



(12) 发明专利

(10) 授权公告号 CN 1950496 B

(45) 授权公告日 2015.02.04

(21) 申请号 200580011116.4

C07K 16/00(2006.01)

(22) 申请日 2005.04.15

(56) 对比文件

(30) 优先权数据

60/562,424 2004.04.15 US

US 6528286 B1, 2003.03.04, 全文.

WO 02080961 A1, 2002.10.17, 全文.

CN 1379811 A, 2002.11.13, 全文.

CN 1331290 A, 2002.01.16, 全文.

(85) PCT国际申请进入国家阶段日

2006.10.13

审查员 董桂灵

(86) PCT国际申请的申请数据

PCT/IB2005/051249 2005.04.15

(87) PCT国际申请的公布数据

W02005/100584 EN 2005.10.27

(73) 专利权人 格利科菲公司

地址 美国新罕布什尔州

(72) 发明人 R·戴维森 T·格恩格洛斯

S·维尔德特 B·-K·蔡 J·内特

P·博布罗维茨 S·哈米尔顿

(74) 专利代理机构 中国专利代理(香港)有限公

司 72001

代理人 权陆军 梁谋

(51) Int. Cl.

C12N 1/19(2006.01)

C12P 21/00(2006.01)

权利要求书2页 说明书39页

序列表33页 附图17页

(54) 发明名称

在低等真核生物中产生半乳糖基化糖蛋白

(57) 摘要

本发明提供了产生人样糖蛋白的新型低等真核生物宿主细胞,所述糖蛋白的特征在于含有末端 β -半乳糖残基并基本上缺少岩藻糖和唾液酸残基。本发明还提供了催化从UDP-半乳糖将半乳糖残基转移至重组低等真核生物宿主细胞的受体底物上的方法,所述底物可被用作治疗性糖蛋白。

1. 一种产生人样糖蛋白的重组酵母宿主细胞,所述宿主细胞被基因工程改造以生产包含具有末端 GlcNAc 残基的 N-连接的寡糖的糖蛋白,所述宿主细胞包含编码融合蛋白的核酸,所述融合蛋白具有在宿主细胞中将 β -半乳糖残基转移至糖蛋白的 N-聚糖的 N-连接的寡糖分支的末端 GlcNAc 残基上的 β 1,4-半乳糖基转移酶活性,且进一步包含编码 UDP-半乳糖转运活性和/或 UDP-半乳糖 C4 差向异构酶活性的核酸,其中所述 N-连接的寡糖分支的特征在于含有末端 β -半乳糖残基并在糖蛋白上缺少岩藻糖和唾液酸残基。

2. 如权利要求 1 所述的宿主细胞,其中所述宿主细胞表达 UDP-半乳糖转运活性。

3. 如权利要求 1 所述的宿主细胞,其中所述宿主细胞表现出升高的 UDP-半乳糖水平。

4. 如上述权利要求任一项所述的宿主细胞,其中所述半乳糖基转移酶活性选自人 GalTI、GalTII、GalTIII、GalTIV、GalTV、GalTVI、GalTVII、牛 GalTI、非洲爪蟾 GalT 和秀丽隐杆线虫 GalTII。

5. 如权利要求 1 所述的宿主细胞,其中所述 UDP-半乳糖转运活性由选自:SpUGT、hUGT1、hUGT2 和 DmUGT 的基因编码。

6. 如权利要求 1 所述的宿主细胞,其中所述 UDP-半乳糖 C4 差向异构酶活性由选自:SpGALE、ScGAL10 和 hGALE 的基因编码。

7. 如权利要求 1 所述的宿主细胞,其中所述 N-连接的寡糖分支选自三甘露糖核心上的 GlcNAc β 1,2-Man α 1,3、GlcNAc β 1,4-Man α 1,3、GlcNAc β 1,2-Man α 1,6、GlcNAc β 1,4-Man α 1,6 和 GlcNAc β 1,6-Man α 1,6。

8. 如权利要求 1 所述的宿主细胞,其中所述宿主细胞产生作为唾液酸的受体底物的糖蛋白。

9. 如权利要求 1-3 和 5-8 任一项的所述的宿主,其中所述宿主细胞在起始关于糖蛋白上聚糖的 1,6 甘露糖基转移酶活性上受损。

10. 如权利要求 1-3 和 5-8 任一项的所述的宿主,其中所述宿主细胞在 Do1-P-Man:Man₅GlcNAc₂-PP-Do1 甘露糖基转移酶活性上被减小或排除。

11. 如权利要求 1-3 和 5-8 任一项的所述的宿主,其中所述宿主细胞表达选自 α -1,2-甘露糖苷酶 I 活性和甘露糖苷酶 II 活性的甘露糖苷酶活性。

12. 如权利要求 1-3 和 5-8 任一项的所述的宿主,其中所述宿主细胞表达选自 GnTI、GnTII、GnTIII、GnTIV、GnTV、GnTVI 和 GnTIX 的 GnT 活性。

13. 如权利要求 1-3 和 5-8 任一项的所述的宿主,其中所述宿主细胞选自毕赤氏酵母属物种、糖酵母属物种和克鲁维氏酵母属物种。

14. 如权利要求 1-3 和 5-8 任一项的所述的宿主,其中所述宿主细胞选自巴斯德毕赤氏酵母、*Pichia finlandica*、喜海藻糖毕赤氏酵母、*Pichia koclamae*、膜醭毕赤氏酵母、*Pichia minuta*、*Pichia opuntiae*、*Pichia thermotolerans*、柳毕赤氏酵母、*Pichia guercuum*、皮杰普氏毕赤氏酵母、具柄毕赤氏酵母、甲醇毕赤氏酵母、啤酒糖酵母、多形汉逊氏酵母、乳克鲁维氏酵母和白色假丝酵母。

15. 如权利要求 1-3 和 5-8 任一项的所述的宿主,其中所述宿主细胞是巴斯德毕赤氏酵母。

16. 如权利要求 4 所述的宿主,其中所述宿主细胞在起始关于糖蛋白上聚糖的 1,6 甘露糖基转移酶活性上受损。

17. 如权利要求 4 所述的宿主,其中所述宿主细胞在 Do1-P-Man :Man₅GlcNAc₂-PP-Do1 甘露糖基转移酶活性上被减小或排除。

18. 如权利要求 4 所述的宿主,其中所述宿主细胞表达选自 α -1,2-甘露糖苷酶 I 活性和甘露糖苷酶 II 活性的甘露糖苷酶活性。

19. 如权利要求 4 所述的宿主,其中所述宿主细胞表达选自 GnTI、GnTII、GnTIII、GnTIV、GnTV、GnTVI 和 GnTIX 的 GnT 活性。

20. 如权利要求 4 所述的宿主,其中所述宿主细胞选自毕赤氏酵母属物种、糖酵母属物种和克鲁维氏酵母属物种。

21. 如权利要求 4 所述的宿主,其中所述宿主细胞选自巴斯德毕赤氏酵母、*Pichia finlandica*、喜海藻糖毕赤氏酵母、*Pichia koclamae*、膜醭毕赤氏酵母、*Pichia minuta*、*Pichia opuntiae*、*Pichia thermotolerans*、柳毕赤氏酵母、*Pichiaguercuum*、皮杰普氏毕赤氏酵母、具柄毕赤氏酵母、甲醇毕赤氏酵母、啤酒糖酵母、多形汉逊氏酵母、乳克鲁维氏酵母和白色假丝酵母。

22. 如权利要求 4 所述的宿主,其中所述宿主细胞是巴斯德毕赤氏酵母。

23. 一种在如上述权利要求任一项的酵母宿主细胞中产生人样糖蛋白组合物的方法,所述方法包括将半乳糖残基在不存在岩藻糖和唾液酸残基时转移至杂合或复合糖蛋白上的步骤。

24. 如权利要求 23 所述的方法,其中所述半乳糖残基被转移至选自 GlcNAcMan₃GlcNAc₂、GlcNAc₂Man₃GlcNAc₂、GlcNAc₃Man₃GlcNAc₂、GlcNAc₄Man₃GlcNAc₂、GlcNAc₅Man₃GlcNAc₂、GlcNAc₆Man₃GlcNAc₂、GlcNAcMan₄GlcNAc₂、GlcNAcMan₅GlcNAc₂、GlcNAc₂Man₅GlcNAc₂ 和 GlcNAc₃Man₅GlcNAc₂ 的糖蛋白中的寡糖上。

25. 如权利要求 23 或 24 所述的方法,其中产生至少 33%半乳糖基化糖蛋白组合物。

26. 如权利要求 23 或 24 所述的方法,其中产生至少 60%半乳糖基化糖蛋白组合物。

27. 如权利要求 23 或 24 所述的方法,其中产生至少 90%半乳糖基化糖蛋白组合物。

28. 如权利要求 23-27 任一项的方法所产生的糖蛋白组合物。

29. 如权利要求 28 的糖蛋白组合物,其中所述糖蛋白组合物是唾液酸的受体底物。

30. 如权利要求 28 的糖蛋白组合物,其中所述糖蛋白选自:促红细胞生成素,细胞因子,凝固因子、抗凝血酶 III、凝血酶、可溶性 IgE 受体 α -链、IgG、IgG 片段、IgG 融合体、IgM、白细胞介素、尿激酶、糜蛋白酶,和胰凝乳蛋白酶抑制剂、IGF- 结合蛋白、表皮生长因子、生长激素释放因子、膜联蛋白 V 融合蛋白、制管张素、血管内皮生长因子-2、骨髓祖先抑制因子-1、骨保护素、 α -1-抗胰蛋白酶、甲胎蛋白、DNA 酶 II、人纤溶酶原 kringle 3、葡糖脑苷脂酶、TNF 结合蛋白 1、促卵泡激素、细胞毒性 T 淋巴细胞相关抗原 4-Ig、跨膜激活剂及钙调节剂及亲环蛋白配体、可溶性 TNF 受体 Fc 融合体、胰高血糖素样蛋白 1、IL-2 受体激动剂。

31. 如权利要求 30 的糖蛋白组合物,其中所述细胞因子选自干扰素- α 、干扰素- β 、干扰素- γ 、干扰素- ω 、粒细胞-CSF 和 GM-CSF,且所述凝固因子选自因子 VIII、因子 IX 和人 C 蛋白。

在低等真核生物中产生半乳糖基化糖蛋白

[0001] 相关申请的交叉参考

[0002] 本申请是美国申请号 10/371,877(2003 年 2 月 20 日提交)的部分接续申请,美国申请号 10/371,877 是美国申请号 09/892,591(2001 年 6 月 27 日提交)的部分接续申请,美国申请号 09/892,591 根据 35U.S.C. § 119(e) 要求美国临时申请号 60/214,358(2000 年 6 月 28 日提交)、美国临时申请号 60/215,638(2000 年 6 月 30 日提交)和美国临时申请号 60/279,997(2001 年 3 月 30 日提交)的利益,上述各申请被全文引入此处以供参考。本申请还是 PCT/US02/41510(2002 年 12 月 24 日提交)的部分接续申请,PCT/US02/41510 要求美国临时申请号 60/344,169(2001 年 12 月 27 日提交)的利益,上述各申请被全文引入此处以供参考。本申请还要求美国临时申请号 60/562,424(2004 年 4 月 15 日提交)的优先权,所述申请被全文引入此处以供参考。

发明领域

[0003] 本发明涉及在低等真核生物中的蛋白质糖基化工程领域,特别地涉及产生含有末端半乳糖残基的糖蛋白。本发明进一步涉及新型宿主细胞,所述细胞包含酶编码基因,所述酶参与聚糖的半乳糖基转移以及尤其可用作治疗剂的糖蛋白的产生。

[0004] 发明背景

[0005] 酵母和丝状真菌均已被成功地用于生产重组蛋白质,包括细胞内和分泌型的(Cereghino, J. L. 和 J. M. Cregg 2000 FEMS Microbiology Reviews 24(1):45-66; Harkki, A. 等, 1989 Bio-Technology 7(6):596; Berka, R. M. 等, 1992 Abstr. Papers Amer. Chem. Soc. 203:121-B10T; Svetina, M. 等, 2000 J. Biotechnol. 76(2-3):245-251)。多种酵母,诸如乳克鲁维氏酵母(*K. Lactis*)、巴斯德毕赤氏酵母(*Pichia pastoris*)、甲醇毕赤氏酵母(*Pichia methanolica*)和多形汉逊氏酵母(*Hansenula polymorpha*)作为真核生物表达系统已经有特别重要的作用,因为它们能够生长至高细胞密度并分泌大量的重组蛋白。同样,丝状真菌,诸如黑色曲霉(*Aspergillus niger*)、镰孢属物种(*Fusarium* sp)、粗糙链孢霉(*Neurospora crassa*)以及其它真菌已被用于以工业规模高效生产糖蛋白。但是,在任何这些真核生物微生物中表达的糖蛋白与动物中的糖蛋白在 N-聚糖结构上存在相当大差异。这已妨碍了酵母或丝状真菌作为宿主在制备糖基化治疗性蛋白质中的应用。

[0006] 目前,表达系统诸如酵母、丝状真菌、植物、藻类和昆虫细胞系(低等真核生物)正被研究用于生产治疗性蛋白质,这些系统与哺乳动物系统相比,更为安全、快速而且生产出的产物效价更高。这些系统共同具有 N-连接的寡糖合成的共有分泌途径。近来,已表明巴斯德毕赤氏酵母的分泌途径可再进行基因工程改造以实现顺序糖基化反应,该反应模拟人类和其它高等哺乳动物中 N-聚糖的早期加工(Choi 等, Proc Natl Acad Sci USA. 2003 Apr 29; 100(9):5022-7。此外,已显示了在巴斯德毕赤氏酵母中通过重新工程改造分泌途径以生产具有缺少半乳糖的复合 N-聚糖的人类糖蛋白(Hamilton 等, Science. 2003 Aug 29; 301(5637):1244-6)。在哺乳动物细胞中,进一步的成熟包括半乳糖转移。因此,来自酵母和低等真核生物的复合糖基化途径的成熟需要 β 1,4-半乳糖基转移酶的功能性表达。

[0007] 在哺乳动物细胞、昆虫细胞（例如，Sf-9）和酵母细胞中已经证实了 UDP-Gal : β GlcNAc β 1,4- 半乳糖基转移酶 (β 1,4GalT) 的重组表达。在甲基营养酵母巴斯德毕赤氏酵母中也已经表达了编码可溶形式的人类 β 1,4- 半乳糖基转移酶 I (EC2. 4. 1. 22) (缺少内源 II 型膜结构域) 的 cDNA。Malissard 等, BiochemBiophysRes Commun. 2000 Jan 7; 267(1):169-73。此外, 编码与人 β 1,4- 半乳糖基转移酶 (Gal-Tf) 催化结构域相融合的 ScMnt1p 的基因融合体已被表达, 从而在酵母高尔基体中显示出一些酶活性, 尽管处于非常低的转化效率。Schwientek 等, J BiolChem. 1996 Feb 16; 271(7):3398-405。因此, 将 β 1,4- 半乳糖基转移酶 (β 1,4GalT) 靶向产生包含末端 GlcNAc 的聚糖的宿主分泌途径, 预期将导致一些半乳糖的转移。但是, 在高等真核生物中复合聚糖的形成要涉及甘露糖苷酶 II 作用, 该酶在哺乳动物细胞中已知与 GalTI 竞争作用 (Fukuta 等, Arch Biochem Biophys. 2001 Aug 1; 392(1):79-86)。由此预期 GalT 的过早作用将阻止分泌途径中复合半乳糖基化糖蛋白的形成以及主要生产杂合聚糖。

[0008] 哺乳动物糖蛋白的 N- 聚糖一般包括半乳糖、岩藻糖和末端唾液酸。这些糖通常不会见于在酵母和丝状真菌中产生的糖蛋白。在人类中, 核苷酸糖前体（例如 UDP-N- 乙酰葡萄糖胺、UDP-N- 乙酰半乳糖胺、CMP-N- 乙酰神经氨酸、UDP- 半乳糖、GDP- 岩藻糖等）在胞质中合成并被转运至高尔基体, 在高尔基体它们由糖基转移酶掺入 N- 聚糖 (Sommers 和 Hirschberg, 1981 J. Cell Biol. 91(2):A406-A406; Sommers 和 Hirschberg 1982 J. Biol. Chem. 257(18):811-817; Perez 和 Hirschberg 1987 Methods in Enzymology 138:709-715)。

[0009] 在异种蛋白表达系统中的糖基化工程改造包括核苷酸糖前体合成中所涉及的各种酶的表达。酶 UDP- 半乳糖 4- 差向异构酶通过 C4 的差向异构化将糖核苷酸 UDP- 葡萄糖转变为 UDP- 半乳糖。该酶还可见于能将半乳糖用作为唯一碳源的生物中。近来, 双功能性酶 Gal10p 已经在啤酒糖酵母 (Saccharomyces cerevisiae) 中得到纯化, 其有 UDP- 半乳糖 4- 差向异构酶和醛糖 1- 差向异构酶两种活性。Majumdar 等, Eur J Biochem. 2004 Feb; 271(4):753-759。

[0010] UDP- 半乳糖转运蛋白 (UGT) 可将 UDP- 半乳糖从胞质转运至高尔基体腔。两种异源基因, 即源自粟酒裂殖糖酵母 (Schizosaccharomyces pombe) 的编码 α 1,2- 半乳糖基转移酶 (α 1,2 GalT) 的 gml2(+) 和编码人 UDP- 半乳糖转运蛋白的 (hUGT2), 已经在啤酒糖酵母中功能性表达, 以检查半乳糖基化所需的细胞内条件。蛋白半乳糖基化作用与 UDP- 半乳糖转运活性之间的关联, 表明 UDP-Gal 转运蛋白的外源供应在啤酒糖酵母有效半乳糖基化中发挥关键作用 (Kainuma, 1999 Glycobiology 9(2):133-141)。同样地, 来自粟酒裂殖糖酵母的 UDP- 半乳糖转运蛋白也已被克隆 (Aoki, 1999 J. Biochem. 126(5):940-950; Segawa, 1999 Febs Letters 451(3):295-298)。

[0011] 糖基转移反应通常会产生副产物, 所述副产物是核苷二磷酸或一磷酸。虽然一磷酸能够通过双向运输机理与核苷二磷酸糖交换直接输出, 但二磷酸核苷（例如, GDP）在被输出之前必须通过磷酸酶（例如, GDP 酶 (GDPase)）切割以产生核苷一磷酸和无机磷酸。该反应对有效糖基化是重要的; 例如, 源自啤酒糖酵母的 GDP 酶已被发现对于甘露糖基化是必需的。但是, GDP 酶对 UDP 的活性减小 90% (Berninsone 等, 1994 J. Biol. Chem. 269(1):207-211)。低等真核生物一般在高尔基体中缺少 UDP- 特异性二磷酸酶活性, 因为它们在基

于高尔基体的糖蛋白合成中不使用 UDP- 糖前体。已发现粟酒裂殖糖酵母,即发现能添加半乳糖残基至细胞壁多糖的酵母(由 UDP- 半乳糖),有特异性的 UDP 酶活性,从而显示了对这种酶的潜在需求(Berninsone 等,1994)。

[0012] 已知 UDP 是糖基转移酶的潜在抑制剂,且除去该糖基化副产物对防止高尔基体腔内糖基转移酶抑制作用是重要的(Khatara 等,1974)。参见 Berninsone, P. 等, 1995. J. Biol. Chem. 270 (24) :14564-14567 ;Beaudet, L. 等,1998Abc Transporters : Biochemical, Cellular, and Molecular Aspects. 292 :397-413。

[0013] 因此,需要这样一种方法,所述方法催化将半乳糖残基从 UDP- 半乳糖的充足库转移至优选的受体底物,以用作治疗性糖蛋白。

发明内容

[0014] 发明概述

[0015] 本发明提供了产生人样糖蛋白的新型低等真核生物宿主细胞,所述糖蛋白特征在于含有末端半乳糖残基并基本上在糖蛋白上缺少岩藻糖和唾液酸残基。在一个实施方案中,本发明提供了产生人样糖蛋白的重组低等真核生物宿主细胞,所述宿主包含编码 β - 半乳糖基转移酶活性的分离的核酸分子和至少编码 UDP- 半乳糖转运活性、UDP- 半乳糖 C4 差向异构酶活性、半乳糖激酶活性或半乳糖 -1- 磷酸尿苷酰转移酶的分离的核酸分子。本发明还提供了产生人样糖蛋白的重组低等真核生物宿主细胞,所述宿主细胞能够将 β - 半乳糖残基转移至包含末端 GlcNAc 残基的糖蛋白的 N- 连接的寡糖分支上,所述 N- 连接的寡糖分支选自三甘露糖核心上的 GlcNAc β 1,2-Man α 1,3、GlcNAc β 1,4-Man α 1,3、GlcNAc β 1,2-Man α 1,6、GlcNAc β 1,4-Man α 1,6 和 GlcNAc β 1,6-Man α 1,6。在另一实施方案中,本发明提供了产生糖蛋白的重组低等真核生物宿主细胞,所述糖蛋白是唾液酸转移的受体底物。

[0016] 在本发明的另一方面中,其中提供了包含人样糖蛋白的组合物,所述糖蛋白特征在于含有末端 β - 半乳糖残基并基本上在糖蛋白上缺少岩藻糖和唾液酸残基。在一个实施方案中,所述糖蛋白包含 N- 连接的寡糖,所述 N- 连接的寡糖选自 :GalGlcNAcMan₃GlcNAc₂、GalGlcNAc₂Man₃GlcNAc₂、Gal₂GlcNAc₂Man₃GlcNAc₂、GalGlcNAc₃Man₃GlcNAc₂、Gal₂GlcNAc₃Man₃GlcNAc₂、Gal₃GlcNAc₃Man₃GlcNAc₂、GalGlcNAc₄Man₃GlcNAc₂、Gal₂GlcNAc₄Man₃GlcNAc₂、Gal₃GlcNAc₄Man₃GlcNAc₂、Gal₄GlcNAc₄Man₃GlcNAc₂、GalGlcNAcMan₅GlcNAc₂、GalGlcNAc₂Man₅GlcNAc₂、Gal₂GlcNAc₂Man₅GlcNAc₂、GalGlcNAc₃Man₅GlcNAc₂、Gal₂GlcNAc₃Man₅GlcNAc₂ 和 Gal₃GlcNAc₃Man₅GlcNAc₂。

[0017] 在另一个实施方案中,提供了在低等真核生物宿主细胞中产生人样糖蛋白的方法,所述方法包括产生高于内源水平的 UDP- 半乳糖的步骤。

[0018] 在另一个实施方案中,提供了在低等真核生物宿主细胞中产生人样糖蛋白组合物的方法,所述方法包括将半乳糖残基在不存在岩藻糖和唾液酸残基时转移至杂合或复合糖蛋白上的步骤。

[0019] 根据本发明的方法,产生至少 10%、优选 33%、更优选 60% 或更高半乳糖基化的糖蛋白组合物。

[0020] 本发明进一步提供了表达 GalNAc 转移酶活性的重组低等真核生物宿主细胞。

[0021] 本发明还提供了表达编码异源 UDP 酶活性的基因的重组低等真核生物宿主细胞。

[0022] 此外,本发明提供了分离的多核苷酸,所述多核苷酸包含或由选自下列的核酸序列组成:(a)SEQ ID NO:14;(b)与SEQ ID NO:13供体核苷酸结合位点的氨基酸残基至少约90%相似的;(c)与SEQ ID NO:14至少92%、至少95%、至少98%、至少99%或至少99.9%相同的核酸序列;(d)编码具有SEQ ID NO:13氨基酸序列的保守多肽的核酸序列;(e)编码与SEQ ID NO:13至少78%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或至少99.9%相同的多肽的核酸序列;(f)在严格条件下与SEQ ID NO:13杂交的核酸序列;和(g)包含(a)-(f)中任一项的长度至少为60个连续核苷酸的片段的核酸序列。

[0023] 本文中还提供了修饰的多核苷酸,所述多核苷酸包含或由选自SEQ IDNO:48-SEQ ID NO:52保守区的核酸序列组成,其中所编码的多肽参与催化UDP-葡萄糖和UDP-半乳糖的相互转变以产生半乳糖基化的糖蛋白。

[0024] 附图简述

[0025] 图1A-1B描述了编码hGalT1的整合载体pXB53质粒图谱的构建。

[0026] 图2描述了编码粟酒裂殖糖酵母Gal差向异构酶(SpGalE)和hGalTI的整合载体pRCD425质粒图谱的构建。

[0027] 图3A-3B描述了编码黑腹果蝇(*D. melanogaster*)UDP-半乳糖转运蛋白(DmUGT)的整合载体pSH263质粒图谱的构建。

[0028] 图4描述了编码hGalTI、SpGalE和DmUGT的整合载体pRCD465质粒图谱的构建。

[0029] 图5描述了编码ScMnn2/SpGalE/hGalTI融合蛋白的整合载体pRCD461质粒图谱的构建。

[0030] 图6A描述了SpGalE的氨基酸序列。图6B描述了SpGALE的编码序列。

[0031] 图7描述了粟酒裂殖糖酵母、人、大肠杆菌(*E. coli*)和啤酒糖酵母差向异构酶的序列比对。

[0032] 图8A是在RDP30-10(用pRCD257转化的RDP27)中产生的从K3所释放的N-聚糖的MALDI-TOF-MS分析,显示出峰值位于1342m/z[A],对应于N-聚糖GlcNAc₂Man₃GlcNAc₂的质量。

[0033] 图8B是在RDP37(用pXB53转化的RDP30-10)中产生的从K3所释放的N-聚糖的MALDI-TOF-MS分析,显示出峰值位于1505m/z[B],对应于N-聚糖GalGlcNAc₂Man₃GlcNAc₂的质量,以及峰值位于1662m/z[C],对应于Gal₂GlcNAc₂Man₃GlcNAc₂的质量。

[0034] 图9A是在用pXB53转化的YSH-44中产生的从K3所释放的N-聚糖的MALDI-TOF-MS分析,显示出峰值位于1501m/z[B],对应于N-聚糖GalGlcNAc₂Man₃GlcNAc₂的质量,以及峰值位于1339m/z[A],对应于GlcNAc₂Man₃GlcNAc₂的质量。

[0035] 图9B是在用pXB53和pRCD395转化的YSH-44中产生的从K3所释放的N-聚糖的MALDI-TOF-MS分析,显示出峰值位于1501m/z[B],对应于N-聚糖GalGlcNAc₂Man₃GlcNAc₂的质量;峰值位于1663m/z[C],对应于Gal₂GlcNAc₂Man₃GlcNAc₂的质量;和峰值位于1339m/z[A],对应于GlcNAc₂Man₃GlcNAc₂的质量。

[0036] 图10A是在用pRCD352和pXB53转化的RDP39-6(巴斯德毕赤氏酵母PBP-3(美国专利申请号20040018590))中产生的从K3所释放的N-聚糖的MALDI-TOF-MS分析,显示出主要峰值位于1622m/z[K],对应于N-聚糖GalGlcNAcMan₅GlcNAc₂的质量,和峰值位于

1460m/z [A], 对应于 $\text{GlcNAcMan}_5\text{GlcNAc}_2$ 的质量。

[0037] 图 10B 是在用 α 1,2 和 β 1,4- 半乳糖苷酶消化后, 在 RDP39-6 中产生的从 K3 所释放的 N- 聚糖的 MALDI-TOF-MS 分析, 显示出主要峰值位于 1461m/z [H], 对应于 N- 聚糖 $\text{GlcNAcMan}_5\text{GlcNAc}_2$ 的质量。

[0038] 图 11 是在各种巴斯德毕赤氏酵母品系中产生的从 K3 所分离的 N- 聚糖的 MALDI-TOF-MS 分析, 其比较了 UDP- 半乳糖转运活性。A 图显示了用编码 Mnn2(s)/hGalTI 和 SpGalE 的载体 pRCD425 转化的巴斯德毕赤氏酵母 YSH-44 (其被命名为 RDP52) 的 N- 聚糖特征。B 图显示了用编码 SpUGT 的载体 pRCD425 和 pRCD393 转化的巴斯德毕赤氏酵母 YSH-44 (其被命名为 RDP69) 的 N- 聚糖特征。C 图显示了用编码 hUGT2 的载体 pRCD425 和 pSH262 转化的巴斯德毕赤氏酵母 YSH-44 (其被命名为 RDP70) 的 N- 聚糖特征。D 图显示了用编码 hUGT1 的载体 pRCD425 和 pSH264 转化的巴斯德毕赤氏酵母 YSH-44 (其被命名为 RDP71) 的 N- 聚糖特征。E 图显示了用编码 DmUGT 的载体 pRCD425 和 pSH263 转化的巴斯德毕赤氏酵母 YSH-44 (其被命名为 RDP57) 的 N- 聚糖特征。

[0039] 图 12 是在各种巴斯德毕赤氏酵母品系中产生的从 K3 所释放的 N- 聚糖的 MALDI-TOF-MS 分析, 其比较了 β -1,4- 半乳糖基转移酶的活性。A 图显示了用编码 DmUGT 的载体 pRCD425 和 pSH263 转化的巴斯德毕赤氏酵母 YSH-44 (其被命名为 RDP57) 的 N- 聚糖特征。B 图显示了用编码 Mnn2(s)/hGalTII 和 SpGalE 的载体 pRCD440 和编码 DmUGT 的载体 pSH263 转化的巴斯德毕赤氏酵母 YSH-44 (其被命名为 RDP72) 的 N- 聚糖特征。C 图显示了用编码 Mnn2(s)/hGalTIII 和 SpGalE 的载体 pRCD443 和编码 DmUGT 的载体 pSH263 转化的巴斯德毕赤氏酵母 YSH-44 (其被命名为 RDP73) 的 N- 聚糖特征。

[0040] 图 13 是在各种巴斯德毕赤氏酵母品系中产生的从 K3 所释放的 N- 聚糖的 MALDI-TOF-MS 分析, 其比较了差向异构酶活性。A 图显示了用编码 Mnn2(s)/hGalTI 和 ScGal110 的载体 pRCD424 和编码 DmUGT 的载体 pSH263 转化的巴斯德毕赤氏酵母 YSH-44 (其被命名为 RDP65) 的 N- 聚糖特征。B 图显示了用编码 DmUGT 载体 pSH263 和载体 pRCD425 顺次转化的巴斯德毕赤氏酵母 YSH-44 (其被命名为 RDP74) 的 N- 聚糖特征。C 图显示了先用载体 pRCD425 然后用编码 DmUGT 的载体 pSH263 顺次转化的巴斯德毕赤氏酵母 YSH-44 (其被命名为 RDP63) 的 N- 聚糖特征。D 图显示了用编码 Mnn2(s)/hGalTI 和 hGalE 的载体 pXB53 和 pRCD438 以及编码 DmUGT 的载体 pSH263 转化的巴斯德毕赤氏酵母 YSH-44 (其被命名为 RDP67) 的 N- 聚糖特征。

[0041] 图 14A 是在 RDP80 (用 pRCD465 转化的巴斯德毕赤氏酵母 YSH-44) 中产生的从 K3 所释放的 N- 聚糖的 MALDI-TOF-MS 分析, 显示主要峰值位于 1663m/z [C], 对应于 N- 聚糖 $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ 的质量。

[0042] 图 14B 是在用 β 1,4- 半乳糖苷酶消化后在 RDP80 (用 pRCD465 转化的巴斯德毕赤氏酵母 YSH-44) 中产生的从 K3 所释放的 N- 聚糖的 MALDI-TOF-MS 分析, 显示主要峰值位于 1340m/z [A], 对应于 N- 聚糖 $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ 的质量。

[0043] 图 14C 是在 RDP80 中产生并在 CMP-NANA 存在时与唾液酸转移酶体外温育所得的从 K3 释放的 N- 聚糖的 MALDI-TOF-MS 分析, 显示主要峰值位于 2227m/z [X], 对应于 N- 聚糖 $\text{NANA}_2\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ 的质量。

[0044] 图 15A 是在巴斯德毕赤氏酵母 YSH-44 (对照) 中产生的从 K3 所释放的 N- 聚糖

GlcNAc₂Man₃GlcNAc₂[A] 的 MALDI-TOF-MS 分析。图 15B 是在 RDP86 (用 pRCD461 (Mnn2(s)/SpGalE/hGalTI 融合体) 转化的巴斯德毕赤氏酵母 YSH-44) 中产生的从 K3 所释放的 N-聚糖的 MALDI-TOF-MS 分析, 显示主要峰值位于 1679m/z[C], 对应于 N-聚糖 Gal₂GlcNAc₂Man₃GlcNAc₂ 的质量。

[0045] 发明详述

[0046] 除非本文另有定义, 本发明相关使用的科学和技术术语具有本领域技术人员通常理解的含义。而且, 除非上下文有其它规定, 单数形式的术语应当包括复数, 而复数形式的术语应当包括单数。本发明的方法和技术通常根据本领域众所周知的常规方法进行。通常, 与本文记载相关使用的命名, 以及本文记载的生物化学、酶学、分子和细胞生物学、微生物学、遗传学以及蛋白质和核酸化学及杂交技术是本领域中众所周知的而且普遍使用的。除非另有说明, 本发明的方法和技术通常按照本技术领域众所周知的以及如贯穿本说明书中所引用和讨论的各种一般性的以及更为具体的参考文献中所记载的常规方法进行。参见例如 Sambrook 等, *Molecular Cloning: A Laboratory Manual*, 2d ed., ColdSpring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989); Ausubel 等, *Current Protocols in Molecular Biology*, Greene Publishing Associates (1992, 以及 2002 增刊); Harlow 和 Lane *Antibodies: A Laboratory Manual* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1990); *Introduction to Glycobiology*, Maureen E. Taylor, Kurt Drickamer, Oxford Univ. Press (2003); *Worthington Enzyme Manual*, Worthington Biochemical Corp. Freehold, NJ; *Handbook of Biochemistry: Section A Proteins Vol I* 1976 CRC Press; *Handbook of Biochemistry: Section A Proteins Vol II* 1976 CRC Press; *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press (1999)。与本文记载相关使用的命名, 以及本文记载的生物化学和分子生物学的实验室操作规程和技术是本领域中众所周知的而且普遍使用的。

[0047] 本文所提及的所有出版物、专利和其它参考文献均被引入作为参考。

[0048] 除非另有说明, 下列术语应被理解成具有以下含义:

[0049] 如本文使用的, 术语“K3”是指人纤溶酶原的 kringle3 结构域。

[0050] 如本文使用的, 术语“N-聚糖”是指 N-连接的寡糖, 例如, 通过天冬酰胺-N-乙酰葡萄糖胺键与多肽天冬酰胺残基附着的寡糖。N-聚糖含有共同的 Man₃GlcNAc₂ 五糖核心 (“Man”表示甘露糖; “Glc”表示葡萄糖; 而“NAc”是指 N-乙酰基; GlcNAc 是指 N-乙酰葡萄糖胺)。N-聚糖的差别在于包含添加至 Man₃GlcNAc₂ (“Man3”) 核心结构的外周糖 (例如, GlcNAc、半乳糖、岩藻糖和唾液酸) 的分支 (触角) 数目不同。N-聚糖可根据其分支的组分进行分类 (例如, 高甘露糖、复合或杂合)。“高甘露糖”型 N-聚糖含有五个或更多个甘露糖残基。“复合”型 N-聚糖通常含有至少一个与 1,3 甘露糖臂附着的 GlcNAc 和至少一个与“三甘露糖”核心的 1,6 甘露糖臂附着的 GlcNAc。“三甘露糖核心”是具有 Man3 结构的五糖核心。其通常被称为“少甘露糖 (paucimannose)”结构。复合 N-聚糖还可含有半乳糖 (“Gal”) 残基, 所述残基任选地被唾液酸或衍生物 (“NeuAc”, 其中“Neu”是指神经氨酸而“Ac”是指乙酰基) 修饰。复合 N-聚糖还可含有包含“二等分 (bisecting)”GlcNAc 和核心岩藻糖 (“Fuc”) 的链内取代。复合 N-聚糖还可含有在“三甘露糖核心”上多个触角, 其通常被称为“多触角性聚糖”。“杂合”N-聚糖在三甘露糖核心的 1,3 甘露糖臂末端至少有一个

GlcNAc, 且在三甘露糖核心的 1,6 甘露糖臂上有零个或多个甘露糖。

[0051] 本文所使用的缩写具有本领域通用的用法, 参见例如, 上述糖的缩写。其它通用的缩写法包括“PNG 酶 (PNGase)”, 其表示肽 N- 糖苷酶 F (EC3. 2. 2. 18); “GalT”, 其表示半乳糖基转移酶; “ β 1,4GalT”, 其表示 UDP- 半乳糖: β -N- 乙酰葡萄糖胺 β 1,4- 半乳糖基转移酶。来源于各种物种的 β - 半乳糖基转移酶缩写如下: “hGalT”是指人 β 1,4- 半乳糖基转移酶, “bGalT”是指牛 β 1,4- 半乳糖基转移酶, “XlGalT”是指非洲爪蟾 (*Xenopus laevis*) β 1,4- 半乳糖基转移酶而“CeGalT”是指秀丽隐杆线虫 (*C. elegans*) β 1,4- 半乳糖基转移酶。“GalNAcT”是指 UDP-GalNAc-GlcNAc β -1,4-N- 乙酰氨基半乳糖基转移酶。

[0052] 如本文使用的, 术语“UGT”是指 UDP- 半乳糖转运蛋白。术语“SpGalE”是指粟酒裂殖糖酵母 UDP- 半乳糖 4- 差向异构酶, “hGalE”是指人 UDP- 半乳糖 4- 差向异构酶, “ScGal10”是指啤酒糖酵母 UDP- 半乳糖 4- 差向异构酶而“EcGalE”是指大肠杆菌 UDP- 半乳糖 4- 差向异构酶。

[0053] 如本文使用的, 术语“UDP-Gal”是指 UDP- 半乳糖而术语“UDP-GalNAc”是指 UDP-N- 乙酰半乳糖胺。

[0054] N- 连接的糖蛋白包含与蛋白质天冬酰胺残基的酰胺氮相连接的 N- 乙酰葡萄糖胺残基。在糖蛋白中发现的主要糖是葡萄糖、半乳糖、甘露糖、岩藻糖、N- 乙酰半乳糖胺 (GalNAc)、N- 乙酰葡萄糖胺 (GlcNAc) 和唾液酸 (例如, N- 乙酰- 神经氨酸 (NANA))。糖基团加工在 ER 腔共翻译地发生, 且在高尔基体中继续而成为 N- 连接的糖蛋白。

[0055] 如本文使用的, 术语“人样”糖蛋白是指与蛋白质共价附着的修饰的 N- 聚糖, 其与人 N- 连接的寡糖合成中发现的糖蛋白相似。复合和杂合 N- 聚糖是人糖基化中发现的中间体。这些中间体的共有结构是 $\text{Man}_3\text{GlcNAc}_2$ 核心结构, 也称少甘露糖核心、五糖核心或简称 Man_3 或 Man_3 。因此, 人样糖蛋白具有至少 Man_3 核心结构。

[0056] 如本文使用的, 术语“起始 1,6 甘露糖基转移酶活性”是指在由含 α 1,6 键的 0ch1p 所起始的外链形成过程中, 酵母特异性聚糖残基通常被添加至三甘露糖核心的 $\text{Man} \alpha$ 1,3 臂上。

[0057] 半乳糖残基至 N- 聚糖的摩尔%转移, 如通过 MALDI-TOF-MS 以正模式测量的, 是指相对于摩尔%总的中性 N- 聚糖转移的摩尔%半乳糖转移。某些阳离子加合物诸如 K^+ 和 Na^+ , 通常与洗脱峰有一定关系, 从而通过各加合物的分子量增加 N- 聚糖的质量。

[0058] 如本文使用的, 术语“分泌途径”是指各种糖基化酶沿着新生多肽链从细胞质流向内质网 (ER) 和高尔基体区室的分子流的装配线, 脂质连接的寡糖前体和 N- 聚糖底物依次暴露于所述酶。这些酶据认为沿着该途径定位。在酶 Y 之前作用于脂质连接的聚糖或 N- 聚糖的酶 X 被称为位于或作用于酶 Y 的“上游”; 相似地, 酶 Y 位于或作用于酶 X 的“下游”。

[0059] 如本文使用的, 术语“突变”是指基因产物例如糖基化相关酶的核酸或氨基酸序列中的任何改变。

[0060] 术语“多核苷酸”或“核酸分子”是指长度至少为 10 个碱基的核苷酸多聚形式。该术语包括 DNA 分子 (例如, cDNA 或基因组或合成 DNA) 和 RNA 分子 (例如, mRNA 或合成 RNA), 以及包含非天然核苷酸类似物、非天然核苷间键或二者兼有的 DNA 或 RNA 类似物。核酸可以是任何拓扑学构象。例如, 核酸可以是单链、双链、三链、四元、部分双链、分支、发夹、环状或扣锁 (padlocked) 构象。该术语包括单链和双链形式的 DNA。

[0061] 除非另有说明,“包含 SEQ ID NO:X 的核酸”是指这样的核酸,该核酸的至少一部分含有 (i) SEQ ID NO:X 的序列,或者 (ii) 与 SEQ ID NO:X 互补的序列。在所述两者之间的选择由上下文确定。例如,如果核酸用作探针,则在所述两者之间的选择由探针与所需靶互补的要求来确定。

[0062] “分离的”或“基本上纯的”核酸或多核苷酸(例如,RNA、DNA 或混合聚合物)是指基本与其它细胞成分分离的核酸或多核苷酸,其它细胞成分是在天然宿主细胞中与天然多核苷酸天然伴随的那些成分,例如与所述核酸天然结合的核糖体、聚合酶和基因组序列。该术语包括如下核酸或多核苷酸,所述核酸或多核苷酸 (1) 已从其天然存在的环境中取出,(2) 与多核苷酸的全部或部分不结合,在自然界中从该多核苷酸发现“分离的多核苷酸”,(3) 可操作地连接于其天然不连接的多核苷酸,或 (4) 不存在于自然界中。术语“分离的”或“基本上纯的”,还用于是指重组或克隆的 DNA 分离物、化学合成的多核苷酸类似物或异源系统生物合成的多核苷酸类似物。

[0063] 然而,“分离的”并非一定要求所述核酸或多核苷酸自身已经从其天然环境中物理地取出。例如,如果存在下述情况,则生物基因组中的内源核酸序列则在本文中被认为是“分离的”,所述情况是将异源序列(即天然不与该内源核酸序列毗邻的序列)置于与所述内源核酸序列相毗邻,由此改变该内源核酸序列的表达。举例说明,可用非天然启动子序列取代(例如通过同源重组)人细胞基因组中基因的天然启动子,从而该基因具有改变的表达模式。该基因现在可成为“分离的”,因为其与至少部分其天然侧接序列相分离。

[0064] 如果核酸包含在基因组中相应核酸天然不存在的任何修饰,则该核酸也被认为是“分离的”。例如,如果内源编码序列包含人工例如通过人为干预,导入的插入、缺失或点突变,则内源编码序列被认为是“分离的”。“分离的核酸”还包括被整合至宿主细胞染色体异源位点的核酸,即表现为附加体的核酸构建体。而且,“分离的核酸”当通过重组技术制备时可基本不含其它细胞材料或基本不含培养基,或当通过化学合成时基本不含化学前体或其它化学制剂。

[0065] 如本文使用的,参照核酸序列的术语“简并变体”包括这样的核酸序列,该核酸序列可根据标准遗传密码进行翻译以提供与所述参照核酸序列翻译的相同的氨基酸序列。

[0066] 核酸序列上下文中的术语“百分比序列同一性”或“同一性”是指当进行最大对应性比对时两条序列中相同的残基。序列同一性比较的长度可以在至少约 9 个核苷酸,通常至少约 20 个核苷酸,更通常至少约 24 个核苷酸,一般地至少约 28 个核苷酸,更一般至少约 32 个核苷酸,以及优选地至少约 36 个或更多个核苷酸的一段序列上。在本技术领域中有许多不同的算法可用于测量核苷酸序列的同一性。例如,多核苷酸序列可以利用 FASTA、Gap 或 Bestfit 来进行比较,FASTA、Gap 或 Bestfit 是 Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin 中的程序。FASTA 提供了在查询和检索序列之间最佳重叠区域的比对和百分比序列同一性 (Pearson, 1990, (在此引入本文作为参考)。例如,核酸序列之间的百分比序列同一性可以利用具有默认参数(字长为 6 和评分矩阵采用 NOPAM 因子)的 FASTA 确定或利用具有 GCG Version 6.1 所提供的具有其默认参数的 Gap 来确定,它们被引入本文作为参考。

[0067] 当涉及核酸或其片段时,术语“相当大同源性”或“相当大相似性”是指当与另一个核酸(或其互补链)进行含有适当核苷酸插入或缺失的最佳序列对比时,如通过任一序

列同一性的公知算法,如上述的 FASTA、BLAST 或 Gap 所测量的,核苷酸序列同一性至为少约 50%,更优选 60%的核苷酸碱基,通常至少约 70%,更通常至少约 80%,优选至少约 90%,且更优选地至少约 95%、96%、97%、98%或 99%的核苷酸碱基。

[0068] 可选地,当核酸或其片段与另一核酸、另一核酸的链或其互补链在严格杂交条件下杂交时,存在相当大的同源性或相似性。核酸杂交试验上下文中的“严格杂交条件”和“严格洗涤条件”依赖于大量不同的物理参数。核酸杂交会受到诸如盐浓度、温度、溶剂、杂交种类的碱基组成、互补区的长度以及杂交核酸之间核苷酸碱基错配数目的条件的影响,这一点本领域技术人员很容易理解。本领域普通技术人员知晓如何改变这些参数以实现特定杂交严格性。

[0069] 一般而言,“严格杂交”在特定组的条件下在低于特定 DNA 杂交体热解链温度 (T_m) 约 25°C 下进行。“严格洗涤”在特定组的条件下在低于特定 DNA 杂交体 T_m 约 5°C 的温度下进行。 T_m 是 50%的靶序列与完全匹配的探针杂交的温度。参见 Sambrook 等,如上文,9.51 页,在此引入作为参考。为了本发明目的,“高严格条件”对溶液相杂交定义为于 65°C 在 6X SSC(其中 20X SSC 包含 3.0MNaCl 和 0.3M 柠檬酸钠),1% SDS 中水性杂交(即不含甲酰胺)8-12 小时,之后为于 65°C 在 0.2X SSC,0.1% SDS 中的两次洗涤,20 分钟。本领域技术人员可以理解于 65°C 的杂交依赖于多种因子,包括杂交序列的长度和百分比同一性而会以不同的速率发生。

[0070] 本发明的核酸(也称为多核苷酸)可包括 RNA、cDNA、基因组 DNA,和上述的合成形式以及混合多聚体的有义链和反义链。它们可以被化学或生化修饰,或可以包含非天然或衍生的核苷酸碱基,本领域技术人员很容易理解这一点。所述修饰包括,例如,标记,甲基化,一个或多个天然存在的核苷酸被其类似物取代,核苷酸间的修饰,诸如非荷电的键(例如,甲基膦酸酯、磷酸三酯、氨基磷酸酯、氨基甲酸酯等),带电的键(例如,硫代磷酸酯、二硫代磷酸酯等),侧基部分(例如,多肽),嵌入剂(例如,吡啶、补骨脂素等),螯合剂,烷化剂,以及修饰的键(例如, α 端基异构核酸等)。还包括在通过氢键和其它化学相互作用结合于指定序列的能力中模拟多核苷酸的合成分子。所述分子是本领域已知的,且包括例如其中分子骨架中磷酸酯键被肽键取代的那些分子。

[0071] 当用于核酸序列时,术语“突变的”是指相对于参照核酸序列而言,核酸序列中的核苷酸可被插入、缺失和改变。可在基因座中作出单个改变(点突变),或可在单一基因座中插入、缺失或改变多个核苷酸。此外,可在核酸序列中任何数目基因座上作出一个或多个改变。核酸序列可采用本领域已知的任何方法进行突变,所述方法包括但不限于诱变技术,诸如“易错 PCR”(在 DNA 聚合酶复制保真度低的条件下进行 PCR 以使沿 PCR 产物全长获得高点突变率的方法。参见例如,Leung, D. W. 等, *Technique*, 1, pp. 11-15(1989) 以及 Caldwell, R. C. & Joyce G. F., *PCR Methods Applic.*, 2, pp. 28-33(1992)); 和“寡核苷酸指导的诱变”(能够在任何克隆的目标 DNA 片段中产生位点特异性突变的方法。参见例如, Reidhaar-Olson, J. F. & Sauer, R. T. 等, *Science*, 241, pp. 53-57(1988))。

[0072] 本文所使用的术语“载体”是指能够转运已经与其连接的另一核酸的核酸分子。载体的一种类型是“质粒”,其是指其中可连接其它 DNA 片段的环状双链 DNA 环。其它载体包括粘粒、细菌人工染色体(BAC)和酵母人工染色体(YAC)。另一类型的载体是病毒载体,其中额外 DNA 片段可连接至病毒基因组中(下面有更详细的讨论)。特定载体能够在其导入

的宿主细胞中自主复制（例如，含有能够在宿主细胞中发挥功能的复制起点的载体）。其它载体可在导入宿主细胞后被整合至宿主细胞基因组中，并由此随宿主基因组一起复制。此外，特定优选的载体能够指导与其可操作地连接的基因的表达。所述载体在本文中被称为“重组表达载体”（或简称为“表达载体”）。

[0073] “可操作地连接的”表达控制序列是指这样一种连接方式，其中表达控制序列邻接目标基因以控制目标基因，以及表达控制序列反式或在远处作用以控制目标基因。

[0074] 本文所使用的术语“表达控制序列”是指影响与其可操作地连接的编码序列表达所必需的多核苷酸序列。表达控制序列是控制核酸序列转录、转录后事件及翻译的序列。表达控制序列包括适宜的转录起始、终止、启动子和增强子序列；有效的 RNA 加工信号，诸如剪接和多腺苷酸化信号；稳定细胞质 mRNA 的序列；增强翻译效率的序列（例如，核糖体结合位点）；增强蛋白质稳定性的序列；以及在需要时，增强蛋白质分泌的序列。所述控制序列的性质依赖于宿主生物而不同；在原核生物中，所述控制序列通常包括启动子、核糖体结合位点和转录终止序列。术语“控制序列”最低限度地意欲包括表达必需的全部组分，且也可包括有益的其它组分，例如前导序列和融合配偶体序列（fusionpartner sequence）。

[0075] 术语“重组低等真核宿主细胞”（或简称为“宿主细胞”），如本文使用的，意指其中已经导入了重组载体的细胞。应当理解，所述术语不仅意指特定受试者细胞而且还意指该细胞的子代。因为由于突变或环境影响，某些修饰会在连续世代中发生，所以所述子代事实上与亲代细胞并不相同，但其仍包含在本发明术语“宿主细胞”范围内。重组宿主细胞可以是培养基中生长的分离的细胞或细胞系，或者可以是位于活体组织或生物中的细胞。重组宿主细胞包括酵母、真菌、领鞭毛虫（collar-flagellates）、微孢子虫、alveolates（例如，甲藻）、stramenopiles（例如，褐藻、原生动物）、红藻门（rhodophyta）（例如，红藻）、植物（例如，绿藻、植物细胞、藓类）以及其它原生生物。

[0076] 本文使用的术语“肽”是指短多肽，例如长度通常小于约 50 个氨基酸以及更通常长度小于约 30 个氨基酸的肽。本文使用的该术语包括模拟结构并由此模拟生物学功能的类似物和模拟物。

[0077] 术语“多肽”，包括天然存在的和非天然存在的蛋白质，以及其片段、突变体、衍生物和类似物。多肽可以是单体或多聚体。而且，多肽可包括大量不同的结构域，每一结构域都具有一种或多种不同活性。

[0078] 术语“分离的蛋白质”或“分离的多肽”是这样的蛋白质或多肽，根据其起源或衍生来源，其（1）不与在天然状态中与其伴随的天然结合的组分结合，（2）当以未见于自然界的纯度存在时，其中纯度可参照其它细胞材料的存在而得以判定（例如，不含来自相同物种的其它蛋白），（3）被来自不同物种的细胞所表达，或者（4）不存在于自然界中（例如，其是在自然界中发现的多肽的片段或其包含未见于自然界的氨基酸类似物或衍生物或包含除标准肽键之外的其它键）。因此，通过化学合成或在与其天然起源的细胞不同的细胞系统合成的多肽，即为与其天然结合的组分“分离的”。还可采用本领域众所周知的蛋白质纯化技术，通过分离使多肽或蛋白质基本上不含天然结合的组分。由此定义，“分离的”并非必然要求所述蛋白质、多肽、肽或寡肽已经从其天然环境中物理地取出。

[0079] 本文使用的术语“多肽片段”是指与全长多肽相比，含有氨基末端和 / 或羧基末端缺失的多肽。在优选的实施方案中，多肽片段是连续的序列，其中所述片段的氨基酸序列与

天然存在的序列中的相应位置相同。片段长度通常为至少 5、6、7、8、9 或 10 个氨基酸,优选至少 12、14、16 或 18 个氨基酸,更优选至少 20 个氨基酸,更优选至少 25、30、35、40 或 45 个氨基酸,甚至更优选至少 50 或 60 个氨基酸,以及甚至更优选至少 70 个氨基酸。

[0080] “修饰的衍生物”是指一级结构序列基本上同源但包含例如体内或体外化学和生化修饰或掺入天然多肽中未见的氨基酸的那些多肽或其片段。所述修饰包括,例如乙酰化、羧基化、磷酸化、糖基化、遍在蛋白化、例如用放射性核素进行的标记、以及各种酶促修饰,这一点本领域技术人员很容易理解。标记多肽的多种方法以及用于该目的的多种取代基或标记是本领域众所周知的,且包括放射性同位素,诸如 ^{125}I 、 ^{32}P 、 ^{35}S 和 ^3H 、结合标记抗配体 (antiligand) (例如抗体) 的配体、荧光团、化学发光剂、酶以及可用作标记的配体的特异结合对成员的抗配体。标记的选择依赖于需要的灵敏度、与引物缀合的简易度、稳定性需要以及可用的手段。标记多肽的方法是本领域众所周知的。参见 Ausubel 等,1992,在此引入作为参考。

[0081] 术语“融合蛋白”是指包含与异源氨基酸序列偶联的多肽或片段的多肽。融合蛋白是有用的,因为它们可被构建为含有来自两种或更多种不同蛋白的两种或更多种所需功能元件。融合蛋白包含来自目标多肽的至少 10 个连续氨基酸,更优选地至少 20 或 30 个氨基酸,甚至更优选地至少 40、50 或 60 个氨基酸,而更优选地至少 75、100 或 125 个氨基酸。融合蛋白可通过构建与编码不同蛋白或肽的核酸序列符合读框的编码多肽或其片段的核酸序列,然后表达融合蛋白而重组性生产。可选地,融合蛋白可通过将多肽或其片段与另一蛋白交联而化学生产。

[0082] 术语“非肽类似物”是指特性与参照多肽特性相类似的化合物。非肽化合物还可被称为“肽的模拟物 (peptide mimetic)”或“肽模拟物 (peptidomimetic)”。参见例如, Jones, (1992) *Amino Acid and Peptide Synthesis*, Oxford University Press; Jung, (1997) *Combinatorial Peptide and Nonpeptide Libraries: A Handbook* John Wiley; Bodanszky 等, (1993) *Peptide Chemistry--A Practical Textbook*, Springer Verlag; ' *Synthetic Peptides: A Users Guide* ', G. A. Grant, Ed, W. H. Freeman and Co., 1992; Evans 等, *J. Med. Chem.* 30 :1229 (1987); Fauchere, J. *Adv. Drug Res.* 15 :29 (1986); Veber 和 Freidinger *TINS* p. 392 (1985); 以及上述文献中所引用的参考文献,所述参考文献在此引入作为参考。所述化合物通常借助计算机分子建模进行开发。与本发明有用的肽结构相似的肽模拟物可用于产生等同作用,并因此被认为构成本发明的一部分。

[0083] “多肽突变体”或“突变蛋白质”是指与天然或野生型蛋白质的氨基酸序列相比,序列包含一个或多个氨基酸的插入、重复、缺失、重排或取代的多肽。突变蛋白质可有一个或多个氨基酸点取代,其中一个位置上的单个氨基酸变成另一氨基酸;突变蛋白质可有一个或多个插入和 / 或缺失,其中在天然存在的蛋白序列中分别插入或缺失一个或多个氨基酸;和 / 或突变蛋白质可有在氨基或羧基末端或者两者的氨基酸序列的截短。突变蛋白质可具有与天然存在的蛋白质相同的生物学活性,但优选具有与之不同的生物学活性。

[0084] 突变蛋白质具有与其野生型对应物至少 70% 的全长序列同源性。甚至更优选的是突变蛋白质具有与野生型蛋白质 80%、85% 或 90% 的全长序列同源性。在甚至更为优选的实施方案中,突变蛋白质表现出 95% 序列同一性,甚至更优选 97%,甚至更优选 98% 以及甚至更优选 99%、99.5% 或 99.9% 的全长序列同一性。序列同源性可通过任何常规的序列

分析算法,诸如 Gap 或 Bestfit 进行测量。

[0085] 优选氨基酸取代是如下的取代,所述取代:(1)降低了对蛋白水解的易感性,(2)降低了对氧化的易感性,(3)改变了用于形成蛋白质复合物的结合亲和性,(4)改变了结合亲和性或酶活性,以及(5)赋予或修饰了所述类似物的其它理化或功能特性。

[0086] 如本文使用的,二十个常规氨基酸及其缩写遵循常规用法。参见 Immunology-A Synthesis(第二版,E. S. Golub 和 D. R. Gren 编, Sinauer Associates, Sunderland, Mass. (1991)),其被引入本文作为参考。二十个常规氨基酸的立体异构体(例如 D-氨基酸),非天然氨基酸如 α -, α -双取代的氨基酸, N-烷基氨基酸,和其它非常规氨基酸也可以是本发明多肽的合适组分。非常规氨基酸的实例包括:4-羟脯氨酸、 γ -羧基谷氨酸、 ϵ -N, N, N-三甲基赖氨酸、 ϵ -N-乙酰赖氨酸、O-磷酸丝氨酸、N-乙酰丝氨酸、N-甲酰甲硫氨酸、3-甲基组氨酸、5-羟赖氨酸、s-N-甲基精氨酸和其它类似氨基酸和亚氨基酸(例如 4-羟脯氨酸)。在本文使用的多肽符号中,按照标准用法和惯例,左手方向是氨基末端方向而右手方向是羧基末端方向。

[0087] 如果编码蛋白质的核酸序列含有与编码第二种蛋白质的核酸序列相似的序列,则该蛋白质与第二种蛋白质有“同源性”或是“同源的”。可选地,如果两种蛋白质具有“相似的”氨基酸序列,则该蛋白质与第二种蛋白质有同源性。(因此,术语“同源的蛋白质”意指两种蛋白质有相似的氨基酸序列)。在优选实施方案中,同源的蛋白质是与野生型蛋白质表现出 60%序列同源性的蛋白质,更优选表现出 70%序列同源性的蛋白质。甚至更优选的是与野生型蛋白质表现出 80%、85%或 90%序列同源性的同源的蛋白质。而在更优选的实施方案中,同源的蛋白质表现出 95%、97%、98%或 99%的序列同一性。如本文使用的,两个氨基酸序列域之间的同源性(尤其是关于预测的结构相似性)被理解为暗示功能上的相似性。

[0088] 当“同源”应用于蛋白质或肽时,应该认识到,不相同的残基位置通常差别在于保守氨基酸取代。“保守氨基酸取代”是其中一氨基酸残基被另一具有相似化学性质(例如电荷或疏水性)的侧链(R基团)的氨基酸残基所取代。通常,保守氨基酸取代基本上不会改变蛋白质功能特性。当两个或更多个氨基酸序列由于保守取代而彼此不同时,百分比序列同一性或同源性程度可被上调以校正取代的保守特性。进行这种调整的手段是本技术领域技术人员所熟知的(参见例如, Pearson 等, 1994, 引入本文作为参考)。

[0089] 以下六组的每一组中包含彼此互为保守取代的氨基酸:1) 丝氨酸(S), 苏氨酸(T); 2) 天冬氨酸(D), 谷氨酸(E); 3) 天冬酰胺(N), 谷酰胺(Q); 4) 精氨酸(R), 赖氨酸(K); 5) 异亮氨酸(I), 亮氨酸(L), 甲硫氨酸(M), 丙氨酸(A), 缬氨酸(V), 和 6) 苯丙氨酸(F), 酪氨酸(Y), 色氨酸(W)。

[0090] 多肽的序列同源性,也被称作百分比序列同一性,一般利用序列分析软件来测量。参见例如, Genetics Computer Group(GCG), University of Wisconsin Biotechnology Center, 910 University Avenue, Madison, Wisconsin 53705 的序列分析软件包。蛋白分析软件利用分配给各种取代、缺失和包括保守氨基酸取代在内的其它修饰的同源性量度来使相似序列匹配。例如, GCG 包括诸如“Gap”和“Bestfit”这样的程序,该程序可用于在默认参数下确定密切相关的多肽之间(诸如来源于生物不同物种的同源多肽之间,或者野生型蛋白与其突变蛋白质之间)的序列同源性或序列同一性。参见例如 GCG 6.1 版。

[0091] 当将抑制性分子序列与含有源自不同生物的大量序列的数据库进行比较时,优选的算法是计算机程序 BLAST(Altschul, S.F. 等 (1990) *J. Mol. Biol.* 215:403-410; Gish 和 States (1993) *Nature Genet.* 3:266-272; Madden, T.L 等 (1996) *Meth. Enzymol.* 266:131-141; Altschul, S.F. 等 (1997) *Nucleic Acids Res.* 25:3389-3402; Zhang, J. 和 Madden, T.L. (1997) *Genome Res.* 7:649-656), 特别是 blastp 或 tblastn(Altschul 等, 1997)。BLASTp 的优选参数是:期望值 (Expectation value):10(默认值);Filter:seg(默认值);打开缺口的成本 (Cost to open a gap):11(默认值);延伸缺口的成本 (Cost to extend a gap):1(默认值);最大比对 (Max. alignments):100(默认值);字长 (Word size):11(默认值);描述数目 (No. of descriptions):100(默认值);罚分矩阵 (Penalty Matrix):BLOWSUM62。

[0092] 进行同源性比较的多肽序列的长度一般是至少约 16 个氨基酸残基,通常至少约 20 个残基,更通常地是至少约 24 个残基,一般至少约 28 个残基,以及优选地多于约 35 个残基。当检索包括源于大量不同生物的序列的数据库时,优选地是进行氨基酸序列比较。采用氨基酸序列的数据库检索可通过本领域中已知的除 blastp 之外的算法进行测量。例如,可以利用 FASTA,一种 GCG 6.1 版程序来比较多肽序列。FASTA 可提供查询和检索序列之间最佳重叠区域的比对和百分比序列同一性 (Pearson (1990), 其被引入本文作为参考)。例如,氨基酸序列之间百分比序列同一性可用具有默认参数的 FASTA(字长为 2 和评分矩阵 PAM250) 来确定,该程序提供在 GCG Version 6.1 中,引入本文作为参考。

[0093] 本文使用的术语“结构域”是指对生物分子已知或推定功能有贡献的生物分子结构。结构域可以是与区域或其部分共生 (coextensive) 的;结构域也可包括生物分子不同的非连续区域。蛋白质结构域的实例包括,但不限于 Ig 结构域、胞外结构域、跨膜结构域和胞质结构域。

[0094] 本文使用的,术语“分子”是指任何化合物,包括但不限于小分子、肽、蛋白质、糖、核苷酸、核酸、脂质等,且这种化合物可以是天然的或合成的。

[0095] 在本说明书以及其实施方案通篇中,词语“包含 (comprise)”或其变体形式诸如“包含 (comprises)”或“包含 (comprising)”,应被理解是指包含所述的整体或整体的组而并不排除任何其它整体或整体的组。

[0096] 除非另有定义,本文所使用的所有技术和科学术语均具有本发明所属领域技术人员通常所理解的含义。以下描述了示例性方法和材料,尽管与本文所述的方法和材料相似或等同的方法和材料也可用于实施本发明并且它们对于本领域技术人员而言是显而易见的。本文所述的所有出版物和其它参考文件均全文引入作为参考。如有冲突,以本说明书(包括定义)为准。材料、方法和实施例仅为举例说明而并非意在限制。

[0097] 工程改造宿主以产生人样半乳糖基化糖蛋白

[0098] 本发明提供了产生人样糖蛋白的重组低等真核宿主细胞,其中所述糖蛋白的特征在于含有末端 β -半乳糖残基并基本上缺少岩藻糖和唾液酸。在一个实施方案中,本发明提供了低等真核生物宿主细胞,所述细胞包含编码 UDP-半乳糖: β -N-乙酰葡萄糖胺 β 1,4-半乳糖基转移酶 (β 1,4GalT) 的分离的核酸分子与至少第二种编码 UDP-半乳糖转运蛋白的分离的核酸分子、编码 UDP-半乳糖 4-差向异构酶的分离的核酸或编码半乳糖激酶或半乳糖-1-磷酸尿苷酰转移酶的分离的核酸的组合。在另一实施方案中, β 1,4GalT 与编

码 UDP- 半乳糖转运蛋白的分离的核酸分子和编码 UDP- 半乳糖 4- 差向异构酶的分离的核酸分子组合表达。还提供了编码上述酶的核酸序列的变体和片段、重组 DNA 分子和包含用于转化的酶的表达载体。

[0099] 本发明的一个方面,提供在低等真核生物宿主细胞中产生人样糖蛋白的方法,所述方法包括如下步骤:通过表达 β 1,4GalT 活性,催化半乳糖残基以 β - 键从 UDP- 半乳糖转移至受体底物上,和向宿主中导入 UDP- 半乳糖 4- 差向异构酶活性、半乳糖激酶活性、半乳糖 -1- 磷酸尿苷酰转移酶活性或 UDP- 半乳糖转运活性。受体底物优选为包含末端 GlcNAc 残基的寡糖组合物,例如,三甘露糖核心上的 GlcNAc β 1,2-Man α 1,3 ;GlcNAc β 1,4-Man α 1,3 ;GlcNAc β 1,2-Man α 1,6 ;GlcNAc β 1,4-Man α 1,6 ; 或 GlcNAc β 1,6-Man α 1,6 分支。

[0100] 受体底物更为优选为与目标蛋白质共价连接(N- 连接)的复合聚糖(例如, GlcNAc₂Man₃GlcNAc₂)、杂合聚糖(例如, GlcNAcMan₅GlcNAc₂)或多触角聚糖(例如, GlcNAc₄Man₃GlcNAc₂)。 β - 半乳糖残基被转移至受体底物上从而形成 β - 糖苷键,所述受体底物包含 2- 乙酰氨基 -2- 脱氧 -D- 葡萄糖(GlcNAc) 碳原子 4 位的羟基。包含能够接受半乳糖残基的末端 GlcNAc 残基的 N- 连接的受体底物包括,但不限于, GlcNAcMan₃GlcNAc₂、GlcNAc₂Man₃GlcNAc₂、GlcNAc₃Man₃GlcNAc₂、GlcNAc₄Man₃GlcNAc₂、GlcNAc₅Man₃GlcNAc₂、GlcNAc₆Man₃GlcNAc₂、GlcNAcMan₄GlcNAc₂、GlcNAcMan₅GlcNAc₂、GlcNAc₂Man₅GlcNAc₂ 和 GlcNAc₃Man₅GlcNAc₂。

[0101] β 1,4- 半乳糖基转移酶基因的克隆

[0102] 采用引物 RCD192(SEQ ID NO :1) 和 RCD186(SEQ ID NO :2) 从人肾 cDNA(marathon ready cDNA, Clontech)PCR 扩增人 β -1,4- 半乳糖基转移酶 I 基因(hGalTI, Genbank AH003575)。将该 PCR 产物克隆至克隆的 pCR2.1(Invitrogen) 并测序。从该克隆实施 PCR 重叠诱变(overlap mutagenesis)。基因 5' 末端至 NotI 位点采用引物 RCD198(SEQ ID NO :3) 和 RCD201(SEQ ID NO :4) 进行扩增,而 3' 末端采用引物 RCD200(SEQ ID NO :5) 和 RCD199(SEQ ID NO :6) 进行扩增。采用 RCD198(SEQ ID NO :3) 和 RCD199(SEQ ID NO :6) 将产物重叠在一起以重新合成含有野生型氨基酸(除 N- 端缺失 43 个氨基酸外)序列但除去了 NotI 位点的 ORF。新的截短的 hGalTI PCR 产物在 pCR2.1 中克隆并测序。导入的 AscI/PacI 位点然后被用于将片段亚克隆至质粒 pRCD259(图 1)(PpURA3/HYG^Rroll-in 载体),从而产生 pRCD260(图 1)(实施例 4)。

[0103] 使用相同的策略克隆人 β 1,4GalTII 和人 β 1,4GalTIII。实施例 4 描述了采用基因特异性引物通过 PCR 扩增人 β 1,4- 半乳糖基转移酶 II 和 III 并随后将其克隆至载体。

[0104] 在低等真核生物中表达 β 1,4- 半乳糖基转移酶活性

[0105] 在低等真核生物宿主细胞(例如巴斯德毕赤氏酵母)中导入并表达编码 β 1,4GalT 活性的基因或编码 β 1,4- 半乳糖基转移酶活性的重组核酸分子,编码 β 1,4GalT 活性的基因融合体(例如, pXB53)(图 1) 或从编码 β 1,4- 半乳糖基转移酶的核酸分子(Genbank AH003575) 的表达,以生成半乳糖基化糖蛋白。可选地,通过激活 β - 半乳糖基转移酶活性,低等真核生物宿主细胞可进行工程改造以生产半乳糖基化糖形(galactosylated glycoform)。催化活性的 β 1,4- 半乳糖基转移酶结构域或其部分催化半乳糖残基从 UDP- 半乳糖转移至寡糖受体底物(例如, GlcNAc₂Man₃GlcNAc₂) 的末端 GlcNAc

残基上,从而形成 β 1,4Gal 糖苷键。根据本发明产生的复合半乳糖基化 N-聚糖基本上缺少岩藻糖和唾液酸(例如, $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$)。包含复合半乳糖基化、无岩藻糖化以及无唾液酸化 N-聚糖的所述糖蛋白组合物可以用作治疗剂。

[0106] 新形成的底物还是在低等真核生物宿主中生产的唾液酸化糖蛋白形成中的优选前体。因此本发明提供了生产人样糖蛋白的方法,其中所述糖蛋白的特征在于含有末端半乳糖残基,所述半乳糖残基是用于在低等真核生物中转移唾液酸的受体底物。

[0107] β 1,4-半乳糖基转移酶的组合物 DNA 文库

[0108] 在本发明的相关方面, β 1,4-半乳糖基转移酶和酵母靶向序列跨膜结构域的组合物 DNA 文库在低等真核生物宿主细胞中生成并表达,如 W002/00879 中所述。

[0109] 因此,生成了与长度:短、中、长靶向肽亚文库相融合的 hGalTI(例如 Genbank 登录号 X55415)亚文库,如 W002/00879 中所述。靶向肽亚文库包括编码靶向信号肽的核酸序列,所述靶向信号肽将使蛋白质在 ER、高尔基体或高尔基体外侧网络中的特定位置定位。这些靶向肽选自要工程改造的宿主生物以及其它相关或不相关的生物。通常,所述序列可分为如下三类:(1)N-末端序列,该序列编码胞质尾(ct)、跨膜结构域(tmd)以及部分或全部的茎区(sr),其一同或独立地将蛋白质锚定在高尔基体内(腔的)膜上;(2)回收信号,其通常见于 C-端,诸如 HDEL 或 KDEL 四肽;和(3)来自各种蛋白质的膜跨越区,例如,核苷酸糖转运蛋白,其已知定位于高尔基体。

[0110] 靶向肽在本文中相对于 II 型膜蛋白的部分而被描述为短(s)、中(m)和长(l)的。描述为短(s)的靶向肽序列对应于膜结合蛋白的跨膜结构域(tmd)。描述为长(l)的靶向肽序列对应于跨膜结构域(tmd)和茎区(sr)的长度。被描述为中(m)的靶向肽序列对应于跨膜结构域(tmd)和茎区(sr)大约一半的长度。在本文中根据相对于其野生型糖基化酶缺失的核苷酸数目而对催化结构域区进行描述。

[0111] 在一个实施方案中,将文库转化至巴斯德毕赤氏酵母中然后在包含潮霉素的基本培养基上选择转化体。与各种前导序列(如下述)融合的 β 1,4-半乳糖基转移酶 I 活性通过半乳糖基化 N-聚糖的生成作为以正模式进行的 MALDI-TOF MS 的读数来进行分析。

[0112] $\square$ -半乳糖基转移酶融合构建体

[0113] 将分离的酵母靶向序列跨膜结构域文库(由 48 个前导序列组成(W002/00879))连接至 pRCD260 中的 NotI/AscI 位点,所述位点位于 hGalTI 基因的上游,以生成质粒 pXB20-pXB67(每个质粒包含一种前导序列)。

[0114] 本发明源自组合物 DNA 文库的 GalT 融合构建体的代表性实例是 pXB53(图 1),其是截短的啤酒糖酵母 Mnn2(s)靶向肽(来自 Genbank NP_009571MNN2 的 1-108 位核苷酸),其与 N-端 43 个氨基酸缺失的人 β 1,4-半乳糖基转移酶 I(Genbank AH003575)符合读框地连接。因此本文使用的命名,将靶向肽/糖基化酶催化结构域区命名为啤酒糖酵母 Mnn2(s)/hGalTI Δ 43。但是,编码的融合蛋白单独不足以生成有大部分半乳糖基化聚糖的 N-聚糖,如图 9A 所示。虽然与 N-聚糖 $\text{Gal}_1\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ [B] 质量相符的峰值在巴斯德毕赤氏酵母 YSH-44 中有 hGalTI 导入时显示,然而,随后消化所述样品显示该峰值与 β -1,4-半乳糖苷酶并不相符(实施例 7)。

[0115] 此外, β -1,4-半乳糖基转移酶活性可能对特定目标蛋白是特异性的。因此,可以进一步理解并非所有靶向肽/半乳糖基转移酶催化结构域融合构建体都同样好地发挥功

能以在目标糖蛋白上生成正确的糖基化。因此,目标蛋白质可被导入用组合 DNA 文库转化的宿主细胞以鉴定出一种或多种表达最适于目标蛋白的半乳糖基转移酶活性的融合构建体。本领域技术人员能够采用本文所述的组合 DNA 文库方法构建并选择最优的融合构建体。

[0116] 此外,展示出局部化活性的半乳糖基转移酶催化结构域(或更通常地,任何酶的结构域)的其它这种融合构建体可采用本文所述技术进行制备,这是显而易见的。制备和使用本发明的组合 DNA 文库以例如,在导入特定宿主细胞的特定表达载体中最优化从融合构建体文库中生产 $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$,这是本领域技术人员的常规试验。

[0117] 在基因改变的巴斯德毕赤氏酵母中产生半乳糖基化 N-聚糖

[0118] 根据本发明方法产生的人样半乳糖基化糖蛋白包括 $\text{GalGlcNAcMan}_3\text{GlcNAc}_2$ 、 $\text{GalGlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ 、 $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ 、 $\text{GalGlcNAc}_3\text{Man}_3\text{GlcNAc}_2$ 、 $\text{Gal}_2\text{GlcNAc}_3\text{Man}_3\text{GlcNAc}_2$ 、 $\text{Gal}_3\text{GlcNAc}_3\text{Man}_3\text{GlcNAc}_2$ 、 $\text{GalGlcNAc}_4\text{Man}_3\text{GlcNAc}_2$ 、 $\text{Gal}_2\text{GlcNAc}_4\text{Man}_3\text{GlcNAc}_2$ 、 $\text{Gal}_3\text{GlcNAc}_4\text{Man}_3\text{GlcNAc}_2$ 、 $\text{Gal}_4\text{GlcNAc}_4\text{Man}_3\text{GlcNAc}_2$ 、 $\text{GalGlcNAcMan}_5\text{GlcNAc}_2$ 、 $\text{GalGlcNAc}_2\text{Man}_5\text{GlcNAc}_2$ 、 $\text{Gal}_2\text{GlcNAc}_2\text{Man}_5\text{GlcNAc}_2$ 、 $\text{GalGlcNAcMan}_5\text{GlcNAc}_2$ 、 $\text{GalGlcNAc}_2\text{Man}_5\text{GlcNAc}_2$ 、 $\text{Gal}_2\text{GlcNAc}_2\text{Man}_5\text{GlcNAc}_2$ 、 $\text{GalGlcNAc}_3\text{Man}_5\text{GlcNAc}_2$ 、 $\text{Gal}_2\text{GlcNAc}_3\text{Man}_5\text{GlcNAc}_2$ 和 $\text{Gal}_3\text{GlcNAc}_3\text{Man}_5\text{GlcNAc}_2$ 。

[0119] 在本发明的一个实施方案中,将包含 MNN2(s)/hGalTI 的质粒 pXB53 转化至巴斯德毕赤氏酵母 RDP30-10,产生 $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ 的宿主中(实施例 5)。催化活性的 β -半乳糖基转移酶结构域催化半乳糖残基转移至含有末端 GlcNAc 残基的受体底物(例如, $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$),以产生半乳糖基化糖形。采用 MALDI-TOFMS,来自巴斯德毕赤氏酵母 RDP37 报告蛋白质所释放的 N-聚糖,显示出峰值位于 1505m/z ,该峰值与 $\text{GalGlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ [B] 的质量相符(图 8B3)。通过包含人啤酒糖酵母 Mnn2(s)/ $\beta 1,4$ -半乳糖基转移酶融合构建体,将半乳糖残基转移至受体底物 $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$,从而产生 $\text{GalGlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ 显示为大约有 10-20%。图 8B 显示了 $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ 在 1662m/z 的对应质量 [C]。因此,两个半乳糖残基转移至 $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ 底物从而产生 $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ 是显而易见的。因此,本发明的宿主表现出人样 N-聚糖上至少 10 摩尔%的半乳糖部分。

[0120] 应当认识到 GalTI 能够将第二个半乳糖残基转移至宿主内的受体底物上,所述受体底物含有第二个末端 GlcNAc 残基,从而产生复合(例如,双触角)聚糖。例如, Mnn2(s)/hGalTI 融合体,其能在聚糖 $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ 的 $\text{GlcNAc } \beta 1,2 \text{ Man } \alpha 1,3$ 臂上用半乳糖残基对末端 GlcNAc 加帽,能在暴露有末端 GlcNAc 残基的其它臂上(例如, $\text{GlcNAc } \beta 1,2 \text{ Man } \alpha 1,6$) 形成至少一个额外的 β -糖苷键,由此在不表达后续半乳糖基转移酶的情况下,产生了半乳糖基化糖形。图 12 显示了 MALDI-TOF MS 所示峰值位于 1663m/z [C],其与 $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ 相对应。结果显示,特定 $\beta 1,4$ -GalT 的底物特异性并不限于催化仅将半乳糖残基转移至聚糖特定臂上,因此第二种半乳糖基转移酶可以不用。因此,在本发明的一个实施方案中,仅表达一个 $\beta 1,4$ -GalT 活性即能够产生单-、双-、三-和四-触角半乳糖基化糖形。在所述实施方案中,聚糖上的半乳糖残基和 GlcNAc 残基之间的所有糖苷键都是相同的。例如, hGalTI 在产生双触角聚糖的宿主中的表达将展示两个末端 $\text{Gal } \beta 1,4$ - $\text{GlcNAc } \beta 1,2$ 键。

[0121] 可选地,不同的 β -半乳糖基转移酶活性(例如 hGalTII)或其催化性活化部分,可在低等真核生物宿主细胞中表达。在一个实施方案中,包含 MNN2(s)/hGalTII 和 SpGALE 的载体 pRCD440 以及包含 DmUGT 的载体 pSH263(图 3B)被转化至宿主巴斯德毕赤氏酵母 YSH-44 中(图 12B)。转化体的 N-聚糖分析显示,产生了 $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ 糖形,从而提示 hGalTII 将两个半乳糖残基转移至受体底物上(图 12B)。二-半乳糖基化结构($\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$)占主要部分。关于%中性聚糖,半乳糖基部分的转移大约是 75%。

[0122] 在另一实施方案中,在低等真核生物宿主细胞中表达编码 hGalTIII 的序列。图 12C 显示了组合的单-和双-半乳糖基化聚糖的半乳糖转移,达到了约 50 至 60 摩尔%。hGalTI、hGalTII 和 hGalTIII 的比较显示出各种水平的半乳糖转移(图 12A-C)。来自巴斯德毕赤氏酵母 RDP71 的 N-聚糖特征(图 12A)显示,hGalTI 表达所产生的半乳糖残基转移对于 K3 报告蛋白质是最佳的(大约 80 摩尔%)。

[0123] 另外的 β 1,4-半乳糖基转移酶的表达

[0124] 在另一实施方案中,分别采用中期和晚期高尔基体靶向序列依次定位和表达 hGalTI 和 hGalTII。例如, hGalTI 定位中期高尔基体,而 hGalTII 定位晚期高尔基体。可选地,为了避免底物与甘露糖苷酶 II 竞争,在另一实施方案中,针对 β -半乳糖基转移酶使用晚期高尔基体前导序列。

[0125] 半乳糖基转移酶活性的表达通常产生单-和双半乳糖基化聚糖。除单-半乳糖基化糖形之外,多触角半乳糖基化糖形通常可在表达半乳糖基转移酶活性的宿主细胞中生成。

[0126] 对本领域技术人员而言,通过采用各种启动子和各种表达载体在重组宿主细胞中最优化半乳糖基转移酶活性或编码蛋白的基因的表达是常规试验。

[0127] 在 N-聚糖产生中定制半乳糖基化糖苷键

[0128] 在本发明的另一方面,采用不同 GalT 产生多触角半乳糖基化糖蛋白,会产生不同的 β -糖苷键。在一个实施方案中,优选所需的 β -糖苷键在低等真核生物宿主细胞中产生。例如,表达 β 1,4GalT 家族中任一种(例如, hGalTI、hGalT2、hGalT3、hGalT4、hGalT5、hGalT6、hGalT7、bGalTI、XlGalT、CeGalTII),以用于产生特征在于含有 β 1,4Gal 糖苷键的半乳糖基化糖蛋白。

[0129] 可选地,通过在低等真核生物宿主细胞(例如巴斯德毕赤氏酵母)中表达其它半乳糖基转移酶,诸如 β 1,3GalT 或 β 1,6GalT 活性(酶、同源物、变体、衍生物及其催化性活化片段),半乳糖残基被转移至中间寡糖受体底物上,从而形成具体所需的 β Gal-糖苷键。作为所需 β -半乳糖基转移酶活性表达的结果,形成了各种末端半乳糖键(例如, β 1,3、 β 1,4 或 β 1,6)。

[0130] GalNAcT 在低等真核生物中的表达

[0131] 已经在人特异性蛋白质上观测到 GalNAc 加帽的聚糖。在本发明的另一方面,编码 GalNAc 转移酶(GalNAcT)的基因在低等真核生物宿主细胞中表达,其将 GalNAc 残基转移至含有末端 GlcNAc 残基的底物上。在一个实施方案中,编码秀丽隐杆线虫 GalNAcT 的基因(Genbank AN NP_490872)催化 GalNAc 残基转移至含有末端 GlcNAc 残基的底物上,从而延伸宿主细胞中所产生的聚糖的寡糖分支。

[0132] 增强的半乳糖基转移

[0133] 如图 12 所示, hGalTI 表达比较显示, β -半乳糖基转移酶的单独表达可能不足以在低等真核生物受体底物上形成 β -Gal-糖苷键。通过加入编码差向异构酶或半乳糖激酶、半乳糖-1-磷酸尿苷酰转移酶的异源基因和/或编码 UGT 的基因, 可增强半乳糖残基的转移。足量的半乳糖基化糖形(例如, $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$) 适宜用作治疗性糖蛋白。因此, 本发明的一个特征是通过另外表达转运活性和/或提升内源 UDP-半乳糖水平, 来增强半乳糖基至聚糖的转移。在一个实施方案中, 在宿主细胞中导入差向异构酶活性以增加 UDP-半乳糖水平。在另一实施方案中, 通过半乳糖激酶或半乳糖-1-磷酸尿苷酰转移酶活性介导增加的 UDP-半乳糖水平。因此, 本发明提供了通过导入和表达 β -半乳糖基转移酶活性组合 UDP-Gal 转运活性和/或通过借助差向异构酶或半乳糖激酶或半乳糖-1-磷酸尿苷酰转移酶提高内源性 UDP-半乳糖的水平, 来而增强半乳糖转移的方法。

[0134] 在人样糖蛋白生产中 UDP-半乳糖转运蛋白 (UGT) 在低等真核生物宿主中的克隆和表达

[0135] 本说明书中, 还公开了在低等真核生物细胞(例如巴斯德毕赤氏酵母)中导入和表达编码 UDP-半乳糖转运蛋白的基因以用于产生人样半乳糖基化糖蛋白的方法。

[0136] 粟酒裂殖糖酵母 UDP-半乳糖转运蛋白的克隆和表达

[0137] 可以设计与粟酒裂殖糖酵母 UDP-半乳糖转运蛋白基因 (GenbankAL022598) 同源区互补的基因特异性引物并从除去单内含子的粟酒裂殖糖酵母基因组 DNA (ATCC24843) 进行 PCR 扩增。引物 RCD164 (SEQ ID NO:7) 和 RCD177 (SEQ ID NO:8) 用于扩增基因的 5' 96bp。引物 RCD176 (SEQ ID NO:9) 和 RCD165 (SEQ ID NO:10) 用于扩增 3' 96bp。引物 RCD164 (SEQ ID NO:7) 和 RCD165 (SEQ ID NO:10) 用于将两个扩增产物重叠成单个 PCR 片段, 所述片段包含一个末端导入有 NotI 和 PacI 位点的连续 ORF。将 PCR 产物克隆至 pCR2.1TA (Invitrogen) 中并测序。将基因产物亚克隆至包含巴斯德毕赤氏酵母 GAPDH 启动子的质粒 pJN335 中 (实施例 2)。

[0138] 因此, 在一个实施方案中, 在产生末端 GlcNAc 残基的宿主(巴斯德毕赤氏酵母 RDP-27 (例如 $\text{GlcNAcMan}_3\text{GlcNAc}_2$)) 中构建并表达编码粟酒裂殖糖酵母 UDP-半乳糖转运蛋白 (Genbank AB023425) 的质粒 pRCD257。

[0139] 各种 UDP-半乳糖转运蛋白的克隆和表达

[0140] 在优选的实施方案中, 在低等真核生物宿主细胞中导入并表达编码黑腹果蝇 UDP-半乳糖转运蛋白的基因。从黑腹果蝇 cDNA 文库 (UC Berkeley Drosophila Genome Project, 卵巢 λ -ZAP 文库 GM) PCR 扩增黑腹果蝇 UGT, 然后将其克隆至 pCR2.1 PCR 克隆载体并测序。引物 DmUGT-5' (SEQ ID NO:11) 和 DmUGT-3' (SEQ ID NO:12) 用于扩增基因, 从而导入 NotI 和 PacI 位点。NotI 和 PacI 位点用于在 pRCD393 NotI/PacI 位点处亚克隆该基因, 从而使其融合至 PpOCH1 启动子下游, 以产生 pSH263 (图 3B)。实施例 2 描述了各种其它 UDP 半乳糖转运蛋白的克隆。

[0141] 图 11 显示了 UDP-转运蛋白活性关于增强半乳糖转移的比较。作为本发明的最佳模式, 从黑腹果蝇中分离的 UDP-半乳糖转运蛋白在巴斯德毕赤氏酵母中表达。图 11E 中显示了与黑腹果蝇 UDP-半乳糖转运蛋白 (DmUGT) 共表达的人 GalTI 基因融合体的活性。令人惊讶地, 表达黑腹果蝇 UGT 的宿主细胞主要产生半乳糖基化糖形, 然而, 来自粟酒裂殖糖

酵母(图 11B)、人 I(图 11C) 和人 II(图 11D) 的 UGT 显示出低于最优转移。双-半乳糖基化、无岩藻糖化且无唾液酸化的糖形 $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ 的产生有显著增加。均一峰值位于 $1664\text{m/z}[\text{C}]$, 对应于聚糖 $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ 的质量。与其它 UDP- 半乳糖转运蛋白相比, 表达 DmUGT 的宿主细胞(例如, 巴斯德毕赤氏酵母) 表现出至少 90 摩尔%的半乳糖转移。

[0142] UDP- 半乳糖转运蛋白多肽

[0143] 本发明另外提供了在低等真核生物宿主细胞(例如, 巴斯德毕赤氏酵母) 中表达的转运蛋白- 转移酶融合体的各种组合。因此, 在一个实施方案中, 本发明提供了一种低等真核生物宿主, 其包含与催化活性的 β - 半乳糖基转移酶结构域框内融合的 UDP- 半乳糖转运蛋白。在另一实施方案中, 产生人样糖蛋白的宿主细胞包含从粟酒裂殖糖酵母中分离的 UDP- 半乳糖转运蛋白和与 hGalTI 催化结构域框内融合的啤酒糖酵母 Mnn2(s) 靶向肽。

[0144] 在人样糖蛋白产生中 UDP- 半乳糖 4- 差向异构酶在低等真核生物宿主中的表达

[0145] 在本发明的另一方面, 提供了通过表达 β 1,4- 半乳糖基转移酶活性和至少 UDP- 半乳糖 4- 差向异构酶活性(酶、同源物、变体、衍生物及其催化活性片段) 在低等真核生物(例如巴斯德毕赤氏酵母) 中产生人样糖蛋白的方法。差向异构酶是催化 UDP- 半乳糖和 UDP- 葡萄糖相互转变的酶。使用本领域众所周知的技术, 可以设计与差向异构酶基因(例如 ScGAL10、SpGALE、hGALE) 同源区互补的基因特异性引物并进行 PCR 扩增(实施例 3)。在一个实施方案中, 在低等真核生物宿主细胞(例如巴斯德毕赤氏酵母) 中导入并表达编码啤酒糖酵母 Gal10 活性的基因或编码差向异构酶的重组核酸分子或从编码差向异构酶活性的核酸分子的表达, 以产生特征在于含有末端 β - 半乳糖残基的人样糖蛋白。可选地, 通过激活差向异构酶活性, 宿主细胞经工程改造以产生水平增高的半乳糖基化糖形。

[0146] 在复合 N- 聚糖的产生中表达 UDP- 半乳糖 4- 差向异构酶

[0147] 在一个实施方案中, 表达编码差向异构酶活性的基因以将 UDP- 葡萄糖转变为 UDP- 半乳糖, 从而产生水平增高的 UDP- 半乳糖用于宿主细胞中的半乳糖基转移。除 β -1,4- 半乳糖基转移酶活性之外差向异构酶活性的表达增加半乳糖基化 N- 聚糖的生成。图 9B 显示了用 Mnn2(s)/hGalTI 融合体组合 pRCD395(编码 ScGal10 的质粒) 转化的、产生复合聚糖的酵母品系(例如, 巴斯德毕赤氏酵母 YSH-44)。添加 ScGal10 差向异构酶增加对于半乳糖转移可用的 UDP- 半乳糖。位于 $1501\text{m/z}[\text{B}]$ 的峰值对应于聚糖 $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ 上的一个半乳糖残基转移, 而位于 $1663\text{m/z}[\text{C}]$ 的峰值对应于聚糖 $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ 上的两个半乳糖残基转移。优选地, 关于%总中性聚糖, 至少 60 摩尔%的半乳糖被转移。因此, 在一个实施方案中, β -1,4- 半乳糖基转移酶活性与差向异构酶活性组合在宿主细胞中表达以产生半乳糖基化糖蛋白(实施例 7)。

[0148] 在杂合 N- 聚糖的产生中表达 UDP- 半乳糖 4- 差向异构酶

[0149] 在另一实施方案中, 导入并表达 ScGAL10 增加了低等真核生物中半乳糖向杂合糖蛋白的转移(实施例 6)。图 10A 显示了表达 Mnn2(m)/hGalTI 融合体组合 ScGal10 差向异构酶的巴斯德毕赤氏酵母品系 RDP39-6, 从而产生杂合半乳糖基化 N- 聚糖。N- 聚糖分析显示峰值位于 $1622\text{m/z}[\text{K}]$, 其对应于聚糖 $\text{GalGlcNAcMan}_5\text{GlcNAc}_2$ 的质量, 从而证实了有一个半乳糖残基的转移, 以及峰值位于 $1460\text{m/z}[\text{H}]$, 其对应于杂合聚糖 $\text{GlcNAcMan}_5\text{GlcNAc}_2$ 的质量。随后的 β 1,4- 半乳糖苷酶消化证实了单半乳糖残基的存在(图 10B)。优选地, 关于%

总中性聚糖,检测到至少 70 摩尔%的半乳糖转移。

[0150] 在宿主细胞中还表达了其它差向异构酶以增加半乳糖转移。实施例 3 描述了构建差向异构酶构建体,而图 13 显示了各种差向异构酶在产生人样 N-聚糖中的活性。图 13A 中 ScGal10 连同 Mnn2(s)/hGalTI 和 DmUGT 的表达,表现出主要的双-半乳糖基化糖形 Gal₂GlcNAc₂Man₃GlcNAc₂。相似地,SpGalE、Mnn2(s)/hGalTI 和 DmUGT 任一次序的转化导致产生出双-半乳糖基化糖形(图 13B 和 C)。加入 hGalE 具有相同的作用(图 13D)。优选地,差向异构酶选自啤酒糖酵母 UDP-半乳糖 4-差向异构酶、粟酒裂殖糖酵母 UDP-半乳糖 4-差向异构酶、大肠杆菌 UDP-半乳糖 4-差向异构酶和智人(*H. sapiens*)UDP-半乳糖 4-差向异构酶。应当预期的是,在宿主细胞中也可以不受限制地选择和表达其它差向异构酶不受限制。

[0151] 编码 SpGALE 的核酸序列

[0152] 本发明还提供了分离的核酸分子,所述核酸分子包括来自粟酒裂殖糖酵母的 GALE 基因及其变体。该基因(编码酶 UDP-半乳糖 4-差向异构酶)的全长核酸序列已经被测序和鉴定,如 Genbank NC_003423 所示。用于扩增来自粟酒裂殖糖酵母基因组 DNA 的 SpGALE 的引物揭示了 175bp 内含子,所述内含子被除去(实施例 3)。克隆的基因组序列中包含粟酒裂殖糖酵母 UDP-半乳糖 4-差向异构酶编码序列。编码的氨基酸序列同样如 SEQ ID NO: 13 所示。SpGALE 基因尤其可用于在宿主细胞中产生用于将半乳糖转移至 N-聚糖上的充足 UDP-半乳糖库。在低等真核生物中表达 SpGALE 基因提供 N-连接的寡糖合成中增加的且有效的半乳糖转移。

[0153] 在一个实施方案中,本发明提供了含有核酸序列的分离的核酸分子,所述核酸序列包括或由 SEQ ID NO:14 所示的 SpGALE 编码序列、其同源物、变体和衍生物组成。在其它实施方案中,本发明提供了核酸分子,所述核酸分子包含如下序列或由如下序列组成,所述序列是具有与野生型基因至少 53%同一性的 SpGALE 基因的变体。所述核酸序列可优选具有与野生型基因至少 70%、75%或 80%同一性。甚至更优选地,所述核酸序列可具有与野生型基因 85%、90%、95%、98%、99%、99.9%或甚至更高的同一性。

[0154] 在另一实施方案中,本发明的核酸分子编码具有 SEQ ID NO:13 氨基酸序列的多肽。还提供了编码多肽序列的核酸分子,所述多肽序列与 SEQ ID NO:13 至少 60%相同。通常,本发明的核酸分子编码与 SEQ ID NO:13 有至少 70%、75%或 80%同一性的多肽序列。优选地,所编码的多肽与 SEQ ID NO:13 为 85%、90%或 95%相同的,且所述同一性可甚至更优选地为 98%、99%、99.9%或甚至更高。

[0155] 用于产生半乳糖基化糖蛋白的 UDP-葡萄糖和 UDP-半乳糖相互转变中所涉及的差向异构酶保守区

[0156] 来自粟酒裂殖糖酵母、人、大肠杆菌的差向异构酶与啤酒糖酵母的起始 362 个氨基酸残基的序列比对,显示出指示存在有若干基序和潜在活性位点的高度保守区(图 7)(实施例 11)。在一个实施方案中,本发明包括多肽,所述多肽包含 SEQ ID NO:13 的氨基酸序列,其在下述序列上含有潜在的 UDP-半乳糖或 UDP-葡萄糖结合基序:

[0157] 9-VLVTGGXGYIGSHT-22(SEQ ID NO:48),

[0158] 83-VIHFAGLKAVGESXQXPLXYY-103(SEQ ID NO:49),

[0159] 127-FSSSATVYGX-136(SEQ ID NO:50),

- [0160] 184-LRYFNPXGAHXSGXXGEDPXGIPNNLXPYXXQVAXGRX-221 (SEQ ID NO :51), 或
- [0161] 224-LXXFGXDYXXXDGTXXRDYIHVXDLAXXHHXAX-256 (SEQ ID NO :52)。
- [0162] 在另一优选的实施方案中, 第一序列 15 位上的氨基酸残基选自 S 和 A。
- [0163] 在另一优选的实施方案中, 第二序列 96 位上的氨基酸残基选自 T 和 V。
- [0164] 在另一优选的实施方案中, 第二序列 98 位上的氨基酸残基选自 V、K 和 I。
- [0165] 在另一优选的实施方案中, 第二序列 101 位上的氨基酸残基选自 S、D、E 和 R。
- [0166] 在另一优选的实施方案中, 第三序列 136 位上的氨基酸残基选自 D 和 N。
- [0167] 在另一优选的实施方案中, 第四序列 190 位上的氨基酸残基选自 G、T、V 和 I。
- [0168] 在另一优选的实施方案中, 第四序列 194 位上的氨基酸残基选自 P 和 A。
- [0169] 在另一优选的实施方案中, 第四序列 197 位上的氨基酸残基选自 E、C、D 和 L。
- [0170] 在另一优选的实施方案中, 第四序列 198 位上的氨基酸残基选自 L、I 和 M。
- [0171] 在另一优选的实施方案中, 第四序列 203 位上的氨基酸残基选自 L 和 Q。
- [0172] 在另一优选的实施方案中, 第四序列 210 位上的氨基酸残基选自 L 和 M。
- [0173] 在另一优选的实施方案中, 第四序列 213 位上的氨基酸残基选自 I、V 和 M。
- [0174] 在另一优选的实施方案中, 第四序列 214 位上的氨基酸残基选自 A 和 S。
- [0175] 在另一优选的实施方案中, 第四序列 218 位上的氨基酸残基选自 V 和 I。
- [0176] 在另一优选的实施方案中, 第四序列 221 位上的氨基酸残基选自 L 和 R。
- [0177] 在另一优选的实施方案中, 第五序列 225 位上的氨基酸残基选自 N、A 和 Y。
- [0178] 在另一优选的实施方案中, 第五序列 226 位上的氨基酸残基选自 V 和 I。
- [0179] 在另一优选的实施方案中, 第五序列 229 位上的氨基酸残基选自 D 和 N。
- [0180] 在另一优选的实施方案中, 第五序列 232 位上的氨基酸残基选自 P 和 D。
- [0181] 在另一优选的实施方案中, 第五序列 233 位上的氨基酸残基选自 T 和 S。
- [0182] 在另一优选的实施方案中, 第五序列 234 位上的氨基酸残基选自 S、E 和 R。
- [0183] 在另一优选的实施方案中, 第五序列 238 位上的氨基酸残基选自 P 和 G。
- [0184] 在另一优选的实施方案中, 第五序列 239 位上的氨基酸残基选自 I 和 V。
- [0185] 在另一优选的实施方案中, 第五序列 246 位上的氨基酸残基选自 C、V 和 M。
- [0186] 在另一优选的实施方案中, 第五序列 250 位上的氨基酸残基选自 E、K 和 D。
- [0187] 在另一优选的实施方案中, 第五序列 251 位上的氨基酸残基选自 A 和 G。
- [0188] 在另一优选的实施方案中, 第五序列 253 位上的氨基酸残基选自 V 和 I。
- [0189] 在另一优选的实施方案中, 第五序列 254 位上的氨基酸残基选自 A 和 V。
- [0190] 在另一优选的实施方案中, 第五序列 256 位上的氨基酸残基选自 L 和 M。
- [0191] 分离的多肽
- [0192] 根据本发明的另一方面, 提供了本发明核酸分子编码的分离的多肽 (包括突变蛋白质、等位基因变体、片段、衍生物和类似物)。在一个实施方案中, 分离的多肽包含与 SEQ ID NO :13 相应的多肽序列。在本发明可选的实施方案中, 分离的多肽包含具有与 SEQ ID NO :13 至少 60% 同一性的多肽序列。优选地, 本发明分离的多肽具有与 SEQ ID NO :13 至少 70%、75% 或 80% 的同一性。更优选地, 同一性为 85%、90% 或 95%, 但与 SEQ ID NO :13 的同一性可以是 98%、99%、99.9% 或甚至更高。
- [0193] 根据本发明的其它实施方案, 提供了包含上述多肽序列的片段的分离的多肽。这

些片段优选包含至少 20 个连续的氨基酸,更优选至少 25、30、35、40、45、50、60、70、80、90、100 或甚至更多个连续的氨基酸。

[0194] 本发明的多肽还包含上述多肽序列和异源多肽之间的融合体。例如,异源序列可以包括经设计促进重组表达的蛋白纯化和 / 或显现的异源序列。

[0195] 蛋白融合体的其它非限制性实例包括允许在噬菌体或细胞表面展示编码的蛋白的融合体,与固有荧光蛋白诸如绿色荧光蛋白 (GFP) 的融合体,以及与 IgG Fc 区的融合体。

[0196] UDP- 半乳糖 4- 差向异构酶 β 1,4- 半乳糖基转移酶融合多肽

[0197] 在本发明的其它方面,生成编码包含差向异构酶和半乳糖基转移酶活性的多肽的基因融合体。在一个实施方案中,生成包含 UDP- 半乳糖 4- 差向异构酶和 β 1,4-GalTI 的融合多肽并将其导入宿主细胞。在更优选的实施方案中,融合多肽进一步包含前导序列。例如,将编码靶向肽的前导序列文库框内连接至 SpGalE/hGalTI 融合体。在甚至更优选的实施方案中,融合多肽包含 ScMnn2(s) 前导序列、SpGalE 差向异构酶和 hGalTI。将融合多肽插入包含 HYG 标记的酵母整合质粒。图 5 中显示了差向异构酶 - 半乳糖基转移酶整合质粒的一个实例,命名为 pRCD461 (实施例 8)。差向异构酶 - 半乳糖基转移酶融合转化体产生大约 70% 半乳糖基化的人样糖蛋白 Gal₂GlcNAc₂Man₃GlcNAc₂ (图 15B)。

[0198] B1,4- 半乳糖基转移酶 ;UDP- 半乳糖 4- 差向异构酶 ;UDP- 半乳糖转运蛋白多肽

[0199] 在本发明的另一方面中,生成了编码多肽的单一构建体,所述多肽包含 β - 半乳糖基转移酶、差向异构酶和 UDP- 半乳糖转运蛋白活性。在一个实施方案中,构建包含人 β 1,4GalT、SpGalE 和 DmUGT (“三元 (triple)”) 的质粒 (实施例 9)。在优选的实施方案中,转移酶多肽进一步包含前导序列,例如,框内连接于 hGalTI 的 ScMnn2(s)。将三种多肽全部插入包含 KAN^R 标记的酵母整合质粒,所述质粒优选含有自身的启动子和终止子。图 4 中显示了该“三元”整合质粒的实例,命名为 pRCD465。在一个实施方案中,在产生末端 GlcNAc 残基的宿主细胞中导入并表达包含融合多肽的“三元”整合质粒。用“三元”整合质粒转化巴斯德毕赤氏酵母 YSH-44 并将其命名为 RDP80。

[0200] 为了评价在品系 RDP80 中产生的 N- 聚糖是否为预期的 Gal₂GlcNAc₂Man₃GlcNAc₂ 种类,在存在 CMP-NANA 的条件下,将 RDP80 所分泌的纯化的 K3 与唾液酸转移酶于体外温育,然后释放所得到的 N- 聚糖。N- 聚糖的 MALDI-TOF MS 分析显示,主要峰值位于 2227m/z [X],其对应于复合物末端唾液酸化的 N- 聚糖 NANA₂Gal₂GlcNAc₂Man₃GlcNAc₂ 的质量 (图 14C)。

[0201] UDP-Gal 的可选择的产生

[0202] 如前所述,半乳糖残基转移至 N- 聚糖上需要活化的半乳糖 (UDP-Gal) 库。在低等真核生物中产生高于内源性水平的所述库的方法之一是表达 UDP- 半乳糖 4 差向异构酶。可选的方法包括在不存在 UDP- 半乳糖 4 差向异构酶的情况下,表达三个独立的基因:质膜半乳糖通透酶、半乳糖激酶和半乳糖 -1- 磷酸尿苷酰转移酶。在不存在 UDP- 半乳糖 4 差向异构酶而存在外源半乳糖来源的情况下,表达 LeLoir 途径的其它三种基因,将可以升高 UDP- 半乳糖的内源性水平 (Ross 等,2004)。此外,在该实施方案中,不存在 UDP- 半乳糖 4 差向异构酶将允许 UDP- 半乳糖水平可通过控制外源半乳糖浓度而得以调节,因为所产生的 UDP- 半乳糖除加入至底物诸如 N- 聚糖之外,其并不能被代谢。

[0203] 半乳糖通透酶是一种质膜己糖转运蛋白,其可从外源性来源中输入半乳糖。在一个实施方案中,使用来自啤酒糖酵母的半乳糖通透酶基因, GAL2 (Genbank :M81879),或编码

能够输入半乳糖的质膜己糖转运蛋白的任何基因。

[0204] 半乳糖激酶是一种催化半乳糖代谢第一步骤（即半乳糖至半乳糖-1-磷酸的磷酸化）的酶。在另一实施方案中，使用来自啤酒糖酵母的 GAL1 基因（Genbank :X76078）。

[0205] 半乳糖-1-磷酸尿苷酰转移酶催化半乳糖代谢的第二步骤，所述步骤是将 UDP-葡萄糖和半乳糖-1-磷酸转化为 UDP-半乳糖和葡萄糖-1-磷酸。在另一实施方案中，可以使用编码半乳糖-1-磷酸尿苷酰转移酶活性的任何基因，包括啤酒糖酵母 GAL7（Genbank : M12348）。

[0206] 在优选的实施方案中，从能够通过 LeLoir 途径代谢半乳糖的低等真核生物中缺失 UDP-半乳糖 4 差向异构酶编码基因。

[0207] 在更优选的实施方案中，在 gal(-) 并且不能内源性表达 LeLoir 途径基因的低等真核生物宿主细胞中，表达半乳糖通透酶、半乳糖激酶和半乳糖-1-磷酸尿苷酰转移酶的编码基因（Hittinger 等，2004）。

[0208] 该可选实施方案的优势在于 UDP-半乳糖 4-差向异构酶的不存在允许通过对低于生长抑制浓度水平的外部半乳糖进行调节来特异性控制内部 UDP-半乳糖浓度。

[0209] 在基因改变的酵母细胞中提高半乳糖基化 N-聚糖的生成

[0210] W000200879A3 和 W003056914A1 中提供了在酵母和真菌宿主中产生人样 N-聚糖的方法，所述文件被引入此处。本领域技术人员能够认识到本文所描述方法的常规修饰组合上述方法可以对目标糖蛋白的产生提供改进的结果。

[0211] 根据本发明的方法，用至少 β -半乳糖基转移酶融合构建体 pXB53（实施例 4）（图 12）转化的巴斯德毕赤氏酵母，在可检测部分产生了复合半乳糖基化聚糖。至少 10% 的半乳糖残基被转移至宿主细胞中的糖蛋白上。在另一实施方案中，至少 40% 的半乳糖残基被转移至宿主细胞中的糖蛋白上。差向异构酶的表达同样增加了半乳糖转移的水平（图 13）。优选地，至少 60% 的半乳糖残基被转移至宿主细胞中的糖蛋白上。另一异源糖基化酶诸如 UGT 的表达进一步增强了所需半乳糖化糖蛋白的细胞生产。令人惊奇地是，一种所述转运蛋白 DmUGT 的表达非常显著地增加了半乳糖转移（图 11）。在实施方案的最佳模式中，用 DmUGT 转化的宿主细胞显示出至少 90% 或更高的半乳糖转移。

[0212] 优选地，将酵母宿主细胞的温度保持在 37°C 以符合酶的最适温度。

[0213] 此外，所述方法还包括分离这些糖蛋白。

[0214] UDP 酶活性的表达

[0215] 如 W002/00879 中所述，在人中，核苷酸糖前体（例如 UDP-N-乙酰葡萄糖胺、UDP-N-乙酰半乳糖胺、CMP-N-乙酰基神经氨酸、UDP-半乳糖等）通常在胞质中合成并被转运至高尔基体中，在那里它们被糖基转移酶附着至核心寡糖。为了在低等真核生物中复制该过程，必须在高尔基体中表达糖核苷特异性转运蛋白以确保足够水平的核苷糖前体（Sommers, 1981 ;Sommers, 1982 ;Perez, 1987）。糖转移至 N-聚糖上的副产物是核苷二磷酸或一磷酸。虽然一磷酸可通过对向运输机制与核苷三磷酸糖交换而被直接输出，但是二磷酸核苷（例如，GDP）在被输出之前必须通过磷酸酶（例如，GDP 酶）进行切割以产生核苷一磷酸和无机磷酸。该反应对有效糖基化是重要的，因为已发现源自啤酒糖酵母的 GDP 酶对于甘露糖基化是必需的。但是，该酶对 UDP 的活性仅为 10%（Berninsone, 1994）。由于低等真核生物在高尔基体中的糖蛋白合成中不使用 UDP-糖前体，故低等真核生物一般在高尔

基体中不具有 UDP 特异性二磷酸酶活性。

[0216] 工程改造的酵母品系包含多个使用 UDP-GlcNAc 或 UDP- 半乳糖作为底物的转移酶。这需要在酵母高尔基体中工程改造合适的底物库,所述酵母高尔基体在大多数种类中不含这些底物。但是,使用 UDP-GlcNAc 或 UDP- 半乳糖的转移酶反应的终产物包括游离的 UDP。该 UDP 可作为使用这些糖核苷酸的大多数转移酶的潜在抑制剂发挥作用。啤酒糖酵母表达具有核苷二磷酸酶活性的两种高尔基体蛋白质。一种 (ScGDA1) 对 GDP 高度特异 (Abeijon 等,1993)。第二种 (ScYND1) 是腺苷三磷酸双磷酸酶,并因此能够水解核苷三磷酸和核苷二磷酸,且对 ADP/ATP、GDP/GTP 和 UDP/UTP 具有相等的特异性 (Gao 等,1999)。然而,因为在野生型高尔基体中缺少 UDP 缀合的糖,且同时缺少产生 UDP 作为终产物的转移酶,所以,在工程改造的酵母品系中 UDP 的可能升高的累积是一个值得关心的问题。

[0217] 因为缺少 UDP 酶会妨碍半乳糖残基从胞质转移至高尔基体,故对于在低等真核生物宿主细胞中有效的半乳糖转移需要基因操作以表达 UDP 酶。因此,在本发明的另一方面中,提供了表达,优选超表达编码 UDP 酶基因的方法。应当预期的是编码 UDP 酶活性的基因的超表达,增加了半乳糖转移至高尔基体中的受体底物上所需的糖核苷酸 UDP- 半乳糖的可用性。为了在酵母高尔基体中升高 UDP 酶活性水平,存在若干种可能性。在一个实施方案中,超表达编码 UDP 酶活性的基因,例如 ScGDA1 (NP_010872),其具有某些 (大约 10%) 针对 UDP 的活性。在另一实施方案中,超表达编码核苷二磷酸酶活性的基因,例如 ScYND1 (NP_010920),该酶对 UDP 的活性比对 GDP 的活性更高,尽管其对于核苷酸二磷酸并非具有特异性。此外,在另一实施方案中,为了在巴斯德毕赤氏酵母中实现较高 UDP 酶活性的目标,表达啤酒糖酵母 GDA1 或 YND1 或超表达这些基因的巴斯德毕赤氏酵母同源物,这些通过 BLAST 同源性检索可很容易得以鉴定。

[0218] 此外,使用这些糖核苷酸的生物能够通过特异针对 UDP 的核苷酸二磷酸酶作用,将 UDP 转化为 UMP。一个实例是由 Wang 以及 Guidotti 鉴定出的人尿苷二磷酸酶 (uridine diphosphatase) (UDP 酶) (AF016032)。然而,该蛋白质包含两个推定的跨膜结构域,一个位于 C- 端而一个位于 N- 端。因此,该蛋白质在酵母高尔基体中的定位,需要将该蛋白质催化结构域与酵母靶向结构域融合。

[0219] 包括乳克鲁维氏酵母和栗酒裂殖糖酵母的其它酵母在其高尔基体中使用 UDP- 糖将 GlcNAc 和半乳糖分别添加至其 N- 聚糖。乳克鲁维氏酵母和栗酒裂殖糖酵母都表达 ScGDA1 的同源物,其分别被命名为 KlGDA1 (Lopez-Avalos 等,2001 ;CAC21576) 和 Spgdal (D' Alessio 等,2003 ;NP593447),其同样具有 UDP 酶活性。在 UDP 在工程改造的酵母品系中累积且证明对工程改造的转移酶有害的情况下,任一或多种上述这些蛋白质的表达可用于增加 UDP 酶活性到可接受的水平。

[0220] 对无唾液酸糖蛋白受体 (ASGR) 的结合亲和性

[0221] 本发明的另一方面提供了针对 ASGR 较低的结合亲和性,ASGR 已知能够清除无唾液酸化糖蛋白以及降低循环系统中治疗性蛋白的半衰期。先前的研究已显示,具有双触角结构的聚糖的清除速度低于具有三或四触角结构的聚糖 (Stockert, Physiol Rev. 1995 Jul ;75 (3) :591-609)。在本发明的一个方面中,提供了具有单糖形的目标蛋白上的聚糖 (例如,双触角结构),其特征在于含有末端半乳糖残基。所述双触角结构并不易于在哺乳动物细胞中产生,因为其它 GnT 催化产生三 - 和四 - 触角分支的反应。通过用半乳糖残基

对含有末端 GlcNAc 残基的底物加帽,其它 GnT(例如, GnTIV、GnTV) 就不能再催化 GlcNAc 转移至半乳糖基化底物上。因此,本发明提供了一种方法,该方法用于产生无唾液酸化糖蛋白,所述无唾液酸化糖蛋白相对于在哺乳动物宿主中产生的那些糖蛋白而言具有针对 ASGR 的较低结合亲和性。在更优选的实施方案中,无唾液酸化糖蛋白的特征在于相对于在哺乳动物中产生的异源糖蛋白而言,其循环半衰期和体内生物活性增高,

[0222] 整合位点

[0223] 优选在基因座中整合编码 UGT、差向异构酶和 β 1,4GalT 的核酸,所述基因座负责甘露糖基转移酶,诸如 1,3 甘露糖基转移酶(例如啤酒糖酵母中的 MNN1)(Graham, 1991)、1,2 甘露糖基转移酶(例如来自啤酒糖酵母的 KTR/KRE 家族)、1,6 甘露糖基转移酶(来自啤酒糖酵母或巴斯德毕赤氏酵母的 OCH1)、甘露糖基磷酸转移酶及其调节物(来自啤酒糖酵母的 MNN4、PNO1 和 MNN6)、液泡蛋白酶(vacuolar proteinase)A(PEP4)、液泡蛋白酶 B(PRBI)GPI- 锚定的天冬氨酸蛋白酶(YPS1),以及涉及异常、免疫原性即非人糖基化反应的其它酶。

[0224] 基因座破坏的突变导致酶活性降低或酶活性完全消除的存活表型。优选地,编码初始 α -1,6 甘露糖基转移酶活性的基因座是编码糖基转移酶活性的基因初始整合的最佳靶位。以相似的方式,人们可以选择其它染色体整合位点的范围,所述整合位点基于在该基因座中的基因破坏事件,预期可以:(1) 改善细胞在更多人样型式中进行糖基化的能力,(2) 改善细胞分泌蛋白质的能力,(3) 降低外源蛋白质的蛋白酶解以及(4) 改善促进纯化过程或发酵过程本身的方法的其它特性。

[0225] 在特别优选的实施方案中,将文库 DNA 整合至宿主染色体非所需基因位点,从而对该基因的破坏或缺失造成影响。例如,整合至 OCH1、MNN1 或 MNN4 基因位点,使所需文库 DNA 能够表达,而抑制酵母中糖蛋白高甘露糖基化(hypermannosylation)所涉及的酶的表达。在其它实施方案中,文库 DNA 可通过核酸分子、质粒、载体(例如,病毒或逆转录病毒载体)、染色体被导入至宿主,并可作为自主核酸分子或通过同源或随机整合至宿主基因组而导入。在任何情况下,通常需要每一文库 DNA 构建体都包含至少一个选择标记基因以便于选择稳定转化的宿主生物。尤其适用的是可再循环标记基因,诸如 URA5(Yeast. 2003 Nov ; 20(15):1279-90),其可被选择或针对其进行选择。

[0226] 产生额外序列多样性

[0227] 当转化至宿主的核酸,例如 DNA 文库包含大量多样性序列,由此增加至少一个转化体表现出所需表型的概率时,本实施方案的方法是最有效的。例如,单氨基酸突变可巨大地改变糖蛋白加工酶的活性(Romero 等,2000)。因此,在转化之前,可对 DNA 文库或组分亚文库实施一种或多种技术以产生额外序列多样性。例如,可以实施一轮或多轮的基因改组(gene shuffling)、易错 PCR、体外诱变或产生序列多样性的其它方法,以获得在融合构建体库中序列的高多样性。

[0228] 密码子最优化

[0229] 还应当预期,还可以对本发明核酸实施导致一级氨基酸序列一种或多种改变的密码子最优化,例如保守氨基酸取代、添加、缺失或其组合。

[0230] 表达控制序列

[0231] 除了上述可读框序列之外,通常优选给每一文库提供含有表达控制序列的构建

体,所述表达控制序列诸如启动子、转录终止子、增强子、核糖体结合位点以及确保融合构建体在转化至宿主生物后融合蛋白的有效转录和翻译所需的其它功能性序列。

[0232] 适宜的载体组分,例如选择标记、表达控制序列(例如,启动子、增强子、终止子等)以及,任选地,宿主细胞自主复制所需序列,可根据所选特定宿主细胞而进行选择。用于特定哺乳动物或低等真核生物宿主细胞的适宜载体组分的选择标准是常规的。本发明优选的低等真核生物宿主细胞包括,巴斯德毕赤氏酵母、*Pichia finlandica*、喜海藻糖毕赤氏酵母(*Pichia trehalophila*)、*Pichiakoclamae*、膜醭毕赤氏酵母(*Pichia membranaefaciens*)、*Pichia minuta*(*Ogataeaminuta*、*Pichia lindneri*)、*Pichia opuntiae*、*Pichia thermotolerans*、柳毕赤氏酵母(*Pichia salictaria*)、*Pichia guercuum*、皮杰普氏毕赤氏酵母(*Pichia pijperi*)、具柄毕赤氏酵母(*Pichia stiptis*)、甲醇毕赤氏酵母、毕赤氏酵母属物种(*Pichiasp.*)、啤酒糖酵母、糖酵母属物种(*Saccharomyces sp.*)、多形汉逊氏酵母、克鲁维氏酵母属物种(*Kluyveromyces sp.*)、乳克鲁维氏酵母、白色假丝酵母(*Candida albicans*)、构巢曲霉(*Aspergillus nidulans*)、黑色曲霉、米曲霉(*Aspergillus oryzae*)、里氏木霉(*Trichoderma reesei*)、*Chrysosporium lucknowense*、镰孢属物种(*Fusarium sp.*)、禾赤镰孢(*Fusarium gramineum*)、*Fusarium venenatum*、*Physcomitrella patens* 和粗糙链孢霉。

[0233] 当宿主是巴斯德毕赤氏酵母时,适宜的启动子包括,例如,AOX1、AOX2、GAPDH、OCH1、SEC4、D2 和 P40 启动子。

[0234] 选择标记

[0235] 还优选为每一构建体提供至少一个选择标记,诸如赋予抗药性的基因或补足宿主代谢损伤的基因。标记的存在可用于随后转化体的选择;例如在酵母中,可使用 URA5、URA3、HIS4、SUC2、G418、BLA 或 SH BLE 基因。多种选择标记是已知的并适用于酵母、真菌、植物、昆虫、哺乳动物和其它真核生物宿主细胞。

[0236] 转化

[0237] 在酵母菌中,可以使用任何方便的 DNA 转移方法,诸如电穿孔、氯化锂法或原生质球法。在丝状真菌和植物细胞中,方便的方法包括粒子轰击、电穿孔和农杆菌介导的转化。为了制备适于高密度培养(例如,酵母发酵)的稳定品系,需要将融合构建体整合至宿主染色体中。在优选的实施方案中,采用本领域众所周知的技术,通过同源重组进行整合。优选地,通过双交换事件对巴斯德毕赤氏酵母进行稳定的基因修饰。Nett 等, *Yeast*. 2003 Nov; 20(15):1279-90。例如,给异源酶活性提供与宿主生物序列同源的侧翼序列,接着使用单一标记进行转化。以这种方式,整合发生在利用可再循环标记的宿主基因组的确定位点。

[0238] 筛选和选择方法

[0239] 用异源酶转化宿主品系后,表现所期望糖基化表型的转化体被选择。选择可采用众多测定或检测方法中的任一种,通过单一步骤或一系列表型富集和/或排除步骤来进行。可以手动或采用自动化高通量筛选设备进行表型表征。通常,宿主微生物在细胞表面(其上定位有各种糖蛋白)展示蛋白质 N-聚糖。

[0240] 例如,人们可以筛选细胞表面有最高浓度末端 GlcNAc 的那些细胞,或可以筛选能分泌具有最高末端 GlcNAc 含量的蛋白的那些细胞。所述筛选可以基于可视化方法,如基于染色方法,基于结合特异性末端 GlcNAc 结合抗体或缀合至标记的凝集素(诸如可购自 E.

Y. Laboratories Inc., San Mateo, CA 的凝集素) 的能力, 基于特异性凝集素结合末端甘露糖残基降低的能力, 基于体外整合放射性标记的糖的能力, 或基于与染料或带电表面结合的改变, 或者可通过荧光辅助细胞分选 (Fluorescence Assisted Cell Sorting (FACS)) 设备联合荧光团标记的凝集素或抗体 (Guillen, 1998) 来实现。

[0241] 因此, 可通过将细胞暴露于特异性结合所需 N- 聚糖的凝集素或抗体, 而对完整细胞筛选所需的糖基化表型。多种寡糖特异性凝集素可在商业上获得 (例如, 购自 EY Laboratories, San Mateo, CA)。可选地, 针对特异性人或动物 N- 聚糖的抗体可在商业上获得或可采用标准技术制得。适宜的凝集素或抗体可以缀合至报告分子, 诸如生色团、荧光团、放射性同位素或具有产色底物的酶 (Guillen 等, 1998. Proc. Natl. Acad. Sci. USA 95(14):7888-7892)。

[0242] 接着, 可采用分析方法, 诸如分光光度法、荧光测定法、荧光激活细胞分类术或闪烁计数实施筛选。在其它情况中, 需要分析从转化的细胞中分离的糖蛋白或 N- 聚糖。可采用本领域已知的技术实施蛋白质分离。在优选的实施方案中, 报告蛋白质被分泌至培养基中, 然后通过亲和层析 (例如 Ni- 亲和或谷胱甘肽-S- 转移酶亲和层析) 进行纯化。在优选分离的 N- 聚糖的情况中, 酶诸如内切 -b-N- 乙酰葡萄糖胺糖苷酶 (Genzyme Co., Boston, MA; New England Biolabs, Beverly, MA) 可用于从糖蛋白切割 N- 聚糖。然后, 可通过液相层析 (例如 HPLC)、质谱法或其它适宜的方法, 分析分离的蛋白质或 N- 聚糖。美国专利号 5, 595, 900 教导了几种可用于鉴定具有所需胞外糖结构的细胞的方法。在优选的实施方案中, MALDI-TOF 质谱分析法用于分析切割的 N- 聚糖。

[0243] 在选择所需转化体之前, 需要排除具有非所需表型的细胞转化群体。例如, 当所述方法用于将功能性甘露糖苷酶活性工程改造至细胞中时, 所需转化体细胞糖蛋白将有低水平的甘露糖。将转化的群体在培养基中暴露于甘露糖的致死放射性同位素排除了具有非所需表型即高水平掺入的甘露糖的转化体群体 (Huffaker TC 和 Robbirs PW., Proc Natl Acad Sci USA. 1983 Dec; 80(24):7466-70)。可选地, 针对非所需 N- 聚糖的细胞毒性凝集素或抗体, 可被用于排除具有非所需表型的转化群体 (例如, Stanley P 和 Siminovitch L. Somatic Cell Genet 1977 Jul; 3(4):391-405)。美国专利号 5, 595, 900 教导了几种可用于鉴定具有所需胞外糖结构的细胞的方法。重复实施该策略使得能够实现低等真核生物越来越多复合聚糖的顺序工程改造。

[0244] 例如, 为了检测在其表面上具有高程度人样 N- 聚糖中间体 $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ 的宿主细胞, 人们可以选择这样的转化体, 所述转化体在体外细胞测定中允许半乳糖被 GalT 从 UDP- 半乳糖最有效的转移。该筛选可通过在选择压力下在琼脂平板上生长含有转化的文库的细胞以及将单个克隆转移至 96- 孔微量滴定板而进行。细胞生长后, 离心细胞, 在缓冲液中重悬细胞, 并在加入 UDP- 半乳糖和 GalT 之后, 通过 HPLC 或 UDP 酶联测定确定 UDP 的释放。可选地, 人们可以使用放射性标记的 UDP- 半乳糖和 GalT, 洗涤细胞, 然后观测由 β - 半乳糖苷酶导致的放射性半乳糖的释放。所有这些均可以手动或通过使用高通量筛选设备自动进行。在第一种测定法中释放更多 UDP 的转化体, 或在第二种测定法中释放更多放射性标记的半乳糖的转化体, 预期在其表面上有较高度度的 $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$, 且由此构成所需表型。相似的测定法同样适用于观测分泌的蛋白质上的 N- 聚糖。

[0245] 可选地,人们可以使用任何其它适宜的筛选,例如凝集素结合测定,所述方法能揭示转化的细胞表面上改变的糖基化模式。在该情况中,凝集素与末端甘露糖特异性结合的降低可以作为适宜的选择工具。雪化莲 (*Galantusnivalis*) 凝集素特异性结合末端 α -1,3 甘露糖,如果高尔基体中存在充足的甘露糖苷酶 II 活性则所述结合预期被降低。人们还可以通过实施层析分离步骤以允许除去具有高末端甘露糖含量的细胞从而富集所需转化体。可采用凝集素柱实施所述分离步骤,所述凝集素柱对具有高末端甘露糖含量的细胞的结合特异性超过对具有低末端甘露糖含量的那些细胞的结合特异性(例如,结合于琼脂糖的雪化莲凝集素, Sigma, St. Louis, MO)。

[0246] 宿主细胞

[0247] 虽然本发明采用巴斯德毕赤氏酵母作为宿主生物进行举例说明,但本领域技术人员能够理解的是,其它真核生物宿主细胞,包括其它酵母和真菌宿主物种,可按照本文所述进行改变以用于产生人样糖蛋白。所述宿主优选包括 *Pichia finlandica*、喜海藻糖毕赤氏酵母、*Pichia koclamae*、膜醭毕赤氏酵母、*Pichia minuta* (*Ogataea minuta*、*Pichia lindneri*)、*Pichia opuntiae*、*Pichiathermotolerans*、柳毕赤氏酵母、*Pichia guercuum*、皮杰普氏毕赤氏酵母、具柄毕赤氏酵母、甲醇毕赤氏酵母、毕赤氏酵母属物种、啤酒糖酵母、糖酵母属物种、多形汉逊氏酵母、克鲁维氏酵母属物种、乳克鲁维氏酵母、白色假丝酵母、构巢曲霉、黑色曲霉、米曲霉、里氏木霉、*Chrysosporium lucknowense*、镰孢属物种、禾赤镰孢、*Fusarium venenatum*、*Physcomitrella patens* 和粗糙链孢霉。

[0248] 本文所描述的用于鉴定和破坏非所需宿主细胞糖基化基因(例如 OCH1)的技术,应被理解为可用于其它真核生物宿主细胞,诸如其它酵母和真菌品系中这些基因和/或其它同源性或功能性相关基因(参见 W002/00879)。此外,其它优选的宿主细胞在 *Alg3p* 中有缺陷,其编码 $\text{Dol-P-Man:Man}_5\text{GlcNAc}_2\text{-PP-Dol}$ 甘露糖基转移酶活性(参见 W003/056914)。

[0249] 优选的宿主细胞是酵母和丝状真菌宿主,其固有缺少 β 1,4-半乳糖键、岩藻糖和末端唾液酸。与哺乳动物糖蛋白的 N-聚糖不同,这些糖通常不存在于酵母和丝状真菌所产生的糖蛋白上。本发明提供了工程改造宿主细胞以在糖蛋白上产生半乳糖残基并且在所述糖蛋白上基本缺少岩藻糖和唾液酸残基的方法。在另一实施方案中,产生岩藻糖或唾液酸的那些宿主细胞可经过修饰以具有降低的或消除的岩藻糖基转移酶活性或唾液酸转移酶活性。因此,由本发明宿主所产生的糖蛋白组合物基本上不含岩藻糖和唾液酸残基。本发明显著的优势在于,无须转体内 (*ex vivo*) 用岩藻糖苷酶和唾液酸酶处理进行修饰,宿主细胞即可产生半乳糖基化、无岩藻糖和无唾液酸的糖蛋白。

[0250] 其它优选的宿主细胞包括缺少关于聚糖的甘露糖基磷酸化的真菌宿主 (US 11/020,808)。而其它优选的宿主细胞包括缺少关于聚糖的 β -甘露糖基化的真菌宿主 (US 60/566,736)。

[0251] 因此,本发明的另一方面涉及表达糖蛋白的非人真核宿主品系,所述糖蛋白包含修饰的 N-聚糖,其与人类细胞所产生的相似。因此,本发明预期且包括在除酵母和真菌细胞之外的物种中实施本发明的方法。应当预期,本发明的组合核酸文库可被用于选择在任何真核宿主细胞系统中修饰糖基化途径的构建体。例如,本发明组合文库还可被用于植物、藻类和昆虫以及包括哺乳动物和人类细胞的其它真核宿主细胞中,以将包括糖基化酶或其催化结构域的蛋白质定位在沿宿主细胞分泌途径的所需位置上。优选地,将糖基化酶或催

化结构域等靶向沿宿主细胞分泌途径的亚细胞位置,在那里它们能发挥功能,且优选它们被设计或选择以最有效地发挥功能。

[0252] 能用根据本发明实施方案的方法改变的糖基化修饰的实例是:(1) 工程改造真核宿主细胞,以从 $\text{Man}_8\text{GlcNAc}_2$ 修整甘露糖残基以产生 $\text{Man}_5\text{GlcNAc}_2\text{N}$ -聚糖;(2) 工程改造真核宿主细胞,以通过 GlcNAc 转移酶 I 的作用在 $\text{Man}_5\text{GlcNAc}_2$ 上添加 N-乙酰葡萄糖胺 (GlcNAc);(3) 工程改造真核宿主细胞,以功能性表达酶,诸如 N-乙酰葡萄糖胺基转移酶 (GnTI、GnTII、GnTIII、GnTIV、GnTV、GnTVI、GnTIX)、甘露糖苷酶 II、岩藻糖基转移酶 (FT)、半乳糖基转移酶 (GalT) 或唾液酸转移酶 (ST)。

[0253] [2S3] 通过重复所述方法,更加复杂的糖基化途径可在靶宿主,诸如低等真核微生物中被工程改造。在一个优选的实施方案中,用包含编码糖基化活性的序列的 DNA 文库转化宿主生物两次或更多次。可以在每轮转化之后或可选地在已进行若干次转化之后,实施所需表型选择。以这种方式可以对复合糖基化途径进行快速工程改造。

[0254] 靶糖蛋白

[0255] 本文所述的方法可用于产生糖蛋白,尤其是治疗性用于人的糖蛋白。例如,具有特定糖形的糖蛋白在治疗性蛋白靶向中尤为有用。例如,甘露糖 -6- 磷酸已经显示出可指导蛋白质至溶酶体,所述溶酶体是几种溶酶体贮积病相关酶的正确功能所必需的,所述溶酶体贮积病诸如 Gaucher 氏病、Hunter 氏病、Hurler 氏病、Scheie 氏病、Fabry 氏病和 Tay-Sachs 病,所描述的仅为少数几个。同样,在聚糖侧链上加入一个或多个唾液酸残基,可以增加治疗性糖蛋白施用后的体内半衰期。因此,宿主细胞(例如,低等真核生物或哺乳动物)可经过基因工程改造以增加细胞所表达的糖蛋白中末端唾液酸的程度。可选地,在施用前,可以采用唾液酸转移酶和适宜的底物将唾液酸与目标蛋白质体外缀合。

[0256] 除了表达人样糖基化所涉及的酶活性之外,可对生长培养基组成进行改变以产生更接近地类似人形式的糖蛋白(S.Weikert 等, *Nature Biotechnology*, 1999, 17, 1116-1121; Werner, Noe 等, 1998 *Arzneimittelforschung* 48(8):870-880; Weikert, Papac 等, 1999; Andersen 和 Goochee 1994 *Cur. Opin. Biotechnol.* 5:546-549; Yang 和 Buttlner 2000 *Biotechnol. Bioengin.* 68(4):370-380)。对单克隆抗体的特异性聚糖修饰(例如加入二等分 GlcNAc)已经显示,可以改善依赖抗体的细胞毒性(UmanaP. 等人, 1999),该改善是抗体或其它治疗性蛋白质生产所需的。

[0257] 治疗性蛋白质通常采用注射、经口、经肺或其它方式施用。可根据本发明生产的适宜靶糖蛋白的实例包括,但不限于:促红细胞生成素,细胞因子诸如干扰素 -a、干扰素 -b、干扰素 -g、干扰素 -w, 和粒细胞-CSF、GM-CSF,凝固因子诸如因子 VIII、因子 IX, 和人 C 蛋白、抗凝血酶 III、凝血酶、可溶性 IgE 受体 α -链、IgG、IgG 片段、IgG 融合体、IgM、白细胞介素、尿激酶、糜蛋白酶,和胰凝乳蛋白酶抑制剂、IGF-结合蛋白、表皮生长因子、生长激素释放因子、膜联蛋白 V 融合蛋白、制管张素、血管内皮生长因子 -2、骨髓祖先抑制因子 -1(myeloid progenitor inhibitory factor-1)、骨保护素(osteoprotegerin)、 α -1-抗胰蛋白酶、甲胎蛋白、DNA 酶 II、人纤溶酶原 kringle 3、葡糖脑苷脂酶、TNF 结合蛋白 1、促卵泡激素、细胞毒性 T 淋巴细胞相关抗原 4-Ig、跨膜激活剂及钙调节剂及亲环蛋白(cyclophilin) 配体、可溶性 TNF 受体 Fc 融合体、胰高血糖素样蛋白 1 以及 IL-2 受体激动剂。

[0258] 分泌性信号序列

[0259] 还优选将编码分泌性信号的核酸序列与编码糖蛋白的目标序列相结合。术语“分泌性信号序列”是指编码多肽(“分泌性肽”)的DNA序列,所述多肽作为更大多肽的组分指导该更大多肽通过合成所述多肽的细胞的分泌途径。所述更大多肽在通过分泌途径转运期间通常被切割以除去所述分泌性肽。为了将多肽指导至宿主细胞的分泌途径中,在表达载体中提供分泌性信号序列(也称为前导序列、前原序列(prepro sequence)或前序列)。分泌性信号序列可以是(但不限于)与糖蛋白相关的野生型序列、啤酒糖酵母 Suc2 信号序列的编码序列、毕赤氏酵母属 Pho2 信号序列的编码序列、毕赤氏酵母属 Prc1 信号序列的编码序列、啤酒糖酵母 α -交配因子(α MF)信号序列的编码序列、牛溶菌酶 C 信号序列的编码序列。分泌性信号序列可操作地连接于核酸序列,亦即两个序列以正确读框连接,并进行定位以指导新合成多肽进入宿主细胞的分泌途径。分泌性信号序列通常位于编码目标多肽的DNA序列的5',尽管某些分泌性信号序列位于目标DNA序列的其它位置(参见例如,Welch等,美国专利号 5,037,743;Holland等,美国专利号 5,143,830)。

[0260] 可选地,本发明多肽所包含的分泌性信号序列可用于指导其它多肽进入分泌途径。本发明提供了所述融合多肽。本发明融合多肽中所包含的分泌性信号序列,优选与其它肽的氨基末端融合以指导该其它肽进入分泌途径。所述构建体可用于本领域已知的多种应用中。例如,这些新分泌性信号序列融合构建体可指导正常非分泌蛋白活性组分的分泌,例如受体。所述融合体可在体内或体外使用,以指导肽通过分泌途径。

[0261] 本发明方法所产生的糖蛋白可通过本领域众所周知的技术进行分离。所需糖蛋白通过诸如分级分离、离子交换、凝胶过滤、疏水性层析以及亲和层析的方法进行纯化和分离。

[0262] 以下是说明本发明组合物和方法的实施例。这些实施例不应被理解为是限制性的;包含该实施例的目的仅在于举例说明。

[0263] 实施例 1**[0264] 构建启动子盒和表达载体**

[0265] 使用引物 RCD48(SEQ ID NO:15)(5'-TATGCGGCCGCGGCTGATGATATTTGCTACGA-3')和 RCD134(SEQ ID NO:16)(5'-CCTCTCGAGTGGACACAGGAGACTCAGAAACAG-3'),扩增 PpOCH1 基因 800bp 的启动子,和使用引物 RCD156(SEQ ID NO:17)(5'-CTTCTCGAGGAAGTAAAGTTGGCGAAACTT-3')和 RCD157(SEQ ID NO:18)(5'-CTTAGCGGCCGCGATTGTTCTGTTTGTAGTTT-3'),扩增 PpSEC4 基因 400bp 的启动子。PCR 产物被克隆到 pCR2.1 克隆载体(Invitrogen)并测序。然后使用导入的 XhoI/NotI 限制位点,将 OCH1 和 SEC4 启动子亚克隆到载体 pJN261(Nett 等, Yeast. 2003 Nov;20(15):1279-90)中取代 GAPDH 启动子,以分别生成质粒 pRCD360 和 pRCD362。

[0266] 使用引物 RCD 152(SEQ ID NO:19)(5'-CTTCTCGAGGGCATTCAAAGAAGCCTTGGG-3')和 RCD153(SEQ ID NO:20)(5'-CTTAGCGGCCGCTGAGTGGTCATGTGGGAACCTT-3')PCR 扩增 PpHIS3 启动子,然后克隆到质粒 pCR2.1 并测序。然后将 XhoI/NotI 位点用于将 PpHIS3 启动子亚克隆到质粒 pTA18,以用较弱的 PpHIS3 启动子替换 PpPMA1 强启动子并生成质粒 pRCD351,该质粒是滚动入(roll into)PpHIS3 启动子的 NAT^R 质粒。

[0267] 使用引物 RCD301(SEQ ID NO:21)(5'-CCTGGATCCAACAGACTACAATGACAGGAG-3')

') 和 RCD302 (SEQ ID NO :22) (5' -CCTGCATGCCTCGAGCTTGCCGGCGTCTAAATAGCCGTTGAA G-3') 扩增部分 PpHIS3 基因, 并使用 BamHI/SphI 限制位点将其插入至 pUC19 以生成质粒 pRCD391。该载体包含 PpHIS3 基因座的 1.2Kb 部分以及工程改造导入至引物 RCD302 (SEQ ID NO :22) 内的 XhoI 和 NgoMIV 位点。G418^R 基因作为来自 pUG6 (Wach 等, 1994) 的 BglIII/SacI 片段插入至 pRCD391 的 BamHI/SacI 位点以生成 pRCD392。

[0268] 用引物 RCD307 (SEQ ID NO :23) (5' -CCTGTCGACGCTGCCGGCAAGCTCGAGTTTAAGCGGTG CTGC-3') 和 RCD308 (SEQ ID NO :24) (5' -CCTGGATCCTTTGGCAAAAACCAGCCCTGGTGAG-3') 从巴斯德毕赤氏酵母基因组 DNA 中扩增 1.2Kb 部分的 PpTRP1 基因。使用 BamHI/SalI 位点将扩增的片段插入到 pUC19 以生成质粒 pRCD399。使用 BglIII/SacI 从质粒 pAG29 (Goldstein 和 McCusker, 1999) 释放赋予对膈丝菌素抗性的 PAT 基因, 并将其插入到用 BamHI/SacI 消化的 pRCD399 以生成 PpTRP1/PAT roll-in 质粒 pRCD401。

[0269] 实施例 2

[0270] 克隆半乳糖转运蛋白

[0271] 粟酒裂殖糖酵母 UDP 半乳糖转运蛋白

[0272] 以两个片段从粟酒裂殖糖酵母基因组 DNA (ATCC24843) PCR 扩增编码 UDP 半乳糖转运蛋白 (SpGMS1+, Genbank AL022598) 的被称为 SpUGT 的粟酒裂殖糖酵母基因, 以除去单个内含子。使用引物 RCD164 (SEQ ID NO :7) (5' -CCTTGCGGCCGCATGGCTG TCAAGGGCGACGATGTCAAA-3') 和 RCD177 (SEQ ID NO :8) (5' -ATTCGAGAATAGTTAAGTGTCAAAA TCAATGCACTATTTT-3') 扩增基因的 5' 端 96bp, 且使用引物 RCD176 (SEQ ID NO :9) (5' -AA AATAGTGCAATTGATTTGACACTTAACCTATTCTCGAAT-3') 和 RCD165 (SEQ ID NO :10) (5' -CCTTTTA ATTAATTAATGCTTATGATCAACGTCCTTAGC-3') 扩增 3' 端 966bp。随后, 使用引物 RCD164 (SEQ ID NO :7) 和 RCD165 (SEQ ID NO :10) 将两个扩增的产物重叠成单个 PCR 片段, 所述 PCR 片段包含一个在末端导入了 NotI 和 PacI 位点的连续可读框。将该 PCR 产物克隆到 pCR2.1 载体 (Invitrogen) 中并测序。然后使用 NotI 和 PacI 位点将该基因亚克隆至质粒 pJN335 中, 所述质粒包含融合有巴斯德毕赤氏酵母 GAPDH 启动子下游基因的表达盒。然后使用引物 RCD202 (SEQ ID NO :25) (5' -TCCTTAATTAAAGAAAGCTAGAGTAAAATAGAT-3') 和 RCD203 (SEQ ID NO :26) (5' -TCCCTCGAGGATCATGTTGATCA ACTGAGACCG-3') PCR 扩增 400bp PpOCH1 转录终止子, 并克隆到 pCR2.1 中。随后, 实施三重连接以作为 XhoI/PacI 片段将 GAPDH 启动子 / SpUGT 基因融合体以及作为 PacI/XhoI 片段将 PpOCH1-TT 插入至质粒 pTA18 的单 XhoI 位点, 以生成质粒 pRCD257。新质粒 pRCD257, 是含 NAT^R 的载体, 该载体包含 GAPDH-SpGALE-OCH1TT 融合体及第二个盒, 所述第二个盒包含由 PpPMAI 启动子所驱动、与 ScVAN1 跨膜结构域相融合的截短型人 GnTIII 基因。

[0273] SpUGT 基因也被插入到 pRCD360 (具有 OCH1 启动子) 以及 pRCD362 (具有 SEC4 启动子) 的 NotI/PacI 位点以分别生成质粒 pRCD385 和 pRCD387。使用 XhoI/NgoMIV 将来自 pRCD385 的 P_{OCH1}-SpUGT-PpCYC1TT 盒和来自 pRCD387 的 P_{SEC4}-SpUGT-PpCYC1TT 盒插入到 pRCD392HIS3/G418^Rroll-in 载体, 以分别生成巴斯德毕赤氏酵母 HIS3/G418^Rroll-in 表达质粒 pRCD393 和 pRCD394。

[0274] 黑腹果蝇 UDP 半乳糖转运蛋白

[0275] 从黑腹果蝇 cDNA 文库 (UCBerkeley Drosophila Genome Project, 卵巢 1-ZAP

文库 GM)PCR 扩增编码 UDP 半乳糖转运蛋白的称为 DmUGT 的黑腹果蝇基因 (Genbank BAB62747), 并将其克隆到 pCR2.1 PCR 克隆载体并测序。使用引物 DmUGT-5' (SEQ ID NO:11) (5' -GGCTCGAGCGGCCGCCACCATGAATAGCATACACAT-GAACGCCAATACG-3') 和 DmUGT-3' (SEQ ID NO:12) (5' -CCCTCGAGTTAAT-TAACTAGACGCGCGGCAGCAGCTTCTCCTCATC G-3') 扩增所述基因, 其分别在 5' 末端和 3' 末端导入 NotI 和 PacI 位点。然后使用 NotI 和 PacI 位点将融合于 PpOCH1 及启动子下游的该基因亚克隆至 pRCD393 NotI/PacI 位点以生成质粒 pSH263。

[0276] 智人 UDP 半乳糖转运蛋白

[0277] 从人前列腺 cDNA (marathon ready dDNA, Clontech) 扩增编码 UDP 半乳糖转运蛋白 1 (Genbank#BAA95615) 和 UDP 半乳糖转运蛋白 2 (Genbank#BAA95614) 的智人基因, 该基因分别称为 hUGT1 和 hUGT2。使用引物 hUGT1-5' (SEQ ID NO:27) (5' -GGCTCGAGCGGCCGCC ACCATG-GCAGCGGTTGGGGCTGGTGGTTC-3') 和 hUGT1-3' (SEQ ID NO:28) (5' -CC-CTCGAGTT AATTAATCACTTCACCAGCACTGACTTTGGCAG-3') 扩增 hUGT1 基因, 并使用引物 hUGT2-5' (SEQ ID NO:29) (5' -GGGCTC-GAGCGGCCGCCACCATGGCAGCGGTTGGGGCTGGTGGTTC-3') 和 hUGT2-3' (SEQ ID NO:30) (5' -CCCTCGAGTTAATTAAGTAACTAGGAACCTTCACCTTG-GTGAGCAAC-3') 扩增 hUGT2 基因。将 PCR 产物克隆到 pCR2.1 载体 (Invitrogen, Carlsbad, CA) 并测序。随后使用 NotI/PacI 将 hUGT1 和 hUGT2 基因插入到 pRCD393 的 PpOCH1 启动子的下游, 以分别生成质粒 pSH264 和 pSH262。

[0278] 实施例 3

[0279] 克隆 UDP- 半乳糖 4- 差向异构酶基因

[0280] 啤酒糖酵母 UDP- 半乳糖 4- 差向异构酶

[0281] 使用引物 RCD270 (SEQ ID NO:31) (5' -TAGCGGCCGCATGACAGCTCAGTTACAAAGTGAAA G-3') 和 RCD271 (SEQ ID NO:32) (5' -CGTTAATTAATCAGGAAAATCTGTAGACAATCTTGG-3') 从啤酒糖酵母基因组 DNA PCR 扩增编码 UDP- 半乳糖 4- 差向异构酶的啤酒糖酵母基因 (ScGAL10)。将得到的 PCR 产物克隆到 pCR2.1 并测序。

[0282] 然后使用 NotI/PacI 位点将 ScGAL10 基因分别亚克隆到质粒 pRCD393 和 pRCD394, 以生成质粒 pRCD395 和 pRCD396, 并且同样将其分别亚克隆到质粒 pRCD402 和 pRCD403, 以生成质粒 pRCD404 和 pRCD405。质粒 pRCD402 和 pRCD403 是表达载体, 它们分别包含巴斯德毕赤氏酵母 OCH1 和 SEC4 启动子, 以及 PpCYC1 终止子以及方便的限制位点, 所述位点可用于将差向异构酶与这些启动子融合以及可用于生成能整体移动至另一质粒的盒。

[0283] 智人 UDP- 半乳糖 4- 差向异构酶

[0284] 使用分别带有 NotI 和 PacI 位点的引物 GD7 (SEQ ID NO:33) 和 GD8 (SEQ ID NO:34), 从人肾 cDNA (marathon ready cDNA, Clontech) PCR 扩增编码 UDP- 半乳糖 4- 差向异构酶的称为 hGALE 的智人基因 (Thoden 等, (2001) JBC Vol. 276 (18) 15131-15136), 将其克隆到 pCR2.1 并测序。然后使用 NotI/PacI 位点将 hGALE 基因亚克隆到质粒 pRCD406 和 pRCD407, 以分别生成质粒 pRCD427 和 pRCD428。

[0285] 粟酒裂殖糖酵母 UDP- 半乳糖 4- 差向异构酶

[0286] 使用引物 GALE2-L (SEQ ID NO:35) 和 GALE2-R (SEQ ID NO:36), 从粟酒裂殖糖酵母 (ATCC24843) 基因组 DNA 扩增 SpGALE 基因。将扩增的产物克隆到 pCR2.1 并测序。测序

显示在 +66 位存在内含子 (175bp)。

[0287] 为了除去内含子,设计上游引物 GD1 (SEQ ID NO :37) (94 个碱基)。其具有 NotI 位点 (内含子上游 66 个碱基),随后是内含子之前的 20 个碱基。GD2 (SEQ ID NO :38) 是下游引物而且具有 PacI 位点。使用引物 GD1 (SEQ ID NO :37) 和 GD2 (SEQ ID NO :38) 从 pCR2.1 亚克隆扩增 SpGALF 无内含子基因,并将产物再次克隆到 pCR2.1 并测序。

[0288] 实施例 4

[0289] 克隆 b-1,4- 半乳糖基转移酶基因

[0290] 智人 b-1,4- 半乳糖基转移酶 I

[0291] 使用引物 RCD192 (SEQ ID NO :1) (5' -GCCGCGACCTGAGCCGCTGCCCCAAC-3') 和 RCD186 (SEQ ID NO :2) (5' -CTAGCTCGGTGTCCCAGTGTCCACTGT-3'), 从人肾 cDNA (marathon ready cDNA, Clontech) PCR 扩增智人 b-1,4- 半乳糖基转移酶 I 基因 (hGalTI, Genbank AH003575)。将该 PCR 产物克隆到 pCR2.1 载体 (Invitrogen, Carlsbad, CA) 并测序。从该克隆实施 PCR 重叠诱变以实现以下三目的:1) 除去可读框内的 NotI 位点同时保持野生型蛋白序列,2) 在紧靠内源跨膜结构域下游处截短蛋白,3) 以及在 5' 和 3' 末端导入 AscI 和 PacI 位点以用于模块克隆。为此,使用引物 RCD198 (SEQ ID NO :3) (5' -CTAGGCGCGCCGCGCGACCTGAGCCGCTGCCC-3') 和 RCD201 (SEQ ID NO :4) (5' -GGGGCATATCTGCCGCCATC-3') 扩增基因 5' 末端直至 NotI 位点,并使用引物 RCD200 (SEQ ID NO :5) (5' -GATGGGCGGCAGATATGCCCC-3') 和 RCD199 (SEQ ID NO :6) (5' -CTTCTTAATTAAGTAGCTCGGTGTCCCAGTGTCCAC-3') 扩增 3' 末端。用引物 198 和 199 将产物重叠在一起以重新合成具有野生型氨基酸序列的 ORF,同时除去了 NotI 位点。新截短的 hGalTIPCR 产物被克隆到 pCR2.1 载体 (Invitrogen, Carlsbad, CA) 并测序。然后使用导入的 AscI/PacI 位点将片段亚克隆到质粒 pRCD259 (其是 PpURA3/HYG^Rroll-in 载体),以生成 pRCD260。将 W002/00879 中所记载的 (引入作为参考) 酵母靶向序列跨膜结构域文库连接到位于 hGalTI 基因上游的 NotI/AscI 位点,以生成质粒 pXB20-pXB67。

[0292] 智人 b-1,4- 半乳糖基转移酶 II

[0293] 使用引物 RCD292 (SEQ ID NO :39) (5' -CTAGGCGCGCCCAGCACCTGGCCTTCTTCAGC-3') 和 RCD293 (SEQ ID NO :40) (5' -CTTGTTAATTAATCAGCCCCGAGGGGGCCACGACGG-3'), 从人肾 cDNA (marathon ready cDNA, Clontech) PCR 扩增截短型智人 b-1,4- 半乳糖基转移酶 II 基因 (hGalTII, Genbank AF038660), 将该基因克隆到质粒 pCR2.1 并测序。使用导入的 AscI/PacI 位点将该截短克隆 (除去了编码 N- 末端跨膜结构域的基因部分) 亚克隆到载体 pXB53 以替代 hGalTI, 以生成质粒 pRCD378。该质粒包含截短的 hGalTII 与啤酒糖酵母 MNN2 基因的跨膜结构域 / 前导序列编码部分的基因融合体, 该质粒由 PpGAPDH 启动子驱动。

[0294] 智人 b-1,4- 半乳糖基转移酶 III

[0295] 使用引物 RCD294 (SEQ ID NO :41) (5' -CTAGGCGCGCCCGAAGTCTCAGTGCCCTATTTGGC-3') 和 RCD295 (SEQ ID NO :42) (5' -CTTGTTAATTAATCAGTGTGAACCTCGGAGGGCTGT-3'), 从人肾 cDNA (marathon ready cDNA, Clontech) PCR 扩增截短型智人 b-1,4- 半乳糖基转移酶 III 基因 (hGalTIII, Genbank AF038661), 将该基因克隆到质粒 pCR2.1 并测序。使用导入的 AscI/PacI 位点将该截短的克隆 (除去了编码 N- 末端跨膜结构域的基因部分) 亚克

隆到载体 pXB53 以代替 hGalTI, 以生成质粒 pRCD381。现在, 该质粒含有由 PpGAPDH 启动子驱动的截短的 hGalTIII 与啤酒糖酵母 MNN2 基因的跨膜结构域 / 前导序列编码部分的基因融合体。

[0296] 实施例 5

[0297] 在产生复合 N- 聚糖的品系中表达 hGalTI 及 SpUGT

[0298] 将包含人 GnTII 基因和 SpGMS1+ 基因 (SpUGT) 的 pRCD257 质粒导入到品系 RDP27 中。RDP27 是巴斯德毕赤氏酵母的具有 och1 和 alg3 缺失的突变株, 且该突变株已转化有质粒 pSH49 和 pPB104 (所述质粒分别包含小鼠甘露糖苷酶 IB 和人 GnTI 活性融合构建体), 以及质粒 pPB103 (所述质粒包含乳克鲁维氏酵母 UDP-GlcNAc 转运蛋白) 和 pBK64 (所述质粒含有报告蛋白 K3) (Choi 等 2003)。在诺尔丝菌素上选择之后, 16 个转化体被选择用于测定所表达的报告蛋白 K3 的糖基化。在两个该转化体中, 观察到预期的复合人糖基化结构 $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$, 且将这些品系命名为 RDP30-10 (图 8A) 和 RDP30-13。将部分 hGalTI 基因 / 前导序列融合体质粒文库转化到品系 RDP30-10 中, 并在含有潮霉素的基本培养基上选择转化体。用 MALDI-TOFMS 分析由所得品系分泌的 K3 所释放的 N- 聚糖。在来自两个不同前导序列构建体 pXB53 和 pXB65 的转化体的 N- 聚糖上观察到与添加一个半乳糖相一致质量的分子移动。第一个 pXB53 由与 ScMnn2(s) 前导序列融合的 hGalTI 组成 (此处称为 ScMnn2(s)/hGalTI), 而另一个则是与 ScMnn1(m) 前导序列的融合体。通过 MALDI-TOF 分析来自 RDP37 (用 pXB53 转化的 RDP30-10) 的 K3 所释放的 N- 聚糖, 揭示约 10-20% $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ 被转化为 $\text{GalGlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ 而较少量 (1-2%) 转化为 $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ (pXB53, 图 8B)。对于第二个融合体 (pXB65), 观察到较小量 (3-5%) 转化为 $\text{GalGlcNAc}_2\text{Man}_3\text{GlcNAc}_2$, 但没有观察到可观测的 $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ 。

[0299] 实施例 6

[0300] 在产生杂合 N- 聚糖的品系中表达 hGalTI 和 ScGAL10

[0301] 用 NotI/PacI 将编码 UDP- 半乳糖 4- 差向异构酶的 ScGAL10 基因亚克隆到 NAT^R 载体 pTA18 和 pRCD351 中以替代 hGnTII, 这将差向异构酶基因分别插入在强 PMAI 启动子和较弱 PpHIS3 启动子前面, 以分别生成质粒 pRCD331 (P_{PMAI} -ScGAL10) 和 pRCD352 (P_{HIS3} -ScGAL10)。将质粒线性化 (pRCD331 在 PpPMA1 启动子中带有 SacI, 而 pRCD352 在 PpHIS3 启动子中带有 BglIII), 然后转化至品系 PBP-3 (美国专利申请号 No. 20040018590)。品系 PBP-3 是巴斯德毕赤氏酵母突变株, 其缺失 och1 并已被质粒 pSH49 和 pPB104 (分别含有小鼠甘露糖苷酶 IB 和人 GnTI 活性融合构建体), 以及质粒 pPB103 (包含乳克鲁维氏酵母 UDP-GlcNAc 转运蛋白) 和质粒 pBK64 (包含报告蛋白 K3) 转化 (Choi 等)。该品系产生位于分泌的蛋白上的结构为 $\text{GlcNAcMan}_5\text{GlcNAc}_2$ 的杂合 N- 聚糖。将在含有诺尔丝菌素的 YPD 培养基上选择得到的转化体, 通过利用引物 RCD285 (SEQ ID NO :43) (5' -TACGAGATTCCCAAATATGATTCC-3') 和 RCD286 (SEQ ID NO :44) (5' -ATAGTGTCTCCATATGGCTTGTTTC-3') 的 PCR、表达报告蛋白 K3 和分析释放的 N- 聚糖进行分析, 以确定所述品系维持了杂合 $\text{GlcNAcMan}_5\text{GlcNAc}_2$ 聚糖结构。将用 pRCD352 (P_{HIS3} -ScGAL10) 构建体转化的一个品系命名为 RDP38-18。在用 SalI (位于 PpURA3 内) 线性化后, 用质粒 pXB53 (包含 Mnn2(s)/hGalTI 融合构建体和 HYG^R 和 PpUR43 基因) 转化该品系。在含有潮霉素的 YPD 培养基上选择转化体并通过表达 K3 和测定 N- 聚糖的大小进行筛选。与来自 RDP38-18 的那些聚糖相比 (其大多是 $\text{GlcNAcMan}_5\text{GlcNAc}_2$), 从

RDP39-6 品系纯化的 K3 所释放的 N- 聚糖的大部分 ($\sim 2/3$) (图 10A) 包含一个额外的己糖 (HexGlcNAcMan₅GlcNAc₂)。此外, 额外的己糖残基可以通过随后与可溶性 b-1, 4- 半乳糖苷酶 (但不是 a-1, 3- 半乳糖苷酶或 a-1, 2- 甘露糖苷酶) 温育而被除去, 这表明在该品系中通过 hGalTI 催化而在 GlcNAc 末端以特异键 (b-1, 4) 添加单个半乳糖。

[0302] 实施例 7

[0303] 在产生复合 N- 聚糖的品系中表达 hGalTI 和 ScGAL10

[0304] 构建巴斯德毕赤氏酵母品系 YSH-44, 该品系展示具有 GlcNAc₂Man₃GlcNAc₂ 结构的复合 N- 聚糖。YSH-44 是巴斯德毕赤氏酵母突变株, 其缺失 och1 并用质粒 pSH49、pPB104、pKD53 和 pTC53 (分别含有小鼠甘露糖苷酶 IB、人 GnTI、黑腹果蝇甘露糖苷酶 II 和人 GnTII 的活性融合构建体) 以及质粒 pPB103 (含有乳克鲁维氏酵母 UDP-GlcNAc 转运蛋白) 和质粒 pBK64 (含有报告蛋白 K3) 转化 (Hamilton 等, Science. 2003 Aug 29 ;301(5637) :1244-6)。该品系用含有 Mnn2(s)/hGalTI 融合构建体的 pXB53 质粒进行转化, 并在含有潮霉素的 YPD 培养基上选择转化体。采用 MALDI-TOFMS 通过纯化 K3 和分析释放的 N- 聚糖来分析几个转化体。经分析的各转化体产生的大部分 N- 聚糖具有 GlcNAc₂Man₃GlcNAc₂ 结构, 且少数 ($\sim 5\%$) 与单个己糖添加 (YSH-71) 一致。然而, 虽然该峰值始终与 hGalTI 的导入相关联, 但其完全与 b-1, 4- 半乳糖苷酶不符。随后, 若干该品系用 BglIII 线性化后的质粒 pRCD395 和 pRCD396 (分别包含 P_{och1}-ScGAL10 和 P_{SEC4}-ScGAL10 的 PpHIS3/G418^R 质粒) 进行转化, 在 (G418 上选择, 并将得到的品系分别命名为 YSH-83 和 YSH-84。从分泌的 K3 释放的 N- 聚糖用 MALDI-TOFMS 进行分析。得到的转化体在含有 G418 的 YPD 培养基上选择, 并使用 MALDI-TOFMS 分析来自这些品系的纯化的分泌的 K3 所释放的 N- 聚糖。来自这些转化体的大部分 N- 聚糖具有三种结构, Gal₂GlcNAc₂Man₃GlcNAc₂ ($\sim 0-25\%$) 或 GalGlcNAc₂Man₃GlcNAc₂ ($\sim 40-50\%$), 剩余 N- 聚糖保持亲本 YSH-44 品系所展示的 GlcNAc₂Man₃GlcNAc₂ 结构。剩余的 N- 聚糖的相对量没有变化, 其与 ScGAL10 差向异构酶基因是由 PpOCH1 启动子 (YSH-83) 驱动还是由 PpSEC4 启动子 (YSH-84) 驱动并不相关。图 9B 显示了从 YSH-84 释放的 N- 聚糖的 MALDI-TOF MS。

[0305] 实施例 8

[0306] 构建差向异构酶 / 转移酶融合构建体

[0307] 使用引物 RCD326 (SEQ ID NO :45) (5' -CTTGGCCGCGCCATGACTGGTGTTCATGAAGGGAC T-3') 和 RCD329 (SEQ ID NO :46) 5' -CCTGGATCCCTTATATGTCTTGGTATGGGTCAG-3') 扩增 SpGALE 基因, 将其克隆到 pCR2.1 载体 (Invitrogen) 并测序。使用引物 RCD328 (SEQ ID NO :47) (5' -CTTGGATCCGGTGGTGGCCGCGACCTGAGC-CGCCTGCCC-3') 和 RCD199 (SEQ ID NO :48) (5' -CTICTTAATTAAGCTCGGTGTCCGATGTCCAC-3') 扩增消除了前 43 个氨基酸的 hGalTI 基因截短部分 (hGalTI Δ 43), 将其克隆到 pCR2.1 载体 (Invitrogen) 并测序。然后用 AscI/BamHI 消化 SpGALE 克隆, 用 BamHI/PacI 消化 hGalTI 克隆, 并将两者插入到用 AscI/PacI 消化的 pRCD452 中。质粒 pRCD452 包含 G418 抗性标记和含有 ScMNN2(s)/hGalTI 融合体的 GAPDH/CYC1 盒。AscI/BamIISpGALE 和 BamHI/PacI hGalTIA43 片段被连接到置换 AscI/PacI 释放的 hGalTI 以生成 pRCD461。该新质粒 pRCD461 包含 ScMNN2(s)/SpGALE/hGalTI 融合体, 其中 SpGalE 和 hGalTI 蛋白编码在单一多肽中, 其中所述多肽由含有 BamHI 位点的四氨基酸 (GSGG) 接头分离并由 PpGAPDH 启动子驱动。

[0308] 实施例 9

[0309] 在产生复合 N-聚糖的品系中表达半乳糖基转移酶、差向异构酶和转运蛋白

[0310] 用与 HYG^R 标记邻近的 XhoI 线性化含有活性 hGalTI-53 基因融合体的质粒 pXB53 以及含有 hGalTII-53 融合体的 pRCD378, 并用 T4 DNA 聚合酶 (New England Biolabs, Beverly, MA) 将其平端化。使用与 URA3 基因邻近的 HindIII 线性化含有 hGalTIII-53 基因融合体的质粒 pRCD381, 并用 T4 聚合酶使其平端化。然后使用 XhoI/SphI 分别从质粒 pRCD404、pRCD406 和 pRCD427 消化三个差向异构酶基因 ScGAL10、SpGALE 和 hGALE, 用 T4DNA 聚合酶使其平端化, 并插入到三个线性化的转移酶质粒中。这产生了九个新的二元转移酶 / 差向异构酶 HYG^R 质粒: 含有 hGalTI-53 和 ScGAL10 的 pRCD424、含有 hGalTI-53 和 SpGALE 的 pRCD425、含有 hGalTI-53 和 hGALE 的 pRCD438、含有 hGalTII-53 和 ScGAL10 的 pRCD439、含有 hGalTII-53 和 SpGALE 的 pRCD440、含有 hGalTII-53 和 hGALE 的 pRCD441、含有 hGalTIII-53 和 ScGAL10 的 pRCD442、含有 hGalTIII-53 和 SpGALE 的 pRCD443 以及含有 hGalTIII-53 和 hGALE 的 pRCD447。随后, 用这些二元 HYG^R 质粒 (用 XbaI 线性化的) 和分别包含 SpUGT、hUGT2、DmUGT 和 hUGT1 UDP-Gal 转运蛋白编码基因的 G418^R 质粒 pRCD393、pSH262、pSH263 和 pSH264 (用 AgeI 线性化的) 顺次转化品系 YSH44。由此, 生成了一系列品系, 其中各品系包含转移酶、差向异构酶和转运蛋白的不同组合。首先, 在包含 hGalTI-53 和 SpGALE 的品系中比较不同的 UDP-Gal 转运蛋白。导入 DmUGT 基因基本产生所有具有两个末端半乳糖残基的复合聚糖 (Gal₂GlcNAc₂Man₃GlcNAc₂), 而其它三个转运蛋白基因产生的复合聚糖特征实际上与仅使用转移酶和差向异构酶所获得的特征相同 (图 11A-11E)。其次, 通过将 pSH263 导入到含 pRCD424、pRCD425 或 pRCD438 的品系而在含有 hGalTI-53 融合体和活性 DmUGT 基因的品系中比较差向异构酶基因。Gal 基因与三个差向异构酶基因每一个的组合, 在分泌的 K3 上产生 Gal₂GlcNAc₂Man₃GlcNAc₂ 复合 N-聚糖方面是等同的。最后, 通过将 pRCD425、pRCD440 和 pRCD443 导入到转化有 pSH263 的品系而在含有 DmUGT 和 SpGALE 的品系中比较三个人转移酶融合构建体 hGalTI-53、hGalTII-53 和 hGalTIII-53。此处, 因为在含有 hGalTI-53 的品系中约 10% 复合 N-聚糖仅具有单个半乳糖 (GalGlcNAc₂Man₃GlcNAc₂), 而含有 hGalTI-53 的品系中全部可以观察到的复合 N-聚糖均为双-半乳糖基化的 Gal₂GlcNAc₂Man₃GlcNAc₂, 故此 hGalTII-53 转移 Gal 的效率轻微较低 (图 12A-12B)。此外, 由于 60-70% 的复合 N-聚糖包含 0-1 个半乳糖残基 (GlcNAc₂Man₃GlcNAc₂ 或 GalGlcNAc₂Man₃GlcNAc₂), 而仅 30-40% 是 Gal₂GlcNAc₂Man₃GlcNAc₂, 故此 hGalTIII-53 的效率显著低于 hGalTI-53 或 hGalTII-53 (图 12A-12C)。

[0311] 实施例 10

[0312] 使用单个质粒构建体表达半乳糖基转移酶、差向异构酶和转运蛋白

[0313] 通过用 SacI 消化将包含 P_{OCH1}-DmUGT 的 (G418^R 质粒 pSH263 线性化, 然后用 T4 DNA 聚合酶 (New England Biolabs) 平端化。用 XhoI/SphI 消化质粒 pRCD405 得到 P_{SEC4}-SpGALE 基因, 并用 T4DNA 聚合酶平端化。然后将平端化的 SpGALE 插入到 pSH263 的平端化 SacI 位点以生成质粒 pRCD446, 该质粒是二元转运蛋白 / 差向异构酶 (G418^R 质粒)。然后用 EcoRI 将 pRCD446 线性化并用 T4 DNA 聚合酶平端化。使用 BglIII/BamHI 从 pXB53 释放 P_{GAPDH}-ScMNN2(s)/hGalTI 融合构建体并用 T4 DNA 聚合酶使其平端化。然后将平端化的 ScMNN2(s)/hGalTI 插入到 pRCD446 的平端化 EcoRI 位点, 以生成质粒 pRCD465, 该

质粒是包含 ScMNN2(s)/hGalTI、SpGALE 和 DmUGT 的三元 G418^R 质粒。用 pRCD465 转化的巴斯德毕赤氏酵母 YSH-44 被命名为 RDP80。N- 聚糖特征显示位于 1663m/z 的单峰值,其对应于 Gal₂GlcNAc₂Man₃GlcNAc₂[C] 的质量 (图 14A)。

[0314] 用 Af1II 线性化包含 hGalTI-53 和 SpGALE 的 HYG^R 质粒 pRCD425 并用 T4 DNA 聚合酶将其平端化。使用 NotI/PacI 从 pSH263 释放 DmUGT 基因,然后将其插入到用 NotI/PacI 消化的质粒 pRCD405 以生成质粒 pRCD468,该质粒包含可作为单个盒释放的 P_{SEC4}-DmUGT-CYC1-TT 融合体。用 XhoI/SalI 消化 pRCD468 以释放 DmUGT 盒,然后用 T4 DNA 聚合酶平端化。将平端化的 DmUGT 插入到 pRCD425 的平端化的 Af1II 位点以生成质粒 pRCD466,该质粒是 hGalTI-53、SpGALE 和 DmUGT 的 HYG^R 三元质粒。

[0315] 用 Af1II 线性化包含 hGalTI-53 和 hGALE 的 HYG^R 质粒 pRCD438 并用 T4 DNA 聚合酶将其平端化。用 XhoI/SalI 消化 pRCD468 以释放 DmUGT 盒,并用 T4 DNA 聚合酶将其平端化。将平端化的 DmUGT 插入到 pRCD438 的平端化的 Af1II 位点以生成质粒 pRCD467,该质粒是 hGalTI-53、hGALE 和 DmUGT 的 HYG^R 三元质粒。

[0316] 体外 b- 半乳糖苷酶消化

[0317] 将来自巴斯德毕赤氏酵母品系 RDP80 的 N- 聚糖 (2 μg) 与 3mU β 1,4 半乳糖苷酶 (QAbio, San Mateo, CA) 在 50mM NH₄HCO₃, pH6.0 于 37℃ 下温育 16-20 小时。图 14B 中的 N- 聚糖分析表明主要峰值位于 1430m/z [A],该峰值对应于 N- 聚糖 GlcNAc₂Man₃GlcNAc₂ 的质量,从而证实图 14A 中的半乳糖转移。

[0318] 体外唾液酸转移酶反应

[0319] 将从品系 RDP80 纯化得到的 K3 (200g) 与 50mg CMP- 唾液酸和 15mU 大鼠重组体 α (2,6)-(N)- 唾液酸转移酶 (Calbiochem) 在 50mM NH₄HCO₃, pH6.0 于 37℃ 下温育 16-20 小时。然后通过 PNG 酶 (PNGase)F 消化释放 N- 聚糖,并在 MALDI-TOFMS 上检测。当与来自 RDP80 的那些聚糖 (图 14A) 比较时,聚糖谱显示在唾液酸转移酶处理后质量增加 (图 14C)。如图 14C 所示的谱显示出主要峰值位于 2227m/z [X],该峰值对应于 N- 聚糖 NANA₂Gal₂GlcNAc₂Man₃GlcNAc₂ 的质量,从而进一步证实品系 RDP80 产生的 N- 聚糖是人型 Gal₂GlcNAc₂Man₃GlcNAc₂。

[0320] 实施例 11

[0321] 差向异构酶序列比对

[0322] 使用 CLUSTAL 实施差向异构酶的序列比对。使用序列表核苷酸序列和 / 或氨基酸序列在 GenBank、SwissProt、BLOCKS 和 Pima II 数据库中查询序列。搜索这些数据库 (其中包含先前鉴定的和注释的序列),使用 BLAST (BasicLocal Alignment Search Tool) 寻找同源性的区域。(参见例如, Altschul, S. F. (1993) J. Mol. Evol 36 :290-300 ;和 Altschul 等 (1990) J. Mol. Biol. 215 :403-410)。BLAST 生成核苷酸序列和氨基酸序列比对以测定序列相似性。当处理一级序列模式和二级结构缺口罚分时,已可以使用其它算法。(参见例如, Smith, T. 等 (1992) Protein Engineering 5 :35-51)。

[0323] 实施例 12

[0324] 材料

[0325] MOPS、二甲胍酸钠、二氯化锰、UDP- 半乳糖和 CMP-N- 乙酰神经氨酸购自 Sigma。TFA 购自 Aldrich。来自牛乳的 β 1,4- 半乳糖基转移酶购自 Calbiochem。蛋白 N- 糖苷酶 F、甘

露糖苷酶和寡糖购自 Glyko (San Rafael, CA)。DEAE ToyoPearl 树脂购自 TosoHaas。金属螯合的“HisBind”树脂购自 Novagen (Madison, WI)。96-孔裂解物清除 (lysate-clearing) 平板购自 Promega (Madison, WI)。蛋白结合 96-孔平板购自 Millipore (Bedford, MA)。盐和缓冲试剂购自 Sigma (St. Louis, MO)。MALDI 基质购自 Aldrich (Milwaukee, WI)。

[0326] 摇瓶培养

[0327] 从包含目标品系的 YPD 平板 (<2 周) 挑取单菌落, 并接种到 50ml “Falcon” 离心管中的 10ml BMGY 培养基中。使培养物在 24℃ 下生长至饱和 (约 48 小时)。将种子培养物转移到包含 150ml BMGY 培养基的 500ml 带挡板的容量瓶中, 并于 24℃ 生长至 OD₆₀₀ 为 5±2 (约 18 小时)。细胞的生长速率作为 OD₆₀₀ 的自然对数对时间的图的斜率来确定。通过于 3000g 离心 10 分钟从生长培养基 (BMGY) 收获细胞, 用 BMMY 洗涤, 然后悬浮于 250ml 带挡板的容量瓶中的 15ml BMMY 中。24 小时后, 通过离心收获表达培养基瓶 (3000g 10 分钟), 并对上清液分析 K3 的产生。

[0328] 生物反应器培养

[0329] 在含有 150ml BMGY 培养基的 500ml 带挡板的容量瓶中接种 1ml 种子培养物 (参见瓶培养)。使接种物于 24℃ 生长至 OD₆₀₀ 为 4-6 (大约 18 小时)。然后离心来自接种物培养物的细胞并将其重悬于 50ml 发酵培养基中 (每升培养基: CaSO₄·2H₂O 0.30g, K₂SO₄ 6.00g, MgSO₄·7H₂O 5.00g, 甘油 40.0g, PTM 盐 2.0ml, 生物素 4×10⁻³g, H₃PO₄ (85%) 30ml, PTM1 盐每升: CuSO₄·H₂O 6.00, NaIO₃ 0.08g, MnSO₄·7H₂O 3.00g, NaMoO₄·2H₂O 0.20g, H₃BO₃ 0.02g, CoCl₂·6H₂O 0.50g, ZnCl₂ 20.0g, FeSO₄·7H₂O 65.0g, 生物素 0.20g, H₂SO₄ (98%) 5.00ml)。

[0330] 在 3 升中凹底 (1.5 升初始进料体积) Applikon 生物反应器中实施发酵。发酵罐在温度 24℃ 补料分批模式 (fed-batch mode) 下进行, 且采用 30% 氢氧化铵将 pH 控制在 4.5±0.1。通过调整搅拌速度 (450-900rpm) 和纯氧供应, 将溶解氧保持在相对于 1 大气压 (atm) 的空气中饱和度的 40% 以上。将空气流速保持在 1vvm。当批次相中初始甘油 (40g/1) 被耗尽时 (这可由 DO 的增加得以指示), 以 12ml/1/h 的给料速度添加包含 12ml/1 PTM 盐的 50% 甘油溶液, 直至达到所需生物量 (biomass) 浓度。在半小时饥饿相 (starvation phase) 之后, 开始甲醇给料 (含有 12ml/1 PTM 的 100% 甲醇)。用甲醇给料速度将发酵罐中的甲醇浓度控制在 0.2 和 0.5% 之间。采用位于发酵罐的出口气 (offgas) 的 TGS 气体传感器 (购自 Figaro Engineering Inc. 的 TGS822) 在线测量甲醇浓度。每隔八小时对发酵罐取样并分析生物量 (OD₆₀₀, 细胞湿重和细胞计数)、残余碳源水平 (采用 Aminex 87H 通过 HPLC 分析甘油和甲醇) 和细胞外蛋白含量 (通过 SDSPage, 和 Bio-Rad 蛋白质测定法)。

[0331] 报告蛋白表达、纯化和 N-连接的聚糖的释放

[0332] 如先前所报道的, 将 K3 结构域 (处于醇氧化酶 1 (AOX1) 启动子控制下) 被用作模型蛋白, 并使用 6x 组氨酸标记进行纯化 (Choi 等, Proc Natl Acad Sci USA. 2003 Apr 29; 100(9):5022-7)。通过改进先前报道的方法从糖蛋白释放和分离聚糖 (Papac 和 Briggs 1998)。在还原和羧甲基化蛋白, 并封闭膜之后, 用水洗涤孔三次。通过加入含有 1 毫单位 N-聚糖酶 (Glyko) 的 30ml 10mM NH₄HCO₃ pH 8.3, 使蛋白去糖基化。于 37℃ 16 小时后, 通过离心取出包含聚糖的溶液, 并蒸发至干燥。

[0333] 蛋白纯化

[0334] 在 Beckman BioMek 2000 样品处理机器人 (Beckman/Coulter Ranch Cucamonga,

CA) 上以 96-孔格式纯化 Kringle 3。使用 C-末端六组氨酸标记从表达培养基中纯化 Kringle 3。机器人纯化是对 Novagen 所提供的规程的 HisBind 树脂的改进。简言之,将 150 μ L (μ L) 固定体积的树脂注入 96-孔裂解物结合平板的孔中,用 3 体积的水洗涤,用 5 体积的 50mM NiSO_4 加载,然后用 3 体积的结合缓冲液 (5mM 咪唑, 0.5M NaCl, 20mM Tris-HCl pH7.9) 洗涤。蛋白表达培养基以培养基/PBS (60mM PO_4 , 16mM KCl, 822mM NaCl pH7.4) 3 : 2 稀释,并加载到柱上。引流后,用 10 体积的结合缓冲液和 6 体积的洗涤缓冲液 (30mM 咪唑, 0.5M NaCl, 20mM Tris-HCl pH7.9) 洗涤柱,并用 6 体积的洗脱缓冲液 (1M 咪唑, 0.5M NaCl, 20mM Tris-HCl pH7.9) 洗脱蛋白。通过冷冻干燥将洗脱的糖蛋白蒸发至干燥。

[0335] N-连接的聚糖的释放

[0336] 通过改进先前报道的方法从糖蛋白释放和分离聚糖 (Papac 等, A. J. S. (1998) *Glycobiology* 8, 445-454)。将 96-孔 MultiScreen IP (Immobilon-P 膜) 板 (Millipore) 的孔用 100 μ L 甲醇弄湿,用 3 $\times$ 150 μ L 水和 50 μ L RCM 缓冲液 (8M 脲, 360mM Tris, 3.2mM EDTA pH8.6) 洗涤,在各次添加之后采用温和真空进行引流。干燥的蛋白样品在 30 μ L RCM 缓冲液中溶解,并转移至含有 10 μ L RCM 缓冲液的孔中。对孔进行引流,并用 RCM 缓冲液洗涤两次。通过加入 60 μ L 溶于 RCM 缓冲液中的 0.1M DTT 于 37 $^\circ\text{C}$ 还原蛋白质 1 小时。用 300 μ L 水洗涤孔三次,然后通过加入 60 μ L 0.1M 碘乙酸于室温下于黑暗中羧甲基化 30 分钟。用水再洗涤孔三次,并通过加入 100 μ L 溶于水中的 1% PVP360 于室温封闭膜 1 小时。对孔进行引流,然后用 300 μ L 水洗涤三次,并通过加入包含一毫单位 N-聚糖酶 (Glyko) 的 30 μ L 10mM NH_4HCO_3 pH 8.3 使之去糖基化。于 37 $^\circ\text{C}$ 16 小时后,通过离心取出包含聚糖的溶液,并蒸发至干燥。

[0337] 其它:根据 Laemmli (Laemmli 1970) 通过 SDS/PAGE 分离蛋白。

[0338] 基质辅助激光解吸电离飞行时间质谱分析法

[0339] 采用 Voyager DE PRO 线性 MALDI-TOF (Applied Biosciences) 质谱仪利用延迟的提取 (Delayed Extraction) 来测定聚糖的分子量。来自各孔的干燥聚糖在 15 μ L 水中溶解,将 0.5 μ L 点样于不锈钢样品平板上,并与 0.5 μ L S-DHB 基质 (溶于 1 : 1 水 / 乙腈 0.1% TFA 中的 9mg/mL 二羟基苯甲酸, 1mg/mL 5-甲氧基水杨酸) 混合,并使之干燥。

[0340] 通过使用 4ns 脉冲时间的脉冲氮激光器 (337nm) 照射产生离子。采用延迟的提取模式 (125ns 延迟,且加速电压为 20kV) 操作所述仪器。栅压 (grid voltage) 为 93.00%,导线电压 (guide wire voltage) 为 0.10%,内压 (internal pressure) 小于 5×10^{-7} 托,低质量闸门 (low mass gate) 为 875Da。由总共 100-200 次激光脉冲生成谱,并用 2GHz 数字转换器获取所述谱。 $\text{Man}_5\text{GlcNAc}_2$ 寡糖用作外部分子量标准。采用所述仪器以正离子模式生成全部谱。谱的估计质量精确度为 0.5%。

[0341] 序列表正文

序列表

<110>GlycoFi, Inc.

Davidson, Robert

Gerngross, Tillman

Wildt, Stefan

Choi, Byung-Kwon

Nett, Juergen

Bobrowicz, Piotr

Hamilton, Stephen

<120> 在低等真核生物中产生半乳糖基化糖蛋白

<130>GFI-12

<160>57

<170>FastSEQ for Windows Version 4.0

<210>1

<211>27

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>1

gccgcgacct gagccgcctg ccccaac 27

<210>2

<211>27

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>2

ctagctcggt gtcccgatgt ccactgt 27

<210>3

<211>36

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>3

cttaggcgcg ccggccgcga cctgagccgc ctgccc 36

<210>4

<211>21

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>4

ggggcatatc tgccgcccat c 21

<210>5

<211>21

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>5

gatgggcggc agatatgccc c 21

<210>6

<211>36

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>6

cttcttaatt aactagctcg gtgtcccgat gtccac 36

<210>7

<211>39

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>7

ccttgcggcc gcatggctgt caagggcgac gatgtcaaa 39

<210>8

<211>40

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>8

attcgagaat agttaagtgt caaatcaat gcactatttt 40

<210>9

<211>40

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>9

aaaatagtg attgattttg acacttaact attctcgaat 40

<210>10

<211>39

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>10

ccttttaatt aattaatgct tatgatcaac gtccttagc 39

<210>11

<211>49

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>11

ggctcgagcg gccgccacca tgaatagcat acacatgaac gccaatacg 49

<210>12

<211>47

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>12

ccctcgagtt aattaactag acgcgcggca gcagcttctc ctcatcg 47

<210>13

<211>355

<212>PRT

<213> 粟酒裂殖糖酵母 (Schizosaccharomyces pombe)

<400>13

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

Asp Val Val Asn Leu Thr Ala Asn Pro Thr Arg Ala Asn Glu Glu Leu
 305 310 315 320
 Lys Trp Lys Thr Ser Arg Ser Ile Tyr Glu Ile Cys Val Asp Thr Trp
 325 330 335
 Arg Trp Gln Gln Lys Tyr Pro Tyr Gly Phe Asp Leu Thr His Thr Lys
 340 345 350
 Thr Tyr Lys
 355

<210>14

<211>1068

<212>PRT

<213> 粟酒裂殖糖酵母 (Schizosaccharomyces pombe)

<400>14

Ala Thr Gly Ala Cys Thr Gly Gly Thr Gly Thr Thr Cys Ala Thr Gly
 1 5 10 15
 Ala Ala Gly Gly Gly Ala Cys Thr Gly Thr Gly Thr Thr Gly Gly Thr
 20 25 30
 Thr Ala Cys Thr Gly Gly Cys Gly Gly Cys Gly Cys Thr Gly Gly Thr
 35 40 45
 Thr Ala Thr Ala Thr Ala Gly Gly Thr Thr Cys Thr Cys Ala Thr Ala
 50 55 60
 Cys Gly Thr Gly Cys Gly Thr Thr Gly Thr Thr Thr Thr Gly Thr Thr
 65 70 75 80
 Ala Gly Ala Ala Ala Ala Ala Gly Gly Ala Thr Ala Thr Gly Ala Thr
 85 90 95
 Gly Thr Thr Gly Thr Ala Ala Thr Thr Gly Thr Cys Gly Ala Thr Ala
 100 105 110
 Ala Thr Thr Thr Ala Thr Gly Cys Ala Ala Thr Thr Cys Thr Cys Gly
 115 120 125
 Cys Gly Thr Thr Gly Ala Ala Gly Cys Cys Gly Thr Gly Cys Ala Cys
 130 135 140
 Cys Gly Cys Ala Thr Thr Gly Ala Ala Ala Ala Ala Cys Thr Cys Ala
 145 150 155 160
 Cys Thr Gly Gly Gly Ala Ala Ala Ala Ala Ala Gly Thr Cys Ala Thr
 165 170 175
 Ala Thr Thr Cys Cys Ala Cys Cys Ala Gly Gly Thr Gly Gly Ala Thr
 180 185 190

Thr Thr Gly Cys Thr Thr Gly Ala Thr Gly Ala Gly Cys Cys Ala Gly		
195	200	205
Cys Thr Thr Thr Gly Gly Ala Cys Ala Ala Gly Gly Thr Cys Thr Thr		
210	215	220
Cys Gly Cys Ala Ala Ala Thr Cys Ala Ala Ala Ala Cys Ala Thr Ala		
225	230	235
240		
Thr Cys Thr Gly Cys Thr Gly Thr Cys Ala Thr Thr Cys Ala Thr Thr		
245	250	255
Thr Thr Gly Cys Thr Gly Gly Thr Cys Thr Cys Ala Ala Ala Gly Cys		
260	265	270
Ala Gly Thr Thr Gly Gly Thr Gly Ala Ala Thr Cys Thr Gly Thr Ala		
275	280	285
Cys Ala Gly Gly Thr Thr Cys Cys Thr Thr Thr Gly Ala Gly Thr Thr		
290	295	300
Ala Thr Thr Ala Cys Ala Ala Ala Ala Ala Thr Ala Ala Cys Ala Thr		
305	310	315
320		
Thr Thr Cys Cys Gly Gly Thr Ala Cys Cys Ala Thr Thr Ala Ala Thr		
325	330	335
Thr Thr Ala Ala Thr Ala Gly Ala Gly Thr Gly Cys Ala Thr Gly Ala		
340	345	350
Ala Gly Ala Ala Gly Thr Ala Thr Ala Ala Thr Gly Thr Ala Cys Gly		
355	360	365
Thr Gly Ala Cys Thr Thr Cys Gly Thr Cys Thr Thr Thr Thr Cys Thr		
370	375	380
Thr Cys Ala Thr Cys Thr Gly Cys Thr Ala Cys Cys Gly Thr Gly Thr		
385	390	395
400		
Ala Thr Gly Gly Cys Gly Ala Thr Cys Cys Thr Ala Cys Thr Ala Gly		
405	410	415
Ala Cys Cys Thr Gly Gly Thr Gly Gly Thr Ala Cys Cys Ala Thr Thr		
420	425	430
Cys Cys Thr Ala Thr Thr Cys Cys Ala Gly Ala Gly Thr Cys Ala Thr		
435	440	445
Gly Cys Cys Cys Thr Cys Gly Thr Gly Ala Ala Gly Gly Thr Ala Cys		
450	455	460
Ala Ala Gly Cys Cys Cys Ala Thr Ala Thr Gly Gly Thr Cys Gly Cys		
465	470	475
480		
Ala Cys Ala Ala Ala Gly Cys Thr Thr Thr Thr Cys Ala Thr Thr Gly		
485	490	495
Ala Ala Ala Ala Thr Ala Thr Cys Ala Thr Thr Gly Ala Gly Gly Ala		

500	505	510
Thr Gly Ala Gly Ala Cys Cys Ala Ala Gly Gly Thr Gly Ala Ala Cys		
515	520	525
Ala Ala Ala Thr Cys Gly Cys Thr Thr Ala Ala Thr Gly Cys Ala Gly		
530	535	540
Cys Thr Thr Thr Ala Thr Thr Ala Cys Gly Cys Thr Ala Thr Thr Thr		
545	550	555
Thr Ala Ala Thr Cys Cys Cys Gly Gly Ala Gly Gly Thr Gly Cys Thr		
565	570	575
Cys Ala Thr Cys Cys Cys Thr Cys Thr Gly Gly Thr Gly Ala Ala Cys		
580	585	590
Thr Cys Gly Gly Thr Gly Ala Ala Gly Ala Thr Cys Cys Thr Cys Thr		
595	600	605
Thr Gly Gly Cys Ala Thr Cys Cys Cys Thr Ala Ala Thr Ala Ala Cys		
610	615	620
Thr Thr Gly Cys Thr Thr Cys Cys Thr Thr Ala Thr Ala Thr Cys Gly		
625	630	635
Cys Gly Cys Ala Ala Gly Thr Thr Gly Cys Thr Gly Thr Ala Gly Gly		
645	650	655
Ala Ala Gly Ala Thr Thr Gly Gly Ala Thr Cys Ala Thr Thr Thr Gly		
660	665	670
Ala Ala Thr Gly Thr Ala Thr Thr Thr Gly Gly Cys Gly Ala Cys Gly		
675	680	685
Ala Thr Thr Ala Thr Cys Cys Cys Ala Cys Ala Thr Cys Thr Gly Ala		
690	695	700
Cys Gly Gly Thr Ala Cys Thr Cys Cys Ala Ala Thr Thr Cys Gly Thr		
705	710	715
Gly Ala Cys Thr Ala Cys Ala Thr Thr Cys Ala Cys Gly Thr Ala Thr		
725	730	735
Gly Cys Gly Ala Thr Thr Thr Gly Gly Cys Ala Gly Ala Gly Gly Cys		
740	745	750
Thr Cys Ala Thr Gly Thr Thr Gly Cys Thr Gly Cys Thr Cys Thr Cys		
755	760	765
Gly Ala Thr Thr Ala Cys Cys Thr Gly Cys Gly Cys Cys Ala Ala Cys		
770	775	780
Ala Thr Thr Thr Thr Gly Thr Thr Ala Gly Thr Thr Gly Cys Cys Gly		
785	790	795
Cys Cys Cys Thr Thr Gly Gly Ala Ala Thr Thr Thr Gly Gly Gly Ala		
805	810	815

Thr Cys Ala Gly Gly Ala Ala Cys Thr Gly Gly Thr Ala Gly Thr Ala		
820	825	830
Cys Thr Gly Thr Thr Thr Thr Thr Cys Ala Gly Gly Thr Gly Cys Thr		
835	840	845
Cys Ala Ala Thr Gly Cys Gly Thr Thr Thr Thr Cys Gly Ala Ala Ala		
850	855	860
Gly Cys Thr Gly Thr Thr Gly Gly Ala Ala Gly Ala Gly Ala Thr Cys		
865	870	875
Thr Thr Cys Cys Thr Thr Ala Thr Ala Ala Gly Gly Thr Cys Ala Cys		
885	890	895
Cys Cys Cys Thr Ala Gly Ala Ala Gly Ala Gly Cys Ala Gly Gly Gly		
900	905	910
Gly Ala Cys Gly Thr Thr Gly Thr Thr Ala Ala Cys Cys Thr Ala Ala		
915	920	925
Cys Cys Gly Cys Cys Ala Ala Cys Cys Cys Cys Ala Cys Thr Cys Gly		
930	935	940
Cys Gly Cys Thr Ala Ala Cys Gly Ala Gly Gly Ala Gly Thr Thr Ala		
945	950	955
Ala Ala Ala Thr Gly Gly Ala Ala Ala Ala Cys Cys Ala Gly Thr Cys		
965	970	975
Gly Thr Ala Gly Cys Ala Thr Thr Thr Ala Thr Gly Ala Ala Ala Thr		
980	985	990
Thr Thr Gly Cys Gly Thr Thr Gly Ala Cys Ala Cys Thr Thr Gly Gly		
995	1000	1005
Ala Gly Ala Thr Gly Gly Cys Ala Ala Cys Ala Gly Ala Ala Gly Thr		
1010	1015	1020
Ala Thr Cys Cys Cys Thr Ala Thr Gly Gly Cys Thr Thr Thr Gly Ala		
1025	1030	1035
Cys Cys Thr Gly Ala Cys Cys Cys Ala Thr Ala Cys Cys Ala Ala Gly		
1045	1050	1055
Ala Cys Ala Thr Ala Thr Ala Ala Gly Thr Ala Ala		
1060	1065	

<210>15

<211>32

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>15

tatgcggccg cggctgatga tatttgctac ga 32

<210>16

<211>33

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>16

cctctcgagt ggacacagga gac tcagaaa cag 33

<210>17

<211>30

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>17

cttctcgagg aagtaaagtt ggcgaaactt 30

<210>18

<211>33

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>18

cttagcggcc gcgattgttc gtttgagtag ttt 33

<210>19

<211>30

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>19

cttctcgagg gcattcaaag aagccttggg 30

<210>20

<211>33

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>20

cttagcggcc gctgagtggc catgtgggaa ctt 33

<210>21

<211>30

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>21

cctggatcca acagactaca atgacaggag 30

<210>22

<211>43

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>22

cctgcatgcc tcgagcttgc cggcgtctaa atagccgttg aag 43

<210>23

<211>41

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>23

cctgtcgacg ctgccggcaa gctcgagttt aagcgggtgct g 41

<210>24

<211>34

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>24

cctggatcct ttggcaaaaa ccagccctgg tgag 34

<210>25

<211>33

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>25

tccttaatta aagaaagcta gagtaaaata gat 33

<210>26

<211>33

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>26

tcctcgagg atcatgttga tcaactgaga ccg 33

<210>27

<211>45

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>27

ggctcgagcg gccgccacca tggcagcggg tggggctggg gggtc 45

<210>28

<211>43

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>28

ccctcgaggtt aattaatcac ttcaccagca ctgactttgg cag 43

<210>29

<211>45

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>29

ggctcgagcg gccgccacca tggcagcggt tggggctggt ggttc 45

<210>30

<211>44

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>30

ccctcgagtt aattaactag gaacccttca ccttggtgag caac 44

<210>31

<211>36

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>31

tagcggccgc atgacagctc agttacaaag tgaaag 36

<210>32

<211>36

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>32

cgttaattaa tcaggaaaat ctgtagacaa tcttgg 36

<210>33

<211>29

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>33

gcggccgcat ggcagagaag gtgctggta 29

<210>34

<211>29

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>34

ttaattaatc aggcttgctg gccaaagcc 29

<210>35

<211>21

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>35

atgactggtg ttcatgaagg g 21

<210>36

<211>21

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>36

ttacttatat gtcttggtat g 21

<210>37

<211>94

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>37

gcggccgcat gactggtgtt catgaaggga ctgtgttggt tactggcggc gctggttata 60
taggttctca tacgtgcgtt gttttgttag aaaa 94

<210>38

<211>29

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>38

ttaattaatt acttatatgt cttggtatg 29

<210>39

<211>33

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>39

cttaggcgcg cccagcacct ggccttcttc agc 33

<210>40

<211>36

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>40

cttgттаatt aatcagcccc gagggggcca cgacgg 36

<210>41

<211>35

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>41

cttaggcgcg cccgaagtct cagtgcccta tttgg 35

<210>42

<211>36

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>42

cttgттаatt aatcagtggtg aacctcggag ggctgt 36

<210>43

<211>24

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>43

tacgagattc ccaaatatga ttcc 24

<210>44

<211>24

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>44

atagtgtctc catatggctt gttc 24

<210>45

<211>35

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>45

cttggcgcgc catgactggt gttcatgaag ggact 35

<210>46

<211>33

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>46

cctggatccc ttatatgtct tggtatgggt cag 33

<210>47

<211>39

<212>DNA

<213> 人工序列

<220>

<223> 简并寡核苷酸引物

<400>47

cttggatccg gtggtggccg cgacctgagc cgcctgccc 39

<210>48

<211>1303

<212>DNA

<213> 粟酒裂殖糖酵母 (Schizosaccharomyces pombe)

<400>48

```

atggtatgaa taactttttt aattaatcaa aagcattctt ttgtatgaaa atactaatta 60
tgattcatag actggtgttc atgaaggac tgtgttggtt actggcggcg ctggttatat 120
aggttctcat acggtacgta gagagcttga agatacagaa gaggattagt aatgtacatg 180
taaagtgttt aagcacgcat cttttgtgaa tatagcttgc tgctctttac tttatacaa 240
tttcgtccat attctataaa gctctttttt gagatatatt gctaaccaca atctgcaata 300
gtgcgttggt ttgttagaaa aaggatatga tgttgtaatt gtcgataatt tatgcaattc 360
tcgcgttgaa gccgtgcacc gcattgaaaa actcactggg aaaaaagtca tattccacca 420
ggtggatttg cttgatgagc cagctttgga caaggtcttc gcaaatcaaa acatatctgc 480
tgtcattcat ttgtctggtc tcaaagcagt tggatgaatc gtacaggttc ctttgagtta 540
ttacaaaaat aacatttccg gtaccattaa tttaatagag tgcatgaaga agtataatgt 600
acgtgacttc gtcttttctt catctgtctac cgtgtatggc gatcctacta gacctggtgg 660
taccattcct attccagagt catgccctcg tgaaggtaga agcccatatg gtcgcacaaa 720
gcttttcatt gaaaatatca ttgaggatga gaccaagggtg aacaaatcgc ttaatgcagc 780
tttattacgc tattttaate ccggagggtgc tcatccctct ggtgaactcg gtgaagatcc 840
tcttggcate cctaataact tgcttccctt tatcgcgcaa gttgctgtag gaagattgga 900
tcatttgaat gtatttggcg acgattatcc cacatctgac ggtactccaa ttcgtgacta 960
cattcacgta tgcgatttgg cagaggtcca tgttgctgct ctcgattacc tgcgccaaca 1020
ttttgttagt tgccgccctt ggaatttggg atcaggaact ggtagtactg tttttcaggt 1080
gctcaatgcg ttttcgaaag ctgttgggaag agatcttcct tataaggcca cccctagaag 1140
agcaggggac gttgttaacc taaccgcca cccactcgc gctaacgagg agttaaaatg 1200
gaaaaccagt cgtagcattt atgaaatttg cgttgacact tggagatggc aacagaagta 1260
tcctatggc tttgacctga ccataccaa gacatataag taa 1303

```

<210>49

<211>355

<212>PRT

<213> 智人 (Homo Sapiens)

<400>49

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

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

<210>50

<211>337

<212>PRT

<213> 智人 (Homo Sapiens)

<400>50

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

Phe	Ile	Phe	Ser	Asp	Val	Asp	Leu	Val	Pro	Met	Asp	Asp	Arg	Asn	Leu	
180					185					190						
Tyr	Arg	Cys	Gly	Asp	Gln	Pro	Arg	His	Phe	Ala	Ile	Ala	Met	Asp	Lys	
195					200					205						
Phe	Gly	Phe	Arg	Leu	Pro	Tyr	Ala	Gly	Tyr	Phe	Gly	Gly	Val	Ser	Gly	
210					215					220						
Leu	Ser	Lys	Ala	Gln	Phe	Leu	Arg	Ile	Asn	Gly	Phe	Pro	Asn	Glu	Tyr	
225					230					235					240	
Trp	Gly	Trp	Gly	Gly	Glu	Asp	Asp	Asp	Ile	Phe	Asn	Arg	Ile	Ser	Leu	
245					250					255						
Thr	Gly	Met	Lys	Ile	Ser	Arg	Pro	Asp	Ile	Arg	Ile	Gly	Arg	Tyr	Arg	
260					265					270						
Met	Ile	Lys	His	Asp	Arg	Asp	Lys	His	Asn	Glu	Pro	Asn	Pro	Gln	Arg	
275					280					285						
Phe	Thr	Lys	Ile	Gln	Asn	Thr	Lys	Leu	Thr	Met	Lys	Arg	Asp	Gly	Ile	
290					295					300						
Gly	Ser	Val	Arg	Tyr	Gln	Val	Leu	Glu	Val	Ser	Arg	Gln	Pro	Leu	Phe	
305					310					315					320	
Thr	Asn	Ile	Thr	Val	Asp	Ile	Gly	Arg	Pro	Pro	Ser	Trp	Pro	Pro	Arg	
325					330					335						
Gly																

<210>51

<211>362

<212>PRT

<213> 智人 (Homo Sapiens)

<400>51

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

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

<210>52

<211>348

<212>PRT

<213> 智人 (Homo Sapiens)

<400>52

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

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

<210>53

<211>699

<212>PRT

<213>啤酒糖酵母 (*Saccharomyces cerevisiae*)

<400>53

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

Phe Asn Pro Ile Gly Ala His Pro Ser Gly Leu Ile Gly Glu Asp Pro		
195	200	205
Leu Gly Ile Pro Asn Asn Leu Leu Pro Tyr Met Ala Gln Val Ala Val		
210	215	220
Gly Arg Arg Glu Lys Leu Tyr Ile Phe Gly Asp Asp Tyr Asp Ser Arg		
225	230	235 240
Asp Gly Thr Pro Ile Arg Asp Tyr Ile His Val Val Asp Leu Ala Lys		
	245	250 255
Gly His Ile Ala Ala Leu Gln Tyr Leu Glu Ala Tyr Asn Glu Asn Glu		
	260	265 270
Gly Leu Cys Arg Glu Trp Asn Leu Gly Ser Gly Lys Gly Ser Thr Val		
	275	280 285
Phe Glu Val Tyr His Ala Phe Cys Lys Ala Ser Gly Ile Asp Leu Pro		
	290	295 300
Tyr Lys Val Thr Gly Arg Arg Ala Gly Asp Val Leu Asn Leu Thr Ala		
305	310	315 320
Lys Pro Asp Arg Ala Lys Arg Glu Leu Lys Trp Gln Thr Glu Leu Gln		
	325	330 335
Val Glu Asp Ser Cys Lys Asp Leu Trp Lys Trp Thr Thr Glu Asn Pro		
	340	345 350
Phe Gly Tyr Gln Leu Arg Gly Val Glu Ala Arg Phe Ser Ala Glu Asp		
	355	360 365
Met Arg Tyr Asp Ala Arg Phe Val Thr Ile Gly Ala Gly Thr Arg Phe		
	370	375 380
Gln Ala Thr Phe Ala Asn Leu Gly Ala Ser Ile Val Asp Leu Lys Val		
385	390	395 400
Asn Gly Gln Ser Val Val Leu Gly Tyr Glu Asn Glu Glu Gly Tyr Leu		
	405	410 415
Asn Pro Asp Ser Ala Tyr Ile Gly Ala Thr Ile Gly Arg Tyr Ala Asn		
	420	425 430
Arg Ile Ser Lys Gly Lys Phe Ser Leu Cys Asn Lys Asp Tyr Gln Leu		
	435	440 445
Thr Val Asn Asn Gly Val Asn Ala Asn His Ser Ser Ile Gly Ser Phe		
	450	455 460
His Arg Lys Arg Phe Leu Gly Pro Ile Ile Gln Asn Pro Ser Lys Asp		
465	470	475 480
Val Phe Thr Ala Glu Tyr Met Leu Ile Asp Asn Glu Lys Asp Thr Glu		
	485	490 495
Phe Pro Gly Asp Leu Leu Val Thr Ile Gln Tyr Thr Val Asn Val Ala		

500							505							510						
Gln	Lys	Ser	Leu	Glu	Met	Val	Tyr	Lys	Gly	Lys	Leu	Thr	Ala	Gly	Glu					
515							520							525						
Ala	Thr	Pro	Ile	Asn	Leu	Thr	Asn	His	Ser	Tyr	Phe	Asn	Leu	Asn	Lys					
530							535							540						
Pro	Tyr	Gly	Asp	Thr	Ile	Glu	Gly	Thr	Glu	Ile	Met	Val	Arg	Ser	Lys					
545							550							555						
Lys	Ser	Val	Asp	Val	Asp	Lys	Asn	Met	Ile	Pro	Thr	Gly	Asn	Ile	Val					
565							570							575						
Asp	Arg	Glu	Ile	Ala	Thr	Phe	Asn	Ser	Thr	Lys	Pro	Thr	Val	Leu	Gly					
580							585							590						
Pro	Lys	Asn	Pro	Gln	Phe	Asp	Cys	Cys	Phe	Val	Val	Asp	Glu	Asn	Ala					
595							600							605						
Lys	Pro	Ser	Gln	Ile	Asn	Thr	Leu	Asn	Asn	Glu	Leu	Thr	Leu	Ile	Val					
610							615							620						
Lys	Ala	Phe	His	Pro	Asp	Ser	Asn	Ile	Thr	Leu	Glu	Val	Leu	Ser	Thr					
625							630							635						
Glu	Pro	Thr	Tyr	Gln	Phe	Tyr	Thr	Gly	Asp	Phe	Leu	Ser	Ala	Gly	Tyr					
645							650							655						
Glu	Ala	Arg	Gln	Gly	Phe	Ala	Ile	Glu	Pro	Gly	Arg	Tyr	Ile	Asp	Ala					
660							665							670						
Ile	Asn	Gln	Glu	Asn	Trp	Lys	Asp	Cys	Val	Thr	Leu	Lys	Asn	Gly	Glu					
675							680							685						
Thr	Tyr	Gly	Ser	Lys	Ile	Val	Tyr	Arg	Phe	Ser										
690							695													

 $\langle 210 \rangle_{54}$

<211>394

 $\langle 212 \rangle$ PRT

〈213〉 智人 (Homo Sapiens)

 $\langle 400 \rangle_{54}$

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

50	55	60
Phe Phe Ala Thr Thr Ala Val Val Met Ala Glu Val Leu Lys Gly Leu		
65	70	75
Thr Cys Leu Leu Leu Leu Phe Ala Gln Lys Arg Gly Asn Val Lys His		80
	85	90
Leu Val Leu Phe Leu His Glu Ala Val Leu Val Gln Tyr Val Asp Thr		95
	100	105
Leu Lys Leu Ala Val Pro Ser Leu Ile Tyr Thr Leu Gln Asn Asn Leu		110
	115	120
Gln Tyr Val Ala Ile Ser Asn Leu Pro Ala Ala Thr Phe Gln Val Thr		125
	130	135
Tyr Gln Leu Lys Ile Leu Thr Thr Ala Leu Phe Ser Val Leu Met Leu		140
145	150	155
Asn Arg Ser Leu Ser Arg Leu Gln Trp Ala Ser Leu Leu Leu Leu Phe		160
	165	170
Thr Gly Val Ala Ile Val Gln Ala Gln Gln Ala Gly Gly Gly Gly Pro		175
	180	185
Arg Pro Leu Asp Gln Asn Pro Gly Ala Gly Leu Ala Ala Val Val Ala		190
	195	200
Ser Cys Leu Ser Ser Gly Phe Ala Gly Val Tyr Phe Glu Lys Ile Leu		205
	210	215
Lys Gly Ser Ser Gly Ser Val Trp Leu Arg Asn Leu Gln Leu Gly Leu		220
225	230	235
Phe Gly Thr Ala Leu Gly Leu Val Gly Leu Trp Trp Ala Glu Gly Thr		240
	245	250
Ala Val Ala Thr Arg Gly Phe Phe Phe Gly Tyr Thr Pro Ala Val Trp		255
	260	265
Gly Val Val Leu Asn Gln Ala Phe Gly Gly Leu Leu Val Ala Val Val		270
	275	280
Val Lys Tyr Ala Asp Asn Ile Leu Lys Gly Phe Ala Thr Ser Leu Ser		285
	290	295
Ile Val Leu Ser Thr Val Ala Ser Ile Arg Leu Phe Gly Phe His Val		300
305	310	315
Asp Pro Leu Phe Ala Leu Gly Ala Gly Leu Val Ile Gly Ala Val Tyr		320
	325	330
Leu Tyr Ser Leu Pro Arg Gly Ala Ala Lys Ala Ile Ala Ser Ala Ser		335
	340	345
Ala Ser Ala Ser Gly Pro Cys Val His Gln Gln Pro Pro Gly Gln Pro		350
	355	360
		365

Pro Pro Pro Gln Leu Ser Ser His Arg Gly Asp Leu Ile Thr Glu Pro
 370 375 380
 Phe Leu Pro Lys Ser Val Leu Val Lys Glx
 385 390

<210>55

<211>397

<212>PRT

<213> 智人 (Homo Sapiens)

<400>55

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

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

<210>56

<211>357

<212>PRT

<213> 黑腹果蝇 (*Drosophila melanogaster*)

<400>56

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

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

<210>57

<211>353

<212>PRT

<213> 粟酒裂殖糖酵母 (Schizosaccharomyces pombe)

<400>57

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

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

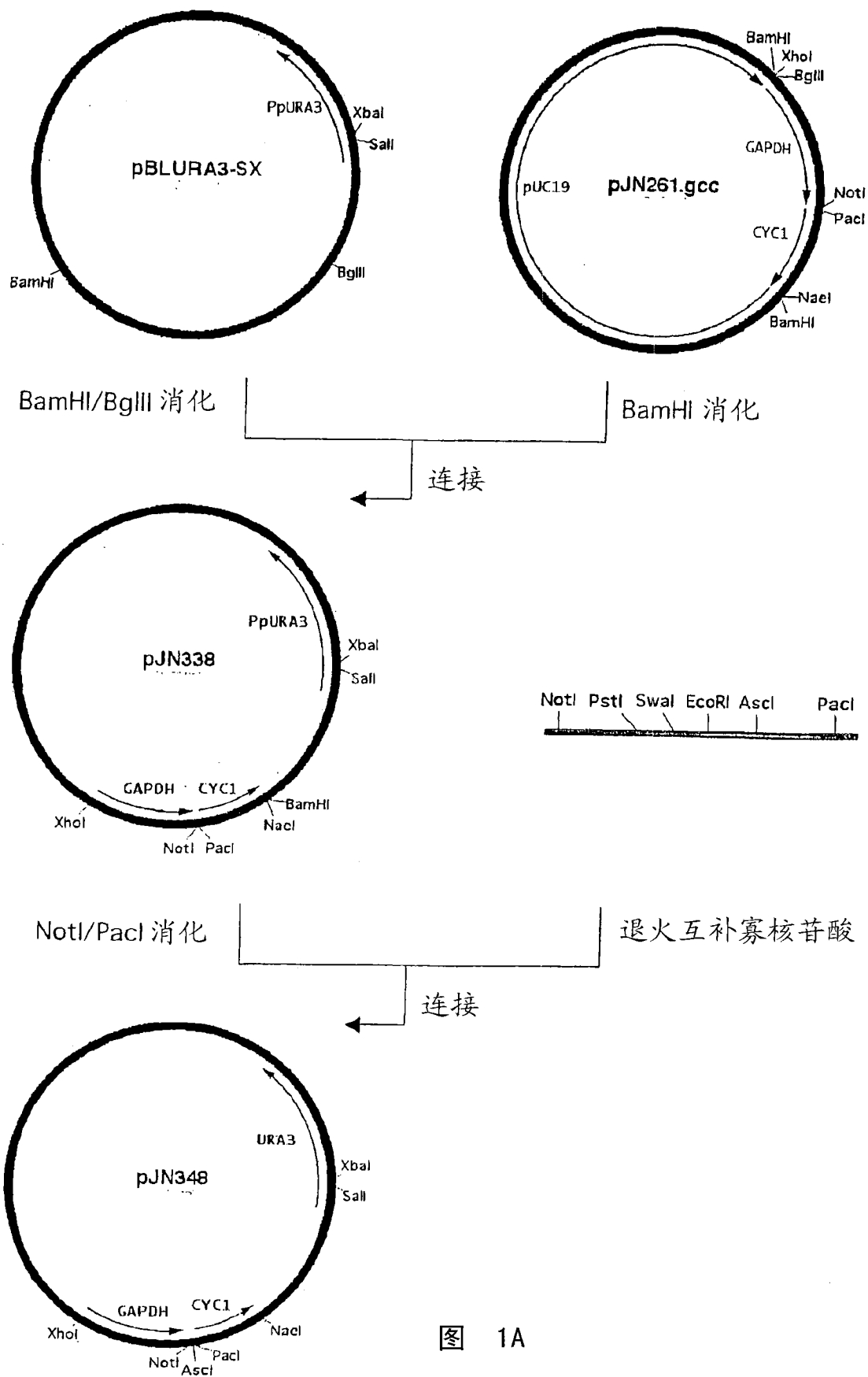


图 1A

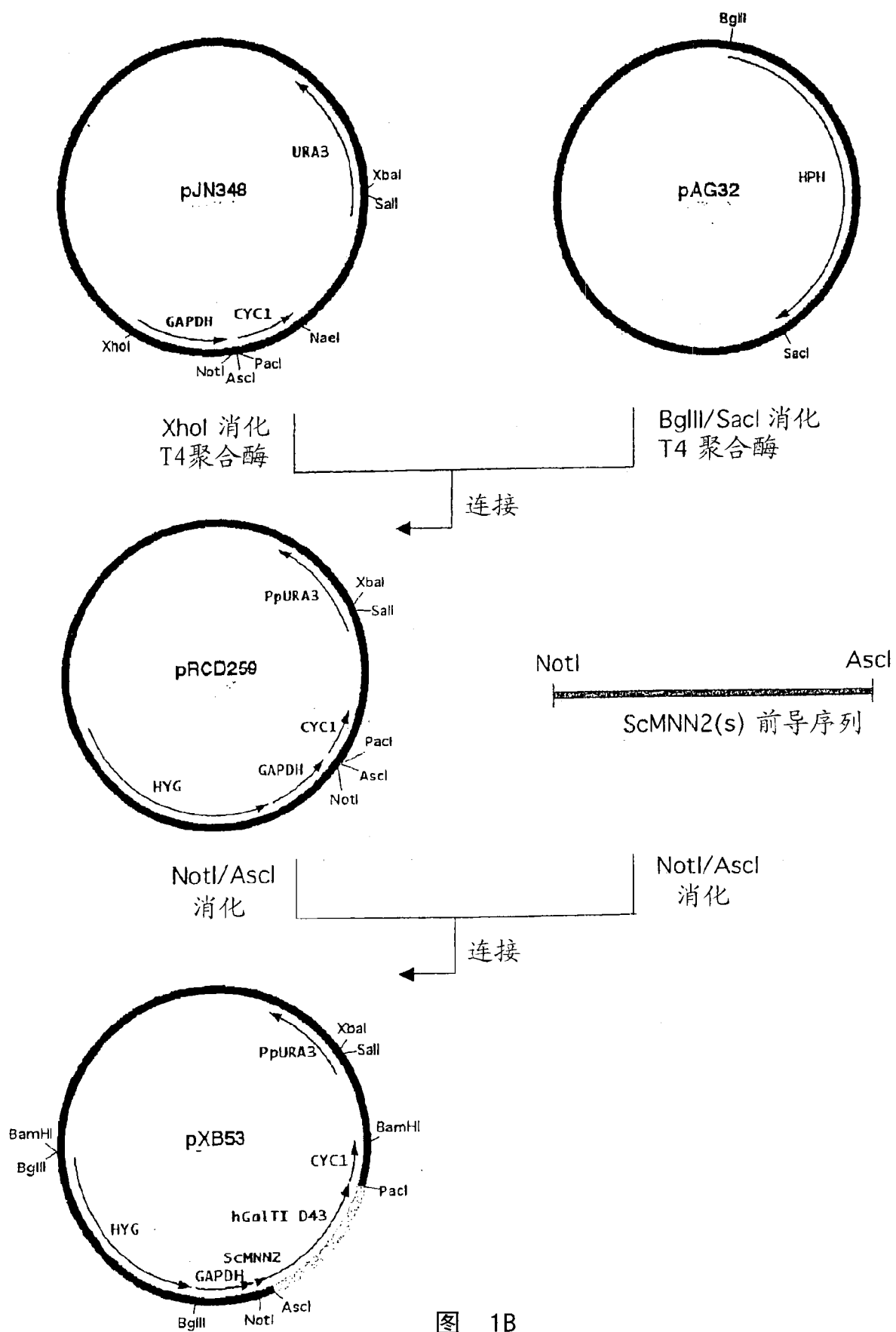


图 1B

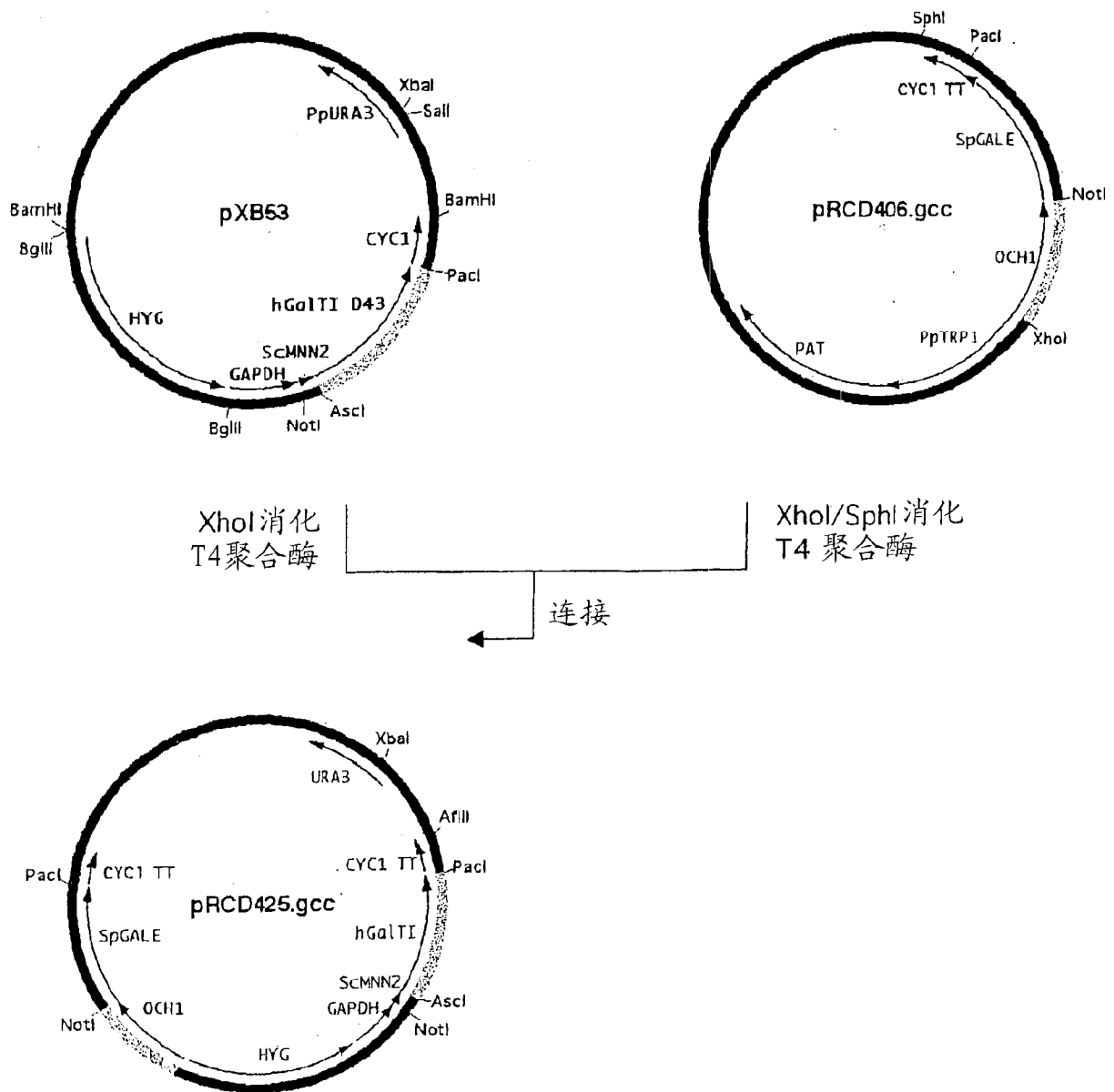


图 2

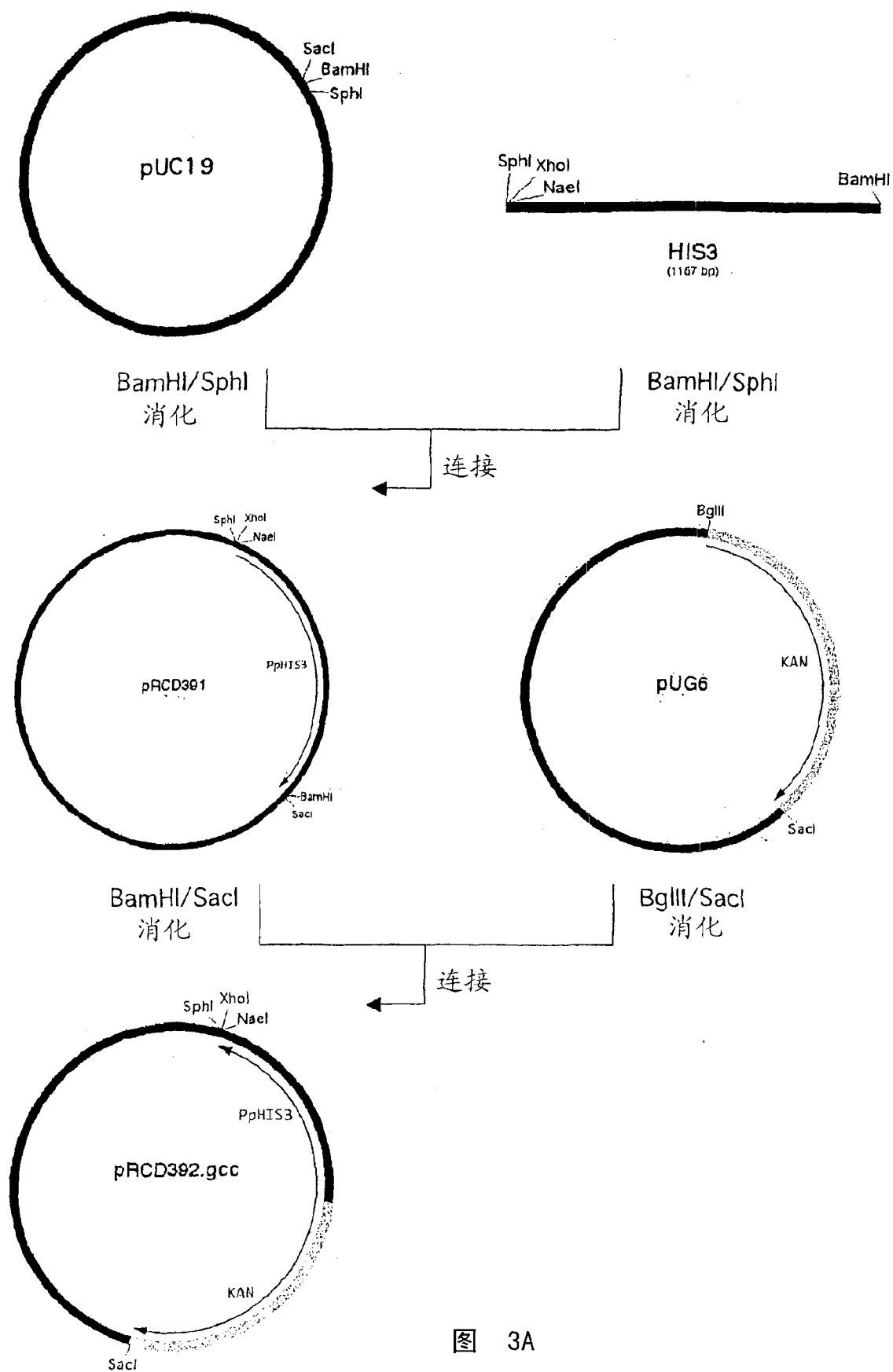


图 3A

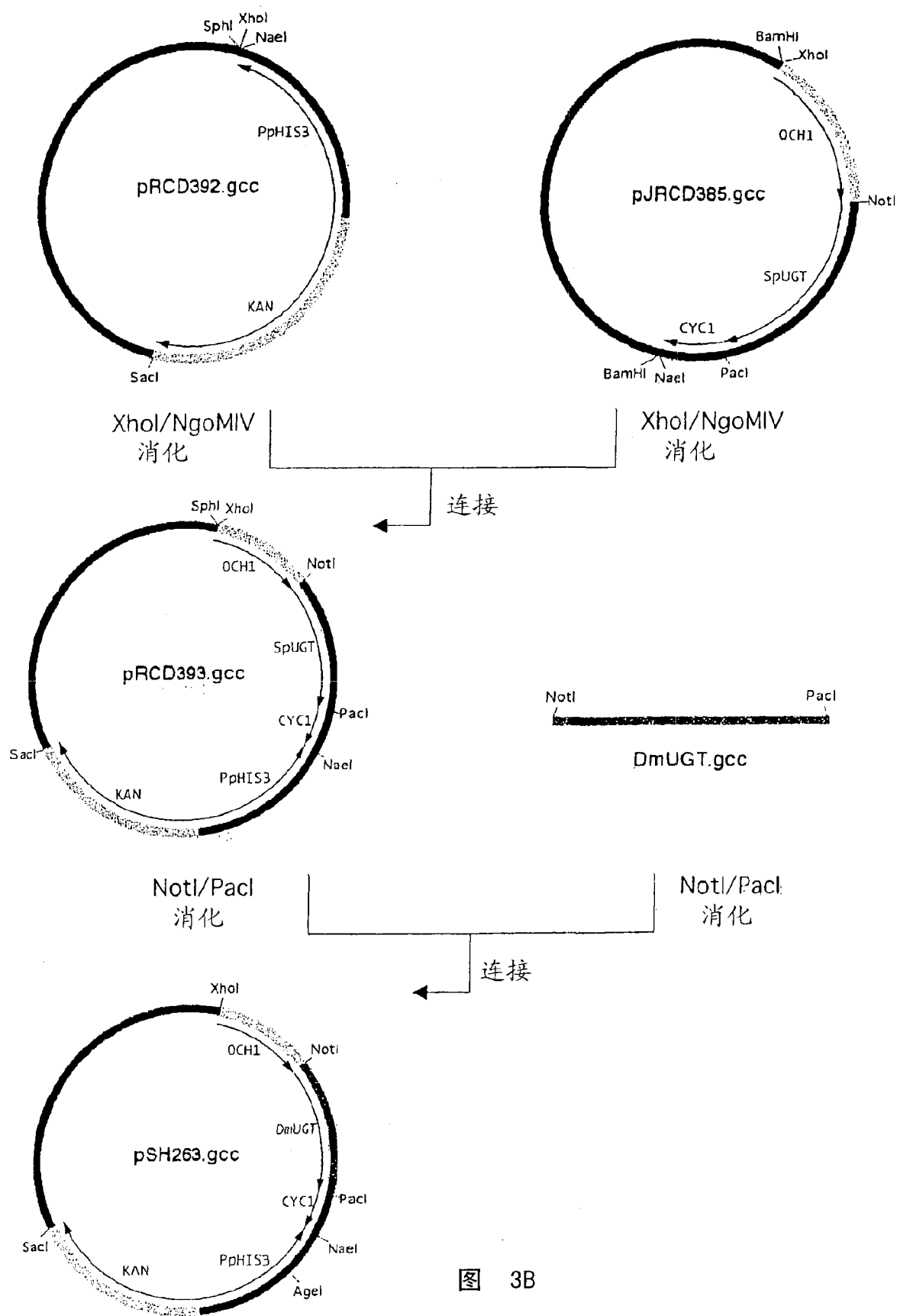


图 3B

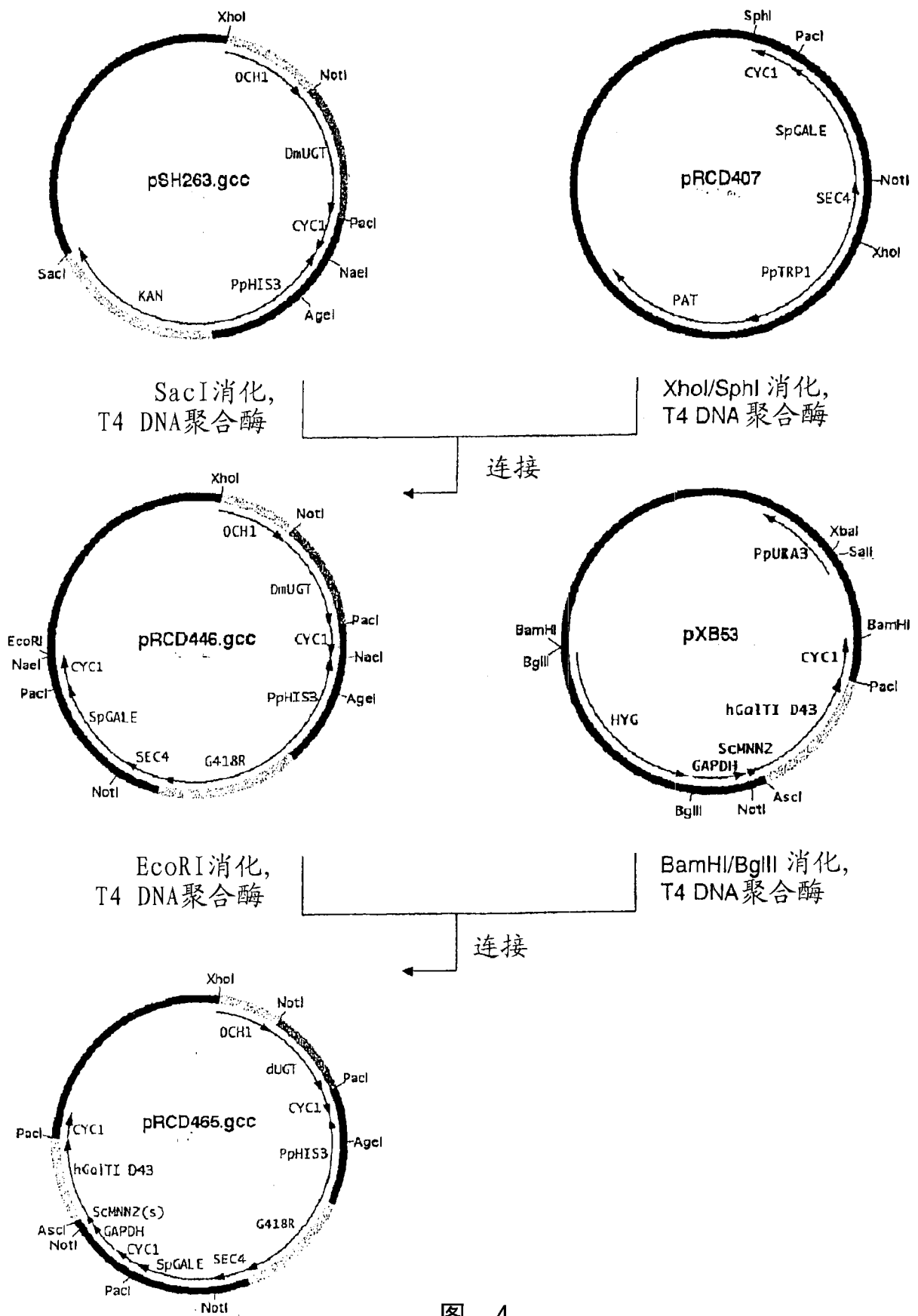


图 4

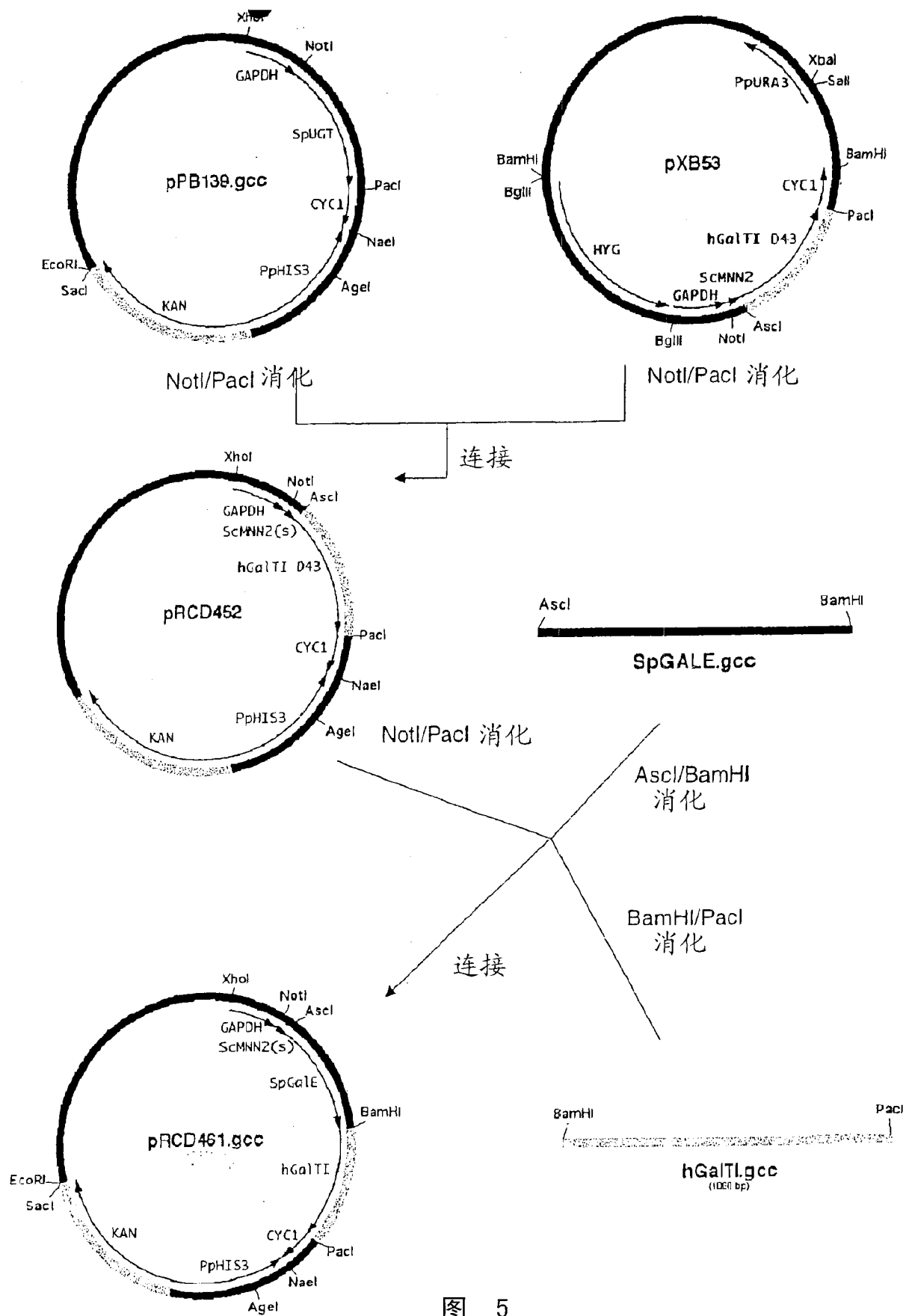


图 5

SpGalE

氨基酸序列

MTGVHEGTVLVTGGAGYIGSHTCVVLLLEKGYDVVIVDNLCNSRVEAVHRIEK
LTGKKVIFHQVDLLDEPALDKVFANQNISAVIHFAGLKAVGESVQVPLSYK
NNISGTINLIECMKKYNVRDFVFSSSATVYGDPTRPGGTIPIPESCPREGTS
PYGRTKLFIENIIEDETKVNKSLNAALLRYFNPGGAHPSGELGEDPLGI PNN
LLPYIAQVAVGRLDHLNVFGDDYPTSDGTPIRDYIHVCDLAEAHVAALDYLR
QHFVSCRPNLGSSTGTVFQVLNAFSKAVGRDLPYKVTPRRAGDVVNL TAN
PTRANEELKWKTSRSIYEICVDTWRWQQKYPYGFDLTHTKTYK

图 6A

SpGalE

1068bp 编码序列

atgactggtgtcatgaaggactgtgttggtactggcggcgtggttatataggttctcatcgtgcgtgtttt
gttagaaaaaggatatgatgttgtaattgicgataatttatgcaattctcgcttgaagccgtgcaccgcattgaa
aaactcactgggaaaaaagtcatttccaccaggtggalttgcttgatgagccagctttggacaaggtcttcgc
aatcaaaacatatctgctgtcattcattttgctggtctcaaagcagttggtgaatctgtacaggttcctttgagtta
ttacaaaaataacatttccggtaccattaatttaataagagtgcataagaagataatgtacgtgacttcgtcttttc
ttcatctgctaccgtgtatggcgatcctactagacctggtggtaccattcctattccagagtcattgcctcgtgaa
ggtacaagcccatatggtcgacaaaagctttcattgaaaatatcattgaggatgagaccaaggtgaacaaat
cgcttaatgcagctttattacgctattttaatcccgagggtgctcatccctctggtgaactcggtgaagatcctctt
ggcatccctaataacttgcttcttatacgcgcaagttgctgtaggaagattggatcattgaaatgtatttggcg
acgattatcccacatctgacggtactccaattcgtgactacattcacgtatgcgatttggcagaggctcatgttg
ctgctctcgattacctgcgccaacattttgtagttgccccttggaatttgggatcaggaactggttagtactgt
tttcaggtgctcaatgcgttttcgaaagctgttggaagagatcttccttataaggtcacccttagaagagcagg
ggacgttgtaacctaacccgccaacccccactcgcgctaacgaggagttaaaatggaaaaccagtcgtagcat
ttatgaaatttgcgttgacacttgagatggcaacagaagtatccctatggctttgacctgacctataccaaga
catataagtaa

图 6B

```

1  MTGVHEGT - - - - - VLVTGGAGYIGSHTCVVLLLEKGYDVVIV SpGalBp
1  MAEK - - - - - VLVTGGAGYIGSHTVLELLEAGYLPVVI hGalBp
1  M - - - - - VLVTGGSGYIGSHTCVQLLQNGHDV IIL EcGalBp
1  MTAQLQSESTSKI VLVTGGAGYIGSHTVVELIENGYDCVVA ScGal10p

37  DNLCNSRV - - - - - EAVHRIEKLGTGKKVIFHQVDLLDEPAL SpGalBp
33  DNFNHNAFRGGGSLPEESLRRVQELTGRSVEFEEMDILDOGAL hGalBp
31  DNLCNSKRSV - - - - - LPVIERLGKHPFTFVEGDIRNEALM EcGalBp
42  DNLSNSTY - - - - - DSVARLEVLTKHHIPFYEVDLCLDRKGL ScGal10p

72  DKVFNANONISAVIHFAGLKAVGESVQVPLSYKNNISGTIN SpGalBp
74  QRLFKKYSFMAVIHFAGLKAVGESVQKPLDYRVNLTGTIQ hGalBp
66  TEILHDHAIDTVIHFAGLKAVGESVQKPLEYYDNNVNGTLR EcGalBp
77  EKVFKKYKIDSVIHFAGLKAVGESTQIPLRYYHNNILGTVV ScGal10p

113  LIECMKKYNVRDFVFSSSATVYGDPTRPGGTIPIPESCPRRE SpGalBp
115  LLEIMKAHGKLVFSSSATVYGNPQY - - - EPLDEAHPTG hGalBp
107  LISAMRAANVKNFVFSSSATVYGDNPK - - - IPYVESHPTEG EcGalBp
118  LLELMQOYNVSKFVFSSSATVYGDATRFPMNIPIPPECPPLG ScGal10p

154  G - TSPYGRTKLFIENIIEDETQVNNK - SLNAALLRYFNPGGA SpGalBp
152  GCTNPPYKSKKEFIEEMIRDLQADK - TWNVLLRYFNPTGA hGalBp
144  TPQSPYKSKLMVEQILTDLQKAP - DWSIALLRYFNPVGA EcGalBp
159  P - TNPYGHTKYATENILNDLYNSDKKSWKFAILRYFNPIGA ScGal10p

193  HPSGELGEDPLGIPNNLLPYIAQVAVGRLDH LNVFGDDYPT SpGalBp
192  HASGCIGEDPQGIPNNLMPYVSQVATGRREALNVFGNDYDT hGalBp
184  HPSGDMGEDPQGIPNNLMPYIAQVAVGRRLSLAIFGNDYPT EcGalBp
199  HPSGLIGEDPLGIPNNLLPYMAQVAVGRREKLYIFGDDYDS ScGal10p

234  SDGTPIRDYIHVQDLAEAHVAALDYL - - - RQH FVS CRPWNL SpGalBp
233  EDGTGVRDYIHVVDLAKGHTAALRKL - - - KEQC GCRIYNL hGalBp
225  EDGTGVRDYIHVMDLADGHVVA MEKL - - - ANKP GVHIYNL EcGalBp
240  RDGTPIRDYIHVVDLAKGHTAALQYIEAYNENEGL CREWNL ScGal10p

272  GSGTGSTVFEQVLNAFSKAVGRDLPYKVTERRAGDVVNLTAN SpGalBp
270  GTGTGYSVLQMVQAMEKASGKIPYKVVARREGDVAAACYAN hGalBp
262  GAGVGN SVLDVVNAFSKACGKPVNYHEAPRR ECDLPA YWAD EcGalBp
281  GSGKGSTVFEVYHAECKASGITDLPYKV TGRRAGDV LNLTA K ScGal10p

313  PTRANEELKWKTSRSIYEICVDTWRRWQKQYLYGFDTLTHTKT SpGalBp
311  PSLAQEELGWTAALGLDFMCB DLWRWQKQNP SGFGTQA hGalBp
303  ASKADRELNW RVTRTLDEMAQD TWHWQSRHP QGY - - PD EcGalBp
322  PDRAKRELKWKTELEQVEDSC KDLWKET TENP FG YQLRGVEA ScGal10p

```

图 7

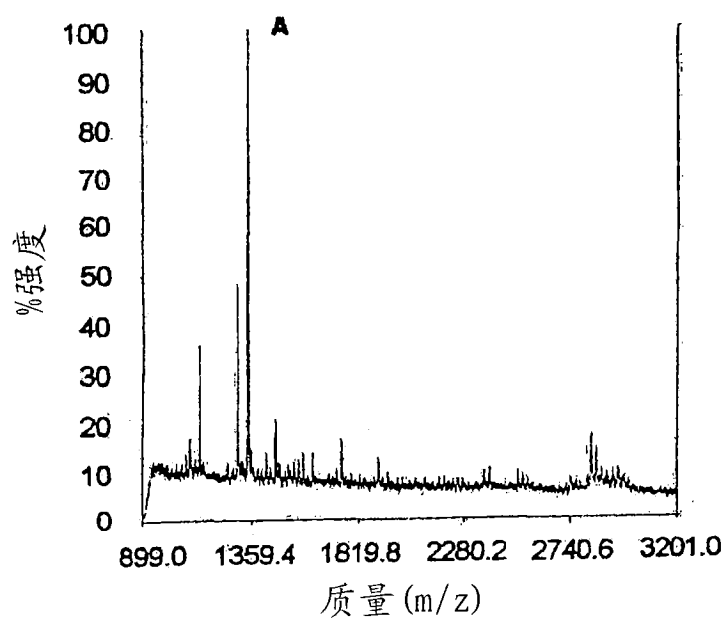


图 8A

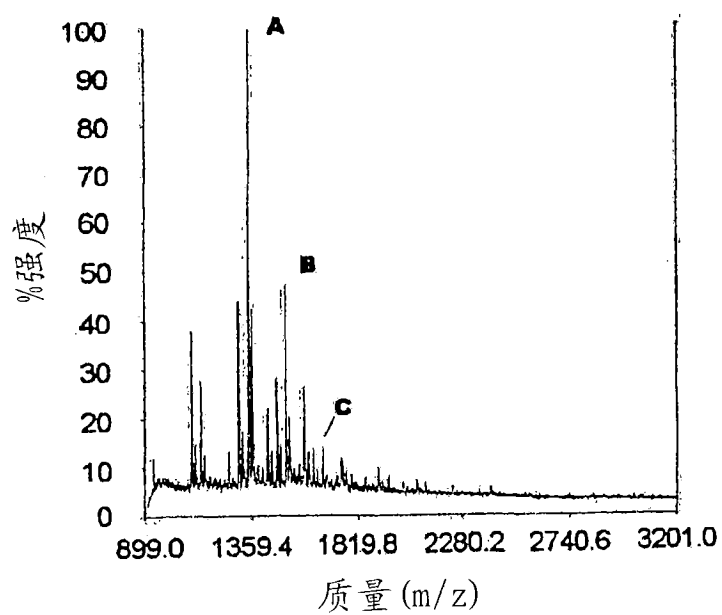
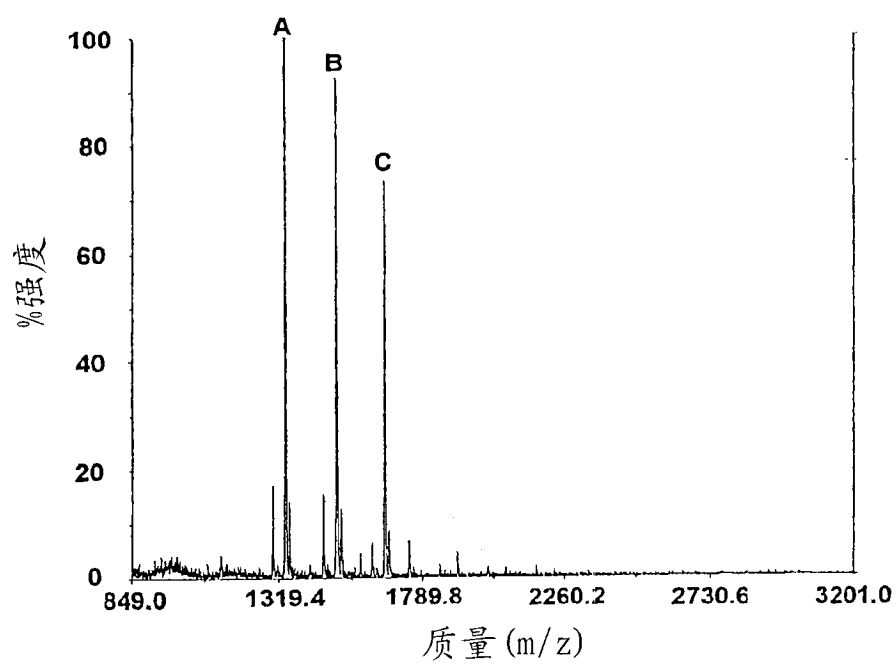
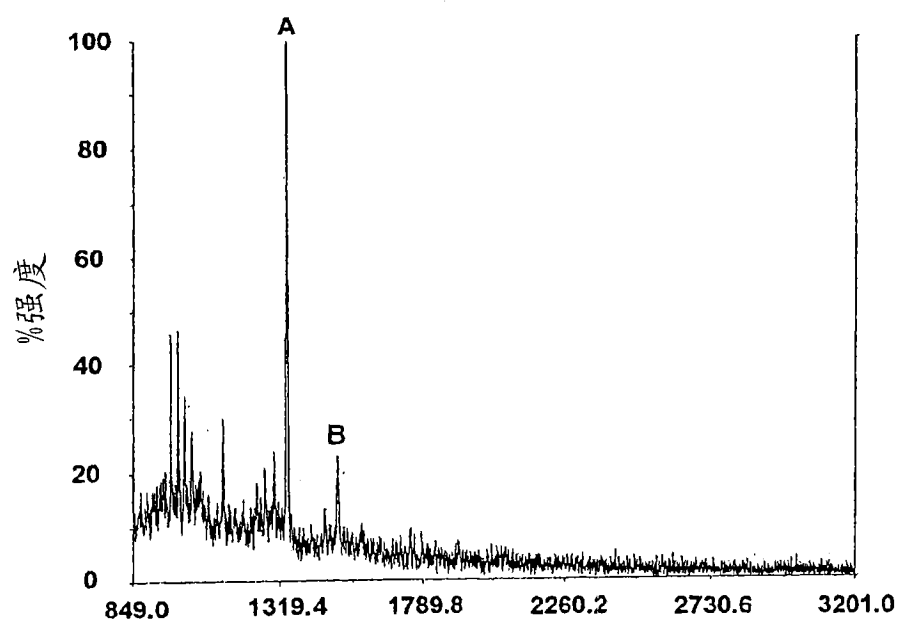


图 8B



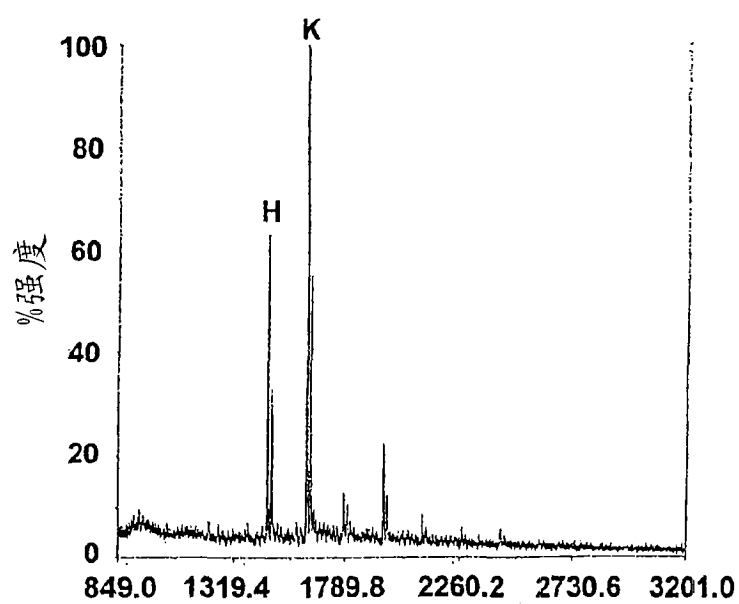


图 10A

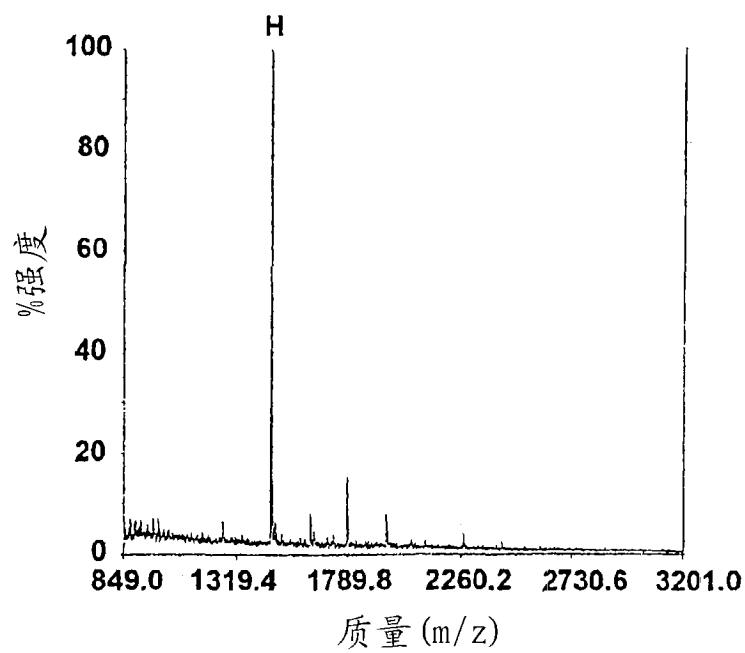


图 10B

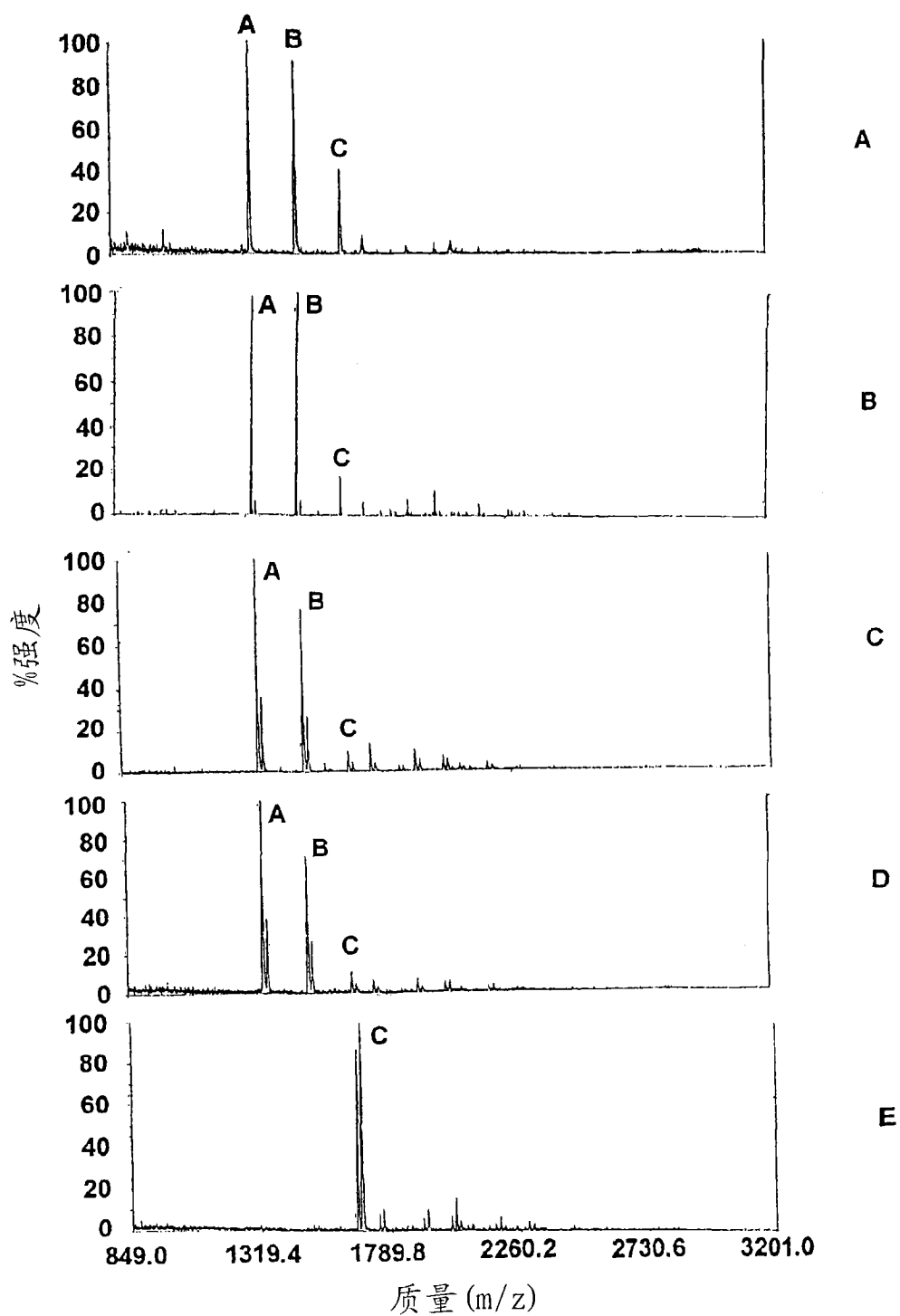


图 11

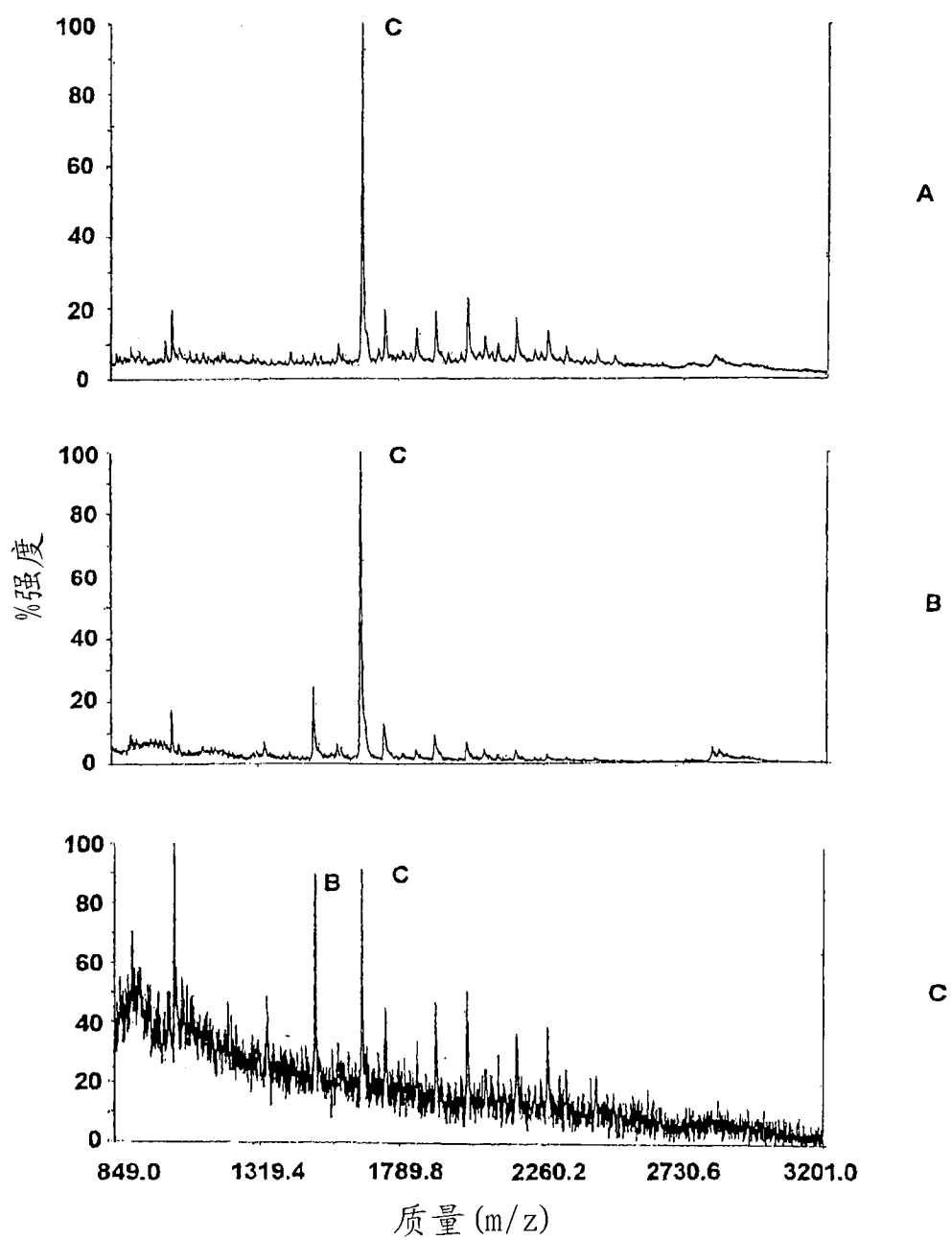


图 12

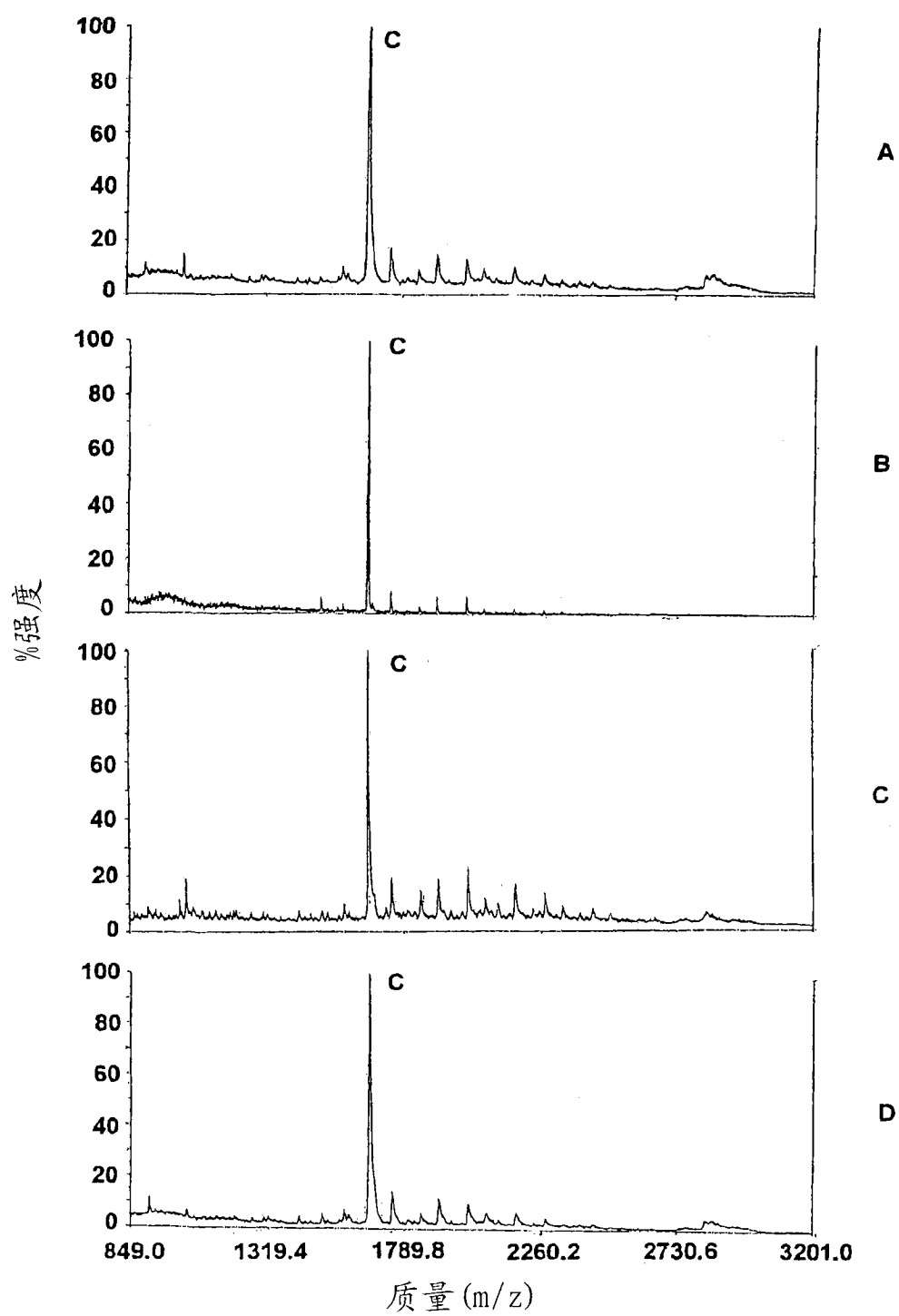


图 13

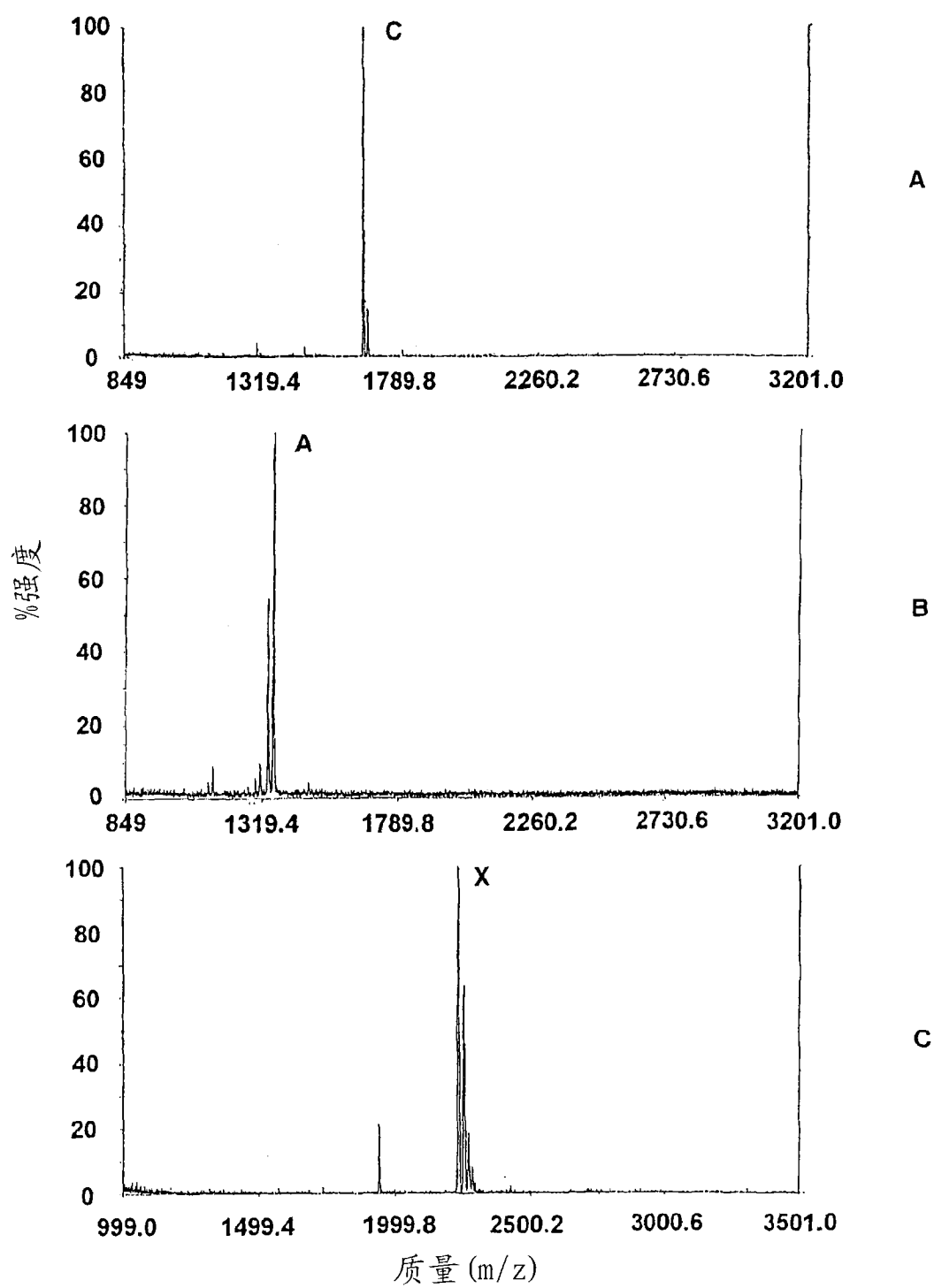


图 14

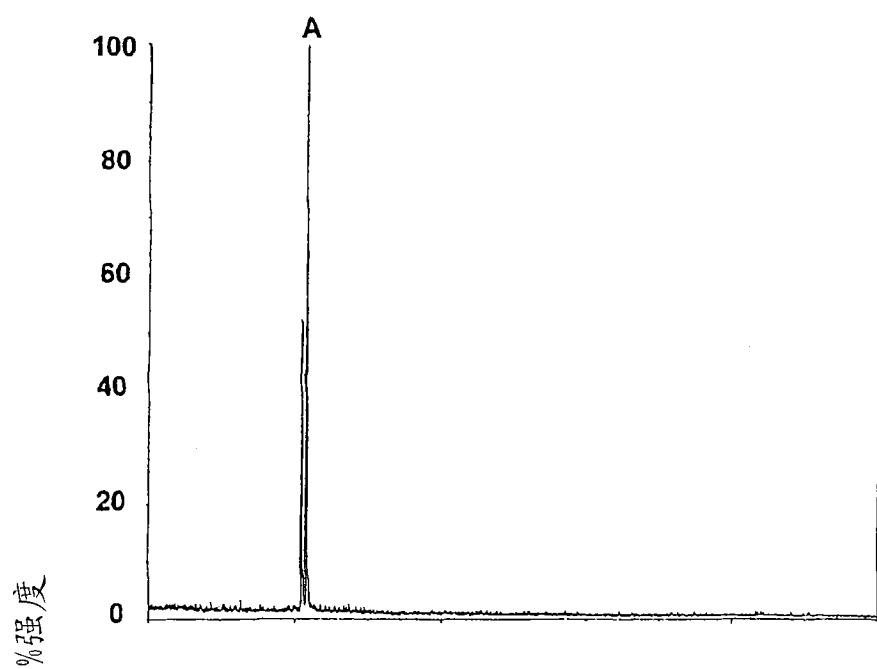


图 15A

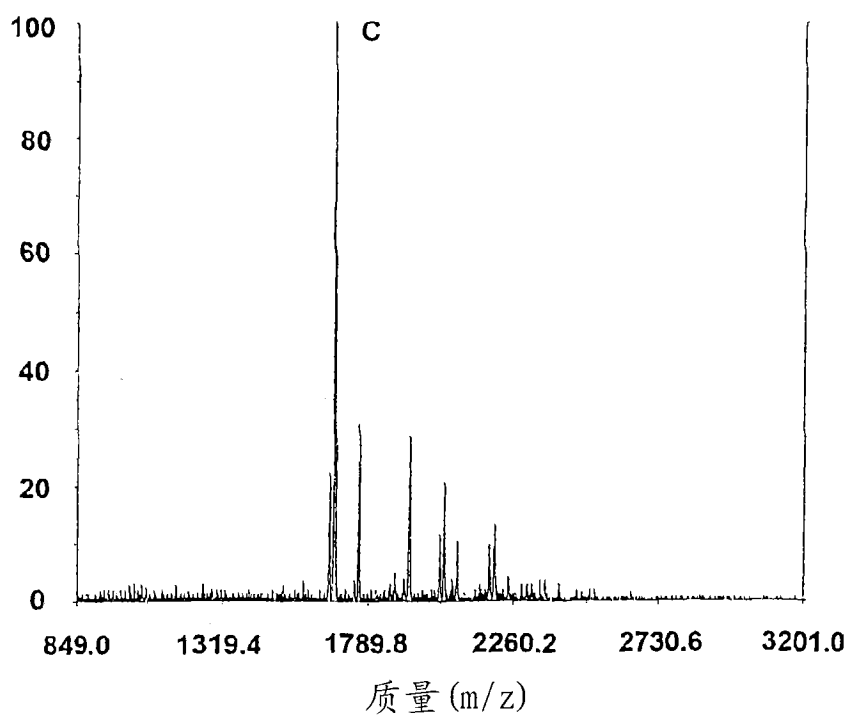


图 15B