



(12) 发明专利

(10) 授权公告号 CN 101855354 B

(45) 授权公告日 2016. 04. 06

(21) 申请号 200880115486. 6

代理人 黄革生 凌立

(22) 申请日 2008. 09. 11

(51) Int. Cl.

(30) 优先权数据

C12N 15/82(2006. 01)

07116174. 9 2007. 09. 12 EP

60/993, 575 2007. 09. 13 US

(56) 对比文件

CN 85100964 A, 1988. 02. 17,

WO 2007/039316 A1, 2007. 04. 12,

CN 85103471 A, 1986. 10. 29,

(85) PCT国际申请进入国家阶段日

2010. 05. 10

审查员 周茂新

(86) PCT国际申请的申请数据

PCT/EP2008/007837 2008. 09. 11

(87) PCT国际申请的公布数据

W02009/033752 EN 2009. 03. 19

(83) 生物保藏信息

DSM 16664 2004. 08. 25

(73) 专利权人 拜尔知识产权有限公司

地址 德国蒙海姆

(72) 发明人 C·弗罗贝格 B·埃西格曼

(74) 专利代理机构 北京市中咨律师事务所

11247

权利要求书2页 说明书47页

序列表43页

(54) 发明名称

合成增加量的葡糖胺聚糖的植物

(57) 摘要

本发明涉及合成增加量的葡糖胺聚糖的植物细胞和植物, 和用于制备此类植物的方法, 以及在这些植物细胞或植物的帮助下制备葡糖胺聚糖的方法。此处, 相比较于野生型植物细胞或野生型植物, 根据本发明的植物细胞或经遗传修饰的植物具有葡糖胺聚糖合酶活性以及额外增加的葡糖胺 6- 磷酸乙酰转移酶活性和增加的 UDP-N- 乙酰基- 葡糖胺焦磷酸化酶活性。本发明还涉及包含具有增加的葡糖胺聚糖合成的植物细胞的组合物。

1. 包含编码葡糖胺聚糖合酶的外来核酸分子的经遗传修饰的植物细胞, 其中所述植物细胞是非繁殖材料, 其还包含编码具有葡糖胺 6- 磷酸乙酰转移酶活性的蛋白质的外来核酸分子以及编码具有单功能 UDP-N- 乙酰基 - 葡糖胺焦磷酸化酶活性的蛋白质的外来核酸分子。

2. 产生植物的方法, 其包括:

a) 遗传修饰植物细胞, 其中所述遗传修饰包括下文的步骤 i 到 iii

i) 向植物细胞中引入编码具有葡糖胺聚糖合酶活性的蛋白质的核酸分子

ii) 向植物细胞中引入编码具有葡糖胺 6- 磷酸乙酰转移酶活性的蛋白质的核酸分子

iii) 向植物细胞中引入编码具有 UDP-N- 乙酰基 - 葡糖胺焦磷酸化酶单功能活性的蛋白质的核酸分子

b) 从来自步骤 a) 的植物细胞再生植物;

c) 适当时, 使用步骤 b) 的植物产生进一步的植物,

其中可以任何顺序进行步骤 a) i 到 a) iii, 并且适当时, 可向根据步骤 b) 或 c) 获得的植物的植物细胞中引入步骤 a) i 到 a) iii 的缺少的外来核酸分子。

3. 权利要求 2 的方法, 其中所述编码具有葡糖胺聚糖合酶活性的蛋白质的核酸分子是 SEQ ID NO :1、SEQ ID NO :3, 或由于遗传密码的简并性与 SEQ ID NO 1 或 SEQ ID NO 3 中显示的核酸序列有所不同的核酸分子;

所述编码具有葡糖胺 6- 磷酸乙酰转移酶活性的蛋白质的核酸分子是 SEQ ID NO :8, 或由于遗传密码的简并性与 SEQ ID NO :8 中显示的核酸序列有所不同的核酸分子, 或 SEQ ID NO :9 所示的氨基酸序列的编码核酸分子;

所述编码具有 UDP-N- 乙酰基 - 葡糖胺焦磷酸化酶单功能活性的蛋白质的核酸分子是 SEQ ID NO :10, 或由于遗传密码的简并性与 SEQ ID NO :10 中显示的核酸序列有所不同的核酸分子, 或 SEQ ID NO :11 所示的氨基酸序列的编码核酸分子。

4. 用于产生葡糖胺聚糖的方法, 其包括从权利要求 1 的经遗传修饰的植物细胞, 或通过权利要求 2 或 3 的方法获得的植物中提取葡糖胺聚糖的步骤。

5. 权利要求 1 的经遗传修饰的植物细胞, 或通过权利要求 2 或 3 的方法获得的植物用于产生葡糖胺聚糖的用途。

6. 包含不再完整的权利要求 1 的经遗传修饰的植物细胞的组合物。

7. 权利要求 6 的组合物, 其中所述编码具有葡糖胺聚糖合酶活性的蛋白质的核酸分子是 SEQ ID NO :1、SEQ ID NO :3, 或由于遗传密码的简并性与 SEQ ID NO 1 或 SEQ ID NO 3 中显示的核酸序列有所不同的核酸分子;

所述编码具有葡糖胺 6- 磷酸乙酰转移酶活性的蛋白质的核酸分子是 SEQ ID NO :8, 或由于遗传密码的简并性与 SEQ ID NO :8 中显示的核酸序列有所不同的核酸分子, 或 SEQ ID NO :9 所示的氨基酸序列的编码核酸分子;

所述编码具有 UDP-N- 乙酰基 - 葡糖胺焦磷酸化酶单功能活性的蛋白质的核酸分子是 SEQ ID NO :10, 或由于遗传密码的简并性与 SEQ ID NO :10 中显示的核酸序列有所不同的核酸分子, 或 SEQ ID NO :11 所示的氨基酸序列的编码核酸分子。

8. 用于制备包含葡糖胺聚糖的组合物方法, 其包括使用权利要求 1 的经遗传修饰的植物细胞, 或通过权利要求 2 或 3 的方法获得的植物。

9. 权利要求 1 的经遗传修饰的植物细胞,或通过权利要求 2 或 3 的方法获得的植物用于制备权利要求 6 或 7 的组合物的用途。

合成增加量的葡糖胺聚糖的植物

[0001] 本发明涉及合成增加量的葡糖胺聚糖的植物细胞和植物,和制备此类植物的方法,还有在这些植物细胞或植物的帮助下制备葡糖胺聚糖的方法。此处,相比较于野生型植物细胞或野生型植物,根据本发明的植物细胞或经遗传修饰的植物具有葡糖胺聚糖合酶活性以及额外增加的葡糖胺 6-磷酸乙酰转移酶活性和增加的 UDP-N-乙酰基-葡糖胺焦磷酸化酶活性。本发明还涉及包含具有增加的葡糖胺聚糖合成的植物细胞的组合物。

[0002] 蛋白聚糖是一类糖蛋白,尤其是软骨的主要成份并且具有由重复的二糖单元组成的附着于核心蛋白质上的葡糖胺聚糖。重复的二糖单元其功能在于通过特征性的碳水化合物结合序列共价附着到核心蛋白质上。根据二糖单元的组成,尤其是在葡糖胺聚糖乙酰肝素/硫酸肝素、硫酸角质素硫酸和软骨素/皮肤素之间进行区分,其二糖单元各包含为葡糖胺或葡糖胺衍生物的分子。

[0003] 乙酰透明质酸是另一种葡糖胺聚糖,也具有作为其二糖单元组分之一一的葡糖胺衍生物(N-乙酰基-葡糖胺),但其在自然界中不附着到蛋白质上。除了乙酰透明质酸,所提及的葡糖胺聚糖是天然硫酸化聚合物。在这些物质中,硫酸根在二糖单元的多个原子或取代基处引入,使得所提及的物质不是单一的聚合物,而是总结于各总称下的聚合物集合。正在讨论的聚合物集合的每一分子可能在硫酸化程度和具有硫酸根的单体的位置上有所不同。

[0004] 乙酰透明质酸是天然产生的、无分支的、线性粘多糖(葡糖胺聚糖),其由交替的葡糖醛酸(GlcA)和N-乙酰基-葡糖胺(GlcNAc)分子组成。乙酰透明质酸的基本结构单元由二糖葡糖醛酸- β -1,3-N-乙酰基-葡糖胺组成。在乙酰透明质酸中,这些重复单元经 β -1,4键相互附着。在药剂学中,经常使用术语透明质酸。由于在大部分情况中乙酰透明质酸作为多聚阴离子而不是游离酸存在,所以在下文中,优选使用术语乙酰透明质酸,但各术语应理解为均包含两种分子形式。

[0005] 乙酰透明质酸具有不一般的物理化学特性,例如聚合电解质的特性、粘弹性特性、结合水的高能力、凝胶形成的特性,除此之外乙酰透明质酸的其他特性都描述于Lapcik等的综述文章中(1998, Chemical Reviews 98(8), 2663-2684)。

[0006] 乙酰透明质酸是脊椎动物胞外结缔组织和体液的组分。在人中,透明质酸由所有体细胞尤其是间充质细胞的细胞膜合成,并广泛存在于身体当中,尤其在结缔组织、细胞外基质、脐带、关节液、软骨组织、皮肤和眼的玻璃体中具有高浓度(Bernhard Gebauer, 1998, Inaugural-Dissertation, Virchow-Klinikum Medizinische **Fakultät** Charité der Humboldt **Universität** zu Berlin; Fraser等, 1997, Journal of Internal Medicine 242, 27-33)。

[0007] 近来,还在动物非脊椎动物有机体(软体动物)中发现了乙酰透明质酸(Volpi和Maccari, 2003, Biochimie 85, 619-625)。

[0008] 此外,由于乙酰透明质酸是非免疫原性物质,一些人类致病性革兰氏阳性细菌(链球菌属(*Streptococcus*) A和C)和革兰氏阴性细菌(巴斯德菌属(*Pasteurella*))合成乙酰透明质酸作为菌表多糖,其保护这些细菌免受宿主免疫系统的攻击。

[0009] 感染小球藻属 (*Chlorella*) 的单细胞绿藻的病毒赋予单细胞绿藻在被所述病毒感染后合成乙酰透明质酸的能力,一些这类病毒作为内共生体存在于草履虫 (*Paramecium*) 种中 (Graves 等,1999,*Virology* 257,15-23)。然而,合成乙酰透明质酸的能力不是用于表征正在讨论的藻类的特征。藻类合成乙酰透明质酸的能力由基因组中具有编码乙酰透明质酸合酶的序列的病毒感染所介导 (DeAngelis,1997,*Science* 278,1800-1803)。

[0010] 乙酰透明质酸合成的催化受单个膜整合或膜结合的酶乙酰透明质酸合酶的影响。

[0011] 将合成的乙酰透明质酸分子穿过细胞质膜转移到细胞周围培养基中的机制仍未完全阐明。初期的假说认为穿过细胞膜的转运受乙酰透明质酸合酶本身的影响。然而,更近期的结果显示乙酰透明质酸分子穿过细胞质膜的转运经负责该运动的转运蛋白质通过能量依赖运输发生。因此,通过突变来产生链球菌属菌株,在所述突变中活性转运蛋白的合成被抑制。这些菌株比相应的野生型细菌菌株合成更少的乙酰透明质酸 (Ouskova 等,2004,*Glycobiology* 14(10),931-938)。在人成纤维细胞中,可证明在特异性抑制已知转运蛋白的试剂的帮助下可以将所产生的乙酰透明质酸的量以及乙酰透明质酸合酶的活性降低 (Prehm 和 Schumacher,2004,*Biochemical Pharmacology* 68,1401-1410)。如果在植物中确实存在转运乙酰透明质酸的转运蛋白质,那么其中量是未知的。

[0012] 乙酰透明质酸的不一般特性提供了在多个领域,例如药剂学、化妆品工业、食品和饲料生产、在技术应用(例如润滑剂)等中应用的广泛可能性。目前使用乙酰透明质酸的最重要的应用就是在医学和化妆品领域(例如见 Lapcik 等,1998,*Chemical Reviews* 98(8),2663-2684,Goa 和 Benfield,1994,*Drugs* 47(3),536-566)。

[0013] 在医学领域,含乙酰透明质酸的产品目前用于关节病的关节内治疗和用于眼的眼外科中。乙酰透明质酸还用于治疗赛马中的关节疾病。此外,乙酰透明质酸是一些眼鼻用药 (rhinologics) 的组分,其例如以滴眼剂和 nasalia 的形式用于润湿干燥粘膜。注射用的含乙酰透明质酸的溶液用作镇痛药和抗风湿剂。在创伤愈合中使用包含乙酰透明质酸或衍生化的乙酰透明质酸的药膏。对于皮肤学,含乙酰透明质酸的凝胶埋植剂用于在整形外科中矫正皮肤畸形。

[0014] 对于药理学应用,优选使用具有高分子量的乙酰透明质酸。

[0015] 在化妆品药物中,乙酰透明质酸制剂是最适合的皮肤填充材料。通过注射乙酰透明质酸有限时间,可以使皱纹平滑或增加嘴唇的体积。

[0016] 在化妆品产品中,尤其是在皮肤乳液和洗液中,由于其高结合水的能力,常使用乙酰透明质酸作为增湿剂。

[0017] 此外,含乙酰透明质酸的制剂也作为所谓的营养制品(食品添加剂)来出售,其也可用于动物中(例如狗、马)用以预防和减轻关节病。

[0018] 目前从动物组织(鸡冠)中分离或使用细菌培养物发酵来制备商业化目的的乙酰透明质酸。

[0019] US 4,141,973 描述了从鸡冠或脐带中分离乙酰透明质酸的方法。除了乙酰透明质酸,动物组织(例如鸡冠、脐带)也包含其他与乙酰透明质酸相关的粘多糖,例如硫酸软骨素 (chondroitin sulfate)、硫酸皮肤素、硫酸角质素、硫酸乙酰肝素和肝素。此外,动物有机体包含蛋白质 (hyaladherins),所述蛋白质与乙酰透明质酸特异性结合并且对于有机体中大多数不同功能,例如肝中乙酰透明质酸的降解、乙酰透明质酸作为细胞迁移的前导

结构的功能、内吞作用的调节、细胞表面上乙酰透明质酸的锚定或乙酰透明质酸网络的形成是必需的 (Turley, 1991, Adv Drug Delivery Rev 7, 257ff. ; Laurent 和 Fraser, 1992, FASEB J. 6, 183ff. ; Stamenkovic 和 Aruffo, 1993, Methods Enzymol. 245, 195ff ; Knudson 和 Knudson, 1993, FASEB 7, 1233ff.)。

[0020] 用于细菌产生乙酰透明质酸的链球菌属菌株全部是对人致病的细菌。在培养过程中, 这些细菌还产生释放到培养基中的 (致热性的) 外毒素和溶血素 (链球菌溶血素, 尤其是 α 溶血素和 β 溶血素) Kilian, M. : Streptococcus and Enterococcus. 出自 Medical Microbiology. Greenwood, D. ; Slack, RCA ; Peutherer, J. F. (编辑). 第 16 章. Churchill Livingstone, Edinburgh, UK : 第 174-188 页, 2002, ISBN 0443070776)。这使纯化和分离在链球菌菌株帮助下制备的乙酰透明质酸变得更加困难。尤其是对于药物应用, 制剂中外毒素和溶血素的存在是个问题。

[0021] US 4, 801, 539 描述了通过诱变处理后的细菌菌株 (Streptococcus zooedemicus) 的发酵来制备乙酰透明质酸。所用的诱变处理细菌菌株不再合成 β 溶血素。所获的产率为每升培养基 3.6g 乙酰透明质酸。

[0022] EP 0694616 描述了培养 Streptococcus zooedemicus 或马链球菌 (Streptococcus equi) 的方法, 其中在所用的培养条件下, 合成了增量的乙酰透明质酸而不产生链球菌溶血素。所获的产率为每升培养基 3.5g 乙酰透明质酸。

[0023] 在培养过程中, 链球菌属菌株向培养基中释放酶透明质酸酶, 其结果是在该生产体系中, 在纯化过程中也降低了分子量。透明质酸酶阴性的链球菌属菌株的用途或制备乙酰透明质酸的方法 (其中在培养过程中抑制了透明质酸酶的产生) 描述于 US 4, 782, 046 中。所获产率至多每升培养基 2.5g 乙酰透明质酸, 并且所获的最高平均分子量为 3.8×10^6 Da, 介于 2.4×10^6 至 4.0×10^6 的分子量分布中。

[0024] US 20030175902 和 WO 03054163 描述了在枯草芽孢杆菌 (Bacillus subtilis) 中在来自类马链球菌 (Streptococcus equisimilis) 的乙酰透明质酸合酶的异源表达的帮助下制备乙酰透明质酸。为实现足量乙酰透明质酸的产生, 除了异源表达乙酰透明质酸合酶之外, 还需要在芽孢杆菌 (Bacillus) 细胞中同时表达 UDP 葡萄糖脱氢酶。US 20030175902 和 WO 03 054163 没有描述在枯草杆菌帮助下的生产中所获的乙酰透明质酸的绝对量。所获的最高平均分子量约为 4.2×10^6 。然而, 该平均分子量仅获自重组芽孢杆菌菌株中, 其中将编码来自类马链球菌的乙酰透明质酸合酶基因的基因和编码来自枯草芽孢杆菌的 UDP 葡萄糖脱氢酶的基因整合入在 amyQ 启动子调控下的枯草芽孢杆菌基因组中, 其中同时失活枯草芽孢杆菌内源 xcpY 基因 (其编码细胞色素 P450 氧化酶)。Chien 和 Lee (2007, Biotechnol. Prog. Online publication, ASAP Article 10.1021/bp070036w, S8756-7938(07)00036-7) 描述了多种的重组枯草芽孢杆菌菌株。已转化有编码乙酰透明质酸合酶的核酸序列和编码 UDP 葡萄糖脱氢酶的核酸序列的菌株合成最高 1.14g/l 的乙酰透明质酸。除转化有刚刚提及的核酸序列外还转化了编码透明颤菌属 (Vitreoscilla) 血红蛋白的核酸序列的菌株合成 1.8g/l 的乙酰透明质酸。

[0025] WO 06 032538 描述了转化有编码乙酰透明质酸合酶的核酸分子的转基因植物。清楚地表明了正在讨论的植物中乙酰透明质酸的合成。

[0026] WO 05 012529 描述了使用编码来自小球藻感染病毒的乙酰透明质酸合酶的核酸

分子转化的转基因烟草植株的产生。在 WO 05 012529 中,一方面使用编码小球藻病毒菌株 CVH1 的乙酰透明质酸合酶的核酸序列,另一方面使用编码小球藻病毒菌株 CVKA1 的乙酰透明质酸合酶的核酸序列来转化烟草植株。仅用编码分离自小球藻病毒菌株 CVKA1 的乙酰透明质酸合酶的核酸转化的植物能证明乙酰透明质酸的合成。对于转化有编码分离自小球藻病毒菌株 CVH1 的乙酰透明质酸合酶的核酸转化的烟草植株,不可能在相应的转基因植物中检测到乙酰透明质酸的合成。WO 05 012529 中唯一产乙酰透明质酸的转基因烟草植株所合成的乙酰透明质酸的量表述为每 ml 测定体积中约 4.2 μ g 的乙酰透明质酸,考虑到对于进行所述实验的描述,其对应于约每克植物材料鲜重最高 12 μ g 乙酰透明质酸的量。

[0027] WO 2007 039314 描述了具有乙酰透明质酸合酶活性和额外增加的谷氨酰胺:果糖 6-磷酸酰胺转移酶 (GFAT) 活性的转基因植物。这些植物相比于仅具有乙酰透明质酸合酶活性的植物合成了增加量的乙酰透明质酸。由这些烟草植株合成的乙酰透明质酸的最大量为用于测定的每克植物材料鲜重的约 0.03% (见 WO 2007 039316 中的图 5)。

[0028] WO 2007 039316 描述了与野生型植物相比具有乙酰透明质酸合酶活性以及额外增加的谷氨酰胺:果糖 6-磷酸酰胺转移酶 (GFAT) 活性和增加的 UDP 葡萄糖脱氢酶 (UDP-Glc-DH) 活性的转基因植物。这些植物相比于具有乙酰透明质酸合酶活性以及同时具有蛋白质 (具有 GFAT 活性) 的活性的植物合成了增加量的乙酰透明质酸。由这些烟草植株合成的乙酰透明质酸的最大量为用于测定的每克植物材料鲜重的 0.2% (见 WO 2007039316 中的图 6)。

[0029] 此外,WO 2007 039316 包含了蛋白质列表,所述蛋白质能在植物细胞中表达以在植物细胞中进一步增加合成的乙酰透明质酸的量。WO 2007039316 中提出的蛋白质为酶的随机列表,所述酶在多种有机体中参与了 UDP-GlcNAc 的合成。列于 WO 2007 039316 中的蛋白质具有多种酶功能。WO 2007 039316 未给予说明列出的酶当在转基因植物中表达时是否可能确实增加乙酰透明质酸的含量或哪一种酶能确实增加乙酰透明质酸的含量。

[0030] 由软骨素合酶从 UDP-葡萄糖醛酸 (UDP-GlcA) 和 UDP-N-乙酰基-半乳糖胺 (UDP-GalNAc) (UDP-N-乙酰基-葡萄糖胺 (UDP-GlcNAc) 的表异构物) 开始来催化软骨素/皮肤素 ($[\beta-1,4]-[\text{葡萄糖醛酸}-\beta-1,4-\text{N-乙酰半乳糖胺}]_n$) 的二糖链的合成 (Kitagawa 等,2001, J Biol Chem 276(42),38721-38726)。通过表异构酶,软骨素的葡萄糖醛酸分子可转化为艾杜糖醛酸。如果超过 10% 的葡萄糖醛酸分子以艾杜糖醛酸存在,则聚合物称为皮肤素。随后经其他酶催化在二糖链的软骨素或皮肤素的多个位置上引入硫酸根,导致硫酸软骨素/皮肤素的形成。此处,硫酸化程度因分子不同而不同。

[0031] 有时候,已将硫酸软骨素作为治疗骨关节炎的潜在有效化合物来讨论 (Clegg 等,2006, The New England Journal of Medicine 354(8),795-808)。

[0032] 肝素/乙酰肝素合酶从 UDP-GlcA 和 UDP-GlcNAc 开始催化肝素/乙酰肝素 (heparosans) 二糖链 ($[\alpha-1,4]-[\text{葡萄糖醛酸}-\beta-1,4-\text{葡萄糖胺}]_n$ 或 $[\alpha-1,4]-[\text{艾杜糖醛酸}-\alpha-1,4-\text{葡萄糖胺}]_n$) 的合成 (DeAngelis 和 White,2004, J. Bacteriology 186(24),8529-8532)。heparosans 的葡萄糖醛酸分子可通过表异构酶转化为艾杜糖醛酸。在乙酰肝素二糖链多个位置上引入硫酸根随后由其他酶催化,导致肝素或硫酸乙酰肝素的形成。硫酸肝素比硫酸乙酰肝素具有显著更多的硫酸根替代。硫酸肝素具有约 90% 的艾杜糖醛酸分子,而在硫酸乙酰肝素中葡萄糖醛酸分子占优势 (Gallagher 等,1992, Int. J. Biochem 24,

553-560)。如在硫酸软骨素 / 硫酸皮肤素的情况中以及硫酸肝素 / 硫酸乙酰肝素的情况下, 硫酸化程度因分子不同而不同。

[0033] 尤其使用硫酸肝素作为抗凝血剂, 例如用于预防和治疗血栓形成。

[0034] 目前, 通过从动物组织中分离来制备硫酸软骨素 / 硫酸皮肤素和硫酸肝素 / 硫酸乙酰肝素。硫酸软骨素主要从牛软骨 (cartilage) 或鲨鱼软骨 (cartilage) 中分离, 而硫酸肝素 / 硫酸乙酰肝素从猪肠或牛肺中分离。由于硫酸软骨素 / 硫酸皮肤素和硫酸肝素 / 硫酸乙酰肝素的二糖链不具有统一的硫酸化结构, 因此难以获得相同的特定产物。因此, 产物通常为具有不同程度硫酸化的分子的混合物。

[0035] 如早已描述, 葡糖胺聚糖, 例如硫酸软骨素或硫酸肝素 / 硫酸乙酰肝素目前从动物组织中分离。除了所需物质外, 这些组织还包含其他葡糖胺聚糖。如完全可能的话, 分离各葡糖胺聚糖是困难且昂贵的。此外, 当使用从动物组织中分离的葡糖胺聚糖时, 可在人中导致疾病的病原微生物和 / 或其他物质例如 BSE 或禽流感病原体对动物组织的潜在污染形成了问题。在患者中, 使用被动物蛋白质污染的医学制剂可导致人体不期望的免疫反应 (对于乙酰透明质酸制剂见例如 US 4, 141, 973), 尤其是如果患者对动物蛋白质过敏。

[0036] 此外, 从动物粗原料中制备的物质对于一些生活方式是无法接受的, 例如对于素食者或 kosher 食物制品。

[0037] 从动物组织中分离葡糖胺聚糖的另一个问题实际在于, 由于动物组织还包含葡糖胺聚糖降解酶, 因此葡糖胺聚糖的分子量常常在纯化过程中下降。

[0038] 可从动物组织中以令人满意的质量和纯度获得的葡糖胺聚糖的量 (产率) 是很低的 (例如来自鸡冠的乙酰透明质酸 :0.079% w/w, EP 0144019, US4, 782, 046), 这意味着需要加工大量的动物组织。

[0039] 由于必须在密闭无菌的容器内并在昂贵的可控培养条件下发酵细菌, 因此通过细菌菌株发酵产生葡糖胺聚糖成本很高 (乙酰透明质酸见例如 US4, 897, 349)。此外, 可通过细菌菌株发酵产生的葡糖胺聚糖的量受每种情况中存在的生产装置的限制。此处, 还必须考虑到作为物理定则的结果, 不能将发酵罐建成具有非常大的培养体积。此处可特别要提及的是, 如果可确保, 仅用高额的技术消费才可确保在大发酵罐中有效生产所需的培养基与从外部填料的物质 (例如细菌必需的营养源、调节 pH 值的试剂、氧气) 的均质混合。

[0040] 植物在天然状态下不产生葡糖胺聚糖, 例如乙酰透明质酸、硫酸肝素 / 硫酸乙酰肝素、硫酸角质素或硫酸软骨素 / 硫酸皮肤素。天然出现的植物自身在其基因组中不具有编码催化葡糖胺聚糖合成的蛋白质的任何核酸, 并且, 尽管已描述并表征了大量植物碳水化合物, 迄今为止还不可能在未感染的天然植物细胞中检测到任何提及的葡糖胺聚糖 (Graves 等, 1999, Virology 257, 15-23)。

[0041] WO 98 35047 (US 6, 444, 878) 描述了在植物细胞中合成 GlcNAc 的代谢途径, 其中葡糖胺经多个连续酶催化反应步骤形成代谢物 GlcNAc、N-乙酰基-葡糖胺 6-磷酸和 N-乙酰基-葡糖胺 1-磷酸并转化为 UDP-GlcNAc。在更高浓度下, 葡糖胺 6-磷酸对植物细胞是有毒性的 (WO 98 35047)。

[0042] 描述的植物备选代谢途径包括果糖 6-磷酸和葡糖胺反应产生葡糖胺 6-磷酸, 其随后经多步连续酶催化反应步骤形成代谢物葡糖胺 1-磷酸和 N-乙酰基-葡糖胺 1-磷酸并转化为 UDP-GlcNAc (Mayer 等, 1968, PlantPhysiol. 43, 1097-1107)。

[0043] 迄今为止,仍不清楚植物中合成 UDP-GlcNAc 的代谢途径中应修饰哪一种蛋白质活性以使植物合成增加量的葡糖胺聚糖。

[0044] 因此,本发明的目标是提供用于在植物中制备有效量的葡糖胺聚糖的备选方法和过程。

[0045] 通过权利要求中涉及的实施方案实现该目标。

[0046] 令人惊奇的是,已发现包含编码葡糖胺聚糖合酶的核酸分子和额外包含编码具有葡糖胺 6-磷酸乙酰转移酶活性的蛋白质的外来核酸分子和编码具有单功能 UDP-N-乙酰基-葡糖胺焦磷酸化酶活性的蛋白质的外来核酸分子的经遗传修饰的植物细胞或经遗传修饰的植物较(仅)具有葡糖胺聚糖合酶活性的植物细胞或植物产生显著更高量的葡糖胺聚糖。

[0047] 因此,本发明涉及包含编码葡糖胺聚糖合酶的外来核酸分子的经遗传修饰的植物细胞或经遗传修饰的植物,其特征在于所述经遗传修饰的植物细胞或所述经遗传修饰的植物还包含编码具有葡糖胺 6-磷酸乙酰转移酶活性的蛋白质的外来核酸分子和编码具有单功能 UDP-N-乙酰基-葡糖胺焦磷酸化酶活性的蛋白质的外来核酸分子。

[0048] 由于必须在密闭无菌的容器内并在昂贵的可控培养条件下发酵细菌,因此通过细菌菌株发酵产生葡糖胺聚糖伴随着高成本(见例如 US4,897,349)。此外,可通过细菌菌株发酵产生的葡糖胺聚糖的量受每种情况中生产装置的限制。例如目前市售乙酰透明质酸的高价格说明这种葡糖胺聚糖虽然具有特定特性(例如粘弹性,结合水的高容量),但不适合于工业应用。

[0049] 因此,与产生葡糖胺聚糖的已知方法相比,根据本发明的植物细胞和根据本发明的植物提供这样的优点,与仅包含葡糖胺聚糖合酶活性的植物细胞或植物相比它们合成了增加量的葡糖胺聚糖(例如乙酰透明质酸)。

[0050] 此处,根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物的遗传修饰可以是导致编码葡糖胺合酶的外来核酸分子及编码具有葡糖胺 6-磷酸乙酰转移酶活性的蛋白质的外来核酸分子及编码具有单功能 UDP-N-乙酰基-葡糖胺焦磷酸化酶活性的蛋白质的外来核酸分子整合到植物细胞或植物的任何遗传修饰。

[0051] 在本发明的上下文中,术语“葡糖胺聚糖合酶”理解为从底物 UDP 葡糖醛酸(UDP-GlcA)和 UDP-N-乙酰基-己糖醛胺(aldohexosamine)(UDP-AldohexNAc)合成葡糖胺聚糖的蛋白质。葡糖胺聚糖的催化根据下文的一般反应方案发生:

[0052]
$$n\text{UDP-GlcA} + n\text{UDP-AldohexNAc} \rightarrow [\text{GlcA-GlcNAc}]_n + 2n\text{UDP}$$

[0053] 优选地,上文反应顺序中制备的 UDP-N-乙酰基-己糖醛胺为 UDP-N-乙酰基-葡糖胺或 UDP-N-乙酰基-半乳糖胺。

[0054] 在优选的实施方案中,本发明涉及根据本发明的植物细胞或根据本发明的植物,其中编码葡糖胺聚糖合酶的外来核酸分子编码乙酰透明质酸合酶或软骨素合酶或肝素/heparosan 合酶。

[0055] 在本发明的上下文中,术语“乙酰透明质酸合酶”(EC 2.4.1.212)理解为从底物 UDP 葡糖醛酸(UDP-GlcA)和 UDP-N-乙酰基-葡糖胺(UDP-GlcNAc)合成乙酰透明质酸的蛋白质。乙酰透明质酸的合成根据下文的反应方案催化:

[0056]
$$n\text{UDP-GlcA} + n\text{UDP-GlcNAc} \rightarrow \beta-1,4-[\text{GlcA}-\beta-1,3-\text{GlcNAc}]_n + 2n\text{UDP}$$

[0057] 可将迄今已研究的乙酰透明质酸合酶划分为两组：I 类乙酰透明质酸合酶和 II 类乙酰透明质酸合酶 (DeAngelis, 1999, CMLS, Cellular and Molecular Life Sciences 56, 670-682)。

[0058] 来自脊椎动物的乙酰透明质酸合酶经鉴定到的同功酶进一步进行区分。多个同功酶以其鉴定的顺序用阿拉伯数字来指代 (例如 hsHAS1、hsHAS2、hsHAS3)。

[0059] 已描述了编码乙酰透明质酸合酶的核酸分子和相应的蛋白质序列, 尤其是以下生物: 兔 (*Oryctolagus cuniculus*) ocHas2 (EMBL AB055978.1, US20030235893)、ocHas3 (EMBL AB055979.1, US 20030235893); 狒狒 (*Papio anubis*) paHas1 (EMBL AY463695.1); 非洲爪蛙 (*Xenopus laevis*) xlHas1 (EMBL M22249.1, US 20030235893)、xlHas2 (DG42) (EMBL AF168465.1)、xlHas3 (EMBL AY302252.1); 人 (*Homo sapiens*) hsHAS1 (EMBL D84424.1, US 20030235893)、hsHAS2 (EMBL U54804.1, US20030235893)、hsHAS3 (EMBL AF232772.1, US 20030235893); 小鼠 (*Mus musculus*) mmHas1 (EMBL D82964.1, US 20030235893)、mmHAS2 (EMBL U52524.2, US 20030235893)、mmHas3 (EMBL U86408.2, US20030235893); 牛 (*Bos taurus*) btHas2 (EMBL AJ004951.1, US20030235893); 鸡 (*Gallus gallus*) ggHas2 (EMBL AF106940.1, US20030235893); 大鼠 (*Rattus norvegicus*) rnHas1 (EMBL AB097568.1, Itano 等, 2004, J. Biol. Chem. 279 (18) 18679-18678)、rnHas2 (EMBL AF008201.1); rnHas3 (NCBI NM_172319.1, Itano 等, 2004, J. Biol. Chem. 279 (18) 18679-18678)、马 (*Equus caballus*) ecHAS2 (EMBL AY056582.1, GI: 23428486)、猪 (*Sus scrofa*) sscHAS2 (NCBI NM_214053.1, GI: 47522921)、sscHas3 (EMBL AB159675)、斑马鱼 (*Danio rerio*) brHas1 (EMBL AY437407)、brHas2 (EMBL AF190742.1) brHas3 (EMBL AF190743.1); 多杀巴斯德氏菌 (*Pasteurella multocida*) pmHas (EMBL AF036004.2); 酿脓链球菌 (*Streptococcus pyogenes*) spHas (EMBL, L20853.1, L21187.1, US 6, 455, 304, US 20030235893); 马链球菌 seHas (EMBL AF347022.1, AY173078.1)、乳房链球菌 (*Streptococcus uberis*) suHasA (EMBL AJ242946.2, US 20030235893)、似马链球菌 (*Streptococcus equisimilis*) seqHas (EMBL AF023876.1, US 20030235893); 硫磺矿硫化叶菌 (*Sulfolobus solfataricus*) ssHAS (US 20030235893)、*Sulfolobus tokodaii* stHas (AP000988.1)、小球藻病毒 1 (*Paramecium bursaria Chlorella virus 1*), cvHAS (EMBL U42580.3, PB42580, US 20030235893)。

[0060] 在优选的实施方案中, 本发明涉及根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物, 其中编码葡萄糖胺聚糖合酶的外来核酸分子的特征在于其编码乙酰透明质酸合酶。正在讨论的编码乙酰透明质酸合酶的外来核酸分子优选为编码病毒乙酰透明质酸合酶的外来核酸分子。优选地, 编码乙酰透明质酸合酶的外来核酸分子编码感染藻类的病毒的乙酰透明质酸合酶。

[0061] 关于藻类感染病毒, 编码乙酰透明质酸合酶的外来核酸分子优选编码小球藻感染病毒 (*Paramecium bursaria Chlorella virus*) 的乙酰透明质酸合酶, 尤其优选为小球藻病毒 1 的乙酰透明质酸合酶并特别优选为 H1 菌株的小球藻病毒的乙酰透明质酸合酶。

[0062] 优选地, 编码乙酰透明质酸合酶的外来核酸分子的特征在于其编码乙酰透明质酸合酶, 所述乙酰透明质酸合酶的氨基酸序列与 SEQ ID NO 2 中显示的氨基酸序列具有至少 70%、优选至少 80%、优选至少 90%、尤其优选为至少 95% 并且特别优选为至少 98% 的同

一性。在尤其优选的实施方案中,编码乙酰透明质酸合酶的外来核酸分子的特征在于其编码具有 SEQ ID NO 2 中显示的氨基酸序列的乙酰透明质酸合酶。

[0063] 在其他实施方案中,编码乙酰透明质酸合酶的外来核酸分子与 SEQ ID NO 1 或 SEQ ID NO 3 中显示的核酸序列具有至少 70%、优选至少 80%、优选至少 90%、尤其优选为至少 95% 并且特别优选为至少 98% 的同一性。在尤其优选的实施方案中,编码乙酰透明质酸合酶的外来核酸分子的特征在于其具有 SEQ ID NO 3 中的核酸序列或特征在于外来核酸分子序列由于遗传密码的简并与 SEQ ID NO 1 或 SEQ ID NO 3 中显示的核酸序列有所不同。

[0064] 2004 年 8 月 25 日,包含编码小球藻病毒乙酰透明质酸合酶的合成核酸分子的质粒 IC 341-222 根据布达佩斯条约以编号 DSM16664 保藏于德国微生物菌种保藏中心 (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH), 德国, 不伦瑞克, Mascheroder Weg 1b, 38124。显示于 SEQ ID NO 2 中的氨基酸序列可来源于整合至质粒 IC 341-222 中的核酸序列的编码区并编码小球藻病毒乙酰透明质酸合酶。

[0065] 因此,本发明还涉及根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物,其中编码乙酰透明质酸合酶的核酸分子的特征在于其编码蛋白质,所述蛋白质的氨基酸序列可来源于插入至质粒 DSM16664 中的核酸序列的编码区,或特征在于其编码蛋白质,所述蛋白质的氨基酸序列与可来源于插入至质粒 DSM16664 中的核酸序列的编码区的氨基酸序列具有至少 70%、优选至少 80%、优选至少 90%、尤其优选为至少 95% 并且特别优选为至少 98% 的同一性。

[0066] 本发明还涉及根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物,其中编码乙酰透明质酸合酶的外来核酸分子的特征在于它是整合至质粒 DSM16664 中的乙酰透明质酸合酶编码核酸序列,或者特征在于它与整合至质粒 DSM16664 中的核酸序列具有至少 70%、优选至少 80%、优选至少 90%、尤其优选至少 95% 并且特别优选至少 98% 的同一性。

[0067] 在本发明的上下文中,术语“软骨素合酶”(EC 2.4.1.175, EC 1.4.1.226) 理解为蛋白质或由两个蛋白质组成的蛋白质复合体,其从底物 UDP 葡萄糖醛酸 (UDP-GlcA) 和 UDP-N-乙酰基-半乳糖胺 (UDP-GalNAc) 合成软骨素。软骨素合成根据下文的反应方案催化:

[0068]
$$n\text{UDP-GlcA} + n\text{UDP-GalNAc} \rightarrow \beta\text{-1,4-[GlcA-}\beta\text{-1,3-GalNAc]}_n + 2n\text{UDP}$$

[0069] 在一些生物体中,由两个不同蛋白质组成的软骨素合酶复合体催化附着在蛋白聚糖上的软骨素分子的延伸。两个蛋白质中的一个蛋白质 N-乙酰基氨基半乳糖基转移酶 II (EC 2.4.1.175) 经 $\beta\text{-1,4-}$ 附着向软骨素分子添加 N-乙酰基-半乳糖胺单体,第二个蛋白质 N-乙酰基氨基半乳糖基蛋白聚糖 3- β 葡萄糖苷酸基转移酶 (EC 2.4.1.226) 经 $\beta\text{-1,3-}$ 附着向软骨素分子添加葡萄糖醛酸单体。然而,本领域技术人员也熟悉双功能蛋白质,在所述双功能蛋白质中,单个蛋白质向软骨素分子添加 N-乙酰基-半乳糖胺单体以及葡萄糖醛酸单体。

[0070] 在其他优选的实施方案中,本发明涉及根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物,其中编码葡糖胺聚糖合酶的外来核酸分子的特征在于其编码软骨素合酶。

[0071] 本发明的优选实施方案涉及根据本发明的经遗传修饰的植物细胞或根据本发明

的经遗传修饰的植物,其中编码软骨素合酶的外来核酸分子编码向软骨素分子附加 N-乙酰基-半乳糖胺单体以及葡糖醛酸单体的双功能软骨素合酶。

[0072] 在本发明的上下文中,术语“双功能软骨素合酶”理解为其中 N-乙酰基氨基半乳糖基转移酶 II (EC 2.4.1.175) 活性以及乙酰氨基半乳糖基蛋白聚糖 3-β-葡糖苷酸基转移酶 (EC 2.4.1.226) 活性存在于一个分子中的蛋白质。

[0073] 例如,已从细菌(例如大肠杆菌 (*Escherichia coli*) US2003109693、EP 1283259) 中描述了编码单功能软骨素合酶的核酸分子和源于其的氨基酸序列。

[0074] 例如,已从哺乳动物中(例如人 WO 03012099、US 2005048604、US2006052335、NCBI acc.No:BC046247.1, BC023531.2; Kitagawa 等, 2001, J. Biol. Chem. 276(42), 38721-38726) 或 *Pasteurella multocida* 中 (US2003104601、EMBL acc.No:AF195517, DeAngelis 和 Padgett-McCue, 2000, J. Biol. Chem. 275(31), 24124-24129) 描述了编码双功能软骨素合酶的核酸分子和源于其的氨基酸序列。

[0075] 编码软骨素合酶的外来核酸分子优选为编码细菌软骨素合酶, 优选编码来自巴斯德菌属的软骨素合酶, 尤其优选编码来自出血败血性巴斯德菌的软骨素合酶的外来核酸分子。

[0076] 优选地, 编码软骨素合酶的外来核酸分子的特征在于其编码软骨素合酶, 所述软骨素合酶的氨基酸序列与 SEQ ID NO 5 中显示的氨基酸序列具有至少 70%、优选为至少 80%、优选至少 90%、尤其优选为至少 95% 并且特别优选为至少 98% 的同一性。在尤其优选的实施方案中, 编码软骨素合酶的外来核酸分子的特征在于它编码具有 SEQ ID NO 5 中显示的氨基酸序列的软骨素合酶。

[0077] 在其他实施方案中, 编码软骨素合酶的外来核酸分子与 SEQ ID NO 4 中显示的核酸序列具有至少 70%、优选为至少 80%、优选至少 90%、尤其优选为至少 95% 并且特别优选为至少 98% 的同一性。在尤其优选的实施方案中, 编码软骨素合酶的核酸分子的特征在于它具有 SEQ ID NO 4 中所示的核酸序列或特征在于外来核酸分子序列由于遗传密码的简并性与 SEQ ID NO 4 中所示的核酸序列有所不同。

[0078] 在本发明的上下文中, 术语“肝素/乙酰肝素合酶”或“heparosan 合酶”(EC 2.4.1.224、EC 2.4.1.225) 理解为从底物 UDP-葡糖醛酸 (UDP-GlcA) 和 UDP-N-乙酰基-葡糖胺 (UDP-GlcNAc) 合成肝素/乙酰肝素的蛋白质或由两个蛋白质组成的酶复合体。肝素/乙酰肝素的合成根据下文的反应方案催化:

[0079]
$$n\text{UDP-GlcA} + n\text{UDP-GlcNAc} \rightarrow \alpha\text{-1,4-}[\text{GlcA-beta-1,4-GlcNAc}]_n + 2n\text{UDP}$$

[0080] 例如, 已从细菌中(出血败血性巴斯德菌 EMBL acc.Nos:AF425591、AF439804、AY292199、AY292200、US 20030099967, 大肠杆菌 EMBL acc.No:X77617.1) 或人中 (NCBI acc.Nos:BC001174.1, NM_207122.1) 描述了编码肝素/乙酰肝素合酶的核酸分子和来源于其的氨基酸序列。

[0081] 在一些生物体中, 由两种不同蛋白质组成的肝素/乙酰肝素合酶酶复合体催化附着在蛋白聚糖的肝素/乙酰肝素分子的延伸。两个蛋白质中的一个蛋白质葡糖苷酸基-N-乙酰基葡糖胺基蛋白聚糖 4-α-N-葡糖胺基转移酶 (EC 2.4.1.224) 经 β-1,4-附着向肝素/乙酰肝素分子添加 N-乙酰基-葡糖胺单体, 第二个蛋白质 N-乙酰基-葡糖胺基蛋白聚糖 4-β-葡糖苷酸基 (glucuronosyl) 转移酶 (EC 2.4.1.225) 经 β-1,3-附着向

肝素 / 乙酰肝素分子添加葡萄糖醛酸单体。然而,本领域技术人员也熟悉双功能蛋白质,在所述双功能蛋白质中,单个蛋白质向肝素 / 乙酰肝素分子添加 N- 乙酰基 - 葡糖胺单体以及葡萄糖醛酸单体。例如已在人中 (Busse 和 Kusche-Gullberg, 2003, J. Biol. Chem. 278(42), 41333-41337) 或巴斯德菌属中 (DeAngelis 和 White, 2004, J. Bacteriology 186(24), 8529-8532) 描述了此类双功能肝素 / 乙酰肝素合酶。具有肝素 / 乙酰肝素合酶活性的双功能蛋白质具有分类为 EC 编号 2. 4. 1. 224 的酶的活性以及分类为 EC 编号 2. 4. 1. 225 的酶的活性。

[0082] 在其他实施方案中,本发明涉及根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物,其中编码葡糖胺聚糖合酶的外来核酸分子的特征在于它编码肝素 / 乙酰肝素合酶。

[0083] 本发明的优选实施方案涉及根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物,其中编码肝素 / 乙酰肝素合酶的外来核酸分子编码向肝素 / 乙酰肝素分子附加 N- 乙酰基 - 葡糖胺单体以及葡萄糖醛酸单体的双功能肝素 / 乙酰肝素合酶。

[0084] 编码肝素 / 乙酰肝素合酶的外来核酸分子优选为编码细菌肝素 / 乙酰肝素合酶,优选编码来自巴斯德菌属的肝素 / 乙酰肝素合酶,尤其优选编码来自出血败血性巴斯德菌的肝素 / 乙酰肝素合酶的外来核酸分子。

[0085] 优选地,编码肝素 / 乙酰肝素合酶的外来核酸分子的特征在于它编码肝素 / 乙酰肝素合酶,所述肝素 / 乙酰肝素合酶的氨基酸序列与 SEQ ID NO 7 中显示的氨基酸序列具有至少 70%、优选为至少 80%、优选至少 90%、尤其优选为至少 95% 并且特别优选为至少 98% 的同一性。在尤其优选的实施方案中,编码肝素 / 乙酰肝素合酶的外来核酸分子的特征在于它编码具有 SEQ ID NO 7 中显示的氨基酸序列的肝素 / 乙酰肝素合酶。

[0086] 在其他实施方案中,编码肝素 / 乙酰肝素合酶的外来核酸分子与 SEQ ID NO 6 中显示的核酸序列具有至少 70%、优选为至少 80%、优选至少 90%、尤其优选为至少 95% 并且特别优选为至少 98% 的同一性。在尤其优选的实施方案中,编码肝素 / 乙酰肝素合酶的核酸分子的特征在于它具有 SEQ ID NO 6 中所示的核酸序列或特征在于外来核酸分子的序列由于遗传密码的简并性而与 SEQ ID NO 6 中所示的核酸序列有所不同。

[0087] 在本发明的上下文中,术语“葡糖胺磷酸 N- 乙酰基转移酶 (乙酰 -CoA :D 葡糖胺磷酸 N- 乙酰基转移酶或 GlcN-P 乙酰基转移酶)” (EC 2. 3. 1. 4) 理解为从底物 D 葡糖胺磷酸 (GlcN-P) 和乙酰 CoA (AcCoA) 合成 N- 乙酰基 -D 葡糖胺磷酸 (GlcNAc-P) 的蛋白质。N- 乙酰基 - 葡糖胺 6- 磷酸的合成根据下文的反应方案催化:

[0088] $\text{GlcN-P} + \text{AcCoA} \rightarrow \text{GlcNAc-P} + \text{CoASH}$

[0089] 在所示的反应方程式中,底物 GlcN-P 可以是葡糖胺 1- 磷酸 (GlcN-1-P) 或葡糖胺 6- 磷酸 (GlcN-6-P)。

[0090] 在正在讨论用于合成 UDP-N- 乙酰基 - 葡糖胺的代谢途径中,所研究的细菌和真核生物间的根本差异在于正在讨论的代谢途径不同的中间产物用作乙酰化反应的底物。在细菌生物中,通过具有 GlcN-1-P 乙酰转移酶 (EC 2. 3. 1. 157) 活性的蛋白质进行 GlcN-1-P 的乙酰化 (Gehring 等, 1996, Biochemistry 35, 579-585), 而在真核动物或真菌中通过具有葡糖胺 6- 磷酸乙酰转移酶 (EC 2. 3. 1. 4) 活性的蛋白质催化 GlcN-6-P 的乙酰化 (Milewski 等, 2006, Yeast 23, 1-14, 在 Wiley InterScience 中在线发表, DOI :10. 1002. /yea. 1337)。

因此,在不同生物中使用不同的底物和不同的蛋白质来合成 UDP-GlcNAc。

[0091] 令人惊奇的是,已发现与现有技术 (WO 2007 023682) 中公开内容不同,不可能通过向植物细胞中引入编码具有 GlcN-P 乙酰转移酶活性的蛋白质的任何核酸分子来在植物细胞中增加合成的葡糖胺聚糖的量。当然,已发现引入外来核酸分子(其编码具有乙酰化 GlcN-1-P 的 GlcN-P 乙酰转移酶活性的蛋白质(例如来自大肠杆菌的 glmu))不能导致植物细胞或植物合成的葡糖胺聚糖量的增加。因此,对于根据本发明的植物细胞或根据本发明的植物而言外来核酸分子编码具有 GlcN-P 乙酰转移酶(其使用 GlcN-6-P 作为乙酰化反应的底物并因此是具有 GlcN-6-P 乙酰转移酶活性的蛋白质 (EC 2.3.1.4)) 活性的蛋白质是至关重要。相反,编码具有 GlcN-1-P 乙酰转移酶活性 (EC 2.3.1.157) (其使用 GlcN-1-P 用于乙酰化反应的) 的蛋白质的外来核酸分子不适合于产生根据本发明的植物细胞或根据本发明的植物。

[0092] 此外,已发现具有外来核酸分子的植物细胞或植物不能合成显著更高量的葡糖胺聚糖,所述外来核酸分子编码葡糖胺 6-磷酸变位酶 (GlcN-6-P 变位酶) (其催化 GlcN-6-P 向 GlcN-1-P 异构化)。

[0093] 在本发明的上下文中,术语“葡糖胺 6-磷酸乙酰转移酶(乙酰 CoA :D 葡糖胺 6-磷酸 N-乙酰基转移酶或 GlcN-6-P 乙酰转移酶)” (EC 2.3.1.4) 应理解为从底物 D-葡糖胺 6-磷酸 (GlcN-6-P) 和乙酰 CoA (AcCoA) 合成 N-乙酰基-D-葡糖胺 6-磷酸 (GlcNAc-6-P) 的蛋白质。N-乙酰基-葡糖胺 6-磷酸的合成根据下文的反应方案催化:

[0094] $\text{GlcN-6-P} + \text{AcCoA} \rightarrow \text{GlcNAc-6-P} + \text{CoASH}$

[0095] 具有 GlcN-6-P 乙酰转移酶活性的蛋白质的功能形式为同质二聚体。单体的三级结构具有中心核区。该核区由具有五条逆平行链 (β 链 1-5) (其由四条 α 螺旋环绕) 和第六条 β 链 (β -6 链) 的 β 折叠片结构组成。在同质二聚体的形成过程中,亚基的 β -6 链与其他各个亚基相应的 β -6 链间具有相互作用。

[0096] SEQ ID No 9 (EMBL acc. No :AB017626.1) 中显示的氨基酸序列编码具有来自酿酒酵母 (*Saccharomyces cerevisiae*) 的 GlcN-6-P 乙酰转移酶活性的蛋白质。在 SEQ ID No 9 中显示的氨基酸序列中,7-11 位氨基酸形成 β -1 链,13-26 位氨基酸形成 α -1 链,37-47 位氨基酸形成 α -2 链,62-69 位氨基酸形成 β -2 链,74-86 位氨基酸形成 β -3 链,92-103 位氨基酸形成 β -4 链,111-125 位氨基酸形成 α -3 链,130-136 位氨基酸形成 β -5 链,139-146 位氨基酸形成 α -3 链以及 154-159 位氨基酸形成 β -6 链。SEQ ID No 9 中显示的序列中的 β -4 链中存在的氨基酸 Glu (98 位)、Asp (99 位) 和 Ile (100 位) 与底物 AcCoA 相互作用,它们将它的羰基键极化并且稳定由 AcCoA 和 GlcN-6-P 以及 GlcN-6-P 乙酰转移酶组成的四聚反应中间产物中的 AcCoA 的氧原子的负电荷。SEQ ID No 9 中显示的序列中的氨基酸 Tyr (143 位) 稳定了待将切除的 CoA 分子的硫醇盐阴离子。SEQ ID No 9 中显示的序列中的氨基酸 Leu (133 位) 支持了反应催化过程中的这些相互作用。在反应催化过程中,GlcN-6-P 结合在同质二聚体的单体间形成的袋之中, β -6 链的氨基酸残基参与了其形成。在反应催化过程中,SEQ ID No9 中显示的序列中的氨基酸 Asp (134 位) 增加了 GlcN-6-P 氨基的亲核性 (Milewski 等,2006,在线发表于 Wiley InterScience, www.interscience.wiley.com, DOI :10.1002/yea.1337)。参与正在讨论的反应催化的具有 GlcN-6-P 乙酰转移酶活性的蛋白质的其他氨基酸描述于 Penneff 等 (2004, J. Biological Chemistry 276(19),

16328-16334, 图 1)。

[0097] 也可以在来自其他生物的编码具有 GlcN-6-P 乙酰转移酶活性的蛋白质的氨基酸序列中鉴定出此处以示例性方式指出的参与催化反应的酿酒酵母氨基酸序列中的氨基酸。例如, 这些是来自白色念珠菌 (*Candida albicans*) (EMBL acc.No :AB017627.1) 的编码具有 GlcN-6-P 乙酰转移酶活性的蛋白质的氨基酸序列中的氨基酸 Glu88、Asp80、Ile90、Asp124 和 Tyr133。

[0098] 已描述了编码具有 GlcN-6-P 乙酰转移酶活性的蛋白质的核酸分子和相应的蛋白质序列, 尤其来自以下生物: 酿酒酵母 (EMBL acc.No :AB017626.1)、裂殖酵母 (*Schizosaccharomyces pombe*) (EMBL acc.No :AB017629.1)、白色念珠菌 (EMBL acc.No :AB017627.1)、米曲霉 (*Aspergillus oryzae*) (EMBL CDS acc.No :BAE62756.1)、秀丽隐杆线虫 (*Caenorhabditis elegans*) (NCBI acc.No :NM_073253.4, EMBL CDS acc.No :BAA63497.1, CAA044531.1)、黑腹果蝇 (*Drosophila melanogaster*) (EMBL CDS acc.No :AAL13916.1)、*Xenopus traopicalis* (EMBL acc.No :CR760021.2)、小鼠 (EMBL CDS acc.No :BAE39886.1)、人 (EMBL CDS acc.No :BAC03482.1)、猩猩 (*Pongo pygmaeus*) (EMBL CDS acc.No :CR858996.1)、*Acanthamoeba polyphaga mimivirus* (EMBL CDS acc.No :AAV50586.1)。尽管如已描述的, 参与催化反应的氨基酸残基在来自多种生物的具有 GlcN-6-P 乙酰转移酶活性的蛋白质中是保守的, 但在一些情况下它们的序列相互间具有非常低的同一性。因此, 来自酿酒酵母的编码具有 GlcN-6-P 乙酰转移酶活性的蛋白质的氨基酸序列 (EMBL acc.No :AB017626.1) 与来自白色念珠菌的相应序列 (EMBL acc.No :AB017627.1) 仅具有 44% 的同一性并且与来自裂殖酵母的相应序列 (EMBL acc.No :AB017629.1) 甚至仅具有 25% 的同一性 (Milewski 等, 2006, 在线发表于 Wiley InterScience, www.interscience.wiley.com, DOI :10.1002/yea.1337)。尽管在正在讨论的氨基酸序列相互间同一性低, 所有上文提及的编码具有 GlcN-6-P 乙酰转移酶活性的蛋白质的序列均适合于产生根据本发明的植物细胞或根据本发明的植物。

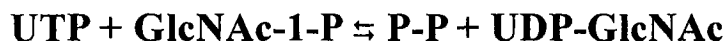
[0099] 根据本发明, 编码具有 GlcN-6-P 乙酰转移酶活性的蛋白质的外来核酸分子可来自任何生物; 优选地, 所述核酸分子来自真菌、动物或植物, 尤其优选来自真菌并特别优选来自酿酒酵母。

[0100] 优选地, 编码 GlcN-6-P 乙酰转移酶的外来核酸分子的特征在于它编码 GlcN-6-P 乙酰转移酶, 其氨基酸序列与 SEQ ID NO 9 中显示的氨基酸序列具有至少 70%、优选为至少 80%、优选至少 90%、尤其优选为至少 95% 并且特别优选为至少 98% 的同一性。在尤其优选的实施方案中, 编码具有 GlcN-6-P 乙酰转移酶活性的蛋白质的外来核酸分子的特征在于它编码具有 GlcN-6-P 乙酰转移酶活性的蛋白质, 所述蛋白质具有 SEQ ID NO 9 中显示的氨基酸序列。

[0101] 在其他实施方案中, 编码具有 GlcN-6-P 乙酰转移酶活性的蛋白质的外来核酸分子与 SEQ ID NO 8 中显示的核酸序列具有至少 70%、优选为至少 80%、优选至少 90%、尤其优选为至少 95% 并且特别优选为至少 98% 的同一性。在尤其优选的实施方案中, 编码 GlcN-6-P 乙酰转移酶的核酸分子的特征在于它具有 SEQ ID NO 8 中显示的核酸序列或特征在于所述外来核酸分子序列由于遗传密码的简并性而与 SEQ ID NO 8 中所示的核酸序列有所不同。

[0102] 在本发明的上下文中,术语“UDP-GlcNAc 焦磷酸化酶 (2- 乙酰胺基 -2- 脱氧 -d- 葡萄糖 1- 磷酸尿苷酰转移酶) (EC 2. 7. 7. 23)”理解为从底物尿苷三磷酸 (UTP) 和 N- 乙酰基 -D- 葡萄糖胺 1- 磷酸 (GlcNAc-1-P) 合成 UDP-N- 乙酰基 - 葡萄糖胺 (UDP-GlcNAc) 的蛋白质。UDP-GlcNAc 的合成根据下文的反应方案催化:

[0103]



[0104] 具有 UDP-GlcNAc 焦磷酸化酶活性的原核生物蛋白质通常为双功能蛋白质,其除了上文显示的反应 (EC 2. 7. 7. 23) 外,还具有葡萄糖胺 1- 磷酸乙酰转移酶 (GlcN-1-P 乙酰转移酶, EC 2. 3. 1. 157) 的功能,即它们催化葡萄糖胺 1- 磷酸 (GlcN-1-P) 向 N- 乙酰基 - 葡萄糖胺 1- 磷酸 (GlcNAc-1-P) 的 N- 乙酰基化 ($\text{GlcN-1-P} + \text{AcCoA} \rightarrow \text{GlcNAc-1-P} + \text{CoASH}$) (Gehring 等, 1996, Biochemistry 35, 579-585)。相反,具有 UDP-GlcNAc 焦磷酸化酶活性的真核生物蛋白质为仅催化上述反应 ($\text{UTP} + \text{GlcNAc-1-P} \rightleftharpoons \text{P-P} + \text{UDP-GlcNAc}$) 的单功能蛋白质 (Mio 等, 1998, J. Biol. Chem. 273 (23), 14392-14397)。

[0105] 在本发明的上下文中,术语“具有 UDP-GlcNAc 焦磷酸化酶活性的单功能蛋白质”理解为催化上文描述的针对具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的反应 ($\text{UTP} + \text{GlcNAc-1-P} \rightleftharpoons \text{P-P} + \text{UDP-GlcNAc}$) 的蛋白质。具有 UDP-GlcNAc 焦磷酸化酶活性的单功能蛋白质不 (额外) 具有催化 GlcN-1-P 向 GlcNAc-1-P 乙酰化的活性。因此,将具有 GlcNAc 焦磷酸化酶活性的单功能蛋白质分类在 EC 编号 2. 7. 7. 23 中,而具有 UDP-GlcNAc 焦磷酸化酶活性的双功能蛋白质具有分类在 EC 编号 2. 7. 7. 23 中的酶的活性以及分类在 EC 编号 2. 3. 1. 157 中的酶的活性。

[0106] 令人惊奇的是,已发现与现有技术中的公开内容 (例如 WO 2007023682) 不同的是,具有编码具有 UDP-GlcNAc 焦磷酸化酶活性和 GlcN-1-P 乙酰转移酶活性的双功能蛋白质 (例如 glmU aus 大肠杆菌, EC2. 7. 7. 23 和 EC 2. 3. 1. 157) 的外来核酸分子的根据本发明的植物细胞或根据本发明的植物不能合成增加量的葡萄糖胺聚糖。因此,编码 UDP-GlcNAc 焦磷酸化酶的外来核酸分子编码具有 UDP-GlcNAc 焦磷酸化酶 (EC 2. 7. 7. 23) 活性的单功能蛋白质对于根据本发明的植物细胞或根据本发明的植物是必需的。因此,编码 UDP-GlcNAc 焦磷酸化酶的外来核酸分子不应该编码蛋白质,其除了刚提及的 UDP-GlcNAc 焦磷酸化酶活性外还具有额外的 GlcN-1-P 乙酰转移酶 (EC 2. 3. 1. 157) 活性。因此,它优选为真核来源的外来核酸分子。此外,令人惊奇的是,已发现与现有技术中的公开内容 (例如 WO 2007 023682) 不同的是,乙酰葡萄糖胺磷酸变位酶 (GlcNAc-P 变位酶, EC 5. 4. 2. 3) 的表达以及具有 GlcN-6-P 乙酰转移酶活性的蛋白质的表达和具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的表达不会导致植物细胞或植物中葡萄糖胺聚糖量的进一步增加。

[0107] 编码具有 UDP-GlcNAc 焦磷酸化酶活性的单功能蛋白质的氨基酸序列包含在蛋白质间高度保守的氨基酸残基。编码具有 UDP-GlcNAc 焦磷酸化酶活性的真核蛋白质的氨基酸序列在每种情况下具有蛋白质间保守的三个结构域。第一个结构域的共有序列为 GlyGly GlnXxxThrArgLeuGlyXxxXxxXxxProLysGly (SEQ ID No 11 中显示的序列的 111-124 位氨基酸),第二个结构域的共有序列为 Pro (Asp 或 Asn)GlyAsn (Gly 或 Ala)GlyXxxXxxXxxAla (SEQ ID No 11 中显示的序列的 219-228 位氨基酸) 并且第三个结构域的共有序列为 LysXxxGluXxxPheXxxPheAspXxxPhe (SEQ ID No 11 中显示的序列的 377-386 位氨基酸),其中 Xxx

为任何氨基酸。具有 UDP-GlcNAc 焦磷酸化酶活性的原核蛋白质（例如 *glmU* 大肠杆菌, EMBL acc. No :EAY46949.1）具有与来自真核生物中相应蛋白质的第一个结构域相似的保守结构域 (GlyXxxGlyThr(Arg 或 Ser)XxxXxxXxxXxxProLys)。对于真核蛋白质的第二个或第三个结构域,在原核蛋白质中未发现相应的结构域。(Mok 和 Edwards,2000, J. Biol. Chem. 280(47),39363-39372)

[0108] SEQ ID No 11 中显示的氨基酸序列中的氨基酸 Gly(112 位)、Gly(114 位)、Thr(115 位)、Arg(116 位)、Pro(122 位)和 Lys(123 位)在编码具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的一级序列中是保守的。SEQ ID No 11 中显示的氨基酸序列中的氨基酸 Gly(112 位)、Arg(116 位)或 Lys(123 位)的替换导致近乎失活的蛋白质。相反,SEQ ID No 11 中显示的氨基酸序列中的氨基酸 Gly(114 位)、Thr(115 位)或 Pro(122 位)的替换仅显示了正在讨论的蛋白质活性的降低。因此,SEQ ID No 11 中显示的氨基酸序列中的氨基酸 Gly(112 位)、Arg(116 位)和 Lys(123 位)是在具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质中具有催化功能的氨基酸 (Mio 等,1998, J. Biol. Chem. 273(23),14392-14397)。

[0109] 在编码来自肠贾第鞭毛虫 (*Giardia intestinales*) 的具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的氨基酸序列中 (EMBL acc. No :AAM54702.1),氨基酸 Gly(108 位)对应于 SEQ ID No 11 中显示的序列的氨基酸 Gly(112 位)。编码来自肠贾第鞭毛虫的具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的氨基酸序列中的氨基酸 Gly(108 位)经氨基酸 Ala 的替换也导致了蛋白质活性的几乎完全降低 (Mok 和 Edwards,2005, J. Biol. Chem. 280(47),39363-39372)。编码来自人的具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的氨基酸序列 (EMBL acc. No :BAA31202.1) 中的氨基酸 Gly(111 位) (其对应于 SEQ ID No 11 中显示的序列中的氨基酸 Gly(112 位)) 的替换也导致了活性的几乎完全降低 (Wang-Gillam 等,2000, J. Biol. Chem. 275(2),1433-1438)。

[0110] 编码来自人的 UDP-GlcNAc 焦磷酸化酶的蛋白质 (EMBL acc. No :BAA31202) 中的氨基酸 Gly(222 位)和来自肠贾第鞭毛虫的相应蛋白质 (EMBL acc. No :AAM54702.1) 中的相应氨基酸 Gly(210 位)的替换在两种情况下均同样导致了活性的几乎完全丧失,这说明所述氨基酸同样为参与了催化的氨基酸 (Mok 和 Edwards,2005, J. Biol. Chem. 280(47),39363-39372)。编码来自人的 UDP-GlcNAc 焦磷酸化酶的蛋白质 (EMBL acc. No :BAA31202) 中的氨基酸 Gly(224 位)的替换导致了蛋白质活性相当大但不完全的丧失,并且氨基酸 Pro(222 位)的替换仅导致了活性轻微下降。从中推断出 SEQ ID No 11 中显示的序列的氨基酸 Gly(221 位)和 Gly(223 位)参与了 UTP 的识别,并且在 SEQ ID No 11 中显示的序列的氨基酸 Gly(111 位)和 Gly(112 位) (在各自一级序列中保守) 参与了结合 GlcNAc-1-P (Wang-Gillam 等,2000, J. Biol. Chem. 275(2),1433-1438)。

[0111] 已描述了编码具有上文提及特性的具有 UDP-GlcNAc 焦磷酸化酶单功能活性的蛋白质的核酸分子和相应蛋白质序列,尤其是在以下生物中:肠贾第鞭毛虫 (*Giardia intestinales*) (EMBL acc. No :AAM54702.1)、酿酒酵母 (EMBL acc. No :X79380.1, NCBI protein ID :accession No :CAA557927)、白色念珠菌 (NCBI acc. No :XM_715480.1)、*Pichia stipitis* (NCBI acc. No :XM_001385151.1)、小鼠 (NCBI acc. No :NM_133806.4)、*Canis lupus* (NCBI acc. No :XM_844774.1)、牛 (NCBI acc. No :NM_001046404.1)、*Xenopus tropicalis* (NM_001011142.1)、非洲爪蛙 (*Xenopus laevis*) (NCBI acc. No :BC077836.1)、

拟南芥 (*Arabidopsis thaliana*) (NCBI acc.No :NM_102845.4)、*Danio rerio* (NCBI acc.No :NM_212621.1)、人 (NCBI acc.No :NM_003115.3, EMBL acc.No. :BAA31202.1)。

[0112] 根据本发明,编码具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的外来核酸分子可来自任何真核生物;优选地,所述核酸分子来自真菌、动物或植物,尤其优选来自真菌,特别优选来自酿酒酵母。

[0113] 优选地,编码具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的外来核酸分子的特征在于它编码 UDP-GlcNAc 焦磷酸化酶,其氨基酸序列与 SEQ ID NO 11 中显示的氨基酸序列具有至少 70%、优选为至少 80%、优选至少 90%、尤其优选为至少 95% 并且特别优选为至少 98% 的同一性。在尤其优选的实施方案中,编码具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的外来核酸分子的特征在于它编码具有 UDP-GlcNAc 焦磷酸化酶活性的且具有 SEQ ID NO 11 中显示的氨基酸序列的蛋白质。

[0114] 在其他实施方案中,编码具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的外来核酸分子与 SEQ ID NO 10 中显示的核酸序列具有至少 70%、优选为至少 80%、优选至少 90%、尤其优选为至少 95% 并且特别优选为至少 98% 的同一性。在尤其优选的实施方案中,编码 UDP-GlcNAc 焦磷酸化酶的核酸分子的特征在于它具有 SEQ ID NO 10 中显示的核酸序列或特征在于所述外来核酸分子序列由于遗传密码的简并性而与 SEQ ID NO 10 中所示的核酸序列有所不同。

[0115] 在本发明的上下文中,术语“外来核酸分子”理解为分子,其不天然存在于相应的野生型植物细胞中或不天然存在于野生型植物细胞中具体的空间排布中或其定位于野生型植物细胞的基因组中其天然不存在的位点上。

[0116] 优选地,外来核酸分子为包含不同元件的重组分子,所述元件的组合或特异的空间排列不天然存在于植物细胞中。

[0117] 在本发明的上下文中,术语“重组核酸分子”理解为包含不同核酸分子(其不以在重组核酸分子中存在的组合天然存在)的核酸分子。因此,重组核酸分子除了包含编码葡糖胺聚糖合酶和/或具有 GlcN-6-P 乙酰转移酶活性的蛋白质和/或具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的核酸分子之外,还包含不与所述核酸分子存在天然组合的核酸序列。以在重组核酸分子上与编码葡糖胺聚糖合酶和/或具有 GlcN-6-P 乙酰转移酶活性的蛋白质和/或具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的核酸分子组合存在的所述额外核酸分子可以是任何序列。例如,它们可以是植物基因组核酸序列。额外的核酸序列优选为调控序列(启动子、终止信号、增强子),尤其优选为在植物组织中有活性的调控序列,特别优选为在植物组织中有活性的组织特异性调控序列。用于产生重组核酸分子的方法为本领域技术人员所知并且包括基因工程方法,例如通过连接来结合核酸分子、遗传重组或核酸分子的从头合成(例如见 Sambrook 等, *Molecular Cloning, A Laboratory Manual*, 第三版 (2001) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. ISBN : 0879695773, Ausubel 等, *Short Protocols in Molecular Biology*, John Wiley & Sons ; 第五版 (2002), ISBN : 0471250929)。

[0118] 本发明优选的实施方案涉及根据本发明的植物细胞或根据本发明的植物,其中外来核酸分子稳定整合到植物细胞或植物的基因组中。

[0119] 其基因组中(稳定)整合了外来核酸分子或者其基因组中整合了多个外来核酸

分子的经遗传修饰的植物细胞和经遗传修饰的植物可以与野生型植物细胞和野生型植物区分开来,尤其是通过它们分别包含野生型植物细胞和野生型植物中天然不存在的外来核酸分子的事实,或因为此分子整合到根据本发明的经遗传修饰的植物细胞基因组或根据本发明的经遗传修饰的植物基因组这样的位点上,(其分别整合到在野生型植物细胞和野生型植物中不发生的位点上,即整合在不同的基因组环境中)的事实而区分开来,所述外来核酸分子编码葡糖胺聚糖合酶和/或具有 GlcN-6-P 乙酰转移酶活性的蛋白质和/或具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质。此外,此类根据本发明的经遗传修饰的植物细胞和根据本发明的经遗传修饰的植物可分别与非遗传修饰的野生型植物细胞和非遗传修饰的野生型植物区分开来,因为如果适当时除了包含天然存在于野生型植物细胞或野生型植物中的该分子的拷贝外,还包含整合到其基因组中的至少一个拷贝的外来核酸分子。如果引入根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物中的外来核酸分子是早已天然存在于野生型植物细胞或野生型植物中的分子的额外拷贝,根据本发明的经遗传修饰的植物细胞和根据本发明的经遗传修饰的植物可分别与非遗传修饰的野生型植物细胞和野生型植物区分开来,尤其是通过该额外拷贝/这些额外拷贝分别定位在野生型植物细胞基因组和野生型植物基因组中非天然存在的位点的事实区分开来。

[0120] 可通过遗传方法和/或分子生物学方法来证实核酸分子向植物细胞或植物的基因组中的整合。核酸分子向植物细胞的基因组或植物基因组中的稳定整合的特征在于在已遗传了所述核酸分子的后代中,稳定整合的核酸分子所处的基因组环境与亲代中的基因组环境相同。使用本领域技术人员已知的方法,尤其是在 RFLP 分析(限制性片段长度多态性)的 DNA 印迹分析(Nam 等,1989, The Plant Cell 1,699-705;Leister 和 Dean,1993, The Plant Journal 4(4),745-750),利用基于 PCR 的方法,例如,扩增片段中长度差异的分析(扩增片段长度多态性,AFLP)(Castiglioni 等,1998,Genetics 149,2039-2056;Meksem 等,2001,Molecular Genetics and Genomics 265,207-214;Meyer 等,1998,Molecular and General Genetics 259,150-160)或使用采用限制性内切酶切割扩增片段(切割扩增多态性序列,CAPS)(Konieczny 和 Ausubel,1993, The Plant Journal 4,403-410;Jarvis 等,1994, Plant Molecular Biology 24,685-687;Bachem 等,1996, The Plant Journal 9(5),745-753)的帮助下证实植物细胞的基因组中或植物的基因组中核酸序列的稳定整合的存在。

[0121] 在本发明的上下文中,术语“野生型植物细胞”理解为植物细胞,其作为用于产生根据本发明的经遗传修饰的植物细胞的起始材料,即除去引入的遗传修饰和导致的编码葡糖胺聚糖合酶和/或具有 GlcN-6-P 乙酰转移酶活性的蛋白质和/或具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的核酸分子的整合外,它们的遗传信息对应于根据本发明的经遗传修饰的植物细胞的遗传信息。

[0122] 在本发明的上下文中,术语“野生型植物”理解为植物,其作为用于产生根据本发明的经遗传修饰的植物的起始材料,即除去引入的遗传修饰和导致的编码葡糖胺聚糖合酶和/或具有 GlcN-6-P 乙酰转移酶活性的蛋白质和/或具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的核酸分子的整合外,它们的遗传信息对应于根据本发明的经遗传修饰的植物的遗传信息。

[0123] 在本发明的上下文中,术语“基因组”理解为存在于植物细胞中的全部遗传物质。

对于本领域的技术人员已知的是,除了细胞核外,其他区室(例如质体、线粒体)也包含遗传物质。

[0124] 可获得向植物宿主细胞中(稳定)整合核酸分子的大量技术。这些技术包括使用根瘤农杆菌(*Agrobacterium tumefaciens*)或发根土壤杆菌(*Agrobacterium rhizogenes*)用 t-DNA 作为转化方法的植物细胞转化、原生质体融合、注射、DNA 电穿孔术、通过生物射弹轰击方法引入 DNA 和其他选项(综述见于“Transgenic Plants”, Leandro 编辑, Humana Press 2004, ISBN 1-59259-827-7)。

[0125] 已深入研究并在 EP 120516 ;Hoekema, IN: The Binary Plant Vector System Offsetdrukkerij Kanters B. V. Alblasterdam(1985), 第五章 ;Fraley 等, Crit. Rev. Plant Sci. 4, 1-46 和 An 等 EMBO J. 4, (1985), 277-287 详尽地描述了农杆菌介导植物细胞转化的用途。对于马铃薯的转化,例如见 Rocha-Sosa 等, EMBO J. 8, (1989), 29-33, 对于番茄植株的转化例如见 US5, 565, 347。

[0126] 也已描述了使用基于农杆菌转化的载体转化单子叶植物(Chan 等, Plant Mol. Biol. 22, (1993), 491-506 ;Hiei 等, Plant J. 6, (1994) 271-282 ;Deng 等, Science in China 33, (1990), 28-34 ;Wilmink 等, Plant Cell Reports 11, (1992), 76-80 ;May 等, Bio/Technology 13, (1995), 486-492 ;Conner 和 Domisse, Int. J. Plant Sci. 153(1992), 550-555 ;Ritchie 等, Transgenic Res. 2, (1993), 252-265)。用于转化单子叶植物的备选系统为使用生物射弹轰击方法的转化(Wan 和 Lemaux, Plant Physiol. 104, (1994), 37-48 ;Vasil 等, Bio/Technology 11(1993), 1553-1558 ;Ritala 等, Plant Mol. Biol. 24, (1994), 317-325 ;Spencer 等, Theor. Appl. Genet. 79, (1990), 625-631)、原生质体转化、部分透化处理细胞的电穿孔术、使用玻璃纤维引入 DNA。具体而言,已在文献中数次描述了玉米的转化(参考,例如, W095/06128、EP0513849、EP0465875、EP0292435 ;Fromm 等, Biotechnology 8, (1990), 833-844 ;Gordon Kamm 等, Plant Cell 2, (1990), 603-618 ;Kozziel 等, Biotechnology 11(1993), 194-200 ;Moroc 等, Theor. Appl. Genet. 80, (1990), 721-726)。也已描述了其他牧草的转化,例如柳枝稷(*Panicum virgatum*, Somleva 等, 2002 Crop Science 42 :2080-2087 ;Richards 等, 2001, Plant Cell Reporters 20, 48-54)或甘蔗(Bower 和 Birch, 1992, Plant Journal 2(3), 409-416 ;Bower 等, 1996 Molecular Breeding 2, 239-249 ;Arencibia 等, 1998, Transgenic Research 7, 213-222)或黍(Casas 等, 1993, PNAS 90, 11212-11216 ;US 6, 369, 298)。

[0127] 还已描述了其他谷类物种的成功转化,例如大麦(Wan 和 Lemaux, s. o. ;Ritala 等, s. o. ;Krens 等, Nature 296, (1982), 72-74)和小麦(Nehra 等, Plant J. 5, (1994), 285-297 ;Becker 等, 1994, Plant Journal 5, 299-307)。所有上述方法均适合于本发明的上下文中。

[0128] 与现有技术相比,根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物提供的优势在于它们比仅具有葡糖胺聚糖合酶活性的植物产生更高量的葡糖胺聚糖(例如乙酰透明质酸)。由于从具有更高葡糖胺聚糖含量的植物中分离葡糖胺聚糖复杂度更低并且更节省成本,因此这允许以很少的花费来生产葡糖胺聚糖。此外,与现有技术中描述的植物相比,使用根据本发明的经遗传修饰的植物需要更少的栽种面积来生产葡糖胺聚糖。这使得提供甚至用于工业应用的足够量的葡糖胺聚糖成为可能,而由于葡糖胺聚糖的缺乏和高价格目前工业应用上是不使用它的。小球藻属的病毒感染植物生物体不适合

于产生相对大量的葡糖胺聚糖（乙酰透明质酸）。在生产葡糖胺聚糖（乙酰透明质酸）中，病毒感染的藻类具有葡糖胺聚糖合酶所需的基因不能稳定整合至它们的基因组中的劣势（Van Etten 和 Meints, 1999, *Annu. Rev. Microbiol.* 53, 447-494），因此，对于生产葡糖胺聚糖（乙酰透明质酸），不得不重复进行病毒感染。因此，不可能分离到连续合成所需品质和数量的葡糖胺聚糖（乙酰透明质酸）的独立小球藻细胞。此外，在病毒感染的小球藻中，葡糖胺聚糖（乙酰透明质酸）仅在有限的时期内产生，并因为病毒引起的裂解，在感染仅约 8 小时后藻类会被杀死（Van Etten 等, 2002, *Arch Virol* 147, 1479-1516）。相反，本发明提供根据本发明的经遗传修饰的植物细胞和根据本发明的经遗传修饰的植物具有可以不受限地以无性或有性方式来繁殖并且它们持续产生葡糖胺聚糖（乙酰透明质酸）的优势。

[0129] 在 WO 05 012529 中描述的具有编码乙酰透明质酸合酶的核酸分子的转基因植物合成相对少量的葡糖胺聚糖（乙酰透明质酸）。相反，本发明提供根据本发明的经遗传修饰的植物细胞和根据本发明的经遗传修饰的植物合成显著更高量的葡糖胺聚糖的优势。

[0130] 因此，本发明还提供合成葡糖胺聚糖的根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物。

[0131] 在优选的实施方案中，根据本发明的植物细胞或根据本发明的植物合成选自软骨素、肝素 / 类肝素和乙酰透明质酸的葡糖胺聚糖。

[0132] 为了测定根据本发明的经遗传修饰的植物中与鲜重相关的葡糖胺聚糖含量，优选使用植物整个地上部分材料，即除了根之外的所有植物部分。

[0133] 可通过分离它们合成的葡糖胺聚糖并确定其结构来鉴定合成葡糖胺聚糖的根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物。由于植物组织具有不含任何葡糖胺聚糖的优势，可使用简易快速的分离方法来证明在根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物中葡糖胺聚糖的存在。由于植物组织还具有不含任何葡糖胺聚糖降解酶的优势，可使用简易快速的分离方法来证明在根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物中葡糖胺聚糖的存在。为此，将水加到待检测的植物组织中，随后将植物组织机械粉碎（例如在砂磨机、Warring Blender、液汁提取机等帮助下）。需要时，然后可向悬浮液中加入更多的水，并随后通过离心去除细胞碎片和水不溶性组分。例如，可使用与相关葡糖胺聚糖（例如乙酰透明质酸）特异性结合的蛋白质来证实离心后获得的上清液中葡糖胺聚糖（例如乙酰透明质酸）的存在。

[0134] 用于不同葡糖胺聚糖的基于免疫学试剂的此类检测试剂盒（ELISA）为本领域技术人员所知并且可通过商业途径获得（例如用于肝素检测试剂盒：Lifespan Technologies, 2401 Foothill Drive, Salt Lake City, UT 84109-1405, 产品号：K-2100）。

[0135] 例如在 US 5, 019, 498 中描述了在特异性结合乙酰透明质酸的蛋白质的帮助下检测乙酰透明质酸的方法。US 5, 019, 498 中描述的进行该方法的检测试剂盒可通过商业途径获得（例如来自 Corgenix, Inc., Colorado, USA, 产品号 029-001 的透明质酸（HA）检测试剂盒；也见于一般方法第四项）。

[0136] 例如在免疫学方法的帮助下检测软骨素（Mizuguchi 等, 2003, *Nature* 423, 443-448）。

[0137] 也可使用其他分析方法，例如 IR、NMR 或质谱来确认离心上清液中葡糖胺聚糖的存在。

[0138] 由于已观察到随着根据本发明的植物发育的时间,葡糖胺聚糖积累于植物组织中,所以尤其优选在收获正在研究的植物时或在收获前(一或两)天来测定根据本发明的经遗传修饰的植物中与鲜重相关的葡糖胺聚糖含量。

[0139] 在优选的实施方案中,本发明涉及根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物,其特征在于与仅具有编码葡糖胺聚糖合酶的外来核酸分子的经遗传修饰的植物细胞或经遗传修饰的植物相比,或与具有编码葡糖胺聚糖合酶的外来核酸分子但不具有编码具有 UDP-GlcNAc 乙酰转移酶活性的外来核酸分子和不具有编码具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的外来核酸分子的经遗传修饰的植物细胞或经遗传修饰的植物相比,它们产生增加量的葡糖胺聚糖。

[0140] 优选地,与(仅)具有葡糖胺聚糖合酶活性的相应经遗传修饰的植物细胞或相应经遗传修饰的植物相比,就根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物中植物材料鲜重产生的葡糖胺聚糖的量增加至少 1.2 倍,优选至少 1.4 倍,尤其优选至少 1.6 倍并且特别优选至少 1.8 倍。为了测定与根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物中植物材料鲜重相关的葡糖胺聚糖含量的增加,优选使用在根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物分别与(仅)具有葡糖胺聚糖合酶活性的相应的植物细胞或植物相比,其中用植物细胞或植物的等量材料(例如叶、块茎)进行比较,其中在相同条件下培养从中获取该材料的所述植物细胞或植物,并且其中将比较具有可比较年龄(发育阶段)的植物材料的葡糖胺聚糖含量。例如,植物的幼叶不应与不同植物的老叶相比较。同样适用于植物整个地上部分的葡糖胺聚糖含量的测定。待用于比较的植物应已在可比较的条件下培养并且具有相同的发育阶段。

[0141] 在优选的实施方案中,本发明涉及根据本发明的植物细胞或根据本发明的植物每克植物材料鲜重(FW)合成至少 160 μ g、优选为至少 180 μ g、尤其优选为至少 200 μ g、特别优选为至少 225 μ g 并且最优选为至少 250 μ g 的葡糖胺聚糖。

[0142] 在其他实施方案中,根据本发明的植物细胞或根据本发明的植物每克植物材料鲜重(FW)合成最多 450 μ g、优选为最多 400 μ g、尤其优选为最多 300 μ g、特别优选为最多 280 μ g 并且最优选为最多 260 μ g 的葡糖胺聚糖。

[0143] 在本发明的其他实施方案中,根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物分别为合成葡糖胺聚糖的绿色陆生植物的植物细胞或绿色陆生植物。

[0144] 在本发明的上下文中,术语“绿色陆生植物(有胚植物)”理解为如在 Strasburger,“Lehrbuch der Botanik”[Textbook of Botany],第 34 版,Spektrum Akad. Verl.,1999,(ISBN 3-8274-0779-6)中所定义的。

[0145] 本发明的优选实施方案涉及是多细胞生物的根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物。因此,该实施方案涉及不来自于单细胞植物(原生生物)或不是原生生物的植物细胞或植物。

[0146] 原则上根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物可以分别是任何植物物种,即单子叶植物和双子叶植物的植物细胞或植物。它们优选为农作物植物,即由人培养用于为人或动物提供食物或用于产生生物量目的和/或制备用于技术、工业目的的物质的植物(例如玉米、稻、小麦、苜蓿、黑麦、燕麦、大麦、木薯、马铃薯、番茄、柳枝稷(*Panicum virgatum*)、西米、绿豆、豌豆、高粱、胡萝卜、茄子、萝卜、油菜、大豆、花

生、黄瓜、南瓜、甜瓜、韭菜、大蒜、卷心菜、菠菜、红薯、芦笋、西葫芦、莴苣、洋葱、甜玉米、防风草、洋葱、菊芋、香蕉、甜菜、甘蔗、甜菜根、绿花椰菜、卷心菜、洋葱、黄甜菜、蒲公英、草莓、苹果、杏、李、桃、葡萄藤、花椰菜、芹菜、bell peppers、swede、大黄)。尤其优选为玉米、甘蔗、红薯或粟糖(sugar millet),非常特别优选为番茄或马铃薯植物。

[0147] 在优选的实施方案中,本发明涉及根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物,其中编码蛋白质的外来核酸分子的特征在于与编码基源生物的各个蛋白质的核酸分子的密码子相比外来核酸分子的密码子被修饰。特别优选地,已将外来核酸分子的密码子修饰使得它们适合于其基因组中整合了或待整合它们的植物细胞或植物的密码子使用频率。

[0148] 由于遗传密码的简并性,氨基酸可由一个或更多个密码子编码。在不同生物中,编码氨基酸的密码子以不同频率使用。将编码核酸序列的密码子与其基因组中将整合待表达序列的植物细胞或植物中密码子使用频率相适应可有助于在特定植物细胞或植物中增加翻译正在讨论的蛋白质的量和/或提高正在讨论的mRNA的稳定性。可由本领域技术人员针对编码特定氨基酸的特定密码子的频率通过测定正在讨论的生物中尽可能多的编码核酸序列来测定正在讨论的植物细胞或植物中密码子的使用频率。特定生物的密码子使用频率为本领域技术人员所知并且可使用计算机程序以简易快速的方式测定。合适的计算机程序是对公众开放的并且尤其是互联网免费提供的(例如<http://gcua.schoedl.de/>; <http://www.kazusa.or.jp/codon/>; <http://www.entelechon.com/eng/cutanalysis.html>)。

[0149] 可通过体外诱变或优选地通过基因序列的从头合成来使编码核酸序列的密码子适应于其基因组将整合待表达序列的植物细胞或植物中的密码子使用频率。用于核酸序列从头合成的方法为本领域技术人员所知。例如可通过开始合成单一核酸寡核苷酸、将其与互补的寡核苷酸杂交使得它们形成DNA双链,并随后将单一双链寡核苷酸连接以获得所需核酸来进行从头合成。还可以将包括与向特定目的生物使用的密码子频率相适应的核酸序列的从头合成交给提供该服务的公司(例如Entelechon GmbH, Regensburg, Germany)。

[0150] 在本发明的上下文中,术语“同一性”指核酸分子的编码区全长或编码蛋白质的氨基酸全长上至少60%的序列相同、尤其为至少70%、优选为至少80%、尤其优选为至少90%并且特别优选为至少95%相同。在本发明的上下文中,术语“同一性”理解为与其他蛋白质/核酸的相同氨基酸/核苷酸的数目(同一性),以百分比表示。优选地,在计算机程序的帮助下通过将其他蛋白质/核酸与SEQ ID NO 2中提供的氨基酸序列相比较来测定与具有乙酰透明质酸合酶活性的蛋白质相关的同一性,并通过与SEQ ID NO 1或SEQ ID NO 3中提供的核酸序列相比较来测定与编码具有乙酰透明质酸合酶活性的蛋白质的核酸分子相关的同一性,通过与SEQ ID NO 5中显示的氨基酸序列相比较来测定与具有软骨素合酶活性的蛋白质相关的同一性或者通过与SEQ ID NO 4中显示的核酸序列相比较来测定与编码具有软骨素合酶活性的蛋白质的核酸分子相关的同一性,通过与SEQ ID NO 7中显示的氨基酸序列相比较来测定与具有肝素/heparosan合酶活性的蛋白质相关的同一性或者通过与SEQ ID NO 6中显示的核酸序列相比较来测定与编码具有肝素/heparosan合酶活性的蛋白质的核酸分子相关的同一性,通过与SEQ ID NO 9中显示的氨基酸序列相比较来测定与具有GlcNAc-6-P乙酰转移酶活性的蛋白质相关的同一性或者通过与SEQ ID NO 8中显示的核酸序列相比较来测定与编码具有GlcNAc-6-P乙酰

转移酶活性的蛋白质的核酸分子相关的同一性,通过与 SEQ ID NO 11 中提供的氨基酸序列相比较来测定与具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质相关的同一性或者通过与 SEQ ID NO 10 中显示的核酸序列相比较来测定与编码具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的核酸分子相关的同一性。如果用于彼此比较的序列长度不同,那么通过测定较短序列与较长序列共有氨基酸数目的同一性百分比来测定同一性。优选地,使用已知且公开可得的计算机程序 ClustalW(Thompson 等, *Nucleic Acids Research* 22(1994), 4673-4680) 来测定同一性。经 Julie Thompson(Thompson@EMBL-Heidelberg.DE) 和 Toby Gibson(Gibson@EMBL-Heidelberg.DE), European Molecular Biology Laboratory, Meyerhofstrasse 1, D 69117 Heidelberg, Germany 使 ClustalW 成为公开可用的。还可以从不同的互联网页上下载 ClustalW,尤其是从 IGBMC(Institut de Génétique et de Biologie Moléculaire et Cellulaire, B.P. 163, 67404 Illkirch Cedex, France ;ftp://ftp-igbmc.u-strasbg.fr/pub/) 和从 EBI(ftp://ftp.ebi.ac.uk/pub/software/) 以及所有 EBI(European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB101SD, UK) 的镜像互联网页上下载。

[0151] 优选地,使用 1.8 版本的 ClustalW 计算机程序来测定本发明上下文中描述的蛋白质与其他蛋白质间的同一性。此处,需将参数按如下设定:KTUPLE = 1, TOPDIAG = 5, WINDOW = 5, PAIRGAP = 3, GAOPEN = 10, GAPEXTEND = 0.05, GAPDIST = 8, MAXDIV = 40, MATRIX = GONNET, ENDGAPS(OFF), NOPGAP, NOHGAP。

[0152] 优选地,使用 1.8 版本的 ClustalW 计算机程序来测定例如本发明上下文中描述的核酸分子的核苷酸序列与其他核酸分子的核苷酸序列间的同一性。此处,需将参数按如下设定:KTUPLE = 2, TOPDIAGS = 4, PAIRGAP = 5, DNAMATRIX:IUB, GAOPEN = 10, GAPEXT = 5, MAXDIV = 40, TRANSITIONS:未加权。

[0153] 此外,同一性表示正在讨论的核酸分子或由其编码的蛋白质间功能和/或结构的等同。与上文描述的分子同源的核酸分子以及这些分子的代表性衍生物通常为代表具有相同生物学功能的修饰的这些分子的变异。它们可以是天然存在的变异,例如来自其他物种的序列,或突变,其中这些突变可能以天然方式发生或通过定向突变来引入。此外,变异可以是合成产生的序列。等位变体可以是天然发生的变体或合成产生的变体或通过重组 DNA 技术产生的变体。衍生物的特殊形式例如为核酸分子,其由于遗传密码的简并性而与本发明上下文中描述的核酸分子有所不同。

[0154] 由不同核酸分子衍生物编码的蛋白质具有某些共有特征。

[0155] 这些例如可以是生物学活性、底物特异性、分子量、免疫学反应性、构象等等。

[0156] 本发明还提供了根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物,其特征在于整合到植物细胞或植物的基因组中的编码葡糖胺聚糖合酶和编码具有 GlcNAc-6-P 乙酰转移酶活性的蛋白质和/或编码具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的外来核酸分子与植物细胞中起始转录的调节元件(启动子)相连接。这些可以是同源或异源启动子。启动子可以是组成型的、组织特异的、发育特异的或受外部因素调节的(例如在应用化学底物后,在非生物因素,如热和/或寒冷、干旱、疾病等的作用下)。此处,整合到根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物的基因组中的编码葡糖胺聚糖合酶或具有 GlcNAc-6-P 乙酰转移酶活性的蛋白质或具有 UDP-GlcNAc

焦磷酸化酶活性的蛋白质的核酸分子在每种情况下可以与相同的启动子结合,或不同启动子可与各条序列相结合。此处,两个或三个不同启动子可以任何组合存在,在每种情况下与根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物中编码葡萄糖胺聚糖合酶或具有 GlcNAc-6-P 乙酰转移酶活性的蛋白质或具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的相关外来核酸分子相结合。

[0157] 本发明的优选实施方案涉及根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物,其中选自编码葡萄糖胺聚糖合酶或具有 GlcNAc-6-P 乙酰转移酶活性的蛋白质或具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的核酸分子中的至少一个外来核酸分子、尤其优选为至少两个外来核酸分子、特别优选为三个外来核酸分子与组织特异的启动子相连接。优选的组织特异启动子为特异地在植物块茎、果实或种子细胞或叶中起始转录的启动子。

[0158] 通常,在植物细胞中具有活性的每一启动子均适合于表达编码葡萄糖胺聚糖合酶或具有 GlcNAc-6-P 乙酰转移酶活性的蛋白质或具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的核酸分子。

[0159] 此处,可选择启动子使得表达为组成型的或仅在特定组织中、在植物特定的发育时间点上或在由外界因素决定的时间点上表达。关于植物和关于待表达的核酸分子,启动子可以是同源的或异源的。

[0160] 例如,合适的启动子为用于组成型表达的花椰菜花叶病毒的 35S RNS 启动子或来自玉米的泛素启动子 (Christensen 和 Quail,1996, Transgenic Research 5(3), 213-218)、来自黍的高粱醇溶蛋白启动子 (De Rose 等,1996, Plant Molecular Biology 321029-1035 ;Mishra 等,2007, Molecular Biology Reports 在线 :2February 2007, DOI : 10.1007/s11033-007-9056-8) 或洋丁香属 (Cestrum) YLCV 启动子 (Yellow Leaf Curling Virus ;WO 0173087 ;Stavolone 等,2003, Plant Mol. Biol. 53,703-713)、马铃薯中用于块茎特异表达的 patatingen 启动子 B33 (Rocha-Sosa 等, EMBO J. 8(1989), 23-29 或番茄果实特异的启动子,例如来自番茄的多聚半乳糖醛酸酶启动子 (Montgomery 等,1993, Plant Cell 5,1049-1062) 或来自番茄的 E8 启动子 (Metha 等,2002, Nature Biotechnol. 20(6), 613-618) 或来自桃的 ACC 氧化酶启动子 (Moon 和 Callahan,2004, J. Experimental Botany 55(402),1519-1528) 或确保仅在光合活性组织中表达的启动子,例如 ST-LS1 启动子 (Stockhaus 等, Proc. Natl. Acad. Sci. USA 84(1987),7943-7947 ;Stockhaus 等人, EMBO J. 8(1989),2445-2451) 或来自小麦的用于胚乳特异表达的 HMWG 启动子、来自玉米的 USP 启动子、菜豆蛋白启动子、玉米醇溶蛋白基因启动子 (Pedersen 等, Cell 29(1982), 1015-1026 ;Quatroccio 等, Plant Mol. Biol. 15(1990),81-93), 谷蛋白启动子 (Leisy 等, Plant Mol. Biol. 14(1990),41-50 ;Zheng 等, Plant J. 4(1993),357-366 ;Yoshihara 等, FEBS Lett. 383(1996),213-218) 或 shrunken-1 启动子 (Werr 等, EMBO J. 4(1985), 1373-1380)。然而,也可使用仅在由外部因素决定的时间点上具有活性的启动子 (例如见 WO 9307279)。此处特别感兴趣的是容许简单诱导的热激蛋白启动子。此外可以使用种子特异性启动子,例如来自蚕豆 (*Vicia faba*) 的 USP 启动子,其确保在蚕豆和其他植物中的种子特异性表达 (Fiedler 等, Plant Mol. Biol. 22(1993),669-679 ;**Bäumlein**等, Mol. Gen. Genet. 225(1991),459-467)。

[0161] 使用藻类感染病毒基因组中存在的启动子也适合于在植物中表达核酸序列 (Mitra 等, 1994, *Biochem. Biophys. Res. Commun.* 204(1), 187-194; Mitra 和 Higgins, 1994, *Plant Mol. Biol.* 26(1), 85-93, Van Etten 等, 2002, *Arch. Virol.* 147, 1479-1516)。

[0162] 在本发明的上下文中, 术语“组织特异性的”理解为将现象 (例如转录起始) 基本限定在特定组织中的涵义。

[0163] 在本发明的上下文中, 术语“块茎、果实或种子细胞”理解为存在于块茎、果实或种子中的所有细胞。

[0164] 在本发明的上下文中, 术语“同源启动子”理解为天然存在于用于产生根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物的植物或植物细胞中 (与植物细胞或植物同源) 的启动子或在从中分离序列的生物中调节基因表达 (与待表达的核酸分子同源) 的启动子。

[0165] 在本发明的上下文中, 术语“异源启动子”理解为不天然存在于用于产生根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物的植物细胞或植物 (与植物细胞或植物异源) 的启动子或在其中分离待表达的核酸序列的生物中不天然存在的用于调节所述核酸序列的表达 (与待表达的核酸分子异源) 的启动子。

[0166] 还存在用于向核酸分子的 mRNA 转录物上添加多聚 A 尾的终止序列 (多聚腺苷酸信号)。认为多聚 A 尾用于稳定转录物。在文献中描述了此类元件 (参考 Gielen 等, *EMBO J.* 8(1989), 23-29) 并且在需要时进行交换。

[0167] 还可以在启动子和编码区或编码蛋白质的外来核酸分子间存在内含子序列。此类内含子序列可在植物中促使表达稳定并增加表达 (Callis 等, 1987, *Genes Devel.* 1, 1183-1200; Luehrsen 和 Walbot, 1991, *Mol. Gen. Genet.* 225, 81-93; Rethmeier 等, 1997; *Plant Journal* 12(4), 895-899; Rose 和 Beliakoff, 2000, *Plant Physiol.* 122(2), 535-542; Vasil 等, 1989, *Plant Physiol.* 91, 1575-1579; XU 等, 2003, *Science in China Series C Vol. 46 No. 6*, 561-569)。例如, 合适的内含子序列是来自玉米的 sh1 基因的第一个内含子、来自玉米的多泛素基因 1 的第一个内含子、来自稻的 EPSPS 基因的第一个内含子或来自拟南芥的 PAT1 基因的前两个内含子。

[0168] 本发明还涉及包含根据本发明的经遗传修饰的植物细胞的植物。通过从根据本发明的经遗传修饰的植物细胞的再生来产生此类植物。

[0169] 在其他实施方案中, 本发明涉及根据本发明的经遗传修饰的植物的可收获植物部分, 如果实、贮藏根和其他根、花、芽、枝条、叶或茎, 优选种子、果实或块茎, 这些可收获部分包含根据本发明的经遗传修饰的植物细胞。

[0170] 在本发明的优选实施方案中, 根据本发明的可收获植物部分是包含本发明经遗传修饰的植物细胞的根据本发明的经遗传修饰的植物的可加工或可消耗部分。

[0171] 在本发明的上下文中, 术语“可加工部分”应理解为用于制备粮食或饲料的植物部分, 其用作工业过程的原材料来源, 制备药剂的原材料来源或制备化妆品产品的原材料来源。

[0172] 在本发明的上下文中, 术语“可消耗部分”应理解为用作人类食物或动物饲料的植物部分。

[0173] 本发明还涉及包含本发明经遗传修饰的植物细胞的根据本发明的经遗传修饰的

植物的繁殖材料。

[0174] 此处，术语“繁殖材料”包含适合经无性途径或有性途径产生后代的植物的那些成分。适合无性繁殖的是例如，切割、愈伤组织培养、根茎或块茎，但还有例如原生质体和细胞培养物。通过有性过程方式产生的繁殖材料包括，例如果实、种子、幼苗等。所述繁殖材料优选采取块茎、果实或种子的形式。

[0175] 优选地，本发明涉及包含葡糖胺聚糖（例如乙酰透明质酸）的根据本发明的繁殖材料或根据本发明的植物的可收获部分。尤其优选地，根据本发明的繁殖材料或根据本发明的植物的可收获部分是合成葡糖胺聚糖的根据本发明的繁殖材料或根据本发明的植物的可收获部分。优选地，所述葡糖胺聚糖是软骨素、肝素 / 乙酰肝素、乙酰透明质酸，尤其优选为乙酰透明质酸。

[0176] 在本发明的上下文中，术语“马铃薯植物”或“马铃薯”应理解为茄属 (*Solanum*) 的植物种，尤其是茄属的块茎产生种，并尤其是马铃薯 (*Solanum tuberosum*)。

[0177] 在本发明的上下文中，术语“番茄植物”或“番茄”应理解为表示番茄属 (*Lycopersicon*) 的植物种，尤其是番茄 (*Lycopersicon esculentum*)。

[0178] 本发明的其他优势是根据本发明的经遗传修饰的植物的可收获部分、繁殖材料、可加工部分或可消耗部分比仅包含编码葡糖胺合酶的外来核酸分子的植物包含更多的葡糖胺聚糖（例如乙酰透明质酸）。因此，根据本发明的经遗传修饰的植物不仅特别适合作为可从其中分离葡糖胺聚糖（例如乙酰透明质酸）的原材料，也可直接用作粮食 / 饲料，或用于制备具有预防或治疗特征（例如用于骨关节炎预防，US 6,607,745）的粮食 / 饲料。由于根据本发明的经遗传修饰的植物比仅具有编码葡糖胺聚糖合酶的外来核酸分子的植物具有更高的葡糖胺聚糖含量，此类粮食 / 饲料的制备需要更少量的根据本发明的经遗传修饰的植物的可收获部分、繁殖材料、可加工部分或可消耗部分。如果根据本发明的经遗传修饰的植物的可消耗部分例如直接作为所谓的“营养制品”被消耗，则甚至可通过摄取相对少量的物质实现积极效果。这尤其在制备动物饲料中具有特殊重要性，因为具有太高植物成分含量的动物饲料不适合用作多种动物种类的饲料。

[0179] 由于葡糖胺聚糖，尤其是乙酰透明质酸结合水的高能力，根据本发明的经遗传修饰的植物的可收获部分、繁殖材料、可加工部分或可消耗部分当产生固体化粮食 / 饲料时还具有需要更少浓缩剂的优势。因此，例如，果冻的制备需要更少的糖，其与对健康的额外积极影响相关。在需要天然植物材料脱水的粮食 / 饲料的产生中，使用根据本发明的经遗传修饰的植物的可收获部分、繁殖材料、可加工部分或可消耗部分的优势在于这样的事实，需要从正在讨论的植物材料中去除更少的水，导致更低的生产成本，并因更温和的生产方法（例如更少的和 / 或更短的热量输入）致使正在讨论的粮食 / 饲料的营养价值提高。因此，例如，在番茄浆的生产中，需要引入更少的能量以达到期望的粘度。

[0180] 本发明还提供用于产生植物的方法，其包括

[0181] a) 遗传修饰的植物细胞，其中所述遗传修饰包括以下步骤 i 到 iii

[0182] i) 向植物细胞中引入编码葡糖胺聚糖合酶的外来核酸分子

[0183] ii) 向植物细胞中引入编码葡糖胺 6- 磷酸乙酰转移酶的外来核酸分子

[0184] iii) 向植物细胞中引入编码 UDP-N- 乙酰基 - 葡糖胺焦磷酸化酶的外来核酸分子

[0185] 其中可以任何顺序进行步骤 i 到 iii，或同时进行单独的步骤 i 到 iii 或其的任何

组合,

[0186] b) 从来自步骤 a)i 和 / 或 a)ii 和 / 或 a)iii 的植物细胞再生植物;

[0187] c) 适当时,使用根据步骤 b) 的植物再生另外的植物,其中,适当时,从步骤 b) 或 c) 的植物中分离植物细胞,并重复方法步骤 a) 到 c) 直至产生具有编码葡糖胺聚糖合酶的外来核酸分子和编码具有葡糖胺 6- 磷酸乙酰转移酶活性的蛋白质的外来核酸分子以及编码具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的外来核酸分子的植物。

[0188] 用于产生植物的根据本发明方法的优选实施方案涉及产生植物的方法,其包括

[0189] a) 遗传修饰植物细胞,其中所述遗传修饰包括单独或同时进行的以下任何顺序的步骤 i 到 iii,或以下步骤 i 到 iii 的任何组合

[0190] i) 向植物细胞中引入编码葡糖胺聚糖合酶的外来核酸分子

[0191] ii) 向植物细胞中引入编码葡糖胺 6- 磷酸乙酰转移酶的外来核酸分子

[0192] iii) 向植物细胞中引入编码 UDP-GlcNAc 焦磷酸化酶的外来核酸分子

[0193] b) 从包含根据以下步骤的遗传修饰的植物细胞中再生植物

[0194] i)a)i

[0195] ii)a)ii

[0196] iii)a)iii

[0197] iv)a)i 和 a)ii,

[0198] v)a)i 和 a)iii,

[0199] vi)a)ii 和 a)iii,或

[0200] vii)a)i 和 a)ii 和 a)iii

[0201] c) 根据以下步骤向植物的植物细胞中引入遗传修饰并再生植物

[0202] i) 根据步骤 a)ii 的 b)i 遗传修饰,

[0203] ii) 根据步骤 a)iii 的 b)i 遗传修饰,

[0204] iii) 根据步骤 a)ii 的 b)i 遗传修饰,和同时根据步骤 a)iii 的遗传修饰,

[0205] iv) 根据步骤 a)i 的 b)ii 遗传修饰,

[0206] v) 根据步骤 a)iii 的 b)ii 遗传修饰,

[0207] vi) 根据步骤 a)i 的 b)ii 遗传修饰,和同时根据步骤 a)iii 的遗传修饰,

[0208] vii) 根据步骤 a)i 的 b)iii 遗传修饰,

[0209] viii) 根据步骤 a)ii 的 b)iii 遗传修饰,

[0210] ix) 根据步骤 a)i 的 b)iii 遗传修饰,和同时根据步骤 a)ii 的遗传修饰,

[0211] x) 根据步骤 a)iii 的 b)iv 遗传修饰,

[0212] xi) 根据步骤 a)ii 的 b)v 遗传修饰,或

[0213] xii) 根据步骤 a)i 的 b)vi 遗传修饰

[0214] d) 根据以下步骤向植物的植物细胞中引入遗传修饰并再生植物

[0215] i) 根据步骤 a)iii 的 c)i 遗传修饰,

[0216] ii) 根据步骤 a)ii 的 c)ii 遗传修饰,

[0217] iii) 根据步骤 a)iii 的 c)iv 遗传修饰,

[0218] iv) 根据步骤 a)ii 的 c)v 遗传修饰,

[0219] v) 根据步骤 a)ii 的 c)vii 遗传修饰,

[0220] vi) 根据步骤 a)i 的 c)vii 遗传修饰,或

[0221] vii) 根据步骤 a)ii 的 c)ix 遗传修饰

[0222] e) 适当时在根据任何步骤 b)vii c)iii、c)vi、c)x、c)xi、c)xii 或根据任何步骤 d)i 到 d)vii 的植物的帮助下进一步产生植物。

[0223] 根据本发明用于产生植物的方法步骤 a), 为了引入外来核酸分子可使用任何可得的方法。用于转化植物细胞的多种方法早已在上文得到描述, 并在此处可以相应方式应用。如果不是同时进行而是连续进行用于产生植物的根据本发明方法的步骤 a) 的方法步骤, 可使用相同或不同方法用于各转化步骤。

[0224] 可通过本领域技术人员已知的方法实现用于产生植物的根据本发明方法的步骤 b) 和适当时步骤 c) 和 d) 的植物再生 (例如描述于“Plant Cell Culture Protocols“, 1999, R. D. Hall 编辑, Humana Press, ISBN0-89603-549-2)。

[0225] 例如可通过无性繁殖 (例如经过切割、块茎或经过愈伤组织培养和完整植株的再生) 或通过有性繁殖进行用于产生植物的根据本发明方法的进一步植物的产生 (根据方法, 根据步骤 c) 或步骤 e))。此处, 有性繁殖优选以可控方式进行, 即具有某些性质的所选植物彼此杂交并繁殖。优选进行选择, 以便所述进一步的植物 (根据方法, 其根据步骤 c) 或步骤 e) 产生) 具有在所述步骤中引入的外来核酸分子。

[0226] 在用于产生植物的根据本发明的方法中, 可同时或以连续步骤并以任何组合进行产生根据本发明经遗传修饰的植物细胞的遗传修饰。可以使用其中还未引入外来核酸分子的野生型植物和野生型植物细胞, 或可以使用早经过遗传修饰并早已向其中引入一个或更多外来核酸分子的植物细胞或植物。

[0227] 在用于产生植物的根据本发明方法的步骤 a) 中, 在向植物细胞或植物中引入外来核酸分子的遗传修饰中, 所述外来核酸分子可以是单个核酸分子或多个核酸分子。因此, 编码葡糖胺聚糖合酶或编码具有 GlcNAc-6-P 乙酰转移酶的酶活性的蛋白质或编码具有 UDP-GlcNAc 焦磷酸化酶的酶活性的蛋白质的外来核酸分子可一起存在于单个核酸分子上, 或者提及的外来核酸分子中的两个可以任何可能组合一起存在于单个核酸分子上并且第三个外来核酸分子存在于另一核酸分子上, 或者所有三个提及的外来核酸分子可存在于各单独的核酸分子上。

[0228] 早已在上文中结合根据本发明的植物细胞或根据本发明的植物描述了外来核酸分子或重组核酸分子的优选性质, 并且在用于产生植物的根据本发明方法的实践中相应地应用它们。

[0229] 在其他优选实施方案中, 用于产生植物的根据本发明的方法用于产生根据本发明的经遗传修饰的植物。

[0230] 本发明还提供通过本发明方法获得的植物, 所述方法用于产生合成乙酰透明质酸的植物。

[0231] 本发明还涉及用于产生葡糖胺聚糖 (例如乙酰透明质酸) 的方法, 其包括从根据本发明的经遗传修饰的植物细胞、根据本发明的经遗传修饰的植物、根据本发明的繁殖材料、根据本发明的可收获植物部分或通过用于产生植物的根据本发明方法获得的植物或这些植物的部分提取葡糖胺聚糖的步骤。优选地, 此方法也包括在提取葡糖胺聚糖 (例如乙酰透明质酸) 之前收获根据本发明的经遗传修饰的植物细胞、根据本发明的经遗传修饰的

植物、根据本发明的繁殖材料、根据本发明的可收获植物部分、根据本发明的可加工植物部分的步骤,并尤其优选地还包括在收获之前培养根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物的步骤。

[0232] 与细菌或动物组织相反,植物组织没有葡糖胺聚糖降解酶(例如透明质酸酶)。因此,如上文早已描述,使用相对简单的方法从植物组织中提取葡糖胺聚糖是可能的。如果需要,可使用本领域技术人员已知的方法,如利用乙醇重复沉淀来进一步纯化含有葡糖胺聚糖的植物细胞或组织的上述水溶性提取物。在一般方法第3项中描述了用于纯化乙酰透明质酸的优选方法。

[0233] 早已描述用于从根据本发明的经遗传修饰的植物细胞或根据本发明的经遗传修饰的植物中提取葡糖胺聚糖的方法也适合于从根据本发明的繁殖材料、根据本发明的可收获植物部分或通过用于制备合成乙酰透明质酸的植物的本发明方法获得的植物或这些植物的部分中分离葡糖胺聚糖(例如乙酰透明质酸)。

[0234] 本发明还提供根据本发明的经遗传修饰的植物细胞、根据本发明的经遗传修饰的植物、根据本发明的繁殖材料、根据本发明的可收获植物部分或通过用于产生制备葡糖胺聚糖的植物的本发明方法可获得的植物的用途。

[0235] 本发明还涉及包含根据本发明的经遗传修饰的植物细胞的组合物。此处,因为植物细胞已经例如通过加工而被破坏,所以无论它们是完整的还是不再完整的,细胞是无形的。所述组合物优选为粮食或饲料,药物产品或化妆品产品。

[0236] 本发明优选提供包含根据本发明的经遗传修饰的植物细胞、根据本发明的经遗传修饰的植物、根据本发明的繁殖材料、根据本发明的可收获植物部分或可通过本发明方法获得的植物的组分,并包含重组核酸分子的组合物,其中所述重组核酸分子的特征在于它们包含编码葡糖胺聚糖合酶和具有 GlcNAc-6-P 乙酰转移酶活性的蛋白质以及具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的核酸分子。

[0237] 外来核酸分子稳定整合到植物细胞或植物的基因组中导致所述外来核酸分子在整合到植物细胞或植物基因组后其侧翼为植物基因组核酸序列。

[0238] 因此,在优选的实施方案中,根据本发明的组合物的特征在于根据本发明组合物中存在的重组核酸分子侧翼为植物基因组核酸序列。

[0239] 此处,所述植物基因组核酸序列可以是在用于制备所述组合物的植物细胞或植物的基因组中天然存在的任何序列。

[0240] 在根据本发明的组合物中存在的重组核酸分子可以是单个核酸分子或不同的重组核酸分子,其中编码葡糖胺聚糖合酶(例如乙酰透明质酸合酶)和具有 GlcN-6-P 乙酰转移酶活性的蛋白质以及具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的核酸分子存在于一个核酸分子中,或者是那些其中在单独核酸分子中存在所述核酸分子。编码葡糖胺聚糖合酶(例如乙酰透明质酸合酶)或编码具有 GlcN-6-P 乙酰转移酶活性的蛋白质或编码具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的核酸分子可以一起存在于单个重组核酸分子上,或所提及的核酸分子中任何可能组合的两个核酸分子可一起存在于单个重组核酸分子上并且第三个核酸分子可以存在于另一重组核酸分子上,或者所有三个所提及的核酸分子在每一种情况下存在于各单独的重组核酸分子上。根据编码葡糖胺聚糖合酶(例如乙酰透明质酸合酶)或编码具有 GlcN-6-P 乙酰转移酶活性的蛋白质或编码具有 UDP-GlcNAc 焦磷酸化

酶活性的蛋白质的核酸分子在根据本发明组合物中如何存在,它们侧翼可以是相同的或不同的植物基因组核酸序列。

[0241] 可使用本领域技术人员已知的方法证明根据本发明的所述组合物包含的重组核酸分子,所述方法如,基于杂交的方法,或优选地使用基于 PCR(聚合酶链式反应)的方法。

[0242] 如上文早已提及,可以使用根据本发明的经遗传修饰的植物细胞、根据本发明的经遗传修饰的植物、根据本发明的繁殖材料、根据本发明的可收获植物部分,或通过根据本发明方法可获得的植物来制备粮食或饲料。然而,在不必提取葡糖胺聚糖(例如乙酰透明质酸)的情况下也可以作为原材料用于工业应用。因此,例如,根据本发明经遗传修饰的植物或根据本发明经遗传修饰的植物的部分可应用于农业培养领域中以实现土壤增加的水结合。此外,根据本发明经遗传修饰的植物或根据本发明经遗传修饰的植物细胞可用于制备干燥剂(例如当航运湿敏性项目使用)或作为液体吸收剂(例如在尿布中或用于吸附泄露的水性液体)。对于此类应用,必要时可以使用根据本发明的完整的经遗传修饰的植物、根据本发明的经遗传修饰的植物的部分或粉碎的(例如研碎的)根据本发明的经遗传修饰的植物或根据本发明的植物部分。适合于其中使用研碎植物或植物细胞的应用尤其是含有葡糖胺聚糖(例如乙酰透明质酸)的植物部分,但仅含有低比例的水。这些优选为谷类植物(玉米、稻、小麦、黑麦、燕麦、大麦、西米或高粱)的谷粒。因为根据本发明经遗传修饰的植物细胞或根据本发明经遗传修饰的植物比仅具有一种编码葡糖胺聚糖合酶的外来核酸分子的植物具有更高的葡糖胺聚糖(例如乙酰透明质酸)含量,所以当使用根据本发明经遗传修饰的植物细胞或根据本发明经遗传修饰的植物时,与这些相比需要使用较少的材料用于工业领域。

[0243] 本发明还提供用于制备根据本发明组合物的方法,其中使用根据本发明的经遗传修饰的植物细胞、根据本发明的经遗传修饰的植物、根据本发明的繁殖材料、根据本发明的可收获植物部分,或通过用于制备植物的根据本发明方法可获得的植物。用于制备根据本发明组合物的方法优选为用于制备粮食或饲料的方法,用于制备药物产品的方法或用于制备化妆品产品的方法。

[0244] 用于制备粮食或饲料的方法为本领域技术人员所知。用于在工业领域使用根据本发明经遗传修饰的植物或根据本发明的植物部分的方法也为本领域技术人员所知,并尤其包括粉碎或碾碎根据本发明经遗传修饰的植物或根据本发明的植物部分;然而,它们不仅限于此。在上文中早已经描述了源于使用根据本发明的物质的一些优势,所述根据本发明的物质用于制备粮食/饲料或用于工业领域中。

[0245] 用于制备组合物的本发明方法尤其优选为用于制备包含葡糖胺聚糖(例如乙酰透明质酸)的组合物的方法。

[0246] 本发明也提供可通过用于制备根据本发明组合物的方法获得的组合物。

[0247] 本发明还涉及根据本发明的经遗传修饰的植物细胞、根据本发明的经遗传修饰的植物、根据本发明的繁殖材料、根据本发明的可收获植物部分或可通过本发明产生植物的方法获得的植物用于制备根据本发明的组合物的用途。优选根据本发明的经遗传修饰的植物细胞、根据本发明的经遗传修饰的植物、根据本发明的繁殖材料、根据本发明的可收获植物部分或可通过本发明产生植物的方法获得的植物用于制备粮食或饲料,用于制备药物产品或化妆品产品的用途。

[0248] 序列描述

[0249] SEQ ID NO 1:编码来自小球藻病毒 1 的乙酰透明质酸合酶的核酸序列。

[0250] SEQ ID NO 2:来自小球藻病毒 1 的乙酰透明质酸合酶的氨基酸序列。所示氨基酸序列可源自 SEQ ID NO 1。

[0251] SEQ ID NO 3:编码来自小球藻病毒 1 的乙酰透明质酸合酶的合成核酸序列。以适合于植物细胞中的密码子使用的方式来合成所示的序列密码子。所示的核酸序列编码具有显示于 SEQ ID NO 2 中的氨基酸序列的蛋白质。

[0252] SEQ ID NO 4:编码具有来自多杀巴斯德氏菌的软骨素合酶活性的蛋白质的核酸序列。

[0253] SEQ ID NO 5:具有来自多杀巴斯德氏菌的软骨素合酶活性的蛋白质的氨基酸序列。所示氨基酸序列可源自 SEQ ID NO 4。

[0254] SEQ ID NO 6:编码具有来自多杀巴斯德氏菌的 heparosan 合酶活性的蛋白质的核酸序列。

[0255] SEQ ID NO 7:具有来自多杀巴斯德氏菌的 heparosan 合酶活性的蛋白质的氨基酸序列。所示氨基酸序列可源自 SEQ ID NO 6。

[0256] SEQ ID NO 8:编码具有来自酿酒酵母的 GlcN-6-P 乙酰转移酶活性的蛋白质的核酸序列。

[0257] SEQ ID NO 9:具有来自酿酒酵母的 GlcN-6-P 乙酰转移酶活性的蛋白质的氨基酸序列。所示氨基酸序列可源自 SEQ ID NO 8。

[0258] SEQ ID NO 10:编码具有来自酿酒酵母的 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的核酸序列。

[0259] SEQ ID NO 11:具有来自酿酒酵母的 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的氨基酸序列。所示氨基酸序列可源自 SEQ ID NO 10。

[0260] SEQ ID NO 12:包含来自 YLCV 启动子、限制性位点、自农杆菌的 ocs 终止子的多腺苷酸化信号序列和来自农杆菌的 nos 终止子的多腺苷酸化信号序列的表达盒的核酸序列。

[0261] SEQ ID NO 13:用于制备 MCS(“多克隆位点”)的合成寡核苷酸。

[0262] SEQ ID NO 14:用于制备 MCS(“多克隆位点”)的合成寡核苷酸。

[0263] SEQ ID NO 15:用于 PCR 反应的引物。

[0264] SEQ ID NO 16:用于 PCR 反应的引物。

[0265] SEQ ID NO 17:用于 PCR 反应的引物。

[0266] SEQ ID NO 18:用于 PCR 反应的引物。

[0267] SEQ ID NO 19:用于 PCR 反应的引物。

[0268] SEQ ID NO 20:用于 PCR 反应的引物。

[0269] SEQ ID NO 21:编码具有来自大肠杆菌的 GlcN-1-P 变位酶活性的蛋白质的核酸序列。

[0270] SEQ ID NO 22:具有来自大肠杆菌的 GlcN-1-P 变位酶活性的蛋白质的氨基酸序列。所示氨基酸序列可源自 SEQ ID NO 21。

[0271] SEQ ID NO 23:用作 PCR 引物的合成寡核苷酸。

[0272] SEQ ID NO 24:用作 PCR 引物的合成寡核苷酸。

[0273] SEQ ID NO 25:编码具有来自大肠杆菌的葡糖胺 1-磷酸乙酰转移酶和 UDP-GlcNAc 焦磷酸化酶双功能活性的蛋白质的核酸序列 (glmu)。

[0274] SEQ ID NO 26:具有来自大肠杆菌的葡糖胺 1-磷酸乙酰转移酶和 UDP-GlcNAc 焦磷酸化酶双功能活性的蛋白质的氨基酸序列。所示氨基酸序列可源自 SEQ ID NO 25。

[0275] SEQ ID NO 27:用于 PCR 反应的引物。

[0276] SEQ ID NO 28:用于 PCR 反应的引物。

[0277] SEQ ID NO 29:编码具有来自酿酒酵母的乙酰葡糖胺磷酸 (GlcN-P) 变位酶活性的蛋白质的核酸序列。

[0278] SEQ ID NO 30:来自酿酒酵母的具有乙酰葡糖胺磷酸变位酶活性的蛋白质的氨基酸序列。所示氨基酸序列可源自 SEQ ID NO 30。

[0279] 一般方法

[0280] 下文描述了可使用的与本发明相关的方法。这些方法是特定的实施方案;然而,本发明不限于这些方法。本领域技术人员已知的是可以通过修饰所描述的方法和/或由备选方法或备选方法的部分替代个别方法或方法的部分的相同方式进行本发明。

[0281] 1. 转化马铃薯植株

[0282] 如在 Rocha-Sosa 等 (EMBO J. 8, (1989), 23-29) 描述的在农杆菌的帮助下转化马铃薯植株。

[0283] 2. 使用乙酰透明质酸作为示例从植物组织中分离葡糖胺聚糖

[0284] 为了检测乙酰透明质酸的存在并测定植物组织中乙酰透明质酸的含量,如下操作植物材料:向约 0.3g 植物材料中加入 200 μ l 水(去矿物质的,导电率 $\geq 18\text{M}\Omega$),并在实验室用丁字球磨机(MM200,来自 Retsch, Germany)(以 30Hz 进行 30 秒)中粉碎混合物。随后再加入 800 μ l 水(去矿物质的,导电率 $\geq 18\text{M}\Omega$),并将混合物充分混合(例如使用 Vortex 混合器)。通过在 16000 $\times g$ 离心 5 分钟从上清液中分离细胞碎片和不溶性组分。为了测定植物整个地上部分中的乙酰透明质酸的量,从培养基质上约 1cm 至 3cm 处切除植株的地上部分,切成小片并随后如一般方法第 3 项中描述的用 Warring 搅拌器粉碎。为了测定乙酰透明质酸含量,可随后从所获的离心上清液中取等分试样(见一般方法第 3 项)。

[0285] 3. 使用乙酰透明质酸作为示例纯化葡糖胺聚糖

[0286] 加入水之后(在每种 100g 植物材料的情况下,约 100ml 水,去矿物质,导电率 $\geq 18\text{M}\Omega$),将粉碎的植物材料或植株的整个地上部分在 Warring 搅拌器中以最高速粉碎约 30 秒。随后使用茶叶滤网去除细胞碎片。将已去除的细胞碎片在 300ml 水(去矿物质,导电率 $\geq 18\text{M}\Omega$)中重悬浮并再次使用茶叶滤网去除。将所获的两份悬浮液(100ml+300ml)合并并在 13000 $\times g$ 下离心 15 分钟。向所获的离心上清液中加入 NaCl 直至已达到 1%的终浓度。当 NaCl 已溶入溶液中后,通过加入两倍体积的乙醇并随后充分混合并在 -20 $^{\circ}\text{C}$ 下孵育过夜进行沉淀。随后将混合物在 13000 $\times g$ 下离心 15 分钟。经该离心后将所获的沉淀物溶于 100ml 缓冲液中(50mM TrisHCl, pH 8, 1mM CaCl_2)并随后加入蛋白酶 K 至 100 $\mu\text{g}/\text{ml}$ 的终浓度并在 42 $^{\circ}\text{C}$ 下孵育该溶液 2 小时。随后在 95 $^{\circ}\text{C}$ 下孵育 10 分钟。再一次向该溶液加入 NaCl 直至已达到 1%的终浓度。当 NaCl 已溶于溶液后,通过加入两倍体积的乙醇、充分混合并在 -20 $^{\circ}\text{C}$ 下孵育约 96 小时进行沉淀。随后在 13000 $\times g$ 下离心 15 分钟。经该离心后将所获的沉淀物溶于 30ml 水中(去矿物质,导电率 $\geq 18\text{M}\Omega$),并再一次加入 NaCl 至终浓度

为 1%。通过加入两倍体积的乙醇、充分混合并在 -20℃ 孵育过夜进行另一次沉淀。将在随后以 13000×g 离心 15 分钟所获的沉淀物溶于 20ml 水（去矿物质，导电率 $\geq 18\text{M}\Omega$ ）中。

[0287] 通过离心过滤进行进一步纯化。为此，在每种情况下将 5ml 溶解的沉淀物加至膜滤器上（CentriconAmicon, 孔径 10 000NMWL, 产品号 UCF8 010 96），并以 2200×g 离心样品直至滤器上仅剩余约 3ml 溶液。随后再进行两次，每次将 3ml 水（去矿物质，导电率 $\geq 18\text{M}\Omega$ ）加入膜上的溶液中并每次在相同条件下再离心直至最终滤器上仅剩余约 3ml 溶液。将离心过滤后仍存在于膜上的溶液去除，并用约 1.5ml 水（去矿物质，导电率 $\geq 18\text{M}\Omega$ ）重复冲洗膜（三至五次）。将仍存在于膜上的所有溶液及从冲洗所获的溶液合并，加入 NaCl 至终浓度为 1%，当 NaCl 溶于溶液后，加入两倍体积的乙醇，将样品混合并在 -20℃ 下过夜保存以获得沉淀。将随后在 13000×g 离心 15 分钟后所获的沉淀溶于 4ml 水（去矿物质，导电率 $\geq 18\text{M}\Omega$ ）中并随后冻干（在 0.37mbar 的压力下在来自 Christ, Osterode, Germany 的冻干仪 Christ Alpha 1-4 上进行 24 小时）。

[0288] 4. 乙酰透明质酸的检测和乙酰透明质酸含量的测定

[0289] 按照生产商说明书使用商品化检测（来自 Corgenix, Inc., Colorado, USA, 产品号 029-001 的透明质酸（HA）检测试剂盒）检测乙酰透明质酸，因此通过将所述说明书合并于此作为参考。检测原理基于特异性结合乙酰透明质酸（HABP）的蛋白质的存在并与酶联免疫吸附测定相似地操作，其中颜色反应指示了所检测样品中乙酰透明质酸的含量。在使用检测试剂盒中包含的确定量的乙酰透明质酸的校准曲线的帮助下来测定乙酰透明质酸的值。因此，对于乙酰透明质酸的定量测定，待测定的样品应该以处于所指明的极限之内的浓度使用（例如：正在讨论的样品的稀释或用更少的水从植物组织中提取乙酰透明质酸均基于是否超过或未达到该极限）。

[0290] 在平行批次中，首先对待测定样品的等分试样进行透明质酸酶消化并随后使用商业化检测（来自 Corgenix, Inc., Colorado, USA, 产品号 029-001 的透明质酸（HA）检测试剂盒）检测。使用 400 μl 的在透明质酸酶缓冲液（0.1M 磷酸氢二钾缓冲液，pH 5.3；150mM NaCl）中的植物提取物，通过加入了 5 μg （ ~ 3 单位）的透明质酸酶（来自 Sigma, 产品号 H2251 的 III 型透明质酸酶）并在 37℃ 温育 30 分钟来进行透明质酸酶消化。

[0291] 每种情况下在 1 : 10 的稀释中，随后所有样品均用于测定乙酰透明质酸含量。

[0292] 5. GFAT 活性的测定

[0293] 如在 Rachel 等 (1996, J. Bacteriol. 178(8), 2320-2327) 中所描述来测定具有 GFAT 活性的蛋白质的活性。

[0294] 为了区分蛋白质是否具有 GFAT-1 或 GFAT-2 活性，使用在 Hu 等 (2004, J. Biol. Chem. 279(29), 29988-29993) 中描述的方法。

实施例

[0295] 1. 制备植物表达质粒 IR 47-71

[0296] 质粒 pBinAR 是双元载体质粒 pBin19 (Bevan, 1984, Nucl Acids Res 12 : 8711-8721) 的衍生物，其如下构建：

[0297] 从质粒 pDH51 (Pietrzak 等, 1986 Nucleic Acids Res. 14, 5858) 中分离包含花椰菜花叶病毒 35S 启动子 6909-7437 位核酸的长度为 529bp 的片段作为 EcoR I/Kpn I 片段，并

且连接至 pUC18 多接头的 EcoR I 和 Kpn I 限制位点之间。以该方式形成了质粒 pUC18-35S。使用限制性内切核酸酶 Hind III 和 Pvu II, 从质粒 pAGV40 (Herrera-Estrella 等, 1983 Nature, 303, 209-213) 分离包括 Ti 质粒 pTiACH5 (Gielen 等, 1984, EMBO Journal 3, 835-846) 的 T-DNA 的章鱼碱合酶基因 (基因 3) 的多聚腺苷酸信号 (3' 末端) (11749-11939 位核酸) 的长度为 192bp 的片段。将 Sph I 接头加入到 Pvu II 限制位点后, 将所述片段连接至 pUC18-35S 的 Sph I 和 Hind III 限制位点中间。这形成了质粒 pA7。此处, 使用 EcoR I 和 Hind III 取出包含 35S 启动子和 Ocs 终止子的整个多接头并将其连接至适当切割的载体 pBin19 中。这产生了植物表达载体 pBinAR (Höfgen 和 Willmitzer, 1990, Plant Science 66, 221-230)。

[0298] 将来自马铃薯 (*Solanum tuberosum*) 的 patatin 基因 B33 的启动子 (Rocha-Sosa 等, 1989, EMBO J. 8, 23-29) 作为 Dra I 片段 (-1512-+14 位核酸) 连接至 Sst I 切割的载体 pUC19 中, 已使用 T4-DNA 聚合酶将该载体末端平端化。这产生了质粒 pUC19-B33。使用 EcoR I 和 Sma I 从该质粒中取出 B33 启动子并连接至适合的限制性载体 pBinAR 中。这产生了植物表达载体 pBinB33。

[0299] 为了便于进一步的克隆步骤, 扩展了 MCS (多克隆位点)。为此, 合成了两条互补的寡核酸, 在 95°C 加热 5 分钟, 缓慢冷却至室温以获得良好的退火并将其克隆至 pBinB33 的 Sal I 和 KpnI 限制位点中。用于此目的的寡核酸显示于 SEQ ID NO 13 和 SEQ ID NO 14 中。所获质粒命名为 IR47-71。

[0300] 2. 制备植物表达载体 pBinARHyg

[0301] 使用限制性内切核酸酶 EcoR I 和 Hind III 从 pA7 中取出包含 35S 启动子、Ocs 终止子和整个多克隆位点的片段并克隆至载体 pBIBHyg (Becker, 1990, Nucleic Acids Res. 18, 203) 中, 已使用相同的限制性内切核酸酶切割该载体。所获质粒命名为 pBinARHyg。

[0302] 3. 制备克隆载体 IC 317-204

[0303] 使用限制性内切核酸酶 Xho I 和 Hind III, 从质粒 IR 47-71 中分离包含 ocs 终止子的核酸片段并将其克隆至载体 pBlueScript KS (来自 Stratagene, 产品号 212207) 中, 已使用相同的限制性内切核酸酶切割该载体。所获质粒命名为 IC 306-204。

[0304] 使用限制性内切核酸酶 Bam HI 和 Eco RI 从质粒 IR 47-71 中分离包含 B33 启动子的核酸片段并将其克隆至载体 pBlueScript KS (来自 Stratagene, 产品号 212207) 中, 已使用相同的限制性内切核酸酶切割该载体。所获质粒命名为 IC 314-204。

[0305] 使用限制性内切核酸酶 Bam HI 从 IC 306-204 中分离 OCS 终止子并将其克隆至质粒 IC 314-204 中, 已使用相同的限制性内切核酸酶切割该质粒。所获质粒命名为 IC 317-204。

[0306] 4. 合成核酸分子

[0307] a) 合成编码来自小球藻病毒 1 的乙酰透明质酸合酶的核酸分子

[0308] 通过 Medigenomix GmbH (Munich, 德国) 合成编码小球藻病毒 1 的乙酰透明质酸合酶的核酸序列并克隆至来自 Invitrogen 的载体 pCR2.1 (产品号 K2000-01) 中。所获质粒命名为 IC 323 215。编码来自小球藻病毒 1 的 HAS 蛋白质的合成核酸序列, 其显示于 SEQ ID NO 3 中。最初分离自小球藻病毒 1 的对应核酸序列显示于 SEQ ID NO 1 中。

[0309] b) 合成包含 YLCV 启动子和 MCS、nos 终止子和 ocs 终止子的核酸序列

[0310] 通过 Entelechon GmbH 合成包含 YLCV 启动子 (Stavolone 等, PlantMolecular Biology 53:703-713, 2003) 和包含限制位点 Sac I 和 Sma I 的 MCS (“多克隆位点”)、nos 终止子和 ocs 终止子的核酸序列并克隆至来自 Invitrogen 的载体 pCR4Topo (产品号 K4510-20) 中。所获质粒命名为 IC389-337。合成核酸序列显示于 SEQ ID NO 12 中。

[0311] 5. 分离核酸分子

[0312] a) 分离并克隆具有 GlcN-6-P 乙酰转移酶活性的蛋白质的编码核酸序列

[0313] 通过 PCR 从酿酒酵母中分离编码具有 GlcN-6-P 乙酰转移酶活性的蛋白质的核酸序列 (gna1) 并克隆至来自 Invitrogen 的载体 pCR 2.1 (产品号 K4510-20) 中。PCR 反应条件如下:

[0314] 1. 步骤:5 分钟 94℃,

[0315] 2. 步骤:45 秒,94℃,

[0316] 3. 步骤:45 秒,59℃,

[0317] 4. 步骤:45 秒,72℃,

[0318] 5. 步骤:10 分钟,72℃

[0319] 6. 步骤:4℃

[0320] 将步骤 2 至步骤 4 重复 35 次,并该操作随后继续进行步骤 5。

[0321] 50 μl 反应批量 (reaction batch) 包含缓冲液 (10mM Tris-HCl, pH 9.0、50mM KCl 和 3mM MgSO₄)、每种情况下 500nM 的扩增引物 (显示于 SEQ ID NO 15 和 SEQ ID NO 16 中)、10 μl Q 溶液 (包含于 Qiagen, 产品号 206143 中)、每种情况下 0.2mM 脱氧核糖核酸、0.5 μl Taq DNA 聚合酶 (Invitrogen, 产品号:11304-011) 和作为模板的 250ng 酵母基因组 DNA。使用来自 Eppendorf 的 Mastercycler 进行 PCR (Prod. NR. 5331 000.010)。

[0322] 从酿酒酵母中分离的编码具有 GlcN-6-P 乙酰转移酶活性的蛋白质的核酸序列显示于 SEQ ID NO 8 中。

[0323] 在将所获片段克隆至载体 pCR 2.1 中并测序确认后,使用限制性内切核酸酶 Kpn I 和 Xba I 分离正在讨论的来自酿酒酵母的编码具有 GlcN-6-P 乙酰转移酶活性的蛋白质的核酸序列并克隆至载体 pA7 中,已使用相同的限制性内切核酸酶切割该载体。所获质粒命名为 IC 298-204。

[0324] b) 分离并克隆编码具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的核酸序列

[0325] 通过 PCR 从酿酒酵母中分离编码具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的核酸序列 (gri) 并克隆至来自 Invitrogen 的载体 pCR 2.1 (产品号 K2000-01) 中。PCR 反应条件如下:

[0326] 1. 步骤:5 分钟 94℃,

[0327] 2. 步骤:45 秒,94℃,

[0328] 3. 步骤:45 秒,59℃,

[0329] 4. 步骤:45 秒,72℃,

[0330] 5. 步骤:30 分钟,72℃

[0331] 6. 步骤:4℃

[0332] 将步骤 2 至步骤 4 重复 35 次,并该操作随后继续进行步骤 5。

[0333] 50 μ l 反应批量包含缓冲液 (10mM Tris-HCl, pH 9.0、50mM KCl 和 3mM MgSO₄)、每种情况下 500nM 的扩增引物 (显示于 SEQ ID NO 17 和 SEQ ID NO 18 中)、10 μ l Q 溶液 (包含于 Qiagen, 产品号 :206143 中)、每种情况下 0.2mM 脱氧核糖核酸、0.5 μ l Taq DNA 聚合酶 (Invitrogen, 产品号 :11304-011) 和作为模板的 250ng 酵母基因组 DNA (Invitrogen 产品号 40802)。使用来自 Eppendorf 的 Mastercycler (产品号 5331 000.010) 进行 PCR。

[0334] 从酿酒酵母中分离的编码具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的核酸序列显示于 SEQ ID NO 10 中。

[0335] 在将所获片段克隆至载体 pCR 2.1 中并测序确认后, 使用限制性内切核酸酶 Kpn I 和 Xba I 分离来自酿酒酵母的所述编码具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的核酸序列并克隆至载体 pA7 中, 已使用相同的限制性内切核酸酶切割该载体。所获质粒命名为 IC 303-204。

[0336] c) 分离并克隆编码具有 GlcNAc-P 变位酶活性的蛋白质的核酸序列

[0337] 通过 PCR 从酿酒酵母中分离编码具有乙酰葡萄糖胺磷酸变位酶活性的蛋白质的核酸序列 (pcm I, EC 5.4.2.3) 并克隆至来自 Invitrogen 的载体 pCR 2.1 (产品号 K2000-01) 中。PCR 反应条件如下:

[0338] 1. 步骤 :5 分钟 94°C,

[0339] 2. 步骤 :45 秒, 94°C,

[0340] 3. 步骤 :45 秒, 59°C,

[0341] 4. 步骤 :45 秒, 72°C,

[0342] 5. 步骤 :30 分钟, 72°C

[0343] 6. 步骤 :4°C

[0344] 将步骤 2 至步骤 4 重复 35 次, 并该操作随后继续进行步骤 5。

[0345] 50 μ l 反应批量包含缓冲液 (10mM Tris-HCl, pH 9.0、50mM KCl 和 3mM MgSO₄)、每种情况下 500nM 的扩增引物 (显示于 SEQ ID NO 27 和 SEQ ID NO 28 中)、10 μ l Q 溶液 (包含于 Qiagen, 产品号 :206143 中)、每种情况下 0.2mM 脱氧核糖核酸、0.5 μ l Taq DNA 聚合酶 (Invitrogen, 产品号 :11304-011) 和作为模板的 250ng 酵母基因组 DNA (Invitrogen 产品号 40802)。使用来自 Eppendorf 的 Mastercycler (产品号 5331 000.010) 进行 PCR。

[0346] 在将所获片段克隆至载体 pCR 2.1 中并测序确认后, 使用限制性内切核酸酶 Kpn I 和 Xba I 分离来自酿酒酵母的所述编码具有乙酰葡萄糖胺磷酸变位酶活性的蛋白质的核酸序列并克隆至载体 pA7 中, 已使用相同的限制性内切核酸酶切割该载体。所获质粒命名为 IC 304-204。

[0347] d) 从大肠杆菌 (*Escherichia coli*) 中分离并克隆编码具有 GlcN-1-P 变位酶活性的蛋白质的核酸序列

[0348] 通过 PCR 分离来自大肠杆菌的编码具有葡萄糖胺 1- 磷酸变位酶 (GlcN-1-P 变位酶) 活性的蛋白质的核酸序列 (glmm) 并克隆至来自 Invitrogen 的载体 pCR 2.1 (产品号 K2000-01) 中。PCR 反应条件如下:

[0349] 1. 步骤 :5 分钟 94°C,

[0350] 2. 步骤 :45 秒, 94°C,

[0351] 3. 步骤 :45 秒, 59°C,

[0352] 4. 步骤 :45 秒,72℃,

[0353] 5. 步骤 :30 分钟,72℃

[0354] 6. 步骤 :4℃

[0355] 将步骤 2 至步骤 4 重复 35 次,并该操作随后继续进行步骤 5。

[0356] 50 μ l 反应批量包含缓冲液 (10mM Tris-HCl,pH 9.0、50mM KCl 和 3mM MgSO₄)、每种情况下 500nM 的扩增引物 (显示于 SEQ ID NO 19 和 SEQ ID NO 20 中)、10 μ l Q 溶液 (包含于 Qiagen,产品号 :206143 中)、每种情况下 0.2mM 脱氧核糖核酸、0.5 μ l Taq DNA 聚合酶 (Invitrogen,产品号 :11304-011) 和作为模板的 250ng 大肠杆菌基因组 DNA。使用来自 Eppendorf 的 Mastercycler (产品号 5331 000.010) 进行 PCR。

[0357] 从大肠杆菌中分离的编码具有葡萄糖胺 1- 磷酸变位酶活性的蛋白质的核酸序列显示于 SEQ ID NO 21 中。

[0358] 在将所获片段克隆至载体 pCR 2.1 中并测序确认后,使用限制性内切核酸酶 Kpn I 和 Xba I 分离来自大肠杆菌的所述编码具有 GlcN-1-P 变位酶活性的蛋白质的核酸序列并克隆至载体 pA7 中,已使用相同的限制性内切核酸酶切割该载体。所获质粒命名为 IC 300-204。

[0359] e) 从大肠杆菌中分离并克隆编码具有 GlcN-1-P 乙酰转移酶和 UDP-GlcNAc-1-P 焦磷酸化酶双功能活性的蛋白质的核酸序列

[0360] 通过 PCR 从大肠杆菌中分离编码具有葡萄糖胺 1- 磷酸乙酰转移酶和 UDP-GlcNAc 焦磷酸化酶活性的双功能蛋白质的核酸序列 (qlmu) 并克隆至来自 Invitrogen 的载体 pCR 2.1 (产品号 K2000-01) 中。PCR 反应条件如下:

[0361] 1. 步骤 :5 分钟 94℃,

[0362] 2. 步骤 :45 秒,94℃,

[0363] 3. 步骤 :45 秒,59℃,

[0364] 4. 步骤 :45 秒,72℃,

[0365] 5. 步骤 :30 分钟,72℃

[0366] 6. 步骤 :4℃

[0367] 将步骤 2 至步骤 4 重复 35 次,并该操作随后继续进行步骤 5。

[0368] 50 μ l 反应批量包含缓冲液 (10mM Tris-HCl,pH 9.0、50mM KCl 和 3mM MgSO₄)、每种情况下 500nM 的扩增引物 (显示于 SEQ ID NO 23 和 SEQ ID NO 24 中)、10 μ l Q 溶液 (包含于 Qiagen,产品号 :206143 中)、每种情况下 0.2mM 脱氧核糖核酸、0.5 μ l Taq DNA 聚合酶 (Invitrogen,产品号 :11304-011) 和作为模板的 250ng 大肠杆菌基因组 DNA。使用来自 Eppendorf 的 Mastercycler (产品号 5331 000.010) 进行 PCR。

[0369] 从大肠杆菌中分离的编码具有葡萄糖胺 1- 磷酸乙酰转移酶和 UDP-GlcNAc 焦磷酸化酶双功能活性的蛋白质的核酸序列显示于 SEQ IDNO 25 中。

[0370] 在将所获片段克隆至载体 pCR 2.1 中并测序确认后,使用限制性内切核酸酶 Kpn I 和 Xba I 从大肠杆菌中分离所述编码具有 GlcN-1-P 乙酰转移酶和 UDP-GlcNAc 焦磷酸化酶活性的双功能蛋白质的核酸序列 (qlmu) 并克隆至载体 pA7 中,已使用相同的限制性内切核酸酶切割该载体。所获质粒命名为 IC 299-204。

[0371] 6. 制备包含来自小球藻病毒 1 的乙酰透明质酸合酶编码核酸序列的植物表达载

体

[0372] 通过用 BamH I 和 Xho I 限制酶切消化从质粒 IC 323-215 中分离包含乙酰透明质酸合酶编码序列的核酸分子并克隆至质粒 IR 47-71 的 BamH I 和 Xho I 限制位点中。所获植物表达载体命名为 IC 341-222。

[0373] 7. 制备包含来自酿酒酵母的具有 GlcN-6-P 乙酰转移酶活性的蛋白质的编码核酸序列的植物表达载体 IC 351-222

[0374] 起始质粒为上述植物表达载体 pUBI bar (WO 9744472), 其中将来自酵母的 gna 基因的编码序列克隆至所述植物表达载体 pUBI bar 的 EcoR I 和 Sda I 限制位点。通过 EcoR I 和 Sda I 限制酶切消化从质粒 IC 298-204 中分离来自酵母的 gna 基因的编码序列。所获载体命名为 IC 351-222。

[0375] 8. 制备包含具有 GlcN-6-P 乙酰转移酶活性的蛋白质和具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的编码核酸序列的植物表达载体 IC 392-337

[0376] 起始质粒为上文中进一步描述的质粒 IC 351-222, 其中将使用限制性内切核酸酶 Eco RI 从质粒 IC 389-337 中分离的 YLCV 启动子和 NOS 终止子和 OCS 终止子的盒克隆至其 Eco RI 限制位点中。所获载体命名为 IC390-337。

[0377] 通过 Sac I 和 Eco RV 限制酶切消化从上述质粒 IC 303-204 中分离 qri 基因的编码序列并连接至载体 IC 390-337 的 Sac I 和 Sma I 限制位点中。所获载体命名为 IC 391-337。

[0378] 为了去除多余的 OCS 终止子, 用 Aat II 消化载体 IC 391-337 并随后重连接。所获植物表达载体命名为 IC 392-337。

[0379] 9. 制备包含来自大肠杆菌的具有 GlcN-1-P 变位酶活性的蛋白质的编码核酸序列, 以及具有 GlcN-1-P 乙酰转移酶和 UDP-GlcNAc 焦磷酸化酶双功能活性的双功能蛋白质的植物表达载体 IC 360-237

[0380] 用于引入来自大肠杆菌的编码具有 GlcN-1-P 乙酰转移酶和 UDP-GlcNAc-1-P 焦磷酸化酶双功能活性的蛋白质的核酸序列的起始质粒为上文中进一步描述的质粒 IC 299-204, 经 Eco RI 限制酶切消化分离所述质粒的编码序列并克隆至 pMCS5 载体 (MoBiTec GmbH, Prod. No. :pMCS5) 的 Eco RI 限制位点中。所获载体命名为 IC 307-204。在下一步中, 用限制性内切核酸酶 Pme I 和 Sma I 消化载体 IC 307-204 并重连接。所获载体命名为 IC 311-204。随后通过用 Bam HI 和 KpnI 限制酶切消化从质粒 IC 311-204 中分离编码具有 GLMU 活性的蛋白质的核酸序列并连接至载体 IC 312-204 的 Bam HI 和 KpnI 限制位点中。所获载体命名为 IC315-204。通过同时连接包括经 Eco RI 和 Sal I 限制酶切消化从质粒 pA7 中分离的 35S 启动子片段、经 Hind III 和 Sal I 限制酶切消化从 IC 309-204 中分离的 ocs 片段和已通过 Eco RI 限制酶切消化切开的载体 IC310-204 的三条片段来制备载体 IC 312-204。质粒 IC310-204 为 pUC 18 载体, 其部分 MCS 已通过 Hind III 和 Ecl 135II 限制酶切消化和随后的重连接被去除。通过使用 Hind III 和 Sal I 从 pA7 中分离 ocs 片段并将其克隆至用 Hind III 和 Sal I 消化的 pBS KS 载体来制备 IC 309-204。

[0381] 通过 Eco RI 限制酶切消化从质粒 IC 315-204 中分离 35S 启动子、来自大肠杆菌的编码具有葡萄糖胺 1-磷酸乙酰转移酶和 UDP-GlcNAc 焦磷酸化酶双功能活性的蛋白质的核酸序列 (qlmu) 和 ocs 终止子, 并将其克隆至 Ubi Bar 载体 (WO 9744472) 的 Eco RI 限制

位点中。所获载体命名为 IC 359-237。

[0382] 用于引入编码具有 GlcN-1-P 变位酶活性的蛋白质的核酸序列的起始质粒为上文中进一步描述的质粒 IC 299-204, 经 Sda I 和 Sma I 限制酶切消化分离其编码序列并连接至 Ubi bar 载体的 Sda I 和 Hpa I 限制位点中。所获载体命名为 IC 355-222。

[0383] 通过 Spe I 和 Dra I 限制酶切消化从质粒 IC 355-222 中分离 glmm 基因的编码序列并将其克隆至 IC 359-237 质粒的 Spe I 和 Pme I 限制位点中。所获载体命名为 IC 360-237。

[0384] 10. 制备包含来自酿酒酵母的具有 GlcN-6-P 乙酰转移酶活性的蛋白质、具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质以及具有 GlcNAc-P 变位酶活性的蛋白质的编码核酸序列的植物表达载体 IC 393-337

[0385] 用于引入来自酿酒酵母的编码 GlcNAc-P 变位酶的核酸序列的起始质粒为上文中进一步描述的质粒 IC 304-204, 经 Eco RI 限制酶切消化分离其编码序列并克隆至 pMCS5 载体 (MoBiTec GmbH, Prod. No. :pMCS5) 的 Eco RI 限制位点中。所获载体命名为 IC 313-204。在下一步中, 经 Pme I 和 Pac I 限制酶切消化从载体 IC 313-204 中分离来自酿酒酵母的编码 GlcNAc-P 变位酶的核酸序列并克隆至已用 Pme I 和 Pac I 消化的载体 IC393-337 中。所获载体命名为 IC 394-337。

[0386] 用于制备质粒 IC 393-337 的起始载体为上文进一步描述的质粒 IC391-337, 其已包含具有 GlcN-6-P 乙酰转移酶活性的蛋白质和具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的核酸序列。为此, 经 Pac I 和 AvrII 限制酶切消化分离上文中进一步描述的 B33 启动子并克隆至已用 Pac I 和 Avr II 消化的载体 IC 391-337 中。所获植物表达载体命名为 IC 393-337。

[0387] 11. 马铃薯植株的转化

[0388] 根据一般方法第一项中描述的方法用植物表达载体 IC 341-222 转化马铃薯植株, 所述载体包含在来自马铃薯的 patatin 基因 B33 的启动子调控下的来自小球藻病毒 1 的乙酰透明质酸合酶的编码核酸序列 (Rocha-Sosa 等, 1989, EMBO J. 8, 23-29)。

[0389] 转化有质粒 IC 341-222 的所获马铃薯植株命名为 365ES X, 其中 X 表示从转化中独立获得的植株。

[0390] 将转化后所获的栽培品种命名为 365ES X, 并分析由正在讨论的植株合成的乙酰透明质酸的量 (也见于 WO 2006032538)。选择栽培品种 365ES13 和 365ES 74 用于下文中描述的转化。

[0391] 根据一般方法第一项中描述的方法用植物表达载体 IC 392-337 或 IC360-237 或 IC 394-337 转化栽培品种 365ES 13 和 365ES 74 的马铃薯植株。

[0392] 将用质粒 IC 392-337 转化的栽培品种 365ES 13 所获的转基因马铃薯植株命名为 437ES X, 其中 X 表示从转化中独立获得的植株。

[0393] 将用质粒 IC 392-337 转化的栽培品种 365ES 74 所获的转基因马铃薯植株命名为 438ES X, 其中 X 表示从转化中独立获得的植株。

[0394] 将用质粒 IC 360-237 转化的栽培品种 365ES 13 所获的转基因马铃薯植株命名为 397ES X, 其中 X 表示从转化中独立获得的植株。

[0395] 将用质粒 IC 360-237 转化的栽培品种 365ES 74 所获的转基因马铃薯植株命名为

398ES X, 其中 X 表示从转化中独立获得的植株。

[0396] 将用质粒 IC 393-337 转化的栽培品种 365ES 13 所获的转基因马铃薯植株命名为 444ES X, 其中 X 表示从转化中独立获得的植株。

[0397] 将用质粒 IC 393-337 转化的栽培品种 365ES 74 所获的转基因马铃薯植株命名为 445ES X, 其中 X 表示从转化中独立获得的植株。

[0398] 12. 分析包含编码乙酰透明质酸合酶和编码具有 GlcNAc-6-P 乙酰转移酶活性的蛋白质和编码具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的外来核酸分子的马铃薯植株

[0399] 在温室中, 将栽培品种 365ES 13、365ES 74、437ES X、438ES X、397ES X、398ES X、444ES X 和 445ES X 的单独植株种植于 6cm 盆中的土壤里。根据一般方法第 3 项中描述的方法从 7 至 9 周龄植株中收获植株的整个地上部分并处理。使用一般方法第 4 项中描述的方法并在标准曲线的帮助下通过测定植物提取物的等分试样中所包含的乙酰透明质酸来测定正在讨论的植物提取物中乙酰透明质酸的量。对于测定乙酰透明质酸的量, 按 1 : 10 稀释来使用离心后所获的上清。对于选择后的植株, 获得以下结果 :

[0400]

植株名	HA [μ g/g fw]
365ES 74-67	120.9
365ES 74-68	125.1
365ES 74-71	129.2
365ES 74-72	118.2
365ES 74-79	129.0
365ES 74-80	140.2
365ES 74-81	92.7
365ES 74-82	100.9
wtDesiree 1	0.2
wtDesiree 2	-0.2
wtDesiree 3	2.1

[0401]

植株名	HA [μ g/g fw]
365ES 13-221	90.6
365ES 13-222	57.9

365ES 13-223	59.0
365ES 13-224	95.3
365ES 13-225	84.0
365ES 13-226	91.7
365ES 13-227	69.1
365ES 13-228	76.2
265ES 13-231	84.3
365ES 13-232	75.5

[0402] 表 1 :正在讨论的植物由仅包含编码乙酰透明质酸合酶的外来核酸分子的栽培品种 365ES 13 和 365ES 74 的独立选择的转基因植株在整个地上部分所产生的乙酰透明质酸的量 (以每克鲜重中的乙酰透明质酸的 μg 量计)。第一列显示了从中收集材料的植株的名称 (“wt Desiree”指栽培品种 Désirée 未转化的野生型植株)。第 2 列显示了基于所用的鲜重的乙酰透明质酸的量。

[0403]

植株名	HA [$\mu\text{g/g fw}$]
438ES 1	183.0
438ES 5	192.5
438ES 10	189.5
438ES 11	168.9
438ES 13	195.5
438ES 14	184.3
438ES 16	168.9
438ES 23	231.2

[0404]

植株名	HA [$\mu\text{g/g fw}$]
437ES 2	105.6

437ES 3	97.6
437ES 6	129.4
437ES 12	103.2
437ES 13	144.2
437ES 14	163.7
437ES 15	128.2
437ES 16	100.3

[0405]

植株名	HA[$\mu\text{g/g fw}$]
438ES 25	169.2
438ES 32	178.6
438ES 33	173.4
438ES 41	178.7
438ES 57	239.4
438ES 62	172.5
438ES 64	199.5
438ES 80	235.4
438ES 84	189.2
438ES 85	168.1
438ES 88	167.4
438ES 97	164.4
438ES 102	160.4
438ES 108	209.8
438ES 112	185.9

[0406]

植株名	HA [μ g/g fw]
437ES 17	186.3
437ES 21	100.7
437ES 23	114.5
437ES 26	105.1
437ES 31	102.0
437ES 34	178.9
437ES 35	104.4
437ES 39	98.2
437ES 40	116.8
437ES 48	125.1
437ES 66	146.8
437ES 69	106.8
437ES 70	115.4
437ES 75	100.3
437ES 76	125.5
437ES 79	102.0
437ES 80	125.7
437ES 82	135.7
437ES 95	100.3
437ES 105	108.8

[0407] 表 2 :正在讨论的植物由栽培品种 437ES 或 438ES 的独立选择的转基因植株在整

个地上部分所产生的乙酰透明质酸的量（以每克鲜重中的乙酰透明质酸的 μg 量计）。第一列显示了从中收集材料的植株的名称。第 2 列显示了基于所用的鲜重的乙酰透明质酸的量。

[0408] 显示的结果说明同时含有编码乙酰透明质酸合酶和编码具有 GlcN-6-P 乙酰转移酶活性的蛋白质和编码具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的外来核酸分子的植株比仅包含编码乙酰透明质酸合酶的外来核酸分子的植株合成明显更大量的乙酰透明质酸。

[0409]

植株名	HA [$\mu\text{g/g fw}$]
397ES 1	4.51
397ES 2	22.54
397ES 3	16.27
397ES 5	10.13
397ES 6	13.60
397ES 7	33.62
397ES 8	19.87
397ES 9	70.37

[0410]

植株名	HA [$\mu\text{g/g fw}$]
398ES 1	53.04
398ES 2	43.64
398ES 3	130.01
398ES 4	89.26
398ES 5	74.35
398ES 6	55.39
398ES 7	99.61
398ES 8	90.82

[0411]

植株名	HA[μ g/g fw]
397ES 10	51.91
397ES 11	78.71
397ES 12	9.76
397ES 13	6.62
397ES 14	32.74
397ES 16	21.34
397ES 17	81.90
397ES 18	33.16
397ES 19	32.18
397ES 20	26.67
397ES 21	63.21
397ES 22	2.66
397ES 24	1.41
397ES 25	37.27
397ES 26	20.34
397ES 27	32.89
397ES 28	9.89
397ES 29	8.31
397ES 30	85.77
397ES 31	47.44
397ES 32	53.47
397ES 33	5.25
397ES 34	17.10

397ES 35	16.80
397ES 36	17.53
397ES 37	25.90
397ES 38	7.68
397ES 39	0.49
397ES 40	0.85
397ES 41	14.65
397ES 42	35.36
397ES 43	49.96
397ES 44	28.78
397ES 45	18.95
397ES 46	7.93
397ES 47	28.28
397ES 48	13.94
397ES 49	60.16
397ES 50	29.77

[0412]

植株名	HA[μ g/g fw]
398ES 9	41.80
398ES 10	79.66
398ES 11	9.57
398ES 12	41.24
398ES 13	89.05

398ES 14	77.19
398ES 15	96.96
398ES 16	84.24
398ES 17	124.63
398ES 18	76.19
398ES 19	71.45
398ES 20	46.17
398ES 21	80.96
398ES 22	31.25
398ES 23	87.35
398ES 24	62.31
398ES 25	58.98
398ES 26	71.14
398ES 27	18.48
398ES 28	105.85
398ES 29	12.24
398ES 30	95.82
398ES 31	18.43
398ES 32	109.41
398ES 33	87.10
398ES 34	88.30
398ES 35	97.85
398ES 36	58.51
398ES 37	78.51

398ES 38	99.67
398ES 39	11.28
398ES 40	29.12
398ES 41	74.15
398ES 42	22.75
398ES 43	77.05
398ES 44	101.34
398ES 45	57.16
398ES 46	54.19
398ES 47	64.35
398ES 48	60.04
398ES 49	77.36
398ES 50	50.84

[0413] 表 3 :正在讨论的植物由栽培品种 497ES 或 498ES 的独立选择的转基因植株在整个地上部分所产生的乙酰透明质酸的量 (以每克鲜重中的乙酰透明质酸的 μg 量计)。第一列显示了从中收集材料的植株的名称。第 2 列显示了基于所用的鲜重的乙酰透明质酸的量。

[0414] 显示的结果说明同时含有编码乙酰透明质酸合酶和编码具有 GlcN-P 变位酶活性的蛋白质和具有 GlcN-1-P 乙酰转移酶和 UDP-GlcNAc 焦磷酸化酶双功能活性的蛋白质的外来核酸分子的植株比仅包含编码乙酰透明质酸合酶的外来核酸分子的植株不合成明显更大量的乙酰透明质酸。

[0415]

植株名	HA [$\mu\text{g/g fw}$]		植株名	HA [$\mu\text{g/g fw}$]
444ES 6	81.06		445ES 1	102.53
444ES 12	95.89		445ES 6	99.65
444ES 14	90.76		445ES 7	152.66
444ES 17	83.20		445ES 8	82.24

444ES 19	69.18		445ES 9	119.74
444ES 23	59.45		445ES 12	82.68
444ES 27	58.93		445ES 16	102.24
444ES 35	58.45		445ES 18	86.47
444ES 37	67.34		445ES 19	103.37
444ES 41	63.27		445ES 20	96.06
444ES 43	60.82		445ES 23	116.64
			445ES 24	95.13
			445ES 34	87.24
			445ES 40	81.74
			445ES 42	98.72
			445ES 47	84.41
			445ES 60	86.71
			445ES 135	94.13

[0416] 表 4 :正在讨论的植物由栽培品种 444ES 或 445ES 的独立选择的转基因植株在整个地上部分所产生的乙酰透明质酸的量 (以每克鲜重中的乙酰透明质酸的 μg 量计)。第一列显示了从中收集材料的植株的名称。第 2 列显示了基于所用的鲜重的乙酰透明质酸的量。

[0417] 显示的结果说明同时含有编码乙酰透明质酸合酶和编码具有 GlcN-P 变位酶活性的蛋白质和具有 GlcN-6-P 乙酰转移酶的蛋白质和具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的外来核酸分子的植株比仅包含编码乙酰透明质酸合酶和具有 GlcN-6-P 乙酰转移酶活性的蛋白质和具有 UDP-GlcNAc 焦磷酸化酶活性的蛋白质的外来核酸分子的植株不合成明显更大量的乙酰透明质酸。

序列表

<110> 拜耳作物科学股份公司 (Bayer CropScience AG)

〈120〉 合成增加量的葡萄糖胺聚糖的植物

<130>BCS 07-5007 PCT

<150>EP07116174.9

<151>2007-09-12

<150>US60/933, 575

<151>2007-09-13

 $\langle 160 \rangle_{30}$

<170>PatentIn 版本 3.3

 $\langle 210 \rangle 1$

<211>1707

$\langle 212 \rangle$ DNA

〈213〉 小球藻病毒 1 (Paramecium bursaria Chlorella Virus 1)

 $\langle 220 \rangle$ $\langle 221 \rangle$ CDS $\langle 222 \rangle (1) \dots (1707)$ $\langle 300 \rangle$

<308>GenBank/U4258 ?

<309>1995-12-24

$\langle 313 \rangle$ (50903) .. (52609)

 $\langle 400 \rangle 1$

atg ggt aaa aat ata atc ata atg gtt tgc tgg tac acc atc ata act 48

Met Gly Lys Asn Ile Ile Ile Met Val Ser Trp Tyr Thr Ile Ile Thr

1 5 10 15

tca aat cta atc gcg gtt gga gga gcc tct cta atc ttg gct ccg gca 96

Ser Asn Leu Ile Ala Val Gly Gly Ala Ser Leu Ile Leu Ala Pro Ala

20 25 30

att act ggg tat gtt cta cat tgg aat att gct ctc tcg aca atc tgg 144

Ile Thr Gly Tyr Val Leu His Trp Asn Ile Ala Leu Ser Thr Ile Trp

35 40 45

gga gta tca gct tat ggt att ttc gtt ttt ggg ttt ttc ctt gca caa 192

Gly Val Ser Ala Tyr Gly Ile Phe Val Phe Gly Phe Phe Leu Ala Gln

50 55 60

gtt tta ttt tca gaa ctg aac agg aaa cgt ctt cgc aag tgg att tct 240

Val Leu Phe Ser Glu Leu Asn Arg Lys Arg Leu Arg Lys Trp Ile Ser

65 70 75 80

ctc aga cct aag ggt tgg aat gat gtt cgt ttg gct gtg atc att gct 288

Leu	Arg	Pro	Lys	Gly	Trp	Asn	Asp	Val	Arg	Leu	Ala	Val	Ile	Ile	Ala	
				85				90				95				
gga	tat	cgc	gag	gat	cct	tat	atg	ttc	cag	aag	tgc	ctc	gag	tct	gta	336
Gly	Tyr	Arg	Glu	Asp	Pro	Tyr	Met	Phe	Gln	Lys	Cys	Leu	Glu	Ser	Val	
				100				105				110				
cgt	gac	tct	gat	tat	ggc	aac	gtt	gcc	cgt	ctg	att	tgt	gtg	att	gac	384
Arg	Asp	Ser	Asp	Tyr	Gly	Asn	Val	Ala	Arg	Leu	Ile	Cys	Val	Ile	Asp	
				115				120				125				
ggt	gat	gag	gac	gat	gat	atg	agg	atg	gct	gcc	gtt	tac	aag	gcg	atc	432
Gly	Asp	Glu	Asp	Asp	Asp	Met	Arg	Met	Ala	Ala	Val	Tyr	Lys	Ala	Ile	
				130				135				140				
tac	aat	gat	aat	atc	aag	aag	ccc	gag	ttt	gtt	ctg	tgt	gag	tca	gac	480
Tyr	Asn	Asp	Asn	Ile	Lys	Lys	Pro	Glu	Phe	Val	Leu	Cys	Glu	Ser	Asp	
145				150				155				160				
gac	aag	gaa	ggt	gaa	cgc	atc	gac	tct	gat	ttc	tct	cgc	gac	att	tgt	528
Asp	Lys	Glu	Gly	Glu	Arg	Ile	Asp	Ser	Asp	Phe	Ser	Arg	Asp	Ile	Cys	
				165				170				175				
gtc	ctc	cag	cct	cat	cgt	gga	aaa	cgg	gag	tgt	ctt	tat	act	ggg	ttt	576
Val	Leu	Gln	Pro	His	Arg	Gly	Lys	Arg	Glu	Cys	Leu	Tyr	Thr	Gly	Phe	
				180				185				190				
caa	ctt	gca	aag	atg	gac	ccc	agt	gtc	aat	gct	gtc	gtt	ctg	att	gac	624
Gln	Leu	Ala	Lys	Met	Asp	Pro	Ser	Val	Asn	Ala	Val	Val	Leu	Ile	Asp	
				195				200				205				
agc	gat	acc	gtt	ctc	gag	aag	gat	gct	att	ctg	gaa	gtt	gta	tac	cca	672
Ser	Asp	Thr	Val	Leu	Glu	Lys	Asp	Ala	Ile	Leu	Glu	Val	Val	Tyr	Pro	
				210				215				220				
ctt	gca	tgc	gat	ccc	gag	atc	caa	gcc	gtt	gca	ggt	gag	tgt	aag	att	720
Leu	Ala	Cys	Asp	Pro	Glu	Ile	Gln	Ala	Val	Ala	Gly	Glu	Cys	Lys	Ile	
225				230				235				240				
tgg	aac	aca	gac	act	ctt	ttg	agt	ctt	ctc	gtc	gct	tgg	cgg	tac	tat	768
Trp	Asn	Thr	Asp	Thr	Leu	Leu	Ser	Leu	Leu	Val	Ala	Trp	Arg	Tyr	Tyr	
				245				250				255				
tct	gcg	ttt	tgt	gtg	gag	agg	agt	gcc	cag	tct	ttt	ttc	agg	act	gtt	816
Ser	Ala	Phe	Cys	Val	Glu	Arg	Ser	Ala	Gln	Ser	Phe	Phe	Arg	Thr	Val	
				260				265				270				
cag	tgc	gtt	ggg	ggg	cca	ctg	ggt	gcc	tac	aag	att	gat	atc	att	aag	864
Gln	Cys	Val	Gly	Gly	Pro	Leu	Gly	Ala	Tyr	Lys	Ile	Asp	Ile	Ile	Lys	
				275				280				285				
gag	att	aag	gac	ccc	tgg	att	tcc	cag	cgc	ttt	ctt	ggt	cag	aag	tgt	912

Glu	Ile	Lys	Asp	Pro	Trp	Ile	Ser	Gln	Arg	Phe	Leu	Gly	Gln	Lys	Cys	
290				295				300								
act	tac	ggt	gac	gac	cgc	cgg	cta	acc	aac	gag	atc	ttg	atg	cgt	ggt	960
Thr	Tyr	Gly	Asp	Asp	Arg	Arg	Leu	Thr	Asn	Glu	Ile	Leu	Met	Arg	Gly	
305				310				315				320				
aaa	aag	gtt	gtg	ttc	act	cca	ttt	gct	gtt	ggt	tgg	tct	gac	agt	ccg	1008
Lys	Lys	Val	Val	Phe	Thr	Pro	Phe	Ala	Val	Gly	Trp	Ser	Asp	Ser	Pro	
325				330				335								
acc	aat	gtg	ttt	cgg	tac	atc	gtt	cag	cag	acc	cgc	tgg	agt	aag	tcg	1056
Thr	Asn	Val	Phe	Arg	Tyr	Ile	Val	Gln	Gln	Thr	Arg	Trp	Ser	Lys	Ser	
340				345				350								
tgg	tgc	cgc	gaa	att	tgg	tac	acc	ctc	ttc	gcc	gcg	tgg	aag	cac	ggt	1104
Trp	Cys	Arg	Glu	Ile	Trp	Tyr	Thr	Leu	Phe	Ala	Ala	Trp	Lys	His	Gly	
355				360				365								
ttg	tct	gga	att	tgg	ctg	gcc	ttt	gaa	tgt	ttg	tat	caa	att	aca	tac	1152
Leu	Ser	Gly	Ile	Trp	Leu	Ala	Phe	Glu	Cys	Leu	Tyr	Gln	Ile	Thr	Tyr	
370				375				380								
ttc	ttc	ctc	gtg	att	tac	ctc	ttt	tct	cgc	cta	gcc	gtt	gag	gcc	gac	1200
Phe	Phe	Leu	Val	Ile	Tyr	Leu	Phe	Ser	Arg	Leu	Ala	Val	Glu	Ala	Asp	
385				390				395				400				
cct	cgc	gcc	cag	aca	gcc	acg	gtg	att	gtg	agc	acc	acg	gtt	gca	ttg	1248
Pro	Arg	Ala	Gln	Thr	Ala	Thr	Val	Ile	Val	Ser	Thr	Thr	Val	Ala	Leu	
405				410				415								
att	aag	tgt	ggg	tat	ttt	tca	ttc	cga	gcc	aag	gat	att	cgg	gcg	ttt	1296
Ile	Lys	Cys	Gly	Tyr	Phe	Ser	Phe	Arg	Ala	Lys	Asp	Ile	Arg	Ala	Phe	
420				425				430								
tac	ttt	gtg	ctt	tat	aca	ttt	gtt	tac	ttt	ttc	tgt	atg	att	ccg	gcc	1344
Tyr	Phe	Val	Leu	Tyr	Thr	Phe	Val	Tyr	Phe	Phe	Cys	Met	Ile	Pro	Ala	
435				440				445								
agg	att	act	gca	atg	atg	acg	ctt	tgg	gac	att	ggc	tgg	ggt	act	cgc	1392
Arg	Ile	Thr	Ala	Met	Met	Thr	Leu	Trp	Asp	Ile	Gly	Trp	Gly	Thr	Arg	
450				455				460								
ggt	gga	aac	gag	aag	cct	tcc	gtt	ggc	acc	cgg	gtc	gct	ctg	tgg	gca	1440
Gly	Gly	Asn	Glu	Lys	Pro	Ser	Val	Gly	Thr	Arg	Val	Ala	Leu	Trp	Ala	
465				470				475				480				
aag	caa	tat	ctc	att	gca	tat	atg	tgg	tgg	gcc	gcg	gtt	gtt	ggc	gct	1488
Lys	Gln	Tyr	Leu	Ile	Ala	Tyr	Met	Trp	Trp	Ala	Ala	Val	Val	Gly	Ala	
485				490				495								
gga	gtt	tac	agc	atc	gtc	cat	aac	tgg	atg	ttc	gat	tgg	aat	tct	ctt	1536

Gly Val Tyr Ser Ile Val His Asn Trp Met Phe Asp Trp Asn Ser Leu
500 505 510
tct tat cgt ttt gct ttg gtt ggt att tgt tct tac att gtt ttt att 1584
Ser Tyr Arg Phe Ala Leu Val Gly Ile Cys Ser Tyr Ile Val Phe Ile
515 520 525
gtt att gtg ctg gtg gtt tat ttc acc ggc aaa att acg act tgg aat 1632
Val Ile Val Leu Val Val Tyr Phe Thr Gly Lys Ile Thr Thr Trp Asn
530 535 540
ttc acg aag ctt cag aag gag cta atc gag gat cgc gtt ctg tac gat 1680
Phe Thr Lys Leu Gln Lys Glu Leu Ile Glu Asp Arg Val Leu Tyr Asp
545 550 555 560
gca act acc aat gct cag tct gtg tga 1707
Ala Thr Thr Asn Ala Gln Ser Val
565

<210>2

<211>568

<212>PRT

<213> 小球藻病毒 1

<400>2

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

Asp Lys Glu Gly Glu Arg Ile Asp Ser Asp Phe Ser Arg Asp Ile Cys		
165	170	175
Val Leu Gln Pro His Arg Gly Lys Arg Glu Cys Leu Tyr Thr Gly Phe		
180	185	190
Gln Leu Ala Lys Met Asp Pro Ser Val Asn Ala Val Val Leu Ile Asp		
195	200	205
Ser Asp Thr Val Leu Glu Lys Asp Ala Ile Leu Glu Val Val Tyr Pro		
210	215	220
Leu Ala Cys Asp Pro Glu Ile Gln Ala Val Ala Gly Glu Cys Lys Ile		
225	230	235
Trp Asn Thr Asp Thr Leu Leu Ser Leu Leu Val Ala Trp Arg Tyr Tyr		
245	250	255
Ser Ala Phe Cys Val Glu Arg Ser Ala Gln Ser Phe Phe Arg Thr Val		
260	265	270
Gln Cys Val Gly Gly Pro Leu Gly Ala Tyr Lys Ile Asp Ile Ile Lys		
275	280	285
Glu Ile Lys Asp Pro Trp Ile Ser Gln Arg Phe Leu Gly Gln Lys Cys		
290	295	300
Thr Tyr Gly Asp Asp Arg Arg Leu Thr Asn Glu Ile Leu Met Arg Gly		
305	310	315
Lys Lys Val Val Phe Thr Pro Phe Ala Val Gly Trp Ser Asp Ser Pro		
325	330	335
Thr Asn Val Phe Arg Tyr Ile Val Gln Gln Thr Arg Trp Ser Lys Ser		
340	345	350
Trp Cys Arg Glu Ile Trp Tyr Thr Leu Phe Ala Ala Trp Lys His Gly		
355	360	365
Leu Ser Gly Ile Trp Leu Ala Phe Glu Cys Leu Tyr Gln Ile Thr Tyr		
370	375	380
Phe Phe Leu Val Ile Tyr Leu Phe Ser Arg Leu Ala Val Glu Ala Asp		
385	390	395
Pro Arg Ala Gln Thr Ala Thr Val Ile Val Ser Thr Thr Val Ala Leu		
405	410	415
Ile Lys Cys Gly Tyr Phe Ser Phe Arg Ala Lys Asp Ile Arg Ala Phe		
420	425	430
Tyr Phe Val Leu Tyr Thr Phe Val Tyr Phe Phe Cys Met Ile Pro Ala		
435	440	445
Arg Ile Thr Ala Met Met Thr Leu Trp Asp Ile Gly Trp Gly Thr Arg		
450	455	460
Gly Gly Asn Glu Lys Pro Ser Val Gly Thr Arg Val Ala Leu Trp Ala		

465	470	475	480
Lys Gln Tyr Leu Ile Ala Tyr Met Trp Trp Ala Ala Val Val Gly Ala			
	485	490	495
Gly Val Tyr Ser Ile Val His Asn Trp Met Phe Asp Trp Asn Ser Leu			
	500	505	510
Ser Tyr Arg Phe Ala Leu Val Gly Ile Cys Ser Tyr Ile Val Phe Ile			
	515	520	525
Val Ile Val Leu Val Val Tyr Phe Thr Gly Lys Ile Thr Thr Trp Asn			
	530	535	540
Phe Thr Lys Leu Gln Lys Glu Leu Ile Glu Asp Arg Val Leu Tyr Asp			
545	550	555	560
Ala Thr Thr Asn Ala Gln Ser Val			

565

<210>3

<211>1707

<212>DNA

<213> 人工的

<220>

<223> 编码小球藻病毒 1 乙酰透明质酸合酶蛋白质的合成核苷酸序列

<400>3

```

atgggtaaga acattatcat tatggtgtcc tggtagacaa ttattacaag taatctcatc 60
gcagttgggtg gtgcatctct tattctcgct ccagctatca ctggatatgt tcttcactgg 120
aacatcgccc tctcaactat ttggggagtt tccgcatatg gtatTTTTgt tttcgggttc 180
TTTTtggtc aggttctgtt ctgagagctc aatcgtaaga gactcaggaa gtggattagc 240
cttagaccaa aggggtggaa tgacgttcgt ctgctgtca ttatcgctgg ctaccgtgaa 300
gatccttaca tgtttcaaaa gtgcttggaa tcagttaggg atagtgatta tggcaacgtc 360
gctagactga tctgtgtgat tgatggagat gaggacgacg atatgaggat ggcagctgtt 420
tataaggcta tctataatga taacattaag aagcctgaat ttgttctttg cgagtctgat 480
gacaaggaag gagaacggat tgattcagat ttctcacgtg atatctgcgt tctccaacct 540
catcgtggga agcgtgaatg tctttataca ggtttccaac tcgccaaaat ggacccatca 600
gtgaacgctg tggttcttat cgatagtgat actgtgctgg agaaagatgc tatcttggag 660
gttgtttacc ctcttgctg tgatcctgaa attcaagctg tggctggaga gtgcaagatc 720
tggaacacag atactcttct ttctctgctt gtcgcatgga gatattactc cgcattctgt 780
gtggagagga gcgctcaatc cttttccgt accgttcaat gcgttgggtg tcctttggga 840
gcttacaaaa ttgatatcat caaggagatt aaggacccat ggattagtca aaggtttctt 900
ggtcagaagt gcacttatgg cgatgatcgt agattgacta acgaaatcct tatgaggggc 960
aagaaagtcg tttttactcc atttgctgtc ggatggtctg attcacctac aaatgttttc 1020
cgttatattg tgcaacaaac acgttggagt aagagctggg ttagggagat ctggtacact 1080
ttgttcgctg cttggaagca cgggcttagc ggaatttggc ttgcttttga atgcctttac 1140

```

```

cagattacat actttttctt ggtgatctat ttgttttcac gtcttgccgt cgaggctgac 1200
cctagagcac agactgcaac tgtgattggt tctactacag tcgcacttat taagtgtggc 1260
tatttcagtt ttagagcaaa agatattaga gccttctatt ttgttttgta cacatttggt 1320
tattttcttt gcatgattcc agctcgtatt accgctatga tgaccttggt ggacatcgga 1380
tggggaacta gaggtggtaa cgaaaagcct tctgtgggaa caagggtggc cctttgggca 1440
aaacaatate tcategccta catgtggtgg gccgctgtcg ttggtgccgg agtgtactca 1500
atcgttcata actggatggt tgactggaac tctttgagct atcgtttcgc tcttgtgggt 1560
atttgttctt acattgtttt categtgatt gtgctcgttg tgtatttcac tggtaaaatc 1620
acaacctgga atttcactaa acttcaaaag gaattgattg aagacagggt tctgtatgat 1680
gctactacca acgcccagtc agtttaa 1707

```

<210>4

<211>2895

<212>DNA

<213> 多杀巴斯德氏菌 (*Pasteurella multocida*)

<220>

<221>CDS

<222>(1).. (2895)

<300>

<308>EMBL/AF195517

<309>2000-09-07

<313>(61).. (2955)

<400>4

```

atg aat aca tta tca caa gca ata aaa gca tat aac agc aat gac tat 48
Met Asn Thr Leu Ser Gln Ala Ile Lys Ala Tyr Asn Ser Asn Asp Tyr
1          5          10          15
gaa tta gca ctc aaa tta ttt gag aag tct gct gaa acc tac ggg cga 96
Glu Leu Ala Leu Lys Leu Phe Glu Lys Ser Ala Glu Thr Tyr Gly Arg
20          25          30
aaa atc gtt gaa ttc caa att atc aaa tgt aaa gaa aaa ctc tcg acc 144
Lys Ile Val Glu Phe Gln Ile Ile Lys Cys Lys Glu Lys Leu Ser Thr
35          40          45
aat tct tat gta agt gaa gat aaa aaa aac agt gtt tgc gat agc tca 192
Asn Ser Tyr Val Ser Glu Asp Lys Lys Asn Ser Val Cys Asp Ser Ser
50          55          60
tta gat atc gca aca cag ctc tta ctt tcc aac gta aaa aaa tta act 240
Leu Asp Ile Ala Thr Gln Leu Leu Leu Ser Asn Val Lys Lys Leu Thr
65          70          75          80
cta tcc gaa tca gaa aaa aac agt tta aaa aat aaa tgg aaa tct atc 288
Leu Ser Glu Ser Glu Lys Asn Ser Leu Lys Asn Lys Trp Lys Ser Ile

```

85	90	95	
act ggg aaa aaa tcg gag aac gca gaa atc aga aag gtg gaa cta gta			336
Thr Gly Lys Lys Ser Glu Asn Ala Glu Ile Arg Lys Val Glu Leu Val			
100	105	110	
ccc aaa gat ttt cct aaa gat ctt gtt ctt gct cca ttg cca gat cat			384
Pro Lys Asp Phe Pro Lys Asp Leu Val Leu Ala Pro Leu Pro Asp His			
115	120	125	
gtt aat gat ttt aca tgg tac aaa aat cga aaa aaa agc tta ggt ata			432
Val Asn Asp Phe Thr Trp Tyr Lys Asn Arg Lys Lys Ser Leu Gly Ile			
130	135	140	
aag cct gta aat aag aat atc ggt ctt tct att att att cct aca ttt			480
Lys Pro Val Asn Lys Asn Ile Gly Leu Ser Ile Ile Ile Pro Thr Phe			
145	150	155	160
aat cgt agc cgt att tta gat ata acg tta gcc tgt ttg gtc aat cag			528
Asn Arg Ser Arg Ile Leu Asp Ile Thr Leu Ala Cys Leu Val Asn Gln			
165	170	175	
aaa aca aac tac cca ttt gaa gtc gtt gtt gca gat gat ggt agt aag			576
Lys Thr Asn Tyr Pro Phe Glu Val Val Val Ala Asp Asp Gly Ser Lys			
180	185	190	
gaa aac tta ctt acc att gtg caa aaa tac gaa caa aaa ctt gac ata			624
Glu Asn Leu Leu Thr Ile Val Gln Lys Tyr Glu Gln Lys Leu Asp Ile			
195	200	205	
aag tat gta aga caa aaa gat tat gga tat caa ttg tgt gca gtc aga			672
Lys Tyr Val Arg Gln Lys Asp Tyr Gly Tyr Gln Leu Cys Ala Val Arg			
210	215	220	
aac tta ggt tta cgt aca gca aag tat gat ttt gtc tcg att cta gac			720
Asn Leu Gly Leu Arg Thr Ala Lys Tyr Asp Phe Val Ser Ile Leu Asp			
225	230	235	240
tgc gat atg gca cca caa caa tta tgg gtt cat tct tat ctt aca gaa			768
Cys Asp Met Ala Pro Gln Gln Leu Trp Val His Ser Tyr Leu Thr Glu			
245	250	255	
cta tta gaa gac aat gat att gtt tta att gga cct aga aaa tat gtg			816
Leu Leu Glu Asp Asn Asp Ile Val Leu Ile Gly Pro Arg Lys Tyr Val			
260	265	270	
gat act cat aat att acc gca gaa caa ttc ctt aac gat cca tat tta			864
Asp Thr His Asn Ile Thr Ala Glu Gln Phe Leu Asn Asp Pro Tyr Leu			
275	280	285	
ata gaa tca cta cct gaa acc gct aca aat aac aat cct tcg att aca			912
Ile Glu Ser Leu Pro Glu Thr Ala Thr Asn Asn Asn Pro Ser Ile Thr			

290	295	300	
tca aaa gga aat ata tcg ttg gat tgg aga tta gaa cat ttc aaa aaa			960
Ser Lys Gly Asn Ile Ser Leu Asp Trp Arg Leu Glu His Phe Lys Lys			
305	310	315	320
acc gat aat cta cgt cta tgt gat tct ccg ttt cgt tat ttt agt tgc			1008
Thr Asp Asn Leu Arg Leu Cys Asp Ser Pro Phe Arg Tyr Phe Ser Cys			
325	330	335	
ggt aat gtt gca ttt tct aaa gaa tgg cta aat aaa gta ggt tgg ttc			1056
Gly Asn Val Ala Phe Ser Lys Glu Trp Leu Asn Lys Val Gly Trp Phe			
340	345	350	
gat gaa gaa ttt aat cat tgg ggg ggc gaa gat gta gaa ttt ggt tac			1104
Asp Glu Glu Phe Asn His Trp Gly Gly Glu Asp Val Glu Phe Gly Tyr			
355	360	365	
aga tta ttt gcc aaa ggc tgt ttt ttc aga gta att gac ggc gga atg			1152
Arg Leu Phe Ala Lys Gly Cys Phe Phe Arg Val Ile Asp Gly Gly Met			
370	375	380	
gca tac cat caa gaa cca cct ggt aaa gaa aat gaa aca gac cgc gaa			1200
Ala Tyr His Gln Glu Pro Pro Gly Lys Glu Asn Glu Thr Asp Arg Glu			
385	390	395	400
gct ggt aaa agt att acg ctt aaa att gtg aaa gaa aag gta cct tac			1248
Ala Gly Lys Ser Ile Thr Leu Lys Ile Val Lys Glu Lys Val Pro Tyr			
405	410	415	
atc tat aga aag ctt tta cca ata gaa gat tca cat att cat aga ata			1296
Ile Tyr Arg Lys Leu Leu Pro Ile Glu Asp Ser His Ile His Arg Ile			
420	425	430	
cct tta gtt tct att tat atc ccc gct tat aac tgt gca aat tat att			1344
Pro Leu Val Ser Ile Tyr Ile Pro Ala Tyr Asn Cys Ala Asn Tyr Ile			
435	440	445	
caa aga tgt gta gat agt gct ctt aat caa act gtt gtc gat ctc gag			1392
Gln Arg Cys Val Asp Ser Ala Leu Asn Gln Thr Val Val Asp Leu Glu			
450	455	460	
gtt tgt att tgt aac gat ggt tca aca gat aat acc tta gaa gtg atc			1440
Val Cys Ile Cys Asn Asp Gly Ser Thr Asp Asn Thr Leu Glu Val Ile			
465	470	475	480
aat aag ctt tat ggt aat aat cct agg gta cgc atc atg tct aaa cca			1488
Asn Lys Leu Tyr Gly Asn Asn Pro Arg Val Arg Ile Met Ser Lys Pro			
485	490	495	
aat ggc gga ata gcc tca gca tca aat gca gcc gtt tct ttt gct aaa			1536
Asn Gly Gly Ile Ala Ser Ala Ser Asn Ala Ala Val Ser Phe Ala Lys			

500	505	510	
ggt tat tac att ggg cag tta gat tca gat gat tat ctt gag cct gat			1584
Gly Tyr Tyr Ile Gly Gln Leu Asp Ser Asp Asp Tyr Leu Glu Pro Asp			
515	520	525	
gca gtt gaa ctg tgt tta aaa gaa ttt tta aaa gat aaa acg cta gct			1632
Ala Val Glu Leu Cys Leu Lys Glu Phe Leu Lys Asp Lys Thr Leu Ala			
530	535	540	
tgt gtt tat acc act aat aga aac gtc aat ccg gat ggt agc tta atc			1680
Cys Val Tyr Thr Thr Asn Arg Asn Val Asn Pro Asp Gly Ser Leu Ile			
545	550	555	560
gct aat ggt tac aat tgg cca gaa ttt tca cga gaa aaa ctc aca acg			1728
Ala Asn Gly Tyr Asn Trp Pro Glu Phe Ser Arg Glu Lys Leu Thr Thr			
565	570	575	
gct atg att gct cac cat ttt aga atg ttt acg att aga gct tgg cat			1776
Ala Met Ile Ala His His Phe Arg Met Phe Thr Ile Arg Ala Trp His			
580	585	590	
tta acg gat gga ttt aac gaa aat att gaa aac gcc gtg gat tat gac			1824
Leu Thr Asp Gly Phe Asn Glu Asn Ile Glu Asn Ala Val Asp Tyr Asp			
595	600	605	
atg ttc ctt aaa ctc agt gaa gtt gga aaa ttt aaa cat ctt aat aaa			1872
Met Phe Leu Lys Leu Ser Glu Val Gly Lys Phe Lys His Leu Asn Lys			
610	615	620	
atc tgc tat aac cgc gta tta cat ggt gat aac aca tcc att aag aaa			1920
Ile Cys Tyr Asn Arg Val Leu His Gly Asp Asn Thr Ser Ile Lys Lys			
625	630	635	640
ctc ggc att caa aag aaa aac cat ttt gtt gta gtc aat cag tca tta			1968
Leu Gly Ile Gln Lys Lys Asn His Phe Val Val Val Asn Gln Ser Leu			
645	650	655	
aat aga caa ggc atc aat tat tat aat tat gac aaa ttt gat gat tta			2016
Asn Arg Gln Gly Ile Asn Tyr Tyr Asn Tyr Asp Lys Phe Asp Asp Leu			
660	665	670	
gat gaa agt aga aag tat atc ttc aat aaa acc gct gaa tat caa gaa			2064
Asp Glu Ser Arg Lys Tyr Ile Phe Asn Lys Thr Ala Glu Tyr Gln Glu			
675	680	685	
gaa atg gat att tta aaa gat ctt aaa ctc att caa aat aaa gat gcc			2112
Glu Met Asp Ile Leu Lys Asp Leu Lys Leu Ile Gln Asn Lys Asp Ala			
690	695	700	
aaa atc gca gtc agt att ttc tat ccc aat aca tta aac ggc tta gtg			2160
Lys Ile Ala Val Ser Ile Phe Tyr Pro Asn Thr Leu Asn Gly Leu Val			

705	710	715	720	
aaa aaa cta aac aat att att gaa tat aat aaa aat ata ttc gtt att				2208
Lys Lys Leu Asn Asn Ile Ile Glu Tyr Asn Lys Asn Ile Phe Val Ile				
725	730	735		
att cta cat gtt gat aag aat cat ctt aca cca gac atc aaa aaa gaa				2256
Ile Leu His Val Asp Lys Asn His Leu Thr Pro Asp Ile Lys Lys Glu				
740	745	750		
ata ttg gct ttc tat cat aag cac caa gtg aat att tta cta aat aat				2304
Ile Leu Ala Phe Tyr His Lys His Gln Val Asn Ile Leu Leu Asn Asn				
755	760	765		
gac atc tca tat tac acg agt aat aga cta ata aaa act gag gca cat				2352
Asp Ile Ser Tyr Tyr Thr Ser Asn Arg Leu Ile Lys Thr Glu Ala His				
770	775	780		
tta agt aat att aat aaa tta agt cag tta aat cta aat tgt gaa tac				2400
Leu Ser Asn Ile Asn Lys Leu Ser Gln Leu Asn Leu Asn Cys Glu Tyr				
785	790	795	800	
atc att ttt gat aat cat gac agc cta ttc gtt aaa aat gac agc tat				2448
Ile Ile Phe Asp Asn His Asp Ser Leu Phe Val Lys Asn Asp Ser Tyr				
805	810	815		
gct tat atg aaa aaa tat gat gtc ggc atg aat ttc tca gca tta aca				2496
Ala Tyr Met Lys Lys Tyr Asp Val Gly Met Asn Phe Ser Ala Leu Thr				
820	825	830		
cat gat tgg atc gag aaa atc aat gcg cat cca cca ttt aaa aag ctg				2544
His Asp Trp Ile Glu Lys Ile Asn Ala His Pro Pro Phe Lys Lys Leu				
835	840	845		
att aaa acc tat ttt aat gac aat gac tta aga agt atg aat gtg aaa				2592
Ile Lys Thr Tyr Phe Asn Asp Asn Asp Leu Arg Ser Met Asn Val Lys				
850	855	860		
ggg gca tca caa ggt atg ttt atg aag tat gcg cta ccg cat gag ctt				2640
Gly Ala Ser Gln Gly Met Phe Met Lys Tyr Ala Leu Pro His Glu Leu				
865	870	875	880	
ctg acg att att aaa gaa gtc atc aca tcc tgc caa tca att gat agt				2688
Leu Thr Ile Ile Lys Glu Val Ile Thr Ser Cys Gln Ser Ile Asp Ser				
885	890	895		
gtg cca gaa tat aac act gag gat att tgg ttc caa ttt gca ctt tta				2736
Val Pro Glu Tyr Asn Thr Glu Asp Ile Trp Phe Gln Phe Ala Leu Leu				
900	905	910		
atc tta gaa aag aaa acc ggc cat gta ttt aat aaa aca tcg acc ctg				2784
Ile Leu Glu Lys Lys Thr Gly His Val Phe Asn Lys Thr Ser Thr Leu				

915	920	925	
act tat atg cct tgg gaa cga aaa tta caa tgg aca aat gaa caa att			2832
Thr Tyr Met Pro Trp Glu Arg Lys Leu Gln Trp Thr Asn Glu Gln Ile			
930	935	940	
caa agt gca aaa aaa ggc gaa aat atc ccc gtt aac aag ttc att att			2880
Gln Ser Ala Lys Lys Gly Glu Asn Ile Pro Val Asn Lys Phe Ile Ile			
945	950	955	960
aat agt ata acg cta			2895
Asn Ser Ile Thr Leu			
965			

<210>5

<211>965

<212>PRT

<213> 多杀巴斯德氏菌

<400>5

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

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

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

Ala Tyr Met Lys Lys Tyr Asp Val Gly Met Asn Phe Ser Ala Leu Thr
820 825 830
His Asp Trp Ile Glu Lys Ile Asn Ala His Pro Pro Phe Lys Lys Leu
835 840 845
Ile Lys Thr Tyr Phe Asn Asp Asn Asp Leu Arg Ser Met Asn Val Lys
850 855 860
Gly Ala Ser Gln Gly Met Phe Met Lys Tyr Ala Leu Pro His Glu Leu
865 870 875 880
Leu Thr Ile Ile Lys Glu Val Ile Thr Ser Cys Gln Ser Ile Asp Ser
885 890 895
Val Pro Glu Tyr Asn Thr Glu Asp Ile Trp Phe Gln Phe Ala Leu Leu
900 905 910
Ile Leu Glu Lys Lys Thr Gly His Val Phe Asn Lys Thr Ser Thr Leu
915 920 925
Thr Tyr Met Pro Trp Glu Arg Lys Leu Gln Trp Thr Asn Glu Gln Ile
930 935 940
Gln SerAla Lys Lys Gly Glu Asn Ile Pro Val Asn Lys Phe Ile Ile
945 950 955 960
Asn Ser Ile Thr Leu
965

<210>6

<211>1953

<212>DNA

<213> 多杀巴斯德氏菌

<220>

<221>CDS

<222>(1).. (1953)

<300>

<308>EMBL/AY292199

<309>2004-11-25

<313>(1).. (1953)

<400>6

atg aag gga aaa aaa gag atg act caa att caa ata gct aaa aat cca 48
Met Lys Gly Lys Lys Glu Met Thr Gln Ile Gln Ile Ala Lys Asn Pro
1 5 10 15
ccc caa cat gaa aaa gaa aat gaa ctc aac acc ttt caa aat aaa att 96
Pro Gln His Glu Lys Glu Asn Glu Leu Asn Thr Phe Gln Asn Lys Ile
20 25 30
gat agt cta aaa aca act tta aac aaa gac att att tct caa caa act 144

Asp Ser Leu Lys Thr Thr Leu Asn Lys Asp Ile Ile Ser Gln Gln Thr	
35 40 45	
cta ttg gca aaa cag gac agt aaa cat ccg cta tcc gca tcc ctt gaa	192
Leu Leu Ala Lys Gln Asp Ser Lys His Pro Leu Ser Ala Ser Leu Glu	
50 55 60	
aac gaa aat aaa ctt tta tta aaa caa ctc caa ttg gtt ctg caa gaa	240
Asn Glu Asn Lys Leu Leu Leu Lys Gln Leu Gln Leu Val Leu Gln Glu	
65 70 75 80	
ttt aaa aaa ata tat acc tat aat caa gca tta gaa gca aag cta gaa	288
Phe Lys Lys Ile Tyr Thr Tyr Asn Gln Ala Leu Glu Ala Lys Leu Glu	
85 90 95	
aaa gat aag caa aca aca tca ata aca gat tta tat aat gaa gtc gct	336
Lys Asp Lys Gln Thr Thr Ser Ile Thr Asp Leu Tyr Asn Glu Val Ala	
100 105 110	
aaa agt gat tta ggg tta gtc aaa gaa acc aac agc gca aat cca tta	384
Lys Ser Asp Leu Gly Leu Val Lys Glu Thr Asn Ser Ala Asn Pro Leu	
115 120 125	
gtc agt att atc atg aca tct cac aat aca gcg caa ttt atc gaa gct	432
Val Ser Ile Ile Met Thr Ser His Asn Thr Ala Gln Phe Ile Glu Ala	
130 135 140	
tct att aat tca tta ttg tta caa aca tat aaa aac ata gaa att att	480
Ser Ile Asn Ser Leu Leu Leu Gln Thr Tyr Lys Asn Ile Glu Ile Ile	
145 150 155 160	
att gta gat gat gat agc tcg gat aat aca ttt gaa att gcc tcg aga	528
Ile Val Asp Asp Asp Ser Ser Asp Asn Thr Phe Glu Ile Ala Ser Arg	
165 170 175	
ata gcg aat aca aca agc aaa gtc aga gta ttt aga tta aat tca aac	576
Ile Ala Asn Thr Thr Ser Lys Val Arg Val Phe Arg Leu Asn Ser Asn	
180 185 190	
cta gga act tac ttt gcg aaa aat aca ggc ata tta aaa tct aaa ggt	624
Leu Gly Thr Tyr Phe Ala Lys Asn Thr Gly Ile Leu Lys Ser Lys Gly	
195 200 205	
gac att att ttc ttt caa gat agt gat gat gta tgt cat cat gaa aga	672
Asp Ile Ile Phe Phe Gln Asp Ser Asp Asp Val Cys His His Glu Arg	
210 215 220	
ata gaa aga tgt gta aat ata tta tta gct aat aaa gaa act att gct	720
Ile Glu Arg Cys Val Asn Ile Leu Leu Ala Asn Lys Glu Thr Ile Ala	
225 230 235 240	
gtt cgt tgt gca tac tca aga cta gca cca gaa aca cag cat atc att	768

Val	Arg	Cys	Ala	Tyr	Ser	Arg	Leu	Ala	Pro	Glu	Thr	Gln	His	Ile	Ile	
245				250				255								
aaa	gtc	aat	aat	atg	gat	tat	aga	tta	ggt	ttt	ata	acc	ttg	ggt	atg	816
Lys	Val	Asn	Asn	Met	Asp	Tyr	Arg	Leu	Gly	Phe	Ile	Thr	Leu	Gly	Met	
260				265				270								
cac	aaa	aaa	gta	ttt	caa	gaa	att	ggt	ttc	ttc	aat	tgt	acg	act	aaa	864
His	Lys	Lys	Val	Phe	Gln	Glu	Ile	Gly	Phe	Phe	Asn	Cys	Thr	Thr	Lys	
275				280				285								
ggc	tca	gat	gat	gag	ttt	ttt	cat	aga	att	gcg	aaa	tat	tat	gga	aaa	912
Gly	Ser	Asp	Asp	Glu	Phe	Phe	His	Arg	Ile	Ala	Lys	Tyr	Tyr	Gly	Lys	
290				295				300								
gaa	aaa	ata	aaa	aat	tta	ctc	ttg	ccg	tta	tac	tac	aac	aca	atg	aga	960
Glu	Lys	Ile	Lys	Asn	Leu	Leu	Leu	Pro	Leu	Tyr	Tyr	Asn	Thr	Met	Arg	
305				310				315				320				
gaa	aac	tct	tta	ttt	act	gat	atg	gtt	gaa	tgg	ata	gac	aat	cat	aac	1008
Glu	Asn	Ser	Leu	Phe	Thr	Asp	Met	Val	Glu	Trp	Ile	Asp	Asn	His	Asn	
325				330				335								
ata	ata	cag	aaa	atg	tct	gat	acc	aga	caa	cat	tat	gca	acc	ctg	ttt	1056
Ile	Ile	Gln	Lys	Met	Ser	Asp	Thr	Arg	Gln	His	Tyr	Ala	Thr	Leu	Phe	
340				345				350								
caa	gcg	atg	cat	aac	gaa	act	gcc	tca	cat	gat	ttc	aaa	aat	ctt	ttt	1104
Gln	Ala	Met	His	Asn	Glu	Thr	Ala	Ser	His	Asp	Phe	Lys	Asn	Leu	Phe	
355				360				365								
caa	ttc	cct	cgt	att	tac	gac	gcc	tta	cca	gta	cca	caa	gaa	atg	agt	1152
Gln	Phe	Pro	Arg	Ile	Tyr	Asp	Ala	Leu	Pro	Val	Pro	Gln	Glu	Met	Ser	
370				375				380								
aag	ttg	tcc	aat	cct	aag	att	cct	gtt	tat	atc	aat	att	tgt	tct	att	1200
Lys	Leu	Ser	Asn	Pro	Lys	Ile	Pro	Val	Tyr	Ile	Asn	Ile	Cys	Ser	Ile	
385				390				395				400				
ccc	tca	aga	ata	gcg	caa	tta	caa	cgt	att	atc	ggc	ata	cta	aaa	aat	1248
Pro	Ser	Arg	Ile	Ala	Gln	Leu	Gln	Arg	Ile	Ile	Gly	Ile	Leu	Lys	Asn	
405				410				415								
caa	tgt	gat	cat	ttt	cat	att	tat	ctt	gat	ggc	tat	gta	gaa	atc	cct	1296
Gln	Cys	Asp	His	Phe	His	Ile	Tyr	Leu	Asp	Gly	Tyr	Val	Glu	Ile	Pro	
420				425				430								
gac	ttc	ata	aaa	aat	tta	ggt	aat	aaa	gca	acc	gtt	gtt	cat	tgc	aaa	1344
Asp	Phe	Ile	Lys	Asn	Leu	Gly	Asn	Lys	Ala	Thr	Val	Val	His	Cys	Lys	
435				440				445								
gat	aaa	gat	aac	tcc	att	aga	gat	aat	ggc	aaa	ttc	att	tta	ctg	gaa	1392

Asp Lys Asp Asn Ser Ile Arg Asp Asn Gly Lys Phe Ile Leu Leu Glu	
450	455
gag ttg att gaa aaa aat caa gat gga tat tat ata acc tgt gat gat	1440
Glu Leu Ile Glu Lys Asn Gln Asp Gly Tyr Tyr Ile Thr Cys Asp Asp	
465	470
gac att atc tat cca agc gat tac atc aat acg atg atc aag aag ctg	1488
Asp Ile Ile Tyr Pro Ser Asp Tyr Ile Asn Thr Met Ile Lys Lys Leu	
485	490
aat gaa tac gat gat aaa gcg gtt att ggt tta cac ggc att ctc ttt	1536
Asn Glu Tyr Asp Asp Lys Ala Val Ile Gly Leu His Gly Ile Leu Phe	
500	505
cca agt aga atg acc aaa tat ttt tcg gcg gat aga ctg gta tat agc	1584
Pro Ser Arg Met Thr Lys Tyr Phe Ser Ala Asp Arg Leu Val Tyr Ser	
515	520
ttc tat aaa cct ctg gaa aaa gac aaa gcg gtc aat gta tta ggt aca	1632
Phe Tyr Lys Pro Leu Glu Lys Asp Lys Ala Val Asn Val Leu Gly Thr	
530	535
gga act gtt agc ttt aga gtc agt ctc ttt aat caa ttt tct ctt tct	1680
Gly Thr Val Ser Phe Arg Val Ser Leu Phe Asn Gln Phe Ser Leu Ser	
545	550
gac ttt acc cat tca ggc atg gct gat atc tat ttc tct ctc ttg tgt	1728
Asp Phe Thr His Ser Gly Met Ala Asp Ile Tyr Phe Ser Leu Leu Cys	
565	570
aag aaa aat aat att ctt cag att tgt att tca aga cca gca aac tgg	1776
Lys Lys Asn Asn Ile Leu Gln Ile Cys Ile Ser Arg Pro Ala Asn Trp	
580	585
cta acg gaa gat aat aga gac agc gaa aca ctc tat cat caa tat cga	1824
Leu Thr Glu Asp Asn Arg Asp Ser Glu Thr Leu Tyr His Gln Tyr Arg	
595	600
gac aat gat gag caa caa act cag ctg atc atg gaa aac ggt cca tgg	1872
Asp Asn Asp Glu Gln Gln Thr Gln Leu Ile Met Glu Asn Gly Pro Trp	
610	615
gga tat tca agt att tat cca tta gtc aaa aat cat cct aaa ttt act	1920
Gly Tyr Ser Ser Ile Tyr Pro Leu Val Lys Asn His Pro Lys Phe Thr	
625	630
gac ctt atc ccc tgt tta cct ttt tat ttt tta	1953
Asp Leu Ile Pro Cys Leu Pro Phe Tyr Phe Leu	
645	650

<210>7

<211>651

<212>PRT

<213> 多杀巴斯德氏菌

<400>7

```

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

```

275	280	285
Gly Ser Asp Asp Glu Phe Phe His Arg Ile Ala Lys Tyr Tyr Gly Lys		
290	295	300
Glu Lys Ile Lys Asn Leu Leu Leu Pro Leu Tyr Tyr Asn Thr Met Arg		
305	310	315
Glu Asn Ser Leu Phe Thr Asp Met Val Glu Trp Ile Asp Asn His Asn		
325	330	335
Ile Ile Gln Lys Met Ser Asp Thr Arg Gln His Tyr Ala Thr Leu Phe		
340	345	350
Gln Ala Met His Asn Glu Thr Ala Ser His Asp Phe Lys Asn Leu Phe		
355	360	365
Gln Phe Pro Arg Ile Tyr Asp Ala Leu Pro Val Pro Gln Glu Met Ser		
370	375	380
Lys Leu Ser Asn Pro Lys Ile Pro Val Tyr Ile Asn Ile Cys Ser Ile		
385	390	395
Pro Ser Arg Ile Ala Gln Leu Gln Arg Ile Ile Gly Ile Leu Lys Asn		
405	410	415
Gln Cys Asp His Phe His Ile Tyr Leu Asp Gly Tyr Val Glu Ile Pro		
420	425	430
Asp Phe Ile Lys Asn Leu Gly Asn Lys Ala Thr Val Val His Cys Lys		
435	440	445
Asp Lys Asp Asn Ser Ile Arg Asp Asn Gly Lys Phe Ile Leu Leu Glu		
450	455	460
Glu Leu Ile Glu Lys Asn Gln Asp Gly Tyr Tyr Ile Thr Cys Asp Asp		
465	470	475
Asp Ile Ile Tyr Pro Ser Asp Tyr Ile Asn Thr Met Ile Lys Lys Leu		
485	490	495
Asn Glu Tyr Asp Asp Lys Ala Val Ile Gly Leu His Gly Ile Leu Phe		
500	505	510
Pro Ser Arg Met Thr Lys Tyr Phe Ser Ala Asp Arg Leu Val Tyr Ser		
515	520	525
Phe Tyr Lys Pro Leu Glu Lys Asp Lys Ala Val Asn Val Leu Gly Thr		
530	535	540
Gly Thr Val Ser Phe Arg Val Ser Leu Phe Asn Gln Phe Ser Leu Ser		
545	550	555
Asp Phe Thr His Ser Gly Met Ala Asp Ile Tyr Phe Ser Leu Leu Cys		
565	570	575
Lys Lys Asn Asn Ile Leu Gln Ile Cys Ile Ser Arg Pro Ala Asn Trp		
580	585	590

Leu Thr Glu Asp Asn Arg Asp Ser Glu Thr Leu Tyr His Gln Tyr Arg
 595 600 605
 Asp Asn Asp Glu Gln Gln Thr Gln Leu Ile Met Glu Asn Gly Pro Trp
 610 615 620
 Gly Tyr Ser Ser Ile Tyr Pro Leu Val Lys Asn His Pro Lys Phe Thr
 625 630 635 640
 Asp Leu Ile Pro Cys Leu Pro Phe Tyr Phe Leu
 645 650

<210>8

<211>480

<212>DNA

<213> 酿酒酵母 (*Saccharomyces cerevisiae*)

<220>

<221>CDS

<222>(1).. (477)

<300>

<308>GenBank/AY558564

<309>2007-04-24

<313>(1).. (477)

<400>8

atg agc tta ccc gat gga ttt tat ata agg cga atg gaa gag ggg gat 48
 Met Ser Leu Pro Asp Gly Phe Tyr Ile Arg Arg Met Glu Glu Gly Asp
 1 5 10 15
 ttg gaa cag gtc act gag acg cta aag gtt ttg acc acc gtg ggc act 96
 Leu Glu Gln Val Thr Glu Thr Leu Lys Val Leu Thr Thr Val Gly Thr
 20 25 30
 att acc ccc gaa tcc ttc agc aaa ctc ata aaa tac tgg aat gaa gcc 144
 Ile Thr Pro Glu Ser Phe Ser Lys Leu Ile Lys Tyr Trp Asn Glu Ala
 35 40 45
 aca gta tgg aat gat aac gaa gat aaa aaa ata atg caa tat aac ccc 192
 Thr Val Trp Asn Asp Asn Glu Asp Lys Lys Ile Met Gln Tyr Asn Pro
 50 55 60
 atg gtg att gtg gac aag cgc acc gag acg gtt gcc gct acg ggg aat 240
 Met Val Ile Val Asp Lys Arg Thr Glu Thr Val Ala Ala Thr Gly Asn
 65 70 75 80
 atc atc atc gaa aga aag atc att cat gaa ctg ggg cta tgt ggc cac 288
 Ile Ile Ile Glu Arg Lys Ile Ile His Glu Leu Gly Leu Cys Gly His
 85 90 95
 atc gag gac att gca gta aac tcc aag tat cag ggc caa ggt ttg ggc 336

Ile Glu Asp Ile Ala Val Asn Ser Lys Tyr Gln Gly Gln Gly Leu Gly	
100	105 110
aag ctc ttg att gat caa ttg gta act atc ggc ttt gac tac ggt tgt	384
Lys Leu Leu Ile Asp Gln Leu Val Thr Ile Gly Phe Asp Tyr Gly Cys	
115	120 125
tat aag att att tta gat tgc gat gag aaa aat gtc aaa ttc tat gaa	432
Tyr Lys Ile Ile Leu Asp Cys Asp Glu Lys Asn Val Lys Phe Tyr Glu	
130	135 140
aaa tgt ggg ttt agc aac gca ggc gtg gaa atg caa att aga aaa tag	480
Lys Cys Gly Phe Ser Asn Ala Gly Val Glu Met Gln Ile Arg Lys	
145	150 155

<210>9

<211>159

<212>PRT

<213> 酿酒酵母

<400>9

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

<210>10

<211>1434

<212>DNA

<213> 酿酒酵母

<220>

<221>CDS

<222>(1).. (1431)

<300>

<308>GenBank/X79380

<309>2005-04-18

<313>(3261).. (4694)

<400>10

atg act gac aca aaa cag cta ttc att gaa gcc gga caa agt caa ctt 48

Met Thr Asp Thr Lys Gln Leu Phe Ile Glu Ala Gly Gln Ser Gln Leu

1 5 10 15

ttc cac aat tgg gaa agc ttg tct cgc aaa gac caa gaa gaa ttg ctt 96

Phe His Asn Trp Glu Ser Leu SerArg Lys Asp Gln Glu Glu Leu Leu

20 25 30

tca aac ctg gag caa ata tct tcc aag agg tcc cct gca aaa cta ctg 144

Ser Asn Leu Glu Gln Ile Ser Ser Lys Arg Ser Pro Ala Lys Leu Leu

35 40 45

gaa gac tgt caa aat gct att aaa ttc tca cta gct aac tct tct aag 192

Glu Asp Cys Gln Asn Ala Ile Lys Phe Ser Leu Ala Asn Ser Ser Lys

50 55 60

gat act ggc gtc gaa att tca cca ttg ccc cct act tcg tac gag tcg 240

Asp Thr Gly Val Glu Ile Ser Pro Leu Pro Pro Thr Ser Tyr Glu Ser

65 70 75 80

ctt att ggc aac agt aag aaa gaa aat gaa tac tgg cgt tta ggc ctt 288

Leu Ile Gly Asn Ser Lys Lys Glu Asn Glu Tyr Trp Arg Leu Gly Leu

85 90 95

gaa gct att ggc aag ggt gaa gtc gca gtg att tta atg gct ggc gga 336

Glu Ala Ile Gly Lys Gly Glu Val Ala Val Ile Leu Met Ala Gly Gly

100 105 110

caa ggt acg cgg tta gga tcc tct caa cca aag ggc tgt tac gac att 384

Gln Gly Thr Arg Leu Gly Ser Ser Gln Pro Lys Gly Cys Tyr Asp Ile

115 120 125

gga ttg cct tct aag aaa tct ctt ttt caa att caa gct gaa aag ttg 432

Gly Leu Pro Ser Lys Lys Ser Leu Phe Gln Ile Gln Ala Glu Lys Leu

130 135 140

atc agg ttg caa gat atg gta aag gac aaa aag gta gaa att cct tgg 480

Ile Arg Leu Gln Asp Met Val Lys Asp Lys Lys Val Glu Ile Pro Trp

145 150 155 160

tat att atg aca tca ggc ccc act aga gcg gct act gag gca tac ttt	528
Tyr Ile Met Thr Ser Gly Pro Thr Arg Ala Ala Thr Glu Ala Tyr Phe	
165 170 175	
caa gaa cac aat tat ttt ggc ttg aat aaa gaa caa att acg ttc ttc	576
Gln Glu His Asn Tyr Phe Gly Leu Asn Lys Glu Gln Ile Thr Phe Phe	
180 185 190	
aac cag gga acc ctg cct gcc ttt gat tta acc ggg aag cat ttc cta	624
Asn Gln Gly Thr Leu Pro Ala Phe Asp Leu Thr Gly Lys His Phe Leu	
195 200 205	
atg aaa gac cca gta aac cta tct caa tca cca gat gga aat ggt gga	672
Met Lys Asp Pro Val Asn Leu Ser Gln Ser Pro Asp Gly Asn Gly Gly	
210 215 220	
ctc tac cgt gcc atc aag gaa aac aag ttg aac gaa gac ttt gat agg	720
Leu Tyr Arg Ala Ile Lys Glu Asn Lys Leu Asn Glu Asp Phe Asp Arg	
225 230 235 240	
aga gga atc aag cat gtt tac atg tac tgt gtc gat aat gtc cta tct	768
Arg Gly Ile Lys His Val Tyr Met Tyr Cys Val Asp Asn Val Leu Ser	
245 250 255	
aaa atc gca gac cct gta ttt att ggt ttt gcc atc aag cat ggc ttc	816
Lys Ile Ala Asp Pro Val Phe Ile Gly Phe Ala Ile Lys His Gly Phe	
260 265 270	
gaa ctg gcc acc aaa gcc gtt aga aag aga gat gcg cat gaa tca gtt	864
Glu Leu Ala Thr Lys Ala Val Arg Lys Arg Asp Ala His Glu Ser Val	
275 280 285	
ggg tta att gct act aaa aac gag aaa cca tgt gtc ata gaa tat tct	912
Gly Leu Ile Ala Thr Lys Asn Glu Lys Pro Cys Val Ile Glu Tyr Ser	
290 295 300	
gaa att tcc aat gaa ttg gct gaa gca aag gat aaa gat ggc tta tta	960
Glu Ile Ser Asn Glu Leu Ala Glu Ala Lys Asp Lys Asp Gly Leu Leu	
305 310 315 320	
aaa cta cgc gca ggc aac att gta aat cat tat tac cta gtg gat tta	1008
Lys Leu Arg Ala Gly Asn Ile Val Asn His Tyr Tyr Leu Val Asp Leu	
325 330 335	
cta aaa cgt gat ttg gat cag tgg tgt gag aat atg cca tat cac att	1056
Leu Lys Arg Asp Leu Asp Gln Trp Cys Glu Asn Met Pro Tyr His Ile	
340 345 350	
gcg aag aag aaa att cca gct tat gat agt gtt acc ggc aag tac act	1104
Ala Lys Lys Lys Ile Pro Ala Tyr Asp Ser Val Thr Gly Lys Tyr Thr	
355 360 365	

aag cct acc gaa cca aac ggt ata aaa tta gag caa ttc ata ttt gat	1152
Lys Pro Thr Glu Pro Asn Gly Ile Lys Leu Glu Gln Phe Ile Phe Asp	
370 375 380	
gtc ttt gac act gta cca ctg aac aag ttt ggg tgc tta gaa gta gat	1200
Val Phe Asp Thr Val Pro Leu Asn Lys Phe Gly Cys Leu Glu Val Asp	
385 390 395 400	
aga tgc aaa gaa ttt tca cct tta aaa aac ggt cct ggt tct aag aac	1248
Arg Cys Lys Glu Phe Ser Pro Leu Lys Asn Gly Pro Gly Ser Lys Asn	
405 410 415	
gat aat cct gag acc agc aga cta gca tat ttg aaa cta gga acc tcg	1296
Asp Asn Pro Glu Thr Ser Arg Leu Ala Tyr Leu Lys Leu Gly Thr Ser	
420 425 430	
tgg ttg gaa gat gca ggc gct att gta aaa gat ggg gta cta gtc gaa	1344
Trp Leu Glu Asp Ala Gly Ala Ile Val Lys Asp Gly Val Leu Val Glu	
435 440 445	
gtt tcc agc aaa ttg agt tat gca ggt gaa aat cta tcc cag ttc aaa	1392
Val Ser Ser Lys Leu Ser Tyr Ala Gly Glu Asn Leu Ser Gln Phe Lys	
450 455 460	
ggt aaa gtc ttt gac aga agt ggt ata gta tta gaa aaa taa	1434
Gly Lys Val Phe Asp Arg Ser Gly Ile Val Leu Glu Lys	
465 470 475	
<210>11	
<211>477	
<212>PRT	
<213> 酿酒酵母	
<400>11	
Met Thr Asp Thr Lys Gln Leu Phe Ile Glu Ala Gly Gln Ser Gln Leu	
1 5 10 15	
Phe His Asn Trp Glu Ser Leu Ser Arg Lys Asp Gln Glu Glu Leu Leu	
20 25 30	
Ser Asn Leu Glu Gln Ile Ser Ser Lys Arg Ser Pro Ala Lys Leu Leu	
35 40 45	
Glu Asp Cys Gln Asn Ala Ile Lys Phe Ser Leu Ala Asn Ser Ser Lys	
50 55 60	
Asp Thr Gly Val Glu Ile Ser Pro Leu Pro Pro Thr Ser Tyr Glu Ser	
65 70 75 80	
Leu Ile Gly Asn Ser Lys Lys Glu Asn Glu Tyr Trp Arg Leu Gly Leu	
85 90 95	
Glu Ala Ile Gly Lys Gly Glu Val Ala Val Ile Leu Met Ala Gly Gly	

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

Asp Asn Pro Glu Thr Ser Arg Leu Ala Tyr Leu Lys Leu Gly Thr Ser
 420 425 430
 Trp Leu Glu Asp Ala Gly Ala Ile Val Lys Asp Gly Val Leu Val Glu
 435 440 445
 Val Ser Ser Lys Leu Ser Tyr Ala Gly Glu Asn Leu Ser Gln Phe Lys
 450 455 460
 Gly Lys Val Phe Asp Arg Ser Gly Ile Val Leu Glu Lys
 465 470 475

<210>12

<211>919

<212>DNA

<213> 人工的

<220>

<223> 包含 YLCV 启动子、克隆位点、ocs- 终止信号和 nos- 终止信号序列的重组核酸序

列

<220>

<221> 启动子

<222> (8).. (352)

<223> YLCV 启动子序列

<220>

<221> 多腺苷酸化信号

<222> (366).. (663)

<223> 来自根瘤农杆菌 (*Agrobacterium tumefaciens*) 的 ocs 基因的多腺苷酸化信号

<220>

<221> 多腺苷酸化信号

<222> (664).. (872)

<223> 来自根瘤农杆菌的 nos 基因的多腺苷酸化信号

<400>12

```

gaattcctgg cagacaaagt ggcagacata ctgtcccaca aatgaagatg gaatctgtaa 60
aagaaaacgc gtgaaataat gcgtctgaca aaggttaggt cggctgcctt taatcaatac 120
caaagtggtc cctaccacga tggaaaaact gtgcagtcgg tttggctttt tctgacgaac 180
aaataagatt cgtggccgac aggtgggggt ccacatgtg aaggcatctt cagactccaa 240
taatggagca atgacgtaag ggcttacgaa ataagtaagg gtagtttggg aaatgtccac 300
tcacccgtca gtctataaat acttagcccc tccctcattg ttaagggagc aagagctcgc 360
ccgggatctc gaatcacgcg ttctaggatc cgaagcagat cggtcaaaca tttggcaata 420
aagtttctta agattgaatc ctgttgccgg tcttgcgatg attatcatat aatttctgtt 480
gaattacgtt aagcatgtaa taattaacat gtaatgcatg acgttattta tgagatgggt 540
ttttatgatt agagtcccgc aattatacat ttaatacgcg atagaaaaca aaatatagcg 600
cgcaaactag gataaattat cgcgcgcggt gtcatctatg ttactagatc gggaagatct 660

```

cgagacgtcg ggacaatcag taaattgaac ggagaatatt attcataaaa atacgatagt 720
aacgggtgat atattcatta gaatgaaccg aaaccggcgg taaggatctg agctacacat 780
gctcaggttt tttaacaagt gcacaacaga attgaaagca aatatcatgc gatcataggc 840
gtctcgcata tctcattaaa gcagggcattg cctgttttaa cattaattaa acctagggtga 900
cgtctaaaag ggcgaattc 919

<210>13

<211>48

<212>DNA

<213> 人工的

<220>

<223> 合成的寡核苷酸

<400>13

tcgacaggcc tggatcctta attaaactag tctcgaggag ctcggtac 48

<210>14

<211>40

<212>DNA

<213> 人工的

<220>

<223> 合成的寡核苷酸

<400>14

cgagctcctc gagactagtt taattaagga tccaggcctg 40

<210>15

<211>29

<212>DNA

<213> 人工的

<220>

<223> 合成的寡核苷酸

<400>15

cttaattaat agttgacgaa cggaagctg 29

<210>16

<211>25

<212>DNA

<213> 人工的

<220>

<223> 合成的寡核苷酸

<400>16

agcatgcttg cagaccgtca ttagg 25

<210>17

<211>26

<212>DNA	
<213> 人工的	
<220>	
<223> 合成的寡核苷酸	
<400>17	
aaagtgcttc ataagtagct caaaca	26
<210>18	
<211>20	
<212>DNA	
<213> 人工的	
<220>	
<223> 合成的寡核苷酸	
<400>18	
aacaccagat cgaactgcaa	20
<210>19	
<211>21	
<212>DNA	
<213> 人工的	
<220>	
<223> 合成的寡核苷酸	
<400>19	
ccgccttttt agccagttat c	21
<210>20	
<211>20	
<212>DNA	
<213> 人工的	
<220>	
<223> 合成的寡核苷酸	
<400>20	
ccactctgtc tgcaaaggaa	20
<210>21	
<211>1338	
<212>DNA	
<213> 大肠杆菌 (Escherichia coli)	
<220>	
<221>CDS	
<222>(1)..(1338)	
<300>	
<308>GenBank /AY616604	

$\langle 400 \rangle_{21}$

atg agt aat cgt aaa tat ttc ggt acc gat ggg att cgt ggt cgt gta	48
Met Ser Asn Arg Lys Tyr Phe Gly Thr Asp Gly Ile Arg Gly Arg Val	
1 5 10 15	
ggg gat gcg ccg atc aca cct gat ttt gtg ctt aag ctg ggt tgg gcc	96
Gly Asp Ala Pro Ile Thr Pro Asp Phe Val Leu Lys Leu Gly Trp Ala	
20 25 30	
gcg ggt aaa gtg ctg gcg cgc cac ggc tcc cgt aag att att att ggt	144
Ala Gly Lys Val Leu Ala Arg His Gly Ser Arg Lys Ile Ile Ile Gly	
35 40 45	
aaa gac acg cgt att tct ggc tat atg ctg gag tca gca ctg gaa gcg	192
Lys Asp Thr Arg Ile Ser Gly Tyr Met Leu Glu Ser Ala Leu Glu Ala	
50 55 60	
ggt ctg gcg gca gcg ggc ctt tcc gca ctc ttc act ggc ccg atg cca	240
Gly Leu Ala Ala Ala Gly Leu Ser Ala Leu Phe Thr Gly Pro Met Pro	
65 70 75 80	
aca ccg gcc gtg gct tat ctg acg cgt acc ttc cgc gca gag gcc gga	288
Thr Pro Ala Val Ala Tyr Leu Thr Arg Thr Phe Arg Ala Glu Ala Gly	
85 90 95	
att gtg ata tct gca tcg cat aac ccg ttc tac gat aat ggc att aaa	336
Ile Val Ile Ser Ala Ser His Asn Pro Phe Tyr Asp Asn Gly Ile Lys	
100 105 110	
ttc ttc tct atc gac ggc acc aaa ctg ccg gat gcg gta gaa gag gcc	384
Phe Phe Ser Ile Asp Gly Thr Lys Leu Pro Asp Ala Val Glu Glu Ala	
115 120 125	
atc gaa gcg gaa atg gaa aag gag atc agc tgc gtt gat tcg gca gaa	432
Ile Glu Ala Glu Met Glu Lys Glu Ile Ser Cys Val Asp Ser Ala Glu	
130 135 140	
ctg ggt aaa gcc agc cgt atc gtt gat gcc gcg ggt cgc tat atc gag	480
Leu Gly Lys Ala Ser Arg Ile Val Asp Ala Ala Gly Arg Tyr Ile Glu	
145 150 155 160	
ttt tgc aaa gcc acg ttc ccg aac gaa ctt agc ctc agt gaa ctg aag	528
Phe Cys Lys Ala Thr Phe Pro Asn Glu Leu Ser Leu Ser Glu Leu Lys	
165 170 175	
att gtg gtg gat tgt gca aac ggt gcg act tat cac atc gcg ccg aac	576
Ile Val Val Asp Cys Ala Asn Gly Ala Thr Tyr His Ile Ala Pro Asn	
180 185 190	

gtg ctg cgc gaa ctg ggg gcg aac gtt atc gct atc ggt tgt gag cca	624
Val Leu Arg Glu Leu Gly Ala Asn Val Ile Ala Ile Gly Cys Glu Pro	
195 200 205	
aac ggt gta aac atc aat gcc gaa gtg ggg gct acc gac gtt cgc gcg	672
Asn Gly Val Asn Ile Asn Ala Glu Val Gly Ala Thr Asp Val Arg Ala	
210 215 220	
ctc cag gct cgt gtg ctg gct gaa aaa gcg gat ctc ggt att gcc ttc	720
Leu Gln Ala Arg Val Leu Ala Glu Lys Ala Asp Leu Gly Ile Ala Phe	
225 230 235 240	
gac ggc gat ggc gat cgc gtg att atg gtt gac cat gaa ggc aat aaa	768
Asp Gly Asp Gly Asp Arg Val Ile Met Val Asp His Glu Gly Asn Lys	
245 250 255	
gtc gat ggc gat cag atc atg tat atc atc gcg cgt gaa ggt ctt cgt	816
Val Asp Gly Asp Gln Ile Met Tyr Ile Ile Ala Arg Glu Gly Leu Arg	
260 265 270	
cag ggc cag ctg cgt ggt ggc gct gtg ggt aca ttg atg agc aac atg	864
Gln Gly Gln Leu Arg Gly Gly Ala Val Gly Thr Leu Met Ser Asn Met	
275 280 285	
ggg ctt gaa ctg gcg ctg aaa cag tta gga att cca ttt gcg cgc gcg	912
Gly Leu Glu Leu Ala Leu Lys Gln Leu Gly Ile Pro Phe Ala Arg Ala	
290 295 300	
aaa gtg ggt gac cgc tac gta ctg gaa aaa atg cag gag aaa ggc tgg	960
Lys Val Gly Asp Arg Tyr Val Leu Glu Lys Met Gln Glu Lys Gly Trp	
305 310 315 320	
cgt atc ggt gca gag aat tcc ggt cat gtg atc ctg ctg gat aaa act	1008
Arg Ile Gly Ala Glu Asn Ser Gly His Val Ile Leu Leu Asp Lys Thr	
325 330 335	
act acc ggt gac ggc atc gtt gct ggc ttg cag gtg ctg gcg gcg atg	1056
Thr Thr Gly Asp Gly Ile Val Ala Gly Leu Gln Val Leu Ala Ala Met	
340 345 350	
gca cgt aac cat atg agc ctg cac gac ctt tgc agc ggc atg aaa atg	1104
Ala Arg Asn His Met Ser Leu His Asp Leu Cys Ser Gly Met Lys Met	
355 360 365	
ttc ccg cag att ctg gtt aac gta cgt tac acc gca ggt agc ggc gat	1152
Phe Pro Gln Ile Leu Val Asn Val Arg Tyr Thr Ala Gly Ser Gly Asp	
370 375 380	
cca ctt gag cat gag tca gtt aaa gcc gtg acc gca gag gtt gaa gct	1200
Pro Leu Glu His Glu Ser Val Lys Ala Val Thr Ala Glu Val Glu Ala	
385 390 395 400	

gcg ctg ggc aac cgt gga cgc gtg ttg ctg cgt aaa tcc ggc acc gaa 1248
 Ala Leu Gly Asn Arg Gly Arg Val Leu Leu Arg Lys Ser Gly Thr Glu
 405 410 415
 ccg tta att cgc gtg atg gtg gaa ggc gaa gac gaa gcg cag gtg act 1296
 Pro Leu Ile Arg Val Met Val Glu Gly Glu Asp Glu Ala Gln Val Thr
 420 425 430
 gaa ttt gca cac cgc atc gcc gat gca gta aaa gcc gtt taa 1338
 Glu Phe Ala His Arg Ile Ala Asp Ala Val Lys Ala Val
 435 440 445

<210>22

<211>445

<212>PRT

<213> 大肠杆菌

<400>22

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

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

<210>23

<211>20

<212>DNA

<213> 人工的

<220>

<223> 合成的寡核苷酸

<400>23

attaccgggc cagaatcact 20

<210>24		
<211>20		
<212>DNA		
<213> 人工的		
<220>		
<223> 合成的寡核苷酸		
<400>24		
gtcaggacgc gtatgttgaa		20
<210>25		
<211>1371		
<212>DNA		
<213> 大肠杆菌		
<220>		
<221>CDS		
<222>(1).. (1368)		
<300>		
<308>GenBank/AP009048.1		
<309>2007-05-19		
<313>(3721481).. (3722851)		
<400>25		
atg ttg aat aat gct atg agc gta gtg atc ctt gcc gca ggc aaa ggc		48
Met Leu Asn Asn Ala Met Ser Val Val Ile Leu Ala Ala Gly Lys Gly		
1 5 10 15		
acg cgc atg tat tcc gat ctt ccg aaa gtg ctg cat acc ctt gcc ggg		96
Thr Arg Met Tyr Ser Asp Leu Pro Lys Val Leu His Thr Leu Ala Gly		
20 25 30		
aaa gcg atg gtt cag cat gtc att gat gct gcg aat gaa tta ggc gca		144
Lys Ala Met Val Gln His Val Ile Asp Ala Ala Asn Glu Leu Gly Ala		
35 40 45		
gcg cac gtt cac ctg gtg tac ggt cac ggc ggc gat ctg cta aaa cag		192
Ala His Val His Leu Val Tyr Gly His Gly Gly Asp Leu Leu Lys Gln		
50 55 60		
gcg ctg aaa gac gac aac ctt aac tgg gtg ctt cag gca gag cag ctg		240
Ala Leu Lys Asp Asp Asn Leu Asn Trp Val Leu Gln Ala Glu Gln Leu		
65 70 75 80		
ggt acg ggt cat gca atg cag cag gcc gca cct ttc ttt gcc gat gat		288
Gly Thr Gly His Ala Met Gln Gln Ala Ala Pro Phe Phe Ala Asp Asp		
85 90 95		
gaa gac att tta atg ctc tac ggc gac gtg ccg ctg atc tct gtc gaa		336

Glu	Asp	Ile	Leu	Met	Leu	Tyr	Gly	Asp	Val	Pro	Leu	Ile	Ser	Val	Glu	
100				105				110								
aca	ctc	cag	cgt	ctg	cgt	gat	gct	aaa	ccg	cag	ggt	ggc	att	ggt	ctg	384
Thr	Leu	Gln	Arg	Leu	Arg	Asp	Ala	Lys	Pro	Gln	Gly	Gly	Ile	Gly	Leu	
115				120				125								
ctg	acg	gtg	aaa	ctg	gat	gat	ccg	acc	ggt	tat	gga	cgt	atc	acc	cgt	432
Leu	Thr	Val	Lys	Leu	Asp	Asp	Pro	Thr	Gly	Tyr	Gly	Arg	Ile	Thr	Arg	
130				135				140								
gaa	aac	ggc	aaa	gtt	acc	ggc	att	gtt	gag	cac	aaa	gat	gcc	acc	ggc	480
Glu	Asn	Gly	Lys	Val	Thr	Gly	Ile	Val	Glu	His	Lys	Asp	Ala	Thr	Gly	
145				150				155				160				
gag	cag	cgt	cag	att	cag	gag	atc	aac	acc	ggc	att	ctg	att	gcc	aac	528
Glu	Gln	Arg	Gln	Ile	Gln	Glu	Ile	Asn	Thr	Gly	Ile	Leu	Ile	Ala	Asn	
165				170				175								
ggc	gca	gat	atg	aaa	cgc	tgg	ctg	gcg	aag	ctg	acc	aac	aat	aat	gct	576
Gly	Ala	Asp	Met	Lys	Arg	Trp	Leu	Ala	Lys	Leu	Thr	Asn	Asn	Asn	Ala	
180				185				190								
cag	ggc	gaa	tac	tac	atc	acc	gac	att	att	gcg	ctg	gcg	tat	cag	gaa	624
Gln	Gly	Glu	Tyr	Tyr	Ile	Thr	Asp	Ile	Ile	Ala	Leu	Ala	Tyr	Gln	Glu	
195				200				205								
ggg	cgt	gaa	atc	gtc	gcc	gtt	cat	ccg	caa	cgt	tta	agc	gaa	gta	gaa	672
Gly	Arg	Glu	Ile	Val	Ala	Val	His	Pro	Gln	Arg	Leu	Ser	Glu	Val	Glu	
210				215				220								
ggc	gtg	aat	aac	cgc	ctg	caa	ctc	tcc	cgt	ctg	gag	cgt	gtt	tat	cag	720
Gly	Val	Asn	Asn	Arg	Leu	Gln	Leu	Ser	Arg	Leu	Glu	Arg	Val	Tyr	Gln	
225				230				235				240				
tcc	gaa	cag	gct	gaa	aaa	ctg	ctg	tta	gca	ggc	gtt	atg	ctg	cgc	gat	768
Ser	Glu	Gln	Ala	Glu	Lys	Leu	Leu	Leu	Ala	Gly	Val	Met	Leu	Arg	Asp	
245				250				255								
cca	gcg	cgt	ttt	gat	ctg	cgt	ggt	acg	cta	act	cac	ggg	cgc	gat	gtt	816
Pro	Ala	Arg	Phe	Asp	Leu	Arg	Gly	Thr	Leu	Thr	His	Gly	Arg	Asp	Val	
260				265				270								
gaa	att	gat	act	aac	gtt	atc	atc	gag	ggc	aac	gtg	act	ctc	ggt	cat	864
Glu	Ile	Asp	Thr	Asn	Val	Ile	Ile	Glu	Gly	Asn	Val	Thr	Leu	Gly	His	
275				280				285								
cgc	gtg	aaa	att	ggc	acc	ggt	tgc	gtg	att	aaa	aac	agc	gtg	att	ggc	912
Arg	Val	Lys	Ile	Gly	Thr	Gly	Cys	Val	Ile	Lys	Asn	Ser	Val	Ile	Gly	
290				295				300								
gat	gat	tgc	gaa	atc	agt	ccg	tat	acc	gtt	gtg	gaa	gat	gcg	aat	ctg	960

Asp Asp Cys Glu Ile Ser Pro Tyr Thr Val Val Glu Asp Ala Asn Leu	
305	310 315 320
gca gcg gcc tgt acc att ggc ccg ttt gcc cgt ttg cgt cct ggt gct	1008
Ala Ala Ala Cys Thr Ile Gly Pro Phe Ala Arg Leu Arg Pro Gly Ala	
325	330 335
gag ttg ctg gaa ggt gct cac gtc ggt aac ttc gtt gag atg aaa aaa	1056
Glu Leu Leu Glu Gly Ala His Val Gly Asn Phe Val Glu Met Lys Lys	
340	345 350
gcg cgt ctg ggt aaa ggc tcg aaa gct ggt catctg act tac ctg ggc	1104
Ala Arg Leu Gly Lys Gly Ser LysAla Gly His Leu Thr Tyr Leu Gly	
355	360 365
gat gcg gaa att ggc gat aac gtt aac atc ggc gcg gga acc att acc	1152
Asp Ala Glu Ile Gly Asp Asn Val Asn Ile Gly Ala Gly Thr Ile Thr	
370	375 380
tgc aac tac gat ggt gcg aat aaa ttt aag acc att atc ggc gac gat	1200
Cys Asn Tyr Asp GlyAla Asn Lys Phe Lys Thr Ile Ile Gly Asp Asp	
385	390 395 400
gtg ttt gtt ggt tcc gac act cag ctg gtg gcc ccg gta aca gta ggc	1248
Val Phe Val Gly Ser Asp Thr Gln Leu Val Ala Pro Val Thr Val Gly	
405	410 415
aaa ggc gcg acc att gct gcg ggt aca act gtg acg cgt aat gtc ggc	1296
Lys Gly Ala Thr Ile Ala Ala Gly Thr Thr Val Thr Arg Asn Val Gly	
420	425 430
gaa aat gca tta gct atc agc cgt gtg ccg cag act cag aaa gaa ggc	1344
Glu Asn Ala Leu Ala Ile Ser Arg Val Pro Gln Thr Gln Lys Glu Gly	
435	440 445
tgg cgt cgt ccg gta aag aaa aag tga	1371
Trp Arg Arg Pro Val Lys Lys Lys	
450	455
<210>26	
<211>456	
<212>PRT	
<213> 大肠杆菌	
<400>26	
Met Leu Asn Asn Ala Met Ser Val Val Ile Leu Ala Ala Gly Lys Gly	
1	5 10 15
Thr Arg Met Tyr Ser Asp Leu Pro Lys Val Leu His Thr Leu Ala Gly	
20	25 30
Lys Ala Met Val Gln His Val Ile Asp Ala Ala Asn Glu Leu Gly Ala	

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

Ala Arg Leu Gly Lys Gly Ser Lys Ala Gly His Leu Thr Tyr Leu Gly
355 360 365
Asp Ala Glu Ile Gly Asp Asn Val Asn Ile Gly Ala Gly Thr Ile Thr
370 375 380
Cys Asn Tyr Asp Gly Ala Asn Lys Phe Lys Thr Ile Ile Gly Asp Asp
385 390 395 400
Val Phe Val Gly Ser Asp Thr Gln Leu Val Ala Pro Val Thr Val Gly
405 410 415
Lys Gly Ala Thr Ile Ala Ala Gly Thr Thr Val Thr Arg Asn Val Gly
420 425 430
Glu Asn Ala Leu Ala Ile Ser Arg Val Pro Gln Thr Gln Lys Glu Gly
435 440 445
Trp Arg Arg Pro Val Lys Lys Lys
450 455

<210>27

<211>20

<212>DNA

<213> 人工的

<220>

<223> 合成的寡核苷酸

<400>27

ggcaaaggtc agcagtaagc 20

<210>28

<211>21

<212>DNA

<213> 人工的

<220>

<223> 合成的寡核苷酸

<400>28

tcaagcagat gccttaacgt g 21

<210>29

<211>1674

<212>DNA

<213> 酿酒酵母

<220>

<221>CDS

<222>(1)..(1671)

<400>29

atg aag gtt gat tac gag caa ttg tgc aaa ctc tac gat gac acg tgc 48

Met Lys Val Asp Tyr Glu Gln Leu Cys Lys Leu Tyr Asp Asp Thr Cys	
1 5 10 15	
cgc acaaag aat gtg cag ttc agt tac ggt acg gcc gga ttc aga acg	96
Arg Thr Lys Asn Val Gln Phe Ser Tyr Gly ThrAla Gly Phe Arg Thr	
20 25 30	
ctg gcc aag aat ttg gat acg gtg atg ttc agt act ggt ata ctg gcg	144
Leu Ala Lys Asn Leu Asp Thr Val Met Phe Ser Thr Gly Ile Leu Ala	
35 40 45	
gtt ctc agg tcg ctg aag ctt cag ggt cag tat gtg ggg gtg atg atc	192
Val Leu Arg Ser Leu Lys Leu Gln Gly Gln Tyr Val Gly Val Met Ile	
50 55 60	
acg gcg tcg cac aac cca tac cag gac aac ggg gtc aag atc gtg gaa	240
Thr Ala Ser His Asn Pro Tyr Gln Asp Asn Gly Val Lys Ile Val Glu	
65 70 75 80	
cca gac gga tcg atg ctt ttg gcc aca tgg gag cca tat gcc atg cag	288
Pro Asp Gly Ser Met Leu Leu Ala Thr Trp Glu Pro Tyr Ala Met Gln	
85 90 95	
ttg gcc aat gcg gcc tct ttt gcc act aat ttt gaa gaa ttt cgt gtt	336
Leu Ala Asn Ala Ala Ser Phe Ala Thr Asn Phe Glu Glu Phe Arg Val	
100 105 110	
gag ttg gcc aag ctg att gaa cac gaa aag att gat ttg aat aca acc	384
Glu Leu Ala Lys Leu Ile Glu His Glu Lys Ile Asp Leu Asn Thr Thr	
115 120 125	
gtc gtg cct cac atc gtg gtt ggg aga gac tct agg gaa agt agt cca	432
Val Val Pro His Ile Val Val Gly Arg Asp Ser Arg Glu Ser Ser Pro	
130 135 140	
tac ttg ctg cgc tgc ttg act tcc tcc atg gcc agc gtc ttc cac gcg	480
Tyr Leu Leu Arg Cys Leu Thr Ser Ser Met Ala Ser Val Phe His Ala	
145 150 155 160	
caa gtt ttg gac cta ggc tgt gtc act acg cct caa ttg cat tac att	528
Gln Val Leu Asp Leu Gly Cys Val Thr Thr Pro Gln Leu His Tyr Ile	
165 170 175	
act gat ttg tcc aac agg cgg aaa ctg gaa gga gac aca gcg cca gtt	576
Thr Asp Leu Ser Asn Arg Arg Lys Leu Glu Gly Asp Thr Ala Pro Val	
180 185 190	
gcc aca gaa cag gac tac tat tcg ttc ttt ata gga gcc ttc aac gag	624
Ala Thr Glu Gln Asp Tyr Tyr Ser Phe Phe Ile Gly Ala Phe Asn Glu	
195 200 205	
ctc ttc gcc acg tat cag ctg gag aag agg ctg tct gtc cca aaa ttg	672

Leu	Phe	Ala	Thr	Tyr	Gln	Leu	Glu	Lys	Arg	Leu	Ser	Val	Pro	Lys	Leu	
210				215				220								
ttc	ata	gac	aca	gcc	aat	ggt	atc	ggt	ggt	cca	cag	ttg	aaa	aaa	cta	720
Phe	Ile	Asp	Thr	Ala	Asn	Gly	Ile	Gly	Gly	Pro	Gln	Leu	Lys	Lys	Leu	
225				230				235				240				
ctg	gcc	tcc	gaa	gat	tgg	gac	gtg	cca	gcg	gag	caa	gtt	gag	gta	atc	768
Leu	Ala	Ser	Glu	Asp	Trp	Asp	Val	Pro	Ala	Glu	Gln	Val	Glu	Val	Ile	
245				250				255								
aac	gac	agg	tcc	gat	gtt	cca	gaa	ctg	ttg	aat	ttt	gaa	tgc	ggt	gcg	816
Asn	Asp	Arg	Ser	Asp	Val	Pro	Glu	Leu	Leu	Asn	Phe	Glu	Cys	Gly	Ala	
260				265				270								
gat	tat	gtg	aag	act	aac	cag	aga	tta	ccc	aag	ggt	ctt	tct	cca	tcc	864
Asp	Tyr	Val	Lys	Thr	Asn	Gln	Arg	Leu	Pro	Lys	Gly	Leu	Ser	Pro	Ser	
275				280				285								
tcg	ttt	gat	tcg	cta	tat	tgc	tcc	ttt	gat	ggt	gac	gca	gac	agg	gtt	912
Ser	Phe	Asp	Ser	Leu	Tyr	Cys	Ser	Phe	Asp	Gly	Asp	Ala	Asp	Arg	Val	
290				295				300								
gtg	ttc	tac	tat	gtc	gac	tca	gga	tca	aaa	ttt	cat	ttg	ttg	gat	ggt	960
Val	Phe	Tyr	Tyr	Val	Asp	Ser	Gly	Ser	Lys	Phe	His	Leu	Leu	Asp	Gly	
305				310				315				320				
gac	aaa	att	tcc	act	ttg	ttt	gca	aag	ttc	ttg	tct	aaa	caa	cta	gaa	1008
Asp	Lys	Ile	Ser	Thr	Leu	Phe	Ala	Lys	Phe	Leu	Ser	Lys	Gln	Leu	Glu	
325				330				335								
ttg	gca	cac	cta	gaa	cat	tct	ttg	aag	att	ggt	gtt	gtg	caa	act	gcc	1056
Leu	Ala	His	Leu	Glu	His	Ser	Leu	Lys	Ile	Gly	Val	Val	Gln	Thr	Ala	
340				345				350								
tat	gca	aac	ggc	agt	tcc	acc	gct	tac	ata	aaa	aat	acg	ttg	cac	tgt	1104
Tyr	Ala	Asn	Gly	Ser	Ser	Thr	Ala	Tyr	Ile	Lys	Asn	Thr	Leu	His	Cys	
355				360				365								
ccc	gtg	tct	tgc	act	aag	aca	ggt	gtt	aaa	cac	ttg	cat	cat	gaa	gct	1152
Pro	Val	Ser	Cys	Thr	Lys	Thr	Gly	Val	Lys	His	Leu	His	His	Glu	Ala	
370				375				380								
gcc	act	cag	tac	gat	att	ggc	att	tat	ttc	gaa	gca	aat	gga	cat	ggt	1200
Ala	Thr	Gln	Tyr	Asp	Ile	Gly	Ile	Tyr	Phe	Glu	Ala	Asn	Gly	His	Gly	
385				390				395				400				
acg	att	ata	ttc	agc	gaa	aaa	ttt	cat	cga	act	atc	aaa	tct	gaa	tta	1248
Thr	Ile	Ile	Phe	Ser	Glu	Lys	Phe	His	Arg	Thr	Ile	Lys	Ser	Glu	Leu	
405				410				415								
tcc	aag	tcc	aag	tta	aat	ggt	gat	acg	tta	gct	ttg	aga	act	ttg	aag	1296

Ser Lys Ser Lys Leu Asn Gly Asp Thr Leu Ala Leu Arg Thr Leu Lys	
420	425
tgt ttc tct gaa ttg att aat cag acc gtg gga gat gct att tca gac	1344
Cys Phe Ser Glu Leu Ile Asn Gln Thr Val Gly Asp Ala Ile Ser Asp	
435	440
atg ctt gct gtc ctt gct act ttg gcg att ttg asa atg tcg cca atg	1392
Met Leu Ala Val Leu Ala Thr Leu Ala Ile Leu Lys Met Ser Pro Met	
450	455
gat tgg gat gaa gag tat act gat ttg ccc aac aag ctg gtt aag tgc	1440
Asp Trp Asp Glu Glu Tyr Thr Asp Leu Pro Asn Lys Leu Val Lys Cys	
465	470
atc gtt cct gat agg tca att ttc caa acc acg gac cag gaa aga aas	1488
Ile Val Pro Asp Arg Ser Ile Phe Gln Thr Thr Asp Gln Glu Arg Lys	
485	490
ttg ctc aat cca gtg ggg ttg caa gac aag ata gat ctt gtg gta gcc	1536
Leu Leu Asn Pro Val Gly Leu Gln Asp Lys Ile Asp Leu Val Val Ala	
500	505
aag tat ccc atg gga aga agc ttt gtc aga gcc agt ggt acg gag gat	1584
Lys Tyr Pro Met Gly Arg Ser Phe Val Arg Ala Ser Gly Thr Glu Asp	
515	520
gcg gtg agg gtt tat gcg gaa tgt aag gac tcc tct aag tta ggt caa	1632
Ala Val Arg Val Tyr Ala Glu Cys Lys Asp Ser Ser Lys Leu Gly Gln	
530	535
ttt tgt gac gaa gtg gtg gag cac gtt aag gca tct gct tga	1674
Phe Cys Asp Glu Val Val Glu His Val Lys Ala Ser Ala	
545	550

<210>30

<211>557

<212>PRT

<213> 酿酒酵母

<400>30

Met Lys Val Asp Tyr Glu Gln Leu Cys Lys Leu Tyr Asp Asp Thr Cys	
1	15
Arg Thr Lys Asn Val Gln Phe Ser Tyr Gly Thr Ala Gly Phe Arg Thr	
20	30
Leu Ala Lys Asn Leu Asp Thr Val Met Phe Ser Thr Gly Ile Leu Ala	
35	45
Val Leu Arg Ser Leu Lys Leu Gln Gly Gln Tyr Val Gly Val Met Ile	
50	60

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

370	375	380
Ala Thr Gln Tyr Asp Ile Gly Ile Tyr Phe Glu Ala Asn Gly His Gly		
385	390	395
Thr Ile Ile Phe Ser Glu Lys Phe His Arg Thr Ile Lys Ser Glu Leu		400
	405	410
Ser Lys Ser Lys Leu Asn Gly Asp Thr Leu Ala Leu Arg Thr Leu Lys		415
	420	425
Cys Phe Ser Glu Leu Ile Asn Gln Thr Val Gly Asp Ala Ile Ser Asp		430
	435	440
Met Leu Ala Val Leu Ala Thr Leu Ala Ile Leu Lys Met Ser Pro Met		445
	450	455
Asp Trp Asp Glu Glu Tyr Thr Asp Leu Pro Asn Lys Leu Val Lys Cys		460
465	470	475
Ile Val Pro Asp Arg Ser Ile Phe Gln Thr Thr Asp Gln Glu Arg Lys		480
	485	490
Leu Leu Asn Pro Val Gly Leu Gln Asp Lys Ile Asp Leu Val Val Ala		495
	500	505
Lys Tyr Pro Met Gly Arg Ser Phe Val Arg Ala Ser Gly Thr Glu Asp		510
	515	520
Ala Val Arg Val Tyr Ala Glu Cys Lys Asp Ser Ser Lys Leu Gly Gln		525
	530	535
Phe Cys Asp Glu Val Val Glu His Val Lys Ala Ser Ala		540
545	550	555