

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷

C12N 9/78

C12N 15/55 C12N 15/63

C12N 15/70 C12P 13/04



[12] 发明专利说明书

[21] ZL 专利号 01105347.X

[45] 授权公告日 2004 年 6 月 9 日

[11] 授权公告号 CN 1152956C

[22] 申请日 2001.2.15 [21] 申请号 01105347.X

[71] 专利权人 中国科学院上海植物生理研究所

地址 200032 上海市枫林路 300 号

[72] 发明人 杨蕴刘 许 祯 白 骅 姜卫红

陶正利 陈 军 焦瑞身

审查员 林峻凯

[74] 专利代理机构 上海专利商标事务所

代理人 徐 迅

权利要求书 1 页 说明书 21 页 附图 8 页

[54] 发明名称 新的乙内酰脲水解酶及其应用

[57] 摘要

本发明提供了一种新的海因酶、编码该海因酶的多核苷酸和经重组技术产生这种海因酶的方法。本发明还公开了含该海因酶的载体和宿主细胞，及其在生产药物中间体 D - 对羟基苯甘氨酸 (D - pHPG) 方面的用途。本发明还公开了可用于生产 D - pHPG 的氨甲酰水解酶及工程菌。

I S S N 1 0 0 8 - 4 2 7 4

- 1.一种分离的海因酶，其特征在于，该酶选自下组：
- (a)具有SEQ ID NO: 4氨基酸序列的多肽；
- 5 (b)将SEQ ID NO:4氨基酸序列经过一个或10个氨基酸残基的取代、缺失或添加而形成的，且具有催化D-对羟基苯乙内酰胺发生不对称水解，形成N-氨甲酰-D-对羟基苯甘氨酸功能的由(a)衍生的多肽。
- 2.如权利要求1所述的海因酶，其特征在于，该酶具有SEQ ID NO: 4氨基酸序列。
- 10 3.一种分离的多核苷酸，其特征在于，它含有一核苷酸序列，该核苷酸序列选自下组：
- (a)编码如权利要求1所述多肽的多核苷酸；
- (b)与多核苷酸(a)互补的多核苷酸。
- 4.如权利要求3所述的多核苷酸，其特征在于，该多核苷酸选自下组：
- 15 (a) 编码具有SEQ ID NO: 4所示氨基酸序列的多核苷酸；
- (b) 具有SEQ ID NO: 3中1758-3128位的核苷酸序列；
- (c) 具有SEQ ID NO: 3中1-3439位的核苷酸序列。
- 5.一种载体，其特征在于，它含有权利要求3所述的多核苷酸。
- 6.一种转化的宿主细胞，其特征在于，它含有权利要求6所述的载体。
- 20 7.如权利要求 6 所述的宿主细胞，其特征在于，它是大肠杆菌 BL21(DE3)-pXZpH2，保藏号为 CGMCC No. 0520.1。
8. 一种海因酶的制备方法，其特征在于，该方法包含：
- (a)在适合表达海因酶的条件下，培养权利要求6所述的宿主细胞；
- (b)从培养物中分离出海因酶。
- 25 9.如权利要求 1 所述的海因酶或权利要求 6 所述的宿主细胞的用途，其特征在于，用于生产 D-对羟基苯甘氨酸的工艺。
10. 如权利要求 9 所述的用途，其特征在于，该生产 D-对羟基苯甘氨酸的工艺还包括：使用具有 SEQ ID NO: 2 所示氨基酸序列的氨甲酰水解酶或表达该氨甲酰水解酶的重组工程菌，将 N-氨甲酰-D-对羟基苯甘氨酸转化为 D-对羟基苯甘氨酸。
- 30

新的乙内酰胺水解酶及其应用

技术领域

5 本发明涉及微生物工程、发酵和药物中间体合成领域。更具体地，本发明涉及一种新的海因酶(又称为乙内酰胺水解酶或乙内酰胺酶)、编码该海因酶的多核苷酸和经重组技术产生这种海因酶的方法。本发明还涉及含该海因酶的载体和宿主细胞及其在生产药物中间体 D-对羟基苯甘氨酸(D-pHPG)方面的用途。本发明还涉及了可用于生产 D-pHPG 的氨甲酰水解酶。

10 背景技术

由于羟氨苄青霉素(Amoxyceillin)、头孢羟氨苄(cefadroxil)、头孢哌酮(cefoperazone)、头孢罗奇(cefaprozil)等半合成 β -内酰胺抗生素，具有抗菌活力强，广谱低毒等特性，已成为临床控制细菌感染的首选药物。作为该类药物合成中间体之一的 D-对羟基苯甘氨酸，也因此受到制药和化工产业的重视。

15 D-对羟基苯甘氨酸(D-pHPG)的传统生产工艺是指通过化学合成和拆分的方法，但该工艺存在收率低，环境污染严重等弊端。

自八十年代以来，酶法制备 D-对羟基苯甘氨酸的研究和应用得到重视。一般而言，先由乙醛酸、苯酚、脲缩合得到 D, L-对羟基苯乙内酰胺(pHPH)，pHPH 可被 D-海因酶(D-hydantoinase, E. C. 3. 5. 2. 2, DHase)立体专一性水解开环生成 D-N-氨甲酰-对羟基苯甘氨酸(D-CpHPG)，在碱性或存在消旋酶条件下，未被开环的 L-型对映体消旋化为 D-型而被进一步转化。因此，通过这种酶促不对称水解混消旋底物，pHPH 可被完全转化，而不象化学拆分法，收率不到 50%。在工业上已有采用固定化的 D-海因酶或固定化细胞进行该步反应的报道。由 D-海因酶催化生成的 D-CpHPG，可通过以下二种方法而脱去氨甲酰基，得到终产物 D-pHPG：(1)HNO₂ 重氮化的化学法，和(2)氨甲酰水解酶(D-Carbamylase, DCase)的生物转化法(图 1)。

S. Takahashi 等曾采用将乙内酰胺化学合成、酶促不对称水解和亚硝酸重氮化脱氨甲酰基三者相结合的酶化学工艺，由 D, L-5-乙内酰胺生产 D-氨基酸，但存在亚硝酸的化学诱变作用而导致的三废污染和劳动防护等问题。

30 以氨甲酰基的酶促水解取代化学重氮化法进行第二步反应的双酶法，不论从可行性和环境保护考虑，都是极有意义的。Kanegafuchi 公司和我国山西太原生物研究所曾分别采用同时产生海因酶和氨甲酰水解酶的 *Agrobacterium radiobacter* 或 *Pseudomonas sp2262* 的休止细胞将 pHPH 转化为 D-pHPG。该

技术的缺点是转化效率低，需要诱导物，诱导物产生的异味损害人体健康并污染环境。

近年来，在分离产酶菌株的同时，已对(1)产 DHase 的 *Pseudomonas fluorescens* DSM84, *Bacillus stearothermophilis* NS1122A, *Pseudomonas putida* CCRC12857, (2)产 DCase 的 *Agrobacterium* sp KNK003, 以及(3)同时产 DHase 和 DCase 的 *Agrobacterium radiobacter* B11921 的相关酶基因进行克隆，得到了其中的相关酶基因，并对克隆后的基因进行分子生物学操作，以改进相关酶的酶学性质。然而，迄今为止还没有克隆出 *Burholderia pickettii*(皮氏伯克霍尔德氏菌)的海因酶和氨甲酰水解酶的基因。

因此，本领域迫切需要开发不同来源的海因酶和氨甲酰水解酶，以及新的生产 D-对羟基苯甘氨酸的工艺。

发明内容

本发明的目的就是提供一种新的海因酶及其片段、类似物和衍生物，及其编码序列。

本发明的另一目的是提供生产海因酶的方法以及其用途。

本发明的另一目的是提供新的高效、无害的生产 D-对羟基苯甘氨酸的工艺。

在本发明的第一方面，提供了新颖的分离出的海因酶，它包含：具有 SEQ ID NO: 4 氨基酸序列的多肽、或其保守性变异多肽、或其活性片段、或其活性衍生物。较佳地，该多肽是具有 SEQ ID NO: 4 氨基酸序列的多肽。

在本发明的第二方面，提供编码分离的海因酶的多核苷酸，该多核苷酸包含一核苷酸序列，该核苷酸序列与选自下组的一种核苷酸序列有至少 70% 相同性：(a)编码上述海因酶的多核苷酸；和(b)与多核苷酸(a)互补的多核苷酸。较佳地，该多核苷酸编码具有 SEQ ID NO: 4 所示氨基酸序列的多肽。更佳地，该多核苷酸的序列是选自下组的一种：具有 SEQ ID NO: 3 中 1758-3128 位的序列；具有 SEQ ID NO: 3 中 1-3439 位的序列。

在本发明的第三方面，提供了含有上述多核苷酸的载体，以及被该载体转化或转导的宿主细胞或者被上述多核苷酸直接转化或转导的宿主细胞。

在本发明的第四方面，提供了制备具有海因酶活性的多肽的方法，该方法包含：(a)在适合表达海因酶的条件下，培养上述被转化或转导的宿主细胞；(b)从培养物中分离出具有海因酶活性的多肽。

在本发明的第五方面，提供了本发明海因酶、其编码序列、载体或宿主细胞的用途。它们可用于生产 D-对羟基苯甘氨酸。

在本发明的第六方面，提供了一种新的生产 D-对羟基苯甘氨酸的工艺以及用于该工艺的氨甲酰水解酶和工程菌。

本发明的其它方面由于本文的技术的公开，对本领域的技术人员而言是显而易见的。

5 附图说明

图 1 是 D-对羟基苯甘氨酸的合成图。

图 2A 是 DCase 表达载体 pXZpC2 的构建示意图。

图 2B 是 DHase 表达载体 pXZpH2 的构建示意图。

图 3 是海因酶和氨甲酰水解酶表达产物的电泳图。其中各泳道是：1. 诱导前的 pXZpH2/BL21；2. 诱导后的 pXZpH2/BL21；3, 7. 蛋白 Maker；4. 空载体；5. 诱导前的 pXZpC2/BL21；6. 诱导后的 pXZpC2/BL21。

图 4 是 DCase 的 DNA 和氨基酸序列。

图 5 是 DHase 的 DNA 和氨基酸序列。

发明内容

15 在本发明中，术语“海因酶”、“乙内酰脲酶”、“乙内酰脲水解酶”或“DHase”可互换使用，都指具有海因酶氨基酸序列(SEQ ID NO:4)的蛋白或多肽。它们包括含有或不含起始甲硫氨酸的海因酶。

如本文所用，“分离的”是指物质从其原始环境中分离出来(如果是天然物质，原始环境即是天然环境)。如活体细胞内的天然状态下的多聚核苷酸和多肽是没有分离纯化的，但同样的多聚核苷酸或多肽如从天然状态中同存在的其他物质中分开，则为分离纯化的。

如本文所用，“分离的海因酶蛋白或多肽”是指海因酶基本上不含天然与其相关的其它蛋白、脂类、糖类或其它物质。本领域的技术人员能用标准的蛋白质纯化技术纯化海因酶。基本上纯的多肽在非还原聚丙烯酰胺凝胶上能产生
25 单一的主带

本发明的海因酶可以是重组的、天然的、合成的，优选是重组的。本发明的海因酶可以是天然纯化的产物，或是化学合成的产物，或使用重组技术从原核或真核宿主(例如，细菌、酵母、高等植物、昆虫和哺乳动物细胞)中产生。根据重组生产方案所用的宿主，本发明的多肽可以是糖基化的，或可以是非糖基化的。本发明的多肽还可包括或不包括起始的甲硫氨酸残基。
30

本发明还包括海因酶的片段、衍生物和类似物。如本文所用，术语“片段”、“衍生物”和“类似物”是指基本上保持本发明的天然海因酶相同的生物学功能或活性的多肽。本发明的多肽片段、衍生物或类似物可以是(i)有一个或多个保守或非保守性氨基酸残基(优选保守性氨基酸残基)被取代的多肽，而这样的

取代的氨基酸残基可以是也可以不是由遗传密码编码的，或(ii)在一个或多个氨基酸残基中具有取代基团的多肽，或(iii)成熟多肽与另一个化合物(比如延长多肽半衰期的化合物，例如聚乙二醇)融合所形成的多肽，或(iv)附加的氨基酸序列融合到此多肽序列而形成的多肽(如前导序列或分泌序列或用来纯化此多肽的序列或蛋白原序列)。根据本文的教导，这些片段、衍生物和类似物属于本领域熟练技术人员公知的范围。

在本发明中，术语“海因酶”指具有海因酶活性的 SEQ ID NO. 4 序列的多肽。该术语还包括具有与海因酶相同功能的、SEQ ID NO.4 序列的变异形式。这些变异形式包括(但并不限于)：若干个(通常为 1-50 个，较佳地 1-30 个，更佳地 1-20 个，最佳地 1-10 个)氨基酸的缺失、插入和/或取代，以及在 C 末端和/或 N 末端添加一个或数个(通常为 20 个以内，较佳地为 10 个以内，更佳地为 5 个以内)氨基酸。例如，在本领域中，用性能相近或相似的氨基酸进行取代时，通常不会改变蛋白质的功能。又比如，在 C 末端和/或 N 末端添加一个或数个氨基酸通常也不会改变蛋白质的功能。该术语还包括海因酶的活性片段和活性衍生物。

该多肽的变异形式包括：同源序列、保守性变异体、等位变异体、天然突变体、诱导突变体、在高或低的严紧度条件下能与海因酶 DNA 杂交的 DNA 所编码的蛋白、以及利用抗海因酶的抗血清获得的多肽或蛋白。本发明还提供了其他多肽，如包含海因酶或其片段的融合蛋白。除了几乎全长的多肽外，本发明还包括了海因酶的可溶性片段。通常，该片段具有海因酶序列的至少约 30 个连续氨基酸，较佳地至少约 50 个连续氨基酸，更佳地至少约 80 个连续氨基酸，最佳地至少约 100 个连续氨基酸。

发明还提供海因酶或多肽的类似物。这些类似物与天然海因酶的差别可以是氨基酸序列上的差异，也可以是不影响序列的修饰形式上的差异，或者兼而有之。这些多肽包括天然或诱导的遗传变异体。诱导变异体可以通过各种技术得到，如通过辐射或暴露于诱变剂而产生随机诱变，还可通过定点诱变法或其他已知分子生物学的技术。类似物还包括具有不同于天然 L-氨基酸的残基(如 D-氨基酸)的类似物，以及具有非天然存在的或合成的氨基酸(如 β 、 γ -氨基酸)的类似物。应理解，本发明的多肽并不限于上述列举的代表性的多肽。

修饰(通常不改变一级结构)形式包括：体内或体外的多肽的化学衍生形式如乙酰化或羧基化。修饰还包括糖基化，如那些在多肽的合成和加工中或进一步加工步骤中进行糖基化修饰而产生的多肽。这种修饰可以通过将多肽暴露于进行糖基化的酶(如哺乳动物的糖基化酶或去糖基化酶)而完成。修饰形式还包括具有磷酸化氨基酸残基(如磷酸酪氨酸，磷酸丝氨酸，磷酸苏氨酸)的序列。

还包括被修饰从而提高了其抗蛋白水解性能或优化了溶解性能的多肽。

在本发明中，“海因酶保守性变异多肽”指与SEQ ID NO: 4的氨基酸序列相比，有至多10个，较佳地至多8个，更佳地至多5个，最佳地至多3个氨基酸被性质相似或相近的氨基酸所替换而形成多肽。这些保守性变异多肽最好根据表A进行氨基酸替换而产生。

表 A

最初的残基	代表性的取代	优选的取代
Ala (A)	Val; Leu; Ile	Val
Arg (R)	Lys; Gln; Asn	Lys
Asn (N)	Gln; His; Lys; Arg	Gln
Asp (D)	Glu	Glu
Cys (C)	Ser	Ser
Gln (Q)	Asn	Asn
Glu (E)	Asp	Asp
Gly (G)	Pro; Ala	Ala
His (H)	Asn; Gln; Lys; Arg	Arg
Ile (I)	Leu; Val; Met; Ala; Phe	Leu
Leu (L)	Ile; Val; Met; Ala; Phe	Ile
Lys (K)	Arg; Gln; Asn	Arg
Met (M)	Leu; Phe; Ile	Leu
Phe (F)	Leu; Val; Ile; Ala; Tyr	Leu
Pro (P)	Ala	Ala
Ser (S)	Thr	Thr
Thr (T)	Ser	Ser
Trp (W)	Tyr; Phe	Tyr
Tyr (Y)	Trp; Phe; Thr; Ser	Phe
Val (V)	Ile; Leu; Met; Phe; Ala	Leu

本发明的多核苷酸可以是 DNA 形式或 RNA 形式。DNA 形式包括 cDNA、基因组 DNA 或人工合成的 DNA。DNA 可以是单链的或是双链的。DNA 可以是编码链或非编码链。编码成熟多肽的编码区序列可以与 SEQ ID NO:3 所示的编码区序列相同或者是简并的变异体。如本文所用，“简并的变异体”在本发明中是指编码具有 SEQ ID NO:4 的蛋白质，但与 SEQ ID NO:3 所示的编码区序列有差别的核酸序列。

编码海因酶成熟多肽的多核苷酸包括：只编码成熟多肽的编码序列；成熟

多肽的编码序列+各种附加编码序列；成熟多肽的编码序列(和任选的附加编码序列)+非编码序列。

术语“编码海因酶的多核苷酸”可以是包括仅编码海因酶的多核苷酸，也可以是还包括附加编码和/或非编码序列的多核苷酸。

5 本发明还涉及上述多核苷酸的变异体，其编码与本发明有相同的氨基酸序列的多肽或多肽的片段、类似物和衍生物。此多核苷酸的变异体可以是天然发生的等位变异体或非天然发生的变异体。这些核苷酸变异体包括取代变异体、缺失变异体和插入变异体。如本领域所知的，等位变异体是一个多核苷酸的替换形式，它可能是一个或多个核苷酸的取代、缺失或插入，但不会从实质上改变其编码的多肽的功能。

10 本发明还涉及与上述的序列杂交且两个序列之间具有至少 50%，较佳地至少 70%，更佳地至少 80%相同性的多核苷酸。本发明特别涉及在严格条件下与本发明所述多核苷酸可杂交的多核苷酸。在本发明中，“严格条件”是指：(1)在较低离子强度和较高温度下的杂交和洗脱，如 $0.2 \times \text{SSC}$, 0.1%SDS, 60°C；或
15 (2)杂交时加有变性剂，如 50%(v/v)甲酰胺, 0.1%小牛血清/0.1% Ficoll, 42°C等；或(3)仅在两条序列之间的相同性至少在 90%以上,更好是 95%以上时才发生杂交。并且，可杂交的多核苷酸编码的多肽与 SEQ ID NO:4 所示的成熟多肽有相同的生物学功能和活性。

20 本发明还涉及与上述的序列杂交的核酸片段。如本文所用，“核酸片段”的长度至少含 15 个核苷酸，较好是至少 30 个核苷酸，更好是至少 50 个核苷酸，最好是至少 100 个核苷酸以上。核酸片段可用于核酸的扩增技术(如 PCR)以确定和/或分离编码海因酶的多聚核苷酸。

25 本发明的海因酶核苷酸全长序列或其片段通常可以用PCR扩增法、重组法或人工合成的方法获得。对于PCR扩增法，可根据本发明所公开的有关核苷酸序列，尤其是开放阅读框序列来设计引物，并用市售的cDNA库或按本领域技术人员已知的常规方法所制备的cDNA库作为模板，扩增而得有关序列。当序列较长时，常常需要进行两次或多次PCR扩增，然后再将各次扩增出的片段按正确次序拼接在一起。

30 一旦获得了有关的序列，就可以用重组法来大批量地获得有关序列。这通常是将其克隆入载体，再转入细胞，然后通过常规方法从增殖后的宿主细胞中分离得到有关序列。

此外，还可用人工合成的方法来合成有关序列，尤其是片段长度较短时。通常，通过先合成多个小片段，然后再进行连接可获得序列很长的片段。

目前，已经可以完全通过化学合成来得到编码本发明蛋白(或其片段，或

其衍生物)的 DNA 序列。然后可将该 DNA 序列引入本领域中已知的各种现有的 DNA 分子(或如载体)和细胞中。此外, 还可通过化学合成将突变引入本发明蛋白序列中。

5 本发明也涉及包含本发明的多核苷酸的载体, 以及用本发明的载体或海因酶编码序列经基因工程产生的宿主细胞, 以及经重组技术产生本发明所述多肽的方法。

通过常规的重组 DNA 技术(Science, 1984; 224: 1431), 可利用本发明的多聚核苷酸序列可用来表达或生产重组的海因酶。一般来说有以下步骤:

- 10 (1).用本发明的编码海因酶的多核苷酸(或变异体), 或用含有该多核苷酸的重组表达载体转化或转导合适的宿主细胞;
- (2).在合适的培养基中培养的宿主细胞;
- (3).从培养基或细胞中分离、纯化蛋白质。

15 本发明中, 海因酶多核苷酸序列可插入到重组表达载体中。术语“重组表达载体”指本领域熟知的细菌质粒、噬菌体、酵母质粒、或其他载体。在本发明中适用的载体包括但不限于: 在细菌中表达的基于 T7 的表达载体(Rosenberg, et al. Gene, 1987, 56:125)。总之, 只要能在宿主体内复制和稳定, 任何质粒和载体都可以用。表达载体的一个重要特征是通常含有复制起点、启动子、标记基因和翻译控制元件。

20 本领域的技术人员熟知的方法能用于构建含海因酶编码 DNA 序列和合适的转录/翻译控制信号的表达载体。这些方法包括体外重组 DNA 技术、DNA 合成技术、体内重组技术等(Sambrook, et al. Molecular Cloning, a Laboratory Manual, cold Spring Harbor Laboratory. New York, 1989)。所述的 DNA 序列可有效连接到表达载体中的适当启动子上, 以指导 mRNA 合成。这些启动子的代表性例子有: 大肠杆菌的 lac 或 trp 启动子; λ 噬菌体 P_L 启动子; 真核启动子包括 CMV 立即早期启动子、HSV 胸苷激酶启动子、早期和晚期 SV40 启动子、反转录病毒的 LTRs 和其他一些已知的可控制基因在原核或真核细胞或其病毒中表达的启动子。表达载体还包括翻译起始用的核糖体结合位点和转录终止子。

30 此外, 表达载体优选地包含一个或多个选择性标记基因, 以提供用于选择转化的宿主细胞的表型性状, 如真核细胞培养用的二氢叶酸还原酶、以及新霉素抗性, 或用于大肠杆菌的四环素或氨苄青霉素抗性。

包含上述的适当 DNA 序列以及适当启动子或者控制序列的载体, 可以用于转化适当的宿主细胞, 以使其能够表达蛋白质。

宿主细胞可以是原核细胞, 如细菌细胞; 或是低等真核细胞, 如酵母细胞;

或是高等真核细胞。较佳地是大肠杆菌。代表性例子有：大肠杆菌，链霉菌属；鼠伤寒沙门氏菌的细菌细胞；真菌细胞如酵母等。

5 用重组 DNA 转化宿主细胞可用本领域技术人员熟知的常规技术进行。当宿主为原核生物如大肠杆菌时，能吸收 DNA 的感受态细胞可在指数生长期后收获，用 CaCl_2 法处理，所用的步骤在本领域众所周知。另一种方法是使用 MgCl_2 。如果需要，转化也可用电穿孔的方法进行。当宿主是真核生物，可选用如下的 DNA 转染方法：磷酸钙共沉淀法，常规机械方法如显微注射、电穿孔、脂质体包装等。

10 获得的转化子可以用常规方法培养，表达本发明的基因所编码的多肽。根据所用的宿主细胞，培养中所用的培养基可选自各种常规培养基。在适于宿主细胞生长的条件下进行培养。当宿主细胞生长到适当的细胞密度后，用合适的方法(如温度转换或化学诱导)诱导选择的启动子，将细胞再培养一段时间。

15 在上面的方法中的重组多肽可在细胞内、或在细胞膜上表达、或分泌到细胞外。如果需要，可利用其物理的、化学的和其它特性通过各种分离方法分离和纯化重组的蛋白。这些方法是本领域技术人员所熟知的。这些方法的例子包括但并不限于：常规的复性处理、用蛋白沉淀剂处理(盐析方法)、离心、渗透破菌、超处理、超离心、分子筛层析(凝胶过滤)、吸附层析、离子交换层析、高效液相层析(HPLC)和其它各种液相层析技术及这些方法的结合。

20 本发明的酶还包括固定在载体上的固定化酶。本领域熟知的各种固定化酶技术都可用于制备本发明的固定化的海因酶和或氨甲酰水解酶。

此外，本发明提供了海因酶、其编码序列、载体或宿主细胞的用途，它们被用于生产 D-对羟基苯甘氨酸的双酶法工艺。该工艺包括步骤：

(a) 在 pH8-9 条件下，通过海因酶的催化作用使 D-对羟基苯乙内酰脲发生不对称水解，形成 N-氨甲酰-D-对羟基苯甘氨酸；

25 (b) N-氨甲酰-D-对羟基苯甘氨酸通过氨甲酰水解酶或表达该氨甲酰水解酶的重组工程菌，被转化为 D-对羟基苯甘氨酸。

此外，L-对羟基苯乙内酰脲被消旋化再形成 D-对羟基苯乙内酰脲，作为步骤(a)的原料(图 1)。

30 因此，本发明还提供了一种新的生产 D-对羟基苯甘氨酸的工艺以及用于该工艺的氨甲酰水解酶和工程菌。一种从皮氏伯克霍尔德氏菌中新分离出的氨甲酰水解酶具有 SEQ ID NO: 2 所示的氨基酸序列，其核酸序列示于 SEQ ID NO: 1，其中阅读框位于第 2065-2976 位。

与上面关于海因酶的描述相同，本发明提供了编码氨甲酰水解酶的多核苷酸，含有该多核苷酸的载体，以及被该载体转化或转导的宿主细胞或者被

上述多核苷酸直接转化或转导的宿主细胞。

鉴于海因酶和氨甲酰水解酶都用于D-对羟基苯甘氨酸的工艺，因此一种优选的工程菌是同时携带海因酶和氨甲酰水解酶基因，并表达海因酶和氨甲酰水解酶的工程菌，例如大肠杆菌。

- 5 本发明的工艺优点在于：克服了原生产工艺中的转化效率低，生产周期长，特别是严重的环境污染，以及工人的劳动保护问题。

具体实施方式

- 下面结合具体实施例，进一步阐述本发明。应理解，这些实施例仅用于说明本发明而不用于限制本发明的范围。下列实施例中未注明具体条件的实验方法，通常按照常规条件如Sambrook等人，分子克隆：实验室手册(New York: Cold Spring Harbor Laboratory Press, 1989)中所述的条件，或按照制造厂商所建议的条件。

实施例 1:

- 15 皮氏伯克霍尔德氏菌 (*Burkholderia pickettii*) 的培养和总 DNA 的分离

- 50ml 的 LB 液体培养基(每升培养基含有 10g 酵母粉, 5g 蛋白胨, 10g 氯化钠, pH 值 7.0)中接种一个单菌落, 28°C 培养 36 小时。将收获的培养液 4000 转离心 10 分钟收集菌体, 悬浮细菌沉淀于 15ml TE 缓冲液中, 加入蛋白酶 K 和 SDS 至终浓度分别为 100ug/ml 和 0.5%。混匀, 并于 37°C 温育过夜至溶液清亮。加入等体积的酚/氯仿/异戊醇(25:24:1), 缓慢混匀后于 8000 转离心 10 分钟, 用大口移液管移出上层水相。反复抽提, 直至没有白色中间层。最后一次将上层转入新管(注意不要带入有机相), 加入 1/10 体积的 3M 的醋酸钠, 再加入 2 倍体积的无水乙醇, 混匀。用封口的玻璃滴管绕出 DNA, 将其转入新瓶中, 用 70%乙醇洗去盐份, 干燥 DNA, 溶解于 TE 缓冲液。在其中加入 RNAase 至终浓度 100ug/ml, 于 37°C 温育 1.5 小时。然后在其中加入 0.2 体积的 10M 的醋酸铵, 再加入 2 倍体积的无水乙醇, 用封口的玻璃滴管绕出 DNA, 将其转入新瓶中, 用 70%乙醇洗去盐份, 干燥 DNA, 溶解于 TE 缓冲液至终浓度 250ng/ul。于 4°C 保存。

- 30 实施例 2:

目的片段的检测

分别取 40ul 总 DNA, 在 200ul 的反应体系中用 8ul ClaI, EcoRI, EcoRV, XhoI 和 20ul 相应的反应缓冲液进行限制性酶切反应。反应两小时后, 从反应体系中取出 5ul, 加入 0.5ul 加样缓冲液, 用 0.8%的琼脂糖进行凝胶电泳以观

察是否酶切完全。待总 DNA 酶切完全后，加入等体积的酚/氯仿/异戊醇 (25:24:1)，混匀后与 12000 转离心 5 分钟，将上层水相转入新管，加入 1/10 体积的 3M 的醋酸钠，再加入 2 倍体积的无水乙醇，混匀后于 -70°C 放置 5 分钟，然后于 12000 转离心 5 分钟。倒去液体，用 70%乙醇洗去盐份，室温晾干，溶于 40 ul 双蒸水。用 0.8%的琼脂糖进行凝胶电泳，1V/cm 电泳过夜。

5 将电泳后的琼脂糖凝胶室温下浸泡在 20 倍体积的 0.25M 盐酸溶液中 5-15 分钟，直至凝胶中的溴酚蓝恰好变为黄色为止。将胶转移至盛有 20 倍体积的 1.5M NaCl, 0.5M NaOH 溶液中，室温下浸泡 20 分钟，更换新液，再浸泡 20 分钟。将胶转移至盛有 20 倍体积的 1.0M Tris-Cl, 1.5M NaCl 溶液 (pH7.5) 中，室温下浸泡 20 分钟，更换新液，再浸泡 20 分钟。将胶底朝上放在水平玻璃板的板面上，赶走两层之间的气泡。裁一张与凝胶一样大小的尼龙膜做好标记，使用前用蒸馏水浸湿，再用 10×SSC 浸没至少 5 分钟，放在凝胶上，赶走两层之间的气泡。裁两张与凝胶一样大小的新华二号滤纸，和尼龙膜进行相同的处理后，放在尼龙膜上，赶走气泡，再在滤纸上放一叠纸巾，在纸巾上方放一块玻璃板和 500 克的重物，转移约两小时或更长。取下转移好的尼龙膜，在 4×SSC 中浸泡 5 分钟，室温干燥 30 分钟，80°C 烘干 30 分钟，供杂交用。

10 15

按照 Roche 公司提供的 DIG High Primer DNA Labeling and Detection Starter Kit II 使用手册进行杂交。加 1ug 模板 DNA 用双蒸水稀释到 16ul，煮沸 10 分钟后迅速冰浴，加入 4ul DIG-High Primer 37°C 放置过夜，再用 65°C 水浴 10 分钟灭活酶。检测标记效率为 100ng/ul。用 DIG Easy Hyb 10ml 42°C 预杂交 30 分钟，换用新的 DIG Easy Hyb 在其中加入 250 ng 已变性的 DIG-标记的 DNA 探针，42°C 放置过夜。68°C 用 0.5×SSC, 0.1% SDS 洗膜两次，每次 15 分钟。用洗涤缓冲液洗 1-5 分钟，用封闭缓冲液室温处理 30 分钟，再用抗体溶液处理 30 分钟，洗涤缓冲液作用 2×15 分钟，检测缓冲液处理 2-5 分钟后，将适量 CSPD 加到膜上，在膜上封口，室温放 15-25 分钟，37°C 放 10 分钟，用 X-光片曝光检测有杂交的条带。其中片段大小合适的是 EcoR V 的杂交条带。

20 25

实施例 3

30 目的片段的分离

取 40ul 总 DNA，在 200ul 的反应体系中用 8ul EcoRV 和 20ul 相应的反应缓冲液进行限制性酶切反应。反应两小时后，从反应体系中取出 5ul，加入 0.5ul 加样缓冲液，用 0.8%的琼脂糖进行凝胶电泳以观察是否酶切完全。待总 DNA 酶切完全后，加入等体积的酚/氯仿/异戊醇 (25:24:1)，混匀后与 12000 转离心

5 分钟，将上层水相转入新管，加入 1/10 体积的 3M 的醋酸钠，再加入 2 倍体积的无水乙醇，混匀后于 -70°C 放置 5 分钟，然后于 12000 转离心 5 分钟。倒去液体，用 70%乙醇洗去盐份，室温晾干，溶于 40 ul 双蒸水。用 0.8%的琼脂糖进行凝胶电泳，1V/cm 电泳过夜。割下电泳后与检测到有杂交的位置相同的琼脂糖条带，利用华舜公司生产的胶回收柱回收目的片段(50ng/ul)。

10 按照《分子克隆》中小量抽提质粒的方法，抽提 pBluKS，将 17ul 抽提好的质粒(100ng/ul)在 20ul 的反应体系中用 1ul EcoRI 和 2ul 相应的反应缓冲液进行限制性酶切反应，反应两小时后用 0.8%的琼脂糖进行凝胶电泳，5V/cm 电泳 30 分钟。割下 3kb 附近的琼脂糖条带，利用华舜公司生产的胶回收柱回收 40ul 的体系中。

取 17ul 回收物加入 2ul 去磷酸化酶缓冲液和 0.1 单位的去磷酸化酶 (Promega 公司)于 37°C 反应 30 分钟，用 0.8%的琼脂糖进行凝胶电泳，5V/cm 电泳 30 分钟。割下 3kb 附近的琼脂糖条带，利用华舜公司生产的胶回收柱回收(60ng/ul)。

15 按照 MBI 公司提供的连接酶最适反应体系 20ul 中加入外源片段 2ul 和载体 1ul 以及缓冲液 2ul 和酶 0.5ul 在 22°C 连接过夜，65°C 灭活 10 分钟。取 2ul 加入 100ul 氯化钙法制备的转化效率达 10^7 的大肠杆菌 DH5 α ，冰上放置 30 分钟，42°C 热激 70 秒，加 700ul LB 液体培养基 37°C 培养 45 分钟复壮，取 200ul 涂在含有规定浓度的 X-gal 和 IPTG 的氨苄青霉素平板。37°C 培养过夜。

20 挑取白斑先接种到氨苄青霉素平板上，再接种到铺有尼龙膜的氨苄青霉素平板。37°C 培养过夜，相对应于每一张滤膜，在一片 Saran 包装膜上制作一个装有 0.5mol/L NaOH 的小洼(0.75ml)，使菌落面朝上，将滤膜放到小洼上，展平 Saran 包装膜，使滤膜均匀湿润，让滤膜留于原处 2-3 分钟，用干的纸巾从滤膜的下方吸干滤膜，重复以上过程。再次吸干滤膜，将滤膜转移到新的带有 1mol/L Tris-Cl (pH7.4) 的 Saran 包装膜小洼上，5 分钟后，吸干滤膜，重复这一步。吸干滤膜，将滤膜转移到新的带有 1.5mol/L NaCl、0.5mol/L Tris-Cl (pH7.4) 的 Saran 包装膜小洼上，5 分钟后，吸干滤膜，把它转移到一张干的 3MM 滤纸上，于室温晾至少 30 分钟，使滤膜干燥。将滤膜夹在两张干的 3MM 滤纸之间，在真空炉中于 80°C 干烤 2 小时，以固定 DNA。

30 按实施例 2 中所述的方法，将膜进行杂交。杂交斑点所对应的克隆即为带有目的片段的克隆。结果得到一个阳性克隆—质粒 pXZ-total。

实施例 4

目的片段的分析

将带有目的片段的克隆用 ClaI, EcoRV, HincII, NotI, SacI, SmaI, SpeI 在 20ul 的反应体系中用 1ul 酶和 2ul 相应的反应缓冲液进行限制性酶切反应，反应两小时后用 0.8% 的琼脂糖进行凝胶电泳，5V/cm 电泳 30 分钟。然后，与实施例 2 进行相同的转膜操作。同样按照 Roche 公司提供的 DIG High Primer DNA Labeling and Detection Starter Kit II 使用手册进行杂交。

结果表明 NotI 所切出的片段适合测序。割下 NotI 切出的三条琼脂糖条带，利用华舜公司生产的胶回收柱回收。按照《分子克隆》中小量抽提质粒的方法抽提 pBluKS。将 17ul 抽提好的质粒(100ng/ul)在 20ul 的反应体系中用 1ul NotI 和 2ul 相应的反应缓冲液进行限制性酶切反应，反应两小时后用 0.8% 的琼脂糖进行凝胶电泳，5V/cm 电泳 30 分钟。割下 3kb 附近的琼脂糖条带，利用华舜公司生产的胶回收柱回收得到 40ul 的体系中。

取 17ul 回收物加入 2ul 去磷酸化酶缓冲液和 0.1 单位的去磷酸化酶 (Promega 公司)于 37°C 反应 30 分钟，用 0.8% 的琼脂糖进行凝胶电泳，5V/cm 电泳 30 分钟。割下 3kb 附近的琼脂糖条带，利用华舜公司生产的胶回收柱回收(60ng/ul)。

按照 MBI 公司提供的连接酶最适反应体系 20ul 中加入外源片段 2ul 和载体 1ul 以及缓冲液 2ul 和酶 0.5ul 在 22°C 连接过夜，65°C 灭活 10 分钟。取 2ul 加入 100ul 转化效率达 10^7 的大肠杆菌 DH5 α ，冰上放置 30 分钟，42°C 热激 70 秒，加 700ul LB 液体培养基 37°C 培养 45 分钟复壮，取 200ul 涂在含有规定浓度的 X-gal 和 IPTG 的氨苄青霉素平板。37°C 培养过夜。挑取白斑用于测序。测序由上海基康公司的测序仪完成。

将分开测序的两目的基因用 BioEdit 分析，拼接为两个完整的读框(见 SEQ ID NO: 1-4)，SEQ ID NO: 1 和 2 为氨甲酰水解酶的 DNA 序列和氨基酸序列，阅读框位于 SEQ ID NO: 1 中第 2065-2976 位，编码一个 347 个氨基酸的氨甲酰水解酶。SEQ ID NO: 3 和 4 为海因酶的 DNA 序列和氨基酸序列，阅读框位于 SEQ ID NO: 3 中第 1758-3128 位，编码一个 457 个氨基酸的氨甲酰水解酶。有趣的是，这两个酶分别位于同一 3439bp 双链 DNA 的两条链上。

用 GCG 进行同源分析，结果表明：皮氏伯克霍尔德氏菌海因酶与 *Agrobacterim. radiobacter* 海因酶在氨基酸水平有 85% 的同源性，与 *Bacillus. sterothermophilus* 海因酶有 47% 的同源性，与 *Pseudomonas. putida* 的海因酶有 35% 的同源性。氨甲酰水解酶的基因与其他物种的氨甲酰水解酶基因高度同源：与 *Agrobacterim. sp* KNK712 有 98% 的同源性 *Agrobacterim. radiobacter* 有 92% 的同源性。

实施例 5

含氨甲酰氨基酸水解酶目的基因表达载体和工程菌的构建

参见图 2A。用 Primer 程序设计引物，氨甲酰氨基酸水解酶用
5'cgagctagcatgacacgtcagatgatac 和 5'aagctttcagagttccgcga 两条引物，以带有目的
5 片段的克隆为模板，94℃变性 5 分钟后，94℃变性 1 分钟，52℃退火 1 分钟，
72℃延伸 1 分钟，共进行 30 个循环，最后 72℃放 10 分钟用于加 A，得到大量目的基因片段。电泳后利用华舜公司生产的胶回收柱回收(50ng/ul) 目的基因片段(外源片段)。

按照 Promega 公司 pGEMT-EASY vector system 提供的最适反应体系 10ul
10 中加入外源片段 1ul 和载体 pGEMT-EASY 1ul 以及缓冲液 5ul 和酶 0.5ul 在 22
℃连接 1 小时，65℃灭活 10 分钟。取 5ul 加入 100ul 转化效率达 10^7 的大肠杆菌
DH5 α ，冰上放置 30 分钟，42℃热激 70 秒，加 500ul LB 液体培养基 37℃
培养 45 分钟复壮，取 200ul 涂在含有规定浓度的 X-gal 和 IPTG 的氨苄青霉
素平板。37℃培养过夜，挑取白斑接入试管。按照《分子克隆》中所述的方法
15 小量抽提质粒 pXZpC1，用 0.5ul NheI 和 0.5ul HindIII 在 20ul 含有相应的反
应缓冲液的反应体系中进行限制性酶切反应，反应两小时后用 0.8%的琼脂糖
进行凝胶电泳，5V/cm 电泳 30 分钟。割下 0.9kb 附近的琼脂糖条带，利用华
舜公司生产的胶回收柱回收得到 20ul 的体系中。

在连接反应体系 10ul 中加入外源片段 4ul 和经过 NheI 和 HindIII 酶切的
20 载体 pET28b 4ul 以及缓冲液 1ul 和酶 0.5ul 在 16℃连接过夜，65℃灭活 10 分
钟。取 5ul 加入 100ul 转化效率达 10^7 的大肠杆菌 DH5 α ，冰上放置 30 分钟，
42℃热激 70 秒，加 500ul LB 液体培养基 37℃培养 45 分钟复壮，取 200ul 涂
在含有规定浓度的卡那霉素平板。37℃培养过夜，挑取单菌落接入试管。按
照《分子克隆》中所述方法小量抽提质粒 pXZpC2，用 0.5ul NheI 和 0.5ul HindIII
25 在 20ul 含有相应的反应缓冲液的反应体系中进行限制性酶切反应，反应两小
时后用 0.8%的琼脂糖进行凝胶电泳，5V/cm 电泳 30 分钟，以验证所挑取的克
隆是否正确。

将构建好的表达载体 pXZpC2 用氯化钙法转化 E.coli BL21(DE3)，得到大
肠杆菌 BL21(DE3)-pXZpC2 工程菌。将获得的携带质粒 pXZpC2 的大肠杆菌于
30 2000 年 12 月 6 日并于 2000 年 11 月 6 日保藏在中国微生物菌种保藏管理委员
会普通微生物中心(CGMCC，中国，北京)，保藏号为 CGMCC No. 0520.2。

实施例 6

含海因酶目的基因表达载体和工程菌的构建

参见图 2B。用 5'gctagcatggacatcattatcaaa 和 5'aagcttaatgccggtttactg 两条引物，以带有目的片段的克隆为模板，94℃变性 5 分钟后，94℃变性 1 分钟，51℃退火 1 分钟，72℃延伸 1.5 分钟，共进行 30 个循环，最后 72℃放 10 分钟用于加 A，得到大量含乙内酰胺水解酶目的基因片段。电泳后利用华舜公司生产的胶回收柱回收(100ng/ul)。

按照 Promega 公司 pGEMT-EASY vector system 提供的最适反应体系 10ul 中加入外源片段 1ul 和载体 1ul 以及缓冲液 5ul 和酶 0.5ul 在 22℃连接 1 小时，65℃灭活 10 分钟。取 5ul 加入 100ul 转化效率达 10^7 的大肠杆菌 DH5 α ，冰上放置 30 分钟，42℃热激 70 秒，加 500ul LB 液体培养基 37℃培养 45 分钟复壮，取 200ul 涂在含有规定浓度的 X-gal 和 IPTG 的氨苄青霉素平板。37℃培养过夜，挑取白斑接入试管。按照《分子克隆》所述的方法小量抽提质粒 pXZpH1，用 0.5ul NheI 和 0.5ul HindIII 在 20ul 含有相应的反应缓冲液的反应体系中进行限制性酶切反应，反应两小时后用 0.8%的琼脂糖进行凝胶电泳，5V/cm 电泳 30 分钟。割下 1.4kb 附近的琼脂糖条带，利用华舜公司生产的胶回收柱回收得到 20ul 的体系中。在连接反应体系 10ul 中加入外源片段 5.5ul 和经过 NheI 和 HindIII 酶切的载体 pET28b 3ul 以及缓冲液 1ul 和酶 0.5ul 在 16℃连接过夜，65℃灭活 10 分钟。取 5ul 加入 100ul 转化效率达 10^7 的大肠杆菌 DH5 α ，冰上放置 30 分钟，42℃热激 70 秒，加 500ul LB 液体培养基 37℃培养 45 分钟复壮，取 200ul 涂在含有规定浓度的卡那霉素平板。37℃培养过夜，挑取单菌落接入试管。按照《分子克隆》中所述的方法小量抽提质粒 pXZpH2，用 0.5ul NheI 和 0.5ul HindIII 在 20ul 含有相应的反应缓冲液的反应体系中进行限制性酶切反应，反应两小时后用 0.8%的琼脂糖进行凝胶电泳，5V/cm 电泳 30 分钟，以验证所挑取的克隆是否正确。

将构建好的表达载体 pXZpH2 用氯化钙法转化 E.coli BL21(DE3)，得到大肠杆菌 BL21(DE3)-pXZpH2 工程菌。将获得的携带质粒 pXZpH2 的大肠杆菌于 2000 年 12 月 6 日并于 2000 年 11 月 6 日保藏在中国微生物菌种保藏管理委员会普通微生物中心(CGMCC，中国，北京)，保藏号为 CGMCC No. 0520.1。

实施例 7

目的基因的表达

挑取单菌落接入含有 50ug/ml 卡那霉素的 LB 试管，培养过夜，再按 1%的接种量接入一管新的含有 50ug/ml 卡那霉素的 LB 试管中，37℃培养至 OD600 = 0.6 时，加入终浓度 1mM 的 IPTG，再培养 2 小时，分别收集诱导前和诱导后的菌体。将菌体全细胞进行 SDS-PAGE，分析蛋白表达量(见图 3)，其中海

因酶的表达量占细菌总蛋白的 30.8%，氨甲酰水解酶的表达量占细菌总蛋白的 42.2%。

实施例 8

5 带有目的片断的菌体活力测定

A. 氨甲酰氨基酸水解酶酶活测定

方法：将过夜培养的 pXZpC2/BL21(DE3)菌液，按 1%接种量接种 48mlLB 液体摇瓶，培养至 OD600=0.6 时，再加入 IPTG,终浓度为 1mM,37℃诱导 2 小时，离心收集菌体用 20ml 生理盐水洗涤，最后悬浮在 3ml pH8.0 NaH₂PO₄-Na₂HPO₄ 缓冲液中，稀释 50 倍测 OD600。在冰箱中冻存过夜，取经 40℃ 预热的菌液 0.2ml 和 0.2ml 1%底物浓度的 pH8.0 的 NaH₂PO₄-Na₂HPO₄ 缓冲液混匀在 40℃反应 30 分钟，反应结束后，立刻离心沉淀菌体，取 100ul 上清加入 100ul pH5.5 的 NaAc-HAc 缓冲液，再加入 3ml 0.5% ninhydrin 和 0.5% hydrindantin 的乙二醇甲醚溶液，60℃反应 30 分钟，然后振荡于室温稳定 1 小时，测定 OD₅₇₀。用已知浓度溶液 100ul 加入 100ul pH5.5 的 NaAc-HAc 缓冲液，再加入 3ml 0.5% ninhydrin 和 0.5% hydrindantin 的乙二醇甲醚溶液，60℃反应 30 分钟，然后振荡于室温稳定 1 小时，测定 OD570，绘制标准曲线，求出 K 值，根据公式，计算酶活。

结果如表 1 所示：

20 表 1: pXZpC2 氨甲酰氨基酸水解酶活力测定

N-氨甲酰-对-羟基苯甘氨酸	121.94 ug/ml · h	65.91 ug/OD · h	0.45U/ml
----------------	------------------	-----------------	----------

测试菌是大肠杆菌-pXZpC2

酶活力(U/ml)=K×OD570×4×15/0.5×48×M

1U 酶活定义为每小时产生 1um 相应产物所需酶量。

25 B. 乙内酰脲水解酶酶活测定

方法：将过夜培养的 pXZpH2/BL21(DE3)菌液，按 1%接种量接种 48mlLB 液体摇瓶，培养至 OD600=0.6 时，再加入 IPTG,终浓度为 1mM,37℃诱导 2 小时，离心收集菌体用 20ml 生理盐水洗涤，最后悬浮在 3ml pH8.0 NaH₂PO₄-Na₂HPO₄ 缓冲液中，稀释 50 倍测 OD600。将菌液分装为每管 1ml，离心，菌体悬浮在 1.2ml 合适底物溶液（1%的对羟基苯海因，海因，0.5%的二氢尿嘧啶）的 pH8.0 的 NaH₂PO₄-Na₂HPO₄ 缓冲液，每管 0.3ml，40℃反应 1 小时，立刻离心取上清 250ul 加入 550ul 5%对二甲胺基苯甲醛的 6N 盐酸溶液，测

定 OD₄₃₈。用已知浓度溶液 250ul 加入 550ul 5% 对二甲氨基苯甲醛的 6N 盐酸溶液，测定 OD₄₃₈，绘制标准曲线，求出 K 值，根据公式，计算酶活。

结果如表 2 所示：

表 2: pXZpH2 乙内酰胺酶活力测定结果

对-羟基苯海因	138.78ug/ml · h	100.48ug/OD · h	0.83U/ml
海因	68.28 ug/ml · h	49.43ug/OD · h	0.58U/ml
二氢尿嘧啶	85.15ug/ml · h	61.65ug/OD · h	0.64U/ml

5 测试菌是大肠杆菌-pXZpH2

酶活力(U/ml)=K×OD₄₃₈×300×12/250×48×M

1U 酶活定义为每小时产生 1um 相应产物所需酶量。

实施例 9

10 双酶法生产 D-对羟基苯甘氨酸(D-pHPG)

取按实施例 8，改用 28℃ 诱导得到的 0.5 克带有 pXZpC2 的大肠杆菌的湿菌体细胞，用 0.035% 的 0.1M pH8.0 的磷酸钠缓冲液进行转化反应，反应分别进行 1、2、3、5 小时后，测定体系中有 0.45mg、0.81mg、1.18mg、1.26mg 的 D-对羟基苯甘氨酸生成。5 小时后转化率为 90.8%。

15 取按实施例 8，改用 28℃ 诱导得到的 0.4 克带有 pXZpH2 的大肠杆菌的湿菌体细胞，用 0.45% 的 0.1M pH8.0 的磷酸钠缓冲液进行转化反应，反应分别进行 1、2、3、5 小时后，测定体系中有 14.64mg、20.89mg、23.23mg、23.75mg 的 N-氨甲酰-对羟基苯甘氨酸生成。5 小时后转化率达 96.7%。

所以当采用双酶法生产 D-对羟基苯甘氨酸时，5 小时后转化率达 87.5%。

20

修改在本发明提及的所有文献都在本申请中引用作为参考，就如同每一篇文章被单独引用作为参考那样。此外应理解，在阅读了本发明的上述讲授内容之后，本领域技术人员可以对本发明作各种改动或修改，这些等价形式同样落于本申请所附权利要求书所限定的范围。

25

序列表

(1)一般信息:

- 5 (ii)发明名称: 新的乙内酰胺水解酶及其应用
(iii)序列数目 :4

(2)SEQ ID NO: 1的信息:

(i)序列特征:

- 10 (A)长度: 3439bp
(B)类型: 核酸
(C)链性: 双链
(D)拓扑结构: 线性

(ii)分子类型: DNA

15 (xi)序列描述: SEQ ID NO: 1:

```

tctcggggtc cgtcaccacg atgaccgaac acggtctgca tgttggagga tgtcgcagcc      60
gactgaccgg cgcctatcag catttccagc gtgcctgggg ggagaggttc ccccgtgtag      120
cgtctgacgg aacgggtgctt cagcatcagc gcgatgcagt ccgacaccag ccgttcgccc      180
gatggtatct gattgaatgt cggtgcgccg taacgcctgt cgaaggcgct tgcaacgctt      240
20 gtcgccgta ggcgcgacga cacttgctga agtgtgtatt gtcacatcg ttctcgtaga      300
atgccggttt actgcttgta tttgcgacgt ttcaggaatt tcccgtcctg cggttcgccc      360
acataggaac cttcgtcgac gataacctg ccacgcagga gcaccgtctt cggcacgccc      420
ttgacctgt gtccctcgta ggaggagtaa tccatggcgt tgtgcatggc ggtctgttgc      480
atcaccattt cggcctcggg gtcccaaagg acgatgtcgg cgtccgaacc gaccgcgac      540
25 gtcccctttt gcggaacat tccgaagacc ttggccgggc gcgtggcgac cagtccaacg      600
aactgggtaa gggaaatccg gccttcgctt acgccctgat agaccatcat caaccgttcc      660
tcgacgcccg gcgcgccgtt cgggatggcg cgaagtcgt tccggccccg gtccttgtgc      720
cccttgaaga gccaggagca atggtcggag gaaaccgtt cgaacacacc gtttctgagt      780
gcgttccaga gaacgtcatg gtctttcttc gcgcgggccg gcggtgtgaa aacgtatttc      840
30 gcaccttga aatccggccg ctccaggtct tccttggta ggtaaaggta atgcgtgcag      900
gtttccgcca gagcgcggac gcctcgcgat tttgcgcgca tcacctcctc aagggactcc      960
tcgcaggtea catggactat gtagatcggg gcgttgacga tttcggccag ggcgagggcc      1020
cgcgcggttg cctcggcttc gaccgggggc gggcggtga gcgcgtggta gatcggcgcg      1080
gttttgcct cggccacgaa cttgtcgcgc agatagtcgg cggcgtcgcc gttttccgcg      1140
35 tgcaccatga cgagcgatcc ggtcttgacc gccttgctga gcgtcttcag cagcgtcacg      1200
tcgtcgatca tgttcatgcc gcgataggcc atgaagacct tgaaggaggt aatgccaaga      1260
tcgggaagca cctccagctc ctcaatcacg ctgtcggctg ggtcagacac gatgatgtgg      1320
tagccgtaat cgatcgccga cttgccgccc gccataccgt cccacttggc gacggcttcc      1380
gccaggctgt ggccgcgacg ctgctgacag aaatcgacga tggttgtcgt tccgccacag      1440

```

```

gcggccgcga ccgtcgtgt tgcgaacgtg tcggccgact gcgtgttgaa gctgaccgtt 1500
tcgacatgcg tgtgaacgtc tatgccgcc gaaagacgt agcggccggc cgcgtcgatc 1560
gtccgctccg ctgggccgag cgcgccgcc atctgggtga tcttgccatc cttgatcccg 1620
agatcgggcg gagaaatgcc atccgcggtc acgatggttc cgtttttgat aatgatgtcc 1680
5 atgatgtcgc tctcaggatt gaagagttt ttcogtata ggaatgtaat tctttataca 1740
agtttgcag gtgttttgaa gctgtcaagc ggaacggcgg cctagcgaag acgattctgt 1800
gcagagcgct ctcgttcgag gcggaacgag ttgaagtcga cggcgggctc gcgcgagagc 1860
ttgtcaagca gcgcaaattc cggttccgct ccggttgaca gatcaaaaat tttacgcctg 1920
ttattgtcgt gctgcatgta atatttcgta ctttatgtag aatttgcatt gcgccgag 1980
10 tcacaaagcc ggttttcggc gatgtgttc acaacgtttt cccggccgct gggccggaca 2040
tcacctagga aggagcagag gttcatgaca cgtcagatga tacttgcagt gggacaacaa 2100
ggtccgatcg cgcgcgcgga gacacgcgaa caggtcgtcg ttcgtcttct ctacatgctg 2160
acgaaagccg cgagccgggg cgcgaatttc attgtcttcc ccgaactcgc gcttacgacc 2220
ttcttcccgc gctggcattt caccgacgag gccgagctcg atagcttcta tgagaccgaa 2280
15 atgccggcc cggtggtccg tccactctt gagaaggccg cggaactcgg gatcggcttc 2340
aatctgggct acgctgaact cgtcgtcga ggcggcgtca agcgtcgtt caacacgtcc 2400
atthtggggg ataagtcagg caagatcgtc ggcaagatc gtaagatcca tttgccgggt 2460
cacaaggagt acgaggccta ccggccgttc cagcatctg aaaagcgtta tttcagccg 2520
ggcgatctcg gcttcccggg ctatgacgtc gacgccgca aaatggggat gttcatctgc 2580
20 aacgatgcc gctggcctga agcctggcgg gtgatgggcc tcaggggccc cgagatcacc 2640
tgcggcggtt acaacacgcc gaccacaat cccctgttc cccagcacga ccacctgacg 2700
tccttcacc atctctatc gatgcaggcc ggtcttacc agaacggggc ctggctccgcg 2760
gcccggggca aggtgggcat ggaggagaac tgcgatctgc tcggccactc ctgcatcgtg 2820
gcgccgaccg gggaaatcgt cgctctcact acgacgttg aagacgaggt gatcaccgcc 2880
25 gccgtcgatc tgatecgtg ccgggaactg cgtgaacaca tcttcaact caagcagcat 2940
cgtcagcccc agcactatgg tctgatcgcg gaactctgag gttgccgaaa agcatgtgtg 3000
tcgttgttct cggcgcctgg gtcacatcca ggcgcgccag ggtgacgtg gtggaatagt 3060
accacgaccg cttcaggcg atccgcaagg agatgcgggt cgcggagcg gcaaagcccg 3120
acattcgttt cgcaccgacg gccgtcgtga actcagacagt ccgcgagaag ggcgtattgc 3180
30 gcggcctgga cctgtacgtg gaactgtagc ccatatatag atttcaaag agtttcggcg 3240
aggecgggcg cgcctagccc catgtgagcg agaaccgtgc ccagatcaaa gaatgagacc 3300
gacgcgccgg ccgcggcaaa ggatgatcct cagggtcggg tctatcgctc cgccctgaag 3360
caggagggcg cacgctggct gcttgacggc ggaggaaggg gttgctggca aagcccaagc 3420
cgcccggcct tgttccggc ~439

```

35

(2)SEQ ID NO: 2的信息:

(i)序列特征:

(A)长度: 304个氨基酸

(B)类型: 氨基酸

(D)拓扑结构: 线性

40

(ii)分子类型: 多肽

(xi)序列描述: SEQ ID NO: 2:

MTRQMILAVG QQGPIARAET REQVVRLLY MLTKAASRGA NFIVFPELAL 50
 TTFPRWHFT DEAELDSFYE TEMPGPVVRP LFEKAAELGI GFNLGYAELV 100
 VEGGVKRRFN TSILVDKSGK IVGKYRKIHL PGHKEYEAYR PFQHLEKRYF 150
 5 EPDGLGFPVY DVDAAKMGFM ICNDRRWPEA WRVMGLRGAE IICGGYNTPT 200
 HNPPVPQHDH LTSFHHLLSM QAGSYQNGAW SAAAGKVGME ENCMLLGHSC 250
 IVAPTGEIVA LTTTLEDEVI TAAVDLDRCR ELREHIFNFK QHRQPQHYGL 300
 IAEL 304

10 (2)SEQ ID NO:3 的信息

(i)序列特征

(A)长度: 3439碱基

(B)类型: 核酸

(C)链性: 单链

15 (D)拓扑结构: 线性

(ii)分子类型: 寡核苷酸

(xi)序列描述: SEQ ID NO : 3:

gccggaacaa ggccgggcg cttgggctt gccagcaacc ccttcctccg ccgtcaagca 60
 gccagcgtgc gccctcctgc ttcaggggcg agcgatagat ccgaccctga ggatcatcct 120
 20 ttgccgcggc cggcgcgctc gtctcattct ttgatctggg cacggttctc gctcacatgg 180
 ggctaggcgc gccgcgcctc gccgaaactc tttggaaatc tatatatggg ctacagtcc 240
 acgtacaggt ccaggccgcg caatacgccc ttctcgcgga ctgtcgagtt cacgacggcc 300
 gtcggtgcga aacgaatgtc gggctttgcc gctccggcga cccgcatctc cttgaggatc 360
 gccctgaagc ggtcgtggta ctattccacc agcgtcacc tggcgcgcct ggatgtgacc 420
 25 caggcggcga gaacaacgac acacatgctt ttcggcaacc tcagagttcc gcgatcagac 480
 catagtctg gggctgacga tgtgcttga agttgaagat gtgttcacgc agttccggc 540
 agcgatcgag atcgacggcg gcggtgatca cctcgtctc cagcgtcgta gtgagagcga 600
 cgatttcccc ggtcggcgcc acgatgcagg agtggccgag cagcatgcag ttctctcca 660
 tgcccacctt gcccggggcc gcggaccagg ccccgttctg ataagacctg gcctgcatcg 720
 30 ataggagatg gtggaaggac gtcagggtgt cgtgctggg aacaggggga ttgtgggtcg 780
 gcgtgttgta gccgccgag atgatctcgg cggccctgag gccatcacc cggcaggctt 840
 caggccagcg gcgatcgtt cagatgaaca tccccattt cgcggcgtc acgtcataga 900
 ccgggaagcc gagatcggc ggctcgaaat aacgcttttc aagatgctgg aacggccggt 960
 aggcctcgta ctcttgtga cccggcaaat ggatcttacg atacttgcc acgatcttgc 1020
 35 ctgacttate caccaaaatg gacgtgttga agcgacgctt gacgccgct tcgacgacga 1080
 gttcagcgta gccagattg aagccgatcc cgagttccgc gcccttctca aagagtggac 1140
 ggaccaccgg gccgggcatc tcggtctcat agaagctatc gagctcggcc tcgctcggta 1200
 aatgccagcg cgggaagaag gtcgtaagcg cgagttcggg gaagacaatg aaattcgcgc 1260
 cccggctcgc ggcttctgtc agcatgtaga gaagacgaac gacgacctgt tcgctgtct 1320
 40 ccgcgcgcg gatcggacct tgttgtccca ctgcaagtat catctgacgt gtcatgaacc 1380

	tctgctcctt cctaggtgat gtccggccca gcgccggga aaacgttgtg aaacacatcg	1440
	ccgaaaaccg gctttgtgac tcgcggcgca atgcaaattc tacataaagt acgaaatatt	1500
	acatgcagca cgacaataac aggcgtaaaa tttttgatct gtcaaccgga gcggaaccgg	1560
	aatttgcgct gcttgacaag ctctcgcgcg agcccgcctg cgacttcaac tcgttccgcc	1620
5	tcgaacgaga gcgctctgca cagaatcgtc ttcgctaggc cgccgttccg cttgacagct	1680
	tcaaaacacc atgcaaactt gtataaagaa ttacattcca tatacggaa aaactcttca	1740
	atcctgagag cgacatcatg gacatcatta tcaaaaacgg aaccatcgtg accgcggatg	1800
	gcatttctcg cgccgatctc gggatcaagg atggcaagat caccagatc ggcggcgcgc	1860
	tcggcccagc ggagcggacg atcgacgcgg ccggccgcta cgtctttccg ggcggcatag	1920
10	acgttcacac gcatgtcgaa acggtcagct tcaacacgca gtcggccgac acgttcgcaa	1980
	cagcgacggt gcgcccgcgc tgtggcggaa cgacaacat cgtcgatttc tgtcagcagg	2040
	atcgcggcca cagcctggcg gaagccgtcg ccaagtggga cggatggcc ggcggcaagt	2100
	cgcgatcga ttacggctac cacatcatcg tgctcgacc gaccgacagc gtgattgagg	2160
	agctggaggt gcttcccgat cttggcatta cctccttcaa ggtcttcatg gcctatcgcg	2220
15	gcatgaacat gatcgacgac gtgacgtgc tgaagacgct cgacaaggcg gtcaagaccg	2280
	gatcgtcgt catggtgcac gcgaaaacg gcgacgccgc cgactatctg cgcgacaagt	2340
	tcgtggccga gggcaaaacc gcgccgatct accacgcgct cagccgcccg ccccggtcg	2400
	aagccgaggc aaccgcgcgg gccctgccc tggccgaaat cgtcaacgcc ccgatctaca	2460
	tagtccatgt gacctgcgag gagtcccttg aggaggtgat gcgcgcaaaa tcgcgaggcg	2520
20	tccgcgtct ggcgaaacc tgcacgatt acctttacct caccaaggaa gacctggagc	2580
	ggccggttt cgaaggtgca aaatacgtt tcacaccgcc ggcccgcgcg aagaaagacc	2640
	atgacgttct ctggaacgca ctcagaaacg gtgtgttga aacggtttcc tccgaccatt	2700
	gctcctggct cttcaagggg cacaaggacc ggggccggaa cgactttcgc gccatcccga	2760
	acggcgcgcc gggcgtcgag gaacggttga tgatgttcta tcagggcgtc aacgaaggcc	2820
25	ggatttcct taccagttc gtggaactgg tcgccacgcg cccggccaag gtcttcggaa	2880
	tgttccgca aaaggggacg atcgcggtcg gttcggacgc cgacatcgtc ctttgggacc	2940
	ccgaggccga aatggtgatc gaacagaccg ccatgcaca cgccatggat tactcctcct	3000
	acgagggaca caaggtcaag ggcgtgccga agacggtgct cctgcgtggc aaggttatcg	3060
	tcgacgaagg ttctatgtc ggcgaaccga cggacgggaa attcctgaaa cgtcgcaaat	3120
30	acaagcagta aaccggcatt ctacgagaac gatgatgaca atacacactt cgacaagtgt	3180
	cgtcgcgcct gacggcgaca agcgttgcaa gcgccttga caggcgttac ggcgcaccga	3240
	cattcaatca gataccatcg ggcgaacggc tgggtcggga ctgcatcgcg ctgatgctga	3300
	agcaccgttc cgtcagacgc tacacggggg aacctctccc cccaggcacg ctggaaatgc	3360
	tgatggcggc gggtcagtcg gctgcgacat cctccaacat gcagaccgtg ttcggtcac	3420
35	gtggtgacgg accccgaga	3439

(2)SEQ ID NO: 4的信息:

(i)序列特征:

(A)长度: 457个氨基酸

40

(B)类型: 氨基酸

(D)拓扑结构: 线性

(ii)分子类型: 多肽

(xi)序列描述: SEQ ID NO: 4:

	MDIIKNGTI	VTADGISRAD	LGIKDGKITQ	IGGALGPAER	TIDAAGRYVF	50
	PGGIDVHTHV	ETVSFNTQSA	DTFATATVAA	ACGGTTTIVD	FCQQDRGHSL	100
	AEAVAKWDGM	AGGKSAIDYG	YHIIIVLDPTD	SVIEELEVLP	DLGITSFKVF	150
5	MAYRGMNMID	DVTLLKTLDK	AVKTGSLVMV	HAENGDAADY	LRDKFVAEGK	200
	TAPIYHALSR	PPRVEAEATA	RALALAEIVN	APIYIVHVTC	EESLEEV MRA	250
	KSRGVRALAE	TCTHYLYLTK	EDLERPDFEG	AKYVFTPPAR	AKKDHDVLWN	300
	ALRNGVFETV	SSDHCSWLFK	GHKDRGRNDF	RAIPNGAPGV	EERLMMVYQG	350
	VNEGRISLTQ	FVELVATRPA	KVFGMFPQKG	TIAVGSDADI	VLWDPEAEMV	400
10	IEQTAMHNAM	DYSSYEGHKV	KGVPKTVLLR	GKVIVDEGSY	VGEPTDGKFL	450
	KRRKYKQ					457

15

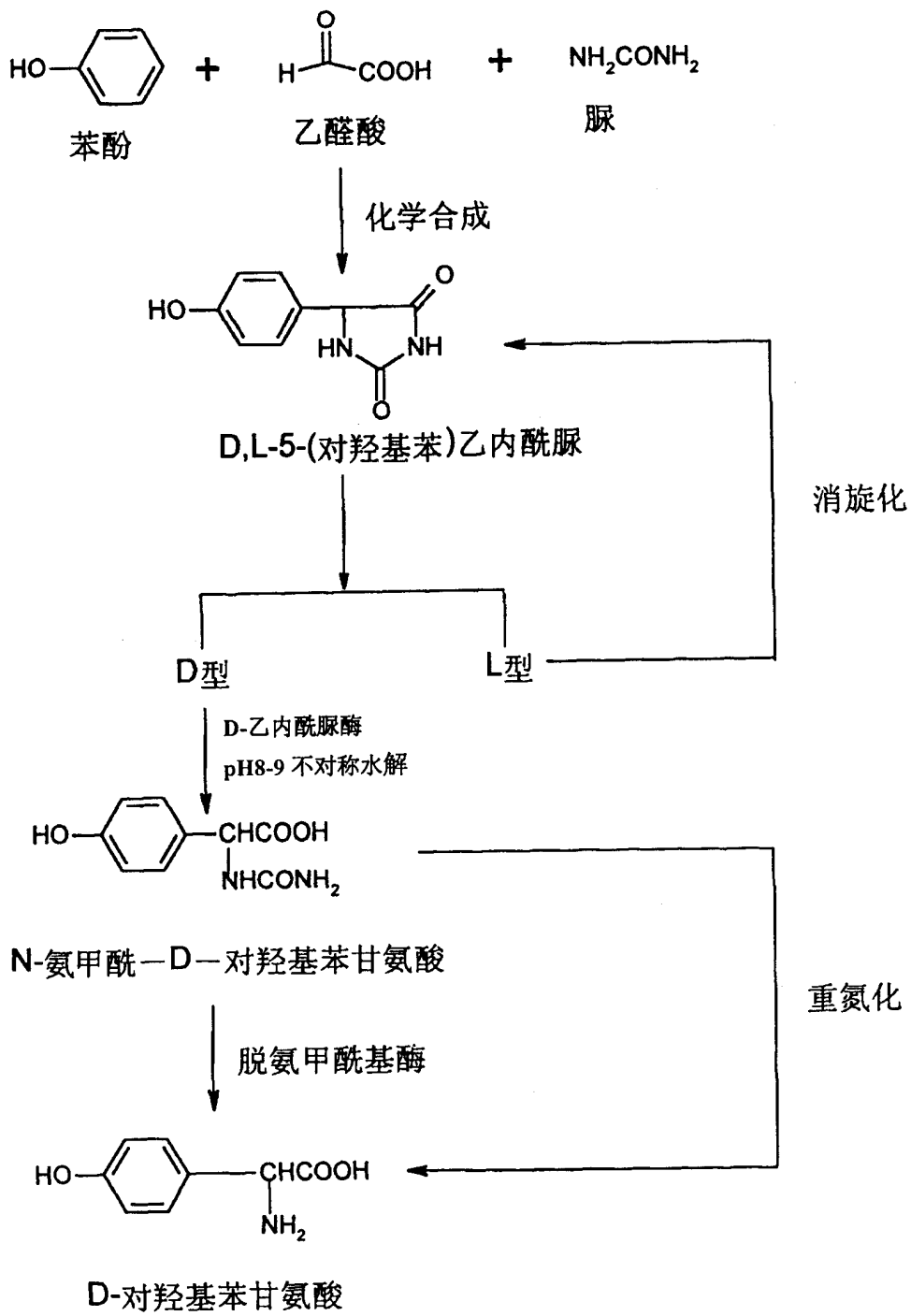


图 1

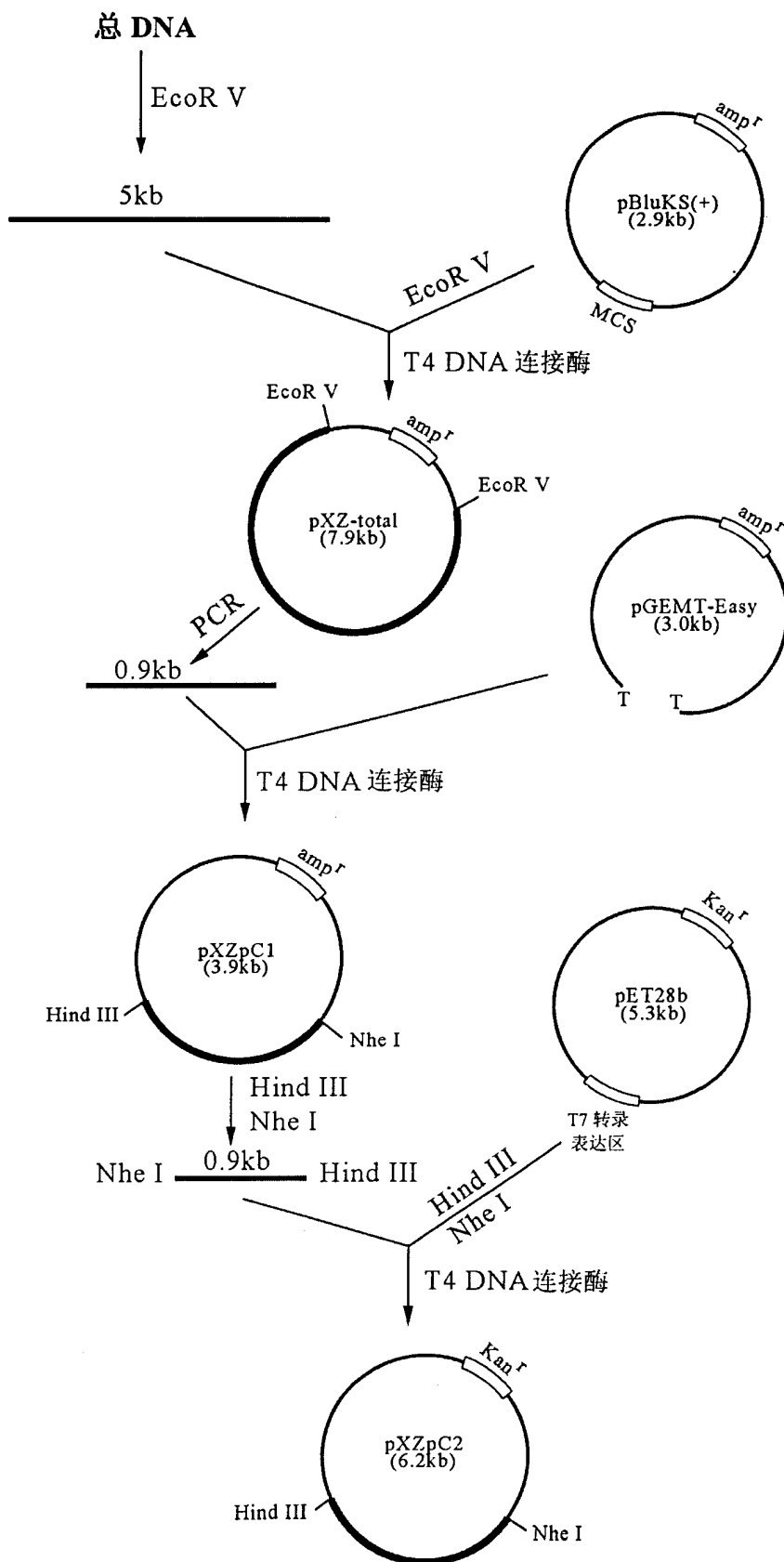


图 2A

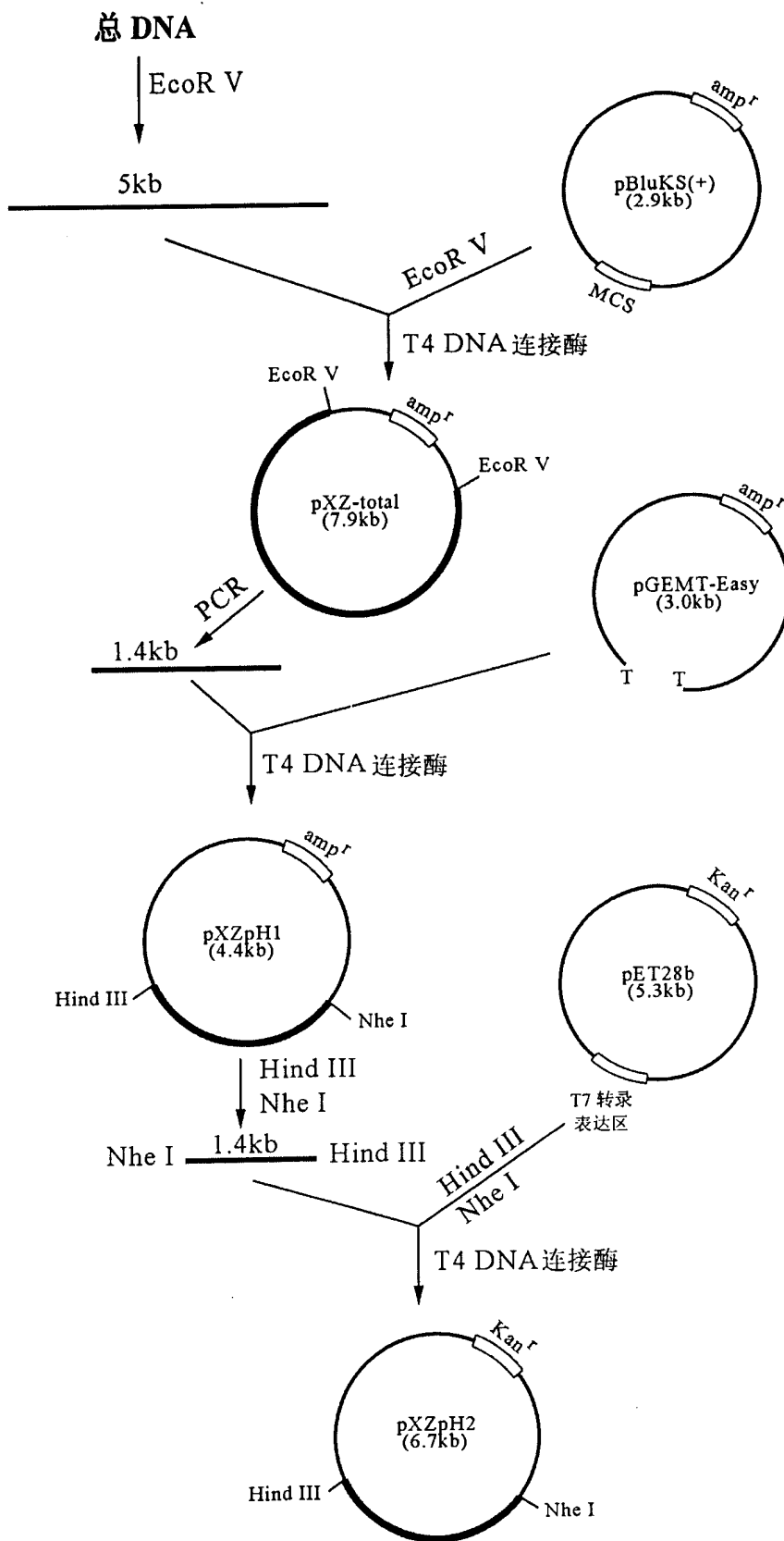


图 2B

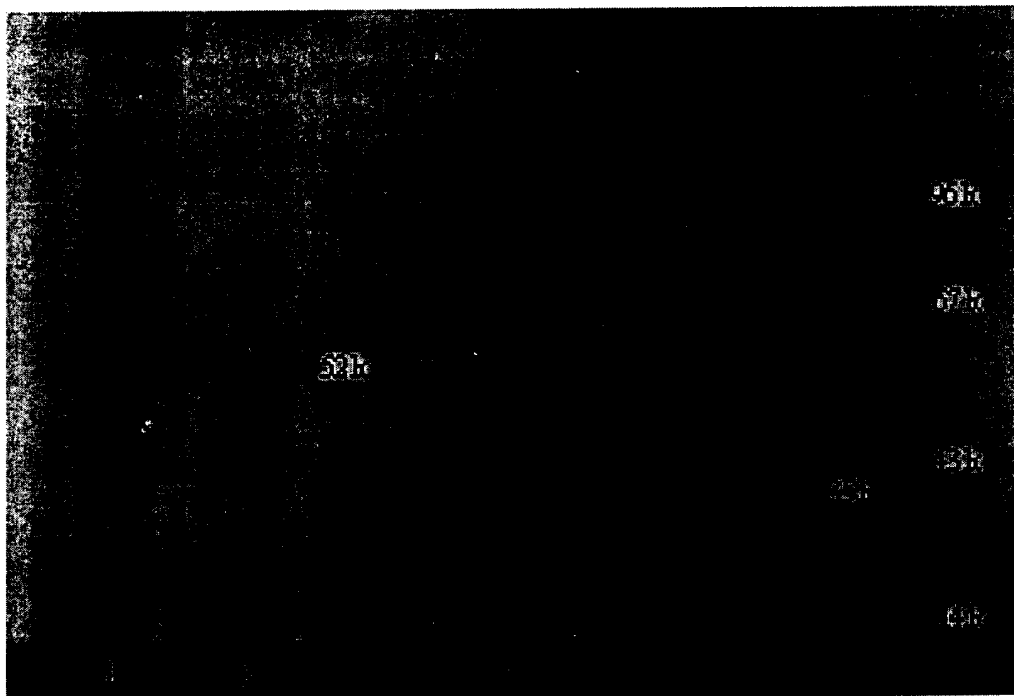


图 3

1	TCT CGG GGT CCG TCA CCA CGA TGA CCG AAC ACG GTC TGC ATG TTG	45
46	GAG GAT GTC GCA GCC GAC TGA CCC GCC GCC ATC AGC ATT TCC AGC	90
91	GTG CCT GGG GGG AGA GGT TCC CCC GTG TAG CGT CTG ACG GAA CGG	135
136	TGC TTC AGC ATC AGC GCG ATG CAG TCC GAC ACC AGC CGT TCG CCC	180
181	GAT GGT ATC TGA TTG AAT GTC GGT GCG CCG TAA CGC CTG TCG AAG	225
226	GCG CTT GCA ACG CTT GTC GCC GTC AGG CGC GAC GAC ACT TGT CGA	270
271	AGT GTG TAT TGT CAT CAT CGT TCT CGT AGA ATG CCG GTT TAC TGC	315
316	TTG TAT TTG CGA CGT TTC AGG AAT TTC CCG TCC GTC GGT TCG CCG	360
361	ACA TAG GAA CCT TCG TCG ACG ATA ACC TTG CCA CGC AGG AGC ACC	405
406	GTC TTC GGC ACG CCC TTG ACC TTG TGT CCC TCG TAG GAG GAG TAA	450
451	TCC ATG GCG TTG TGC ATG GCG GTC TGT TCG ATC ACC ATT TCG GCC	495
496	TCG GGG TCC CAA AGG ACG ATG TCG GCG TCC GAA CCG ACC GCG ATC	540
541	GTC CCC TTT TGC GGA AAC ATT CCG AAG ACC TTG GCC GGG CGC GTG	585
586	GCG ACC AGT TCC ACG AAC TGG GTA AGG GAA ATC CGG CCT TCG TTG	630
631	ACG CCC TGA TAG ACC ATC ATC AAC CGT TCC TCG ACG CCC GGC GCG	675
676	CCG TTC GGG ATG GCG CGA AAG TCG TTC CCG CCC CGG TCC TTG TGC	720
721	CCC TTG AAG AGC CAG GAG CAA TGG TCG GAG GAA ACC GTT TCG AAC	765
766	ACA CCG TTT CTG AGT GCG TTC CAG AGA ACG TCA TGG TCT TTC TTC	810
811	GCG CGG GCC GGC GGT GTG AAA ACG TAT TTC GCA CCT TCG AAA TCC	855
856	GGC CGC TCC AGG TCT TCC TTG GTG AGG TAA AGG TAA TGC GTG CAG	900
901	GTT TCC GCC AGA GCG CGG ACG CCT CGC GAT TTT GCG CGC ATC ACC	945
946	TCC TCA AGG GAC TCC TCG CAG GTC ACA TGG ACT ATG TAG ATC GGG	990
991	GCG TTG ACG ATT TCG GCC AGG GCG AGG GCC CGC GCG GTT GCC TCG	1035
1036	GCT TCG ACC CGG GGC GGG CGG CTG AGC GCG TGG TAG ATC GGC GCG	1080
1081	GTT TTG CCC TCG GCC ACG AAC TTG TCG CGC AGA TAG TCG GCG GCG	1125
1126	TCG CCG TTT TCC GCG TGC ACC ATG ACG AGC GAT CCG GTC TTG ACC	1170
1171	GCC TTG TCG AGC GTC TTC AGC AGC GTC ACG TCG TCG ATC ATG TTC	1215
1216	ATG CCG CGA TAG GCC ATG AAG ACC TTG AAG GAG GTA ATG CCA AGA	1260
1261	TCG GGA AGC ACC TCC AGC TCC TCA ATC ACG CTG TCG GTC GGG TCG	1305
1306	AGC ACG ATG ATG TGG TAG CCG TAA TCG ATC GCC GAC TTG CCG CCG	1350
1351	GCC ATA CCG TCC CAC TTG GCG ACG GCT TCC GCC AGG CTG TGG CCG	1395
1396	CGA TCC TGC TGA CAG AAA TCG ACG ATG GTT GTC GTT CCG CCA CAG	1440
1441	GCG GCC GCG ACC GTC GCT GTT GCG AAC GTG TCG GCC GAC TGC GTG	1485
1486	TTG AAG CTG ACC GTT TCG ACA TGC GTG TGA ACG TCT ATG CCG CCC	1530
1531	GGA AAG ACG TAG CGG CCG GCC GCG TCG ATC GTC CGC TCC GCT GGG	1575
1576	CCG AGC GCG CCG CCG ATC TGG GTG ATC TTG CCA TCC TTG ATC CCG	1620
1621	AGA TCG GCG CGA GAA ATG CCA TCC GCG GTC ACG ATG GTT CCG TTT	1665
1666	TTG ATA ATG ATG TCC ATG ATG TCG CTC TCA GGA TTG AAG AGT TTG	1710
1711	TTC CGT ATA TGG AAT GTA ATT CTT TAT ACA AGT TTG CAT GGT GTT	1755
1756	TTG AAG CTG TCA AGC GGA ACG GCG GCC TAG CGA AGA CGA TTC TGT	1800
1801	GCA GAG CGC TCT CGT TCG AGG CGG AAC GAG TTG AAG TCG ACG GCG	1845
1846	GGC TCG CGC GAG AGC TTG TCA AGC AGC GCA AAT TCC GGT TCC GCT	1890
1891	CCG GTT GAC AGA TCA AAA ATT TTA CGC CTG TTA TTG TCG TGC TGC	1935
1936	ATG TAA TAT TTC GTA CTT TAT GTA GAA TTT GCA TTG CGC CGC GAG	1980
1981	TCA CAA AGC CGG TTT TCG GCG ATG TGT TTC ACA ACG TTT TCC CGG	2025
2026	CCG CTG GGC CGG ACA TCA CCT AGG AAG GAG CAG AGG TTC ATG ACA	2070
	M T	2
2071	CGT CAG ATG ATA CTT GCA GTG GGA CAA CAA GGT CCG ATC GCG CGC	2115
3	R Q M I L A V G Q Q G P I A R	17
2116	GCG GAG ACA CGC GAA CAG GTC GTC GTT CGT CTT CTC TAC ATG CTG	2160
18	A E T R E Q V V V R L L Y M L	32
2161	ACG AAA GCC GCG AGC CGG GGC GCG AAT TTC ATT GTC TTC CCC GAA	2205
33	T K A A S R G A N F I V F P E	47

图 4

2206	CTC GCG CTT ACG ACC TTC TTC CCG CGC TGG CAT TTC ACC GAC GAG	2250
48	L A L T T F F P R W H F T D E	62
2251	GCC GAG CTC GAT AGC TTC TAT GAG ACC GAA ATG CCC GGC CCG GTG	2295
63	A E L D S F Y E T E M P G P V	77
2296	GTC CGT CCA CTC TTT GAG AAG GCC GCG GAA CTC GGG ATC GGC TTC	2340
78	V R P L F E K A A E L G I G F	92
2341	AAT CTG GGC TAC GCT GAA CTC GTC GTC GAA GGC GGC GTC AAG CGT	2385
93	N L G Y A E L V V E G G V K R	107
2386	CGC TTC AAC ACG TCC ATT TTG GTG GAT AAG TCA GGC AAG ATC GTC	2430
108	R F N T S I L V D K S G K I V	122
2431	GGC AAG TAT CGT AAG ATC CAT TTG CCG GGT CAC AAG GAG TAC GAG	2475
123	G K Y R K I H L P G H K E Y E	137
2476	GCC TAC CGG CCG TTC CAG CAT CTT GAA AAG CGT TAT TTC GAG CCG	2520
138	A Y R P F Q H L E K R Y F E P	152
2521	GGC GAT CTC GGC TTC CCG GTC TAT GAC GTC GAC GCC GCG AAA ATG	2565
153	G D L G F P V Y D V D A A K M	167
2566	GGG ATG TTC ATC TGC AAC GAT CGC CGC TGG CCT GAA GCC TGG CGG	2610
168	G M F I C N D R R W P E A W R	182
2611	GTG ATG GGC CTC AGG GGC GCC GAG ATC ATC TGC GGC GGC TAC AAC	2655
183	V M G L R G A E I I C G G Y N	197
2656	ACG CCG ACC CAC AAT CCC CCT GTT CCC CAG CAC GAC CAC CTG ACG	2700
198	T P T H N P P V P Q H D H L T	212
2701	TCC TTC CAC CAT CTC CTA TCG ATG CAG GCC GGG TCT TAT CAG AAC	2745
213	S F H H L L S M Q A G S Y Q N	227
2746	GGG GCC TGG TCC GCG GCC GCG GGC AAG GTG GGC ATG GAG GAG AAC	2790
228	G A W S A A A G K V G M E E N	242
2791	TGC ATG CTG CTC GGC CAC TCC TGC ATC GTG GCG CCG ACC GGG GAA	2835
243	C M L L G H S C I V A P T G E	257
2836	ATC GTC GCT CTC ACT ACG ACG CTG GAA GAC GAG GTG ATC ACC GCC	2880
258	I V A L T T T L E D E V I T A	272
2881	GCC GTC GAT CTC GAT CGC TGC CGG GAA CTG CGT GAA CAC ATC TTC	2925
273	A V D L D R C R E L R E H I F	287
2926	AAC TTC AAG CAG CAT CGT CAG CCC CAG CAC TAT GGT CTG ATC GCG	2970
288	N F K Q H R Q P Q H Y G L I A	302
2971	GAA CTC TGA GGT TGC CGA AAA GCA TGT GTG TCG TTG TTC TCG GCG	3015
303	E L *	
3016	CCT GGG TCA CAT CCA GGC GCG CCA GGG TGA CGC TGG TGG AAT AGT	3060
3061	ACC ACG ACC GCT TCA GGG CGA TCC GCA AGG AGA TGC GGG TCG CCG	3105
3106	GAG CGG CAA AGC CCG ACA TTC GTT TCG CAC CGA CGG CCG TCG TGA	3150
3151	ACT CGA CAG TCC GCG AGA AGG GCG TAT TGC GCG GCC TGG ACC TGT	3195
3196	ACG TGG AAC TGT AGC CCA TAT ATA GAT TTC CAA AGA GTT TCG GCG	3240
3241	AGG CGC GGC GCG CCT AGC CCC ATG TGA GCG AGA ACC GTG CCC AGA	3285
3286	TCA AAG AAT GAG ACC GAC GCG CCG GCC GCG GCA AAG GAT GAT CCT	3330
3331	CAG GGT CGG ATC TAT CGC TCC GCC CTG AAG CAG GAG GGC GCA CGC	3375
3376	TGG CTG CTT GAC GGC GGA GGA AGG GGT TGC TGG CAA AGC CCA AGC	3420
3421	CGC CCG GCC TTG TTC CGG C	3439

图 4(续)

		GC	2
3	CGG AAC AAG GCC GGG CGG CTT GGG CTT TGC CAG CAA CCC CTT CCT	47	
48	CCG CCG TCA AGC AGC CAG CGT GCG CCC TCC TGC TTC AGG GCG GAG	92	
93	CGA TAG ATC CGA CCC TGA GGA TCA TCC TTT GCC GCG GCC GGC GCG	137	
138	TCG GTC TCA TTC TTT GAT CTG GGC ACG GTT CTC GCT CAC ATG GGG	182	
183	CTA GGC GCG CCG CGC CTC GCC GAA ACT CTT TGG AAA TCT ATA TAT	227	
228	GGG CTA CAG TTC CAC GTA CAG GTC CAG GCC GCG CAA TAC GCC CTT	272	
273	CTC GCG GAC TGT CGA GTT CAC GAC GGC CGT CGG TGC GAA ACG AAT	317	
318	GTC GGG CTT TGC CGC TCC GGC GAC CCG CAT CTC CTT GCG GAT CGC	362	
363	CCT GAA GCG GTC GTG GTA CTA TTC CAC CAG CGT CAC CCT GGC GCG	407	
408	CCT GGA TGT GAC CCA GGC GCC GAG AAC AAC GAC ACA CAT GCT TTT	452	
453	CGG CAA CCT CAG AGT TCC GCG ATC AGA CCA TAG TGC TGG GGC TGA	497	
498	CGA TGC TGC TTG AAG TTG AAG ATG TGT TCA CGC AGT TCC CGG CAG	542	
543	CGA TCG AGA TCG ACG GCG GCG GTG ATC ACC TCG TCT TCC AGC GTC	587	
588	GTA GTG AGA GCG ACG ATT TCC CCG GTC GGC GCC ACG ATG CAG GAG	632	
633	TGG CCG AGC AGC ATG CAG TTC TCC TCC ATG CCC ACC TTG CCC GCG	677	
678	GCC GCG GAC CAG GCC CCG TTC TGA TAA GAC CCG GCC TGC ATC GAT	722	
723	AGG AGA TGG TGG AAG GAC GTC AGG TGG TCG TGC TGG GGA ACA GGG	767	
768	GGA TTG TGG GTC GGC GTG TTG TAG CCG CCG CAG ATG ATC TCG GCG	812	
813	CCC CTG AGG CCC ATC ACC CGC CAG GCT TCA GGC CAG CGG CGA TCG	857	
858	TTG CAG ATG AAC ATC CCC ATT TTC GCG GCG TCG ACG TCA TAG ACC	902	
903	GGG AAG CCG AGA TCG CCC GGC TCG AAA TAA CGC TTT TCA AGA TGC	947	
948	TGG AAC GGC CGG TAG GCC TCG TAC TCC TTG TGA CCC GGC AAA TGG	992	
993	ATC TTA CGA TAC TTG CCG ACG ATC TTG CCT GAC TTA TCC ACC AAA	1037	
1038	ATG GAC GTG TTG AAG CGA CGC TTG ACG CCG CCT TCG ACG ACG AGT	1082	
1083	TCA GCG TAG CCC AGA TTG AAG CCG ATC CCG AGT TCC GCG GCC TTC	1127	
1128	TCA AAG AGT GGA CGG ACC ACC GGG CCG GGC ATT TCG GTC TCA TAG	1172	
1173	AAG CTA TCG AGC TCG GCC TCG TCG GTG AAA TGC CAG CGC GGG AAG	1217	
1218	AAG GTC GTA AGC GCG AGT TCG GGG AAG ACA ATG AAA TTC GCG CCC	1262	
1263	CGG CTC GCG GCT TTC GTC AGC ATG TAG AGA AGA CGA ACG ACG ACC	1307	
1308	TGT TCG CGT GTC TCC GCG CGC GCG ATC GGA CCT TGT TGT CCC ACT	1352	
1353	GCA AGT ATC ATC TGA CGT GTC ATG AAC CTC TGC TCC TTC CTA GGT	1397	
1398	GAT GTC CGG CCC AGC GGC CGG GAA AAC GTT GTG AAA CAC ATC GCC	1442	
1443	GAA AAC CGG CTT TGT GAC TCG CGG CGC AAT GCA AAT TCT ACA TAA	1487	
1488	AGT ACG AAA TAT TAC ATG CAG CAC GAC AAT AAC AGG CGT AAA ATT	1532	
1533	TTT GAT CTG TCA ACC GGA GCG GAA CCG GAA TTT GCG CTG CTT GAC	1577	
1578	AAG CTC TCG CGC GAG CCC GCC GTC GAC TTC AAC TCG TTC CGC CTC	1622	
1623	GAA CGA GAG CGC TCT GCA CAG AAT CGT CTT CGC TAG GCC GCC GTT	1667	
1668	CCG CTT GAC AGC TTC AAA ACA CCA TGC AAA CTT GTA TAA AGA ATT	1712	
1713	ACA TTC CAT ATA CGG AAC AAA CTC TTC AAT CCT GAG AGC GAC ATC	1757	
1758	ATG GAC ATC ATT ATC AAA AAC GGA ACC ATC GTG ACC GCG GAT GGC	1802	
1	M D I I I K N G T I V T A D G	15	
1803	ATT TCT CGC GCC GAT CTC GGG ATC AAG GAT GGC AAG ATC ACC CAG	1847	
16	I S R A D L G I K D G K I T Q	30	
1848	ATC GGC GGC GCG CTC GGC CCA GCG GAG CGG ACG ATC GAC GCG GCC	1892	
31	I G G A L G P A E R T I D A A	45	
1893	GGC GCG TAC GTC TTT CCG GGC GGC ATA GAC GTT CAC ACG CAT GTC	1937	
46	G R Y V F P G G I D V H T H V	60	
1938	GAA ACG GTC AGC TTC AAC ACG CAG TCG GCC GAC ACG TTC GCA ACA	1982	
61	E T V S F N T Q S A D T F A T	75	
1983	GCG ACG GTC GCG GCC GCC TGT GGC GGA ACG ACA ACC ATC GTC GAT	2027	
76	A T V A A A C G G T T T I V D	90	

图 5

2028	TTC TGT CAG CAG GAT CGC GGC CAC AGC CTG GCG GAA GCC GTC GCC	2072
91	F C Q Q D R G H S L A E A V A	105
2073	AAG TGG GAC GGT ATG GCC GGC GGC AAG TCG GCG ATC GAT TAC GGC	2117
106	K W D G M A G G K S A I D Y G	120
2118	TAC CAC ATC ATC GTG CTC GAC CCG ACC GAC AGC GTG ATT GAG GAG	2162
121	Y H I I V L D P T D S V I E E	135
2163	CTG GAG GTG CTT CCC GAT CTT GGC ATT ACC TCC TTC AAG GTC TTC	2207
136	L E V L P D L G I T S F K V F	150
2208	ATG GCC TAT CGC GGC ATG AAC ATG ATC GAC GAC GTG ACG CTG CTG	2252
151	M A Y R G M N M I D D V T L L	165
2253	AAG ACG CTC GAC AAG GCG GTC AAG ACC GGA TCG CTC GTC ATG GTG	2297
166	K T L D K A V K T G S L V M V	180
2298	CAC GCG GAA AAC GGC GAC GCC GCC GAC TAT CTG CCG GAC AAG TTC	2342
181	H A E N G D A A D Y L R D K F	195
2343	GTG GCC GAG GGC AAA ACC GCG CCG ATC TAC CAC GCG CTC AGC CGC	2387
196	V A E G K T A P I Y H A L S R	210
2388	CCG CCC CGG GTC GAA GCC GAG GCA ACC GCG CGG GCC CTC GCC CTG	2432
211	P P R V E A E A T A R A L A L	225
2433	GCC GAA ATC GTC AAC GCC CCG ATC TAC ATA GTC CAT GTG ACC TGC	2477
226	A E I V N A P I Y I V H V T C	240
2478	GAG GAG TCC CTT GAG GAG GTG ATG CGC GCA AAA TCG CGA GGC GTC	2522
241	E E S L E E V M R A K S R G V	255
2523	CGC GCT CTG GCG GAA ACC TGC ACG CAT TAC CTT TAC CTC ACC AAG	2567
256	R A L A E T C T H Y L Y L T K	270
2568	GAA GAC CTG GAG CGG CCG GAT TTC GAA GGT GCG AAA TAC GTT TTC	2612
271	E D L E R P D F E G A K Y V F	285
2613	ACA CCG CCG GCC CGC GCG AAG AAA GAC CAT GAC GTT CTC TGG AAC	2657
286	T P P A R A K K D H D V L W N	300
2658	GCA CTC AGA AAC GGT GTG TTC GAA ACG GTT TCC TCC GAC CAT TGC	2702
301	A L R N G V F E T V S S D H C	315
2703	TCC TGG CTC TTC AAG GGG CAC AAG GAC CGG GGC CGG AAC GAC TTT	2747
316	S W L F K G H K D R G R N D F	330
2748	CGC GCC ATC CCG AAC GGC GCG CCG GGC GTC GAG GAA CGG TTG ATG	2722
331	R A I P N G A P G V E E R L M	345
2793	ATG GTC TAT CAG GGC GTC AAC GAA GGC CGG ATT TCC CTT ACC CAG	2837
346	M V Y Q G V N E G R I S L T Q	360
2838	TTC GTG GAA CTG GTC GCC ACG CGC CCG GCC AAG GTC TTC GGA ATG	2882
361	F V E L V A T R P A K V F G M	375
2883	TTT CCG CAA AAG GGG ACG ATC GCG GTC GGT TCG GAC GCC GAC ATC	2927
376	F P Q K G T I A V G S D A D I	390
2928	GTC CTT TGG GAC CCC GAG GCC GAA ATG GTG ATC GAA CAG ACC GCC	2972
391	V L W D P E A E M V I E Q T A	405
2973	ATG CAC AAC GCC ATG GAT TAC TCC TCC TAC GAG GGA CAC AAG GTC	3017
406	M H N A M D Y S S Y E G H K V	420
3018	AAG GGC GTG CCG AAG ACG GTG CTC CTG GGC AAG GTT ATC GTC	3062
421	K G V P K T V L L R G K V I V	435
3063	GAC GAA GGT TCC TAT GTC GGC GAA CCG ACG GAC GGG AAA TTC CTG	3107
436	D E G S Y V G E P T D G K F L	450
3108	AAA CGT CGC AAA TAC AAG CAG TAA ACC GGC ATT CTA CGA GAA CGA	3152
451	K R R K Y K Q *	
3153	TGA TGA CAA TAC ACA CTT CGA CAA GTG TCG TCG CGC CTG ACG GCG	3197
3198	ACA AGC GTT GCA AGC GCC TTC GAC AGG CGT TAC GGC GCA CCG ACA	3242
3243	TTC AAT CAG ATA CCA TCG GGC GAA CGG CTG GTG TCG GAC TGC ATC	3287
3288	GCG CTG ATG CTG AAG CAC CGT TCC GTC AGA CGC TAC ACG GGG GAA	3332
3333	CCT CTC CCC CCA GGC ACG CTG GAA ATG CTG ATG GCG CCG GGT CAG	3377
3378	TCG GCT GCG ACA TCC TCC AAC ATG CAG ACC GTG TTC GGT CAT CGT	3422
3423	GGT GAC GGA CCC CGA GA	3439

图 5(续)