

[19] 中华人民共和国国家知识产权局



[12] 发明专利说明书

专利号 ZL 200380104850.6

[51] Int. Cl.

C12N 9/54 (2006.01)
C12N 15/57 (2006.01)
C12N 1/21 (2006.01)
C12P 21/02 (2006.01)
C12R 1/19 (2006.01)
C12R 1/125 (2006.01)

[45] 授权公告日 2008年7月9日

[11] 授权公告号 CN 100400660C

[22] 申请日 2003.12.26

[21] 申请号 200380104850.6

[30] 优先权

[32] 2002.12.26 [33] JP [31] 376054/2002

[32] 2003.12.18 [33] JP [31] 420887/2003

[86] 国际申请 PCT/JP2003/016936 2003.12.26

[87] 国际公布 WO2004/058960 日 2004.7.15

[85] 进入国家阶段日期 2005.6.2

[73] 专利权人 协和发酵工业株式会社

地址 日本东京

[72] 发明人 桥本信一 田畑和彦 羽田文

池田创

[56] 参考文献

WO0003009A2 2000.1.20

Bacillus subtilis genome project : cloning and sequence of the 97kb region from 325degrees to 333 degrees. . Glaser P et al. , . , Vol. 2, No. 2, (1993), p. 371. 384. Mol Microbiol, Vol. 10 No. 2. 1993

The complete genome sequence of the gram - positive bacterium Bacillus subtilis. . Kunst F et al. Nature, Vol. 390 No. 6657. 1997

审查员 王大鹏

[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 程泳

权利要求书 5 页 说明书 104 页 附图 2 页

[54] 发明名称

生产二肽的方法

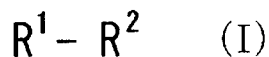
[57] 摘要

本发明提供了具有合成与 L-Ala-L-Ala 不同的二肽的活性的蛋白和用于合成二肽的蛋白、用于生产具有二肽合成活性的蛋白的方法、使用具有二肽合成活性的蛋白或用于二肽合成的蛋白生产二肽的方法、和使用能生产具有二肽合成活性的蛋白或用于二肽合成的蛋白的微生物培养物等作为酶源生产二肽的方法。

1. 选自下面的[1]至[4]的蛋白, 条件是不包括由 SEQ ID NO: 1 所示的氨基酸序列组成的蛋白以及数据库 Swiss-Prot 中登录号为 Q8KWT3 的蛋白质:

[1] 由 SEQ ID NO: 2 至 8 中的任一个所示的氨基酸序列组成的蛋白;

[2] 含有氨基酸序列的蛋白, 其中在 SEQ ID NO: 1 至 8 中的任一个所示的氨基酸序列中缺失、取代或添加了一个或多个氨基酸残基, 且其具有合成式(I)所示的二肽的活性:



其中, R^1 是 L-Ala、Gly、L-Met、L-Ser、L-Thr 或 β -Ala, R^2 是 L-Gln、L-Glu、Gly、L-Val、L-Leu、L-Ile、L-Pro、L-Phe、L-Trp、L-Met、L-Ser、L-Thr、L-Cys、L-Asn、L-Tyr、L-Lys、L-Arg、L-His、L-Asp、L- α -氨基丁酸、 β -Ala、L-重氮丝氨酸、L-茶氨酸、L-4-HYP、L-3-HYP、L-Orn、L-Cit 或 L-6-重氮-5-氧-正亮氨酸; 或 R^1 是 L-Gln, R^2 是 L-Phe 或 L- α -氨基丁酸; 或 R^1 是 L-Phe, R^2 是 L-Gln; 或 R^1 是 L-Trp, R^2 是 Gly; 或 R^1 是 L-Cys, R^2 是 L-Ala、L-Gln、Gly 或 L-Met; 或 R^1 是 L-Lys, R^2 是 L-Ala、Gly 或 L-Met; 或 R^1 是 L-Arg, R^2 是 L- α -氨基丁酸; 或 R^1 是 L-His, R^2 是 L-Met; 或 R^1 是 L- α -氨基丁酸, R^2 是 L-Ala、L-Gln、Gly、L-Ser、L-Thr、L-Arg 或 L-氨基丁酸;

[3] 蛋白, 其含有的氨基酸序列表现出与 SEQ ID NO: 1 至 8 中的任一个所示的氨基酸序列具有 90%或更多的相似性, 且其具有合成式(I)所示的二肽的活性; 和

[4] 蛋白, 其含有的氨基酸序列表现出与 SEQ ID NO: 17 所示的氨基酸序列具有 90%或更多的相似性, 且其具有合成式(I)所示的二肽的活性。

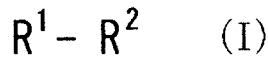
2. 选自下面的[1]至[4]的 DNA, 条件是不包括由 SEQ ID NO: 9

所示的核苷酸序列组成的 DNA:

[1] 编码根据权利要求 1 的蛋白的 DNA;

[2] 由 SEQ ID NO: 10 至 16 和 36 中的任一个所示的核苷酸序列组成的 DNA;

[3] DNA, 其能与由 SEQ ID NO: 9 至 16 和 36 中的任一个所示的核苷酸序列组成的 DNA 杂交, 杂交条件为: 在存在 0.7 至 1.0 mol/l 氯化钠的情况下, 在 65℃ 进行杂交, 然后在 65℃ 用 0.1 至 2 倍浓度的 SSC 溶液洗涤, 且其编码具有合成式 (I) 所示的二肽的活性的蛋白:



其中, R^1 是 L-Ala、Gly、L-Met、L-Ser、L-Thr 或 β -Ala, R^2 是 L-Gln、L-Glu、Gly、L-Val、L-Leu、L-Ile、L-Pro、L-Phe、L-Trp、L-Met、L-Ser、L-Thr、L-Cys、L-Asn、L-Tyr、L-Lys、L-Arg、L-His、L-Asp、L- α -氨基丁酸、 β -Ala、L-重氮丝氨酸、L-茶氨酸、L-4-HYP、L-3-HYP、L-Orn、L-Cit 或 L-6-重氮-5-氧-正亮氨酸; 或 R^1 是 L-Gln, R^2 是 L-Phe 或 L- α -氨基丁酸; 或 R^1 是 L-Phe, R^2 是 L-Gln; 或 R^1 是 L-Trp, R^2 是 Gly; 或 R^1 是 L-Cys, R^2 是 L-Ala、L-Gln、Gly 或 L-Met; 或 R^1 是 L-Lys, R^2 是 L-Ala、Gly 或 L-Met; 或 R^1 是 L-Arg, R^2 是 L- α -氨基丁酸; 或 R^1 是 L-His, R^2 是 L-Met; 或 R^1 是 L- α -氨基丁酸, R^2 是 L-Ala、L-Gln、Gly、L-Ser、L-Thr、L-Arg 或 L-氨基丁酸;

[4] DNA, 其由与 SEQ ID NO: 18 所示的核苷酸序列有 90% 或更多的相似性的核苷酸序列组成, 且编码具有合成式 (I) 所示的二肽的活性的蛋白。

3. 重组 DNA, 其含有根据权利要求 2 的 DNA。

4. 转化体, 其携带根据权利要求 3 的重组 DNA。

5. 根据权利要求 4 的转化体, 其中该转化体是使用微生物作为宿主得到的转化体。

6. 根据权利要求 5 的转化体, 其中该微生物是埃希氏菌属的微生物。

7. 生产根据权利要求1的蛋白的方法,其包括:在培养基中培养根据权利要求4至6中的任一项的转化体,使根据权利要求1的蛋白在培养物中形成和积累,和从培养物中回收蛋白。

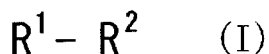
8. 生产根据权利要求1的蛋白的方法,其包括:在培养基中培养能生产根据权利要求1的蛋白的微生物,使所述蛋白在培养物中形成和积累,和从培养物中回收所述蛋白。

9. 根据权利要求8的方法,其中该微生物是芽孢杆菌属的微生物。

10. 根据权利要求9的方法,其中该芽孢杆菌属的微生物是能生产芽孢杆菌溶素的芽孢杆菌属的微生物。

11. 根据权利要求10的方法,其中所述能生产芽孢杆菌溶素的芽孢杆菌属的微生物是选自下面种属的微生物:枯草芽孢杆菌、解淀粉芽孢杆菌、凝结芽孢杆菌、地衣芽孢杆菌、巨大芽孢杆菌和短小芽孢杆菌。

12. 生产式(I)所示的二肽的方法:



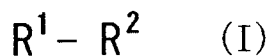
其中, R^1 是 L-Ala、Gly、L-Met、L-Ser、L-Thr 或 β -Ala, R^2 是 L-Gln、L-Glu、Gly、L-Val、L-Leu、L-Ile、L-Pro、L-Phe、L-Trp、L-Met、L-Ser、L-Thr、L-Cys、L-Asn、L-Tyr、L-Lys、L-Arg、L-His、L-Asp、L- α -氨基丁酸、 β -Ala、L-重氮丝氨酸、L-茶氨酸、L-4-HYP、L-3-HYP、L-Orn、L-Cit 或 L-6-重氮-5-氧-正亮氨酸; 或 R^1 是 L-Gln, R^2 是 L-Phe 或 L- α -氨基丁酸; 或 R^1 是 L-Phe, R^2 是 L-Gln; 或 R^1 是 L-Trp, R^2 是 Gly; 或 R^1 是 L-Cys, R^2 是 L-Ala、L-Gln、Gly 或 L-Met; 或 R^1 是 L-Lys, R^2 是 L-Ala、Gly 或 L-Met; 或 R^1 是 L-Arg, R^2 是 L- α -氨基丁酸; 或 R^1 是 L-His, R^2 是 L-Met; 或 R^1 是 L- α -氨基丁酸, R^2 是 L-Ala、L-Gln、Gly、L-Ser、L-Thr、L-Arg 或 L-氨基丁酸;

使根据权利要求1的蛋白或序列为 SEQ ID NO: 1 的蛋白、至少2个相同或不同的氨基酸、和 ATP 存在于水性介质中;

使二肽在介质中形成和积累; 和

从介质中回收二肽。

13. 生产式(I)所示的二肽的方法:



其中, R^1 是 L-Ala、Gly、L-Met、L-Ser、L-Thr 或 β -Ala, R^2 是 L-Gln、L-Glu、Gly、L-Val、L-Leu、L-Ile、L-Pro、L-Phe、L-Trp、L-Met、L-Ser、L-Thr、L-Cys、L-Asn、L-Tyr、L-Lys、L-Arg、L-His、L-Asp、L- α -氨基丁酸、 β -Ala、L-重氮丝氨酸、L-茶氨酸、L-4-HYP、L-3-HYP、L-Orn、L-Cit 或 L-6-重氮-5-氧-正亮氨酸; 或 R^1 是 L-Gln, R^2 是 L-Phe 或 L- α -氨基丁酸; 或 R^1 是 L-Phe, R^2 是 L-Gln; 或 R^1 是 L-Trp, R^2 是 Gly; 或 R^1 是 L-Cys, R^2 是 L-Ala、L-Gln、Gly 或 L-Met; 或 R^1 是 L-Lys, R^2 是 L-Ala、Gly 或 L-Met; 或 R^1 是 L-Arg, R^2 是 L- α -氨基丁酸; 或 R^1 是 L-His, R^2 是 L-Met; 或 R^1 是 L- α -氨基丁酸, R^2 是 L-Ala、L-Gln、Gly、L-Ser、L-Thr、L-Arg 或 L-氨基丁酸;

使酶源和至少 2 个相同或不同的氨基酸存在于水性介质中, 所述的酶源是选自下面的 [1] 至 [3] 的培养物或处理过的培养物:

[1] 根据权利要求 4 至 6 中的任一项的转化体的培养物, 或处理过的所述培养物;

[2] 能生产根据权利要求 1 的蛋白的微生物的培养物, 或处理过的所述培养物; 和

[3] 能生产 SEQ ID NO: 1 所示氨基酸序列的微生物的培养物, 或处理过的所述培养物;

使二肽在介质中形成和积累; 和

从介质中回收二肽。

14. 根据权利要求 13 的方法, 其中能生产根据权利要求 1 的蛋白的微生物是芽孢杆菌属的微生物。

15. 根据权利要求 14 的方法, 其中所述芽孢杆菌属的微生物是能生产芽孢杆菌溶素的芽孢杆菌属的微生物。

16. 根据权利要求 15 的方法, 其中能生产芽孢杆菌溶素的芽孢杆菌属的微生物是选自下面种属的微生物: 枯草芽孢杆菌、解淀粉芽孢

杆菌、凝结芽孢杆菌、地衣芽胞杆菌、巨大芽孢杆菌和短小芽孢杆菌。

17. 根据权利要求 13 的方法，其中能生产 SEQ ID NO: 1 所示氨基酸序列的微生物是用包含含有 SEQ ID NO: 9 所示的核苷酸序列的 DNA 的重组 DNA 转化的微生物，或枯草芽孢杆菌属的微生物。

18. 根据权利要求 17 的方法，其中携带包含含有 SEQ ID NO: 9 所示的核苷酸序列的 DNA 的重组 DNA 的微生物是埃希氏菌属的微生物。

19. 根据权利要求 13 至 18 中的任一项的方法，其中处理过的所述培养物是浓缩的培养物、干燥的培养物、通过离心培养物得到的细胞、通过提取细胞得到的酶制品或通过对细胞进行下述操作得到的产物：干燥，冻干，用表面活性剂、超声破碎、机械摩擦处理，用溶剂处理，酶处理，蛋白分级或固定化。

生产二肽的方法

技术领域

本发明涉及具有二肽合成活性的蛋白和用于二肽合成的蛋白、用于生产具有二肽合成活性的蛋白的方法、使用具有二肽合成活性的蛋白或用于二肽合成的蛋白生产二肽的方法、能生产具有二肽合成活性的蛋白或用于二肽合成的蛋白的微生物或转化体、和使用这样的微生物或转化体生产二肽的方法。

背景技术

对于大规模的肽合成，可以得到化学合成方法（液相方法和固相方法）、酶合成方法和使用重组 DNA 技术的生物合成方法。目前，酶合成方法和生物合成方法用于合成超过 50 个残基的长链肽，化学合成方法和酶合成方法主要用于合成二肽。

但是，在用化学合成方法合成二肽时，必须进行为官能团导入和去除保护基的操作，还会形成外消旋物。因而认为化学合成方法在成本和效率方面是不利的。从环境卫生的角度看，它们也是不利的，因为需要使用大量的有机溶剂等。

关于酶法合成二肽，已知有下述的方法：使用蛋白酶的逆反应的方法 (J. Biol. Chem., 119, 707-720 (1937)); 使用热稳定的氨酰基 t-RNA 合成酶的方法 (日本公开的未审查的专利申请号 146539/83, 日本公开的未审查的专利申请号 209991/83, 日本公开的未审查的专利申请号 209992/83, 和日本公开的未审查的专利申请号 106298/84); 和使用非核糖体的肽合成酶(以下简称为 NRPS)的方法 (Chem. Biol., 7, 373-384 (2000), FEBS Lett., 498, 42-45 (2001), 美国专利号 5,795,738 和美国专利号 5,652,116)。

但是，使用蛋白酶的逆反应的方法需要为用作底物的氨基酸的官

能团导入和去除保护基，这会造成提高肽合成反应的效率和防止肽降解反应方面的困难。使用热稳定的氨酰基 t-RNA 合成酶的方法具有下述缺陷：难以表达酶和难以预防副反应形成目标产物以外的副产物。使用 NRPS 的方法的效率较差，由于难以通过重组 DNA 技术表达酶，因为它的酶分子较大，且必须提供辅酶 4'-磷酸泛酰巯基乙胺。

另一方面，有一组肽合成酶，其具有比 NRPS 更小的酶分子量，且不需要辅酶 4'-磷酸泛酰巯基乙胺；例如， γ -谷氨酰半胱氨酸合成酶、谷胱甘肽合成酶、D-丙氨酸-D-丙氨酸 (D-Ala-D-Ala) 连接酶、和聚- γ -谷氨酸合成酶。这些酶中的大多数使用 D-氨基酸作为底物，或催化在 γ -羧基处形成肽键。因为这样的性质，它们不能用于通过在 L-氨基酸的 α -羧基处形成肽键来合成二肽。

已知的能通过 L-氨基酸的 α -羧基处形成肽键的活性来合成二肽的酶的唯一实例是芽孢杆菌溶素 (来自芽孢杆菌属微生物的二肽抗生素) 合成酶。已知芽孢杆菌溶素合成酶具有合成芽孢杆菌溶素 [L-丙氨酰基-L-抗英膜菌素 (L-Ala-L-抗英膜菌素)] 和 L-丙氨酰基-L-丙氨酸 (L-Ala-L-Ala) 的活性，但是没有关于它的合成其它肽的信息 (J. Ind. Microbiol., 2, 201-208 (1987) 和 Enzyme. Microbial. Technol., 29, 400-406 (2001))。

关于已经清楚了其整体基因组信息的杆草芽孢杆菌 168 的芽孢杆菌溶素生物合成酶基因 (Nature, 390, 249-256 (1997))，已知通过扩增含有 ORF ywfA-F (W000/03009 小册子) 的芽孢杆菌溶素操纵子能提高芽孢杆菌溶素的生产力。但是，尚不清楚这些 ORF 是否含有能编码蛋白的 ORF，所述蛋白具有通过肽键连接 2 个或多个氨基酸的活性，如果含有，则是哪个 ORF 编码该蛋白？

本发明的一个目的是提供：具有合成二肽的活性的蛋白，所述二肽与 L-Ala-L-Ala 不同，迄今为止还没有提出关于它的使用肽合成酶等的酶合成方法；用于合成该二肽的蛋白；编码具有二肽合成活性的蛋白的 DNA；含有该 DNA 的重组 DNA；携带重组 DNA 的转化体；生产具有二肽合成活性的蛋白的方法；使用具有二肽合成活性的蛋白或用于

二肽合成的蛋白来合成二肽的酶促方法；和使用微生物培养物作为酶源来生产二肽的方法，所述微生物具有生产具有二肽合成活性的蛋白或用于二肽合成的蛋白等的的能力。

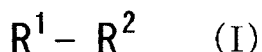
发明概述

本发明涉及下面的(1)至(21)。

(1) 选自下面的[1]至[4]的蛋白，条件是不包括由 SEQ ID NO: 1 所示的氨基酸序列组成的蛋白：

[1] 含有 SEQ ID NO: 2 至 8 中的任一个所示的氨基酸序列的蛋白；

[2] 含有氨基酸序列的蛋白，其中在 SEQ ID NO: 1 至 8 中的任一个所示的氨基酸序列中缺失、取代或添加了一个或多个氨基酸残基，且其具有合成式(I)所示的二肽的活性：

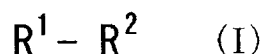


(其中 R^1 和 R^2 可以相同或不同，每个代表一个氨基酸残基，条件是 R^1 和 R^2 不同时代表 L-丙氨酸)；

[3] 蛋白，其含有的氨基酸序列表现出与 SEQ ID NO: 1 至 8 中的任一个所示的氨基酸序列具有 65%或更多的相似性，且其具有合成式(I)所示的二肽的活性；

[4] 蛋白，其含有的氨基酸序列表现出与 SEQ ID NO: 17 所示的氨基酸序列具有 80%或更多的相似性，且其具有合成式(I)所示的二肽的活性。

(2) 用于合成式(I)所示的二肽的蛋白：



(其中 R^1 和 R^2 可以相同或不同，每个代表一个氨基酸残基，条件是 R^1 和 R^2 不同时代表 L-丙氨酸)，其具有 SEQ ID NO: 1 所示的氨基酸序列。

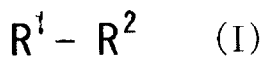
(3) 选自下面的[1]至[4]的 DNA，条件是不包括由 SEQ ID NO: 9

所示的核苷酸序列组成的 DNA:

[1] 能编码根据上面(1)的蛋白的 DNA;

[2] 含有 SEQ ID NO: 10 至 16 和 36 中的任一个所示的核苷酸序列的 DNA;

[3] DNA, 其能在严格条件下与含有 SEQ ID NO: 9 至 16 和 36 中的任一个所示的核苷酸序列的互补链的 DNA 杂交, 且其能编码具有合成式(I)所示的二肽的活性的蛋白:



(其中 R^1 和 R^2 可以相同或不同, 每个代表一个氨基酸残基, 条件是 R^1 和 R^2 不同时代表 L-丙氨酸); 和

[4] DNA, 其含有与 SEQ ID NO: 18 所示的核苷酸序列有 80% 或更多的相似性的核苷酸序列, 且能编码具有合成式(I)所示的二肽的活性的蛋白。

(4) 重组 DNA, 其含有根据上面(3)的 DNA。

(5) 转化体, 其携带根据上面(4)的重组 DNA。

(6) 根据上面(5)的转化体, 其中该转化体是使用微生物作为宿主得到的转化体。

(7) 根据上面(6)的转化体, 其中该微生物是埃希氏菌属的微生物。

(8) 生产根据上面(1)的蛋白的方法, 其包括: 在培养基中培养根据上面(5)至(7)中的任一项的转化体, 使根据上面(1)的蛋白在培养基中形成和积累, 和从培养基中回收蛋白。

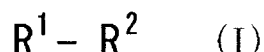
(9) 生产根据上面(1)的蛋白的方法, 其包括: 在培养基中培养能生产根据上面(1)的蛋白的微生物, 使蛋白在培养基中形成和积累, 和从培养基中回收蛋白。

(10) 根据上面(9)的方法, 其中该微生物是芽孢杆菌属的微生物。

(11) 根据上面(10)的方法, 其中该芽孢杆菌属的微生物是能生产芽孢杆菌溶素的芽孢杆菌属的微生物。

(12) 根据上面(11)的方法,其中所述能生产芽孢杆菌溶素的芽孢杆菌属的微生物是选自下面种属的微生物: 杆草芽孢杆菌 (*Bacillus subtilis*), 解淀粉芽孢杆菌 (*Bacillus amyloliquefaciens*), 凝结芽孢杆菌 (*Bacillus coagulans*), 地衣芽胞杆菌 (*Bacillus licheniformis*), 巨大芽孢杆菌 (*Bacillus megaterium*) 和短小芽孢杆菌 (*Bacillus pumilus*)。

(13) 生产式(I)所示的二肽的方法:



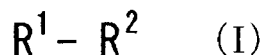
(其中 R^1 和 R^2 可以相同或不同, 每个代表一个氨基酸残基, 条件是 R^1 和 R^2 不同时代表 L-丙氨酸), 其包含:

使根据上面(1)的蛋白或根据上面(2)的用于二肽合成的蛋白、至少 2 个可以相同或不同的氨基酸、和 ATP 存在于水性介质中;

使二肽在介质中形成和积累; 和

从介质中回收二肽。

(14) 生产式(I)所示的二肽的方法:



(其中 R^1 和 R^2 可以相同或不同, 每个代表一个氨基酸残基, 条件是 R^1 和 R^2 不同时代表 L-丙氨酸), 其包含:

使酶源和至少 2 个可以相同或不同的氨基酸存在于水性介质中, 所述的酶源是选自下面的 [1] 至 [3] 的培养物或处理过的培养物:

[1] 根据上面(5)至(7)中的任一项的转化体的培养物, 或处理过的所述培养物;

[2] 能生产根据上面(1)的蛋白的微生物的培养物, 或处理过的所述培养物; 和

[3] 能生产根据上面(2)的用于二肽合成的蛋白的微生物, 或处理过的所述培养物;

使二肽在介质中形成和积累; 和

从介质中回收二肽。

(15) 根据上面(14)的方法,其中能生产根据上面(1)的蛋白的微生物是芽孢杆菌属的微生物。

(16) 根据上面(15)的方法,其中芽孢杆菌属的微生物是能生产芽孢杆菌溶素的芽孢杆菌属的微生物。

(17) 根据上面(16)的方法,其中能生产芽孢杆菌溶素的芽孢杆菌属的微生物是选自下面种属的微生物:杆草芽孢杆菌,解淀粉芽孢杆菌,凝结芽孢杆菌,地衣芽胞杆菌,巨大芽孢杆菌和短小芽孢杆菌。

(18) 根据上面(14)的方法,其中能生产根据上面(2)的用于二肽合成的蛋白的微生物是用重组DNA(其包含含有SEQ ID NO: 9所示的核苷酸序列的DNA)转化的微生物,或属于杆草芽孢杆菌的微生物。

(19) 根据上面(18)的方法,其中携带重组DNA(其包含含有SEQ ID NO: 9所示的核苷酸序列的DNA)的微生物是埃希氏菌属的微生物。

(20) 根据上面(14)至(19)中的任一项的方法,其中处理过的培养物是浓缩的培养物、干燥的培养物、通过离心培养物得到的细胞、通过提取细胞得到的酶制品或通过对细胞进行下述操作得到的产物:干燥,冻干,用表面活性剂、超声破碎、机械摩擦处理,用溶剂处理,酶处理,蛋白分级或固定化。

(21) 根据上面(13)至(20)中的任一项的方法,其中该二肽是式(II)代表的二肽:



(其中 R^3 和 R^4 可以相同或不同,每个代表选自下述的氨基酸残基: L-丙氨酸, L-谷氨酰胺, L-谷氨酸, 甘氨酸, L-缬氨酸, L-亮氨酸, L-异亮氨酸, L-脯氨酸, L-苯丙氨酸, L-色氨酸, L-甲硫氨酸, L-丝氨酸, L-苏氨酸, L-半胱氨酸, L-天冬酰胺, L-酪氨酸, L-赖氨酸, L-精氨酸, L-组氨酸, L-天门冬氨酸, L- α -氨基丁酸, β -丙氨酸, L-重氮丝氨酸, L-茶氨酸, L-4-羟基脯氨酸, L-3-羟基脯氨酸, L-鸟氨酸, L-瓜氨酸和 L-6-重氮-5-氧-正亮氨酸,条件是 R^3 和 R^4 不同时代

表 L-丙氨酸)。

根据本发明，可以生产出具有合成二肽的活性的蛋白，所述二肽与 L-Ala-L-Ala 不同，迄今为止还没有提出关于它的酶促合成方法。使用该蛋白、或能生产该蛋白的转化体或微生物可以生产出非 L-Ala-L-Ala 的二肽。

附图说明

图 1 显示了构建质粒 pPE43 的步骤。

图 2 显示了构建质粒 pQE60ywfE 的步骤。

符号解释

ywfE: 源自杆草芽孢杆菌 168 的 ywfE 基因

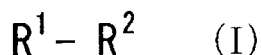
P_{trp}: 色氨酸启动子基因

PT5: T5 启动子基因

发明详述

本发明的蛋白包括:

- [1] 含有 SEQ ID NO: 2 至 8 中的任一个所示的氨基酸序列的蛋白;
- [2] 由氨基酸序列构成的蛋白，其中在 SEQ ID NO: 1 至 8 中的任一个所示的氨基酸序列中缺失、取代或添加了一个或多个氨基酸残基，且其具有合成式 (I) 所示的二肽的活性:



(其中 R^1 和 R^2 可以相同或不同，每个代表一个氨基酸残基，条件是 R^1 和 R^2 不同时代表 L-丙氨酸);

- [3] 蛋白，其含有的氨基酸序列表现出与 SEQ ID NO: 1 至 8 中的任一个所示的氨基酸序列具有 65%或更多的相似性，且其具有合成式 (I) 所示的二肽的活性; 和

- [4] 蛋白，其含有的氨基酸序列表现出与 SEQ ID NO: 17 所示的

氨基酸序列具有 80%或更多的相似性,且其具有合成式(I)所示的二肽的活性,条件是不包括由 SEQ ID NO: 1 所示的氨基酸序列组成的蛋白。

用于合成本发明的式(I)所示的二肽的蛋白的实例是由 SEQ ID NO: 1 所示的氨基酸序列组成的蛋白。

在下文中,本发明的上述蛋白和用于合成式(I)所示的二肽的蛋白可以一起称作本发明的蛋白。

可以得到上面的由氨基酸序列组成的蛋白,其中缺失、取代或添加了一个或多个氨基酸残基,且其具有合成式(I)所示的二肽的活性,例如通过下述描述的定点诱变将定位突变导入能编码由 SEQ ID NO: 1 至 8 中的任一个所示的氨基酸序列组成的蛋白的 DNA 中: Molecular Cloning, A Laboratory Manual, 第 2 版, Cold Spring Harbor Laboratory Press (1989) (以下简称为 Molecular Cloning, 第 2 版); Current Protocols in Molecular Biology, John Wiley & Sons (1987-1997) (以下简称为 Current Protocols in Molecular Biology); Nucleic Acids Research, 10, 6487 (1982); Proc. Natl. Acad. Sci. USA, 79, 6409 (1982); Gene, 34, 315 (1985); Nucleic Acids Research, 13, 4431 (1985); Proc. Natl. Acad. Sci. USA, 82, 488 (1985), 等。

缺失、取代或添加的氨基酸残基的数目没有特别限制,但是在通过已知方法(例如上面的定点诱变)可以进行缺失、取代或添加的范围内。合适的数目是 1 至数打(dozens)、优选 1 至 20、更优选 1 至 10、再优选 1 至 5。

表述“在 SEQ ID NO: 1 至 8 中的任一个所示的氨基酸序列中缺失、取代或添加了一个或多个氨基酸残基”是指该氨基酸序列可以在 SEQ ID NO: 1 至 8 中的任一个所示的氨基酸序列的任意位置含有一个或多个氨基酸残基的缺失、取代或添加。

可以取代的氨基酸残基是,例如,当使用已知的比对软件对比序列时,不是在 SEQ ID NO: 1 至 8 所示的所有氨基酸序列中都保守的那些。已知的比对软件的实例是包含在基因分析软件 Genetyx (Software

Development Co., Ltd.)中的比对分析软件。关于分析软件的分析参数，可以使用默认值。

可以含有氨基酸残基的缺失或添加，例如，在 SEQ ID NO: 1 至 8 中的任一个所示的氨基酸序列的 N-末端区或 C-末端区。

缺失、取代或添加可以同时包含在一个序列中，要取代或添加的氨基酸可以是天然的或非天然的。天然氨基酸的实例是 L-丙氨酸，L-天冬酰胺，L-天门冬氨酸，L-谷氨酰胺，L-谷氨酸，甘氨酸，L-组氨酸，L-异亮氨酸，L-亮氨酸，L-赖氨酸，L-甲硫氨酸，L-苯丙氨酸，L-脯氨酸，L-丝氨酸，L-苏氨酸，L-色氨酸，L-酪氨酸，L-缬氨酸和 L-半胱氨酸。

下面是能够相互取代的氨基酸的实例。在同一组中的氨基酸可以相互取代。

组 A: 亮氨酸，异亮氨酸，正亮氨酸，缬氨酸，正缬氨酸，丙氨酸，2-氨基丁酸，甲硫氨酸，O-甲基丝氨酸，叔丁基甘氨酸，叔丁基丙氨酸，环己基丙氨酸

组 B: 天门冬氨酸，谷氨酸，异天门冬氨酸，异谷氨酸，2-氨基己二酸，2-氨基辛二酸

组 C: 天冬酰胺，谷氨酰胺

组 D: 赖氨酸，精氨酸，鸟氨酸，2,4-二氨基丁酸，2,3-二氨基丙酸

组 E: 脯氨酸，3-羟基脯氨酸，4-羟基脯氨酸

组 F: 丝氨酸，苏氨酸，高丝氨酸

组 G: 苯丙氨酸，酪氨酸

为了使本发明的蛋白可以具有合成式 (I) 所示的二肽的活性，理想的是，其氨基酸序列与 SEQ ID NO: 1 至 8 中的任一个 (优选 SEQ ID NO: 1) 所示的氨基酸序列的相似性是 65%或更高、优选 75%或更高、更优选 85%或更高、再优选 90%或更高、特别优选 95%或更高、和最优选 98%或更高。

可以使用 Karlin 和 Altschul 的算法 BLAST [Proc. Natl. Acad.

Sci. USA, 90, 5873 (1993)]和 FASTA [Methods Enzymol., 183, 63 (1990)]检测氨基酸序列和核苷酸序列中的相似性。在算法 BLAST 的基础上,已经开发了 BLASTN 和 BLASTX 等程序[J. Mol. Biol., 215, 403 (1990)]。当在 BLAST 的基础上使用 BLASTN 分析核苷酸序列时,各参数例如如下:分值=100 和字长=12。当在 BLAST 的基础上使用 BLASTX 分析氨基酸序列时,各参数例如如下:分值=50 和字长=3。当使用 BLAST 和 Gapped BLAST 程序时,使用每个程序的默认参数。关于这些分析的具体技术是已知的(<http://www.ncbi.nlm.nih.gov>)。

本发明的蛋白还包括与 SEQ ID NO: 1 至 8 中的任一个、优选 SEQ ID NO: 1 所示的氨基酸序列具有 65%或更高、优选 75%或更高、更优选 85%或更高、再优选 90%或更高、特别优选 95%或更高、最优选 98%或更多的相似性的由氨基酸序列组成的蛋白,且其具有合成式(I)所示的二肽的活性(条件是不包括由 SEQ ID NO: 1 所示的氨基酸序列组成的蛋白)。如上所述,可以使用 BLAST 或 FASTA 检测氨基酸序列中的相似性。

SEQ ID NO: 17 所示的氨基酸序列是在具有 SEQ ID NO: 1 至 7 所示的氨基酸序列的蛋白中保守的区域,也是与源自各种微生物的具有 Ala-Ala 连接酶活性的蛋白的共同序列相对应的区域。

本发明的蛋白还包括这样的蛋白,其氨基酸序列与 SEQ ID NO: 17 所示的氨基酸序列具有 80%或更高、优选 90%或更高、再优选 95%或更多的相似性,且其具有合成式(I)所示的二肽的活性(条件是不包括由 SEQ ID NO: 1 所示的氨基酸序列组成的蛋白)。

为了使其氨基酸序列与 SEQ ID NO: 17 所示的氨基酸序列具有 80%或更高、优选 90%或更高、再优选 95%或更多的相似性的蛋白可以具有合成式(I)所示的二肽的活性,理想的是,其氨基酸序列与 SEQ ID NO: 1 至 8 中的任一个所示的氨基酸序列的相似性是至少 80%或更高、通常 90%或更高和特别地 95%或更高。

如上所述,可以使用 BLAST 或 FASTA 检测氨基酸序列中的相似性。可以证实本发明的蛋白是具有合成上面的式(I)所示的二肽的活

性的蛋白，例如，以下面的方式。也就是说，通过重组 DNA 技术制备能表达本发明的蛋白的转化体，使用该转化体生产本发明的蛋白，然后使本发明的蛋白、至少 2 个可以相同或不同的氨基酸(条件是，L-丙氨酸与另一种 L-氨基酸组合使用)和 ATP 存在于水性介质中，然后进行 HPLC 分析等，以确定是否在水性介质中形成和积累了上面的式 (I) 所示的二肽。

本发明的 DNA 包括：

[5] DNA，其能编码根据上面的 [1] 至 [4] 中的任一项的本发明的蛋白；

[6] DNA，其含有 SEQ ID NO: 10 至 16 和 36 中的任一个所示的核苷酸序列；

[7] DNA，其能在严格条件下与含有 SEQ ID NO: 9 至 16 和 36 中的任一个所示的核苷酸序列的互补链的 DNA 杂交，且其能编码具有合成式 (I) 所示的二肽的活性的蛋白，条件是不包括由 SEQ ID NO: 9 所示的核苷酸序列组成的 DNA，优选地条件是不包括能编码由 SEQ ID NO: 1 所示的氨基酸序列组成的蛋白的 DNA；和

[8] DNA，其含有与 SEQ ID NO: 18 所示的核苷酸序列具有 80% 或更多的相似性的核苷酸序列，且能编码具有合成式 (I) 所示的二肽的活性的蛋白，条件是不包括由 SEQ ID NO: 9 所示的核苷酸序列组成的 DNA，优选地条件是不包括能编码由 SEQ ID NO: 1 所示的氨基酸序列组成的蛋白的 DNA。

可以用在生产本发明的式 (I) 所示的二肽的方法中的 DNA 包括根据上面 [5] 至 [8] 的 DNA 和由 SEQ ID NO: 9 所示的核苷酸序列组成的 DNA。

上面的能在严格条件下杂交的 DNA 是指，例如，使用部分或全部具有 SEQ ID NO: 9 至 16 和 36 中的任一个所示的核苷酸序列的 DNA 作为探针，通过菌落杂交、噬菌斑杂交、DNA 印迹杂交等得到的 DNA。这种 DNA 的具体实例是可以通过下述方法鉴别出的 DNA：使用其上面固定有菌落-或噬菌斑-衍生的 DNA 的滤膜，在存在 0.7 至 1.0 mol/l

氯化钠的情况下, 在 65°C 进行杂交, 然后在 65°C 用 0.1 至 2-倍浓度的 SSC 溶液(1 倍浓度的 SSC 溶液: 150 mmol/l 氯化钠和 15 mmol/l 柠檬酸钠) 洗涤滤膜。可以根据 Molecular Cloning, 第 2 版; Current Protocols in Molecular Biology; DNA Cloning 1: Core Techniques, A Practical Approach, 第 2 版, Oxford University (1995) 等所述的方法进行杂交。更具体地, 可杂交的 DNA 包括与 SEQ ID NO: 9 至 16 和 36 中的任一个所示的核苷酸序列具有至少 75% 或更多的相似性、优选 85% 或更多的相似性、再优选 90% 或更多的相似性、特别优选 95% 或更多的相似性的 DNA, 如通过使用上述的 BLAST 或 FASTA、在上述参数的基础上计算的。

可以证实能在严格条件下与具有 SEQ ID NO: 9 至 16 和 36 中的任一个所示的核苷酸序列的 DNA 杂交的 DNA 是能编码具有合成式 (I) 所示的二肽的活性的蛋白的 DNA, 例如通过重组 DNA 技术生产该 DNA 编码的蛋白, 并如上所述检测该蛋白的活性。

(i) 本发明的 DNA 和用在生产本发明的蛋白或二肽的方法中的 DNA 的制备

可以得到本发明的 DNA 和用在生产本发明的蛋白或二肽的方法(在下文中, 也称作本发明的生产方法)中的 DNA, 例如, 通过使用在 SEQ ID NO: 9 至 16 中的任一个所示的核苷酸序列的基础上设计的探针、对来自芽孢杆菌属微生物的染色体 DNA 文库进行 DNA 杂交, 或通过使用在 SEQ ID NO: 9 至 16 中的任一个所示的核苷酸序列的基础上设计的引物 DNA、以芽孢杆菌属微生物的染色体 DNA 作为模板进行 PCR [PCR Protocols, Academic Press (1990)]。

通过在各种基因序列数据库中搜索与能编码 SEQ ID NO: 1 至 8 和 17 中的任一个所示的氨基酸序列的 DNA 的核苷酸序列具有 75% 或更多的相似性、优选 85% 或更多的相似性、更优选 90% 或更多的相似性、再优选 95% 或更多的相似性、特别优选 98% 或更多的相似性的序列, 并在搜索得到的核苷酸序列的基础上, 根据上述方法从具有该核苷酸序列的生物的染色体 DNA 或 cDNA 文库得到理想的 DNA, 也可以得到本发

明的 DNA 和用在本发明的生产方法中的 DNA。

用常规方法将得到的 DNA 以原样或在用合适的限制酶剪切后插入到载体中，并将得到的重组 DNA 导入宿主细胞中。然后，可以通过常规测序方法例如双脱氧法 [Proc. Natl. Acad. Sci., USA, 74, 5463 (1977)]，或通过使用核苷酸测序仪例如 373A DNA 测序仪 (Perkin-Elmer Corp.) 来检测 DNA 的核苷酸序列。

当通过分析核苷酸序列发现得到的 DNA 是部分 DNA 时，可以使用该部分 DNA 作为探针、通过染色体 DNA 文库的 DNA 杂交得到全长 DNA。

在检测的 DNA 的核苷酸序列的基础上，还可以使用 DNA 合成仪 (例如 Model 8905, PerSeptive Biosystems) 化学合成需要的 DNA。

可以通过上述方法获得的 DNA 的实例是具有 SEQ ID NO: 9 至 16 所示的核苷酸序列的 DNA。

用于插入本发明的 DNA 或用在本发明的生产方法中的 DNA 的载体的实例包括 pBluescriptII KS(+) (Stratagene), pDIRECT [Nucleic Acids Res., 18, 6069 (1990)], pCR-Script Amp SK(+) (Stratagene), pT7 Blue (Novagen, Inc.), pCR II (Invitrogen Corp.) 和 pCR-TRAP (Genhunter Corp.)。

可以使用埃希氏菌属等的微生物作为宿主细胞。埃希氏菌属的微生物的实例包括大肠杆菌 (*Escherichia coli*) XL1-Blue, 大肠杆菌 XL2-Blue, 大肠杆菌 DH1, 大肠杆菌 MC1000, 大肠杆菌 ATCC 12435, 大肠杆菌 W1485, 大肠杆菌 JM109, 大肠杆菌 HB101, 大肠杆菌 No. 49, 大肠杆菌 W3110, 大肠杆菌 NY49, 大肠杆菌 MP347, 大肠杆菌 NM522 和大肠杆菌 ME8415。

可以通过能将 DNA 导入上述宿主细胞的任何方法进行重组 DNA 的导入，例如，使用钙离子的方法 [Proc. Natl. Acad. Sci. USA, 69, 2110 (1972)]，原生质体方法 (日本公开的未审查的专利申请号 248394/88) 和电穿孔 [Nucleic Acids Res., 16, 6127 (1988)]。

通过上述方法得到的携带着用在本发明的生产方法中的 DNA 的微生物的实例是大肠杆菌 NM522/pPE43，它是用含有具有 SEQ ID NO: 1

所示的序列的 DNA 的重组 DNA 转化的微生物。

(ii) 生产本发明的蛋白的方法

使用 Molecular Cloning, 第 2 版, Current Protocols in Molecular Biology 等所述的方法, 例如以下述方式, 在宿主细胞中表达通过在上面 (i) 所述的方法得到的本发明的 DNA 或用在本发明的生产方法中的 DNA, 从而可以生产本发明的蛋白。

在本发明的 DNA 或用在本发明的生产方法中的 DNA 的基础上, 根据需要制备了含有能编码本发明的蛋白的区域的合适长度的 DNA 片段。通过替换蛋白编码区的核苷酸序列中的核苷酸, 使密码子最适于在宿主细胞中表达, 可以增强蛋白的生产力。含有优化了密码子的核酸序列的 DNA 会被认为是现在公开的发明的一部分。

可以将 DNA 片段插入合适表达载体中的启动子的下游, 以制备重组 DNA。

通过将重组 DNA 导入适合于表达载体的宿主细胞, 可以得到能生产本发明的蛋白的转化体。

可以使用能表达目标基因的任何细菌细胞、酵母细胞、动物细胞、昆虫细胞、植物细胞等作为宿主细胞。

可以使用的表达载体是能自生复制或整合进上面的宿主细胞中的染色体、且在适用于转录本发明的 DNA 或转录用在本发明的生产方法中的 DNA 的位置含有启动子的那些。

当使用原核生物例如细菌作为宿主细胞时, 优选地, 含有本发明的 DNA 或用在本发明的生产方法中的 DNA 的重组 DNA 是能在原核生物中自生复制的重组 DNA, 和含有例如启动子、核糖体结合序列、本发明的 DNA 或用在本发明的生产方法中的 DNA 和转录终止序列的重组 DNA。重组 DNA 还可以含有能调控启动子的基因。

合适的表达载体的实例是 pBTrp2、pBTac1 和 pBTac2 (Boehringer Mannheim GmbH 的产品), pHelix1 (Roche Diagnostics Corp.), pKK233-2 (Amersham Pharmacia Biotech), pSE280 (Invitrogen Corp.), pGEMEX-1 (Promega Corp.), pQE-8 (Qiagen, Inc.), pET-3

(Novagen, Inc.), pKYP10 (日本公开的未审查的专利申请号 110600/83), pKYP200 [Agric. Biol. Chem., 48, 669 (1984)], pLSA1 [Agric. Biol. Chem., 53, 277 (1989)], pGEL1 [Proc. Natl. Acad. Sci. USA, 82, 4306 (1985)], pBluescript II SK(+), pBluescript II KS(-) (Stratagene), pTrs30 [从大肠杆菌 JM109/pTrS30 (FERM BP-5407) 制备], pTrs32 [从大肠杆菌 JM109/pTrS32 (FERM BP-5408) 制备], pPAC31 (W098/12343), pUC19 [Gene, 33, 103 (1985)], pSTV28 (Takara Shuzo Co., Ltd.), pUC118 (Takara Shuzo Co., Ltd.) 和 pPA1 (日本公开的未审查的专利申请号 233798/88)。

可以使用能在宿主细胞例如大肠杆菌中发挥作用的任何启动子作为启动子。例如,可以使用源自大肠杆菌或噬菌体的启动子,例如 trp 启动子 (P_{trp}), lac 启动子 (P_{lac}), P_L 启动子、 P_R 启动子和 P_{SE} 启动子、 $SP01$ 启动子、 $SP02$ 启动子和 $penP$ 启动子。还可以使用人工设计的和改进的启动子,例如其中 2 个 P_{trp} 串连组合的启动子、tac 启动子、 $lacT7$ 启动子和 $letI$ 启动子等。

还有用的是,用于在芽孢杆菌属微生物中表达的 $xylA$ 启动子 [Appl. Microbiol. Biotechnol., 35, 594-599 (1991)]和用于在棒状杆菌属微生物中表达的 P_{54-6} 启动子 [Appl. Microbiol. Biotechnol., 53, 674-679 (2000)]。

优选地,使用 SD 序列(核糖体结合序列)和起始密码子之间的距离调节至合适长度(例如 6 至 18 核苷酸)的质粒。

在其中本发明的 DNA 或用在本发明的生产方法中的 DNA 连接到表达载体的重组 DNA 中,转录终止序列不是重要的,但是优选地使转录终止序列紧挨结构基因的下游。

这样的重组 DNA 的实例是 pPE43。

合适的原核生物的实例包括下面种属的微生物:埃希氏菌属,沙雷氏菌属,芽孢杆菌属,短杆菌属,棒状杆菌属,微杆菌属,假单胞菌属,土壤杆菌属,脂环酸杆菌属,鱼腥蓝细胞属,组囊蓝细菌属,节杆菌属,固氮菌属,着色菌属,欧文氏菌属,甲基杆菌属,席蓝细

胞属, 红细菌属, 红假单孢菌属, 红螺菌属, 栅藻属 (*Scenedesmus*), 链霉菌属, 聚球蓝细菌属和发酵单胞菌属。具体的实例是大肠杆菌 XL1-Blue, 大肠杆菌 XL2-Blue, 大肠杆菌 DH1, 大肠杆菌 DH5 α , 大肠杆菌 MC1000, 大肠杆菌 KY3276, 大肠杆菌 W1485, 大肠杆菌 JM109, 大肠杆菌 HB101, 大肠杆菌 No. 49, 大肠杆菌 W3110, 大肠杆菌 NY49, 大肠杆菌 MP347, 大肠杆菌 NM522, 杆草芽孢杆菌 ATCC 33712, 巨大芽孢杆菌, 芽孢杆菌属 FERM BP-6030, 解淀粉芽孢杆菌, 凝结芽孢杆菌, 地衣芽胞杆菌, 短小芽孢杆菌, 产氨短杆菌 (*Brevibacterium ammoniagenes*), *Brevibacterium immariophilum* ATCC 14068, 解糖短杆菌 (*Brevibacterium saccharolyticum*) ATCC 14066, 黄色短杆菌 (*Brevibacterium flavum*) ATCC 14067, 谷氨酸棒杆菌 (*Brevibacterium lactofermentum*) ATCC 13869, 谷氨酸棒杆菌 (*Corynebacterium glutamicum*) ATCC 13032, 谷氨酸棒杆菌 (*Corynebacterium glutamicum*) ATCC 14297, 嗜乙酰乙酸棒杆菌 (*Corynebacterium acetoacidophilum*) ATCC 13870, 嗜氨微杆菌 (*Microbacterium ammoniophilum*) ATCC 15354, 无花果沙雷氏菌 (*Serratia ficaria*), 居泉沙雷氏菌 (*Serratia fonticola*), 液化沙雷氏菌 (*Serratia liquefaciens*), 粘质沙雷氏菌 (*Serratia marcescens*), 假单胞菌属 (*Pseudomonas* sp.) D-0110, 放射形土壤杆菌 (*Agrobacterium radiobacter*), 发根土壤杆菌 (*Agrobacterium rhizogenes*), 悬钩子土壤杆菌 (*Agrobacterium rubi*), 柱胞鱼腥蓝细菌 (*Anabaena cylindrica*), 桶形鱼腥蓝细菌 (*Anabaena doliolum*), 水华鱼腥蓝细菌 (*Anabaena flos-aquae*), 金黄节杆菌 (*Arthrobacter aurescens*), 柠檬节杆菌 (*Arthrobacter citreus*), 球形节杆菌 (*Arthrobacter globiformis*), *Arthrobacter hydrocarboglutamicus*, 迈索尔节杆菌 (*Arthrobacter mysorens*), 烟草节杆菌 (*Arthrobacter nicotianae*), 石蜡节杆菌 (*Arthrobacter paraffineus*), 原玻璃蝇节杆菌 (*Arthrobacter protophormiae*), 玫瑰色石蜡节杆菌 (*Arthrobacter roseoparaffinus*), 硫磺节杆菌

(*Arthrobacter sulfureus*), 产豚节杆菌 (*Arthrobacter ureafaciens*), 巴氏着色菌 (*Chromatium buderi*), 微温着色菌 (*Chromatium tepidum*), 酒色着色菌 (*Chromatium vinosum*), 沃氏着色菌 (*Chromatium warmingii*), *Chromatium fluviatile*, 噬夏孢欧文氏菌 (*Erwinia uredovora*), 胡萝卜软腐欧文氏菌 (*Erwinia carotovora*), 菠萝泛菌 (*Erwinia ananas*), 草生欧文氏菌 (*Erwinia herbicola*), *Erwinia punctata*, *Erwinia terreus*, 罗得西亚甲基杆菌 (*Methylobacterium rhodesianum*), 扭脱甲基杆菌 (*Methylobacterium extorquens*), 席蓝细菌属 (*Phormidium* sp.) ATCC 29409, 荚膜红细菌 (*Rhodobacter capsulatus*), *Rhodobacter sphaeroides*, *Rhodopseudomonas blastica*, 海红菌 (*Rhodopseudomonas marina*), 血色红假单胞菌 (*Rhodopseudomonas palustris*), 深红红螺菌 (*Rhodospirillum rubrum*), 需盐红螺菌 (*Rhodospirillum salexigens*), 盐场红螺菌 (*Rhodospirillum salinarum*), 产二素链霉菌 (*Streptomyces ambofaciens*), 金霉素链霉菌 (*Streptomyces aureofaciens*), 金色链霉菌 (*Streptomyces aureus*), 杀真菌素链霉菌 (*Streptomyces fungicidicus*), 灰产色链霉菌 (*Streptomyces griseochromogenes*), 灰色链霉菌 (*Streptomyces griseus*), 浅青紫链霉菌 (*Streptomyces lividans*), 橄榄烬灰链霉菌 (*Streptomyces olivogriseus*), 枝链霉菌 (*Streptomyces rameus*), 田无链霉菌 (*Streptomyces tanashiensis*), 酒红链霉菌 (*Streptomyces vinaceus*) 和运动发酵单胞菌 (*Zymomonas mobilis*)。

可以通过能将 DNA 导入上述宿主细胞的任何方法进行重组 DNA 的导入, 例如, 使用钙离子的方法 [Proc. Natl. Acad. Sci. USA, 69, 2110 (1972)], 原生质体方法 (日本公开的未审查的专利申请号 248394/88) 和电穿孔 [Nucleic Acids Res., 16, 6127 (1988)]。

当用酵母菌株作为宿主细胞时, 可以使用 YEp13 (ATCC 37115), YEp24 (ATCC 37051), YCp50 (ATCC 37419), pHS19, pHS15 等作为表

达载体。

可以使用能在酵母菌株中发挥作用的任何启动子作为启动子。合适的启动子包括 PH05 启动子、PGK 启动子、GAP 启动子、ADH 启动子、gal 1 启动子、gal 10 启动子、热休克多肽启动子、MF α 1 启动子和 CUP 1 启动子。

合适的宿主细胞的实例是下列种属的酵母菌株：糖酵母属，裂殖糖酵母属，克鲁维氏酵母属，丝孢酵母属，许旺氏酵母属，毕赤酵母属和假丝酵母属，更具体地，啤酒糖酵母 (*Saccharomyces cerevisiae*)，粟酒裂殖糖酵母 (*Schizosaccharomyces pombe*)，乳克鲁维氏酵母 (*Kluyveromyces lactis*)，茁芽丝孢酵母 (*Trichosporon pullulans*)，河岸许旺氏酵母 (*Schwanniomyces alluvius*)，巴斯德毕赤氏酵母 (*Pichia pastoris*) 和产朊假丝酵母 (*Candida utilis*)。

可以通过能将 DNA 导入酵母的任何方法进行重组 DNA 的导入，例如，电穿孔 [Methods Enzymol., 194, 182 (1990)]，原生质球方法 [Proc. Natl. Acad. Sci. USA, 81, 4889 (1984)] 和乙酸锂方法 [J. Bacteriol., 153, 163 (1983)]。

当用动物细胞作为宿主细胞时，可以使用 pCDNA1，pCDM8 (可以从 Funakoshi Co., Ltd. 商业获得)，pAGE107 (日本公开的未审查的专利申请号 22979/91)，pAS3-3 (日本公开的未审查的专利申请号 227075/90)，pCDM8 [Nature, 329, 840 (1987)]，pCDNA1/Amp (Invitrogen Corp.)，pREP4 (Invitrogen Corp.)，pAGE103 [J. Biochem., 101, 1307 (1987)]，pAGE210，pAMo，pAMoA，等作为表达载体。

可以使用能在动物细胞中发挥作用的任何启动子作为启动子。合适的启动子包括：巨细胞病毒 (CMV) 的 IE (立即早期) 基因的启动子、SV40 早期启动子、金属硫蛋白启动子、逆转录病毒的启动子、热休克启动子、SR α 启动子等。人 CMV 的 IE 基因的增强子可以与启动子组合使用。

合适的宿主细胞的实例是小鼠骨髓瘤细胞，大鼠骨髓瘤细胞，小鼠杂交瘤细胞，人源的 Namalwa 细胞和 Namalwa KJM-1 细胞，人胚肾细胞，人白血病细胞，非洲绿猴肾细胞，中国仓鼠源的 CHO 细胞，和 HBT5637（日本公开的未审查的专利申请号 299/88）。

小鼠骨髓瘤细胞包括 SP2/0 和 NS0；大鼠骨髓瘤细胞包括 YB2/0；人胚肾细胞包括 HEK293 (ATCC CRL-1573)；人白血病细胞包括 BALL-1；非洲绿猴肾细胞包括 COS-1 和 COS-7。

可以通过能将 DNA 导入动物细胞的任何方法进行重组 DNA 的导入，例如，电穿孔 [Cytotechnology, 3, 133 (1990)]，磷酸钙方法 (日本公开的未审查的专利申请号 227075/90)，脂转染 [Proc. Natl. Acad. Sci. USA, 84, 7413 (1987)]，和 Virology, 52, 456 (1973) 中描述的方法。

当用昆虫细胞作为宿主细胞时，可以使用 Baculovirus Expression Vectors, A Laboratory Manual, W. H. Freeman and Company, New York (1992)；Current Protocols in Molecular Biology；Molecular Biology, A Laboratory Manual；Bio/Technology, 6, 47 (1988) 等所述的方法生产蛋白。

也就是说，将重组基因转移载体和杆状病毒共转染进昆虫细胞中，在昆虫细胞的培养上清液中得到重组病毒，然后用重组病毒感染昆虫细胞，由此可以生产蛋白。

可以用于该方法的基因转移载体包括 pVL1392, pVL1393 和 pBlueBacIII (Invitrogen Corp. 的产品)。

杆状病毒的实例是苜蓿银纹夜蛾核型多角体病毒，它是一种能感染 Barathra 家族的昆虫的病毒。

昆虫细胞的实例是草地夜蛾 (*Spodoptera frugiperda*) 的卵巢细胞、银纹夜蛾 (*Trichoplusia ni*) 的卵巢细胞和培养的源自蚕卵巢的细胞。

草地夜蛾的卵巢细胞包括 Sf9 和 Sf21 (Baculovirus Expression Vectors, A Laboratory Manual)；银纹夜蛾的卵巢细胞包括 High 5

和 BTI-TN-5B1-4 (Invitrogen Corp.); 培养的源自蚕卵巢的细胞包括家蚕 (*Bombyx mori*) N4。

通过磷酸钙方法(日本公开的未审查的专利申请号 227075/90)、脂转染[Proc. Natl. Acad. Sci. USA, 84, 7413 (1987)]等,可以将上面的重组基因转移载体和上面的杆状病毒共转染进用于制备重组病毒的昆虫细胞中。

当用植物细胞作为宿主细胞时,可以使用 Ti 质粒、烟草花叶病毒载体等作为表达载体。

可以使用能在植物细胞中发挥作用的任何启动子作为启动子。合适的启动子包括花椰菜花叶病毒(CaMV)的 35S 启动子、稻肌动蛋白(rice actin) 1 启动子等。

合适的宿主细胞的实例是烟草、马铃薯、番茄、胡萝卜、大豆、油菜、苜蓿、稻、小麦和大麦等植物的细胞。

可以通过能将 DNA 导入植物细胞的任何方法进行重组载体的导入,例如,使用土壤杆菌的方法(日本公开的未审查的专利申请号 140885/84 和 70080/85, W094/00977),电穿孔(日本公开的未审查的专利申请号 251887/85)和使用微粒枪(基因枪)的方法(日本专利号 2606856 和 2517813)。

当在酵母、动物细胞、昆虫细胞或植物细胞中表达 DNA 时,可以得到糖基化的蛋白。

通过在培养基中培养如上得到的转化体、使本发明的蛋白在培养基中形成和积累、并从培养基中回收蛋白,可以生产出本发明的蛋白。

用于生产本发明的蛋白的上述转化体的宿主可以是任意的细菌、酵母、动物细胞、昆虫细胞、植物细胞等,但是优选细菌、更优选埃希氏菌属的微生物、再优选属于大肠杆菌的微生物。

通过用于培养宿主的常规方法,可以在培养基中培养上述的转化体。

为了培养使用原核生物(例如大肠杆菌)或真核生物(例如酵母)作为宿主得到的转化体,可以使用任何天然培养基和合成培养基,只

要它是适用于有效培养该转化体的培养基，它含有能被使用的宿主同化的碳源、氮源、无机盐等。

关于碳源，可以使用能被宿主同化的任何碳源。合适的碳源的实例包括碳水化合物，例如葡萄糖、果糖、蔗糖、含有它们的糖蜜、淀粉和淀粉水解产物；有机酸，例如醋酸和丙酸；和醇类，例如乙醇和丙醇。

关于氮源，可以使用氨、有机酸或无机酸的铵盐，例如氯化铵、硫酸铵、醋酸铵和磷酸铵，和其它的含氮化合物，以及蛋白胨、肉膏、酵母提取物、玉米浆、酪蛋白水解产物、豆饼、豆饼水解产物和各种发酵的微生物细胞和其消化的产物。

无机盐的实例包括磷酸二氢钾、磷酸氢二钾、磷酸镁、硫酸镁、氯化钠、硫酸亚铁、硫酸锰、硫酸铜和碳酸钙。

通常在需氧条件下进行培养，例如通过振荡培养或在通气下的浸没旋转培养。培养温度优选地为 15 至 40°C，培养周期通常是 5 小时至 7 天。在培养过程中，将 pH 维持在 3.0 至 9.0。使用有机酸或无机酸、碱溶液、脲、碳酸钙、氨等进行 pH 调节。

如果需要，在培养过程中还可以向培养基中加入抗生素，例如氨基青霉素和四环素。

在培养用含有可诱导的启动子的表达载体转化的微生物时，如果需要，还可以向培养基中加入诱导物。例如，在培养用含有 lac 启动子的表达载体转化的微生物的情况下，可以向培养基中加入异丙基-β-D-硫代吡喃型半乳糖苷等；在培养用含有 trp 启动子的表达载体转化的微生物的情况下，可以向培养基中加入吲哚丙烯酸等。

为了培养使用动物细胞作为宿主细胞得到的转化体，可以使用常用的培养基，例如 RPMI1640 培养基 [J. Am. Med. Assoc., 199, 519 (1967)], Eagle's MEM [Science, 122, 501 (1952)], DMEM [Virology, 8, 396 (1959)] 和 199 培养基 [Proc. Soc. Biol. Med., 73, 1 (1950)], 通过向这些培养基中添加胎牛血清制备的培养基等。

通常在 pH 6 至 8、在 25 至 40°C、在存在 5% CO₂ 的条件下培养 1

至 7 天。

如果需要，在培养过程中可以向培养基中加入抗生素，例如卡那霉素、青霉素和链霉素。

为了培养使用昆虫细胞作为宿主细胞得到的转化体，可以使用常用的培养基，例如 TNM-FH 培养基 (PharMingen, Inc.)，Sf-900 II SFM 培养基 (Life Technologies, Inc.)，ExCell 400 和 ExCell 405 (JRH Biosciences, Inc.) 和 Grace's 昆虫培养基 [Nature, 195, 788 (1962)]。

通常在 pH 6 至 7、在 25 至 30°C 培养 1 至 5 天。

如果需要，在培养过程中可以向培养基中加入抗生素，例如庆大霉素。

使用植物细胞作为宿主细胞得到的转化体可以以原来的细胞形式培养，或者在分化成植物细胞或植物器官后再培养。为了培养这样的转化体，可以使用常用的培养基，例如 Murashige-Skoog (MS) 培养基和 White 培养基，通过向这些培养基中加入植物激素制备的培养基，例如生长素和细胞分裂素。

通常在 pH 5 至 9、在 20 至 40°C 培养 3 至 60 天。

如果需要，在培养过程中可以向培养基中加入抗生素，例如卡那霉素和潮霉素。

如上所述，根据常规培养方法，通过培养源自微生物、昆虫细胞、动物细胞或植物细胞的、且携带通过将本发明的 DNA 或用在本发明的生产方法中的 DNA 连接到表达载体上而制备的重组 DNA 的转化体，使蛋白形成和积累，并从培养基中回收蛋白，从而可以生产出本发明的蛋白。

通过宿主细胞的胞内生产、宿主细胞的胞外分泌或在宿主细胞外膜上的生产，可以生产出本发明的蛋白，并且根据选择的方法，能合适地修饰生产的蛋白的结构。

当本发明的蛋白是在宿主细胞内或在宿主细胞的外膜上生产时，可以通过采用下述方法来促使蛋白分泌到宿主细胞外：Paulson 等的

方法[J. Biol. Chem., 264, 17619 (1989)], Lowe 等的方法[Proc. Natl. Acad. Sci. USA, 86, 8227 (1989); Genes Develop., 4, 1288 (1990)],或日本公开的未审查的专利申请号 336963/93, W094/23021 等描述的方法。

也就是说,通过重组 DNA 技术将信号肽添加到了含有本发明的蛋白的活性位点的蛋白的上游,通过以这样的蛋白形式来生产它,可以使宿主细胞细胞外地分泌本发明的蛋白。

根据日本公开的未审查的专利申请号 227075/90 所述的方法,利用使用二氢叶酸还原酶基因等的基因放大系统,还可以提高蛋白的生产。

而且,可以使用通过再分化含有导入的基因的动物细胞或植物细胞生成的含有导入的基因的动物(非人类的转基因动物)或含有导入的基因的植物(转基因植物)来生产本发明的蛋白。

当能生产本发明的蛋白的转化体是动物或植物时,可以如下生产蛋白:以通常方式饲养或培养动物或植物,使蛋白在其中形成和积累,并从该动物或植物回收蛋白。

可以使用动物生产本发明的蛋白,例如,根据已知的方法[Am. J. Clin. Nutr., 63, 639S (1996); Am. J. Clin. Nutr., 63, 627S (1996); Bio/Technology, 9, 830 (1991)],在通过导入基因构建的动物中生产蛋白。

在动物的情况下,可以生产本发明的蛋白,例如通过饲养携带导入的本发明的 DNA 或用在本发明的生产方法中的 DNA 的非人类的转基因动物,使蛋白在动物中形成和积累,并从该动物回收蛋白。蛋白形成和积累的地方包括动物的奶(日本公开的未审查的专利申请号 309192/88)、蛋等。关于该方法中的启动子,可以使用能在动物中发挥作用的任何启动子。优选的启动子包括乳腺细胞特异性的启动子,例如 α 酪蛋白启动子、 β 酪蛋白启动子、 β 乳球蛋白启动子和乳清酸性蛋白启动子。

可以使用植物生产本发明的蛋白,例如,根据已知的方法[Soshiki

Baiyo (Tissue Culture), 20 (1994); Soshiki Baiyo, 21 (1995); Trends Biotechnol., 15, 45 (1997)], 通过培养携带着导入的能编码本发明的蛋白的 DNA 的转基因植物, 使蛋白在植物中形成和积累, 并从该植物回收蛋白。通过本领域的普通技术人员已知的方法, 也可以将组织特异性的启动子用于植物生产。

通过用于分离和纯化酶的常规方法, 可以分离和纯化使用能生产本发明的蛋白的转化体生产的本发明的蛋白。

例如, 当在细胞中以可溶形式生产本发明的蛋白时, 可以在完成培养后通过离心收集细胞, 并悬浮到水性缓冲液中, 随后使用超声波仪、弗氏压碎器 (French press)、Manton Gaulin 匀浆机, Dynomill 等进行破碎, 以得到无细胞的提取物。

可以如下得到纯化的蛋白制品: 离心无细胞的提取物, 得到上清液, 然后对上清液进行普通的分离和纯化酶的方法, 例如单独地或组合地用溶剂提取, 用硫酸铵等盐析, 脱盐, 用有机溶剂沉淀, 使用树脂进行阴离子交换色谱, 例如二乙基氨基乙基 (DEAE)-Sephacel 和 DIAION HPA-75 (Mitsubishi Chemical Corporation), 使用树脂进行阳离子交换色谱, 例如 S-Sephacel FF (Pharmacia), 使用树脂进行疏水色谱, 例如丁基 Sepharose 和苯基 Sepharose, 使用分子筛进行凝胶过滤, 亲和色谱, 色谱聚焦, 和电泳例如等电聚焦。

当蛋白是作为包涵体在细胞中生产时, 类似地回收和破碎细胞, 然后进行离心, 得到沉淀部分。通过常规方法从沉淀中回收蛋白后, 用蛋白变性剂溶解蛋白的包涵体。

用不含有蛋白变性剂的溶液、或含有低浓度的不会造成蛋白变性的蛋白变性剂的溶液, 稀释或渗析溶解的蛋白溶液, 由此使蛋白复性, 具有正常的高级结构。然后, 通过上述的相同的分离和纯化步骤可以得到纯化的蛋白制品。

当细胞外地分泌本发明的蛋白或其衍生物 (例如糖基化形式) 时, 可以从培养上清液中回收蛋白或其衍生物 (例如糖基化形式)。

也就是说, 通过上述的相同方法 (例如离心) 处理培养物, 得到

可溶的部分。使用上述的相同的分离和纯化方法，可以从可溶的部分得到纯化的蛋白制品。

以上述方式得到的蛋白的实例是分别由 SEQ ID NO: 1 至 8 所示的氨基酸序列组成的蛋白。

还可以作为与另一种蛋白的融合蛋白来生产本发明的蛋白，使用对融合蛋白具有亲合力的物质，通过亲合色谱进行纯化。例如，根据 Lowe 等的方法 [Proc. Natl. Acad. Sci. USA, 86, 8227 (1989); Genes Develop., 4, 1288 (1990)] 和日本公开的未审查的专利申请号 336963/93 和 W094/23021 所述的方法，本发明的蛋白可以作为与蛋白 A 的融合蛋白来生产，且可以通过使用免疫球蛋白 G 的亲合色谱进行纯化。

而且，可以作为与标记 (Flag) 肽的融合蛋白来生产本发明的蛋白，使用抗-标记的抗体，通过亲合色谱进行纯化 [Proc. Natl. Acad. Sci. USA, 86, 8227 (1989); Genes Develop., 4, 1288 (1990)]。还可以使用针对所述蛋白的抗体，通过亲合色谱纯化该蛋白。

基于上面得到的蛋白的氨基酸序列信息，还可以通过化学合成方法来生产本发明的蛋白，例如 Fmoc 方法 (苄基甲氧基羰基方法) 和 tBoc 方法 (叔丁氧基羰基方法)。而且，还可以使用来自 Advanced ChemTech, Perkin-Elmer, Pharmacia, Protein Technology Instrument, Synthecell-Vega, PerSeptive, Shimadzu Corporation 等的肽合成仪，化学合成本发明的蛋白。

(iii) 生产本发明的二肽的方法

(1) 酶方法

生产二肽的酶方法的一个实例是这样的方法，其包括：使本发明的蛋白、至少 2 个可以相同或不同的氨基酸和 ATP 存在于水性介质中；使式 (I) 所示的二肽在介质中形成和积累；和从介质中回收二肽。

在上述方法中用作底物的至少 2 个氨基酸、优选一种或两种氨基酸选自：氨基酸，优选 L-氨基酸，甘氨酸 (Gly) 和 β -丙氨酸 (β

-Ala), 且可以任意组合地使用, 除了使用 L-丙氨酸作为一种氨基酸之外。L-氨基酸的实例是 L-丙氨酸(L-Ala), L-谷氨酰胺(L-Gln), L-谷氨酸(L-Glu), L-缬氨酸(L-Val), L-亮氨酸(L-Leu), L-异亮氨酸(L-Ile), L-脯氨酸(L-Pro), L-苯丙氨酸(L-Phe), L-色氨酸(L-Trp), L-甲硫氨酸(L-Met), L-丝氨酸(L-Ser), L-苏氨酸(L-Thr), L-半胱氨酸(L-Cys), L-天冬酰胺(L-Asn), L-酪氨酸(L-Tyr), L-赖氨酸(L-Lys), L-精氨酸(L-Arg), L-组氨酸(L-His), L-天门冬氨酸(L-Asp), L- α -氨基丁酸(L- α -AB), L-重氮丝氨酸, L-茶氨酸, L-4-羟基脯氨酸(L-4-HYP), L-3-羟基脯氨酸(L-3-HYP), L-鸟氨酸(L-Orn), L-瓜氨酸(L-Cit)和 L-6-重氮-5-氧-正亮氨酸。

在上面方法中更优选地使用的氨基酸包括下面的: 组合中的一种氨基酸选自 L-Ala, Gly, L-Met, L-Ser, L-Thr 和 β -Ala, 且另一种氨基酸选自 L-Ala, L-Gln, L-Glu, Gly, L-Val, L-Leu, L-Ile, L-Pro, L-Phe, L-Trp, L-Met, L-Ser, L-Thr, L-Cys, L-Asn, L-Tyr, L-Lys, L-Arg, L-His, L-Asp, L- α -AB, β -Ala, L-重氮丝氨酸, L-茶氨酸, L-4-HYP, L-3-HYP, L-Orn, L-Cit 和 L-6-重氮-5-氧-正亮氨酸(不包括 L-Ala 和 L-Ala 的组合); L-Gln 和 L-Phe 的组合; L- α -AB 和 L-Gln, L-Arg 或 L- α -AB 的组合。再优选的氨基酸是: L-Ala 和一种选自 L-Gln, Gly, L-Val, L-Leu, L-Ile, L-Phe, L-Trp, L-Met, L-Ser, L-Thr, L-Cys, L-Asn, L-Tyr, L-Lys, L-Arg, L-His, L- α -AB, L-重氮丝氨酸, L-Cit 和 L-茶氨酸的氨基酸的组合; Gly 和一种选自 L-Gln, Gly, L-Phe, L-Trp, L-Met, L-Ser, L-Thr, L-Cys, L-Tyr, L-Lys, L-Arg, L- α -AB 和 L-Cit 的氨基酸的组合; L-Met 和一种选自 L-Phe, L-Met, L-Ser, L-Thr, L-Cys, L-Tyr, L-Lys 和 L-His 的氨基酸的组合; L-Ser 和一种选自 L-Gln, L-Phe, L-Ser, L-Thr, L-Tyr, L-His 和 L- α -AB 的氨基酸的组合; L-Thr 和一种选自 L-Gln, L-Phe, L-Leu, L-Thr 和 L- α -AB 的氨基酸的组合; L-Gln 和 L-Phe 的组合; β -Ala 和一种选自 L-Phe, L-Met, L-His 和 L-Cit 的氨基酸的组合; 以及 L- α -AB 和 L-Gln, L-Arg 或 L- α -AB 的组合。

在上面的方法中，对于每 mg 用作底物的氨基酸，本发明的蛋白的加入量为 0.01 至 100 mg、优选 0.1 至 10 mg。

在上面的方法中，用作底物的氨基酸是在开始时或者在反应过程中加入水性介质中，产生 0.1 至 500 g/l、优选 0.2 至 200 g/l 的浓度。

在上面的方法中，以 0.5 mmol 至 10 mol/l 的终浓度使用用作能源的 ATP。

在上面的方法中使用的水性介质可以含有任意的组分，且可以具有任意的组成，只要不抑制二肽形成反应即可。合适的水性介质包括水和缓冲液，例如磷酸盐缓冲液、碳酸盐缓冲液、醋酸盐缓冲液、硼酸盐缓冲液、柠檬酸盐缓冲液和 Tris 缓冲液。水性介质可以含有醇类例如甲醇和乙醇，酯类例如醋酸乙酯，酮类例如丙酮，和酰胺类例如乙酰胺。

二肽形成反应是在水性介质中、在 pH 5 至 11、优选 pH 6 至 10、在 20 至 50°C、优选 25 至 45°C 进行 2 至 150 小时、优选 6 至 120 小时。

通过上面的方法生产的二肽包含式 (I) 代表的二肽。优选的二肽是式 (I) 代表的那些，其中 R^1 和 R^2 可以相同或不同，每个代表选自 L-Ala, L-Gln, L-Glu, Gly, L-Val, L-Leu, L-Ile, L-Pro, L-Phe, L-Trp, L-Met, L-Ser, L-Thr, L-Cys, L-Asn, L-Tyr, L-Lys, L-Arg, L-His, L-Asp, L- α -AB, β -Ala, L-重氮丝氨酸, L-茶氨酸, L-4-HYP, L-3-HYP, L-Orn, L-Cit 和 L-6-重氮-5-氧-正亮氨酸的氨基酸(不包括其中 R^1 和 R^2 同时是 L-Ala)。更优选的是这样的二肽，其中 R^1 是 L-Ala, Gly, L-Met, L-Ser, L-Thr 或 β -Ala, 且 R^2 是 L-Gln, L-Glu, Gly, L-Val, L-Leu, L-Ile, L-Pro, L-Phe, L-Trp, L-Met, L-Ser, L-Thr, L-Cys, L-Asn, L-Tyr, L-Lys, L-Arg, L-His, L-Asp, L- α -AB, β -Ala, L-重氮丝氨酸, L-茶氨酸, L-4-HYP, L-3-HYP, L-Orn, L-Cit 或 L-6-重氮-5-氧-正亮氨酸。再优选的二肽是：二肽，其中 R^1 是 L-Ala, 且 R^2 是 L-Gln, Gly, L-Val, L-Leu, L-Ile, L-Phe, L-Trp, L-Met,

L-Ser, L-Thr, L-Cys, L-Asn, L-Tyr, L-Lys, L-Arg, L-His, L- α -AB, L-重氮丝氨酸, L-茶氨酸或 L-Cit; 二肽, 其中 R¹是 Gly, 且 R²是 L-Gln, Gly, L-Phe, L-Trp, L-Met, L-Ser, L-Thr, L-Cys, L-Tyr, L-Lys, L-Arg, L- α -AB 或 L-Cit; 二肽, 其中 R¹是 L-Met, 且 R²是 L-Phe, L-Met, L-Cys, L-Tyr, L-Lys 或 L-His; 二肽, 其中 R¹是 L-Ser, 且 R²是 L-Gln, Gly, L-Phe, L-Met, L-Ser, L-Thr, L-Tyr, L-His 或 L- α -AB; 二肽, 其中 R¹是 L-Thr, 且 R²是 L-Gln, L-Leu, L-Phe, L-Met, L-Ser, L-Thr 或 L- α -AB; 二肽, 其中 R¹是 L-Gln, 且 R²是 L-Phe 或 L- α -AB; 二肽, 其中 R¹是 L-Phe, 且 R²是 L-Gln; 二肽, 其中 R¹是 L-Trp, 且 R²是 Gly; 二肽, 其中 R¹是 L-Cys, 且 R²是 L-Ala, L-Gln, Gly 或 L-Met; 二肽, 其中 R¹是 L-Lys, 且 R²是 L-Ala, Gly 或 L-Met; 二肽, 其中 R¹是 β -Ala, 且 R²是 L-Phe, L-Met 或 L-His; 二肽, 其中 R¹是 L-Arg, 且 R²是 L- α -AB; 二肽, 其中 R¹是 L-His, 且 R²是 L-Met; 和二肽, 其中 R¹是 L- α -AB, 且 R²是 L-Ala, L-Gln, Gly, L-Ser, L-Thr, L-Arg 或 L- α -AB.

(2) 使用转化体或微生物的培养物、或处理过的培养物作为酶源的方法

使用转化体或微生物的培养物、或处理过的培养物作为酶源生产二肽的方法的一个实例是这样的方法, 其包括: 使酶源和至少 2 个可以相同或不同的氨基酸存在于水性介质中, 所述的酶源是能生产本发明的蛋白的转化体的培养物、或能生产本发明的蛋白的微生物或处理过的所述培养物; 使式 (I) 所示的二肽在培养基中形成和积累; 和从培养基中回收二肽。

用于上述方法的转化体包括能生产本发明的蛋白的转化体, 其可以通过上面 (ii) 的方法生产。可以使用细菌、酵母、动物细胞、昆虫细胞、植物细胞等作为转化体的宿主。优选的宿主是细菌, 其中更优选埃希氏菌属、芽孢杆菌属和棒状杆菌属的微生物。

优选的埃希氏菌属的微生物包括大肠杆菌; 优选的芽孢杆菌属微

生物包括杆草芽孢杆菌和巨大芽孢杆菌；和优选的棒状杆菌属微生物包括谷氨酸棒杆菌和产氨短杆菌。

在上面方法中使用的微生物可以是能生产本发明的蛋白的任何微生物，但是优选地为芽孢杆菌属微生物，更优选具有芽孢杆菌溶素合成能力的芽孢杆菌属微生物，再优选选自下列种属的微生物：杆草芽孢杆菌，解淀粉芽孢杆菌，凝结芽孢杆菌，地衣芽胞杆菌，巨大芽孢杆菌和短小芽孢杆菌，最优选选自下列的菌株：杆草芽孢杆菌 ATCC 15245，杆草芽孢杆菌 ATCC 6633，杆草芽孢杆菌 IAM 1213，杆草芽孢杆菌 IAM 1107，杆草芽孢杆菌 IAM 1214，杆草芽孢杆菌 ATCC 9466，杆草芽孢杆菌 IAM 1033，杆草芽孢杆菌 ATCC 21555，解淀粉芽孢杆菌 IFO 3022 和短小芽孢杆菌 NRRL B-12025。

处理过的培养物包括浓缩的培养物、干燥的培养物、通过离心培养物得到的细胞、通过提取细胞得到的酶制品或通过各种方法处理细胞得到的产物，例如干燥，冻干，用表面活性剂、超声破碎、机械摩擦处理，用溶剂处理，酶处理，蛋白分级和固定化。

在上面的方法中，用作底物的氨基酸的种类、它们的浓度、它们的添加时间、生产的二肽与在上面 (iii) (1) 所述的酶方法中的类似。

在使用微生物的培养物或处理过的所述培养物作为酶源的方法中，除了在上面 (iii) (1) 所述的酶方法中使用的水性介质外，还可以使用用作酶源的转化体或微生物的培养物作为水性介质。

而且，在上面的方法中，可以根据需要向水性介质中加入 ATP 或能被转化体或微生物代谢产生 ATP 的化合物作为 ATP 源，例如，糖类例如葡萄糖，醇类例如乙醇，和有机酸例如醋酸。

如果需要，还可以向水性介质中加入表面活性剂或有机溶剂。可以使用能促进二肽形成的任何表面活性剂。合适的表面活性剂包括：非离子的表面活性剂例如聚氧乙烯十八烷基胺（例如 Nymeen S-215, NOF Corporation），阳离子表面活性剂例如溴化十六烷基三甲基铵和烷基二甲基苄基氯化铵（例如 Cation F2-40E, NOF Corporation），阳离子表面活性剂例如月桂酰肌氨酸盐，和叔胺例如烷基二甲胺（例如

Tertiary Amine FB, NOF Corporation), 它们可以单独使用或组合使用。表面活性剂通常以 0.1 至 50 g/l 的浓度使用。关于有机溶剂, 通常以 0.1 至 50 ml/l 的浓度使用二甲苯、甲苯、脂族醇、丙酮、醋酸乙酯等。

当使用培养物或处理过的培养物作为酶源时, 根据其具体活性等改变要添加的酶源的量, 但是, 其量例如为 5 至 1000 mg (湿细胞重)、优选 10 至 400 mg /mg 用作底物的氨基酸。

二肽形成反应是在水性介质中、在 pH 5 至 11、优选 pH 6 至 10、在 20 至 65°C、优选 25 至 55°C、更优选 30 至 45°C 进行 1 分钟至 150 小时、优选 3 分钟至 120 小时、更优选 30 分钟至 100 小时。

在上面 (iii) (1) 和 (2) 所述的方法中, 使用活性炭、离子交换树脂等, 或通过有机溶剂提取、结晶、薄层色谱和高效液相色谱等方法, 可以通过普通方法回收在水性介质中形成和积累的二肽,

在下面的实施例中解释了本发明的某些实施方案。这些实施例不用于限制本发明的范围。

实施例 1

使用数据库搜索具有二肽合成活性的蛋白

使用源自杆草芽孢杆菌 168 [Nature, 390, 249-256 (1997)] 的 D-Ala-D-Ala 连接酶基因的氨基酸序列作为查询序列, 使用杆草芽孢杆菌 168 的基因组 DNA 的数据库 Subtilist (<http://genolist.pasteur.fr/SubtiList/>) 的相似性检索功能, 搜索杆草芽孢杆菌 168 的基因组 DNA 序列中存在的编码具有相似性的蛋白的基因。

从作为检索结果得到的序列中, 排除了能编码如 SEQ ID NO: 33、34 和 35 所示的 D-Ala-D-Ala 连接酶基序 [Biochemistry, 30, 1673 (1991)] 的氨基酸序列的基因, 这些基因编码已经清楚其功能的蛋白。在剩余的序列中, 选择了与 D-Ala-D-Ala 连接酶基序具有最高相似性 (29.1%) 的序列作为已知功能的基因 ywfE。

ywfE 的核苷酸序列如 SEQ ID NO: 9 所示, 该核苷酸序列编码的蛋白的氨基酸序列如 SEQ ID NO: 1 所示。

实施例 2

能表达 ywfE 基因的菌株的构建

在实施例 1 得到的核苷酸序列信息的基础上, 以下述方式得到了杆草芽孢杆菌的 ywfE 基因片段。

也就是说, 将杆草芽孢杆菌 168 接种到 LB 培养基中 [10 g/l Bacto-胰蛋白胨 (Difco), 5 g/l 酵母提取物 (Difco) 和 5 g/l 氯化钠], 在 30°C 静置培养过夜。培养后, 根据使用 Current Protocols in Molecular Biology 所述的饱和苯酚的方法, 分离和纯化微生物的染色体 DNA。

使用 DNA 合成仪 (Model 8905, PerSeptive Biosystems, Inc.), 合成了具有 SEQ ID NO: 19 至 22 所示的核苷酸序列的 DNA (以下分别简称为引物 A, 引物 B, 引物 C 和引物 D)。引物 A 具有这样的序列, 其中含有 XhoI 识别序列的核苷酸序列被添加到了含有 ywfE 的起始密码子的杆草芽孢杆菌染色体 DNA 区的 5' 末端。引物 B 具有这样的序列, 其中含有 BamHI 识别序列的核苷酸序列被添加到了与含有 ywfE 的终止密码子的序列互补的核苷酸序列的 5' 末端。引物 C 具有这样的序列, 其中含有 EcoRI 识别序列的核苷酸序列被添加到了含有 trp 启动子 [从大肠杆菌 JM109/pTrS30 (FERM BP-5407) 制备] 的表达载体 pTrS30 的 trp 启动子区的核苷酸序列的 5' 末端区。引物 D 具有这样的序列, 其中含有 XhoI 识别序列的核苷酸序列被添加到了与含有 trp 启动子的表达载体 pTrS30 的 trp 启动子区的序列互补的序列的 5' 末端区。

使用上述的引物 A 和引物 B, 使用杆草芽孢杆菌的染色体 DNA 作为模板, 通过 PCR 扩增了 ywfE 基因片段。使用引物 C 和引物 D, 使用 pTrS30 作为模板, 通过 PCR 扩增了 trp 启动子区片段。PCR 进行了 30 个循环, 一个循环包括: 在 94°C 反应 1 分钟, 在 55°C 反应 2 分钟, 在 72°C 反应 3 分钟, 使用的 40 μ l 反应混合物含有作为模板的 0.1 μ g

染色体 DNA 或 10 ng pTrs30、0.5 $\mu\text{mol}/\text{l}$ 每种引物、2.5 单位的 Pfu DNA 聚合酶 (Stratagene)、4 μl Pfu DNA 聚合酶的缓冲液 (10 x) (Stratagene) 和 200 $\mu\text{mol}/\text{l}$ 每种 dNTP (dATP, dGTP, dCTP 和 dTTP)。

对得到的每种反应混合物的十分之一进行琼脂糖凝胶电泳, 以确信在使用引物 A 和引物 B 的 PCR 中和在使用引物 C 和引物 D 的 PCR 中分别扩增了与 ywfE 基因片段相对应的约 1.4 kb DNA 片段和与 trp 启动子区片段相对应的约 0.3 kb DNA 片段。然后, 将剩余的混合物与等量的以 TE [10 mmol/l Tris-HCl (pH 8.0), 1 mmol/l EDTA] 饱和的苯酚/氯仿 (1 体积/1 体积) 相混合。离心得到的溶液, 将得到的上层与 2 倍体积的冷乙醇混合, 在 -80°C 静置 30 分钟。离心得到的溶液以沉淀 DNA, 将得到的 DNA 溶解到 20 μl TE 中。

对这样得到的溶液 (每次 5 μl) 分别进行下述反应: 使用限制酶 XhoI 和 BamHI 剪切用引物 A 和引物 B 扩增的 DNA, 和使用限制酶 EcoRI 和 XhoI 剪切用引物 C 和引物 D 扩增的 DNA。通过琼脂糖凝胶电泳分离 DNA 片段, 并使用 GENECLAN II 试剂盒 (BIO 101) 分别回收含有 ywfE 的 1.4 kb 片段和含有 trp 启动子区的 0.3 kb 片段。

使用限制酶 EcoRI 和 BamHI 剪切了 pTrs30 [从大肠杆菌 JM109/pTrs30 (FERM BP-5407) 制备的含有 trp 启动子的表达载体, 0.2 μg]。通过琼脂糖凝胶电泳分离 DNA 片段, 以与上面相同的方式回收 4.5 kb DNA 片段。

使用连接试剂盒 (Takara Shuzo Co., Ltd.), 使上面得到的含有 ywfE 的 1.4 kb 片段、含有 trp 启动子区的 0.3 kb 片段和 4.5 kb DNA 片段在 16°C 进行连接反应 16 小时。

根据使用钙离子的方法 [Proc. Natl. Acad. Sci. USA, 69, 2110 (1972)], 使用反应混合物转化了大肠杆菌 NM522 (Stratagene), 涂布在含有 50 $\mu\text{g}/\text{ml}$ 氨苄青霉素的 LB 琼脂培养基上, 在 30°C 培养过夜。

根据已知的方法, 从在该培养基上生长的转化体菌落提取质粒, 使用限制酶分析了质粒的结构, 由此确认得到了表达载体 pPE43, 其含有连接在 trp 启动子下游的 ywfE (图 1)。

实施例 3

二肽的生产

将在实施例 2 得到的携带 pPE43 的大肠杆菌 NM522(大肠杆菌 NM522/pPE43)接种到装在大试管中的含有 50 μ g/ml 氯苄青霉素的 8 ml LB 培养基中, 在 28 $^{\circ}$ C 培养 17 小时。离心得到的培养物, 得到湿细胞。

制备了包含 60 mg/ml (终浓度)湿细胞、120 mmol/l 磷酸钾缓冲液 (pH 7.4)、60 mmol/l 氯化镁、60 mmol/l ATP、30 mmol/l L-Ala、30 mmol/l L-Gln 和 0.4% Nymeen S-215 的反应混合物(0.1 ml), 在 37 $^{\circ}$ C 反应 3 分钟。

反应结束后, 通过二硝基酚方法衍生生化反应产物, 然后通过 HPLC 分析。使用 Lichrosorb-RP-18 柱(Kanto Kagaku)作为分离柱、1% (v/v) 磷酸和 25% (v/v) 乙腈作为洗脱剂, 以 0.7 ml/min 的流速, 进行了 HPLC 分析。结果, 确认在反应混合物中形成和积累了 120 mg/l L-丙氨酰基-L-谷氨酰胺 (L-Ala-L-Gln)。

当使用仅携带载体的对照菌大肠杆菌 NM522/pTrS31 细胞进行反应时, 没有观察到 L-Ala-L-Gln 的形成。

实施例 4

C-末端 His-标记的重组二肽合成酶的纯化

使用上面的 DNA 合成仪, 合成了具有 SEQ ID NO: 23 和 24 所示的核苷酸序列的 DNA (以下分别简称为引物 E 和引物 F)。引物 E 具有含有这样的区域的核苷酸序列, 其中 ywfE 起始密码子(atg)被取代为 NcoI 识别序列(ccatgg)。引物 F 具有含有这样的区域的核苷酸序列, 其中 ywfE 终止密码子被取代为 BamHI 识别序列(ggatcc)。

使用杆草芽孢杆菌 168(ATCC 23857)的染色体 DNA 作为模板、上面的引物 E 和引物 F 作为一组引物, 进行了 PCR。也就是说, PCR 进行了 30 个循环, 一个循环包括: 在 94 $^{\circ}$ C 反应 1 分钟, 在 55 $^{\circ}$ C 反应 2 分

钟，在 72℃ 反应 3 分钟，使用的 40 μl 反应混合物含有 0.1 μg 染色体 DNA、0.5 μmol/l 每种引物、2.5 单位的 Pfu DNA 聚合酶、4 μl Pfu DNA 聚合酶的缓冲液 (10 x) 和 200 μmol/l 每种 dNTP。

对得到的反应混合物的十分之一进行琼脂糖凝胶电泳，以确信扩增了与 ywfE 片段相对应的约 1.4 kb 片段。然后，将剩余的混合物与等量的以 TE 饱和的苯酚/氯仿相混合。离心得到的溶液，将得到的上层与 2 倍体积的冷乙醇混合，在 -80℃ 静置 30 分钟。离心得到的溶液，将得到的 DNA 沉淀溶解到 20 μl TE 中。

对这样得到的溶液 (5 μl) 使用限制酶 NcoI 和 BamHI 剪切扩增的 DNA。通过琼脂糖凝胶电泳分离 DNA 片段，并使用 GENECLAN II 试剂盒回收含有 ywfE 的 1.4 kb DNA 片段。

使用限制酶 NcoI 和 BamHI 剪切了 C-末端 His-标记的重组表达载体 pQE60 (Qiagen, Inc.) (0.2 μg)。通过琼脂糖凝胶电泳分离 DNA 片段，以与上面相同的方式回收 3.4 kb DNA 片段。

使用连接试剂盒，使上面得到的含有 ywfE 的 1.4 kb DNA 片段和 3.4 kb DNA 片段在 16℃ 连接反应 16 小时。

根据使用钙离子的方法，使用连接反应混合物转化了大肠杆菌 NM522，涂布在含有 50 μg/ml 氨苄青霉素的 LB 琼脂培养基上，在 30℃ 培养过夜。

根据已知的方法，从在该培养基上生长的转化体菌落提取质粒，使用限制酶分析了质粒的结构，由此确认得到了 pQE60ywfE，它是 C-末端 His-标记的 ywfE 表达载体 (图 2)。

将携带 pQE60ywfE 的大肠杆菌 NM522 (大肠杆菌 NM522/pQE60ywfE) 接种到装在大试管中的含有 50 μg/ml 氨苄青霉素的 8 ml LB 培养基中，在 28℃ 培养 17 小时。将得到的培养物接种到装在 250-ml 锥形瓶中的含有 50 μg/ml 氨苄青霉素的 50 ml LB 培养基中，在 30℃ 培养 3 小时。然后，加入异丙基-β-D-硫代吡喃型半乳糖苷 (IPTG)，得到 1 mmol/l 的终浓度，随后进一步在 30℃ 培养 4 小时。离心得到的培养物，以得到湿细胞，使用 HisTrap (His-标记的蛋白纯化试剂盒, Amersham

Pharmacia Biotech), 根据所附的说明书, 从湿细胞纯化 His-标记的重组酶。

实施例 5

使用 His-标记的重组酶 (1) 生产二肽

(i) 制备了含有 0.04 mg 在实施例 4 中得到的纯化的 His-标记的重组酶、100 mmol/l Tris-HCl (pH 8.0)、60 mmol/l 氯化镁、60 mmol/l ATP、30 mmol/l L-Ala 和 30 mmol/l L-Gln 的反应混合物 (0.1 ml), 在 37°C 反应 16 小时。

反应结束后, 与上面实施例 3 相同的方式分析了反应产物, 由此确信在反应混合物中形成和积累了 3.7 g/l L-Ala-L-Gln 和 0.3 g/l L-丙氨酰基-L-丙氨酸 (L-Ala-L-Ala)。

(ii) 使用与上面 (i) 的反应混合物相同组成的反应混合物, 但是使用 0.01 mg 酶、并分别用 L-Phe、L-Met、L-Leu 和 L-Val 代替 L-Gln, 在与上面 (i) 相同的条件下进行了反应。

反应结束后, 与上面实施例 3 相同的方式分析了反应产物, 由此确信在各反应混合物中形成和积累了下面的二肽: 7.0 g/l 单独的 L-丙氨酰基-L-苯丙氨酸 (L-Ala-L-Phe); 7.0 g/l L-丙氨酰基-L-甲硫氨酸 (L-Ala-L-Met) 和 0.03 g/l L-Ala-L-Ala; 5.0 g/l L-丙氨酰基-L-亮氨酸 (L-Ala-L-Leu) 和 0.2 g/l L-Ala-L-Ala; 和 1.6 g/l L-丙氨酰基-L-缬氨酸 (L-Ala-L-Val) 和 0.3 g/l L-Ala-L-Ala。

(iii) 使用与上面 (i) 的反应混合物相同组成的反应混合物, 但是使用 0.01 mg 酶、并用 Gly 代替 L-Ala、用 L-Phe 和 L-Met 分别代替 L-Gln, 在与上面 (i) 相同的条件下进行了反应。

反应结束后, 与上面实施例 3 相同的方式分析了反应产物, 由此确信在各反应混合物中形成和积累了 5.2 g/l 甘氨酰基-L-苯丙氨酸 (Gly-L-Phe) 和 1.1 g/l 甘氨酰基-L-甲硫氨酸 (Gly-L-Met)。

当上面的反应混合物的组成不包括 ATP 时, 没有形成二肽。

上面的结果表明, 在存在 ATP 时, ywfE 基因产物具有生产下述二肽的活性: 从 L-Ala+L-Gln, L-Phe, L-Met, L-Leu 或 L-Val 合成 L-Ala-L-Gln+L-Ala-L-Ala, L-Ala-L-Phe, L-Ala-L-Met+L-Ala-L-Ala, L-Ala-L-Leu+L-Ala-L-Ala, 或 L-Ala-L-Val+L-Ala-L-Ala; 以及从 Gly+L-Phe 或 L-Met 合成 Gly-L-Phe 或 Gly-L-Met。

实施例 6

使用 His-标记的重组酶 (2) 生产二肽

制备了含有 0.04 mg 在实施例 4 中得到的纯化的 His-标记的重组酶、100 mmol/l Tris-HCl (pH 8.0)、60 mmol/l 氯化镁和 60 mmol/l ATP 的反应混合物 (0.1 ml)。向该混合物中分别加入表 1 第一行所示的 L-氨基酸、Gly 或 β -Ala 和表 1 最左一列所示的 L-氨基酸、Gly 或 β -Ala 的组合, 产生 30 mmol/l 每种氨基酸的浓度, 使得到的混合物在 37°C 反应 16 小时。反应结束后, 通过 HPLC 分析了反应产物, 由此确信形成了表 1 所示的二肽。

表 1-1

	Ala	Gln	Glu	Gly	Val	Leu	Ile	Pro
Ala	AlaAla	AlaGln AlaAla	AlaAla	AlaGly AlaAla	AlaVal AlaAla	AlaLeu AlaAla	AlaIle AlaAla	AlaAla
Gln		×	×	GlyGln GlyGly	×	×	×	×
Glu				GlyGly				
Gly				GlyGly				GlyGly
Val								
Leu								
Ile								
Pro								
Phe								
Trp								
Met								
Ser								
Thr								
Cys								
Asn								
Tyr								
Lys								
Arg								
His								
Asp								
α AB								
β -Ala								
Cit								

表 1-2

	Phe	Trp	Met	Ser	Thr	Cys	Asn	Tyr
Ala	AlaPhe AlaAla	AlaTrp AlaAla	AlaMet	AlaSer AlaAla	AlaThr AlaAla	AlaAla ○	AlaAsn AlaAla	AlaTyr AlaAla
Gln	○	×	MetMet	SerGln SerSer	ThrGln ThrThr	○	×	×
Glu								
Gly	GlyPhe	GlyGly ○	GlyMet GlyGly	GlySer GlyGly SerGly SerSer	GlyThr GlyGly ThrGly ThrThr	GlyGly ○	GlyGly	GlyTyr GlyGly
Val			×					
Leu			MetMet		ThrLeu			
Ile			MetMet					
Pro			MetMet	SerSer	ThrThr			
Phe			MetPhe MetMet	SerPhe	ThrPhe ThrThr			
Trp								
Met			MetMet	SerMet	ThrMet ThrThr	MetMet ○		MetTyr MetMet
Ser				SerSer	SerThr SerSer ThrSer ThrThr			SerTyr SerSer
Thr					ThrThr			
Cys								
Asn								
Tyr								
Lys								
Arg								
His								
Asp								
α - AB								
β -Ala								
Cit								

表 1-3

	Lys	Arg	His	Asp	α -AB	β -Ala	Cit	重氮丝氨酸	茶氨酸
Ala	AlaAla ○	AlaArg AlaAla	AlaHis AlaAla	AlaAla	AlaAla ○		AlaAla ○	AlaAla ○	AlaAla ○
Gln	×	×	×	×	○				
Glu									
Gly	GlyGly ○	GlyArg GlyGly	GlyGly	GlyGly	GlyGly ○		○		
Val									
Leu									
Ile									
Pro									
Phe				×		○			
Trp									
Met	MetMet ○		MetMet ○			○			
Ser			SerHis		SerSer ○				
Thr					ThrThr ○				
Cys									
Asn									
Tyr									
Lys									
Arg					○				
His						β -AlaHis			
Asp									
α -AB					○				
β -Ala									
Cit						○			

在表的各单元格中显示了使用表 1 的第 1 行和最左一列所示的 2 (或 1) 种 L-氨基酸、Gly 或 β -Ala 作为底物形成的二肽。在表中, ○表示形成了二肽, 但是未鉴别其序列; ×表示没有确定形成了二肽; 空格表示没有进行反应。

实施例 7

使用能表达 His-标记的重组酶的菌株生产二肽

将在实施例 4 得到的大肠杆菌 NM522/pQE60ywfE 接种到装在大试管中的含有 50 μ g/ml 氨苄青霉素的 8 ml LB 培养基中, 在 28 $^{\circ}$ C 培养 17 小时。将得到的培养物接种到装在 250-ml 锥形瓶中的含有 50 μ g/ml 氨苄青霉素的 50 ml LB 培养基中, 在 30 $^{\circ}$ C 培养 3 小时。然后, 加入异丙基- β -D-硫代吡喃型半乳糖苷 (IPTG), 得到 1 mmol/l 的终浓度, 随后进一步在 30 $^{\circ}$ C 培养 4 小时。离心得到的培养物, 以得到湿细胞。

将包含 200 g/l 湿细胞、50 g/l 葡萄糖、5 g/l 植酸(用 33%浓氢氧化钠溶液稀释至中性)、15 g/l 磷酸二氢钾、5 g/l 硫酸镁七水合物、4 g/l Nymeen S-215、10 ml/l 二甲苯、200 mmol/l L-Ala 和 200 mmol/l L-Gln 的反应混合物(20 ml, pH 7.2)置于 50-ml 烧杯中, 在 32 $^{\circ}$ C、900 rpm 反应 2 小时。在反应过程中, 使用 2 mol/l 氢氧化钾将反应混合物的 pH 维持在 7.2。

将与上面实施例 3 相同的方法分析了反应产物, 由此确信积累了 25 mg/l L-Ala-L-Gln。

实施例 8

从芽孢杆菌属的各种微生物克隆与 ywfE 基因相对应的基因并分析在 SEQ ID NO: 9 所示的核苷酸序列的基础上, 以下述方式得到了存在于杆草芽孢杆菌 ATCC 15245、ATCC 6633、IAM 1213、IAM 1107、IAM 1214、ATCC 9466、IAM 1033 和 ATCC 21555、解淀粉芽孢杆菌 IFO 3022 和短小芽孢杆菌 NRRL B-12025 中的与 ywfE 基因相对应的基因。

也就是说, 将杆草芽孢杆菌 ATCC 15245、ATCC 6633、IAM 1213、IAM 1107、IAM 1214、ATCC 9466、IAM 1033 和 ATCC 21555、解淀粉芽孢杆菌 IFO 3022 和短小芽孢杆菌 NRRL B-12025 分别接种到 LB 培养基中, 在 30 $^{\circ}$ C 静置培养过夜。培养后, 根据使用 Current Protocols in Molecular Biology 所述的饱和苯酚的方法, 分离和纯化各微生物的染色体 DNA。

使用 DNA 合成仪 (Model 8905, PerSeptive Biosystems, Inc.), 合成了具有 SEQ ID NO: 25 和 26 所示的核苷酸序列的 DNA (以下分别简称为引物 G 和引物 H)。引物 G 具有这样的序列, 其中含有位于杆草芽孢杆菌 168 的染色体 DNA 的 ywfE 的起始密码子上游的区域, 引物 H 具有这样的序列, 其与含有位于 ywfE 的终止密码子下游的区域的序列互补。

使用杆草芽孢杆菌 ATCC 15245、ATCC 6633、IAM 1213、IAM 1107、IAM 1214、ATCC 9466、IAM 1033 和 ATCC 21555 和解淀粉芽孢杆菌 IFO 3022 的染色体 DNA 中的每一个作为模板, 使用上面的引物 G 和引物 H 作为一组引物, 进行了 PCR。也就是说, PCR 进行了 30 个循环, 一个循环包括: 在 94°C 反应 1 分钟, 在 55°C 反应 2 分钟, 在 72°C 反应 3 分钟, 使用的 40 μ l 反应混合物含有 0.1 μ g 染色体 DNA、0.5 μ mol/l 每种引物、2.5 单位的 Pfu DNA 聚合酶、4 μ l Pfu DNA 聚合酶的缓冲液 (10 x) 和 200 μ mol/l 每种 dNTP。

对得到的每种反应混合物的十分之一进行琼脂糖凝胶电泳, 以确信扩增了与 ywfE 片段相对应的约 1.4 kb 片段。然后, 将剩余的混合物与等量的以 TE 饱和的苯酚/氯仿相混合。离心得到的溶液, 将得到的上层与 2 倍体积的冷乙醇混合, 在 -80°C 静置 30 分钟。离心得到的溶液, 将得到的 DNA 沉淀溶解到 20 μ l TE 中。

对这样得到的每种源自各菌株的染色体 DNA 的 1.4 kb DNA 片段和 pCR-blunt (Invitrogen Corp.), 使用连接试剂盒在 16°C 进行连接反应 16 小时。

根据使用钙离子的方法, 使用每种连接反应混合物转化了大肠杆菌 NM522, 涂布在含有 50 μ g/ml 氨苄青霉素的 LB 琼脂培养基上, 在 30°C 培养过夜。

根据已知的方法, 从在该培养基上生长的每种转化体菌落提取质粒, 使用限制酶分析了质粒的结构, 结果, 确认得到了下述质粒, 它们含有与 ywfE 基因相对应的基因: pYWFE1 (源自 ATCC 15245 (SEQ ID NO: 36)), pYWFE2 (源自 ATCC 6633 (SEQ ID NO: 10)), pYWFE3 (源

自 IAM 1213 (SEQ ID NO: 11)), pYWFE4 (源自 IAM 1107 (SEQ ID NO: 12)), pYWFE5 (源自 IAM 1214 (SEQ ID NO: 13)), pYWFE6 (源自 ATCC 9466), pYWFE7 (源自 IAM 1033 (SEQ ID NO: 36)), pYWFE8 (源自 ATCC 21555 (SEQ ID NO: 14)) 和 pYWFE9 (源自 IFO 3022 (SEQ ID NO: 15))。

另一方面, 以下述方式得到与源自短小芽孢杆菌 NRRL B-12025 (SEQ ID NO: 16) 的 ywfE 相对应的基因。

使用上面制备的 NRRL B-12025 菌株的染色体 DNA 作为模板、分别由 SEQ ID NO: 27 和 28 所示的核苷酸序列组成的 DNA 作为一组引物, 进行了 PCR。也就是说, PCR 进行了 30 个循环, 一个循环包括: 在 98°C 反应 5 秒, 在 55°C 反应 30 秒, 在 72°C 反应 1 分钟, 使用的 50 μ l 反应混合物含有 0.1 μ g 染色体 DNA、0.5 μ mol/l 每种引物、2.5 单位的 Z-taq 聚合酶 (Takara Shuzo Co., Ltd.)、5 μ l Z-taq 聚合酶的缓冲液 (10 x) (Takara Shuzo Co., Ltd.) 和 200 μ mol/l 每种 dNTP。

对得到的反应混合物的十分之一进行琼脂糖凝胶电泳, 以确信扩增了大约 0.8 kb 片段。然后, 将剩余的反应混合物与等量的以 TE 饱和的苯酚/氯仿相混合。离心得到的混合物, 将得到的上层与 2 倍体积的冷乙醇混合, 在 -80°C 静置 30 分钟。离心得到的溶液, 将得到的 DNA 沉淀溶解到 20 μ l TE 中。

对这样得到的源自染色体 DNA 的 0.8 kb 片段和 pGEM T-easy (Promega Corp.), 使用连接试剂盒在 16°C 进行连接反应 16 小时。

根据使用钙离子的方法, 使用反应混合物转化了大肠杆菌 DH5 α , 涂布在含有 50 μ g/ml 氨苄青霉素的 LB 琼脂培养基上, 在 30°C 培养过夜。

从上面得到的转化体提取质粒, 检测大约 0.8 kb DNA 插入物的核苷酸序列, 由此确信了来自 SEQ ID NO: 16 所示的核苷酸序列的核苷酸 358 至 1160 的序列。

用 EcoRI 剪切了上面的质粒, 随后进行琼脂糖凝胶电泳以分离 DNA 片段。使用 GENECLEAN II 试剂盒纯化了 DNA 片段, 使用 DIG-High Prime DNA Labeling & Detection Starter Kit I (Roche Diagnostics

Corp.)，根据所附的说明书，DIG-标记了约 0.5 μg 纯化的 DNA 片段。

使用上面得到的 DIG-标记的 DNA，对 NRRL B-12025 菌株的染色体 DNA 进行了 Southern 分析。

分别用 BamHI、EcoRI、HindIII、KpnI、PstI、SacI、SalI 和 SphI 彻底消化了 NRRL B-12025 菌株的染色体 DNA，并进行琼脂糖凝胶电泳以分离 DNA 片段，随后根据普通方法转移到带电荷的尼龙膜 (Roche Diagnostics Corp.) 上。

通过紫外照射将 DNA 片段固定到尼龙膜上以后，使用上面的探针 DNA 和尼龙膜进行了 DNA 杂交。杂交如下进行：使尼龙膜与探针 DNA 在 65 $^{\circ}\text{C}$ 接触 16 小时，在室温用由 0.1% SDS 和 2 x SSC 组成的溶液洗涤尼龙膜 2 次、每次 5 分钟，再在 65 $^{\circ}\text{C}$ 用由 0.1% SDS 和 0.5 x SSC 组成的溶液洗涤尼龙膜 2 次、每次 15 分钟。根据上述的 DIG-High Prime DNA Labeling & Detection Starter Kit I 所附的说明书，对杂交的 DNA 进行了其它的操作、调节和检测。

结果，在用 HindIII 和 PstI 彻底消化的约 3.5 kbp 片段观察到了显色。

随后，分别用 HindIII 和 PstI 彻底消化了 NRRL B-12025 菌株的染色体 DNA，并进行琼脂糖凝胶电泳以分离 DNA 片段。使用 GENE CLEAN II 试剂盒，从各个限制酶-消化的 DNA 中纯化出了 3-4 kbp 片段，随后使用连接试剂盒进行自动环化。

在上面检测的 0.8 kb DNA 片段的核苷酸序列的基础上，设计和合成了 SEQ ID NO: 29 和 30 所示的核苷酸序列，使用它们作为引物、上面得到的环化的 DNA 作为模板，进行 PCR。PCR 进行了 30 个循环，一个循环包括：在 98 $^{\circ}\text{C}$ 反应 5 秒，在 55 $^{\circ}\text{C}$ 反应 30 秒，在 72 $^{\circ}\text{C}$ 反应 3 分钟 30 秒，使用的 50 μl 反应混合物含有 10 ng 环化的 DNA、0.5 $\mu\text{mol/l}$ 每种引物、2.5 单位的 pyrobest 聚合酶 (Takara Shuzo Co., Ltd.)、5 μl pyrobest 聚合酶的缓冲液 (10 x) (Takara Shuzo Co., Ltd.) 和 200 $\mu\text{mol/l}$ 每种 dNTP。

对得到的反应混合物的十分之一进行琼脂糖凝胶电泳，以确信扩

增了大约 3.0 kb 片段。然后，将剩余的反应混合物与等量的以 TE 饱和的苯酚/氯仿相混合。离心得到的混合物，将得到的上层与 2 倍体积的冷乙醇混合，在 -80°C 静置 30 分钟。离心得到的溶液，将得到的 DNA 沉淀溶解到 20 μ l TE 中。

使用连接试剂盒，对这样得到的 DNA 片段和 Zero Blunt PCR 克隆试剂盒 (Invitrogen Corp.) 进行连接反应。

根据使用钙离子的方法，使用反应混合物转化了大肠杆菌 NM522，涂布在含有 50 μ g/ml 氨苄青霉素的 LB 琼脂培养基上，在 30°C 培养过夜。

根据已知的方法，从在该培养基上生长的转化体菌落提取质粒，使用限制酶分析了质粒的结构。结果，确认得到了质粒 pYWFE10 (源自 NRRL B-12025)，其含有与 ywfE 基因相对应的基因。

使用 373A DNA 测序仪，检测了上面得到的分别包含在质粒 pYWFE1 至 pYWFE10 中的与 ywfE 基因相对应的基因的核苷酸序列。

由分别包含在 pYWFE1、pYWFE6 和 pYWFE7 中的基因编码的蛋白的氨基酸序列与 ywfE 基因编码的蛋白的氨基酸序列相同，然而，由分别包含在 pYWFE2、pYWFE3、pYWFE4、pYWFE5、pYWFE8、pYWFE9 和 pYWFE10 中的基因编码的蛋白的氨基酸序列与 ywfE 基因编码的蛋白的氨基酸序列不同。

由包含在 pYWFE2、pYWFE3、pYWFE4、pYWFE5、pYWFE8、pYWFE9、pYWFE10、和 pYWFE1 和 pYWFE7 中并与 ywfE 相对应的基因编码的蛋白的氨基酸序列分别如 SEQ ID NO: 2 至 8 和 1 所示，这些基因的核苷酸序列分别如 SEQ ID NO: 10 至 16 和 36 所示。

实施例 9

C-末端 His-标记的重组二肽合成酶的纯化

使用每种杆草芽孢杆菌 ATCC 15245、ATCC 6633、IAM 1213、IAM 1107、IAM 1214、ATCC 9466、IAM 1033 和 ATCC 21555 和解淀粉芽孢杆菌 IFO 3022 的染色体 DNA 作为模板、实施例 2 所述的引物 A 和引物

B 作为一组引物，进行了 PCR。也就是说，PCR 进行了 30 个循环，一个循环包括：在 94℃ 反应 1 分钟，在 55℃ 反应 2 分钟，在 72℃ 反应 3 分钟，使用的 40 μ l 反应混合物含有 0.1 μ g 染色体 DNA、0.5 μ mol/l 每种引物、2.5 单位的 Pfu DNA 聚合酶、4 μ l Pfu DNA 聚合酶的缓冲液 (10 x) 和 200 μ mol/l 每种 dNTP。

当短小芽孢杆菌 NRRL B-12025 的染色体 DNA 用作模板时，使用分别具有 SEQ ID NO: 31 和 32 所示的核苷酸序列的 DNA 作为一组引物，在与上面相同的条件下进行 PCR。

对得到的反应混合物的十分之一进行琼脂糖凝胶电泳，以确信扩增了与 ywfE 片段相对应的大约 1.4 kb DNA 片段。然后，将剩余的反应混合物与等量的以 TE 饱和的苯酚/氯仿相混合。离心得到的混合物，将得到的上层与 2 倍体积的冷乙醇混合，在 -80℃ 静置 30 分钟。离心得到的溶液，将得到的 DNA 沉淀溶解到 20 μ l TE 中。

对这样得到的每种溶液 (5 μ l) 使用限制酶 NcoI 和 BamHI 剪切扩增的 DNA。通过琼脂糖凝胶电泳分离 DNA 片段，并使用 GENE CLEAN II 试剂盒回收含有与 ywfE 相对应的基因的 1.4 kb DNA 片段。

随后，使用限制酶 NcoI 和 BamHI 剪切了 0.2 μ g C-末端 His-标记的重组表达载体 pQE60。通过琼脂糖凝胶电泳分离 DNA 片段，以与上面相同的方式回收 3.4 kb DNA 片段。

使用连接试剂盒，使上面得到的含有与杆草芽孢杆菌 168 的 ywfE 相对应的基因的 1.4 kb DNA 片段和 3.4 kb DNA 片段在 16℃ 进行连接反应 16 小时。

根据使用钙离子的方法，使用每种连接反应混合物转化了大肠杆菌 NM522，涂布在含有 50 μ g/ml 氨苄青霉素的 LB 琼脂培养基上，在 30℃ 培养过夜。

根据已知的方法，从在该培养基上生长的转化体菌落提取质粒，使用限制酶分析了每种质粒的结构。结果，确认得到了下面的 C-末端 His-标记的基因表达载体：pQE60ywfE1 (该载体含有源自 ATCC 15245 的基因)，pQE60ywfE2 (该载体含有源自 ATCC 6633 的基因)，pQE60ywfE3 (该载体含有源自 IAM 1213 的基因)，pQE60ywfE4 (该载

体含有源自 IAM 1107 的基因), pQE60ywfE5 (该载体含有源自 IAM 1214 的基因), pQE60ywfE6 (该载体含有源自 ATCC 9466 的基因), pQE60ywfE7 (该载体含有源自 IAM 1033 的基因), pQE60ywfE8 (该载体含有源自 ATCC 21555 的基因), pQE60ywfE9 (该载体含有源自 IFO 3022 的基因)和 pQE60ywfE10 (该载体含有源自 NRRL B-12025 的基因)。

将上面得到的大肠杆菌 NM522/pQE60ywfE1 至 NM522/pQE60ywfE10 菌株分别接种到装在大试管中的含有 50 μ g/ml 氨苄青霉素的 8 ml LB 培养基中, 在 28 $^{\circ}$ C 培养 17 小时。将得到的每种培养物接种到装在 250-ml 锥形瓶中的含有 50 μ g/ml 氨苄青霉素的 50 ml LB 培养基中, 在 30 $^{\circ}$ C 培养 3 小时。然后, 加入异丙基- β -D-硫代吡喃型半乳糖苷, 得到 1 mmol/l 的终浓度, 随后进一步在 30 $^{\circ}$ C 培养 4 小时。离心得到的培养物, 以得到湿细胞, 使用 HisTrap, 根据所附的说明书, 从各湿细胞纯化 His-标记的重组酶。

实施例 10

使用纯化的酶生产二肽

制备了含有 0.04 mg 在实施例 9 中得到的各种重组酶、100 mmol/l Tris-HCl (pH 8.0)、60 mmol/l 氯化镁、60 mmol/l ATP、30 mmol/l L-Ala 和 30 mmol/l L-Gln 的反应混合物(每种 0.1 ml), 在 37 $^{\circ}$ C 反应 16 小时。

反应结束后, 通过实施例 3 所述的方法分析了反应产物, 由此确信形成和积累了 3.0 至 3.5 g/l L-Ala-L-Gln 和 0.25 至 0.3 g/l L-Ala-L-Ala。

当上面的反应混合物的组成不包括 ATP 时, 根本没有形成 L-Ala-L-Gln 或 L-Ala-L-Ala。

上面的结果表明, 在存在 ATP 时, 所有在实施例 8 得到的基因的产物都具有从 L-Ala 和 L-Gln 生产 L-Ala-L-Gln 和 L-Ala-L-Ala 的活性。

序列表中的文字说明

SEQ ID NO: 19 -人工序列的描述: 引物

SEQ ID NO: 20 -人工序列的描述: 引物

SEQ ID NO: 21 -人工序列的描述: 引物

SEQ ID NO: 22 -人工序列的描述: 引物

SEQ ID NO: 23 -人工序列的描述: 引物

SEQ ID NO: 24 -人工序列的描述: 引物

SEQ ID NO: 25 -人工序列的描述: 引物

SEQ ID NO: 26 -人工序列的描述: 引物

SEQ ID NO: 27 -人工序列的描述: 引物

SEQ ID NO: 28 -人工序列的描述: 引物

SEQ ID NO: 29 -人工序列的描述: 引物

SEQ ID NO: 30 -人工序列的描述: 引物

SEQ ID NO: 31 -人工序列的描述: 引物

SEQ ID NO: 32 -人工序列的描述: 引物

SEQ ID NO: 33 -人工序列的描述: 在数据库搜索中使用的氨基酸序列

SEQ ID NO: 34 -人工序列的描述: 在数据库搜索中使用的氨基酸序列

SEQ ID NO: 35 -人工序列的描述: 在数据库搜索中使用的氨基酸序列

序列表

<110> KYOWA HAKKO KOGYO CO., LTD.

<120> 生产二肽的方法

<130> 11524W01

<150> JP 2002-376054

<151> 2002-12-26

<150> JP 2003-420887

<151> 2003-12-18

<160> 36

<170> PatentIn Ver. 2.1

<210> 1

<211> 472

<212> PRT

<213> 枯草芽孢杆菌 168

<400> 1

Met Glu Arg Lys Thr Val Leu Val Ile Ala Asp Leu Gly Gly Cys Pro
1 5 10 15

Pro His Met Phe Tyr Lys Ser Ala Ala Glu Lys Tyr Asn Leu Val Ser
20 25 30

Phe Ile Pro Arg Pro Phe Ala Ile Thr Ala Ser His Ala Ala Leu Ile
35 40 45

Glu Lys Tyr Ser Val Ala Val Ile Lys Asp Lys Asp Tyr Phe Lys Ser
50 55 60

Leu Ala Asp Phe Glu His Pro Asp Ser Ile Tyr Trp Ala His Glu Asp
65 70 75 80

His Asn Lys Pro Glu Glu Glu Val Val Glu Gln Ile Val Lys Val Ala
85 90 95

Glu Met Phe Gly Ala Asp Ala Ile Thr Thr Asn Asn Glu Leu Phe Ile
100 105 110

Ala Pro Met Ala Lys Ala Cys Glu Arg Leu Gly Leu Arg Gly Ala Gly
115 120 125

Val Gln Ala Ala Glu Asn Ala Arg Asp Lys Asn Lys Met Arg Asp Ala
130 135 140

Phe Asn Lys Ala Gly Val Lys Ser Ile Lys Asn Lys Arg Val Thr Thr
145 150 155 160

Leu Glu Asp Phe Arg Ala Ala Leu Glu Glu Ile Gly Thr Pro Leu Ile
165 170 175

Leu Lys Pro Thr Tyr Leu Ala Ser Ser Ile Gly Val Thr Leu Ile Thr
180 185 190

Asp Thr Glu Thr Ala Glu Asp Glu Phe Asn Arg Val Asn Asp Tyr Leu
195 200 205

Lys Ser Ile Asn Val Pro Lys Ala Val Thr Phe Glu Ala Pro Phe Ile
210 215 220

Ala Glu Glu Phe Leu Gln Gly Glu Tyr Gly Asp Trp Tyr Gln Thr Glu
225 230 235 240

Gly Tyr Ser Asp Tyr Ile Ser Ile Glu Gly Ile Met Ala Asp Gly Glu
245 250 255

Tyr Phe Pro Ile Ala Ile His Asp Lys Thr Pro Gln Ile Gly Phe Thr
260 265 270

Glu Thr Ser His Ile Thr Pro Ser Ile Leu Asp Glu Glu Ala Lys Lys
275 280 285

Lys Ile Val Glu Ala Ala Lys Lys Ala Asn Glu Gly Leu Gly Leu Gln
290 295 300

Asn Cys Ala Thr His Thr Glu Ile Lys Leu Met Lys Asn Arg Glu Pro
305 310 315 320

Gly Leu Ile Glu Ser Ala Ala Arg Phe Ala Gly Trp Asn Met Ile Pro
325 330 335

Asn Ile Lys Lys Val Phe Gly Leu Asp Met Ala Gln Leu Leu Leu Asp

Pro His Met Phe Tyr Lys Ser Ala Ala Glu Lys Tyr Asn Leu Val Ser
 20 25 30

Phe Ile Pro Arg Pro Phe Ala Ile Thr Ala Ser His Ala Ala Leu Ile
 35 40 45

Glu Lys Tyr Ser Val Ala Val Ile Lys Asp Lys Asp Tyr Phe Lys Ser
 50 55 60

Leu Ala Asp Phe Glu His Pro Asp Ser Ile Tyr Trp Ala His Glu Asp
 65 70 75 80

His Asn Lys Pro Glu Glu Val Val Glu Gln Ile Val Lys Val Ala
 85 90 95

Glu Met Phe Gly Ala Asp Ala Ile Thr Thr Asn Asn Glu Leu Phe Ile
 100 105 110

Ala Pro Met Ala Lys Ala Cys Glu Arg Leu Gly Leu Arg Gly Ala Gly
 115 120 125

Val Gln Ala Ala Glu Asn Ala Arg Asp Lys Asn Lys Met Arg Asp Ala
 130 135 140

Phe Asn Lys Ala Gly Val Lys Ser Ile Lys Asn Lys Arg Val Thr Thr
 145 150 155 160

Leu Glu Asp Phe Arg Ala Ala Leu Glu Glu Ile Gly Thr Pro Leu Ile
 165 170 175

Leu Lys Pro Thr Tyr Leu Ala Ser Ser Ile Gly Val Thr Leu Ile Thr
 180 185 190

Asp Thr Glu Thr Ala Glu Asp Glu Phe Asn Arg Val Asn Asp Tyr Leu
 195 200 205

Lys Ser Ile Asn Val Pro Lys Ala Val Thr Phe Glu Ala Pro Phe Ile
 210 215 220

Ala Glu Glu Phe Leu Gln Gly Glu Tyr Gly Asp Trp Tyr Gln Thr Glu
 225 230 235 240

Gly Tyr Ser Asp Tyr Ile Ser Ile Glu Gly Ile Met Ala Asp Gly Glu

<210> 4

<211> 472

<212> PRT

<213> 枯草芽孢杆菌 IAM1107

<400> 4

Met Glu Arg Lys Thr Val Leu Val Ile Ala Asp Leu Gly Gly Cys Pro
1 5 10 15

Pro His Met Phe Tyr Lys Ser Ala Ala Glu Lys Tyr Asn Leu Val Ser
 20 25 30

Phe Ile Pro Arg Pro Phe Ala Ile Thr Ala Ser His Ala Ala Leu Ile
 35 40 45

Glu Lys Tyr Ser Val Ala Val Val Lys Asp Lys Asp Tyr Phe Lys Ser
 50 55 60

Leu Ala Asp Phe Glu His Pro Asp Ser Ile Tyr Trp Ala His Glu Asp
65 70 75 80

His Asn Lys Pro Glu Glu Glu Val Val Glu Gln Ile Val Lys Val Ala
 85 90 95

Glu Met Phe Gly Ala Asp Ala Ile Thr Thr Asn Asn Glu Leu Phe Ile
 100 105 110

Ala Pro Met Ala Lys Ala Cys Glu Arg Leu Gly Leu Arg Gly Ala Gly
 115 120 125

Val Gln Ala Ala Glu Asn Ala Arg Asp Lys Asn Lys Met Arg Asp Ala
 130 135 140

Phe Asn Lys Ala Gly Val Lys Ser Ile Lys Asn Lys Arg Val Thr Thr
145 150 155 160

Leu Glu Asp Phe Arg Ala Ala Leu Glu Glu Ile Gly Thr Pro Leu Ile
 165 170 175

Leu Lys Pro Thr Tyr Leu Ala Ser Ser Ile Gly Val Thr Leu Ile Thr
 180 185 190

Asp Thr Glu Thr Ala Glu Asp Glu Phe Asn Arg Val Asn Asp Tyr Leu

195	200	205
Lys Ser Ile Asn Val Pro Lys Ala Val Thr Phe Glu Ala Pro Phe Ile		
210	215	220
Ala Glu Glu Phe Leu Gln Gly Glu Tyr Gly Asp Trp Tyr Gln Thr Glu		
225	230	235
Gly Tyr Ser Asp Tyr Ile Ser Ile Glu Gly Ile Met Ala Asp Gly Glu		
	245	250
		255
Tyr Phe Pro Ile Ala Ile His Asp Lys Thr Pro Gln Ile Gly Phe Thr		
	260	265
		270
Glu Thr Ser His Ile Thr Pro Ser Ile Leu Asp Glu Glu Ala Lys Lys		
	275	280
		285
Lys Ile Val Glu Ala Ala Lys Lys Ala Asn Glu Gly Leu Gly Leu Gln		
	290	295
		300
Asn Cys Ala Thr His Thr Glu Val Lys Leu Met Lys Asn Arg Glu Pro		
305	310	315
		320
Gly Leu Ile Glu Ser Ala Ala Arg Phe Ala Gly Trp Asn Met Ile Pro		
	325	330
		335
Asn Ile Lys Lys Val Phe Gly Leu Asp Met Ala Gln Leu Leu Leu Asp		
	340	345
		350
Val Leu Cys Phe Gly Lys Asp Ala Asp Leu Pro Asp Gly Leu Leu Asp		
	355	360
		365
Gln Glu Pro Tyr Tyr Val Ala Asp Cys His Leu Tyr Pro Gln His Phe		
	370	375
		380
Lys Gln Asn Gly Gln Ile Pro Glu Thr Ala Glu Asp Leu Val Ile Glu		
385	390	395
		400
Ala Ile Asp Ile Pro Asp Gly Leu Leu Lys Gly Asp Thr Glu Ile Phe		
	405	410
		415
Ser Phe Ser Ala Ala Ala Pro Gly Thr Ser Val Asp Leu Thr Leu Phe		
	420	425
		430

Glu Ala Phe Asn Ser Ile Ala Ala Phe Glu Leu Lys Gly Ser Asn Ser
 435 440 445

Gln Asp Val Ala Glu Ser Ile Arg Gln Ile Gln Gln His Ala Lys Leu
 450 455 460

Thr Ala Lys Tyr Val Leu Pro Val
 465 470

<210> 5

<211> 472

<212> PRT

<213> 枯草芽孢杆菌 IAM1214

<400> 5

Met Glu Arg Lys Thr Val Leu Val Ile Ala Asp Leu Gly Gly Cys Pro
 1 5 10 15

Pro His Met Phe Tyr Lys Ser Ala Ala Glu Lys Tyr Asn Leu Val Ser
 20 25 30

Phe Ile Pro Arg Pro Phe Ala Ile Thr Ala Ser His Ala Ala Leu Ile
 35 40 45

Glu Lys Tyr Ser Val Ala Val Ile Lys Asp Lys Asp Tyr Phe Gln Ser
 50 55 60

Leu Ala Asp Phe Glu His Pro Asp Ser Ile Tyr Trp Ala His Glu Asp
 65 70 75 80

His Asp Lys Pro Glu Glu Glu Val Val Glu Gln Ile Val Lys Val Ala
 85 90 95

Gln Met Phe Glu Ala Asp Ala Ile Thr Thr Asn Asn Glu Leu Phe Ile
 100 105 110

Ala Pro Met Ala Lys Ala Cys Glu Arg Leu Gly Leu Arg Gly Ala Gly
 115 120 125

Val Gln Ala Ala Glu Asn Ala Arg Asp Lys Asn Lys Met Arg Asp Ala
 130 135 140

Phe Asn Lys Ala Gly Val Lys Ser Ile Lys Asn Lys Arg Val Thr Thr

145 150 155 160
 Leu Glu Asp Phe Arg Ala Ala Leu Glu Glu Ile Gly Thr Pro Leu Ile
 165 170 175
 Leu Lys Pro Thr Tyr Leu Ala Ser Ser Ile Gly Val Thr Leu Ile Thr
 180 185 190
 Asp Thr Glu Thr Ala Glu Asp Glu Phe Asn Arg Val Asn Asp Tyr Leu
 195 200 205
 Lys Ser Ile Asn Val Pro Lys Ala Val Thr Phe Glu Ala Pro Phe Ile
 210 215 220
 Ala Glu Glu Phe Leu Gln Gly Glu Tyr Gly Asp Trp Tyr Gln Thr Glu
 225 230 235 240
 Gly Tyr Ser Asp Tyr Ile Ser Ile Glu Gly Ile Met Ala Asp Gly Glu
 245 250 255
 Tyr Phe Pro Ile Ala Ile His Asp Lys Thr Pro Gln Ile Gly Phe Thr
 260 265 270
 Glu Thr Ser His Ile Thr Pro Ser Ile Leu Asp Glu Glu Ala Lys Lys
 275 280 285
 Lys Ile Val Glu Ala Ala Lys Lys Ala Asn Glu Gly Leu Gly Leu Gln
 290 295 300
 Asn Cys Ala Thr His Thr Glu Ile Lys Leu Met Lys Asn Arg Glu Pro
 305 310 315 320
 Gly Leu Ile Glu Ser Ala Ala Arg Phe Ala Gly Trp Asn Met Ile Pro
 325 330 335
 Asn Ile Lys Lys Val Phe Gly Leu Asp Met Ala Gln Leu Leu Leu Asp
 340 345 350
 Val Leu Cys Phe Gly Lys Asp Ala Asp Leu Pro Asp Gly Leu Leu Asp
 355 360 365
 Gln Glu Pro Tyr Tyr Val Ala Asp Cys His Leu Tyr Pro Gln His Phe
 370 375 380

Lys Gln Asn Gly Gln Ile Pro Glu Thr Ala Glu Asp Leu Val Ile Glu
385 390 395 400

Ala Ile Asp Ile Pro Asp Gly Leu Leu Lys Gly Asp Thr Glu Ile Val
 405 410 415

Thr Phe Ser Ala Ala Ala Pro Gly Thr Ser Val Asp Leu Thr Leu Phe
 420 425 430

Glu Ala Phe Asn Ser Ile Ala Ala Phe Glu Leu Lys Gly Ser Asn Ser
 435 440 445

Gln Asp Val Ala Glu Ser Ile Arg Gln Ile Gln Gln His Ala Lys Leu
 450 455 460

Thr Ala Lys Tyr Val Leu Pro Val
465 470

<210> 6

<211> 472

<212> PRT

<213> 枯草芽孢杆菌 ATCC21555

<400> 6

Met Glu Arg Lys Thr Val Leu Val Ile Ala Asp Leu Gly Gly Cys Pro
1 5 10 15

Pro His Met Phe Tyr Lys Ser Ala Ala Glu Lys Tyr Asn Leu Val Ser
 20 25 30

Phe Ile Pro Arg Pro Phe Ala Ile Thr Ala Ser His Ala Ala Leu Ile
 35 40 45

Glu Lys Tyr Ser Ile Ala Val Ile Lys Asp Lys Asp Tyr Phe Lys Ser
 50 55 60

Leu Ala Asp Phe Glu His Pro Asp Ser Ile Tyr Trp Ala His Glu Asp
65 70 75 80

His Asp Lys Pro Glu Glu Val Val Glu Glu Ile Val Lys Val Ala
 85 90 95

Asp Met Phe Gly Val Asp Ala Ile Thr Thr Asn Asn Glu Leu Phe Ile

100	105	110
Ala Pro Met Ala Lys Ala Cys Lys Arg Leu Gly Leu Arg Gly Ala Gly		
115	120	125
Val Gln Ala Ala Glu Asn Ala Arg Asp Lys Asn Lys Met Arg Ala Ala		
130	135	140
Phe Asn Arg Ala Gly Val Lys Ser Ile Lys Asn Lys Arg Val Thr Thr		
145	150	155
Leu Glu Asp Phe Arg Ala Ala Leu Gln Glu Ile Gly Thr Pro Leu Ile		
165	170	175
Leu Lys Pro Thr Tyr Leu Ala Ser Ser Ile Gly Val Thr Leu Ile Lys		
180	185	190
Glu Met Glu Thr Ala Glu Ala Glu Phe Asn Arg Val Asn Glu Tyr Leu		
195	200	205
Lys Ser Ile Asn Val Pro Lys Ala Val Thr Phe Glu Ala Pro Phe Ile		
210	215	220
Ala Glu Glu Phe Leu Gln Gly Glu Tyr Asp Asp Trp Tyr Glu Thr Ser		
225	230	235
Gly Tyr Ser Asp Tyr Ile Ser Ile Glu Gly Ile Met Ala Asp Gly Glu		
245	250	255
Tyr Phe Pro Val Ala Ile His Asp Lys Thr Pro Gln Ile Gly Phe Thr		
260	265	270
Glu Thr Ala His Ile Thr Pro Ser Ile Leu Asp Asp Asp Ala Lys Arg		
275	280	285
Lys Ile Val Glu Ala Ala Lys Lys Ala Asn Glu Gly Leu Gly Leu Glu		
290	295	300
Asn Cys Ala Thr His Thr Glu Ile Lys Leu Met Lys Asn Arg Glu Ala		
305	310	315
Gly Leu Ile Glu Ser Ala Ala Arg Phe Ala Gly Trp Asn Met Ile Pro		
325	330	335

Asn Ile Lys Lys Val Phe Gly Val Asp Met Ala Gln Leu Leu Leu Asp
 340 345 350

Val Leu Cys Tyr Gly Lys Glu Ala Asp Leu Pro Lys Gly Leu Leu Glu
 355 360 365

Gln Glu Pro Cys Tyr Val Ala Asp Cys His Leu Tyr Pro Gln His Phe
 370 375 380

Lys Glu Asn Gly Gln Leu Pro Glu Thr Val Val Asp Phe Val Ile Glu
 385 390 395 400

Ser Ile Glu Ile Pro Asp Gly Val Leu Lys Gly Asp Thr Glu Leu Val
 405 410 415

Ser Phe Ser Ala Ala Glu Ala Gly Thr Ser Val Asp Leu Arg Leu Phe
 420 425 430

Glu Ala Phe Asn Ser Ile Ala Ala Phe Glu Leu Lys Gly Ser Asn Ser
 435 440 445

Asn Asp Val Ala Glu Ser Ile Lys Gln Ile Gln Gln Gln Ala Lys Leu
 450 455 460

Thr Ala Lys Tyr Ala Leu Ser Val
 465 470

<210> 7

<211> 472

<212> PRT

<213> 解淀粉芽孢杆菌 IF03022

<400> 7

Met Glu Arg Lys Thr Val Leu Val Ile Ala Asp Leu Gly Gly Cys Pro
 1 5 10 15

Pro His Met Phe Tyr Lys Ser Ala Ala Glu Lys Tyr Asn Leu Val Ser
 20 25 30

Phe Ile Pro Arg Pro Phe Ala Ile Thr Ala Ser His Ala Ala Leu Ile
 35 40 45

Glu Lys Tyr Ser Val Ala Val Ile Lys Asp Lys Asp Tyr Phe Lys Ser

Lys Ile Val Glu Ala Ala Lys Lys Ala Asn Glu Gly Leu Gly Leu Glu
290 295 300

Asn Cys Ala Thr His Thr Glu Ile Lys Leu Met Lys Asn Arg Glu Ala
305 310 315 320

Gly Leu Ile Glu Ser Ala Ala Arg Phe Ala Gly Trp Asn Met Ile Pro
325 330 335

Asn Ile Lys Lys Val Phe Gly Val Asp Met Ala Gln Leu Leu Leu Asp
340 345 350

Val Leu Cys Phe Gly Lys Glu Ala Asp Leu Pro Lys Gly Leu Leu Glu
355 360 365

Gln Glu Pro Cys Tyr Val Ala Asp Cys His Leu Tyr Pro Gln His Phe
370 375 380

Lys Glu Asn Gly Gln Leu Pro Glu Thr Ala Val Asp Phe Val Ile Glu
385 390 395 400

Ser Ile Asp Ile Pro Asp Gly Val Leu Lys Gly Asp Thr Glu Ile Val
405 410 415

Ser Phe Ser Ala Ala Glu Ala Gly Thr Ser Val Asp Leu Arg Leu Phe
420 425 430

Glu Ala Phe Asn Ser Ile Ala Ala Phe Glu Leu Lys Gly Ser Asn Ser
435 440 445

Gly Asp Val Ala Glu Ser Ile Lys Gln Ile Gln Gln Gln Ala Lys Leu
450 455 460

Thr Ala Lys Tyr Ala Leu Pro Val
465 470

<210> 8

<211> 476

<212> PRT

<213> 短小芽孢杆菌 NRRL B-12025

<400> 26

Val Leu Ser Leu Ser Lys Lys Thr Val Leu Val Ile Ala Asp Leu Gly

1	5	10	15
Gly Cys Pro Pro His Met Phe Tyr Glu Ser Val Ala Ala Ser Tyr His	20	25	30
Ile Val Ser Tyr Ile Pro Arg Pro Phe Ala Ile Thr Lys Gly His Ala	35	40	45
Glu Leu Ile Glu Lys Tyr Ser Ile Ala Val Ile Lys Asp Arg Asp Tyr	50	55	60
Phe Glu Thr His Pro Ser Phe Glu His Pro Asp Ser Ile Tyr Trp Ala	65	70	75
His Asp Asp Tyr Pro Lys Ser Glu Glu Glu Val Val Glu Asp Phe Ile	85	90	95
Arg Val Ala Ser Phe Phe Lys Ala Asp Ala Ile Thr Thr Asn Asn Glu	100	105	110
Leu Phe Ile Ala Pro Met Ala Lys Ala Ala Glu Arg Leu Gly Leu Arg	115	120	125
Gly Ala Gly Val Lys Ala Ala Glu Met Ala Arg Asp Lys Ser Gln Met	130	135	140
Arg Ala Ala Phe Asn Ala Ser Gly Val Lys Ala Val Lys Thr Gln Pro	145	150	155
Val Thr Thr Leu Ser Asp Phe Gln Gln Ala Ile Glu Ser Ile Gly Thr	165	170	175
Pro Leu Ile Leu Lys Pro Thr Tyr Leu Ala Ser Ser Ile Gly Val Thr	180	185	190
Leu Phe His Asp Lys Ala Gly Ser Asp Asp Leu Phe Leu Gln Val Gln	195	200	205
Ser Tyr Leu Glu Thr Ile Pro Val Pro Asp Ala Val Thr Tyr Glu Ala	210	215	220
Pro Phe Val Ala Glu Thr Tyr Leu Glu Gly Ala Tyr Glu Asp Trp Tyr	225	230	235
			240

Glu Asp Glu Gly Tyr Ala Asp Tyr Val Ser Val Glu Gly Leu Val Val
 245 250 255

Glu Gly Glu Tyr Leu Pro Phe Val Ile His Asp Lys Thr Pro Gln Ile
 260 265 270

Gly Phe Thr Glu Thr Ala His Ile Thr Pro Thr Ile Leu Asp Asn Glu
 275 280 285

Ala Lys Gln Ile Ile Ile Glu Ala Ala Arg Lys Ala Asn Glu Gly Leu
 290 295 300

Gly Leu Glu His Cys Ala Thr His Thr Glu Ile Lys Leu Met Lys Asn
 305 310 315 320

Arg Glu Thr Gly Leu Ile Glu Ala Ala Ala Arg Phe Ala Gly Trp Asn
 325 330 335

Met Ile Pro Asn Ile Lys Lys Val Phe Gly Val Asp Met Ala Lys Leu
 340 345 350

Leu Ile Asp Val Leu Val Asp Gly Lys Lys Ala Val Leu Pro Lys Gln
 355 360 365

Leu Leu Ser Gly His Thr Phe Tyr Val Ala Asp Cys His Leu Tyr Pro
 370 375 380

Gln His Phe Lys Glu Ser Gly Leu Ile Pro Pro Glu Ala Thr His Ile
 385 390 395 400

Thr Ile Asp His Val Ser Ile Pro Gln Glu Ala Phe Val Gly Asp Thr
 405 410 415

Ala Ile Val Ser Gln Ser Phe Pro Ala Lys Gly Thr Ile Val Asp Leu
 420 425 430

Glu Leu Phe Glu Ala Phe Asn Gly Ile Val Ser Leu Glu Leu Lys Gly
 435 440 445

Ser Ser Ser Gln Asp Val Ala Ala Ser Ile Arg Asn Ile Gln Lys Gln
 450 455 460

Ala Thr Ile Gln Leu Met Asp Glu Leu Val Lys Gly
 465 470 475

<210> 9

<211> 1416

<212> DNA

<213> 枯草芽孢杆菌 168

<400> 9

atg gag aga aaa aca gta ttg gtc atc gct gat ctt gga ggc tgc ccg 48
Met Glu Arg Lys Thr Val Leu Val Ile Ala Asp Leu Gly Gly Cys Pro

1 5 10 15

ccg cac atg ttt tat aaa agc gct gct gaa aaa tat aac ctg gtc agc 96
Pro His Met Phe Tyr Lys Ser Ala Ala Glu Lys Tyr Asn Leu Val Ser

20 25 30

ttt att cca aga cct ttt gca att aca gcc tcc cat gca gca ttg att 144
Phe Ile Pro Arg Pro Phe Ala Ile Thr Ala Ser His Ala Ala Leu Ile

35 40 45

gaa aaa tac tcg gtc gcg gtc ata aaa gat aaa gac tat ttt aag agt 192
Glu Lys Tyr Ser Val Ala Val Ile Lys Asp Lys Asp Tyr Phe Lys Ser

50 55 60

tta gct gat ttt gaa cac cct gat tcc att tat tgg gcg cat gaa gat 240
Leu Ala Asp Phe Glu His Pro Asp Ser Ile Tyr Trp Ala His Glu Asp

65 70 75 80

cat aac aag cct gag gaa gag gtc gtc gag caa atc gtc aag gtt gcc 288
His Asn Lys Pro Glu Glu Glu Val Val Glu Gln Ile Val Lys Val Ala

85 90 95

gaa atg ttt ggg gcg gat gcc atc aca aca aac aat gaa tta ttc att 336
Glu Met Phe Gly Ala Asp Ala Ile Thr Thr Asn Asn Glu Leu Phe Ile

100 105 110

gct ccg atg gcg aaa gcc tgt gaa cgt ctg ggc ttg aga ggt gcc gcc 384
Ala Pro Met Ala Lys Ala Cys Glu Arg Leu Gly Leu Arg Gly Ala Gly

115 120 125

gtg cag gca gcc gaa aat gcc aga gat aaa aat aaa atg agg gac gct 432
Val Gln Ala Ala Glu Asn Ala Arg Asp Lys Asn Lys Met Arg Asp Ala

130 135 140

ttt aat aag gcc gga gtc aaa tcg atc aaa aac aaa cga gtc aca act	480
Phe Asn Lys Ala Gly Val Lys Ser Ile Lys Asn Lys Arg Val Thr Thr	
145 150 155 160	
ctt gaa gat ttc cgt gct gct ctt gaa gag atc ggc aca cct ctt atc	528
Leu Glu Asp Phe Arg Ala Ala Leu Glu Glu Ile Gly Thr Pro Leu Ile	
165 170 175	
tta aag cct aca tac tta gcg agt tct atc ggt gta acg ctg att acg	576
Leu Lys Pro Thr Tyr Leu Ala Ser Ser Ile Gly Val Thr Leu Ile Thr	
180 185 190	
gac act gag acg gca gaa gat gaa ttt aac aga gtc aat gac tat ctg	624
Asp Thr Glu Thr Ala Glu Asp Glu Phe Asn Arg Val Asn Asp Tyr Leu	
195 200 205	
aaa tca att aac gtg cca aag gcg gtt acg ttt gaa gcg ccg ttt atc	672
Lys Ser Ile Asn Val Pro Lys Ala Val Thr Phe Glu Ala Pro Phe Ile	
210 215 220	
gct gaa gaa ttt tta cag ggt gag tac gga gac tgg tat caa aca gaa	720
Ala Glu Glu Phe Leu Gln Gly Glu Tyr Gly Asp Trp Tyr Gln Thr Glu	
225 230 235 240	
ggg tac tcc gac tat atc agt ata gaa ggc atc atg gct gac ggt gag	768
Gly Tyr Ser Asp Tyr Ile Ser Ile Glu Gly Ile Met Ala Asp Gly Glu	
245 250 255	
tat ttc ccg atc gcc att cat gat aaa acg ccg caa atc ggg ttt aca	816
Tyr Phe Pro Ile Ala Ile His Asp Lys Thr Pro Gln Ile Gly Phe Thr	
260 265 270	
gag aca tcc cac att acg ccg tcc att ctg gat gaa gag gca aaa aag	864
Glu Thr Ser His Ile Thr Pro Ser Ile Leu Asp Glu Glu Ala Lys Lys	
275 280 285	
aaa att gtc gaa gct gcc aaa aag gca aat gaa ggg ctt gga ctg caa	912
Lys Ile Val Glu Ala Ala Lys Lys Ala Asn Glu Gly Leu Gly Leu Gln	
290 295 300	
aat tgc gca aca cat aca gag atc aag cta atg aaa aac aga gaa ccg	960
Asn Cys Ala Thr His Thr Glu Ile Lys Leu Met Lys Asn Arg Glu Pro	
305 310 315 320	

ggt tta ata gag tcg gca gcc aga ttt gcc ggc tgg aat atg atc ccc 1008
 Gly Leu Ile Glu Ser Ala Ala Arg Phe Ala Gly Trp Asn Met Ile Pro
 325 330 335

aat att aaa aag gtc ttt ggc ctt gat atg gcg caa tta tta tta gat 1056
 Asn Ile Lys Lys Val Phe Gly Leu Asp Met Ala Gln Leu Leu Leu Asp
 340 345 350

gtc ctc tgt ttc gga aaa gac gcc gat ctg ccg gac gga tta ttg gat 1104
 Val Leu Cys Phe Gly Lys Asp Ala Asp Leu Pro Asp Gly Leu Leu Asp
 355 360 365

caa gag cct tat tat gtt gcc gac tgc cat ttg tac ccg cag cat ttc 1152
 Gln Glu Pro Tyr Tyr Val Ala Asp Cys His Leu Tyr Pro Gln His Phe
 370 375 380

aaa caa aat ggc caa att cct gaa act gct gag gat ttg gtc att gaa 1200
 Lys Gln Asn Gly Gln Ile Pro Glu Thr Ala Glu Asp Leu Val Ile Glu
 385 390 395 400

gcg atc gat att ccg gac ggg ctt tta aaa ggg gat act gaa atc gtt 1248
 Ala Ile Asp Ile Pro Asp Gly Leu Leu Lys Gly Asp Thr Glu Ile Val
 405 410 415

tct ttt tcg gcc gca gca cca ggc act tca gtt gat ttg aca ttg ttt 1296
 Ser Phe Ser Ala Ala Ala Pro Gly Thr Ser Val Asp Leu Thr Leu Phe
 420 425 430

gaa gct ttc aat tcc att gct gca ttt gaa ctg aaa ggc agt aat tca 1344
 Glu Ala Phe Asn Ser Ile Ala Ala Phe Glu Leu Lys Gly Ser Asn Ser
 435 440 445

cag gat gtg gct gaa tca atc aga caa att cag cag cat gcg aag ctg 1392
 Gln Asp Val Ala Glu Ser Ile Arg Gln Ile Gln Gln His Ala Lys Leu
 450 455 460

acg gca aag tat gtg ctg cca gta 1416
 Thr Ala Lys Tyr Val Leu Pro Val
 465 470

<210> 10

<211> 1416

<212> DNA

<213> 枯草芽孢杆菌 ATCC6633

<400> 10

```

atg gag aga aaa aca gta ttg gtc atc gct gat ctt gga ggc tgc ccg 48
Met Glu Arg Lys Thr Val Leu Val Ile Ala Asp Leu Gly Gly Cys Pro
  1           5           10           15

ccg cac atg ttt tat aaa agc gct gct gaa aaa tat aac ctg gtt agc 96
Pro His Met Phe Tyr Lys Ser Ala Ala Glu Lys Tyr Asn Leu Val Ser
           20           25           30

ttt att ccg aga cct ttt gca ata aca gcc tcc cat gca gca ctg att 144
Phe Ile Pro Arg Pro Phe Ala Ile Thr Ala Ser His Ala Ala Leu Ile
           35           40           45

gaa aaa tac tcg gtc gcg gtc ata aaa gat aaa gac tat ttt cag agc 192
Glu Lys Tyr Ser Val Ala Val Ile Lys Asp Lys Asp Tyr Phe Gln Ser
           50           55           60

tta gct gat ttt gag cat ccc gat tca att tat tgg gcg cat gag gat 240
Leu Ala Asp Phe Glu His Pro Asp Ser Ile Tyr Trp Ala His Glu Asp
           65           70           75           80

cat gac aag cct gaa gaa gag gtt gtc gag caa atc gtc aag gtt gcc 288
His Asp Lys Pro Glu Glu Glu Val Val Glu Gln Ile Val Lys Val Ala
           85           90           95

caa atg ttt gag gcg gac gcc atc aca aca aac aat gaa tta ttc att 336
Gln Met Phe Glu Ala Asp Ala Ile Thr Thr Asn Asn Glu Leu Phe Ile
           100           105           110

gcc ccg atg gcg aaa gcc tgt gaa cgc ctt ggc ctg agg ggc gcc gga 384
Ala Pro Met Ala Lys Ala Cys Glu Arg Leu Gly Leu Arg Gly Ala Gly
           115           120           125

gtg cag gca gcg gaa aat gcc aga gat aaa aat aaa atg agg gac gct 432
Val Gln Ala Ala Glu Asn Ala Arg Asp Lys Asn Lys Met Arg Asp Ala
           130           135           140

ttt aat aag gcg gga gtc aaa tcg atc aaa aac aaa cga gtc aca act 480
Phe Asn Lys Ala Gly Val Lys Ser Ile Lys Asn Lys Arg Val Thr Thr
           145           150           155           160

ctt gag gat ttt cgt gct gca ctt gaa gag atc ggc aca cct cta atc 528

```

Leu Glu Asp Phe Arg Ala Ala Leu Glu Glu Ile Gly Thr Pro Leu Ile
 165 170 175

tta aag cct aca tac tta gcg agt tca atc ggc gta acg ctg att acc 576
 Leu Lys Pro Thr Tyr Leu Ala Ser Ser Ile Gly Val Thr Leu Ile Thr
 180 185 190

gac acg gag acg gca gaa gat gaa ttt aac aga gtc aat gac tac ctg 624
 Asp Thr Glu Thr Ala Glu Asp Glu Phe Asn Arg Val Asn Asp Tyr Leu
 195 200 205

aaa tcg att aac gtg ccg aag gcg gtc aca ttt gaa gca ccg ttt att 672
 Lys Ser Ile Asn Val Pro Lys Ala Val Thr Phe Glu Ala Pro Phe Ile
 210 215 220

gct gag gaa ttt tta cag ggt gag tac gga gac tgg tat caa aca gaa 720
 Ala Glu Glu Phe Leu Gln Gly Glu Tyr Gly Asp Trp Tyr Gln Thr Glu
 225 230 235 240

ggg tac tcc gac tat atc agc ata gaa ggc att atg gca gat ggt gag 768
 Gly Tyr Ser Asp Tyr Ile Ser Ile Glu Gly Ile Met Ala Asp Gly Glu
 245 250 255

tat ttt ccg atc gcc att cat gac aaa acg ccg caa att gga ttt aca 816
 Tyr Phe Pro Ile Ala Ile His Asp Lys Thr Pro Gln Ile Gly Phe Thr
 260 265 270

gag aca tca cat att acg cca tcc att ctg gat gaa gag gcg aaa aag 864
 Glu Thr Ser His Ile Thr Pro Ser Ile Leu Asp Glu Glu Ala Lys Lys
 275 280 285

aaa att gtc gaa gcg gct aaa aag gca aat gaa ggg ctt gga ctg caa 912
 Lys Ile Val Glu Ala Ala Lys Lys Ala Asn Glu Gly Leu Gly Leu Gln
 290 295 300

aat tgc gca aca cat aca gaa atc aag cta atg aaa aac aga gaa ccg 960
 Asn Cys Ala Thr His Thr Glu Ile Lys Leu Met Lys Asn Arg Glu Pro
 305 310 315 320

ggt tta ata gag tcg gct gcc aga ttc gca ggc tgg aat atg att cct 1008
 Gly Leu Ile Glu Ser Ala Ala Arg Phe Ala Gly Trp Asn Met Ile Pro
 325 330 335

aac att aaa aag gtt ttc ggc ctt gat atg gcg caa tta tta tta gat 1056

Asn Ile Lys Lys Val Phe Gly Leu Asp Met Ala Gln Leu Leu Leu Asp
 340 345 350

gtt ctc tgt ttc gga aaa gat gct gat ctg ccg gac ggg tta ttg gat 1104
 Val Leu Cys Phe Gly Lys Asp Ala Asp Leu Pro Asp Gly Leu Leu Asp
 355 360 365

caa gag cct tac tat gtt gct gac tgc cat ctg tac cct cag cat ttc 1152
 Gln Glu Pro Tyr Tyr Val Ala Asp Cys His Leu Tyr Pro Gln His Phe
 370 375 380

aaa caa aat ggc cag atc cct gaa act gcc gag gat ttg gta atc gaa 1200
 Lys Gln Asn Gly Gln Ile Pro Glu Thr Ala Glu Asp Leu Val Ile Glu
 385 390 395 400

gcg atc gat att ccg gat ggg ctt ttg aag ggt gat aca gaa atc gtt 1248
 Ala Ile Asp Ile Pro Asp Gly Leu Leu Lys Gly Asp Thr Glu Ile Val
 405 410 415

act ttt tcg gct gcg gca cca gga aca tca gtt gat ttg aca ctg ttt 1296
 Thr Phe Ser Ala Ala Ala Pro Gly Thr Ser Val Asp Leu Thr Leu Phe
 420 425 430

gaa gcc ttc aac tcc att gct gca ttt gaa ctg aaa ggc agc aat tca 1344
 Glu Ala Phe Asn Ser Ile Ala Ala Phe Glu Leu Lys Gly Ser Asn Ser
 435 440 445

cag gat gtg gct gaa tca atc aga caa att cag cag cat gcg aag ctg 1392
 Gln Asp Val Ala Glu Ser Ile Arg Gln Ile Gln Gln His Ala Lys Leu
 450 455 460

acg gca aag tat gtg ctg cca gta 1416
 Thr Ala Lys Tyr Val Leu Pro Val
 465 470

<210> 11
 <211> 1416
 <212> DNA
 <213> 枯草芽孢杆菌 IAM1213

<400> 11
 atg gag aga aaa aca gta ttg gtc atc gct gat ctt gga ggc tgc ccg 48
 Met Glu Arg Lys Thr Val Leu Val Ile Ala Asp Leu Gly Gly Cys Pro

1	5	10	15	
ccg cac atg ttt tat aaa agc gct gct gaa aaa tat aac ctg gtc agc				96
Pro His Met Phe Tyr Lys Ser Ala Ala Glu Lys Tyr Asn Leu Val Ser				
	20	25	30	
ttt att cca aga cct ttt gca att aca gcc tcc cat gca gca ttg att				144
Phe Ile Pro Arg Pro Phe Ala Ile Thr Ala Ser His Ala Ala Leu Ile				
	35	40	45	
gaa aaa tac tcg gtc gcg gtc ata aaa gat aaa gac tat ttt aag agt				192
Glu Lys Tyr Ser Val Ala Val Ile Lys Asp Lys Asp Tyr Phe Lys Ser				
	50	55	60	
tta gct gat ttt gag cat cct gac tcc att tat tgg gcg cat gag gat				240
Leu Ala Asp Phe Glu His Pro Asp Ser Ile Tyr Trp Ala His Glu Asp				
	65	70	75	80
cat aac aag cct gag gaa gag gtc gtc gag caa atc gtc aag gtt gcc				288
His Asn Lys Pro Glu Glu Glu Val Val Glu Gln Ile Val Lys Val Ala				
	85	90	95	
gaa atg ttt ggg gcg gat gcc atc aca aca aac aat gaa tta ttc att				336
Glu Met Phe Gly Ala Asp Ala Ile Thr Thr Asn Asn Glu Leu Phe Ile				
	100	105	110	
gct ccg atg gcg aaa gcc tgt gaa cgt ctg ggc ctg aga ggt gcc ggc				384
Ala Pro Met Ala Lys Ala Cys Glu Arg Leu Gly Leu Arg Gly Ala Gly				
	115	120	125	
gtg cag gca gcc gaa aat gcc aga gat aaa aat aaa atg agg gac gct				432
Val Gln Ala Ala Glu Asn Ala Arg Asp Lys Asn Lys Met Arg Asp Ala				
	130	135	140	
ttt aat aag gcc gga gtc aaa tcg atc aaa aac aaa cga gtc aca act				480
Phe Asn Lys Ala Gly Val Lys Ser Ile Lys Asn Lys Arg Val Thr Thr				
	145	150	155	160
ctc gaa gat ttc cgt get gct ctt gaa gag atc ggc aca cct ctt atc				528
Leu Glu Asp Phe Arg Ala Ala Leu Glu Glu Ile Gly Thr Pro Leu Ile				
	165	170	175	
tta aag cct aca tac tta gcg agt tca atc ggt gta acg ctg att acg				576
Leu Lys Pro Thr Tyr Leu Ala Ser Ser Ile Gly Val Thr Leu Ile Thr				

gac act gag acg gca gaa gat gaa ttt aac aga gtc aat gac tat ctg	624
Asp Thr Glu Thr Ala Glu Asp Glu Phe Asn Arg Val Asn Asp Tyr Leu	
195 200 205	
aaa tca att aac gtg cca aag gcg gtt acg ttt gaa gcg ccg ttt atc	672
Lys Ser Ile Asn Val Pro Lys Ala Val Thr Phe Glu Ala Pro Phe Ile	
210 215 220	
gct gaa gaa ttt tta cag ggt gag tac gga gac tgg tat caa aca gaa	720
Ala Glu Glu Phe Leu Gln Gly Glu Tyr Gly Asp Trp Tyr Gln Thr Glu	
225 230 235 240	
ggg tac tcc gac tat atc agt ata gaa ggc atc atg gct gac ggt gag	768
Gly Tyr Ser Asp Tyr Ile Ser Ile Glu Gly Ile Met Ala Asp Gly Glu	
245 250 255	
tat ttc ccg atc gcc att cat gat aaa acg ccg caa atc ggg ttt aca	816
Tyr Phe Pro Ile Ala Ile His Asp Lys Thr Pro Gln Ile Gly Phe Thr	
260 265 270	
gag aca tcc cac att acg ccg tcc att ctg gat gaa gag gca aaa aag	864
Glu Thr Ser His Ile Thr Pro Ser Ile Leu Asp Glu Glu Ala Lys Lys	
275 280 285	
aaa att gtc gaa gct gcc aaa aag gca aat gaa ggt ctt ggc ctg caa	912
Lys Ile Val Glu Ala Ala Lys Lys Ala Asn Glu Gly Leu Gly Leu Gln	
290 295 300	
aat tgc gca aca cat aca gag atc aag cta atg aaa aat aga gaa ccg	960
Asn Cys Ala Thr His Thr Glu Ile Lys Leu Met Lys Asn Arg Glu Pro	
305 310 315 320	
ggt tta att gaa tcg gca gcc aga ttc gcc ggc tgg aat atg atc ccc	1008
Gly Leu Ile Glu Ser Ala Ala Arg Phe Ala Gly Trp Asn Met Ile Pro	
325 330 335	
aat att aaa aag gtc ttt ggc ctt gat atg gcg caa tta tta tta gat	1056
Asn Ile Lys Lys Val Phe Gly Leu Asp Met Ala Gln Leu Leu Leu Asp	
340 345 350	
gtc ctt tgt ttc gga aaa gac gcc gat ctg ccg gac gga tta ttg gat	1104
Val Leu Cys Phe Gly Lys Asp Ala Asp Leu Pro Asp Gly Leu Leu Asp	

355	360	365	
caa gag cct tat tat gtt gcc gac tgc cat ttg tac ccg caa cat ttc			1152
Gln Glu Pro Tyr Tyr Val Ala Asp Cys His Leu Tyr Pro Gln His Phe			
370	375	380	
aaa caa aat ggc cag att cca gaa act gct gag gat ttg gtc att gaa			1200
Lys Gln Asn Gly Gln Ile Pro Glu Thr Ala Glu Asp Leu Val Ile Glu			
385	390	395	400
gcg atc gat ctg cct gac ggg ctt tta aaa ggg gat act gag atc gtt			1248
Ala Ile Asp Leu Pro Asp Gly Leu Leu Lys Gly Asp Thr Glu Ile Val			
405	410	415	
tct ttt tcg gcc gca gca cca gga act tca gtt gat ttg aca ttg ttt			1296
Ser Phe Ser Ala Ala Ala Pro Gly Thr Ser Val Asp Leu Thr Leu Phe			
420	425	430	
gaa gct ttc aat tcc att gct gca ttt gaa ctg aaa ggc agt aat tca			1344
Glu Ala Phe Asn Ser Ile Ala Ala Phe Glu Leu Lys Gly Ser Asn Ser			
435	440	445	
cag gat gtg gct gaa tca atc aga caa att cag cag cat gcg aag ctg			1392
Gln Asp Val Ala Glu Ser Ile Arg Gln Ile Gln Gln His Ala Lys Leu			
450	455	460	
acg gca aag tat gtg ctg cca gta			1416
Thr Ala Lys Tyr Val Leu Pro Val			
465	470		
<p><210> 12 <211> 1416 <212> DNA <213> 枯草芽孢杆菌 IAM1107</p>			
<p><400> 12</p>			
atg gag aga aaa aca gta ttg gtc atc gct gat ctt gga ggc tgc ccg			48
Met Glu Arg Lys Thr Val Leu Val Ile Ala Asp Leu Gly Gly Cys Pro			
1	5	10	15
ccg cac atg ttt tat aaa agc gct gct gaa aaa tat aac ctg gtc agc			96
Pro His Met Phe Tyr Lys Ser Ala Ala Glu Lys Tyr Asn Leu Val Ser			
20	25	30	

ttt att cca aga cct ttt gca att aca gcc tcc cat gca gca ttg att	144
Phe Ile Pro Arg Pro Phe Ala Ile Thr Ala Ser His Ala Ala Leu Ile	
35 40 45	
gaa aaa tac tcg gtc gcg gtc gta aaa gat aaa gac tat ttt aag agt	192
Glu Lys Tyr Ser Val Ala Val Val Lys Asp Lys Asp Tyr Phe Lys Ser	
50 55 60	
tta gct gat ttt gag cat cct gac tcc att tat tgg gcg cat gag gat	240
Leu Ala Asp Phe Glu His Pro Asp Ser Ile Tyr Trp Ala His Glu Asp	
65 70 75 80	
cat aac aag cct gag gaa gag gtc gtc gag caa atc gtc aag git gcc	288
His Asn Lys Pro Glu Glu Glu Val Val Glu Gln Ile Val Lys Val Ala	
85 90 95	
gaa atg ttc ggg gcg gat gcc atc aca aca aac aat gaa tta ttc att	336
Glu Met Phe Gly Ala Asp Ala Ile Thr Thr Asn Asn Glu Leu Phe Ile	
100 105 110	
gct ccg atg gcg aaa gcc tgt gaa cgt ctg ggc ttg aga ggt gcc ggc	384
Ala Pro Met Ala Lys Ala Cys Glu Arg Leu Gly Leu Arg Gly Ala Gly	
115 120 125	
gtg cag gca gcc gaa aat gcc aga gat aaa aat aaa atg agg gac gct	432
Val Gln Ala Ala Glu Asn Ala Arg Asp Lys Asn Lys Met Arg Asp Ala	
130 135 140	
ttt aat aag gcc gga gtc aaa tcg atc aaa aac aaa cga gtc aca act	480
Phe Asn Lys Ala Gly Val Lys Ser Ile Lys Asn Lys Arg Val Thr Thr	
145 150 155 160	
ctt gaa gat ttc cgt gct gct ctt gaa gag atc ggc aca cct ctt atc	528
Leu Glu Asp Phe Arg Ala Ala Leu Glu Glu Ile Gly Thr Pro Leu Ile	
165 170 175	
tta aag cct aca tac tta gcg agt tct atc ggt gta acg ctg att acg	576
Leu Lys Pro Thr Tyr Leu Ala Ser Ser Ile Gly Val Thr Leu Ile Thr	
180 185 190	
gac act gag acg gca gaa gat gaa tit aac aga gtc aat gac tat ctg	624
Asp Thr Glu Thr Ala Glu Asp Glu Phe Asn Arg Val Asn Asp Tyr Leu	
195 200 205	

aaa tca att aac gtg cca aag gcg gtt acg ttt gaa gcg ccg ttt atc 672
 Lys Ser Ile Asn Val Pro Lys Ala Val Thr Phe Glu Ala Pro Phe Ile
 210 215 220

gct gaa gaa ttt tta cag ggt gag tac gga gac tgg tat caa aca gaa 720
 Ala Glu Glu Phe Leu Gln Gly Glu Tyr Gly Asp Trp Tyr Gln Thr Glu
 225 230 235 240

ggg tac tcc gac tat atc agt ata gaa ggc atc atg gct gac ggt gag 768
 Gly Tyr Ser Asp Tyr Ile Ser Ile Glu Gly Ile Met Ala Asp Gly Glu
 245 250 255

tat ttc ccg atc gcc att cat gat aaa acg ccg caa atc ggg ttt aca 816
 Tyr Phe Pro Ile Ala Ile His Asp Lys Thr Pro Gln Ile Gly Phe Thr
 260 265 270

gag aca tcc cac att acg ccg tcc att ctg gat gaa gag gca aaa aag 864
 Glu Thr Ser His Ile Thr Pro Ser Ile Leu Asp Glu Glu Ala Lys Lys
 275 280 285

aaa att gtc gaa gct gcc aaa aag gca aat gaa ggg ctt ggc ctg caa 912
 Lys Ile Val Glu Ala Ala Lys Lys Ala Asn Glu Gly Leu Gly Leu Gln
 290 295 300

aat tgc gca aca cat aca gag gtc aag cta atg aaa aac aga gaa ccg 960
 Asn Cys Ala Thr His Thr Glu Val Lys Leu Met Lys Asn Arg Glu Pro
 305 310 315 320

ggt tta att gaa tgc gca gcc aga ttt gcc ggc tgg aat atg atc cct 1008
 Gly Leu Ile Glu Ser Ala Ala Arg Phe Ala Gly Trp Asn Met Ile Pro
 325 330 335

aac att aaa aag gtt ttc ggc ctt gat atg gcg caa tta tta tta gat 1056
 Asn Ile Lys Lys Val Phe Gly Leu Asp Met Ala Gln Leu Leu Leu Asp
 340 345 350

gtc ctc tgt ttc gga aaa gat gcc gat ctg ccg gac gga tta ttg gat 1104
 Val Leu Cys Phe Gly Lys Asp Ala Asp Leu Pro Asp Gly Leu Leu Asp
 355 360 365

caa gag cct tac tat gtc gcc gac tgc cat ttg tac ccg cag cat ttc 1152
 Gln Glu Pro Tyr Tyr Val Ala Asp Cys His Leu Tyr Pro Gln His Phe
 370 375 380

aaa caa aat ggc cag att cca gaa acc gct gag gat ttg gtc att gaa 1200
 Lys Gln Asn Gly Gln Ile Pro Glu Thr Ala Glu Asp Leu Val Ile Glu
 385 390 395 400

gcg atc gat att ccg gac ggg ctt tta aaa ggg gat act gaa atc ttt 1248
 Ala Ile Asp Ile Pro Asp Gly Leu Leu Lys Gly Asp Thr Glu Ile Phe
 405 410 415

tct ttt tcg gcc gca gca cca ggc act tca gtt gat ttg aca ttg ttt 1296
 Ser Phe Ser Ala Ala Ala Pro Gly Thr Ser Val Asp Leu Thr Leu Phe
 420 425 430

gaa gct ttc aat tcc att gct gca ttt gaa ctg aaa ggc agt aat tca 1344
 Glu Ala Phe Asn Ser Ile Ala Ala Phe Glu Leu Lys Gly Ser Asn Ser
 435 440 445

cag gat gtg gct gaa tca atc aga caa att cag cag cat gcg aag ctg 1392
 Gln Asp Val Ala Glu Ser Ile Arg Gln Ile Gln Gln His Ala Lys Leu
 450 455 460

acg gca aag tat gtg ctg cca gta 1416
 Thr Ala Lys Tyr Val Leu Pro Val
 465 470

<210> 13

<211> 1416

<212> DNA

<213> 枯草芽孢杆菌 IAM1214

<400> 13

atg gag aga aaa aca gta ttg gtc atc gct gat ctt gga ggc tgc ccg 48
 Met Glu Arg Lys Thr Val Leu Val Ile Ala Asp Leu Gly Gly Cys Pro
 1 5 10 15

ccg cac atg ttt tat aaa agc gct gct gaa aaa tat aac ctg gtt agc 96
 Pro His Met Phe Tyr Lys Ser Ala Ala Glu Lys Tyr Asn Leu Val Ser
 20 25 30

ttt att ccg aga cct ttt gca ata aca gcc tcc cat gca gca ctg att 144
 Phe Ile Pro Arg Pro Phe Ala Ile Thr Ala Ser His Ala Ala Leu Ile
 35 40 45

gaa aaa tac tcg gtc gcg gtc ata aaa gat aaa gac tat ttt cag agc 192
 Glu Lys Tyr Ser Val Ala Val Ile Lys Asp Lys Asp Tyr Phe Gln Ser
 50 55 60

tta gct gat ttt gag cat ccc gat tca att tat tgg gcg cat gag gat 240
 Leu Ala Asp Phe Glu His Pro Asp Ser Ile Tyr Trp Ala His Glu Asp
 65 70 75 80

cat gac aag cct gaa gaa gag gtt gtc gag caa atc gtc aag gtt gcc 288
 His Asp Lys Pro Glu Glu Glu Val Val Glu Gln Ile Val Lys Val Ala
 85 90 95

caa atg ttt gag gcg gac gcc atc aca aca aac aat gaa tta ttc att 336
 Gln Met Phe Glu Ala Asp Ala Ile Thr Thr Asn Asn Glu Leu Phe Ile
 100 105 110

gcc ccg atg gcg aaa gcc tgt gaa cgc ctt ggc ctg agg gcc gcc gga 384
 Ala Pro Met Ala Lys Ala Cys Glu Arg Leu Gly Leu Arg Gly Ala Gly
 115 120 125

gtg cag gca gcg gaa aat gcc aga gat aaa aat aaa atg agg gac gct 432
 Val Gln Ala Ala Glu Asn Ala Arg Asp Lys Asn Lys Met Arg Asp Ala
 130 135 140

ttt aat aag gcg gga gtc aaa tcg atc aaa aac aaa cga gtc aca act 480
 Phe Asn Lys Ala Gly Val Lys Ser Ile Lys Asn Lys Arg Val Thr Thr
 145 150 155 160

ctt gag gat ttt cgt gct gca ctt gaa gag atc ggc aca cct cta atc 528
 Leu Glu Asp Phe Arg Ala Ala Leu Glu Glu Ile Gly Thr Pro Leu Ile
 165 170 175

tta aag cct aca tac tta geg agt tca atc ggc gta acg ctg att acc 576
 Leu Lys Pro Thr Tyr Leu Ala Ser Ser Ile Gly Val Thr Leu Ile Thr
 180 185 190

gac acg gag acg gca gaa gat gaa ttt aac aga gtc aat gac tac ctg 624
 Asp Thr Glu Thr Ala Glu Asp Glu Phe Asn Arg Val Asn Asp Tyr Leu
 195 200 205

aaa tcg att aac gtg ccg aag gcg gtc aca ttt gaa gca ccg ttt att 672
 Lys Ser Ile Asn Val Pro Lys Ala Val Thr Phe Glu Ala Pro Phe Ile
 210 215 220

gct gag gaa ttt tta cag ggt gag tac gga gac tgg tat caa aca gaa 720
 Ala Glu Glu Phe Leu Gln Gly Glu Tyr Gly Asp Trp Tyr Gln Thr Glu
 225 230 235 240

ggg tac tcc gac tat atc agc ata gaa ggc att atg gca gat ggt gag 768
 Gly Tyr Ser Asp Tyr Ile Ser Ile Glu Gly Ile Met Ala Asp Gly Glu
 245 250 255

tat ttt ccg atc gcc att cat gac aaa acg ccg caa att gga ttt aca 816
 Tyr Phe Pro Ile Ala Ile His Asp Lys Thr Pro Gln Ile Gly Phe Thr
 260 265 270

gag aca tca cat att acg cca tcc att ctg gat gaa gag gcg aaa aag 864
 Glu Thr Ser His Ile Thr Pro Ser Ile Leu Asp Glu Glu Ala Lys Lys
 275 280 285

aaa att gtc gaa gcg gct aaa aag gca aat gaa ggg ctt gga ctg caa 912
 Lys Ile Val Glu Ala Ala Lys Lys Ala Asn Glu Gly Leu Gly Leu Gln
 290 295 300

aat tgc gca aca cat aca gaa atc aag cta atg aaa aac aga gaa ccg 960
 Asn Cys Ala Thr His Thr Glu Ile Lys Leu Met Lys Asn Arg Glu Pro
 305 310 315 320

ggt tta ata gag tcg gct gcc aga ttc gca ggc tgg aat atg att cct 1008
 Gly Leu Ile Glu Ser Ala Ala Arg Phe Ala Gly Trp Asn Met Ile Pro
 325 330 335

aac att aaa aag gtt ttc ggc ctt gat atg gcg caa tta tta tta gat 1056
 Asn Ile Lys Lys Val Phe Gly Leu Asp Met Ala Gln Leu Leu Leu Asp
 340 345 350

gtt ctc tgt ttc gga aaa gat gct gat ctg ccg gac ggg tta ttg gat 1104
 Val Leu Cys Phe Gly Lys Asp Ala Asp Leu Pro Asp Gly Leu Leu Asp
 355 360 365

caa gag cct tac tat gtt gct gac tgc cat ctg tac cct cag cat ttc 1152
 Gln Glu Pro Tyr Tyr Val Ala Asp Cys His Leu Tyr Pro Gln His Phe
 370 375 380

aaa caa aat ggc cag atc cct gaa act gcc gag gat ttg gta atc gaa 1200
 Lys Gln Asn Gly Gln Ile Pro Glu Thr Ala Glu Asp Leu Val Ile Glu
 385 390 395 400

gcg atc gat att ccg gat ggg ctt ttg aag ggt gat aca gaa atc gtt 1248
 Ala Ile Asp Ile Pro Asp Gly Leu Leu Lys Gly Asp Thr Glu Ile Val
 405 410 415

act ttt tcg gct gcg gca cca gga aca tca gtt gat ttg aca ctg ttt 1296
 Thr Phe Ser Ala Ala Ala Pro Gly Thr Ser Val Asp Leu Thr Leu Phe
 420 425 430

gaa gcc ttc aac tcc att gct gca ttt gaa ctg aaa ggc agc aat tca 1344
 Glu Ala Phe Asn Ser Ile Ala Ala Phe Glu Leu Lys Gly Ser Asn Ser
 435 440 445

cag gat gtg gct gaa tca atc aga caa att cag cag cat gcg aag ctg 1392
 Gln Asp Val Ala Glu Ser Ile Arg Gln Ile Gln Gln His Ala Lys Leu
 450 455 460

acg gca aag tat gtg ctg cca gta 1416
 Thr Ala Lys Tyr Val Leu Pro Val
 465 470

<210> 14
 <211> 1416
 <212> DNA
 <213> 枯草芽孢杆菌 ATCC21555

<400> 14
 atg gag aga aaa aca gta ttg gtt atc gct gat ctt ggg ggc tgc ccg 48
 Met Glu Arg Lys Thr Val Leu Val Ile Ala Asp Leu Gly Gly Cys Pro
 1 5 10 15

ccg cat atg ttt tac aaa agc gca gcc gaa aaa tac aac ctc gtc agc 96
 Pro His Met Phe Tyr Lys Ser Ala Ala Glu Lys Tyr Asn Leu Val Ser
 20 25 30

ttt att ccg aga ccc ttt gca att aca gcc tct cat gcg gcc tta att 144
 Phe Ile Pro Arg Pro Phe Ala Ile Thr Ala Ser His Ala Ala Leu Ile
 35 40 45

gaa aaa tac tcg att gcg gtc att aaa gat aaa gac tat ttt aag agt 192
 Glu Lys Tyr Ser Ile Ala Val Ile Lys Asp Lys Asp Tyr Phe Lys Ser
 50 55 60

ctg gct gat ttt gaa cat ccc gat tcg att tat tgg gct cat gaa gat 240

Gly Tyr Ser Asp Tyr Ile Ser Ile Glu Gly Ile Met Ala Asp Gly Glu	
245	250
255	
tac ttc ccc gtt gcg atc cat gat aaa aca ccg caa atc gga ttc acg	816
Tyr Phe Pro Val Ala Ile His Asp Lys Thr Pro Gln Ile Gly Phe Thr	
260	265
270	
gag aca gcg cat att acg ccg tcc atc ctg gat gat gac gcc aag cgg	864
Glu Thr Ala His Ile Thr Pro Ser Ile Leu Asp Asp Asp Ala Lys Arg	
275	280
285	
aaa atc gtc gaa gct gcc aag aag gcg aat gaa gga ctc ggc ctc gaa	912
Lys Ile Val Glu Ala Ala Lys Lys Ala Asn Glu Gly Leu Gly Leu Glu	
290	295
300	
aac tgt gca acg cat aca gaa ata aaa tta atg aaa aac cgg gaa gcc	960
Asn Cys Ala Thr His Thr Glu Ile Lys Leu Met Lys Asn Arg Glu Ala	
305	310
315	320
gga ctg att gag tca gcg gcc aga ttc gcg gga tgg aat atg att ccg	1008
Gly Leu Ile Glu Ser Ala Ala Arg Phe Ala Gly Trp Asn Met Ile Pro	
325	330
335	
aat att aaa aag gtc ttc ggc gtt gat atg gcg cag cta tta ttg gat	1056
Asn Ile Lys Lys Val Phe Gly Val Asp Met Ala Gln Leu Leu Leu Asp	
340	345
350	
gtt ctc tgt tac gga aaa gaa gct gat ctg ccg aaa gga tta ttg gag	1104
Val Leu Cys Tyr Gly Lys Glu Ala Asp Leu Pro Lys Gly Leu Leu Glu	
355	360
365	
cag gag cca tgc tat gtc gca gac tgc cac ttg tat cct cag cat ttc	1152
Gln Glu Pro Cys Tyr Val Ala Asp Cys His Leu Tyr Pro Gln His Phe	
370	375
380	
aaa gag aac ggc cag ctg cct gag acg gtt gtc gat ttc gtc att gaa	1200
Lys Glu Asn Gly Gln Leu Pro Glu Thr Val Val Asp Phe Val Ile Glu	
385	390
395	400
agc att gaa att cct gac ggc gtc tta aag gga gac act gaa ctc gtt	1248
Ser Ile Glu Ile Pro Asp Gly Val Leu Lys Gly Asp Thr Glu Leu Val	
405	410
415	
tct ttc tca gcg gct gag gcg ggt acg tca gtg gat ctg cgg ctg ttc	1296

Ser Phe Ser Ala Ala Glu Ala Gly Thr Ser Val Asp Leu Arg Leu Phe
420 425 430

gaa gcg ttc aac agc att gcg gcg ttt gag ctg aaa gga agc aat tcg 1344
Glu Ala Phe Asn Ser Ile Ala Ala Phe Glu Leu Lys Gly Ser Asn Ser
435 440 445

aac gac gtg gcc gaa tca atc aaa caa att cag cag cag gcg aag ctg 1392
Asn Asp Val Ala Glu Ser Ile Lys Gln Ile Gln Gln Gln Ala Lys Leu
450 455 460

act gca aag tat gcg tta tcg gta 1416
Thr Ala Lys Tyr Ala Leu Ser Val

<210> 15
<211> 1416
<212> DNA
<213> 解淀粉芽孢杆菌 IF03022

<400> 15

atg gag aga aaa aca gta ttg gtt atc gct gac ctt ggg gga tgc ccg 48
Met Glu Arg Lys Thr Val Leu Val Ile Ala Asp Leu Gly Gly Cys Pro
1 5 10 15

ccg cat atg ttt tac aaa agc gca gcc gaa aaa tac aac ctc gtc agc 96
Pro His Met Phe Tyr Lys Ser Ala Ala Glu Lys Tyr Asn Leu Val Ser
20 25 30

ttt att ccg aga cct ttt gca att aca gcc tct cat gcg gca tta att 144
Phe Ile Pro Arg Pro Phe Ala Ile Thr Ala Ser His Ala Ala Leu Ile
35 40 45

gaa aaa tac tcg gtc gcg gtc ata aaa gat aaa gac tat ttt aag agt 192
Glu Lys Tyr Ser Val Ala Val Ile Lys Asp Lys Asp Tyr Phe Lys Ser
50 55 60

ctg gct gat ttt gag cat ccc gat tcg att tac tgg gct cat gaa gat 240
Leu Ala Asp Phe Glu His Pro Asp Ser Ile Tyr Trp Ala His Glu Asp
65 70 75 80

cat gac aaa cct gag gaa gaa gta gtc gaa gaa atc gtc aag gtg gcc 288
His Asp Lys Pro Glu Glu Glu Val Val Glu Glu Ile Val Lys Val Ala
85 90 95

ggc atg ttc gcg gtt gac gcc att acg acc aac aat gaa ctg ttt atc 336
 Gly Met Phe Ala Val Asp Ala Ile Thr Thr Asn Asn Glu Leu Phe Ile
 100 105 110

gct ccg atg gca aaa gcg tgt gaa cgt etc ggc ctg cgg gga gcg ggc 384
 Ala Pro Met Ala Lys Ala Cys Glu Arg Leu Gly Leu Arg Gly Ala Gly
 115 120 125

gta cag gcc gct gaa aat gcc aga gat aaa aac aaa atg aga gcc gct 432
 Val Gln Ala Ala Glu Asn Ala Arg Asp Lys Asn Lys Met Arg Ala Ala
 130 135 140

ttc aac cgg gcc ggc gtc aag tct atc aaa aac aga cgg gtg acg acg 480
 Phe Asn Arg Ala Gly Val Lys Ser Ile Lys Asn Arg Arg Val Thr Thr
 145 150 155 160

ctg gaa gat ttc cgc gcc gcg ctt cag gaa atc gga acg ccg etc att 528
 Leu Glu Asp Phe Arg Ala Ala Leu Gln Glu Ile Gly Thr Pro Leu Ile
 165 170 175

ctg aag cct aca tat ctg gcg agc tcc atc ggc gtg acg etc atc aaa 576
 Leu Lys Pro Thr Tyr Leu Ala Ser Ser Ile Gly Val Thr Leu Ile Lys
 180 185 190

gag agg gaa acg gcc gaa gcc gaa ttt aac aga gtc aat gaa tac ctg 624
 Glu Arg Glu Thr Ala Glu Ala Glu Phe Asn Arg Val Asn Glu Tyr Leu
 195 200 205

aag tcg atc aac gta ccg aaa gcg gtc acg ttt gaa gcg ccg ttt atc 672
 Lys Ser Ile Asn Val Pro Lys Ala Val Thr Phe Glu Ala Pro Phe Ile
 210 215 220

gcg gaa gaa ttt ttg cag ggc gag tat gac gac tgg tac gaa aca agc 720
 Ala Glu Glu Phe Leu Gln Gly Glu Tyr Asp Asp Trp Tyr Glu Thr Ser
 225 230 235 240

ggt tat tcc gac tat atc agc ata gaa ggc atc atg gcc gac gga gaa 768
 Gly Tyr Ser Asp Tyr Ile Ser Ile Glu Gly Ile Met Ala Asp Gly Glu
 245 250 255

tac ttc cct gtc gca att cat gat aaa aca ccg caa atc gga ttc acg 816
 Tyr Phe Pro Val Ala Ile His Asp Lys Thr Pro Gln Ile Gly Phe Thr
 260 265 270

gag aca tgc cat att acg ccg tcc atc ctg gat gat gac gcg aag cgg 864
 Glu Thr Ser His Ile Thr Pro Ser Ile Leu Asp Asp Asp Ala Lys Arg
 275 280 285

aaa atc gtc gaa gca gcc aaa aag gcg aat gaa gga ctc ggc ctc gaa 912
 Lys Ile Val Glu Ala Ala Lys Lys Ala Asn Glu Gly Leu Gly Leu Glu
 290 295 300

aac tgc gca acc cat aca gag att aaa tta atg aaa aac cgg gaa gcc 960
 Asn Cys Ala Thr His Thr Glu Ile Lys Leu Met Lys Asn Arg Glu Ala
 305 310 315 320

gga ctg att gaa tca gcg gca cga ttt gcg ggc tgg aac atg att ccg 1008
 Gly Leu Ile Glu Ser Ala Ala Arg Phe Ala Gly Trp Asn Met Ile Pro
 325 330 335

aat att aaa aag gtc ttc ggc gtc gat atg gcg cag ctg tta ttg gat 1056
 Asn Ile Lys Lys Val Phe Gly Val Asp Met Ala Gln Leu Leu Leu Asp
 340 345 350

gtt ctc tgt ttc gga aaa gaa gcc gat ctg ccg aaa gga tta ttg gag 1104
 Val Leu Cys Phe Gly Lys Glu Ala Asp Leu Pro Lys Gly Leu Leu Glu
 355 360 365

cag gag ccg tgc tat gtc gcc gac tgc cac ttg tat cct cag cat ttc 1152
 Gln Glu Pro Cys Tyr Val Ala Asp Cys His Leu Tyr Pro Gln His Phe
 370 375 380

aaa gag aac ggc cag ctg cct gag acg gct gtc gat ttc gtc att gaa 1200
 Lys Glu Asn Gly Gln Leu Pro Glu Thr Ala Val Asp Phe Val Ile Glu
 385 390 395 400

agc att gac att ccc gac ggc gtc tta aag gga gac acc gaa atc gtt 1248
 Ser Ile Asp Ile Pro Asp Gly Val Leu Lys Gly Asp Thr Glu Ile Val
 405 410 415

tct ttc tgc gcg gcc gag gcg ggt aca tcc gtg gat ctg cgg ctg ttc 1296
 Ser Phe Ser Ala Ala Glu Ala Gly Thr Ser Val Asp Leu Arg Leu Phe
 420 425 430

gaa gcg ttc aac agc att gcg gcg ttc gag ctg aaa gga agc aat tgc 1344
 Glu Ala Phe Asn Ser Ile Ala Ala Phe Glu Leu Lys Gly Ser Asn Ser
 435 440 445

ggt gac gtg gcc gaa tca atc aaa caa att cag cag cag gcg aag ctg 1392
 Gly Asp Val Ala Glu Ser Ile Lys Gln Ile Gln Gln Gln Ala Lys Leu
 450 455 460

act gca aag tat gcg tta ccg gta 1416
 Thr Ala Lys Tyr Ala Leu Pro Val

<210> 16
 <211> 1428
 <212> DNA
 <213> 短小芽孢杆菌 NRRL B-12025

<400> 16
 gtg ctt tca ttg agt aaa aaa act gta ctt gtc att gct gac tta gga 48
 Val Leu Ser Leu Ser Lys Lys Thr Val Leu Val Ile Ala Asp Leu Gly
 1 5 10 15

ggg tgc ccg ccc cat atg ttt tat gaa agc gtg gcg gca tca tac cat 96
 Gly Cys Pro Pro His Met Phe Tyr Glu Ser Val Ala Ala Ser Tyr His
 20 25 30

atc gtt tct tat atc cca aga ccc ttt gcg att aca aag gga cat gcc 144
 Ile Val Ser Tyr Ile Pro Arg Pro Phe Ala Ile Thr Lys Gly His Ala
 35 40 45

gag cta atc gaa aaa tac tcc att gcc gtc atc aaa gac cgt gat tat 192
 Glu Leu Ile Glu Lys Tyr Ser Ile Ala Val Ile Lys Asp Arg Asp Tyr
 50 55 60

ttt gag aca cac cct tct ttt gaa cac cct gat tct att tac tgg gca 240
 Phe Glu Thr His Pro Ser Phe Glu His Pro Asp Ser Ile Tyr Trp Ala
 65 70 75 80

cat gat gat tat cca aaa tca gaa gaa gaa gtt gtg gaa gac ttc att 288
 His Asp Asp Tyr Pro Lys Ser Glu Glu Glu Val Val Glu Asp Phe Ile
 85 90 95

cga gta gct tcc ttt ttc aaa gca gat gca atc acg acc aat aat gaa 336
 Arg Val Ala Ser Phe Phe Lys Ala Asp Ala Ile Thr Thr Asn Asn Glu
 100 105 110

tta ttc att gca ccg atg gca aag gcc gct gaa cgt ctt ggg cta cga 384

Leu Phe Ile Ala Pro Met Ala Lys Ala Ala Glu Arg Leu Gly Leu Arg
 115 120 125

ggt gcc ggt gtc aag gca gcc gaa atg gcg cgt gat aaa agc caa atg 432
 Gly Ala Gly Val Lys Ala Ala Glu Met Ala Arg Asp Lys Ser Gln Met
 130 135 140

agg gct gca ttc aat gcc tct ggc gtc aaa gcg gtg aaa act cag cct 480
 Arg Ala Ala Phe Asn Ala Ser Gly Val Lys Ala Val Lys Thr Gln Pro
 145 150 155 160

gtc acg act tta tct gat ttc caa caa gcc att gag tct atc gga aca 528
 Val Thr Thr Leu Ser Asp Phe Gln Gln Ala Ile Glu Ser Ile Gly Thr
 165 170 175

ccg ctc att tta aag cct aca tat tta gcc agt tct att ggc gtc acc 576
 Pro Leu Ile Leu Lys Pro Thr Tyr Leu Ala Ser Ser Ile Gly Val Thr
 180 185 190

ttg ttt cat gac aaa gcc gga agt gat gac ttg ttt tta caa gta caa 624
 Leu Phe His Asp Lys Ala Gly Ser Asp Asp Leu Phe Leu Gln Val Gln
 195 200 205

tcg tat ttg gaa acc ata cca gtc cca gac gct gtc acg tat gaa gca 672
 Ser Tyr Leu Glu Thr Ile Pro Val Pro Asp Ala Val Thr Tyr Glu Ala
 210 215 220

ccg ttt gtc gct gaa aca tat tta gag ggt gct tac gaa gat tgg tat 720
 Pro Phe Val Ala Glu Thr Tyr Leu Glu Gly Ala Tyr Glu Asp Trp Tyr
 225 230 235 240

gaa gac gaa gga tat gct gat tat gtc agt gta gaa ggg ctg gtc gta 768
 Glu Asp Glu Gly Tyr Ala Asp Tyr Val Ser Val Glu Gly Leu Val Val
 245 250 255

gag ggc gaa tat ctc cct ttt gtc ata cat gat aaa acc cct caa atc 816
 Glu Gly Glu Tyr Leu Pro Phe Val Ile His Asp Lys Thr Pro Gln Ile
 260 265 270

ggc ttt aca gaa acg gct cat atc act ccg acg atc tta gac aat gaa 864
 Gly Phe Thr Glu Thr Ala His Ile Thr Pro Thr Ile Leu Asp Asn Glu
 275 280 285

gcc aag caa atc atc att gaa gca gca agg aag gca aat gaa ggg cta 912

Ala Lys Gln Ile Ile Ile Glu Ala Ala Arg Lys Ala Asn Glu Gly Leu
 290 295 300

ggt ctt gaa cat tgt gca acc cat aca gaa atc aaa ctc atg aaa aat 960
 Gly Leu Glu His Cys Ala Thr His Thr Glu Ile Lys Leu Met Lys Asn
 305 310 315 320

cga gaa act gga ctg atc gag gca gcg gct cga ttc gct ggc tgg aat 1008
 Arg Glu Thr Gly Leu Ile Glu Ala Ala Ala Arg Phe Ala Gly Trp Asn
 325 330 335

atg atc ccg aat att aaa aaa gtc ttt ggc gtc gat atg gcg aag cta 1056
 Met Ile Pro Asn Ile Lys Lys Val Phe Gly Val Asp Met Ala Lys Leu
 340 345 350

ttg att gat gta tta gtt gat ggt aaa aag gct gta ctg cca aaa cag 1104
 Leu Ile Asp Val Leu Val Asp Gly Lys Lys Ala Val Leu Pro Lys Gln
 355 360 365

ctg ctt tct gga cat aca ttt tat gta gcg gac tgc cac ctg tac cct 1152
 Leu Leu Ser Gly His Thr Phe Tyr Val Ala Asp Cys His Leu Tyr Pro
 370 375 380

cag cat ttt aaa gag agt ggg ctt atc ccg cct gaa gcc aca cat att 1200
 Gln His Phe Lys Glu Ser Gly Leu Ile Pro Pro Glu Ala Thr His Ile
 385 390 395 400

acc att gat cat gtg tct att ccg cag gaa gca ttc gtt gga gat act 1248
 Thr Ile Asp His Val Ser Ile Pro Gln Glu Ala Phe Val Gly Asp Thr
 405 410 415

gcg att gtc agt caa tca ttc cct gcc aaa ggg act att gtg gat ctt 1296
 Ala Ile Val Ser Gln Ser Phe Pro Ala Lys Gly Thr Ile Val Asp Leu
 420 425 430

gaa tta ttt gaa gct ttt aat gga atc gta tct ctt gaa tta aaa gga 1344
 Glu Leu Phe Glu Ala Phe Asn Gly Ile Val Ser Leu Glu Leu Lys Gly
 435 440 445

tca tcc tca caa gat gtt gcc gcg tcc atc cgc aac att cag aaa cag 1392
 Ser Ser Ser Gln Asp Val Ala Ala Ser Ile Arg Asn Ile Gln Lys Gln
 450 455 460

gca acg att cag tta atg gat gaa tta gtg aag gga 1428

Val Thr Thr Leu Glu Asp Phe Arg Ala Ala Leu Glu Glu Ile Gly Thr
 35 40 45

cct ctt atc tta aag cct aca tac tta gcg agt tct atc ggt gta acg 192
 Pro Leu Ile Leu Lys Pro Thr Tyr Leu Ala Ser Ser Ile Gly Val Thr
 50 55 60

ctg att acg gac act gag acg gca gaa gat gaa ttt aac aga gtc aat 240
 Leu Ile Thr Asp Thr Glu Thr Ala Glu Asp Glu Phe Asn Arg Val Asn
 65 70 75 80

gac tat ctg aaa tca att aac gtg cca aag gcg gtt acg 279
 Asp Tyr Leu Lys Ser Ile Asn Val Pro Lys Ala Val Thr
 85 90

<210> 19

<211> 30

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述: 合成的 DNA

<400> 19

attctcgagt agagaaggag tgttttacct 30

<210> 20

<211> 30

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述: 合成的 DNA

<400> 20

ttaggatcct catactggca gcacatactt 30

<210> 21

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述: 合成的 DNA

<400> 21

caagaattct catgtttgac agct

24

<210> 22

<211> 28

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述: 合成的 DNA

<400> 22

taactcgaga ttcccttttt acgtgaac

28

<210> 23

<211> 27

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述: 合成的 DNA

<400> 23

ttaacccatgg agagaaaaac agtattg

27

<210> 24

<211> 30

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述: 合成的 DNA

<400> 24

atatggatcc tactggcagc acatactttg

30

<210> 25

<211> 21

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述: 合成的 DNA

<400> 25

caccgcagac ggaggataca c

21

<210> 26

<211> 22

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述: 合成的 DNA

<400> 26

cggacgtcac ccaataatcg tg

22

<210> 27

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述: 合成的 DNA

<400> 27

ccgatggcra aagcstgtra acg

23

<210> 28

<211> 21

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述: 合成的 DNA

<400> 28

cggcagater gedttcttttc c

21

- <210> 29
<211> 26
<212> DNA
<213> 人工序列
- <220>
<223> 人工序列的描述: 合成的 DNA
- <400> 29
gctaggtcctt gaacattgtg caaccc 26
- <210> 30
<211> 23
<212> DNA
<213> 人工序列
- <220>
<223> 人工序列的描述: 合成的 DNA
- <400> 30
ggtgttccga tagactcaat ggc 23
- <210> 31
<211> 44
<212> DNA
<213> 人工序列
- <220>
<223> 人工序列的描述: 合成的 DNA
- <400> 31
catgccatgg agaaaaaac tgtacttgctc attgctgact tagg 44
- <210> 32
<211> 38
<212> DNA
<213> 人工序列
- <220>
<223> 人工序列的描述: 合成的 DNA

<400> 32

cgcggatccc ttcaactaatt catccattaa ctgaatcg

38

<210> 33

<211> 12

<212> PRT

<213> 人工序列

<220>

<221> UNSURE

<222> (3)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<221> UNSURE

<222> (4)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<221> UNSURE

<222> (9)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<221> UNSURE

<222> (10)

<223> Xaa 代表选自下列的任何氨基酸: Leu, Ile, Val, Met 和 Ala

<220>

<221> UNSURE

<222> (11)

<223> Xaa 代表 Glu, Ser 或 Ala

<220>

<221> UNSURE

<222> (12)

<223> Xaa 代表 Gly, Ser 或 Ala

<220>

<223> 人工序列的描述: 用于数据库搜索的氨基酸序列

<400> 33

His Gly Xaa Xaa Gly Gln Asp Gly Xaa Xaa Xaa Xaa
5 10

<210> 34

<211> 28

<212> PRT

<213> 人工序列

<220>

<221> UNSURE

<222> (1)

<223> Xaa 代表 Leu Ile 或 Val

<220>

<221> UNSURE

<222> (2)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<221> UNSURE

<222> (3)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<221> UNSURE

<222> (4)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<221> UNSURE

<222> (5)

<223> Xaa 代表 Gly 或 Ala

<220>

<221> UNSURE

<222> (6)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<221> UNSURE

<222> (7)

<223> Xaa 代表选自下列的任何氨基酸: Gly, Ser, Ala, Ile 和 Val

<220>

<221> UNSURE

<222> (9)

<223> Xaa 代表选自下列的任何氨基酸: Leu, Ile, Val, Met, Cys 和 Ala

<220>

<221> UNSURE

<222> (11)

<223> Xaa 代表选自下列的任何氨基酸: Leu, Ile, Val, Met, Phe 和 Ala

<220>

<221> UNSURE

<222> (12)

<223> Xaa 代表选自下列的任何氨基酸: Leu, Ile, Val, Met, Phe 和 Ala

<220>

<221> UNSURE

<222> (13)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<221> UNSURE

<222> (14)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<221> UNSURE

<222> (15)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<221> UNSURE

<222> (16)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<221> UNSURE

<222> (17)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<221> UNSURE

<222> (18)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<221> UNSURE

<222> (19)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<221> UNSURE

<222> (20)

<223> Xaa 代表 Leu, Ile 或 Val

<220>

<221> UNSURE

<222> (21)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<221> UNSURE

<222> (23)

<223> Xaa 代表选自下列的任何氨基酸: Leu, Ile, Val, Ala 和 Pro

<220>

<221> UNSURE

<222> (25)

<223> Xaa 代表 Ser, Thr 或 Pro

<220>

<221> UNSURE

<222> (26)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His,

Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<221> UNSURE

<222> (5)

<223> Xaa 代表 Gly 或 Ala

<220>

<221> UNSURE

<222> (6)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<221> UNSURE

<222> (7)

<223> Xaa 代表选自下列的任何氨基酸: Gly, Ser, Ala, Ile 和 Val

<220>

<221> UNSURE

<222> (9)

<223> Xaa 代表选自下列的任何氨基酸: Leu, Ile, Val, Met, Cys 和 Ala

<220>

<221> UNSURE

<222> (11)

<223> Xaa 代表选自下列的任何氨基酸: Leu, Ile, Val, Met, Phe 和 Ala

<220>

<221> UNSURE

<222> (12)

<223> Xaa 代表选自下列的任何氨基酸: Leu, Ile, Val, Met, Phe 和 Ala

<220>

<221> UNSURE

<222> (13)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<221> UNSURE

<222> (14)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<221> UNSURE

<222> (15)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<221> UNSURE

<222> (16)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<221> UNSURE

<222> (17)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<221> UNSURE

<222> (18)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<221> UNSURE

<222> (19)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<221> UNSURE

<222> (20)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<221> UNSURE

<222> (21)

<223> Xaa 代表选自下列的任何氨基酸: Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr 和 Val

<220>

<212> DNA

<213> 枯草芽孢杆菌 ATCC 15245 和 枯草芽孢杆菌 IAM 1033

<400> 36

```

atg gag aga aaa aca gta ttg gtc atc gct gat ctt gga ggc tgc ccg 48
Met Glu Arg Lys Thr Val Leu Val Ile Ala Asp Leu Gly Gly Cys Pro
  1           5           10           15

ccg cac atg ttt tat aaa agc gct gct gaa aaa tat aac ctg gtc agc 96
Pro His Met Phe Tyr Lys Ser Ala Ala Glu Lys Tyr Asn Leu Val Ser
          20           25           30

ttt att cca aga cct ttt gca att aca gcc tcc cat gca gca ttg att 144
Phe Ile Pro Arg Pro Phe Ala Ile Thr Ala Ser His Ala Ala Leu Ile
          35           40           45

gaa aaa tac tcg gtc gcg gtc ata aaa gat aaa gac tat ttt aag agt 192
Glu Lys Tyr Ser Val Ala Val Ile Lys Asp Lys Asp Tyr Phe Lys Ser
          50           55           60

tta gct gat ttt gag cat cct gat tcc att tat tgg gcg cat gag gat 240
Leu Ala Asp Phe Glu His Pro Asp Ser Ile Tyr Trp Ala His Glu Asp
          65           70           75           80

cat aac aag cct gag gaa gag gtc gtc gag caa atc gtc aag gtt gcc 288
His Asn Lys Pro Glu Glu Val Val Glu Gln Ile Val Lys Val Ala
          85           90           95

gaa atg ttt ggg gcg gat gcc atc aca aca aac aat gaa tta ttc att 336
Glu Met Phe Gly Ala Asp Ala Ile Thr Thr Asn Asn Glu Leu Phe Ile
          100           105           110

gct ccg atg gcg aaa gcc tgt gaa cgt ctg ggc ctg aga ggt gcc ggc 384
Ala Pro Met Ala Lys Ala Cys Glu Arg Leu Gly Leu Arg Gly Ala Gly
          115           120           125

gtg cag gca gcc gaa aat gcc aga gat aaa aat aaa atg agg gac gct 432
Val Gln Ala Ala Glu Asn Ala Arg Asp Lys Asn Lys Met Arg Asp Ala
          130           135           140

ttt aat aag gcc gga gtc aaa tcg atc aaa aac aaa cga gtc aca act 480
Phe Asn Lys Ala Gly Val Lys Ser Ile Lys Asn Lys Arg Val Thr Thr
          145           150           155           160

```

ctt gaa gat ttc cgt gct gct ctt gaa gag atc ggc aca cct ctt atc	528
Leu Glu Asp Phe Arg Ala Ala Leu Glu Glu Ile Gly Thr Pro Leu Ile	
165 170 175	
tta aag cct aca tac tta gcg agt tca atc ggt gta acg ctg att acg	576
Leu Lys Pro Thr Tyr Leu Ala Ser Ser Ile Gly Val Thr Leu Ile Thr	
180 185 190	
gac act gag acg gca gaa gat gaa ttt aac aga gtc aat gac tat ctg	624
Asp Thr Glu Thr Ala Glu Asp Glu Phe Asn Arg Val Asn Asp Tyr Leu	
195 200 205	
aaa tca att aac gtg cca aag gcg gtt acg ttt gaa gcg ccg ttt atc	672
Lys Ser Ile Asn Val Pro Lys Ala Val Thr Phe Glu Ala Pro Phe Ile	
210 215 220	
gct gaa gaa ttt tta cag ggt gag tac gga gac tgg tat caa aca gaa	720
Ala Glu Glu Phe Leu Gln Gly Glu Tyr Gly Asp Trp Tyr Gln Thr Glu	
225 230 235 240	
ggg tac tcc gac tat atc agt ata gaa ggc atc atg gct gac ggt gag	768
Gly Tyr Ser Asp Tyr Ile Ser Ile Glu Gly Ile Met Ala Asp Gly Glu	
245 250 255	
tat ttc ccg atc gcc att cat gat aaa acg ccg caa atc ggg ttt aca	816
Tyr Phe Pro Ile Ala Ile His Asp Lys Thr Pro Gln Ile Gly Phe Thr	
260 265 270	
gag aca tcc cac att acg ccg tcc att ctg gat gaa gag gca aaa aag	864
Glu Thr Ser His Ile Thr Pro Ser Ile Leu Asp Glu Glu Ala Lys Lys	
275 280 285	
aaa att gtc gaa gct gcc aaa aag gca aat gaa ggg ctt ggc ctg caa	912
Lys Ile Val Glu Ala Ala Lys Lys Ala Asn Glu Gly Leu Gly Leu Gln	
290 295 300	
aat tgc gca aca cat aca gag atc aag cta atg aaa aac aga gaa ccg	960
Asn Cys Ala Thr His Thr Glu Ile Lys Leu Met Lys Asn Arg Glu Pro	
305 310 315 320	
ggt tta ata gag tcg gca gcc aga ttc gca ggc tgg aat atg att cct	1008
Gly Leu Ile Glu Ser Ala Ala Arg Phe Ala Gly Trp Asn Met Ile Pro	
325 330 335	

aat att aaa aag gtc ttt ggc ctt gat atg gcg caa tta tta tta gat	1056
Asn Ile Lys Lys Val Phe Gly Leu Asp Met Ala Gln Leu Leu Leu Asp	
340 345 350	
gtc ctc tgt ttc gga aaa gac gcc gat ctg ccg gac gga tta ttg gat	1104
Val Leu Cys Phe Gly Lys Asp Ala Asp Leu Pro Asp Gly Leu Leu Asp	
355 360 365	
caa gag cct tat tat gtt gcc gac tgc cat ttg tac ccg cag cat ttc	1152
Gln Glu Pro Tyr Tyr Val Ala Asp Cys His Leu Tyr Pro Gln His Phe	
370 375 380	
aaa caa aat ggc cag att cca gaa acc gct gag gat ttg gtc att gaa	1200
Lys Gln Asn Gly Gln Ile Pro Glu Thr Ala Glu Asp Leu Val Ile Glu	
385 390 395 400	
gcg atc gat att ccg gac ggg ctt tta aaa ggg gat act gaa atc gtt	1248
Ala Ile Asp Ile Pro Asp Gly Leu Leu Lys Gly Asp Thr Glu Ile Val	
405 410 415	
tca ttt tca gcc gca gca cca ggc act tca gtt gat ttg aca ttg ttt	1296
Ser Phe Ser Ala Ala Ala Pro Gly Thr Ser Val Asp Leu Thr Leu Phe	
420 425 430	
gaa gct ttc aat tcc att gct gca ttt gaa ctg aaa ggc agt aat tca	1344
Glu Ala Phe Asn Ser Ile Ala Ala Phe Glu Leu Lys Gly Ser Asn Ser	
435 440 445	
cag gat gtg gct gaa tca atc aga caa att cag cag cat gca aag ctg	1392
Gln Asp Val Ala Glu Ser Ile Arg Gln Ile Gln Gln His Ala Lys Leu	
450 455 460	
acg gca aag tat gtg ctg cca gta	1416
Thr Ala Lys Tyr Val Leu Pro Val	
465 470	

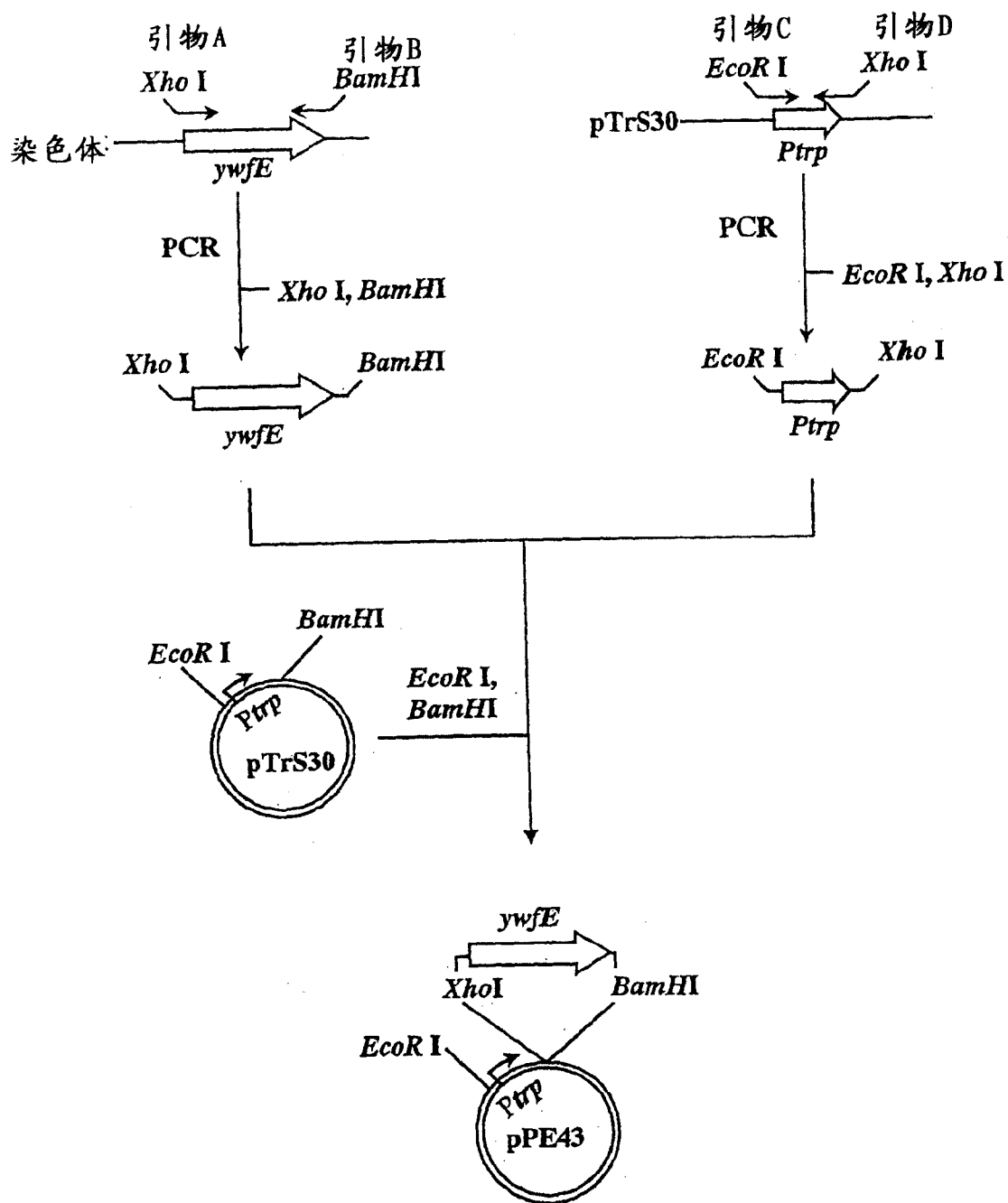


图 1

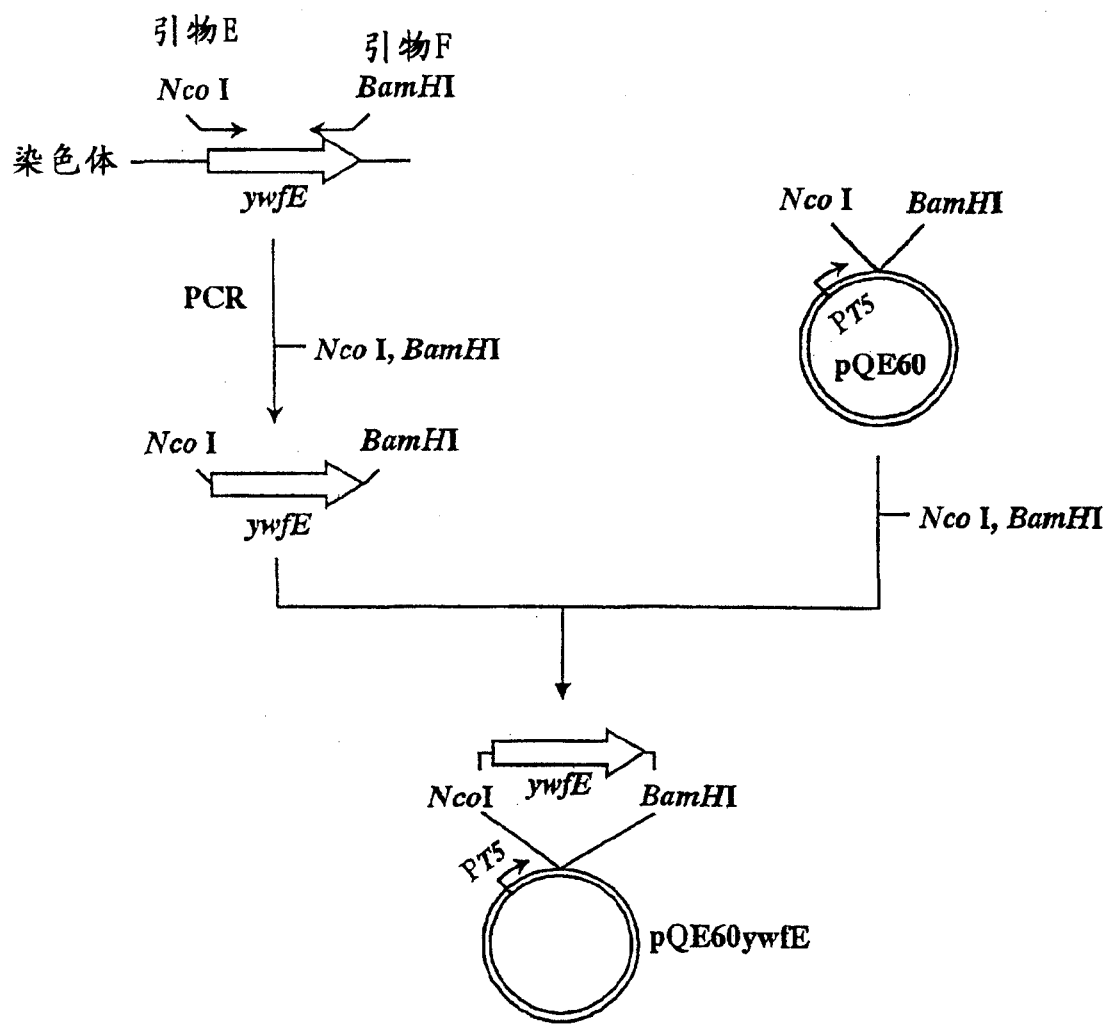


图2