

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公布说明书

[21] 申请号 200580011440.6

[51] Int. Cl.

C12N 15/80 (2006.01)

C07K 14/47 (2006.01)

[43] 公开日 2007年4月4日

[11] 公开号 CN 1942586A

[22] 申请日 2005.4.18

[21] 申请号 200580011440.6

[30] 优先权

[32] 2004.4.16 [33] EP [31] 04076175.1

[32] 2004.4.16 [33] EP [31] 04076180.1

[32] 2004.4.16 [33] EP [31] 04076182.7

[32] 2004.4.16 [33] EP [31] 04076179.3

[32] 2004.4.16 [33] EP [31] 04076176.9

[86] 国际申请 PCT/EP2005/051696 2005.4.18

[87] 国际公布 WO2005/100573 英 2005.10.27

[85] 进入国家阶段日期 2006.10.16

[71] 申请人 帝斯曼知识产权资产管理有限公司

地址 荷兰海尔伦

[72] 发明人 蒂鲍特·乔斯·维恩策尔

诺林·尼古拉斯·玛丽亚·艾里萨伯
斯·佩伊·范 阿恩·斯达姆

[74] 专利代理机构 北京东方亿思知识产权代理有限
责任公司

代理人 肖善强

权利要求书 2 页 说明书 33 页 附图 8 页

[54] 发明名称

用于在真菌细胞中表达基因的真菌启动子

[57] 摘要

本发明涉及经过分离的真菌启动子 DNA 序列，涉及包含这些启动子的 DNA 构建体、载体和真菌宿主细胞，所述启动子与编码多肽的编码序列可操作地相连。本发明还涉及用经过分离的新启动子表达基因和/或生产多肽的方法。本发明还涉及用本发明的新启动子改变转录水平和/或对内源基因的调控的方法。

1. 启动子 DNA 序列，选自由如下物质构成的组：
 - (a) DNA 序列，其中包含 SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4、SEQ ID NO: 5 的核苷酸序列，
 - (b) DNA 序列，其能与 (a) 的 DNA 序列杂交，
 - (c) DNA 序列，其与 (a) 的 DNA 序列至少 50%同源，
 - (d) (a) 至 (c) 所述的任何 DNA 序列的变体，以及
 - (e) (a) 至 (d) 所述的 DNA 序列中任何一种的亚序列。
2. DNA 构建体，其中包含如权利要求 1 所述的启动子 DNA 序列，以及与所述启动子 DNA 序列可操作地相连的编码序列，从而所述编码序列可在所述启动子 DNA 序列控制下被表达。
3. 宿主细胞，包含如权利要求 2 所述的 DNA 构建体。
4. 如权利要求 3 所述的宿主细胞，其中，所述宿主细胞是 *Aspergillus*、*Penicillium* 或 *Trichoderma* 的物种。
5. 如权利要求 4 所述的宿主细胞，其中，所述 *Aspergillus* 是 *Aspergillus niger* 或 *Aspergillus oryzae* 的物种。
6. 一种方法，用于在宿主细胞中表达基因，所述方法包括：
 - (a) 提供如权利要求 2 所述的 DNA 构建体，
 - (b) 用所述 DNA 构建体去转化合适的宿主细胞，以及
 - (c) 在所述启动子 DNA 序列控制下表达所述基因。
7. 一种方法，用于在合适的宿主中生产多肽，所述方法包括：
 - (a) 提供如权利要求 2 所述的 DNA 构建体，其中包含编码所述多肽的编码序列，
 - (b) 用所述 DNA 构建体去转化合适的宿主细胞，
 - (c) 在有益于所述多肽表达的培养条件下培养所述合适的宿主，以及
 - (d) 从培养物中回收所述多肽。
8. 一种方法，用于在合适的宿主中生产次级代谢产物：
 - (a) 提供如权利要求 2 所述的 DNA 构建体，其中包含编码对次级代

谢产物的生产中所涉及的酶的编码序列，

(b) 用所述 DNA 构建体去转化合适的宿主细胞，

(c) 在有益于所述次级代谢产物生产的培养条件下培养所述合适的宿主，以及

(d) 从培养物中回收所述次级代谢产物。

用于在真菌细胞中表达基因的真菌启动子

技术领域

本发明涉及 DNA 序列，具体而言，涉及经过分离的真菌启动子，并涉及包含这些启动子的 DNA 构建体、载体和真菌宿主细胞，其中，所述启动子与编码多肽的编码序列可操作地相连。本发明还涉及表达基因和/或生产多肽的方法。

发明背景

在真菌宿主细胞中生产重组多肽通常是通过下述方法完成的：构建表达盒，其中编码多肽的 DNA 处于适用于该宿主细胞的启动子的表达控制下。可通过质粒或载体介导的转化将表达盒引入宿主细胞。然后可通过在表达盒所含有的启动子正确发挥功能所必要的诱导条件下培养经过转化的宿主细胞来生产多肽。

对每种真菌宿主细胞而言，通过转化引入到真菌宿主中的编码序列的表达和该编码序列编码的重组多肽的生产需要获得功能性的启动子。已知有大量启动子在真菌宿主细胞中是有功能的。下面是启动子在物种间交叉使用的例子：已知 *Aspergillus nidulans* (*A. nidulans*) *gpdA* 基因的启动子在 *Aspergillus niger* (*A. niger*) 中有功能 (J Biotechnol. 1991 Jan;17(1):19-33. Intracellular and extracellular production of proteins in *Aspergillus* under the control of expression signals of the highly expressed *A. nidulans* *gpdA* gene. Punt PJ, Zegers ND, Busscher M, Pouwels PH, van den Hondel CA)。另一个例子是用于 *A. niger* 和 *A. nidulans* 的 *A. niger* beta-木糖苷酶 *xlnD* 启动子 (Transcriptional regulation of the xylanolytic enzyme system of *Aspergillus*, van Peij, NNME, PhD-thesis Landbouwniversiteit Wageningen, the Netherlands, ISBN 90-5808-154-0)，以及 *Escherichia coli* beta-葡糖醛酸酶基因在 *A. niger*、*A. nidulans* 和 *Cladosporium fulvum* 中的表达，见 Curr

Genet. 1989 Mar;15(3):177-80: Roberts IN, Oliver RP, Punt PJ, van den Hondel CA. "Expression of the Escherichia coli beta-glucuronidase gene in industrial and phytopathogenic filamentous fungi"所述。

但是，人们仍需要经过改进的启动子，用于控制被引入的基因的表达，用于控制内源基因的表达水平，用于对内源基因的表达加以调控或者用于介导内源基因的失活，或用于生产多肽，或用于前述应用的组合。上述经过改进的启动子可以，例如比以前已知的那些更强。它们也可以是可由特定的方便的底物或化合物诱导的。当需要在单种真菌宿主中同时过量表达多种不同的基因，知道若干种功能性启动子是有益的。为防止压制（特定转录因子的滴定（titration）），优选使用多种不同的启动子，对每种待表达基因使用一种特定的启动子。

附图说明

图 1 描述了 pGBTOPGLA 的质粒图谱，其是整合型葡糖淀粉酶表达载体。

图 2 描述了 pGBTOPGLA-2 的质粒图谱，其是具有多克隆位点的整合型葡糖淀粉酶表达载体。

图 3 描述了 pGBTOPGLA-3 的质粒图谱，其是一种整合型葡糖淀粉酶表达载体，其中含有与葡糖淀粉酶编码序列可操作地相连的本发明启动子。

图 4 描述了通过单同源重组进行整合的示意图。

图 5: WT1、WT2 和多种 pGBTOPGLA 载体转化子的葡糖淀粉酶活性。显示的是标准化的活性，其中，WT1 在第 4 天的活性被设置为 100%。

图 6 描述了 pGBDEL-PGGLAA 的质粒图谱，其是置换型载体。

图 7 描述了启动子置换的示意图。

图 8 描述了通过同源重组的整合示意图。

发明详述

本发明涉及选自由如下物质所构成的组：

- (a) DNA 序列，其中包含 SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4、SEQ ID NO: 5 的核苷酸序列，
- (b) DNA 序列，其能与 (a) 的 DNA 序列杂交，
- (c) DNA 序列，其与 (a) 的 DNA 序列至少 50%同源，
- (d) (a) 至 (c) 的任何 DNA 序列的变体，以及
- (e) (a) 至 (d) 的 DNA 序列中任何一种的亚序列 (subsequence)。

在本发明的上下文中，启动子 DNA 序列是：当该启动子 DNA 序列与编码序列可操作地相连时，能控制该编码序列表达的 DNA 序列。术语“可操作地相连”在本文中被定义为下述结构，其中启动子 DNA 被放置于相对编码序列来说合适的位置，使得启动子 DNA 序列能指导被编码序列编码的多肽的生产。

术语“编码序列”在本文中被定义为：在合适的控制序列的控制下，能转录为 mRNA，能翻译为多肽的核酸序列。编码序列的边界通常是 ATG 起始密码子（通常是 mRNA 5'末端的开放读码框的开始）和转录终止子序列（恰好处于 mRNA 3'末端开放读码框的下游）界定的。编码序列可以包括但不限于，基因组 DNA，cDNA，半合成、合成和重组核酸序列。

更具体地，术语“启动子”在本文中被定义为下述 DNA 序列，其能与 RNA 聚合酶结合，指导聚合酶到编码多肽的编码序列正确的下游转录开始位点，来初始化转录。RNA 聚合酶能有效地催化与编码区域适当 DNA 链互补的信使 RNA 的装配。术语“启动子”还将被理解为：包括用于转录为 mRNA 之后进行翻译的 5'非编码区域（启动子和翻译起始之间）、顺式作用的转录控制元件（例如增强子）以及其它能与转录因子反应的核苷酸序列。

在一种优选的实施方式中，本发明的启动子 DNA 是根据 SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5 的序列。

根据另一种优选的实施方式，本发明的启动子 DNA 序列是能与 SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5 杂交且仍然保持启动子活性的 DNA 序列。

优选通过测量与启动子可操作地相连的编码序列编码的蛋白的浓度来确定启动子活性。或者，通过测量与启动子可操作地相连的编码序列编码的蛋白的酶活来测定启动子浓度。根据一种优选的实施方式，通过测量 *lacZ* 报道基因 (reporter gene) 编码序列的表达来测定启动子活性 (及长度) (见 Luo (Gene 163 (1995) 127-131.) 根据另一种优选的实施方式，通过使用绿荧光蛋白作为编码序列来测定启动子活性 (见 Microbiology. 1999 Mar;145 (Pt 3):729-34. Santerre Henriksen AL, Even S, Muller C, Punt PJ, van den Hondel CA, Nielsen J.Study))。

此外，可以通过测量启动子控制下产生的转录子的 mRNA 水平来确定启动子活性。mRNA 水平可以，例如，通过 Northern 杂交印记来测量 (J. Sambrook, E. F. Fritsch, and T. Maniatus, 1989, Molecular Cloning, A Laboratory Manual, 2d edition, Cold Spring Harbor, N.Y.)。

本发明包括能在非常低严谨度条件下、优选低严谨度条件下、更优选中等严谨度条件下、更优选中高严谨度条件下、进一步更优选高严谨度条件下以及最优选非常高严谨度条件下与相应于如下分子的核酸探针杂交的 (经过分离的) 启动子 DNA 序列，所述分子是 (i) SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5 的第 1 至 2000 位核苷酸，优选地，第 100 至 1990 位核苷酸，更优选地，第 200 至 1980 位核苷酸，进一步更优选地，第 300 至 1970 位核苷酸，进一步更优选地，第 350 至 1950 位核苷酸，最优选地，第 360 至 1900 位核苷酸，(ii) 是 (i) 的亚序列；或 (iii) 是 (i)、(ii) 互补链 (J. Sambrook, E. F. Fritsch, and T. Maniatis, 1989, Molecular Cloning, A Laboratory Manual, 2d edition, Cold Spring Harbor, N.Y.)。SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5 的亚序列可以是至少 100 个核苷酸，优选至少 200 个核苷酸，更优选至少 300 个核苷酸，进一步更优选至少 400 个核苷酸，以及最优选地，至少 500 个核苷酸。

SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5 或其亚序列的核酸序列可被用于设计核酸探针，以按照本领域公知的方法来鉴定和克隆来自不同属或种的菌株的 DNA 启动子。具体而言，此类探针可被用于按照标准 Southern 杂交印记流程与目标属或种的基因组或 cDNA 杂交，以鉴定和分离其中相应的基因。此类探针可比全长序列短很多，但是长度应当有至少 15 个，优选至少 25 个，更优选至少 35 个核苷酸。此外，此类探针可被用于通过 PCR 来扩增 DNA 启动子。通过 PCR 克隆启动子的例子在本文中有所描述（见实施例 1.3）。更长的探针也可以使用。DNA、RNA 或肽核酸（PNA）探针可以使用。典型地，对探针进行标记，用于探测相应的基因（例如，用³²P、³³P、³H、³⁵S、生物素或亲和素或荧光标记来标记）。此类探针被包括在本发明中。

因此，可对从上述其它生物制得的基因组 DNA 或 cDNA 文库加以筛选，选出能与上述探针杂交并能编码多肽的 DNA。可通过琼脂糖或聚丙烯酰胺凝胶电泳，或其它分离技术来分离来自此类其它生物的基因组或其它 DNA。可将来自文库的 DNA 或分离的 DNA 转移并固定到硝酸纤维素或其它合适的载体物质上。为鉴定出与 SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5 或其亚序列同源的克隆或 DNA，可将载体物质用于 Southern 杂交印记。

就本发明的目的而言，杂交指：核酸序列与相应于 SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5 示出的核酸序列、它们的互补链或其亚序列的经过标记的核酸探针在非常低至非常高严谨度条件下杂交。用例如 X 射线膜来探测上述条件下与核酸探针杂交的分子。其它杂交技术也可使用，例如，使用荧光来探测以及玻璃侧面（glass side）和/或 DNA 微阵列作为支持物的技术。FEMS Yeast Res. 2003 Dec;4(3):259-69 (Daran-Lapujade P, Daran JM, Kotter P, Petit T, Piper MD, Pronk JT. "Comparative genotyping of the *Saccharomyces cerevisiae* laboratory strains S288C and CEN.PK113-7D using oligonucleotide microarrays"中给出了 DNA 微阵列杂交探测的例子。此外，PNA 微阵列用于杂交的用途被描

述于 *Nucleic Acids Res.* 2003 Oct 1 ;31 (19):e119 (Brandt O, Feldner J, Stephan A, Schroder M, Schnolzer M, Arlinghaus HF, Hoheisel JD, Jacob A. PNA microarrays for hybridisation of unlabelled DNA samples 中。

在一种优选的实施方式中，核酸探针是 SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5 的核酸序列。在另一种优选的实施方式中，核酸探针是具有 SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5 第 20 至 1980 位核苷酸，更优选地，SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5 的第 500 至 1950 位核苷酸，进一步更优选地，SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5 的第 800 至 1920 位核苷酸，以及最优选地，SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5 的第 900 至 1900 位核苷酸的序列。另一种优选的探针是 DNA 序列转录位点前的部分。

对于长度为至少 100 个核苷酸的长探针而言，非常低至非常高严谨度的条件被定义为：按照标准 Southern 杂交印记流程，于 42 摄氏度进行预杂交和杂交，在下述物质中进行：5 倍 SSPE，0.3% SDS，200 $\mu\text{g/ml}$ 经修剪和变性的鲑鱼精 DNA，以及，对于非常低和低严谨度而言，25%甲酰胺；对于中等和中高严谨度而言，35%甲酰胺；或对于高和非常高严谨度而言，50%甲酰胺。

对于长度为至少 100 个核苷酸的长探针而言，载体物质被洗三次，每次 15 分钟，其中使用 2 倍 SSC，0.2% SDS，优选在至少 45 摄氏度（非常低严谨度），更优选在至少 50 摄氏度（低严谨度），更优选在至少 55 摄氏度（中等严谨度），更优选在至少 60 摄氏度（中高严谨度），进一步更优选在至少 65 度（高严谨度）以及最优选在 70 摄氏度（非常高严谨度）进行。

对于长度为大约 15 个核苷酸至大约 70 个核苷酸的短探针而言，严谨度条件被定义为：预杂交、杂交和杂交后洗涤在比根据 Bolton and McCarthy (1962, *Proceedings of the National Academy of Sciences USA*

48:1390)的算式计算得到的 T_m 低 5 摄氏度至 10 摄氏度的温度下, 在 0.9 M NaCl、0.09 M Tris HCl pH 7.6、6 mM EDTA、0.5% NP-40、1 倍 Denhardt's 溶液、1 mM 焦磷酸钠、1 mM 磷酸二氢钠、0.1 mM ATP 和每 ml 0.2 mg 酵母 RNA 中, 按照标准 Southern 杂交印记流程来进行。

对于长度为大约 15 个核苷酸至大约 70 个核苷酸的短探针而言, 对载体物质的洗涤如下安排: 使用 6 倍 SSC 加 0.1% SDS 洗一次, 15 分钟, 用 6 倍 SSC 洗两次, 每次 15 分钟, 洗涤在比计算出的 T_m 低 5 至 10 摄氏度的温度下进行。

根据另一种优选的实施方式, SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5 首先被用于克隆与之可操作地相连的天然基因、编码序列或其部分。这可以用 SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5 或前文定义的其亚序列来开始, 用该序列作为探针。将探针与给定宿主 (*Aspergillus niger* 或本申请中定义的任何其它真菌宿主) 的 cDNA 或基因组文库杂交。一旦天然基因或其部分已被克隆, 随后就可以通过本文所述的杂交实验, 用其自身作为探针去克隆来自其它真菌的其同源基因。

在本发明的上下文中, 同源基因指与天然基因至少 50%同源 (相同) 的基因。优选地, 同源基因与天然基因至少 55%同源, 更优选地, 至少 60%, 更优选地, 至少 65%, 更优选地, 至少 70%, 进一步更优选地, 至少 75%, 优选地, 大约 80%, 更优选地, 大约 90%, 进一步更优选地, 大约 95%, 以及最优选地, 大约 97%同源。

同源基因编码序列上游的序列是本发明所包括的启动子。或者, 可以用 SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5 或前文定义的其亚序列去搜索基因组数据库, 例如, 使用本文所述的 BLAST 算法或比对方法, 对与本发明启动子可操作地相连的天然基因的序列、编码序列或其部分加以鉴定。该鉴定出的序列可被用于鉴定本申请所定义的任何其它真菌宿主中的直向同源物或同源基因。鉴定出的直向同源物或同源基因编码序列上游的序列是本发明所包括的启动子。

根据另一种优选的实施方式，本发明的启动子 DNA 序列是（经过分离的）DNA 序列，其与 SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5 至少 50%同源（相同）。优选地，该 DNA 序列与 SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5 至少 55%同源，更优选地，至少 60%，更优选地，至少 65%，更优选地，至少 70%，进一步更优选地，至少 75%，优选地，大约 80%，更优选地，大约 90%，进一步更优选地，大约 95%，以及最优选地，大约 97%同源。

就本发明的目的而言，两条核酸序列之间同源性（相同性）的程度优选通过 BLAST 程序来确定。用于进行 BLAST 分析的软件是公众可通过 National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>) 获得的。BLAST 算法参数 W、T 和 X 确定了比对的灵敏度和速度。BLAST 程序使用缺省值，字长 (W) 为 11，BLOSUM62 分数矩阵（见 Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA 89: 10915 (1989)）比对 (B) 为 50，期望 (E) 为 10，M=5，N=-4，以及对两条链都加以比对。

在另一种优选的实施方式中，启动子是 SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5 的亚序列，该亚序列仍然具有启动子活性。该亚序列优选含有至少大约 100 个核苷酸，更优选地，至少大约 200 个核苷酸，以及最优选地，至少大约 300 个核苷酸。

在另一种优选的实施方式中，亚序列是 SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5 所包含的核酸序列（除了来自 5'和/或 3'末端的一个或多个核苷酸已缺失之外），所述 DNA 序列仍具有启动子活性。

在另一种优选的实施方式中，启动子序列是“经修整的 (trimmed)”序列，即，来自翻译起始处和/或转录起始处上游的序列片段。对启动子进行修整以及功能分析的例子见 Gene. 1994 Aug 5;145(2): 179-87: the effect of multiple copies of the upstream region on expression of the *Aspergillus niger*

glucoamylase-encoding gene. Verdoes JC, Punt PJ, Stouthamer AH, van den Hondel CA 所述。

在本发明的另一种实施方式中，启动子 DNA 序列是 SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5 的变体。

术语“变体”或“变体启动子”在本文中被定义为具有下述核苷酸序列的启动子，所述核苷酸序列包含对亲本启动子一个或多个核苷酸的取代、缺失和/或插入，其中，变体启动子较之相应的亲本启动子具有更多或更少的活性。术语“变体启动子”将包括天然变体和体外产生的变体，它们是用本领域公知的方法，例如经典诱变、定点诱变和 DNA 改组（shuffling）获得的。变体启动子可以具有一处或多处突变。每处突变是独立的对核苷酸的取代、缺失和/或插入。

根据一种优选的实施方式，变体启动子是：较之最初鉴定出的启动子序列（SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5）具有至少一处经修饰的调控位点的启动子。此类调控位点可以被整体去除或按照上文的解释进行特定的突变。对此类启动子变体的调控由此被修饰，从而，例如，其不再被葡萄糖所诱导。此类启动子变体和如何获得它们的技术的例子被描述于 EP 673 429 或 WO 94/04673 中。这些专利通过引用并入本文。

启动子变体可以是等位基因启动子。等位基因变体指，占据相同染色体基因座的基因的两种或多种替代性形式中的任何形式。等位基因变化是通过突变天然产生的，其能导致种群多态性。可通过下述方法来获得变体启动子：（a）在非常低、低、中等、中高、高或非常高严谨度条件下，将 DNA 与（i）SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5，（ii）（i）的亚序列或（iii）（i）、（ii）的互补链杂交，以及（b）从 DNA 分离变体启动子。严谨度和洗涤条件如本文所述。

本发明的启动子可以是随序列提供有接头的启动子，所述接头用于如下目的：引入特定限制位点，协助启动子序列与编码多肽的核酸序列编码

区域之间的连接。

本文提供的序列信息不应被狭义理解为需要包括进被错误鉴定的碱基。本文公开的特定序列可容易地用于从有丝真菌，特别是 *Aspergillus niger* 分离原始 DNA 序列，再对其进行进一步的序列分析，由此鉴定出序列错误。

除非另有指明，本文中通过对 DNA 分子进行测序测定的所有核苷酸序列都是用自动 DNA 测序仪测定的。因此，如本领域关于通过该自动方法测定的任何 DNA 序列所已知的，本文测定的任何核苷酸序列可能含有一些错误。通过自动方法测定的核苷酸序列典型地与被测序 DNA 分子的实际核苷酸序列至少大约 90%相同，更典型地，至少大约 95%相同至至少大约 99.9%相同。可以通过其它方法，包括本领域公知的手动 DNA 测序方法对实际序列进行更为精确地测定。

本领域技术人员能鉴定出此类被错误鉴定的碱基，并且知道如何更正此类错误。

本发明包括启动子的功能等同物，其典型地含有不会改变目标启动子生物学功能的突变。术语“功能等同物”还包括 *A. niger* DNA 序列的直向同源物。*A. niger* DNA 序列的直向同源物是从其它菌株或物种中分离出来的，具有相似或相同生物活性的 DNA 序列。

本发明的启动子可从任何属的微生物中获得。就本发明的目的而言，本文中所用的与一给定来源相关的术语“从……获得”将表示：多肽是该来源生产的，或是被来自该来源的基因所插入的细胞生产的。

启动子序列可从真菌来源获得，优选从酵母菌株获得，例如 *Candida*、*Hansenula*、*Kluyveromyces*、*Pichia*、*Saccharomyces*、*Schizosaccharomyces* 或 *Yarrowia* 的菌株，更优选地，从 *Saccharomyces carlsbergensis*、*Saccharomyces cerevisiae*、*Saccharomyces diastaticus*、*Saccharomyces douglasii*、*Saccharomyces kluyveri*、*Saccharomyces norbensis* 或 *Saccharomyces oviformis* 的菌株获得。

在另一种优选的实施方式中，启动子序列从有丝真菌菌株获得，例如 *Acremonium*、*Aspergillus*、*Aureobasidium*、*Cryptococcus*、*Filibasidium*、

Fusarium、*Humicola*、*Magnaporthe*、*Mucor*、*Myceliophthora*、*Neocallimastix*、*Neurospora*、*Paecilomyces*、*Penicillium*、*Piromyces*、*Schizophyllum*、*Talaromyces*、*Thermoascus*、*Thielavia*、*Tolyocladium* 或 *Trichoderma* 的菌株，更优选地，从 *Aspergillus aculeatus*、*Aspergillus awamori*、*Aspergillus foetidus*、*Aspergillus japonicus*、*A. nidulans*、*A. niger*、*Aspergillus oryzae* (*A. oryzae*)、*Humicola insolens*、*Humicola lanuginosa*、*Mucor miehei*、*Myceliophthora thermophila*、*Neurospora crassa*、*Penicillium purpurogenum*、*Trichoderma harzianum*、*Trichoderma koningii*、*Trichoderma longibrachiatum*、*Trichoderma reesei* 或 *Trichoderma viride* 的菌株获得。

在另一种优选的实施方式中，启动子序列从 *Fusarium bactridioides*、*Fusarium cerealis*、*Fusarium crookwellense*、*Fusarium culmorum*、*Fusarium graminearum*、*Fusarium graminum*、*Fusarium heterosporum*、*Fusarium negundi*、*Fusarium oxysporum*、*Fusarium reticulatum*、*Fusarium roseum*、*Fusarium sambucinum*、*Fusarium sarcochroum*、*Fusarium sporotrichioides*、*Fusarium sulphureum*、*Fusarium torulosum*、*Fusarium trichothecioides*、*Fusarium venenatum* 获得。

应理解，对上述物种而言，本发明包括完美的和不完美的状态以及其它分类学等同物，例如，无性型（anamorph），而不管它们是作为什么名字被知道的。本领域技术人员容易对合适的等同物加以鉴定。这些物种的菌株是公众容易从大量的培养单位获得的，例如，American Type Culture Collection（ATCC）、Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH（DSM）、Centraalbureau Voor Schimmelcultures（CBS）和 Agricultural Research Service Patent Culture Collection, Northern Regional Research Center（NRRL）。

此外，可使用上文提到的探针，从其它来源（包括从自然界（例如，土壤、肥料、水等）分离的微生物）鉴定并获得根据本发明的启动子序列。用于从自然环境分离微生物的技术是本领域公知的。然后可以通过对其它微生物基因组 DNA 文库进行类似筛选来获得核酸序列。一旦用探针

探测出了编码启动子的核酸序列，就可以利用本领域普通技术人员已知的技术来分离或克隆该序列（例如，见上述 Sambrook et al., 1989）。

在本发明中，启动子 DNA 序列还可以是杂交启动子，其中包含本发明的一种或多种启动子的一部分；本发明的启动子的一部分以及另外的已知启动子的一部分，例如，一种启动子的引导子（leader）序列和来自其它启动子的转录起始位点；或者本发明的一种或多种启动子的一部分以及一种或多种其它启动子的一部分。其它启动子可以是任何在选用的真菌宿主细胞中显示出转录活性的启动子序列，包括变体、截短的和杂交启动子，其可以从编码对于宿主细胞来说同源或异源的、细胞外或细胞内多肽的基因获得。其它启动子序列可以是对编码多肽的核酸序列来说天然的或外源的，并且对细胞来说可以是天然的或外源的。

作为一种优选的实施方式，鉴定出的启动子的重要调控序列可以与其它“基础”启动子融合，以增强它们的启动子活性（例如，见 Mol Microbiol. 1994 May;12(3):479-90. Regulation of the xylanase- encoding xlnA gene of *Aspergillus tubigenensis*. de Graaff LH, van den Broeck HC, van Ooijen AJ, Visser J.所述）。

可用于与本发明的启动子一起构建杂交启动子的其它启动子的例子包括：从 *A. oryzae* TAKA 淀粉酶、*Rhizomucor miehei* 天冬氨酸蛋白酶、*A. niger* 中性 alpha 淀粉酶、*A. niger* 酸稳定性 alpha 淀粉酶、*A. niger* 或 *Aspergillus awamori* 葡糖淀粉酶（glaA）、*A. niger* gpdA、*A. niger* 葡萄糖氧化酶 goxC、*Rhizomucor miehei* 脂酶、*A. oryzae* 碱性蛋白酶、*A. oryzae* 磷酸丙糖异构酶、*A. nidulans* 乙酰胺酶和 *Fusarium oxysporum* 类胰岛素蛋白酶（WO 96/00787）的基因获得的启动子，以及 NA2-tpi 启动子（来自用于（*A. niger* 中性 alpha 淀粉酶和 *A. oryzae* 磷酸丙糖异构酶的基因的启动子的杂交体）、*Saccharomyces cerevisiae* 烯醇酶（ENO-1）、*Saccharomyces cerevisiae* 半乳糖激酶（GAL1）、*Saccharomyces cerevisiae* 醇脱氢酶/甘油醛-3-磷酸酯脱氢酶（ADH2/GAP）和 *Saccharomyces cerevisiae* 3-磷酸甘油酯激酶的基因获得的启动子以及它们的突变的、截短的和杂交的启动子。其它可用于酵母宿主细胞的启动子见 Romanos et al., 1992, Yeast 8: 423-488

所述。

在本发明中，启动子 DNA 序列还可以是“串联启动子”。“串联启动子”在本文中被定义为两个或多个启动子序列，它们中的每一个都与编码序列可操作地相连，并对编码序列向 mRNA 的转录加以调节。

串联启动子包含：本发明的两种或多种启动子或本发明的一种或多种启动子与一种或多种其它已知的启动子，例如，上文例示的用于构建杂交启动子的那些。串联启动子的两种或多种启动子序列可以同时促进核酸序列的转录。或者，串联启动子的一种或多种启动子序列可以促进核酸序列在细胞不同生长阶段或者菌丝体不同形态部位的转录。

在本发明中，启动子对于编码目标多肽的编码序列和/或真菌宿主细胞来说可以是外源的。本发明的变体、杂交或串联启动子应当被理解为：对于编码多肽的编码来说是外源的，即使野生型启动子是编码序列或真菌宿主细胞天然的。

本发明的变体、杂交或串联启动子具有：具有 SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5 的启动子的启动子活性的至少大约 20%，优选至少大约 40%，更优选至少大约 60%，更优选至少大约 80%，更优选至少大约 90%，更优选至少大约 100%，进一步更优选至少大约 200%，最优选至少大约 300%，以及进一步最优选至少大约 400%。

本发明还涉及一种 DNA 构建体，其中包含至少一种上文定义的启动子 DNA 序列，以及与所述启动子 DNA 序列可操作地相连的编码序列，从而该编码序列可于给定的真菌宿主细胞中在启动子 DNA 序列的控制下被表达。

编码序列编码多肽，其对于目标真菌宿主细胞来说可以是天然的或异源的。

术语“多肽”在本文中并不代表被编码产物的特定长度，因此其包括肽、寡肽和蛋白。术语“异源多肽”在本文中被定义为：并非真菌细胞天然的多肽，经过修饰改变了天然序列的天然多肽，或通过重组 DNA 技术对真菌细胞进行操作导致表达从数量上被改变的天然多肽。例如，天然多

肽可以通过下述方法被重组生产，所述方法例如，将编码多肽的序列置于本发明启动子的控制之下，以增强多肽的表达，用信号序列加速目标天然多肽被运出细胞的过程，以及增加编码通常由该细胞产生的多肽的基因的拷贝数。真菌细胞可以含有一个或多个拷贝的、编码多肽的编码序列。

优选地，编码序列编码肽激素或其变体、酶、细胞内蛋白、分泌过程涉及的蛋白、折叠过程涉及的蛋白、分子伴侣（chaperone）、肽氨基酸转运蛋白、糖基化因子、转录因子、受体或其一部分、抗体或其一部分或报道基因。

在一种优选的实施方式中，多肽是细胞外分泌的。

在一种更优选的实施方式中，多肽是酶。酶的例子是纤维素酶，例如，内切葡聚糖酶、 β -葡聚糖酶、纤维二糖水解酶或 β -葡糖苷酶；半纤维素酶或果胶水解（pectinolytic）酶，例如，木聚糖酶、木糖苷酶、甘露聚糖酶、半乳糖酶、半乳糖苷酶、果胶甲酯酶、果胶裂解酶、果胶酸裂解酶、内切聚半乳糖醛酸酶、外切聚半乳糖醛酸酶、鼠李半乳糖醛酸酶、阿拉伯聚糖酶、阿拉伯呋喃糖苷酶、阿拉伯木聚糖水解酶、半乳糖醛酸酶，裂解酶或淀粉水解酶；磷酸酶，例如，植酸酶，酯酶，例如脂酶，蛋白水解酶，例如蛋白酶，肽酶，氧化还原酶，例如氧化酶，转移酶或异构酶。

或者，编码序列可以编码细胞内蛋白，例如，分子伴侣或转录因子。这方面的例子被描述于 Appl Microbiol Biotechnol. 1998 Oct; 50(4):447-54 ("Analysis of the role of the gene bipA, encoding the major endoplasmic reticulum chaperone protein in the secretion of homologous and heterologous proteins in black Aspergilli. Punt PJ, van Gemeren IA, Drint-Kuijvenhoven J, Hessing JG, van Muijlwijk-Harteveld GM, Beijersbergen A, Verrips CT, van den Hondel CA)中。这可用于：例如，提高宿主细胞作为蛋白生产者的效率，如果该编码序列，例如分子伴侣或转录因子已知是蛋白生产限制因素的话。

编码序列还可编码初级或次级代谢产物合成中涉及的酶，所述代谢产物例如，有机酸、类胡萝卜素，（ β -内酰胺）抗生素、维生素。

编码目标多肽的编码序列可从任何原核生物、真核生物或其它来源获

得。

或者，编码序列可针对反义 RNA 和/或 RNAi (RNA 干扰) 构建体的表达进行编码。编码反义-RNA 的例子见 Appl Environ Microbiol. 2000 Feb;66(2):775-82. (Characterization of a foldase, protein disulfide isomerase A, in the protein secretory pathway of *Aspergillus niger*. Ngiam C, Jeenes DJ, Punt PJ, Van Den Hondel CA, Archer DB)或(Zrenner R, Willmitzer L, Sonnewald U. Analysis of the expression of potato uridinediphosphate-glucose pyrophosphorylase and its inhibition by antisense RNA. Planta. (1993);190(2):247-52)。对基因表达完全失活可用于，例如，用于对控制不想要的代谢途径支链的基因失活，例如，增加对特定次级代谢产物（例如（beta-内酰胺）抗生素或类胡萝卜素）的生产。完全失活还可用于减少毒性或不想要的化合物的产生（*Penicillium* 中的黄青霉素、*Aspergillus* 中的黄曲霉素: MacDonald KD et al.; heterokaryon studies and the genetic control of penicillin and chrysogenin production in *Penicillium chrysogenum*. J Gen Microbiol. (1963) 33:375-83）。完全失活还可用于改变生物的形态，从而改进发酵过程和下游加工。

本发明的另一种实施方式涉及对真菌细胞进行广泛的代谢重构（reprogramming）或工程改造。将全新的途径引入和/或对不想要的途径的修饰将提供被特异性地改造过的细胞，所述改造针对特定化合物（例如蛋白质或代谢产物）的生产。

在本发明的方法中，当编码序列编码多肽时，所述多肽还可以包括融合或杂交多肽，其中，另一种多肽融合到前述多肽或其片段的 N-末端或 C-末端上。融合的多肽可以通过将编码一种多肽的核酸序列（或其一部分）融合到编码另一种多肽的核酸序列（或其一部分）来生产。用于生产融合多肽的技术是本领域内已知的，其包括将编码多肽的编码序列相连，使得它们处于同一读码框内（in frame），并且融合多肽的表达处于同样的启动子和终止子的控制之下。杂交多肽可以包含从至少两种不同的多肽（其中，一种或多种可以是真菌细胞异源的）获得的部分或全部多肽序列的组合。

除启动子 DNA 序列之外，DNA 构建体还可以包含一种或多种控制序列，它们能在与控制序列兼容的条件下，指导编码序列在合适的宿主细胞中的表达。表达应被理解为包括：生产多肽所涉及的任何步骤，其包括但不限于，转录、转录后修饰、翻译、翻译后修饰以及分泌。一种或多种控制序列可以是编码序列或宿主天然的。或者，一种或多种控制序列可被对核酸序列来说外源的一种或多种控制序列置换，用于提高编码序列在宿主细胞中的表达。

“DNA 构建体”在本文中被定义为下述核酸分子，其是单链或双链的，是从天然存在的基因分离出来的，或已经过修饰，以便含有以非天然存在的方式结合及并置的核酸片断。当 DNA 构建体含有编码序列以及表达该编码序列所需的所有控制序列时，术语“DNA 构建体”与术语“表达盒”同义。

术语“控制序列”在本文中被定义为包括：对于表达编码序列来说必要或有益的所有组件，其中包括本发明的启动子。每种控制序列可以是编码多肽的核酸序列天然的或外源的。此类控制序列包括但不限于，引导子（leader）、优化的翻译初始化序列（如 Kozak, 1991, J. Biol. Chem. 266:19867-19870 所述）、聚腺苷化序列、前肽序列、信号肽序列、上游活化序列、本发明的启动子（包括从其获得的变体、片段和杂交体，以及串联启动子）和转录终止子。至少，控制序列包括转录和翻译终止信号以及本发明的启动子（的一部分）。控制序列是可随接头提供，所述接头用于如下目的：引入特定限制位点，协助控制序列与编码多肽的核酸序列编码区域之间的连接。

控制序列可以是合适的终止子序列，即，可被宿主细胞识别以终止转录的序列。终止子序列与编码多肽的编码序列的 3' 末端可操作地相连。任何在宿主细胞中具有功能的终止子都可用于本发明。

用于有丝真菌宿主细胞的优选终止子可从 *A. oryzae* TAKA 淀粉酶、*A. niger* 葡糖淀粉酶、*A. nidulans* 氨基苯甲酸盐/酯合酶、*A. niger* alpha-葡糖苷酶、trpC 基因以及 *Fusarium oxysporum* 类胰岛素蛋白酶的基因获得。

用于酵母宿主细胞的优选终止子可从 *Saccharomyces cerevisiae* 烯醇

酶、*Saccharomyces cerevisiae* 细胞色素 C (CYC1) 和 *Saccharomyces cerevisiae* 甘油醛-3-磷酸酯脱氢酶的基因获得。其它可用于酵母宿主细胞的终止子如前述 Romanos et al, 1992 所述。

控制序列可以是合适的引导子序列，即，对宿主细胞的翻译来说重要的 mRNA 5'非翻译区域。引导子序列可与编码多肽的核酸序列的 5'末端可操作地相连。在选用的宿主细胞中有功能的任何引导子序列可用于本发明。

优选用于有丝真菌宿主细胞的引导子获得自 *A. oryzae* TAKA 淀粉酶、*A. nidulans* 磷酸丙糖异构酶和 *A. niger* glaA 的基因。

用于酵母宿主细胞的合适的引导子是从 *Saccharomyces cerevisiae* 烯醇酶 (ENO-1)、*Saccharomyces cerevisiae* 3-磷酸甘油酯激酶、*Saccharomyces cerevisiae* alpha-因子和 *Saccharomyces cerevisiae* 醇脱氢酶/甘油醛-3-磷酸酯脱氢酶 (ADH2/GAP) 基因获得的。

控制序列还可以是聚腺苷化序列，这是与核酸序列 3'末端可操作地相连的序列，当转录时，其可被宿主细胞作为信号识别，用于向转录的 mRNA 添加聚腺苷残基。在选用的宿主细胞中有功能的任何聚腺苷化序列可用于本发明。

优选用于有丝真菌宿主细胞的聚腺苷化序列是从 *A. oryzae* TAKA 淀粉酶、*A. niger* 葡糖淀粉酶、*A. nidulans* 氨基苯甲酸盐/酯合酶、*Fusarium oxysporum* 类胰岛素蛋白酶和 *A. niger* alpha-葡糖苷酶的基因获得的。

可用于酵母宿主细胞的聚腺苷化序列见 Guo and Sherman, 1995, *Molecular Cellular Biology* 15: 5983-5990 所述。

控制序列还可以是信号肽编码区域，其编码与多肽氨基末端相连的氨基酸序列，并能指导编码的多肽向细胞中分泌的途径。核酸序列的编码序列的 5'末端可以遗传性地含有下述信号肽编码区域，其天然就与编码被分泌多肽的编码区域片断以同一翻译读码框的方式相连。或者，编码序列的 5'可以含有对编码序列来说是外源的信号肽编码区域。当编码区域天然不含信号肽编码区域的情况下，可能会需要外源信号肽编码区域。或者，外源信号肽编码区域可以简单地置换天然信号肽编码区域，以提高多肽的分

泌。但是，能指导被表达的多肽进入选用的宿主细胞的分泌途径的任何信号肽编码区域都可用于本发明。

对于有丝真菌宿主细胞而言，有效的信号肽编码区域是从 *A. oryzae* TAKA 淀粉酶、*A. niger* 中性淀粉酶、*A. ficuum* 植酸酶、*A. niger* 葡糖淀粉酶、*A. niger* 内切木聚糖酶、*Rhizomucor miehei* 天冬氨酸蛋白酶、*Humicola insolens* 纤维素酶和 *Humicola lanuginosa* 脂酶基因获得的信号肽编码区域。

可用于酵母宿主细胞的信号肽是从 *Saccharomyces cerevisiae* alpha 因子和 *Saccharomyces cerevisiae* 转化酶的基因获得的。其它有用的信号肽区域见前述 Romanos et al., 1992 所述。

控制序列还可以是前肽编码区域，其编码位于多肽氨基末端的氨基酸序列。得到的多肽被称为酶原或前多肽（或者，某些情况下，酵素原（zymogen））。前多肽通常没有活性，可通过从前多肽上催化或自动催化切割下前肽，将其转化为成熟的活性多肽。前肽编码区域可从 *Bacillus subtilis* 碱性蛋白酶（aprE）、*Bacillus subtilis* 中性蛋白酶（nprT）、*Saccharomyces cerevisiae* alpha 因子、*Rhizomucor miehei* 天冬氨酸蛋白酶、*Myceliophthora thermophila* 漆酶（WO 95/33836）和 *A. niger* 内切木聚糖酶（endo1）的基因获得。

当信号肽和前肽区域都存在于多肽氨基末端时，前肽区域位于多肽氨基末端邻侧，信号肽区域位于前肽区域氨基末端邻侧。

添加允许对宿主细胞生长相关多肽的表达进行调控的调控序列也可能是人们需要的。调控系统的例子是，导致基因表达应答于化学或物理刺激（包括调控化合物的存在）被打开或关闭的那些。原核细胞系统中的调控系统包括 *lac* 和 *trp* 操纵子系统。酵母中，ADH2 系统或 GAL1 系统可以使用。有丝真菌中，TAKA alpha-淀粉酶启动子、*A. niger* 葡糖淀粉酶启动子、*A. oryzae* 葡糖淀粉酶启动子、*A. tubingensis* 内切木聚糖酶（*xlnA*）启动子、*A. niger* 硝酸盐还原酶（*niaD*）启动子、*Trichoderma reesei* 纤维二糖水解酶启动子和 *A. nidulans* 醇和醛脱氢酶（分别为 *alcA* 和 *aldA*）启动子（如 US 5,503,991 所述）可用作为调控序列。调控序列的其它例子是允

许基因扩增的那些。在真核系统中，它们包括二氢叶酸还原酶基因（在甲氨蝶呤存在的情况下被扩增）以及金属硫蛋白基因（在重金属存在的情况下被扩增）。在这些情况下，编码多肽的核酸序列将与可调控的序列可操作地相连。

重要的可能是去除 *creA* 结合位点（如以前在 EP 673 429 中所述，碳代谢物遏制），改变 *pacC* 和 *areA*（针对 pH 和氮调控）。

本发明还涉及重组表达载体，其中包含本发明启动子、编码多肽的编码序列和转录和翻译终止信号。上文所述的多种编码和控制序列都可以一起加入，产生包括一种或多种方便的限制性位点的重组表达载体，所述位点用于允许启动子和/或编码多肽的编码序列在此类位点插入或取代。或者，编码序列和启动子的融合可通过，例如，使用 PCR 进行序列重叠延伸（SOE-PCR）来进行，如 *Gene*. 1989 Apr 15;77(1):51-9. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR "Site-directed mutagenesis by overlap extension using the polymerase chain reaction"所述，或通过用 Gateway™ 克隆系统（Invitrogen）开展克隆来进行。或者，可以通过将编码序列或包含启动子和/或编码序列的 DNA 构建体插入到用于表达的合适载体中，来表达编码序列。在制造表达载体的过程中，编码序列被放置于载体中，使得编码序列与本发明的启动子和一种或多种用于表达的合适控制序列可操作地相连。

重组表达载体可以是任何载体（例如，质粒或病毒），其可方便地经历重组 DNA 流程，并可使得编码序列表达。典型地，对载体的选择将依赖于载体与载体将被引入的宿主细胞之间的兼容性。载体可以是线性的或闭合环状质粒。

载体可以是自主复制载体，即，作为染色体外的整体存在，其复制不依赖于染色体复制，例如，质粒、染色体外元件、微型染色体或人工染色体。对自主复制而言，载体可以包含能使载体在所用的宿主细胞中自主复制的复制原点（origin）。可用于酵母宿主细胞中的复制原点的例子是 2 微米复制原点、ARS1、ARS4、ARS1 和 CEN3 的组合以及 ARS4 和 CEN6 的组合。复制原点可以是一种具有突变的原点，使得其在宿主细胞中的功能

对温度敏感（见，例如 Ehrlich, 1978, Proceedings of the National Academy of Sciences USA 75:1433）。在有丝真菌中自主保持的克隆载体的例子是包含 AMA1 序列的克隆载体。AMA1 是从 *A. nidulans* 分离的 6.0 kb 的基因组 DNA 片段，其能自主保持于 *Aspergillus* 中（见，例如 Aleksenko and Clutterbuck (1997), Fungal Genet. Biol. 21: 373- 397）。

或者，载体可以是如下载体：当引入宿主细胞时，其能整合进基因组，并随着其被整合进的染色体一起复制。此外，可以使用单种载体或质粒，或者一起含有将被引入到宿主细胞基因组的完整 DNA 的两种或多种载体或质粒，或转座子。

本发明的载体优选含有一种或多种选择性标记，其允许对转化的细胞加以容易的选择。宿主可以被至少两种载体共转化，其中一种包含选择标记。选择性标记是一种基因，其产物提供针对生物杀灭剂或病毒的抗性、对重金属的抗性、针对营养缺陷的原营养型等。用于酵母宿主细胞的合适标记是 ADE2、HIS3、LEU2、LYS2、MET3、TRP1 和 URA3。用于有丝真菌宿主细胞的选择性标记包括但不限于，amdS（乙酰胺酶）、argB（鸟氨酸氨基甲酰转移酶）、bar（草胺磷乙酰转移酶）、hygB（潮霉素磷酸转移酶）、niaD（硝酸盐还原酶）、pyrG（乳清苷-5'-磷酸脱羧酶）、sC（硫酸盐腺苷转移酶）、trpC（氨基苯甲酸盐/酯合酶）以及其等同物。赋予针对例如，腐草霉素、潮霉素 B 或 G418 抗性的标记也可使用。优选用于 *Aspergillus* 细胞的是 *A. nidulans* 或 *A. oryzae* 的 amdS 和 pyrG 基因，以及 *Streptomyces hygroscopicus* 的 bar 基因。amdS 标记优选按照 EP 635 574 或 WO 97/0626 所描述的技术来使用。优选的选择标记基因是与 *A. nidulans* *gpdA* 启动子融合的 *A. nidulans* amdS 编码序列（EP635 574）。来自其它有丝真菌的 amdS 基因也可使用（WO 97/06261）。

为整合进宿主细胞基因组，载体可以依赖于启动子序列和/或编码多肽的编码序列或载体的其它任何元件，用于通过同源或非同源重组将载体稳定整合进基因组。或者，载体可以含有额外的核酸序列，用于指导通过同源重组向宿主细胞基因组的整合。额外的核酸序列能使载体整合进宿主细胞基因组染色体上的精确位置。为增加在精确位置上整合的可能性，整合

元件应当优选含有足够数量的、与相应目标序列高度同源的核酸，例如 100 至 1500 个碱基对，优选地，400 至 1500 个碱基对，更优选地，800 至 1500 个碱基对，以及最优选地，至少 2 kb，以增大同源重组的可能性。整合元件可以是与宿主细胞基因组中的目标基因座同源的任意序列。整合元件可以是非编码或编码的核酸序列。为促进定位整合，优选在转化宿主细胞之前对克隆载体进行线性化。线性化优选被进行至下述程度：使得克隆载体的至少一端或优选全部两端的侧翼都有与目标基因座同源的序列。

优选地，克隆载体中与目标基因座同源的整合元件来自高度表达的基因座，这意味着它们来自能在真菌宿主细胞中高水平表达的基因。能高水平表达的基因，即高度表达的基因，在本文中被定义为其 mRNA 占细胞总 mRNA 的至少 0.5% (w/w) 的基因，例如，在诱导条件下的，或者被定义为其基因产物占细胞总蛋白的至少 1% (w/w) 的基因，或者，在分泌出的基因产物的情况下，可分泌至至少 0.1 g/l 的水平（如 EP 357 127 B1 所述）。下面举例给出了大量优选的高度表达的真菌基因：来自 *Aspergilli* 或 *Trichoderma* 的淀粉酶、葡糖淀粉酶、醇脱氢酶、木聚糖酶、甘油醛-磷酸酯脱氢酶或纤维二糖水解酶的基因。就上述目的而言，最优选的高度表达的基因是葡糖淀粉酶基因（优选地，*A. niger* 葡糖淀粉酶基因），*A. oryzae* TAKA 淀粉酶基因，*A. nidulans* *gpdA* 基因，SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5 的基因座，SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5 的 *A. niger* 基因座，或 *Trichoderma reesei* 纤维二糖水解酶基因。

另一方面，载体可通过非同源重组整合进宿主细胞的基因组。

可以有超过一个拷贝的、编码多肽的核酸序列被插入到宿主细胞中，以增加基因产物的产生。这可以通过下述方法来实现：优选地，将 DNA 序列的拷贝整合进其基因组，更优选地，将 DNA 序列整合到高度表达的基因座，优选地，整合到葡糖淀粉酶基因座或 SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5 的基因座。或者，这可以通过下述方法来实现：将可扩增的选择性标记基因包括到核酸

序列，在该处，含有选择标记基因扩增拷贝以及核酸序列的额外拷贝的细胞可被选择出来，这是通过在有合适的选择试剂存在的情况下对细胞加以培养来实现的。为进一步增加将被过量表达的 DNA 序列的拷贝数，如 WO98/46772 中所述的基因转化技术可以使用。

用于将上述元件连接起来以构建本发明的重组表达载体的流程是本领域技术人员公知的（见，例如前述 Sambrook et al., 1989）。

本发明还涉及重组宿主细胞，其中包含与编码多肽的编码序列可操作地相连的本发明的启动子序列，所述细胞可有利地用于生产多肽。包含与编码多肽的编码序列可操作地相连的本发明的启动子的载体被引入宿主细胞，使得该载体被保持为染色体成分或如前文所述的自主复制型染色体外载体。术语“宿主细胞”包括：由于复制期间发生的突变，故而与亲本细胞不相同的、亲本细胞的任何后代。对宿主细胞的选择将在很大程度上依赖于编码多肽的基因及其来源。

本发明还涉及重组宿主细胞，其中包含多于一种本发明的启动子 DNA 序列，每种启动子都与编码多肽的编码序列可操作地相连。此类宿主细胞可以有利地用于对至少一种多肽的重组生产。优选地，载体中存在至少一种启动子及其相连的编码序列。载体被引入到宿主细胞中，使得其被保持为染色体成分或如前文所述的自主复制型染色体外载体。

根据另一种优选的实施方式，宿主细胞被用于生产特定的初级或次级代谢产物，例如，（beta-内酰胺）抗生素、维生素或类胡萝卜素。

宿主细胞可以是有用于本发明的方法中的任何真菌细胞。本文中使用的“真菌”包括 phyla Ascomycota、Basidiomycota、Chytridiomycota 和 Zygomycota（如 Hawksworth et al., Ainsworth and Bisby's Dictionary of The Fungi, 8th edition, 1995, CAB International, University Press, Cambridge, UK 所定义的）以及 Oomycota（如 Hawksworth et al., 1995, supra, page 171 中提到的）以及所有的有丝分裂孢子真菌（前述 Hawksworth et al., 1995）。

在一种优选的实施方式中，真菌宿主细胞是酵母细胞。用于本文中的“酵母”包括产子囊酵母（Endomycetales）、担孢子产芽孢（basidiosporogenous）酵母和属于 Fungi Imperfecti 的酵母

(*Blastomycetes*)。因为对酵母的分类未来可能会有改变,就本发明的目的而言,酵母将如 *Biology and Activities of Yeast* (Skinner, F. A., Passmore, S. M., and Davenport, R. R., eds, Soc. App. Bacteriol. Symposium Series No. 9, 1980)中所定义。

在一种更优选的实施方式中,酵母宿主细胞是 *Candida*、*Hansenula*、*Kluyveromyces*、*Pichia*、*Saccharomyces*、*Schizosaccharomyces* 或 *Yarrowia* 细胞。

在一种最为优选的实施方式中,酵母宿主细胞是 *Saccharomyces carlsbergensis*、*Saccharomyces cerevisiae*、*Saccharomyces diastaticus*、*Saccharomyces douglasii*、*Saccharomyces kluyveri*、*Saccharomyces norbensis* 或 *Saccharomyces oviformis* 细胞。在另一种最为优选的实施方式中,酵母宿主细胞是 *Kluyveromyces lactis* 细胞。在另一种最为优选的实施方式中,酵母宿主细胞是 *Yarrowia lipolytica* 细胞。

在另一种优选的实施方式中,真菌宿主细胞是有丝真菌细胞。“有丝真菌细胞”包括 *Eumycota* 和 *Oomycota* 亚门的所有有丝形式(如前述 *Hawksworth et al.*, 1995 所定义的)。有丝真菌具有如下特征:由几丁质、纤维素、葡聚糖、壳聚糖、甘露聚糖和其它复杂多糖构成的菌丝体壁。营养生长是通过菌丝伸长进行的,碳代谢是专性需氧的。与之相对,酵母,例如 *Saccharomyces cerevisiae* 的营养生长是通过单细胞原植体出芽来进行的,碳代谢可以是发酵的。

在一种优选的实施方式中,有丝真菌宿主细胞是下述物种的细胞,但不限于此: *Acremonium*、*Aspergillus*、*Fusarium*、*Humicola*、*Mucor*、*Myceliophthora*、*Neurospora*、*Penicillium*、*Thielavia*、*Tolyocladium* 或 *Trichoderma*。

在一种最优选的实施方式中,有丝真菌细胞是 *Aspergillus awamori*、*Aspergillus foetidus*、*Aspergillus japonicus*、*A. nidulans*、*A. niger* 或 *A. oryzae* 的细胞。在另一种最优选的实施方式中,有丝真菌细胞是 *Fusarium bactridioides*、*Fusarium cerealis*、*Fusarium crookwellense*、*Fusarium culmorum*、*Fusarium graminearum*、*Fusarium graminum*、*Fusarium*

heterosporum、*Fusarium negundi*、*Fusarium oxysporum*、*Fusarium reticulatum*、*Fusarium roseum*、*Fusarium sambucinum*、*Fusarium sarcochromum*、*Fusarium sporotrichioides*、*Fusarium sulphureum*、*Fusarium torulosum*、*Fusarium trichothecioides* 或 *Fusarium venenatum* 的细胞。在另一种最优选的实施方式中，有丝真菌细胞是 *Humicola insolens*、*Humicola lanuginosa*、*Mucor miehei*、*Myceliophthora thermophila*、*Neurospora crassa*、*Penicillium purpurogenum*、*Thielavia terrestris*、*Trichoderma harzianum*、*Trichoderma koningii*、*Trichoderma longibrachiatum*、*Trichoderma reesei* 或 *Trichoderma viride* 的细胞。

可以用本身已知的技术，通过下述方法来转化真菌细胞，所述方法涉及到原生质体形成、原生质体转化以及细胞壁再生。合适的用于转化 *Aspergillus* 宿主细胞的方法被描述于 EP 238 023 and Yelton et al., 1984, Proceedings of the National Academy of Sciences USA 81 : 1470-1474 中。合适的、用 *Agrobacterium tumefaciens* 转化 *Aspergillus* 和其它有丝真菌宿主细胞的方法被描述于，例如 Nat Biotechnol. 1998 Sep;16(9):839-42. Erratum in: Nat Biotechnol 1998 Nov;16(11):1074. *Agrobacterium tumefaciens*-mediated transformation of filamentous fungi, de Groot MJ, Bundock P, Hooykaas PJ, Beijersbergen AG. Unilever Research Laboratory Vlaardingen, The Netherlands 中。合适的用于转化 *Fusarium* 的种的方法被描述于 Malardier et al., 1989, Gene 78: 147-156 和 WO 96/00787 中。可以用 Becker and Guarente, In Abelson, J. N. and Simon, M. I., editors, Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology, Volume 194, pp 182-187, Academic Press, Inc., New York、Ito et al., 1983, Journal of Bacteriology 153: 163 和 Hinnen et al., 1978, Proceedings of the National Academy of Sciences USA 75: 1920 所述的流程来转化酵母。

本发明还涉及在宿主细胞中表达编码序列的方法。所述方法包括如下步骤：

(a) 提供一种 DNA 构建体，所述构建体包含本发明的启动子 DNA 序列以及上述编码序列，

(b) 用所述的 DNA 构建体去转化合适的宿主细胞，以及

(c) 在所述启动子 DNA 序列的控制下表达编码序列。

本发明还涉及一种方法，用于在合适的真菌宿主中生产编码序列所编码的多肽，所述编码序列处于本发明启动子的控制下。所述方法包括如下步骤：

(a) 提供一种 DNA 构建体，所述构建体包含本发明的启动子 DNA 序列以及编码上述次级代谢产物的生产中所涉及的酶的序列，

(b) 用所述的 DNA 构建体去转化合适的真菌宿主细胞，以及

(c) 在有益于表达所述多肽的合适培养条件下培养所述合适的真菌宿主，

(d) 从培养物中回收所述多肽。

本发明还涉及一种方法，用于在合适的宿主中生产次级代谢产物，所述方法包括：

(a) 提供一种 DNA 构建体，所述构建体包含本发明的启动子 DNA 序列以及编码上文定义的酶的序列，

(b) 用所述的 DNA 构建体去转化合适的真菌宿主细胞，以及

(c) 在有益于生产所述次级代谢产物的合适培养条件下培养所述合适的真菌宿主，

(d) 从培养物中回收所述多肽。

在本发明的生产方法中，细胞被培养于适合用于用本领域已知的方法来生产多肽或代谢产物的营养培养基中。例如，可通过在合适的培养基中于允许编码序列得以表达和/或多肽得以被分离的条件下进行摇瓶培养、实验室或工业发酵罐中的小规模或大规模发酵（包括连续、分批、补料分批或固态发酵）来培养细胞。培养发生于包含碳和氮源以及无机盐的合适营养培养基中，使用本领域已知的流程来进行培养。合适的培养基可以从厂商处购得，或者按照公开的组分配方来制备（例如，在 American Type Culture Collection 的目录中）。如果多肽或代谢产物被分泌到营养培养基中，可以直接从培养基中回收多肽或代谢产物。如果多肽或代谢产物不分泌，可以从细胞裂解物中对其进行回收。

可以用本领域已知的、特定用于多肽的方法来探测多肽。这些探测方法可以包括：使用特异性抗体、酶产物的形成或酶底物的消失。

可以用本领域已知的方法来回收得到的多肽或代谢产物。例如，可以通过传统流程，其包括但不限于，离心、过滤、萃取、喷雾干燥、蒸发或沉淀，从营养培养基中回收多肽或代谢产物。

可以用本领域已知的多种流程来纯化多肽，其包括但不限于，色谱（例如，离子交换、亲和、疏水、层析聚焦和尺寸排除）、电泳流程（例如，制备性等电聚焦）、差异溶解（例如，硫酸铵沉淀）、SDS-PAGE 或萃取（见，例如 *Protein Purification*, J. -C. Janson and Lars Ryden, editors, VCH Publishers, New York, 1989）。

本发明还涉及 DNA 构建体，其用于改变真菌宿主细胞内源的、编码多肽的编码序列的表达。所述构建体可以含有用于改变内源基因表达所需的最小数量的组分。

在一种实施方式中，核酸构建体优选含有（a）定位序列，（b）本发明的启动子 DNA 序列，（c）外显子以及（d）剪接供体位点。将核酸构建体引入细胞后，该构建体通过同源重组整合进细胞基因组的内源基因位点。定位序列引导着元件（a）-（d）整合进内源基因，使得元件（b）-（d）与内源基因可操作地相连。

在另一种实施方式中，核酸构建体优选含有（a）定位序列，（b）本发明的启动子 DNA 序列，（c）外显子，（d）剪接供体位点，（e）内含子以及（f）剪接受体位点，其中，定位序列引导着元件（a）-（f）的整合，使得元件（b）-（f）与内源基因可操作地相连。但是，所述构建体还可以含有额外元件，例如选择性标记。前文已描述过可以使用的选择性标记。

在两种实施方式中，这些元件的引入导致了新的转录单元的产生，其中，内源基因的表达被改变。大体上，新的转录单元是定位构建体引入的序列和内源基因的融合产物。在一种实施方式中，其中内源基因被改变，基因即被活化。在该实施方式中，通过插入调控序列，同源重组被用于置换、打断或失活在正常情况下与亲本细胞内源基因相连的调控区域，这导

致：较之在相应亲本细菌中的情况，基因以更高的水平被表达。

定位序列可以位于内源基因内，紧邻该基因，在上游基因中，或在内源基因上游并与其有一段距离。可以使用一种或多种定位序列。例如，环状质粒或 DNA 片段优选使用单种定位序列，而线状质粒或 DNA 片段优选使用两种定位序列。

构建体还含有内源基因的一个或多个外显子。外显子被定义为下述 DNA 序列：其会被拷贝为 RNA，并存在于成熟 mRNA 分子中，使得外显子序列与内源基因的编码序列处于同一读码框。外显子可选地可以含有编码一个或多个氨基酸和/或部分编码氨基酸的 DNA。或者，外显子含有相应于 5'非编码区域的 DNA。当外源外显子或外显子编码一个或多个氨基酸和/或氨基酸的一部分时，核酸构建体被设计为：转录和剪接后，读码框与内源基因的编码区域处于同一读码框，使得从第二个外显子获得的 mRNA 的部分的合适读码框不被改变。构建体的剪接供体位点指导从一个外显子到另一个外显子的剪接。典型地，第一个外显子处于第二个外显子的 5'侧，重叠且侧翼于第一个外显子 3'侧的剪接供体位点能识别在第二个外显子 5'侧侧翼于第二个外显子的剪接受体位点。剪接受体位点，和剪接供体位点一样，都是能指导从一个外显子到另一个外显子的剪接的序列。剪接装置与剪接供体位点联合作用，使用剪接受体位点来实现内含子的去除。

用于改变给定 DNA 序列表达的优选策略包括：缺失给定的 DNA 序列和/或用经修饰的启动子 DNA 序列，例如本发明的启动子，去置换给定 DNA 序列的内源启动子序列。优选通过 EP 0 357 127 所述的基因置换技术来进行缺失和置换。优选使用 *amdS* 基因作为选择标记基因按照 EP 635 574 所述来进行对基因和/或启动子序列的特异性缺失。通过 EP 635 574 所述的在氟乙酰胺培养基上进行的反向选择，得到的菌株是不含选择标记的，其可用于进一步的基因修饰。

作为替代，或与上面提到的技术组合，还可以使用基于在 *E. coli* 中进行粘粒体内重组的技术，如 A rapid method for efficient gene replacement in the filamentous fungus *A. nidulans* (2000) Chaveroche, M-K., Ghico, J-M. and d'Enfert C; Nucleic acids Research, vol 28, no 22 所述。该技术可用于其它有

丝真菌，例如 *A. niger*。本文中描述并要求保护的本发明不仅限于本文所公开的特定实施方式的范围，因为这些实施方式仅被用于对本发明的若干方面加以阐述。任何等同的实施方式都被包括在本发明的范围内。事实上，基于前述的说明书，除本文展示和描述过的之外，本发明的多种改良形式将是本领域技术人员显而易见的。此类改良形式也将在所附权利要求的范围内。在有争论的情况下，本发明的公开文本，包括定义，将起控制作用。

本文提到的所有专利和出版物，包括这些专利和出版物内所公开的所有序列和方法，都通过引用被明确并入本文。这些专利和公开文本包括：EP 357 127、EP 673 429、EP 635 574、WO 97/06261、WO 98/46772、WO 94/04673。

实施例

实验信息

菌株

WT1：该 *A. niger* 菌株被用作为野生型菌株。该菌株被保藏于 CBS 研究所，保藏号为 CBS 513.88。

WT2：该 *A. niger* 菌株是包含编码葡糖淀粉酶的基因 (*glaA*) 缺失的 WT1 菌株。WT2 是按照 EP 0 635 574 所述使用“MARKER-GENE FREE”方法构建的。在该专利中，广泛地描述了如何在 CBS 513.88 基因组中缺失 *glaA* 特异性 DNA 序列的方法。该流程产生了不含标记基因的 *glaA* 重组 *A. niger* CBS513.88 菌株，该菌株最终并不具有任何外源 DNA 序列。

葡糖淀粉酶活性检验

按照 WO 98/46772 所述，使用对硝基苯 α -D-吡喃葡萄糖苷 (Sigma) 来测定葡糖淀粉酶活性。

实施例 1 构建 DNA 构建体，其中包含与编码序列可操作地相连的本发明的启动子

本实施例描述了对在本发明启动子控制下的表达构建体的构建。此处使用的编码序列或报道构建体是编码 *A. niger* 葡糖淀粉酶的 *glaA* 基因。葡糖淀粉酶被用作能测量本发明启动子活性的报道酶。

1.1 对整合型葡糖淀粉酶表达载体 (pGBTOPGLA) 的描述

将葡糖淀粉酶启动子和来自 *A. niger* 的葡糖淀粉酶编码基因 *glaA* 克隆进表达载体 pGBTOP-8 (其被描述于 WO99/32617 中)。按照已知流程和常规克隆技术来进行克隆, 得到质粒 pGBTOPGLA (见图 1)。大体上, 该表达载体包含葡糖淀粉酶启动子、编码序列和终止子区域, 侧翼有 *E. coli* 载体中的 3' 和 3'' *glaA* 定位序列。

1.2 构建带有多克隆位点 MCS 的整合型葡糖淀粉酶表达载体 (pGBTOPGLA-2)

使用 SEQ ID NO 6 示出的寡核苷酸

5'-

ATgCggCCgCCTCgAgTTAATTAAggCCAaggCCggCCggCgCgCCTCAgCAA
TgTCgTTCCgA-3'

和 SEQ ID NO 7 示出的

5'-AGCCATTGACTTCTTCCCAG-3'

以及 1 ng pGBTOPGLA 载体作为模板, 产生了含有 *glaA* 编码序列一部分的 PCR 片段。用 *XhoI* 和 *BglII* 来消化该片段, 将其引入到 *XhoI* 和 *BglII* 消化过的载体 pGBTOPGLA 中, 得到载体 pGBTOPGLA-2 (见图 2)。通过序列分析, 证实了包含 MCS 和 *glaA* 编码序列一部分的、引入的 PCR 片段的序列。

1.3 构建带有本发明的启动子的整合型表达载体 (pGBTOPGLA-3), 所述启动子与葡糖淀粉酶编码序列可操作地相连

对菌株 CBS513.88 的基因组 DNA 进行测序和分析。使用下表展示的寡核苷酸组合物 (寡核苷酸 SEQ ID NO:), 用菌株 CBS513.88 的基因组

DNA 作为模板，通过 PCR 扩增将合适的限制性位点连到本发明的启动子上。被示为 SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4 或 SEQ ID NO: 5 的序列包含大约 2 kb 的获得的片段序列，如下表所示。用 *AscI* 和 *XhoI* 去消化全部五段得到的片段，将它们引入到用 *AscI* 和 *XhoI* 消化过的载体 pGBTOPGLA-2 中，得到载体 pGBTOPGLA-4、pGBTOPGLA-5、pGBTOPGLA-7、pGBTOPGLA-9 或 pGBTOPGLA-10，如下表所示（载体名）。在图 3 中，可以找到用于说明全部五种载体的布局结构的图。通过序列分析，证实了多种包含本发明启动子的引入的 PCR 片段的序列。

寡核苷酸 SEQ ID NO:	寡核苷酸 SEQ ID NO:	载体名	被克隆进载体的启动 子的 SEQ ID NO:
8	9	pGBTOPGLA-7	1
10	11	pGBTOPGLA-9	2
12	13	pGBTOPGLA-4	3
14	15	pGBTOPGLA-10	4
16	17	pGBTOPGLA-5	5

实施例 2 用 DNA 构建体转化过的真菌宿主细胞

为在 WT2 中引入 pGBTOPGLA、pGBTOPGLA-4、pGBTOPGLA-5、pGBTOPGLA-7、pGBTOPGLA-9 或 pGBTOPGLA-10 载体，按照 WO98/46772 和 WO99/32617 所述，进行转化和随后的转化子选择。原理上，在用 *NotI* 消化之后，分离出所有载体的线性 DNA，与含有 *amdS* 选择性标记基因的载体（被命名为 pGBAAS-1，按照 EP 635574 所述来构建）一起进行共转化。两种载体都包含与 *A. niger* 宿主菌株 *glaA* 基因座同源的两种 DNA 结构域，以指导向 WT2 中被截短的 *glaA* 基因座上的定位。按照标准流程，在乙酰胺培养基上对转化子进行选择，并对菌落加以纯化。将孢子涂布到氟乙酰胺培养基上，选出丢失了 *amdS* 标记的菌株。针对在 *glaA* 基因座上的整合和拷贝数，对生长的菌落加以诊断分析。选出具有相似评估拷贝数的 pGBTOPGLA、pGBTOPGLA-4、pGBTOPGLA-5、

pGBTOPGLA-7、pGBTOPGLA-9 或 pGBTOPGLA-10 的转化子。优选地，选出单拷贝的少数转化子（1A、1B、1C）以及，可能地，具有多个拷贝的转化子（2A）。

此外，选择性标记基因和本发明启动子控制的目标基因已在同一构建体上。关于此的一个例子显示于图 4 中。

实施例 3 在真菌宿主细胞中生产葡糖淀粉酶多肽，所述多肽由 *glaA* 编码序列编码，该序列处于本发明启动子控制下

用按照上文所述的选出的大量 WT2 的转化子以及菌株 WT1 和 WT2 来进行摇瓶实验，其中，使用 500 ml 带挡板摇瓶，于 34°C 和 170 rpm 于培养摇床上，按照 EP 635 574，在 100 ml 培养基中来进行该实验。4 和 5 天发酵之后，取样品来测定葡糖淀粉酶活性，如上文所述。将葡糖淀粉酶活性标准化为 WT1 在第 4 天的活性。图 5 展示了 WT1、WT2 和针对 pGBTOPGLA、pGBTOPGLA-5、pGBTOPGLA-7 和 pGBTOPGLA-10 的选出的大量转化子的标准化活性。

从对用作报道的葡糖淀粉酶的测量活性可以得出结论，本发明提供了用于在真菌细胞中高水平表达目标基因的强启动子。同样，本发明提供了在真菌细胞中高水平表达目标基因的替代性启动子或额外启动子。

实施例 4 构建包含本发明启动子的启动子置换构建体 pGBDEL-PGLAA

为改变给定基因在宿主细胞中的表达水平，本发明的启动子可置换所述给定基因的内源启动子。在本实施例中，本发明的启动子置换了真菌宿主细胞中编码葡糖淀粉酶的 *glaA* 基因的启动子。实施例 4、5 和 6 描述了该过程中使用的大量不同步骤。

按照已知的原理来设计用于葡糖淀粉酶启动子的置换载体，按照常规克隆流程来对其进行构建（见图 6）。大体上，*glaA* 启动子置换载体 pGBDEL-PGLAA 包含 *glaA* 启动子序列（其将通过在预定基因组基因座上进行同源重组被 SEQ ID NO: 4 所包含的本发明启动子所置换）的大约 1000 bp 的侧翼区域。此处使用的侧翼区域（见图 6）是 *glaA* 启动子的 5'

上游区域和 *glaA* 编码序列的一部分。此外，置换载体含有 *A. nidulans* 双向 *amdS* 选择标记，其位于正向重复之间。用于本实施例的正向重复是 *glaA* 编码序列的一部分。对上述缺失载体的常用设计在 EP635574 和 WO 98/46772 中已有描述。

实施例 5 在真菌宿主细胞中用本发明的启动子去置换 *glaA* 启动子

分离出用 *NotI* 消化过的缺失载体 pGBDEL-PGLAA 的线性 DNA，将其用于转化 WT 1 (CBS513.88)。该线性 DNA 能整合进基因组的 *glaA* 基因座，由此用含有 *amdS* 和本发明启动子的构建体取代 *glaA* 启动子区域（见图 7）。按照标准流程在乙酰胺培养基上选择转化子，对菌落进行纯化。通过 PCR，针对 *glaA* 基因座处的整合，对生长的菌落加以诊断分析。通过具有本发明启动子所特有的大小的条带的扩增，以及 *glaA* 启动子特有的条带的丢失，可对 *glaA* 启动子的缺失加以探测。将孢子涂布到氟乙酰胺培养基上，选出丢失了 *amdS* 标记的菌株。用 Southern 分析，针对葡糖淀粉酶启动子的正确缺失，以及本发明启动子（由 SEQ ID NO: 4 包含）的正确置换，对候选菌株加以检验。菌株 dPGLAA 被选作为 *glaA* 启动子被本发明启动子置换、并具有重建的功能性 *glaA* 编码序列的代表菌株（见图 7）。

实施例 6 在真菌宿主细胞中，生产由 *glaA* 编码序列编码的葡糖淀粉酶多肽，该编码序列处于在本发明的置换启动子控制下

将选出的 dPGLAA (WT1 的正确 pGBDEL-PGLAA 转化子，在实施例 5 中分离的) 和菌株 WT1 用于进行摇瓶实验，其中，使用 500 ml 带挡板摇瓶，于 34°C 和 170 rpm 于培养摇床上，按照 EP 635 574，在 100 ml 培养基中进行该实验。其它条件和活性测量方法如实施例 3 中所述。在两种发酵天数取样测量时，选出的 WT1 的 pGBDEL-PGLAA 转化子中，葡糖淀粉酶活性较之针对 WT1 测得的活性升高（数据未示出）。

实施例 7 在真菌宿主细胞中加入额外的、处于本发明启动子控制下 *glaA*

基因

为改变给定基因在宿主细胞中的表达水平，可将与本发明启动子可操作地相连的所述基因的多个额外拷贝加入到内源给定的基因中。在本实施例中，本发明的启动子（SEQ ID NO: 4 包含的，与 *glaA* 编码序列可操作地相连）被引入到真菌宿主细胞中内源存在的 *glaA* 基因（编码葡糖淀粉酶）旁边。实施例 7 和 8 描述了本过程中的大量步骤。

分离出图 8 所示的环状构建体，用其转化 WT1（CBS513.88）。该线性 DNA 能整合到基因组中 *glaA* 编码序列处，由此加入处于本发明启动子（在选择性标记 *amdS* 旁边）控制下的第二个 *glaA* 基因（见图 8）。按照标准流程，在乙酰胺培养基上选择转化子，对菌落加以纯化。通过 PCR，针对在 *glaA* 基因座处的整合，对生长的菌落加以诊断分析。通过 PCR 和 Southern 分析，探测到在 *glaA* 基因座处的整合。菌株 P2GLAA 被选为：具有整合到 *glaA* 基因座上的、本发明启动子控制下的至少第二个 *glaA* 基因的代表菌株。

实施例 8 在真菌宿主细胞中，生产由处于本发明启动子和内源 *glaA* 启动子控制下的 *glaA* 编码序列编码的葡糖淀粉酶多肽

实施例 7 中分离出的选出的 P2GLAA 菌株，和菌株 WT1 被用于进行摇瓶实验，该实验在 100 ml 培养基中按照实施例 3 所述来进行。4 和 6 天的发酵后，取样品来测定葡糖淀粉酶活性。四或五天发酵后，较之对 WT1 进行测量得到的结果，选出的 WT1 P2GLAA 转化子中葡糖淀粉酶活性增加。作为报道的葡糖淀粉酶增加的活性表明，本发明的启动子提供了在真菌细胞中进一步增加已在强启动子下表达的目标基因表达的方法。

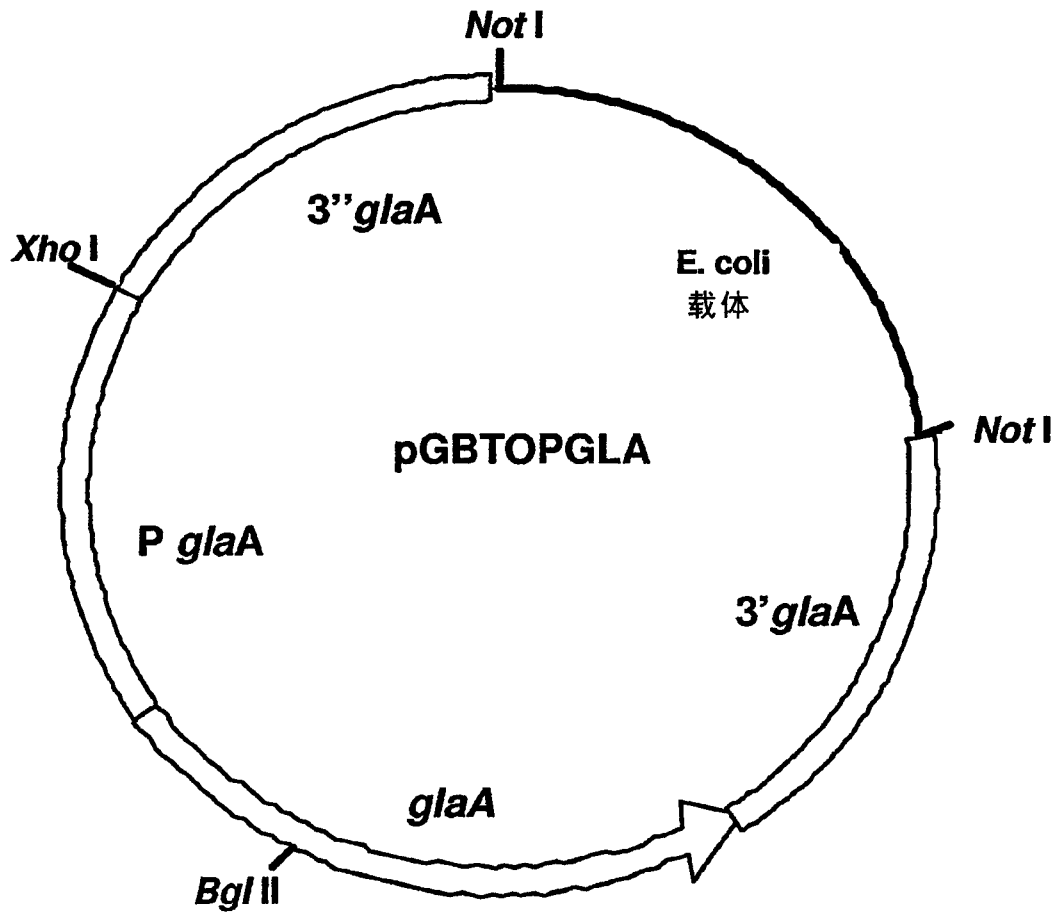


图1

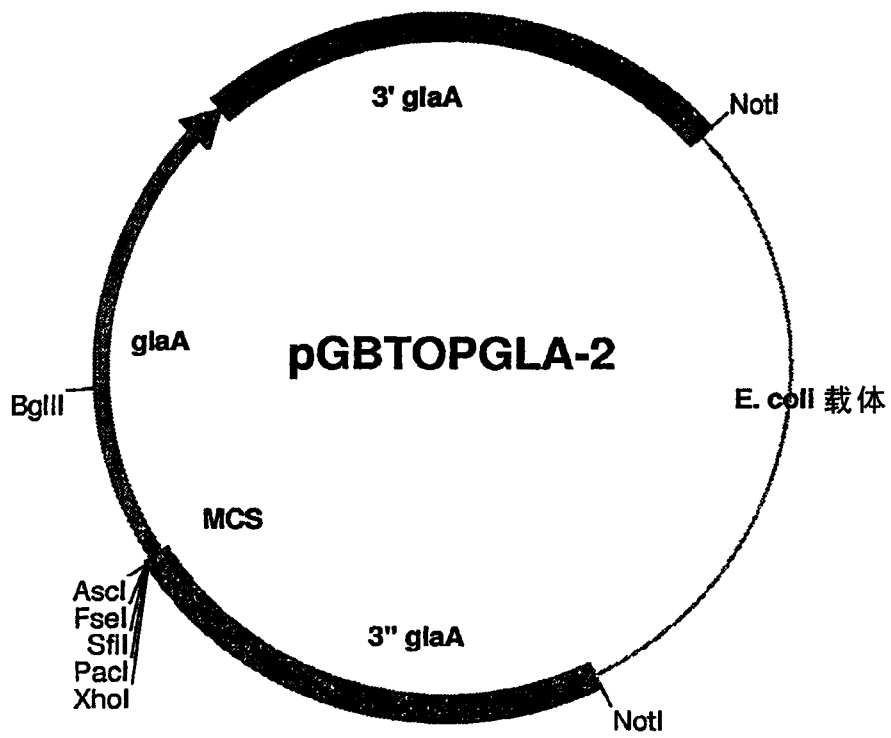


图2

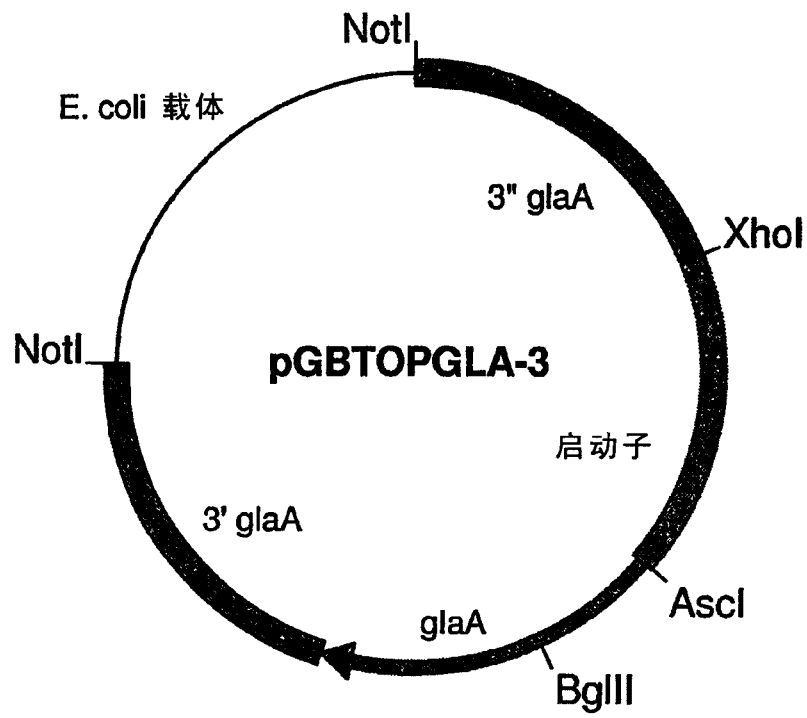


图3

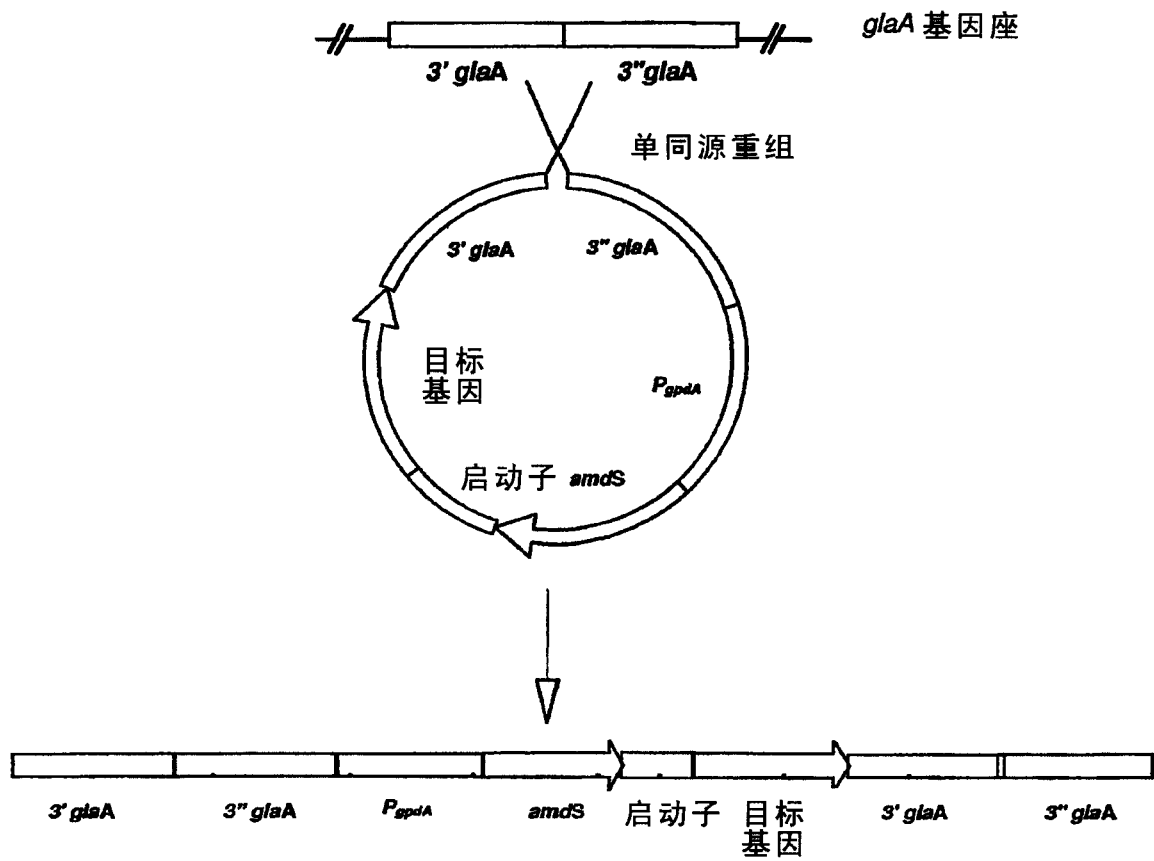


图4

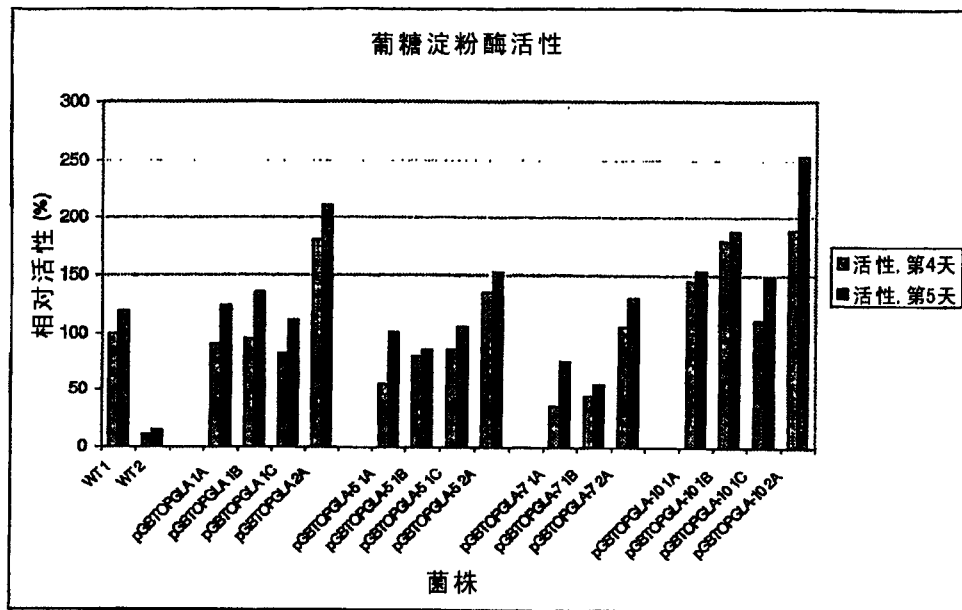


图5

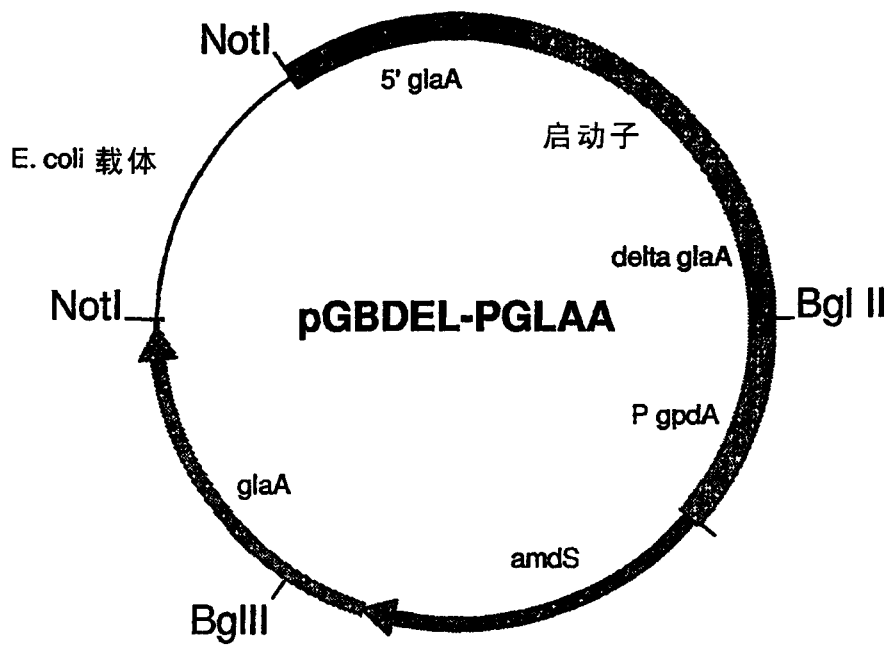


图6

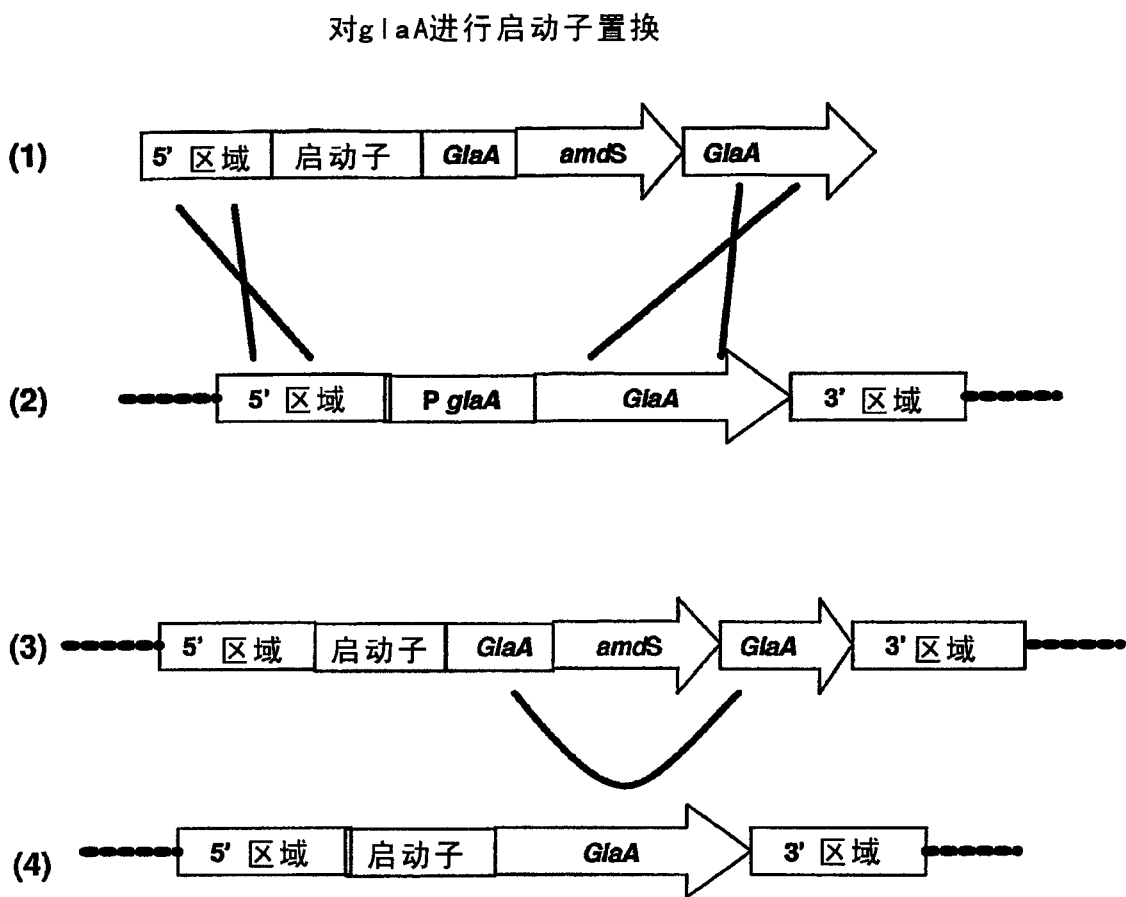


图7

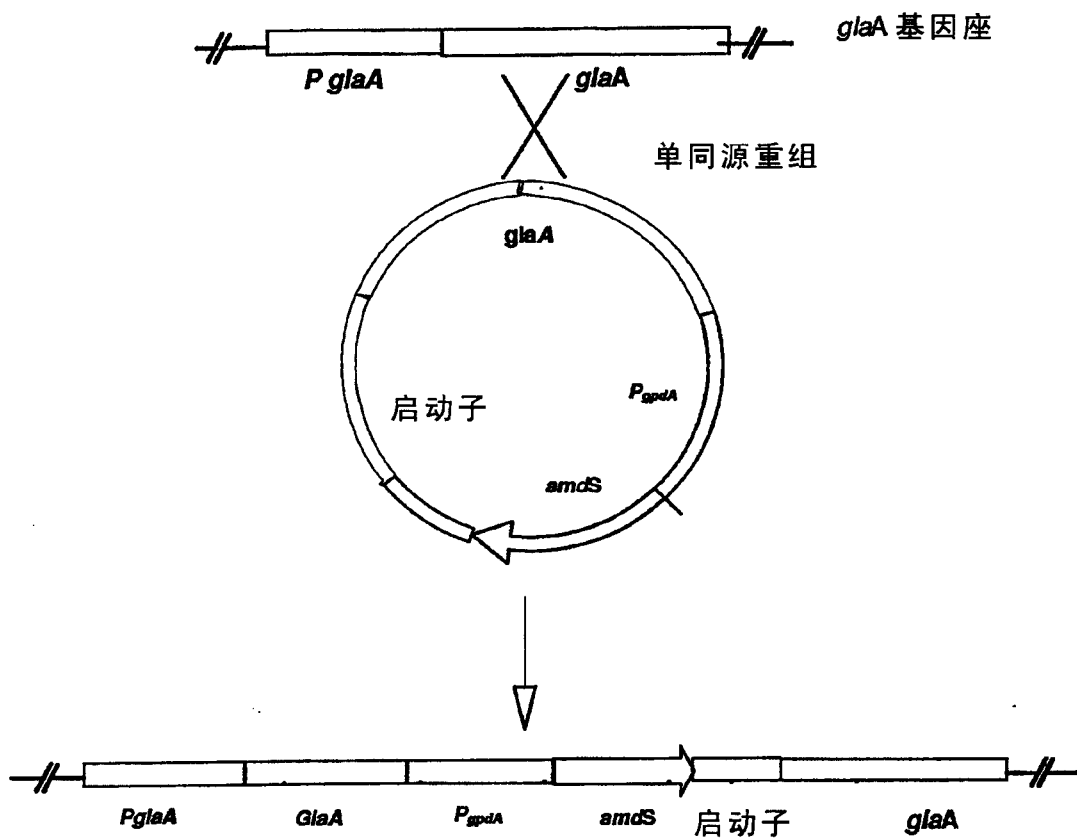


图8