

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷

C07K 14/36

C12P 21/04

C12N 15/68

C12Q 1/68



[12] 发明专利申请公开说明书

[21] 申请号 03817132.5

[43] 公开日 2005 年 9 月 14 日

[11] 公开号 CN 1668639A

[22] 申请日 2003.6.18 [21] 申请号 03817132.5

[30] 优先权

[32] 2002.6.21 [33] FR [31] 02/07728

[86] 国际申请 PCT/FR2003/001851 2003.6.18

[87] 国际公布 WO2004/000879 法 2003.12.31

[85] 进入国家阶段日期 2005.1.18

[71] 申请人 原子能委员会

地址 法国巴黎

共同申请人 国家研究中心

[72] 发明人 M·贡德里 R·热内 S·洛特吕

J-L·佩尔诺代

[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 唐伟杰

权利要求书 4 页 说明书 24 页 序列表 10 页
附图 7 页

[54] 发明名称 涉及二酮哌嗪衍生物合成的多核苷
酸和由所述多核苷酸编码的多肽

[57] 摘要

本发明涉及二酮哌嗪衍生物合成中涉及的新的分离的天然或合成的多核苷酸和由所述多核苷酸编码的多肽、含有所述多核苷酸的载体、用所述多核苷酸转化的微生物、所述多核苷酸和所述多肽的用途、以及合成二酮哌嗪衍生物的方法，所述二酮哌嗪衍生物包含环二肽和 3- 及 6- 位被 α ， β 不饱和氨基酸侧链取代的二酮哌嗪衍生物。

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

1. 分离的天然或合成的多核苷酸，其特征在于它包含至少对应于序列 SEQ ID NO.1、SEQ ID NO.2 和 SEQ ID NO.3 的三个开放阅读框。

2. 如权利要求 1 所请求保护的多核苷酸，其特征在于它还包含对应于序列 SEQ ID NO.4 的开放阅读框。

3. 分离的天然或合成的多核苷酸，其特征在于它对应于序列 SEQ ID NO.5。

4. 分离的天然或合成的多核苷酸，其特征在于它包含对应于序列 SEQ ID NO.2、SEQ ID NO.3 和 SEQ ID NO.4 的三个开放阅读框中的至少一个。

5. 对应于序列 SEQ ID NO.2、SEQ ID NO.3 和 SEQ ID NO.4 之中任一个的分离的天然或合成的多核苷酸。

6. 载体，其特征在于它含有权利要求 1 至 5 中任一项所请求保护的多核苷酸。

7. 如权利要求 6 所请求保护的载体，其特征在于它是质粒、粘粒、细菌人工染色体（BAC）、放线菌的整合元件、病毒或噬菌体。

8. 至少一种如权利要求 1 至 5 中任一项所请求保护的多核苷酸、或其具有至少 15 个核苷酸的片段之一、或如权利要求 6 或 7 任一项所请求保护的载体作为探针的用途。

9. 至少一种如权利要求 1 至 5 中任一项所请求保护的多核苷酸、或其具有至少 15 个核苷酸的片段之一、或如权利要求 6 或 7 任一项所请求保护的载体之一作为扩增核苷酸序列的引物的用途。

10. 分离的天然或合成的多肽，其特征在于它包含至少序列 SEQ ID NO.7 至 SEQ ID NO.10 之中的任一序列。

11. 分离的天然或合成的多肽，其特征在于它对应于序列 SEQ ID NO.7 至 SEQ ID NO.10 之中的任一序列。

12. 分离的天然或合成的多肽，其特征在于它由如权利要求 1 至 5 中任一项所请求保护的多核苷酸之一、或如权利要求 6 或 7 任一项

所请求保护的载体之一编码。

13. 如权利要求 1 至 5 中任一项所请求保护的多核苷酸、或如权利要求 6 或 7 任一项中所述的载体用于制备如权利要求 10 至 12 之中任一项所述的多肽的用途。

14. 至少一种如权利要求 10 至 12 之中任一项所请求保护的多肽单独或组合用于（特别是体外用于）制备环二肽和/或在 3-及 6-位被 α , β 不饱和氨基酸侧链取代的二酮哌嗪衍生物（特别是白诺斯菌素）的用途。

15. 至少一种如权利要求 1 至 5 中任一项所请求保护的多核苷酸、或如权利要求 6 或 7 任一项所述的载体用于制备经修饰的生物系统或经修饰的体外非细胞系统的用途。

16. 如权利要求 15 所请求保护的用途,其特征在于该经修饰的生物系统是一种微生物或是以原核细胞或真核细胞为宿主的异源表达系统。

17. 经修饰的生物系统,其特征在于它包含至少一种如权利要求 1 至 5 中任一项所述的多核苷酸和/或至少一种如权利要求 6 或 7 任一项所述的载体。

18. 如权利要求 17 中所述的生物系统,其特征在于它由微生物、或以原核细胞或真核细胞为宿主的异源表达系统、或体外非细胞系统构成。

19. 如权利要求 18 中所述的生物系统,其特征在于所述微生物是例如大肠杆菌或变青链霉菌的细菌。

20. 经修饰的体外非细胞系统,其特征在于它包含至少一种如权利要求 1 至 5 中任一项所述的多核苷酸和/或至少一种如权利要求 6 或 7 任一项所述的载体。

21. 至少一种如权利要求 17 至 19 中任一项所请求保护的经修饰的生物系统或如权利要求 20 所请求保护的经修饰的体外非细胞系统用于制备环二肽和/或在 3-及 6-位被 α , β 不饱和氨基酸侧链取代的二酮哌嗪衍生物（特别是白诺斯菌素）的用途。

22. 体外合成环二肽的方法，其特征在于：

(1) 在适宜条件下使两种可以相同或不同的氨基酸与多肽 A1bC (SEQ ID NO.9) 接触，和

(2) 纯化所得的环二肽。

23. 体外合成 3-及 6-位被氨基酸侧链取代的 α , β 不饱和二酮哌嗪衍生物的方法，其特征在于：

(1) 在适宜条件下使两种可以相同或不同的氨基酸与多肽 A1bC (SEQ ID NO.9) 接触并纯化所得的环二肽，和

(2) 使步骤 (1) 中获得的环二肽与 A1bA (SEQ ID NO.6)、A1bB1 (SEQ ID NO.7) 和 A1bB2 (SEQ ID NO.8) 接触，并纯化所得的 α , β 不饱和二酮哌嗪衍生物。

24. 如权利要求 23 所请求保护的方法，其特征在于该方法在步骤 (2) 中还包含多肽 A1bD (SEQ ID NO.10)。

25. 如权利要求 22 至 24 之中任一项所请求保护的方法，其特征在于多肽的量介于 0.1 nM 和 10 μ M 之间，优选介于 10 nM 和 1 μ M 之间。

26. 合成环二肽的方法，其特征在于：

(1) 在适宜培养所述已选择的生物系统的条件下接触至少含有多核苷酸 a1bC (SEQ ID NO.3) 的生物系统，和

(2) 纯化所得的环二肽。

27. 如权利要求 26 所请求保护的方法，其特征在于该生物系统还包含多核苷酸 a1bD (SEQ ID NO.4)。

28. 合成 3-及 6-位被 α , β 不饱和氨基酸侧链取代的二酮哌嗪衍生物的方法，其特征在于：

(1) 在适宜培养所述已选择的生物系统的条件下接触包含一种至少含有 a1bA、a1bB 和 a1bC (SEQ ID NO.1-3) 的多核苷酸的生物系统，和

(2) 纯化所得的 α , β 不饱和二酮哌嗪衍生物。

29. 如权利要求 28 所请求保护的方法，其特征在于该生物系统还

包含多核苷酸 a1bD (SEQ ID NO.4)。

30. 如权利要求 26 至 29 之中任一项所请求保护的方法，其特征在于该生物系统是一种微生物，例如如大肠杆菌或变青链霉菌的细菌，或任一已知的以原核细胞或真核细胞为宿主的异源表达系统，或者甚至是体外非细胞系统。

31. 如权利要求 22 至 30 之中任一项所请求保护的方法，其特征在于氨基酸的量介于 0.1 mM 和 100 mM 之间，优选介于 1 mM 和 10 mM 之间。

涉及二酮哌嗪衍生物合成的 多核苷酸和由所述多核苷酸编码的多肽

本发明涉及二酮哌嗪衍生物合成中涉及的新的分离的天然或合成的多核苷酸和由所述多核苷酸编码的多肽、含有所述多核苷酸的载体、用所述多核苷酸转化的微生物、所述多核苷酸和所述多肽的应用、以及合成二酮哌嗪衍生物的方法，所述二酮哌嗪衍生物包含环二肽和 3-及 6-位被 α 、 β 不饱和氨基酸侧链取代的二酮哌嗪衍生物。

为了本发明的目的，术语“二酮哌嗪衍生物”意指具有在 3-和 6-位被氨基酸取代的二酮哌嗪环（哌嗪-2,5-二酮或 2,5-二氧代哌嗪或 2,5-DKP）的分子。环状二氨基酸（环二肽或环状二肽）这一特殊情况下，3-和 6-位的取代基为氨基酸侧链。在双脱氢环状二氨基酸（双脱氢环状二肽）这一特殊情况下，3-和 6-位的取代基为 α 、 β -不饱和氨基酸侧链（附图 1）。

二酮哌嗪衍生物由一个化合物家族构成，这些化合物主要由微生物例如细菌、酵母、丝状真菌和地衣产生。也已从海洋生物，例如海绵和海星分离出了其它化合物。在人类中已经证实了这些衍生物的一个例子：环（L-His-L-Pro）。

二酮哌嗪衍生物具有从简单环二肽到复杂得多的结构的非常多变的结构。

简单环二肽仅构成二酮哌嗪衍生物中的一小部分，绝大多数二酮哌嗪衍生物具有更为复杂的结构，其中主环和/或侧链包含多种修饰：导入基于碳的、羟基、硝基、环氧基、乙酰基或甲氧基，和形成二硫桥或杂环。在两个碳原子间形成双键也是非常普遍的。源自海洋的特定衍生物中掺入了卤素原子。

下表 I 中给出了一些掺入环二肽的氨基酸例子：

表 I

环二肽	生物
环 (Gly-L-Pro)	<i>Luidia clathrata</i>
环 (L-Pro-L-Leu)	褐座坚壳菌 (<i>Rosellinia necatrix</i>)
环 (L-Ala-L-Val)	棕曲霉 (<i>Aspergillus ochraceus</i>)
环 (L-Ala-L-Leu)	黑曲霉 (<i>Aspergillus niger</i>)
环 (D-Ala-N-甲基-L-Leu)	<i>Beauveria nivea</i>
环 (L-Pro-L-Val)	赤者色 (<i>Aspergillus ochraceus</i>)
环 (L-Pro-L-Leu)	褐座坚壳菌
环 (D-Val-L-Trp)	薛氏曲霉 (<i>Aspergillus chevalieri</i>)
环 (L-Phe-L-Phe)	变黑青霉 (<i>Penicillium nigricans</i>) 诺尔斯链霉素
环 (Δ Phe- Δ Leu) (白诺斯菌素)	诺尔斯链霉菌
环 (L-Pro-L-Tyr)	链格孢菌 (<i>Alternaria alternata</i>)
环 (L-Pro-L-Trp)	短密青霉 (<i>Penicillium brevicompactum</i>)
环 (L-Ser-L-Ser)	<i>Streptomyces orchidaceus</i>
环 (L-Arg-D-Pro)	假单孢菌属 (<i>Pseudomonas</i> sp.)
苯基组氨酸 roquefortine	娄格法尔特氏青霉 (<i>Penicillium roquefortii</i>)
环 (L-Trp- Δ Aba)	壮观链霉菌 (<i>Streptomyces spectabilis</i>)
环 (4-甲基-D-Pro-L-Nva)	<i>Calyx</i> cf. <i>podatypa</i>
环 (Δ Ala-L-Val)	绿脓杆菌 (<i>Pseudomonas aeruginosa</i>)

关于二酮哌嗪衍生物的生理作用已知得非常少。已描述, 由绿脓杆菌产生的环 (Δ Ala-L-Val) 可能涉及细菌间的信号交流。其它化合物被描述成涉及病原微生物毒力或与铁结合或具有神经生物学特性。

二酮哌嗪衍生物自发现以来就已被证明是有益的, 其中一些具有如, 例如抗细菌、抗真菌、抗病毒、免疫抑制或抗肿瘤活性的生物学特性。

下表 II 给出了一些具有已知生物活性的二酮哌嗪衍生物的例子:

表 II

分子	生物	活性
Ambewelamides A 和 B	Usnea sp.	细胞毒性
阿拉诺丁	Arachniotus aureus	抗病毒
双环霉素	Streptomyces sapporonensis	抗细菌 (抑制转录终止)
环(Δ -Ala-L-Leu)	青霉属菌 (F70614)	抑制 α -葡萄糖苷酶
环(N-甲基-Tyr),	灰色链霉菌 (Streptomyces griseus)	抑制钙蛋白酶
环(Trp- Δ -Aba)	壮观链霉菌 (Streptomyces spectabilis)	抑制谷胱甘肽-S-转移酶
支霉菌素	黄曲霉 (Aspergillus flavus)	除草、抗真菌、抗细菌、抗病毒
Haematocin	Nectria haematococca	抗真菌
Hyalodendrin	不整青霉 (Penicillium turbatum)	抗生素
Mycelianamide	青霉属菌	抗细菌 (抑制丁基胆碱脂酶)
苯基组氨酸	焦曲霉 (Aspergillus ustus) (NSC-F038)	抑制微管多聚化
Tan-1496 A、C 和 E	Microphaeropsis sp. (FL-16144)	抑制拓扑异构酶 抗细菌 (革兰氏阳性)
Verticillin A	粘帚霉菌属 (Gliocladium sp.) (SCF-1168)	抑制 c-fos 原癌基因的诱导
XR334	链霉菌属 (X01/4/100)	抑制 PAI-I

尽管对这些分子已经展开了广泛的研究,但是有关它们的合成却所知甚少。通常已知的是,在细菌和真菌中,这些分子经非核糖体生物合成途径产生。在某些情况中,有可能显示在分子中发生二酮哌嗪

环形成，该分子已由酶经硫酯键预激活，而且对于该分子由于脯氨酸残基的存在而促成了环化反应所必需的肽键顺式构象。在其它情况中，已经证明氨基酸残基的 N-烷化，特别式 N-甲基化，也促成肽键的顺式构象。

因此，截至目前完成的所有这些研究已经证明，构调节前体分子的构象的该前体分子的一级结构是二酮哌嗪环的形成得以发生、以及导致产生最终的二酮哌嗪衍生物的过程的基础。

然而，存在不含脯氨酸残基或 N-烷基化残基的二酮哌嗪衍生物。为对这样的衍生物举例，可以提及白诺斯菌素、或环 (Δ Phe- Δ Leu)、由诺尔斯链霉菌 (*Streptomyces noursei*) 产生的抗生素。已知诺尔斯链霉菌中存在一种酶活性，它催化产生白诺斯菌素的最后步骤，也即 α , β -不饱和残基的形成 (Gondry 等, Eur. J. Biochem., 2001, 268, 1712-1721)。然而，这一酶活性需要环状形式的底物，环 [L-Phe-L-Leu]，该底物不含脯氨酸残基或任何 N-烷基化残基，而且它的合成途径是未知的。

因此，二酮哌嗪衍生物呈现极高的结构多样性和富于变化的生物活性，这使得它们成为发现和开发新的药物的有益分子。

为此，必需能得到大量这样的分子。

诚然，已经描述了二酮哌嗪衍生物的化学合成途径，但是对于大多数复杂衍生物来说，产量很低且该过程并非总被能产业化。

对二酮哌嗪衍生物，尤其是环二肽，的天然合成途径的理解可能会使在生产者生物中进行合理的遗传改良成为可能，并可以开拓通过生产和纯化产量的优化来取代或改良已有合成方法（经化学或生物技术途径）的前景。此外，对涉及二酮哌嗪衍生物生物合成途径的酶的本性和/或特异性的修饰可能创造出具有原始分子结构以及优化的生物特性的新衍生物。

本发明包含在此范围内。

在对白诺斯菌素合成途径的研究中，本发明人证实了一条多核苷酸（下文称作 BamH1 多核苷酸 (SEQ ID NO.5)），该多核苷酸包含四

个开放阅读框，各阅读框各自编码负责在诺尔斯链霉菌和异源宿主例如变青链霉菌中从 L-苯丙氨酸和 L-亮氨酸合成白诺斯菌素以及运输白诺斯菌素各步骤的多肽（参见附图 2 和 3）。

本发明人已能显示：

- 第一个开放阅读框 orf1 (albA, SEQ ID NO.1) 编码一个涉及如在 Gondry 等, (Eur. J. Biochem., 2001, 268, 1712-1721 (α , β -不饱和)) 中所描述的二肽氧化酶 (CDO) 活性的多肽 (AlbA, SEQ ID NO.6) ;

- 第二个开放阅读框 orf2 (albB, SEQ ID NO.2) 编码一个被翻译成 AlbA 多肽活性所需的两个同种型 (AlbB₁, SEQ ID NO.7 和 AlbB₂, SEQ ID NO.8) 的多肽。基本以等量表达的这两个 AlbB 同种型彼此间的区别在于由于使用两个不同的起始密码子而产生的位于 AlbB₁ N-末端的 5 个附加氨基酸。对于 AlbB₂, 去除了起始甲硫氨酸;

- 第三个开放阅读框 orf3 (albC, SEQ ID NO.3) 编码一个与肽合成酶无相似性并能催化两个氨基酸残基缩合以形成环状二肽的多肽 (AlbC, SEQ ID NO.9)。例如在诺尔斯链霉菌中, albC 催化 L-苯丙氨酸与 L-亮氨酸或两个苯丙氨酸的缩合从而形成环状二肽 (L-Phe-L-Leu), 该环状二肽是形成白诺斯菌素和环状二肽环 (L-Phe-L-Phe) 所需的前体。在这种特别情况中, albC 催化既非脯氨酸也非 N-烷基化残基的氨基酸残基的环化, 和

- 第四个开放阅读框 orf4 (albD, SEQ ID NO.4) 编码一个不直接涉及导致从氨基酸形成 α , β -不饱和二酮哌嗪衍生物的反应过程的多肽 (AlbD, SEQ ID NO.10), 但是该多肽可能涉及所述衍生物的运输机制。

本发明人由此表明, 对于 α , β -不饱和二酮哌嗪衍生物的合成, 特别是对于白诺斯菌素在诺尔斯链霉菌中的合成来说, 仅有 albA、albB 和 albC 这三个开放阅读框是绝对必需的。

因此, 本发明的主题是一种分离的天然或合成的多核苷酸, 其特征在于它包含至少分别对应于序列 SEQ ID NO.1、SEQ ID NO.2 和 SEQ

ID NO.3 的 a1bA、a1bB 和 a1bC 这三个开放阅读框。

该多核苷酸编码合成 α , β -不饱和二酮哌嗪衍生物所必需的酶。

根据本发明的有利实施方式, 所述多核苷酸还包含对应于序列 SEQ ID NO.4 的开放阅读框 a1bD。

该多核苷酸编码 α , β -不饱和二酮哌嗪衍生物的合成、运输及分泌所需的酶。

根据本发明的一种特定形式, 该多核苷酸对应于序列 SEQ ID NO.5。该多核苷酸 (BamH1 多核苷酸) 含 a1bA、a1bB、a1bC 和 a1bD 这四个开放阅读框, 并由此编码 α , β -不饱和二酮哌嗪衍生物的合成、运输及分泌所需的酶。

本发明的主题还是一种分离的天然或合成的多核苷酸, 其特征在于它包含分别对应于序列 SEQ ID NO.2、SEQ ID NO.3 和 SEQ ID NO.4 的三个开放阅读框 a1bB、a1bC 和 a1bD 中的至少一个。

因此, 包含开放阅读框 a1bC (SEQ ID NO.3) 的多核苷酸编码一种允许两个可以相同或不同的氨基酸发生环化以形成环状二肽的酶。包含开放阅读框 a1bC 和 a1bD (SEQ ID NO.3 和 SEQ ID NO.4) 的多核苷酸编码一种首先允许两个可以相同或不同的氨基酸发生环化以形成环状二肽、其次允许所述多肽运输的酶。

本发明的主题还是一种对应于序列 SEQ ID NO.2 (a1bB, 318 核苷酸)、SEQ ID NO.3 (a1bC, 720 核苷酸) 或 SEQ ID NO.4 (a1bD, 834 核苷酸) 中任一序列的分离的天然或合成的多核苷酸。

本发明的主题还是如上述定义的多核苷酸的片段。术语“片段”意指任一具有至少 15 个核酸的序列。

根据本发明的多核苷酸可以获自 DNA 文库, 特别是微生物 DNA 文库, 更特别是诺尔斯链霉菌 DNA 文库。本发明的多核苷酸还可通过对诺尔斯链霉菌总 DNA 进行聚合酶链式反应 (PCR) 获得。本发明的多核苷酸可通过对诺尔斯链霉菌总 RNA 进行 RT-PCR 获得。

本发明的主题也是其中插入了上述任一多核苷酸的载体。因此, 本发明的载体可以包含: 含有三个或四个分别相应于序列 SEQ ID

NO. 1、SEQ ID NO. 2、SEQ ID NO. 3 和/或 SEQ ID NO. 4 的开放阅读框 a1bA、a1bB、a1bC 和/或 a1bD 的多核苷酸,含有分别对应于序列 SEQ ID NO. 2、SEQ ID NO. 3 和 SEQ ID NO. 4 的这三个开放阅读框 a1bB、a1bC 或 a1bD 中至少一个的多核苷酸,对应于至少分别对应于序列 SEQ ID NO. 2、SEQ ID NO. 3 和 SEQ ID NO. 4 的这三个开放阅读框 a1bB、a1bC 或 a1bD 中至少一个的任一多核苷酸,对应于序列 SEQ ID NO. 5 的 BamH1 多核苷酸,或所述多核苷酸的片段。

所用载体可以是现有技术中已知的任一载体。作为可以在本发明中使用的载体,值得特别提及的是质粒、粘粒、细菌人工染色体(BAC)、放线菌的整合元件、病毒或噬菌体。

所述载体还可以含有任何对于载体复制和/或由多核苷酸编码的多肽的表达所需的调控序列(启动子、终止位点,等)。

本发明的主题还是至少一种如上述所定义的多核苷酸、或其片段作为在生物中探测对应序列的探针或作为扩增这些序列的引物的用途。

当所述多核苷酸或所述片段用作引物时,它们还包含反义序列。

上述探针或引物的一种优选用途是在其它生物中调查与开放阅读框 a1bA、a1bB、a1bC 或 a1bD 的序列同源的多核苷酸序列,尤其是为了证实二酮哌嗪衍生物的新合成途径。

本发明的主题还是一种分离的天然或合成的多肽,其特征在于它包含分别对应于多肽 A1bB₁、a1bB₂、a1bC 或 A1bD 的序列 SEQ ID NO. 7 至 SEQ ID NO. 10 中的至少一个序列。

本发明的主题还是一种分离的天然或合成的多肽,其特征在于它对应于序列 SEQ ID NO. 7 (A1bB₁)、SEQ ID NO. 8 (A1bB₂)、SEQ ID NO. 9 (A1bC) 或 SEQ ID NO. 10 (A1bD) 中的任一序列。

本发明还涉及由本发明任一多核苷酸,尤其是选自序列 SEQ ID NO. 2 (a1bB)、SEQ ID NO. 3 (a1bC) 或 SEQ ID NO. 4 (a1bD) 的任一多核苷酸,编码的多肽。

有利的是,根据本发明的多肽可以分离自例如微生物(诺尔斯链

霉菌)，或通过化学合成或生物技术途径从本发明的多核苷酸获得，例如，从在正常条件下不表达所述多肽的经修饰的微生物获得。

本发明的主题还是一种分离的多肽，其序列与如上述所定义的序列 SEQ ID NO.7 至 SEQ ID NO.10 中至少一个序列基本同源。

在此，当多肽的氨基酸序列与序列 SEQ ID NO.7 至 SEQ ID NO.10 中至少一个的氨基酸序列表现出 80% 的相似性时，则视该多肽为具有基本同源的序列。

“多肽 P 与序列 SEQ ID NO.7 至 10 间 80% 的相似性”这一表述意指，当将两个多肽进行比对时，P 中 80% 的氨基酸与序列 SEQ ID NO.7 至 10 中对应的氨基酸相同或由同组氨基酸替代。

“同组氨基酸”这一表述意指具有基本相同的化学特性的氨基酸。尤其是，该术语意指具有基本相同的电荷、和/或相同的大小、和/或相同的亲水性或疏水性、和/或相同的芳香结构的氨基酸。

这样的氨基酸组尤其包含：

- (i) 甘氨酸，丙氨酸
- (ii) 异亮氨酸，亮氨酸，缬氨酸
- (iii) 色氨酸，酪氨酸，苯丙氨酸
- (iv) 天冬氨酸，谷氨酸
- (v) 精氨酸，赖氨酸，组氨酸
- (vi) 丝氨酸，苏氨酸。

可以预想其它的取代，其中氨基酸被其它可比的但非天然的氨基酸（羟基脯氨酸、正亮氨酸、鸟氨酸、瓜氨酸、环己氨酸、右旋氨基酸，等）取代。

本发明的主题还是上述本发明的多核苷酸或载体用于合成与序列 SEQ ID NO.7 至 10 对应的多肽的用途。

本发明的主题还包括根据本发明的多肽单独或组合用于（特别是体外用于）制备环二肽和/或 3-及 6-位被 α ， β 不饱和氨基酸侧链取代的二酮哌嗪衍生物（特别是白诺斯菌素）的用途。

本发明的主题还包括本发明的多肽单独或组合用于通过修饰生物

分子的结构来修饰其药理学活性的用途，对生物分子结构的修饰例如使侧链（特别是氨基酸侧链）脱氢或环化（特别是肽分子的环化）。

本发明的主题还是经修饰的生物系统，其中导入了至少一种根据本发明的多核苷酸或至少一个根据本发明的载体。

这样的生物系统可以是任何已知的用原核细胞或真核细胞作为宿主的异源表达系统。例如，可以提到的是微生物，例如大肠杆菌或变青链霉菌这样的细菌，或动物或昆虫细胞。

本发明的主题还是经修饰的体外非细胞系统，其中导入了至少一种根据本发明的多核苷酸或至少一种根据本发明的载体。

本发明的主题还是至少一种根据本发明的多核苷酸和/或至少一种根据本发明的载体用于制备经修饰的生物系统的用途，对所述系统来说它可能是微生物，例如大肠杆菌或变青链霉菌这样的细菌，或任何已知的用原核细胞或真核细胞作为宿主的异源表达系统，或经修饰的体外非细胞系统。

可以用已知的任何方法将根据本发明的多核苷酸和/或载体导入宿主经修饰的生物系统，例如，转染、感染、融合、电穿孔、微注射或其它生物射弹技术。

本发明的主题还包括如上所述的至少一种经修饰的生物系统或至少一种经修饰的体外非细胞系统用于制备环二肽和/或 3-及 6-位被 α , β 不饱和氨基酸侧链取代的二酮哌嗪衍生物（特别是白诺斯菌素）的用途。

该生物系统适于以高产量合成如上述所定义的环二肽 α , β -不饱和和二酮哌嗪衍生物。

当经修饰的生物系统为微生物时，它可选地允许将本发明的二酮哌嗪衍生物分泌至培养基中，从而使其更易于提取和纯化。在如微生物这样的生物系统中存在 A1bD 使由此而分泌到培养基中的衍生物便于提取和纯化，从而构成了对工业过程有利的步骤。

本发明的主题还是一种体外合成环二肽的方法，其特征在于：

(1) 在适宜条件下使两种可以相同或不同的氨基酸与 A1bC

(SEQ ID NO.9) 接触, 和

(2) 纯化所得的环二肽。

本发明的主题还是一种体外合成 3-及 6-位被 α , β 不饱和氨基酸侧链取代的二酮哌嗪衍生物的方法, 其特征在于:

(1) 在适宜条件下使两种可以相同或不同的氨基酸与 A1bC (SEQ ID NO.9) 接触, 和

(2) 使步骤 (1) 中获得的环二肽与 A1bA (SEQ ID NO.6)、A1bB1 (SEQ ID NO.7) 和 A1bB2 (SEQ ID NO.8) 接触, 并纯化所得的 α , β 不饱和二酮哌嗪衍生物。该方法在步骤 (2) 中还包含 A1bD (SEQ ID NO.10)。该方法可以可选地在步骤 (1) 和步骤 (2) 之间包含一个纯化从步骤 (1) 获得的环二肽的附加步骤。

当然, 该方法也可以在一个步骤内完成, 其中在适宜条件下使两种可以相同或不同的氨基酸与 A1bA (SEQ ID NO.6)、a1bB1 (SEQ ID NO.7)、a1bB2 (SEQ ID NO.8) 和 A1bC (SEQ ID NO.9)、可选的 A1bD (SEQ ID NO.10) 进行接触, 并纯化所得的 α , β -不饱和二酮哌嗪衍生物。

术语“适宜条件”优选意指孵育以下物质时所处的条件:

- 多肽 (A1bA、a1bB、a1bC 和/或 A1bD), 其浓度介于 0.1 nM 和 10 μ M 之间, 优选介于 10 nM 和 1 μ M 之间;

- 存在浓度介于 0.1 mM 和 100 mM 之间、优选介于 1 mM 和 10 mM 之间的氨基酸, 所述氨基酸可以相同或不同;

- 于 0.1 M Tris-HCl 缓冲液中, pH 介于 6.8 和 8.0, 温度介于 28 $^{\circ}$ C 和 40 $^{\circ}$ C, 时间介于 2 小时至 48 小时之间。

本发明的主题还是一种合成环二肽的方法, 其特征在于:

(1) 在适宜培养所述已选择的生物系统的条件下接触含有至少多核苷酸 SEQ ID NO.3 (a1bC, 编码 A1bC) 的生物系统, 和

(2) 纯化所得的环二肽。该生物系统还可以包含多核苷酸 SEQ ID NO.4 (编码 A1bD)。

本发明的主题还是一种合成 3-及 6-位被 α , β 不饱和氨基酸侧链取代的二酮哌嗪衍生物的方法, 其特征在于:

(1) 在适宜培养所述已选择的生物系统的条件下接触含有对应于序列 SEQ ID NO.1-3 的多核苷酸 (编码 A1bA、A1bB 和 A1bC) 的生物系统, 和

(2) 纯化所得的 α , β 不饱和二酮哌嗪衍生物。该生物系统还可以包含对应于 SEQ ID NO.4 的多核苷酸 (编码 A1bD)。

将“适于培养所述已选择的生物系统的条件”这一表述理解为意指, 在培养所述已选择的生物系统的条件 (包含含有大大过量的氨基酸的培养基) 下实施该方法。例如, 如果该生物系统是例如大肠杆菌这样的微生物, 那么适宜的条件就是通常用来培养该细菌的那些条件。对于变青链霉菌或当生物系统为真核细胞时也是如此。

根据本发明的方法, 可以相同或不同的氨基酸以介于 0.1 mM 和 100 mM 之间、优选介于 1 mM 和 10 mM 之间的量存在。

类似地, 根据本发明的方法, 多肽 A1bA、a1bB、a1bC 和 a1bD 以介于 0.1 nM 和 10 μ M 之间、优选介于 10 nM 和 1 μ M 之间的量存在。

环二肽和 α , β -不饱和二酮哌嗪衍生物的纯化可以直接从体内或体外合成物中通过本领域技术人员已知的液相萃取技术或沉淀、或薄层或液相层析技术, 特别是反相 HPLC, 或其它适宜纯化肽的方法完成。

可以在适宜的生物系统中, 尤其是在例如微生物 (例如大肠杆菌或变青链霉菌这样的细菌) 这样的宿主、或任何已知的用原核细胞和真核细胞作为宿主的异源表达系统、或甚至在体外非细胞系统中实施本发明的方法。

除了上述方案, 本发明还包含其它可以从以下描述中显现的方案, “以下描述”是指实施本发明的实施例以及附图, 其中:

-附图 1 代表二酮哌嗪环 (A) 和白诺斯菌素 (B) 的化学结构。

-附图 2 代表诺尔斯链霉菌的包含白诺斯菌素合成所需基因簇的基因组区域的图表, 其中 orf1 对应于 a1bA, orf2 对应于 a1bB, orf3 对应于 a1bC, 以及 orf4 对应于 a1bD。

-附图 3 表示假设的诺尔斯链霉菌白诺斯菌素合成途径。

-附图 4 表示所制备的用于向大肠杆菌或变青链霉菌中导入各种开放阅读框（或 orf）的特定质粒构建体。

-附图 5 表示对用质粒 pSL128 (A)、pUWL201 (B) 和 pSL129 (C) 转化的变青链霉菌的培养基的分析结果。

-附图 6 表示对用质粒 pSL168 和 pSL159 转化的变青链霉菌、或未经转化的变青链霉菌的培养基的分析结果：

A: 变青链霉菌 [pSL168]，在 CD0 存在下培养；

B: 变青链霉菌 [pSL168]，在 CD0 不存在下培养；

C: 变青链霉菌 [pSL159]，不存在或存在 CD0；

D: 在 CD0 存在下培养的变青链霉菌 TK21。

-附图 7 表示对在大肠杆菌 BL21 (DE3) -plysS 中从多核苷酸 albA-albB 表达 AlbA 和 AlbB₁ 和 AlbB₂ 的十二烷基硫酸钠/聚丙烯酰胺凝胶 (SDS-PAGE) 分析。用考马斯亮蓝 R-250 染色。泳道 S: 分子量标准，泳道 1: 全细胞质提取物 (约 10 μ g)，泳道 2: 用 Ni-琼脂糖凝胶柱纯化的酶级分 (约 10 μ g)。

以下实施例对本发明进行举例说明但不限制本发明。

实施例 1: 从诺尔斯链霉菌中分离本发明的多核苷酸

用以 PCR 基因扩增为基础的方法从诺尔斯链霉菌全基因组中分离出在诺尔斯链霉菌中生物合成白诺斯菌素所需的含有全部遗传信息的多核苷酸。

- 获得环二肽氧化酶的部分肽序列：

根据 M. Gondry 等 (Gondry 等, 2001, 上述) 中描述的方案，通过用 Edman 法对纯化的环二肽氧化酶经胰蛋白酶水解所得的多肽进行直接测序得到在诺尔斯链霉菌中催化环二肽环 (L-Phe-L-Leu) 向白诺斯菌素转化的酶，即环二肽氧化酶 (CD0)。经 15% 聚丙烯酰胺凝胶电泳和蛋白质的考马斯亮蓝染色分离出酶级分组成后，切出一条含有分子量约为 21000 道尔顿的蛋白质的凝胶带，并将其在 1 ml 50 mM Tris-HCl 缓冲液，pH 8，胰蛋白酶 (相对胰蛋白酶/底物浓度=1/50)

存在下于 37℃ 孵育 20 小时。之后用反相高效液相层析 (HPLC) (μ RPC C₂/C₁₈ SC21/10 柱, Pharmacia) 以 0% 至 76% 乙腈的线性梯度在 62 分钟内对所得多肽进行分离 (溶剂: 0.1% 三氟乙酸; 流速: 1 ml/分钟)。之后在已用含 30% 乙腈和 0.1% 三氟乙酸的缓冲液平衡的 superdex 肽 PC32/30 柱 (Pharmacia) 上进行凝胶渗透层析以纯化各分离的肽。最后经 Edman 法自动测序 (型号 477A 测序仪, Applied Biosystems) 和 MALDI-TOF 质谱法分析所得的三种多肽。表 1 给出了所得的肽序列以及由其推导出的核苷酸序列。

表 I:

肽序列 推导出的核苷酸序列

<u>EPVDDALIEQLLEAMLAPT</u> (SEQ ID N°11)	GARCCSGTSGACGACGC (oligo1f) (SEQ ID N°14)
	GCGTCGTCSACSGGYTC (oligo1r) (SEQ ID N°15)
<u>NEVVNYEXWGNR</u> (SEQ ID N°12)	AACGARGTSGTSAACTACGA (oligo2f) (SEQ ID N°16)
	TCGTAGTTSACSACYTCGTT (oligo2r) (SEQ ID N°17)
<u>QAXSFMVVR</u> (SEQ ID N°13)	CAGGCSTGGWSSTTCATGGT (oligo3f) (SEQ ID N°18)
	ACCATGAASSWCCASGCCTG (oligo3r) (SEQ ID N°19)

X: 未确定的氨基酸 (R = A 或 G; S = C 或 G; Y = C 或 T 和 W = A 或 T)。

对对应于序列 SEQ ID NO. 13 的多肽的试验和理论质量进行比较使得有可能鉴定 3-位的色氨酸残基。根据链霉菌中典型的密码子使用, 下划线和粗体的序列用于表示 6 条有义 (1f、2f 和 3f) 和反义 (1r、2r 和 3r) 简并寡核苷酸。由于缺少关于这些多肽各自在蛋白质序列中的位置的信息, 将这六个寡核苷酸组合用于克隆。由于所测试的 PCR 条件导致大量的扩增核苷酸片段产生, 因此开发了反转录 (RT-PCR) 策略。

- 由 RT-PCR 扩增寡核苷酸片段:

从根据 Kieser 等 (Practical Streptomyces Genetics. The John

Innes Foundation Norwich, U.K. (2000)) 描述的方法在培养基 5 (ATCC 培养基) 中的培养 24 小时的培养物中提取诺尔斯链霉菌的总 RNA。用 DNase I 进行附加处理以将 DNA 完全去除。由 Sigma Genosys Ltd 合成简并的寡核苷酸(有义和反义), 以及根据标准说明用 Titan™ One Tube RT-PCR 试剂盒 (Boehringer Mannheim) 在各反应中使用 1 μ g 总 RNA 进行 RT-PCR。反转录在 50℃ 下进行 30 分钟, 之后的 PCR 条件如下: 初始 97℃ 变性 4 分钟, 之后 45 循环的 95℃ 1 分钟、50℃ 1 分钟和 68℃ 1 分钟, 和最后于 68℃ 进行 10 分钟聚合反应。反应产物经 1.5% 琼脂糖凝胶电泳分析, 之后纯化 (DNA 和 Gel Band 纯化试剂盒, Pharmacia)。

用六个寡核苷酸组合进行的 RT-PCR 扩增出一个带有寡核苷酸 3f 和 2r 的约 400 碱基对的片段。将该片段克隆进载体 pGEM-T 便利载体并进行测序, 从而有可能确认在同一阅读框中真正使用的寡核苷酸引物。将该核苷酸片段用作筛选在粘粒 pWED1 中制备的诺尔斯链霉菌基因组 DNA 文库的探针。

- 构建诺尔斯链霉菌基因组 DNA 文库:

用 0.33 U 的 BamHI 对以标准程序 (Kieser 等, 前述; J. Sambrook 等, Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, N.Y. (1989)) 从诺尔斯链霉菌提取的基因组 DNA (2.5 μ g) 进行部分消化, 产生一个约 35 至 45 kb 的 DNA 片段。通过连接将这些片段导入预先用 BamHI 消化并脱磷酸化的粘粒载体 pWED1。在体外将连接产物包装进入噬菌体 (Packagene Lambda DNA 包装系统, Promega), 并经转染导入大肠杆菌 (SURE) 菌株。

- 筛选诺尔斯链霉菌基因组 DNA 文库:

之后, 由 RT-PCR 扩增获得的核苷酸片段用 T7 Quick Prime 试剂盒 (Pharmacia) 通过随机引发以 [α -³²P]-dCTP 进行标记, 并用作筛选文库的探针。根据标准方法用菌落杂交测试了约 2000 个克隆 (J. Sambrook 等, 前述), 并选择出 12 个克隆。提取对应的粘粒 (指 pSL110

至 pSL121)，用 BamHI 消化并以 RT-PCR 片段为探针作 DNA 印迹技术分析。该探针使得有可能分离出对所有粘粒来说都相同、且存在于用 BamHI 消化的诺尔斯链霉菌基因组 DNA 中的约 3.8 kb 的核苷酸片段。从粘粒 pSL117 中分离该片段，并将其克隆进载体 pBC SK⁺，从而得到载体 pSL122（所用载体的定义：参见表 II）。

实施例 2：对本发明多核苷酸序列的分析

在 ABI PRISM Genetic Analyzer (Perkin Elmer) 上用 DYEnamic ET 终止子循环试剂盒 (Pharmacia) 或由公司 Genome Express 完成对本发明多核苷酸的自动测序。用 Frame (Bibb, M. J. 等, Gene, 30, 157-166(1984))、BLAST 和 FASTA(Altschul, S. F. 等, Nucleic Acids Res., 25, 3389-3402(1997); Pearson, W. R., Methods in Enzymology, 183, 63-98(1990)) 程序对序列进行计算机分析并与数据库进行比较。

用 FRAME 程序对 BamHI 多核苷酸 (SEQ ID NO.5) 所作的分析揭示了四个完整的开放阅读框，即以相同方向转录的 orf1 至 orf4 (albA 至 albD, SEQ ID No.1 至 4)，和一个末端被截断的开放阅读框 (1119 bp)，即 orf5 (参见附图 2)。orf5 的翻译产物与源自蓝色链霉菌 (*Streptomyces coelicolor*) 的 NADP-特异性谷氨酸脱氢酶 N 末端部分表现出高度相似性 (根据 BLAST 程序为 78% 同一性和 86% 相似性)。

第一个开放阅读框，orf1 (albA, SEQ ID NO.1)，包含由 RT-PCR 扩增的片段的核苷酸序列，而且推导的肽序列确实包含最初分离的 3 个胰蛋白酶肽。因此，orf1 产物对应于分离纯化自诺尔斯链霉菌 (M. Gondry 等, 前述) 的、质量约为 21 千道尔顿的酶蛋白。因此，该基于无疑涉及白诺斯菌素的生物合成，且将被称为 albA。

对 albA (orf1) 序列的分析表明，3 个密码子，两个 GUG 和一个 AUG，可被视为 albA 翻译的起始密码子，它们产生 219、204 或 196 个氨基酸的蛋白质。由于存在对末端的转录后修饰而未能成功确定 N-末端的肽序列，因此为 albA (SEQ ID NO.6) 选择了最长的序列 (657 核苷酸) (第一个起始密码子位于距 BamHI 片段末端 20 个核苷酸的位置)。

置, 因此该 BamHI 片段不包含必需位于该基因更上游的启动子区)。

从 *albA* 推导的肽序列 (*AlbA*, SEQ ID NO.6) 与数据库进行比较显示了与来自 *Archaeoglobus fulgidus* 的 NADH 氧化酶最大程度的相似性 (根据 BLAST 程序为 32% 同一性和 52% 相似性), 对保守结构域的查询显示它具有一个大的硝基还原酶类型结构域 (pfam00881, 151 个氨基酸)。

与 *albA* 相邻但阅读框不同的 *orf2* (*albB*, SEQ ID NO.2) 也表现出典型的链霉菌密码子使用。根据起始密码子 AUG 或 GUG, 考虑了 *orf2*, *albB* 翻译成 *AlbA* 多肽活性所需的两个同种型 (*AlbB*₁, SEQ ID NO.7 和 *AlbB*₂, SEQ ID NO.8)。基本以等量表达的这两个 *AlbB* 同种型彼此间的区别在于由于使用两个不同的起始密码子而产生的位于 *AlbB*₁ N-末端的 5 个附加氨基酸。对于 *AlbB*₁, 去除了起始甲硫氨酸。

这两种可能性与用 FRAME 程序对序列进行分析是相容的。BLAST 和 FASTA 程序未显示由 *orf2* 推导的肽序列与数据库中的蛋白质有特别的同源性。

类似地, 以分别推导自 *orf3* (*albC*, SEQ ID NO.3) 和 *orf4* (*albD*, SEQ ID NO.4) 的多肽序列为基础进行数据库查询未显示与功能已知的蛋白间的显著同源性。*orf3* 以 ATG 起始密码子开始并编码 239 个氨基酸的多肽 *albC* (SEQ ID NO.9), 它与两种功能未知的假定蛋白质表现出具有低相似性: 来自结核杆菌 (*Mycobacterium tuberculosis*) 的 Rv2275 (根据 BLAST 程序为 34% 同一性和 53% 相似性) 和来自枯草芽孢杆菌 (*Bacillus subtilis*) 的 YvmC (根据 BLAST 程序为 29% 同一性和 46% 相似性)。*Orf4* 编码 277 个氨基酸的蛋白质 *AlbE* (SEQ ID NO.10), 它含有一个如同用 TMHMM 程序 (Krogh, A. 等, J. Mol. Biol. 305, (2001)) 对其进行序列分析所示的跨膜结构域, 并与蓝色链霉菌的一个功能未知的跨膜蛋白表现出弱同源性 (根据 BLAST 程序为 54% 同一性和 67% 相似性)。

实施例 3: 多核苷酸 *albA*、*albB*、*albC* 和 *albD* 的克隆和表达载体的构建

根据 Sambrook 等（前述）和 Kieser 等（前述）描述的标准程序完成提取和制备 DNA、转化大肠杆菌和变青链霉菌 TK21 菌株、以及制备原生质体的方法。

表 II 给出了所制备的用于操作作为本发明主题的多核苷酸的所有质粒和粘粒载体（也参见附图 4）。

表 II：所述的菌株和载体

细菌	特性	来源/参考
大肠杆菌 DH52	用于克隆的标准菌株	Invitrogen
大肠杆菌 SURE	用于粘粒文库的菌株	Stratagene
变青链霉菌 TK21	用于克隆 alb 基因的链霉菌菌株	Hopwood, D. A. 等, J. Gen. Microbol. 129 , 2257-2269 (1983)。
诺尔斯链霉菌 ATCC1145	产白诺斯菌素的野生型菌株	ATCC
载体		
pGEM-T easy	用于克隆 PCR 产物的载体, Amp ^r	Promega
pWED1	衍生自 pWED15 的粘粒, Amp ^r	Gourmelen, A. 等, Antimicrobial Agents Chemotherapy, 42, 2612-2619 (1998)
pBC SK ⁺	克隆载体, Cm ^r	Stratagene
pHP45 Ωacc	用于获得 Ωacc 盒的质粒, Amp ^r , Apr ^r	Blondelet-Rouaul t, M. H. 等, Gene, 190 , 315-317 (1997)
pUWL201	大肠杆菌/链霉菌穿梭载体, 含有在链霉菌中表达克隆基因的 ErmeE*启动子, Amp ^r , Thio ^r	Doumith, M., 等, Mol. Gen. Genet., 264 , 477-485 (2000)
pET-28a	表达载体	Novagen
pSL117	来自诺尔斯链霉菌文库的粘粒, 用于 RT-PCR 产物	
pSL122	克隆进 pBC SK ⁺ 的 3.8kb BamHI 片段	
pSL127	pSL122 衍生物, 含有 ApaI 片段的一个内部缺失, 去除了 orf2、orf3、orf4 和 orf5	
pSL128	克隆进 pUWL201 中的 pSL122 的 3.8kb BamHI 片段, 受 Erme*p 控制	
pSL129	克隆进 pUWL201 中的 pSL122 的	

	3.8kb BamHI 片段, 含有与 pSL128 相反反向的插入	
pSL138	pSL122 衍生物, 含有 EcoRI 片段的一个内部缺失, 去除了 orf3、orf4 和 orf5, 并带有 Ω acc 盒插入	
pSL140	pSL122 衍生物, 含有 NdeI/EcoRV 片段的一个内部缺失, 去除了 orf5, 并带有 Ω acc 盒插入	
pSL142	克隆进 pUWL201 的, pSL138 的 Asp718/Klenow/BamHI 片段(含有 orf1、orf2 和 Ω acc 盒)	
pSL144	克隆进 pUWL201 的, pSL140 的 Asp718/Klenow/BamHI 片段(含有 orf1 至 orf4 和 Ω acc 盒)	
pSL145	pSL122 衍生物, 含有 EcoRI 片段的一个内部缺失, 去除了 orf3、orf4 和 orf5	
pSL150	克隆进表达载体 pET-28a 的 orf1 至 orf2 的 PCR 扩增产物	
pSL157	克隆进表达载体 pGEM-T easy 的, orf4 的 PCR 扩增产物	
pSL159	克隆进 pUWL201 的, pSL157 的 Pst1/Klenow/BamHI 片段	
pSL165	克隆进 pGEM-T easy 的, 的 orf3 PCR 扩增产物	
pSL166	克隆进 pGEM-T easy 的, 的 orf3+orf4 PCR 扩增产物	
pSL167	克隆进 pUWL201 的, pSL166 的 Pst1/Klenow/BamHI 片段	
pSL168	克隆进 pUWL201 的, pSL165 的 Pst1/Klenow/BamHI 片段	

pSL117 是含诺尔斯链霉菌基因组 DNA 文库的粘粒。

pSL122 包含被克隆进克隆载体 pBC SK⁺的、源自诺尔斯链霉菌的 3.8kb BamHI 多核苷酸 (SEQ ID NO.5), 该多核苷酸是本发明一个主题。经 ApaI 或 EcoRI 消化 pSL122 和再连接分别构建了 pSL127 和 pSL145。

BamHI 多核苷酸还被以适于将所有基因置于 ermE*启动子控制下的方向 (pSL128) 或以相反方向 (pSL129) 克隆进了大肠杆菌/链霉菌

穿梭载体 pUWL201。

两步构建了 pSL142 和 pSL144: 先用 EcoRI 和 Klenow 酶、或 NdeI 和 Klenow 酶消化 pSL122, 之后将这些片段与已用 HindIII-Klenow 消化的 Ω aac 盒连接从而产生质粒 pSL138 和 pSL140。之后将这些质粒用 Asp718、Klenow 和 BamHI 消化, 首先将所得的包含 orf1 (albA, SEQ ID NO.1)、orf2 (albB, SEQ ID NO.2) 和 Ω aac 盒的片段, 以及其次将 orf1 至 orf4 (albA 至 albD) 和 Ω aac 盒克隆进已用 XbaI-Klenow-BamHI 消化过的载体 pUWL201。

用下列引物 PCR 扩增 orf3 (albC, SEQ ID NO.3)、orf4 (albD, SEQ ID NO.4) 和 (orf3+orf4):

对 orf3

sylv24 : (SEQ ID N°20):

5'-CGGCTGCAGGAGAAGGGAGCGGACATATGCTTGCAGGCTTAGTTCCC
-3', (PstI 位点有下划线);

sylv22 : (SEQ ID N°21):

5'-CGGTCCCGTGGATCCAAGCTTCTAGGCCGCGTCGGCCAGCTC-3',
(BamHI 位点有下划线);

对 orf4

sylv19 : (SEQ ID N°22):

5'-GAGCGGGATCCTGCAGTGTTCATGGGGAGGACAGGAC-3',
(PstI 位点有下划线);

sylv18 : (SEQ ID N°23):

5'-CGATCACGTGGATCCAAGCTTGCCAATCCTGTACGCGATTT-3',
(BamHI 位点有下划线);

对 (orf3+orf4): sylv 24 和 sylv 18。

仅由 37 个核苷酸分离 orf2 和 orf3, sylv24 中包含了一个合成的核糖体结合位点以确保 orf3 的正确翻译。之后将 PCR 扩增的片段克隆进载体 pGEM-T easy (Promega), 以形成 pSL165、pSL157 和 pSL166。之后将从这三个质粒得到的 PstI-BamHI 片段克隆进已用 PstI-BamHI 消化的载体 pUWL201, 以形成 pSL168、pSL159 和 pSL167。

实施例 4: 在导入了 BamHI 多核苷酸 (SEQ ID NO.5) 的转化异源宿主变青链霉菌中生产二酮哌嗪衍生物环 (Δ Phe- Δ Leu) (白诺斯菌素)

根据 Sambrook 等 (前述) 和 Kieser 等 (前述) 描述的标准过程, 用 pSL128 (含有 BamHI 多核苷酸 (SEQ ID NO.5))、pSL129 或 pUWL201 (对照) 转化变青链霉菌 TK21 原生质体。在 28℃ 至 30℃ 间的温度下, 将这三个菌株在 M5 培养基 (ATCC 培养基) 中培养三天, 该 M5 培养基是一种含有大大过量氨基酸的丰富培养基。在下列条件下用反相 HPLC 分析这三个转化菌株的培养物上清液: 过滤培养物上清液 (500 μ l) (ultrafree-MC 10 kDa, Millipore) 并直接注入 HPLC (Vydac C18 柱 (4.6 $\times$ 250 mm): 流速: 1 ml/分钟; 洗脱: 在 0.1% 三氟乙酸中 0 至 45% 乙腈的线性梯度, 45 分钟内)。用多波长探测器监测 200 和 600 nm 间的洗脱。

在变青链霉菌 [pSL128] 中检测到了两种立体异构形式的白诺斯菌素的生产 (出现在 38.3 分钟和 40.5 分钟的两个峰; $\lambda_{\max}$ = 318 nm; m = 256.4 Da) (附图 5A)。

在对照菌株变青链霉菌 [pUWL201] 和变青链霉菌 [pSL129] (附图 5B 和 5C) 中, 未检测到白诺斯菌素的生产。

来自诺尔斯链霉菌的 BamHI 多核苷酸 (SEQ ID NO.5) 包含白诺斯菌素生产的遗传信息。

实施例 5: 用经多核苷酸 albA 和 albB 插入转化的异源宿主, 大肠杆菌, 将环 (L-Trp-L-Trp) 向环 (Δ Trp- Δ Trp) 的转化在培养皿中可视化, 以此证明多核苷酸 albA 和 albB 的功能。

为了直接在分离的克隆上探测由环二肽向双脱氢环二肽的转化, 开发了一种在培养皿中的快速测试。该测试是以在溶液中无色的环二肽环 (L-Trp-L-Trp) 向黄色的不溶产物环 (Δ Trp- Δ Trp) ($\lambda_{\max}$ = 367 nm 和 450 nm) 的转化为基础的, 它使得表现 CDO 活性的克隆呈现亮黄色。

直接在含有 0.5 mM 环 {L-Trp-L-Trp} 的 LB 培养基的平皿中测试已用根据实施例 3 详细说出的程序制备的质粒转化的大肠杆菌。

37℃ 下培养 16 小时后, 大肠杆菌 [pSL122] (包含 BamHI 多核苷酸) 和大肠杆菌 [pSL145] (包含带有 orf3 至 orf5 缺失的 BamHI 片段) 菌

株表现强黄色染色，而大肠杆菌 [pBC SK⁺]（包含完整的克隆载体）和大肠杆菌 [pSL127]（包含带有 orf2 至 orf5 缺失的 BamHI 片段）不被染色。

结果证实，orf1 (albA, SEQ ID NO.1) 和 orf2 (albB, SEQ ID NO.2) 这两个基因涉及与白诺斯菌素生产相关的环二肽氧化酶活性。

实施例 6: 环二肽氧化酶 (CDO) 酶活性在导入了 orf1 和 orf2 (albA 和 albB, SEQ ID NO.1 和 SEQ ID NO.2) 的转化异源宿主变青链霉菌中的表达。

在实施例 4 中所述的条件下培养经含 BamHI 多核苷酸的质粒 pSL142 转化的变青链霉菌 TK21 菌株，该 BamHI 多核苷酸中删除了 orf3 (albC, SEQ ID NO.3)、orf4 (albD, SEQ ID NO.4) 和 orf5，对该菌株上清液所作的 HPLC 分析证实未出现白诺斯菌素的生产，并由此表明 orf3 和/或 orf4 在二酮哌嗪衍生物生产中的相关性。

然而，在 Gondry 等（前述）描述的标准条件下向上清液中加入环 (L-Phe-L-Leu)，明显导致白诺斯菌素的生产。这暗示着 orf3 和/或 orf4 在环二肽环 (L-Phe-L-Leu) 生物合成中的相关性。

实施例 7: 证明环 (L-Phe-L-Leu) 和环 (L-Phe-L-Phe) 在经 orf3 (albC, SEQ ID NO.3) 的导入而转化的异源宿主变青链霉菌中的体内生产。

为确认实施例 6 中所描述的结果，将 orf3 (albC) 和 orf4 (albD) 分别克隆进穿梭质粒 pUWL201 而分别产生 pSL168 (orf3) 和 pSL159 (orf4)，根据 Sambrook 等（前述）和 Kieser 等（前述）描述的标准方法将它们导入变青链霉菌 TK21。

变青链霉菌 [pSL168] 和变青链霉菌 [pSL159] 在实施例 4 中所述的条件下培养后，在实施例 4 中所述的标准条件下对其培养物上清液作 HPLC 分析，以证实环二肽环 (L-Phe-L-Leu) 的生产。由于该化合物的低摩尔吸收系数 (254 nm 处 $\epsilon_{\text{mol}} \approx 100 \text{ M}^{-1}\text{cm}^{-1}$) 和培养基的复杂性使得难以直接检测该化合物，因此通过加入已根据 Gondry 等（前述）描述的方法纯化的 CDO 酶在将环二肽转化为白诺斯菌素（环

($\Delta\text{Phe}-\Delta\text{Leu}$) , 318 nm 处 $\varepsilon_{\text{mol}} = 25120 \text{ M}^{-1} \cdot \text{cm}^{-1}$) 后进行证实。

过滤培养物上清, 之后与 4.1×10^{-3} 酶单位的纯化 CD0 于 30℃ 孵育 10 至 15 小时。比较与 CD0 或未与 CD0 孵育的培养物上清液的 HPLC 分析。用质谱法 (Quattro II, Micromass) 测定所产生的代谢物的分子量。

结果示于附图 6:

在与 CD0 孵育的变青链霉菌 [pSL168] 培养物上清液中检测到, 白诺斯菌素 (环 ($\Delta\text{Phe}-\Delta\text{Leu}$)) (峰值出现在 40.5 分钟; $\lambda_{\text{max}} = 318 \text{ nm}$; $m = 256.4 \text{ Da}$) 和环 ($\Delta\text{Phe}-\Delta\text{Phe}$) (峰值出现在 44.1 分钟; $\lambda_{\text{max}} = 338 \text{ nm}$; $m = 290.3$) (图版 A);

未检测到代谢物的:

- 不含 CD0 时的变青链霉菌 [pSL168] 培养物上清液 (图版 B);
- 含或不含 CD0 的变青链霉菌 [pSL159] 培养物上清液 (图版 C);
- 与 CD0 孵育的变青链霉菌 TK21 的培养物上清液 (图版 D)。

该结果清楚地证实了 orf3 (albC) 涉及白诺斯菌素的前体分子环二肽环 (L-Phe-L-Leu)、和第二个代谢产物环 ($\Delta\text{Phe}-\Delta\text{Phe}$) 的前体分子环 (L-Phe-L-Phe) 最初一同在诺尔斯链霉菌中生产 (Khokhlov A. S. 等, Tetrahedron Lett., 27, 1881 (1963))。

实施例 8: 证明环 (L-Phe-L-Leu) 和环 (L-Phe-L-Phe) 在经多核苷酸 orf3-orf4 (albC-albD, SEQ ID NO.3-SEQ ID NO.4) 的导入而转化的异源宿主变青链霉菌中的体内生产。

根据对实施例 7 的修改, 将多核苷酸 orf3-orf4 (albC-albD) 克隆进穿梭质粒 pUWL201, 并根据 Sambrook 等 (前述) 和 Kieser 等 (前述) 描述的标准方法将所得的质粒 pSL167 导入变青链霉菌 TK21。

对经与实施例 7 所述相同的方式处理的培养物上清液所作的 HPLC 分析表明, 在存在 CD0 时培养的变青链霉菌 [pSL167] 的上清液中检测到了白诺斯菌素 (环 ($\Delta\text{Phe}-\Delta\text{Leu}$)) 和环 ($\Delta\text{Phe}-\Delta\text{Phe}$), 在没有 CD0 的变青链霉菌 [pSL167] 的培养物上清液中未检测到代谢物, 在存在 CD0 时培养的变青链霉菌 TK21 的培养物上清液中也没有检测到代谢

物。

这些结果证实，在变青链霉菌中如同在诺尔斯链霉菌中，orf5基因的产物不直接参与白诺斯菌素的合成，而orf4(albC)对于生产前体环二肽环(L-Phe-L-Leu)和环(L-Phe-L-Phe)时必需且充分的。

实施例 9: AlbA 和 AlbB 从多核苷酸 albA-albB (SEQ ID NO.1-SEQ ID NO.2) 的过表达

用含有 N-末端和 C-末端聚组氨酸序列(His 标签)的载体 pET-28a (+)构建一个含有多核苷酸 albA-albB 的表达载体，以便于纯化重组蛋白。选择最短的 albA 序列，它编码 196 个氨基酸残基的多肽。用设计成含有 NdeI (有义)和 XhoI (反义)克隆位点的寡核苷酸引物 PCR 扩增该多核苷酸。

PCR 条件如下：94℃初始变性 4 分钟，之后进行 10 个循环：94℃ 1 分钟、45℃ 1 分钟和 72℃ 1.5 分钟，之后 20 个循环的 94℃ 1 分钟、50℃ 1 分钟和 72℃1.5 分钟，以及最后在 72℃下进行 10 分钟的聚合反应。用 NdeI 和 XhoI 消化反应产物，并将片段亚克隆进载体 pET-28a 以得到 pSL150。用 DYEnamic ET 终止子循环试剂盒(Amersham Pharmacia Biotech)经自动测序 (ABI PRISM Genetic analyzer, Perkin Elmer) 核实插入序列。

使用了标准表达条件：

- 生长温度：20℃，
- 表达的诱导：0.6 mM IPTG，
- 诱导时间：16 小时。

将 pSL150 导入大肠杆菌 BL21 (DE3) -plysS。菌株于 20℃培养于 LB 培养基直至吸收为 0.6，之后诱导表达。之后于 4℃在 4190 g 下离心培养物 15 分钟。之后将细胞重悬于抽提缓冲液 (100 mM Tris-HCl, pH 8.0, 1 μM 磷酸二肽, 1 mM PMSF 和 5%丙三醇)，并用 Eaton press 研磨。30℃下在 benzonase (25 U/ml) 存在下孵育蛋白质提取物 10 分钟，之后于 4℃在 11300 g 离心 15 分钟。根据 Gondry 等 (前述) 描述的标准检测方法测定酶活性。将一个单位的酶活性定

义为，每分钟催化产生 1 μmol 产物的酶量，且特异酶活性以每毫克蛋白质的酶单位表示。亲和层析（柱：螯合了 HP（1 ml）的 HiTrap, Amersham Pharmacia Biotech; Ni^{2+} 离子平衡于含有 0.5 M NaCl 和 10 mM 咪唑、pH 为 8.0 的 100 mM Tris-HCl 缓冲液中；洗脱：35 分钟内 在 0.3 至 1 M 咪唑梯度中；流速：1 ml/min）纯化后，酶提取物的特异酶活性提高了 50 倍（ $A_s = 2 \text{ U/mg}$ ）。

对大肠杆菌 [pSL150] 的纯化级分所作的十二烷基硫酸钠/聚丙烯酰胺凝胶（SDS-PAGE）（12%）分析证实 AlbA 和 AlbB 以非化学计量比例同时出现，并证实 AlbB 表达为两个同种型（SDS-PAGE 中的两个条带）。用质谱法对纯化的 AlbB 所作的分析和对多肽序列的 N-末端测序显示，这两个同种型对应于源自其序列中鉴定的两个起始密码子的产物 AlbB₁ 和 AlbB₂（参见上述）。

实施例 10：由重组 AlbA-AlbB 对环（L-Phe-L-His）和环（L-Phe-L-Leu）的体外转化

在 Gondry 等（前述）描述的标准条件下孵育根据实施例 9 中描述的方法纯化的酶制品，催化环肽向单脱氢和双脱氢环二肽的转化。

于 30°C 在环（L-Phe-L-His）底物存在下孵育纯化的酶制品 72 小时。反应产物在实施例 4 中描述的条件下作反相 HPLC 分析，并根据它们由质谱分析确定的光谱特征和分子量进行鉴定。反应产物如下：

-源自环（L-Phe-L-His）：环（ $\Delta\text{Phe-L-His}$ ）（ $\lambda_{\text{max}} = 297 \text{ nm}$ 和 $m = 282 \text{ Da}$ ）和环（ $\Delta\text{Phe-}\Delta\text{-His}$ ）（ $\lambda_{\text{max}} = 338 \text{ nm}$ 和 $m = 280 \text{ Da}$ ），

-源自环（L-Phe-L-Leu）：环（ $\Delta\text{Phe-L-Leu}$ ）（ $\lambda_{\text{max}} = 297 \text{ nm}$ 和 $m = 258 \text{ Da}$ ）和环（ $\Delta\text{Phe-}\Delta\text{-Leu}$ ）（ $\lambda_{\text{max}} = 316 \text{ nm}$ 和 $m = 256 \text{ Da}$ ）。

这些结果证实，由将多核苷酸 albA-albB 克隆进表达载体、并将该表达载体导入异源宿主而得到的酶制品在体外催化环二肽向 α, β -脱氢的二酮哌嗪衍生物的转化。

<110> Commissariat à l'Energie Atomique
 Centre National de la Recherche Scientifique
 GONDRY Muriel
 GENET Roger
 LAUTRU Sylvie
 PERNODET Jean-Luc

<120> 涉及二酮哌嗪衍生物合成的多核苷酸和由所述多核苷酸编码的多肽

<130> CGA263/83FR

<140>

<141>

<160> 23

<170> PatentIn Ver. 2.1

<210> 1

<211> 657

<212> ADN

<213> 诺尔斯链酶菌

<400> 1

```
gtgaggcgcc acccatcgca ttcgccgtac cgcgggcggt gtgagggtgcg cccaaaaaga 60
aggggattga tgttagctca cagttcatct gaatcgccgc cggaatcctt gccggacgcg 120
tggacggtec tcaaaaccgc taccgccgtc cgcaattacg cgaaagagcc ggtcgacgac 180
gcgctgatcg agcagctggt ggaggccatg ctcgccgcgc cgaccgcctc caaccggcag 240
gcgtggtcgt tcatggtggt ggcagggccc gccgcggtcc gccggctgcg cgcgttctcg 300
cccggggtgc tgggaacccc cgccttcttc gtcgtggcct gcgtcgaccg cagtctgacc 360
gacaacctct ccccgaagct ctgcgagaag atctacgaca ccagcaagct ctgtgtcgcc 420
atggcggtgg agaacctgct gctcgcggcg cacgcggcgc gcctgggcgg atgcccggtg 480
ggcagcttca ggtccgacat cgtcaccagc atgctcggtg tcccgaaca catcgagccg 540
atgctcgtgg tcccgatcgg ccgtcccgcg acagccctcg tcccctccca gcgccgcgcc 600
aagaatgagg tcgtcaacta tgaatccttg ggaaaccgtg ctgccgcccc aactgcg 657
```

<210> 2

<211> 318

<212> ADN

<213> 诺尔斯链酶菌

<400> 2

```
atgaatcctg gggaaaccgt gctgccgccc caactgcgtg aggagatcgc gtcctcgcgc 60
gtctatctgc tcagcagcgg ccgcggactc ctggaggagc cggccgacta cggaatttac 120
cgctgtaccg acggggcccg tcgggcgctc caactcctcg acgaacacgg cgggagcacg 180
gcacggctga ccgccgtccg cgagcgtctc gacgaggtca tggtcgcgcc gatgggcgag 240
gaccgggaca tgggcgcgat tctggacgac ctgtgtcgcc aaatggcaga cgctcttcgc 300
gaaattgaaa ccccctga
```

318

<210> 3

<211> 720

<212> ADN

<213> 诺尔斯链酶菌

<400> 3

```

atgcttgccag gcttagttcc cgcgccggac cacggaatgc gggaagaaat acttggcgac 60
cgcagccgat tgatccggca acgcggtgag cagccctca tcggaatcag tgcgggcaac 120
agttatttca gccagaagaa caccgtcatg ctgctgcaat gggccgggca gcgtttcgag 180
cgcaccgatg tcgtctatgt cgacaccac atcgacgaga tgctgatcgc cgacggccgc 240
agcgcgcagg aggccgagcg gtcggtcaaa cgcacgtca aggatctgcg gcgcagactc 300
cggcgctcgc tggagagcgt gggcgaccac gccgagcggt tccgtgtccg gtccctgtcc 360
gagctccagg agaccctga gtaccgggccc gtacgcgagc gcaccgaccg ggccttcgag 420
gaggacgccg aattcgccac cgcctgcgag gacatggtgc gggccgtggt gatgaaccgg 480
cccggtgacg gcgtcggcat ctccgcgga cacctgcggg ccggtctgaa ctacgtgctg 540
gccgaggccc cgctcttcgc ggactcgcgc ggagtcttct ccgtcccctc ctcggtgctc 600
tgctaccaca tcgacacccc gatcacggcg ttctgtccc ggcgcgagac cggtttcgag 660
gcggccgagg gacaggcgta cgtcgtcgtc agggcccagg agctggccga cgcggcctag 720

```

<210> 4

<211> 834

<212> ADN

<213> 诺尔斯链酶菌

<400> 4

```

atgtcatggg gaggacagga cacttgctca tggtgcgga cggggcccct cggcgaagct 60
gaagacgtag gaagacagca cacgtcgac gccgggggac ccgtcatgac tcaagccgcc 120
accgtcaccg ccaccacgag ccagggcagg gcaactctgc ggagcctgac gccgctgttc 180
gtggacgccg cgatcccgtc cggctcgtac ttctcctcg ccgagggctt cggcatgagc 240
acggtcgcgc cgctggcctg gagcagcgtg gtcccggcgc tgcgcacgat ctggggcctg 300
gtccgggagc ggacgggtcaa cggcctcgcg ctgctgatcc tcgtcgtcaa cgtggtgggg 360
ctggcgacga gcaccctgac cggcgatgcc cggctgatga tggccaagga cagcggcgtc 420
agcagcgtcg tcgggatcgc gatcctgctc tcggtgcgcg gccggcgccc gctgatgacc 480
gccggactcc ggccctgggt gaccaaggga agcccggagg ggaacgcgcg atgggaccgg 540
ctgtggggcg gcagcgcgcg gttccggcaa ctggagcggc gattctcgac ggtctggggg 600
agcgcctgct tgatcgagtg cgtggtcaag gtcgtcggtg cgtacgtcct gccggtgcac 660
accatggtgt ggctgggcac ggtgctgacg gtggtggcga tcctgctggc catggtggtc 720
gcgggcggcg gcagcgccga gccgatggag cggatggtca aggccgaggt cggggccgcc 780
ggcgaggccg ccacggcggg gaacgcggag ccggcgccgg ccgccgcggc ctga 834

```

<210> 5

<211> 3839

<212> ADN

<213> 诺尔斯链酶菌

<400> 5

```

ggatccgtcc cgacgggcgg gaaccgggtga ggcgccaccc atcgcatcgc ccgtaccgcg 60
gcgggtgtga ggtgcgccc aaaagaagg gattgatgtt agctcacagt tcatctgaat 120
cgccgcggga atccttgccg gacgcgtgga cggctctcaa aaccgctacc gccgtccgca 180
attacgcgaa agagccggtc gacgacgcgc tgatcgagca gctgttgagg gccatgctcg 240
ccgcgccgac cgcctccaac cggcaggcgt ggtcgttcat ggtggtgcgc aggcccgccg 300
cggtcgcgcg gctgcgcgcg ttctcgcgcg ggggtgctggg aacccccgcc ttcttcgctc 360
tggcctgctg cgaccgcagt ctgaccgaca acctctcccc gaagctctcg cagaagatct 420
acgacaccag caagctctgt gtcgccatgg cgggtggagaa cctgctgctc gcggcgcacg 480

```

cggccggcct	ggcgggatgc	ccggtgggca	gcttcaggte	cgacatcgte	accagcatgc	540
tcggtatccc	ggaacacatc	gagccgatgc	tcgtggtecc	gatcggccgt	cccgcgacag	600
ccctcgtecc	ctcccagcgc	cgcgccaaga	atgaggtegt	caactatgaa	tcctggggaa	660
accgtgctgc	cgccccaaact	gcgtgaggag	atcgcgctcc	tcgcegtcta	tctgctcagc	720
agcgcccgcg	gactcctgga	ggagccggcc	gactacggaa	tttaccgctg	taccgacggg	780
gcccgtcggg	cgctccaact	cctcgacgaa	cacggcgga	gcacggcacg	gctgacggcc	840
gtccgcgagc	gtctcgacga	ggtcatgttc	gcgccgatgg	gcgaggaccg	ggacatgggc	900
gcgattcttg	acgacctgtg	tcgccaaatg	gcagacgctc	ttccggaaat	tgaaaccccc	960
tgacggctgt	ccggggcaac	cccaaaaggga	cttcttagca	tgcttgacagg	cttagttccc	1020
gcgccggacc	acggaatgcg	ggaagaaata	cttggcgacc	gcagccgatt	gatccggcaa	1080
cgcggtgagc	acgccctcat	cggaatcagt	gcgggcaaca	gttatttcag	ccagaagaac	1140
accgtcatgc	tgctgcaatg	ggccgggcag	cgtttcgagc	gcaccgatgt	cgctctatgtc	1200
gacacccaca	tcgacgagat	gctgatcgcc	gacggccgca	gcgcgcagga	ggccgagcgg	1260
tcggtcaaac	gcacgctcaa	ggatctgcgg	cgcagactcc	ggcgctcgct	ggagagcgtg	1320
ggcgaccacg	ccgagcgggt	ccgtgtccgg	tcctgtcccg	agctccagga	gacccttgag	1380
taccggggccg	tacgcgagcg	caccgaccgg	gccttcgagg	aggacgccga	attcgccacc	1440
gcctgcgagg	acatgggtgcg	ggccgtgggtg	atgaaccggc	ccggtgacgg	cgctggcatc	1500
tccgcggaac	acctgcgggc	cggtctgaac	tacgtgctgg	ccgaggcccc	gctcttcgcg	1560
gactcgcccc	gagtcttctc	cgctccctcc	tcggtgctct	gctaccacat	cgacaccccc	1620
atcacggcgt	tcctgtcccg	gcgcgagacc	ggtttccggg	cgcccgaggg	acaggcgta	1680
gtcgctcgta	ggccccagga	gctggccgac	gcggcctagt	tgggggcgtc	cgcgggcgga	1740
cctgcctccc	caccgcctcc	cggtgcgggc	gccgggcattg	acaaatgtca	tggggaggac	1800
aggacacttg	ctcatggtgc	ggaacggggc	ccctcggcga	agctgaagac	gtaggaagac	1860
agcacacgtc	gcacgcgggg	ggaccctca	tgactcaagc	cgccaccgtc	accgccacca	1920
cgagccaggg	cagggcactc	ctgcggagcc	tgacgcgcgt	gttcgtggac	gccgcgatac	1980
cgctcggtc	gtacttcctc	ctcgccgagg	gcttcggcat	gagcacggtc	gccgcgctgg	2040
cctggagcag	cgtaggtccc	gcgctgcgca	cgatctgggg	cctggtccgg	gagcggacgg	2100
tcaacggcct	cgcgctgctg	atcctcgctc	tcaacgtgg	ggggctggcg	acgagcacc	2160
tgaccggcga	tgcccggtcg	atgatggcca	aggacagcgg	cgtcagcagc	gtcgctggga	2220
tcgcatcct	gctctcggtg	cgcgccgggc	gcccgctgat	gaccgcggga	ctccggccct	2280
gggtgaccac	gggaagcccc	gaggggaacg	ccgcatggga	ccggtgtgg	gcgcgcagcg	2340
cgcggttccg	gcaactggag	cggcgattct	cgacggctctg	ggggagcgcc	ctgctgatcg	2400
agtgcgtgg	caaggctcgtc	ggtgcgtacg	tcctgcgggt	gcacaccatg	gtgtggctgg	2460
gcacggtgct	gacggtgggtg	gcgatcctgc	tggccatggt	ggtcgcgggc	ggcggcagcg	2520
ccgagccgat	ggagcggatg	gtcaaggccg	aggtcggggc	cgccggcgag	gccgccacgg	2580
cggggaacgc	cgagccggcg	ccggccggcc	cggcctgaga	ccgcgcggcg	ggggagttgg	2640
ggaaatcgcg	tacaggattg	gcgcgtcgag	cacccccgcc	ctcgatagg	cgggcccccg	2700
gcgcataatg	tcggcgatgc	gacgggacat	cgagccccg	cgtcgacggt	tcaacggcga	2760
tccggacggc	acgcggcttt	cgctggccac	gaagggaaacg	gaagtcatgt	cgactgttca	2820
cactggggtc	acgcagagcg	gtctcaccgc	cgagctggcc	tcctgcacg	ccgagctcgt	2880
ccgtcggaat	cccgggtgaag	cggagttcca	ccaggcggcc	ctggagggtcc	tcgaaacgct	2940
ggcaccgggtg	ctcaccgccc	ggccggaggtt	cgccgacgcc	aaggtcctgg	agcggatcgt	3000
cgagccggag	cggcagatca	tggtccgcgt	gccctggcag	gacgactccg	gcacgatccg	3060
ggtcaaccgc	ggcttccggg	tggagttcaa	cagcgcgctc	ggcccctaca	agggcgccct	3120
gcggttccac	gcgtccgtca	acctcggcac	cgtagaagttc	ctcggttccg	agcagatctt	3180
caagaacgcc	ctgaccgggc	tgaacatcgg	cgccggcaag	ggcggcagcg	acttcgaccc	3240
gcacggcagg	tcggacgcgc	aggtgatgcg	cttctgccag	tccttcatga	ccgagctgca	3300
ccgtcacctg	ggcgagcaca	ccgacgtgcc	ggctggcgac	atcggcgtcg	gcggccggga	3360
gatcggtac	ctcttcggcc	agtaccggcg	gatcaccaac	cgctgggagg	ccggcgtcct	3420
gaccggcaag	ggcctggcgt	ggggcggtc	caaggcccgt	acggaggcca	ccggttacgg	3480
caatgtgctg	ttcaccgagg	agatgctcaa	gcagcgcggc	gaggagctgg	acggccagca	3540
ggtggtggte	tccgggtccg	gcaacgtcgc	catctacacc	atcgagaagg	cccaggcgct	3600
cggcgccaac	gtcctgacgg	tctcggaactc	cggcggctac	gtcgtcgacg	agaaggcat	3660
cgacctggcg	ctgctcaagc	aggtcaagga	ggtcgagcgc	ggccgggtcg	gcgactacgc	3720
ccagcggcgc	ggcagttcgg	cgaagtacgt	cgccggcggg	agcgtgtggg	acgtcgctcg	3780
tgacgtggcg	ctgccgtcgg	ccaccagaa	cgagctcgac	gcggacgccg	cccggatcc	3839

<210> 6

<211> 219

<212> PRT

<213> 诺尔斯链酶菌

<400> 6

Met	Arg	Arg	His	Pro	Ser	His	Ser	Pro	Tyr	Arg	Gly	Gly	Cys	Glu	Val
1				5					10					15	

Arg	Pro	Lys	Arg	Arg	Gly	Leu	Met	Leu	Ala	His	Ser	Ser	Ser	Glu	Ser
			20					25					30		

Pro	Pro	Glu	Ser	Leu	Pro	Asp	Ala	Trp	Thr	Val	Leu	Lys	Thr	Arg	Thr
		35					40					45			

Ala	Val	Arg	Asn	Tyr	Ala	Lys	Glu	Pro	Val	Asp	Asp	Ala	Leu	Ile	Glu
	50					55					60				

Gln	Leu	Leu	Glu	Ala	Met	Leu	Ala	Ala	Pro	Thr	Ala	Ser	Asn	Arg	Gln
65					70					75					80

Ala	Trp	Ser	Phe	Met	Val	Val	Arg	Arg	Pro	Ala	Ala	Val	Arg	Arg	Leu
				85					90						95

Arg	Ala	Phe	Ser	Pro	Gly	Val	Leu	Gly	Thr	Pro	Ala	Phe	Phe	Val	Val
			100					105					110		

Ala	Cys	Val	Asp	Arg	Ser	Leu	Thr	Asp	Asn	Leu	Ser	Pro	Lys	Leu	Ser
		115					120					125			

Gln	Lys	Ile	Tyr	Asp	Thr	Ser	Lys	Leu	Cys	Val	Ala	Met	Ala	Val	Glu
	130						135				140				

Asn	Leu	Leu	Leu	Ala	Ala	His	Ala	Ala	Gly	Leu	Gly	Gly	Cys	Pro	Val
145					150					155					160

Gly	Ser	Phe	Arg	Ser	Asp	Ile	Val	Thr	Ser	Met	Leu	Gly	Ile	Pro	Glu
				165					170					175	

His	Ile	Glu	Pro	Met	Leu	Val	Val	Pro	Ile	Gly	Arg	Pro	Ala	Thr	Ala
			180					185					190		

Leu	Val	Pro	Ser	Gln	Arg	Arg	Ala	Lys	Asn	Glu	Val	Val	Asn	Tyr	Glu
		195					200					205			

Ser	Trp	Gly	Asn	Arg	Ala	Ala	Ala	Pro	Thr	Ala
	210					215				

<210> 7

<211> 104

<212> PRT

<213> 诺尔斯链酶菌

<400> 7

```

Asn Pro Gly Glu Thr Val Leu Pro Pro Gln Leu Arg Glu Glu Ile Ala
 1              5              10              15

Leu Leu Ala Val Tyr Leu Leu Ser Ser Gly Arg Gly Leu Leu Glu Glu
      20              25              30

Pro Ala Asp Tyr Gly Ile Tyr Arg Cys Thr Asp Gly Ala Arg Arg Ala
      35              40              45

Leu Gln Leu Leu Asp Glu His Gly Gly Ser Thr Ala Arg Leu Thr Ala
 50              55              60

Val Arg Glu Arg Leu Asp Glu Val Met Phe Ala Pro Met Gly Glu Asp
 65              70              75              80

Arg Asp Met Gly Ala Ile Leu Asp Asp Leu Cys Arg Gln Met Ala Asp
      85              90              95

Ala Leu Pro Glu Ile Glu Thr Pro
      100

```

<210> 8

<211> 99

<212> PRT

<213> 诺尔斯链酶菌

<400> 8

```

Met Leu Pro Pro Gln Leu Arg Glu Glu Ile Ala Leu Leu Ala Val Tyr
 1              5              10              15

Leu Leu Ser Ser Gly Arg Gly Leu Leu Glu Glu Pro Ala Asp Tyr Gly
      20              25              30

Ile Tyr Arg Cys Thr Asp Gly Ala Arg Arg Ala Leu Gln Leu Leu Asp
      35              40              45

Glu His Gly Gly Ser Thr Ala Arg Leu Thr Ala Val Arg Glu Arg Leu
 50              55              60

Asp Glu Val Met Phe Ala Pro Met Gly Glu Asp Arg Asp Met Gly Ala
 65              70              75              80

Ile Leu Asp Asp Leu Cys Arg Gln Met Ala Asp Ala Leu Pro Glu Ile
      85              90              95

Glu Thr Pro

```

<210> 9

<211> 239

<212> PRT

<213> 诺尔斯链酶菌

<400> 9

```

Met Leu Ala Gly Leu Val Pro Ala Pro Asp His Gly Met Arg Glu Glu
 1              5              10              15

Ile Leu Gly Asp Arg Ser Arg Leu Ile Arg Gln Arg Gly Glu His Ala
      20              25              30

Leu Ile Gly Ile Ser Ala Gly Asn Ser Tyr Phe Ser Gln Lys Asn Thr
      35              40              45

Val Met Leu Leu Gln Trp Ala Gly Gln Arg Phe Glu Arg Thr Asp Val
      50              55              60

Val Tyr Val Asp Thr His Ile Asp Glu Met Leu Ile Ala Asp Gly Arg
      65              70              75              80

Ser Ala Gln Glu Ala Glu Arg Ser Val Lys Arg Thr Leu Lys Asp Leu
      85              90              95

Arg Arg Arg Leu Arg Arg Ser Leu Glu Ser Val Gly Asp His Ala Glu
      100             105             110

Arg Phe Arg Val Arg Ser Leu Ser Glu Leu Gln Glu Thr Pro Glu Tyr
      115             120             125

Arg Ala Val Arg Glu Arg Thr Asp Arg Ala Phe Glu Glu Asp Ala Glu
      130             135             140

Phe Ala Thr Ala Cys Glu Asp Met Val Arg Ala Val Val Met Asn Arg
      145             150             155             160

Pro Gly Asp Gly Val Gly Ile Ser Ala Glu His Leu Arg Ala Gly Leu
      165             170             175

Asn Tyr Val Leu Ala Glu Ala Pro Leu Phe Ala Asp Ser Pro Gly Val
      180             185             190

Phe Ser Val Pro Ser Ser Val Leu Cys Tyr His Ile Asp Thr Pro Ile
      195             200             205

Thr Ala Phe Leu Ser Arg Arg Glu Thr Gly Phe Arg Ala Ala Glu Gly
      210             215             220

Gln Ala Tyr Val Val Val Arg Pro Gln Glu Leu Ala Asp Ala Ala
      225             230             235

```

<210> 10

<211> 277

<212> PRT

<213> 诺尔斯链酶菌

<400> 10

```

Met Ser Trp Gly Gly Gln Asp Thr Cys Ser Trp Cys Gly Thr Gly Pro
 1              5              10              15

```

Leu Gly Glu Ala Glu Asp Val Gly Arg Gln His Thr Ser His Ala Gly
 20 25 30

Gly Pro Val Met Thr Gln Ala Ala Thr Val Thr Ala Thr Thr Ser Gln
 35 40 45

Gly Arg Ala Leu Leu Arg Ser Leu Thr Pro Leu Phe Val Asp Ala Ala
 50 55 60

Ile Pro Leu Gly Ser Tyr Phe Leu Leu Ala Glu Gly Phe Gly Met Ser
 65 70 75 80

Thr Val Ala Ala Leu Ala Trp Ser Ser Val Val Pro Ala Leu Arg Thr
 85 90 95

Ile Trp Gly Leu Val Arg Glu Arg Thr Val Asn Gly Leu Ala Leu Leu
 100 105 110

Ile Leu Val Val Asn Val Val Gly Leu Ala Thr Ser Thr Leu Thr Gly
 115 120 125

Asp Ala Arg Leu Met Met Ala Lys Asp Ser Gly Val Ser Ser Val Val
 130 135 140

Gly Ile Ala Ile Leu Leu Ser Val Arg Gly Arg Arg Pro Leu Met Thr
 145 150 155 160

Ala Gly Leu Arg Pro Trp Val Thr Lys Gly Ser Pro Glu Gly Asn Ala
 165 170 175

Ala Trp Asp Arg Leu Trp Ala Arg Ser Ala Arg Phe Arg Gln Leu Glu
 180 185 190

Arg Arg Phe Ser Thr Val Trp Gly Ser Ala Leu Leu Ile Glu Cys Val
 195 200 205

Val Lys Val Val Gly Ala Tyr Val Leu Pro Val His Thr Met Val Trp
 210 215 220

Leu Gly Thr Val Leu Thr Val Val Ala Ile Leu Leu Ala Met Val Val
 225 230 235 240

Ala Gly Gly Gly Ser Ala Glu Pro Met Glu Arg Met Val Lys Ala Glu
 245 250 255

Val Gly Ala Ala Gly Glu Ala Ala Thr Ala Gly Asn Ala Glu Pro Ala
 260 265 270

Pro Ala Ala Ala Ala
 275

<210> 11

<211> 20

<212> PRT

<213> 诺尔斯链酶菌

<400> 11

Glu Pro Val Asp Asp Ala Leu Ile Glu Gln Leu Leu Glu Ala Met Leu
 1 5 10 15

Ala Ala Pro Thr

20

<210> 12

<211> 12

<212> PRT

<213> 诺尔斯链酶菌

<400> 12

Asn Glu Val Val Asn Tyr Glu Xaa Trp Gly Asn Arg
 1 5 10

<210> 13

<211> 9

<212> PRT

<213> 诺尔斯链酶菌

<400> 13

Gln Ala Xaa Ser Phe Met Val Val Arg
 1 5

<210> 14

<211> 17

<212> ADN

<213> 人工序列

<220>

<223> 人工序列的描述: 引物

<400> 14

garccsgtsg acgacgc

17

<210> 15

<211> 17

<212> ADN

<213> 人工序列

<220>

<223> 人工序列的描述: 引物

<400> 15

gcgtcgtsa csggytc

17

<210> 16

<211> 20

<212> ADN

<213> 人工序列	
<220>	
<223> 人工序列的描述: 引物	
<400> 16	
aacgargtsg tsaactacga	20
<210> 17	
<211> 20	
<212> ADN	
<213> 人工序列	
<220>	
<223> 人工序列的描述: 引物	
<400> 17	
tcgtagtttsa csacytcggt	20
<210> 18	
<211> 20	
<212> ADN	
<213> 人工序列	
<220>	
<223> 人工序列的描述: 引物	
<400> 18	
caggcstggw ssttcacggt	20
<210> 19	
<211> 20	
<212> ADN	
<213> 人工序列	
<220>	
<223> 人工序列的描述: 引物	
<400> 19	
accatgaass wccasgcctg	20
<210> 20	
<211> 47	
<212> ADN	
<213> 人工序列	
<220>	
<223> 人工序列的描述: 引物	
<400> 20	
cggctgcagg agaagggagc ggacatatgc ttgcaggctt agttccc	47

<210> 21
<211> 42
<212> ADN
<213> 人工序列

<220>
<223> 人工序列的描述: 引物

<400> 21
cgggtcccgtg gatccaagct tctaggccgc gtcggccagc tc 42

<210> 22
<211> 36
<212> ADN
<213> 人工序列

<220>
<223> 人工序列的描述: 引物

<400> 22
gagcgggatc ctgcagtgtc atggggagga caggac 36

<210> 23
<211> 41
<212> ADN
<213> 人工序列

<220>
<223> 人工序列的描述: 引物

<400> 23
cgatcacgtg gatccaagct tgccaatcct gtacgcgatt t 41

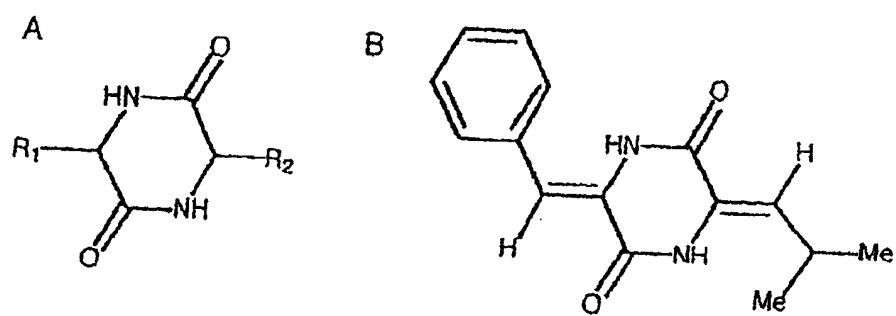


图1

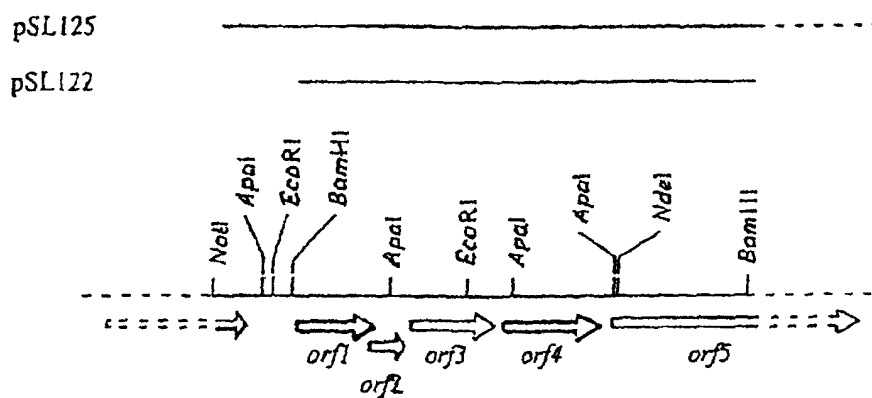


图2

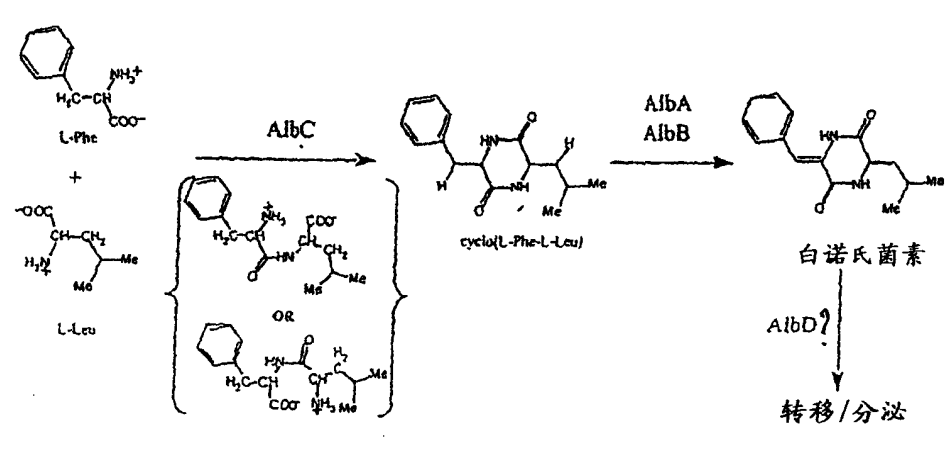
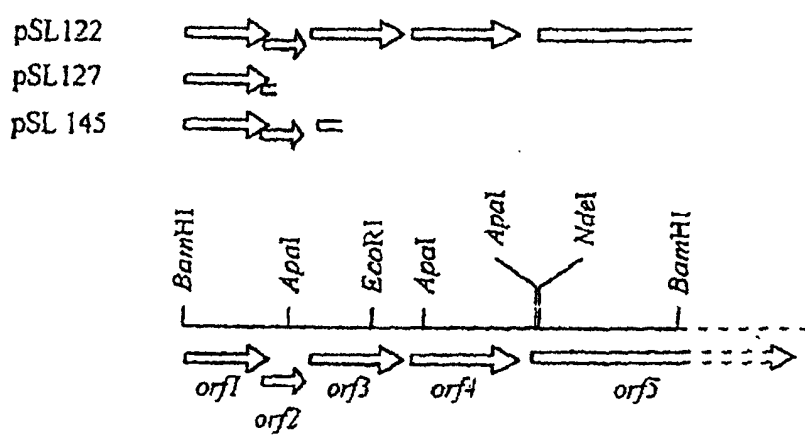


图 3

In 大肠杆菌



In 变青链霉菌

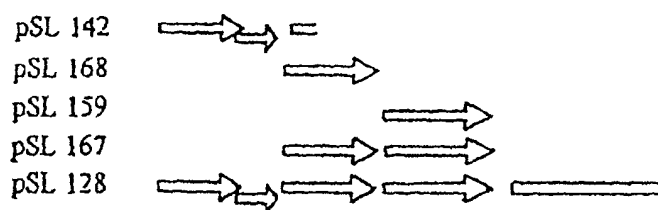
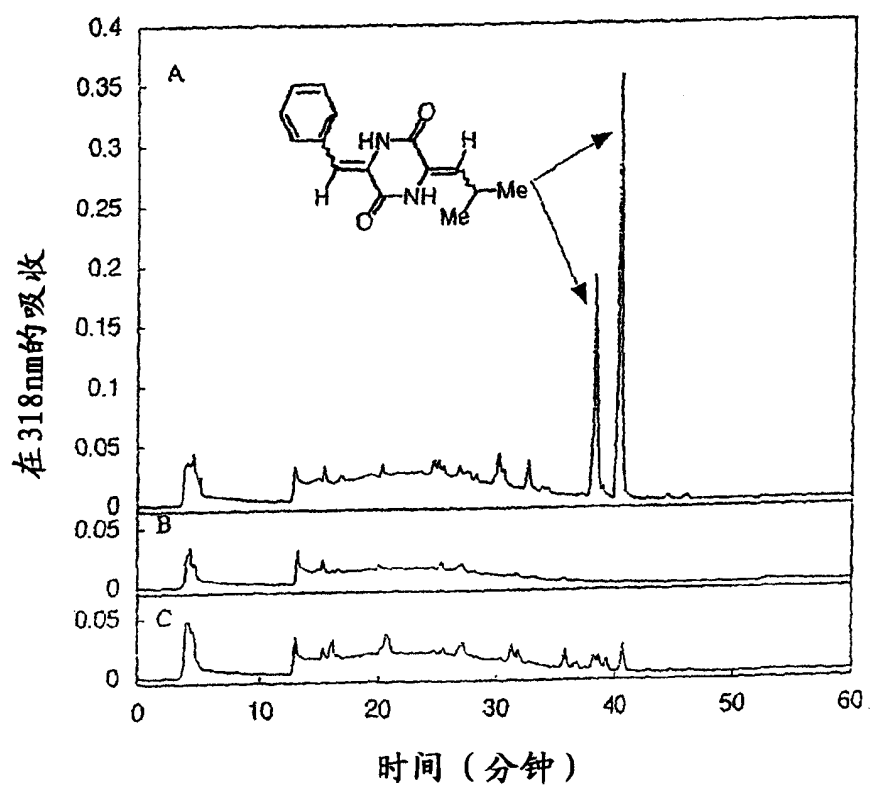


图 4



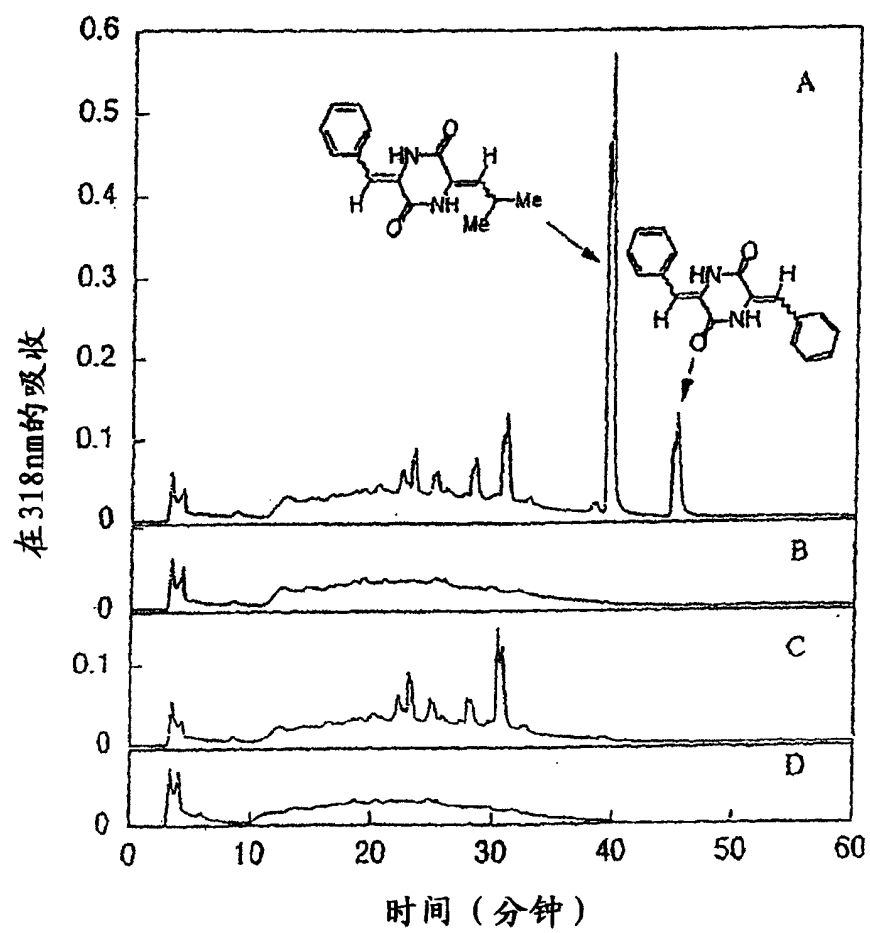


图6

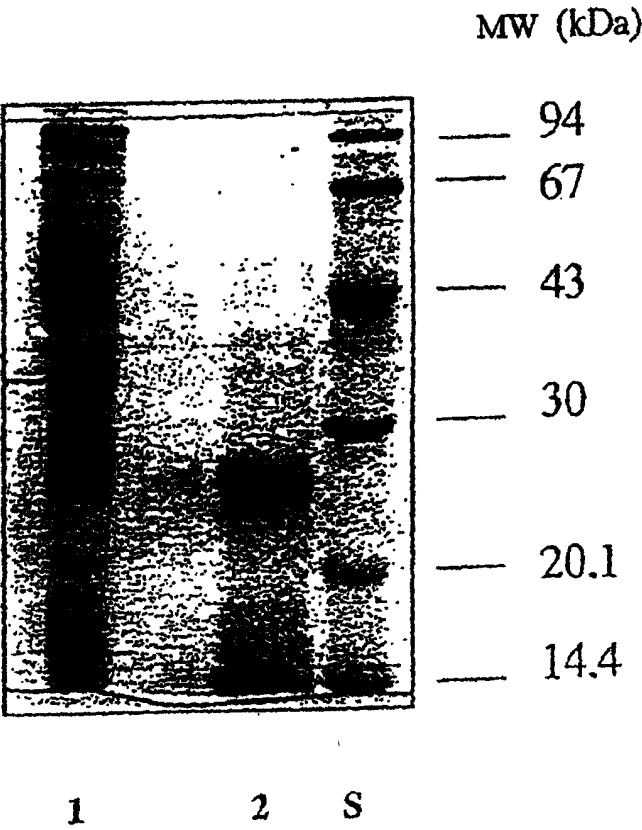


图 7