPEDSVDFINPGNLKIESDENFLLSSNTLQIKHENND  
YITTAPSGSINDFFNSYDISESNRLHHPAAPFTANAFDLNDFVF  
FQE

图27B

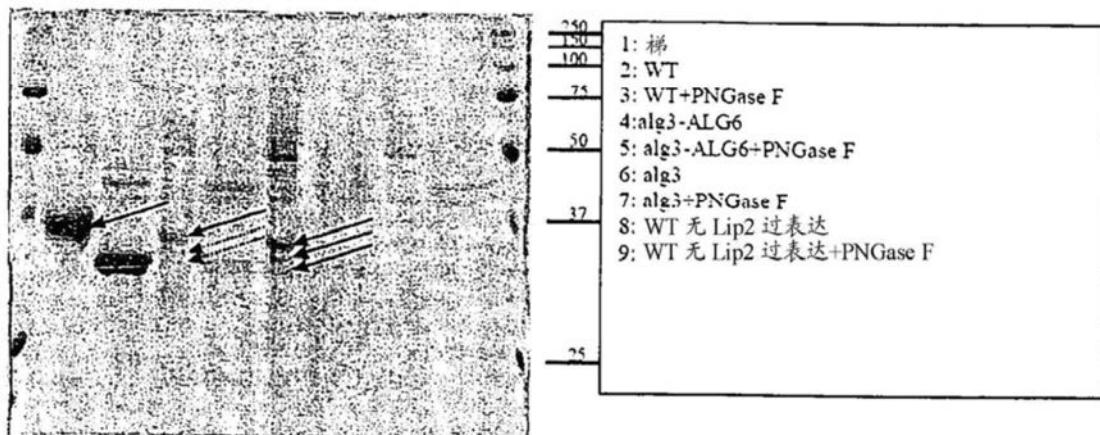


图28

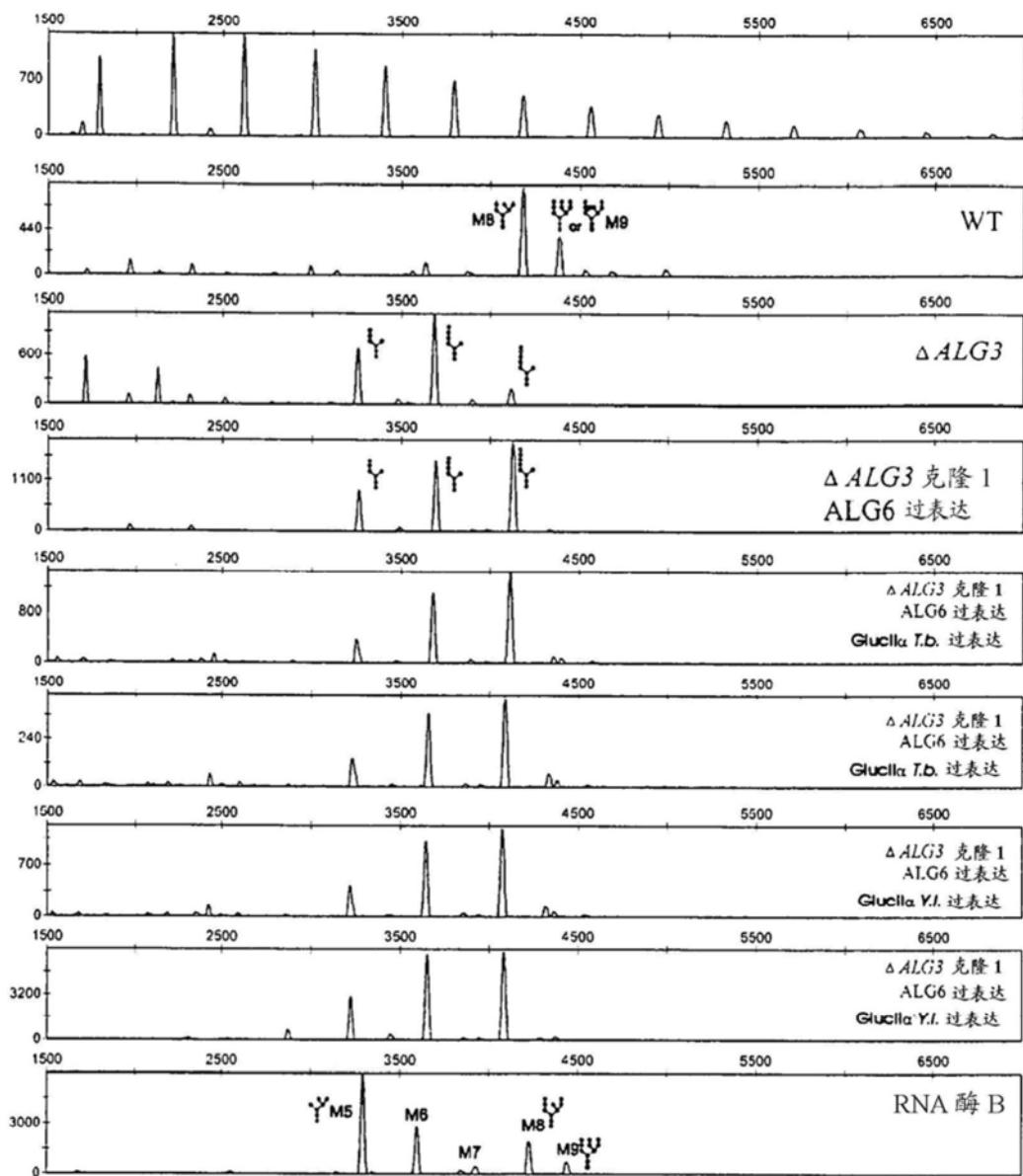


图29

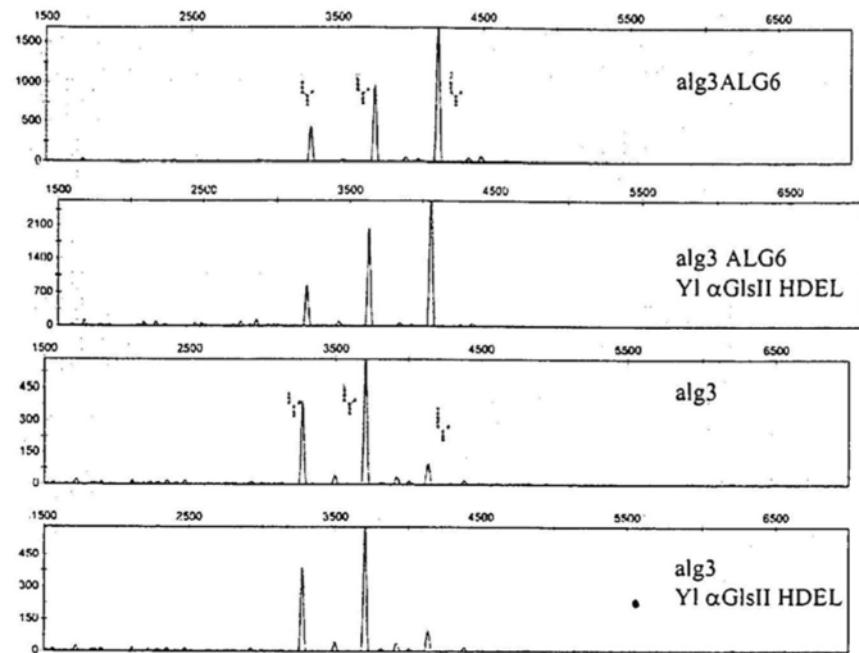


图30

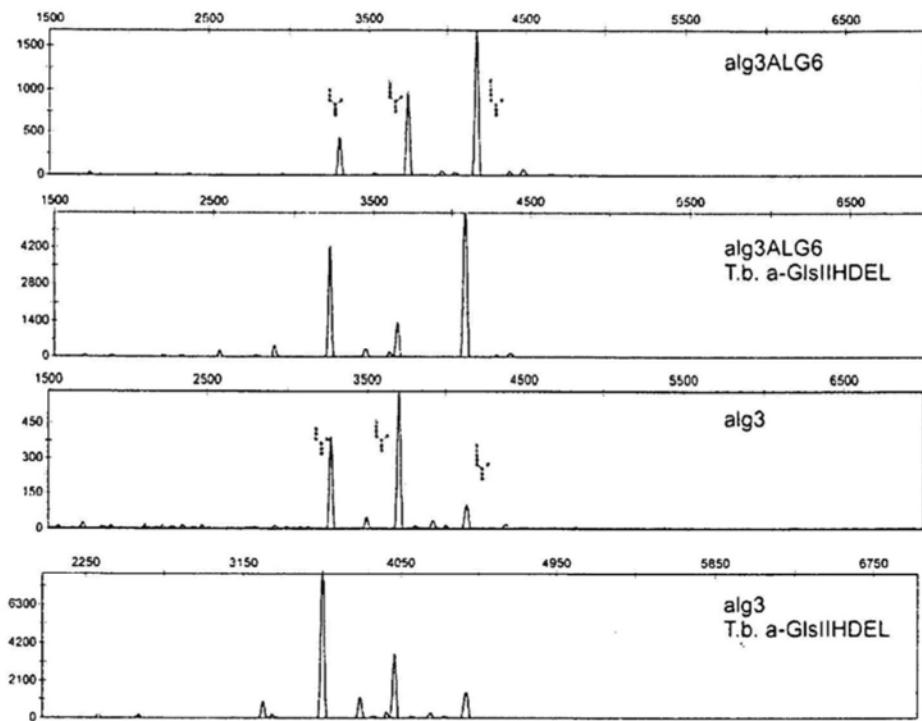


图31

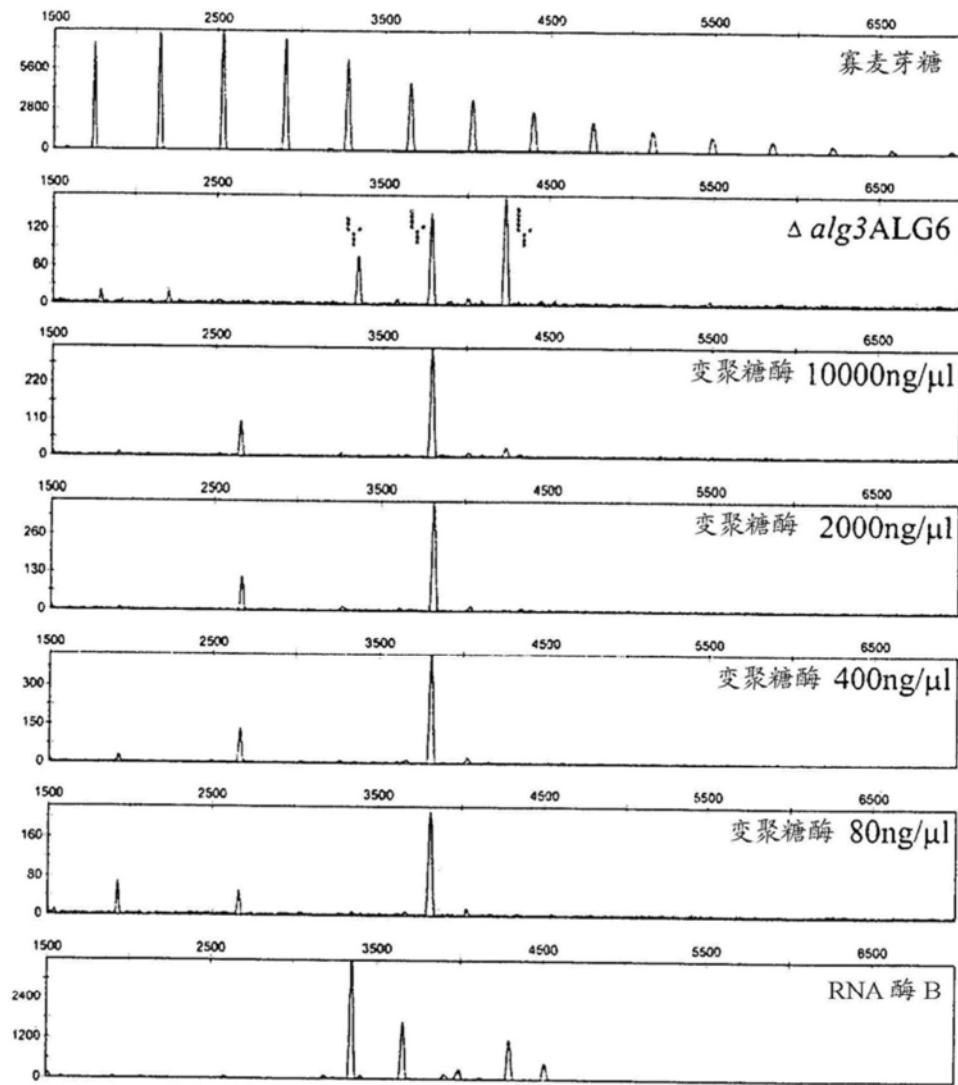


图32

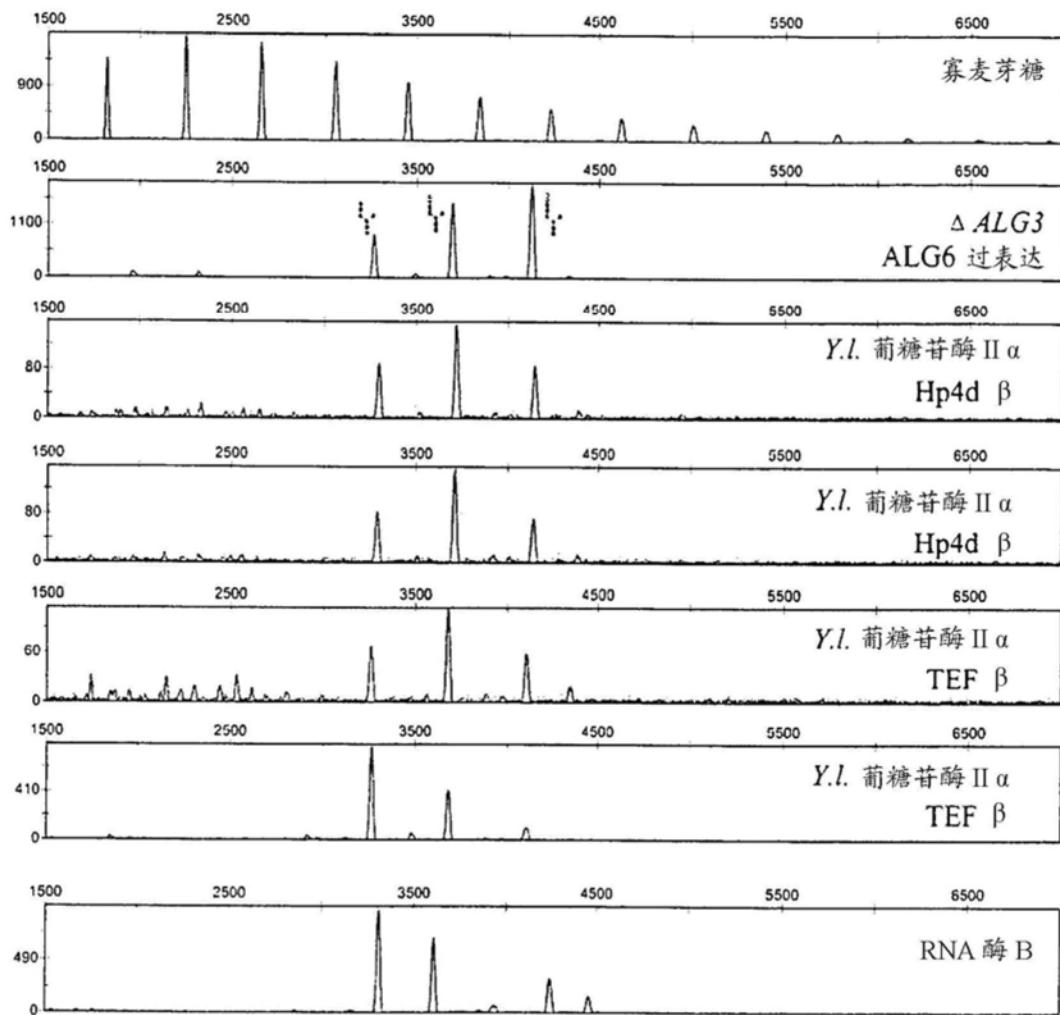


图33

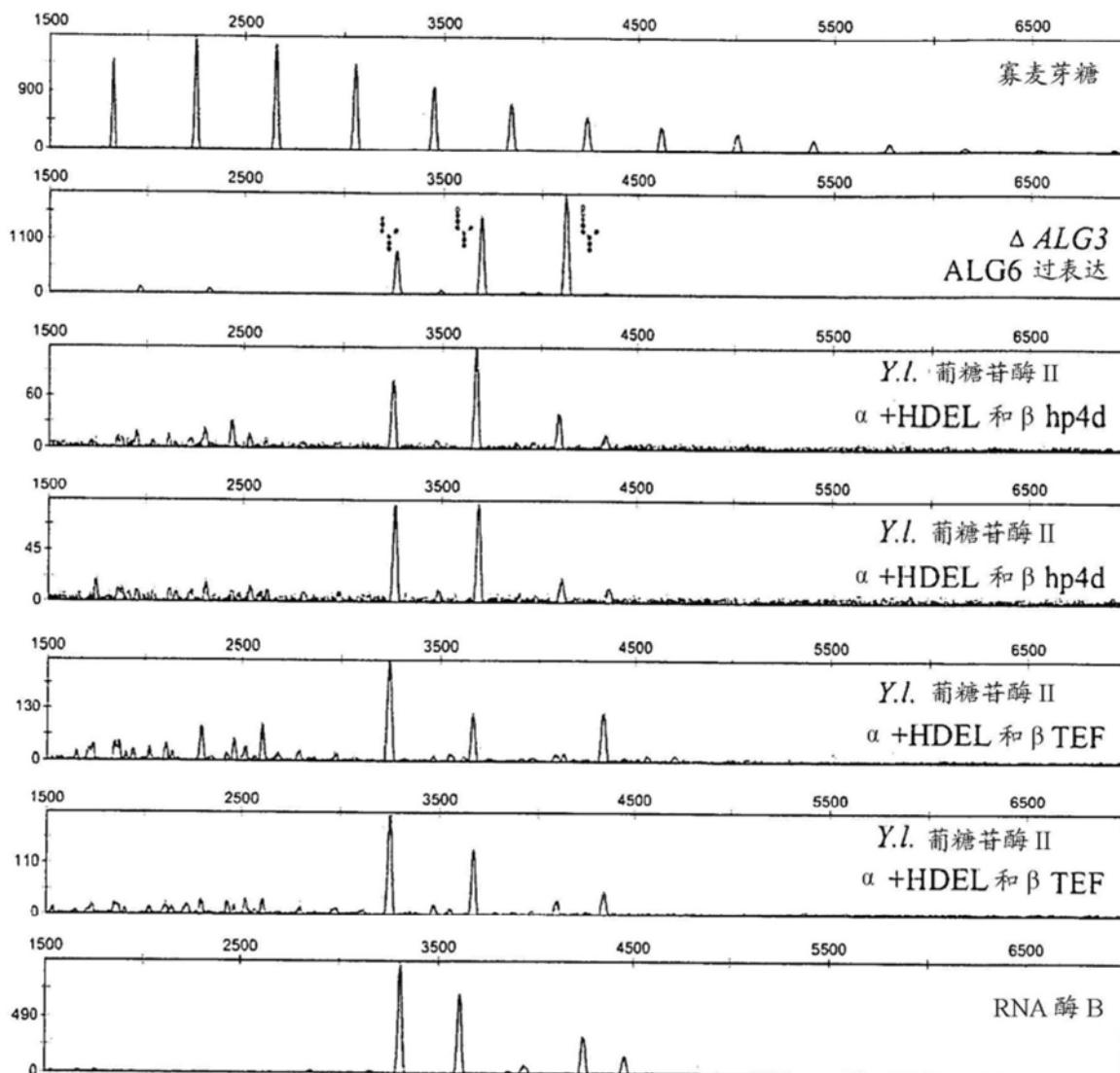


图34

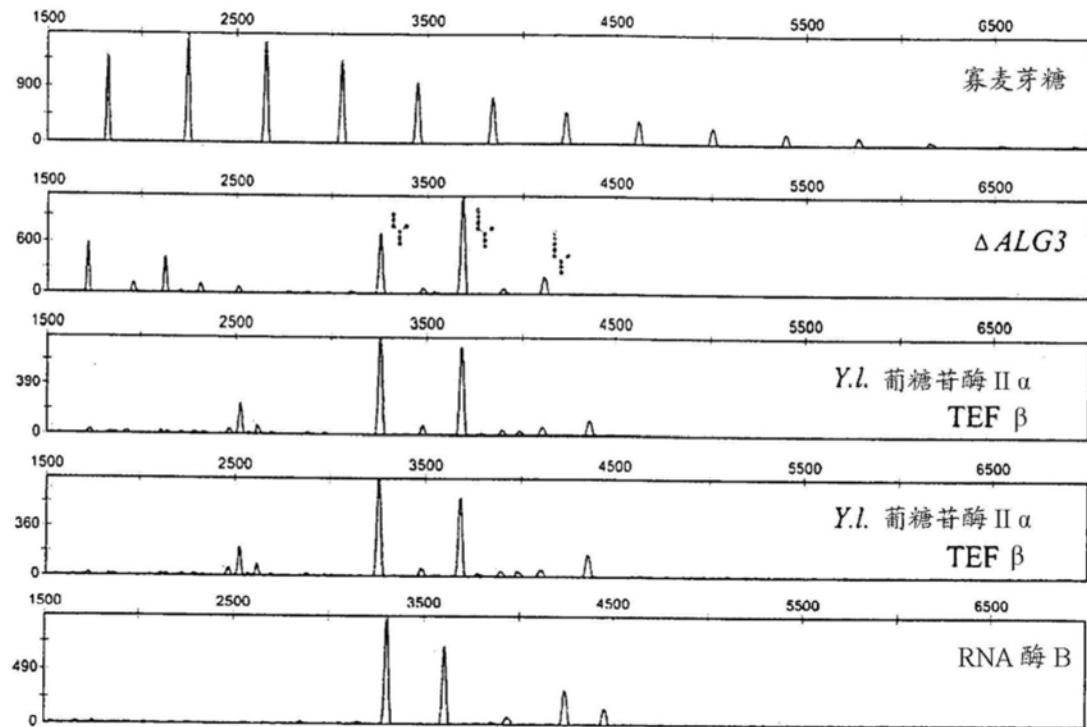


图35

gtaaaggacgagaactcaagaccgttctcagtcggctctgttaagcgaaaccgagccctgcgcacgcgcgcgtgcccagggtcttc  
ttgggcctccctacgagctggacictctatccaggcggccagctgcacggcaccatctgaagtctgtctccaaacga  
gaaggtgaagctgccccgtgggtcttgcggagctggccgctcgagtggtggacgaggagaagcgaatgaacggcgacat  
ccagctgcgacacgactctaaggcccggaaaggagcgataacaacgaggccgagaagtgggtctggcggccgtgatccgacacgc  
cccttctcgccgactcaagcgagatggccagaccacgtcagctgaacaacaaggctacatggagcactggcgaccca  
agggtggaggtggagggcgagggtagcagcagcacccaggaggacgagttacccgtggacgagtcttcggcgcaacaccgaca  
ccaagccccgaggccccgagtcgtggcctggacatcacccctccggctacaagcacgtgtcggcatccccgagcacgcgactcc  
ctgtctctgaaggagacccgaggcgccgaggcaaccacgaggccctaccgaatgtacaacgcccacgtgtcggactacgagctgc  
ctctccatgaccctgtacggcgccatccctcatgcaggcccaccgaaaggacttacccgtggcgttctggctgaacgcccggag  
acctggggacatcgtgaagtctacccatcccaacccctggccctggcgtggagccaccaccgacaccagactctactggtctc  
tgagtcgtggccagctggacgtgtcgttctggcccccaccccccaggagattctaagacctacggcagctgaccggctacccag  
ctgcccccagcacttcgcccattgcctaccaccagtgtcgtgaaactacccgacgaggacgtgaaggagggtggaccgaaacttcgac  
aagtaccagatccctacgacgtatccgtggctggacatcgagttacccgacgaccgaaagtacttccatggcccccgttctcc  
ccccatctatggaggagcagctggcagactgtgagcgaaagctggcgtgtcatcgaccccccacatcaagaaccaggacaactc  
tatcgtgcaggagatgaagtctaaggacccgtggccaccaagaacaaggacggcagattacgacggctgggttggccggcttctac  
tggatcgacacccatcaaccccgctggccatcaagtgggggtctctgttcaagttcgtgacaaggccaccctgtctaaactgttcatct  
ggaacgacatgaacgagccctgtgtcaacggccccggagaccaccatgcccaggacaaccctgcaccacggcaactggagcacc  
agacatccacaacgtgcacggcatccctggtaacgccacccatcgacccctgtggagcgaagaaggcgagatccgacgaccc  
ttcatctgacccgatcttactacgcccgtggccagcgaatgtgtccatgtggaccggcacaaccaggccaccctggagcaccctggcc  
gcctctatcccatggcgtgaacaacggaaatcgccgttcccttcgcccgtggccacgtggggggcttccagaacccctctaagg  
agctgtgacccgatggatcaggccggcatctgttccatccgtggccacccatcgacccgacgacccgacgagagccctac  
gatcgccgagcccccaccgatctatctcaggccatccgactgcgtatccgtggccaccatcgacccgacgacgagagccctac  
tctgtgaacggcatccatcgccgacccctggtaactacgcccacccctggacgaggccggcttcggccatcgacgaccagctgtac  
ggctctaccggcgtggccaaagccctgggtctgaggaggccaccaccggccgacatctacccgtggccatcgacgacgagaactacga

图36A

ctacttcgactacaccgtgtaccaggcgctggcaagcgacacaccgtccccatggagaccgtgcggccatgtcgtgcaggcg  
gccacgtatccccgaaaggaccgaccccgacgttctgcctgtatgcgtggaccctacaccctgggtggctggacaaga  
acggccaggccgacggctctgtacgtggacgacggcgagacctcgactacgagcgaggcgctacatccatcgacgattccgattcc  
aggagtctgcccggctccgaggacgtggcaccaagggcccaagaccggcagttaccaagaccatggccaacgtgcgagtgga  
gcgagtggtggctggaccccccaggagtggcaggcaagacccctgtgaccgtatcgaggacggcgctctgcgccttacccg  
cctccatgcagtagccatctcagcccacggcaagggccctacccgtggtaagaaccacgtggcatggcaagacccatggcg  
aatcgagttctaatacgcttag

图36B

agcgctggcgacgccttctgaccccgaggcggtggcccgagttcgccaagtttacaaggacaccaccacccatgttatctc  
accccgccatccagattccctctgcgtgaacgacgactactgtgactgtcccgacggctgtacgagccggcacctgcctgtgcc  
tttctgtctcgaaactctgcctgaccccggtgagcgacccggatctgacgacccgtgactgacccctgcctgccttactgtaa  
aacaaggccacaagccggctacgtgcctccagcgagttacatggcatctgtgactacgagctgttgtgacggatggatgag  
ggcgtgacccggcggaaaccaagtgtgaggacaagtgttaaggagatggcaaggagtggcggaaagaaggaggagaagcgacagaa  
gtctatgaccggccgtctgaagaagaaggacccgtggcggacgacagcagaaggaggcggaggacaacatcaacgc  
actggaggtggagatccaggcccaggagctgaaggtaacgcacccgtggcggagctggaggaggtggagcaggaggccctaa  
ggcgtcaaggcaagaccggcaaggtgaacgtgtggctggccctcgccaaactctgagttcaagggtggagtacaacccaa  
cttcaacgcacggcgtgaagcgagccgtgcacccgtggctggactacgcggccaaaggccaccctggaggcggccgtgaacaacgc  
ccaggaccgagacccgtggacgagatcgccaagccgacgacgagaaggccggcatcaactggagcagtggagaacgaggaggac  
ggcgtgaggctggctgggttaccagctggccgcctacctgccttctgtggagttcatcgaggcaagggtgtcgag  
gcctgctggaggacaacggcatccgtccaaaggccggagacctctacctgtttcaagggtgtctggaggccgaggccgt  
aagtctgtccgagaaggagctggcgacaaggcagaacgacgactgtgagttcaaggaccacaactgtgacccgtggctcta  
tctccgagccctgaagggcgtgttatctctaaggactctggcagttacacccgtggactgtttctggaccagccaagcagatcccc  
aagaaggcggaggcttacccgaatggcaagtacacccgtggcatcggtgtgtggacgtgtgaacgaggccggcggagatgt  
gcctgaggatcgagtgaccctgcagttacgcacggccagggtgtggaaacggaccctggatctaccaccgtgtaccctgt  
gcgaggaggacccatccgtgaaggtggccgaggacgagaagtgtgtactctatgcacgtgacccctccggcgtgtccggcgg  
acgaggccaccggccaccggccatccgttaatagccta

图37

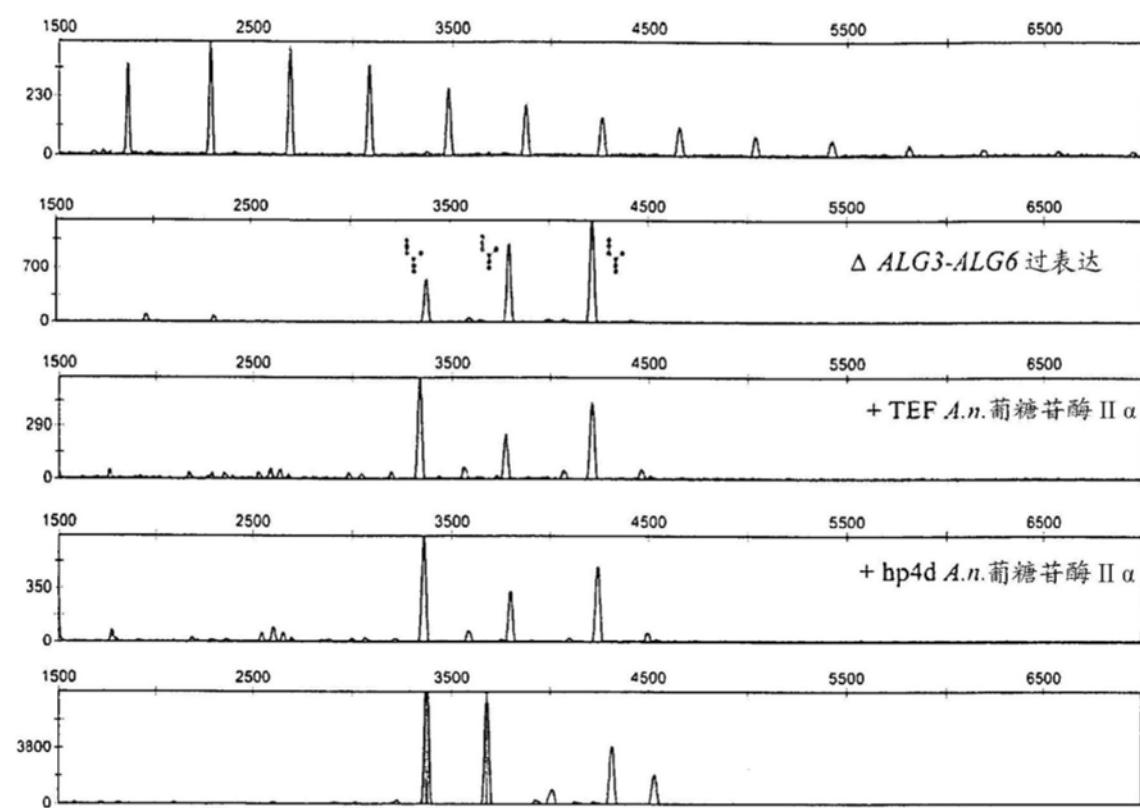


图38

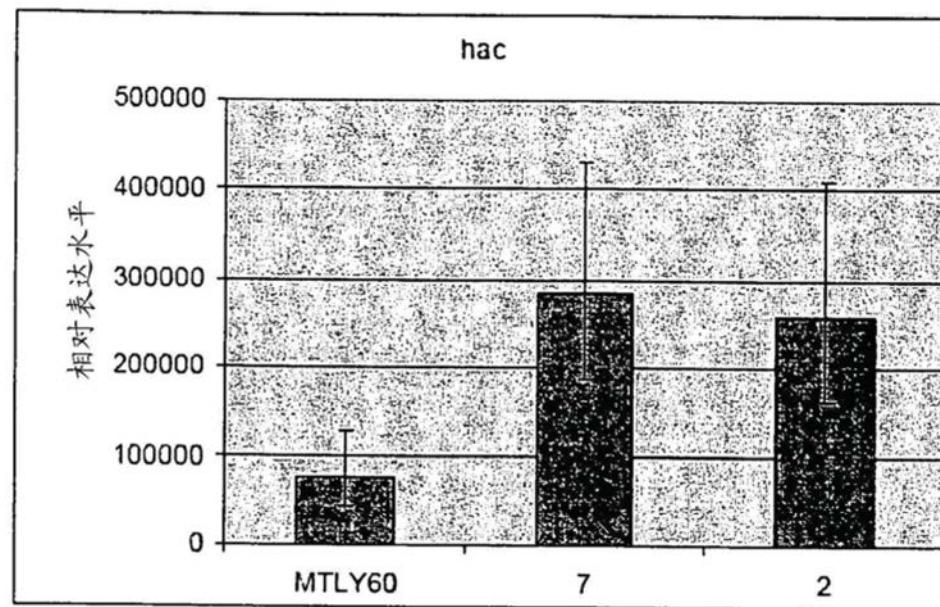


图39A

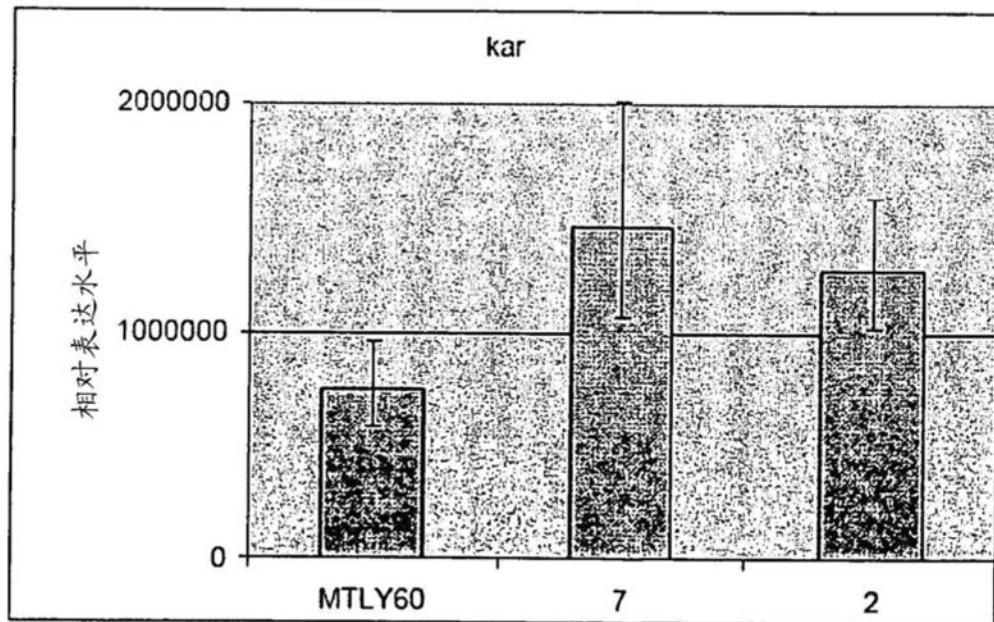


图39B

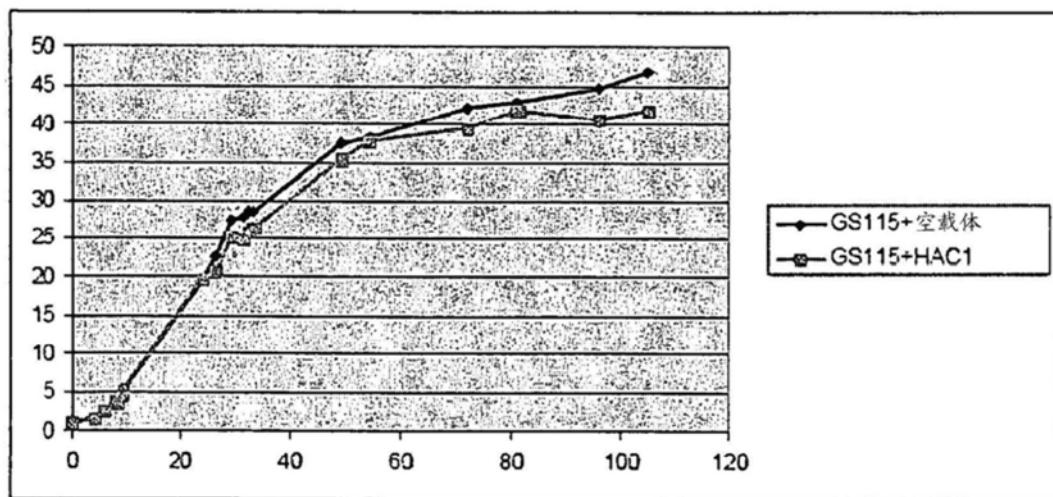


图40

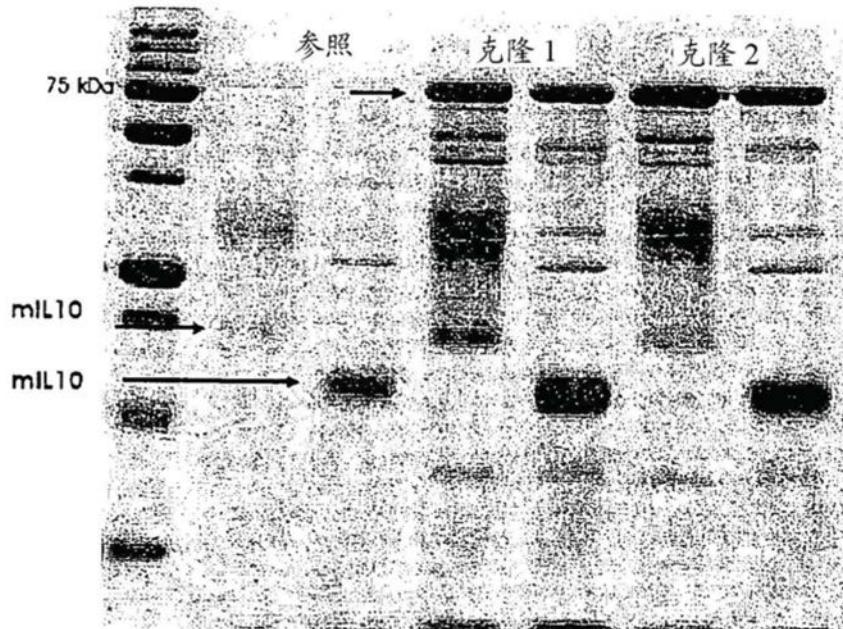


图41

图42

cggcggactgcgtccgaaccagctccagcagcggtttccggccattgagccgactgcgaccccgcacgtgtctggcccacgcact  
catgcatagttgggttgtggaggccacttttaagttagcacaaggcacctagctcgagcaaggltgtccgaaccaaagaagcggctgcagt  
ggtgcaaaacggggcggaaacggcggaaaaagccacggggcacgaattgaggcacgcctcgaattttagacgagtcacggcccc  
attcgcccgcgcaatggctcgccaacgcccgtcttgcaccacatcaggtaacccaaacccattgtgtaaaaagcttaacatatta  
taccgaacgttagttggcgggcttgcctgtccaggcaacattataagggtctgcatcgccggctcaattgaatcttttcttctt  
ctctctatattcatttgaattaaacacacatcaaca

图43

图44A

图44B

cagttacccctcgaaaaaagagttggtagcttgcgtatccggcaaaaccaccgctggtagcggcggttttgtcaagcagcagattac  
gcccggaaaaaggatctcaagaagatccttgatctttctactgaacggtagccccaccggaaattggccgtcgaaaaccggc  
ttcagggtggAACAGGACACCTCCCTGACTTCTGGTATATCAGTAGGTGATTCTAGTGGGTTTCATAATAATTACTAAGCGC  
AGGCAACATTCACTCGGCTAAACGCAAAACGGACCAGCTGTGATATCTCTGACGCATTGACCACCGAGAAAATAGTGTAGTACCGG  
TTATGTCCTACACAGGCGACGCCATGCTAGGTGATGTTACTAATCAGATTCTACACTACCTACCCATCCCTGGTACGCACAAAGC  
ACTTGCTAGATAGAGTCGACAAAGGCGGGCCCCCTCGAGATTACCCGTATCCCTACATAACTCGTATAGCATAACATTACGAAGT  
TAATTCCGAGAAACACAACATGCCCATGGACAGACCATGCGGATACAGGTTGCGAGTACCCATAACTCGGATCAGACAGGTC  
GTCTGACCATCATAAGCTGAACAGCGCTCCATACTTGACGCTCTATATACACAGTTAAATTACATATCCATAGTCACCTCAACAG  
TAATCTCTGGTAAGCCTCCAGCCAGCCTCTGGTATCGCTGGCTCTCAATAGGACTCGGTTGGCCGTACAGACCTCGGCCGACA  
ATTGATATCCGTCCGGTAGACATGACATCCTCAACAGTCGGTACTGCTGCCAGAGCGCTCCCTGTCGTCAAGACCCACCCGGGG  
GCAATAAGCCAGTCCTCAGAGTCGCCCTAGGTGGTCTGGCAATGAAGCCAACCACAAACTGGGGTGGATGGCAAGCTCAATGG  
TCTGCTGGAGTACTCGCCAGTGGCCAGAGAGGCCCTGCAAGACAGCTGGCCAGCATGAGCAGACCTCTGCCAGCTCTCGTGG  
GGGGACTAGGAACTCTGTACTGGGAGTCTCTGCTAGTCAGAGACGTCCTCCCTCTGCTAGAGACAGTTCTCGGCCAGCTCG  
GCCAGCAATGATCCGGTCCGGGLACACCGTGGCGTGGTGGTGGTGGAGTATCGGACCCTCGCGATTGGTGGACACCCGGTACT  
GGTGTGGCAATATCTGCAACTTCTGCTCTGCAACAGGAAGAAACCGTGTCTAAGAGCAAGTCTGAGGGGGAGCACAGTGG  
GGTGAAGTCGTCAATGATGTCGATATGGGTCTGATCATGCACACATAAGGTCCGACCTTATCGGCAAGCTCAATGAGC  
ACATCCAGAGAAGCACACAGGTGGTTCTGGCTGCCACGAGCTTGAGCAGCTCGAGCGGCAAAAGCGGGACTTGGACGTT  
CTCGTAGGAGGGCATTGGTGGTGAAGAGGAGACTGAAATAATTAGTCAGAACCTTATCGGAAC

图44C

图45A

图45B

aaggatctcaagaagatccttgcactttcttactgaacggatccccaccggaaattcgccgcttcggaaaccgcgtcagggtggaa  
caggacaccccttgcacttcttgatatactcgtataggctgtattcatagtggtttcataataaattactaacggcaggcaacattc  
actcggtttaaacgcggaaaacggaccgtcttgcattctgtacgcattgaccaccgagaaatagtgttagttaccgggtgagttattgtcttcta  
cacaggcgacgcccattgttagtgtactaactcagatttactacccatccctggtaacgcacaaagcacttgcata  
gtcgacaaaggcgccccccctcgagattaccctgttatccctacataactcgtatagcatacattatacgaagttattcgaattccggcca  
gagagccattgacgttcttctaatttgaccgatagccgtatagccgtatctataagttcaactaactcgtacttaccataacataact  
tcactgccccagataagggtccgataaaaaagtctgcagactaaattttcgttcttccaccaccaaattcccttacgaagctg  
gctaacgtccacaagtccgccttgcgcgtcgagtgctcaagctcggtggcagccaagaaaaccacccgtgtgcctctggatgttaccacc  
accaaggagctattgagcttgcgataaggctggaccctatgtgtgcattgtcaagacccatatcgacatcattgacgacttacccatcgcc  
ggcactgtgtccccctcaaggaacttgccttaagcacggtttccctgttcgaggacagaaattcgcagatattggcaacactgtcaagc  
accagtaccgggtcaccgaatcgccgagtggtccgatattcacaacggccacgggttacccggaaaccggaaatcatgtggcctgcgag  
ctggtgccgaggaaactgtcttgcatacagaagaaggaggacgtctctgactacgagaactcccgatattcacaagggttccctgt  
caacgagaagctggccagaggctgtcatgtggccgagctgtcatgtggccacttggcgagttactccaagcagaccatt  
gagcttgtcccgatccgaccccgagtttggttgttgcattgtccagaaccgacctaaggcgacttgcaggacttgcattattctgaccc  
cggtgggttgttgcacgacaaggagacgtctcgacagcagtgacttgcaggatgtcatgttaccggaaacggatataatt  
gtcgccgaggctgtacggccagaaccgagatccattgaggaggccaagcgataccagaaggctggctggaggcttaccagaagat  
taactgttaggttagactatggataatgttaactgtgtatatactggagcgtcaagtatggagcgtcgttgcacgttgcattgt  
gacccgtctgtatcgagttatgttgcacaaccctgtgtatcccatggctgtccatggcatgttgcacgttgcattgt  
acttcgtata

图45C

图46A

gaatctgacgtgcaggtcaggccaacggcgcaacaagttcggtcaacaccgaggcccaccccttatccatcgatcttctcgacg  
 aggccgcccacccatgacgagctgtatacgcttagggtggtatctaagctatttatcactttacaacttctacactatct  
 actttaataaaatgaatatcgttattctctatgattactgtatgcgttcctctaagacaatcgtaacactcgctctggagagtta  
 gtcatccgacaggtaacttaatcccaacacccattactctcgtaactgtactcttcgtccacgtcatcaatccctgtca  
 tcatctgctggattgttgtatcgltggctctaatacalttatgtttatggccaaacaacttcattgcacgtaagtgaattgtttataacagcgt  
 tcgccaattgtcgccatcgltccggctgtccctaccgttaggttagtgcgtcacactaccgaggtaactagagtggaaagcgatac  
 tgctcgacacaccacccatggcttacgactgcagagagaatcggttacccctcacaaggccctcagtgcggccgccccgggtggc  
 gaagaactccacgatgagatccccgcgtggaggatcatccagccgggtcccgaaaacgattccgaagcccaacccatagaagg  
 cggcggtggaaatcgaaatctcgatggcagggtggcgctcgcttgcgttgcatttcgaaccccagactccgcataagaactcgta  
 agaaggcgatagaaggcgatgcgtcgaaatcggtatgcggccacaccgcggccacagtcgtatgcgtatccagaaaagcgg  
 ccatttcaccatgatattcgcaagcaggcatcgccatggtcacgacgagatccgcggcatgcgcgccttgagccctggcga  
 acagttcggtggcgagccctgtatgcgtcccgatcatctgtatgcgacaaagccggcttccatccgactacgtgcgtcgatg  
 cgatgttcgttgcgttgcgtatggcaggtagccggatcaagcgatgcggccgcattgcgtatcagccatgtatggataacttcggc  
 aggagcaaggtgagatgacaggagatccgtggccacttcgccaatagcagccactcccccgttcgttgcgtatgcgtccgc  
 agctgcgtcgaaacgcggccgtggccagccacgatagccgcgtgcctcgctgcgttgcgttgcgttgcgttgcgttgc  
 gacaaaaagaacccggccgtggccatcgatccgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc  
 aatagccctccaccaagccggccggagaacctcgatgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc  
 ttgtatccctgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc  
 ggcaatccgggttcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc  
 gtttccctgtcccgatagcccgatgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc  
 gcccttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc  
 ggcgttccgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc  
 ttatccacagaatcagggataacgcagaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgc  
 tggcggtttccataggctccggcccccgtacgagcatcacaatcgacgctcaagtgcggcggaaacccgacaggactataag

图46B

ataccaggcggttccccctggaaagccctcgctcgctccgttccgaccctgccgttaccggataacctgcgccttcgccttcggga  
agcgtggcgcttctcatagctcacgctgttaggtatctcagttcggttaggtcgctccaaagctggctgtgcacgaaccccccttc  
agcccggccgtgcgccttatccgtaactatcgttgcgttccaaacccggtaagacacgacttacgcactggcagcagccactggta  
caggattagcagagcgaggtatgttaggcgggtcacagagtt  
tctcgctctgttgcgttccaaacccggtaactatcgttgcgttccaaacccggtaactacggctacactagaaggacagtttttttt  
tgcaaggcagcagaltacgcgcagaaaaaaaaggatctcaagaagatccttgcgttccaaacccggtaactacggctacactagaagg  
gctgtcgggaaccgcgttcagggttgcgttccaaacccggtaactatcgttgcgttccaaacccggtaactacggctacactagaagg  
aataaaatttactaacggcaggcaacattcactcggcttacggccaaacggaccgttgcgttccaaacccggtaactacggctacactaga  
tgttagttaccgggttagtt  
ggtagcgcacaaaggacttgcgttccaaacccggtaactacggcttacggccaaacggaccgttgcgttccaaacccggtaactacgg  
cattatacgaagtatctgaattccgcggccagagagccattgacgttgcgttccaaacggaccgttgcgttccaaacccggtaactacgg  
actaactcgtaactattaccataacatataacttcacttgcgttccaaacggccatgttgcgttccaaacccggtaactacggcttac  
ccacccaaatgcgccttacgaagctcgagcttgcgttccaaacggaccgttgcgttccaaacccggtaactacggcttac  
caacccgtgtgcgttgcgttccaaacggaccgttgcgttccaaacggaccgttgcgttccaaacccggtaactacggcttac  
atcgacatcattgacgacttgcgttccaaacggaccgttgcgttccaaacggaccgttgcgttccaaacggaccgttgcgttccaaac  
agttgcgttccaaacggaccgttgcgttccaaacggaccgttgcgttccaaacggaccgttgcgttccaaacggaccgttgcgttcc  
ccactggcgacttccaaacggaccgttgcgttccaaacggaccgttgcgttccaaacggaccgttgcgttccaaacggaccgttgc  
cgactctgaggactggcttattctgacggaccgttgcgttccaaacggaccgttgcgttccaaacggaccgttgcgttccaaacgg  
gtcatgttccaaacggaccgttgcgttccaaacggaccgttgcgttccaaacggaccgttgcgttccaaacggaccgttgcgttcc  
aggctggctgggaggctaccagaagatgttgcgttccaaacggaccgttgcgttccaaacggaccgttgcgttccaaacggacc  
agcgtgttgcgttccaaacggaccgttgcgttccaaacggaccgttgcgttccaaacggaccgttgcgttccaaacggacc  
aagacgaacaagtgttgcgttccaaacggaccgttgcgttccaaacggaccgttgcgttccaaacggaccgttgcgttccaaac

图46C

图47A

图47B

aacggcaggcaacattcactcggttaaacgc当地有大量地名、人名、物名等，这些地名、人名、物名等在文中多次出现，且具有一定的规律性，如“a”、“t”、“c”、“g”等字母的出现频率较高，且分布较为均匀。整个文本由大量的字母组成，没有明显的语法结构或语义意义。

图47C

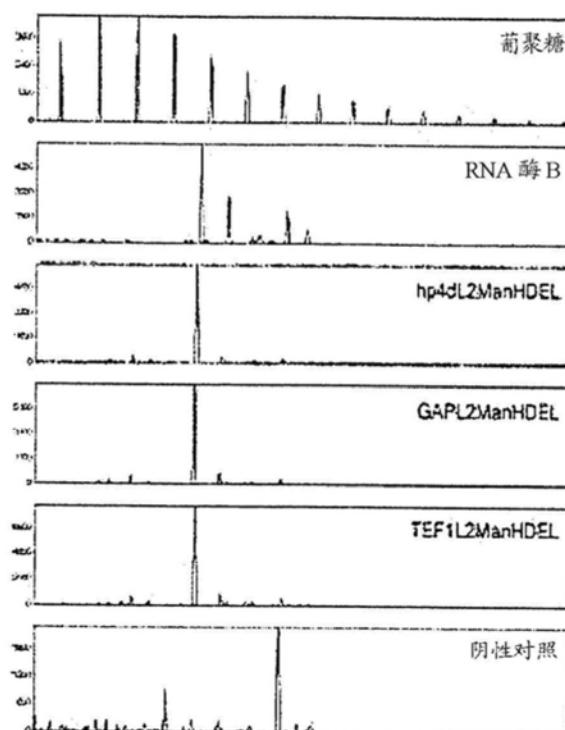


图48

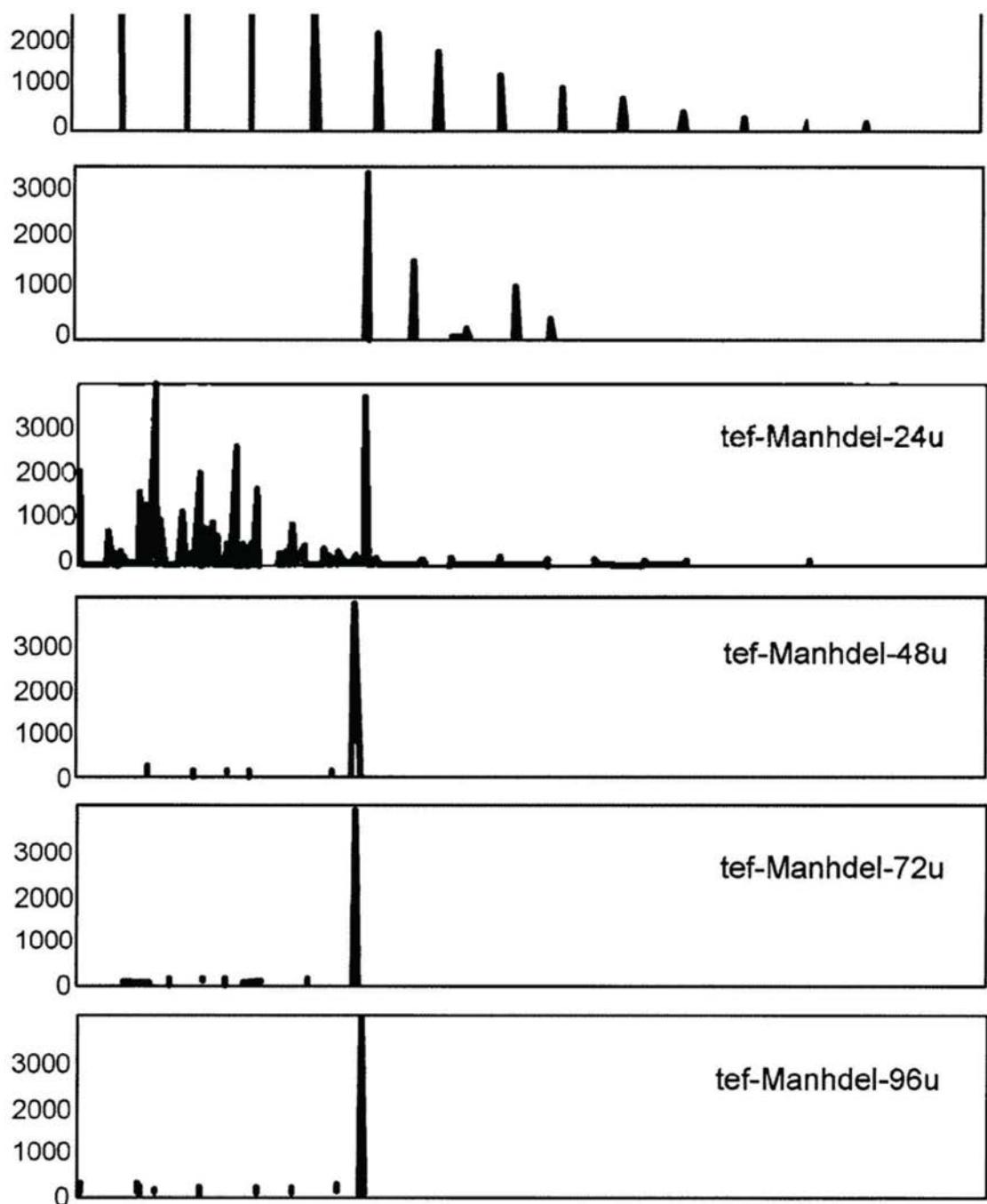


图49

gagctcagcgctcgaccctgtatccccaaagtcttcggctactcttctgtgggtgtgtgtgtaacgccacctacttgactcggtcgacccccc  
 caccctccccggccctgggcacccctctcgatacggatctacccgatctggccgacgaatggagctgtctatgggccccatccaggccaac  
 cacaccggcaccggcctgctgaccctgcagccgagcagaagttccagaaggtaagggcttcggggcatggctacaacatcatcc  
 ctggccctgaacatcctggccctgtctcccccgtcagaacccctgtctgaagttacttctctgaggagggcatggctacaacatcatcc  
 gagtgccatggcccttgacttctatccgaacctacacccctacccgacaccccccacgacttccagctgcacaacttctccctggcc  
 aggaggacaccaagctgaagatccccctgatccaccggagccctgcagctggccagcaccggctgtctgcggccctccctggaccc  
 ctcccacctggctcaagaccaacggccgtgaacggcaagggtctctgaagggccagccggcgcacatctaccaccagacctggcc  
 ccgatactcgtgaagttccctggacgcctacgcccggcacaagctgcagttctggccgtgaccggcggagaacggccctctgtggact  
 gctgtccggctaccccttcaggctgtctggcttcacccggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcgg  
 ctctacccaccacaacgtgcactgtctgtatgtggacgaccaggcactgtctgtccctactggccaaagggtggctgtgaccggaccc  
 agggccgg  
 actgtccgg  
 caactgggtgcaactcgactctccatcatccaacctgttaccacgtggggctggggactggaaacctggccatcgccctgatgcaccc  
 acttctctaagttcatccccggggctctcagcgactggggctggggctctcagaagaacgacctggacggccgtggccctgatgcaccc  
 cgacggctctggctggggctggggctgtgaccatctaaaggacgtggccctgaccatcaaggacccggcgtggccctggagacc  
 atctctccggctactctatccacacccatgtggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcgg  
 accatggggatggggatggggatggggatggggatggggatggggatggggatggggatggggatggggatggggatggggatgggg

图50

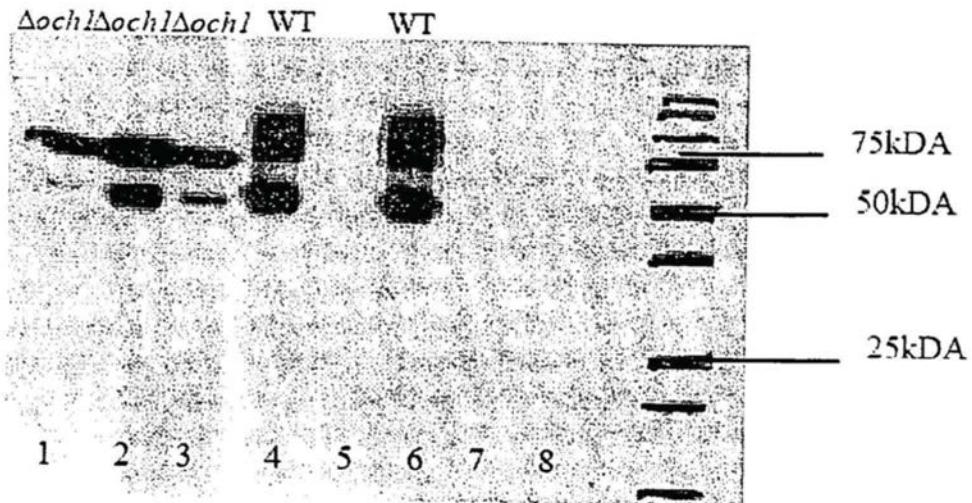


图51

gctcccccccactgtgactctcgagtgtggAACGATAACCTGTGGAAAGCCAAGGAAGGCCGAGAACATCACCACCGCTGTGCC  
gagcacgttcgtgaacgagaacattaccgtgcccgcacaccaaggtaacttctacgcctgtggAAAGCAGCAGCACCC  
gtggagggtgtggcaggactggctctgtgtggccgtgtcgaggacaggcttcctggtaacttctcagccctgggagcccc  
tgcaacgtggacaaggccgtgtggccgtgcataccatcaccggcggacaccctccgaaagctgttccgagtgactctaacttctcggaggc  
ccccgacgccgcctgtgtccccctgcgaaccatcaccggcggacaccctccgaaagctgttccgagtgactctaacttctcggaggc  
aagctgaagctgtacaccggcgaggctgtcgaaaccggcgaccg

图52

ctggacaacggccctggcccaaccccccaccaatgggctggcgtgcacgtggaggcattcatgtgtaaacctggactgtcaggaagagccccga  
ctcttgtatctctgagaagctgtcatggaaatggcccgagctgtatgggtgtcgagggtgtggaaaggacgcggctacgagtacatgttatcg  
acgactgtggatggccccccagcgagactctgagggccgactccaggccgaccccccagcgattccccacggcatccgacagctgc  
caactacgtcactctaaggccctgaagctggcattacggcgtggcaacaagaccgtgtccggctttccggctttcgctac  
tagacatcgacgcccagaccctcgccgactggggctggacactgtgaagttcgacggctgtactgtactctctcgagaacctggccg  
acggctacaagcacatgtctggccctgaaccgaaccggccgatctatcgactcttgtgagtgccctgtacatgtggcccttcaga  
agcccaactacaccgagatccgacagactgtaaaccactggcggacatcgacgacgtggaaagtctatcaagtctattctg  
gacgtggaccctttcaaccaggagcgaatcgacgtcgccggaccggatggaaacgaccggacatgtgggtatcgccacttc  
ggccgtgtggaaaccaggcaggtgaccctggccctgtggctatcatggctggccctgttcatgtctacgaccgtgcgacacatct  
ccccaggccaaggccctgctccaggacaaggacgtatcgccatcaaccaggacccctggcaagcaggctaccagctccgacag  
ggcggacaacttcgagggtgtggagcgcacccctgtctggctggccctggccatgtcaaccgacaggagatggcggaccc  
gatcttacaccatcgccgtggccctggaaagggcgtggctgttaaccggccctgttcatcaccctggccctgtgaagcggaaa  
gctgggattctacgagtgacactcgactgtcgatctcacatcaacccaccggcaccgtgtccagctcgagaacaccatcgagatg  
tctctgtgaaggacactgt

图53

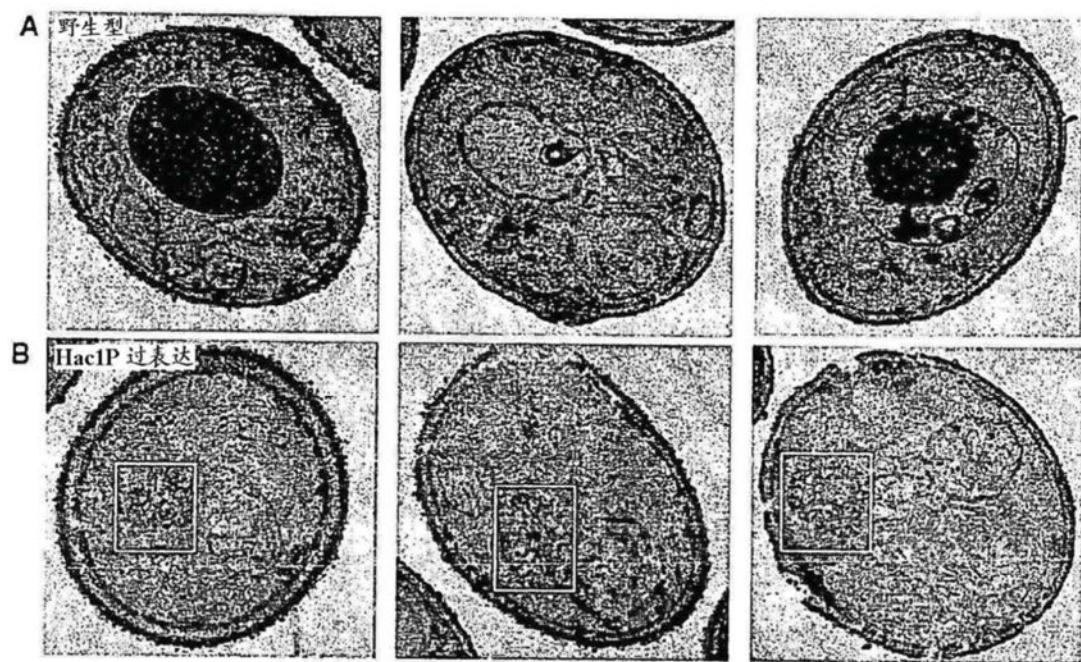


图54

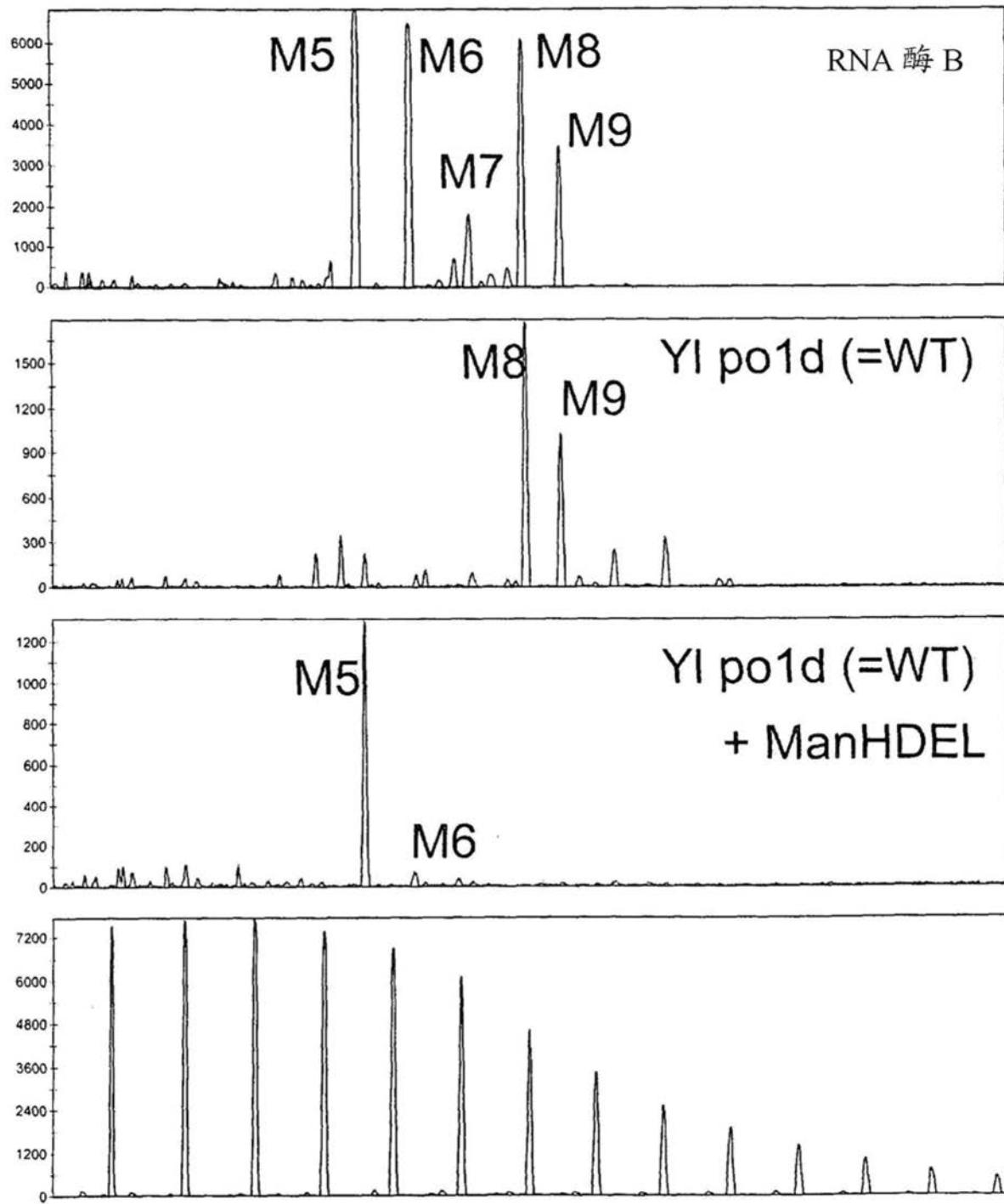


图55