



(12) 发明专利申请



(10) 申请公布号 CN 119110849 A

(43) 申请公布日 2024.12.10

(21) 申请号 202380031039.7

M·萨洛埃莫

(22) 申请日 2023.01.26

(74) 专利代理机构 北京汉智嘉成知识产权代理

(30) 优先权数据

有限公司 11682

63/303,517 2022.01.27 US

专利代理师 方珉 鲁山

(85) PCT国际申请进入国家阶段日

(51) Int.Cl.

2024.09.26

C12P 21/00 (2006.01)

(86) PCT国际申请的申请数据

C12N 9/10 (2006.01)

PCT/IB2023/050667 2023.01.26

C07K 14/37 (2006.01)

(87) PCT国际申请的公布数据

C12N 15/80 (2006.01)

W02023/144739 EN 2023.08.03

(71) 申请人 二进国际(美国)有限公司

地址 美国佛罗里达州

(72) 发明人 A·于斯科南 R·切莱特

M·A·艾玛法布

N·瓦布伦纳克斯波

权利要求书4页 说明书28页

序列表(电子公布) 附图12页

(54) 发明名称

THERMOTHELOMYCES HETEROHALLICA的糖工程

(57) 摘要

提供了被遗传修饰以产生具有哺乳动物蛋白质(特别是人类、伴侣动物和其他动物蛋白质)的N-聚糖的糖蛋白的Thermothelomyces heterothallica(原名嗜热毁丝霉(Myceliophthora thermophila)),其包含alg3基因的缺失或破坏、靶向ER的甘露糖苷酶1(α -1,2-甘露糖苷酶)的表达和靶向ER的葡萄糖苷酶2 α 亚基的表达。所述Th.heterothallica还可以进一步包含异源GlcNAc转移酶1(GNT1)、GlcNAc转移酶2(GNT2)、异源寡糖基转移酶的STT3亚基和半乳糖基转移酶。

1. 一种 *Thermothelomyces heterothallica*, 其被遗传修饰以产生具有哺乳动物 N-聚糖的糖蛋白, 其中所述遗传修饰包括:

(i) *alg3* 基因的缺失或破坏, 使得所述 *Th.heterothallica* 不能产生功能性 α -1,3-甘露糖基转移酶;

(ii) 靶向 ER 的甘露糖苷酶 1 (α -1,2-甘露糖苷酶) 的表达; 和

(iii) 靶向 ER 的葡萄糖苷酶 2 α 亚基的表达。

2. 根据权利要求 1 所述的 *Th.heterothallica*, 其中所述甘露糖苷酶 1 是里氏木霉 (*Trichoderma reesei*) 甘露糖苷酶。

3. 根据权利要求 1 或 2 中任一项所述的 *Th.heterothallica*, 其中所述葡萄糖苷酶 2 α 亚基是 *Th.heterothallica* 葡萄糖苷酶 2 α 亚基。

4. 根据权利要求 1-3 中任一项所述的 *Th.heterothallica*, 其中所述遗传修饰进一步包括异源 GlcNAc 转移酶 1 (GNT1) 和 GlcNAc 转移酶 2 (GNT2) 的表达。

5. 根据权利要求 4 所述的 *Th.heterothallica*, 其中所述异源 GNT1 和 GNT2 是动物来源的。

6. 根据权利要求 5 所述的 *Th.heterothallica*, 其中所述动物来源的 GNT1 是人 GNT1。

7. 根据权利要求 6 所述的 *Th.heterothallica*, 其中所述动物来源的 GNT1 是进一步包含 *Th.heterothallica* 高尔基体定位信号的人 GNT1。

8. 根据权利要求 7 所述的 *Th.heterothallica*, 其中所述 *Th.heterothallica* 高尔基体定位信号来自 C1 蛋白 KRE2。

9. 根据权利要求 8 所述的 *Th.heterothallica*, 其中所述包含来自 C1 蛋白 KRE2 的高尔基体定位信号的人 GNT1 包含 SEQ ID NO:6 中阐述的氨基酸序列。

10. 根据权利要求 9 所述的 *Th.heterothallica*, 其中所述包含来自 C1 蛋白 KRE2 的高尔基体定位信号的人 GNT1 由引入到所述 *Th.heterothallica* 中的包含 SEQ ID NO:5 中阐述的序列的外源多核苷酸编码。

11. 根据权利要求 5 所述的 *Th.heterothallica*, 其中所述动物来源的 GNT1 是牛 GNT1。

12. 根据权利要求 11 所述的 *Th.heterothallica*, 其中所述动物来源的 GNT1 是包含 *Th.heterothallica* 高尔基体定位信号的牛 GNT1。

13. 根据权利要求 12 所述的 *Th.heterothallica*, 其中所述包含来自 *Th.heterothallica* 蛋白 KRE2 的高尔基体定位信号的牛 GNT1 包含 SEQ ID NO:10 中阐述的氨基酸序列。

14. 根据权利要求 12 所述的 *Th.heterothallica*, 其中所述包含来自 *Th.heterothallica* 蛋白 KRE2 的高尔基体定位信号的牛 GNT1 由引入到所述 *Th.heterothallica* 中的包含 SEQ ID NO:9 中阐述的序列的外源多核苷酸编码。

15. 根据权利要求 5 所述的 *Th.heterothallica*, 其中所述动物来源的 GNT2 是大鼠 GNT2。

16. 根据权利要求 15 所述的 *Th.heterothallica*, 其中所述大鼠 GNT2 包含 SEQ ID NO:15 中阐述的氨基酸序列。

17. 根据权利要求 15 所述的 *Th.heterothallica*, 其中所述大鼠 GNT2 由引入到所述 *Th.heterothallica* 中的包含 SEQ ID NO:16 中阐述的序列的外源多核苷酸编码。

18. 根据权利要求 5 所述的 *Th.heterothallica*, 其中所述动物来源的 GNT1 是包含来自

Th.heterothallica蛋白KRE2的高尔基体定位信号的人GNT1,并且所述动物来源的GNT2是大鼠GNT2。

19.根据权利要求5所述的Th.heterothallica,其中所述动物来源的GNT1是包含来自Th.heterothallica蛋白KRE2的高尔基体定位信号的牛GNT1,并且所述动物来源的GNT2是大鼠GNT2。

20.根据前述权利要求中任一项所述的Th.heterothallica,其中所述遗传修饰进一步包括内源翻转酶的过表达或异源翻转酶的表达。

21.根据权利要求20所述的Th.heterothallica,其被遗传修饰以表达外源翻转酶,其中所述翻转酶是Th.heterothallica RFT1翻转酶。

22.根据权利要求21所述的Th.heterothallica,其中所述Th.heterothallica RFT1翻转酶包含SEQ ID NO:8中阐述的氨基酸序列。

23.根据权利要求21所述的Th.heterothallica,其中所述Th.heterothallica RFT1翻转酶由引入到所述Th.heterothallica中的包含SEQ ID NO:7中阐述的序列的外源多核苷酸编码。

24.根据权利要求20所述的Th.heterothallica,其被遗传修饰以过表达所述内源Th.heterothallica RFT1翻转酶。

25.根据前述权利要求中任一项所述的Th.heterothallica,其中所述遗传修饰进一步包括异源寡糖基转移酶的STT3亚基的表达。

26.根据权利要求25所述的Th.heterothallica,其中所述STT3是大型利什曼原虫(*Leishmania major*)STT3。

27.根据权利要求26所述的Th.heterothallica,其中所述大型利什曼原虫(*Leishmania major*)STT3包含SEQ ID NO:12中阐述的氨基酸序列。

28.根据权利要求26所述的Th.heterothallica,其中所述大型利什曼原虫(*Leishmania major*)STT3由引入到所述Th.heterothallica中的包含SEQ ID NO:11中阐述的序列的外源多核苷酸编码。

29.根据前述权利要求中任一项所述的Th.heterothallica,其中所述遗传修饰进一步包括异源半乳糖基转移酶的表达。

30.根据权利要求29所述的Th.heterothallica,其中所述异源半乳糖基转移酶是动物来源的半乳糖基转移酶。

31.根据权利要求30所述的Th.heterothallica,其中所述动物来源的半乳糖基转移酶是人半乳糖基转移酶。

32.根据权利要求31所述的Th.heterothallica,其中所述包含酿酒酵母(*S.cerevisiae*)KRE2高尔基体定位信号的人半乳糖基转移酶包含SEQ ID NO:14中阐述的氨基酸序列。

33.根据权利要求31所述的Th.heterothallica,其中所述包含酿酒酵母(*S.cerevisiae*)KRE2高尔基体定位信号的人半乳糖基转移酶由引入到所述Th.heterothallica中的包含SEQ ID NO:13中阐述的序列的外源多核苷酸编码。

34.根据前述权利要求中任一项所述的Th.heterothallica,其中所述Th.heterothallica是Th.heterothallica C1,其中所述C1是被修饰以缺失一个或多个编

码内源蛋白酶的基因的菌株。

35. 根据前述权利要求中任一项所述的 *Th.heterothallica*, 其中所述 *Th.heterothallica* 是 *Th.heterothallica* C1, 其中所述 C1 是被修饰以缺失编码内源几丁质酶的基因的菌株。

36. 根据权利要求 1-35 中任一项所述的 *Th.heterothallica*, 其中所述 *Th.heterothallica* 是 *Th.heterothallica* C1, 其中所述 C1 是选自下述的菌株: 野生型 C1 保藏号 VKM F-3500D, UV13-6 保藏号 VKM F-3632D, NG7C-19 保藏号 VKM F-3633D, UV18-25 保藏号 VKM F-3631D, W1L#100I (prt- Δ alp1 Δ chil Δ alp2 Δ pyr5) 保藏号 CBS141153, UV18-100f (prt- Δ alp1, Δ pyr5) 保藏号 CBS141147, W1L#100I (prt- Δ alp1 Δ chil Δ pyr5) 保藏号 CBS141149, 和 UV18-100f (prt- Δ alp1 Δ pep4 Δ alp2 Δ prt1 Δ pyr5) 保藏号 CBS141143。

37. 根据前述权利要求中任一项所述的 *Th.heterothallica*, 其被进一步遗传修饰以表达异源哺乳动物糖蛋白。

38. 根据权利要求 37 所述的 *Th.heterothallica*, 其中所述异源哺乳动物糖蛋白是抗体或其抗原结合片段。

39. 一种用于产生生产具有哺乳动物 N-聚糖的糖蛋白的 *Th.heterothallica* 的方法, 所述方法包括:

(a) 缺失或破坏所述 *Th.heterothallica* 的 *alg3* 基因, 使得所述 *Th.heterothallica* 不能产生功能性 α -1,3-甘露糖基转移酶; 和

(b) 在所述 *Th.heterothallica* 中引入编码靶向 ER 的甘露糖苷酶 1 (α -1,2-甘露糖苷酶) 和靶向 ER 的葡萄糖苷酶 2 α 亚基的外源多核苷酸。

40. 根据权利要求 39 所述的方法, 其进一步包括在所述 *Th.Heterothallica* 中引入: 编码内源翻转酶的外源多核苷酸, 以诱导所述内源翻转酶在所述 *Th.Heterothallica* 中的过表达; 或编码异源翻转酶的外源多核苷酸, 以诱导所述异源翻转酶在所述 *Th.Heterothallica* 中的表达。

41. 一种用于生产具有哺乳动物 N-聚糖的糖蛋白的方法, 所述方法包括:

(a) 提供根据权利要求 1-40 中任一项所述的遗传修饰的 *Th.heterothallica*;

(b) 将所述 *Th.heterothallica* 在适合于表达所述糖蛋白的条件下培养; 和

(c) 回收所述糖蛋白。

42. 根据权利要求 41 所述的方法, 其中所述糖蛋白是在所述 *Th.heterothallica* 中重组表达的异源哺乳动物糖蛋白。

43. 根据权利要求 41 所述的方法, 其中所述异源哺乳动物糖蛋白是在所述 *Th.heterothallica* 中重组表达的人糖蛋白。

44. 根据权利要求 41 所述的方法, 其中所述异源哺乳动物糖蛋白是抗体或其抗原结合片段。

45. 一种重组糖蛋白, 其由根据权利要求 1-39 中任一项所述的遗传修饰的 *Th.heterothallica* 产生, 其中所述糖蛋白包含 GlcNAc₂Man₃GlcNAc₂ (G0) 聚糖。

46. 一种重组糖蛋白, 其由根据权利要求 1-39 中任一项所述的遗传修饰的 *Th.heterothallica* 产生, 其中所述糖蛋白包含 Gal₁GlcNAc₂Man₃GlcNAc₂ (G1) 聚糖、Gal₂GlcNAc₂Man₃GlcNAc₂ (G2) 聚糖或其组合。

47. 根据权利要求45或权利要求46所述的重组糖蛋白,其中所述糖蛋白是医药级的。

THERMOTHELOMYCES HETEROTHALLICA的糖工程

技术领域

[0001] 本发明涉及遗传修饰的*Thermothelomyces heterothallica* (原名嗜热毁丝霉 (*Myceliophthora thermophila*)), 其中在对内源基因破坏最小的情况下对蛋白质糖基化途径进行工程化改造, 以产生具有与哺乳动物蛋白质、特别是人类蛋白质相似的N-聚糖的蛋白质。

背景技术

[0002] 大多数治疗性蛋白需要糖基化以确保正确的折叠、功能和活性。治疗性蛋白的糖基化对其免疫原性也特别重要。因此, 此类蛋白质不能在缺乏必需的糖基化机制的标准原核表达系统中生产。由于糖基化和其他翻译后修饰对治疗性糖蛋白至关重要, 因此它们中的大多数目前在哺乳动物细胞中生产。然而, 基于哺乳动物细胞培养 (例如CHO、鼠类或人类细胞) 的发酵过程通常非常缓慢, 需要昂贵的营养物和辅因子 (例如胎牛血清或特定生长因子), 通常得到低产物滴度, 并且也容易受到可能污染所得的蛋白质产品的传染源感染。因此, 正在越来越多地转向无血清表达系统。具体而言, 酵母和真菌正被开发为替代蛋白质表达系统。

[0003] 作为真核生物, 酵母和真菌能够进行翻译后修饰, 包括N-和O-糖基化, 但酵母和真菌中的蛋白质糖基化与哺乳动物细胞中的蛋白质糖基化有很大不同。为了克服这些问题, 已探索了重新工程化改造N-糖基化途径的可能性, 特别是在最常用于生产异源蛋白质的物种中 (例如酿酒酵母 (*S. cerevisiae*)、巴斯德毕赤酵母 (*Pichia pastoris*)、解脂耶氏酵母 (*Yarrowia lipolytica*)、多形汉逊酵母 (*Hansenula polymorpha*) 以及曲霉属和木霉属物种)。然而, 蛋白质产量仍需要改进, 特别是需要改进糖基化分步, 使得所生产的蛋白质中有很高的比例携带所需的糖型, 即哺乳动物 (特别是人类和伴侣动物) 蛋白质的糖型。

[0004] Parsaie Nasab等, 2013, *Appl Environ Microbiol.*, 79(3): 997-1007记载了一种合成的N-糖基化途径, 以在酿酒酵母 (*Saccharomyces cerevisiae*) 中产生携带人N-聚糖的重组蛋白。使用了 $\Delta alg3 \Delta alg11$ 双突变菌株, 对其进行了进一步遗传修饰, 以表达人工翻转酶、原生动物寡糖基转移酶和靶向高尔基体的人N-乙酰氨基葡萄糖转移酶I和II。结果证实, 在突变菌株中重组表达的分泌型单克隆抗体上存在复杂的人N-聚糖结构 $GlcNAc_2Man_3GlcNAc_2$ 。然而, 由于高尔基体定位的甘露糖基转移酶的干扰, 观察到N-连接聚糖的非均质性。

[0005] Parsaie Nasab等人的工作也被记载在US 2011/0207214中, 其公开了被修饰以表达脂质连接寡糖 (LLO) 翻转酶活性的细胞, 所述活性能够将包含1个甘露糖残基、2个甘露糖残基和3个甘露糖残基数的LLO从胞质一侧翻转到细胞内细胞器的管腔一侧, 并与其他相关研究一起在De Wachter等, 2018, “酵母糖蛋白表达的工程化” (Engineering of Yeast Glycoprotein Expression, 在《生物化学工程/生物技术进展》(Advances in Biochemical Engineering/Biotechnology), Springer, Berlin, Heidelberg中) 中进行了进一步综述。

[0006] US 7,029,872、US 7,326,681、US 7,629,163、US 7,981,660公开了具有遗传修饰

的糖基化途径的细胞系,这些途径允许它们进行一系列模拟人类中糖蛋白的加工的酶促反应。通常产生含高甘露糖的N-聚糖的真核生物例如单细胞和多细胞真菌被修饰,以产生N-聚糖例如Man₅GlcNAc₂或沿着人类糖基化途径的其他结构。

[0007] US 7,449,308、US 7,935,513公开了具有修饰的寡糖的真核宿主细胞,其可以通过一组糖基转移酶、糖转运蛋白和甘露糖苷酶的异源表达进一步修饰,变成用于生产哺乳动物例如人治疗性糖蛋白的宿主菌株。在所述工程化宿主细胞中制造的N-聚糖具有Man₅GlcNAc₂核心结构,其然后通过一种或多种酶例如糖基转移酶、糖转运蛋白和甘露糖苷酶的异源表达进一步修饰,产生类人糖蛋白。

[0008] US 7,795,002公开了产生类人糖蛋白的真核宿主细胞例如酵母和丝状真菌,所述类人糖蛋白的特征是具有末端 β -半乳糖残基,基本上缺乏岩藻糖和唾液酸残基。进一步公开了一种在重组真核宿主细胞中催化半乳糖残基从UDP-半乳糖转移到受体底物上的方法,所述受体底物可用作治疗性糖蛋白。

[0009] US 8,986,949公开了非哺乳动物真核生物的遗传工程菌株,其表达催化活性的内切甘露糖苷酶基因以增强N-连接聚糖结构的加工,总体目标是获得更多的类人聚糖分布。此外,还公开了新型人和小鼠内切甘露糖苷酶的克隆和表达。

[0010] US 9,359,628公开了能够产生具有较小聚糖的蛋白质的毕赤酵母遗传工程菌株。具体而言,所述遗传工程菌株能够表达 α -1,2-甘露糖苷酶和葡萄糖苷酶II中的任一者或两者。所述遗传工程菌株可以被进一步修饰,使得OCH1基因被破坏。还提供了使用此类毕赤酵母遗传工程菌株生产具有较小聚糖的糖蛋白的方法。

[0011] US 9,695,454公开了包含具有降低的蛋白酶活性并表达岩藻糖基化途径的丝状真菌细胞(例如木霉真菌细胞)的组合物。进一步描述了使用遗传修饰的丝状真菌细胞(例如木霉真菌细胞)作为表达系统生产具有岩藻糖基化N-聚糖的糖蛋白的方法。

[0012] *Thermothelomyces heterothallica* (Th.heterothallica) 菌株C1(最近从嗜热毁丝霉(*Myceliophthora thermophila*)重新命名,后者从*Chrysosporium lucknowense*重新命名)是一种生产高水平纤维素酶的耐热子囊菌类丝状真菌,使其对在商业规模上生产这些和其他酶具有吸引力。

[0013] 例如,美国专利第8,268,585和8,871,493号公开了在丝状真菌宿主领域中的一种用于表达和分泌异源蛋白或多肽的转化系统。还公开了一种以经济的方式生产大量多肽或蛋白质的方法。所述系统包括转化或转染的金孢霉属(*Chrysosporium*)、更具体来说是*Chrysosporium lucknowense*的真菌菌株及其突变体或衍生菌株。还公开了含有金孢霉属(*Chrysosporium*)编码序列以及金孢霉属(*Chrysosporium*)基因的表达调控序列的转化体。

[0014] 野生型C1根据布达佩斯公约以编号VKMF-3500D保藏,保藏日期为1996年8月29日。高纤维素酶(HC)和低纤维素酶(LC)菌株也已被保藏,正如在美国专利第8,268,585号中所述。例如:菌株UV13-6,保藏号VKM F-3632 D;菌株NG7C-19,保藏号VKM F-3633 D;菌株UV18-25,保藏号VKM F-3631 D。

[0015] 已被保藏的其他改进的C1菌株包括:(i) HC菌株UV18-100f (Δ alp1 Δ pyr5),保藏号CBS141147;(ii) HC菌株UV18-100f (Δ alp1 Δ pep4 Δ alp2 Δ pyr5 Δ prt1),保藏号CBS141143;(iii) LC菌株W1L#100I (Δ chi1 Δ alp1 Δ alp2 Δ pyr5),保藏号CBS141153;和(iv) LC菌株W1L#100I (Δ chi1 Δ alp1 Δ pyr5),保藏号CBS141149。

[0016] EP 2505651公开了一种已被突变或选择成具有低蛋白酶活性的分离的真菌,其中与未突变真菌相比,所述真菌具有低于50%的蛋白酶活性。所述真菌属于金孢霉属(*Chrysosporium*),优选地它是*Chrysosporium lucknowense*菌株。

[0017] 属于本发明的申请人的W0 2021/094935公开了遗传修饰的*Thermothelomyces heterothallica*,其中蛋白质糖基化途径已被工程化改造,以产生具有与人类相似的N-聚糖的蛋白质。W0 2021/094935公开了alg3和alg11基因的缺失或破坏、翻转酶的过表达以及异源GlcNAc转移酶1(GNT1)和GlcNAc转移酶2(GNT2)的表达。

[0018] 需要额外的可选的表达系统来生产重组人、伴侣动物和其他哺乳动物蛋白质,这些蛋白质能够产生高产率的具有哺乳动物蛋白质、特别是人和伴侣动物蛋白质的N-聚糖的糖蛋白,使得所述蛋白质适合于人类、伴侣动物和其他哺乳动物中的治疗性用途。

发明内容

[0019] 本发明提供了被遗传修饰以产生具有哺乳动物蛋白质的N-聚糖、特别是人类蛋白质的N-聚糖的糖蛋白的*Thermothelomyces heterothallica*。本发明的*Th.heterothallica*的遗传修饰包括alg3基因的缺失或破坏,靶向ER的甘露糖苷酶1(α -1,2-甘露糖苷酶)和靶向ER的葡萄糖苷酶2 α 亚基的异源表达或过表达。所述*Th.heterothallica*的遗传修饰还可以进一步包括异源GlcNAc转移酶1(GNT1)和GlcNAc转移酶2(GNT2)的表达。在一些实施方案中,所述遗传修饰进一步包括异源寡糖基转移酶(OST)的STT3亚基的表达。在其他实施方案中,所述遗传修饰进一步包括异源半乳糖基转移酶的表达。在一些实施方案中,所述遗传修饰进一步包括内源翻转酶的过表达或异源翻转酶的表达。

[0020] 本发明部分是基于下述发现,即如本文所公开的遗传修饰的*Th.heterothallica*产生其中所需的哺乳动物/人N-聚糖占糖蛋白上存在的N-聚糖的90%以上,在某些情况下甚至占N-聚糖的95%以上或98%以上的糖蛋白。此外,当进一步修饰以表达异源哺乳动物糖蛋白(例如抗体)时,如本文所公开的遗传修饰的*Th.heterothallica*产生高水平的异源糖蛋白,其中所需的哺乳动物/人N-聚糖占其N-聚糖的90%以上。这与迄今为止记载的表达系统形成对比,本申请在得到的N-聚糖方面产生很大的变动。引人注目的是,没有观察到任何所做的修饰对细胞存活率具有重大负面影响。

[0021] 进一步公开了靶向ER的甘露糖苷酶1(α -1,2-甘露糖苷酶)和靶向ER的葡萄糖苷酶2 α 亚基的表达不降低细胞存活率或异源糖蛋白的生产水平。现在已公开,出乎意料的是,在所生产的糖蛋白中实现了所需的哺乳动物/人N-聚糖,而不需要破坏内源 α -1,2-甘露糖基转移酶(ALG11)的表达。

[0022] 值得注意的是,与大多数真菌和酵母不同,*Th.heterothallica*不具有可能含有多达50个甘露糖残基的高甘露糖基化N-聚糖,而是具有含有3-9个甘露糖残基的“寡甘露糖”N-聚糖和结构尚未完全表征的含有甘露糖和HexNAc残基的杂合型N-聚糖。由于所述杂合N-聚糖的结构和合成途径尚未完全表征,因此尚不清楚是否可以使用本文描述的遗传修饰来消除此类聚糖。令人惊讶的是,根据本发明的遗传修饰足以导致这些结构的基本消除,其中90%以上的N-糖型是所需的哺乳动物/人N-聚糖,而不需要降低alg11的表达。

[0023] 有利情况下,本发明的*Th.heterothallica*细胞产生高产量的蛋白质。使用本发明的*Th.heterothallica*细胞获得的蛋白质水平远高于使用例如酵母获得的蛋白质水平。

[0024] 因此,本发明提供了一种用于生产具有所需N-聚糖、适合于人类中的治疗性用途的糖蛋白的方法。

[0025] 根据一个方面,本发明提供了一种*Thermothelomyces heterothallica*,其被遗传修饰以产生具有哺乳动物N-聚糖的糖蛋白,其中所述遗传修饰包括:

[0026] (i) *alg3*基因的缺失或破坏,使得所述*Th.heterothallica*不能产生功能性 α -1,3-甘露糖基转移酶;

[0027] (ii) 靶向ER的甘露糖苷酶1(α -1,2-甘露糖苷酶)的表达;和

[0028] (iii) 靶向ER的葡萄糖苷酶2 α 亚基的表达。

[0029] 在一些实施方案中,所述甘露糖苷酶1是里氏木霉(*Trichoderma reesei*)甘露糖苷酶1。在其他实施方案中,所述甘露糖苷酶1是*Th.heterothallica*甘露糖苷酶1。在一些实施方案中,所述葡萄糖苷酶2 α 亚基选自*Th.Heterothallica*、里氏木霉(*T.reesei*)和黑曲霉(*Aspergillus niger*)葡萄糖苷酶2 α 亚基。在其他实施方案中,所述遗传修饰进一步包括葡萄糖苷酶 β 亚基的表达。

[0030] 在一些实施方案中,所述靶向ER的里氏木霉(*Trichoderma reesei*)甘露糖苷酶1包含SEQ ID NO:2中阐述的氨基酸序列。在一些实施方案中,靶向ER的里氏木霉(*Trichoderma reesei*)甘露糖苷酶1由引入到所述*Th.heterothallica*中的包含SEQ ID NO:1中阐述的序列或其具有至少90%序列同一性的类似物或衍生物的外源多核苷酸编码。

[0031] 在一些实施方案中,所述靶向ER的*Th.heterothallica*葡萄糖苷酶2 α 亚基包含SEQ ID NO:4中阐述的氨基酸序列。在一些实施方案中,所述靶向ER的*Th.heterothallica*葡萄糖苷酶2 α 亚基由引入到所述*Th.heterothallica*中的包含SEQ ID NO:3中阐述的序列或其具有至少90%序列同一性的类似物或衍生物的外源多核苷酸编码。

[0032] 在一些实施方案中,所述靶向ER的甘露糖苷酶1(α -1,2-甘露糖苷酶)和/或靶向ER的葡萄糖苷酶2 α 亚基被整合到*Th.Heterothallica*基因组内的 α lp3蛋白酶基因座。在某些实施方案中,所述靶向ER的甘露糖苷酶1(α -1,2-甘露糖苷酶)和靶向ER的葡萄糖苷酶2 α 亚基两者均被整合到*Th.Heterothallica*基因组内的 α lp3蛋白酶基因座。

[0033] 在一些实施方案中,所述遗传修饰进一步包括异源GlcNAc转移酶1(GNT1)和GlcNAc转移酶2(GNT2)的表达。

[0034] 在一些实施方案中,根据本发明的异源GNT1和GNT2是动物来源的。

[0035] 在一些实施方案中,根据本发明的动物来源的GNT1包含异源高尔基体定位信号。

[0036] 在一些实施方案中,根据本发明的动物来源的GNT1是人GNT1。在一些实施方案中,根据本发明的动物来源的GNT1是包含异源高尔基体定位信号的人GNT1。

[0037] 在一些实施方案中,根据本发明的动物来源的GNT1是牛GNT1。在一些实施方案中,根据本发明的动物来源的GNT1使包含异源高尔基体定位信号的牛GNT1。

[0038] 在一些实施方案中,根据本发明的异源高尔基体定位信号是*Th.heterothallica*高尔基体定位信号。在一些实施方案中,所述*Th.heterothallica*高尔基体定位信号来自*Th.heterothallica*蛋白KRE2。

[0039] 在其他实施方案中,根据本发明的异源高尔基体定位信号是酵母高尔基体定位信号。在一些实施方案中,所述酵母高尔基体定位信号来自酵母蛋白KRE2。

[0040] 在一些实施方案中,根据本发明的动物来源的GNT2是大鼠GNT2。在其他实施方案

中,根据本发明的动物来源的GNT2是人GNT2。

[0041] 在一些实施方案中,所述动物来源的GNT1是人GNT1,并且所述动物来源的GNT2是大鼠GNT2。在一些实施方案中,所述人GNT1包含Th.heterothallica高尔基体定位信号。在一些实施方案中,所述Th.heterothallica高尔基体定位信号来自C1 Th.heterothallica蛋白KRE2。在其他实施方案中,所述人GNT1包含酵母高尔基体定位信号。

[0042] 在一些实施方案中,所述动物来源的GNT1是人GNT1,并且所述动物来源的GNT2是大鼠GNT2。在一些实施方案中,所述人GNT1包含Th.heterothallica高尔基体定位信号。在一些实施方案中,所述Th.heterothallica高尔基体定位信号来自C1蛋白KRE2。在其他实施方案中,所述人GNT1包含酵母高尔基体定位信号。

[0043] 在一些实施方案中,根据本发明的Th.heterothallica被遗传修饰以过表达内源Th.heterothallica RFT1翻转酶。

[0044] 在其他实施方案中,根据本发明的Th.heterothallica被遗传修饰以表达异源翻转酶,其中所述异源翻转酶是酵母FLC2p翻转酶。

[0045] 在一些实施方案中,根据本发明的遗传修饰进一步包括异源寡糖基转移酶的STT3亚基(异源STT3)的表达。在一些实施方案中,根据本发明的异源STT3是大型利什曼原虫(*Leishmania major*)STT3。

[0046] 在一些实施方案中,根据本发明的遗传修饰进一步包括异源半乳糖基转移酶的表达。在一些实施方案中,所述异源半乳糖基转移酶是动物来源的半乳糖基转移酶。

[0047] 在一些实施方案中,所述动物来源的半乳糖基转移酶是人半乳糖基转移酶。在一些实施方案中,根据本发明的动物来源的半乳糖基转移酶是包含异源高尔基体定位信号、例如包含Th.heterothallica KRE2高尔基体定位信号的人半乳糖基转移酶。

[0048] 在其他实施方案中,所述动物来源的半乳糖基转移酶是热带爪蟾(*Xenopus tropicalis*)半乳糖基转移酶。在一些实施方案中,根据本发明的动物来源的半乳糖基转移酶是包含异源高尔基体定位信号、例如包含酿酒酵母(*S.cerevisiae*)KRE2高尔基体定位信号的热带爪蟾(*Xenopus tropicalis*)半乳糖基转移酶。

[0049] 在一些实施方案中,所述Th.heterothallica是Th.heterothallica C1。

[0050] 在一些实施方案中,所述C1是被修饰以缺失一个或多个编码内源蛋白酶的基因的菌株。

[0051] 在一些实施方案中,所述C1是被修饰以缺失编码内源几丁质酶的基因的菌株。

[0052] 在一些实施方案中,所述C1是选自下述的菌株:野生型C1保藏号VKM F-3500 D, UV13-6保藏号VKM F-3632 D, NG7C-19保藏号VKM F-3633 D, UV18-25, 保藏号VKM F-3631 D, W1L#100I (prt- Δ alp1 Δ chi1 Δ alp2 Δ pyr5) 保藏号CBS141153, UV18-100f (prt- Δ alp1, Δ pyr5) 保藏号CBS141147, W1L#100I (prt- Δ alp1 Δ chi1 Δ pyr5) 保藏号CBS141149, 和 UV18-100f (prt- Δ alp1 Δ pep4 Δ alp2 Δ prt1 Δ pyr5) 保藏号CBS141143。每种可能性代表本发明的独立实施方案。根据某些实施方案,所述C1菌株具有至少8种、至少9种、至少10种、至少11种、至少12种、至少13种、至少14种或更多种蛋白酶的降低的表达和/或活性。根据某些示例性实施方案,所述C1菌株具有ALP1、ALP2、PEP4、PRT1、SRP1、ALP3、PEP1和MTP2的降低的表达或活性(Δ alp Δ alp2 Δ pep4 Δ prt1 Δ srp1 Δ alp3 Δ pep1 Δ mtp2)。

[0053] 根据一些实施方案,所述Th.heterothallica能够产生异源哺乳动物糖蛋白,其哺

乳动物/人N-聚糖占所述糖蛋白上存在的N-聚糖的90%、95%或98%以上。

[0054] 在一些实施方案中,所述*Th.heterothallica*被进一步遗传修饰以表达异源哺乳动物糖蛋白。在一些实施方案中,所述异源哺乳动物糖蛋白是抗体或其抗原结合片段。

[0055] 根据一些实施方案,所述异源哺乳动物糖蛋白包含的哺乳动物/人N-聚糖占其总N-聚糖的90%以上。根据一些实施方案,所述异源哺乳动物糖蛋白包含的哺乳动物/人N-聚糖占其总N-聚糖的92%、94%、96%、98%或99%以上。

[0056] 根据一些实施方案,所述异源哺乳动物糖蛋白包含的哺乳动物/人N-聚糖的量为由哺乳动物/人细胞产生的相同糖蛋白中存在的N-聚糖的量的至少80%、85%、90%、95%或更多。

[0057] 根据另一方面,本发明提供了一种用于产生生产具有哺乳动物N-聚糖的糖蛋白的*Th.heterothallica*的方法,所述方法包括:

[0058] (a) 缺失或破坏所述*Th.heterothallica*的 $\alpha 1,3$ 基因,使得所述*Th.heterothallica*不能产生功能性 $\alpha 1,3$ -甘露糖基转移酶;和

[0059] (b) 向所述*Th.heterothallica*中引入编码靶向ER的甘露糖苷酶1($\alpha 1,2$ -甘露糖苷酶)的外源多核苷酸;和

[0060] (c) 向所述*Th.heterothallica*中引入编码靶向ER的葡萄糖苷酶2 α 亚基的外源多核苷酸。

[0061] 根据一些实施方案,所述方法进一步包括向所述*Th.Heterothallica*中引入下述外源多核苷酸的步骤:编码内源翻转酶的外源多核苷酸,以诱导所述内源翻转酶在所述*Th.Heterothallica*中的过表达;或编码异源翻转酶的外源多核苷酸,以诱导所述异源翻转酶在所述*Th.Heterothallica*中的表达。

[0062] 根据一些实施方案,所述靶向ER的甘露糖苷酶1($\alpha 1,2$ -甘露糖苷酶)和靶向ER的葡萄糖苷酶2 α 亚基使用编码两种酶的单个外源多核苷酸引入。

[0063] 在一些实施方案中,所述甘露糖苷酶1是里氏木霉(*Trichoderma reesei*)甘露糖苷酶1。在一些实施方案中,所述葡萄糖苷酶2 α 亚基是*Th.heterothallica*葡萄糖苷酶2 α 亚基。

[0064] 根据另一方面,本发明提供了一种用于生产具有哺乳动物N-聚糖的糖蛋白的方法,所述方法包括:

[0065] (a) 提供根据本发明的遗传修饰的*Th.heterothallica*;

[0066] (b) 将所述*Th.heterothallica*在适合于表达所述糖蛋白的条件下培养;和

[0067] (c) 回收所述糖蛋白。

[0068] 在一些实施方案中,所述糖蛋白是在所述(*Th.heterothallica*)中重组表达的异源哺乳动物糖蛋白。在一些特定实施方案中,所述糖蛋白是在所述*Th.heterothallica*中重组表达的人蛋白。在其他实施方案中,所述糖蛋白是在所述*Th.heterothallica*中重组表达的伴侣动物蛋白质。在一些实施方案中,所述异源哺乳动物糖蛋白是抗体或其抗原结合片段。

[0069] 根据一些实施方案,所述异源哺乳动物糖蛋白包含的哺乳动物/人N-聚糖占其总N-聚糖的90%、92%、94%、96%、98%或99%。根据其他实施方案,所述异源哺乳动物糖蛋白包含的哺乳动物/人N-聚糖的量为由哺乳动物/人细胞产生的相同糖蛋白中存在的N-聚

糖的量的至少80%、85%、90%、95%或更多。

[0070] 根据另一方面,本发明提供了由根据本发明的遗传修饰的Th.heterothallica产生的重组糖蛋白,其中所述糖蛋白包含GlcNAc₂Man₃GlcNAc₂(G0)聚糖。

[0071] 根据另一方面,本发明提供了由根据本发明的遗传修饰的Th.heterothallica产生的重组糖蛋白,其中所述糖蛋白包含Gal₁GlcNAc₂Man₃GlcNAc₂(G1)聚糖、Gal₂GlcNAc₂Man₃GlcNAc₂(G2)聚糖或其组合。

[0072] 在一些实施方案中,由根据本发明的遗传修饰的Th.heterothallica产生的重组糖蛋白是医药级糖蛋白。

[0073] 根据一些实施方案,所述异源哺乳动物糖蛋白包含的哺乳动物/人N-聚糖占其总N-聚糖的90%、92%、94%、96%、98%或99%。根据其他实施方案,所述异源哺乳动物糖蛋白包含的哺乳动物/人N-聚糖的量为由哺乳动物/人细胞产生的相同糖蛋白中存在的N-聚糖的量的至少80%、85%、90%、95%或更多。

[0074] 应当理解,本文公开的每个方面和实施方案的任何组合都明确包含在本发明的公开范围内。

[0075] 从下面的详细描述、实施例和权利要求,本发明的这些和其他方面和特征将变得显而易见。

附图说明

[0076] 图1A-ID.来自糖修饰菌株的蛋白A亲和纯化的纳武单抗(Nivolumab)的经释放的N-聚糖的分析。作为实例示出了来自每个转化的一个色谱图。文本中描述的所有菌株的主要N-聚糖的结果显示在色谱图下方的相应表格中。图1A-在摇瓶中培养的菌株M4855和M4856(表达里氏木霉(T.reesei)Mns1-HDEL和C1 gls2a-HDEL的基于M3291的菌株)。示出了M4855的色谱图以及M4855和M4856两者主要释放的N-聚糖的表格。图1B-1D-在摇瓶中培养的G1/2菌株M5129至M5132(表达G1/2机制的基于M4855的菌株)。示出了M5130(图1B)和M5132(图1C)的色谱图以及M5129至M5132主要释放的N-聚糖的表格(图1D)。

[0077] 图2.由在摇瓶中培养的菌株M5129至M5132(表达G1/2机制的基于M4855的菌株)产生的纳武单抗的释放的N-聚糖的量的比较。使用添加到每个用于分析的样品中的固定量的内标,在样品之间对释放的N-聚糖的量进行归一化。将参比蛋白Opdivo的响应%设置为100%。

[0078] 图3A-3D.来自三种糖修饰菌株的发酵样品蛋白A亲和纯化的纳武单抗的释放的N-聚糖的分析。本文中描述的所有菌株的主要N-聚糖的结果显示在色谱图下方的相应表格中。图3A-仅缺失alg3的起始菌株M3291。示出了M3291的色谱图和在各种发酵条件下主要释放的N-聚糖的表格。图3B-3D-G1/2菌株M5130和M5132。示出了M5130(图3B)和M5132(图3C)的色谱图以及在多种发酵条件下主要释放的N-聚糖的表格(图3D)。

[0079] 图4.由在多种发酵条件下培养的菌株M5130和M5132产生的纳武单抗的释放的N-聚糖的量的比较。使用添加到每个用于分析的样品中的固定量的内标,在样品之间对释放的N-聚糖的量进行归一化。将参比蛋白Opdivo的响应%设置为100%。

[0080] 图5A-5C.来自糖修饰菌株的总分泌蛋白的释放的N-聚糖的分析。作为实例示出了来自每个转化的一个色谱图。文本中描述的所有菌株的主要N-聚糖的结果显示在表格(图

5C)和色谱图(图5A-5B)中。M6589和M6596是具有G1/2机制的菌株,其中TrMns1-HDEL在泛素样蛋白启动子下,而M6590和M6597是具有G1/2机制的菌株,其中TrMns1-HDEL在bg18启动子下。

具体实施方式

[0081] 本发明涉及对真菌*Thermothelomyces heterothallica*、特别是菌株C1的遗传修饰,以产生具有哺乳动物蛋白质的N-聚糖、特别是人、伴侣动物和其他哺乳动物蛋白质的N-聚糖的糖蛋白。

[0082] 由如本文所述遗传修饰的*Th.heterothallica*产生的糖蛋白,适合于人类、伴侣动物(例如狗、猫和马)和其他哺乳动物中的治疗性用途。

[0083] 蛋白质糖基化,即将寡糖共价附连到细胞中新合成的多肽链的侧链,是真核细胞中的一个有序过程,涉及一系列依次添加和去除糖部分的酶。N-糖基化是一种将寡糖附连到天冬酰胺残基的侧链的过程,特别是出现在序列Asn-Xaa-Ser/Thr中的天冬酰胺,其中Xaa表示除Pro之外的任何氨基酸。

[0084] N-糖基化在内质网(ER)中启动,其中将寡糖Glc₃Man₉GlcNAc₂组装在脂质载体焦磷酸多萜醇上,随后转移到已进入ER内腔的多肽的选定天冬酰胺残基上。脂质连接的寡糖的生物合成需要几种特定糖基转移酶(例如ALG1、ALG2和ALG3)的活性。它起始于ER膜的胞质侧,终止于内腔中,在此处寡糖基转移酶(OST)选择初生多肽的N-X-S/T序列子,并在天冬酰胺的侧链酰胺与寡糖之间产生N-糖苷键。脂质连接的寡糖从ER外部向内部的翻转由位于ER膜上的翻转酶进行。在转移到初生多肽后,所述寡糖通常被葡萄糖苷酶和甘露糖苷酶修剪,然后所述初生糖蛋白被转移到高尔基体进行进一步加工。

[0085] 在所有已知的真核生物中,焦磷酸多萜醇结合的寡糖的合成基本上是类似的。然而,随着糖蛋白沿着分泌途径移动,所述寡糖的进一步加工在低等真核生物例如真菌或酵母与高等真核生物例如动物和植物之间差异很大。因此,在各种生物体之间糖侧链的最终组成是不同的,并且取决于宿主。

[0086] 在微生物例如酵母中,通常添加额外的甘露糖和/或甘露糖磷酸酯糖,产生可能含有多达30-50个甘露糖残基的“高甘露糖”型N-聚糖。

[0087] 在包括人、伴侣动物和其他哺乳动物细胞在内的动物细胞中,初生糖蛋白被转移到高尔基体,在那里甘露糖残基被高尔基体特异性1,2-甘露糖苷酶去除。随着蛋白质通过高尔基体,由大量修饰酶继续进行加工,修饰酶包括N-乙酰氨基葡萄糖转移酶(GnT I、GnT II、GnT III、GnT IV、GnT V、GnT VI)、甘露糖苷酶II和岩藻糖基转移酶,它们添加和去除特定糖残基。最后,N-聚糖被半乳糖基转移酶(GalT)和唾液酸基转移酶(ST)作用,并且最终的糖蛋白从高尔基体中释放。动物糖蛋白的N-聚糖具有双、三或四触角结构,并且通常可能包括半乳糖、岩藻糖和N-乙酰氨基葡萄糖。通常,N-聚糖的末端残基由唾液酸组成。

[0088] 与大多数真菌和酵母不同,*Th.heterothallica*不具有高甘露糖基化的N-聚糖,而是具有“寡甘露糖”聚糖(Man₃至Man₈₋₉)和含有Man和HexNAc残基的杂合型聚糖(Man₃HexNAc-Man₈HexNAc)。这些杂合聚糖的确切结构尚不完全清楚。所述杂合聚糖具有典型的甘露糖残基,但另外还具有通过尚未表征的键附连的未知HexNAc。

[0089] 由于所述杂合聚糖的结构和合成途径尚未完全表征,因此尚不清楚此类聚糖是否

可以使用本文描述的遗传修饰来消除。令人惊讶的是,根据本发明的遗传修饰引起这些结构的基本消除,其中90%以上、通常98%以上的N-糖型是所需的哺乳动物/人聚糖。

[0090] 本发明旨在对*Th.heterothallica*中的N-糖基化途径进行遗传修饰,使得它产生高百分率的具有哺乳动物N-聚糖、特别是人N-聚糖例如GlcNAc₂Man₃GlcNAc₂(“G0”)、GlcNAc₂Man₃GlcNAc₂(Fuc)(“FG0”)、Gal₁₋₂GlcNAc₂Man₃GlcNAc₂(“G1”/“G2”)和Gal₁₋₂GlcNAc₂Man₃GlcNAc₂(Fuc)(“FG1”/“FG2”)的糖蛋白。

[0091] 具体而言,在一些实施方案中,所述*Th.heterothallica*中N-糖基化途径的遗传修饰包括下述修饰:

[0092] 1.C1 *alg3*基因(编码 α -1,3-甘露糖基转移酶)的缺失;

[0093] 2.靶向ER的里氏木霉(*Trichoderma reesei*)甘露糖苷酶1(α -1,2-甘露糖苷酶)的表达;

[0094] 3.靶向ER的C1葡萄糖苷酶2 α 亚基的表达;

[0095] 4.异源GlcNAc转移酶1(GNT1)的表达;

[0096] 5.异源GlcNAc转移酶2(GNT2)的表达;和

[0097] 6.异源半乳糖基转移酶1(GalT1)的表达。

[0098] *alg3*的缺失将N-聚糖前体的合成终止在具有1或2个末端葡萄糖的Man₅GlcNAc₂处。这种聚糖作为被引入到*Th.heterothallica*中的GNT1和GNT2的底物。额外的遗传修饰可能包括引入来自人、伴侣动物和其他哺乳动物糖基化途径的额外的酶,例如半乳糖基转移酶和/或岩藻糖基转移酶。

[0099] 上述异源酶与靶向肽一起表达,使得表达的酶被靶向特定的细胞区室。

[0100] 当在本文中使用时,在提及酶时,它涵盖其酶活性片段及其酶活性变体。

[0101] 本发明具体而言涉及*Th.heterothallica*的N-糖基化途径的工程化改造。值得注意的是,0-聚糖可能存在,也可能通过所述*Th.heterothallica*的进一步遗传修饰而去除或改变。

[0102] 应当理解,根据本发明的遗传修饰使得所述遗传修饰的*Th.heterothallica*能够以适合其预期用途的足够速率生长。

[0103] 本文所使用的“C1”或“*Thermothelomyces heterothallica* C1”或“*Th.heterothallica* C1”均指*Thermothelomyces heterothallica*菌株C1。*Thermothelomyces*属及其种的描述可以在例如Marin-Felix Y(2015.*Mycologica* 107(3): 619-632)和van den Brink J等(2012,*Fungal Diversity* 52(1):197-207)中找到。

[0104] 值得注意的是,上述作者(Marin-Felix等,2015)建议根据最适生长温度、分生孢子形态和有性繁殖周期细节的差异将毁丝霉(*Myceliophthora*)属分拆。根据所提出的标准,C1明确属于新建立的*Thermothelomyces*属(其包含以前的耐热的毁丝霉物种),而不是毁丝霉(*Myceliophthora*)属(其仍然包括非耐热的物种)。由于C1可以与具有相反交配型的一些其他*Thermothelomyces*(以前称为毁丝霉属(*Myceliophthora*))菌株形成子囊孢子,因此C1最好被分类为*Th.heterothallica*菌株C1而不是*Th.thermophila*C1。

[0105] 还必须认识到,真菌分类学在过去也在不断变化,因此上文列出的当前名称可能以前具有除嗜热毁丝霉(*Myceliophthora thermophila*)(van Oorschot,1977.*Persoonia* 9(3):403)之外的各种旧名称,它们现在被认为是同义语。例如,*Thermothelomyces*

heterothallica (Marin-Felix等, 2015. *Mycologica*, 3: 619-63) 与 *Corynascus heterothallica* (von Arx等, 1983)、*Thielavia heterothallica* (von Klopotek, 1976. *Archives of Microbiology* 107 (2), 223-224)、*Chrysosporium lucknowense* 和 *thermophile* (von Klopotek, 1974. *Archives of Microbiology* 98 (1), 365-369) 以及 *Sporotrichium thermophile* (Alpinis 1963. *Nova Hedwigia* 5: 74) 同义。

[0106] 应该进一步明确理解的是, 本发明涵盖含有与 SEQ ID NO: 22 显示出 99% 或更高同源性的核糖体 DNA (rDNA) 序列的任何菌株, 并且所有这些菌株都被认为是与 *Thermothelomyces heterothallica* 同种的。

[0107] SEQ ID NO: 22 与在 *Th. heterothallica/thermophila* (被列为嗜热毁丝霉 (*Myceliophthora thermophila*)) ATCC 42464 rDNA 序列 (ncbi.nlm.nih.gov/nucleotide/CP003008.1) 的第 7 号染色体上发现的 rDNA 序列具有 99.98% 的同一性。

[0108] *Th. heterothallica* 菌株 C1 (作为 *Chrysosporium lucknowense* 菌株 C1) 根据布达佩斯公约以编号 VKM F-3500 D 保藏, 保藏日期为 1996 年 8 月 29 日。

[0109] 上述术语还涵盖了源自于野生型菌株的遗传修饰的亚株, 其已使用随机或定点方法例如使用 UV 诱变或通过缺失一个或多个内源基因进行突变。例如, 所述 C1 菌株可以是指被修饰以缺失一个或多个编码内源蛋白酶的基因和/或一个或多个编码内源几丁质酶的基因的野生型菌株。例如, 本发明涵盖的 C1 菌株 (亚株) 包括: UV18-25, 保藏号 VKM F-3631 D; 菌株 NG7C-19, 保藏号 VKM F-3633 D; 和菌株 UV13-6, 保藏号 VKM F-3632 D。根据本发明的教导可以使用的其他 C1 菌株包括: HC 菌株 UV18-100f, 保藏号 CBS141147; HC 菌株 UV18-100f, 保藏号 CBS141143; LC 菌株 W1L#100I, 保藏号 CBS141153; 和 LC 菌株 W1L#100I, 保藏号 CBS141149。

[0110] 当在适合条件下生长时, 与酵母菌株相比, 总的来说 *Th. heterothallica* 真菌 (具体而言菌株 C1) 显示出更高的生物质产量。*Th. heterothallica* 真菌可以在大体积三维 (3D) 液体培养物中以及固体培养基上生长。与常规酵母和其他真菌相比, 由本发明的申请人开发的几种菌株对在真菌生长培养基中作为碳源存在的葡萄糖和其他可发酵糖的反馈阻遏不太敏感, 并且可以耐受碳源的高进料速率, 从而获得高产率。此外, 与用非葡萄糖阻遏的野生型 *Th. heterothallica* 真菌或已知用于蛋白质生产的其他丝状真菌获得的高黏度相比, 这些菌株中的一些菌株当在商业发酵罐中生长时提供了显著降低的培养基黏度。所述低黏度可归因于菌株的形态变化, 从亲本菌株中具有长而高度交织的菌丝变为所开发的菌株中具有短而较少交织的菌丝。低培养基黏度在发酵罐中的大规模工业生产中是非常有利的。例如, 对葡萄糖阻遏显示出降低的敏感性的 *Th. heterothallica* C1 菌株 UV18-25 (保藏号 VKM F-3631D) 已在工业上生长, 以在超过 100,000 升的体积中生产重组酶。

[0111] 在一些实施方案中, 本发明的 C1 菌株是被修饰以缺失多个 (即至少两个) 编码内源蛋白酶的基因的菌株。在一些实施方案中, 所述 C1 菌株是被修饰以缺失至少四个编码内源蛋白酶的基因的菌株。在其他实施方案中, 所述 C1 菌株是被修饰以缺失至少五个编码内源蛋白酶的基因的菌株。在一些特定实施方案中, 所述 C1 菌株是被修饰以缺失至少六个编码内源蛋白酶的基因的菌株。在其他特定实施方案中, 所述 C1 菌株是被修饰以缺失至少八个编码内源蛋白酶的基因的菌株。在其他特定实施方案中, 所述 C1 菌株是被修饰以缺失至少 8、9、10、11、12、13、14 个或更多个编码内源蛋白酶的基因的菌株。在某些示例性实施方案

中,所述C1菌株是被修饰以缺失至少13或14个编码内源蛋白酶的基因的菌株。

[0112] 应当明确地理解,本发明的教导涵盖了*Th.heterothallica* C1菌株的突变体、衍生菌株、后代、克隆体和类似物,只要这些衍生菌株、后代、克隆体和类似物在根据本发明的教导的遗传修饰时能够生长并产生具有本文所描述的N-聚糖的蛋白质即可。

[0113] 应当明确地理解,指称真菌株系的术语“衍生菌株”涵盖任何具有修饰的真菌亲本株系,所述修饰正面影响产物产量、效率或功效,或影响任何改善所述真菌衍生菌株作为工具以生产本文所描述的具有哺乳动物蛋白质,特别是人、伴侣动物和其他哺乳动物蛋白质的N-聚糖的异源蛋白质的性状。本文所使用的术语“后代”是指来自亲本真菌株系的未修饰的后代,例如来自细胞的细胞。

[0114] 本文所使用的“聚糖”是指可以连接到载体例如氨基酸、肽、多肽、脂质或还原末端偶联物的寡糖链。本发明特别涉及通过与天冬酰胺残基(Asn)的侧链酰胺氮的N-连接键而与多肽N-糖基化位点例如-Asn-Xxx-Ser/Thr-偶联的N-连接聚糖(“N-聚糖”),其中Xxx是除Pro之外的任何氨基酸残基。本发明还可以涉及作为多萜醇-磷酸-寡糖(Dol-P-P-OS)前体脂质结构的一部分的聚糖,所述结构是真核细胞的内质网中N-连接聚糖的前体。所述前体寡糖通过其还原末端结合到多萜醇脂质上的两个磷酸残基。

[0115] 在哺乳动物糖蛋白中发现的通常构成N-聚糖的单糖包括但不限于N-乙酰葡萄糖胺(缩写为“GlcNAc”)、甘露糖(缩写为“Man”)、葡萄糖(缩写为“Glc”)、半乳糖(缩写为“Gal”)、唾液酸(缩写为“Neu5Ac”)和岩藻糖(简称为“Fuc”)。

[0116] N-聚糖共享被称为“核心”结构的共用的五糖 $\text{Man}_3\text{GlcNAc}_2$ (缩写为“ Man_3 ”)。本发明的重要靶聚糖结构包括在核心结构的末端1,3甘露糖臂上具有一个GlcNAc残基并在核心结构的末端1,6甘露糖臂上具有一个GlcNAc残基的N-聚糖。此类N-聚糖包括: $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ (被称为“G0”糖型), $\text{Gal}_{1-2}\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ (根据半乳糖残基的数量被称为“G1”或“G2”糖型),以及它们的核心岩藻糖基化糖型: $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2(\text{Fuc})$ (“G0F”或“FG0”)和 $\text{Gal}_{1-2}\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2(\text{Fuc})$ (“G1F”和“G2F”,或“FG1”和“FG2”)。

[0117] 术语“alg3基因”是指编码 α -1,3-甘露糖基转移酶的基因。术语“ α -1,3-甘露糖基转移酶”是指多萜醇-P-Man: $\text{Man}_5\text{GlcNAc}_2$ -PP-多萜醇 α -1,3-甘露糖基转移酶(EC 2.4.1.258),它是一种ER驻留酶,催化下述反应:

[0118] 多萜醇 β -D-甘露糖基磷酸+D-Man- α -(1→2)-D-Man- α -(1→2)-D-Man- α -(1→3)-[D-Man- α -(1→6)]-D-Man- β -(1→4)-D-GlcNAc- β -(1→4)-D-GlcNAc-二磷酸多萜醇

[0119] →

[0120] D-Man- α -(1→2)-D-Man- α -(1→2)-D-Man- α -(1→3)-[D-Man- α -(1→3)-D-Man- α -(1→6)]-D-Man- β -(1→4)-D-GlcNAc- β -(1→4)-D-GlcNAc-二磷酸多萜醇+多萜醇磷酸

[0121] 在一些特定实施方案中,“alg3基因”是编码C1的 α -1,3-甘露糖基转移酶的基因(JGI *M.thermophila*基因组的直系同源物(mycocosm.JGI.doe.gov),登录号2310419)。

[0122] 本发明的*Th.heterothallica*通过alg3基因的缺失或破坏进行遗传修饰,使得所述*Th.heterothallica*不能产生功能性 α -1,3-甘露糖基转移酶。本发明的*Th.heterothallica*不表现出可检测的 α -1,3-甘露糖基转移酶活性。

[0123] 术语“甘露糖苷酶1”(α-1,2-甘露糖苷酶)缩写为“MDS1”或“MNS1”,其催化下述反

应:

[0124] $3xH_2O+N_4-(\alpha-D-Man-(1\rightarrow2)-\alpha-D-Man-(1\rightarrow2)-\alpha-D-Man-(1\rightarrow3)-[\alpha-D-Man-(1\rightarrow3)-[\alpha-D-Man-(1\rightarrow2)-\alpha-D-Man-(1\rightarrow6)]-\alpha-D-Man-(1\rightarrow6)]-\beta-D-Man-(1\rightarrow4)-\beta-D-GlcNAc-(1\rightarrow4)-\alpha-D-GlcNAc)-L-天冬酰胺基-[蛋白质](N-葡聚糖甘露糖异构体8A1,2,3B1,3)\rightarrow3\beta-D-甘露糖+N_4-(\alpha-D-Man-(1\rightarrow3)-[\alpha-D-Man-(1\rightarrow3)-[\alpha-D-Man-(1\rightarrow6)]-\alpha-D-Man-(1\rightarrow6)]-\beta-D-Man-(1\rightarrow4)-\beta-D-GlcNAc-(1\rightarrow4)-\beta-D-GlcNAc)-L-天冬酰胺基-[蛋白质](N-葡聚糖甘露糖异构体5A1,2)。$

[0125] $\alpha-1,2$ -甘露糖苷酶的示例性登录号是Uniprot Q9P8T8。

[0126] 术语“葡萄糖苷酶2 α 亚基”(GLS2- α 或GLS2 α)是指一种从未成熟糖蛋白的Glc2Man9GlcNAc2寡糖前体上顺序切割2个最靠内的 $\alpha-1,3$ -连接的葡萄糖残基的酶。

[0127] 术语“翻转酶”(EC 7.6.2.1)是指一种将脂质连接的聚糖前体在其在ER中的合成期间从ER的胞质侧转移到管腔侧的酶。

[0128] 术语“GlcNAc转移酶1”缩写为“GNT1”(也被称为“GnTI”),是指 $\alpha-1,3$ -甘露糖基-糖蛋白2- β -N-乙酰氨基葡萄糖转移酶(EC 2.4.1.101),这是一种高尔基体驻留酶,将GlcNAc残基从UDP-GlcNAc转移到受体底物Man₅GlcNAc₂,以产生GlcNAcMan₅GlcNAc₂。在本发明中,鉴于alg3的缺失和 $\alpha-1,2$ -甘露糖苷酶和葡萄糖苷酶2 α 亚基的表达,N-聚糖前体的合成产生Man3GlcNAc2,因此聚糖Man3GlcNAc2充当GNT1的底物,产生GlcNAcMan3GlcNAc2。

[0129] 术语“GlcNAc转移酶2”缩写为“GNT2”(也被称为“GnTII”),是指 $\alpha-1,6$ -甘露糖基-糖蛋白2- β -N-乙酰氨基葡萄糖转移酶(EC 2.4.1.143),这是一种高尔基体驻留酶,将GlcNAc残基从UDP-GlcNAc转移到GlcNAcMan₃GlcNAc₂中的游离末端甘露糖残基,以产生GlcNAc₂Man₃GlcNAc₂。

[0130] 术语“寡糖基转移酶的STT3亚基”、“STT3蛋白”或简称“STT3”在本文中可互换使用,是指多萜醇-二磷酸寡糖-蛋白质糖基转移酶亚基(EC:2.4.99.18)。它是寡糖基转移酶(OST)复合体的催化亚基,催化所限定的聚糖(真核生物中的Glc₃Man₉GlcNAc₂)从脂质载体焦磷酸多萜醇向初生多肽链中Asn-X-Ser/Thr共有基序内的天冬酰胺残基的初始转移,这是蛋白质N-糖基化的第一步。STT3蛋白催化下述反应:

[0131] 多萜醇二磷酸寡糖+L-天冬酰胺基-[蛋白质]

[0132] $\rightarrow$

[0133] 多萜醇二磷酸+H⁺+N⁴-(寡糖-(1 $\rightarrow$ 4)-N-乙酰-p-D-氨基葡萄糖-(1 $\rightarrow$ 4)-N-乙酰- β -D-氨基葡萄糖)-L-天冬酰胺基-[蛋白质]

[0134] 术语“半乳糖基转移酶”(EC 2.4.1.38)是指一种高尔基体驻留酶,它将p-连接的半乳糖基残基转移到末端N-乙酰氨基葡萄糖。

[0135] 当提及基因、酶、蛋白质或肽序列例如亚细胞定位信号时,术语“异源的”在本文中用于描述在C1中不天然存在或表达的基因、酶、蛋白质或肽序列。当提及亚细胞定位信号时,所述术语还描述了与相应蛋白质中天然存在的不同的亚细胞定位信号。

[0136] 当提及基因、酶、蛋白质或肽序列例如亚细胞定位信号时,术语“内源的”在指在C1中天然存在的基因、酶、蛋白质或肽序列。

[0137] 当提及多核苷酸时,术语“外源的”在本文中用于描述通过转化外源性引入到C1中的合成多核苷酸。所述外源多核苷酸可以以稳定或瞬时的方式引入到C1中,从而产生核糖

核酸(RNA)分子并随后产生多肽分子。

[0138] 表达构建体和载体

[0139] 术语“表达构建体”、“DNA构建体”或“表达盒”在本文中可互换使用,并且是指人工组装或分离的核酸分子,其包含编码目标蛋白质的核酸序列,并被组装成使得所述目标蛋白质在靶宿主细胞中表达。表达构建体通常包含与编码所述目标蛋白质的核酸序列可操作连接的适合的调控序列。表达构建体还可以包含编码选择标记物的核酸序列。

[0140] 术语“核酸序列”、“核苷酸序列”和“多核苷酸”在本文中用于指称脱氧核糖核苷酸(DNA)、核糖核苷酸(RNA)及其修饰的形式的聚合物,其采取独立片段的形式或作为更大构建体的组分。核酸序列可以是编码序列,即在细胞中编码最终产物例如蛋白质的序列。核酸序列也可以是调控序列,例如启动子。

[0141] 术语“肽”、“多肽”和“蛋白质”在本文中用于指称氨基酸残基的聚合物。术语“肽”通常表示由2至50个氨基酸组成的氨基酸序列,而“蛋白质”表示由超过50个氨基酸残基组成的氨基酸序列。

[0142] 与参比序列“同源的”序列(例如核酸序列和氨基酸序列)在本文中是指所述序列之间的百分比同一性,其中所述百分比同一性为至少75%,优选地至少80%、至少85%、至少90%、至少95%、至少98%或至少99%。每种可能性代表本发明的独立实施方案。本文描述的序列的同源物被涵盖在本发明内。只要蛋白质同源物保持原始蛋白质的活性,它们就被涵盖在内。同源核酸序列包括与密码子用法和遗传密码的简并性相关的变化。序列同一性可以使用本领域中已知的核苷酸/氨基酸序列比较算法来确定。

[0143] 编码本发明的多肽的核酸序列可以被优化用于表达。此类序列修饰的实例包括但不限于改变G/C含量以更接近在*Th.heterothallica*中通常存在的含量,以及去除在这种真菌中非典型存在的密码子,这通常被称为密码子优化。

[0144] 短语“密码子优化”是指选择接近目标生物体内的密码子用法的适合的DNA核苷酸用于结构基因或其片段中,和/或是指通过将本源序列的至少一个密码子(例如约或超过约1、2、3、4、5、10、15、20、25、50个或更多个密码子)用在目标宿主细胞的基因中更频繁或最频繁使用的密码子代替并同时维持本源氨基酸序列,来修饰核酸序列以增强在所述宿主细胞中的表达的过程。各种不同物种对特定氨基酸的某些密码子表现出特定偏倚。密码子偏倚(生物体之间密码子用法的差异)通常与信使RNA(mRNA)的翻译效率相关,信使RNA(mRNA)的翻译效率进而据信取决于尤其是正在翻译的密码子的性质和特定转运RNA(tRNA)分子的可利用性。细胞中所选tRNA的主导性通常反映了蛋白质合成中最频繁使用的密码子。因此,可以在密码子优化的基础上对基因进行定制,以在给定生物体中获得最佳基因表达。因此,优化的基因或核酸序列是指其中本源或天然存在的基因的核苷酸序列已被修饰,以利用所述生物体内统计学优选的或统计学喜爱的密码子的基因。本发明明确涵盖了编码本文所公开的目标酶的多核苷酸,其被密码子优化以用于在*Th.heterothallica*中表达。

[0145] 术语“调控序列”是指控制编码序列的表达(转录)的DNA序列,例如启动子和终止子。

[0146] 术语“启动子”是指在体内或体外控制或指导另一个DNA序列的转录的调控DNA序列。通常,启动子位于被转录序列的5'区域中(即之前,位于上游)。启动子可以整体源自于天然来源,或者可以由源自于自然界中存在的不同启动子的不同元件组成,或甚至包含合

成的核苷酸区段。启动子可以是组成型(即启动子激活不受诱导剂调控,因此转录速率恒定)或诱导型(即启动子激活受诱导剂调控)的。在大多数情况下,调控序列的精确边界尚未被完全定义,并且在某些情况下不能被完全定义,因此某些变异的DNA序列可能具有相同的启动子活性。

[0147] 术语“终止子”是指调控转录终止的另一种调控DNA序列。终止子序列被可操作连接对待转录的核酸序列的3'端。

[0148] 术语“*Th.heterothallica*启动子”和“*Th.heterothallica*终止子”是指适合在*Th.heterothallica*中使用,即能够在*Th.heterothallica*中指导基因表达的启动子和终止子序列。在一些特定实施方案中,使用C1启动子和C1终止子,它们是指能够在C1中指导基因表达的启动子和终止子序列。

[0149] 根据一些实施方案,所述*Th.heterothallica*启动子/终止子源自*Th.heterothallica*的内源基因。根据其他实施方案,所述*Th.heterothallica*启动子/终止子源自进入*Th.heterothallica*的外源基因。

[0150] 适合的组成型启动子和终止子包括例如下述基因的启动子和终止子:C1糖酵解基因,例如磷酸甘油酸激酶基因(PGK)(Uniprot:G2QLD8,NCBI参考序列:XM_003665967)、甘油醛3-磷酸脱氢酶(GPD)(Uniprot:G2QPQ8,NCBI参考序列:XM_003666768)、磷酸果糖激酶(PFK)(Uniprot:G2Q605,NCBI参考序列:XM_003659879);或 β -葡萄糖苷酶1基因**bg11**(登录号:XM_003662656);或磷酸丙糖异构酶(TPI)(Uniprot:G2QBR0,NCBI参考序列:XM_003663200);或肌动蛋白(ACT)(Uniprot:G2Q7Q5,NCBI参考序列:XM_003662111);或C1 **cbh1**启动子(GenBank AX284115)或C1 **chi1**启动子(GenBank HI550986)。可以使用的其他启动子是黑曲霉(*Aspergillus nidulans*) **gpdA**启动子,以及在Rantasalo等(2018 NAR46(18):e111)中描述的合成启动子。可用于本发明的合成启动子被进一步记载在WO 2017/144777中。作为示例性终止子,可以使用C1几丁质酶1基因**chi1**(GenBank HI550986)、纤维二糖水解酶1 **cbh1**(GenBank AX284115)的终止子或酵母**adh1**终止子。

[0151] 示例性启动子和终止子序列以及启动子/终止子对被提供在下文的实施例部分中。在一些实施方案中,用于本发明的启动子序列选自如下文中所示例的启动子-8、**bg18**启动子、启动子-9、启动子-3、启动子-1和TEF1A启动子。每种可能性代表本发明的独立实施方案。

[0152] 术语“可操作连接”意味着所选核酸序列与调控元件(启动子或终止子)邻近,以允许所述调控元件调控所选核酸序列的表达。

[0153] 术语“定位信号”、“定位序列”、“亚细胞靶向肽/信号/序列”等在本文中可互换使用,并且是指蛋白质序列中包含的短肽序列(通常为5-30个氨基酸长)(通常存在于蛋白质的一个末端例如N-端处),所述序列将所述蛋白质引导到细胞内的特定亚细胞定位。例如,高尔基体定位信号将蛋白质靶向高尔基体。“异源定位信号”例如“异源高尔基体定位信号”表示不是所述蛋白质中天然存在的定位信号。在一些实施方案中,“异源的”是指来自另一种生物体的定位信号。

[0154] 在一些实施方案中,在根据本发明的*Th.heterothallica*中表达的蛋白质的定位信号源自*Th.heterothallica*的内源基因。例如,在一些实施方案中,使用来自C1蛋白**KRE2a**的高尔基体定位信号(JGI *M.thermophila*基因组(mycocosm.jgi.doe.gov)的直系同源物,

登录号2300989)。

[0155] 在其他实施方案中,在根据本发明的*Th.heterothallica*中表达的蛋白质的定位信号源自*Th.heterothallica*的外源基因(异源定位信号)。例如,在一些实施方案中,在根据本发明的*Th.heterothallica*中表达的动物来源的酶使用其自身天然存在的高尔基体定位信号表达。作为另一个实例,在一些实施方案中,使用来自酵母蛋白质例如酿酒酵母(*S.cerevisiae*)蛋白KRE2(GenBank登录号CAA44516)的高尔基体定位信号。

[0156] 根据一些实施方案,所述在*Th.heterothallica*中表达的蛋白质包含ER靶向序列。在某些实施方案中,所述ER靶向序列是序列HDEL。

[0157] 当在本文中使用时,当提及一个或多个核酸序列时,术语“框内”表示这些序列被连接成使得它们的正确阅读框得以保留。

[0158] 根据本发明的一些实施方案的表达构建体包含可操作地连接到编码酶(例如翻转酶、GNT1或GNT2)的核酸序列的*Th.heterothallica*启动子序列和*Th.heterothallica*终止子序列。在一些特定实施方案中,本发明的表达构建体包含可操作地连接到编码酶(例如翻转酶、GNT1或GNT2)的核酸序列的C1启动子序列和C1终止子序列。

[0159] 特定的表达构建体可以通过各种不同的方法组装,包括常规的分子生物学方法,例如聚合酶链反应(PCR)、限制性内切酶消化、体外和体内组装方法以及基因合成方法或其组合。示例性的表达构建体及其构建方法在下文的实施例部分中提供。

[0160] alg3的缺失

[0161] 基因缺失技术能够部分或完全去除基因,从而消除其表达。在此类方法中,基因的缺失可以通过使用质粒进行同源重组来实现,所述质粒已被构建成连续地含有所述基因两侧的5'和3'区。

[0162] 基因缺失也可以通过在所述基因中插入破坏性核酸构建体(在本文中也称为缺失构建体)来进行。破坏性构建体可能只是带有与所述基因同源的5'和3'区的选择标记物基因。所述选择标记物能够鉴定含有被破坏基因的转化体。可选地或此外,所述破坏性核酸构建体可以包含一个或多个编码将要在宿主细胞中表达的异源蛋白质的多核苷酸。

[0163] 用于alg3的示例性缺失构建体和执行所述缺失的程序被记载在例如WO 2021/094935中。可以利用在所述破坏性构建体两侧的适合引物,使用PCR来确认所述缺失。

[0164] 在一些实施方案中,本发明的*Th.heterothallica*被遗传修饰以表达异源翻转酶。在一些特定实施方案中,所述异源翻转酶是酵母翻转酶。在其他特定实施方案中,所述酵母翻转酶是酿酒酵母(*S.cerevisiae*)突变翻转酶FLC2p,其是酿酒酵母(*S.cerevisiae*)ER定位翻转酶FLC2的C-端截短的版本。

[0165] 在其他实施方案中,本发明的*Th.heterothallica*被遗传修饰以过表达内源*Th.heterothallica* RFT1翻转酶。在一些特定实施方案中,*Th.heterothallica* C1被遗传修饰以过表达内源C1 RFT1翻转酶。在根据本发明的*Th.heterothallica*中RFT1的过表达可以通过引入编码RFT1的外源多核苷酸来进行,所述外源多核苷酸包含在*Th.heterothallica*中可操作的与调控序列可操作连接的编码RFT1的核酸序列。*rft1*的示例性核苷酸序列阐述在SEQ ID NO:7中。*rft1*的示例性氨基酸序列阐述在SEQ ID NO:8中。

[0166] 在一些示例性实施方案中,所述*Th.Heterothallica*被遗传修饰,以缺失或破坏alg3、表达甘露糖苷酶1(α -1,2-甘露糖苷酶)和靶向ER的C1葡萄糖苷酶2 α 亚基以及过表达

内源Th.heterothallica RFT1翻转酶,并被进一步遗传修饰以表达包含异源高尔基体定位信号的动物来源的GNT1和动物来源的GNT2,例如表达包含来自酵母蛋白KRE2的高尔基体定位信号的人GNT1和人GNT2。

[0167] 在一些示例性实施方案中,所述Th.heterothallica被遗传修饰,以缺失或破坏alg3、表达甘露糖苷酶1(α -1,2-甘露糖苷酶)和葡萄糖苷酶2 α 亚基并表达酵母FLC2p翻转酶,并被进一步遗传修饰以表达包含异源高尔基体定位信号的动物来源的GNT1和动物来源的GNT2,例如表达包含来自Th.heterothallica蛋白KRE2的高尔基体定位信号的人GNT1和大鼠GNT2。

[0168] 在其他示例性实施方案中,所述Th.heterothallica通过下述方式进行遗传修饰:alg3的缺失或破坏;甘露糖苷酶1(α -1,2-甘露糖苷酶)和靶向ER的C1葡萄糖苷酶2 α 亚基的表达;内源Th.heterothallica RFT1翻转酶的过表达;包含Th.heterothallica KRE2a高尔基体定位信号的人GNT1和大鼠GNT2的表达;和大型利什曼原虫(Leishmania major)STT3的表达。在一些实施方案中,通过人半乳糖基转移酶或热带爪蟾(Xenopus tropicalis)半乳糖基转移酶的表达,对这种Th.heterothallica进行进一步遗传修饰。

[0169] 在其他示例性实施方案中,所述Th.heterothallica通过下述方式进行遗传修饰:alg3的缺失或破坏;甘露糖苷酶1(α -1,2-甘露糖苷酶)和靶向ER的C1葡萄糖苷酶2 α 亚基的表达;内源Th.heterothallica RFT1翻转酶的过表达;包含Th.heterothallica KRE2高尔基体定位信号的牛GNT1和大鼠GNT2的表达;和大型利什曼原虫(Leishmania major)STT3的表达。

[0170] 甘露糖苷酶1(α -1,2-甘露糖苷酶)和葡萄糖苷酶2 α 亚基

[0171] 根据一些实施方案,将靶向ER的甘露糖苷酶1(α -1,2-甘露糖苷酶)和靶向ER的C1葡萄糖苷酶2 α 亚基的表达构建体整合到Thermothelomyces heterothallica的alp3蛋白酶基因座。

[0172] 靶向ER的里氏木霉(Trichoderma reesei)甘露糖苷酶1的示例性核苷酸序列被阐述在SEQ ID NO:1中。靶向ER的里氏木霉(Trichoderma reesei)甘露糖苷酶1的示例性氨基酸序列被阐述在SEQ ID NO:2中。

[0173] 靶向ER的C1葡萄糖苷酶2 α 亚基(gls2a-HDEL)的示例性核苷酸序列被阐述在SEQ ID NO:3中。靶向ER的C1葡萄糖苷酶2 α 亚基的示例性氨基酸序列被阐述在SEQ ID NO:4中。

[0174] GNT1和GNT2

[0175] 在一些实施方案中,本发明的Th.heterothallica被遗传修饰以表达异源GNT1和GNT2。在一些实施方案中,所述异源GNT1和GNT2是动物来源的。本文所使用的“动物来源”涵盖哺乳动物源,包括例如伴侣动物(例如狗和猫)和其他哺乳动物(例如马)。正如下文中所示例的,动物来源包括例如大鼠源。术语“动物来源”进一步涵盖人类来源,正如下文中进一步示例的。

[0176] 所述异源GNT1和GNT2可以通过引入一个或多个编码GNT1和GNT2的外源多核苷酸,在根据本发明的Th.heterothallica中表达,所述多核苷酸包含与在Th.heterothallica中可操作的调控序列可操作连接的编码GNT1和GNT2的核酸序列。在一些实施方案中,所述编码GNT1和GNT2的核酸序列被包含在引入到所述Th.heterothallica中的单个多核苷酸中。在其他实施方案中,所述编码GNT1和GNT2的核酸序列被包含在引入到所述

Th.heterothallica中的两个不同多核苷酸中。

[0177] 在一些实施方案中,所述GNT1使用其自身天然存在的高尔基体定位信号在所述Th.heterothallica中表达。在其他实施方案中,所述GNT1使用异源高尔基体定位信号在所述Th.heterothallica中表达。

[0178] 在一些实施方案中,所述异源高尔基体定位信号是酵母高尔基体定位信号。在一些特定实施方案中,所述异源高尔基体定位信号来自酵母蛋白KRE2 α -1,2-甘露糖基转移酶。在一些示例性实施方案中,所述异源高尔基体定位信号来自酿酒酵母(S.cerevisiae)的KRE2。

[0179] 在其他实施方案中,所述异源高尔基体定位信号来自丝状真菌。在一些实施方案中,所述异源高尔基体定位信号来自Th.heterothallica。在一些特定实施方案中,所述异源高尔基体定位信号来自酵母蛋白KRE2的C1同源物。

[0180] 在一些实施方案中,所述GNT1是人GNT1。在一些实施方案中,所述引入到Th.heterothallica中的人GNT1包含异源高尔基体定位信号。在一些实施方案中,所述GNT1是包含酵母高尔基体定位信号的人GNT1。在一些特定实施方案中,所述GNT1是包含来自酿酒酵母(S.cerevisiae)的蛋白KRE2的高尔基体定位信号的人GNT1。KRE2信号-GNT1的示例性核苷酸信号被阐述在SEQ ID NO:5中。KRE2信号-GNT1的示例性氨基酸序列被阐述在SEQ ID NO:6中。

[0181] 在一些实施方案中,所述GNT2是大鼠GNT2。GNT2通常使用其自身天然存在的高尔基体定位信号表达。大鼠GNT2的氨基酸序列被阐述在SEQ ID NO:16中。根据本发明使用的编码大鼠GNT2的多核苷酸的示例性核酸序列被阐述在SEQ ID NO:15中。

[0182] 在其他实施方案中,所述GNT2的人GNT2。GNT2通常使用其自身天然存在的高尔基体定位信号表达。

[0183] 根据本发明的GNT1和GNT2的示例性而非限制性组合包括:

[0184] -具有酵母KRE2高尔基体定位信号的人GNT1和人GNT2;

[0185] -具有酵母KRE2高尔基体定位信号的人GNT1和大鼠GNT2;

[0186] -具有C1KRE2 α 高尔基体定位信号的人GNT1和人GNT2;

[0187] -具有C1KRE2 α 高尔基体定位信号的人GNT1和大鼠GNT2;

[0188] -具有C1 KRE2 α 高尔基体定位信号的牛GNT1和大鼠GNT2。

[0189] 每种组合代表本发明的独立实施方案。

[0190] 半乳糖基转移酶

[0191] 在一些实施方案中,本发明的Th.Heterothallica被遗传修饰以表达异源半乳糖基转移酶。在一些实施方案中,所述异源半乳糖基转移酶是动物来源的。

[0192] 半乳糖基转移酶可以通过引入编码半乳糖基转移酶的外源多核苷酸而在根据本发明的Th.heterothallica中表达,所述多核苷酸包含与在Th.heterothallica中可操作的调控序列可操作连接的编码半乳糖基转移酶的核酸序列。

[0193] 在一些实施方案中,所述半乳糖基转移酶使用其自身天然存在的高尔基体定位信号在所述Th.heterothallica中表达。在其他实施方案中,所述半乳糖基转移酶使用异源高尔基体定位信号在所述Th.heterothallica中表达。

[0194] 在一些实施方案中,所述半乳糖基转移酶是人半乳糖基转移酶(huGalT1)。在一些

实施方案中,所述被引入到Th.heterothallica中的人半乳糖基转移酶包含异源高尔基体定位信号。在一些特定实施方案中,所述人半乳糖基转移酶包含酿酒酵母(S.cerevisiae) KRE2高尔基体定位信号。具有来自酿酒酵母(S.cerevisiae)的蛋白KRE2的高尔基体定位信号的人半乳糖基转移酶的氨基酸序列被阐述在SEQ ID NO:14中。根据本发明使用的编码具有来自酿酒酵母(S.cerevisiae)的蛋白KRE2的高尔基体定位信号的人半乳糖基转移酶的示例性核酸序列被阐述在SEQ ID NO:13中。

[0195] 根据其他实施方案,所述半乳糖基转移酶来自热带爪蟾(Xenopus tropicalis) (XtGalT1)。在一些实施方案中,所述被引入到Th.heterothallica中的来自热带爪蟾(Xenopus tropicalis)的半乳糖基转移酶包含异源高尔基体定位信号。在一些特定实施方案中,所述来自热带爪蟾(Xenopus tropicalis)的半乳糖基转移酶包含来自酿酒酵母(S.cerevisiae)的蛋白KRE2的高尔基体定位信号。

[0196] STT3寡糖基转移酶

[0197] 在一些实施方案中,本发明的Th.heterothallica被遗传修饰,以表达的异源寡糖基转移酶的STT3亚基。在一些特定实施方案中,所述异源STT3是大型利什曼原虫(Leishmania major)STT3。大型利什曼原虫(Leishmania major)STT3的氨基酸序列被阐述在SEQ ID NO:12中。大型利什曼原虫(Leishmania major)STT3可以通过引入编码大型利什曼原虫(Leishmania major)STT3的外源多核苷酸而在根据本发明的Th.heterothallica中表达,所述多核苷酸包含与在Th.heterothallica中可操作的调控序列可操作连接的编码大型利什曼原虫(Leishmania major)STT3的核酸序列。编码根据本发明使用的大型利什曼原虫(Leishmania major)STT3的示例性核酸序列被阐述在SEQ ID NO:11中。

[0198] 遗传修饰的Th.heterothallica

[0199] 根据本发明的被遗传修饰以产生具有哺乳动物蛋白质(特别是人和伴侣动物蛋白质)的N-聚糖的糖蛋白的Th.heterothallica细胞如下产生:修饰、例如缺失Th.heterothallica alg3的内源基因,使得所述基因不能产生功能性蛋白质,并表达编各种酶的外源多核苷酸。

[0200] 应当理解,本文所公开的Th.heterothallica的遗传修饰不一定要求对所述遗传修饰的Th.heterothallica的每个细胞进行修饰,只要获得本文公开的产生具有哺乳动物蛋白质(特别是人和伴侣动物蛋白质)的N-聚糖的糖蛋白的所需结果即可。

[0201] 在一些实施方案中,所述Th.heterothallica被进一步遗传修饰以表达异源哺乳动物糖蛋白。在一些实施方案中,所述异源哺乳动物糖蛋白是抗体或其抗原结合片段。

[0202] 在一些示例性实施方案中,所述Th.heterothallica被遗传修饰以表达纳武单抗或其抗原结合片段。在其他示例性实施方案中,所述Th.heterothallica被遗传修饰以表达具有Th.heterothallica CBH1信号序列的纳武单抗轻链和具有Th.heterothallica CBH1信号序列的纳武单抗重链。

[0203] 在一些实施方案中,本发明提供了一种如本文所公开的遗传修饰的Th.heterothallica细胞。

[0204] 外源多核苷酸的表达通过将包含编码将要在C1中表达的蛋白质的核酸的表达构建体引入到Th.heterothallica细胞中、特别是Th.heterothallica细胞的核中进行。具体而言,根据本发明的遗传修饰意指将所述表达构建体引入到宿主基因组中。

[0205] 表达构建体在*Th.heterothallica*细胞中的引入,即*Th.heterothallica*的转化,可以通过转化丝状真菌的方法来进行。

[0206] 为了便于容易地选择转化的细胞,可以将选择标记物转化到所述*Th.heterothallica*细胞中。“选择标记物”是指编码基因产物的多核苷酸,所述基因产物赋予在非转化细胞中不存在的特定类型的表型,例如抗生素抗性(抗性标记)、利用某种资源的能力(利用/营养缺陷标记)或可以例如通过光谱测量检测的报告蛋白的表达。在食品或制药行业中营养缺陷标记物通常优选作为选择手段。所述选择标记物可以在与所述表达构建体共转化的单独多核苷酸上,也可以在所述表达构建体的同一多核苷酸上。转化后,根据所选的选择标记物,通过在例如选择性培养基上培养所述C1细胞来选择阳性转化体。在某些情况下,使用分拆的标记系统,其中选择标记物被分拆在两个质粒中,只有当所述两个质粒被共转化并通过同源重组连接在一起时,才能形成功能性选择标记物。

[0207] 当使用合成的表达系统时,将编码适合的合成转录因子(sTF)的表达盒引入到宿主细胞中。

[0208] 所述转化的DNA可以通过同源重组或非同源末端连接整合到*Th.heterothallica*染色体中。为了便于靶向整合到基因组中的特定基因座中,将与所述靶基因座对应的序列与所述表达构建体并入到同一多核苷酸中。

[0209] 然后生长所选的克隆体,并检查其是否产生具有所需N-糖型的蛋白质。将所述遗传修饰的*Th.heterothallica*在适合的条件培养。根据某些实施方案,将所述真菌在约20℃至约45℃的温度和约4.0至约8.0的培养基pH下生长。可以根据最终产品的监管要求选择特定的培养基类型。可以分离并分析所产生的糖蛋白。

[0210] GNT1、GNT2和任选的其他酶例如半乳糖基转移酶在*Th.heterothallica*中的表达可以通过由所述C1产生的N-聚糖的结构分析来确定。

[0211] 根据本发明的遗传修饰的*Th.heterothallica*产生G0 (Man₃GlcNAc₂) 作为最终的N-聚糖结构或中间的N-聚糖结构。

[0212] 在一些实施方案中,通过alg3的缺失或破坏、靶向ER的里氏木霉(*Trichoderma reesei*)甘露糖苷酶1(α -1,2-甘露糖苷酶)和靶向ER的C1葡萄糖苷酶2 α 亚基的表达、任选的异源翻转酶的表达或内源翻转酶的过表达、异源GNT1和GNT2的表达以及任选的异源STT3寡糖基转移酶的表达进行遗传修饰的*Th.heterothallica*产生分泌糖蛋白,其中G0占所述分泌糖蛋白上的N-聚糖的至少80%,优选为所述分泌糖蛋白上的N-聚糖的至少85%、至少90%、至少91%、至少92%、至少93%、至少94%或甚至至少95%。每个值代表本发明的独立实施方案。

[0213] 在一些实施方案中,被进一步遗传修饰以表达异源半乳糖基转移酶的*Th.heterothallica*产生分泌糖蛋白,其中G1和G2(G1和G2两者的总和)占所述分泌糖蛋白上的N-聚糖的至少75%,优选为所述分泌糖蛋白上的N-聚糖的至少80%、至少85%、至少90%、至少91%、至少92%、至少93%、至少94%或甚至至少95%。每个值代表本发明的独立实施方案。

[0214] 在一些实施方案中,所述*Th.heterothallica*的遗传修饰不包括异源寡糖基转移酶(OST)的表达。

[0215] 为了更充分地说明本发明的某些实施方案,给出了下述实施例。然而,它们绝不应

该被解释为限制本发明的广泛范围。本领域技术人员可以容易地设计出本文公开的原理的许多变化和修改,而不脱离本发明的范围。

[0216] 实施例

[0217] 实施例1-分三步产生具有人源化G1/2聚糖的制备纳武单抗的菌株

[0218] 产生制备人型半乳糖基化聚糖的*Thermothelomyces heterothallica* C1菌株中的第一步是从携带8种蛋白酶基因缺失并表达纳武单抗的菌株中缺失alp3基因(记载在WO 2021/094935中)。在alp3缺失后,下一个C1糖工程步骤是将靶向ER的里氏木霉(*Trichoderma reesei*)甘露糖苷酶1(α -1,2-甘露糖苷酶)和靶向ER的C1葡萄糖苷酶2 α 亚基整合到alp3蛋白酶基因座,以将所述聚糖前体修剪成适合合成人型聚糖的Do1P-GlcNAc₂-Man₃。然后,将含有人或牛GNT1(GlcNAc转移酶I)、大鼠GNT2(GlcNAc转移酶II)、C1翻转酶RFT1、来自大型利什曼原虫(*Leishmania major*)的寡糖基转移酶STT3和人GalT1(半乳糖基转移酶I)的表达盒整合到alp6蛋白酶基因座。这些基因的表达导致在GlcNAc₂Man₃聚糖的两个分支上添加GlcNAc残基,产生G0(GlcNAc₂Man₃GlcNAc₂)聚糖,随后向G0聚糖的一个或两个分支添加半乳糖残基,导致产生G1(GlcNAc₂Man₃GlcNAc₂Gal)和G2(GlcNAc₂Man₃MlcNAc₂Gal₂)聚糖。

[0219] 在第一个转化步骤中,将被设计用于整合到alp3基因座(JGI *M. thermophila*基因组数据库ID2306020)并同时表达里氏木霉(*Tr*)mns1-HDEL(JGI里氏木霉基因组数据库ID45717)和C1gls2a-HDEL(JGI *M. thermophila*基因组数据库ID2125259)的DNA构建体分两部分构建到两个单独的质粒中。HDEL是四个氨基酸的C-端ER定位信号。第一个质粒含有用于整合的alp3 5'侧翼区片段、用于Tr-mns1-HDEL的表达盒(其中所述基因位于C1 bg18启动子和终止子之间(JGI *M. thermophila*基因组数据库ID115968)、合成转录因子(sTF,用于3'质粒中的合成启动子)、C1 cbh1终止子的直接重复序列(JGI *M. thermophila*基因组数据库ID109566)和amdS标记基因(编码构巢曲霉(*Aspergillus nidulans*)的乙酰胺酶)的前2/3。第二个质粒含有amdS标记的最后2/3、靶向sTF盒末端的直接重复片段(用于通过重组去除amdS标记)、合成的AnSES启动子(Rantasalo A等,2018,用于真菌的通用基因表达系统(A universal gene expression system for fungi),*Nucl. Acids Research* 46(18):e111)和chi1终止子(JGI *M. thermophila*基因组数据库ID50608)之间的用于C1 gls2a-HDEL的表达盒,以及用于整合的alp3 3'侧翼区片段。这两个质粒中的amdS标记片段彼此交叠,并且该区域在所述质粒之间经历C1中的同源重组,同时5'和3'侧翼区域片段与alp3基因两侧上的基因组DNA重组。选择标记片段之间的重组使所述标记基因功能化,并使转化体能够在选择下生长。

[0220] 含有alp3 5'侧翼区片段、用于Tr mns1-HDEL的表达盒、sTF和amdS标记的前2/3的构建体,被阐述在SEQ ID NO:17(pMYT1288)中。5'侧翼序列对应于SEQ ID NO:17的第9-1196位。bg18启动子序列对应于SEQ ID NO:17的第1202-2593位。编码具有人工HDEL ER驻留信号的基因组里氏木霉(*Trichoderma reesei*)甘露糖苷酶1的核酸序列对应于SEQ ID NO:17的第2594-4420位。编码具有HDEL信号的里氏木霉(*T. reesei*)MNS1的核酸序列也被阐述为SEQ ID NO:1(Tr MNS1-HDEL nt),氨基酸序列为SEQ ID NO:2(Tr MNS1-HDEL aa)。bg18终止子序列对应于SEQ ID NO:17的第4421-4887位。合成转录因子盒的核酸序列对应于SEQ ID NO:17的第4901-6550位。amdS标记基因的前2/3对应于SEQ ID NO:17的第7060-

9126位。所述片段通过PCR产生,在琼脂糖凝胶中分离,纯化,并通过Gibson组装(NEBuilder® HiFi DNA组装克隆试剂盒,New England Biolabs)方法克隆到骨架载体pRS426中,得到5'臂表达载体pMYT1288,其通过测序进行验证。

[0221] 含有amdS标记的最后2/3、用于C1 gls2a-HDEL的表达盒和用于整合的alp3 3'侧翼区片段的构建体,被阐述在SEQ ID NO:18(pMYT0721)中。amdS标记基因的2/3对应于SEQ ID NO:18的第17-1738位。合成的AnSES启动子序列对应于SEQ ID NO:18的第2255-2747位。编码具有人工HDEL ER驻留信号的C1葡萄糖苷酶2 α 的序列对应于SEQ ID NO:18的第2748-5760位。编码具有HDEL信号的C1 GLS2 α 的核酸序列也被阐述为SEQ ID NO:3(C1 gls2a-HDEL nt),氨基酸序列为SEQ ID NO:4(C1 gls2a-HDEL aa)。chi1终止子序列对应于SEQ ID NO:18的第5761-6406位。alp3 3'侧翼序列对应于SEQ ID NO:18的第6415-7548位。所述片段通过PCR产生,在琼脂糖凝胶中分离,纯化,并通过Gibson组装(NEBuilder® HiFi DNA组装克隆试剂盒,New England Biolabs)方法克隆到骨架载体pRS426中,得到3'臂表达载体pMYT0721,其通过测序进行验证。

[0222] 对于第二个转化步骤,将被设计用于整合到alp6基因座(JGI M.thermophila基因组数据库ID94536)并同时表达GNT1和GNT2以及RFT1、STT3和GalT1的DNA构建体分两部分构建到两个单独的质粒中。第一个质粒含有用于整合的alp6 5'侧翼区片段、用于GNT1的表达盒(其中与C1 KRE2高尔基体定位信号(JGI M.thermophila基因组数据库ID2300989)融合的人或牛GNT1基因位于C1 bg18启动子和bg18终止子之间)、位于泛素样蛋白基因(JGI M.thermophila基因组数据库ID2315548)的启动子和终止子之间的C1翻转酶RFT1(JGI M.thermophila基因组数据库ID2307799),以及C1 pyr4标记基因的前2/3(JGI M.thermophila基因组数据库ID2311494)。第二个质粒含有C1 pyr4标记基因的最后2/3、靶向pyr4盒的开端的直接重复片段(用于通过重组去除pyr4标记)、位于C1假想蛋白(JGI M.thermophila基因组数据库ID2315630)的启动子和终止子之间的用于来自大型利什曼原虫(Leishmania major)的STT3的反向表达盒、位于泛素样蛋白基因的启动子与bg18终止子之间的用于与酿酒酵母(Saccharomyces cerevisiae)KRE2高尔基体定位信号融合的人GalT1的表达盒、位于翻译延伸因子1A(JGI M.thermophila基因组数据库ID2298136)的启动子和C1假想蛋白(JGI M.thermophila基因组数据库ID114107)的终止子之间的来自大鼠的GNT2基因,最后是用于整合的alp6 3'侧翼区片段。这两个质粒中的C1 pyr4标记片段彼此交叠,并且该区域在所述质粒之间经历C1中的同源重组,同时5'和3'侧翼区域片段与alp6基因两侧上的基因组DNA重组。选择标记片段之间的重组使所述标记基因有功能,并使转化体能够在选择下生长。

[0223] 含有alp6 5'侧翼区片段、用于人GNT1、C1 RFT1的表达盒和C1 pyr4标记基因的前2/3的第一个5'构建体被阐述为SEQ ID NO:19(pMYT1451)。所述5'侧翼序列对应于SEQ ID NO:19的第8-1156位。bg18启动子序列对应于SEQ ID NO:19的第1165-2556位。编码与C1 KRE2高尔基体定位信号融合的人GNT1的核酸序列对应于SEQ ID NO:19的第2557-4042位,其中第2557-2821位编码C1 KRE2定位信号,并且第2822-4042位编码人GNT1。编码与C1 KRE2高尔基体定位信号融合的人GNT1的核酸序列也被阐述为SEQ ID NO:5(C1 kre2-huGNT1nt)。编码人GNT1的核酸序列针对M.thermophila C1进行了密码子优化,并由GenScript(USA)合成。在所述合成的序列中,人GNT1的第1-100个氨基酸的区域被C1 KRE2

高尔基体定位信号(第1-70个氨基酸)替换。C1 KRE2-GNT1的完整氨基酸序列被阐述为SEQ ID NO:6(C1 KRE2-huGNT1 aa)。bgl8终止子序列对应于SEQ ID NO:19的第4046-4512位。泛素样蛋白启动子序列对应于SEQ ID NO:19的第4513-5532位。C1翻转酶rft1基因对应于SEQ ID NO:19的第5533-7449位。编码C1 RFT1的核酸序列也被阐述为SEQ ID NO:7(C1 RFT1 nt),氨基酸序列为SEQ ID NO:8(C1 RFT1 aa)。泛素样蛋白终止子序列对应于SEQ ID NO:19的第7458-7953位。C1 pyr4标记基因的前2/3对应于SEQ ID NO:19的第7969-9748位。

[0224] 含有alp6 5'侧翼区片段、用于牛GNT1、C1 RFT1的表达盒和C1pyr4标记基因的前2/3的第二个5'构建体被阐述为SEQ ID NO:20(pMYT1452)。所述5'侧翼序列对应于SEQ ID NO:20的第8-1156位。bgl8启动子序列对应于SEQ ID NO:20的第1165-2556位。编码与C1 KRE2高尔基体定位信号融合的牛GNT1的核酸序列对应于SEQ ID NO:20的第2557-4051位,其中第2557-2821位编码C1 KRE2定位信号,第2822-4051位编码牛GNT1。编码与C1 KRE2高尔基体定位信号融合的牛GNT1的核酸序列也被阐述为SEQ ID NO:9(C1 kre2-boGNT1 nt)。编码牛GNT1的核酸序列针对M.thermophila C1进行了密码子优化,并由GenScript(USA)合成。在所述合成的序列中,牛GNT1的第1-38个氨基酸的区域在表达质粒克隆后被去除,并被C1 KRE2高尔基体定位信号(第1-70个氨基酸)替换。C1 KRE2-GNT1的完整氨基酸序列被阐述为SEQ ID NO:10(C1 KRE2-boGNT1 aa)。bgl8终止子序列对应于SEQ ID NO:20的第4052-4518位。泛素样蛋白启动子序列对应于SEQ ID NO:20的第4524-5543位。为pMYT1451描述的C1翻转酶RFT1基因对应于SEQ ID NO:20的第5544-7460位。泛素样蛋白终止子序列对应于SEQ ID NO:20的第7469-7964位。C1 pyr4标记基因的前2/3对应于SEQ ID NO:20的第7980-9759位。用于两个5'质粒的片段通过PCR产生,在琼脂糖凝胶中分离,纯化,并通过Gibson组装(NEBuilder® HiFi DNA组装克隆试剂盒,New England Biolabs)方法克隆到骨架载体pRS426中,得到5'臂表达载体pMYT1451和pMYT1452,其通过测序进行验证。

[0225] 含有C1 pyr4标记基因的最后2/3、用于来自大型利什曼原虫(Leishmania major)的STT3的反向表达盒、用于人GalT1的反向表达盒、用于大鼠GNT2基因的反向表达盒和用于整合的alp6 3'侧翼区片段的构建体被阐述在SEQ ID NO:21(pMYT1453)中。C1 pyr4标记基因的2/3对应于SEQ ID NO:21的第17-1273位。假想蛋白(ID2315630)启动子序列对应于SEQ ID NO:21的第5965-4957位。编码来自大型利什曼原虫(Leishmania major)的STT3的序列针对M.thermophila C1进行了密码子优化并由GenScript(USA)合成,其对应于SEQ ID NO:21的第4956-2383位。编码大型利什曼原虫(Leishmania major)STT3基因的核酸序列也被阐述为SEQ ID NO:11(LmSTT3nt)。大型利什曼原虫STT3的完整氨基酸序列被阐述为SEQ ID NO:12(LmSTT3 aa)。假想蛋白(ID2315630)终止子序列对应于SEQ ID NO:21的第2374-1795位。编码与Sc KRE2高尔基体定位信号融合的人GalT1的序列对应于SEQ ID NO:21的第7704-6439位,其中第7704-7405位编码Sc KRE2定位信号,第7404-6439位编码人GalT1。泛素样蛋白启动子序列对应于SEQ ID NO:21的第8724-7705位。编码与Sc KRE2高尔基体定位信号融合的人GalT1的核酸序列也被阐述为SEQ ID NO:13(Sckre2-huGalT1 nt)。编码人GalT1的核酸序列针对M.thermophila C1进行了密码子优化,并由GenScript(USA)合成。在所述合成的序列中,人GalT1的第1-77个氨基酸的区域在表达质粒克隆后被移除,并被Sc KRE2高尔基体定位信号(第1-100个氨基酸)替换。Sc KRE2-GalT1的完整氨基酸序列被阐述为SEQ ID NO:14(ScKRE2-huGalT1 aa)。bgl8终止子序列对应于SEQ ID NO:21的第6438-

5972位。翻译延伸因子1A启动子序列对应于SEQ ID NO:21的第11618-10562位。针对 *M.thermophila* C1进行密码子优化并由GenScript (USA) 合成的编码大鼠GNT2的序列对应于SEQ ID NO:21的第10561-9233位。编码大鼠GNT2的核酸序列也被阐述为SEQ ID NO:15 (大鼠GNT2nt)。大鼠GNT2的完整氨基酸序列被阐述为SEQ ID NO:16 (大鼠GNT2aa)。假想蛋白 (ID114107) 终止子序列对应于SEQ ID NO:21的第9232-8729位。alp6 3' 侧翼序列对应于SEQ ID NO:21的第11626-12681位。片段通过PCR产生,在琼脂糖凝胶中分离,纯化,并通过Gibson组装(**NEBuilder[®]** HiFi DNA组装克隆试剂盒,New England Biolabs)方法克隆到骨架载体pRS426中,得到3' 臂表达载体pMYT1453,其通过测序进行验证。

[0226] 为了获得具有G1/2N-聚糖的制备纳武单抗的菌株,将上述表达质粒相继转化到pyr4-阴性制备纳武单抗的alg3缺失菌株M3291中(描述在WO 2021/094935中)。在每次转化中,使用成对的一个5' 臂表达载体和一个3' 臂表达载体(用MssI从表达质粒骨架中切下)。所述对是:

[0227] • 第一轮:SEQ ID NO:17(pMYT1288) (alp3 5' 侧翼区-Th Mns1-HDEL盒-sTF盒-2/3amdS)+SEQ ID NO:18(pMYT0721) (2/3 amdS-C1 gls2a-HDEL盒-alp3 3' 侧翼片段)

[0228] • 第二轮:SEQ ID NO:19(pMYT1451) 或SEQ ID NO:20(pMYT1452) (alp6 5' 侧翼区-人或牛GNT1盒-C1 RFT1盒-2/3pyr4)+SEQ ID NO:21(pMYT1453) (2/3pyr4-LmSTT3盒-人GalT1盒-大鼠GNT2盒-alp6 3' 侧翼片段)

[0229] C1转化和转化体选择基本上如WO 2021/094935中所述来进行。对于第一轮转化而言,选择是基于功能性amdS基因,对于第二轮而言是基于功能性pyr4基因。通过PCR筛选转化体,找到其中alp3缺失位点(第一轮) 或alp6基因(第二轮) 被构建体替换的克隆体。

[0230] 使用特定引物验证了质粒/基因与转化体基因组的正确整合。具有表达构建体的正确整合和进一步alp3基因座的缺失的第一轮转化体被储存为M4855和M4856。将菌株M4855用于第二轮转化。具有构建体的正确整合和alp6基因的丧失的第二轮转化体分别被储存为M5129和M5130(具有人GNT1) 以及M5131和M5132(具有牛GNT1)。

[0231] 如WO 2021/094935中对亲本菌株M3291的摇瓶培养所述,将来自两轮转化的构建的C1菌株在250ml摇瓶中,在50ml液体培养基中生长。培养在35℃、~200RPM下进行4天。通过离心去除菌丝体,并使用**AKTA** Start自动HPLC系统(Cytiva) 和HiTrap MabSelect Sure或HiTrap MabSelect PrismA预装1ml柱,根据制造商(Cytiva) 的说明书将培养的上清液用于纳武单抗的蛋白A亲和纯化。如WO 2021/094935的实施例1中所述,对来自所有样品的峰级分进行释放N-聚糖分析。图1中总结了所有菌株的结果。图1A显示,通过表达TrMNS1-HDEL和C1 GLS2A-HDEL两者,纳武单抗上的主要N-聚糖是Man3 (M3),其含量接近90%,而仅有非常少的更高甘露糖结构(GlcNAc2Man4、GlcNAc2Man5),几乎没有Hex6 (GlcNAc2Man5Glc),后者被认为阻断了N-聚糖修饰到G0及以上。图1B-1D显示,通过使用本实施例中描述的alp6靶向表达载体,在靶mAb上达到了非常高的G1 (GlcNAc2Man3GlcNAc2Gal) 和G2 (GlcNAc2Man3GlcNAc2Gal2) N-聚糖水平。向人型G1/2N-聚糖的转化率为88-98%,即使用这种方法几乎完全转化为人型聚糖。除了非常高水平的人型N-聚糖之外,图2还显示了使用添加到用于分析的每个样品的固定量的内标进行归一化的样品之间释放N-聚糖的量。来自不同纳武单抗样品的释放N-聚糖的百分率与参考Opdivo (在哺乳动物系统中生产的商业化纳武单抗,Bristol-Myers Squibb) 大致在相同水平上,

表明这种人源化糖基化的方法不对N-糖基化位点的占有率产生负面影响。

[0232] 接下来,在以酵母提取物为有机氮源和葡萄糖为碳源的培养基中,使用补料分批法在ambr250或1L生物反应器中培养C1菌株M3291、M5130和M5132。培养在38℃下进行7天。在培养结束后,通过以4000g离心20分钟去除菌丝体,添加浓度为1-2mM的苯甲基磺酰氟(PMSF)以抑制得到的培养上清液中的蛋白酶活性,并将样品储存在-80℃下。使用与上述摇瓶样品纯化基本上相同的蛋白A亲和纯化方法,从第七天的发酵样品中纯化纳武单抗。对来自所有样品的峰级分进行释放N-聚糖分析。图3总结了所有三个菌株的结果。图3A显示,尽管亲本菌株M3291具有超过80%的所需前体Man3(M3),但它也具有超过3.5%的Hex6(M6),其被认为阻止N-聚糖人源化(转化到G0和进一步修饰)和一些更高的甘露糖结构(Man4、Man5)。图3B-3D显示,靶mAb上的释放N-聚糖接近完全人源化,即检测到的99%或更多的N-聚糖属于G0至G2种类。图4显示了使用添加到用于分析的每个样品的固定量的内标(唾液酸基聚糖肽)进行归一化的样品之间释放N-聚糖的量。与摇瓶中一样,来自具有人源化N-聚糖的两种纳武单抗菌株的释放N-聚糖的百分率似乎与参考Opdivo(在哺乳动物系统中生产的商业化纳武单抗)在相同水平上,进一步证实了对N-聚糖位点占有率没有负面影响。

[0233] 实施例2-一步产生具有人源化G1/2聚糖的空菌株

[0234] 在该第二步中,将实施例1中描述的生成人型半乳糖基化聚糖的*Thermothelomyces heterothallica* C1菌株所需的所有三个糖修饰步骤合并。简而言之,将携带14个蛋白酶基因的缺失和较弱启动子下的kex2蛋白酶基因的菌株中alg3基因的缺失,与靶向ER的里氏木霉(*Trichoderma reesei*)甘露糖苷酶1(α -1,2-甘露糖苷酶)、靶向ER的C1葡萄糖苷酶2 α 亚基、人GNT1(GlcNAc转移酶I)、大鼠GNT2(GlcNAc转移酶II)、来自大型利什曼原虫(*Leishmania major*)的寡糖基转移酶STT3和人GalT1(半乳糖基转移酶I)在alg3基因座中的整合相组合。这些基因的表达首先导致将聚糖前体修剪成适合合成人型聚糖的Do1P-GlcNAc2-Man₃,然后在所述GlcNAc₂Man₃聚糖的两个分支上添加GlcNAc残基,产生G0(GlcNAc₂Man₃GlcNAc₂)聚糖,然后向G0聚糖的一个或两个分支添加半乳糖残基,导致产生G1(GlcNAc₂Man₃GlcNAc₂Gal)和G2(GlcNAc₂Man₃MlcNAc₂Gal₂)聚糖。

[0235] 将被设计用于整合到alg3基因座(JGI *M. thermophila*基因组数据库ID 2310419)并同时表达里氏木霉(Tr)mns1-HDEL(JGI里氏木霉(*T. reesei*)基因组数据库ID45717)、C1 gls2a-HDEL(JGI *M. thermophila*基因组数据库ID2125259)、人GNT1、大鼠GNT2以及大型利什曼原虫(*Leishmania major*) (Lm) STT3和人GalT1的DNA构建体分两部分构建到两个单独的质粒中。HDEL代表四个氨基酸的C-端ER定位信号。

[0236] 第一个质粒含有用于整合的alg3 5'侧翼区片段、用于GNT1的表达盒(其中在C1 bgl8启动子和bgl8终止子(JGI *M. thermophila*基因组数据库ID115968)之间将人GNT1基因与C1KRE2高尔基体定位信号(JGI *M. thermophila*基因组数据库ID2300989)融合)、位于泛素样蛋白基因的启动子(JGI *M. thermophila*基因组数据库ID2315548)和C1假想蛋白的终止子(JGI *M. thermophila*基因组数据库ID2302731)之间的用于与酿酒酵母(*Saccharomyces cerevisiae*)KRE2高尔基体定位信号融合的人GalT1的表达盒、位于合成的AnSES启动子(Rantasalo A等,2018,用于真菌的通用基因表达系统(A universal gene expression system for fungi),*Nucl. Acids Research* 46(18):e111)和chi1终止子(JGI *M. thermophila*基因组数据库ID50608)之间的用于C1 gls2a-HDEL的表达盒、用于合成的转

录因子 (sTF, 用于C1 gls2a-HDEL的合成启动子) 的表达盒以及pyr4标记基因的前2/3 (JGI M.thermophila基因组数据库ID2311494)。

[0237] 第二个质粒含有C1 pyr4标记基因的最后2/3、靶向pyr4盒的开端的直接重复片段 (用于通过重组去除pyr4标记)、位于C1假想蛋白 (JGI M.thermophila基因组数据库ID2315630) 的启动子和终止子之间的用于大型利什曼原虫 (*Leishmania major*) 的STT3的反向表达盒、用于Tr mns1-HDEL的反向表达盒 (其中所述基因位于C1 bgl8启动子 (JGI M.thermophila基因组数据库ID115968) 或泛素样蛋白基因的启动子 (JGI M.thermophila基因组数据库ID2315548) 与泛素样蛋白质基因终止子 (JGI M.thermophila基因组数据库ID2315548) 之间)、位于翻译延伸因子1A的启动子 (JGI M.thermophila基因组数据库ID2298136) 与C1假想蛋白的终止子 (JGI M.thermophila基因组数据库ID114107) 之间的用于大鼠GNT2基因的反向表达盒, 最后是用于整合的alg3 3'侧翼区片段。这两个质粒中的C1 pyr4标记片段彼此交叠, 并且该区域在所述质粒之间经历C1中的同源重组, 同时5' 和3' 侧翼区域片段与alg3基因两侧上的基因组DNA重组。选择标记片段之间的重组使所述标记基因有功能, 并使转化体能够在选择下生长。

[0238] 含有alg3 5'侧翼区片段、用于人GNT1、人GalT1、C1 gls2a-HDEL、sTF的表达盒和pyr4标记的前2/3的构建体被阐述在SEQ ID NO:23 (pMYT1974) 中。所述alg3 5'侧翼序列对应于SEQ ID NO:23的第9-1010位。bgl8启动子序列对应于SEQ ID NO:23的第1018-2409位。编码与C1 KRE2高尔基体定位信号融合的人GNT1的核酸序列对应于SEQ ID NO:23的第2410-3898位, 其中第2410-2674位编码C1 KRE2定位信号, 第2675-3898位编码人GNT1。编码与C1 KRE2高尔基体定位信号融合的人GNT1的核酸序列也被阐述为SEQ ID NO:5 (C1 kre2-huGNT1 nt)。编码人GNT1的核酸序列针对M.thermophila C1进行了密码子优化, 并由GenScript (USA) 合成。在所述合成的序列中, 人GNT1的第1-100个氨基酸的区域被C1 KRE2高尔基体定位信号 (第1-70个氨基酸) 替换。C1 KRE2-GNT1的完整氨基酸序列被阐述为SEQ ID NO:6 (C1KRE2-huGNT1 aa)。bgl8终止子序列对应于SEQ ID NO:23的第3899-4365位。泛素样蛋白启动子序列对应于SEQ ID NO:23的第4374-5393位。编码与Sc-KRE2高尔基体定位信号融合的人GalT1的序列对应于SEQ ID NO:23的第5394-5693位, 其中第5394-5693位编码Sc KRE2定位信号, 第5694-6659位编码人GalT1。编码与Sc KRE2高尔基体定位信号融合的人GalT1的核酸序列也被阐述为SEQ ID NO:13 (Sckre2-huGalT1 nt)。编码人GalT1的核酸序列针对M.thermophila C1进行了密码子优化, 并由GenScript (USA) 合成。在所述合成的序列中, 人GalT1的第1-77个氨基酸的区域在表达质粒克隆后被移除, 并被Sc KRE2高尔基体定位信号 (第1-100个氨基酸) 替换。Sc KRE2-GalT1的完整氨基酸序列被阐述为SEQ ID NO:14 (ScKRE2-huGalT1 aa)。假想蛋白 (ID2302731) 的终止子序列对应于SEQ ID NO:23的第6660-7065位。合成的AnSES启动子序列对应于SEQ ID NO:23的第7073-7565位。编码具有人工HDEL ER驻留信号的C1葡萄糖苷酶2 α 的序列对应于SEQ ID NO:23的第7566-10578位。编码具有HDEL信号的C1GLS2 α 的核酸序列也被阐述为SEQ ID NO:3 (C1 gls2a-HDEL nt), 氨基酸序列为SEQ ID NO:4 (C1 GLS2A-HDEL aa)。chi1终止子序列对应于SEQ ID NO:23的第10579-11224位。合成转录因子盒的核酸序列对应于SEQ ID NO:23的第11225-12874位。C1 pyr4标记基因的前2/3对应于SEQ ID NO:23的第12888-14667位。片段通过PCR产生, 在琼脂糖凝胶中分离, 纯化, 并使用酵母重组法逐步克隆到骨架载体pRS426中, 得到最终的5' 臂表

达载体pMYT1974,其通过测序进行验证。

[0239] 含有C1 pyr4标记基因的最后2/3、用于标记去除的直接重复序列、用于来自大型利什曼原虫 (*Leishmania major*) 的STT3的反向表达盒、用于Tr Mns1-HDEL的反向表达盒、用于大鼠GNT2基因的反向表达盒和用于整合的alg3 3'侧翼区片段的两个平行构建体分别被阐述为SEQ ID NO:24(pMYT1963)和SEQ ID NO:25(pMYT1964)。C1pyr4标记基因的最后2/3对应于SEQ ID NO:24的第17-1273位。靶向pyr4盒的开端的直接重复片段对应于SEQ ID NO:24的第1290-1786位。假想蛋白 (ID2315630) 启动子序列对应于SEQ ID NO:24的第5965-4957位。针对*T.heterothallica* C1进行了密码子优化并由GenScript (USA) 合成的编码大型利什曼原虫 (*Leishmania major*) STT3基因的序列对应于SEQ ID NO:24的第4956-2383位。编码大型利什曼原虫 (*Leishmania major*) STT3基因的核酸序列也被阐述为SEQ ID NO:11 (LmSTT3 nt)。大型利什曼原虫 (*L.major*) STT3的完整氨基酸序列被阐述为SEQ ID NO:12 (LmSTT3aa)。假想蛋白 (ID2315630) 终止子序列对应于SEQ ID NO:24的第2374-1795位。泛素样蛋白启动子序列对应于SEQ ID NO:24的第9316-8297位。编码具有人工HDEL ER驻留信号的基因组里氏木霉 (*Trichoderma reesei*) 甘露糖苷酶1的核酸序列对应于SEQ ID NO:24的第8296-6470位。编码具有HDEL信号的里氏木霉MNS1的核酸序列也被阐述为SEQ ID NO:1 (Tr mns1-HDEL nt), 氨基酸序列为SEQ ID NO:2 (Tr MNS1-HDEL aa)。泛素样蛋白终止子序列对应于SEQ ID NO:24的第6469-5974位。翻译延伸因子1A启动子序列对应于SEQ ID NO:24的第12210-11154位。针对*T.heterothallica* C1进行密码子优化并由GenScript (USA) 合成的编码大鼠GNT2的序列对应于SEQ ID NO:24的第11153-9825位。编码大鼠GNT2的核酸序列也被阐述为SEQ ID NO:15 (大鼠GNT2 nt)。大鼠GNT2的完整氨基酸序列被阐述为SEQ ID NO:16 (大鼠GNT2 aa)。假想蛋白 (ID114107) 终止子序列对应于SEQ ID NO:24的第9824-9321位。alg3 3'侧翼序列对应于SEQ ID NO:24的第12218-13217位。通过PCR产生片段,在琼脂糖凝胶中分离,纯化并使用酵母重组法逐步克隆到骨干载体pRS426中,得到最终的3' arm表达载体pMYT1963,并通过测序进行验证。片段通过PCR产生,在琼脂糖凝胶中分离,纯化,并使用酵母重组法逐步克隆到骨架载体pRS426中,得到最终的3'臂表达载体pMYT1963,其通过测序进行验证。

[0240] 在SEQ ID NO:25(pMYT1964)中,C1 pyr4标记基因的最后2/3对应于第17-1273位。靶向pyr4盒的开端的直接重复片段对应于SEQ ID NO:25的地1290-1786位。假想蛋白 (ID2315630) 启动子序列对应于SEQ ID NO:25的第5965-4957位。针对*T.heterothallica* C1进行了密码子优化并由GenScript (USA) 合成的编码来自大型利什曼原虫 (*Leishmania major*) 的STT3的序列对应于SEQ ID NO:25的第4956-2383位。编码大型利什曼原虫 (*Leishmania major*) STT3基因的核酸序列也被阐述为SEQ ID NO:11 (LmSTT3nt)。大型利什曼原虫 (*L.major*) STT3的完整氨基酸序列被阐述为SEQ ID NO:12 (LmSTT3 aa)。假想蛋白 (ID2315630) 终止子序列对应于SEQ ID NO:25的第2374-1795位。bgl8启动子序列对应于SEQ ID NO:25的第9688-8297位。编码具有人工HDEL ER驻留信号的基因组里氏木霉 (*Trichoderma reesei*) 甘露糖苷酶1的核酸序列对应于SEQ ID NO:25的第8296-6470位。编码具有HDEL信号的里氏木霉 (*T.reesei*) MNS1的核酸序列也被阐述为SEQ ID NO:1 (Tr mns1-HDEL nt), 氨基酸序列为SEQ ID NO:2 (Tr MNS1-HDEL aa)。泛素样蛋白终止子序列对应于SEQ ID NO:25的第6469-5974位。翻译延伸因子1A启动子序列对应于SEQ ID NO:25的

第12584-11528位。针对*M. thermophila* C1进行了密码子优化并由GenScript (USA) 合成的编码大鼠GNT2的序列对应于SEQ ID NO:25的第11527-10199位。编码大鼠GNT2的核酸序列也被阐述为SEQ ID NO:15 (大鼠GNT2 nt)。大鼠GNT2的完整氨基酸序列被阐述为SEQ ID NO:16 (大鼠GNT2 aa)。假想蛋白 (ID114107) 终止子序列对应于SEQ ID NO:25的第10198-9695位。*alg3* 3' 侧翼序列对应于SEQ ID NO:25的第12592-13591位。片段通过PCR产生,在琼脂糖凝胶中分离,纯化,并使用酵母重组法逐步克隆到骨架载体pRS426中,得到最终的3'臂表达载体pMYT1964,其通过测序进行验证。

[0241] 为了获得两种不同的具有G1/2N-聚糖的空菌株(不表达异源哺乳动物糖蛋白),将上述表达质粒一步转化到携带14种蛋白酶基因的缺失和在较弱启动子下的*kex2*蛋白酶基因的*pyr4*⁻菌株中。在每次转化中,使用成对的一个5'臂表达载体和一个3'臂表达载体(用*MssI*从表达质粒骨架中切下)。所述对是:SEQ ID NO:23 (pMYT1974;*alg3* 5'侧翼片段-人GNT1盒-人GalT1盒-C1 *gls2a*-HDEL盒-sTF盒-2/3 *pyr4*)+SEQ ID NO:24 (pMYT1963)或SEQ ID NO:25 (pMYT1964) (2/3*pyr4*-DR-LmSTT3盒-TrMns1-HDEL盒-大鼠GNT2盒-*alg3* 3'侧翼片段)。C1转化和转化体选择基本上如WO 2021/094935中所述进行。转化的选择是基于功能性*pyr4*基因。通过PCR筛选转化体,以找到其中*alg3*基因被构建体替换的克隆。用特定引物验证质粒/基因与转化体基因组的正确整合。具有构建体的正确整合和*alg3*基因丧失的转化体分别被储存为M6589和M6596 (具有用于TrMns1-HDEL的泛素样蛋白启动子)以及M6590和M6597 (具有用于TrMns1-HDEL的*bg18*启动子)。

[0242] 如WO 2021/094935中为菌株培养所述,将来自两种转化的构建的C1菌株在24孔深孔板中,在3.5ml液体培养基中生长。培养在35℃、800RPM、80%湿度下进行4天。通过离心去除菌丝体,并如WO 2021/094935的实施例1中所述对培养上清液进行释放N-聚糖分析。图5总结了所有四种菌株的结果。图5显示,通过从*alg3*基因座表达影响聚糖结构的所有六个基因,总分泌蛋白上的主要N-聚糖80%以上是人型的(属于G0至G2种类的N-聚糖),并且没有可检测的Hex6 (GlcNAc₂Man₅Glc),后者被认为阻断了N-聚糖修饰到G0及以上。同样值得注意的是最终的G2N-聚糖的量。已发现在总分泌蛋白上人型N-聚糖的量(以%计)总是低于在任何纯化的靶蛋白(例如实施例1中的纳武单抗)上观察到的量。

[0243] 序列

[0244] SEQ ID NO:1-里氏木霉(*T.reesei*)*mns1*-HDEL nt (编码序列,即去除内含子的,1684bp)

[0245] SEQ ID NO:2-里氏木霉(*T.reesei*)*MNS1*-HDEL aa (527 aa)

[0246] SEQ ID NO:3-C1 *gls2a*-HDEL nt (编码序列,即去除内含子的,2961 bp)

[0247] SEQ ID NO:4-C1 *GLS2A*-HDEL aa (986 aa)

[0248] SEQ ID NO:5-C1 *kre2*-huGNT1 nt (编码序列,即去除内含子的,1434bp)

[0249] SEQ ID NO:6-C1 *KRE2*-huGNT1 aa (477 aa)

[0250] SEQ ID NO:7-C1 *rft1*nt (编码序列,即去除内含子的,1839bp)

[0251] SEQ ID NO:8-C1 *RFT1* aa (612 aa)

[0252] SEQ ID NO:9-C1 *kre2*-boGNT1 nt (编码序列,即去除内含子的,1440bp)

[0253] SEQ ID NO:10-C1 *KRE2*-boGNT1 aa (479 aa)

[0254] SEQ ID NO:11-LmSTT3 nt (2574bp)

- [0255] SEQ ID NO:12-LmSTT3 aa(857 aa)
- [0256] SEQ ID NO:13-Sckre2-huGalT1 nt(1266bp)
- [0257] SEQ ID NO:14-ScKRE2-huGalT1 aa(421 aa)
- [0258] SEQ ID NO:15-大鼠GNT2 nt(1329bp)
- [0259] SEQ ID NO:16-大鼠GNT2 aa(442 aa)
- [0260] SEQ ID NO:17-pMYT1288(14648bp)
- [0261] SEQ ID NO:18-pMYT0721(13062bp)
- [0262] SEQ ID NO:19-pMYT1451(15270bp)
- [0263] SEQ ID NO:20-pMYT1452(15281bp)
- [0264] SEQ ID NO:21-pMYT1453(18195bp)
- [0265] SEQ ID NO:22-Thermothelomyces heterothallica C1的rDNA
- [0266] SEQ ID NO:23-pMYT1974(20189bp)
- [0267] SEQ ID NO:24-pMYT1963(18731bp)
- [0268] SEQ ID NO:25-pMYT1964(19105bp)

[0269] 特定实施方案的上述描述充分揭示了本发明的总体性质,使得其他人可以不需过多实验并且不背离一般性概念,通过应用当前知识容易地修改和/或改编此类特定实施方案以适用于各种不同应用,因此,此类改编和修改应当并且旨在被涵盖在所公开的实施方案的等同物的含义和范围之内。应当理解,本文中使用的短语或术语是出于描述而不是限制的目的。用于执行各种公开的化学结构和功能的手段、材料和步骤可以采取各种不同的替代形式,而不背离本发明。

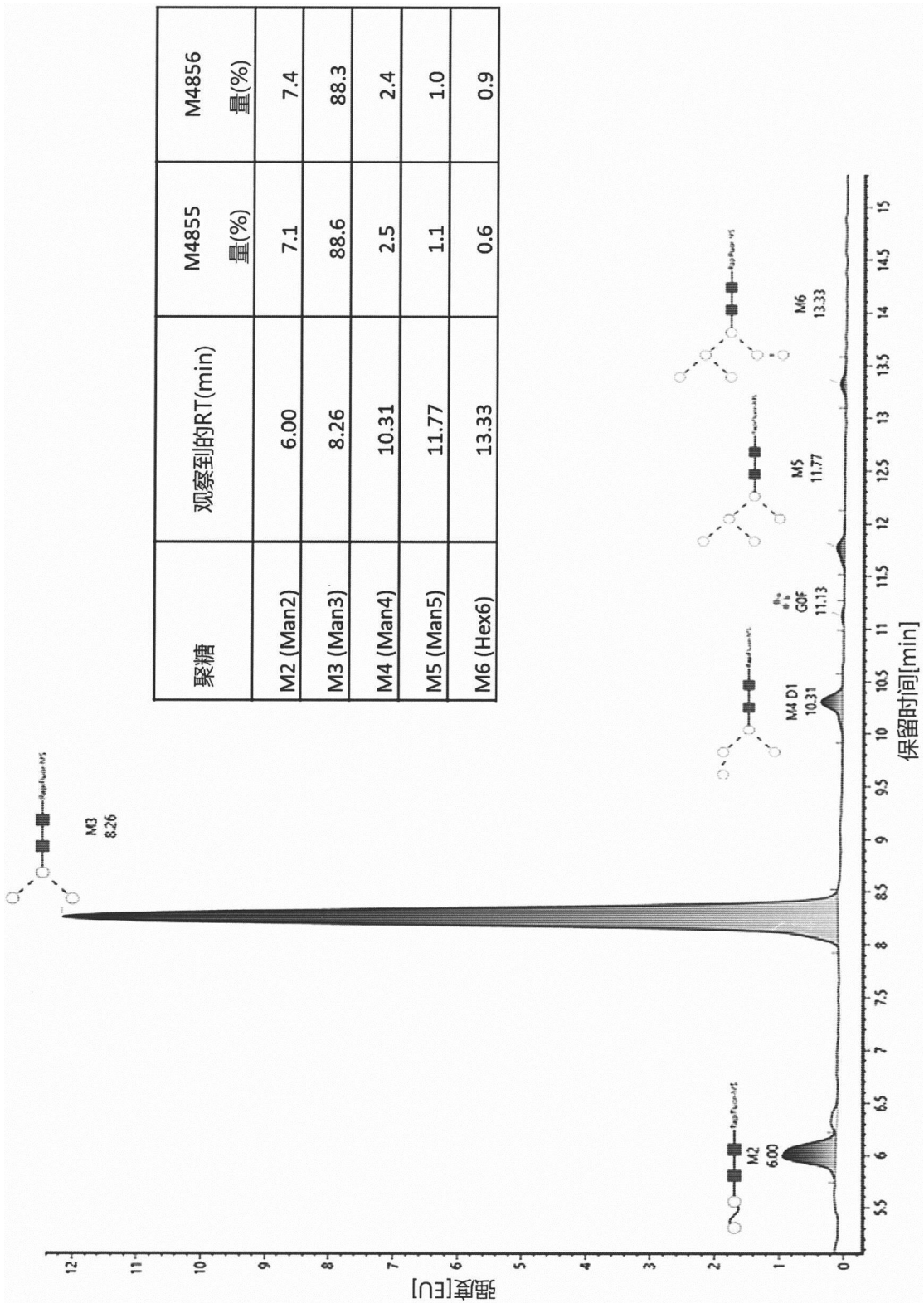


图1A

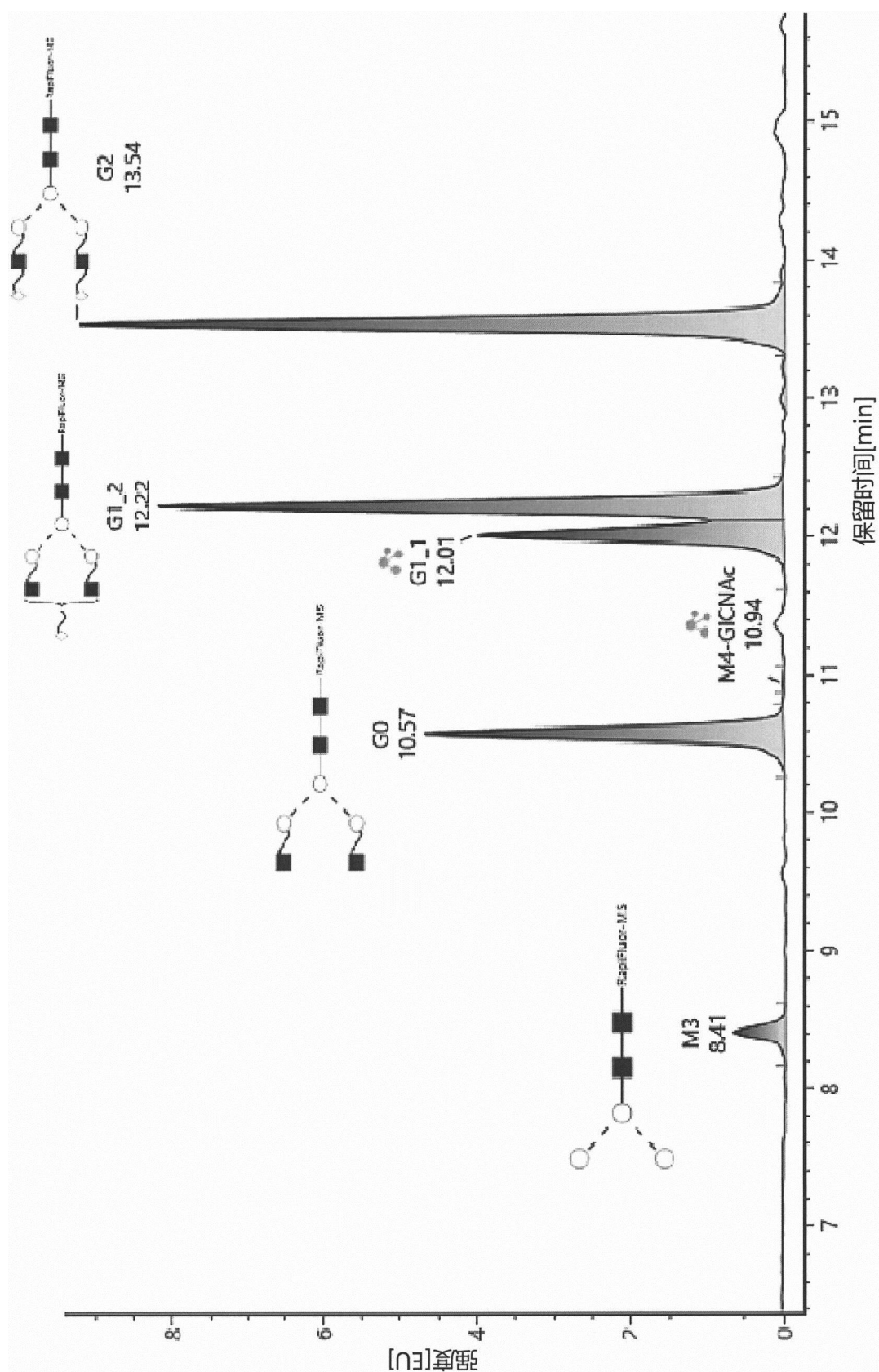


图1B

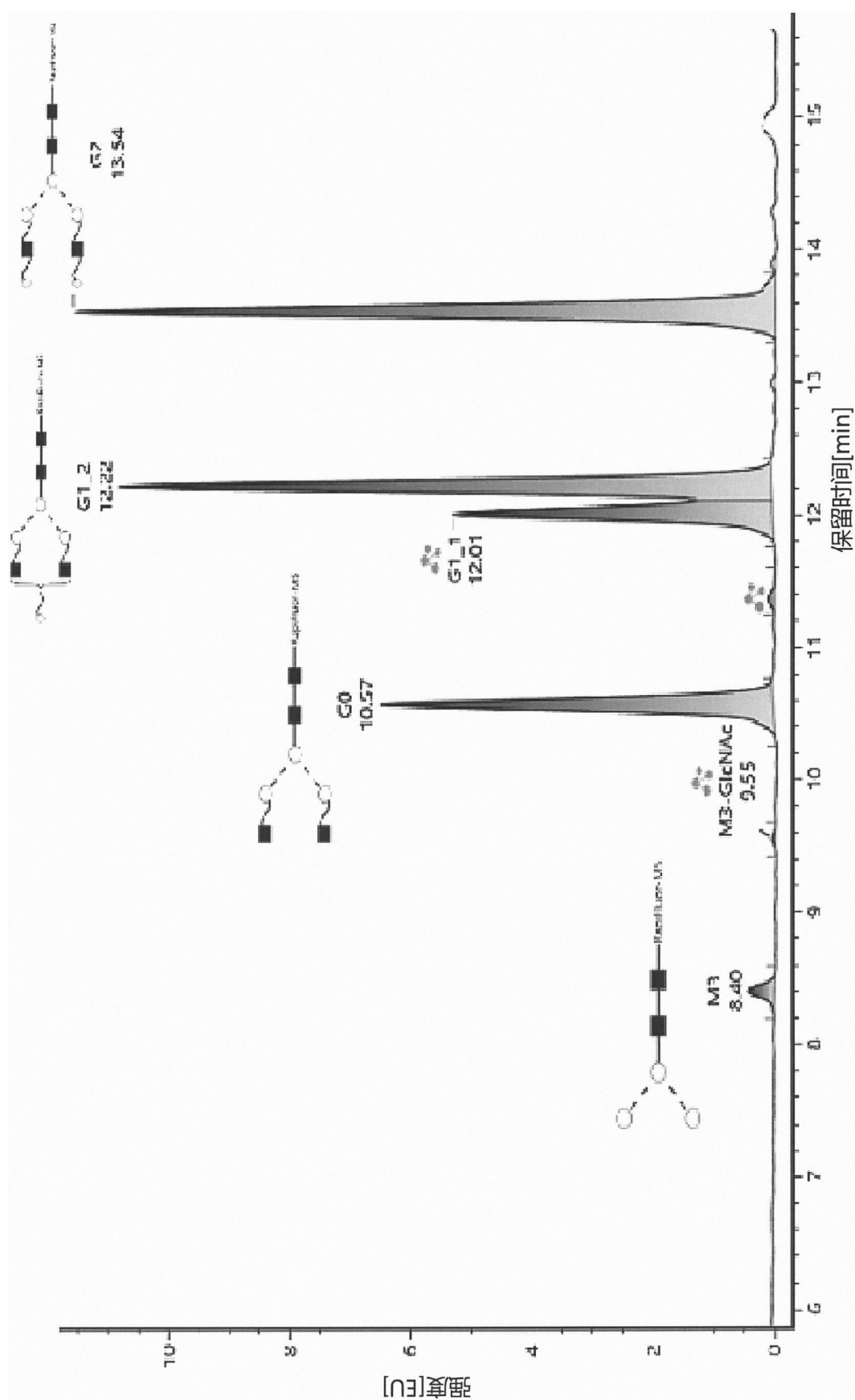


图1C

聚糖	观察到的RT(min)	M5129 量(%)	M5130 量(%)	M5131 量(%)	M5132 量(%)
M3 (Man3)	8.39	6.6	2.6	1.4	1.4
M3-GlcNAc	9.54	0.4		0.3	0.2
G0	10.56	14.6	17.4	17.0	18.6
M3_Gal_GlcNAc	11.34	5.0		0.4	0.4
G1_1	12.00	15.2	15.8	15.4	15.6
G1_2	12.21	25.1	29.4	30.2	30.2
G2	13.53	33.1	34.8	35.3	33.7
人型N-聚糖的总量 (G0至G2, %)		88.0	97.4	98.0	98.1

图1D

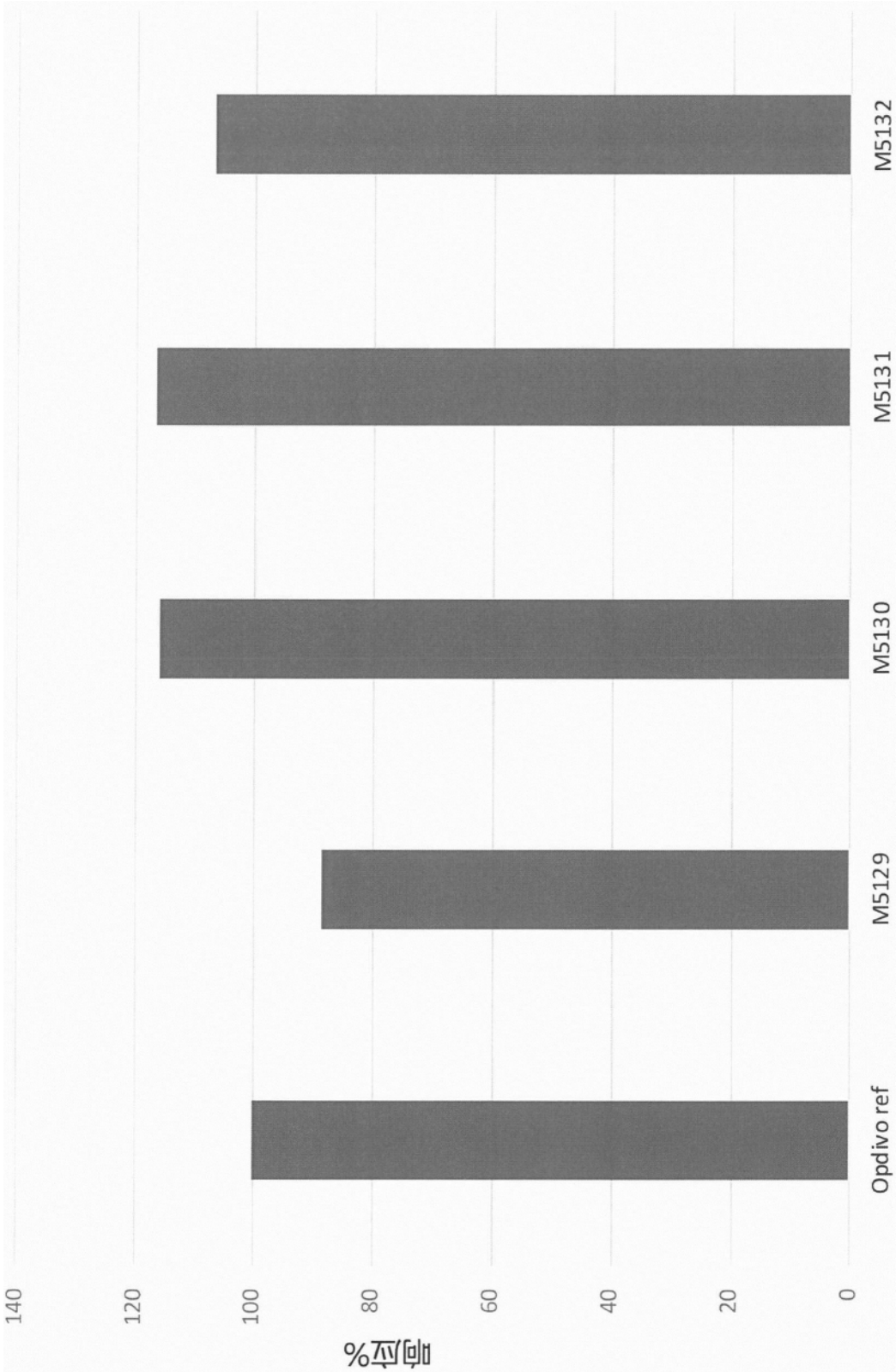


图2

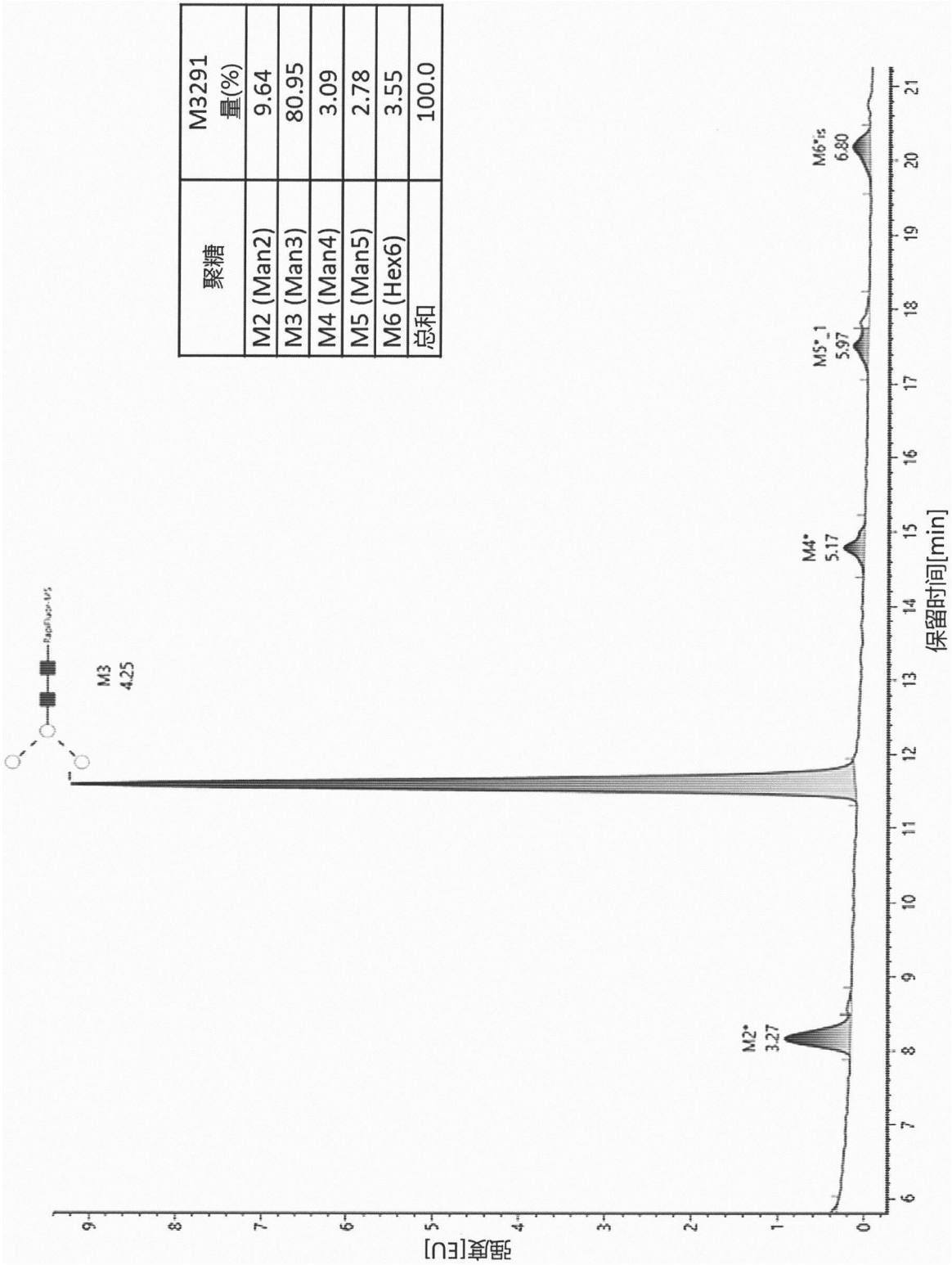


图3A

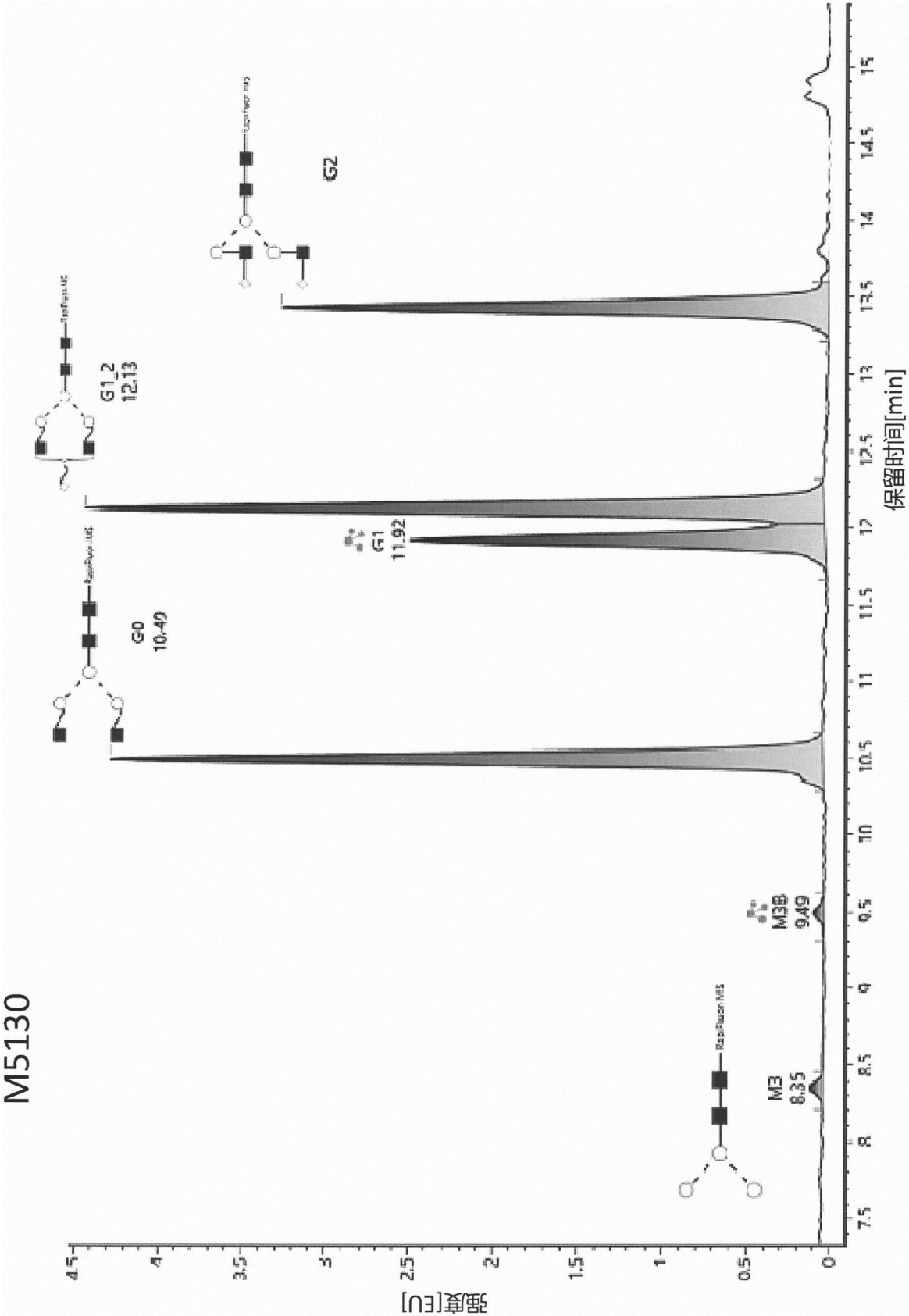


图3B

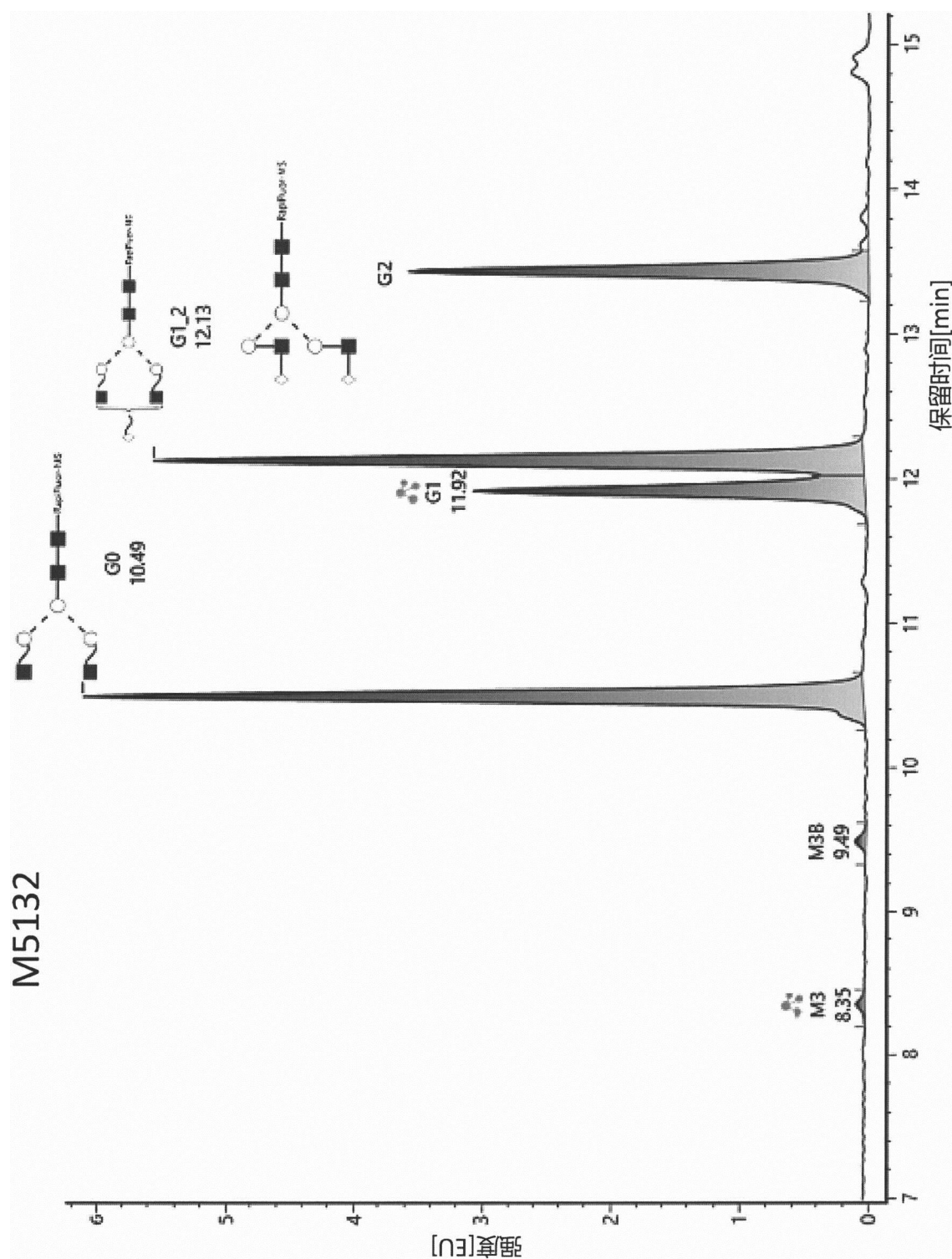


图3C

聚糖	观察到的RT(min)	M5130 量(%)	M5132 量(%)
M3 (Man3)	8.35	0.51	0.35
M3B	9.49	0.46	0.45
G0	10.49	28.67	32.54
G1	11.92	17.56	17.21
G1_2	12.13	29.77	29.51
G2	13.43	23.02	19.94
人型N-聚糖的总量 (G0至G2, %)		99.0	99.2

图3D

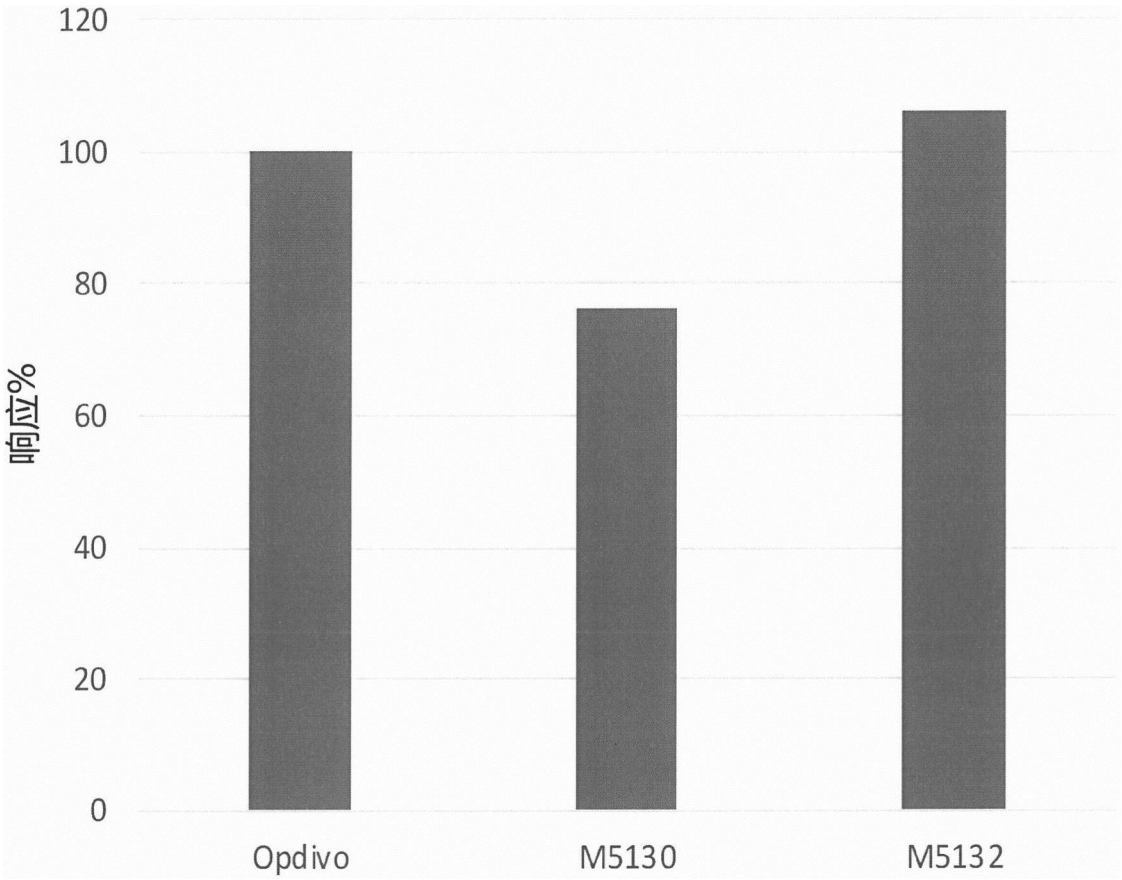


图4

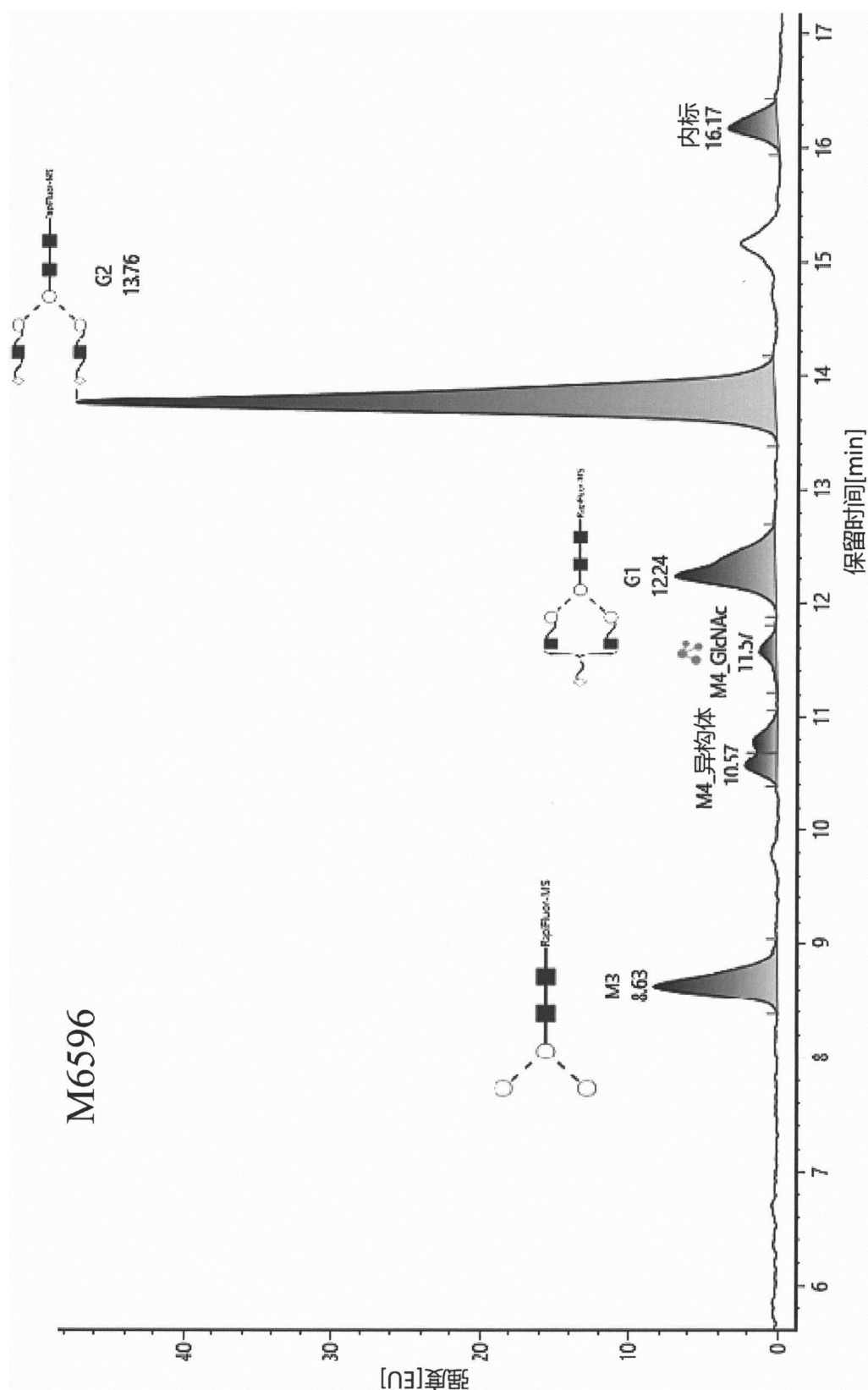


图5A

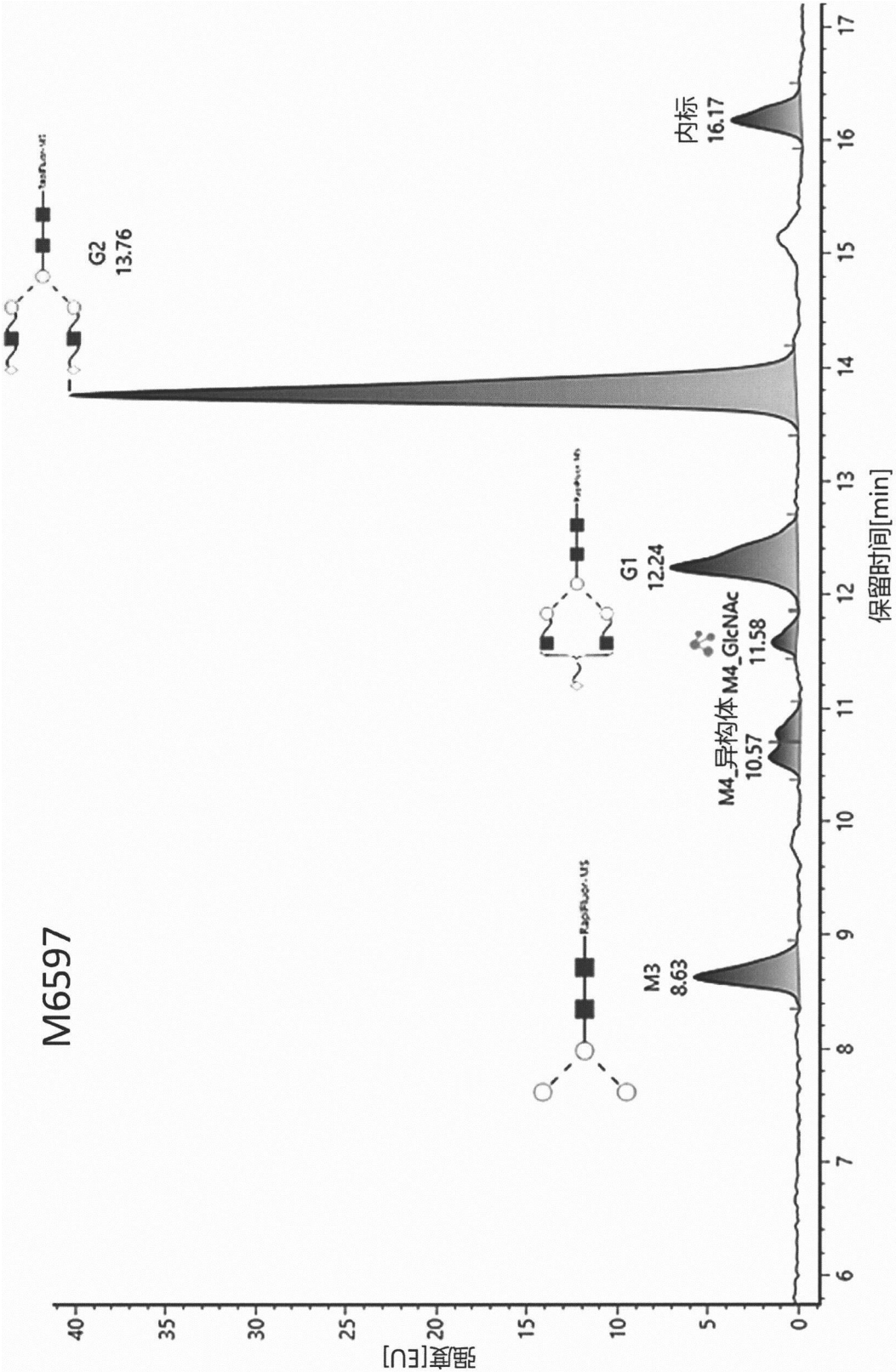


图5B

聚糖	观察到的RT (min)	M6589 量,%	M6596 量,%	M6597 量,%	M6590 量,%
M3	8,63	11,92	11,87	9,91	11,24
M4异构体	10,57	3,26	2,55	2,76	2,56
G0	10,78	2,52	2,18	2,15	2,33
M4_GlcNAc	11,58	1,62	1,57	2,08	2,46
G1	12,24	14,34	12,97	15,82	16,07
G2	13,76	66,35	68,86	67,28	65,35
人型N-聚糖(G0至G2), %		83,20	84,01	85,25	83,75

图5C