(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 101084233 B
(45)授权公告日 2012.08.15

(21)申请号 200580013792.5
(22)申请日 2005.04.29
(30)优先权数据
60/620,186 2004.10.18 US
(85)PCT申请进入国家阶段日
2006.10.30
(86)PCT申请的申请数据
PCT/IB2005/051415 2005.04.29
(87)PCT申请的公布数据
WO2005/106010 EN 2005.11.10
(73)专利权人 格利科菲公司
地址 美国新罕布什尔州
(72)发明人 P·波布罗维茨
(74)专利代理机构 中国专利代理(香港)有限公司 72001
代理人 程淼 梁谋
(51)Int.Cl.
C12N 15/00(2006.01)
C07H 21/04(2006.01)
C12N 1/19(2006.01)
C07K 16/00(2006.01)
(56)对比文件
US 5272066 A,1993.12.21,
审查员 王少华

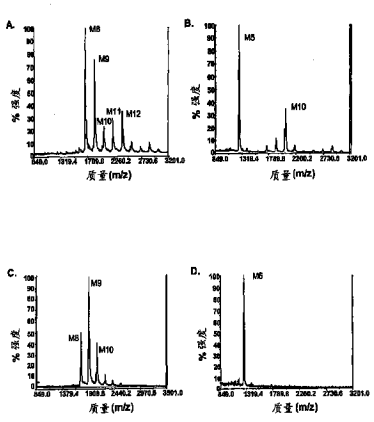
权利要求书 1 页 说明书 20 页
序列表 12 页 附图 9 页

(54)发明名称

用于在糖蛋白产生中减少或消除α-甘露糖苷酶抗性聚糖的方法

(57)摘要

本发明提供减少或消除酵母中糖蛋白上α-甘露糖苷酶抗性聚糖的方法。糖蛋白上α-甘露糖苷酶抗性聚糖的减少或消除产生自新分离的编码β 1,2-甘露糖基转移酶的P.pastoris AMR2基因的破坏。本发明还公开了与聚糖的α-甘露糖苷酶抗性相关的新的基因、多肽、抗体、载体和宿主细胞。



CN 101084233 B

1. 用于在酵母或丝状真菌宿主细胞中产生糖蛋白组合物的方法，所述糖蛋白组合物具有减少的高甘露糖聚糖的量，所述方法包含减少或消除所述糖蛋白上 α－甘露糖苷酶抗性聚糖的存在，其中减少或消除糖蛋白上 α－甘露糖苷酶抗性聚糖的步骤包括通过参与 N- 聚糖的 β－甘露糖基化的基因的破坏、缺失或突变而修饰宿主细胞，并且其中该基因如 SEQ ID NO :11 所示。

2. 权利要求 1 的方法，其中 α－甘露糖苷酶抗性的聚糖包括 β－甘露糖、支链的高甘露糖或 α-1,4 甘露糖残基。

3. 权利要求 1 或 2 的方法，其中宿主细胞选自由以下构成的组 :巴斯德毕赤氏酵母 (Pichia pastoris)、Pichia finlandica、喜海藻糖毕赤氏酵母 (Pichia trehalophila)、Pichia koclamae、膜醭毕赤氏酵母 (Pichia membranaefaciens)、Pichia methanolica、Pichia minuta(Ogataea minuta，Pichia lindneri)、Pichia opuntiae、耐热毕赤氏酵母 (Pichia thermotolerans)、柳毕赤氏酵母 (Pichi salictaria)、Pichiaguercum、皮杰普氏毕赤氏酵母 (Pichia pijperi)、具柄毕赤氏酵母 (Pichia stiptis)、毕赤氏酵母物种 (Pichia sp.)、Saccharomyces castellii、酿酒酵母 (Saccharomyces cerevisiae)、Saccharomyces kluyveri、糖酵母属物种 (Saccharomyces sp.)、多形汉逊酵母 (Hansenula polymorpha)、克卢费氏酵母属物种 (Kluyveromyces sp.)、乳酸克卢费氏酵母 (Kluyveromyces lactis)、白假丝酵母 (Candida albicans)、假丝酵母属物种 (Candida sp.)、烟曲霉 (Aspergillus fumigatus)、构巢曲霉 (Aspergillus nidulans)、黑曲霉 (Aspergillus niger)、米曲霉 (Aspergillusoryzae)、Trichoderma reesei、Chrysosporium lucknowense、镰刀菌属物种 (Fusarium sp.)、Fusarium gramineum、Fusarium venenatum，Physcomitrella patens 和粗糙链孢霉 (Neurospora crassa)。

4. 权利要求 1 的方法，其中该宿主细胞进一步包括编码 α-1,6- 甘露糖基转移酶活性的功能性基因产物的缺失。

5. 权利要求 1 的方法，其中该宿主细胞进一步包括编码甘露糖基磷酸转移酶活性的功能性基因产物的缺失。

6. 通过权利要求 1-5 的任何一项的方法产生的宿主细胞。

7. 通过权利要求 6 的宿主细胞产生的糖蛋白组合物。

8. 权利要求 7 的糖蛋白组合物，其中该糖蛋白包括 N- 连接的聚糖或 O- 连接的聚糖。

9. 修饰的酵母或丝状真菌宿主细胞，特征在于该宿主细胞已经过修饰而减少参与 N- 聚糖的 β－甘露糖基化的基因的表达，其中所述修饰的宿主细胞包含 SEQ ID NO :11 所示的基因中的破坏、缺失或突变。

10. 分离的核酸分子，其选自由以下构成的组 :

(a)SEQ ID NO :11 的核苷酸序列所示的核酸分子 ;

(b) 是 SEQ ID NO :11 的核苷酸序列的简并变体的核酸分子 ;和

(c) 编码 SEQ ID NO :12 的氨基酸序列所示的多肽的核酸分子。

11. 包含权利要求 10 的核酸序列的载体。

12. 权利要求 11 的载体，其中该核酸序列表达具有 β－甘露糖基转移酶活性的多肽。

13.SEQ ID NO :12 的氨基酸序列所示的分离的多肽。

用于在糖蛋白产生中减少或消除 α－甘露糖苷酶抗性聚糖的方法

[0001] 相关申请的交叉引用

[0002] 本申请要求 2004 年 10 月 18 日提交的美国临时申请系列 No.60/620,186 的优先权。

发明领域

[0003] 本发明涉及低等真核生物中蛋白质糖基化工程领域。本发明进一步涉及用于治疗性糖蛋白产生的酵母和丝状真菌宿主细胞的工程改造。尤其，本发明涉及糖蛋白上 α－甘露糖苷酶抗性聚糖的减少或消除，以及用于减少或除去酵母细胞中参与聚糖上 α－甘露糖苷酶抗性的产生的基因。

[0004] 发明背景

[0005] 产生重组人蛋白质的能力引起人类健康护理的很大进步并且维持了活跃的药物发现领域。许多治疗性的蛋白质需要聚糖翻译后添加至蛋白质特定的天冬酰胺残基 (N- 糖基化）以确保适当的结构－功能活性以及在人血清中随后的稳定性。对于人中治疗性的用途，糖蛋白需要人－样 N- 糖基化。可模拟人－样糖蛋白加工的哺乳动物细胞系（例如，CHO 细胞、人视网膜细胞）具有一些缺点，包括低的蛋白质效价、长的发酵时间、不同种类的产物以及连续的病毒抑制。因此希望使用不仅在短的发酵时间内产生高蛋白质，还可以产生人－样糖蛋白的表达系统。

[0006] 真菌宿主如亲甲基醇酵母巴斯德毕赤氏酵母具有用于治疗性蛋白质表达的显著优点－例如其不分泌大量的内源蛋白质，其具有强的诱导型启动子，它可在确定的化学培养基中生长，以及其可产生高效价的重组蛋白质 (Cregg 等.,2000)。然而，P.pastoris 中表达的糖基化蛋白质包含产生“高甘露糖”聚糖的另外的甘露糖，以及在糖蛋白上提供负电荷的磷酸甘露糖基团。具有高甘露糖聚糖或带电甘露聚糖的糖蛋白为禁止人中免疫应答的高风险 (Takeuchi,1997 ;Rosenfeld 和 Ballou,1974)。因此，希望在真菌宿主系统中产生治疗性的糖蛋白使得糖基化模式与人中的相同或至少类似。

[0007] 有些真菌宿主包含糖蛋白的聚糖上免疫原性的 β－甘露糖基化。为了避免抗原性，希望在人－样糖蛋白产生中消除 β－甘露糖基化。具有 β-1,2- 键的寡甘露糖苷首先由 Shibata 等.,(1985) 描述，其通过磷酸二酯桥与白色念珠菌细胞壁磷肽甘露聚糖相连。随后，已在念珠菌属细胞壁甘露聚糖的侧链中鉴定了三种类型的 β-1,2 键。第一种为位于磷酸二酯化低聚糖部分的 β-1,2- 连接的甘露－寡聚物，其为一些念珠菌属种类的甘露聚糖中共同的表位 (Shibata 等.1993a)。第二种类型为附着于白色念珠菌、tropicalis 和薄荷的甘露聚糖中 α-1,2 寡甘露糖基侧链未还原末端的 β-1,2- 连接的甘露糖单元 (Kobayashi 等.,1989,1992 和 1994)。第三种类型的 β-1,2 键在季也蒙假丝酵母中发现并且包含附着于 α-1,3 连接的甘露糖单元的 β-1,2 连接的甘露糖单元 (Shibata 等.,1993b)。尽管有这些发现，对 β-1,2 键的研究受到对鉴定 β-1,2 甘露糖基－转移酶基因的失败尝试的限制。Suzuki 等.,(1997) 描述了季也蒙假丝酵母中 β-1,2- 甘露糖基转移

酶的存在，然而该酶的基因还有待克隆。

[0008] 在C.albicans酵母中，组成磷酸甘露聚糖复合物酸－不稳定区的β－寡甘露糖苷和组成该复合物酸－稳定区的α－寡甘露糖苷用作这些致病酵母细胞附着至宿主脾脏和淋巴结巨噬细胞的粘合素(Cutler，2001)。令人感兴趣的是，保护免受各种形式念珠菌病的抗体识别β－连接的甘露三糖，而不是更大甘露糖链长的寡甘露糖苷(Han等.，1997)。据报导发展由C.albicans深度组织侵袭的患者不具有对于β－连接的寡甘露糖苷特异的可检测抗体效价，而所述抗体存在于健康个体中(Jouault等.，1997)。

[0009] 几乎没有来自P.pastoris的糖蛋白上β－连接的甘露糖残基的例子。在1986年，Kobayashi等描述了用于主要由β-1，2连接的甘露糖残基组成的甘露－低聚糖分离的具有温和条件的改进的乙酰解方法。在2003年，Trimble等报道了P.pastoris中表达的重组人胆盐－刺激的脂肪酶(hBSSL)中β-1，2-连接的甘露糖残基的存在。如由未感染的个体中保护性抗体存在所证实的，β－连接的甘露聚糖很可能为免疫原性的。另外，治疗性蛋白质上暴露的甘露糖基团由巨噬细胞上的甘露糖受体快速清除，导致低的药效。因此，考虑到其免疫原性潜力和其结合清除因子的能力，并不希望在真菌宿主例如，P.pastoris中表达的异种治疗性蛋白质N-或O-连接的聚糖上存在β－连接的甘露糖残基。

[0010] 因此需要的是用于除去糖蛋白上不想要的甘露糖残基而用于治疗性糖蛋白产生的方法。

[0011] 发明概述

[0012] 因此，本发明提供用于在酵母或丝状真菌宿主细胞中产生糖蛋白组合物的方法，所述糖蛋白组合物具有高甘露糖聚糖减少的量，所述方法包含减少或消除所述糖蛋白上α－甘露糖苷酶抗性聚糖的存在。某些实施方案，减少或消除糖蛋白上α－甘露糖苷酶抗性聚糖由通过参与N-聚糖甘露糖基化基因的破坏、删除或突变而修饰宿主细胞来实现。所述基因可包括例如，此处作为SEQ ID NO:11或其变体而描述的基因序列。

[0013] 某些实施方案中，α－甘露糖苷酶抗性的聚糖可包括β－甘露糖、分枝的高甘露糖或α-1，4甘露糖残基。

[0014] 本发明的宿主细胞优选酵母和/或丝状真菌来源的。本发明中有用的宿主细胞可包括下列科、属和种：巴斯德毕赤氏酵母、Pichiafinlandica、Pichia trehalophila、Pichia koclamae、膜醭毕赤氏酵母、Pichia methanolica、Pichia minuta(Ogataea minuta，Pichia lindneri)、Pichia opuntiae、耐热毕赤氏酵母、Pichi salictaria、Pichia guercum、Pichia pijeri、Pichia stiptis、Pichia sp.、Saccharomyces castellii、酿酒酵母、Saccharomyces kluyveri、Saccharomyces sp.、多形汉逊酵母、克卢费氏酵母属、乳酸克卢费氏酵母、白假丝酵母、假丝酵母属、烟曲霉、构巢曲霉、黑曲霉、米曲霉、Trichoderma reesei、Chrysosporiumlucknowense、镰刀菌属、Fusarium gramineum、Fusarium venenatum、Physcomitrella patens和粗糙链孢霉。

[0015] 某些实施方案中，本发明的宿主细胞可包括编码α-1，6-甘露糖基转移酶活性的功能基因，如OCH1的缺失体。在另一个实施方案中，本发明的宿主细胞可进一步包括编码甘露糖基磷酸转移酶活性的基因，如PNO1和MNN4B。

[0016] 本发明还提供具有对α－甘露糖苷酶抗性的高甘露糖聚糖结构减少量的糖蛋白组合物，其可通过本发明的方法产生。所述糖蛋白组合物可包括N-连接的聚糖和/或O-连

接的聚糖。

[0017] 本发明进一步提供分离的核酸序列，其参与 α－甘露糖苷酶抗性聚糖的产生。这些分离的核酸序列包括此处作为SEQ ID NO:11或其变体而描述的序列。本发明还包括与上述序列呈现结构相似性的另外的核酸序列。这些可包括例如，SEQ ID NO:11 的简并变体以及与上述序列具有高水平核苷酸序列同一性的核酸序列。包括在本发明中的核酸序列因此将包括具有与SEQ ID NO:11 的序列至少72％、75％、80％或85％、90％、95％、99％、99.9％同一性的那些核酸序列，以及编码具有通过SEQ ID NO:11所产生氨基酸序列的多肽的核酸序列；编码具有与通过 SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15 产生的氨基酸序列至少 72％、75％、80％或 85％、90％、95％、98％、99％、99.9％同一性的多肽的核酸序列；以及在严谨条件下与SEQ ID NO:11所述的核酸序列杂交的核酸序列。本发明还包括包含至少 60 个连续核苷酸长度的任何上述核酸序列的片段的核酸序列。

[0018] 仍然在另一个实施方案中，本发明提供修饰的酵母或丝状真菌宿主细胞。本发明的宿主细胞可表征为经修饰而降低 SEQ ID NO:11 所示核酸序列的功能基因产物的表达。某些实施方案中，修饰的宿主细胞包括SEQ ID NO:11所示核酸序列中的破坏、缺失或突变。另一个实施方案中，修饰的宿主细胞包括选自 SEQ ID NO:11 或其变体的核酸的功能基因产物表达的细胞抑制剂。可用于本发明表达的细胞抑制剂包括例如，反义 DNA 和短的干扰 RNA。

[0019] 另一个实施方案中，本发明可包括修饰的酵母或丝状真菌宿主细胞，其能表达具有高甘露糖聚糖减少量的糖蛋白组合物，所述糖蛋白组合物包括 α－甘露糖苷酶抗性的聚糖，例如－甘露糖基残基存在的减少。

[0020] 附图简述

[0021] 图 1 描述了 P.pastoris AMR2 基因的核酸和氨基酸序列。

[0022] 图 2A. 显示 P.Dastoris YAS137(Δoch1、Δpno1、Δmnn4b)N-聚糖的 MALDI-TOF MS。B. 显示 P.pastoris YAS137 的 N-聚糖用 α-1,2 甘露糖苷酶消化后的 MALDI-TOF MS。C. 显示 P.pastoris PBP130(Δoch1、Δpno1、Δmnn4b、Δamr2)N-聚糖的 MALDI-TOF MS。D. 显示 P.pastoris PBP130(Δoch1、Δpno1、Δmnn4b、Δamr2) 的 N-聚糖用 α-1,2 甘露糖苷酶消化后的 MALDI-TOF MS。

[0023] 图 3P.pastoris YAS137 的 N-聚糖用 α-1,2 甘露糖苷酶和刀豆甘露糖苷酶消化后的 MALDI-TOF MS 谱。

[0024] 图 4 说明了利用抗药性标记，卡那霉素的 P.pastoris AMR2 的融合 PCR 敲除策略。

[0025] 图 5P.pastoris AMR2 基因推导的氨基酸序列与来自白假丝酵母的 AMR2 同源物相比对而显示。

[0026] 发明详述

[0027] 除非此处另外定义，所用与本发明有关的科学和技术术语应当具有本领域普通技术人员通常了解的含义。此外，除非上下文另外要求，单数术语应当包括复数情况并且复数术语应当包括单数情况。通常，此处所述的生物化学、酶学、分子和细胞生物学、微生物学、遗传学以及蛋白质和核酸化学和杂交的技术以及所用与此相关的命名为本领域所熟知并且通常在本领域中使用。本发明的方法和技术通常根据本领域熟知的常规方法以及在各种常规和更具体的参考文献中所述的而进行，除非另有陈述，其在整个说明书中引用并且

讨论。参见例如.,Sambrook et al.Molecular Cloning:A Laboratory Manual,2d ed.,ColdSpring Harbor Laboratory Press,Cold Spring Harbor,N.Y.(1989);Ausubel 等.,Current Protocols in Molecular Biology,GreenePublishing Associates(1992,and SuPPlements to 2002);Harlow andLane,Antibodies:A Laboratory Manual,Cold Spring HarborLaboratory Press,Cold Spring Harbor,N.Y.(1990);Taylor andDrickamer,Introduction to Glycobiology,Oxford Univ.Press(2003);Worthington Enzyme Manual,Worthington Biochemical Corp.,Freehold,NJ;Handbook of Biochemistry:Section A Proteins,Vol I,CRC Press(1976);Handbook of Biochemistry:Section A Proteins,Vol II,CRC Press(1976);Essentials of Glycobiology,Cold SpringHarbor Laboratory Press(1999)。

[0028] 此处提及的所有出版物、专利和其他的参考文献在此通过引用整体引入。

[0029] 除非另有陈述,下列术语应该理解为具有下列含义:

[0030] 术语“多核苷酸”或“核酸分子”指至少 10 个碱基长度的核苷酸的聚合形式。该术语包括 DNA 分子(例如,cDNA 或基因组或合成的 DNA)和 RNA 分子(例如,mRNA 或合成的 RNA),以及包含非天然核苷酸类似物、非天然核苷酸间键或两者的 DNA 或 RNA 类似物。核酸可以是任何拓扑构象。例如,核酸可以是单链的、双链的、三链的、四体的、部分双链的、分枝的、发夹的、环状的或挂锁构象的。

[0031] 除非另有陈述,“包括 SEQ ID NO:X 的核酸”指其至少一部分具有 (i)SEQ ID NO:X 的序列,或 (ii)SEQ ID NO:X 互补序列的核酸。两者间的选择通过上下文规定。例如,如果该核酸用作探针,两者间的选择通过探针与所希望的靶标互补的要求而限定。

[0032] “分离的”或“基本上纯的”核酸或多核苷酸(例如,RNA、DNA 或混合聚合物)为基本上与其天然宿主细胞中天然伴随该天然 (native) 多核苷酸的其他细胞组分相分离的核酸或多核苷酸,例如其天然结合的核糖体、聚合酶和基因组序列。该术语包含这样的核酸或多核苷酸,其 (1) 已从其天然存在的环境移出,(2) 与自然界中该“分离的多核苷酸”发现于其中的多核苷酸的所有或一部分不相结合,(3) 与自然状态下其不与之连接的多核苷酸可操作连接或 (4) 自然界不存在。术语“分离的”或“基本上纯的”还可用于指重组或克隆的 DNA 分离物、化学合成的多核苷酸类似物或通过异种系统生物学合成的多核苷酸类似物。

[0033] 然而,“分离的”不一定要求如此所述的核酸或多核苷酸本身从其自然环境中物理地移出。例如,如果异源序列置于内源核酸序列的附近,使得该内源核酸序列的表达改变,则生物体基因组中的该内源核酸序列在此认为是“分离的”。在这里,异源序列为不天然接近该内源核酸序列的序列,而不管该异源序列本身是否是内源的(源于相同的宿主细胞或其子代)或外源的(源于不同的宿主细胞或其子代)。举例来说,启动子序列可用于代替(例如,通过同源重组)宿主细胞基因组中天然的基因启动子,使得该基因具有改变的表达模式。该基因目前将成为“分离的”,因为其与天然位于其侧翼的至少一些序列分离开。

[0034] 如果核酸包含基因组中相应核酸并不天然存在的任何修饰,则其也认为是“分离的”。例如,如果内源的编码序列包含人工,例如通过人为干预引入的插入、缺失或点突变,则其认为是“分离的”。“分离的核酸”还包括在异种 (heterologous) 部位整合入宿主细胞染色体的核酸以及作为游离基因存在的核酸构建体。此外,“分离的核酸”可基本上无其他的细胞材料,或当通过重组技术产生时基本上无培养基,或当化学合成时基本上无化学前

体或其他化学试剂。

[0035] 如此处所用的，短语对照核酸序列的“简并变体”包括可根据标准遗传密码翻译而提供与翻译自对照核酸序列的氨基酸序列相同的氨基酸序列的核酸序列。术语“简并寡核苷酸”或“简并引物”用于表示能与靶核酸序列杂交的寡核苷酸，其不一定在序列上相同但在一个或多个特定的片段内相互是同源的。

[0036] 在核酸序列的上下文中，术语“序列同一性的百分比”或“同一的”指两个序列中的残基当用于最大对应比对时是相同的。序列同一性比较的长度可超过至少约九个核苷酸，通常至少约 20 个核苷酸，更通常至少约 24 个核苷酸，典型地至少约 28 个核苷酸，更典型地至少约 32 个核苷酸，并且优选至少约 36 或更多核苷酸的范围。本领域有许多已知可用于测量核苷酸序列同一性的不同算法。例如，多核苷酸序列可利用 FASTA、Gap 或 Bestfit 比较，其为 Wisconsin PackageVersion 10.0，Genetics Computer Group(GCG)，Madison，Wisconsin 中的程序。FASTA 提供查询和搜索序列之间最佳重叠区域的比对以及序列同一性百分比。Pearson，Methods Enzymol.183 :63-98(1990)（在此通过引用整体引入）。例如，核酸序列之间的序列同一性百分比可利用 FASTA 以缺省参数（字长为 6 以及计分矩阵的 NOPAM 因数）或利用 Gap 以 GCG Version 6.1 提供的缺省参数确定，在此通过引用引入。或者，序列可利用计算机程序，BLAST(Altschul 等.，J.Mol.Biol.215 :403-410(1990) ;Gish and States，Nature Genet.3 :266-272(1993) ;Madden 等.，Meth.Enzymol.266 :131-141(1996) ;Altschul 等.，Nucleic Acids Res.25 :3389-3402(1997) ;Zhang andMadden，Genome Res.7 :649-656(1997))，尤其是 blastp 或 tblastn(Altschul 等.，Nucleic Acids Res.25 :3389-3402(1997)) 比较。

[0037] 如上所述，当指核酸或其片段时，术语“显著的同源性”或“显著的相似性”表明，当与另一个核酸（或其互补链）以适当的核苷酸插入或缺失而最佳比对时，如通过任何熟知的序列同一性算法测量的，如 FASTA、BLAST 或 Gap，存在核苷酸碱基的至少约 50％、更优选核苷酸碱基的 60％、通常至少约 70％、更通常至少约 80％、优选至少约 90％以及更优选至少约 95％、96％、97％、98％或 99％的核苷酸序列同一性。

[0038] 或者，当核酸或其片段在严谨杂交条件下杂交至另一核酸、至另一核酸的链，或至其互补链时，存在显著的同源性或相似性。核酸杂交实验上下文中的“严谨杂交条件”和“严谨洗涤条件”取决于许多不同的物理参数。如本领域技术人员很容易理解的，核酸杂交将受以下条件如盐浓度、温度、溶剂、杂交种类的碱基组成、互补区域的长度以及杂交核酸之间错配核苷酸碱基的数目的影响。本领域普通技术人员知道如何变化这些参数以实现特定的杂交严谨度。

[0039] 通常，“严谨杂交”在特定的条件设置下，在低于具体 DNA 杂交物的热熔点 (Tm) 约 25℃时进行。“严谨洗涤”在特定的条件设置下，在比具体 DNA 杂交物 Tm 低约 5℃的温度进行。Tm 为 50％的靶序列杂交至完全匹配的探针时的温度。参见 Sambrook 等.，MolecularCloning :A Laboratory Manual，2d ed.，Cold Spring Harbor LaboratoryPress，Cold Spring Harbor，N.Y.(1989)，page 9.51，在此通过引用引入。对于此处的目的，液相杂交的“严谨条件”定义为在 6X SSC（其中 20X SSC 包含 3.0M NaCl 和 0.3M 柠檬酸钠），1％ SDS 中，在 65℃ 8-12 小时的水相杂交（即无甲酰胺），随后为 0.2X SSC，0.1％ SDS 中 65℃ 20 分钟两次洗涤。技术工作者将理解取决于许多因数包括杂交序列

的长度和同一性百分比，在65℃将存在不同速率的杂交。

[0040] 本发明的核酸（也称为多核苷酸）可包括RNA、cDNA、基因组DNA和上述的合成形式和混合聚合物的正义和反义链。如本领域技术人员很容易理解的，其可经化学或生化修饰或可包含非天然的或衍生的核苷酸碱基。所述修饰包括，例如，标记、甲基化、一种或多种天然存在的核苷酸用类似物的取代、核苷酸间修饰如不带电的连接（例如，膦酸甲酯、磷酸三酯、氨基磷酸酯、氨基甲酸酯等等）、带电的连接（例如，硫代磷酸酯、二硫代磷酸酯等等）、悬挂的部分（pendentmoieties）（例如，多肽）、嵌入剂（例如，吖啶、补骨脂素等等）、螯合剂、烷化剂和经修饰的连接（例如，α 异头核酸等等）。此外包括的为合成的分子，其模拟多核苷酸通过氢键作用和其他化学相互作用结合指定序列的能力。所述分子在本领域中已知并且包括例如，其中肽键代替该分子主链中磷酸酯键的那些。其他的修饰可包括例如类似物，其中核糖环包含桥部分或其他结构如“挂锁”（locked）核酸中发现的修饰。

[0041] 术语“突变的”当应用于核酸序列时指相比对照核酸序列，可插入、缺失或改变核酸序列中的核苷酸。在一个位点可产生单个改变（点突变）或在单个位点可插入、缺失或改变多个核苷酸。此外，可在核酸序列内的任何数目的位点产生一个或多个改变。核酸序列可通过本领域已知的任何方法突变，包括但不限于诱变技术如“易错PCR”（用于在其中DNA聚合酶的复制保真度较低，使得顺着PCR产物的全长获得高比率的点突变的条件下进行PCR的方法；参见例如.，Leung等1.，Technique，1：11-15(1989)和Caldwell and Joyce，PCR MethodsApplic.2：28-33(1992)）；以及“寡核苷酸指导的诱变”（能在任何克隆的目的DNA片段中产生位点特异性突变的方法；参见例如，Reidhaar-Olson and Sauer，Science 241：53-57(1988)）。

[0042] 如此处所用的术语“载体”用来指能转运与之连接的另一个核酸的核酸分子。载体的一种类型为“质粒”，指另外的DNA片段可以连接进入其中的环状双链DNA环。其他的载体包括粘粒、细菌人工染色体(BAC)和酵母人工染色体(YAC)。载体的另一种类型为病毒载体，其中另外的DNA片段可连接进入该病毒基因组（底下更详细地论述）。某些载体能在它所导入的宿主细胞中自主复制（例如，具有在宿主细胞中起作用的复制起点的载体）。其他的载体可在导入宿主细胞后整合入宿主细胞的基因组中，并且由此随着宿主基因组一起重复。此外，某些优选的载体能指导与之可操作连接的基因的表达。所述载体此处称为“重组表达载体”（或简单地称为“表达载体”）。

[0043] 如此处所用的，术语“目的序列”或“目的基因”指核酸序列，一般编码通常不在宿主细胞中产生的蛋白质。此处公开的方法允许一个或多个目的序列或目的基因整合入宿主细胞基因组中。优选的整合位点为AMR2基因座。目的序列的非-限制性例子包括编码具有酶活性的一种或多种多肽的序列，例如影响N-聚糖在宿主中合成的酶如甘露糖基转移酶、N-乙酰基葡糖胺转移酶、UDP-N-乙酰葡糖胺转运蛋白、半乳糖基转移酶以及唾液酸转移酶。

[0044] 如此处所用的，术语“治疗性糖蛋白”包括促红细胞生成素、细胞因子如干扰素-α、干扰素-β、干扰素-γ、干扰素-ω和粒细胞-CSF、GM-CSF、凝血因子如凝血因子III、凝血因子IX以及人蛋白质C、抗凝血酶III、凝血酶、可溶的IgE受体α-链、IgG、IgG片段、IgG融合蛋白、IgM、白介素、尿激酶、Chymase、和尿素胰蛋白酶抑制剂、IGF-结合蛋白、表皮生长因子、生长激素释放因子、annexin V融合蛋白、血管抑素、血管内皮生长因子-2、

骨髓祖抑制因子-1、骨保护素(osteoprotegerin)、α-1-抗胰蛋白酶、α-胚胎蛋白、DNase II、人纤溶酶原的kringle 3、葡糖脑苷脂酶、TNF结合蛋白1、促卵泡激素、细胞毒性T淋巴细胞相关抗原4-Ig、跨膜活化剂和钙调节剂以及亲环素配体、可溶的TNF受体Fc融合蛋白、高血糖素-样蛋白1、IL-2受体激动剂。

[0045] “可操作连接的”表达调控序列指其中表达调控序列与目的基因相邻近以控制目的基因的连接，以及以反式或在一定的距离内作用而控制目的基因的表达调控序列。

[0046] 如此处所用的术语“表达调控序列”指影响与之可操作连接的编码序列的表达所必需的多核苷酸序列。表达调控序列为控制核酸序列转录、转录后事件以及翻译的序列。表达调控序列包括合适的转录起始、终止、启动子和增强子序列；有效的RNA加工信号如剪接和聚腺苷酸化信号；稳定胞质mRNA的序列；增强翻译效率的序列（例如，核糖体结合位点）；增强蛋白质稳定性的序列；以及当需要时，增强蛋白质分泌的序列。所述调控序列的性质因宿主生物体而不同；原核生物中，所述调控序列通常包括启动子、核糖体结合位点和转录终止序列。术语“调控序列”用来包括，至少其存在是用于表达必不可少的所有组分，并且还可以包括其存在是有利的另外的组分，例如前导序列和融合配偶体序列。

[0047] 如此处所用的，术语“重组宿主细胞”（或简单地称“宿主细胞”）用来指其中已导入重组载体的细胞。应该理解的是所述术语不仅用来指特定的主题细胞(subiect cell)而且指所述细胞的子代。因为由于突变或者环境影响而造成某些修饰可能在继代中存在，实际上所述子代可能不同于亲本细胞，但仍然包括在如此处所用的术语“宿主细胞”的范围内。重组宿主细胞可以是分离的细胞或培养基中生长的细胞系或可以是存在于活组织或生物体中的细胞。

[0048] 如此处所用的术语“肽”指短的多肽，例如通常小于约50个氨基酸长并且更通常小于约30个氨基酸长的多肽。如此处所用的术语包括模拟结构并且因此具有生物学功能的类似物和模拟物。

[0049] 术语“多肽”包括天然-存在的和非-天然-存在的蛋白质，以及其片段、突变体、衍生物和类似物。多肽可以是单体的或多聚体的。此外，多肽可包括许多不同的结构域，每个具有一种或多种不同的活性。

[0050] 术语“分离的蛋白质”或“分离的多肽”为根据其来源或衍生来源的蛋白质或多肽，其(1)与在天然状态下伴随其的天然缔合的组分不相缔合，(2)以自然界未发现的纯度存在，其中可根据其他细胞材料的存在判定纯度（例如，无来自相同种的其他蛋白质）(3)通过来自不同种的细胞表达，或(4)自然界不存在（例如，其为自然界中发现的多肽的片段或其包括自然界中未发现的氨基酸类似物或衍生物或不是标准肽键的连接）。因此，化学合成的或在不同于多肽天然来源的细胞的细胞系统中合成的多肽将是与其天然缔合组分相“分离的”。利用本领域熟知的蛋白质纯化技术，还可通过分离使多肽或蛋白质基本上无天然缔合组分(naturally associated components)。如由此定义的，“分离的”不一定要求如此描述的该蛋白质、多肽、肽或寡肽已从其天然环境中物理移出。

[0051] 如此处所用的术语“多肽片段”指具有缺失，例如相比全长多肽具有氨基-端和/或羧基-端缺失的多肽。在一优选的实施方案中，该多肽片段为其中该片段的氨基酸序列与天然存在序列的相应位点相同的连续序列。片段通常至少5、6、7、8、9或10个氨基酸长，优选至少12、14、16或18个氨基酸长，更优选至少20个氨基酸长，更优选至少25、30、35、

40 或 45 个氨基酸，甚至更优选至少 50 或 60 个氨基酸长，并且甚至更优选至少 70 个氨基酸长。

[0052] “修饰的衍生物”指多肽或其片段，其初级结构序列基本上同源但其包括例如，体内或体外的化学和生化修饰或其掺入天然多肽中未发现的氨基酸。如本领域技术人员很容易理解的，所述修饰包括例如，乙酰化、羧基化、磷酸化、糖基化、泛素化、标记，例如用放射性核素标记，以及各种酶促修饰。用于标记多肽的各种方法以及可用于所述目的的取代基或标记物为本领域所熟知，并且包括放射性同位素如 ^{125}I、^{32}P、^{35}S 和 ^{3}H、结合被标记抗配体的配体（例如，抗体）、荧光团、化学发光剂、酶以及可用作被标记配体的特异性结合对成员的抗配体。标记物的选择取决于所需要的灵敏度、与引物结合的容易度、稳定性要求和可得到的设备。用于标记多肽的方法为本领域所熟知。参见，例如 .，Ausubel 等 .，Current Protocols in Molecular Biology，Greene Publishing Associates(1992，and Supplements to 2002)（在此通过引用引入）。

[0053] 术语“融合蛋白”指包括偶联至异种氨基酸序列的多肽或片段的多肽。融合蛋白由于其可被构建而包含来自两种或更多不同的蛋白质的两种或更多所需的功能元件而是有用的。融合蛋白包括来自目的多肽的至少 10 个连续的氨基酸，更优选至少 20 或 30 个氨基酸，甚至更优选至少 40、50 或 60 个氨基酸，而更优选至少 75、100 或 125 个氨基酸。包括本发明蛋白质整体的融合蛋白具有特定的功用。包括在本发明融合蛋白内的异种多肽为至少 6 个氨基酸长度，常常至少 8 个氨基酸长度，并且有用地至少 15、20 和 25 个氨基酸长度。包括更大多肽如 IgG Fc 区域，以及甚至整个蛋白质的融合蛋白，如包含绿色荧光蛋白(“GFP”) 生色团的蛋白质具有特定的功用。融合蛋白可通过将编码多肽或其片段的核酸序列与编码不同蛋白质或肽的核酸序列构建在读框内并且随后表达该融合蛋白而重组产生。或者，融合蛋白可通过将多肽或其片段交联至另一种蛋白质而以化学方法产生。

[0054] 如此处所用的，术语“抗体”指其至少一部分由至少一种免疫球蛋白基因编码的多肽，或其片段，并且其可特异性结合所想要的靶分子。该术语包括天然 - 存在的形式，以及片段和衍生物。

[0055] 在该术语“抗体”范围内的片段包括通过用各种蛋白酶消化产生的、通过化学裂解和 / 或化学解离产生的以及重组产生的，只要该片段保持能特异性结合靶分子即可。所述片段为 Fab、Fab’、Fv、F(ab’)2 和单链 Fv(scFv) 片段。

[0056] 在该术语范围内的衍生物包括在序列上已被修饰，但仍然能特异性结合靶分子的抗体（或其片段），包括：种间嵌合的和人源化的抗体；抗体融合物；异抗体复合物和抗体融合物，如双抗体（双特异性抗体）、单链双抗体和内体（参见，例如 .，Intracellular Antibodies：Research andDisease Applications，(Marasco，ed.，Springer-Verlag New York，Inc.，1998)，其公开内容在此通过引用整体引入）。

[0057] 如此处所用的，抗体可通过任何已知的技术产生，包括收集自天然 B 淋巴细胞的细胞培养物，收集自杂交瘤、重组表达系统和噬菌体展示的培养物。

[0058] 术语“非 - 肽模拟物”指具有类似于对照多肽性质的化合物。非 - 肽化合物还可称为“肽模拟物”或“模拟肽”。参见，例如 .，Jones，Amino Acid and Peptide Synthesis，Oxford University Press(1992)；Jung，Combinatorial Peptide and Nonpeptide Libraries：A Handbook，John Wiley(1997)；Bodanszky 等 .，Peptide Chemistry--A

PracticalTextbook, Springer Verlag(1993) ;Synthetic Peptides :A Users Guide, (Grant, ed., W.H.Freeman and Co.,1992) ;Evans 等 ., J.Med.Chem.30 :1229(1987) ; Fauchere, J.Adv.Drug Res.15 :29(1986) ;Veber and Freidinger, Trends Neurosci.,8 : 392-396(1985) ;以及以上引用的每篇参考文献,其在此通过引用引入。所述化合物常常借助于计算机分子建模开发。结构类似于本发明有用的肽的肽模拟物可用于产生同等的作用并且因此设想属于本发明。

[0059] “多肽突变体”或“突变蛋白”指相比天然的或野生型蛋白质的氨基酸序列,其序列包含一个或多个氨基酸的插入、重复、缺失、重排或取代的多肽。突变蛋白可具有一个或多个氨基酸点取代,其中该位点的单个氨基酸已改变至另一种氨基酸,一个或多个插入和/或缺失,其中一个或多个氨基酸分别在天然-存在的蛋白质序列中插入或缺失,和/或该氨基酸序列氨基或羧基末端或两者的截短物。相比天然-存在的蛋白质,突变蛋白可具有相同的但优选具有不同的生物学活性。

[0060] 突变蛋白具有与其野生型对应物至少65%总的序列同源性。甚至更优选的为具有与野生型蛋白质至少70%、75%、80%、85%或90%总的序列同源性的突变蛋白。在更优选的实施方案中,突变蛋白呈现至少95%的序列同一性,甚至更优选98%,甚至更优选99%并且甚至更优选99.9%总的序列同一性。序列同源性可通过任何常用的序列分析算法,如Gap或Bestfit测量。

[0061] 氨基酸取代可包括那些,其:(1)降低对蛋白质水解的敏感性,(2)降低对氧化作用的敏感性,(3)改变结合亲和力而用于形成蛋白质复合物,(4)改变结合亲和力或酶活性,以及(5)给予或改进所述类似物的其他物理化学的或功能性的性质。

[0062] 如此处所用的,二十种常规氨基酸及其缩写遵循常规用法。参见Immunology-A Synthesis(Golub and Gren eds.,Sinauer Associates,Sunderland,Mass.,2nd ed.1991),其在此通过引用引入。二十种常规氨基酸的立体异构体(例如,D-氨基酸),非天然氨基酸如α-、α-二取代的氨基酸、N-烷基氨基酸以及其他的非常规氨基酸也可以为本发明多肽的合适组分。非常规氨基酸的例子包括:4-羟脯氨酸、γ-羧基谷氨酸盐、ε-N,N,N-三甲基赖氨酸、ε-N乙酰赖氨酸、O-磷酸丝氨酸、N-乙酰丝氨酸、N-甲酰甲硫氨酸、3-甲基组氨酸、5-羟基赖氨酸、N-甲基精氨酸以及其他的类似氨基酸和亚氨基酸(例如,4-羟脯氨酸)。在此处所用的多肽注释中,根据标准用法和惯例,左-侧末端对应氨基-末端而右-侧末端对应羧基-末端。

[0063] 如果编码蛋白质的核酸序列具有与编码第二种蛋白质的核酸序列类似的序列,则该蛋白质与第二种蛋白质具有“同源性”或与之“同源的”。或者,如果两种蛋白质具有“类似的”氨基酸序列,则一种蛋白质与第二种蛋白质具有同源性。(因此,术语“同源蛋白质”定义指两种蛋白质具有类似的氨基酸序列。)在一优选的实施方案中,同源蛋白质对野生型蛋白质呈现至少72%的序列同源性,更优选至少75%的序列同源性。甚至更优选的为具有与野生型蛋白质至少80%、85%、90%或95%序列同源性的同源蛋白质。在又一个更优选的实施方案中,同源蛋白质呈现至少96%、98%、99%或99.9%的序列同一性。如此处所用的,两个氨基酸序列区域之间的同源性(尤其关于预测的结构相似性)解释为暗示功能的相似性。

[0064] 当“同源的”用于指蛋白质或肽时,公认的是不相同的残基位点常常以保守的氨基

酸取代而不同。“保守的氨基酸取代”为其中氨基酸残基由具有类似化学性质（例如，电荷或疏水性）侧链（R 基团）的另一种氨基酸残基取代的取代。通常，保守的氨基酸取代基本上不会改变蛋白质的功能性质。如果两个或更多氨基酸序列彼此通过保守取代而不同，可调高序列同一性的百分比或同源性程度以修正保守的取代性质。用于进行该调节的方法为本领域技术人员所熟知。参见，例如 .，Pearson，1994，Methods Mol. Biol. 24 :307-31 and 25 :365-89（此处通过引用引入）。

[0065] 下列六组每组包含彼此为保守取代的氨基酸 :1) 丝氨酸 (S)、苏氨酸 (T) ;2) 天冬氨酸 (D)、谷氨酸 (E) ;3) 天冬酰胺 (N)、谷氨酰胺 (Q) ;4) 精氨酸 (R)、赖氨酸 (K) ;5) 异亮氨酸 (I)、亮氨酸 (L)、甲硫氨酸 (M)、丙氨酸 (A)、缬氨酸 (V)，和 6) 苯丙氨酸 (F)、酪氨酸 (Y)、色氨酸 (W)。

[0066] 多肽的序列同源性，也称为序列同一性百分比，通常利用序列分析软件测量。参见，例如 .，Genetics Computer Group(GCG) 的序列分析软件包，University of Wisconsin Biotechnology Center，910University Avenue，Madison，Wisconsin 53705。蛋白质分析软件利用指定至各种取代、缺失和其他修饰，包括保守氨基酸取代的同源性测量而匹配相似的序列。例如，GCG 包含如“Gap”和“Bestfit”的程序，其以缺省参数使用以确定紧密相关的多肽之间，如来自不同生物体种类的同源多肽或野生型蛋白质和其突变蛋白之间的序列同源性或序列同一性。参见，例如 .，GCG Version 6.1。

[0067] 当将特定的多肽序列与包含来自不同生物体的大量序列的数据库比较时优选的算法为计算机程序 BLAST(Altschul 等 .，J. Mol. Biol. 215 :403-410(1990) ;Gish and States，Nature Genet. 3 :266-272(1993) ;Madden 等 .，Meth. Enzymol. 266 :131-141(1996) ;Altschul 等 .，NucleicAcids Res. 25 :3389-3402(1997) ;Zhang and Madden，Genome Res. 7 :649-656(1997))，尤其是 blastp 或 tblastn(Altschul 等 .，Nucleic AcidsRes. 25 :3389-3402(1997))。

[0068] 用于 BLASTp 的优选的参数为 :

[0069] 期望值 :10（缺省）;过滤器 :seg（缺省）;开放间隙的成本 :11（缺省）;延长间隙的成本 :1（缺省）;最大对比 :100（缺省）;字长 :11（缺省）;描述的编号 :100（缺省）;罚分矩阵 :BLOWSUM62。

[0070] 比较同源性的多肽序列长度通常为至少约 16 个氨基酸残基，通常至少约 20 个残基，更通常至少约 24 个残基，一般至少约 28 个残基，并且优选超过约 35 个残基。当搜索包含来自大量不同生物体的序列的数据库时，优选比较氨基酸序列。利用氨基酸序列搜索数据库可通过本领域已知的 blastp 以外的算法测量。例如，多肽序列可利用 FASTA 比较，GCG Version 6.1 中的程序。FASTA 提供查询和搜索序列之间最佳重叠区域的比对和序列同一性百分比。Pearson，MethodsEnzymol. 183 :63-98(1990)（此处通过引用引入）. 例如，氨基酸序列之间的序列同一性百分比可利用如 GCG Version 6.1 提供的 FASTA 以缺省参数（字长为 2 以及 PAM250 计分矩阵）确定，在此通过引用引入。

[0071] “特异性结合”指两种分子先于结合环境中其他的分子而相互结合的能力。一般，“特异性结合”区别超过反应中偶发的结合至少两倍，更通常至少 10 倍，常常至少 100 倍。通常，如通过解离常数定量的，特异性结合反应的亲和力或亲合力为约 10^{-7}M 或更强（例如，约 10^{-8}M，10^{-9}M 或甚至更强）。

[0072] 如此处所用的术语“区域”指生物分子一级结构的物理上连续的部分。在蛋白质的情况下，区域定义为蛋白质氨基酸序列的连续部分。

[0073] 如此处所用的术语“结构域”指对生物分子已知或怀疑的功能有贡献的生物分子的结构。结构域可以是与区域或其部分具有共同边界 (coextensive)；也可包括生物分子不同的、非－连续的区域。蛋白质结构域的例子包括但不限于，Ig 结构域、胞外域、跨膜结构域和胞质结构域。

[0074] 如此处所用的，术语“分子”指任何化合物，包括但不限于，小分子、肽、蛋白质、糖、核苷酸、核酸、脂类等等，并且所述化合物可以为天然的或合成的。

[0075] 如此处所用的术语“高甘露糖”指糖蛋白上的聚糖结构，其通过酵母和／或丝状真菌种类天然产生，并且通常具有八个或更多的甘露糖残基。

[0076] 如此处所用的，术语“分子”指任何化合物，包括但不限于，小分子、肽、蛋白质、糖、核苷酸、核酸、脂类等等，并且所述化合物可以为天然的或合成的。

[0077] 术语“α－甘露糖苷酶抗性的聚糖”和“抗性的 α－甘露糖聚糖”在此可互换使用并且指糖蛋白的聚糖结构，其对 α－甘露糖苷酶，如 α-1,2；α-1,3；和／或 α-1,6- 甘露糖苷酶裂解为全部或部分抗性的。抗性的 α－甘露糖聚糖可包括 β－甘露糖、分枝的高甘露糖、α-1,4 甘露糖或未表征的甘露糖。这些聚糖结构因此提供提高的高甘露糖聚糖结构的存在。

[0078] 除非另外定义，此处使用的所有技术和科学术语具有本发明所涉及技术领域的普通技术人员通常理解的相同的意思。典型的方法和材料描述如下，虽然与所述的那些相似的或等效的方法和材料也可用于本发明的实践并且其对本领域技术人员是显而易见的。此处提及的所有出版物、专利和其他的参考文献在此通过引用整体引入。如有冲突，以包括定义的本说明书为准。材料、方法和实施例仅是解释性的而并不试图作为限制。

[0079] 贯穿本说明书和权利要求书，单词“包括”或变体如“包括”或“包括”将理解为暗示包含所述的整数或整数组而不是排除任何其他的整数或整数组。

[0080] α－甘露糖苷酶抗性基因的鉴定

[0081] 在致力于在真菌系统中表达人－样糖蛋白的努力中，需要消除非－人糖基化。真菌糖基化以高甘露糖／非－人聚糖为特征。确定甘露糖基化的类型通常为消除不想要的糖基化事件的第一步。P.pastoris 中 OCHI 的缺失引起高甘露糖基化的降低 (Choi 等.，2003)。通过负性 MALDI-TOF MS 的分析揭示负电荷的存在与这些甘露糖基团有关。推定的甘露糖基磷酸盐转移酶基因，PNO1 和 MNN4B 随后的缺失产生甘露糖基磷酸盐缺乏的菌株（图 2A）（美国申请系列 No.11/020,808）。残留的甘露糖基团通过 α-1,2 甘露糖苷酶和刀豆甘露糖苷酶的消化进一步降低甘露聚糖的大小（图 2B，图 3）（实施例 1）。尽管这些基因用 α－甘露糖苷酶破坏和消化，多至 20％的 N- 聚糖仍然维持抗性。根据糖成分分析，以及这些甘露糖基团对 α-1,2 甘露糖苷酶和刀豆甘露糖苷酶（α-1,2/α-1,3/α-1,6）的抗性，推断这些甘露糖结构为分枝的 β－甘露糖、高甘露糖、α-1,4 甘露糖或未表征的甘露糖。

[0082] 最初假定这些抗性甘露聚糖产生于甘露糖基转移酶活性，并且因此推测负责这些抗性甘露聚糖的该基因（或多个基因）与其他的甘露糖基转移酶基因有一些同源性。来自酿酒酵母 MNN4 基因的 C 末端序列用于探测 P.pastoris 的基因组（实施例 2）。通过选择编码推定的 II 型膜蛋白的基因（并且因此，可在高尔基体膜上发现），鉴定了一个基因，当其

被破坏时，在用 α-1,2 甘露糖苷酶消化后，消除抗性的甘露聚糖（图 2D）。我们命名该基因为 AMR2（α－甘露糖苷酶抗性的）。因此，如图 1 中所述，本发明公开了参与糖蛋白甘露糖基化的 P.pastoris 基因。

[0083] 核酸序列

[0084] 本发明的一个方面中，鉴定并且测序了 P.pastoris 中参与 N-聚糖（其对已知的 α－甘露糖苷酶具有抗性）甘露糖基化的基因（图 1，实施例 2）。在一实施方案中，提供编码 P.pastoris AMR2 基因及其变体的核酸序列。P.pastoris AMR2 基因的破坏尤其利于酵母菌株中糖蛋白上 α－甘露糖苷酶－抗性聚糖的减少或消除。在另一实施方案中，本发明提供包括或由序列组成的核酸分子，该序列为与 SEQ ID NO:11 具有至少 72％同一性的 P.pastoris AMR2 基因的变体。核酸序列优选具有与野生型基因的至少 75％、80％或 85％的同一性。更优选，核酸序列具有对 SEQ ID NO:11 的 90％、95％、99％、99.9％或甚至更高的同一性。

[0085] 根据本发明其他的实施方案，本发明的核酸分子编码具有图 1 中显示的氨基酸序列的多肽。还提供编码对 SEQ ID NO:12 至少 72％同一性的多肽序列的核酸分子。通常，本发明核酸分子编码对 SEQ IDNO:12 至少 75％、80％或 85％同一性的多肽序列。更优选，编码的多肽具有对 SEQ ID NO:12 的 90％、95％、98％、99％、99.9％或甚至更高的同一性。

[0086] 本发明的另一个方面中，参与 N-聚糖的 α－甘露糖苷酶－抗性甘露糖基化的 AMR2 基因与 P.pastoris 中的三个另外的基因同源 -AMR1、AMR3、AMR4（图 5），并且具有对下列种类中基因的同源区域：白假丝酵母中的 11 个基因，（图 5 中显示八个）；Saccharomyces castellii 株系 NRRL Y-12630 中的八个基因；Saccharomyces kluyveri 株系 NRRL Y-12651 中的两个基因，以及烟曲霉中的三个基因（实施例 3）。对于减少或消除其他物种中抗性的 α－甘露糖聚糖，本领域技术人员可鉴定来自给定物种基因组的 AMR2 同源物，并且破坏该同源基因或多个基因。本领域技术人员理解破坏任何给定的物种中发现的所有同源基因，或同源基因的组合可能是必要的。更具体地，熟练技术人员认识到为了减少或消除其他物种中 α－甘露糖苷酶抗性的 N-聚糖，可根据保守序列设计用于 AMR2 同源物 PCR 克隆的简并引物。或者还可构建用于 AMR2 同源物杂交的探针。

[0087] 宿主细胞

[0088] 本发明的另一方面中，产生糖蛋白的宿主细胞，其通常具有 α－甘露糖苷酶－抗性的聚糖，已经过工程改造而产生糖蛋白但无 α－甘露糖苷酶－抗性的 N-聚糖。在一实施方案中，产生治疗性糖蛋白的宿主细胞，其通常具有 α－甘露糖苷酶－抗性的聚糖，已经过工程改造而产生治疗性糖蛋白但无甘露糖苷酶－抗性的聚糖。在一优选的实施方案中，本发明的宿主细胞已通过与本发明分离核酸的破坏、缺失或突变的重组而突变以使宿主细胞中聚糖上的 α－甘露糖基化相比缺乏该突变的宿主细胞减少。更优选，消除 N-聚糖上的 α－甘露糖苷酶－抗性。本发明的宿主细胞优选为巴斯德毕赤氏酵母或 Pichia methanolica，但其他的宿主细胞，尤其酵母细胞也包括在本发明范围内。

[0089] 一些实施方案中，酵母宿主细胞中的 AMR2 基因通过实施例 3 中论述的并且在图 4 中显示的 PCR 敲除策略破坏。本发明的另一些实施方案中，α－甘露糖苷酶抗性活性缺陷型宿主细胞被用于利用本发明的核酸分子和／或方法将一个或多个目的序列或基因整合入宿主细胞基因组中。在一优选的实施方案中，整合目的序列或基因以便破坏宿主细胞的

内源基因。例如，AMR2 基因被目的序列破坏。包含所述整合的宿主细胞很容易通过选择标记物鉴定，其在酵母中通常为营养缺陷型的基因，其允许转化的细胞在缺乏特定氨基酸或核苷酸的培养基上生长，或抗生素抗性基因，其允许在包含相应抗生素的培养基上生长。

[0090] 本发明的另一个方面中，提供用本发明的核酸分子或载体转化的宿主细胞以及其子代。本发明的一些实施方案中，这些细胞携带载体上的本发明的核酸序列，其可以是但不需要是自主复制载体。本发明的另一些实施方案中，核酸已经整合入所述宿主细胞的基因组中。

[0091] 编码参与糖蛋白的聚糖上的 α－甘露糖抗性活性的破坏的 AMR2 基因优选为来自毕赤氏酵母属的酵母菌株。属于本发明毕赤氏酵母属的酵母包括，但不限于例如，巴斯德毕赤氏酵母、Pichia finlandica、Pichiatrehalophila、Pichia koclamae、膜醭毕赤氏酵母、Pichia methanolia、Pichia minuta(Ogataea minuta，Pichia lindneri)、Pichia opuntiae、Pichia thermotolerans、Pichi salictaria、Pichia guercum、Pichia pijperi、Pichia stiptis、Pichia sp.，以及其他的酵母。又一个实施方案中，携带 AMR2 活性和／或同源性的基因可在下列宿主的一种中破坏：Saccharomyces castellii、酿酒酵母、Saccharomyces kluyveri、Saccharomyces sp.，多形汉逊酵母、克卢费氏酵母属、乳酸克卢费氏酵母、白假丝酵母、假丝酵母属、烟曲霉、构巢曲霉、黑曲霉、米曲霉、Trichoderma reesei、Chrysosporium lucknowense、镰刀菌属、Fusariumgramineum、Fusarium venenatum、Physcomitrella patens 以及粗糙链孢霉。

[0092] 编码 β－甘露糖基化活性的 AMR2 基因

[0093] 观察到具有该 AMR2 基因同源物的一些酵母种类也具有 β－甘露糖基化活性。因此，本发明的另一个方面中，所述 AMR2 基因编码 β－甘露糖基转移酶活性。在一实施方案中，编码 β－甘露糖基转移酶活性的 AMR2 单独和／或与 P.pastoris 或其他的真菌物种中任何其同源物一起破坏，导致糖蛋白上 α－甘露糖基化抗性聚糖的减少或消除。

[0094] O－聚糖上 β－甘露糖基化的存在在 Trimble 等.，2003 的报告中论述。因此，另外的实施方案中，P.pastoris 以及其他种类中的 α－甘露糖苷酶抗性 O－聚糖用 AMR2 基因及其同源物的破坏而减少或消除。

[0095] 下列为举例说明本发明组合物和方法的实施例。这些实施例不应认为是限制－所述实施例仅包括用于例证性的目的。

[0096] 实施例 1

[0097] N－聚糖的刀豆和 α－1,2 甘露糖苷酶消化

[0098] 标准的 N－连接的低聚糖 (20mg) 重构于 100ml HPLC 级水中。10ml 的等分试样添加至 0.6ml 的硅化管。蒸干样品。10ml 50mM 乙酸铵与刀豆甘露糖苷酶 (0.03U) 或来自 Trichoderma reseei 的 α－1,2 甘露糖苷酶 (0.03mU，Dr Contreras R 赠送，Unit of Fundamental andApplied Molecular Biology，Department of Molecular Biology，GhentUniversity，Ghent，Belgium) 一起添加至样品。样品在 37℃与酶一起温育 16 至 24h。随后蒸干样品。样品重构于 10ml 水中。样品随后通过 MALDI-TOF MS 分析。

[0099] 基质辅助的激光解吸飞行电离时间质谱法

[0100] 利用 Voyager DE PRO linear MALDI-TOF(Applied Biosciences) 质谱仪利用延缓的提取确定所述聚糖的分子量。每孔的干燥的聚糖溶于 15μL 水并且 0.5μL 点样于不

锈钢样品平板上且与0.5μL S-DHB基质(1∶1水/乙腈0.1% TFA中的9mg/mL双羟基苯甲酸,1mg/mL 5-甲氧水杨酸)混合并且干燥。

[0101] 通过用脉冲氮激光器(337nm)以4ns脉冲时间的照射产生离子。所述仪器以125ns延缓和20kV加速电压的延缓提取的方式操作。栅极电压为93.00%,导线电压为0.10%,内压小于5X 10-7托,并且低质量的闸为875Da。从100-200个激光脉冲的总和产生谱并且用2GHz的数字转换器获得。$Man_5GlcNAc_2$低聚糖用作外部分子量标准。所有光谱用所述仪器以正离子方式产生。评估的所述光谱的质量准确度为0.5%。

[0102] 实施例2

[0103] 来自巴斯德毕赤氏酵母AMR2基因的鉴定和序列分析

[0104] 包含富集赖氨酸和谷氨酸的重复序列的酿酒酵母MNN4基因(Genbank登录号#P36044)的C末端部分用作探针而对P.pastoris基因组序列(Integrated Genomics,Chicago,IL)进行blast。鉴定具有ORF的数个DNA片段,该ORF编码具有类似的赖氨酸和谷氨酸富集重复体的蛋白质。其中发现一个ORF编码具有推定的结构类似于酿酒酵母Mnn4p的N末端跨膜结构域和富集赖氨酸和谷氨酸的C末端尾的644个氨基酸的蛋白质。基于携带突变等位基因的菌株的表型分析,该基因命名为AMR2(α-甘露糖苷酶抗性的)。随后P.pastoris基因组序列的blast搜索揭示存在与AMR2紧密有关的另外三个基因。

[0105] 实施例3

[0106] YAS137菌株中AMR2基因的缺失

[0107] P.pastoris菌株YAS137(Δoch1、Δpno1、Δmnn4b)(美国申请系列No.11/020,808)用作AMR2基因敲除实验的宿主。YAS137产生无外部链的无电荷N-聚糖(图2A)。当用α-1,2-甘露糖苷酶(图2B)和刀豆(具有α-1,2/α-1,3/α-1,6活性)甘露糖苷酶消化时,纯化自菌株YAS137的N-聚糖的显著百分比可转换成$Man_5GlcNAc_2$。然而,上至20%的总N-聚糖对这些α-甘露糖苷酶具有抗性(图3)。通过PCR重叠方法产生amr2缺失等位基因(amr2::Kan^R)(图4)。与PBS2-KO2(SEQ ID NO:2)(5′-GTGCTACCTAAAT-CGTATGTGTCGTTGAAGCTTCCCAATGATAGC-3′)成对的引物PBS1-2-C3(SEQ ID NO:1)(5'-TAATAGTGGAGAAA-CTTGCAAAGG-3'),以及与PBS1-2-KO4(SEQ ID NO:4)(5′-CTTGG-TTCAACGCAGCACTTTG-AC-3′)成对的PBS1-2-KO3(SEQ ID NO:3)(5'-CTCCCTATAGTGAGTCGTATTCATATGAT-GGGTGTTTGCTCACTC-3')用于扩增来自基因组DNA(分离自菌株NRRL-Y11430的基因组DNA)的AMR2基因的5'和3'侧翼区。与PR32(SEQ ID NO:6)(5'-AATACGACTCACTATAGG-GAG-3')成对的引物PR29(SEQID NO:5)(5'-CACATACGATTTAG-GTGACAC-3')用于扩增来自载体pUG6(Goldstein and McCusker,1999)的Kan(G418)抗性标记物。随后,引物PBS1-2-C3和PBS1-2-KO4用于用来自第一轮PCR反应的所有三种产物的第二个反应以产生重叠产物。得到的融合PCR产物用于转化菌株YAS137。通过添加乙酸钠至终浓度0.3M而制备用于转化的DNA。随后向DNA样品添加百分之百冰冷的乙醇至终浓度70%。DNA通过离心沉淀(12000g×10min)并且用70%冰冷的乙醇洗涤两次。干燥DNA并且随后重悬于50ml的10mM Tris中,pH8.0。通过在BMGY中(缓冲最小量的甘油:100mM磷酸钾,pH6.0;1.34%酵母氮基质;4×10^{-5}%生物素;1%甘油)扩大酵母培养至~2-6的O.D.而制备YAS137。酵母通过在1M山梨糖醇中洗涤3次并且重悬于~1-2mls 1M山梨糖醇中而制备为电感受态。DNA(1-2mg)与50ml感受态酵母混合并且冰上孵育1分钟。酵母

随后用 BTX ElectrocellManiPulator 600 利用下列参数电穿孔：1.5kV，129ohms 以及 25mF。一毫升 YPDS(1％酵母提取物，2％蛋白胨，2％右旋糖，1M 山梨糖醇）添加至电穿孔的细胞。转化的酵母随后在选择性琼脂平板上铺板。在含有 200mg/ml G418 硫酸盐，GIBCO™ 的 YPD 培养基上选择转化体。通过 PCR 证实缺失等位基因 amr2::Kan^R 的正确整合。通过敲除构建体的 5’和 3’部分的 PCR 扩增进行敲除的筛选。PBS1-C5(SEQ IDNO：7)(5’-TTTTCCTCA-AGCCTTCAAA-GACAG 3’）和 PTEF(SEQ ID NO：8)(5’-AGCTGCGCA-CGTCAAGACTGTCAA-GG-3’）引物用于筛选敲除构建体的 5’部分而 PBS1-2-C2(SEQ ID NO：9)(5’-TACCGATACATAC-GTAGCCAACAC-3’）和 KAN10(SEQ IDNO：10)(5’-TCGCTATACTGCTG-TCGATTCGATAC-3’）引物用于筛选敲除构建体的 3’部分。在两个筛选中观察到 PCR 产物为 AMR2 基因成功敲除的指示，因为引物 PTEF 和 KAN10 分别在抗药性标记物序列的 5’和 3’端退火；并且 PBS1-C5 和 PBS1-2-C2 与位于用于敲除构建体中的 DNA 的 5’和 3’区域侧翼的基因组中的序列互补。新的（Δoch1、Δpno1、Δmnn4b、Δamr2）菌株命名为 PBP130。

[0108] PCR 扩增

[0109] Eppendorf Mastercycler 用于所有的 PCR 反应。PCR 反应包含模板 DNA，125mM dNTP，0.2mM 每种正向和反向引物，Ex Taq 聚合酶缓冲液 (Takara Bio Inc.)，以及 Ex Taq 聚合酶 (Takara Bio Inc.)。扩增预测的 AMR2 基因 5’，预测的 AMR2 基因的 3’的 DNA 片段以及抗药性标记物，以 30 个循环的 98℃ 10 秒，52℃ 10 秒和 72℃ 2min，94℃ 2min 的初始变性步骤以及 72℃ 10min 的最终延伸步骤。通过琼脂糖凝胶电泳分开 PCR 样品且用 Qiagen 的 Gel Extraction Kit 抽提和纯化 DNA 条带。所有 DNA 纯化物在 10mM Tris，pH8.0 中洗脱。

[0110] 搜索和比对

[0111] 如本发明详述中论述的，进行 BLAST 搜索以获得来自 NCBI 的完全和不完全的真菌基因组的序列，导致 AMR2 同源物的鉴定。利用 Megalign 程序 (DNAStar) 和 ClustalW 算法构建图 5 中显示的比对。

[0112] 实施例 4

[0113] P.pastoris 中 α－甘露糖苷酶抗性 N-聚糖的确定

[0114] YAS137 和 PBP130 中 N-连接的聚糖通过分泌在甲醇诱导型 AOX1 启动子调控下表达的 His-标记的报告蛋白而分析。报告蛋白 K3，包含单个 N-糖基化位点。简要地，含有 BMGY 的摇瓶用 YAS-130 的新鲜培养物接种并且生长至～ 20 的 O.D.。离心培养物并且用 BMMY（缓冲的最小量甲醇：除了 0.5％甲醇代替 1％甘油外，与 BMGY 相同）洗涤细胞沉淀。细胞沉淀重悬于 BMMY 至原始 BMGY 培养物的 1/5 体积并且置于振荡器中 24h。通过离心沉淀生物物质并且转移培养基至新的管而收集分泌蛋白。His-标记的 K3 蛋白随后在 Ni-亲和柱上纯化并且用 PNGase 消化 (Choi 等.，2003)。

[0115] 实施例 5

[0116] Man10 和 Man11/12 的分析

[0117] 对 α－甘露糖苷酶消化抗性的聚糖的结构分析揭示了连接至核心甘露糖低聚糖的至少一个 β1，2-甘露糖基残基。结构 I 显示在分枝 5 呈现的建议的 β－甘露糖基残基 13，其可明确连接至该链，并且很可能为高－甘露糖结构的不可分割的部分。下列聚糖结构上（结构 I）的残基基于 Ziegler 等的论文编号 (Ziegler，F.D.，J.Cavanagh，C.Lubowski，

R.B.Trimble,1999,Glycobiology 9:497-505),任意添加另外的编号(13、14、15)。

[0118] 结构 I

[0119] 9

[0120] .....[Man a(1-2)]Man a(1-6)6

[0121] \

[0122] Man a(1-6)4

[0123] /\

[0124] .....Man a(1-3)7\321

[0125] 15Man bβ(1-4)GlcNAcβ1-4)GlcNAc

[0126] [Man a]-[Man a(1-6)]12/

[0127] \/

[0128] Man a(1-3)5

[0129] 14 13 11 8/

[0130] ...[Man a(1-2)]Man b(1-2)Manαa(1-2)Man a(1-2)

[0131] 对 β-甘露糖基残基的存在分析来自三个样品的数据。

[0132] 样品 1:‘man10’-此为作为 man9/10 混合物描述的原始样品

[0133] 样品 2:‘man11’-此为作为来自 P4 馏分 63-66 的 man11/man12 混合物而描述的第二种制备物,估计 50%的 man11,15%的 man12。

[0134] 样品 3:‘man10 消化物’-此为其中原始的 man11/man12 混合物用 1,2-甘露糖苷酶处理而调整回 man9 或 man10 的第三种制备物。

[0135] 还检验了包含核心 man8/man9 混合物的第四种样品。

[0136] 从 D20 冻干样品,并且再溶入 200μL 或 500μL 的 D20。

[0137] 在 25C 收集 Varian Inova 600MHz 和 800MHz 分光计上的 NMR 数据。

[0138] 来自 Varian library(梯度 COSY、TOCSY、NOESY、梯度 HSQC 和 HMBC)的标准实验用于表征。

[0139] β-甘露糖基残基的证据:

[0140] 所有三种样品具有与 β-端基异构体一致的 NMR 信号。三种光谱特征支持该论据:化学位移、H1-H2 标量耦合和分子内 NOE 测量。

[0141] 质子化学位移。

[0142] α-甘露糖基残基异头质子的化学位移值通常大于 5.0ppm,而 β-甘露糖基残基的通常小于 5.0ppm。也有例外,如 α-甘露糖基残基 3、4 和 12,因此仅此不能证实异头构型。

[0143] 显示四种样品质子 1D 光谱异头区域的分析揭示两个峰 13 和 13t,其对应于 1-2 连接至 11 的 β-甘露糖基残基,13t 为末端残基,并且 13 似乎用 α-甘露糖基残基 14 另外取代。可将这些光谱与来自 ziegler 等中发现的高-甘露糖结构的 NMR 数据比较,其仅在低于 5ppm 的化学位移区域中显示核心残基 3 和 4,以及残基 12。

[0144] 从 2D TOCSY 数据提取另外的化学位移,其显示来自残基 13 的不同的 H1、H2、H3、H4 和 H5 信号,其与 Trinel 等.,JBC 1999 和 Nitz 等.,JBC 2002 中报道的数据紧密对应。来自末端残基 13t 的信号并不同样分开的但具有类似的位移(参见表 1)。

[0145] 2D TOCSY(质子-质子关联图)显示属于 β-Man 13 残基的交叉峰(数据没有显示)。β-甘露糖基残基的化学位移和文献值列于表 1 中。

[0146] 表 1

[0147]

	H1	H2	H3	H4	H5
b-man 13	1.90	1.35	1.75	1.64	1.40
b-man 13t	1.86	1.27	1.67	1.47	1.40
Trinel et al.	1.90-5	1.16-4	1.63-3	1.47-3	1.41
	.05	.41	.73	.59	
Nitz et al.	1.95	1.42	1.62	1.56	

[0148] α-甘露糖基残基的化学位移通常具有相同的等级,但 H2 的值通常为低～0.1ppm,H3 为高～0.1ppm 并且 H5 为高～0.2ppm。

[0149] (Trinel, P.-A., Y.Plancke, P.Gerold, T.Jouault, E.Delplace, R.T.Schwarz, G.Strecker, 和 D.Poulain,1999, J.Biol.Chem.274 :30520-30526)(Nitze.M, C.-C.Ling, A.Otter, J.E.Cutler, D.R.Bundle,2002, J.Biol.Chem.,277 :3440-3446)

[0150] NOE 交叉峰。

[0151] β-异头质子位于吡喃糖环的与 H3 和 H5 相同的面并且约 2.4 埃间距,而 α-异头质子位于背面,并且离 H3 或 H5 接近 4 埃。因此,预期 β-异头物中的 H1 和 H3 以及 H5 之间有强的 NOE 信号。这在 NOESY 数据中清楚地显示(未呈现),其中光谱显示 NOE 交叉峰,其通过空间相互作用而产生并且为高度距离依赖性的。4.895ppm 的水平线与 β-man 13 残基的 H1 对应(数据没有显示)。如通过顶部板块中的 TOCSY 试验确定的,垂直线证实这些峰赋予特定的共振。强的峰 H3 和 H5 应归于 H1 和 H3 或 H5 之间的残基内的 NOE,证实 β-异头赋值(数据没有显示)。

[0152] H1 至 H2 的标量耦合(峰的'分裂')

[0153] 产生共振'分裂'和多重结构的质子标量耦合 (scalar coupling) 的数值与所涉及质子的相对方位有关。其因此能成为异头构型的指示。

[0154] α-甘露糖基残基中的 H1-H2 标量耦合为约 1.5Hz,而 β-异头物的为< 1Hz。在具有非常窄线的光谱中可看到 α-异头物的分裂,而 β-异头物太小并且信号看起来像单峰。这些数据中,行距不是最佳的,因此两组峰看起来像单峰。然而,2D TOCSY 中峰的模式和强度(数据没有显示)还取决于偶合的数值。TOCSY 光谱的区域显示其中水平线与 H1 信号对应,并且交叉峰来自各自残基中的其他质子。与 β-Man 残基 (3、13、13t) 有关的 H2 区域的第一个峰比 α-Man 峰(例如 4) 的强度更低。强度为信号转移效率以及因此耦合常数量值的大致指示。

[0155] α-异头物还呈现 H1 至 H3 以及有时 H4 的另外的交叉峰-此表示 H1 和 H2 之间的标量耦合足够大以允许信号转移超过 H2。相反,β-异头物仅呈现 H1 至 H2 交叉峰;因为偶合如此小而信号转移非常低效并且未延续至环中的其他质子。13、13t 和 3 的模式相同,但

明显不同于另一种甘露糖基残基。

[0156] β-甘露糖基 13(t) 键合位点的证据

[0157] H1 和 H2 化学位移的变化。

[0158] 对于已知的高-甘露糖结构，通常足以将 H1 和 H2 质子化学位移与标准的文献值比较而得到一种结构。在这种情况下，不规则的信号需要更多的工作来建立键合。然而，存在一些峰，其可分配给核心组织，如呈现用于 Man8/9 光谱。比较 Man8/9 和 Man9/10 光谱（数据没有显示），有许多差异，包括新信号 13 和 13t，其已在上面如 β-甘露糖基残基所述。

[0159] 比较样品 Man9/10 和 Man11/12 的光谱，类似于 9 的末端 α-甘露糖基残基返回。此提示 Man9/10 光谱中标记为 6、12 的峰很可能为末端 6，因为其在用另一个 α-甘露糖取代后预计会移位。其回到甘露糖苷酶消化样品，Man10- 消化物的起始位置。来自 man11 和 Man10- 消化物样品的碳-质子相关数据 (HSQC，HMBC 光谱，未显示）证实该残基以 1-6 键连接。

[0160] 如在典型的 Man9 结构中看到的，标记为 14 的信号提示为连接至 13 的 α-甘露糖基残基，而不是取代的甘露糖基残基 7。此以底下论述的 NOE 数据为基础。然而，为了解释异头质子的化学位移，其可能为另外取代的。

[0161] NOE 数据分析。

[0162] 用于确定键合位点的初始数据来自 NOE 光谱，其显示空间上接近的质子。因此，除了残基内的 NOE 交叉峰，还可观察到残基间的交叉峰，其通常显示键合位点。在 Man(1-2) 键合的情况下，观察到相同的残基中 H1 和 H2 之间，糖苷连接的残基的 H1 和 H2 之间并且常常糖苷连接的残基的 H1 和 H1 之间的交叉峰。

[0163] 分析来自 man10 样品 NOESY 和 TOCSY 光谱的区域（数据没有显示）。上平板中，峰代表鉴定的甘露糖基残基 H1 和 H2 之间的相互关系。中间平板中，与其赋值一致，标记为 11 的信号显示甘露糖基残基 8 的 H1 和 H2 之间的 NOE 相互关系 (box)。如果检验来自残基 13 和 13t 的异头质子的 NOE 交叉峰，其与甘露糖基残基 11 的 H2 有相互关系。此将 β-甘露糖基残基连接至核心寡糖的 1-3 支链。底下平板显示 13t 的 H1 和 11 的 H1 之间另外的强 NOE 相互关系，也与 b(1-2) 键合一致。

[0164] 对该结构片段的另外的支持来自甘露糖基残基 11 的 H2 的异常化学位移；其值与来自 Trimble 等 (Trimble，R.B.，C.Lubowski，C.HauerIII，R.Stack，L.McNaughton，T.Gemmill，和 S.Anand Kumar，2004，Glycobiology 14：265-274)) 的数据一致，其显示 β-man 取代的 α-甘露糖基残基 2 的 4.26ppm 的化学位移。此外，质子-碳相关数据 (HSQC 和 HMBC 光谱，未显示）表明甘露糖基 11（以及例如 8 和 5) 的 C2 的碳化学位移为糖苷键中碳的高值（～82-84ppm) 特性。

[0165] 推断至少一种 β-甘露糖基残基可显示 1-2 连接至核心甘露糖低聚糖。可能有 β-甘露糖基残基的多个键合位点，并且其本身可被取代。

[0166] 参考文献

[0167] Choi，B-K. 等 .2003.Use of combinatorial genetic libraries tohumanize N-linked glycosylation in yeast Pichiapastoris.Proc.Natl.Acad.Sci.100；5022-5027.

[0168] Cregg, J.M. 等 .2000.Recombinant protein expression in Pichiapastoris. Mol.Biotechnol.16 ;23-52.

[0169] Cutler, J.E.2001 N-glycosylation of yeast, with emphasis onCandida albicans.Med.Mycology.39 ;S75-S86.

[0170] Han, Y., Kanbe, T., Cherniak, R.and Cutler, J.E.1997. Biochemicalcharacterization of Candida albicans epitopes that can elicit protectiveand nonprotective antibodies.Infect.Immun.65 ;4100-4107.

[0171] Joualt, T., Delaunoy, C.Sen did, B., Ajana, R.and Poulain, D.1997. Differential humoral response against α-and b-linked mannoseresidues associated with tissue invasion by Candidaalbicans.Clin.Diagn.Lab.Immunol.4 ; 328-333.

[0172] Kobayashi,H.,Shibata,N.and Suzuki,S.1986.Acetolysis of Pichiapastoris IFO 0948Strain Mannan Containing α-1,2and β-1,2LinkagesUsing Acetolysis medium of Low Sulfuric AcidConcentration.Arch.Biochem.Biophys.245 ;494-508.

[0173] Kobayashi, H 等 .1989.Structural study of phosphomannan ofyeast-form cells of Candida albicans J-1012 strain with specialreference to application of mild acetolysis.Arch.Biochem.Biophys.272 ;364-375.

[0174] Kobayashi, H. 等 .1992.Structural study of a cell wallmannan-protein complex of the pathogenic yeast Candida glabrata IFO0622strain.Arch.Biochem. Biophys.294 ;662-669.

[0175] Kobayashi, H. 等 .1994.Structural modification of cell wallmannans of Candida albicans serotype A strains grown in yeastextract-Sabouraud liquid medium under acidicconditions.Infect.Immun.62 ;968-973.

[0176] Rosenfeld, L.and Ballou, C.1974.Genetic Control of Yeast MannanStructure.J.Biol.Chem.249 ;2319-2321.

[0177] Shibata, N., Ichikawa, T., Tojo, M. 等 .1985.Immunochemical studyon the mannans of Candida albicans NIH A-207, NIH B-792and J-1012strains prepared by fractional precipitation withcetyltrimethylammonium bromide.Arch.Biochem. Biophys.243 ;338-348.

[0178] Shibata, N., Hisamichi, K., Kobayashi, H., and Suzuki, S.1993a.Complete assignment of 1H and 13C nuclear magneticresonance chemical shifts of beta-1, 2-linked mannooligosaccharidesisolated from the phosphomannan of the pathogenic yeast Candidaalbicans NIH B-792strain.Arch Biochem Biophys.302 ;113-117.

[0179] Shibata, N., Hisamichi, K., Kobayashi, H., and Suzuki, S.1993b.Structural study of a cell-wall mannan of Saccharomyceskluyveri IFO 1685 strain.Presence of a branched side chain andbeta-1,2-linkage.Eur J Biochem.217 ;1-12.

[0180] Suzuki, A. 等 .1997.Charcterization of β-1,2Mannosyltransferasein Candida guilliermondii and Its Utilization in the Synthesis of Novel Oligosaccharides.J.Biol.Chem.272 ;16822-16828.

[0181] Takeuchi, Makato. 1997. Trial for Molecular Breeding of Yeast forthe Production of Glycoprotein Therapeutics. Trends in Glycoscienceand Glycotechnology. 9 ;S29-S35.

[0182] Trimble, R.B. 等 . 2004. Characterization of N-and O-linkedglycosylation of recombinant humah bile salt-stimulated lipasesecreted by Pichia pastoris. Glycobiology. 14 ;265-274.

[0001] 序列表

[0002] <110>GlycoFi, Inc.

[0003] Bobrowicz, Piotr

[0004] <120> 用于在糖蛋白产生中减少或消除 a- 甘露糖苷酶抗性聚糖的方法

[0005] <130>PCT-GF0004Y

[0006] <140>PCT/IB05/51415

[0007] <141>2005-04-29

[0008] <160>15

[0009] <170>FastSEQ for Windows Version 4.0

[0010] <210>1

[0011] <211>24

[0012] <212>DNA

[0013] <213> 人工序列

[0014] <220>

[0015] <223> 简并寡核苷酸引物

[0016] <400>1

[0017] taatagtgga gaaacttgca aagg 24

[0018] <210>2

[0019] <211>45

[0020] <212>DNA

[0021] <213> 人工序列

[0022] <220>

[0023] <223> 简并寡核苷酸引物

[0024] <400>2

[0025] gtgctaccta aatcgtatgt gtcgttgaag cttcccaatg atagc 45

[0026] <210>3

[0027] <211>45

[0028] <212>DNA

[0029] <213> 人工序列

[0030] <220>

[0031] <223> 简并寡核苷酸引物

[0032] <400>3

[0033] ctccctatag tgagtcgtat tcatatgatg ggtgtttgct cactc 45

[0034] <210>4

[0035] <211>24

[0036] <212>DNA

[0037] <213> 人工序列

[0038] <220>

[0039] <223> 简并寡核苷酸引物

[0040] <400>4

[0041] cttggttcaa cgcagcactt tgac 24

[0042] <210>5

[0043] <211>21

[0044] <212>DNA

[0045] <213> 人工序列

[0046] <220>

[0047] <223> 简并寡核苷酸引物

[0048] <400>5

[0049] cacatacgat ttaggtgaca c 21

[0050] <210>6

[0051] <211>21

[0052] <212>DNA

[0053] <213> 人工序列

[0054] <220>

[0055] <223> 简并寡核苷酸引物

[0056] <400>6

[0057] aatacgactc actataggga g 21

[0058] <210>7

[0059] <211>24

[0060] <212>DNA

[0061] <213> 人工序列

[0062] <220>

[0063] <223> 简并寡核苷酸引物

[0064] <400>7

[0065] ttttcctcaa gccttcaaag acag 24

[0066] <210>8

[0067] <211>26

[0068] <212>DNA

[0069] <213> 人工序列

[0070] <220>

[0071] <223> 简并寡核苷酸引物

[0072] <400>8

[0073] agctgcgcac gtcaagactg tcaagg 26

[0074] <210>9

[0075] <211>24

[0076] <212>DNA

[0077] <213> 人工序列

[0078] <220>

[0079] <223> 简并寡核苷酸引物

[0080] <400>9

[0081] taccgataca tacgtagcca acac 24

[0082] <210>10

[0083] <211>26

[0084] <212>DNA

[0085] <213> 人工序列

[0086] <220>

[0087] <223> 简并寡核苷酸引物

[0088] <400>10

[0089] tcgctatact gctgtcgatt cgatac 26

[0090] <210>11

[0091] <211>1935

[0092] <212>DNA

[0093] <213>Pichia pastoris

[0094] <400>11

[0095] atgagaacac gactcaactt cctgctgctc tgtattgcca gtgttttgtc tgtgatttgg 60

[0096] atcggagtcc tccttacttg gaatgataat aatcttggcg gaatctccct aaacggaggc 120

[0097] aaggattctg cctatgatga tctgctatca ttgggaagct tcaacgacat ggaggtcgac 180

[0098] tcctatgtca ccaacatcta cgacaatgct ccagtgctag gatgtacgga tttgtcttat 240

[0099] catggattgt tgaaagtcac cccaaagcat gacttagctt gcgatttgga gttcataaga 300

[0100] gctcagattt tggacattga cgtttactcc gccataaaag acttagaaga taaagccttg 360

[0101] actgtaaaac aaaaggttga aaaacactgg tttacgtttt atggtagttc agtctttctg 420

[0102] cccgaacacg atgtgcatta cctggttaga cgagtcatct tttcggctga aggaaaggcg 480

[0103] aactctccag taacatctat catagttgct cagatatatg acaaaaactg gaacgagtta 540

[0104] aatggccatt tcttggacat cctgaaccca aatactggga aggtccagca caacacgttt 600

[0105] ccacaagttc ttcctattgc aaccaatttt gtcaaaggta agaagtttcg tggggcagaa 660

[0106] gatcctagag ttgttttgag aaagggccgt tttggacctg atcctttggt gatgttcaac 720

[0107] tccctaactc aagataacaa acgtaggaga atttttacca tttctccatt tgaccagttc 780

[0108] aaaacagtca tgtacgacat taaagactat gagatgccca ggtatgaaaa gaactgggtc 840

[0109] ccatttttct taaaagacaa tcaggaggca gttcattttg tttactcttt caaccctctg 900

[0110] agagtactca aatgcagtct tgatgacggc tcatgtgata ttgtgtttga gataccgaaa 960

[0111] gttgactcca tgtcgtctga gttgcgtggt gccacaccta tgatcaatct tcctcaggca 1020

[0112] attccgatgg cgaaggacaa agagatctgg gtttcattcc ccagaacgag aattgcaaat 1080

[0113] tgtggttgct ccaggacgac atacagacca atgctgatgc tctttgtcag agaaggttca 1140

[0114] aatttctttg ttgaactctt gtccacctct cttgattttg gtctggaggt tttaccgtat 1200

[0115] tcaggaaacg gattaccatg cagtgcggac cattccgttt taatcccaaa tagcattgat 1260

[0116] aactgggaag tcgtagatag caatggagac gatatcttga cattgtcatt cagtgaggcg 1320

[0117] gacaagagta cctctgtgat tcatatcaga gggctgtata actatctatc tgaactggat 1380

[0118] ggctatcaag gtccagaagc agaggatgaa cataatttcc agcgtatcct gagtgactta 1440

[0119] cattttgaca acaaaaccac ggtaaacaat tttataaaag tacaatcatg tgcactagat 1500

[0120] gctgccaaag gttattgtaa agaatatggg cttactcgtg gggaggcaga acgaagaagg 1560

[0121] agggtcgctg aggagagaaa gaagaaggag aaagaggaag aagaaaaaaa gaaaaagaaa 1620

[0122] gaaaaagaag aagaagagaa aaagaggatt gaagaggaga agaagaagat tgaagaaaag 1680

[0123] gaacgaaagg agaaagagaa agaagaagcg gagagaaaaa agctgcaaga aatgaaaaag 1740

[0124] aaacttgagg aaatcacaga aaaacttgaa aaaggccaga ggaataaaga gatagatcca 1800

[0125] aaagagaagc aaagggaaga agaagaaaga aaggagagag tcaggaaaat agcggagaaa 1860

[0126] caaaggaagg aggcggaaaa gaaggaggct gaaaaaaagg caaatgacaa aaaggatcta 1920

[0127] aaaataagac agtag 1935

[0128] <210>12

[0129] <211>644

[0130] <212>PRT

[0131] <213>Pichia pastoris

[0132] <400>12

[0133] Met Arg Thr Arg Leu Asn Phe Leu Leu Leu Cys Ile Ala Ser Val Leu

[0134] 1 5 10 15

[0135] Ser Val Ile Trp Ile Gly Val Leu Leu Thr Trp Asn Asp Asn Asn Leu

[0136] 20 25 30

[0137] Gly Gly Ile Ser Leu Asn Gly Gly Lys Asp Ser Ala Tyr Asp Asp Leu

[0138] 35 40 45

[0139] Leu Ser Leu Gly Ser Phe Asn Asp Met Glu Val Asp Ser Tyr Val Thr

[0140] 50 55 60

[0141] Asn Ile Tyr Asp Asn Ala Pro Val Leu Gly Cys Thr Asp Leu Ser Tyr

[0142] 65 70 75 80

[0143] His Gly Leu Leu Lys Val Thr Pro Lys His Asp Leu Ala Cys Asp Leu

[0144] 85 90 95

[0145] Glu Phe Ile Arg Ala Gln Ile Leu Asp Ile Asp Val Tyr Ser Ala Ile

[0146] 100 105 110

[0147] Lys Asp Leu Glu Asp Lys Ala Leu Thr Val Lys Gln Lys Val Glu Lys

[0148] 115 120 125

[0149] His Trp Phe Thr Phe Tyr Gly Ser Ser Val Phe Leu Pro Glu His Asp

[0150] 130 135 140

[0151] Val His Tyr Leu Val Arg Arg Val Ile Phe Ser Ala Glu Gly Lys Ala

[0152] 145 150 155 160

[0153] Asn Ser Pro Val Thr Ser Ile Ile Val Ala Gln Ile Tyr Asp Lys Asn

[0154] 165 170 175

[0155] Trp Asn Glu Leu Asn Gly His Phe Leu Asp Ile Leu Asn Pro Asn Thr

[0156] 180 185 190

[0157] Gly Lys Val Gln His Asn Thr Phe Pro Gln Val Leu Pro Ile Ala Thr

[0158] 195 200 205

[0159] Asn Phe Val Lys Gly Lys Lys Phe Arg Gly Ala Glu Asp Pro Arg Val

[0160] 210 215 220

[0161] Val Leu Arg Lys Gly Arg Phe Gly Pro Asp Pro Leu Val Met Phe Asn

[0162] 225 230 235 240

[0163] Ser Leu Thr Gln Asp Asn Lys Arg Arg Arg Ile Phe Thr Ile Ser Pro

[0164] 245 250 255

[0165] Phe Asp Gln Phe Lys Thr Val Met Tyr Asp Ile Lys Asp Tyr Glu Met

[0166] 260 265 270

[0167] Pro Arg Tyr Glu Lys Asn Trp Val Pro Phe Phe Leu Lys Asp Asn Gln

[0168] 275 280 285

[0169] Glu Ala Val His Phe Val Tyr Ser Phe Asn Pro Leu Arg Val Leu Lys

[0170] 290 295 300

[0171] Cys Ser Leu Asp Asp Gly Ser Cys Asp Ile Val Phe Glu Ile Pro Lys

[0172] 305 310 315 320

[0173] Val Asp Ser Met Ser Ser Glu Leu Arg Gly Ala Thr Pro Met Ile Asn

[0174] 325 330 335

[0175] Leu Pro Gln Ala Ile Pro Met Ala Lys Asp Lys Glu Ile Trp Val Ser

[0176] 340 345 350

[0177] Phe Pro Arg Thr Arg Ile Ala Asn Cys Gly Cys Ser Arg Thr Thr Tyr

[0178] 355 360 365

[0179] Arg Pro Met Leu Met Leu Phe Val Arg Glu Gly Ser Asn Phe Phe Val

[0180] 370 375 380

[0181] Glu Leu Leu Ser Thr Ser Leu Asp Phe Gly Leu Glu Val Leu Pro Tyr

[0182] 385 390 395 400

[0183] Ser Gly Asn Gly Leu Pro Cys Ser Ala Asp His Ser Val Leu Ile Pro

[0184] 405 410 415

[0185] Asn Ser Ile Asp Asn Trp Glu Val Val Asp Ser Asn Gly Asp Asp Ile

[0186] 420 425 430

[0187] Leu Thr Leu Ser Phe Ser Glu Ala Asp Lys Ser Thr Ser Val Ile His

[0188] 435 440 445

[0189] Ile Arg Gly Leu Tyr Asn Tyr Leu Ser Glu Leu Asp Gly Tyr Gln Gly

[0190] 450 455 460

[0191] Pro Glu Ala Glu Asp Glu His Asn Phe Gln Arg Ile Leu Ser Asp Leu

[0192] 465 470 475 480

[0193] His Phe Asp Asn Lys Thr Thr Val Asn Asn Phe Ile Lys Val Gln Ser

[0194] 485 490 495

[0195] Cys Ala Leu Asp Ala Ala Lys Gly Tyr Cys Lys Glu Tyr Gly Leu Thr
[0196] 500 505 510
[0197] Arg Gly Glu Ala Glu Arg Arg Arg Arg Val Ala Glu Glu Arg Lys Lys
[0198] 515 520 525
[0199] Lys Glu Lys Glu Glu Glu Glu Lys Lys Lys Lys Lys Glu Lys Glu Glu
[0200] 530 535 540
[0201] Glu Glu Lys Lys Arg Ile Glu Glu Glu Lys Lys Lys Ile Glu Glu Lys
[0202] 545 550 555 560
[0203] Glu Arg Lys Glu Lys Glu Lys Glu Glu Ala Glu Arg Lys Lys Leu Gln
[0204] 565 570 575
[0205] Glu Met Lys Lys Lys Leu Glu Glu Ile Thr Glu Lys Leu Glu Lys Gly
[0206] 580 585 590
[0207] Gln Arg Asn Lys Glu Ile Asp Pro Lys Glu Lys Gln Arg Glu Glu Glu
[0208] 595 600 605
[0209] Glu Arg Lys Glu Arg Val Arg Lys Ile Ala Glu Lys Gln Arg Lys Glu
[0210] 610 615 620
[0211] Ala Glu Lys Lys Glu Ala Glu Lys Lys Ala Asn Asp Lys Lys Asp Leu
[0212] 625 630 635 640
[0213] Lys Ile Arg Gln
[0214] <210>13
[0215] <211>652
[0216] <212>PRT
[0217] <213>Pichia pastoris
[0218] <400>13
[0219] Met Val Asp Leu Phe Gln Trp Leu Lys Phe Tyr Ser Met Arg Arg Leu
[0220] 1 5 10 15
[0221] Gly Gln Val Ala Ile Thr Leu Val Leu Leu Asn Leu Phe Val Phe Leu
[0222] 20 25 30
[0223] Gly Tyr Lys Phe Thr Pro Ser Thr Val Ile Gly Ser Pro Ser Trp Glu
[0224] 35 40 45
[0225] Pro Ala Val Val Pro Thr Val Phe Asn Glu Ser Tyr Leu Asp Ser Leu
[0226] 50 55 60
[0227] Gln Phe Thr Asp Ile Asn Val Asp Ser Phe Leu Ser Asp Thr Asn Gly
[0228] 65 70 75 80
[0229] Arg Ile Ser Val Thr Cys Asp Ser Leu Ala Tyr Lys Gly Leu Val Lys
[0230] 85 90 95
[0231] Thr Ser Lys Lys Lys Glu Leu Asp Cys Asp Met Ala Tyr Ile Arg Arg
[0232] 100 105 110
[0233] Lys Ile Phe Ser Ser Glu Glu Tyr Gly Val Leu Ala Asp Leu Glu Ala

[0234] 115 120 125

[0235] Gln Asp Ile Thr Glu Glu Gln Arg Ile Lys Lys His Trp Phe Thr Phe

[0236] 130 135 140

[0237] Tyr Gly Ser Ser Val Tyr Leu Pro Glu His Glu Val His Tyr Leu Val

[0238] 145 150 155 160

[0239] Arg Arg Val Leu Phe Ser Lys Val Gly Arg Ala Asp Thr Pro Val Ile

[0240] 165 170 175

[0241] Ser Leu Leu Val Ala Gln Leu Tyr Asp Lys Asp Trp Asn Glu Leu Thr

[0242] 180 185 190

[0243] Pro His Thr Leu Glu Ile Val Asn Pro Ala Thr Gly Asn Val Thr Pro

[0244] 195 200 205

[0245] Gln Thr Phe Pro Gln Leu Ile His Val Pro Ile Glu Trp Ser Val Asp

[0246] 210 215 220

[0247] Asp Lys Trp Lys Gly Thr Glu Asp Pro Arg Val Phe Leu Lys Pro Ser

[0248] 225 230 235 240

[0249] Lys Thr Gly Val Ser Glu Pro Ile Val Leu Phe Asn Leu Gln Ser Ser

[0250] 245 250 255

[0251] Leu Cys Asp Gly Lys Arg Gly Met Phe Val Thr Ser Pro Phe Arg Ser

[0252] 260 265 270

[0253] Asp Lys Val Asn Leu Leu Asp Ile Glu Asp Lys Glu Arg Pro Asn Ser

[0254] 275 280 285

[0255] Glu Lys Asn Trp Ser Pro Phe Phe Leu Asp Asp Val Glu Val Ser Lys

[0256] 290 295 300

[0257] Tyr Ser Thr Gly Tyr Val His Phe Val Tyr Ser Phe Asn Pro Leu Lys

[0258] 305 310 315 320

[0259] Val Ile Lys Cys Ser Leu Asp Thr Gly Ala Cys Arg Met Ile Tyr Glu

[0260] 325 330 335

[0261] Ser Pro Glu Glu Gly Arg Phe Gly Ser Glu Leu Arg Gly Ala Thr Pro

[0262] 340 345 350

[0263] Met Val Lys Leu Pro Val His Leu Ser Leu Pro Lys Gly Lys Glu Val

[0264] 355 360 365

[0265] Trp Val Ala Phe Pro Arg Thr Arg Leu Arg Asp Cys Gly Cys Ser Arg

[0266] 370 375 380

[0267] Thr Thr Tyr Arg Pro Val Leu Thr Leu Phe Val Lys Glu Gly Asn Lys

[0268] 385 390 395 400

[0269] Phe Tyr Thr Glu Leu Ile Ser Ser Ser Ile Asp Phe His Ile Asp Val

[0270] 405 410 415

[0271] Leu Ser Tyr Asp Ala Lys Gly Glu Ser Cys Ser Gly Ser Ile Ser Val

[0272] 420 425 430

[0273] Leu Ile Pro Asn Gly Ile Asp Ser Trp Asp Val Ser Lys Lys Gln Gly
[0274] 435 440 445
[0275] Gly Lys Ser Asp Ile Leu Thr Leu Thr Leu Ser Glu Ala Asp Arg Asn
[0276] 450 455 460
[0277] Thr Val Val Val His Val Lys Gly Leu Leu Asp Tyr Leu Leu Val Leu
[0278] 465 470 475 480
[0279] Asn Gly Glu Gly Pro Ile His Asp Ser His Ser Phe Lys Asn Val Leu
[0280] 485 490 495
[0281] Ser Thr Asn His Phe Lys Ser Asp Thr Thr Leu Leu Asn Ser Val Lys
[0282] 500 505 510
[0283] Ala Ala Glu Cys Ala Ile Phe Ser Ser Arg Asp Tyr Cys Lys Lys Tyr
[0284] 515 520 525
[0285] Gly Glu Thr Arg Gly Glu Pro Ala Arg Tyr Ala Lys Gln Met Glu Asn
[0286] 530 535 540
[0287] Glu Arg Lys Glu Lys Glu Lys Lys Glu Lys Glu Ala Lys Glu Lys Leu
[0288] 545 550 555 560
[0289] Glu Ala Glu Lys Ala Glu Met Glu Glu Ala Val Arg Lys Ala Gln Glu
[0290] 565 570 575
[0291] Ala Ile Ala Gln Lys Glu Arg Glu Lys Glu Glu Ala Glu Gln Glu Lys
[0292] 580 585 590
[0293] Lys Ala Gln Gln Glu Ala Lys Glu Lys Glu Ala Glu Glu Lys Ala Ala
[0294] 595 600 605
[0295] Lys Glu Lys Glu Ala Lys Glu Asn Glu Ala Lys Lys Lys Ile Ile Val
[0296] 610 615 620
[0297] Glu Lys Leu Ala Lys Glu Gln Glu Glu Ala Glu Lys Leu Glu Ala Lys
[0298] 625 630 635 640
[0299] Lys Lys Leu Tyr Gln Leu Gln Glu Glu Glu Arg Ser
[0300] 645 650
[0301] <210>14
[0302] <211>594
[0303] <212>PRT
[0304] <213>Pichia pastoris
[0305] <400>14
[0306] Met Arg Ile Arg Ser Asn Val Leu Leu Leu Ser Thr Ala Gly Ala Leu
[0307] 1 5 10 15
[0308] Ala Leu Val Trp Phe Ala Val Val Phe Ser Trp Asp Asp Lys Ser Ile
[0309] 20 25 30
[0310] Phe Gly Ile Pro Thr Pro Gly His Ala Val Ala Ser Ala Tyr Asp Ser
[0311] 35 40 45

[0312] Ser Val Thr Leu Gly Thr Phe Asn Asp Met Glu Val Asp Ser Tyr Val
[0313] 50 55 60
[0314] Thr Asn Ile Tyr Asp Asn Ala Pro Val Leu Gly Cys Tyr Asp Leu Ser
[0315] 65 70 75 80
[0316] Tyr His Gly Leu Leu Lys Val Ser Pro Lys His Glu Ile Leu Cys Asp
[0317] 85 90 95
[0318] Met Lys Phe Ile Arg Ala Arg Val Leu Glu Thr Glu Ala Tyr Ala Ala
[0319] 100 105 110
[0320] Leu Lys Asp Leu Glu His Lys Lys Leu Thr Glu Glu Glu Lys Ile Glu
[0321] 115 120 125
[0322] Lys His Trp Phe Thr Phe Tyr Gly Ser Ser Val Phe Leu Pro Asp His
[0323] 130 135 140
[0324] Asp Val His Tyr Leu Val Arg Arg Val Val Phe Ser Gly Glu Gly Lys
[0325] 145 150 155 160
[0326] Ala Asn Arg Pro Ile Thr Ser Ile Leu Val Ala Gln Ile Tyr Asp Lys
[0327] 165 170 175
[0328] Asn Trp Asn Glu Leu Asn Gly His Phe Leu Asn Val Leu Asn Pro Asn
[0329] 180 185 190
[0330] Thr Gly Lys Leu Gln His His Ala Phe Pro Gln Val Leu Pro Ile Ala
[0331] 195 200 205
[0332] Val Asn Trp Asp Arg Asn Ser Lys Tyr Arg Gly Gln Glu Asp Pro Arg
[0333] 210 215 220
[0334] Val Val Leu Arg Arg Gly Arg Phe Gly Pro Asp Pro Leu Val Met Phe
[0335] 225 230 235 240
[0336] Asn Thr Leu Thr Gln Asn Asn Lys Leu Arg Arg Leu Phe Thr Ile Ser
[0337] 245 250 255
[0338] Pro Phe Asp Gln Tyr Lys Thr Val Met Tyr Arg Thr Asn Ala Phe Lys
[0339] 260 265 270
[0340] Met Gln Thr Thr Glu Lys Asn Trp Val Pro Phe Phe Leu Lys Asp Asp
[0341] 275 280 285
[0342] Gln Glu Ser Val His Phe Val Tyr Ser Phe Asn Pro Leu Arg Val Leu
[0343] 290 295 300
[0344] Asn Cys Ser Leu Asp Asn Gly Ala Cys Asp Val Leu Phe Glu Leu Pro
[0345] 305 310 315 320
[0346] His Asp Phe Gly Met Ser Ser Glu Leu Arg Gly Ala Thr Pro Met Leu
[0347] 325 330 335
[0348] Asn Leu Pro Gln Ala Ile Pro Met Ala Asp Asp Lys Glu Ile Trp Val
[0349] 340 345 350
[0350] Ser Phe Pro Arg Thr Arg Ile Ser Asp Cys Gly Cys Ser Glu Thr Met

[0351] 355 360 365

[0352] Tyr Arg Pro Met Leu Met Leu Phe Val Arg Glu Gly Thr Asn Phe Phe

[0353] 370 375 380

[0354] Ala Glu Leu Leu Ser Ser Ser Ile Asp Phe Gly Leu Glu Val Ile Pro

[0355] 385 390 395 400

[0356] Tyr Thr Gly Asp Gly Leu Pro Cys Ser Ser Gly Gln Ser Val Leu Ile

[0357] 405 410 415

[0358] Pro Asn Ser Ile Asp Asn Trp Glu Val Thr Gly Ser Asn Gly Glu Asp

[0359] 420 425 430

[0360] Ile Leu Ser Leu Thr Phe Ser Glu Ala Asp Lys Ser Thr Ser Val Val

[0361] 435 440 445

[0362] His Ile Arg Gly Leu Tyr Lys Tyr Leu Ser Glu Leu Asp Gly Tyr Gly

[0363] 450 455 460

[0364] Gly Pro Glu Ala Glu Asp Glu His Asn Phe Gln Arg Ile Leu Ser Asp

[0365] 465 470 475 480

[0366] Leu His Phe Asp Gly Lys Lys Thr Ile Glu Asn Phe Lys Lys Val Gln

[0367] 485 490 495

[0368] Ser Cys Ala Leu Asp Ala Ala Lys Ala Tyr Cys Lys Glu Tyr Gly Val

[0369] 500 505 510

[0370] Thr Arg Gly Glu Glu Asp Arg Leu Lys Asn Lys Glu Lys Glu Arg Lys

[0371] 515 520 525

[0372] Ile Glu Glu Lys Arg Lys Lys Glu Glu Glu Arg Lys Lys Lys Glu Glu

[0373] 530 535 540

[0374] Glu Lys Lys Lys Lys Glu Glu Glu Glu Lys Lys Lys Lys Glu Glu Glu

[0375] 545 550 555 560

[0376] Glu Glu Glu Glu Lys Arg Leu Lys Glu Leu Lys Lys Lys Leu Lys Glu

[0377] 565 570 575

[0378] Leu Gln Glu Glu Leu Glu Lys Gln Lys Asp Glu Val Lys Asp Thr Lys

[0379] 580 585 590

[0380] Ala Lys

[0381] <210>15

[0382] <211>503

[0383] <212>PRT

[0384] <213>Pichia pastoris

[0385] <400>15

[0386] Met Lys Leu Asp Thr Gln Gln Ile Ser His Leu Leu Ser Arg Gln Met

[0387] 1 5 10 15

[0388] Tyr His Leu Ala Pro Arg Lys Lys Leu Leu Ile Trp Gly Gly Ser Leu

[0389] 20 25 30

[0390] Gly Phe Val Leu Leu Leu Leu Ile Val Ala Ser Ser His Gln Arg Ile
[0391] 35 40 45
[0392] Arg Ser Thr Ile Leu His Arg Thr Pro Ile Ser Thr Leu Pro Val Ile
[0393] 50 55 60
[0394] Ser Gln Glu Val Ile Thr Ala Asp Tyr His Pro Thr Leu Leu Thr Gly
[0395] 65 70 75 80
[0396] Phe Ile Pro Thr Asp Ser Asp Asp Ser Asp Cys Ala Asp Phe Ser Pro
[0397] 85 90 95
[0398] Ser Gly Val Ile Tyr Ser Thr Asp Lys Leu Val Leu His Asp Ser Leu
[0399] 100 105 110
[0400] Lys Asp Ile Arg Asp Ser Leu Leu Lys Thr Gln Tyr Lys Asp Leu Val
[0401] 115 120 125
[0402] Thr Leu Glu Asp Glu Glu Lys Met Asn Ile Asp Asp Ile Leu Lys Arg
[0403] 130 135 140
[0404] Trp Tyr Thr Leu Ser Gly Ser Ser Val Trp Ile Pro Gly Met Lys Ala
[0405] 145 150 155 160
[0406] His Leu Val Val Ser Arg Val Met Tyr Leu Gly Thr Asn Gly Arg Ser
[0407] 165 170 175
[0408] Asp Pro Leu Val Ser Phe Val Arg Val Gln Leu Phe Asp Pro Asp Phe
[0409] 180 185 190
[0410] Asn Glu Leu Lys Asp Ile Ala Leu Lys Phe Ser Asp Lys Pro Asp Gly
[0411] 195 200 205
[0412] Thr Val Ile Phe Pro Tyr Ile Leu Pro Val Asp Ile Pro Arg Glu Gly
[0413] 210 215 220
[0414] Ser Arg Trp Leu Gly Pro Glu Asp Ala Lys Ile Ala Val Asn Pro Glu
[0415] 225 230 235 240
[0416] Thr Pro Asp Asp Pro Ile Val Ile Phe Asn Met Gln Asn Ser Val Asn
[0417] 245 250 255
[0418] Arg Ala Met Tyr Gly Phe Tyr Pro Phe Arg Pro Glu Asn Lys Gln Val
[0419] 260 265 270
[0420] Leu Phe Ser Ile Lys Asp Glu Glu Pro Arg Lys Lys Glu Lys Asn Trp
[0421] 275 280 285
[0422] Thr Pro Phe Phe Val Pro Gly Ser Pro Thr Thr Val Asn Phe Val Tyr
[0423] 290 295 300
[0424] Asp Leu Gln Lys Leu Thr Ile Leu Lys Cys Ser Ile Ile Thr Gly Ile
[0425] 305 310 315 320
[0426] Cys Glu Lys Glu Phe Val Ser Gly Asp Asp Gly Gln Asn His Gly Ile
[0427] 325 330 335
[0428] Gly Ile Phe Arg Gly Gly Ser Asn Leu Val Pro Phe Pro Thr Ser Phe

[0429] 340 345 350

[0430] Thr Asp Lys Asp Val Trp Val Gly Phe Pro Lys Thr His Met Glu Ser

[0431] 355 360 365

[0432] Cys Gly Cys Ser Ser His Ile Tyr Arg Pro Tyr Leu Met Val Leu Val

[0433] 370 375 380

[0434] Arg Lys Gly Asp Phe Tyr Tyr Lys Ala Phe Val Ser Thr Pro Leu Asp

[0435] 385 390 395 400

[0436] Phe Gly Ile Asp Val Arg Ser Trp Glu Ser Ala Glu Ser Thr Ser Cys

[0437] 405 410 415

[0438] Gln Thr Ala Lys Asn Val Leu Ala Val Asn Ser Ile Ser Asn Trp Asp

[0439] 420 425 430

[0440] Leu Leu Asp Asp Gly Leu Asp Lys Asp Tyr Met Thr Ile Thr Leu Ser

[0441] 435 440 445

[0442] Glu Ala Asp Val Val Asn Ser Val Leu Arg Val Arg Gly Ile Ala Lys

[0443] 450 455 460

[0444] Phe Val Asp Asn Leu Thr Met Asp Asp Gly Ser Thr Thr Leu Ser Thr

[0445] 465 470 475 480

[0446] Ser Asn Lys Ile Asp Glu Cys Ala Thr Thr Gly Ser Lys Gln Tyr Cys

[0447] 485 490 495

[0448] Gln Arg Tyr Gly Glu Leu His

[0449] 500

```
1 ATG AGA ACA CGA CTC AAC TTC CTG CTG CTC TGT ATT GCC AGT GTT TTG TCT GTG ATT TGG ATC GGA GTC CTC CTT ACT
1 Met Arg Thr Arg Leu Asn Phe Leu Leu Leu Cys Ile Ala Ser Val Leu Ser Val Ile Trp Ile Gly Val Leu Leu Thr

79 TGG AAT GAT AAT AAT CTT GGC GGA ATC TCC CTA AAC GGA GGC AAG GAT TCT GCC TAT GAT GAT CTG CTA TCA TTG GGA
27 Trp Asn Asp Asn Asn Leu Gly Gly Ile Ser Leu Asn Gly Gly Lys Asp Ser Ala Tyr Asp Asp Leu Leu Ser Leu Gly

157 AGC TTC AAC GAC ATG GAG GTC GAC TCC TAT GTC ACC AAC ATC TAC GAC AAT GCT CCA GTG CTA GGA TGT ACG GAT TTG
53 Ser Phe Asn Asp Met Glu Val Asp Ser Tyr Val Thr Asn Ile Tyr Asp Asn Ala Pro Val Leu Gly Cys Thr Asp Leu

235 TCT TAT CAT GGA TTG TTG AAA GTC ACC CCA AAG CAT GAC TTA GCT TGC GAT TTG GAG TTC ATA AGA GCT CAG ATT TTG
79 Ser Tyr His Gly Leu Leu Lys Val Thr Pro Lys His Asp Leu Ala Cys Asp Leu Glu Phe Ile Arg Ala Gln Ile Leu

313 GAC ATT GAC GTT TAC TCC GCC ATA AAA GAC TTA GAA GAT AAA GCC TTG ACT GTA AAA CAA AAG GTT GAA AAA CAC TGG
105 Asp Ile Asp Val Tyr Ser Ala Ile Lys Asp Leu Glu Asp Lys Ala Leu Thr Val Lys Gln Lys Val Glu Lys His Trp

391 TTT ACG TTT TAT GGT AGT TCA GTC TTT CTG CCC GAA CAC GAT GTG CAT TAC CTG GTT AGA CGA GTC ATC TTT TCG GCT
131 Phe Thr Phe Tyr Gly Ser Ser Val Phe Leu Pro Glu His Asp Val His Tyr Leu Val Arg Arg Val Ile Phe Ser Ala

469 GAA GGA AAG GCG AAC TCT CCA GTA ACA TCT ATC ATA GTT GCT CAG ATA TAT GAC AAA AAC TGG AAC GAG TTA AAT GGC
157 Glu Gly Lys Ala Asn Ser Pro Val Thr Ser Ile Ile Val Ala Gln Ile Tyr Asp Lys Asn Trp Asn Glu Leu Asn Gly

547 CAT TTC TTG GAC ATC CTG AAC CCA AAT ACT GGG AAG GTC CAG CAC AAC ACG TTT CCA CAA GTT CTT CCT ATT GCA ACC
183 His Phe Leu Asp Ile Leu Asn Pro Asn Thr Gly Lys Val Gln His Asn Thr Phe Pro Gln Val Leu Pro Ile Ala Thr

625 AAT TTT GTC AAA GGT AAG AAG TTT CGT GGG GCA GAA GAT CCT AGA GTT GTT TTG AGA AAG GGC CGT TTT GGA CCT GAT
209 Asn Phe Val Lys Gly Lys Lys Phe Arg Gly Ala Glu Asp Pro Arg Val Val Leu Arg Lys Gly Arg Phe Gly Pro Asp

703 CCT TTG GTG ATG TTC AAC TCC CTA ACT CAA GAT AAC AAA CGT AGG AGA ATT TTT ACC ATT TCT CCA TTT GAC CAG TTC
235 Pro Leu Val Met Phe Asn Ser Leu Thr Gln Asp Asn Lys Arg Arg Arg Ile Phe Thr Ile Ser Pro Phe Asp Gln Phe

781 AAA ACA GTC ATG TAC GAC ATT AAA GAC TAT GAG ATG CCC AGG TAT GAA AAG AAC TGG GTC CCA TTT TTC TTA AAA GAC
261 Lys Thr Val Met Tyr Asp Ile Lys Asp Tyr Glu Met Pro Arg Tyr Glu Lys Asn Trp Val Pro Phe Phe Leu Lys Asp

859 AAT CAG GAG GCA GTT CAT TTT GTT TAC TCT TTC AAC CCT CTG AGA GTA CTC AAA TGC AGT CTT GAT GAC GGC TCA TGT
287 Asn Gln Glu Ala Val His Phe Val Tyr Ser Phe Asn Pro Leu Arg Val Leu Lys Cys Ser Leu Asp Asp Gly Ser Cys

937 GAT ATT GTG TTT GAG ATA CCG AAA GTT GAC TCC ATG TCG TCT GAG TTG CGT GGT GCC ACA CCT ATG ATC AAT CTT CCT
313 Asp Ile Val Phe Glu Ile Pro Lys Val Asp Ser Met Ser Ser Glu Leu Arg Gly Ala Thr Pro Met Ile Asn Leu Pro

1015 CAG GCA ATT CCG ATG GCG AAG GAC AAA GAG ATC TGG GTT TCA TTC CCC AGA ACG AGA ATT GCA AAT TGT GGT TGC TCC
339 Gln Ala Ile Pro Met Ala Lys Asp Lys Glu Ile Trp Val Ser Phe Pro Arg Thr Arg Ile Ala Asn Cys Gly Cys Ser

1093 AGG ACG ACA TAC AGA CCA ATG CTG ATG CTC TTT GTC AGA GAA GGT TCA AAT TTC TTT GTT GAA CTC TTG TCC ACC TCT
365 Arg Thr Thr Tyr Arg Pro Met Leu Met Leu Phe Val Arg Glu Gly Ser Asn Phe Phe Val Glu Leu Leu Ser Thr Ser

1171 CTT GAT TTT GGT CTG GAG GTT TTA CCG TAT TCA GGA AAC GGA TTA CCA TGC AGT GCG GAC CAT TCC GTT TTA ATC CCA
391 Leu Asp Phe Gly Leu Glu Val Leu Pro Tyr Ser Gly Asn Gly Leu Pro Cys Ser Ala Asp His Ser Val Leu Ile Pro

1249 AAT AGC ATT GAT AAC TGG GAA GTC GTA GAT AGC AAT GGA GAC GAT ATC TTG ACA TTG TCA TTC AGT GAG GCG GAC AAG
417 Asn Ser Ile Asp Asn Trp Glu Val Val Asp Ser Asn Gly Asp Asp Ile Leu Thr Leu Ser Phe Ser Glu Ala Asp Lys

1327 AGT ACC TCT GTG ATT CAT ATC AGA GGG CTG TAT AAC TAT CTA TCT GAA CTG GAT GGC TAT CAA GGT CCA GAA GCA GAG
443 Ser Thr Ser Val Ile His Ile Arg Gly Leu Tyr Asn Tyr Leu Ser Glu Leu Asp Gly Tyr Gln Gly Pro Glu Ala Glu

1405 GAT GAA CAT AAT TTC CAG CGT ATC CTG AGT GAC TTA CAT TTT GAC AAC AAA ACC ACG GTA AAC AAT TTT ATA AAA GTA
469 Asp Glu His Asn Phe Gln Arg Ile Leu Ser Asp Leu His Phe Asp Asn Lys Thr Thr Val Asn Asn Phe Ile Lys Val

1483 CAA TCA TGT GCA CTA GAT GCT GCC AAA GGT TAT TGT AAA GAA TAT GGG CTT ACT CGT GGG GAG GCA GAA CGA AGA AGG
495 Gln Ser Cys Ala Leu Asp Ala Ala Lys Gly Tyr Cys Lys Glu Tyr Gly Leu Thr Arg Gly Glu Ala Glu Arg Arg Arg

1561 AGG GTC GCT GAG GAG AGA AAG AAG AAG GAG AAA GAG GAA GAA GAA AAA AAG AAA AAG AAA GAA AAA GAA GAA GAA GAG
521 Arg Val Ala Glu Glu Arg Lys Lys Lys Glu Lys Glu Glu Glu Glu Lys Lys Lys Lys Lys Glu Lys Glu Glu Glu Glu

1639 AAA AAG AGG ATT GAA GAG GAG AAG AAG AAG ATT GAA GAA AAG GAA CGA AAG GAG AAA GAG AAA GAA GAA GCG GAG AGA
547 Lys Lys Arg Ile Glu Glu Glu Lys Lys Lys Ile Glu Glu Lys Glu Arg Lys Glu Lys Glu Lys Glu Glu Ala Glu Arg

1717 AAA AAG CTG CAA GAA ATG AAA AAG AAA CTT GAG GAA ATC ACA GAA AAA CTT GAA AAA GGC CAG AGG AAT AAA GAG ATA
573 Lys Lys Leu Gln Glu Met Lys Lys Lys Leu Glu Glu Ile Thr Glu Lys Leu Glu Lys Gly Gln Arg Asn Lys Glu Ile

1795 GAT CCA AAA GAG AAG CAA AGG GAA GAA GAA GAA AGA AAG GAG AGA GTC AGG AAA ATA GCG GAG AAA CAA AGG AAG GAG
599 Asp Pro Lys Glu Lys Gln Arg Glu Glu Glu Glu Arg Lys Glu Arg Val Arg Lys Ile Ala Glu Lys Gln Arg Lys Glu

1873 GCG GAA AAG AAG GAG GCT GAA AAA AAG GCA AAT GAC AAA AAG GAT CTA AAA ATA AGA CAG TAG
625 Ala Glu Lys Lys Glu Ala Glu Lys Lys Ala Asn Asp Lys Lys Asp Leu Lys Ile Arg Gln •••
```

图 1

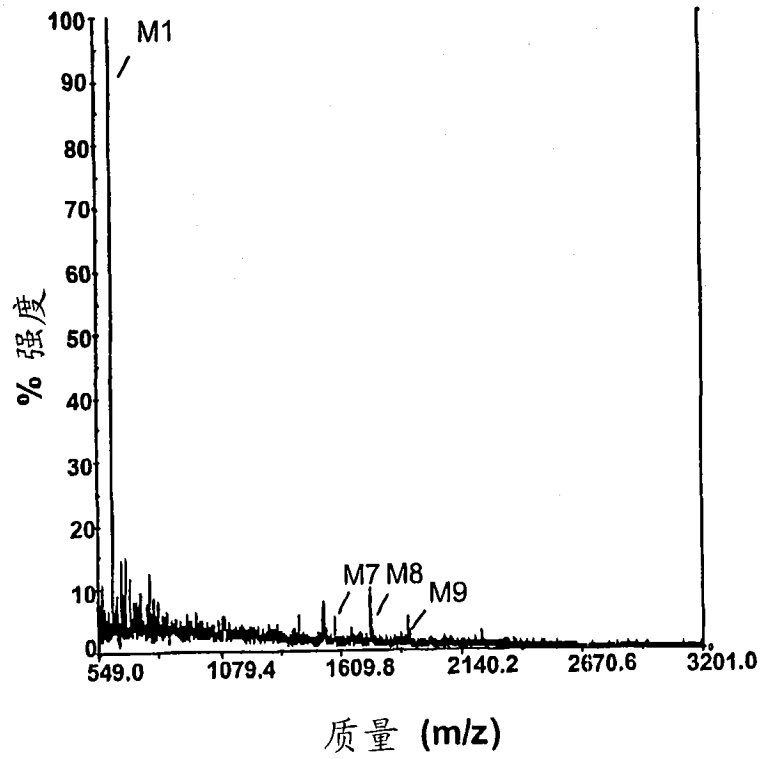

100
90
80
70
60
50
40
30
20
10
0
% 强度
M1
M7
M8
M9
549.0
1079.4
1609.8
2140.2
2670.6
3201.0
质量 (m/z)

图 3

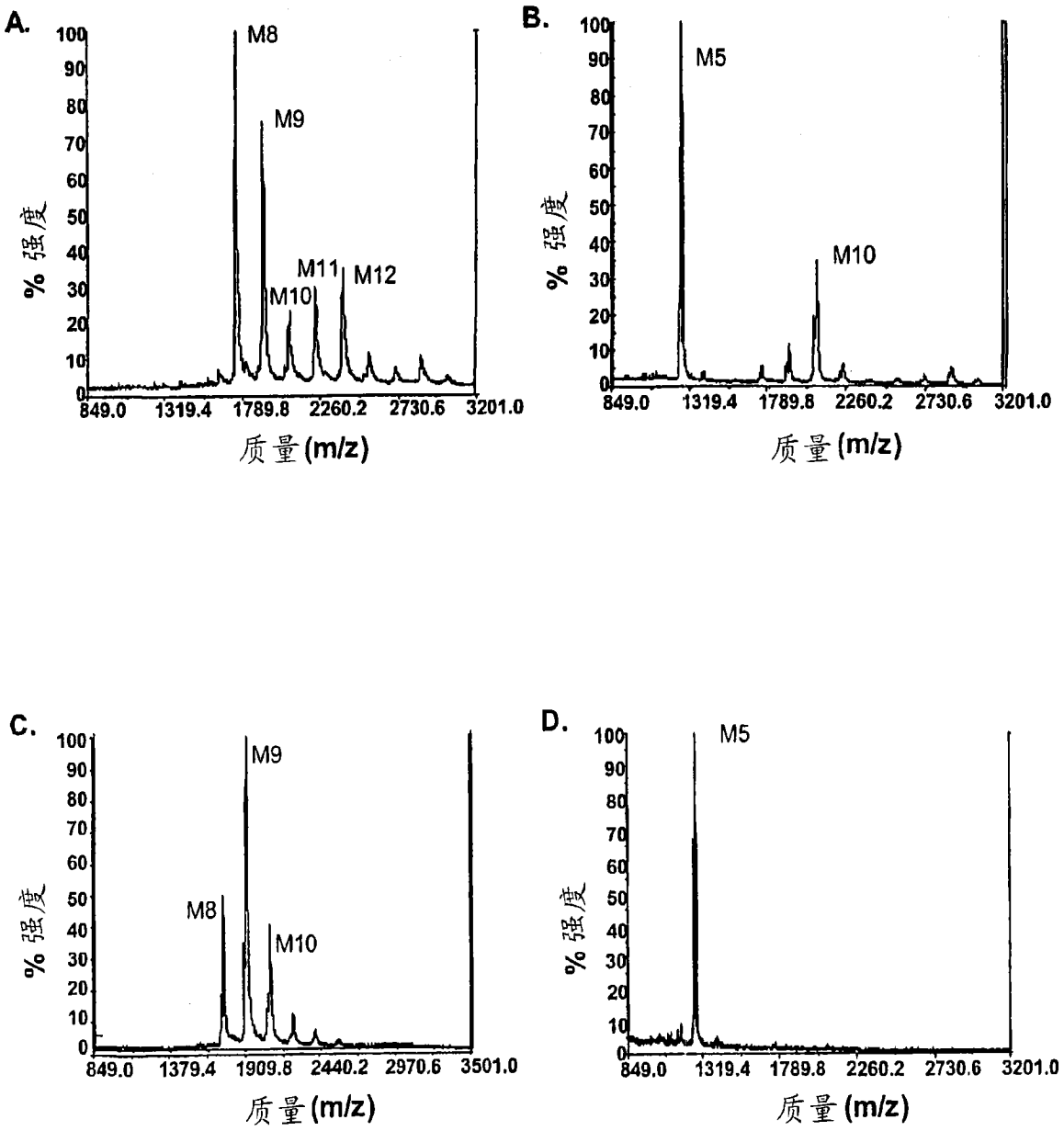

A.
M8
M9
M10
M11
M12
% 强度
质量(m/z)
849.0 1319.4 1789.8 2260.2 2730.6 3201.0
B.
M5
M10
% 强度
质量(m/z)
849.0 1319.4 1789.8 2260.2 2730.6 3201.0
C.
M8
M9
M10
% 强度
质量(m/z)
849.0 1379.4 1909.8 2440.2 2970.6 3501.0
D.
M5
% 强度
质量(m/z)
849.0 1319.4 1789.8 2260.2 2730.6 3201.0

图 2

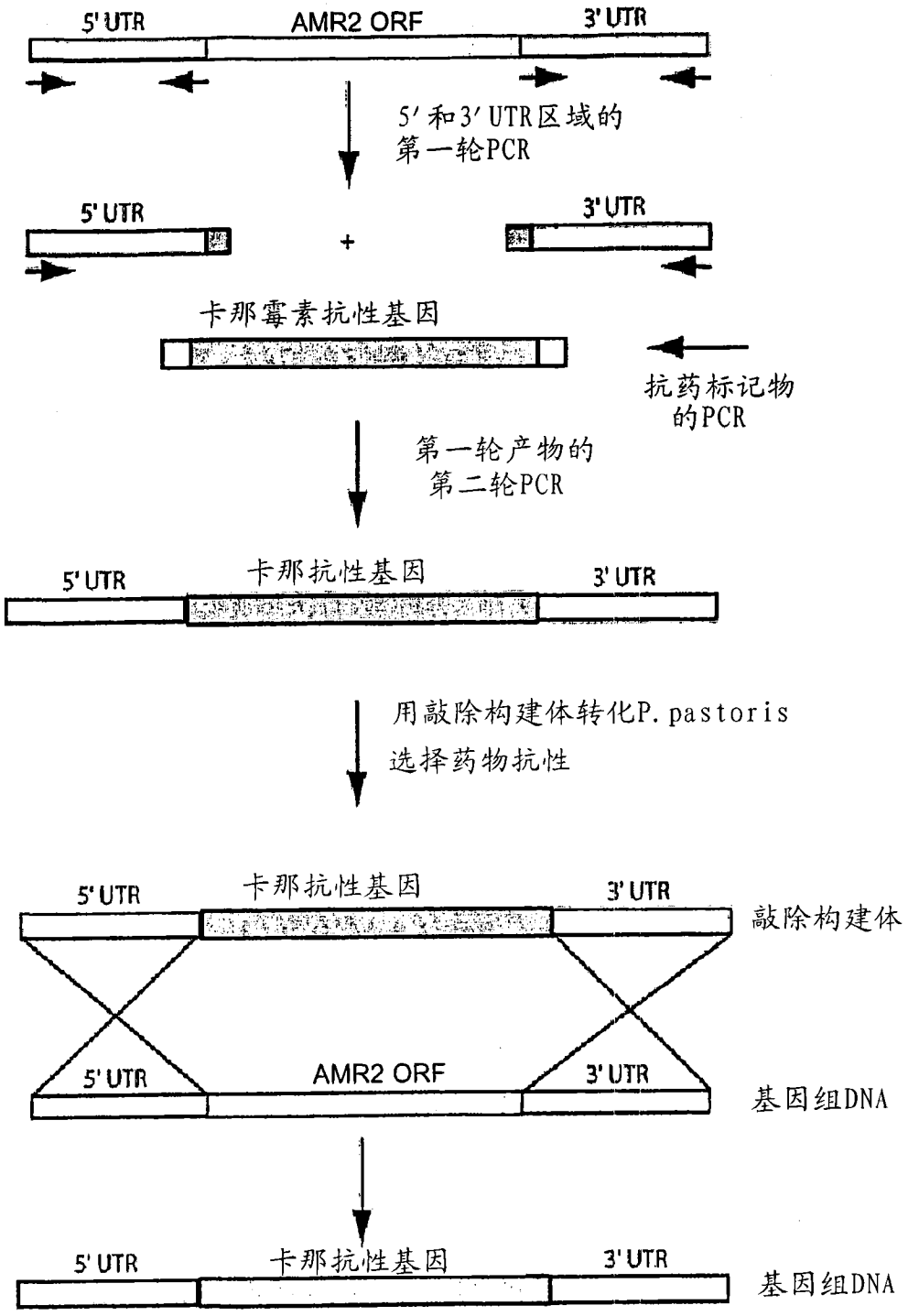

5' UTR
AMR2 ORF
3' UTR
5′和3′UTR区域的
第一轮PCR
5' UTR
3' UTR
+
卡那霉素抗性基因
抗药标记物
的PCR
第一轮产物的
第二轮PCR
5' UTR
卡那抗性基因
3' UTR
用敲除构建体转化P. pastoris
选择药物抗性
5' UTR
卡那抗性基因
3' UTR
敲除构建体
5' UTR
AMR2 ORF
3' UTR
基因组DNA
5' UTR
卡那抗性基因
3' UTR
基因组DNA

图 4

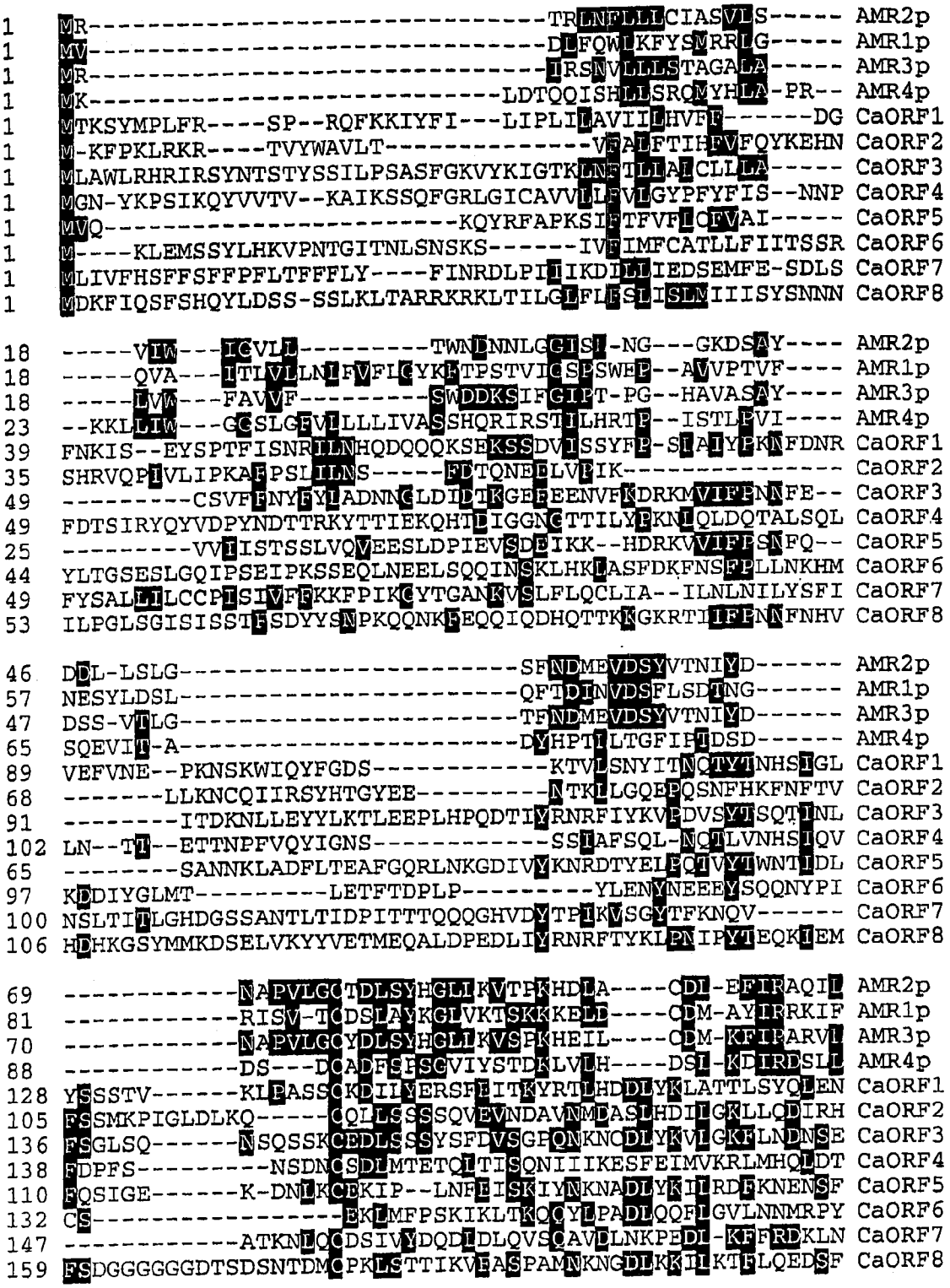
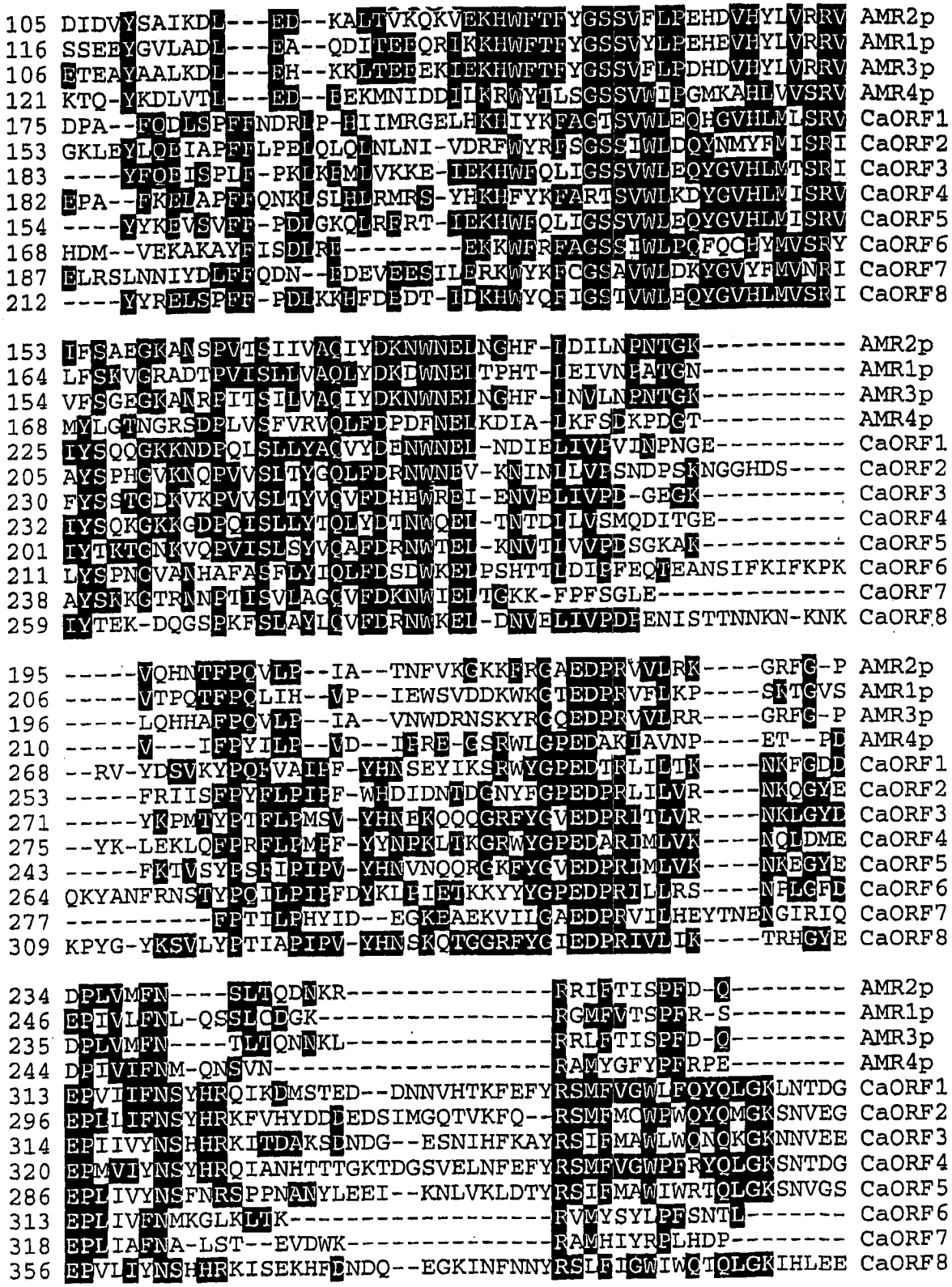

AMR2p
AMR1p
AMR3p
AMR4p
CaORF1
CaORF2
CaORF3
CaORF4
CaORF5
CaORF6
CaORF7
CaORF8

图 5

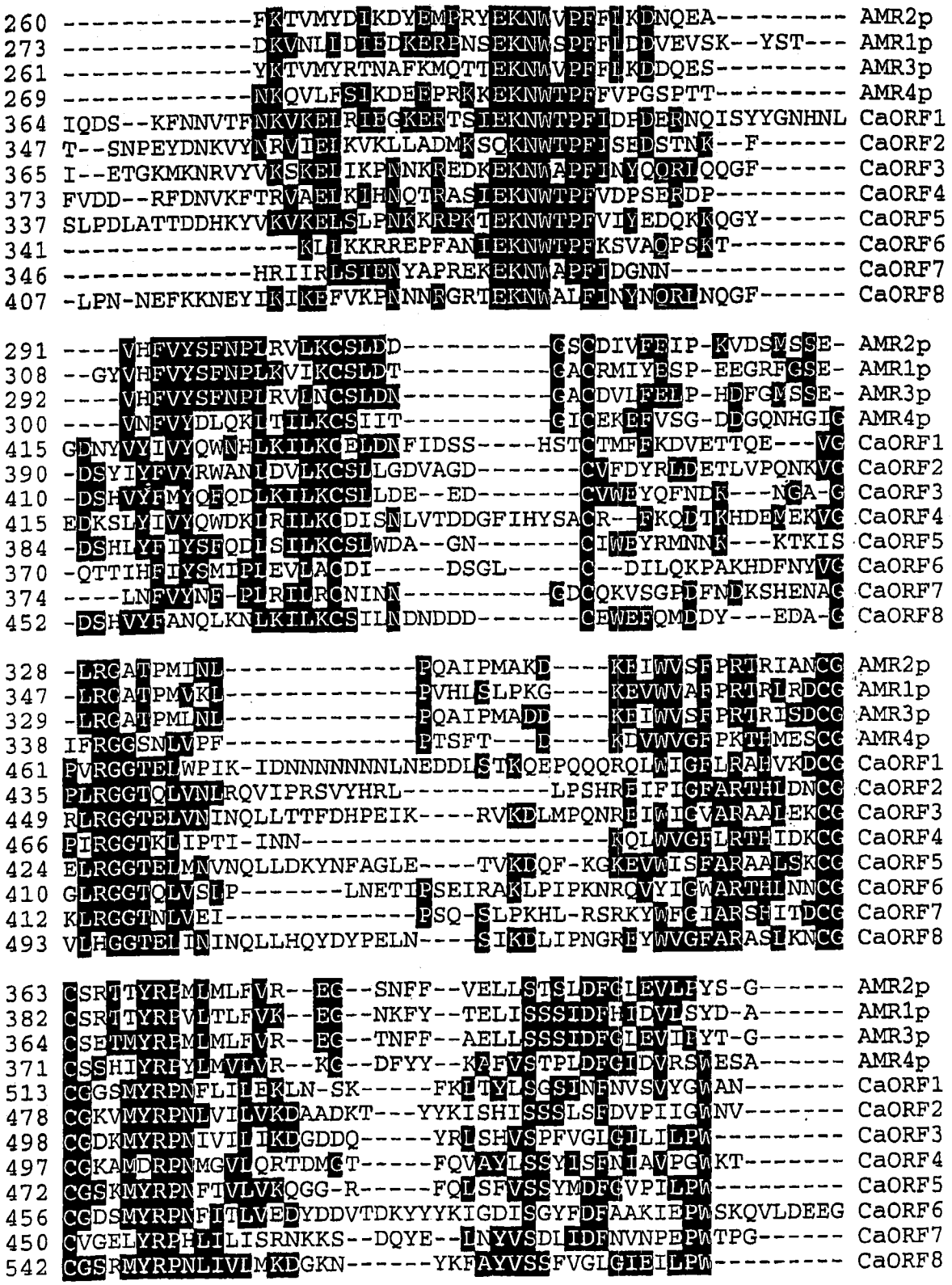

105 DIDVYSAIKDL---ED--KALTVKQKVEKHWFTFYGSSVFLPEHDVHYLVRRV AMR2p
116 SSEEYGVLADL---EA--QDITEEQRIKKHWFTFYGSSVYLPEHEVHYLVRRV AMR1p
106 ETEAYAALKDL---EH--KKLTEEEKIEKHWFTFYGSSVFLPDHDVHYLVRRV AMR3p
121 KTQ-YKDLVTL---ED--EEKMNIDDILKRWYTLSGSSVWIPGMKAHLVVSRV AMR4p
175 DPA--FQDLSPFFNDRLP-HIIMRGELHKHIYKFAGTSVWLEQHGVHLMLSRV CaORF1
153 GKLEYLQEIAPFFLPELQLQLNLNI-VDRFWYRFSGSSIWLDQYNMYFMISRI CaORF2
183 ----YFQEISPLF-PKLKEMLVKKE-IEKHWFQLIGSSVWLEQYGVHLMTSRI CaORF3
182 EPA--FKELAPFFQNKLSIHLRMRS-YHKHFYKFARTSVWLKDYGVHLMISRV CaORF4
154 ----YYKEVSVFF-PDLGKQLRFRT-IEKHWFQLIGSSVWLEQYGVHLMISRV CaORF5
168 HDM--VEKAKAYFISDIRE--------EKKWFRFAGSSIWLPQFQCHYMVSRY CaORF6
187 ELRSLNNIYDLFFQDN--EDEVEESILERKWYKFCGSAVWLDKYGVYFMVNRI CaORF7
212 ----YYRELSPFF-PDLKKHFDEDT-IDKHWYQFIGSTVWLEQYGVHLMVSRI CaORF8

153 IFSAEGKANSPVTSIIVAQIYDKNWNELNGHF-LDILNPNTGK---------- AMR2p
164 LFSKVGRADTPVISLLVAQLYDKDWNEITPHT-IEIVNPATGN---------- AMR1p
154 VFSGEGKANRPITSILVAQIYDKNWNELNGHF-INVLNPNTGK---------- AMR3p
168 MYLGTNGRSDPLVSFVRVQLFDPDFNELKDIA-LKFSDKPDGT---------- AMR4p
225 IYSQQGKKNDPQLSLLYAQVYDENWNEL-NDIELIVPVINPNGE--------- CaORF1
205 AYSPHGVKNQPVVSLTYGQLFDRNWNEV-KNINLLVPSNDPSKNGGHDS---- CaORF2
230 FYSSTGDKVKPVVSLTYVQVFDHEWREI-ENVELIVPD-GEGK---------- CaORF3
232 IYSQKGKKGDPQISLLYTQLYDTNWQEL-TNTDILVSMQDITGE--------- CaORF4
201 IYTKTGNKVQPVISLSYVQAFDRNWTEL-KNVTIVVPDSGKAK---------- CaORF5
211 LYSPNGVANHAFASFLYIQLFDSDWKELPSHTTLDIPFEQTEANSIFKIFKPK CaORF6
238 AYSKKGTRNNPTISVLAGQVFDKNWIELTGKK-FPFSGLE------------- CaORF7
259 IYTEK-DQGSPKFSLAYLQVFDRNWKEL-DNVELIVPDPENISTTNNKN-KNK CaORF8

195 -----VQHNTFPQVLP--IA--TNFVKGKKFRGAEDPRVVLRK----GRFG-P AMR2p
206 -----VTPQTFPQLIH--VP--IEWSVDDKWKGTEDPFVFLKP-----SKTGVS AMR1p
196 -----LQHHAFPQVLP--IA--VNWDRNSKYRGQEDPRVVLRR----GRFG-P AMR3p
210 -----V----IFPYILP--VD--IERE-GSRWLGPEDAKIAVNP----ET--PD AMR4p
268 --RV-YDSVKYPQFVAIPF-YHNSEYIKSRWYGPEDTRLILTK----NKFGDD CaORF1
253 -----FRIISFPYFLPIPF-WHDIDNTDGNYFGPEDPRLILVR----NKQGYE CaORF2
271 -----YKPMTYPTFLPMSV-YHNFKQQQGRFYGVEDPRITLVR----NKLGYD CaORF3
275 --YK-LEKLQFPRFLPMPF-YYNPKLTKGRWYGPEDARIMLVK----NQLDME CaORF4
243 -----FKTVSYPSFIPIPV-YHNVNQQRGKFYGVEDPRIMLVK----NKEGYE CaORF5
264 QKYANFRNSTYPQILPIPFDYKLPIETKKYYYGPEDPRILLRS----NPLGFD CaORF6
277 ----------FPTILPHYID--EGKEAEKVILGAEDPRVILHEYTNENGIRIQ CaORF7
309 KPYG-YKSVLYPTIAPIPV-YHNSKQTGGRFYGIEDPRIVLIK----TRHGYE CaORF8

234 DPLVMFN----SLTQDNKR--------------RRIFTISPFD-Q-------- AMR2p
246 EPIVLFNL-QSSLCDGK---------------RGMFVTSPFR-S-------- AMR1p
235 DPLVMFN----TLTQNNKL--------------RRLFTISPFD-Q-------- AMR3p
244 DPIVIFNM-QNSVN------------------RAMYGFYPFRPE-------- AMR4p
313 EPVIIFNSYHRQIKDMSTED--DNNVHTKFEFYRSMFVGWLFQYQLGKLNTDG CaORF1
296 EPLIIFNSYHRKFVHYDDLEDSIMGQTVKFQ--RSMFMCWPWQYQMGKSNVEG CaORF2
314 EPIIVYNSHHRKITDAKSDNDG--ESNIHFKAYRSIFMAWLWQNQKGKNNVEE CaORF3
320 EPMVIYNSYHRQIANHTTTGKTDGSVELNFEFYRSMFVGWPFRYQLGKSNTDG CaORF4
286 EPLIVYNSFNRSPPNANYLEEI--KNLVKLDTYRSIFMAWIWRTQLGKSNVGS CaORF5
313 EPLIVFNMKGLKLIK------------------RVMYSYLPFSNTL------- CaORF6
318 EPLIAFNA-LST--EVDWK--------------RAMHIYRPLHDP-------- CaORF7
356 EPVLIYNSHHRKISEKHFDNDQ--EGKINFNNYRSLFIGWIWQTQLGKIHLEE CaORF8

图 5


260 ----------FKTVMYDIKDYEMPRYEKNWVPFFLKDNQEA--------- AMR2p
273 ----------DKVNLIDIEDKERENSEKNWSPFFLDDVEVSK--YST--- AMR1p
261 ----------YKTVMYRTNAFKMQTTEKNWVPFFLKDDQES--------- AMR3p
269 ----------NKQVLFSIKDEFPRKKEKNWTPFFVPGSPTT--------- AMR4p
364 IQDS--KFNNVTFNKVKEIRIEGKERTSIEKNWTPFIDPDERNQISYYGNHNL CaORF1
347 T--SNPEYDNKVYNRVIEIKVKLLADMKSQKNWTPFISEDSTNK--F------ CaORF2
365 I--ETGKMKNRVYVKSKEIIKPNNKREDKEKNWAPFINYQQRIQQGF------ CaORF3
373 FVDD--RFDNVKFTRVAEIKIHNQTRASIEKNWTPFVDPSERDP--------- CaORF4
337 SLPDLATTDDHKYVKVKELSLPNKKRPKTEKNWTPFVIYEDQKKQGY------ CaORF5
341 ----------KLIKKRREPFANIEKNWIPFKSVAQPSKT-------- CaORF6
346 ----------HRIIRLSIENYAPREKEKNWAPFIDGNN---------- CaORF7
407 -LPN-NEFKKNEYIKIKEFVKPNNNRGRTEKNWALFINYNQRLNQGF------ CaORF8

291 ----VHFVYSFNPIRVLKCSLDD----------GSCDIVFEIP-KVDSMSSE- AMR2p
308 --GYVHFVYSFNPLKVIKCSLDT----------GACRMIYESP-EEGRFGSE- AMR1p
292 ----VHFVYSFNPLRVLNCSLDN----------GACDVLFELP-HDFGMSSE- AMR3p
300 ----VNFVYDLQKLTILKCSIIT----------GICEKEFVSG-DDGQNHGIG AMR4p
415 GDNYVYIVYQWNHLKILKCELDNFIDSS----HSTCTMFFKDVETTQE---VG CaORF1
390 -DSYIYFVYRWANLDVLKCSLLGDVAGD-------CVFDYRLDETLVPQNKVC CaORF2
410 -DSHVYFMYQFQDLKILKCSLLDE--ED-------CVWEYQFNDK---NGA-G CaORF3
415 EDKSLYIVYQWDKLRILKCDISNLVTDDGFIHYSACR--FKQDTKHDEMEKVG CaORF4
384 -DSHLYFIYSFQDLSILKCSLWDA--GN-------CIWEYRMNNK---KTKIS CaORF5
370 -QTTIHFIYSMIPLEVLACDI-----DSGL-----C--DILQKPAKHDFNYVG CaORF6
374 ----LNFVYNE-PLRILRCNINN----------GDCQKVSGPDFNLKSHENAG CaORF7
452 -DSHVYFANQLKNLKILKCSILNDNDDD-------CEWEFQMLDY---EDA-G CaORF8

328 -LRCATPMINL----------PQAIPMAKD----KEIWVSFPRTRIANCG AMR2p
347 -LRCATPMVKL----------PVHLSLPKG----KEVWVAFPRTRLRDCG AMR1p
329 -LRCATPMINL----------PQAIPMADD----KEIWVSFPRTRISDCG AMR3p
338 IFRGGSNLVPF----------PTSFT---D----KDVWVGFPKTHMESCG AMR4p
461 PVRGGTELWPIK-IDNNNNNNNNLNEDDLSTKQEPQQQRQLWIGFLRAHVKDCG CaORF1
435 PLRGGTQLVNLRQVIPRSVYHRL----------LPSHREIFIGFARTHLDNCG CaORF2
449 RLRGGTELVNINQLLTTFDHPEIK----RVKDLMPQNREIWICVARAALEKCG CaORF3
466 PIRGGTKLIPTI-INN--------------------KQLWVGFLRTHIDKCG CaORF4
424 ELRGGTELMNVNQLLDKYNFAGLE----TVKDQF-KGKEVWISFARAALSKCG CaORF5
410 GLRGGTQLVSLP-------LNETIPSEIRAKLPIPKNRQVYIGWARTHLNNCG CaORF6
412 KLRGGTNLVEI----------PSQ-SLPKHL-RSRKYWFGIARSHITDCG CaORF7
493 VLHGGTELININQLLHQYDYPELN----SIKDLIPNGREYWVGFARASLKNCG CaORF8

363 CSRTTYREMLMLFVR--EG--SNFF--VELLSTSLDFGIEVLEYS-G------ AMR2p
382 CSRTTYRPVLTLFVK--EG--NKFY--TELISSSIDFHIDVLSYD-A------ AMR1p
364 CSETMYRPMLMLFVR--EG--TNFF--AELLSSSIDFGIEVIEYT-G------ AMR3p
371 CSSHIYRPYLMVLVR--KG--DFYY--KAFVSTPLDFGIDVRSWESA------ AMR4p
513 CGGSMYRPNFLILEKLN-SK-----FKLTYLSGSINFNVSVYGWAN------- CaORF1
478 CGKVMYRPNLVILVKDAADKT---YYKISHISSSLSFDVPIIGWNV------- CaORF2
498 CGDKMYRPNIVILIKDGDDQ-----YRLSHVSPFVGLGILILPW--------- CaORF3
497 CGKAMDRPNMGVLQRTDMGT-----FQVAYLSSYISFNIAVPGWKT------- CaORF4
472 CGSKMYRPNFTVLVKQGG-R-----FQLSFVSSYMDFGVPILPW--------- CaORF5
456 CGDSMYRPNFITLVEDYDDVTDKYYYKIGDISGYFDFAAKIEPWSKQVLDEEG CaORF6
450 CVGELYRPHLILISRNKKS--DQYE--LNYVSDLIDFNVNPEPWTPG------ CaORF7
542 CGSRMYRPNLIVLMKDGKN------YKFAYVSSFVGLGIEILPW--------- CaORF8

图 5

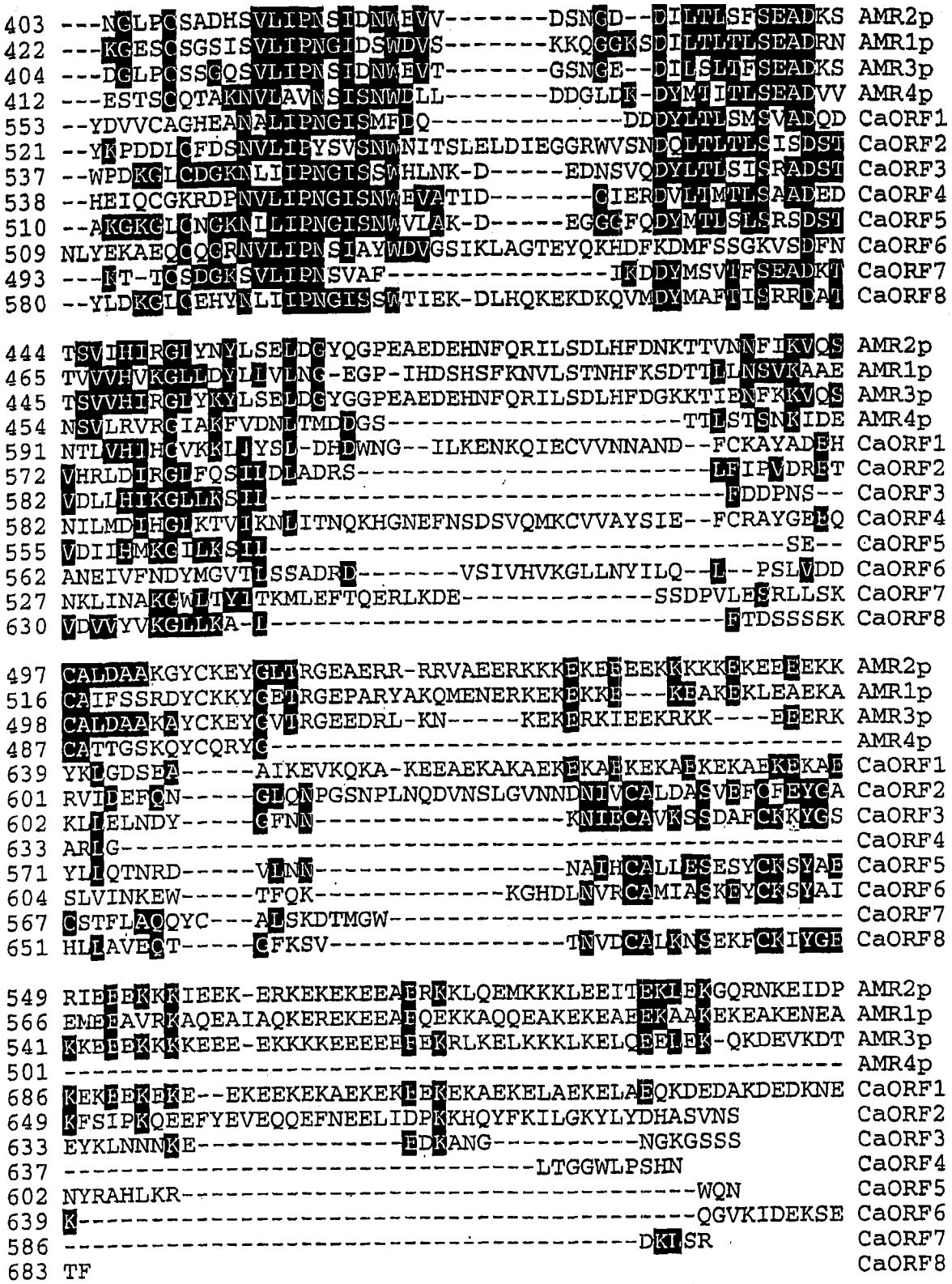

403 ---NGLPCSADHSVLIPNSIDNWEVV-------DSNGD--DILTLSFSEADKS AMR2p
422 ---KGESCSGSISVLIPNGIDSWDVS-------KKQGGKSDILTLTLSEADRN AMR1p
404 ---DGLPCSSGQSVLIPNSIDNWEVT-------GSNGE--DILSLTFSEADKS AMR3p
412 ---ESTSCQTAKNVLAVNSISNWDLL-------DDGLDK-DYMTITLSEADVV AMR4p
553 --YDVVCAGHEANALIPNGISMFDQ----------DDDYLTLSMSVADQD CaORF1
521 --YKPDDLCFDSNVLIPYSVSNWNITSLELDIEGGRWVSNDQLTLTLSISDST CaORF2
537 --WPDKGLCDGKNLIIPNGISSWHLNK-D-----EDNSVQDYLTLSISRADST CaORF3
538 --HEIQCGKRDPNVLIPNGISNWEVATID-------CIERDVLTMTLSAADED CaORF4
510 --AKGKGLCNGKNLLIPNGISNWVLAK-D-----EGGGFQDYMTLSLSRSDST CaORF5
509 NLYEKAEQCQCRNVLIPNSIAYWDVGSIKLAGTEYQKHDFKDMFSSGKVSDFN CaORF6
493 ---FT-TCSDGKSVLIPNSVAF----------IKDDYMSVTFSEADKT CaORF7
580 --YLDKGLCEHYNLIIPNGISSWTIEK-DLHQKEKDKQVMDYMAFTISRRDAT CaORF8
444 TSVIHIRGIYNYLSELDGYQGPEAEDEHNFQRILSDLHFDNKTTVNNFIKVQS AMR2p
465 TVVVHVKGLLDYLIVLNG-EGP-IHDSHSFKNVLSTNHFKSDTILLNSVKAAE AMR1p
445 TSVVHIRGLYKYLSELDGYGGPEAEDEHNFQRILSDLHFDGKKTIENFKKVQS AMR3p
454 NSVLRVRGIAKFVDNLTMDDGS----------TTLSTSNKIDE AMR4p
591 NTLVHIHGVKKLIYSL-DHDWNG--ILKENKQIECVVNNAND--FCKAYADEH CaORF1
572 VHRLDIRGLFQSIIDLADRS----------LFIPVDRET CaORF2
582 VDLLHIKGLLKSII----------FDDPNS-- CaORF3
582 NILMDIHGIKTVIKNLITNQKHGNEFNSDSVQMKCVVAYSIE--FCRAYGEEQ CaORF4
555 VDIIHMKGILKSII----------SE-- CaORF5
562 ANEIVFNDYMGVTLSSADRD-------VSIVHVKGLLNYILQ--L--PSLVDD CaORF6
527 NKLINAKGWLTYITKMLEFTQERLKDE----------SSDPVLESRLLSK CaORF7
630 VDVVYVKGLLKA-I----------FTDSSSSK CaORF8
497 CALDAAKGYCKEYGLTRGEAERR-RRVAEERKKKEKEEEKKKKKEKEEEEKK AMR2p
516 CAIFSSRDYCKKYGETRGEPARYAKQMENERKEKEKKE---KEAKEKLEAEKA AMR1p
498 CALDAAKAYCKEYGVTRGEEDRL-KN-----KEKERKIEEKRKK----EEERK AMR3p
487 CATTGSKQYCQRYG---------- AMR4p
639 YKLGDSEA-----AIKEVKQKA-KEEAEKAKAEKEKAEKEKAEKEKAEKEKAE CaORF1
601 RVIEEFQN-----GLQNPGSNPLNQDVNSLGVNNDNIVCALDASVEFCFEYGA CaORF2
602 KLIELNDY-----GFNN----------KNIECAVKSSDAFCKKYGS CaORF3
633 ARLG---------- CaORF4
571 YLLQTNRD-----VLNN----------NAIHCALLESESYCKSYAE CaORF5
604 SLVINKEW-----TFQK----------KGHDLNVRCAMIASKEYCKSYAI CaORF6
567 CSTFLAQQYC---ALSKDTMGW---------- CaORF7
651 HLLAVECT-----GFKSV----------TNVDCALKNSEKFCKIYGE CaORF8
549 RIEEEKKKIEEK-ERKEKEKEEAERKKLQEMKKKLEEITEKLEKGQRNKEIDP AMR2p
566 EMEEAVRKAQEAIAQKEREKEEAEQEKKAQQEAKEKEAEEKAAKEKEAKENEA AMR1p
541 KKEEEKKKKEEE-EKKKKEEEEEEKRLKELKKKLKELQEELEK-QKDEVKDT AMR3p
501 ---------- AMR4p
686 KEKEEKEKE--EKEEKEKAEKEKLEKEKAEKELAEKELAEQKDEDAKDEDKNE CaORF1
649 KFSIPKQEEFYEVEQQEFNEELIDPKKHQYFKILGKYLYDHASVNS CaORF2
633 EYKLNNNKE----------EDKANG----------NGKGSSS CaORF3
637 ----------LTGGWLPSHN CaORF4
602 NYRAHLKR----------WQN CaORF5
639 K----------QGVKIDEKSE CaORF6
586 ----------DKISR CaORF7
683 TF CaORF8

图 5

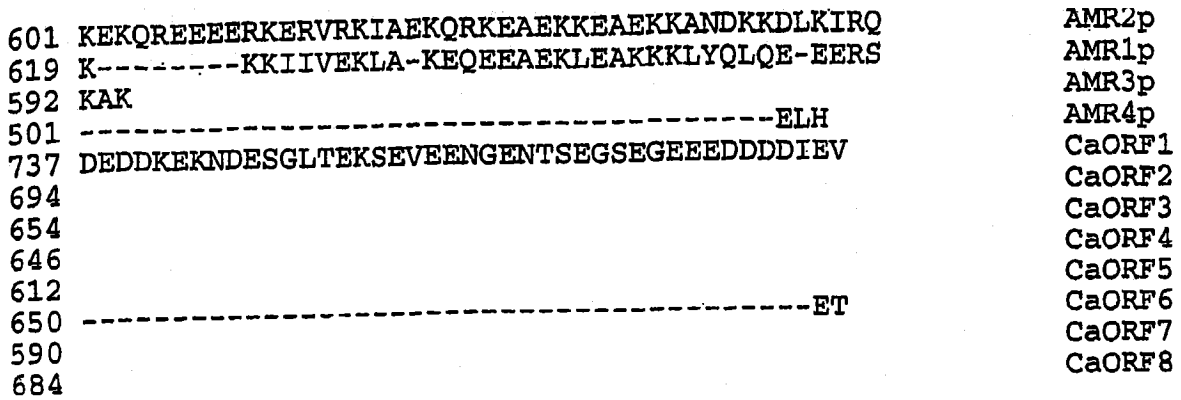

601 KEKQREEEERKERVRKIAEKQRKEAEKKEAEKKANDKKDLKIRQ AMR2p
619 K--------KKIIVEKLA-KEQEEAEKLEAKKKLYQLQE-EERS AMR1p
592 KAK AMR3p
501 ----------ELH AMR4p
737 DEDDKEKNDESGLTEKSEVEENGENTSEGSEGEEEDDDDIEV CaORF1
694 CaORF2
654 CaORF3
646 CaORF4
612 CaORF5
650 ----------ET CaORF6
590 CaORF7
684 CaORF8

图 5