

[19]中华人民共和国国家知识产权局

[51]Int. Cl⁶

C12N 15/53

C12N 9/06 C12N 15/90

[12] 发明专利申请公开说明书

[21] 申请号 98800601.4

[43]公开日 1999年8月25日

[11]公开号 CN 1226932A

[22]申请日 98.4.17 [21]申请号 98800601.4

[30]优先权

[32]97.4.18 [33]ES [31]P9700833

[86]国际申请 PCT/ES98/00101 98.4.17

[87]国际公布 WO98/48019 西 98.10.29

[85]进入国家阶段日期 99.1.6

[71]申请人 抗生素有限公司

地址 西班牙马德里

[72]发明人 何塞·曼努埃尔·费尔南德斯·卡农

马特·罗德里格斯·塞斯

布鲁诺·戴伊斯·加西亚

何塞·路易斯·巴雷多·富恩特

亚历杭德罗·维塔莱·阿尔瓦

福兰西斯科·塞托马尔多纳多

米格尔·安赫尔·佩尼亚尔瓦·索托

何塞·曼努埃尔·明戈特·阿森桑

[74]专利代理机构 中科专利代理有限责任公司

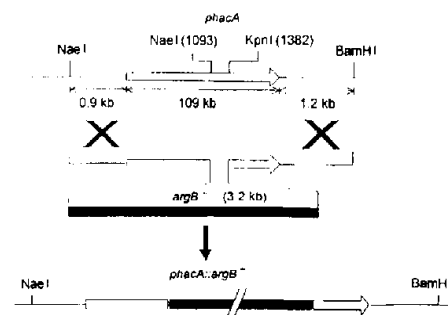
代理人 李悦

权利要求书 3 页 说明书 27 页 附图页数 7 页

[54]发明名称 使编码乙酸苯酯分解代谢酶的基因失活的方法和质粒以及用其转化的菌株

[57]摘要

本发明涉及一种使编码乙酸苯酯分解代谢酶的基因失活的方法和含有它的质粒 以及其转化的菌株。该方法选择性地用于构巢曲霉 *pahcA* 基因及产黄青霉的 *pahA* 基因。这两种基因分别编码同青霉素合成酶竞争底物的竞争酶。使这两种酶不表达将有利于这种抗生素的合成,即在青霉素生产中在培养其中加入少量的乙酸苯酯即可增加青霉素的产量。



ISSN 1008-4274

权 利 要 求 书

1. 在微生物中使一些基因失活的方法, 这些基因编码乙酸苯酯分解代谢酶, 该酶与青霉素生物合成酶竞争乙酸苯酯, 该方法包括通过同源重组整合转化, 这种重组发生在至少一个外源 DNA 化合物和至少待失活基因的部分序列之间。

2. 根据权利要求 1 所述的方法, 其中转化体 DNA 化合物是环状分子, 至少含有一个待失活的基因的表达片断。

3. 根据权利要求 1 所述的方法, 其中转化体 DNA 化合物是线性分子, 至少含有一个不包括在待失活基因的序列中的 DNA 片断, 但至少含有一个转化标记, 该标记中断其编码序列。

4. 根据前面几项权利要求所述的方法, 其中转化体 DNA 分子, 含有全部或部分待失活基因的序列的拷贝, 失活的方法是突变, 优选地是读码框架移动突变, 没有有义或缺失, 这将导致所述基因失去功能性表现型。

5. 根据前面的任何一项权利要求所述的方法, 其中能够产生青霉素 G 或 V 的微生物是转化的目标, 在转化中基因失活。

6. 根据权利要求 5 所述的方法, 其中产生青霉素 G 或 V 的微生物是真菌。

7. 根据权利要求 1 到 6 所述的方法, 其中转化的微生物是构巢曲霉。

8. 根据权利要求 1 到 6 所述的方法, 其中转化的微生物是产黄青霉。

9. 根据权利要求 1 到 7 所述的方法, 其中所用的转化体 DNA 化合物可以全部或部分地包括在载体中, 优选地在质粒 pPhacA::argB 中。

10. 根据权利要求 1 到 6 和 8 所述的方法, 其中所用的转化体 DNA 化合物可以全部或部分地包括在载体中, 优选地在质粒 pALP696 中。

11. 根据权利要求 1 到 7 和 9 所述的方法, 其中失活的基因通过 SEQ ID NO:1, 它的突变基因序列和/或同源基因代表。

12. 根据权利要求 11 所述的方法, 其中失活的基因编码 SEQ ID NO:2 表示的多肽, 或表达 P450 乙酸苯酯 2-羟化酶活性的相似的序列。



13. 根据权利要求 1 到 6, 8 和 10 所述的权利要求, 其中失活的基因通过 SEQ ID NO:3, 它的突变基因序列和/或同源基因代表。

14. 根据权利要求 13 所述的方法, 其中失活基因编码 SEQ ID NO:4 代表的多肽或表达 P450 乙酸苯酯 2-羟化酶活性的相似的序列。

15. 构巢曲霉的转化菌株和起源于该菌株的突变体, 该菌株至少掺入了一个外源 DNA 化合物, 该 DNA 化合物包括在至少一个基因的截短的, 不完全的或失活的序列中, 该基因编码乙酸苯酯分解代谢酶, 在它同源重组整合后, 使内源基因失活。

16. 根据权利要求 15 所述的构巢曲霉的转化菌株, 其中待失活的基因编码传导乙酸苯酯羟化作用的酶。

17. 根据权利要求 15 和 16 所述的构巢曲霉的转化菌株, 其中失活基因是 *phacA* 和在 SEQ ID NO:1 中, 它的突变基因序列和同源基因叙述的该基因的序列和侧接区。

18. 根据权利要求 17 所述的构巢曲霉的转化菌株, 其中失活基因与 *phacA* 同源, 来源不是构巢曲霉, 该基因与构巢曲霉的 *phacA* 基因的 DNA 序列的同源性足以介导与构巢曲霉的内源基因的同源重组。

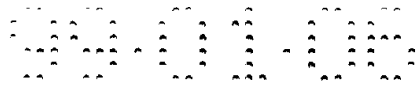
19. 根据权利要求 15 到 18 所述的构巢曲霉的转化菌株, 其中失活的基因编码蛋白质序列 SEQ ID NO:2 代表的多肽, 或表达 P450 乙酸苯酯 2-羟化酶活性的相似的序列。

20. 根据权利要求 15 到 19 所述的转化菌株, 特征是它包括在构巢曲霉 CECT20195 纯化菌株, 它的突变体和/或转化衍生物中。

21. 如图 6 的限制图谱所示, 失活构巢曲霉的 *phcA* 基因的质粒 pPhacA::argB, 它包括在质粒 pUC18 中, 该质粒含有 6.8kb EcoRI 插入片断, 该 3.2kb 片断含有构巢曲霉的 *phacA* 基因, 该基因通过插入 3.2kb 的片断而失活, 该片断含有构巢曲霉的 *argB* 基因。

22. 产黄青霉的转化菌株和起源于该菌株的突变体, 含有包括在截短, 不完全或失活的基因的序列中的 DNA 化合物, 该基因编码乙酸苯酯分解代谢酶, 并在通过同源重组整合该基因后, 使内源基因失活。

23. 根据权利要求 22 所述的产黄青霉的转化菌株, 其中待失活的基因编码传导乙酸苯酯羟化作用的酶。



24. 根据权利要求 22 和 23 所述的产黄青霉的转化菌株，其中基因是 *phcA* 和在该基因 SEQ ID NO: 3，它的突变基因序列和/或同源基因所述序列和它的侧接区。

25. 根据权利要求 22 到 24 所述的产黄青霉的转化菌株，其中失活的基因编码蛋白质序列 SEQ ID NO: 4 代表的多肽，或表达确定的 P450 乙酸苯酯 2-羟化酶活性的相似序列。

26. 根据权利要求 24 所述的产黄青霉的转化菌株，其中该基因是构巢曲霉的 *phacA* 基因或任何其它从不同于产黄青霉的来源中分离的同源 DNA 化合物，它与产黄青霉的 *pahA* 的 DNA 序列的同源性足以传导与产黄青霉的内源基因的同源重组。

27. 图 7 限制图谱所表明的失活产黄青霉 *pahA* 基因的质粒 pALP696，该质粒包括质粒 pBC KS+，pBC KS+ 包含具有产黄青霉的 *pahA* 基因的 7.9kb 的 *Sa*II 插入片断，然后通过插入 2.0kb 含有产黄青霉的 *gdh* 启动子控制下表达的 *S. hindustanus* 的 *ble^R* 基因使 *PahA* 基因失活。

说明书

使编码乙酸苯酯分解代谢酶的基因失活的方法和质粒以及用其转化的菌株

发明领域

本发明的基础是鉴定了新的 DNA 化合物和重组 DNA 分子，它们编码从丝状真菌构巢曲霉 (*Aspergillus nidulans*) 和产黄青霉 (*Penicillium chrysogenum*) 得到的乙酸苯酯 2-羟化酶，用前述 DNA 化合物转化生产青霉素的菌株能使这一抗生素的产量提高。

已有技术

在某些丝状真菌中，苯青霉素(青霉素 G)生物合成途径的最后步骤包括在异青霉素 N 转换成青霉素 G 的步骤中。这包括酰基移转反应，在这个反应中，用一个苯乙酸替代了异青霉素 N 中的 L-氨基己二酸侧链。催化这一反应的酶(酰基-辅酶 A: 异青霉素 N 酰基转移酶)利用活化的以辅酶 A (CoA) 的硫酯形式的苯乙酸作为它的底物之一。

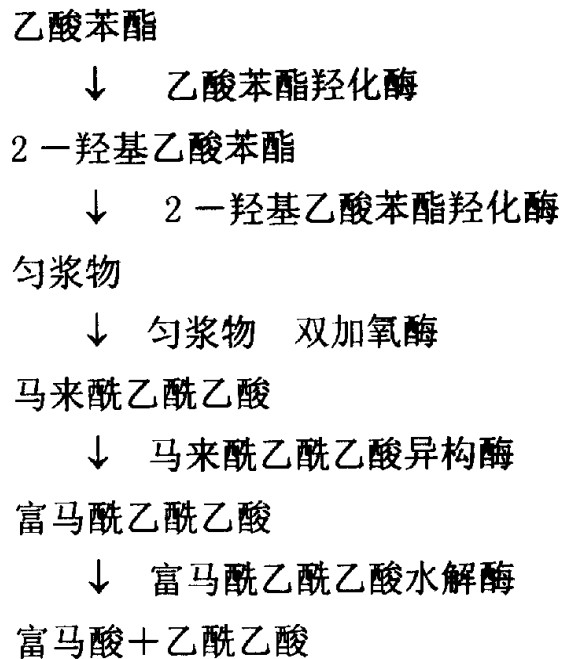
用于工业生产青霉素的真菌产黄青霉，和构巢曲霉不能合成乙酸苯酯。结果是，为了促进青霉素 G 的生物合成，必须在产黄青霉的工业培养物中加入过量的乙酸苯酯。过量的乙酸苯酯能防止不需要的其它具有脂肪族侧链的青霉素的合成，但增加了发酵过程的最终成本。

加入到青霉素生产的培养物中的部分乙酸苯酯可能参与了代谢。在产黄青霉发酵中积累相当量的 2-羟基乙酸苯酯是一个完全确定的事实。这一侧链的前体组分显然在青霉素 G 的生物合成中没有贡献。

构巢曲霉和产黄青霉都通过同样的酶步骤从三个前体氨基酸合成青霉素。另外，当在培养物中加入过量的乙酸苯酯时，构巢曲霉和产黄青霉一样，将乙酸苯酯转换成 2-羟基乙酸苯酯。构巢曲霉除了将乙酸苯酯转换成 2-羟基乙酸苯酯，也可以分解代谢这一最后化合物，事实上，构巢曲霉能利用乙酸苯酯作为唯一的碳源。



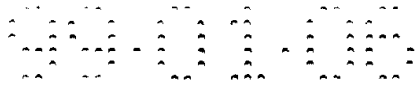
构巢曲霉中的分解代谢途径如下图所示：



这一途径的第一步是乙酸苯酯的邻位羟基化作用。第二个羟基化反应将2-羟基乙酸苯酯转换成2,5-二羟基乙酸苯酯(=匀浆物)。匀浆物通过富马酰乙酰乙酸分解代谢成富马酸和乙酰乙酸，这两个产物参与了Krebs循环(参见 Fernandez Canon y Penalva, 美国核酸科学进程 92 : 9132 - 9136, 1995)。

所述的丝状真菌部分地或完全分解代谢乙酸苯酯的能力是青霉素生产过程中的不利特征。由于这一原因，完全或部分地去除这一特征将产生青霉素G生产能力提高的菌株。

在生产青霉素的真菌中进行经典的诱变去除这一不利特征(也就是部分或完全分解代谢乙酸苯酯的能力)需要在大量的菌株中进行艰巨的选择工作。不管怎样，所述的诱变经常产生继发突变，这些继发突变可能会去除亲本菌株的有用的特征，例如产生营养缺陷型突变或生长活力或孢子形成减弱的突变，而生长活力和孢子的形成是工业发酵的关键。所以，为了部分或完全去除分解代谢乙酸苯酯的能力，使用基因工程技术是恰当的，这些技术没有所述的局限性。进行涉及的基因工程的基本需要是克隆和鉴定基因，该基因或这些基因传递前面所述的生产青霉素的真菌的不利特



征。

发明的详细说明

本发明的目的是解决本领域现存的前面所述的问题。包括鉴定真菌中的基因，该基因编码乙酸苯酯 2-羟化酶活性，催化曲霉和青霉中乙酸苯酯分解代谢的第一个酶解步骤的酶，和使用该基因通过重组 DNA 技术去除生产青霉素的真菌的基因组中存在的该基因。本专利叙述了在基因工程构建的菌株中这一基因是如何失活的。这一基因工程菌株不能分解代谢乙酸苯酯，产生的青霉素水平比亲本菌株高。另外，这一重组菌株的最大青霉素产量需要的乙酸苯酯超额量比亲本菌株需要的小。利用上面的 DNA 化合物使乙酸苯酯分解代谢失活，导致青霉素产量的明显提高。

SEQ ID NO:1 显示了来自构巢曲霉的 1986 个碱基对 (bp) 的基因组 DNA 片断的序列，该序列中包括用于本发明的新基因。将该基因称为 phacA (phac 是由于使用乙酸苯酯 phenylacetate)，它编码将乙酸苯酯邻位羟基化的酶 (这一反应是构巢曲霉中乙酸苯酯分解代谢的第一步)。与基因组 DNA 互补的 DNA (cDNA) 克隆的核苷酸序列覆盖了整个编码区和它们随后的排列，并暴露了这一基因具有下面的特征：

—该基因编码具有 518 个氨基酸的多肽。第一个甲硫氨酸由位置 82 的 ATG 三联体编码，而终止密码 (TAG) 位于位置 1810。这些位置与 SEQ ID NO:1 中显示的核苷酸序列匹配。

—编码区被三个长度为 65，56 和 53 个核苷酸的内含子 (SEQ ID NO:1) 中断。

—在 SEQ ID NO:1 中，以对应的外显子下面的三字母氨基酸密码表示了推断的 518 个残基的多肽，该多肽也在 SEQ ID NO:2 中单独以氨基酸密码表示了。推断的蛋白质的分子量是 58,495 克/摩尔。利用国家生物技术信息中心 (NCBI, 美国) 的公共可得服务 BLAST，在蛋白质序列的数据库 (如 SwissProt 和 PIR) 中，和在 DNA 序列的数据库 (如，基因库和 EMBL 数据库) 的概念性翻译中，对六个可能的读码框架进行的研究表明该氨基酸序列与细胞色素 P450 家族的成员 (血红蛋白) 相似。这些蛋白质通常参与氧化过程。事实上，与残基 431 和 439 之间的序列相对应的含有氨基酸半胱氨酸的肽，Gly-X-Gly-X-X-X-Cys-X-Gly (其中 X 指任何氨基酸)，该肽

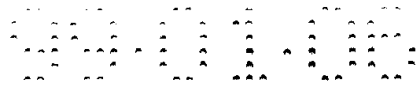


参与了血红素基团的结合，这是这些血硫蛋白的特征。

新的 DNA 化合物可以从天然微生物中分离，也可以利用自动 DNA 合成仪全部合成，SEQ ID NO:1 详细叙述了它的结构。另外，由于遗传密码的简并性，新 DNA 化合物编码的蛋白质可以由替代的 DNA 序列编码。这些也包括在本发明中。另一方面，任何起源于上面叙述的新 DNA 化合物的天然遗传突变体将认为是它的等同物。这些遗传突变体包括从进化角度上与构巢曲霉密切相关的有机体如产黄青霉中的同源基因。如下所述，利用来自构巢曲霉的 DNA 化合物作为杂交探针可以容易地鉴定这些基因。

可以将构巢曲霉的 *phacA* 基因用作分子探针通过杂交在其它真菌品种的基因组 DNA 和 cDNA 文库中寻找功能同源基因。所以，例如，本发明叙述了利用该基因对产黄青霉 DNA 文库的筛选，表明从产黄青霉的工业菌株中可以分离到与构巢曲霉的 *phacA* 同源的基因。SEQ ID NO: 3 显示了来自产黄青霉的基因组的 2558bp 的 DNA 片断的序列，该片断是通过与 *phacA* 基因探针杂交分离得到的。将与构巢曲霉的 *phacA* 同源的产黄青霉基因命名为 *pahA* (由于乙酸苯酯羟基化作用 *phenylacetate hydroxylation*)。产黄青霉的 *pahA* 基因编码含有 516 个氨基酸的多肽，该多肽显示了与 *PhacA* 的氨基酸序列有 84 % 的同一性，*PhacA* 的氨基酸序列即构巢曲霉的 *phacA* 基因的蛋白质产物。SEQ ID NO:4 显示了产黄青霉的 *pahA* 基因的产物的序列。如前所述，这一从产黄青霉分离的新 DNA 化合物也包括在本发明中，如其它同源基因一样，这些同源基因是通过同样的方法从真菌的其它种类中分离得到的。

通过反向遗传学可以将已经克隆的这些基因用于产生失去功能的突变。为了这一目的，可以在携带截短的该真菌基因的大肠杆菌载体中构建新的重组 DNA 分子，这样的载体可以通过转化使该内源基因失活。例如，这一重组质粒可以含有构巢曲霉的 *phacA* 基因的 5' 区，接着对 *phacA* 基因编码区修饰，该修饰包括用含有构巢曲霉的 *argB* 基因的 3.2kb *XbaI* 片断替代 289bp 的 *Nae I-kpnI* 片断(参见图 1)。然后是 *phacA* 基因的 3' 区。这一突变的 *phacA* 基因的表达将产生 *PhacA* 蛋白，该蛋白在残基 297 处截短，结果是缺乏 221 个羧基末端的残基。突变蛋白中缺乏的后面的这些残基包括前面提到的肽，该肽包含参与血红素基团的结合的 Cys 残基，该残



基是这一类型的蛋白质活性的关键。

通过适当的限制酶可以从载体的序列中分离到含有上面提到的区域的线性 DNA 片断，然后通过标准技术纯化。这一线性片断可以用于将构巢曲霉的 *argB*⁻ 菌株转化成精氨酸原养型。由于用于转化的 DNA 缺乏自我复制需要的序列。原养型转化体是从将该 DNA 片断整合进基因组得到的。如图 1 所示，一个可能的整合方法是在线性 DNA 分子与存在于真菌基因组中的 *phacA* 座位之间的双交叉连接，这一方法将产生 *argB*⁻ 表现型。这一双交叉连接发生导致体外产生的截短的基因替代了原初的 *phacA* 基因，引起 *phacA* 功能的丢失。通过它们的杂交方式可以识别携带这一替代(即失去功能的突变)基因的转化体，在这一方法中，利用了适当的限制酶消化基因组 DNA，然后利用 ³²P 放射性标记的 *phacA* 或 *argB* 基因作为探针，通过 Southern 技术杂交分析。通过这种方式，可以容易地将其它类型的转化体的杂交方式与对应于基因替代的那些杂交方式区别开。利用标准技术可以纯化通过 *phacA* 基因提供的缺乏功能的转化体，用于进一步测试。

与野生型菌株不同，通过反向遗传学产生的构巢曲霉转化体缺乏 *phacA* 基因的功能，所以这些转化体在乙酸苯酯作为唯一碳源时不能生长。但是，它们却可以在 2-羟基乙酸苯酯或 2, 5-二羟基乙酸苯酯中生长。这种情况表明新 DNA 化合物(构巢曲霉的 *phacA* 基因)编码乙酸苯酯的邻位羟基化所需要的酶活性，乙酸苯酯的邻位羟基化是构巢曲霉中利用乙酸苯酯的基本步骤。产黄青霉确实不利用乙酸苯酯作为唯一碳源，但如前面所示，产黄青霉含有编码酶 PhacA 的同系物的基因。所以，新的 DNA 化合物提供了去除侧向的代谢转化的方法，该方法可以降低青霉素生物合成中乙酸苯酯的积累。

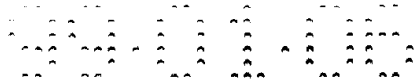
在产黄青霉中，利用与前叙的相似的方法使 *pahA* 基因失活，其中利用了本发明中包括的新产黄青霉 DNA 化合物和任何已经得到的产黄青霉转化标记，例如 *trpC* 基因。同样令人满意的是除了 *argB*(在构巢曲霉中)或 *trpC*(在产黄青霉中)以外的转化标记可以用于中断新 DNA 化合物的编码区。这些转化标记包括其它营养缺陷型标记(例如 *pyrG* 或 *riboB*)和抗生素抗性基因(例如在真菌中授予 *flomycin* 或潮霉素 B 抗性的基因)。这些方法基本上与这里所用的并包括在本发明中的那些方法相当。图 2 显示了

实施例 1. 克隆和鉴定构巢曲霉的 phacA 基因

采取了下述的步骤通过从 cDNA 文库特异选择可以分离对应于 phacA 基因的 cDNA 克隆:

a) 获得对应于基因转录物的 cDNA 群体, 当真菌在存在乙酸苯酯时生长时, 这些基因转录物得到优选表达(用作筛选 DNA 文库的“+”探针)。

将构巢曲霉 A26 菌株在 37 °C, 在适当补充的最小培养基中培养, 该菌株来自真菌遗传原种收藏中心(堪萨斯大学的药学中心, 微生物系), 该培养基含有(克/升)KPO₄H₂, 13.6; (NH₄)₂SO₄, 2.0; MgSO₄ × 7H₂O, 0.25 和 Fe × 7H₂O, 0.0005, 碳源是 0.3%(w/v)的葡萄糖。通过分析培养基中的葡萄糖浓度确定利用葡萄糖的时间(利用酶试剂盒)。这通常是在不断搅拌温育 18 小时后。2 小时后在培养物中加入乙酸苯酯, 直到最后浓度 10 毫摩尔/升, 在 37 °C, 再搅拌 1 小时, 目的是诱导所述的基因的表达。在这一时间段后, 过滤收集菌丝体, 用水洗涤, 在液氮中冷冻。冷冻干燥用于分离总 RNA。将总 RNA 制备物通过 oligo-dT 纤维素亲和层析分离得到 poly(A⁺) mRNA。在存在禽成髓细胞瘤病毒逆转录酶(商业来源)时制备了单链 cDNA 作为引物, 这一制备过程利用了 2 微克来自乙酸苯酯诱导的菌丝体和 oligo-dT(15 聚合体)分离的所述 mRNA 作为模板。将反应物在 42 °C, 在缓冲液中温育 1 小时, 缓冲液含有 10 毫摩尔/升 Tris-HCl, pH8.8(在 25 °C), 50 毫摩尔/升 KCl, 0.1% Triton X-100, 5 毫摩尔/升 MgCl₂, 10 毫摩尔/升每种 dNTP, 和 0.5 单位的 RNasin。在第一条链合成后, 在 60 °C, 与 3 摩尔/升 NaOH 温育 1 小时去除模板 RNA。在中和(用酸)后, 通过用 2.5 倍体积的纯乙醇在 -80 °C 沉淀 2 小时, 并离心回收单链 DNA。用 30 倍过量的 poly(A⁺) RNA 除去这种 cDNA, poly(A⁺) RNA 是从开始利用葡萄糖时回收的菌丝体分离得到的, 所用的方法如 Sargent, T. D., 酶学方法 152: 423 - 432, 1987 所述。在 68 °C, 通过羟磷灰石层析分离 cDNA-RNA 杂合分子, 并丢弃。如上所述将留下的 cDNA 与过量的 poly(A⁺) RNA 杂交, 这时的 poly(A⁺) RNA 来自缺乏乙酸苯酯时培养的菌丝体。再次丢弃。如上所述, 回收剩余的单链 cDNA 群体, 这些 cDNA 代表了在存在乙酸苯酯时优先表达的基因的转录物。利用 [α-³²P]dCTP 和过量的所有其它 dNTP, 在存在 DNA 聚合酶的 Klenow 片断和作为引物的任意序列的六核苷



酸时均一地标记这种 cDNA。将得到的 ^{32}P 标记的 cDNA 群体(特异活性 $>10^8\text{cpm/毫克}$)用作探针,去筛选如 c) 部分所述构建的 cDNA 文库。

b) 得到对应于缺乏乙酸苯酯时转录的基因的 cDNA 探针(“—”探针)。

如 a) 部分所述合成单链 cDNA 并用 ^{32}P 标记,但是利用的 mRNA 来自在含有葡萄糖培养基中的葡萄糖用尽时并在缺乏乙酸苯酯情况,在 37°C 再温育 1 小时所得到的菌丝体。

c) 在 $\lambda\text{gt}10$ 载体中构建 cDNA 文库,构建过程利用的 cDNA 来自加入 10 毫摩尔/升乙酸苯酯作为唯一碳源时培养的菌丝体。

为了构建这种 cDNA 文库,如前面所述方法进行第一条 cDNA 链的合成反应。将得到的 cDNA-RNA 杂合体转化成双链 cDNA,方法是利用 Rnase H 在 RNA 链中导入任意裂口,将这些裂口用作引物点,并利用大肠杆菌 DNA 聚合酶 I。为了在这一末端平齐化的 cDNA 制备物中加上 EcoRI 末端,将含有平齐化磷酸化末端的合成受体具有双链 cDNA 的 EcoRI 突出端在加入 T4 DNA 连接酶情况下温育。然后纯化具有 EcoRI 末端的 cDNA,并用 T4 多聚核苷酸激酶和 ATP 磷酸化。将磷酸化的 cDNA 与 $\lambda\text{gt}10$ 载体的群体混合,这一载体已经预先用 EcoRI 消化并脱磷酸化,用混合物与 T4 DNA 连接酶一起温育。利用商业包装提取物,体外包装重组 DNA 分子。由于在大肠杆菌 hf1 菌株(F⁻, thi-1, leuB6, lacY1, tonA21, supE44, hf1A150, [chr::Tn10]中的溶菌表现型选择出重组噬菌体,在这些噬菌体中 cDNA 插入片断已经使 cI 基因失活(其中存在 EcoRI 位点)。通过这种方法,得到了总共 10^7 个重组克隆。

d) cDNA 文库筛选

利用在大肠杆菌 C600hf1⁺菌株中形成的溶菌斑选出 cDNA 文库,并将得到的溶菌斑转印到硝基纤维滤膜上。一个复膜与“+” cDNA 探针(a)中所述)杂交,另一个复膜与“—” cDNA 探针(b)中所述)杂交。然后选择与“+”探针杂交,与“—”探针不杂交的克隆并纯化。

e) 对应于 phacA 基因的 cDNA 鉴定

利用了包含在 EcoRI 接合体的 NotI 切割位点,通过用内切核酸酶 NotI 消化,从载体中切出了 cDNA 插入片断。在用 NotI 消化的 pBluescript SK(+) 质粒选择克隆的插入片断被亚克隆化,然后通过标准方法对它们测序。在



六个可能的读码框架中翻译 cDNA 插入片断的序列，并将该序列与利用 BLAST 公式，在公共可得的 DNA 序列数据库 GenBank 和 EMBL 的六个读码框架中的概念性翻译比较或与蛋白质序列数据库 SwissProt 和 PIR 比较。cDNA 插入片断的一个序列产生了扩展的开放读码框架，这一框架没有被终止密码子中断。推断的氨基酸序列表现了与细胞色素 P450 家族中的蛋白质的氨基酸序列明显的同一性，该家族蛋白通常参与氧化反应。正如后面将显示的(参见实施例 3)，该开放读码框架编码乙酸苯酯 2-羟化酶。通过杂交，利用这一 cDNA 插入片断得到了对应于这一基因的新 cDNA 克隆。对这些克隆的两条链测序，得到的核苷酸序列显示了长度为 1554 个 bp 的完全开放的读码框架(没有翻译终止密码)，这一读码框架编码 518 个氨基酸的多肽(SEQ ID NO:2)。这一序列与细胞色素 P450 蛋白质家族成员的同一性明确地决定了这一推断的蛋白质代表这一家族的新成员。

f) phacA 基因的克隆

利用对应于几乎完全的转录物的 ^{32}P 一标记的 cDNA 探针在 λ EMBL4 载体中筛选构建的构巢曲霉基因组 DNA 文库。纯化阳性克隆，利用限制性酶消化和与上面提到的探针杂交鉴定它们的插入片断。以这种方法，将两个邻接的长度为 2.4kb 和 1.9kb 的 BamHI 片断的杂交区图谱定位。图 4 显示了含有这些片断的基因组区的限制图谱。将所述的 BamHI 片断亚克隆进 pBluescript SK(+). 将含有 2.4kb 的 BamHI 片断的重组质粒命名为 pBS-FG4A，而携带 1.9kb BamHI 片断的质粒命名为 pBS-FG4B。

实施例 2. 克隆和鉴定产黄青霉的 pahA 基因，该基因是构巢曲霉的 phacA 基因的功能同系物

a) 克隆和确定产黄青霉的 pahA 基因的核苷酸序列

利用 (^{32}P) 放射性标记的探针筛选产黄青霉威斯康星 54-1255 的基因组 DNA 文库，该文库是利用克隆进 λ EMBL4 的 BamHI 位点的部分 Sau3A 片断构建的，该探针含有长度差不多完全的 phacA 基因的 cDNA。通过标准技术约将 12,000 个克隆转移到纤维素滤膜上。然后，在 37 °C，在缓冲液中杂交 24 小时，缓冲液含有 50 % 甲酰胺，5 × Denhart 溶液，5 × SSC，



0.1%SDS, 50 微克/毫升超声波处理的鲑鱼精子单链 DNA 和 50 纳克标记探针。杂交滤过物在 $0.2 \times \text{SSC}$, 0.1%SDS 中, 在 37°C 进行最后洗涤。用这一方法纯化了 10 个克隆, 分离它们的 DNA。然后, 利用限制酶消化和与前面所述探针杂交鉴定所述克隆的插入片断。杂交区定位于 2.55kb XhoI 片断中, 图 5 显示了这一片断的限制性图谱。将这一 DNA 片断亚克隆进质粒 pBluescript SK(+) 中, 产生的质粒称为 pALP520。通过二脱氧核苷酸方法可以确定 XhoI 片断的核苷酸序列, 该片断含有 pahA 基因(图 5) (Sanger, F., Nicklen, S. 和 Coulson, A. R. (1977) 美国核酸科学进程 74, 5463 - 5467)。

b) 在青霉素生产条件下, 从纯化的 RNA 构建产黄青霉 cDNA 文库和鉴定对应于 pahA 基因的 cDNA。

产黄青霉的工业菌株的发酵罐培养物在青霉素 G 生产条件下温育 100, 124 和 148 小时后, 收集生长的菌丝体, 并利用 Poly(A)Quick mRNA 纯化试剂盒(Stratagene) 从中得到 poly(A⁺) mRNA。根据制造商的说明, 利用 Zap-cDNA 合成试剂盒(Stratagene), 将上面 poly(A⁺) mRNA 作为模板构建 cDNA 文库。将此双链 cDNA 插入在噬菌体载体 λ ZAP 的 EcoRI 和 XhoI 限制位点之间, 转录物的 5' 区定位于 EcoRI 位点一侧。利用 Gigapack II Gold 包装提取物(Stratagene) 包装 cDNA 文库, 产生约 10^6 个独立的克隆。

利用标准方法 (Sambrook, J. Fritsch, E. F. 和 Maniatis, T. (1989), 分子克隆: 实验室手册, 第二版, 冷泉港实验室出版, 纽约) 进行这种 cDNA 文库的筛选, 其中利用了含有 pahA 基因的 1174bp 的 EcoRV 片断作为探针(图 5)。以这种方法, 分离了 12 个阳性重组克隆(称为 # 1 到 # 12), 选择含有最大尺寸(约 1600bp) 的插入片断的克隆用于鉴定。然后确定这一克隆的全部核苷酸序列, 从它的序列观察到它包括 1548 个 bp 的开放读码框架(ORF), 这一 ORF 编码 516 个氨基酸的多肽, 多肽的分子量为 58, 112 道尔顿, 与构巢曲霉的 phacA 基因编码的蛋白质有 84 % 的同一性。正如它的同源构巢曲霉基因的情况一样, 数据库的研究表明在 pahA 蛋白和 P450 蛋白家族之间有相当大的同一性。PahA 蛋白含有序列 Gly-X-Gly-X-X-X-Cys-X-Gly(残基 430 - 438, SEQ ID NO: 4), 这一序列是这一类型的蛋白质的特征(参见本发明的详细说明)。从这些结果可以得出



结论，产黄青霉的 *pahA* 基因是构巢曲霉的 *pahA* 基因的同系物。

实施例 3. 通过反向遗传学使构巢曲霉的 *phacA* 基因失活

用突变等位基因替代构巢曲霉的 *phacA* 基因的野生型等位基因，在突变等位基因中，可能是它的功能的关键的编码区的一部分 (SEQ ID NO:1 中的残基 298 到 392) 受到抑制，被 *argB*⁻ 基因替代。这一遗传操作在密码子 297 将 *phacA* 基因截短，产生了无效的突变等位基因。

另外，从 pBS-FG4B 纯化出 1.7kb 的 KpnI-EcoRI 片断 (图 4) 并亚克隆进 pUC18，该质粒已经用上面两种酶消化，产生质粒 pUCB。然后从 pBS-FG4A 纯化出 1.9kb 的 NaeI 片断 (图 4) 并克隆进用 SmaI 消化的 pBluescript SK(+) 中。选择质粒，该质粒中的不完全的 *phacA* 基因的编码链 (从 NaeI 位点开始) 的定向与载体中的编码 β 一半乳糖苷酶的基因相同取向。这一质粒称为 pBSA。从 pBSA 中纯化 XbaI-HindIII DNA 片断，插入用 XbaI-HindIII 消化的 pUCB 中，产生质粒 pUCA-B。最后，在 pUCA-B 的唯一 XbaI 位点插入含有构巢曲霉的 *argB*⁻ 基因的 3.2kb 的 DNA 片断，产生 pPhacA::*argB*⁻。从 pUC18 的 *lacZ* 启动子开始，pPhacA::*argB*⁻ 包括 *phacA* 基因的 5' 区的 0.94kb 的片断，随后接着质粒的基因组序列开始的 297 个密码子，含有 *argB*⁻ 基因的 3.2kb 的片断，对应于密码子 393 - 518 的 *phacA* 基因的基因组序列和 *phacA* 基因的 3' 区 1.2kb。通过 EcoRI 消化，从质粒中纯化出含有所有这些区域的线性片断 (图 6)。

同样，转化方法可以用于构建含有破坏一缺失突变的 *phacA* 基因的构巢曲霉的菌株，其中利用了上面的线性 EcoRI 片断和上面所述的图中概括的方案。为了这一目的，利用 2 微克所述的 DNA 片断转化构巢曲霉 *biA1*, *methG1*, *argB2* 的菌株的原生质体，所用的方案来自 Tilburn, J., Scazzocchio, C., Taylor, G.G., Zabicky-Zissman, J.H., Lockington, R.A. 和 Davies, R.W. (1983) 基因 26: 205 - 211。利用适当的选择培养基选择精氨酸原养型转化体，然后纯化。从对应于一系列转化体的菌丝体中分离 DNA，用 PstI 消化，利用特异的 *phacA* 和 *argB* 基因探针，通过 Southern 杂交分析。如图 1 所示，许多转化体显示了希望得到的成功的双互交的杂交方式。例如，利用 *phacA* 基因的 0.9kb 的 PstI 片断作为探

针(参见图 4), 发现在转化体中 3.8kb 的带已经替代了受体菌株中单一的 0.9kb PstI 杂交带, 这展示了图 1 显示的整合类型。这一带也与 argB 探针杂交。选择这些转化体中的两个, 随后鉴定, 命名为 Δ phacA#1 和 Δ phacA#2。这些转化体菌株在将丙氨酸苯酯, 2-羟基乙酸苯酯或 2, 5-一二羟基乙酸苯酯作为唯一碳源的培养基中具有生长能力, 但与此相反, 它们在将乙酸苯酯作为唯一碳源的培养基中不能生长。这支持了这样的结论, 即构巢曲霉的 phacA 基因编码的细胞色素 P450 酶(扩展到产黄青霉的 pahA 基因编码的一个)具有将乙酸苯酯羟基化成为 2-羟基乙酸苯酯的活性, 这一活性催化构巢曲霉中乙酸苯酯利用途径的第一个步骤(参见以前的技术部分)。构巢曲霉菌株 Δ phacA#1 与 Δ phacA#2 在表现型上有区别, 与构巢曲霉 biA1 veA1 methG1 argB2 phacA::[pPhacA::argB⁺]一样, 构巢曲霉菌株 Δ phacA#1 于 1996 年 6 月 19 日保藏在西班牙培养物类型收藏处(CECT), Valencia 大学, 研究大楼, Burjasot Campus, 46100, Valencia, 登记号为 CECT 20195。获得含有失活的 pahA 基因的产黄青霉转化体的方法是本领域任何技术人员所非常熟悉的。简单地, 该方法将包括分离产黄青霉的 pahA 基因, 方法是与存在于转化体 CECT 20195 中的构巢曲霉的 phacA 基因杂交或选择性地与根据 SEQ ID NO:1 合成的低聚核苷酸杂交。如本发明的图 7 所示, 利用质粒 pAlfleo7(保藏于西班牙培养物类型收藏处, Valencia 大学, 研究大楼, Burjasot Campus, 46100, Valencia, 1997 年 2 月 20 日, 登记号为 CECT 4849), 可以构建质粒 pALP696。

实施例 4: 失活 phacA 基因的功能和增加青霉素产量的实验

为了证实是否去除由 phacA 基因编码的乙酸苯酯 2-羟化酶活性可以使青霉素产量提高, 我们进行了实验, 测量菌株 Δ phacA 生产的青霉素水平并和菌株 phacA⁺比较。在烧瓶中的最小培养基中发酵比较两个菌株, 最小培养基含有 2.5%(w/v)乳糖作为基本碳源, 和 2.5%(w/v)玉米浸渍液, 和如标明的不同量的乙酸苯酯钠(w/v)。在所有情况中, 都用相等数目的活的分生孢子(2×10^6 个/毫升)接种培养物, 在 37 °C, 有规律地搅拌(250rpm)温育。在各种时间取样, 通过 Miracloth 过滤, 上清液用于生物测试, 测量生产的青霉素。

在生物测试中，将生长到 $OD^{600} = 6$ 的藤黄微球菌的培养物 1 毫升与 Antibiotic 1 号培养基(Difco) 1 升在 50 °C 混合。将这一混合物注入 13.6 厘米的培养皿(65 毫升/盘)中。将 100 微升来自培养物的上清液的样品(适当用 10 毫摩尔/升, pH6.8 磷酸钠缓冲液稀释)分配在 8 毫米直径的孔中。平板在 4 °C, 温育 2 小时, 然后在 37 °C 再温育 22 小时。最后, 测量由于样品中存在的青霉素引起的细菌生长抑制区的直径, 与校正线比较估算抗生素的量, 校正线是利用不同量的青霉素 G 钾制作的。

图 3 显示了在生产培养基中利用 0.125% 的乙酸苯酯培养 24 小时后, 得到与菌株 $phacA^+$ 相对应的菌株最高水平的青霉素产量。这时得到的青霉素产量是 1.8 微克/毫升。将乙酸苯酯的浓度减半, 使这一菌株产生的青霉素的最高水平减少到 1.1 微克/毫升。

图 3 也显示, 在菌株 $\Delta phacA$ 的培养物中青霉素的最高水平与亲本菌株有明显差别, 达到 5.6 微克/毫升, 即比亲本菌株高 3.1 倍。另外, 作为失活 $phacA$ 基因的结果, 将乙酸苯酯的开始浓度减少到 0.0625% 不影响青霉素产量的增加。

结论是如图 1 所示通过反向遗传学技术进行编码乙酸苯酯 2-羟化酶活性的 $phacA$ 基因的破坏—缺失突变, 至少在构巢曲霉生产青霉素过程中产生两个优点: (i) 青霉素水平有相当大的提高和(ii) 青霉素水平更少地依赖于乙酸苯酯的外部供应。

附图的详细说明

图 1: 通过同源重组在构巢曲霉的 $phacA$ 基因中产生破坏—缺失突变的方案。用黑体的大叉表示推断的交互连接。

图 2: 通过同源重组在产黄青霉的 $pahA$ 基因中产生破坏—缺失突变的方案。大的虚线叉表示了推断的交互连接。

图 3: 与野生型菌株($phacA^+$)比较构巢曲霉的转化 $phacA::argB$ 菌株的青霉素产量。横坐标: 时间, 小时(h); 纵坐标: 青霉素产量(微克/毫升)。连续的实线: 乙酸苯酯浓度为 0.125%; 断线: 乙酸苯酯浓度为 0.0625%。

图 4: 含有构巢曲霉的 $phacA$ 基因的基因组区的限制图谱, 显示了亚克隆产生 pFG4A 和 pFG4B 的片断, 黑区是 3 个内含子。

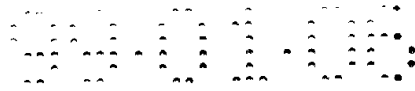


图 5：含有产黄青霉的 *pahA* 基因的基因组区的限制图谱。显示了亚克隆产生 pALP520 的 2.55kb 的 XhoI 片断。黑区是 3 个内含子。

图 6：pPhacA::argB^r失活构建体的限制图谱和相关的特征。黑区是含有 *phacA* 基因的测序区；空白区是在 *phacA* 基因前后的侧接区；条纹区是 *argB^r* 区。

图 7：pALP696 失活构建体的限制图谱和相关特征。黑区是含有 *pahA* 基因的测序区(1745bp)；空白区是 *pahA* 基因前后的侧接区(4143bp)；条纹区是 *bleomycin* 抗性区，基因 *ble^R*(2048bp)。

序 列 表

总的情况：

申请人：

姓名：抗生素有限公司 (ANTIBIOTICOS, S. A. U)

街道：Avda. de Burgos, 8-A

城市：马德里

州或省：马德里

国家：西班牙

邮政编码：28036

电话：91 - 3841200

传真：91 - 3841220

题目：“使编码乙酸苯酯分解代谢酶的基因失活的方法，包括质粒，和用它们转化的菌株”

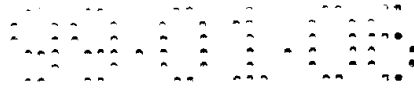
序列数：4

通讯地址：

收信人：抗生素有限公司 (ANTIBIOTICOS, S. A. U)

街道：Avda. de Burgos, 8-A

城市：马德里



州或省：马德里
国家：西班牙
邮政编码：28036

机器可读形式：

介质类型：3.5”软盘
计算机：PC
操作系统：WINDOWS
软件：WORD

律师/代理人的情况：

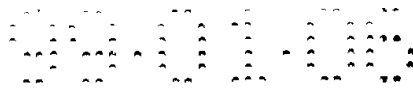
姓名：阿伯特 伊丽沙白 (ALBERTO DE ELZABURU)
登记号：232/1
登记号：PCT - 42

电信资料：

电话：91 7009400
传真：91 3193810
电子信箱：elzaburu@elzaburu.es

SEQ ID NO:1 的特征：

长度：1986 个碱基对
类型：核苷酸
链数：2
构型：线性
分子类型：基因组 DNA
假设：没有
反义链：没有
原始来源：构巢曲霉
直接来源：质粒 pPhacA::argB



基因组的位置：未知

特征：

名称/关键词：编码序列

定位：820 ... 1810

其它信息：phacA 基因

序列描述：SEQ ID NO： 1

```

C CAGCCAGATA TAAAGGCCGT 21
CCTTAGACGA CTTGATCTTA CTCTGGTTTC TAAAAGTTCA CTGATTATCG CAAGGTATCC 81
ATG TCT CTT CAA ACA ATC GGG ATC GCC GCT GTC GCG GTG GTC TAT TTT 129
Met Ser Leu Gln Thr Ile Gly Ile Ala Ala Val Ala Val Val Tyr Phe
      5                10                15
CTC ATC CGC TAC TTC AAC CGC ACA GAC ATC CCA AAG ATC AAG GGT CTC 177
Leu Ile Arg Tyr Phe Asn Arg Thr Asp Ile Pro Lys Ile Lys Gly Leu
      20                25                30
CCT GAA GTT CCA GGC GTA CCG ATC TTT GGC AAC CTC ATC CAG CTC GGT 225
Pro Glu Val Pro Gly Val Pro Ile Phe Gly Asn Leu Ile Gln Leu Gly
      35                40                45
GAC CAG CAC GCG ACC GTA GCG CAG AAA TGG GCG AAG AAA TTT GGA CCT 273
Asp Gln His Ala Thr Val Ala Gln Lys Trp Ala Lys Lys Phe Gly Pro
      50                55                60
GTT TTC CAG GTT CGC ATG GGG AAT AAA GTGAGTTCAG TGTCTGCATT TGTAAC 326
Val Phe Gln Val Arg Met Gly Asn Lys
65                70
```



AGAC GACAAATATTG CGAGAATAAT TCTGACCTAA CACAG CGC GTT GTC TTC GCA 380
Arg Val Val Phe Ala
75

AAC ACC TTC GAC TCT GTC CGT CAG CTA TGG ATC AAA GAT CAG TCC GCG 428
Asn Thr Phe Asp Ser Val Arg Gln Leu Trp Ile Lys Asp Gln Ser Ala
80 85 90

CTC ATC TCC CGG CCG ACC TTT CAC ACC TTC CAC AGT GTA GTT TCC AGC 476
Leu Ile Ser Arg Pro Thr Phe His Thr Phe His Ser Val Val Ser Ser
95 100 105 110

TCT CAG GGA TTC ACC ATC GGA ACG TCG CCG TGG GAC GAG TCG TGC AAG 524
Ser Gln Gly Phe Thr Ile Gly Thr Ser Pro Trp Asp Glu Ser Cys Lys
115 120 125

CGT CGT CGG AAG GCT GCA GCT ACA GCC TTG AAC CGC CCG GCT ACC CAG 572
Arg Arg Arg Lys Ala Ala Ala Thr Ala Leu Asn Arg Pro Ala Thr Gln
130 135 140

TCG TAT ATG CCT ATT ATC GAT CTT GAG TCG ATG TCG AGT ATC CGG GAA 620
Ser Tyr Met Pro Ile Ile Asp Leu Glu Ser Met Ser Ser Ile Arg Glu
145 150 155

TTG CTC AGG GAT AGC GCG AAT GGA ACA ATG GAT ATC AAC CCG ACA GCG 668
Leu Leu Arg Asp Ser Ala Asn Gly Thr Met Asp Ile Asn Pro Thr Ala
160 165 170

TAC TTC CAG CGG TTC GCG TTG AAC ACG AGC TTA ACA TTG AAC TAT GGA 716
Tyr Phe Gln Arg Phe Ala Leu Asn Thr Ser Leu Thr Leu Asn Tyr Gly
175 180 185 190

ATC CGA ATC GAG GGC AAT GTG AAC GAT GAG CTT TTG CGC GAA ATT GTC 764
Ile Arg Ile Glu Gly Asn Val Asn Asp Glu Leu Leu Arg Glu Ile Val
195 200 205


GAT GTC GAG CGC GGG GTG TCG AAC TTC CGG AGT ACC AGC AAC CAG TGG 812
Asp Val Glu Arg Gly Val Ser Asn Phe Arg Ser Thr Ser Asn Gln Trp
210 215 220

CAG GAC TAT ATC CCG CTC CTG AGA ATC TTC CCG AAG ATG AAC CGC GAG 860
Gln Asp Tyr Ile Pro Leu Leu Arg Ile Phe Pro Lys Met Asn Arg Glu
225 230 235

GCT GAG GAG TTC CGG GTG CGG AGA GAC AAG TAT CTT ACC TAT CTT TTG 908
Ala Glu Glu Phe Arg Val Arg Arg Asp Lys Tyr Leu Thr Tyr Leu Leu
240 245 250

GAT GTT CTC AAG GAT CGC ATT GCA AAG GGA ACC GAC AAG CCC TGT ATT 956
Asp Val Leu Lys Asp Arg Ile Ala Lys Gly Thr Asp Lys Pro Cys Ile
255 260 265 270

ACT GGA AAC ATC CTC AAA GAC CCT GAG GCT AAG CTC AAT GAT G GTATG 1004
Thr Gly Asn Ile Leu Lys Asp Pro Glu Ala Lys Leu Asn Asp
275 280



CCGTC CGGTCCTGTC CGCGCTTGAA GGAAAATAAA GATTAACAGA GTCTAG CC GAG 1060
Ala Glu
285

ATT AAA TCG ATC TGC TTG ACC ATG GTC TCC GCC GGC CTC GAT ACT GTT 1108
 Ile Lys Ser Ile Cys Leu Thr Met Val Ser Ala Gly Leu Asp Thr Val
290 295 300

CCG GGC AAC TTG ATC ATG GGC ATC GCG TAC CTC GCC TCC GAA GAC GGC 1156
 Pro Gly Asn Leu Ile Met Gly Ile Ala Tyr Leu Ala Ser Glu Asp Gly
305 310 315

CAA AGG ATC CAG AAG CGC GCC CAC GAC GAG ATC ATG AAA GTC TAC CCG 1204
 Gln Arg Ile Gln Lys Arg Ala His Asp Glu Ile Met Lys Val Tyr Pro
320 325 330

GAC GGC GAT GCA TGG GAG AAA TGC CTG CTC GAA GAG AAA GTC CCC TAC 1252
 Asp Gly Asp Ala Trp Glu Lys Cys Leu Leu Glu Glu Lys Val Pro Tyr
335 340 345 350

GTC ACA GCT TTG GTC AAA GAA ACC CTC CGC TTC TGG ACT GTC ATT CCC 1300
 Val Thr Ala Leu Val Lys Glu Thr Leu Arg Phe Trp Thr Val Ile Pro
355 360 365

ATC TGT CTG CCT AGA GAA AAC ACC AAG GAT ATT GTC TGG AAC GGA GCC 1348
 Ile Cys Leu Pro Arg Glu Asn Thr Lys Asp Ile Val Trp Asn Gly Ala
370 375 380

GTT ATC CCT AAG GGA ACG ACC TTT TTC ATG AAC GCC TAC GCT GCA GAC 1396
 Val Ile Pro Lys Gly Thr Thr Phe Phe Met Asn Ala Tyr Ala Ala Asp
385 390 395

TAC GAC GAA ACA CAC TTC ACC AAT CCA CAC GCC TTT GAA CCA GAA CGC 1444
 Tyr Asp Glu Thr His Phe Thr Asn Pro His Ala Phe Glu Pro Glu Arg
400 405 410

TAC CTC ACC GCC TCA TCT GAC GGC TCC GGC ACT CCA CAC TAC GGC TAC 1492
 Tyr Leu Thr Ala Ser Ser Asp Gly Ser Gly Thr Pro His Tyr Gly Tyr
415 420 425 430

GGC GCG GGC TCG CGC ATG GTACCTCCCC CACAACCCAC ATGTTATATA GACAAT 1546
 Gly Ala Gly Ser Arg Met
435

ACTG ATTTATTATG AAG TGC GCT GGC TCG CAT CTT GCG AAC CGC GAG CTC 1596
Cys Ala Gly Ser His Leu Ala Asn Arg Glu Leu
440 445

TTC ACC GCT TAC GTC CGT CTG ATC ACA GCA TTT ACC ATG CAT CCA GCT 1644
 Phe Thr Ala Tyr Val Arg Leu Ile Thr Ala Phe Thr Met His Pro Ala
450 455 460

AAG AGG GCT GAG GAT CGG CCG ATC CTC GAT GCG ATT GAG TGT AAT GCG 1692
 Lys Arg Ala Glu Asp Arg Pro Ile Leu Asp Ala Ile Glu Cys Asn Ala
465 470 475

```

ATC CCG ACC GCT TTG ACG ACG GAG CCG AAG CCC TTC AAG GTT GGG TTT      1740
Ile Pro Thr Ala Leu Thr Thr Glu Pro Lys Pro Phe Lys Val Gly Phe
480                      485                      490                      495
AAA CCG AGA GAT CCT GTC TTG GTA AGG AAA TGG ATT GCG GAG AGT GAG      1788
Lys Pro Arg Asp Pro Val Leu Val Arg Lys Trp Ile Ala Glu Ser Glu
                    500                      505                      510
GAG AGG ACG AAG CAC CTG AAT TAGTTTTTCT TTTCTTTCTT GCGTGCAAGT TAC      1842
Glu Arg Thr Lys His Leu Asn
                    515
AGCGCAATTT ATACTTAGCT GGGACATGGT CGATTGGAGT ATACTGATTT CAGCAACAGC      1902
GCAAATAAAAA ATAAGGCAAC CAATAGAAAT GCAACAATAT GCAATCTCCC AAGACCATTA      1962
AATGGTTGAT CAGATGATCC CAAG                                          1986
    
```

SEQ ID NO: 2 的资料:

序列的特征:

长度: 518 个氨基酸

类型: 氨基酸

链数: 1

构型: 线性

分子类型: 肽

原始来源: 构巢曲霉

特征:

其它信息: 乙酸苯酯 2-羟化酶的氨基酸序列

特征:

其它信息: 分子量 58495 道尔顿

序列描述: SEQ ID NO: 2

```

Met Ser Leu Gln Thr Ile Gly Ile Ala Ala Val Ala Val Val Tyr
                    5                      10                      15
Phe Leu Ile Arg Tyr Phe Asn Arg Thr Asp Ile Pro Lys Ile Lys
                    20                      25                      30
Gly Leu Pro Glu Val Pro Gly Val Pro Ile Phe Gly Asn Leu Ile
                    35                      40                      45
    
```

1000

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

2558

Phe	Trp	Thr	Val	Ile	Pro	Ile	Cys	Leu	Pro	Arg	Glu	Asn	Thr	Lys
				365					370					375
Asp	Ile	Val	Trp	Asn	Gly	Ala	Val	Ile	Pro	Lys	Gly	Thr	Thr	Phe
				380					385					390
Phe	Met	Asn	Ala	Tyr	Ala	Ala	Asp	Tyr	Asp	Glu	Thr	His	Phe	Thr
				395					400					405
Asn	Pro	His	Ala	Phe	Glu	Pro	Glu	Arg	Tyr	Leu	Thr	Ala	Ser	Ser
				410					415					420
Asp	Gly	Ser	Gly	Thr	Pro	His	Tyr	Gly	Tyr	Gly	Ala	Gly	Ser	Arg
				425					430					435
Met	Cys	Ala	Gly	Ser	His	Leu	Ala	Asn	Arg	Glu	Leu	Phe	Thr	Ala
				440					445					450
Tyr	Val	Arg	Leu	Ile	Thr	Ala	Phe	Thr	Met	His	Pro	Ala	Lys	Arg
				455					460					465
Ala	Glu	Asp	Arg	Pro	Ile	Leu	Asp	Ala	Ile	Glu	Cys	Asn	Ala	Ile
				470					475					480
Pro	Thr	Ala	Leu	Thr	Thr	Glu	Pro	Lys	Pro	Phe	Lys	Val	Gly	Phe
				485					490					495
Lys	Pro	Arg	Asp	Pro	Val	Leu	Val	Arg	Lys	Trp	Ile	Ala	Glu	Ser
				500					505					510
Glu	Glu	Arg	Thr	Lys	His	Leu	Asn							
				515										

SEQ ID NO: 3 的资料:

序列的特征

长度: 2558 个碱基对

类型: 核苷酸

链数: 2

构型: 线性

分子类型: 基因组 DNA

假设: 没有

反义链: 没有

原始来源: 产黄青霉

直接来源: 质粒 pALP520

基因组中的位置: 未知

特征:

名称/关键词: 编码序列

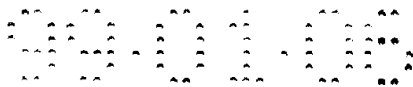


定位: 371 ... 2094

其它信息: pahA 基因

序列描述: SEQ ID NO: 3

```
CTCGAGAAGG TCTGTTGACA CGAGTCCACA AGATGTTGCG AAGTATATAA TATGAGAAAT 60
GGATATACAG CGGAAAATTG AAGTTTCATT CCGCGGGGGC GGGGAACAAG AGGAAGCGCG 120
GAATGAACAT TCCGTCTATG CGCGGTGGAG AGATCTTCGA ACTGGGGACT TGTGGACTTG 180
GACCTATGAG AGAGCTGCCT TTA CTGATTG TGGGACTTGT GGCCAATGAA ACATGTCATG 240
GATGTTACAG CTTTCTTTAT GCGACTCTCT AA ACTCGGAA ATATGCCGTA GCCGAAATGC 300
AGGAAAGGCT GGGCTATAAG AGATGCATGC TCCCACCGAA CGTCTGCAAT TCCTTGTTTC 360
ACTCTATATC ATG GCC ATC CAA ACA CTC GCA GTC GCC GTG ATC ACG GTG GTC 412
Met Ala Ile Gln Thr Leu Ala Val Ala Val Ile Thr Val Val
                    5                      10
TAT TTC GTC ATT CGA TAC TTC AAC CGC ACT GAT ATC CCT AAG ATT AAA 460
Tyr Phe Val Ile Arg Tyr Phe Asn Arg Thr Asp Ile Pro Lys Ile Lys
15                      20                      25                      30
GGC CTC CCG GAG ATT CCT GGT ATA CCC ATA TTT GGC AAT CTA TTG CAG 508
Gly Leu Pro Glu Ile Pro Gly Ile Pro Ile Phe Gly Asn Leu Leu Gln
35                      40                      45
CTA GGA GAT CAA CAT GCC ACA GTC ACG GGG AAA TGG GCA AAG AAA TTT 556
Leu Gly Asp Gln His Ala Thr Val Thr Gly Lys Trp Ala Lys Lys Phe
50                      55                      60
GGC CCA GTT TTC CAA GTG CGC ATG GGA AAC AAG GTAAGTAATC AAATAATCTT 609
Gly Pro Val Phe Gln Val Arg Met Gly Asn Lys
65                      70
TTAAATCGGA CATTCTGATA ATCTAATGCC CTTTTAG CGC ATC GTG TTC GCC AAC 664
Arg Ile Val Phe Ala Asn
75
GGC TTT GAC TCC GTT CGT CAA TTG TGG ATT AAG GAC TCG TCG GCT CTG 712
Gly Phe Asp Ser Val Arg Gln Leu Trp Ile Lys Asp Ser Ser Ala Leu
80                      85                      90                      95
ATC TCC CGC CCA ACT TTC CAC ACT TTC CAC AGC GTC GTC TCC AGT TCA 760
Ile Ser Arg Pro Thr Phe His Thr Phe His Ser Val Val Ser Ser Ser
100                      105                      110
CAG GGC TTC ACG ATC GGA ACT TCC CCG TGG GAT GAT TCC TGT AAG AAA 808
Gln Gly Phe Thr Ile Gly Thr Ser Pro Trp Asp Asp Ser Cys Lys Lys
115                      120                      125
```



CGT CGC AAG GCA GCT GCC ACT GCG CTG AAT CGA CCG GCC GTG CAA TCG 856
Arg Arg Lys Ala Ala Ala Thr Ala Leu Asn Arg Pro Ala Val Gln Ser
130 135 140

TAT ATG CCC ATC ATC GAT CTT GAA TCC AAT TCC AGC ATC AAA GAA CTG 904
Tyr Met Pro Ile Ile Asp Leu Glu Ser Asn Ser Ser Ile Lys Glu Leu
145 150 155

TAC CGG GAC AGC CAA AAT GGC AAA CGT GAT GTG AAT CCC ACT GCA TAC 952
Tyr Arg Asp Ser Gln Asn Gly Lys Arg Asp Val Asn Pro Thr Ala Tyr
160 165 170 175

TTC CAG CGA TAC GCT TTC AAC ACC AGT TTG ACT TTG AAC TAT GGA TTC 1000
Phe Gln Arg Tyr Ala Phe Asn Thr Ser Leu Thr Leu Asn Tyr Gly Phe
180 185 190

CGC ATC GAG GGC AAT GTG GAT GAT ACG CTG CTG CAT GAG ATT GTG GAT 1048
Arg Ile Glu Gly Asn Val Asp Asp Thr Leu Leu His Glu Ile Val Asp
195 200 205

GTG GAG CGT GGT GTG TCC AAC TTC CGC AGC ACT TCG AAC AAC TGG CAG 1096
Val Glu Arg Gly Val Ser Asn Phe Arg Ser Thr Ser Asn Asn Trp Gln
210 215 220

GAC TAC ATT CCC CTA TTG CGC ATT TTC CCC AAG ATG AAC AAT GAG GCC 1144
Asp Tyr Ile Pro Leu Leu Arg Ile Phe Pro Lys Met Asn Asn Glu Ala
225 230 235

GCT GAC TTC CGG GGT CGC CGT GAT AAA TAT CTG ACC TAC TTG CTC GAT 1192
Ala Asp Phe Arg Gly Arg Arg Asp Lys Tyr Leu Thr Tyr Leu Leu Asp
240 245 250 255

ATG CTC AAG GAT CGA ATC GCC AAG GGA ACC GAT AAG CCT TGC ATC ACT 1240
Met Leu Lys Asp Arg Ile Ala Lys Gly Thr Asp Lys Pro Cys Ile Thr
260 265 270

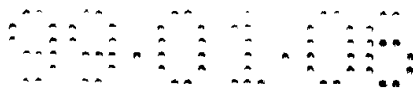
GGT AAT ATC TTG AAG GAT CCC GAG GCT AAG CTG AAT GAT G GTGAGTTACA 1290
Gly Asn Ile Leu Lys Asp Pro Glu Ala Lys Leu Asn Asp
275 280

AAATCTCGCG AAATCTTTGT GAATGTTGGT ATTGAGTACT CATCTATCGT CTAG CC 1346
Ala
285

GAG GTG AAA TCG ATC TGT TTG ACT ATG GTG TCG GCT GGC CTT GAC ACC 1394
Glu Val Lys Ser Ile Cys Leu Thr Met Val Ser Ala Gly Leu Asp Thr
290 295 300

GTT CCG GGC AAC TTG ATT ATG GGA ATT GCC TAC CTG GCA TCT GAA GAC 1442
Val Pro Gly Asn Leu Ile Met Gly Ile Ala Tyr Leu Ala Ser Glu Asp
305 310 315

GGC CAA AGA ATT CAG AAA AAG GCC TAT GAT GCA ATC ATG GAG GTA TAC 1490
Gly Gln Arg Ile Gln Lys Lys Ala Tyr Asp Ala Ile Met Glu Val Tyr
320 325 330



CCG GAC GGC GAT GCT TGG GAG AAA TGT TTG GTG GAG GAG AAG GTC CCT 1538
Pro Asp Gly Asp Ala Trp Glu Lys Cys Leu Val Glu Glu Lys Val Pro
335 340 345

TAT GTG ACG GCA CTG GTC AAG GAG GTC TTG CGC TTC TGG ACG GTC ATT 1586
Tyr Val Thr Ala Leu Val Lys Glu Val Leu Arg Phe Trp Thr Val Ile
350 355 360 365

CCT ATT TGT CTG CCA CGC GAG AGC ACC AAG GAT ATC CAG TGG AAT GGA 1634
Pro Ile Cys Leu Pro Arg Glu Ser Thr Lys Asp Ile Gln Trp Asn Gly
370 375 380

GCT ACA ATC CCT GCC GGG ACG ACC TTT TTC ATG GTATGTTGCG CCTTGCTTCT 1687
Ala Thr Ile Pro Ala Gly Thr Thr Phe Phe Met
385 390

GTCCCTTTTCG GGACGTTGCT AACAGGATCT GATAG AAT GTT TGG GCT GCC GAT 1740
Asn Val Trp Ala Ala Asp
395

TAC GAT GAA GAT CAC TTT AAG GAT GCC GAT AAA TTC ATC CCG GAG CGT 1788
Tyr Asp Glu Asp His Phe Lys Asp Ala Asp Lys Phe Ile Pro Glu Arg
400 405 410

TAT TTG GAG GCC AGT GAG GGT GCA GGC ACT CCA CAC TAT GCA TAT GGA 1836
Tyr Leu Glu Ala Ser Glu Gly Ala Gly Thr Pro His Tyr Ala Tyr Gly
415 420 425 430

GCG GGA TCA CGT ATG TGT GCA GGC TCA CAT CTC GCC AAT CGC GAG CTG 1884
Ala Gly Ser Arg Met Cys Ala Gly Ser His Leu Ala Asn Arg Glu Leu
435 440 445

TTC ACT GCT TTT ATC CGC CTC GTC ACT GCT TTC AAC ATG CAC ACG GCG 1932
Phe Thr Ala Phe Ile Arg Leu Val Thr Ala Phe Asn Met His Thr Ala
450 455 460

AAG GAG ACG GCC GAC CGA CCG ATT CTG AAT GCA ATT GAG TGC AAT TTG 1980
Lys Glu Thr Ala Asp Arg Pro Ile Leu Asn Ala Ile Glu Cys Asn Leu
465 470 475

ATT CCG ACA GCC TTG ACA ACC GAG CCG AAG CCA TTC AAG GTT GGC TTT 2028
Ile Pro Thr Ala Leu Thr Thr Glu Pro Lys Pro Phe Lys Val Gly Phe
480 485 490

AGT GCA CGC GAC CCT AAG AAG CTT GAG CAG TGG ATT GCT GAG AGC GAT 2076
Ser Ala Arg Asp Pro Lys Lys Leu Glu Gln Trp Ile Ala Glu Ser Asp
495 500 505 510

GAA CGG ACC AAA GAT CTA TAAGAGGAAG TTTTATCTGA TATGATTCCC TTTTTTTTTT
Glu Arg Thr Lys Asp Leu
515

TCGGAGGCTA TTTATGTATC TACTTGAATA TATGTGAAAT AAAGGCAATG AAGTATAGAT 2194
ATAGACCTGC CCAGGTGAGA CCTACATGTA TGTAATCGAC GTCTCCCAGT ACCTATTTAG 2254
GAACCGAATA GAAATTTGAG TGATGAAGGG TGAAGAATAA CCTCAGAGAA CATAGGTCTA 2314



CTAAGATCTG CTAATTCTTA GACTCCTTTC CCTGAATTAT TGCCAATTTTC CCCAGTTGTC 2374
TCTCTCCTCT ATGACTCCCG GCCTTGTCAC ACCGGCGGAA TCTCCCCGCA TTTTCCTCGA 2434
CCTCACCCCTT TTCTTCCTCT TCACCCTCTC CCCATCCACT TGAAATGAAC CTCTGATCGC 2494
AATCAATTGC ATCCCGAAGT CATTTTGGAT GATGAATAAT CCGCAGATGT ACCCTGTCCT 2554
CGAG 2558

SEQ ID NO: 4 的资料:

序列的特征:

长度: 516 个碱基对

类型: 氨基酸

链数: 1

构型: 线性

分子类型: 肽

特征:

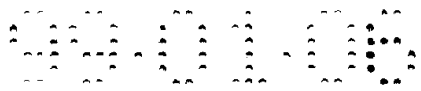
其它信息: 乙酸苯酯 2-羟化酶的氨基酸序列

特征:

其它信息: 分子量 58112 道尔顿

序列的描述: SEQ ID NO: 4

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



Arg Lys Ala Ala Ala Thr Ala Leu Asn Arg Pro Ala Val Gln Ser Tyr
130 135 140

Met Pro Ile Ile Asp Leu Glu Ser Asn Ser Ser Ile Lys Glu Leu Tyr
145 150 155 160

Arg Asp Ser Gln Asn Gly Lys Arg Asp Val Asn Pro Thr Ala Tyr Phe
165 170 175

Gln Arg Tyr Ala Phe Asn Thr Ser Leu Thr Leu Asn Tyr Gly Phe Arg
180 185 190

Ile Glu Gly Asn Val Asp Asp Thr Leu Leu His Glu Ile Val Asp Val
195 200 205

Glu Arg Gly Val Ser Asn Phe Arg Ser Thr Ser Asn Asn Trp Gln Asp
210 215 220

Tyr Ile Pro Leu Leu Arg Ile Phe Pro Lys Met Asn Asn Glu Ala Ala
225 230 235 240

Asp Phe Arg Gly Arg Arg Asp Lys Tyr Leu Thr Tyr Leu Leu Asp Met
245 250 255

Leu Lys Asp Arg Ile Ala Lys Gly Thr Asp Lys Pro Cys Ile Thr Gly
260 265 270

Asn Ile Leu Lys Asp Pro Glu Ala Lys Leu Asn Asp Ala Glu Val Lys
275 280 285

Ser Ile Cys Leu Thr Met Val Ser Ala Gly Leu Asp Thr Val Pro Gly
290 295 300

Asn Leu Ile Met Gly Ile Ala Tyr Leu Ala Ser Glu Asp Gly Gln Arg
305 310 315 320

Ile Gln Lys Lys Ala Tyr Asp Ala Ile Met Glu Val Tyr Pro Asp Gly
325 330 335

Asp Ala Trp Glu Lys Cys Leu Val Glu Glu Lys Val Pro Tyr Val Thr
340 345 350

Ala Leu Val Lys Glu Val Leu Arg Phe Trp Thr Val Ile Pro Ile Cys
355 360 365

Leu Pro Arg Glu Ser Thr Lys Asp Ile Gln Trp Asn Gly Ala Thr Ile
370 375 380

Pro Ala Gly Thr Thr Phe Phe Met Asn Val Trp Ala Ala Asp Tyr Asp
385 390 395 400

Glu Asp His Phe Lys Asp Ala Asp Lys Phe Ile Pro Glu Arg Tyr Leu
405 410 415

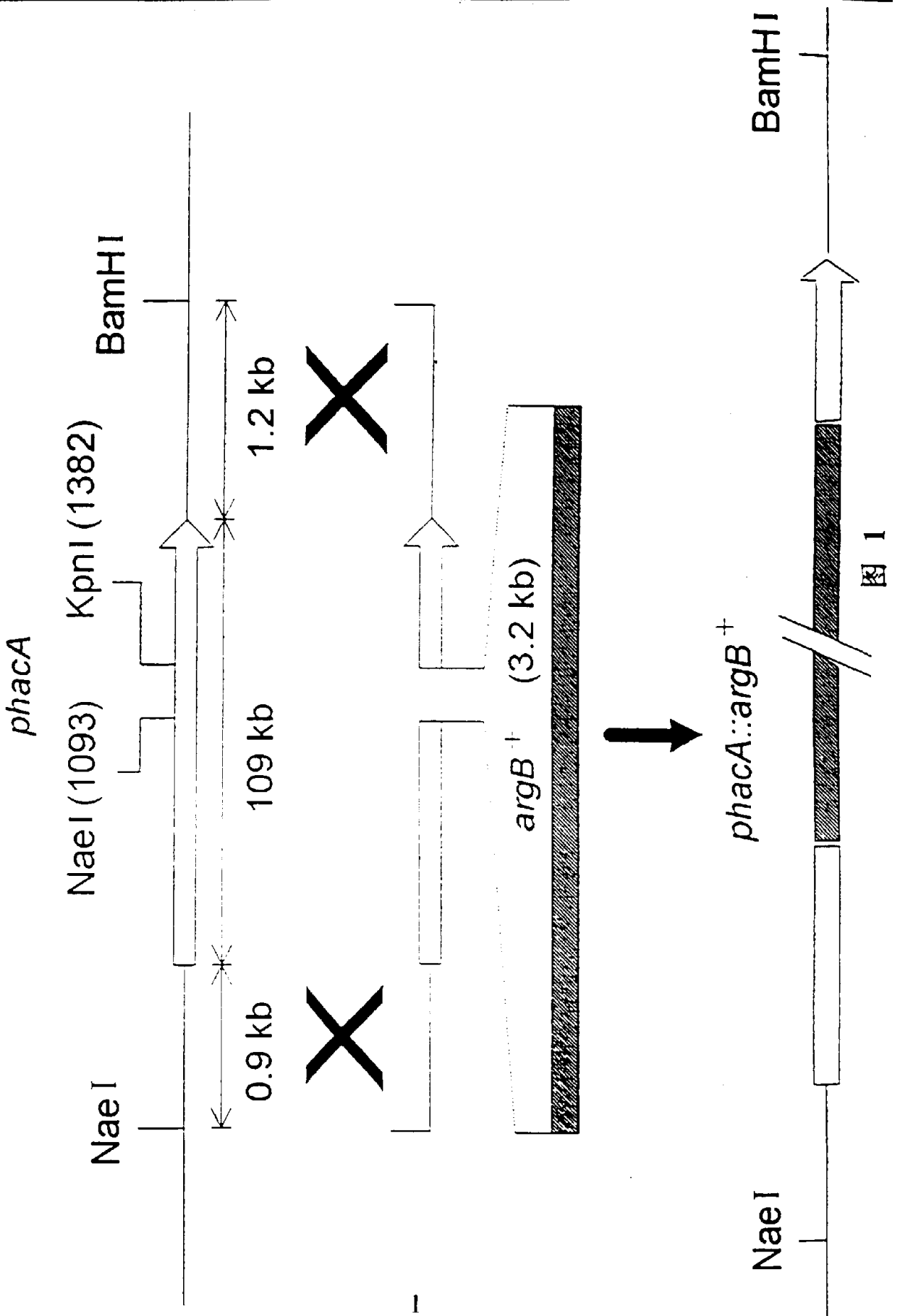
Glu Ala Ser Glu Gly Ala Gly Thr Pro His Tyr Ala Tyr Gly Ala Gly
420 425 430

Ser Arg Met Cys Ala Gly Ser His Leu Ala Asn Arg Glu Leu Phe Thr
435 440 445

Ala Phe Ile Arg Leu Val Thr Ala Phe Asn Met His Thr Ala Lys Glu
450 455 460

9010

Thr	Ala	Asp	Arg	Pro	Ile	Leu	Asn	Ala	Ile	Glu	Cys	Asn	Leu	Ile	Pro
465					470					475					480
Thr	Ala	Leu	Thr	Thr	Glu	Pro	Lys	Pro	Phe	Lys	Val	Gly	Phe	Ser	Ala
				485					490					495	
Arg	Asp	Pro	Lys	Lys	Leu	Glu	Gln	Trp	Ile	Ala	Glu	Ser	Asp	Glu	Arg
			500					505					510		
Thr	Lys	Asp	Leu												
			515												



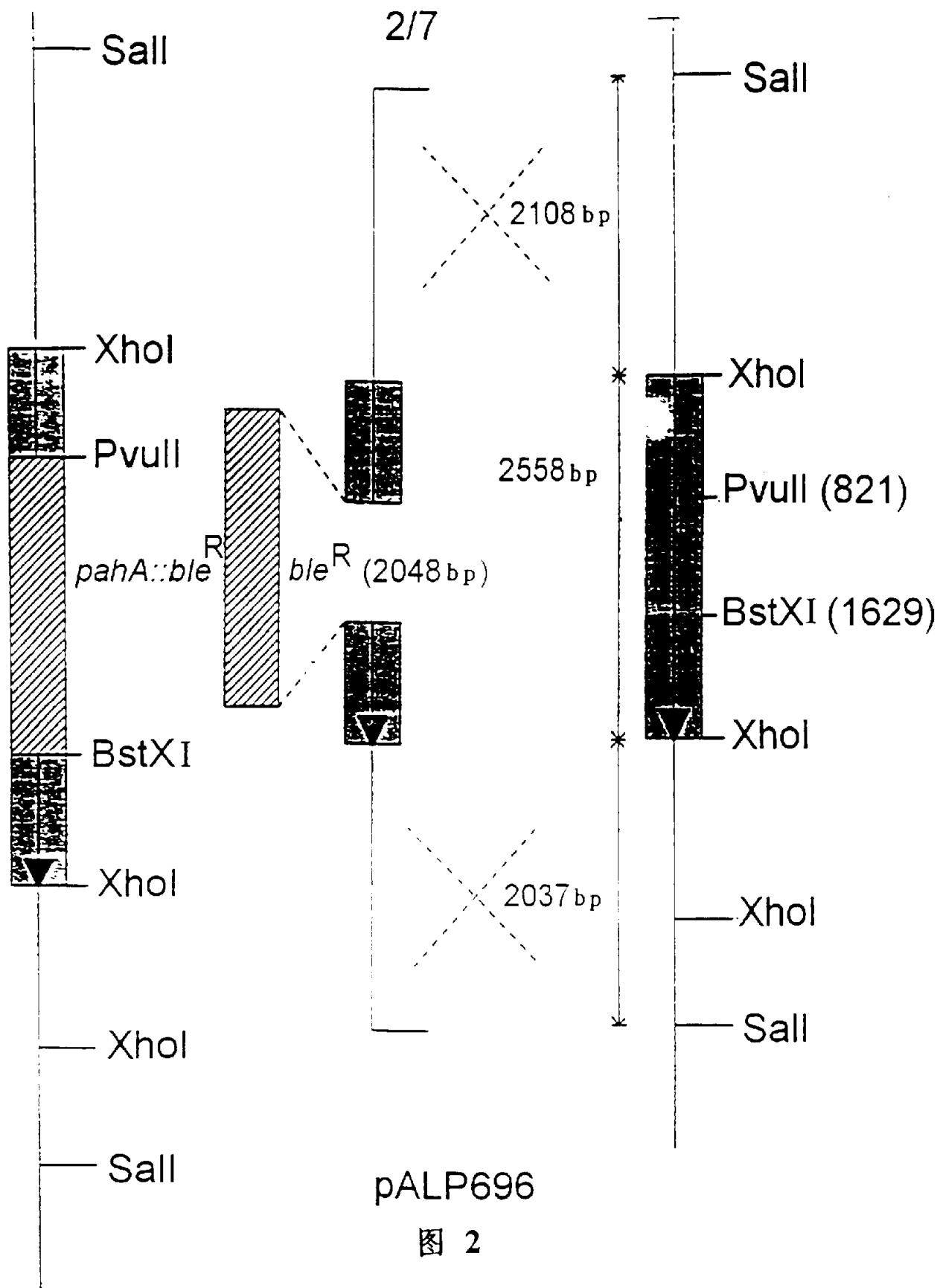


图 2

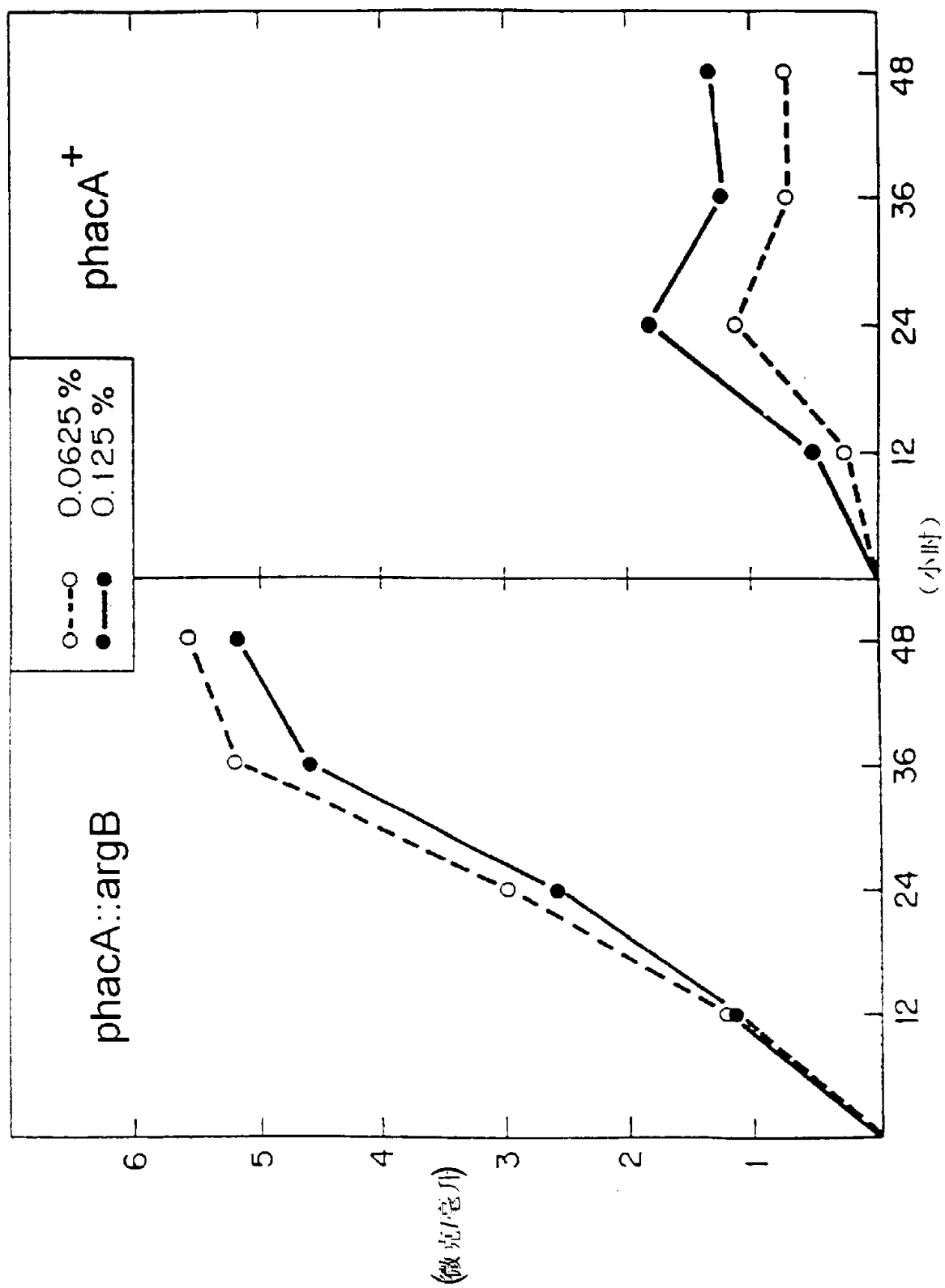


图 3

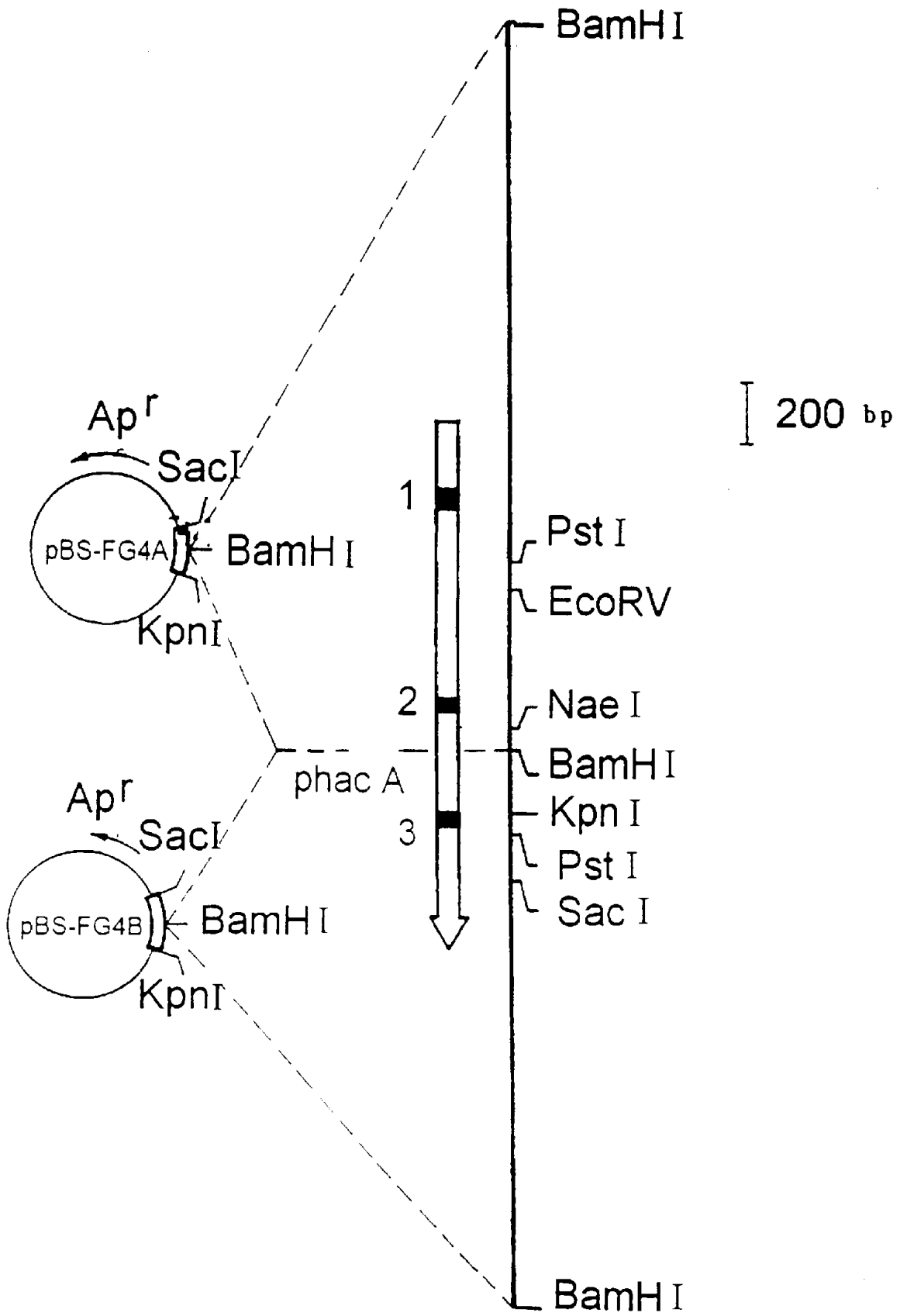


图 4

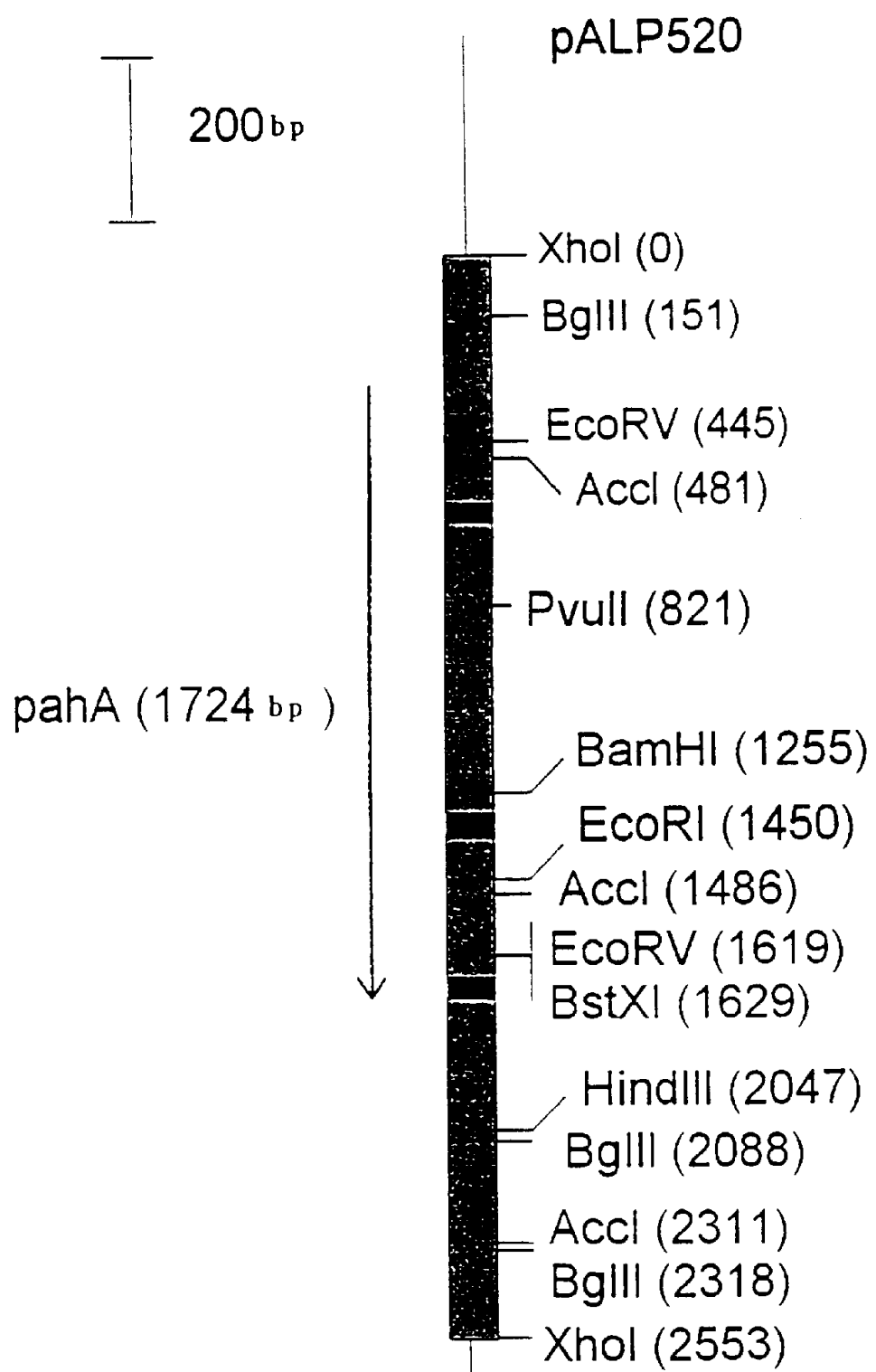


图 5

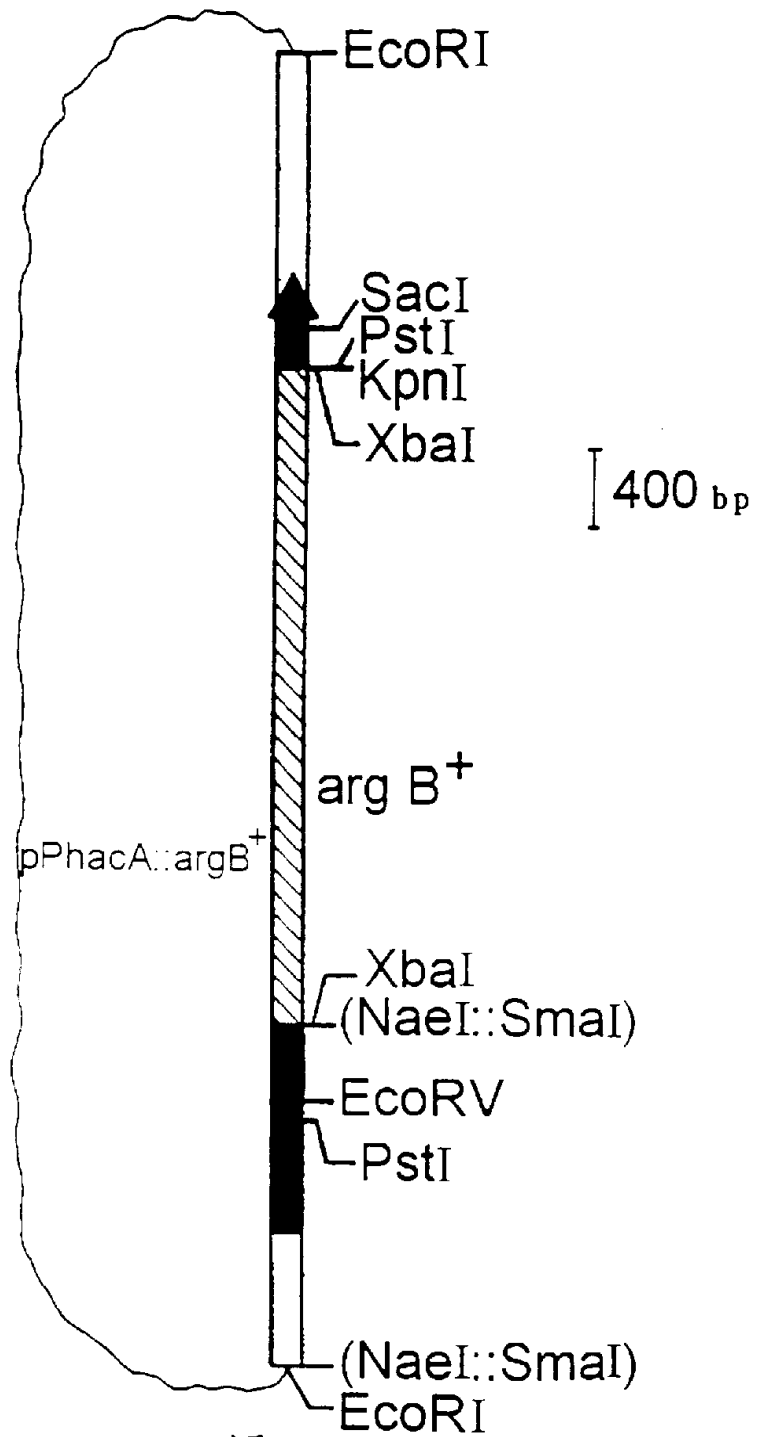


图 6

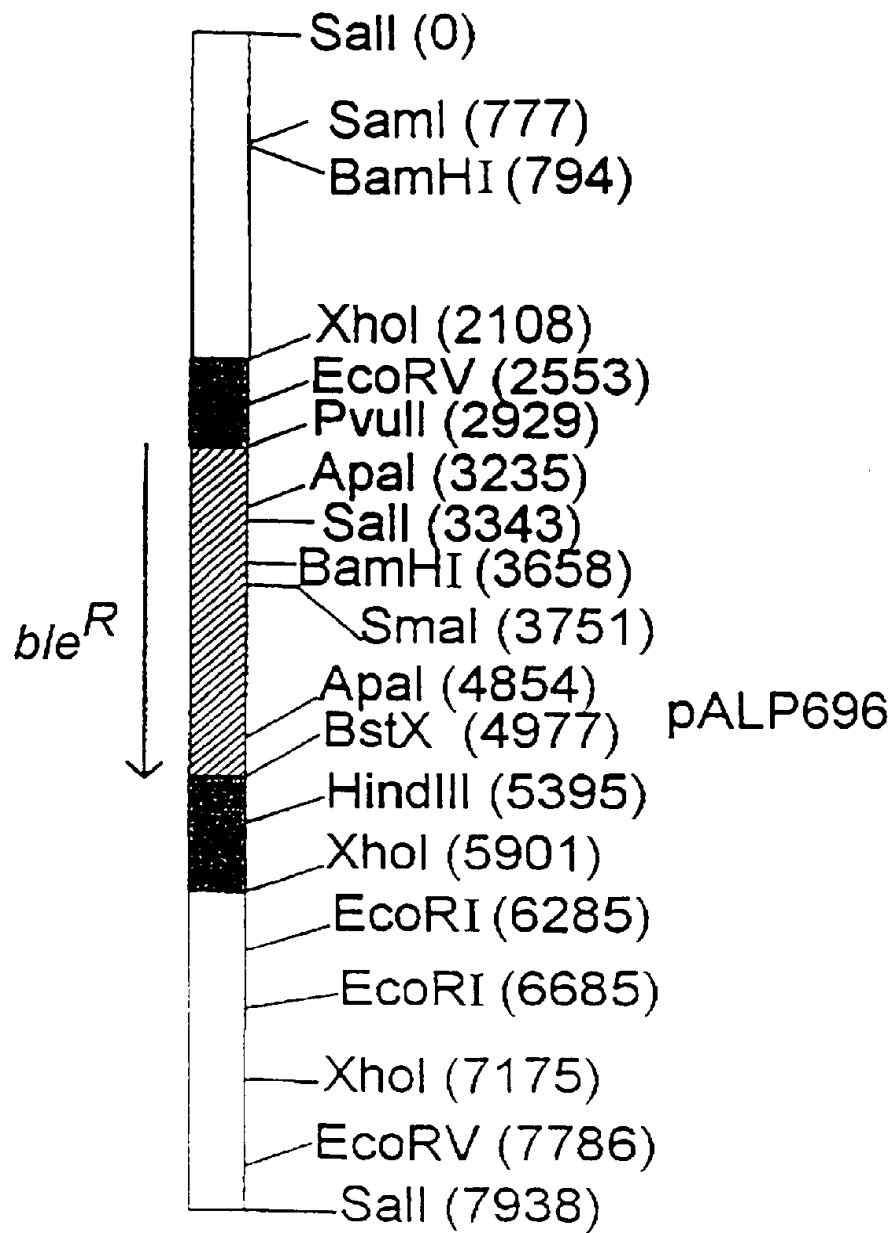


图 7