

[19] 中华人民共和国国家知识产权局



[12] 发明专利说明书

专利号 ZL 95193631. X

[51] Int. Cl.

C12N 15/80 (2006.01)

C12N 1/15 (2006.01)

C07K 14/38 (2006.01)

C07H 21/04 (2006.01)

C12N 15/67 (2006.01)

[45] 授权公告日 2006 年 4 月 26 日

[11] 授权公告号 CN 1253574C

[22] 申请日 1995.6.19 [21] 申请号 95193631. X

[30] 优先权

[32] 1994.6.17 [33] DK [31] 0717/94

[86] 国际申请 PCT/DK1995/000254 1995.6.19

[87] 国际公布 WO1995/035385 英 1995.12.28

[85] 进入国家阶段日期 1996.12.16

[71] 专利权人 诺沃奇梅兹有限公司

地址 丹麦巴格斯瓦尔德

[72] 发明人 T·克里斯坦森 M·J·海尼斯

审查员 孙广秀

[74] 专利代理机构 中国国际贸易促进委员会专利
商标事务所

代理人 李 瑛

权利要求书 2 页 说明书 21 页 附图 6 页

[54] 发明名称

来自米曲霉的 areA 基因和其中 areA 基因已被
修饰的真菌

[57] 摘要

本发明涉及不产生蛋白酶的真菌。本发明的
真菌可作为宿主用于生产可被通常产生的蛋白酶降
解的蛋白质，因而本发明还包括利用本发明的真菌
以高产率生产所研究的蛋白质的方法。本发明也
包括生产这样的真菌的方法以及用于这些方法的
DNA 构建体。

1. 一种真菌，其选自米曲霉、黑曲霉、泡盛曲霉、海枣曲霉、日本曲霉、臭曲霉、*T.reesei*、*T.harzianum*、*H.insulens*、柔毛腐质霉、禾木科镰孢、腐皮镰孢和产黄青霉，其中由重组 DNA 技术得到的 *areA* 基因已得到修饰，使得其不能以提供功能性 *areA* 基因产物的方式表达，其中所述功能性 *areA* 基因产物的失活作用通过全部或部分 *areA* 基因的缺失、干扰调控 *areA* 基因本身表达的表达信号的调控作用、使用反义技术或在 *areA* 基因中插入额外的 DNA 而得到。

2. 生产依据权利要求 1 之真菌的方法，其中所述失活作用已通过 *AreA* 基因的缺失而获得，该方法包括：

(i) 克隆真菌的 *areA* 基因，其中真菌选自米曲霉、黑曲霉、泡盛曲霉、海枣曲霉、日本曲霉、臭曲霉、*T.reesei*、*T.harzianum*、*H.insulens*、柔毛腐质霉、木材科镰孢、腐皮镰孢和产黄青霉，

(ii) 生产含有 *areA* 基因的 DNA 构建体，其中基因内在部分已被取代、删除、或者已插入额外的 DNA，

(iii) 用该构建体转化所述真菌，和

(iv) 选择出为 *areA*⁻ 的转化体。

3. 生产依据权利要求 1 之真菌的方法，其中所述失活作用已利用反义技术而获得，该方法包括

(i) 构建表达质粒，该质粒可引起互补于转录自 *areA* 基因的 mRNA 的 RNA 分子的合成，

(ii) 用所述表达质粒及适宜的标记物来转化宿主真菌，所述表达质粒和标记物既可以是在独立的质粒上，也可以在同一质粒上，并且所述真菌选自米曲霉、黑曲霉、泡盛曲霉、海枣曲霉、日本曲霉、臭曲霉、*T.reesei*、*T.harzianum*、*H.insulens*、柔毛腐质霉、木材科镰孢、腐皮镰孢和产黄青霉，

(iii) 利用所述标记物选择转化体，

(iv) 从所选择的转化体筛选出 *AreA* 基因产物合成降低的菌株。

4. 一种生产所需基因产物的方法，该方法包括在适宜的生长培养基中，在适宜的条件下培养依据权利要求 1 的真菌，对所需的基因产物进行回收并提纯，其中所述真菌选自米曲霉、黑曲霉、泡盛曲霉、海枣曲霉、日本曲霉、臭曲霉、*T.reesei*、*T.harzianum*、*H.insulens*、柔毛腐质霉、木材科镰孢、腐皮镰孢和产黄青霉。

5. 一种生产所需基因产物的方法，该方法包括使依据权利要求 1 的真菌被转化，以功能性方式将编码所需基因产物的 DNA 序列整合到真菌的基因组中，在适宜的生长培养基中和适宜的条件下培养真菌，回收所需的基因产物并提纯，其中所述真菌选自米曲霉、黑曲霉、泡盛曲霉、海枣曲霉、日本曲霉、臭曲霉、*T.reesei*、*T.harzianum*、*H.insulens*、柔毛腐质霉、木材科镰孢、腐皮镰孢和产黄青霉。

6. 权利要求 5 的方法，其中真菌中通过重组 DNA 技术得到的 *areA* 基因已被修饰，使其不能以提供功能性 *areA* 基因产物的方式被表达，其中所述功能性 *areA* 基因产物的失活作用是通过全部或部分 *areA* 基因的缺失、干扰调控 *areA* 基因本身表达的表达信号的调控作用、使用反义技术或在 *areA* 基因中插入额外的 DNA 而得到。

来自米曲霉的 areA 基因和其中 are A 基因已被修饰的真菌

发明领域

本发明涉及不产生蛋白酶的真菌。本发明的真菌可作为宿主用于生产可被通常产生的蛋白酶降解的蛋白质,因而本发明还包括利用本发明的真菌以高产率生产所研究的蛋白质的方法。本发明也包括产生这些真菌的方法以及用于这些方法的 DNA 构建体。

发明背景

真菌特别是丝状真菌被广泛应用于商业领域,因为他们具有分泌极大量蛋白质的能力。

在曲霉属的丝状真菌种类中,在生产内源蛋白质的商业用途方面已有很长的历史,近来也用于生产异源蛋白质。

用于生产蛋白质的大多数微生物的一个缺点在于内在生成蛋白酶,它会使所研究的蛋白质产物因蛋白水解而发生降解。

人们已经设想出避免这个缺点的多种方法。在另外一些解决方法中,有人提议可删除或破坏多种蛋白酶的编码基因。遗憾的是,真菌可产生大量的蛋白酶,这使该方法多少有些不够实用。

因而我们非常需要一种不产生或只产生极少量蛋白酶的丝状真菌菌株。

多年来人们知道在无冠构巢曲霉中介导氮代谢物阻抑的调控基因 areA 可影响胞外蛋白酶的产生 [Arst & Cove, molec, gen. Genet. 126, (1973)111 - 141]。

人们对来自无冠构巢曲霉的 areA 基因进行了克隆 [Caddick 等人, EMBO Journa(5, (1986)1087 - 1090], 并对其进行了多种修饰以评估在由该基因编码的激活蛋白中的不同区域的功能 [Stankovitch 等人, Mol. Microbiol. 7, (1993)81 - 87]。另外,编码烟曲霉中相应功能的基因最近显然已进行了克隆 [Heasel 等人, 第二届欧洲真菌遗

传学会议,1994年4月28至5月1日,摘要本,E11]。

从文献中也得知了基因型 arg B areA1 的无冠构巢曲霉菌株作为生产 t-PA 的宿主的专一用途[Upshall 等人, *Biotechnology* 5, (1987)1301-1304]。在这个例子中,只有 argB 基因型通过其精氨酸原养型被用作选择标记物,而 areA 基因型不过是一个并发物。

本发明的目的在于缓解对无蛋白酶丝状真菌的需求。

发明概述

本发明涉及一种真菌,其中由重组 DNA 技术产生的 areA 基因已进行了修饰,使得其不能以提供功能性 AreA 激活物的方式进行表达。

本发明还涉及生产这些真菌的方法,是通过 areA 基因的缺失而获得的。

其可以通过一种方法而获得,该方法包括

- i)对来自所研究真菌的 areA 基因进行克隆,
- ii)生产含有 areA 基因的 DNA 构建体,其中基因内在部分已被取代、删除,或已插入额外的 DNA,
- iii)用该构建体来转化所述真菌,和
- iv)选择出为 areA⁻的转化体。

由上述对 areA 基因克隆而得到的信息也可与人们熟悉的反义技术结合用于构建可合成互补于由 areA 基因转录的 mRNA 的 RNA 分子的表达质粒,并以此来转化所研究的真菌。

本发明还涉及拟用于上述方法的 DNA 构建体。

本发明亦涉及生产所需蛋白质或基因产物(特别是分泌性蛋白质)的方法,即:用 DNA 构建体(至少含有所研究的蛋白质或基因产物的 DNA 编码序列)修饰并选择性转化真菌宿主,在适宜的条件下在适宜的生长培养基中进行培养,回收所需要的基因产物并提纯。

当实施本发明时,人们惊奇地发现,本发明的真菌产生了这样的分泌性蛋白质,而且产率大大提高。

还令人惊讶的是,能够提供米曲霉 areA⁻ 菌株良好生长的唯一氮源为谷氨酰胺。

最后,本发明涉及通过上述方法生产的蛋白质产物。

附图简述

以下部分将参照实施例及附图进一步详述本发明,其中

图 1 显示了 HowB101 构建的步骤,

图 2 显示了 pSK5 及 pSK9 构建的步骤,

图 3a 及 3b 显示了 pToC266 构建的步骤,

图 4 显示了 pMT1606 构建的步骤,

图 5 显示了 pToC56 构建的步骤。

定义

在本说明书中用到了以下定义。

areA Δ 意指一种菌株,其中缺失了 areA 基因。

areA⁻意指一种菌株,其不产生功能性 AreA 激活物。术语“丧失功能”也经常用于这方面。

“反义技术”描述了如美国专利 5,190,931 中所公开的几种方法。

发明详述

如所述,本发明的第一方面涉及真菌,其中通过重组 DNA 技术得到的 areA 基因已被修饰,使得其不能以提供功能性 AreA 激活物的方式进行表达。

这个目的尤其可通过 areA 基因的缺失或破坏而达到。

areA 基因的克隆见实施例中所述。

来自其它真菌的 AreA 同源物也可以进行克隆,方法既可以是与已知基因之一的交叉杂交,也可以是 areA 突变体的互补作用;例如,本申请中所述的无冠构巢曲霉 areA-18 或米曲霉 areA 缺失的菌株。

删除或破坏基因的方法具体描述于 WO90/00192(Genencor)。

取代基因中 DNA 的方法也是人们普遍知道的,其可以通过取代基因的一个或多个连续部分而实现,但其也可以通过位点针对性诱变产生编码无功能的 AreA 激活变异体的 DNA 序列而获得。

达到这种目的的另一方法是使用反义技术。

反义技术及其如何应用详见上述美国专利 5,190,931(纽约大学)。

获得所述失活作用的另一个方法是在 areA 基因内部插入额外的 DNA,从而引起机能障碍性激活蛋白的表达。

与该方法相关,由克隆提供的信息也能用于建造可以整合入 areA 基因中的 DNA 构建体,甚至可用另一个基因(如 pyrG 基因)来取代它。

避免 areA 激活物出现的另一个方法是干扰用于调控 areA 基因本身表达的表达信号的调控作用。

依据本发明,真菌优选属于选自下列的属:曲霉属、木霉属、腐质霉属、假丝酵母属、Acremonium、镰孢属及青霉属。

在这些属中,选自下列的菌种是优选的:米曲霉、黑曲霉、泡盛曲霉、海枣曲霉、日本曲霉、臭曲霉、无冠构巢曲霉、T. reesei、T. harzianum、H. insulens、柔毛腐质霉、禾木科镰孢、腐皮镰孢、产黄青霉等等。

如所述,本发明还包括生产本发明第一方面的真菌的方法,且其中所述失活作用已通过 areA 基因的缺失而获得,该方法包括

- i)来自所研究真菌的 areA 基因同源物的克隆化,
- ii)生产含有 areA 基因的 DNA 构建体,其中内在部分已被取代、删除,或者已插入额外的 DNA,
- iii)用该构建体来转化上述真菌,和
- iv)选择出为 areA⁻的转化体。

还包括产生该真菌的方法,其中失活作用已通过反义技术获得。这种方法包括

- i)一种表达质粒的构建,该质粒可引起互补于转录自 areA 基因的 MRA 的 RNA 分子的合成,
- ii)用所述表达质粒及适宜的标记物来转化宿主真菌,可在独立的质粒上进行,或在相同的质粒上进行,
- iii)利用上述标记物来选择转化体,和
- iv)筛选出在 AreA 产物的合成中表现出还原作用的菌株转化

体,例如可通过对各种氮源上生长的分析而获得。

本发明的另一个方面包括用于上述方法的 DNA 构建体。

就前者的方法而言,所述 DNA 构建体可含有 areA 基因,其中内在部分已被取代、删除,或已插入额外的 DNA。

DNA 构建体还可含有编码所研究的蛋白质产物的 DNA 序列,如后面所述及。

就后者的反义方法而言,DNA 构建体可含有与功能性启动子相连的 areA 基因的反向 DNA 序列,使得 mRNA 至少与来自 areA 基因的 mRNA 部分互补。

本发明的另一方面涉及生产所需要的基因产物(优选为分泌性基因产物)的方法,即依据本发明的真菌在适宜的条件下在适宜的培养基中进行培养,回收所需的基因产物并提纯。

在由异源基因所表达的基因产物方面,编码所需基因产物的 DNA 序列可以是用于产生所述真菌的 DNA 构建体的一部分。

不过,一般而言,本发明中真菌的独立转化作用的发生是为了使真菌能够生产所需要的产物。

转化真菌的方法是本领域人士所熟悉的,例如可参见 EP 0 184 A2 (Gist - Brocades N. V.) 及欧洲专利申请 87103806 (Novo Nordisk A/S)。

对于固有产物来说,当然这并不是必需的,但是为了增加产量,提供拟掺入宿主的编码所研究蛋白质的基因可能是有好处的。

所需基因产物通常为肽或蛋白质,优选为酶。

在酶中,其优选的是选自蛋白酶,如胰蛋白酶和凝乳酶、脂酶、几丁质酶、纤维素酶、木聚糖酶、漆酶、果胶酶等等。

另一类所需的基因产物一般为具有治疗活性的肽或蛋白质。

在具有治疗活性的肽或蛋白质中,蛋白质优选地选自胰岛素、生长激素、胰高血糖素、生长激素抑制素、干扰素、PDGF、V II 因子、V III 因子、尿激酶、t-PA、CSF、乳铁蛋白、TPO 等等。

以下给出的实施例将对本发明做进一步的详尽阐述。不过,这些实施例不应以任何方式限制如所附权利要求书中所限定的本发明

的范围。

实施例

材料及方法

菌株

米曲霉, IFO4177: 可从日本大阪发酵研究所; 日本大阪 17-25 Juso Hammachi Z-chome Yodogawa-Ku 得到。

ToC913: 该菌株的构建如实施例所述。

基因:

areA: 该基因编码能控制氮代谢的调控蛋白。

pyrG: 该基因编码乳清酸核苷-S-磷酸脱羧酶, 这是一种尿苷合成中所涉及的酶。

bar: 该基因最初分离自吸水链霉菌, 并为 phosphinothricin 乙酰转移酶编码。该酶可修饰 phosphinothricin (= glufosinate), 从而使这种对细菌、真菌及植物有毒性的化合物失活。

质粒

pUC118: Viera and Mesing J. Meth. Enzymol. 1987 153 3-11

pSO2: 该质粒的构建如实施例所述。

pJers4: pUC118 中 pSO2 的 2.0kb 亚克隆。pJers4 含有功能性米曲霉 pyrG 基因

pSO5: 该来自 pSO2 的质粒的构建如实施例所述。

pToC56: 该质粒的构建如欧洲专利申请 87103806 所述。

pToC266: 该质粒的构建如实施例所述。

pMT1606: 该来自 pBP1T(B. Straubinger 等人, Fungal Genetics Newsletter 39(1992):82-83) 及 p775(欧洲专利申请 87103806) 的质粒的构建如实施例所述。

p777: 该质粒的构建如欧洲专利申请 87103806 所述。

pHW470: 该质粒的构建如实施例所述。

实施例 1 米曲霉 areA Δ 菌株的构建

areA Δ 菌株是通过下述步骤构建的。对米曲霉 pyrG 基因进行克隆, 分离出米曲霉 pyrG 突变株。对来自米曲霉的 areA 基因进行

克隆。用携带着 pyrG 基因(插在 areA 基因的 DNA 片段上游与下游之间)来转化 pyrG 基因突变体。在质粒上并没有 areA 的编码区域。根据转化体在没有尿苷和存在氯酸盐的条件下生长的能力将它们选择出来。这种双重选择既可以选择功能性 pyrG 基因,也可以选择 areA 负型。最后利用 Southern 分析来筛选用这种选择程序所得到的菌株以鉴定其中染色体 areA 基因被 pyrG 基因所取代的菌株。

米曲霉 pyrG 基因的克隆化

通过与黑曲霉 pyrG 基因的交叉杂交对米曲霉 pyrG 基因进行克隆(W. van Hartingsvelolt 等人, Mol. Gen. Genet 206:71 - 75 (1987))。利用来自黑曲霉 pyrG 基因的 1kb DNA 片段在低严格性条件下探查部分 Sau III A 消化的米曲霉 IFO4177 DNA 的 λ 文库。将来自阳性克隆的 3.8kb Hind III 片段亚克隆到 pUC118 载体中。所生成的质粒 pSO₂ 通过黑曲霉 pyrG 突变体的互补作用而表现出含有 pyrG 基因。

米曲霉 pyrG 负型菌株的构建

pyrG 缺失质粒 pSO₅ 在每一末端含有约 1kb 的 pyrG 侧翼序列,该质粒是从质粒 pSO₂ 构建而来的。用该构建体转化米曲霉 IFO4177,并通过对 5-氟-乳清酸的抗性(一种具有 pyrG 突变体特性的表型)而选择出转化体。通过 Southern 分析表明了转化体 HowB101 在 pyrG 基因座具有预期的缺失。做为 pyrG 突变体,HowB101 需要有尿苷才能生长。通过无尿苷条件下生长能力的选择,用 wtpyrG 基因可转化 HowB101。

HowB101 构建中所涉及的步骤如图 1 所述。

areA 基因的克隆化

通过与无冠构巢曲霉 areA 基因的交叉杂交克隆了米曲霉 areA 基因[B. Kudla 等人, EMBOJ. 9: 1355 - 1364(1990)]。通过用 Sau III A 部分消化染色体 DNA 并将得到的 DNA 片段克隆到载体 λ GEM- II (得自 Promega)中,制备出米曲霉 IFO4177 的基因组文库。在 40% 的甲酰胺中,于 37℃ 进行该文库与无冠构巢曲霉 areA 基因的交叉杂交。从这些片段中分离出杂交性 λ 克隆,再亚克隆到

载体 pBluescript SK⁺ (得自 Stratagene) 中, 得到如图 2 所述的质粒 pSK5 及 pSK9。经克隆的基因能补足无冠构巢曲霉 areA 突变体, 这说明其的确是米曲霉 areA 同源物。对 5643bp 的克隆进行序列测定, 对米曲霉及无冠构巢曲霉 areA 基因序列的比较表明: 它们具有较高的同源性。米曲霉 areA 基因的序列如 SEQ ID NO.1 所示。

areA 缺失质粒的构建

为了从米曲霉染色体中删除掉 areA 基因, 构建了质粒 pToC266。pToC266 含有源于 areA 基因(分离自 pSK5)上游的 -2.1kb 的 DNA 片段以及源于 areA 基因(分离自 pSK9)下游的 -1.4kb 的 DNA 片段。这两个片段在基因组中相隔约 3.2kb。编码区域则位于该基因的这个部分。将来自 pJers4 的米曲霉 pyrG 基因插入 areA 上游及下游 DNA 片段之间。pToC266 的构建如图 3a 及 3b 所示。pToC266 具有一独特位点, 在用做转化作用之前, 可用该限制性内切酶的切割使其线性化。

米曲霉 areA Δ 菌株的选择

用线性化 pToC266 来转化米曲霉 HowB101。在基本平板[Cove Biochem. biophg. Acta(1966)113 51-56]上选择出转化体, 该平板含有 5% 的氯化钠、0.5mM 的硫酸铵和 1% 的葡萄糖。对转化体进行双重选择, 即从通过可在不加尿苷的条件下生长而得出 pyrG 基因以及氯化盐抗性两方面来选择。氯化盐抗性是无冠构巢曲霉 areA 突变体表型之一[H.N.Arst 及 D.J.Cove, MGG 126:111-141 (1973)]。在相同类型的平板上对生长性差的转化体重复分离两次。在不同的氮源上对三种独立的转化体 ToC913、ToC919 及 ToC920 进行生长测试。它们在谷氨酰胺上生长良好, 但在其它测试的氮源(包括氮)上生长较差。Southern 分析表明, 这三个菌株都缺失了 areA 结构基因, 后者已被 pyrG 基因所取代了。

areA Δ 菌株也可以通过在含有谷氨酰胺作为氮源的基本平板上对线性化 pToC266 转化体的选择而得到。在一个这样的实验中, 25 个转化体中有一个是 areA Δ 菌株。

实施例 2

pMT1606 的构建

构建 - 含有来自吸水链霉菌的 bar 基因 [C. J. Thompson 等人, EMBO J. 6:2519 - 2523 (1987)] 的质粒, bar 基因被插入在米曲霉 TAKA - 淀粉酶启动子之后, 其后是含有转录终止子的片段和来自黑曲霉 gla 基因的聚腺苷酸化信号。

质粒 pMT1606 可用于选择米曲霉的 glufosinate 抗性突变体。通过从质粒 pBP1T [B. Straubinger 等人, Fungal Genetics Newsletter 39:82 - 83 (1993)] 中分离出 bar 基因并将其克隆到真菌表达质粒 p775 (见欧洲专利申请 87103806) 中而构建出 pMT1606。图 4 示 pMT1606 的构建。

实施例 3

ToC913 (米曲霉 IFO4177 areA Δ) 中的凝乳酶的生产。

通过 pMT1606 的共转化作用用质粒 pToC56 转化米曲霉 areA Δ 菌株 ToC913 (图 5), 质粒 pToC56 是一种用于哺乳动物酶凝乳酶的真菌表达质粒。质粒 pToC56 的构建如欧洲专利申请 87103806 所述。

通过其产生凝乳酶的能力, 依据在基本培养基 (含有 10mM 的氨及 1mg/ml 的 glufosinate) 上的生长而选择转化体, 并依据 pToC56 的存在而进行筛选。使三个转化体在摇瓶中的基本培养基 (含有麦芽糖糊精及谷氨酰胺) 上于 30 $^{\circ}$ C 生长 4 天。使在 IFO4177 (如欧洲专利 87103806 所述而得到) 中的两个 pToC56 转化体以及未转化的 IFO4177 和 ToC913 与 ToC913 转化体共同生长。

每天抽取发酵液样品并用于 SDS - Page 及 Western 印迹法。将印迹膜与凝乳酶特异性兔抗体并随后与偶联在过氧化酶上的山羊兔抗体一起温育。膜的染色表明, 在发酵作用发生的第一天和第二天, 来自 IFO4177 转化体的上清液含有少量的凝乳酶或其降解产物, 而在其后的发酵中并不含这些物质。

ToC913 转化体含有至少 10 倍以上的整体凝乳酶。上清液中的凝乳酶量在头 2 - 3 天内是增加的, 其后则维持不变。

将 IFO4177、ToC913、ToC913 中 pToC56 的转化体以及

IFO4177 中的转化体发酵的第 3 天和第 4 天的上清液用于等电子聚焦凝胶,进行电泳操作。PH 梯度为 3.5-9.5。电泳之后,用 PH 为 7.0 的缓冲液(含 2mMZn²⁺)洗涤凝胶,再用含 0.5%酪蛋白的琼脂涂盖。在 45℃ 下对凝胶进行温育,直至可见到蛋白酶活性为止。

在 IFO4177 的样品中,可以看到三条具有蛋白酶活性的带,一个具碱性 pI,两个具酸性 pI。

在 IFO4177 的 pToC56 转化体的样品中,可见到来自凝乳酶的微弱反应,它被见于未转化的 IFO4177 中的酸性带之一部分地覆盖,具强酸性 pI 的蛋白酶几乎未见到,而具碱性 pI 的蛋白酶却伴随着具几乎中性的 pI 的一个或多个带而清晰可见。

在 ToC913 样品中,未检测到任何蛋白酶活性,而 ToC913 的 pToC56 转化体的样品却表现出强烈的凝乳酶信号。在该转化体样品中未检测到其它蛋白酶。

实施例 4

ToC913(米曲霉 IFO4177 areAΔ)中人胰蛋白酶 I 的生产

使用标准程序及由 M. Emi 等人发表的序列[Gene(1986)41:305-310,参见丹麦专利申请 no.693/95]来分离出编码人胰蛋白酶原 I(TRYI)的 cDNA。将 BamHI 位点(GGATCC)直接引入其间具有短序列 ACC 的起始密码子[ATG(Met)]的上游。

在如欧洲专利申请 no.87103806 所述的真菌表达质粒 p777 中用该 BamHI 位点使 cDNA 与 Taka-淀粉酶启动子中的 BamHI 接头融合。cDNA 的 3'末端在终止密码子下游的 41bp 与 p777 中的 NruI 位点融合。这样将 TRYICDNA 插在米曲霉 Taka-淀粉酶启动子与黑曲霉葡糖淀粉酶转录终止子之间。所生成的质粒称为 pHW470(参见丹麦专利申请 no.693/95)。

利用质粒 pMT1606 共转从作用将 pHW470 转化到 ToC913 中。通过分生孢子对 BASTA 抗性转化体进行两次重复分离。在 30℃ 条件下使 8 个转化体生长 4 天,培养基为 YPM[YPD(Sherman, F, 等人(1981),见“酵母遗传学方法”,冷泉港实验室,纽约冷泉港),其中用 2%的麦芽糖代替葡萄糖]。利用 SDS-PAGE,随后利用 Western

印迹法和与抗猪胰蛋白酶的兔抗体一起温育来分析上清液中人胰蛋白酶的含量。然后将印迹膜与偶联在过氧化酶上的山羊抗兔抗体一起温育并与 3'-氨基-9-乙基吖唑进行反应。三个转化体的上清液含有预期大小的染色带。三种阳性上清液中的胰蛋白浓度为 2-5mg/l。

通过将上清液样品与 L-苯甲酰-精氨酸-对硝基苯胺(L-BAPNA)一起温育进一步证实了胰蛋白酶的存在。三种免疫阳性菌株样品裂解了底物,引起了黄色的显色。来自 ToC913 及 IFO4177 的样品并未显示对该底物的丝毫活性。我们对该检测中人胰蛋白酶的特异活性尚不得而知,因而不可能从这些数据中计算出上清液中胰蛋白酶的浓度。

也在野生型菌株 IFO4177 中制备出 pHW470 转化体。我们看到了 20 多个 L-BAPAN 阳性转化体,但尚不可能从这些转化体的上清液中检测到任何免疫活性带。在该检测中检测限度约为 0.5mg/l。

序列表

(1)一般信息

(i) 申请人

(A) 名称: Novo Nordisk A/S

(B) 街道: Novo Alle

(C) 城市: Bagsvaerd

(E) 国家: 丹麦

(F) 邮编: Dk - 2880

(G) 电话: + 45 44422668

(H) 传真: + 45 44493256

(ii) 发明名称: 新型微生物

(iii) 序列数: 2

(iv) 计算机可读形式:

(A) 媒介类型: 软盘

(B) 计算机: IBM PC 兼容

(C) 操作系统: PC - DOS/MS - DOS

(D) 软件: PatentIn Release # 1.0, Version # 1.25(EPO)

(2) SEQ ID NO:1 的信息:

(i) 序列特性

(A) 长度 5643 个碱基对

(B) 类型: 核酸

(C) 链型: 单

(D) 拓扑学: 线性

(ii) 分子类型: DNA(基因组)

(iii) 假拟: 无

(iii) 反义: 无

(vi) 来源:

(A) 生物体: 米曲霉

(B) 菌株: IFO4177

(ix) 特征:

(A)名称/关键词:内含子

(B)位置:2701..2769

(ix)特征:

(A)名称/关键词:CDS

(B)位置:连接处(2282..2700,2770..4949)

(xi)序列描述:SEQ ID NO:1:

```

AAGCTTCGTC CTCGCATCTC GGCCGGGTGA GTAAGGTATG GTATTATTCA TGAAGGGATC      60
TCGTTGGTTA CCGTTGTCTA TCCCTAAACA AAGGATTCAA GAGAACAAC TCGAATGCTC      120
CCTCCGCTTA AACCCCTTGA CTCACTGATG GTGTATGTAC TATGGGTACG ACGTTCGGGA      180
TGTGGACTAC CAACCAGAGA GTGATTAGAG AGTCCGGGTT CTCAGTCCAT GATTTTTCGA      240
TCTTTGAAAC AGACGATGCG GAGCGGTCAT TGGCGGAGTT TACTCCCAA TACGGCCGAA      300
CGGGGTACTT TAAGTGGAAT CTCCGATTTT GGATCTAAGC TCATGAAGGA AAAGTACTAC      360
TAATGCGTAC CTGTGCCTAA TGTTAGTGCT AGTTCGTCTG TTGCATTTTA CCCGTCGGTT      420
AAGACGAATG GATCCGTTCA GGTTTTAAAA TAACTATCTA TGAAATATTT TAGATTTCCC      480
GACATAGTGG TTGGGATGTC TCGATTAACA CTAGGTACAT CAGGCTCAAT TGATTTTGGT      540
TTTAACGAAA CATGATATAG GTCAGGGTCG TGGACCACCC TCCGCCAGGG ATCAGGGGAC      600
GGTTACATGC GAAGGATTCT GATTATATTC ATGATTATGT CAAGCCTTTT CTCTCGTGTG      660
AAGAGGAGCA GAGAATCCGT ACGGGTTTAA TTTAATTTAG CGCCCTGCAG CTTCGAGAAC      720
ATCCCCAGCA ACGTTAAAAA CCACGAGCTA AAATGGGTCTG CCACCGGAAG CACTCGAGTC      780
GAGAGATCGG TCGGCTCAGT ATTCGTAATA CCTGCGTTCC AGACGGTTTT GGTCGTTGGT      840
TTCACTCAGG GAACTTAATT CCAGCGGGAC CCAATATAAT TTGAATGATT CATGATACAT      900
CCATTCGTTT GAACCGATCC TGCAAGAGTT CTGTCTGATT TGGTCAACAT AGTTTTCCTC      960
TGGGGGAGAC TGGGGAAGAG TCAACACAAT GGTCAGGGAG AGAAGAATGA AAGCTCTCGC     1020
AAGTGGATGA TCATGCTACG TACTGTAGGA ATAAAATTAA TTAATGCGAG GCTGCAAGTA     1080
TCCCTGCGCC GATTTTCTCT TCTTACGGCG GGAACCAAAA AATGTGACGC TGTGATTTTC     1140
TGGAAAAGGT AAGGATGTTT AGTTTCCCAG GATTATTACT GGTTCGGTAT GTGTATGTGT     1200
ATGGATATCA TTCCGTATGG ATACGCCCCT TTCCTCCGCC CAGAACCAGT CCGTCATCCA     1260
TCCTCCACTC TTTCTTCTCT TAGAGCCTTT CCACCTCTCT TCACTTTCTT TTTCTTTCCC     1320
CCCTCCCTCT TTGCTTTCCC TCTCCAGTA TTATTCTTAT ATTATCGGTT TGACCGTTCG     1380
CTCAGTATCG GCCCCCCGTG AATCACTTTT CGTTTCTCTT GTATTTTACT TTCCTATCTG     1440
GGATTGCTCC TCGATTAGCA GCTCTACTTC ATTCGGCCAT GTGCGTCTAG AGGGTCTAGC     1500
CCCTCTCTCT CTTTGCACTG ACTGTCAGCC ATACCATAGT ATCATCCCGG AATTAAGAAA     1560
AAAAAAGAAA TTATTCTACC TCCGATCTGG ACAAATTATA ACCAGGAGAA AATCAAGCGA     1620
AAGAGGGGCA AAGGAGGAGA CACCATTAAT ACTGGGTCTG GTTTGATTCA TGACATACAT     1680

```

TCGTCGTCTT GAATTTCAAT AGGTACGGAC TGATGCATTC CACTCGAGCC TTTTtagctg	1740
CGTGTCGGTC TCCAATCGCA CTTCTTTTCT TATTTCCCTTG TGGGATAAAT TGATTATTTA	1800
CCGTTTCGTT TTCTCTATAT TGCGGTGGTG GTGCGACCCA TCCAAC TATT ATTATTATAA	1860
TTGGAATTTG ATTTGGATTT TGATTCCTGT GACGGATCTC AGACCAAGTG CCTAAACTAT	1920
AACTGACTTG GACCCCTTC AGATCCTAGC TTCCCGATTC TTTCCACCA CTGCTGCATC	1980
CTCTTCCTGC ACGCAGCGTT CGTTTAGGGC GGGTAGACTG GAATTTATTC CTTGCGCCAC	2040
GGACCAATCG CTCCCTCGAC GCTCTCATTCT CTGCGTCGAG CTCTTTTTCC CTCGACTCTC	2100
ATTGCTTGCT GGGCTGGTTC TTGAACCTCT TCAATCGTCC TTATCTCTTT CCCCCATCC	2160
GGCCTGTGAT TCCTATCTTT CCTTTTTTTC TTCCCTTCT TGTTTGATCC CCCCTCCTCC	2220
CCGTCTTATC GCCTACTATC GTGATCCCCG CCCTTCCCAA TAAAGAGTAG GCGGTGTGAA	2280
C ATG TCC GGG TTA ACC CTC GGG CGA GGC CCT GGG GGC GTG CGA CCG Met Ser Gly Leu Thr Leu Gly Arg Gly Pro Gly Gly Val Arg Pro	2326
1 5 10 15	
ACT CAA ACC GCA ACT TTT ACC ACC CAC CAC CCG TCC GCC GAT GCT GAC Thr Gln Thr Ala Thr Phe Thr Thr His His Pro Ser Ala Asp Ala Asp	2374
20 25 30	
CGC TCC TCC AAC AAC CTC CCC CCT ACC TCC TCG CAG CTG TCC GAT GAC Arg Ser Ser Asn Asn Leu Pro Pro Thr Ser Ser Gln Leu Ser Asp Asp	2422
35 40 45	
TTT TCT TTC GGT TCC CCT CTG AGC CCC GCC GAC TCA CAG GCC CAT GAC Phe Ser Phe Gly Ser Pro Leu Ser Pro Ala Asp Ser Gln Ala His Asp	2470
50 55 60	
GGC CTA CTT CAG GAC TCC CTC TTC CCT GAA TGG GGG TCT GGT GCG CCT Gly Leu Leu Gln Asp Ser Leu Phe Pro Glu Trp Gly Ser Gly Ala Pro	2518
65 70 75	
CGA CCC GGC ATT GAC AGT CCG GAT GAG ATG CAG AGG CAA GAT CCG CTA Arg Pro Gly Ile Asp Ser Pro Asp Glu Met Gln Arg Gln Asp Pro Leu	2566
80 85 90 95	
GCG ACT CAA ATA TGG AAG CTC TAT TCT AGG ACC AAG GCC CAG TTG CCC Ala Thr Gln Ile Trp Lys Leu Tyr Ser Arg Thr Lys Ala Gln Leu Pro	2614
100 105 110	
AAC CAG GAG CGT ATG GAA AAC CTG ACC TGG CGG ATG ATG GCG ATG AGT Asn Gln Glu Arg Met Glu Asn Leu Thr Trp Arg Met Met Ala Met Ser	2662
115 120 125	
TTG AAA CGT AAG GAG CGG GAA CGT GCT CAA CAG TCC AT GTAGGTGTTC Leu Lys Arg Lys Glu Arg Glu Arg Ala Gln Gln Ser Met	2710
130 135 140	
TCCCTCTGTA GAGGAACGGC TGGACCCGCT CATCATTAAT TTTTTTTTGG TCTGTGAAG G	2770
TTT CCT GCG AGA CGC GGT AGC GCT GGC CCC AGT GGT ATC GCT CAA CTG Phe Pro Ala Arg Arg Gly Ser Ala Gly Pro Ser Gly Ile Ala Gln Leu	2818
145 150 155	
CGC ATT TCC GAC CCG CCC GTT GCC ACC GGT AAC CCT CAG TCA ACC GAC Arg Ile Ser Asp Pro Pro Val Ala Thr Gly Asn Pro Gln Ser Thr Asp	2866
160 165 170	

CTG ACC GCC GAC CCT ATG AAC CTC GAC GAT TTC ATC GTG CCC TTC GAA Leu Thr Ala Asp Pro Met Asn Leu Asp Asp Phe Ile Val Pro Phe Glu 175 180 185	2914
TCT CCT TCG GAC CAC CCC TCG CCC AGT GCC GTC AAG ATT TCC GAC TCC Ser Pro Ser Asp His Pro Ser Pro Ser Ala Val Lys Ile Ser Asp Ser 190 195 200	2962
ACG GCG TCC GCG GCC ATT CCC ATC AAG TCC CGG AAA GAC CAG CTG AGA Thr Ala Ser Ala Ala Ile Pro Ile Lys Ser Arg Lys Asp Gln Leu Arg 205 210 215 220	3010
GAT TCT ACC CCG GTG CCG GCC TCG TTC CAC CAT CCG GCT CAG GAT CAA Asp Ser Thr Pro Val Pro Ala Ser Phe His His Pro Ala Gln Asp Gln 225 230 235	3058
CGG AAG AAC AGT GAA TTT GGC TAC GTC CCC CGT CGC GTG CGC AAG ACG Arg Lys Asn Ser Glu Phe Gly Tyr Val Pro Arg Arg Val Arg Lys Thr 240 245 250	3106
AGT ATC GAC GAG CGT CAA TTT TTC TCA CTG CAG GTG CCG ACC CGA AAG Ser Ile Asp Glu Arg Gln Phe Phe Ser Leu Gln Val Pro Thr Arg Lys 255 260 265	3154
CGA CCG GCC GAA TCC TCG CCC CAG GTA CCC CCC GTT TCC AAC TCG ATG Arg Pro Ala Glu Ser Ser Pro Gln Val Pro Pro Val Ser Asn Ser Met 270 275 280	3202
TTG GCC CAC GAT CCG GAC CTC GCT TCC GGC GTG CCC GAT TAT GCC TTG Leu Ala His Asp Pro Asp Leu Ala Ser Gly Val Pro Asp Tyr Ala Leu 285 290 295 300	3250
GAC GCC CCG TCC TCG GCC TTT GGC TTC CAT CAG GGT AAC CAC CAT CCG Asp Ala Pro Ser Ser Ala Phe Gly Phe His Gln Gly Asn His His Pro 305 310 315	3298
GTC AAT CAT CAC AAC CAC ACC TCC CCC GGG GCA CCG TTT GGC TTG GAT Val Asn His His Asn His Thr Ser Pro Gly Ala Pro Phe Gly Leu Asp 320 325 330	3346
ACG TTC GGC CTG GGA GAT GAT CCA ATC TTG CCC TCC GCG GGC CCC TAC Thr Phe Gly Leu Gly Asp Asp Pro Ile Leu Pro Ser Ala Gly Pro Tyr 335 340 345	3394
CAG TCG CAA TTC ACC TTC TCA CCC AGC GAG TCT CCG ATG GCC TCC GGT Gln Ser Gln Phe Thr Phe Ser Pro Ser Glu Ser Pro Met Ala Ser Gly 350 355 360	3442
CAT CCG TTT GCG AAC CTC TAT TCG CAT ACC CCG GTG GCT TCG TCC CTC His Pro Phe Ala Asn Leu Tyr Ser His Thr Pro Val Ala Ser Ser Leu 365 370 375 380	3490
AAC TCG ACG GAT TTC TTC TCT CCA CCG CCA TCA GGC TAC CAG TCC ACG Asn Ser Thr Asp Phe Phe Ser Pro Pro Pro Ser Gly Tyr Gln Ser Thr 385 390 395	3538
GCA TCC ACG CCG CAG CCC ACC TAC GAC GGG GAC CAT TCC GTT TAT TTC Ala Ser Thr Pro Gln Pro Thr Tyr Asp Gly Asp His Ser Val Tyr Phe 400 405 410	3586
GAT ATG CCG TCG GGC GAC GCG CGC ACC CAG CGC CGC ATT CCG AAC TAT Asp Met Pro Ser Gly Asp Ala Arg Thr Gln Arg Arg Ile Pro Asn Tyr 415 420 425	3634
ATT TCG CAT CGG TCC AAC TTG TCT GCT TCG CTG CAG CCT CGG TAT ATG Ile Ser His Arg Ser Asn Leu Ser Ala Ser Leu Gln Pro Arg Tyr Met 430 435 440	3682

TTC Phe 445	AAC Asn 445	CAG Gln 445	AAC Asn 445	AAC Asn 445	CAT His 450	GAA Glu 450	CAG Gln 450	GCC Ala 450	AGT Ser 455	TCG Ser 455	TCG Ser 455	ACG Thr 455	GTG Val 455	CAT His 460	TCG Ser 460	3730
CCG Pro 465	AGC Ser 465	TAC Tyr 465	CCC Pro 465	ATT Ile 465	CCC Pro 465	CAG Gln 465	CCG Pro 465	CAA Gln 465	CAT His 470	GTG Val 470	GAC Asp 470	CCC Pro 470	ACT Thr 475	CAG Gln 475	GTG Val 475	3778
TTG Leu 480	AAC Asn 480	GCC Ala 480	ACC Thr 480	AAT Asn 480	TAC Tyr 480	TCG Ser 485	ACC Thr 485	GGC Gly 485	AAC Asn 485	TCC Ser 485	CAC His 485	CAT His 485	ACC Thr 490	GGC Gly 490	GCC Ala 490	3826
ATG Met 495	TTT Phe 495	TCA Ser 495	TTT Phe 495	GGA Gly 495	GCC Ala 495	GAT Asp 500	TCA Ser 500	GAT Asp 500	AAC Asn 500	GAG Glu 500	GAT Asp 505	GAC Asp 505	GAT Asp 505	GGT Gly 505	CAT His 505	3874
CAG Gln 510	CTG Leu 510	TCC Ser 510	GAG Glu 510	CGG Arg 510	GCT Ala 510	GGT Gly 515	CTG Leu 515	GCG Ala 515	ATG Met 515	CCG Pro 515	ACT Thr 520	GAA Glu 520	TAT Tyr 520	GGG Gly 520	GAC Asp 520	3922
GAG Glu 525	GAC Asp 525	GGG Gly 525	TTC Phe 525	TCG Ser 530	TCG Ser 530	GGC Gly 530	ATG Met 530	CAG Gln 530	TGG Trp 530	GAT Asp 535	GGG Gly 535	CAG Gln 535	TTC Phe 535	CCG Pro 535	GGC Gly 540	3970
TCC Ser 545	TTC Phe 545	CAT His 545	TCG Ser 545	CTG Leu 545	CCG Pro 545	GGC Gly 545	TTT Phe 545	GGC Gly 545	CCT Pro 550	CAA Gln 550	CAT His 550	CGC Arg 550	AAG Lys 550	CAT His 555	GTT Val 555	4018
ACC Thr 560	ATC Ile 560	GGG Gly 560	TCC Ser 560	ACG Thr 560	GAC Asp 560	ATG Met 565	ATG Met 565	GAC Asp 565	ACC Thr 565	CCC Pro 565	GAG Glu 565	GAG Glu 565	TGG Trp 570	AAT Asn 570	CAC His 570	4066
GGT Gly 575	GGC Gly 575	AGT Ser 575	TTG Leu 575	GGT Gly 575	CGG Arg 575	ACT Thr 580	CAT His 580	GGG Gly 580	TCG Ser 580	GTG Val 580	GCT Ala 585	TCG Ser 585	GTC Val 585	AGT Ser 585	GAG Glu 585	4114
GTG Val 590	CGC Arg 590	AAC Asn 590	CGA Arg 590	GAG Glu 590	CAG Gln 590	GAC Asp 595	CCT Pro 595	CGC Arg 595	CGG Arg 595	CAG Gln 595	AAG Lys 600	ATT Ile 600	GCC Ala 600	CGC Arg 600	ACC Thr 600	4162
ACG Thr 605	TCC Ser 605	ACC Thr 605	CCC Pro 605	AAT Asn 605	ACG Thr 610	GCC Ala 610	CAG Gln 610	CTG Leu 610	TTG Leu 610	CGC Arg 615	CAA Gln 615	AGC Ser 615	ATG Met 615	CAC His 615	TCT Ser 620	4210
AAT Asn 625	AAC Asn 625	AAT Asn 625	ACG Thr 625	TCT Ser 625	CAT His 625	ACC Thr 625	TCC Ser 625	CCT Pro 630	AAT Asn 630	ACG Thr 630	CCG Pro 630	CCC Pro 630	GAG Glu 635	TCC Glu 635	GCC Ala 635	4258
CTG Leu 640	AGC Ser 640	AGC Ser 640	GCA Ala 640	GTT Val 640	CCG Pro 640	TCC Ser 645	CGC Arg 645	CCG Pro 645	GCC Ala 645	AGT Ser 645	CCC Pro 645	GGG Gly 650	GGC Gly 650	AGC Ser 650	AAG Lys 650	4306
AAC Asn 655	GGC Gly 655	GAC Asp 655	CAA Gln 655	GGC Gly 655	AGC Ser 655	AAC Asn 660	GGA Gly 660	CCG Pro 660	ACC Thr 660	ACC Thr 660	TGC Cys 665	ACG Thr 665	AAC Asn 665	TGC Cys 665	TTC Phe 665	4354
ACT Thr 670	CAA Gln 670	ACC Thr 670	ACT Thr 670	CCG Pro 670	CTG Leu 670	TGG Trp 675	CGT Arg 675	CGG Arg 675	AAC Asn 675	CCA Pro 675	GAG Glu 680	GGC Gly 680	CAG Gln 680	CCA Pro 680	CTG Leu 680	4402
TGC Cys 685	AAT Asn 685	GCC Ala 685	TGC Cys 685	GGG Gly 685	TTG Leu 690	TTT Phe 690	TTG Leu 690	AAA Lys 690	TTG Leu 690	CAC His 695	GGT Gly 695	GTC Val 695	GTG Val 695	CGC Arg 695	CCT Pro 700	4450
CTG Leu 705	TCC Ser 705	CTG Leu 705	AAA Lys 705	ACG Thr 705	GAC Asp 705	GTT Val 705	ATC Ile 705	AAA Lys 705	AAG Lys 710	CGC Arg 710	AAC Asn 710	CGT Arg 710	AGC Ser 710	AGT Ser 710	GCC Ala 710	4498

AAC AGC TTG GCG GTT GGG ACC TCC CGT GCG TCG AAG AAG ACA GCC CGC Asn Ser Leu Ala Val Gly Thr Ser Arg Ala Ser Lys Lys Thr Ala Arg 720 725 730	4546
AAG AAC TCG GTG CAG CAA GCA TCC GTC ACG ACT CCG ACA TCA AGC CGC Lys Asn Ser Val Gln Gln Ala Ser Val Thr Thr Pro Thr Ser Ser Arg 735 740 745	4594
GCT CAG AAT GGG ACT TCC TTC GAA TCC CCG CCC GCC GGC TTT AGT GCT Ala Gln Asn Gly Thr Ser Phe Glu Ser Pro Pro Ala Gly Phe Ser Ala 750 755 760	4642
GCC GCG GGA CGG TCG AAT GGG GTG GTA CCC ATT GCC GCC GCT CCT CCG Ala Ala Gly Arg Ser Asn Gly Val Val Pro Ile Ala Ala Ala Pro Pro 765 770 775 780	4690
AAG GCA GCT CCC TCC GCA GCC GCC TCC CCT AGC ACG GGC CAG ACC CGC Lys Ala Ala Pro Ser Ala Ala Ala Ser Pro Ser Thr Gly Gln Thr Arg 785 790 795	4738
AAC CCG ATC CAG GCT GCC CCG AAA CGT CAA CGA CCG CTG GAA AAG GCC Asn Pro Ile Gln Ala Ala Pro Lys Arg Gln Arg Arg Leu Glu Lys Ala 800 805 810	4786
ACG GAG ATG GAA ACG GAC GAG GCT AAC AAG TCC GCG GGA GGC CGA TCC Thr Glu Met Glu Thr Asp Glu Ala Asn Lys Ser Ala Gly Gly Arg Ser 815 820 825	4834
AAG GTG GTG CCT CTG GCA CCC GCC ATG CCA CCG GCA GCA GCC AAT CCG Lys Val Val Pro Leu Ala Pro Ala Met Pro Pro Ala Ala Ala Asn Pro 830 835 840	4882
GCG AAC CAT AGT ATT GCC GGA GGC CAA GGG GCT AGT CAG GAA TGG GAG Ala Asn His Ser Ile Ala Gly Gly Gln Gly Ala Ser Gln Glu Trp Glu 845 850 855 860	4930
TGG TTG ACG ATG AGT CTGTAATGGC CGCGCTTACC TCTCTACTTC TCTACTCTCG Trp Leu Thr Met Ser Leu 865	4985
TTTCTTAATA TCTTTCTTGA ACCCCCCCTT ATATTTTCCC ACCGTTGATG CTACGCCATG	5045
ACCGATAGAG ATGATGAATA CTGCAACCAA TGGAATCTCG CTAGACGAGA GGTGTTAGAT	5105
GACGTGGCCC GCGATGCACT TAATGAGATA CGAGGAGGTG CAATGCGTTG GTTACGCTAG	5165
TTTAATGGTA ACATGACGAG GGATATTCGC TCTGTTATTT CGGGCTTTGA TCTGTTTCAG	5225
TCTGCGATTT AACAGCGACT GATCCTCTGC TGTGACAATA CACAGCTTGT CTTGTGGTTC	5285
TGTTGTGGCT TTCTGTTTGT TTGGCTGATT TGATTTATGC TTGATACAAT CGCGTCTGTC	5345
CGGACCCCGG CCTTTGTTTT GTTTTCAATT CTGATTCTTC ACTGTTTCTG ATTCTCTTGT	5405
TCATGTTTTT GATTTGTTCA AGGCTTGGGG CCGGGCAGAA GTGCGCATCT CTGCTTTGTG	5465
TTTTCCGTCA CCGTGATAG ACGCTGTATG TATATGCTAC AGCAAGATTC TACTTATCCA	5525
GTCTGAGCCT GTATTCATTG AAGTGTAGCC AGCTGTGCGAA TGAGCTTTTT AACGATATTG	5585
TTTTGTTGAG TAGTCAACAA GTAGTATCTG TATATTCCGG AGTCTAAGTA AGACACTT	5643

(2)SEQ ID NO:2 的信息

(i)序列特性:

(A)长度:866 个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:蛋白质

(xi)序列描述:SEQ ID NO:2

```

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

```

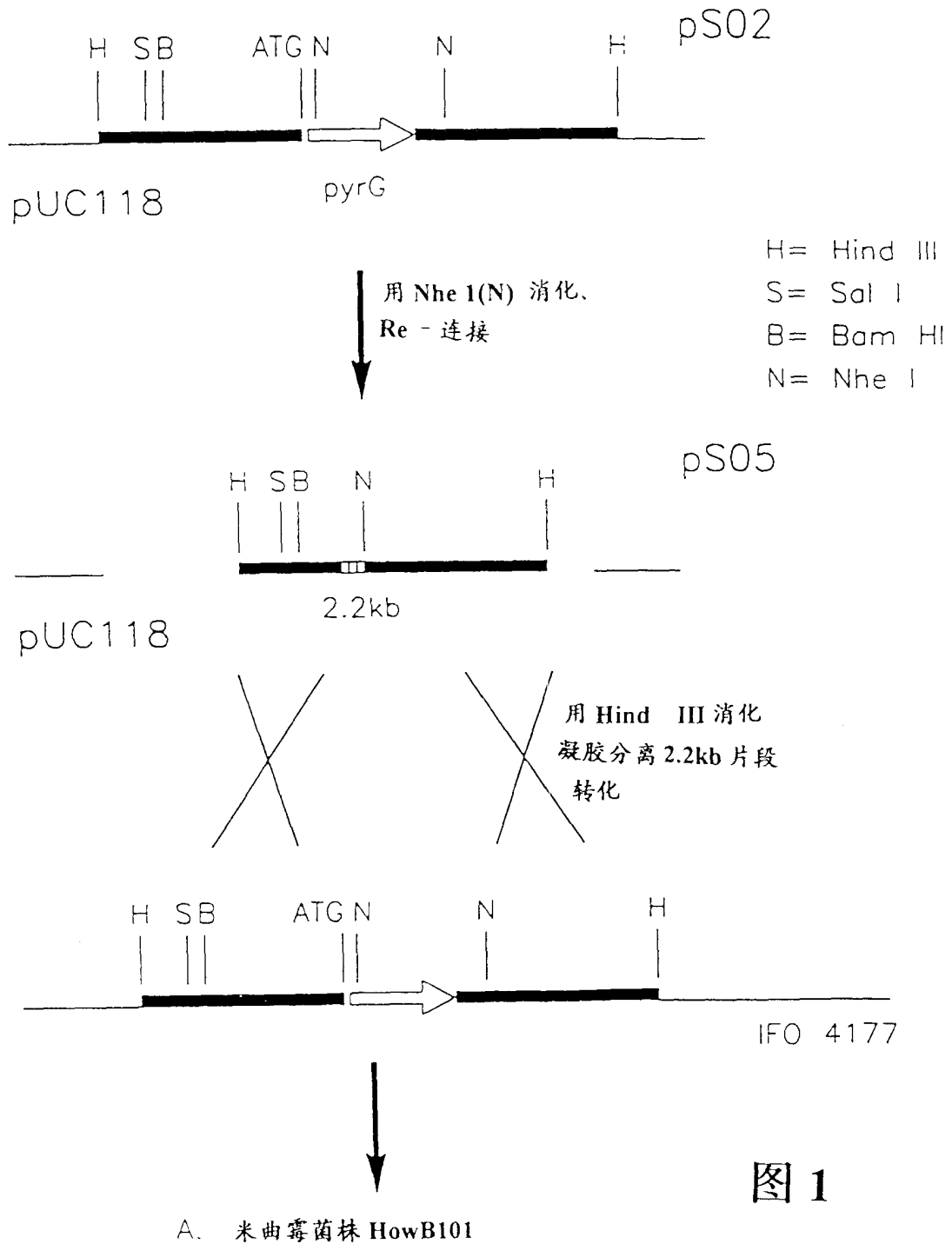
Val Pro Ala Ser Phe His His Pro Ala Gln Asp Gln Arg Lys Asn Ser
 225 230 235 240
 Glu Phe Gly Tyr Val Pro Arg Arg Val Arg Lys Thr Ser Ile Asp Glu
 245 250 255
 Arg Gln Phe Phe Ser Leu Gln Val Pro Thr Arg Lys Arg Pro Ala Glu
 260 265 270
 Ser Ser Pro Gln Val Pro Pro Val Ser Asn Ser Met Leu Ala His Asp
 275 280 285
 Pro Asp Leu Ala Ser Gly Val Pro Asp Tyr Ala Leu Asp Ala Pro Ser
 290 295 300
 Ser Ala Phe Gly Phe His Gln Gly Asn His His Pro Val Asn His His
 305 310 315 320
 Asn His Thr Ser Pro Gly Ala Pro Phe Gly Leu Asp Thr Phe Gly Leu
 325 330 335
 Gly Asp Asp Pro Ile Leu Pro Ser Ala Gly Pro Tyr Gln Ser Gln Phe
 340 345 350
 Thr Phe Ser Pro Ser Glu Ser Pro Met Ala Ser Gly His Pro Phe Ala
 355 360 365
 Asn Leu Tyr Ser His Thr Pro Val Ala Ser Ser Leu Asn Ser Thr Asp
 370 375 380
 Phe Phe Ser Pro Pro Pro Ser Gly Tyr Gln Ser Thr Ala Ser Thr Pro
 385 390 395 400
 Gln Pro Thr Tyr Asp Gly Asp His Ser Val Tyr Phe Asp Met Pro Ser
 405 410 415
 Gly Asp Ala Arg Thr Gln Arg Arg Ile Pro Asn Tyr Ile Ser His Arg
 420 425 430
 Ser Asn Leu Ser Ala Ser Leu Gln Pro Arg Tyr Met Phe Asn Gln Asn
 435 440 445
 Asn His Glu Gln Ala Ser Ser Ser Thr Val His Ser Pro Ser Tyr Pro
 450 455 460
 Ile Pro Gln Pro Gln His Val Asp Pro Thr Gln Val Leu Asn Ala Thr
 465 470 475 480
 Asn Tyr Ser Thr Gly Asn Ser His His Thr Gly Ala Met Phe Ser Phe
 485 490 495
 Gly Ala Asp Ser Asp Asn Glu Asp Asp Asp Gly His Gln Leu Ser Glu
 500 505 510
 Arg Ala Gly Leu Ala Met Pro Thr Glu Tyr Gly Asp Glu Asp Gly Phe
 515 520 525

Ser Ser Gly Met Gln Trp Asp Gly Gln Phe Pro Gly Ser Phe His Ser
 530 535 540
 Leu Pro Gly Phe Gly Pro Gln His Arg Lys His Val Thr Ile Gly Ser
 545 550 555 560
 Thr Asp Met Met Asp Thr Pro Glu Glu Trp Asn His Gly Gly Ser Leu
 565 570 575
 Gly Arg Thr His Gly Ser Val Ala Ser Val Ser Glu Val Arg Asn Arg
 580 585 590
 Glu Gln Asp Pro Arg Arg Gln Lys Ile Ala Arg Thr Thr Ser Thr Pro
 595 600 605
 Asn Thr Ala Gln Leu Leu Arg Gln Ser Met His Ser Asn Asn Asn Thr
 610 615 620
 Ser His Thr Ser Pro Asn Thr Pro Pro Glu Ser Ala Leu Ser Ser Ala
 625 630 635 640
 Val Pro Ser Arg Pro Ala Ser Pro Gly Gly Ser Lys Asn Gly Asp Gln
 645 650 655
 Gly Ser Asn Gly Pro Thr Thr Cys Thr Asn Cys Phe Thr Gln Thr Thr
 660 665 670
 Pro Leu Trp Arg Arg Asn Pro Glu Gly Gln Pro Leu Cys Asn Ala Cys
 675 680 685
 Gly Leu Phe Leu Lys Leu His Gly Val Val Arg Pro Leu Ser Leu Lys
 690 695 700
 Thr Asp Val Ile Lys Lys Arg Asn Arg Ser Ser Ala Asn Ser Leu Ala
 705 710 715 720
 Val Gly Thr Ser Arg Ala Ser Lys Lys Thr Ala Arg Lys Asn Ser Val
 725 730 735
 Gln Gln Ala Ser Val Thr Thr Pro Thr Ser Ser Arg Ala Gln Asn Gly
 740 745 750
 Thr Ser Phe Glu Ser Pro Pro Ala Gly Phe Ser Ala Ala Ala Gly Arg
 755 760 765
 Ser Asn Gly Val Val Pro Ile Ala Ala Ala Pro Pro Lys Ala Ala Pro
 770 775 780
 Ser Ala Ala Ala Ser Pro Ser Thr Gly Gln Thr Arg Asn Pro Ile Gln
 785 790 795 800
 Ala Ala Pro Lys Arg Gln Arg Arg Leu Glu Lys Ala Thr Glu Met Glu
 805 810 815
 Thr Asp Glu Ala Asn Lys Ser Ala Gly Gly Arg Ser Lys Val Val Pro
 820 825 830

Leu Ala Pro Ala Met Pro Pro Ala Ala Ala Asn Pro Ala Asn His Ser
835 840 845

Ile Ala Gly Gly Gln Gly Ala Ser Gln Glu Trp Glu Trp Leu Thr Met
850 855 860

Ser Leu
865



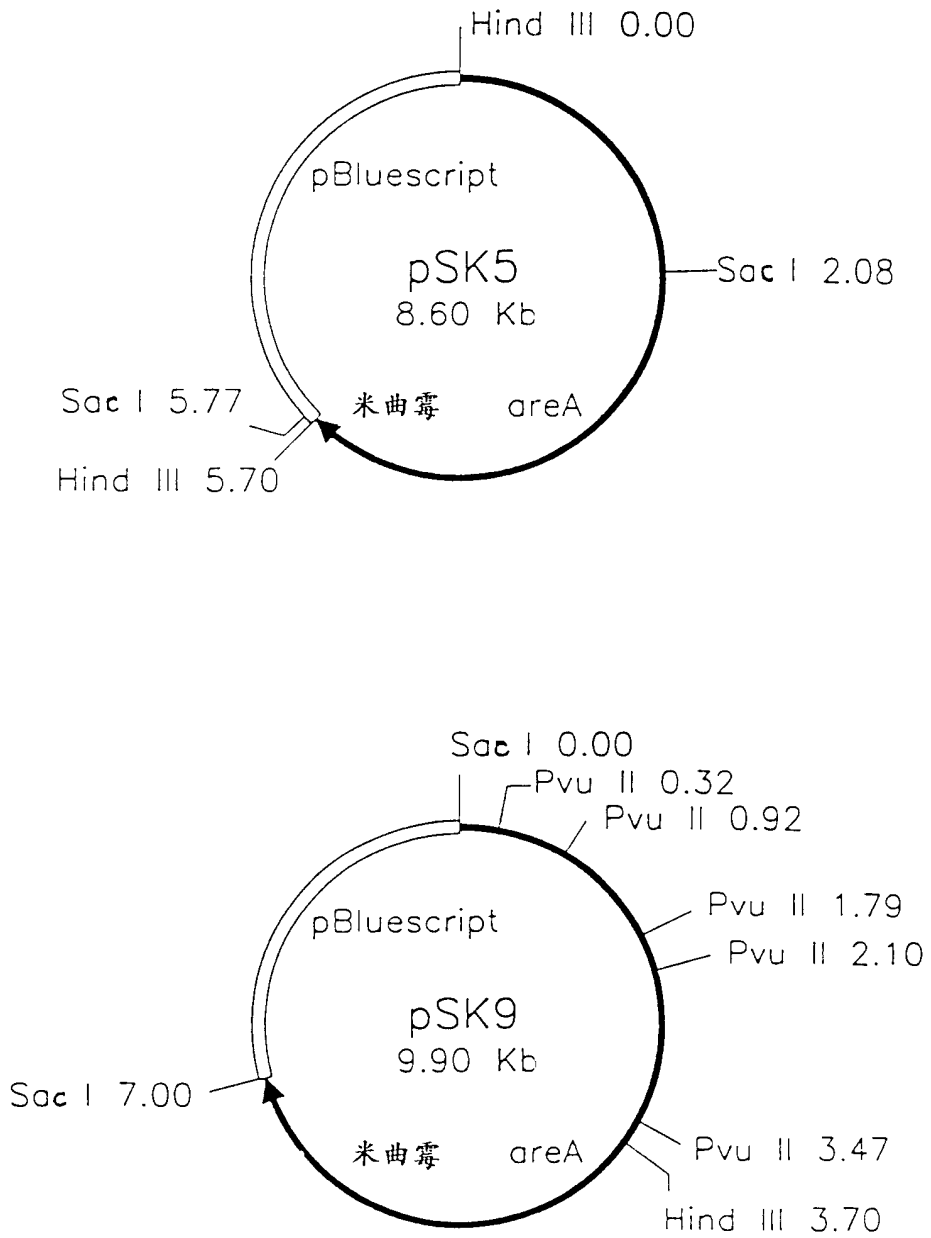
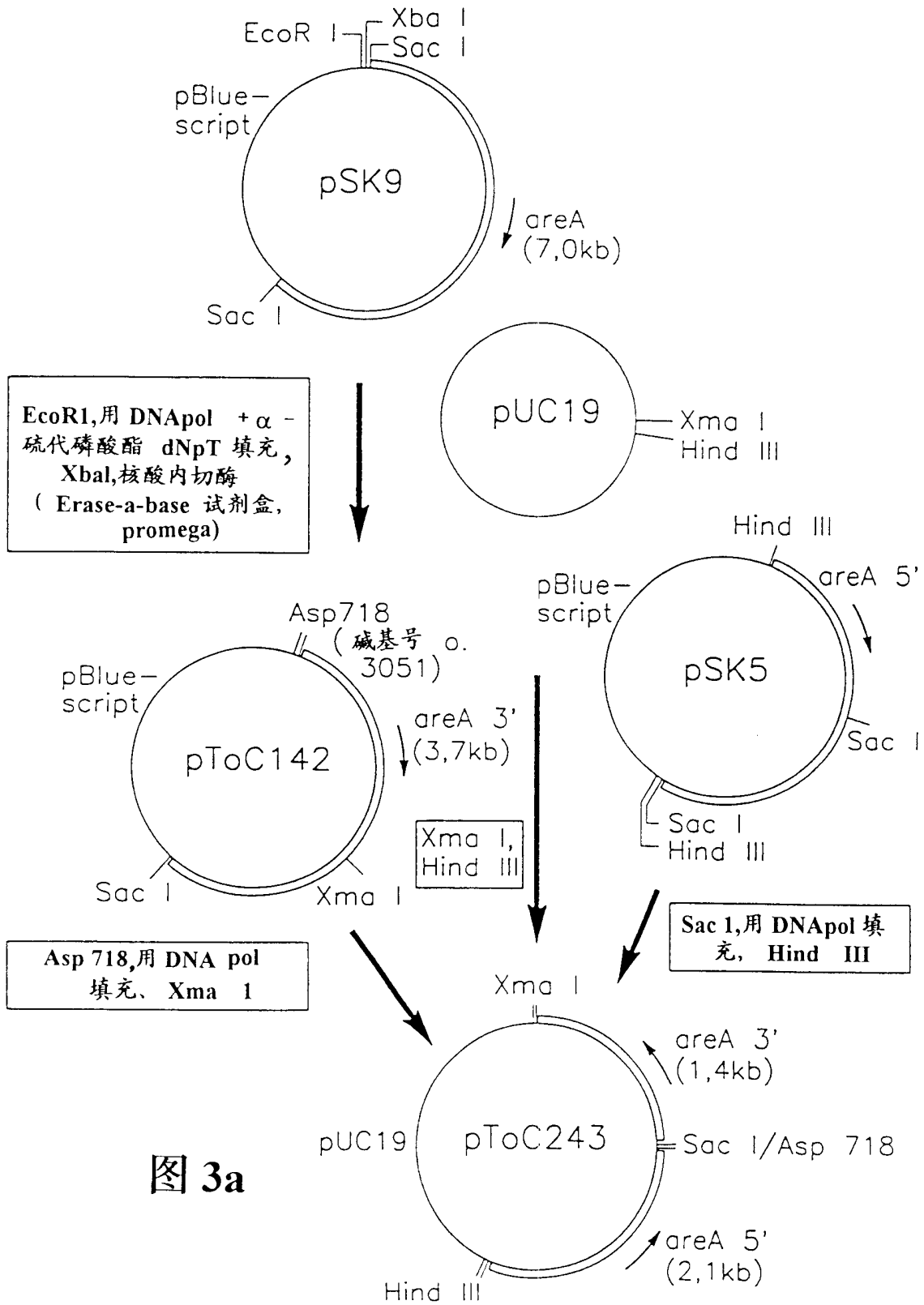


图 2



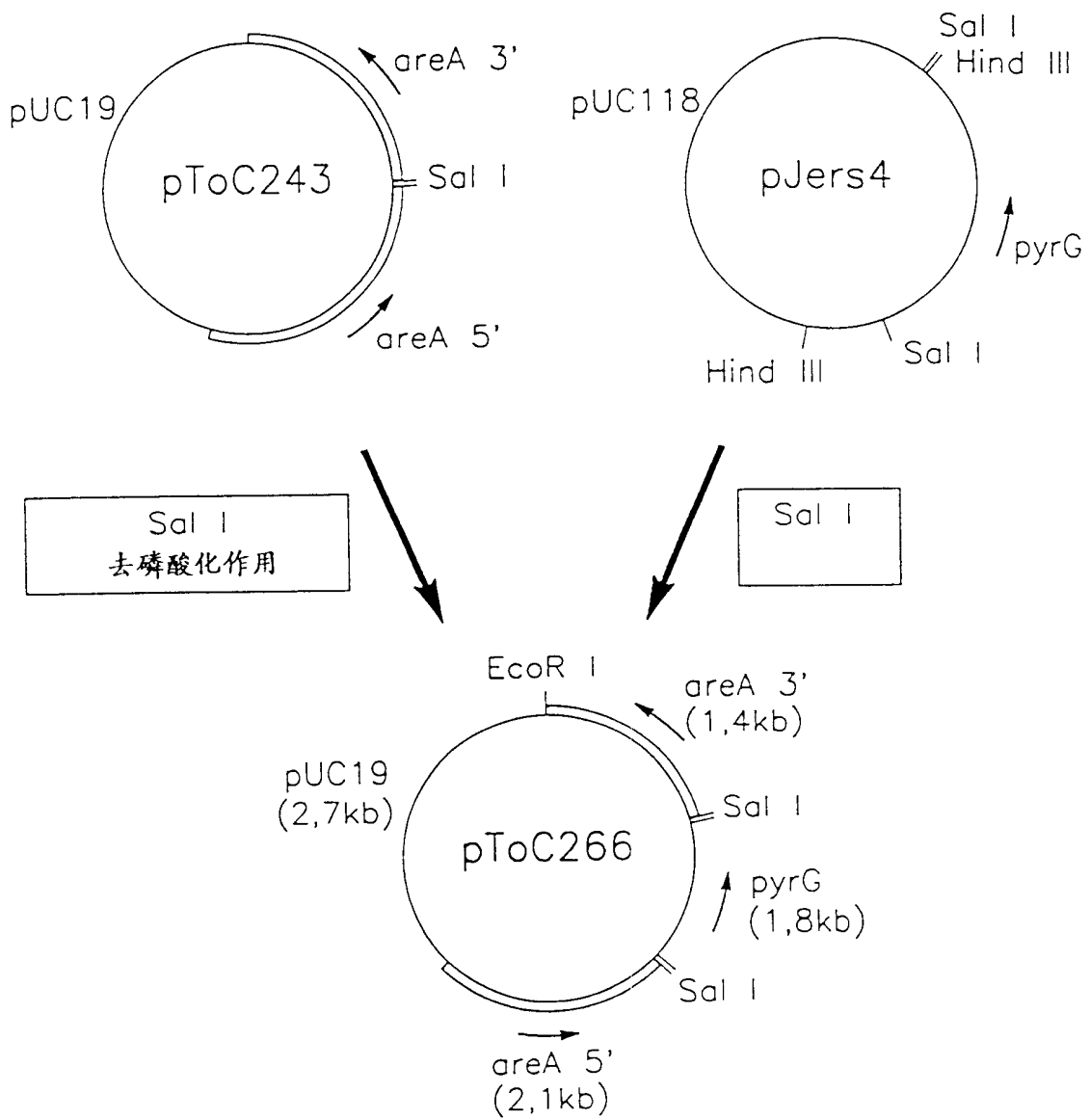


图 3b

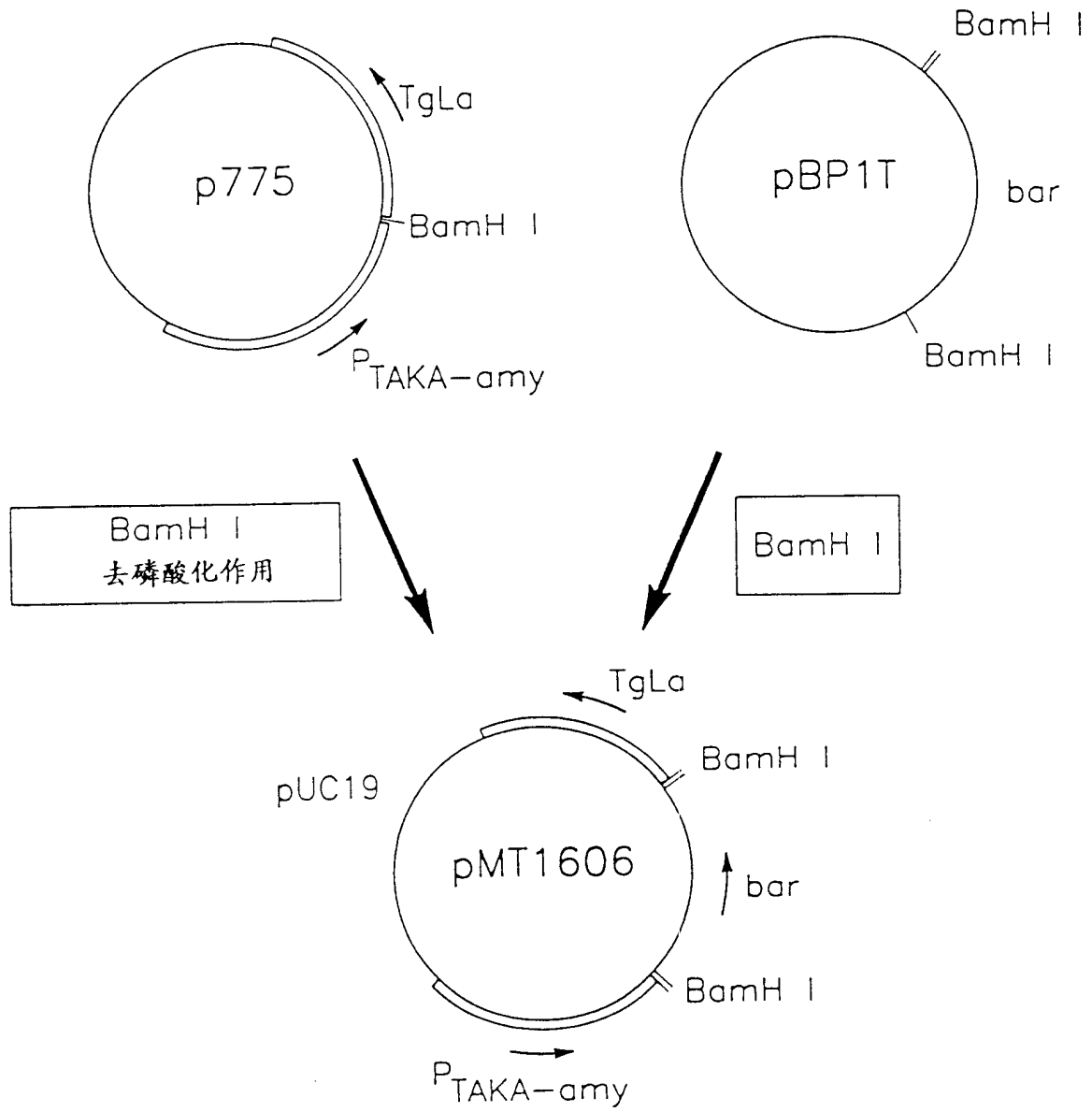


图 4

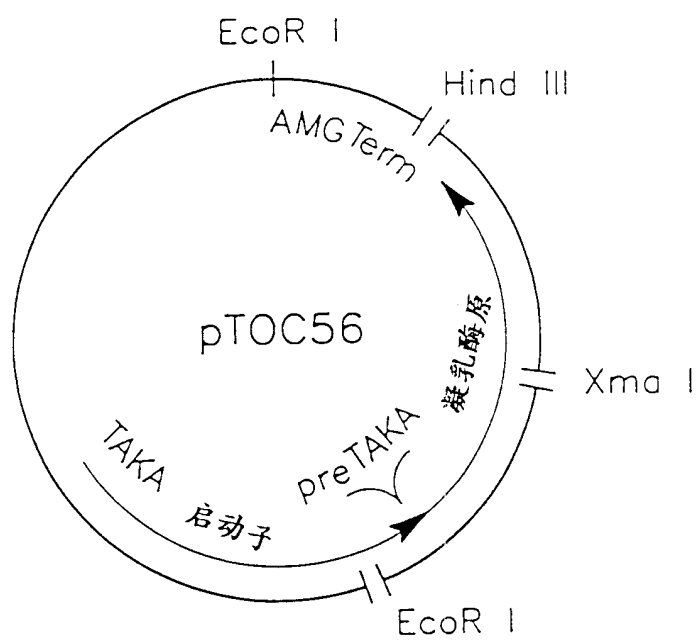


图 5