



(12)发明专利

(10)授权公告号 CN 104379162 B

(45)授权公告日 2017.03.15

(21)申请号 201380024824.6  
(22)申请日 2013.03.14  
(65)同一申请的已公布的文献号  
    申请公布号 CN 104379162 A  
(43)申请公布日 2015.02.25  
(30)优先权数据  
    61/611,485 2012.03.15 US  
(85)PCT国际申请进入国家阶段日  
    2014.11.12  
(86)PCT国际申请的申请数据  
    PCT/IB2013/000912 2013.03.14  
(87)PCT国际申请的公布数据  
    W02013/136189 EN 2013.09.19  
(73)专利权人 奥克西雷恩英国有限公司  
    地址 英国曼彻斯特  
(72)发明人 W.沃维肯 K.C.T.A.M.皮恩斯  
    J.R.L.斯陶特 G.N.派纳尔特  
(74)专利代理机构 北京市柳沈律师事务所  
    11105  
    代理人 张文辉

(51)Int.Cl.  
    A61K 38/47(2006.01)  
    C12N 9/26(2006.01)  
(56)对比文件  
    WO 2011039634 A2,2011.04.07,  
    YUNXIANG ZHU et al.Glycoengineered  
    acid [alpha]-glucosidase with improved  
    efficacy at correcting the metabolic  
    aberrations and motor function deficits  
    in a mouse model of pompe disease.  
    《MOLECULAR THERAPY》.2009,第17卷(第6期),  
    MORELAND RODNEY J et al.lysosomal  
    acid alpha-glucosidase consists of four  
    different peptides processed from a  
    single chain precursor.《JOURNAL OF  
    BIOLOGICAL CHEMISTRY, AMERICAN SOCIETY  
    FOR BIOCHEMISTRY AND MOLECULAR BIOLOGY》  
    .2005,第280卷(第8期),  
    审查员 吴希哲

权利要求书2页 说明书32页  
序列表18页 附图10页

(54)发明名称  
    用于治疗蓬佩氏病的方法和材料

(57)摘要  
    本文件涉及具有酸性 $\alpha$ 1pha葡糖苷酶活性和  
    至少一处修饰的分子复合物,所述修饰导致分子  
    复合物被转运到哺乳动物细胞内部的能力增强。

1. 一种用于生成分子复合物的方法,所述方法包括使与SEQ ID NO:1中所列的氨基酸序列具有至少95%序列同一性的多肽接触与SEQ ID NO:8中所列的氨基酸序列具有至少90%序列同一性的蛋白酶,其中所述蛋白酶在氨基酸50和氨基酸74之间的一个或多个位点处切割所述多肽,其中所述分子复合物具有GAA活性。

2. 权利要求1的方法,其中所述多肽具有SEQ ID NO:1中所列的氨基酸序列。

3. 权利要求1或2任一项的方法,其中在体外实施所述接触。

4. 权利要求1或2任一项的方法,其中在重组真菌细胞中发生所述接触。

5. 权利要求4的方法,其中所述真菌细胞是解脂耶氏酵母(*Yarrowia lipolytica*)、*Arxula adeninivorans*、或甲基营养酵母(*methylotrophic yeast*),例如选自假丝酵母属(*Candida*)、汉逊酵母属(*Hansenula*)、*Oogataea*、毕赤酵母属(*Picha*)和球拟酵母属(*Torulopsis*)的甲基营养酵母。

6. 权利要求1至5任一项的方法,其中所述蛋白酶在氨基酸60和氨基酸65之间的一个或多个位点处切割所述多肽。

7. 权利要求1-6任一项的方法,进一步包括在氨基酸719和氨基酸746之间的一个或多个位点处和/或氨基酸137和氨基酸151之间的一个或多个位点处蛋白水解所述多肽。

8. 权利要求1-7任一项的方法,进一步包括使用至少一个修饰改变所述分子复合物,所述修饰导致所述分子复合物被运输到哺乳动物细胞内部的能力增强。

9. 权利要求8的方法,其中所述至少一个修饰包括胞外受体的配体,或其中所述至少一个修饰包括人胰岛素样生长因子II的识别域。

10. 权利要求8的方法,其中所述分子复合物的至少一种多肽包括一个或多个磷酸化N-聚糖,并且其中所述修饰包括至少一个磷酸化N-聚糖的脱帽和脱甘露糖基化。

11. 权利要求10的方法,其中所述多肽上至少40%的N-聚糖是脱帽且脱甘露糖基化的,或其中所述多肽上至少60%的N-聚糖是脱帽且脱甘露糖基化的,或其中所述多肽上至少80%的N-聚糖是脱帽且脱甘露糖基化的,或其中所述多肽上至少90%的N-聚糖是脱帽且脱甘露糖基化的。

12. 一种分离的真菌细胞,其包含编码与SEQ ID NO:1中所列的GAA氨基酸序列具有至少95%序列同一性的GAA氨基酸序列的核酸,和编码与SEQ ID NO:8中所列的氨基酸序列具有至少90%序列同一性的碱性蛋白酶的核酸,其中所述真菌细胞生成具有GAA活性且包含至少两个多肽的分子复合物,每个多肽与SEQ ID NO:1中所列的氨基酸序列的区段具有至少95%序列同一性,每个区段通过所述碱性蛋白酶对与SEQ ID NO:1中所列的氨基酸序列具有至少95%序列同源性的氨基酸序列在氨基酸50和氨基酸74之间的一个或多个位点处的蛋白水解衍生。

13. 权利要求12的分离的真菌细胞,每个区段通过所述碱性蛋白酶对与SEQ ID NO:1中所列的氨基酸序列具有至少95%序列同一性的氨基酸序列在氨基酸60和氨基酸65之间的一个或多个位点处的蛋白水解衍生。

14. 权利要求12或权利要求13的真菌细胞,所述真菌细胞进一步包含:

1) 编码甘露糖苷酶的核酸,所述甘露糖苷酶能够(i)将甘露糖-1-磷酸-6-甘露糖模块水解成甘露糖-6-磷酸及(ii)水解末端 $\alpha$ -1,2甘露糖、 $\alpha$ -1,3甘露糖和/或 $\alpha$ -1,6甘露糖连接;或

2) 编码甘露糖苷酶的核酸,所述甘露糖苷酶能够将甘露糖-1-磷酸-6-甘露糖模块水解成甘露糖-6-磷酸;或

3) 编码甘露糖苷酶的核酸,所述甘露糖苷酶能够水解末端 $\alpha$ -1,2甘露糖、 $\alpha$ -1,3甘露糖和/或 $\alpha$ -1,6甘露糖连接;或

4) 编码能够促进甘露糖基磷酸化的多肽的核酸。

15. 权利要求12-14中任一项的真菌细胞,其中所述真菌细胞被遗传工程化改造为OCH1活性缺陷。

## 用于治疗蓬佩氏病的方法和材料

[0001] 对相关申请的交叉引用

[0002] 本申请要求2012年3月15日提交的美国临时申请流水号61/611,485的权益。认为先前申请的公开内容是本申请的公开内容(并且通过提及并入其中)的一部分。

### 发明领域

[0003] 本发明涉及具有酸性 $\alpha$ 葡糖苷酶活性的分离的分子复合物,且更具体地涉及包含通过蛋白水解从前体分子衍生的至少两个多肽的分子复合物,其中所述分子复合物包含至少一处修饰,其导致所述分子复合物被转运到哺乳动物细胞内部的能力增强。

[0004] 发明背景

[0005] 蓬佩(Pompe)氏病(又称为糖原贮积病(glycogen-storage disease) II型或酸性麦芽糖酶缺陷)是一种罕见的常染色体隐性病症,其由于酸性 $\alpha$ 葡糖苷酶(GAA)缺陷而导致糖原在溶酶体中的积累。糖原的积累引起遍及全体的进行性肌无力(progressive muscle weakness)(肌病),并且影响各个身体组织,包括心脏、骨骼肌、肝、和神经系统。

[0006] 蓬佩氏病被宽泛地分成婴儿期发作性和迟发性形式。在婴儿期发作性形式中,婴儿通常在婴儿期早期(4-8个月龄)期间呈现无力和懒散(floppiness),并且不能举起他们的手,而且不能进行对于其年龄而言常见的其他运动任务,诸如翻身。在不治疗的情况下,具有蓬佩氏病的婴儿通常由于心力衰竭(heart failure)和呼吸虚弱(respiratory weakness)而在12个月龄前死亡。参见United Pompe Foundation。迟发性形式(包括青少年和成年形式)比婴儿期形式具有更晚的发作并且更慢地进展。重组人GAA(Myozyme®或Lumizyme®)用于治疗蓬佩氏病。然而,Myozyme®或Lumizyme®都是非常昂贵的,每年费用完全超过300,000美元。因而,需要改善的蓬佩氏病治疗。

[0007] 发明概述

[0008] 在一个方面,本文件的特征在于一种分离的分子复合物,其具有酸性 $\alpha$ 葡糖苷酶(GAA)活性,且包含至少两个多肽(例如至少三个或至少四个多肽),每个多肽与SEQ ID NO:1中所列的氨基酸序列的区段具有至少85%(例如至少90%、95%、99%、或100%)序列同一性,每个区段通过在氨基酸50和氨基酸74之间(例如氨基酸56和氨基酸68之间或氨基酸60和氨基酸65之间)的一个或多个位点处对SEQ ID NO:1中所列的氨基酸序列的蛋白水解衍生。所述分子复合物包含至少一处修饰,其导致所述分子复合物被转运到哺乳动物细胞内部的能力增强。对SEQ ID NO:1中所列的氨基酸序列的蛋白水解进一步可以包括在SEQ ID NO:1中所列的氨基酸序列的氨基酸719和氨基酸746之间的一个或多个位点处的切割或在SEQ ID NO:1中所列的氨基酸序列的氨基酸137和氨基酸151之间的一个或多个位点处的切割。蛋白水解进一步可以包括在SEQ ID NO:1中所列的氨基酸序列的氨基酸719和氨基酸746之间的一个或多个位点处的切割和在SEQ ID NO:1中所列的氨基酸序列的氨基酸137和氨基酸151之间的一个或多个位点处的切割。

[0009] 在本文中描述的任何分子复合物中,至少一个多肽可以包含一个或多个磷酸化的N-聚糖,且所述修饰可以包含至少一个磷酸化的N-聚糖的脱帽(uncapping)和脱甘露糖基

化(demansylation)。至少一个多肽上的至少40% (例如至少60%、80%、90%、95%、或99%)的N-聚糖可以是脱帽且脱甘露糖基化的。

[0010] 在本文中描述的任何分子复合物中,对于所述至少两个多肽之一,所述区段包含SEQ ID NO:1的氨基酸22至57,且其中对于所述至少两个多肽之一,所述区段包含SEQ ID NO:1的氨基酸66至896。

[0011] 在含有至少三个多肽的本文中描述的任何分子复合物中,对于所述至少三个多肽之一,所述区段包含SEQ ID NO:1的氨基酸22至57,其中对于所述至少三个多肽之一,所述区段包含SEQ ID NO:1的氨基酸66至726,且其中对于所述至少三个多肽之一,所述区段包含SEQ ID NO:1的氨基酸736至896。

[0012] 在含有至少四个多肽的本文中描述的任何分子复合物中,对于所述至少四个多肽之一,所述区段包含SEQ ID NO:1的氨基酸22至57,其中对于所述至少四个多肽之一,所述区段包含SEQ ID NO:1的氨基酸66至143,其中对于所述至少四个多肽之一,所述区段包含SEQ ID NO:1的氨基酸158至726,且其中对于所述至少四个多肽之一,所述区段包含SEQ ID NO:1的氨基酸736至896。

[0013] 在本文中描述的任何分子复合物中,所述至少一处修饰可以包括与分子复合物中至少一个多肽融合的下列至少一种:胞外受体的配体,结合靶细胞表面上的受体胞外域的靶向域,尿激酶型纤维蛋白溶酶原受体,或人胰岛素样生长因子II的识别域。

[0014] 本文件的特征还在于包含本文中描述的任何分子复合物的组合物,其中分子复合物是冻干的。所述组合物可以以一次性管形瓶(single use vial)包装。

[0015] 本文件的特征还在于药物组合物,其包含本文中描述的任何分子复合物和药学可接受载体。所述组合物可以被配制用于静脉内或皮下施用。所述组合物可以被配制用于静脉内输注。

[0016] 在另一个方面,本文件的特征在于治疗蓬佩氏病的方法。所述方法包括对诊断为蓬佩氏病的患者施用本文中描述的任何组合物。所述患者可以诊断为婴儿期发作性蓬佩氏病或迟发性蓬佩氏病。

[0017] 本文件的特征还在于生成分子复合物的方法。所述方法包括使与SEQ ID NO:1中所列的氨基酸序列具有至少85%序列同一性的多肽与蛋白酶接触,所述蛋白酶与SEQ ID NO:8中所列的氨基酸序列具有至少85% (例如至少90%,至少95%,至少99%,或100%)序列同一性,其中所述蛋白酶在氨基酸50和氨基酸74之间(例如氨基酸56和氨基酸68之间或氨基酸60和氨基酸65之间)的一个或多个位点处切割所述多肽。可以在体外实施接触步骤。

[0018] 本文件的特征还在于用于生成包含脱帽且脱甘露糖基化的磷酸化N-聚糖的分子复合物的方法。该方法包括使分子复合物与甘露糖苷酶接触,所述甘露糖苷酶能够(i)将甘露糖-1-磷酸-6-甘露糖模块水解成甘露糖-6-磷酸及(ii)水解末端alpha-1,2甘露糖,alpha-1,3甘露糖和/或alpha-1,6甘露糖连接,所述分子复合物具有GAA活性且包含至少两个多肽,每个多肽与SEQ ID NO:1中所列的氨基酸序列的区段具有至少85%序列同一性,每个区段通过在氨基酸50和氨基酸74之间(例如氨基酸56和氨基酸68之间或氨基酸60和氨基酸65之间)的一个或多个位点处对SEQ ID NO:1中所列的氨基酸序列的蛋白水解衍生,其中在接触前,至少一个所述多肽包含含有一个或多个甘露糖-1-磷酸-6-甘露糖模块的磷酸化N-聚糖。甘露糖苷酶可以是家族38糖基水解酶(例如直生刀豆(*Canavalia ensiformis*)甘

露糖苷酶或解脂耶氏酵母 (*Yarrowia lipolytica*) 甘露糖苷酶)。可以在表达甘露糖苷酶的重组真菌细胞中发生所述接触。

[0019] 本文件的特征还在于生成包含脱帽且脱甘露糖基化的磷酸化N-聚糖的分子复合物的方法。该方法包括使分子复合物与甘露糖苷酶接触,所述甘露糖苷酶能够水解末端 $\alpha$ -1,2甘露糖、 $\alpha$ -1,3甘露糖和/或 $\alpha$ -1,6甘露糖连接,所述分子复合物具有GAA活性且包含至少两个多肽,每个多肽与SEQ ID NO:1中所列的氨基酸序列的区段具有至少85%序列同一性,每个区段通过在氨基酸50和氨基酸74之间(例如氨基酸56和氨基酸68之间或氨基酸60和氨基酸65之间)的一个或多个位点处对SEQ ID NO:1中所列的氨基酸序列的蛋白水解衍生,其中至少一个多肽在接触前包含含有脱帽的甘露糖-6-磷酸模块的磷酸化N-聚糖。甘露糖苷酶可以是家族47糖基水解酶(例如斋藤曲霉 (*Aspergillus satoi*) 甘露糖苷酶),家族92糖基水解酶(例如纤维化纤维微细菌 (*Cellulosimicrobium cellulans*) 甘露糖苷酶),或家族38糖基水解酶(例如直生刀豆甘露糖苷酶)。可以在表达甘露糖苷酶的重组真菌细胞中发生所述接触。

[0020] 本文件的特征还在于生成包含脱帽且脱甘露糖基化的磷酸化N-聚糖的分子复合物的方法。该方法包括使分子复合物与甘露糖苷酶接触,所述甘露糖苷酶能够将甘露糖-1-磷酸-6-甘露糖模块水解成甘露糖-6-磷酸,所述分子复合物具有GAA活性并且包含至少两个多肽,每个多肽与SEQ ID NO:1中所列的氨基酸序列的区段具有至少85%序列同一性,每个区段通过在氨基酸50和氨基酸74之间(例如氨基酸56和氨基酸68之间或氨基酸60和氨基酸65之间)的一个或多个位点处对SEQ ID NO:1中所列的氨基酸序列的蛋白水解衍生,其中在接触前,至少一个多肽包含一个或多个甘露糖-1-磷酸-6-甘露糖模块。甘露糖苷酶可以是家族38糖基水解酶(例如直生刀豆甘露糖苷酶或解脂耶氏酵母甘露糖苷酶)。

[0021] 在另一个方面,本文件的特征在于生成包含脱帽且脱甘露糖基化的磷酸化N-聚糖的分子复合物的方法。该方法包括a)使分子复合物与能够将甘露糖-1-磷酸-6-甘露糖模块水解成甘露糖-6-磷酸的甘露糖苷酶接触以使所述分子复合物中的至少一个多肽上的甘露糖-6-磷酸模块脱帽,所述分子复合物具有GAA活性且包含至少两个多肽,每个多肽与SEQ ID NO:1中所列的氨基酸序列的区段具有至少85%序列同一性,每个区段通过在氨基酸50和氨基酸74之间(例如氨基酸56和氨基酸68之间或氨基酸60和氨基酸65之间)的一个或多个位点处对SEQ ID NO:1中所列的氨基酸序列的蛋白水解衍生;并b)使所述分子复合物与能够水解末端 $\alpha$ -1,2甘露糖、 $\alpha$ -1,3甘露糖和/或 $\alpha$ -1,6甘露糖连接的甘露糖苷酶接触。可以通过两种不同酶催化或者通过单一酶催化步骤(a)和步骤(b)。接触步骤可以一起或分开且以任一次序实施。可以在重组真菌宿主细胞中发生接触,所述真菌宿主细胞表达能够催化步骤(a)的甘露糖苷酶和能够催化步骤(b)的甘露糖苷酶。可以在重组真菌宿主细胞中发生接触,所述真菌宿主表达能够催化步骤(a)和(b)的甘露糖苷酶。

[0022] 可以使用包含至少一个脱帽且脱甘露糖基化的N-聚糖的本文中描述的任何分子复合物接触哺乳动物细胞,其中在接触后,分子复合物以增强的效率转运到哺乳动物细胞的内部。哺乳动物细胞可以是人细胞。

[0023] 本文件的特征还在于将具有GAA活性的分子复合物转运到细胞内部的方法。所述方法包括使哺乳动物细胞与所述分子复合物接触,所述分子复合物包含至少两个多肽,每个多肽与SEQ ID NO:1中所列的氨基酸序列的区段具有至少85%序列同一性,每个区段通

过在氨基酸50和氨基酸74之间(例如氨基酸56和氨基酸68之间或氨基酸60和氨基酸65之间)的一个或多个位点处对SEQ ID NO:1中所列的氨基酸序列的蛋白水解衍生;其中至少一个所述多肽上的磷酸化N-聚糖已经如本文中描述的方法中所列的那样脱帽并脱甘露糖基化。哺乳动物细胞可以在体外或在哺乳动物受试者中。哺乳动物细胞可以是人细胞。

[0024] 在另一个方面,本文件的特征在于将具有GAA活性的分子复合物转运到细胞内部的方法。所述方法包括使哺乳动物细胞与所述分子复合物接触,所述分子复合物包含至少两个多肽,每个多肽与SEQ ID NO:1中所列的氨基酸序列的区段具有至少85%序列同一性,每个区段通过在氨基酸50和氨基酸74之间(例如氨基酸56和氨基酸68之间或氨基酸60和氨基酸65之间)的一个或多个位点处对SEQ ID NO:1中所列的氨基酸序列的蛋白水解衍生,所述分子复合物包含至少一处修饰,其导致分子复合物被转运到哺乳动物细胞内部的能力增强。哺乳动物细胞可以在体外或在哺乳动物受试者中。哺乳动物细胞可以是人细胞。所述修饰可以包括与分子复合物中至少一个多肽融合的下列任一种:胞外受体的配体,结合靶细胞表面上的受体胞外域的靶向域,尿激酶型纤维蛋白溶酶原受体,或人胰岛素样生长因子II的识别域。

[0025] 在另一个方面,本文件的特征在于包含编码碱性蛋白酶的外源核酸的分离的真菌细胞,所述碱性蛋白酶与SEQ ID NO:8中所列的氨基酸序列具有至少85%序列同一性。

[0026] 本文件的特征还在于分离的真菌细胞,其包含编码SEQ ID NO:1中所列的GAA氨基酸序列的核酸和编码与SEQ ID NO:8中所列的氨基酸序列具有至少85%序列同一性的碱性蛋白酶的核酸。所述真菌细胞生成具有GAA活性且包含至少两个多肽的分子复合物,每个多肽与SEQ ID NO:1中所列的氨基酸序列的区段具有至少85%序列同一性,每个区段通过所述碱性蛋白酶在氨基酸50和氨基酸74之间(例如氨基酸56和氨基酸68之间或氨基酸60和氨基酸65之间)的一个或多个位点处对SEQ ID NO:1中所列的氨基酸序列的蛋白水解衍生。在一些实施方案中,真菌细胞进一步包含编码甘露糖苷酶的核酸,所述甘露糖苷酶能够将甘露糖-1-磷酸-6-甘露糖模块水解成甘露糖-6-磷酸。在一些实施方案中,真菌细胞进一步包含编码甘露糖苷酶的核酸,所述甘露糖苷酶能够水解末端 $\alpha$ -1,2甘露糖, $\alpha$ -1,3甘露糖和/或 $\alpha$ -1,6甘露糖连接。在一些实施方案中,真菌细胞进一步可以包含编码甘露糖苷酶的核酸,所述甘露糖苷酶能够(i)将甘露糖-1-磷酸-6-甘露糖模块水解成甘露糖-6-磷酸并(ii)水解末端 $\alpha$ -1,2甘露糖, $\alpha$ -1,3甘露糖和/或 $\alpha$ -1,6甘露糖连接。任何此类真菌细胞进一步可以包括编码能够促进甘露糖基磷酸化的多肽的核酸和/或遗传工程化改造为OCH1活性缺陷。

[0027] 除非另有定义,本文中使用的所有技术和科学术语与本发明所属领域的普通技术人员的通常理解具有相同的意义。虽然与本文中描述的方法和材料相似或等同的方法和材料可以用于实施或测试本发明,下文描述了例示性的方法和材料。本文中提及的所有出版物,专利申请,专利,Genbank®登录号,和其它参考文献通过提及完整并入。在冲突的情况下,应以本申请(包括定义)为准。材料,方法和例子仅仅是例示的,而并不意图为限制的。

[0028] 本发明的其它特征和优点从以下详细描述及从权利要求书看会是显而易见的。

[0029] 附图简述

[0030] 图1是切割信号序列后人酸性 $\alpha$ 葡糖苷酶(GAA)的氨基酸序列(SEQ ID NO:1)的图。

[0031] 图2A是DsbA-纤维化纤维微细菌甘露糖苷酶5 (CcMan5) 的可读框 (ORF) 的核苷酸序列 (SEQ ID NO:2) 的图。

[0032] 图2B是具有粗体的信号序列的CcMan5多肽的氨基酸序列 (SEQ ID NO:3) 的图。

[0033] 图2C是没有信号序列的CcMan5多肽的氨基酸序列 (SEQ ID NO:4) 的图。没有信号序列的CcMan5多肽的预测分子量是173kDa。

[0034] 图3A和3B是一系列电泳图,其描绘了用CcMan5和JbMan处理的rhGAA的N-聚糖分析。使用DNA测序仪辅助的、荧光团辅助的碳水化合物电泳 (DSA-FACE) 实施分析。Y轴代表相对荧光单位,作为每个N-聚糖结构的量的指示。X轴代表每个N-聚糖结构通过毛细管的相对迁移率。在图3A和图3B两者中,小图A是含有用PNGaseF从RNaseB释放的N-聚糖的参照样品。在图3A中,小图B和C含有分别用CcMan5和JbMan处理之前和之后来自huGAA (76kDa变体) 的N-聚糖分析。在图3B中,小图B、C、和D分别含有来自huGAA76kDa形式、95kDa形式、和110kDa形式的N-聚糖分析。

[0035] 图4是使用家兔肝糖原作为底物用Myozyme (•), 76kDa GAA (▲), 95kDa GAA (▼), 和110kDa GAA (◆) 每分钟形成的葡萄糖量的线图。

[0036] 图5A含有心脏中各个小鼠的糖原水平 ( $\mu\text{g}/\text{mg}$ 蛋白质) 的两幅图。图5B含有骨骼肌中各个小鼠的糖原水平 ( $\mu\text{g}/\text{mg}$ 蛋白质) 的两幅图。红色点是雌性,黑色点是雄性。线代表每组的中值。

[0037] 图6含有解脂耶氏酵母AMS1甘露糖苷酶的氨基酸序列 (SEQ ID NO:5) 的图。

[0038] 图7含有斋藤曲霉甘露糖苷酶的氨基酸序列 (SEQ ID NO:6) 的图。

[0039] 图8含有纤维化纤维微细菌甘露糖苷酶4 (CcMan4, SEQ ID NO:7) 的氨基酸序列的图,信号序列为粗体。没有信号序列的CcMan4多肽的预测分子量是184kDa。

[0040] 图9含有包含信号肽 (21个氨基酸), 前肽 (pro-peptide) (100个氨基酸) 和成熟蛋白质 (282个氨基酸) 的米曲霉 (*Aspergillus oryzae*) 碱性蛋白酶的氨基酸序列 (SEQ ID NO:9) 的图。

[0041] 图10含有融合构建体的核苷酸序列的图,所述融合构建体含有编码米曲霉碱性蛋白酶的解脂耶氏酵母密码子优化的序列 (SEQ ID NO:10)。用于克隆的限制性位点加下划线。编码接头的核苷酸序列为粗体,并且编码His标签 (10个His残基) 的核苷酸序列为斜体。

[0042] 发明详述

[0043] 一般而言,本文件提供了分离的分子复合物,其具有酸性 $\alpha$ -葡糖苷酶 (GAA) 活性和至少一处修饰,该修饰导致被转运到哺乳动物细胞内部的能力增强。GAA以含有N-连接的聚糖的110kDa前体合成。前体被蛋白水解加工以除去信号序列,然后进一步蛋白水解加工成95kDa, 76kDa, 和70kDa的主要种类。然而,从前体释放的至少一些肽仍然与主要种类联合。参见例如Moreland等, *J. Biol. Chem.*, 280:6780-6791 (2005)。如此,具有本文中描述的GAA活性的分子复合物包含至少两个多肽 (至少两个,三个,或四个多肽),其通过在一个或多个位点处对前体分子的蛋白水解切割衍生。分子复合物中的至少两个多肽源自前体中一个或多个位点处的蛋白水解切割。例如,SEQ ID NO:1中所列的氨基酸序列的蛋白水解可以介于氨基酸50和氨基酸74之间,例如在氨基酸56和氨基酸68之间或在氨基酸60和氨基酸65之间,以生成至少两个多肽。含有两个多肽的分子复合物在本文中称为95kDa形式。

[0044] 在一些实施方案中,分子复合物中的至少三个多肽源自前体中两个或更多个位点

的蛋白水解切割。例如,SEQ ID NO:1中所列的氨基酸序列的蛋白水解在氨基酸50和氨基酸74之间(例如氨基酸50和氨基酸74之间或氨基酸60和氨基酸65之间)的切割外可以包括SEQ ID NO:1中所列的氨基酸序列的氨基酸719和氨基酸746之间的一个或多个位点处的切割或氨基酸137和氨基酸151之间的一个或多个位点处的切割。含有三个多肽的分子复合物在本文中称为76kDa形式。

[0045] 在一些实施方案中,分子复合物中的至少四个多肽源自前体中三个或更多个位点处的蛋白水解切割。例如,SEQ ID NO:1中所列的氨基酸序列的蛋白水解在氨基酸50和氨基酸74之间(例如氨基酸56和氨基酸68之间或氨基酸60和氨基酸65之间)的切割外可以包括SEQ ID NO:1中所列的氨基酸序列的氨基酸719和氨基酸746之间的一个或多个位点处的切割或SEQ ID NO:1中所列的氨基酸序列的氨基酸137和氨基酸151之间的一个或多个位点处的切割。含有四个多肽的分子复合物在本文中称为70kDa形式。

[0046] 应当领会,可以在一个分子中的一个或多个位点处发生切割,且切割位点在不同分子中可以是不同的。

[0047] 可以使用含有来自米曲霉的蛋白酶的商品化蛋白酶混合物(例如来自Sigma或NovozymesCorp)在氨基酸50和74之间,例如在氨基酸56和68之间或氨基酸60和65之间切割SEQ ID NO:1中所列的氨基酸序列。或者,可以使用与来自米曲霉的碱性蛋白酶(SEQ ID NO:8)具有至少85%(例如,至少90%、95%、97%、98%、99%、或100%)序列同一性的碱性蛋白酶。例如,如本文中描述的,可以使具有SEQ ID NO:1中所列的氨基酸序列的GAA多肽与蛋白酶接触,所述蛋白酶与SEQ ID NO:8或SEQ ID NO:9中所列的氨基酸序列具有至少85%序列同一性。SEQ ID NO:8是成熟米曲霉碱性蛋白酶的氨基酸序列。SEQ ID NO:9是包含信号肽、前肽和成熟蛋白的米曲霉蛋白酶的氨基酸序列。可以使用自米曲霉分离或已经重组生成的蛋白酶在体外发生接触。或者,可以将真菌宿主工程化改造为使得GAA多肽和碱性蛋白酶都被分泌到培养基中,其中碱性蛋白酶可以在氨基酸50和氨基酸74之间(例如氨基酸56和68之间或氨基酸60和氨基酸65之间)切割SEQ ID NO:1中所列的氨基酸序列。

[0048] 本文中描述的分离的分子复合物具有至少一处修饰,该修饰导致被转运到哺乳动物细胞内部的能力增强。增强复合物被转运到哺乳动物细胞内部的能力的修饰的非限制性例子包括磷酸化的N-聚糖的脱帽和脱甘露糖基化或促进转运的肽标签。本文中描述了用于制备含有标签或脱帽且脱甘露糖基化的N-聚糖的分子复合物的方法和材料。

[0049] 本文中描述的分离的分子复合物特别可用于治疗蓬佩氏病患者,包括诊断为蓬佩氏病,即婴儿期发作性蓬佩氏病和迟发性蓬佩氏病两者的患者。蓬佩氏病由于GAA缺陷而导致糖原在溶酶体中的积累。糖原的积累引起遍及全身的进行性肌无力(肌病),并且影响各个身体组织,包括心脏、骨骼肌、肝、和神经系统。

[0050] 分子复合物中的每个多肽与SEQ ID NO:1中所列的氨基酸序列的区段具有至少85%序列同一性(例如至少90%、95%、97%、98%、99%、或100%),每个区段通过在氨基酸50和氨基酸74之间(例如氨基酸56和氨基酸68之间或氨基酸60和氨基酸65之间)的一个或多个位点处对SEQ ID NO:1中所列的氨基酸序列的蛋白水解衍生。可以如下测定特定的氨基酸序列和SEQ ID NO:1中所列的氨基酸序列之间的百分比同一性。首先,使用来自含有BLASTP第2.0.14版的BLASTZ单机版本的BLAST 2 Sequences (B12seq)程序比对氨基酸序列。BLASTZ的此单机版本可以获自Fish&Richardson站点(例如www.fr.com/blast/)或美国

政府的国立生物技术信息中心站点([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov))。解释如何使用B12seq程序的指导可以参见BLASTZ所附的自述文件。B12seq使用BLASTP算法实施两种氨基酸序列之间的比较。为了比较两种氨基酸序列,B12seq的选项如下设置:-i设置为含有要比较的第一氨基酸序列的文件(例如C:\seq1.txt);-j设置为含有要比较的第二氨基酸序列的文件(例如C:\seq2.txt);-p设置为blastp;-o设置为任何期望的文件名称(例如C:\output.txt);并且所有其他选项保留为其缺省设置。例如,可以使用以下命令来产生含有两种氨基酸序列见的比较的输出文件:C:\B12seq-ic:\seq1.txt-j c:\seq2.txt-p blastp-o c:\output.txt。若两个比较序列共享同源性,则指定的输出文件会呈现与比对序列的同源性的那些区域。若两个比对序列不共享同源性,则指定的输出文件不会呈现比对序列。核酸序列可以遵循相似的规程,只是使用blastn。

[0051] 在比对后,通过计算两个序列中呈现相同氨基酸残基的位置的数目测定匹配数目。通过用匹配的数目除以全长多肽氨基酸序列的长度,接着用所得的数值乘以100测定百分比同一性。

[0052] 注意到,百分比同一性数值四舍五入到最近的十分位。例如,78.11,78.12,78.13,和78.14向下四舍五入到78.1,而78.15,78.16,78.17,78.18,和78.19向上四舍五入到78.2。还注意到长度数值会总是整数。

[0053] 应当领会,许多核酸可以编码具有特定氨基酸序列的多肽。遗传密码的简并性是本领域公知的;即对于许多氨基酸,存在有超过一个充当氨基酸密码子的核苷酸三联体。例如,可以修饰给定GAA多肽的编码序列中的密码子,使得特定物种(例如细菌或真菌)中的最佳表达使用适合于所述物种的密码子偏爱表获得。

[0054] 在一个实施方案中,分子复合物可以包含至少两个多肽,其中一个多肽包含SEQ ID NO:1的氨基酸22至57,且另一个多肽包含SEQ ID NO:1的氨基酸66至896。

[0055] 在一个实施方案中,分子复合物可以包含至少三个多肽,其中一个多肽包含SEQ ID NO:1的氨基酸22至57,一个多肽包含SEQ ID NO:1的氨基酸66至726,且一个多肽包含SEQ ID NO:1的氨基酸736至896。

[0056] 在一个实施方案中,分子复合物可以包含至少四个多肽,其中一个多肽包含SEQ ID NO:1的氨基酸22至57,一个多肽包含SEQ ID NO:1的氨基酸66至143,一个多肽包含SEQ ID NO:1的氨基酸158至726,且一个多肽包含SEQ ID NO:1的氨基酸736至896。

[0057] 相对于SEQ ID NO:1中所列的序列,GAA的生物学活性变体可以含有添加,缺失,或取代。具有取代的GAA蛋白一般会具有不超过10个(例如不超过1、2、3、4、5、6、7、8、9、或10)个保守氨基酸取代。保守取代是用一个氨基酸取代具有相似特征的另一个。保守取代包括下组内的取代:缬氨酸,丙氨酸和甘氨酸;亮氨酸,缬氨酸和异亮氨酸;天冬氨酸和谷氨酸;天冬酰胺和谷氨酰胺;丝氨酸,半胱氨酸,和苏氨酸;赖氨酸和精氨酸;和苯丙氨酸和酪氨酸。非极性疏水性氨基酸包括丙氨酸、亮氨酸、异亮氨酸、缬氨酸、脯氨酸、苯丙氨酸、色氨酸和甲硫氨酸。极性中性氨基酸包括甘氨酸、丝氨酸、苏氨酸、半胱氨酸、酪氨酸、天冬酰胺和谷氨酰胺。带正电荷的(碱性)氨基酸包括精氨酸、赖氨酸和组氨酸。带负电荷的(酸性)氨基酸包括天冬氨酸和谷氨酸。将上文提及的极性、碱性或酸性组的一个成员用相同组的另一成员的任何取代可以视为保守取代。比较而言,非保守取代是用一个氨基酸取代具有不同特征的另一个。

[0058] 在一些实施方案中,GAA多肽可以是与异源氨基酸序列的融合蛋白,所述异源氨基酸序列诸如用于纯化重组蛋白的序列(例如FLAG,多组氨酸(例如六组氨酸),血凝素(hemagglutinin)(HA),谷胱甘肽-S-转移酶(GST),或麦芽糖结合蛋白(MBP))。

[0059] 在一些实施方案中,使用异源氨基酸序列增强将分子复合物转运到哺乳动物细胞中的效率。例如,可以将复合物中的至少一个多肽与胞外受体的配体,结合靶细胞表面上的受体的胞外域的靶向域,尿激酶型纤维蛋白溶酶原受体,或结合甘露糖-6-磷酸受体的人胰岛素样生长因子II的结构域(例如人胰岛素样生长因子的氨基酸1-67或1-87;至少氨基酸48-55;至少氨基酸8-28和41-61;或至少氨基酸8-87;人胰岛素样生长因子III的其序列变体(例如R68A)或人胰岛素样生长因子的截短形式(例如从第62位起的C端截短))融合。异源氨基酸序列可以在多肽的N端或C端融合。在一个实施方案中,通过间隔物(例如5-30个氨基酸或10-25个氨基酸)将肽标签与多肽的N或C端融合。参见例如美国专利No.7,785,856。

[0060] 异源氨基酸序列还可以是可用于诊断或可检测标志物的蛋白质,例如萤光素酶,绿色荧光蛋白(GFP),或氯霉素乙酰基转移酶(CAT)。

[0061] 在某些宿主细胞(例如酵母宿主细胞)中,可以通过使用异源信号序列提高靶蛋白的表达和/或分泌。在一些实施方案中,融合蛋白可以含有可用于例如引发免疫应答以生成抗体的载体(例如KLH)或内质网或高尔基体保留信号。异源序列可以具有不同的长度,并且在一些情况下可以是比与异源序列衔接的全长靶蛋白更长的序列。

[0062] 使糖蛋白脱甘露糖基化,或脱帽并脱甘露糖基化的方法

[0063] 可以使含有磷酸化的N-聚糖的糖蛋白脱甘露糖基化,并且可以使含有含甘露糖-1-磷酸-6-甘露糖连接或模块的磷酸化的N-聚糖的糖蛋白脱帽并脱甘露糖基化,其通过使糖蛋白与甘露糖苷酶接触进行,所述甘露糖苷酶能够(i)将甘露糖-1-磷酸-6-甘露糖连接或模块水解成甘露糖-6-磷酸及(ii)水解末端alpha-1,2甘露糖,alpha-1,3甘露糖和/或alpha-1,6甘露糖连接或模块。每个甘露糖苷酶的非限制性例子包括直生刀豆(刀豆(Jack bean))甘露糖苷酶和解脂耶氏酵母甘露糖苷酶(例如AMS1)。刀豆和AMS1甘露糖苷酶两者是家族38糖苷水解酶。

[0064] 刀豆甘露糖苷酶以硫酸铵悬浮液(产品目录编号M7257)和蛋白质组学级制剂(产品目录编号M5573)可购自例如Sigma-Aldrich(St.Louis,MO)。此类商业制剂可以例如通过凝胶过滤层析进一步纯化以除去污染物,诸如磷酸酶。刀豆甘露糖苷酶含有具有以下氨基酸序列NKIPRAGWQIDPFGHSAVQG(SEQ ID NO:11)的区段。参见Howard等,J.Biol.Chem.,273(4):2067-2072,1998。

[0065] 解脂耶氏酵母AMS1甘露糖苷酶可以是重组生成的。AMS1多肽的氨基酸序列在SEQ ID NO:5中列出(还可见图6)。

[0066] 在一些实施方案中,脱帽和脱甘露糖基化步骤通过两种不同的酶催化。例如,可以使用来自纤维化纤维微细菌的甘露糖苷酶(例如CcMan5)实施甘露糖-1-磷酸-6甘露糖连接或模块的脱帽。编码CcMan5多肽的核苷酸序列在SEQ ID NO:2中列出(参见图2A)。含有信号序列的CcMan5多肽的氨基酸序列在SEQ ID NO:3中列出(参见图2B)。没有信号序列的CcMan5多肽的氨基酸序列在SEQ ID NO:4中列出(参见图2C)。在一些实施方案中,使用CcMan5多肽的生物学活性片段。例如,生物学活性片段可以包含SEQ ID NO:4中所列的氨基酸序列的残基1-774。还可见WO 2011/039634。CcMan5甘露糖苷酶是家族92糖苷水解酶。

[0067] 可以使用来自斋藤曲霉 (As) (又称为海枣曲霉 (*Aspergillus phoenicis*)) 的甘露糖苷酶或来自纤维化纤维微细菌的甘露糖苷酶 (例如 CcMan4) 催化脱帽的糖蛋白或糖蛋白的分子复合物的脱甘露糖基化。斋藤曲霉甘露糖苷酶是家族 47 糖苷水解酶, 而 CcMan4 甘露糖苷酶是家族 92 糖苷水解酶。斋藤曲霉甘露糖苷酶的氨基酸序列在图 7 (SEQ ID NO:6) 和 GenBank 登录号 BAA08634 中列出。CcMan4 多肽的氨基酸序列在图 8 (SEQ ID NO:7) 中列出。

[0068] 还可以使用来自家族 38 糖苷水解酶的甘露糖苷酶, 诸如直生刀豆 (刀豆) 甘露糖苷酶或解脂耶氏酵母甘露糖苷酶 (例如 AMS1) 催化脱帽的糖蛋白或糖蛋白的分子复合物的脱甘露糖基化。例如, 可以使用 CcMan5 使糖蛋白 (或糖蛋白的分子复合物) 上的甘露糖-1-磷酸-6 甘露糖模块脱帽, 并且可以使用刀豆甘露糖苷酶使脱帽的糖蛋白 (或糖蛋白的分子复合物) 脱甘露糖基化。

[0069] 为了生成脱甘露糖基化的糖蛋白 (或糖蛋白的分子复合物), 或脱帽并脱甘露糖基化的糖蛋白 (或糖蛋白的分子复合物), 在合适的条件下使含有甘露糖-1-磷酸-6 甘露糖连接或模块的靶分子 (或分子复合物) 与合适的甘露糖苷酶和/或含有合适的重组生成的甘露糖苷酶的细胞裂解物接触。上文描述了合适的甘露糖苷酶。细胞裂解物可以来自任何遗传工程细胞, 包括真菌细胞, 植物细胞, 或动物细胞。动物细胞的非限制性例子包括线虫、昆虫、植物、禽、爬行动物、和哺乳动物诸如小鼠、大鼠、兔、仓鼠、沙鼠、犬、猫、山羊、猪、牛、马、鲸、猴、或人。

[0070] 在使靶分子 (例如分子复合物) 与纯化的甘露糖苷酶和/或细胞裂解物接触后, 可以将甘露糖-1-磷酸-6-甘露糖连接或模块水解成磷酸-6-甘露糖, 并可以水解此类含有磷酸根的聚糖的末端  $\alpha$ -1,2 甘露糖,  $\alpha$ -1,3 甘露糖和/或  $\alpha$ -1,6 甘露糖连接或模块以生成脱帽并脱甘露糖基化的靶分子。在一些实施方案中, 使用一种甘露糖苷酶, 其催化脱帽和脱甘露糖基化步骤两者。在一些实施方案中, 使用一种甘露糖苷酶催化脱帽步骤, 并使用不同甘露糖苷酶催化脱甘露糖基化步骤。在甘露糖苷酶处理后, 可以分离靶分子或分子复合物。

[0071] 保持裂解物中甘露糖苷酶活性的活性或完整性的适合于获得细胞裂解物的方法可以包括使用合适的缓冲液和/或抑制剂, 包括核酸酶、蛋白酶和磷酸酶抑制剂, 其保持细胞裂解物中的 N-糖基化活性或者使细胞裂解物中的 N-糖基化活性的变化最小化。此类抑制剂包括例如螯合剂诸如乙二胺四乙酸 (EDTA)、乙二醇二 (P-氨基基醚) N,N,N1,N1-四乙酸 (EGTA)、蛋白酶抑制剂诸如苯甲磺酰氟 (PMSF)、抑肽酶、亮抑酶肽 (leupeptin)、抗蛋白酶 (antipain) 等等, 及磷酸酶抑制剂诸如磷酸盐、氟化钠、钒酸盐, 等等。适合于获得包含酶促活性的裂解物的缓冲液和条件记载于例如 Ausubel 等 Current Protocols in Molecular Biology (Supplement 47), John Wiley & Sons, New York (1999); Harlow and Lane, Antibodies: A Laboratory Manual Cold Spring Harbor Laboratory Press (1988); Harlow and Lane, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Press (1999); Tietz Textbook of Clinical Chemistry, 3rd ed. Burtis and Ashwood, eds. W.B. Saunders, Philadelphia, (1999)。

[0072] 在适当时, 可以将细胞裂解物进一步加工以消除干扰性物质或使干扰性物质的存在最小化。若想要的话, 可以通过本领域技术人员公知的多种方法将细胞裂解物分级, 所述方法包括亚细胞分级、和层析技术诸如离子交换、疏水性和反相、大小排阻、亲和、疏水性电

荷诱导层析,等等。

[0073] 在一些实施方案中,可以制备细胞裂解物,其中整个细胞细胞器保持完整和/或功能性。例如,裂解物可以含有下列一种或多种:完整的糙面内质网、完整的光面内质网、或完整的高尔基体。适合于制备含有完整细胞细胞器的裂解物及测试细胞器功能性的方法记载于例如Moreau等(1991) *J. Biol. Chem.* 266 (7) :4329-4333; Moreau等(1991) *J. Biol. Chem.* 266 (7) :4322-4328; Rexach等(1991) *J. Cell Biol.* 114 (2) :219-229; and Paulik等(1999) *Arch. Biochem. Biophys.* 367 (2) :265-273。

[0074] 在哺乳动物细胞与含有脱帽且脱甘露糖基化的磷酸化N-聚糖的分子复合物接触后,分子复合物可以被转运到哺乳动物细胞(例如人细胞)内部。具有脱帽的,但未脱甘露糖基化的磷酸化N-聚糖的分子复合物实质上不结合哺乳动物细胞上的甘露糖-6-磷酸受体,并且因此没有被有效转运到细胞的内部。如本文中使用的,“实质上不结合”意指小于15%(例如小于14%,12%,10%,8%,6%,4%,2%,1%,0.5%,或更小,或0%)的糖蛋白分子结合哺乳动物细胞上的甘露糖-6-磷酸受体。然而,若使此类分子复合物与能够在下层甘露糖被磷酸化时水解末端 $\alpha$ -1,2甘露糖连接或模块的甘露糖苷酶接触,生成脱甘露糖基化的糖蛋白,其实质上结合哺乳动物细胞上的甘露糖-6-磷酸受体,并且被有效转运到细胞内部。如本文中使用的,“实质上结合”意指15%或更多(例如大于16%,18%,20%,25%,30%,35%,40%,45%,50%,55%,60%,65%,70%,75%,80%,85%,90%或95%)的分子复合物结合哺乳动物细胞上的甘露糖-6-磷酸受体。应当理解,含有使磷酸化的N-聚糖脱帽但不脱甘露糖基化的酶的制备物(例如重组宿主细胞或无细胞制备物)可以被使磷酸化N-聚糖脱甘露糖基化的酶污染。与此类制备物接触后的靶蛋白样品可以含有具有仅脱帽的一些磷酸化N-聚糖和脱帽并脱甘露糖基化的其它磷酸化N-聚糖的蛋白质分子。自然地,含有脱帽并脱甘露糖基化的磷酸化N-聚糖的那些蛋白质分子实际上可以结合甘露糖-6-磷酸受体。“实际上不结合”的上述定义不适用于此类靶蛋白样品,因为蛋白质分子上的磷酸化N-聚糖不能表征为脱帽但未脱甘露糖基化的。

[0075] 如此,本文件提供了将分子复合物从不结合哺乳动物细胞上的甘露糖-6-磷酸受体的第一种形式转化为确实结合哺乳动物细胞上的甘露糖-6-磷酸受体的第二种形式的方法。如下的分子复合物为第一种形式,其中复合物中的至少一个多肽包含含有一个或多个甘露糖残基的一个或多个N-聚糖,所述甘露糖残基在1位与在6位含有磷酸根残基的甘露糖残基连接。在此类方法中,使分子复合物的第一种形式与使末端甘露糖残基脱甘露糖基化的甘露糖苷酶接触,从而导致在6位含有磷酸根的甘露糖变为末端甘露糖。在一些实施方案中,甘露糖苷酶具有脱帽和脱甘露糖基化活性两者(例如直生刀豆(Jack bean)或解脂耶氏酵母AMS1甘露糖苷酶)。在一些实施方案中,甘露糖苷酶没有脱帽活性(例如来自斋藤曲霉的甘露糖苷酶或来自纤维化纤维微细菌的甘露糖苷酶(例如CcMan4))。

[0076] 可以使用细胞摄取测定法评估糖蛋白或分子复合物对细胞内部的转运。例如,可以使哺乳动物细胞和含有脱帽且脱甘露糖基化的磷酸化N-聚糖的分子复合物温育,然后将细胞清洗并裂解。细胞裂解物可以评估GAA复合物的存在(例如通过Western印迹)或通过细胞裂解物中的GAA活性评估。例如,可以在酸性 $\alpha$ 葡糖苷酶活性缺陷的成纤维细胞中评估摄取。可以使用4-甲基伞形酮基- $\alpha$ -D-葡糖吡喃糖苷(4-methylumbelliferyl- $\alpha$ -D-glucopyranoside, 4-MUG)测定法评估 $\alpha$ 葡糖苷酶的胞内活性。葡糖苷酶对底

物4-MUG的切割导致产生荧光生成性产物4-MU,其可以通过用UV光照射显现或检测。

[0077] 多肽的重组生成

[0078] 可以通过标准技术生成编码多肽(例如甘露糖苷酶、碱性蛋白酶、或GAA或其片段)的分离的核酸分子。术语“核酸”和“多核苷酸”在本文中可互换使用,指RNA和DNA两者,包括cDNA、基因组DNA、合成的DNA、和含有核酸类似物的DNA(或RNA)。多核苷酸可以具有任何三维结构。核酸可以是双链或单链(即,有义链或反义链)。多核苷酸的非限制性例子包括基因、基因片段、外显子、内含子、信使RNA(mRNA)、转移RNA、核糖体RNA、siRNA、微小RNA、核酶、cDNA、重组多核苷酸、分支多核苷酸、质粒、载体、任何序列的分离的DNA、任何序列的分离的RNA、核酸探针、和引物,以及核酸类似物。

[0079] “分离的核酸”指与存在于天然存在的基因组的其它核酸分子分开的核酸,包括通常在天然存在的基因组(例如,酵母基因组)中的核酸的一侧或两侧侧翼的核酸。如本文中所使用的,术语“分离的”就核酸而言还包括任何非天然存在的核酸序列,因为此类非天然存在的序列不存在于自然界,并且在天然存在的基因组中没有刚好连续的序列。

[0080] 分离的核酸可以是例如,DNA分子,只要通常刚刚在天然存在的基因组中的DNA分子侧翼找到的核酸序列之一被除去或缺乏。如此,分离的核酸包括但不限于作为不依赖于其它序列的不同分子(例如,化学合成的核酸、或通过PCR或限制性内切核酸酶处理生成的cDNA或基因组DNA片段)存在的DNA分子以及掺入载体、自主复制型质粒、病毒(例如,任何副粘液病毒、逆转录病毒、慢病毒、腺病毒、或疱疹病毒)中,或掺入原核生物或真核生物的基因组DNA中的DNA。另外,分离的核酸可以包括经工程化改造的核酸诸如作为杂合或融合核酸的一部分的DNA分子。认为存在于例如cDNA文库或基因组文库,或含有基因组DNA限制性消化物的凝胶切片内的数百至数百万个其它核酸间的核酸不是分离的核酸。

[0081] 如本文中所使用的,术语“外源的”就核酸和特定宿主细胞而言意指不存在于(并且不能获自)如自然界中找到的所述特定细胞的任何核酸。如此,认为一旦导入宿主细胞中,非天然存在的核酸对于宿主细胞而言是外源的。重要的是,注意到非天然存在的核酸可以含有自然界中找到的核酸亚序列或核酸序列片段,只要整个核酸不存在于自然界中。例如,在表达载体内含有基因组DNA序列的核酸分子是非天然存在的核酸,并且如此一旦导入宿主细胞中,对于宿主细胞而言是外源的,因为整个核酸分子(基因组DNA加载体DNA)不存在于自然界中。如此,认为作为整体不存在于自然界中的任何载体、自主复制型质粒、或病毒(例如,逆转录病毒、腺病毒、或疱疹病毒)是非天然存在的核酸。由此可见,认为通过PCR或限制性内切核酸酶处理产生的基因组DNA片段及cDNA是非天然存在的核酸,因为它们以自然界中找不到的分开的分子存在。由此还可见,以自然界中找不到的排列含有启动子序列和多肽编码序列(例如,cDNA或基因组DNA)的任何核酸是非天然存在的核酸。天然存在的核酸可以是特定细胞外源的。例如,一旦自酵母x细胞分离的整个染色体被导入酵母细胞中,所述染色体对于酵母y细胞而言是外源核酸。

[0082] 可以使用聚合酶链式反应(PCR)技术来获得含有本文中所描述的核苷酸序列的分离的核酸。可以使用PCR从DNA及RNA(包含来自总基因组DNA或总细胞RNA的序列)扩增特定序列。一般地,采用来自感兴趣区域末端或之外的序列信息来设计与要扩增的模板的相反链在序列上相同或相似的寡核苷酸引物。多种PCR策略也是可用的,通过所述PCR策略可以将位点特异性核苷酸序列修饰引入模板核酸中。分离的核酸也可以作为单一核酸分子(例

如使用亚磷酰胺 (phosphoramidite) 技术以3' 至5' 方向使用自动化DNA合成) 或者作为一系列寡核苷酸化学合成。例如, 可以合成含有期望序列的一对或多对长的寡核苷酸 (例如, 大于100个核苷酸), 其中每对含有短的互补性区段 (例如, 约15个核苷酸), 使得在寡核苷酸对退火时形成双链体。使用DNA聚合酶来延长寡核苷酸, 每个寡核苷酸对生成单一的、双链核酸分子, 然后, 可以将其连接到载体中。也可以通过例如对天然存在的DNA的诱变来获得分离的核酸。

[0083] 为了重组生成多肽 (例如甘露糖苷酶, 碱性蛋白酶, 或GAA或其片段), 使用含有与编码多肽的核酸可操作连接的启动子的载体。如本文中所使用的, “启动子” 指使基因能够得到转录的DNA序列。RNA聚合酶识别启动子, 然后其启动转录。如此, 启动子含有由RNA聚合酶直接结合的, 或者牵涉RNA聚合酶募集的DNA序列。启动子序列还可以包含“增强子区”, 其是一种或多种可以与蛋白质 (即, 反式作用因子, 很像一组转录因子) 结合以增强基因簇中基因的转录水平 (因此得名) 的DNA区。虽然通常在编码区的5' 端, 增强子也可以与启动子序列分开, 并且可以例如在基因的内含子区内或在基因编码区的3' 。

[0084] 如本文中所使用的, “可操作连接” 意指掺入遗传构建体 (例如, 载体) 中, 使得表达控制序列有效控制感兴趣的编码序列表达。

[0085] 可以将表达载体导入宿主细胞中 (例如, 通过转化或转染来进行) 以表达编码的多肽, 然后, 可以将所述编码的多肽纯化。可以用于多肽 (例如甘露糖苷酶, 碱性蛋白酶, 或GAA或其片段) 的小规模或大规模生成的表达系统包括但不限于微生物体, 诸如用含有核酸分子的重组噬菌体DNA、质粒DNA、或粘粒DNA表达载体转化的细菌 (例如, 大肠杆菌 (*E. coli*)), 和用含有核酸分子的重组真菌表达载体转化的真菌 (例如解脂耶氏酵母、*Arxula adenivorans*、巴斯德毕赤酵母 (*Pichia pastoris*)、多形汉森酵母 (*Hansenula polymorpha*)、*Ogataea minuta*、甲醇毕赤酵母 (*Pichia methanolica*)、黑曲霉 (*Aspergillus niger*)、里氏木霉 (*Trichoderma reesei*)、和酿酒酵母 (*Saccharomyces cerevisiae*))。有用的表达系统还包括用含有核酸分子的重组病毒表达载体 (例如杆状病毒) 感染的昆虫细胞系统、和用重组病毒表达载体 (例如, 烟草花叶病毒) 感染或用含有核酸分子的重组质粒表达载体 (例如Ti质粒) 转化的植物细胞系统。也可以使用哺乳动物表达系统来生成甘露糖苷酶或碱性蛋白酶多肽, 所述哺乳动物表达系统包括含有重组表达构建体的细胞 (例如, 永生化细胞系诸如COS细胞、中国仓鼠卵巢细胞、HeLa细胞、人胚肾293细胞、和3T3L1细胞), 所述重组表达构建体含有源自哺乳动物细胞基因组 (例如, 金属硫蛋白启动子) 或哺乳动物病毒 (例如腺病毒晚期启动子和巨细胞病毒启动子) 的启动子。

[0086] 通常, 可以用异源氨基酸序列诸如FLAG、多组氨酸 (例如, 六组氨酸)、血凝素 (HA)、谷胱甘肽-S-转移酶 (GST) 或麦芽糖结合蛋白 (MBP) 给重组多肽诸如甘露糖苷酶加标签以帮助纯化蛋白质。用于纯化蛋白质的其它方法包括层析技术诸如离子交换、疏水性和反相、大小排阻、亲和、疏水性电荷诱导层析, 等等 (参见例如Scopes, *Protein Purification: Principles and Practice*, third edition, Springer-Verlag, New York (1993); Burton and Harding, *J. Chromatogr. A* 814:71-81 (1998))。

[0087] 使糖蛋白脱帽和脱甘露糖基化的体内方法

[0088] 可以使用本文中描述的遗传工程化细胞生成具有GAA活性的分子复合物。例如, 可以使用遗传工程化细胞生成分子复合物, 该分子复合物具有GAA活性并且包含至少两个多

肽,每个多肽与SEQ ID NO:1中所列的氨基酸序列的区段具有至少85%序列同一性,每个区段通过在氨基酸50和氨基酸74之间(例如氨基酸56和氨基酸68之间或氨基酸60和氨基酸65之间)的一个或多个位点处对SEQ ID NO:1中所列的氨基酸序列的蛋白水解衍生。例如,真菌细胞可以工程化改造为包含编码SEQ ID NO:1中所列的氨基酸序列的核酸和编码与SEQ ID NO:8中所列的氨基酸序列具有至少85%序列同一性的碱性蛋白酶的核酸,使得每个编码的多肽被分泌到培养基中,其中碱性蛋白酶可以切割SEQ ID NO:1中所列的氨基酸序列。如实施例12中描述的,在重组GAA被分泌到具有碱性蛋白酶的培养基中时,完成将110kDa前体加工成95kDa形式,即没有检出110kDa前体。

[0089] 还可以使用本文中描述的遗传工程化细胞生成脱帽且脱甘露糖基化的分子复合物。此类遗传工程化细胞可以包含编码具有SEQ ID NO:1中所列的氨基酸序列的多肽的核酸,编码如本文中描述的甘露糖苷酶的核酸,和任选编码碱性蛋白酶的核酸,所述碱性蛋白酶与SEQ ID NO:8中所列的氨基酸序列具有至少85%序列同一性。

[0090] 生成脱帽且脱甘露糖基化的分子复合物的基于细胞的方法可以包括导入真菌细胞中,所述真菌细胞经遗传工程化改造以包含编码能够将甘露糖-1-磷酸-6-甘露糖连接或模块水解成磷酸-6-甘露糖的甘露糖苷酶的核酸,编码具有SEQ ID NO:1中所列的氨基酸序列的多肽的核酸和任选编码与SEQ ID NO:8中所列的氨基酸序列具有至少85%序列同一性的碱性蛋白酶的核酸,其中所述细胞生成包含脱帽的磷酸化N-聚糖的本文中描述的分子复合物。如上文描述的,可以使此类磷酸化N-聚糖脱甘露糖基化。在一些实施方案中,编码甘露糖苷酶和GAA的核酸含有分泌序列,使得甘露糖苷酶和GAA被共分泌。在包含编码碱性蛋白酶的核酸的遗传工程化细胞中,分子复合物可以被加工成95kDa形式。

[0091] 另一种基于细胞的方法可以包括导入真菌细胞中的步骤,所述真菌细胞经遗传工程化改造以包含编码甘露糖苷酶的核酸,编码具有SEQ ID NO:1中所列的氨基酸序列的多肽的核酸,和任选编码与SEQ ID NO:8中所列的氨基酸序列具有至少85%序列同一性的碱性蛋白酶的核酸,其中所述细胞生成脱帽的且脱甘露糖基化的分子复合物,所述甘露糖苷酶能够(i)将甘露糖-1-磷酸-6-甘露糖连接或模块水解成磷酸-6-甘露糖,并(ii)水解含有磷酸根的聚糖的末端 $\alpha$ -1,2甘露糖, $\alpha$ -1,3甘露糖和/或 $\alpha$ -1,6甘露糖连接或模块。在一些实施方案中,编码甘露糖苷酶和GAA的核酸含有分泌序列,使得甘露糖苷酶和靶分子被共分泌。在包含编码碱性蛋白酶的核酸的遗传工程化细胞中,分子复合物可以被加工成95kDa形式。

[0092] 适合于靶分子或分子复合物的体内生成的细胞可以是真菌起源的,包括解脂耶氏酵母、*Arxula adeninivorans*、甲基营养酵母(methylotrophic yeast)(诸如念珠菌属(*Candida*)、汉森酵母属(*Hansenula*)、*Oogataea*、毕赤酵母属(*Pichia*)或球拟酵母属(*Torulopsis*)的甲基营养酵母)或曲霉属、木霉属(*Trichoderma*)、脉孢菌属(*Neurospora*)、镰孢菌属(*Fusarium*)、或金孢子菌属(*Chrysosporium*)的丝状真菌。例示性真菌物种包括但不限于异常毕赤酵母(*Pichia anomala*)、异常毕赤酵母(*Pichia anomala*)、牛毕赤酵母(*Pichia bovis*)、加拿大毕赤酵母(*Pichia canadensis*)、卡氏毕赤酵母(*Pichia carsonii*)、粉状毕赤酵母(*Pichia farinose*)、发酵毕赤酵母(*Pichia fermentans*)、*Pichia fluxuum*、膜醭毕赤酵母(*Pichia membranaefaciens*)、膜醭毕赤酵母(*Pichia membranaefaciens*)、粗状假丝酵母(*Candida valida*)、白念珠菌(*Candida albicans*)、

*Candida ascalaphidarum*、*Candida amphixiae*、南极假丝酵母 (*Candida Antarctica*)、大西洋假丝酵母 (*Candida atlantica*)、*Candida atmosphaerica*、蟑螂假丝酵母 (*Candida blattae*)、果生假丝酵母 (*Candida carpophila*)、*Candida cerambycidarum*、*Candida chauliodes*、延胡索假丝酵母 (*Candida corydalis*)、*Candida dosseyi*、杜氏假丝酵母 (*Candida dubliniensis*)、*Candida ergatensis*、*Candida fructus*、光滑假丝酵母 (*Candida glabrata*)、发酵假丝酵母 (*Candida fermentati*)、吉利蒙假丝酵母 (*Candida guilliermondii*)、黑马朗假丝酵母 (*Candida haemulonii*)、*Candida insectamens*、*Candida insectorum*、中型假丝酵母 (*Candida intermedia*)、*Candida jeffresii*、乳酒假丝酵母 (*Candida kefir*)、克鲁斯氏念珠菌 (*Candida krusei*)、葡萄牙假丝酵母 (*Candida lusitaniae*)、*Candida lyxosophila*、麦芽糖假丝酵母 (*Candida maltosa*)、膜醭假丝酵母 (*Candida membranifaciens*)、梅林假丝酵母 (*Candida milleri*)、橄榄假丝酵母 (*Candida oleophila*)、俄勒冈假丝酵母 (*Candida oregonensis*)、近平滑念珠菌 (*Candida parapsilosis*)、桔假丝酵母 (*Candida quercitrusa*)、休哈塔假丝酵母 (*Candida shehatei*)、*Candida temnochilae*、纤维假丝酵母 (*Candida tenuis*)、热带假丝酵母 (*Candida tropicalis*)、*Candida tsuchiyaе*、*Candida sinolaborantiu*、大豆假丝酵母 (*Candida sojae*)、维斯假丝酵母 (*Candida viswanathii*)、产朊假丝酵母 (*Candida utilis*)、*Oogataea minuta*、膜醭毕赤酵母 (*Pichia membranaefaciens*)、*Pichia silvestris*、膜醭毕赤酵母、*Pichia chodati*、膜醭毕赤酵母、膜醭毕赤酵母、*Pichia minuscule*、巴斯德毕赤酵母 (*Pichia pastoris*)、*Pichia pseudopolymorpha*、*Pichia quercuum*、*Pichia robertsii*、*Pichia saitoi*、*Pichia silvestrisi*、*Pichia strasburgensis*、陆生毕赤酵母 (*Pichia terricola*)、*Pichia vanriji*、*Pseudozyma Antarctica*、圆红冬孢酵母菌 (*Rhodosporidium toruloides*)、粘红酵母 (*Rhodotorula glutinis*)、贝酵母 (*Saccharomyces bayanus*)、贝酵母 (*Saccharomyces bayanus*)、*Saccharomyces momdshuricus*、葡萄汁酵母 (*Saccharomyces uvarum*)、贝酵母、酿酒酵母 (*Saccharomyces cerevisiae*)、二孢酵母 (*Saccharomyces bisporus*)、谢瓦利埃酵母 (*Saccharomyces chevalieri*)、德尔布酵母 (*Saccharomyces delbrueckii*)、少孢酵母 (*Saccharomyces exiguous*)、发酵性酵母 (*Saccharomyces fermentati*)、脆壁酵母 (*Saccharomyces fragilis*)、*Saccharomyces marxianus*、蜂蜜酵母 (*Saccharomyces mellis*)、罗斯酵母 (*Saccharomyces rosei*)、鲁酵母 (*Saccharomyces rouxii*)、葡萄汁酵母 (*Saccharomyces uvarum*)、魏氏酵母 (*Saccharomyces willianus*)、路氏类酵母 (*Saccharomycodes ludwigii*)、荚复膜孢酵母 (*Saccharomycopsis capsularis*)、扣囊复膜酵母 (*Saccharomycopsis fibuligera*)、扣囊复膜酵母、*Endomyces hordei*、*Endomycopsis fobuligera*、*Saturnispora saitoi*、八孢裂殖酵母 (*Schizosaccharomyces octosporus*)、粟酒裂殖酵母 (*Schizosaccharomyces pombe*)、西方许旺酵母 (*Schwanniomyces occidentalis*)、德尔布有孢圆酵母 (*Torulaspora delbrueckii*)、德尔布有孢圆酵母、*Saccharomyces dairensis*、德尔布有孢圆酵母、*Torulaspora fermentati*、发酵性酵母 (*Saccharomyces fermentati*)、德尔布有孢圆酵母、罗斯有孢圆酵母 (*Torulaspora rosei*)、罗斯有孢圆酵母、德尔布有孢圆酵母、罗斯酵母、德尔布有孢圆酵母、德尔布酵母、德尔布有孢圆酵母、德尔布酵母、*Zygosaccharomyces mongolicus*、*Dorulaspora globosa*、

球形德巴利酵母 (*Debaryomyces globosus*)、球拟圆酵母 (*Torulopsis globosa*)、皮状丝孢酵母 (*Trichosporon cutaneum*)、三角酵母 (*Trigonopsis variabilis*)、*Williopsis californica*、土星拟威尔酵母 (*Williopsis saturnus*)、二孢接合酵母 (*Zygosaccharomyces bisporus*)、二孢接合酵母、*Debaryomyces dispersa*、二孢酵母 (*Saccharomyces bisporas*)、二孢接合酵母、二孢酵母、*Zygosaccharomyces mellis*、*Zygosaccharomyces priorianus*、*Zygosaccharomyces rouxiim*、鲁氏接合酵母 (*Zygosaccharomyces rouxii*)、巴格式结合酵母 (*Zygosaccharomyces barkeri*)、鲁酵母、鲁氏接合酵母、酱醪接合酵母 (*Zygosaccharomyces majori*)、*Saccharomyces rousii*、异常毕赤酵母 (*Pichia anomala*)、牛毕赤酵母 (*Pichia bovis*)、加拿大毕赤酵母 (*Pichia Canadensis*)、卡氏毕赤酵母 (*Pichia carsonii*)、粉末毕赤酵母 (*Pichia farinose*)、发酵毕赤酵母 (*Pichia fermentans*)、*Pichia fluxuum*、膜醭毕赤酵母 (*Pichia membranaefaciens*)、*Pichia pseudopolymorpha*、*Pichia quercuum*、*Pichia robertsii*、*Pseudozyma Antarctica*、圆红冬孢酵母菌 (*Rhodospiridium toruloides*)、圆红冬孢酵母菌、胶红类酵母菌 (*Rhodotorula glutinis*)、贝酵母 (*Saccharomyces bayanus*)、贝酵母、二孢酵母、酿酒酵母、谢瓦利埃酵母、德尔布酵母、发酵性酵母、脆壁酵母、路氏类酵母 (*Saccharomycodes ludwigii*)、粟酒裂殖酵母、西方许旺酵母、德尔布有孢圆酵母、球有孢圆酵母 (*Torulaspora globosa*)、变异三角酵母 (*Trigonopsis variabilis*)、*Williopsis californica*、土星拟威尔酵母、二孢接合酵母、*Zygosaccharomyces mellis*、鲁氏接合酵母。例示性的丝状真菌包括曲霉属 (*Aspergillus*) 的多种物种,包括但不限于浅蓝灰曲霉 (*Aspergillus caesiellus*)、亮白曲霉 (*Aspergillus candidus*)、肉色曲霉 (*Aspergillus carneus*)、棒曲霉 (*Aspergillus clavatus*)、弯头曲霉 (*Aspergillus deflectus*)、黄曲霉 (*Aspergillus flavus*)、烟曲霉 (*Aspergillus fumigatus*)、灰绿曲霉 (*Aspergillus glaucus*)、构巢曲霉 (*Aspergillus nidulans*)、黑曲霉、赭曲霉 (*Aspergillus ochraceus*)、米曲霉 (*Aspergillus oryzae*)、寄生曲霉 (*Aspergillus parasiticus*)、帚状曲霉 (*Aspergillus penicilloides*)、局限曲霉 (*Aspergillus restrictus*)、酱油曲霉 (*Aspergillus sojae*)、聚多曲霉 (*Aspergillus sydowi*)、溜曲霉 (*Aspergillus tamari*)、土曲霉 (*Aspergillus terreus*)、焦曲霉 (*Aspergillus ustus*)、或花斑曲霉 (*Aspergillus versicolor*)。在如本文中规定的遗传工程前,此类细胞可以获自多种商业来源和研究资源机构,诸如例如,美国典型培养物保藏中心 (Rockville, MD)。

[0093] 在编码甘露糖苷酶、GAA、和/或碱性蛋白酶的外源核酸外,细胞的遗传工程可以包括一处或多处遗传修饰,诸如:(i) 编码外部链延长 (Outer Chain elongation, OCH1) 蛋白的内源基因的缺失;(ii) 引入编码能够促进甘露糖基磷酸化的多肽 (例如来自解脂耶氏酵母、酿酒酵母、*Ogataea minuta*、巴斯德毕赤酵母、或白念珠菌的MNN4多肽,或来自巴斯德毕赤酵母的PN01多肽) 的重组核酸以提高甘露糖残基的磷酸化;(iii) 干扰 OCH1 蛋白的功能性表达的 RNA 分子的引入或表达;(iv) 引入编码具有 N-糖基化活性的野生型 (例如内源或外源) 蛋白质的重组核酸 (即表达具有 N-糖基化活性的蛋白质);或 (v) 改变编码具有 N-糖基化活性的蛋白质的一个或多个内源基因的启动子或增强子元件,如此改变其编码蛋白质的表达。RNA 分子包括例如小干扰 RNA (siRNA), 短发夹 RNA (shRNA), 反义 RNA, 或微小 RNA (miRNA)。遗传工程化还包括改变编码具有 N-糖基化活性的蛋白质的内源基因以生成具有添加 (例如

异源序列)、缺失、或取代(例如突变,诸如点突变,保守或非保守突变)的蛋白质。突变可以特异性引入(例如通过定点诱变或同源重组)或可以随机引入(例如,细胞可以化学诱变,如记载于例如Newman and Ferro-Novick(1987) *J. Cell Biol.* 105 (4) :1587的)。

[0094] 本文中描述的遗传修饰可以导致下列一项或多项:(i) 经遗传修饰的细胞中的一种或多种活性的升高,(ii) 经遗传修饰的细胞中的一种或多种活性的降低,或(iii) 经遗传修饰的细胞中的一种或多种活性的定位或胞内分布的变化。应当理解,特定活性(例如促进甘露糖基磷酸化)的量增加可以由于过表达一种或多种能够促进甘露糖基磷酸化的蛋白质、内源基因的拷贝数目增加(例如,基因复制)、或刺激由基因编码的蛋白质表达升高的内源基因的启动子或增强子的变化所致。一种或多种特定活性的降低可以是由于突变体形式(例如,显性负性形式)的过表达、导入或表达降低具有特定活性的一种或多种蛋白质的表达的一种或多种干扰RNA分子,或删除编码具有特定活性的蛋白质的一种或多种内源基因所致。

[0095] 为了通过同源重组破坏基因,“基因替换”载体可以以如下的方式构建以包含选择标志物基因。选择标志物基因可以在5' 和3' 端两者与足以介导同源重组的长度的基因的部分可操作连接。选择标志物可以是互补宿主细胞营养缺陷型或提供抗生素抗性的众多基因之一,包括URA3、LEU2和HIS3基因。其它合适的选择标志物包括CAT基因,其对酵母细胞赋予氯霉素抗性,或lacZ基因,其由于表达 $\beta$ -半乳糖苷酶而导致蓝色菌落。然后,使用本领域中公知的方法(见下文)将基因替换载体的线性化DNA片段导入细胞中。线性片段对基因组的整合和基因的破坏可以基于选择标志物测定,并且可以例如通过Southern印迹分析确认。可以通过例如Cre-loxP系统(参见下文)从宿主细胞的基因组中除去选择标志物。

[0096] 或者,基因替换载体可以以如下的方式构建,使得包含要破坏的基因部分,该部分缺乏任何内源基因启动子序列并编码无一基因编码序列或基因编码序列的无活性片段。“无活性片段”指编码具有从基因的全长编码序列生成的蛋白质的活性的例如小于约10%(例如小于约9%,小于约8%,小于约7%,小于约6%,小于约5%,小于约4%,小于约3%,小于约2%,小于约1%,或0%)的蛋白质的基因片段。将此类基因的部分以如下的方式在载体中插入,使得无已知启动子序列与基因序列可操作连接,但是终止密码子和转录终止序列与基因序列的部分可操作连接。随后,可以将此载体在基因序列的该部分中线性化,并转化到细胞中。通过单同源重组,此线性化载体然后被整合在基因的内源对应物中。

[0097] 表达载体可以是自主的或整合的。可以将重组核酸(例如编码甘露糖苷酶,GAA,或碱性蛋白酶的重组核酸)以表达载体形式诸如质粒、噬菌体、转座子、粘粒或病毒颗粒导入细胞中。可以将重组核酸在染色体外维持或者可以将其整合入酵母细胞染色体DNA中。表达载体可以含有编码在选定的条件下对于细胞存活力需要的蛋白质的选择标志物基因(例如,URA3,其编码尿嘧啶生物合成必需的酶,或TRP1,其编码色氨酸生物合成需要的酶)以容许检测和/或选择用期望的核酸转化的那些细胞(见例如美国专利No. 4,704,362)。表达载体还可以包含自主复制序列(ARS)。例如,美国专利No. 4,837,148描述了提供适合于维持巴斯德毕赤酵母中的质粒的手段自主复制序列。

[0098] 整合性载体披露于例如美国专利No. 4,882,279。整合性载体一般包含至少第一可插入DNA片段、选择标志基因、和第二可插入DNA片段的连续排列序列。第一和第二可插入DNA片段各为约200(例如,约250、约300、约350、约400、约450、约500、或约1000或更多)个核

苷酸的长度,并且具有与要转化的物种的基因组DNA的一部分同源的核苷酸序列。含有表达的感兴趣基因(例如,编码GAA的基因)的核苷酸序列在此载体中在第一和第二可插入DNA片段间插入,无论在标志物基因之前还是之后。可以在酵母转化前将整合性载体线性化以便于感兴趣的核苷酸序列对宿主细胞基因组的整合。

[0099] 表达载体的特征可以是在酵母(例如,解脂耶氏酵母、*Arxula adenivorans*、巴斯德毕赤酵母、或其它合适的真菌物种)启动子(使它们能够在真菌细胞中表达)的控制下的重组核酸。合适的酵母启动子包括例如ADC1、TPI1、ADH2、hp4d、POX、和Gal10(参见例如Guarente等(1982)Proc.Natl.Acad.Sci.USA 79(23):7410)启动子。其它合适的启动子记载于例如Zhu and Zhang(1999)Bioinformatics 15(7-8):608-611及美国专利No.6,265,185。

[0100] 启动子可以是组成性或诱导型(条件性的)。组成性启动子理解为其表达在标准的培养条件下恒定的启动子。诱导型启动子是响应一种或多种诱导信号的启动子。例如,可以将诱导型启动子化学调节(例如,其转录活性受到化学诱导剂诸如醇、四环素、类固醇、金属、或其它小分子的存在或缺乏调节的启动子)或物理调节(例如,其转录活性受到物理诱导物诸如光或高或低温的存在或缺乏调节的启动子)。也可以通过自身受到化学或物理信号直接调节的一种或多种转录因子间接调节诱导型启动子。

[0101] 应当理解,其它遗传工程修饰也可以是条件性的。例如,可以使用例如,位点特异性DNA重组酶诸如Cre-loxP系统(见例如Gossen等(2002)Ann.Rev.Genetics 36:153-173及美国申请公开文本No.20060014264)来条件性删除基因。

[0102] 可以使用多种方法诸如原生质球技术或全细胞氯化锂酵母转化方法来将重组核酸导入本文中所述的细胞中。其它可用于将质粒或线性核酸载体转化入细胞中的方法记载于例如,美国专利No.4,929,555;Hinnen等(1978)Proc.Nat.Acad.Sci.USA 75:1929;Ito等(1983)J.Bacteriol.153:163;美国专利No.4,879,231;及Sreekrishna等(1987)Gene 59:115,每篇的公开内容通过提及完整并入本文。也可以使用电穿孔和PEG1000全细胞转化规程,如由Cregg and Russel,Methods in Molecular Biology:Pichia Protocols, Chapter 3, Humana Press, Totowa, N.J., pp.27-39(1998)所描述的。

[0103] 可以通过使用合适的技术,包括但不限于在转化后在缺乏需要的生物化学产物(由于细胞的营养缺陷型所致)的情况下培养营养缺陷型细胞,选择并检测新表型,或在存在抗生素的情况下培养来选择经转化的真菌细胞,所述抗生素在缺乏转化体中含有的抗性基因的情况下对于酵母是毒性。也可以通过将表达盒整合入基因组中来选择和/或确认转化体,其可以通过例如Southern印迹或PCR分析来评估。

[0104] 在将载体导入感兴趣的靶细胞中前,可以将载体在细菌细胞诸如大肠杆菌(*Escherichia coli*,*E.coli*)中增加(例如,扩增),如上文描述的。可以通过导致载体DNA自细菌环境(milieu)纯化的本领域中已知的任何方法来自细菌细胞分离载体DNA。可以用酚、氯仿、和乙醚来广泛提取纯化的载体DNA,以确保质粒DNA制备物中不存在大肠杆菌蛋白质,因为这些蛋白质对于哺乳动物细胞可以是毒性的。

[0105] 在一些实施方案中,遗传工程化真菌细胞缺乏OCH1基因或其基因产物(例如mRNA或蛋白质),并且是OCH1活性缺陷的。在一些实施方案中,遗传工程化细胞表达能够促进甘露糖基磷酸化的多肽(例如来自解脂耶氏酵母、酿酒酵母、*Ogataea minuta*、巴斯德毕赤酵

母、或白念珠菌的MNN4多肽,或来自巴斯德毕赤酵母的PNO1多肽)。例如,真菌细胞可以表达来自解脂耶氏酵母的MNN4多肽(Genbank® 登录号:XM\_503217, Genolevures Ref: YALI0D24101g)。在一些实施方案中,遗传工程化细胞是OCH1活性缺陷的,并且表达能够促进甘露糖基磷酸化的多肽。

[0106] 在脱帽和脱甘露糖基化后,可以分离分子复合物。在一些实施方案中,将分子复合物在酵母细胞内维持,并且在细胞裂解后释放。在一些实施方案中,经由由指导分子从细胞中分泌的编码序列(对于外源核酸而言天然的或工程化改造到表达载体中)提供的机制将分子复合物分泌到培养基中。可以通过多种用于检测分子存在的标准方案,例如用对GAA特异性的抗体进行的免疫印迹或放射性免疫沉淀,或测试酶比活(specific enzyme activity)(例如糖原降解)确认细胞裂解物或培养基中脱帽且脱甘露糖基化的分子复合物的存在。

[0107] 在一些实施方案中,在分离后,可以将脱帽且脱甘露糖基化的靶分子或分子复合物与异源模块附接,例如使用酶促或化学手段。“异源模块”指与改变的靶分子连接(例如共价或非共价)的任何组成,该组成不同于最初存在于改变的靶分子上的组成。异源模块包括例如聚合物,载体,佐剂,免疫毒素,或可检测(例如荧光,发光,或放射性)模块。在一些实施方案中,可以将额外的N-聚糖添加到改变的靶分子。

[0108] 用于检测分子的糖基化的方法包括DNA测序仪辅助(DSA)、荧光团辅助碳水化合物电泳(FACE)或表面增强的激光解吸/离子化飞行时间质谱术(SELDI-TOF MS)。例如,分析可以利用DSA-FACE,其中例如,将糖蛋白变性,接着在例如膜上固定化。然后,可以用合适的还原剂诸如二硫苏糖醇(DTT)或β-巯基乙醇来还原糖蛋白。可以使用酸诸如碘乙酸来将蛋白质的硫氢基基团羧化。接着,可以使用酶诸如N-糖苷酶F自蛋白质释放N-聚糖。任选地,可以将N-聚糖重建,并通过还原性胺化来衍生化。例如,可以通过还原胺化在还原端用荧光团,如APTS(8-氨基苊-1,3,6-三磺酸(8-aminopyrene-1,3,6-trisulfonic acid))标记释放的N-聚糖。标记的化学计量学使得仅将一个APTS分子附接于每个寡糖分子。然后,可以将衍生化的N-聚糖浓缩。适合于N-聚糖分析的设备包括例如ABI PRISM® 377 DNA测序仪(Applied Biosystems)。可以使用例如GENESCAN® 3.1软件(Applied Biosystems)来实施数据分析。可以用一种或多种酶诸如小牛肠磷酸酶进一步处理分离的甘露糖蛋白(mannoprotein)以确认其N-聚糖状态。N-聚糖分析的其它方法包括例如质谱术(例如MALDI-TOF-MS)、正常相的高压液相层析(HPLC)、反相和离子交换层析(例如,在未标记聚糖时通过脉冲安培计检测,且若将聚糖适当标记,则通过UV吸光度或荧光)。还可见Callewaert等(2001)Glycobiology 11(4):275-281及Freire等(2006)Bioconjug.Chem.17(2):559-564。

[0109] 工程化细胞的培养

[0110] 本文件还提供了本文中所描述的任何遗传工程化细胞的基本上纯的培养物。如本文中所使用的,遗传工程化细胞的“基本上纯的培养物”是所述细胞的如下培养物,其中培养物中活细胞总数的小于约40%(即,小于约:35%;30%;25%;20%;15%;10%;5%;2%;1%;0.5%;0.25%;0.1%;0.01%;0.001%;0.0001%;或甚至更少)是与遗传工程化细胞不同的活细胞,例如,细菌、真菌(包括酵母)、支原体、或原生动细胞。在本上下文中的术

语“约”意味着相关百分比可以是在规定的百分比之上或之下的规定百分比的15%百分比。如此,例如,约20%可以是17%至23%。遗传工程化细胞的此类培养物包括细胞和生长、贮存、或转运培养基。培养基可以是液体、半固体(例如,凝胶状培养基)、或冷冻的。培养基包括在液体中或在半固体培养基中/上生长或在贮存或转运培养基,包括冷冻贮存或转运培养基中贮存或转运的细胞。培养物在培养容器或贮存容器或基片(例如,培养皿、烧瓶、或管或贮存小瓶或管)中。

[0111] 可以将本文中所描述的遗传工程化细胞例如,以冷冻的细胞悬浮液在例如含有冻干保护剂诸如甘油或蔗糖的缓冲液中以冻干细胞贮存。或者,可以将它们例如以例如通过流化床干燥或喷雾干燥,或任何其它合适的干燥方法获得的干燥细胞制备物贮存。

[0112] 药物组合物 and 治疗方法

[0113] 可以将本文中描述的GAA分子和分子复合物,例如含有至少一处增强对哺乳动物细胞内部转运的修饰的分子复合物掺入药物组合物中,所述药物组合物含有治疗有效量的分子和一种或多种佐剂、赋形剂、载体、和/或稀释剂。可接受的稀释剂、载体和赋形剂通常没有不利地影响接受体的稳态(例如,电解质平衡)。可接受载体包括生物相容性、惰性或生物可吸收盐、缓冲剂、寡或多糖、聚合物、粘性改善剂、防腐剂等。一种例示性的载体是生理盐水(0.15M NaCl, pH 7.0至7.4)。另一种例示性的载体是50mM磷酸钠、100mM氯化钠。关于配制和施用药物组合物的技术的更多详情可参见例如Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, Pa.)。补充性活性化合物也可以掺入组合物中。

[0114] 含有具有本文中描述的一处或多处修饰的分子复合物的药物组合物的施用可以是系统的或局部的。可以如下配制药组合物,使得它们适合于胃肠外和/或非胃肠外施用。具体的施用模式包括皮下、静脉内、肌肉内、腹膜内、经皮、鞘内、口服、直肠、含服、表面、鼻、眼、关节内、动脉内、蛛网膜下、支气管、淋巴系统、阴道、和子宫内施用。

[0115] 施用可以通过周期性注射药物组合物的推注或者可以通过自外部(例如,IV袋)或内部(例如,生物可蚀性植入物、生物人工器官、或植入的经改变的N-糖基化分子生成细胞的集落)的储库的静脉内或腹膜内施用而不是不中断的或连续的。见例如美国专利4,407,957,5,798,113和5,800,828。可以使用合适的投递手段来实现药物组合物的施用,诸如:泵(见例如Annals of Pharmacotherapy, 27:912 (1993); Cancer, 41:1270 (1993); Cancer Research, 44:1698 (1984); 微囊化(microencapsulation) (见例如美国专利4,352,883; 4,353,888; 和5,084,350); 连续释放聚合物植入物(见例如Sabel, 美国专利No. 4,883,666); 宏囊化(macroencapsulation) (见例如美国专利5,284,761, 5,158,881, 4,976,859和4,968,733及公布的PCT专利申请W092/19195, W0 95/05452); 注射(皮下、静脉内、动脉内、肌肉内、或对其它合适的部位); 或口服施用(在胶囊、液体、片剂、丸剂、或延长释放的配制剂中)。

[0116] 胃肠外投递系统的例子包括乙烯-乙烯基乙酸酯共聚物颗粒、渗透泵、可植入输注系统、泵投递、包囊细胞投递、脂质体投递、针投递注射、无针注射、喷雾器、气雾器、电穿孔、和经皮贴片。

[0117] 适合于胃肠外施用的配制剂通常含有经改变的N-糖基化分子的无菌水性制备物,其优选地是与接受体的血液等张的(例如,生理盐水溶液)。配制剂可以以单位剂量或多剂形式呈现。

[0118] 适合于口服施用的配制剂可以以离散单元呈现,诸如胶囊、扁囊剂、片剂、或锭剂,它们各自含有预定量的经改变的N-糖基化分子;或水性液体或非水性液体中的悬浮液,诸如糖浆剂、酞剂、乳剂、或顿服水剂(draught)。

[0119] 可以对哺乳动物(例如,人患者)以例如膏剂、喷雾剂、泡沫、凝胶、油膏、软膏、或干擦(dry rub)施用适合于表面施用的含有至少一处增强复合物对哺乳动物细胞内部转运的修饰的分子复合物。干擦在施用部位可以是再水合的。此类分子也可以直接输注入(例如,浸入并干燥)绷带、纱布、或贴片,其然后可以直接表面应用。此类分子也可以以半液体、胶化、或完全液体状态在供表面施用的绷带、纱布或贴片中维持(见例如美国专利No.4,307,717)。

[0120] 可以以本领域技术人员可确定的剂量方案对此需要的受试者施用治疗有效量的药物组合物。例如,可以对受试者以每剂0.01 $\mu\text{g}/\text{kg}$ 至10,000 $\mu\text{g}/\text{kg}$ 受试者体重的剂量系统性施用组合物。在另一个例子中,剂量是每剂1 $\mu\text{g}/\text{kg}$ 至100 $\mu\text{g}/\text{kg}$ 受试者体重。在另一个例子中,剂量是每剂1 $\mu\text{g}/\text{kg}$ 至30 $\mu\text{g}/\text{kg}$ 受试者体重,例如,每剂3 $\mu\text{g}/\text{kg}$ 至10 $\mu\text{g}/\text{kg}$ 受试者体重。

[0121] 为了优化治疗效力,首先可以以不同给药方案施用本文中描述的分子复合物。单位剂量和方案取决于如下因素,其包括例如,哺乳动物物种、其免疫状态、哺乳动物的体重。通常,可以使用合适的筛选测定法作为临床测试规程的一部分来监测此类分子复合物在组织中的水平,例如以测定给定治疗方案的功效。

[0122] 本文中描述的分子复合物的给药频率在医学从业人员(例如医生或护士)的技术和临床判断内。通常,通过可以建立最佳施用参数的临床试验来建立施用方案。然而,从业人员可以依照受试者的年龄、健康、重量、性别和医学状态来改变此类施用方案。可以根据治疗是预防性还是治疗性来改变给药频率。

[0123] 可以通过例如细胞培养物或实验动物中的已知药学规程来确定此类分子复合物或其药物组合物的毒性和治疗效力。例如,可以使用这些规程来测定LD50(对50%的群体致死的剂量)和ED50(在50%的群体中治疗有效的剂量)。毒性与治疗效果之间的剂量比率是治疗系数,并且其可以表示为比率LD50/ED50。优选展现出高的治疗系数的药物组合物。虽然可以使用展现出毒性副作用的药物组合物,但是要注意设计如下的投递系统,其将此类化合物靶向受累组织位置以使对正常细胞(例如,非靶定细胞)的潜在损害最小化,由此降低副作用。

[0124] 可以在配制合适受试者(例如,人患者)用剂量范围中使用自细胞培养测定法和动物研究获得的数据。此类药物组合物的剂量一般位于包括具有少许毒性或没有毒性的ED50的循环浓度范围内。剂量可以在此范围内变化,这取决于所采用的剂量形式和所利用的施用路径。对于如本文中所描述的那样使用的药物组合物(例如,用于治疗受试者中的代谢紊乱),最初可以自细胞培养测定法评估治疗有效剂量。可以在动物模型中配制剂量以达到包括IC50(即,实现对症状的半最大抑制的药物组合物的浓度)的循环血浆浓度范围,如在细胞培养中所测定的。可以使用此类信息来更精确地确定人中有用的剂量。可以例如通过高效液相层析来测量血浆中的水平。

[0125] 如本文中所定义的,分子复合物的“治疗有效量”是能够在经治疗的受试者中产生医学上期望的结果(例如,改善蓬佩氏病的一种或多种症状)的复合物的量。治疗有效量(即,有效剂量)可以包括每千克受试者或样品重量毫克或微克量的化合物(例如,每千克约

1微克至每千克约500毫克、每千克约100微克至每千克约5毫克、或每千克约1微克至每千克约50微克)。

[0126] 受试者可以是任何哺乳动物,例如,人(例如,人患者)或非人灵长类(例如,黑猩猩、狒狒、或猴)、小鼠、大鼠、兔、豚鼠、沙鼠、仓鼠、马、一类牲畜(例如,母牛、猪、绵羊、或山羊)、犬、猫、或鲸。

[0127] 本文中描述的分子复合物或其药物组合物可以作为与用于蓬佩氏病的另一种治疗的联合疗法对受试者施用。例如,联合疗法可以包括对受试者(例如人患者)施用一种或多种别的药剂,其对已经形成或有风险形成(由于编码GAA的基因中的突变)蓬佩氏病的受试者提供治疗益处。如此,化合物或药物组合物和一种或多种别的药剂可以同时施用。或者,可以首先施用分子复合物,接着施用一种或多种别的药剂,或反之亦然。

[0128] 本文中描述的任何分子复合物可以是冻干的。

[0129] 本文中所描述的任何药物组合物可以与施用用法说明书一起包含在容器、包装、或分配器中。

[0130] 本文中描述的任何药物组合物可以与施用说明书一起包含在容器,包装,或分配器中。在一些实施方案中,组合物可以以一次性管形瓶包装。

[0131] 以下是本发明实施的实施例。它们不应解释为以任何方式限制本发明的范围。

## 实施例

[0132] 实施例1

[0133] 用CcMan5和刀豆 $\alpha$ -甘露糖苷酶对重组huGAA的脱帽和脱甘露糖基化

[0134] 如W02011/039634中描述的,使用解脂耶氏酵母生产菌株OXY1589生成重组人GAA(rhGAA),所述解脂耶氏酵母生产菌株OXY1589含有三个拷贝的人 $\alpha$ 1,4-甘露糖苷酶基因(又称为酸性 $\alpha$ 1,4-甘露糖苷酶或酸性麦芽糖酶EC3.2.1.3)和两个拷贝的解脂耶氏酵母MNN4基因。人GAA的氨基酸序列在图1中列出。菌株OXY1589的基因型如下:

[0135] MatA, leu2-958, ura3-302, xpr2-322,

[0136] gut2-744, ade2-844

[0137] POX2-Lip2pre-huGAA:URA3Ex::zeta

[0138] POX2-Lip2pre-huGAA:LEU2Ex::zeta

[0139] POX2-Lip2pre-hGM-CSF:GUTEx::zeta

[0140] Y1MNN4-POX2-hp4d-YLMNN4:ADE2::PT targeted

[0141] 用纤维化纤维细菌甘露糖苷酶(CcMan5)和刀豆 $\alpha$ 甘露糖苷酶(JbMan)(Sigma Product M7257, 3.0M磷酸铵悬浮液)使RhGAA脱帽并脱甘露糖基化。如下重组生成CcMan5, 首先将编码CcMan5多肽的核酸(图2A)克隆到载体pLSAH36中,该载体pLSAH36含有DsbA信号序列,并且导致具有N端HIS标签的蛋白质的表达。图2B和2C分别含有具有和没有信号序列的CcMan5多肽的氨基酸序列。将质粒pLSAH36克隆到大肠杆菌B21细胞中,并使用Talon柱将驻留于周质中的蛋白质分离并纯化。在使用JbMan的硫酸铵悬浮液前,将其通过凝胶过滤过Superdex 200柱以除去污染性磷酸酶活性来进一步纯化。

[0142] 通过以对于huGAA:CcMan5:JbMan为100:5:10的w:w比率与CcMan5(磷酸盐缓冲盐水(PBS)中约0.15-0.30mg/mL)和JbMan(PBS中约0.5-1mg/mL)一起温育使RhGAA(20mM乙酸

钠(NaOAc)缓冲液,pH 5.0中约5mg/mL的浓度)脱帽并脱甘露糖基化。用500mM NaOAc缓冲液,pH 5.0和100mM CaCl<sub>2</sub>稀释总反应体积以获得100mM NaOAc和2mM CaCl<sub>2</sub>的终浓度。将反应混合物于30°C温育16小时。

[0143] 为了评估脱帽过程及分析纯化的huGAA的N-聚糖概况,将5μg糖蛋白的N-聚糖用N-糖苷酶F(PNGaseF)释放,用APTS(8-氨基-1,3,6-萘三磺酸;三钠盐)标记,随后在DSA-FACE(DNA测序仪辅助荧光团辅助碳水化合物电泳)上分析。该方法基本上遵循Laroy等,Nature Protocols,1:397-405(2006)中描述的方案。

[0144] 图3A中呈现了用CcMan5和JbMan处理之前(小图B)和之后(小图C)的来自huGAA(76kD形式)的N-聚糖的DSA-FACE电泳图。小图A是含有用PNGaseF从RNaseB释放的N-聚糖的参照样品。从加帽的huGAA释放的N-聚糖混合物主要由ManP-Man<sub>8</sub>GlcNAc<sub>2</sub>和(ManP)<sub>2</sub>-Man<sub>8</sub>GlcNAc<sub>2</sub>组成(图3A,小图B)。比ManP-Man<sub>8</sub>GlcNAc<sub>2</sub>略快运行的峰可以归为ManP-Man<sub>7</sub>GlcNAc<sub>2</sub>。脱帽和脱甘露糖基化后观察到的主峰可以归为双重磷酸化的P<sub>2</sub>-Man<sub>6</sub>GlcNAc<sub>2</sub>和单磷酸化的P-Man<sub>4</sub>GlcNAc<sub>2</sub>、P-Man<sub>5</sub>GlcNAc<sub>2</sub>和P-Man<sub>6</sub>GlcNAc<sub>2</sub>(小图C)。

[0145] huGAA的不同经加工形式的脱帽对于76kD形式(小图B),95kD形式(小图C)和110kD形式(小图D)产生相同的N-聚糖概况(图3B)。

[0146] 实施例2

[0147] 110kDa rhGAA的纯化

[0148] 如下从菌株OXY1589分离rhGAA的110kDa形式。在收获后,将培养液离心,并使用Durapore膜(Merck Millipore)过滤。将硫酸铵(AMS)添加至1M的浓度,并过滤溶质,之后在疏水性相互作用层析(HIC)柱上加载,该柱在20mM磷酸钠pH 6,1M硫酸铵中平衡。用20mM磷酸钠pH 6洗脱产物。

[0149] 在第二层析柱上加载前,首先将产物经由再生纤维素膜上的切向流过滤(TFF)浓缩,然后从缓冲液更换为20mM乙酸钠pH 4.5。在用20mM乙酸钠pH4.5平衡的阳离子交换层析(CEX)柱上加载此材料。在柱加载后,将其用平衡缓冲液清洗,直到UV吸光度信号达到基线,然后用20mM乙酸钠pH 4.5,50mM NaCl清洗。将产物在20mM乙酸钠pH 4.5,150mM NaCl中洗脱,并浓缩,并从缓冲液更换成20mM乙酸钠pH 5。(参见下文)

[0150] 如实施例中描述的那样使样品脱帽并脱甘露糖基化,然后将D-甘露醇添加至100mM的浓度。将所述材料的四分之三使用具有10kDa分子量截留(MWCO)的再生纤维素膜经由TFF在柱中减少,并在Superdex 200柱上经由大小排阻层析(SEC)进一步纯化,所述Superdex 200柱于4°C用25mM磷酸钠pH6,150mM NaCl,100mM D-甘露醇平衡。然后,在变性条件下在Cibacron蓝染色的聚丙烯酰胺凝胶上对级分筛选纯度。将合并的级分经由TFF浓缩,并使用10kD MWCO(再生纤维素膜,Merck Millipore)的Amicon离心过滤器超速离心。

[0151] 实施例3

[0152] 110kDa rhGAA的纯化

[0153] 如下从菌株OXY1589分离rhGAA的110kDa形式。在收获后,将材料离心并过滤,之后将AMS的浓度提高到1M。将溶质再次过滤,并将产物在用20mM磷酸钠pH 6,1M AMS平衡的HIC柱上捕捉,并在20mM磷酸钠pH 6缓冲液中在1至0M AMS的逐步梯度中释放。

[0154] 将洗脱液浓缩,并在Vivaflow 200型(PES膜,10kD MWCO,Sartorius)上经由TFF将缓冲液更换为10mM BIS-TRIS,pH 6。将经脱盐的材料引入阴离子交换层析(AEC)柱上。在柱

清洗到UV信号几乎达到基线后,应用两相连续盐梯度;第一梯度从0变为0.3M NaCl,第二梯度从0.3变为1M NaCl。将级分在梯度期间收集,并经由定性4-甲基伞形酮基- $\alpha$ -D-葡萄糖吡喃糖苷(4-MUG)筛选GAA。在4-MUG测定法中,通过将反应缓冲液添加至10 $\mu$ l洗脱级分开始反应,所述反应缓冲液由10:1体积比例的0.35mM 4-MUG,0.1%BSA和100mM乙酸钠pH 4组成。于37 $^{\circ}$ C温育30分钟至1小时后,添加等体积的100mM甘氨酸pH 11以停止反应。在UV光下观察荧光生成反应产物4-甲基伞形酮(4-methylumbelliferone)的释放。

[0155] 将含有GAA的级分合并,并通过Vivaflow 200模块(PES膜,10kD MWC0,Sartorius)上的TFF和使用10kD MWC0的Amicon离心滤器(再生纤维素膜,Merck Millipore)进行的超速离心浓缩。

[0156] 将浓缩的材料分成两份,并在Superdex 200柱上进一步纯化,所述Superdex 200柱于4 $^{\circ}$ C用50mM磷酸钠pH 6,250mM NaCl,0.05%Tween-20平衡。随后,在变性条件下在Cibacron-蓝染色的聚丙烯酰胺凝胶上对级分筛选纯度,并使用对硝基苯磷酸盐(其在405nm波长测量黄色的对硝基苯酚盐反应产物的酶促释放)使用比色测试测定磷酸酶含量。

[0157] 从含有GAA的级分制备实验用(pilot)合并物。经由Bradford测定法测定实验用合并物的总蛋白质。将选择的级分合并以浓缩到Vivaflow 200TFF模块(PES膜,10kD MWC0,Sartorius)上。使用10kD MWC0的15ml Amicon离心滤器(再生纤维素膜,Merck Millipore)进一步减少体积。

[0158] 将浓缩的材料进行第二轮大小排阻层析(SEC),使用与第一SEC步骤相同的条件进行。在变性条件下在Cibacron-蓝染色的聚丙烯酰胺凝胶上对级分再洗筛选纯度。将级分依照选择的实验用合并物合并,并在15ml Amicon离心滤器(10kD MWC0,再生纤维素膜,Merck Millipore)上浓缩。

[0159] 实施例4

[0160] 95kDa rhGAA的纯化

[0161] 如下从菌株OXY1589分离rhGAA的95kDa形式。在收获后,将培养液离心,并使用Durapore膜(Merck Millipore)过滤。然后,在具有10kD分子量截留(MWC0)的改良聚醚砜(PES)膜上经由TFF浓缩产物。将AMS添加至1M的浓度,并过滤溶质,之后在HIC柱上加载,该HIC柱在20mM磷酸钠pH 6,1M AMS中平衡。将产物用水洗脱,将洗脱液的pH通过将100mM BIS-TRISpH 6的储液缓冲液添加至10mM的浓度调节,并将其于4 $^{\circ}$ C贮存13天。

[0162] 在AEX柱上加载前,将产物在具有10kD MWC0的再生纤维素膜上经由TFF浓缩,并将缓冲液更换为10mM BIS-TRIS pH 6。将经脱盐的材料经由AEX层析进一步加工,所述AEX层析如实施例3中描述的那样实施。将级分在梯度期间收集,并经由定性4-MUG测定法筛选GAA。将含有GAA的级分合并以进一步纯化。

[0163] 对于第三个层析步骤,将AMS的浓度提高到1M,并且在过滤后,经由HIC进一步纯化材料。在梯度过程中收集级分期间应用1至0M AMS的连续盐梯度。经由定性测定法对所有级分筛选GAA,并且将那些含有GAA的级分合并以进一步加工。

[0164] 将合并物使用10kD MWC0再生纤维素膜的15ml Amicon离心滤器经由超速离心浓缩,并经由SEC使用与实施例3中所述相同的规程进一步纯化。然后,在变性条件下在经Cibacron-蓝染色的聚丙烯酰胺凝胶上对级分筛选纯度。将90%纯的GAA级分合并,并首先在TFF Vivaflow 200模块(PES膜,10kDMWC0,Sartorius)上浓缩,然后使用15ml Amicon离

心滤器(10kD MWC0,再生纤维素膜,Merck Millipore)进行超速离心。将浓缩的材料进行第二轮SEC,使用与第一SEC步骤相同的条件进行。使用定性4-MUG GAA测定法对级分筛选GAA。将具有GAA活性的级分合并并浓缩。

[0165] 在如实施例1中所列的脱帽和脱甘露糖基化后,将D-甘露醇添加至100mM的浓度,并将体积再次减少,之后加载到最终的Superdex 200凝胶过滤柱上,该凝胶过滤柱于4℃用25mM磷酸钠pH 6,150mM NaCl,100mM D-甘露醇平衡。使用4-MUG定性测定法对级分筛选GAA,并将那些含有产物的级分合并并浓缩。

[0166] 实施例5

[0167] 95-110kDa rhGAA混合物的纯化

[0168] 如下从菌株OXY1589分离rhGAA的110kDa前体和95kDa形式两者。在收获后,将材料进行第二层析步骤,如实施例2中描述的。在HIC步骤后,将产物浓缩,并经由TFF将缓冲液更换成10mM BIS-TRIS pH 6,并在AEX柱上加载。将产物在0至300mM NaCl的单一步骤中于pH 6(10mM BIS-TRIS)洗脱,然后使用Pellicon XL50TFF模块(具有10kD MWC0的再生纤维素膜)浓缩。将一半的材料经由大小排阻层析进一步纯化。如实施例3中描述的那样实施层析步骤,但是仅基于变性条件下经Cibacron-蓝染色的聚丙烯酰胺凝胶上的纯度进行级分的选择以进一步加工。

[0169] 将一半的合并物浓缩,并与AEX材料的剩余部分组合。在脱帽和脱甘露糖基化后,将D-甘露醇的浓度提高到100mM,并且遵循与实施例2中的描述相同的规程进行随后的浓缩和SEC步骤。将级分基于4-MUG定性测定法合并,并与来自实施例6的脱帽产物合并。

[0170] 实施例6

[0171] 95kDa rhGAA的纯化

[0172] 如下从菌株OXY1589分离rhGAA的95kDa形式。在收获后,将材料加工,直到且包括AEX步骤,如实施例3中描述的。在AEX步骤中,相当大量的产物由于加载过程中电导率的增加而驻留于流过级分中。因此,将流过材料再次渗滤到合适的缓冲液,并进行第二轮AEX层析。在分开加工后,两个量(批次A和批次B)都来自这里。

[0173] 将批次A与来自实施例5的SEC合并物的剩余部分和来自实施例2的CEX合并物的剩余部分组合,随后将合并物浓缩并渗滤到含有10mM BIS-TRIS pH 6,300mM NaCl的缓冲液。将材料于30℃温育65h。然后,通过将1M乙酸钠储液缓冲液pH 5添加至125mM的浓度将pH降低到pH 5,并将样品于30℃再次温育。在24h后,使用40:1重量:重量比产物的总蛋白含量对蛋白酶混合物用Flavourzyme (Novozymes Corp)(一种来自米曲霉的蛋白酶混合物)处理产物,并且最后一次于30℃温育。在过夜温育后,经由第一SEC步骤纯化材料,所述第一SEC步骤在与实施例3中的描述相同的条件下实施。将级分合并,该级分基于还原性条件下经Cibacron-蓝染色的聚丙烯酰胺凝胶经评估含有纯的产物。在浓缩并缓冲液更换为20mM乙酸钠pH 5后,将材料脱帽并脱甘露糖基化,如实施例1中所列的。将D-甘露醇添加至100mM的浓度,并将材料与来自实施例5的脱帽并脱甘露糖基化的材料合并。如实施例2中所描述的那样实施最终的SEC步骤及随后的样品操作。

[0174] 将批次B与来自实施例3的终产物合并,并将合并物浓缩,并渗滤到含有10mM BIS-TRIS pH 6,300mM NaCl的缓冲液。然后,使用与实施例5中的描述相同的重量比将产物于30℃用米曲霉蛋白酶混合物处理过夜温育时段,然后经由第一SEC步骤纯化,实施第一SEC步

骤在如实施例3中描述的相同条件下实施。如实施例5中描述的那样完成产物的进一步加工。

[0175] 在最终批次中,将来自批次A和批次B的产物以在GAA含量上14:1的比率混合。

[0176] 实施例7

[0177] 76kDa rhGAA的纯化

[0178] 如下从菌株OXY1589分离rhGAA的76kDa形式。在收获后,将培养物进行二轮连续离心。将上清液合并,并将AMS引入到约1M的浓度。在溶解后,将在20mM磷酸钠pH 6,1M AMS中预平衡的1个体积的HIC树脂在搅动的情况下添加到50个体积的上清液,以在批次吸收模式(batch uptake mode)中结合产物。将所得的浆体于4℃在不搅动的情况下贮存过夜。在此时段期间,褐色层停留在溶质上方,该溶质经由温和抽吸在早餐除去。将树脂在每个轮次中用三个体积的前导缓冲液(20mM磷酸钠pH 6,1M AMS)清洗三次,之后将其填充到柱中。将填充树脂清洗,直到UV信号几乎达到基线,然后,用水洗脱产物。通过将100mM BIS-TRIS pH 6的储液缓冲液添加到10mM的浓度调节洗脱液的pH。然后,将材料在袋中无菌过滤,并且于4℃贮存11天的时段。

[0179] 如实施例4中所描述的,实施第二和第三层析步骤和附随的材料操作。首先,将第三层析步骤后的合并物在两个平衡偶联的TFF Vivaflow 200模块(PES膜,10kD MWC0, Sartorius)上浓缩约7倍,然后使用10kD MWC0的15ml Amicon离心滤器(再生纤维素膜, Merck Millipore)超速离心。将浓缩的材料分成两份,并使用与实施例4中的描述相同的条件经由SEC进一步纯化。然后,在变性条件下在经Cibacron-蓝染色的聚丙烯酰胺凝胶上对级分筛选纯度。将选择的级分合并以浓缩到两个平行偶联的Vivaflow 200 TFF模块(PES膜,10kD MWC0, Sartorius)上。使用10kD MWC0的15ml Amicon离心滤器(再生纤维素膜, Merck Millipore)进一步减少体积。

[0180] 在脱帽和脱甘露糖基化后,将D-甘露醇添加到100mM的浓度,并将样品在Vivaflow 50 TFF模块(PES膜,10kD MWC0, Sartorius)上再次浓缩,之后加载到最终的SEC柱上,其以与实施例4中的描述相同的方式实施。将含有产物的级分合并并浓缩。

[0181] 实施例8

[0182] 使用人工生色底物对硝基苯基- $\alpha$ -D-葡萄糖吡喃糖苷对huGAA的不同变体

[0183] (76,95和110kD变体)的酶促表征

[0184] 使用人工生色底物对硝基苯基- $\alpha$ -D-葡萄糖吡喃糖苷(PNPG)测定实施例2中获得的未加工的huGAA(110kDa)和实施例7(76kDa),实施例6(95kDa)和实施例4(95kDa)中获得的经加工的huGAA变体的动力学参数。用商业人 $\alpha$ -葡萄糖苷酶Myozyme®(阿葡萄糖苷酶alpha, Genzyme)进行比较。

[0185] 将酶在含有0.1%BSA和100mM AMS的100mM乙酸钠缓冲液pH 4.0(反应缓冲液)中稀释至20 $\mu$ g/ml。将10 $\mu$ l酶溶液一式三份添加到96孔板。将PNPG底物(Sigma)在反应缓冲液中稀释至多个底物浓度(10,6,4,2,1.6,1.2,0.8,和0.4mM),并将90 $\mu$ l稀释底物添加至每孔。将酶促反应于37℃温育60分钟,接着添加100 $\mu$ l 10%碳酸钠,pH 12以淬灭反应。于405nm测量吸光度。测量用对硝基酚(PNP)得到的标准曲线以计算每分钟形成的产物量。以 $\mu$ M/min表示的速度以产生Michaelis-Menten曲线的不同底物浓度的函数绘图。依照对Michaelis-Menten等式的直接拟合,使用GraphPad Prism计算Vmax和Km。计算催化常数

kcat和催化效率kcat/Km。各种酶的比活以U/mg报告,其中1个单位定义为在100mM乙酸钠缓冲液,pH 4.0+0.1%BSA和100mMAMS中在2mM底物浓度每分钟催化1nmol底物水解的酶量。表1中显示了结果。

[0186] 表1

[0187]

	Myozyme	95 kDa (实施例4)	76 kDa (实施例7)	110 kDa (实施例2)	95 kDa (实施例6)
<b>Vmax</b> ( $\mu\text{M}/\text{min}$ )	12	12	14	13	13
<b>Km (mM)</b>	4.4	4.4	4.3	4.4	4.7
<b>kcat (<math>\text{min}^{-1}</math>)</b>	660	677	770	688	730
<b>kcat/Km (<math>\text{min}^{-1}\text{mM}^{-1}</math>)</b>	150	154	179	156	155
<b>比活</b> ( <b>U/mg</b> )	2000	1910	2415	1935	1980

[0188] huGAA的未加工的和经加工的形式和Myozyme对底物PNPG具有相当的动力学参数。这与在关于来自小鼠乳和中国仓鼠卵巢(CHO)培养基的人酸性 $\alpha$ -葡糖苷酶的文献中报告的数据一致(Bijvoet et al(1998), Human Molecular Genetics,7,1815-1824)。报告了未加工的(110kD)和经加工的(76kD)形式对人工底物4-甲基伞形酮基- $\alpha$ -D-葡糖吡喃糖苷具有相同的Km和kcat值。

[0189] 实施例9

[0190] 使用兔肝糖原作为底物对huGAA的不同变体(76,95和110kD变体)的酶

[0191] 促表征

[0192] 使用兔肝糖原(批次N°099K37931V, Sigma)测试未加工的huGAA(110kD变体;实施例2)和经加工的huGAA变体(76kDa, 实施例7;和95kD, 实施例6)的酶促活性。用商业人 $\alpha$ -葡糖苷酶Myozyme®(阿葡糖苷酶alpha, Genzyme)进行比较。将酶在100mM乙酸钠缓冲液pH 4.0中稀释至500ng/ml。将50 $\mu$ l酶溶液一式两份添加到96孔板。将糖原底物在乙酸盐缓冲液中稀释至多个底物浓度(250, 200, 150, 100, 75, 50, 25mg/ml),并将100 $\mu$ l稀释的底物添加到每孔。将酶促反应于37°C温育60分钟。用amplex红底物使用葡萄糖氧化酶法测量葡萄糖的量。

[0193] 测量葡萄糖标准曲线以计算每分钟形成的产物量。以 $\mu\text{M}/\text{min}$ 表示的酶速度以产生Michaelis-Menten曲线的不同底物浓度的函数绘图。依照对Michaelis-Menten等式的直接拟合,使用GraphPad Prism计算Vmax和Km。计算催化常数kcat和催化效率kcat/Km。各种酶的比活以U/mg报告,其中1个单位定义为在100mM乙酸钠缓冲液,pH 4.0中在50mg/ml最终底物浓度每分钟催化形成1 $\mu\text{mol}$ 葡萄糖的酶量。表2中显示了结果。

[0194] 表2

	<b>Myozyme</b>	<b>76 kDa (实施例7)</b>	<b>95 kDa (实施例6)</b>	<b>110 kDa (实施例2)</b>
<b>Vmax (<math>\mu\text{M}/\text{min}</math>)</b>	32 $\pm$ 6	15 $\pm$ 2	13 $\pm$ 1	11 $\pm$ 1
[0195] <b>Km (mM)</b>	600 $\pm$ 140	100 $\pm$ 10	93 $\pm$ 8	162 $\pm$ 17
<b>kcat (<math>\text{min}^{-1}</math>)</b>	21100	10000	8600	7260
<b>kcat/Km (<math>\text{min}^{-1}\text{mM}^{-1}</math>)</b>	35	100	92	45
<b>比活 (U/mg)</b>	14	32	27	16

[0196] 在此实验中,由于兔糖原的有限溶解度而不能达到底物饱和(图4)。对于Myozyme,仅计算表观Km和kcat值。对于三种huGAA变体,测定较低的表现Km值。经加工的形式催化效率比未加工的huGAA和Myozyme的催化效率高2倍。

[0197] 实施例10

[0198] 在蓬佩病的小鼠模型中酸性alpha葡糖苷酶对糖原清除的影响

[0199] 对蓬佩病的小鼠模型施用来自实施例7(76kDa,脱帽且脱甘露糖基化),实施例6(95kDa,脱帽且脱甘露糖基化),和实施例2(110kDa,脱帽且脱甘露糖基化)的GAA产物以测定糖原从骨骼肌和心脏中的清除。

[0200] 用FVB GAA敲除小鼠和FVB野生型小鼠实施实验。选择此动物模型作为测试系统,因为它是早发性婴儿期疾病形式的一种良好代表。从出生起,KO小鼠具有溶酶体糖原的全身化(generalized)且进行性的积累(Bijvoet等,1998,见上文)。雄性和雌性FVB GAA KO小鼠获自Erasmus University,Rotterdam。在处理开始时,小鼠是26-49周龄。

[0201] 通过在尾静脉中以10ml/kg体重(bw)的剂量体积一周一次持续四周缓慢推注静脉内施用测试物质或媒介物。在验尸前16小时使小鼠空腹。在最后一次注射后4天处死动物。表3中显示了研究组的详情。

[0202] 表3

[0203]

组/颜色编码	剂量水平 (mg/kg bw)	剂量体积 (ml/kg bw)	小鼠类型	小鼠数目
1/白色	0	10	WT	9
2/蓝色	0	10	KO	16
3/绿色	20 mg/kg 76 kDa	10	KO	16
4/红色	20 mg/kg 95 kDa	10	KO	16
5/黄色	20 mg/kg 110 kDa	10	KO	16
6/橙色	20 mg/kg Myozyme	10	KO	16

[0204] 器官的灌注和均质化

[0205] 在用PBS灌注后分离心脏和骨骼肌(四头肌,两侧)。通过使用ultra-turrax将组织在10个重量体积的冰冷的PBS中均质化。此后,将匀浆以16微米在冰上超声处理两次,持续15分钟。在以12000g离心30分钟后,收集上清液以测量糖原。

[0206] 生物分析

[0207] 使用确认的定量酶促测定法测量每个个体小鼠的心脏和骨骼肌中的糖原含量。在煮沸组织后,体外添加淀粉葡萄糖苷酶和 $\alpha$ -淀粉酶的混合物以将糖原降解为葡萄糖。用amplex红底物使用葡萄糖氧化酶方法测量葡萄糖的量。糖原的量以 $\mu\text{g}$ 糖原/mg蛋白质报告。

[0208] 统计学分析

[0209] 通过ANOVA分析来自组2,3,4,5的心脏中的糖原含量,接着通过Dunnet氏检验与组6(Myozyme)及与组2(安慰剂)进行事后(post hoc)比较。组1离开统计学分析,并且用作WT小鼠模型中糖原贮积缺乏的质量检查。

[0210] 由于四头肌数据中存在无关观察,使用Kruskal-Wallis检验评估不同产物的糖原含量数据的潜在差异分布。

[0211] 用Wilcoxon秩和检验实施四头肌数据的事后分析。将每个产物组和Myozyme组与安慰剂(组2)组比较,并将每个产物组与Myozyme比较。

[0212] 结果

[0213] 表4显示了每组16只小鼠的心脏(A)和骨骼肌(B)中的平均糖原水平( $\mu\text{g}/\text{mg}$ 蛋白质)。

[0214] 表4

[0215] A. 心脏中的平均糖原水平

[0216]

汇总	均值	sd
WT	0.58	0.95
KO/安慰剂	525.47	67.75
KO/76kDa	377.75	80.20
KO/95kDa	380.56	78.30
KO/110kDa	416.56	106.77

KO/Myozyme	475.83	98.16
------------	--------	-------

[0217] B. 骨骼肌中的平均糖原水平

[0218]

汇总	均值	sd
WT	2.22	0.66
KO/安慰剂	191.80	34.75
KO/76kDa	152.27	35.35
KO/95kDa	169.27	46.68
KO/110kDa	160.39	36.46
KO/Myozyme	186.49	40.61

[0219] 图5显示了心脏 (5A) 和骨骼肌 (5B) 中各个小鼠的糖原水平 ( $\mu\text{g}/\text{mg}$  蛋白质)。结果显示了在以  $20\text{mg}/\text{kg}$  的四次静脉内注射后, 本文中生成的GAA产物 (110kDa, 95kDa, 和76kDa) 与经安慰剂处理的小鼠相比统计学降低心脏中的糖原水平。相同的 Myozyme® 剂量没有减少心脏中的糖原量。76kDa产物和95kDa处理组两者中的糖原水平与 Myozyme® 处理组相比在统计学上不同。在统计学上, 本文中生成的三种不同GAA产物之间没有差异。

[0220] 与经安慰剂处理的或经 Myozyme® 处理的小鼠相比, 本文中生成的76kDa产物还统计学减少骨骼肌中的糖原量。95kDa和110kDa产物两者中的糖原水平与经安慰剂和 Myozyme® 处理的小鼠相比不是统计学不同的, 这可能是由于各个小鼠间的较高偏差所致。与安慰剂相比,  $20\text{mg}/\text{kg}$  的 Myozyme® 不能减少骨骼肌中的糖原水平。

[0221] 实施例11

[0222] 来自米曲霉的蛋白酶的鉴定

[0223] GAA于酸性pH在与少量Flavourzyme (Novozymes Corp) (一种来自米曲霉的蛋白酶混合物) 一起温育后经历特异性蛋白水解切割。所得的GAA产物在还原性条件下在SDS-PAGE上具有约95kD的分子量。在含有来自生产菌株 (解脂耶氏酵母) 的背景蛋白质的某些部分纯化的GAA制备物中观察到类似的蛋白水解活性。

[0224] 为了评估蛋白水解事件, 在蛋白水解处理后使用EndoH除去GAA的N-聚糖, 每个N-糖基化位点单个N-乙酰基葡萄糖胺。这容许通过SDS PAGE进行更准确的评估。然后, 将GAA产物与Flavourzyme蛋白酶混合物或其纯化的样品一起温育。于  $30^\circ\text{C}$  在  $100\text{mM}$  乙酸钠缓冲液 pH 5 中实施反应。将样品在不同时间点时采集, 并在还原性条件下经由SDS-PAGE分析。每道加载含有  $0.5\mu\text{g}$  GAA的体积。

[0225] 为了调查哪个蛋白酶家族负责蛋白酶混合物中对GAA的特异性蛋白水解, 测定法中包括蛋白酶抑制剂, 其对限定的蛋白酶家族是特异性的, 以促进蛋白酶的鉴定。如上文描述的那样实施反应, 只是现在将蛋白酶抑制剂添加至反应混合物。在蛋白水解反应前, 使不可逆抑制剂PMSF (Sigma-Aldrich prod.nr.E5134-500G) 和E-64 (Calbiochem prod.nr.CALB324890-5) 分别以  $1\text{mM}$  和  $10\mu\text{M}$  的浓度与稀释的蛋白酶混合物一起温育。将可逆抑制剂糜蛋白酶抑制素 (chymostatin) (Calbiochem prod.nr.CALB230790-5)、EDTA、和亮抑酶肽 (Calbiochem prod.nr.CALB108976-10MG) 分别以  $60\mu\text{g}/\text{ml}$ ,  $50\text{mM}$  和  $100\mu\text{M}$  的浓度直接添加到反应混合物。

[0226] PMSF和糜蛋白酶抑制素(消除丝氨酸和半胱氨酸蛋白酶的活性的蛋白酶抑制剂)抑制GAA的特异性蛋白水解。不可逆抑制剂E-64(其抑制半胱氨酸蛋白酶)不阻断蛋白水解。这些数据提示了特异性蛋白水解是丝氨酸蛋白酶家族成员。其他测定法提供了支持此假设的更多证据,其中将蛋白酶混合物与PMSF和氧化还原剂二硫苏糖醇(DTT)(其还原二硫键)一起温育。此还原剂的添加还原共价的无活性半胱氨酸蛋白酶:PMSF加合物,这恢复半胱氨酸蛋白酶活性。在被PMSF抑制时,丝氨酸蛋白酶的活性不能通过DTT恢复。使用此性能差异进一步区别对GAA起作用的丝氨酸和半胱氨酸蛋白酶。

[0227] PMSF抑制的蛋白酶与DTT一起温育没有恢复蛋白酶混合物的GAA特异性蛋白水解活性。GAA特异性蛋白水解也不受到金属蛋白酶抑制剂EDTA和广谱抑制剂亮抑酶肽抑制。所有数据指示丝氨酸蛋白酶负责此GAA蛋白水解事件。

[0228] 为了鉴定来自混合物的蛋白酶,使用一系列层析步骤纯化蛋白酶。第一层析步骤使用阴离子交换层析树脂(Q-Sepharose FF,GE healthcare)。在加载前将蛋白酶混合物材料在20mM TRIS-HCl缓冲液pH 7中稀释。收集在20mMTRIS-HCl缓冲液中100mM,300mM和500mM NaCl的流过液和洗脱液。使用如上文描述的测定法分析所有流过液和洗脱液级分。对GAA起作用的蛋白酶存在于运行的流过液级分中,并且与起始材料相比显著富集。

[0229] 经由阳离子交换层析(SP sepharose XL(GE Healthcare)于pH510mM乙酸钠;用0-300mM NaCl洗脱)进一步加工流过液材料。将洗脱级分收集,并经由立即蓝染色SDS PAGE(instant blue stained SDS PAGE)分析,并使用如上文描述的测定法测定感兴趣的蛋白酶的存在。

[0230] 大部分活性存在于CEX层析洗脱物的最后级分中。将最后两个级分合并,并如下经由质谱术分析。将蛋白质混合物脱盐,还原并烷基化,之后胰蛋白酶消化,随后进行LC-MS/MS方法。使用Mascot算法将获得的谱匹配到NCBI数据库上。应用以下设置:

[0231] • 胰蛋白酶,胰凝乳蛋白酶(容许多达4次错误切割(miscleavage))

[0232] • 氧化(M,W),脱酰胺(N,Q)(可变修饰)

[0233] • 脲基甲基化(Carbamidomethylation)(固定修饰)

[0234] • 分类学:真核生物

[0235] • MS容差(tolerance):0.05Da,MS/MS容差:0.05Da

[0236] 从搜索鉴定来自曲霉的碱性蛋白酶(GenBank登录号BAA00258.1;gi217809)。成熟蛋白酶的序列是:

[0237] >gi|217809|dbj|BAA00258.1|碱性蛋白酶[米曲霉]

[0238]

GLTTQKSAPWGLGSI SHKGQQSTDYIYDTSAGEGYAYVVDSGVNVDHEEFEGRASKAYNAAGGQHVDSIGHGTHVS  
GTIAGKTYGI AKKASILSVKVFQGESSTSVILDGFNWAANDIVSKKRTSKAAINMSLGGGYSKAFNDAVENAFEQG  
VLSVVAAGNENS DAGQTSAPAPDAITVAAIQKSNRASFSNFGKVVDFAPGQDILSAWIGSSSATNTISGTSMAT  
PHIVGLSLYLAALENLDGPAAVTKRIKELATKDVVKDVKGSPNLLAYNGNA (SEQ ID NO:8)。

[0239] 来自米曲霉的纯化的蛋白酶的SDS-PAGE凝胶分析显示了存在30kDa(成熟蛋白酶)左右的MW的条带和具有20和10kDa之间的MW的几个条带。将低MW条带从凝胶切出,用胰蛋白酶消化,并通过nano-LC-MS/MS分析。这些条带鉴定为来自米曲霉碱性蛋白酶的产物,这指示来自米曲霉的碱性蛋白酶易受自身蛋白水解。

[0240] 实施例12

[0241] 在解脂耶氏酵母中表达米曲霉蛋白酶

[0242] 本实施例描述了表达成熟蛋白ALP的解脂耶氏酵母的构建。将编码来自米曲霉的碱性蛋白酶(ALP)(EC.3.4.21.63)的基因密码子优化以进行解脂耶氏酵母表达,并以融合构建体化学合成。融合构建体编码酶的整个可读框(ORF),包括信号肽(21个氨基酸),前肽(100个氨基酸)和成熟蛋白(282个氨基酸),接着是接头(SGGG)和His标签(10x His残基)。参见图9。图10中显示了融合构建体的完整核苷酸序列。

[0243] 将ALP的合成ORF以BamHI/AvrII片段克隆到pPT载体系列中,以靶向整合到解脂耶氏酵母基因组中,其利用不同基因座进行表达盒的稳定整合。在pPT载体中,细菌模块源自质粒pHSS6,并且包含细菌复制起点(ori)和赋予卡那霉素抗性的卡那霉素抗性基因(KanR)。整合盒包含a)用于转化到解脂耶氏酵母的选择标志物(URA3;LEU2;ADE2),b)由启动子(POX2;Hp4d)构成的表达盒,c)插入ALP合成构建体的多克隆位点(MCS)和d)Y1LIP2基因终止子。整合盒侧翼有特定基因座的上游(P)和下游(T)序列,以通过同源重组实现对解脂耶氏酵母基因组的稳定单拷贝靶向整合。两个NotI限制性位点在转化前实现表达盒的分离以避免细菌模块的整合。

[0244] 用于解脂耶氏酵母的培养基和技术由Barth和Gaillardin(FEMS Microbiol Rev.,19(4):219-37,1997)描述,使用1 $\mu$ g纯化的整合盒和用于大肠杆菌的标准技术通过由Le Dall等(Curr Genet.,26(1):38-44,1994)描述的乙酸锂方法转化酵母细胞。

[0245] 基于如下的实情在一个自由的基因座及在2个特定的基因座实施表达盒ALP的整合:插入提供发酵过程期间不必要的高度分泌性蛋白质(脂肪酶2和脂肪酶8)的表达消除。最终菌株OXYY2184含有由半组成性Hp4D启动子驱动的ALP的3个表达盒。

[0246] OXYY2184生成重组米曲霉ALP成熟形式(35kDa),平均产生约2至2.5g/L发酵液。使用Bradford技术测定总蛋白质,并使用偶氮酪蛋白(azocasein)作为底物的测定法测量蛋白酶活性。蛋白酶将偶氮酪蛋白消化成酪蛋白和游离的偶氮染料。经消化的蛋白质的沉淀和离心容许游离的偶氮染料在碱性条件测量,这指示蛋白水解活性。在OD 440nm测量此产物的吸光度。可以通过与偶氮酪蛋白稀释系列关联计算消化的偶氮酪蛋白的量,所述偶氮酪蛋白稀释系列具有其吸光度在OD 440nm测量的已知浓度。

[0247] 将菌株OXYY2184的培养上清液中的ALP通过SDS-PAGE评估,并用抗His多克隆抗体免疫检测。解脂耶氏酵母中生成的重组ALP是有活性的,并且与纯化的天然酶具有相似的特性。重组ALP的这些酶特性容许其用于加工rhuGAA前体以获得95kDa rhuGAA形式。

[0248] 将菌株OXYY2122构建以共表达ALP和rhuGAA。将一个拷贝的ALP表达盒整合到表达rhuGAA(4个拷贝的rhuGAA)的接受菌株中。编码rhuGAA和ALP的两个基因都在诱导型POX2启动子下驱动。所得的菌株OXYY2122生成ALP的成熟形式及rhuGAA前体(110Kda)。将菌株OXYY2122的培养上清液中的重组rhuGAA通过SDS-PAGE,接着通过免疫印迹测定,并且确认rhuGAA被加工成上清液中的95kDa形式。此加工是完成的;未检出110kDa形式,而在没有ALP的菌株的相同的培养中,不发生加工。

[0249] 实施例13

[0250] 用解脂耶氏酵母中表达的米曲霉碱性蛋白酶处理rhGAA发酵液后获得的95kDa rhGAA的纯化

[0251] 如下从菌株OXYY1589分离rhGAA的95kDa形式。在收获后,使用陶器膜(Pa11 Corporation)使培养液澄清。经由具有10kD分子量截留(MWCO)的中空纤维膜浓缩产物。将AMS添加至1M的浓度,并将溶质加热至30℃,之后过滤。将过滤物用解脂耶氏酵母(菌株OXYY2184)中重组表达的米曲霉碱性蛋白酶处理,并在没有任何进一步纯化的情况下在发酵液的澄清后使用。总蛋白质:蛋白酶的200:1重量:重量比和于30℃温育16小时导致完全蛋白水解成95kDa产物。

[0252] 进一步纯化后和磷酸化N-聚糖的脱帽和脱甘露糖基化后的分析揭示在PNPG上与表1中的报告具有相似比活的95kDa GAA产物(如SDS-PAGE上观察到的)。

[0253] 实施例14

[0254] 用米曲霉碱性蛋白酶(ALP)处理后rhGAA中的蛋白水解切割位点的鉴定

[0255] 将rhGAA用米曲霉ALP处理,并进一步纯化,如上述实施例中描述的。为了便于序列分析,将纯化的样品用PNGaseF处理以使rhGAA脱糖基化,因为PNGase F使序列中的N-糖基化的天冬酰胺残基脱酰胺成天冬氨酸。

[0256] 为了确认rhGAA的序列,在还原二硫化物桥和烷基化半胱氨酸残基后使用胰蛋白酶消化脱糖基化的蛋白质。将所得的肽混合物进行LC-MS和MS/MS,并将数据匹配到基因编码的蛋白质序列上,由此测定身份。精确的质量(<10ppm)和片段化谱是用于决定鉴定的标准。

[0257] 几乎完全的序列覆盖获自肽混合物(残基23-60,65-535,和538-898),并且蛋白水解切割位点测定为在氨基酸60和65之前(依照SEQ ID NO:1的序列编号)。rhGAA序列中残基60和65之间的缺口可以源自Gly62之前和/或Gly65之前的蛋白水解事件。文献中报告了来自米曲霉的碱性蛋白酶在Leu和Gly之间降解合成的肽Ileu-Gln-Asn-Cys-Pro-Leu-Gly-NH<sub>2</sub>(参见Nakadai等,1973,Agr.Biol.Chem.,37,2685-2694)。

[0258] 此实验中测定的蛋白水解切割位点与溶酶体中观察到的GAA的蛋白水解加工一致。参见Moreland等,2005,J.Biol.Chem.,280,6780-6791,其中对于95kDa多肽,在氨基酸59和氨基酸68之间(依照SEQ ID NO:1的序列编号)鉴定切割位点。切割的N端肽经由链间二硫键仍然结合。

[0259] 其它实施方案

[0260] 虽然本发明已经结合其详细描述进行了描述,前述描述意图例示而不限制本发明的范围,其以所附权利要求书的范围限定。其它方面,优点,和修改在所附权利要求书的范围内。

序列表

<110> 奥克西雷恩英国有限公司

<120> 用于治疗蓬佩氏病的方法和材料

<130> 18990-0034W01

<150> 61/611,485

<151> 2012-03-15

<160> 12

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 898

<212> PRT

<213> 人

<400> 1

Ala Ala Gln Gln Gly Ala Ser Arg Pro Gly Pro Arg Asp Ala Gln Ala  
 1 5 10 15  
 His Pro Gly Arg Pro Arg Ala Val Pro Thr Gln Cys Asp Val Pro Pro  
 20 25 30  
 Asn Ser Arg Phe Asp Cys Ala Pro Asp Lys Ala Ile Thr Gln Glu Gln  
 35 40 45  
 Cys Glu Ala Arg Gly Cys Cys Tyr Ile Pro Ala Lys Gln Gly Leu Gln  
 50 55 60  
 Gly Ala Gln Met Gly Gln Pro Trp Cys Phe Phe Pro Pro Ser Tyr Pro  
 65 70 75 80  
 Ser Tyr Lys Leu Glu Asn Leu Ser Ser Ser Glu Met Gly Tyr Thr Ala  
 85 90 95  
 Thr Leu Thr Arg Thr Thr Pro Thr Phe Phe Pro Lys Asp Ile Leu Thr  
 100 105 110  
 Leu Arg Leu Asp Val Met Met Glu Thr Glu Asn Arg Leu His Phe Thr  
 115 120 125  
 Ile Lys Asp Pro Ala Asn Arg Arg Tyr Glu Val Pro Leu Glu Thr Pro  
 130 135 140  
 His Val His Ser Arg Ala Pro Ser Pro Leu Tyr Ser Val Glu Phe Ser  
 145 150 155 160  
 Glu Glu Pro Phe Gly Val Ile Val Arg Arg Gln Leu Asp Gly Arg Val  
 165 170 175  
 Leu Leu Asn Thr Thr Val Ala Pro Leu Phe Phe Ala Asp Gln Phe Leu  
 180 185 190  
 Gln Leu Ser Thr Ser Leu Pro Ser Gln Tyr Ile Thr Gly Leu Ala Glu  
 195 200 205  
 His Leu Ser Pro Leu Met Leu Ser Thr Ser Trp Thr Arg Ile Thr Leu  
 210 215 220  
 Trp Asn Arg Asp Leu Ala Pro Thr Pro Gly Ala Asn Leu Tyr Gly Ser  
 225 230 235 240  
 His Pro Phe Tyr Leu Ala Leu Glu Asp Gly Gly Ser Ala His Gly Val  
 245 250 255  
 Phe Leu Leu Asn Ser Asn Ala Met Asp Val Val Leu Gln Pro Ser Pro  
 260 265 270  
 Ala Leu Ser Trp Arg Ser Thr Gly Gly Ile Leu Asp Val Tyr Ile Phe  
 275 280 285  
 Leu Gly Pro Glu Pro Lys Ser Val Val Gln Gln Tyr Leu Asp Val Val  
 290 295 300  
 Gly Tyr Pro Phe Met Pro Pro Tyr Trp Gly Leu Gly Phe His Leu Cys  
 305 310 315 320  
 Arg Trp Gly Tyr Ser Ser Thr Ala Ile Thr Arg Gln Val Val Glu Asn  
 325 330 335  
 Met Thr Arg Ala His Phe Pro Leu Asp Val Gln Trp Asn Asp Leu Asp  
 340 345 350  
 Tyr Met Asp Ser Arg Arg Asp Phe Thr Phe Asn Lys Asp Gly Phe Arg  
 355 360 365  
 Asp Phe Pro Ala Met Val Gln Glu Leu His Gln Gly Gly Arg Arg Tyr  
 370 375 380  
 Met Met Ile Val Asp Pro Ala Ile Ser Ser Ser Gly Pro Ala Gly Ser

[0001]

385 390 395 400  
 Tyr Arg Pro Tyr Asp Glu Gly Leu Arg Arg Gly Val Phe Ile Thr Asn  
 405 410 415  
 Glu Thr Gly Gln Pro Leu Ile Gly Lys Val Trp Pro Gly Ser Thr Ala  
 420 425 430  
 Phe Pro Asp Phe Thr Asn Pro Thr Ala Leu Ala Trp Trp Glu Asp Met  
 435 440 445  
 Val Ala Glu Phe His Asp Gln Val Pro Phe Asp Gly Met Trp Ile Asp  
 450 455 460  
 Met Asn Glu Pro Ser Asn Phe Ile Arg Gly Ser Glu Asp Gly Cys Pro  
 465 470 475 480  
 Asn Asn Glu Leu Glu Asn Pro Pro Tyr Val Pro Gly Val Val Gly Gly  
 485 490 495  
 Thr Leu Gln Ala Ala Thr Ile Cys Ala Ser Ser His Gln Phe Leu Ser  
 500 505 510  
 Thr His Tyr Asn Leu His Asn Leu Tyr Gly Leu Thr Glu Ala Ile Ala  
 515 520 525  
 Ser His Arg Ala Leu Val Lys Ala Arg Gly Thr Arg Pro Phe Val Ile  
 530 535 540  
 Ser Arg Ser Thr Phe Ala Gly His Gly Arg Tyr Ala Gly His Trp Thr  
 545 550 555 560  
 Gly Asp Val Trp Ser Ser Trp Glu Gln Leu Ala Ser Ser Val Pro Glu  
 565 570 575  
 Ile Leu Gln Phe Asn Leu Leu Gly Val Pro Leu Val Gly Ala Asp Val  
 580 585 590  
 Cys Gly Phe Leu Gly Asn Thr Ser Glu Glu Leu Cys Val Arg Trp Thr  
 595 600 605  
 Gln Leu Gly Ala Phe Tyr Pro Phe Met Arg Asn His Asn Ser Leu Leu  
 610 615 620  
 Ser Leu Pro Gln Glu Pro Tyr Ser Phe Ser Glu Pro Ala Gln Gln Ala  
 625 630 635 640  
 Met Arg Lys Ala Leu Thr Leu Arg Tyr Ala Leu Leu Pro His Leu Tyr  
 645 650 655  
 Thr Leu Phe His Gln Ala His Val Ala Gly Glu Thr Val Ala Arg Pro  
 660 665 670  
 [0002] Leu Phe Leu Glu Phe Pro Lys Asp Ser Ser Thr Trp Thr Val Asp His  
 675 680 685  
 Gln Leu Leu Trp Gly Glu Ala Leu Leu Ile Thr Pro Val Leu Gln Ala  
 690 695 700  
 Gly Lys Ala Glu Val Thr Gly Tyr Phe Pro Leu Gly Thr Trp Tyr Asp  
 705 710 715 720  
 Leu Gln Thr Val Pro Val Glu Ala Leu Gly Ser Leu Pro Pro Pro Pro  
 725 730 735  
 Ala Ala Pro Arg Glu Pro Ala Ile His Ser Glu Gly Gln Trp Val Thr  
 740 745 750  
 Leu Pro Ala Pro Leu Asp Thr Ile Asn Val His Leu Arg Ala Gly Tyr  
 755 760 765  
 Ile Ile Pro Leu Gln Gly Pro Gly Leu Thr Thr Thr Glu Ser Arg Gln  
 770 775 780  
 Gln Pro Met Ala Leu Ala Val Ala Leu Thr Lys Gly Gly Gln Ala Arg  
 785 790 795 800  
 Gly Glu Leu Phe Trp Asp Asp Gly Glu Ser Leu Glu Val Leu Glu Arg  
 805 810 815  
 Gly Ala Tyr Thr Gln Val Ile Phe Leu Ala Arg Asn Asn Thr Ile Val  
 820 825 830  
 Asn Glu Leu Val Arg Val Thr Ser Glu Gly Ala Gly Leu Gln Leu Gln  
 835 840 845  
 Lys Val Thr Val Leu Gly Val Ala Thr Ala Pro Gln Gln Val Leu Ser  
 850 855 860  
 Asn Gly Val Pro Val Ser Asn Phe Thr Tyr Ser Pro Asp Thr Lys Val  
 865 870 875 880  
 Leu Asp Ile Cys Val Ser Leu Leu Met Gly Glu Gln Phe Leu Val Ser  
 885 890 895  
 Trp Cys

<210> 2  
 <211> 5060  
 <212> DNA  
 <213> 纤维化纤维微生物

```

<400> 2
atgaaaaaga ttggctggc goiggtggt itagttttag cgttttagcc atcggccggc 60
caicaccatc aicaccacgf gggcccggc tccgacgaag tggatgcacc ggaacctccg 120
agegcagait atgcaagcct ggtttagtft ittgittgca ccgaagggtg ttttggtaat 180
gatatgccfg cagcacaggg accgaatggt ctggcaaaag ttaatccggg taccacaccg 240
ggtcgtaata ataccggtta tgattatgco cagagcaaaa ttacgggttt taccatface 300
aatciggatg gtgttggtag tagcgggtgt ggtgggtgac tctgtgtgti tccgaccagc 360
ggtagctata ccgcacgtcc gggtagcagg acciatgacac atccgtttag ccatgatgat 420
gaagatgcag gtccgggttt ttatagcgti ggtctgggta atgtfgcagg caccgatggt 480
gcaaitaccg gtgtcccggg tacaaitgaa gcagaagtig cagcagcaac ccgtagcggf 540
gticactgft atgcatttcc ggcaggttag accccgagcc tggttgttga tctggaaacc 600
aataatacca gccgtcttag cagcagcgtt caggttgaaa ccctgicaga tggcaccgtt 660
gaacigagcg gtccagttlac cggctatlll lataaigcag cctalacccl gtallatacc 720
gcacgcaccg gtccagcttc aaccgttcag accigggttg atgatgatcc tctgtttgat 780
gcaaccgcac aggatggtgt tgaatccgti gcaatctiga cctttgatcc ggcagatgcc 840
ggtagaatig gtctgcaggt taccctgtct ccggttagcg ttgaaccagg acctatigat 900
cagcaggttg aactgggtga tctgagcttt gatgcaattc gtgatctga ccgtgcagaa 960
tggaaigcaa accctggctg ttttgcaati gatgcaagca ccgcaaccga tccgaccggt 1020
gaactgcagc gtctgtttta taccctatcg tatcgcatgt ttgcaatgcc gatgaatgca 1080
accagacca gccgcaccct tctgtgtgtt gatggtgcag ttcatgcagc acagggcctt 1140
acctatfatg atagctgggc aacctgggat gatcttcgca aatttagcgt gatfgcctat 1200
attgatccgg cactgtatcg tgatagtgti cagagccctgg ttaccctgtt tgcagatgca 1260
gaagcaaccg gtacagcggg tggctgggti ggttttgttc atagcgttcc gaccgttctg 1320
tgggaacgta gcagcgttgt tttgtagat gcaattgccca aaggctttga tggttttgat 1380
cgtctggatg aagcatalcc ggcactgcag cccttggttg gtcatatag cgcagatgaa 1440
ctgcctcgig ttgatgttgc aggtaatccg ggtgcaagcg ttcagcgttg ttatgatcag 1500
fatggtctga cgtttatitg cgaatgaacg ggtctgaccg aagaagcaga aaccttgcgc 1560
gaacagggca gctggccgat tgaanaactg accaaaccgg gtcatggac cgcagcagat 1620
gglacacagg ttggctctct gacaccgcti gcagccgatg glagclggca gaggcagat 1680
caltgccaaal ttgaaagcag aggtctgtat cagggcaccc tgggtagta lcallggtat 1740
gatgcctatg atatggaatg actggtttaa gcaatgggtg gtcatgaagc agccctctct 1800
ggtatgcctc atatgtttgg tgaacatgca ccgatgatg gttaaagcaat gctgcatagc 1860
aatgccaatg aatitgatct gcaggcaacc taccgtttta atataccgg tgaaccgagc 1920
ctgacccaga aatggcagc tgcaalltat accaaagaaa ccctggaatcg ctatattgca 1980
accggttagc gctctgcagt tccgtcaggt ggtggtgaa ttacacctcc gctganaacc 2040
aaagtttatc gtctggacc ccgtgtgtatg ctgccgacca tggataatga tgcaggtaca 2100
atgagcacca tgtttgttgc agcagccgtt ggtctgttcc cggttaccgc aggttagcgc 2160
cagtttcagg ttgttagccc gttttttgat agcaccacca ttacctatga tgatggtagc 2220
gcattaccg ttaccgcaga tgggttttag gaagatgctt ttatgttca gaggcaacc 2280
ctggatgggt caacctitgg taataccgtt gttgattatg caaccgttgt tgggttggtca 2340
gatctggcat ttctgatggg tgaacagccg agcgtttggg gcaccgatac cgcaccggca 2400
tttagcatta gcaccgccac cgaatgaacc gcagaagctc ctccggttag cgcagaaccg 2460
accaccgtgc agaccggiga tgggtgtgca ctggaigcaa ccgttaccct gacactggat 2520
ggcgcacgtc tggcagcacc ggcaggtaca gatctgttta ccagcgttgc agcaagcgtt 2580
gttggctctg ccgatgggtt taccgcagca gttaccgttg caagcccagc cgcactgacc 2640
gttagcttga ccggcaccgc atcagcagat gcacgttttt ttgtgcatct gcgtgatgca 2700
gcactggccc atgggttltc agccgcaagc ctgcagggtc aggggttag cgttcttct 2760
ccctctgctc tgagcgtltc aagcgcagaa ctgatgacac tggcagcacl ggtltagat 2820
gcocttcttg ttgctatgg taattatagc agcgttacct ttgatcgttt agcaaccctc 2880
tgacaagaag acaggaagca ctggcgagc aagcagcaac cagcattgca ctgcgttttg 2940
cagcagatcg tctgggtgca gcagcagatg caciggatct gaccggtgtt ggttatcgt 3000
ccctggaagc agaacagagc gaagcatgtt ctgggtgtga actgaaaaat gaagccaata 3060
gcagcagcgg taatctgggt ggtgttctga ccggtagctg ggtcagtat ccgatatga 3120
ccittgaaac cgcagccggt gatacacctc ccggtttctt gaccgttctt tatgatacca 3180
gctttgcacc gaccgatacc ccgagcaccg tctgttctca tgcgggtgat gttcttggic 3240
cggttgttgc aacctttgat ctgaaagcca ccagcgttgg ggttaaatat accgaagtta 3300
ccgcagaact ggttgatgti caggccctgg ttgatgccc ggtgttacc tttgaactgc 3360
tggcaccgag ccgtcttagc tgggttggta attttagttg gtttctctt agcgcagaag 3420
atccggcagc accgggtcag ccgtgtgaaa gcccgaccgt taccatgaa gccgaagatt 3480
ggaccgcaag cagcgtctgt ggtctgaaaa aagaagcagc cacttggacc agcctccgg 3540
tgaccaatgt tgggtgtlaca gcagatggtg atttgattgc ctatggtgaa gttgatctgg 3600
gtgaactgcc gctggggcaa ctgagcgttc attatgttca taatagcaat cgcagcgtta 3660
ataatagcgc actgagcgtt tatctggaig cattttagcc ggctaatccg ggtgaaccgt 3720
ttgttaccgt tcccttccc accaccggtg gcagltggac cgcagatgce acagccaccg 3780
ttgttctgoc ggaacccgtg cagggcacc atgaagtttt tgtctgtctg agcaccgaac 3840
cgtatgcaga ttatccgtat gttgcaaatc tggatagcct gacctttgca ccgggtgttc 3900
cgaccagcgt tgtgtttgaa agcgaagcct ggaccagcaa tctgtgtctg ggcctgaaaa 3960
atgaattctt taccgtgacc tctgttccgg ttacaantgt ggttggcacc gctgatggcg 4020
atiggctgce atatggcgaa atlgaatctg gcagcagcag actggatcag ctgtctgtgc 4080
atfatgttca taattctaaf ccctctggic gtaattctgc actgtctgtg tatctggatg 4140
ccittgatcc ggcaantccg ggtgaaccgt ttgtgacagt gccctggcca aataccggtt 4200
gctcttggac caccgatggt actgcagttg tggatctgcc gctctaccgt cgtgttaaac 4260

```

[0003]

```

atcaggittg ggittgctcg tciaccgaag caiaigecga tcaiccgtat gfggccaatc 4320
tggattctat gcgctttttt accgatgcat atgatgttga agttcctccg accgatacag 4380
cagcaactggc agccgittgt gaigcagcag gtacaccgga agcagaaatf gcaegttafg 4440
gtctgataga tgcctgittt ttaccctgig aactggcagc agcactagc gtictggtcg 4500
atgccggtgc aacacaggca cagccagatg aaccgctcg tctctgggti ctggcaaccg 4560
atcagctggt tccggcagaa cgtcgtcgtc tggaaaatct ggttgccagc gcagaageac 4620
tgaccgacga aggtatitct ccgaaaagct ggcaggcatt tctiaccgca ctggctgctg 4680
caaccggcac cctggatgat gcagcagcat ctgatgaagc actgcatgat gcaegtctgg 4740
cgctgcaggg tgcagttgat gcactggaag aaccggcaga tcttctctcg gttgaagttg 4800
aagtttctcc gcgttctcig gcaggtaaac cgtatgttgc cgttctctca gttaatgitt 4860
ctgatcagc cgttgatgtt gaactggcaa gctctciggg caccctagc ttgttggig 4920
tggcaccggg tgcgagcgca taicagagct ttgcagcccg tagccgaacc ggtgatctgg 4980
atgttaccgt gaccgaacc ggigcagatg gtactcagac cgttgaacag gttgigaccg 5040
ttccgagctg tagetaataa 5060

```

```

<210> 3
<211> 1665
<212> PRT
<213> 纤维化纤维细菌

```

```

<400> 3
Ala Leu Ala Val Val Gly Leu Ala Pro Ala Thr Ala Ala Ser Ala Ala
1 5 10 15
Pro Glu Pro Pro Ser Ala Asp Tyr Ala Ser Leu Val Asp Val Phe Val
20 25 30
Gly Thr Glu Gly Asp Phe Gly Asn Asp Met Pro Ala Ala Gln Ala Pro
35 40 45
Asn Gly Leu Ala Lys Val Asn Pro Arg Thr Thr Pro Gly Arg Asn Asn
50 55 60
Thr Gly Tyr Asp Tyr Ala Gln Ser Lys Ile Ser Gly Phe Thr His Thr
65 70 75 80
Asn Leu Asp Gly Val Gly Gly Ser Gly Gly Gly Asp Leu Leu Val
85 90 95
Val Pro Thr Ser Gly Ser Tyr Thr Ala Arg Pro Gly Thr Gly Thr Tyr
100 105 110
Ala His Pro Phe Ser His Asp Asp Glu Asp Ala Gly Pro Gly Phe Tyr
115 120 125
Ser Val Gly Leu Gly Asn Val Ala Gly Thr Asp Gly Ala Ile Thr Gly
130 135 140
Ala Pro Gly Thr Ile Glu Ala Glu Val Ala Ala Thr Arg Ser Gly
145 150 155 160
Val His Arg Tyr Ala Phe Pro Ala Gly Ser Thr Pro Ser Leu Val Val
165 170 175
Asp Leu Glu Thr Asn Asn Thr Ser Arg Arg Ser Ser Ser Val Gln Val
180 185 190
Glu Thr Arg Ala Asp Gly Thr Val Glu Leu Ser Gly Gln Val Thr Gly
195 200 205
Tyr Phe Tyr Asn Ala Ala Tyr Thr Leu Tyr Tyr Thr Ala Arg Thr Leu
210 215 220
Gln Pro Ala Thr Val Gln Thr Trp Gly Asp Asp Asp Arg Leu Val Asp
225 230 235 240
Ala Thr Ala Gln Asp Gly Val Asp Thr Gly Ala Ile Leu Thr Phe Asp
245 250 255
Pro Ala Asp Ala Gly Glu Ile Gly Leu Gln Val Thr Leu Ser Pro Val
260 265 270
Ser Val Glu Gln Ala Arg Ile Asp Gln Gln Val Glu Leu Gly Asp Leu
275 280 285
Ser Phe Asp Ala Ile Arg Asp Arg Thr Arg Ala Glu Trp Asn Ala Thr
290 295 300
Leu Gly Arg Val Ala Ile Asp Ala Ser Thr Ala Thr Asp Pro Thr Gly
305 310 315 320
Glu Leu Gln Arg Leu Phe Tyr Thr His Leu Tyr Arg Met Phe Ala Met
325 330 335
Pro Met Asn Ala Thr Ser Thr Ser Gly Thr Tyr Arg Gly Val Asp Gly
340 345 350
Ala Val His Ala Ala Gln Gly Phe Thr Tyr Tyr Asp Ser Trp Ala Thr
355 360 365
Trp Asp Asp Phe Arg Lys Phe Ser Val Ile Ala Tyr Ile Asp Pro Ala
370 375 380
Leu Tyr Arg Asp Met Val Gln Ser Leu Val Tyr Leu Phe Ala Asp Ala
385 390 395 400
Glu Ala Thr Gly Thr Gly Gly Gly Leu Gly Gly Phe Val His Ser Val

```

[0004]

[0005]

405 410 415  
 Pro Thr Val Arg Trp Glu Arg Ser Ser Val Val Val Ala Asp Ala Ile  
 420 425 430  
 Ala Lys Gly Phe Asp Gly Phe Asp Arg Leu Asp Glu Ala Tyr Pro Ala  
 435 440 445  
 Leu Gln Arg Leu Val Gly Gln Tyr Ser Ala Asp Glu Leu Arg Arg Gly  
 450 455 460  
 Tyr Val Ala Gly Asn Pro Gly Ala Ser Val Gln Arg Gly Tyr Asp Gln  
 465 470 475 480  
 Tyr Gly Leu Ser Val Ile Ala Asp Glu Leu Gly Leu Thr Glu Glu Ala  
 485 490 495  
 Glu Thr Leu Arg Glu Gln Ala Ser Trp Pro Ile Gln Lys Leu Thr Lys  
 500 505 510  
 Pro Gly Ala Trp Thr Ala Ala Asp Gly Thr Gln Val Gly Leu Leu Thr  
 515 520 525  
 Pro Arg Ala Ala Asp Gly Ser Trp Gln Ser Ala Asp His Ala Lys Phe  
 530 535 540  
 Glu Ala Ala Gly Leu Tyr Gln Gly Thr Leu Trp Gln Tyr His Trp Tyr  
 545 550 555 560  
 Asp Ala Tyr Asp Met Asp Ala Leu Val Glu Ala Met Gly Gly His Glu  
 565 570 575  
 Ala Ala Arg Leu Gly Met Arg His Met Phe Gly Glu His Ala Pro Asp  
 580 585 590  
 Asp Gly Lys Ala Met Leu His Ser Asn Ala Asn Glu Ile Asp Leu Gln  
 595 600 605  
 Ala Pro Tyr Leu Phe Asn Tyr Thr Gly Glu Pro Ser Leu Thr Gln Lys  
 610 615 620  
 Trp Ala Arg Ala Ile Tyr Thr Lys Glu Thr Trp Asn Arg Tyr Ile Ala  
 625 630 635 640  
 Thr Gly Ser Ser Ser Ala Val Pro Ser Gly Gly Gly Glu Phe Thr Pro  
 645 650 655  
 Pro Leu Lys Thr Lys Val Tyr Arg Leu Asp Pro Arg Gly Met Leu Pro  
 660 665 670  
 Thr Met Asp Asn Asp Ala Gly Thr Met Ser Thr Met Phe Val Ala Ala  
 675 680 685  
 Ala Val Gly Leu Phe Pro Val Thr Ala Gly Ser Ser Gln Phe Gln Val  
 690 695 700  
 Gly Ser Pro Phe Phe Asp Ser Thr Thr Ile Thr Tyr Asp Asp Gly Ser  
 705 710 715 720  
 Ala Phe Thr Val Thr Ala Asp Gly Val Ser Glu Asp Ala Phe Tyr Val  
 725 730 735  
 Gln Ser Ala Thr Leu Asp Gly Ala Thr Phe Gly Asn Thr Trp Val Asp  
 740 745 750  
 Tyr Ala Thr Val Val Gly Gly Ala Asp Leu Ala Phe Arg Met Gly Glu  
 755 760 765  
 Gln Pro Ser Asp Trp Gly Thr Asp Thr Ala Pro Ala Phe Ser Met Ser  
 770 775 780  
 Thr Ala Thr Asp Glu Pro Ala Glu Gly Pro Arg Val Ser Ala Glu Pro  
 785 790 795 800  
 Thr Thr Val Gln Thr Gly Asp Gly Gly Ala Leu Asp Ala Thr Val Thr  
 805 810 815  
 Leu Thr Leu Asp Gly Ala Arg Leu Ala Ala Pro Ala Gly Thr Asp Leu  
 820 825 830  
 Val Thr Ser Gly Ala Ala Ser Val Val Gly Leu Pro Asp Gly Val Thr  
 835 840 845  
 Ala Ala Val Thr Val Ala Ser Pro Thr Ala Leu Thr Val Ser Leu Thr  
 850 855 860  
 Gly Thr Ala Ser Ala Asp Ala Arg Phe Phe Val His Leu Arg Asp Ala  
 865 870 875 880  
 Ala Leu Ala Asp Gly Val Ala Ala Ala Ser Leu Gln Gly Gln Gly Val  
 885 890 895  
 Ser Val Arg Ser Pro Leu Arg Leu Ser Val Ala Ser Ala Glu Arg Asp  
 900 905 910  
 Ala Leu Ala Ala Leu Val Asp Asp Ala Val Leu Val Arg His Gly Asn  
 915 920 925  
 Tyr Ser Ser Val Thr Phe Asp Arg Phe Ser Thr Ala Leu Thr Lys Ala  
 930 935 940  
 Gln Glu Ala Leu Gly Asp Glu Ala Ala Thr Ser Ile Ala Leu Arg Phe  
 945 950 955 960  
 Ala Ala Asp Arg Leu Gly Ala Ala Ala Asp Ala Leu Asp Leu Thr Gly  
 965 970 975  
 Gly Gly Tyr Arg Thr Leu Glu Ala Glu Gln Ser Glu Ala Trp Ser Gly

[0006]

	980		985		990										
Gly	Glu	Leu	Lys	Asn	Glu	Ala	Asn	Ser	Ser	Ser	Gly	Asn	Leu	Gly	Gly
	995						1000					1005			
Val	Arg	Ser	Gly	Ser	Trp	Val	Gln	Tyr	Arg	Asp	Met	Thr	Phe	Glu	Thr
	1010						1015					1020			
Ala	Ala	Gly	Asp	Thr	Pro	Pro	Arg	Phe	Leu	Thr	Val	Arg	Tyr	Asp	Thr
	1025						1030				1035				1040
Ser	Phe	Ala	Pro	Thr	Asp	Thr	Pro	Ser	Thr	Val	Arg	Val	His	Ala	Gly
			1045						1050						1055
Asp	Val	Ser	Gly	Pro	Val	Val	Ala	Thr	Val	Asp	Leu	Lys	Gly	Thr	Ser
			1060						1065						1070
Gly	Trp	Gly	Lys	Tyr	Thr	Glu	Val	Thr	Ala	Glu	Leu	Gly	Asp	Val	Gln
			1075						1080						1085
Ala	Leu	Val	Asp	Ala	Gln	Val	Val	Thr	Phe	Glu	Leu	Leu	Ala	Pro	Ser
	1090								1095						1100
Gly	Arg	Ser	Trp	Val	Gly	Asn	Phe	Asp	Trp	Phe	Arg	Phe	Ser	Ala	Glu
	1105								1110						1120
Asp	Pro	Ala	Ala	Pro	Gly	Gln	Pro	Gly	Glu	Ser	Pro	Thr	Val	Thr	Ile
				1125					1130						1135
Glu	Ala	Glu	Asp	Trp	Thr	Ala	Ser	Ser	Gly	Arg	Gly	Leu	Lys	Lys	Glu
			1140						1145						1150
Ser	Ser	Thr	Trp	Thr	Ser	Gly	Pro	Val	Thr	Asn	Val	Gly	Gly	Thr	Ala
			1155						1160						1165
Asp	Gly	Asp	Trp	Ile	Ala	Tyr	Gly	Glu	Val	Asp	Leu	Gly	Glu	Leu	Pro
	1170								1175						1180
Leu	Gly	Glu	Leu	Ser	Val	His	Tyr	Val	His	Asn	Ser	Asn	Arg	Ser	Gly
	1185								1190						1200
Asn	Asn	Ser	Ala	Leu	Ser	Val	Tyr	Leu	Asp	Ala	Phe	Asp	Pro	Ala	Asn
				1205					1210						1215
Pro	Gly	Glu	Pro	Phe	Val	Thr	Val	Pro	Leu	Pro	Thr	Thr	Gly	Ser	Ser
			1220						1225						1230
Trp	Thr	Ala	Asp	Gly	Thr	Ala	Thr	Val	Val	Leu	Pro	Glu	Thr	Val	Gln
			1235						1240						1245
Gly	Thr	His	Glu	Val	Phe	Val	Arg	Leu	Ser	Thr	Glu	Pro	Tyr	Ala	Asp
	1250								1255						1260
His	Pro	Tyr	Val	Ala	Asn	Leu	Asp	Ser	Leu	Thr	Phe	Ala	Pro	Gly	Gly
	1265								1270						1280
Pro	Thr	Ser	Val	Val	Val	Glu	Ser	Glu	Ala	Trp	Thr	Ser	Asn	Ser	Gly
				1285					1290						1295
Arg	Gly	Leu	Lys	Asn	Glu	Ser	Ser	Thr	Trp	Thr	Ser	Gly	Pro	Val	Thr
			1300						1305						1310
Asn	Val	Gly	Gly	Thr	Ala	Asp	Gly	Asp	Trp	Leu	Ala	Tyr	Gly	Glu	Ile
	1315								1320						1325
Asp	Leu	Gly	Ser	Ala	Ala	Leu	Asp	Gln	Leu	Ser	Val	His	Tyr	Val	His
	1330								1335						1340
Asn	Ser	Asn	Arg	Ser	Gly	Arg	Asn	Ser	Ala	Leu	Ser	Val	Tyr	Leu	Asp
	1345								1350						1360
Ala	Phe	Asp	Pro	Ala	Asn	Pro	Gly	Glu	Pro	Phe	Val	Thr	Val	Pro	Leu
				1365					1370						1375
Ala	Asn	Thr	Gly	Ser	Ser	Trp	Thr	Thr	Asp	Gly	Thr	Ala	Val	Val	Asp
			1380						1385						1390
Leu	Pro	Ser	Thr	Val	Arg	Gly	Lys	His	Gln	Val	Trp	Val	Arg	Leu	Ser
			1395						1400						1405
Thr	Glu	Ala	Tyr	Ala	Asp	His	Pro	Tyr	Val	Ala	Asn	Leu	Asp	Ser	Met
	1410								1415						1420
Arg	Phe	Phe	Thr	Asp	Ala	Tyr	Asp	Val	Glu	Val	Pro	Pro	Thr	Asp	Thr
	1425								1430						1440
Ala	Ala	Leu	Ala	Ala	Val	Val	Asp	Ala	Ala	Gly	Thr	Pro	Glu	Ala	Glu
				1445					1450						1455
Ile	Ala	Arg	Tyr	Gly	Arg	Ile	Asp	Ala	Arg	Val	Phe	Thr	Arg	Glu	Leu
			1460						1465						1470
Ala	Ala	Ala	Arg	Ser	Val	Leu	Ala	Asp	Ala	Gly	Ala	Thr	Gln	Ala	Gln
	1475								1480						1485
Ala	Asp	Glu	Arg	Ala	Arg	Arg	Leu	Gly	Leu	Ala	Thr	Asp	Gln	Leu	Val
	1490								1495						1500
Pro	Ala	Glu	Arg	Arg	Arg	Leu	Glu	Asn	Leu	Val	Ala	Ser	Ala	Glu	Ala
	1505								1510						1520
Leu	Thr	Asp	Glu	Gly	Tyr	Ser	Pro	Glu	Ser	Trp	Gln	Ala	Phe	Arg	Thr
				1525					1530						1535
Ala	Leu	Ala	Ala	Ala	Thr	Gly	Thr	Leu	Asp	Asp	Ala	Ala	Ala	Ser	Asp
			1540						1545						1550
Glu	Ala	Leu	His	Asp	Ala	Arg	Leu	Ala	Leu	Gln	Gly	Ala	Val	Asp	Ala

1555                      1560                      1565  
 Leu Glu Glu Pro Ala Asp Val Val Leu Val Glu Val Glu Val Ser Pro  
 1570                      1575                      1580  
 Arg Cys Leu Ala Gly Lys Pro Tyr Val Ala Val Arg Ala Val Asn Val  
 1585                      1590                      1595                      1600  
 Ser Asp Ala Ala Val Asp Val Glu Leu Ala Ser Ser Leu Gly Thr Arg  
                             1605                      1610                      1615  
 Ser Phe Val Gly Val Ala Pro Gly Ala Ser Ala Tyr Gln Ser Phe Ala  
                             1620                      1625                      1630  
 Ala Arg Ser Ala Thr Gly Asp Leu Asp Val Thr Val Thr Ala Thr Gly  
                             1635                      1640                      1645  
 Ala Asp Gly Thr Gln Thr Val Glu Gln Val Val Thr Val Pro Ser Cys  
                             1650                      1655                      1660  
 Ser  
 1665

<210> 4  
 <211> 1650  
 <212> PRT  
 <213> 纤维化纤维细菌

[0007]

<400> 4  
 Ala Pro Glu Pro Pro Ser Ala Asp Tyr Ala Ser Leu Val Asp Val Phe  
 1                              5                              10                              15  
 Val Gly Thr Glu Gly Asp Phe Gly Asn Asp Met Pro Ala Ala Gln Ala  
                             20                              25                              30  
 Pro Asn Gly Leu Ala Lys Val Asn Pro Arg Thr Thr Pro Gly Arg Asn  
                             35                              40                              45  
 Asn Thr Gly Tyr Asp Tyr Ala Gln Ser Lys Ile Ser Gly Phe Thr His  
                             50                              55                              60  
 Thr Asn Leu Asp Gly Val Gly Gly Ser Gly Gly Gly Asp Leu Leu  
 65                              70                              75                              80  
 Val Val Pro Thr Ser Gly Ser Tyr Thr Ala Arg Pro Gly Thr Gly Thr  
                             85                              90                              95  
 Tyr Ala His Pro Phe Ser His Asp Asp Glu Asp Ala Gly Pro Gly Phe  
                             100                              105                              110  
 Tyr Ser Val Gly Leu Gly Asn Val Ala Gly Thr Asp Gly Ala Ile Thr  
                             115                              120                              125  
 Gly Ala Pro Gly Thr Ile Glu Ala Glu Val Ala Ala Thr Arg Ser  
                             130                              135                              140  
 Gly Val His Arg Tyr Ala Phe Pro Ala Gly Ser Thr Pro Ser Leu Val  
 145                              150                              155                              160  
 Val Asp Leu Glu Thr Asn Asn Thr Ser Arg Arg Ser Ser Val Gln  
                             165                              170                              175  
 Val Glu Thr Arg Ala Asp Gly Thr Val Glu Leu Ser Gly Gln Val Thr  
                             180                              185                              190  
 Gly Tyr Phe Tyr Asn Ala Ala Tyr Thr Leu Tyr Tyr Thr Ala Arg Thr  
                             195                              200                              205  
 Leu Gln Pro Ala Thr Val Gln Thr Trp Gly Asp Asp Asp Arg Leu Val  
                             210                              215                              220  
 Asp Ala Thr Ala Gln Asp Gly Val Asp Thr Gly Ala Ile Leu Thr Phe  
 225                              230                              235                              240  
 Asp Pro Ala Asp Ala Gly Glu Ile Gly Leu Gln Val Thr Leu Ser Pro  
                             245                              250                              255  
 Val Ser Val Glu Gln Ala Arg Ile Asp Gln Gln Val Glu Leu Gly Asp  
                             260                              265                              270  
 Leu Ser Phe Asp Ala Ile Arg Asp Arg Thr Arg Ala Glu Trp Asn Ala  
                             275                              280                              285  
 Thr Leu Gly Arg Val Ala Ile Asp Ala Ser Thr Ala Thr Asp Pro Thr  
                             290                              295                              300  
 Gly Glu Leu Gln Arg Leu Phe Tyr Thr His Leu Tyr Arg Met Phe Ala  
 305                              310                              315                              320  
 Met Pro Met Asn Ala Thr Ser Thr Ser Gly Thr Tyr Arg Gly Val Asp  
                             325                              330                              335  
 Gly Ala Val His Ala Ala Gln Gly Phe Thr Tyr Tyr Asp Ser Trp Ala  
                             340                              345                              350  
 Thr Trp Asp Asp Phe Arg Lys Phe Ser Val Ile Ala Tyr Ile Asp Pro  
                             355                              360                              365  
 Ala Leu Tyr Arg Asp Met Val Gln Ser Leu Val Tyr Leu Phe Ala Asp  
                             370                              375                              380  
 Ala Glu Ala Thr Gly Thr Gly Gly Gly Leu Gly Gly Phe Val His Ser

[0008]

385 Val Pro Thr Val Arg Trp Glu Arg Ser Ser Val Val Val Ala Asp Ala  
 405 410 415  
 420 425 430  
 435 440 445  
 450 455 460  
 465 470 475 480  
 485 490 495  
 500 505 510  
 515 520 525  
 530 535 540  
 545 550 555 560  
 565 570 575  
 580 585 590  
 595 600 605  
 610 615 620  
 625 630 635 640  
 645 650 655  
 660 665 670  
 675 680 685  
 690 695 700  
 705 710 715 720  
 725 730 735  
 740 745 750  
 755 760 765  
 770 775 780  
 785 790 795 800  
 805 810 815  
 820 825 830  
 835 840 845  
 850 855 860  
 865 870 875 880  
 885 890 895  
 900 905 910  
 915 920 925  
 930 935 940  
 945 950 955 960  
 Gly Gly Gly Tyr Arg Thr Leu Glu Ala Glu Gln Ser Glu Ala Trp Ser



1540 1545 1550  
 Ala Leu Glu Glu Pro Ala Asp Val Val Leu Val Glu Val Glu Val Ser  
 1555 1560 1565  
 Pro Arg Cys Leu Ala Gly Lys Pro Tyr Val Ala Val Arg Ala Val Asn  
 1570 1575 1580  
 Val Ser Asp Ala Ala Val Asp Val Glu Leu Ala Ser Ser Leu Gly Thr  
 1585 1590 1595 1600  
 Arg Ser Phe Val Gly Val Ala Pro Gly Ala Ser Ala Tyr Gln Ser Phe  
 1605 1610 1615  
 Ala Ala Arg Ser Ala Thr Gly Asp Leu Asp Val Thr Val Thr Ala Thr  
 1620 1625 1630  
 Gly Ala Asp Gly Thr Gln Thr Val Glu Glu Val Val Thr Val Pro Ser  
 1635 1640 1645  
 Cys Ser  
 1650

<210> 5  
 <211> 1079  
 <212> PRT  
 <213> 解脂耶氏酵母

<400> 5  
 Met Tyr Ser His Phe Asn Asn Glu Pro Val Ala Lys Arg Val Asn Asn  
 1 5 10 15  
 Leu Phe Thr Asp Arg Leu Arg Gln Phe Thr Ser Asp Gly Glu Tyr Arg  
 20 25 30  
 Ser Leu Asn Leu Pro Ala Phe Tyr Gln Arg Glu Arg Leu Asp Gly Lys  
 35 40 45  
 Asn His Val Ala Ile Glu Thr Tyr Ala Val Ser Asp Leu Arg Arg Pro  
 50 55 60  
 Leu Phe Lys Asp Ala Leu Lys Glu Ala Asp Gly His Trp Lys Pro Ala  
 65 70 75 80  
 Lys Lys Gly Ser Glu Tyr Gly Pro Ser Trp Ala Thr His Trp Phe Lys  
 85 90 95  
 Ile Gln Val Cys Val Pro Pro Glu Trp Lys Lys Asn Tyr Tyr Lys Lys  
 100 105 110  
 Gly Asp Leu Val Val Phe Asn Trp Asn Leu Asn Cys Glu Gly Leu Val  
 115 120 125  
 Phe Ser Glu Ser Gly Glu Ala Leu Ile Gly Leu Ser Gly Glu Glu Arg  
 130 135 140  
 Arg Glu Trp Pro Ile Pro Asp Asn Trp Phe Asp Gly Lys Cys His Thr  
 145 150 155 160  
 Phe Tyr Ile Glu Ala Ser Cys Asn Gly Met Phe Gly Asn Ala Thr Gly  
 165 170 175  
 Ser Ser Ile Gln Pro Pro Ser Asp Asn Arg Tyr Phe Arg Leu Asp Ser  
 180 185 190  
 Ala Asp Leu Val Val Ile Asn Ser Gln Ala Arg His Leu Phe Val Asp  
 195 200 205  
 Phe Trp Ile Ile Gly Asp Ala Ala Arg Glu Phe Pro Gly Asp Ser Trp  
 210 215 220  
 Gln Arg Gly Lys Ala Leu Asp Val Ala Asn Lys Ile Met Asp Ala Phe  
 225 230 235 240  
 Asp Pro Glu Asn Pro Asp Glu Ser Ile Ala Glu Gly Arg Lys Leu Ala  
 245 250 255  
 Lys Glu Tyr Leu Gly Asp Thr Thr Lys Ala Tyr Lys Gln Gln Leu Pro  
 260 265 270  
 Phe Ala Asp Gly Leu Val Tyr Ala Leu Gly Asn Cys His Ile Asp Thr  
 275 280 285  
 Ala Trp Leu Trp Pro Phe Ala Glu Thr Arg Arg Lys Ala Gly Arg Ser  
 290 295 300  
 Trp Ala Ser Gln Leu Glu Leu Ile Asp Lys Tyr Pro Glu Tyr Val Phe  
 305 310 315 320  
 Val Ala Ser Gln Ala Gln Gln Phe Lys Trp Leu Lys Glu Asp Tyr Pro  
 325 330 335  
 Asp Leu Phe Ala Lys Ile Gln Lys Gln Ala Lys Lys Gly Arg Phe Leu  
 340 345 350  
 Pro Val Gly Gly Ala Trp Thr Glu Cys Asp Thr Asn Leu Pro Ser Gly  
 355 360 365  
 Glu Ser Leu Leu Arg Gln Phe Leu Leu Gly Gln Arg Phe Phe Leu Glu  
 370 375 380  
 His Phe Gly Ser Leu Ser Asp Thr Phe Trp Leu Pro Asp Thr Phe Gly

[0010]

[0011]

```

385          390          395          400
Tyr Ser Ala Gln Val Pro Gln Leu Cys Arg Leu Ala Gly Met Asp Arg
          405          410          415
Phe Leu Thr Gln Lys Leu Ser Trp Asn Asn Ile Asn Ser Phe Pro Asn
          420          425          430
Ser Thr Phe Asn Trp Val Ala Leu Asp Gly Ser Gln Val Leu Cys His
          435          440          445
Met Pro Pro Asn Asn Thr Tyr Thr Ser Met Ala Asn Phe Gly Asp Val
          450          455          460
Ser Arg Thr Gln Lys Gln Asn Lys Asn Leu Asp Thr Thr Arg Asn Ser
          465          470          475          480
Leu Met Leu Tyr Gly His Gly Asp Gly Gly Gly Pro Thr Ala Glu
          485          490          495
Met Leu Glu Lys Leu Arg Arg Cys Arg Gly Val Ser Asn Thr Val Gly
          500          505          510
Glu Leu Pro Pro Val Ile Gln Gly Gln Ser Val Thr Asp Phe Tyr Asn
          515          520          525
Glu Leu Leu Asp Gln Thr Asn Asn Gly Lys Asp Leu Val Thr Trp Val
          530          535          540
Gly Glu Leu Tyr Phe Glu Phe His Arg Gly Thr Tyr Thr Ser Gln Ala
          545          550          555          560
Gln Thr Lys Lys Gly Asn Arg Val Ser Glu Asn Leu Leu His Asp Val
          565          570          575
Glu Leu Leu Ala Thr Leu Ala Ser Ile Arg Asp Ser Ser Tyr Lys Tyr
          580          585          590
Pro Phe Ala Gln Leu Glu Ser Leu Trp Glu Asp Val Cys Leu Cys Gln
          595          600          605
Phe His Asp Val Leu Pro Gly Ser Cys Ile Glu Met Val Tyr Lys Asp
          610          615          620
Val Lys Lys Ile His Gly Arg Val Ile Asp Thr Ala Ser His Leu Ile
          625          630          635          640
Asp Lys Ala Ala Ser Ala Leu Gly Leu Ser Gly His Pro Ser Lys Asp
          645          650          655
Ser Phe Asp Cys Thr Pro Val Ala Leu Asn Thr Met Pro Trp Ser Arg
          660          665          670
Thr Glu Val Val Ala Val Pro Gln Pro His Trp Asp Ala Thr Val Glu
          675          680          685
Leu Ala Glu Gly Val Glu Ile Gln Glu Asp Ser Gly Asn Ala Leu Val
          690          695          700
Met Met Ser Glu Ser Gly Pro Val Val Thr Thr Gln Ser Val Asp Leu
          705          710          715          720
Phe Lys Ser Glu Asp Ala Tyr Ile Leu Glu Asn Ser Gln Val Lys Val
          725          730          735
Thr Ile Cys Lys Asp Asp Gly Thr Leu Thr Ser Ile Tyr Asp Lys Glu
          740          745          750
Asn Asp Arg Arg Val Leu Ser Gly Thr Gly Asn Arg Leu Val Leu Phe
          755          760          765
Asp Asp Gln Pro Leu Ser Trp Gln Ala Trp Asp Thr Glu Val Phe Ser
          770          775          780
Leu Gly Lys Lys Gln Tyr Ile Gly Ala Glu Asn Val Thr Arg His Ser
          785          790          795          800
Ile Val Ser Ser Gly Pro Leu Arg Ser Thr Val Ala Phe Thr Tyr Glu
          805          810          815
Phe Asn Lys Ser Val Val Thr Thr Glu Ile Ser Leu Asp Ala Asn Ser
          820          825          830
Pro Leu Val Thr Phe Asn Thr Arg Ala Asp Trp His Glu Thr Cys Lys
          835          840          845
Phe Leu Lys Val Glu Phe Pro Val Asp Val His Ser Glu Ser Ala Ser
          850          855          860
Tyr Glu Ser Gln Phe Gly Val Val Lys Arg Pro Thr His Tyr Asn Thr
          865          870          875          880
Ser Trp Asp Val Ala Lys Phe Glu Val Cys Cys His Lys Phe Ala Asp
          885          890          895
Leu Ser Glu Leu Asp Tyr Gly Val Ser Ile Leu Asn Asp Cys Lys Tyr
          900          905          910
Gly Phe Ala Thr His Gly Asn Leu Met Arg Leu Ser Leu Leu Arg Ala
          915          920          925
Pro Lys Ala Pro Asp Ala His Ala Asp Met Gly His His Glu Phe Lys
          930          935          940
Tyr Gly Val Leu Ala His Lys Gly Pro Leu Gly Ala Thr Thr Val Arg
          945          950          955          960
Ala Ala Tyr Asn Phe Asn Asn Pro Leu Arg Val Lys Tyr Val Gly Leu
    
```

Ser Glu Val Ser Thr Lys Gln Ala Phe Ser Leu Lys Gly Pro Ala Asn  
 965 970 975  
 980 985 990  
 Leu Val Leu Ser Gln Val Lys Arg Ala Glu Val Asp Arg Ser Lys Lys  
 995 1000 1005  
 Ser Thr Asn Val Ile Leu Arg Val Tyr Glu Ala Leu Gly Gly Arg Thr  
 1010 1015 1020  
 Arg Gly Lys Leu Val Ile Asp Leu Pro Asn Val Val Ser Val Thr Lys  
 1025 1030 1035 1040  
 Thr Cys Ala Leu Glu Tyr Ser Lys Glu Lys Gln Val Val Ala Lys Ser  
 1045 1050 1055  
 Glu Gly Val Thr Ser Val Asp Ile Ser Leu Arg Ala Phe Glu Val Ala  
 1060 1065 1070  
 Thr Tyr Lys Val Glu Leu Ala  
 1075

<210> 6  
 <211> 513  
 <212> PRT  
 <213> 斋藤曲霉

<400> 6  
 Met His Leu Pro Ser Leu Ser Leu Thr Ala Leu Ala Ile Ala  
 1 5 10 15  
 Ser Pro Ser Ala Ala Tyr Pro His Phe Gly Ser Ser Gln Pro Val Leu  
 20 25 30  
 His Ser Ser Ser Asp Thr Thr Gln Ser Arg Ala Asp Ala Ile Lys Ala  
 35 40 45  
 Ala Phe Ser His Ala Trp Asp Gly Tyr Leu Gln Tyr Ala Phe Pro His  
 50 55 60  
 Asp Glu Leu His Pro Val Ser Asn Gly Tyr Gly Asp Ser Arg Asn Gly  
 65 70 75 80  
 Trp Gly Ala Ser Ala Val Asp Ala Leu Ser Thr Ala Val Ile Met Arg  
 85 90 95  
 Asn Ala Thr Ile Val Asn Gln Ile Leu Asp His Val Gly Lys Ile Asp  
 100 105 110  
 Tyr Ser Lys Thr Asn Thr Thr Val Ser Leu Phe Glu Thr Thr Ile Arg  
 115 120 125  
 Tyr Leu Gly Gly Met Leu Ser Gly Tyr Asp Leu Leu Lys Gly Pro Val  
 130 135 140  
 Ser Asp Leu Val Gln Asn Ser Ser Lys Ile Asp Val Leu Leu Thr Gln  
 145 150 155 160  
 Ser Lys Asn Leu Ala Asp Val Leu Lys Phe Ala Phe Asp Thr Pro Ser  
 165 170 175  
 Gly Val Pro Tyr Asn Asn Leu Asn Ile Thr Ser Gly Gly Asn Asp Gly  
 180 185 190  
 Ala Lys Thr Asn Gly Leu Ala Val Thr Gly Thr Leu Ala Leu Glu Trp  
 195 200 205  
 Thr Arg Leu Ser Asp Leu Thr Gly Asp Thr Thr Tyr Ala Asp Leu Ser  
 210 215 220  
 Gln Lys Ala Glu Ser Tyr Leu Leu Asn Pro Gln Pro Lys Ser Ala Glu  
 225 230 235 240  
 Pro Phe Pro Gly Leu Val Gly Ser Asn Ile Asn Ile Ser Asn Gly Gln  
 245 250 255  
 Phe Thr Asp Ala Gln Val Ser Trp Asn Gly Gly Asp Asp Ser Tyr Tyr  
 260 265 270  
 Glu Tyr Leu Ile Lys Met Tyr Val Tyr Asp Pro Lys Arg Phe Gly Leu  
 275 280 285  
 Tyr Lys Asp Arg Trp Val Ala Ala Ala Gln Ser Thr Met Gln His Leu  
 290 295 300  
 Ala Ser His Pro Ser Ser Arg Pro Asp Leu Thr Phe Leu Ala Ser Tyr  
 305 310 315 320  
 Asn Asn Gly Thr Leu Gly Leu Ser Ser Gln His Leu Thr Cys Phe Asp  
 325 330 335  
 Gly Gly Ser Phe Leu Leu Gly Gly Thr Val Leu Asn Arg Thr Asp Phe  
 340 345 350  
 Ile Asn Phe Gly Leu Asp Leu Val Ser Gly Cys His Asp Thr Tyr Asn  
 355 360 365  
 Ser Thr Leu Thr Gly Ile Gly Pro Glu Ser Phe Ser Trp Asp Thr Ser  
 370 375 380  
 Asp Ile Pro Ser Ser Gln Gln Ser Leu Tyr Glu Lys Ala Gly Phe Tyr

[0012]

385 390 395 400  
 Ile Thr Ser Gly Ala Tyr Ile Leu Arg Pro Glu Val Ile Glu Ser Phe  
 405 410 415  
 Tyr Tyr Ala Trp Arg Val Thr Gly Gln Glu Thr Tyr Arg Asp Trp Ile  
 420 425 430  
 Trp Ser Ala Phe Ser Ala Val Asn Asp Tyr Cys Arg Thr Ser Ser Gly  
 435 440 445  
 Phe Ser Gly Leu Thr Asp Val Asn Ala Ala Asn Gly Gly Ser Arg Tyr  
 450 455 460  
 Asp Asn Gln Glu Ser Phe Leu Phe Ala Glu Val Met Lys Tyr Ser Tyr  
 465 470 475 480  
 Met Ala Phe Ala Glu Asp Ala Ala Trp Gln Val Gln Pro Gly Ser Gly  
 485 490 495  
 Asn Gln Phe Val Phe Asn Thr Glu Ala His Pro Val Arg Val Ser Ser  
 500 505 510  
 Thr

<210> 7  
 <211> 1810  
 <212> PRT  
 <213> 纤维化纤维微生物

<400> 7  
 Met Thr Arg Pro Leu Pro Pro Gly Arg Ala Val Ala Arg Ser Gly Ser  
 1 5 10 15  
 Gly Arg Ala Arg Pro Leu Gly Leu Val Leu Ala Ala Ala Leu Ala Val  
 20 25 30  
 Pro Leu Gly Val Pro Leu Ala Ala Pro Ala Gly Ala Leu Ala Ala Ala  
 35 40 45  
 Pro Ala Ala Ala Ala Glu Pro Gly Asp Phe Ser Ser Ser Phe Glu Ser  
 50 55 60  
 Gly Asp Pro Ala Ala Leu Pro Thr Thr Val Ala Glu Arg Asp Gly Ala  
 65 70 75 80  
 Pro Trp Gln Ala Asn Val Gly Ser Phe Thr Ala Gly Leu Pro Gly Ser  
 85 90 95  
 Val Leu Gly Gln Leu Lys Gly Val Thr Ala Ser Ala Gln Asn Leu Pro  
 100 105 110  
 Asn Glu Gly Ala Ala Asn Leu Ala Asp Gly Ser Ser Gly Thr Lys Trp  
 115 120 125  
 Leu Ala Phe Ala Ser Thr Gly Trp Val Arg Tyr Glu Phe Ala Glu Pro  
 130 135 140  
 Val Ser Phe Val Ala Tyr Thr Met Thr Ser Gly Asp Asp Ala Ala Gly  
 145 150 155 160  
 Arg Asp Pro Lys Thr Trp Thr Val Glu Gly Ser Asn Asp Gly Ser Thr  
 165 170 175  
 Trp Ala Ala Leu Asp Arg Arg Thr Asp Glu Asp Phe Pro Asn Arg Gln  
 180 185 190  
 Gln Thr Arg Thr Phe Glu Leu Glu Ala Pro Thr Ala Ala Tyr Thr Tyr  
 195 200 205  
 Leu Arg Leu Asn Val Thr Ala Asn Ser Gly Asp Ser Ile Val Gln Leu  
 210 215 220  
 Ala Gly Trp Asp Leu Ser Ala Asp Leu Ser Ala Gly Pro Ser Ala Ala  
 225 230 235 240  
 Pro Met Thr Thr Lys Val Gly Thr Gly Pro Arg Val Ser Phe Thr Asn  
 245 250 255  
 Lys Ala Gly Val Gly Phe Ser Gly Leu His Ser Leu Arg Tyr Asp Gly  
 260 265 270  
 Ser His Leu Ala Asp Gly Glu Thr Tyr Ala Thr Asn Val Leu Tyr Asp  
 275 280 285  
 Asp Val Asp Val Val Val Gly Glu Asp Thr Arg Leu Ser Tyr Thr Ile  
 290 295 300  
 Phe Pro Glu Leu Leu Asp Asp Leu Gln Tyr Pro Ser Thr Tyr Ala Ala  
 305 310 315 320  
 Val Asp Val Leu Phe Thr Asp Gly Thr Tyr Leu Ser Asp Leu Gly Ala  
 325 330 335  
 Arg Asp Ala His Glu Thr Val Ala Thr Ala Gln Ala Gln Gly Glu Gly  
 340 345 350  
 Lys Ile Leu Tyr Ala Asp Gln Trp Asn Ser Val Arg Val Asp Leu Gly  
 355 360 365  
 Asp Val Ala Glu Gly Lys Thr Val Asp Gln Val Leu Leu Gly Tyr Asp

[0013]

[0014]

```

370          375          380
Asn Pro Gly Gly His Ala Gly Thr Lys Phe Ala Gly Trp Leu Asp Asp
385          390          395          400
Val Glu Ile Thr Ala Glu Pro Ala Thr Ile Asp Gly Ser Ser Leu Ala
          405          410          415
Asn Tyr Val Asp Thr Arg Arg Gly Thr Leu Ala Ser Gly Ser Phe Ser
          420          425          430
Arg Gly Asn Asn Ile Pro Ala Thr Ala Thr Pro Asn Gly Phe Asn Phe
          435          440          445
Trp Thr Pro Tyr Thr Asn Ala Ser Ser Gln Ser Trp Leu Tyr Glu Tyr
          450          455          460
His Lys Ala Asn Asn Ala Asn Asn Lys Pro Val Leu Gln Gly Phe Gly
465          470          475          480
Ile Ser His Glu Pro Ser Pro Trp Met Gly Asp Arg Asn Gln Leu Thr
          485          490          495
Phe Leu Pro Ser Thr Ala Ser Gly Thr Pro Asp Ala Thr Leu Ser Thr
          500          505          510
Arg Gly Leu Glu Phe Asp His Ala Asp Glu Thr Ala Arg Pro Asp Tyr
          515          520          525
Tyr Gly Val Thr Phe Thr Asn Gly Ser Ala Ile Glu Ala Thr Pro Thr
          530          535          540
Asp His Gly Ala Val Leu Arg Phe Ser Tyr Pro Gly Ala Lys Gly His
545          550          555          560
Val Leu Val Asp Lys Val Asp Gly Ser Ser Lys Leu Thr Tyr Asp Gln
          565          570          575
Ala Thr Gly Thr Ile Ser Gly Trp Val Glu Asn Gly Ser Gly Leu Ser
          580          585          590
Val Gly Arg Thr Arg Met Phe Val Ala Gly Thr Phe Asp Arg Ser Pro
          595          600          605
Thr Ala Val Gly Thr Ala Ala Gly Asn Arg Ala Asp Ala Arg Phe Ala
          610          615          620
Thr Phe Glu Thr Ser Ser Asp Lys Thr Val Glu Leu Arg Val Ala Thr
625          630          635          640
Ser Phe Ile Ser Leu Asp Gln Ala Arg Lys Asn Leu Asp Leu Glu Val
          645          650          655
Thr Gly Lys Thr Phe Thr Glu Val Lys Ala Ala Ala Ala Gln Ala Trp
          660          665          670
Asn Asp Arg Leu Gly Val Ile Glu Val Glu Gly Ala Ser Glu Asp Gln
          675          680          685
Leu Val Thr Leu Tyr Ser Asn Leu Tyr Arg Leu Asn Leu Tyr Pro Asn
          690          695          700
Ser Gln Phe Glu Asn Thr Gly Thr Ala Gln Glu Pro Val Tyr Arg Tyr
705          710          715          720
Ala Ser Pro Val Ser Ala Thr Thr Gly Ser Ala Thr Asp Thr Gln Thr
          725          730          735
Asn Ala Lys Ile Val Asp Gly Lys Ile Tyr Val Asn Asn Gly Phe Trp
          740          745          750
Asp Thr Tyr Arg Thr Ala Trp Pro Ala Tyr Ser Leu Leu Tyr Pro Glu
          755          760          765
Leu Ala Ala Glu Leu Val Asp Gly Phe Val Gln Gln Tyr Arg Asp Gly
          770          775          780
Gly Trp Ile Ala Arg Trp Ser Ser Pro Gly Tyr Ala Asp Leu Met Thr
785          790          795          800
Gly Thr Ser Ser Asp Val Ala Phe Ala Asp Ala Tyr Leu Lys Gly Ser
          805          810          815
Leu Pro Thr Gly Thr Ala Leu Glu Ala Tyr Asp Ala Ala Leu Arg Asn
          820          825          830
Ala Thr Val Ala Pro Pro Ser Asn Ala Val Gly Arg Lys Gly Leu Gln
          835          840          845
Thr Ser Pro Phe Leu Gly Phe Thr Pro Glu Ser Thr His Glu Ser Val
          850          855          860
Ser Trp Gly Leu Glu Gly Leu Val Asn Asp Phe Gly Ile Gly Asn Met
865          870          875          880
Ala Ala Ala Leu Ala Glu Asp Pro Ala Thr Pro Glu Glu Arg Arg Glu
          885          890          895
Thr Leu Arg Glu Glu Ser Ala Tyr Phe Leu Glu Arg Ala Thr His Tyr
          900          905          910
Val Glu Leu Phe Asp Pro Glu Val Asp Phe Phe Val Pro Arg His Glu
          915          920          925
Asp Gly Thr Trp Ala Val Asp Pro Glu Thr Tyr Asp Pro Glu Ala Trp
          930          935          940
Gly Gly Gly Tyr Thr Glu Thr Asn Gly Trp Asn Phe Ala Phe His Ala

```

945 950 955 960  
 Pro Gln Asp Gly Gln Gly Leu Ala Asn Leu Tyr Gly Gly Lys Gln Gly  
 965 970 975  
 Leu Glu Asp Lys Leu Asp Glu Phe Phe Ser Thr Pro Glu Lys Gly Ala  
 980 985 990  
 Gly Asn Gly Gly Ile His Glu Gln Arg Glu Ala Arg Asp Val Arg Met  
 995 1000 1005  
 Gly Gln Trp Gly Met Ser Asn Gln Val Ser His His Ile Pro Trp Leu  
 1010 1015 1020  
 Tyr Asp Ala Ala Gly Ala Pro Ser Lys Ala Gln Glu Lys Val Arg Glu  
 1025 1030 1035 1040  
 Val Thr Arg Arg Leu Phe Val Gly Ser Glu Ile Gly Gln Gly Tyr Pro  
 1045 1050 1055  
 Gly Asp Glu Asp Asn Gly Glu Met Ser Ser Trp Trp Ile Phe Ala Ser  
 1060 1065 1070  
 Leu Gly Phe Tyr Pro Leu Gln Val Gly Ser Asp Gln Tyr Ala Val Gly  
 1075 1080 1085  
 Ser Pro Leu Phe Asp Lys Ala Thr Val His Leu Pro Asp Gly Asp Leu  
 1090 1095 1100  
 Val Val Asn Ala Glu Asn Asn Ser Val Asp Asn Val Tyr Val Gln Ser  
 1105 1110 1115 1120  
 Leu Ala Val Asp Gly Glu Ala Arg Thr Ser Thr Ser Leu Ser Gln Ala  
 1125 1130 1135  
 Asp Leu Ser Gly Gly Thr Thr Leu Asp Phe Val Met Gly Pro Glu Pro  
 1140 1145 1150  
 Ser Asp Trp Gly Thr Gly Glu Asp Asp Ala Pro Pro Ser Leu Thr Glu  
 1155 1160 1165  
 Gly Asp Glu Pro Pro Thr Pro Val Gln Asp Ala Thr Thr Ala Gly Leu  
 1170 1175 1180  
 Gly Thr Thr Thr Val Ala Asp Gly Asp Ala Thr Thr Ser Ala Ala Ala  
 1185 1190 1195 1200  
 Leu Thr Asp Asn Thr Ser Gly Thr Arg Thr Thr Phe Ala Thr Thr Thr  
 1205 1210 1215  
 Pro Ser Ile Thr Trp Ala Gly Asn Gly Ile Arg Pro Thr Val Gly Ser  
 1220 1225 1230  
 Tyr Thr Leu Thr Ser Gly Ala Ser Gly Thr Ala Ser Pro Ser Ala Trp  
 1235 1240 1245  
 Thr Leu Glu Gly Ser Asp Asp Gly Glu Thr Trp Thr Thr Leu Asp Glu  
 1250 1255 1260  
 Arg Ser Gly Glu Gln Phe Arg Trp Ala Leu Gln Thr Arg Pro Phe Thr  
 1265 1270 1275 1280  
 Val Ala Glu Pro Thr Ala Phe Ala Arg Tyr Arg Val Thr Val Thr Ala  
 1285 1290 1295  
 Thr Ser Gly Ser Gly Ala Leu Ser Leu Ala Glu Val Glu Leu Leu Ala  
 1300 1305 1310  
 Asp Pro Lys Glu Ser Gly Ala Glu Glu Leu Thr Leu Ser Ala Ala Pro  
 1315 1320 1325  
 Asp Arg Asp Gly Val Thr Gly Arg Glu Val Ser Gly Ser Phe Ala Thr  
 1330 1335 1340  
 Leu Thr Gly Val Glu Gly Asp Val Ala Ala Leu Asp Val Gln Val Ala  
 1345 1350 1355 1360  
 Phe Gly Asp Gly Ser Glu Pro Val Ala Gly Thr Leu Arg Ala Gly Ala  
 1365 1370 1375  
 Phe Gly Gly Tyr Ala Val Asp Ala Ala His Thr Trp Thr Ala Pro Gly  
 1380 1385 1390  
 Val Tyr Pro Val Thr Val Thr Val Ser Gly Glu Gly Ile Glu Thr Val  
 1395 1400 1405  
 Ser Ala Ser Ser Tyr Val Ser Val Ser Leu Leu Arg Glu Gly Ser Leu  
 1410 1415 1420  
 Leu Ala Ala Tyr Asp Asn Val Cys Ile Gly Asp Ala Gly Thr Thr Val  
 1425 1430 1435 1440  
 Gly Ser Cys Asp Gly Gln Gly Val Phe Phe Asp Arg Ala Gln Leu Ala  
 1445 1450 1455  
 Ala Lys Gly Phe Val Gln Gly Glu Arg Ala Thr Val Pro Gly Thr Asp  
 1460 1465 1470  
 Leu Ala Phe Asp Val Pro Ala Val Pro Ala Gly Gln Pro Asp Asn Ala  
 1475 1480 1485  
 Thr Gly Asp Gly Gln Thr Ile Glu Leu Asp Val Pro Ala Asp Ala Glu  
 1490 1495 1500  
 Gln Leu Ser Val Ile Gly Thr Gly Thr Glu Lys Asn Gln Gln Ala Thr  
 1505 1510 1515 1520  
 Gly Thr Leu Thr Phe Asp Asp Gly Ser Thr Gln Pro Ile Asp Leu Ser

[0015]

1525 1530 1535  
 Phe Gly Asp Trp Ser Gly Ala Ala Arg Asn Pro Val Phe Gly Asn Ile  
 1540 1545 1550  
 Pro Val Ala Val Thr Asp Ser Arg Leu Arg Gly Gly Ser Pro Gln Thr  
 1555 1560 1565  
 Gly Thr Pro Ala Ala Phe Phe Ala Thr Ala Pro Ile Thr Leu Pro Glu  
 1570 1575 1580  
 Gly Lys Arg Pro Val Ser Leu Thr Leu Pro Asp Gln Pro Gly Glu Leu  
 1585 1590 1595 1600  
 Ser Arg Asp Gly Arg Ile His Val Val Ala Val Ala His Asp Gly Thr  
 1605 1610 1615  
 Phe Ala Glu His Pro Ala Leu Glu Val Thr Ala Ala Glu Gly Val Thr  
 1620 1625 1630  
 Leu Ala Val Gly Gln Thr Ser Asp Val Ala Leu Ala Gln Val Ala Gly  
 1635 1640 1645  
 Gly Arg Glu Gly Ala Asp Leu Arg Ala Ala Val Thr Trp Gly Asp Gly  
 1650 1655 1660  
 Ser Asp Val Ala Ala Gly Ala Val Thr Asp Gly Ser Val Ser Gly Ser  
 1665 1670 1675 1680  
 His Ala Tyr Thr Ala Ala Gly Thr Tyr Thr Ala Tyr Val Val Val Asp  
 1685 1690 1695  
 Asp Gly Trp Thr Ser Gln Val Val Glu Val Pro Val Thr Val Thr Glu  
 1700 1705 1710  
 Ala Glu Pro Ala Leu Ala Val Asp Val Thr Val Ser Thr Arg Cys Leu  
 1715 1720 1725  
 Ala Gly Lys Ala Tyr Val Ala Val Arg Ala Glu Asn Gly Glu Asp Val  
 1730 1735 1740  
 Pro Leu Ala Ile Arg Leu Val Thr Pro Phe Gly Thr Lys Glu Val Ala  
 1745 1750 1755 1760  
 Ala Val Ala Pro Gly Ala Asn Ala Tyr Ser Phe Ala Thr Arg Val Thr  
 1765 1770 1775  
 Ala Val Glu Ala Gly Thr Val Thr Val Glu Ala Thr Arg Gly Thr Gly  
 1780 1785 1790  
 Asp Glu Glu Val Thr Ala Ser Ile Gln Ala Asp Tyr Ala Ala Val Thr  
 1795 1800 1805  
 Cys Gly  
 1810

[0016]

<210> 8  
 <211> 282  
 <212> PRT  
 <213> 米曲霉

<400> 8  
 Gly Leu Thr Thr Gln Lys Ser Ala Pro Trp Gly Leu Gly Ser Ile Ser  
 1 5 10 15  
 His Lys Gly Gln Gln Ser Thr Asp Tyr Ile Tyr Asp Thr Ser Ala Gly  
 20 25 30  
 Glu Gly Thr Tyr Ala Tyr Val Val Asp Ser Gly Val Asn Val Asp His  
 35 40 45  
 Glu Glu Phe Glu Gly Arg Ala Ser Lys Ala Tyr Asn Ala Ala Gly Gly  
 50 55 60  
 Gln His Val Asp Ser Ile Gly His Gly Thr His Val Ser Gly Thr Ile  
 65 70 75 80  
 Ala Gly Lys Thr Tyr Gly Ile Ala Lys Lys Ala Ser Ile Leu Ser Val  
 85 90 95  
 Lys Val Phe Gln Gly Glu Ser Ser Ser Thr Ser Val Ile Leu Asp Gly  
 100 105 110  
 Phe Asn Trp Ala Ala Asn Asp Ile Val Ser Lys Lys Arg Thr Ser Lys  
 115 120 125  
 Ala Ala Ile Asn Met Ser Leu Gly Gly Gly Tyr Ser Lys Ala Phe Asn  
 130 135 140  
 Asp Ala Val Glu Asn Ala Phe Glu Gln Gly Val Leu Ser Val Val Ala  
 145 150 155 160  
 Ala Gly Asn Glu Asn Ser Asp Ala Gly Gln Thr Ser Pro Ala Ser Ala  
 165 170 175  
 Pro Asp Ala Ile Thr Val Ala Ala Ile Gln Lys Ser Asn Asn Arg Ala  
 180 185 190  
 Ser Phe Ser Asn Phe Gly Lys Val Val Asp Val Phe Ala Pro Gly Gln  
 195 200 205  
 Asp Ile Leu Ser Ala Trp Ile Gly Ser Ser Ser Ala Thr Asn Thr Ile

210 215 220  
 Ser Gly Thr Ser Met Ala Thr Pro His Ile Val Gly Leu Ser Leu Tyr  
 225 230 235 240  
 Leu Ala Ala Leu Glu Asn Leu Asp Gly Pro Ala Ala Val Thr Lys Arg  
 245 250 255  
 Ile Lys Glu Leu Ala Thr Lys Asp Val Val Lys Asp Val Lys Gly Ser  
 260 265 270  
 Pro Asn Leu Leu Ala Tyr Asn Gly Asn Ala  
 275 280

<210> 9  
 <211> 403  
 <212> PRT  
 <213> 米曲霉

<400> 9  
 Met Gln Ser Ile Lys Arg Thr Leu Leu Leu Leu Gly Ala Ile Leu Pro  
 1 5 10 15  
 Ala Val Leu Gly Ala Pro Val Gln Glu Thr Arg Arg Ala Ala Glu Lys  
 20 25 30  
 Leu Pro Gly Lys Tyr Ile Val Thr Phe Lys Pro Gly Ile Asp Glu Ala  
 35 40 45  
 Lys Ile Gln Glu His Thr Thr Trp Ala Thr Asn Ile His Gln Arg Ser  
 50 55 60  
 Leu Glu Arg Arg Gly Ala Thr Gly Gly Asp Leu Pro Val Gly Ile Glu  
 65 70 75 80  
 Arg Asn Tyr Lys Ile Asn Lys Phe Ala Ala Tyr Ala Gly Ser Phe Asp  
 85 90 95  
 Asp Ala Thr Ile Glu Glu Ile Arg Lys Asn Glu Asp Val Ala Tyr Val  
 100 105 110  
 Glu Glu Asp Gln Ile Tyr Tyr Leu Asp Gly Leu Thr Thr Gln Lys Ser  
 115 120 125  
 Ala Pro Trp Gly Leu Gly Ser Ile Ser His Lys Gly Gln Ser Thr  
 130 135 140  
 Asp Tyr Ile Tyr Asp Thr Ser Ala Gly Glu Gly Thr Tyr Ala Tyr Val  
 145 150 155 160  
 Val Asp Ser Gly Val Asn Val Asp His Glu Glu Phe Glu Gly Arg Ala  
 165 170 175  
 Ser Lys Ala Tyr Asn Ala Ala Gly Gly Gln His Val Asp Ser Ile Gly  
 180 185 190  
 His Gly Thr His Val Ser Gly Thr Ile Ala Gly Lys Thr Tyr Gly Ile  
 195 200 205  
 Ala Lys Lys Ala Ser Ile Leu Ser Val Lys Val Phe Gln Gly Glu Ser  
 210 215 220  
 Ser Ser Thr Ser Val Ile Leu Asp Gly Phe Asn Trp Ala Ala Asn Asp  
 225 230 235 240  
 Ile Val Ser Lys Lys Arg Thr Ser Lys Ala Ala Ile Asn Met Ser Leu  
 245 250 255  
 Gly Gly Gly Tyr Ser Lys Ala Phe Asn Asp Ala Val Glu Asn Ala Phe  
 260 265 270  
 Gln Gln Gly Val Leu Ser Val Val Ala Ala Gly Asn Glu Asn Ser Asp  
 275 280 285  
 Ala Gly Gln Thr Ser Pro Ala Ser Ala Pro Asp Ala Ile Thr Val Ala  
 290 295 300  
 Ala Ile Gln Lys Ser Asn Asn Arg Ala Ser Phe Ser Asn Phe Gly Lys  
 305 310 315 320  
 Val Val Asp Val Phe Ala Pro Gly Gln Asp Ile Leu Ser Ala Trp Ile  
 325 330 335  
 Gly Ser Ser Ser Ala Thr Asn Thr Ile Ser Gly Thr Ser Met Ala Thr  
 340 345 350  
 Pro His Ile Val Gly Leu Ser Leu Tyr Leu Ala Ala Leu Glu Asn Leu  
 355 360 365  
 Asp Gly Pro Ala Ala Val Thr Lys Arg Ile Lys Glu Leu Ala Thr Lys  
 370 375 380  
 Asp Val Val Lys Asp Val Lys Gly Ser Pro Asn Leu Leu Ala Tyr Asn  
 385 390 395 400  
 Gly Asn Ala

[0017]

<210> 10

<211> 1272  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 编码碱性蛋白酶的密码子优化序列

<400> 10  
 ggatccatgc agtecatata gcaactctg ctgctgctgg gagccattct gcccgccgig 60  
 ctgggagccc ccgttcagga gacccgacga gccgcccaga agctccccgg caagtacatt 120  
 gtcaccttca agccctggtat cgacgagget aagattcagg agcacaccac ttgggccacc 180  
 aacatccatc agcgatccct cgagcgacga ggagccaccg gcggtgacct gectgtggga 240  
 atcgagcgaa actacaagat taacaagtfc gccgcttac ctggatcttt tgacgatgcc 300  
 accatcgagg agattcgaaa gaacgaggac gtcgcttacg tggaggaaga ccagatctac 360  
 tacctcgaig gtcgaccac tcagaagtec gctccttggg gccctggctc catctctcac 420  
 aagggacagc agtcgactga ctacatctac gatacctccg ctggcgaggg tacttacgcc 480  
 taegtctggg actccggtgt taacgtcgat cagcaggagt ttgagggagc agcctctaag 540  
 gcctacaacg ccgctggagg ccagcatgtg gactctatcg gacacggcac ccattgttcc 600  
 ggtactatlg ccggaagac ctacggcatc gccaaagaagg cttctattct ctccgtgaag 660  
 gtttccagg gagagctctc ttccacctct gtcatectgg acggctttaa ctgggccgct 720  
 aacgataitg tgcctaagaa gcgaacctcg aaggccgcta tcaacatgic cctcgggtga 780  
 ggctactcta aggccttcaa cgacgtgtti gagaacgctt ttgagcaggg tctccigtct 840  
 gttgtggctg ctggtaacga gaactctgac gctggacaga cctccccctc ttctgtcct 900  
 gatgccatca ctgtggccgc tatteagaag tccaacaacc gagcttctgt ctccaacttt 960  
 ggcaaggigg ttgacgtttt ccccccgga caggataacc tctctgcttg gattggctcc 1020  
 tcttcggcca ccaacactat ctccggcacc tccatggcca ctccccacat tctcggctcg 1080  
 tccctctacc tggctgctct ggagaacctg gacggacctg ccgctgttac caagcgaatc 1140  
 aaggagctgg ctactaagga cgtcgtgaag gatgtcaagg gttctcctaa cctgctcgcc 1200  
 tacaacggca acgctctcgg ccgcccagga caccaccacc atcaccatca ccaccatcat 1260  
 tgataacctt gg 1272

[0018]

<210> 11  
 <211> 20  
 <212> PRT  
 <213> 人工序列

<220>  
 <223> 来自刀豆(Jack bean)甘露糖苷酶的肽

<400> 11  
 Asn Lys Ile Pro Arg Ala Gly Trp Gln Ile Asp Pro Phe Gly His Ser  
 1 5 10 15  
 Ala Val Gln Gly  
 20

<210> 12  
 <211> 7  
 <212> PRT  
 <213> 人工序列

<220>  
 <223> 碱性蛋白酶底物

<400> 12  
 Ile Gln Asn Cys Pro Leu Gly  
 1 5

AAQQGASRPGPRDAQAHFGRPRAVPTQCDVPPNSRFDCAFDKAITQEQCEARGCCYIPAKQGLQ  
GAQMGQPWCFPPSYPSYKLENLSSSEMGYTATLRTTPTFFPKDILTLRLDVMETENRLHFTIKD  
PANRRYEVPLETPHVHSRAPSPLYSVEFSEEPFGVIVRRQLDGRVLLNTTVAPLFFADQFLQLSTSLPS  
QYITGLAEHLSPLMLSTSWTRITLWNRDLAPTGANLYGSHPFYLALEDGGS AHGVFLLNSNAMDV  
VLQSPALSWRSTGGILDVYIFLGPEPKSVVQQYLDVVGYPFMPYWG LGFHLCRWGYSSTAITRQ  
VVENMTRAHFPLDVQWNDLDYMDSRDFTFNKDGFRDFPAMVQELHQGGRRYMMIVDPAISSS  
GPAGSYRPHYDEGLRRGVFITNETGQPLIGKVWPGSTAFPDTNPTALAWWEDMVAEFHDQVPFD  
GMWIDMNEPSNFIRGSEDGCPNNELENPYVPGVVG GTLQAATICASSHQFLSTHYNLHNLVGLTE  
AIASHRALVKARGTRPFVISRSTFAGHG RYAGHWTGDVWSSWEQLASSVPEILQFNLLGVPLVGAD  
VCGFLGNTSEELCVRWTQLGAFYPFMRNHNSLLSLPQEPYSFSEPAQQAMRKALTRYALLPHLYTL  
FHQAHVAGETVARPLFLEFPKDSSTWTVDHQLLWGEALLITPV LQAGKAEVTGYFPLGTWYDLQTV  
PVEALGSLPPPPAAPREPAIHSEGQWVTLPA PLDTINVHLRAGYIIP LQGPGLTTTESRQQPMALAV  
ALTKGGEARGELFWDDGESLEVLERGAYTQVIFLARNNTIVNELVRVTSEGAGLQLQKVTVLGVATA  
PQQVLSNGVPVSNFTYSPDTKVL DICVSLLMGEQFLVSWC\* (SEQ ID NO:1)

图1



ACTGAGCGTTTATCTGGATGCATTTGATCCGGCTAATCCGGGTGAACCGTTTGTACCCTCCGCTGCCGACCACCG  
GTAGCAGTTGGACCGCAGATGGCACAGCCACCGTTGTTCTGCCGAAACCGTGCAGGGCACCACATGAAGTTTTGT  
TCGTCTGAGCACCAGAACCGTATGCAGATCATCCGTATGTTGCAAATCTGGATAGCCTGACCTTTCACCCGGTGGTC  
CGACCAGCGTTTGGTTGAAAGCGAAGCCTGGACCAGCAATCTGGTCTGGCCTGAAAAATGAATCTTCTACCTG  
GACCTCTGGTCCGGTTACAAATGTGGGTGGCACCCTGATGGCGATTGGCTGGCATATGGCGAAATTGATCTGGGC  
AGCGCAGCACTGGATCAGCTGCTGTGCATTATGTTCAATTCCTAATCGCTCTGGTCTAATTCTGCACTGCTGTG  
TATCTGGATGCCCTTTCATCCGGCAAATCCGGGTGAACCGTTTGTGACAGTCCCGCTGGCAAATACCGGTAGCTCTG  
GACCACCGATGGTACTGCAAGTTGTGGATCTGCCGTCTACCCTCTGGTAAACATCAGGTTTGGGTTCTGTCTCTA  
CCGAAAGCATATGECGATCATCCGTATGTGGCAAATCTGGATTCTATGCGCTTTTACCAGATGCATATGATGTTGAAG  
TTCTCCGACCGATACAGCAGCACTGGCAGCCGTTGTTGATGCAGCAGGTACACCGGAAGCAGAAATTGCACGTTAT  
GGTCTGATTGATGCCCGTGTTCCTTACCCTGAACTGGCAGCAGCACGTAGCGTTCTGGCCGATGCCGGTGCAACACA  
GGCACAGCAGATGAACGTGCTCGTCTGGGTCTGGCAACCGATCAGCTGGTCCGGCAGAACGTGCTGCTG  
GAAAATCTGGTTGCCAGCGCAGAAGCACTGACCGACGAAGGTTATTCCTCCGAAAGCTGGCAGGCATTTCTACCG  
CACTGGCTGCTGCAACCGGCACCCCTGGATGATGCAGCAGCATCTGATGAAGCACTGCATGATGCACGTCTGGCGCT  
GCAGGGTGCAGTTGATGCACTGGAAGAACCAGCAGATGTTGTTCTGGTTGAAGTTGAAGTTTCTCCGCGTTGTCTG  
GCAGGTAACCGTATGTTGCCGTTCTGTCAGTTAATGTTCTGATGCAGCCGTTGATGTTGAACTGGCAAGCTCTCT  
GGGCACCCGTAGCTTTGTTGGTGTGGCACCAGGTTGCGAGCGCATATCAGAGCTTTCAGCCCGTAGCGCAACCGGT  
GATCTGGATGTTACCCTGACCGCAACCGGTGCAGATGGTACTCAGACCCTTGAACAGGTTGTGACCCTCCGAGCT  
GTAGCTAATAA [SEQ ID NO:2]

图2A (续)

**ALAVVGLAPATAASA** APEPPSADYASLVDVDFVGTGDFGNDMPAAQAPNGLAKVNPRTT  
 PGRNNTGYDYAQS KISGFTHTNL DGVGGSGGGDLLVPTSGSYTARFGTGT YAHPPSH  
 DDEDAGPGFYSVGLGNVAGTDG AITGAPGTIEAEVAAATRS GVVHRYAFFPAGSTPSLVVD  
 LETNNTSRRSSSVQVETRADGTVELSGQVTGYFYNAAYTLYYTARTLQFPATVQTWGDDD  
 RLVDATAQDGVDTGAILTFDPADAGEIGLQVTLSPVSVSEQARI DQQVELGDLSEFDAIRD  
 RTRAENWATLGRVAIDASTATDPTGELQRLFYTHLYRMEAMP MNATSTSGTYRGVDGAV  
 HAAQGFTYYDSWATWDDFRKFSVIAYIDPALYRDMVQSLVYL FADAEATGTGGSLGGFV  
 HSVPTVRWERSSSVVVADAIKGFDFGFDRLDEAYFALQRLVGGQYSADELRRGYVAGNPGA  
 SVQSGYDQYGLSVIADDELGLTEEAETLREQASWFI EKLTTPGAWTAADGTQVGLLT PRA  
 ADGSWQSADHAKFEAAGLYQGTLLWQYHWHYDAYDMDALVEAMGGHEAARLGMRRHMFGEHA  
 PDDGKAMLHNSANEIDLQAPYLFNYTGEPSLTQKWARAIYTKETWNR YIATGSSSAVPS  
 GGGEFTPPLKTKVYRLDFRGM LPTMDNDAGTMSTMEVAAAVGLFPVTAGSSQFQVGSFF  
 EDSTTIYDDGSAFTVTADGVSEDAFYVQSATLDGATFGNTWVDYATV VGGADLAFRMG  
 EQPSDWGTD TAPAFSMSTATDEPAEGPRVSAEPTTVQTDG GGDALDATVTLTLDGARLAA  
 PAGTDLVTS GAASVVGLPDGVTAAVTVASPTALT VSLTGTASADARFFVHLRDAALADG  
 VAAASLQGGQVSVRSP LRLSVASAERDALAALVDDAVLVRHGNYSSVTFDRFPSTALTKA  
 QEALGDEAATSIALRFAADRLGAAADALDLTGGGYRTLEAEQSEAWSGGELKNEANSSS  
 GNLGGVRSVSWVQYRDMTFETAAGDTPPRFLTVRYDTSFAPTDT PSTVRVHAGDVSGPV  
 VATVDLKTSGWNGKYTEVTAELGDVQALVDAQVVT FELLAPSGRSWVGNFDWFRFSAED  
 PAAPGQPGESPTVTIEAEDWTASSGRGLKKESS TWTSGPVTNVGGTADGDWLAYGEVDL  
 GELPLGELSVHYVHNSNRSGNNSALS VYLDAFDPANPGE PFVTVPLPTTGSSWTADGTA  
 TVVLPETVQGTHEVFVRLSTEPYADHPYVANLDSLTFAPGGPTSVVVESEAWTNSGRG  
 LKNESS TWTSGPVTNVGGTADGDWLAYGEIDLGSAALDQLSVHYVHNSNRSGRNSALS  
 YLDAFDPANPGE PFVTVPLANTGSSWTTDGTA VVDLPSTVRGKHQVWVRLSTEPYADHP  
 YVANLDSMRFFT DAYDVEVPPTDTAALAAVVDAAGTPEAEIARYGRI DARVFTRELAAA  
 RSVLADAGATQAQADERARRLGLATDQLVPAERRRLENLVASAEALTDEGYSPESWQAF  
 RTALAAATGTLDDAAASDEALHDARLALQGAVDAL EEPADVVLVEVEVSPRCLAGKPYV  
 AVRAVNVSDAAVDVELASSLGTRSFVGVAPGASAYQSEFAAR SATGDLDVTVTATGADGT  
 QTVEQVTVFVSCS (SEQ ID NO:3)

图2B

APEPPSADYASLVDVVFVGTGEGDFGNDMPAAQAENGLAKVNPRTT PGRNNTGYDYAQSKI SGF  
 THTNLDGVGSSGGGDLVVPVPTSGSYTPARPGTGTYYAH PFSHDDDEDAGPGFYSVGLGNVAGTD  
 GAITGAPGTIEAEVAAATRSQVHRYAFPAGSTPSLVVDLETNNTSRRSSSVQVETRADGTVE  
 LSGQVTGYFYNAAYTLYYTARTLQFATVQVQWGGDDRLVDATAQDGVDTGAILTFDPADAGEI  
 GLOVTLSPVSVQEARIDQQVELGDLSPDAIRDRTRAEWNATLGRVAIDASTATDPTGELQRL  
 FYTHLYRMFAMPMNATSTSGTYRGVDGAVHAAQGFYYDSWATWDDFRKFSVIAYIDPALYR  
 DMVQSLVYLEADAENATGTGGGLGGFVHVSVPVTRWERSVVVVADAIKGFEDGDFDRLDEAYPAL  
 QRLVGGQYSADELRRGYVAGNFGASVQRGYDQYGLSVIADELGLTTEEAETLREQASWPIEKLT  
 KPGAWTAADGTQVGLLT PRAADGGSWQSADHAKFEAAGLYQGT LWQYHWYDAYDMDALVEAMG  
 GHEAARLQMRHMFGEHAPDDGKAMLHNSANEIDLQAPYLENYTGEPSLTQKWARAIYTKETW  
 NRYIATGSSSAVPSGGGEPTPPLKTKVYRLDPRGMLPTMDNDAGTMSTMFVAAA VGLFPVTA  
 GSSQFQVGSPPFDSTTITYYDGS AFTVTADGVSEDAFYVQSATLDGATFGNTWVDYATVVG  
 ADLAFRMGEQPSDWGTDTPAFMSMSTATDEPAEGPRVSAEPTTVQTDGGALDATVTLTLDG  
 ARLAAPAGTDLVTSGAASVVG LFDGVTA AVTVASPTALT VSLTGTASADARFFVHLRDAALA  
 DGVAASLQGGQVSVRSPLRLSVASAERDALAALVDDAVLVRHGNYSVTFDRFSTALT KQA  
 EALGDEAATSIALRFAADRLGAAADALDLTGGGYRTLEAEQSEAWSSGGELKNEANSSSSGNLG  
 GVRSGSWVQYRDMTFETAAGDTPPRFLTVRYDTSFAPTDTPSTVRVHAGDVS GPVVATVDLK  
 GTSGWKGYTEVTAELGDVQALVDAQVVT FELLAPSGRSWVGNFDWFRFSAEDPAAPGQPGES  
 PTVTIEAEDWTASSGRGLKKESSWTSGFVTNVGGTADGDWIA YGEVDLGELPLGELSVHYV  
 HNSNRSGNNSALSVYLD AFDPANPGEPFVTVPLPTT GSSWTADGTATVVL PETVQGTHEV FV  
 RLSTEPYADHPYVANLDSLTFAPGGPTSVVVESEAWT SNSGRGLKNESSWTSGFVTNVGGT  
 ADGDWLAYGEIDLGSAALDQLSVHYVHNSNRSGRNSALSVYLD AFDPANPGEPFVTVPLANT  
 GSSWTTDGTAVVDLPSTVRGKHQVWVRLSTEAYADHPYVANLDSMRFFTDAYDVEVPPTDTA  
 ALAAVVDAAGTPEAEIARYGRIDARVFTRELA AARSVLADAGATQAQADERARRLGLATDQL  
 VPAERRRLENLVAEALTDGYSPE SWQAFRTALAAATGTLDDAAASDEALHDARLALQGA  
 VDALEEPADVVLVEVEVSPRCLAGKPYAVRAVNVSDAAVDVELASSLGT RSVFGVAPGASA  
 YQSFAARSATGDLVTVTATGADGTQVVEQVVTVPSCS (SEQ ID NO:4)

图2C

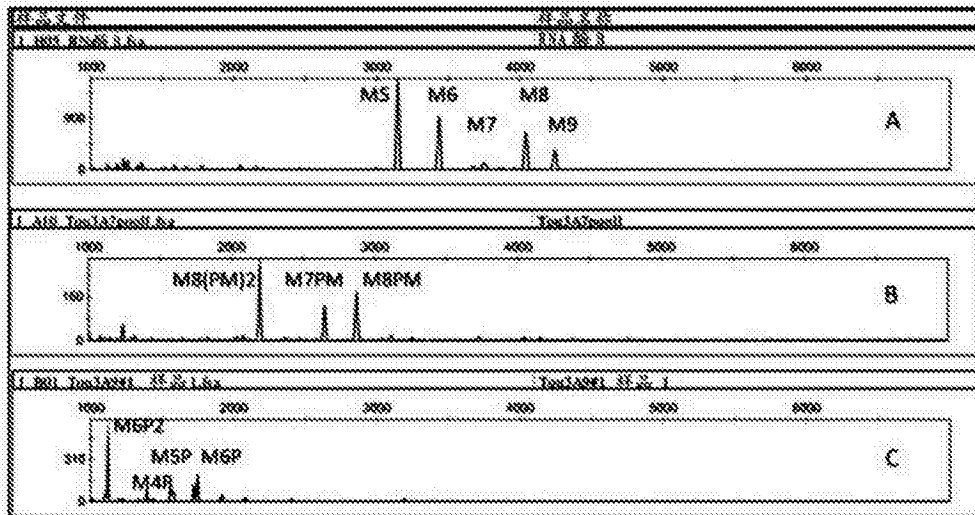


图3A

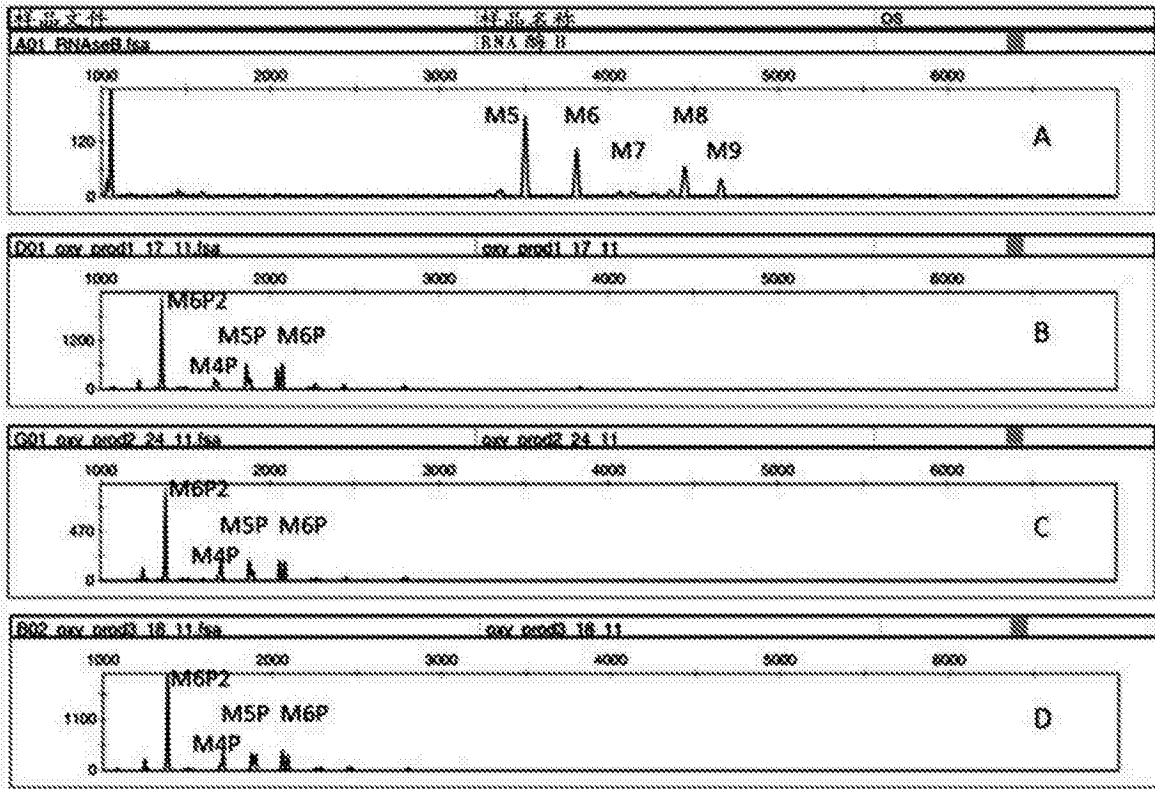


图3B

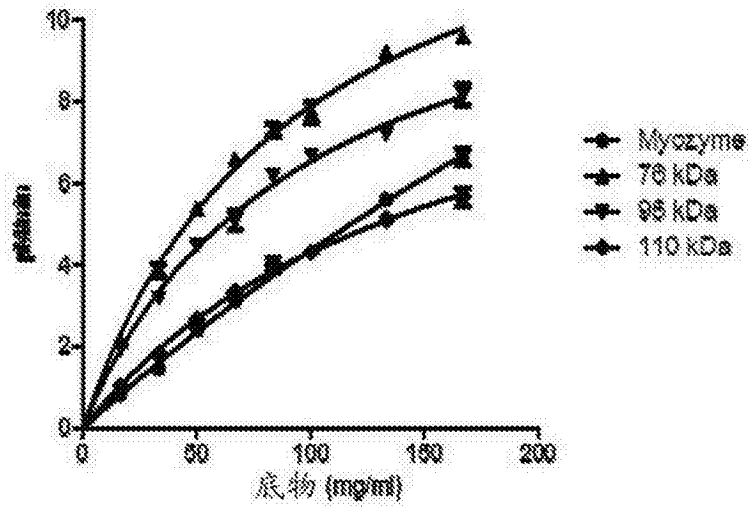


图4

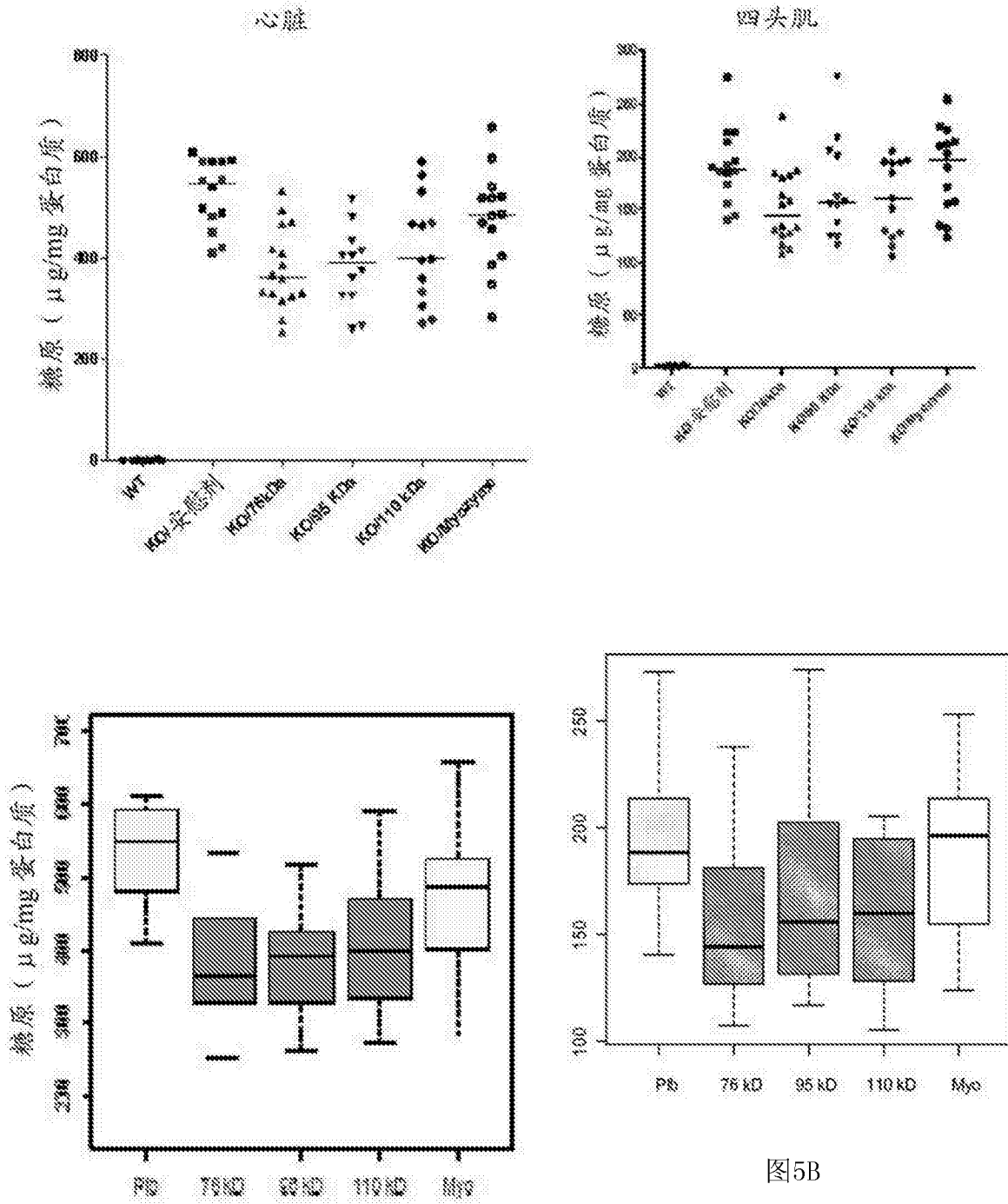


图5A

图5B

```

MYSHFNNEPVAKRVNNLFTDRLRQFTSDGEYRSLNLPAFYERERLDGKNHVAIETAVSD
LRRPLFKDALKEADGHWKEFAKKGSEYGFPSWATHWFKIQVCVPEWKKNYKKGDIVVFNW
NLNCEGLVFSSESSEALIGLSGEEERREWPI PDNWFEDGKCHTFYIEASCNCFGNATGSSIQ
PPSDNRYFRLDSADLVVINSEARHLFVDFWIIGDAAREFPEDSWQKALDVANKIMDAF
DPENPDESLAEGRIKLAKKEYLGDTTKAYKQQLPEADGLVYALGNCHIDTAWLWFFAETRRK
AGRSWASQLELIDKYPEYVFFVASQAQQFKWLKEDYDLEFAKIQKQAKKGRFLPVGGAWTE
CDTNLPSGESLLRQFLLLGQRFPLEHFGSLSDTFWLPDTEFGYSAQVPLCRLAGMDRFLTQ
KLSWNNINSPNSFTNWVALDGSQVLCHMFPNNTYTSMAFNGDVSRTQKQKNLDTTRNS
IMLYGHGDGGGGPTAEMLEKLRRCRGVSNVVGELFPVVIQGGQSVTDFYNELLDQTNNGKDL
VTWVGELYFEPHRCGTYSQAQTKKGNRVSENLLHDVELLATLASTRDSYKYPPFAQLES
WEDVCLCQPHDVLPGSCIEMVYKDVKKIHGRVIDTASHLIDKAASALGLSGHPSKDSFDC
TPVALNTMPWSRTEVAVPQPHWDATVELAEGVEIQEDSGNALVMMSSESGPVVTTQSVDL
FKSEDAIILENSQVKVTICKDDGFLTSIYDKENDRRVLSGTGNRLVLFDDQPLESWQAWDT
EVFSLGKKQYIGAENVTRHSIVSSGPLRSTVAFTYEFNKSVVTEISLDANSPLVTFNTR
ADWHETCKFLKVEFPVDVHSESASYESQFGVVKRPTHYNTSWDVAKFEVCCCHKFADLSEL
DYGVSILNDCKYGFATHGNIMRLSLLRAFKAFDAHADMGHHEFKYCVLANKGFLGATTVR
AAYNFNNPLRVKYVGLSEVSTKQAFSLKGPANLVLSQVKRAEVDRSKKSTNVI LRVYEAL
GGRTRGKLVIDLNVVSVTKTCALEYSKEKQVAKSEGVTSEVDISLRAFEVATYKVELA
(SEQ ID NO:5)

```

图6

```

1 mhlpalalal talaiaspaa ayphfgessqp vhhsssdttq sradaikaaf shawdgylyg
61 afphdelhgv sngygdarng wqasavdals tavimnati vngildhvgk idyxtnttv
121 slfettiryl ggmisgydli kgpvedivqn eskidvlltq sknladvlkf afdtpsgvpy
181 nnlnitagnn dgaktnglav tglalewtr ledltgdttty adlsqkaesy llnpgpkxae
241 pfpglvgsni nisngqftda qvswnggdds yyeylikmyv ydpkrfglyk drwvaaqst
301 mqhlaashpaa rpditflasy nngtlglssq hltcfdgqaf llggtvinrt dfinfgldly
361 sqchdtynst ltgigpafsa wdtedipssq qalyekagfy itsgayilrp eviesfyyaw
421 rvtggetyrd wiwsafsavn dycrtaagfa gltdvnaang gerydnqesf lfaevmkysy
481 mafaedaawq vqpgsgnqfv fteahpvrv sst (SEQ ID NO:6)

```

图7

**MTRPLPPGRAVARSGSGRRARPLGLVLAALAVPLGVPLAAPAGALAAAPAAAAPGDFSS**  
 SSFESGDPALPTTVAERDGAFWQANVGSPTAGLPGSVLGQLKGVNTASAQNLNNEGAAN  
 LADGSSGTFKWLAFASSTGWVRYEEAEFVSFVAYTMTSGDDAAGRDPKTTTVEGSDNGSTW  
 AALDRRTDEDFENRQQTRTFELEAPTAAYTYLRLNVTANSNGDSIVQLAGWDLADLSAG  
 PSAAPMTTKVGTGPRVSFTNKAGVGFSGGLHSLRYDGSGLADGETYATNVLYDDVDVVVG  
 EDTRLSTYTFPELLDDLQYFSTYAAMDVLEFDGTYLSDLGARDAHETVATAQAQEGEKI  
 LYADQWNSVRVDLGDVAEGKTVDQVLLGYDNPGGHAGTKFAGWLDDEVEITAEPATIDGS  
 SLANYVDTRRGTLAGSFSRGNNI PATATPENGFNFTFYTNASSQSWLYEYHKANNANN  
 KPVLOGGFGISHEPSPFWMGDRNQLTFLFSTASGTFDATLSTRGLEFDHHADETPARPDYGV  
 TPTNGSAIEATPTDHGAVLRFVSYPGAKGHVLDKVDGSSKLTVDQATGTISGWVENGSG  
 LSVGRTRMFVAGTFDRSPTAVGTAAGNRADARFATFETSSDKTVELRVATSPISLDQAR  
 KNLDLEVTGKTFTEVKAAAQAQWNRDLGVIEVEGASEDQLVTLYSNLYRLNLYPNSQFE  
 NTGTAQEPVYRYASFVSATPFGSATDTQTNAKIVDGKIYVNNGFWDTPYRTAWPAYSLLYP  
 ELAAELVDGFVQOYRDGGWIARWSSPGYADLMTGTSDDVAFAADAYLKGSLPTGTALEAY  
 DAALRNATVAPPSNAVGRKGLQTSFPLGFTPESTHESVSWGLEGLVNDFGIGNMAAALA  
 EDPATPEERRETLREESAYFLERATHYVELEDFEVDFFVPRHEDGTWAVDPEYDFEAW  
 GGGYTEPNGWNFAFHAPQDQGLANLYGGKQGLEDKLDEFFSTPEKGAGNGGIEHQREA  
 RDVRMGQWGMNSNQVSHHI FWLYDAAGAPSKAQEKVREVTRRLFVNSEIGQYPGDEDNG  
 EMSSWNI FASLGFYPLQVGSQYAVGSPLEFKATVHLPDGDLVVNAENNSVDNVYVQSL  
 AVDGEARTSTSLSQADLSGGTTLDFFVMGFEPSDWTGDEDDAPFSLTEGDEFFTFVQDAT  
 TAGLGTTTVADGDATTSAAALTDNTSGTRTFATTTPSITWAGNGIRPTVGSYTLTSGA  
 SGTASPSAWTLEGSDDGETWTTLDERSGEGFRWALQTRPPTVAEPTAFARYRVTVTATS  
 GSGALSLAEVELLADPKEGAEELTLSAAPDRDGVGTGREVSGSFATLTGVEGDVAALDV  
 QVAFGDGSEPVAGTLRAGAFGGYAVDAHTWTAPGVYFVTVTVSSEGIETVSASSYVSV  
 SLLREGSLLAAYDNVCI GDAGTTVGS CDGQGVFFDRAQLAAKGFVQGERATVPGTDLAF  
 DVPVAVPAGQPDNATGDGQTI ELDPADAEQLSVIGTGTEKNQQATGTLTFDDGSTQPID  
 LSFQDWSGAARNPVFGNI PVAVTDSRLRGGSPOTGT PAAFFATAPITLPEGKRFVSLTL  
 PDQPGELSRDGRIHVVAVAHDGTFAEHPALEVTAAGVTLAVGQTS DVALAQVAGGREG  
 ADLRAAVTWGSDGSDVAAGAVTDGSSVSGSHAYTAAGTYTAYVVDDGWTSQVVEVPVTVT  
 EAEFALAVDVTVSTRCLAGKAYVAVRAENGEDVFLAIRLVTPEFGTKEVAAVAPGANAYS  
 PATRVTAVEAGTVTVEATRGTGDEEVTASIQADYAAVTCG (SEQ ID NO:7)

图8

**MQSIKRTL LLLGAILPAVLGAPVQETRRAAEKLPKYIVTFKPGIDEAKIQEHTTWATNI**  
 HQRSLERRGATGGDLPVGIERNYKINKFAAYAGSFDDATIEEIRKNEDVAYVEEDQIYYL  
 DGLTTPQKSAPWGLGSI SHKQQSTDYIYDTSAGEGTYAYVVDSCVNVDHEEFEGRASKAY  
 NAAGGQHVDSIGHGTHVSGTIAGKTYGI AKKASILSVKVFQGESSTSVILDFGNWAAND  
 IVSKKRTEKAAINMSLGGYSKAFNDAVENAFEQGVLSVVAAGNENS DAGQTS PASAPDA  
 ITVAAIQKSNNRASFSNFGKVVDVFAPQDILSAWIGSSSATNTISGTSMATPHIVGLSL  
 YLAALENLDGPAAVTKRIKELATKDVVKDVKGSPNLLAYNGNA (SEQ ID NO:9)

图9

```

GGATCCATGCAGTCCATTAAGCGAACTCTGCTGCTGCTGGGAGCCATTCTGCCCGCCGTG
CTGGGAGCCCCCGTTTCAGGAGACCCGACGAGCCCGCGAGAAGCTCCCCGGCAAGTACATT
GTCACCTTCAAGCCTGGTATCGACGAGGCTAAGATTTCAGGAGCACACCACCTGGGCCACC
AACATCCATCAGCGATCCCTCGAGCGACGAGGAGCCACCCGGCGGTGACCTGCCTGTGSSA
ATCGAGCGAAACTACAAGATTAACAAGTTCGCCGCTTACGCTGGATCTTTTGACGATGCC
ACCATCGAGGAGATTGAAAGAACGAGGACGTCGCTTACGCTGGAGGAAGACCAGATCTAC
TACCTCGATGGTCTGACCACTCAGAAGTCCGCTCCTTGGGGCCTGGGCTCCATCTCTCAC
AAGGGACAGCAGTCTGACTGACTACATCTACGATACCTCCGCTGGCGAGGGTACTTACGCC
TACGTCGTGGACTCCGGTGTAAACGTTCGATCACGAGGAGTTTGAGGGACGAGCCTCTAAG
GCTTACAACGCCGCTGGAGGCCAGCATGTGGACTCTATCGGACACGGCACCCATGTTTCG
GGTACTATTGCCGAAAGACCTACGGCATCGCCAAGAAGGCTTCTATTCTCTCGGTGAAG
GTTTTCCAGGGAGAGTCCCTCTTCGACCTCTGTATCCTGGACGGCTTTAACTGGGCCGCT
AACGATATTGTGTCTAAGAAGCGAACCTCGAAGGCCGCTPATCAACATGTCCCTCGGTGGA
GGCTACTCTAAGGCCCTTCAACGACGCTGTTGAGAACGCCCTTTGAGCAGGGTGTCTGTCT
GTTGTGGCTGCTGGTAACGAGAACTCTGACGCTGGACAGACCTCCCCTGCTTCTGCTCCT
GATGCCATCACTGTGGCCGCTATTTCAGAACTCCAACAACCGAGCTTCGTTCTCCAACCTT
GGCAAGGTGGTTGACGTTTTCCGCCCGGACAGGATATCCTCTCTGCTTGGATTGGCTCC
TCTTCGGCCACCAACACTATCTCGGGCACCTCCATGGCCACTCCCCACATTGTCCGTCTG
TCCCTCTACCTGGCTGCTCTGGAGAACCTGGACCGACCTGCCGCTGTTACCAAGCGAATC
AAGGAGCTGGCTACTAAGGACGTCGTGAAGGATGTCAAGGGTTCTCCTAACCTGCTCGCC
TACAACGGCAACGCTTCTGGCGGCGGAGGACATCACCAACCATCACCATCACCAACCATCAT
TGATAACCTAGG (SEQ ID NO:10)

```

图10