



(12)发明专利申请

(10)申请公布号 CN 107429220 A

(43)申请公布日 2017.12.01

(21)申请号 201680018548.6

(74)专利代理机构 永新专利商标代理有限公司  
72002

(22)申请日 2016.03.25

代理人 区斌

(30)优先权数据

62/139,312 2015.03.27 US

(51)Int.Cl.

*C12N 1/19*(2006.01)

(85)PCT国际申请进入国家阶段日

*C12N 15/56*(2006.01)

2017.09.26

*C12P 7/06*(2006.01)

(86)PCT国际申请的申请数据

*C12R 1/865*(2006.01)

PCT/US2016/024249 2016.03.25

(87)PCT国际申请的公布数据

W02016/160584 EN 2016.10.06

(71)申请人 嘉吉公司

地址 美国明尼苏达州

(72)发明人 C·K·米勒 A·内格雷特-雷蒙德

J·维尔多豪斯 A·瓦什

权利要求书4页 说明书29页  
序列表30页 附图3页

(54)发明名称

经葡糖淀粉酶修饰的酵母菌株和用于产生生物产物的方法

(57)摘要

本发明描述了含异源葡糖淀粉酶的基因工程化酵母以及发酵方法。所述工程化酵母可具有多个外源核酸序列,所述多个外源核酸序列各自具有不同的序列,但编码与所述酵母异源的相同或相似葡糖淀粉酶蛋白。所述工程化酵母在发酵过程中表现出期望的生物产物产量曲线。可使含淀粉材料的发酵培养基与所述工程化酵母发酵,以提供高乙醇滴度、低甘油滴度或这两者。

1. 一种工程化酵母,包含至少第一外源核酸、第二外源核酸、第三外源核酸和第四外源核酸,所述外源核酸各自包含编码与酵母属(*Saccharomyces*)菌种异源的共有葡糖淀粉酶多肽或编码彼此具有90%或更高序列同一性的两种或更多种不同葡糖淀粉酶多肽的序列,其中所述第一外源核酸、所述第二外源核酸、所述第三外源核酸和所述第四外源核酸具有彼此不同的核酸序列。

2. 根据权利要求1所述的工程化酵母,其中与所述酵母属菌种异源的所述葡糖淀粉酶多肽来自酵母或真菌菌种。

3. 根据权利要求2所述的工程化酵母,其中所述葡糖淀粉酶多肽来自选自下列的酵母或真菌生物:树脂枝孢霉(*Amorphotheca resiniae*)、黑曲霉(*Aspergillus niger*)、泡盛曲霉(*Aspergillus awamori*)、米曲霉(*Aspergillus oryzae*)、川地曲霉(*Aspergillus kawachii*)、*Aspergillus shirousami*、*Blastobotrys adenivorans*、白假丝酵母(*Candida albicans*)、米根霉(*Rhizopus oryzae*)、粟酒裂殖酵母(*Schizosaccharomyces pombe*)和扣囊复膜酵母(*Saccharomycopsis fibuligera*)。

4. 根据权利要求1所述的工程化酵母,其中所述葡糖淀粉酶多肽为来自扣囊复膜酵母的葡糖淀粉酶多肽的酶促活性部分。

5. 根据权利要求4所述的工程化酵母,其中所述葡糖淀粉酶多肽与SEQ ID NO:11具有90%或更高的序列同一性,并且任选地包含处于SEQ ID NO:11中一个或多个以下氨基酸位置处的一个或多个氨基酸取代:1M-5T、7F、9T、10A、15V、16A、18C、19V、21V、22E、24D、27N-30H、32Q、36G、38T、40A、42S-50E、52P、53A、56W、64D、71K-74K、77V、79V、86E、90F、97T、103S-106A、108V、111H、112S、114S、121V、127S-129T、131T、135V、141N、144S、145P、148D、157V、159D、160T、164A、165S、187A、189A-191H、193N、197L、199A-203G、205P、206Y、209A、210S、214W、215K、222Q、223H、225S-227H、229S-231S、242T、247A、251L、255S、257G-261S、263T-265N、267P、268P、270T-272W、274E-270A、281N、284I-286S、292S-295K、300S、302Q、304S、307G、316A、317A、319I、325D-329Y、333N、337S、341N、343L、345Y、348V、352N、355K、356I、358G、359N、361K、366V、377V、379T、386Q、393、G、395T、396F、398T、402N-411V、413E、415L、419L、420Y、422S、423F、425A、429K、431D、433S、435A、437K、440L、442L-444Y、448N-451N、453I-455S、457L、458Q、465K、467L、472D、474N、476Q、478T、480E、481I、487F-489A、492V、500S、503S、505N、507A、511L-513E和515L。

6. 根据权利要求5所述的工程化酵母,其中所述葡糖淀粉酶多肽与SEQ ID NO:11具有95%或更高、96%或更高、97%或更高、98%或更高、或者99%或更高的序列同一性。

7. 根据权利要求6所述的工程化酵母,其中所述葡糖淀粉酶多肽包含SEQ ID NO:11。

8. 根据权利要求1所述的工程化酵母,其中所述第一外源核酸、所述第二外源核酸、所述第三外源核酸和所述第四外源核酸编码彼此具有95%或更高、96%或更高、97%或更高、98%或更高、或者99%或更高序列同一性的两种或更多种不同葡糖淀粉酶多肽。

9. 根据权利要求1所述的工程化酵母,其中所述第一外源核酸、所述第二外源核酸、所述第三外源核酸和所述第四外源核酸编码各自与SEQ ID NO:11具有90%或更高序列同一性的两种或更多种不同葡糖淀粉酶多肽。

10. 根据权利要求1所述的工程化酵母,其中所述葡糖淀粉酶多肽包含与SEQ ID NO:12或SEQ ID NO:13具有90%或更高同一性的分泌信号氨基酸序列。

11. 根据权利要求10所述的工程化酵母,其中所述葡糖淀粉酶多肽包含SEQ ID NO:12或SEQ ID NO:13的分泌信号氨基酸序列。

12. 根据权利要求1所述的工程化酵母,其中所述第一外源核酸、所述第二外源核酸、所述第三外源核酸和所述第四外源核酸彼此之间的差别为至少10%。

13. 根据权利要求12所述的工程化酵母,其中所述第一外源核酸、所述第二外源核酸、所述第三外源核酸和所述第四外源核酸彼此之间的差别量在10%至30%的范围内。

14. 根据权利要求13所述的工程化酵母,其中所述第一外源核酸、所述第二外源核酸、所述第三外源核酸和所述第四外源核酸彼此之间的差别量在15%至30%的范围内。

15. 根据权利要求14所述的工程化酵母,其中所述第一外源核酸、所述第二外源核酸、所述第三外源核酸和所述第四外源核酸彼此之间的差别量在19%至23%的范围内。

16. 根据权利要求1所述的工程化酵母,其中所述第一核酸包含与SEQ ID NO:5具有85%或更高、90%或更高、95%或更高、98%或更高、99%或更高或者100%同一性的核酸。

17. 根据权利要求1所述的工程化酵母,其中所述第二核酸包含与SEQ ID NO:6具有85%或更高、90%或更高、95%或更高、98%或更高、99%或更高或者100%同一性的核酸。

18. 根据权利要求1所述的工程化酵母,其中所述第三核酸包含与SEQ ID NO:7具有85%或更高、90%或更高、95%或更高、98%或更高、99%或更高或者100%同一性的核酸。

19. 根据权利要求1所述的工程化酵母,其中所述第四核酸包含与SEQ ID NO:8具有85%或更高、90%或更高、95%或更高、98%或更高、99%或更高或者100%同一性的核酸。

20. 根据前述权利要求中任一项所述的工程化酵母,所述工程化酵母具有总共四种外源核酸,所述外源核酸各自编码所述共有葡糖淀粉酶多肽。

21. 根据前述权利要求中任一项所述的工程化酵母,其中所述第一核酸和所述第二核酸整合到所述酵母属菌种中非必需基因的第一等位基因中,并且所述第三核酸和所述第四核酸整合到所述酵母属菌种中非必需基因的第二等位基因中。

22. 根据权利要求21所述的工程化酵母,其中所述非必需基因为L-乳酸细胞色素c氧化还原酶(CYB2)。

23. 根据前述权利要求中任一项所述的工程化酵母,所述第一核酸和所述第三核酸受第一共有启动子的控制。

24. 根据前述权利要求中任一项所述的工程化酵母,所述第一核酸和所述第三核酸受第一共有启动子的控制。

25. 根据权利要求24所述的工程化酵母,其中所述第一共有启动子包含甘油醛-3-磷酸脱氢酶(TDH3)启动子核酸序列。

26. 根据前述权利要求中任一项所述的工程化酵母,其中所述第二核酸和所述第四核酸受第二共有启动子的控制。

27. 根据权利要求26所述的工程化酵母,其中所述第二共有启动子包含磷酸甘油酸激酶(PGK)启动子核酸序列。

28. 根据前述权利要求中任一项所述的工程化酵母,所述工程化酵母选自酿酒酵母(*Saccharomyces cerevisiae*)、东方伊萨酵母(*Isaatchenkia orientalis*)、产朊假丝酵母(*Candida utilis*)、树干毕赤酵母(*Pichia Stipitis*)、解脂耶氏酵母(*Yarrowia lipolytica*)、马克斯克鲁维酵母(*Kluyveromyces marxianus*)和乳酸克鲁维酵母

(*Kluyeromyces lactis*)。

29. 根据前述权利要求中任一项所述的工程化酵母,所述工程化酵母为酿酒酵母。

30. 一种工程化酵母,包含第一外源核酸、第二外源核酸、第三外源核酸和第四外源核酸,所述外源核酸各自包含编码包含SEQ ID NO:11的多肽的序列,其中所述第一核酸包含SEQ ID NO:5,所述第二核酸包含SEQ ID NO:6,所述第三核酸包含SEQ ID NO:7,并且所述第四核酸包含SEQ ID NO:8。

31. 根据前述权利要求中任一项所述的工程化酵母,所述工程化酵母与不包含所述核酸的亲本菌株相比在相同发酵条件下产生的乙醇量更高。

32. 根据前述权利要求中任一项所述的工程化酵母,所述工程化酵母与不包含所述核酸的对照菌株相比在相同发酵条件下产生的甘油更少。

33. 一种发酵培养基,包含根据前述权利要求中任一项所述的工程化酵母。

34. 根据权利要求33所述的发酵培养基,包含浓度为约90g/L或更高的乙醇。

35. 根据权利要求34所述的发酵培养基,包含浓度在90g/L至170g/L范围内的乙醇。

36. 一种用于产生生物产物的方法,包括:

使包含淀粉材料和根据权利要求1至31中任一项所述的工程化酵母的液体培养基发酵,其中所述发酵产生所述生物产物。

37. 根据权利要求36所述的方法,其中所述生物产物为乙醇,并且发酵在所述液体培养基中提供90g/L或更高的浓度。

38. 根据权利要求36或37所述的方法,其中所述淀粉材料以30至37重量%(干固体)范围内的浓度存在于所述液体培养基中,并且所述淀粉材料具有在45至65范围内的右旋糖当量。

39. 根据权利要求36至38中任一项所述的方法,其中所述发酵提供在90g/L至170g/L范围内的乙醇。

40. 根据权利要求36至39中任一项所述的方法,其中所述发酵提供在110g/L至170g/L范围内的乙醇。

41. 根据权利要求36至40中任一项所述的方法,其中所述发酵提供在125g/L至170g/L范围内的乙醇。

42. 根据权利要求36至41中任一项所述的方法,其中所述发酵提供在140g/L至170g/L范围内的乙醇。

43. 根据权利要求36至42中任一项所述的方法,其中所述微生物与不包含所述核酸的对照酵母相比产生的甘油更少。

44. 根据权利要求43所述的方法,其中所述生物产物为乙醇,并且所述工程化酵母的乙醇与甘油之比(mol:mol)大于所述对照酵母的乙醇与甘油之比。

45. 根据权利要求44所述的方法,其中所述工程化酵母的乙醇与甘油之比大于13:1或大于14:1。

46. 根据权利要求45所述的方法,其中所述乙醇与甘油之比在14:1至25:1的范围内。

47. 根据权利要求36所述的方法,其中所述生物产物为乳酸。

48. 一种核酸,所述核酸与SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:7或SEQ ID NO:8具有85%或更高、90%或更高、95%或更高、98%或更高、或者99%或更高的同一性。

49. 根据权利要求48所述的核酸,还包含操作地连接至所述序列的启动子、终止子或这两者。

50. 一种载体,包含根据权利要求48或49所述的核酸。

51. 一种用于产生乙醇的发酵方法,包括:

提供包含部分水解淀粉和工程化酵母的发酵培养基,其中在整个发酵周期中向所述发酵培养基添加包含水解淀粉的液体培养基,其中所述工程化酵母包含至少一种提供分泌到所述培养基中的异源葡糖淀粉酶的基因修饰,所述至少一种基因修饰在所述发酵周期内是稳定的,其中所述发酵周期为五天或更长的时间段,并且其中所述异源葡糖淀粉酶有利于产生在所述发酵培养基中浓度达130g/L或更高的乙醇。

52. 根据权利要求51所述的发酵方法,其中所述部分水解淀粉包括部分水解的玉米淀粉。

## 经葡糖淀粉酶修饰的酵母菌株和用于产生生物产物的方法

[0001] 相关申请的交叉引用

[0002] 本申请要求2015年3月27日提交的标题为Glucoamylase-Modified Yeast Strains and Methods for Bioproduct Production (经葡糖淀粉酶修饰的酵母菌株和用于产生生物产物的方法)的共同拥有的美国临时申请序列号62/139,312的权益,该申请全文以引用方式并入本文。创建于2015年3月26日,标题为“CAR0196P1\_Sequence\_Listing\_ST25.txt”并且具有79千字节大小的ASCII文本文件全文以引用方式并入本文。

### 技术领域

[0003] 本发明涉及具有多个编码异源葡糖淀粉酶基因的核酸的经基因修饰的酵母、用于产生生物产物诸如乙醇并且还减少甘油产生的发酵方法。

### 背景技术

[0004] 通过发酵产生乙醇是众所周知的工业方法。然而,增加乙醇产率在技术上可能是困难的。有许多因素使得微生物在设计用于增加乙醇产量的发酵条件下生长具有挑战性。例如,发酵培养基可具有更高的底物浓度以促进乙醇产生,但是这些条件可能对细胞生长具有负面影响。另外,增加的乙醇浓度和不期望的副产物的积聚也可能不利于细胞健康。酵母菌株经选择为对这些条件具有耐受性,这可使得乙醇产率提高。具体地,酿酒酵母(*Saccharomyces cerevisiae*)的乙醇耐受性菌株已在工业环境中作为用于产生乙醇的主要微生物。

[0005] 分子技术已导致对乙醇耐受性相关基因的鉴定。例如,Kajiwara (Appl.Microbiol.Biotechnol.2000;53:568-74. (《应用微生物学与生物技术》,2000年;第53卷:第568-574页))报道了不饱和脂肪酸(UFA)合成中涉及的OLE1基因的过表达导致细胞中的不饱和脂肪酸水平提高,并且乙醇产量提高。其它研究已发现,因海藻糖水解酶酸性海藻糖酶(ATH)的破坏而引起的海藻糖积聚(Kim et al.,Appl.Environ.Microbiol.1996;62:1563-1569 (Kim等人,《应用与环境微生物学》,1996年;第62卷:第1563-1569页))或由携带PRO1  $\gamma$ -谷氨酰激酶突变的菌株导致的脯氨酸L-脯氨酸的积聚(Takagi,et al.,Appl.Environ.Microbiol.2005;71:8656-8662. (Takagi等人,《应用与环境微生物学》,2005年;第71卷:第8656-8662页))增强酵母的乙醇耐受性。麦角固醇与酿酒酵母的乙醇耐受性密切相关(Inoue,et al.,Biosci.Biotechnol Biochem.2000;64:229-236 (Inoue等人,《生物科学、生物技术与生物化学》,2000年;第64卷:第229-236页))。虽然在这一领域已取得了进展,但是单独使用显示出乙醇耐受性的经基因修饰的菌株可能不足以在发酵过程中提供所需水平的乙醇。

[0006] 除了发酵微生物的遗传特征图外,发酵培养基的组分可能对乙醇产量具有重大影响。在发酵过程中,培养基中存在碳水化合物或碳水化合物混合物。淀粉是被广泛使用的廉价碳水化合物源。它可以从各种各样的植物来源诸如玉米、小麦、大米、大麦等获得。许多生物不能直接代谢淀粉,也不能缓慢而低效地代谢淀粉。

[0007] 因此,通常在将淀粉供料到发酵过程中之前对其进行处理,以便将淀粉分解成单糖,使生物可以容易地发酵。通常,淀粉被水解以形成主要含有葡萄糖(即,右旋糖)的混合物。然而,在发酵准备中对淀粉组合物的预处理可能昂贵且费力,因为其通常涉及向淀粉材料中添加纯化的淀粉降解酶,并且在进行发酵之前需要额外的步骤。此外,对葡萄糖的完全水解增加了大量成本,因此大多数可商购获得的葡萄糖产品往往含有少量的各种低聚多糖。

[0008] 产生淀粉基乙醇的成本的很大一部分用于将淀粉分解成可发酵糖的酶。已对酿酒酵母尝试了多种分子技术以减少或消除在发酵培养基中添加淀粉分解酶的需要,但这些方法的成功程度各有不同。影响工程化菌株的商业可行性的潜在限制因素是酿酒酵母分泌大量异种蛋白质的能力。

## 发明内容

[0009] 本发明涉及用于产生生物产物的发酵方法,该生物产物包括具有异源葡糖淀粉酶的工程化酵母。在一个方面,本发明涉及表达异源葡糖淀粉酶或者彼此具有高度序列同一性的两种或更多种异源葡糖淀粉酶的工程化酵母菌株,其中所述一种或多种葡糖淀粉酶由彼此不同的多种外源核酸编码。工程化酵母菌株表达异源葡糖淀粉酶,并且能够在发酵产品中产生高水平的生物产物,诸如乙醇。

[0010] 在一个方面,本发明提供了工程化酵母,该工程化酵母包含至少第一外源核酸、第二外源核酸、第三外源核酸和第四外源核酸。在一些情况下,每种外源核酸可包含编码与酵母菌种异源的共有葡糖淀粉酶多肽的序列,并且第一外源核酸、第二外源核酸、第三外源核酸和第四外源核酸具有彼此不同的核酸序列。在其它情况下,第一外源核酸、第二外源核酸、第三外源核酸和第四外源核酸是不同的核酸序列,并编码与酵母菌种异源且彼此序列同一性大于90%的两种或更多种葡糖淀粉酶多肽。任选地,微生物可包含额外的核酸(第五核酸、第六核酸等),所述额外的核酸编码共有葡糖淀粉酶多肽或者两种或更多种具有高度同一性的葡糖淀粉酶多肽。在一些方面,工程化酵母是酵母属(*Saccharomyces*)菌种。

[0011] 在示例性方面,一种或多种葡糖淀粉酶多肽与扣囊复膜酵母(*Saccharomycopsis fibuligera*)菌株的葡糖淀粉酶中的SEQ ID NO:11具有至少90%同一性。在一些方面,第一外源核酸、第二外源核酸、第三外源核酸和第四外源核酸彼此之间的差别为至少10%,例如,差别量在10%至30%的范围内。在一些方面,第一外源核酸、第二外源核酸、第三外源核酸和第四外源核酸是独特的密码子优化序列,其编码葡糖淀粉酶并在宿主细胞中提供高效的核酸表达。

[0012] 在一些方面,第一核酸和第二核酸整合到酵母中非必需基因(诸如酵母属菌种中的CYB2)的第一等位基因中,并且第三核酸和第四核酸整合到酵母中非必需基因(诸如酵母属菌种中的CYB2)的第二等位基因中。在一些方面,另外两组核酸(例如,第一核酸和第三核酸,和/或第二核酸和第四核酸)受共有启动子(诸如,TDH3或PGK启动子)的控制。

[0013] 在一些方面,本发明提供了用于产生生物产物的方法,该方法包括使包含淀粉材料和工程化酵母的液体培养基发酵,该工程化酵母包含至少第一外源核酸、第二外源核酸、第三外源核酸和第四外源核酸。在一些方面,生物产物为乙醇,并且发酵方法在液体培养基中提供约90g/L或更高的乙醇浓度,诸如其量在约90g/L至约170g/L的范围内。在一些方面,

本发明提供了包含工程化微生物的发酵培养基,该工程化微生物具有本文所述的任一基因修饰。

[0014] 在一些方面,本发明提供了工程化酵母和用于发酵的方法,其中细胞在发酵方法中产生比亲本菌株更少的甘油。所述工程化酵母可包含本公开的一个或多个基因修饰,所述基因修饰使得细胞与不包含一个或多个基因修饰的亲本菌株相比产生更少的甘油。

[0015] 在一些方面,产生生物产物的方法还包括减少发酵方法中产生的甘油量,其中该方法包括培养具有本公开的一个或多个基因修饰的微生物,该微生物与不包含基因修饰的亲本菌株相比产生更少的甘油。

[0016] 在一些方面,本发明提供与以下序列中的任一个具有85%或更高序列同一性的核酸:SEQ ID NO:5-8。本发明还提供了核酸以及包含该核酸的载体,该核酸具有操作地连接至核酸序列的启动子、终止子或这两者。

[0017] 本发明还提供了工程化微生物,该工程化微生物为酵母属菌种,其包含至少第一外源核酸、第二外源核酸、第三外源核酸和第四外源核酸,所述核酸编码与酵母属菌种异源的葡糖淀粉酶多肽,其中葡糖淀粉酶多肽具有经修饰的N末端,该N末端含有酿酒酵母 $\alpha$ 交配因子2 (ScMfa2) 分泌信号或酿酒酵母可抑制酸性磷酸酶 (ScPho5) 分泌信号。

[0018] 在另一方面,本发明提供用于产生乙醇的发酵方法。在该方法中,提供了包含部分水解淀粉(诸如,部分水解的玉米淀粉)和工程化酵母的发酵培养基。在整个发酵周期中向发酵培养基添加包含水解淀粉的液体培养基。工程化酵母包含至少一种提供分泌到培养基中的异源葡糖淀粉酶的基因修饰,所述至少一种基因修饰在发酵周期内是稳定的。发酵周期为五天或更长的时间段。在发酵方法中,异源葡糖淀粉酶有利于产生在发酵培养基中浓度达130g/L或更高的乙醇。

[0019] 在另一方面,本发明提供可用于制备饲料组合物的方法和组合物。作为由发酵培养基获得的发酵培养基联产品的饲料组合物包含本公开的工程化酵母。例如,在完成发酵过程之后,可以从发酵培养基中去除生物产物中的一些或全部,从而得到包含非生物产物固体的精制组合物。该非生物产物固体可包含工程化酵母、培养基中未被酵母利用的原料以及发酵副产物。精制组合物可用于形成饲料组合物,例如牲畜饲料组合物。包含非生物产物固体的精制组合物可以提供碳水化合物和蛋白质补充,以改善饲料组合物的营养含量。

## 附图说明

[0020] 图1为工程化CYB2基因座的示意图,其中示出编码Sf GA和标记基因(菌株1-5)的四个独特密码子优化核酸序列的排列方式。

[0021] 图2为在发酵培养基中添加和不添加商用葡糖淀粉酶的情况下,与(未工程化)亲本菌株相比,具有编码Sf GA的四个独特密码子优化核酸序列的乙醇耐受性菌株的乙醇产量的图表。

[0022] 图3为示出与具有相同Sf GA序列但在天然基因座处恢复URA3的菌株相比,具有编码Sf GA的四个独特密码子优化核酸序列的乙醇耐受性菌株的乙醇产量的图表。

## 具体实施方式

[0023] 下文所述的本发明的方面并非旨在为穷举性的或将本发明限制为以下详细描述



中所公开的确切形式。相反,所选择和所描述的方面的目的在于可促进本领域其它技术人员对本发明的原理和实践的了解和理解。

[0024] 本发明的方面涉及表达异源葡糖淀粉酶或者两种或更多种异源葡糖淀粉酶的工程化酵母菌株,其中葡糖淀粉酶由彼此不同的多种外源核酸编码。该工程化酵母菌株能够在发酵过程中产生高水平的生物产物,诸如乙醇。此外,该工程化酵母菌株能够在发酵过程中产生较低水平的甘油。

[0025] 葡糖淀粉酶可以从细胞分泌到发酵培养基中,在此处,酶可以对存在于发酵培养基中的葡萄糖聚合物发挥淀粉分解活性。继而,该酶可以引起葡萄糖聚合物降解为葡萄糖,葡萄糖可进入细胞并作为碳源用于产生目标化合物,诸如乙醇。

[0026] 用于表达这些酶的模板的核酸以及包含这些核酸的构建体也是本发明的方面。

[0027] 如本文所用的术语“外源”意指将诸如核酸的分子或诸如酶活性的活性引入宿主生物中。外源核酸可通过熟知的技术引入宿主生物中,并且可以保持在宿主染色体物质的外部(例如,保持在非整合载体上),或者可以整合到宿主的染色体中,诸如,通过重组事件来实现。外源核酸可以编码与宿主生物同源或异源的酶或其部分。

[0028] 术语“异源”(例如,“非天然”)是指来自不同于所引用的分子或生物的来源的分子或活性。因此,与所引用的生物异源的基因或蛋白质是在该生物中未发现的基因或蛋白质。在本公开的上下文中,“异源葡糖淀粉酶”是指与宿主生物中天然存在的葡糖淀粉酶多肽不同的葡糖淀粉酶多肽。例如,存在于第一真菌菌种中并且外源引入作为宿主生物的第二真菌菌种中的特异性葡糖淀粉酶基因与该第二真菌生物“异源”。

[0029] 葡糖淀粉酶(E.C.3.2.1.3)是从D-葡萄糖释放的寡糖链和多糖链的非还原末端依次水解1,4-连接的 $\alpha$ -D-葡萄糖基残基的淀粉分解酶。葡糖淀粉酶还可以裂解支链淀粉分支点的 $\alpha$ -1,6键。如本文所用,参考异源葡糖淀粉酶的术语“淀粉分解活性”涉及这些酶催化机理。葡糖淀粉酶多肽可以是天然存在的葡糖淀粉酶的变体或天然存在的葡糖淀粉酶的一部分(诸如,在其N末端、其C末端或这两者处被截短的葡糖淀粉酶),其中葡糖淀粉酶多肽保持淀粉分解活性。

[0030] 葡糖淀粉酶的别称包括淀粉转葡糖苷酶、 $\gamma$ -淀粉酶、溶酶体 $\alpha$ -葡糖苷酶、酸性麦芽糖酶、外切-1,4- $\alpha$ -葡糖苷酶、葡萄糖淀粉酶、 $\gamma$ -1,4-葡聚糖葡萄糖水解酶、酸性麦芽糖酶、1,4- $\alpha$ -D-葡聚糖葡萄糖水解酶。

[0031] 大多数葡糖淀粉酶是多域酶。许多葡糖淀粉酶包含淀粉结合域,该淀粉结合域经由O-糖基化接头区连接至催化域。淀粉结合域可以折叠为反平行 $\beta$ 桶形,并且可具有针对淀粉或 $\beta$ -环糊精的两个结合位点。然而,一些葡糖淀粉酶不包含淀粉结合域(例如,参见 Hostinova et al., Archives of Biochemistry and Biophysics, 411:189-195, 2003 (Hostinova等人,《生物化学与生物物理学集刊》,第411卷:第189-195页,2003年)),或包含非经典淀粉结合域。例如,米根霉(*Rhizopus oryzae*)葡糖淀粉酶具有N末端生淀粉结合域,并且扣囊复膜酵母 IFO 0111 葡糖淀粉酶缺少明确的淀粉结合域(Hostinova等人,出处同上)。

[0032] 葡糖淀粉酶还可具有催化域,该催化域具有扭转的( $\alpha/\alpha$ )(6)-桶形,并具有中心漏斗形活性位点。葡糖淀粉酶可具有结构保守的大约450个残基的催化域。在一些葡糖淀粉酶中,催化域之后通常是由连接至大约100个残基的淀粉结合域的30和80个残基组成的接头

区。

[0033] 葡糖淀粉酶性质可能与其结构特征相关。基于结构的多序列比对使用来自催化和淀粉结合域模型的信息进行构建(参见例如Coutinho,P.M.,and Reilly,P.J.,1994.*Protein Eng.*7:393-400 and 749-760(Coutinho,P.M.和Reilly,P.J.,1994年,《蛋白质工程》,第7卷:第393-400页和第749-760页))。已证实,基于结构-功能关系研究,催化和淀粉结合域在功能上是独立的,并且在微生物葡糖淀粉酶中存在结构相似性。通过其它研究,特异性葡糖淀粉酶残基已示出与直接蛋白质构型改变、底物结合、热稳定性和催化活性相关(参见例如Sierks,M.R.,et al.1993.*Protein Eng.*6:75-79(Sierks,M.R.等人,1993年,《蛋白质工程》,第6卷:第75-79页);和Sierks,M.R.,and Svensson,B.1993.*Biochemistry* 32:1113-1117(Sierks,M.R.和Svensson,B.,1993年,《生物化学》,第32卷:第1113-1117页))。因此,葡糖淀粉酶序列和蛋白质功能之间的相关性在本领域中得到理解,并且技术人员可以设计和表达具有一个或多个氨基酸缺失、取代和/或添加的淀粉分解活性葡糖淀粉酶的变体。

[0034] 在本公开的优选方面,酵母属宿主至少具有第一外源核酸、第二外源核酸、第三外源核酸和第四外源核酸(并且它们彼此不同),所述外源核酸各自包含编码至少一种葡糖淀粉酶多肽的序列,这些序列与扣囊复膜酵母菌株IFO 0111的G1m葡糖淀粉酶具有至少90%的同源性。Hostinova等人(*Archives of Biochemistry and Biophysics*,411:189-195,2003(《生物化学与生物物理学集刊》,第411卷:第189-195页,2003年))描述了扣囊复膜酵母菌株IFO 0111中葡糖淀粉酶基因G1m(即,“Sf GA”多肽)的核苷酸序列。根据Hostinova等人所述,扣囊复膜酵母G1m基因被转录为1.7kb RNA转录物,该基因编码515个氨基酸的蛋白质并且由SEQ ID NO:11表示。在长度为515个氨基酸的多肽链中,26个N末端氨基酸残基构成信号肽,并且随后的489个氨基酸残基构成成熟蛋白质。成熟G1m缺少信号序列且长度为489个氨基酸,其去糖基化形式的预测分子量为54,590Da。在与其它葡糖淀粉酶的比对中,G1m被示出具有催化域同源性(例如,参见SEQ ID NO:1的第239位和第493位的谷氨酸)。

[0035] Itoh等人(*J.Bacteriol.*169:4171-4176(《细菌学杂志》,第169卷:第4171-4176页))描述了扣囊复膜酵母中的另一种葡糖淀粉酶基因GLU1的核苷酸序列。扣囊复膜酵母GLU1基因被转录为2.1kb RNA转录物,该基因编码519个氨基酸的蛋白质并且分子量为57,000Da。GLU1具有四个潜在糖基化位点(针对分子量为2000Da的天冬酰胺连接的糖苷)。GLU1具有四个潜在糖基化位点(针对分子量为2000Da的天冬酰胺连接的糖苷)。GLU1具有用于分泌的天然信号序列,其可能在蛋白质输出期间被切除。切除位点之前是碱性氨基酸Lys-Arg,所述氨基酸被认为是用于生成成熟蛋白质的蛋白分解加工信号。

[0036] Itoh等人(出处同上)还描述了来自酵母和真菌的葡糖淀粉酶的氨基酸序列的比对。对扣囊复膜酵母(*Saccharomycopsis fibuligera*)、黑曲霉(*A.niger*)、米根霉(*Rhizopus oryzae*)和糖化酵母(*Saccharomyces diastaticus*)和酿酒酵母(*Saccharomyces cerevisiae*)进行了比对,显示出五个高度同源的片段(S1-S5)。这些相应保守片段部分经表明在构象上彼此相似。通常位于羧基末端的S5片段似乎对于淀粉分解活性并非关键,因为来自酵母属菌种的葡糖淀粉酶缺乏该区域。

[0037] 就这一点而言,本发明还设想了具有葡糖淀粉酶活性的Sf GA的变体和部分。表1和表2示出了各种真菌和细菌葡糖淀粉酶基因的列表,包括天然信号序列的氨基酸位置,以

及一些序列中的葡糖淀粉酶多肽的前肽。

[0038] 表1:真菌葡糖淀粉酶

名称	入藏号	生物	信号肽	前肽	链
GAMP ( AMYG_AMORE )	Q03045	树脂枝孢霉 ( <i>Amorphotheca resinae</i> ) ( 木焦油真菌 ) ( 树脂枝孢霉 ( <i>Hormoconis resinae</i> ) )	1-29		30-616
GLAA ( AMYG_ASPNG )	P69328	黑曲霉 ( <i>Aspergillus niger</i> )	1-18	19-24	25-640
STA1 ( AMYH_YEASX )	P04065	酿酒酵母 ( <i>Saccharomyces cerevisiae</i> )	1-21		22-767
STA2 ( AMYI_YEASX )	P29760	酿酒酵母 ( <i>Saccharomyces cerevisiae</i> )	1-21		22-768
GLAA ( AMYG_ASPAW )	P69327	泡盛曲霉 ( <i>Aspergillus awamori</i> ) ( 黑曲霉菌 )	1-18	19-24	25-640

glaA ( AMYG_ASPOR )	P36914	米曲霉 ( <i>Aspergillus oryzae</i> ) ( 菌株 ATCC 42149/RIB 40 ) ( 黄曲霉菌 )	1-19	20-25	26-612
GAA ( AMYG_BLAAD )	P42042	<i>Blastobotrys adeninivorans</i> ( 酵母 ) ( <i>Arxula adeninivorans</i> )	1-18		19-624
GAMI ( AMYG_SCHOC )	P22861	西方许旺酵母 ( <i>Schwanniomyces occidentalis</i> ) ( 酵母 ) ( 德巴利酵母 ( <i>Debaryomyces occidentalis</i> ) )	1-22		23-958
gal ( AMYG_ASPKA )	P23176	川地曲霉 ( <i>Aspergillus kawachii</i> ) ( 白曲霉菌 ) ( 泡盛曲霉川地变种 ( <i>Aspergillus awamori</i> var. <i>kawachi</i> ) )	1-18	19-24	25-639
glaA ( AMYG_ASPSH )	P22832	<i>Aspergillus shirousami</i>	1-18	19-24	25-639
GAMI ( AMYG_CANAL )	O74254	白假丝酵母 ( <i>Candida albicans</i> ) ( 菌株 SC5314/ATCC MYA-2876 )	1-20		21-946
AMYG_RHIOR	P07683	米根霉 ( <i>Rhizopus oryzae</i> ) ( 毛霉菌病病因 ) ( 少根根霉戴尔变种 ( <i>Rhizopus arrhizus</i> var. <i>delemar</i> ) )	1-25		26-604
meu17 ( mAMYG_SCHPO )	O60087	粟酒裂殖酵母 ( <i>Schizosaccharomyces pombe</i> ) ( 菌株 972/ATCC 24843 ) ( 分裂酵母 )	1-16	17-28	29-450
	I2K2N7	布鲁塞尔酒香酵母 ( <i>Brettanomyces bruxellensis</i> ) AWRI1499	1-21		22-575
SGA1	A0A0H5C316	杰丁塞伯林德纳氏酵母 ( <i>Cyberlindnera jadinii</i> ) ( 圆酵母 ) ( 杰丁毕赤酵母 ( <i>Pichia jadinii</i> ) )	1-16		17-577
GLA1 ( AMYH_SACFI )	P26989	扣囊复膜酵母 ( <i>Saccharomycopsis</i>	1-27		28-519

[0040]

		<i>fibuligera</i> ) ( Hostinova 等人, 2001 年 )			
[0041]	GLUI AMYG_SACFI	P08017.1 扣囊复膜酵母 ( <i>Saccharomycopsis fibuligera</i> ) ( Itoh 等人, 1987 年 )	1-27		28-519
	Glm SEQ ID NO:11	CAC83969 扣囊复膜酵母 ( <i>Saccharomycopsis fibuligera</i> ) IFO 0111 ( Hostinova 等人, 2003 年 )	1-26		27-515

[0042] 表2:细菌葡糖淀粉酶

淀粉酶基因	入藏号	生物	信号肽	前肽	链
[0043]	SusB ( SUSB_BACTN )	G8JZS4 多形拟杆菌 ( <i>Bacteroides thetaiotaomicron</i> ) ( 菌株 ATCC 29148/DSM 2079/NCTC 10582/E50/VPI-5482 )	1-21		22-738
	cga ( AMYG_CLOS0 )	P29761 梭菌属 ( <i>Clostridium sp.</i> ) ( 菌株 G0005 )	1-21		22-702

[0044] 功能性葡糖淀粉酶的任何多肽序列可以由四个或更多个不同的葡糖淀粉酶核酸序列表达。在一些方面,葡糖淀粉酶核酸序列可以编码葡糖淀粉酶多肽的一个或多个天然(“野生型”)序列(即,葡糖淀粉酶多肽序列与任何氨基酸位置上的天然序列并无不同)。在其它方面,葡糖淀粉酶多肽的序列与天然序列在一个或多个氨基酸位置处存在差异(例如,天然葡糖淀粉酶多肽的“变体”)。这些差异可以是例如(a)从野生型序列中去除一个或多个氨基酸、(b)向野生型序列中添加一个或多个氨基酸、(c)野生型序列的取代、(a)和(c)的组合、或(b)和(c)的组合。

[0045] 在一些方面,被引入酵母属菌种中的葡糖淀粉酶核酸编码一种(即,共有葡糖淀粉酶)或多种葡糖淀粉酶多肽,所述葡糖淀粉酶多肽与SEQ ID NO:11(来自扣囊复膜酵母 IFO 0111的Glm葡糖淀粉酶)具有约90%或更高的序列同一性。在更具体的方面,被引入酵母属菌种中的葡糖淀粉酶核酸编码一种或多种葡糖淀粉酶多肽,所述葡糖淀粉酶多肽与SEQ ID NO:11具有95%或更高、96%或更高、97%或更高、98%或更高、或者99%或更高的序列同一性。与SEQ ID NO:11的同一性差异可归因于多肽的一个或多个区域中的一个或多个氨基酸取代,例如在被认为对酶活性重要的那些区域之外和/或基于与一种或多种其它葡糖淀粉酶的比对的保守区域之外的区域。

[0046] 例如,参考SEQ ID NO:11,在一个方面,所述一个或多个取代处于以下位置中的任一位置处,所述任一位置表示SEQ ID NO:11与其它葡糖淀粉酶之间具有较低同一性程度的氨基酸位置。例如,SEQ ID NO:11与其它葡糖淀粉酶的比对在Hostinova和Gasperik (Biologia 65:559-568;2010(《生物学》,第65卷:第559-568页;2010年))提供的主要结构比对中示出。根据SEQ ID NO:11的取代位置为:1M-5T、7F、9T、10A、15V、16A、18C、19V、21V、22E、24D、27N-30H、32Q、36G、38T、40A、42S-50E、52P、53A、56W、64D、71K-74K、77V、79V、86E、

90F、97T、103S-106A、108V、111H、112S、114S、121V、127S-129T、131T、135V、141N、144S、145P、148D、157V、159D、160T、164A、165S、187A、189A-191H、193N、197L、199A-203G、205P、206Y、209A、210S、214W、215K、222Q、223H、225S-227H、229S-231S、242T、247A、251L、255S、257G-261S、263T-265N、267P、268P、270T-272W、274E-270A、281N、284I-286S、292S-295K、300S、302Q、304S、307G、316A、317A、319I、325D-329Y、333N、337S、341N、343L、345Y、348V、352N、355K、356I、358G、359N、361K、366V、377V、379T、386Q、393、G、395T、396F、398T、402N-411V、413E、415L、419L、420Y、422S、423F、425A、429K、431D、433S、435A、437K、440L、442L-444Y、448N-451N、453I-455S、457L、458Q、465K、467L、472D、474N、476Q、478T、480E、481I、487F-489A、492V、500S、503S、505N、507A、511L-513E和515L。

[0047] 如果葡糖淀粉酶多肽序列与野生型序列不同,则葡糖淀粉酶多肽可包含例如参考SEQ ID NO:11所列出的低同一性位置中任一位置处的最多约5、10、15、20、25、30、35、40、45或50个氨基酸取代。在一些情况下,所述取代可以是保守取代,诸如其中一个特定类型(例如,极性、非极性/脂族、带正电/碱性、带负电/酸性)的氨基酸被取代为相同类型的氨基酸。

[0048] 如果四个(或更多个)外源核酸序列编码彼此不同的两种或更多种葡糖淀粉酶多肽,则所述两种或更多种葡糖淀粉酶多肽也可具有诸如本文所述的取代,并且与SEQ ID NO:11具有约90%或更高的序列同一性。也就是说,具有不同序列的外源核酸序列可以编码同样具有不同序列但同一性足够接近以发挥类似功能的葡糖淀粉酶,从而具有葡糖淀粉酶活性。

[0049] 在其它方面,葡糖淀粉酶序列任选地包含天然葡糖淀粉酶多肽中不存在的另外的序列。所述另外的序列可为葡糖淀粉酶提供天然多肽中不存在的功能。另外的功能包括例如蛋白酶位点或针对其它蛋白质或物质的结合位点。

[0050] 可能不存在于天然葡糖淀粉酶多肽中但可以加入的另外序列的示例是标签序列。标签序列可以位于葡糖淀粉酶序列的C末端,并且这样的蛋白质可以如下注释:[GA]-[T],其中“T”表示提供标签序列的一个或多个氨基酸。示例性肽标签包含最多5、10、15或20个氨基酸。该肽标签可用于各种目的中的任一种或多种。例如,标签可以允许通过标签结合成员与标签特异性地相互作用的能力从培养基中纯化酶。标签还可以使用具有可检测标签的标签结合成员来检测或鉴定蛋白质。示例性短肽标签为poly-Arg、FLAG、poly-His、c-myc、S和Strep II。

[0051] 本公开的葡糖淀粉酶多肽也可使得天然葡糖淀粉酶多肽的一个或多个区域缺失,其中缺失不影响多肽的淀粉分解活性。该缺失可以基于关于天然葡糖淀粉酶的结构和功能的已知信息,包括突变研究和序列比对(例如,参见Coutinho,出处同上;以及Sierks,出处同上)。在一些方面,葡糖淀粉酶多肽的氨基酸序列具有最多1%、最多2%、最多4%、最多6%、最多8%、最多10%、最多12%、最多14%、最多16%、最多18%、最多20%或最多25%的缺失。例如,在一些方面,葡糖淀粉酶多肽可具有对应于天然葡糖淀粉酶多肽的C末端的一部分的缺失。

[0052] 葡糖淀粉酶的截短形式已被生成,并且经表明具有酶活性。例如,Evans等人(Gene, 91:131;1990(《基因》,第91卷:第131页;1990年))生成了一系列截短形式的葡糖淀粉酶,以研究O-糖基化区域中有多少是GAI的活性或稳定性所必需的,GAI是缺少生淀粉结合域的酶的完全活性形式。据发现,C末端的很大一部分可以从GAI中去除,且对酶的活

性、热稳定性或分泌有很小影响。

[0053] 与引起葡糖淀粉酶活性变化相关联的各种氨基酸取代也是本领域已知的。葡糖淀粉酶序列中各个位置处的氨基酸取代经表明会影响诸如热稳定性、淀粉水解活性、底物利用和蛋白酶抗性性质。因此,本公开设想具有多种独特外源核酸的宿主细胞,每种外源核酸表达一个葡糖淀粉酶序列,所述葡糖淀粉酶序列在多肽的葡糖淀粉酶部分中包含一个或多个氨基酸取代,其中取代不同于葡糖淀粉酶的野生型序列。

[0054] 例如,美国专利No.8,809,023描述了在淀粉水解期间降低异麦芽糖合成与淀粉水解活性之比(IS/SH比)的方法。具体地,描述了在如下氨基酸位置处进行了修饰的里氏木霉(*Trichoderma reesei*)葡糖淀粉酶(Tr GA)(总长度为632个氨基酸,具有含信号肽的N末端):D44R和A539R;或D44R、N61I和A539R。据报道,该葡糖淀粉酶变体在淀粉水解期间表现出与所述亲本葡糖淀粉酶相比降低的IS/SH比率。例如,本公开设想了对应于D44R和A539R的氨基酸取代;或经修饰的Tr GA的D44R、N61I和A539R取代。在广义上,本公开提供了具有多种独特外源核酸的宿主细胞,每种外源核酸表达一个具有如下氨基酸取代的葡糖淀粉酶变体:D44R和A539R;或D44R、N61I和A539R,这些位置与TrGA序列中的相应位置对应,其中所述葡糖淀粉酶变体与TrGA序列的整个长度具有至少90%的氨基酸序列同一性。模板葡糖淀粉酶序列与TrGA序列的对应“相应位置”可以通过例如已知的葡糖淀粉酶多肽序列(用于构建葡糖淀粉酶基因的模板)与TrGA序列的序列比对来理解。

[0055] 又如,美国专利No.8,592,194描述了与野生型葡糖淀粉酶变体相比具有增加的热稳定性的葡糖淀粉酶变体。本公开还描述了在天然Tr GA序列的第10位、第14位、第15位、第23位、第42位、第45位、第46位、第59位、第60位、第61位、第67位、第68位、第72位、第73位、第97位、第98位、第99位、第102位、第108位、第110位、第113位、第114位、第122位、第124位、第125位、第133位、第140位、第144位、第145位、第147位、第152位、第153位、第164位、第175位、第182位、第204位、第205位、第214位、第216位、第219位、第228位、第229位、第230位、第231位、第236位、第239位、第240位、第241位、第242位、第244位、第263位、第264位、第265位、第268位、第269位、第276位、第284位、第291位、第300位、第301位、第303位、第310位、第311位、第313位、第316位、第338位、第342位、第344位、第346位、第349位、第359位、第361位、第364位、第379位、第382位、第390位、第391位、第393位、第394位、第408位、第410位、第415位、第417位和第418位具有一个或多个氨基酸取代的里氏木霉葡糖淀粉酶。例如,本公开设想了还具有被证实提供增加的热稳定性的氨基酸取代中任何一个或多个的Sf GA序列。在广义上,本公开提供了具有多种独特外源核酸的宿主细胞,每种外源核酸表达具有提供增加的热稳定性的氨基酸取代的葡糖淀粉酶变体,发生氨基酸取代的位置与Tr GA序列中的相应位置相对应。

[0056] 来自两种或更多种葡糖淀粉酶的“对应”氨基酸的测定可以通过比对其全部或部分氨基酸序列来确定。序列比对和序列同一性的产生包括全局比对和局部比对,所述比对通常使用计算方法。为了提供全局比对,使用跨所有查询序列的整个长度的全局优化强制序列比对。相比之下,在局部比对中,识别长序列内较短的相似性区域。

[0057] 如本文所用,“等同位置”意指基于一种葡糖淀粉酶的氨基酸序列比对或根据三维结构的对齐,两个序列(例如,具有所需取代的Sf GA序列和GA序列)共有的位置。因此,可以使用序列比对或结构对齐或这两者来确定等同性。

[0058] 在一些实践模式中, BLAST算法用于比较和确定序列相似性或同一性。此外, 可以确定可分配权重或分数的序列中的间隙的存在或显著性。这些算法也可用于确定核苷酸序列相似性或同一性。用于确定相关性的参数是基于用于计算统计相似性的领域已知方法和所确定的匹配的显著性来计算的。预期相关的基因产物具有高度相似性, 诸如大于50%的序列同一性。用于使用BLAST算法确定两个或更多个序列的相关性的示例性参数可以如下所述。

[0059] 在一些实践模式中, 使用BLAST(国家生物技术信息中心(NCBI)基本局部比对搜索工具)2.2.29版软件和默认参数进行比对。使用BLAST2.2.29版算法和默认参数相对于参考序列具有XX%(例如, 80%)的同一性分数的序列被认为与参考序列至少XX%相同或者换句话说讲具有XX%序列同一性。可使用全局比对来比对与例如扣囊复膜酵母G1m葡糖淀粉酶(SEQ ID NO:11)具有显著同一性的序列, 以确定在使用葡糖淀粉酶变体的情况下靶序列(例如, 葡糖淀粉酶同源物)中的哪个(些)对应氨基酸位置可以被一种或多种氨基酸取代。

[0060] 如本文和表1和表2所指出, 来自各种真菌和细菌菌种的葡糖淀粉酶通常还包含天然的“信号序列”。可使用各种其它术语指示本领域已知的“信号序列”, 诸如其中字词“信号”替换为“分泌”或“靶向”或“定位”或“转运”或“先导”, 并且字词“序列”替换为“肽”或“信号”。通常, 信号序列是位于新合成的蛋白质的氨基末端的短氨基酸片段(通常长度在5-30个氨基酸的范围内)。大多数信号肽包含碱性N末端区域(n区域)、中心疏水区域(h区域)和极性C末端区域(c区域)(例如, 参见von Heijne, G. (1986) *Nucleic Acids Res.* 14, 4683-4690 (von Heijne, G., 1986年, 《核酸研究》, 第14卷: 第4683-4690页))。信号序列可以将蛋白质靶向细胞的某个部分, 或者可以靶向蛋白质以从细胞分泌。例如, 经表明, 糖化酵母葡糖淀粉酶STAI基因的天然N末端信号序列可以将其靶向分泌装置的内质网(例如, 参见Yamashita, I. et al., (1985) *J. Bacteriol.* 161, 567-573 (Yamashita, I. 等人, 1985年, 《细菌学杂志》, 第161卷, 第567-573页))。

[0061] 在一些方面, 被引入宿主细胞的外源核酸编码带Sc MF $\alpha$ 2或Sc PH05分泌信号的异源葡糖淀粉酶。在此方面, 酶可被称为“融合蛋白”, 因为其包含来自两个不同多肽的部分, 并且可以如下注释: [Sc MF $\alpha$ 2-SS]-[GA]和[Sc PH05-SS]-[GA]。

[0062] 酿酒酵母交配因子 $\alpha$ 2(Sc MF $\alpha$ 2)分泌信号在美国专利No. 4, 546, 082 (Kurjan等人)中有所描述。Sc MF $\alpha$ 2 SS序列如下所示: MKFISTFLTFILAAVSVA (SEQ ID NO:12)。Sc MF $\alpha$ 2序列来自基因YGL089C (YGL089C), 而MF $\alpha$ 1由基因YPL187W MF $\alpha$ 1编码, 并且MF $\alpha$ 2是MAT $\alpha$ 细胞所分泌的信息素。在一些方面, 本公开提供了具有多种不同序列的外源核酸的宿主细胞, 每种外源核酸编码一种葡糖淀粉酶融合蛋白, 该葡糖淀粉酶融合蛋白包含与SEQ ID NO:12具有90%或更高同一性的分泌信号序列。

[0063] 酿酒酵母可抑制酸性磷酸酶(Sc PH05)分泌信号在美国专利No 5, 521, 086 (Scott等人)和Meyhack等人(EMBO J. 6:675-680, 1982(《欧洲分子生物学学会杂志》, 第6卷: 第675-680页, 1982年))中有所描述, 并且具有如下所示的氨基酸序列: MFKSVVYSILAASLANA (SEQ ID NO:13)。Sc PH05序列来自PH05, 其是编码酿酒酵母酸性磷酸酶的结构基因, 该酸性磷酸酶受到培养基中无机磷酸盐(P<sub>i</sub>)浓度的调控。

[0064] 在一些方面, 本公开提供了具有多种不同序列的外源核酸的宿主细胞, 每种外源核酸编码一种葡糖淀粉酶融合蛋白, 该葡糖淀粉酶融合蛋白包含与SEQ ID NO:13具有90%



或更高同一性的分泌信号序列。例如,一种SEQ ID NO:13的氨基酸可被诸如保守氨基酸的氨基酸取代。

[0065] 可执行分子技术构建核酸序列,该核酸序列是用于表达Sc MF $\alpha$ 2 SS或Sc PH05 SS-葡糖淀粉酶基因的模板(如果葡糖淀粉酶的蛋白质/核苷酸序列是本领域已知的)。可制备独特核酸(四种或更多种),用以编码包含Sc PH05 SS-或Sc MF $\alpha$ 2 SS序列和葡糖淀粉酶序列的蛋白质。也就是说,这些核酸序列彼此不同,但仍然编码相同的Sc PH05 SS-或Sc MF $\alpha$ 2 SS-葡糖淀粉酶多肽,所述不同例如基于核酸序列编码区中的密码子简并性。

[0066] 例如,在一个方面,葡糖淀粉酶的天然序列可在添加Sc MF $\alpha$ 2 SS或Sc PH05 SS序列之前,在其N末端进行改变。在一些方面,天然葡糖淀粉酶信号序列的全部或一部分在连接Sc MF $\alpha$ 2 SS或Sc PH05 SS序列之前去除。例如,葡糖淀粉酶的天然先导序列的一部分可以通过缺失天然分泌信号的一个或多个但不是全部的氨基酸来改变(例如,缺失天然先导序列的最多50%、60%、70%、80%、90%或95%)。天然先导序列的这种一部分的缺失可引起天然葡糖淀粉酶先导物损失其天然功能,以Sc MF $\alpha$ 2或Sc PH05分泌信号提供的功能代替。在其它方面,所有天然分泌信号可从葡糖淀粉酶多肽中去除,并由Sc MF $\alpha$ 2 SS或Sc PH05 SS序列取代。

[0067] 例如并参考表1,在制备融合蛋白构建体中,去除扣囊复膜酵母IFO 0111葡糖淀粉酶(G1m)的前18个氨基酸,这些氨基酸对应于使用CBS预测服务器预测的先导序列(即,SEQ ID NO:11的氨基酸1-18)。因此,扣囊复膜酵母葡糖淀粉酶天然分泌信号的一部分被Sc MF $\alpha$ 2 SS序列(SEQ ID NO:12;19个氨基酸)或Sc PH05 SS序列(SEQ ID NO:13;17个氨基酸)取代,这些序列然后可以直接或间接连接至扣囊复膜酵母葡糖淀粉酶多肽的其余部分(例如,SEQ ID NO:11的氨基酸19-515)。这提供516个氨基酸的Sc MF $\alpha$ 2 SS-Sf GA(SEQ ID NO:14)或514个氨基酸的Sc PH05 SS-Sf GA(SEQ ID NO:15)。在已知所需氨基酸序列后,可制备编码该多肽的独特核酸序列。

[0068] 可能不存在于天然葡糖淀粉酶多肽中的额外序列或者Sc PH05 SS-或Sc MF $\alpha$ 2 SS序列可以作为接头或间隔序列添加。接头序列可以位于Sc PH05 SS-或Sc MF $\alpha$ 2 SS序列与葡糖淀粉酶序列之间。这种融合多肽可以按照如下所示注释:[Sc MF $\alpha$ 2-SS]-[L]-[GA]和[Sc PH05-SS]-[L]-[GA],其中“L”表示将信号序列连接至葡糖淀粉酶的一个或多个氨基酸。示例性接头包含最多5、10、15、20、25、30、35或40个氨基酸。接头可包含使接头变得刚性并防止分泌信号与葡糖淀粉酶的其它部分之间发生相互作用的氨基酸。刚性接头可包含诸如Pro、Arg、Phe、Thr、Glu和Gln的残基。另选地,融合多肽可包含柔性接头。柔性接头可包含甘氨酸残基,并将信号序列连接至融合蛋白的葡糖淀粉酶部分,而不干扰它们各自的功能。在一些方面,该多肽包含具有蛋白酶酶切序列的接头。示例性蛋白酶酶切序列包含凝血酶、Xa因子,鼻病毒3C、TEV蛋白酶、Ssp DnaB、内含肽、Sce VMA1内含肽、肠激酶和KEX2的酶切序列(参见例如Waugh,D.S.,*Protein Expr Purif.* 80 (2):283-293,2011(Waugh,D.S.,《蛋白质表达与纯化》,第80卷第2期:第283-293页,2011年);Zhou et al.,*Microbial Cell Factories* 13:44,2014(Zhou等人,《微生物细胞工厂》,第13卷:第44页,2014年);以及Bourbonnais et al.,*J.Bio.Chem.* 263 (30):15342,1988(Bourbonnais等人,《生物化学杂志》,第263卷第30期:第15342页,1988年))。

[0069] 用于共有葡糖淀粉酶多肽的模板的外源核酸可被引入宿主细胞中。提供至少四种

各自具有不同核酸序列但编码相同葡糖淀粉酶多肽的核酸,并将这些核酸引入宿主细胞中进行表达。换句话说讲,尽管外源葡糖淀粉酶基因可在一些方面编码相同的多肽,但它们不具有相同的核酸序列。在其它方面,至少四种多肽各自具有不同的编码两种或更多种葡糖淀粉酶多肽的核酸序列,优选彼此具有高度同一性(例如,至少90%,并且优选高于95%)的序列。

[0070] 核酸序列的差异可以至少基于氨基酸密码的简并性。也就是说,葡糖淀粉酶基因的形式之间的核酸序列差异是葡糖淀粉酶序列中至少一个密码子三联体中的第二核苷酸和/或第三核苷酸变化的结果。优选的是,核酸序列的差异基于序列中多个密码子的差异。

[0071] 核酸序列的差异可以表示为两个不同葡糖淀粉酶核酸序列之间的核酸数量的百分比差异,或者可表示为在两种不同葡糖淀粉酶核酸序列之间不同的密码子数量之间的百分比差异。例如,如果在两个特异性葡糖淀粉酶核苷酸序列GA<sub>1</sub>和GA<sub>2</sub>之间存在175个核苷酸的差异,每个序列具有1545个核苷酸并编码共有的GA多肽,则其核苷酸序列百分比差异将为大约11.3%(或反过来讲,同一性百分比将为88.7%)。在一些方面,第一外源核酸、第二外源核酸、第三外源核酸和第四外源核酸彼此之间的差别为至少10%。

[0072] 当每个配对提供不同的序列差异时,差异可以表示为一个范围。例如,该范围可以是大约最小差异量到最大差异量。因此,在一些方面,第一外源核酸、第二外源核酸、第三外源核酸和第四外源核酸可以彼此不同,其差异量在约10%至约30%、约15%至约30%、约17%至约27%、约18%至约25%或者约19%至约23%范围内。

[0073] 下面的表3是四种编码葡糖淀粉酶的外源核酸GA<sub>1</sub>-GA<sub>4</sub>的序列差异百分比的示例。表中详细列出了每种可能的核酸配对之间的核苷酸百分比差异。在一些方面,每种核酸配对的百分比差异不同,而在其它方面,两种或更多种核酸配对的百分比差异相同。

[0074] 表3: 示例性序列差异

[0075]

	GA <sub>1</sub>	GA <sub>2</sub>	GA <sub>3</sub>	GA <sub>4</sub>
GA <sub>1</sub>	0	22	19	20
GA <sub>2</sub>	22	0	22	22
GA <sub>3</sub>	19	22	0	23
GA <sub>4</sub>	20	22	23	0

[0076] 另选地,第一外源核酸、第二外源核酸、第三外源核酸和第四外源核酸之间的关系可以解读为同一性百分比。同一性百分比为1-序列差异%,例如,1-0.22=78%序列同一性。

[0077] 在其它方面,第一外源核酸、第二外源核酸、第三外源核酸和第四外源核酸之间的关系可以解读为密码子之间的同一性百分比的百分比差异。例如,如果在两个特异性葡糖淀粉酶核苷酸序列GA<sub>1</sub>和GA<sub>2</sub>之间存在100个密码子的差异(例如,由于175个核苷酸的变化),每个序列具有1545个核苷酸并编码共有的GA多肽,则其密码子百分比差异将为大约6.5%(或反过来讲,同一性百分比将为93.5%)。

[0078] 在一些实践模式中,编码葡糖淀粉酶多肽的核酸序列以及任何调控序列(例如,终止子、启动子等)和载体序列(例如,包含选择标记、整合标记、复制序列等)可以使用已知的分子技术制备。用于制备DNA构建体的方法的一般指南可见于Sambrook et al Molecular

Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989 (Sambrook等人,《分子克隆实验室手册》,冷泉港实验室出版社,纽约州冷泉港,1989年);以及Ausubel et al. Current Protocols in Molecular Biology, Greene Publishing and Wiley-Interscience, New York, N.Y., 1993 (Ausubel等人,《分子生物学现代方法》,格林出版与威立国际科学,纽约州纽约,1993年)。

[0079] 在PCR中可使用少量的葡糖淀粉酶模板DNA作为起始材料,以生成相对大量的特异性DNA片段,该特异性DNA片段包含葡糖淀粉酶基因,任选具有任何另外的所需序列。

[0080] PCR技术可用于扩增或修饰天然葡糖淀粉酶核酸序列以添加任选的异源序列,诸如Sc MF $\alpha$ 2 SS或Sc PH05 SS序列,或将一个或多个突变引入葡糖淀粉酶核酸序列中以提供变体。PCR技术在例如以下文献中有所描述:Higuchi, (1990) in PCR Protocols, pp.177-183, Academic Press (Higuchi, 1990年,《PCR实验手册》,第177-183页,美国学术出版社); Ito et al. (1991) Gene 102:67-70 (Ito等人, 1991年,《基因》,第102卷:第67-70页); Bernhard et al. (1994) Bioconjugate Chem. 5:126-132 (Bernhard等人, 1994年,《生物共轭化学》,第5卷:第126-132页);以及Vallette et al. (1989) Nuc. Acids Res. 17:723-733 (Vallette等人, 1989年,《核酸研究》,第17卷:第723-733页)。这些技术可以任选地包括先前制备的编码葡糖淀粉酶多肽的DNA的定点(或寡核苷酸介导的)诱变、PCR诱变和盒式诱变。

[0081] 另选地,可通过定制基因合成提供商生成核酸分子,诸如DNA2.0公司(加利福尼亚州门洛帕克(Menlo Park, CA))或GeneArt公司(赛默飞世尔科技公司旗下的生命技术公司(Life Technologies, Thermo Fisher Scientific))。

[0082] 可构建表达载体,以包含可操作地连接至在宿主生物中发挥功能的表达控制序列的一个或多个葡糖淀粉酶核酸序列。载体可包含可操作以稳定整合到宿主染色体中的选择序列或标记。另外,载体可包含一个或多个选择性标记基因和适当的表达控制序列。也可包含选择性标记基因,以例如提供对抗生素或毒素的抗性、补充营养缺陷或者提供培养基中不存在的关键营养物质。表达控制序列可包含本领域所熟知的组成型和诱导型启动子、转录增强子、转录终止子等等。

[0083] 在一些方面,核酸可进行密码子优化。用于葡糖淀粉酶编码区的核酸模板可以是编码葡糖淀粉酶的天然DNA序列,或者该模板可以是针对在期望的宿主细胞中表达而优化的密码子优化版本。在一些方面,葡糖淀粉酶序列经过密码子优化,以在酵母属菌种诸如,酿酒酵母中表达。提供关于特定宿主生物中期望的密码子使用的信息的数据库是本领域已知的。

[0084] 在一些方面,葡糖淀粉酶编码区经过密码子优化以在酿酒酵母中表达,并具有以下序列:SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:7和SEQ ID NO:8。

[0085] 根据本公开的一个方面,包含 $\alpha$ -葡糖淀粉酶的DNA构建体可操作地连接至启动子序列,其中启动子序列在精选的宿主细胞中发挥作用。在一些方面,启动子在真菌宿主细胞中显示出转录活性,并且可以衍生自编码与宿主细胞同源或异源的蛋白质的基因。在一些方面,启动子可用于酿酒酵母中的表达。众所周知的组成型启动子的示例包括但不限于细胞色素c启动子(pCYC)、翻译延伸因子启动子(pTEF)、甘油醛-3-磷酸脱氢酶启动子(pGPD)(也称为TDH3[磷酸丙糖脱氢酶]启动子)、磷酸甘油酸激酶启动子(PGK)和乙醇脱氢酶启动

子(pADH1)。任选地,也可以在载体中包含控制表达的另外的因子,诸如增强子等。

[0086] 在一些方面,特定启动子种类(启动子<sub>1</sub>)提供对两个或更多个葡糖淀粉酶核酸序列的表达控制。例如,本公开的生物可包含两个DNA构建体,一个构建体具有受第一启动子(例如,-[启动子<sub>a</sub>]-[GA<sub>1</sub>]-)控制的葡糖淀粉酶基因,并且另一个构建体具有同样受第一启动子(例如,-[启动子<sub>a</sub>]-[GA<sub>3</sub>]-)控制的另一种葡糖淀粉酶基因,其中GA<sub>1</sub>和GA<sub>3</sub>编码共有的葡糖淀粉酶,但具有不同的核酸序列。在一些方面,用于GA<sub>1</sub>和GA<sub>3</sub>的启动子选自甘油醛-3-磷酸脱氢酶(TDH3)启动子和磷酸甘油酸激酶(PGK)启动子。在一些方面,一种DNA构建体(-[启动子<sub>a</sub>]-[GA<sub>1</sub>]-)在生物基因组的一个等位基因(geneX<sub>1</sub>)处整合,并且其它构建体(-[启动子<sub>a</sub>]-[GA<sub>3</sub>]-)在生物基因组的对应等位基因(geneX<sub>2</sub>)处整合。

[0087] 在一些方面,特定启动子种类(启动子<sub>1</sub>)提供对两种或更多种葡糖淀粉酶核酸序列的表达控制,并且另一种特定启动子种类(启动子<sub>2</sub>)提供对另外两种或更多种葡糖淀粉酶核酸序列的表达控制。例如,本公开的生物可包含四种DNA构建体,一种构建体具有受第一启动子(例如,-[启动子<sub>a</sub>]-[GA<sub>1</sub>]-)控制的葡糖淀粉酶基因,另一种构建体具有同样受第一启动子(例如,-[启动子<sub>a</sub>]-[GA<sub>3</sub>]-)控制的另一种葡糖淀粉酶基因,另一种构建体具有受第二启动子(例如,-[启动子<sub>b</sub>]-[GA<sub>2</sub>]-)控制的葡糖淀粉酶基因,并且另一种构建体具有同样受第二启动子(例如,-[启动子<sub>b</sub>]-[GA<sub>4</sub>]-)控制的另一种葡糖淀粉酶基因,其中GA<sub>1</sub>、GA<sub>2</sub>、GA<sub>3</sub>和GA<sub>4</sub>编码共有的葡糖淀粉酶,但具有不同的核酸序列。在一些方面,用于GA<sub>1</sub>和GA<sub>3</sub>的启动子是甘油醛-3-磷酸脱氢酶(TDH3)启动子,并且用于GA<sub>2</sub>和GA<sub>4</sub>的启动子是磷酸甘油酸激酶(PGK)启动子。在一些方面,两种DNA构建体(-[启动子<sub>a</sub>]-[GA<sub>1</sub>]-和-[启动子<sub>b</sub>]-[GA<sub>2</sub>]-)在生物基因组的一个等位基因(geneX<sub>1</sub>)处整合,并且另外两种DNA构建体(-[启动子<sub>a</sub>]-[GA<sub>3</sub>]-和-[启动子<sub>b</sub>]-[GA<sub>4</sub>]-)在生物基因组的对应等位基因(geneX<sub>2</sub>)处整合。

[0088] 包含葡糖淀粉酶基因的表达构建体还可包含在宿主细胞中发挥功能的任何终止序列。例如,终止序列和启动子序列可以来自相同的基因,或者终止序列与启动子序列是异源的(即,来自不同的基因)。在一些实践模式中,终止子包括核糖体蛋白L3(RPL3)终止子或细胞色素C(CYC1)终止子。

[0089] 在一些方面,彼此异源的特定启动子和终止子种类(启动子<sub>a</sub>-终止子<sub>c</sub>)对提供对两个或更多个葡糖淀粉酶核酸序列的表达控制,并且另一个彼此异源的特定启动子和终止子种类(启动子<sub>b</sub>-终止子<sub>d</sub>)对提供对另外两个或更多个葡糖淀粉酶核酸序列的表达控制。例如,本公开的生物可包含四种DNA构建体,一种构建体具有受第一启动子和第一终止子(例如,-[启动子<sub>a</sub>]-[GA<sub>1</sub>]-[终止子<sub>c</sub>])控制的葡糖淀粉酶基因,另一种构建体具有同样受第一启动子和第一终止子(例如,-[启动子<sub>a</sub>]-[GA<sub>3</sub>]-[终止子<sub>c</sub>])控制的另一种葡糖淀粉酶基因,另一种构建体具有受第二启动子和第二终止子(例如,-[启动子<sub>b</sub>]-[GA<sub>2</sub>]-[终止子<sub>d</sub>])控制的葡糖淀粉酶基因,并且另一种构建体具有同样受第二启动子和第二终止子(例如,-[启动子<sub>b</sub>]-[GA<sub>4</sub>]-[终止子<sub>d</sub>])控制的另一种葡糖淀粉酶基因,其中GA<sub>1</sub>、GA<sub>2</sub>、GA<sub>3</sub>和GA<sub>4</sub>编码共有的葡糖淀粉酶,但具有不同的核酸序列。在一些方面,用于GA<sub>1</sub>和GA<sub>3</sub>的启动子和终止子是甘油醛-3-磷酸脱氢酶(TDH3)启动子和细胞色素C(CYC1)终止子,并且用于GA<sub>2</sub>和GA<sub>4</sub>的启动子和终止子是磷酸甘油酸激酶(PGK)启动子和核糖体蛋白L3(RPL3)终止子。在一些方面,两种DNA构建体(-[启动子<sub>a</sub>]-[GA<sub>1</sub>]-[终止子<sub>c</sub>]和-[启动子<sub>b</sub>]-[GA<sub>2</sub>]-[终止子<sub>d</sub>])在生物基因组的一个等位基因(geneX<sub>1</sub>)处整合,并且另外两种DNA构建体(-[启动子<sub>1</sub>]-[GA<sub>3</sub>]-[终止子<sub>c</sub>]和-

[启动子<sub>2</sub>]-[GA<sub>4</sub>]-[终止子<sub>d</sub>]在生物基因组的对应等位基因(geneX<sub>2</sub>)处整合,其中生物为二倍体。

[0090] DNA构建体可以使用载体引入宿主细胞中。载体可以是被引入宿主细胞中时稳定引入的任何载体。在一些方面,载体被整合到宿主细胞基因组中并被复制。载体包括克隆载体、表达载体、穿梭载体、质粒、噬菌体颗粒、表达盒等。在一些方面,载体是包含可操作地连接至葡糖淀粉酶编码序列的调控序列的表达载体。

[0091] 包含葡糖淀粉酶基因的DNA构建体还可包含选择性标记,从而便于宿主细胞中的选择。例如,选择性标记可用于转化的酵母。酵母选择性标记的示例包括通常用于选择转化的酵母细胞的标记。可使用控制营养缺陷型的基因来使用营养缺陷型标记,意味着该基因能够使酵母产生酵母生长所需的营养物质。控制营养缺陷型的示例性基因包括亮氨酸营养缺陷型(LEU2)、组氨酸营养缺陷型(HIS3)、尿嘧啶营养缺陷型(URA3,URA5)和色氨酸营养缺陷型(TRP1)。

[0092] DNA构建体可以是整合到基因组中并与整合其的染色体一起复制的构建体。例如,真菌细胞可以用编码葡糖淀粉酶的DNA构建体转化,并将DNA构建体整合到宿主染色体中。这种整合通常被认为是有利的,因为DNA序列更可能被稳定地保持。在本公开的方法中,整合的葡糖淀粉酶构建体可以在发酵周期过程中稳定地保持,这意味着该构建体不从其整合位点丢失。因此,整合的葡糖淀粉酶能够不断地产生葡糖淀粉酶并将葡糖淀粉酶分泌到发酵培养基中,在此它可以作用于水解的淀粉底物。将DNA构建体整合到宿主染色体中可以根据常规方法进行,诸如通过同源或异源重组。

[0093] 在一种实践模式中,包含葡糖淀粉酶基因的一种或多种DNA构建体整合在遗传基因座处,其中该整合对细胞的健康没有显著的不利影响。例如,整合可以发生在不知道具有任何多肽编码序列的基因组的基因座处,或者发生在基因组的对于在所需生长条件下(诸如,在使用淀粉或淀粉衍生产品作为能源的发酵条件下)发挥功能并非必需的基因的基因座处。例如,对于酿酒酵母,有关于其基因组中开放阅读框(ORF)的重要性的信息是可用的。参见例如[http://www-sequence.stanford.edu/group/yeast\\_deletion\\_project/deletions3.html](http://www-sequence.stanford.edu/group/yeast_deletion_project/deletions3.html)。鉴于本领域已知的信息,技术人员可以选择一种或多种非必需基因作为用于整合包含葡糖淀粉酶基因的一种或多种DNA构建体的靶标。基因是否“必需”可以使用富含葡萄糖的培养基在生长条件下确定。

[0094] 在一些方面,包含葡糖淀粉酶基因的一种或多种DNA构建体整合在非必需基因处,所述非必需基因为L-乳酸细胞色素c氧化还原酶(CYB2)。

[0095] 在一些实践模式中,第一核酸和第二核酸(GA<sub>1</sub>和GA<sub>2</sub>)首先整合到酵母属菌株中非必需基因的第一等位基因中,并且第三核酸和第四核酸(GA<sub>3</sub>和GA<sub>4</sub>)整合到酵母属菌株中非必需基因的第二等位基因中。在一些实践模式中,CYB2基因的两个等位基因被外源葡糖淀粉酶基因破坏。

[0096] 在一些方面,整合策略是使两个不同DNA构建体共转化到基因的一个等位基因中,其中第一构建体携带GA<sub>1</sub>基因,并且第二构建体携带GA<sub>2</sub>基因。这之后是另外两个不同DNA构建体的另一共转化,第三构建体携带GA<sub>3</sub>基因,并且第四构建体携带GA<sub>4</sub>基因。第一构建体包含:被靶向整合的基因的一部分(“GTI”;例如,CYB2基因的一部分),其优选位于构建体的5'端;GA<sub>1</sub>基因,其侧接所需启动子和终止子序列;以及选择性标记基因,例如URA3。第二构建

体包含:选择性标记基因(例如,URA3),其优选位于构建体的5'端;GA<sub>2</sub>基因,其侧接与GA<sub>1</sub>基因不同的所需启动子和终止子序列;被靶向整合的基因的一部分(GTI;例如,CYB2基因的一部分)。第三DNA构建体和第四DNA构建体可以类似于第一DNA构建体和第二DNA构建体排列,但使用不同的选择性标记(例如,来自曲霉属(*Aspergillus*)的amdS基因,而不是URA3)。另外,选择性标记可以被截短,使得共转化的两个转化片段本身都不能产生功能性基因,但是通过同源重组可以产生全长功能性选择性标记。

[0097] 示例性第一DNA构建体具有以下遗传因子排列:5'-[GTI部分]-[启动子<sub>a</sub>]-[GA<sub>1</sub>]-[终止子<sub>c</sub>]-[选择性标记<sub>1</sub>]-3'。示例性第二DNA构建体具有以下遗传因子排列:5'-[选择性标记<sub>1</sub>]-[启动子<sub>b</sub>]-[GA<sub>2</sub>]-[终止子<sub>d</sub>]-[GTI部分]-3'。示例性第三DNA构建体具有以下遗传因子排列:5'-[GTI部分]-[启动子<sub>a</sub>]-[GA<sub>3</sub>]-[终止子<sub>c</sub>]-[选择性标记<sub>2</sub>]-3'。示例性第四DNA构建体具有以下遗传因子排列:5'-[选择性标记<sub>2</sub>]-[启动子<sub>b</sub>]-[GA<sub>4</sub>]-[终止子<sub>d</sub>]-[GTI部分]-3'。

[0098] 可用包含一个或多个葡糖淀粉酶基因的DNA构建体转化各种宿主细胞。在一些方面,DNA构建体存在于细菌细胞中,其可用于例如核酸序列的扩增或用于产生大量多肽。

[0099] 在其它方面,宿主细胞是真核细胞,例如真菌细胞。

[0100] 在本公开的一些方面,葡糖淀粉酶核酸可以在酵母菌株中进行工程化。例如,酵母可以是酵母属、伊萨酵母属(*Issatchenkia*)、假丝酵母属(*Candida*)、毕赤酵母属(*Pichia*)、耶氏酵母属(*Yarrowia*)或克鲁维酵母属(*Kluyveromyces*)菌种。示例性酵母菌种包括酿酒酵母、东方伊萨酵母(*Issatchenkia orientalis*)、产朊假丝酵母(*Candida utilis*)、树干毕赤酵母(*Pichia stipitis*)、解脂耶氏酵母(*Yarrowia lipolytica*)、马克斯克鲁维酵母(*Kluyveromyces marxianus*)和乳酸克鲁维酵母(*Kluyveromyces lactis*)。这些菌种的工程化细胞在本领域已有所描述(例如,参见WO 2007032792; Tamakawa, H. et al. (2011) *Biosci Biotechnol Biochem.* 75:1994-2000 (Tamakawa, H. 等人, 2011年,《生物科学、生物技术与生物化学》,第75卷:第1994-2000页); Ilmén, M., et al. (2007) *Appl Environ Microbiol.* 73:117-123 (Ilmén, M. 等人, 2007年,《应用与环境微生物学》,第73卷:第117-123页); Xie, D., et al. (2015) *Appl Microbiol Biotechnol.* 99:1599-1610 (Xie, D. 等人, 2015年,《应用微生物学与生物技术》,第99卷:第1599-1610页); Zhang, J. et al. (2014) *Bioresour Technol.* 152:192-201 (Zhang, J. 等人, 2014年,《生物资源技术》,第152卷:第192-201页); 以及González-Siso, M. I., et al. (2015) *Microb Biotechnol.* 8:319-330 (González-Siso, M. I. 等人, 2015年,《微生物生物技术》,第8卷:第319-330页)。

[0101] 在一些方面,宿主细胞对发酵培养基中较高量的生物衍生产物诸如乙醇具有耐受性。在一些方面,宿主细胞是“工业酵母”,其是指在乙醇发酵中便于使用的任何酵母。示例包括清酒酵母、烧酒酵母、葡萄酒酵母、啤酒酵母、面包酵母等等。清酒酵母表现出高乙醇发酵能力以及高乙醇耐受性和遗传稳定性。通常,工业酵母具有高乙醇耐受性,并且优选在10%或更高的乙醇浓度下存活。

[0102] 在示例性方面,宿主细胞是酿酒酵母。一些酿酒酵母对乙醇具有高耐受性。多种乙醇耐受性酵母菌株可商购获得,诸如RED STAR™和ETHANOL RED™酵母(美国Fermentis/Lesaffre公司(Fermentis/Lesaffre, USA))、FALI™(美国Fleischmann's酵母公司(Fleischmann's Yeast, USA))、SUPERSTART™和THERMOSACC™酵母(美国威斯康星州的乙醇

技术公司(Ethanol Technology,Wis.,USA)、BIOFERM™AFT和XR™(美国佐治亚州NABC北美生物产品公司(NABC--North American Bioproducts Corporation,GA,USA))、GERT STRAND™(瑞典Gert Strand AB公司(Gert Strand AB,Sweden))和FERMIOL™(帝斯曼配料部(DSM Specialties))。

[0103] 工业酵母通常是原养型的,并因此没有适于选择转化体的营养缺陷型标记。如果宿主细胞不具有原本在转化后促进细胞内葡糖淀粉酶基因的选择或保留的遗传背景,则宿主细胞可经工程化以引入一种或多种基因突变,从而建立与细胞中的葡糖淀粉酶基因相关联的标记基因的使用并维持细胞中的葡糖淀粉酶基因。例如,在将葡糖淀粉酶基因引入细胞之前,可对可商购获得的乙醇耐受性酵母细胞进行基因修饰。

[0104] 可通过破坏控制营养缺陷型的基因来提供不同营养缺陷型的标记。在一种实践模式中,乙醇耐受性酵母菌株经工程化以破坏控制营养缺陷型的一种或多种基因的拷贝,诸如LYS2、LEU2、HIS3、URA3、URA5和TRP1。在提供尿嘧啶营养缺陷型的情况下,例如,乙醇耐受性酵母的正常ura3基因可以被从尿嘧啶营养缺陷突变体(例如,酿酒酵母MT-8菌株)获得的ura3<sup>-</sup>片段取代,以破坏正常URA3基因。就URA3被破坏的菌株而言,可通过利用URA3基因被破坏的菌株能够在含有5-氟乳清酸(5-FOA)的培养基中生长,而正常的URA3菌株(野生型酵母或常用工业酵母)不能在其中生长这一事实优点来容易地鉴定或选择标记的存在/不存在。就LYS2基因被破坏的菌株而言,可通过利用LYS2被破坏的菌株能够在含有 $\alpha$ -氨基己二酸的培养基中生长,而正常的LYS2菌株(野生型酵母或常用工业酵母)不能在其中生长这一事实优点来容易地鉴定或选择标记的存在/不存在。根据所采用的营养缺陷型,可以使用用于破坏营养缺陷型控制基因和选择性分离营养缺陷型控制基因突变体的方法。另选地,可使用显性选择标记,诸如来自构巢曲霉(*Aspergillus nidulans*)的amdS(美国专利No.5,876,988),其可在乙酰胺作为唯一氮源的条件生长;或ARO4-OPF,其可在存在氟苯丙氨酸的条件下生长(Fukuda等人)。这些标记可以使用可再循环的cre-loxP系统重复使用,或者可以用于生成允许使用额外标记的营养缺陷型菌株。

[0105] 在宿主细胞经工程化以便为引入葡糖淀粉酶基因提供期望的遗传背景之后,将一种或多种基因构建体引入细胞以整合到基因组中或稳定保持并允许用于表达。将基因构建体引入宿主细胞的方法包括转化、转导、转染、共转染、电穿孔。具体地,可使用乙酸锂法、原生质体法等进行酵母转化。待引入的基因构建体可以质粒的形式或通过插入宿主的基因中或通过宿主基因的同源重组而整合到染色体中。其中引入了基因构建体的转化酵母可以用选择性标记(例如,如上所述的营养缺陷型标记)进行选择。可通过测量表达的蛋白质的活性来进一步确认。

[0106] 包含葡糖淀粉酶基因的外源核酸序列的转化可以使用本领域熟知的方法确认。这样的方法包括例如核酸分析(诸如Northern blots或mRNA的聚合酶链反应(PCR)扩增),或用于表达基因产物的免疫印迹法,或其它用于测试引入的核酸序列的表达或其相应基因产物的合适分析方法。本领域技术人员应理解,外源核酸以足够的量表达以产生所需产物,并且还理解,可使用本领域熟知的和本文所公开的方法优化表达水平以获得足够的表达。

[0107] 本公开的工程化酵母可以任何合适的形式提供。在一些方面,工程化酵母被脱水以形成干酵母组合物。干酵母组合物相比于湿组合物可具有延长的储存寿命。

[0108] 表达异源葡糖淀粉酶或者两种或更多种异源葡糖淀粉酶的工程化酵母菌株可在

发酵过程中用于制备产物,其中葡糖淀粉酶由彼此不同的多种外源核酸编码。发酵产物(在本文中也称为“生物产物”)可以是可通过葡糖淀粉酶进行的淀粉材料酶促降解、葡萄糖形成和葡萄糖发酵来制备的任何产物。在一些方面,发酵产物选自:氨基酸、有机酸、醇、二醇、多元醇、脂肪酸、单酰基甘油酯、二酰基甘油酯、三酰基甘油酯、甜味剂、槐糖脂以及它们的混合物。

[0109] 为有机酸或氨基酸的其它示例性生物产物包括乳酸、柠檬酸、丙二酸、羟基丁酸、己二酸、赖氨酸、酮戊二酸、戊二酸、3-羟基丙酸、琥珀酸、苹果酸、富马酸、衣康酸、粘康酸、甲基丙烯酸和乙酸,以及其衍生物和其盐。

[0110] 在一些实践模式中,具有编码异源GA的多种外源核酸的工程化酵母菌株可在发酵过程中用于产生乳酸。在这方面,可以进一步修饰经葡糖淀粉酶修饰的酵母菌株以包含用于增强乳酸产生的异源乳酸脱氢酶(LDH)基因。转化到酵母菌株中的异源LDH基因在例如WO 99/14335、WO 00/71738和WO 02/42471中有所描述。

[0111] 工程化酵母还可包含一种或其它不同于用异源信号序列修饰葡糖淀粉酶的基因修饰。例如,一种或多种另外的修饰可包括提供不同的多糖降解酶的那些,诸如外源或经修饰的 $\alpha$ -淀粉酶、 $\beta$ -淀粉酶、支链淀粉酶、异淀粉酶或环糊精葡萄糖基转移酶;外源或经修饰的糖转运蛋白基因(诸如,异麦芽糖转运蛋白);和/或将低分子量非葡萄糖的糖转化为葡萄糖的外源或经修饰的基因,诸如异麦芽糖酶。

[0112] 使用表达葡糖淀粉酶基因的宿主细胞的发酵可以在含淀粉和/或糖的植物材料的存在下进行,所述植物材料涉及可得自任何植物和植物部分的含淀粉和/或糖的植物材料,诸如块茎、根、茎、叶和种子。含淀粉和/或糖的植物材料可以从谷物诸如大麦、小麦、玉米、裸麦、高粱、粟、大麦、马铃薯、木薯或大米及它们的任何组合获得。可加工含淀粉和/或糖的植物材料,诸如通过诸如研磨、用麦芽处理或部分用麦芽处理的方法。在一些方面,淀粉材料来自玉米粉、经碾磨的玉米胚乳、高粱粉、大豆粉、小麦粉、生物质源淀粉、大麦粉以及它们的组合。

[0113] 在一些方面,发酵培养基包含经处理的淀粉。例如,发酵培养基可包含部分水解的淀粉。部分水解的淀粉包括经过极少水解的制剂(例如,DE为5,含少量右旋糖)到经过大量水解的制剂(例如,DE为95,主要为右旋糖)。在一些制剂中,部分水解的淀粉可包含高分子量糊精和高分子量麦芽糖糊精。在一些实践模式中,在发酵培养基中使用右旋糖当量(“DE”)在约5至约95或更高、或在一些实施方案中约45至约65范围内的部分水解淀粉产品。部分水解淀粉的其它DE范围包括5-25、15-35、25-45、35-55、55-75、65-85和75-95。

[0114] 部分水解淀粉及其制剂是本领域所熟知的。部分水解淀粉可通过用酸(诸如,盐酸或硫酸)在高温下加热淀粉,然后用合适的碱(诸如,碳酸钠)中和水解混合物来制备。另选地,部分水解淀粉可以通过酶法制备,诸如通过向淀粉制剂中添加 $\alpha$ -淀粉酶来制备。 $\alpha$ -淀粉酶可以引起含有三个或更多个(1 $\rightarrow$ 4)- $\alpha$ -连接的D-葡萄糖单元的多糖中的(1 $\rightarrow$ 4)- $\alpha$ -D-糖苷键的内水解。例如,通过酶处理部分水解的玉米淀粉是经过液化并用 $\alpha$ -淀粉酶处理过的液化玉米淀粉。可使用具有期望范围内的量的淀粉和淀粉降解产物的部分水解淀粉产品。

[0115] 在本公开的方面,考虑到葡糖淀粉酶从工程化酵母产生并分泌到发酵培养基中,发酵方法可以省略向培养基中添加纯化或富集的商用葡糖淀粉酶,或至少允许大量减少用于发酵方法的商用葡糖淀粉酶。例如,本公开的工程化酵母可允许添加待消除或至少减少



约50%、60%、70%、80%、90%或95%的商用葡糖淀粉酶。在不使用分泌葡糖淀粉酶的工程化酵母的发酵方法中,将使用典型量在约7个单位至约50个单位/升范围内的葡糖淀粉酶。

[0116] 可使用玉米湿法研磨工艺来得到可用于发酵的浸泡水。玉米谷粒可以被浸泡,然后研磨,并分离成其主要成分。轻质浸泡水是浸泡工艺的副产物,并含有可溶性蛋白质、氨基酸、有机酸、碳水化合物、维生素和矿物质的混合物。

[0117] 发酵液包含水,并且优选包含营养物质,诸如氮源(诸如蛋白质)、维生素和盐。缓冲剂也可以存在于发酵培养基中。其它成分也可以在发酵周期后存在于发酵液中,诸如可随着发酵进程而积聚的发酵产物和其它代谢物。任选地,发酵液可以用碱(诸如,氢氧化钙或碳酸钙、氨或氢氧化铵、氢氧化钠或氢氧化钾)缓冲,以维持使生物正常运作的pH。

[0118] 本公开的工程化酵母还可以从工程化酵母的生长速率方面进行描述。酵母的生长速率可以通过 $L = \log(\text{数量})$  对比 $T(\text{时间})$  来定义,其中数量是每单位体积(mL)形成的酵母细胞数量。

[0119] 发酵在可进行发酵的条件下进行。虽然条件可以根据特定的生物和所需的发酵产物而变化,但典型条件包括约20°C或更高的温度,并且更典型地在约30°C至约50°C的范围内。在发酵期间,可混合或搅拌反应混合物。在一些实践模式中,混合或搅拌可以通过使气体在发酵液中鼓泡的机械动作来实现。另选地,可以在发酵过程中使用直接机械搅拌,诸如通过叶轮或通过其它装置进行。

[0120] 工程化酵母可以是具有增强的耐受性,以在比诸如酿酒酵母的酵母通常生长的温度更高的温度下生长的酵母。例如,酿酒酵母通常在30°C-33°C的温度范围内具有最佳生长。在一些方面,本公开的工程化酵母显示出改善的耐受性,以在34°C-40°C范围内的温度下生长。

[0121] 例如,与未经过基因修饰的参考酵母相比,本公开的工程化酵母可在34°C-40°C范围内的温度下具有一定生长速率,该生长速率比未经过基因修饰的参考酵母的生长速率高10%、20%、30%、40%或50%。

[0122] 在一些情况下,在工业产能发酵罐中进行发酵,以便实现商业规模的经济效益和控制。在一个方面,发酵在容量为约10,000升或更高的发酵罐中进行。

[0123] 可调节发酵培养基的pH,为葡糖淀粉酶活性、细胞生长和发酵活性提供最佳条件,从而得到期望的产物,诸如乙醇。例如,可将溶液的pH调节至3至5.5的范围内。在一种实践模式中,发酵培养基的pH在4至4.5的范围内。

[0124] 如上文所述,本发酵过程使用表达葡糖淀粉酶基因的经基因修饰的微生物,并且能够分泌产生到发酵培养基中的酶。因此,这些酶直接暴露于发酵液条件下并影响发酵培养基中的碳水化合物组成。在发酵培养基中,葡糖淀粉酶可以通过酶切 $\alpha$ -(1,4)糖苷键和 $\alpha$ -(1,6)糖苷键而引起淀粉或相关寡糖和多糖分子的非还原性末端的D-葡萄糖的水解和释放。

[0125] 淀粉也可受到存在于发酵培养基中的一种或多种其它淀粉酶(例如, $\alpha$ -淀粉酶)的作用。例如,如果 $\alpha$ -淀粉酶存在于发酵培养基中,则它引起前体淀粉的部分水解,并通过水解内部的 $\alpha$ -(1,4)-键来引起淀粉分子的部分分解。

[0126] 在一些实践模式中,发酵以单一批次进行,直到完成。

[0127] 在其它实践模式中,发酵以分批补料发酵过程进行。在这种实践模式中,将待发酵

的淀粉材料的总量的第一部分添加到发酵培养基中,其中葡糖淀粉酶作用于淀粉以引起用作发酵底物的葡萄糖的形成。以一个或多个部分添加另外的淀粉材料,以便为培养基中的葡糖淀粉酶提供更多的底物。可对淀粉的添加进行调控,并可监测葡萄糖的形成以提供有效的发酵。

[0128] 优选的是,发酵以连续的操作模式进行。在这种模式下,多个发酵罐串联运行,其中在第一发酵罐中供应淀粉水解产物,将其送入第二发酵罐,以此类推,直到淀粉水解产物转化为乙醇。连续操作可以使用2至7个发酵罐进行。

[0129] 在发酵模式中,可以在“整个”发酵周期中添加包含部分水解淀粉的液体培养基,意味着部分水解的淀粉在此期间不止一次地添加。这可以通过在整个周期中向发酵培养基中周期性地添加多个批次的液体培养基、通过在整个周期中连续添加包含部分水解淀粉的液体培养基或这两种方式的组合来实现。

[0130] 在一些实践模式中,使用可变速率添加系统将淀粉材料的总量的一部分添加到发酵液中。这种系统的示例包括可变速度泵或可操作地连接至泵的计量阀(诸如,节流阀),该泵或阀可用于改变随时间推移引入发酵液中的淀粉材料的量。在一些实践模式中,在添加一部分淀粉材料的过程中,通过实时监测系统监测葡萄糖浓度。

[0131] 实时监测系统包括直接监测葡萄糖浓度的系统和间接监测葡萄糖浓度的系统。通常直接监测葡萄糖浓度的实时监测系统的示例包括基于红外(IR)光谱的系统、近红外(NIR)光谱系统、傅里叶变换红外(FTIR)系统、基于折射率的系统、基于酶的自动化测量系统(诸如,由YSI生命科学系统公司(YSI Life Sciences systems)出售的YSI 2950生化分析仪(YSI 2950Biochemistry Analyzer))、基于高效液相色谱(HPLC)的系统、基于气相色谱(GC)的系统以及本领域技术人员已知的其它实时监测系统。另外,间接监测/测量发酵过程葡萄糖浓度的实时监测系统可以通过确定特定发酵过程中典型的碳分布并将发酵液中存在的葡萄糖浓度与发酵所表现出的另一个参数相关联来开发,例如,将发酵液中存在的葡萄糖水平与二氧化碳释放速率的测量结果和来自发酵容器的废气流中存在的二氧化碳的量相关联。二氧化碳可易于通过使用质谱仪或用于测量废气流的组分的其它合适仪器技术来测量。在优选方面,通过使用红外光谱的实时监测系统监测葡萄糖浓度。在另一个优选方面,通过使用近红外光谱的实时监测系统监测葡萄糖浓度。实时监测系统与控制发酵液中淀粉材料的引入的设备相连,以调节葡萄糖的形成从而在发酵液中的期望浓度。

[0132] 在发酵过程中,可以取发酵培养基的样品来测定培养基中葡糖淀粉酶活性的量。培养基中的葡糖淀粉酶活性的量可以被称为胞外葡糖淀粉酶活性,因为其对应于从工程化酵母分泌的葡糖淀粉酶。在一些测量模式中,培养基中葡糖淀粉酶活性的量可以通过每体积培养基中每一定量的生物质的葡糖淀粉酶活性的量来确定。

[0133] 如本文所用,“生物质”是指工程化酵母的重量,其可以每升培养基的干燥细胞重量的克数测量(DCW/L)。

[0134] GA活性的单位(U)可以定义为催化从淀粉释放1mg葡萄糖/min的酶量。葡糖淀粉酶活性可以通过在两步终点测定法中将淀粉水解物偶联到HXK/G6PDH反应混合物(Sigma G3293),从而在浓缩发酵液中测得。发酵液可以从使用非葡萄糖碳源(即,棉子糖)生长的预定量的细胞中浓缩,以避免干扰测定。

[0135] 比活性等于给定体积的发酵液中的活性除以相同体积的发酵液中的细胞的湿重。

比活性具有以下单位：每克生物质的以U计的GA活性(U/g生物质)。在测定法中使用的生物质的量可以通过经过滤或离心去除发酵液后测定湿细胞重量来测量。

[0136] 通过将1.1g玉米淀粉(S4126,西格玛公司(Sigma))溶解于50mL接近沸腾的水中,然后加入1mL 3M醋酸钠(pH 5.2),从而制备淀粉溶液。将通常在1-20 $\mu$ L范围内的一定体积的浓缩发酵液( $V_b$ ) (通过使用10Kb Kd截留柱Millipore#UFC901008制备)加入淀粉浆料( $V_s$ )中,总体积为200 $\mu$ L,并使其在37 $^{\circ}$ C下温育一段特定的时间(T),通常为5至60分钟。选择参数,使得葡萄糖形成在所需时间内是线性的。每个样品取20 $\mu$ L加入到2 $\mu$ L 0.6N NaOH中并充分混合。然后加入200 $\mu$ L HXK/G6PDH混合物,并在30 $^{\circ}$ C下温育30分钟。使用分光光度计(SpectraMax<sup>TM</sup> M2)测量340nm处的吸光度。通过采用已知葡萄糖标准品的回归分析计算每个样品释放的葡萄糖的量。每克生物质的酶比活性(U/g生物质)可以通过获得浓缩前使用的样品的重量(克)来计算。活性单位=(mg葡萄糖/T)  $\times$  (( $V_b+V_s$ )/( $V_b$ ))  $\times$  (222/20)。比活性=活性单位/g生物质。

[0137] 在其它方面,可以相对于参考酵母描述由本公开的工程化酵母提供的发酵培养基中的葡糖淀粉酶活性的量。例如,表达由多种外源核酸编码的异源葡糖淀粉酶(例如,与SEQ ID NO:11具有90%或更高的同一性)的工程化酵母的葡糖淀粉酶活性的量可以与不含外源核酸的原本相同的酵母比较。

[0138] 发酵培养基中葡糖淀粉酶活性的测定可以在发酵期间的所需时间点进行。例如,发酵培养基的样品可以在发酵过程进程的约1/10、约2/10、约3/10、约4/10、约5/10、约6/10、约7/10、约8/10、约9/10处获取,或者在发酵过程终点处获取,并且可测试样品的葡糖淀粉酶的活性。

[0139] 在一些实践模式中,发酵周期为约30小时或更长、约40小时或更长、约50小时或更长、或约60小时或更长,诸如在约40至约160小时、50至约150小时、约60至约140小时、约60至约120小时或者约60至约120小时范围内的时间段。在实践模式中,发酵周期为约5天或约6天。

[0140] 发酵产物(在本文中也称为“生物衍生产物”或“生物产物”)可以是可通过葡糖淀粉酶进行的淀粉材料酶促降解、葡萄糖形成和葡萄糖发酵来制备的任何产物。在一些方面,发酵产物选自:氨基酸、有机酸、醇、二醇、多元醇、脂肪酸、脂肪酸烷基酯(诸如,脂肪酸甲酯或脂肪酸乙酯(例如,C6至C12脂肪酸甲酯(优选C8至C10脂肪酸甲酯)))、单酰基甘油酯、二酰基甘油酯、三酰基甘油酯以及它们的混合物。优选的发酵产物是有机酸、氨基酸、脂肪酸烷基酯(诸如,脂肪酸甲酯(例如,C8至C12脂肪酸甲酯(优选C8至C10脂肪酸甲酯)))以及它们的盐,并且尤其是其中有机酸选自羟基羧酸(包括单羟基单羧酸、二羧酸和三羧酸以及二羟基单羧酸、二羧酸和三羧酸)、单羧酸、二羧酸和三羧酸,以及它们的混合物。通过本发明的方法制备的发酵产物的示例是有机酸或氨基酸诸如乳酸、柠檬酸、丙二酸、羟基丁酸、己二酸、赖氨酸、酮戊二酸、戊二酸、3-羟基丙酸、琥珀酸、苹果酸、富马酸、衣康酸、粘康酸、甲基丙烯酸、乙酸、己酸甲酯、辛酸甲酯、壬酸甲酯、癸酸甲酯、十二烷酸甲酯、己酸乙酯、辛酸乙酯、壬酸乙酯、癸酸乙酯、十二烷酸乙酯,以及它们的混合物和它们的衍生物和它们的盐。在优选方面,本公开的发酵方法产生乙醇作为生物产物。

[0141] 在一些实践模式中,可以从发酵液中回收发酵产物。实现这一点的方式将取决于具体产物。然而,在一些实践模式中,通常通过过滤步骤或离心步骤将生物与液相分离,并

通过例如蒸馏、萃取、结晶、膜分离、渗透、反渗透或其它合适的技术回收产物。

[0142] 本发明的方法提供以优异的产率和纯度在生产规模水平上制备发酵产物的能力。在一个方面,该方法在至少25,000加仑的量的发酵液中进行。在一个方面,执行分批处理以产生至少25,000加仑的最终发酵液的批次。添加连续方法,至少200,000加仑的容器

[0143] 包含分泌的葡糖淀粉酶的组合物可以任选地与不同于葡糖淀粉酶的以下酶的任何一种或任何组合结合使用。示例性的其它酶包括 $\alpha$ -淀粉酶、 $\beta$ -淀粉酶、肽酶(蛋白酶、蛋白水解酶、内肽酶、外肽酶)、支链淀粉酶、异淀粉酶、纤维素酶、半纤维素酶、内切葡聚糖酶和相关 $\beta$ -葡聚糖水解辅助酶、木聚糖酶和木聚糖酶辅助酶、乙酰乳酸脱羧酶、环糊精糖基转移酶、脂酶、植酸酶、漆酶、氧化酶、酯酶、角质酶、颗粒淀粉水解酶和其它葡糖淀粉酶。

[0144] 在一些方面,由重组细胞产生的葡糖淀粉酶可用于淀粉转化工艺,诸如用于产生用于果糖糖浆的右旋糖、特种糖和乙醇以及其它最终产物(例如,有机酸、抗坏血酸和氨基酸)。使用本公开的葡糖淀粉酶通过淀粉底物的发酵产生醇可包括产生燃料醇或饮用醇。

[0145] 与亲本菌株相比,当在相同条件下使用具有至少四种外源核酸(各自具有不同的核酸序列,但编码相同或类似的葡糖淀粉酶多肽)的酵母属菌种时乙醇产量更高。例如,使用本公开的工程化酵母属时,乙醇产量增加,其可以是野生型菌株乙醇产量的1.1倍或更高、1.2倍或更高、1.3倍或更高、1.4倍或更高、1.5倍或更高、1.6倍或更高、1.7倍或更高、1.7倍或更高、1.8倍或更高、1.9倍或更高、2.0倍或更高、2.1倍或更高、2.2倍或更高、2.3倍或更高、2.4倍或更高、或者2.5倍或更高。

[0146] 在一些方面,本公开提供了通过发酵产生乙醇的方法,其中乙醇以90g/L或更高的浓度存在于发酵培养基中。在该方法中,使包含淀粉材料和工程化酵母的液体培养基发酵,所述工程化酵母诸如为具有至少四种外源核酸的工程化酵母属菌种,所述核酸各自具有不同的核酸序列,但编码相同或相似葡糖淀粉酶多肽。发酵可以在液体培养基中提供约90g/L或更高的乙醇浓度,诸如在约90g/L至约170g/L的范围内,在约110g/L至约170g/L的范围内,在约125g/L至约170g/L的范围内,在约130g/L至约170g/L的范围内,或在约140g/L至约170g/L的范围内。

[0147] 在一些方面,发酵方法使用工程化酵母,该工程化酵母具有提供异源葡糖淀粉酶的至少一种基因修饰,并且在整个发酵周期中向发酵培养基添加包含部分水解淀粉的液体培养基。异源葡糖淀粉酶分泌到培养基中,并且基因修饰在发酵周期内稳定,以在五天或更长的发酵周期后在发酵培养基中提供浓度为130g/L或更高的乙醇。

[0148] 本公开的菌株也可用于发酵过程中,以提供与野生型菌株相比更少量的甘油产量。甘油是乙醇厌氧产生期间的重要副产物,并且希望最小化甘油产量,因为这样使从所需生物产物(诸如,乙醇)的产生中被转移用于甘油产生的糖原料更少。令人惊讶的是,本公开的菌株在发酵过程中表现出甘油的显著降低。例如,在发酵过程中,菌株可以产生少于10g/L的甘油、少于约9g/L的甘油、少于约8g/L的甘油或少于约7g/L的甘油。在发酵过程中,使用本公开的生物得到的甘油产量也可以相对于一种或多种生物产物(诸如,乙醇)表示。例如,在一些实践模式中,在发酵过程中,乙醇与甘油之比大于约13:1,或大于约14:1,诸如在约14:1至约25:1的范围内。

[0149] 该方法包括使包含淀粉材料和工程化酵母的液体培养基发酵,所述工程化酵母包含编码多肽的外源核酸,该多肽包含葡糖淀粉酶部分的与葡糖淀粉酶异源的信号序列,其

中所述发酵在液体培养基中提供浓度为90g/L或更高的乙醇。

[0150] 发酵产物首先可以在处理系统中用一种或多种试剂处理。经处理的发酵产物然后可以送到蒸馏系统中。在蒸馏系统中,发酵产物可以被蒸馏并脱水为乙醇。在一些方面,从发酵培养基中去除的组分包括水、可溶性组分、油和未发酵固体。这些组分中的一些可以用于其它目的,诸如用于动物饲料产品。其它联产物(例如,糖浆)可以从釜馏物中回收。

[0151] 回糟(Backset)是在蒸馏过程中去除生物产物(例如,乙醇)之后的剩余发酵液。洗涤水是从洗涤器收集的液体。本公开的常用术语的定义可见于The Alcohol Textbook, 4<sup>th</sup> Edition, 1995(《酒精教科书》,第4版,1995年)。

[0152] 本公开还提供了用于产生食品、饲料或饮料产品的方法,所述饮料产品诸如为酒精或非酒精饮料,诸如基于谷物或麦芽的饮料如啤酒或威士忌,诸如葡萄酒、苹果酒、醋、米酒、酱油或果汁,所述方法包括用本文所述的组合物处理含淀粉和/或糖的植物材料的步骤。在另一方面,本发明还涉及包含本公开的葡糖淀粉酶或本文所设想的组合物的试剂盒;以及所述葡糖淀粉酶或组合物的使用说明。本发明还涉及通过使用葡糖淀粉酶的方法产生的发酵饮料。

[0153] 发酵过程完成后,可以使用存在于发酵培养基中的材料。在一些方面,在完成发酵过程之后,或者当发酵过程正在进行时,可以从发酵培养基中去除生物产物中的一些或全部,从而得到包含非生物产物固体的精制组合物。该非生物产物固体包含工程化酵母、培养基中未被酵母利用的原料以及发酵联产物。这些材料可以提供碳水化合物源和蛋白质源,所述碳水化合物源和蛋白质源可用作补充剂来提高饲料组合物的营养含量。饲料材料可以是来自发酵过程的联产物,诸如釜馏物(全釜馏物、稀釜馏物等)或由其制备的组合物,包括干酒糟(DDG)、含可溶物的干酒糟(DDGS)、湿酒糟(DWG)和酒糟可溶物(DS)。

[0154] 可进一步处理任选地去除了部分或全部目标生物产物的发酵培养基(诸如)以便去除水,或引起非生物产物固体从培养基中沉淀出来或与培养基分离。在一些情况下,通过冷冻干燥或烘箱干燥处理培养基。处理后,精制组合物可以是例如液体浓缩物、半湿滤饼或干燥固体的形式。精制组合物可以用作饲料组合物本身,或者用于制备饲料组合物的成分。在优选的制剂中,饲料组合物是诸如用于羊、牛、猪等的牲畜饲料组合物。

[0155] 发酵培养基中的固体可以提供一种或多种氨基酸的来源。引入动物饲料中之后,发酵联产物可以对于一种或多种必需氨基酸提供增强的氨基酸含量。必需氨基酸可包括组氨酸、异亮氨酸、赖氨酸、甲硫氨酸、苯丙氨酸、苏氨酸和色氨酸。这些氨基酸可作为游离氨基酸存在于饲料组合物中,或者可以来源于富含氨基酸的蛋白质或肽。发酵培养基中的固体可以提供一种益生元的来源,这种益生元是不易消化的食物物质,诸如不易消化的寡糖,其选择性地刺激肠道中的有利细菌菌种的生长,从而有益于宿主。发酵培养基中的固体可以提供植酸酶、 $\beta$ -葡聚糖酶、蛋白酶和木聚糖酶的来源。

[0156] 饲料组合物可用于水产养殖,以养殖水生生物诸如鱼类、贝类或植物。水产养殖包括海洋物种和淡水物种的养殖,并且范围可从陆地到远洋产生。

[0157] 除了从发酵培养基获得的材料之外,饲料组合物可包含一种或多种饲料添加剂。可使用饲料添加剂以例如帮助提供均衡的饮食(例如,维生素和/或微量矿物质)、保护动物免受疾病和/或应激(例如,抗生素、益生菌)以及/或者刺激或控制生长和行为(例如激素)。添加剂产品成分可包括例如:生长促进剂、药用物质、缓冲剂、抗氧化剂、酶、防腐剂、粒料粘

结剂、直接饲喂的微生物等。添加剂产品成分还可包括例如离子载体(例如,莫能菌素、拉沙里菌素、莱特洛霉素等)、 $\beta$ -激动剂(齐帕特罗、莱克多巴胺等)、抗生素(例如,氯四环素(CTC)、土霉素、杆菌肽、泰乐菌素、金霉素)、益生菌和酵母培养物、抗球虫剂(例如,氨丙啉、癸氧喹酯、拉沙里菌素、莫能菌素)和激素(例如,生长激素或抑制发情和/或排卵的激素,诸如醋酸美仑孕酮)、信息素、营养品、药物、类黄酮、营养性和非营养性补充剂、解毒剂等。一些可商购获得的添加剂以商品名Rumensin<sup>®</sup>、Bovatec<sup>®</sup>、Deccox<sup>®</sup>、Tylan<sup>®</sup>、Optaflexx<sup>®</sup>和MGA<sup>®</sup>出售。

[0158] 实施例1

[0159] 淀粉分解酿酒酵母菌株的制备

[0160] 菌株1-3:ura3 $\Delta$ 酿酒酵母基础菌株

[0161] 用SEQ ID NO:1转化菌株1(Ethanol Red<sup>™</sup>)。SEQ ID NO:1包含以下要素:i)来自酿酒酵母的3-脱氧-D-阿拉伯庚酮糖-7-磷酸(DAHP)合酶基因的突变体形式的表达盒(ARO4-OPF);以及ii)用于将靶向染色体整合到URA3基因座中的侧接DNA。在包含3.5g/L对氟苯丙氨酸和1g/L L-酪氨酸的完全合成培养基(ScD-PFP)上选择转化体。将所得转化体在ScD-PFP上划线培养,分离出单个菌落。选择单个菌落。通过对单个菌落的PCR验证SEQ ID NO:1是否正确整合到URA3的一个等位基因中。该经PCR验证的分离物称为菌株1-1。

[0162] 用SEQ ID NO:2转化菌株1-1。SEQ ID NO:2包含以下要素:i)来自构巢曲霉的乙酰胺酶(amdS)基因的表达盒;以及ii)用于将靶向染色体整合到URA3基因座中的侧接DNA。在包含80mg/L尿嘧啶和1g/L乙酰胺作为唯一氮源的酵母氮源基础(不含硫酸铵或氨基酸)上选择转化体。将所得转化体在包含80mg/L尿嘧啶和1g/L乙酰胺作为唯一氮源的酵母氮源基础(不含硫酸铵或氨基酸)上划线培养,分离出单个菌落。选择单个菌落。通过对单个菌落的PCR验证SEQ ID NO:2是否正确整合到基因座A的第二个等位基因中。该经PCR验证的分离物称为菌株1-2。

[0163] 用SEQ ID NO:3和SEQ ID NO:4共转化到菌株1-2。SEQ ID NO:3包含以下要素:i)来自P1噬菌体的cre重组酶的开放阅读框,以及ii)与SEQ ID NO:4同源的侧接DNA。SEQ ID NO:4包含以下要素:i)2 $\mu$ 复制起点;ii)来自酿酒酵母的URA3选择性标记;以及iii)包含来自酿酒酵母的PGK启动子和CYC1终止子的侧接DNA。在不含尿嘧啶的缺陷型合成培养基(ScD-Ura)上选择转化体。将所得转化体在ScD-Ura上划线培养,分离出单个菌落。选择单个菌落。将分离出的菌落在ScD-PFP和包含80mg/L尿嘧啶和1g/L乙酰胺作为唯一氮源的酵母氮源基础(不含硫酸铵或氨基酸)上筛选出生长缺乏的菌落。通过PCR验证ARO4-OPF和amdS基因的丢失。将经PCR验证的分离物在包含5-FOA的YNB上划线培养,以选择出丢失2 $\mu$ 质粒的分离物。该经PCR验证的分离物称为菌株1-3。

[0164] 菌株1-4:表达在CYB2的一个等位基因处的扣囊复膜酵母葡糖淀粉酶的两个密码子优化变体的酿酒酵母

[0165] 用SEQ ID NO:16和SEQ ID NO:17共转化菌株1-3。SEQ ID NO:16包含以下要素:i)与天然CYB2基因的5'区同源的DNA;以及ii)扣囊复膜酵母葡糖淀粉酶的独特密码子优化变体的表达盒,受TDH3启动子和CYC1终止子控制;以及iii)URA3启动子以及URA3基因的一部分。SEQ ID NO:17包含以下要素:i)URA3基因的一部分和终止子;以及ii)扣囊复膜酵母葡

糖淀粉酶的独特密码子优化变体的表达盒,受PGK启动子和RPL3终止子控制;以及iii)与天然CYB2基因的3'区同源的DNA。在ScD-Ura上选择转化体。将所得转化体在ScD-Ura上划线培养,分离出单个菌落。选择单个菌落。通过PCR验证SEQ ID NO:16和SEQ ID NO:17在CYB2的一个等位基因处是否正确整合。该经PCR验证的分离物称为菌株1-4。

[0166] 菌株1-5:表达在CYB2的两个等位基因处的扣囊复膜酵母葡糖淀粉酶的四个密码子优化变体的酿酒酵母

[0167] 用SEQ ID NO:18和SEQ ID NO:19共转化菌株1-4。SEQ ID NO:18包含以下要素:i)与天然CYB2基因的5'区同源的DNA;以及ii)扣囊复膜酵母葡糖淀粉酶的独特密码子优化变体的表达盒,受TDH3启动子和CYC1终止子控制;以及iii)TEF1启动子和构巢曲霉乙酰胺酶基因(amdS)的一部分。SEQ ID NO:19包含以下要素:i)构巢曲霉乙酰胺酶基因(amdS)的一部分和ADH1终止子;以及ii)扣囊复膜酵母葡糖淀粉酶的独特密码子优化变体的表达盒,受PGK启动子和RPL3终止子控制;以及iii)与天然CYB2基因的3'区同源的DNA。在包含80mg/L尿嘧啶和1g/L乙酰胺作为唯一氮源的酵母氮源基础(不含硫酸铵或氨基酸)上选择转化体。将所得转化体在包含80mg/L尿嘧啶和1g/L乙酰胺作为唯一氮源的酵母氮源基础(不含硫酸铵或氨基酸)上划线培养,分离出单个菌落。选择单个菌落。通过PCR验证SEQ ID NO:18和SEQ ID NO:19在CYB2的其余等位基因处是否正确整合。该经PCR验证的分离物称为菌株1-5。图1是工程化CYB2基因座的示意图,示出密码子优化GA变体和标记基因的排列方式。

[0168] 菌株1-6:通过菌株1-5中的cre重组酶再循环利用URA3和amdS标记

[0169] 用SEQ ID NO:9转化菌株1-5。SEQ ID NO:9包含以下要素:i)来自酿酒酵母的3-脱氧-D-阿拉伯庚酮糖-7-磷酸(DAHP)合酶基因的突变体形式的表达盒(ARO4-OFP);2)来自P1噬菌体的cre重组酶的表达盒;3)包含天然URA3的表达盒;以及4)酿酒酵母CEN6着丝粒。在包含3.5g/L对氟苯丙氨酸和1g/L L-酪氨酸的完全合成培养基(ScD-PFP)上选择转化体。将所得转化体在ScD-PFP上划线培养,分离出单个菌落。选择单个菌落。该经PCR验证的分离物称为菌株1-6。

[0170] 菌株1-7:恢复菌株1-6中原始基因座处的天然URA3

[0171] 用SEQ ID NO:10转化菌株1-6。SEQ ID NO:10包含以下要素:1)天然URA3的表达盒,其5'区和3'区与菌株1-6中被破坏的URA3基因座同源。在ScD-ura上选择转化体。将所得转化体在ScD-ura上划线培养,分离出单个菌落。选择单个菌落。该经PCR验证的分离物称为菌株1-7。

[0172] 实施例2

[0173] 通过同时糖化和发酵摇瓶分析评估包含扣囊复膜酵母葡糖淀粉酶的独特密码子优化变体的多个拷贝的淀粉分解酿酒酵母菌株

[0174] 使用部分水解的玉米淀粉进行摇瓶评估

[0175] 将菌株1-5和菌株1在ScD-Ura板上划线,并在30°C下温育,直至可以看到单个菌落(1-2天)。刮取ScD-Ura板上的细胞,放入10mL无菌水中,并测量光密度(OD<sub>600</sub>)。使用Genesys20型分光光度计(赛默科技公司(Thermo Scientific)),以1cm光程长度,在600nm波长下测量光密度。

[0176] 将细胞浆液接种到摇瓶中,达到0.1-0.3的初始OD<sub>600</sub>。在即将接种之前,将50mL摇瓶培养基加入配有带透气密封件的螺旋盖(corning 1395-45LTMC)的250mL无挡板摇瓶

(Corning 4995-250) 中。摇瓶培养基由850g部分水解的玉米淀粉、150g经过滤的轻质浸泡水、25g葡萄糖和1g尿素组成。在装有菌株1的烧瓶中,于接种后立即加入17.5 $\mu$ L Dupont Distillase葡萄糖淀粉酶(每千克淀粉大约1.7g酶)。

[0177] 将接种后的烧瓶置于轨道式振荡器中,在30 $^{\circ}$ C、100rpm的转速下震荡温育67小时。获取样品,并通过高效液相色谱法结合示差折光检测器分析发酵期间发酵液中的乙醇浓度。

[0178] 如图2所示,摇瓶分析的结果显示出,与培养基中未添加商用葡萄糖淀粉酶的菌株1(即,未经过基因修饰)和培养基中添加了商用葡萄糖淀粉酶的菌株1相比,具有扣囊复膜酵母葡萄糖淀粉酶的四种独特密码子优化变体的菌株1-5在提供高乙醇滴度方面的有效性。

[0179] 实施例3

[0180] 通过同时糖化和发酵摇瓶分析评估包含扣囊复膜酵母葡萄糖淀粉酶的独特密码子优化变体的多个拷贝且恢复了URA3的淀粉分解酿酒酵母菌株

[0181] 使用部分水解的玉米淀粉进行摇瓶评估

[0182] 将菌株1-5和菌株1-7在ScD-Ura板上划线,并在30 $^{\circ}$ C下温育,直至可以看到单个菌落(1-2天)。刮取ScD-Ura板上的细胞,放入10mL无菌水中,并测量光密度(OD<sub>600</sub>)。使用Genesys20型分光光度计(赛默科技公司(Thermo Scientific)),以1cm光程长度,在600nm波长下测量光密度。

[0183] 将细胞浆液接种到摇瓶中,达到0.1-0.3的初始OD<sub>600</sub>。在即将接种之前,将50mL摇瓶培养基加入配有带透气密封件的螺旋盖(corning 1395-45LTM)的250mL无挡板摇瓶(Corning 4995-250)中。摇瓶培养基由850g部分水解的玉米淀粉、150g经过滤的轻质浸泡水、25g葡萄糖和1g尿素组成。

[0184] 将接种后的烧瓶置于轨道式振荡器中,在30 $^{\circ}$ C、100rpm的转速下震荡温育64小时。获取样品,并通过高效液相色谱法结合示差折光检测器分析发酵期间发酵液中的乙醇浓度。

[0185] 如图3所示,摇瓶的结果显示出,包含扣囊复膜酵母葡萄糖淀粉酶的四种独特密码子优化变体且在工程化CYB2基因座处具有标记的菌株1-5与包含扣囊复膜酵母葡萄糖淀粉酶的四种独特密码子优化变体且在天然基因座处恢复了URA3的菌株1-7相比的乙醇产量曲线。

[0186] 实施例4

[0187] 在连续同时糖化和发酵系统中评估菌株1和菌株1-5

[0188] 连续糖化和发酵

[0189] 连续同时糖化和发酵系统由糖化罐和串联运行的4个发酵罐(New Brunswick BioFlo 310)组成。糖化罐的工作体积为2.1-2.3升,并保持在60 $^{\circ}$ C,并连续送入液化玉米淀粉水解产物(34%溶解固体,3%葡萄糖),加入葡萄糖淀粉酶(Dextrozyme 1.5X,诺维信公司(Novozymes)),以在7小时内获得所需的右旋糖转化率(以转化为右旋糖的淀粉的重量百分比表示)。将来自糖化罐的液流的大约60%送入第一发酵罐(Ferm1)中,而40%被引入第二发酵罐(Ferm2)中。将灭菌后的轻质浸泡水(sLSW,由原始轻质浸泡水、回糟和洗涤水组成)送入Ferm1中,以获得约600ppm的游离氨基氮(FAN)目标。加入釜馏物和洗涤水作为稀释剂,以在设定的流速和保留时间下实现系统FAN和葡萄糖目标。Ferm1的工作体积为2.2升,保持在30 $^{\circ}$ C,以100RPM搅拌,并且停留时间为6.7小时。Ferm1以0.25体积空气/体积培养基(VVM)



的流速供应空气,通过发酵罐底部的环形分布器引入。Ferm1的液流流向Ferm2。Ferm2的工作体积为3.9-4.1升,保持在30℃,以100RPM搅拌,并且停留时间为8.7-9.1小时。Ferm2的液流流向发酵罐3 (Ferm3),该发酵罐的工作体积为5.4升,保持在30℃,以100RPM搅拌,并且停留时间为11.8小时。Ferm3的液流流向发酵罐4 (Ferm4),该发酵罐的工作体积为13.8升,保持在30℃,以100RPM搅拌,并且停留时间为30.5小时。

[0190] 将PDA板上的单个菌落在装有50mL酵母霉菌(YM)发酵液的250mL锥形摇瓶中,以250RPM搅拌,并且温度为30℃,使菌株1或菌株1-5生长过夜。为了使用菌株1-5接种连续系统,将来自过夜摇瓶的发酵液加入到Ferm1中以达到0.2的起始OD<sub>600</sub>,并且以批次(最初补充50g/L葡萄糖)运行,直至二氧化碳释放速率(CER)达到最高水平并开始呈现向下趋势,此后液流开始持续运行。为了使用菌株1接种连续系统,将来自过夜摇瓶的发酵液加入到Ferm1中以达到0.2的起始OD<sub>600</sub>。接种后,立即向Ferm1中以0.5g/Kg淀粉(干基)的速率添加商用葡萄糖淀粉酶(Distillase SSF,杜邦公司(Dupont)),并按针对菌株1-5所述的批次操作。对于菌株1-5,未向系统的发酵部分添加额外的GA。一旦参数保持设定条件,系统就可以持续运行至少五天。每天收集样品两次,最少收集五天,并通过HPLC(克/升)和细胞干重(CDW,克/升)分析。其结果为平均值,并示于下表中。

[0191] 表4至表6所示的发酵结果显示,相对于菌株1,菌株1-5具有相似的乙醇滴度(EtOH)、右旋糖滴度(Dx)和生长情况(CDW),但菌株1-5产生的甘油较少。

[0192] 表4:菌株1(67%+/-5%右旋糖糖化材料)连续发酵的每个阶段中相关代谢物的含量

[0193]

	Ferm1	Ferm2	Ferm3	Ferm4
Dx (g/L)	71.4+/-11.9	74.5+/-17.7	45.1+/-21.4	6.0+/-3.99
EtOH(g/L)	46.4+/-2.29	67.2+/-8.83	95.0+/-2.00	127.0+/-7.73
甘油(g/L)	6.1+/-0.56	7.3+/-0.74	8.8+/-0.59	10.2+/-0.47
CDW(g/L)	6.8+/-1.8	7.9+/-3.0	8.9+/-2.1	7.3+/-1.8

[0194] 表5:菌株1-5(62%+/-0.5%右旋糖糖化材料)连续发酵的每个阶段中相关代谢物的含量

[0195]

	Ferm1	Ferm2	Ferm3	Ferm4
Dx (g/L)	77.2+/-3.00	97.3+/-2.61	65.9+/-2.62	21.0+/-1.91
EtOH(g/L)	47.3+/-1.00	67.3+/-0.81	100.3+/-0.81	131.8+/-1.08
甘油(g/L)	4.4+/-0.15	6.0+/-0.20	5.7+/-0.21	8.8+/-0.21
CDW(g/L)	9.0+/-1.8	8.9+/-0.7	9.0+/-0.7	7.8+/-1.0

[0196] 表6:菌株1-5(45%+/-2%右旋糖糖化材料)连续发酵的每个阶段中相关代谢物的含量

[0197]

	Ferm1	Ferm2	Ferm3	Ferm4
Dx (g/L)	28.9+/-2.17	54.1+/-3.61	38.1+/-4.45	5.1+/-2.57
EtOH(g/L)	50.1+/-1.73	69.4+/-3.01	101.4+/-3.39	132.2+/-1.30

甘油 (g/L)	3.1+/-0.15	4.6+/-0.24	5.7+/-0.21	6.6+/-0.34
CDW (g/L)	9.2+/-1.3	9.5+/-0.7	9.8+/-0.4	7.9+/-0.6

## 序列表

&lt;110&gt; 嘉吉公司

C·K·米勒

&lt;120&gt; 经葡糖淀粉酶修饰的酵母菌株和用于产生生物产物的方法

&lt;130&gt; CAR0196/P1

&lt;160&gt; 19

&lt;170&gt; PatentIn 3.5版

&lt;210&gt; 1

&lt;211&gt; 3182

&lt;212&gt; DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; pAV18转化片段

&lt;400&gt; 1

```

cctactgcgc caattgatga caatacagac gatgataaca aaccgaagtt atctgatgta 60
gaaaaggatt aaagatgcta agagatagtg atgatatttc ataaataatg taattctata 120
tatgttaatt accttttttg cgaggcataat ttatgggtgaa ggataagttt tgaccatcaa 180
agaaggttaa tgtggctgtg gtttcagggt ccataaagct tttcaattca tctttttttt 240
ttttgttctt ttttttgatt ccggtttctt tgaattttt ttgattcggg aatctccgag 300
cagaaggaag aacgaaggaa ggagcacaga cttagattgg tatatatacg catatgtggg 360
gttgaagaaa catgaaattg cccagtattc ttaaccaac tgcacagaac aaaaacctgc 420
aggaaacgaa gataaagcgg ccgcataact tcgtataatg tatgctatac gaagttatct 480
gccagtatac agctagcctt gaaagtgatg gaaaacattg tcatcggcac ataaataaaa 540
aaattatgaa tcacgtgatc aacagcaaat tatgtactcg tatatatgca agcgcattcc 600
ttatattgac actctttcat tgggcatgag gctgtgtaaa cataagctgt aacggctctca 660
cggaacactg tgtagttgca ttactgtcag gcagttatgt tgcttaatat aaaggcaaag 720
gcatggcaga atcaacttaa aacgtggccc caccgctgc accctgtgca tttgttacgt 780
tactgcgaaa tgactcaacg atgaaatgaa aaaattttgc ttgaaatttt gaaaaaaga 840
tgtgcgggac gcattgttag ctattgaaat acatcgtgat cgaatccaat caatgtttaa 900
tttcatatta atacagaaac tttttctcat actttctct tcttttcatt ggtatattat 960
ctatatatcg tgtaattcc tctttcgtca tttttagcat cgttataaga gtaattaaga 1020
ataactagaa gagtctctct ttatattcgt ttattttata tatttaaccg ctaaatttag 1080
taaacaaaag aatctatcag aatgagtga atctccaatg ttcgctgcca acggcatgcc 1140
aaaggtaaat caaggtgctg aagaagatgt cagaatttta ggttacgacc cattagcttc 1200
tccagctctc cttcaagtgc aaatcccage cacaccaact tctttgaaa ctgccaagag 1260
aggtagaaga gaagctatag atattattac cggtaaagac gacagagttc ttgtcattgt 1320
cggtccttgt tccatccatg atctagaagc cgctcaagaa tacgctttga gattaaagaa 1380
attgtcagat gaattaaag gtgatttata catcattatg agagcatact tggagaagcc 1440

```

aagaacaacc gtcggctgga aaggtctaataa taatgaccct gatgttaaca acactttcaa 1500  
catcaacaag ggtttgcaat ccgctagaca attgtttgte aacttgacaa atatcggttt 1560  
gccaattggt tctgaaatgc ttgataccat ttctcetaaa tacttggctg atttggctctc 1620  
cttcgggtgcc attgggtgcca gaaccaccga atctcaactg cacagagaat tggcctccgg 1680  
tttgtctttc ccagttggtt tcaagaacgg taccgatggt accttaaatg ttgctgtgga 1740  
tgcttgctcaa gccgctgctc attctcacca tttcatgggt gttactaagc atgggtgtgc 1800  
tgctatcacc actactaagg gtaacgaaca ctgcttcggt attctaagag gtggtaaaaa 1860  
gggtaccaac tacgacgcta agtccgttgc agaagctaag gctcaattgc ctgccggttc 1920  
caacgggtcta atgattgact actctcacgg taactccaat aaggatttca gaaaccaacc 1980  
aaagggtcaat gacgttgttt gtgagcaaat cgtaacgggt gaaaacgcca ttaccgggtg 2040  
catgattgaa tcaaacatca acgaaggtaa ccaaggcatc ccagccgaag gtaaagccgg 2100  
cttgaaatat ggtgtttcca tcaactgatgc ttgtataggt tgggaaacta ctgaagacgt 2160  
cttgaggaaa ttggctgctg ctgtcagaca aagaagagaa gttaacaaga aatagatggt 2220  
tttttaataga tatatgtaac gtacattett tctctacca ctgccaattc ggtattattt 2280  
aattgtgttt agcgcatttt actaattaac tagaaactca atttttaag gcaaagctcg 2340  
ctgacctttc actgatttcg tggatgttat actatcagtt actcttctgc aaaaaaaaaat 2400  
tgagtcatat cgtagctttg ggattatfff tctctctctc cacggctaata taggtgatca 2460  
tgaaaaaaaaatg aaaaattcat gagaaaagag tcagacatcg aaacatacat aagttgatat 2520  
tcctttgata tcgacgacta ctcaatcagg ttttaaaaga aaagaggcag ctattgaagt 2580  
agcagtatcc agtttaggtt ttttaattat ttacaagtaa agaaaaagag aatgccggtc 2640  
gttcacgata acttcgtata atgtatgcta tacgaagtta tgcggcccgag agaagatgag 2700  
gccagcaaaa ctaaaaaact gtattataag taaatgcatg taaactaac tcacaaatta 2760  
gagcttcaat ttaattatat cagttattac ccgggaatct cggtcgtaat gatttctata 2820  
atgacgaaaa aaaaaaaaaatt ggaaagaaaa agcttcatgg cttttataaa aaggaactat 2880  
ccaatacctc gccagaacca agtaacagta ttttacgggg cacaaatcaa gaacaataag 2940  
acaggactgt aaagatggac gcattgaact ccaagaaca acaagagttc caaaaagtag 3000  
tggaacaaaa gcaaatgaag gatttcatgc gtttgactc taatctggta gaaagatggt 3060  
tcacagactg tgtcaatgac ttcaaacat caaagctaac caataaggaa caaacatgca 3120  
tcatgaagtg ctcaaaaaag ttcttgaagc atagcgaacg tgtagggcag cgtttccaag 3180  
ag 3182  
<210> 2  
<211> 3275  
<212> DNA  
<213> 人工序列  
<220>  
<223> pCM520转化片段  
<400> 2  
cctactgcgc caattgatga caatacagac gatgataaca aaccgaagtt atctgatgta 60  
gaaaaggatt aaagatgcta agagatagtg atgatatttc ataaataatg taattctata 120

tatgttaatt accttttttg cgaggcatat ttatggtgaa gaataagttt tgaccatcaa 180  
 agaaggtaa tgtggctgtg gtttcagggt ccataaagct tttcaattca tcattttttt 240  
 tttattcttt tttttgattc cggtttcett gaaatttttt tgattcggta atctccgaac 300  
 agaaggaaga acgaaggaag gagcacagac ttagattggt atatatacgc atatgtagtg 360  
 ttgaagaaac atgaaattgc ccagtattct taaccaact gcacagaaca aaaatctgca 420  
 ggaaacgaag ataaagcggc cgcataactt cgtatagcat acattatacgc aagttatcgc 480  
 ctgttaagat ataactgaaa aaagagggga atttttagat actgaaatga tatttttagaa 540  
 taaccagact atataaaagg ataaattaca aaaaattaac taatagataa gatttaaata 600  
 taaaagatat gcaactagaa aagtccttate aatctcetta tggagtgcgc acgttaccga 660  
 acaatttacc gacttcttcg gcgatagcca aagttctctc ttcggacaat cttctacca 720  
 taacttgaac agcaacagga gcaccgtgat aagcctctgg gtcgtattct tcttgaacca 780  
 aagcatccaa ttcggaaaca gctttaaag attcgttctt cttatcaata ttcttatcag 840  
 cgaaagtgc tgggacgaca acagaggtga aatccaata gtttaataac gaggcgtaac 900  
 cgtagtatct gaattgatcg tgtctgacag cggcggtagg agtaattgga gcgataatag 960  
 cgtccaattc cttaccagct ttttcttcag cttcacgcca cttttccaag tattccattt 1020  
 gatagttcca cttttgtaa tgagtgtccc acaattcgtt catgttaaca gccttaatat 1080  
 ttgggttcaa caagtcctta atgttaggga tggctggctc accagaggca gaaatgtctc 1140  
 tcatgacgtc ggcagaacca tcagcagcat agatgtggga aatcaagtca tgaccgaaat 1200  
 catgcttgta tggagtccat ggagtaacgg tgtgaccagc cttggccaaa gcggcaacgg 1260  
 tagtttcgac accacgtaa attggtgggt gtggcaagac gttaccgtcg aaattgtaat 1320  
 aaccaatgtt caaaccacca ttcttaatct tagaggcaat gatgtcagat tcagattgtc 1380  
 tccatggcat tgggatgacc ttagagtcgt acttccaagg ttcttgacc aagacagatt 1440  
 tggatgaaaa tctcaagtct tcgacggagt gagtgatagg accaacgacg gagtgaacgg 1500  
 tttcttgacc ttccatagag ttagccattt tagcatatgg caatctaccg tgagatggtc 1560  
 tcaaacgta taaaagttg aaagcagctg ggactctaat ggaaccacca atgtcagtac 1620  
 cgacaccaat aacaccacct ctaatacaca caatagcacc ttcaccacca gaagaaccac 1680  
 cacaggacca atttttgttt cttggattga cagttctacc aatgatgttg ttgacggttt 1740  
 cacagaccat caaggtttgt gggacagagg tcttaacgta gaaaacagca ccagcttttc 1800  
 tcaacatggt ggtaagacg gaatcacct catcgtattt gtttaaccag gaaatgtaac 1860  
 ccatggaggt ttcgtaacce ttaacacgca attggtcctt taaagagatt ggtaaaccgt 1920  
 gtaatggacc aactggctc ttatgcttag cgtagtattc atctaattct ctagcttgag 1980  
 ctaaagcagc atctgggaag aattcgtgag cacagttggt taattgttga gcaatagcag 2040  
 ctctcttaca aaaagccaaa gtgacttcaa cagaagtaa ctcaccagcg gccaaacttg 2100  
 agaccaaate agcagcagag gcttcggtaa tcttcaattc agcctcagac aaaataccgg 2160  
 acttctttgg gaaatcaata acggaatctt cggcaggcaa agtttgaacc ttccattcgt 2220  
 caggaatggt tttagccaaa cgggcacggt tgtcggcgge caattcttc caggattgtg 2280  
 gcattttgta attaaaactt agattagatt gctatgcttt ctttctaag agcaagaagt 2340  
 aaaaaagtt gtaatagaac aagaaaaacg aaactgaaac ttgagaaatt gaagaccatt 2400  
 tattaactta aatatcaatg ggaggtcatc gaaagagaaa aaaatcaaaa aaaaaatttt 2460

tcaagaaaaa gaaacgtgat aaaaattttt attgcctttt tcgacgaaga aaaagaaacg 2520  
 aggcggtctc ttttttcttt tccaaacctt tagtacgggt aattaacgcc accctagagg 2580  
 aagaaagagg ggaaatttag tatgctgtgc ttgggtgttt tgaagtggta cggcgatgcg 2640  
 cggagtccga gaaaatctgg aagagtaaaa aaggagtaga aacattttga agctatgggtg 2700  
 tgtgggggat cacttgtggg ggattgggtg tgatgtaagg ataacttcgt atagcataca 2760  
 ttatacgaag ttatgcggcc gcgagaagat gcggccagca aactaaaaa actgtattat 2820  
 aagtaaatgc atgtatacta aactcacaaa ttagagcttc aatttaatta tatcagttat 2880  
 taccgggaa tctcggtcgt aatgattttt ataatgacga aaaaaaaaa attggaaaga 2940  
 aaaagcttca tggcctttat aaaaaggaac catccaatac ctcgccagaa ccaagtaaca 3000  
 gtattttacg gggcacaaaat caagaacaat aagacaggac tgtaaagatg gacgcattga 3060  
 actccaaaga acaacaagag ttccaaaaag tagtggaaca aaagcaaatg aaggatttca 3120  
 tgcgtttgta ctctaactcg gtagaaagat gttttacaga ctgtgtcaat gacttcacaa 3180  
 catcaaagct aaccaataag gaacaaacat gcacatgaa gtgctcagaa aagttcttga 3240  
 agcatagcga acgtgtaggg cagcgtttcc aagag 3275

<210> 3

<211> 1132

<212> DNA

<213> 人工序列

<220>

<223> cre PCR片段

<400> 3

ctctttttta cagatcatca aggaagtaat tatctacttt ttacaagaat tcatgtctaa 60  
 ttacttact gttcaccaaa acttgctgc attaccagtt gacgcaacct ccgatgaagt 120  
 cagaaagaac cttatggata tgtttagaga tagacaagct ttctccgaac atacttggaa 180  
 aatgttatta tccgtttgta gatcctgggc cgcttgggtg aacttaaca atagaaaatg 240  
 gtttcctgct gaaccagaag acgtcagaga ttacttactt tacttacaag ctagaggttt 300  
 ggctgttaaa actatccaac aacacttagg tcaattgaat atgttacaca gaagatccgg 360  
 ttaccaaga ccatccgatt ccaacgcagt ttcccttgtt atgagaagaa ttagaaaaga 420  
 aaatgttgac gctggtgaaa gagctaaaca agcattagca ttgaaagaa ccgatttcga 480  
 tcaagttaga tccttaatgg aaaattccga tagatgteaa gatattagaa acttagcttt 540  
 cttaggtatt gcttacaaca cattattaag aatcgctgaa attgctagaa ttagagttaa 600  
 agatatttca agaaccgatg gcggtagaat gttaatccac attggcagaa caaaaacctt 660  
 agtctccaca gcaggcgtcg aaaaagcatt atcattaggt gttactaat tagttgaacg 720  
 ttggatttcc gtttccgggt ttgcagatga cccaacaac tacttattct gtcgtgttag 780  
 aaaaaatggg gttgccgctc cttccgetac ctcaatta tccacaagag cattagaagg 840  
 catttttgaa gctaccaca gaettattta tggtgcaaaa gacgattccg gtcaaagata 900  
 tttagcttgg tctggtcatt ccgetagagt tgggtccgca agagacatgg caagagctgg 960  
 tgtttctatt cctgaaatta tgcaagccgg tggttggact aatgttaaca ttgttatgaa 1020  
 ctatatcaga aacttagatt ccgaaacagg tgctatggtt agattacttg aagacggtga 1080

ttaagctagc taagatccgc tctaaccgaa aaggaaggag ttagacaacc tg 1132  
 <210> 4  
 <211> 6376  
 <212> DNA  
 <213> 人工序列  
 <220>  
 <223> pGP52片段  
 <400> 4  
 ctagctaaaga tccgctctaa ccgaaaagga aggagttaga caacctgaag tctaggtccc 60  
 tatttatttt tttatagtta tgtagtatt aagaacgta tttatatttc aaatTTTTct 120  
 tttttttctg tacagacgcg tgtacgcatg taacattata ctgaaaacct tgcttgagaa 180  
 ggTTTTggga cgctcgaaga tccagctgca ttaatgaate ggccaacgcg cggggagagg 240  
 cggTTTgcgt attgggcgct ctcccgcttc ctccgctact gactcctgctc gctcggctct 300  
 tcggctgcgg cgagcgggat cagctcactc aaaggcggta atacggttat ccacagaatc 360  
 aggggataac gcaggaaaga acatgtgagc aaaaggccag caaaaggcca ggaaccgtaa 420  
 aaaggccgcg ttgctggcgt tttccatag gctccgcccc cctgacgagc atcacaaaaa 480  
 tcgacgctca agtcagaggt ggcgaaaacc gacaggacta taaagatacc aggcgtttcc 540  
 ccctggaagc tccctcgtgc gctctcctgt tccgacctg ccgcttaccg gatacctgtc 600  
 cgcctttctc ccttcgggaa gcgtggcgct ttctcatagc tcacgctgta ggtatctcag 660  
 ttcggtgtag gtcgttcgct ccaagctggg ctgtgtgcac gaacccccg ttcagccccg 720  
 ccgctgcgcc ttatccgta actatcgtct tgagtccaac ccgtaagac acgacttate 780  
 gccactggca gcagccactg gtaacaggat tagcagagcg aggtatgtag gcggtgctac 840  
 agagtctctg aagtgggtgc ctaactacgg ctacactaga aggacagtat ttggtatctg 900  
 cgctctgctg aagccagtta ccttcggaaa aagagttggt agctcttgat ccggcaaaca 960  
 aaccaccgct ggtagcggtg gttttttgt ttgcaagcag cagattacgc gcagaaaaaa 1020  
 aggatctcaa gaagatcctt tgatcttttc tacggggtct gacgctcagt ggaacgaaaa 1080  
 ctacagttaa gggattttgg tcatgagatt atcaaaaagg atcttcacct agatcctttt 1140  
 aaattaaaaa tgaagtttta aatcaatcta aagtataat gagtaaactt ggtctgacag 1200  
 ttaccaatgc ttaatcagtg aggcacctat ctacagatc tgtctatttc gttcatccat 1260  
 agttgcctga ctccccgctg ttagataaac tacgatacgg gagggettac catctggccc 1320  
 cagtgtgca atgataccgc gagaccacg ctaccggct ccagatttat cagcaataaa 1380  
 ccagccagcc ggaaggccg agcgcagaag tggctctgca actttatccg cctccatcca 1440  
 gtctattaat tgttgccggg aagctagagt aagtagttcg ccagttaata gtttgcgcaa 1500  
 cgttgttgcc attgctacag gcacgtgggt gtcacgctcg tcgtttggta tggcttcatt 1560  
 cagctccggt tccaacgat caaggcgagt tacatgatec cccatgttgt gcaaaaaagc 1620  
 ggtagctcc ttcggtcctc cgatcgttgt cagaagtaag ttggcccgag tgttatact 1680  
 catggttatg gcagcactgc ataattctct tactgtcatg ccatccgtaa gatgcttttc 1740  
 tgtgactggg gactactcaa ccaagtcatt ctgagaatag tgtatgcggc gaccgagttg 1800  
 ctcttgcccc gcgtcaatac gggataatac cgcgccacat agcagaactt taaaagtgt 1860

catcattgga aaacgttctt cggggcgaaa actetcaagg atettaccgc tgttgagatc 1920  
 cagttcgatg taaccacctc gtgcaccaa ctgatcttca gcattcttta ctttcaccag 1980  
 cgtttctggg tgagcaaaaa caggaaggca aaatgccgca aaaaaggaa taaggcgac 2040  
 acggaaatgt tgaatactca tactcttctt tttcaatat tattgaagca tttatcaggg 2100  
 ttattgtctc atgagcggat acatatttga atgtatttag aaaaataaac aaatagggtt 2160  
 tccgcgcaca tttccccgaa aagtgccacc tgaacgaagc atctgtgctt cattttgtag 2220  
 aacaaaaatg caacgcgaga gcgctaattt ttcaaacaaa gaatctgagc tgcattttta 2280  
 cagaacagaa atgcaacgcg aaagcgtctat ttaccaacg aagaatctgt gcttcatttt 2340  
 tgtaaaacaa aaatgcaacg cgagagcgtt aatttttcaa acaaagaatc tgagctgcat 2400  
 ttttacagaa cagaaatgca acgcgagagc gctattttac caacaaagaa tctatacttc 2460  
 tttttgttc tacaaaaatg catcccgaga gcgtattttt tctaacaaag catcttagat 2520  
 tacttttttt ctctttgtg cgtctataaa tgcagtctct tgataacttt ttgcaactga 2580  
 ggtccgttaa ggttagaaga aggetacttt ggtgtctatt ttctcttcca taaaaaagc 2640  
 ctgactccac ttcccgcgtt tactgattac tagcgaagct gcgggtgcat tttttcaaga 2700  
 taaaggcatc cccgattata ttctataacc atgtggattg cgcatacttt gtgaacagaa 2760  
 agtgatagcg ttgatgattc ttcatgggtc agaaaattat gaacggtttc ttctattttg 2820  
 tctctatata ctacgtatag gaaatgttta cattttcgta ttgttttcca ttactctat 2880  
 gaatagttct tactacaatt tttttgtcta aagagtaata ctagagataa acataaaaaa 2940  
 tgtagaggtc gaggtttagat gcaagttcaa ggagcgaaag gtggatgggt aggttatata 3000  
 gggatatagc acagagatat atagcaaaga gatacttttg agcaatgttt gtggaagcgg 3060  
 tattcgaat atttttagtag ctctgttacag tccggtgcgt ttttggtttt ttgaaagtgc 3120  
 gtcttcagag cgcttttggg tttcaaaagc gctctgaagt tctatactt tctagagaat 3180  
 aggaacttcg gaataggaac ttcaaagcgt ttccgaaaac gagecgttcc gaaaatgcaa 3240  
 cgcgagctgc gcacatacag ctcaactgtc acgtcgcacc tatatctgcg tgttgccctgt 3300  
 atatatatat acatgagaag aacggcatag tgcgtgttta tgcttaaatt cgtacttata 3360  
 tgcgtctatt tatgtaggat gaaaggtagt ctagtacctc ctgtgatatt atcccattcc 3420  
 atgcggggta tcgtatgctt cttcagcac tacccttag ctgttctata tgctgccact 3480  
 cctcaattgg attagtctca tcttcaatg ctatcatttc ctttgatatt ggatcact 3540  
 aagaaacctt tattatcatg acattaacct ataaaaatag gcgtatcacg aggcccttc 3600  
 gtctcgcgcg tttcgggtgat gacggtgaaa acctctgaca catgcagctc ccggagacgg 3660  
 tcacagcttg tctgtaagcg gatgccggga gcagacaagc ccgtcagggc gcgtcagcgg 3720  
 gtgttgccgg gtgtcggggc tggettaact atgcggcatc agagcagatt gtactgagag 3780  
 tgcaccatac cacagctttt caattcaatt catcattttt tttttattct ttttttgat 3840  
 ttcggtttct ttgaaatttt tttgattcgg taatctccga acagaaggaa gaacgaagga 3900  
 aggagcacag acttagattg gtatatatac gcataatgtag tgttgaagaa acatgaaatt 3960  
 gccagatatt cttaacccaa ctgcacagaa caaaaacctg caggaaacga agataaatca 4020  
 tgtcgaaagc tacatataag gaacgtgctg ctactcatcc tagtctgtt gctgccaagc 4080  
 tatttaatat catgcacgaa aagcaacaaa acttgtgtgc ttcatggat gttcgtacca 4140  
 ccaaggaatt actggagtta gttgaagcat taggtcccaa aatttgttta ctaaaaacac 4200



atgtggatat cttgactgat ttttccatgg agggcacagt taagccgcta aaggcattat 4260  
 ccgccaagta caatTTTTTA ctcttcgaag acagaaaatt tgctgacatt ggtaatacag 4320  
 tcaaattgca gtactctgcg ggtgtataca gaatagcaga atgggcagac attacgaatg 4380  
 cacacgggtg ggtgggcca ggtattgta gcggtttgaa gcaggcggca gaagaagtaa 4440  
 caaaggaacc tagaggcctt ttgatgtag cagaattgtc atgcaagggc tccctatcta 4500  
 ctggagaata tactaagggt actgttgaca ttgcgaagag cgacaaagat tttgttatcg 4560  
 gctttattgc tcaaagagac atgggtggaa gagatgaagg ttacgattgg ttgattatga 4620  
 caccgggtg gggtttagat gacaaggag acgcattggg tcaacagtat agaaccgtgg 4680  
 atgatgtgg ctctacagga tctgacatta ttattgttg aagaggacta tttgcaaagg 4740  
 gaagggatgc taaggtagag ggtgaacgtt acagaaaagc aggctgggaa gcatatttga 4800  
 gaagatgcgg ccagcaaaac taaaaaactg tattataagt aatgcatgt atactaaact 4860  
 cacaaattag agcttcaatt taattatate agttattacc ctatgcggtg tgaataaccg 4920  
 cacagatgcg taaggagaaa ataccgcate aggaaattgt aaacgttaat attttgttaa 4980  
 aattcgcgtt aaatTTTTgt taaatcaget cattttttaa ccaataggcc gaaatcgga 5040  
 aaatccctta taaatcaaaa gaatagaccg agatagggtt gagtgttgtt ccagtttggga 5100  
 acaagagtcc actattaaag aacgtggact ccaacgtcaa agggcgaaaa accgtctatc 5160  
 agggcgatgg cccactacgt gaaccatcac cctaatcaag ttttttgggg tcgaggtgcc 5220  
 gtaaagcact aaatcggaac cctaaaggga gccccgatt tagagcttga cggggaaagc 5280  
 cggcgaacgt ggcgagaaa gaagggaaga aagcgaagg agcgggcgct agggcgctgg 5340  
 caagtgtagc ggtcacgctg cgcgtaacca ccacaccgc cgcgcttaat gcgccctac 5400  
 agggcgctc cattcgccat tcaggctgcg caactgttg gaaggcgat cgggtcgggc 5460  
 ctcttcgcta ttaccgcagc tgaattggag cgacctatg ctatactga gaaagcaacc 5520  
 tgacctacag gaaagagtta ctcaagaata agaattttcg ttttaaacc taagagtcac 5580  
 tttaaaattt gtatacactt atttttttaa taactattt aataataaaa atcataaatc 5640  
 ataagaaatt cgcttattta gaagtgtcaa caacgtatct accaacgatt tgaccctttt 5700  
 ccatcttttc gtaaatttct ggcaaggtag acaagccgac aaccttgatt ggagacttga 5760  
 ccaaacctct ggcaagaat tgtaattaa gccagaaaa ggaagtgtt cctccttct 5820  
 tgaattgatg ttacctcat aaagcacgtg gcctcttate gagaaagaaa ttaccgtcgc 5880  
 tcgtgatttg tttgcaaaaa gaacaaaact gaaaaaacc agacacgctc gacttctctg 5940  
 cttcctattg attgcagctt ccaatttctg cacacaaca ggtcctagcg acggctcaca 6000  
 ggttttgtaa caagcaatcg aaggttctgg aatggcgga aagggttag taccacatgc 6060  
 tatgatgcc actgtgatct ccagagcaaa gttcgttcga tcgtactgtt actctctctc 6120  
 tttcaaacag aattgtccga atcgtgtgac aacaacagcc tgttctcaca cactcttttc 6180  
 ttctaacc aa ggggttggt tagtttagta gaacctcgtg aaacttacet ttacatatat 6240  
 ataaacttgc ataaattgg caatgcaaga aatacatatt tggcttttcc taattcgtag 6300  
 ttttcaagt tcttagatgc tttcttttcc ttttttttac agatcatcaa ggaagtaatt 6360  
 atctactttt tacaag 6376

<210> 5

<211> 1548

<212> DNA

<213> 人工序列

<220>

<223> Sf GA变体#1

<400> 5

```

atgattagat taaccgtatt cctcactgca gtttttgcag cagtcgcttc ctgtgttcca 60
gttgaattgg ataagagaaa tacaggccat ttccaagcat attctggta caccgtagct 120
agatcaaact ttactcaatg gattcacgag caaccagecg tatcatggta ctatttgctt 180
cagaatatag actatccaga aggacaattc aagtctgcca agccaggggt cgttgtggct 240
tccccittta catccgaacc tgattacttc taccaatgga ctagagatac tgctatcacc 300
ttcttgtcac ttatcgcgga agttgaggat cattcttttt caaatactac actagccaag 360
gtggttgaat actacatctc taatacttac acattacaaa gagtttccaa cccatctggt 420
aacttcgaca gtccaaatca cgacggtttg ggagaaccaa agtttaatgt tgatgataca 480
gcttatactg catcttgggg tagaccacaa aatgatggcc cagcgttgag agcatacga 540
atccaagat accttaacgc agtagcaaaa cacaacaacg gtaagtact gctcgtgga 600
caaaacggta ttccttactc ttcagettct gatatctact ggaagattat caagccagat 660
cttcaacatg tgcaacceca ttggtctaca tctggttttg atttgtggga agagaatcag 720
ggaacacatt tctttactgc gttggtccag ctaaaagcac ttagttacgg cattccttta 780
agtaagacct acaacgatcc tggtttact agttggctag aaaagcaaaa ggatgcttta 840
aactcttata tcaacagctc tggtttcgta aactctggca aaaagcatat agtggagagc 900
cctcaactat cttcaagagg agggttggat agcgcacat acattgcage cttaatcaca 960
catgatattg gcgacgacga cacttacaca ctttcaacg ttgacaactc ctatgtcttg 1020
aactcactgt attaccttct agtcgataac aaaaaccgtt acaaaatcaa tggtaactac 1080
aaggccgggtg ctgctgttgg tagataccca gaggatgttt acaacgggtg tgggacatca 1140
gaaggcaatc catggcaatt agctacagcc tacgccggcc aacatttta cacactggct 1200
tacaactcat tgaaaaacaa aaaaaactta gtgattgaaa agttgaacta cgacctctac 1260
aattctttca tagcagattt atccaagatc gatagttctt acgcatcaaa agactccttg 1320
actttgacct acggttctga caactacaaa aacgtcataa agtcactatt acagtttggga 1380
gattcattcc tgaaggtcct gctcgatcac attgatgata atggacaatt aacagaagag 1440
atcaatagat acacagggtt ccaggtggt gctgtagtt tgacatggtc ctctggttca 1500
ttactttcag caaacctgct gagaaataag ttgattgaac tattgtag 1548

```

<210> 6

<211> 1548

<212> DNA

<213> 人工序列

<220>

<223> Sf GA变体#2

<400> 6

```

atgatcagac ttacagtttt cctaacagcc gttttcgccg ccggttgcate atgtgtccca 60

```

gtagaattgg ataagagaaa caccggccat ttccaagcat attcaggata caccgttgca 120  
 cgttctaatt tcacacaatg gattcatgag cagcctgctg tgtcctggta ctacttatta 180  
 caaacattg attatcctga gggacaattc aagtcagcga aaccaggcgt tgtggttgct 240  
 tctccatcca cttcagaacc agactacttc taccagtgga cccgtgacac agcaataact 300  
 ttcttatctt tgatagcaga agtagaagat cactcatttt caaatacaac tctagctaag 360  
 gttgtcgaat actacatctc taacacatac accctacaaa gagtttctaa cccatctggt 420  
 aatttcgata gcccaaatca cgatggctctg ggtgaaccaa agttcaacgt tgacgacact 480  
 gcttacactg catcatgggg cagacctcaa aacgacggtc cagccttaag agcttacgcg 540  
 atctcaagat atttgaacgc agttgccaag cataacaacg gtaagctatt gctcgcgggt 600  
 caaaatggta ttccttactc atctgcatca gatatctact ggaagattat caagccagat 660  
 ttacaacatg taagtactca ctggagtaca tctggttttg acttatggga agagaatcaa 720  
 ggtacacatt tctttactgc acttgtccag ttaaaagctc tttcatacgg tatacctttg 780  
 tctaagacat ataacgatcc aggatttact tcttggttgg aaaagcagaa ggatgccttg 840  
 aactcttaca tcaattccag cggettctgc aactccggga aaaagcacaat tgtcgaatct 900  
 cctcaattat ctagtagagg gggctcttgat agecetaact acatcctgctc tctaattaca 960  
 catgatattg gtgatgatga tacatacaact ccttttaacg tagataatc ttatgtgctg 1020  
 aactctttat actatctgct tntagacaac aaaaacagat acaagatcaa cgggaactac 1080  
 aaagcaggag ctgcagttgg tagataccca gaagatgtgt acaatggagt gggaacctca 1140  
 gagggaaacc catggcaatt ggcgacagca tacgccggcc aaacctttta cactctgct 1200  
 tacaattctc tcaaaaacaa aaaaaatttg gttattgaga agttgaatta cgatctatac 1260  
 aactccttta tagctgactt aagtaagatt gactcctett acgcttctaa ggattcattg 1320  
 acattgacct acggctcaga taactacaaa aatgctatta agtcaactttt acaattcggg 1380  
 gattctttct taaaagtctt gttggaccat attgatgata atggtcagct aacagaggaa 1440  
 atcaacagat atacaggttt tcaagctggc gcagtttccc tcaacttgag tagtggttca 1500  
 ctcttatctg caaacagagc cagaacaag ttgatcgaat tgctttag 1548

<210> 7

<211> 1548

<212> DNA

<213> 人工序列

<220>

<223> Sf GA变体#3

<400> 7

atgatcagac ttactgtttt cctcacagec gtttttgcag cagtagcttc ttgtgttcca 60  
 gttgaattgg ataagagaaa tacaggtcat ttccaagctt actctgggta cactgtggct 120  
 agatctaact tcacacaatg gattcatgaa cagcctgccg tgagttggta ctatttgcta 180  
 caaacattg attaccctga gggteaattc aaateagcta agccagggtgt tgttgctcgcg 240  
 agcccatcaa cttctgaacc agattacttc taccaatgga ctagagatac cgcaataacc 300  
 ttcttatctc taatcgcaga ggtagaagat cactcttttt caaatactac cctggcaaaa 360  
 gtggctcagat actacatctc aaacacatac accttgcaga gagtctcaaa cccatcagga 420

aacttcgatt ctccaatca tgacggctta ggagaaccaa agtttaatgt tgacgatacc 480  
gcttatactg catcttgggg tagaccacag aatgatggcc ctgccttacg tgcatacgcc 540  
atttccagat atctcaacgc tgtagcgaag cacaacaacg gtaagctgct tttagctggt 600  
caaaatggga taccatactc ttccgcttca gacatttact ggaagattat caaaccagac 660  
ttgcagcatg tcagtacaca ttgggtcaact tctggttttg atttgtggga agagaaccaa 720  
ggcactcact tctttacagc cttggttcaa ctaaaggcat tgtcttacgg aatccctttg 780  
tccaagacat acaatgatcc tggattcact agttggctag aaaagcaaaa ggatgcactg 840  
aactcataca ttaacagttc aggctttgtg aactccggta aaaagcatat tgttgaaagc 900  
ccacaactat ctagcagagg tggtttagat tctgcaacct acatagcagc cttgatcaca 960  
cacgacattg gggatgacga tacatacaca ccattcaacg tcgacaattc atacgttttg 1020  
aatagcttat actacctact ggtagataac aaaaacagat ataagatcaa tggcaactac 1080  
aaggccggtg ctgccgtagg aagataccct gaagatgtct acaacggagt tggtaacatca 1140  
gaaggtaacc catggcaatt agcaacagea tatgcccggc agacatttta cactttggct 1200  
tacaattcat tgaaaaacaa aaaaaattta gtgatagaaa agcttaacta tgacctttac 1260  
aactctttca ttgccgattt atccaagatt gattctctct acgcatcaaa ggactccttg 1320  
acacttacat acggttctga caactacaaa aatgttatca agtctctctt gcaatttggg 1380  
gattctttct tgaaggtttt actcgatcat atcgatgata atgggtcaact aactgaggaa 1440  
atcaacagat aactgggtt ccaagctgga gctgtctctt taacatggag ttcaggaggt 1500  
ttgttatctg ctaacagagc gcgtaacaaa cttattgagc ttctgtag 1548

<210> 8

<211> 1548

<212> DNA

<213> 人工序列

<220>

<223> Sf GA变体#4

<400> 8

atgattagat taacagtatt tcttacagcc gtttctcgag ccgtcgcatac ctgtgttcca 60  
gtagaattag ataagcgtaa tacaggacat tttcaagctt actctggcta tacagttgag 120  
agatctaact ttacacaatg gattcacgaa cagccagcag tttcttggtg ctatttgctc 180  
caaaacatcg actaccctga aggccaattc aagtctgcaa agccaggagt ggtcgtcgtc 240  
tctcctagta cttcagaacc agattacttc taccagtgga caagagacac tgctattacc 300  
ttcctgagct taatcgctga agttgaagat cactcttttt ctaatacaac actggccaaa 360  
gtagttgagt actacatctc taacaettac actctacaaa gagtgtcaaa cccttctggg 420  
aacttcgaca gcccacaa tgatggtttg ggggagccaa aattcaacgt tgatgataca 480  
gcctacaccg catcttgggg tagaccacaa aacgacggac cagctttaag agcatacgca 540  
atatctcgtt accttaatgc tgttgcaaaag cacaataatg gaaagtgtt gttggctggt 600  
caaaacggta ttccttactc ttcagcatct gatattctact ggaagattat caagccagat 660  
cttcaacacg tatccacaca ttgggtcaacc tccggtctcg atttatggga ggaaaatcag 720  
ggtacacatt tcttcaccgc tctagtgcaa ttgaaggctt tgagttacgg cattccattg 780

tctaagactt acaacgatcc tggtttcacc tcatggcttg aaaagcagaa ggatgccctg 840  
aatagctaca tcaactcadc tggttttggt aactcagga aaaagcatat agttgaatcc 900  
ccacaactat catcaagagg aggttttagac tccgccacat acattgctgc cttgattaca 960  
catgatattg gggatgatga cacatatact ccatttaacg tcgataacag ttatgtcctt 1020  
aattccttat actatttggt ggtcgataac aaaaatagat acaaaatcaa cggcaactac 1080  
aaggctggcg cagcgggtgg tagataccct gaggatgttt acaatgggtg aggtacatct 1140  
gaaggcaatc catggcaatt agcgactgct tacgctggac aaactttcta cacacttgcg 1200  
tacaactcat tgaaaaacaa aaaaaaccta gtcattgaaa agttgaatta cgatctgtac 1260  
aactctttca tcgcagacct atcaaagatt gactcatctt atgcaagtaa agattcacta 1320  
actttaacct acggtagtga taactacaaa aacgttatea agtctttact ccagtttgggt 1380  
gattcattct tgaagggtgt gttagatcat atagacgaca atggtcaact cacagaggag 1440  
ataaacagat acaactgggtt tcaagcagga gctgtttcac ttacttggtc aagtggttct 1500  
ttgctttccg ccaacagagc cagaaacaag ctcatcgaat tactatag 1548

<210> 9

<211> 8719

<212> DNA

<213> 人工序列

<220>

<223> pAV47

<400> 9

atcacatagg aagcaacagg cgcgttgac ttttaatttt cgaggaccgc gaatccttac 60  
atcacacca atccccaca agtgatcccc cacacacat agettcaaaa tgtttctact 120  
ccttttttac tcttcagat tttctcggac tccgcgcatc gccgtaccac ttcaaaacac 180  
ccaagcacag catactaaat ttcccctctt tcttcteta ggggtgctgtt aattaccctg 240  
actaaagggt tggaaaagaa aaaagagacc gcctcgtttc tttttcttcg tcgaaaaggg 300  
caataaaaaat ttttatcacg tttcttttctc ttgaaaattt ttttttttga tttttttctc 360  
tttcgatgac ctcccattga tatttaagtt aataaacgggt cttcaatttc tcaagtttca 420  
gtttcatttt tcttgttcta ttacaacttt ttttacttct tgctcattag aaagaaagca 480  
tagcaatcta atctaagttt taattacaaa tctagaatga gtgaatctcc aatgttcgct 540  
gccaacggca tgccaaaggt aaatcaaggt gctgaagaag atgtcagaat ttaggttac 600  
gaccatttag cttctccagc tctccttcaa gtgcaaatcc cagccacacc aacttctttg 660  
gaaactgcca agagaggtag aagagaagct atagatatta ttaccggtaa agacgacaga 720  
gttcttgtca ttgtcggtec ttgttccate catgatcttg aagccgctca agaatacgtc 780  
ttgagattaa agaaattgtc agatgaatta aaaggtgatt tatccatcat tatgagagca 840  
tacttggaga agccaagaac aaccgtcggc tggaaaggtc taattaatga ccctgatgtt 900  
aacaacactt tcaacatcaa caagggtttg caatccgcta gacaattggt tgtaacttg 960  
acaaatatcg gtttgccaat tggttctgaa atgcttgata ccatttctcc taaatacttg 1020  
gctgatattg tctccttcgg tgccattgggt gccagaacca ccgaatctca actgcacaga 1080  
gaattggcct ccggtttgct tttcccagtt ggtttcaaga acggtaccga tggtaacctta 1140

aatgttgctg tggatgcttg tcaagccgct gctcattctc accatttcat ggggtgttact 1200  
aagcatgggtg ttgctgctat caccactact aagggtaacg aacactgctt cgttattcta 1260  
agagggtgta aaaagggtag caactacgac gctaagtccg ttgcagaagc taaggctcaa 1320  
ttgectgccg gttccaacgg tctaagtatt gactactctc acggtaactc caataaggat 1380  
ttcagaaacc aaccaaaggt caatgacggt gtttgtgagc aaatcgctaa cggtgaaaac 1440  
gccattaccg gtgtcatgat tgaatcaaac atcaacgaag gtaaccaagg catcccagcc 1500  
gaaggtaaag ccggcttgaa atatggtggt tccatcactg atgcttgat aggttgggaa 1560  
actactgaag acgtcttgag gaaattggct gctgctgtca gacaaagaag agaagttaac 1620  
aagaaataga tgttttttta atgatatatg taacgtacat tctttctct accactgcca 1680  
attcgggtatt atttaattgt gtttagcgtt atttactaat taactagaaa ctcaattttt 1740  
aaaggcaaag ctgctgacc tttcactgat ttcgtggatg ttatactatc agttactctt 1800  
ctgcaaaaaa aaattgagtc atatcgtagc tttgggatta tttttctctc tctccacggc 1860  
taattaggtg atcatgaaaa aatgaaaaat tcatgagaaa agagtcagac atcgaaacat 1920  
acataagttg atattccttt gatatcgacg actactcaat caggttttaa aagaaaagag 1980  
gcagctattg aagtagcagt atccagttta ggttttttaa ttatttaca gtaaagaaaa 2040  
agagaatgcc ggtcgttcac ggcggccgcg ccagaaaaag gaagtgttc cctccttctt 2100  
gaattgatgt taccctcata aagcacgtgg cctcttatcg agaaagaaat taccgtcgtc 2160  
cgtgatttgt ttgcaaaaag aacaaaactg aaaaaacca gacacgctcg acttctgtc 2220  
ttcctattga ttgcagcttc caatttcgtc acacaacaag gtccctagcga cggctcacag 2280  
gttttgtaac aagcaatcga aggttctgga atggcgggaa aggttttagt accacatgct 2340  
atgatgcca ctgtgatctc cagagcaaag ttcgttcgat cgtactgtta ctctctctct 2400  
ttcaaacaga attgtccgaa tcgttgaca acaacagcct gttctcacac actcttttct 2460  
tctaaccaag ggggtggttt agtttagtag aacctcgtga aacttacatt tacatatata 2520  
taaacttgca taaattggtc aatgcaagaa atacatattt ggtcttttct aattcgtagt 2580  
ttttcaagtt cttagatgct ttcttttct cttttttaca gatcatcaac tcttttttac 2640  
agatcatcaa ggaagtaatt atctactttt tacaagaatt catgtctaatt ttacttactg 2700  
ttcaccaaaa ctgacctgca ttaccagttg acgcaacctc cgatgaagtc agaaagaacc 2760  
ttatggatat gtttagagat agacaagctt tctccgaaca tacttgaaa atgttattat 2820  
ccgtttgtag atcctgggcc gcttggtgta aacttaaca tagaaaatgg tttcctgctg 2880  
aaccagaaga cgtcagagat tacttacttt acttacaagc tagaggtttg gctgttaaaa 2940  
ctatccaaca aacttaggt caattgaata tgttacacag aagatccggt ttaccaagac 3000  
catccgattc caacgcagtt tcccttgta tgagaagaat tagaaaagaa aatgttgacg 3060  
ctggtgaaaag agctaaacaa gcattagcat ttgaaagaac cgatttcgat caagtttagat 3120  
ccttaatgga aaattccgat agatgtcaag atattagaaa cttagctttc ttaggtattg 3180  
cttacaacac attattaaga atcgetgaaa ttgctagaat tagagttaa gatatttcaa 3240  
gaaccgatgg cggtagaatg ttaatccaca ttggcagaac aaaaacctta gtctccacag 3300  
caggcgtcga aaaagcatta tcattaggtg ttactaaatt agttgaacgt tggatttccg 3360  
tttccggtgt tgcagatgac ccaaacact acttattctg tcgtgttaga aaaaatggtg 3420  
ttgccgctcc ttccgctacc tcacaattat ccacaagagc attagaagc atttttgaag 3480

ctaccacag acttatttat ggtgcaaaag acgattccgg tcaaagatat ttagcttggt 3540  
ctggtcattc cgctagagtt ggtgccgcaa gagacatggc aagagctggt gtttctattc 3600  
ctgaaattat gcaagccggt ggttgacta atgtaacat tgttatgaac tatatcagaa 3660  
acttagattc cgaaacaggt gctatggta gattacttga agacggtgat taagctagct 3720  
aagatccgct ctaaccgaaa aggaaggagt tagacaacct gaagtctagg tccctattta 3780  
tttttttata gttatgtag tattaagaac gttatttata tttcaaattt ttcttttttt 3840  
tctgtacaga cgcgtgtacg catgtaacat tatactgaaa accttgcttg agaaggtttt 3900  
gggacgctcg aaggagctcc aattcgcct atagtgagtc gtattacaat tcaactggccg 3960  
tcgttttaca acgctgtgac tgggaaaacc ctggcgctac ccaacttaat cgccttgacg 4020  
cacatcccc cttcgcagc tggcgtaata gcgaagagge ccgcaccgat cgcccttccc 4080  
aacagttgcg cagcctgaat ggcgaatggc gcgacgcgc ctgtageggc gcattaagcg 4140  
cggcggggt ggtggttacg cgcagcgtga ccgctacact tgccagegcc ctagegcccc 4200  
ctcctttcgc tttcttcct tcctttctcg ccacgttcgc cggtttccc cgtcaagctc 4260  
taaactcggg gctccctta gggttccgat ttagtgcttt acggcacctc gaccccaaaa 4320  
aacttgatta gggtagtggt tcacgtagtg ggccatcgcc ctgatagacg gtttttcgcc 4380  
ctttgacgtt ggagtccacg ttctttaata gtggactctt gttccaaact ggaacaacac 4440  
tcaaccctat ctcggtctat tcttttgatt tataaggat tttgccgatt tcggcctatt 4500  
ggttaaaaaa tgagctgatt taacaaaaat ttaacgcgaa ttttaacaaa atattaacgt 4560  
ttacaatttc ctgatgcggt atttctcct tacgcatctg tgcggtattt cacaccgcag 4620  
ggtaataact gatataatta aattgaagct ctaatttggt agtttagtat acatgcattt 4680  
acttataata cagtttttta gttttgctgg ccgcatctc tcaaataatgc tteccagcct 4740  
gcttttctgt aacgctcacc ctctacctta gcatecctc cctttgcaaa tagtctctt 4800  
ccaacaataa taatgtcaga tcctgtagag accacatcat ccacggttct atactgttga 4860  
cccaatgcgt ctcccttgc atctaaacc acaccgggtg tcataatcaa ccaatcgtaa 4920  
ccttcatctc ttccaccat gtctcttga gcaataaagc cgataacaaa atctttgtcg 4980  
ctcttcgcaa tgtcaacagt acccttagta tattctccag tagataggga gcccttgcat 5040  
gacaattctg ctaacatcaa aaggcctcta ggttccttg ttacttcttc tgccgcctgc 5100  
ttcaaaccgc taacaatacc tgggccacc acaccgtgt cattcgtaat gtctgcccatt 5160  
tctgctattc tgtatacacc cgcagagtac tgcaatttga ctgtattacc aatgtcagca 5220  
aattttctgt cttcgaagag taaaaaattg tacttggegg ataatgcctt tagcggctta 5280  
actgtgccct ccatggaaaa atcagtcaag atateccat gtgttttttag taacaaatt 5340  
ttgggacctt atgcttcaac taactccagt aattccttg tggtagaac atccaatgaa 5400  
gcacacaagt ttgtttgett ttcgtgcatg atattaata gcttggcagc aacaggacta 5460  
ggatgagtag cagcacgttc cttatatgta gcttcgaca tgatttatct tcgtttctcg 5520  
caggtttttg ttctgtgcag ttgggttaag aatactgggc aatttcatgt ttcttcaaca 5580  
ctacatatgc gtatatatac caatetaagt ctgtgetct tccttcttc ttcttctgt 5640  
tcggagatta ccgaatcaaa aaaatttcaa agaaaccgaa atcaaaaaaa agaataaaaa 5700  
aaaaatgatg aattgaattg aaaagcgtgg tgcactctca gtacaatctg ctctgatgcc 5760  
gcatagttaa gccagcccc acaccgcca acaccgctg acgcgcctg acgggcttgt 5820

ctgctcccgg catccgctta cagacaagct gtgaccgtct ccgggagctg catgtgtcag 5880  
 aggttttcac cgtcatcacc gaaacgcgcg agacgaaagg gcctcgtgat acgcctattt 5940  
 ttataggtta atgtcatgat aataatggtt tcttaggacg gatcgttgc ctgtaactta 6000  
 cacgcgcctc gtatctttta atgatggaat aatttgggaa tttactctgt gtttatttat 6060  
 ttttatgttt tgtatttga ttttagaaag taaataaaga aggtagaaga gttacggaat 6120  
 gaagaaaaaa aaataaacia aggtttaaaa aatttcaaca aaaagcgtac tttacatata 6180  
 tatttattag acaagaaaag cagattaaat agatatacat tcgattaacg ataagtaaaa 6240  
 tgtaaaatca caggatttcc gtgtgtggtc ttctacacag acaagatgaa acaattcggc 6300  
 attaatacct gagagcagga agagcaagat aaaaggtagt atttgttggc gatcccccta 6360  
 gagtctttta catcttcgga aaacaaaaac tttttttct ttaatttctt tttttacttt 6420  
 ctatttttaa tttatatatt tatattaaaa aatttaaatt ataattattt ttatagcacg 6480  
 tgatgaaaag gaccagggtg gcacttttcg gggaaatgtg cgcggaacc cttttgttt 6540  
 atttttctaa atacattcaa atatgtatec gctcatgaga caataaccct gataaatgct 6600  
 tcaataatat tgaaaaagga agagtatgag tattcaacat ttccgtgtcg cccttattcc 6660  
 cttttttgcg gcattttgcc ttctgtttt tgetcaccia gaaacgctgg tgaagtaaa 6720  
 agatgctgaa gatcagttgg gtgcacgagt gggttacatc gaactggatc tcaacagcgg 6780  
 taagatcctt gagagtttcc gccccgaaga acgttttcca atgatgagca cttttaaagt 6840  
 tctgctatgt ggcgcggtat tatcccgtat tgacgccggg caagagcaac tcggtcgccg 6900  
 catacactat tctcagaatg acttggttga gtactacca gtcacagaaa agcatcttac 6960  
 ggatggcatg acagtaagag aattatgcag tgctgccata accatgagtg ataactgc 7020  
 ggccaactta ctctgacaa cgatcggagg accgaaggag ctaaccgett ttttccaaa 7080  
 catgggggat catgtaactc gccttgatcg ttgggaaccg gagctgaatg aagccatacc 7140  
 aaacgacgag cgtgacacca cgatgcctgt agcaatggca acaacgttgc gaaactatt 7200  
 aactggcgaa ctacttactc tagcttcccg gcaacaatta atagactgga tggaggcgga 7260  
 taaagttgca ggaccacttc tgcgctcggc cttccggct ggctggttta ttgctgataa 7320  
 atctggagcc ggtgagcgtg ggtctcgcgg tatcattgca gcactggggc cagatggtaa 7380  
 gccctcccgt atcgtagtta tctacacgac gggcagtcag gcaactatgg atgaacgaaa 7440  
 tagacagatc gctgagatag gtgcctcact gattaagcat tggttaactgt cagaccaagt 7500  
 ttactcatat aacttttaga ttgatttaa acttcatttt taatttaaaa ggatctaggt 7560  
 gaagatcctt tttgataate tcatgaccaa aatccctaa cgtgagtttt cgttccactg 7620  
 agcgtcagac cccgtagaaa agatcaaagg atcttcttga gatccttttt ttctgcgcgt 7680  
 aatctgctgc ttgcaaaaaa aaaaaccacc gctaccagcg gtggtttgtt tgccggatca 7740  
 agagctacca actcttttcc cgaaggtaac tggettccagc agagcgcaga taccaatac 7800  
 tgtccttcta gtgtagccgt agttaggcca ccaettcaag aactctgtag caccgcctac 7860  
 atacctcgtc ctgctaatcc tgttaccagt ggetgctgcc agtggcgata agtcgtgtct 7920  
 taccgggttg gactcaagac gatagttacc ggataaggcg cagcggtcgg gctgaacggg 7980  
 gggttcgtgc acacagccca gcttggagcg aacgacctac accgaactga gatactaca 8040  
 gcgtgagcat tgagaaaagc ccacgettcc cgaagggaga aaggcggaca ggtatccggt 8100  
 aagcggcagg gtcggaacag gagagcgcac gagggagctt ccagggggga acgcctggtta 8160



tctttatagt cctgtcgggt ttcgccacct ctgacttgag cgtegat ttt tgtgatgctc 8220  
 gtcagggggg ccgagcctat ggaaaaacgc cagcaacgcg gcctttttac ggttcctggc 8280  
 cttttgctgg ccttttgctc acatgttctt tcttgcgta tcccctgatt ctgtggataa 8340  
 ccgtattacc gcctttgagt gagctgatac cgctcgccgc agccgaacga ccgagcgag 8400  
 cgagtcagtg agcgaggaag cggaagagcg cccaatacgc aaaccgcctc tccccgcgcg 8460  
 ttggccgatt cattaatgca gctggcacga caggtttccc gactggaaag cgggcagtga 8520  
 gcgcaacgca attaatgtga gttacctcac tcattaggca ccccaggctt tacactttat 8580  
 gcttccggct cctatgttgt gtggaattgt gageggataa caatttcaca caggaaacag 8640  
 ctatgaccat gattacgcca agctcggaat taacctcac taaagggaac aaaagctggg 8700  
 taccgggccc cccctcgag 8719

<210> 10

<211> 1632

<212> DNA

<213> 人工序列

<220>

<223> pTAN08

<400> 10

ggcaacggtt catcatctca tggatctgca catgaacaaa caccagagtc aaacgacgtt 60  
 gaaattgagg ctactgcgcc aattgatgac aatacagacg atgataacaa accgaagtta 120  
 tctgatgtag aaaaggatta gagatgctaa gagatagtga tgatatttca taaataatgt 180  
 aattctatat atgttaatta ccttttttgc gaggcatatt tatggtgaag gataagtttt 240  
 gaccatcaaa gaaggtaat gtggctgtgg tttcagggtc cataaagctt ttcaattcat 300  
 cttttttttt tttgttcttt tttttgattc cggtttcttt gaaatttttt tgattcggta 360  
 atctccgagc agaaggaaga acgaaggaag gagcacagac ttagattggt atatatacgc 420  
 atatgtggtg ttgaagaaac atgaaattgc ccagtattct taaccaact gcacagaaca 480  
 aaaacctgca ggaaacgaag ataaatcatg tcgaaagcta catataagga acgtgctgct 540  
 actcatccta gtccctgttc tgccaagcta ttaatatca tgcacgaaaa gcaaacaaac 600  
 ttgtgtgctt cattggatgt tcgtaccacc aaggaattac tggagttagt tgaagcatta 660  
 ggtcccaaaa tttgtttact aaaaacacat gtggatatct tgactgattt ttccatggag 720  
 ggcacagtta agccgctaaa ggcattatec gccaaagta attttttact cttcgaagac 780  
 agaaaatttg ctgacattgg taatacagtc aaattgcagt actctgcggg tgtatacaga 840  
 atagcagaat gggcagacat tacgaatgca cacggtgtgg tgggcccagg tattgttagc 900  
 ggtttgaagc aggcggcgga agaagtaaca aaggaaccta gaggcctttt gatgttagca 960  
 gaattgtcat gcaagggctc cctagetact ggagaatata ctaagggtac tgttgacatt 1020  
 gcgaagagcg acaaagattt tgttatcggc tttattgctc aaagagacat ggggtggaaga 1080  
 gatgaagggt acgattgggt gattatgaca cccggtgtgg gtttagatga caagggagac 1140  
 gcattgggtc aacagtatag aaccgtggat gatgtggtct ctacaggatc tgacattatt 1200  
 attgttggaagaggactatt tgcaaaggga agggatgcta aggtagaggg tgaacgttac 1260  
 agaaaagcag gctgggaagc atatttgaga agatgcggcc agcaaaacta aaaactgta 1320

ttataagtaa atgcatgtat actaaactca caaattagag cttcaattta attatatcag 1380  
 ttattaccgc ggaatctcgg tcgtaatgat ttttataatg acgaaaaaaaa aaaattgga 1440  
 aagaaaaagc ttcattggcct ttataaaaag gaaccatcca atacctcgcc agaaccaagt 1500  
 aacagtattt tacggggcac aaatcaagaa caataagaca ggactgtaaa gatggacgca 1560  
 ttgaactcca aagaacaaca agagttccaa aaagtagtgg aacaaaagca aatgaaggat 1620  
 ttcatgcggt tg 1632

<210> 11

<211> 515

<212> PRT

<213> 扣囊复膜酵母(Saccharomycopsis fibuligera)

<400> 11

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

Ser Thr His Trp Ser Thr Ser Gly Phe Asp Leu Trp Glu Glu Asn Gln  
 225 230 235 240  
 Gly Thr His Phe Phe Thr Ala Leu Val Gln Leu Lys Ala Leu Ser Tyr  
 245 250 255  
 Gly Ile Pro Leu Ser Lys Thr Tyr Asn Asp Pro Gly Phe Thr Ser Trp  
 260 265 270  
 Leu Glu Lys Gln Lys Asp Ala Leu Asn Ser Tyr Ile Asn Ser Ser Gly  
 275 280 285  
 Phe Val Asn Ser Gly Lys Lys His Ile Val Glu Ser Pro Gln Leu Ser  
 290 295 300  
 Ser Arg Gly Gly Leu Asp Ser Ala Thr Tyr Ile Ala Ala Leu Ile Thr  
 305 310 315 320  
 His Asp Ile Gly Asp Asp Asp Thr Tyr Thr Pro Phe Asn Val Asp Asn  
 325 330 335  
 Ser Tyr Val Leu Asn Ser Leu Tyr Tyr Leu Leu Val Asp Asn Lys Asn  
 340 345 350  
 Arg Tyr Lys Ile Asn Gly Asn Tyr Lys Ala Gly Ala Ala Val Gly Arg  
 355 360 365  
 Tyr Pro Glu Asp Val Tyr Asn Gly Val Gly Thr Ser Glu Gly Asn Pro  
 370 375 380  
 Trp Gln Leu Ala Thr Ala Tyr Ala Gly Gln Thr Phe Tyr Thr Leu Ala  
 385 390 395 400  
 Tyr Asn Ser Leu Lys Asn Lys Lys Asn Leu Val Ile Glu Lys Leu Asn  
 405 410 415  
 Tyr Asp Leu Tyr Asn Ser Phe Ile Ala Asp Leu Ser Lys Ile Asp Ser  
 420 425 430  
 Ser Tyr Ala Ser Lys Asp Ser Leu Thr Leu Thr Tyr Gly Ser Asp Asn  
 435 440 445  
 Tyr Lys Asn Val Ile Lys Ser Leu Leu Gln Phe Gly Asp Ser Phe Leu  
 450 455 460  
 Lys Val Leu Leu Asp His Ile Asp Asp Asn Gly Gln Leu Thr Glu Glu  
 465 470 475 480  
 Ile Asn Arg Tyr Thr Gly Phe Gln Ala Gly Ala Val Ser Leu Thr Trp  
 485 490 495  
 Ser Ser Gly Ser Leu Leu Ser Ala Asn Arg Ala Arg Asn Lys Leu Ile  
 500 505 510  
 Glu Leu Leu  
 515

<210> 12

<211> 19

<212> PRT

<213> 人工序列

<220>

<223> MFa2信号序列

<400> 12

Met Lys Phe Ile Ser Thr Phe Leu Thr Phe Ile Leu Ala Ala Val Ser  
1                   5                   10                   15

Val Thr Ala

<210> 13

<211> 17

<212> PRT

<213> 人工序列

<220>

<223> PH05信号序列

<400> 13

Met Phe Lys Ser Val Val Tyr Ser Ile Leu Ala Ala Ser Leu Ala Asn  
1                   5                   10                   15

Ala

<210> 14

<211> 516

<212> PRT

<213> 人工序列

<220>

<223> MFa2- Sf GA蛋白

<400> 14

Met Lys Phe Ile Ser Thr Phe Leu Thr Phe Ile Leu Ala Ala Val Ser  
1                   5                   10                   15

Val Thr Ala Val Pro Val Glu Leu Asp Lys Arg Asn Thr Gly His Phe  
                  20                   25                   30

Gln Ala Tyr Ser Gly Tyr Thr Val Ala Arg Ser Asn Phe Thr Gln Trp  
                  35                   40                   45

Ile His Glu Gln Pro Ala Val Ser Trp Tyr Tyr Leu Leu Gln Asn Ile  
                  50                   55                   60

Asp Tyr Pro Glu Gly Gln Phe Lys Ser Ala Lys Pro Gly Val Val Val  
65                   70                   75                   80

Ala Ser Pro Ser Thr Ser Glu Pro Asp Tyr Phe Tyr Gln Trp Thr Arg  
                  85                   90                   95

Asp Thr Ala Ile Thr Phe Leu Ser Leu Ile Ala Glu Val Glu Asp His

	100		105		110
Ser Phe Ser Asn Thr Thr Leu Ala Lys Val Val Glu Tyr Tyr Ile Ser					
	115		120		125
Asn Thr Tyr Thr Leu Gln Arg Val Ser Asn Pro Ser Gly Asn Phe Asp					
	130		135		140
Ser Pro Asn His Asp Gly Leu Gly Glu Pro Lys Phe Asn Val Asp Asp					
145		150		155	160
Thr Ala Tyr Thr Ala Ser Trp Gly Arg Pro Gln Asn Asp Gly Pro Ala					
	165		170		175
Leu Arg Ala Tyr Ala Ile Ser Arg Tyr Leu Asn Ala Val Ala Lys His					
	180		185		190
Asn Asn Gly Lys Leu Leu Leu Ala Gly Gln Asn Gly Ile Pro Tyr Ser					
	195		200		205
Ser Ala Ser Asp Ile Tyr Trp Lys Ile Ile Lys Pro Asp Leu Gln His					
210		215		220	
Val Ser Thr His Trp Ser Thr Ser Gly Phe Asp Leu Trp Glu Glu Asn					
225		230		235	240
Gln Gly Thr His Phe Phe Thr Ala Leu Val Gln Leu Lys Ala Leu Ser					
	245		250		255
Tyr Gly Ile Pro Leu Ser Lys Thr Tyr Asn Asp Pro Gly Phe Thr Ser					
	260		265		270
Trp Leu Glu Lys Gln Lys Asp Ala Leu Asn Ser Tyr Ile Asn Ser Ser					
275		280		285	
Gly Phe Val Asn Ser Gly Lys Lys His Ile Val Glu Ser Pro Gln Leu					
290		295		300	
Ser Ser Arg Gly Gly Leu Asp Ser Ala Thr Tyr Ile Ala Ala Leu Ile					
305		310		315	320
Thr His Asp Ile Gly Asp Asp Asp Thr Tyr Thr Pro Phe Asn Val Asp					
	325		330		335
Asn Ser Tyr Val Leu Asn Ser Leu Tyr Tyr Leu Leu Val Asp Asn Lys					
	340		345		350
Asn Arg Tyr Lys Ile Asn Gly Asn Tyr Lys Ala Gly Ala Ala Val Gly					
	355		360		365
Arg Tyr Pro Glu Asp Val Tyr Asn Gly Val Gly Thr Ser Glu Gly Asn					
370		375		380	
Pro Trp Gln Leu Ala Thr Ala Tyr Ala Gly Gln Thr Phe Tyr Thr Leu					
385		390		395	400
Ala Tyr Asn Ser Leu Lys Asn Lys Lys Asn Leu Val Ile Glu Lys Leu					
	405		410		415

Asn Tyr Asp Leu Tyr Asn Ser Phe Ile Ala Asp Leu Ser Lys Ile Asp  
 420 425 430  
 Ser Ser Tyr Ala Ser Lys Asp Ser Leu Thr Leu Thr Tyr Gly Ser Asp  
 435 440 445  
 Asn Tyr Lys Asn Val Ile Lys Ser Leu Leu Gln Phe Gly Asp Ser Phe  
 450 455 460  
 Leu Lys Val Leu Leu Asp His Ile Asp Asp Asn Gly Gln Leu Thr Glu  
 465 470 475 480  
 Glu Ile Asn Arg Tyr Thr Gly Phe Gln Ala Gly Ala Val Ser Leu Thr  
 485 490 495  
 Trp Ser Ser Gly Ser Leu Leu Ser Ala Asn Arg Ala Arg Asn Lys Leu  
 500 505 510  
 Ile Glu Leu Leu  
 515  
 <210> 15  
 <211> 514  
 <212> PRT  
 <213> 人工序列  
 <220>  
 <223> PH05 - Sf GA蛋白  
 <400> 15  
 Met Phe Lys Ser Val Val Tyr Ser Ile Leu Ala Ala Ser Leu Ala Asn  
 1 5 10 15  
 Ala Val Pro Val Glu Leu Asp Lys Arg Asn Thr Gly His Phe Gln Ala  
 20 25 30  
 Tyr Ser Gly Tyr Thr Val Ala Arg Ser Asn Phe Thr Gln Trp Ile His  
 35 40 45  
 Glu Gln Pro Ala Val Ser Trp Tyr Tyr Leu Leu Gln Asn Ile Asp Tyr  
 50 55 60  
 Pro Glu Gly Gln Phe Lys Ser Ala Lys Pro Gly Val Val Val Ala Ser  
 65 70 75 80  
 Pro Ser Thr Ser Glu Pro Asp Tyr Phe Tyr Gln Trp Thr Arg Asp Thr  
 85 90 95  
 Ala Ile Thr Phe Leu Ser Leu Ile Ala Glu Val Glu Asp His Ser Phe  
 100 105 110  
 Ser Asn Thr Thr Leu Ala Lys Val Val Glu Tyr Tyr Ile Ser Asn Thr  
 115 120 125  
 Tyr Thr Leu Gln Arg Val Ser Asn Pro Ser Gly Asn Phe Asp Ser Pro  
 130 135 140

Asn His Asp Gly Leu Gly Glu Pro Lys Phe Asn Val Asp Asp Thr Ala  
 145 150 155 160  
 Tyr Thr Ala Ser Trp Gly Arg Pro Gln Asn Asp Gly Pro Ala Leu Arg  
 165 170 175  
 Ala Tyr Ala Ile Ser Arg Tyr Leu Asn Ala Val Ala Lys His Asn Asn  
 180 185 190  
 Gly Lys Leu Leu Leu Ala Gly Gln Asn Gly Ile Pro Tyr Ser Ser Ala  
 195 200 205  
 Ser Asp Ile Tyr Trp Lys Ile Ile Lys Pro Asp Leu Gln His Val Ser  
 210 215 220  
 Thr His Trp Ser Thr Ser Gly Phe Asp Leu Trp Glu Glu Asn Gln Gly  
 225 230 235 240  
 Thr His Phe Phe Thr Ala Leu Val Gln Leu Lys Ala Leu Ser Tyr Gly  
 245 250 255  
 Ile Pro Leu Ser Lys Thr Tyr Asn Asp Pro Gly Phe Thr Ser Trp Leu  
 260 265 270  
 Glu Lys Gln Lys Asp Ala Leu Asn Ser Tyr Ile Asn Ser Ser Gly Phe  
 275 280 285  
 Val Asn Ser Gly Lys Lys His Ile Val Glu Ser Pro Gln Leu Ser Ser  
 290 295 300  
 Arg Gly Gly Leu Asp Ser Ala Thr Tyr Ile Ala Ala Leu Ile Thr His  
 305 310 315 320  
 Asp Ile Gly Asp Asp Asp Thr Tyr Thr Pro Phe Asn Val Asp Asn Ser  
 325 330 335  
 Tyr Val Leu Asn Ser Leu Tyr Tyr Leu Leu Val Asp Asn Lys Asn Arg  
 340 345 350  
 Tyr Lys Ile Asn Gly Asn Tyr Lys Ala Gly Ala Ala Val Gly Arg Tyr  
 355 360 365  
 Pro Glu Asp Val Tyr Asn Gly Val Gly Thr Ser Glu Gly Asn Pro Trp  
 370 375 380  
 Gln Leu Ala Thr Ala Tyr Ala Gly Gln Thr Phe Tyr Thr Leu Ala Tyr  
 385 390 395 400  
 Asn Ser Leu Lys Asn Lys Lys Asn Leu Val Ile Glu Lys Leu Asn Tyr  
 405 410 415  
 Asp Leu Tyr Asn Ser Phe Ile Ala Asp Leu Ser Lys Ile Asp Ser Ser  
 420 425 430  
 Tyr Ala Ser Lys Asp Ser Leu Thr Leu Thr Tyr Gly Ser Asp Asn Tyr  
 435 440 445  
 Lys Asn Val Ile Lys Ser Leu Leu Gln Phe Gly Asp Ser Phe Leu Lys

450	455	460			
Val Leu Leu Asp His Ile Asp Asp Asn Gly Gln Leu Thr Glu Glu Ile					
465	470	475	480		
Asn Arg Tyr Thr Gly Phe Gln Ala Gly Ala Val Ser Leu Thr Trp Ser					
	485	490	495		
Ser Gly Ser Leu Leu Ser Ala Asn Arg Ala Arg Asn Lys Leu Ile Glu					
	500	505	510		
Leu Leu					
<210> 16					
<211> 4632					
<212> DNA					
<213> 人工序列					
<220>					
<223> pTAN29					
<400> 16					
cagagcctct tatattcaact ctgttcctcc atcgectatt gagaaacgtt ggaataaaac 60					
tctaaaaata tcatctagtt ggtagttttt tattttacca gtacattgtc acttgcggag 120					
ggaggatgac ataaagattg agacgcagtc atttaatgaa gtttaaacgc aggtatttga 180					
taaagtaata cgatattgaa tcatgacgta taaagtgaaa tgaacaaatg attacgtaa 240					
aaatgtcgat tttctcttga gagactccca tagcctctaa gaggccttct actacgttcc 300					
atatatctaa gaatggggcc atatccagtg gaatcccagc aattatttaa ggatcaccta 360					
tttctcagcc gatattttag caaaatcaact accaatatca gggggcaata gttgatcgcc 420					
tactttaaca aaaaatgttg ctcacgtatt aacacaggca acaaaaagga tattacgcaa 480					
gaacgtagta tccacatgcc atcctccttg ttgcatettt ttttttccga aatgattccc 540					
tttctgcac aacacgagat ctttcacgca tacatcgtaa ggatcacccc ccaactcaagt 600					
cgttgcattg ctaacatgtg gcattctgcc catttttttc acgaaaattc tctctctata 660					
atgaagaccc ttgtgccctg gactctgtaa tacttgaaac tacttctca ataatcgctt 720					
ggagacctac cccacgctt ttcaaacaag gcgctagcaa aaagcctgcc gatatctcct 780					
tgccccctcc ttctgttcga gagaactacg acccgaccaa taataatgtc atacaagaac 840					
cgccaagaac caactgctga accttagatc tccaatactt cagttggagt atgtgaatat 900					
ataagtacct ggtcgactaa tcttcttgea tcttttctga ttcttacatc ctatgtcgct 960					
aatacagttc ccgcatagag aagaaagcaa acaaaagtag tcaactcgaga tctcccagat 1020					
ttatcattat caatactgcc atttcaaaga atacgtaaat aattaatagt agtgattttc 1080					
ctaactttat ttagtcaaaa aattggcett ttaattctgc tgtaaccctg acatgcccac 1140					
aatagggggc gggttacaca gaatatataa catcataggt gtctgggtga acagtttatt 1200					
cctggcatcc actaaatata atggagcccc ctttttttaa gctggcatcc agaaaaaaaa 1260					
agaatcccag caccaaaata ttgttttett caccaacctt cagttcatag gtccattctc 1320					
ttagcgaac tacacagaac aggggcacaa acaggcaaaa aacgggcaca acctcaatgg 1380					
agtgatgcaa cctgcttggg gtaaatgatg acacaaggca attgacctac gcatgtatct 1440					



atctcatttt cttacacctt ctattacctt ctgctctctc tgatttgtaa aaagctgaaa 1500  
aaaaagggtg aaaccagttc cctgaaatta ttccctatt tgactaataa gtatataaag 1560  
acggtaggta ttgattgtaa ttctgtaa atctatttetta aacttcttaa attctacttt 1620  
tatagttagt ctttttttta gtttaaaaca ccaagaactt agtttcgaat aaacacacat 1680  
aaacaaacaa atctagaatg attagattaa ccgtattcct cactgcagtt tttgcagcag 1740  
tcgcttcctg tgttccagtt gaattggata agagaaatac aggccatttc caagcatatt 1800  
ctggttacac cgtagctaga tcaaacttta ctcaatggat tcacgagcaa ccagccgtat 1860  
catggtacta tttgcttcag aatataact atccagaagg acaattcaag tctgccaagc 1920  
caggggtcgt tgtggcttcc cttctacat ccgaacctga ttactttctac caatggacta 1980  
gagatactgc tatcaccttc ttgtcactta tcgctggaagt tgaggatcat tctttttcaa 2040  
atactacact agccaagggt gttgaatact acatctctaa tacttacaca ttacaaagag 2100  
tttccaacc atctggtaac ttcgacagtc caaatcacga cggtttggga gaaccaaagt 2160  
ttaatgttga tgatacagct tatactgeat cttgggtag accacaaaat gatggcccag 2220  
cgttgagagc atacgcaatt tcaagatacc ttaacgcagt agcaaacac aacaacggta 2280  
agttactgct cgctggacaa aacggtatte cttactette agcttctgat atctactgga 2340  
agattatcaa gccagatctt caacatgtgt caaccattg gtctacatct ggttttgatt 2400  
tgtgggaaga gaatcagga acacatttct ttactgcgtt ggtccagcta aaagcactta 2460  
gttacggcat tcctttaagt aagacctaca acgatcctgg tttcactagt tggctagaaa 2520  
agcaaaagga tgctttaaac tcttataca acagctctgg tttcgtaaac tctggcaaaa 2580  
agcatatagt ggagagccct caactatctt caagaggagg gttggatagc gccacataca 2640  
ttgcagcctt aatcacacat gatattggcg acgacgacac ttacacacct ttcaacgttg 2700  
acaactccta tgtcttgaac tcactgtatt accttctagt cgataacaaa aaccgttaca 2760  
aaatcaatgg taactacaag gccggtgctg ctgttgtag ataccagag gatgtttaca 2820  
acgggtgttg gacatcagaa ggcaatccat ggcaattagc tacagcctac gccggccaaa 2880  
cattttacac actggcttac aactcattga aaaacaaaa aacttagtg attgaaaagt 2940  
tgaactacga cctctacaat tctttcatag cagatttacc caagatcgat agttcttacg 3000  
catcaaaaga ctcttgact ttgacctagc gttctgacaa ctacaaaaac gtcataaagt 3060  
cactattaca gtttgagat tcattctga aggtcttgcg cgatcacatt gatgataatg 3120  
gacaattaac agaagagatc aatagataca cagggttcca ggctggtgct gttagtttga 3180  
catggtcctc tggttcatta cttcagcaa accgtgcgag aaataagttg attgaactat 3240  
tgtagttaat taaacaggec cttttcctt tgctgatata atgtaattag ttatgtcacg 3300  
cttacattca cgccctctc ccacatccgc ttaaccgaa aaggaaggag ttagacaacc 3360  
tgaagtctag gtccctattt atttttttat agttatgtta gtattaagaa cgttattttat 3420  
atttcaaatt tttctttttt ttctgtacaa acgcgtgtac gcatgtaacg ggcagacggc 3480  
cggccataac ttcgtataat gtatgtata cgaagttagt gcaacggttc atcatctcat 3540  
ggatctgcac atgaacaaac accagagtea aacgacgttg aaattgagge tactgcgcca 3600  
attgatgaca atacagacga tgataacaaa ccgaagttag ctgatgtaga aaaggattag 3660  
agatgctaag agatagtgat gatatttcat aaataatgta attctatata tgttaattac 3720  
cttttttgcg aggcatattt atgggtgaagg ataagttttg accatcaaag aaggttaatg 3780

tggctgtggt ttcagggtcc ataaagcttt tcaattcacc tttttttttt ttgttctttt 3840  
 ttttgattcc ggtttctttg aaattttttt gattcggtaa tctccgagca gaaggaagaa 3900  
 cgaaggaagg agcacagact tagattggta tatatacgca tatgtggtgt tgaagaaaca 3960  
 tgaattgcc cagtattctt aaccaactg cacagaacaa aaacctgcag gaaacgaaga 4020  
 taaatcatgt cgaaagctac atataaggaa cgtgctgcta ctcatcctag tcctgttgct 4080  
 gccaaactat ttaatatcat gcacgaaaag caaacaaact tgtgtgcttc attggatggt 4140  
 cgtaccacca aggaattact ggagttagtt gaagcattag gtcccaaat ttgtttacta 4200  
 aaaacacatg tggatatctt gactgatttt tccatggagg gcacagttaa gccgctaaag 4260  
 gcattatccg ccaagtacaa ttttttactc ttcgaagaca gaaaatttgc tgacattggt 4320  
 aatacagtca aattgcagta ctctgcgggt gtatacagaa tagcagaatg ggcagacatt 4380  
 acgaatgcac acggtgtggt gggcccaggt attgttagcg gtttgaagca ggcggcgaa 4440  
 gaagtaacaa aggaacctag aggctttttg atgttagcag aattgtcatg caagggtcc 4500  
 ctactactg gagaatatac taagggtact gttgacattg cgaagagcga caaagatttt 4560  
 gttatcggtt ttattgctca aagagacatg ggtggaagag atgaaggta cgattggttg 4620  
 attatgacac gc 4632

<210> 17

<211> 4363

<212> DNA

<213> 人工序列

<220>

<223> pTAN30

<400> 17

ggccgctcca tggagggcac agttaagccg ctaaaggcat tatccgcca gtacaatttt 60  
 ttactcttcg aagacagaaa atttgctgac attgtaata cagtcaaatt gcagtactct 120  
 gcgggtgtat acagaatagc agaatgggca gacattacga atgcacacgg tgtggtgggc 180  
 ccaggatttg ttagcggttt gaagcaggcg gcggaagaag taacaaagga acctagaggc 240  
 cttttgatgt tagcagaatt gtcattgcaag ggctccctag ctactggaga atatactaag 300  
 ggtactgttg acattgcgaa gagcgacaaa gatthttgta tcggctttat tgctcaaaga 360  
 gacatgggtg gaagagatga aggttacgat tggttgatta tgacaccggg tgtgggttta 420  
 gatgacaagg gagacgcatt gggtaacag tatagaaccg tggatgatgt ggtctctaca 480  
 ggatctgaca ttattattgt tggaagagga ctatttgcaa agggaaggga tgctaaggta 540  
 gaggtgaac gttacagaaa agcaggctgg gaagcatatt tgagaagatg cggccagcaa 600  
 aactaaaaaa ctgtattata agtaaatgca tgtatactaa actcacaat tagagcttca 660  
 atttaattat atcagttatt acccgggaat ctcggtcgta atgattttta taatgacgaa 720  
 aaaaaaaaaa ttgaaagaa aaagcttcat ggcctttata aaaaggaacc atccaatacc 780  
 tcgccagaac caagtaacag tattttacgg ggcacaaatc aagaacaata agacaggact 840  
 gtaaagatgg acgcattgaa ctccaaagaa caacaagagt tccaaaagt agtgaacaa 900  
 aagcaaatga aggatttcat gcgtttgata acttcgtata atgtatgcta tacgaagtta 960  
 tctcgagggc cagaaaaagg aagtgtttcc ctcttcttg aattgatggt acctcataa 1020

agcacgtggc ctcttatcga gaaagaaatt accgctcctc gtgatttggt tgcaaaaaga 1080  
 acaaaactga aaaaaccag acacgctcga ctctctgtct tctgttgat tgcagcttcc 1140  
 aatttcgtca cacaacaagg tcctagcgac ggctcacagg ttttgtaaca agcaatcgaa 1200  
 ggttctggaa tggcgggaaa gggtttagta ccacatgcta tgatgccac tgtgatctcc 1260  
 agagcaaagt tcgttcgac gtactgttac tctctctctt tcaaacagaa ttgtccgaat 1320  
 cgtgtgacaa caacagcctg ttctcacaca ctcttttctt ctaaccaagg ggggtggtta 1380  
 gtttagtaga acctcgtgaa acttacattt acatataat aaacttgcatt aaattggtca 1440  
 atgcaagaaa tacatatttg gtcttttcta attcgtagtt tttcaagttc ttagatgctt 1500  
 tctttttctc ttttttacag atcatcaagg aagtaattat ctacttttta caagtctaga 1560  
 atgatcagac ttacagtttt cctaacagcc gtttctgccc ccgttgcatc atgtgtccca 1620  
 gtagaattgg ataagagaaa caccggccat ttccaagcat attcaggata caccgttgca 1680  
 cgttctaatt tcacacaatg gattcatgag cagcctgctg tgtctggta ctacttatta 1740  
 caaaacattg attatcctga gggacaatte aagtcagcga aaccaggcgt tgtggttgct 1800  
 tctccatcca ctccagaacc agactaette taccagtgga cccgtgacac agcaataact 1860  
 ttcttatctt tgatagcaga agtagaagat cactcatttt caaatacaac tctagctaag 1920  
 gttgtcgaat actacatctc taacacatac acctacaaa gagtttctaa cccatctggt 1980  
 aatttcgata gcccaaatca cgatggctctg ggtgaaccaa agttcaacgt tgacgacact 2040  
 gcttacactg catcatgggg cagacctcaa aacgacggtc cagccttaag agcttacgcg 2100  
 atctcaagat atttgaacgc agttgccaag cataacaac gtaagctatt gctcgcgggt 2160  
 caaaatggta ttcttactc atctgcatca gatcttact ggaagattat caagccagat 2220  
 ttacaacatg taagtactca ctggagtaca tctggttttg acttatggga agagaatcaa 2280  
 ggtacacatt tctttactgc acttgtccag ttaaaagctc tttcatacgg tatacctttg 2340  
 tctaagacat ataacgatcc aggatttact tcttggttgg aaaagcagaa ggatgccttg 2400  
 aactcttaca tcaattccag cggtctcgtc aactccggga aaaagcagat tgtcgaatct 2460  
 cctcaattat ctagtagagg gggctttgat agcgtctact acatcgctgc tctaattaca 2520  
 catgatattg gtgatgatga tacatacact ctttttaacg tagataattc ttatgtgctg 2580  
 aactctttat actatctgct ttagacaac aaaaacagat acaagatcaa cgggaactac 2640  
 aaagcaggag ctgcagttgg tagatacca gaagatgtgt acaatggagt gggaacctca 2700  
 gagggaaacc catggcaatt ggcgacagca tacgccggcc aaacctttta cacactggct 2760  
 tacaattctc tcaaaaacaa aaaaaatttg gttattgaga agttgaatta cgatctatac 2820  
 aactccttta tagctgactt aagtaagatt gactctctct acgcttctaa ggattcattg 2880  
 acattgacct acggctcaga taactacaaa aatgtcatta agtcaactttt acaattcggg 2940  
 gattctttct tgaagttctt gttggacct attgatgata atggctcagct aacagaggaa 3000  
 atcaacagat atacaggttt tcaagctggc gcagtttccc tcaactggag tagtggttca 3060  
 ctcttatctg caaacagagc cagaaacaag ttgatcgaat tgcttttagtt aattaagaag 3120  
 ttttgtaga aaataaatca ttttttaatt gageattctt attcctattt tatttaata 3180  
 gttttatgta ttgttagcta catacaacag tttaaatcaa attttctttt tcccaagtcc 3240  
 aaaatggagg tttattttga tgaccgcat gcgattatgt tttgaaagta taagactaca 3300  
 tacatgtaca tatatttaaa catgtaaacc cgtccattat attgccgggc agacggccgg 3360

ccttatagcc tagctttaag gctactttaa aaacttttta tttattcata cacatatatt 3420  
 atcgaacatt cgtataactt aatatcattc aaaaaaaaaa aaaaaaaaaa aagaaaacat 3480  
 atacacatat atatttatgt ttatagagag agagagagaa aatttgaatt tttgaatcat 3540  
 ttgcaaagtt atatgtttta tacattatit attcattttt tttgggtgctg aggacattgt 3600  
 gctgttcaga gaaccactta aaatacgcac cgttctgtaa atatccactt tcattaataaa 3660  
 ccttattcac ttctaacttt gccttcaact ccttcttggg gttttctccc ttttttttct 3720  
 gaacaagctc aaccagatat aatggttcgt tcttttcgaa ctttgccttt acatatattt 3780  
 cctcctttgt acctcttctc tttcccacat aaacagctcc cttttcaata aaacgagaga 3840  
 aataccagaa aagtagcgag agaacaaaat atgcgcctac caaaagcttt tgatacgtaa 3900  
 caatctgac tctctcaaat tttttateca agaagaaact caaacagct acaacagcta 3960  
 tggaataacc tatgtacaat ttagcatcga gtaaagcgta tgatctctcg taatttaac 4020  
 tcgcgaaaac agaaggtagg gcttcateta aagcttgggt caactccggg attgaatata 4080  
 cattaatagg tttagcagaa ctcatcttga acaggegtct cttttcttta caataacttg 4140  
 tgcttttct tctataatc cgtttcaacg tgtacaattg tcattttttg tctggatga 4200  
 ttttgagaa ctgaaaaaat ctcttaaatg ttccgctca tcaagaagc atattccttt 4260  
 acaaaagtac attgatctta caagaagcta gctaattgta ctatttaaaa aacaactaca 4320  
 ctccatcaat acataaaaatt gttatgatag acttgaggga cgg 4363

<210> 18

<211> 5015

<212> DNA

<213> 人工序列

<220>

<223> pTAN37

<400> 18

cagagcctct tatattcact ctgttctcc atgcctatt gagaaacgtt ggaataaaac 60  
 tctaaaaata tcatctagtt ggtagtttt tattttacca gtacattgtc acttgcggag 120  
 ggaggatgac ataaagattg agacgcagtc atttaatgaa gtttaaaccg aggtatttga 180  
 taaagtaata cgatattgaa tcatgacgta taaagtgaat tgaacaaatg attacgtaaa 240  
 aaatgtcgat tttctcttga gagactccca tagcctetaa gaggccttct actacgttcc 300  
 atatatctaa gaatggggcc atatccagtg gaatcccagc aattatttaa ggatcaccta 360  
 tttctcagcc gatattttag caaaatcact accaatatac gggggcaata gttgatcgcc 420  
 tactttaaca aaaaatgttg ctcacgtatt aacacaggca acaaaaagga tattacgcaa 480  
 gaacgtagta tccacatgcc atctctcttg ttgcatcttt ttttttccga aatgatcccc 540  
 tttctgcac aacacgagat ctttcacgca tacatcgtaa ggatcacccc cactcaagt 600  
 cgttgcatg ctaacatgtg gcattctgac ctttttttc acgaaaatc tctctctata 660  
 atgaagacc ttgtgccctg gactctgtaa tacttgaaac tacttctca ataategett 720  
 ggagacctac ccccacgett ttcaaacaag gcgctagcaa aaagcctgcc gatctctct 780  
 tgccccctc ttctgttcca gagaactacg acccgaccaa taataatgtc atacaagaac 840  
 cgccaagaac caactgctga accttagatc tccaatactt cagttggagt atgtgaatat 900

ataagtacct ggtcgactaa tcttcttgca tcttttcgta ttettacate ctatgtcget 960  
 aatacagttc ccgcatagag aagaaagcaa acaaaagtag tcaactcgaga tctcccagagt 1020  
 ttatcattat caatactgcc atttcaaaga atacgtaaat aattaatagt agtgattttc 1080  
 ctaactttat ttagtcaaaa aattggcctt ttaattctgc tgtaaccctg acatgcccac 1140  
 aatagggggc gggttacaca gaatatataa catcataggt gtctgggtga acagtttatt 1200  
 cctggcatcc actaaatata atggagcccc ctttttttaa gctggcatcc agaaaaaaaa 1260  
 agaatcccag caccaaaaata ttgttttctt caccaacctt cagttcatag gtccattctc 1320  
 ttagcgcaac tacacagaac aggggcacaa acaggcaaaa aacgggcaca acctcaatgg 1380  
 agtgatgcaa cctgcttggg gtaaatgatg acacaaggca attgacctac gcatgtatct 1440  
 atctcatttt cttacacctt ctattacctt ctgctctctc tgatttggaa aaagctgaaa 1500  
 aaaaagggtg aaaccagttc cctgaaatta tteccctatt tgactaataa gtatataaag 1560  
 acggtaggta ttgattgtaa ttctgtaaat ctatttetta aacttettaa attctacttt 1620  
 tatagttagt ctttttttta gtttaaaaca ccaagaactt agtttcgaat aaacacacat 1680  
 aaacaaacaa atctagaatg atcagaetta ctgttttctt cacagccgtt tttgcagcag 1740  
 tagcttcttg tgttccagtt gaattggata agagaaatac aggtcatttc caagcttact 1800  
 ctggttacac tgtggctaga tctaacttca cacaatggat tcatgaacag cctgccgtga 1860  
 gttggtacta tttgctacaa aacattgatt acctgaggg tcaattcaaa tcagctaagc 1920  
 cagggtgtgt tgtcgcgagc ccatcaactt ctgaaccaga ttacttctac caatggacta 1980  
 gagataccgc aataaccttc ttatctctaa tcgcagaggt agaagatec tctttttcaa 2040  
 atactaccct ggcaaaagtg gtcgagtact acatctcaaa cacatacacc ttgcagagag 2100  
 tctcaaaccc atcaggaaac ttcgattctc ctaatcatga cggetttaga gaaccaaaagt 2160  
 ttaatgttga cgataccgct tatactgcat cttggggtag accacagaat gatggccctg 2220  
 ccttacgtgc atacgccatt tccagatata tcaacgtgt agcgaagcac aacaacggta 2280  
 agctgctttt agctgggtcaa aatgggatac catactcttc cgcttcagac atttactgga 2340  
 agattatcaa accagacttg cagcatgtca gtacacattg gtcaacttct ggttttgatt 2400  
 tgtgggaaga gaaccaaggc actcacttct ttacagcctt gttcaacta aaggcattgt 2460  
 cttacggaat cccttgtcc aagacataca atgatcctgg attcactagt tggctagaaa 2520  
 agcaaaaagga tgcaactgaac tcatacatta acagttcagg ctttgtgaac tccggtaaaa 2580  
 agcatattgt taaaagccca caactatcta gcagaggtgg tttagattct gcaacctaca 2640  
 tagcagcctt gatcacacac gacattgggg atgacgatac atacacacca ttcaacgtcg 2700  
 acaattcata cgttttgaat agcttatact acctactggt agataacaaa aacagatata 2760  
 agatcaatgg caactacaag gccggtgctg ccgtaggaag ataccctgaa gatgtctaca 2820  
 acggagttgg tacatcagaa ggtaacctat ggcaattagc aacagcatat gcgggccaga 2880  
 cattttacac tttggcttac aattcattga aaaacaaaaa aaatttagtg atagaaaagc 2940  
 ttaactatga cctttacaac tctttcattg ccgatttate caagattgat tectctacg 3000  
 catcaaagga ctcttgaca cttacatacg gttctgacaa ctacaaaaat gttatcaagt 3060  
 ctctcttgca atttgggtgat tctttcttga aggttttact cgatcatata gatgataatg 3120  
 gtcaactaac tgaggaaaac aacagataca ctgggttcca agctggagct gtctctttaa 3180  
 catggagttc agggagtttg ttatctgcta acagagcgcg taacaaactt attgagcttc 3240

tgtagttaat taaacaggcc ccttttcctt tgtegatate atgtaattag ttatgtcacg 3300  
 cttacattca cgccctcctc ccacatccgc tctaaccgaa aaggaaggag ttagacaacc 3360  
 tgaagtctag gtccctatth atttttttat agttatgta gtattaagaa cgttatthtat 3420  
 atttcaaatt tttctttttt ttctgtacaa acgcgtgtac gcatgtaacg ggcagacggc 3480  
 cggccataac ttctgtataat gtatgtata cgaagttatc cttacatcac acccaatccc 3540  
 ccacaagtga tccccacac accatagctt caaaatgttt ctactccttt tttactcttc 3600  
 cagatthttct cggactccgc gcatcgccgt accacttcaa aacaccaag cacagcatac 3660  
 taaatthccc ctctthcttc ctctagggtg gcgttaatta cccgtactaa aggtthtgaa 3720  
 aagaaaaaag agaccgcctc gthctthttt ctctgtcgaa aaaggcaata aaaatthtta 3780  
 tcacgthttct thttcttgaa aaatthtttt thtgatthtt ttctctthtcg atgacctccc 3840  
 attgatathtt aagttaataa atggtcttca atthctcaag thttagthtt gthttthcttg 3900  
 thctattaca actthtttta ctctctgtct attagaaaga aagcatagca atctaatcta 3960  
 agthtttaatt acaaaatgcc acaatctctg gaagaattgg ccgccgacaa acgtgccctg 4020  
 thggctaaaa ccttctctga cgaatggaag gthcaactt tgcttgcga agatthcctt 4080  
 attgatthcc caaagaagtc cggatthttg tctgagctg aattgaagat taccgaagcc 4140  
 tctgtctctg atthggctc caagthggcc gctggctgag tgacttctgt tgaagctact 4200  
 thggctthtt gtaagagagc tgctatthct caacaattaa ccaactgtgc tcacgaattc 4260  
 thcccagatg ctgctthtagc tcaagctaga gaattagatg aatactacgc taagcataag 4320  
 agaccagthg gtccattaca cggthttacca atctctthta aggaccaatt gcgtgttaag 4380  
 ggttacgaaa cctccatggg thacatthcc tggthaaaca aatactatga aggtgattcc 4440  
 gtcttaacca ccatgttgag aaaagctggg gctgthttct acgttaagac ctctgtccca 4500  
 caaaccttga tggctctgtga aaccgtcaac aacatcattg gtagaactgt caatccaaga 4560  
 aacaaaaaatt ggtcctgtgg tggthcttct ggtgggtgaag gtgctatthg tggthattaga 4620  
 ggtgggtgta thgggtctcg tactgacatt ggtggthcca thtagatccc agctgctthc 4680  
 aactthtttat acggtthtag accatctcac ggtagatthc catatgctaa aatggctaac 4740  
 tctatggaag gtcaagaaac cgttctctc gctgthggc ctatctctca ctccgtcgaa 4800  
 gacttgagat thttcaccaa atctgtcttg ggtcaagaac ctthggaagta cgactctaac 4860  
 gtcactccca tgccatggag acaatctgaa tctgacatca thgctctaa gattaagaat 4920  
 ggtggthttga acatthgth thacaatthc gacggtaac thttgccaca cccaccaatt 4980  
 thactggthg tcgaaactac cgtthccctt thggc 5015  
 <210> 19  
 <211> 4771  
 <212> DNA  
 <213> 人工序列  
 <220>  
 <223> pTAN39  
 <400> 19  
 ggccgcgaag gtgctatthg thgtattaga ggtgggtgth thgggtgtcgg tactgacatt 60  
 ggtggthcca thtagatccc agctgctthc aactthtttat acggtthtag accatctcac 120

ggtagattgc catatgctaa aatggctaac tctatggaag gtcaagaaac cgttcactcc 180  
 gtcgttggtc ctatcactca ctccgtcgaa gacttgagat tgttcaccaa atctgtcttg 240  
 ggtcaagaac cttggaagta cgactctaag gtcateccaa tgccatggag acaatctgaa 300  
 tctgacatca ttgcctctaa gattaagaat ggtgggttga acattgggta ttacaatttc 360  
 gacggtaacg tcttgccaca cccaccaatt ttacgtgggtg tcgaaactac cgttgccgct 420  
 ttggccaagg ctggtcacac cgttactcca tggactccat acaagcatga tttcggctcat 480  
 gacttgattt cccacatcta tgctgctgat ggttctgccg acgtcatgag agacatttct 540  
 gcctctgggtg agccagccat ccctaacatt aaggacttgt tgaacccaaa tattaaggct 600  
 gttaacatga acgaattgtg ggacactcat ttacaaaagt ggaactatca aatggaatac 660  
 ttgaaaaagt ggcgtgaagc tgaagaaaaa gctggtaagg aattggacgc tattatcgct 720  
 ccaattactc ctaccgccgc tgtcagacac gatcaattca gatactacgg ttacgcctcc 780  
 gttattaact tattggattt cacctctgtt gtcgtcccag tcaacttctgc tgataagaat 840  
 attgataaga agaacgaate ttttaaaget gtttccgaat tggatgcttt ggttcaagaa 900  
 gaatacgacc cagaggetta tcacggtget cctgttgetg ttcaagttat tggtagaaga 960  
 ttgtccgaag agagaacttt ggetatcgcc gaagaagtcg gtaaattgtt gggtaacgct 1020  
 gtcactccat aagcgaattt cttatgattt atgattttta ttattaaata agttataaaa 1080  
 aaaataagtg tatacaaatt ttaaagtgac tcttaggttt taaaacgaaa attcttattc 1140  
 ttgagtaact ctttctgta ggtcaggttg ctttctcagg tatagcatga ggtcgtcttt 1200  
 attgaccaca cctctaccgg catgccgagc aaatgcctgc aaatcgtcc ccatttcacc 1260  
 caattgtaga tatgctaact ccagcaatga gttgatgaat ctcggtgtgt attttatgtc 1320  
 ctcagaggac aacacataac ttcgtataat gtatgctata cgaagttatc tcgagggcca 1380  
 gaaaaaggaa gtgtttccct ctttcttgaa ttgatgttac cctcataaag cacgtggcct 1440  
 cttatcgaga aagaaattac cgtcgtcgtg gatttgtttg caaaaagaac aaaactgaaa 1500  
 aaaccagac acgctcgact tctgtcttc ctgttgattg cagcttccaa tttcgtcaca 1560  
 caacaaggtc ctagcgacgg ctcacagggt ttgtaacaag caatcgaagg ttctggaatg 1620  
 gcgggaaagg gtttagtacc acatgctatg atgccactg tgatctccag agcaaagtcc 1680  
 gttcgatcgt actgttactc tctctcttc aaacagaatt gtccgaatcg tgtgacaaca 1740  
 acagcctgtt ctcacacact ctttctctt aaccaagggg gtggtttagt ttagtagaac 1800  
 ctcgtgaaac ttacatttac atatatataa acttgcataa attggtcaat gcaagaaata 1860  
 catatttggc ctttctctaat tcgtagtttt tcaagttctt agatgcttc ttttctctt 1920  
 ttttacagat catcaaggaa gtaattatct actttttaca agtctagaat gattagatta 1980  
 acagtatttc ttacagccgt tttcgcagcc gtcgcactct gtgttccagt agaattagat 2040  
 aagcgtaata caggacattt tcaagettac tetggtata cagttgcgag atctaacttt 2100  
 acacaatgga ttacgaaca gccagcagtt tcttggtact atttgcctca aaacatcgac 2160  
 taccctgaag gccaatcaa gtctgcaaag ccaggagtgg tcgtccttc tctagtact 2220  
 tcagaaccag attacttcta ccagtggaaca agagacactg ctattacctt cctgagetta 2280  
 atcgtgaag ttgaagatca ctcttttct aatacaacac tggccaaagt agttgagtac 2340  
 tacatctcta aacttacac tctacaaaga gtgtcaaacc cttctgggaa cttcgacagc 2400  
 ccaaacatg atggtttggg ggagccaaaa ttcaacgttg atgatacagc ctacaccgca 2460

tcttggggta gaccacaaaa cgacggacca gctttaagag catacgcaat atctcgttac 2520  
cttaatgctg ttgcaaagca caataatgga aagttgttgt tggctggtea aaacggattt 2580  
ccttactctt cagcatctga tatctactgg aagattatca agccagatct tcaacacgta 2640  
tccacacatt ggtaaacctc cggttcgat ttatggggagg aaaatcaggg tacacatttc 2700  
ttcaccgctc tagtgcaatt gaaggctttg agttacggca ttccattgtc taagacttac 2760  
aacgatcctg gtttcacctc atggcttgaa aagcagaagg atgccctgaa tagctacatc 2820  
aacatcatctg gttttgttaa ctccaggaaa aagcatatag ttgaatcccc acaactatca 2880  
tcaagaggag gtttagactc cgccacatac attgctgect tgattacaca tgatattggg 2940  
gatgatgaca cataactcc atttaacgtc gataacagtt atgtccttaa ttccttatac 3000  
tatttgttgg tcgataacaa aaatagatac aaaatcaacg gcaactacaa ggctggcgca 3060  
gcggtgggta gataccctga ggatgtttac aatgggttag gtacatctga aggcaatcca 3120  
tggcaattag cgactgctta cgctggacaa actttctaca cacttgcgta caactcattg 3180  
aaaaacaaaa aaaacctagt cattgaaaag ttgaattacg atctgtacaa ctctttcatt 3240  
gcagacctat caaagattga ctcatcttat gcaagtaaag attcactaac tttaacctac 3300  
ggtagtgata actacaaaaa cgttatcaag tctttactec agtttgggta ttctattctg 3360  
aagggttgt tagatcata agacgacaat ggtcaactca cagaggagat aaacagatac 3420  
actggttttc aagcaggagc tgtttcactt acttggctca gtggttcttt gctttccgcc 3480  
aacagagcca gaaacaagct catcgaatta ctatagttaa ttaagaagtt ttgttagaaa 3540  
ataaatcatt ttttaattga gcattcttat tctattttta tttaaatagt tttatgtatt 3600  
gtagctaca tacaacagtt taaatcaaat tttctttttc ccaagtccaa aatggaggtt 3660  
tattttgatg acccgcagc gattatgttt tgaaagtata agactacata catgtacata 3720  
tatttaaaaa tgtaaaccg tccattatat tgccgggcag acggccggcc ttatagccta 3780  
gctttaaggc tactttaaaa actttttatt tattcataca catatattat cgaacattcg 3840  
tataacttaa tatcattcaa aaaaaaaaaa aaaaaaaaaa gaaaacatat acacatatat 3900  
atztatgttt atagagagag agagagaaaa tttgaatttt tgaatcattt gcaaagttat 3960  
atgttttata cattatttat tcattttttt tgggtctcag gacattgtgc tgttcagaga 4020  
accacttaaa atacgcacg tctctgtaaat atccacttc attaaaaacc ttattcactt 4080  
ctaactttgc ctcaactcc ttcttgaggt tttctcctt tttttctga acaagctcaa 4140  
ccagatataa tggttcgttc tttcgaact ttgtctttac atatatttcc tctttgtac 4200  
ctcttctctt tccacataa acagtccct tttcaataa acgagagaaa taccagaaaa 4260  
gtagcgagag aacaaaatat gcgctacca aaagcttttg atacgtaaca atctgatctc 4320  
tctcaaatTT tttatccaag aagaaactca aaccagctac aacagctatg gaataaccta 4380  
tgtacaattt agcatcgagt aaagcgtatg atctctcgta atttaatctc gcgaaaacag 4440  
aaggtagggc ttcatctaaa gcttggttca actccgggat tgaatataca ttaataggtt 4500  
tagcagaact catcttgaac aggcgtctct tttcttaca ataacttgtg cttttctctc 4560  
tataattccg tttcaacgtg tacaattgtc attttttgtc tgggtatgatt ttgcagaact 4620  
gaaaaaatct cttaaagtgt ccgctcctc aagaaggcat attcctttac aaaagtacat 4680  
tgatcttaca agaagctagc taatgtact atttaaaaa cactacact ccatcaatac 4740  
ataaaattgt tatgatagac ttgagggac g 4771



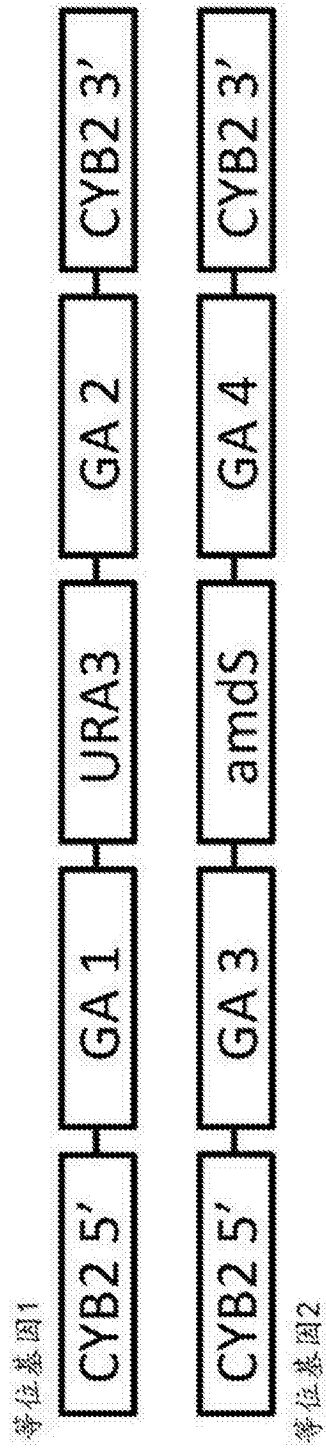


图1

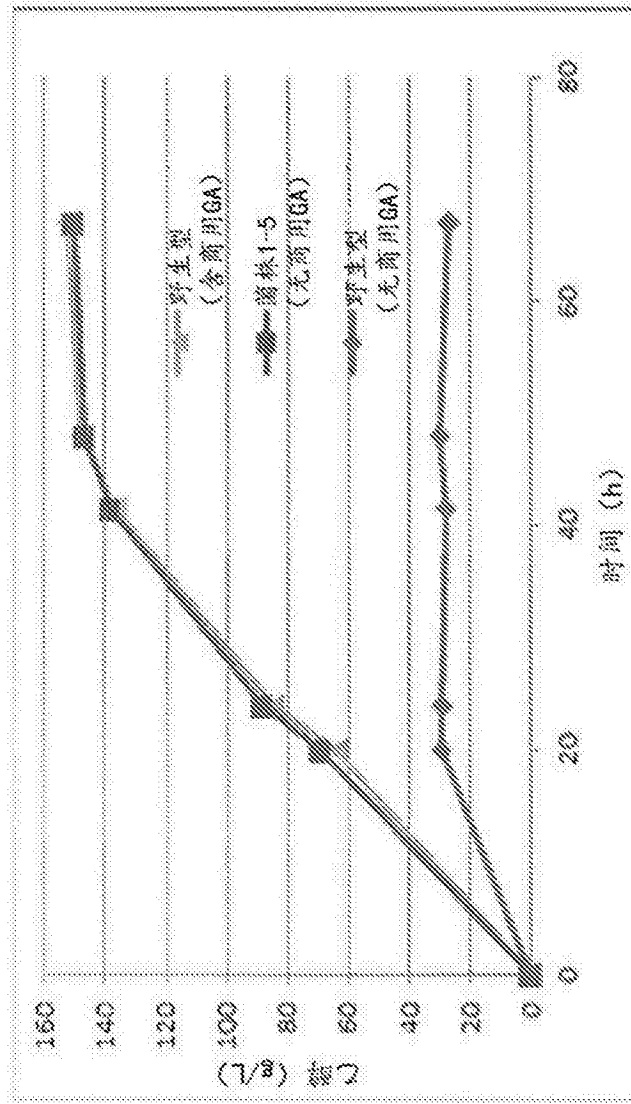


图2

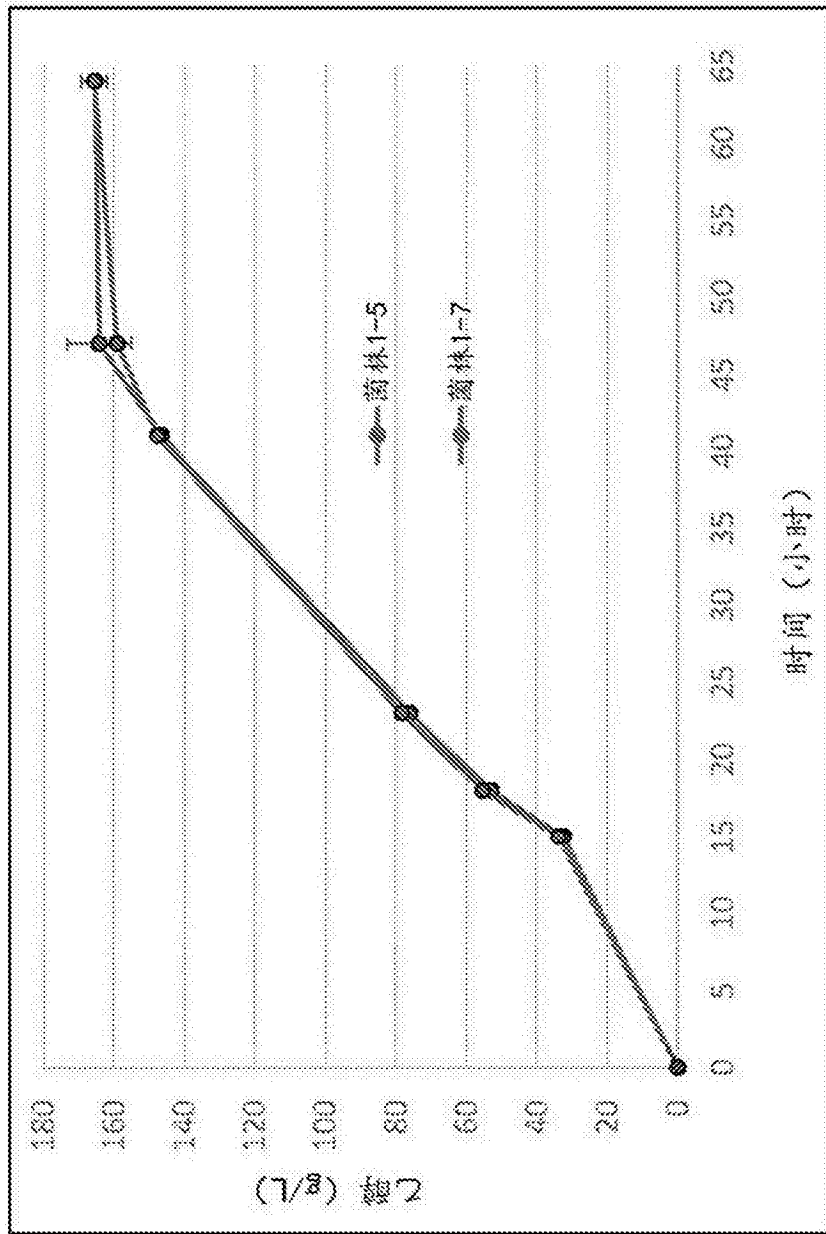


图3